

A PROJECT ON
SMART CITY DEVELOPMENT: ENHANCING URBAN PLANNING IN
SEATTLE THROUGH ELECTRICAL PERMIT ANALYSIS AND
PREDICTIVE MODELING

BY
CHIZURUMOKE DAVID OKAFOR

002245333

EBENEZER OPOKU-MANU

NORTHEASTERN UNIVERSITY

VANCOUVER CAMPUS

IE7945 PROJECT

18TH AUGUST, 2024

Introduction

Electrical energy is a vital resource for many different types of businesses, including big, small, residential, and agricultural ones. The persistent technical improvement in the important areas will eventually lead to an increase in the need for electrical energy (Subramani et al., 2024). As a result, electrical energy has come to be seen as essential to daily life.

Every year, the City of Seattle grants a significant number of electrical permits for both residential and non-residential buildings. Regulating and forecasting permit applications, issuances, and completions, however, is a difficult process that is frequently hampered by inefficiencies and erratic swings.

Seattle's sustainable development and public safety depend heavily on the efficient administration and control of electrical licenses. This project is highly relevant to society for several reasons:

- ▶ Urban Planning and Development: Improved Resource Allocation, Informed Decision-Making
- ▶ Economic Impact: Cost Management, Boosting Efficiency
- ▶ Public Safety and Compliance: Ensuring Safety Standards, Timely Inspections
- ▶ Environmental Sustainability: Energy Efficiency, Sustainable Urban Growth
- ▶ Community Benefits: Transparency and Accountability, Enhanced Quality of Life
- ▶ Technological Advancement: Innovation in Public Services, Smart City Initiatives

Electrical permits are required in Seattle to ensure the safety and conformity of electrical installations. The Seattle Department of Construction and Inspections is in charge of these permits, which are required for any electrical work, including installations, alterations, additions, or connections to electrical equipment (SDCI).

In Seattle, there are two different kinds of electrical permits: Over-the-Counter (OTC) permits and Plan Review permits. Online self-issuance is available for Over-the-Counter (OTC) permits, which are typically given for smaller projects. Installations in duplexes or single-family homes, temporary electricity installations for events under 600 amps, fire alarm systems with six or fewer devices, and photovoltaic systems rated less than 7.7 kW are a few examples. On the other hand, more complicated projects need a plan review permit. Installations in academic, institutional, or medical settings; projects larger than 5,000 square feet; fire alarm systems including seven or more devices; and renewable energy systems with a rating of 12 kW or above are a few examples. Through the Seattle Services Portal, applicants can submit applications for permits. The application procedure for OTC permits is simple and can be finished online.

Applicants seeking plan review permits must provide comprehensive plans for SDCI to review. Depending on their complexity, certain projects may also require pre-application meetings or extra documentation.

The size and complexity of the project, the value of the work, any extra permits needed, the time needed for application and plan reviews, inspection, and technology fees are some of the variables that affect the price of electrical permits in Seattle.

Certain permits, like the Subject-to-Field-Inspection permits, which cost 40% of the plan review price, may only require partial payment; the SDCI offers tools to estimate permit payments. Inspections are required following permit acquisition to guarantee adherence to the Seattle Electrical Code. Several inspections may be needed at various phases of the installation, depending on the project. To prevent delays, the SDCI offers comprehensive guidance on what must be prepared for each inspection. In order to guarantee the dependability and safety of the electrical systems in homes, businesses, and public spaces, electrical licenses in Seattle are an essential requirement. In order to ensure the safe and sustainable growth of the city, property owners and contractors must adhere to the set procedures and get the relevant licenses.

The development of smart cities and urban planning are crucial for the efficient and sustainable expansion of cities. In cities like Seattle, where population and technology are advancing at a rapid pace, creative solutions are desperately needed to manage urban infrastructure efficiently. The goal of this project is to develop a smart city application that incorporates information from electrical permits that Seattle has issued or is currently processing. This project integrates data analysis, predictive modeling, and application development to offer useful tools and insights to city planners, developers, and citizens.

Importance of Electrical Permits in Urban Planning

Electrical permits are necessary to ensure the safety and compliance of electrical installations in residential, commercial, and industrial establishments. These licenses ensure that installations adhere to local building codes, help with routine maintenance, and reduce the likelihood of electrical risks. Analyzing and projecting trends in electrical permit issuance can significantly enhance urban planning projects. This makes it possible to build infrastructure more proactively, make educated decisions, and use resources more effectively (City of Seattle, 2023).

Project Objectives

The primary objectives of this project are:

1. **Trend Analysis:** To analyze historical data on electrical permits to identify seasonal patterns, trends, and significant changes over time.
2. **Predictive Modeling:** To develop predictive models that can forecast future permit issuance based on historical data and various influencing factors.
3. **Geospatial Analysis:** To identify hotspots for electrical permits and visualize their spatial distribution within the city.
4. **Application Development:** To create a user-friendly smart city application that provides interactive maps, dashboards, and search functionalities for users.

Literature Review

In order to comprehend previous behaviors and forecast future patterns, it is essential to analyze historical trends in data. Analyzing trends entails spotting cyclical tendencies, seasonal patterns, and any notable departures from projected trends throughout time. In order to identify trends and seasonality in a variety of areas, including construction permits, time series analysis has been widely used in urban planning and infrastructure management. According to research by Chatfield (2004), good forecasting requires a grasp of temporal trends, particularly in industries like construction and electrical licenses that are subject to regulatory changes and economic cycles. With this method, urban planners can forecast times of peak demand and distribute resources appropriately.

Studies have indicated that there are frequently significant seasonal patterns in the issuing of electrical permits and other relevant activities related to building. For instance, Lee et al. (2017) discovered that in temperate regions, building activities typically peak during the warmer months, which has a direct impact on the number of licenses given. Planners must take this seasonality into account when estimating future permit volumes, planning inspections, and resource requirements.

Research has shown how regulatory changes and the state of the economy affect the issuing of construction and electrical permits. Examples of these studies are those conducted by Gyourko and Saiz (2006). Building activity and permit applications often rise during times of economic expansion, although they may fall during recessions. Furthermore, modifications to energy efficiency standards or building rules may result in brief spikes and drops in permit applications.

An advanced analytical technique called predictive modeling is used to project future patterns from historical data. These models support resource planning and decision-making by forecasting permit issuance in the context of electrical permits.

Because machine learning approaches can handle big datasets and reveal intricate patterns, they are becoming more and more common in predictive modeling. Time series data in urban contexts have been predicted using methods like ARIMA, SARIMA, and, more recently, machine learning models like Random Forests, Support Vector Machines (SVM), and Neural Networks (Hyndman & Athanasopoulos, 2018).

The study conducted by Hsiao et al. (2020) suggests that machine learning methods can improve the prediction accuracy of construction permit issuing more than traditional statistical models. Their study revealed that combining historical data with external factors like the state of the economy and weather might enhance forecast performance. Given all the external factors that could influence the process, this method is particularly helpful for forecasting when electrical permits will be issued.

Accurate forecasting depends on the incorporation of external factors into predictive models. Research has indicated that a number of factors, including economic expansion, housing market dynamics, and energy laws, have a major impact on the number of permits that are issued (Lütkepohl, 2005). Models that incorporate these variables are able to produce forecasts that are more accurate and better reflect the dynamics of permit issuing.

