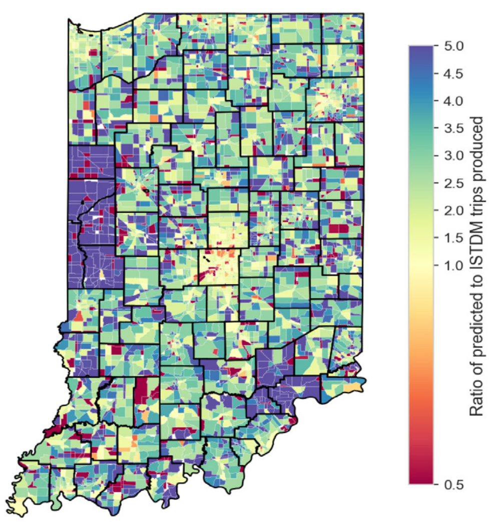
JOINT TRANSPORTATION
RESEARCH PROGRAM

INDIANA DEPARTMENT OF TRANSPORTATION
AND PURDUE UNIVERSITY

Forecasting Shifts in Hoosiers' Travel Demand and Behavior



Rajat Verma, Hao Luo, Saloni Deodhar, Eunhan Ka,
Ricardo Chahine, Pallavi Natu, Harshit Malhotra,
Vineeth Polisetty, Dutt J. Thakkar, Satish V. Ukkusuri, Hua Cai,
Steven R. Dunlop, Ananth V. Iyer, Konstantina Gkritza

SPR-4608 • Report Number: FHWA/IN/JTRP-2023/28 • DOI: 10.5703/1288284317685

RECOMMENDED CITATION

Verma, R., Luo, H., Deodhar, S., Ka, E., Chahine, R., Natu, P., Malhotra, H., & Polisetty, V., Thakkar, D. J., Ukkusuri, S. V., Cai, H., Dunlop, S. R., Iyer, A. V., & Gkritza, K. (2023). *Forecasting shifts in Hoosiers' travel demand and behavior* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2023/28). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317685

AUTHORS

Rajat Verma
Saloni Deodhar
Eunhan Ka
Ricardo Chahine
Graduate Research Assistants
Lyles School of Civil Engineering
Purdue University

Hao Luo
Graduate Research Assistant
Environmental and Ecological Engineering
Purdue University

Pallavi Natu
Graduate Research Assistant
Mitchell E. Daniels, Jr. School of Business
Purdue University

Harshit Malhotra
Graduate Research Assistant
Mitchell E. Daniels, Jr. School of Business
Purdue University

Vineeth Polisetty
Graduate Research Assistant
Mitchell E. Daniels, Jr. School of Business
Purdue University

Dutt J. Thakkar
Associate Project Manager of DCMME
Mitchell E. Daniels, Jr. School of Business
Purdue University

Satish V. Ukkusuri, PhD, F.ASCE
Reilly Professor of Civil Engineering
Lyles School of Civil Engineering
Purdue University

Hua Cai, PhD
Associate Professor
School of Industrial Engineering
Environmental and Ecological Engineering
Purdue University

Steven R. Dunlop
Managing Director of DCMME and GSCMI
Mitchell E. Daniels, Jr. School of Business
Purdue University

Ananth V. Iyer, PhD
Professor of Operations Management and Strategy
University at Buffalo

Konstantina Gkritza, PhD, F. ASCE
Professor of Civil Engineering and
Agricultural and Biological Engineering
Lyles School of Civil Engineering
Purdue University
(765) 494-4597
nadia@purdue.edu

JOINT TRANSPORTATION RESEARCH PROGRAM

The Joint Transportation Research Program serves as a vehicle for INDOT collaboration with higher education institutions and industry in Indiana to facilitate innovation that results in continuous improvement in the planning, design, construction, operation, management and economic efficiency of the Indiana transportation infrastructure. https://engineering.purdue.edu/JTRP/index_html

Published reports of the Joint Transportation Research Program are available at http://docs.lib.purdue.edu/jtrp/.

NOTICE

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/IN/JTRP-2023/28		

4. Title and Subtitle	5. Report Date
Forecasting Shifts in Hoosiers' Travel Demand and Behavior	November 2023
	6. Performing Organization Code
7. Author(s) Rajat Verma, Hao Luo, Saloni Deodhar, Eunhan Ka, Ricardo Chahine, Pallavi Natu, Harshit Malhotra, Vineeth Polisetty, Dutt J. Thakkar, Satish V. Ukkusuri, Hua Cai, Steven R. Dunlop, Ananth V. Iyer, and Konstantina Gkritza	**8. Performing Organization Report No.** FHWA/IN/JTRP-2023/28
9. Performing Organization Name and Address Joint Transportation Research Program Hall for Discovery and Learning Research (DLR), Suite 204 207 S. Martin Jischke Drive West Lafayette, IN 47907	**10. Work Unit No.** **11. Contract or Grant No.** SPR-4608
12. Sponsoring Agency Name and Address Indiana Department of Transportation (SPR) State Office Building 100 North Senate Avenue Indianapolis, IN 46204	**13. Type of Report and Period Covered** Final Report **14. Sponsoring Agency Code**

15. Supplementary Notes

Conducted in cooperation with the US Department of Transportation and Federal Highway Administration.

16. Abstract

The changing landscape of transportation technology and traveler behavior, accelerated by recent events like COVID-19, has led to significant shifts in travel demand and vehicle miles traveled in Indiana. This study seeks to understand the long-term implications of these changes and their potential impact on passenger, freight, and micro-mobility movements across the state. To achieve this objective, this project focused on forecasting future transportation demand conditions and carrying out long-range scenario planning by accomplishing four tasks: forecasting travel demand shifts based on location-based data, evaluating medium-term inter- and intra-urban transportation demand shifts, forecasting county-level industry shifts using scenario-based growth models, and providing recommendations and guidance to the Indiana Department of Transportation (INDOT) based on the study results. Results offer improved planning for infrastructure investments and operations, the incorporation of emerging technologies into transportation planning processes, and an enhanced understanding of passenger and freight movements at the statewide and regional levels. Deliverables from this study include valuable tools and models that can help INDOT navigate potential transportation system changes and accommodate the evolving needs of the future.

17. Key Words	18. Distribution Statement
travel demand, transport planning, travel behavior, large-scale data, freight, e-commerce	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	66 including appendices	

Form DOT F 1700.7 (8-72)	Reproduction of completed page authorized

EXECUTIVE SUMMARY

Introduction

The main goal of this project was to develop long-term future forecasts of transportation demand conditions for people and freight—both conventional and micro-freight (e.g., individual meal or small package delivery). To achieve this goal, this project focused on three objectives.

The first objective was forecasting travel demand to determine which locations in Indiana will experience long-term increases in traffic flows. This was achieved using high-resolution, individual trip-level data derived from the GPS geolocation data of mobile phones. The use of GPS data provided some key advantages over travel surveys, including scalability and temporal variance (longitudinal data), which are important factors in trip generation modeling. The primary outcome of this objective was the generation of origin-destination matrices representing travel patterns for a typical weekday and weekend in three future years—2025, 2035, and 2045.

The second objective involved forecasting and quantifying the travel demand impact of e-commerce on Indiana residents by considering different future adoption scenarios. Using Indianapolis as a case study region, we first conducted a market segmentation analysis for Hoosier's heterogenous shopping behaviors (e.g., frequency) based on the 2017 National Household Travel Survey (NHTS) data and consumer panel data. Subsequently, travel demand changes resulting from e-commerce service (i.e., substitute in-store shopping trips) were integrated into a travel demand simulation model to estimate the net vehicle miles traveled (VMT) impacts. The model accounted for personal travel flow, mode choice, and goods delivery. Different scenarios representing the forward trend of e-commerce service were also analyzed to forecast future travel demand impacts.

The third objective focused on forecasting the growth in freight and passenger traffic at the county level, incorporating macroeconomic and demographic data, such as Gross Domestic Product (GDP) and population, using a scenario-based approach. A regression analysis was employed, using data obtained from several government sources. Additionally, an optimization model was developed to identify the road infrastructure upgrades needed to support the projected growth in traffic while enhancing connectivity and commute. The optimization model used *opensolver* engine, which allowed users to change the input variables and make the model more dynamic.

Findings

This study forecasts a nearly constant overall trip count growth rate over the coming decades, but with spatial variations in this growth pattern. Urban areas, particularly the Indianapolis Metropolitan Statistical Area (MSA) outside of Marion County, are predicted to experience the highest growth. Furthermore, most of the trip flow growth is expected to originate from major urban centers across the state.

The analysis of e-commerce patterns in Indianapolis revealed that different shopper categories correlate with household size, internet access, and education level. Promoting the adoption of e-commerce is expected to yield benefits to the transportation system by reducing personal travel demand, as well as encouraging Hoosiers to choose more sustainable travel modes. Additionally, centralizing delivery services can improve system efficiency by enabling route optimization.

The freight and passenger traffic forecasts at the county level are influenced by county population and industry GDP, and Freight Analysis Framework (FAF) domestic regions, respectively, as revealed through regression analysis. The study identified five industries—manufacturing, construction, retail trade, wholesale trade, and transportation and warehousing—which contribute significantly to freight traffic based on the North American Industry Classification System (NAICS) codes. Specific counties were identified for these five industries using industry attractiveness criteria, such as current industry giants, workforce availability, industry growth potential, and highway proximity. Optimization models were developed for these counties to determine the bottlenecks in road infrastructure that could arise from future traffic growth led by economic activities.

Implementation

This study leveraged a diverse set of data sets and modeling techniques. The overall origin-destination matrices were developed using multiple linear regression-based trip generation and gravity-based trip distribution models. These models were trained on trips extracted from mobile phone geolocation data.

The development of an e-commerce for urban transportation sustainability emphasizes transitioning from in-store to online shopping. Key strategies include building accessible and affordable digital infrastructure to reach traditional shoppers and prioritizing cyber security. The reduction of private car use and the development of green delivery services, including the use of low-emission vehicles, can minimize traffic congestion and emissions. Insights from an agent-based travel demand simulation model can guide infrastructure investment and long-range transportation planning. This approach aims to reduce urban traffic burdens and enhance environmental sustainability through digital innovation and shared green delivery systems.

Based on the regression analysis, freight, and passenger traffic forecasts for up to 2050 across various Shared Socioeconomic Pathway (SSP) scenarios provide key information for INDOT's capacity planning. Identifying counties that are attractive for specific industries enables INDOT to plan for the growth of those industries in said counties and further allows road infrastructure projects to align with the expected growth in industry and the associated freight and passenger traffic. The optimization model uses the traffic forecasts and the industry-attractive counties to determine whether existing capacity is enough to support future growth.

LIST OF ABBREVIATIONS

AADT	Annual Average Daily Traffic
ACS	American Community Survey
BTS	Bureau of Transportation Statistics
DOT	Department of Transportation
FAF	Freight Analysis Framework
FHWA	Federal Highway Administration
FWTDM	Fort Wayne Travel Demand Model
GDP	Gross Domestic Product
GPS	Global Positioning System
HPMS	Highway Performance Monitoring System
INDOT	Indiana Department of Transportation
ISTDM	Indiana Statewide Travel Demand Model
ITTTS	Integrated Traditional and Transformative Transportation System Use Model
LBS	Location-Based Services
LCA	Latent Class Analysis
MPO	Metropolitan Planning Organization
MSA	Metropolitan Statistical Area
MSE	Mean Squared Error
NHTS	National Household Travel Survey
ODM	Origin-Destination Matrix
SEA	Socioeconomic Attributes
SSP	Shared Socioeconomic Pathway
STDM	Statewide Travel Demand Model
STEEP	Social, Technical, Economic, Environmental, and Political
TAZ	Traffic Analysis Zone
TDM	Travel Demand Model
US	United States
USDOT	United States Department of Transportation
VMT	Vehicle Miles Traveled
VRP	Vehicle Routing Problem
VRPPDTW	VRP with Pick-up/Drop-off Demand and Time Window Constraints
ZINBR	Zero-Inflated Negative Binomial Regression

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The main goal of this project is to develop long-term future forecasts of transportation demand for people and freight—both conventional and micro-freight (e.g., individual meal or small package delivery) to inform the Indiana Department of Transportation (INDOT). This would ultimately allow the agency to (1) properly adjust its business practices, planning models/applications (e.g., Indiana Statewide Travel Demand Model or traffic microsimulation), and investment decisions to effectively cater to future transportation needs, and (2) execute long-range scenario planning.

1.1 Background and Motivation

Transportation planning is a complex yet routinely exercised activity to estimate the state of traffic distribution in a region (typically a city or a neighborhood). This is especially useful to assess the impact of new infrastructure, such as a new transit line, prominent constructions, or residential/commercial complexes on regional traffic congestion, vehicular pollution, and other transportation externalities. Transportation planners have long relied on a four-stage planning scheme to solve this complex problem, though newer methods, such as activity-based planning, are also gaining popularity in recent years.

The first two stages of traditional four-stage transportation planning involve (1) trip generation: the estimation of the total traffic demand generation and attraction in all the traffic analysis zones (TAZs) of a region, and (2) trip distribution—the splitting of that demand into each interzonal trip interchanges. Traditionally, the first step of trip generation is conducted using land use and employment spatial distribution from sources like population surveys, such as the United States (US) decennial Census or the American Community Survey (ACS). Similarly, the second stage of trip distribution generally involves the solution of the matrix balancing problem where interzonal trip interchanges are identified given the total trip generation and attraction in each zone. This problem is typically solved using simple models, such as the gravity or the radiation model.

While the traditional models are sufficient approximations of trip origin and destination assignment using limited data, researchers have recognized the need for more realistic data sources that rely on high quality and quantity data of human movement. Traditional land use-based models do not capture mobility patterns as effectively because of blanket assumptions based on land use. This is particularly important in the context of long-term transportation planning since a less careful assessment of the current mobility distribution can lead to significantly erroneous predictions in the long future.

Human movement patterns, specifically commute and routine travel patterns, are typically observed on large scales using travel surveys, such as Federal Highway Administration (FHWA)'s National Household Travel Survey (NHTS). While travel surveys capture a reasonable number of high-quality responses with detailed travel diaries, they are quite expensive to undertake and are not easily scalable.

In this context, the emergence of large-scale mobile phone geolocation data has become notable. The sharp rise of smartphone penetration in the recent years has led researchers to use location-based service (LBS) data to meet various ends such as estimating traffic demand variation, designing customized services, and analyzing long and short-term human mobility patterns within and among cities. Such data, usually provided in the form of anonymized timestamped geographic coordinates from Global Positioning System (GPS) devices embedded in smartphones, provide quantitative information about mobility at the expense of personal information, such as demographics and travel preferences. Since such geolocation data provide an easy way to obtain large-scale movement patterns over time, they help overcome the main limitations of the data used in traditional trip generation and distribution modeling.

In addition to emerging data sources in the US, the last few years have seen significant changes in travel demand due to adoption of new transportation technologies (automation, shared vehicles, micro-mobility, etc.) and in traveler behavior. These changes have been accelerated due to COVID-19 with work from home policies, online shopping, and curb side pick-up/deliveries. In specific, in the US, e-commerce sales rose to \$870 billion in 2021, a 14.2% increase over 2020 and a 50.5% increase over 2019. E-commerce represented 13.2% of all retail sales in 2021 (Goldberg, 2022). Additionally, the pandemic had a devastating impact on retail sales and store closures. In 2020, 5,463 net retail stores closed in the US and retailers like Nordstrom saw 54% of their sales happening online (Sapling, 2021). The growing demand for e-commerce delivery is set to result in 36% more delivery vehicles in inner cities by 2030, leading to a rise in both emissions and traffic congestion (Handy et al., 2013). However, the propensity of individuals to use online delivery services from a transport perspective is not well understood to date. A study on understanding user preferences and usage of such services would inform the modeling of micro-freight delivery flows and allow for such estimates to be included in current and future policy decisions.

These changes in traveler and shopping behavior not only affect intra urban flows but also statewide travel patterns. Indiana, known as the Crossroads of America, has transformed into a global center of transportation and logistics for the 21st century. Indiana's robust multi-modal transportation infrastructure is critical to provide companies with a competitive advantage in manufacturing and distribution. Indiana has been recorded as the fifth busiest state for commercial freight traffic as it serves regional, national, and international markets due to its strategic location. Indiana Department of Transportation (INDOT) is responsible for maintaining and regulating transportation and transportation-related infrastructure in the state. INDOT plays a significant role in enabling Indiana's industry level competitiveness by supporting freight logistics across Indiana through multi-modal freight systems. According to the agency's 2019 Strategic Plan, INDOT is responsible for more than 4,000 freight railroad miles, over 110 public access airports and maintains more than 11,000 centerline miles and observes more than 1.5 billion tons or $495 billion worth of annual freight movement throughout Indiana. INDOT seeks guidance regarding the medium (decade out) and long-term (25 to 35 years out) shifts in the way people and goods move in and across our state, by whatever modes are present today or can be reasonably expected to be relevant 10 to 35 years in the future.

1.2 Objectives and Deliverables

This project develops a suite of modeling and analysis tools for forecasting transportation demand conditions for people and freight—both conventional and micro-freight (e.g., individual meal or small package delivery) that can support INDOT to understand and plan for the potential transportation system changes due to shifts in passenger and freight movements. INDOT's understanding of likely or potential future fundamental changes in travel demand, both people and freight, may substantially affect how the agency adjusts practices and infrastructure investments to serve the related transportation needs. That matters today given the manner we plan and design those services are often based on the predicted environment 20 and 30 years into the future rather than current conditions. The results of this study can be useful inputs to INDOT's business practices, planning models/applications, and investment decisions. The results of this study can also help metropolitan planning organizations (MPOs) in the study areas to anticipate the demand for passenger, freight, and micro-mobility flows. In specific, the quantification of the impacts associated with shift in passenger, freight and micro-mobility flows can help INDOT to plan accordingly for existing and forthcoming funding needs for operations and infrastructure investment. Additionally, a better understanding of the use and impacts of emerging transportation options in Indiana for statewide and regional passenger and intra-urban freight movements can facilitate the incorporation of these emerging technologies into the Transportation Planning Process and the long-range transportation plan. The objectives of this project are encompassed in the following three tasks.

1.2.1 Forecasting Travel Demand Using Geolocation Data

The main objective of this task is to predict inter-zonal trip interchanges in Indiana in future decades up to 2050. These include both passenger and freight trips using all travel modes along roads, including cars, trucks, buses, and walking. The main deliverable of this task in this project is a set of six origin-destination matrices (ODMs) separately for a representative weekday and weekend of three future years—2025, 2035, and 2045. Each ODM is a square matrix with cell $A_{i,j}$ representing the number of predicted trips from a traffic analysis zone (TAZ) i to TAZ j for all the n TAZs of Indiana. The three target years are selected specifically because of the presence of reasonable estimates of the influential factors of the TAZs in only these future years based on data from the current Indiana Statewide Travel Demand Model (ISTDM). To this end, raw geolocation data from LBSs on smartphones are used to estimate travel patterns. Chapter 2 provides the description of the data generation and processing, the modeling procedure for trip generation and distribution, and the interpretation and discussion of the projection results.

1.2.2 Estimating the Impact of E-Commerce on Travel Demand

This task aims to understand and forecast the impact of e-commerce's future adoption on travel demand. We extend the Integrated Traditional and Emerging Transportation System Use Model (ITTTS) developed for Indianapolis as part of Luo et al. (2022) to evaluate the medium and long-term transportation pattern changes and the associated vehicle-miles-travelled (VMT) and vehicle ownership change. The model that is being developed in Luo et al. (2022) has the capability to simulate transportation mode choice considering not only the traditional transportation modes (e.g., private vehicle, bus, taxis, walking/biking) but also the emerging shared and active transportation mobility options (e.g., shared bikes, shared e-scooters, ride hailing). This task enhances the modeling capability to also include freight in terms of food and products delivery services to evaluate the substitution of individual consumers' shop-visit trip/leg in the trip chain being replaced by multi-stop trip chains of the delivery vehicle/trucks and shifting VMTs from passenger vehicles miles to delivery vehicle/truck miles.

The expanded model has the capability to perform scenario analysis for medium and long-term scenarios with different levels of e-commerce adoption, and travel behavior changes. The model is calibrated using the NHTS data to reflect the mode choice and travel behavior shifts (dynamic demand and trip patterns) that have been observed. Scenario analysis results on personal and delivery travel demand can inform the trip generation module of ISTDM.

1.2.3 Forecasting County-Level Shifts Based on Scenarios

The objective of this task is to employ a scenario-based approach to understand possible evolutions of future industry growth by county, through 2050, using accepted scenario-based growth models. Using a county level approach, we define the parameters that impact the freight and passenger traffic and use globally accepted forecasts of these parameters to forecast the growth in traffic in Indiana at county level. We then focus on identifying potential industry desirable locations, and evaluation of their attractiveness on a variety of industry relevant dimensions.

1.3 Report Organization

The subsequent sections of the report delve into the three main tasks of this study. Chapter 2 describes the first task "forecasting travel demand using geolocation data," encompassing an overview (Section 2.1), literature review (Section 2.2), data description (Section 2.3), methodology (Section 2.4), results (Section 2.5), and a chapter summary (Section 2.6). Similarly, Chapter 3 presents the second task "estimating the future impact of e-commerce on travel demand," comprising an overview (Section 3.1), literature review (Section 3.2), data description (Section 3.3), methods (Section 3.4), results (Section 3.5), and a chapter summary (Section 3.6). Chapter 4 then describes the third task "forecasting county-level shifts based on scenarios," encompassing an overview (Section 4.1), literature review (Section 4.2), data description (Section 4.3), methods (Section 4.4), optimization model (Section 4.5), and a chapter summary (Section 4.6). Lastly, the concluding Chapter 5 synthesizes key findings across the tasks, offering insights for implementation. The specific strategies and recommendations from each task are detailed in Sections 5.5.1, 5.5.2, and 5.5.3. Additionally, the report includes a references section and an extensive appendix (Section A) offering supplementary details related to each task.

2. FORECASTING TRAVEL DEMAND USING GEOLOCATION DATA

2.1 Overview

In this chapter, we focus on the first two steps of the four-step planning model, namely trip generation and trip distribution. Trip generation via GPS data involves inferring the trips made by the mobile device users and modeling their relationship with socioeconomic attributes (SEAs) of the TAZ of their home location estimated from their GPS traces. The forecasts of the SEAs are considered as inputs that are typically modeled based on demographic and economic growth models. The inferred trips are also used to compute the ODM of Indiana. The developed trip generation model is used to predict the trip production and attraction patterns of the TAZs in the future, while a trip distribution model is used to predict the future trip interchanges between the TAZs based on these trips. Further, the predicted travel patterns are analyzed and validated using both the (currently travel survey-based) ISTDM as well as a traditional four-step model for the Fort Wayne metropolitan area of Indiana. In doing so, we show how large-scale mobile phone GPS data can be used to refine the four-step conventionally used in the ISTDM, including using observed trip ODMs and interzonal costs in terms of travel times rather than Euclidean distances. We also discuss the limitations of this approach and provide directions for integrating it with the traditional four-step planning process.

The rest of the chapter is organized as follows. The latest relevant research in regard to four-step planning and GPS data in transport planning is reviewed in Section 2.2. The GPS and the ISTDM data are described in Section 2.3. The methods of the data processing and the trip generation and distribution modeling are shown in Section 2.4. Their results and their implications are discussed in Section 2.5. A chapter summary is provided in Section 2.6.

2.2 Literature Review

2.2.1 Travel Demand Forecasting

Long-term travel demand forecasts assume significant importance, undergirding strategic planning and investment decision-making procedures (He & Hong, 2018). It has traditionally relied on data sources like household travel surveys which often require significant financial, human, and time investments for thorough data collection (Cheng et al., 2020; Cools et al., 2010). The emergence of location-based data given its ubiquity and substantial volume, has significantly bolstered our capacity to conduct expansive spatial and temporal pattern analyses within urban environment (Hasan & Ukkusuri, 2014; Wang et al., 2021). This type of high-resolution data source presents new opportunities for comprehending and predicting mobility patterns, thereby refining the precision of travel demand forecasting.

2.2.2 Four-Step Planning Models

The four-step planning model is a commonly used methodology in transportation planning, which involves the following four stages: trip generation, trip distribution, mode choice, and trip assignment. Trip generation, the first step in the four-step planning model, calculates the number of trips in a specific area based on socio-demographics, employment, and land use, using statistical models. Trip distribution, the second step, predicts the volume of trips between different origin-destination pairs. This involves modeling travel patterns and estimating trip interchanges ("flows") between zones. It is often based on the gravity model introduced by Wilson (1967), which accounts for factors like distance and attractiveness of destinations. The third step, mode choice, determines the transport mode for each trip considering travel time, cost, convenience, and personal preferences. McFadden et al. (1973) introduced the conditional logit model, which is widely used to model mode choice behavior. Finally, traffic assignment specifies the routes that travelers will take, and the traffic volumes thus generated on each network link based on Wardrop's user equilibrium principle (Wardrop, 1952), which suggests travelers choose cost-minimizing routes.

2.2.3 GPS Data in Transport Planning

Travel surveys are extensively utilized but come with certain limitations, such as high cost, non-response rates, low sample sizes and data inaccuracies. In this context GPS technology emerges as a potential alternative offering more reliable and precise data (Sarmiento et al., 2013).

Chung and Shalaby (2005) proposed a trip reconstruction tool for GPS-based personal travel surveys, which extracts travel information such as trip purpose, mode of transportation, and destination from GPS data (Chung & Shalaby, 2005). Bohte and Maat (2009) presented a methodology for deriving and validating trip purposes and travel modes from multi-day GPS-based travel surveys (Bohte & Maat, 2009). The authors posited that GPS data can yield more accurate and comprehensive travel data than traditional survey methods, but reliable methods for processing and analyzing the data are crucial. Hong et al. (2021) provided insights on data quality from a large-scale application of smartphone-based travel survey technology in the Phoenix metropolitan area (Hong et al., 2021). They showed the importance of data completeness, accuracy, and representativeness, and highlighted challenges associated with smartphone-based travel surveys.

Based on previous studies, we summarize the advantages of using GPS data in transport planning as follows: (1) GPS data offer comprehensive coverage, with substantial representation of the population (Stopher et al., 2008), (2) these data are available at large scales and relatively low cost (Hong et al., 2021), and (3) the sufficient temporal variability and spatial density of GPS data enable more granular inferences regarding a variety of attributes such as trip chains, trip purpose, and travel modes with reasonable confidence (Bohte & Maat, 2009; Shen & Stopher, 2014; Zou et al., 2022). However, GPS data also bear several limitations, such as (1) data processing challenges, including requiring careful selection of methods, as well as their parameters and computing resources (Bohte & Maat, 2009; Chung & Shalaby, 2005) and (2) the need for validation through more direct methods, such as travel surveys (Hong et al., 2021).

Though GPS data are increasingly used in large-scale travel demand models, such as STDMs, this effort is not evenly distributed across the states. This task is poised to illustrate the incorporation of GPS data in travel demand prediction and advance the field of large-scale geolocation data analysis for transportation planning.

2.3 Data Description

There are two main datasets used in this task: socioeconomic data layers and geolocation data, which are described in the subsequent sections. Additionally, other publicly available geographic datasets are used and described accordingly.

2.3.1 ISTDM Socioeconomic Data

The ISTDM (Fricker, 2007), developed in collaboration with INDOT, is used to forecast future travel patterns in Indiana. It is a comprehensive modeling system that integrates various data sources, such as demographic, land use, and transportation data, to develop an understanding of travel patterns in Indiana. This ISTDM is based on a four-step planning model. The demographic and economic factors are provided for 4,726 TAZs within Indiana.

The ISTDM is used as a reference for both the TAZs considered as well as the socioeconomic features used in the trip generation models. ISTDM data were provided by an officer at INDOT upon request. These include four geographic tables for the years 2015, 2025, 2035, and 2045 in the form of shapefiles which includes 4,867 TAZs based on the US Census Bureau, of which 4,726 lie within Indiana's boundaries. Since this task considers predicting only the traffic demand distribution, only Indiana's TAZs are considered. For reference, the boundaries of these TAZs are compared with those of counties, census tracts, block groups, and zip code tabulation areas (ZCTAs) in Figure A.1. The variables present in the ISTDM data are shown in Table 2.1 and explained in (Fricker, 2007).

2.3.2 Geolocation Data

The geolocation data are provided to us by Quadrant Inc. These data contain GPS records (hereby referred to as "pings") for the month of March in the years 2019, 2020, and 2021. Though the methods described in the subsequent sections of this chapter are applicable with even a few days' worth of data, we used all the three years' 1-month data to reduce the probability of regression to the mean of the daily mobility patterns and increase the robustness of our estimates. The geolocation data are generally distributed relatively evenly across the 31 days of March in the three study years (see Section A.1.2).

The rows in each day's table have five pertinent columns, and a sample is shown in Figure 2.1.

- device_id: Anonymized alphanumeric code depicting a unique device ID.
- lat, lon: Geographical coordinates (latitude and longitude) of the ping in degrees.
- timestamp: Unix-style timestamp (number of milliseconds since 1/1/1970 UTC 00:00:00.000).
- error_radius: approximate radius of error (in meters) in the location of the ping as estimated by the number of satellites detected and the strength of the GPS signal.

2.4 Methodology

2.4.1 Framework

The central idea in this task is to use geolocation data to estimate current travel patterns and use relevant

TABLE 2.1
Description of the variables in the original ISTDM data tables

Name	Description	Summary Statistics			
TAZ_2010	Geographic identifier (geoid) of the TAZ as per the US Census Bureau 2010 classification	Total 4,726 TAZs within Indiana			
COUNTY	Name of the enclosing county	All 92 counties of Indiana (141 counties in surrounding states of IL, KT, MI, and OH removed from the analysis)			
NEW_AT	Rurality class of the TAZ	Rural: 2,743 TAZs Suburban: 1,155 TAZs Urban: 828 TAZs			
		Max.	**Median**	**Mean**	**Std. dev.**
AGG_VEH	Total no. of vehicles owned by the population	13,225	389	975.04	1,587.54
POP	Population of TAZ	19,069	499.5	1,389.42	2,393.03
GQPOP	Population living in government quarters	10,826	0	39.55	302.14
HH	No. of households	7,762	188	528.91	923.54
HHPOP	Household population	18,942	491	1,349.87	2,319.99
MEANHHINC	Mean household income (US $)	199,989	59,561	61,906.67	18,646.07
WORKERS	Total no. of workers	10,073	233	650.71	1,127.87
TOTEMP	Total employed people (may not be same as workers)	42,257	120	759.74	1,931.05
		Total Employed People in These Sectors			
AGCON	AGCON: agriculture	6,267	12	47.76	145.07
INDUST	INDUST: industry	8,525	17	165.14	507.71
RETAIL	RETAIL: retail	4,500	7	95.12	280.69
FOODLD	FOODLD: food and lodging	4,245	1	61.40	189.53
PROSRV	PROSRV: professional services	35,200	20	264.29	971.86
GOVNMT	GOVNMT: government	11,080	0	39.16	246.91
OTHSRV	OTHSRV: other services	5,415	12	86.87	264.77

	user_id	lat	lon	error_radius	timestamp
0	7ae54ad5-4b58-4bf7-ab7d-e75913dd66b4	37.909910	-87.329320	9045.0	1551447494000
1	a2f60c7f-599b-4a51-90f5-6a3ed39f0121	41.512287	-87.253174	10.0	1551456710000
2	dc26b24f-9e09-4d51-8f56-1859ae52decc	38.303307	-86.948150	15.0	1551409291000
3	7f06cda6-93ec-4b7b-b023-3a73ab090650	41.136870	-85.199340	15.0	1551445799000

Figure 2.1 A sample of the smartphone geolocation data table.

socioeconomic attributes in the current and the future years to estimate the travel patterns (in the form of ODMs). This process is illustrated in Figure 2.2 and described in the following sections.

After the smartphone data is collected, they are first cleaned and are then used to estimate the device owners' home zone and identify trip segments, where each segment represents the reconstruction of a person's trip by any single travel mode. The origins and destinations of these trips are used to generate ODMs for different base years, segregated by day type. These constitute the current ODMs, called OD_{now}. By aggregating the trip count by origin and destination zones using ISTDM TAZ classification, the current trip attraction and generation are computed, referred to as D_{now} in Figure 2.2. Patterns from these trips are visualized. Then, the current socioeconomic data attributes of ISTDM zones, X_{now}, are used to fit a set of trip generation models by day type.

2.4.2 Data Collection and Processing

The GPS data described in Section 2.3.2 are first filtered for high accuracy. This is important as spatially inaccurate pings act as noise in the downstream tasks of home/work location estimation and trip segmentation. For this purpose, all records with a GPS spatial error radius of more than 50 m are removed.

