NATIONAL CENTER FOR UNDERSTANDING FUTURE
TRAVEL BEHAVIOR AND DEMAND

Final Project Report

# Machine Learning Based Analysis of Activity Patterns to Assess Travel Behavior in Five Boroughs of New York City

***BY***

**Mahdieh Allahviranloo**
Email: mallahviranloo@ccny.cuny.edu

**Nikhita Kannam**
Email: nkannam000@citymail.cuny.edu

Grove School of Engineering, Department of Civil Engineering
The City College of New York
160 Convent Avenue, New York, NY, 10031

October 2025

## TECHNICAL REPORT DOCUMENTATION PAGE

| | | | |
|---|---|---|---|
| **1. Report No.**<br>N/A | **2. Government Accession No.**<br>N/A | **3. Recipient's Catalog No.**<br>N/A | |
| **4. Title and Subtitle**<br>Machine Learning Based Analysis of Activity Patterns to Assess Travel Behavior in Five Boroughs of New York City | | **5. Report Date**<br>October 3, 2025<br>**6. Performing Organization Code**<br>N/A | |
| **7. Author(s)**<br>Mahdieh Allahviranloo, https://orcid.org/0009-0001-8089-5267<br>Nikhita Kannam, https://orcid.org/0009-0000-1085-2483 | | **8. Performing Organization Report No.**<br>N/A | |
| **9. Performing Organization Name and Address**<br>The city college of New York<br>160 Convent Avenue, New York, NY, 10031 | | **10. Work Unit No. (TRAIS)**<br>N/A<br>**11. Contract or Grant No.**<br>69A3552344815 and 69A3552348320 | |
| **12. Sponsoring Agency Name and Address**<br>U.S. Department of Transportation,<br>University Transportation Centers Program,<br>1200 New Jersey Ave, SE, Washington, DC 20590 | | **13. Type of Report and Period Covered**<br>Final Report, 2024-2025<br>**14. Sponsoring Agency Code**<br>USDOT OST-R | |
| **15. Supplementary Notes**<br>N/A or Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration. | | | |
| **16. Abstract**<br>This study examines changes in urban mobility patterns in New York City between 2019 and 2022 using machine learning techniques. Analysis of the Citywide Mobility Survey data (n=85,000) reveals a reduction from seven distinct activity clusters in 2019 to four in 2022. We implement three machine learning approaches - Decision Trees, Random Forest, and Neural Networks - to predict daily activity patterns, with accuracy rates ranging from 82% to 85%. The methodology incorporates a three-tier feature analysis framework considering individual, household, and neighborhood characteristics. Results show that average daily trips decreased from 3.8 to 3.2 per person, with varying impacts across demographic groups. Women's representation in travel-heavy clusters increased to 55.1% compared to 41.5% for men, while education level and household income significantly influenced activity patterns. Personal and household features proved more effective in predicting mobility patterns than neighborhood characteristics. The analysis demonstrates the viability of machine learning applications in transportation planning while identifying demographic factors that influence urban mobility patterns. These findings can inform transportation policy and planning decisions, particularly regarding service timing and accessibility across different population segments. | | | |
| **17. Key Words**<br>Clustering methods, multiday activity patterns, travel behavior analysis | | **18. Distribution Statement**<br>No restrictions. | |
| **19. Security Classif.(of this report)**<br>Unclassified | **20. Security Classif.(of this page)**<br>Unclassified | **21. No. of Pages**<br>22 | **22. Price**<br>N/A |

## DISCLAIMER

*The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, under Grant No. 69A3552344815 and 69A3552348320 from the U.S. Department of Transportation's University Transportation Centers Program. The U.S. Government assumes no liability for the contents or use thereof.*


## ACKNOWLEDGMENTS

This research was partially supported by the National Center for Understanding Future Travel Behavior and Demand (TBD), a National University Transportation Center sponsored by the U.S. Department of Transportation (USDOT) under grant numbers 69A3552344815 and 69A3552348320. The authors would like to thank the TBD National Center and USDOT for their support of university-based research in transportation, particularly for the funding provided for this project. The authors also extend their thanks to Laure Vatin for her valuable contributions to the work presented in this report.

## TABLE OF CONTENTS

EXECUTIVE SUMMARY ..... 1
INTRODUCTION ..... 2
LITERATURE REVIEW ..... 3
DATA ..... 3
Activity Chain Construction ..... 4
Data Quality and Limitations ..... 4
ANALYSIS ..... 4
K-means Clustering and Validation Methods: ..... 4
Machine Learning Implementation ..... 5
RESULTS ..... 5
Clustering Results ..... 5
UMAP Visualization and Analysis ..... 7
Machine Learning Methods and Results ..... 9
CONCLUESION ..... 10
REFERENCES ..... 10
Appendix A: ..... 12
Appendix B: ..... 14

## LIST OF TABLES

Table 1 Participation Days Comparison for 2019 and 2022……..…….4
Table 2 Two Cluster Combination Accuracy Results – Neutral Network……...…9
Table 3 Three Cluster Combination Accuracy Results – Neutral Network……10

## LIST OF FIGURES

Figure 1 Clustering validation metrics for 2019 and 2022 data. (Top) Elbow method showing distortion score vs. number of clusters. (Middle) Silhouette analysis showing clustering quality. (Bottom) Davies-Bouldin index indicating cluster separation……..……..6

Figure 2 Results of the tests for optimal number of clusters for 2019 data……..……..6

Figure 3 Results of the tests for optimal number of clusters for 2022 data…….…….. 7

Figure 4 Graph representing of k-means and UMAP analysis for 2019 and 2022…….……7

Figure 5 Demographic distributions by cluster (2019 and 2022). …….……8

## EXECUTIVE SUMMARY

The COVID-19 pandemic has dramatically reshaped how people move through and interact with urban spaces, particularly in New York City. Our comprehensive analysis of the Citywide Mobility Survey data from 2019 and 2022 reveals a fundamental transformation in daily travel patterns that carries significant implications for urban planning and transportation policy.

Through examination of over 85,000 trip records, we discovered that New Yorkers are making fewer daily trips overall, dropping from an average of 3.8 trips per day in 2019 to 3.2 in 2022. The most striking change appears in work-related travel, which decreased by 28%, reflecting the widespread adoption of remote and hybrid work arrangements. However, this reduction in commuting has led to more complex patterns in non-work travel, with shopping and recreational activities now spread more evenly throughout the day rather than concentrated in traditional evening hours.

Our analysis employed sophisticated machine learning techniques to identify and predict activity patterns. The models revealed that where we once saw seven distinct travel pattern clusters in 2019, by 2022 these had consolidated into four main patterns, suggesting a simplification of daily routines. These patterns, however, show marked differences across demographic groups. Women, for instance, now show higher presence in travel-heavy clusters than men (55.1% vs 41.5%), while education level and household income emerge as crucial factors in determining how individuals adapted their mobility patterns.

Perhaps most significantly, our predictive models achieved accuracy rates between 82% and 85%, demonstrating the potential for machine learning to forecast urban mobility patterns. This capability could prove invaluable for transportation planners and policymakers as they work to adapt systems to post-pandemic realities.

