INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

EXPLORATORY DATA ANALYSIS ON ELECTRIC VECHICLE

Name : AHMED
LinkedIn:https://www.linkedin.com/in/ahmedbaymeen/
Github:https://github.com/Ahmed900786

Objective

Uncovering Patterns and Trends in the EV Market

- Worked with a large dataset comprising 112,635 rows and 17 features including crucial variables such as VIN, Electric Vehicle Type, Make, Model Year, Electric Range, and Geographic Location (County, City, State).

-Utilized advanced visualization techniques to create engaging, interactive visual reports that allow users to explore the data trends and correlations intuitively, enhancing the data storytelling process.

-Leveraged PlotlyExpress to create an interactive Choropleth map visualizing the density of electric vehicles across geographic regions, mapping EV adoption trends at a granular level using census tract data and vehicle location coordinates.

-Developed a dynamic Racing Bar Plot using the bar-chart-race library to visualize the growth and evolution of EV manufacturers over time.

Summary of the Data

Vehicle Identification and Location

•VIN (1-10):This is the Vehicle Identification Number, a unique code assigned to each vehicle.

•County, City, State: These columns specify the geographic location of the vehicle.

•Postal Code: The ZIP code corresponding to the vehicle's location.

Vehicle Details

•Model Year: The year the vehicle was manufactured.

•M ake :The brand or manufacturer of the vehicle.

•Model:The specific model or series of the vehicle.

•Electric Vehicle Type:The type of electric vehicle (e.g., battery electric, plug-in hybrid).

•Clean Alternative Fuel Vehicle (CAFV) Eligibility:Indicates whether the vehicle qualifies for incentives or benefits as a clean alternative fuel ve hic le.

•Electric Range:The estimated distance the vehicle can travel on a single charge.

•Base MSRP:The manufacturer's suggested retail price for the vehicle.

Government and Utility Information

•Legislative District: The legislative district in which the vehicle is registered.

•DOL Vehicle ID: A unique identifier assigned by the Department of Labor.

•Vehicle Location: The specific location or address of the vehicle.

•Electric Utility: The electric utility company serving the vehicle's location.

•2020 Census Tract: The census tract in which the vehicle is located.

Data Types

The image also specifies the data types for each column:

•object:This typically represents text or categorical data, such as VIN, Make, Model, Electric Vehicle Type, County, City, State, Vehicle Location, and Electric Utility.
•int64:This indicates integer values, like Model Year, Postal Code, Electric Range, Base MSRP, DOL Vehicle ID, and 2020 Census Tract.

•float64:This represents floating-point numbers, such as Legislative District.

Potential Uses of This Dataset

This dataset could be used for various purposes, including:

•Electric Vehicle Research:Analyzing trends in electric vehicle ownership, usage, and location.

•Policy Development: Evaluating the effectiveness of incentives and policies related to electric vehicles.

•Market Analysis: Understanding consumer preferences and market demand for different types of electric vehicles.

•Geographic Analysis: Mapping the distribution of electric vehicles across regions.

•Utility Planning: Assessing the impact of electric vehicle adoption on energy demand and infrastructure.

Exploratory Data Analysis:

Data Cleaning Steps :

- **To converting object to datetime with help of the function**
- **In this data set I got 3 datetime dtypesbut it was in object data types. I converted object to datetime with help of pandas library has function pd.datetime**

CheckedNull Values:

- **If null values are presented in our dataset we have different techniques to treat missing values but if bull values are**
- **present in greater than 70 precent Its better to remove the column**

Duplicated Values:

It is important step to check duplicated values if duplicated values are present in data no matter what its better to

- **remove duplicated values**
- **Outliers :**

Outliers is nothing its far of other values again how it is treat if greater than 70 percent we need to remove less than 70 treat the outliers to know to find outliers by using box plot

-
-

Data Manipulation Steps

Data Cleaning

•Handling Missing Data : Identify and fill missing values or drop rows/columns with excessive missing data.
•Example methods: Imputation (mean, median, mode), dropping missing data, or using algorithms that handle missing data.

