Forecasting Electric Vehicle Adoption and Charging Infrastructure Demand

Aruna Jithesh, Chinmayi Karumuri, Dilip Kumar Kasina, Meghana Doddapuneni, Shanmukha Raj ,Vaishnavi Samboji

Abstract— **The rapid expansion of electric vehicles in Washington State emphasizes the crucial need for scaleable and equitable Charging Infrastructure to drive range anxiety off the table and to meet emission reduction goals. This study combines several Machine Learning techniques, SARIMAX, and Auto ARIMA in a Time-Series Forecast to analyze more than 1 million records about EV registrations from Data.gov and data from 2,759 charging station sites provided by the NREL API. This study researched trends across counties from 2010 to 2024 to identify the areas most in need, showing disparity between urban and rural regions. Urban counties like King have higher rates of EV Adoption and dense infrastructure, while rural counties are falling behind, exposing gaps in investments. Key metrics, such as "EVs per Charger," will help drive strategic decisions, enhance Charging Infrastructure, and ensure equity. This research aligns with the goals of Sustainable Transportation set by Washington State and offers actionable insights for policymakers, businesses, and stakeholders. It leverages data-driven solutions to address infrastructure gaps, underlining the transformational potential of Machine Learning in supporting fair EV Adoption and advancing Sustainable Transportation planning. The proposed methodology will be scalable and thus can be adopted in other regions to foster data-informed EV infrastructure development.**

Key words **—Electric Vehicles, Charging Infrastructure, Machine Learning, SARIMAX, Auto ARIMA, Time-Series Forecasting, Sustainable Transportation, EV Adoption**

I. INTRODUCTION

The uptake of electric vehicles (EVs) has increased considerably, acting as a key element in cutting down greenhouse gas emissions, particularly in the transportation sector. While earlier investigations, like those by Khaki et al. (2019), examined machine learning methods to enhance EV charging schedules, these analyses mainly concentrated on current infrastructure without anticipating future requirements or tackling regional inequalities. Expanding on this groundwork, our initiative utilizes SARIMAX and Auto ARIMA models to forecast future trends in EV adoption and the demand for charging stations in Washington State. By leveraging over 1 million EV registration data (2010–2024) along with information on 2,759 charging stations, we deliver detailed, county-specific forecasts that highlight areas lacking adequate service, especially in rural locales. To accomplish this, we targeted the top 10 counties with the highest EV purchase numbers, utilizing over 1 million EV registration records from 2010 to 2024. Furthermore, data from 2,759 currently operational charging stations was incorporated into our assessment. Through a time-series modeling technique, we projected the anticipated number of EVs in each county and evaluated these estimates against the existing charging infrastructure. By pinpointing areas with insufficient service, particularly in rural settings, we facilitate a focused strategy for charging station locations. Our models, which employ Auto ARIMA and SARIMAX, accurately project EV demand throughout different counties, allowing for data-informed decisions regarding resource distribution and fair planning of charging infrastructure. This proactive strategy reinforces Washington's climate objectives and creates a scalable model for the development of sustainable EV infrastructure. The methodology used not only fills the voids in current research but also coincides with Washington's goals for transportation electrification by emphasizing underserved regions and enhancing the durability of the charging infrastructure.

A. *Problem Statement*

The swift uptake of electric vehicles (EVs) in Washington State requires predictive models to anticipate future adoption patterns and charging station needs, guaranteeing fair access to infrastructure. This initiative seeks to pinpoint underserved regions and direct resource distribution by examining historical EV registration data and present charging station placements. Although machine learning models such as SARIMAX and Auto ARIMA deliver practical insights with considerable accuracy, obstacles like incomplete datasets, regional inequities, and possible biases need to be confronted. Moreover, ethical factors and advancing EV technologies highlight the necessity of integrating these results with human knowledge to ensure equity, dependability, and alignment with long-term sustainability objectives.

B. *Project Background*

The electric vehicle (EV) registration dataset offers an extensive overview of EV adoption patterns throughout Washington State from 2010 to 2024, containing over 1 million records and 33 attributes. This dataset, along with charging station information from the NREL API, provides valuable insights into regional differences in EV adoption and the distribution of infrastructure. Analyzing this information allows policymakers, businesses, and researchers to pinpoint areas that lack service, evaluate the efficiency of existing infrastructure, and formulate strategies for fair resource distribution. Moreover, examining these trends aids in broader conversations regarding sustainable transportation and the socio-economic effects of moving to EVs. The results can also act as a reference framework for other states or regions looking to meet emission reduction targets and encourage widespread EV adoption.

II. PROPOSED METHOD

The diagram displayed in Figure 1 offers a visual depiction of the machine learning model development lifecycle. We are incorporating the concepts of the Cross-Industry Standard Process for Data Mining (CRISP-DM), which is a six-phase model designed for data mining initiatives. The process commences with Data Collection, corresponding to CRISP-DM's initial phase of Business Understanding, during which objectives are outlined, and data needs are determined. This phase emphasizes grasping project aims, such as predicting EV adoption trends and pinpointing underrepresented areas, and translating this insight into a formalized problem statement and an initial strategy for realizing these goals.

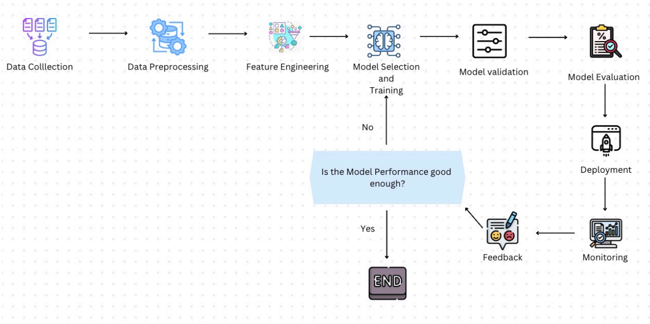

Fig. 1. The diagram of the machine learning model development lifecycle.

Following collection phase, the next stage is Data Preprocessing. The gathered data is examined to obtain insights and is then cleaned and formatted, resolving issues like missing values, duplicates, and inconsistencies in both temporal and categorical attributes. Ensuring that the dataset is dependable and uniform involves standardizing date formats, eliminating duplicates, and addressing missing values by either filling them in or removing them. The following phase, Feature Engineering, entails utilizing domain knowledge to derive and create pertinent features from the raw data. Features that provide valuable information, such as "EVs per Charger," temporal elements including year, month, and day, as well as the categorical encoding of vehicle types, are developed. This step represents the essential part of the Data Preparation phase in CRISP-DM, guaranteeing that the dataset is tailored for machine learning algorithms.

Model Selection and Training is the subsequent phase, where time-series forecasting models like SARIMAX and Auto ARIMA are assessed and trained with the processed dataset. This phase corresponds to the Modeling stage in CRISP-DM, concentrating on choosing the most appropriate algorithms and fine-tuning their parameters for improved performance. Model Validation verifies that the models perform adequately on new data, employing evaluation metrics such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and residual diagnostics to evaluate their efficacy. These processes assist in accurately capturing the underlying patterns of the data, facilitating iterative enhancements whenever necessary.

The Model Evaluation phase, corresponding to the Evaluation stage of CRISP-DM, verifies that the models meet the project's goals, like accurately forecasting EV trends and pinpointing infrastructure deficiencies. This stage starts with Seasonal Decomposition to reveal trends, seasonality, and residual patterns within the data. The dataset is partitioned into training and testing subsets to assess the model's effectiveness on previously unseen data, ensuring durability. Auto ARIMA facilitates parameter tuning for the best outcomes, while SARIMAX models are set up to encompass seasonal and external effects. The models are subsequently evaluated using predictions from test data, and confidence intervals are scrutinized to address forecasting uncertainty. Ultimately, the models are retrained on the complete dataset to improve future predictions. If performance falls short, the process reverts to the Modeling stage for enhancement. Once the models exhibit acceptable results, they are launched for practical applications, providing actionable insights, accompanied by ongoing monitoring and feedback to adapt to changing trends and maintain long-term significance.

This combination of the machine learning workflow with CRISP-DM guarantees a systematic and repetitive method for model creation, highlighting both technical efficiency and consistency with business goals. By tackling issues like insufficient data and differences in infrastructure, this initiative provides a strong framework for data-driven decision-making aimed at promoting fair and sustainable EV infrastructure planning.

III. Data Preparation

Our data preparation method commenced with gathering dependable information from authoritative governmental sources, such as EV registration data from Data. gov and charging station details from the NREL API. To guarantee precision and uniformity, unnecessary columns like base MSRP and transaction_year were eliminated, while absent values in essential fields like county were dealt with by discarding incomplete rows. Duplicate records, particularly those only varying in fields such as county or city, were managed by keeping the most recent entry to ensure relevance. Temporal fields, including transaction_date, were normalized and transformed into a datetime format for efficient trend evaluation. This thorough cleaning and preprocessing assured the dataset's quality, preparing it for analysis and predicting future EV adoption trends and charging infrastructure needs.

A. Historical Dataset Description

The dataset utilized for this project encompasses "Electric Vehicle Title and Registration Data" from Data. gov along with charging station information from the National Renewable Energy Laboratory (NREL) API. The EV dataset, formatted in CSV and exceeding 100MB in size, contains 1,048,576 records and 33 attributes, reflecting electric vehicle registrations in Washington State between 2010 and 2024. The charging station dataset features 2,759 operational stations complete with location and capacity information. Although the EV registration data is consistently refreshed, certain fields, like county, have missing values, which were addressed during the preprocessing phase. The charging station data offers spatial insights but is devoid of real-time usage data. While derived from official sources, minor errors, such as transcription mistakes or reporting discrepancies, might be present. In spite of these challenges, the datasets are trustworthy and deliver a solid basis for predicting EV adoption trends and evaluating the adequacy of charging infrastructure. Figure 2 presents a detailed account of the dataset's columns, featuring column names, data types, and non-null counts for each attribute, thereby ensuring a comprehensive grasp of the dataset's organization.