Urban planning uses geospatial analysis to map and examine the spatial distribution of different operations, such as the granting of electrical licenses. City planners can more efficiently deploy resources and make plans for future growth and development by identifying hotspots.

The spatial distribution of infrastructure projects, particularly electricity permits, is being studied more and more using geospatial analysis. Research like Anselin's (1995) has highlighted the significance of spatial autocorrelation in figuring out where there is a lot of activity and how permits cluster together. Patterns that are not immediately obvious in conventional tabular data can be found using this method. Methods for locating spatial hotspots were first presented in research by Getis and Ord (1992), and they have subsequently been frequently used in urban studies. City planners can use hotspot analysis to pinpoint places with concentrated permit activity, which may point to areas experiencing rapid development or higher infrastructure demands. These understandings are essential for designing infrastructure and allocating resources proactively.

Geospatial analysis results must be communicated effectively through visualization. User-friendly methods of presenting spatial data include heat maps, choropleth maps, and interactive dashboards. These visualizations are crucial for helping non-experts, including legislators and city planners, comprehend intricate spatial patterns and make defensible decisions, claims MacEachren (1994).

Modern urban planning requires the creation of smart city applications with interactive dashboards, maps, and search features. Real-time data interaction between users and these applications promotes improved decision-making and public participation.

Applications for "smart cities" have become more popular as cities look to use data to plan and run their operations more effectively. These programs usually combine information from several sources, such as economic, temporal, and geographic data, to give a thorough picture of urban activity. Batty et al. (2012) conducted research that emphasizes how smart city applications might enhance urban planning procedures by increasing the accessibility and actionability of data.

The usability of smart city applications is a key factor in their success. Nielsen (1993) and Shneiderman (2004) conducted studies that highlight the significance of user-centered design in creating apps that work.

A user-friendly interface that facilitates data navigation, report generation, and trend visualization is essential for guaranteeing the application's widespread acceptance and efficient use. When predictive modeling and geographic analytics are combined into one tool, urban planners can gain a lot of useful information. According to Goodchild (2018), combining these two analytical modalities allows for a more thorough comprehension of temporal and spatial dynamics, facilitating more informed decision-making.

Numerous case studies, including Chicago's "OpenGrid" and New York City's "NYC Planning Labs," show how smart city technologies can be beneficial for urban planning. These programs offer interactive tools for examining many facets of urban infrastructure, such as public safety, transit, and construction permits. The advantages of creating a comparable program for examining electrical permits in Seattle are demonstrated by these case studies.

An effective framework for assessing and projecting electrical permit activities is produced by combining trend analysis, predictive modeling, geospatial analysis, and application development. Through the utilization of these approaches, urban planners can acquire significant knowledge about historical and prospective patterns, pinpoint regions with significant activity, and formulate plans for more effective distribution of resources. The accessibility and usefulness of these insights are further improved by the creation of an intuitive application, which gives decision-makers the capacity to efficiently plan for the expansion and infrastructure requirements of the city.

Data overview

The dataset for this project is sourced from the City of Seattle's Open Data Portal, which provides comprehensive data on all electrical permits issued or in process. The dataset includes detailed information on permit number, issue date, permit type, contractor details, project description, project valuation, and location coordinates. This rich dataset allows for in-depth analysis and robust predictive modeling.

Data source: Seattle open data

Link to dataset: https://data.seattle.gov/Permitting/Electrical-Permits/c4tj-daue/about_data

The dataset as at the time of this project contained over 442k rows and 20 columns

Github repository:

<https://github.com/Chizurumoke/ENHANCING-URBAN-PLANNING-IN-SEATTLE-THROUGH-ELECTRICAL-PERMIT-ANALYSIS-AND-PREDICTIVE-MODELING.git>

Data overview before cleaning and feature engineering

	PermitNum	PermitClass	PermitClassMapped	PermitTypeMapped	\
AppliedDate					
2002-07-15	6001044-EL	Institutional	Non-Residential	Electrical	
2002-07-16	6001080-EL	Commercial	Non-Residential	Electrical	
2002-07-17	6001190-EL	Commercial	Non-Residential	Electrical	
2002-07-18	6001254-EL	Multifamily	Residential	Electrical	
2002-07-18	6001255-EL	Institutional	Non-Residential	Electrical	
					Description \
AppliedDate					
2002-07-15	400 AMP SERVICE, FEEDERS TO PANELS, BRANCH CIR...				
2002-07-16	SERVICE CHANGE TO 400 AMP SINGLE PHASE\r\n8/30...				
2002-07-17	ELECTRICAL EQUIPMENT REPAIR AND REPLACE (TRAND...				
2002-07-18	WIRE THREE UNIT TOWNHOUSE				
2002-07-18	INTERIOR ALTERATIONS , CONVERSION TO CLASSROOMS				
					EstProjectCost IssuedDate ExpiresDate CompletedDate \
AppliedDate					
2002-07-15	5000.0	2002-07-17	2003-07-17	2002-11-06	
2002-07-16	7200.0	2002-07-16	2003-07-16	2004-06-03	
2002-07-17	130000.0	2002-07-23	2003-07-23	2002-10-09	
2002-07-18	9500.0	2002-07-19	2003-07-19	2002-10-02	
2002-07-18	84000.0	2002-08-01	2003-08-01	2003-01-27	
					StatusCurrent ... ApplicationToIssueTime IssueToCompletionTime \
AppliedDate					
2002-07-15	Completed	...	2	112	
2002-07-16	Completed	...	0	688	
2002-07-17	Completed	...	6	78	
2002-07-18	Completed	...	1	75	
2002-07-18	Completed	...	14	179	

AppliedDate	TotalPermitTime	AppliedYear	AppliedMonth	AppliedDay	Year	Month	\
2002-07-15	114	2002	7	15	2002	7	
2002-07-16	688	2002	7	16	2002	7	
2002-07-17	84	2002	7	17	2002	7	
2002-07-18	76	2002	7	18	2002	7	
2002-07-18	193	2002	7	18	2002	7	

AppliedDate	DayOfWeek	Quarter
2002-07-15	0	3
2002-07-16	1	3
2002-07-17	2	3
2002-07-18	3	3
2002-07-18	3	3

Methodology

Data collection

As stated earlier, the data set used for this project was obtained from Seattle open data which provides comprehensive data on all electrical permits issued or in process since the year 2002.

Data cleaning

The data was relatively clean except for the estimated project cost column that had lots of missing data. To clean the data and prepare it for analysis, the following procedures were followed:

- Different libraries were imported: pandas for data manipulation, Linear Regression from sklearn for modeling, and numpy for numerical operations.
- The dataset Electrical_permits.csv was loaded into a Data Frame df.
- All leading or trailing spaces in column names were cleaned to prevent issues with accessing columns along the line. Then column names were printed to verify they are correct and clean.
- The necessary columns (Permit Class, Permit Class Mapped, and Permit Type Mapped) were checked to ensure their presence in the Data frame and to raise an error if any of the columns were missing.
- The data frame was Split into non-missing data (rows where the estimated project cost not missing) and missing data (rows where estimated project cost is missing).
- Using get_dummies categorical variables (Permit Class, Permit Class Mapped, Permit Type Mapped) were converted to dummy/indicator variables in both the missing data and non-missing data dataframes.
- We made sure that both data frames had the same columns by reindexing missing data to match the columns of non-missing data, filling any missing columns with zeros.
- A linear regression model was initialized and trained using the predictor columns to predict estimated project cost.
- The trained model was used to predict estimated project cost for the rows with missing values.
- The missing estimated project cost values in the original data frame were replaced with the predicted values.
- The count of remaining missing values in the estimated project cost column was printed to ensure they have been filled.
- Backfill (bfill) and forward fill (ffill) methods were used to fill any other missing values in the data frame.

