Electric Vehicle Market Segmentation

Anmol Mani Dubey

22-08-2024

Objective: - Market Segment Analysis of EV Market in India.

Dataset Description: - Dataset 1 - The data_bw dataset comprises 845 rows of data capturing user reviews and ratings for Electric Vehicles (EVs). It includes a mix of textual feedback and numerical ratings, offering insights into various aspects of EVs.

1. **Review**: Contains the written feedback or review provided by the respondent about their experience with the EV. This column provides qualitative insights into user perceptions and satisfaction.
2. **Used it for**: Indicates the primary use of the EV (e.g., daily commute, long-distance travel).
3. **Owned for**: Specifies the duration for which the respondent has owned the EV. This provides context on whether longer ownership affects satisfaction and perceptions.
4. **Ridden for**: Represents the duration the respondent has used the EV.
5. **Rating**: The overall rating given by the respondent to the EV. This is a crucial metric for evaluating the general satisfaction level of the vehicle.
6. **Visual Appeal**: Rating for the visual aesthetics of the EV, assessing how the design and appearance influence user satisfaction.
7. **Reliability**: Rating for the reliability of the EV, reflecting the vehicle's performance consistency and dependability.
8. **Performance**: Rating for the overall performance of the EV, including acceleration, handling, and driving experience.
9. **Service Experience**: Rating for the service experience, including aspects like customer support and service quality.
10. **Extra Features**: Rating for the additional features offered by the EV, such as advanced technology or convenience features.
11. **Maintenance Cost**: Rating for the maintenance costs associated with the EV, reflecting the cost of upkeep and repairs.
12. **Value for Money**: Rating for the overall value of the EV relative to its cost, indicating whether respondents feel the vehicle offers good value for the price paid.

```
[228]: data_bw.head()
```

[228]:

	review	Used it for	Owned for	Ridden for	rating	Visual Appeal	Reliability	Performance	Service Experience	Extra Features	Comfort	Maintenance cost	Value for Money	Model Name
0	We all checked the bike's capacity to be 150 k...	Daily Commute	Never owned	NaN	1	3.0	4.0	NaN	NaN	NaN	4.0	NaN	1.0	TVS iQube
1	Performance is very poor on this bike. The cha...	Everything	> 1 yr	< 5000 kms	1	3.0	1.0	NaN	1.0	NaN	3.0	NaN	3.0	TVS iQube
2	I purchased this in April 2022 and the sales s...	Daily Commute	< 3 months	< 5000 kms	3	4.0	4.0	NaN	2.0	NaN	5.0	NaN	2.0	TVS iQube
3	If any issues come in scooty parts not availab...	Daily Commute	6 months-1 yr	5000-10000 kms	1	1.0	1.0	NaN	1.0	NaN	1.0	NaN	1.0	TVS iQube
4	Don't buy this vehicle unless you have a near ...	Daily Commute	6 months-1 yr	< 5000 kms	1	3.0	4.0	NaN	1.0	NaN	3.0	NaN	2.0	TVS iQube

Dataset 2 - The dataset, evdata_df, contains responses from 540 individuals regarding their demographic, psychographic, and behavioural preferences, with a particular focus on their views about adopting Electric Vehicles (EVs). The dataset features a combination of numerical and categorical variables that capture a wide array of factors influencing EV adoption.

1. **Age**: Type: Numerical
   Description: Represents the age of the respondents. This data helps in understanding age-related trends and preferences in EV adoption.
2. **City**: Type: Categorical
   Description: Indicates the city where the respondent resides. This column provides geographic context for analysing regional preferences and trends.
3. **Profession**: Type: Categorical
   Description: Represents the professional background of the respondents. Understanding professions helps in assessing how occupation impacts attitudes towards EVs.
4. **Marital Status**: Type: Categorical
   Description: Indicates whether the respondent is married or single. Marital status may influence vehicle needs and preferences.
5. **Education**: Type: Categorical
   Description: Represents the highest level of education attained by the respondent. Education level can affect attitudes towards technology and sustainability.
6. **No. of Family Members**: Type: Numerical
   Description: Represents the size of the respondent's family. This data can provide insights into family-related factors affecting vehicle preferences.
7. **Annual Income**: Type: Numerical
   Description: Indicates the respondent's annual income. Income level is crucial for understanding budget constraints and purchasing power for EVs.
8. **Would you prefer replacing all your vehicles with Electric Vehicles?** : Type: Categorical
   Description: Captures the respondent's willingness to switch from existing vehicles to EVs. This data is vital for assessing overall adoption intent.
9. **If Yes/Maybe, what type of EV would you prefer?** : Type: Categorical
   Description: Specifies the preferred type of EV for respondents who are open to or interested in EVs (e.g., two-wheelers, sedans, SUVs).
10. **Do you think Electric Vehicles are economical?** : Type: Categorical
    Description: Reflects the respondent's perception of the cost-effectiveness of EVs. This column provides insight into perceived financial benefits.
11. **Which brand of vehicle do you currently own?** : Type: Categorical
    Description: Indicates the current vehicle brand owned by the respondent. Understanding brand ownership can highlight potential brand loyalty and market opportunities.
12. **How much money could you spend on an Electric Vehicle?** : Type: Categorical

Description: Represents the budget range the respondent is willing to allocate for purchasing an EV. This helps in assessing price sensitivity and potential market segments.

13. **Preference for wheels in EV**: Type: Numerical
    Description: Shows the respondent's preference for the number of wheels in an EV. This data can inform product development and feature offerings.
14. **Do you think Electric Vehicles will replace fuel cars in India?** : Type: Categorical
    Description: Reflects the respondent's opinion on the future potential of EVs replacing traditional fuel vehicles in India. This provides insight into long-term market outlook and acceptance.

```
[131]: evdata_df.head()
```

[131]:

	Age	City	Profession	Marital Status	Education	No. of Family members	Annual Income	Would you prefer replacing all your vehicles to Electronic vehicles?	If Yes/Maybe what type of EV would you prefer?	Do you think Electronic Vehicles are economical?	Which brand of vehicle do you currently own?	How much money could you spend on an Electronic vehicle?	Preference for wheels in EV	Do you think Electronic vehicles will replace fuel cars in India?
0	30	Nabha	NaN	Single	Graduate	5	1.193876e+06	Maybe	SUV	Yes	Hyundai	<5 lakhs	2	I don't think so
1	27	Pune	NaN	Single	Graduate	4	1.844540e+06	Yes	SUV	Yes	Honda	<15 lakhs	4	Yes, in <20years
2	32	Kashipur	NaN	Single	Graduate	4	2.948150e+06	Yes	Hatchback	Yes	KIA	<15 lakhs	4	Yes, in <20years
3	55	Pune	Business	Single	Graduate	3	2.832380e+06	Maybe	Hatchback	No	Hyundai	<5 lakhs	4	Yes, in <10 years
4	26	Satara	NaN	Single	Graduate	4	2.638751e+06	Yes	Sedan	Yes	McLaren	<15 lakhs	4	Yes, in <20years

Libraries used: -

1. **NumPy (numpy)**:

- A fundamental library for numerical computing in Python. It provides support for arrays, matrices, and a wide range of mathematical functions. In the context of your project, NumPy is used for efficient numerical operations and data manipulation.

