Decoding EV Adoption

Authors:

Amit Singh, amitsingh@iisc.ac.in

Anant M Nambiar, anantnambiar@iisc.ac.in

Shweta R Joshi, shwetajoshi@iisc.ac.in

Shubham Yadav, shubhamy@iisc.ac.in

Problem Statement:

As the global focus shifts toward sustainability, the adoption of Electric Vehicles (EVs) is gaining momentum. However, effective planning for infrastructure, policy, and market strategies requires a deep understanding of EV adoption trends and pricing dynamics. This project aims to analyze historical EV data to uncover patterns in adoption, forecast future EV sales, and predict pricing based on key features such as electric range, vehicle type, and location. Leveraging advanced data science techniques, this study seeks to generate actionable insights that can guide stakeholders in promoting sustainable transportation solutions while addressing consumer needs and market demands.

Objectives:

- Understanding and forecasting EV Sales
- Analyze the past EV data to understand the trend in EV adoption
- Predict price details based on input features (year, range, make and model, EV type, Clean Alternative Fuel Vehicle eligibility)

Data Description:

The data is taken from data.gov . Dataset is provided by the state of Washington. This dataset shows the Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) that are currently registered through Washington State Department of Licensing (DOL). There are ~2.1L rows spanning models produced since 1999-2024.

It has the following fields:

- VIN (1-10): Vehicle Identification Number (categorical)
- County: Registration county (categorical)
- City: Registration city (categorical)
- State: Registration state (categorical)
- Postal Code: Registration postal code (numerical - float)
- Model Year: Vehicle manufacture year (numerical - int)

- Make: Vehicle manufacturer/brand (categorical)
- Model: Specific vehicle model (categorical)
- EVType: Electric vehicle classification (categorical)
- CAFV: Clean Alternative Fuel Vehicle eligibility (categorical)
- Electric Range: Maximum electric-only travel distance in miles (numerical - float)
- Base MSRP: Manufacturer's Suggested Retail Price (numerical - float)
- Legislative District: Registration legislative district (numerical - float)
- DOL Vehicle ID: Department of Licensing identifier (numerical - int)
- Vehicle Location: Specific registration location details (categorical)
- Electric Utility: Power provider for the vehicle's location (categorical)
- 2020 Census Tract: Census geographic area identifier (numerical - float)

Data cleaning and preprocessing:

Among these, some of the fields have considerable null values. Details like range and price were missing on some of the data points. Using details like model, make, and model year, we fetched the accurate price and range details from other API sources and imputed it into our dataset.

“Clean Alternative Fuel Vehicle (CAFV) Eligibility” was simplified into a True/False value for easier analysis.

“DOL Vehicle ID”, “2020 Census Tract” columns were dropped since they provide no useful insight to our analysis.

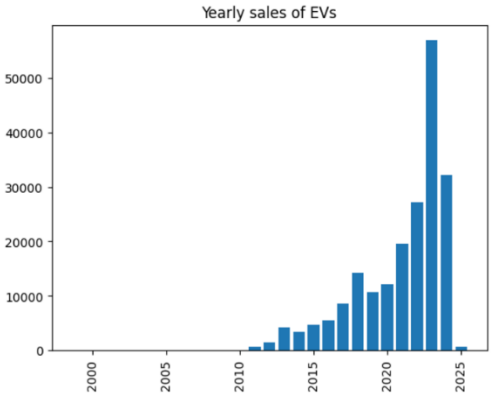
Post EDA, we have also removed outlier data so that the samples are better represented

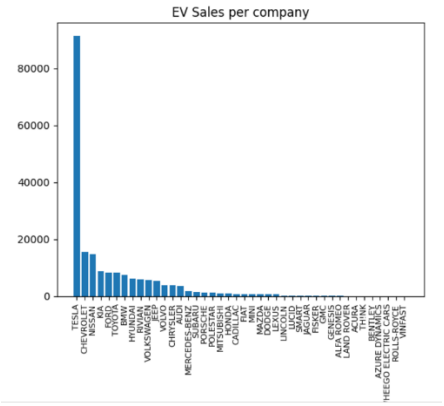
Exploratory Data Analysis:

Sales Insights:

Year sales data shows an exponential increase in the sale of EVs indicating an increasing demand for electric vehicles.

One company, Tesla, accounted for most of the EV sales over the years.





