PROJECT 2:- MARKET SEGMENTATION

ANALYSING ELECTRIC VEHICLE MARKET USING SEGMENTATION ANALYSIS

BY: - SAHIL NAIK, BIPRAJIT NAMASUTRA, HIMANSHI JAIN

Overview

The global electric vehicle (EV) market is developing at a rapid pace. According to EV volumes, overall electric vehicle reached a global share of 8.3% (including battery electric vehicles [BEVs] and Plug- in hybrid electric vehicles [PHEVs]) in 2021 from 4.2% in 2020 with 6.75 million vehicles on the road. This is an increase of 108% as of 2020.

Segmentation analysis is an important step before we embark on a marketing plan. It is important to learn how to analyze your audience and market. Electric Vehicle Market is segmented by Type (Battery Electric Vehicle and Plug-in Hybrid Electric Vehicle), Clean Alternative Fuel Vehicle (CAFV) Eligibility, Model, Make, and Electric Range etc.

Data Collection

The data used in the report is obtained from the following source:

Kaggle-https://www.kaggle.com/code/saibhargav2810/eda-electric-vehicle-poulation/data

This dataset presents information about the Electric Vehicle Market**.**

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legislative District	DOL Vehicle ID	Vehicle Location
0	JTMEB3FV6N	Monroe	Key West	FL	33040	2022	TOYOTA	RAV4 PRIME	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible	42	0	NaN	198968248	POINT (-81.80023 24.5545)
1	1G1RD6E45D	Clark	Laughlin	NV	89029	2013	CHEVROLET	VOLT	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible	38	0	NaN	5204412	POINT (-114.57245 35.16815)
2	JN1AZ0CP8B	Yakima	Yakima	WA	98901	2011	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	73	0	15.0	218972519	POINT (-120.50721 46.60448)
3	1G1FW6S08H	Skagit	Concrete	WA	98237	2017	CHEVROLET	BOLT EV	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	238	0	39.0	186750406	POINT (-121.7515 48.53892)
4	3FA6P0SU1K	Snohomish	Everett	WA	98201	2019	FORD	FUSION	Plug-in Hybrid Electric Vehicle	Not eligible due to low battery range	26	0	38.0	2006714	POINT (-122.20596 47.97659)

Glimpse of original data.

After that we Drop few columns that are not important for our study such as VIN,POSTAL CODE,BASE MSRP,LEGISLATICE DISTRICT,DOL VEHICLE ID,ELECTIRIC UTILITY,2020 CENSUS TRACT.

```
df.tail()
```

	County	City	State	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Vehicle Location
112629	King	Duvall	WA	2022	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	POINT (-121.98609 47.74068)
112630	San Juan	Friday Harbor	WA	2019	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	150	POINT (-123.01648 48.53448)
112631	King	Vashon	WA	2022	FORD	ESCAPE	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible	38	POINT (-122.4573 47.44929)
112632	King	Covington	WA	2018	KIA	NIRO	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range	26	POINT (-122.09124 47.33778)
112633	King	Covington	WA	2022	VOLVO	XC90	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range	18	POINT (-122.09124 47.33778)

Above data we use for modeling and visualization.

