

# **Electric Vehicle Adoption in Washington State**

## **Abstract:**

This comprehensive study explored the elaborate landscape of electric vehicle (EV) adoption in Washington state, employing three robust models—Random Forest, Logistic Regression, and Naive Bayes. The methodology involved the steps of data cleaning, preparation, exploratory data analysis (EDA), and model predictions. The research aimed to offer meaningful insights for stakeholders.

Notably, the Random Forest model exhibited exceptional accuracy, achieving 99.97%. By addressing critical questions about EV adoption rates, preferences, and economic considerations, this research makes a substantial contribution to the promotion of sustainable transportation and aids informed decision-making for policymakers, manufacturers, and electric utility providers.

Key findings of the study include identifying the most significant predictors of EV adoption, understanding regional variations in EV preferences, and uncovering the impact of economic incentives on consumer choices. The research also sheds light on the role of charging infrastructure availability in driving EV adoption rates.

In essence, this study not only provides a comprehensive analysis of the current state of EV adoption in Washington but also paves the way for future research and policy initiatives aimed at accelerating the transition to sustainable transportation.

## Table of Contents

<b>Abstract:</b> .....	<b>2</b>
<b>1. Background, Introduction, and Project Goal</b> .....	<b>4</b>
1.1 Background .....	4
1.2 Introduction .....	4
1.3 Project Goal.....	5
1.4 Research Questions .....	5
1.5 Methodology used in the following report.....	6
1.4.1 Data Cleaning: .....	6
1.4.2 Data Preparation: .....	6
1.4.3 Exploratory Data Analysis (EDA):.....	6
1.4.4 Model Prediction:.....	6
<b>2. Data</b> .....	<b>7</b>
2.1 Dataset source:.....	7
2.2 Data Cleaning: .....	8
2.3 Feature Engineering .....	9
<b>3 Exploratory Data Analysis (EDA)</b> .....	<b>10</b>
3.1 EDA -Univariate .....	10
3.2 EDA-Bivariate .....	13
<b>4 Research Questions Analysis</b> .....	<b>15</b>
4.1 EV Adoption Over Time .....	15
4.2 EV Type Preference.....	16
4.3 Price vs. Electric Range .....	17
4.4 CAFV Eligibility Impact:.....	17
4.5 Utility Analysis: .....	18
<b>5. Modelling</b> .....	<b>19</b>
5.1 Naïve Bayes:.....	20
5.2 Logistic Regression .....	21
5.3 Random Forest .....	22
5.4 Comparison of three models .....	22
5.6 Predictions based on best model .....	25
<b>6. Discussions</b> .....	<b>25</b>
6.1 From Research Questions .....	25
6.2 From EDA, Modeling.....	26
<b>7. Conclusions</b> .....	<b>26</b>
<b>8. Recommendations</b> .....	<b>27</b>

## 1. Background, Introduction, and Project Goal

### 1.1 Background

Within the expanding electric vehicle market, our project addresses the critical necessity of intricate classification for Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs). This scientific precision is fundamental for stakeholders, delivering nuanced insights crucial for informed decision-making in manufacturing, policy formulation, and consumer engagement. This endeavour propels advancements in electric vehicle research and development, contributing to the scientific evolution of sustainable transportation

### 1.2 Introduction

Washington's electric vehicle revolution is a testament to its pioneering spirit in embracing clean energy and sustainable practices. The state's commitment to reducing carbon emissions is evident in its aggressive policies and incentives designed to promote the adoption of EVs. These policies include tax rebates, grants for charging infrastructure, and preferential parking for EVs, making them an increasingly attractive option for consumers.

In the relentless pursuit of sustainable and eco-friendly transportation, the automotive industry is undergoing a transformative shift from conventional gasoline-powered vehicles to the revolutionary era of Electric Vehicles (EVs). This paradigm shift is not merely a technological evolution but a profound response to environmental challenges, presenting a pragmatic solution to curb emissions and liberate society from the shackles of fossil fuel reliance.

Washington, a state renowned for its commitment to environmental progress, stands at the forefront of this eco-friendly surge. The adoption of EVs in the state has witnessed a noteworthy rise, reflecting a collective consciousness towards a cleaner and greener future. This transition aligns with the state's ethos of embracing innovative solutions for a sustainable tomorrow.

Fueling this transformation is a groundbreaking project focusing on intricate classification methodologies for Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs). A meticulous analysis of EV registrations lies at the heart of this initiative, delving into crucial attributes such as make, model, electric utility, and Clean Alternative Fuel Vehicles (CAFEVs) eligibility. This data-driven approach not only unveils the current landscape of EV usage in Washington but also provides a comprehensive understanding of the diverse factors influencing their adoption.

The environmental impact of this shift to EVs in Washington is profound. By reducing reliance on fossil fuels, the state is significantly cutting down its greenhouse gas emissions, contributing to cleaner air and a healthier environment. This transition also aligns with Washington's broader environmental goals, including the use of renewable energy sources and conservation initiatives.

The strategic significance of this initiative goes beyond mere data collection. It empowers policymakers and manufacturers with pivotal insights, offering a roadmap for strategic decision-making in the ever-evolving EV landscape of Washington state. By understanding the intricacies of EV registrations, stakeholders can make informed choices that influence manufacturing practices, policy formulations, and consumer engagement.

In essence, this project is not just a scientific endeavor; it is a catalyst for change, propelling advancements in electric vehicle research and development. As Washington embraces this innovative approach, it contributes significantly to the scientific evolution of sustainable transportation, paving the way for a cleaner, greener, and more sustainable future.

In conclusion, Washington's journey towards an EV-dominated future is a multifaceted endeavor. It's a blend of technological innovation, policy initiatives, consumer education, infrastructure development, and environmental stewardship. As the state continues to lead the way in sustainable transportation, it sets a precedent for others to follow, paving the way for a greener and more sustainable future globally

### 1.3 Project Goal

The primary objective of this project is to conduct a comprehensive analysis of electric vehicle (EV) registration data in Washington state. By meticulously examining this data, the project aims to identify key trends, consumer preferences, and the various factors that influence the adoption of EVs. This analysis is not just about numbers; it's about understanding the story behind the data, the motivations of EV buyers, and the barriers they face.

Through this analysis, we aim to provide actionable insights for stakeholders to promote sustainable transportation.

### 1.4 Research Questions

We as a group want to tackle the below research questions.

#### 1. EV Adoption Over Time

Question: How has the annual adoption rate of electric vehicles in Washington state evolved over the years?

#### 2. EV Type Preference:

Question: Which type of electric vehicle (e.g., hybrid, full electric) is most popular in Washington state?

#### 3. Price vs. Electric Range:

Question: Is there a correlation between the base MSRP (price) of an electric vehicle and its electric range?

#### 4. CAFV Eligibility Impact:

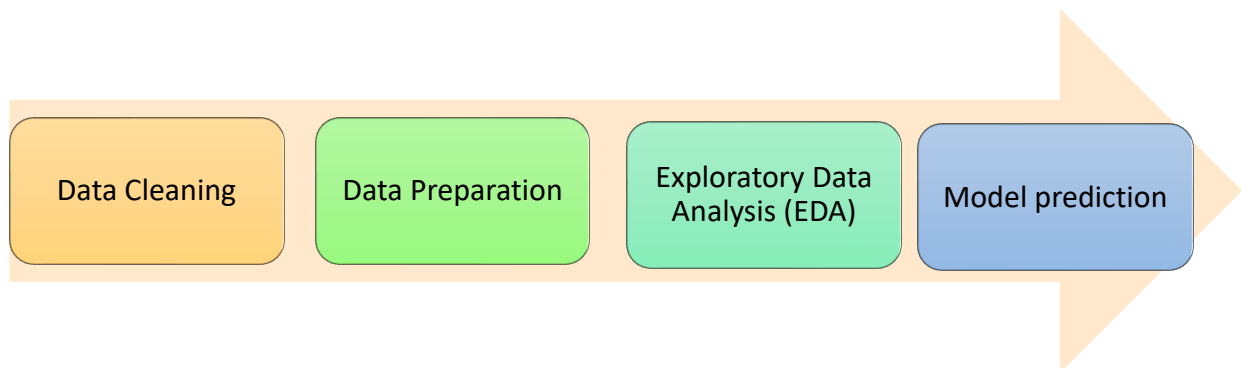
Question: Does the Clean Alternative Fuel Vehicle (CAFV) eligibility influence its electric range?

#### 5. Utility Analysis:

Question: Which electric utilities support the highest number of electric vehicles? Is there a relationship between the electric utility and the type of EV adopted?

## 1.5 Methodology used in the following report.

Following methodology has been used in this project.



### 1.4.1 Data Cleaning:

To ensure the dataset is free from errors, inconsistencies, and inaccuracies, following steps were followed:

- Missing Values: Check for any missing values in the dataset. If found, decide on a strategy to handle them.
- Duplicate Rows: Identify and remove any duplicate rows.
- Outliers: Detect outliers using methods like the IQR (Interquartile Range) or Z-score. Decide on a strategy to handle them.
- Data Types: Ensure that each column is of the correct data.
- Inconsistent Data: Check for inconsistencies in categorical.

### 1.4.2 Data Preparation:

To transform the data into a format suitable for analysis following steps were followed:

- Feature Engineering: Create new features that might be relevant for analysis and prediction.
- Scaling: Standardize or normalize numerical features so they're on a similar scale, especially if you're using algorithms sensitive to feature scales.
- Data Splitting: Split the data into training and testing sets for model validation.

### 1.4.3 Exploratory Data Analysis (EDA):

To understand the data's underlying structure and relationships. Following steps were followed:

- Descriptive Statistics: Calculate measures like mean, median, standard deviation, etc.
- Visualizations: Use plots such as histograms, box plots, scatter plots, and correlation matrices to visualize data distributions and relationships.

### 1.4.4 Model Prediction:

To develop a predictive model based on the data following steps were followed:

- a. Model Selection: Choose a suitable algorithm based on the problem type (e.g., regression, classification etc).
- b. Training: Use the training data to train the model.
- c. Validation: Validate the model's performance using the test data.

## 2. Data

### 2.1 Dataset source:

The dataset for this project is taken from the data.gov website:

(<https://catalog.data.gov/dataset/electric-vehicle-population-data>)

- a. VIN (1-10): The first 10 characters of the Vehicle Identification Number (VIN), which is a unique code used by the automotive industry to identify individual motor vehicles.
- b. County: The county in Washington state where the vehicle is registered.
- c. City: The city in Washington state where the vehicle is registered
- d. State: The states where the vehicle is registered. In this dataset, it should be "Washington."
- e. Postal Code: The postal or ZIP code corresponding to the vehicle's registration address.
- f. Model Year: The year when the vehicle model was manufactured.
- g. Make: The brand or manufacturer of the vehicle (e.g., Tesla, Nissan).
- h. Model: The specific model of the vehicle (e.g., Model S, Leaf).
- i. Electric Vehicle Type: The type of electric vehicle, which could be categories like hybrid, plug-in hybrid etc.
- j. . Clean Alternative Fuel Vehicle (CAFV) Eligibility: Indicates whether the vehicle is eligible for Clean Alternative Fuel Vehicle benefits, which might include tax incentives, rebates, or access to carpool lanes.
- k. 11. Electric Range: The maximum distance the vehicle can travel on a single charge is typically measured in kilometers.
- l. 12. Base MSRP: The Manufacturer's Suggested Retail Price (MSRP) for the base model of the vehicle.
- m. 13. Legislative District: The legislative district in Washington state where the vehicle is registered.
- n. 14. DOL Vehicle ID: A unique identification number is assigned by the Department of Licensing (DOL) to registered vehicle.