Next, only high-quality device users are filtered so that there is reasonable confidence in the subsequent steps. To achieve this, a double temporal frequency matrix is constructed for each base year (2019–2021), based on the work of Handy et al. (2013). An example frequency matrix is shown in Figure 2.3 which shows

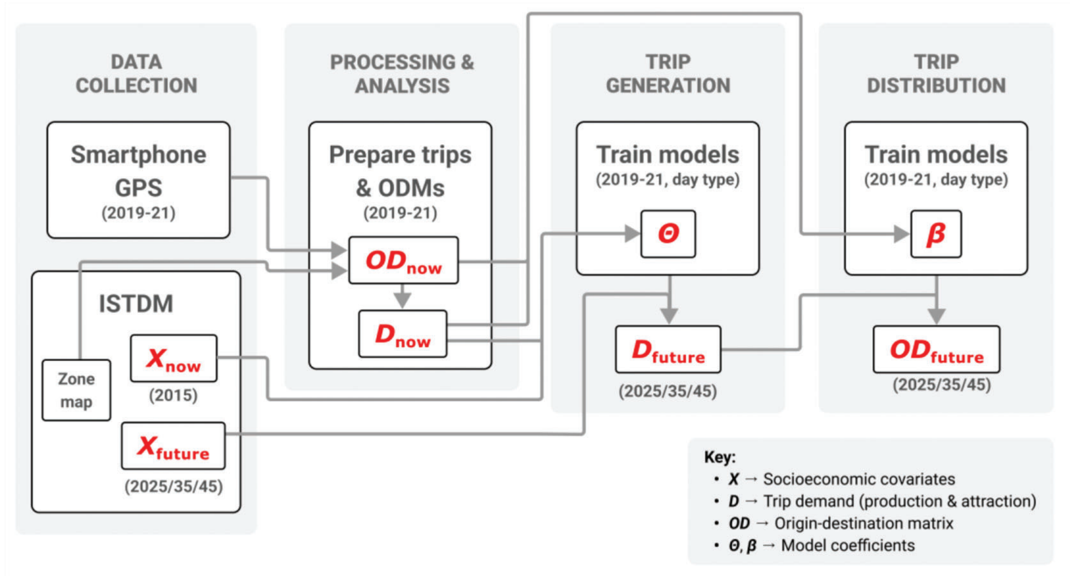


Figure 2.2 Overall framework for trip forecasting using mobile phone data.

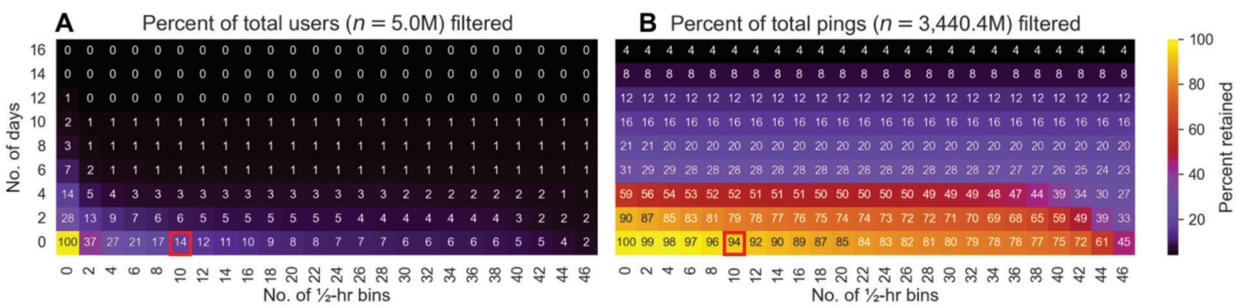


Figure 2.3 Data quality-quantity tradeoff matrices for the 2021 geolocation data showing: (A) the number of users filtered by applying different quality thresholds, and (B) the amount of ping data retained from the users filtered in panel A.

the distribution of the number of filtered users in Indiana in 2021 for different combinations of frequency thresholds. In panel A, a cell $A_{i,j}$ represents the percentage of users having at least one GPS ping in at least j half-hour periods (at most 48) on at least i days within March of a given year. Similarly, in panel B, the cell $A_{i,j}$ represents the percentage of total pings retained from the users filtered for the corresponding cell in panel A. This dual filter criterion ensures that not only users with more data are filtered, but also that those data are not concentrated in just one short period over those days.

This is illustrated in Figure 2.3 where the total number of users and pings are shown in the panel titles. For example, the cell $A_{10,0}$ highlighted in red indicates that 14% of the total unique users (i.e., $\approx$ 700,000) have at least one ping in more than 10 half-hour bins/periods on more than zero (i.e., at least one) day. These 14% users, however, contain about 94% of the total pings of the 2021 dataset (see the cell $A_{i,j}$ in panel B. As a general pattern similar to the Pareto principle, some high-quality users contain most of the data, as is evident in the difference of the gradients of the two panels. To make sure all the subsequent analysis relies on high-quality data, only the users included in the cell $A_{10,0}$ are filtered.

2.4.3 Data Representativeness Using Home Detection

Once the data has been filtered, it is important to compute their representativeness to see if there are any specific regions which show problematic representation of or bias in its population in the dataset (Grantz et al., 2020). In this project, data representativeness for a region is defined as the ratio of the number of users whose homes are detected in that region to that region's population as of the 5-year estimate of the ACS of that year.

For home location detection, which is a common technique used in GPS geolocation data analysis for human mobility, we use an adaptation of the method described in (Yabe et al., 2022). In summary, this method involves finding the centroid of the largest cluster of pings observed during the nights (considered as 9 PM–6 AM in this project), with the assumption that people are typically most likely to be found at their homes during the nighttime. For this task, mean shift clustering (Comaniciu & Meer, 2002) is used, with the main parameter being the radius of a flat kernel, set to 200 m (656 ft).

The representativeness figures obtained in this way are shown in Figure 2.4. The values are computed for two scales—(a) counties and (b) TAZs using the unfiltered dataset of 2020. At the county level, the unfiltered dataset includes information about 20% to 40% of the population, which is sufficient for most of the analysis even after considerable filtering (Erlich et al., 2018). This variance increases significantly at the TAZ level, with values ranging from just 10% to about 90% of the population.

For this analysis, these values are computed for all the TAZs. Then, they are used as simple scaling factors for upscaling the number of computed trips in the OD matrices as explained below.

2.4.4 Trip Segmentation and OD Matrix Preparation

The main step in this task is to identify trips, whose process is illustrated in Figure 2.5. First, the GPS records of only the high-quality users (as described in the previous section) are filtered. Next, spatiotemporally close points are clustered together to create "stay regions" that serve as potential trip endpoints, based on the algorithm presented in (Sadeghinasr et al., 2019). To do this, the daily records of each user are first spatially clustered using mean shift clustering with the same parameters as used in home detection in the previous section. Then, these are further divided into smaller clusters based on temporal discontinuity to make sure that the generated virtual regions are spatiotemporally close. Subsequently, these regions are labeled as either "stay" or "enroute" based on a minimum dwell time parameter (taken as 30 minutes in this study (Li et al., 2008). Virtual regions, for which the time difference between their earliest and latest ping is greater than this minimum dwell, are considered "stay regions," while the remaining ones are labeled "enroute." Finally, a trip is generated between each pair of consecutive stay regions such that it includes the latest ping of the sequentially earlier stay region, the earliest ping of the subsequent stay region, and all the pings of all the enroute regions in between, if any. In the illustration in Figure 2.5, for example, a trip starts from point 3 of the first stay region, ends at point 7, the first point of the next stay region, and includes all the points in between.

Once the trip segments are generated, their endpoints (origin and destination) are used to compute the total number of trips to and from different TAZs. These are then upscaled using the representativeness figures as explained in the previous section.

2.4.5 Trip Generation

A linear regression-based trip generation model is developed in this task to predict the production and attraction of home-based and work-based trips. To compute the attractiveness of the TAZs, the socioeconomic and land use indicators of the zones are assessed. The predictor variables are computed from the measures available in the ISTDM shapefiles for the four study years—2015, 2025, 2035, and 2045.

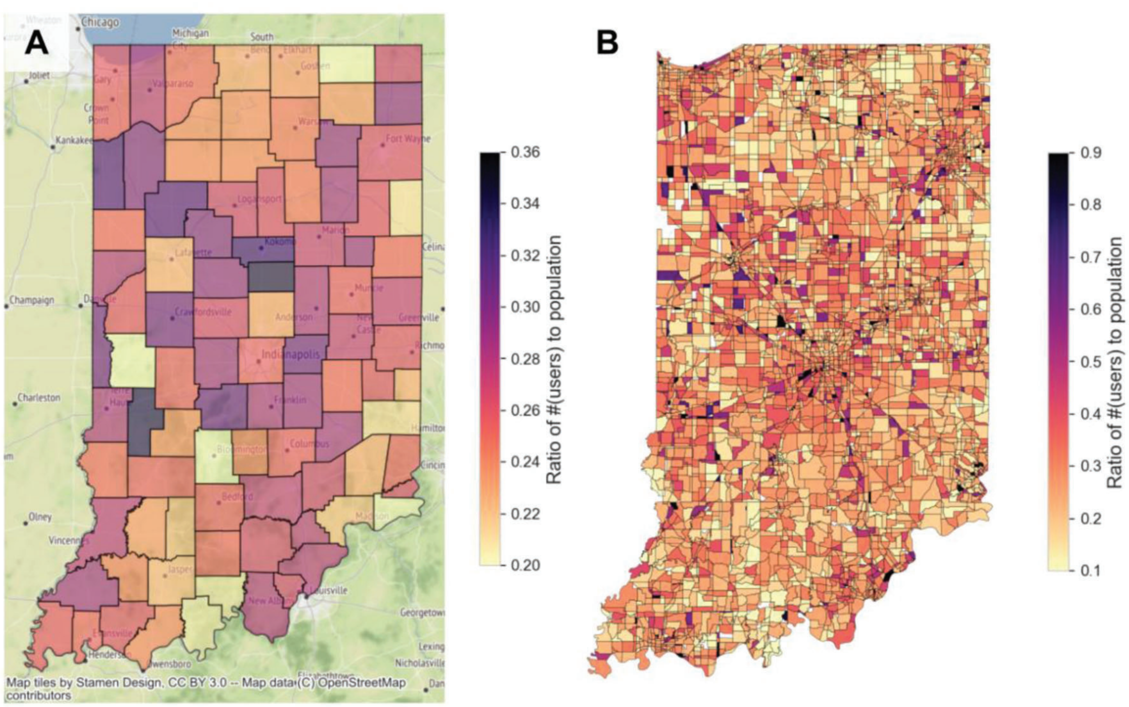


Figure 2.4 Representativeness values in Indiana at the (A) county level, and (B) TAZ level in 2020.

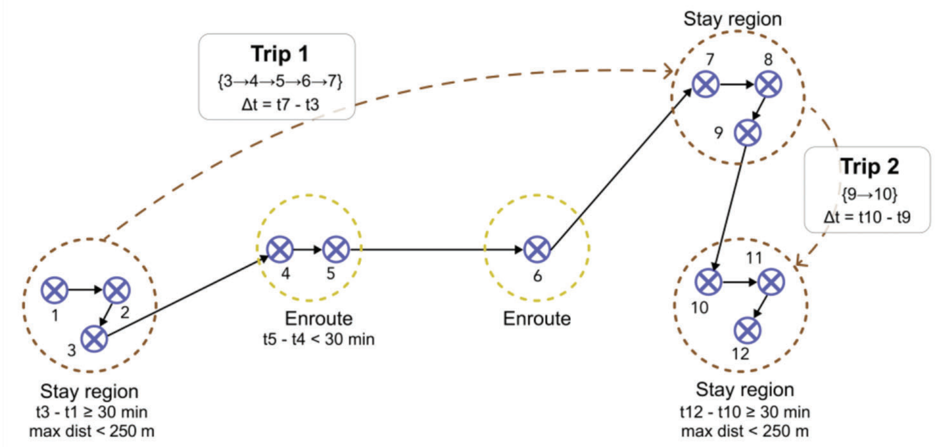


Figure 2.5 Overview of the trip segmentation process.

These variables are computed so that they have minimal correlation among themselves. Here, population density is highly correlated to the overall population of the TAZs ($\rho = 0.61$). All the other variable pairs have a correlation with a magnitude less than 0.5 and are thus considered safe for inclusion in the subsequent trip generation model.

2.4.6 Trip Distribution

The main purpose of a trip distribution model is to estimate an OD matrix from a set of row and column sums, given by the total zonal production and attraction obtained via the trip generation model. The gravity model is a traditional trip distribution model which assumes a negative exponential dependence of trip interchange probability on the impedance between two zones (Duffus et al., 1987). In this task, a variant of the standard gravity model is used, given by the following equation:

$$\hat{N}_{ij}(\mathbf{P},\mathbf{A},\mathbf{D}|\beta) = \frac{A_j D_{ij}^{-\beta}}{\sum_{k=1}^{n} A_k D_{ij}^{-\beta}} \cdot P_i \,\forall\, i,j \in 1:n \qquad \text{(Eq. 2.1)}$$

Here, $\hat{N}_{ij}$ is the number of predicted trips between zones i and j, P_i is the number of trips produced in zone i, A_j is the number of trips attracted by zone j, D_{ij} is the friction factor (alternatively called the impedance or cost functions) between zones i and j, n is the total number of zones, β is the gravity exponent, the only parameter of the model. This model does not explicitly consider the socioeconomic adjustment factor between each pair of zones which is otherwise included in some implementations of the gravity model (Duffus et al., 1987).

The computation of the friction factors is contextualized to suit the given geolocation data in this project. Instead of using Euclidean distances between two zones (or a nonvariant equivalent conversion to travel time) as the measure of friction factor, we use the mean or median of the distance or travel time values of all the trips made in the study period between any two zone pairs. This method is better than the static conversion in that it accounts for the actual path of the trip along the roadway instead of the as-the-crow-flies path and also, the temporal heterogeneity in trip-making behavior, such as weekdays versus weekends and peak and off-peak hours. Thus, four cost functions are tested in the model calibration process—(1) mean travel time, (2) median travel time, (3) mean trip length, and (4) median trip length of the trips obtained in the previous step.

Once the data for the gravity model are obtained, the only parameter of the model—the gravity exponent, β, is calibrated using line search optimization. Its estimator is given by the following equation:

$$\beta = \arg \min_{\beta \in [\beta_0, \beta_1]} \|\mathbf{N} - \widehat{\mathbf{N}}(\beta|\mathbf{P},\mathbf{A},\mathbf{D})\|_2 \qquad \text{(Eq. 2.2)}$$

Herein, $\mathbf{N}$ is the observed OD matrix as per the computed trip endpoints, $\widehat{\mathbf{N}}$ is the fitted OD matrix for a given value of β with the computed trip generation vectors ($\mathbf{P}$ and $\mathbf{A}$) and the friction factors ($\mathbf{D}$). The L_2 matrix norm $\|\cdot\|$ computes the mean squared error (MSE) of the flattened input matrix. The parameter test range $[\beta_0, \beta_1]$ is set to [0.5, 3.0] based on the accepted range in the literature (Black, 1973; Erlander & Stewart, 1990).

2.5 Results

2.5.1 Trip Generation

Trip generation models for trip production and attraction are developed as a set of linear regression models for a typical weekday and weekend at different spatial scales for different years. These models primarily make use of socioeconomic characteristics that have already been predicted for the future years in the ISTDM. The factors listed in Equation 2.2 are normalized with respect to the population to make them comparable across the different regions of comparison (e.g., counties or ISTDM TAZs) and converted into the following 12 explanatory variables: (1) total population, (2) population density, (3) percentage of population living in government quarters, (4) average household income, (5) average number of vehicles, (6) percentage of employed population, (7) percentage of employer in agriculture sector, (8) percentage of employer in industry sector, (9) percentage of employer in retail sector, (10) percentage of employer in food and lodging sector, (11) percentage of employer in professional services sector, and (12) percentage of employer in government sector.

The correlation half-matrices for these variables at the county and TAZ-level are shown in Figure 2.6. Except for total population and population density, the descriptive variables do not have significant correlation since the magnitudes of correlation coefficients are less than 0.5. This correlation is particularly high in the case of TAZs because unlike the Census-delineated regions, which focus on creating regions based on the uniformity of population distribution, TAZs typically focus on creating more uniformly distributed land areas. This results in large TAZs even in densely populated regions, causing a strong correlation between total population and population density. To avoid the issue of multicollinearity in the subsequent models, we did not include population density in the models.

The choice of spatial scale of modeling and the year of data to be used for modeling is considered important in the analysis. Therefore, for selecting the optimal trip generation model, we conducted a scenario analysis of the coefficients of the model for different combinations of years and spatial scales. The results of the resulting coefficients are provided in Figure A.4. Among these models, the ones for the year of 2021 at the TAZ scale were selected. This is because the models based on the 2019 data have low fitness (R^2) values and there might be possible confounding with the 2020 data because of partial substantial mobility pattern disruptions during the month of March due to COVID-19 (Verma et al., 2021).

The coefficients of the covariates and the model fitness statistics for the four finalized models are shown in Figure 2.2. Across all the models, population is one of the most important predictive factors of trip counts. This is consistent with many existing trip generation models (Hooper, 2017). Also, all the other covariates have a strong positive impact on trip production and generation at the 95% confidence level. Some of the

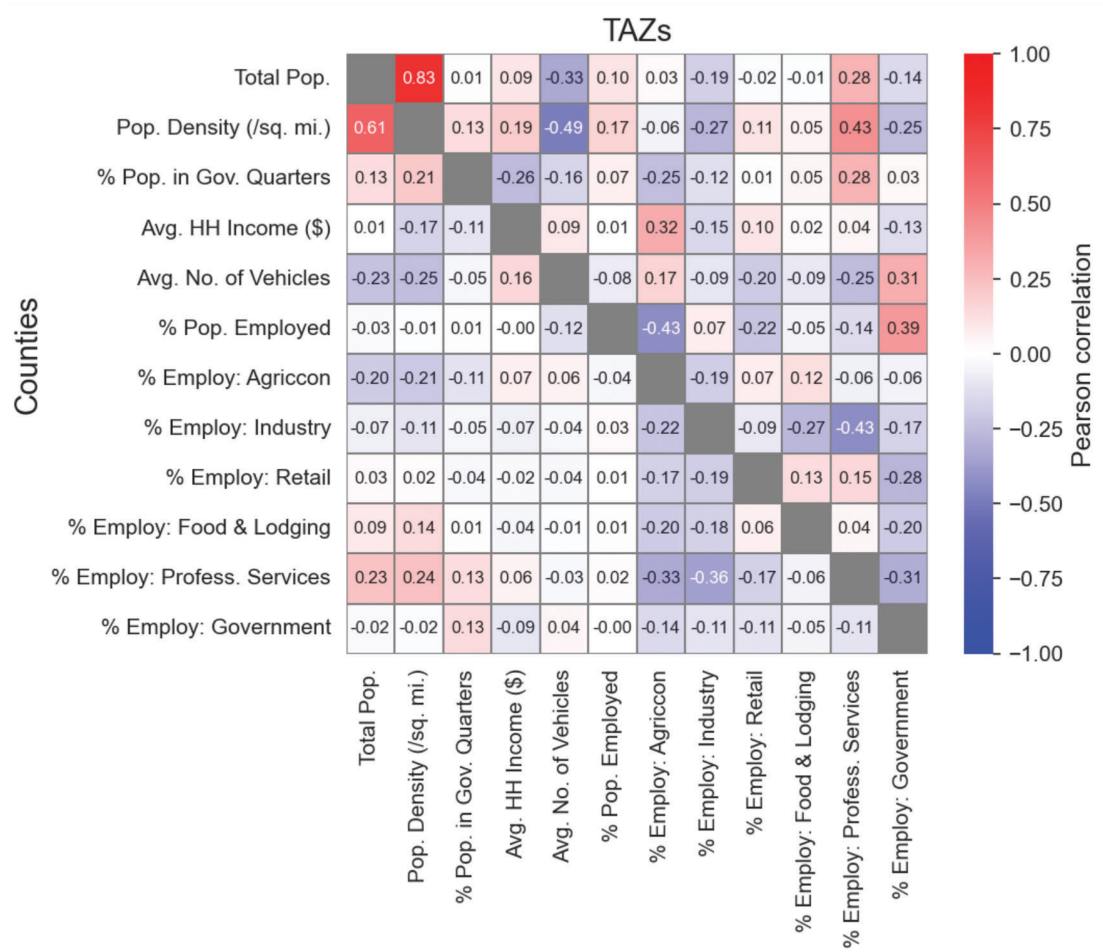


Figure 2.6 Pearson correlation between the covariates used in the trip generation model at the county and TAZ level.

TABLE 2.2
Coefficients of the finalized trip generation models for the total number of population-adjusted trips on a typical weekday/weekend

Model → Covariate ↓	Production Weekday	Attraction Weekday	Production Weekend	Attraction Weekend
Constant	-526.146*** (88.72)	-516.757*** (91.62)	-374.123*** (66.50)	–369.917***(72.78)
Total Population	0.483*** (0.00)	0.482*** (0.00)	0.356*** (0.00)	0.361*** (0.00)
Percent Population in Government Quarters	1,344.826*** (150.08)	1,411.755*** (154.98)	475.288*** (112.48)	611.110*** (123.11)
Avg. Household Income ($)	0.003*** (0.00)	0.002*** (0.00)	0.003*** (0.00)	0.002*** (0.00)
Avg. Number of Vehicles	214.119** (84.64)	212.882** (87.40)	185.729*** (63.44)	188.649*** (69.43)
Percent of Population Employed	3.406*** (0.55)	3.600*** (0.57)	1.764*** (0.41)	2.006*** (0.45)
Percent of Employed: Industry	366.771*** (66.24)	371.531*** (68.40)	129.971*** (49.65)	137.377** (54.34)
Percent of Employed: Retail	637.368*** (82.63)	673.878*** (85.33)	514.551*** (61.93)	598.531*** (67.78)
Percent of Employed: Food & Lodging	896.288*** (107.66)	945.145*** (111.18)	789.855*** (80.69)	884.140*** (88.32)
Percent of Employed: Professional Services	381.680*** (68.99)	390.159*** (71.25)	123.195** (51.71)	124.241** (56.60)
Adjusted R^2	0.770	0.758	0.761	0.733
F-Statistic (Probability)	1,357.0 (0.0)	1,272.0 (0.0)	1,291.0 (0.0)	1,117.0 (0.0)

Note: The asterisks next to the coefficients denote the p-value level: *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$. The values in the parentheses denote the standard error of the covariates.

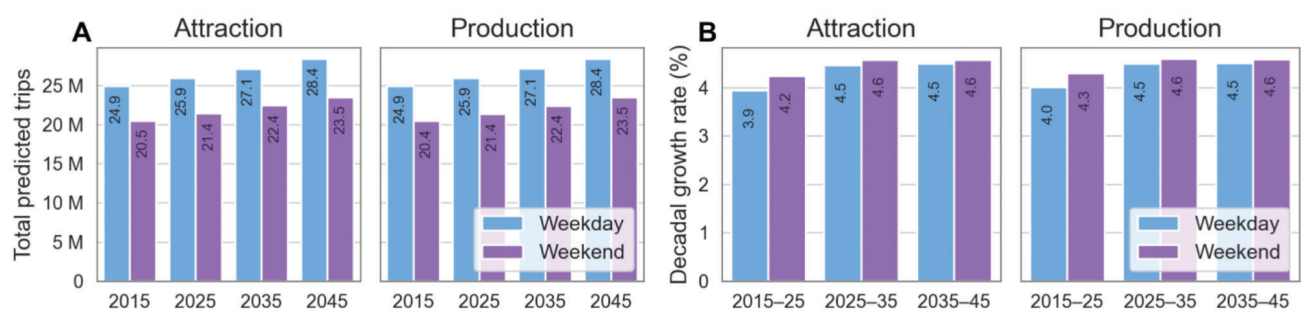


Figure 2.7 Predicted (A) total trip production and attraction in the TAZs, and (B) decadal growth rate.

industry-specific employment covariates such as percent employment in industry (INDUST), retail (RETAIL), food and lodging (FOODLD), and professional services (PROSRV) are found to be significant indicators of trip count, while others such as the percent employment in agriculture (AGCON), government (GOVNMT), and other services (OTHSRV) are not found to be significant.

2.5.1.1 Predicted trip generation. The finalized linear regression models, described in Figure 2.7, were used to predict the trip generation and attraction for three future and one past year in the ISTDM dataset—2015, 2025, 2035, and 2045. For these years, the total attraction and production by weekday and weekend are shown in Figure 2.7(a). Across Indiana, the models predict a total of 22.8 and 17.0 million trips on a typical weekday and weekend in 2015, respectively. According to the models, these numbers are expected to grow at a steady decadal growth rate of 4% to 5% (panel B) over the three decades, leading to a total of 26.2 and 19.8 million trips by 2045. Notably, the differences in total trip attraction and production are not substantial.

The trip generation resulting in values and growth also show considerable spatial heterogeneity over the years. Figure 2.8 shows the trip production at the TAZ level, i.e., the number of predicted trips produced in each TAZ. The broad patterns of trip density remain largely the same between 2015 (panel A) and 2045 (panel B) in that a huge proportion of the trips are centered in the major urban regions, primarily the Indianapolis and Chicago-Gary metropolitan statistical areas (MSAs). These are consistent with the population density distribution of the state at the TAZ level.

However, there are significant differences even at the TAZ level. This is shown explicitly in panel C which highlights the zones by their trip production growth rate between these three decades. The green regions that show positive growth are primarily centered around the northern belt from Chicago to Fort Wayne as well as in the suburbs of Indianapolis. Particularly, the regions close to Indianapolis core but outside its parent Marion County show some of the highest growth during this period. This intuitively makes sense since cities generally grow outward in the suburbs (Reia et al., 2022). However, it is not explicitly checked if this growth occurs due to the increase in population of the suburbs or because of the change in the employment and sectoral distribution of the jobs.

Conversely, the zones inside Marion County broadly show a small decline in trip production, suggesting a general decline in the level of activity in the region. Similar patterns are also seen in the western region of the state, close to the border with Illinois, and in the

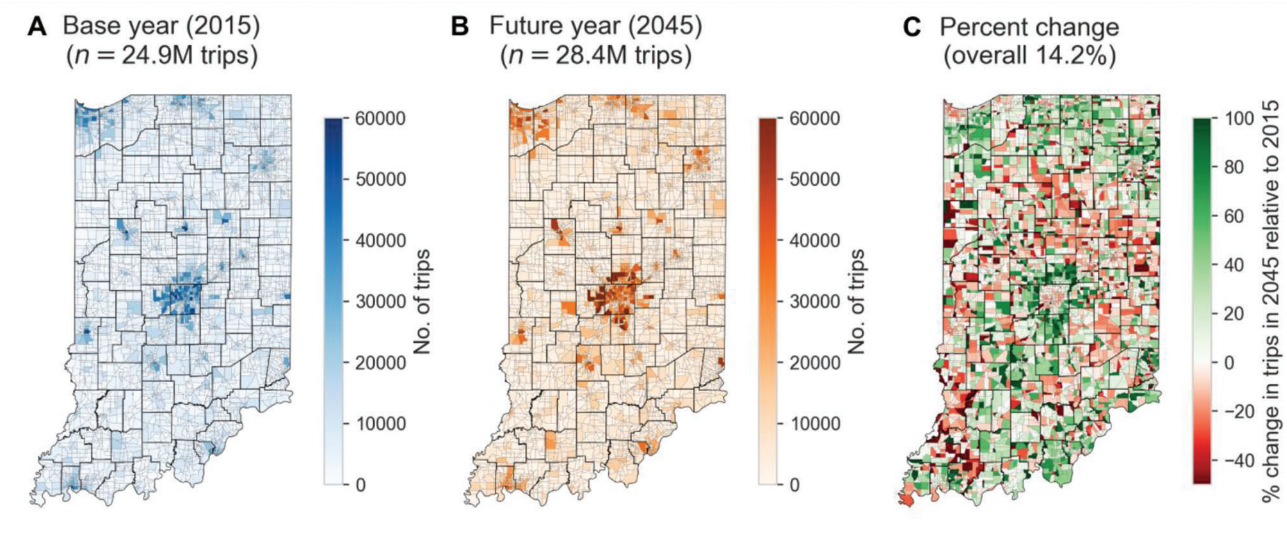


Figure 2.8 Comparison of weekday trip production in TAZs between (A) 2015, and (B) 2045. (C) The panel shows the percent growth in trip counts.

belt between the northern corridor around Fort Wayne and the center around Indianapolis.

2.5.1.2 Regions with the greatest and least growth. Figure 2.9 shows the top and bottom performing counties in Indiana by the percent growth rate of predicted trips between 2015 and 2045. For both trip production and attraction, the top five counties with the greatest growth are in the Indianapolis MSA, with upwards of 40% growth over the three decades. Notably, Marion County, the central area of the city, is associated with a small decline in the trip count. This indicates a substantial shift of population and opportunities from the city core to the suburbs of the largest city of Indiana. This is consistent with the trend of other major cities in the US over the years (Reia et al., 2022).

2.5.2 Trip Distribution

2.5.2.1 Model calibration. The gravity models for the trip distribution modeling were calibrated using line search optimization because of the presence of ground truth data and there being only one parameter in each model—the gravity exponent, β. The model was calibrated for each of the three base years (2019, 2020, and 2021), both day types (weekday and weekend), and additionally four cost functions (median/mean of travel time/trip length). The test range for the gravity exponent was taken as 0.1 to 3.0. The results of the line search optimization, observed in terms of MSE of the fitted model's trip count from the ground truth data of 2021, are shown in Figure A.5. Based on the preference of distance over travel time and the general preference of the median over the mean in the case of skewed distributions, the models with median distance were chosen. This led to a value of $\hat{\beta} = 1.6$ for a typical weekday and $\hat{\beta} = 1.4$ for a weekend. This higher value of the gravity exponent for weekdays compared to weekends is in alignment with the literature (Zhao et al., 2023) as it indicates the impact of travel time on commute trips on weekdays compared to more leisure trips on weekends.

2.5.2.2 Change of trip flows. The inter-county flows predicted from the calibrated model are shown in Figure 2.10. Most of the heavy traffic corridors are predicted to be between Indianapolis and the major urban centers across the state, particularly the Chicago-Gary MSA and Fort Wayne (panel A). These major corridors remain uncontested across the years and contribute to the greatest growth in flows between 2015 and 2045 (panel B). However, when taking into account the growth rate (i.e., normalizing with respect to base year's flows), the distribution changes completely (panel C). There seems to be substantial heterogeneity in the growth rate across different OD pairs, but most of these are inconsequential in absolute numbers, as seen in the differences in the flow maps in panels B and C.

To look at this issue in more depth, we also looked at the distribution of the TAZ OD pairs. The distributions of the predicted OD flows between the TAZs of Indiana for the prediction years (2025, 2035, 2045) and the past year 2015 are shown on logarithmic scale in Figure 2.11. Though the values over the years are similar for less popular zone pairs (i.e., the left end of the distribution), there is substantial difference for the more popular zone pairs (and thus possibly the routes between them), as evidenced on the right part of the distribution. Particularly, there are more OD pairs with higher flows between them in 2045 compared to 2015. This implies that the bulk of the growth in trip flows happening in Indiana between these years is predicted to happen between these zone pairs.

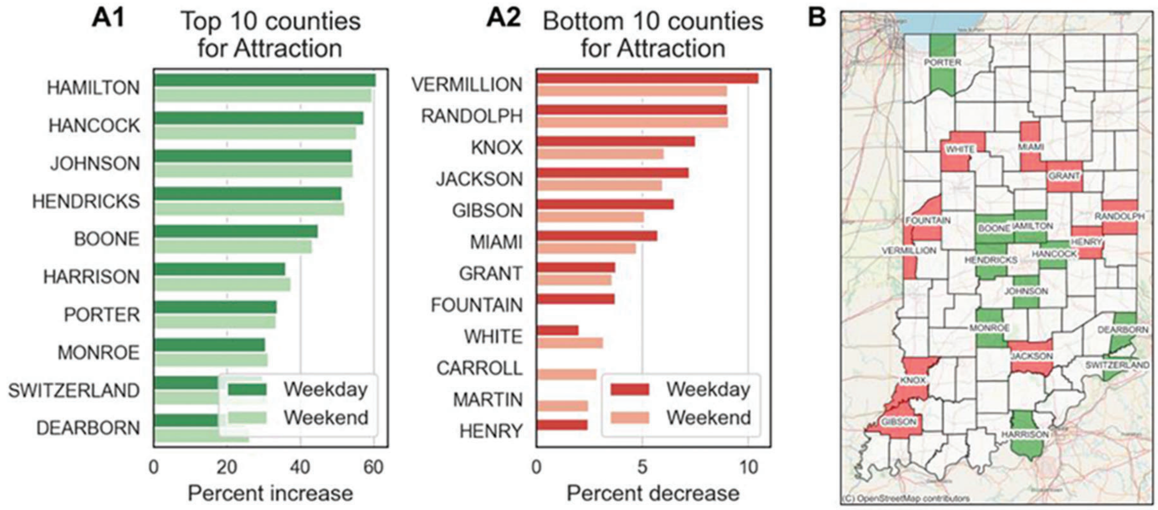


Figure 2.9 Top and bottom 10 counties in Indiana by the percent change in predicted trips between 2015 and 2045.