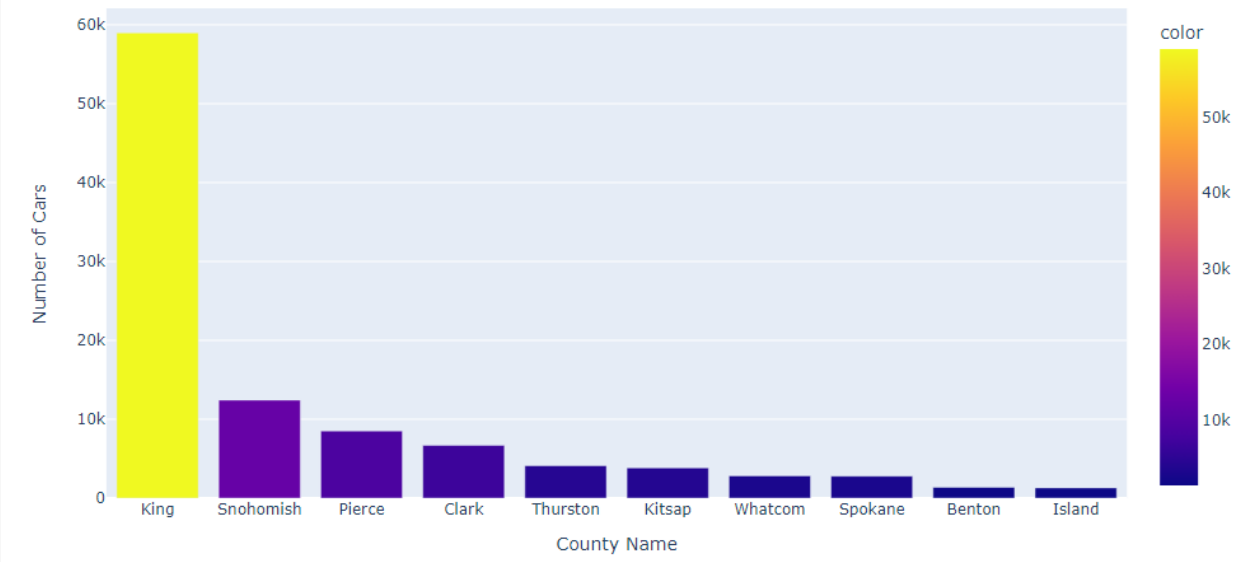
Python libraries such as NumPy, Pandas and Scikit-learn etc. are used for the workflow and the results obtained are ensured to be reproducible.

Exploratory Data Analysis

Before we start data analysis or run our data through a machine learning algorithm, one must clean the data and make sure it is in a suitable form. Further, it is essential to know any

recurring patterns and significant correlations that might be present in the data. The process of getting to know the data in depth is called Exploratory Data Analysis.

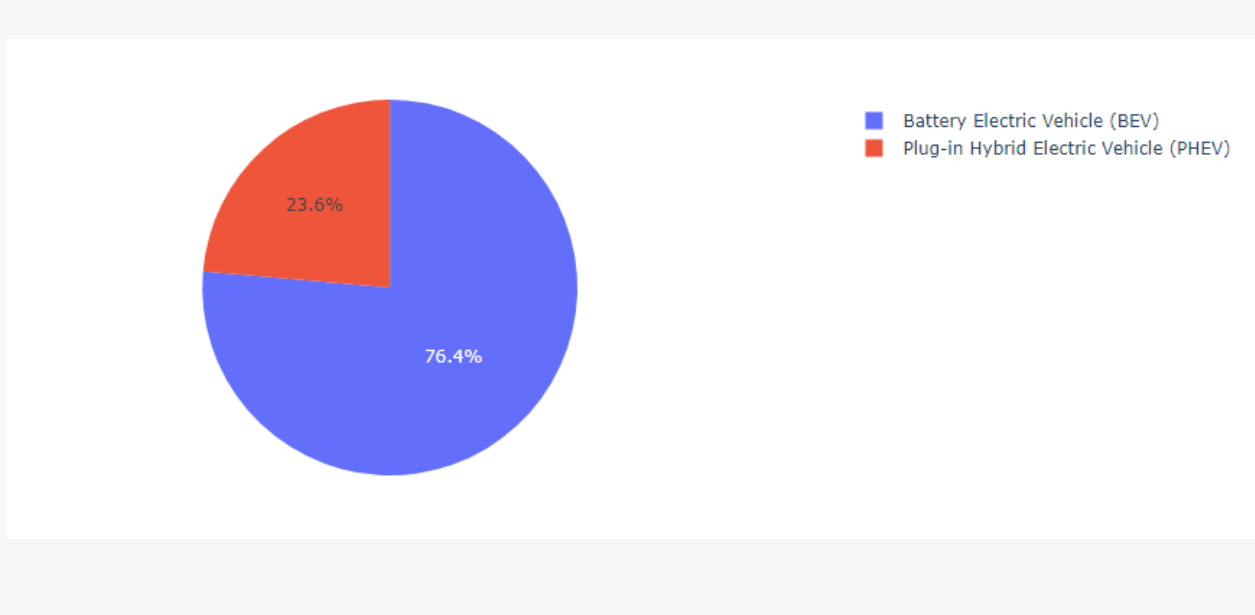
Firstly, we check from which county we have more entries from (Top 10).