- o. 15. Vehicle Location: The precise location or coordinates where the vehicle is registered.
- p. 16. Electric Utility: The electric utility company serving the area where the vehicle is registered.
- q. 17. 2020 Census Tract: A geographic region defined for the purpose of the 2020 census

## 2.2 Data Cleaning:

During the data preprocessing phase, a rigorous approach was undertaken to enhance dataset integrity. Initial measures involved the elimination of instances with missing values within the specific County and City subset, resulting in a refined dataset.

Subsequent attention was directed towards numeric columns, notably Legislative District, where missing values were identified. To preserve statistical integrity, a judicious imputation strategy was applied. The missing values in Legislative District were addressed through a discerning selection between the median and mean methods.

Concurrently, the treatment of missing values extended to categorical columns, such as Model and Vehicle Location. For these variables, missing values were substituted with the mode value for each respective column. This method ensured the replacement of missing values with the most frequently occurring categorical values within the dataset.

The implementation of these systematic approaches for missing value handling facilitated a comprehensive refinement of the dataset. This process ensures that subsequent analyses and modeling endeavours are founded upon a robust dataset characterized by completeness and accuracy. Such meticulous attention to data quality not only elevates the reliability of the dataset but also enhances the overall robustness and validity of the analytical insights derived from it.

```
#dataset info
dataframe.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150482 entries, 0 to 150481
Data columns (total 17 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   VIN (1-10)            150482 non-null object
 1   County                150479 non-null object
 2   City                  150479 non-null object
 3   State                 150482 non-null object
 4   Postal Code           150479 non-null float64
 5   Model Year            150482 non-null int64
 6   Make                  150482 non-null object
 7   Model                 150482 non-null object
 8   Electric Vehicle Type 150482 non-null object
 9   Clean Alternative Fuel Vehicle (CAFV) Eligibility 150482 non-null object
10   Electric Range         150482 non-null int64
11   Base MSRP              150482 non-null int64
12   Legislative District    150141 non-null float64
13   DOL Vehicle ID         150482 non-null int64
14   Vehicle Location       150475 non-null object
15   Electric Utility        150479 non-null object
16   2020 Census Tract      150479 non-null float64
dtypes: float64(3), int64(4), object(10)
memory usage: 19.5+ MB
```

```
# show missing values of data
dataframe.isnull().sum()

VIN (1-10)                0
County                    3
City                      3
State                     0
Postal Code                3
Make                      0
Model                     0
Electric Vehicle Type      0
Clean Alternative Fuel Vehicle (CAFV) Eligibility 0
Electric Range             0
Base MSRP                  0
Legislative District       341
DOL Vehicle ID             0
Vehicle Location           7
Electric Utility            3
2020 Census Tract          3
dtype: int64
```



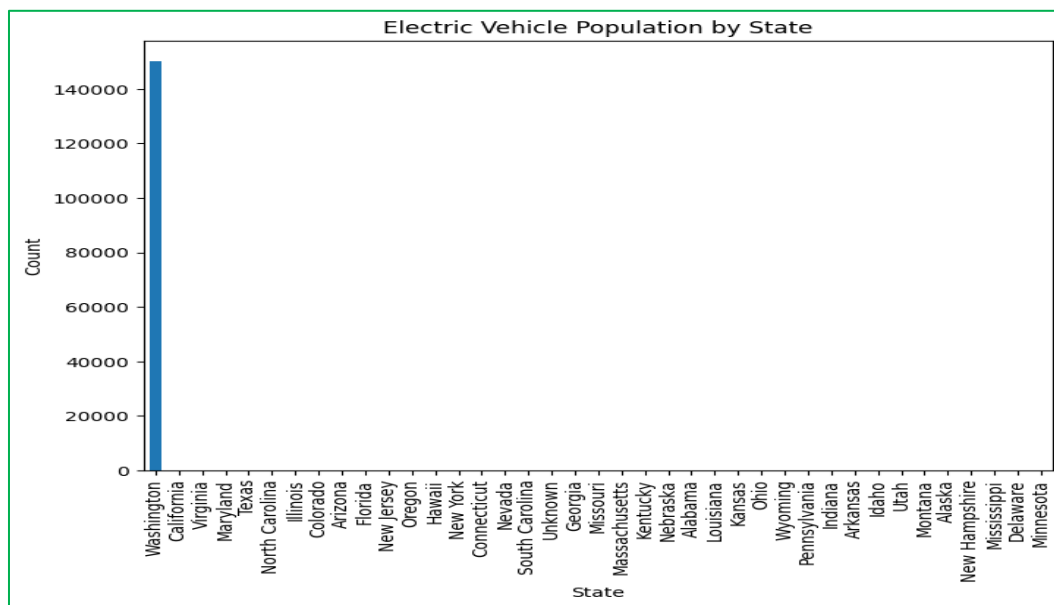
## 2.3 Feature Engineering

- a. **Electric Vehicle Type Updated:** Replaced values within the 'Electric Vehicle Type' column, specifically abbreviating 'Plug-in Hybrid Electric Vehicle (PHEV)' to 'PHEV' and 'Battery Electric Vehicle (BEV)' to 'BEV' for brevity and clarity.
- b. **Column Renamed:** Underwent a renaming process for the 'Clean Alternative Fuel Vehicle (CAFV) Eligibility' column, resulting in its new nomenclature as 'Clean Alternative Fuel Vehicle Eligibility' to align with improved clarity and consistency.
- c. **Longitude and Latitude Extracted:** Executed an extraction procedure to derive longitude and latitude information from the 'Vehicle Location' column, subsequently creating distinct 'Longitude' and 'Latitude' columns to enhance spatial data representation.
- d. **State Codes Mapped to Names** Employed a mapping mechanism to associate state codes present in the 'State' column with their corresponding state names, ensuring a more comprehensible and informative dataset.
- e. **Substring Extraction:** Implemented substring extraction techniques on the 'Electric Utility' column, focusing on capturing the initial substring to augment data clarity and facilitate more precise analysis.
- f. **Categorical to Numerical Transformation:** For the purpose of modeling we need the categorical data in numerical format.
  - i. **Ordinal Encoding Applied:** Acknowledging the modeling requirements, employed scikit-learns Ordinal Encoder to transform categorical data in the 'State,' 'Make,' 'Electric Vehicle Type,' and 'Clean Alternative Fuel Vehicle Eligibility' columns into ordinal numeric values. Ensured consistency by converting the encoded columns to integers for uniform numerical representation
  - ii. **Frequency Encoding Applied:** Leveraged the capabilities of the category\_encoders library to execute frequency encoding on pertinent columns such as 'VIN (1-10),' 'County,' 'City,' 'Model,' and 'Electric Utility.' Updated the original dataset with the frequency-encoded values, enhancing its numerical representation.
  - iii. **Resulting Dataset:** The 'train' dataset now incorporates the applied ordinal and frequency encoding, enhancing its numerical features.
- g. Also note that for EDA (Univariate and bivariate analysis) the original data frame was used. Train dataset was used in Modeling.

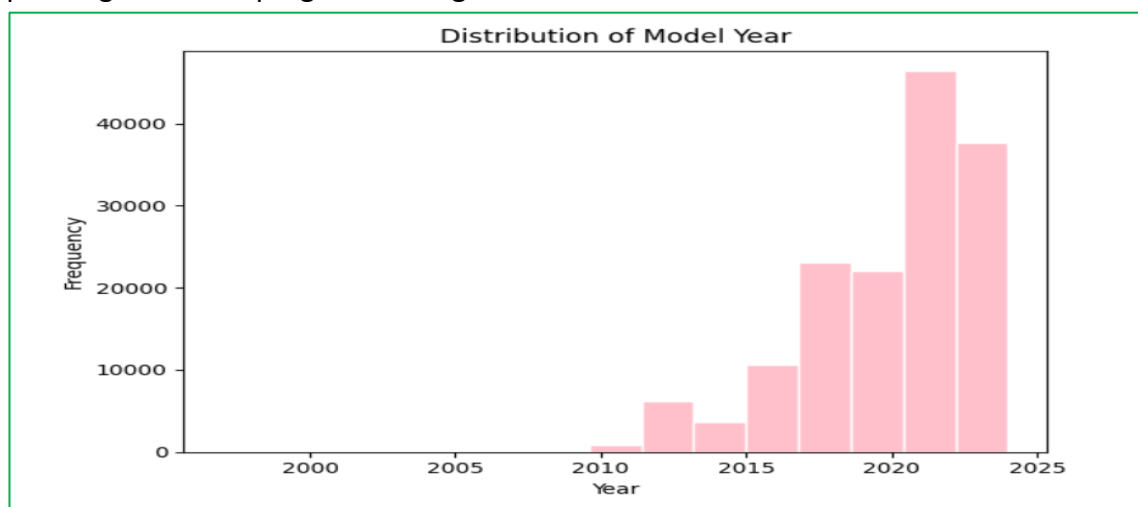
### 3 Exploratory Data Analysis (EDA)

#### 3.1 EDA -Univariate

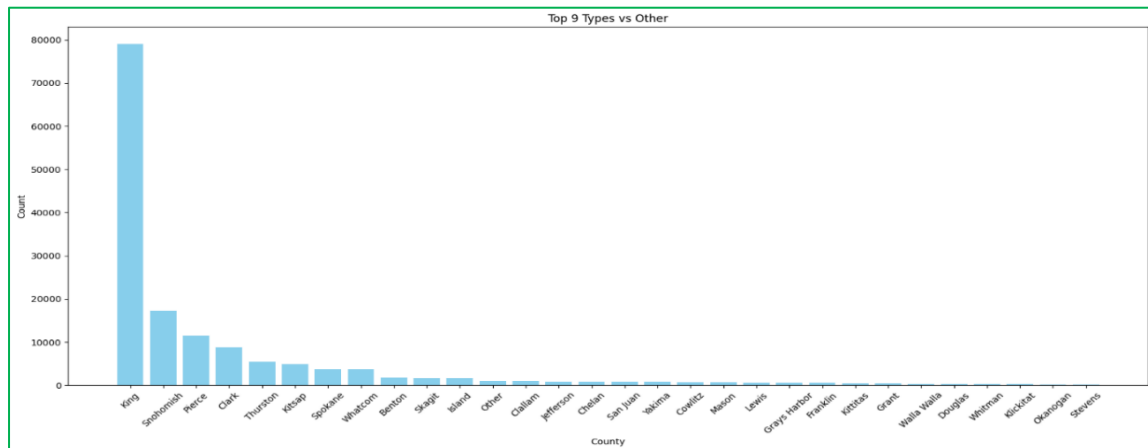
- a) The electric vehicle population by state vs count chart: The chart depicting the distribution of electric vehicle populations across states reveals a notable trend, with Washington emerging as the leader in terms of the sheer count of electric vehicles. This data visualization underscores that Washington boasts the highest number of electric vehicles compared to other states, signifying a substantial presence and adoption of electric mobility within the region. The prominence of Washington in this context suggests a robust embrace of electric vehicles, indicative of a progressive stance towards sustainable and eco-friendly transportation alternatives.



