IS 680 (72318)

Data Science I: Introduction to Data Science and Visualization

Final Report

Electric Vehicle Population Data

December 10, 2024

Group Members:

Abhijeet Hariyani
Bhavik Chopra
Chitra Dusane
Prachi Holkar
Shreeya Desai
Yogesh Gemani

- ## Introduction

1. ### How does your project help those who are interested in this topic?

Recent years witnessed a surge in electric vehicle (EV) popularity, attributed to battery technology advancements and advanced data analytics. This fusion has elevated EV market share in the automotive industry. The harmonious integration of electronic sensors, actuators, controls, software, and robust electrical propulsion systems enables secure data transfer and analysis in modern EVs. This integration not only facilitates data-driven advancements but also offers a comprehensive solution for transportation challenges.

2. ### Why is data analysis needed for this?

Since the 1990s, electric vehicles (EVs) have grappled with challenges like limited range and cold weather issues, with a range of around 50 miles. In contrast, Internal Combustion Engine (ICE) cars gained popularity in the 20th century due to fewer constraints. Governments incentivize EV use through laws and initiatives. Analyzing the correlation between policy changes and EV registrations aids in evaluating their impact. This data-driven approach supports future decision-making on policies, infrastructure, and charging stations, guiding growth planning. As EV technology improved, addressing limitations, it emerged as a safer and more eco-friendly choice.

3. ### How do you expect your analysis to improve decision-making in this area?

Our research aims to leverage technological advancements to support electric vehicle (EV) enthusiasts. Through comprehensive data analysis, we seek to enhance our understanding of EV economy, battery performance, and user behavior. This study is vital for refining EV design, overcoming early challenges, and predicting future trends. Our initiative serves as a valuable resource for manufacturers, governments, and consumers, empowering them to make informed decisions based on empirical evidence. Manufacturers can improve cars, governments can shape infrastructure and regulations, and consumers can make informed choices. Our analysis aims to provide information and enhance decision-making in the EV market, promoting the adoption of greener and more efficient transportation alternatives.

- ## Data

What is the source we used?

Source: We have fetched Electric Vehicle Data from Electric Vehicle Population Data

The dataset contains electric car data, and in this project, we are targeting to perform data analysis on this data. The attributes of every completely electric passenger car are listed in this dataset. The dataset contains data on two types of electric vehicles i.e., Battery Electric Vehicles (BEV) and Plug-in Hybrid Electric Vehicles (PHEV) which are registered through the Washington State Department of Licensing (DOL).

All the variables in the dataset are mentioned below:

- VIN, Country, City, State, Zip code, Model year, Make, Model, Electric Vehicle Type, Electric Range, dol Vehicle ID, Vehicle ID, Electric Utility

EXPLORING THE DATASET

Importing the dataset of electric vehicle population data. After importing the dataset it has **159,467 rows and 17 columns.**

Findings:

- Total rows: 159467
- Most column data types are as expected; categorical variables are 'chr,' and numerical ones like 'Postal Code' are 'dbl.' Considering converting 'Postal Code' to 'chr' for non-mathematical use. 'Model Year' in 'dbl' may be more suitable as 'int.' 'Legislative District' in 'dbl' could be converted to 'int' if they are whole numbers. Use a function to count missing values in 'evp' columns: 'County,' 'City,' 'Electric Utility,' 'Postal Code' (4 each), 'Vehicle Location' (9), and 'Legislative District' (361). Count unique values in each column of the dataset.

DATA CLEANING PROCESS:

We have started by updating the names of the columns, removing unwanted fields, removing all the blank values, and converting the data type as required.

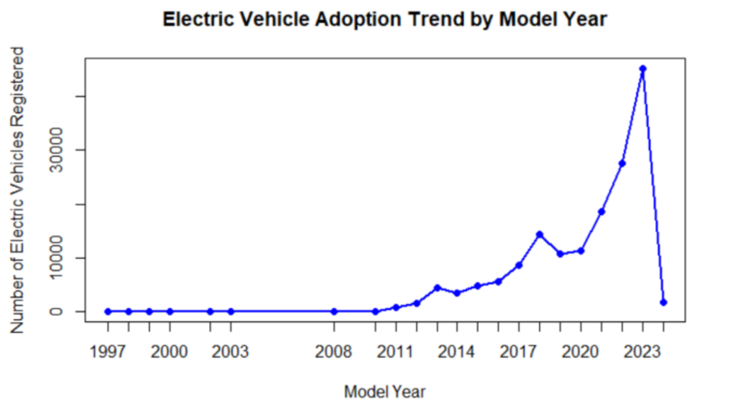
- **Exploratory Data Analysis**

 # IMP:

While analyzing the data, we found there are no explicit columns for seasonal variations, government incentives, or charging infrastructure availability, which leaves us no way to find answers to the questions stated in the proposal. We would introduce 3 new research questions instead.