•Removing Duplicates : Check for and remove duplicate rows to prevent data redundancy.

•Handling Outliers: Detect and handle outliers that may skew analysis, either by removal or transformation

•I filled Missing Values by using fillna(“ffill”) in columns like **Model, Legislative District, Vehicle Location and Electric Utility**

Univariate Analysis Steps

What is univariate analysis

How the data distributed in single column, we have different techniques to perform this univariant analysis

- Central measure of tendency
- Mean, median, mode
- Mean is talking about a average of a data
- It is applicable for only numerical data

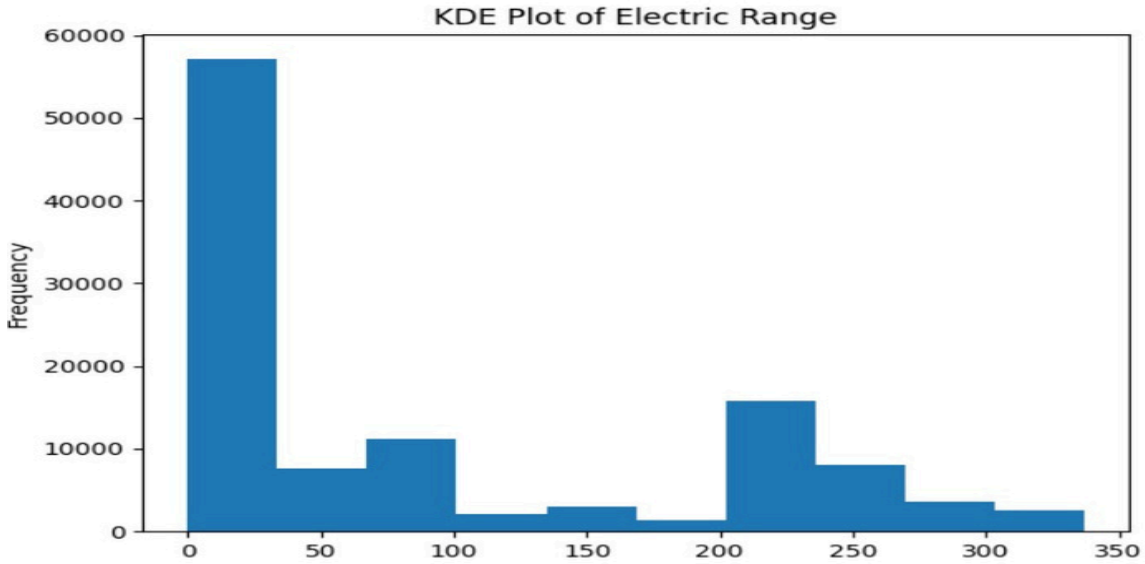
  It is effected by outliers
- Median is talking about a middle value of the data
- It also applicable for numerical data only
- It is not effected by outliers
- Note If mean = median it means there is no outliers in our data, although if mean =/ median there is a outliers in a data
-

Mode is talking about frequency of values are present in data
It is applicable for only categorical data

-

```
********** Base MSRP **********
min            0.000000
max       845000.000000
mean        1793.439681
median         0.000000
std        10783.753486
skew          10.100497
kurt         371.695682
Name: Base MSRP, dtype: float64
```

- The Base MSRP column shows a significant right-skewed distribution with a large number of vehicles priced at $0 (likely indicating errors or incomplete data). The average price is $1793, but the median is $0, suggesting the distribution is heavily influenced by outliers. The standard deviation of $10783 further highlights the wide range of prices.



The graph shows a KDE plot of the electric range of electric vehicles. The x-axis represents the electric range in miles, and the y-axis represents the frequency or density of vehicles at each range.

O bse rvat ions :

Right-skewed distribution:The majority of vehicles have a lower electric range, while a smaller number have a higher range. This indicates that there are more vehicles with shorter ranges compared to longer ranges.