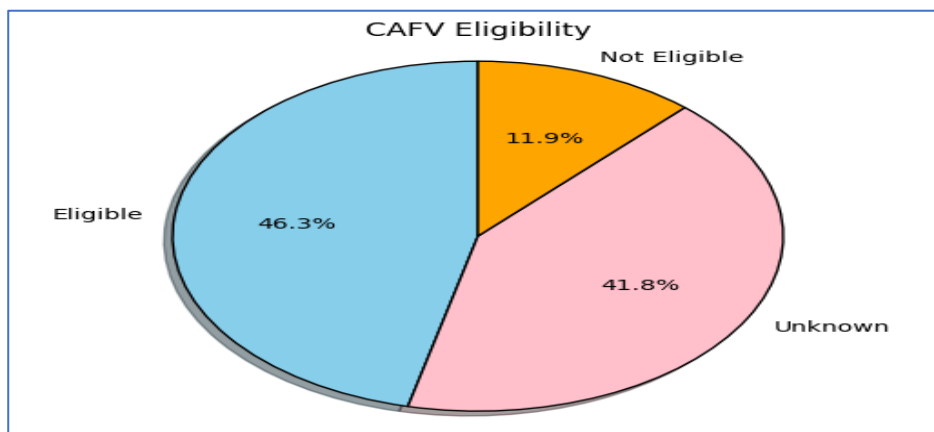
- b) The distribution of the car vs follows exponential growth. : The analysis of car distribution reveals a noteworthy pattern characterized by exponential growth over the observed period. This exponential trend underscores a significant and consistent increase in the number of cars, pointing towards a progressive surge in automobile occurrence.



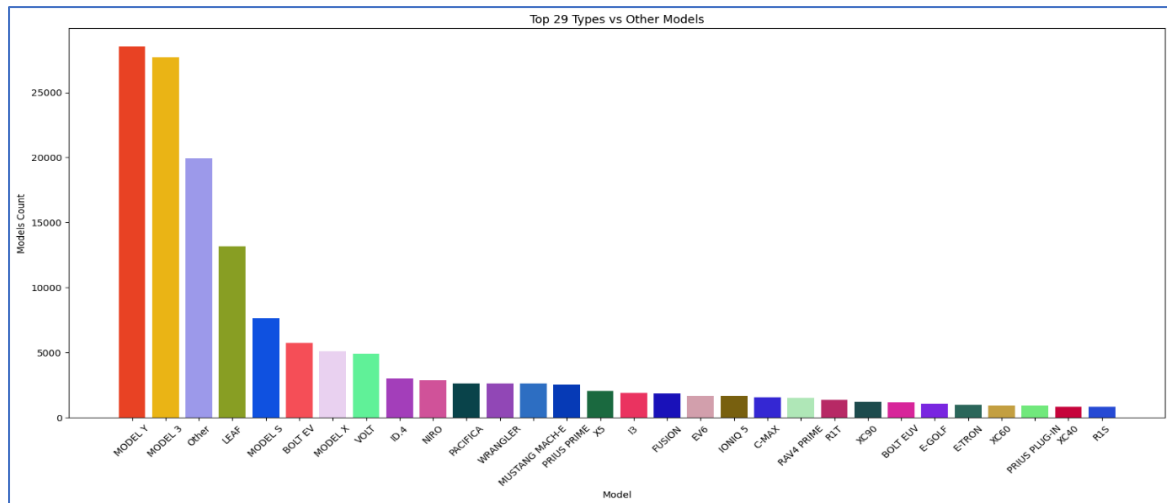
c) **Top county vs frequency chart:** The frequency chart depicting the top counties and their respective frequencies highlights King County as the frontrunner with a substantial count of 79,075. Following closely are Snohomish County with 17,307, Pierce County with 11,542, Clark County with 8,855, Thurston County with 5,403, and Kitsap County with 4,923. This distribution provides a clear ranking of the most prevalent counties, emphasizing the dominance of King County in terms of the frequency of occurrences.



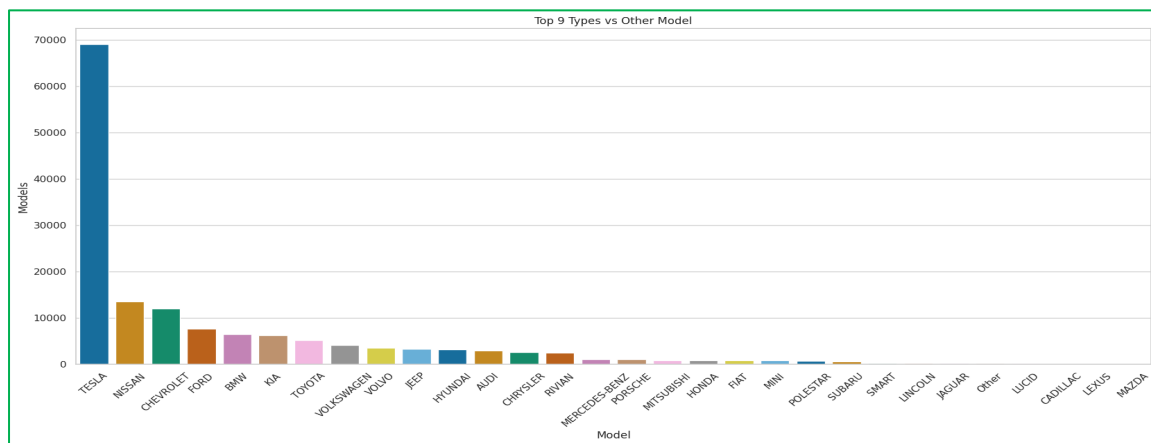
d) **Clean air fuel eligibility:** The calculation of Clean Air Fuel Vehicle (CAFV) eligibility reveals that approximately 47% of the vehicles in the dataset meet the criteria for CAFV status. This percentage signifies the proportion of vehicles considered environmentally friendly based on their adherence to clean air fuel standards. The dataset's composition indicates a considerable presence of vehicles classified as CAFV-eligible, reflecting a significant commitment to and adoption of cleaner and more sustainable fuel options.



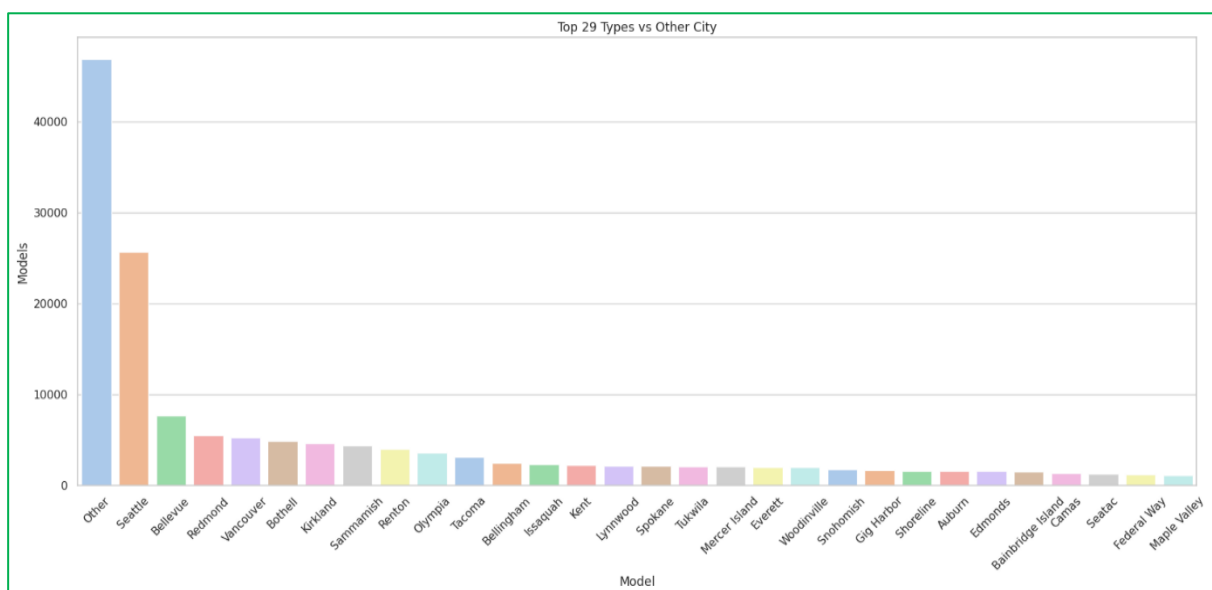
e) **Model vs Model count graph.** The graph depicting the count of different models reveals that Model Y and Model 4 stand out as the most popular choices.



f) Model vs models: The model distribution indicates Tesla as the most popular, trailed by Nissan and Chevrolet in terms of popularity.

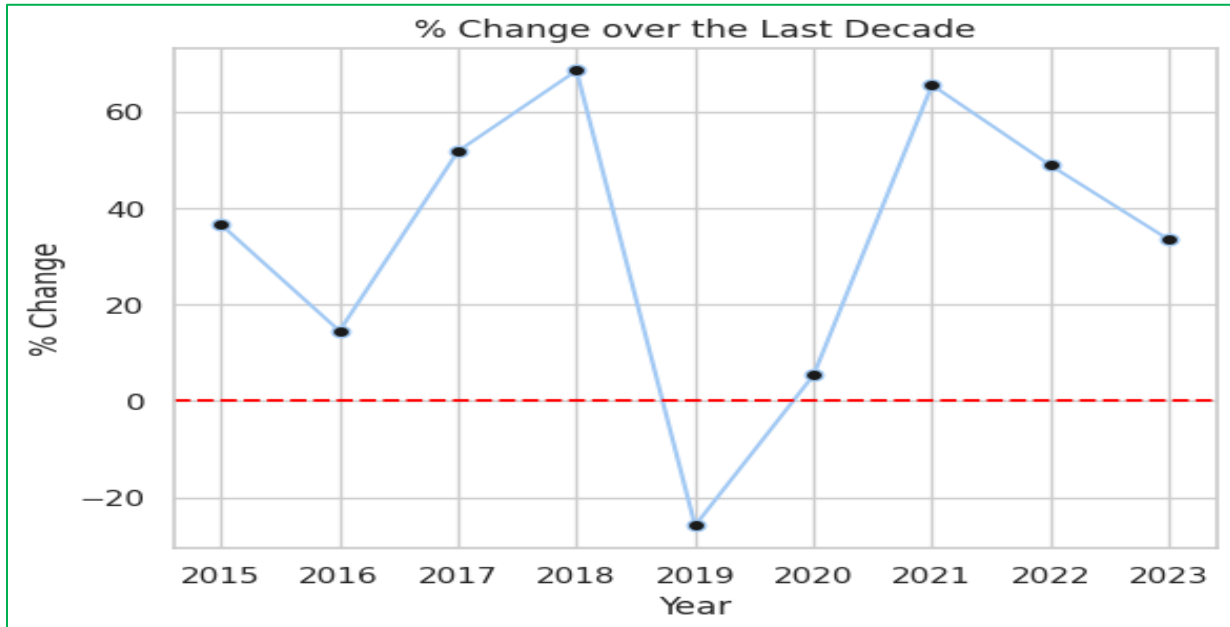


g) Model's vs model count chart is shown below.