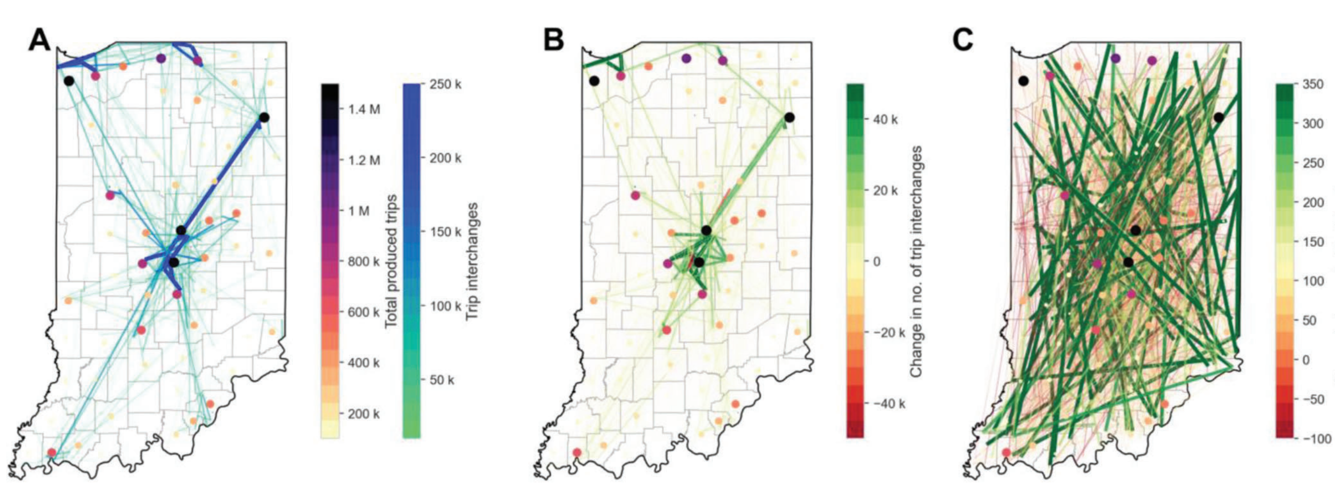


Figure 2.10 Maps of the predicted weekday inter-county flows and changes over the years: (A) flows in 2045, (B) change in flows between 2015 and 2045, and (C) percent change between 2015 and 2045. The total county trip production of 2045 is shown in each panel by colored dots, with their legend in panel A.

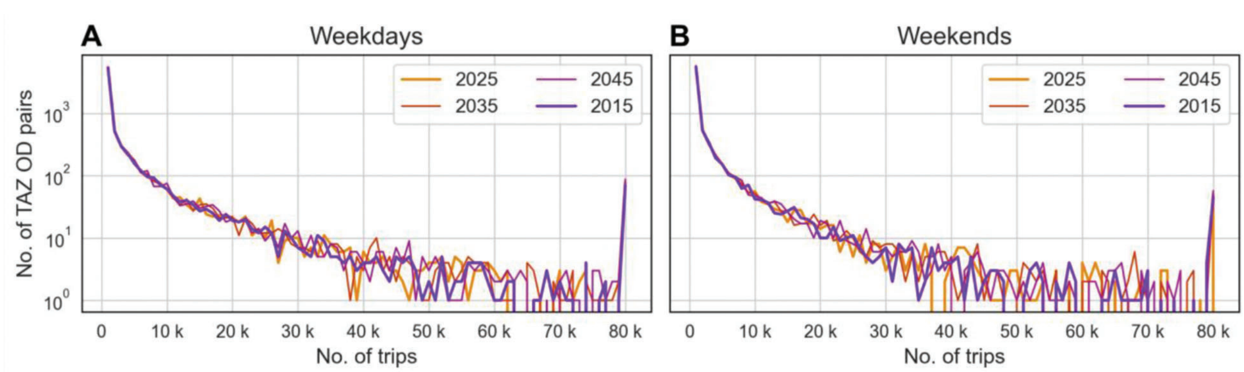


Figure 2.11 Distribution of the predicted OD flows for the target years for a typical (A) weekday, and (B) weekend.

2.5.3 Comparison and Validation

2.5.3.1 Comparison with ISTDM. To strengthen the validity of our results from the GPS data, we compared the trip generation predictions with the ISTDM's trip demand figures. We obtained the ISTDM's ODM for 2015 from the Indiana DOT containing a symmetric ODM for the TAZs for automobiles. The relationship between the trip production from ISTDM and our model is shown in Figure 2.12(a) at the TAZ-level. Their spatial relationship is shown in panel B.

In this figure, the geolocation-based predictions are strongly correlated ($\rho = 0.82$) with the ISTDM figures. However, they are also observed to be consistently overestimated compared to ISTDM's predictions. The fit of a univariate linear regression shows a slope of 1.152, indicating the GPS data-based predictions being overestimated by about 15%. It should be noted, however, that the underlying trip count for the GPS-based model depends on several parameters that may be tuned specifically for a target region so as to achieve a target prediction count. This includes the clustering parameters (such as the cluster radius or the maximum inter-cluster distance) (Yabe et al., 2019), the duration of the stay region, or a minimum number of pings or data quality for the extracted trips.

In addition to the observation of a slight overestimation of the overall trip count from the GPS-based trip generation model, it is also observed that these values are also spatially heterogeneously distributed. In Figure 2.12(b), it can be seen that there is a much higher number of trips predicted by the GPS-based model in the western and the southeastern portions of the state, which tend to be rural. It might be attributable to inadequate representativeness figures of these regions which highly skew the results. Notably, the predicted figures match very closely with the ISTDM figures (ratio close to unity) in Marion County, where there are sufficient mobile phone users to be adequately represented in the GPS dataset. This points out to the general observation that larger and more representative GPS data tend to show better results in predicting human mobility patterns (Blumenstock et al., 2015; Wang et al., 2018).

2.5.3.2 Validation with Fort Wayne Model. To strengthen the validity of our results from the GPS data, we compare the results of the developed trip generation models with the results of the Fort Wayne Travel Demand Model (FWTDM) (Avery, 2019) whose planning region and period closely resemble those used in this task. The Fort Wayne Transportation Plan of 2040 was developed on the basis of 2015 data. Its planning area includes most of Allen County, which includes most of the Fort Wayne City boundary, and parts of nearby rural counties—Whitley and Huntington, shown in Figure 2.13(a). Trip productions and attractions were developed for different trip purposes. To perform a fair comparison of the FWTDM with our model's results in the absence of a boundary GIS layer, we include all of Allen County and compare linearly interpolated figures between 2015 and 2045.

It can be seen in Figure 2.13(b) that our model consistently overestimates the number of trips by about 5%–10% compared to the FWTDM. This is similar to the results of the previous sections and might be attributed to the choice of data filtering and trip segmentation parameters. Further, our model also predicts a smaller growth in overall travel in the Fort Wayne area over the coming decades compared to the FWTDM, though the difference is small. For example, in 2015, our model overestimates ≈97,800 trips, but this difference reduces to ≈32,000 by 2045. This might be caused due to a huge number of differences in the two

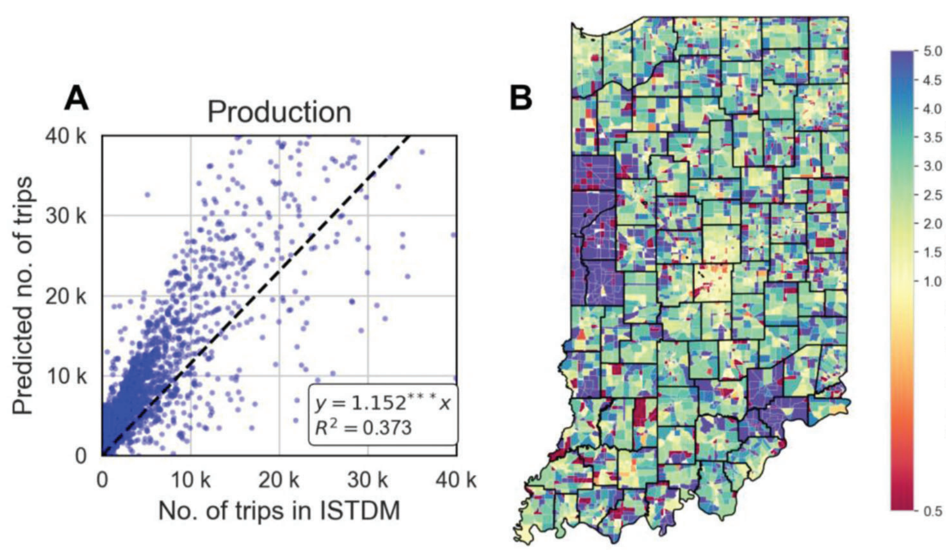


Figure 2.12 Comparison of trip production predicted by the trip generation model with ISTDM trips for 2015. (A) Scatter plot showing a linear model fit, with statistically significant slope at 95% confidence level. (B) Map showing the ratio of predicted ISTDM values for each TAZ, with black lines showing the county borders.

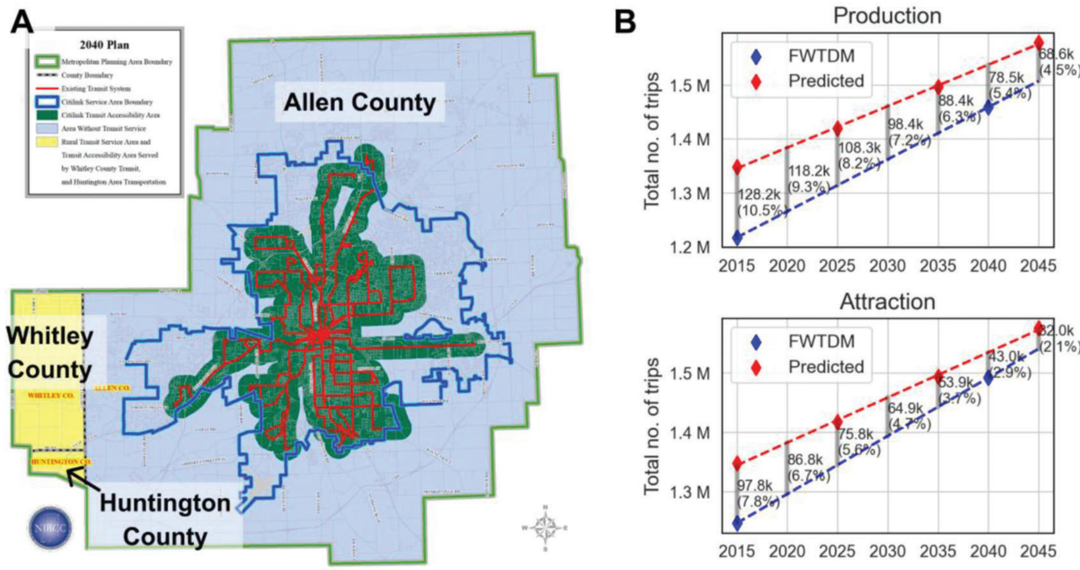


Figure 2.13 Comparison of predicted trip generation values with the values from the FWTDM. (A) Map of the planning area in FWTDM, per Avery, 2019. (B) Year-wise differences in the trip counts.

models, such as different SEA data and region boundary. However, we deem this difference small enough to consider our predictions reasonable.

2.6 Chapter Summary

This task extends the traditional four-step model for large-scale travel demand forecasting using GPS geolocation data from mobile phones, such as for statewide travel demand models (STDMs). Though multiple US states have traditionally used household travel surveys to obtain trip making behavior, the reluctance to integrate large-scale geolocation data analysis into this mobility assessment persists, including in Indiana. In this task, we show methods for cleaning and processing the GPS data to predict current trip patterns and use them in aiding the four-step model. In doing so, we argue that the trips obtained from the GPS data provide a better representation of the population at large scale, the data quality can be sufficient to make accurate current trip counts and use the estimated trip durations as the cost functions for the trip distribution model.

Our findings from this analysis are multifaceted and overarching. We find several factors, particularly total population, and percent employment in specific sectors, like industry and retail, to be major contributors to both trip production and attraction. In the trip generation predictions, the growth of movement in the state is expected to be at a nearly constant rate albeit with substantial spatial heterogeneity. Travel within the cores of the largest urban centers is predicted to decline mildly, but travel will increase the most to and from their suburbs. For example, all five of the highest-growing counties are situated within the Indianapolis MSA, whereas its core county, Marion County, is expected to experience a slight decline in overall trip production and generation. The same pattern is found in the inter-city flows, where the suburban cities are predicted to have the greatest growth in absolute counts, but this pattern vanishes when compared in terms of relative growth.

We also partially validate our results with the forecasts of other models. Particularly, we compare our predictions with those of both Indiana's ISTDMs as well as Fort Wayne's regional Travel Demand Model (TDM). In both cases, our model shows an over-prediction of the trips by 5%–15%. This difference is the least in the main urban areas which have on average higher representation in the GPS data. We argue that the methods (and their parameters) used to identify the trips using the GPS data play a significant role in influencing the overall trip counts.

2.6.1 Limitations and Future Work

We also recognize the limitations of our work. Importantly, we do not include trip rates by different travel modes and purposes in this study, such as home-based and work-based trips. Mode and trip purpose identification from geolocation data are subjects of active research, including some of the work in our research group. However, we do not include these in this study because of the scope of this project. Moreover, in the lack of ground truth data of trip patterns given GPS trajectories, we make numerous assumptions about the data processing methods. For example, we currently ignore night-shift workers in the step of home detection, which may possibly skew representativeness figures in commercial and urban areas. Similarly, we decide thresholds for trip segmentation based on literature rather than true trip patterns, which in some cases leads to erroneous classification of stay points as trip endpoints.

This study could benefit significantly by the availability of GPS-enabled travel diary surveys, albeit at a smaller level in a place similar in mobility characteristics to Indiana. In the presence of such travel diaries, we could develop supervised machine learning-driven models to predict trip origins/destinations by mode and purpose. Though conventional travel surveys such as the NHTS include details about the travelers along with the trips they make, a lack of geolocation data for them prevents them being useful to such supervised methods.

3. ESTIMATING THE FUTURE IMPACT OF E-COMMERCE ON TRAVEL DEMAND

3.1 Overview

To estimate how e-commerce technology could affect Hoosiers in the future, this chapter investigates the various preferences of Indiana households on e-commerce and forecasts travel demand changes under different adoption scenarios. Indianapolis was selected as the case study region because of data availability. We first review the related literature in Section 3.2 to summarize prior work findings and identify research gaps that motivate this study. The data and methods are discussed in detail in Sections 3.3 and 3.4, respectively. In this chapter, we first used the NHTS and consumer shopping record data with econometric models to identify the heterogenous shopping behaviors and market segmentations for e-commerce. We then applied the Integrated Traditional and Transformative Transportation System Use Model (ITTTS) developed in a previous INDOT project (Luo et al., 2022) and extended it with e-commerce features to estimate the travel demand impact of e-commerce under different future adoption scenarios. Major findings from the analysis in terms of market segmentation of shopping and travel demand impacts of e-commerce are presented in Section 3.5. Lastly, we summarize our major conclusions and discuss the practical implications of our findings in Section 3.6.

3.2 Literature Review

Prior work has examined the market segmentation of e-commerce and the correlation between the frequencies of online and in-store shopping trips in an effort to investigate the travel demand impact of e-commerce services. Dias et al. (2020) estimated a multivariate ordered probit model to analyze the differences in frequency between online and in-store activities for meal, grocery, and non-grocery shopping. The authors used revealed preferences data about activity travel behaviors and production information of online shopping to estimate the correlation between shopping frequency and household characteristics. The study also explored the complementary and substitution interactions between online and in-person shopping activities. The findings showed that online shopping is more prevalent for utilitarian shopping activities, while in-person shopping is more common for hedonic shopping activities. Shah et al. (2021) estimated a latent class model to analyze the clusters of shoppers based on their online shopping and daily travel patterns and used a multinomial logit model to examine the correlations between external characteristics (e.g., demographics, built environment, lifestyle, etc.) and different shopper types. Combining the travel behavior information from the NHTS and household demographic and socio-economic information from the ACS, they also observed that both complementary and substitution effects existed between online and in-store shopping. The authors pointed out that frequent online shoppers could travel less for in-store shopping purposes, showing the substitution relationship. But they also found that dual-channel shoppers may increase their in-store shopping frequency with online shopping. Cao et al. (2012) conducted a shopping survey in Minneapolis–St. Paul to understand the substitution and synergistic interactions between e-shopping and traditional in-store shopping and how online-shopping could affect travel behaviors. They built a structural equation model to estimate the interconnections of the frequency among online searching, online buying, and traditional shopping and their correlations with demographic characteristics. This study found that there are complimentary effects between online-shopping and traditional in-store shopping for new adopters, indicating that instead of replacing in-person trips, online shopping could stimulate more shopping trips, increasing vehicle usage. However, with the increased share of online purchases in overall shopping activities, online shopping could have a substitution effect on shopping trip reduction. Their findings show that different market penetration rates could lead to different impacts on travel demand. Kim and Wang (2021) built a two-step simultaneous equation model to estimate the factors that affect the delivery frequency for different store types (retail, grocery, and food) and their correlations with in-store shopping frequency by different travel modes (driving and walking). By analyzing the 1-day trip diaries for New York City residents, the authors found the existence of both positive and negative correlations between in-store shopping and delivery frequencies, depending on the type of store and travel mode of in-store trips. Their findings indicate the complex competing and complementary effects between in-store and online shopping, leading to the pendent impacts on travel demand.

These studies estimated the complementary and substitution effects of online shopping on in-store shopping and quantified the heterogenous impacts from the different market segments. However, travel demand impact assessment requires a quantitative analysis based on detailed shopping trips, including trip origin-destination, trip distance, and mode choice. Without linking the frequency changes and market segmentation with trip information, we cannot understand how online shopping could change travel demand and cannot support transportation planning.

Combining shopping behaviors and detailed travel demand data, some existing studies have also examined how online shopping services affect VMT. Suel and Polak (2017) developed discrete choice models (multinomial logit and nested logit) to estimate the joint choice of channel, shopping destination (store), and travel mode in response to the development of online-shopping services. They found that the availability of online shopping provides an attractive alternative among early adopters for high-income people. Their mode choice analysis showed that deliveries mostly reduced driving trips but replaced walking and public transport trips to a lesser extent. Although their study analyzed the VMT impact of e-commerce based on trip-level information and mode choice, this study only considered the avoided VMT from substituted online-shopping but ignored the additional VMT that delivery vehicles could potentially bring to the city. The net impact on travel demand is still unclear.

One study (Stinson et al., 2019) analyzed the net effect of e-commerce on regional mobility and energy consumption, considering whether the increase in delivery trips outweighs the reduction in household shopping travel. The authors first developed a bilevel multinomial probit model to estimate e-commerce demand at the household level based on the household travel survey in Seoul, South Korea. This was done to account for household decision rules on whether to use online-shopping and how frequently to order online based on their individual household demographic characteristics. They also used the modeling tool POLARIS and delivery data from UPS to build an agent-based simulation to model the travel demand changes under different shopping demand scenarios. Their results confirmed that, on the one hand, e-commerce reduces the demand for passenger car shopping trips, which can lead to a reduction in VMT and energy consumption; while on the other hand, e-commerce increases the demand for delivery truck trips, which can lead to an increase in VMT and energy consumption. As a result, the net effect of e-commerce on VMT and energy consumption depends on several factors, including the type of goods being purchased online, the distance between households and delivery locations, and the mode of transportation used for delivery. Their study conducted a comprehensive analysis of the travel demand impact of e-commerce, integrating shopping behavior with individual travel demand for the in-store shopping substitution benefits and estimating the net impact on VMT of delivery service. However, transportation planning requires forward-looking with future travel demand analysis. Their model and analysis were built based on the historical e-commerce market and demand, which cannot shed light on how future e-commerce would shift the travel demand. Previous studies have already shown that the market penetration of e-commerce could result in various impacts on urban transportation (Cao et al., 2012). Given the rapid adoption of e-commerce, it is necessary to understand the travel demand impact in future scenarios to support urban planning and infrastructure investment.

To fill the aforementioned research gaps, this chapter presents a modeling framework (Figure 3.1) to estimate the future travel demand impact of e-commerce in Indiana. The modeling framework first examines the market segmentation of shoppers to understand Indiana residents' heterogenous shopping behavior (Sections 3.4.1 and 3.4.2). Second, a batch of future scenarios are developed to model the future adoption of e-commerce, varying the market penetration and delivery service for different types of goods (Section 3.4.3.2). Finally, the model combines the e-commerce scenarios with a TDM and a goods delivery model to quantify the VMT changes from substituted in-person shopping trips and involved delivery tours (Section 3.4.3.3 and Section 3.4.3.4).

3.3 Data Description

3.3.1 Data Used for Modeling Online Deliveries Using NHTS 2017

We first used the 2017 NHTS data to develop trip rates for e-commerce in the Indianapolis-Carmel-Anderson MSA and the broader state of Indiana. The estimated population of the Indianapolis MSA and

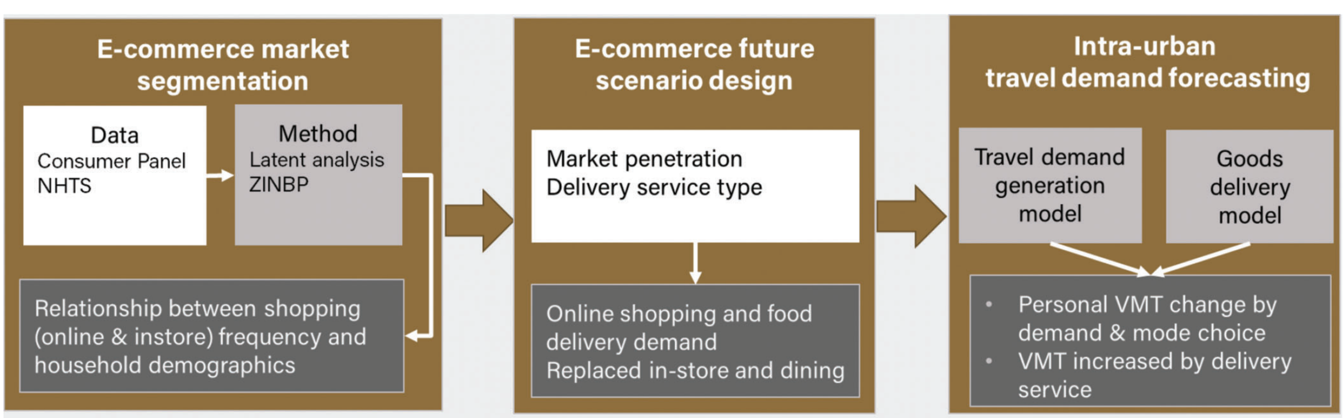


Figure 3.1 A modeling framework for travel demand impact analysis of e-commerce.

Indiana in 2017 is 2.02 million and 6.65 million, respectively. While the Indianapolis MSA primarily consists of urban geographies, the state of Indiana is largely comprised of rural, agricultural, and industrial clusters. To further justify analyzing the two settings separately, a descriptive analysis is subsequently conducted.

The data provides information on the households and the number of times respondents ordered goods online in the last 30 days of the survey. The histogram (Figure 3.2) shows that the use of online orders is over-dispersed in both settings. This over-dispersion suggests the use of a model such as a negative binomial Poisson regression.

Many respondents also reported no usage of online orders, likely due to a lack of access to the internet or awareness of online shopping. This suggests the use of a zero-inflated negative binomial regression (ZINBR) model.

Some individuals skipped the question due to reasons such as not engaging in the activity or the question not being applicable to them. To that end, about 15.6% and 13.1% of the responses have skipped the question for Indianapolis MSA and Indiana, respectively. Considering the relatively low skip rate, we have deemed it appropriate to exclude these records from our analysis. Thus, the subsequent analysis is conducted after excluding such responses. Additionally, the data underwent cleaning and preprocessing to facilitate its input into NLOGIT4. It should be noted that the dataset exhibits a panel-like structure, given that multiple people from the same household responded to the survey. This can be accounted for by using random effects in the model, but it is currently out of the scope of the project. To ensure data integrity, no missing values were found for any of the observations. Furthermore, to mitigate the impact of multicollinearity, we examined the correlation matrix with a threshold of 0.7 for correlation. This step was taken to refrain from using highly correlated variables in the model (Section 3.4.1), as it would lead to unstable results and overfitting. With the cleaned dataset, some new variables are created to aid the modeling, as described in Table 3.1.

3.3.2 Nielsen Consumer Panel Data for E-Commerce Market Segmentation

NHTS data only provides limited information about delivery usage frequency. Understanding shopping behaviors and e-commerce adoption requires the analysis on multiple dimensions, such as shopping distance and shopping cost, with detailed shopping records. We used the Nielsen consumer panel dataset, a cross-sectional survey with an annual shopping diary, as complementary information to analyze the current shopping pattern for the entire US (Nielsen, 2022). We obtained the data for 2017 that includes shopping records of 1,496 households in Indiana. Households participating in the consumer panel provide information to Nielsen about their shopping trips, including the place and date of purchase, the products bought, and their price. Each shopping trip includes the type of retailer, such as restaurant, grocery, beverage, and online. The dataset also provides information about household demographics (income level, household size, household members' age, education, gender, etc.) and geographic location (county and zip code).

To ensure the representativeness of the households in Indiana, we use the projection factors (sampling weights) for each household provided by the data collector to correct the potential sampling bias of the dataset. The weighted household demographics are comparable to the census values for Indiana state as shown in Table 3.2.

3.3.3 Travel Demand and Transportation System Data

We utilize a travel demand generation model developed in a previous INDOT project (Luo et al., 2022) to represent the daily travel demand in Indianapolis.

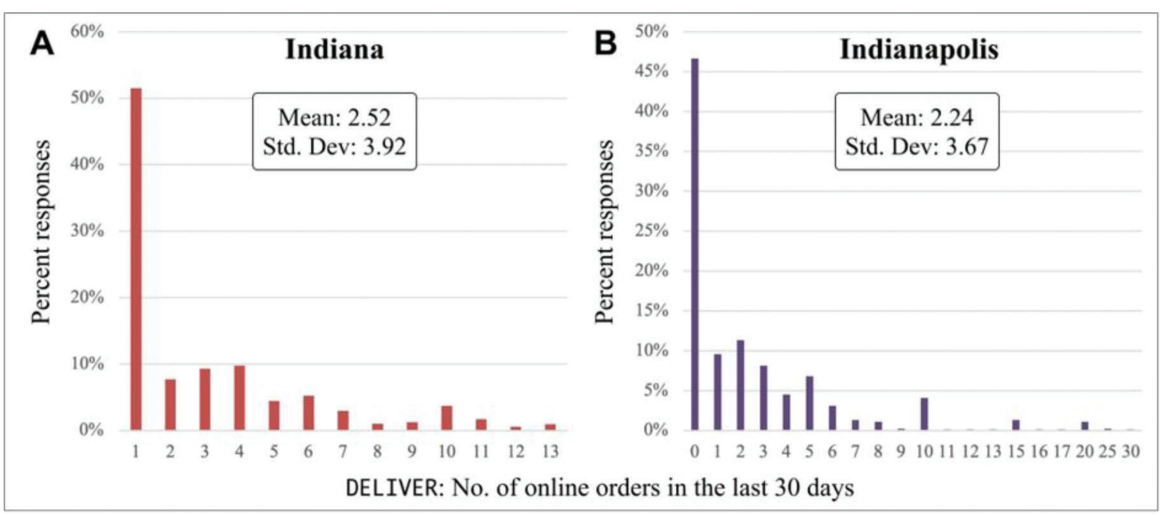


Figure 3.2 Histogram of the deliver variable in the NHTS for (A) Indiana, and (B) Indianapolis area.

TABLE 3.1
Derived variables from the NHTS and their descriptive statistics

Variable	Definition	Percentage (Indianapolis)	Percentage (Indiana)
OLD	1 if respondent's age > 60; 0 otherwise	25.20	25.76
RICH	1 if respondent's annual income > $75,000; 0 otherwise	29.07	34.09
DRIVER	1 if respondent drives; 0 otherwise	89.27	89.53
DAILYPC	1 if respondent uses PC daily; 0 otherwise	73.73	73.17
MALE	1 if male; 0 otherwise	47.21	47.56
EDUCATED	1 if higher education than high school; 0 otherwise	63.18	65.19
SPARSE	1 if population density of the block group of the household's home location is less than 100 persons per sq. mi.: 0 otherwise	2.37	19.22
DENSE	1 if population density of the block group of the household's home location is more than 4,000 persons per sq. mi.: 0 otherwise	28.69	19.86
NO CAR	1 if household vehicle count is 0; 0 otherwise	4.42	2.52

TABLE 3.2
Comparison of household demographic statistics for Nielsen 2017 weighted data and ACS 2018 5-year estimates

Characteristic	Category	Nielsen 2017 (%)	ACS 2018 (%)	Difference (%)
Age (years)	<18	25.4	26.1	-0.70
	19–24	6.3	6.9	-0.60
	25–34	13.4	12.9	0.50
	35–44	13.4	12.8	0.60
	45–54	11.7	12.0	-0.30
	55–64	12.7	12.8	-0.10
	>65	17.1	16.5	0.60
Gender	Male	47.4	49.6	-2.20
	Female	52.6	50.4	2.20
Household Income (thousand US $)	<10	5.2	5.7	-0.50
	10–15	4.5	3.8	0.70
	15–25	7.3	7.8	-0.50
	25–35	9.4	9.5	-0.10
	35–50	12.8	12.6	0.20
	50–75	19.0	19.3	-0.30
	75–100	12.9	13.9	-1.00
	100–150	16.4	15.5	0.90
	150–200	5.5	6.3	-0.80
	>200	7.0	5.6	1.40

The model uses data from the ACS to generate households (n = 780,000) in Indianapolis, with detailed socio-demographic information for each household and household member. Each household member is then assigned a chain of trips to represent the daily travel demand, including the trip origin and destination (OD) (with longitude and latitude), start/end time, trip purpose (e.g., home, shopping, dining, work), and mode choice. The trip chain is sampled based on trip diary data from the NHTS 2017. Detailed trip OD locations are inferred using NHTS travel diary data and local land use data from the (Indy GIS, 2022). It is worth noting that the shopping trips generated by this model are currently limited to in-store shopping trips. However, as online shopping can change shopping travel demand and the rest of daily trip chains, a detailed method of adjusting travel demand to account for the use of online shopping is discussed later in Section 3.4.3.3.

3.4 Methods

This section introduces the detailed sub-models proposed in our framework. First, the prevalence of online delivery in a household is modeled using a ZINBR model based on the NHTS data in Section 3.3.1. A latent class analysis (LCA) is then performed using consumer panel data to understand the households' heterogeneous shopping behaviors more comprehensively in Section 3.4.2. Linking the shopping behavior shifts with individual's travel demand, an agent-based model from a previous INDOT project

(Luo et al., 2022) is expanded to quantify the net impact on VMT from personal travel and goods delivery in Section 3.4.3. Finally, different adoption scenarios are designed to evaluate the future impacts of e-commerce in Section 3.4.3.

3.4.1 Online Delivery Modeling

Since the dependent variable DELIVER is of count data type and contains many zeros, we use the ZINBR model. ZINBR model assumes that the events, $Y = (y_1, y_2, \ldots, y_n)$ are independent. The model form is:

$$P(y_i = 0) = p_i + (1 - p_i)\left[\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i}\right]^{\alpha^{-1}} \quad \text{(Eq. 3.1)}$$

$$P(y_i = y) = (1 - p_i)\left[\frac{\Gamma(\alpha^{-1} + y)u_i^{\alpha^{-1}}(1 - u_i)^y}{\Gamma(\alpha^{-1})y!}\right] \quad \forall\, y = 1,2,3\ldots \quad \text{(Eq. 3.2)}$$

Here, p_i is the probability of DELIVER = 0 and y is the number of events (i.e., using delivery service for e-commerce) per period. To test the appropriateness of the ZINBR model, we use Vuong's statistic and the overdispersion factor, α, to select the final model (Table 3.3). The table provides a model-selection guideline to determine the appropriate model to consider. For instance, when both the Vuong statistic and the overdispersion parameter exceed 1.96, the ZINBR model is preferred over the Poisson or the regular Negative Binomial (NB) model. More details on the model are discussed subsequently (Section 3.5.1)

3.4.2 Online Shopping Market Segmentation

In order to assess the heterogeneity of shopping behaviors of Hoosiers, based on their similar shopping habits, we classify the households into multiple classes using an LCA approach, as illustrated in Figure 3.3.

