UNIVERSITY OF TEXAS ARLINGTON

TxDOT Report 0-7184-01

Develop an Interactive Unit Price Estimation and Visualization Tool

Mohsen Shahandashti
Mahmut Yasar
Anita Bhattarai
Santosh Acharya
Mahnaz Paydarzarnaghi
Kayvon Khodahemmati
Sushil Bhatta
Abhijit Roy

Report Publication Date:	Submitted: August 2025 Published: September 2025
Project:	0-7184
Project Title:	Develop an Interactive Unit Price Estimation and Visualization Tool

Technical Report Documentation Page

1. Report No. FHWA/TX-26/0-7184-1	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Develop an Interactive Unit Price Estimation and Visualization Tool: Final Report		5. Report Date July 2025 6. Performing Organization Code
7. Author(s) Mohsen Shahandashti, Mahmut Yasar, Anita Bhattarai, Santosh Acharya, Abhijit Roy, Sushil Bhatta, Mahnaz Paydarzarnaghi, Kayvon Khodahemmati		8. Performing Organization Report No. 0-7184-7
9. Performing Organization Name and Address The University of Texas at Arlington Department of Civil Engineering P.O. Box 19308 Arlington, TX 76019		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 0-7184
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Division P.O. Box 5080 Austin, TX 78763-5080		13. Type of Report and Period Covered Technical Report August 2023–August 2025 14. Sponsoring Agency Code

15. Supplementary Notes

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.

16. Abstract

Texas Department of Transportation (TxDOT) determines unit prices of pay items using the historical bids-based estimation method and then develops an engineer's project appraisal. The engineer's estimate is used to assess the bids and select the bidder. However, the unit price of a work item is heavily affected by various project-specific and external factors, including but not limited to the project location, the quantity of the work, project complexity, time factors, and macroeconomic conditions. Therefore, accurate and reliable unit price estimation based on these project-specific and external factors is vital for the optimum use of the available project budget. The project objectives included: (1) conducting an overview analysis of factors affecting unit prices, (2) identifying factors affecting unit prices in Texas, (3) creating a unit price estimation database, (4) creating a spatio-temporal unit price estimation model considering the factors affecting unit prices, (5) developing a GIS-based visualization tool, and (6) implementing, demonstrating, and validating the interactive unit price estimation and GIS-based visualization tool on six Receiving Agency's projects.

The models that are developed based on the identified factors affecting unit prices help enhance accurate and reliable unit price estimation. Moreover, the developed GIS-based unit price visualization tool can be used for a quick retrieval of unit price values across various geographical locations. The tool helps track changes in data over time, by county, and across projects, as well as changes in the quantity of work items.

17. Key Words Unit price estimation, GIS-based tool, visualization	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161; www.ntis.gov.

19. Security Classif. (of report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	113	

DEVELOP AN INTERACTIVE UNIT PRICE ESTIMATION AND VISUALIZATION TOOL

FINAL REPORT

Prepared by:

Mohsen Shahandashti, Ph.D., P.E.

Mahmut Yasar, Ph.D.

Anita Bhattarai

Santosh Acharya

Abhijit Roy, Ph.D.

Sushil Bhatta

Mahnaz Paydarzarnaghi, Ph.D.

Kayvon Khodahemmati

The University of Texas at Arlington

Prepared for:

Research and Technology Implementation Division (RTI)

Texas Department of Transportation (TxDOT)

Project 0-7184

Report 0-7184-1

Project Title: Develop an Interactive Unit Price Estimation and Visualization Tool

July 2025

UNIVERSITY OF TEXAS ARLINGTON | Department of Civil Engineering

425 Nedderman Dr., 416 Yates St.

Box 19308, Arlington, TX 76019

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Texas Department of Transportation (TxDOT). This report does not constitute a standard, specification, or regulation.

This report is not intended for construction, bidding, or permit purposes.

The United States Government and the State of Texas do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

ACKNOWLEDGEMENT

This research project is supported by the Research and Technology Implementation (RTI) Division of the Texas Department of Transportation (TxDOT). We would like to thank the Project Monitoring Committee (PMC) members: Danny Souraphath, Andrea Ranft, Sam Kent, William Allen, Alonzo Balderaz, James Padilla, and Nehal Atlam for their coordination with the research team in ensuring the smooth execution of this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Texas Department of Transportation (TxDOT) determines unit prices of pay items using the historical bids-based estimation method and then develops an engineer's project appraisal. The engineer's estimate is used to assess the bids and select the bidder. However, the unit price of a work item is heavily affected by various project-specific and external factors, including but not limited to the project location, the quantity of the work, project complexity, time factors, and macroeconomic conditions. Therefore, accurate and reliable unit price estimation based on these project-specific and external factors is vital for the optimum use of the available project budget.

The project objectives included: (1) conducting an overview analysis of factors affecting unit prices, (2) identifying factors affecting unit prices in Texas, (3) creating a unit price estimation database, (4) creating a spatio-temporal unit price estimation model considering the factors affecting unit prices, (5) developing a GIS-based visualization tool, and (6) implementing, demonstrating, and validating the interactive unit price estimation and GIS-based visualization tool on six Receiving Agency's projects.

The factors that affect unit prices of construction line items were identified with an extensive literature review. Then, we collected data on various project-specific factors and external factors (i.e., independent variables) and the bidder's estimate and the TxDOT engineer's estimate (i.e., dependent variables). Using that dataset, spatiotemporal unit price estimation models were developed to predict the TxDOT engineer's unit price estimate and the bidder's average unit bid price. Finally, a GIS-based unit price estimation and visualization tool (UPEVT) was created, and we used the tool for estimating unit prices of a few line items in six Texas projects to demonstrate the tool's application. The UPEVT enables TxDOT personnel to visualize the historic unit price data. Implementing this system facilitates quick data retrieval and visualization across different geographic locations.

CHAPTER 1. INTRODUCTION

Texas is a large state with an expansive construction budget. In fiscal year 2024, TxDOT achieved a record milestone with $13.7 billion awarded for state highway improvement projects, and the agency plans to invest 104 billion dollars over the next 10 years from 2025 to 2034 (TxDOT 2025). Hence, accurate and reliable unit price estimation is vital for the optimum use of the available project budget. Moreover, a GIS-based tool is essential to facilitate quick retrieval and visualization of unit price data across different geographic locations.

The objectives of this project are: (1) conducting an overview analysis of factors affecting unit prices, (2) identifying factors affecting unit prices in Texas, (3) creating a unit price estimation database, (4) creating a spatio-temporal unit price estimation model considering the factors affecting unit prices, (5) developing a GIS-based visualization tool, and (6) implementing, demonstrating, and validating the interactive unit price estimation and GIS-based visualization tool on six Receiving Agency's projects.

This technical report explains all the tasks performed in the development of the Unit Price Estimation and Visualization Tool (UPEVT). The report is organized as follows:

- **Chapter 1** is this introductory chapter.
- **Chapter 2** explains factors affecting unit prices and the state DOT's unit price estimation methods.
- **Chapter 3** explains the data collection of historical unit prices and factors that potentially affect unit prices.
- **Chapter 4** explains machine learning model development for unit price estimation.
- **Chapter 5** explains mixed-effects model development for unit price estimation.
- **Chapter 6** describes the creation of the Unit Price Estimation and Visualization Tool (UPEVT).
- **Chapter 7** explains the implementation of the developed GIS-based visualization tool for Texas projects to validate the framework's performance.
- **Chapter 8** explains the process of handing over the developed GIS-based visualization tool and the models to TxDOT.

CHAPTER 2. OVERVIEW ANALYSIS OF FACTORS AFFECTING UNIT PRICES AND THE STATE DOT'S UNIT PRICE ESTIMATION METHODS

2.1. FACTORS AFFECTING UNIT PRICES

The unit price of a work item is heavily affected by various project-specific and external factors, including but not limited to the project location, the quantity of the line item, time factors, and macroeconomic conditions. **Figure 1** shows the gaps between the unit prices estimated by TxDOT engineers (Workbook: Bid Tabulations 2025) and the actual unit prices reported by Engineering News Record (ENR 2023) of the Portland cement line item in Dallas from January 2022 to December 2023. It should be noted that the TxDOT Engineer's Estimate of unit prices is determined by calculating the average of different unit prices of Portland cement (bid item 275-6001) on projects let in the same month in Dallas.

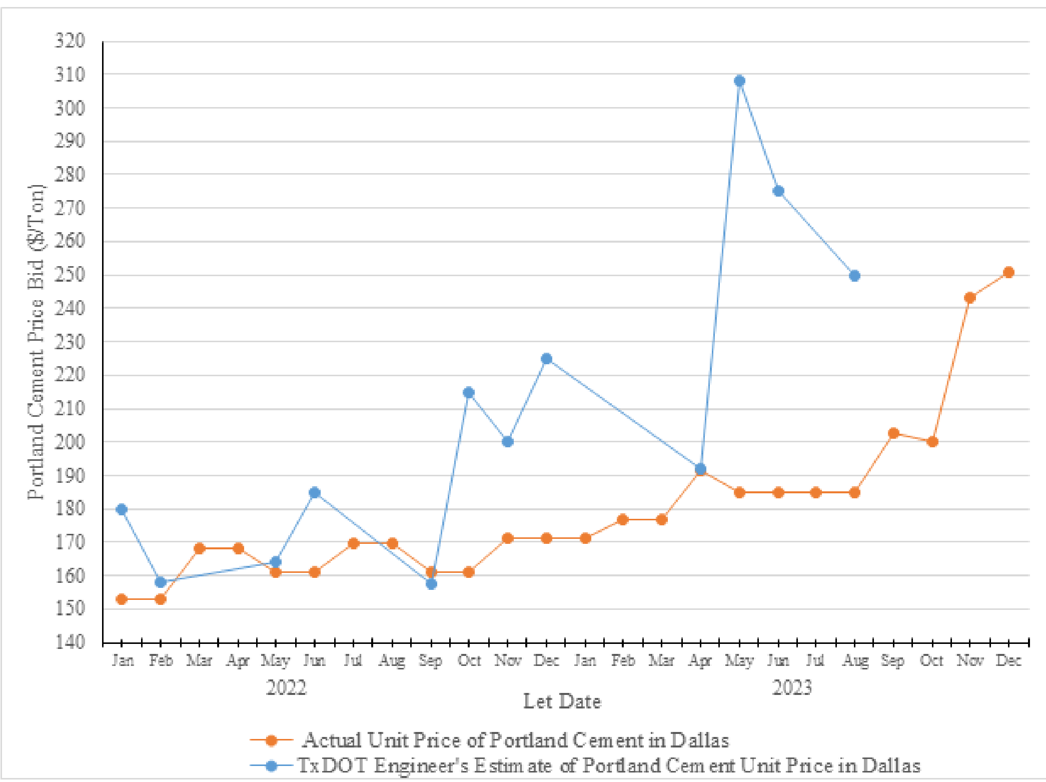


Figure 1 Gaps between the TxDOT Engineer's Estimate average unit price and actual unit price for the Portland cement line item in Dallas

The unit prices in the TxDOT Engineer's Estimate vary significantly from actual unit prices, which may result in a highly biased and inconsistent project cost estimation. Previous studies show that project-specific and external factors affect the unit prices of work items in highway projects. For example, unexpected risks such as the 2021 Texas Winter Storm inflated the unit price of Dallas pipe material items by up to 10 percent (Kim and Shahandashti 2022). Highway construction costs are found to be highly related to macroeconomic factors such as crude oil prices (Shahandashti and Ashuri 2016). Moreover, as the quantity of goods or services requested in a bid increases, suppliers or contractors may offer lower bid unit prices because larger quantities often lead to economies of scale and volume discounts (Baek and Ashuri 2019). Construction market conditions have a significant impact on a highway construction cost forecast (Mahadavian et al. 2021). Different project-specific and external factors were considered to estimate unit prices in highway projects, as summarized in **Table 1**.

Table 1 Studies on cost estimation in highway projects

Research	Project location	Project type	Project complexity	Project size	Project duration	Bid quantities	Number of bidders	Temporal factors	Macroeconomic conditions	Regional conditions	Unexpected risks
Shrestha et. al (2025)	✓	✓	-	✓	-	-	-	✓	-	-	-
Dong et. al (2025)	-	-	-	-	-	-	-	✓	-	-	✓
Meng et. al. (2024)	✓	✓	-	✓	-	-	-	✓	-	-	-
Kim and Shahandashti (2023)	-	-	-	-	-	-	-	-	-	-	✓
Awuku et al. (2022)	✓	-	-	-	-	-	-	-	-	-	-

Research	Project location	Project type	Project complexity	Project size	Project duration	Bid quantities	Number of bidders	Temporal factors	Macroeconomic conditions	Regional conditions	Unexpected risks
Kim and Shahandashti (2022)	✓	-	-	-	-	-	-	✓	-	-	✓
Meharie et al. (2022)	-	✓	-	✓	-	-	-	-	✓	-	-
Wang et al. (2022)	✓	✓	-	-	✓	-	-	-	✓	-	-
Mohamed and Moselhi (2022)	✓	✓	✓	-	-	-	-	-	-	-	-
Mahadavian et al. (2021)	-	-	-	-	-	-	-	✓	✓	-	-
Karaca et al (2020)	-	✓	✓	✓	-	-	-	-	-	-	-
Baek and Ashuri (2019)	✓	-	✓	✓	-	✓	✓	-	✓	✓	-
Meharie et al. (2019)	✓	✓	✓	✓	✓	-	-	-	✓	✓	-
Le et al. (2019)	✓	-	-	-	-	✓	-	-	-	-	-
Shrestha and Jeong (2019)	✓	-	-	-	-	-	-	-	-	-	-
Baek and Ashuri (2018a)	✓	-	-	-	-	✓	✓	-	✓	✓	-
Baek and Ashuri (2018b)	✓	-	-	-	-	✓	-	-	✓	✓	✓
Cao et al.(2018)	✓	-	-	✓	✓	✓	✓	✓	✓	✓	-

Research	Project location	Project type	Project complexity	Project size	Project duration	Bid quantities	Number of bidders	Temporal factors	Macroeconomic conditions	Regional conditions	Unexpected risks
Rafiei and Adeli (2018)	✓	-	✓	✓	✓	-	-	-	✓	✓	-
Bhargava et al. (2017)	✓	✓	-	-	-	-	-	-	-	-	-
Gardner et al. (2017)	-	✓	✓	-	-	-	-	-	-	✓	-
Zhang et al. (2017)	-	-	-	-	-	-	✓	-	✓	✓	-
Swei et al. (2017)	-	-	-	-	-	✓	✓	-	-	-	-
Hannan et al. (2016)	-	-	-	-	-	✓	-	-	-	-	✓
Hyari et al. (2016)	✓	-	-	-	-	✓	✓	-	-	-	-
Shahandashti and Ashuri (2016)	-	-	-	-	-	-	-	✓	✓	-	-
Cirilovic et al. (2014)	-	-	-	-	-	-	✓	-	✓	✓	-
Cheng (2014)	✓	✓	✓	✓	-	-	-	-	✓	✓	✓
Hegazy and Ayed (1998)	✓	✓	-	✓	✓	-	-	✓	-	✓	-

The factors affecting unit prices of work items are classified into two groups: project-specific and external factors (Baek and Ashuri 2019). **Figure 2** shows the different project-specific and external factors that can affect the unit prices of a work item.

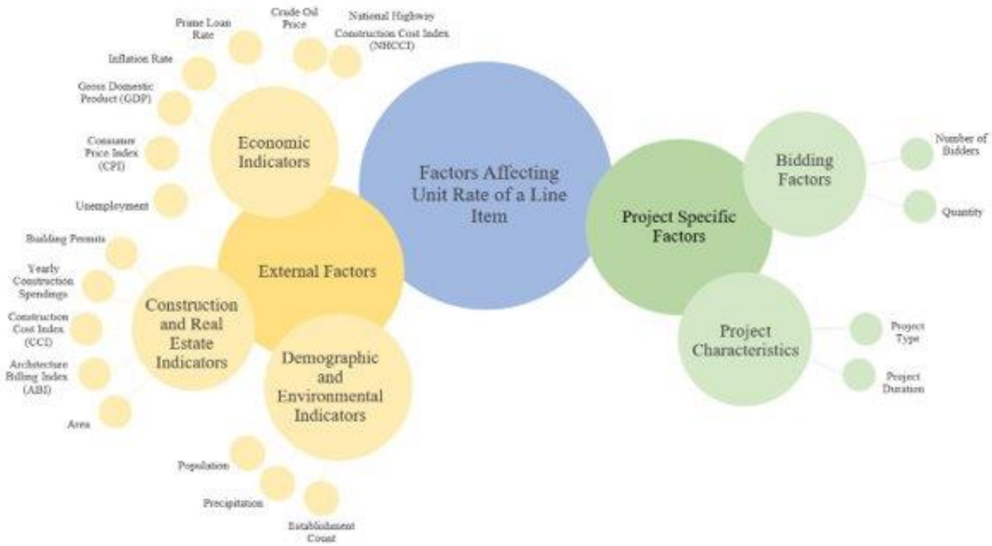


Figure 2 Potential project-specific and external factors affecting the unit price of a work item

2.1.1. PROJECT-SPECIFIC FACTORS

Estimating unit prices for highway projects involves considering a range of project-specific factors that can impact the highway project cost. Project-specific factors include project features, time factors, site conditions, bid conditions, and legal regulations.

Project Features

Project-specific characteristics, such as project type, duration, complexity, and size, are significantly correlated with construction costs (Baek and Ashuri 2019). The complexity of the project design and scope can impact unit prices as more complex projects may require specialized equipment and skilled labor, which can increase unit prices. The complex engineering and logistical challenges, such as excavating a deep tunnel beneath the city in the Alaskan Way Viaduct replacement tunnel project in Seattle, Washington, contributed to the higher unit prices, requiring specialized and custom-built tunneling equipment (Riddle and Whittington 2022). Also, projects requiring extended periods of construction will reflect higher bid prices because suppliers do not normally guarantee the same prices for extended periods of time, and the contractor(s) will usually hedge their bid prices for protection against any increase in unit prices (Kim et al. 2022; Thomason 2017). Moreover, projects at places where equipment and construction materials are not easily accessible are likely to have more construction costs than projects in well-facilitated areas (Baek and Ashuri 2019; Wang et al. 2022).

Time Factors

Project time factors, such as month, season, and year, are closely related to unit prices as the unit prices of a work item fluctuate over time with a trend and seasonality (Kim et al. 2020, Shahandashti and Ashuri 2013). The time of year a project is to be let for contract and the estimated time required for completion can significantly impact prices, as seasonal or weather factors (e.g., inclement weather that may necessitate project suspension or delay) can affect unit bid prices (Thomason 2017). The project timeline and work schedule can also impact unit prices, as accelerated timelines or work schedules may require additional resources or overtime pay, leading to unit price inflation (Levy 2018).

Site Conditions

Project site factors, such as location, distance to manufacturing plants, lack of access to the job site, and source of materials, can cause significant differences in unit prices for highway pavement construction projects (Cao et al. 2018). The project location can significantly impact the unit price of a work item, as it affects transportation costs, availability of materials and labor, and environmental factors (Ahmed and Arocho 2021). Site conditions, such as soil type, terrain, and geology, can impact unit prices, as they can affect the ease and cost of construction (Meharie et al. 2022). Work that is normally easy to accomplish on level terrain or gentle slopes may be almost impossible on steep slopes (Thomason 2017). Sites that require implementation measures to mitigate adverse environmental impacts, safeguard species, or protect archaeological structures may have increased unit costs.

Bid Conditions

Bid conditions, such as the number of bidders and bid quantities, can impact unit prices for highway construction projects (Mahadavian et al. 2021). These bid conditions can affect the competitive landscape, the bidding process, and the final price that is agreed upon. For example, when there are more bidders, competition is typically higher, which can drive down prices. Conversely, when there are fewer bidders, the competition is lower, which can lead to higher prices. The quantity of bid items shows a significant correlation with unit prices (Baek and Ashuri

2019). When the bid quantity is large, it can provide economies of scale, allowing contractors to achieve cost savings through bulk purchasing or more efficient use of resources. This can lead to lower unit prices. Bid qualifications and requirements can also impact unit prices. When a project has strict requirements or requires specialized skills, the number of qualified bidders may be limited. This can result in higher unit prices due to less competition among the bidders.

Legal Regulations

Project-specific legal regulations can also affect unit prices, as complying with these requirements can require additional resources and expenses. For example, changes in earthquake safety specifications right after the 1994 Northridge earthquake required design and material changes, increasing the unit prices and resulting in significant cost overruns for ongoing projects (Danisworo and Latief 2019; Raetz et al. 2020). Also, the type of contract used for the project can impact unit prices, as different types of contracts have different risk profiles and cost structures (Awuku et al. 2022).

2.1.2. EXTERNAL FACTORS

External factors include macroeconomic conditions, regional construction market conditions, regional economic conditions, unexpected external risks, and national highway construction cost variations represented by the national highway construction cost index. Regional construction market conditions, such as construction demand and supply, can determine the unit price of work items for highway construction projects (Cao et al. 2018). Also, overall changes in national highway construction costs can impact the unit price of work items in highway projects (Shahandashti and Ashuri 2016).

Macroeconomic Conditions

Macroeconomic conditions can significantly impact unit prices of work items since they influence the cost and availability of construction labor, materials, and equipment (Kim et al. 2024; Ashuri et al. 2012). For example, the cost of steel during the Interstate 4 Ultimate project in Orlando, Florida, increased by over 40% due to high inflation rates in the construction industry between 2015 and 2019, which led to higher unit prices for steel-related work items such as reinforcing bars and bridge components (FDOT 2023). Also, inflation can lead to higher interest rates, which can increase the cost of financing, leading to unit price inflation (Musarat et al. 2021). The energy costs can affect the unit prices of work items (Ashworth and Perera 2015). Government policies and regulations can also impact the overall unit price structure of the industry. The contribution of the inflation rate to predict in the total costs of highway construction projects is approximately 45% of all the estimation parameters taken into account (Meharie et al. 2022).

Regional Construction Market

Regional construction market factors could also have a significant impact on unit prices for highway construction projects since they affect the availability and cost of labor, materials, and equipment within a specific geographic area. For example, the unit prices for construction materials and labor vary over regions due to different material and labor market conditions, including the supply of materials and labor, transportation and logistics costs, prevailing wages, and unionization rates (Kim et al. 2022). The level of regional construction demand also affects unit prices for highway construction projects. Markets with high construction demand have higher unit prices (Ahmadi and Shahandashti 2018).

Regional Economy

Local economic conditions, such as regional GDP and economic growth, can impact unit prices for highway construction projects (Baek and Ashuri 2019; Cao et al. 2018). Regions with strong economic growth tend to have higher unit prices due to increased resource competition (Shiha et al. 2020). Regional budget expenditure in transportation projects also impacted the unit prices of pavement materials (Kim et al. 2020).

Unexpected Risks

Unexpected risks, such as disasters, supply chain disruption, and pandemics, can have a significant impact on unit prices for highway construction projects (Khodahemmati and Shahandashti 2020; Ahmadi and Shahandashti 2020). Disasters such as hurricanes and tornadoes have damaged or destroyed critical infrastructure, increased demand for reconstruction, and inflated labor wages in the US highways, roads, and bridges sector by 20% (Pradhan and Arneson 2021). Hurricanes Katrina and Rita caused bid prices for superpave asphaltic concrete line items for highway projects in Louisiana to be significantly higher than their bid prices before the hurricanes (Baek and Ashuri

2018b). Also, the global pandemic significantly influenced pavement material unit prices and project costs in California (Kim et al. 2020).

National Highway Construction Cost Variations

The National Highway Construction Cost Index (NHCCI) measures the costs of materials and labor used in highway construction projects (Shahandashti and Ashuri 2015). It is used by the Federal Highway Administration (FHWA) to adjust the costs of highway construction projects for inflation and other cost factors. As such, the NHCCI is highly correlated with the unit prices of work items for highway projects (Shahandashti and Ashuri 2016).

CHAPTER 3. DATA COLLECTION OF HISTORICAL BID PRICES AND FACTORS AFFECTING UNIT PRICES

3.1. INTRODUCTION

This chapter provides a concise overview of data collection procedures and variable sources. The relevance of the data to be collected is established in the previous chapter (Chapter 2) through a review of studies.

3.2. DATA SOURCES

A dataset is created by collecting data on dependent and independent variables from publicly available sources. **Tables 2 and 3** display the list of dependent and independent variables, respectively, with their sources and the granularity of space and time in which they are collected.

Table 2 Data sources for dependent variables

Dependent Variable	Definition	Link to Data Source	Data level
TxDOT Engineer's Estimate	Preliminary estimated value of an item by TxDOT's engineer in a construction project that serves as a baseline for budgeting purposes.	https://tableau.txdot.gov/views/BidTabulationstxdot_gov	Item level

Table 3 Data sources for independent variables

Independent Variable	Definition	Link of Data Source	Data level
Building Permit	The total dollar value of new privately owned residential construction	https://www.census.gov/	county level, monthly
Gross Domestic Product	The inflation-adjusted value of the goods and services produced by labor and property	https://apps.bea.gov/itable/	county level, monthly
Establishment Counts	The total number of private construction establishments	https://data.bls.gov/maps/	county level, quarterly

Independent Variable	Definition	Link of Data Source	Data level
Unemployment Rate	A measure of the percentage of unemployed people	https://data.bls.gov/lausmap/	county level, monthly
Precipitation	Any form of water, whether liquid or solid, that falls from the atmosphere and reaches the ground, including rain, snow, sleet, hail, and drizzle.	https://www.ncei.noaa.gov/	county level, monthly
Population	The total number of people or inhabitant in a county.	https://www.census.gov/data/datasets/	county level, yearly
Area of a County	The total land and water surface of the county	https://data.census.gov/profile?g=040XX00US48	county level
Inflation	The overall general upward price movement of goods and services in the US economy	https://www.bls.gov/charts/	national level, monthly
Architecture Billings Index (ABI)	An economic indicator for nonresidential construction activities	https://www.aia.org/partner-aia	national level, monthly
Prime loan rate	The lowest interest rate at which money can be borrowed commercially	https://fred.stlouisfed.org/series/MPRIME	national level, monthly
Consumer Price Index (CPI)	Index measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services	https://www.ssa.gov/oact/STATS/cpiw.html	national level, monthly
Crude Oil Price of West Texas Intermediate (WTI)	The price of West Texas Intermediate (WTI) crude oil, which serves as one of the main benchmarks for oil pricing	https://www.eia.gov/	national level, monthly

Independent Variable	Definition	Link of Data Source	Data level
Construction Spending	The total amount of money spent on construction related activities	https://www.census.gov/econ/	national level, monthly
Construction Cost Index (CCI)	An overview of construction cost trends across 20 designated cities in the USA	https://www.enr.com/Cost-Data-Dashboard	national level, monthly
Project duration	Estimated number of working days duration of a project	https://tableau.txdot.gov/views/BidTabulations/	project level
Length of the Project	The total linear distance that a project covers along its main alignment	https://tableau.txdot.gov/views/BidTabulations/	project level
The number of items in a project	The number of line items in a project	https://tableau.txdot.gov/views/BidTabulations/	project level
Quantity	The quantity/amount of a bid item that is required for the project	https://tableau.txdot.gov/views/BidTabulations/	Item level
Number of Bidders	The number of bidders who participated in the bid	https://tableau.txdot.gov/views/BidTabulations/	project level

A detailed explanation of the process of data collection from data sources is provided in **Appendix A**.

CHAPTER 4. MACHINE LEARNING MODEL DEVELOPMENT

4.1. INTRODUCTION

This chapter outlines the development, training, and deployment of machine learning models designed to estimate the engineer's unit price of highway construction line items for the Texas Department of Transportation (TxDOT). This chapter explains the critical steps in developing the machine learning forecasting framework. The objective of the research was to design data-driven, reliable, and reproducible estimation systems to assist TxDOT engineers and planners in estimating future construction unit prices.

Machine learning (ML) algorithms are among the most advanced alternatives for making cost predictions (Cao et al. 2018). Machine learning models have become a viable and robust alternative to traditional statistical approaches for exploring nonparametric and nonlinear relationships (Shahandashti et al. 2023). ML methods are capable of learning from input data and generating predictions based on that information. Hegazy and Ayed (1998) utilized a neural network to create a parametric cost estimating model for 18 highway projects. To estimate the initial costs of highway construction projects, Al-Tabtabai et al. (1999) employed a neural network-based model. According to Emsley et al. (2002), neural networks outperformed linear regression models by better handling the nonlinear characteristics of data in construction cost prediction. Wilmot and Cheng (2003) employed an artificial neural network model utilizing subitem cost information to project the Louisiana Highway Construction Cost Index. Wilmot and Mei (2005) developed an artificial neural network model to replicate past trends in highway construction costs in Louisiana. Petroutsatou et al. (2012) employed neural network techniques to develop models that forecast initial cost estimates for both road and tunnel construction projects. Gardner et al. (2016) proposed cost estimation models for highway projects by applying both artificial neural networks and multiple regression methods. Ashuri et al. (2018) developed predictive models to estimate the costs of the lump sum pay items for Traffic Control and Grading Complete. Previous research offers a promising foundation for enhancing the accuracy of construction cost estimation in highway projects using machine learning models.

In this project, two machine learning models (Deep Neural Network models and Ensemble models) were employed to predict unit prices of construction line items for highway projects in Texas. Both models leveraged a wide range of input features, including project-specific and external variables, encompassing both numerical and categorical types. Deep Neural Network (DNN) models and Ensemble models were developed using a structured pipeline that combines feature engineering, categorical embeddings, dimensionality reduction using Principal Component Analysis (PCA), hyperparameter tuning, and model validation. Finally, the developed models were used to provide an estimated unit price for TxDOT engineers for construction line items.

Figure 3 illustrates a typical machine learning workflow, starting with raw data processing, followed by iterative data preprocessing until the data is ready for modelling. Once prepared, machine learning algorithms are applied, and models are evaluated to identify the best-performing models. The model is then deployed for practical application.