2. **Pandas (pandas)**:

- A powerful data manipulation and analysis library. Pandas offers data structures like DataFrames and Series for handling and analyzing structured data. It is used for loading, cleaning, and preparing the datasets for analysis.

3. **Matplotlib (matplotlib.pyplot)**:

- A plotting library for creating static, animated, and interactive visualizations in Python. It is used to create various types of plots and charts to visualize data distributions and relationships.

4. **Seaborn (seaborn)**:

- A statistical data visualization library built on top of Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Seaborn is used for creating advanced visualizations and exploring data patterns.

5. **Missingno (missingno)**:

- A library for visualizing missing data in Python. It helps in understanding the patterns and distribution of missing values in the dataset, aiding in data cleaning and preprocessing.

6. **NLTK (nltk)**:

- The Natural Language Toolkit is a library for processing and analyzing human language data (text). It includes tools for text processing, sentiment analysis, and other natural language processing tasks. In your project, it is used for sentiment analysis of textual reviews.

7. **SentimentIntensityAnalyzer (nltk.sentiment)**:
- A specific tool from the NLTK library for performing sentiment analysis. It assesses the sentiment of textual data and is used to analyze the sentiment of customer reviews in your dataset.

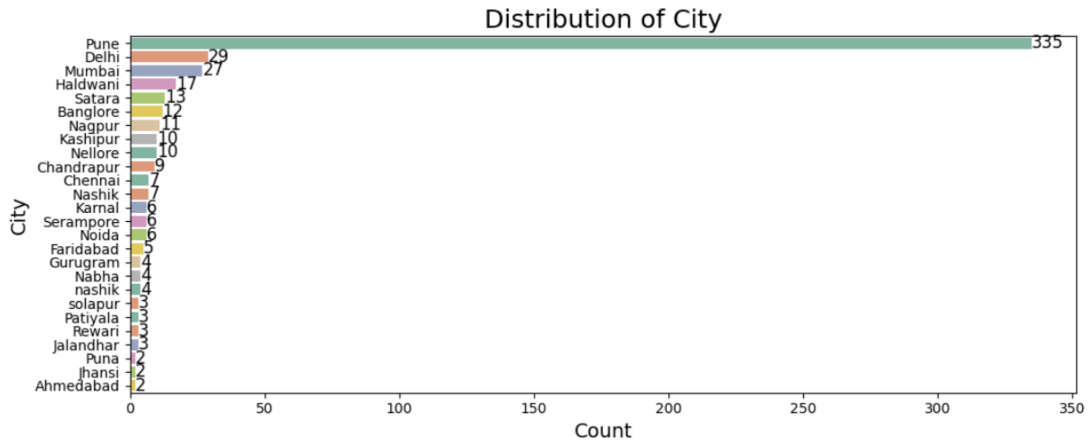
8. **Scikit-learn (sklearn.preprocessing, sklearn.decomposition, sklearn.cluster)**:
- A machine learning library that provides simple and efficient tools for data mining and data analysis. It includes modules for preprocessing data, dimensionality reduction, and clustering.
- **StandardScaler (sklearn.preprocessing.StandardScaler)**: Used to standardize features by removing the mean and scaling to unit variance.
- **PCA (sklearn.decomposition.PCA)**: Used for Principal Component Analysis to reduce the dimensionality of the data while preserving as much variance as possible.
- **KMeans (sklearn.cluster.KMeans)**: A clustering algorithm used to partition the dataset into distinct groups based on feature similarity.

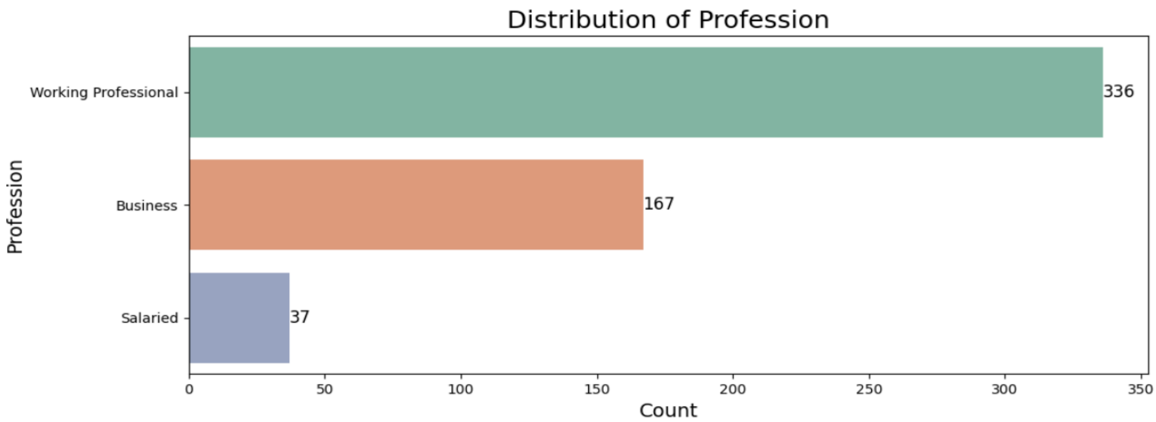
9. **Warnings (warnings)**: A standard library used to manage warnings in Python. It is used to filter out warnings during the execution of the code to ensure that only relevant messages are displayed.