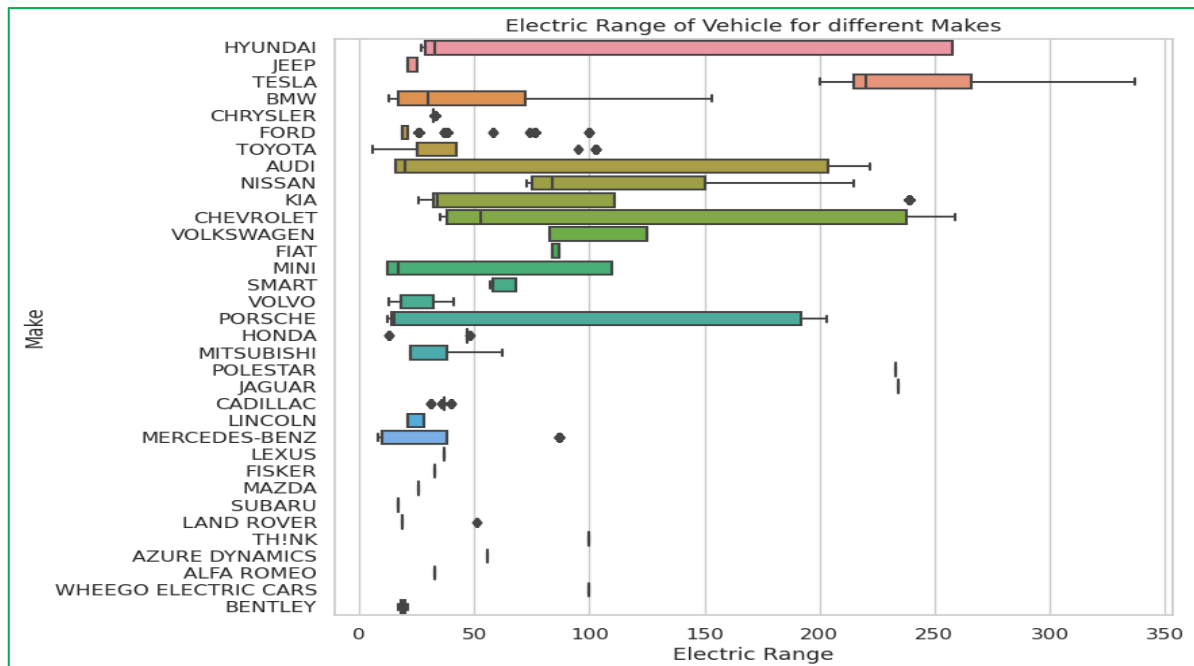


### 3.2 EDA-Bivariate

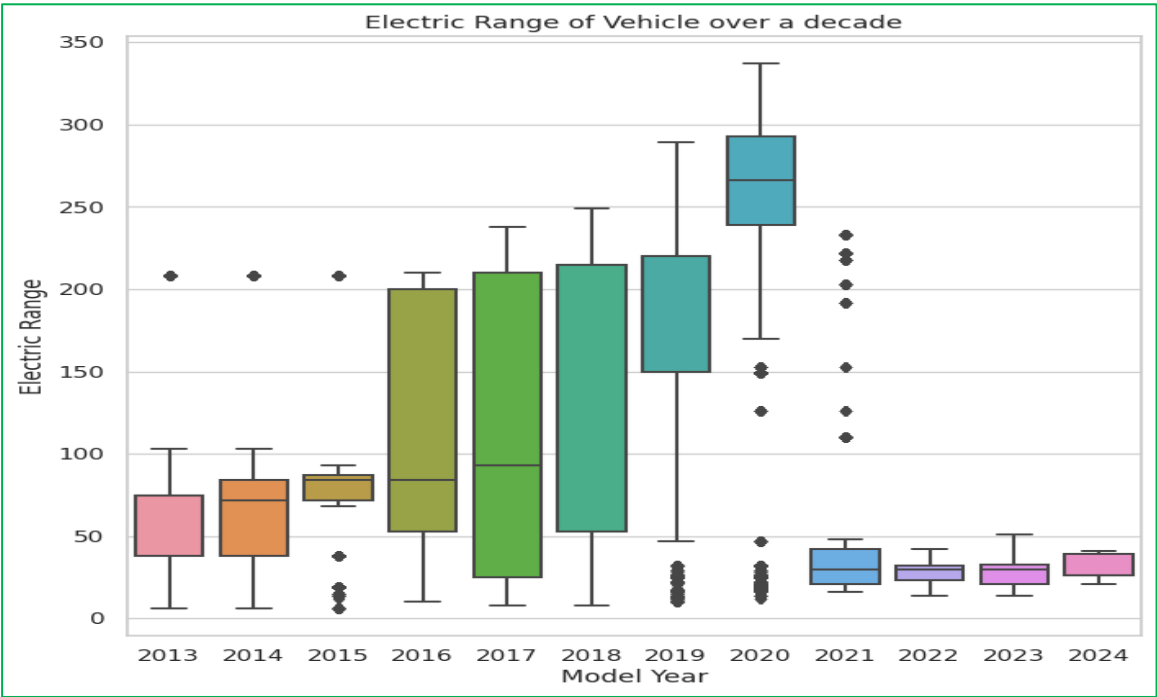
- a) The percentage change over the last decade: The longitudinal assessment of percentage change over the past decade unveils a persistent upward trajectory, marred only by a substantial dip in 2019, promptly succeeded by a vigorous recovery. However, the current analysis of percentage change in 2023 indicates a discernible downturn.



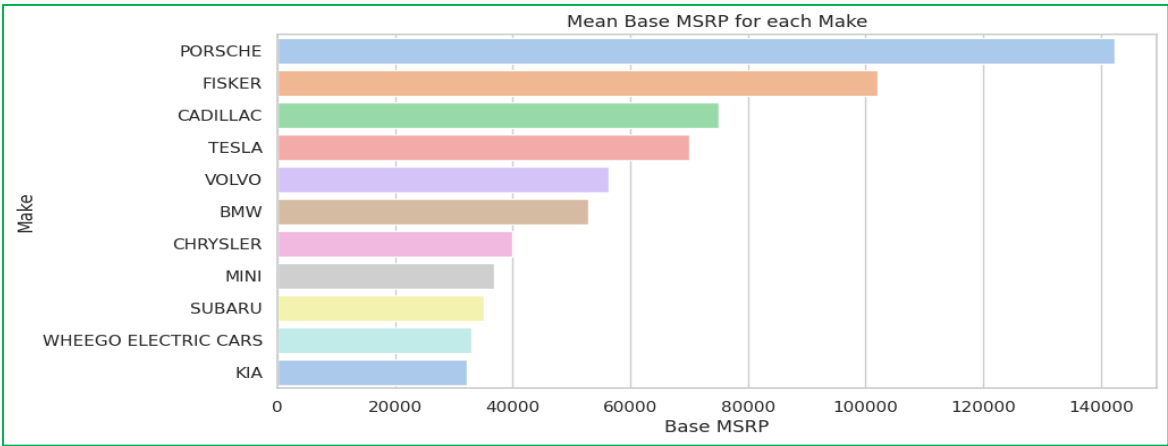
- b) Electric range vs make of the vehicles: In examining the electric range across various vehicle makes, it is evident that Tesla surpasses other manufacturers, boasting the highest electric range among the observed models.



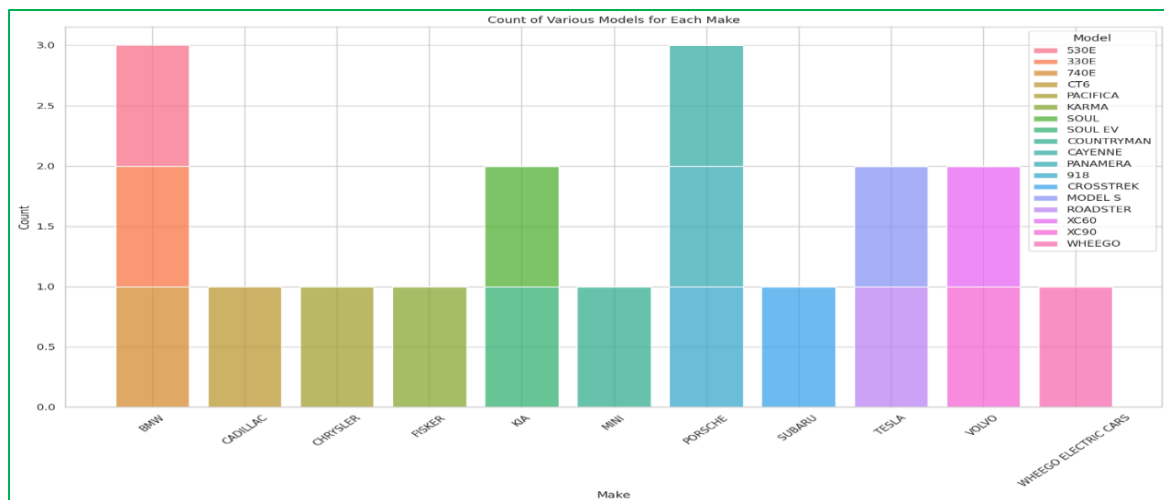
c) Electric range vs Model year: Looking at the electric range in relation to the model year, we notice a significant increase in range from 2016 to 202



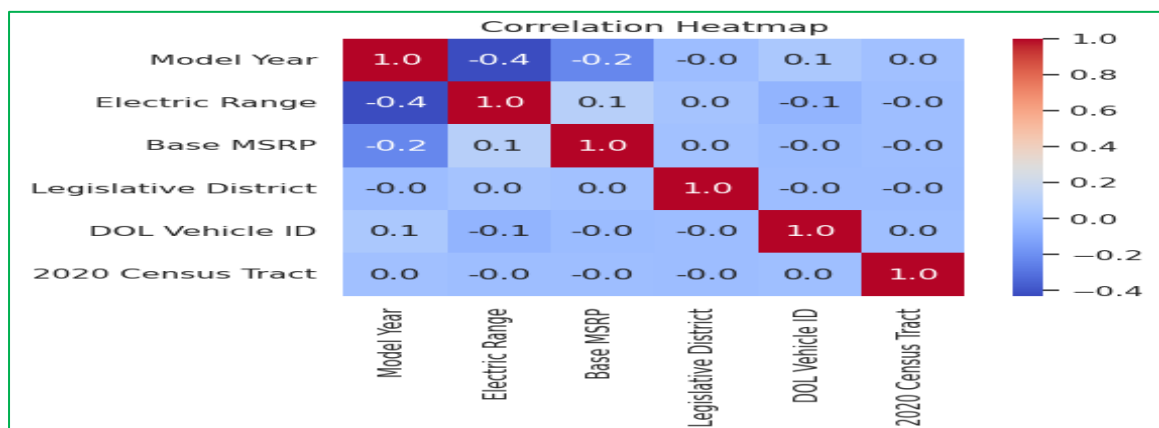
d) Make vs base MSRP: In the chart of Vehicle make and base Manufacturer's Suggested Retail Price (MSRP), it is evident that Porsche commands the highest base MSRP, with Fisker and Cadillac following closely in the rankings.



e) Count of various model of each make : The chart provides an overview of the number of models offered by each car company in the current dataset, highlighting the existence of multiple variations within each company.



f) Correlation Heatmap of the data: The chart indicates a lack of strong linear correlations between numerous features and the target variable. This implies that the dataset predominantly comprises non-linear correlations.