Multimodal:The plot shows multiple peaks, suggesting that there might be distinct groups or clusters of vehicles with different range characteristics. This could be due to factors like battery technology, vehicle size, or intended use.

Concentration around 0-50 miles:There is a significant concentration of vehicles in the range of 0-50 miles. This might indicate the presence of a large number of hybrid or plug-in hybrid vehicles with smaller electric batteries.

Gaps in the distribution:There are noticeable gaps in the distribution, particularly between 100-150 miles and 250-300 miles. This suggests that there might be fewer vehicles available in these range segments.

Uni-Variant categorical

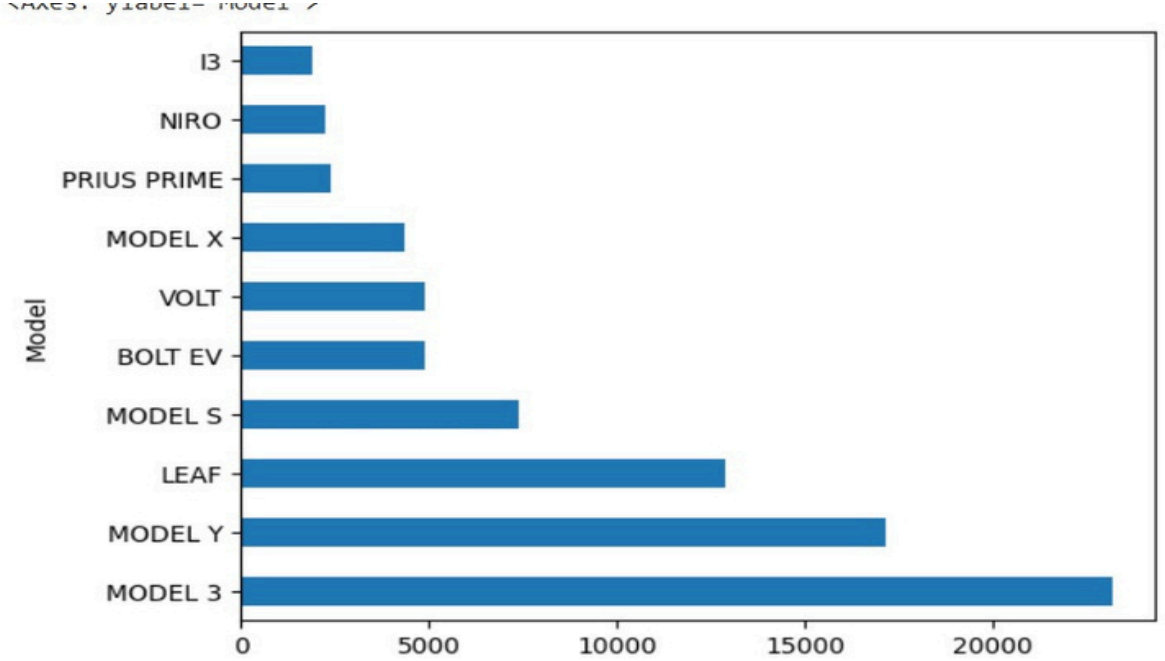
```
------------------------------------------------------
********** Make **********
Unique values: ['TOYOTA' 'CHEVROLET' 'NISSAN' 'FORD' 'TESLA' 'KIA' 'AUDI' 'FIAT' 'BMW'
 'PORSCHE' 'CADILLAC' 'HONDA' 'MITSUBISHI' 'CHRYSLER' 'RIVIAN' 'HYUNDAI'
 'VOLVO' 'VOLKSWAGEN' 'MERCEDES-BENZ' 'JEEP' 'MINI' 'SMART' 'SUBARU'
 'POLESTAR' 'LUCID MOTORS' 'LINCOLN' 'JAGUAR' 'FISKER' 'LAND ROVER'
 'LEXUS' 'TH!NK' 'GENESIS' 'BENTLEY' 'AZURE DYNAMICS']
Nunique values: 34
```

Variety of Manufacturers:The dataset includes a wide range of car makes, from popular brands like Toyota, Chevrolet, and Nissan to more niche manufacturers like Lucid Motors and Azure Dynamics.