We utilize shopping trip-related variables to categorize households by shopping behavior, including the total number of online and in-store shopping trips in a year and the average money spent (in US \$). The socio-demographic characteristics of each household are used as covariates in the LCA model to examine their correlations with the identified latent classes. The LCA model is estimated based on the following equation:

$$P(Y_i|X_i) = \sum_{c=1}^{c} P(c|X_i)\prod_{j=1}^{j} P(Y_{ij}|c) \quad \text{(Eq. 3.3)}$$

where each household i = 1, 2, …, N includes socio-demographic attributes X_i and indicator variables j. If household i is classified to be latent class c based on indicator variable j, Y_{ij} is equal to 1, otherwise 0.

3.4.3 Travel Demand Modeling Under the Impact of E-Commerce: An Agent-Based Approach

In order to assess how online shopping may impact travel demand, it is crucial to consider trip-level changes, including trip chaining, mode choice, and trip chains. For this study, we extended the ITTTS developed in a previous project to consider the impact of e-commerce on transportation (Luo et al., 2022). Figure 3.4 provides an overview of the general framework, illustrating the existing model and its relationship with the model extension, as well as the main outputs. We first provide a brief introduction to the existing agent-based model that forms the basis of our model (Section 3.4.3.1). We then outline the design of the future e-commerce adoption scenarios that serve as inputs for the model (Section 3.4.3.2). We further explain how the personal trip chaining is adjusted when a portion of the initial in-store shopping trips are replaced by e-commerce (Section 3.4.3.3). This adjustment allows us to capture the changes in travel patterns resulting from the adoption of online shopping. Additionally, we incorporate a goods delivery model to quantify the truck VMT associated with the delivery service, ensuring a comprehensive analysis of the entire system (Section 3.4.3.4).

3.4.3.1 Integrated Traditional and Transformative Transportation System Use Model. The existing model adopted from Luo et al. (2022) consists of a travel demand generation model and an agent-based simulation model for mode choice (Figure 3.5). The travel demand generation model generates households by sampling socio-demographic attributes from Census data in Indianapolis and assigns a trip chain for each household member by sampling from the NHTS trip data. NHTS data includes the trip purpose, mode, start/

TABLE 3.3
Vuong statistic vs. the overdispersion parameter

		t Statistic of the NB Overdispersion Parameter α	
		< | 1.96 |	> | 1.96 |
Vuong statistic for ZINBR (f_1(.)) and NB (f_2(.)) comparison	< | 1.96 |	ZIP or Poisson as alternative to NB	NB
	> | 1.96 |	ZIP	ZINB

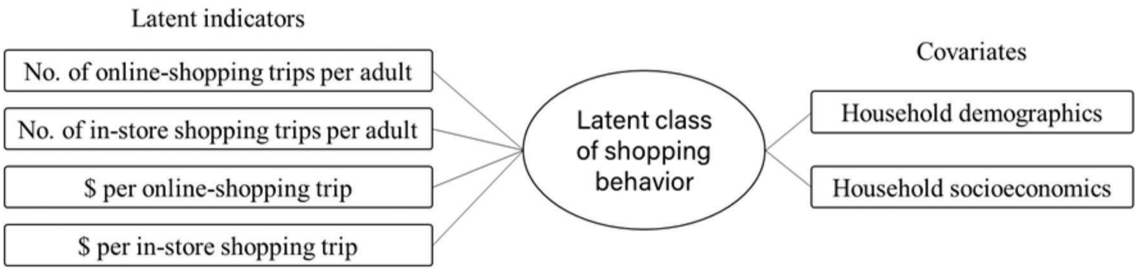


Figure 3.3 LCA model with covariates for shopping behaviors.

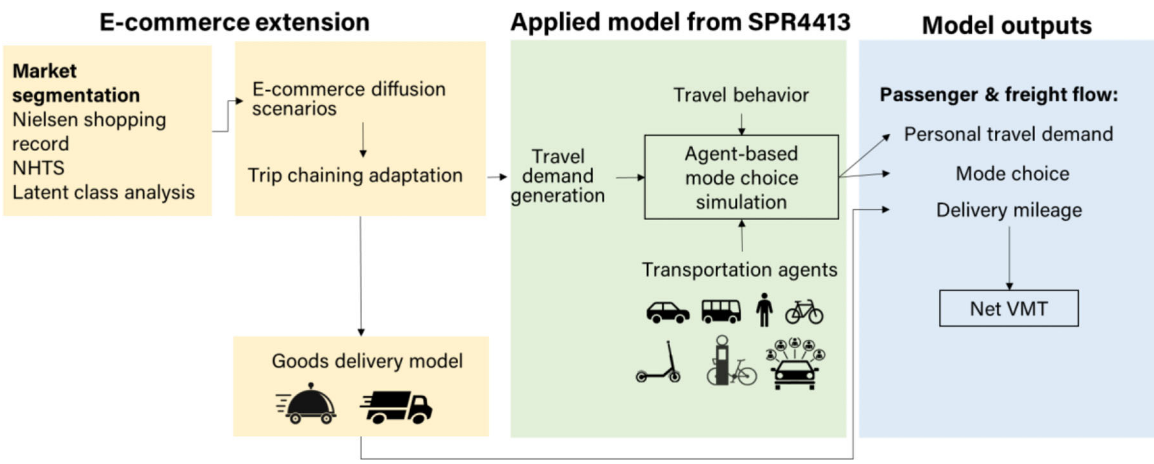


Figure 3.4 Extended travel demand generation and mode choice model.

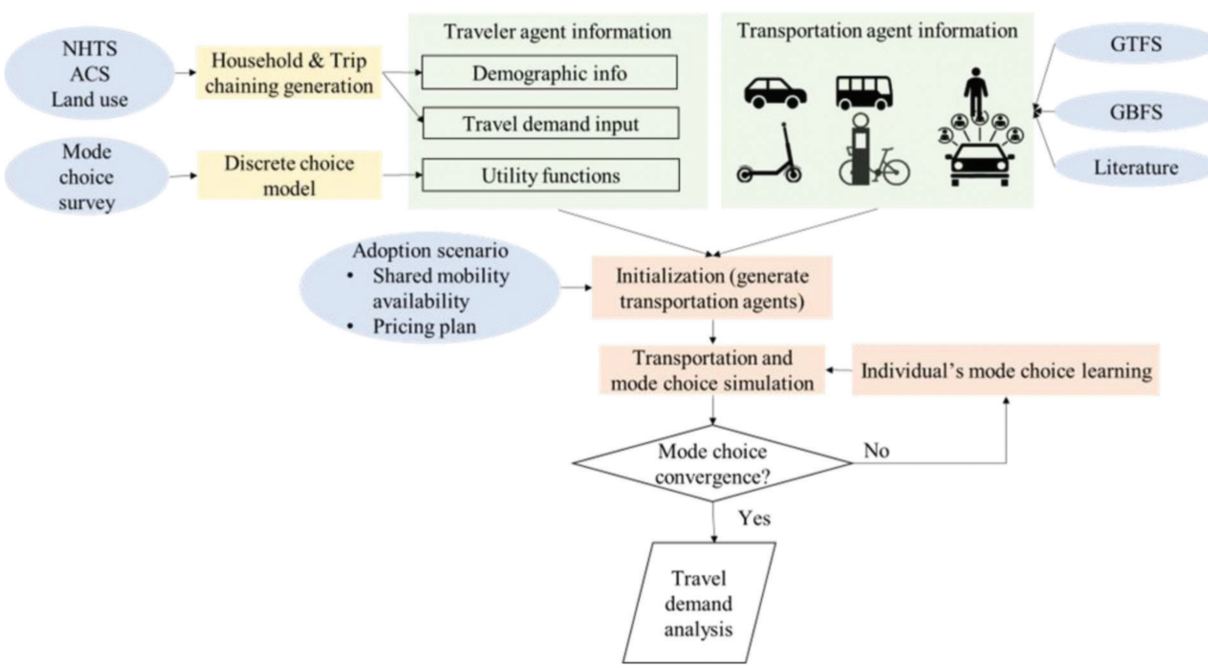


Figure 3.5 Integrated traditional and transformative transportation system use model (Luo et al., 2022).

end time, and distance, with high-resolution trip OD information estimated based on the trip chain information and local land use data (Indy GIS, 2022; McGuckin & Fucci, 2018). The generated travel demand reflects the chain of trips for Indianapolis residents and their traditional mode choices. The generated travel demand has been validated with the ISTDM OD matrix at the census tract level.

This agent-based simulation model incorporates two classes of agents. First, individual travelers are represented as agents, encompassing sociodemographic information and trip-chaining details obtained from the travel demand generation model. Each traveler agent is equipped with mode choice decision rules in the form of utility functions, derived from the local choice experiment survey from (Luo et al., 2023). Second, the model also includes various transportation systems as agents, such as public transit, private cars, bike-sharing, shared e-scooters, ride-hailing, and multimodal systems. The transportation agents receive inputs based on their historical system usage data, including factors like availability, pricing, travel speed, and more. The agent-based model simulates the final mode choice for each trip segment, considering the availability of shared mobility and multimodal options within the given adoption scenarios. The model generates outputs of the selected modes for each trip, enabling further analysis of travel demand and vehicle use under different scenarios.

3.4.3.2 E-commerce adoption scenario design. To estimate how online shopping services would change the travel demand in Indiana, we need to consider the potential increased use of online shopping in the next 30 years. Previous studies have shown that online shopping users are growing in the US and different delivery services are developing to satisfy various demands. To capture the different future adoption scenarios of online shopping, we create a Cartesian grid of 100 scenarios combining different percentages of market penetration of e-commerce (10%, 20%, …, 100%), which is the proportion of in-store shopping trips that would be replaced by online shopping, and the market share of centralized delivery (10%, 20%, …, 100%). In this study, we only consider e-commerce's impact on inter-city delivery VMT. City-to-city (e.g., hub-to-hub) delivery demand may not affect VMT significantly since courier firms, such as UPS, run fixed routes with pre-established schedules.

Market penetration is defined as the proportion of e-commerce shopping trips (i.e., orders) to the total shopping trips that shifted from in-store shopping trips. E-commerce has a substitution impact on in-store shopping when shoppers shift from in-store to e-commerce. We assume that each in-store shopping trip is replaced by one single online shopping order and delivery trip. We notice that there are many factors that can affect e-commerce adoption, such as energy cost, interest rate, economic recession, and delivery fees. To understand the impact of all possible scenarios under these factors, we consider the market penetration rate from 0% to 100% scenarios. We recognize that e-commerce adoption has several prerequisites, such as credit card, smartphone, and internet access. The simulated penetration scenarios assume these prerequisites will be considered in the future, although a considerable effort might be required to achieve this goal since Indiana ranks 7th lowest in credit card ownership and 18% of households remain without internet access (Tatham, 2018).

The likelihood of each shopper (i.e., household member) in different classes shifting from in-store shopping to e-commerce is based on the market segmentation results (Figure 3.6). We assume that the probability of new e-commerce trips taken by each shopper class is based on the relative online-shopping frequency from the 2017 consumer panel data. Based on the frequency of e-commerce use, *frequent online shopper* has a higher adoption rate with e-commerce diffusion, and *traditional shopper* is the last group that adopts e-commerce. As the market penetration increases, new e-commerce trips are more likely to substitute in-store shopping trips for *frequent online shopper*, followed by *dual-channel shopper*, *infrequent shopper*, and finally *traditional shopper*.

The delivery service type can be classified as centralized delivery (e.g., UPS, Amazon, Walmart Express) and decentralized delivery (e.g., Instacart, UberEATS). For the retail delivery service, we assume that orders for centralized delivery require to be delivered by 8 PM every day and each delivery vehicle can carry at most 50 collective orders, while orders for decentralized delivery require to be delivered within 1 hour and each delivery vehicle has a capacity of at most four orders.

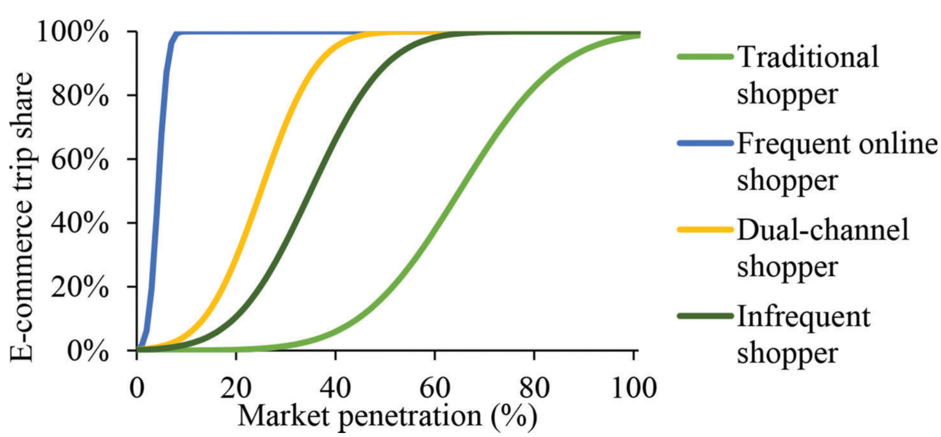


Figure 3.6 E-commerce trip share by shopper group.

3.4.3.3 Personal trip chaining adjustment to account for e-commerce adoption. The travel demand generation and mode choice model presented earlier is built on the assumption that all residents' shopping trips are in-store. To account for the impact of online shopping services on travel demand and mode choice, in addition to replacing previous in-store shopping or dining trip with online shopping, the change of the entire chain of trips and mode choice also need to be considered.

Figure 3.7 illustrates the adjustments made to two different trip types, to account for the impact of e-commerce. For round trip scenarios where the initial in-store shopping/dining trip is present, both trips are removed while retaining the rest of the trip chain, if applicable (panel A). In cases of chained trips where the initial trip is an in-store one, the inbound shopping trip and the trip leaving the shopping location are eliminated, and a new trip is added from the pre-shopping location to the post-shopping location (panel B). The updated trip chain represents the travel demand impact of replacing in-store shopping with online shopping. It encompasses more than the inversion of the conventional home-based in-store shopping trip, by integrating the distance required for delivery vehicles to travel from their starting location to the shopping place for goods collection. The delivery commuting distance may be longer or shorter than in-store shopping VMT depending on the delivery routes, thus resulting in mixed impacts on the net VMT that need further investigation.

In addition to the number of trips, mode choice for the entire daily trip chain can also be affected by the replacement of in-store shopping trips with online shopping, due to the differences in trip origin-destination locations and the availability of different travel modes (such as shared mobility and public transit). For instance, a traveler may drive a car for all-day's trips due to the in-store shopping activity. However, if this trip is replaced by online shopping, the traveler no longer has to drive a car and may shift to other modes, such as public transit and shared mobility. This shift in mode choice can also affect VMT. Therefore, in this study, we input the updated travel demand data to the existing agent-based model to re-simulate the mode choice given the adjusted chain of trips to estimate how online shopping could change VMT.

3.4.3.4 Intra city goods delivery model for parcel and food. In addition to personal travel, e-commerce can also impact a city's travel demand by increasing the VMT resulting from goods delivery services. Delivery vehicles pick up online-ordered goods from local stores (i.e., the same location as in-person shopping) and restaurants and deliver them to residents' homes within certain required delivery time window. In this study, we apply an optimization model based on the vehicle routing problem (VRP) (Toth & Vigo, 2002) that considers pick-up/drop-off demand and time window constraints (VRPPDTW).

Under different market penetration scenarios, the delivery order information can be identified from the selected in-store shopping trips that would be replaced by online shopping services. For each sampled online shopping trip, the delivery order includes a goods pick-up trip from local stores and a drop-off trip to the home. With the information on the delivery orders extracted, an algorithm is applied to define the graph, time window, and demand to formulate VRPPDTW that is then used to estimate the delivery vehicle routes and additional VMT. The algorithm is adapted from (Luo et al., 2020). The objective function is to minimize the total travel distance to finish the delivery work for all vehicles. The algorithm includes several main assumptions: (1) each node (stores and home locations) can only be visited once by one delivery vehicle; (2) the route for each vehicle needs to start from the depot, visit a sequence of nodes, and return to the depot; (3) each node must be visited within the pre-assigned time window; and (4) there is an upper limit to the number orders that each delivery vehicle can carry.

The VRPPDTW algorithm is applied to each of the developed scenarios, which vary in terms of the types of goods, market penetration, and the type of delivery service. For parcel delivery service, it is assumed that the time window constraint requires orders to be picked up between 9 AM and 5 PM and delivered to residents before 8 PM. For food delivery, a more stringent time window is imposed. Each food delivery order must be picked up after the order has been placed (i.e., visit time should be after initial dining time) and delivered within 1 hour.

Figure 3.7 Examples of trip chain adaptation for replaced in-store shopping/dining: (A) round trips, and (B) chained trips. White, dashed arrows represent the initial, eliminated trips and black, solid arrows represent added trips.

3.5 Results

3.5.1 Modeling Online Deliveries

The prevalence of online delivery in households is modeled using a ZINBR model based on the NHTS data. The model includes demographic and spatial variables to gain further insights into the trends of online delivery usage. The coefficients of the final model are shown in Table 3.4.

3.5.1.1 Indianapolis. With 266 observations, the model had a $\rho^2 = 0.46$. Critical t-statistics values were considered at significance levels of 0.1%, 1%, and 5%. The t-statistic for DRIVER and EDUCATED was significant at 0.1%. The t-statistic for MALE was significant at 1% and the rest of the variables were significant at 5%., RICH and DAILYPC were marginally significant. The dispersion factor and the zero-inflation factor were significant at 0.1% which gives us the indication that zero inflated negative binomial was the correct equation to use. The Vuong statistic of 3.642 also gives a strong indication of the zero-inflated model being appropriate.

The model revealed that various demographics significantly influence the frequency of online delivery usage in households in Indianapolis. Results showed that individuals aged 60 and over are less likely to use online delivery services. Drivers were found to be more likely to use online delivery services. The model also indicated that males were less likely to use online delivery services compared to females, while individuals with higher education levels were more likely to use online delivery services. The population density variables also play a role, with higher density areas showing lower online delivery usage. However, PC users and income status were not found to significantly influence online delivery usage.

3.5.1.2 Indiana. With 911 observations, the estimated model for Indiana explains 41% of the variance in the data ($\rho^2 = 0.41$). The t-statistic for all the parameters was above the critical value at 0.1%, except for SPARSE, which was significant at 5% and the constant term. The dispersion factor, alpha and the zero-inflation factor were significant at 0.1% affirming the suitability of the zero-inflated negative binomial. The Vuong statistic of 6.708 further supported the appropriateness of the zero-inflated model.

The second model showed similar results to the one estimated for Indianapolis, with consistent variables influencing online delivery usage. However, the daily usage of PC and income status became significant predictors in the second model, indicating the positive influence of daily computer usage and negative influence of higher income on online delivery frequency. Additionally, location sparsity was also liked with higher online delivery usage.

3.5.2 Market Segmentation of Shoppers

3.5.2.1 Latent class analysis for shopping behavior class. To select the model with the best estimation performance and efficiency, we ran multiple LCA models with varying numbers of latent classes. First, we found that only the models with two to six classes achieved estimation convergence. Second, we finalized the model selection based on the goodness-of-fit

TABLE 3.4
Parameter estimation results for online delivery prevalence model

Variable	Indianapolis Estimate (t-statistic)	Indiana Estimate (t-statistic)
Constant	-0.781 (-1.879)	-0.019 (-0.151)
OLD (1 if age>60, 0 otherwise)	-0.791 (-4.812)	-0.526 (-6.363)
RICH (income>$75,000)	0.179 (1.127)	0.418 (5.569)
DRIVER (1 if respondent drives; 0 otherwise)	1.520 (3.502)	0.545 (4.681)
DAILYPC (1 if respondent uses PC daily; 0 otherwise)	0.214 (1.083)	0.387 (4.505)
MALE (1 if make, 0 if female)	-0.365 (-2.536)	-0.421 (-5.903)
EDUCATED (1 if higher education than high school, 0 otherwise)	0.651 (3.799)	0.560 (7.197)
SPARSE (1 if population density (persons per square mile) in the census block group of the household's home location is less than 100, 0 otherwise)	–	0.166 (1.732)
DENSE (1 if population density (persons per square mile) in the census block group of the household's home location is more than 4,000, 0 otherwise)	-0.281 (-1.602)	–
Dispersion Factor, α	0.998 (5.533)	0.894 (9.865)
Zero-Inflation Factor	-1.33 (-3.743)	-1.254 (-7.474)
Vuong Statistic	3.642	6.708
Rho-Squared	0.460	0.410
Adjusted Rho-Squared	0.450	0.407
Restricted Log Likelihood	-847.575	-2907.671
Log Likelihood	-457.124	-1715.148
Log Likelihood Ratio k	780.902	2,385.04
	9	9
Number of Observations	266	911

statistics for each model. The four-class model had the lowest Akaike information criterion (AIC) and Bayesian information criterion (BIC) values, indicating that it was the best model with a trade-off between model fitness and complexity (as shown in Figure 3.8).

Descriptive statistics for each estimated class can help us understand the shopping behaviors of different household classes, taking into consideration shopping frequency and cost (Figure 3.9).

1. *Traditional shopper:* A large group (64%) of Hoosier households belong to this class. The primary characteristic of this class is that they engage in minimal online shopping and prefer in-store shopping. 93.4% of traditional shoppers never used online shopping during 2017. Even when they used online shopping, their frequency was less than once per month, and they spent very little online compared with other classes. Most of their shopping demand is satisfied by in-store shopping. Compared with other classes, traditional shoppers also have median values of annual in-store shopping frequency (3–6 times per week) and per-time shopping spending. In summary, this class of households has a moderate shopping demand with a preference for in-store shopping.
2. *Frequent online shopper:* It is the smallest group (7%), indicating that the online shopping market penetration rate in Indiana is still low and has the potential to increase in the future. This class has the most usage in e-commerce, with 24% of households shopping online more than once a week. However, they still have a moderate in-store shopping frequency. Over 50% of households go in-store shopping about two to three times a week. With the highest online shopping frequency, households in this class spend less money for each online shopping order. The statistics show that this class of households is becoming accustomed to shopping online, but the current online shopping services in Indiana cannot fully satisfy their various shopping demands.
3. *Dual-channel shopper:* This class (12%) represents households that have an average level of frequency and cost for both online shopping and in-store shopping. Most households shop online every 2 to 4 weeks, while they still rely on a lot of in-store shopping compared with the “frequent online shopper" class. In summary, the “dual-channel shopper" represents households who have started to use online shopping services but still rely on in-store shopping to cover most of their shopping demand.
4. *Infrequent shopper:* Households in this group not only barely shop online but also have the lowest in-store shopping frequency. Most households shop less than 1.5 times per week in 2017. However, their per-time cost is relatively higher than other classes of households. Over 40% and 37% of households spend more than $75 on each trip for in-store and online shopping, respectively, while the average per-time shopping spend is $56. This indicates that infrequent shoppers prefer to buy more things each time to avoid many shopping trips.

3.5.2.2 Household characteristics of shopping classes. Understanding the correlation between shopping behaviors and household characteristics can help estimate which shopping class a household may belong to. Third, our findings suggest that households with younger household heads, higher education levels, and higher incomes are more likely to use online shopping services more frequently. Using online shopping services requires access to a PC/smartphone, credit card, internet, and mailing service, as well as knowledge of using digital services. Therefore, it is not surprising that households with higher levels of education and income are more likely to use online shopping services more frequently. Table 3.5 lists the model estimation results of the demographic characteristics with the assigned latent class. We use Class 1 (*traditional shopper*) as the reference class to identify how other shopping behavior classes differ from them. The findings indicate that compared with Class 1, the other three latent classes may have smaller household sizes, internet access, younger age, and higher education level.

First, access to the internet is a critical determinant of whether a household can use online shopping or not. According to our consumer panel dataset, over 17% of households in Indiana did not have internet access by 2017 (as shown in Figure 3.10), which could be a significant obstacle for future online shopping market penetration. The lack of internet infrastructure/equipment or lack of knowledge to use the internet may be reasons for this issue. Therefore, without improving access to the Internet, the online shopping market may not be able to cover those households. When affordable, convenient, and user-friendly internet services can be provided, it is still possible for those households to replace some of their in-store shopping trips with e-commerce.

Second, we found that households with smaller sizes are more likely to be frequent online shoppers and dual-channel shoppers, while larger households are more likely to be traditional shoppers. Larger households may include multiple household members (e.g., children) who may have various types of goods and consume the goods (e.g., food) more quickly than other households. In addition, visiting the stores could also be a family activity and an entertainment activity for children. Such consumption characteristics encourage them to shop more frequently. Therefore, visiting local shopping stores is a more convenient way for such households to save time searching for goods online.

Third, our findings suggest that households with younger household heads, higher education levels, and higher incomes are more likely to use online shopping services more frequently. Using online shopping services requires access to a PC/smartphone, credit card, internet, and mailing service, as well as knowledge of using digital services. Therefore, it is not surprising that households with higher levels of education and income are more likely to use online shopping services more frequently.

3.5.3 Travel Demand Impact of E-Commerce Under Different Scenarios

We have analyzed a series of scenarios varying the market penetration and delivery service type to study

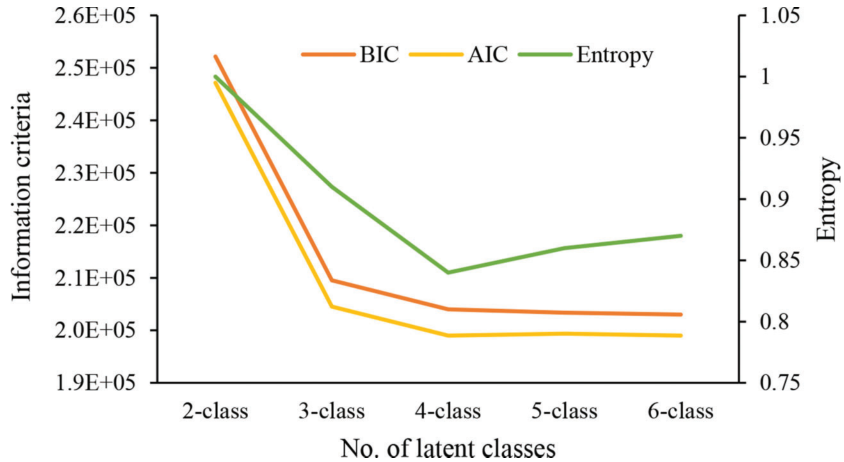


Figure 3.8 LCA model selection criteria for different numbers of latent classes.

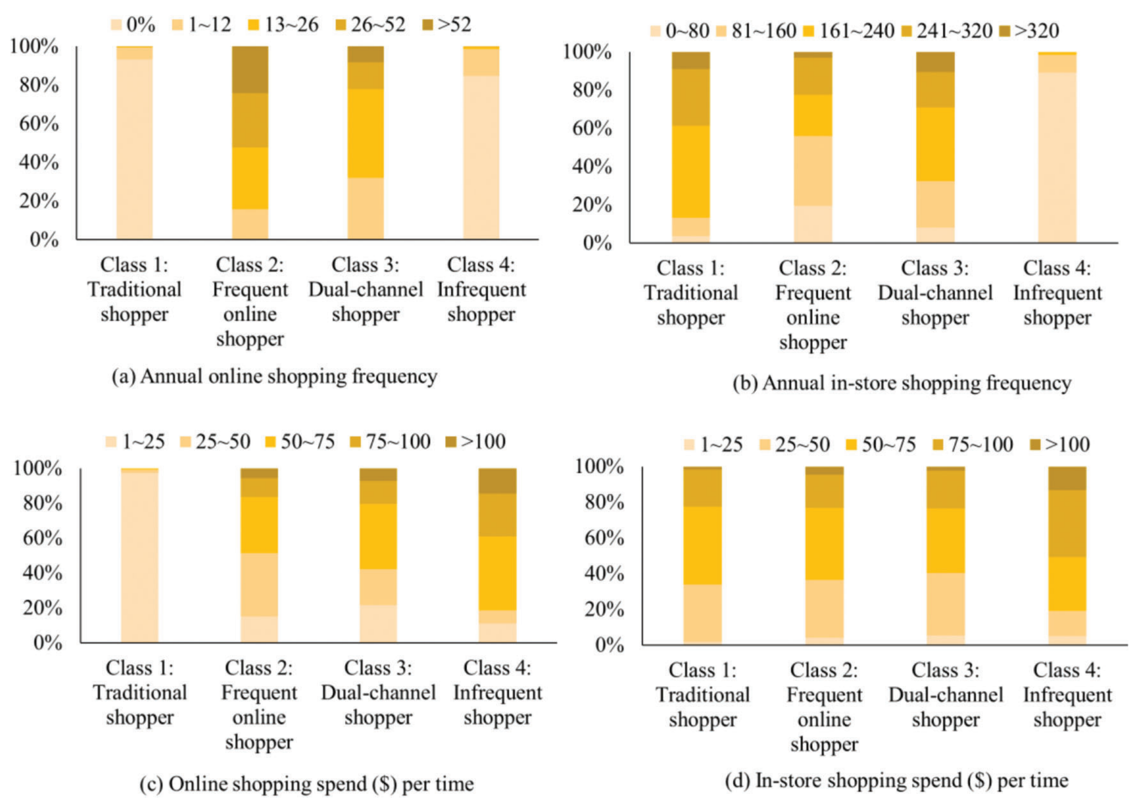


Figure 3.9 Statistical description of four latent classes based on the shopping behavior.

their impacts on VMT in Indianapolis. The changes in VMT result from both the replaced in-store shopping trips and the distances traveled by delivery vehicles.

Figure 3.10 displays the net VMT changes compared with scenarios with no online shopping for both parcel delivery and food delivery. Overall, our findings show that parcel delivery could have a greater impact on VMT than food delivery. As the market penetration rate increases, the changes in VMT resulting from parcel delivery become more significant than those resulting from food delivery. If all in-store shopping trips were to be replaced by online shopping and delivery, this could potentially result in a change in shopping-related VMT of ±16%, depending on the market share of the delivery service. The impact is equivalent to around ±300 thousand VMT per day. Specifically, the best scenario happens when all in-store trips are replaced by e-commerce (market penetration 100%) and customers only choose centralized delivery (100%), while the worst scenario happens when customers only use decentralized delivery service. However, take-out orders and food delivery services are unlikely to have a significant impact on shopping related VMT, with changes ranging from ±2% (±36,000 VMT per day), regardless of the market penetration rate or delivery service type.

TABLE 3.5
Latent class model's covariates (Reference class: Class 1: Traditional shopper)

		Online-Shopping Adopter		Dual-Channel Shopper		Infrequent Shopper	
Variable	**Category**	**Coefficient**	***t*-statistic**	**Coefficient**	***t*-statistic**	**Coefficient**	***t*-statistic**
Intercept	–	2.23	17.26	0.96	15.28	0.43	10.85
Household General	Household size in 2017	-0.45	-5.85	-0.3	-5.39	-0.02	-6.05
	Number of children <18	0.88	2.09	0.71	1.91	0.04	-1.94
	Number of workers	1.05	3.28	0.89	3.58	-0.01	-2.64
	Internet access in 2017	-2.35	-2.22	-2.24	-2.14	-0.48	-1.71
Household Income (2017 US$) (reference: $50–100k)	<$50k	-0.67	-3.18	-0.52	-3.15	-0.12	-3.56
	>$100k	0.4	2.43	0.26	2.23	-0.11	-2.55
Household Composition (reference: married)	Living alone	1.01	4.02	0.5	3.57	0.89	4.59
	Living with others	-0.34	-4.21	-0.17	-3.29	-0.12	-404
Age (years) (reference: >55)	**Male Head Age**						
	18–34	-0.83	-2.65	-0.07	-1.97	-0.61	-4.13
	34–54	-2.15	-5.45	-0.26	-4.48	-1.98	-5.20
	Female Head Age						
	18–34	-0.84	3.15	-0.2	-3.80	-0.59	-3.02
	34–54	-2.93	-1.56	-0.18	-2.76	-2.07	-4.21
Education (reference: high school)	**Male Head Education**						
	<High school	-1.63	-1.61	-1.47	-3.21	0.06	4.01
	College or higher	0.32	4.07	0.28	4.02	0.29	1.88
	Female Head Education						
	<High school	-1.78	-2.68	-1.61	-2.87	0.53	3.90
	College or higher	-0.15	-3.81	-0.03	-2.62	-0.18	-2.03
Employment	**Male Head Employment**						
	Full/part-time employment	1.58	1.96	1.45	3.82	0.17	1.85
	Unemployed	0.52	2.75	0.99	3.53	-0.77	-3.28
	Student	1.82	3.06	1.76	2.08	1.29	2.15
	Other	0.79	1.99	1.07	3.73	-0.62	2.38
	Female Head Employment						
	Full/part-time employment	1.61	3.85	1.5	2.21	0.16	4.27
	Unemployed	0.55	3.28	1.02	1.98	-0.61	-4.14
	Student	1.91	2.15	1.86	3.09	1.44	2.09
	Other	0.81	2.38	1.19	2.36	-0.6	3.33

Note:
Log-likelihood: -24823
AIC: 19917
BIC: 20564

The difference in elasticity between VMT changes resulting from online shopping and takeout delivery can be attributed to the difference in trip distances. As depicted in Figure 3.11, the trip distance for shopping trips is generally longer than that for dining trips. Replacing in-store shopping with online shopping could potentially result in significant VMT savings if a majority of customers opt for a centralized delivery service, but it could also lead to an increase in VMT due to delivery service if most customers choose a decentralized delivery service. Our analysis suggests that a 70% centralized delivery rate is the breakeven point for the VMT impact of online shopping.