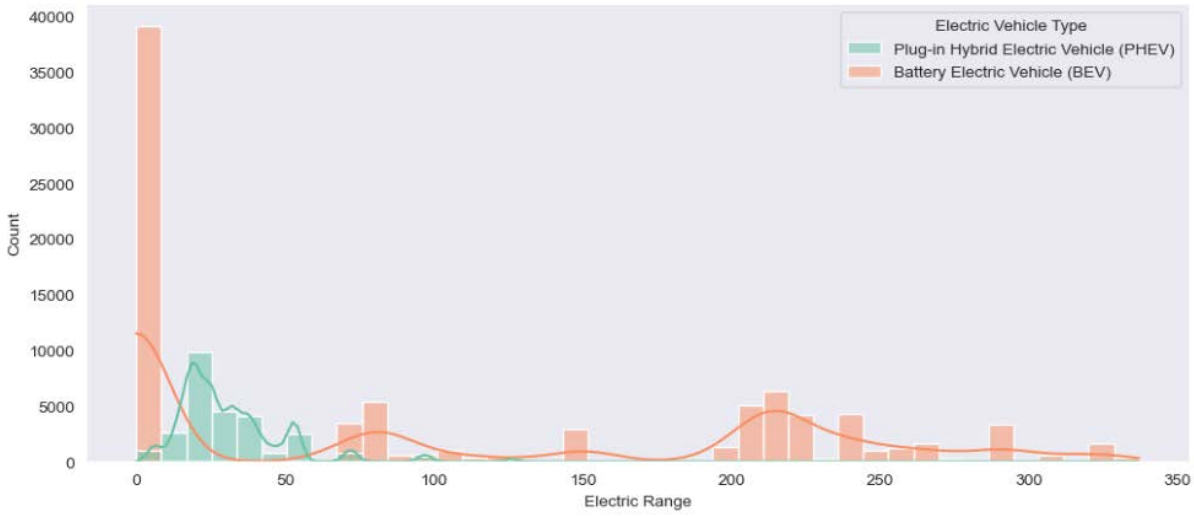
After that, we find out the information about top selling model for top company.

```
Top selling model for TESLA is -----------> MODEL 3
Top selling model for NISSAN is -----------> LEAF
Top selling model for CHEVROLET is -----------> BOLT EV
Top selling model for FORD is -----------> FUSION
Top selling model for BMW is -----------> I3
Top selling model for KIA is -----------> NIRO
Top selling model for TOYOTA is -----------> PRIUS PRIME
Top selling model for VOLKSWAGEN is -----------> ID.4
Top selling model for AUDI is -----------> E-TRON
Top selling model for VOLVO is -----------> XC90
```

An electric vehicle converts over 50% of the electrical energy from the grid to power at the wheels, whereas the gas-powered vehicle only manages to convert about 17%–21% of the energy stored in gasoline. The demand for fuel-efficient vehicles has increased recently, owing to rise in price of petrol and diesel. This is attributed to depleting fossil fuel reserves and growth in tendency of companies to gain maximum profit from these oil reserves. Thus, these factors give rise to the need for advanced fuel-efficient technologies, leading to surge in demand for electrically powered vehicles for travel.



Percentage distribution between BEVs and PHEVs.



The data shows that BEVs will give more range but there are some vehicles which give almost 0 range on PHEV.

Modeling

As the part discussed in the above EDA section which deals with data exposure to various angles. In this part we are focused on analysis of the data, extracting useful insights from it and transforming it to form suitable for the machine learning model.

Features Extraction

```
df.select_dtypes('object')
```

	County	City	Make	Model	Vehicle Location	Electric Utility
2	Yakima	Others	NISSAN	LEAF	POINT (-120.50721 46.60448)	PACIFICORP
3	Skagit	Others	CHEVROLET	BOLT EV	POINT (-121.7515 48.53892)	PUGET SOUND ENERGY INC
4	Snohomish	Everett	FORD	FUSION	POINT (-122.20596 47.97659)	PUGET SOUND ENERGY INC
5	Snohomish	Bothell	TESLA	MODEL 3	POINT (-122.18384 47.8031)	PUGET SOUND ENERGY INC
6	Snohomish	Everett	NISSAN	LEAF	POINT (-122.23019 47.94949)	PUGET SOUND ENERGY INC
...	...	...	...	...	...	...
112629	King	Others	TESLA	MODEL Y	POINT (-121.98609 47.74068)	PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)
112630	Others	Others	NISSAN	LEAF	POINT (-123.01648 48.53448)	BONNEVILLE POWER ADMINISTRATION||ORCAS POWER &...
112631	King	Others	FORD	ESCAPE	POINT (-122.4573 47.44929)	PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)

In the above Evs dataset, we have to predict which car model shall the company make for its profitable business growth. So, except for the Make and the Model column, all other columns are its features which we shall be providing to the machine learning model. Now these columns need to be properly analysed and checked, if any information can be extracted.