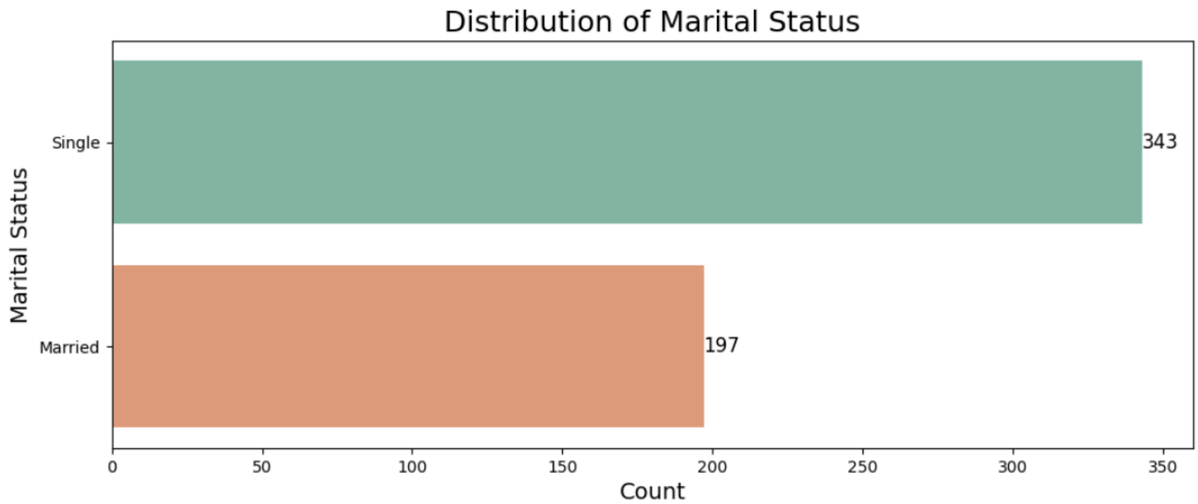
Data Exploration

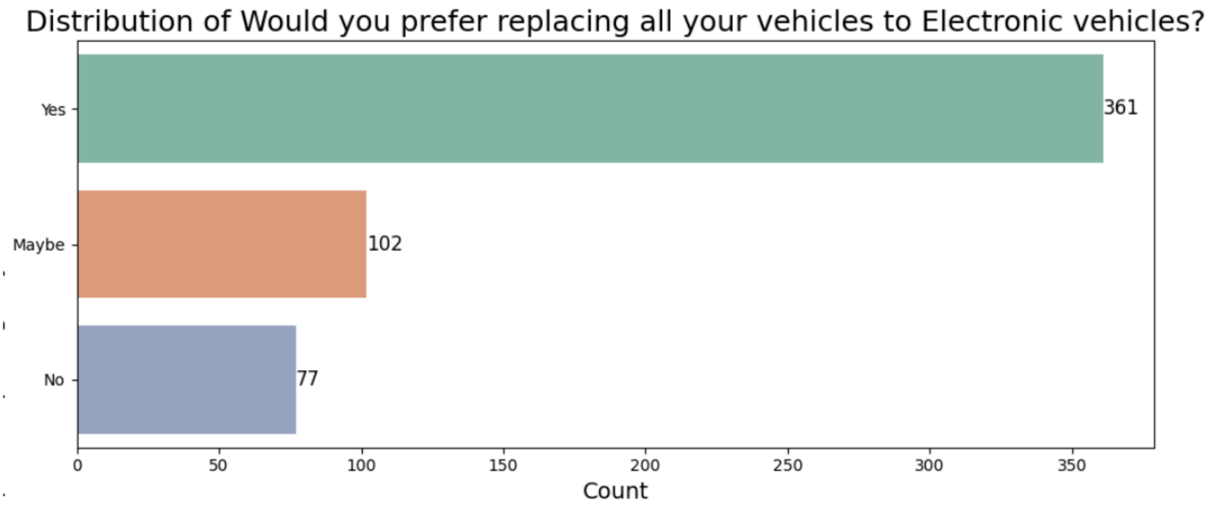
```
[133]: evdata_df.keys()
```

```
[133]: Index(['Age', 'City', 'Profession', 'Marital Status', 'Education',
              'No. of Family members', 'Annual Income',
              'Would you prefer replacing all your vehicles to Electronic vehicles?',
              'If Yes/Maybe what type of  EV would you prefer?',
              'Do you think Electronic Vehicles are economical?',
              'Which brand of vehicle do you currently own?',
              'How much money could you spend on an Electronic vehicle?',
              'Preference for wheels in EV',
              'Do you think Electronic vehicles will replace fuel cars in India?'],
             dtype='object')
```