When comparing different scenarios of delivery service types for online shopping, we found that centralized delivery service is more efficient than decentralized delivery, as it can result in greater reductions in VMT from parcel delivery. The increased efficiency of centralized delivery is due to the collective orders, as depicted in Figure 3.12, which shows the average load per delivery truck (number of orders/truck-mile) for different delivery services. On average, a centralized

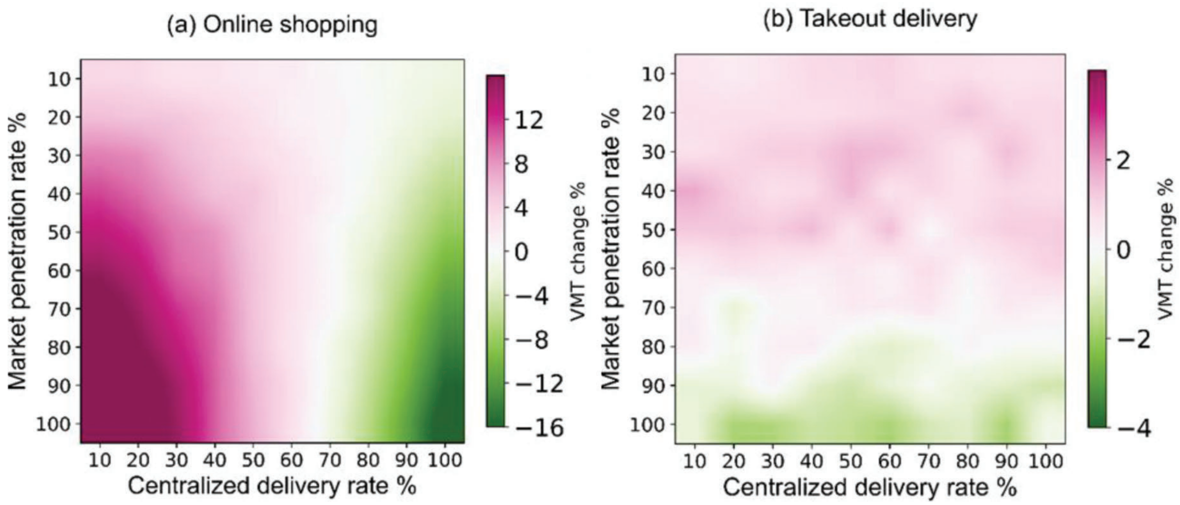


Figure 3.10 Net VMT changes under different scenarios: (A) online shopping, and (B) takeout food.

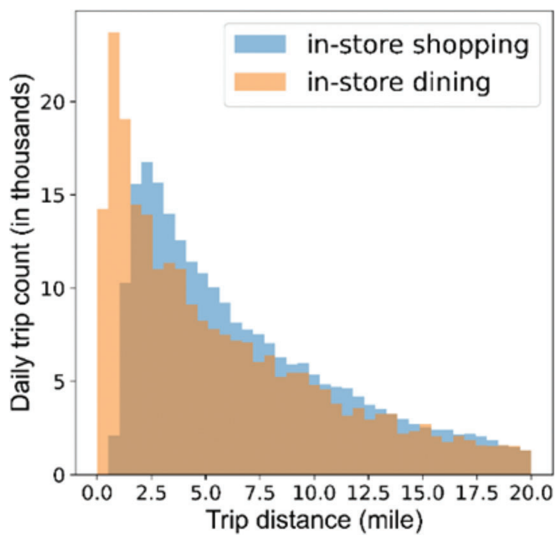


Figure 3.11 Trip distance distribution of in-store shopping and dining trips.

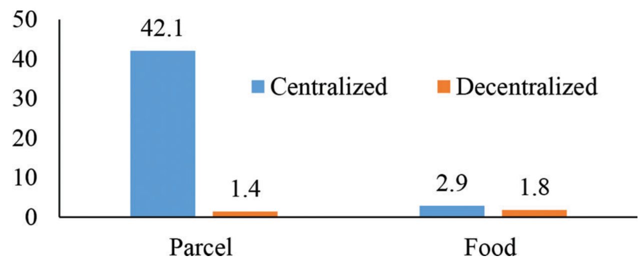


Figure 3.12 Average load for delivery services (no. of orders / mile).

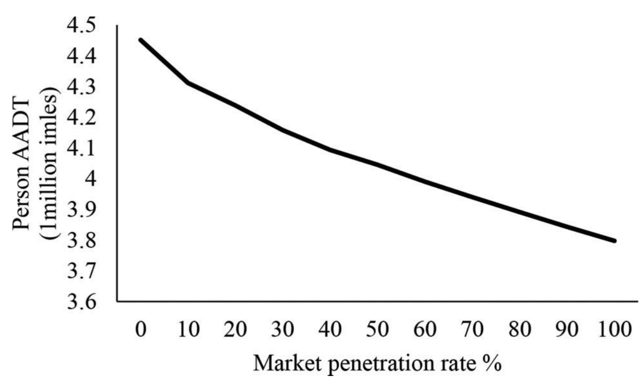


Figure 3.13 Personal travel distance by e-commerce market penetration.

parcel delivery can carry up to 42 packages per truck, while a decentralized parcel delivery can only carry less than two packages at a time. The disparity is due to the different time requirements for the two services. Centralized parcel delivery allows for a wider time window, allowing for the pickup of orders from stores during regular business hours and delivery to customers any time before 8 PM. The relaxed time constraints allow each centralized delivery truck to better plan delivery routes to pick up as many packages as possible along the way while minimizing VMT. In contrast, decentralized parcel delivery requires packages to be delivered within a 1-hour time constraint. In such a restricted time window, delivery trucks have to travel between stores and customers more frequently and cannot deliver multiple packages at the same time. Similarly, for food delivery services, both centralized and decentralized services have restricted delivery time constraints, which lead to low truck loads. This explains why the market share of delivery types for food delivery does not have a significant impact on travel demand changes.

We further conduct a detailed analysis of the impact of e-commerce on travel demand, specifically focusing on personal travel distance, mode shift, and delivery distance. Figure 3.13 presents the changes in Indianapolis Annual Average Daily Traffic (AADT) corresponding to varying market penetration rates. The figure highlights that the adoption of e-commerce has the potential to decrease personal travel distance by at most 14% from 4.45 million miles to 3.82 million miles per day.

This reduction in personal travel translates to an annual reduction of 0.95 million gallons of gasoline (FHWA, 2020) and 83,000 t $CO_{2\text{-eq}}$ of GHG emissions (U.S. EPA, 2023). Moreover, our observations indicate that the effectiveness of reducing personal travel demand is more pronounced at lower market penetration rates. This diminishing marginal effect arises from the diffusion of e-commerce technology among different customer groups. In our model, we incorporated the current frequency of e-commerce usage to capture the diffusion process. As the market penetration increases, new e-commerce trips are more likely to substitute in-store shopping trips for "frequent online shopper," followed by "dual-channel shopper," "infrequent shopper," and finally "traditional shopper."

In Figure 3.14, the distribution of shopping trips among different customer groups reveals that "frequent online shopper" tend to travel longer distances for in-store shopping compared with other groups, whereas "traditional shopper" has the shortest in-store shopping trips. When the market penetration of e-commerce is low, it exhibits a greater potential to replace long in-store shopping trips for "frequent online shoppers." Based on our market segmentation analysis (Section 3.5.2), we found that "online shoppers" prefer to purchase a larger quantity of goods with less frequency. E-commerce not only offers them a convenient alternative to avoid lengthy shopping trips but also effectively reduces their personal travel demand. However, as e-commerce begins to reach the "traditional shopper," who exhibits regular and frequent shopping behavior due to larger household sizes, its ability to substitute longer in-store trips diminishes. Consequently, the marginal benefits of reducing personal travel decline over time.

The change in personal travel resulting from e-commerce can also lead to a shift in the mode choice for travelers, subsequently impacting the city's VMT. Figure 3.15 provides the mode share distribution corresponding to different levels of e-commerce market penetration. Our findings show that the mode share of private cars can potentially decrease by up to 5% in terms of the number of trips, reducing from 90% to 85%. With the substitution of in-store shopping demand and altered trip chains, individuals exhibit a greater willingness to utilize non-car travel modes. Specifically, we observed that the mode share of multimodal transportation, walking, and public transit can increase from 0.7%, 5%, and 2% to 2%, 7%, and 3%, respectively. The adoption of e-commerce not only has the potential to reduce personal travel demand but also encourages people to opt for greener transportation modes. This shift in travel behavior contributes to the promotion of transportation sustainability.

Apart from its impact on personal travel demand, e-commerce also encompasses the delivery service, which involves the pickup of goods or food from stores and their subsequent delivery to customers. However, this aspect introduces an additional traffic burden due to the usage of delivery trucks.

Figure 3.16 illustrates the daily average mileage associated with parcel delivery and food delivery under various e-commerce market penetration scenarios. In general, parcel delivery services tend to entail higher delivery mileage compared with food delivery services. This discrepancy arises from the relatively long distances between customers' homes and shopping destinations, as depicted in Figure 3.11. To address this issue, implementing centralized delivery services can yield significant reductions in truck VMT for parcel deliveries within the city.

3.6 Chapter Summary

This chapter presented the data, methods, and estimation results pertaining to forecasting and quantifying the travel demand impact of e-commerce on Indiana residents under different future adoption scenarios. Using Indian as a case study region, we first conducted a market segmentation analysis for Hoosier's heterogeneous shopping behaviors (e.g., frequency) based on 2017 NHTS data and then, focused on Indianapolis residents based on consumer panel data. Second, we integrated the travel demand changes from e-commerce service (i.e., substitute in-store shopping trips) into a travel demand simulation model to estimate the net VMT impacts from personal travel flow, mode choice, and goods delivery. Different scenarios of potential e-commerce service adoption were also analyzed to estimate the future travel demand impact.

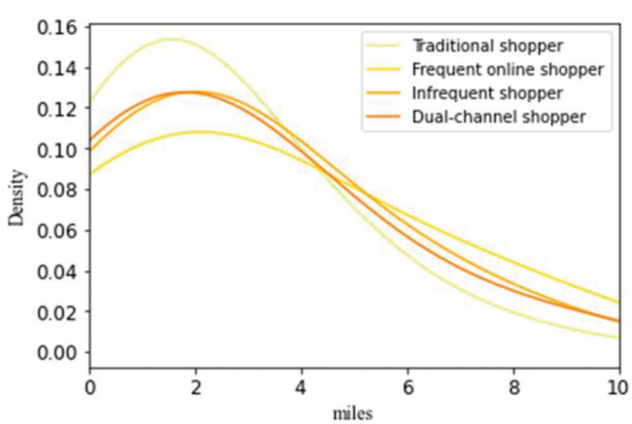


Figure 3.14 In-store trip distance by customer groups.

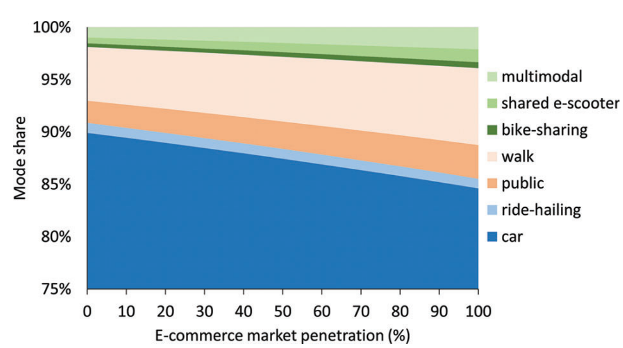


Figure 3.15 Mode split by e-commerce market penetration.

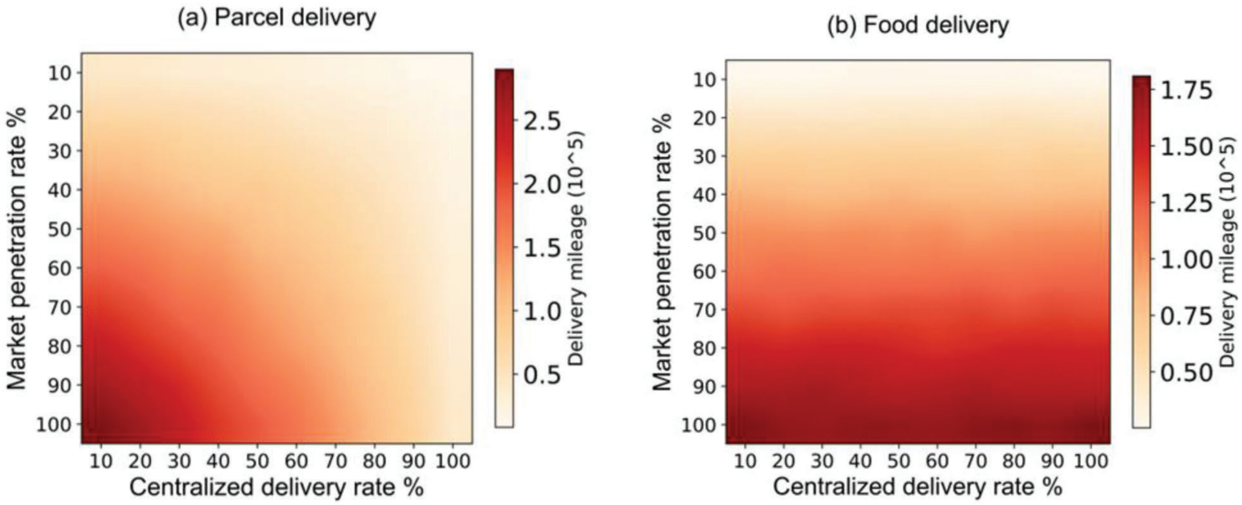


Figure 3.16 Delivery travel distance by scenarios.

NHTS data analysis showed that individuals aged 60 and over are less likely to shop online (negative coefficient for "old"). Higher-income individuals tend to order more goods online (positive coefficient for "rich"). Daily computer users exhibit a positive tendency towards e-commerce. Education and driving are also associated with a propensity for online shopping. The built environment influences online shopping behavior, with less dense areas showing higher online shopping propensity (<100 persons per sq mile) and dense neighborhoods showing lower propensity (>4,000 persons per sq mile). This confirms online shopping's potential to address accessibility gaps in rural areas. The model fit was good, explaining 41% and 46% of the variance for Indiana and Indianapolis, respectively.

Results from the consumer panel data analysis show that the current market penetration rate of e-commerce in Indianapolis is about 2%, indicating that only 2% of current shopping demand is satisfied by e-commerce services. Over 64% of Indianapolis residents are "traditional shoppers" who have never used e-commerce before. Market segmentation analysis highlights that internet access is a prerequisite for e-commerce. In Indiana, 18% of households are without internet connectivity, impeding their adoption of e-commerce. Results suggest a correlation between larger household sizes and the absence of internet access. Additionally, individuals with lower education levels are more likely to be the "traditional shoppers." Households with smaller household sizes and higher education levels are more likely to use online shopping.

Considering the future e-commerce adoption and its impact on travel demand, we did a scenario analysis using an agent-based simulation model. We found that e-commerce market penetration could reduce personal travel demand by 14% (about 600,000 miles/day) at most due to the transition from in-person to online shopping. The reduced travel demand can also encourage a shift in travel mode choice by decreasing the mode share of private cars by 4%, shifting to green transportation modes such as public transit and multimodal systems. Truck delivery can increase VMT by a maximum of 300,000 miles/day from parcel delivery and 180,000 miles/day from food delivery. In general, a centralized delivery system would consolidate orders and relax time constraints, enabling route optimization and VMT savings.

3.6.1 Limitations and Future Work

This study is subject to several limitations that are worth further investigation to better understand the travel demand impact of e-commerce. First, this study analyzes overall VMT changes without considering the transfer of personal vehicle mileage to 4-tire commercial vehicles (4-TCV) or non-freight trucks (NFT). Advanced analyses could assist INDOT in formulating informed policies by gauging e-commerce's nuanced effects. Second, the reliance on 2017 data from the NHTS and Nielsen consumer panel may not reflect the current situation in Indiana, given the swift evolution in e-commerce. However, the proposed model is adaptable and can be updated as new data emerges. Third, this study assumes that the in-store and online shopping trips are one-to-one substitutions, disregarding evidence that e-commerce could either replace or augment physical shopping trips. Future analysis of how e-commerce adoption substitutes and stimulates in-store shopping demand for different market segmentations can address this limitation in our assumption. Finally, externalities like crime, weather, technological access (credit card ownership, smartphone usage, internet access, e-commerce literacy), and economic climate (energy prices, interest rates, delivery costs, economic downturns) also shape travel patterns and e-commerce adoption but were not accounted for due to data constraints. A survey study in Indiana can collect the necessary information to complement the data input in this study and support modeling more complex consumer behaviors. In addition, incorporating economic factors into the scenario development can help build realistic e-commerce scenarios with a more comprehensive consideration of social and economic conditions.

4. FORECASTING COUNTY-LEVEL SHIFTS BASED ON SCENARIOS

4.1 Overview

In this chapter, we present a comprehensive implementation of a scenario-based approach, utilizing accepted growth models, to explore potential evolutions of future industry growth by county until 2050. Through a detailed county-level analysis, our focus centers on identifying favorable locations for industries and assessing their attractiveness based on various industry-relevant dimensions to examine the freight traffic. Additionally, we conduct a review of relevant research (Section 4.2) involving travel demand models, transportation plans, and freight plans from diverse states. In Section 4.3, we provide a thorough description of the data utilized in this study, encompassing AADT, freight traffic, passenger traffic, specified analysis zones, scenarios, as well as GDP and population data across those scenarios. The methods employed in our analysis are elaborated in Section 4.4, and the corresponding results and key findings are presented. Moreover, we introduce an optimization model in Section 4.5, which aims at determining whether the ramps, near industry cluster, in the identified counties that are attractive for industries, have enough traffic capacity to support the future growth in freight and passenger traffic across the scenarios. The optimization model uses *opensolver* to minimize the aggregate of transportation, construction, and delay cost to find the optimal solution. Finally, the chapter concludes with a summary in Section 4.6, encapsulating the main insights from our investigation.

4.2 Literature Review

Travel demand models, transportation plans, and freight plans of various states were studied and analyzed to understand the several factors these studies considered to forecast the vehicular movements for the future. This helped us identify the factors we could consider forecasting the passenger and freight movement in Indiana.

4.2.1.1 Washington STDM. We examined the STDM of Washington State (Metropolitan Washington Council of Governments, n.d.b) It inputs forecasts of future population, households, and employment in the region, which are updated periodically through the Council of Governments' (COG) Cooperative Forecasting Program (Metropolitan Washington Council of Governments, n.d.a). This data reflects the best judgments of local officials regarding the location of future housing, commercial, and industrial development within the region. Additionally, information about future transportation networks, including planned changes in infrastructure, is considered as input. The output of the modeling process includes estimates of vehicular volume and traffic speed on road segments, zone-to-zone trips by travel mode, origin-destination patterns by zone and travel mode, person trips/volumes on transit links, and car/truck emissions.

The model is a traditional four-step trip-based TDM, estimating travel at an aggregate zone level. Steps include trip generation, trip distribution, mode choice, and trip assignment. The model estimates vehicular volume, traffic speed, zone-to-zone trips by travel mode, origin-destination patterns, person trips on transit links, and car/truck emissions. The model's geographic unit of analysis is the TAZ, and it covers a modeled area of 6,800 mi^2 with ~3,700 zones. From this model, we considered factors such as population, employment data, and household data in our study, which can be used to forecast the travel demand.

4.2.1.2 Georgia STDM. Georgia uses a similar travel demand model to that of Washington, incorporating transportation network, land use, demographic, and socioeconomic attributes (SEA) data as inputs. Land use is represented by SEA statistics such as population, household, and employment by industry. On current land uses, including land development, as well as statewide or regional estimates of population, household, and employment, future year projections of SEA data are predicted. Forecasts for subsequent years consider important planned transportation enhancements, such as multi-model transportation improvements and upgrades to road capacity (HNTB Corporation, 2019).

The goal of the GSTDM is to create analytical tools that can assess the effects of modal diversion for people and products, significant adjustments to economic and land use regulations, and alternate modes of transport for people. The model enables the investigation of the effects of upcoming transportation infrastructure plans and investments. The model may also be used to test various project/corridor alternatives, aid with future MPO model upgrades, and update Georgia Department of Transportation's STDM. The GSTDM is composed of two main parts: a freight model and a passenger model. Except for the traffic assignment, when freight trucks, non-freight trucks, and passenger cars are merged to reflect the overall highway traffic conditions, each model completes the modeling processes independently. To assess future traffic conditions, the base-year model undergoes calibration and validation. Two scenarios are created for the future, with varying project changes while keeping other components consistent. The latest upgrade is projected for the year 2050. To create scenarios for the following years, socioeconomic data, freight inputs, and other model inputs are updated. The 2050 GSTDM socioeconomic data is derived from the Regional Economic Model (REMI) total population forecast and 2015 GSTDM SEA data. MPO zonal SEA data is incorporated for TAZs within the MPO modeling area to account for detailed growth trends. The freight component of the model requires additional future-year data for the 2050 scenario.

The GSTDM supplies external travel for the MPO models and anticipated future demand for travel, including internal-external travel as well as pass-through travel. It determines the effects of significant corridor upgrades, such as enlarging interstates or building new facilities. This can assist with policy-level analysis, such as study of toll facilities and freight diversion between truck and rail. Overall, this TDM helped us determine how socioeconomic data can be used to predict the future travel demand and how the model can be used to forecast travel demand separately for freight trucks and passenger cars.

4.2.1.3 Minnesota transportation planning to support development. The study aims to improve our understanding of the relationship between transportation and economic development by looking into how firms in competitive industry clusters use transportation networks and what role these networks play in the clusters' establishment and expansion. The method incorporates both quantitative and qualitative strategies for locating competitive industry clusters spatially and investigating the role of transportation. In 25 Minnesota metropolitan and micropolitan regions, the US Cluster Mapping tool is used to identify competitive clusters based on employment location quotients. Twelve competitive clusters were chosen for further investigation, and in-depth interviews and site visits with enterprises in each cluster were undertaken to determine the competitive importance of various modes of transportation. These methodologies can provide useful information about how transportation acts as an input within competitive industry clusters and how it can be used to guide economic development policies suited to specific areas and industries (Munnich et al., 2015).

The method used in this study, which involved examining industry clusters in various state regions and distinguishing between clusters that are more state-wide in scope, multi-regional clusters, and clusters that are focused on a single industry, suggests that various types of lessons can be learned from each. For instance, transportation planners may be able to establish a basis for broad-based regulations relating to freight movement or maintenance operations by considering the transportation demands of businesses in competing industry clusters spread throughout a state. The analysis of businesses inside single-industry clusters, on the other hand, may reveal information on highly specialized and regionally focused policies, such as initiatives to enhance regional airline service. Additional policy areas that may benefit from an industry cluster-based approach have also been highlighted, such as opportunities for cooperation between transportation authorities and local industry stakeholders on crucial funding and investment decisions, intermodal freight connections, and other policy areas.

4.2.1.4 Arizona State Freight Plan. The Arizona State Freight Plan identifies short- and long-term investment priorities and policies that maximized Arizona's economic return while simultaneously promoting other important transportation system goals, such as the MAP-21 (Moving Ahead for Progress in the 21st Century Act)-identified global objectives (ADOT, 2016). The plan identifies the Arizona freight transportation infrastructure that is essential to the state's economic development and gives investments in those facilities the proper level of priority. Scenario planning imagines potential futures and "back casts" them to the present, as opposed to forecasting, which projects historical tendencies into the future. Scenario planning offers a way to get ready for potential futures rather than attempting to foretell them, as is done in forecasting. The scenarios are designed to be quite extreme to cover a wide range of potential futures and give Arizona DOT a number of contexts in which to situate upcoming infrastructure improvements.

The competitiveness and expansion of Arizona's freight industries and related freight flows, as well as the state's overall economy, are influenced by a variety of factors. These elements can be categorized using the STEEP (Social, Technical, Economic, Environmental, and Political) drivers' framework. Five consequences of freight flow caused by STEEP drivers are further discussed in the report (ADOT, 2016). The impact of various future scenarios is also evaluated considering these five consequences. This is accomplished by evaluating each effect's projected influence in relation to the Base Case scenario estimates. To that end, this study discusses the different scenarios that can be used while forecasting the freight travel demand. This helped us to determine the scenarios that can be used while forecasting AADT for Indiana. This study also sheds light on the factors which can be considered to predict the freight flow for future.

4.3 Data Description: County-Level Analysis

In contrast to the analysis presented in the previous chapters, the data presented herein is at the macro-level (county-level). This provides INDOT with information that can be used to make the macro-level decisions. Such decisions would be applicable to larger and multiple regions, yielding benefits. Additionally, the data considered for forecasting and in the optimization model, such as GDP and population, are only current to 2021 or 2022, depending on the availability of the data for the parameters. The analysis can be updated as new data become available.

4.3.1 Freight Analysis Framework

The Freight Analysis Framework (FAF), produced through a partnership between Bureau of

Transportation Statistics (BTS) and FHWA, integrates data from a variety of sources to create a comprehensive picture of freight movement among states and major metropolitan areas by all modes of transportation. It provides data for the goods based on the 2017 Commodity Flow Survey (Hu et al., 2022). There are 132 mutually exclusive domestic regions defined across the US. These regions, referred to as the major regional economic centers, are selected by BTS and Census to provide a graphical representation and data of freight movement at a more granular level. Indiana comprises four major regional economic centers—(1) Chicago, (2) Indianapolis, (3) Fort Wayne, and (4) the rest of Indiana. These are described in appendices.

4.3.2 Annual Average Daily Traffic

The AADT data for all the routes in Indiana for the year 2018 are extracted from the Highway Performance Monitoring System (HPMS) database shape files (ArcGIS, 2018). The important variables of this dataset are defined in Table 4.1.

Factors like GDP and population were considered to identify the factors that impact traffic—both freight and passenger. These factors are used to determine the future traffic trends in Indiana. To proactively plan for future needs, it is pragmatic to consider these trends across various scenarios.

4.3.3 Shared Socioeconomic Pathways

In the late 2000s, teams of researchers, modelers, and climate scientists around the world began the process of developing new scenarios to explore how the world might evolve over the rest of the 21st century. These scenarios, which are coordinated with those used by International Protocol for Climate Change (IPCC), have been named Shared Socioeconomic Pathways (SSPs) (O'Neill et al., 2017). They represent different possible evolutions of the world's economies based on geo-political choices. It includes factors like population, economic growth, education, urbanization, and the rate of technological development. These SSP scenarios have been translated into quantitative estimates of future growth, enabling us to use them to build scenarios for future growth. As per (O'Neill et al., 2017), these are summarized as follows.

1. ***SSP 1*** *(A world of sustainability-focused growth and equality):* This scenario poses low challenges to mitigation and low challenges to adaptation. This future emphasizes human well-being and global population peaks mid-century. It boasts environmentally friendly technologies and renewable energy.
2. ***SSP 2*** *(A "middle of the road" world where trends broadly follow their historical patterns):* It poses moderate challenges to mitigation and moderate challenges to adaptation. The population growth stabilizes toward the end of the century. Current social, economic, and technological trends continue in the future for this scenario.
3. ***SSP 3*** *(A fragmented world of "resurgent nationalism"):* It poses high challenges to mitigation and high challenges to adaptation. This future emphasizes national issues due to regional conflicts and nationalism. The population growth continues with high growth in developing countries in this scenario. Economic development is slow and fossil fuel dependence is high in this future.
4. ***SSP 4*** *(A world of ever-increasing inequality):* It poses low challenges to mitigation and high challenges to adaptation. This future sees a growing divide between globally connected, well-educated society and fragmented lower income societies. The population growth stabilizes toward the end of the century. Unrest and conflict become more common in this scenario.
5. ***SSP 5*** *(A world of rapid and unconstrained growth in economic output and energy use):* This scenario poses high challenges to mitigation and low challenges to adaptation. This future emphasizes economic growth and technological progress and global population peaks mid-century. It boasts global adoption of resource- and energy-intensive lifestyles with a lack of environmental awareness.

The SSP public database provides GDP forecasts for the world and the US from 2010 through 2100 at an interval of 5 years. Their trends over the 21st century are shown in Figure 4.1.

TABLE 4.1
Important variables in the AADT data for 2018 obtained from the HPMS

Variable	Description
	Original Variables
ROUTE_ID	Identifier of the route segment (facility)
COUNTY_CODE	Identifier of the county containing this facility
AADT	AADT on the entire facility
AADT_COMBINATION	AADT of combination trucks
AADT_SINGLE_UNIT	AADT of single unit trucks
	Derived Variables
LENGTH	Length of the route (miles), derived from the shapefile in ArcGIS
TRUCK_AADT	Total truck traffic, computed as AADT_SINGLE_UNIT + AADT_COMBINATION
PAX_AADT	Total passenger traffic, computed as AADT – TRUCK_AADT
{TRUCK/PAX}_DVMT	Passenger or truck vehicle miles traveled, given by {TRUCK/PAX}_AADT * LENGTH

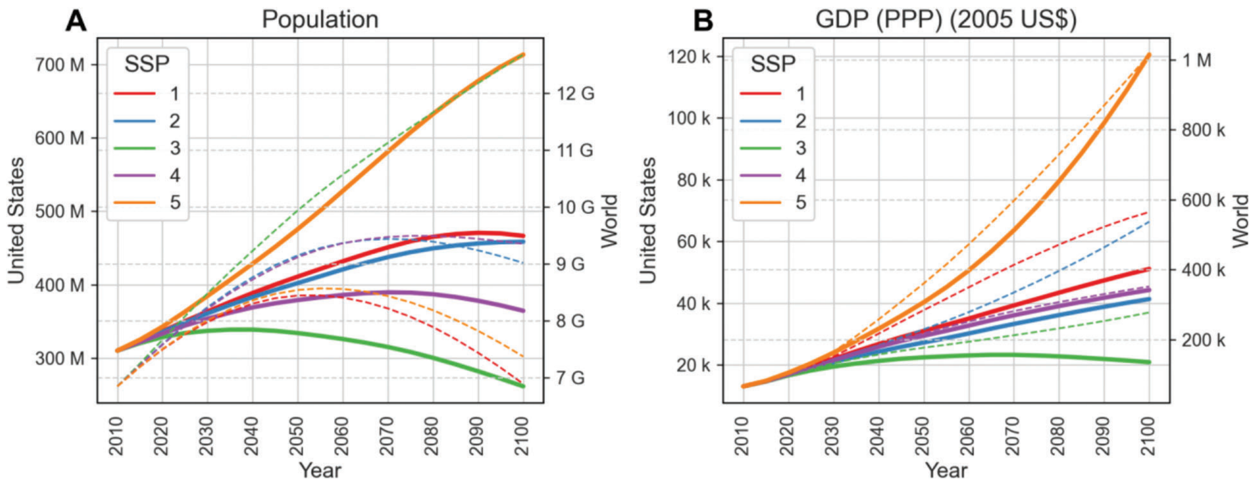


Figure 4.1 Forecasts of (A) population and (B) GDP of the US (solid) and the world (dashed) across the five SSP scenarios based on the OECD model. Adapted from the SSP database.

4.4 Methods

The goal of the task is to identify the trend in both freight and passenger traffic and forecast AADT for future to plan and implement construction projects proactively.

4.4.1 Forecasting Freight Traffic

To understand freight traffic, we analyze the freight associated with every industry. From the literature, we determine that freight traffic is dependent on socio-economic parameters like GDP. The GDP of a region is made up of several factors including economic activity due to freight and passenger traffic. Within this framework, the region's GDP is considered as the driving factor for freight traffic; an increase in GDP correlates with an increase in employment, economic activity, and production of goods and services. This in turn can lead to an increase in freight movement. To forecast the freight traffic across five SSP scenarios, we first calculate the industry level GDP across the five scenarios.

Historical GDP data for US and Indiana, as shown in Figure 4.2(a), was obtained from (BEA, 2023b). We fit a linear regression model to obtain the mean proportion of the GDP of the US contributed by Indiana. This comes out to be 1.84% (see Figure 4.2b).

We then scale down the US-level GDP figures for the five SSP scenarios by this percentage to estimate Indiana's future GDP across these SSP scenarios. These are shown in data as represented in Figure 4.3.