Looking at the column “Vehicle Location”

Information regarding the geographic location where a particular car model is sold can be extracted. This is achieved using split operation as shown below.

```
df['lat']=df['Vehicle Location'].apply(lambda x:x.split('(')[1].split(')')[0].split(' ')[0])
df['lon']=df['Vehicle Location'].apply(lambda x:x.split('(')[1].split(')')[0].split(' ')[1])
```

Data Transformation

After extracting the geographic information (latitude and longitude), this extracted information is in string object form. So, they are converted to float and then to integer, because machine learning models cannot understand string objects.

```
df['lon'] = df['lon'].astype(float)
df['lat'] = df['lat'].astype(float)
```

Also, some of the column values are presented in categorical order. This needs to be converted to integer. This is accomplished using the get_dummies() method from the pandas library.

But in doing so, there is a problem as there are a huge number of values or categories. So, the categories which occur maximum are considered and the rest are grouped as “others” categories.

```
df['County']= df['County'].replace(['Clallam','San Juan','Cowlitz','Mason','Lewis','Grays Harbor',
                    'Kittitas','Franklin','Grant','Walla Walla','Douglas','Whitman',
                    'Klickitat','Okanogan','Pacific','Skamania','Stevens','Asotin',
                    'Wahkiakum','Adams','Pend Oreille','Lincoln','Ferry','Columbia','Garfield'],'Others')
```

```
df['City']=df['City'].replace(df['City'].value_counts().iloc[25:].index.tolist(),'Others')
```

After this, the data are converted as:

```
dummies = pd.get_dummies(df['County'], drop_first=True)
df = pd.concat([df.drop('County', axis = 1), dummies], axis=1)
```

```
dummies = pd.get_dummies(df['City'], drop_first=True)
df = pd.concat([df.drop('City', axis = 1), dummies], axis=1)
```

Dropping the unwanted column:

Further analysis is done to find-out the useful column and eliminate the column which is less important or no important information can be extracted.

Here we also drop columns for which the dummy variables are created.

Creating the model:

Once all the features columns are converted to suitable form for our ML model, it is divided into training and test dataset. This is done using a train test split method from the sklearn library.

```
from sklearn.model_selection import train_test_split
```

```
X = df.drop(['Make', 'Model'], axis = 1).values
```

```
y= df[['Make', 'Model']].values
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=101)
```

Here X represents the actual data on which the model is trained and tested. ‘y’ represents the validation data upon which the model performance is tested.

Choosing the right ML algorithm and fitting it to our dataset:

As some labeled data is present, our model will be a supervised ML model. Out of all the many supervised algorithms, we choose “Decision Tree” as our output is some categorical values and Decision Tree is known for providing a good performance regarding categorical value.

We import it from the sklearn library inside the tree module, then fit it to our dataset.

```
from sklearn.tree import DecisionTreeClassifier
```

```
dtree=DecisionTreeClassifier()
```

```
dtree.fit(X_train, y_train)
```

```
DecisionTreeClassifier()
```

Testing the performance of our model:

Once the model is trained, its prediction is tested using the test dataset.

```
predictions = dtree.predict(X_test)
```

```
print(predictions.shape)
```

```
(33646, 2)
```

```
print(predictions)
```

```
[['TESLA' 'MODEL X']
 ['POLESTAR' 'PS2']
 ['CHEVROLET' 'BOLT EV']
 ...
 ['TESLA' 'MODEL 3']
 ['TESLA' 'MODEL 3']
 ['VOLVO' 'C40']]
```