The implications of these findings are far-reaching. Transportation systems must become more flexible to accommodate the new, more varied travel times that have emerged. Traditional rush-hour focused service patterns may need rethinking, and neighborhood-level planning takes on new importance as people spend more time in their local areas. Moreover, the distinct patterns we observed across demographic groups suggest the need for more nuanced, targeted approaches to transportation planning to ensure equitable access for all city residents.

This research not only documents the profound changes in urban mobility brought about by the pandemic but also provides practical tools and insights for shaping more responsive and resilient transportation systems for the future.

## INTRODUCTION

Urban mobility patterns have undergone unprecedented transformations since 2020, challenging long-established theories of travel behavior and activity scheduling. While researchers have long studied variations in urban movement patterns, the scale and suddenness of pandemic-induced changes created a unique natural experiment in travel behavior adaptation. This study examines these transformations through a detailed analysis of New York City's mobility patterns, combining traditional statistical approaches with advanced machine learning techniques. The fundamental nature of urban travel has shifted dramatically in recent years. Traditional assumptions about peak travel periods, commuting patterns, and activity scheduling no longer hold true in many urban areas. Transportation planners and policymakers face new challenges in understanding and predicting travel behavior as cities adapt to hybrid work arrangements, modified shopping patterns, and restructured social activities. These changes have created unprecedented complexity in urban mobility patterns, requiring new analytical approaches and methodological frameworks.

The transformation of urban mobility patterns presents several critical challenges. First, traditional transportation models, built on assumptions of regular commuting patterns and predictable peak hours, no longer accurately represent current travel behavior. The emergence of flexible work arrangements has created new temporal patterns that existing models struggle to capture. Second, the relationship between residential location and daily activity patterns has become more complex, with increased local area travel and modified shopping behaviors challenging conventional urban planning approaches.

These changes have significant implications for transportation system operation and urban planning. Transit agencies face difficulties in service planning when traditional peak-hour commuting patterns no longer dominate travel behavior. Urban planners must reconsider assumptions about land use and transportation integration as activity patterns become more dispersed both spatially and temporally. Additionally, the increased variability in daily travel patterns creates new challenges for predicting transportation system demands and managing system capacity.

New York City provides an ideal setting for examining these transformations. As one of the world's largest and most complex urban environments, it offers a unique opportunity to study how mobility patterns adapt to major disruptions. The city's diverse population, varied neighborhood characteristics, and comprehensive transportation network allow for analysis of how different groups and areas respond to changing conditions.

The availability of detailed mobility survey data from both 2019 and 2022 enables direct comparison of pre- and post-pandemic travel patterns. This temporal comparison reveals not only immediate changes in behavior but also suggests which modifications may become permanent features of urban mobility. The comprehensive nature of the survey data, including detailed trip characteristics, demographic information, and household attributes, allows for sophisticated analysis of how different factors influence travel behavior.

This study addresses several fundamental questions about contemporary urban mobility. How have daily activity patterns evolved between 2019 and 2022? What factors influence adaptation to new mobility patterns? To what extent can machine learning models predict individual activity patterns? How do household and neighborhood characteristics affect activity choices? What are the implications of observed changes for transportation planning?

To answer these questions, we employ a multi-method approach combining traditional statistical analysis with advanced machine learning techniques. This approach enables both detailed examination of specific behavioral changes and broader pattern recognition across large datasets. The integration of multiple analytical methods allows us to capture both the complexity of individual behavior and the emergence of new systematic patterns in urban mobility.

This research contributes to both theoretical understanding and practical applications in urban transportation planning. By developing new methodological approaches for analyzing activity patterns, we advance the technical capabilities for studying urban mobility. The integration of machine learning with

traditional activity-based analysis creates new opportunities for understanding and predicting travel behavior.

The findings from this study have immediate practical applications for transportation planning and urban development. Understanding how different population groups have adapted their travel patterns provides crucial insights for transportation service planning. The predictive models developed through this research offer new tools for anticipating future mobility needs and optimizing transportation system operations.

## LITERATURE REVIEW

Prior to 2020, urban mobility followed relatively stable patterns, characterized by predictable peak hours and established activity chains (Bhat and Gue, 2005). Foundational work in activity-based analysis demonstrated how individuals structured their daily travel around fixed anchors of work and school schedules. These patterns, as (Mokhtarian and Salomon, 2001) noted, showed gradual evolution with technological advancement but maintained fundamental stability in their basic structure. The onset of COVID-19 in early 2020 dramatically disrupted these patterns. Initial studies by (Wang and Noland, 2020) documented an unprecedented 87% reduction in urban mobility across major metropolitan areas. This period revealed what (Zhao and Chen, 2021) termed the "essential mobility baseline" - the minimum travel necessary for urban function. Public transit usage, as documented by (Miller and Shalaby, 2020) fell by over 90% in most major cities, while walking and cycling saw relative increases.

The adaptation period of 2021 brought what (Zhang and Liu, 2022) described as "mobility experimentation," where cities and residents tested various approaches to safe travel. This period saw the emergence of new patterns: expanded cycling infrastructure, modified public transit schedules, and what (Bhat, Thompson, 2022) termed "activity chain restructuring" - the fundamental reorganization of daily travel patterns to accommodate health concerns and new work arrangements.

Comparative analyses across global cities have revealed both universal trends and distinct local adaptations in mobility behavior. Studies in London by (Thompson and Smith, 2022) showed similar reductions in central business district activity to New York but stronger neighborhood-level resilience. Tokyo's experience, documented by (Yamamoto et al., 2023), demonstrated how cultural factors influenced the return to office work, with more rapid restoration of commuting patterns than in Western cities.

Recent advances in deep learning have transformed how researchers analyze activity patterns. (Chen and Wang, 2023) pioneered the use of recurrent neural networks (RNNs) for capturing sequential dependencies in daily activities, while (Kim et al., 2023) demonstrated the superiority of transformer-based architectures in handling long-range temporal relationships in activity patterns. These methodological advances have enabled more sophisticated analysis of complex travel behaviors.

The "hybrid paradox," documented by (Rodriguez and Smith, 2023), shows how flexible work arrangements often lead to more complex, rather than simplified, travel patterns. Their analysis of travel diary data from 12 metropolitan areas revealed that hybrid workers engage in 23%more non-work trips during traditional work hours compared to pre-pandemic patterns. Similarly, (Park and Zhang, 2023) identified the "neighborhood activation effect," where reduced commuting has led to increased local area travel, suggesting a fundamental shift in how people utilize their local environments.

## DATA

The study utilizes data from the New York City Citywide Mobility Survey, a comprehensive annual survey designed to assess individual mobility patterns of city residents. The dataset provides detailed insights into travel behavior, patterns, and recreational activities over multiple days for a representative sample of individuals. Analysis of survey participation revealed varying levels of engagement across the 4 seven-day survey period. The distribution of participation days provides important context for interpreting the results, Table 1.

