

Market Segment Analysis of EV Market

Github: <https://github.com/riya2025/Electric-Vehicles-Market-Segmentation>

Which ML Algorithm we used?

Step1: Inferences from dataset

Dataset1: https://github.com/Rohit-Rannavre/Feynn-Labs-Internship-2022/blob/main/Project%202.1%3A%20Market%20Segmentation%20of%20Electric%20Vehicles%20in%20India/Datasets/1_ev_charger_dataset.csv

Dataset2: https://github.com/Rohit-Rannavre/Feynn-Labs-Internship-2022/blob/main/Project%202.1%3A%20Market%20Segmentation%20of%20Electric%20Vehicles%20in%20India/Datasets/3_ev_market_india_dataset.xlsx

using simple visualisation methods present in pandas we grabbed some inferences from data.

Inferences:

From dataset 1,

1. Maharastra has highest 2W EV
2. Uttar Pradesh has highest 3W EV
3. Maharastra has highest 4W EV
4. Maharasthra has highest charging spots

From dataset2,

1. Tesla is the most car producing brands
2. SUV is most comfortable body styles of cars
3. Cars produced by tesla gets fast charging
4. Cars manufactured by tesla has highest speed

Step 2 :Clustering Data

Dataset3: <https://www.kaggle.com/datasets/deadprstkrish/ev-cars-user-reviews-india>

This dataset has reviews of electric vehicles users in India for 39 different model cars.

Step -1: Data Preprocessing

1. Filtering Data: Top10 model cars data
2. Removing unnecessary columns
3. Removing null values in numeric columns
4. Transforming string columns to numeric columns

5. Transforming object columns to string columns

Step-2: Converting reviews to word embeddings

1. Do stemming, stop words and special characters removal
2. Use word2vec for word embeddings and do avg word2vec to get one word embedding for each feature

Step-3: PCA

So out 24 columns we need to choose best features for this PCA is used, we got top 7 features here.

Step-4: K Means Clustering

K means as dataset is small (498 rows x 7 columns)

Assuming, 3 cluster labels (positive, negative, neutral)

From electric vehicles reviews from customers, we can which model of car is highly preferable.

Step 3: Predicting Price

Dataset4:

Step1: Data Preprocessing

1. Use price range column in this data which has prices like

0	₹ 13.99 - 17.4 L
1	₹ 12.49 - 13.64 L
2	₹ 17.74 - 19.24 L
3	₹ 21.99 - 25.88 L
4	₹ 23.79 - 23.98 L
5	₹ 1.06 - 1.12 Cr
6	₹ 1.8 Cr
7	₹ 29.15 L
8	₹ 1 Cr
9	₹ 1.16 Cr
10	₹ 1.5 Cr
11	₹ 1.01 - 1.19 Cr

Name: PriceRange, dtype: object

To get one value-> using regex we get only numbers if it's a price range take avg of both

