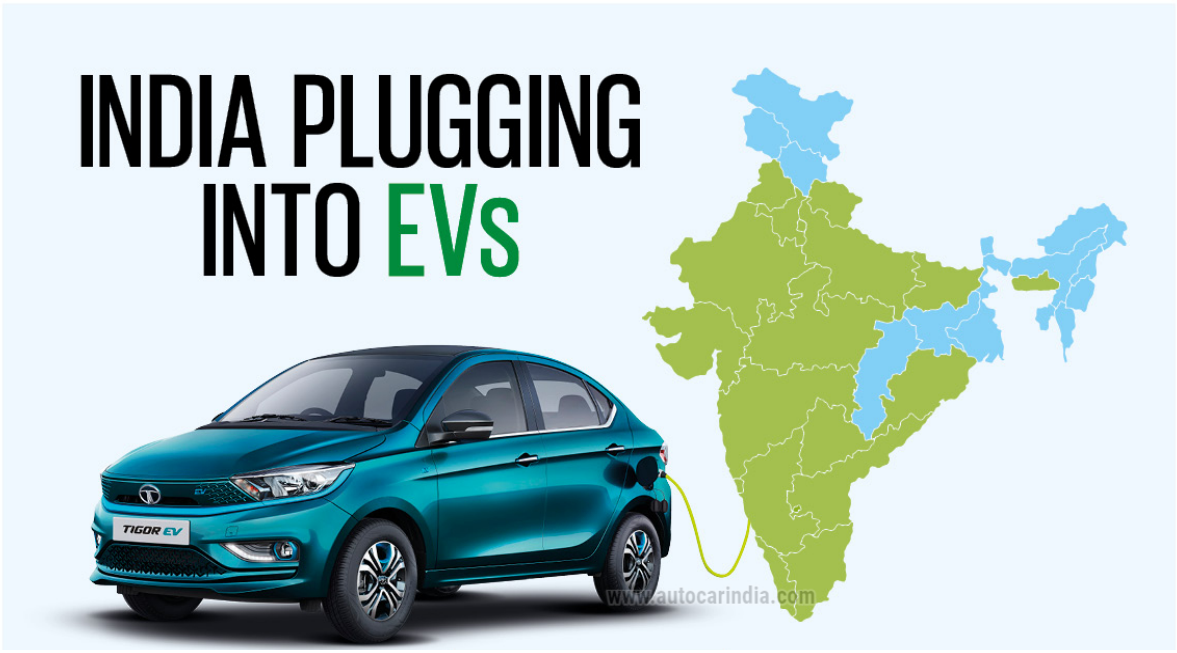
Feynn Labs Project

EV Market Segmentation in India



~ Aditya Kendre

Problem Statement:

To analyze the EV Market in India using Segmentation analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use Electric vehicles

0.Fermi Estimation (Breakdown of Problem Statement)

India's Population = **130 Crore**

Number of States and UTs in India = **28 + 8**

The number of registered vehicles across India was **295 million** in 2019.

Vehicle registrations compound annual growth rate = **10%** (2007-19)

The electric vehicle market in India is expected to increase at a compounded annual growth rate of **49%** between 2021 to 2030 while the EV segment's volume may cross annual sales of **17 million** by 2030.

Then we can assume a **8%** Vehicle registrations compound annual growth rate after 2030. Also we will **target the car segment** in the **Electric Vehicles Specifically**.

Why will the EV Market rise and is it stable in the long run?

Yes, it will be stable in the long run because it is renewable, environment friendly and will be economical in the long run than fossil fuels.

We will see the statewise sale of vehicles and also see which company sells the most cars and generates more revenue including price range, safety and customer satisfaction.

Then we will move to the KM range, durability, safety, charging time, customer satisfaction, battery etc.

1. Data Sources

https://www.kaggle.com/datasets/Vehicle Dataset

https://data.gov.in/resource/all-india-level-composition-vehicle-population-during-2015-2016

https://data.gov.in/

https://pib.gov.in/PressReleasePage.aspx?PRID=1842704

2. Data Pre-processing (Steps and Libraries used)

1)Imported Libraries like Pandas, Numpy, Seaborn, Matplotlib to analyze the data. Other libraries were imported as required.

2)We took a dataset of EV cars and converted the Euro Prices to INR. We use SKLearn library mostly for this project

3)Performed Feature Selection, One hot encoding and removed null values and checked if there are any missing values.

Q)Which location in India is most suitable to create the early market?