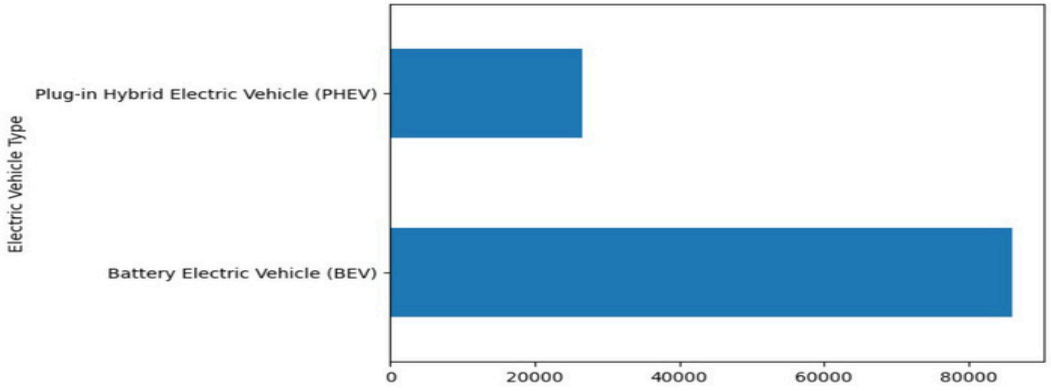
Dominance of Established Brands: Established automotive brands like Toyota, Chevrolet, and Nissan appear to be well-represented in the data.

Presence of Newer Brands:Newer electric vehicle manufacturers like Tesla, Rivian, and Lucid Motors are also included, suggesting the dataset contains information on recent automotive trends.

D iver sity :The presence of 34 unique manufacturers indicates a diverse range of car models in the dataset.



The popularity of different electric car models. The most popular model is the Model 3, followed by the Model Y. The least popular models are the i3 and Niro. The overall popularity of electric cars seems to be increasing, as evidenced by the relatively high sales of the Model 3 and Model Y



The distribution of electric vehicle types. The majority of electric vehicles are Battery Electric Vehicles (BEV), while a smaller portion are Plug-in Hybrid Electric Vehicles (PHEV). This suggests that BEVs are currently more popular among consumers, potentially due to factors such as longer range and lower operating costs.

INNOMATICS RESEARCH LABS

Bi Variant Analysis

- The Relationship between Numeric Vs Numeric
- To finding Relationship between numerical vs Numerical by Person correlation coefficient

Negative Correlation:The negative value (-0.2884) indicates a weak negative correlationbetween electric range and model year. This means that as the model year increases (i.e., newer models), the electric range tends to decrease slightly.

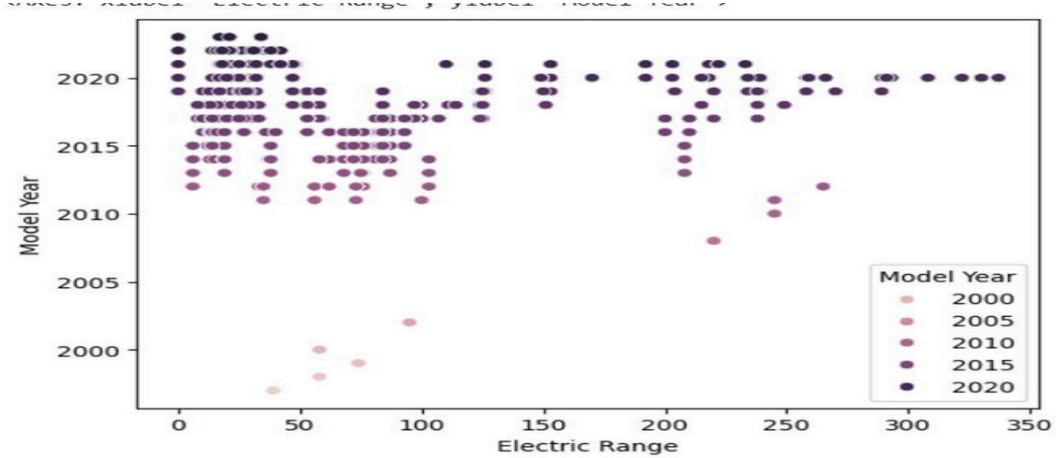
Weak Relationship: The correlation coefficient is relatively close to zero, suggesting a weak relationshipbetween the two variables. This implies that while there's a slight tendency for newer models to have shorter ranges, the relationship is not strong and other factors likely play a more significant role in determining electric range.