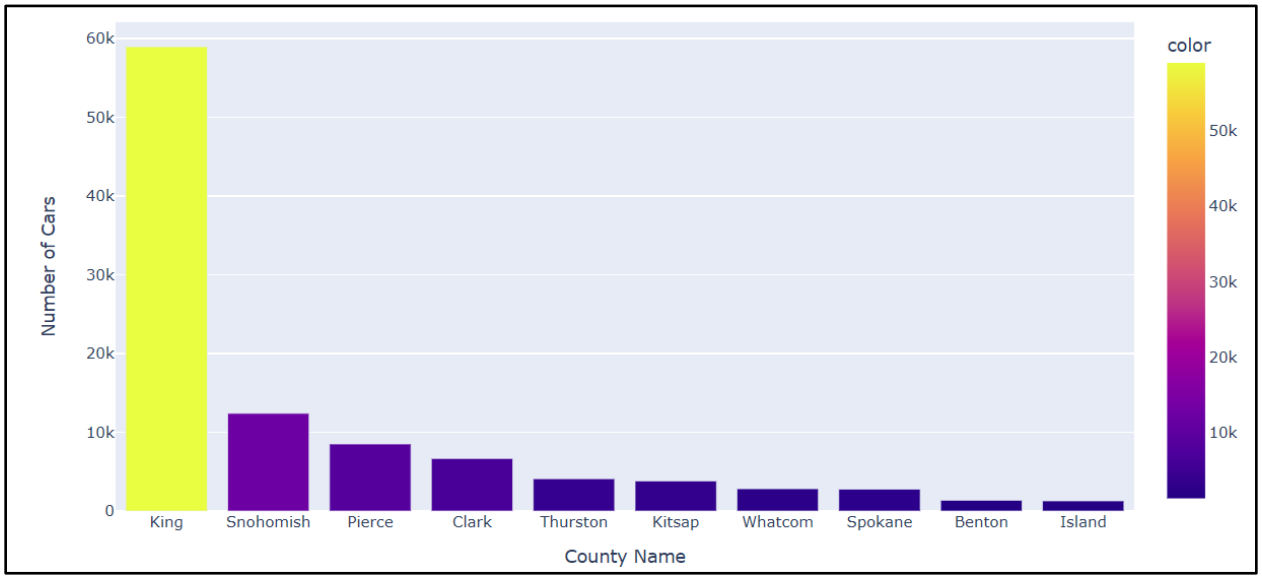
So, we can observe that the model suggests some of the Electric car's Make and Model (in ascending order) which the company can implement. These car's Make and Model are suggested based on the dataset and features provided. Therefore, to the best knowledge of the provided dataset, these suggested models are supposed to provide maximum profit for the company.

Describing Segments

In segment profiling we will be understanding and studying different segments of the given data, which will include to understand the market and the requirements of the customer.

There are various factors that are involve in deciding a segment. For instance, the variable named 'Make' tells us about the company which made the EV, this can be a key factor in deciding and understanding which companies dominate the EV market. Another variable called 'Electric Range' can also be helpful since it let us know the distance each EV model can travel in a single charge; this can be used to understand which model made by which company has the highest electric range and are able to dominate the EV market. Let's breakdown and understand each key variable one by one:

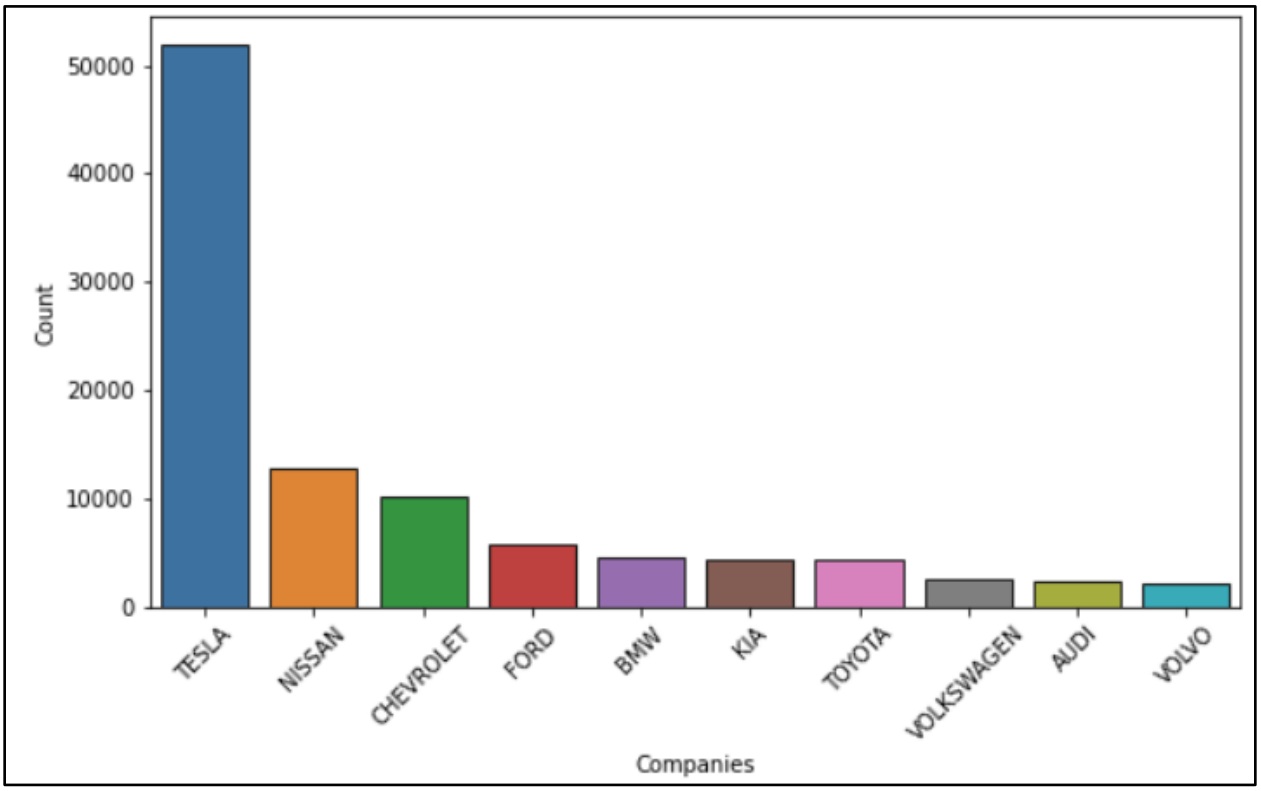
1. **County having most number of EV cars:** By looking at the geographic data of each EV models, we can see that the county named 'King' contains the most number of EV cars i.e. 58.98K followed by 'Snohomish' having only 12.412K number of EV cars.



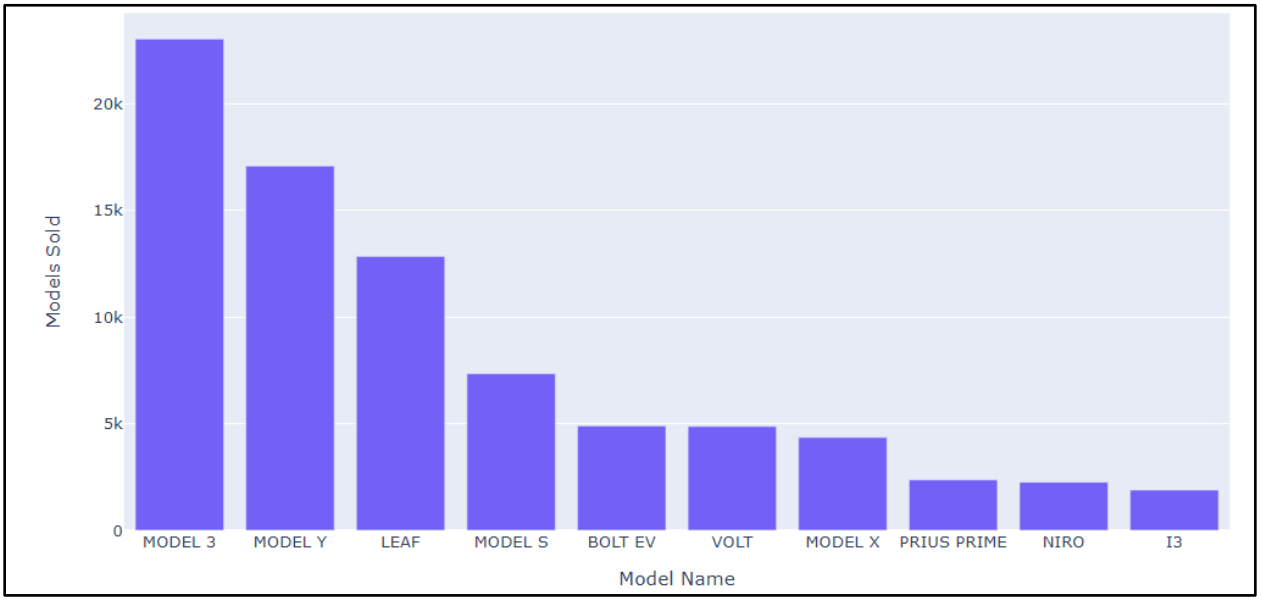
When we look at the location of 'King county' we can see that it is located near Seattle,USA, and it is no doubtably an urban metropolitan city, which explains the presence of enormous amount of EV models in that county. According to wiki, King County is located in the U.S. state of Washington. The population was 2,269,675 in the 2020 census, making it the most populous county in Washington, and the 13th-most populous in the United States. The county seat is Seattle, also the state's most populous city. Talking more about this county we know that it is the eight largest bus transit networks in USA when compared to another county. In essence if we know each attribute that contribute to the enormous population of EV models in 'King County', we can then use this data to compare it with any other city data from India. At this stage, while looking at the primary attributes, we can say that cities of India like Kolkata, New Mumbai, Pune, Chennai and Delhi are a best fit to start the EV market.