2. Transforming string columns to numeric columns

Step 2: Linear Regression

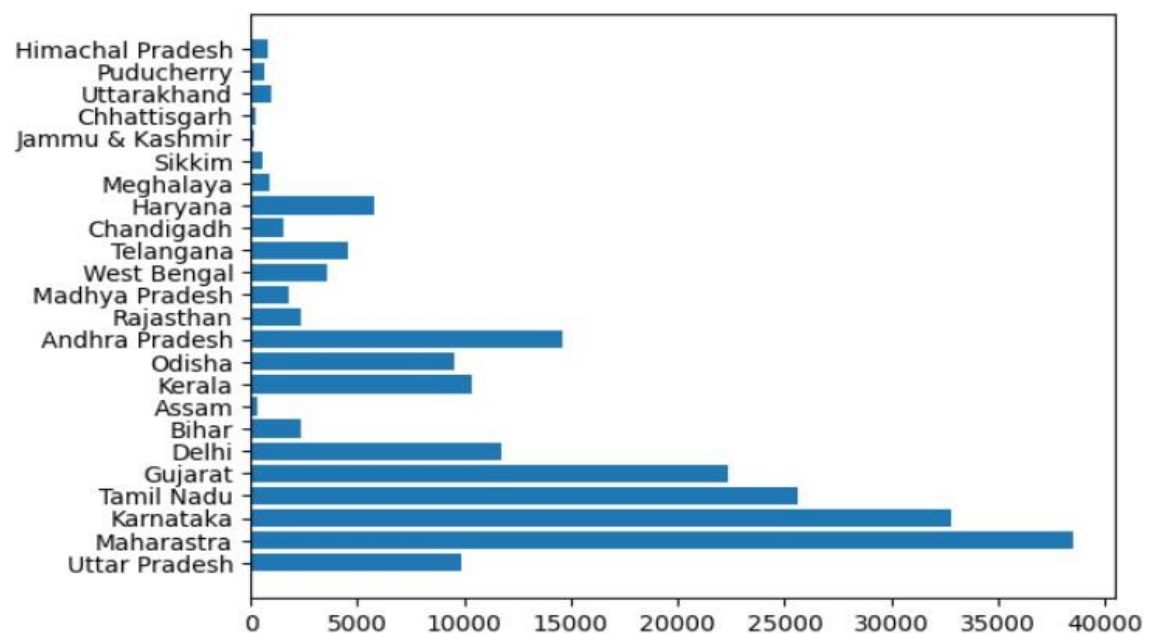
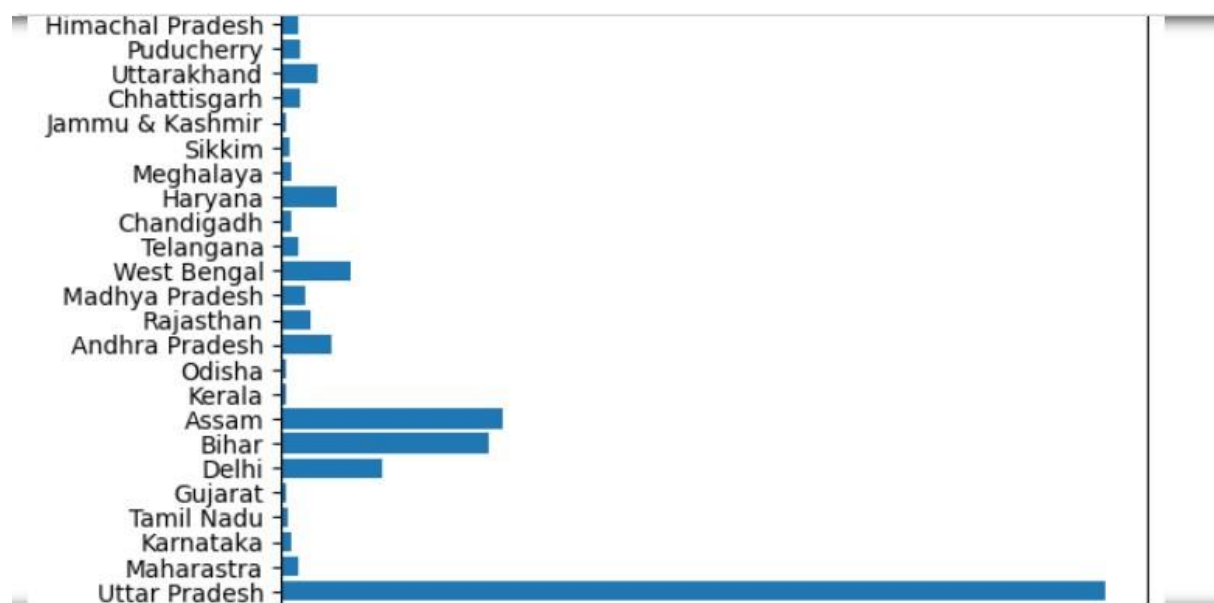
1. Use style column and price range to do linear regression

2. This helps in predicting prices based on Style Name

[14.0175, 13.065, 14.0175, 14.0175, 14.0175, 1.09, 1.8, 29.15, 14.0175, 1.16, 1.5, 14.0175]