Do seasonal variations impact EV adoption?
What role do government incentives and policies play in EV adoption?
How does the availability of charging infrastructure affect EV adoption?

1. **What are the current trends in electric vehicle adoption in the dataset's region?**



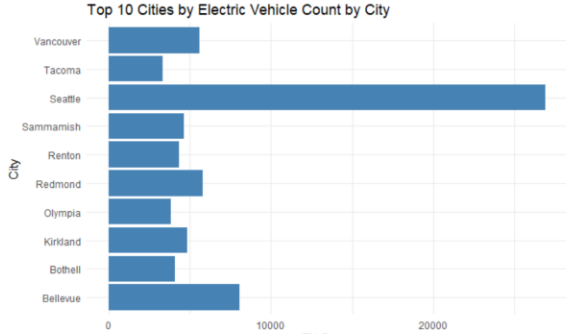
Observation: The trend data and the corresponding line chart show the progression of electric vehicle registrations by model year, indicating a general increase in EV adoption over time, with notable growth starting from 2011 onwards.

2. **Are there demographic patterns that influence EV adoption?**

EV adoption is influenced by several aspects, including income levels, geographic areas, and environmental consciousness. Additional aspects encompass government incentives, accessibility of charging infrastructure, cultural attitudes towards sustainability, car performance, and market availability of various EV models. Each of these factors influences customer behavior and preferences about electric automobiles.

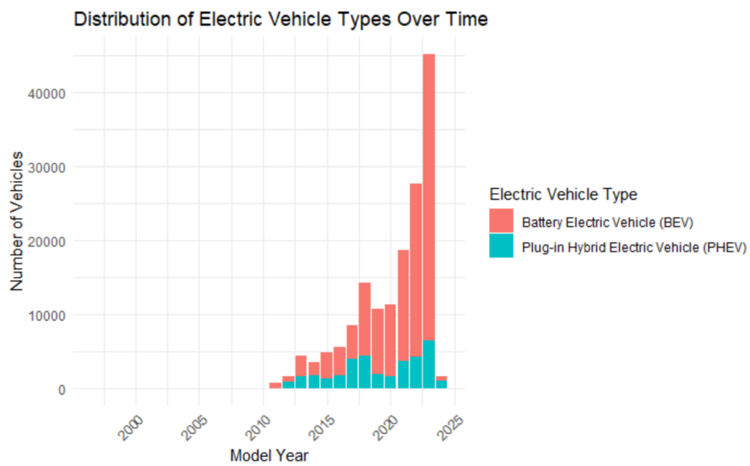
County <chr>	City <chr>	EV Count <int>
King	Seattle	26932
King	Bellevue	8075
King	Redmond	5796
Clark	Vancouver	5599
King	Kirkland	4859
King	Sammamish	4657
King	Renton	4366
Snohomish	Bothell	4106
Thurston	Olympia	3848
Pierce	Tacoma	3333

1-10 of 10 rows



3. What is the distribution of electric vehicle types (e.g., Battery Electric Vehicle, Plug-in Hybrid Electric Vehicle) in the dataset, and how has it evolved over time?

Observation: The chart above displays the distribution of electric vehicle types over time, categorized by model year. Each color represents a different type of electric vehicle, such as a Battery Electric Vehicle (BEV) or Plug-in Hybrid Electric Vehicle (PHEV). Overall, the graph shows that electric vehicles are becoming increasingly popular. This is likely due to several factors, including rising gas prices, concerns about climate change, and government incentives. The number of BEVs has increased by more than 30 times since 2000, while the number of PHEVs has increased by more than 20 times. The number of electric vehicles on the road is expected to continue to grow in the coming years.



4. What is the average electric range of vehicles by make, and how does this vary among the top manufacturers?

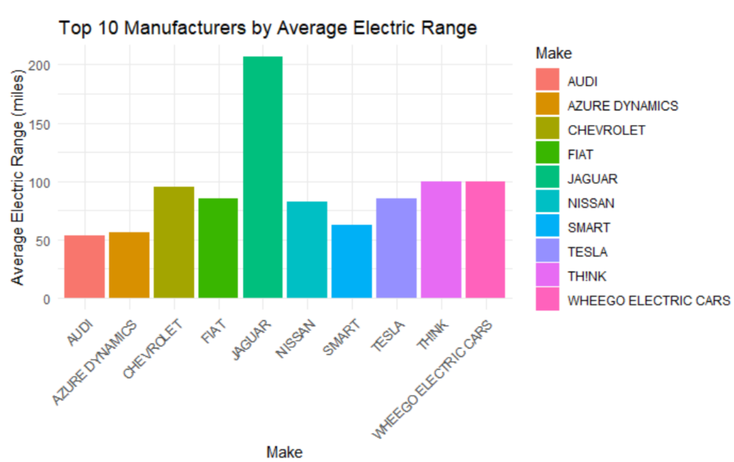
Observation: The bar chart displays the top 10 manufacturers by average electric range. Each bar represents the average electric range of vehicles produced by that manufacturer.