- The count of remaining missing values in the entire data frame was printed to ensure all missing values have been addressed.
- The cleaned data frame was saved to a new CSV file named Cleaned_Electrical_permits5.csv.

This process ensured that missing values in the estimated project cost column were predicted using a regression model based on related categorical variables, and any other missing values were filled using forward and backward filling techniques. The cleaned data was then saved for further use.

Feature engineering

Feature engineering was performed on the data Frame df by creating new columns based on the existing date columns. The following columns were created:

- ApplicationToIssueTime: Number of days between application and issue dates.
- IssueToCompletionTime: Number of days between issue and completion dates.
- TotalPermitTime: Total number of days from application to completion.
- AppliedYear: Year part of the application date.
- AppliedMonth: Month part of the application date.
- AppliedDay: Day part of the application date.

Data overview after cleaning and feature engineering

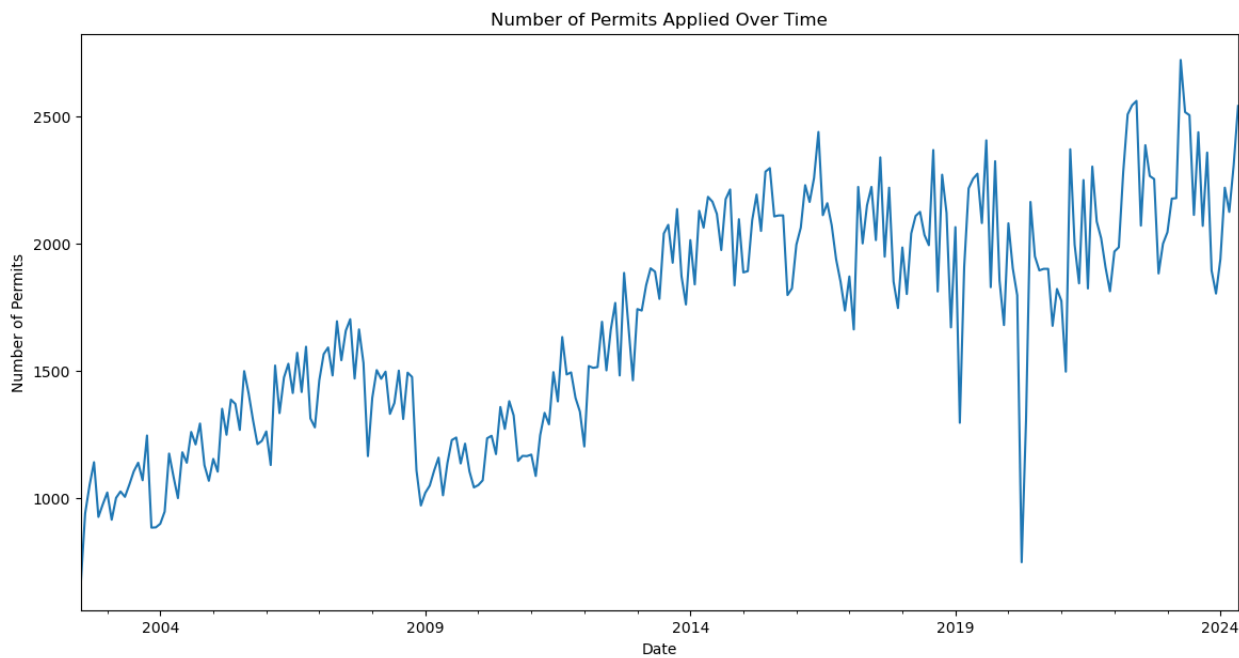
AppliedDate	PermitNum	PermitClass	PermitClassMapped	\	
2002-07-13	6001001-EL	Single Family/Duplex	Residential		
2002-07-13	6001002-EL	Institutional	Non-Residential		
2002-07-13	6001003-EL	Commercial	Non-Residential		
2002-07-13	6001004-EL	Commercial	Non-Residential		
2002-07-13	6001005-EL	Single Family/Duplex	Residential		
AppliedDate	PermitTypeMapped	\			
2002-07-13	Electrical				
2002-07-13	Electrical				
2002-07-13	Electrical				
2002-07-13	Electrical				
2002-07-13	Electrical				
AppliedDate	Description	\			
2002-07-13	125 AMP SERVICE AND PICK-UP EXISTING CIRCUITS;...				
2002-07-13	REMOVE/REINSTALL CLASS ROOM, CORRIDOR LIGHT FI...				
2002-07-13	PROVIDE CIRCUIT FOR NEW MIXER AND HOOK-UP.				
2002-07-13	BASEMENT LIGHTING, SWITCHING, RECEPTACLES, 1ST...				
2002-07-13	NEW 200 AMP SERVICE AND REWIRE FOR REMODEL AND...				
AppliedDate	EstProjectCost	IssuedDate	ExpiresDate	CompletedDate	\
2002-07-13	21198.061491	2002-07-13	2002-07-13	2005-03-10	
2002-07-13	300674.061491	2002-07-13	2003-07-13	2002-08-07	
2002-07-13	106070.061491	2002-07-13	2003-07-13	2004-06-03	
2002-07-13	106070.061491	2002-07-13	2003-07-13	2002-11-06	
2002-07-13	21198.061491	2002-07-13	2003-07-13	2003-07-10	

AppliedDate	StatusCurrent	...	ApplicationToIssueTime	IssueToCompletionTime	\
2002-07-13	Completed	...	0	971	
2002-07-13	Completed	...	0	25	
2002-07-13	Completed	...	0	691	
2002-07-13	Completed	...	0	116	
2002-07-13	Completed	...	0	362	

AppliedDate	TotalPermitTime	AppliedYear	AppliedMonth	AppliedDay	Year	Month	\
2002-07-13	971	2002	7	13	2002	7	
2002-07-13	25	2002	7	13	2002	7	
2002-07-13	691	2002	7	13	2002	7	
2002-07-13	116	2002	7	13	2002	7	
2002-07-13	362	2002	7	13	2002	7	