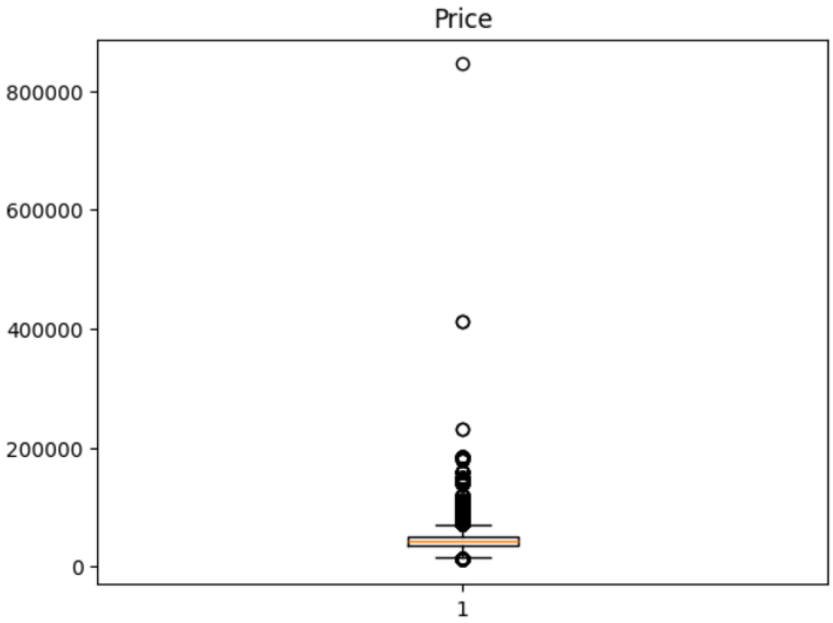
Univariate Analysis:

Numerical features:

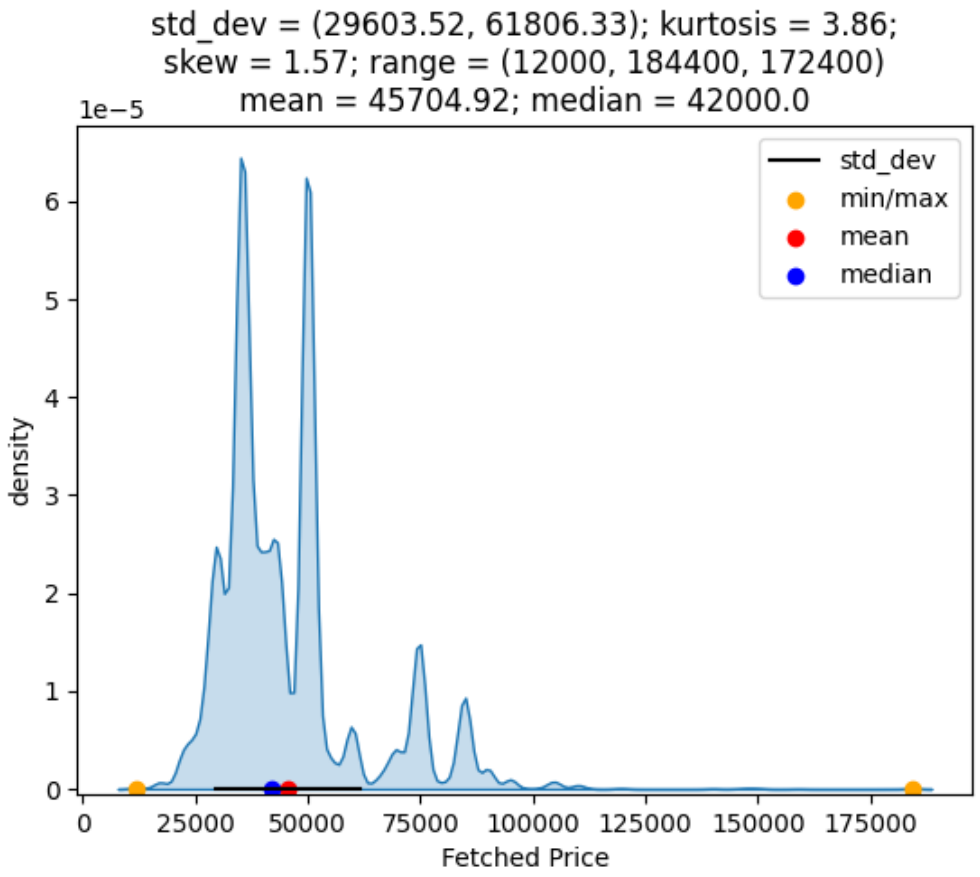
Price:

As seen from the boxplot, we see some outliers in the price distribution.