	Column	Non-Null Count	Dtype
0	electric_vehicle_type	948,994	object
1	vin_1_10	948,994	object
2	dol_vehicle_id	948,994	int64
3	model_year	948,994	int64
4	make	948,994	object
5	model	948,994	object
6	vehicle_primary_use	948,994	object
7	electric_range	948,984	float64
8	odometer_reading	948,994	int64
9	odometer_code	948,994	object
10	new_or_used_vehicle	948,994	object
11	sale_price	948,994	float64
12	date_of_vehicle_sale	264,908	object
13	base_msrp	948,984	float64
14	transaction_type	948,994	object
15	transaction_date	948,994	object
16	transaction_year	948,994	int64
17	county	948,977	object
18	city	948,952	object
19	state_of_residence	948,994	object
20	zip	948,967	float64
21	hb_2042_clean_alternative_fuel_vehicle_cafv_eligibility	948,994	object
22	meets_2019_hb_2042_electric_range_requirement	948,994	bool
23	meets_2019_hb_2042_sale_date_requirement	948,994	bool
24	meets_2019_hb_2042_sale_price_value_requirement	948,994	bool
25	_2019_hb_2042_battery_range_requirement	948,994	object
26	_2019_hb_2042_purchase_date_requirement	948,994	object
27	_2019_hb_2042_sale_price_value_requirement	948,994	object
28	electric_vehicle_fee_paid	948,994	object
29	transportation_electrification_fee_paid	878,082	object
30	hybrid_vehicle_electrification_fee_paid	878,082	object
31	census_tract_2020	948,977	float64
32	legislative_district	947,512	float64
33	electric_utility	948,977	object

Fig. 2. The overview of the columns of the dataset.

B. Exploratory Data Analysis

During the exploratory data analysis stage of my project, different visualizations were created to uncover patterns and insights associated with electric vehicle adoption and infrastructure throughout Washington State.

The bar chart shows that 75% of the vehicles are Battery Electric Vehicles (BEVs), which operate exclusively on electric batteries, whereas Plug-in Hybrid Electric Vehicles (PHEVs), utilizing both electric and internal combustion engines, account for 24% of the dataset, as illustrated in Fig. 3. This underscores the increasing transition towards complete electrification in the market.

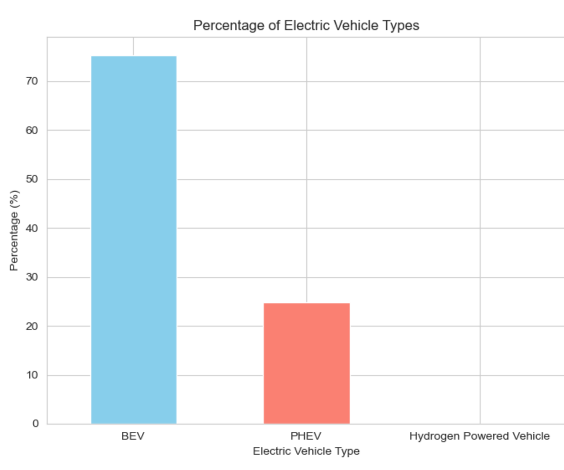

Fig. 3. Electric Vehicle Types Distribution

The rapid increase in electric vehicle adoption, illustrated in Fig. 4, from 2010 to 2024 showcases the influence of technological progress and policy encouragement. The addition of a 6-month rolling average validates a steady upward trajectory, highlighting possibilities for infrastructure investments to support this expansion.

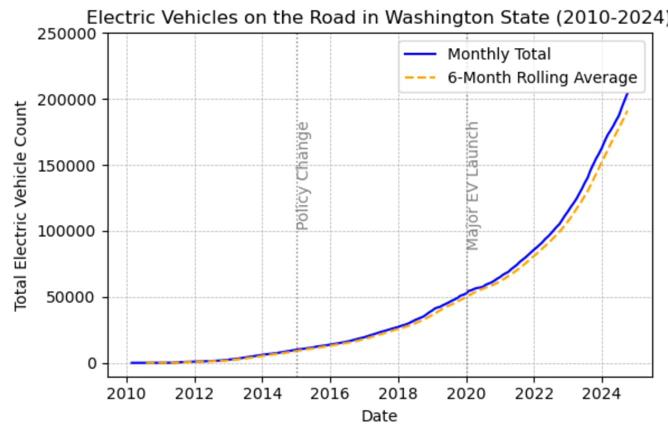

Fig. 4. Exponential Growth of EVs in Washington State

The graph below illustrates exponential expansion in EV adoption from 2010 to 2024, with King County significantly ahead. Suburban counties such as Snohomish and Pierce demonstrate consistent growth, while smaller counties trail behind. This emphasizes regional differences and the necessity for focused infrastructure investments.

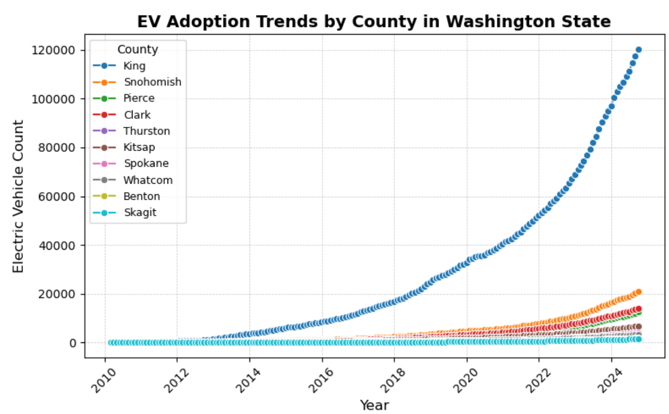

Fig. 5. County-Wise Adoption Trends

The chart in Fig. 6 demonstrates exponential growth in electric vehicles in counties like Snohomish and Pierce between 2010 and 2024. Other counties exhibit a slower pace of adoption, highlighting the necessity for focused expansion of infrastructure in suburban and rural areas.

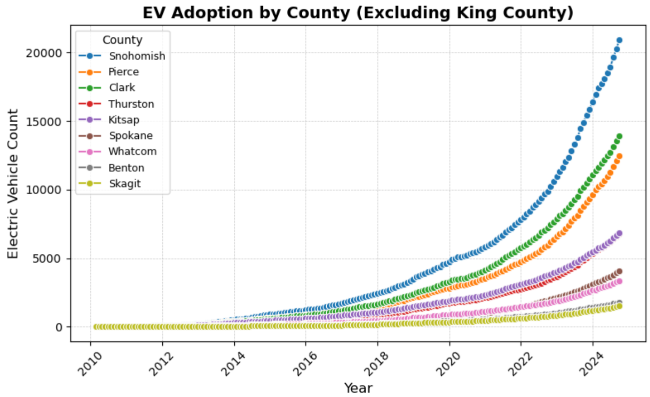

Fig. 6. EV Trends Excluding King County

The map showcases concentrated EV charger groupings in cities like King County, influenced by elevated adoption rates. Suburban locations, including Olympia and Tacoma, demonstrate moderate availability of chargers, whereas rural and eastern counties encounter limited coverage. This underscores the necessity for focused infrastructure development in underserved regions to alleviate range anxiety and foster statewide EV expansion. Funding in high-demand corridors can guarantee accessibility and sustainability.

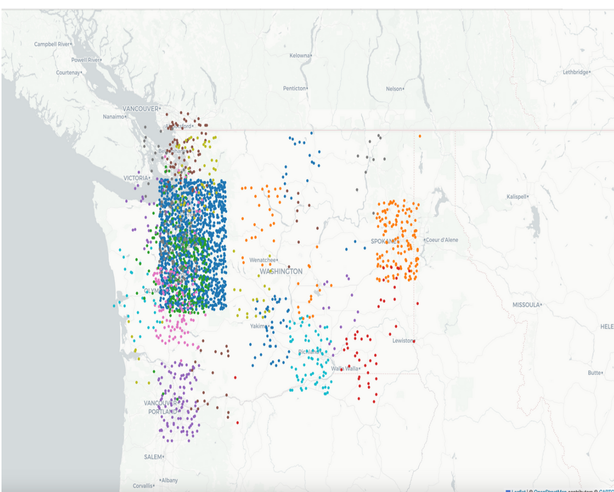

Fig. 7. Analysis of Charger Distribution

King County leads, as illustrated in Fig. 8, possessing 1,377 chargers, which represent almost half of the state's overall total. Following behind are Pierce and Snohomish counties with considerably fewer chargers, and the notable decrease in charger access in other counties emphasizes the necessity for equitable infrastructure distribution to facilitate broad EV adoption.

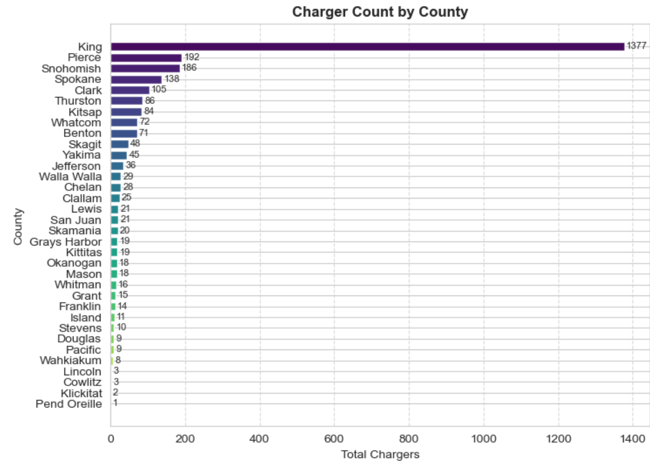

Fig. 8. Total Charger Count by County

C. Data Validation Approaches

In the data preparation phase, numerous validation steps were carried out to guarantee the integrity and usability of the dataset. These steps involved validating date and time fields, resolving missing data, managing duplicate entries, and ensuring the consistency of categorical and geographical data.

1. Validation of Missing Data

To assess the dataset's quality, missing values were detected through a null_counts summary. This process revealed which columns contained null values, assisting in determining how to address them.

```
[19]: # Calculate the count of null values for each column
      null_counts = df.isnull().sum()

      # Display the count of null values
      print(null_counts)
```

Fig. 9. Null Value Count Per Column

- Action on Columns with High Missing Values: Columns like date_of_vehicle_sale contained a large percentage of missing values and were eliminated since they were not critical to the analysis.
- Action on Columns with Minimal Missing Values: Columns such as city and zip had a limited count of null values, which were substituted with "Unknown “to maintain data integrity.

2. Validation of Categorical Data

Categorical variables were examined to confirm they included solely pertinent categories. For instance, the transaction type column was checked to ensure it contained only "Original Title" and "Transfer Title," which denote significant transactions for the analysis.

```
df = df[df['transaction_type'].isin(['Original Title', 'Transfer Title'])]
```

Fig. 10. Filtering Relevant Transaction Types

This step filtered out irrelevant transaction types such as renewals, focusing now on the purchases and transfers.

3. Validation of Duplicate Entries

Duplicates were identified and removed to ensure data reliability. The duplication check was performed on subsets of critical columns, including month_year, dol_vehicle_id, and county.

```
# Identifying duplicates by counting each unique combination
duplicates = df[df.duplicated(subset=['month_year', 'dol_vehicle_id', 'county'], keep=False)]
print(f"Number of duplicate entries: {duplicates.shape[0]}")
```

Fig. 11. Duplicate Entry Validation and Removal

We ensured analysis be non-skewed by removing redundant entries and duplicate records.