Sr. No.	State Name	Total Electric Vehicle	Total Non-Electric Vehicle	Total
1	Andaman & Nicobar Island	162	1,46,945	1,47,107
2	Arunachal Pradesh	20	2,52,965	2,52,985
3	Assam	64,766	46,77,053	47,41,819
4	Bihar	83,335	1,04,07,078	1,04,90,413
5	Chandigarh	2,812	7,46,881	7,49,693

6	Chhattisgarh	20,966	68,36,200	68,57,166
7	Delhi	1,56,393	76,85,600	78,41,993
8	Goa	3,870	10,71,570	10,75,440
9	Gujarat	45,272	2,06,05,484	2,06,50,756
10	Haryana	37,035	1,07,78,270	1,08,15,305
11	Himachal Pradesh	1,175	19,64,754	19,65,929
12	Jammu and Kashmir	2,941	18,69,962	18,72,903
13	Jharkhand	16,811	64,86,937	65,03,748
14	Karnataka	1,20,532	2,68,70,303	2,69,90,835

15	Kerala	30,775	1,57,74,078	1,58,04,853
16	Ladakh	26	38,302	38,328
17	Maharashtra	1,16,646	3,10,58,990	3,11,75,636
18	Manipur	586	4,99,324	4,99,910
19	Meghalaya	49	4,59,001	4,59,050
20	Mizoram	21	3,15,626	3,15,647
21	Nagaland	58	3,39,129	3,39,187
22	Odisha	23,371	98,45,073	98,68,444

23	**Puducherry**	**2,149**	**12,13,735**	**12,15,884**
24	**Punjab**	**14,804**	**1,24,63,019**	**1,24,77,823**
25	**Rajasthan**	**81,338**	**1,73,27,388**	**1,74,08,726**
26	**Sikkim**	**21**	**97,189**	**97,210**
27	**Tamil Nadu**	**82,051**	**2,98,42,376**	**2,99,24,427**
28	**Tripura**	**9,262**	**6,50,026**	**6,59,288**
29	**UT of DNH and DD**	**183**	**3,07,671**	**3,07,854**
30	**Uttarakhand**	**31,008**	**33,12,041**	**33,43,049**

31	**Uttar Pradesh**	**3,37,180**	**4,00,92,490**	**4,04,29,670**
32	**West Bengal**	**48,767**	**1,41,34,171**	**1,41,82,938**
Grand Total		**13,34,385**	**27,81,69,631**	**27,95,04,016**

Table : The number of electric vehicles currently being used on the roads of India, State wise as on 14-07- 2022

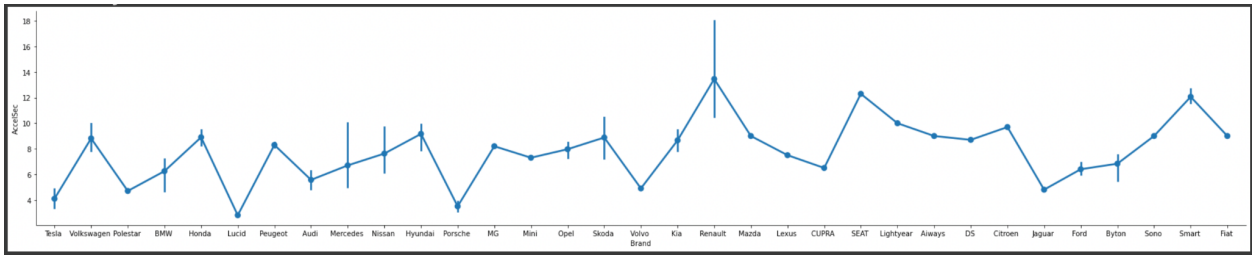
From the above table, we should target states like **UP, Bihar, Delhi, Karnataka and Maharashtra** as they quickly adopt to innovative technologies.

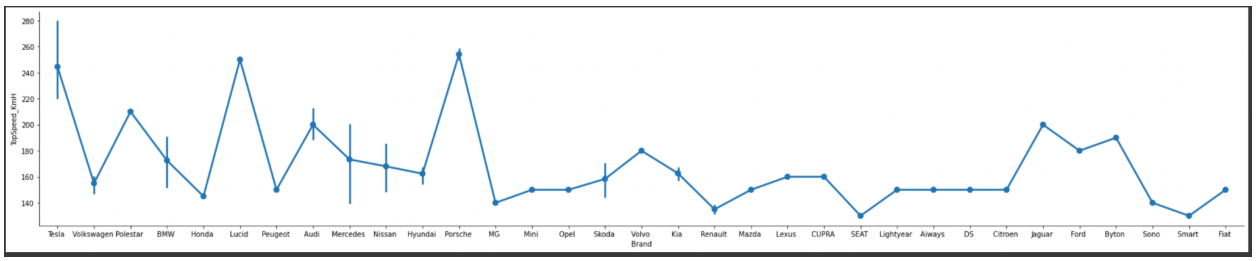
Q)Which demographic, psychographic, behavioural or other factors your team will target based on Data Analysis of available datasets. In the event of unavailability of proper datasets, how your team will base your decisions as accurate and unbiased as possible.