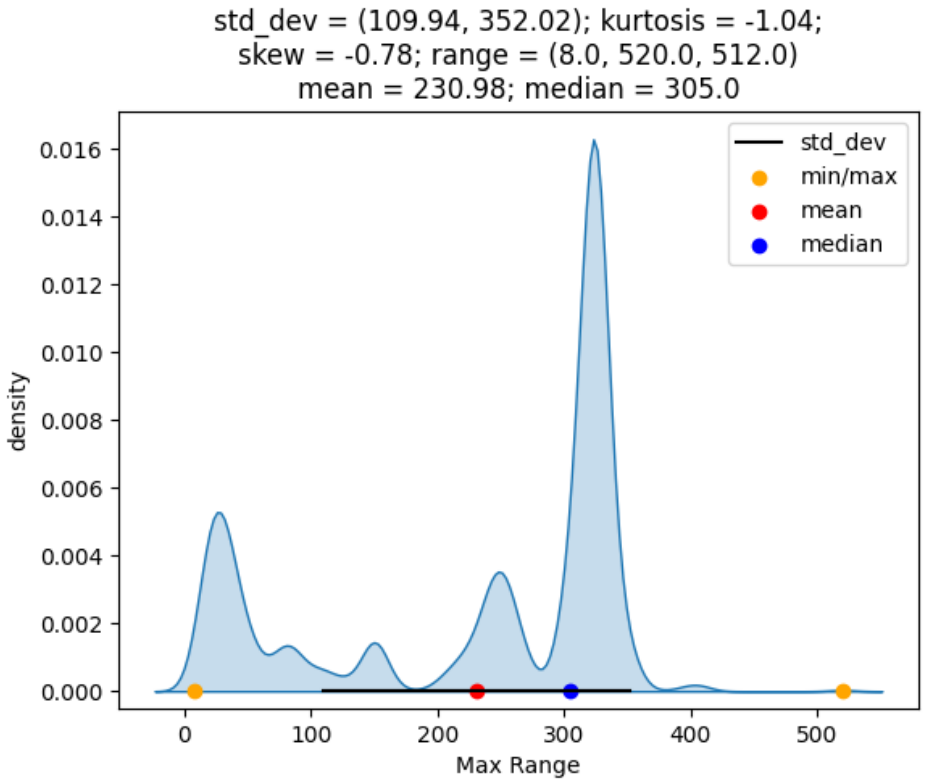
Outlier handling: Consider prices above $200000 as outliers and remove during analysis.



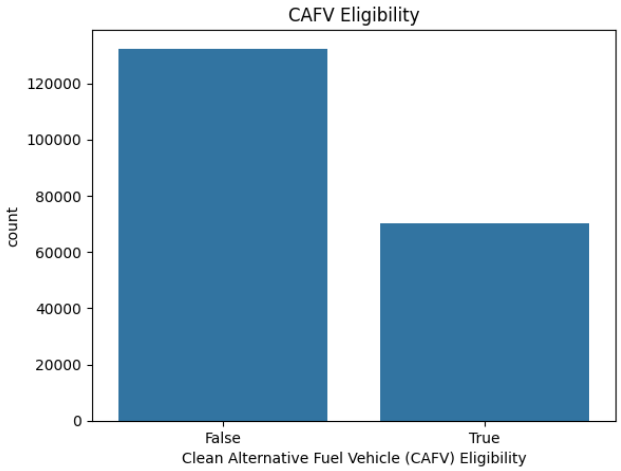
Distribution after outlier handling is as seen below with average price of EVs being around $45,000

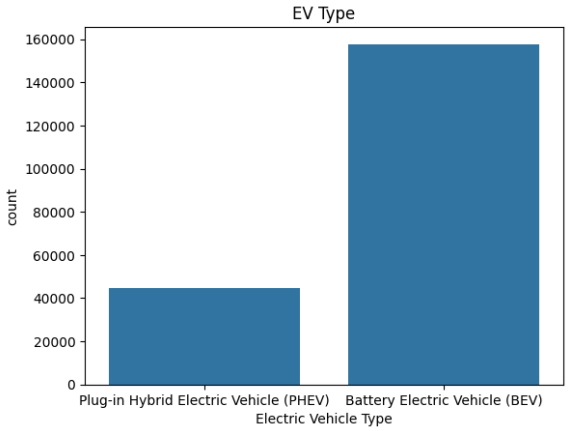


Range: non-uniform distribution with average range of 230 miles



Categorical features:

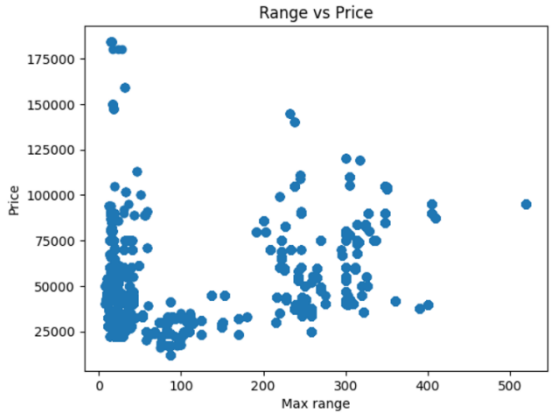


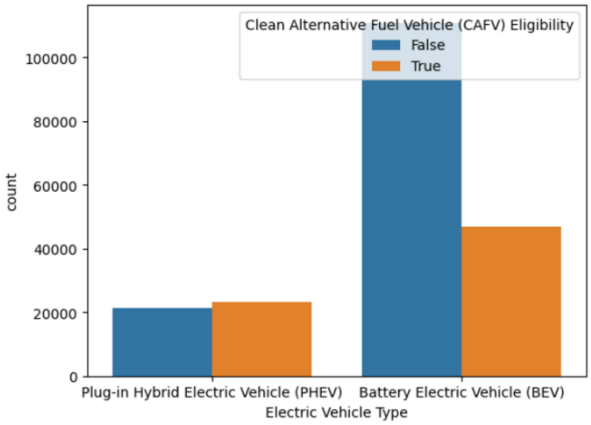


CAFV eligibility: Most people choose to not opt for CAFV eligibility. Around a third of EV buyers choose model with CAFV eligibility.