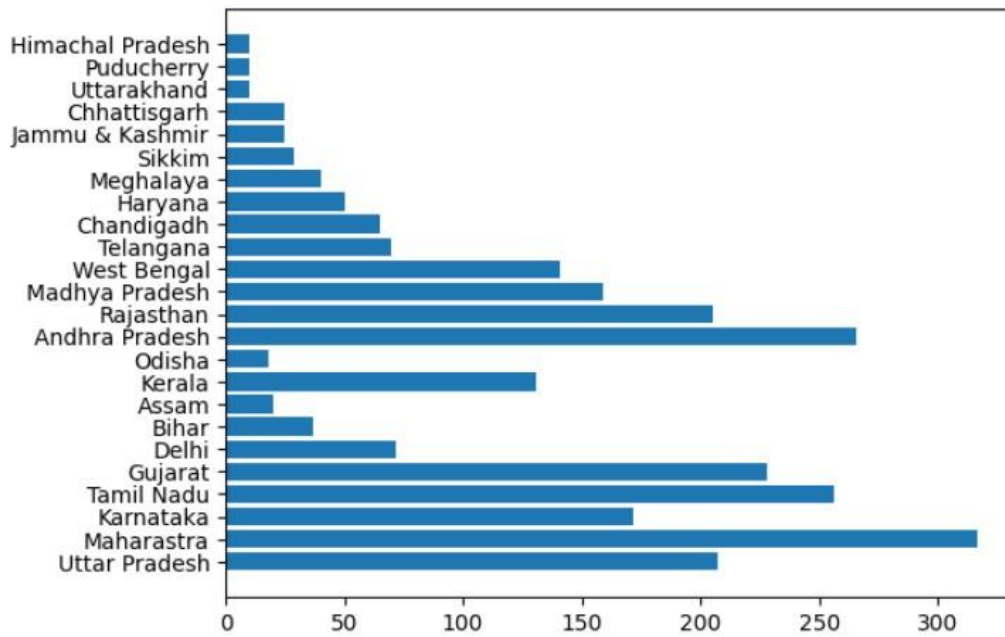
2. Final Conclusion and Insights

Uttar Pradesh has highest 3W EV



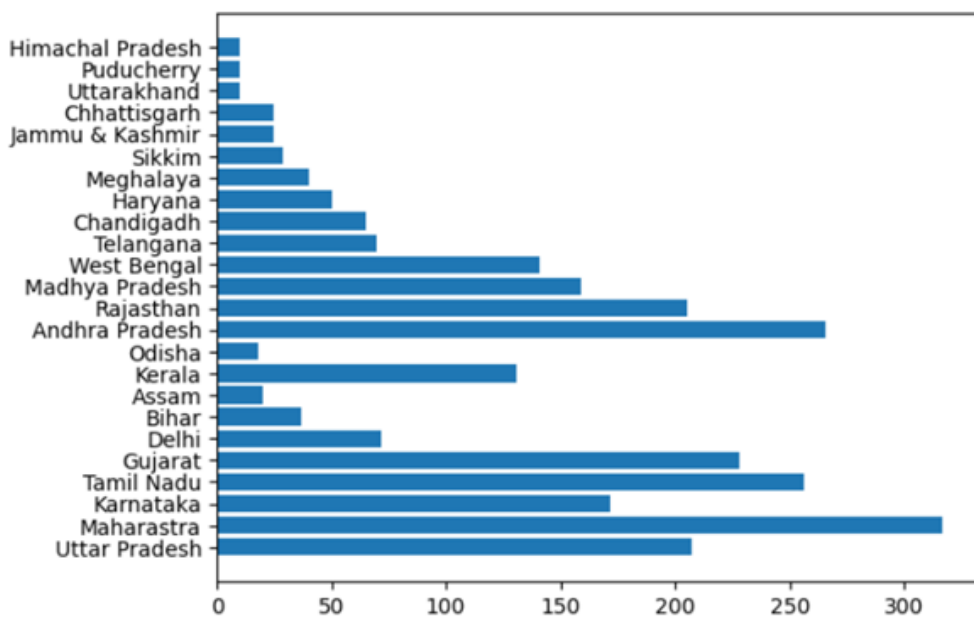