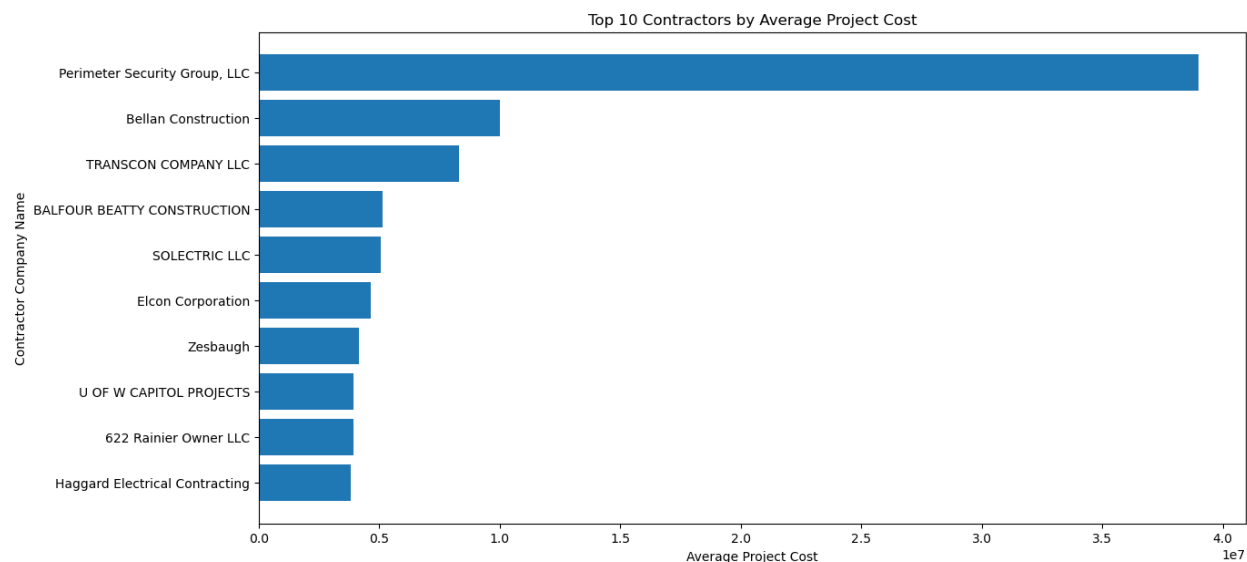
AppliedDate	DayOfWeek	Quarter
2002-07-13	5	3
2002-07-13	5	3
2002-07-13	5	3
2002-07-13	5	3
2002-07-13	5	3

Exploratory Data Analysis



The number of permits sought for shows a distinct rising trend over time, indicating a rise in permit applications over time. There are variations in the number of licenses applied for within specific time

periods, as evidenced by the non-linear trend. Over the whole time, there are observable highs and lows. This means that the number of permits issued varies according on the season or other cyclical factors. There is a noticeable decline in 2009 and a further notable decline in 2020, which could be connected to economic occurrences such as the COVID-19 epidemic and the global financial crisis. Though there are still oscillations, starting in 2014, the number of permits stabilizes at a higher level than in prior years. Near the end of the timeframe, roughly in 2023–2024, there is a peak that denotes a large number of recently submitted permits. The global impact of the COVID-19 epidemic, which impacted numerous businesses, including construction and other sectors requiring permits, is probably to blame for the rapid decline that began around 2020. In conclusion, the graph shows a long-term rise in the quantity of permit applications, with notable short-term variations that are probably caused by outside and economic variables.



Here is a bar graph that shows the "Top 10 Contractors by Average Project Cost."

At almost \$40 million, Perimeter Security Group, LLC is the company with the highest average project cost. Bellam construction trails closely behind Perimeter Security Group, LLC, with an average project cost that is somewhat less. The fact that these two contractors have substantially higher project expenses than the rest suggests that they are capable of handling more complicated or large-scale projects. At \$15 million to \$20 million on average, Transcon Company LLC and Balfour Beatty Construction had the next highest project costs. The fact that these contractors are grouped in the center indicates that their project sizes are significant but not elite.

Haggard Electrical Contracting comes in at number ten, with an average project cost that is the lowest of the top ten, but still close to \$5 million. With typical project prices of less than \$10 million, 622 Rainier

Owner LLC and U OF W Capitol Projects are likewise seen as being on the lower end. The graph shows how the average project costs of the top two contractors fell sharply from there to the bottom of the list. The dominance of high-cost projects by Perimeter Security Group, LLC and Bellan construction suggests that these companies may specialize in large-scale or highly specialized projects. The center of the pack includes contractors like Balfour Beatty Construction and Solelectric LLC, indicating that they work on a variety of mid- to large-scale projects. There is a skewness in the average project cost distribution among these ten top contractors; some of them handle much more expensive projects than others. The chart can be used to determine which contractors are most likely to be working on large-scale projects and may also serve as a sign of their specialization or experience with particular kinds of work. For stakeholders wishing to hire contractors for projects with certain financial specifications, this chart offers helpful information.



The expected project expenditures are shown in this box plot chart, which is divided into various permission classes.

Class of Commercial Permit: The largest variety of project expenses may be seen in this category, with a few outliers reaching sums of \$100 million or more. Comparatively speaking to other categories, the median cost seems to be higher.

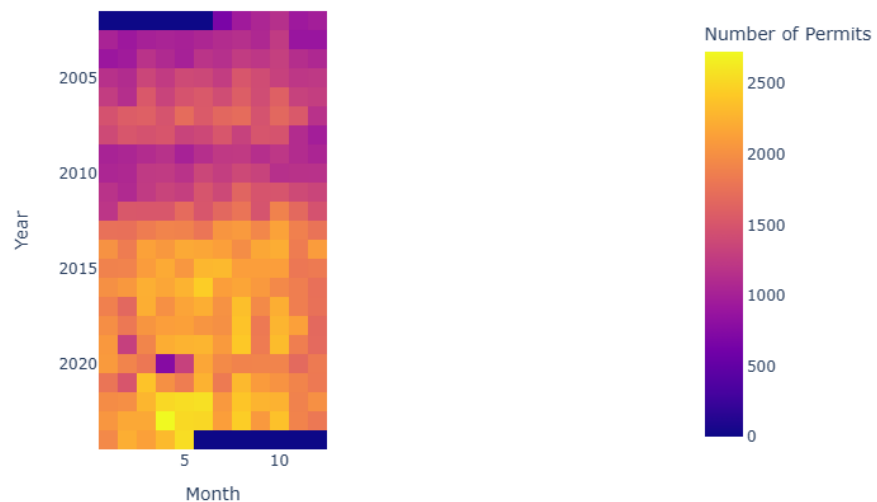
Multifamily: An additional group with a broad distribution and a few notable outliers; The anticipated costs of some projects exceed 150 million.

Vacant Land, Single Family/Duplex, Industrial, and Institutional: There are fewer outliers and lower median expenses in these areas. Within these groups, they are more closely clustered, suggesting more constant project costs.