Early models might have had slightly longer ranges:As electric vehicle technology has evolved, there might have been a focus on improving other aspects like battery efficiency or vehicle performance, which could have led to a slight decrease in range for newer models.

Other factors influence electric range:Factors such as battery capacity, vehicle size, driving conditions, and charging infrastructure likely have a stronger impact on electric range than the model year alone.

It's important to note that correlation does not imply causation.While the negative correlation suggests a relationship between electric range and model year, it doesn't necessarily mean that one causes the other. Further analysis would be needed to establish a causalrelationship.

INNOMATICS RESEARCH LABS



Positive Trend: There's a general trend of increasing electric range with newer model years. This suggests that electric vehicle manufacturers have been making advancements in battery technology, resulting in vehicles with longer ranges over time.
Clustering:The data points tend to cluster in specific regions. For example, there's a cluster of points around 2015-2020 with electric ranges between 200 and 300 miles, indicating a significant number of vehicles with similar characteristics were introduced during this period.

Outliers:A few data points, particularly in the earlier model years, have relatively high electric ranges. These could be outliers or represent vehicles with unique features or battery technologies.

Variation Within Model Years:While there's a general trend, there's also variation in electric range within each model year. This suggests that factors other than model year, such as battery capacity, vehicle size, and specific manufacturer choices, influence the electricrange

Numerical Vs Categorical

Make	Electric Range
AUDI	62.876930
AZURE DYNAMICS	56.000000
BENTLEY	18.666667
BMW	46.657479
CADILLAC	35.537037
CHEVROLET	109.766549
CHRYSLER	32.361204
FIAT	85.624088
FISKER	33.000000
FORD	16.848084

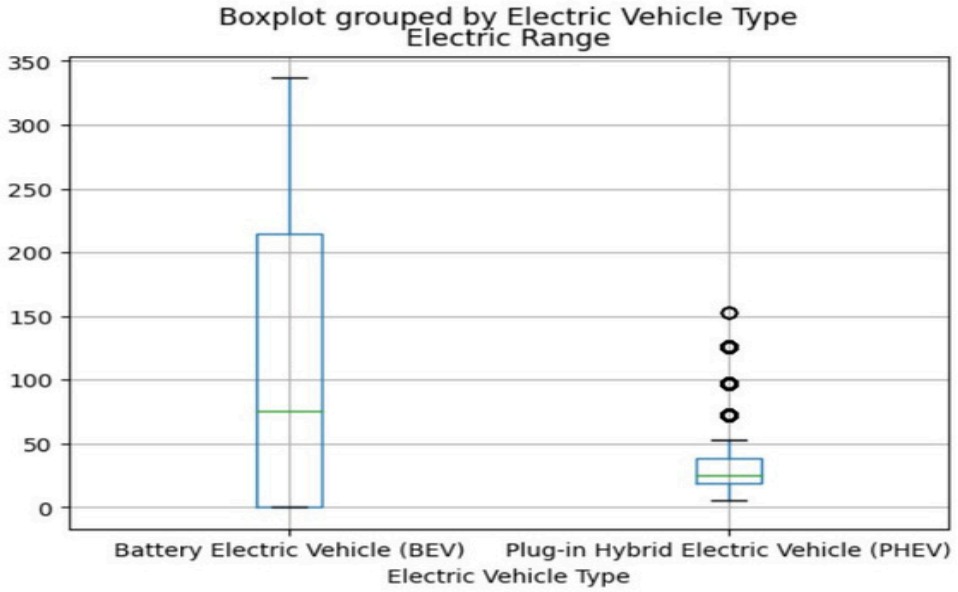
Range Variation: The electric range varies significantly across different car makes. Chevrolet has the highest average range (109.77 miles), while Bentley has the lowest (18.67 miles).