Jaguar leads with the highest average electric range, slightly above 200 miles.

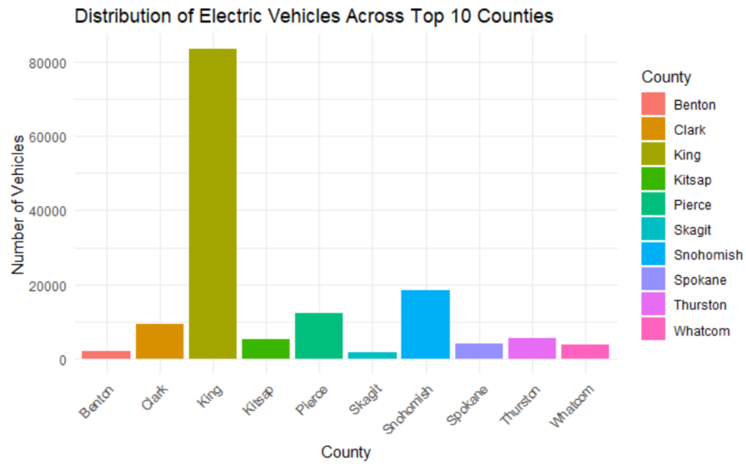
Tesla follows as the second highest, with a range between 150 and 175 miles.

Audi and Azure Dynamics have the lowest average ranges, under 50 miles.

Other manufacturers like Chevrolet, Fiat, Nissan, Smart, Think, and Wheego Electric Cars have ranges mostly between 75 to 125 miles.



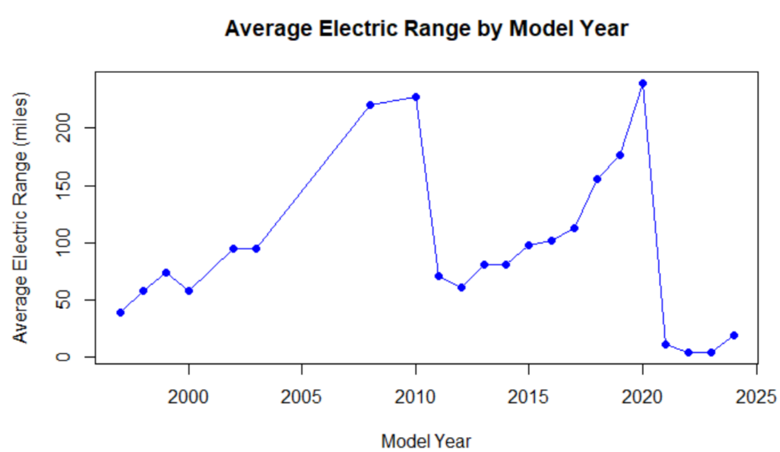
5. How does the distribution of electric vehicles vary across different counties and cities within the dataset's region?



Observation: The chart above illustrates the top 10 counties with the highest number of electric vehicles. Each bar represents the number of electric vehicles registered in that county. This chart provides a clear visual representation of the disparity in electric vehicle distribution among these counties, with one county leading by a substantial margin.

The key takeaway is that King County has by far the most electric vehicles, with almost 80,000. Snohomish County has the second most with around 20,000 vehicles. The distribution is highly skewed towards King County.

6. Is there a relationship between "range anxiety" and EV adoption?



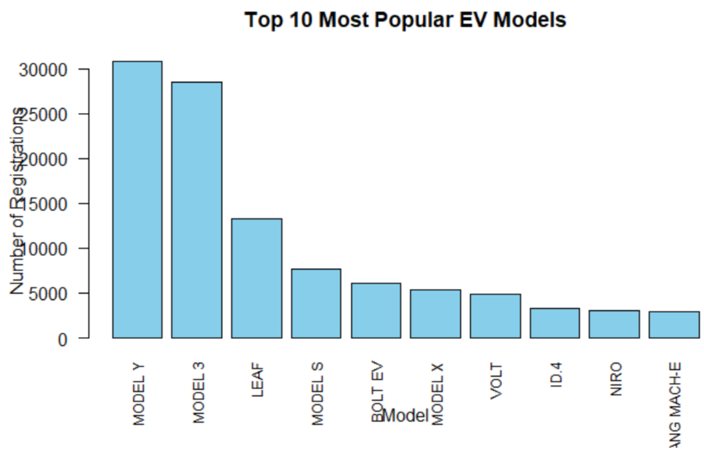
The examination of the electric car dataset uncovers a compelling pattern in the electric range over time, which might be linked to the notion of range anxiety.