Validation of Geographical Data

We checked geographical data, such as city and zip, for missing or invalid values. Missing entries were replaced with placeholders like "Unknown" to maintain the data's usability.

```
df['city'].fillna('Unknown', inplace=True)
df['zip'].fillna('Unknown', inplace=True)
```

Fig. 12. Handling Missing Geographical Data

We made the data consistent for location- based insights and ensured geographical consistency.

Data Cleaning

During the data cleaning process, we simplified the dataset by removing unnecessary columns to make it more suitable for analysis. For example, fields like electric_vehicle_fee_paid, hb_2042_clean_alternative_fuel_vehicle_cafv_eligibility, and census_tract_2020 were excluded because they were not relevant to the project's goals. By eliminating these fields, as shown in Figure 13, we reduced complexity and ensured that the analysis focused on the most important attributes needed for effective modeling.

```
cols_dropping = ['electric_vehicle_fee_paid',
            'hb_2042_clean_alternative_fuel_vehicle_cafv_eligibility',
            'meets_2019_hb_2042_electric_range_requirement',
                '_2019_hb_2042_battery_range_requirement',
                '_2019_hb_2042_sale_price_value_requirement',
                '_2019_hb_2042_purchase_date_requirement',
                'census_tract_2020',
            'meets_2019_hb_2042_sale_date_requirement',
                'electric_utility',
            'meets_2019_hb_2042_sale_price_value_requirement',
            'transportation_electrification_fee_paid',
            'hybrid_vehicle_electrification_fee_paid', 'legislative_district']
df.drop(cols_dropping, axis=1, inplace=True)
df.head()
```

Fig. 13. Dropping irrelevant columns

Several key columns in the dataset had missing values, including *date_of_vehicle_sale* (842,439 missing entries), *county* (61 missing entries), and *city* (108 missing entries). The *date_of_vehicle_sale* column was removed because it wasn't relevant to the analysis. For less critical missing data, placeholders like "Unknown" were used in fields such as *county* and *city.* Meanwhile, numerical columns like *electric_range* and *base_msrp,* which had 11 missing entries each, were filled in with reasonable estimates to maintain data integrity.

```
[29]: # Calculate the count of null values for each column
null_counts = df.isnull().sum()

# Display the count of null values
print(null_counts)
```

```
electric_vehicle_type        0
vin_1_10                     0
dol_vehicle_id               0
model_year                   0
make                         0
model                        0
vehicle_primary_use          0
electric_range              11
odometer_reading             0
odometer_code                0
new_or_used_vehicle          0
sale_price                   0
date_of_vehicle_sale    842439
base_msrp                   11
transaction_type             0
transaction_year             0
county                      61
city                       108
```

Fig. 14. Missing Value Counts

During the data cleaning process, we tackled missing values and eliminated columns that were either redundant or irrelevant to our analysis. For instance, the date_of_vehicle_sale column had a substantial number of missing entries (842,439), making it unreliable for use. Since the transaction_date column already captured detailed temporal information, we opted to remove date_of_vehicle_sale Similarly, we removed the transaction_year column because it duplicated information already available in transaction_date. Additionally, we decided to drop the base_msrp column, as the sale_price offered more accurate and relevant pricing data for our analysis. To focus the dataset on meaningful transactions, we filtered the data to include only entries that reflected changes in EV ownership, such as "Original Title" and "Transfer Title." We excluded transactions like renewals, as they didn't align with our objective of tracking vehicle movements and new registrations.

```
df['date_of_vehicle_sale'].isna().sum()

842439

redundant_columns = ['transaction_year', 'base_msrp', 'date_of_vehicle_sale']
df.drop(columns=redundant_columns, inplace=True)

redundant_columns = ['transaction_year', 'base_msrp', 'date_of_vehicle_sale']
df.drop(columns=redundant_columns, inplace=True)

df = df[df['transaction_type'].isin(['Original Title', 'Transfer Title'])]
```

Fig. 15. Dataset Feature Reduction

We identified and resolved duplicate entries by using a combination of fields, including month, year, **dol_vehicle_id**, and county, as shown in Figure 16. In total, 196 duplicate rows were detected, differing only in minor details like city or zip code. To maintain accuracy, we retained the most recent transaction for each duplicate, assuming it to be the most reliable version of the data. Certain columns, including **county**, **city**, and **zip**, had missing values, with 61, 23, and 15 entries missing, respectively. To address this, we filled these gaps with placeholders like "Unknown," ensuring the dataset remained complete without impacting the quality of the analysis.

```
# Identifying duplicates by counting each unique combination
duplicates = df[df.duplicated(subset=['month_year', 'dol_vehicle_id', 'county'], keep=False)]
print(f"Number of duplicate entries: {duplicates.shape[0]}")
```

```
Number of duplicate entries: 196
```

```
# Dropping duplicates while keeping the latest entry within each group
df.drop_duplicates(subset=['month_year', 'dol_vehicle_id', 'county'], keep='last', inplace=True)
```

```
df.drop_duplicate`s(subset=['month_year', 'dol_vehicle_id'], keep='last', inplace=True)
```

```
odometer_reading        0
odometer_code           0
new_or_used_vehicle     0
sale_price              0
transaction_type        0
county                  8
city                   23
state_of_residence      1
zip                    15
month_year              0
```

Fig. 16. Removed duplicates

To improve consistency and improve readability, we standardized the model column by capitalizing the first letter of each word. This step resolved inconsistencies in text formatting, creating a uniform and professional presentation of the model names of model names.

Data Transformation

To enhance the dataset's structure and usability for temporal and analytical operations, we implemented several key transformations. These included setting the **date_of_purchase** column as the DataFrame's index to enable efficient time-series analysis, creating a **month_year** column to simplify monthly trend tracking, and renaming and reformatting columns to improve clarity and consistency.

1. Setting the Date as Index

The **date_of_purchase** column was set as the DataFrame's index to streamline time-series operations. This transformation optimized the dataset for temporal analysis, enabling smooth integration with time-based computation techniques.

```
df.set_index('date_of_purchase', inplace=True)
```

Fig. 17. Indexing Dataset with Purchase Dates

By indexing the dataset with dates, it became better structured for operations such as trend analysis and forecasting.

2. Creating a month_year Column

We created a new month_year column by extracting data from the date_of_purchase column using the .strftime('%m-%Y') function. This transformation made it easier to group and analyze data by monthly trends, provided by a clearer view of time based patterns in the dataset.

```
df['month_year'] = df['date_of_purchase'].dt.strftime("%m-%Y")
```

Fig. 18. Month-Year Extraction

This feature simplified the process of analyzing temporal trends and comparing data across different months.

3. Rename and Reformat Columns

To improve clarity and consistency, the transaction_date column was renamed to date_of_purchase. This change better reflects the context of the data and makes it more intuitive for analysis. Additionally, the date_of_purchase column was reformatted into a consistent datetime format using pd.to_datetime(). This standardization ensures accurate temporal analysis by providing a uniform structure, making it easier to perform operations like filtering, grouping, and identifying trends.

```
df.rename(columns={'transaction_date': 'date_of_purchase'}, inplace=True)
df['date_of_purchase'] = pd.to_datetime(df['date_of_purchase'])
```

Fig. 19. Renaming and Standardizing Date Columns

This transformation enhances the dataset's usability for time-based analyses and ensures it integrates seamlessly with various analytical techniques.

IV. MODEL DEVELOPMENT

Model Proposals

This project utilizes advanced machine learning models, including Auto ARIMA and SARIMAX, to analyze historical data and predict future trends, providing reliable, data-driven forecasts to support sustainable EV infrastructure development. Figure 13 outlines the main flow of the proposed model, with detailed descriptions provided in the following sections.

Auto ARIMA is a sophisticated time-series forecasting model designed to automate the process of selecting optimal parameters, simplifying the creation of accurate predictive models. It evaluates various combinations of autoregressive (p), differencing (d), and moving average (q) parameters, selecting the best configuration based on metrics such as the Akaike Information Criterion (AIC). This automation not only streamlines the modeling process but also improves forecasting accuracy by reducing the need for manual intervention. Auto ARIMA is particularly effective for datasets with strong short-term dependencies and moderate seasonality, making it an ideal tool for quickly generating reliable forecasts of dynamic trends like EV adoption.

SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous variables) builds on the ARIMA framework by incorporating seasonal patterns and external explanatory variables. This allows it to effectively model datasets that exhibit periodic trends and are influenced by additional factors. SARIMAX is especially suited for complex, long-term datasets where seasonal cycles, trends, and external variables interact. By capturing both short-term dependencies and seasonal variations, SARIMAX provides a robust approach to forecasting. It has been widely applied in areas that require a deep understanding of temporal dynamics, such as transportation planning and urban sustainability projects. For EV infrastructure forecasting, SARIMAX offers actionable insights that can guide policy decisions and optimize resource allocation, thanks to its ability to handle diverse seasonal patterns and external influences.

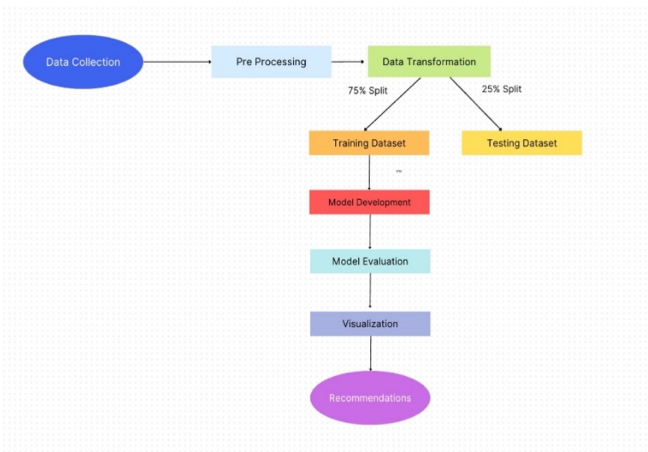

Fig. 20. An Overview of the Proposed Framework.

A. Model Deployment

We created a new **month_year** column by extracting data from the **date_of_purchase** column using the .strftime('%m-%Y') function. This transformation made it easier to group and analyze data by monthly trends, providing a clearer view of time-based patterns in the dataset..

B. Model Integration

Model Implementation: SARIMAX and Auto ARIMA models were developed and executed in a Python environment using libraries like **statsmodels** and **pmdarima**. These models were trained and tested on historical data to produce accurate predictions.