EV Type: Fully electric EVs are more popular than hybrid kind.

Bivariate Analysis:



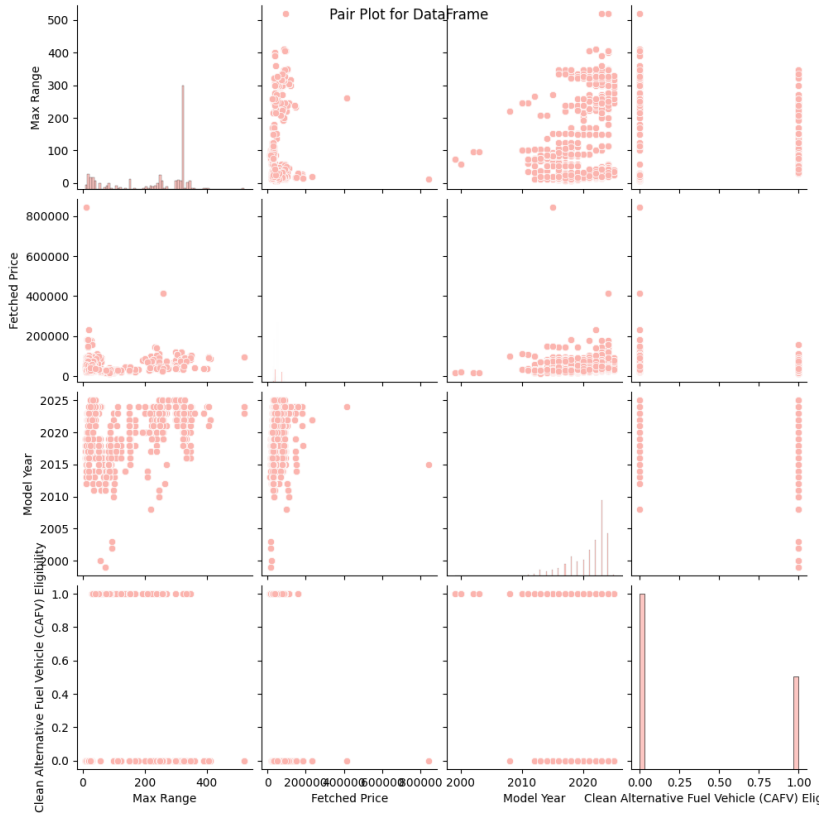


Range and Price do not show a strong correlation. Higher price does not guarantee higher range. Mid-priced and low-priced EVs have good electric range.

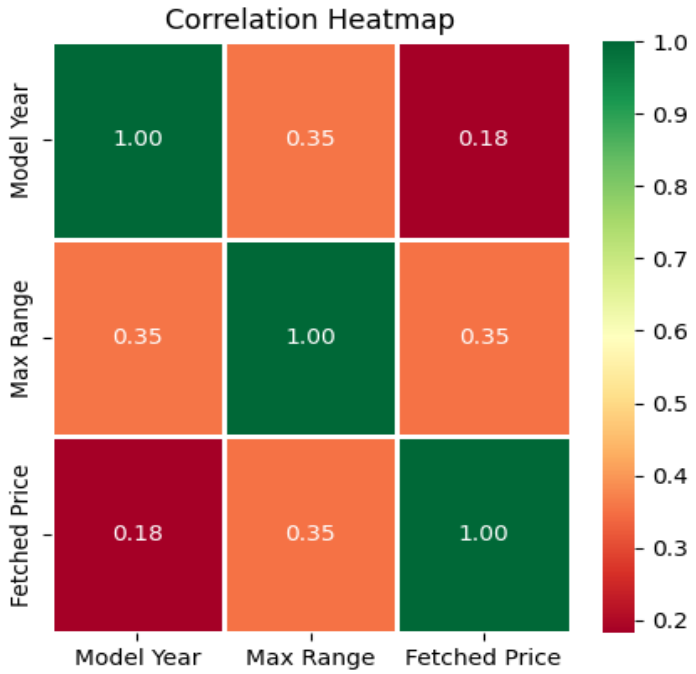
Most of the battery EVs also do not have CAFV eligibility.