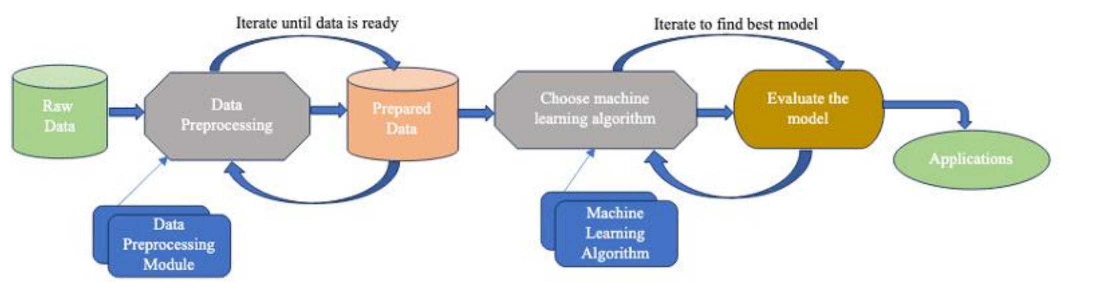


Figure 3 Typical machine learning modeling process

4.1.1 DEEP NEURAL NETWORK MODELS

Deep learning neural networks have gained substantial attention in recent years due to their capacity to model complex patterns in data (Ortiz-Garcia et al. 2016) and capture spatiotemporal relationships among variables (Hussain et al. 2022). Goodfellow et al. (2016) characterize deep learning as an advanced form of machine learning that builds knowledge hierarchically, where higher-level abstractions are constructed from layers of simpler, lower-level features. This layered representation enables the learning of highly complex functions (Hosein and Hosein 2017).

In the context of highway construction unit price estimation, Deep Neural Networks (DNNs) offer a powerful and flexible modeling approach. Unlike traditional models that often rely on expertdesigned rules or handcrafted features, DNNs excel due to their ability to automatically identify and learn meaningful patterns from raw input data. This capability of DNN allows them to uncover complex structures in the data by leveraging large datasets, leading to more accurate and generalizable models (Sze et al. 2020). These deep neural networks are composed of multiple interconnected layers that extract, learn, and interpret hierarchical representations of the input, enabling them to handle diverse and high-dimensional datasets effectively (Chauhan and Singh, 2018). Moreover, DNNs can model nonlinear relationships that are difficult to capture with conventional linear approaches, particularly when dealing with spatiotemporal variations in highway construction unit price estimation data. By integrating both spatial inputs (e.g., county-level geographic information) and temporal factors (e.g., inflation rates, construction cost index trends), DNNs can enhance the accuracy, depth, and robustness of cost predictions, making them a valuable tool for data-driven unit price estimation.

Figure 4 illustrates the general architecture of a deep neural network, where a set of k input features is processed through multiple hidden layers to estimate the unit prices of a work item as an output.

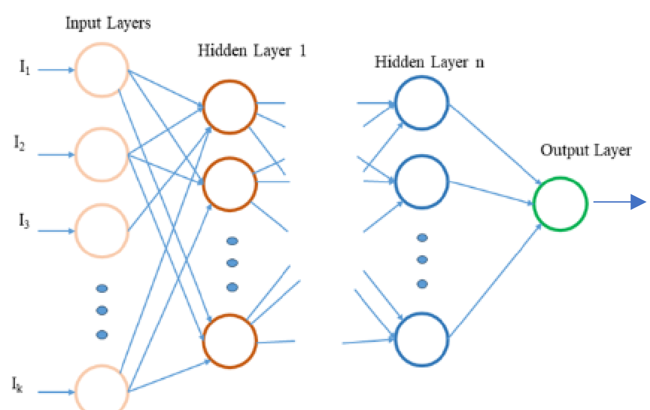


Figure 4 Architecture of a deep neural network model

4.1.2 ENSEMBLE MODEL

Ensemble-based machine learning methods are gaining traction as advanced tools for construction cost estimation (Meharie et al. 2022). While prior studies have explored these methods at a broader project level scope, their application specifically to unit price estimation remains limited, presenting a key area where this study contributes. Ensemble methods are machine learning techniques that combine outputs of multiple base learners and make predictions on new data by combining their outputs, typically through a weighted voting approach (Dietterich 2000). Ensemble machine learning models combine predictions from several base models to produce results that are more accurate and reliable than those from any single model (Xiao et al. 2018; Han et al. 2020). Chou and Lin (2012) predicted the likelihood of disputes in public–private partnership projects using early-stage project parameters. According to their findings, the ensemble approach that merged an artificial neural network (ANN), a support vector machine (SVM), and a decision tree provided the most precise predictions, with an accuracy close to 84%. Williams and Gong (2014) utilized an ensemble staking model combining K-Star, Ridor, and Radial Basis Function (RBF) neural networks to predict project cost overruns using contract documents, achieving the best performance among tested models despite a relatively low accuracy of around 44%. Cao et al. (2018) combined two layers of prediction models to predict unit price bids of resurfacing highway projects. The first layer made predictions from three machine learning algorithms (extreme gradient boosting, random forest, and gradient boosting), and the second layer was made of a neural network to make the final predictions. The developed ensemble model outperformed multiple regression and Monte Carlo simulations.

In this research, an ensemble model was developed to combine the strengths of Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) at Layer 1 and a Neural Network at Layer 2 to create models that involve both temporal (sequence) and spatial (pattern) information. LSTM extracts the temporal dependencies in the dataset, and CNN captures the spatial features. In this context, the ensemble models combine spatial information from different counties, project and construction line-item types, with temporal trends leading to a robust spatiotemporal model.

Figure 5 illustrates the architecture of an ensemble model that is composed of an LSTM and CNN in the first layer and a neural network in the second layer. The model processes spatiotemporal data as input, starting with Layer 1. The output from the LSTM and CNN (first layer) is fed into the neural network (second layer) to make the final prediction. The model's final output represents the estimated engineer's unit price, reflecting the influence of both spatial and temporal factors within the input data.

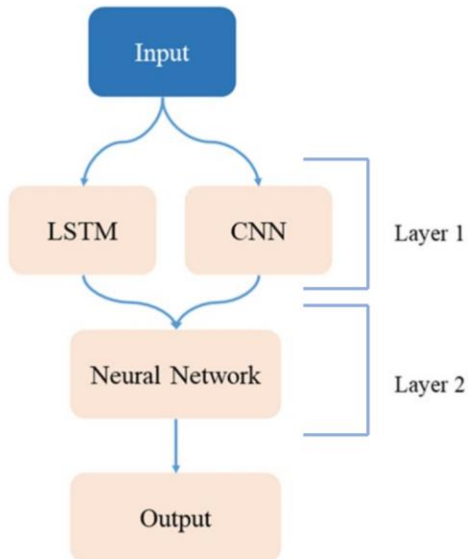


Figure 5 Architecture of an ensemble model

4.2. DATASET FOR MACHINE LEARNING MODELING

The dataset consisted of engineers' unit price estimates over 2 years, from March 4, 2022, to September 24, 2024. The dataset contains 13,398 unique observations, covering 115 distinct construction line-item types, 299 different construction project types, and 216 counties across Texas. However, the data coverage is uneven. First, not all of Texas's 254 counties are represented, indicating incomplete geographic coverage. Second, even among the included counties, not all types of construction projects and construction line-item types are represented consistently over the two years. This results in an imbalanced dataset with gaps in construction project and line-item type representation across counties and time.

4.3. MODEL DEVELOPMENT

This section describes a systematic process for developing machine learning models for eight line items for the Texas highway projects, using both DNN models and Ensemble models. **Figure 6** depicts the overall methodology adopted for developing the machine learning models.

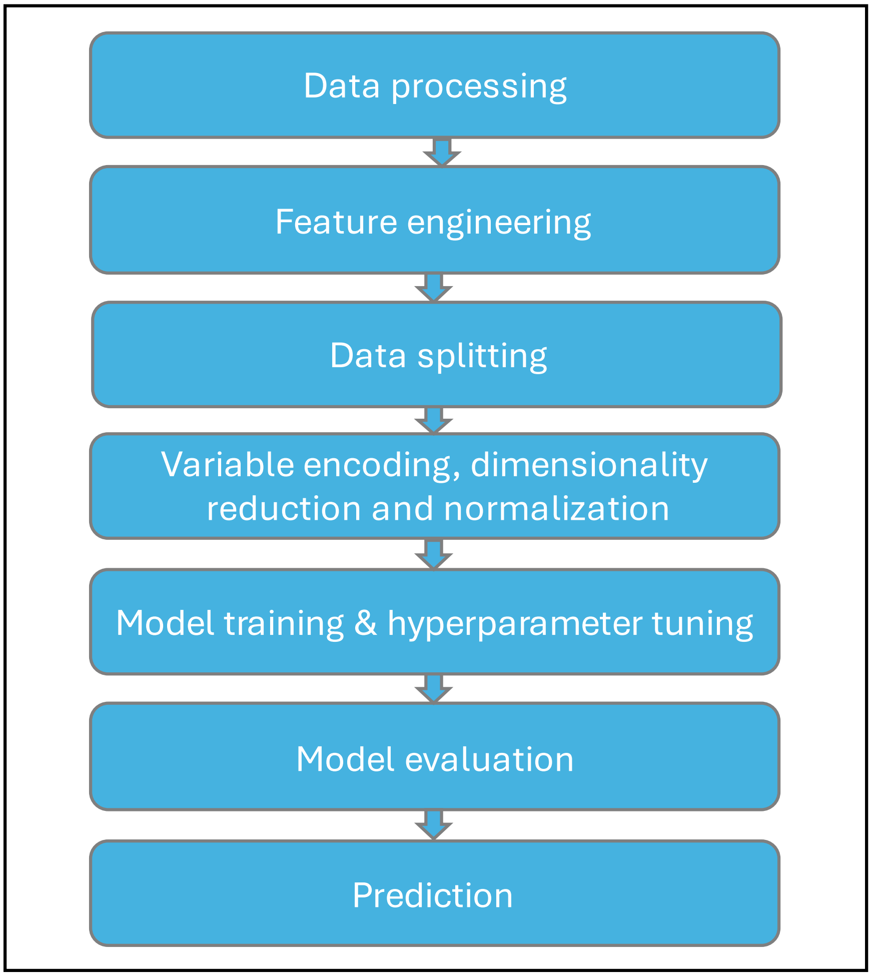


Figure 6 Overview of the process for developing machine learning models

The subsection below describes the step-by-step pipeline for developing the two machine learning models. It begins with data processing, followed by feature engineering to extract meaningful variables. The data is then split into training and testing sets and subjected to variable encoding, dimensionality reduction, and normalization to prepare it for modeling. Next, the model undergoes training and hyperparameter tuning to optimize performance and accuracy. The trained model is then evaluated using evaluation metrics and finally used for prediction on new or unseen data.

4.3.1 DATA PROCESSING

Poor data quality can significantly compromise the performance of forecasting model development. Therefore, the modeling process began with a rigorous data preparation step, using a curated dataset of estimated records for the top eight highway construction line items. These top eight items were selected based on their frequency across all counties over the two years. The selected line items were: "Excavation", "Backfill", "Asphalt", "Barricades, Signs and Traffic Handling", "Truck-Mounted Attenuator (TMA)", "Mobilization", "Portable Changeable Message Sign", and "Vegetative Watering".

This dataset incorporated both project-level bid information (such as engineer's estimated unit price, estimated duration for the project, bid date), bidder's information's relevant to engineer's unit price estimate (such as number of bidders for the project, average of the lowest three bidders), and external factors (such as NHCCI, CCI, CPI, GDP). The average of the lowest three bidders was included in the machine learning models in addition to the variables described in Chapter 2. To ensure unbiased estimation of unit prices, the total project cost was excluded from the model as an external variable.

Table 4 depicts the project-specific factors used in developing the machine learning models. The table also indicates the spatial and temporal data levels associated with each variable.

Table 4 Project-specific factors used in the development of the machine learning models

Variables	Definition	Data level
Project duration	Estimated number of working days duration of a project.	Project level
Length of the Project	The total linear distance that a project covers along its main alignment.	Project level
The number of items in a project	The number of line items in a project.	Project level
Quantity	The quantity/amount of a bid item that is required for the project.	Item level
Number of Bidders	The number of bidders who participated in the bid.	Project level
Average of three lowest bidders	The average of the three lowest bidders is the arithmetic mean of the three smallest bid amounts submitted for the project.	Project level

Table 5 shows the external influencing factors used in the development of the machine learning models. The table indicates the data levels associated with each variable. Refer to **Chapter 2** for a detailed explanation of the variables used.

Table 5 External factors used in the development of the machine learning models

Variables	Definition	Data level
Building Permit	The total dollar value of new privately owned residential construction.	County level, monthly
Gross Domestic Product	The inflation-adjusted value of the goods and services produced by labor and property.	County level, monthly

Establishment Counts	The total number of private construction establishments.	County level, quarterly
Unemployment rate	A measure of the percentage of unemployed people.	County level, monthly
Precipitation	Any form of water, whether liquid or solid, that falls from the atmosphere and reaches the ground, including rain, snow, sleet, hail, and drizzle.	County level, monthly
Population	The total number of people or inhabitants in a county.	County level, yearly
Area of a county	The total land and water surface of the county.	County level
Inflation rate	The overall general upward price movement of goods and services in the US economy.	national level, monthly
Architecture Billings Index (ABI)	An economic indicator for nonresidential construction activities.	national level, monthly
Prime loan rate	The lowest interest rate at which money can be borrowed commercially.	national level, monthly
Consumer Price Index (CPI)	Index measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services.	national level, monthly
Crude Oil Price of West Texas Intermediate (WTI)	The price of West Texas Intermediate (WTI) crude oil, which serves as one of the main benchmarks for oil pricing.	national level, monthly
Construction Cost Index (CCI)	An overview of construction cost trends across 20 designated cities in the USA.	national level, monthly

4.3.2 DATA CLEANING AND STRUCTURING

The dataset was filtered to isolate line-item-specific records. Irrelevant entries were removed from the dataset. Another prominent step in data cleaning is handling the outliers. Outliers refer to the data points that significantly deviate from most of the data points. The presence of outliers can have a detrimental impact on the effectiveness and accuracy of the models, and it is often required to detect and handle the outliers to improve the performance of the model. The handling of the outliers must be done before training the machine learning models.

To improve the model's performance, stabilize variance, and handle extreme outliers, the dependent variable (the engineer's unit price estimate) underwent key preprocessing steps:

Winsorization

To mitigate the influence of extreme outliers, the engineer's estimated unit prices were winsorized by capping the engineer's estimated unit price distribution. Winsorization is a technique used to handle outliers by replacing extreme values with the nearest value of the threshold (Nyitrai and Virag 2019). This method involves defining upper and lower threshold limits based on the percentile of the empirical distribution and substituting any values that are outside these limits with the nearest threshold value. During the development of the models, the range varied from the 1st to 5th percentile (as lower threshold) and the 95th to the 99th percentile (as upper threshold). This ensures that extremely low or high estimates do not disproportionately affect model training.

Log Transformation

A logarithmic transformation was applied to the winsorized engineer's estimates to reduce skewness and approximate a more normal distribution. This is particularly effective in dealing with right-skewed cost data and enhances model convergence and interpretability.

4.3.3 FEATURE EXTRACTION

To capture the trends in the dataset, four feature extraction steps were performed.

Temporal Features

To capture seasonality and trends over time, the following features were extracted:

- Calendar features: Year, month, and quarter from the bid date.
- Cyclic encodings: Sine and cosine transformations of month and quarter to model periodic behavior effectively.

Lagged and Rolling Statistics

To provide historical context of unit bid prices at the county level, the following features were derived from the past bid information:

- Lag features: Lagged engineer's estimates at 1, 3, 6, and 12 months.
- Rolling aggregates: Mean and standard deviation over 3- and 6-month windows.
- Trend estimation: Local temporal slope from linear regression over the prior 6 months.

Monthly Aggregation

Average unit prices were aggregated monthly across key dimensions to enrich temporal signals:

- Item code and month
- Project type and month
- County and month

Interaction Features

To capture complex cross-variable dependencies, interaction terms were created and encoded:

- Item type and project type
- Item type and county
- Item type and year
- Item type and month

A safe encoding approach was applied to prevent information leakage by ensuring encodings were learned only from training data. All these temporal features were used in addition to the project-specific factors and external factors during the machine learning models.

4.3.4. CORRELATION ANALYSIS AND DIMENSIONALITY REDUCTION

To manage the high-dimensional nature of the external dataset and reduce the risk of multicollinearity, a two-step process was followed: (1) correlation analysis to understand the relationships among variables, and (2) Principal Component Analysis (PCA) to transform the data into a lower-dimensional, uncorrelated feature space.

Correlation Analysis

The initial step involved computing the Pearson correlation coefficient between all pairs of external variables. This analysis helped identify variables that were strongly linearly related. A threshold of 0.8 was selected as a cutoff value to flag high correlations. Variables that exhibit such strong correlations can introduce multicollinearity, which may negatively impact model training by inflating variance and reducing interpretability.

Figure 7 shows a heatmap of the Pearson correlation coefficients across all external variables. The heatmap reveals multiple clusters of high correlation among external variables. Notably, the correlation between "Establishment Counts" and "GDP" exceeds 0.8, as does the correlation between "Construction Spending" and variables like "CPI" and "CCI." These correlations indicate that several variables are capturing overlapping information. However, rather than removing individual variables manually, Principal Component Analysis (PCA) was performed to reduce the number of variables automatically.

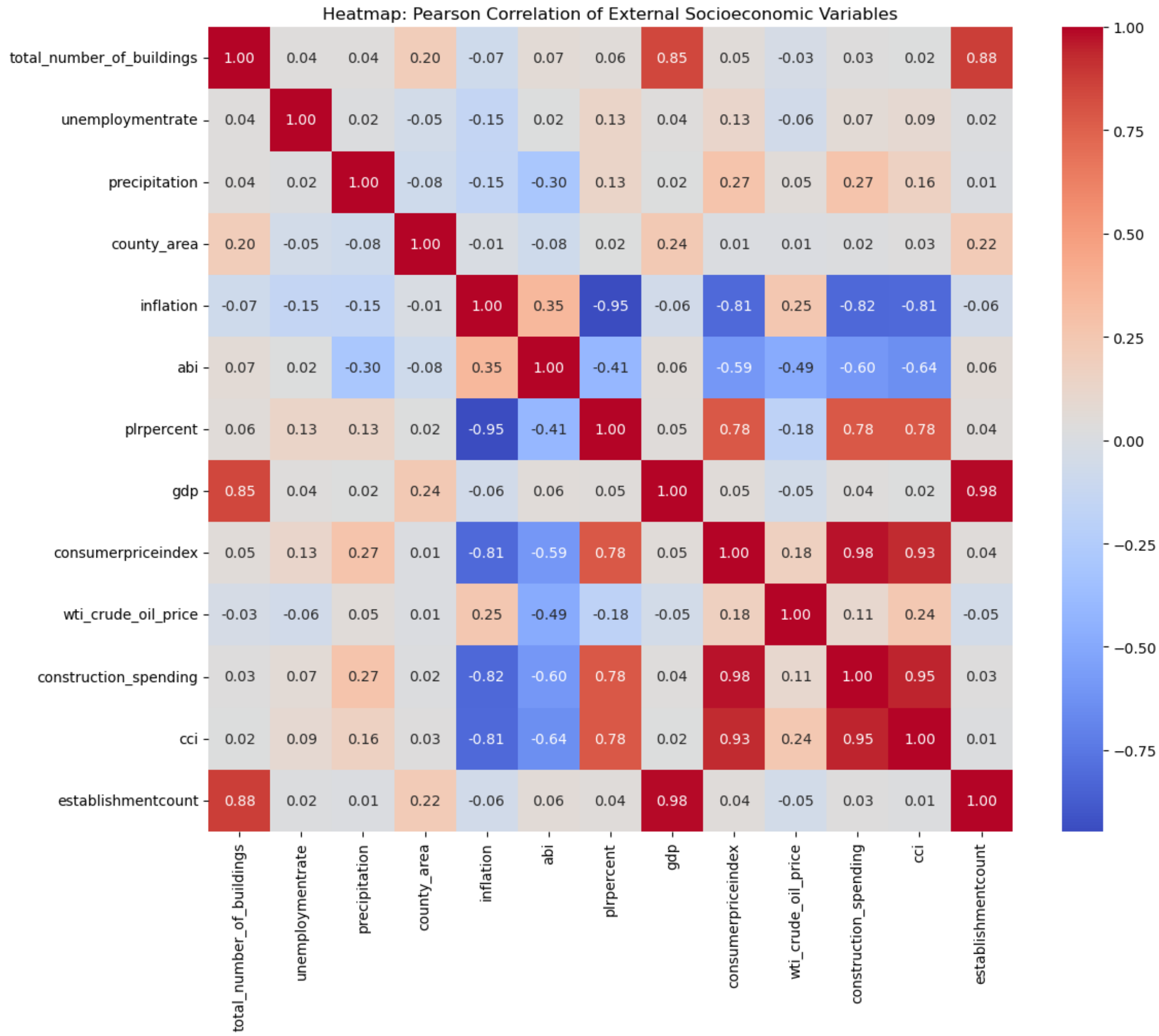


Figure 7 Heatmap of Pearson correlation coefficients among external variables

Principal Component Analysis (PCA)

To systematically address the observed multicollinearity and reduce the number of input features while retaining as much information as possible, Principal Component Analysis (PCA) was applied to the entire set of external variables. PCA is a linear transformation technique that converts a correlated set of variables into a smaller number of orthogonal (uncorrelated) principal components**.** Each component captures a portion of the total variance in the original dataset, allowing the model to focus on the most informative dimensions.

Figure 8 displays the variance explained by each principal component, shown here for the Excavation line item as an example. The first principal component accounts for approximately 38% of the total variance, and the second component adds over 20%, indicating that a significant portion of the information is concentrated in just a few components. To determine how many components to retain, we used a cumulative variance threshold of 98%**,** ensuring that the vast majority of the original data's variability was preserved. This typically resulted in a considerable reduction in dimensionality, simplifying the input space for subsequent modeling.

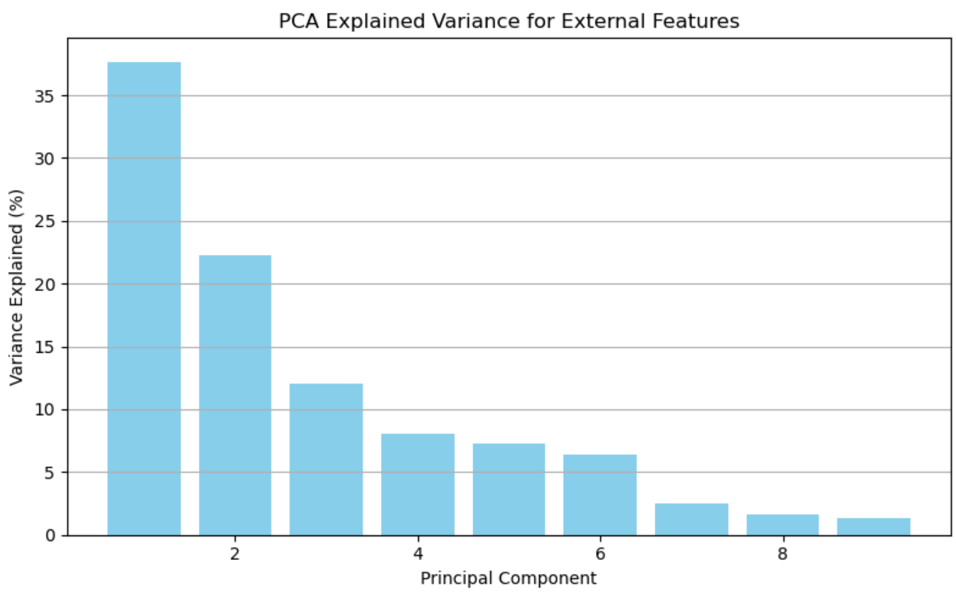


Figure 8 Variance explained by the principal components (Example representation for excavation line item)

By retaining the principal components that explain the majority of the variance, the model benefits from a more compact**,** informative**,** and noise-reduced representation of the input data. This also helps improve model performance, reduce overfitting, and lower computational cost without sacrificing predictive power.

4.4. MODELING FRAMEWORK

This section outlines the modeling strategy used to predict the engineer's estimated unit prices for highway construction line items, emphasizing temporal integrity, scalability, and reproducibility for DNN models and ensemble models.

4.4.1 DATA SPLITTING AND VALIDATION DATASET

To ensure temporal consistency during model development, a 5-fold time series cross-validation strategy was applied for model validation and hyperparameter tuning. In this method, the dataset is split into sequential folds while preserving the chronological order of the data. Each fold expands the training window and uses the subsequent period as the validation set. This approach allows the model to be validated on future data in a realistic, time-aware manner, avoiding any look-ahead bias.

Unlike traditional K-fold cross-validation, which randomly shuffles data, time series cross-validation respects the temporal structure, an essential requirement when working with sequential data where current values depend on past observations. This method was used solely during model training and validation to evaluate generalization performance and optimize model parameters.

The final evaluation of the model was performed separately on a hold-out test set, consisting of the most recent three months of data, which remained completely unseen during the training and validation phases. The remaining data was used for training. This combination of time-based validation and out-of-sample testing ensures that the model development process aligns with the sequential nature of the construction data and supports reliable forecasting by evaluating on the most recent, unseen data. **Figure 9** shows the time series cross-validation split of the training dataset.

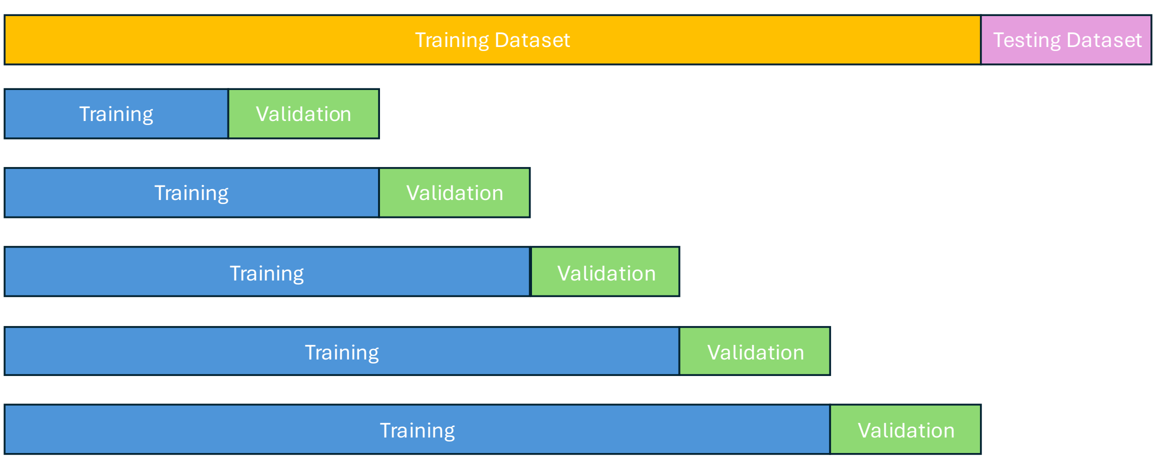


Figure 9 Time series cross-validation split

4.4.2 FEATURE ENCODING AND NORMALIZATION

Categorical Encoding: High-cardinality variables and interaction terms were label encoded. Lower-cardinality features (e.g., item type, unit) were one-hot encoded for efficiency.

Numerical Scaling: Continuous features were standardized using statistics from the training set to ensure consistency across validation and test folds.

4.4.3 MODEL ARCHITECTURE

The predictive modeling framework incorporated both deep neural network (DNN) and ensemble learning approaches, designed to integrate heterogeneous input modalities via dedicated processing components. The following subsections detail the architectural configurations used for each model type.

Deep Neural Network (DNN) Architecture

The DNN model was constructed to handle mixed-type data inputs and capture complex nonlinear relationships among variables:

- **Input Representation**

Categorical features and the interaction terms were mapped into dense vectors using trainable embedding layers, while one-hot encoded categorical variables and normalized numerical features were directly passed into the dense layers.

- **Embedding Layers**

Each categorical and interaction feature was assigned a dedicated embedding layer, with dimensionality ranging from 4 to 16 depending on the feature's cardinality.

- **Hidden Layers**

The combined feature vector consisting of embeddings, numeric values, and one-hot encodings was passed through a sequence of fully connected layers:

  - Number of layers: 3 to 10
  - Units per layer: 2 to 512
  - Activation functions: ReLU, Swish, Tanh, and ELU

- Regularization: Dropout (0.2 to 0.6) and L2 weight decay
- Learning rate: 0.00001 to 0.1

- **Output Layer**

A single linear neuron was used to predict the log-transformed engineer's unit price.

Ensemble Model Architecture

To improve predictive performance, an ensemble model architecture was developed by combining convolutional, recurrent, and feedforward components:

- **Input Representation**
  Categorical features and engineered interactions were mapped into dense vectors using trainable embedding layers, while one-hot encoded categorical variables and normalized numerical features were directly passed into the dense layers.
- **Embedding Layers**
  Each categorical and interaction feature was assigned a dedicated embedding layer, with dimensionality ranging from 4 to 16 depending on the feature's cardinality.
- **CNN and LSTM Branches**
  Normalized numerical inputs were reshaped and processed via two parallel modules:
  - 1D Convolutional Neural Network (CNN): Captured local spatial trends using convolution and global max pooling.
  - Long Short-Term Memory (LSTM): Modeled temporal dependencies over time-engineered features.
- **Dense Stack and Output**
  - Outputs from embeddings, CNN, LSTM, and one-hot layers were merged and passed through 2 to 10 fully connected layers.
  - Layer sizes (8 to 128 units), activation functions (ReLU, Swish, Tanh, ELU), dropout (0.2 to 0.5), and L2 regularization were tuned.

The final output was a single neuron predicting the log-transformed engineer unit price.