Brand-Specific Performance:Some brands consistently demonstrate higher electric ranges than others. This could be attributed to factors such as battery technology, vehicle design, and manufacturer focus on electric vehicles.

Outliers:There might be outliers within each make that influence the average electric range. For example, a single vehicle with an exceptionally high or low range could significantly impact the av era ge.

Visualize the data: Creating a bar chart or box plot could provide a clearer visual representation of the range variation across different makes.

Consider additional factors: Explore other factors that might influence electric range, such as battery capacity, vehicle size, and driving co ndi tion s.



Median Range: The median electric range for BEVs is significantly higher than that for PHEVs. This indicates that BEVs generally have a longer range compared to P HE Vs.

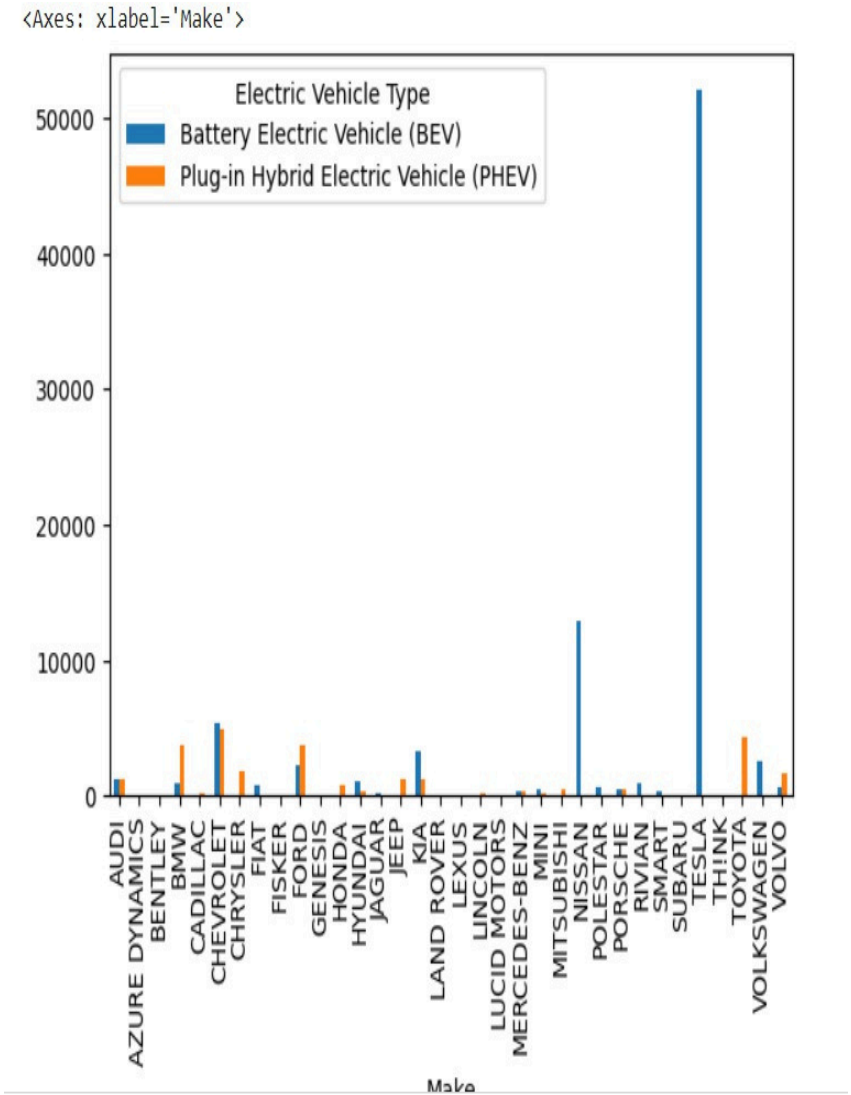
Range Variability:The interquartile range (IQR) for BEVs is also larger than that for PHEVs. This suggests that there is more variation in the electric range among BEVs compared to PHEVs.