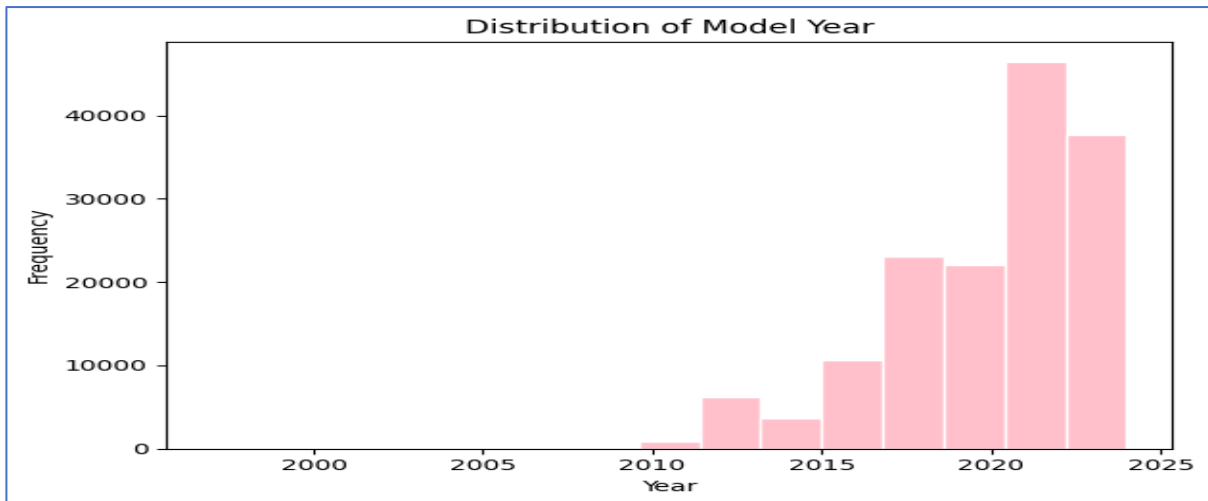


## 4 Research Questions Analysis

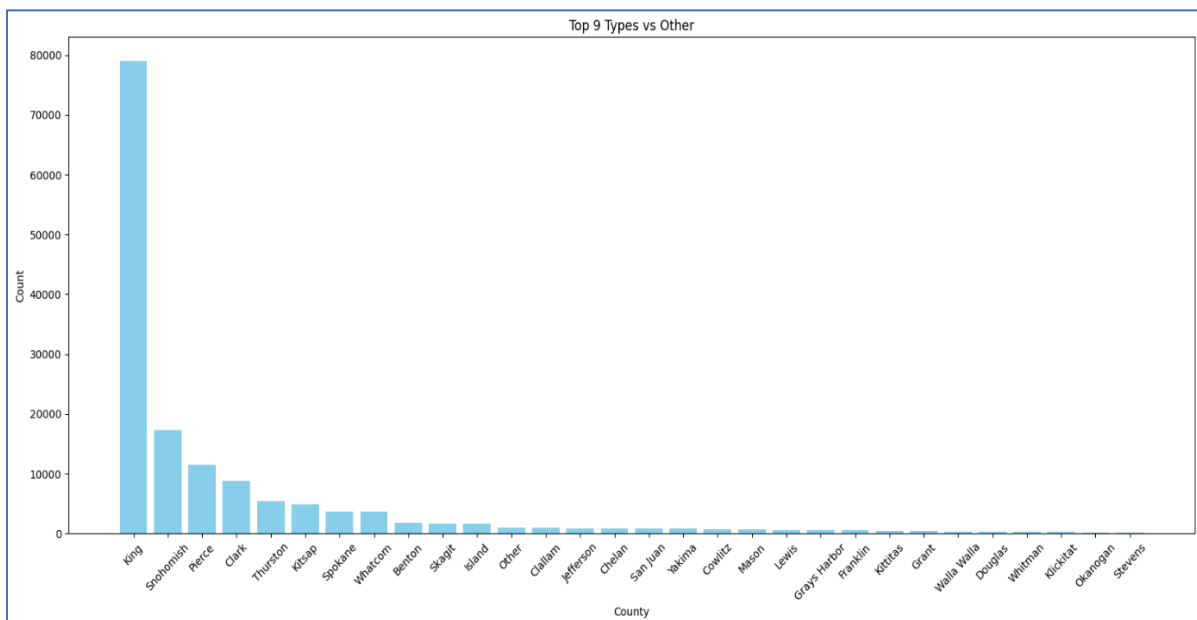
### 4.1 EV Adoption Over Time

**Question:** How has the annual adoption rate of electric vehicles in Washington state evolved over the years?

The annual adoption rate has closely mirrored exponential growth, with a minor dip observed in 2019. It's important to note that 2023 data is still ongoing, preventing a conclusive analysis of the last bar and any generalized statements about the current year.



The counties with the highest count, signalling substantial presence, are King, Snohomish, and Pierce.

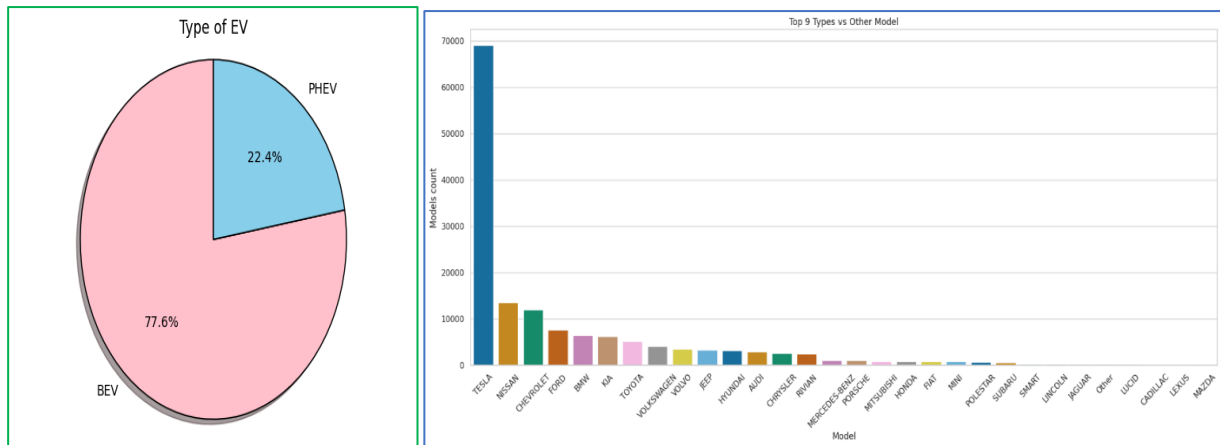


## 4.2 EV Type Preference

**Question:** Which type of electric vehicle is most popular in Washington state?

BEV emerges as the more popular electric vehicle, capturing a significant 77.6%. Within the electric vehicle landscape, Tesla takes the lead, closely followed by Nissan, making them the most renowned electric vehicle manufacturers.

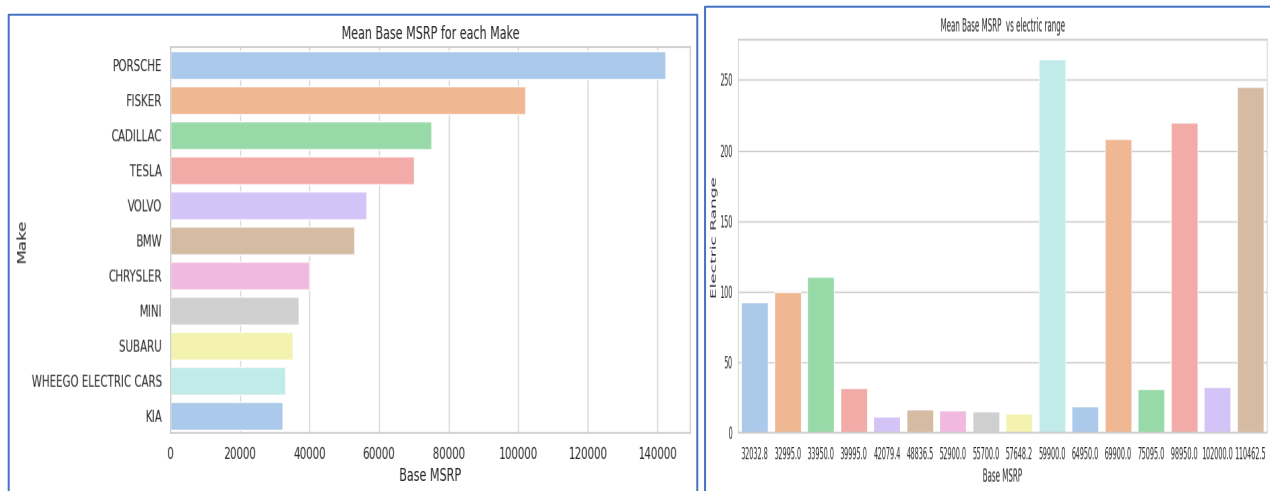




#### 4.3 Price vs. Electric Range

**Question:** Is there a correlation between the base MSRP (price) of an electric vehicle and its electric range?

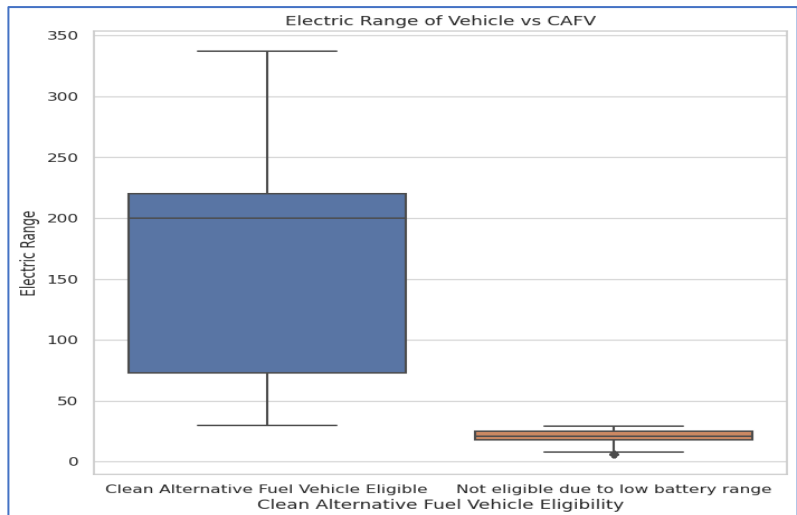
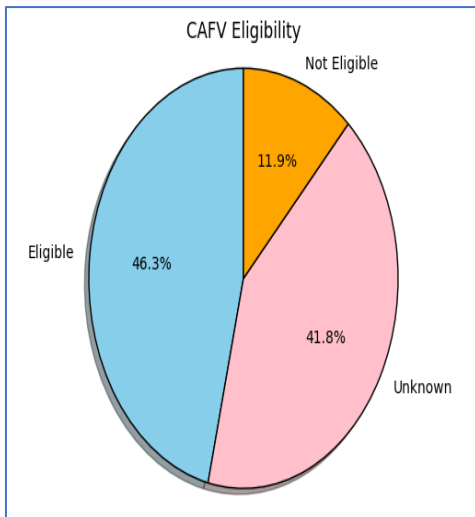
The elevated Base Manufacturer's Suggested Retail Price (MSRP) for Porsche, Fisker, and Cadillac highlights their premium positioning in the market. Additionally, the positive correlation between higher electric range and increased MSRP emphasizes the significance of pricing about the extended electric range, indicating a potential premium associated with enhanced electric capabilities.