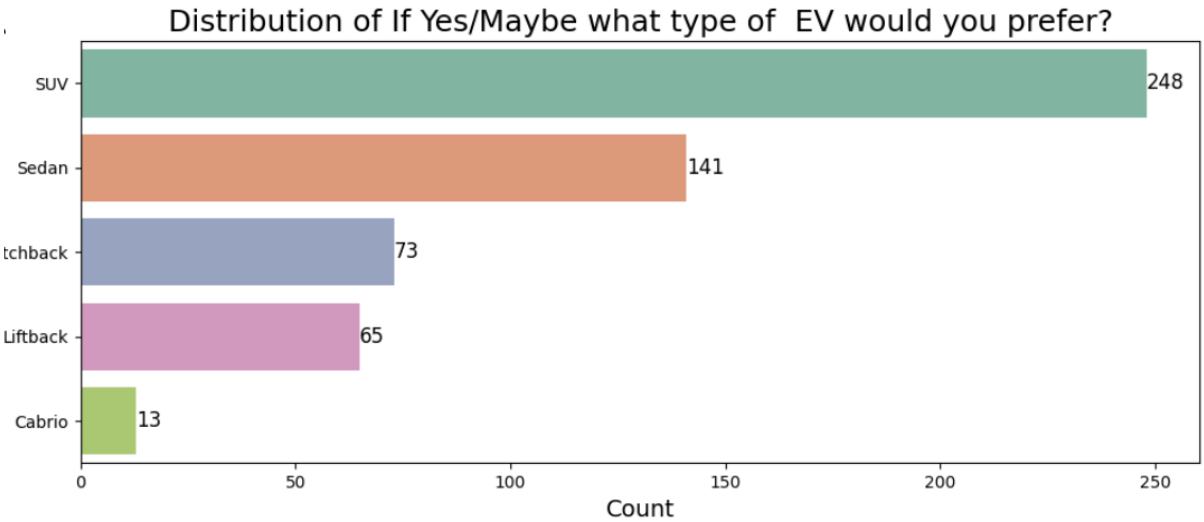




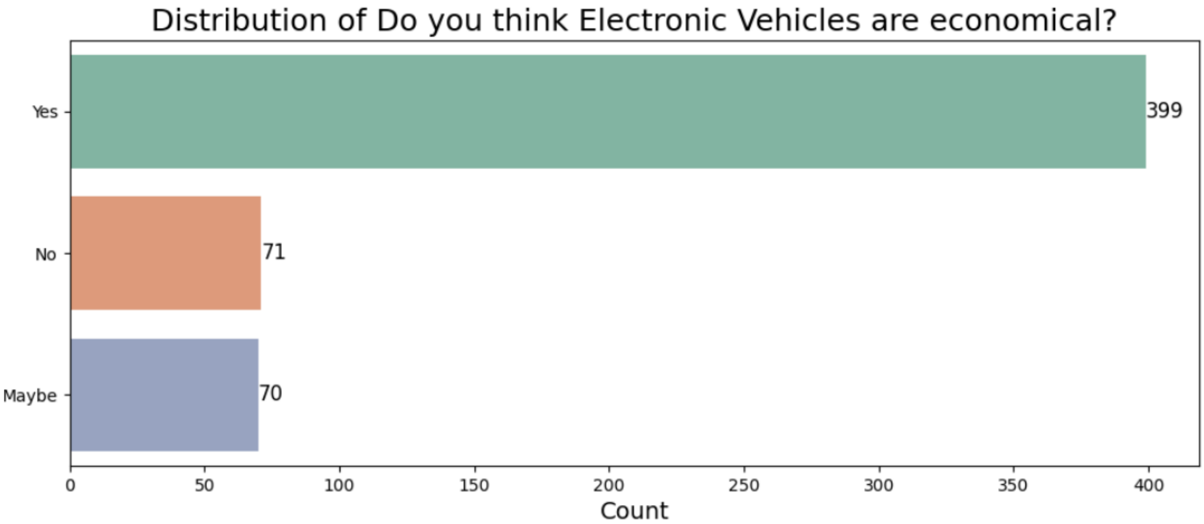




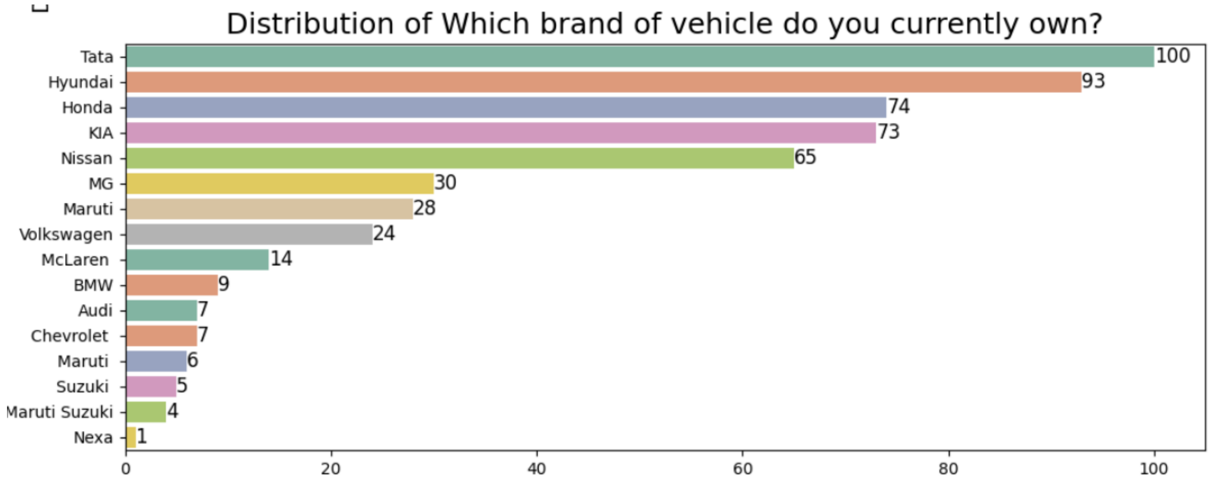
Distribution of Would you prefer replacing all your vehicles to Electronic vehicles?
Yes
Maybe
No
361
102
77
0
50
100
150
200
250
300
350
Count

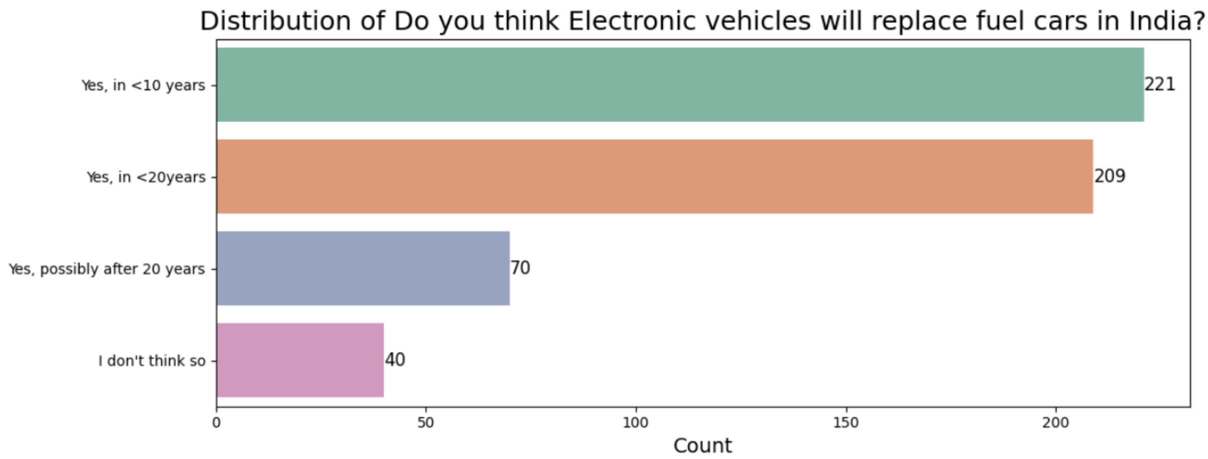


Distribution of If Yes/Maybe what type of EV would you prefer?
SUV
Sedan
tchback
Liftback
Cabrio
248
141
73
65
13
0
50
100
150
200
250
Count



Distribution of Do you think Electronic Vehicles are economical?
Yes
No
Maybe
399
71
70
0
50
100
150
200
250
300
350
400
Count