Outliers:There are a few outliers (individual data points outside the whiskers) in the PHEV category, indicating that some PHEVs have significantly higher or lower electric ranges compared to the majority of the vehicles in that category.

Distribution Shape:The boxplots show a slight right-skewness for both BEVs and PHEVs, indicating that there are a few vehicles with higher electric ranges that are pulling the median to the right.

Overall, the boxplot suggests that BEVs generally have a higher median electric range and a wider range of values compared toPHEVs. However, there are some outliers in the PHEV category, indicating that there might be a few PHEVs with exceptionally high or low electricranges.

Categorical Vs Categorical



BEV Dominance:Many car makes have a higher number of BEVs compared to PHEVs. This suggests that BEVs are becoming more popular than PHEVs in the market.

Make-Specific Trends:Some car makes have a significantly higher number of BEVs or PHEVs. For example, Tesla has a very high number of BEVs, while Chrysler has a relatively higher number of PHEVs.

Limited PHEV Offerings:Several car makes have a limited number of PHEVs available. This could be due to various factors, such as market demand, technological limitations, or manufacturer focus on BEVs.

Emerging Trends:Newer car makes, such as Rivianand Lucid Motors, are primarily focused on BEVs, indicating a shift towards fully electric vehicles.

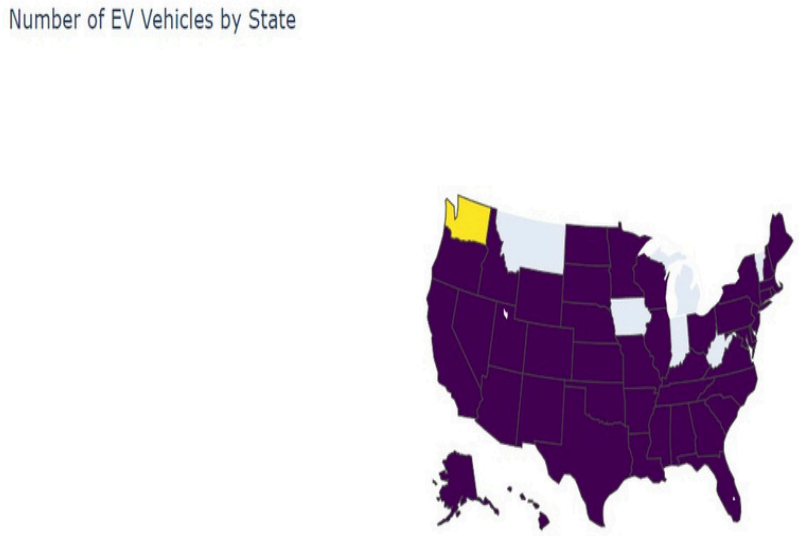
Analyze market share:Calculate the market share of each car make for BEVs and PHEVs to compare their relative popularity.

Consider factors influencing choice:Explore factors that might influence consumers' choice between BEVs and PHEVs, such as range, price, and charging infrastructure.

Track trends over time:Analyze the data over time to identify changes in the distribution of BEVs and PHEVs and understand the evolving market dynamics.