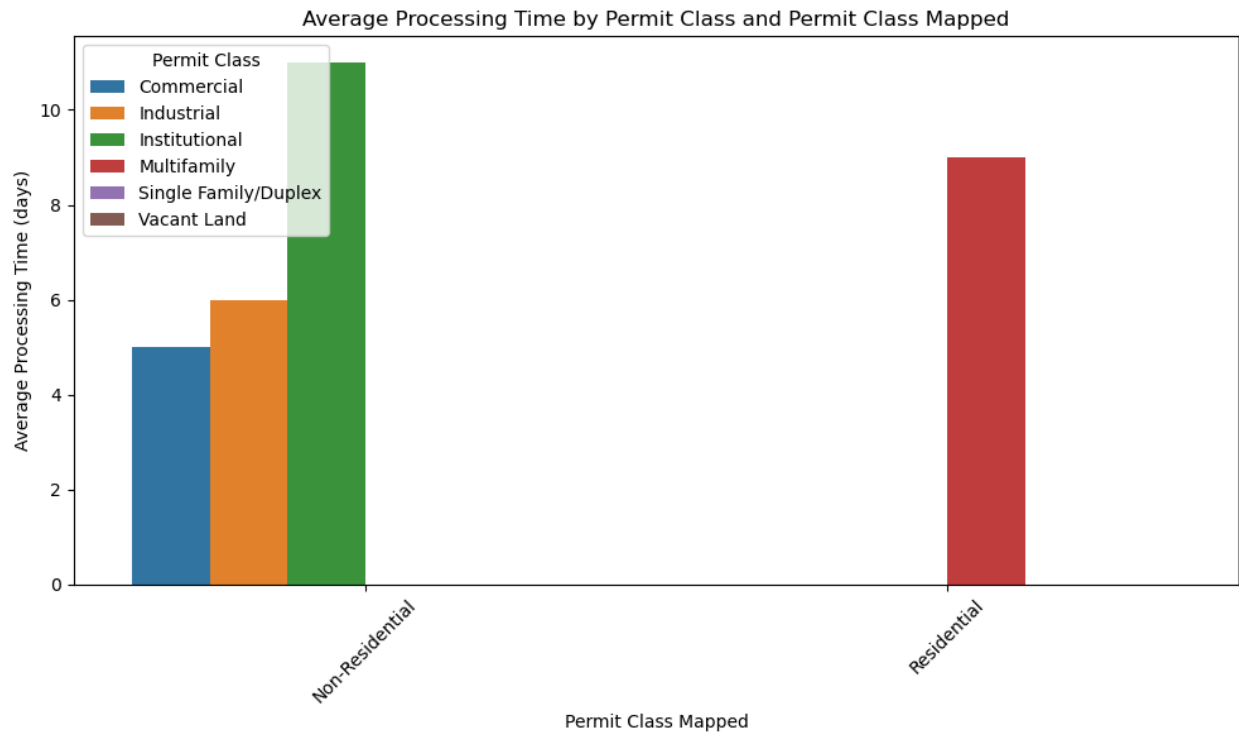
Unoccupied Space: The predicted project costs in this category are the lowest and most closely clustered.

It looks that project costs are highest and most variable in the commercial and multifamily categories, with some very expensive projects indicated by significant outliers. The project prices in the other categories are less exorbitant and more consistent.

Heatmap of Permits by Year and Month



This heatmap visualizes the number of permits issued over different years and months. Here's a detailed analysis of the chart: Recent years (closer to 2022) generally show more activity (more yellow and orange), indicating an increase in the number of permits issued in recent times. Earlier years (closer to 2005) have more purple and dark blue, suggesting fewer permits were issued during these times. Certain months show more activity across multiple years. For example, the mid-year months (around month 6-7) tend to have more permits issued than the beginning or end of the year. There is a noticeable drop in permit issuance around 2020, which might be related to the COVID-19 pandemic. Some months and years repeatedly show high activity, suggesting possible seasonal or cyclical factors influencing permit issuance.

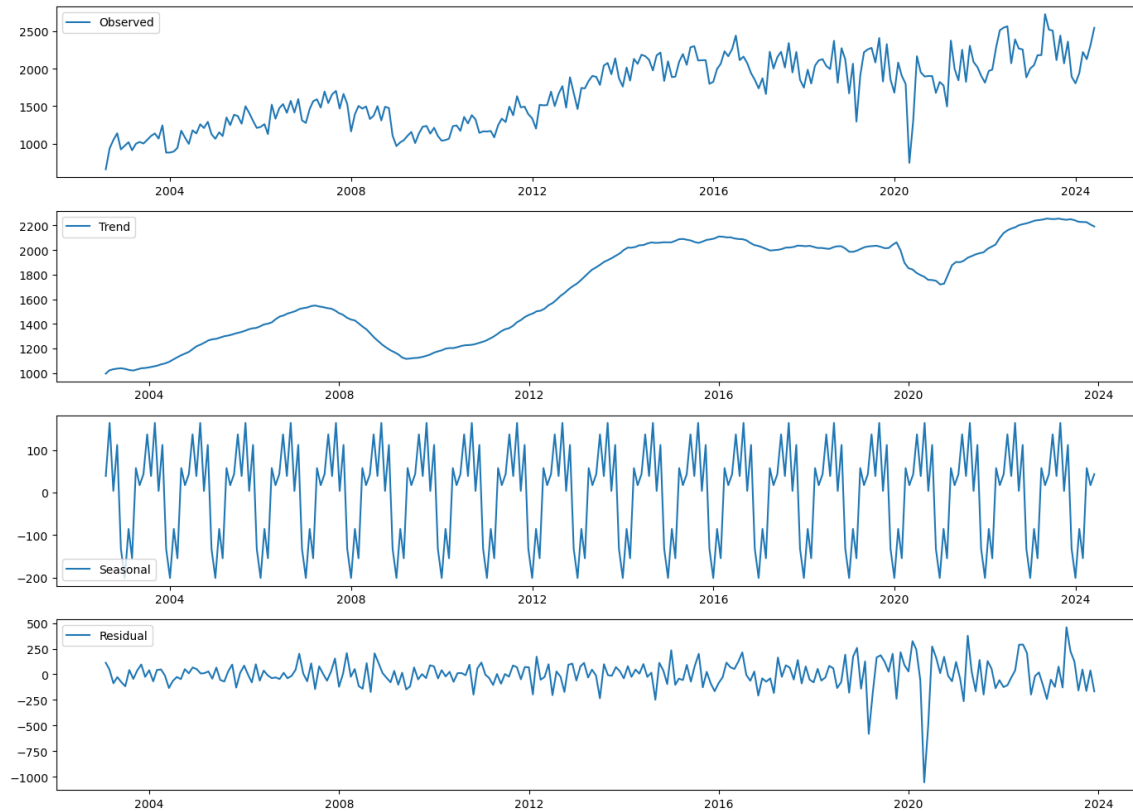


This bar graph illustrates the "Average Processing Time by Permit Class and Permit Class Mapped," emphasizing the duration required to process permits in various categories.

At about 11 days, the Institutional (green bar) non-residential permit has the longest average processing time. Industrial (orange bar) comes next, taking roughly 7 days on average to process. In the non-residential category, Commercial (blue bar) has the fastest processing time, taking about 5 days. For single family/duplex, and vacant land, the processing time is in hours. While for multifamily building, the processing time for electrical permits is 8 days. In order to shorten processing times, this chart offers a clear picture of potential time lags in the permit processing pipeline and may be useful for allocating resources or implementing process improvement projects.

Trend Analysis

Trend analysis involves examining historical data to identify patterns and peak periods in permit issuance. This analysis helps in understanding the temporal distribution of permits and any underlying factors driving these trends.



The image displays the decomposition of a time series into its three main components: Trend, Seasonal, and Residual:

1. Observed

This plot shows the original time series data. It represents the raw data as it was recorded over time. The observed data appears to show an increasing trend with some seasonal fluctuations and variations. There are notable spikes and drops, especially around the year 2020.