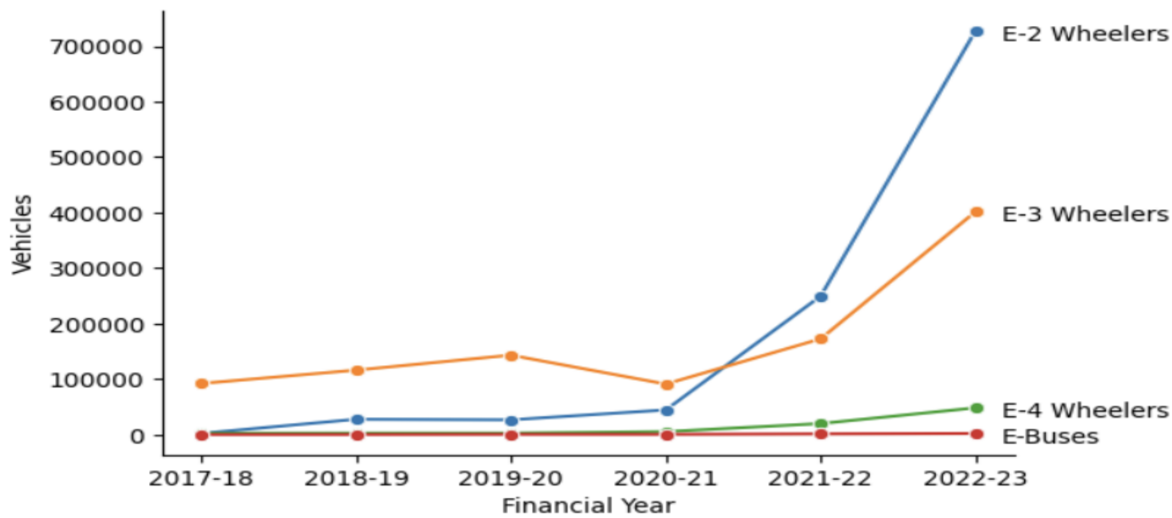


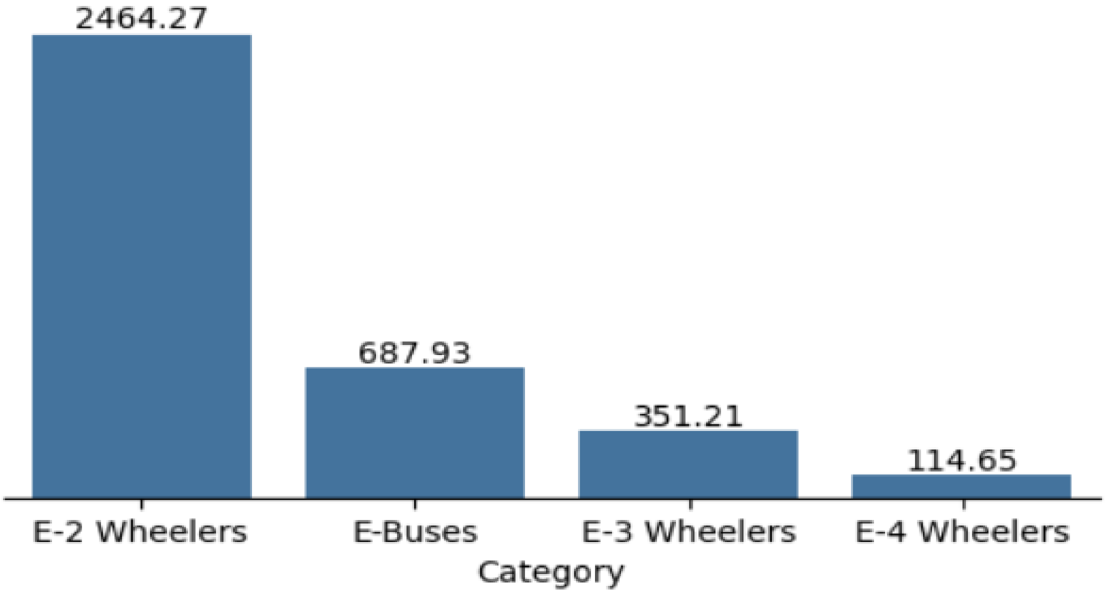



```
data_smev.keys()
```

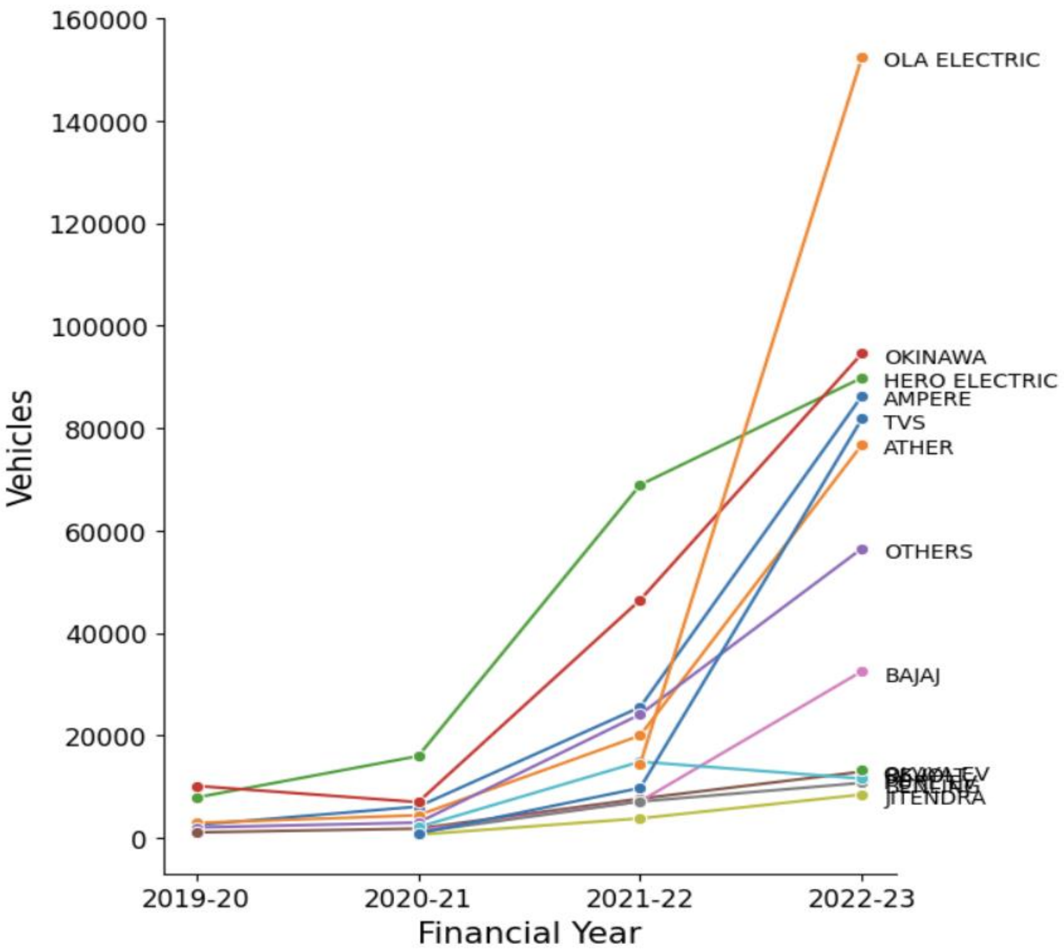
```
dict_keys(['EV 2W FY 19-20', 'EV 2W FY 20-21', 'EV 2W FY 21-22', 'EV 2W FY 22-23', 'EV Market', 'Electric Bus Sales', 'Electric 4-Wheeler Sales', 'Electr
ic 3-Wheeler Sales', 'Electric 2-Wheeler Sales', 'EV Industries'])
```

Dataset 3 – data_smev Consists data of number of electric vehicles produced by each manufacturer in respective year. This dataset enables the visualization of growth in ev manufacturing from FY 17 – 23.





Above Figure delved into the market's financial perspective, representing the industry's total value in crores. Notably, two-wheelers emerged as the primary revenue generators, highlighting their economic significance.



Above Figure honed in on specific electric two-wheeler companies, with Ola Electric emerging as the market leader in 2023, illustrating industry leadership and market competitiveness.

EV Market Segmentation

Principal Component Analysis (PCA):

[169]:
```
pca = PCA(random_state = 42)
pca.fit(data_scaled)
```

[169]:
```
PCA
PCA(random_state=42)
```

[170]:
```
data_pca = pca.transform(data_scaled)
```

[171]:
```
df_pca = pd.DataFrame(data_pca, columns = [f'PC{x +1}' for x in range(len(data_segment.columns))])
```

[172]:
```
df_pca.head()
```

[172]:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
0	-0.291227	-1.038055	0.354864	0.623469	1.102720	0.169287	0.438012	-0.957827
1	-0.710801	-1.394405	-0.360466	0.621671	-0.320899	0.086053	0.426279	0.149917
2	0.849149	-1.189765	0.167683	0.410898	0.409054	0.191904	0.329993	-0.830738
3	-1.967022	-0.878935	-0.100197	-0.330003	0.075822	-0.069599	-0.013068	-0.011328
4	-0.078940	-1.017161	-0.079210	0.324132	0.961554	-0.005517	0.107260	-0.427972

[173]:
```
pca_summary = pd.DataFrame({'Standard Deviation':df_pca.std(),
                            'Proportion of Variance': pca.explained_variance_ratio_,
                            'Cumultive Proportion': pca.explained_variance_ratio_.cumsum()})
```