4.4.4 HYPERPARAMETER TUNING

Both the DNN and ensemble architectures' hyperparameters were optimized using the Optuna hyperparameter search framework as defined by Akiba et al. (2019). A 5-fold time series cross-validation was applied to preserve the temporal sequence of project data and avoid data leakage. The tuning objective was to minimize Mean Absolute Error (MAE). Key hyperparameters explored included:

- Number of hidden layers: 3 to 10
- Units per layer: 2 to 512 for DNN models and 8 to 128 for Ensemble models
- Activation functions: ReLU, Swish, Tanh, and ELU
- Dropout rates: 0.2 to 0.6 for DNN models and 0.2 to 0.5 for Ensemble models
- L2 regularization strength
- Learning rate: 0.00001 to 0.1

For the ensemble model, additional tuning involved the integration of CNN and LSTM branches with dense layers. Early stopping and pruning strategies were employed in both models to enhance training efficiency and reduce overfitting. The best-performing configuration for each model was selected based on the lowest average mean absolute error across validation folds.

4.4.5 MODEL EVALUATION

The results obtained from the developed machine learning models were evaluated using three metrics (Mean Absolute Error, Root Mean Squared Error, and Mean Absolute Percentage Error) on training and testing datasets. The following three metrics were used to evaluate the performance of machine learning models on in-sample (training) and out-of-sample (testing) datasets.

Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) represents the average of the absolute differences between predicted and actual values, disregarding whether the predictions are over or under the actual values. As a linear metric, it assigns equal weight to all errors, making it a straightforward and interpretable measure of model accuracy. The metric is defined as Equation 1.1 below.

$$\mathbf{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\mathbf{y} - \mathbf{y}_i| \qquad \text{Equation 1.1}$$

where y is the actual value, y_i is the predicted value, and n is the number of observations.

Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) calculates the square root of the average of the squared differences between predicted and actual values. Unlike MAE, RMSE gives greater weight to larger errors, making it particularly sensitive to outliers and useful for highlighting models that occasionally make large mistakes. The metric is defined as Equation 1.2 below.

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathbf{y} - \mathbf{y}_i)^2} \qquad \text{Equation 1.2}$$

where y is the actual value, y_i is the predicted value, and n is the number of observations.

Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) quantifies the average absolute difference between predicted and actual values as a percentage of the actual values. This metric provides an intuitive, scale-independent measure of prediction accuracy, making it especially useful for comparing model performance across different datasets or units. The metric is defined as Equation 1.3 below.

$$\mathbf{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|\mathbf{y} - \mathbf{y}_i|}{y} \qquad \text{Equation 1.3}$$

where y is the actual value, y_i is the predicted value, and n is the number of observations.

4.4.6 FORECASTING FUTURE PRICES

The target or dependent variable (the engineer's unit price) was log-transformed to stabilize variance and normalize skewness during the model development. To return predictions to the original scale (in dollar units), the inverse transformation is applied. Furthermore, to improve numerical stability and avoid the influence of extreme or spurious predictions, all predicted dollar values were clipped to lie within the 1st and 99th or 5th and 95th percentiles of the training target distribution. This approach ensures that future price estimates remain within the plausible range observed during model training. A structured pipeline was followed to predict prices for the next three months:

- Generated all combinations of county × project type × item code × month
- Imputed missing values using historical monthly averages
- Applied trained encoders and scalers
- Removed unseen categorical values
- Predicted log values and converted them to the dollar scale
- Exported the predicted engineer's estimate to Excel

4.4.7 AUTOMATION AND REPRODUCIBILITY

The full pipeline from data processing to forecasting is modular and fully automated. The process is easily replicable across all line items using preprocessed datasets, enabling scalable forecasting for large infrastructure datasets. The models can be generalized with the addition of more datasets.

4.5. RESULTS OF THE MACHINE LEARNING MODELS

This subsection depicts the training and testing errors for individual line-item models as well as a multi-task model developed to estimate unit prices for TxDOT highway construction line items. Each model's performance is evaluated using three common metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics are reported separately for both the training and testing datasets to assess the model's generalization ability.

Table 6 presents the performance of the deep neural network models for the eight highway construction line items and a multi-task model. Performance was evaluated using MAE, RMSE, and MAPE for both training and testing phases.

Table 6 Train and test results for deep neural network models

Line Item	Training Errors			Testing Errors		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Excavation	0.23	0.32	8.04	0.34	0.45	11.25
Backfill	0.21	0.31	4.46	0.24	0.35	5.24
Asphalt	0.12	0.16	14.22	0.22	0.28	26.97
Barricades, Signs and Traffic Handling	0.33	0.43	3.75	0.35	0.44	3.93
TMA	0.27	0.38	5.44	0.28	0.39	5.50
Vegetative Watering	0.36	0.51	12.76	0.43	0.59	12.37

Mobilization	0.41	0.55	4.57	0.63	0.87	7.43
Portable Changeable Message Sign	0.29	0.42	4.68	0.45	0.64	7.61
Multi-task Model	0.34	0.52	7.77	0.53	0.74	12.97

Table 6 shows that the model for Barricades, Signs, and Traffic Handling achieved a training MAE of 0.33, RMSE of 0.43, and MAPE of 3.75%, while its testing performance remained stable with MAE of 0.35, RMSE of 0.44, and MAPE of 3.93%. The Backfill model showed similarly strong results, with a training MAE of 0.21, RMSE of 0.31, and MAPE of 4.46%, and a testing MAE of 0.24, RMSE of 0.35, and MAPE of 5.24%. Likewise, the TMA model exhibited consistent behavior, with training errors of MAE 0.27, RMSE 0.38, and MAPE 5.44%, and testing errors of MAE 0.28, RMSE 0.39, and MAPE 5.50%.

Other models showed more variability between training and testing phases. The Portable Changeable Message Sign model had a training MAE of 0.29, RMSE of 0.42, and MAPE of 4.68%, but testing metrics increased to MAE 0.45, RMSE 0.64, and MAPE 7.61%. The Excavation model reported a training MAE of 0.23, RMSE of 0.32, and MAPE of 8.04%, with testing values of MAE 0.34, RMSE 0.45, and MAPE 11.25%. The Vegetative Watering model had generally higher errors, with training MAE 0.36, RMSE 0.51, and MAPE 12.76%, and testing MAE 0.43, RMSE 0.59, and MAPE 12.37%. The Mobilization model showed notable differences between training and testing, increasing from MAE 0.41, RMSE 0.55, and MAPE 4.57% in training to MAE 0.63, RMSE 0.87, and MAPE 7.43% in testing.

The Asphalt model, despite a low training MAE of 0.12 and RMSE of 0.16, with a MAPE of 14.22%, experienced a substantial rise in testing errors, reaching MAE 0.22, RMSE 0.28, and a much higher MAPE of 26.97%, indicating high variability in asphalt unit price data. Finally, the multi-task model, which predicted all line items jointly, showed moderate training errors (MAE 0.34, RMSE 0.52, and MAPE 7.77%) but an apparent increase in testing errors (MAE 0.53, RMSE 0.74, and MAPE 12.97%), reflecting the limitations of using a single generalized model across diverse item types.

These results suggest that line-item-specific models generally outperform the multi-task model, particularly in terms of prediction accuracy and generalization. The variation in errors across line items underscores the importance of accounting for item-specific characteristics such as unit pricing variability, data volume, and temporal patterns when developing cost estimation models for highway construction projects.

Similarly, **Table 7** presents the performance of the ensemble model, which integrates LSTM, CNN, and neural network architectures, across eight highway construction line items. Performance was evaluated using MAE, RMSE, and MAPE for both training and testing phases.

Table 7 Train and test results for ensemble models

Line Item	**Training Errors**			**Testing Errors**		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Excavation	0.35	0.49	12.17	0.58	0.76	19.68
Backfill	0.22	0.30	4.42	0.35	0.48	7.02
Asphalt	0.21	0.27	31	0.37	0.55	49.53
Barricades, Signs and Traffic Handling	0.39	0.49	4.57	0.38	0.49	4.40
TMA	0.31	0.42	6.15	0.38	0.48	7.21
Vegetative Watering	0.53	0.69	17.73	0.62	0.80	18.28
Mobilization	0.89	1.14	9.54	0.97	1.21	10.26
Portable Changeable Message Sign	0.33	0.56	5.44	0.72	0.27	10.72
Multi-task Model	0.40	0.58	7.50	0.60	0.84	10.89

Table 7 shows that the Barricades, Signs, and Traffic Handling model exhibited the most stable and accurate performance, with a training MAE of 0.39, RMSE of 0.49, and MAPE of 4.57%, and slightly improved testing values of MAE 0.38, RMSE 0.49, and MAPE 4.40%. The Backfill model also performed well, with a training MAE of 0.22, RMSE 0.30, and MAPE 4.42%, and testing metrics of MAE 0.35, RMSE 0.48, and MAPE 7.02%. The TMA model maintained balanced results, with training MAPE 6.15% and testing MAPE 7.21%, reflecting consistent learning across both phases.

However, other models demonstrated more variation. The Excavation model showed a notable increase in error from training (MAPE 12.17%) to testing (MAPE 19.68%), while the Vegetative Watering model remained relatively consistent with high MAPE values of 17.73% (training) and 18.28% (testing). The Mobilization model had relatively high errors as well, with testing MAE 0.97, RMSE 1.21, and MAPE 10.26%. The Asphalt model experienced significant overfitting, increasing from a training MAPE of 31% to a testing MAPE of 49.53%. The Portable Changeable Message Sign model showed a training MAPE of 5.44% and a notable increase in testing MAPE to 10.72%, indicating possible instability or overfitting. The multi-task model, which attempts to predict all line items simultaneously, produced moderate results with a testing MAE of 0.60, RMSE of 0.84, and MAPE of 10.89%.

The ensemble model offered improved generalization for some items, such as Barricades and TMA, while showing increased error for others, such as Asphalt and Excavation, in comparison to the DNN models. In general, DNN models had lower testing MAPEs for line items like Backfill, Asphalt, and Portable Signs. Conversely, the ensemble model outperformed in terms of stability for items with strong temporal or sequential patterns. These findings suggest that model selection should be guided by the specific characteristics of each line item, where ensemble models may be better suited for capturing temporal complexity, while DNNs may generalize better across more stable or well-represented features.

CHAPTER 5. MIXED-EFFECTS MODEL DEVELOPMENT

5.1. INTRODUCTION

A mixed-effects model is a statistical model that integrates fixed effects, which represent the average relationship between predictors and the outcome across all observations, with random effects, which address unobserved heterogeneity and correlation within groups or clusters (Laird and Ware 1982; Meteyard and Davies 2020). In our dataset, individual items are nested within projects, which are nested within counties. Mixed-effects models allow for intercepts and slopes across groups to vary while leveraging the overall population strength to enhance estimation efficiency (Meteyard and Davies 2020). Hence, a mixed-effects model is suitable for our hierarchical data structure.

5.2. DATASET FOR MODEL DEVELOPMENT

The dataset comprises 31 months of data, from March 4, 2022, to September 25, 2024, encompassing 13,398 observations categorized into 115 item types, 299 project types, and 212 counties. The top eight most repeated line items were selected from the pool of construction line items based on their repetition over all the counties and over the two years.

- Truck-Mounted Attenuator (TMA)
- Backfill
- Excavation
- Vegetative Watering
- Mobilization
- Barricades, Signs and Traffic Management
- Portable Changeable Message Signs
- Asphalt Operations

With additional years of data, models can be developed to estimate unit prices of more line items.

5.3. VARIABLE ANALYSIS FOR MODEL DEVELOPMENT

The variables in the model are collected at different levels (project-level, item-level, county-level, and time-level). **Table 8** shows variable definitions, acronyms, and the granularity of data used for model development.

Table 8 Description of variables

Variable name	Acronym	Description	Level	Frequency
Engineer's estimate	log(EngEst)	Natural logarithm of the preliminary estimated value of an item by TxDOT's engineer, used as a budgeting baseline.	Item	--
Quantity	log(Qty)	Natural logarithm of the quantity of a bid item required in the project.	Item	--
Project duration	log(Dur)	Natural logarithm of the estimated number of working days to complete a project.	Project	--
Number of items in project	log(Items)	Natural logarithm of the number of bid line items in a project.	Project	--
Number of bidders	log(Bids)	Natural logarithm of the number of contractors who submitted bids for the project.	Project	--
Unemployment rate	UnempRate	Percentage of unemployed individuals in the civilian labor force.	County	Monthly
Establishment counts	log(Estab)	Natural logarithm of the number of private construction establishments in the county.	County	Quarterly
Precipitation	Precip	Any form of water from the atmosphere that reaches the ground (e.g., rain, snow, hail).	County	Monthly
County area	log(Area)	Natural logarithm of the total land and water area of the county.	County	--
Crude oil price	log(WTI)	Natural logarithm of the West Texas Intermediate crude oil price, a U.S. benchmark.	National	Monthly
Construction Cost Index	log(CCI)	Natural logarithm of the construction cost index for 20 major U.S. cities.	National	Monthly

In order to study the distribution and central tendencies of the variables included in the model, as defined in **Table 8**, descriptive statistics are performed. **Table 9** presents the descriptive statistics of the variables.

Table 9 Descriptive statistics of variables

Variable	Mean	Median	Std. Dev.	Min	Max
log(EngEst)	6.733	5.858	3.194	-2.303	17.823
log(Qty)	3.404	3.178	2.714	-4.605	14.055
log(Dur)	4.979	5.03	0.917	1.946	7.805
log(Items)	3.704	3.807	1.061	0	6.753
log(Bids)	1.242	1.386	0.517	0	2.565
UnempRate	4.11	4	0.923	2.1	11
log(Estab)	5.648	5.768	1.892	0	9.057
Precip	2.927	2.17	2.662	0	18.13
log(Area)	6.861	6.829	0.406	5.003	8.731
log(WTI)	4.358	4.358	0.068	4.252	4.687
log(CCI)	9.505	9.51	0.012	9.457	9.52

The engineer's estimate (log(EngEst)) exhibits considerable variation (SD = 3.194), indicating diverse project scales, market conditions, item types, and complexity levels across the hierarchical structure of items nested within projects. The quantity of bid items, log(Qty), demonstrates high variability (SD = 2.714), indicating a wide range of construction volumes, while project duration (log(Dur)) displays moderate variance (SD = 0.917) with a mean of 4.979. The number of items per project (log(Items)) shows variability (SD = 1.061), but the number of bidders (log(Bids)) has less dispersion (SD = 0.517), with minimum values of zero indicating that some projects attracted only a single bidder, as the natural logarithm of one equals zero. The economic and environmental variables capture diverse conditions across counties and periods. Unemployment rates (UnempRate) vary modestly around a mean of 4.110 (SD = 0.923), spanning from 2.100% to 11.000%. Meanwhile, establishment counts, log(Estab), exhibit substantial heterogeneity (SD = 1.892), reflecting varying levels of local construction industry presence. Precipitation (Precip) exhibits high variability (SD = 2.662), suggesting diverse weather conditions that may impact construction activities. The county area, log(Area), exhibits relatively low variation (SD = 0.406), indicating that the sample comprises counties of similar geographic scope. In contrast, the macroeconomic indicators, like West Texas Intermediate crude oil price, log(WTI), and the

Construction Cost Index, log(CCI), display minimal variation (SD = 0.068 and 0.012, respectively), consistent with their role as time-varying factors affecting all observations uniformly within specific periods. The wide ranges across most variables underscore the dataset's comprehensive coverage of diverse market conditions and economic environments. **Table 10** displays the correlation matrix of all variables, illustrating the pairwise correlations among the explanatory and control variables in our study.

Table 10 Correlation matrix

Variables	log(EngEst)	log(Qty)	log(Dur)	log(Items)	log(Bids)	UnempRate	log(Estab)	Precip	log(Area)	log(WTI)	log(CCI)
log(EngEst)	1										
log(Qty)	-0.809	1									
log(Dur)	-0.113	0.304	1								
log(Items)	-0.084	0.254	0.494	1							
log(Bids)	0.015	-0.026	-0.1	-0.038	1						
UnempRate	0	-0.006	0.016	0.007	-0.081	1					
log(Estab)	-0.004	0.005	0.114	0.073	0.013	-0.088	1				
Precip	-0.014	-0.004	-0.029	-0.037	0.044	0.119	0.045	1			
log(Area)	0.001	0.01	0.041	0.033	-0.072	-0.033	0.159	-0.048	1		
log(WTI)	-0.002	0.015	0.009	0.022	-0.017	-0.053	-0.064	0.08	-0.015	1	
log(CCI)	0.043	-0.022	0.011	-0.039	0.05	0.112	0.023	0.118	-0.011	0.062	1

From **Table 10**, it can be noted that the item characteristic quantity is negatively correlated with both the engineer's estimate (-0.809) and the average bid (-0.778), suggesting that larger quantities are associated with lower unit costs. A moderate positive correlation (0.494) exists between the number of items and project duration, indicating that projects with a greater number of line items tend to require more time to complete.

5.4. MODEL DEVELOPMENT

The final mixed-effects model is developed through a systematic evaluation of alternative formulations, each designed to address the econometric challenges posed by the hierarchical structure of the data. This section describes the model progression and the corresponding results.

5.4.1. BASELINE MODELS

To establish baseline estimates before addressing hierarchical structure and heterogeneity concerns, standard regression methods are employed.

Ordinary Least Squares (OLS) Regression

The ordinary least squares (OLS) model is used as our first framework, considering all observations as independent and disregarding the hierarchical data structure. The model estimates coefficients for all variables without considering the potential correlation of observations within the same project or county. The pooled OLS model serves as the starting point, treating all observations as independent and ignoring the nested data structure. **Table 11** presents the results of the OLS for the engineer's estimates.

Table 11 Results from Ordinary Least Squares (OLS) regression

Variable	log(EngEst)	VIF
log(Qty)	-1.015*** (-0.007)	1.12
log(Dur)	0.383*** (-0.02)	1.41
log(Items)	0.251*** (-0.018)	1.35
log(Bids)	0.037 (-0.032)	1.03
UnempRate	-0.034** (-0.017)	1.05
log(Area)	-0.033*** (-0.009)	1.04
Precip	-0.015** (-0.006)	1.04
log(WTI)	0.039 (-0.041)	1.02
log(Estab)	0.29 (-0.233)	1.06
log(CCI)	7.775*** (-1.372)	1.04
Intercept	-67.757*** (-13.028)	--
Mean VIF	--	1.12
Number of observations	13,398	--
R-squared	0.6794	--
Item types	115	--
Project types	299	--
Counties	212	--
Months	25	--

*Note: Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.*

Table 11 also presents Variance Inflation Factors (VIF) and mean VIF statistics to assess multicollinearity concerns among the explanatory variables. The VIF indicates low collinearity concerns, with individual VIF values below two and a mean VIF of 1.12, which is well below the conventional threshold of concern (VIF > 5). The models represent substantial variation in the dependent variables, with R-squared values of 0.6794 for engineers' estimates, based on 13,398 observations across 115 item types, 299 project types, 212 counties, and 31 months. However, since OLS overlooks the nested structure of items clustered within projects and counties, it is likely to result in biased standard errors and estimates. Also, while providing initial coefficient estimates, this specification fails to account for correlation within counties and projects, likely producing biased standard errors and inefficient estimates.

Random Effects Regression

A single-level random effects model includes random variation at one hierarchical level, such as random effects for item types, project types, or counties, but not multiple levels simultaneously. This approach allows group-specific intercepts that vary randomly around the overall mean. **Table 12** presents three separate single-level random effects specifications for both dependent variables: item-type random effects, project-type random effects, and county-level random effects.

Table 12 Random effects regression results

	Item level	**Project level**	**County level**
Variable	**log(EngEst)**	**log(EngEst)**	**log(EngEst)**
log(Qty)	-0.407*** -0.007	-1.046*** -0.006	-1.018*** -0.006
log(Dur)	0.273*** -0.014	0.583*** -0.024	0.380*** -0.021
log(Items)	0.327*** -0.012	0.014 -0.026	0.257*** -0.018
log(Bids)	0.017 -0.019	0.049 -0.033	0.082** -0.032
UnempRate	-0.039*** -0.011	-0.028 -0.017	-0.051** -0.024
log(Estab)	-0.052*** -0.005	-0.015* -0.009	-0.046** -0.017
Precip	0.001 -0.004	-0.006 -0.006	-0.005 -0.006
log(Area)	0.035 -0.024	0.009 -0.039	0.023 -0.064
log(WTI)	0.003 -0.143	0.021 -0.232	0.192 -0.236
log(CCI)	6.874*** -0.849	9.406*** -1.384	8.281*** -1.403
Intercept	-61.195*** -8.05	-82.161*** -13.109	-71.961*** -13.292

Number of groups	115	299	212
Within R-squared	0.263	0.702	0.6853
R-squared	0.6217	0.6742	0.6792
Rho	0.5733	0.1289	0.0233
p-value	0	0	0
Observations	13,398	13,398	13,398

*Note: Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.*

The intraclass correlation coefficients (rho) in **Table 12** show considerable variation in grouping effects across specifications, ranging from 0.0233 for county-level to 0.5733 for item-type random effects, indicating that between 2.33% and 57.33% of the total variance goes to the corresponding group-level differences. For example, the rho value of 0.5733 for item-type random effects means that 57.33% of the total variance in the log(EngEst) is due to differences between different item types. However, our data format is hierarchical; thus, these single-level random effect models are insufficient to address the multi-level hierarchy of items within projects across counties.

Fixed Effects Regression for Addressing Heterogeneity Concerns

Heterogeneity concern arises when individuals or groups have unobserved differences that are not accounted for in the model. Different construction line items, such as concrete work, asphalt paving, and steel structures, exhibit different cost structures, technical requirements, and markets. Additionally, prices may vary across geographic locations (such as counties) and periods due to local economic conditions, seasonal effects, and fluctuations in material costs. We developed models that include fixed effects for item types and project types, in addition to various combinations of temporal (month) and spatial (county) controls to prepare for the multi-dimensional heterogeneity of our data, as detailed in the following subsections.

I. Item Level Heterogeneity and Temporal Spatial Variation

This specification introduces item-specific controls and temporal-spatial factors to capture variation in item characteristics and location-time effects. **Table 13** presents results for the engineer's estimates (log(EngEst)), indicating improvements in model performance when considering item-level heterogeneity and temporal-spatial variation.

Table 13 Item-level heterogeneity and temporal spatial variation for log(EngEst)

	log(EngEst)				
log(Qty)	–	-0.423*** -0.007	-0.450*** -0.015	-0.402*** -0.007	-0.450*** -0.007
log(Dur)	0.132*** -0.018	0.368*** -0.017	0.407*** -0.031	0.259*** -0.014	0.422*** -0.023
log(Items)	0.135*** -0.022	0.155*** -0.02	0.207*** -0.035	0.340*** -0.012	0.178*** -0.027
log(Bids)	-0.044* -0.024	0.035 -0.023	0.145** -0.045	0.068** -0.021	0.015 -0.03
UnempRate	–	0.051 -0.054	–	0.046 -0.054	0.101 -0.069
log(Estab)	–	0.095 -0.316	–	-0.208 -0.309	0.602 -0.458
Precip	–	-0.003 -0.006	–	-0.001 -0.006	0.003 -0.008
log(Area)	–	-3.324 -11.598	–	8.864 -10.89	3.799 -3.778
log(WTI)	–	5.613 -6.186	–	6.233 -6.398	–
log(CCI)	–	34.122 -27.58	–	28.978 -28.121	–
Intercept	5.686*** -0.096	-321.263 -297.476	5.695*** -0.164	-357.513 -302.483	-25.636 -29.055
Fixed Effects	ItemType×Month & ProjType	ItemType & ProjType	ItemType×Month×County & ProjType	ItemType	ItemType×Month & ProjType×Month
Dummy Variables	–	Month & County	–	Month & County	County
Observations	13,099	13,360	5,643	13,365	13,075
Adjusted R-squared	0.8622	0.8973	0.8949	0.8844	0.9033
AIC	40,797.92	38,383.80	12,841.36	40,275.86	35,317.26

*Note: Significance levels are indicated as follows: *** p < 0.01, ** p < 0.05, * p < 0.10.*

Table 13 demonstrates an increase in adjusted R-squared values across models as compared to OLS models in **Table 11** and the RE model in **Table 12**, suggesting that item-type heterogeneity accounts for a considerable portion of the variation in both dependent variables. The AIC values indicate substantial improvements, reflecting the model's advantages compared to simpler models. These models address item-level variation but do not fully capture the complete hierarchical structure where items are nested within projects within counties.

II. Project Type Heterogeneity with Incremental Spatial and Temporal Controls

The heterogeneity of project types, together with spatial and temporal controls, indicates an advancement in our approach by considering the cost structures, purchasing patterns, and market dynamics associated with various construction project categories, such as highway construction, bridge building, and utility work, which may also differ over time and across geographic locations. This approach strengthens the item-level heterogeneity controls from the previous models by including project-type fixed effects, along with various spatial (county) and temporal (month) controls, which address the multidimensional aspects of construction pricing heterogeneity. The models change from basic project-type-month interactions (ProjType × Month) to more complex structures that incorporate distinct project-type effects (ProjType) with interaction terms. Ultimately, we included additional dummy variables for months and counties to account for broader temporal trends and geographic influences that operate independently of project-type-specific variations. **Table 14** presents the results for the engineer's estimates (log(EngEst)), showing the effects of adjusting for project-type heterogeneity through incremental spatial and temporal controls.

Table 14 Project type heterogeneity with incremental spatial and temporal controls for log(EngEst)

	log(EngEst)				
log(Qty)			-1.083*** -0.006	-1.060*** -0.006	-1.060*** -0.006
log(Dur)	-0.011 -0.064	-0.188** -0.083	0.609*** -0.044	0.640*** -0.027	0.642*** -0.027
log(Items)	-0.453*** -0.076	-0.289** -0.094	-0.052 -0.05	-0.095** -0.033	-0.092** -0.032
log(Bids)	-0.127 -0.083	0.115 -0.123	0.112* -0.065	0.052 -0.038	0.068* -0.037
UnempRate				0.037 -0.09	-0.04 -0.045
log(Estab)				-0.087 -0.522	-0.152 -0.517
Precip				0 -0.01	0.006 -0.007
log(Area)				-3.279 -16.378	-0.993 -16.222
log(WTI)				2.251 -10.305	-0.185 -0.248
log(CCI)				21.623 -45.937	11.158*** -1.503
Intercept	8.627*** -0.337	8.608*** -0.456	7.448*** -0.24	-184.783 -489.903	-89.657 -110.281
Fixed Effects	ProjType×Month	ProjType×Month + ProjType	ProjType×Month + ProjType	ProjType	ProjType
Dummy Variables			Item	Month + County	County
Observations	13,363	13,303	13,303	13,391	13,390
Within R-sq.	0.0039	0.0025	0.7248	0.7119	0.7148
AIC	67,742.05	67,202.29	50,075.40	52,234.22	52,216.21

*Note: Significance levels are indicated as follows: *** p < 0.01, ** p < 0.05, * p < 0.10.*

The findings from **Table 14** indicate considerable improvements in model fit with the incremental inclusion of spatial and temporal controls. The differing significance patterns of key variables across various fixed-effects models indicate that project-type heterogeneity interacts closely with spatial and temporal factors. This shows the necessity for modeling approaches that can simultaneously address multiple causes of variation within the hierarchical data structure.

III. County-Level Heterogeneity and Temporal Specification

County-level heterogeneity and temporal specification are essential for addressing geographic clustering and unobserved county characteristics that affect the dependent variables. **Table 15** displays the effect of considering county-level heterogeneity through the models that combine fixed-effects structures.

Table 15 County-level heterogeneity and temporal specification

log(EngEst)		
log(Qty)	-1.024*** (-0.006)	-1.060*** (-0.006)
log(Dur)	0.381*** (-0.022)	0.640*** (-0.027)
log(Items)	0.251*** (-0.018)	-0.095** (-0.033)
log(Bids)	0.107** (-0.034)	0.052 (-0.038)
UnempRate	-0.003 (-0.089)	0.037 (-0.09)
log(Estab)	-0.278 (-0.507)	-0.087 (-0.522)
Precip	0.006 (-0.009)	0 (-0.01)
log(Area)	3.265 (-15.236)	-3.279 (-16.378)
log(WTI)	3.922 (-10.529)	2.251 (-10.305)
log(CCI)	19.67 (-46.278)	21.623 (-45.937)
Intercept	-217.349 (-493.44)	-184.783 (-489.903)
Fixed Effects	–	ProjType
DummyVariables	County + Month	County + Month
Observations	13,398	13,391
Within R-sq.	0.6927	0.7119
AIC	53,814.81	52,234.22

*Note: Significance levels are indicated as follows: *** p < 0.01, ** p < 0.05, * p < 0.10.*

This specification acknowledges that counties may exhibit variations in economic conditions, regulatory environments, material availability, labor markets, and contractor networks, necessitating consideration of spatial dependencies within the modeling framework. The model includes county and month dummy variables to adjust for cross-sectional geographic differences and temporal trends that consistently influence all observations within specified time frames.

The differing significance patterns of key variables across various fixed-effects models indicate that project-type heterogeneity interacts closely with spatial and temporal factors. This shows the necessity for modeling approaches that can simultaneously address multiple causes of variation within the hierarchical data structure.

5.4.2. MIXED EFFECTS MODEL

We developed a mixed-effects model to control for county, project type, and item type heterogeneity. This approach strengthens the item-level heterogeneity controls from the previous models by including project-type fixed effects, along with various spatial (county) and temporal (month) controls, which address the multidimensional aspects of construction pricing heterogeneity. Seven mixed effects models are developed that change from basic project-type-month interactions (ProjType × Month) to more complex structures that incorporate distinct project type effects (ProjType) with interaction terms. Ultimately, they include additional dummy variables for months and counties to account for broader temporal trends and geographic influences that operate independently of project-type-specific variations. **Table 16** presents results for engineer's estimates, log(EngEst), showing the effects of adjusting for project-type heterogeneity through incremental spatial and temporal controls.