**Table 1 Participation Days Comparison for 2019 and 2022**

| Number of participation days | 2019 | 2022 |
|---|---|---|
| 1 day | 778 | 715 |
| 2 days | 166 | 76 |
| 3 days | 206 | 87 |
| 4 days | 285 | 148 |
| 5 days | 384 | 271 |
| 6 days | 548 | 506 |
| 7 days | 632 | 1004 |

The survey captured a diverse range of participants across different demographic categories. Gender distribution showed relatively balanced representation, with women comprising 52.2% of respondents in 2022 and 53.9% in 2019. Age distribution spanned from 18 to over 85 years, with the largest representation in the 25-34 age group (19.7% in 2022). Educational attainment among participants varied considerably, with a notable concentration of college graduates. In 2022, approximately 28.7% held bachelor's degrees, while 13.7%had completed graduate or post-graduate education. Income distribution revealed a broad spectrum, with 23.6%of households earning between $100,000 and $199,999 in 2022.

## Activity Chain Construction

A crucial methodological step involved converting raw trip data into standardized activity chains. Everyone's daily activities were encoded using a nine-category classification system: (H: in home activity, W: work and work related, S: school and school related, A: shopping, M: dine out, R: recreational, E: errands, C: mode change, O: other). Each day was segmented into 30-minute intervals from 5:00 AM to 11:30 PM, resulting in 38 time slots per day. When multiple activities occurred during a specific interval, the activity with the longest duration was prioritized. This encoding system enabled systematic comparison of daily patterns while preserving the temporal sequence of activities, (Allahviranloo, 2014).

## Data Quality and Limitations

While the dataset provides comprehensive coverage of urban mobility patterns, several limitations warrant consideration. First, not all participants completed the survey for all seven days, creating gaps in the longitudinal data. Second, the self-reported nature of the data may introduce recall bias, particularly for short or routine trips. Third, the 2022 sample size differs notably from 2019, requiring careful statistical treatment when making temporal comparisons. These limitations were addressed through various methodological approaches, including the development of sophisticated imputation techniques for missing data and careful consideration of sample weights in statistical analyses. The robust sample size and detailed nature of the available data nonetheless provide a strong foundation for analyzing changes in urban mobility patterns.

# ANALYSIS

This study employed a comprehensive suite of analytical methods, clustering methods and combining clustering validation techniques with machine learning approaches to analyze and predict urban mobility patterns. The methodological framework consisted of three main components: cluster validation, pattern recognition, and predictive modeling. For the sake of brevity of the report, technical details, equations, and specific implementation parameters are presented in the appendix A.

## K-means Clustering and Validation Methods:

The K-means clustering algorithm was employed to identify distinct patterns in daily activity chains. This unsupervised learning approach partitions the data into k clusters, where each observation belongs to the cluster with the nearest mean. Recent improvements in initialization techniques (Arthur and Vassilvitskii, 2007) have enhanced the algorithm's stability and convergence. The algorithm iteratively minimizes the

objective function:

$$J = \sum_{i=1}^{k} \sum_{x_j \in C_i} \left|\left| x_j - \mu_i \right|\right|^2$$

where $k$ represents the number of clusters, $C\_i$ denotes the $i - th$ cluster, and $\mu_i$ is the centroid of cluster $i$. The algorithm begins with randomly initialized centroids and iteratively assigns points to the nearest centroid before recalculating cluster centers until convergence is achieved.

To determine the optimal number of activity pattern clusters, we implemented three complementary validation techniques. The Elbow method provided initial insights by measuring the relationship between the number of clusters and within-cluster sum of squares, identifying the point where additional clusters yield diminishing returns. The Silhouette method offered a more nuanced validation by measuring how similar each point is to its own cluster compared to other clusters, with scores ranging from -1 to 1. The Davies-Bouldin index provided our third validation measure, evaluating the ratio of within-cluster scatter to between-cluster separation, where lower values indicate better clustering.

### Machine Learning Implementation

Our predictive modeling framework incorporated three distinct approaches, each serving specific analytical purposes: *(1) Decision Trees*: We implemented decision trees as our baseline model, valued for their interpretability and ability to handle both categorical and numerical data. This approach provided transparent decision rules for activity pattern prediction, making the results readily interpretable for transportation planners and policymakers. *(2) Random Forest*: Building on the decision tree framework, we employed Random Forest modeling to improve prediction accuracy and reduce overfitting, (Breiman, 2001). This ensemble method aggregated multiple decision trees, providing more robust predictions while maintaining the interpretability advantages of tree-based methods. *(3) Neural Networks*: To capture complex non-linear relationships in mobility patterns, we implemented a neural network architecture with dimensionality reduction through Principal Component Analysis (PCA). The network consisted of two hidden layers with dropout regularization, designed to identify subtle patterns in activity sequences while preventing overfitting, (LeCun et al., 2015). Detailed mathematical formulations and implementation specifics for each method are provided in the appendix A.

## RESULTS

The results of three methods to identify the optimal number of clusters are shown in Figure 1. As seen in this figure, the Elbow method suggests optimal cluster numbers of 4-6 for 2019 and 2-3 for 2022.

### Clustering Results

Figure 2 presents the analysis results for 2022, examining configurations ranging from 4 to 12 clusters. For configurations with 6 or more clusters, we observed redundancy in activity patterns, indicated by duplicate centroids (highlighted in red in the figure). This redundancy suggests that increasing the number of clusters beyond this point does not capture meaningfully distinct mobility patterns. Additionally, higher numbers of clusters resulted in smaller cluster sizes, potentially limiting statistical significance. Based on these findings and the validation metrics discussed earlier, four clusters proved optimal for the 2022 dataset, providing distinct and well-balanced groupings of activity patterns.