Data Preparation: Cleaned and preprocessed datasets served as inputs for the models, ensuring reliability. The resulting outputs, such as forecasts and confidence intervals, were exported for further visualization and analysis.Visualization and Insights

Interactive Dashboards: Model results, including predicted EV adoption trends and error metrics, were visualized through interactive plots and graphs created in Python. These visualizations offered stakeholders clear, actionable insights.

County-Level Analysis: Forecasts were detailed at the county level, showcasing localized trends and enabling targeted, practical recommendations for EV infrastructure planning.Accessibility.

Streamlined Deployment: Model results were shared through static visualizations and reports, making them easily accessible to stakeholders. No integration with external systems or cloud services was required. By focusing on straightforward visualizations, the project effectively provided insights into EV adoption trends and infrastructure needs without relying on complex deployment processes.

C. Model Support

The project was developed on multiple platforms, including Windows and MacOS, using **Jupyter Notebook** and **Google Colab** as the primary programming environments. Python (version 3.9) was chosen as the core programming language for its extensive library ecosystem, which supports data analysis, visualization, and machine learning. **Pandas** was used for data manipulation and preprocessing, allowing efficient handling of missing values, duplicate records, and transformations necessary for time-series analysis. Visualizations were created using **Matplotlib** and **Seaborn**, enabling a clear analysis of historical trends and forecast results. For time-series forecasting, the project utilized **pmdarima** for Auto ARIMA and **statsmodels** for SARIMAX, both tailored to capture seasonal patterns and external factors in the dataset.

While cross-validation was not explicitly applied, model evaluation relied on **AIC** and **BIC** metrics to achieve a balance between model complexity and predictive accuracy. Google Colab provided cloud-based computational support for resource-intensive tasks, facilitating an iterative development process. This comprehensive technical setup supported effective modeling and evaluation, delivering actionable insights into EV adoption trends and optimizing plans for charging infrastructure development.

D. Model Evaluation Methods

The **Model Evaluation** phase was crucial to ensuring the models aligned with the project's goals, such as accurately predicting EV adoption trends and identifying infrastructure gaps. To guarantee their reliability and robustness, the models were thoroughly evaluated and validated through a structured process involving the following steps:

1. Seasonal Decomposition

The first step in analyzing the dataset was **seasonal decomposition**, which broke the data into three key components:

- **Trend:** Reflecting the overall growth trajectory of EV adoption over time.
- **Seasonality:** Highlighting recurring patterns, such as yearly cycles in EV registrations.
- **Residuals:** Representing random variations not explained by the trend or seasonality.

This decomposition was instrumental in identifying both cyclical patterns and long-term trends, which were critical for selecting suitable forecasting models. For instance, Snohomish County displayed a strong annual seasonal pattern alongside a clear upward trend.

2. Data Splitting:

For more robust evaluation and to avoid overfitting, the dataset was split into 75% training data and 25% test data. In this way, the models will get a large part of the data to train on but also leave some data unseen for validation. Training data will range from 2010 to 2021, while testing data from 2021 to 2024. In this way, it would surely allow the models to generalize well to future trends..

3. Auto ARIMA Parameter Selection

p, and the best parameters were chosen via auto ARIMA, which automatically chose the best values of p, d, and q by trying plenty of combinations and minimizing metrics, such as Akaike Information Criterion. In such a way, Auto ARIMA prevented sophisticated parameter tuning and allowed considerable model performance. Such an approach was especially good in applications that show moderate seasonality and great dependency on the short-term influences, which is just the case of Snohomish County.

4. SARIMAX Model Configuration

The SARIMAX model was manually configured to account for seasonal trends and external factors. Key parameters, including (p, d, q) for short-term dependencies and (P, D, Q, m) for seasonal effects, were carefully fine-tuned to match the dataset's unique characteristics. For instance, a seasonal order of (0, 0, 2, 12) was applied to capture annual seasonal patterns. This configuration enabled SARIMAX to effectively model both short-term variations and recurring trends, providing valuable insights into EV adoption dynamics.

5. Forecasting on Test Data

The models were validated using a testing dataset by comparing the forecasted EV adoption numbers against actual values. Confidence intervals were used to evaluate the uncertainty of the predictions, while visual comparisons demonstrated a strong alignment with observed trends. This was especially evident in Snohomish County, further confirming the reliability of the models.

6. Performance Metrics

The models were evaluated using several metrics to assess their accuracy and reliability:

- **Akaike Information Criterion (AIC):** This metric balanced model complexity with goodness-of-fit, where lower values indicated better performance.
- **Bayesian Information Criterion (BIC):** Similar to AIC, but it applied a stricter penalty for model complexity, promoting simpler models.
- **Log-Likelihood:** This measured how well the model explained the observed data, with higher values indicating a better fit.
- **Root Mean Squared Error (RMSE):** This metric quantified the size of prediction errors, with lower values reflecting greater model reliability.
- **Mean Absolute Percentage Error (MAPE):** Provided a relative measure of accuracy by calculating the average percentage error, making it useful for comparing prediction performance across datasets.

7. Residual Diagnostics

Residual diagnostics were conducted to ensure that errors were random and normally distributed, which is essential for model validity. Residual Plots: Analyzed for randomness around zero, indicating no systematic biases. Quantile-Quantile (Q-Q) Plot: Examined residuals for deviations from normality. Autocorrelation Function (ACF): Checked for significant lags to confirm the absence of autocorrelation.

8. Forecasting on Complete Dataset

Following validation, the models were retrained using the complete dataset spanning 2010 to 2024. This approach ensured that all available historical data was incorporated, enhancing the accuracy of long-term forecasts. The retrained models generated comprehensive projections for EV adoption, providing valuable insights to guide strategic planning for charging infrastructure. In Snohomish County, the retrained SARIMAX models offered detailed predictions with confidence intervals, effectively capturing potential uncertainties in future trends.

9. Visualization

Visualizations were integral to the evaluation process:

• Trend Analysis: Forecasts and actual values were plotted for each county to visually inspect their alignment and identify deviations.

• Diagnostic Plots: Residual and Q-Q plots offered insights into model performance and areas for refinement.

• Confidence Intervals: Visual displaying along with predictions, to show how uncertain or reliable the forecast is.

IV. EVALUATION RESULTS

Forecasts of time-series models of EV adoption trends were implemented to predict and plan required improvements to the charging infrastructure within County. With the historical trends given, county-specific trends and seasonality, gaps in infrastructure are to be estimated using the Auto ARIMA and SARIMAX. To this day, we have systematically split the given data into training and testing data subsets to ensure that model fits are robust. Various performance metrics were considered for the study: AIC, BIC, RMSE, MAPE, and visual diagnostics by residual analysis. Results have also been analyzed post-forecasting to see how predicted values align with real-world requirements of infrastructure, hence enabling actionable insights into under-served regions and future planning of infrastructure. In addition to an overview, each individual county is given full focus as results are drawn together through emphasis on unique learnings brought forward through the models related to predicted accuracies, seasonal trends, and strategic recommendations concerning developments related to EV infrastructure. Additional sections outline the specific ramifications arising regarding actionable ways targeted resources would then foster outcomes focused on sustainable goals surrounding e-transportation adoption.

1. Snohomish County:

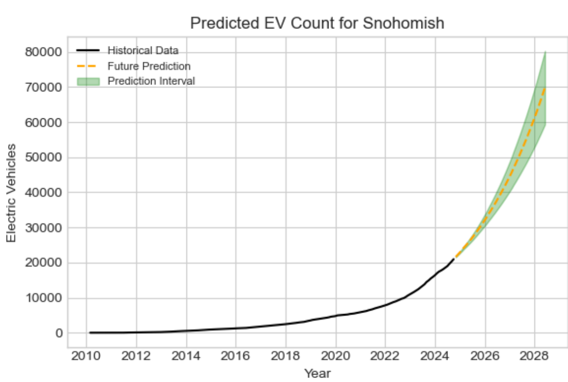

Fig. 21. Snohomish County EV Adoption Forecast

The SARIMAX model specified by the non-seasonal order of 1,1,2 and seasonal order of 0,0,2,12 was a perfect fit for both short-term dependencies and yearly seasonality for Snohomish County, evinced from the univariate analysis through seasonal decomposition of the time series data. Then, the data were split into 75% training and 25% testing to ensure reliability in the evaluation; the training data for fitting the model, while holding out the test data for validation. The SARIMAX model fits fairly well, but the residual analysis showed very little autocorrelation, slight non-normality, heteroskedasticity, and high variance; hence, there's room for improvement. The forecasting on the test data using it gives good growth in front of the EVs coming up in November 2026 to about 43325. It's an excellent upward trajectory. Diagnostic Metrics Low values of AIC and BIC for Auto ARIMA have shown its better performance in automatic parameter selection to deliver smooth and efficient forecasts. Likewise, diagnostic plots such as residual plots and trend patterns proved that the model was able to catch the seasonality combined with the long-term slope properly.

This "EVs per Charger" ratio is very high at 233 and makes for a strong demand to develop the charging infrastructure further. Business districts, transit routes along Highway 2, and Paine Field Airport are also listed as key locations for priority infrastructure siting. Insights like these place Snohomish County in a strategic position for investment in EV charging, ensuring equitable access to support sustainable adoption goals.

2. King County:

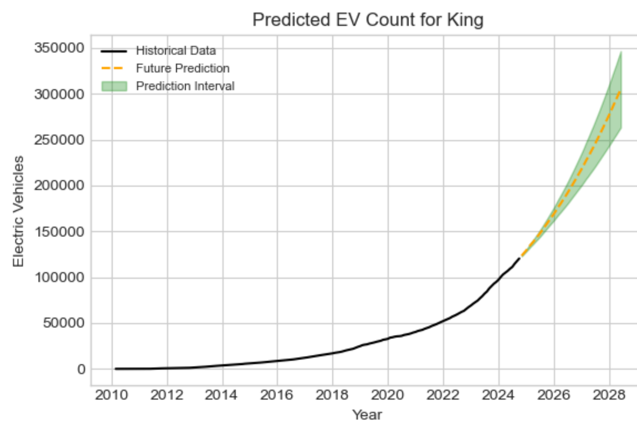

Fig. 22. King County EV Adoption Forecast

We tune King County's SARIMAX by setting the non-seasonal order to (1, 1, 1) and the seasonal order to (1, 0, 1, 12). This setting has captured the short-term dependency and yearly seasonality. The model was estimated using historical data from February 2010 to January 2021, with a total of 132 observations, and out-of-sample testing was conducted from February 2021 to September 2024. Model diagnostics gave a Log Likelihood of -748.495; AIC, BIC, and HQIC were 1506.989, 1520.800, and 1512.596, respectively. These metrics are indicative of a strong balance between model complexity and goodness of fit. Residual diagnostics showed very low autocorrelation, with the Ljung-Box (Q) p-value equal to 0.57, but showed areas for improvement: slight non-normality, heteroskedasticity, and a very high variance, $\sigma^2 = 19{,}940$.