#### 4.4 CAFV Eligibility Impact:

**Question:** Does the Clean Alternative Fuel Vehicle (CAFV) eligibility influence its electric range?

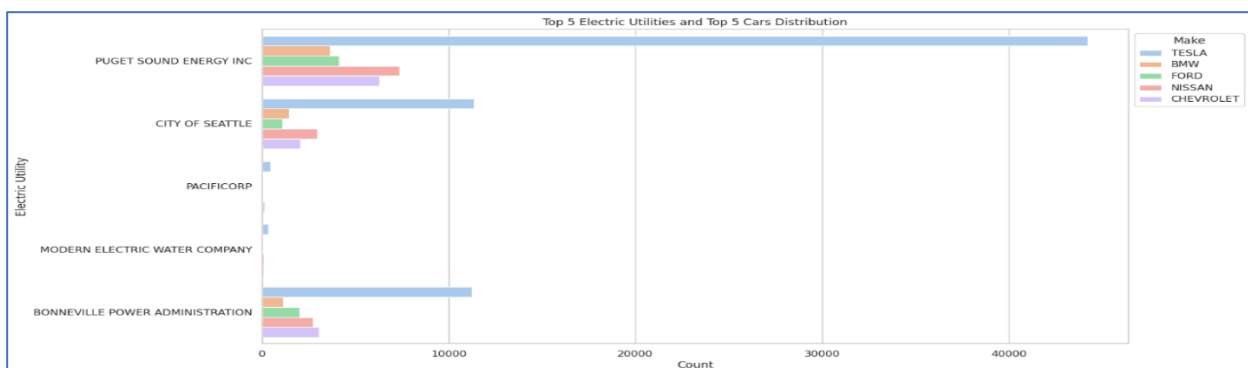
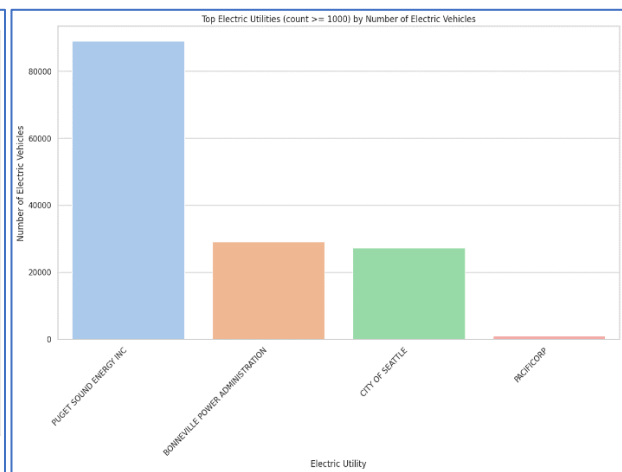
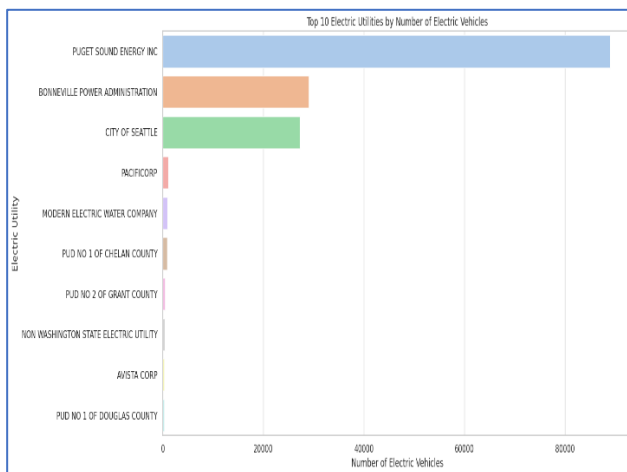
Knowing that about 46.3% of vehicles are eligible for Clean Alternative Fuel Vehicle (CAFV) highlights a considerable portion adhering to eco-friendly standards. Interestingly, CAFV eligibility is more common among vehicles with a longer electric range compared to those that are not eligible. This emphasizes the significance of considering CAFV eligibility, particularly when prioritizing vehicles with extended electric capabilities. The higher prevalence of CAFV eligibility in such vehicles suggests a collective commitment towards adopting cleaner and more sustainable fuel options in the electric vehicle domain.



#### 4.5 Utility Analysis:

**Question:** Which electric utilities support the highest number of electric vehicles? Is there a relationship between the electric utility and the type of EV adopted?

The graph illustrates the top 10 electric utilities in relation to the number of electric vehicles. Puget Sound Energy Inc. emerges as the most utilized utility, indicating its prominence in facilitating the charging infrastructure for electric vehicles.



## 5. Modelling

Three distinct models were subjected to rigorous testing to discern their efficacy in predicting and understanding patterns within electric vehicle (EV) data:

- **Naïve Bayes:** Naïve Bayes is a probabilistic algorithm, grounded in Bayes' theorem, that assumes conditional independence among features. In the context of EV data analysis, this model predicts adoption patterns by considering factors such as location, specific model characteristics, and eligibility criteria. By leveraging these features, Naïve Bayes provides valuable insights into the nuanced dynamics influencing EV adoption, making it particularly adept at handling probabilistic scenarios.
- **Logistic Regression:** Logistic Regression is a statistical method designed for modeling the log-odds of binary outcomes using the logistic function. In the realm of EV data, this model excels in binary classification tasks, proficiently forecasting categorical outcomes, such as regional EV adoption. By capturing the relationship between various input features and the likelihood of a specific event occurring, Logistic Regression proves to be a robust tool for understanding and predicting adoption patterns in EV datasets.
- **Random Forest:** Random Forest stands out as an ensemble learning technique that harnesses the power of multiple decision trees. In the context of EV data analysis, this model shines in uncovering intricate relationships and patterns that influence adoption. By accommodating various factors and their interactions, Random Forest provides a comprehensive understanding of the multifaceted influences on EV adoption, making it particularly suitable for capturing the complexity inherent in the dataset.

These three models, each with its unique strengths and characteristics, were carefully chosen and evaluated to offer a holistic perspective on the diverse aspects of electric vehicle adoption patterns. The combination of probabilistic modeling, statistical analysis, and ensemble learning provides a robust framework for gaining valuable insights into the dynamics shaping the adoption landscape in the realm of electric vehicles.

Following metrics collectively provide a comprehensive evaluation of the model's accuracy, discrimination, and effectiveness in classifying instances within the dataset.

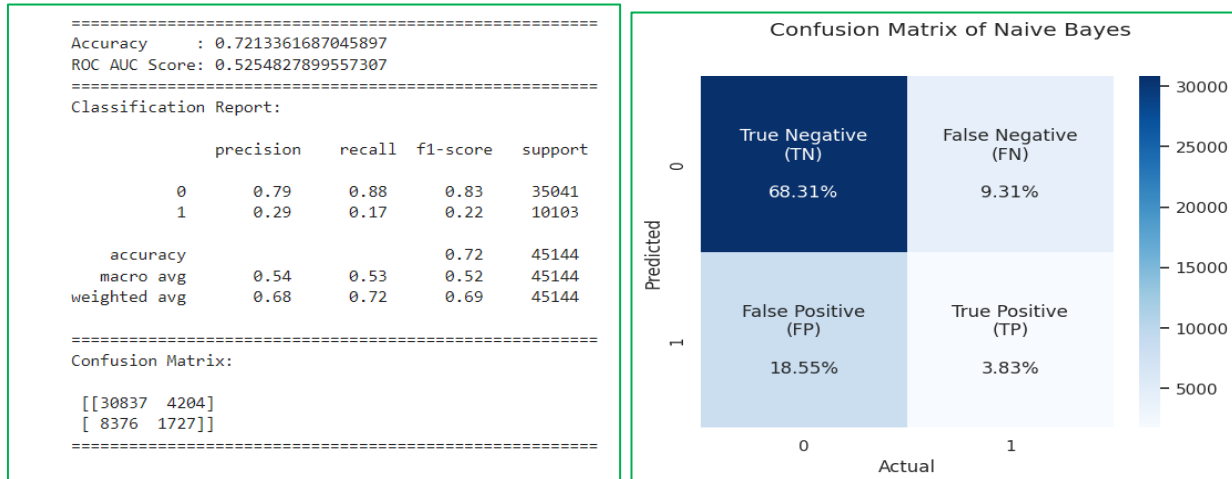
- **Accuracy:** Overall correctness of the model's predictions.
- **ROC AUC Score:** Measures the model's discrimination ability between classes.
- **Classification Report:** Summarizes precision, recall, and F1 score for each class.
- **Precision:** Accuracy of positive predictions.
- **Recall:** Model's ability to capture true positive instances.
- **F1 Score:** Harmonic mean of precision and recall, balancing both.
- **Confusion Matrix:** Tabular representation of predicted versus actual classifications.

Fine-tuning post-hyperparameter adjustment involves iteratively refining model parameters for optimal performance. After initial hyperparameter tuning, assess results, and iteratively train the model. Evaluate on a validation set, monitor learning curves, and consider

grid/random search. Apply regularization techniques and ensemble methods, ensuring stability via cross-validation. This process maximizes model accuracy and generalization.

## 5.1 Naïve Bayes:

The following results were obtained for Naïve Bayes model:



The cross-validation accuracy and ROC AUC score for untuned Naive Bayes after a 10 fold cross validation:  
Accuracy : 0.5800949316829318  
ROC AUC : 0.7912524530200896

After finetuning the above adjusting hyper parameters and regularization techniques.