[174]:
```
pca_summary
```

[174]:

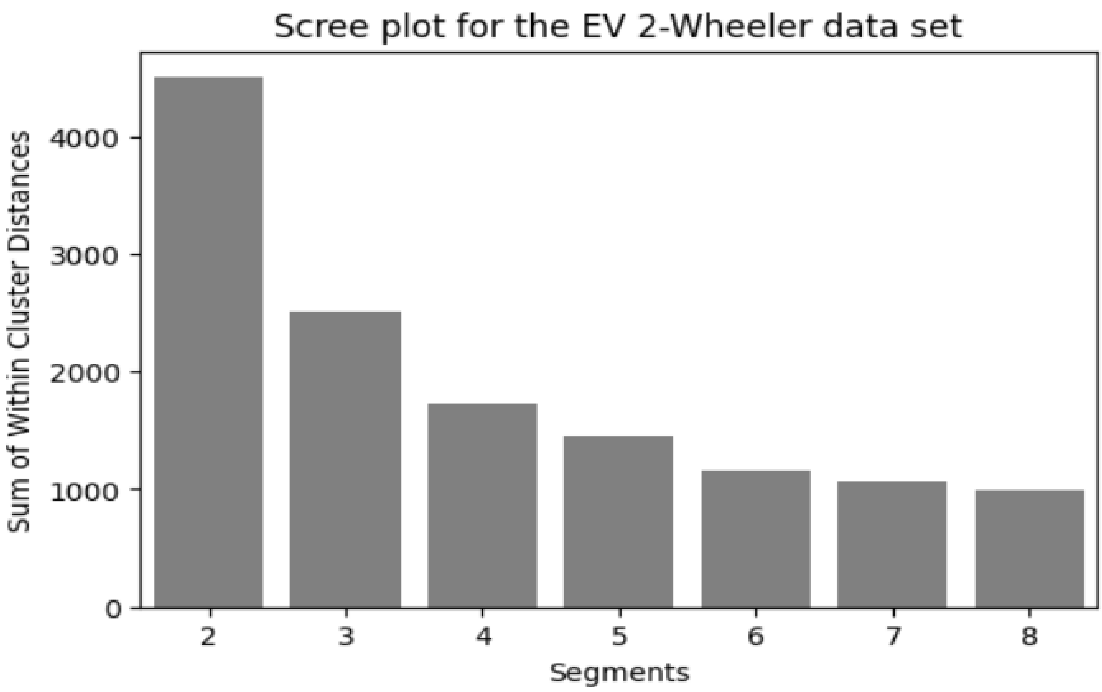
	Standard Deviation	Proportion of Variance	Cumultive Proportion
PC1	1.845017	0.425007	0.425007
PC2	1.736646	0.376546	0.801553
PC3	0.903486	0.101915	0.903468
PC4	0.517750	0.033468	0.936936
PC5	0.405239	0.020503	0.957439
PC6	0.379558	0.017987	0.975426
PC7	0.337883	0.014254	0.989680
PC8	0.287510	0.010320	1.000000

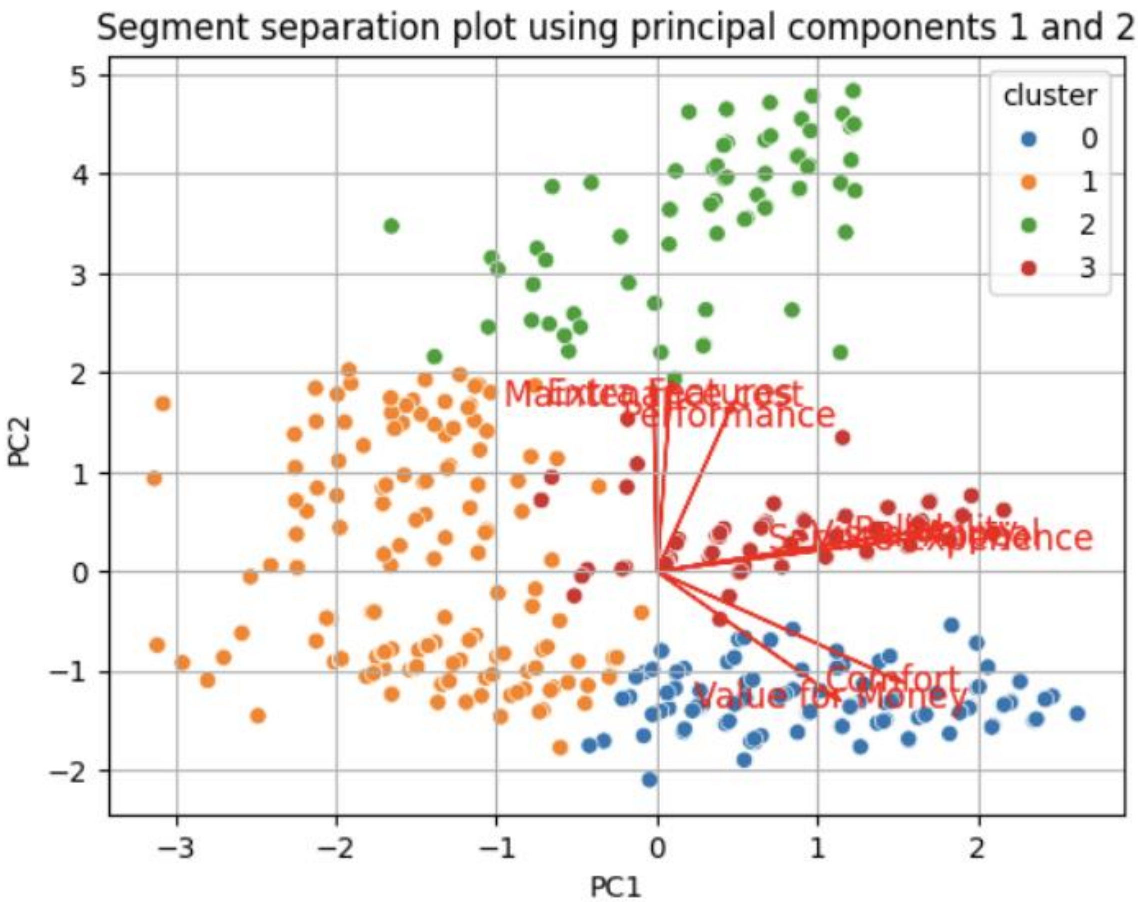
```
[175]: factor_loadings = pd.DataFrame(pca.components_, columns = data_segment.columns, index = df_pca.columns).T
```

```
[176]: factor_loadings.style.background_gradient(cmap = 'Blues')
```

[176]:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Visual Appeal	0.480170	0.117814	0.063320	0.730598	-0.247014	-0.105903	-0.375474	-0.067539
Reliability	0.494758	0.124910	-0.002776	-0.152447	0.819319	-0.060484	-0.117211	-0.166384
Performance	0.128721	0.459145	0.574833	0.005549	0.019902	0.025704	0.288468	0.598232
Service Experience	0.486499	0.100691	-0.054176	-0.653781	-0.470391	-0.052432	-0.311210	0.044129
Extra Features	0.024373	0.519633	-0.364578	0.023208	-0.116821	-0.559390	0.456829	-0.246323
Comfort	0.418255	-0.304266	0.249807	0.020111	-0.172621	0.296656	0.623271	-0.404238
Maintenance cost	-0.005912	0.513208	-0.386495	0.054822	-0.020302	0.762039	0.003360	-0.055435
Value for Money	0.309572	-0.351548	-0.563840	0.107598	0.046688	-0.009572	0.260855	0.617065





Above Figure, utilizing principal components, further emphasizes the differences among segments. Notably, Segment 1, despite being the largest segment, lacks specific opinions, making them unique in their lack of satisfaction.