Forecasts for the future, using the entire dataset, 2010–2024, are for a significant increase in the adoption of EVs, hence an upward trend. The SARIMAX model is reliable in its predictions by showing residuals which were random with minimal autocorrelation. The diagnostic plots further established that the model was efficient both in the seasonal pattern and long-run. However, normality has some deviations, as depicted in the Q-Q plot, meaning the model will face difficulties when it deals with extreme values. The forecasted confidence intervals provide actionable insights while acknowledging uncertainty in the long-term prediction.

By comparing the performances of various models, SARIMAX outperformed Auto ARIMA for King County. While Auto ARIMA had efficiently automated the parameter selection process, SARIMAX gave a better AIC and BIC, capturing seasonality. Besides, the high "EVs per Charger" ratio underlines that charging infrastructure in focal areas like urban hubs, transit corridors, and high-density residential zones urgently needs to be extended. These results identify King County as a strategic priority of targeted EV infrastructure investments into the future to meet demand and support equity.

This comprehensive evaluation illustrates King County's unique EV adoption trajectory, providing valuable insights for future planning and implementation.

3. Pierce County:

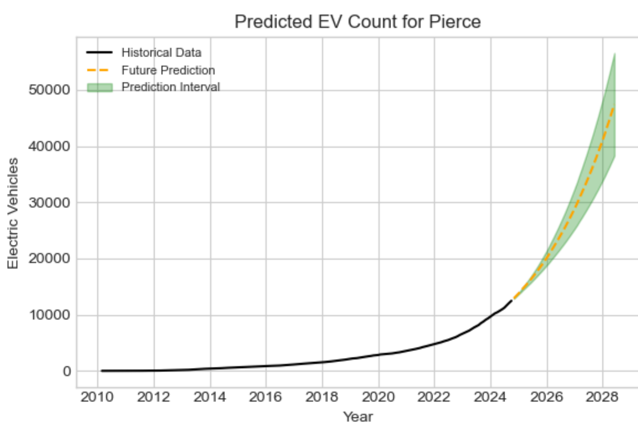

Fig. 23. Pierce County EV Adoption Forecast

Accordingly, the configuration of the order of the SARIMAX model for Pierce County became (3, 1, 0) and of a seasonal order (0, 0, [1], 12); it thus captured the dependencies at short-term intervals and for annual seasonality. Its best Log Likelihood was -459.502. The AIC, BIC, and HQIC equated to 929.003, 942.857, and 934.628, respectively; thus, it assured of its best fit with an optimum level of complexity. The key coefficients, which have contributed much to model performance, are AR terms, L1: 0.4517, and L3: 0.3928, and the seasonal MA term: S.L12: 0.3681, confirmed by their p-values (<0.01). Residual diagnostic analysis revealed very little evidence of autocorrelation since the Ljung-Box test p-value was 0.94, but there was non-normality, Jarque-Bera p-value = 0.00, as well as heteroskedasticity, H-statistic p-value = 0.00. The residual variance, σ², was 139.2904 with a moderate variability in the error terms. Accordingly, the configuration of the order of the SARIMAX model for Pierce County became (3, 1, 0) and of a seasonal order (0, 0, [1], 12); it thus captured the dependencies at short-term intervals and for annual seasonality. Its best Log Likelihood was -459.502. The AIC, BIC, and HQIC equated to 929.003, 942.857, and 934.628, respectively; thus, it assured of its best fit with an optimum level of complexity. The key coefficients, which have contributed much to model performance, are AR terms, L1: 0.4517, and L3: 0.3928, and the seasonal MA term: S.L12: 0.3681, confirmed by their p-values (<0.01). Residual diagnostic analysis revealed very little evidence of autocorrelation since the Ljung-Box test p-value was 0.94, but there was non-normality, Jarque-Bera p-value = 0.00, as well as heteroskedasticity, H-statistic p-value = 0.00. The residual variance, σ², was 139.2904 with a moderate variability in the error terms. A higher ratio of "EVs per Charger" in Pierce County develops greater necessity for the building-up process of charging infrastructures when demand for EVs increases. Key areas are defined as business districts, transit corridors, and highly densified residential zones as crucial zones where infrastructure needs to be developed. Its best performance is now certain from comprehensive evaluation and diagnosis regarding the utility of the method of SARIMAX strategic planning of an electric car charging network in the zone of Pierce County.

4. Clark County:

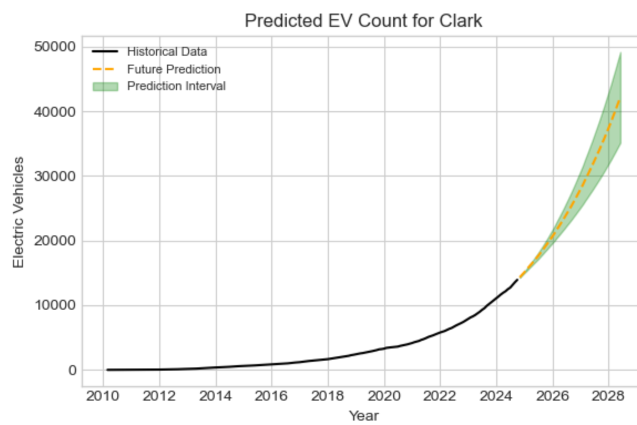

Fig. 24. Clark County EV Adoption Forecast

In this SARIMAX model, (4, 1, 0) represents non-seasonal order and has captured the short-term dependences, while the configuration of (2, 0, 0, 12) reflects yearly seasonality in the seasonal order. To be specific, it was based on 132 observations from Feb 2010 to Jan 2021 with values of Log Likelihood=-420.794, AIC = 855.589, BIC = 874.032, HQIC=863.059; model complexity and quality of fit were balanced. Residual diagnostics show very low autocorrelation, with the Ljung-Box Q p-value being 0.79, which indicates that this model specification is appropriate to catch the temporal dependencies. However, there is non-normality (Jarque-Bera p-value = 0.00) and heteroskedasticity (H = 14.88), indicating variability in error terms. Residuals are slightly positively skewed, 0.59, and leptokurtic, 6.14, which may indicate further directions for improvement in the handling of extreme values. Despite these facts, the model captured well seasonal patterns, as shown by significant on a statistical basis seasonality AR terms: ar.S.L12 = 0.6762, $p < 0.01$ and seasonality AR: ar.S.L24 = 0.4325, $p < 0.01$. Forecast evaluation shows that the SARIMAX model is reliable, as the diagnostic plots show that residuals are random and have low autocorrelation. Model performance comparisons It has been identified that SARIMAX outperforms Auto ARIMA for clark County, as indicated by the lower AIC and BIC. Predictive analytics show a strong upward trajectory in EV adoption, focusing on strategic investments in charging infrastructure along transit corridors and urban hubs. The insight from SARIMAX shall be invaluable to guide EV infrastructure planning within Clark County for the foreseeable future.

5. Thurston County:

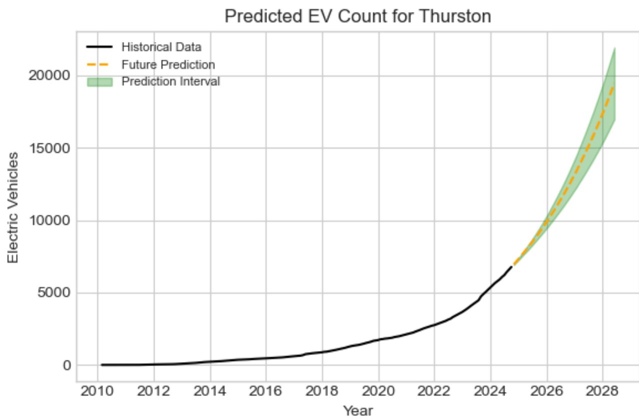

Fig. 25. Thurston County EV Adoption Forecast

In this regard, the SARIMAX model for Thurston County has been configured with a non-seasonal order of 1,1,1 and a seasonal order of 1,0,,12 to efficiently capture both the short-term dependencies and the annual seasonality. Thus, the model trained on observations from February 2010 to January 2021 provides insight into the EV adoption trends. The key diagnostics show that Log Likelihood = -423.245, while AIC, BIC, and HQIC are correspondingly 854.490, 865.573, and 858.990, indicating that this model represents a nice balance between fit and parsimony.

The residual diagnostics indicate a very low level of autocorrelation, as the Ljung-Box (Q) p-value is 0.60, indicating the reliability of the forecast from this model. On the other hand, the Jarque-Bera test results in a p-value of 0.00, which implies non-normality of the residuals. The heteroskedasticity is represented with the value of 8.14 (p-value = 0.00), suggesting some variability in error terms over time. These limitations notwithstanding, the Q-Q plots assure that the residuals from this model are for the most part well-behaved, with deviations in normality having a limited impact on predictive accuracy.

The model effectively captures the seasonality in the trend, and this is reflected in a statistically significant autoregressive term: ar.L1 = 1.0031, $p < 0.01$, and a moving average term: ma.L1 = -0.7013, $p < 0.01$. The model diagnostics have also been validated with a randomness of residuals and low autocorrelation. This lends credence to the strength of the model in predicting the future EV adoption trends. The results of the SARIMAX model for future predictions give a consistent upward trajectory of EV adoption in Thurston. County, aligning with the statewide emphasis on constructing EV infrastructure. The comparison of the results from SARIMAX and Auto ARIMA shows that SARIMAX outperforms the other, as it has lower AIC and BIC with better incorporation of seasonal components. These results position the SARIMAX as a key tool in the planning and implementation of focused EV infrastructure investments in Thurston County, helping policymakers meet the growing adoption needs with efficiency.