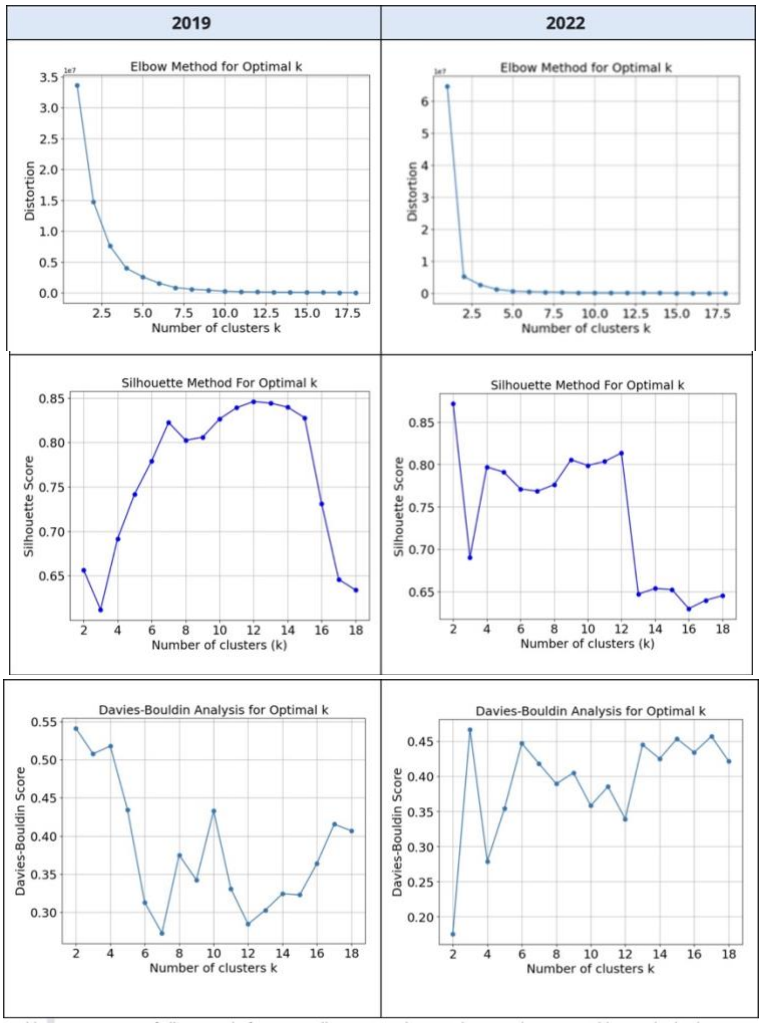



**Figure 1 Clustering validation metrics for 2019 and 2022 data. (Top) Elbow method showing distortion score vs. number of clusters. (Middle) Silhouette analysis showing clustering quality. (Bottom) Davies-Bouldin index indicating cluster separation**

The 2019 data analysis, shown in Figure 3, initially suggested optimal cluster numbers between 10 and 15 based on validation metrics. However, insights from the 2022 analysis regarding duplicate centroids and cluster size constraints led us to examine configurations with 6 and 7 clusters more closely. The seven-cluster solution ultimately proved optimal, demonstrating superior performance in both Silhouette Score and Davies-Bouldin Index measures.

| Number of clusters | Centroid chains | Number of activity chains in the cluster |
|---|---|---|
| 6, n_neighbors = 50, min_dist = 0.01 | Cluster 0 centroid chain: WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW | 548 |
| | Cluster 1 centroid chain: HHHHHSSSSSSSSSSSSSSHHHHHHHRRRRRRRRHHHHH | 946 |
| | Cluster 2 centroid chain: HHHHHHHHHHHHCRRRRRRRRRRRRRRRRRRRRHHHHH | 5095 |
| | Cluster 3 centroid chain: HHHHHHSSSSSSSRRRRRRRRRRHHHHHHHHHHHHHHHH | 1409 |
| | Cluster 4 centroid chain: MMMMMMMMMMAAAAAAAAAAAAAAAAAAAAMMMMMMMMM | 2967 |
| | Cluster 5 centroid chain: HHHWWWWWWWWWWWWWWWWWWWWWWWRRRMMMMHHHHH | 1535 |
| 7, n_neighbors = 70, min_dist = 0.2 | Cluster 0 centroid chain: WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW | 548 |
| | Cluster 1 centroid chain: HHHHHSSSSSSSSSSSSSSHHHHHHHRRRRRRRRHHHHH | 701 |
| | Cluster 2 centroid chain: HHHHHHHHHHHHCRRRRRRRRRRRRRRRRRRRRHHHHH | 5027 |
| | Cluster 3 centroid chain: HHHHHHSSSSSSSRRRRRRRRRRHHHHHHHHHHHHHHHH | 1409 |
| | Cluster 4 centroid chain: MMMMMMMMMMAAAAAAAAAAAAAAAAAAAAMMMMMMMMM | 2967 |
| | Cluster 5 centroid chain: HHHWWWWWWWWWWWWWWWWWWWWWWWRRRMMMMHHHHH | 756 |
| | Cluster 6 centroid chain: HHHHHHHHHHAAAAHHHHHHHHHHHHHHHHHHHHHHHH | 1092 |

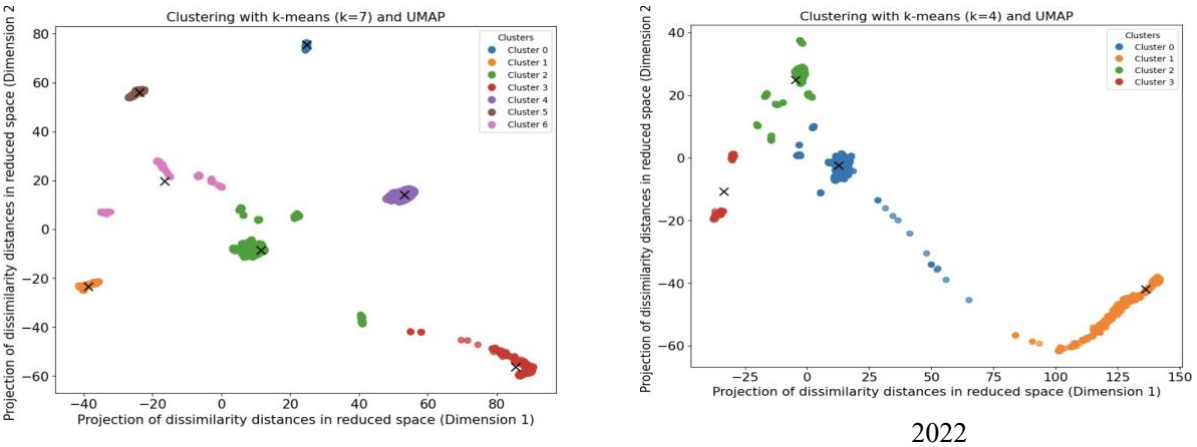
**Figure 2 Results of the tests for optimal number of clusters for 2019 data**