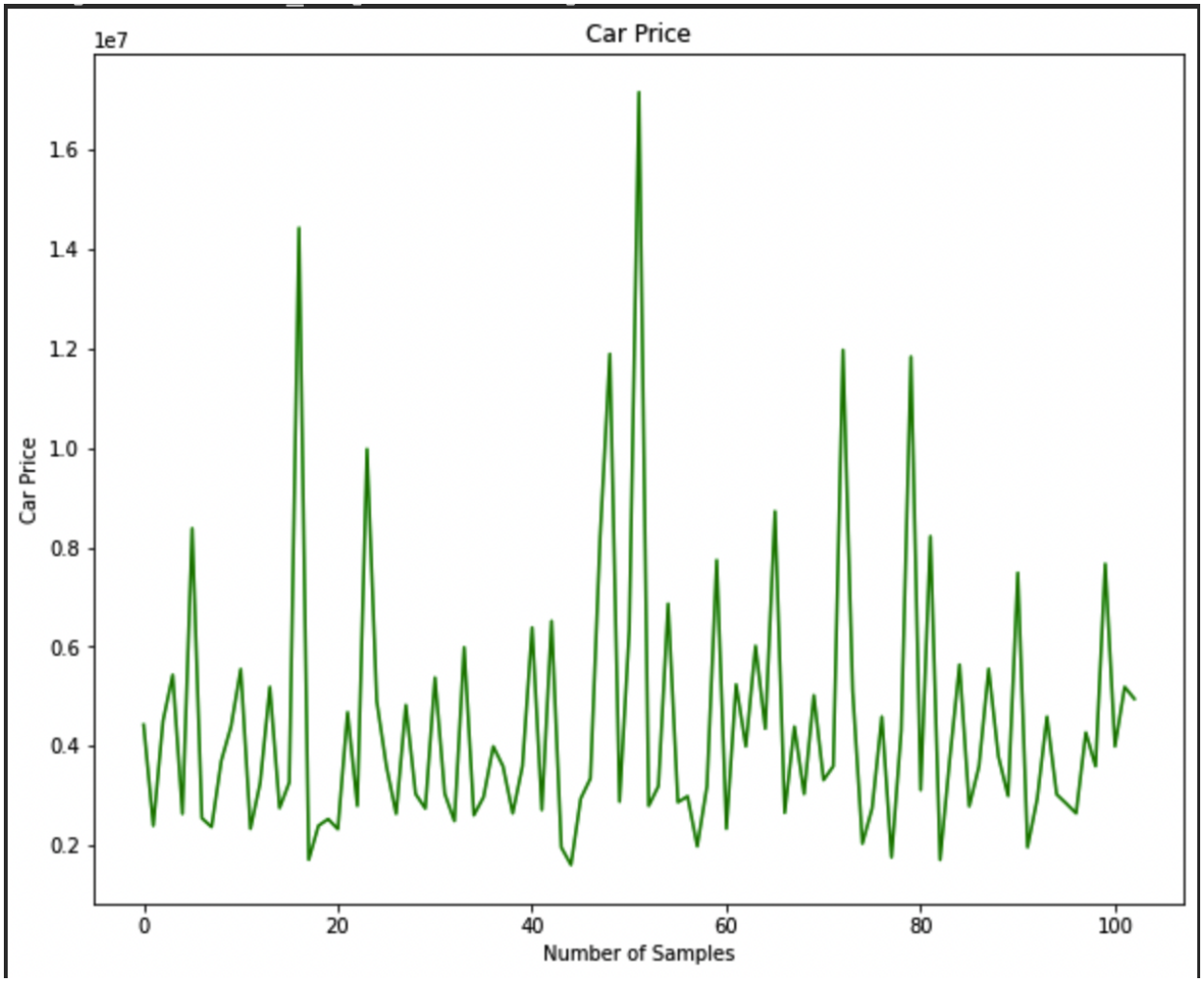
Because we were unable to obtain customer survey data from the Indian market, we had to use a sample data set from Kaggle for our market segment analysis.