2. Trend

This plot represents the underlying trend in the data. The trend component is the long-term movement in the data, showing the overall direction in which the data is moving. The trend shows a general increase

from 2000 to around 2008, followed by a decline and then a steady rise again from around 2012 to 2024. This indicates long-term growth with a dip during the middle period.

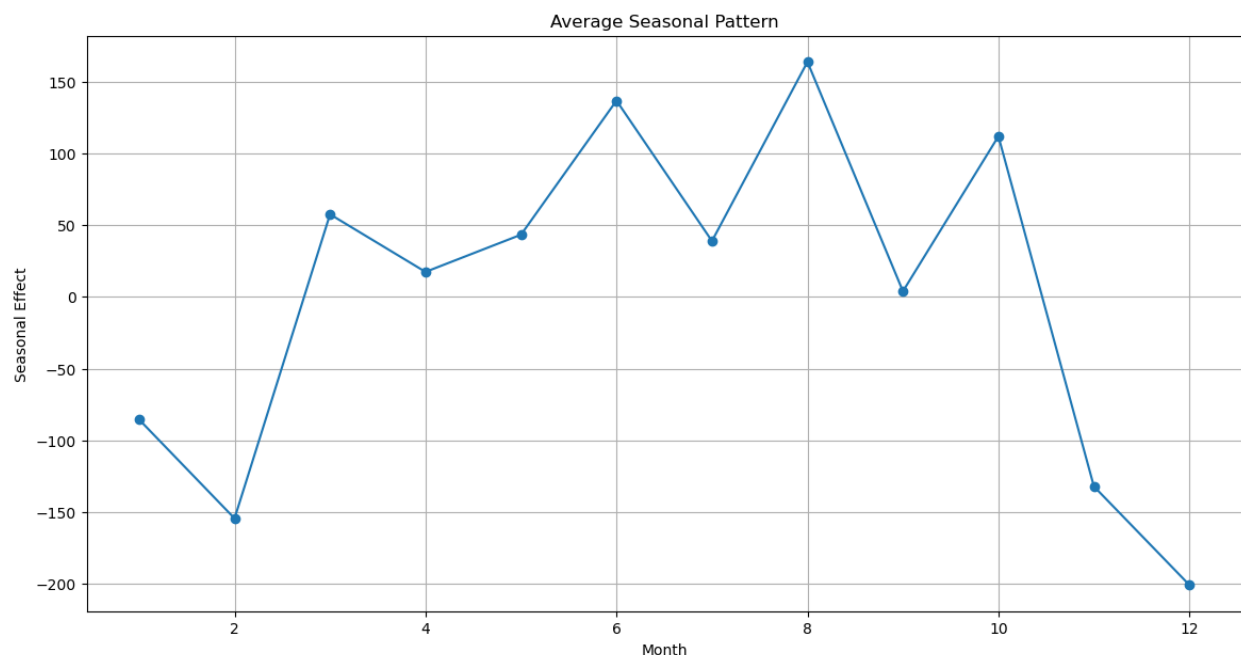
3. Seasonal

This plot shows the seasonal component of the data. Seasonal variations are patterns that repeat at regular intervals, such as monthly or yearly cycles. The seasonal plot displays regular, repeating patterns over each year, indicating consistent seasonal effects throughout the period. These patterns repeat in a predictable manner, suggesting a strong seasonal influence in the data.

4. Residual

This plot shows the residual or irregular component, which is what remains after removing the trend and seasonal components from the observed data. The residuals represent random noise and any other irregular variations. The residual plot shows fluctuations around zero, with some larger deviations especially noticeable around 2020. This indicates periods of irregular behavior that are not captured by the trend or seasonal components, possibly due to unusual events or anomalies.

This decomposition helps in understanding the different underlying patterns in the time series data and is useful for forecasting, anomaly detection, and analyzing the behavior of the data over time.



Insights from the Seasonal Pattern

1. **January to February:** There is a significant drop in the seasonal effect in January and it reaches its lowest point in February. This indicates a seasonal low during these months.
2. **March to June:** The seasonal effect increases sharply from February to March and continues to rise, peaking in June. This suggests a strong upward seasonal effect during the spring and early summer months.
3. **July to August:** The effect fluctuates but remains relatively high during these months. This indicates that the mid-summer period still maintains a strong seasonal influence.
4. **September:** There is a dip in the seasonal effect in September, indicating a slight decrease in seasonal activity.
5. **October:** The effect rises again in October, showing a temporary increase.
6. **November to December:** There is a sharp decline from October to December, reaching the second lowest point in December, indicating a seasonal low at the end of the year.

Predictive Modeling

The machine learning technique known as XGBoost, or eXtreme Gradient Boosting, is a potent and popular tool that is particularly effective in settings involving structured or tabular data. It operates by building an ensemble of decision trees, each of which tries to rectify the mistakes caused by the preceding trees. It is a member of the gradient boosting algorithm family. XGBoost is especially renowned for its effectiveness, speed, and capacity to manage big datasets, missing values, and intricate correlations between variables.

Variables Used in the XGBoost Model

In this model, the following variables were used as features to predict the target variable, which is likely related to the timing of permit issuance:

1. **ApplicationToIssueTime:** The time taken from the application submission to the issuance of the permit. This variable captures the efficiency of the permit process and could be influenced by various factors such as the complexity of the application and the workload of the issuing body.
2. **IssueToCompletionTime:** The time between the issuance of the permit and the completion of the project. This variable helps understand the duration of the actual work and may be affected by project size, complexity, and external factors like weather.

3. **TotalPermitTime:** The total time from the application submission to the project's completion. It is a cumulative measure that provides an overall view of the permit process's duration.

4. **AppliedYear:** The year in which the permit application was submitted. This variable could capture trends over time, such as changes in regulations, economic conditions, or seasonal patterns.

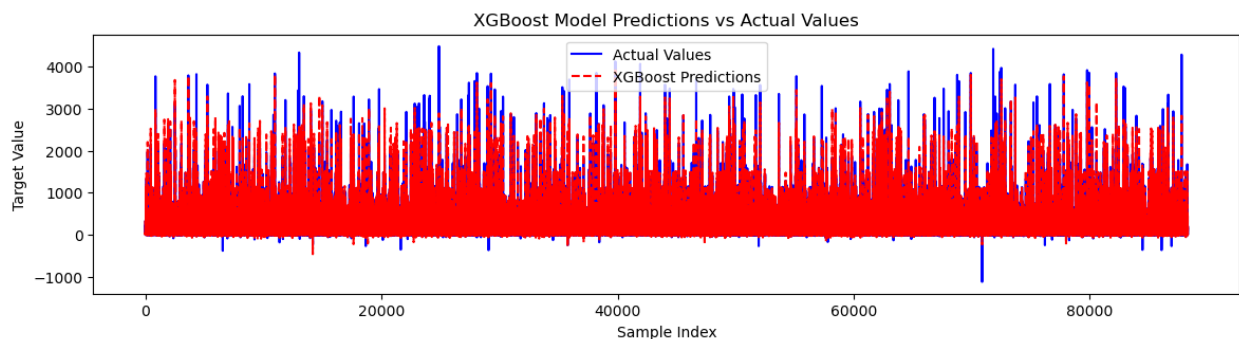
5. **AppliedMonth:** The month in which the permit application was submitted. This variable can reveal seasonal trends, as certain types of construction may peak during specific months.