Multivariate Analysis:

Pairplot:



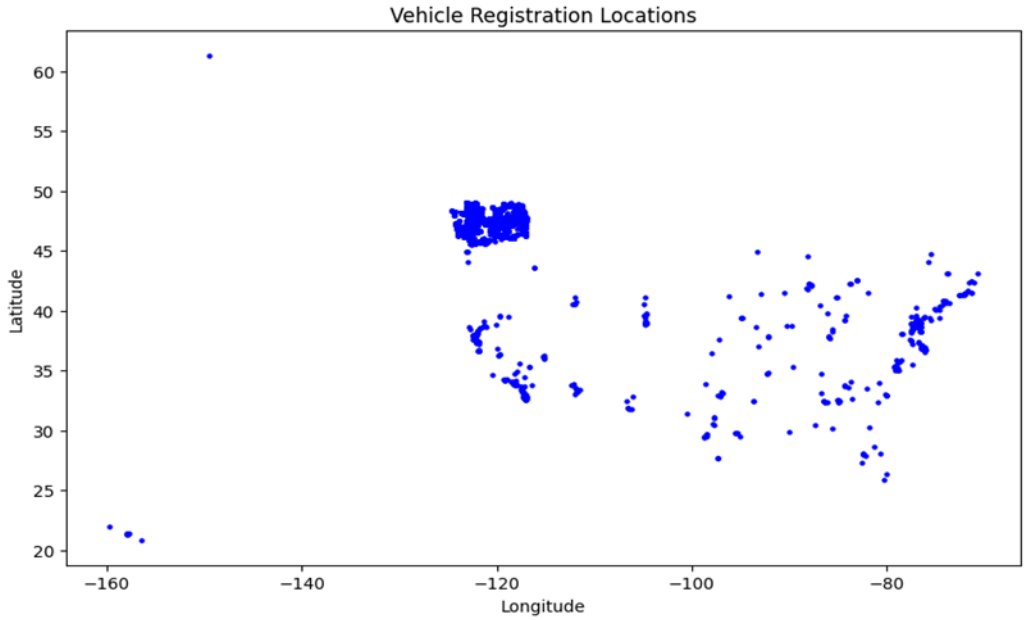
Heatmap:



Low correlation between price, range and sales

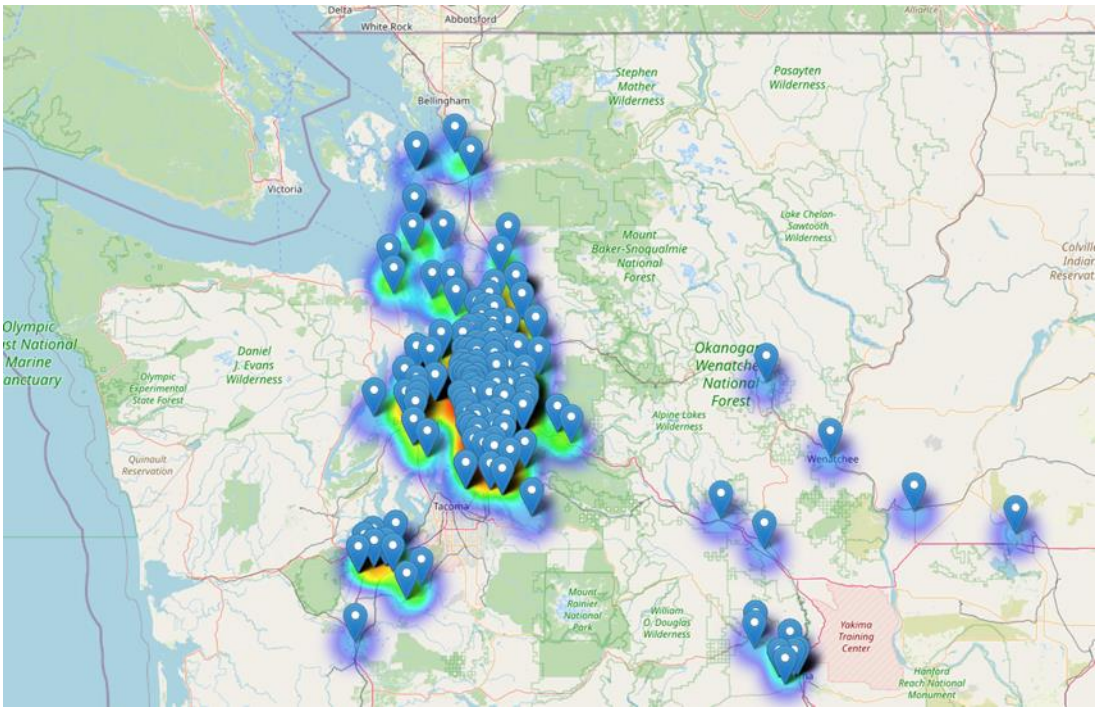
Geospatial Analysis

- Vehicle Registration location (on static map):