EDA before clustering

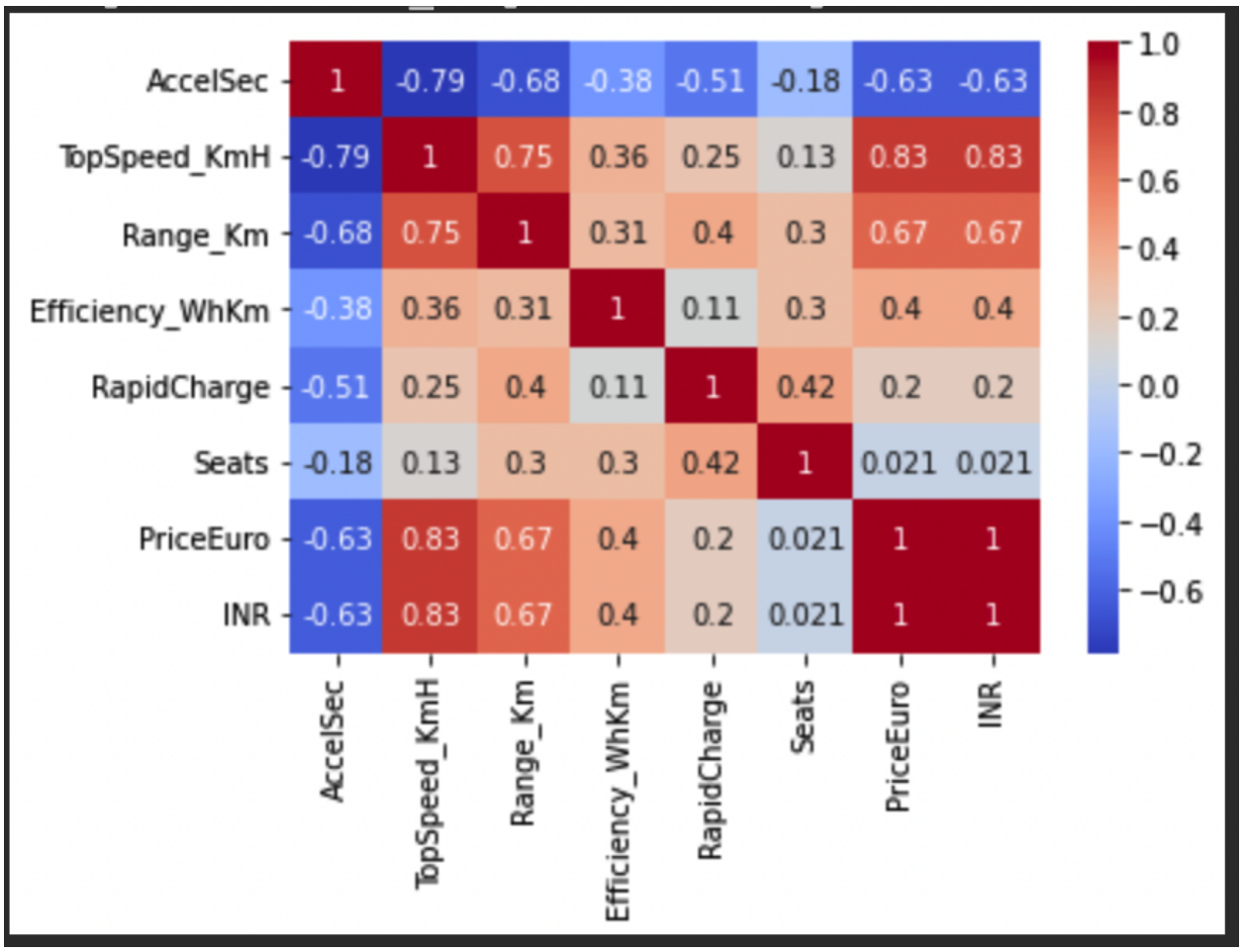
Brands having fastest Speed and acceleration:







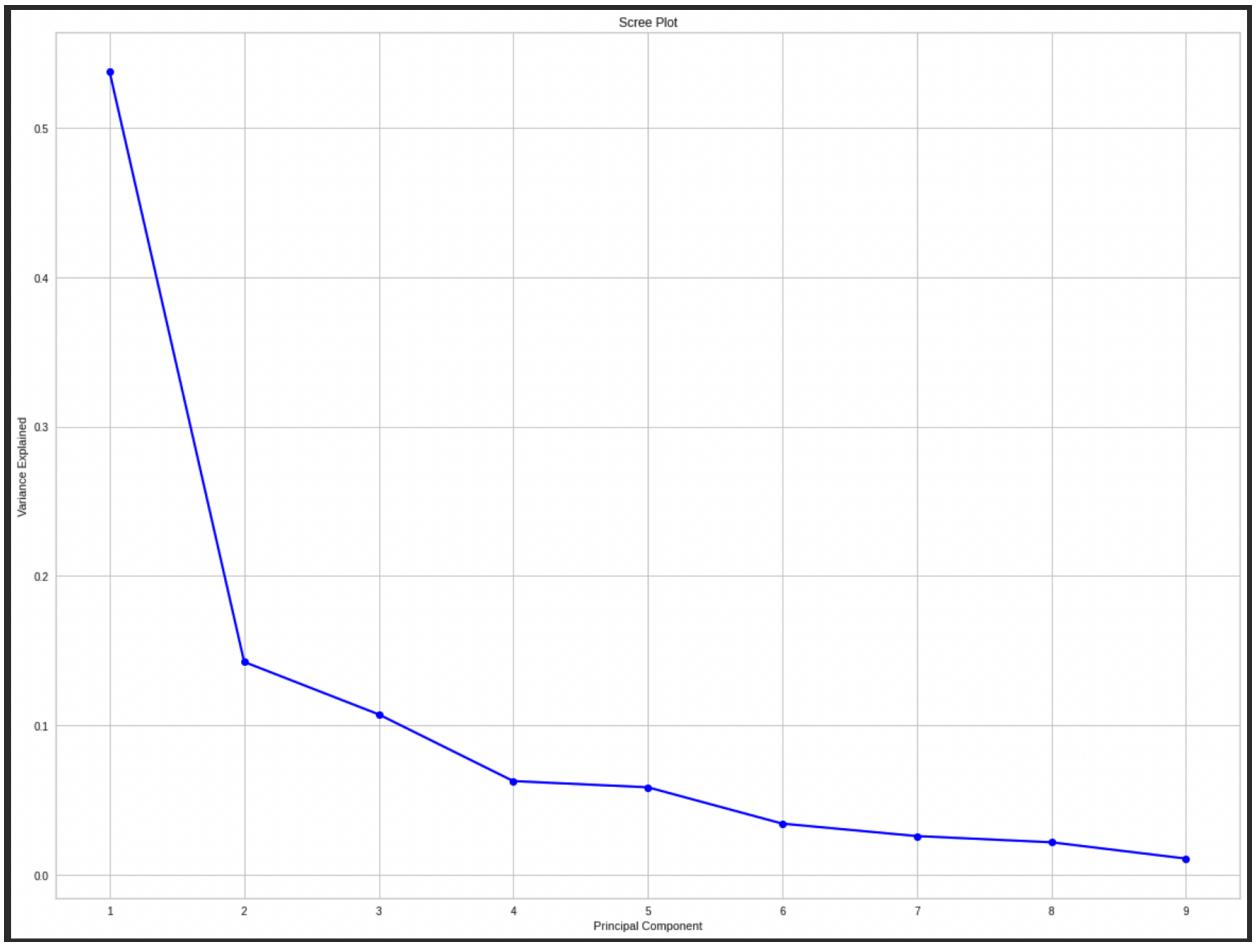
Correlation matrix for dataset



PCA and scree plot:

A Scree Plot helps you determine the number of PCs to retain by displaying the eigenvalues for each PC on a graph. The plot shows the eigenvalues on the y-axis, and the number of factors on the x-axis. The first component generally explains a great deal of variability, while the next few components explain a moderate amount. The final components only explain a small fraction of the overall variability. To use the scree plot criterion, look for the “elbow” in the curve and select all components just before the line flattens out.

The proportion of variance plot: The selected PCs should be able to describe at least 80% of the variance.