Table 16 Mixed effects models controlling for county, project type, and item type heterogeneity for log(EngEst)

	Log(EngEst)						
Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
log(Qty)	-0.939*** -0.006	-0.939*** -0.006	-0.525*** -0.007	-0.525*** -0.007	-0.422*** -0.007	-0.430*** -0.007	-0.443*** -0.007
log(Dur)	0.474*** -0.02	0.470*** -0.02	0.404*** -0.015	0.402*** -0.015	0.372*** -0.017	0.356*** -0.016	0.409*** -0.021
log(Items)	0.208*** -0.018	0.208*** -0.019	0.164*** -0.017	0.161*** -0.017	0.155*** -0.019	0.190*** -0.018	0.167*** -0.025
log(Bids)	0.093*** -0.028	0.083*** -0.028	0.056*** -0.02	0.041** -0.02	0.032 -0.021	0.002 -0.021	-0.028 -0.027
UnempRate	-0.036 -0.023	-0.04 -0.028	-0.01 -0.024	0.046 -0.048	-0.033** -0.016	-0.031** -0.012	-0.046*** -0.015
log(Estab)	-0.060*** -0.016	-0.059*** -0.016	-0.086 -0.281	-0.049 -0.284	-0.050*** -0.011	-0.042*** -0.006	-0.040*** -0.007
Precip	-0.005 -0.005	-0.011 -0.007	0.005 -0.004	-0.003 -0.005	0.003 -0.004	-0.004 -0.005	-0.002 -0.006
log(Area)	0.03 -0.064	0.031 -0.064			0.016 -0.042	0.012 -0.024	-0.017 -0.03
log(WTI)	0.2 -0.183		-0.149 -0.133		-0.095 -0.146		
log(CCI)	11.048*** -1.101		10.536*** -0.81		9.589*** -0.874		
Intercept	-98.955*** -10.426	7.128*** -0.484	-93.444*** -7.848	5.668*** -2.091	-80.606*** -8.266	9.491*** -0.22	5.256*** -0.546
Random Effects	County, ProjType, ItemType	County, ProjType, ItemType	ProjType, ItemType	ProjType, ItemType	County	ProjType	ItemType
Fixed Effects		Month	County	County + Month	ItemType + ProjType	ItemType×Month	ProjType× Month
Observations	13,398	13,398	13,398	13,398	13,398	13,398	13,398
Log Likelihood	-25,397.60	-25,421.80	-20,515.20	-20,539.60	-19,309.60	-19,170.80	-18,790.10
AIC	50,825.20	50,917.60	41,478.40	41,571.20	**39,465.30**	39,891.60	40,074.30
Parameters	15	37	224	246	423	775	1,247

Table 16 displays the results of the mixed-effects model for the engineer's estimate (log(EngEst)) across seven specifications, each including different combinations of fixed and random effects structures. The specifications range from models with random effects for county, project type, and item type, without fixed effects, to a more complex model that includes month fixed effects in the second model. The third and fourth models use random effects for project type and item type, while county is considered as a fixed effect. The latter model includes month fixed effects, too.

The fifth specification employs random effects for counties, while project type and item type are treated as fixed effects. The sixth specification includes fixed effects for item type and month, with random effects for project type. In contrast, the seventh model analyzes the interaction between project type and month as fixed effects, with random effects for item type.

Model selection based on the Akaike Information Criterion indicates that the fifth model has the lowest AIC value (39,465.3), representing the optimal balance between model fit and complexity. This model, which accounts for county-level heterogeneity through random effects while controlling for project and item type characteristics through fixed effects, proved to be the most appropriate model.

5.4.3. MODEL COMPARISON AND SELECTION CRITERIA

The results of the developed models in **Section 5.4.2** suggest the limitations of regression methods in the context of hierarchical construction data. The OLS model ignores the nested structure, while single-level random effects models just consider the association of variables within groups. The fixed-effects models address the heterogeneity by using various combinations of fixed effects for item types, project types, counties, and temporal controls. While these models suggest improvements in fit statistics and uncover relationships, they primarily consider the hierarchical structure as a problem to be managed rather than as a fundamental aspect of the data creation process. Additionally, they are not efficient in managing the large number of parameters needed when all group-specific effects are treated as fixed. This limitation presents issues when particular groups have limited observations, resulting in imprecise estimates and increased standard errors.

5.4.4. PREDICTION AND PERFORMANCE EVALUATION

Model 5 from **Table 16** for engineers' estimates (log(EngEst)) is the selected model for which we utilize mixed-effects models to enhance predictive accuracy by considering the hierarchical structure of our data. In contrast to conventional methods that assume independence among observations, these mixed-effects models utilize both population-level fixed effects and group-specific random effects to produce more precise predictions. The evaluation of predictions for engineers' estimates indicates that the mixed-effects model yields a Symmetric Mean Absolute

Percentage Error (SMAPE) of 66.2% and a Root Mean Square Error (RMSE) of 1.00 on the logarithmic scale.

5.4.5. RESULT INTERPRETATION

In model 5, the quantity variable (log(Qty) shows a strong negative relationship with the engineer's estimate, with a coefficient of -0.422, indicating that a 1% increase in quantity is associated with approximately a 0.42% decrease in the unit price estimate, possibly because of production efficiency and economy of scale (PennDOT 2025).

The project duration (log(Dur)) demonstrates a positive and significant effect (0.372), indicating that longer projects tend to have higher unit prices, potentially because accelerated work schedules may require additional resources or overtime pay, leading to unit price inflation (Levy 2006).

The number of items per project (log(Items)) exhibits a positive relationship (0.155), indicating that projects with more line items tend to have higher unit estimates.

The number of bidders (log(Bids)) shows an insignificant positive coefficient (0.032), which contradicts the idea that more competition should reduce prices (Zhang et al. 2023). However, this finding is consistent with Model 5's specification, which includes fixed effects for project type and item type. After accounting for differences among project and item types, the remaining variation in bidder participation may indicate project attractiveness rather than mere competitive pressure. More technically demanding projects within the same category could attract a greater number of bidders and require higher estimates.

The negative coefficient (-0.033) of unemployment rates (UnempRate)) is economically rational, because high unemployment generally results in decreased labor costs and greater contractor availability, which decreases project estimates (Jiang et al. 2022).

Establishment counts (log(Estab)) exhibit a significant negative coefficient (-0.050), indicating that counties with a higher number of construction experience increased competitive pressure, resulting in lower estimates (Jiang et al., 2022).

Environmental and geographic factors, such as precipitation (Precip) and county area (log(Area)), display statistically insignificant effects. This indicates that these variables do not impact engineers' estimates in model 5, which includes county-level random effects that account for much of the geographic and environmental variation typically associated with these variables.

The construction cost index (log(CCI)) has a significant positive association (9.589), which matches expectations that rises in aggregate construction costs result in higher project estimates (Zhang et al. 2023).

CHAPTER 6. DEVELOPMENT OF A GIS-BASED UNIT PRICE ESTIMATION AND VISUALIZATION TOOL

6.1. INTRODUCTION

This chapter defines the development of a web-based ArcGIS tool that facilitates easy visualization of the unit prices of construction line items. We used ESRI ArcGIS Pro for data visualization because of its ability to effectively represent spatial relationships and create time-series maps that help access changes in data over time. Thematic mapping and interactive visualization are some additional features of ESRI ArcGIS Pro that qualify its use in this project.

6.2. MAP-BASED INTERFACE TO VISUALIZE PRICE DATA

Users can access the map-based application using the web address: "https://axb9823.uta.cloud/UPEVT/login.php". The widgets included in the map-based interface is shown in **Figure 10**.

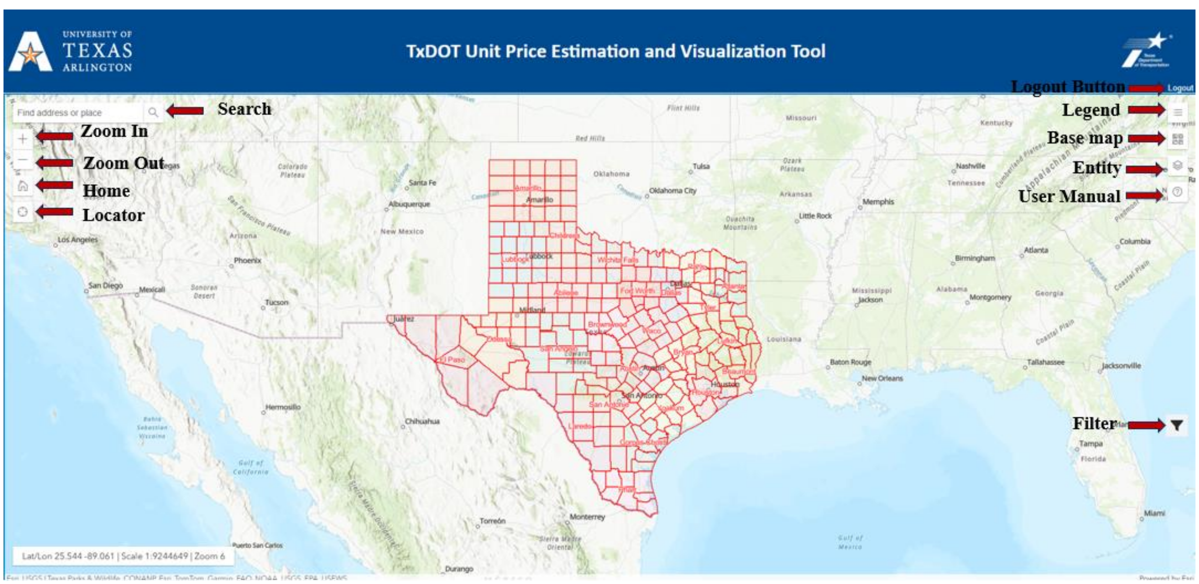


Figure 10 Features of map-based interface

The features pointed by red arrows are the widgets of the tool. Each widget is designed to enable users to customize the display by selecting the specific information they want to view. **Table 17** provides a concise overview of the widgets and their functionalities.

Table 17 Description of widgets in the map-based interface

Widget Name	Description
Search	This widget helps to find a specific location in the map-based interface.
Zoom In	This widget helps to zoom in on the map view in the map-based interface.
Zoom Out	This widget helps to zoom out of the current map view in the map-based interface.
Home	This widget brings the map view to the initial view extent.
Locator	This widget helps to find the location of the user.
Basemap	This widget allows the user to select the base map to be displayed in the map view of the map-based interface.
Entity	This widget displays the list of spatial data entities. Users can choose to display or remove the entities from the map in the map-based interface.
Editor	This widget allows the user to add or edit a slope failure feature in the map-based interface.
Legend	This widget displays the legends of the spatial data entities which are displayed in the map-based interface.
User Manual	This widget allows the user to access the user manual for the map-based interface.

Table 17 summarizes the key functionalities of various widgets within the map-based interface, highlighting their roles in enhancing user interaction and navigation. The widgets that facilitate map navigation options are Search, Zoom In, Zoom Out, Home, and Locator. Other widgets enable data interaction.

6.3. USE CASES

A use case is a set of possible sequences of interactions between a user and a system. The use case clearly indicates what action the system takes in as response to what action is taken by the user. A use-case diagram is a graphical table of contents for individual use cases and defines a system boundary. **Figure 11** represents the use case diagram for Unit Price Estimation and Visualization Tool (UPEVT).

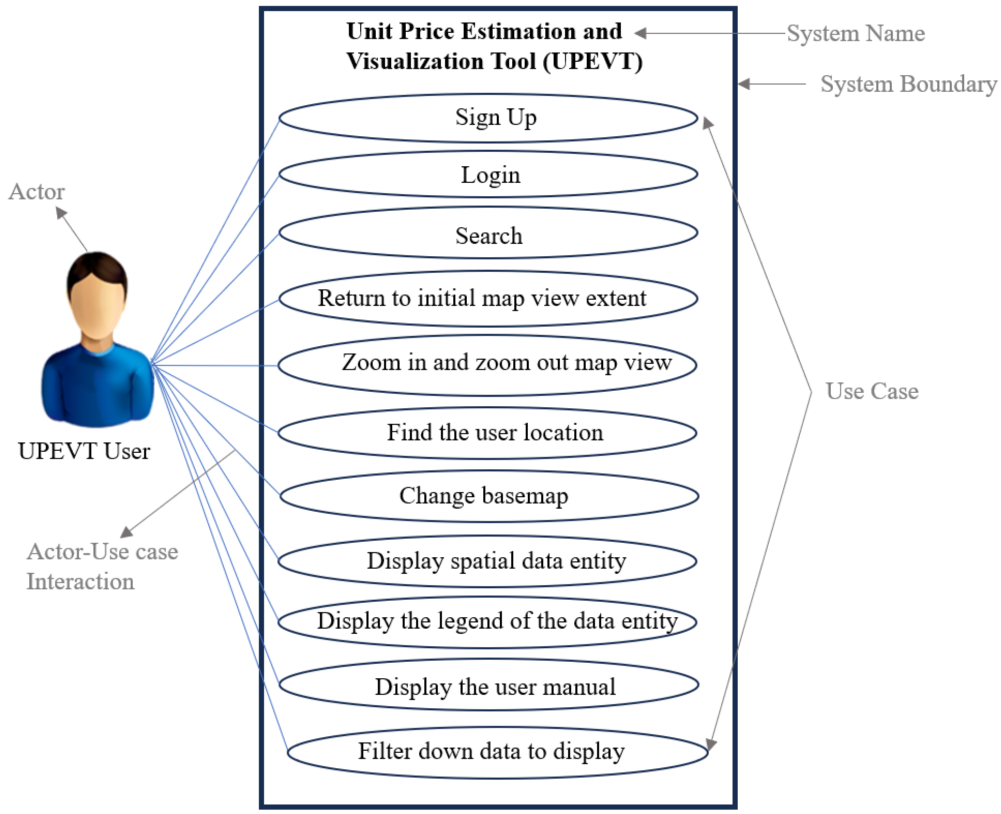


Figure 11 Use case diagram

Table 18 UC1: Signup

Actor: User	**System: UPEVT**
	0. The browser displays a web page.
1. The user enters the web address in the address bar and presses enter. I. URL: I. https://axb9823.uta.cloud/UPEVT/register.php	2. The system displays the signup page which prompts the user to sign up for a new account.
3. The user fills in the information (First Name, Last Name, Email address, Password) requested on the sign-up page and clicks the Sign-Up button to complete the process.	4. The system sends an email to the user's email address for activation of the account.
5. The user opens the email and clicks the activation link to activate the account.	6. The system registers the user and displays the confirmation of registration.

Table 19 UC2: Login

Actor: User	**System: UPEVT**
	0. The browser displays a web page.
1. The user enters the web address in the address bar and press enter. URL: https://axb9823.uta.cloud/UPEVT/login.php	2. The system displays the login page which prompts the user to login using a username and password.
3. The user enters the username and password and then clicks the login button.	4. The system displays a. The map-based interface if username and password are entered correctly. b. The message requesting to recheck inputs if the username or password is incorrect.
5. The user sees the map-based interface or login error is displayed.	

Table 20 UC3: Search

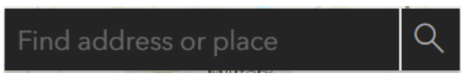
Actor: User	**System: UPEVT**
	0. The system displays the map-based interface.
1. The user enters the location on the search bar. 	2. The system displays the searched location.

Table 21 UC4: Return to initial map view extent

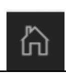
Actor: User	**System: UPEVT**
	0. The system displays the map-based interface.
1. The user clicks the home button. 	2. The system returns to the initial map view extent.

Table 22 UC5: Zoom in and zoom out of the map view

Actor: User	**System: UPEVT**
	0. The system displays the map-based interface.
1. The user clicks the zoom-in or zoom-out button. 	2. The system zooms in or zooms out in the map view of the map-based interface.

Table 23 UC6: Find the user's location

Actor: User	**System: UPEVT**
	0. The system displays the map-based interface.

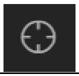
1. The user clicks the locator widget. 	2. The system displays the location of the user in the map-based interface.

Table 24 UC7: Change basemap

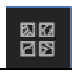
Actor: User	System: UPEVT
	0. The system displays the map-based interface.
1. The user clicks the base map widget. 	2. The system displays the available base maps from which the user can make the selection.
3. The user clicks on the desired base map.	4. The system changes the existing base map to the base map selected by the user.
5. The user clicks on the base map widget.	6. The system closes the expanded base map widget.

Table 25 UC8: Display spatial data entity

Actor: User	System: UPEVT
	0. The system displays the map-based interface.
1. The user clicks the entity widget. III. 	2. The system expands the entity widget and displays the list of spatial data entities.
3. The user clicks on the entity to turn it on and off.	4. The system displays or removes the entity from the map view.
5. The user clicks on the entity widget to close the list of entities.	6. The system closes the expanded entity widget.

Table 26 UC9: Display the legend of the data entity

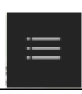
Actor: User	System: UPEVT
	0. The system displays the map-based interface.
1. The user clicks the legend widget. 	2. The system displays the legend of the entities displayed in the map-based interface.

Table 27 UC10: Display the user manual

Actor: User	System: UPEVT
	0. The system displays the map-based interface.
1. The user clicks the legend widget. 	2. The system expands the widget and provides the user an option to “ click here” to open the user manual.

3. The user clicks on “click here” to open the user manual.	4. The system opens the user manual.

Table 28 UC11: Filter Down the Data

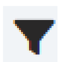
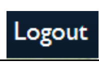
Actor: User	**System: UPEVT**
	0. The system displays the map-based interface.
1. The user clicks the filter widget. 	2. The system expands the widgets and provides the user with an option to select item, year, data type, and data details.
3. The user selects the line item, year, data type, and details to visualize on the interface.	4. The system displays engineer’s estimate or unit average bid (either average value or values with information on quantity) of the selected line item in the selected year across Texas counties in the map.

Table 29 UC12: Logout

Actor: User	**System: UPEVT**
	0. The system displays the map-based interface.
1. The user clicks the logout button on the application. 	2. The system exits the application.

A comprehensive summary of usermanual is presented in **Appendix B**. The usermanual provides necessary information for users to effectively access, navigate, and use the map-based interface for the Unit Price Estimation and Visualization Tool (UPEVT).

CHAPTER 7. IMPLEMENTATION OF THE DEVELOPED GIS-BASED VISUALIZATION TOOL ON SIX TXDOT PROJECTS

7.1. INTRODUCTION

This chapter defines how the GIS-based visualization tool we developed can visualize the unit prices of construction line items used in Texas projects via case studies of six projects whose estimation details are provided by TxDOT.

7.2. CASE STUDIES

This chapter presents the case studies conducted for six different projects in Texas. The research team received project cost estimates for six projects from six counties (i.e., Denton County, Runnels County, Tarrant County, Midland County, Terrell County, and San Angelo) to develop case studies and showcase the successful implementation of the visualization tool in cooperation with TxDOT. Each project has a key descriptor called Construction Control Section Job (CCSJ), which provides a record of that project in the Design and Construction Information System (DCIS) of TxDOT. Thus, we use CCSJ as an identifier for every project. **Figure 12** shows the locations of the projects for which estimation documents were provided by TxDOT.

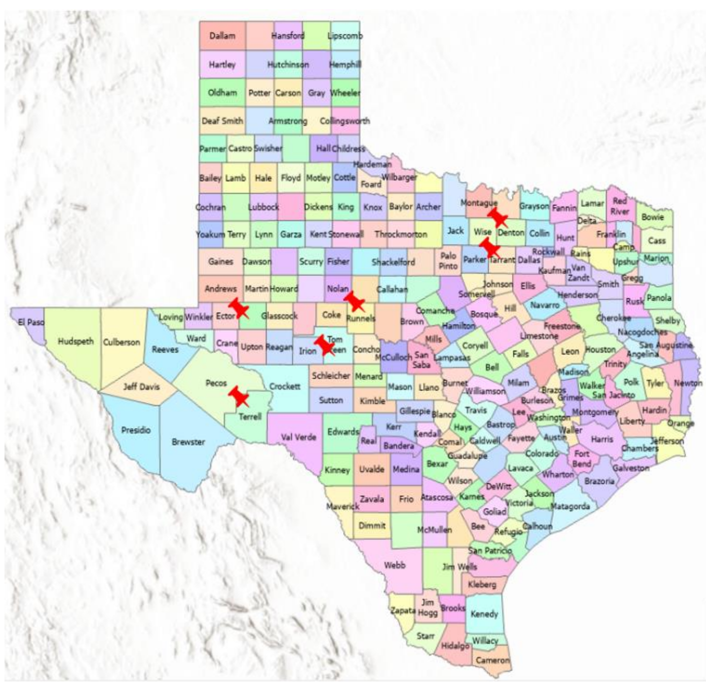


Figure 12 Locations of projects for case studies

The goal of conducting case studies is to show the performance of the developed tool. The estimated unit price values in project documents are compared with the predicted values of the estimates by visualizing historical data using the tool. First, all the available unit price values for an item are visualized along with the corresponding quantity. Then, the data is filtered based on the project type and the year. For every county in the case study, the research team tried to match all information about the item, project type, and time, and analyze unit price values. If data were unavailable in the same county, the research team analyzed the values in neighboring counties. The following subsections elaborate on the case studies of the projects in detail.

7.2.1 CASE 1: PROJECT 2054-01-018

The basic characteristics of the project 2054-01-018 are summarized in **Table 30**.

Table 30 Summary of the project 2054-01-018

CCSJ (Construction Control Section job)	Location	Highway Name	Project Type	Let Date	Number of Bidders	Number of Line Items	Total Estimated Cost of the Project
2054-01-018	Denton County	FM 2164	HIGHWAY IMPROVEMENT, WIDEN ROAD - ADD SHOULDERS	1/8/2025	5	105	$10,474,532.26

We selected three line items and compared their prices with the estimated prices obtained from the visualization of the historical data in the tool. We also compared the values with the estimated values from the developed models. The values are shown in **Table 31**.

Table 31 Comparison of engineer's estimate of example line items used in the project 2054-01-018

Item	Bid Quantity	Actual Bid Cost	TxDOT Engineer's Estimate Used in this Project	Predicted Value of Engineer's Estimate using the Visualization tool	Predicted Engineer's Estimate by the Statistical Model	Predicted Engineer's Estimate by ML Model
1007001: EXCAVATION (ROADWAY)(CY)	13224	$17.00	$16.00	$12.00	$13.67	$20.12
1347004: BACKFILL (TY A OR B) (STA)	361.62	$250	$300.00	$298.30	$304.47	$258
5057002: TMA (MOBILE OPERATION))(DAY)	250	$110.00	$250.00	$250.00	$203.00	$266.03

A detailed explanation of the price estimation using the visualization tool is provided in **Figures 13**, **14**, and **15**.

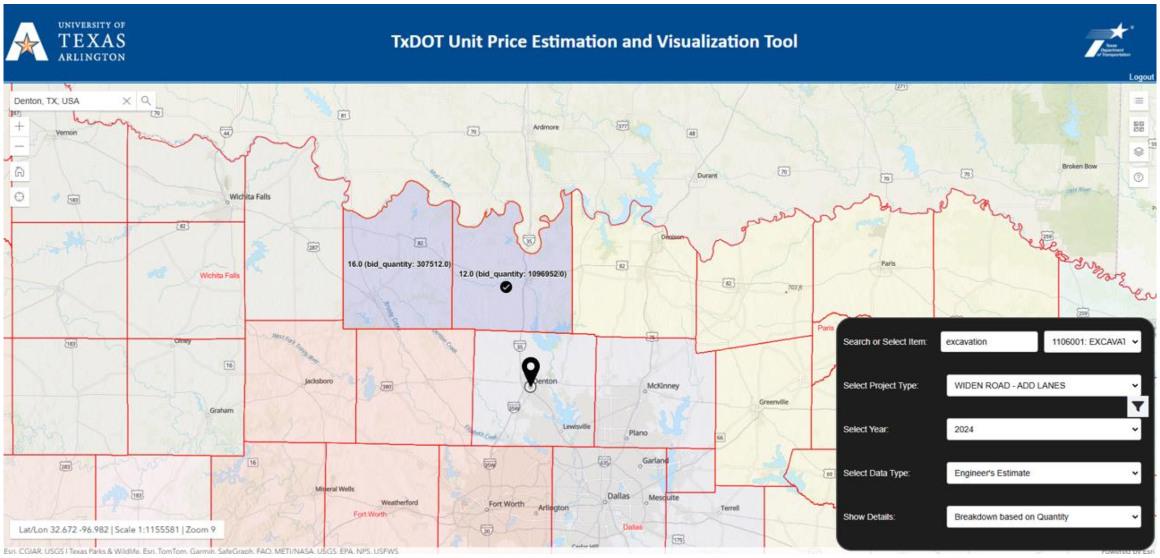


Figure 13 The estimation of unit price of EXCAVATION (ROADWAY)(CY) for the project 2054-01-018 using the GIS-based visualization tool

The unit price of EXCAVATION (ROADWAY)(CY) wasn't available for Denton County, so the value from the nearest neighboring county, Cooke County (marked by ✔), was taken. The predicted unit price from the tool is $12, while the value used in this project was $16.

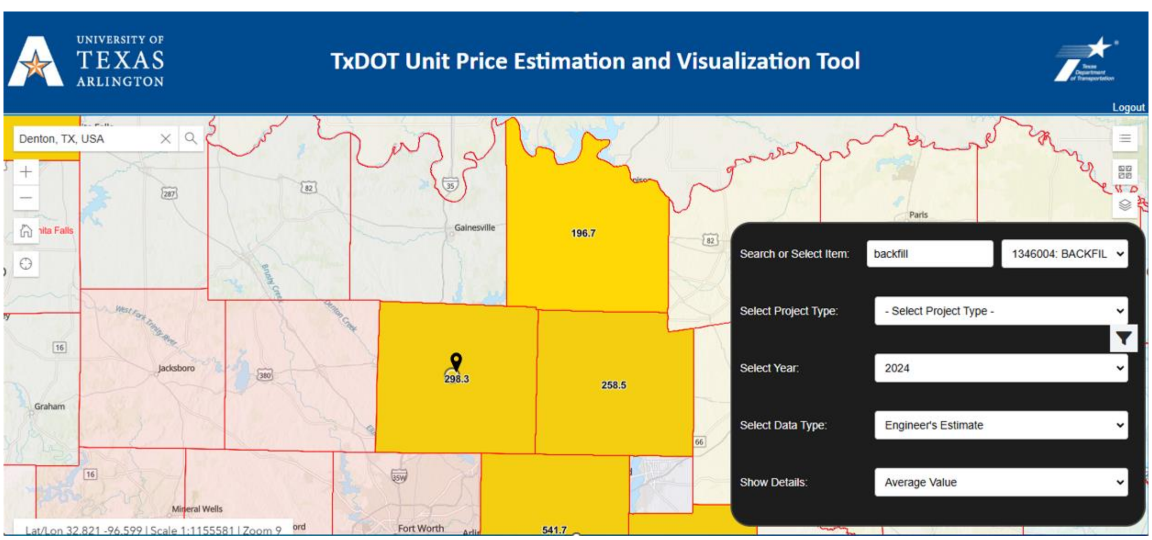


Figure 14 The estimation of unit price of BACKFILL (TY A OR B)(STA) for the project 2054-01-018 using the GIS-based visualization tool

The unit price of BACKFILL (TY A OR B)(STA) was available for Denton County, but the criteria for the project type were not met. Hence, we took the average value, i.e., 298.3$ in that county, while the value used in this project was $300.

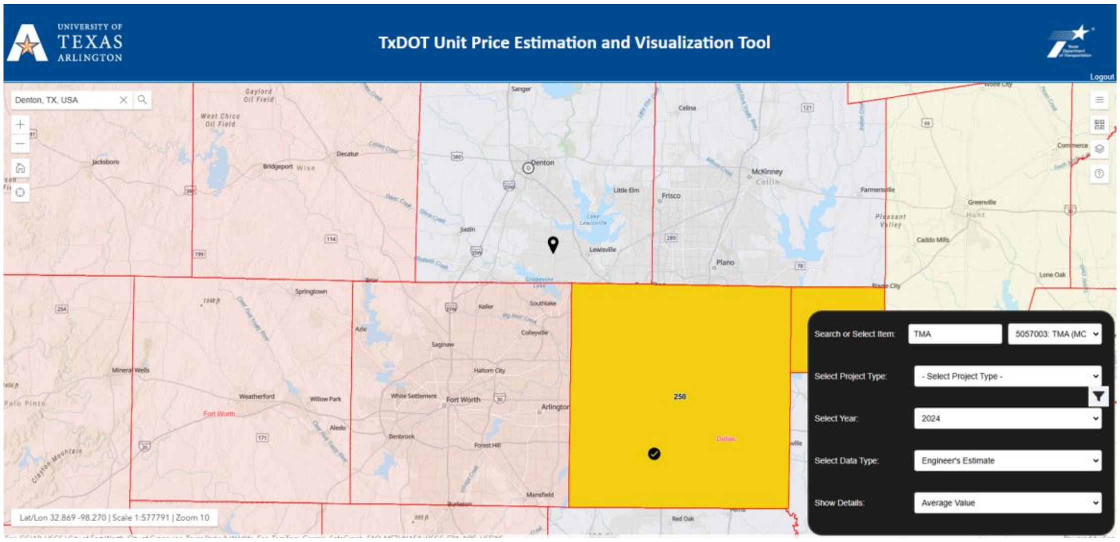


Figure 15 The estimation of unit price of TMA (MOBILE OPERATION)(DAY) for the project 2054-01-018 using the GIS-based visualization tool

The unit price of TMA (MOBILE OPERATION)(DAY) wasn't available for Denton County, and the criterion for project type wasn't met. Hence, the average value from the nearest neighboring county, Dallas (marked by ✔), was taken. The unit price is $250, and the value used in this project was also $250.

7.2.2 CASE 2: PROJECT 0907-13-017

The basic characteristics of the project 0907-13-017 are summarized in **Table 32**.

Table 32 Summary of the project 0907-13-017

CCSJ (Construction Control Section job)	Location	Project Type	Let Date	Number of Bidders	Number of Line Items	Total Estimated Cost of the Project
0907-13-017	Runnels County	BRIDGE REPLACEMENT	1/8/2025	8	49	$2,664,343.88

We selected two line items and compared their prices with the estimated prices obtained from the visualization of the historical data in the tool. We also compared the values with the estimated values from the developed models. The values are shown in **Table 33**.