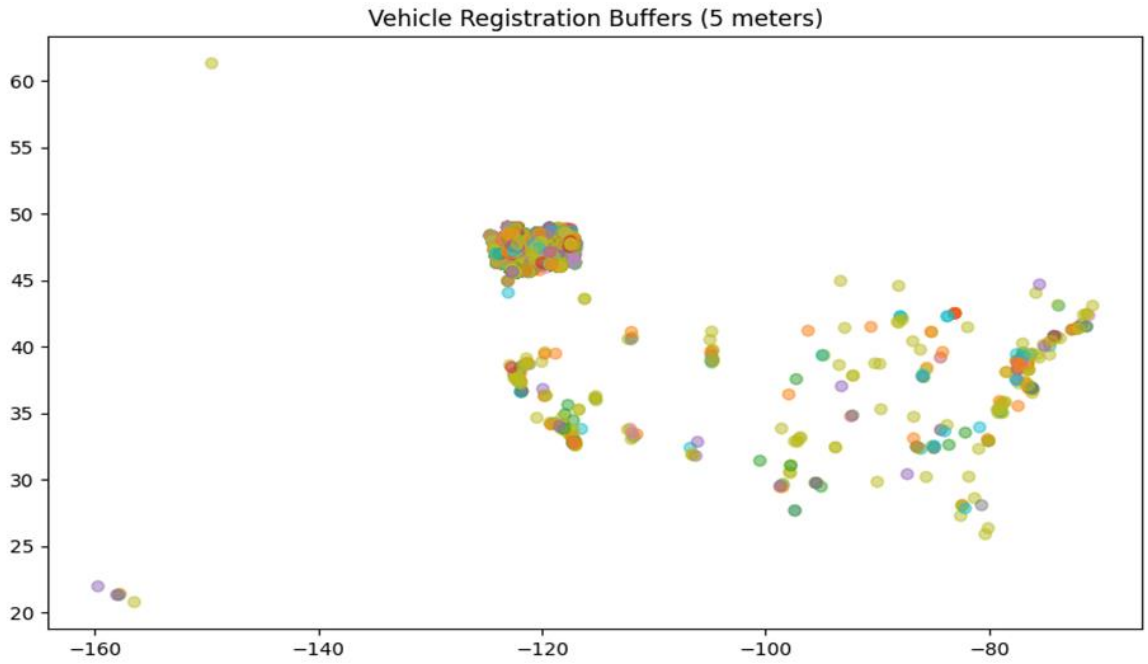
Analysis: - Good number of vehicles are registered in a particular area.

- Vehicle location Heatmap (on interactive map):



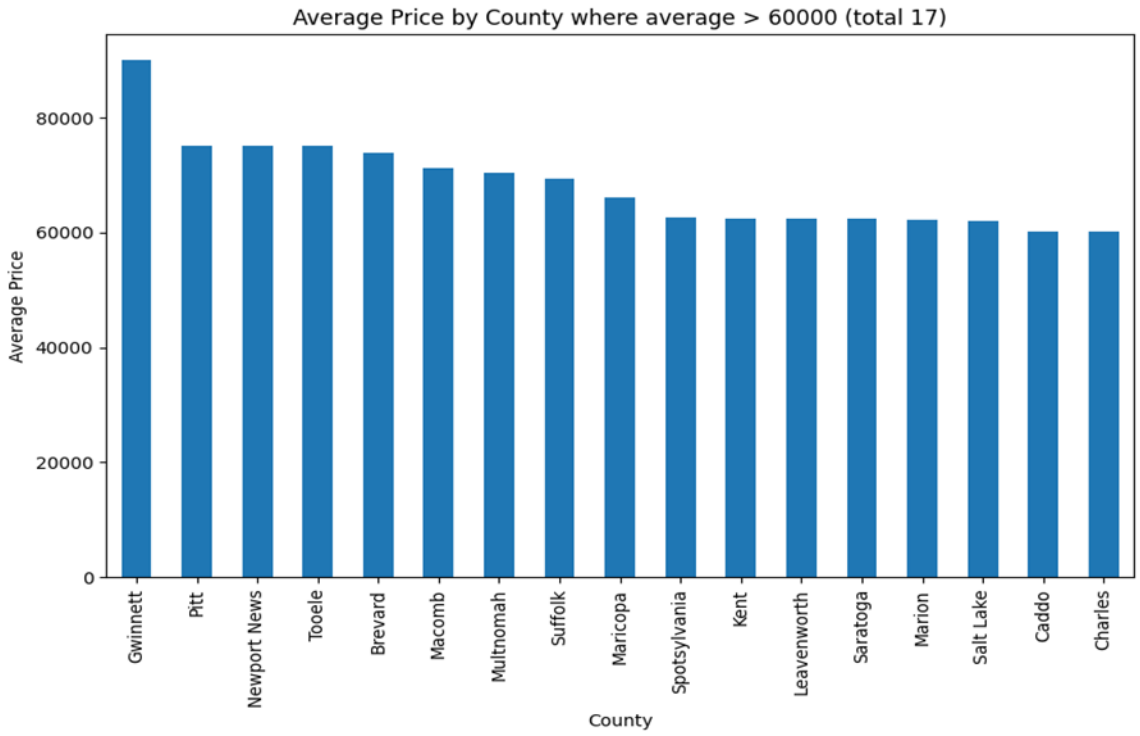
Analysis: - EVs are successfully adopted in few areas/regions.