6. **AppliedDay:** The day of the month the application was submitted. This variable could be useful if there are day-specific patterns, such as deadlines or peaks in applications at the beginning or end of the month.

Target Variable

TotalPermitTime: The model is designed to predict the total time required for a permit to go through the application, issuance, and completion stages. This variable is crucial for understanding and optimizing the permit process, providing insights into areas where improvements can be made to reduce delays.

Model results



XGBoost MAE: 3.2917100662335854

XGBoost MSE: 1253.4774678480308

XGBoost RMSE: 35.40448372520112

XGBoost R-squared: 0.9805822024416042

The high R-squared value (0.9806) shows that the model explains a significant amount of variance in permit issuance, meaning it has captured the patterns in the data very well.

The relatively low MAE (3.29) and RMSE (35.40) indicate that the model's predictions are close to the actual values, with only a small average error. The MAE is particularly useful here because it gives a clear idea of the average error in the same units as the target variable.

The MSE is higher than MAE because it squares the errors, emphasizing larger errors more. This indicates there might be some outliers or particularly challenging predictions in the dataset.

The XGBoost model appears to be a robust and reliable predictor of future permit issuance, with excellent predictive power as indicated by the high R-squared value and relatively low error metrics. However, the MSE suggests there may be some outliers or more complex cases where the model's predictions are less accurate. Despite this, the model is likely to perform very well in practical applications for forecasting permit issuance

Cross validated result

Cross-Validated MAE: 7.550955058813495

Cross-Validated MSE: 3546.339712550991

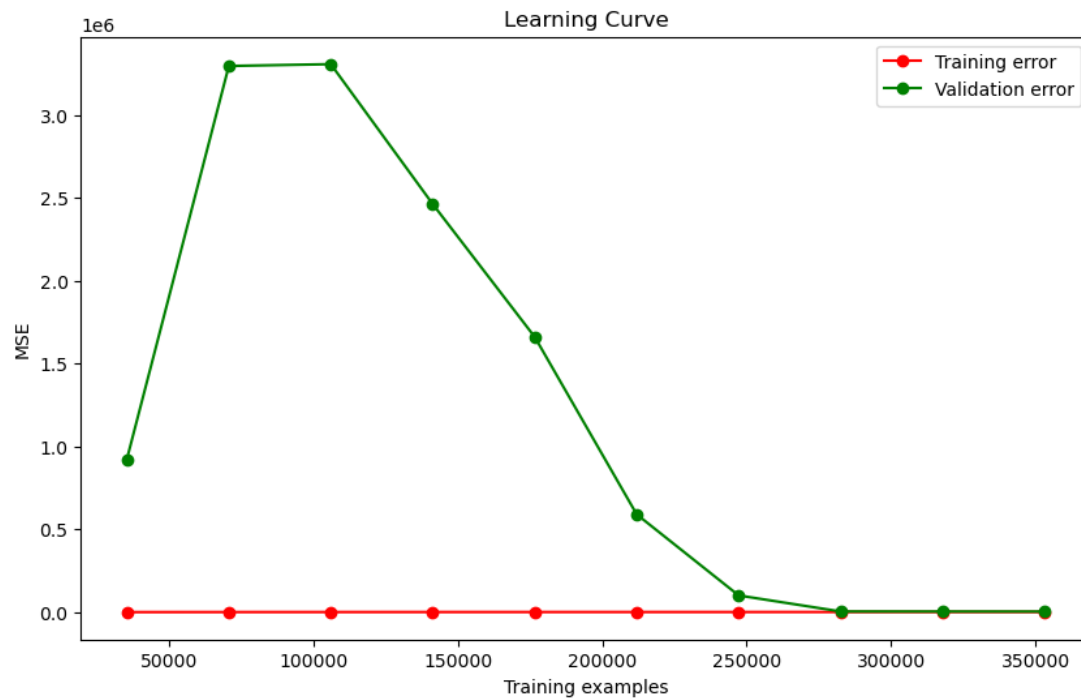
Cross-Validated R-squared: 0.9496339383761738

The cross-validation results are slightly less favorable than the initial train-test split results, which is common because cross-validation provides a more rigorous test of the model's generalization capability.

The increase in MAE and MSE indicates that there are certain subsets of the data where the model struggles a bit more, possibly due to variability in those subsets that the model wasn't fully able to capture.

Despite the slight increases in error metrics, the R-squared value remains very strong, confirming that the model generally fits the data well and can explain most of the variance.

Learning curve



For all training set sizes, the training error is nearly constant and minimal. This shows that there is very little inaccuracy in the model's fit to the training set. It might also imply that the model is overfitting, particularly if the training error is nearly zero.

The validation error starts relatively high and increases sharply as the training size increases up to around 100,000 examples. This increase might suggest that the model initially struggles to generalize from a small amount of training data.

After the peak, the validation error begins to decrease steadily as more training data is added. This behavior suggests that as the model is exposed to more data, it learns to generalize better.

Towards the end, the validation error drops significantly and begins to flatten out as it approaches the training error. The convergence of the training and validation errors at a low level suggests that with sufficient data, the model is able to generalize well to unseen data.

The flat and low training error line, combined with a high initial validation error, suggests that the model might initially overfit the training data. However, as the training size increases, this overfitting issue diminishes.