When we look at the next county i.e., Snohomish, we know from wiki that, Snohomish County is a county located in the U.S. state of Washington. With a population of 827,957 as of the 2020 census, it is the third-most populous county in Washington, after nearby King and Pierce counties, and the 75th-most populous in the United States. Even though the population of EV models is very few as compared to 'King County', it is still to be noted that it stands in 2nd place amongst every other county. The city is an urban metropolitan with average amount of bus and train transits facilities.

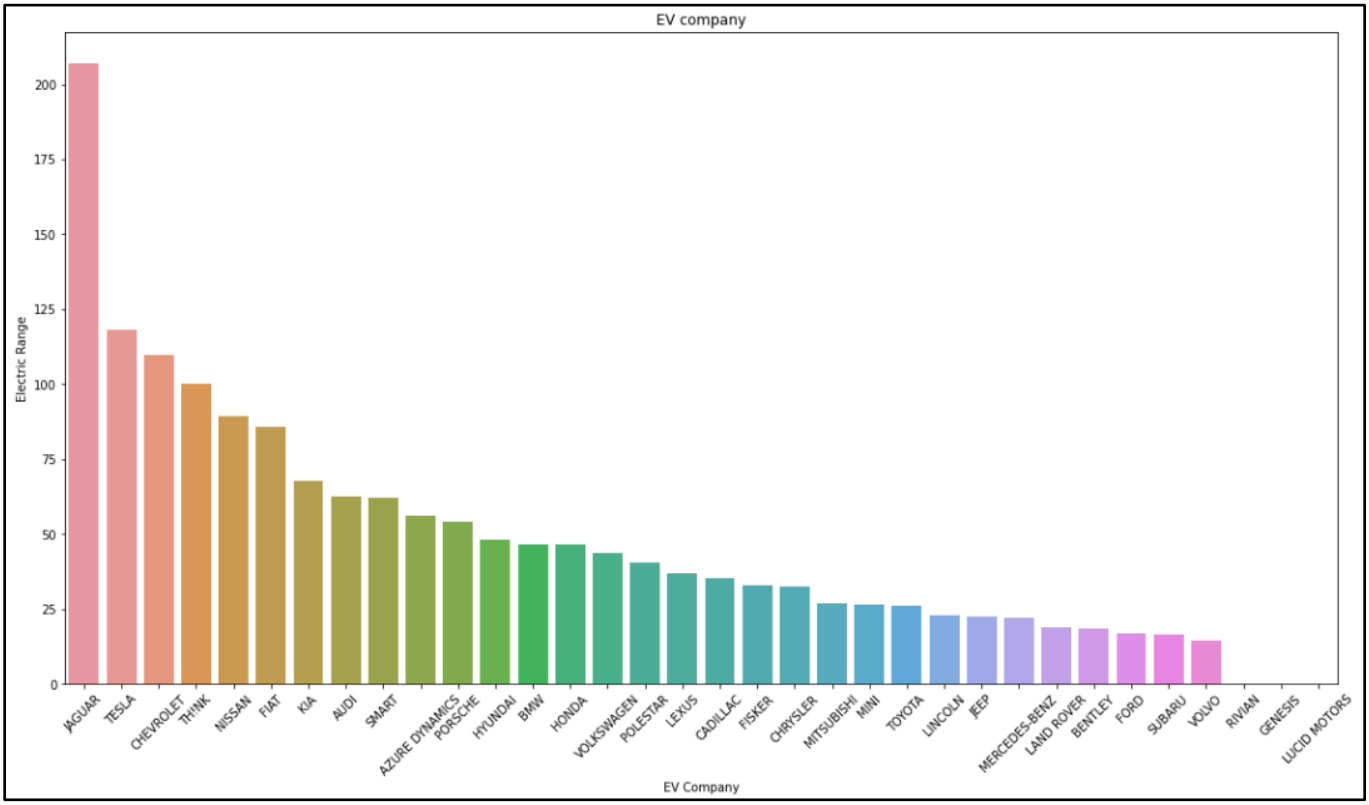
2. **Companies dominating EV market:**So far, we understood which attributes of a particular location from India must be target to start the EV market, now let's understand which EV company will be more suitable for starting the EV market in India.
   While starting an EV market in any region, we first need to understand the region's requirement and the customer attributes to help us better elaborate the current situation of EV market with current EV companies. For this dataset, the most dominant EV company was Tesla, which managed to sold more than 50,000 units, followed by Nissan with sales unit just reaching over 10,000. By looking at the graph below, we can easily visualize the amount of market Tesla has gained over other companies. Other EV companies like the BMW, Audi and Volvo lies at the bottom part of the EV market.