- Vehicle Registration buffer:

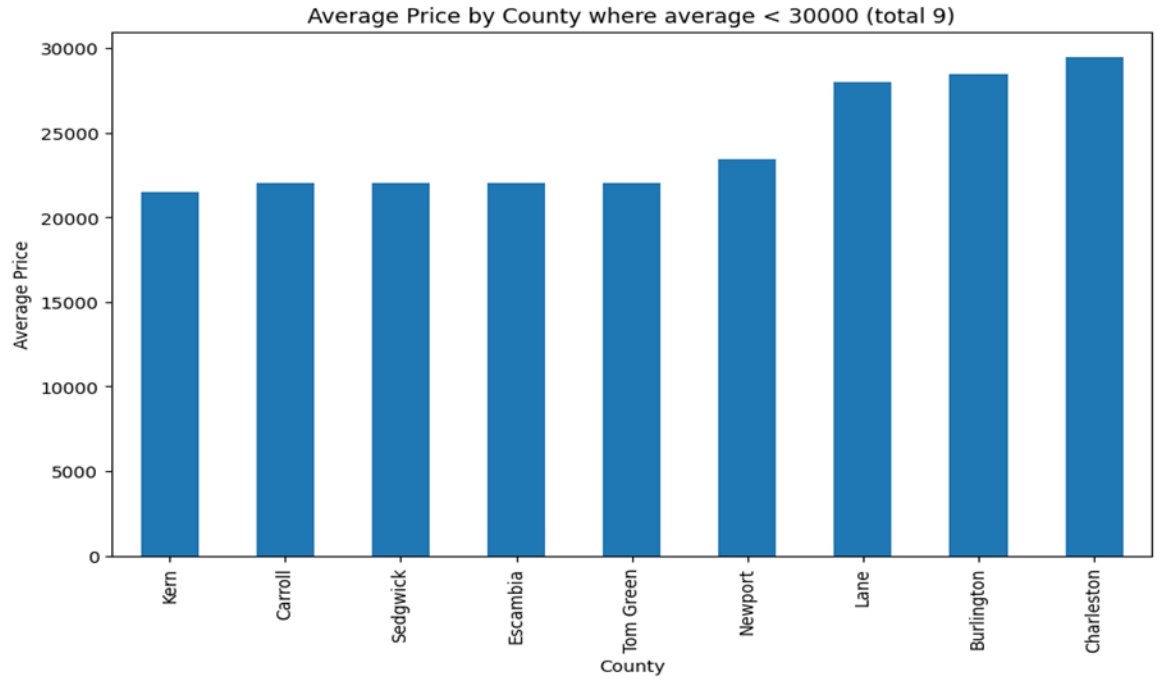


Vehicle location buffer can be used in several applications such as traffic analysis, emission analysis, urban planning, service accessibility etc.