Best hyperparameters for accuracy and ROC AUC:

```

Accuracy :
{'var_smoothing': 1e-09}
ROC AUC :
{'var_smoothing': 1e-09}

```

The cross-validation accuracy and ROC AUC score for tuned Naive Bayes after a 10 fold cross validation:  
Accuracy : 0.5800949316829318  
ROC AUC : 0.7912524530200896



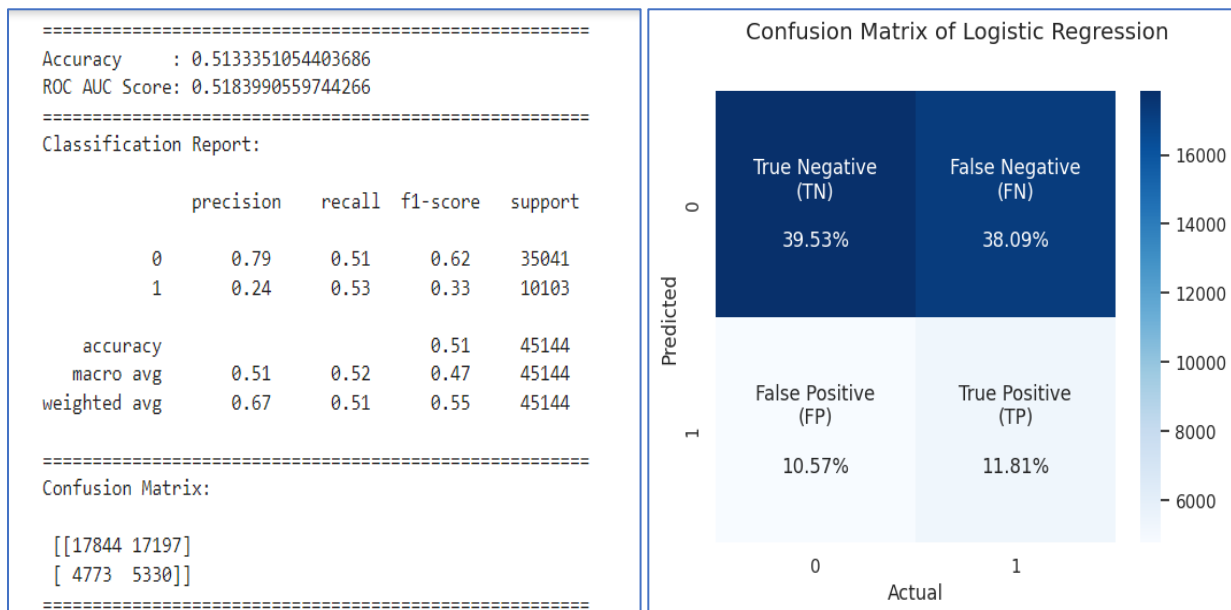
```

=====
Train Data
Accuracy : 0.5800949316829318
ROC AUC  : 0.7912524530200896
=====
Test Data
Accuracy : 0.7755184976382079
ROC AUC  : 0.7152844042989862

```

## 5.2 Logistic Regression

Following results were obtained for Logistic regression model:



The cross-validation accuracy and ROC AUC score for untuned Logistic Regression after a 10 fold cross validation:  
Accuracy : 0.5952402288978342  
ROC AUC : 0.5796176387162966

After finetuning the above adjusting hyper parameters and regularization techniques.

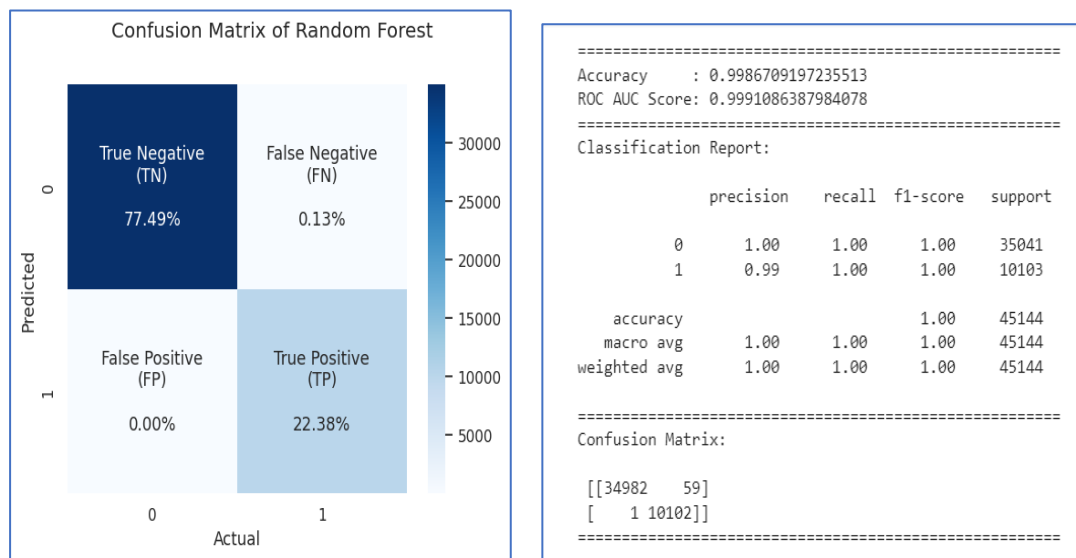
```
Best hyperparameters for accuracy and ROC AUC:
Accuracy :
{'C': 0.1, 'class_weight': 'balanced', 'fit_intercept': True, 'max_iter': 100, 'penalty': 'l2', 'random_state': None, 'warm_start': True}
ROC AUC :
{'C': 0.1, 'class_weight': 'balanced', 'fit_intercept': True, 'max_iter': 100, 'penalty': 'l2', 'random_state': None, 'warm_start': True}

The cross-validation accuracy and ROC AUC score for tuned Logistic Regression after a 10 fold cross validation:
Accuracy : 0.5954310595770547
ROC AUC : 0.5796190602949384
```

```
=====
Train Data results for Logistic Regression
Accuracy : 0.5954310595770547
ROC AUC : 0.5796190602949384
=====
Test Data results for Logistic Regression
Accuracy : 0.6383933680426246
ROC AUC : 0.5270226101152191
=====
```

### 5.3 Random Forest

Following results were obtained for the Random Forest model:



The cross-validation accuracy and ROC AUC score for untuned Random Forest after a 10 fold cross validation:

```
Accuracy : 0.9996181901702068
ROC AUC : 0.9999761045231779
```

After finetuning the above adjusting hyper parameters and regularization techniques.

Best hyperparameters for accuracy and ROC AUC:

```
Accuracy :
{'bootstrap': False, 'max_depth': None, 'n_estimators': 100}
ROC AUC :
{'bootstrap': True, 'max_depth': 10, 'n_estimators': 150}
```

The cross-validation accuracy and ROC AUC score for tuned Random Forest after a 10-fold cross-validation:

```
Accuracy : 0.9997454616131819
ROC AUC : 0.9999902195268531
```

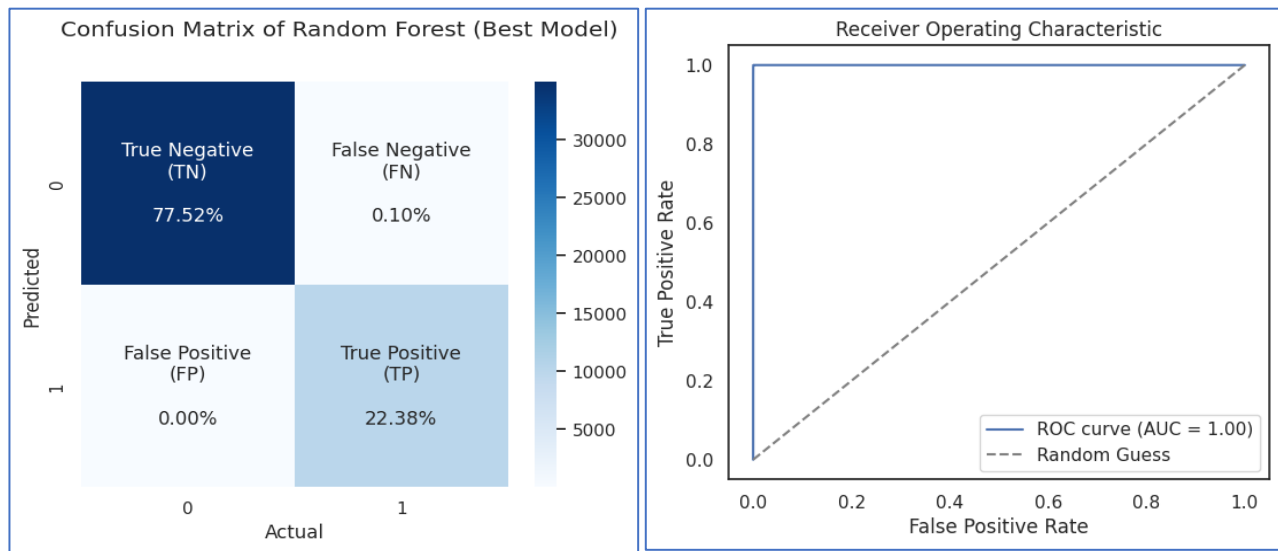
```
=====
Train Data results for Random Forest
Accuracy : 0.9997454616131819
ROC AUC : 0.9999765187990942
=====
Test Data results for Random Forest
Accuracy : 0.9997120218088575
ROC AUC : 0.9999809242392395
```

### 5.4 Comparison of three models

The table provides detailed insights into the modeling outcomes, including accuracy scores and ROC-AUC for both training and testing sets, for three distinct models—Naive Bayes, Random Forest, and Logistic Regression. Upon thorough examination, it becomes apparent that Random Forest emerges as the most effective approach, exhibiting superior performance in terms of accuracy and AUC-ROC compared to Naive Bayes and Logistic Regression.

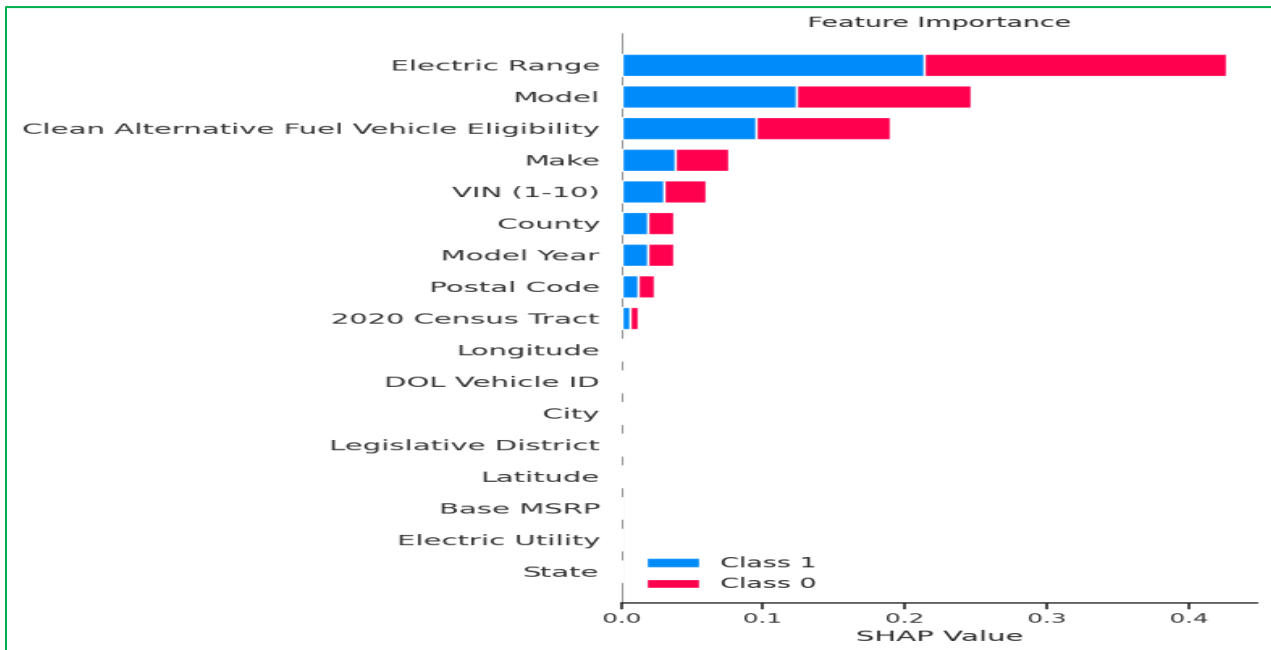
	Model	Accuracy (train)	Accuracy (test)	ROC-AUC (train)	ROC-AUC (test)
0	Random Forest	0.999745	0.999734	0.999976	0.999995
1	Naive Bayes	0.580095	0.775518	0.791252	0.715284
2	Logistic Regression	0.595431	0.638393	0.579619	0.527023

Confusion matrix for the best model and operating characteristic.