3. **Most sold EV model:** After looking at the total sales of Tesla's EV model, one can easily say that the Tesla is the best EV company out there to start the EV market with, but we also need to consider which EV model of Tesla is the most successful from the provided data.From the graph below we can easily say that the Model 3 and Model Y are the two top most successful EV model in market since their count of unit sold is very large. If we look at these model and try to find out which EV company made these models, then we would find that the Tesla has made the Model 3, Model Y, Model S and the Model X EVs. According to the results we can suggest that the Tesla's Model 3 and Model Y will be a good fit to start the EV market in India. Since we already know that the US market is flooded with Tesla EV models, we can say the same for the Indian market and customer requirements.



4. **Electric range of EVs:** Now we know that the most EVs are sold by the Tesla and the most used Tesla model is the Model 3. Now we need to find which EV car company actually manufactures an EV model which can travel the furthest distance in single charge. From the dataset provided the EV company 'Jaguar' stands first when it comes to the furthest distance traveled in single charge followed by Tesla. This means that even though the sales of Jaguar's EV car are not that much, but their cars can travel the longest distance in single charge whereas Tesla's are on second place and we already know that their sales are sky rocketing and they have established a dominatingwin in the EV market. The average distance Jaguar's model can go in single charge is just above 200 Miles, whereas Tesla's model can go roughly 110 Miles in single charge. Meaning Tesla's will require more charging stations and will also travel less distance as compared to Jaguar's model. The figure below illustrates the Electric range of EV models in a single charge manufactured by each company from the dataset.



Selecting the Target Segments

So far, by looking at the previous EV market data we can say that in order to start EV market in India we have to first target Urban Metropolitan cities like Mumbai, New Delhi, Kolkata, Chennai and Pune.

From the data it was clear that the most dominant EV company is the Tesla which also happens to produce the most sold EV in entire USA i.e., Model 3. The Tesla's Model 3 is the most successful EV out there which has exceptional travel distance in single charge. Even though we saw that the Jaguar's EV model also had the longest distance traveled in single charge, but looking at the sales and the Price, we can say that the Jaguar's EV are not that much successful and is only just a luxury commodity which is target for only 1% of the people, making it an unfit for starting the EV market in India.

Insights and Recommendation

Electric vehicles are not just the wave of the future, they are saving lives today. Electric vehicles have a smaller carbon footprint than gasoline-powered cars, no matter where your electricity comes from .Because electric vehicles are more efficient in converting energy to power cars and

trucks, electricity across the board is cleaner and cheaper as a fuel for vehicles, even when that electricity comes from the dirtiest grid.

Running electric or hybrid cars on the grid in any state has lower greenhouse gas emissions than gasoline-powered cars. Planning now by states and utilities to build infrastructure for charging electric vehicles will go a long way.

The acknowledgement of finite energy options clearly indicates that innovative companies unlocking sustainable long – term energy solutions like EV, solar and alternate energy will score better commercial results and are also equipped to tap into the sector's long term potential .As an investment opportunity , sustainability , specifically clean energy is at cusp of becoming one of the capital generating long – term assets like gold , real estate and equities. The EV industry in India can create some 10 million or 1 crore indirect jobs by 2030, according to estimates by the Ministry of Skill Development and Entrepreneurship.

Project Code Link

https://github.com/Sahil-Naik/Feynn_Lab/blob/main/T2/EV_final.ipynb