Maharastra has highest 2W EV

Maharastra has highest 4W EV

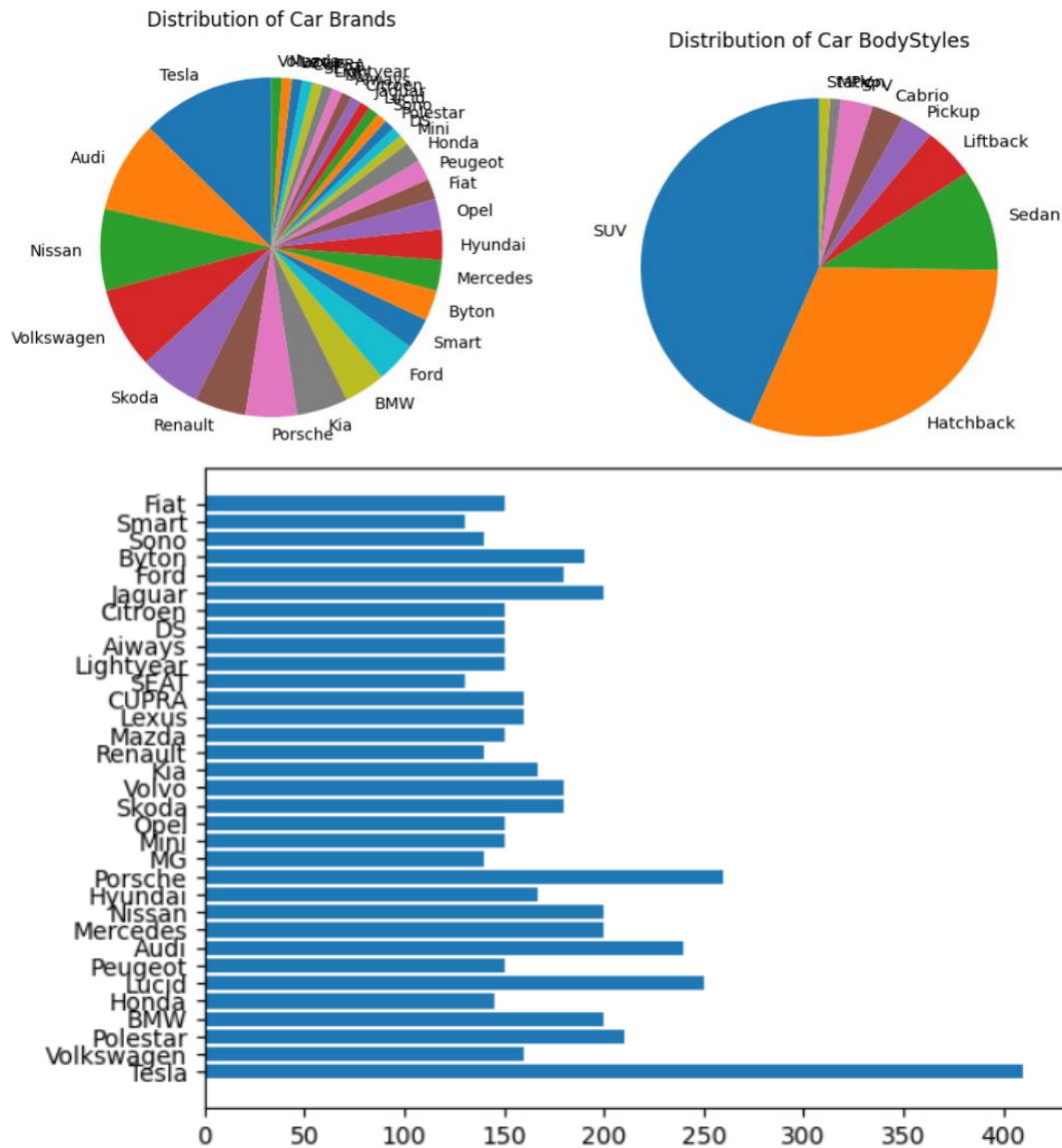


Maharastra has highest charging spots.