## 5.5 Feature importance

In the preceding analysis, SHAP (SHapley Additive exPlanations) values were employed to interpret predictions made by the Random Forest model for electric vehicle types. A SHAP explainer object was defined to extract values reflecting the impact of various features. Subsequently, a bar plot was generated to visually depict the importance of each feature in the model's decision-making process. Features such as VIN (1-10), County, City, etc., were evaluated for their influence on predictions. This method facilitated a comprehensive understanding of the factors driving the model's outputs, enhancing transparency and trust in the model's predictive capabilities. The resulting summary plot served as a valuable tool for interpreting the Random Forest model and informing data-driven decision-making in the realm of electric vehicle predictions.



Notably, the 'Electric Range' exhibited the highest importance, contributing approximately 43.88% to the model's decision-making. Following closely were 'Model' and 'Clean Alternative Fuel Vehicle Eligibility,' contributing 18.84% and 16.87%, respectively. 'Make' and 'VIN (1-10)' also played roles, contributing 6.64% and 4.07%, respectively. These findings underscore the key determinants in the model's predictions, with electric range, specific vehicle models, and eligibility for clean alternative fuel vehicles standing out as pivotal factors influencing the outcomes.





## 5.6 Predictions based on best model

The prediction code generated a DataFrame to store predictions made by the best-performing Random Forest model on the original training data. The predictions were then concatenated with relevant columns from the initial dataset, facilitating a comprehensive comparison between predicted and actual electric vehicle types.

The resulting 'result' DataFrame included crucial details such as VIN, location, and vehicle characteristics. Numerical labels are converted to categorical types for better interpretation, illustrating the model's accuracy in predicting electric vehicle types. This process served to evaluate the model's past performance, offering valuable insights into its predictive accuracy and enhancing our understanding of electric vehicle adoption patterns.

A has feature names, but RandomForestClassifier was fitted without feature names

	Predicted	Actual	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	...	Electric Range	Base MSRP	Legislative District	DOL Vehicle ID	Electric Utility	2020 Census Tract	Longitude	Latitude	Electric Vehicle Type (Actual)	Electric Vehicle Type (Predicted)
0	0	0	2.0	79075.0	25675.0	37.0	98103	2020	13.0	208.0	...	258	0	43	249675142.0	27269.0	53033004800	-122.343010	47.659185	BEV	BEV
1	1	1	7.0	79075.0	4861.0	37.0	98011	2022	15.0	663.0	...	25	0	1	233928502.0	88942.0	53033021804	-122.205780	47.762405	PHEV	PHEV
2	1	1	4.0	835.0	552.0	37.0	98908	2023	15.0	663.0	...	25	0	14	229675939.0	1102.0	53077002900	-120.602720	46.596562	PHEV	PHEV
3	0	0	245.0	79075.0	4622.0	37.0	98034	2018	31.0	27708.0	...	215	0	45	104714466.0	88942.0	53033021903	-122.209285	47.711240	BEV	BEV
4	1	1	8.0	5403.0	3634.0	37.0	98501	2018	4.0	1917.0	...	97	0	22	185498386.0	88942.0	53067010700	-122.896920	47.043535	PHEV	PHEV

The table showcases the comparison between predicted and actual electric vehicle types. The 'Predicted' column contains the model's predictions, while the 'Actual' column denotes the true electric vehicle types. Relevant details, including VIN, location (County, City, State, Postal Code), model year, make, and model, are presented for each entry. Electric vehicle characteristics such as electric range and base MSRP are also included.

	Predicted	Actual	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	...	Electric Range	Base MSRP	Legislative District	DOL Vehicle ID	Electric Utility	2020 Census Tract	Longitude	Latitude	Electric Vehicle Type (Actual)	Electric Vehicle Type (Predicted)
0	0	0	KM8K33AGXL	King	Seattle	Washington	98103	2020	HYUNDAI	KONA	...	258	0	43	249675142.0	CITY OF SEATTLE	53033004800	-122.343010	47.659185	BEV	BEV
1	1	1	1C4RJYB61N	King	Bothell	Washington	98011	2022	JEEP	GRAND CHEROKEE	...	25	0	1	233928502.0	PUGET SOUND ENERGY INC	53033021804	-122.205780	47.762405	PHEV	PHEV
2	1	1	1C4RJYD61P	Yakima	Yakima	Washington	98908	2023	JEEP	GRAND CHEROKEE	...	25	0	14	229675939.0	PACIFICORP	53077002900	-120.602720	46.596562	PHEV	PHEV
3	0	0	5YJ3E1EA7J	King	Kirkland	Washington	98034	2018	TESLA	MODEL 3	...	215	0	45	104714466.0	PUGET SOUND ENERGY INC	53033021903	-122.209285	47.711240	BEV	BEV
4	1	1	WBY7Z8C5XJ	Thurston	Olympia	Washington	98501	2018	BMW	I3	...	97	0	22	185498386.0	PUGET SOUND ENERGY INC	53067010700	-122.896920	47.043535	PHEV	PHEV

5 rows x 21 columns

The table provides a detailed overview, enabling a comprehensive assessment of the model's performance in capturing real-world electric vehicle adoption patterns.

EV type predicted by the model has higher accuracy.

## 6. Discussions

### 6.1 From Research Questions

1. The annual adoption rate of the Electric vehicle in Washington state has closely mirrored exponential growth, with a minor dip observed in 2019. It's important to note that 2023 data is still ongoing, preventing a conclusive analysis and any generalized statements about the current year.

2. BEV emerges as the more popular electric vehicle, capturing a significant 77.6%. Within the electric vehicle landscape, Tesla takes the lead, closely followed by Nissan, making them the most renowned electric vehicle manufacturers.

3. The elevated Base Manufacturer's Suggested Retail Price (MSRP) for Porsche, Fisker, and Cadillac highlights their premium positioning in the market. Additionally, the positive correlation between higher electric range and increased MSRP emphasizes the significance of pricing about the extended electric range, indicating a potential premium associated with enhanced electric capabilities.

4. 46.3% of vehicles are eligible for Clean Alternative Fuel Vehicle (CAFV) highlights a considerable portion adhering to eco-friendly standards. Interestingly, Clean Alternative Fuel Vehicle (CAFV) eligibility is more common among vehicles with a longer electric range compared to those that are not eligible. This emphasizes the significance of considering CAFV eligibility, particularly when prioritizing vehicles with extended electric capabilities. The higher prevalence of CAFV eligibility in such vehicles suggests a collective commitment towards adopting cleaner and more sustainable fuel options in the electric vehicle domain.

5. Puget Sound Energy Inc. emerges as the most utilized utility, indicating its prominence in facilitating the charging infrastructure for electric vehicles.

## 6.2 From EDA, Modeling

- The predominant electric vehicle type is BEV (Battery Electric Vehicle), constituting 77.6% of the dataset.
- King County leads in electric vehicle numbers with 79,075, followed by Snohomish with 17,307.
- Washington state emerges as the leading region with the highest electric vehicle count.
- Tesla dominates the market for electric vehicles in Washington.
- Puget Sound Energy electric utilities are the most commonly associated with electric vehicles.
- Electric vehicles exhibit a high likelihood of meeting clean fuel (low emission) eligibility requirements.
- The Random Forest model attains exceptional performance, boasting accuracy and ROC-AUC scores of 0.999745 and 0.999977, respectively.
- Overall, predictions demonstrate a high level of accuracy, reinforcing the model's reliability.

## 7. Conclusions

The rise of electric vehicles (EVs) has brought about a transformative shift in the automotive sector, with Washington state leading the charge towards sustainability. Our comprehensive analysis aimed to uncover patterns and factors influencing EV adoption, offering a valuable roadmap for stakeholders in promoting sustainable transportation practices. By addressing key questions on adoption rates, preferences, and economic considerations, the research contributes to

environmental sustainability and informed decision-making. Notably, Battery Electric Vehicles (BEVs) constitute a significant 77.6% of the EV landscape, and King County leads in EV numbers with 79,075, closely followed by Snohomish with 17,307. Tesla's dominance in EV production in Washington, along with the common association of Puget Sound Energy electric utilities with clean fuel eligible EVs, underscores the state's commitment to innovation and eco-friendly mobility. Additionally, electric vehicles exhibit a high likelihood of meeting clean fuel (low emission) eligibility requirements. The Random Forest model attains exceptional performance, boasting accuracy and ROC-AUC scores of 0.999745 and 0.999977, respectively. Overall, predictions demonstrate a high level of accuracy, reinforcing the model's reliability in interpreting and predicting electric vehicle adoption patterns. This research provides insights and metrics that not only capture the current EV scenario but also lay the groundwork for strategic interventions, fostering a vision of sustainable and electrified transportation in Washington and beyond.

## **8. Recommendations**

### **1. Expand Charging Infrastructure:**

- Develop a strategic plan for charging station placement, focusing not only on high-density urban areas but also on highways and rural locations to facilitate long-distance travel.
- Implement fast-charging stations along major transportation corridors to cater to drivers on longer journeys.
- Encourage businesses and workplaces to install charging stations, offering incentives or subsidies for such initiatives.

### **2. Incentivize BEV Adoption:**

- Introduce a tiered incentive program based on the EV's battery range and efficiency to promote the adoption of more environmentally friendly models.
- Partner with local dealerships to offer discounts or special financing options for BEV purchases.
- Develop a trade-in program that offers additional incentives for consumers who switch from a gasoline-powered vehicle to a BEV.

### **3. Collaborate with Utility Companies:**

- Work with utilities to develop off-peak charging programs that offer reduced rates during certain hours, encouraging EV owners to charge their vehicles at times that are beneficial for the power grid.
- Support initiatives for the integration of smart grid technologies, allowing for more efficient energy distribution and usage.

- Promote the installation of solar panels and home energy storage systems for EV owners, creating a more sustainable and self-sufficient ecosystem.

#### **4. Educational Campaigns:**

- Organize community events and EV showcases to provide hands-on experiences and direct interaction with various EV models.
- Collaborate with schools and universities to include EV technology and sustainability topics in their curricula.
- Leverage social media and digital platforms to reach a broader audience, particularly targeting younger demographics.

#### **5. Support EV Manufacturers:**

- Offer tax breaks or subsidies for EV manufacturers to set up research and development centers in the state, fostering innovation in EV technology.
- Encourage public-private partnerships for the development of new EV technologies and battery recycling methods.
- Create a business-friendly environment that attracts startups and established companies in the EV sector, promoting job creation and economic growth.