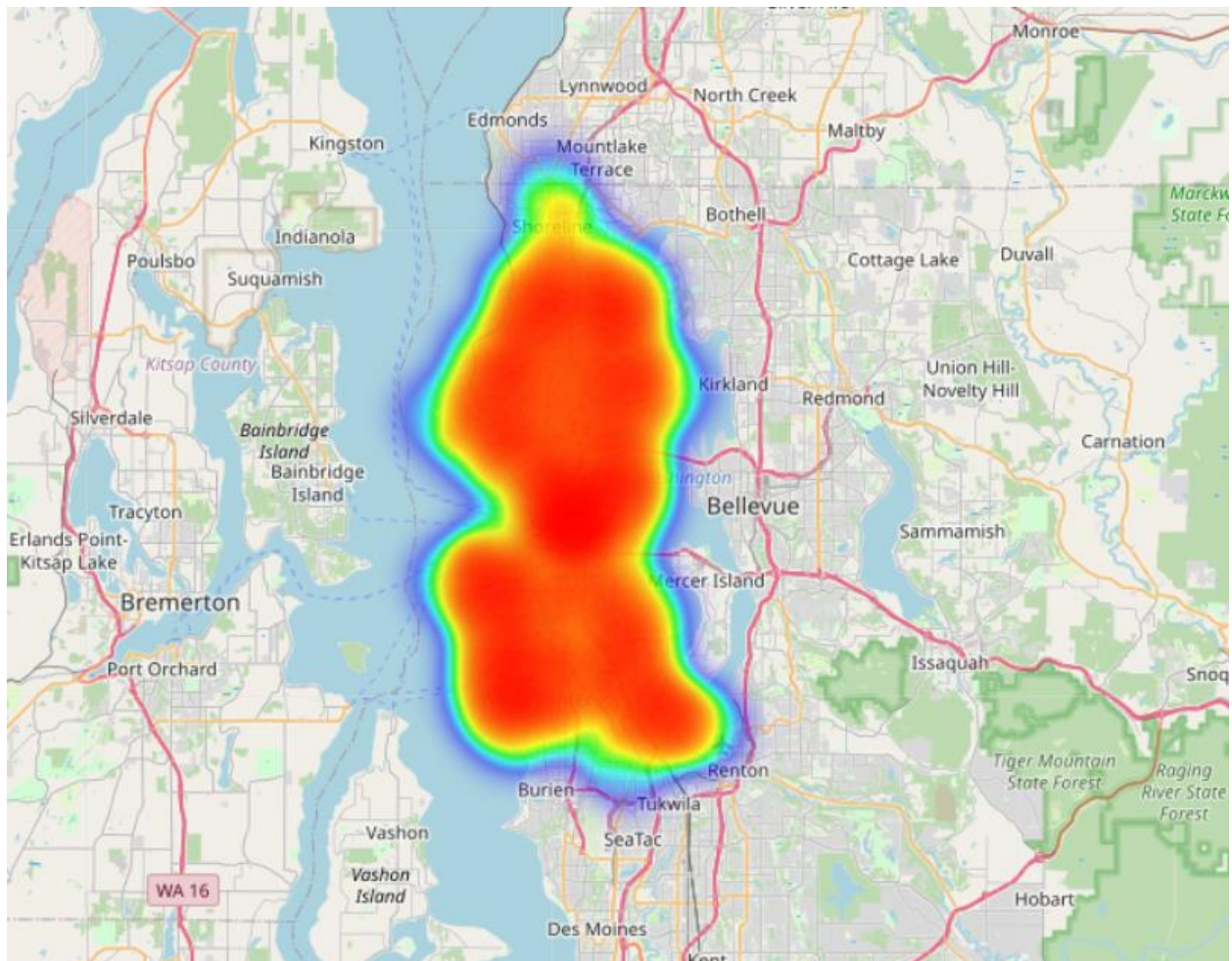
The learning curve shows that more training data significantly improves the model's performance. Beyond a certain point (around 250,000 examples), adding more data provides diminishing returns, as the validation error stabilizes.

The final gap between the training and validation error is minimal, which is a good sign. It suggests that with a large enough training set, the model is not overfitting and is likely to perform well on

Overall, the learning curve suggests that your model benefits greatly from a larger dataset and is likely to perform well with a sufficient amount of training data.

Geospatial Analysis

Geospatial analysis helps in identifying hotspots for electrical permits and visualizing their spatial distribution. By mapping the permits' locations, city planners can identify areas with high permit activity and allocate resources more effectively.



This heatmap represents the distribution of electrical permits across the Seattle metropolitan area. The areas with the most intense red coloration are potential hotspots where there's a higher concentration of electrical permit activity.

City planners can use this data to allocate inspection resources more effectively, focusing on the areas with the highest permit activity. Identifying hotspots can help in planning upgrades or maintenance for electrical infrastructure, ensuring that areas with rapid development or higher electrical activity are adequately supported. The heatmap can aid in predicting where future growth might occur, allowing for proactive measures in infrastructure and service provisioning. If this map is used alongside other data, like construction trends or population growth, it could provide even more insights into the underlying causes of these hotspots.

Dashboard

A dashboard for city planners will allow them to monitor current permit statuses, processing times, and compliance was created.

A Python script for a Streamlit web application designed to analyze and forecast electrical permit data in Seattle was made. The script imports several libraries including `'streamlit'` for creating the web app, `'pandas'` and `'numpy'` for data manipulation, `'sklearn'` for model evaluation metrics, `'pmdarima'` for time series forecasting, `'matplotlib'` for plotting, `'folium'` for map visualization, and `'Prophet'` for time series forecasting.

The script used in the dashboard acts a comprehensive tool for analyzing electrical permit data, providing city planners with insights into trends, costs, top contractors, permit classes, processing times, and geospatial distributions, while also forecasting future permit activities.

Significance and Impact

The findings from this project are expected to have significant implications for various stakeholders. For city planners, the insights can help in better resource planning and allocation. For contractors, understanding permit trends can aid in project planning and management. Policy-makers can use the data-driven insights to formulate policies that streamline the permitting process and address emerging challenges.

Moreover, the predictive models developed in this project can serve as valuable tools for anticipating future permit demands, allowing for proactive measures to accommodate growth and ensure compliance with safety standards. Ultimately, this project contributes to a safer, more efficient, and well-planned urban environment in Seattle.

Conclusion

In conclusion, the Seattle Electrical Permits project leverages data analytics to provide a comprehensive understanding of electrical permit issuance trends and their underlying factors. By combining historical data analysis with predictive modeling, actionable insights were gotten that can enhance the efficiency and effectiveness of the permitting process. The ultimate goal of supporting the city's growth and development while ensuring the safety and reliability of its electrical infrastructure can be met if the insights from this analysis are put to use.

References

1. Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2), 93-115.
2. Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics*, 214(1), 481-518.
3. Chatfield, C. (2004). *The Analysis of Time Series: An Introduction*. Chapman & Hall/CRC.
4. City of Chicago. *OpenGrid - Case Study*.
5. City of Seattle. (2023). Electrical Permit Process and Regulations. Retrieved from [seattle.gov](https://www.seattle.gov/)
6. Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3), 189-206.
7. Goodchild, M. F. (2018). Geographical data science and spatial data analytics: An evolving research agenda. *Foundations and Trends in Databases*, 10(1-2), 1-92.
8. Gyourko, J., & Saiz, A. (2006). Construction costs and the supply of housing structure. *Journal of Regional Science*, 46(4), 661-680.
9. Hsiao, H., Lin, S., & Wei, C. (2020). Improving the accuracy of construction permit forecasts using machine learning approaches. *Automation in Construction*, 120.
10. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
11. Lee, S., Lee, C., & Park, H. (2017). Seasonal effects on construction activity and labor productivity. *Journal of Construction Engineering and Management*, 143(5).
12. Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.
13. MacEachren, A. M. (1994). *Some Truth with Maps: A Primer on Symbolization and Design*. University of Chicago Press.
14. New York City Planning Labs. *NYC Planning Labs - Case Study*.
15. Nielsen, J. (1993). *Usability Engineering*. Academic Press.
16. Seattle Department of Construction and Inspections. (2023). Trends in construction and permitting. Retrieved from [seattle.gov](https://www.seattle.gov/sdci/)

17. Seattle Department of Construction & Inspections - Electrical Permit Information: [seattle.gov](https://www.seattle.gov/sdci/permits/permits-we-issue-(a-z)/electrical-permit)
18. Seattle Open Data Portal. (2023). Electrical Permits Dataset. Retrieved from data.seattle.gov
19. Seattle Services Portal - Help Center: seattlegov.zendesk.com
20. SDCI Permit Cost Estimator: [seattle.gov](https://www.seattle.gov/sdci/permits/how-much-will-your-permit-cost)
21. Shneiderman, B. (2004). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson Education.
22. Subramani, K., & Subramani, A. (2024). Management of electrical appliances for smart home to regulate energy consumption using Arduino and GSM.
23. Washington State Department of Commerce. (2023). Urban growth and development in Seattle. Retrieved from [commerce.wa.gov](https://www.commerce.wa.gov/)