Clustering used for segmentation of data set into sub-groups

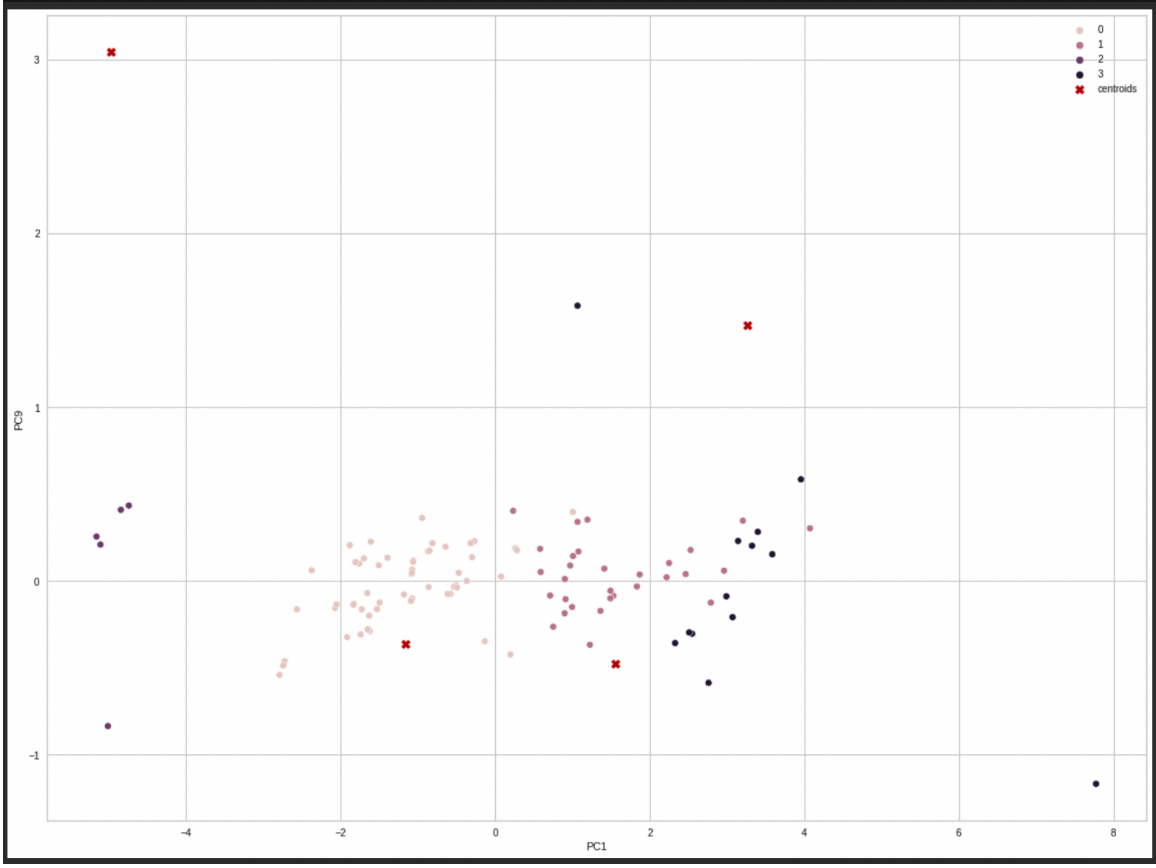
```
#K-means clustering

kmeans = KMeans(n_clusters=4, init='k-means++', random_state=0).fit(t)
df['cluster_num'] = kmeans.labels_
print (kmeans.labels_)
print (kmeans.inertia_)
print(kmeans.n_iter_)
print(kmeans.cluster_centers_)
```

```
[3 0 1 0 0 3 0 0 0 1 1 0 0 0 0 0 3 0 0 0 0 1 0 3 3 0 0 1 0 0 1 0 0 1 0 0 0
 0 0 1 3 0 1 0 0 0 0 3 3 0 1 3 0 0 1 0 0 2 0 3 0 1 0 1 0 3 0 1 2 1 0 1 3 1
 0 0 1 2 1 3 0 1 2 0 1 0 1 1 1 0 1 2 0 1 0 0 0 0 0 1 1 1 1]
354.31005524281215
5
[[-1.16608327e+00 -3.62170519e-01 -4.91076568e-01  1.49737097e-01
  -9.70304598e-02  1.21469554e-01 -4.17088637e-02  1.77314086e-02
  -2.90178110e-02]
 [ 1.55200814e+00 -4.80550170e-01  8.47934234e-01 -1.49972561e-01
  -1.21774934e-01 -1.99516710e-01  1.59031956e-01  3.54881637e-02
   3.75207776e-02]
 [-4.97443663e+00  3.04020235e+00  1.84589535e+00  2.36595708e-01
   1.16330644e+00 -5.35493280e-01  7.53336157e-02  5.64018077e-02
   9.27772433e-02]
 [ 3.26511684e+00  1.47191322e+00 -5.89099425e-01 -3.78410925e-01
   2.44107008e-01  1.52472480e-01 -2.19510712e-01 -1.78606263e-01
   4.97696817e-04]]
```

We have used k-means Clustering here: According to the Elbow method, here we take K=4 clusters to train KMeans model. The derived clusters are shown in the following:

```
#Visulazing clusters
sns.scatterplot(data=data2, x="PC1", y="PC9", hue=kmeans.labels_)
plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1],
            marker="X", c="r", s=80, label="centroids")
plt.legend()
plt.show()
```



Prediction of Prices most used cars:

```
X_train, X_test, y_train, y_test = train_test_split(X, y,test_size=0.4, random_state=101)
lm=LinearRegression().fit(X_train,y_train)
```

```
print(lm.intercept_)
```

```
4454320.851553398
```

```
lm.coef_
```

```
array([ 1007709.35075752,   942692.80890578,   194217.42459894,
        -145469.53935564,   232034.78635588,  1507079.7865495 ,
       -1284714.57184456,  -300556.25007253,  1149218.20693473])
```

```
X_train.columns
```

```
Index(['PC1', 'PC2', 'PC3', 'PC4', 'PC5', 'PC6', 'PC7', 'PC8', 'PC9'], dtype='object')
```

```
cdf=pd.DataFrame(lm.coef_, X.columns, columns=['Coeff'])
cdf
```