Table 33 Comparison of engineer's estimate of example line items used in the project 2054-01-018

Item	Bid Quantity	Actual Bid Cost	TxDOT Engineer's Estimate Used in this Project	Predicted Value of Engineer's Estimate using the Visualization tool	Predicted Engineer's Estimate by the Statistical Model	Predicted Engineer's Estimate by ML Model
5027001: BARRICADES, SIGNS AND TRAFFIC HANDLING (MO)	22	$5,000	$9,609.00	$10,000.00	$7,542.05	$6,916.91
1007001: EXCAVATION (ROADWAY) (CY)	312	$25	$45.00	$25.00	$25.70	$34.63

A detailed explanation of the price estimation using the visualization tool is provided in **Figures 16** and **17**.

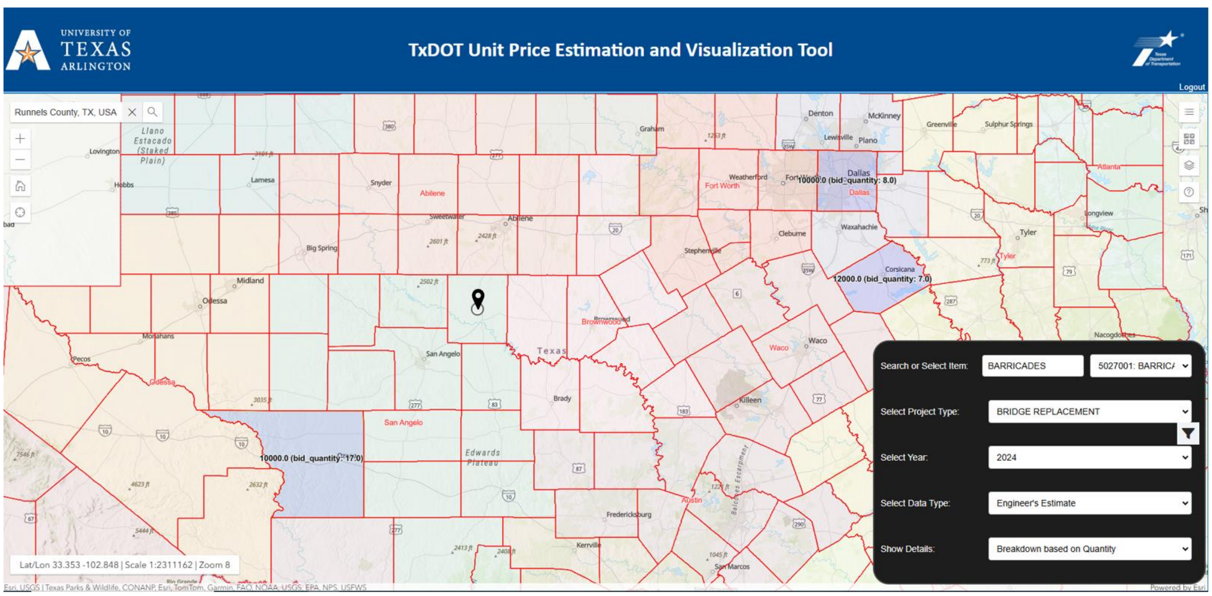


Figure 16 The estimation of the unit price of BARRICADES, SIGNS AND TRAFFIC HANDLING (MO) for the project 0907-13-017 using the GIS-based visualization tool

The unit price of BARRICADES, SIGNS AND TRAFFIC HANDLING (MO) wasn't available for Runnels County. Considering the criteria for project type, time (year), and bid quantity, most counties had the value $10,000, so this value was considered as the predicted value, while the value used in this project was $9609.

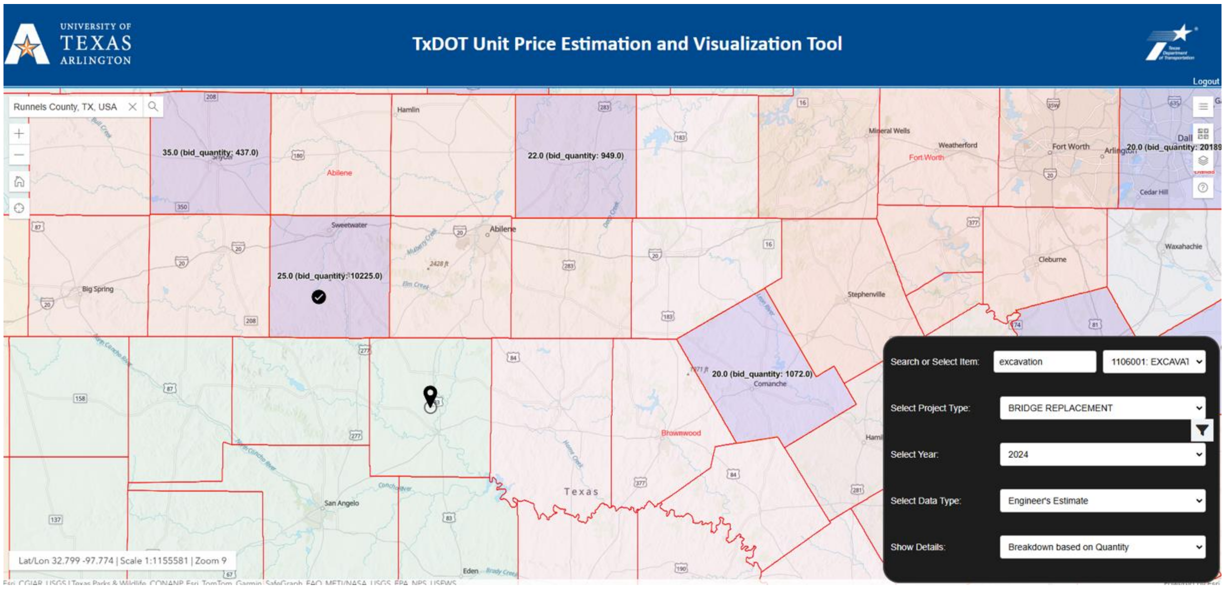


Figure 17 The estimation of unit price of EXCAVATION (ROADWAY)(CY) for the project 0907-13-017 using the GIS-based visualization tool

The unit price of EXCAVATION (ROADWAY)(CY) was not available in Runnels County, so the unit price value from the nearest neighboring county, Nolan (marked by ✔), was taken, which is $25, while the value used in this project was $45.

7.2.3 CASE 3: PROJECT 2208-01-071

The basic characteristics of Project 2208-01-071 are summarized in **Table 34.**

Table 34 Summary of the project 2208-01-071

CCSJ (Construction Control Section job)	Location	Project Type	Let Date	Number of Bidders	Number of Line Items	Total Estimated Cost of the Project
2208-01-071	Tarrant County	BRIDGE MAINTAINANCE	01/08/2025	7	97	$4,810,282

We selected three line items and compared their prices with the estimated prices obtained from the visualization of the historical data in the tool. We also compared the values with the estimated values from the developed models. **Table 35** compares price data used in the project with unit prices obtained from the tool.

Table 35 Price comparison of example line items used in the project 2208-01-071

Item	Bid Quantity	Actual Bid Cost	TxDOT Engineer's Estimate Used in this Project	Predicted Value of Engineer's Estimate using the Visualization tool	Predicted Engineer's Estimate by the Statistical Model	Predicted Engineer's Estimate by ML Model
4027001: TRENCH EXCAVATION PROTECTION(LF)	390	$25	$22.00	$25.00	$14	$23.72
1687001: VEGETATIVE WATERING (MG)	562	$60	$26	$23.75	$23	$28.98
5057001: TMA (STATIONARY) (DAY)	16	$150	$175.00	$250.00	$182	$194.56

A detailed explanation of the price estimation using the visualization tool is provided in **Figures 18**, **19**, and **20**.

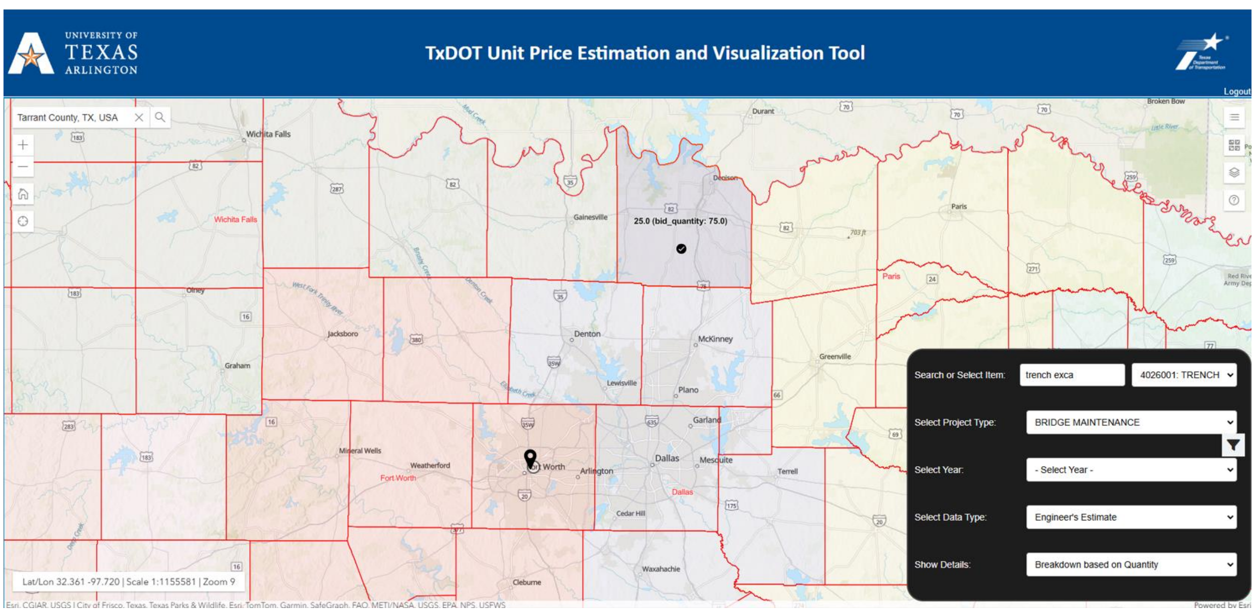


Figure 18 The estimation of unit price of TRENCH EXCAVATION PROTECTION(LF) for the project 2208-01-071 using the GIS-based visualization tool

The unit price of TRENCH EXCAVATION PROTECTION(LF) wasn't available in Tarrant County, so the value from the nearest neighboring county, Grayson County (marked by ✔), was taken, which is $25, while the value used in this project was $22.

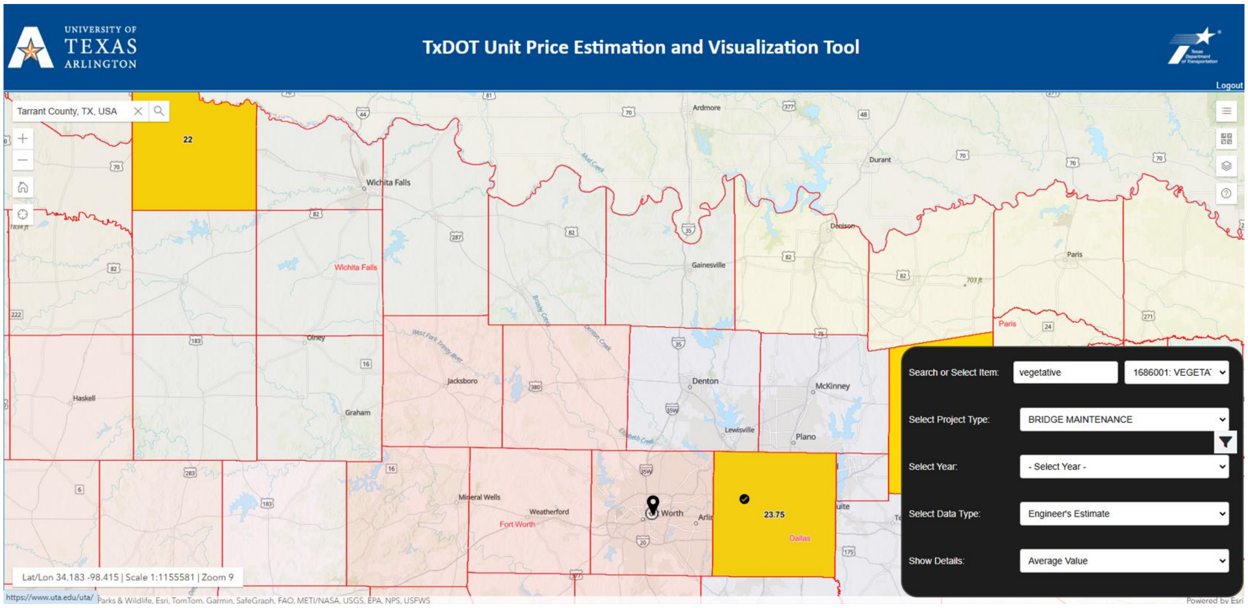


Figure 19 The estimation of unit price of VEGETATIVE WATERING (MG) for the project 2208-01-071 using the GIS-based visualization tool

The unit price of VEGETATIVE WATERING (MG) wasn't available in Tarrant County, so the nearest neighboring county, Dallas, was considered as a reference for price estimation. There were two values in Dallas, so the average value (\$23.75) was taken (marked by ✔). The value used in this project was \$26.

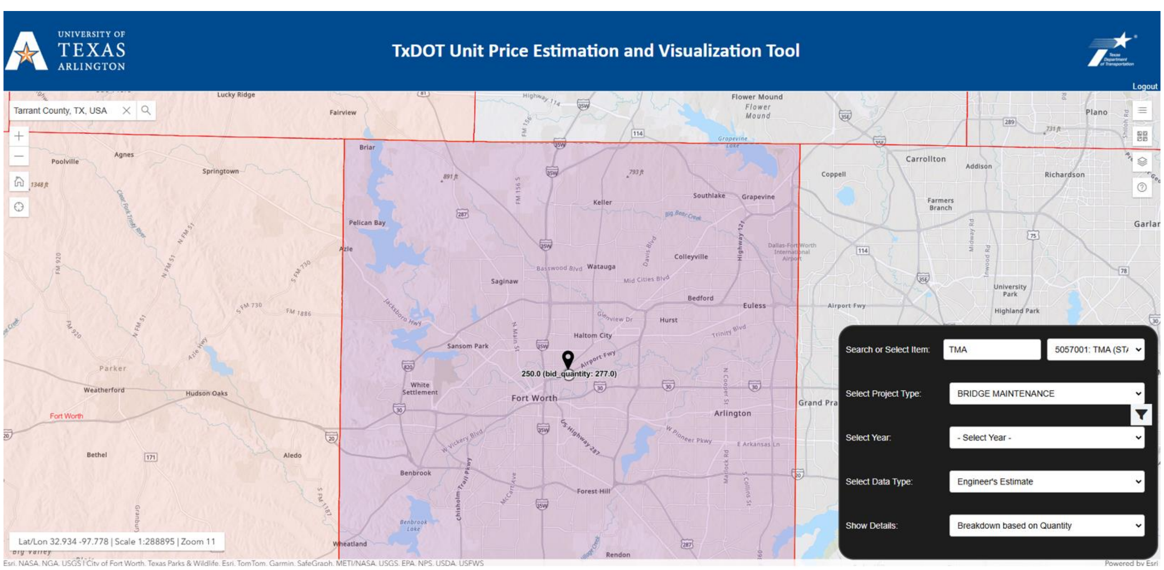


Figure 20 The estimation of unit price of TMA STATIONARY(DAY) for the project 2208-01-071 using the GIS-based visualization tool

The unit price of TMA STATIONARY(DAY) was available on Tarrant County, which is $250. However, since this unit price value isn't for the year 2020 and the bid quantity 16 or around 16, it is more than the estimate used in the project which was $175.

7.2.4 CASE 4: PROJECT 0380-09-104

The basic characteristics of Project 0380-09-104 are summarized in **Table 36.**

Table 36 Summary of the project 0380-09-104

CCSJ (Construction Control Section job)	Location	Project Type	Let Date	Number of line items	Total Estimated Cost of the Project
0380-09-104	Midland County	OVERLAY	to be let in Sep 2025	74	$7,890,199.35

It is to be noted that the project hasn't been let yet. However, we can still compare the unit prices of this project with the latest price data (end of 2024, if available) in the tool and with the estimated values by the statistical and machine learning models. **Table 37** shows some examples of comparisons of unit prices.

Table 37 Price comparison of example line items used in the project 0380-09-104

Item	Bid Quantity	Actual Bid Cost	TxDOT Engineer's Estimate Used in this Project	Predicted Value of Engineer's Estimate using the Visualization tool	Predicted Engineer's Estimate by the Statistical Model	Predicted Engineer's Estimate by ML Model
1347002: BACKFILL (TY B) (STA)	266	$635	$175	$282.50	$261.82	$244.43
5027001: BARRICADES, SIGNS AND TRAFFIC HANDLING (MO)	7	$13,600	$15,000	$15,000	$12,483	$9,433.57
5057001: TMA (STATIONARY) (DAY)	236	$165	$200	$250.00	$202.60	$204.22
5057003: TMA (MOBILE OPERATION) (DAY)	60	$935	$300	$254.18	$569.75	$272.35

A detailed explanation of the price estimation using the visualization tool is provided in **Figures 21, 22**, **23**, and **24**.

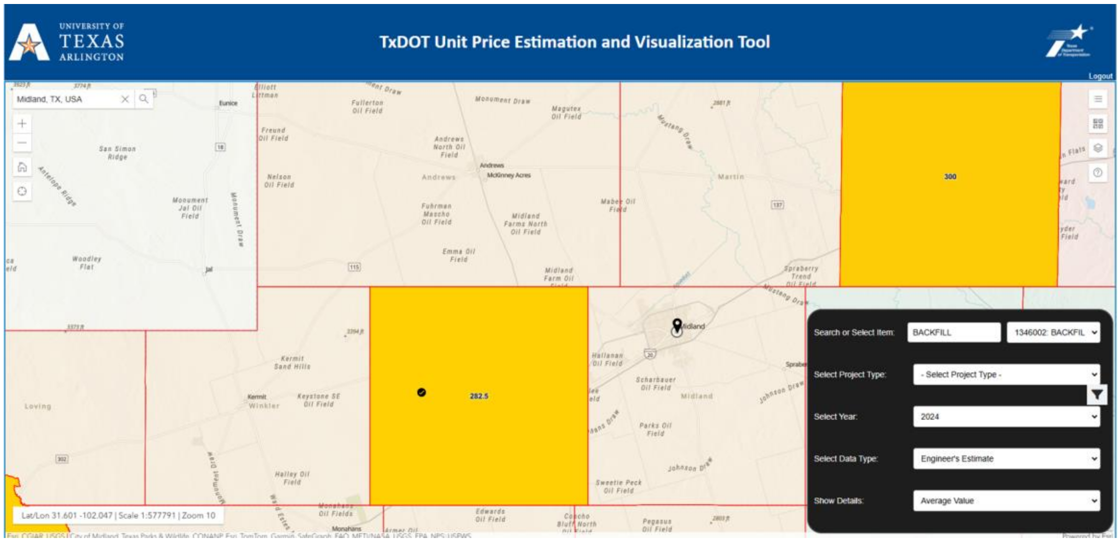


Figure 21 The estimation of unit price BACKFILL (TY B)(STA) for the project 0380-09-104 using the GIS-based visualization tool

The unit price of BACKFILL (TY B)(STA) wasn't available for Midland County, and the criterion for project type wasn’t met. Hence, the average value from the nearest neighboring county, Ector (marked by ⬤), was taken. The unit price value is $282.5, while the value used in this project was $175.

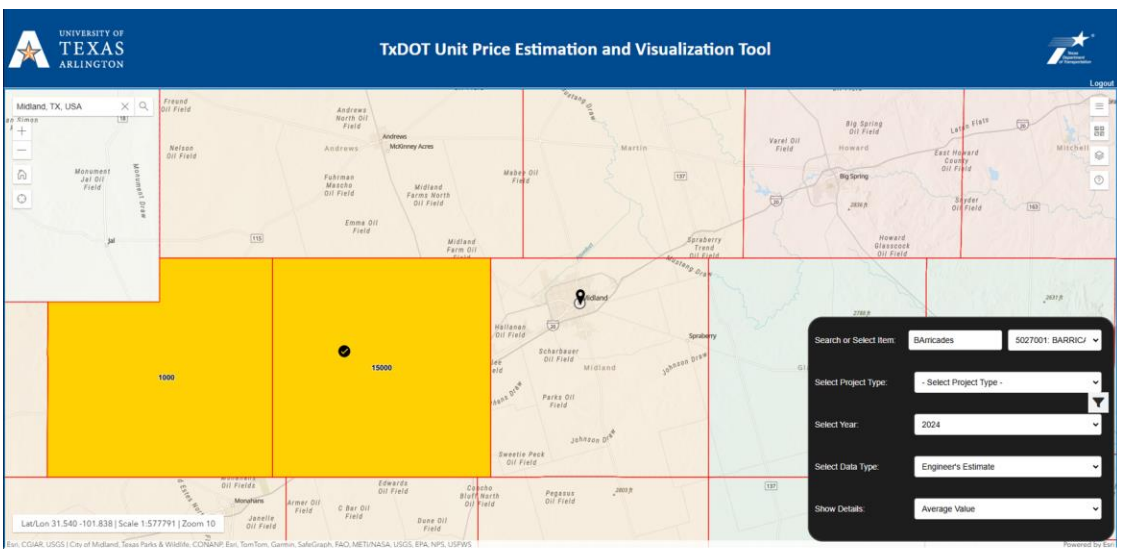


Figure 22 The estimation of unit price BARRICADES, SIGNS AND TRAFFIC HANDLING(MO) for the project 0380-09-104 using the GIS-based visualization tool

The unit price of BARRICADES, SIGNS AND TRAFFIC HANDLING(MO) wasn't available for Midland County, and the criterion for project type wasn't met. Hence, the average value from the nearest neighboring county, Ector (marked by ⊘), was taken. The unit price value is $15000, and the value used in this project was also $15000.

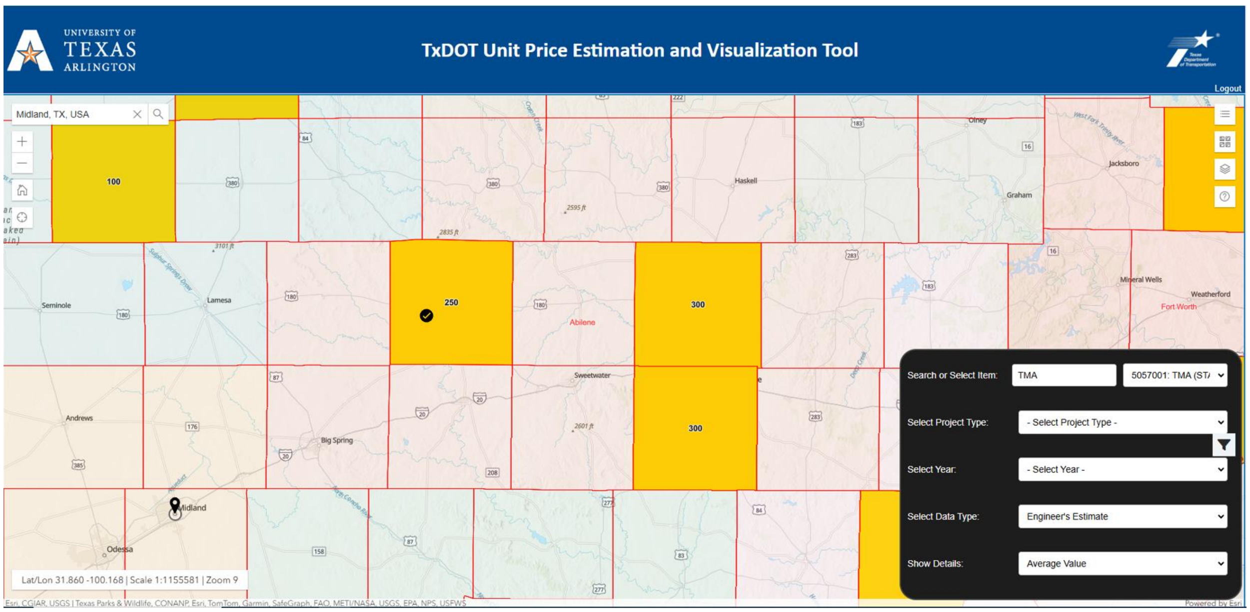


Figure 23 The estimation of unit price TMA (STATIONARY)(DAY) for the project 0380-09-104 using the GIS-based visualization tool

The unit price of TMA (STATIONARY)(DAY) wasn't available for Midland County, and the criterion for project type wasn't met. Hence, the average value from the nearest neighboring county, Scurry (marked by ⊘), was taken. The unit price value is $250, while the value used in this project was $200.

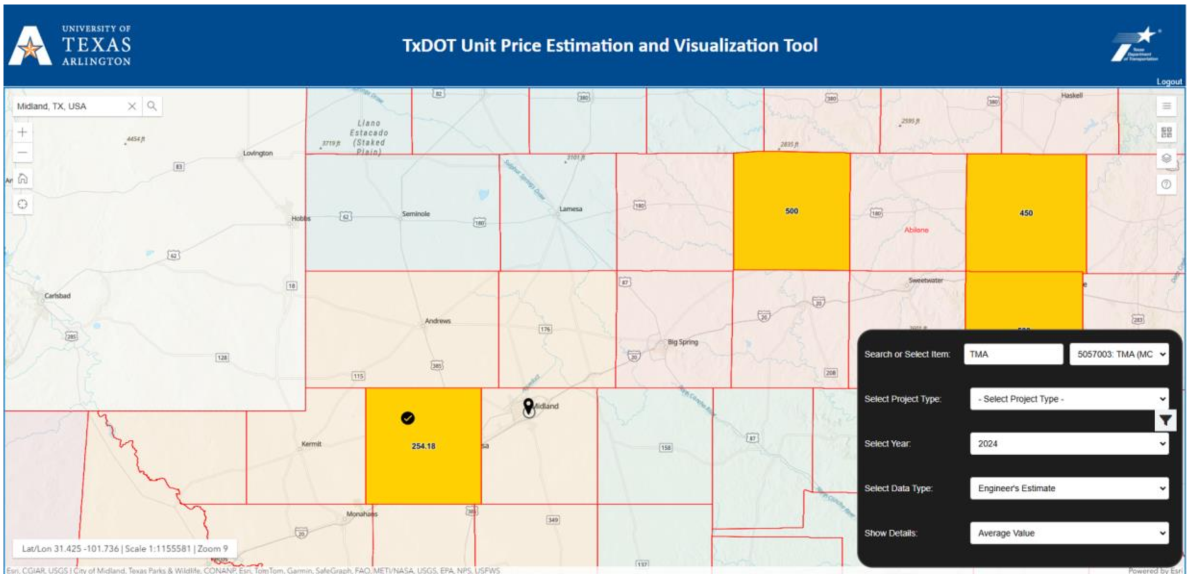


Figure 24 The estimation of unit price TMA (MOBILE OPERATION)(DAY) for the project 0380-09-104 using the GIS-based visualization tool

The unit price of TMA (MOBILE OPERATION)(DAY) wasn't available for Midland County, and the criterion for project type wasn't met. Hence, the average value from the nearest neighboring county, Ector (marked by ✔), was taken. The unit price value is $254.18, while the value used in this project was $300.

7.2.5 CASE 5: PROJECT 0022-01-034

The basic characteristics of the project 0022-01-034 are summarized in **Table 38.**

Table 38 Summary of the project 0022-01-034

CCSJ (Construction Control Section job)	Location	Project Type	Let Date	Number of Line Items	Total Estimated Cost of the Project
0022-01-034	Terrell County	BRIDGE REPLACEMENT	January 2026	34	$8,776,758

Table 38 shows that the project will be let in 2026, so we don't have actual price data to use as a comparison reference in the tool. However, we can use the latest values for comparison (price data for 2024, if available). We also don't have actual bid price data.**Table 39** shows examples of comparing the price data used in the project with the values estimated using the tool. We also compared the values with the estimated values from the developed models.

Table 39 Price comparison of example line items used in the project 0022-01-034

Item	**Bid Quantity**	**TxDOT Engineer's Estimate Used in this Project**	**Predicted Value of Engineer's Estimate using the Visualization Tool**	**Predicted Engineer's Estimate by the Statistical Model**	**Predicted Engineer's Estimate by ML Model**
5027001: BARRICADES, SIGNS AND TRAFFIC HANDLING(MO)	15	$20,000	$10,000	$19,387	$6621.42
5057001: TMA (STATIONARY) (DAY)	440	$280	$250	$281	$265.40
5057003: TMA (MOBILE OPERATION) (DAY)	16	$500	$350	$343	$297.77
1007001: EXCAVATION (ROADWAY)(CY)	890	$20	$30	$16	$22.22

A detailed explanation of the price estimation using the visualization tool is provided in **Figures 25, 26**, **27**, and **28**.

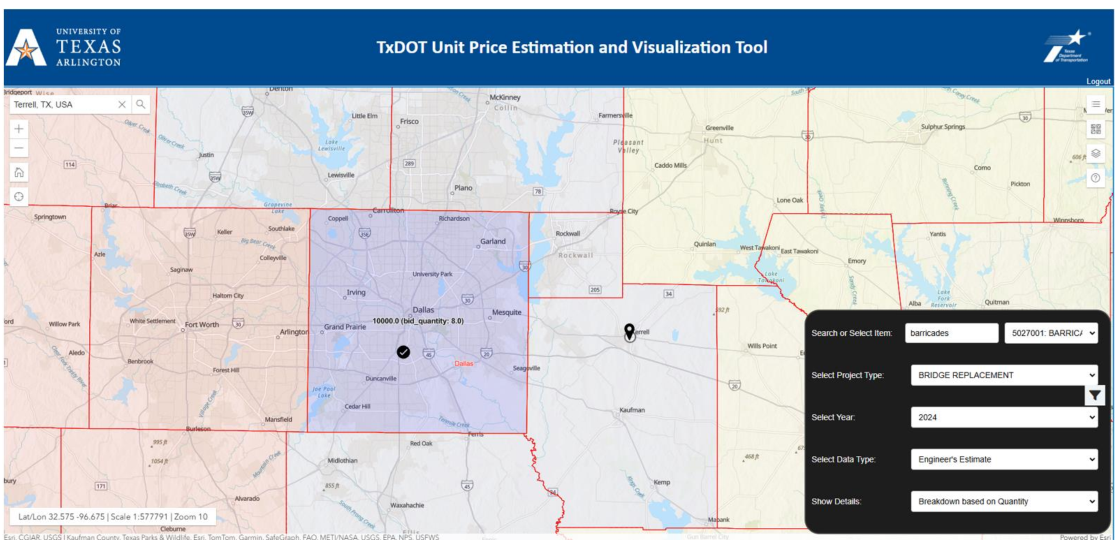


Figure 25 The estimation of unit price BARRICADES, SIGNS AND TRAFFIC HANDLING(MO) for the project 0022-01-034 using the GIS-based visualization tool

The unit price of BARRICADES, SIGNS AND TRAFFIC HANDLING(MO) wasn't available for Terrell County. Hence, the value from the nearest neighboring county, Dallas (marked by ●), was used as a reference for estimation. The value is $10,000, while the value used in this project was $20000.