6. Kitsap County:

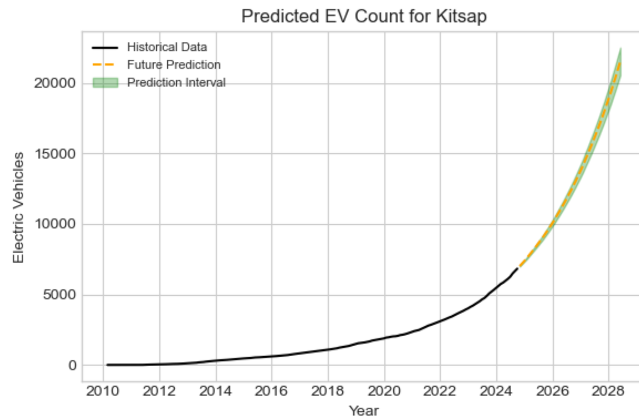

Fig. 26. Kitsap County EV Adoption Forecast

The non-seasonal order (1,1,2) and the seasonal order of (1, 0, [], 12) were set for this SARIMAX model of Kitsap County, capturing short-term dependencies and seasonality fairly well. The model was estimated with a sample of 132 observations, ranging from February 2010 to January 2021. The key diagnostics metrics include Log Likelihood, -411.160; AIC, 832.320; BIC, 846.173; and HQIC, 837.945, proving the model's balance between goodness of fit and complexity. Residual diagnostics are as follows, highlighting the excellent quality of the model fitted.

The Ljung-Box (Q) p-value of 0.99 implies that the autocorrelation is very low, temporal dependencies were well captured accordingly. Residuals are non-normal, as evidenced by the Jarque-Bera p-value of 0.00, and since the Heteroskedasticity value is 6.42, hence, there is some variability in error terms over time. Values of skewness at -0.13 and kurtosis of 4.79 further demonstrate the non-normality. From a statistical point of view, significant parameters include AR(1) and seasonal AR(12), which justifies the model for considering both autoregressive and seasonal components effectively. This would also be in line with the wider efforts across the globe for the use of Electric Vehicles.

The SARIMAX model gives a very consistent upward-going trend and confirms Kitsap county upward growth. Diagnostic plots confirm reliability: showing residual randomness with little to no signs of autocorrelation and slight deviation from residual normality shown by Q-Q plots, thus presenting room for improvement. For Kitsap County, SARIMAX outperforms Auto ARIMA since explicit modeling of seasonality yields lower results in AIC, BIC, and HQIC. In that respect, we place our bets on ARIMAX as a strong and fine-tuned forecasting tool for such a seasonal dataset to provide meaningful insights about investments to be made within Kitsap County regarding their EV infrastructure.

7. Spokane County:

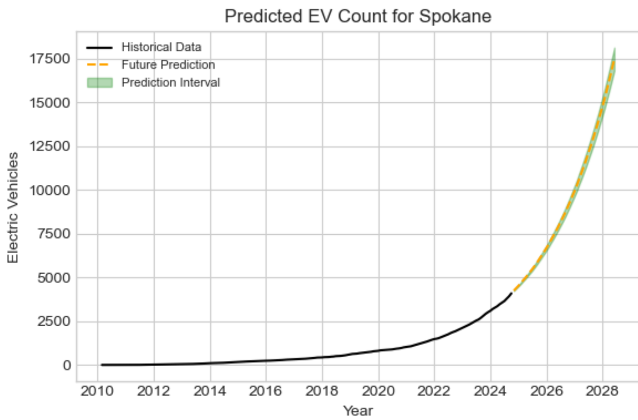

Fig. 27. Spokane County EV Adoption Forecast

Spokane County's SARIMAX model was configured with a non-seasonal order of (1, 1, 2) and a seasonal order of (2, 0, [], 12), effectively capturing short-term dependencies and annual seasonality. The model was trained on 132 observations spanning from February 2010 to January 2021. Key diagnostic metrics included a Log Likelihood of -321.326, with AIC, BIC, and HQIC values of 654.651, 670.632, and 661.128, respectively, indicating a strong balance between model fit and complexity. Residual diagnostics indicated that there was no significant autocorrelation, as evidenced by the 0.83 Ljung-Box (Q) p-value, suggesting thereby that most temporal dependencies have been successfully captured by the model.

The residuals were normally distributed, as reflected by the Jarque-Bera p-value of 0.61, whereas heteroskedasticity, with a heteroskedasticity value of 6.06, still remained within manageable limits, which means error terms may vary some over time but not strongly. Statistically significant parameters, including AR(1), MA(1), and seasonal terms of AR.S.L12 and AR.S.L24, were indicative that the model effectively captured the temporal and seasonal patterns in the data.

The SARIMAX model shows consistent growth upwards in the adoption of EVs in Spokane County and, importantly, underlines the role of planning for expanding infrastructure to meet demand. The diagnostic plots for the models showed that the model is reliable, as the residuals are random; there is little autocorrelation, but there is some heteroskedasticity, which is an issue for long-term forecasting. Against Auto ARIMA, SARIMAX performed better since the model explicitly contained seasonal components, reflected by the much lower values for both AIC and BIC. Hence, the SARIMAX will be a robust model in Spokane County, as this would yield very accurate and actionable insights that may drive infrastructure planning in light of growing EV adoption.

These analyses give meaning to the capability of the model to provide strategic insights on a trade-off basis with good fit and forecast accuracy.

8.Whatcom County:

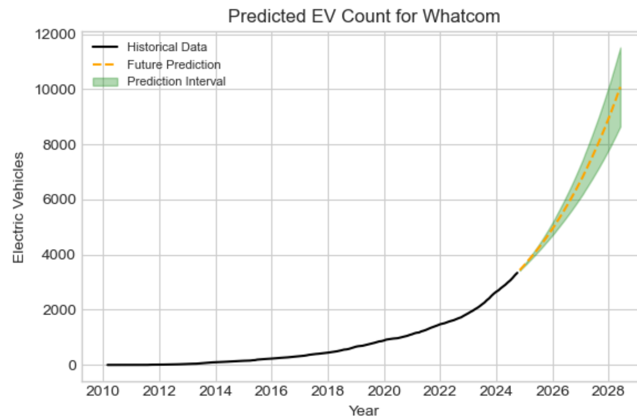

Fig. 28. Whatcom County EV Adoption Forecast

The Whatcom County SARIMAX model was then set with the non-seasonal order: (2, 1, 1) and seasonal order: (1, 0, [], 12) to capture short-term dependencies along with yearly seasonality. Fitting the model on 132 observations ranging from February 2010 to January 2021 yielded a very good fit. It had a Log Likelihood of -341.571, AIC of 693.141, BIC of 706.952, and HQIC of 698.748, which shows that the model has a very good balance between complexity and performance.

The significant parameters were AR(1) and AR(2), which represented the autoregressive pattern; MA(1), that effectively modeled the dependencies on the forecast errors of the past, capturing yearly cycles with its length-12 seasonal autoregression, which has clearly improved the predictive accuracy in.

Residual diagnostics indicated very little auto-correlation: Ljung-Box p-value equaled.77, suggesting temporal dependencies be pretty well captured by this model, though residuals are slightly non-normal (Jarque Bera p-value=.00,) and heteroskedasticity in nature-it follows that error terms varying with time yields a 'Heteroskedasticy' of 12.50; values pointing out the skewness given at -0.23 and kurtosis about 5.14 leads to the very same remark, potentially indicating room for one's further development.

The SARIMAX model projects a continuing, solid upward trajectory in EV adoption for Whatcom County. The diagnostic plots support the validity of the model, where one can see the correspondence of the predicted versus the historical data, and the confidence intervals account for the uncertainty in the forecast.

9.Benton County:

The SARIMAX model for Benton County of the non-seasonal order of (1,1,1) best defined the EV adoption patterns at the county level. Trained on 132 observations (from February 2010 to January 2021), the model had a Log likelihood of -336.515

when fitted.

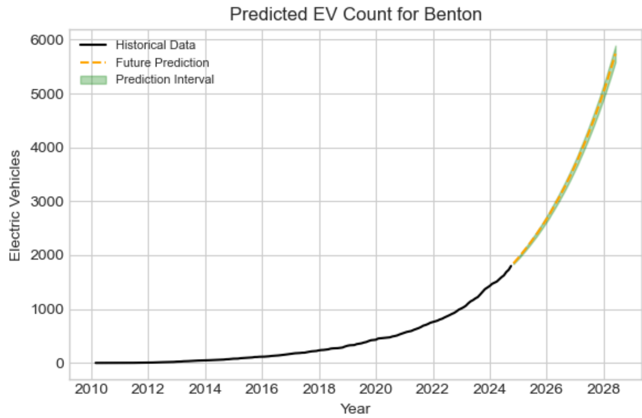

Fig. 29. Benton County EV Adoption Forecast

The key evaluation metrics included AIC: 679.030, BIC: 687.610, and HQIC: 682.516; these showed the best model balance between complexity and fit quality. The significant AR(1) and MA(1) parameters underlined the strength of the model in capturing short-run dependencies, while the no seasonal order underlined the relative low seasonal variation of the County. Residual diagnostics indicated that the model was robust in capturing temporal dependencies. The Ljung-Box test had a p-value of 0.59, confirming that the autocorrelation in residuals was minimal. However, slight non-normality was observed, as highlighted by the Jarque-Bera test with a p-value of 0.00, suggesting refinement in order to address tail behavior. Besides, the residuals showed heteroskedasticity-the H-value equaled to 31.49 with the p-value = 0.00. In that respect, error terms would be unstable over time and thus long-run forecasting may be compromised. Diagnostic plots also confirmed that the SARIMAX model captured the historical pattern satisfactorily and provided reliable forecasts. While the projections show a relatively steady and sharp increase of EVs in Benton County, confidence intervals provide information on the uncertainty of their prediction. Further, evidence of the superiority of the best SARIMAX model over that of Auto ARIMA had been shown with lower results for AIC, BIC, and HQIC parameters. The forecasts from the SARIMAX model have given great emphasis on strategic EV charging infrastructure, thus enabling the policymakers to take critical insights for planning and scaling of the stations to meet up the demand in an effective manner.

10. Skagit County:

The SARIMAX model for Skagit County was configured with a non-seasonal order of (4, 1, 0), effectively modeling the county’s electric vehicle adoption trends by capturing short-term dependencies through autoregressive terms.

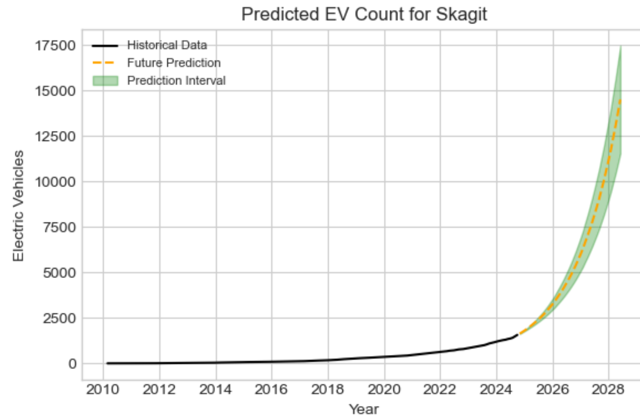

Fig. 30. Skagit County EV Adoption Forecast

This model was estimated based on 132 observations, from February 2010 to January 2021, which provided a Log Likelihood of -278.208 and a reflection of a very good fit of the model. These come via the AIC value of 566.416, BIC of 580.637, and HQIC of 572.194-all of which together provide an indication of a rather balanced model in terms of the level of complexity against the level of performance. The significant AR terms in this model are L1, L2, and L4, with AR.L1 and AR.L2 having p-values of 0.000, proving their statistical significance. A minimal autocorrelation was then found in residual diagnostics through a Ljung-Box p-value of 0.91, which means that the model is enough to capture most of the dependencies in the data.