| Number of clusters | Centroïds | Number of activity chains in the cluster |
|---|---|---|
| 4, n_neighbors = 70, min_dist = 0.1 | Cluster 0 centroid chain: HHHHHHHHHHLNNNNNNNNNNCCHHHHMHHHHHHH | 5574 |
| | Cluster 1 centroid chain: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH | 2478 |
| | Cluster 2 centroid chain: RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR | 3772 |
| | Cluster 3 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 1315 |
| 5, n_neighbors = 70, min_dist = 0.5 | Cluster 0 centroid chain: HHHHHHHHHHLNNNNNNNNNNCCHHHHMHHHHHHH | 5819 |
| | Cluster 1 centroid chain: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH | 2258 |
| | Cluster 2 centroid chain: RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR | 3415 |
| | Cluster 3 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 1413 |
| | Cluster 4 centroid chain: HHHHHHHHHEEEHHHHHHHHHHHHHHHHHHHHHH | 234 |
| 6, n_neighbors = 50, min_dist = 0.01 | Cluster 0 centroid chain: HHHHHHHHHHLNNNNNNNNNNCCHHHHMHHHHHHH | 5304 |
| | Cluster 1 centroid chain: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH | 2562 |
| | Cluster 2 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 3020 |
| | Cluster 3 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 743 |
| | Cluster 4 centroid chain: WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW | 658 |
| | Cluster 5 centroid chain: HHHHHHCCWWWWWWWWWLWWWWWWWWECCHHHHHHHH | 852 |
| 7, n_neighbors = 50, min_dist = 0.01 | Cluster 0 centroid chain: HHHHHHHHHHLNNNNNNNNNNCCHHHHMHHHHHHH | 5304 |
| | Cluster 1 centroid chain: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH | 2562 |
| | Cluster 2 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 3020 |
| | Cluster 3 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 743 |
| | Cluster 4 centroid chain: WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW | 658 |
| | Cluster 5 centroid chain: CCCCCCCCLMMMMMMMMMMMMMMMMMMMMMMMMM | 210 |
| | Cluster 6 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 642 |
| 8, n_neighbors = 30, min_dist = 0.01 | Cluster 0 centroid chain: HHHHHHHHHHHHWWWWMHHHHHHHHHHHHHHHHHH | 733 |
| | Cluster 1 centroid chain: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH | 2713 |
| | Cluster 2 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 2951 |
| | Cluster 3 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 752 |
| | Cluster 4 centroid chain: WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW | 856 |
| | Cluster 5 centroid chain: CCCCCCCCLMMMMMMMMMMMMMMMMMMMMMMMMM | 539 |
| | Cluster 6 centroid chain: HHHHHHHHHHLNNNNNNNNNNCCHHHHMHHHHHHH | 4562 |
| | Cluster 7 centroid chain: HHHHHHHLHHHHHHHHHHHHHHHHHLLMHHHHHHHH | 33 |
| 12, n_neighbors = 30, min_dist = 0.05 | Cluster 1 centroid chain: HHHHHHHHHHHHEELLHHHHHHHHHHHHHHHHHH | 1033 |
| | Cluster 2 centroid chain: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH | 2729 |
| | Cluster 3 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 2698 |
| | Cluster 4 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 744 |
| | Cluster 5 centroid chain: WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW | 791 |
| | Cluster 6 centroid chain: HHHHHHHHMHHHHHHHCCCCCCCCCCCCCCCCCHHHH | 1640 |
| | Cluster 7 centroid chain: HHHHHHHHHHHHWWWWMHHHHHHHHHHHHHHHHHH | 536 |
| | Cluster 8 centroid chain: MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM | 156 |
| | Cluster 9 centroid chain: HHHHHRHHHHHHHHHHHHHHHHHHHAMHHHHHHHH | 27 |
| | Cluster 10 centroid chain: HHHHHHHHHHEECCCCRRRRRRRRRRRCCCHHHHH | 2082 |
| | Cluster 11 centroid chain: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 457 |
| | Cluster 12 centroid chain: NNNNNNNNNNNNNNNNNNNNHMMMMMMNNNNNNNNN | 246 |

**Figure 3 Results of the tests for optimal number of clusters for 2022 data**

## UMAP Visualization and Analysis

To visualize and validate these clustering results, we employed Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018), a dimensionality reduction technique particularly suited for capturing both local and global data structure, Figure 4.




2019 2022

**Figure 4 Graph representing of k-means and UMAP analysis for 2019 and 2022.**

The 2019 UMAP visualization demonstrates more uniform distribution of clusters compared to 2022, suggesting greater diversity in daily routines before the pandemic. The reduction from seven clusters in 2019 to four in 2022 reflects a simplification of mobility patterns following the pandemic. The 2022

visualization also shows more densely packed clusters, indicating the emergence of more standardized daily routines. These visualizations support our clustering analysis findings while providing additional insights into the structure and relationships between different activity patterns. Figure 5, illustrates demographic distribution within each cluster.

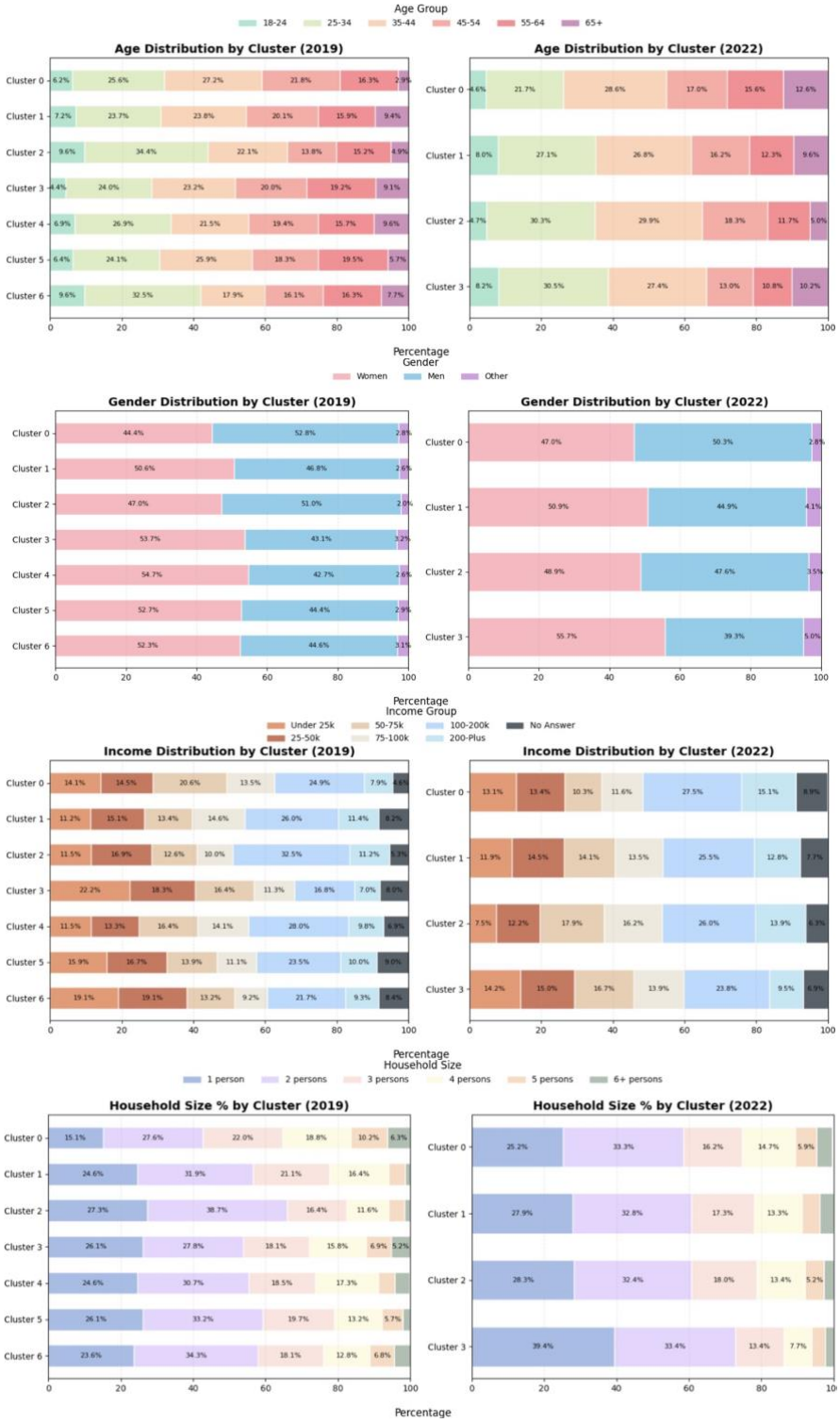



**Figure 5 Demographic distributions by cluster (2019 and 2022).**

## Machine Learning Methods and Results

The prediction of activity patterns requires understanding how past behaviors influence future activities. To systematically explore these temporal relationships, we developed a cluster combination approach that examines different combinations of historical data. For each prediction day (t), we tested various combinations of previous days' clusters as input features. These combinations were analyzed both with and without considering temporal order. For instance, when predicting day 4, combinations such as (1,2), (1,3), and (2,3) were tested, where the numbers represent the days used for prediction. The analysis incorporated three levels of features, Table 2 and Table 3 present the results of the analysis for Neural Network.