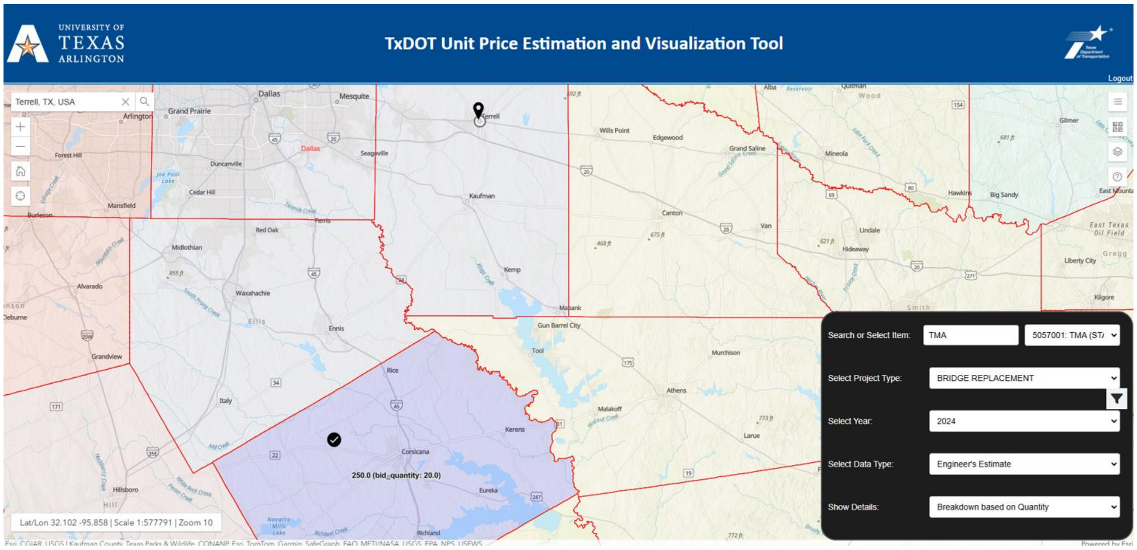


Figure 26 The estimation of unit price TMA (STATIONARY)(DAY) for the project 0022-01-034 using the GIS-based visualization tool

The unit price of TMA (STATIONARY)(DAY) wasn't available for Terrell County. Hence, the value from the nearest neighboring county, Navarro (marked by ✔), was used as a reference for estimation. The value is $250, while the value used in this project was $280.

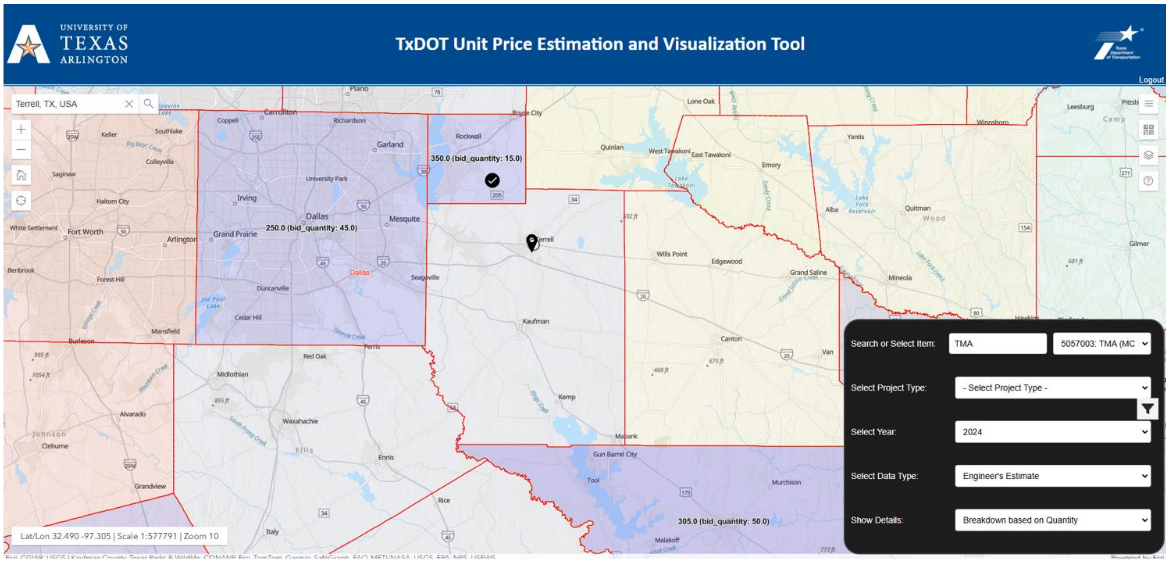


Figure 27 The estimation of unit price TMA (MOBILE OPERATION)(DAY) for the project 0022-01-034 using the GIS-based visualization tool

The unit price of TMA (MOBILE OPERATION)(DAY) wasn't available for Terrell County. Hence, the value from the nearest neighboring county, Rockwall (marked by ⊙), was used as a reference for estimation. This county was selected among all the neighboring counties because of the similarity in bid quantity used in the project. The value is $350, while the value used in this project was $500.

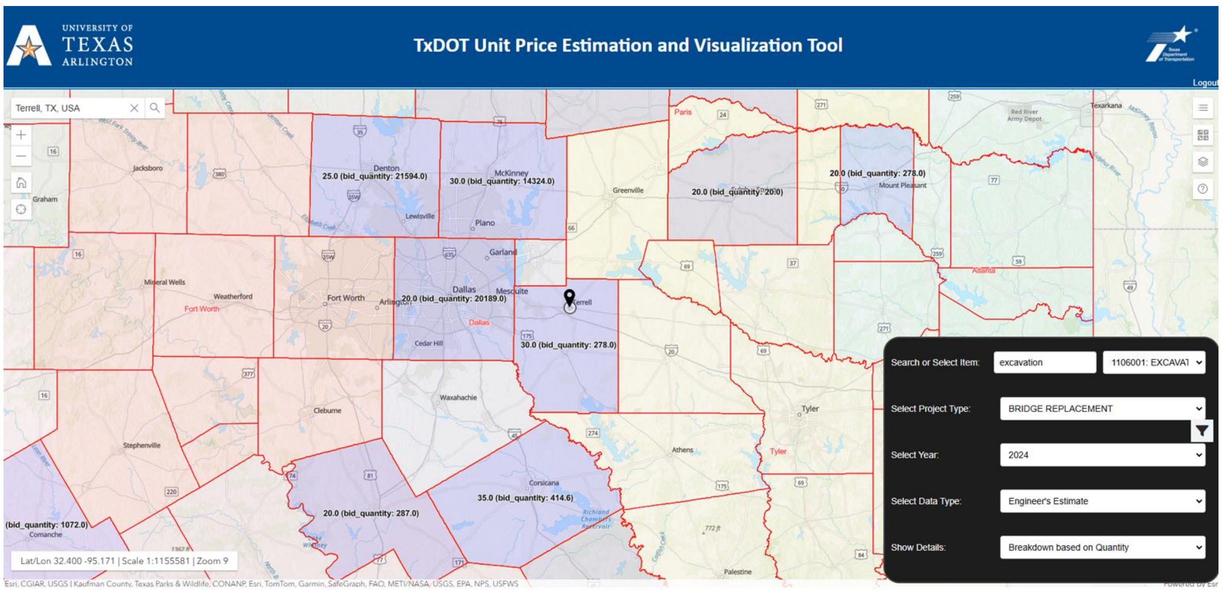


Figure 28 The estimation of unit price EXCAVATION (ROADWAY)(CY) for the project 0022-01-034 using the GIS-based visualization tool

The unit price of EXCAVATION (ROADWAY)(CY) was available for Terrell County. However, the quantity match wasn't exact, so the value from the tool was more than the one used in this project. We tried to adjust the price for the quantity, but the relationship between the price and the quantity wasn't consistent when we analyzed price data in other counties. Hence, we used the value $30, while the value used in this project was $20.

7.2.6 CASE 6: PROJECT 6435-42-001

The basic characteristics of Project 6435-42-001 are summarized in **Table 40**.

Table 40 Summary of the project 6435-42-001

CCSJ (Construction Control Section job)	Location	Project Type	Let Date	Number of Line Items	Total Estimated Cost of the Project
6435-42-001	San Angelo	Culvert Lengthening and Replacement	2025	46	$456,935

Table 40 shows that the project was let in 2025, so we don't have actual price data to use as a comparison reference in the tool. However, we can use the latest values for comparison (price data for 2024, if available). We also don't have bid price data available. **Table 41** shows examples of comparing the price data used in the project with the values estimated using the tool.

Table 41 Price comparison of example line items used in the project 6435-42-001

Item	Bid Quantity	Predicted Value of Engineer's Estimate using the visualization tool	TxDOT Engineer's Estimate Used in this Project	Predicted Engineer's Estimate by the Statistical Model	Predicted Engineer's Estimate by ML Model
4027001: TRENCH EXCAVATION PROTECTION (LF)	28	$45.57	$50	$56	$52.88
5057001: TMA STATIONARY (DAY)	9	$280	$300	$299	$269.63

A detailed explanation of the price estimation using the visualization tool is provided in **Figures 29** and **30.**

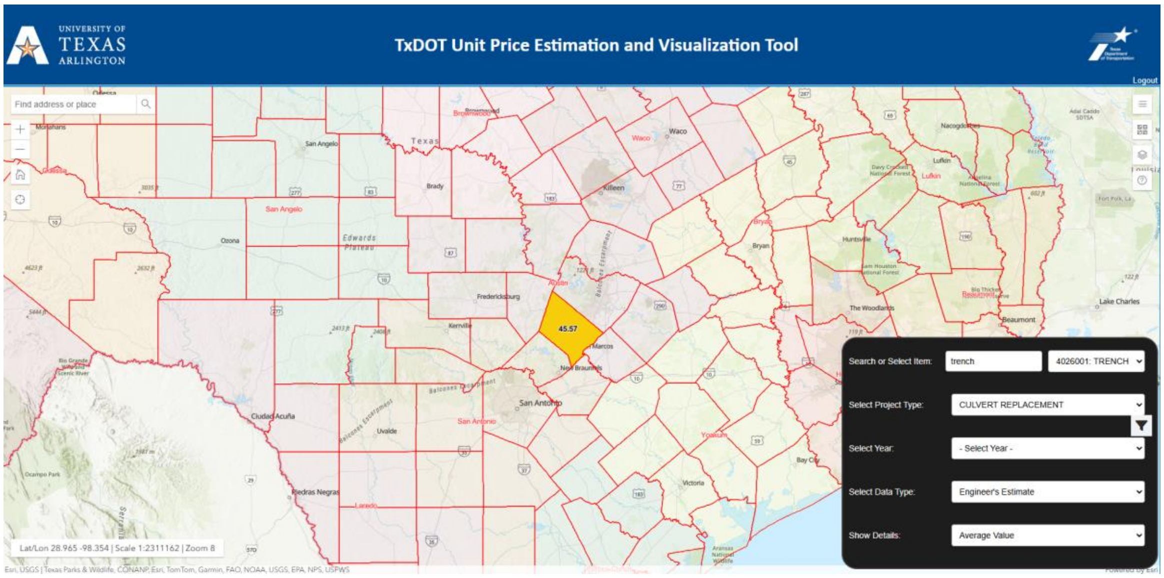


Figure 29 The estimation of unit price TRENCH EXCAVATION PROTECTION (LF) for the project 6435-42-001 using the GIS-based visualization tool

The unit price of TRENCH EXCAVATION PROTECTION (LF) wasn't available for San Angelo. Hence, the value from the nearest neighboring county was used as a reference for estimation. The value is $45.57.

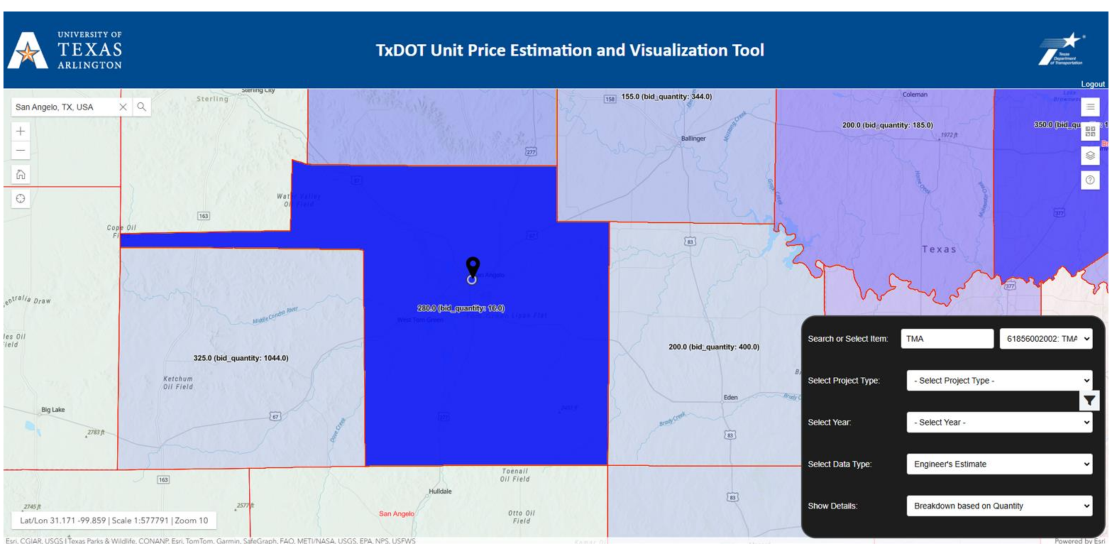


Figure 30 The estimation of unit price TMA STATIONARY (DAY) for the project 6435-42-001 using the GIS-based visualization tool

The unit price of TMA STATIONARY (DAY) was available for San Angelo, so we estimated the value $280, while the value used in this project was $300.

CHAPTER 8. TECHNOLOGY TRANSFER

8.1. INTRODUCTION

Technology transfer refers to the formal handover of the developed web-based visualization tool and the unit price estimation models to the Texas Department of Transportation (TxDOT). This chapter describes the delivered components of the GIS-based visualization tool and the models, along with the procedures and considerations necessary for their successful implementation on a cloud platform or a computing environment.

8.2. DELIVERED COMPONENTS OF THE DEVELOPED GIS-BASED TOOL AND MODELS

This section details the codebase and files that were handed over to TxDOT for the successful implementation of the GIS-based visualization tool and the models.

8.2.1. DELIVERED COMPONENTS OF THE GIS-BASED TOOL

The tool facilitates the visualization of the unit price of construction line items, illustrated for the data collected from TxDOT projects (2022 to 2024). The tool allows users to filter the unit price based on different criteria related to the line items (such as quantity). Since the tool is GIS-based, the database is submitted as a geodatabase file (GIS_Data.gdb). Then, the codebase for developing the tool was zipped into a file (0-7184_webapp_ historical_price_code.zip) and handed over to TxDOT.

8.2.2. DELIVERED COMPONENTS OF ESTIMATION MODELS

Machine learning models (DNN models and Ensemble models) and statistical models (Mixed-effects Models) were developed for estimating the unit prices and illustrated for estimating the top 8 most commonly used construction line items for the past two years (2022 to 2024). The components of the developed models include a database, codebase, and files with predicted values. There are two different folders for each machine learning model (DNN and Ensemble models). In each folder, there are nine folders where the database, codebase, and prediction files are saved. Similarly, for the statistical models (Mixed-effects Model), the database, codebase, and prediction file are saved under the folder for the mixed-effects model.

Delivered Database of Estimation Models

A dataset (final_df_8.xlsx) containing the top 8 most commonly used construction line items was created, which is a subset of the 2-year dataset (1_dataset.xlsx). Both of these Excel files were transferred, while the one used for developing both the machine learning models and statistical models is the dataset for the top eight line items (final_df_8.xlsx).

Delivered Codebase and Generated Prediction of Machine Learning Models

The codebase (in Python) and prediction files (in Excel) were provided for all eight line items for each model. For running the model for each item, a Python file (.ipynb) was handed over. Then, a Keras file was handed over for saving the model in a Python file. Finally, the saved predicted values were provided in an Excel file (.xlsx) for each model.

Delivered Codebase and Generated Prediction of the Mixed Effects Model

For preparing the database before running the model in STATA, a codebase (prep_regression_data.do) was handed over. Then, another STATA code (mixed_effect_model.do) was provided for running the model. This code generates seven models, and model 5 is concluded to be the best model among those models. Hence, for generating prediction values, model 5 was used. A codebase (model5_prediction_eng_est.do) for generating the predicted values and an Excel file (model5_predictions_eng_est.xlsx) with the saved predicted values were handed over.

8.3. PROCEDURES AND CONSIDERATIONS FOR THE IMPLEMENTATION OF THE DEVELOPED GIS-BASED TOOL

This section details the steps to be followed for the integration of the tool into a cloud platform and for running the model in a computing environment.

8.3.1. STEPS FOR INTEGRATION OF THE GIS-BASED TOOL

The following steps detail the measures to be followed for the integration of the tool into the TxDOT cloud platform:

- Open the folder (GIS_Data.gdb) in ArcGIS Pro and upload the geodatabase file (GIS_data) into the folder. Then, share the uploaded layer as a feature layer in ArcGIS Online. The folder (GIS_Data.gdb) has 50 companion files.

- Unzip the file (0-7184_webapp_ historical_price_code.zip) and upload it to the TxDOT cloud platform. In the zipped files, there are CSS files, PHP files, and an HTML file (with JavaScript). The CSS files help with the design of the web interface layout. The PHP files control the registration of user accounts and their login to and logout from the app. The HTML file has a script container that houses most of the display codes and links to the ArcGIS Online web layer. The link should be updated to point to the web layer created in the previous step.

8.3.2. STEPS FOR RUNNING MACHINE LEARNING MODELS

There are two separate folders for two machine learning models (DNN Model and Ensemble Models). Each folder has a codebase for all eight line items, including a multi-task model. All the files follow consistent naming, so the steps to run the model are the same for all the items for both models. The following steps detail the measures to be followed for running the machine learning models for one of the line items: EXCAVATION.

- A dataset with the top 8 most common line items (final_df_8.xlsx) should be developed, which is a subset of the 2-year dataset (1_dataset.xlsx). However, since both of these files are provided, one can directly use the dataset (final_df_8.xlsx) for running the model.
- Run the model with Python code (Excavation.ipynb).
- Save the model in a Python file (Excavation.keras).
- Save the prediction values in an Excel file (excavation_eng_est_predictions.xlsx).

8.3.3. STEPS FOR RUNNING MIXED-EFFECTS MODELS

The following steps detail the measures to be followed for running the mixed-effects model.

- A dataset with the top 8 most common line items (final_df_8.xlsx) should be developed, which is a subset of the 2-year dataset (1_dataset.xlsx). However, since both of these files are provided, one can directly use the dataset (final_df_8.xlsx) for running the model.
- Prepare data for modeling with STATA codes (prep_regression_data.do).

- Run the model with STATA codes (mixed_effect_model.do). This code develops seven statistical models.
- Generate predicted values for model 5 with STATA codes (model5_prediction_eng_est.do). Save the predicted values in an Excel file (model5_predictions_eng_est.xlsx). We selected model 5 as the best model since it has the least AIC (Akaike Information Criterion) value.

8.4. TECHNICAL DOCUMENTATION

This subsection summarizes the documentation provided to guide the users to implement and use the developed GIS-based visualization tool and to run the models. A presentation (0-7184_June2025_Tool_Transfer_Meeting_Final.pptx) was prepared and handed over, where the names of the delivered files are clearly mentioned along with the overview of the steps to be followed for implementation. Moreover, Educational Material (0-7184_April2025_EM.pptx), Video Training Material (0-7184_April2025_VTM.mp4), and User manual (0-7184_April2025_Usermanual.docx) were developed for TxDOT staff to learn how to estimate the unit price of each work item, visualize unit prices, and analyze the results using the GIS-based visualization tool.

CHAPTER 9. SUMMARY AND CONCLUSION

The goal of this project (TxDOT 0-7184) was to develop spatiotemporal models for estimating unit prices of construction line items for TxDOT projects and create a Unit Price Estimation and Visualization Tool (UPEVT) to visualize unit price data across Texas counties.

The unit price of a work item is heavily affected by various project-specific and external factors, including but not limited to the project location, quantity of the work, project duration, time factors, site conditions, market conditions, and macroeconomic conditions. Moreover, unit prices of work items are subject to significant variations from project to project and over time. Hence, firstly, the identification of potential factors that affect unit prices is necessary, and it was achieved through a literature review. The practices and recommendations from several State Departments of Transportation (State DOTs) regarding adjusting unit prices considering various factors were reviewed as well. Texas Department of Transportation (TxDOT) recommends adjusting unit prices for work quantity, project type, site conditions, and inflation, but hasn't mentioned clearly their process for adjusting the unit prices.

After identifying potential factors that affect unit prices, data were collected from publicly available sources. The unit price data were collected for about two years (March 4, 2022, to September 25, 2024), and other data from publicly available resources were created in the same timeframe. Then, a database was created by merging all the collected data. A collection of data from each data source is briefly explained in this report.

Machine learning models (DNN models and Ensemble models) were developed and illustrated for estimating the unit prices of the top eight commonly used construction line items using two years of data (March 2022 to September 2024). The modeling process involved data preprocessing, incorporation of external factors, feature extraction, and training both individual and multi-task models using DNN and Ensemble architectures. The last three months of the dataset were reserved for testing, while the remaining data was used for training to ensure a realistic evaluation of model performance. Results showed that DNN models generally offered better generalization for those line items with stable or less complex patterns, while Ensemble models may better capture temporal dependencies where such patterns exist. Overall, the models provided accurate and robust estimation tools based on the available two-year dataset. Since machine learning models heavily rely on the data used during training, expanding the dataset with more historical records, broader project coverage, and greater variability could further enhance model performance. TxDOT could benefit even more from these models by fine-tuning them with a larger and more diverse dataset.

Statistical models (Mixed-effects Model) were developed for estimating the unit prices of the top eight commonly used construction line items using two years of data (March 2022 to September 2024). The mixed-effects models successfully addressed the multidimensional aspects of the heterogeneity involved in the dataset by utilizing both population-level fixed effects and group-specific random effects and produced more precise predictions. Reliable predictions from mixed-effects models could be observed via case studies. However, the models were developed with only two years of data, so the model performance could be improved with an expanded dataset covering more line items, projects, counties, and historical records.

A GIS-based Unit Price Estimation and Visualization Tool (UTEVT) was developed to visualize unit prices of construction line items across Texas counties. Spatial data such as unit price, quantity, and the Texas county map were collected and stored in a geodatabase. Two years of historical unit price data were visualized using the tool. TxDOT determines unit prices of pay items using the historical bid-based estimation method and then develops an engineer's project appraisal. The historical value of an engineer's estimate is used to assess the bids and select the bidder. Therefore, this tool can help users access historical values in an efficient way. Moreover, a user manual can be accessed through the tool for users to be able to understand the functions of the tool easily.

To demonstrate the application of the developed GIS-based visualization tool, case studies of six projects from different Texas counties were conducted. Since the projects were of different types and let in different counties, the research team could demonstrate the use of the developed tool to analyze and estimate unit prices in different scenarios. In many cases, all criteria for quantity of work item, county, project type, and time couldn't be satisfied. Even so, reasonable estimates could be achieved based on the nearby/relatable criteria.

In addition to the usual manual, Educational Material (EM) and Video Training Material (VTM) were created to help potential users learn how to estimate the unit price of each work item, visualize unit prices, and analyze the results data using the GIS-based visualization tool. A PowerPoint presentation was provided as the EM, and a video was provided as the VTM. These materials helped document the process of estimating unit prices using the developed GIS-based visualization tool.

Finally, the tool and the developed model, database, codebase, and other files were transferred to TxDOT. The components of the delivered products (a tool and models) and their integration and computation procedures were thoroughly documented and presented.

It is expected that this research project's findings will assist TxDOT engineers and managers in decision-making by considering various factors that potentially affect unit prices in estimating unit prices. Moreover, the visualization of unit prices using the GIS-based tool will help them visually access price values in developing estimates.

APPENDIX A

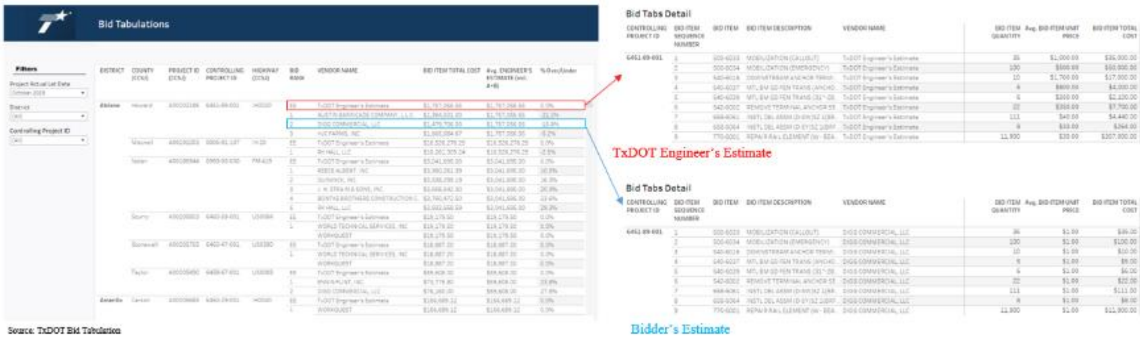


Figure A1: Bid information for TxDOT engineer's estimate and bidders' estimate

Note: TxDOT illustrates bidding information on projects statewide and districtwide along with TxDOT Engineer's Estimate for the past 24 months on the Bid Tabulations dashboard (Workbook: Bid Tabulations (Txdot.Gov) v3.0, 2025).

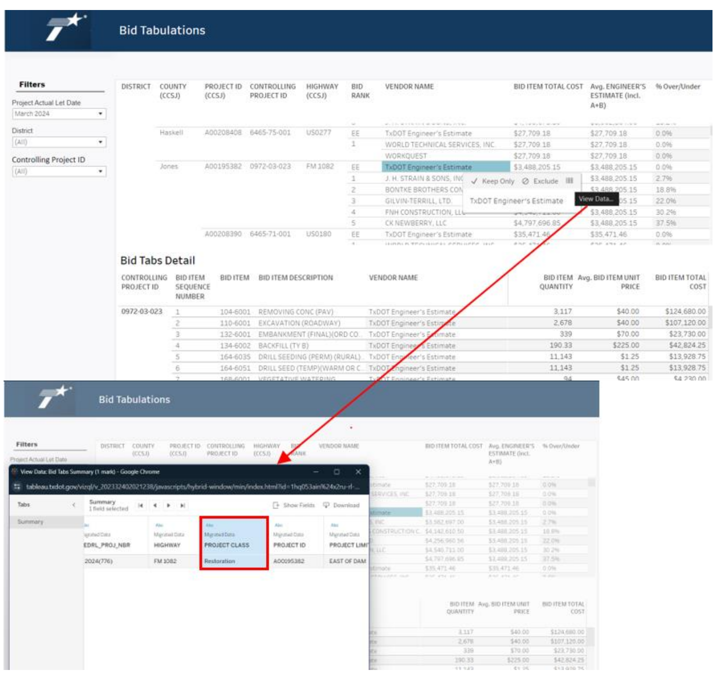


Figure A2: Bid Tabulations dashboard showing details on project type

Note: The Bid Tabulations dashboard displays information on project type under the "PROJECT CLASS" heading.

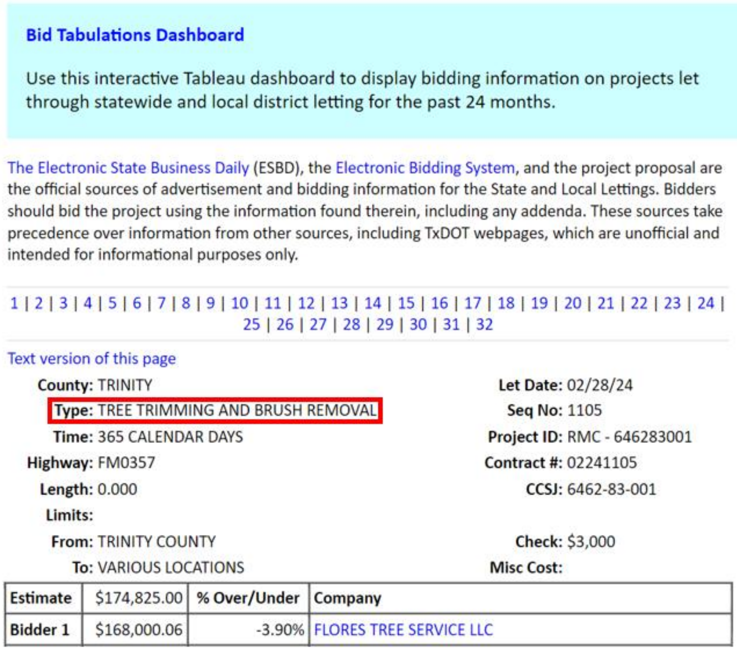


Figure A3: Bid Tabulations dashboard showing details on project duration

Note: The Bid Tabulations dashboard also displays information on project duration.