By improving some charging spots in Himachal Pradesh will help as 4W users here are high



Tesla is the heighest car producing brand



Cars produced by tesla gets fast charging
SUV is most comfortable body styles of cars

Cluster labels for Revolt RV 400: review cluster_label

17 wanna buy bike gave basic inform compani respo...	2
18 simpli amaz except good price littl bit high c...	2
19 batteri life good like bike tension refuel cit...	1
20 seen bike road nice bike good drive pollut bik...	0
21 rv400 bike year bike gave averag 70 km singl c...	1
22 buy experi revolt 400 bike usual tamulpur assa...	1
23 ll bike meet interest indian factor exclud lik...	2
24 buy experi awesom ride perform great oil bike ...	0
25 valu money thumb bike petrol price gone bike l...	1
26 protect batteri neighbor search batteri stole ...	1

27	s great bike daili use problem go use travel d...	1
28	bad custom servic book revolt 9 month ago rece...	1
29	best bike royal bike feel like young energet c...	1
30	e bike look amaz bike video epic e bike look l...	0
31	price vehicl good s offer abl price feel ride ...	0
32	absolut terribl revolt respond queri wait till...	2
33	awesom bike awesom featur need bike stadium da...	1
34	amaz buy experi revolt team cooper explain fea...	2
35	nice bike electr good look easi charg best per...	1
36	bad experi start book oct 21 3 month wait peri...	2
37	nice bike better ride experi better technolog ...	1
38	t buy revolt s worthless worst buy experi took...	1
39	revolt motor fraud compani provid book decemb ...	2
40	good bike ride comfort bike friendli environ p...	0
41	detail look perform pollut invest valu money q...	1
42	s good eco friendli bike mainten s like ride n...	1
43	revolt bike best budget friendli bike india 20...	1
44	1st total visual appeal awesom ergonom built q...	2
45	1 term power design bike compar 150 cc bike av...	2
46	bike use pollut free featur good like ai app c...	0
47	own revolt bike seen think buy look awesom per...	1
48	electr bike india revolt rv400 base ai easi ch...	2
49	book 5th oct got dhantera nice experi ride lik...	1
50	enjoy bike citi use perfect bike use accord ne...	0
51	batteri life good puck extraordinarili like bi...	1
52	seen video revolt rv 400 sure bring chang mark...	2
53	ridden bike look good dash want bike home town...	2
54	bike kirti nagar delhi industri area drive bik...	2
55	purchas bike 1 week 37 emi s total cost 1 5 la...	1
56	look like good bike dash look mesmer like fz y...	0
57	rv400 good e bike share good bad experi bike p...	2
58	biggest issu go electr rang perform bike medio...	2
59	comfort superb ride mileag long simpl use low ...	1
60	seen featur amaz 7 half year batteri warranti ...	1
61	s amaz bike type servic nice look better bike ...	0
62	easi buy good rid experi nice look easi mainte...	0
63	buy experi great book bike price 1 38lac road ...	1
64	s ponder consid purchas revolt rv 400 terribl ...	2

65	good perform bike electron bike scooti india t...	1
66	sale servic poor part avail revolt showroom se...	2
67	detail look perform best low buy bike best rid...	0
68	bike good citi ride bike lot issu like 1 remot...	1

Above output shows that data points got cluster 1.If we assume 1 as positive then customers gave positive response for Revolt RV 400

3.How To Improve?

1.We need a labelled dataset with reviews columns to do Sentiment analysis and know on which customers are really showing interest.

2. Need a huge dataset for cost predictions in India.

Columns:

Reviews

Price Range

Car Brand

Regions

Body Style

Speed

Charging spots

No. of 3W,2W,4W

ML algorithms:

Large dataset we can use DBSCAN for clustering.

If it is labelled dataset, large dataset-neural networks, random forest classifier.

4.Estimated Market Size

It can be some 50,000 vehicles as from above dataset which tells total number of vehicles used in India we can estimate some 50,000 to 60,000 vehicles.

5.Optimal features?

Columns:

Reviews

Price Range

Car Brand

Regions

Body Style

Speed

-Roopali.B