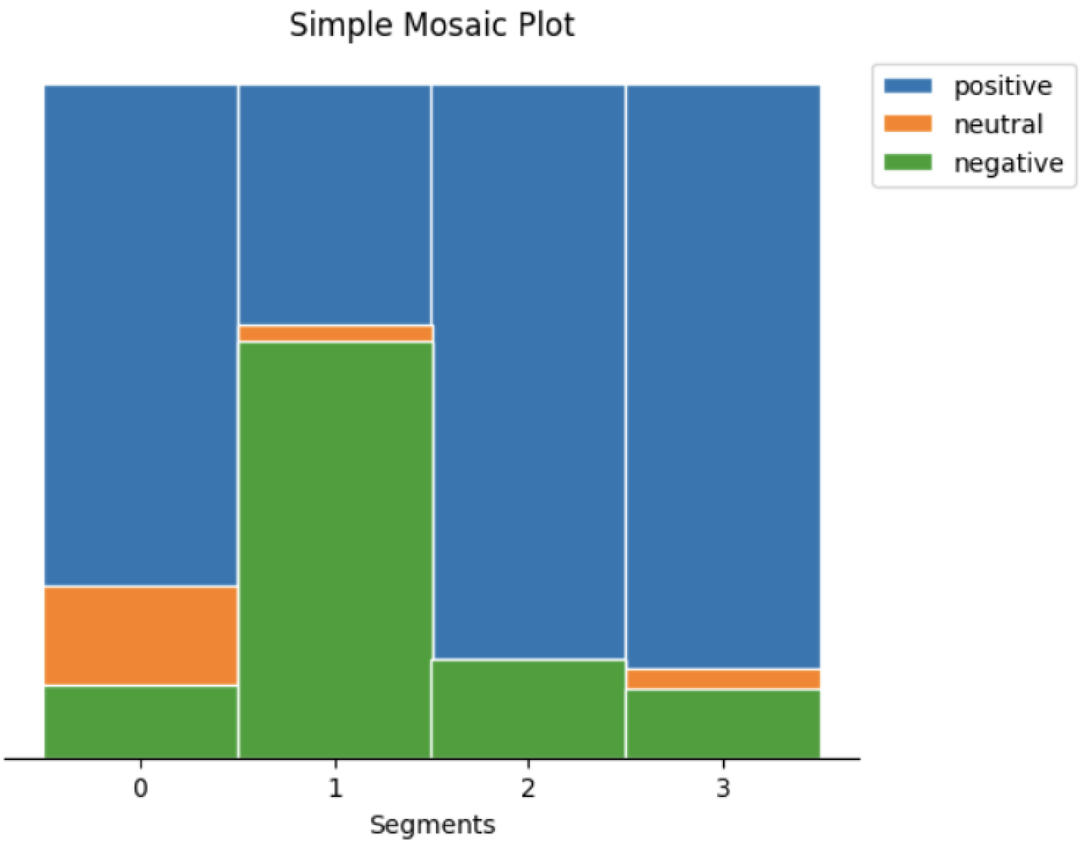
Describing Segments

```
data_desc.columns
```

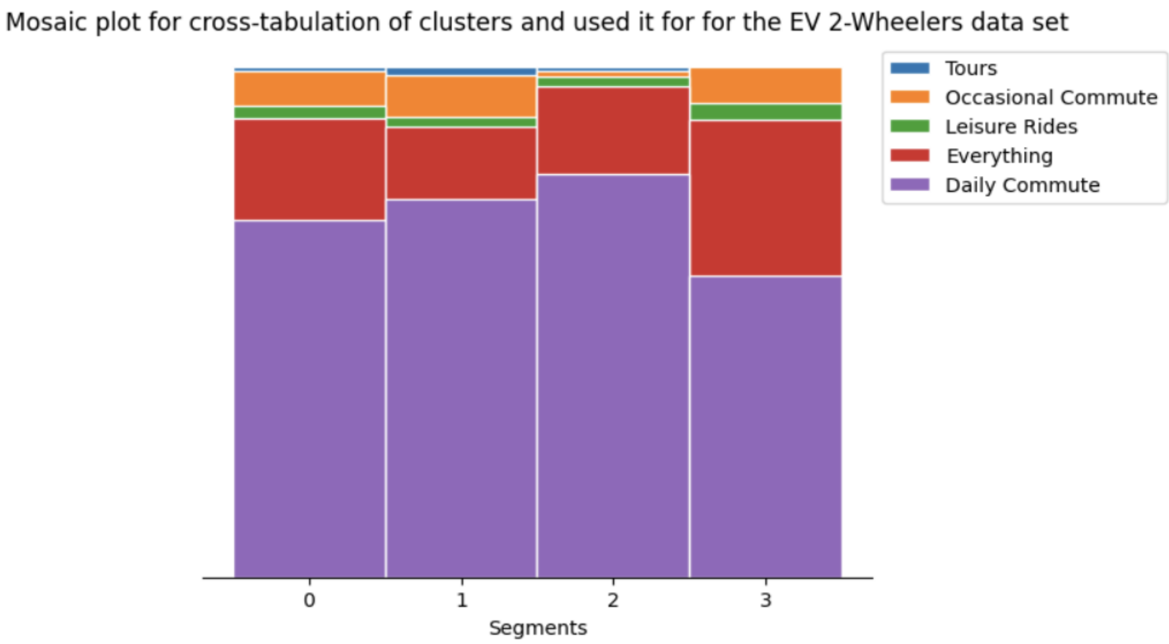
```
Index(['review', 'Used it for', 'Owned for', 'Ridden for', 'rating',
       'Visual Appeal', 'Reliability', 'Performance', 'Service Experience',
       'Extra Features', 'Comfort', 'Maintenance cost', 'Value for Money',
       'Model Name', 'Price', 'Riding Range (km)', 'Top Speed (kmph)',
       'Weight (kg)', 'Battery Charging Time (hrs)', 'Rated Power (W)',
       'sentiment', 'cluster'],
      dtype='object')
```