Further, we determine the forecasted GDP contributions of each county in Indiana across the five scenarios. We obtained county GDP data, in thousands of current dollars, from BEA from 2001 through 2021 (BEA, 2023a). We calculated the contribution of each county GDP to the Indiana GDP by considering the weighted average of the county-wise GDP of the last 5 years, with the weights assumed such that more weightage is given to more recent year, as shown in Table 4.2. The resulting weights of the counties are shown in Figure 4.4.

Using the forecasted Indiana GDP across the scenarios and the weighted average county GDP contribution ratio, we determined the forecasted county GDP across SSPs. A sample of this data is shown in Figure 4.5.

We obtained Indiana's historical GDP data from the US Bureau of Economic Analysis (BEA, 2023a) for the years 2001 to 2020. This data provides detailed industry contribution to each County's GDP in Indiana. The GDP is in thousands of current dollars (not adjusted for inflation). The industries are identified based on the 2012 North American Industry Classification System (NAICS). In the dataset, the "all industry total" is divided into "private industries" and "government and government enterprises."

This data was obtained for all the 92 counties in Indiana from 2001 through 2021. Once the industry-level GDP data was obtained, the contribution of each industry GDP to county GDP was calculated, determining the industry GDP ratio. A weighted average of the industry GDP ratio was determined using the weights in Table 4.2. These ratios, used with forecasted county GDP, provide forecasted industry-level county GDP from 2025 to 2050 across five SSP scenarios.

To forecast industry-wise freight AADT, a linear regression was performed where the industry-level county GDP and FAF zone data is considered as the dependent variable for the independent variable – freight AADT. This regression yields the below equation with an $R^2 = 0.81525$.

$$Freight\ AADT = -377.74 * GDP_{construction} + 26.23 * GDP_{Manufacturing} + 1{,}327.16 * GDP_{Retail\ Trade} + 440.72 * GDP_{Transportation\ \&\ Warehousing} - 492.11 * GDP_{Wholesale\ Trade} + 2{,}242.56 * Zone_1 + 921.99 * Zone_2 + 1{,}576.11 * Zone_3 + 741.64 * Zone_4 \quad \text{(Eq. 4.1)}$$

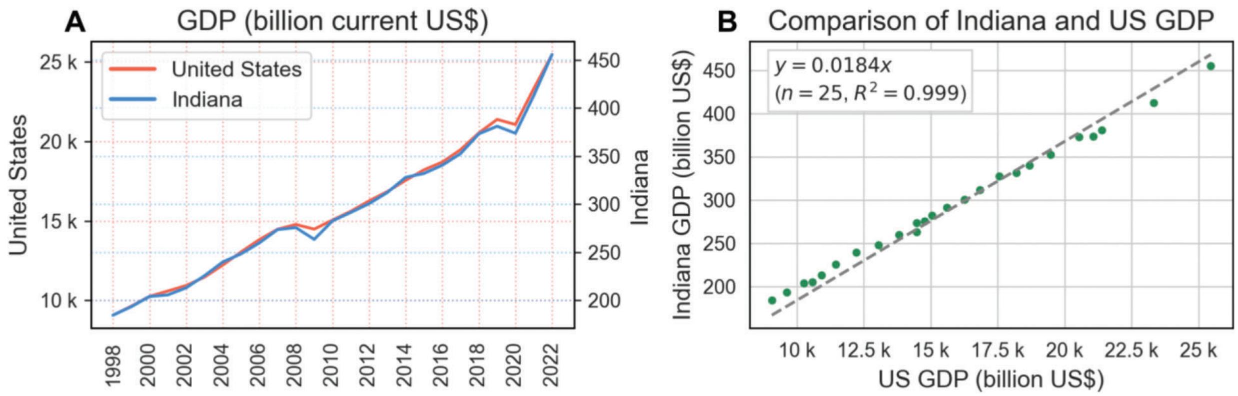


Figure 4.2 Relationship between the historical GDP of Indiana and the US (Data source: US Bureau of Economic Analysis, 2023b): (A) Annual GDP figures, and (B) results of ordinary linear regression.

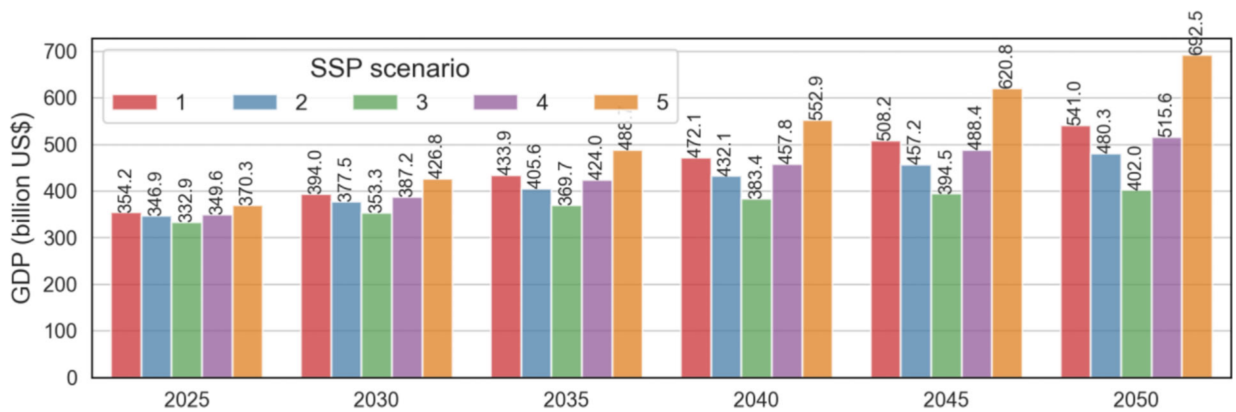


Figure 4.3 Predicted GDP of Indiana up to 2050 for the five SSP scenarios.

TABLE 4.2
Weights for the past 5 years of GDP

Year	Weight
2021	0.45
2020	0.25
2019	0.15
2018	0.10
2017	0.05

The forecasted freight AADT based on county-level industry GDP is presented in Section 4.5 and used in the optimization model.

4.4.2 Forecasting Passenger Traffic

A similar approach to forecasting freight traffic in Section 4.4.1 is used to forecast passenger traffic. Based on the literature review, we determine that population has a very strong linear relation with passenger traffic. This is verified by performing a county-level linear regression analysis of the Indiana population with passenger traffic, which is obtained from Section 4.3.2, for the year 2018, as shown in Figure 4.6.

Using this relationship, future forecasts for county-level passenger AADT are developed across the five SSP scenarios.

Scenario-wise population data is obtained from the SSP database for world- and US-level, as shown in Figure 4.1A. This data is used to establish a relationship between the US and Indiana population, which is then used to forecast IN population for the SSP scenarios. The historical population data for the US and Indiana is obtained from the US Census Bureau for the years 2000–2021. Figure 4.7 represents the data and the relationship between US and IN population.

This relationship is used to determine the future IN population across SSP scenarios. Table 4.3 represents the forecasted IN population from 2025 through 2050 across five SSP scenarios at an interval of 5 years.

The county-wise population for Indiana counties is also obtained from the US Census Bureau. The population ratio of the county population to state population is determined for 2000–2021. A weighted average population ratio is calculated for each county for the last 5 years, using the weights in Table 4.2. Based on the county population ratio, the county population is forecasted across the five SSP scenarios using IN population shown in Table 4.3. Using the county population and passenger AADT relationship defined in Figure 4.6, the county-level passenger AADT is forecasted for five SSP scenarios from 2025 through 2050. The forecasted passenger AADT based on

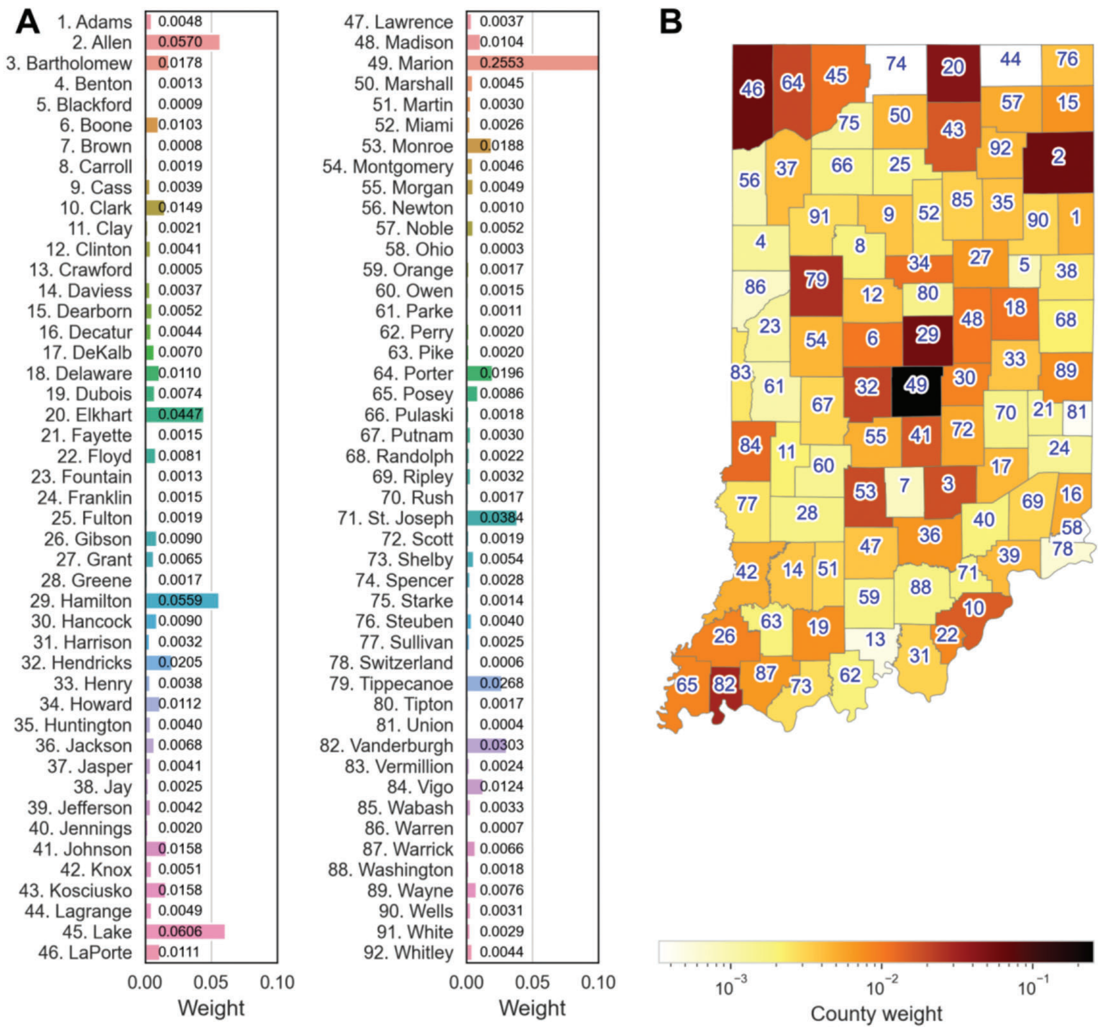


Figure 4.4 GDP county-weight distribution in Indiana: (A) list, and (B) map.

Scenario	County	Unit	2025	2030	2035	2040	2045	2050
SSP1	Adams, IN	Billion Current Dollars	$ 1.707	$ 1.899	$ 2.091	$ 2.275	$ 2.449	$ 2.607
SSP2	Adams, IN	Billion Current Dollars	$ 1.672	$ 1.819	$ 1.955	$ 2.083	$ 2.204	$ 2.315
SSP3	Adams, IN	Billion Current Dollars	$ 1.604	$ 1.703	$ 1.782	$ 1.848	$ 1.901	$ 1.938
SSP4	Adams, IN	Billion Current Dollars	$ 1.685	$ 1.866	$ 2.044	$ 2.207	$ 2.354	$ 2.485
SSP5	Adams, IN	Billion Current Dollars	$ 1.785	$ 2.057	$ 2.355	$ 2.665	$ 2.992	$ 3.337

Figure 4.5 Forecasted county GDP across SSPs.

population is provided in Section 4.5 and used in the optimization model.

4.4.3 Industry Attractiveness Criteria

Industry attractiveness illustrates the profit potential for a particular industry, wherein a new business can set foot and compete in that industry. The higher the potential of the industry, the higher will be the attractiveness of that specific geographical area. A superior industry attractiveness implies that more investors are attracted to invest in that industry in that county, since it has a higher potential to yield profits.

The industry attractiveness analysis aims to identify Indiana's fastest growing and promising counties for specific industries. We focus on industry-wise desirable locations and evaluation of their attractiveness on a variety of industry-relevant factors. Factors that make up this industry attractiveness criteria and are common across all the industries are workforce engaged in the industry, largest companies in the sector, largest growth rate for the past 20 years, weekly wages, and proximity to interstates passing through the state. Figure 4.8 provides information about the largest firms in Indiana by employment in five major sectors. Figure 4.9 displays the top counties by percentage employment

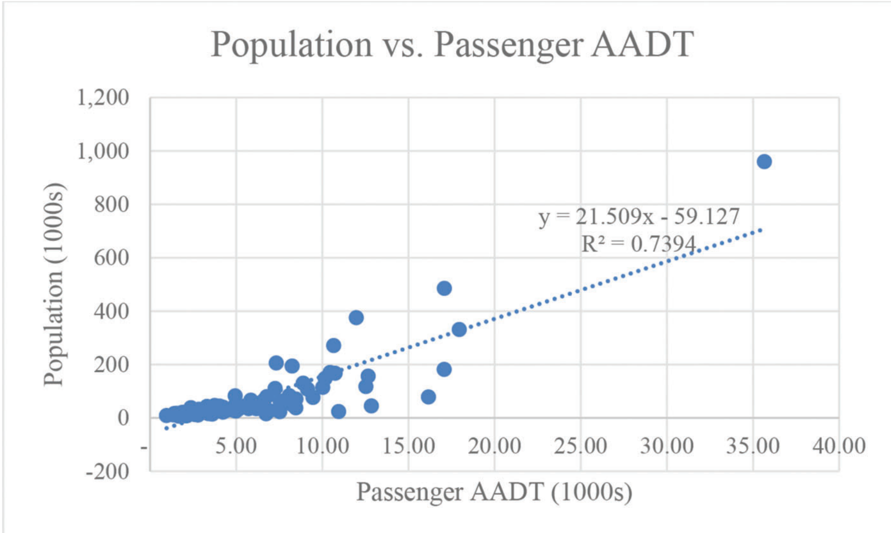


Figure 4.6 Population vs. passenger AADT relationship.

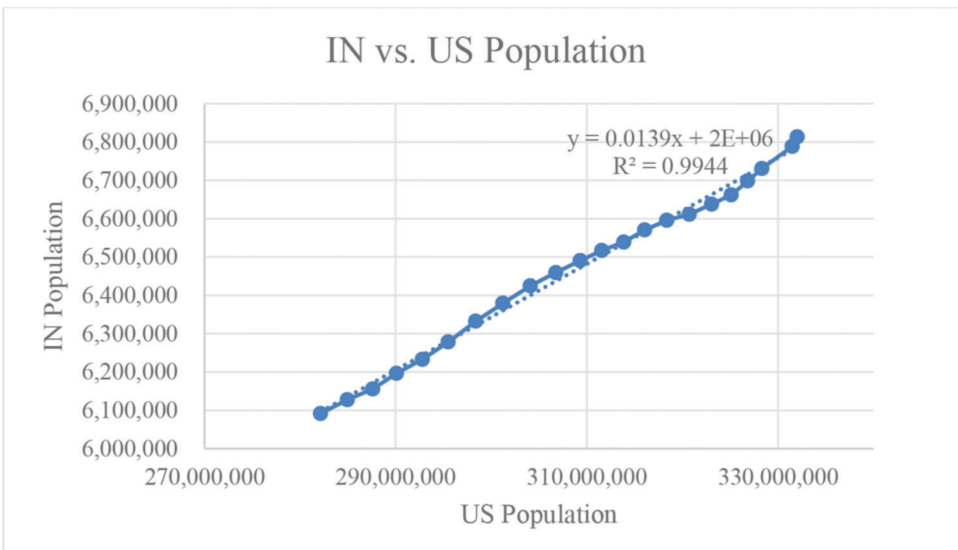


Figure 4.7 IN vs. US population.

TABLE 4.3
Forecasted population of Indiana for different SSP scenarios

Scenario/Year	2025	2030	2035	2040	2045	2050
SSP1	7,042,182	7,226,058	7,403,834	7,574,079	7,732,674	7,882,504
SSP2	7,018,233	7,189,242	7,348,942	7,496,930	7,632,635	7,761,212
SSP3	6,818,130	6,861,073	6,882,546	6,882,838	6,860,102	6,817,904
SSP4	6,959,646	7,088,582	7,204,029	7,302,405	7,380,317	7,440,397
SSP5	7,221,342	7,518,956	7,819,586	8,126,993	8,444,074	8,778,476

in the five major sectors. While Figure 4.10 illustrates top 20 counties by growth rate in the past 20 years across the five major sectors.

This analysis will consequently facilitate INDOT to synchronize planning of their future infrastructure development projects with anticipated industry growth plans to ensure both efficiency and industry competitiveness.

The top 5 freight traffic contributing industries and the corresponding most suitable counties in the state were identified based on industry GDP data and the industry attractiveness criteria. The top industries are

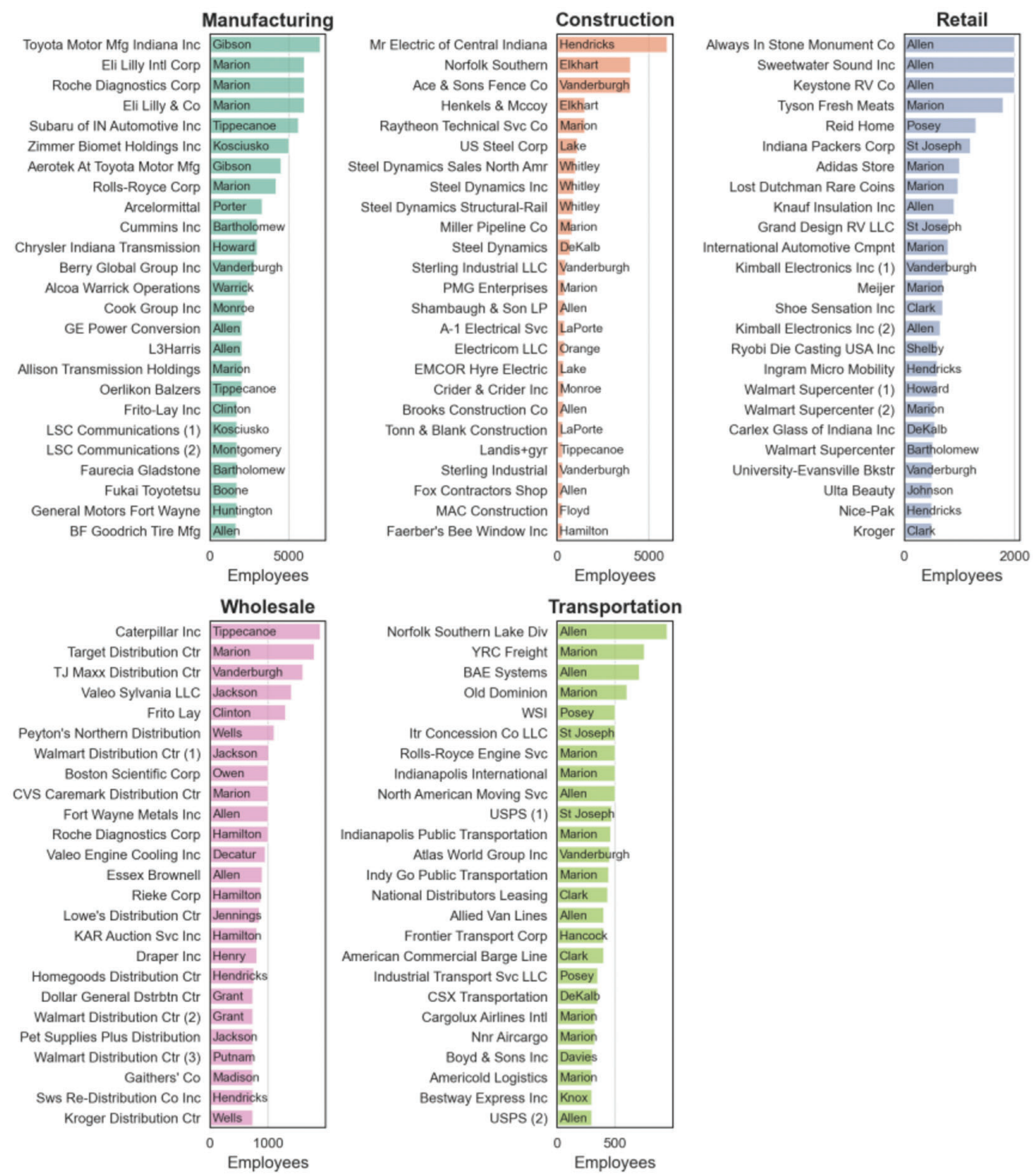


Figure 4.8 Largest firms in Indiana by employment in five major sectors, along with the county of their head office/location.

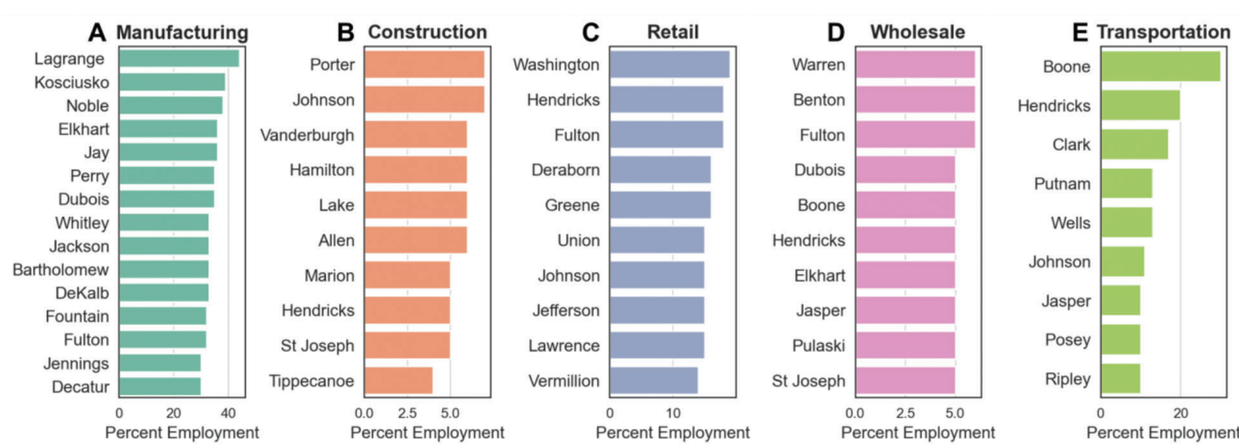


Figure 4.9 Top Indiana counties by percentage employment in five major sectors.

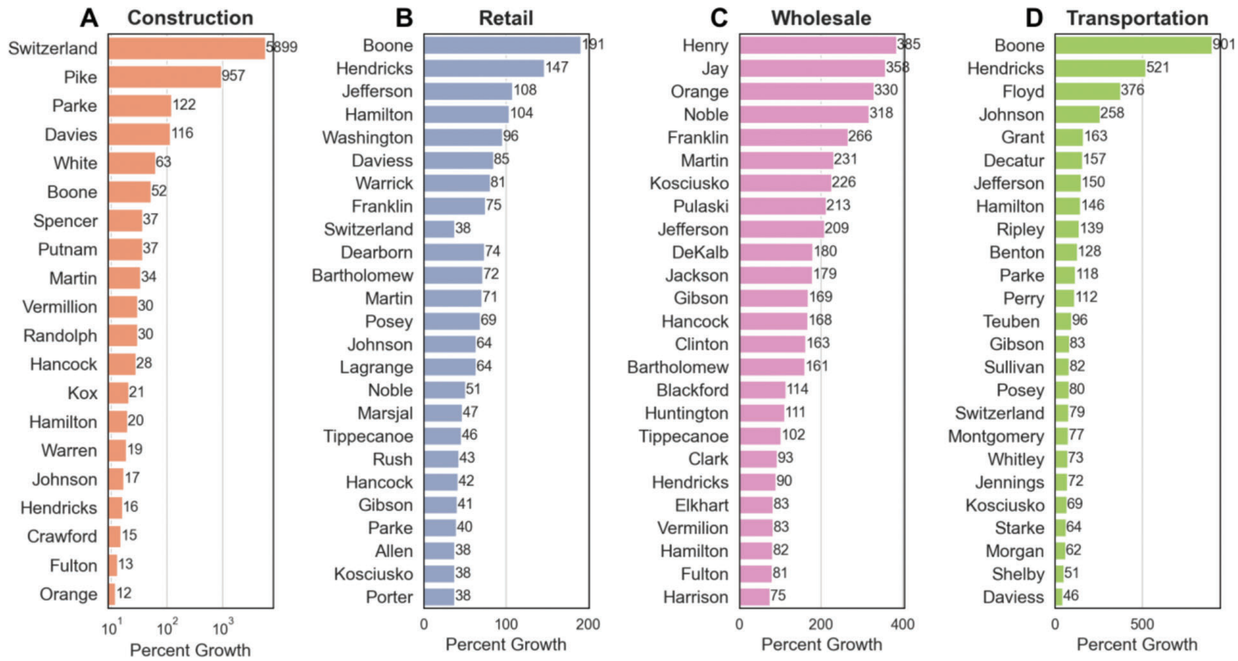


Figure 4.10 Top 20 counties by growth rate in the past 20 years in five major sectors.

construction, manufacturing, retail trade, wholesale trade, transportation, and warehousing.

4.4.3.1 Manufacturing industry. The manufacturing industry in Indiana has seen tremendous growth over the past couple of decades. The state boasts of more than 7,000 manufacturing firms—large scale and small scale, and over 90,000 manufacturing jobs. According to the 2022 report released by Conexus Indiana (Conexus Indiana, 2022), the state has the largest number of manufacturing jobs in the country, with over 520,000 Hoosiers employed in the sector. According to East Central Indiana Regional Partnership (n.d.), Indiana is one of the best states in the country to do business since the Indiana department of revenue reported in July 2021 that the corporate income tax rate dropped to as low as 4.9%. As a result of the low tax rate, coupled with generous grants, which continued even during the Covid-19 pandemic, the manufacturing industry has continued to grow exponentially. Indiana Economic development Corporation (IEDC) reported that a $10 million grant was offered by the Economic Activity Stabilization and Enhancement program to promote research and technological advancements in the industry.

To determine the counties that display tremendous growth potential for manufacturing industry over the next 20–30 years, we looked at factors like workforce availability, locations of existing manufacturing firms, weekly wages, and growth of the manufacturing industry for the past 20 years. Counties that looked promising across all criteria and would contribute the most to freight and passenger traffic over the next 30 years, were identified as attractive counties for the manufacturing industry.

Largest firms: The state is home to some of the biggest manufacturing firms across USA, with companies like Eli Lily and Co. (>10,000 employees), Lippert Components, Inc. (7,500 employees), Subaru of Indiana Automotive, Inc. (6,200 employees), Toyota Motors (5,400 employees), etc.

Gibson County tops the chart with a 330% industry growth over the last 20 years and for being home to Toyota Motor Manufacturing, one of the largest manufacturing firms located in Indiana. Gibson County is followed by Boone (146% growth, Fukai Toyotetsu), Porter (118% growth, Arcelormittal), Tippecanoe (107% growth, Subaru, Oerlikon Balzers), Bartholomew (90% growth, Cummins, Faurecia) and so on.

Wages: Counties with higher weekly wages tend to attract more people into the workforce. Bartholomew, Tippecanoe, Kosciusko are some counties that consistently feature in the above criteria and report high weekly wages as well as per the US Bureau of Labor Statistics.

Employment: Lagrange County is the leader with 44% of its workforce engaged in manufacturing, followed by Kosciusko with 39%, followed by Noble with 38%, Elkhart with 36% and so on. The availability of skilled labor facilitates the growth of respective industries.

Proximity to highways: The interstate highway system in the US is one of the biggest transportation networks in the world. According to the public purpose (Cox & Love, 1996) this highway system facilitates the movement of nearly 23% of all roadway traffic, coming on top of rail services by a huge margin. Indiana's robust interstate network acts as a catalyst for up-and-coming businesses, existing businesses, and economic

Figure 4.11 Interstate system of Indiana highlighted in red, along with the counties chosen for attractiveness for the five major industries (data source for roads and county boundaries: Indiana map.org, n.d).

growth. Figure 4.11 represents the interstate and highway network in Indiana.

Bartholomew County, Jackson County, and Perry County are highlighted in the figure as the counties with the highest potential for future expansions in the manufacturing sector due to their proximity to interstates and consequently convenience for transportation of goods. Perry County has a high unemployment rate, which could be another factor contributing to industry attractiveness.

Takeaway: Based on the criteria mentioned above, Bartholomew County, Elkhart County, Jackson County, Perry County, and Whitley County are the counties with potentially the highest industry attractiveness for the manufacturing sector for the next three decades.

4.4.3.2 Construction industry. Construction and infrastructure development are indicators of the economic health of a state. Over the last couple of years, Indiana has seen huge investments in construction, ranging from residential construction to non-residential constructions—for a variety of upcoming industries. The capital invested in non-residential construction averaged approximately $500 billion per month in the US for the first half of 2022, as reported by Statista (Statista Research Department, 2023). As per Build Indiana Council (BIC, 2023), INDOT has invested over $2.3 billion in construction projects and bids in 2022. This capital investment is expected to shoot up to $2.846 billion in 2023.

Trends in construction are critical since they facilitate job creation which translates into higher revenue for the state government. These trends mirror the trends in other dependent industries like manufacturing, warehousing and transportation, real estate, supply chain, etc. A report published by the Economic Policy Institute in 2019 (Bivens, 2019) published data which stated that for every 100 jobs created in the construction sector, 88 additional jobs are created in allied industries, which include goods and services providers, building materials suppliers, etc. and a total of about 226 indirect jobs.

Largest firms: Mr. Electric of Central Indiana from Hendricks County tops the list with 6,000 employees, followed by Norfolk Southern from Elkhart with 4,000 employees, followed by Ace and Sons Fence Co. from Vanderburgh County with 4,000 employees along with others.

Growth rate: Figure 4.10 illustrates the counties in Indiana that demonstrated the greatest expansion in the construction industry during the past two decades. Switzerland County tops the list, with an astounding growth rate of 5,899%. It is followed by Pike, Parke, Davies, and White counties with 957%, 122%, 116%, and 63% growth, respectively.

Employment: Porter County has the highest number of workers, consisting of 4,206 individuals, who are employed in the construction industry. This constitutes approximately 7% of Porter's total workforce. Johnson county also has 7% of its entire workforce employed in the construction industry. Vanderburgh, Hamilton, Lake, and Allen counties also feature among the top 6 counties where construction is a significant industry, with 6% of their total employees working in this field.

Proximity to highways: Marion, Hamilton, and Hendricks counties are well-connected to the rest of the state with I-69 passing through Hamilton and I-70, I-74 passing through Hendricks. Vanderburgh and Lake counties have close proximity to I-164 and I-80, respectively. This strong network of roadways provides efficient travelling options and helps the construction industry maintain its competitive edge.

Takeaway: Considering the above factors, Marion County, Lake County, Hendricks County, Hamilton County and Vanderburgh County are the counties with potentially the highest industry attractiveness for the construction sector.

4.4.3.3 Transportation and warehousing industry. The warehousing and transportation sector was identified as the fastest growing in the country as well as in Indiana over the past 10 years. According to Langham Logistics (2019), Indiana has quickly become a hub for warehousing and transportation needs in the country. The Indiana Business Research Center reported (Strange, 2023) that this sector grew by 36.8% from 2011 to 2021, which translated into 46,116 new jobs in Indiana.

The state has an extensive network of major highways, rail lines, and a good international port system. The port system works efficiently due to access to the Great Lakes and the Ohio-Mississippi rivers. With over 12 airports, 14 interstates and more than 4,000 miles of rail availability, warehousing and transportation is a promising sector of growth for the state.

Largest firms: Norfolk Southern Lake Division is the largest firm in this sector based out of Allen County, with 950 employees and annual sales of over $150,000,000. Norfolk is followed by YRC Freight, situated in Marion County with 750 employees, followed by BAE Systems from Allen County with 710 employees, and Old Dominion from Marion County with 600 employees among the top. It is evident that Marion and Allen counties are home to some of the largest companies from this sector and provide the bulk of the employment opportunities in the industry.

Growth rate: Boone County saw the highest growth with a staggering growth of 901% (Figure 4.10d). The top counties comprise of Hendricks County, Floyd County, Johnson County, and Grant County, respectively. A healthy growth rate of over 50% is demonstrated by the top 25 counties in this category, which is promising for the future.