Residual analysis showed that the errors from this model are slightly non-normal and heteroskedastic. The results from the Jarque-Bera test, JB = 10.54, p-value = 0.01, indicate that residuals are not normally distributed. The heteroskedasticity test resulted in an H-statistic of 3.33 with a p-value of 0.00, showing that residual variance varied over time. Notwithstanding the limitation, the model tended to be strongly predictive; the diagnostic plots indicated historical data aligned very well with forecasted values. This indicates that Skagit County would see a steep rise in electric vehicle adoption and, by inference, an urgent need for focused investment in electric vehicle infrastructure.

This analysis calls for focusing on expanding charging stations in high-demand areas, such as business districts and along transit routes, but also in underserved regions to meet equity demands. SARIMAX outperformed Auto ARIMA in this context, returning better AIC, BIC, and HQIC values; hence, it is reliable for capturing the unique trends of Skagit County. With such strength in the model's robustness, it therefore positions itself as an important tool to guide strategic infrastructure development, with readiness against the growing demand for electric vehicles in the region.

County	EV Count for 2024-09-30	EV Prediction for 2026-11-30	Existing Charger Count	Chargers per EV (2026)	EVs per Charger (2026)	EVs Added (2024-2026)
Clark	13928.0	27249.0	105.0	0.004	260.0	13321.0
Snohomish	20954.0	43325.0	186.0	0.004	233.0	22371.0
Kitsap	6827.0	13405.0	84.0	0.006	160.0	6578.0
Thurston	6770.0	13405.0	86.0	0.006	156.0	6635.0
King	120356.0	213316.0	1377.0	0.006	155.0	92960.0
Pierce	12473.0	28013.0	192.0	0.007	146.0	15540.0
Skagit	1564.0	5685.0	48.0	0.008	118.0	4121.0
Whatcom	3341.0	6505.0	72.0	0.011	90.0	3164.0
Spokane	4095.0	9621.0	138.0	0.014	70.0	5526.0
Benton	1800.0	3573.0	71.0	0.020	50.0	1773.0

Fig. 31. EV Adoption and Charger Infrastructure Projections (2024-2026)

The table offers a comprehensive snapshot of electric vehicle (EV) adoption and charging infrastructure projections across various counties, shedding light on both current trends and future needs.

Current EV Counts (2024-09-30): This column indicates the number of EVs currently on the road in each county. King County leads with 120,550 EVs, showcasing its position as a frontrunner in EV adoption. In contrast, Benton County has the lowest count at 1,800, reflecting a smaller EV market in that region.

Projected EV Counts (2026-11-30): Over the next two years, EV adoption is expected to grow significantly. King County is projected to reach 213,610 EVs, marking a substantial increase. Snohomish County also anticipates notable growth, nearly **doubling** from 20,954 to 43,255 EVs. Meanwhile, Benton County shows a modest increase to 3,630 EVs, highlighting the disparity in adoption rates across counties. These projections underscore the need for tailored infrastructure expansion to meet local demands.

Existing Charger Count: This metric reflects the current number of charging stations. King County has 1,377 chargers, indicating better preparedness for its growing EV population, whereas Benton County, with just 71 chargers, faces a notable infrastructure gap that could become critical as EV adoption accelerates.

EVs per Charger (2026): This ratio measures the adequacy of charging infrastructure against projected EV counts. Spokane County exhibits a balanced ratio of 0.014 EVs per charger, while Benton County's higher ratio of 0.220 EVs per charger signals potential strain on its charging network, emphasizing the need for infrastructure upgrades.

EV Growth (2024–2026): This column highlights the net increase in EVs over two years. King County is forecasted to add 92,060 EVs, representing the highest growth. Snohomish and Pierce Counties will also see significant increases of 22,370 and 28,140 EVs, respectively. Conversely, Benton County's addition of only 1,830 EVs indicates slower growth. These trends suggest that counties with rapid adoption, such as King and Snohomish, will need substantial charging infrastructure expansions, while smaller counties like Benton may benefit from focused investments to address specific gaps.

These insights emphasize the varying dynamics of EV adoption and the critical importance of strategic infrastructure planning to support the growing demand.

The data highlights the critical need for strategic planning in developing charging infrastructure. High-growth counties like King and Snohomish must implement scalable solutions to prevent congestion and ensure accessibility for their expanding EV populations. In contrast, counties with fewer EVs, such as Benton, should concentrate on optimizing their current resources to support gradual growth. The projections also underscore the importance of equitable distribution of charging infrastructure to facilitate a smooth and inclusive transition to electric mobility across all regions.

VI. Visualizations And Key Insights

Figure 31 clearly shows that Clark, Snohomish, and Kitsap counties have the highest EVs-to-charger ratios, indicating substantial gaps in their charging infrastructure. To address these disparities and accommodate the increasing EV adoption, these counties have been selected for visualization. This will help provide a clearer understanding of their current charging network demands and future infrastructure needs.

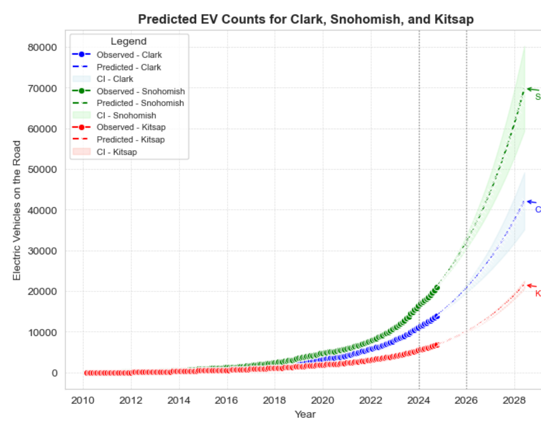

Fig. 32. Predicted EV Growth Trends for Clark, Snohomish, and Kitsap Counties (2010-2026)

This line graph depicts the projected growth in electric vehicle (EV) counts for Snohomish, Clark, and Kitsap counties in Washington State through 2026. It integrates observed data with forecasts, accompanied by shaded confidence intervals for each county, showcasing both the expected trends and the uncertainty of the predictions. By 2026, Snohomish is anticipated to lead with 69,814 EVs, followed by Clark with 42,153, and Kitsap with 21,522 EVs, highlighting varying growth rates across the counties. The graph uses solid lines to represent historical data, while dashed lines extend into the forecasted period. Shaded confidence intervals around the forecasted lines illustrate the range within which future EV counts are expected to fall. The strong alignment between observed and predicted trends confirms the reliability of the models, with Snohomish County showing the steepest growth trajectory.In addition to EV growth, the current charging infrastructure provides critical insights into readiness for future demand. Snohomish County, with 186 chargers, is projected to maintain a balanced EV-to-charger ratio of approximately 0.004 by 2026, indicating sufficient capacity. Similarly, Clark County, with 105 chargers, will achieve a comparable ratio of 0.004, suggesting its infrastructure is adequately prepared for growth. However, Kitsap County, with only 50 chargers, will face a higher EV-to-charger ratio of 0.008, signaling potential strain on its charging network as EV adoption increases.

This visualization underscores the urgent need for strategic planning to expand charging infrastructure in line with rising EV adoption. High-growth areas like Snohomish will need scalable solutions to maintain accessibility, while Kitsap's limited charging capacity calls for targeted investments to ensure equitable access. These insights highlight the importance of proactive planning to meet future EV infrastructure demands effectively.

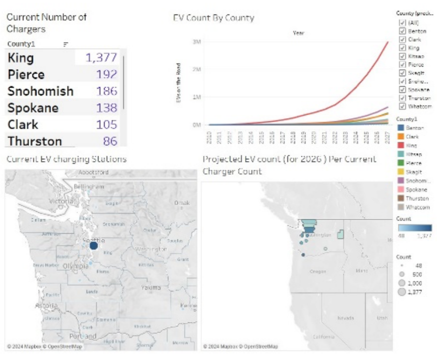

Fig.33. Washington Counties EV Charging Analysis

This visualization provides an overview of the current EV charging infrastructure and future projections across counties in Washington. King County leads with 1,377 chargers, followed by Pierce and Snohomish. The anticipated EV growth by 2026, especially in Snohomish and Clark counties, highlights the increasing demand for charging stations. Maps showcase the geographic distribution of existing chargers and projected EV-to-charger ratios, emphasizing the critical need for strategic infrastructure investments in counties experiencing high EV adoption but limited charging capacity.

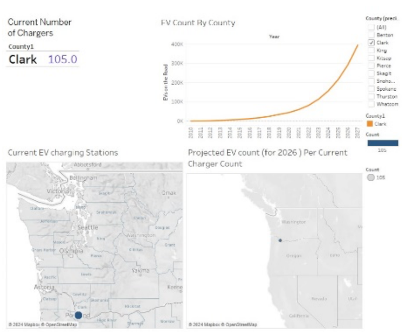

Fig.34. EV Charging Infrastructure Analysis for Clark County

This analysis focuses on Clark County's EV charging infrastructure and future requirements. The top-left panel displays the current availability of 105 charging stations, while the top-right chart projects a substantial increase in EV adoption by 2026. The bottom maps offer a spatial perspective, highlighting the current distribution of chargers and the projected EV-to-charger ratio for 2026. These insights underscore the necessity for strategic infrastructure expansion to accommodate the growing demand effectively.

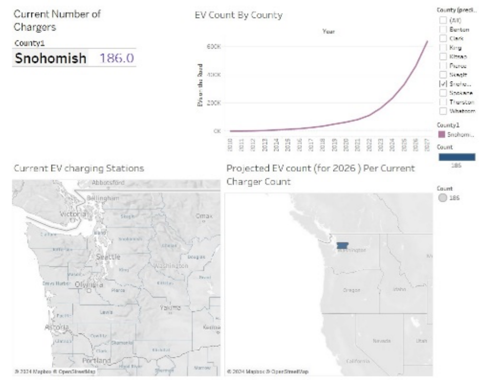

Fig.35. EV Charging Infrastructure Analysis for Snohomish County

Snohomish County currently hosts 186 EV charging stations, as shown in the top-left figure. The top-right chart projects a significant increase in EV adoption by 2026. The bottom maps provide a geographic view of the current charger distribution and the forecasted EV-to-charger ratio for 2026, pinpointing areas where infrastructure expansion will be critical to meet future demand.