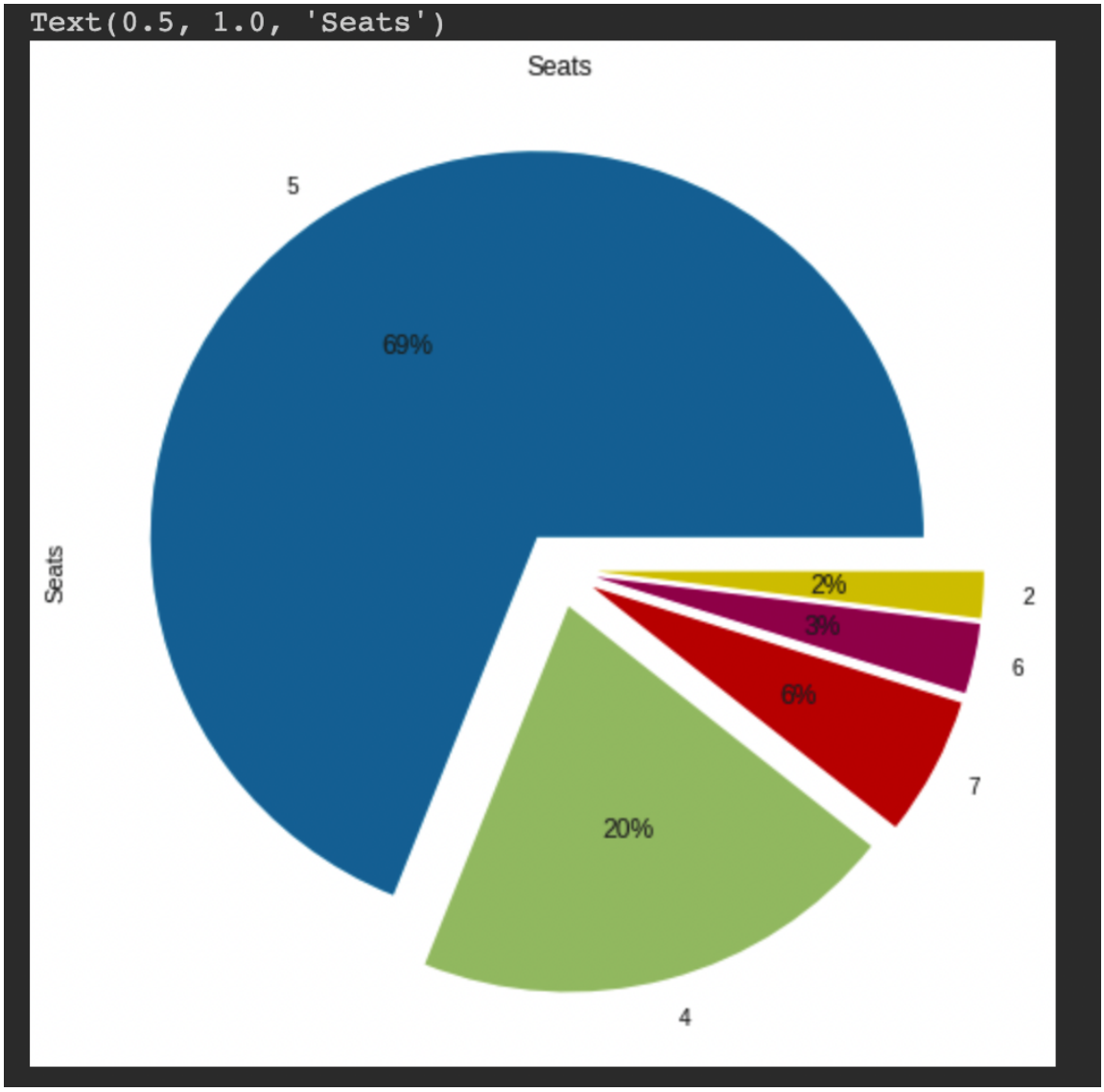
	Coeff
PC1	1.007709e+06
PC2	9.426928e+05
PC3	1.942174e+05
PC4	-1.454695e+05
PC5	2.320348e+05
PC6	1.507080e+06
PC7	-1.284715e+06
PC8	-3.005563e+05
PC9	1.149218e+06