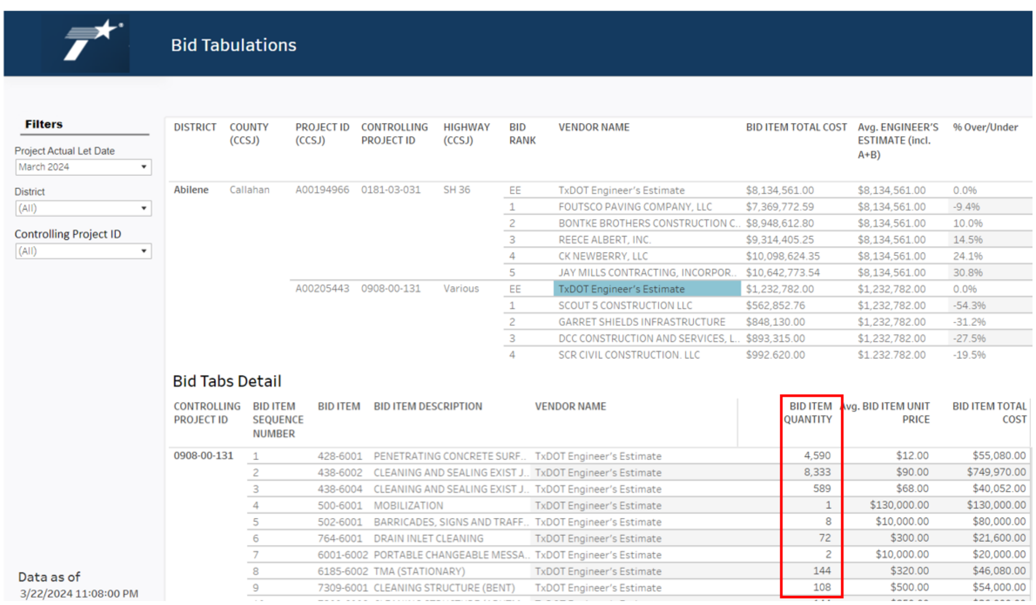


Figure A4: Bid Tabulations dashboard showing details on bid quantity

Note: The Bid Tabulations dashboard also displays information on bid quantity

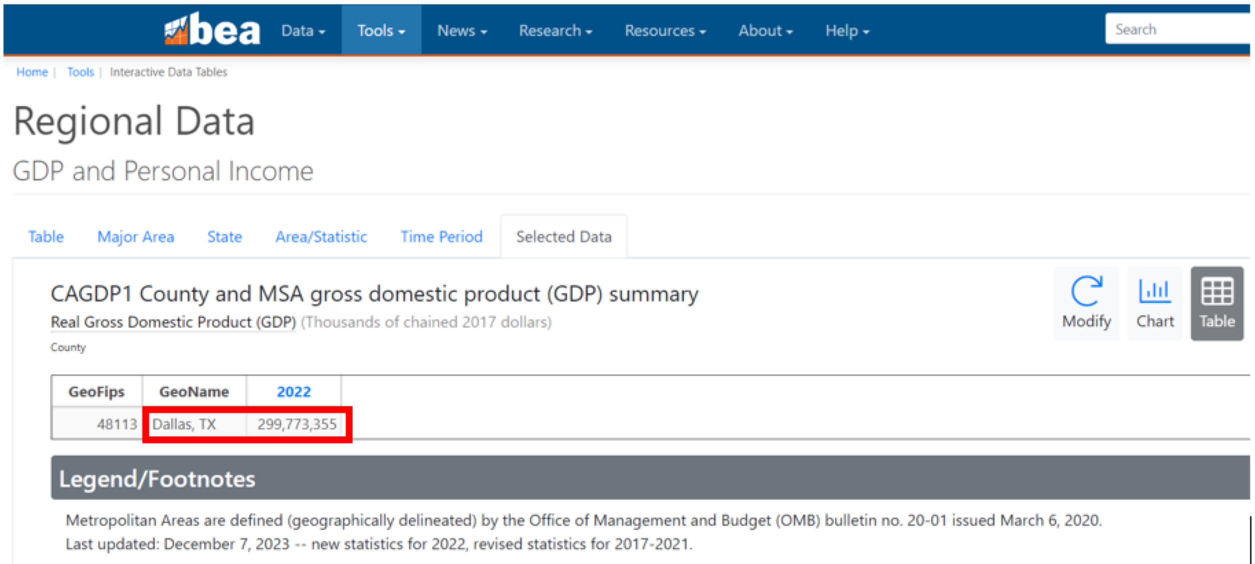
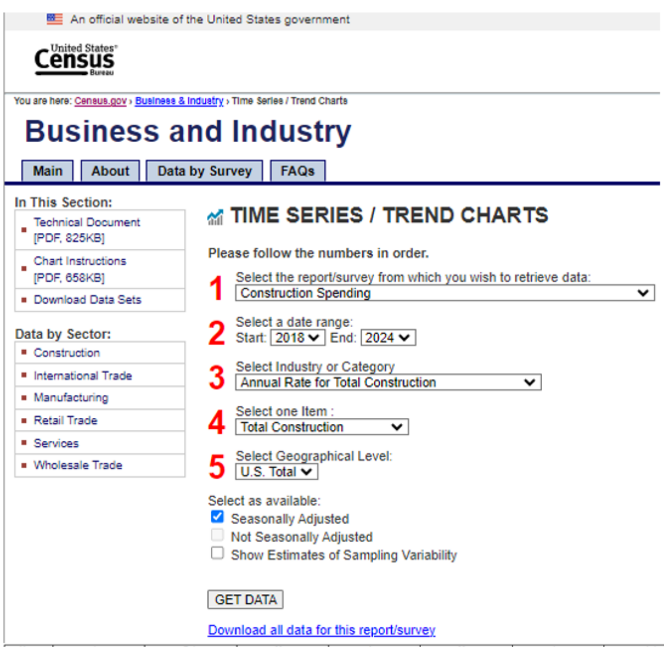


Figure A5: Real Gross Domestic Product of Dallas County in the year 2022 as shown by US Bureau of Labor Statistics

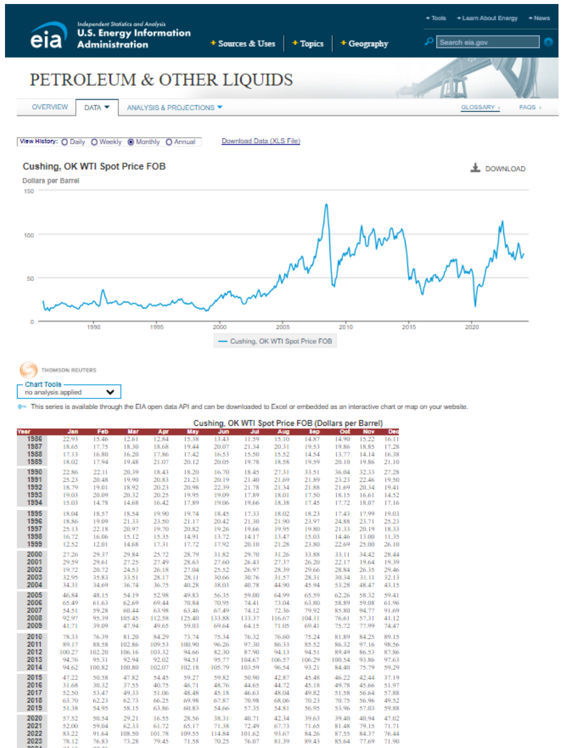
Note: The US Bureau of Labor Statistics provides data on the Gross Domestic Product.



Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2018	1,336,469	1,355,105	1,344,968	1,355,101	1,362,364	1,343,214	1,336,631	1,340,449	1,326,111	1,310,946	1,300,694	1,289,930
2019	1,296,049	1,312,040	1,322,532	1,353,657	1,364,421	1,383,309	1,410,749	1,426,091	1,438,119	1,442,150	1,463,036	1,466,415
2020	1,494,117	1,501,790	1,506,335	1,460,471	1,452,086	1,459,188	1,470,971	1,489,640	1,510,182	1,534,900	1,550,976	1,572,760
2021	1,606,561	1,585,044	1,613,045	1,615,638	1,628,281	1,639,779	1,658,346	1,666,756	1,667,332	1,685,243	1,726,586	1,754,426
2022	1,796,460	1,823,728	1,849,637	1,882,753	1,880,896	1,873,184	1,869,262	1,847,285	1,836,930	1,830,477	1,842,206	1,840,896
2023	1,882,199	1,889,562	1,901,401	1,907,837	1,946,733	1,956,226	1,969,005	2,010,143	2,017,510	2,058,903	2,082,923	2,105,791
2024	2,102,434	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Figure A6: U.S. Total Construction Spending data monthly from 2018 to 2024 as published by United States Census Bureau

Note: United States Census Bureau publishes monthly data on construction spending from January 2002.



Cushing, OK WTI Spot Price FOB (Dollars per Barrel)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1986	22.93	15.46	12.61	12.84	15.38	13.43	11.59	15.10	14.87	14.90	15.22	16.11
1987	18.65	17.75	18.30	18.68	19.44	20.07	21.34	20.31	19.53	19.86	18.85	17.28
1988	17.13	16.80	16.20	17.86	17.42	16.53	15.50	15.52	14.54	13.77	14.14	16.38
1989	18.02	17.94	19.48	21.07	20.12	20.05	19.78	18.58	19.59	20.10	19.86	21.10
1990	22.86	22.11	20.39	18.43	18.20	16.70	18.45	27.31	33.51	36.04	32.33	27.28
1991	25.23	20.48	19.90	20.83	21.23	20.19	21.40	21.69	21.89	23.23	22.46	19.50
1992	18.79	19.01	18.92	20.23	20.98	22.39	21.78	21.34	21.88	21.69	20.34	19.41
1993	19.03	20.09	20.32	20.25	19.95	19.09	17.89	18.01	17.50	18.15	16.61	14.52
1994	15.03	14.78	14.68	16.42	17.89	19.06	19.66	18.38	17.45	17.72	18.07	17.16
1995	18.04	18.57	18.54	19.90	19.74	18.45	17.33	18.02	18.23	17.43	17.99	19.03
1996	18.86	19.09	21.33	23.50	21.17	20.42	21.30	21.90	23.97	24.88	23.71	25.23
1997	25.13	22.18	20.97	19.70	20.82	19.26	19.66	19.95	19.80	21.33	20.19	18.33
1998	16.72	16.06	15.12	15.35	14.91	13.72	14.17	13.47	15.03	14.46	13.00	11.35
1999	12.52	12.01	14.68	17.31	17.72	17.92	20.10	21.28	23.80	22.69	25.00	26.10
2000	27.26	29.37	29.84	25.72	28.79	31.82	29.70	31.26	33.88	33.11	34.42	28.44
2001	29.59	29.61	27.25	27.49	28.63	27.60	26.43	27.37	26.20	22.17	19.64	19.39
2002	19.72	20.72	24.53	26.18	27.04	25.52	26.97	28.39	29.66	28.84	26.35	29.46
2003	32.95	35.83	33.51	28.17	28.11	30.66	30.76	31.57	28.31	30.34	31.11	32.13
2004	34.31	34.69	36.74	36.75	40.28	38.03	40.78	44.90	45.94	53.28	48.47	43.15
2005	46.84	48.15	54.19	52.98	49.83	56.35	59.00	64.99	65.59	62.26	58.32	59.41
2006	65.49	61.63	62.69	69.44	70.84	70.95	74.41	73.04	63.80	58.89	59.08	61.96
2007	54.51	59.28	60.44	63.98	63.46	67.49	74.12	72.36	79.92	85.80	94.77	91.69
2008	92.97	95.39	105.45	112.58	125.40	133.88	133.37	116.67	104.11	76.61	57.31	41.12
2009	41.71	39.09	47.94	49.65	59.03	69.64	64.15	71.05	69.41	75.72	77.99	74.47
2010	78.33	76.39	81.20	84.29	73.74	75.34	76.32	76.60	75.24	81.89	84.25	89.15
2011	89.17	88.58	102.86	109.53	100.90	96.26	97.30	86.33	85.52	86.32	97.16	98.56
2012	100.27	102.20	106.16	103.32	94.66	82.30	87.90	94.13	94.51	89.49	86.53	87.86
2013	94.76	95.31	92.94	92.02	94.51	95.77	104.67	106.57	106.29	100.54	93.86	97.63
2014	94.62	100.82	100.80	102.07	102.18	105.79	103.59	96.54	93.21	84.40	75.79	59.29
2015	47.22	50.58	47.82	54.45	59.27	59.82	50.90	42.87	45.48	46.22	42.44	37.19
2016	31.68	30.32	37.55	40.75	46.71	48.76	44.65	44.72	45.18	49.78	45.66	51.97
2017	52.50	53.47	49.33	51.06	48.48	45.18	46.63	48.04	49.82	51.58	56.64	57.88
2018	63.70	62.23	62.73	66.25	69.98	67.87	70.98	68.06	70.23	70.75	56.96	49.52
2019	51.38	54.95	58.15	63.86	60.83	54.66	57.35	54.81	56.95	53.96	57.03	59.88
2020	57.52	50.54	29.21	16.55	28.56	38.31	40.71	42.34	39.63	39.40	40.94	47.02
2021	52.00	59.04	62.33	61.72	65.17	71.38	72.49	67.73	71.65	81.48	79.15	71.71
2022	83.22	91.64	108.50	101.78	109.55	114.84	101.62	93.67	84.26	87.55	84.37	76.44
2023	78.12	76.83	73.28	79.45	71.58	70.25	76.07	81.39	89.43	85.64	77.69	71.90
2024	74.15	77.25										

Figure A7: Monthly crude oil price of West Texas Intermediate

Note: U.S. Energy Information Administration publishes monthly data on West Texas Intermediate (WTI).

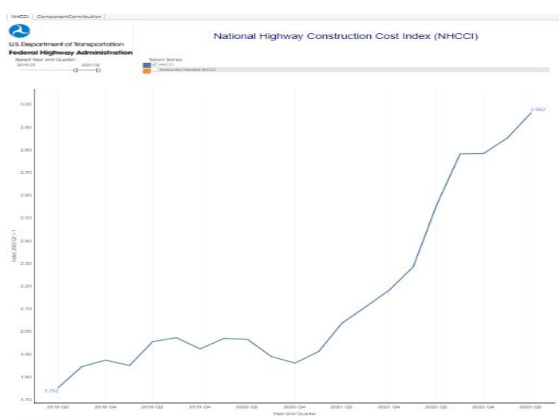


Figure A8: Quarterly value of National Highway Construction Cost Index

Note: The US Bureau of Transportation Statistics publishes the quarterly value of the National Highway Construction Cost Index (NHCCI).

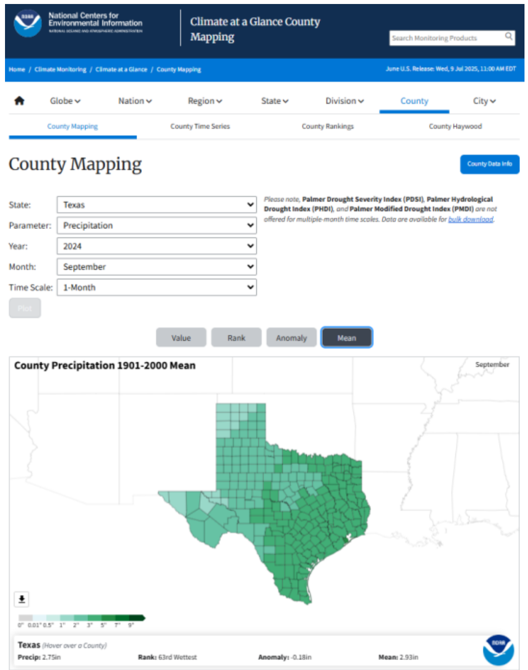


Figure A9: Mean monthly precipitation over Texas counties for September 2024

Note: The National Centers for Environmental Information, NOAA publishes precipitation data for different granularities of time and space.

Figure A10: US Census Bureau with annual estimates of the resident population

Note: Clicking on the any one of the states directly downloads annual population data for all the counties in that state.

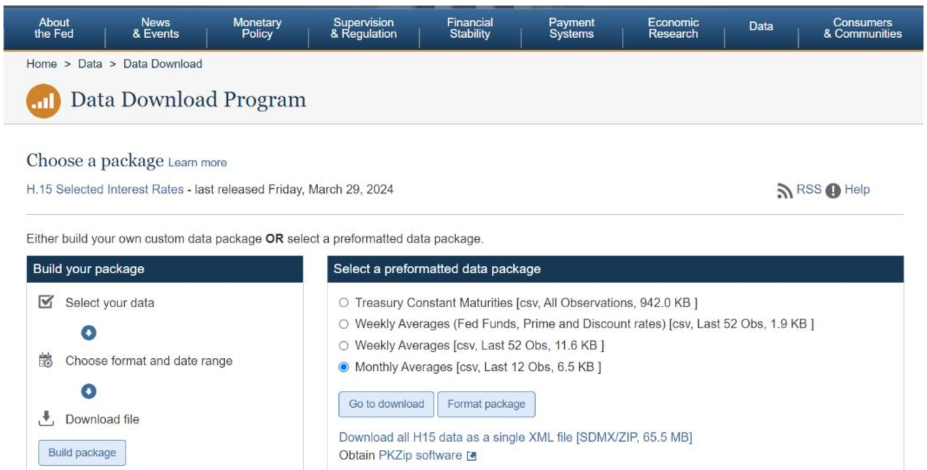


Figure A11: Federal Reserve website for collecting data on monthly prime loan rate

Note: The Board of Governors of the Federal Reserve System publishes data on weekly, monthly, or annual Prime Bank Loan Rate (PLR).

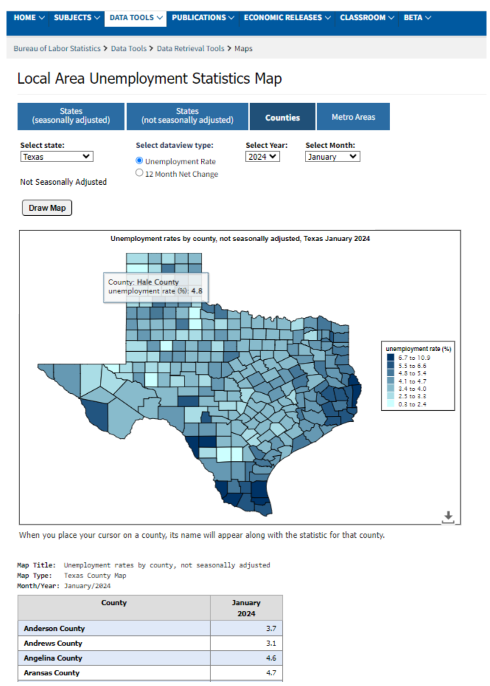
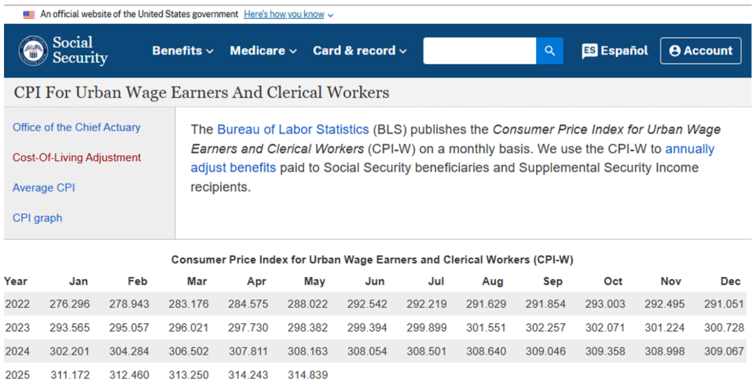


Figure A12: Unemployment rate of Texas counties in January 2024

Notes: The US Bureau of Labor Statistics (BLS) publishes the unemployment rate from 1990 to the present.



An official website of the United States government Here's how you know

Social Security Benefits Medicare Card & record Español Account

CPI For Urban Wage Earners And Clerical Workers

Office of the Chief Actuary

Cost-Of-Living Adjustment

Average CPI

CPI graph

The Bureau of Labor Statistics (BLS) publishes the *Consumer Price Index for Urban Wage Earners and Clerical Workers* (CPI-W) on a monthly basis. We use the CPI-W to annually adjust benefits paid to Social Security beneficiaries and Supplemental Security Income recipients.

Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2022	276.296	278.943	283.176	284.575	288.022	292.542	292.219	291.629	291.854	293.003	292.495	291.051
2023	293.565	295.057	296.021	297.730	298.382	299.394	299.899	301.551	302.257	302.071	301.224	300.728
2024	302.201	304.284	306.502	307.811	308.163	308.054	308.501	308.640	309.046	309.358	308.998	309.067
2025	311.172	312.460	313.250	314.243	314.839							

Figure A13: Consumer Price Index (CPI) for Urban Wage Earners and Clerical Workers (CPI-W) monthly from 2022 to 2025

Note: U.S. Bureau of Labor Statistics publishes monthly data on CPI from 1974.

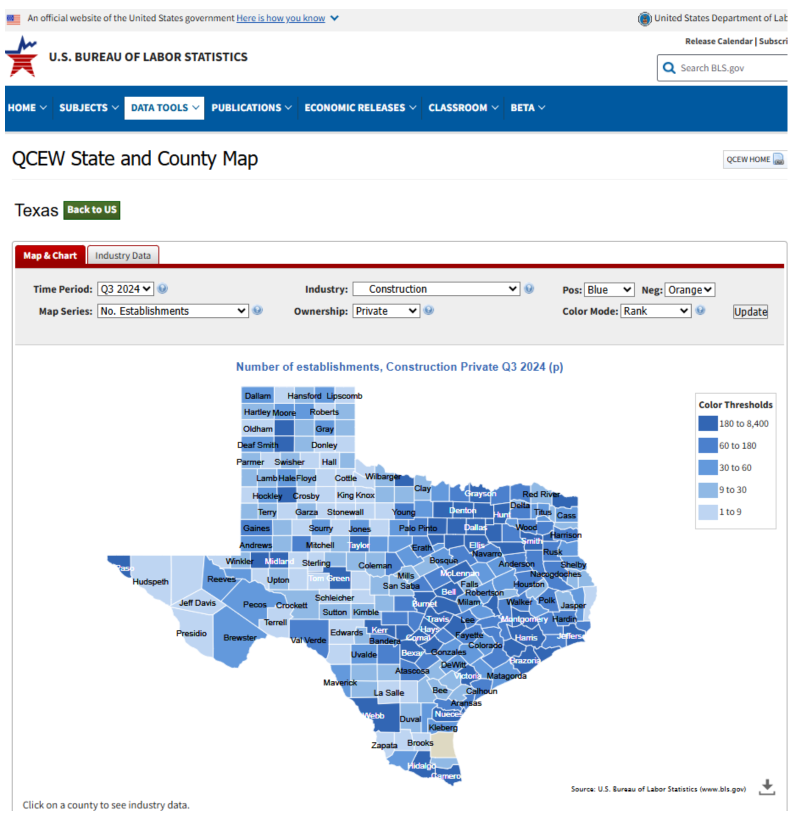
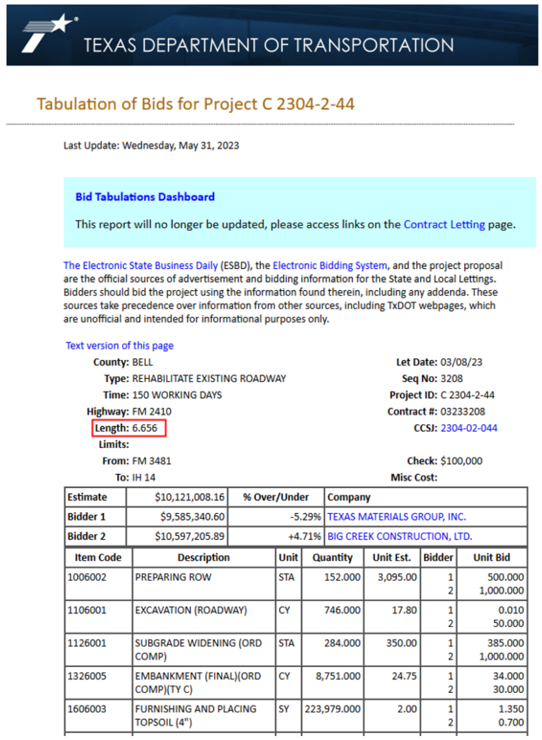


Figure A14: Quarterly data of establishment counts across Texas counties

Note: U.S. Bureau of Labor Statistics publishes quarterly data on CPI from 2001.



TEXAS DEPARTMENT OF TRANSPORTATION

Tabulation of Bids for Project C 2304-2-44

Last Update: Wednesday, May 31, 2023

Bid Tabulations Dashboard

This report will no longer be updated, please access links on the Contract Letting page.

The Electronic State Business Daily (ESBD), the Electronic Bidding System, and the project proposal are the official sources of advertisement and bidding information for the State and Local Lettings. Bidders should bid the project using the information found therein, including any addenda. These sources take precedence over information from other sources, including TxDOT webpages, which are unofficial and intended for informational purposes only.

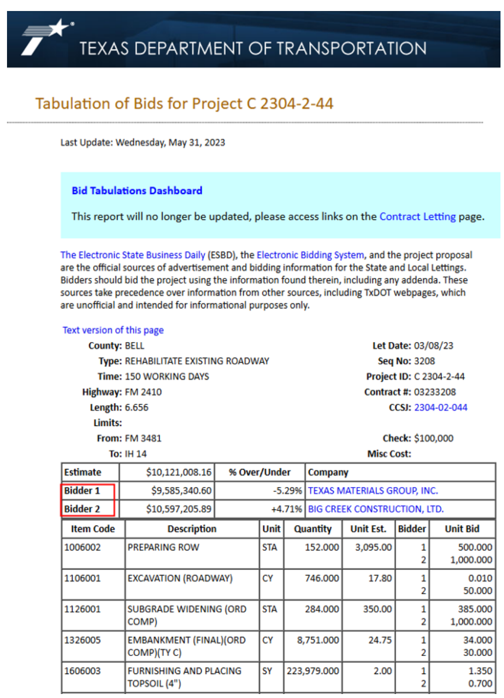
Text version of this page

County: BELL **Let Date:** 03/08/23
Type: REHABILITATE EXISTING ROADWAY **Seq No:** 3208
Time: 150 WORKING DAYS **Project ID:** C 2304-2-44
Highway: FM 2410 **Contract #:** 03233208
Length: 6.656 **CCSJ:** 2304-02-044
Limits:
From: FM 3481 **Check:** $100,000
To: IH 14 **Misc Cost:**

Estimate	$10,121,008.16	% Over/Under	Company
Bidder 1	$9,585,340.60	-5.29%	TEXAS MATERIALS GROUP, INC.
Bidder 2	$10,597,205.89	+4.71%	BIG CREEK CONSTRUCTION, LTD.

Item Code	Description	Unit	Quantity	Unit Est.	Bidder	Unit Bid
1006002	PREPARING ROW	STA	152.000	3,095.00	1 2	500.000 1,000.000
1106001	EXCAVATION (ROADWAY)	CY	746.000	17.80	1 2	0.010 50.000
1126001	SUBGRADE WIDENING (ORD COMP)	STA	284.000	350.00	1 2	385.000 1,000.000
1326005	EMBANKMENT (FINAL)(ORD COMP)(TY C)	CY	8,751.000	24.75	1 2	34.000 30.000
1606003	FURNISHING AND PLACING TOPSOIL (4")	SY	223,979.000	2.00	1 2	1.350 0.700

Figure A15: Bid Tabulations dashboard showing details on project length



TEXAS DEPARTMENT OF TRANSPORTATION

Tabulation of Bids for Project C 2304-2-44

Last Update: Wednesday, May 31, 2023

Bid Tabulations Dashboard

This report will no longer be updated, please access links on the Contract Letting page.

The Electronic State Business Daily (ESBD), the Electronic Bidding System, and the project proposal are the official sources of advertisement and bidding information for the State and Local Lettings. Bidders should bid the project using the information found therein, including any addenda. These sources take precedence over information from other sources, including TxDOT webpages, which are unofficial and intended for informational purposes only.

Text version of this page

County: BELL **Let Date:** 03/08/23
Type: REHABILITATE EXISTING ROADWAY **Seq No:** 3208
Time: 150 WORKING DAYS **Project ID:** C 2304-2-44
Highway: FM 2410 **Contract #:** 03233208
Length: 6.656 **CCSJ:** 2304-02-044
Limits:
From: FM 3481 **Check:** $100,000
To: IH 14 **Misc Cost:**

Estimate	$10,121,008.16	% Over/Under	Company
Bidder 1	$9,585,340.60	-5.29%	TEXAS MATERIALS GROUP, INC.
Bidder 2	$10,597,205.89	+4.71%	BIG CREEK CONSTRUCTION, LTD.