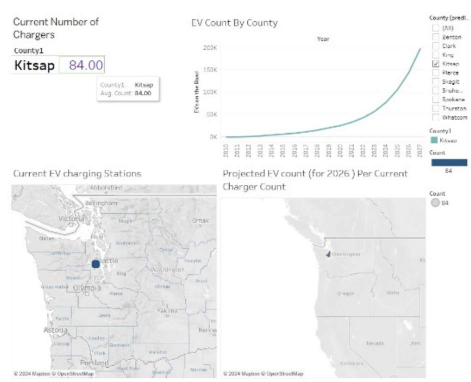

Fig.36. EV Charging Infrastructure Analysis for Kitsap County

Kitsap County currently has 84 charging stations supporting its EV population. Projections for 2026 indicate a significant increase in EV adoption, emphasizing the importance of expanding charging infrastructure. The charts and maps showcase the existing charger locations and the forecasted EV-to-charger ratio, highlighting key areas that should be prioritized for future development

VII. CONCLUSION AND RECOMMENDATIONS

The rapid growth in electric vehicle (EV) adoption across Washington counties underscores an urgent need for strategic expansion of EV charging infrastructure to support this transition. As Washington State positions itself as a leader in transportation electrification, fueled by technological innovations and supportive policies, this analysis highlights the importance of county-specific investments to meet varying demands. The rising popularity of EV models like the Tesla Model Y and Model X demonstrates a consumer shift toward high-performance vehicles, further emphasizing the critical need for reliable and accessible charging networks.

County-Specific Recommendations for EV Charging Infrastructure:

Snohomish County: Snohomish County is projected to lead in EV adoption with an estimated 43,325 EVs by November 2026, but it faces a significant challenge with a high "EVs per Charger" ratio of 233, among the highest in the region. This disparity highlights an urgent need to expand the charging infrastructure, particularly in high-density urban areas, business hubs, and key transit routes like Highway 2. Targeting strategic locations such as Paine Field Airport and commercial zones can maximize the impact of new installations. Additionally, fostering public-private partnerships and promoting workplace charging initiatives can help mitigate the anticipated strain on the charging network.

Pierce County: Pierce County, with its diverse population and wide geographical reach, requires a comprehensive approach to **Thurston County** : Forecasted growth in EV adoption underscores the necessity for expanded infrastructure to meet rising demand in Thurston County. Efforts should focus on urban centers, shopping districts, and densely populated residential areas. Adding chargers along state highways and near government buildings will further enhance accessibility. Collaborating with private businesses to incorporate EV chargers into existing parking facilities can significantly increase capacity and readiness for the growing EV population.

Benton County: Benton County's projections call for a focused expansion of EV charging infrastructure in urban areas, particularly around shopping centers and residential neighborhoods. Developing chargers along agricultural routes and at recreational sites will cater to rural EV users. Incentivizing businesses and property developers to incorporate chargers into new projects can promote adoption while meeting the county's increasing infrastructure needs.

Whatcom County: Whatcom County's proximity to the Canadian border highlights the importance of supporting cross-border travel. Strategic placement of chargers at border crossings and along international routes will facilitate seamless travel. Additional infrastructure in urban centers and commercial zones, especially near Bellingham, will address growing EV adoption. expanding its EV charging infrastructure. Strategic placement of chargers in business districts, near schools, and along major highways will cater to daily commuter needs. Installing chargers in recreational areas and parks can encourage EV adoption among rural residents. Offering incentives for private charger installations at workplaces and residential properties will further support the public infrastructure, ensuring a balanced and accessible charging network across the county.

Kitsap County: With a projected EV count of 21,522 by 2026, Kitsap County's "EVs per Charger" ratio highlights an urgent need for targeted investments. Prioritizing charging stations at transit hubs and ferry terminals, which are integral to the county's transportation network, will be essential. Strategically placing chargers near long-term parking areas, urban centers, and major commuter routes will improve accessibility. Expanding workplace and residential charging options, particularly in multi-family housing complexes, will help distribute demand more evenly across the region

Clark County: Clark County, with a projected EV count of 42,153 by 2026, demonstrates significant growth potential. To meet this demand, the existing charging infrastructure must be expanded to cover high-traffic areas such as transit corridors, shopping centers, and residential neighborhoods. Its location near the Oregon border further emphasizes the importance of strategically placing chargers along cross-state travel routes. Prioritizing suburban areas for new installations and taking advantage of federal and state funding programs can help accelerate infrastructure development and ensure the county is prepared for the rise in EV adoption.

Spokane County: As a growing EV market, Spokane requires targeted infrastructure development to support its increasing adoption rates. Installing chargers along interstate highways, urban centers, and popular recreational destinations will enhance accessibility for a broad range of users. Ensuring rural areas within the county are prioritized will help achieve equitableaccess. Partnering with local businesses to co-locate chargers with existing services can efficiently expand coverage and meet demand.

Skagit County: Skagit County's increasing EV adoption highlights the need for targeted infrastructure expansion. Key areas for development include ferry terminals, urban centers, and popular tourist destinations. Strategically placing chargers along state highways and near major retail hubs will maximize their utility. Additionally, prioritizing underserved rural areas will promote equitable access to charging facilities, aligning with the county's commitment to sustainability and inclusivity.

Overall Recommendations:

To address the growing demand for EV infrastructure across Washington State, leveraging public-private partnerships and incentive programs will be key to accelerating development. Integrating renewable energy into charging stations can align operational needs with the state's sustainability goals. Continuous data-driven monitoring of EV adoption trends and charger utilization will ensure the infrastructure adapts to evolving demands while maintaining equity. By focusing on county-specific priorities and strategic investments, Washington can establish itself as a leader in creating a seamless and sustainable EV ecosystem.

VII. FUTURE SCOPE

The future development of this project aims to enhance the EV Charger Prediction System to better address the dynamic needs of electric vehicle (EV) infrastructure planning. By incorporating additional data sources—such as demographic trends, energy grid capacity, and consumer behavior—the system can generate more detailed and comprehensive forecasts. Advanced machine learning techniques, including deep learning models, can also be integrated to better capture complex non-linear patterns, further improving prediction accuracy.

Real-time data ingestion from IoT devices and connected vehicles will enable the system to dynamically update forecasts, providing stakeholders with timely insights into shifts in EV adoption and charger utilization. Expanding cloud integration across platforms like AWS, Azure, and Google Cloud will enhance the system's scalability and efficiency, allowing it to process larger datasets and conduct more sophisticated analyses.

The visualization capabilities can be refined with interactive dashboards, enabling stakeholders to simulate various scenarios, such as the effects of policy changes or incentives on EV adoption. Geospatial analytics can also be incorporated, providing precise recommendations for charger placement by considering factors such as highway proximity, urban density, and access to renewable energy sources.

Moreover, the system's scope can extend beyond Washington State to a national or global scale, adapting to different geographical and policy environments. By partnering with government agencies and private organizations, this project can contribute to accelerating sustainable EV adoption and infrastructure development worldwide, aligning with global efforts to reduce greenhouse gas emissions and combat climate change.

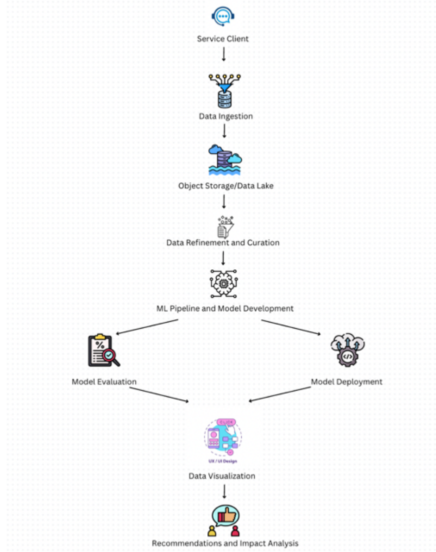

Fig.37. End-to-End Machine Learning Workflow for Data-Driven Decision Making

The flowchart outlines a robust pipeline for analyzing EV adoption trends and forecasting charger demand. The process begins with **Service Clients**, where data is collected from multiple sources, including traffic sensors, weather APIs, EV adoption trends, and charger usage logs.

Data Ingestion utilizes real-time and batch streaming tools such as Kafka, Kinesis, Azure Event Hubs, and Pub/Sub to ensure seamless data flow. The ingested data is then stored in **Object Storage/Data Lakes** using scalable platforms like Amazon S3, Azure Blob Storage, or Google Cloud Storage.

In the **Data Refinement & Curation** phase, preprocessing and real-time transformation are conducted using tools like Snowflake and Spark to ensure the data is clean and ready for analysis. The **ML Pipeline & Model Development** leverages advanced models like Auto-ARIMA and SARIMAX to predict EV adoption and charger demand. These models are evaluated using metrics such as MAPE, RMSE, and R^2 to ensure accuracy. Deployment of the models is handled via platforms like H2O.ai or APIs, facilitating batch and real-time predictions using cloud integration tools like AWS Lambda and Google Cloud Functions.

The **Visualization & Insights** phase provides stakeholders with intuitive dashboards to monitor trends and receive actionable recommendations for infrastructure planning. This pipeline is supported by **Cloud Integration**, ensuring scalability and efficient orchestration of data across the entire system.

REFERENCES

[1] Sakib Shahriar; A. R. Al-Ali; Ahmed H. Osman; Salam Dhou; Mais Nijim (2021). Prediction of EV Charging Behavior Using Machine Learning. *IEEE Access, 9*, 111576–111586

[2] Hyndman, R. J., & Athanasopoulos, G. (2018): Explains the automation capabilities and forecasting accuracy of Auto ARIMA, emphasizing its role in simplifying time-series analysis. *Forecasting:*

Principles and Practice (OTexts).

[3] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020): Highlights the strengths of SARIMAX in addressing seasonal and complex datasets for forecasting purposes. *International Journal of Forecasting, 36*(1), 54-74.

[4] Zhang, J., & He, Y. (2018): Demonstrates SARIMAX's application in transportation planning, focusing on its ability to integrate external variables. *IEEE Transactions on Transportation Electrification, 4*(2), 456-464.

[5] García et al. (2020): Discusses the application of time-series models like SARIMAX in sustainable urban planning for EV infrastructure. *Sustainability, 12*(9), 3604.

APPENDIX

Code: Link to AutoArima and Serimax

Dashboard Link