INNOMATICS RESEARCH LABS

Number of EV Vehicles by State



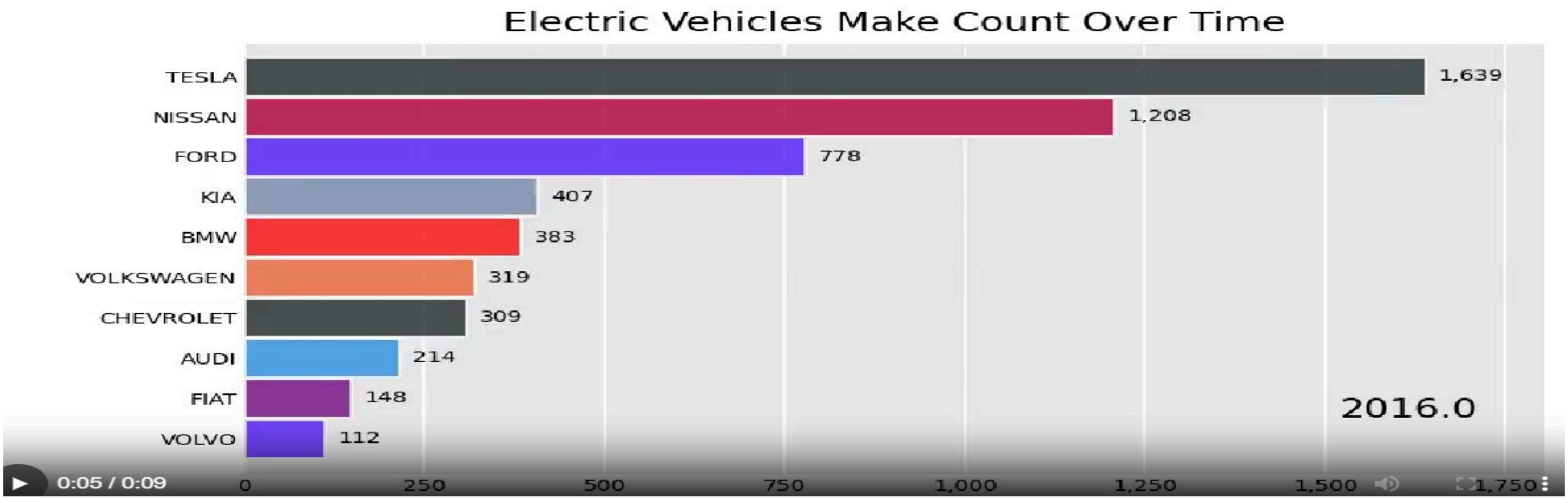
Regional Variation:There are significant regional differences in the density of EVs. States in the western US, particularly California and Washington, have a much higher concentration of EVs compared to states in the Midwest and South.

California Dominance:California stands out with the highest number of EVs, likely due to factors such as favorable policies, a large population, and a strong focus on environmental sustainability.

Clustering:There are clusters of states with similar EV densities. For example, states in the Pacific Northwest (Washington, Oregon) and the Northeast (New York, Massachusetts) tend to have higher EV adoption ra tes.

Lower Density:
States in the Midwest and South have lower EV densities, which could be attributed to factors such as lower population, less favorable policies, and different driving habits.

INNOMATICS RESEARCH LABS



Tesla Dominance:Tesla leads the market with a significantly higher number of EVs compared to other manufacturers. This dominance is evident in the length of the Tesla bar, which is much longer than the others.

Nissan and Ford:Nissan and Ford follow Tesla with a considerable number of EVs, indicating their strong presence in the electric vehicle market.

Other Manufacturers:
The remaining manufacturers have a relatively lower number of EVs, suggesting that the market is still dominated by a few key players.

Time Period:The chart is likely for a specific year, which is indicated as 2016.0. This information is important for understanding the market dynamics at that particular point in time.

Conclusion

Reading is a powerful tool that can expand your knowledge, stimulate your imagination, and improve your critical thinking skills. It allows you to explore different worlds, learn about diverse cultures, and gain new perspectives. Whether you prefer fiction or non-fiction, there's a book out there for everyone.

THANK YOU

INNOMATICS RESEARCH LABS