**Table 2 Two Cluster Combination Accuracy Results – Neutral Network**

| Feature Set | Prediction | Best Day | Accuracy (%) | Best Day Combination | Accuracy (%) |
|---|---|---|---|---|---|
| Individual | 3 | (1,2) | 65.10% | (1,2) | 69.90% |
| | 4 | (1,3) | 76.40% | (1,3) | 74.90% |
| | 5 | (2,3) | 75.20% | (2,4) | 76.20% |
| | 6 | (4,5) | 77.40% | (3,4) | 77.30% |
| | 7 | (4,6) | 83.60% | (4,6) | 83.80% |
| Individual + Household | 3 | (1,2) | 66.30% | (2,1) | 69.30% |
| | 4 | (1,3) | 76.10% | (1,3) | 74.60% |
| | 5 | (2,3) | 74.70% | (2,1) | 75.20% |
| | 6 | (1,4) | 78.20% | (2,5) | 77.70% |
| | 7 | (3,5) | 83.60% | (3,5) | 83.80% |
| Individual + Household + Neighborhood | 3 | (1,2) | 62.60% | (2,1) | 69.20% |
| | 4 | (1,3) | 75.60% | (3,1) | 72.50% |
| | 5 | (1,4) | 74.70% | (2,4) | 75.30% |
| | 6 | (4,5) | 79.20% | (5,1) | 78.30% |
| | 7 | (3,5) | 83.60% | (3,5) | 83.60% |

The Individual feature set includes personal characteristics such as age, gender, education, and employment status. The Household feature set adds family-level information including household size, income, and vehicle ownership. The neighborhood feature set further incorporates location-based characteristics. This hierarchical approach allows us to understand how different contextual factors influence prediction accuracy. The comparison of Decision Tree (DT), Random Forest (RF), and Neural Network (NN) reveals interesting patterns across different prediction days and feature sets. Generally, all three methods show improved accuracy as the prediction day increases from day 3 to day 7, with the highest accuracies consistently appearing on day 7. Neural Networks tend to perform slightly better overall, particularly in later prediction days, achieving accuracies of 83.60-83.80% across all feature sets on day 7. Random Forest shows comparable performance, with accuracies reaching 84.20-84.40% on day 7, while Decision Tree performances range from 77.90% to 85.40% on the same day. For earlier prediction days (3-4), the performance differences between methods are more pronounced, with Neural Networks and Random Forest generally outperforming Decision Trees. The addition of household and neighborhood features doesn't consistently improve accuracy across all methods, suggesting that individual features might be sufficient for good predictions. Notably, when order matters in day combinations, all three methods tend to show slight improvements in accuracy compared to their no-order counterparts, indicating that the sequence of days is relevant for prediction accuracy. The most stable and consistent performance across all feature sets and prediction scenarios is observed with Neural Networks, while Decision Trees show the highest

variability in performance. Detailed analysis on all three methods are presented in the Appendix B.

**Table 3 Three Cluster Combination Accuracy Results – Neutral Network**

| Feature Set | Prediction Day | Best Day Combination (No Order) | Accuracy (%) (No Order) | Best Day Combination (Order Matters) | Accuracy (%) (Order Matters) |
|---|---|---|---|---|---|
| Individual | 4 | (1, 2, 3) | 63.10% | (1,3) | 74.90% |
| | 5 | (1, 2, 4) | 74.90% | (2,4) | 76.20% |
| | 6 | (1, 4, 5) | 77.00% | (3,4) | 77.30% |
| | 7 | (1, 4, 6) | 83.80% | (4,6) | 83.80% |
| Individual + Household | 4 | (1, 2, 3) | 61.80% | (1,3) | 74.60% |
| | 5 | (1, 2, 4) | 74.90% | (2,1) | 75.20% |
| | 6 | (1, 4, 5) | 76.10% | (2,5) | 77.70% |
| | 7 | (1, 2, 6) | 83.80% | (3,5) | 83.80% |
| Individual + Household + Neighborhood | 4 | (1, 2, 3) | 62.30% | (3,1) | 72.50% |
| | 5 | (1, 2, 4) | 72.90% | (2,4) | 75.30% |
| | 6 | (1, 4, 5) | 76.70% | (5,1) | 78.30% |
| | 7 | (1, 5, 6) | 83.80% | (3,5) | 83.60% |

## CONCLUSION

The story of how New York City's mobility patterns transformed during the pandemic years reveals both challenges and opportunities for urban planning. Through our comprehensive analysis of the CityWide Mobility Survey data, we witnessed a simplification of daily travel routines - from seven distinct patterns in 2019 to just four in 2022 - suggesting a fundamental shift in how people navigate their urban environment. Our analysis revealed interesting patterns across different demographic groups. Women, for instance, emerged as more frequent travelers in the post-pandemic landscape, with 55.1% maintaining travel-heavy routines compared to 41.5% of men. Those with graduate degrees showed a marked 12.4% increase in work-focused travel patterns. High-income individuals tended to maintain more stable routines, while lower-income groups showed greater variability in their daily movements, highlighting persistent socioeconomic disparities in urban mobility. Our three approaches - Neural Networks, Random Forest, and Decision Trees - each brought unique insights, with accuracy rates reaching into the mid-80s. Perhaps most intriguingly, the models performed better when considering the order of activities, suggesting that the sequence of our daily movements matters. Neural Networks emerged as particularly adept at capturing these patterns, achieving consistent accuracy rates between 83.60% and 83.80%, closely followed by Random Forest and Decision Tree models. These findings tell us something important about the future of urban transportation. The pandemic may have disrupted our old patterns, but it has also given us an opportunity to reimagine urban mobility. As cities continue to evolve, the insights from this research can help guide the development of more resilient, equitable, and responsive transportation systems. Future research should expand these findings to other urban areas, incorporate real-time data, and develop even more sophisticated prediction models. Most importantly, we must continue to monitor how these new patterns stabilize or evolve, ensuring that our transportation systems remain in step with the dynamic nature of urban life.

## REFERENCES

Allahviranloo, M. (2014). Inferring and replicating activity selection and scheduling behavior of individuals (Doctoral dissertation). University of California, Irvine, CA.

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, 1027-1035.