- Counties with high purchase power analysis: -



- Counties with low purchase power analysis: -



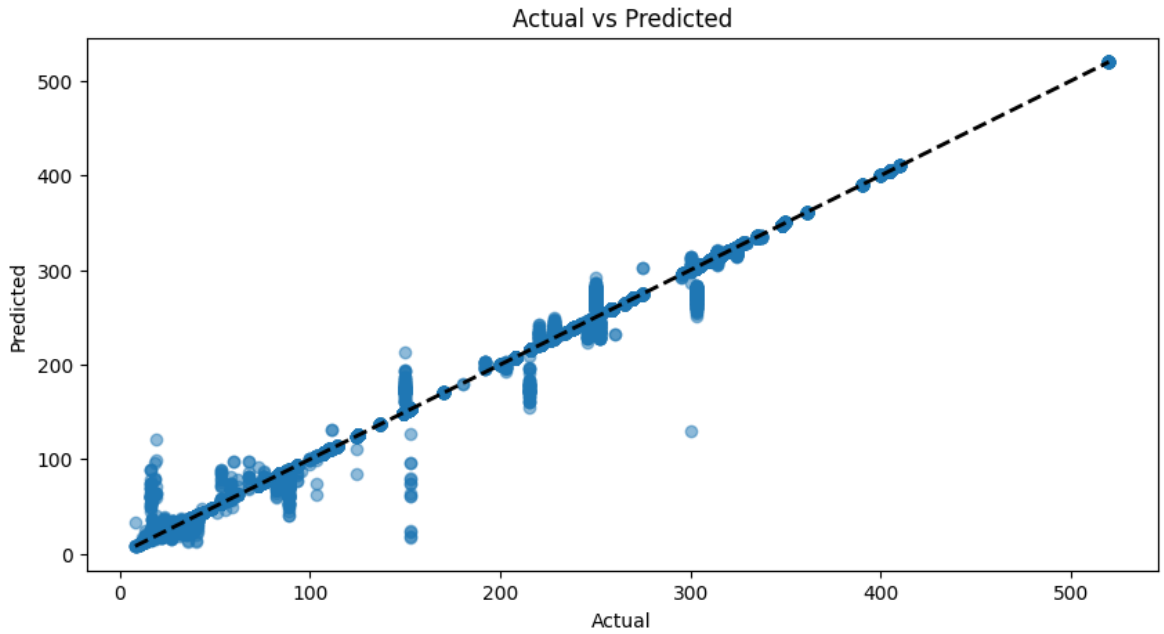
Methodology

- Random Forest Regression (Range Prediction):
  The Random Forest Regressor is an ensemble machine learning algorithm primarily used for regression tasks. It is based on the Random Forest method, which builds multiple decision trees and aggregates their results for prediction.
- Gradient Boosting Regression (Price Prediction):
  Gradient Boosting Regression is a powerful ensemble learning technique used for regression tasks. It builds a predictive model by combining the strengths of multiple "weak learners," typically decision trees, to improve accuracy and robustness.

Analysis and Results

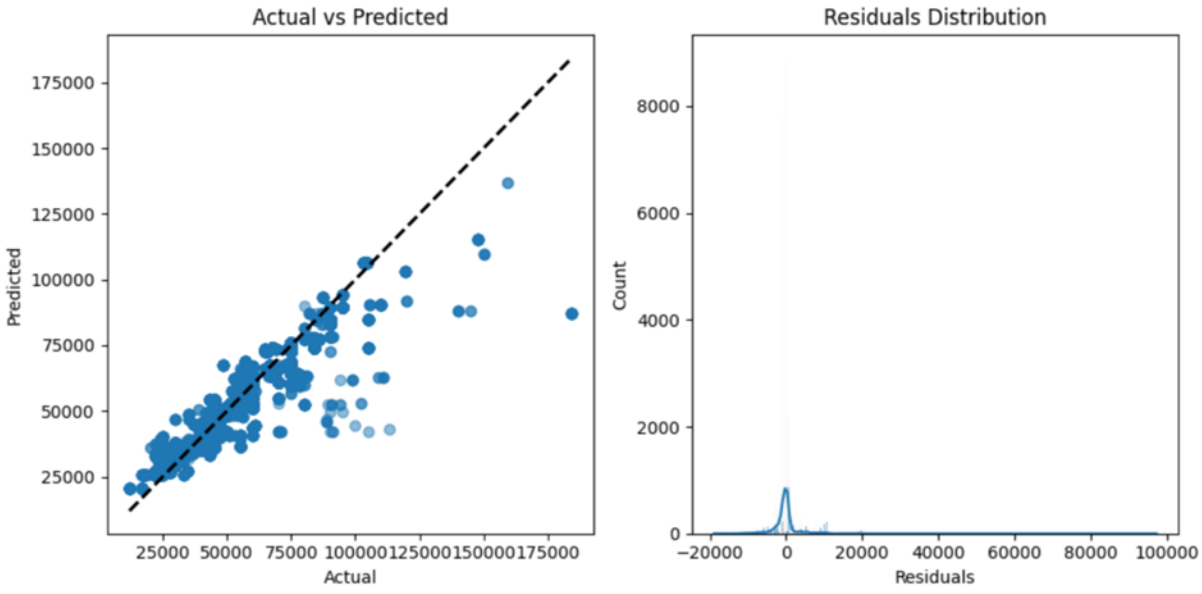
- Range Prediction:

```
Evaluation Matrix
Mean Absolute Error: 1.187043
Mean Squared Error: 30.738393
Root Mean Squared Error: 5.544222
R-squared Error: 0.947912
MAPE: 6.626591%
```



- Price Prediction:

```
features = [
    "County",
    "City",
    "State",
    "Model Year",
    "Make",
    "Model",
    "Electric Vehicle Type",
    "Clean Alternative Fuel Vehicle (CAFV) Eligibility",
    "Max Range",
]
target = "Fetched Price"
```




```
Mean Squared Error: 17370466.769106243
R^2 Score: 0.9331644118865465
```

References

- Dataset - Department of Licensing, US Govt - https://catalog.data.gov/dataset/electric-vehicle-population-data
- Edmunds API - https://developer.edmunds.com/index.html