Item Code	Description	Unit	Quantity	Unit Est.	Bidder	Unit Bid
1006002	PREPARING ROW	STA	152.000	3,095.00	1 2	500.000 1,000.000
1106001	EXCAVATION (ROADWAY)	CY	746.000	17.80	1 2	0.010 50.000
1126001	SUBGRADE WIDENING (ORD COMP)	STA	284.000	350.00	1 2	385.000 1,000.000
1326005	EMBANKMENT (FINAL)(ORD COMP)(TY C)	CY	8,751.000	24.75	1 2	34.000 30.000
1606003	FURNISHING AND PLACING TOPSOIL (4")	SY	223,979.000	2.00	1 2	1.350 0.700

Figure A16: Bid Tabulations dashboard showing details on number of bidders

APPENDIX B

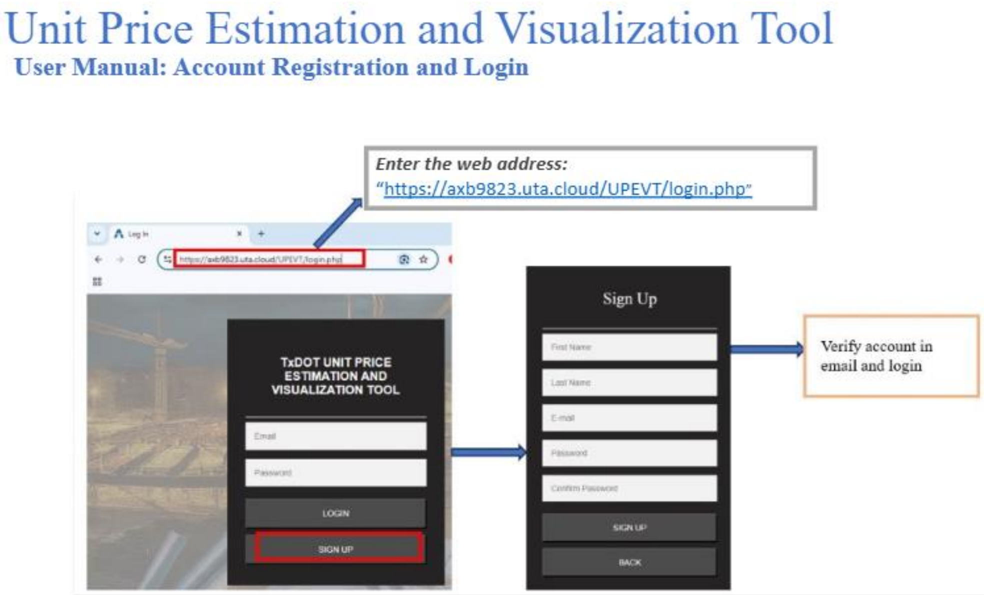
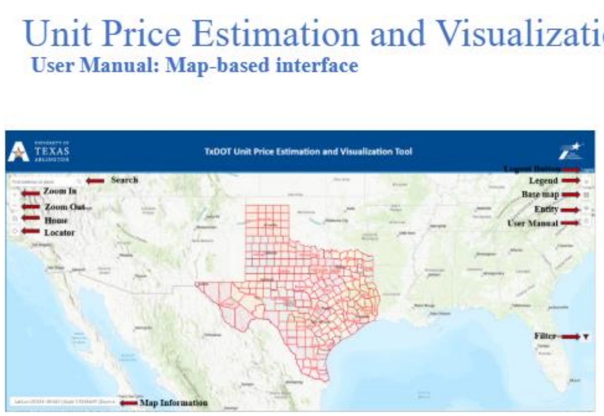


Figure B1: Account registration and sign up process for the developed tool



Layout window of Unit Price Estimation and Visualization Tool interface

Widget Name	Description
Search	This widget helps to find a specific location in the map-based interface.
Zoom In	This widget helps to zoom in on the map view in the map-based interface.
Zoom Out	This widget helps to zoom out of the current map view in the map-based interface.
Home	This widget brings the map view to the initial view extent.
Locator	This widget helps to find the location of the user.
Legend	This widget displays the legends of the spatial data entities which are displayed in the map-based interface.
Basemap	This widget allows users to select the base map to be displayed in the map view of the map-based interface.
Entity	This widget displays the list of spatial data entities. Users can choose to display or remove the entities from the map in the map-based interface.
User Manual	This widget allows users to access the user manual for the map-based interface.
Filter	This widget allows users to sort down data to be displayed based on specific criteria.

Description of the widgets in the map-based interface

Figure B2: Layout window and description of the widgets in the developed tool

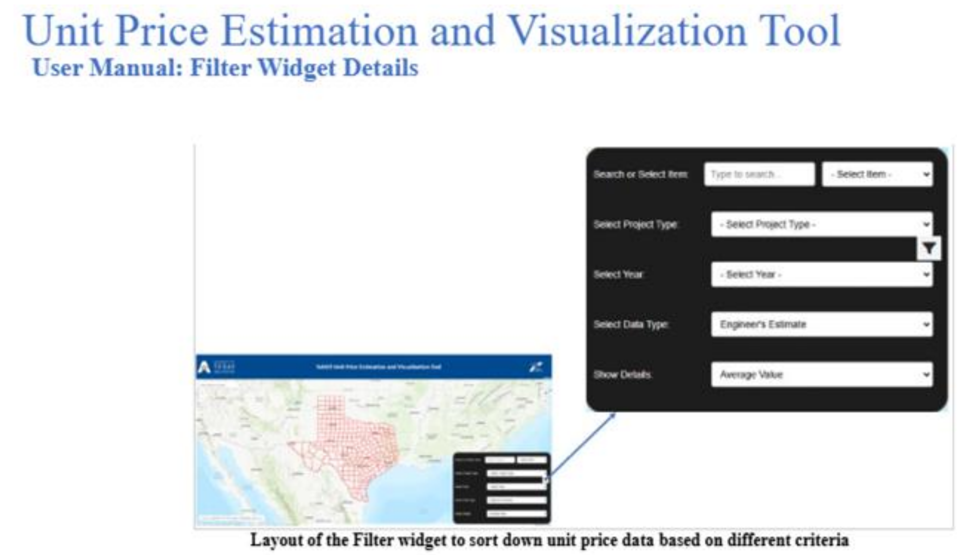


Figure B3: Filter widget details

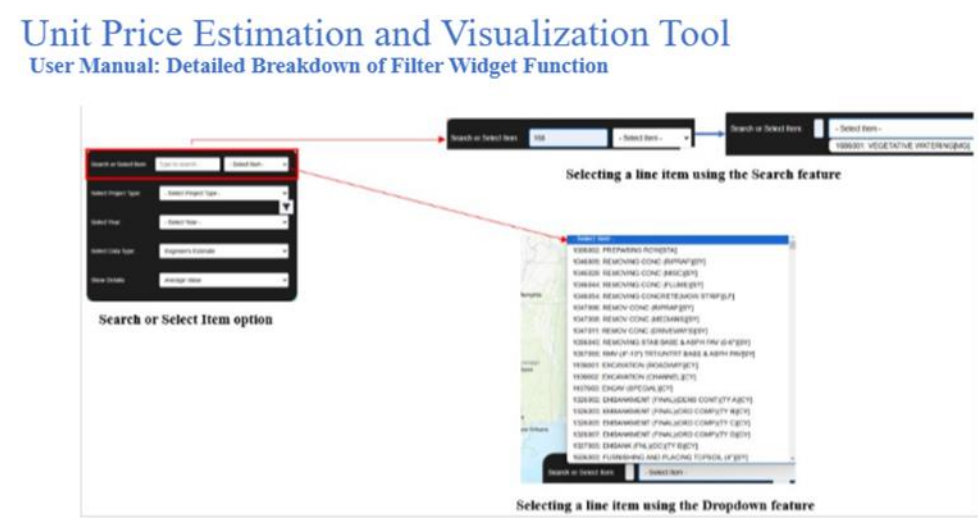


Figure B4: Search or select item function using the Filter widget

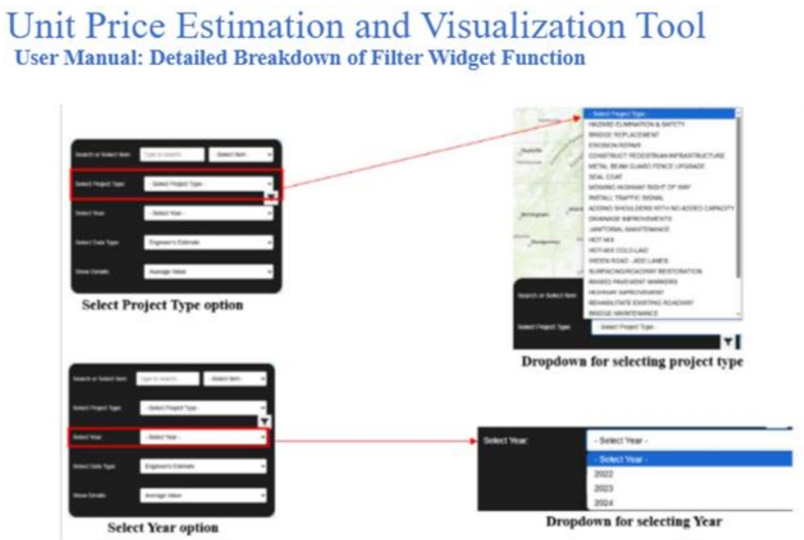


Figure B5: Functions for selecting project type and year using the Filter widget

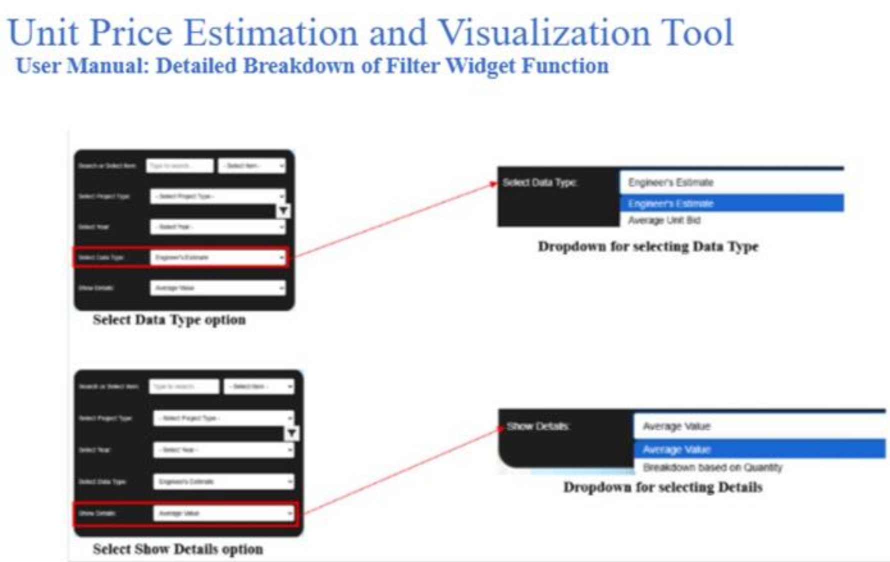


Figure B6: Functions for selecting data type and data details using the Filter widget

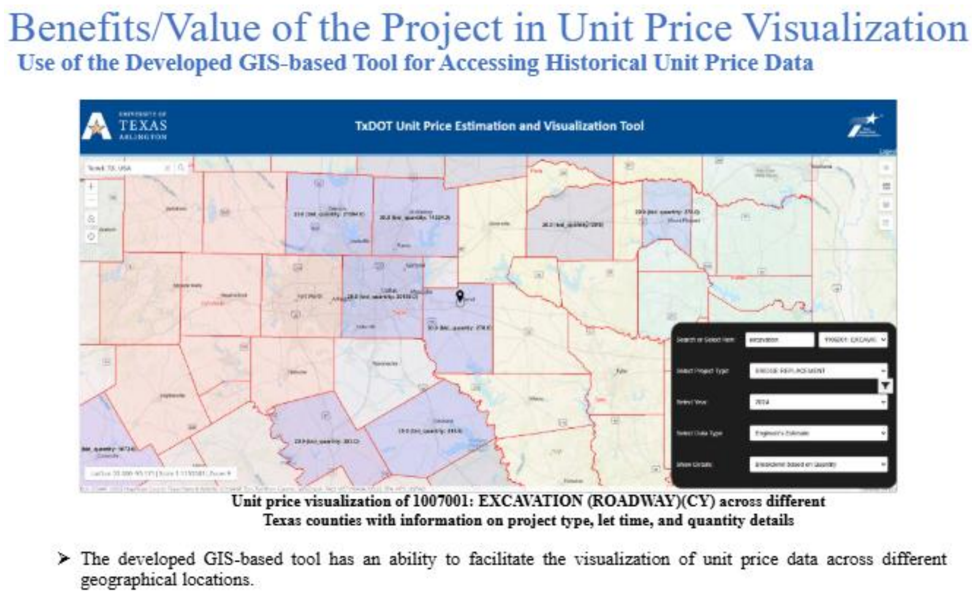


Figure B7: Visualization of unit prices of a line item across Texas counties for demonstrating the value of the project for unit price visualization

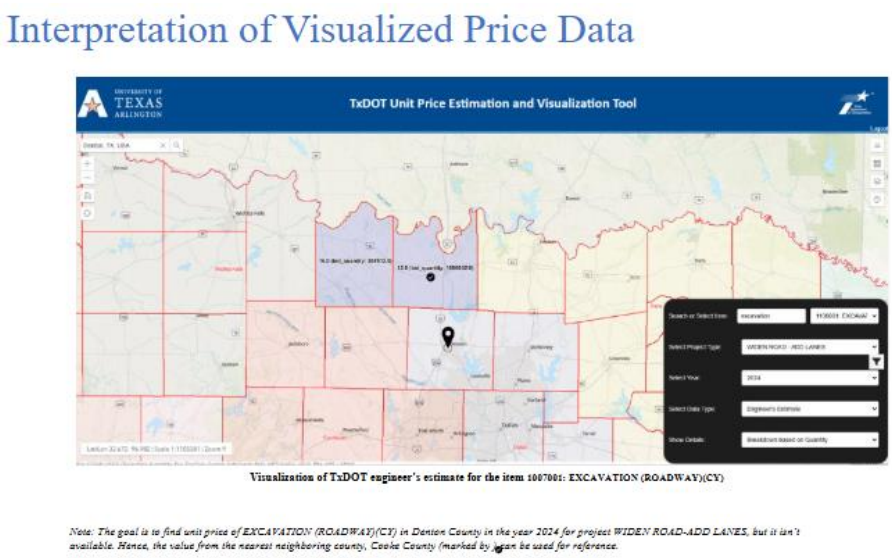


Figure B8: Demonstration of the tool use for estimating the unit price of a construction line item

REFERENCES

Abd, A. M., Kareem, Y. A., and Zehawi, R. N. (2024). Prediction and Estimation of Highway Construction Cost Using Machine Learning. *Engineering, Technology Applied Science Research*, 14(5):17222–17231.

Ahmadi, N., and Shahandashti, M. (2020), Characterizing Construction Demand Surge Using Spatial Panel Data Models, Natural Hazards Review, ASCE, 21(2), 04020008.

Ahmed, S., & Arocho, I. (2021). Analysis of cost comparison and effects of change orders during construction: Study of a mass timber and a concrete building project. Journal of Building Engineering, 33, 101856.

Al-Tabtabai, H., Alex, A. P., & Tantash, M. (1999). Preliminary Cost Estimation of Highway Construction Using Neural Networks. Cost Engineering, 41(3), 19.

Ashuri, B., Baek, M., & Li, M. (2022). Enhancing the Accuracy of Construction Cost Estimates for Major Lump Sum (LS) Pay Items and Generating a More Accurate List of Pay Items throughout the Design Development Process (No. FHWA-GA-22-2017). Georgia. Department of Transportation. Office of Performance-Based Management & Research.

Ashuri, B., Shahandashti, S. M., and Lu, J. (2012), Empirical Tests for Identifying Leading Indicators of ENR Construction Cost Index, Construction Management and Economics, Taylor & Francis, 30(11), 917-927.

Ashworth, A., & Perera, S. (2015). Cost Studies of Buildings (0 ed.). Routledge. https://doi.org/10.4324/9781315708867

Awuku, B., Asa, E., & Baffoe-Twum, E. (2022). Conceptual cost estimation of highway bid unit prices using ordinary kriging. International Journal of Construction Management, 1–10. https://doi.org/10.1080/15623599.2022.2156844

Baek, M. and Ashuri, B. (2019). Analysis of the variability of submitted unit price bids for asphalt line items in highway projects. *Journal of construction engineering and management*, 145(4):04019020.

Baek, M., & Ashuri, B. (2018a). Assessment of spatial correlation patterns of unit price bids and external factors. Proc., 54th Associated Schools of Construction (ASC) Annual Int. Conf, 18–21.

Baek, M., & Ashuri, B. (2018b). Profile Monitoring for Examining Impact of Hurricane Katrina and Rita on Highway Construction Cost. Transportation Research Record: Journal of the Transportation Research Board, 2672(51), 79–87. https://doi.org/10.1177/0361198118777619

Bhargava, A., Labi, S., Chen, S., Saeed, T. U., & Sinha, K. C. (2017). Predicting cost escalation pathways and deviation severities of infrastructure projects using risk-based econometric models and Monte Carlo simulation. Computer-Aided Civil and Infrastructure Engineering, 32(8), 620–640.

Board of Governors of the Federal Reserve System (U.S.). (n.d.). *Bank Prime Loan Rate [MPRIME]*. Retrieved July 11, 2025, from Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/MPRIME

Cao, Y., Ashuri, B., & Baek, M. (2018). Prediction of unit price bids of resurfacing highway projects through ensemble machine learning. Journal of Computing in Civil Engineering, 32(5), 04018043.

Chauhan, N. K., & Singh, K. (2018, September). A review on conventional machine learning vs deep learning. In 2018 International conference on computing, power and communication technologies (GUCON) (pp. 347-352). IEEE.

Cheng, Y. M. (2014). An exploration into cost-influencing factors on construction projects. International Journal of Project Management, 32(5), 850-860.

Choi, C-Y., Ryu, K.R., and Shahandashti, M. (2021), Predicting City-level Construction Cost Index using Linear Forecasting Models, Journal of Construction Engineering and Management, ASCE, 147(2), 04020158.

Chou, J. S., & Lin, C. (2013). Predicting disputes in public-private partnership projects: Classification and ensemble models. Journal of Computing in Civil Engineering, 27(1), 51-60.

Cirilovic, J., Vajdic, N., Mladenovic, G., & Queiroz, C. (2014). Developing cost estimation models for road rehabilitation and reconstruction: Case study of projects in Europe and Central Asia. Journal of Construction Engineering and Management, 140(3), 04013065.

Danisworo, B., & Latief, Y. (2019). Estimation model of Jakarta MRT phase 1 project cost overrun for the risk based next phase project funding purpose. IOP Conference Series: Earth and Environmental Science, 258, 012049. https://doi.org/10.1088/1755-1315/258/1/012049

Dietterich, T. G. 2000. "Ensemble methods in machine learning." In Proc., Int. Workshop on Multiple Classifier Systems, 1–15. Berlin: Springer.

Dong, R., Muhammad, A., and Nauman, U. (2025). The influence of weather conditions on time, cost, and quality in successful construction project delivery. *Buildings*, 15(3):474.

Emsley, M. W., Lowe, D. J., Duff, A. R., Harding, A., & Hickson, A. (2002). Data modelling and the application of a neural network approach to the prediction of total construction costs. Construction Management & Economics, 20(6), 465-472.

Engineering News-Record. (n.d.). *Engineering News-Record*. ENR. https://www.enr.com/

Gardner, B. J., Gransberg, D. D., & Jeong, H. D. (2016). Reducing data-collection efforts for conceptual cost estimating at a highway agency. Journal of Construction Engineering and Management, 142(11), 04016057.

Ghimire, P., Pokharel, S., Kim, K., and Barutha, P. (2023). Machine learning-based pre- diction models for budget forecast in capital construction. In *Proceedings of the 2nd International Conference on Construction, Energy, Environment & Sustainability, Funchal, Portugal*, pages 27–30.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2). Cambridge: MIT press.

Han, T., Siddique, A., Khayat, K., Huang, J., & Kumar, A. (2020). An ensemble machine learning approach for prediction and optimization of modulus of elasticity of recycled aggregate concrete. Construction and Building Materials, 244, 118271.

Hannan, A., Ahmed, A., Ashraf, T., & Bai, Q. (2016). Estimation of highway project cost using probabilistic technique. DEStech Transactions on Engineering and Technology Research.

Hegazy, T., & Ayed, A. (1998). Neural network model for parametric cost estimation of highway projects. Journal of construction engineering and management, 124(3), 210-218.

Hyari, K. H., Al-Daraiseh, A., & El-Mashaleh, M. (2016). Conceptual Cost Estimation Model for Engineering Services in Public Construction Projects. Journal of Management in Engineering, 32(1), 04015021. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000381

Idaho Transportation Department (2019). *Construction Cost Estimating Guide*. Idaho Transportation Department. Published online; includes monthly equipment-rental and cost calculations.

Indiana Department of Transportation (2020). *Chapter 20: Cost Estimating*. Indiana De- partment of Transportation. Refer to the chapter in the Indiana Design Manual.

Jiang, F., Awaitey, J., and Xie, H. (2022). Analysis of construction cost and investment planning using time series data. *Sustainability*, 14(3):1703.

Karaca, I., Gransberg, D. D., and Jeong, H. D. (2020). Improving the accuracy of early cost estimates on transportation infrastructure projects. *Journal of Management in Engineering*, 36(5):04020063.

Khodahemmati, N., and Shahandashti, M. (2020), Diagnosis and Quantification of Post-Disaster Construction Material Cost Fluctuations, Natural Hazards Review, ASCE, 21(3), DOI: 10.1061/(ASCE)NH.1527-6996.0000381.

Kim, S., & Shahandashti, M. (2023). Examining Pipe Cost Changes after Various Disasters in Los Angeles, California. In *Pipelines 2023* (pp. 481-491).

Kim, S., Abediniangerabi, B., and Shahandashti, M. (2020). Forecasting pipeline construction costs using time series methods. In *Pipelines 2020*, pages 198–209. American Society of Civil Engineers Reston, VA.

Kim, S., Choi, C.-Y., Shahandashti, M., and Ryu, K. R. (2022). Improving accuracy in predicting city-level construction cost indices by combining linear arima and nonlinear anns. *Journal of Management in Engineering*, 38(2):04021093.

Kim, S., Makhmalbaf, A., and Shahandashti, M. (2024), Forecasting the Architecture Billing Index (ABI) using Machine Learning Prediction Models, Engineering, Construction, and Architectural Management, Emerald, DOI: 10.1108/ECAM-06-2023-0544.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.

Le, C., Le, T., Jeong, H. D., and Lee, E.-B. (2019). Geographic information system–based framework for estimating and visualizing unit prices of highway work items. *Journal of Construction Engineering and Management*, 145(8):04019044.

Levy, S. M. (2006). *Project Management in Construction (McGraw-Hill Professional Engineering)*. McGraw-Hill Professional.

Mahdavian, A., Shojaei, A., Salem, M., Yuan, J. S., and Oloufa, A. A. (2021). Data-driven predictive modeling of highway construction cost items. *Journal of Construction Engineering and Management*, 147(3):04020180.

Massaoudi, M., Abu-Rub, H., Refaat, S. S., Chihi, I., & Oueslati, F. S. (2021). Deep learning in smart grid technology: A review of recent advancements and future prospects. *IEEE Access*, *9*, 54558-54578.

Meharie, M. G., Mengesha, W. J., Gariy, Z. A., & Mutuku, R. N. (2022). Application of stacking ensemble machine learning algorithm in predicting the cost of highway construction projects. Engineering, Construction and Architectural Management, 29(7), 2836-2853.

Meng, C., Qu, D., and Duan, X. (2024). Cost estimation of metro construction projects using interpretable machine learning. *Journal of Computing in Civil Engineering*, 38(6):04024038.

Meteyard, L. and Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112:104092.

Mohamed, B., & Moselhi, O. (2022). Conceptual estimation of construction duration and cost of public highway projects. J. Inf. Technol. Constr, 27(29), 595-618.

Morabito, F. C., et al. 2017. "Deep learning representation from electroencephalography of early-stage Creutzfeld-Jakob disease and features for differentiation from rapidly progressive dementia." Int. J. Neural Syst. 27 (2): 1650039. https://doi.org/10.1142/S0129065716500398.

Musarat, M. A., Alaloul, W. S., & Liew, M. S. (2021). Impact of inflation rate on construction projects budget: A review. Ain Shams Engineering Journal, 12(1), 407–414.

National Centers for Environmental Information. (n.d.). *National Centers for Environmental Information*. National Oceanic and Atmospheric Administration. Retrieved July 11, 2025, from https://www.ncei.noaa.gov/

Nyitrai, T., & Virág, M. (2019). The effects of handling outliers on the performance of bankruptcy prediction models. Socio-Economic Planning Sciences, 67, 34-42.

Ortega-Zamorano, F., Jerez, J. M., Gómez, I., & Franco, L. (2017). Layer multiplexing FPGA implementation for deep back-propagation learning. *Integrated Computer-Aided Engineering*, *24*(2), 171-185.

Ortiz-Garcia, A., J. Munilla, J. M. Gorriz, and J. Ramirez. 2016. "Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease." Int. J. Neural Syst. 26 (7): 1650025. https://doi.org/10.1142/S0129065716500258.

Paik, Y., Chung, F., and Ashuri, B. (2025). Preliminary cost estimation of pavement maintenance projects through machine learning: Emphasis on trees algorithms. *Journal of Management in Engineering*, 41(4):04025027.

Pennsylvania Department of Transportation (PennDOT) (2025). *Publication 352: Estimating Manual*. Pennsylvania Department of Transportation. Effective for projects initiated after October 1, 2025.

Petroutsatou, K., Georgopoulos, E., Lambropoulos, S., & Pantouvakis, J. P. (2012). Early cost estimating of road tunnel construction using neural networks. Journal of construction engineering and management, 138(6), 679-687.

Pradhan, S., & Arneson, E. (2021). Postdisaster Labor-Demand Surge in the US Highways, Roads, and Bridges Construction Sector. Journal of Management in Engineering, 37(1), 04020102. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000869

Raetz, H., Forscher, T., Kneebone, E., & Reid, C. (2020). The Hard Costs of Construction: Recent Trends in Labor and Materials Costs for Apartment Buildings in California. https://ternercenter.berkeley.edu/wpcontent/uploads/pdfs/Hard_Construction_Costs_March_2020.pdf

Rafiei, M. H., & Adeli, H. (2018). Novel Machine-Learning Model for Estimating Construction Costs Considering Economic Variables and Indexes. Journal of Construction Engineering and Management, 144(12), 04018106. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001570

Riddle, M., & Whittington, J. (2022). A tunnel beneath Seattle: The megaproject to replace the Alaskan Way Viaduct. In Megaprojects for Megacities (pp. 275–305). Edward Elgar Publishing.

Shahandashti, M., Abediniangerabi, B., Zahed, E., & Kim, S. (2023). Construction analytics: Forecasting and investment valuation. Berlin: Springer.

Shahandashti, M., and Ashuri, B. (2015), Highway Construction Cost Variation Forecasting Using Vector Error Correction Models, Journal of Management in Engineering, ASCE, 32(2), 04015040.

Shahandashti, S. M. and Ashuri, B. (2013), Forecasting ENR Construction Cost Index Using Multivariate Time Series Models, Journal of Construction Engineering and Management, ASCE, 139(9), 1237-1243.

Shiha, A., Dorra, E. M., & Nassar, K. (2020). Neural Networks Model for Prediction of Construction Material Prices in Egypt Using Macroeconomic Indicators. Journal of Construction Engineering and Management, 146(3), 04020010. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001785

Shrestha, B., Shrestha, P. P., Maharjan, R., & Gransberg, D. (2022). Cost, Change Order, and Schedule Performance of Highway Projects. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 14(1), 04521044. https://doi.org/10.1061/(ASCE)LA.1943-4170.0000523

Shrestha, K. J., Paul, P., and Uddin, M. M. (2025). Parametric cost estimating models for resurfacing projects. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 13(4):04525058.

Simi´c, N., Ivaniˇsevi´c, N., Nedeljkovi´c, , Seni´c, A., Stojadinovi´c, Z., and Ivanovi´c, M. (2023). Early highway construction cost estimation: Selection of key cost drivers. *Sustainability*, 15(6):5584.

Swei, O., Gregory, J., & Kirchain, R. (2017). Construction cost estimation: A parametric approach for better estimates of expected cost and variation. Transportation Research Part B: Methodological, 101, 295–305. https://doi.org/10.1016/j.trb.2017.04.013

Texas Department of Transportation (2022). Plan review guidelines. https://ftp.txdot.gov/pub/txdot-info/sat/specinfo/plan-review.pdf. Accessed: 2025-06-24.

Texas Department of Transportation (2025). Transportation funding in texas: January 2025 edition (state fiscal years 2026–2027). Accessed: July 3, 2025.

Texas Department of Transportation. (n.d.). *Bid tabulations dashboard*. Tableau Public. https://tableau.txdot.gov/views/BidTabulations/BidTabulations

Bureau of Labor Statistics. (n.d.). *Home page*. U.S. Department of Labor. Retrieved July 4, 2025, from https://www.bls.gov/

Thomason, C. (2017, October 26). PS&E Preparation Manual: Prices. http://onlinemanuals.txdot.gov/txdotmanuals/pse/prices.htm

Tong, B., Guo, J., and Fang, S. (2021). Predicting budgetary estimate of highway construction projects in china based on gra-lasso. *Journal of Management in Engineering*, 37(3):04021012.

U.S. Bureau of Economic Analysis. (n.d.). *U.S. Bureau of Economic Analysis*. Retrieved July 11, 2025, from https://www.bea.gov/

U.S. Census Bureau. (n.d.). *U.S. Census Bureau*. Retrieved July 11, 2025, from https://www.census.gov/

U.S. Census Bureau. (n.d.). *U.S. Census Bureau*. Retrieved July 11, 2025, from https://www.census.gov/

U.S. Department of Transportation, Office of Inspector General (2019). Fhwa lacks adequate oversight and guidance for engineer's estimates. Technical Report ST2019020, U.S. Department of Transportation. Report No. ST2019020, issued March 13, 2019.

U.S. Energy Information Administration. (n.d.). *U.S. Energy Information Administration*. Retrieved July 11, 2025, from https://www.eia.gov/

Virginia Department of Transportation (2025). *VDOT Cost Estimating Manual*. Virginia Department of Transportation, Richmond, VA. Published March 25, 2025; accessed July 6, 2025.

Wang, R., Asghari, V., Cheung, C. M., Hsu, S. C., & Lee, C. J. (2022). Assessing effects of economic factors on construction cost estimation using deep neural networks. Automation in Construction, 134, 104080.

Williams, T. P., & Gong, J. (2014). Predicting construction cost overruns using text mining, numerical data and ensemble classifiers. Automation in Construction, 43, 23-29.

Wilmot, C. G., & Cheng, G. (2003). Estimating future highway construction costs. Journal of Construction Engineering and Management, 129(3), 272-279.

Wilmot, C. G., & Mei, B. (2005). Neural network modeling of highway construction costs. Journal of construction engineering and management, 131(7), 765-771.

Xiao, Y., Wu, J., Lin, Z., & Zhao, X. (2018). A deep learning-based multi-model ensemble method for cancer prediction. Computer methods and programs in biomedicine, 153, 1-9.

Zhang, Y., Minchin Jr, R. E., Flood, I., and Ries, R. J. (2023). Preliminary cost estimation of highway projects using statistical learning methods. *Journal of Construction Engineering and Management*, 149(5):04023026.