- During the first period (1997-2003), the typical electric cars had a rather limited electric range, usually below 100 miles.
- There has been a substantial rise in the average distance an electric vehicle can go on a single charge starting from 2010, with a particularly noticeable leap occurring around 2018.
- The years 2019 and 2020 have the greatest average ranges, suggesting that newer models possess much-improved ranges, which might potentially alleviate concerns about limited range.

The observed pattern indicates that over time, electric vehicles are being engineered with increased driving distances, presumably as a reaction to consumer apprehensions over limited range. Nevertheless, the data from 2021-2024 indicates a decline in the average range. This decline warrants more examination to comprehend the underlying factors, such as introducing new models with varying range capabilities or potential data abnormalities.

The information provides indirect evidence indicating that manufacturers are placing a growing emphasis on enhancing the electric range of vehicles. This is a critical aspect in the adoption of electric vehicles and in addressing concerns about limited driving distance.

7. Are certain EV models preferred over others by consumers?



Observation: The top 10 most popular EV models by number of registrations have been identified. "MODEL Y" and "MODEL 3" are the most popular EV models based on the number of registrations, with a significant lead over the other models listed. There is a clear drop in popularity after the top three models, with the remaining seven models having relatively similar but lower numbers of registrations.

- ## Predictive Analysis

The execution of the k-nearest Neighbors (k-NN) order model holds huge ramifications for foreseeing electric vehicle models, offering different applications inside the business. This predictive capability is especially important for assignments going from **inventory management** to **demand forecasting** and **personalized recommendations**. In the space of inventory management, the model empowers key inventory management by precisely expecting interest in explicit electric vehicle models. This improvement not just guarantees a sufficient stock of well-known models yet in addition limits the gamble of overloading less-requested ones, prompting cost reserve funds and uplifted functional proficiency across the production network. Besides, the k-NN model's expectations assume a crucial part of popular estimating, utilizing verifiable information on electric vehicle inclinations to project future patterns.

Makers can powerfully change creation plans, adjust production schedules, and proactively answer developing business sector elements, acquiring a competitive edge in a quickly developing industry. The model's usefulness also extends to personalized recommendations for potential buyers of electric vehicles. It makes use of features like county, vehicle type, electric range, and car make to tailor recommendations based on individual preferences and geographic considerations.

This capacity improves the general client experience, encourages satisfaction, and holds the possibility to help deals through designated and applicable recommendations.

```
 Electric.Range     Base.MSRP      Legislative.District DOL.Vehicle.ID
Vehicle.Location   Electric.Utility
 Min.   :  0.00   Min.   :      0   Min.   : 1.00       Min.   :     4385
Length:159467      Length:159467
 1st Qu.:  0.00   1st Qu.:      0   1st Qu.:18.00       1st Qu.:173101620   Class
:character   Class :character
 Median : 14.00   Median :      0   Median :33.00       Median :219844964   Mode
:character   Mode  :character
 Mean   : 64.28   Mean   :   1228   Mean   :29.26       Mean   :214024243
 3rd Qu.: 84.00   3rd Qu.:      0   3rd Qu.:43.00       3rd Qu.:244836346
 Max.   :337.00   Max.   :845000   Max.   :49.00       Max.   :479254772
                                    NA's   :361
 X2020.Census.Tract
 Min.   :1.081e+09
 1st Qu.:5.303e+10
 Median :5.303e+10
 Mean   :5.297e+10
 3rd Qu.:5.305e+10
 Max.   :5.603e+10
 NA's   :4
[1] "County"                "Electric.Vehicle.Type" "Electric.Range"        "Make"
"Model"
[1] 0
[1] 0
Optimal k: 29
Accuracy: 77.95%
```

Notably, **this code's implemented k-NN model is 77.95% accurate**. This quantitative measure provides valuable insights into the model's performance, guiding further refinements and adjustments for enhanced predictive capabilities. All in all, the k-NN calculation, esteemed for its straightforwardness and interpretability, fills in as a flexible device in the electric vehicle industry, giving significant experiences to guide decision-making, upgrade competitiveness, and satisfy developing business sector needs.

- **Conclusion**

The conducted Exploratory Data Analysis (EDA) on the dataset reveals a complex relationship of factors influencing the adoption of electric vehicles (EVs) in the region. The investigation, utilizing sophisticated statistical techniques and visualizations, illustrates the substantial impact of demographic trends, geographical disparities, and vehicle specifications on the patterns of electric vehicle (EV) adoption. The results highlight the crucial importance of government incentives and infrastructure development, namely charging networks, in influencing customer choices and rates of adoption. This comprehensive report offers essential insights for stakeholders in the electric vehicle (EV) industry, emphasizing key areas that require strategic attention to expedite the shift toward sustainable transportation.

References Source: We have fetched the Electric Vehicle Data from https://catalog.data.gov/dataset/electric-vehicle-population-data