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

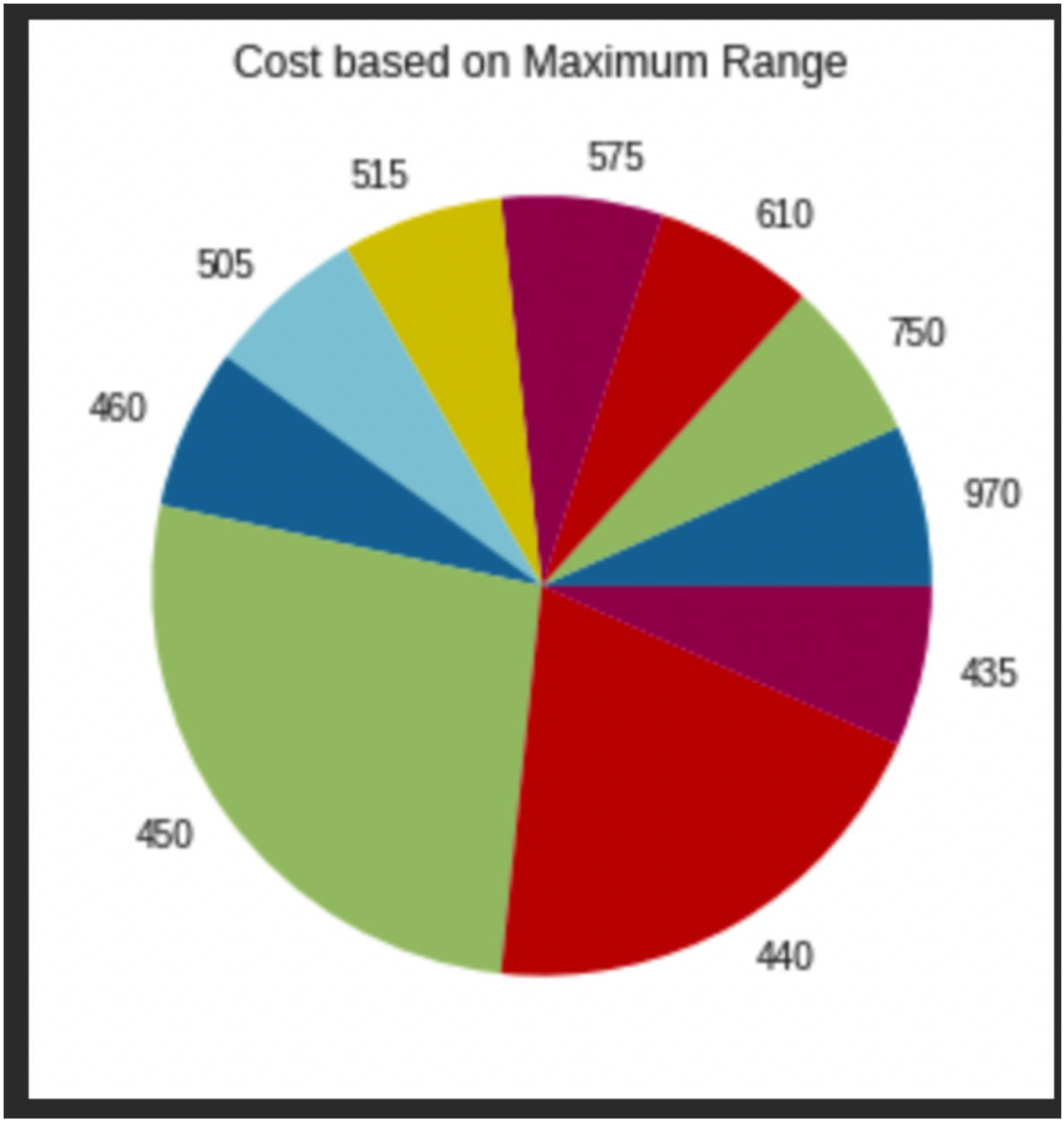
```
array([ 3591450.        ,  2394300.        ,  5020049.        ,
        3111552.47      ,  2939960.97      ,  5237132.2       ,
        2785369.        ,  3192400.        ,  3790975.        ,
        2488795.04      ,  2545939.        ,  3591450.        ,
        1958058.54      , 14428131.61000001,  6340505.45      ,
        3041160.05      ,  4269835.        ,  2750172.79      ,
        3591450.        , 17159150.        ,  4678462.2       ,
        5430272.40000001,  5637060.11000001,  1978489.9       ,
       11891689.99999999,  8216040.45      , 11835902.81000001,
        3192400.        ,  4504476.4       ,  5187650.        ,
        2939960.97      ,  2992875.        ,  2711225.51      ,
        9976250.        ,  2633730.        ,  2793350.        ,
        3024799.        ,  4823476.97      ,  2326142.26      ,
```

Profiling and Describing the Segments:

```
Text(0.5, 1.0, 'Seats')
```



Majority of customers want 5 seats



Most customers prefer a price range of 440-450k while buying cars

Q)What can be the strategic pricing range of products with understanding of early market psychographics?

Since we do not have relevant data from the Indian market which can be analyzed for customer psychographic behaviors, we have taken a sample of data which indicates that the price must be around 15 lacs.

Link to github profile with codes and datasets well documented.

https://github.com/AdityaKendre02/EV-Market-Segmentation