Employment: With 30%, Boone County has the highest percentage of workforce engaged in the transportation and warehousing sector, followed by counties Hendricks (20%), Clark (17%), Putnam (13%), and Wells (13%) (Figure 4.9e).

Proximity to highways: Marion County is at the epicenter of the network of interstates and is very well connected to the other states in the US. The interstates I-70 and I-74 pass through Hendricks County, while I-64 passes through Posey County (Figure 4.11). Allen County is well linked as well with I-69 passing through.

Job availability: Indiana has over 127,000 jobs in transportation and logistics, which is 44% above the national average. A huge chunk of these jobs, i.e., over 45% are restricted to the truck transportation subsector, which is central to the forecast model we've built. The top five counties with the highest availability of jobs in this sector are Marion, Hendricks, Lake, Allen, and Clark counties, respectively (Figure A.7).

Takeaway: Considering the factors mentioned above, Marion County, Hendricks County, Allen County, and Posey County are the counties with potentially highest industry attractiveness for transportation and warehousing sectors.

4.4.3.4 Retail industry. Retail is the largest private sector employer in the US, which is mirrored in Indiana as well. Retail trade is central to Indiana's economic growth and its flourishing economy. As per data released by the National Retail Federation (NRF, 2014) in 2013 alone, a record 623,000 jobs were created in the state in the retail and restaurant business.

This industry shares a symbiotic relationship with other sectors—for every 1,000 retail jobs, 384 jobs are created in allied sectors, according to the same report. Retailers based out of Indiana also share over 100 subsidiaries in foreign countries.

Largest firms: Always in Stone Monument Co. is the largest retailer from Elkhart County with 2,000 employees and over $312,834,000 in annual sales. It is followed by Sweetwater Sound Inc. from Allen County and Keystone RV Co. from Elkhart County among the top retailers. The list also features some big retailers from Marion and Hendricks counties like Adidas Store, Lost Dutchman Rare Coins, Ingram Micro Mobility, among others.

Highest growth: Figure 4.10 displays the counties with the highest growth percentage in the past two decades. Boone County tops the list with a 191% growth, followed by Hendricks County with 147% growth, Jefferson County with 108% growth, Hamilton County with 104%, and Washington County (96% growth) which make up the top 5 growing counties. Dearborn County also features in the list in the tenth position with a growth of 74% in retail in the last 20 years.

Employment: Boone County tops the list with 30% of the workforce involved in the retail trade sector in 2021 followed by Hendricks, Clark, Putnam, and Wells counties with 20%, 17%, 13% of the workforce, respectively.

Retail trade is one of the largest private employment sectors in Indiana, which is depicted in the healthy workforce employed numbers in this sector in Figure 4.8 and Figure 4.9, 19% of the total employee workforce in Washington County is engaged in retail trade. Washington is closely followed by Hendricks and Fulton counties with 18% of its total workforce employed in this sector, followed by Dearborn and Greene counties with 16% each.

Over 17 counties in Indiana display great affiliation to this sector with more than 10% of their workforce employed in retail trade.

Proximity to highways: As mentioned in the earlier sections, Marion and Hendricks counties are very well connected to the rest of the states by virtue of the network of interstates passing through the counties. I-74 passes through Dearborn County, making it a lucrative option for the retail industry.

Takeaway: Based on the four key factors discussed, the counties with the maximum growth potential in retail trade in the subsequent years are Marion County, Dearborn County, Washington County and Hendricks County.

4.4.3.5 Wholesale trade. The wholesale trade sector consists of institutions involved in wholesaling goods, and providing services allied with the sale of the goods. Wholesale trade can be viewed as an intermediate step in distribution of the goods to the end consumer. This trade typically functions from the bounds of a warehouse, with minimal or no display of merchandise.

Largest firms: Caterpillar Inc., located in Tippecanoe County, is the largest wholesale trader in Indiana with 1,900 employees and 361,429,000 in total annual sales (Figure 4.8). Target Distribution Ctr, TJ Maxx Distribution Ctr, Valeo Sylvania LLC, Frito Ray, Walmart are some other big wholesale traders in the State. Walmart Distribution Ctr. is in Jackson County.

Highest growth: Figure 4.10 depicts the counties with the highest growth percentage in the last 20 years. Henry, Jay, Orange, Noble, and Franklin counties feature in the top five of this category with 385%, 358%, 330%, 318%, and 266% growth, respectively. Jackson County (179%) and Tippecanoe County (102%) also display healthy growth rates.

Employment: Warren, Benton, and Fulton counties share the top spot with 6% of their total workforce engaged in wholesale trading industry in 2021. There are several counties with 5% of their workforce engaged in this sector like Jackson, Dubois, Boone, etc.

Proximity to highways: I-65 passes through Tippecanoe while Jackson County is at the crossroads of I-65 and US Hwy 50.

Takeaway: Tippecanoe County and Jackson County have numerous factors acting in their favor, which makes them the top picks for anticipated industry growth in wholesale trade in Indiana.

4.5 Optimization Model

4.5.1 Overview

The primary goal of this task is to identify the need for increasing or adding ramp capacity to meet the increasing future demand driven by the increasing economic growth in various counties across the state of Indiana. This is done while minimizing the overall costs associated with construction and travel, including delay cost due to congestion on ramps. To achieve this goal, we have developed an optimization model that considers the specific industry and population clusters in the county. The model developed essentially considers the interchanges/major highways around the industry and population clusters, assuming that all the freight traffic related to the industry and the passenger traffic related to the population operates on those roads. By finding efficient solutions to mitigate congestion-related expenses, this endeavor aims to improve transportation efficiency, support economic growth, and enhance the overall quality of life in the region. The optimization model uses *opensolver* which offers multiple solvers but primarily uses COIN-OR CBC optimization engine, which is an open-source solver.

To ensure a comprehensive understanding of the problem's dynamic nature, we have incorporated five potential scenarios representing different ways in which the world may evolve in the projected years. It is often challenging to accurately predict any one scenario. Hence, considering multiple scenarios and developing a model that would provide results for any scenario that might occur would enhance decision making and reduce the margin of error. These scenarios serve as valuable insights for better preparedness in addressing future challenges effectively. The models have been designed with an equal probability for all scenarios, ensuring a fair assessment of each possibility.

Identifying the most suitable county for each industry sector's freight traffic is essential. Based on our analysis, the major contributors to freight traffic in Indiana are retail, manufacturing, construction, transportation and warehousing, and wholesale trade industries. To determine which counties, attract freight traffic from these industries, we established the "industry attractiveness" parameter. This parameter comprises four criteria that assist in selecting the most suitable county for each industry sector's freight traffic. The criteria involve evaluating the presence of the top 25 companies of a specific industry, assessing the industry's growth over the past 20 years, understanding the workforce distribution, and identifying counties with favorable highway proximity.

Note that the model developed for the study only considers a single industry for traffic data. The model can further be expanded to consider multiple industries which have potential for growth in the counties. Further, this chapter only discusses the analysis and results for a single county as a sample. INDOT can use this optimization model and the approach to develop the analysis for other counties in the state as part of future scope based on counties of interest or priority.

4.5.2 Model Description

The following Mixed Integer Programming (MIP) model is defined to minimize the overall cost across all scenarios.

$$\min\left\{\sum_{i,j} Y_{i,j}P_{i,j} + \sum_{i,j,m,k,t} CD_{i,j,m}X_{i,j,m,k,t} + UD \cdot \sum_{i,j,k,t} EX_{i,j,k,t}\right\} \quad \text{(Eq. 4.2)}$$

$$\sum_{j} X_{i,j,m,k,t} = F_{i,m,k,t}\ \forall\ i,m,k,t \quad \text{(Eq. 4.3)}$$

$$EX_{i,j,k,t} \geq \sum_{m} X_{i,j,m,k,t} - (C^0_{i,j} + Y_{i,j} \cdot CN_{i,j})\ \forall j,k,t \quad \text{(Eq. 4.4)}$$

$$AXE_{i,j,k,t} \geq EX_{i,j,k,t} - CX_{i,j}\ \forall j,k,t \quad \text{(Eq. 4.5)}$$

$$\sum_{i,j,k,t} q_k \cdot AXE_{i,j,k,t} \leq E_{cong} \quad \text{(Eq. 4.6)}$$

$$Y_{i,j} \in \{0,1\}\ \forall\ i,j \quad \text{(Eq. 4.7)}$$

$$X_{i,j,k,m,t} \geq 0\ \forall\ i,j,k,m,t \quad \text{(Eq. 4.8)}$$

$$EX_{i,j,k,t} \geq 0\ \forall\ i,j,k,t \quad \text{(Eq. 4.9)}$$

$$AXE_{i,j,k,t} \geq 0\ \forall\ i,j,k,t \quad \text{(Eq. 4.10)}$$

Where:

- i: County index.
- j: Index of ramp into county i; j = 1, 2, …, $R(i)$.
- k: Scenario index; k = 1, 2, …, K.
- t: Time step; t = 1, 2, …, T.
- m: Index of industry and population cluster.
- q_k: Probability of occurrence of scenario k.
- $F_{i,m,k,t}$: Traffic headed to cluster m in county i in period t for scenario k.
- $X_{i,j,m,t}$: Flow decision, denoting the flow through ramp j in county i headed to industry cluster m for scenario k at time period t.
- $CD_{i,j,m}$: Cost associated with the distance traveled from ramp j to cluster m in county i.
- $C^0_{i,j}$: Base capacity of ramp j in county i.
- $CN_{i,j}$: Capacity that can be added to each ramp j in county i. Additional capacity available for t periods and all scenarios.
- $P_{i,j}$: Cost associated with increasing capacity of each ramp j in county i by $CN_{i,j}$.
- $Y_{i,j}$: A binary variable when set to 1, increases capacity by $CN_{i,j}$.
- $EX_{i,j,k,t}$: Surplus flow beyond the capacity on ramp j in county i during period t and under scenario k. It quantifies the level of congestion, where a higher value indicates a greater degree of congestion.
- UD: Disutility associated with congestion. It is introduced to encourage flows to stay within capacity limits, aiming to mitigate congestion effects.
- $CX_{i,j}$: Target congestion for ramp j in county i.
- $AXE_{i,j,k,t}$: Observed downside i.e., congestion beyond the level $CX_{i,j}$ in period t.
- E_{cong}: Projected level of congestion that surpasses the threshold values $CX_{i,j}$ considering all scenarios and exits.

In the optimization model, each ramp is initially assigned a base capacity, measured in vehicles per hour. Using traffic data, congested ramps can be upgraded to increase their vehicle capacity, incurring an associated ramp cost. Additional capacity and ramp cost are vital user-input variables that depend on various factors. CX, another significant parameter, defines an acceptable delay that avoids negative impacts on competitiveness. Alongside CX, UD and E_{cong} are also crucial user-input parameters. UD represents the cost per unit of freight that surpasses capacity, acting as a penalty for lateness, while E_{cong} reflects the planned and expected congestion beyond the acceptable level CX. The population cluster is treated as another industry cluster for the optimization model.

To calculate transportation costs and identify ramps near industry and population clusters, a distance matrix has been established where the distance between the industry cluster and the ramp is identified along the highway network. Two critical variables, EX and AXE, are computed for each ramp and across all three periods. EX represents the flow in excess capacity on a ramp, whereas AXE measures congestion beyond the specified level C, target congestion. Subsequently, the binary variable Y is used to determine which ramps experience congestion and require capacity expansion. The capacity that can be added to congested ramps is represented by CN. The total construction cost is calculated by multiplying the binary variable with the corresponding cost of expansion or modification at the specific ramp.

Finally, the delay cost is computed by summing up the EX -variables, weighted by the probabilities of each scenario across the periods and multiplied by UD. By adding the transportation (travel) and construction costs to the delay cost, the overall cost is calculated. As stated, the goal is to minimize this overall cost (aggregation of transportation cost, construction cost, delay cost).

4.5.3 Retail Trade Industry Model

Based on the industry attractiveness, Hendricks County was selected for the retail trade industry to develop the optimization model. The industry attractiveness analysis yielded the top giants of retail trade in Hendricks County, which was used to determine the industry cluster in the county. For passenger traffic, a populous cluster was assumed at the center of the most populous city in the county—which is Plainfield City in Hendricks County. Once these clusters are identified, nearby exit ramps on interstates, along with the associated capacity are identified from the *Highway Capacity Manual* (TRB, 1997). For the Hendricks County clusters, the nearby interstate is I-70. Table 4.4 represents the exit ramps of interest and their capacity in terms of AADT. The geocoordinates of the industry and population clusters are identified and the distance from the ramp exit to the clusters is calculated as shown in Table 4.4.

Based on the forecast for freight and passenger AADT, Table 4.5 represents the data used for Hendricks County for the optimization model.

4.5.4 Results

The optimization model was run for multiple iterations, changing the user inputs available. This displayed how the results can vary when the policies and needs for INDOT change in the future.

For the report, certain assumptions were considered.

- Construction cost of each ramp lane is assumed to be $1.5 million.
- Additional capacity, in terms of lane, added to the ramp after construction is assumed to be 20 AADT.
- It is assumed a maximum of one lane can be added to each ramp.
- The target congestion is set at 10 AADT.
- Each scenario is assigned an equal probability of 0.1 of occurring in the future.

The optimization model is run for various values of UD and ECONG—the parameters that can be adjusted by INDOT to decide the level of congestion or level of service required for the industry cluster. For retail trade industry, a high level of service is industry standard, especially for retailers like Walmart, Amazon, Target, etc. To attract these retailers and catalyze the growth of the retail trade industry, it is important to allow as little congestion as possible and as high level of service as possible. Hendricks County shares borders with Marion County, the largest and busiest county in Indiana, experiences high traffic volume and congestion. Thus, a better level of service is required for passenger vehicles too.

As defined previously, *UD* is the cost per unit of freight congestion. Increasing *UD* increases the delay cost, forcing the model to find route without congestion. E_{cong} is the measure of flow over acceptable congestion. As E_{cong} increases, the flow allowed beyond acceptable congestion increases, thereby causing delays, and increasing the cost due to the delay. This, however, prevents the need for new construction. Increasing *UD* increases the penalty for flow over the capacity and thus after a certain *UD* value, the model recommends new construction. The optimization model for Hendricks County suggests that there is enough capacity, in terms of exit ramp AADT, around the industry and population cluster to support the growth in freight and passenger traffic until 2045 across various scenarios. The capacity becomes insufficient for an exceedingly high value of *UD* (10^{20}), the penalty associated with delay.

4.6 Chapter Summary

This chapter presented a detailed analysis of the current and future growth in freight and passenger traffic and inform INDOT about road infrastructure needs that it can assist in ensuring efficiency and economic competitiveness in Indiana.

First, the relationship of freight traffic with industry-level county GDP was shown using linear regression analysis. It was established that five industries—construction, manufacturing, retail trade, wholesale trade,

TABLE 4.4
Selected exit ramps in Hendricks County on I-70, along with their AADT capacity and distance to the population and retail industry cluster of the county

				Distance of Ramp to Cluster (mi)	
Ramp ID	**Latitude**	**Longitude**	**AADT Capacity**	**Retail**	**Population**
70-066A	39.66688735	-86.37135791	1,729.6	5.2	4.1
70-066B	39.66693538	-86.37013554	1,767.9	5.1	8.2
70-066B	39.66942016	-86.36611385	1,793.1	4.8	7.9
70-066C	39.66944541	-86.36983037	1,792.6	3.9	3.9
70-066D	39.66687179	-86.37444005	1,757.1	18.9	16.2
70-066K	39.66605012	-86.36929798	1,880.7	5.1	8.2
70-068A	39.68235673	-86.32936345	1,682.9	5	8.1
70-068B	39.68420671	-86.32472766	1,690.1	5.7	9.1
70-068C	39.68779849	-86.32528634	1,719.0	2.3	5.4
70-068D	39.68565648	-86.33282011	1,638.2	6.1	6.2
70-068F	39.68485977	-86.32670251	1,642.6	2.4	5.5
70-068H	39.68615533	-86.33197589	1,719.1	5.3	8.4
70-069A	39.68691811	-86.32366947	1,858.7	5.5	8.9

TABLE 4.5
Forecasted freight and passenger AADT (vehicles per day) for Hendricks County

Scenario	Freight 2025	Freight 2035	Freight 2045	Passenger 2025	Passenger 2035	Passenger 2045
SSP1	2,448.8	2,603.4	2,747.5	7,204.5	7,552.7	7,869.3
SSP2	2,434.6	2,548.4	2,648.7	7,181.4	7,499.8	7,773.0
SSP3	2,407.4	2,478.8	2,526.9	6,988.7	7,050.8	7,029.2
SSP4	2,439.8	2,584.2	2,709.3	7,125.0	7,360.3	7,530.1
SSP5	2,480.1	2,709.7	2,966.1	7,377.0	7,953.0	8,554.3

and transportation and warehousing, contributed to the maximum freight traffic on road. Based on data from the FAF, economic zones were identified for each county and dummy variable was added to the regression analysis for a better fit. Using linear regression analysis, a relationship between passenger traffic and county-level population was also determined.

For future forecasts, multiple SSP scenarios, prepared by the International Institute for Applied Systems Analysis (IIASA) were considered. These scenarios include future predictions across scenarios for various characteristics like GDP, population, energy, emissions, etc. The forecasted world- and US-level GDP for various SSP scenarios was extracted and used with the historical Indiana GDP to determine the scenario-based future GDP forecast using linear regression. GDP contribution ratio of every county and industries in each county was used to forecast the industry-level county GDP through 2050. Similarly, world and US population forecasts from SSP database were used along with Indiana and county population from BEA to forecast the future county-level population growth. The forecasted GDP and population were used to predict scenario-based future freight and passenger vehicle traffic based on the previously established relationship using linear regression analysis.

It was important to identify what parameters are considered by different industries when planning for relocation or expansion that would make a region/county attractive for business for the industry. Four parameters—growth in past 20 years, top 20 industry giants, weekly wages and employment, and proximity to interstates and highways were identified. Using these four parameters, attractive, potential counties were identified for the five industries that are established as the highest freight traffic contributing industries. For instance, Marion County, Dearborn County, Washington County, and Hendricks County are the counties identified for the retail trade industry.

After identifying the attractive counties, an optimization model was developed to identify the ramps near industry cluster that would experience the maximum freight traffic and evaluate whether the ramps provide enough capacity to support the growth of industries in the county. The objective function of the optimization is to minimize the aggregate of construction cost, delay cost, transportation cost, and downside due to congestion. The optimization allows for user input where INDOT can change certain variables according to their need and this makes the model dynamic. From the Hendricks County model for retail industry, it can be concluded that the ramps near the industry and population cluster have enough capacity to support the growth in traffic. The need for construction of new ramps and addition of capacity arises when the penalty for congestion is very high.

These analyses and results can be used for informed development and improvement of road infrastructure. INDOT can target counties for industries of interest and ensure there is enough ramp capacity to support the predicted growth. In addition, INDOT can align their infrastructure projects with the growth models for efficiency and improved service. INDOT can leverage these analyses to plan proactively and stay competitive for business, leading to economic growth.

4.6.1 Limitations and Future Work

This analysis is not without limitations, which are subsequently discussed to set the stage for areas of future work. First, data such as the GDP and population data obtained for the forecast calculation are current only until 2020 and 2021 since at the time of this analysis data was only available for that time frame. Due to that, the analysis might not reflect any current or latest developments. The analysis can be replicated with current data to obtain results reflecting the latest data. Moreover, the model developed for the study only considers single industry for traffic data. The model can further be expanded to consider multiple industries which have potential for growth in the counties. Furthermore, the optimization model is developed for a single county, single industry scenario. This can be further expanded, as part of the future scope, to include multiple industries and the analysis can be performed for multiple counties of interest. The optimization model in this chapter of the study only discusses the analysis and results for a single county as a sample. INDOT can use this optimization model and the approach to develop the analysis for other counties in the state as part of future scope based on counties of interest or priority. Lastly, this chapter only focuses on county-level analysis. Future research can replicate the macro-level analysis for micro-levels by further disaggregating the data to the micro-level (ZIP code- or ISTDM-level).

5. CONCLUSION

This study explored the change in freight and passenger travel demand and travel behavior of Hoosiers while considering different levels of analysis (macro- and micro-level) and drawing on diverse relevant data sources. Even though some of the data considered for each level are obtained from different sources, the study aims to provide INDOT the ability to understand the travel behavior at different granularities. The results across these multiple scales aim to complement each other.

5.1 Forecasting Travel Demand Using Geolocation Data

One of the project objectives was to predict interzonal trips in Indiana up to 2050 by leveraging traditional transportation planning models and incorporating large-scale mobile phone geolocation data. The four-stage planning scheme, which includes trip generation and trip distribution, is enhanced by using GPS data to estimate current travel patterns and relevant socioeconomic attributes. Data collection and processing involve filtering for high accuracy and identifying high-quality device users. Home detection is used to assess data representativeness, while trip segmentation and OD matrix preparation identify and aggregate trips across different TAZs. Trip generation and distribution models are developed using linear regression and gravity models, respectively. Socioeconomic and land use indicators are considered to predict the production and attraction of home-based and work-based trips. The gravity exponent is calibrated using line search optimization. By combining traditional transportation planning methods with large-scale mobile phone geolocation data, this task provides a more comprehensive and accurate estimation of future interzonal trip interchanges in Indiana.

5.2 Estimating the Impact of E-Commerce on Travel Demand

The analyses turn into multiple key findings. First, there exist multiple types of shoppers in Indianapolis, varying by frequency and mode preference. The diversified shopper classes are correlated with their household socio-demographic characteristics, especially household size, internet access, and education level. Second, the urban transportation system could reap benefits from increasing e-commerce adoption (i.e., a higher market penetration rate). Substituted in-person shopping trips can not only decrease personal travel demand but also encourage travelers to choose more sustainable travel modes. We found that the e-commerce market penetration could reduce personal travel demand by 14% (about 600,000 miles/day) at most due to the transition from in-person to online shopping. The changed travel demand can also encourage a shift in travel mode choice by decreasing the mode share of private cars by 4%, shifting to green transportation modes like public transit and multimodal systems. Third, since goods delivery service may have a negative impact on the transportation system, we find that the net VMT change depends on delivery type. In the 100% e-commerce adoption scenario, truck delivery can increase VMT by a maximum of 300,000 miles/day from parcel delivery and 180,000 miles/day from food delivery. Centralized delivery service allows delayed delivery and can carry multiple orders at the same time, enabling delivery route optimization and improving efficiency. Due to data limitations, this study did not include many factors (e.g., crime and economic conditions) that may change shopping behavior and our analysis did not differentiate the VMT impacts from different vehicle types (e.g., 4-tire commercial vehicles, non-freight trucks, etc.). The simulation model proposed in this study is adaptable and can be updated as more recent data becomes available.

5.3 Forecasting County-Level Industry Shifts Based on Scenarios

One of the objectives of the project was to understand the freight and passenger traffic at the county and industry level and provide recommendations for the road infrastructure based on growth in the traffic. Using linear regression models, a relationship for freight and passenger traffic was identified with industry GDP and population, respectively, among many other variables. Once the relationship was established, future growth in freight and passenger traffic was predicted across multiple scenarios and thereby future growth in traffic was predicted across scenarios. Next, industry attractiveness parameters were used to identify the counties that would attract the five maximum freight traffic contributing industries. To determine the need for improvements in road infrastructure, and help INDOT better plan and align their projects, an optimization model was developed that provided insight to whether the existing capacity in counties is enough to support the growing traffic in Indiana. This optimization model allowed users to provide input and change the levers according to their needs. The retail industry model for Hendricks County showed that ramps in Hendricks County near the industry and population cluster have enough capacity to support the future growth in freight and passenger traffic.

5.4 Synthesis of Key Findings

The three tasks complement each other by providing comprehensive insights into different aspects of travel demand forecasting (Figure 5.1). The first task focuses on utilizing geolocation data and short-term forecasts to finetune parameters in long-term models, enhancing the predictive power of travel demand estimates. The second task explores the impact of e-commerce on travel demand, offering recommendations to encourage sustainable urban transportation through digital infrastructure development and shared delivery services. Additionally, the task assesses how e-commerce adoption affects travel demand for freight trucks and passenger cars. The third task involves forecasting county-level shifts based on scenarios, incorporating economic and demographic parameters to understand freight and passenger traffic growth. It utilizes regression analysis and scenario-based forecasts to predict future traffic conditions, supporting informed road

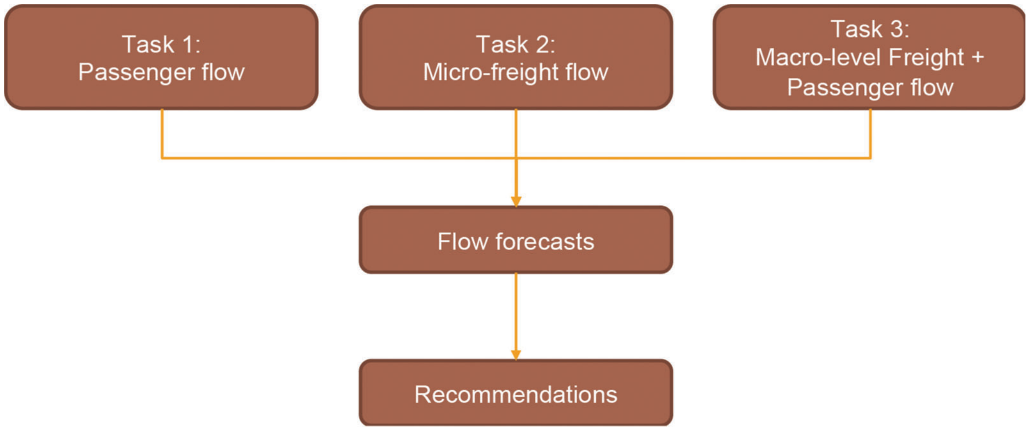


Figure 5.1 Relationships between the key project tasks.

infrastructure development and efficiency improvements for INDOT.

For example, the long-term predictions of total trips made in the first task may be used to compute the total predicted change in VMT over the decades. These predictions may then be combined with the VMT change predictions with the change in market penetration of e-commerce (shown in Figure 3.13 of Task 2). Such analysis will help analysts identify the trends of e-commerce activity in the future and help assess the extent of e-commerce market penetration required for targeting a specific level of intra- and inter-county travel. For example, INDOT may want to target an increase in the number of freight movement trips from Marion County to Lake County to achieve a specific increase in freight revenue. To achieve this, the level of market penetration of e-commerce activity will provide information about that increase in trip count.

Similarly, the total trips predicted in Task 1 may be evaluated with respect to the total freight movement predicted in Task 3 across the different SSP scenarios. Notably, the overall county freight and passenger trips predicted across the five SSP scenarios may be used as inputs to the trip distribution model of Task 1 to identify the change in passenger and freight movement across the major corridors of Indiana.

5.5 Implementation

5.5.1 Forecasting Travel Demand Using Geolocation Data

GPS data provide an excellent opportunity for INDOT to consider developing short-term forecasts to finetune the parameters of the long-term models used in Chapter 2. This could be made possible by repeating the modeling exercise with one to 2-year forecasts and testing which combination of parameters of data processing and the models fits best with the inferred trip count of the subsequent years. This is feasible since the socioeconomic attributes are available every year in the form of ACS 1-year estimates. For example, with the 2022 ACS data and GPS data, forecasts may be made for 2023. Then, the predictions may be tested against the trip counts inferred from the 2023 GPS data. This process may sufficiently narrow the cone of uncertainty of the predictive power of the long-term forecasts.

There are some technical requirements for INDOT to adequately implement the framework proposed in Chapter 2. First, it is important to establish business contracts with data vendors to obtain the mobile phone geolocation data needed for this task. Second, geolocation data processing requires sufficient computing power and storage. For reference, the original files of the 3-month data used in this study (Indiana, March 2019–2021) occupy ~246 GB storage, which reduces to ~43.5 GB with appropriate compression. Third, data processing is best achieved using in-memory, parallel processing frameworks like Hadoop or Apache Spark. To aid users with the heavy processing of data, our team has developed an open-source toolkit called mobilkit+ (Verma & Ukkusuri, 2023). It is a Python and Spark-based package that contains commonly used routines to process the geolocation data.

5.5.2 Estimating the Impact of E-Commerce on Travel Demand

In developing the e-commerce platform for improved urban transportation sustainability, it is essential to encourage the transition from in-store shopping to online platforms with decreased private car use and an optimized delivery system. Results from Chapter 3 provide some insights for planners, business sectors, and governments. First, investing more resources to build an accessible and affordable digital infrastructure by the government can promote e-commerce adoption. Market segmentation result shows that there still exist many households that do not have access to the internet. Households with lower education levels and lower income levels are mostly “traditional shoppers.” Providing affordable and easy-to-use services to residents who currently do not have internet access, either because they cannot afford to or do not have enough knowledge to use it. Prioritizing the security of online payment systems and personal privacy can also attract traditional shoppers to use e-commerce. In addition, providing low-income households with credit cards to other payment systems to facilitate their use of online shopping can benefit social equity by increasing their accessibility to food, pharmacy, and other essential services. By establishing a robust and accessible digital platform, more residents in Indianapolis could be encouraged to adopt e-commerce as a convenient alternative to replace traditional in-store shopping. The avoided shopping car trips can further reduce urban traffic burdens and benefit environmental sustainability. Encouraging the shift from in-store shopping to e-commerce may reduce travel demand under the condition that most users choose a centralized delivery system. Second, developing green shared delivery services, where multiple retailers can combine their deliveries, allows for optimized delivery routes, and reduced overall transportation costs. This approach minimizes the number of individual delivery trips and contributes to reducing traffic congestion and emissions. In addition, replacing current fossil fuel power delivery vehicles with other low-emission vehicles (e.g., electric vehicles) can further reduce the GHG emissions from delivery services. The results from the agent-based travel demand simulation model from Chapter 3 can help Indiana MPOs adjust the travel demand model and account for the impact of future e-commerce adoption. Insights from the model can also provide guidance for the infrastructure investment plan and land use changes. The developed modeling framework can also be applied in other cities to generate city-specific results as more data becomes available. Results from this project can also inform future long-range transportation plan updates.

5.5.3 Forecasting County-Level Shifts Based on Scenarios

Chapter 4 presented Indiana demographic data, industry-level county GDP and population, to forecast freight and passenger traffic for up to 2050. These forecasts, which encompass multiple SSP socioeconomic impact scenarios, provide excellent flexibility for INDOT to plan for the future traffic trends. These forecasts can be incorporated to enhance INDOT's long range 2045 plan. The industry attractiveness analysis allows INDOT to identify the regions which are more attractive and suitable for specific industries than others, based on major parameters considered across industries. Determination of these regions can help INDOT plan for expansion or development of infrastructure according to needs. Since INDOT road infrastructure projects are long-term projects and need to be determined well in advance, this analysis helps INDOT plan proactively and align the projects with prospective industry growth for efficiency and better outcome. The results of the optimization model allow INDOT to ascertain whether the ramp traffic capacity in the attractive counties is enough to support the increase in traffic due to GDP-based industry and population growth. Results from this optimization model would help INDOT identify the ramps in the counties that would experience congestion and help INDOT prioritize their plan for expansion where required. This would allow INDOT to help Indiana be competitive and prevent ramps from being bottlenecks and hampering the growth in the state.

REFERENCES

ADOT. (2016, May 12). *Arizona state freight plan: Future scenarios and implications for freight transportation.* Arizona Department of Transportation. https://azdot.gov/sites/default/files/2019/08/ph-6-arizona-freight-modeling-and-forecasting-reports.pdf

ArcGIS. (2018). *Indiana 2018 PR* [Web server]. https://geo.dot.gov/server/rest/services/Hosted/Indiana_2018_PR/FeatureServer

BEA. (2023a, December 7). *CAGDP2 gross domestic product (GDP) by county and metropolitan area 1.* US Bureau of Economic Analysis. https://apps.bea.gov/iTable/?reqid=70&step=30&isuri=1&major_area=5&area=38340&year=-1&tableid=501&category=2501&area_type=5&year_end=-1&state=5&classification=naics&statistic=-1&yearbegin=-1&unit_of_measure=levels

BEA. (2023b, December 14). *SASUMMARY state annual summary statistics: Personal income, GDP, consumer spending, price indexes, and employment.* Bureau of Economic Analysis. https://apps.bea.gov/itable/index.html?appid=70&stepnum=40&Major_Area=3&State=00000&Area=XX&TableId=600&Statistic=15&Year=-1&YearBegin=-1&Year_End=-1&Unit_Of_Measure=Levels&Rank=0&Drill=1&nRange=5&AppId=70

BIC. (2023). *INDOT construction investment: INDOT construction, preliminary engineering and development opportunities data*. Build Indiana Council. https://www.buildindianacouncil.org/construction-investment/

Bivens, J. (2019, January 23). *Updated employment multipliers for the U.S. economy*. Economic Policy Institute. https://epi.org/160282

Black, W. R. (1973). An analysis of gravity model distance exponents. *Transportation*, *2*(3), 299–312. https://doi.org/10.1007/BF00243358

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, *350*(6264), 1073–1076. https://www.sciencemag.org/lookup/doi/10.1126/science.aac4420

Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*, *17*(3), 285–297.