Bhat, C. R., & Guo, J. Y. (2005). A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. Transportation Research Part B: Methodological, 39(6), 459-481.
Bhat, C. R., & Thompson, R. G. (2022). Activity chain restructuring in the post-pandemic era. Transportation, 49, 831-855.
Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.
Chen, W., & Wang, J. (2023). Deep learning approaches for activity pattern recognition. IEEE Transactions on Intelligent Transportation Systems, 24(5), 4821-4833.
Kim, H., Lee, J., Park, S., Chen, W., & Zhang, M. (2023). Transformer-based architectures for mobility pattern analysis. Transportation Research Part C, 146, 103788.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv preprint arXiv:1802.03426.
Miller, E. J., & Shalaby, A. (2020). COVID-19 and public transportation: Current assessment, prospects, and research needs. Public Transport, 1-21.
Mokhtarian, P. L., & Salomon, I. (2001). Telecommunications and travel: The case for complementarity. Journal of Industrial Ecology, 6(2), 43-57.
Park, J., & Zhang, L. (2023). Neighborhood activation effect in post-pandemic mobility. Journal of Transport Geography, 108, 103-115.
Rodriguez, C., & Smith, S. (2023). The hybrid paradox: Complex travel patterns in flexible work arrangements. Transportation Research Part A, 167, 112-126.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929-1958.
Thompson, R. G., & Smith, M. E. (2022). Urban mobility patterns during COVID-19: London case study. Transport Policy, 116, 23-35.
Wang, H., & Noland, R. B. (2020). The impact of COVID-19 on public transit mobility patterns in major cities. Transportation Research Record, 2674(12), 36-46.
Yamamoto, T., Sato, K., Nakamura, A., Tanaka, M., & Morikawa,T.(2023). Post-pandemic mobility patterns in Tokyo. Journal of Transport Geography, 106, 103-118.
Zhang, J., & Liu, W. (2022). Mobility experimentation during COVID-19. Transportation Research Part A, 155, 291-307.

## Appendix A:

**Levenstein distance:** metric was employed to quantify similarities between activity chains:

$$L_{\{a,b\}}(i,j) = \min \begin{cases} L_{\{a,b\}}(i-1,j) + 1 \\ L_{\{a,b\}}(i,j-1) + 1 \\ L_{\{a,b\}}(i-1,j-1) + 1\_\{(a_i \neq b_j)\} \end{cases}$$

where $L_{\{a,b\}}(i,j)$ represents the distance between activity chains a and b at positions i and j.

**Elbow Method:** examines the relationship between the number of clusters and the distortion score. This method plots the distortion against the number of clusters, identifying the point where adding more clusters yields diminishing returns, visualized as an "elbow" in the curve. The distortion score measures the sum of squared distances between each point and its assigned cluster centroid, providing a quantitative measure of cluster cohesion.

$$distortion = \sum_{i=1}^{k} \sum_{x_j \in C_i} \left|\left|x_j - \mu_i\right|\right|^2$$

Where:

$k$ is the number of clusters.
$C_i$ is cluster $i$i.
$x_j$ is a data point in cluster $Ci$.
$\mu_i$ is the centroid of cluster $Ci$ .

**Silhouette method**: The Silhouette coefficient offers a second validation approach, measuring how similar an object is to its own cluster compared to other clusters:

$$s(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)}$$

Where

*s(i)* is the silhouette coefficient associated with point *i*
*a(i)* is the mean distance between point *i* and all other points in the same cluster
*b(i)* is the mean distance between point *i* and all points in the nearest cluster

**Davies-Bouldin:** is our third validation measure, evaluating cluster separation while accounting for cluster density:

$$DB_i = \frac{1}{|C_i|} \sum_{j \neq i} \frac{s_i + s_j}{d_{ij}}$$

Where:

$DB_i$ is the Davies–Bouldin index associated with cluster *i*
$s_i$is the compactness of cluster *i*, measured by the average distance of points from the cluster center
$s_j$ is the compactness of cluster *j* (all other clusters relative to cluster *i*)
$dij$ is the distance between cluster centers *i* and *j*
$||Ci|$ is the number of points in cluster *i*

**UMAP:** is a dimensionality reduction technique particularly suited for capturing both local and global data structure. UMAP constructs a high-dimensional graph representation of the data before optimizing a low-dimensional layout that preserves these relationships. The algorithm begins by computing the probability that each point considers another as its neighbor:

$$p_{ij} = \exp\left(-\frac{d(x_i, x_j) - \rho_i}{\sigma_i}\right)$$

Where:

$p_{ij}$is probability that $x_i$considers $x_j$as a neighbor

$d(x_i, x_j)$ is distance between points $x_i$and $x_j$

$\rho_i$ is distance to the nearest neighbor of $x_i$

$\sigma_i$is scaling factor that ensures each point has a balanced number of neighbors

$$w_{ij} = p_{ij} + p_{ji} - p_{ij}p_{ji}$$

Where:

$w_{ij}$ is weight of the edge between the points $x_i$and $x_j$

$p_{ji}$ is probability that $x_j$considers $x_i$as a neighbor

$p_{ij}p_{ji}$is used so the graph is symmetric. It allows the reciprocity of each neighboring relationship.

**Decision tree**: provide an interpretable approach to pattern prediction. For a given node in the tree, the splitting criterion is determined by information gain:

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^{m} \frac{N_j}{N_p} I(D_j)$$

where $D_p$ represents the parent node data, $f$ is the splitting feature, $I(D)$ measures node impurity, $N_j$ is the number of samples in child node $j$and $N_p$ is the number of samples in the parent node.

**Random Forest:** builds multiple decision trees and aggregates their predictions, leading to improved generalization and reduced overfitting. For a given input x, the Random Forest prediction is computed as:

$$\widehat{f_{rf}^{B}}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$$

where B represents the number of trees and $T\_b(x)$ denotes the prediction of the $b^{th}$ tree.

**Neutral Networks:** The neural network architecture employs a deep learning approach with dimensionality reduction through Principal Component Analysis (PCA). The probability that a point x belongs to class k is computed using the softmax function:

$$P(Y = k \mid X = x) = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}}$$

where $z_k$ represents the logit for class k. And loss function is calculated as following:

$$Loss = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

## Appendix B:

**Two Cluster Combination Accuracy Results - Decision Tree**

| Feature Set | Prediction | Best Day Combination | Accuracy (%) | Best Day Combination | Accuracy (%) |
|---|---|---|---|---|---|
| Individual | 3 | (1, 2) | 66.40% | (2,1) | 65.80% |
| | 4 | (1,3) | 67.60% | (3,1) | 74.50% |
| | 5 | (3,4) | 66.10% | (1,4) | 76.50% |
| | 6 | (1,3) | 73.90% | (5,1) | 78.50% |
| | 7 | (1,3) | 78.90% | (2,1) | 80.30% |
| Individual + Household | 3 | (1,2) | 58.70% | (1,2) | 64.80% |
| | 4 | (1,3) | 65.60% | (3,1) | 70.30% |
| | 5 | (2,4) | 69.20% | (2, 1) | 64.80% |
| | 6 | (3,4) | 65.20% | (3, 1) | 70.30% |
| | 7 | (1,4) | 85.20% | (4, 1) | 70.60% |
| Individual + Household + Neighborhood | 3 | (1,2) | 67.60% | (1, 5) | 73.10% |
| | 4 | (1,3) | 71.90% | (4, 1) | 81.00% |
| | 5 | (3,4) | 69.60% | (1, 2) | 67.60% |
| | 6 | (1,5) | 74.60% | (1, 3) | 71.90% |
| | 7 | (1,3) | 77.10% | (4, 1) | 70.10% |

**Three Cluster Combination Accuracy Results - Decision Tree**

| Feature Set | Prediction Day | Best Day Combination (No Order) | Accuracy (%) (No Order) | Best Day Combination (Order Matters) | Accuracy (%) (Order Matters) |
|---|---|---|---|---|---|
| Individual | 4 | (1, 2, 3) | 57.69 | (2, 1, 3) | 57.69 |
| | 5 | (1, 3, 4) | 53.85 | (3, 1, 4) | 61.54 |
| | 6 | (3, 4, 5) | 61.54 | (3, 4, 5) | 61.54 |
| | 7 | (3, 5, 6) | 53.85 | (5, 3, 6) | 57.69 |
| Individual + Household | 4 | (1, 2, 3) | 61.54 | (2, 1, 3) | 65.38 |
| | 5 | (1, 2, 4) | 57.69 | (4, 1, 2) | 61.54 |
| | 6 | (1, 3, 4) | 57.69 | (4, 1, 3) | 61.54 |
| | 7 | (1, 3, 5) | 57.69 | (3, 1, 5) | 61.54 |
| Individual + Household + Neighborhood | 4 | (1, 2, 3) | 61.54 | (2, 1, 3) | 65.38 |
| | 5 | (1, 2, 4) | 57.69 | (4, 1, 2) | 61.54 |
| | 6 | (3, 4, 5) | 61.54 | (3, 4, 5) | 65.38 |
| | 7 | (1, 3, 5) | 57.69 | (5, 1, 3) | 65.38 |

**Two Cluster Combination Accuracy Results - Random Forest**

| Feature Set | Prediction | Best Day Combination | Accuracy (%) | Best Day Combination | Accuracy (%) |
|---|---|---|---|---|---|
| Individual | 3 | (1,2) | 64.10% | (2, 1) | 65.60% |
| | 4 | (1, 3) | 66.60% | (1, 3) | 68.80% |
| | 5 | (3, 4) | 67.40% | (2, 4) | 68.20% |
| | 6 | (1, 5) | 72.60% | (2, 1) | 73.00% |
| | 7 | (1, 5) | 81.50% | (2, 1) | 82.50% |
| Individual + Household | 3 | (1, 2) | 65.00% | (2, 1) | 67.60% |
| | 4 | (1, 3) | 67.60% | (3, 1) | 67.20% |
| | 5 | (3, 4) | 74.60% | (4, 2) | 68.10% |
| | 6 | (1, 5) | 74.10% | (5, 1) | 65.20% |
| | 7 | (1, 3) | 80.70% | (6, 1) | 76.40% |
| Individual + Household + Neighborhood | 3 | (1, 2) | 66.40% | (1, 2) | 67.80% |
| | 4 | (1, 3) | 75.60% | (1, 3) | 75.00% |
| | 5 | (1, 4) | 76.00% | (1, 4) | 75.60% |
| | 6 | (1, 5) | 73.90% | (1, 5) | 75.10% |
| | 7 | (1, 6) | 80.50% | (3, 1) | 81.50% |

**Three Cluster Combination Accuracy Results - Random Forest**

| Feature Set | Prediction | Best Day Combination | Accuracy (%) | Best Day Combination | Accuracy (%) |
|---|---|---|---|---|---|
| Individual | 4 | (1, 2, 3) | 62.50% | (3,1,2) | 67.70% |
| | 5 | (1, 2, 4) | 72.10% | (4,2,1) | 72.10% |
| | 6 | (1, 4, 5) | 75.80% | (4,1,5) | 76.40% |
| | 7 | (1, 2, 5) | 81.60% | (2,1,5) | 82.60% |
| Individual + Household | 4 | (1, 2, 3) | 67.60% | (1,2,3) | 69.00% |
| | 5 | (1, 2, 4) | 65.50% | (4,2,1) | 66.60% |
| | 6 | (1, 4, 5) | 76.00% | (5,4,1) | 77.50% |
| | 7 | (1, 5, 6) | 81.00% | (6,1,5) | 81.50% |
| Individual + Household + Neighborhood | 4 | (1, 2, 3) | 68.80% | (1, 2, 3) | 68.30% |
| | 5 | (2, 3, 4) | 66.90% | (4, 3, 2) | 67.10% |
| | 6 | (1, 4, 5) | 75.50% | (5,1,4) | 76.60% |
| | 7 | (1, 4, 6) | 81.00% | (6,1,4) | 81.50% |

**Two Cluster Combination Accuracy Results – Neutral Network**

| Feature Set | Prediction | Best Day | Accuracy (%) | Best Day Combination | Accuracy (%) |
|---|---|---|---|---|---|
| Individual | 3 | (1,2) | 65.10% | (1,2) | 69.90% |
| | 4 | (1,3) | 76.40% | (1,3) | 74.90% |
| | 5 | (2,3) | 75.20% | (2,4) | 76.20% |
| | 6 | (4,5) | 77.40% | (3,4) | 77.30% |
| | 7 | (4,6) | 83.60% | (4,6) | 83.80% |
| Individual + Household | 3 | (1,2) | 66.30% | (2,1) | 69.30% |
| | 4 | (1,3) | 76.10% | (1,3) | 74.60% |
| | 5 | (2,3) | 74.70% | (2,1) | 75.20% |
| | 6 | (1,4) | 78.20% | (2,5) | 77.70% |
| | 7 | (3,5) | 83.60% | (3,5) | 83.80% |
| Individual + Household + Neighborhood | 3 | (1,2) | 62.60% | (2,1) | 69.20% |
| | 4 | (1,3) | 75.60% | (3,1) | 72.50% |
| | 5 | (1,4) | 74.70% | (2,4) | 75.30% |
| | 6 | (4,5) | 79.20% | (5,1) | 78.30% |
| | 7 | (3,5) | 83.60% | (3,5) | 83.60% |

**Three Cluster Combination Accuracy Results – Neutral Network**

| Feature Set | Prediction | Best Day | Accuracy (%) | Best Day Combination | Accuracy (%) |
|---|---|---|---|---|---|
| Individual | 4 | (1, 2, 3) | 63.10% | (1,3) | 74.90% |
| | 5 | (1, 2, 4) | 74.90% | (2,4) | 76.20% |
| | 6 | (1, 4, 5) | 77.00% | (3,4) | 77.30% |
| | 7 | (1, 4, 6) | 83.80% | (4,6) | 83.80% |
| Individual + Household | 4 | (1, 2, 3) | 61.80% | (1,3) | 74.60% |
| | 5 | (1, 2, 4) | 74.90% | (2,1) | 75.20% |
| | 6 | (1, 4, 5) | 76.10% | (2,5) | 77.70% |
| | 7 | (1, 2, 6) | 83.80% | (3,5) | 83.80% |
| Individual + Household + Neighborhood | 4 | (1, 2, 3) | 62.30% | (3,1) | 72.50% |
| | 5 | (1, 2, 4) | 72.90% | (2,4) | 75.30% |
| | 6 | (1, 4, 5) | 76.70% | (5,1) | 78.30% |
| | 7 | (1, 5, 6) | 83.80% | (3,5) | 83.60% |