Cao, X. J., Xu, Z., & Douma, F. (2012). The interactions between e-shopping and traditional in-store shopping: An application of structural equations model. *Transportation*, *39*(5), 957–974. https://doi.org/10.1007/s11116-011-9376-3

Cheng, Z., Jian, S., Rashidi, T. H., Maghrebi, M., & Waller, S. T. (2020, June). Integrating household travel survey and social media data to improve the quality of OD matrix: A comparative case study. *IEEE Transactions on Intelligent Transportation Systems*, *21*(6), 2628–2636.

Chung, E.-H., & Shalaby, A. (2005). A trip reconstruction tool for GPS-based personal travel surveys. *Transportation Planning and Technology*, *28*(5), 381–401.

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(5), 603–619. https://doi.org/10.1109/34.1000236

Conexus Indiana. (2022). *A glimpse into Indiana's factory of the future: Companies scale-up industry 4.0 technology investments to enhance efficiency and optimize productivity*. https://www.conexusindiana.com/wp-content/uploads/2022/12/2022-Conexus-Tech-Adoption-Report.pdf

Cools, M., Moons, E., & Wets, G. (2010). Assessing the quality of origin–destination matrices derived from activity travel surveys: Results from a Monte Carlo experiment. *Transportation Research Record*, *2183*(1), 49–59. https://doi.org/10.3141/2183-06

Cox, W., & Love, J. (1996). *The best investment a nation ever made*: A tribute to the Dwight D. Eisenhower system if interstate & defense highways. http://www.publicpurpose.com/freeway1.htm

Dias, F. F., Lavieri, P. S., Sharda, S., Khoeini, S., Bhat, C. R., Pendyala, R. M., Pinjari, A. R., Ramadurai, G., & Srinivasan, K. K. (2020). A comparison of online and in-person activity engagement: The case of shopping and eating meals. *Transportation Research Part C: Emerging Technologies*, *114*, 643–656. https://doi.org/10.1016/j.trc.2020.02.023

Duffus, L. N., Alfa, A. S., & Soliman, A. H. (1987). The reliability of using the gravity model for forecasting trip distribution. *Transportation*, *14*, 175–192.

East Central Indiana Regional Partnership. (n.d.). *Cost of doing business* [Webpage]. Retrieved August 20, 2023, from https://www.ecirp.org/cost-of-doing-business

Erlander, S., & Stewart, N. F. (1990). The gravity model in transportation analysis: Theory and extensions. *Environment and Planning B: Planning and Design*, *3*(2), 232–233.

Erlich, A., Jung, D. F., Long, J. D., & McIntosh, C. (2018). The double-edged sword of mobilizing citizens via mobile phone in developing countries. *Development Engineering*, *3*, 34–46. https://doi.org/10.1016/J.DEVENG.2017.11.001

FHWA. (2020, February). *Average fuel economy by major vehicle category*. Federal Highway Administration. https://afdc.energy.gov/data/10310

Fricker, J. D., & Jin, L. (2008, October). *Development of an integrated land-use transportation model for Indiana* (Report No. FHWA/IN/JTRP-2008/15). https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2669&context=jtrp

Goldberg, J. (2022, February 18). *E-commerce sales grew 50% to $870 billion during the pandemic*. Forbes. https://www.forbes.com/sites/jasongoldberg/2022/02/18/e-commerce-sales-grew-50-to-870-billion-during-the-pandemic/?sh=6d3650fb4e83

Grantz, K. H., Meredith, H. R., Cummings, D. A. T. T., Metcalf, C. J. E., Grenfell, B. T., Giles, J. R., Mehta, S., Solomon, S., Labrique, A., Kishore, N., Buckee, C. O., & Wesolowski, A. (2020). The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature Communications*, *11*(4961). https://doi.org/10.1038/s41467-020-18190-5

Handy, S., Shafizadeh, K. R., & Schneider, R. (2013, March). *California smart-growth trip generation rates study*. https://nacto.org/docs/usdg/smart_growth_trip_generation_rates_handy.pdf

Hasan, S., & Ukkusuri, S. V. (2014). Urban activity pattern classification using topic models from online geo-location data. *Transportation Research Part C: Emerging Technologies*, *44*, 363–381.

He, Q.-C., & Hong, T. (2018). Integrated facility location and production scheduling in multi-generation energy systems. *Operations Research Letters*, *46*(1), 153–157.

HNTB Corporation. (2019, September). *Georgia statewide travel demand model report*. Georgia Department of Transportation. https://www.dot.ga.gov/InvestSmart/TravelDemandModels/Development of Statewide Model.pdf

Hong, S., Zhao, F., Livshits, V., Gershenfeld, S., Santos, J., & Ben-Akiva, M. (2021). Insights on data quality from a large-scale application of smartphone-based travel survey technology in the Phoenix metropolitan area, Arizona, USA. *Transportation Research Part A: Policy and Practice*, *154*(Issue C), 413–429.

Hooper, K. G. (2017). *Trip generation handbook* (3rd ed.). Institute of Transportation Engineers.

Hu, P., Schmitt, R., Fan, C.-C., Parker, J., Wingfield, A., & Kweon, Y.-J. (2022). *Comparison of commodity flow survey (CFS) and freight analysis framework (FAF) for 2017 domestic freight flows*. https://doi.org/10.21949/1526410

Indiana map.org. (n.d). *Indiana road map* [Webpage]. https://www.indiana-map.org/road-map.htm

Indy GIS. (2022, July 25). *Indianapolis and Marion County current land use*. City of Indianapolis and Marion County. https://data.indy.gov/datasets/current-land-use/explore

Kim, W., & Wang, X. (2021). To be online or in-store: Analysis of retail, grocery, and food shopping in New York city. *Transportation Research Part C: Emerging Technologies*, *126*, 103052. https://doi.org/10.1016/j.trc.2021.103052

Langham Logistics. (2019, July 15). *Shipping: Why Indiana needs to be your logistics hub*. https://www.elangham.com/2019/07/15/why-indiana-needs-to-be-your-logistics-hub/

Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W., & Ma, W.-Y. (2008). Mining user similarity based on location history. *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, *183*, 1–10. https://doi.org/10.1145/1463434.1463477

Luo, H., Chahine, R., Gkritza, K., & Cai, H. (2023). *What motivates the use of shared mobility systems and their integration with public transit: Evidence from a choice experiment study*. http://dx.doi.org/10.2139/ssrn.4330300

Luo, H., Chahine, R., Rambaram, A., Rosenzweig, E. T., Gkritza, K., & Cai, H. (2022). *Assessing the travel demand and mobility impacts of transformative transportation technologies in Indiana* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2022/11). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317374

Luo, H., Zhao, F., Chen, W.-Q., & Cai, H. (2020). Optimizing bike sharing systems from the life cycle greenhouse gas emissions perspective. *Transportation Research Part C: Emerging Technologies*, *117*, 102705. https://doi.org/10.1016/j.trc.2020.102705

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics* (pp. 105–142). Academic Press. https://www.econbiz.de/Record/conditional-logit-analysis-of-qualitative-choice-behavior-mcfadden-daniel/10002395479

McGuckin, N., & Fucci, A. (2018, July). *Summary of travel trends: 2017 National household travel survey* (Report No. FHWA-PL-18-019). Federal Highway Administration Department of Transportation. https://nhts.ornl.gov/assets/2017_nhts_summary_travel_trends.pdf

Metropolitan Washington Council of Governments. (n.d.a). *Community cooperative forecast* [Webpage]. https://www.mwcog.org/community/planning-areas/cooperative-forecast/

Metropolitan Washington Council of Governments. (n.d.b). *Transportation modeling* [Webpage]. https://www.mwcog.org/transportation/data-and-tools/modeling/

Munnich, L. W., Jr., Iacono, M., Dworin, J., & Brandt-Sargent, B. (2015). *Transportation planning to support economic development: An exploratory study of competitive industry clusters and transportation in Minnesota* (Report No. MN/RC 2015-02). Humphrey School of Public Affairs. https://cts-d8resmod-prd.oit.umn.edu/pdf/mndot-2015-02.pdf

NIRCC. (2019). *Travel forecast: 2040 travel demands* Northeastern Indiana Regional Coordinating Council. https://www.nircc.com/uploads/1/2/9/8/129837621/2040planamended9_19.pdf

NRF. (2014). *Trade matters for Indiana*. National Retail Federation. https://tradepartnership.com/wp-content/uploads/2015/04/Indiana_NRF_Report.pdf

Nielsen. (2022). *Nielsen IQ* [Webpage]. https://www.chicagobooth.edu/research/kilts/datasets/nielsen

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., & Solecki, W. (2017). The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, *42*, 169–180.

Reia, S. M., Rao, P. S. C., Barthelemy, M., & Ukkusuri, S. V. (2022). Spatial structure of city population growth. *Nature Communications*, *13*(1), 5931. https://doi.org/10.1038/s41467-022-33527-y

Sadeghinasr, B., Akhavan, A., & Wang, Q. (2019). Estimating commuting patterns from high resolution phone GPS data. *Computing in Civil Engineering 2019: Data, Sensing, and Analytics* (pp. 9–16). https://doi.org/10.1061/9780784482438.002

Sapling, C. (2021, January 19). *What 6 charts say about the pandemic's impact on retail*. https://Www.Retaildive.Com/News/What-6-Charts-Say-about-the-Pandemics-Impact-on-Retail/593102/

Sarmiento, I., González-Calderón, C., Córdoba, J., & Díaz, C. (2013). Important aspects to consider for household travel surveys in developing countries. *Transportation Research Record*, *2394*(1), 128–136. https://doi.org/10.3141/2394-16

Shah, H., Carrel, A. L., & Le, H. T. K. (2021). What is your shopping travel style? Heterogeneity in US households' online shopping and travel. *Transportation Research Part A: Policy and Practice*, *153*(July), 83–98. https://doi.org/10.1016/j.tra.2021.08.013

Shen, L., & Stopher, P. R. (2014). Review of GPS travel survey and GPS data-processing methods. *Transport Reviews*, *34*(3), 316–334. https://doi.org/10.1080/01441647.2014.903530

Statista Research Department. (2023, December 18). *Construction industry in the U.S.–statistics & facts* [Webpage]. https://www.statista.com/topics/974/construction/#topicOverview

Stinson, M., Enam, A., Moore, A., & Auld, J. (2019, September). Citywide impacts of e-commerce: Does parcel delivery travel outweigh household shopping travel reductions? *Proceedings of the 2nd ACM/EIGSCC Symposium on Smart Cities and Communities*. https://doi.org/10.1145/3357492.3358633

Stopher, P., FitzGerald, C., & Zhang, J. (2008). Search for a global positioning system device to measure person travel. *Transportation Research Part C: Emerging Technologies*, *16*(3), 350–369.

Strange, R. (2023). *Employment growth within Indiana's industry sectors*. INcontext. https://www.incontext.indiana.edu/2023/jan-feb/article1.asp

Suel, E., & Polak, J. W. (2017). Development of joint models for channel, store, and travel mode choice: Grocery shopping in London. *Transportation Research Part A: Policy and Practice*, *99*, 147–162. https://doi.org/10.1016/j.tra.2017.03.009

Tatham, M. (2018, February 21). *Which states have the most credit cards?* Experian Research. Retrieved January 28, 2023, from https://www.experian.com/blogs/ask-experian/which-states-have-the-most-credit-cards/

Toth, P., & Vigo, D. (2002). *The vehicle routing problem*. SIAM.

TRB. (1997). *Highway capacity manual 2000*. Transportation Research Board. https://sjnavarro.files.wordpress.com/2008/08/highway_capacital_manual.pdf

U.S. EPA. (2023). *GHG emission factors hub* [Webpage]. https://www.epa.gov/climateleadership/ghg-emission-factors-hub

Verma, R., & Ukkusuri, S. V. (2023, April 3). *Mobilkit Plus: A toolkit for processing large-scale mobile phone GPS data*. https://github.com/rvanxer/mobilkit

Verma, R., Yabe, T., & Ukkusuri, S. V. (2021). Spatiotemporal contact density explains the disparity of COVID-19 spread in urban neighborhoods. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-90483-1

Wang, Z., He, S. Y., & Leung, Y. (2018). Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, *11*, 141–155. https://doi.org/10.1016/j.tbs.2017.02.005

Wang, Z., Wang, S., & Lian, H. (2021). A route-planning method for long-distance commuter express bus service based on OD estimation from mobile phone location data: The case of the Changping Corridor in Beijing. *Public Transport*, *13*(6), 101–125. https://doi.org/10.1007/s12469-020-00254-w

Wardrop, J. G. (1952, May). Road paper. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, *1*(3), 325–362.

Wilson, A. G. (1967). A statistical theory of spatial distribution models. *Transportation Research*, *1*(3), 253–269. https://doi.org/10.1016/0041-1647(67)90035-4

Yabe, T., Jones, N. K. W., Rao, P. S. C., Gonzalez, M. C., & Ukkusuri, S. V. (2022). Mobile phone location data for disasters: A review from natural hazards and epidemics. *Computers, Environment and Urban Systems*, *94*, 101777. https://doi.org/10.1016/J.COMPENVURBSYS.2022.101777

Yabe, T., Ukkusuri, S. V., & Rao, P. S. C. (2019). Mobile phone data reveals the importance of pre-disaster inter-city social ties for recovery after Hurricane Maria. *Applied Network Science*, *4*(98), 1–18. https://doi.org/10.1007/s41109-019-0221-5

Zhao, P., Hu, H., Zeng, L., Chen, J., & Ye, X. (2023). Revisiting the gravity laws of inter-city mobility in megacity regions. *Science China Earth Sciences*, *66*(2), 271–281.

Zou, X., Zhang, S., Zhang, C., Yu, J. J. Q., & Chung, E. (2022). Long-term origin-destination demand prediction with graph deep learning. *IEEE Transactions on Big Data*, *8*(6), 1481–1495. https://doi.org/10.1109/TBDATA.2021.3063553

APPENDICES

Appendix A. Supplementary Material

APPENDIX A. SUPPLEMENTARY MATERIAL

A.1 Forecasting Travel Demand Using Geolocation Data

A.1.1 Spatial Scales of Analysis in Indiana

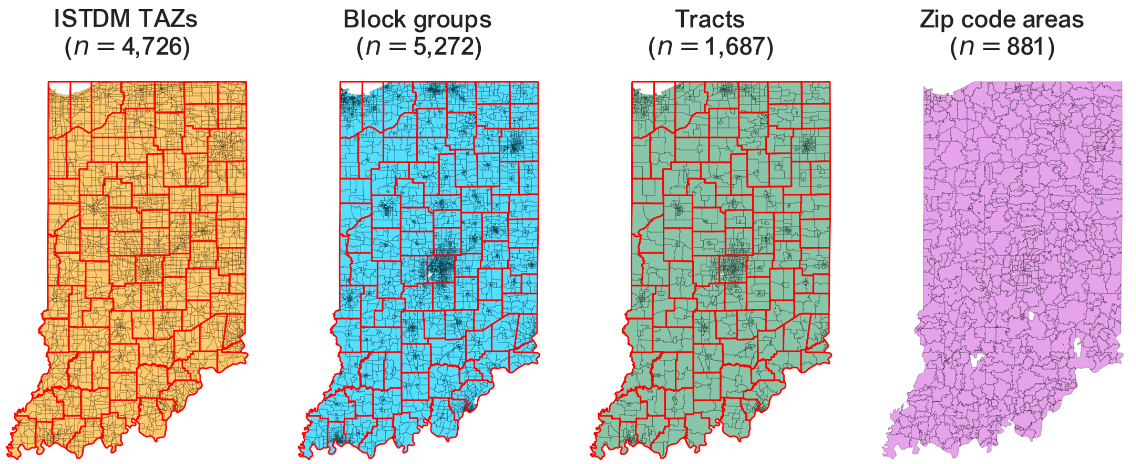


Figure A.1 ISTDM TAZs in Indiana compared with the US Census designation areas. The outlines in red delineate the county boundaries. The number of zones at each scale is given in parentheses.

A.1.2 GPS Data Distribution

The geolocation data are generally distributed relatively evenly across the 31 days of March in the three study years. This is shown in Figure A.2. It is also clear that the data density of 2021 is much higher than in 2020 which itself is much higher than 2019. This is primarily because the data provider increased its data collection and aggregation schemes to many more application programming interfaces in 2020 and 2021 compared to 2019. The consequence of this difference in the original ping density is the difference in the quality of the outcomes from the data, including the observation of poor trip and home detection quality in 2019 compared to 2020 and 2021 (both of which are comparable as explored outside this project).

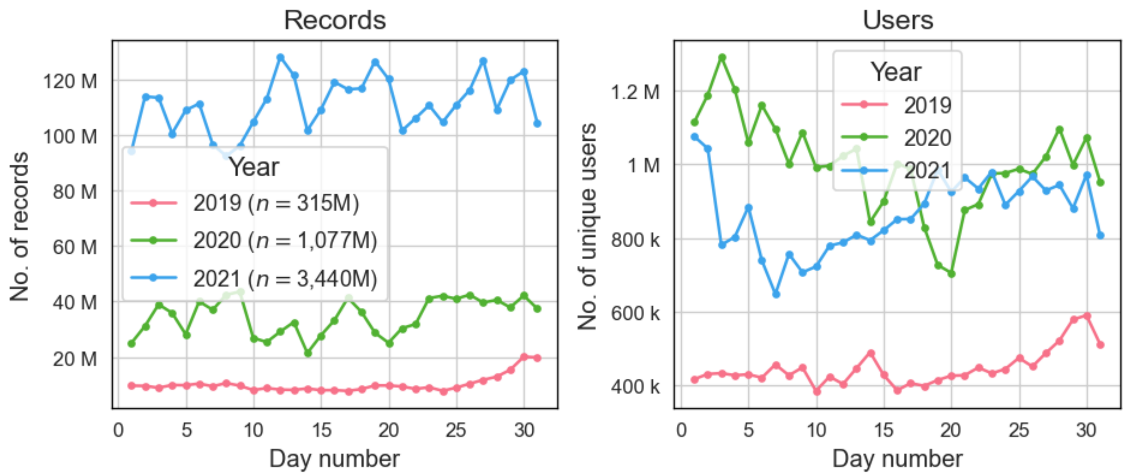


Figure A.2 Daily count of records ("pings") and unique mobile device users in the unfiltered geolocation dataset for the month of March in the study years. The total number of pings are given in the legend.

In addition to the temporal availability of the data over the years, another important factor to consider the quality of the data is ‘sampling interval’, which is the time difference between two consecutive pings. It is generally considered ideal to have consistently low sampling intervals because that instills confidence about the data continuity which is important for trip segmentation (Paipuri et al., 2020). We see in Figure A.3 that the sampling rate is low for most ping segments (i.e., pairs of consecutive pings), though for some segments, it can reach as high as 20 minutes. That is possible when a user’s device is either unavailable (switched off, for example) or the user explicitly disables location tracking, such as during nights at home.

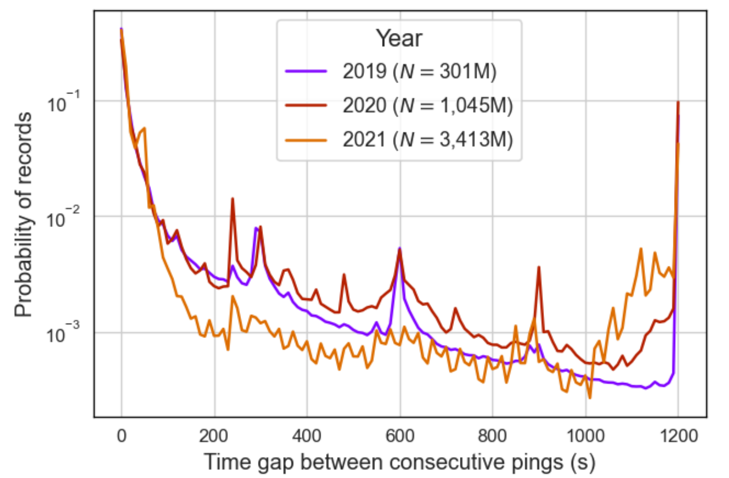


Figure A.3 Distribution of the sampling interval of the original GPS data across the study years, capped at 20 minutes. The total number of pings in each year’s dataset are shown in parentheses.

Overall, based on the given data quality in Figure A.2 and Figure A.3, we ascertain that the geolocation data are temporally reliable for further analysis, at least for the years 2020 and 2021.

A.1.3 Trip Generation Models for Indiana

The choice of spatial scale of modeling and the year of data to be used for modeling is considered important in the analysis, as illustrated in Section 2.5.1. For selecting the optimal trip generation model, we conducted a scenario analysis of the coefficients of the model with 48 scenarios, given by the Cartesian product of the following controls:

1. 2 demand outputs – trip production and attraction;
2. 2 day types – weekday and weekend;
3. 3 years – 2019, 2020, 2021;
4. 4 zone scales – TAZ, BG (block group), tract, and county.

Figure A.4 shows the signs and significance levels (in terms of p-value) of the covariates of the models of these scenarios. It can generally be seen that many descriptors in the models of 2019 are statistically insignificant at the significance level of 0.1. The models of 2020 have reasonable coefficients, but caution is to be exercised when using the data of 2020 for making long-term predictions. This is because the data are for the March of 2020 when substantial mobility pattern disruptions occurred around the time of a statewide lockdown due to COVID-19. It is for this reason that we use the models of 2021 at the ISTDM TAZ scale (highlighted in red in Figure A.4).

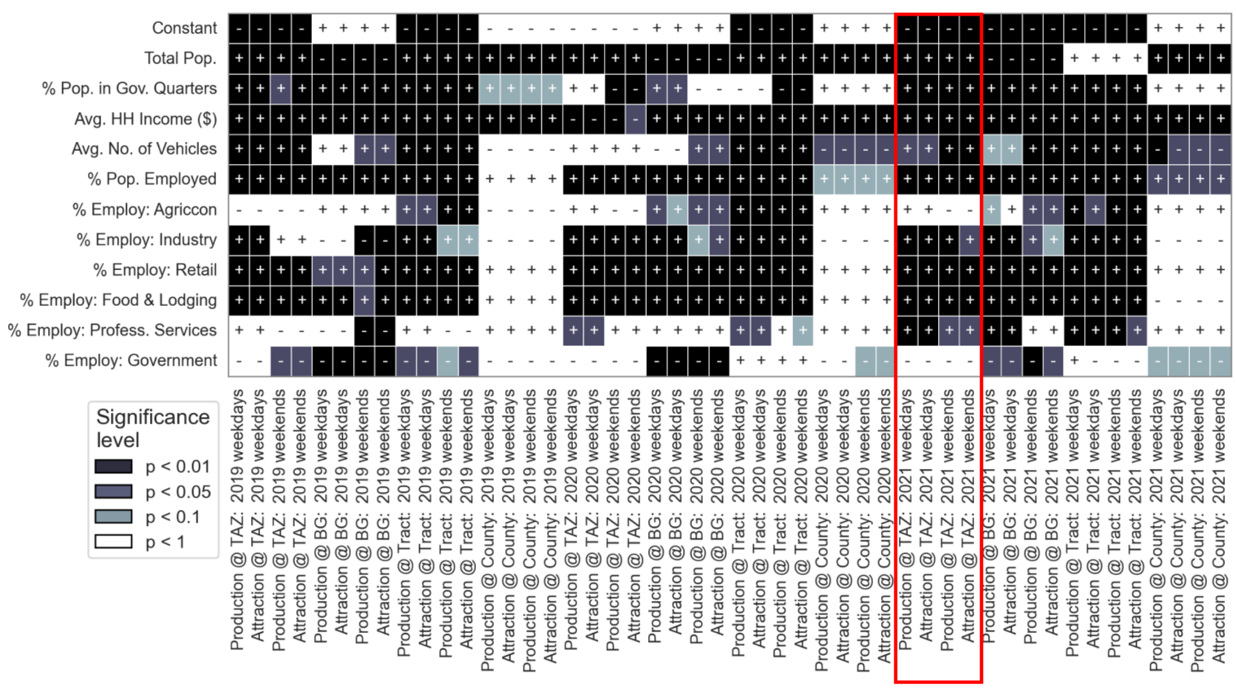


Figure A.4: Signs and significance levels of the covariates (y-axis) in the different trip generation model scenarios (x-axis). The finally selected set of models for trip production and attraction by day type are highlighted in red.

A.1.2 Gravity Model Calibration

The results of the line search optimization for finding the optimum gravity exponent (as explained in Section 2.4.6) are shown in Figure A.5. It is generally seen that the travel time MSE curves of travel time aggregations (median and mean) are continuously increasing and thus have the best fitted β values at the range start (i.e., $\beta = 0.1$), whereas the ones for distance (trip length)-based aggregations have distinct minima at around $\beta = 1.5$.

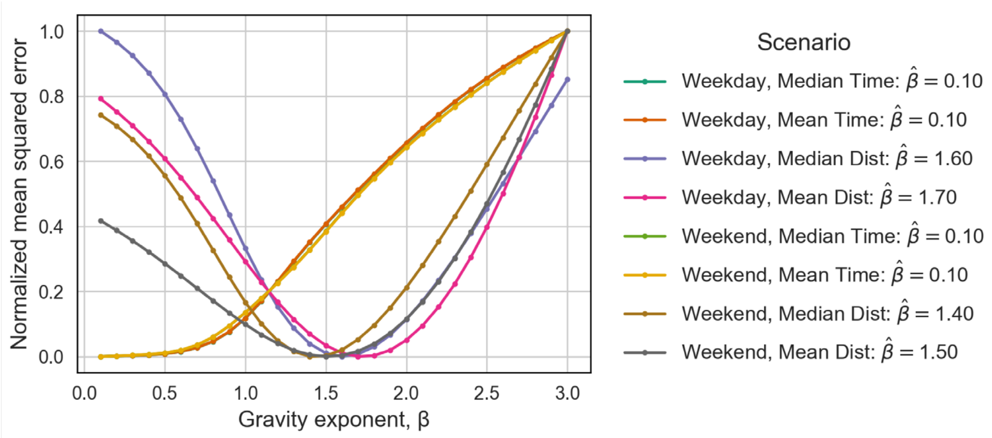


Figure A.5 Calibration results of the gravity models of the different scenarios. The fitted gravity exponent, $\hat{\beta}$, for each model is shown in the legend.

A.2 Estimating the future impact of e-commerce on travel demand

A.2.1 Market Segmentation of Hoosier's Shopping Behavior

Table A.1 Market segmentation of the latent class model

	Class 1: Traditional shopper	**Class 2: Frequent online shopper**	**Class 3: Dual-channel shopper**	**Class 4: Infrequent shopper**
No. of households for each class	951	102	186	257
Online shopping frequency				
0	93.4%	0.0%	0.0%	84.6%
1~12	6.2%	15.7%	42.1%	13.9%
13~26	0.3%	31.9%	35.7%	1.4%
26~52	0.1%	28.1%	14.0%	0.1%
>52	0.0%	24.3%	8.2%	0.0%
In-store shopping frequency				
0~80	3.5%	10.5%	8.1%	89.1%
81~160	9.6%	19.6%	24.2%	9.4%
161~240	48.4%	28.4%	38.5%	1.5%
241~320	29.6%	23.6%	18.7%	0.0%
>320	8.9%	17.9%	10.5%	0.0%
Average cost per online-shopping trip				
1~25	97.2%	15.0%	21.7%	11.0%
25~50	2.1%	36.6%	20.5%	7.6%
50~75	0.7%	31.9%	37.6%	42.5%
75~100	0.0%	10.8%	12.8%	24.4%
>100	0.0%	5.7%	7.4%	13.5%

Average cost per instore-shopping trip				
1~25	2.0%	4.3%	5.4%	5.1%
25~50	31.8%	32.1%	35.0%	14.0%
50~75	43.7%	40.4%	36.0%	30.1%
75~100	20.9%	18.7%	21.1%	37.4%
>100	1.6%	4.5%	2.5%	13.4%

A.3 Forecasting county-level shifts based on scenarios

A.3.1 Major Regional Economic Centers of Indiana

As per the Freight Analysis Framework described in Section 4.3.1, Indiana comprises four major regional economic centers containing the following counties:

- ***181: Chicago****-Naperville, IL-IN-WI*: Jasper, La Porte, Lake, Newton, Porter counties in Indiana.
- ***182: Indianapolis****-Carmel-Muncie, IN*: Bartholomew, Boone, Brown, Decatur, Delaware, Hamilton, Hancock, Hendricks, Henry, Jackson, Jennings, Johnson, Madison, Marion, Montgomery, Morgan, Putnam, Shelby counties.
- ***183: Fort Wayne****-Huntington-Auburn, IN*: Adams, Allen, DeKalb, Huntington, Noble, Steuben, Wells, Whitley counties.
- ***189****: Remainder of Indiana*

These are illustrated in Figure A.6.

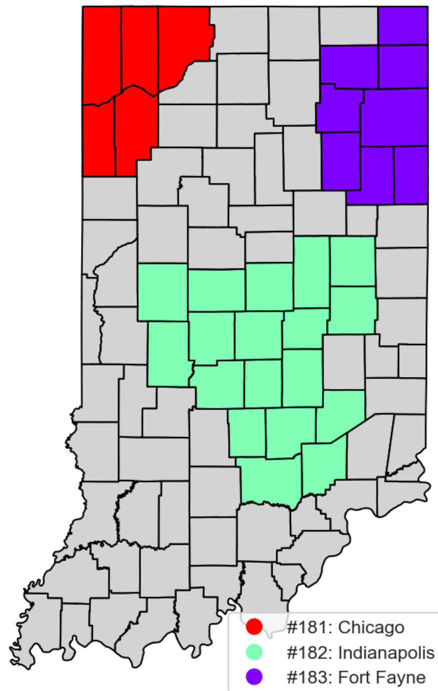


Figure A.6 The four FAF zones in Indiana, including three CSAs.

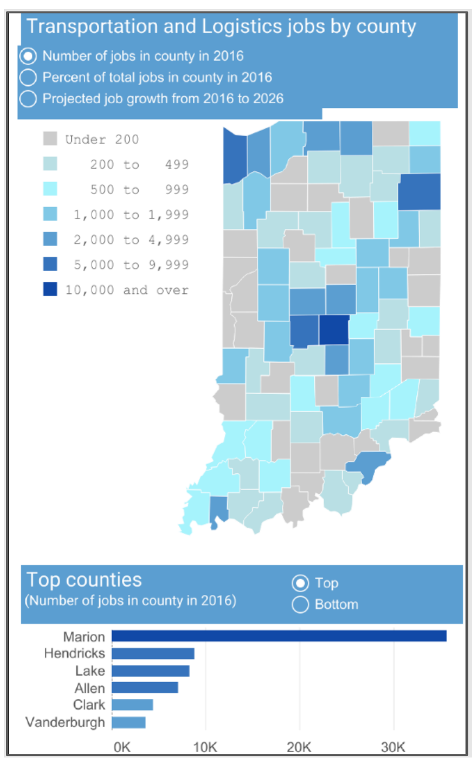


Figure A.7 Transportation and logistics jobs in Indiana. Source: (Hoosier Data, 2016).

About the Joint Transportation Research Program (JTRP)

On March 11, 1937, the Indiana Legislature passed an act which authorized the Indiana State Highway Commission to cooperate with and assist Purdue University in developing the best methods of improving and maintaining the highways of the state and the respective counties thereof. That collaborative effort was called the Joint Highway Research Project (JHRP). In 1997 the collaborative venture was renamed as the Joint Transportation Research Program (JTRP) to reflect the state and national efforts to integrate the management and operation of various transportation modes.

The first studies of JHRP were concerned with Test Road No. 1—evaluation of the weathering characteristics of stabilized materials. After World War II, the JHRP program grew substantially and was regularly producing technical reports. Over 1,600 technical reports are now available, published as part of the JHRP and subsequently JTRP collaborative venture between Purdue University and what is now the Indiana Department of Transportation.

Free online access to all reports is provided through a unique collaboration between JTRP and Purdue Libraries. These are available at http://docs.lib.purdue.edu/jtrp.

Further information about JTRP and its current research program is available at http://www.purdue.edu/jtrp.

About This Report

An open access version of this publication is available online. See the URL in the citation below.

Verma, R., Luo, H., Deodhar, S., Ka, E., Chahine, R., Natu, P., Malhotra, H., & Polisetty, V., Thakkar, D. J., Ukkusuri, S. V., Cai, H., Dunlop, S. R., Iyer, A. V., & Gkritza, K. (2023). *Forecasting shifts in Hoosiers' travel demand and behavior* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2023/28). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317685