Parameters in dataset

Simple Mosaic Plot of reviews performing sentiment analysis

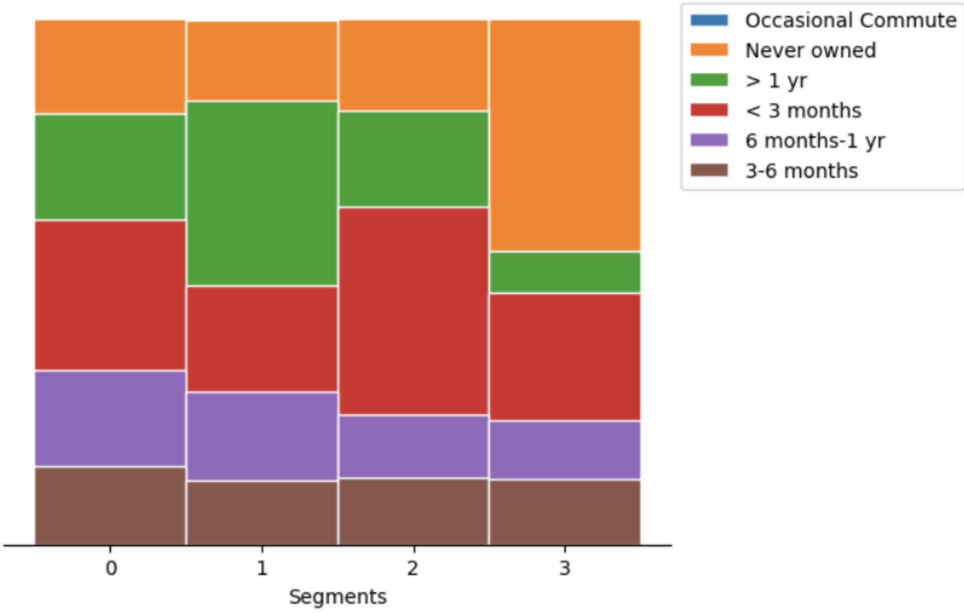


Above mosaic plot, explores consumer sentiments, revealing that all segments, except Segment 1, exhibit positive sentiments. Segment 1 consumers stand out with negative sentiments, indicating dissatisfaction across various aspects



Above mosaic plot illustrates that all segments predominantly use electric vehicles for daily commuting, with limited usage for tours, occasional commuting, and leisure rides.

Mosaic plot for cross-tabulation of clusters and owned for for the EV 2-Wheelers data set



This mosaic plot delineates the ownership duration of electric vehicles among segments. Segment 1 stands out, owning electric vehicles for more than a year, while Segment 0 has no prior ownership experience. Segment 2 members moderately own vehicles ranging from less than 3 months to over a year, and Segment 3 consumers have owned electric vehicles for a few days to less than 3 months.

Clustering

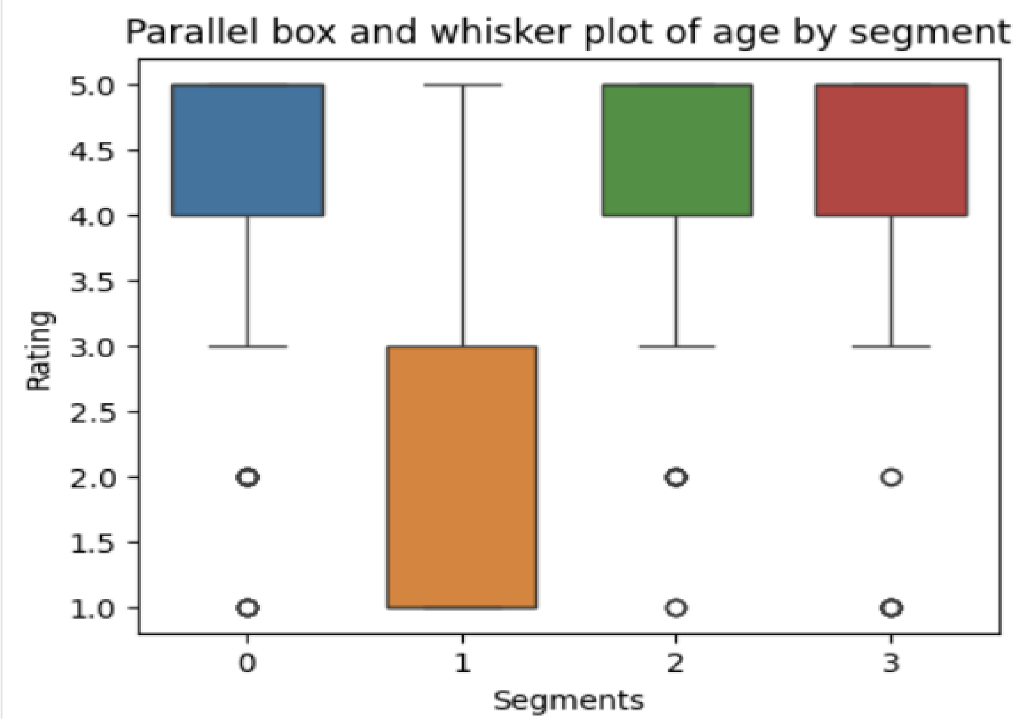
```
[209]: ridden_cluster = pd.crosstab(index = data_desc['cluster'], columns = data_desc['Ridden for'])
```

```
[210]: ridden_cluster
```

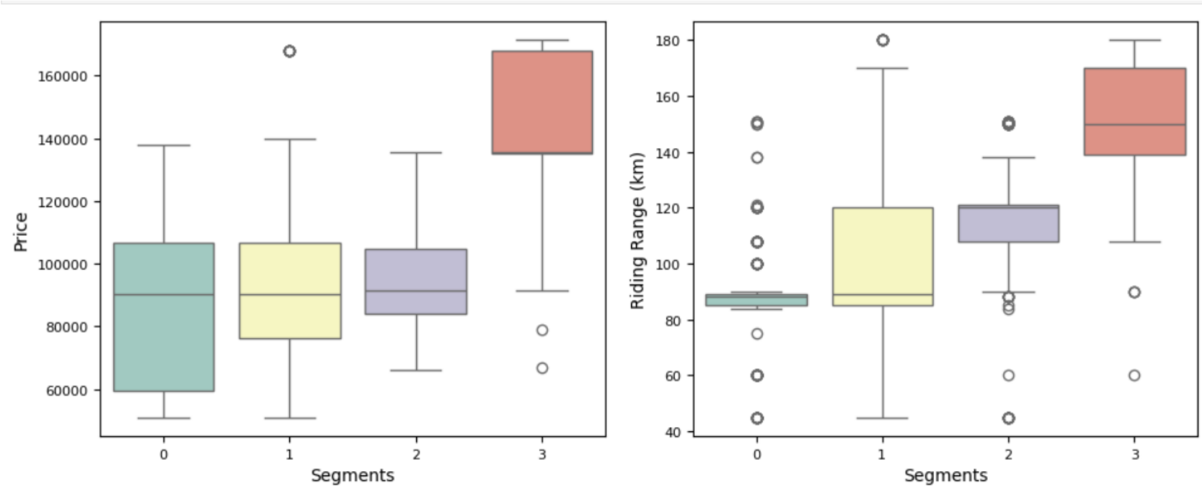
[210]:

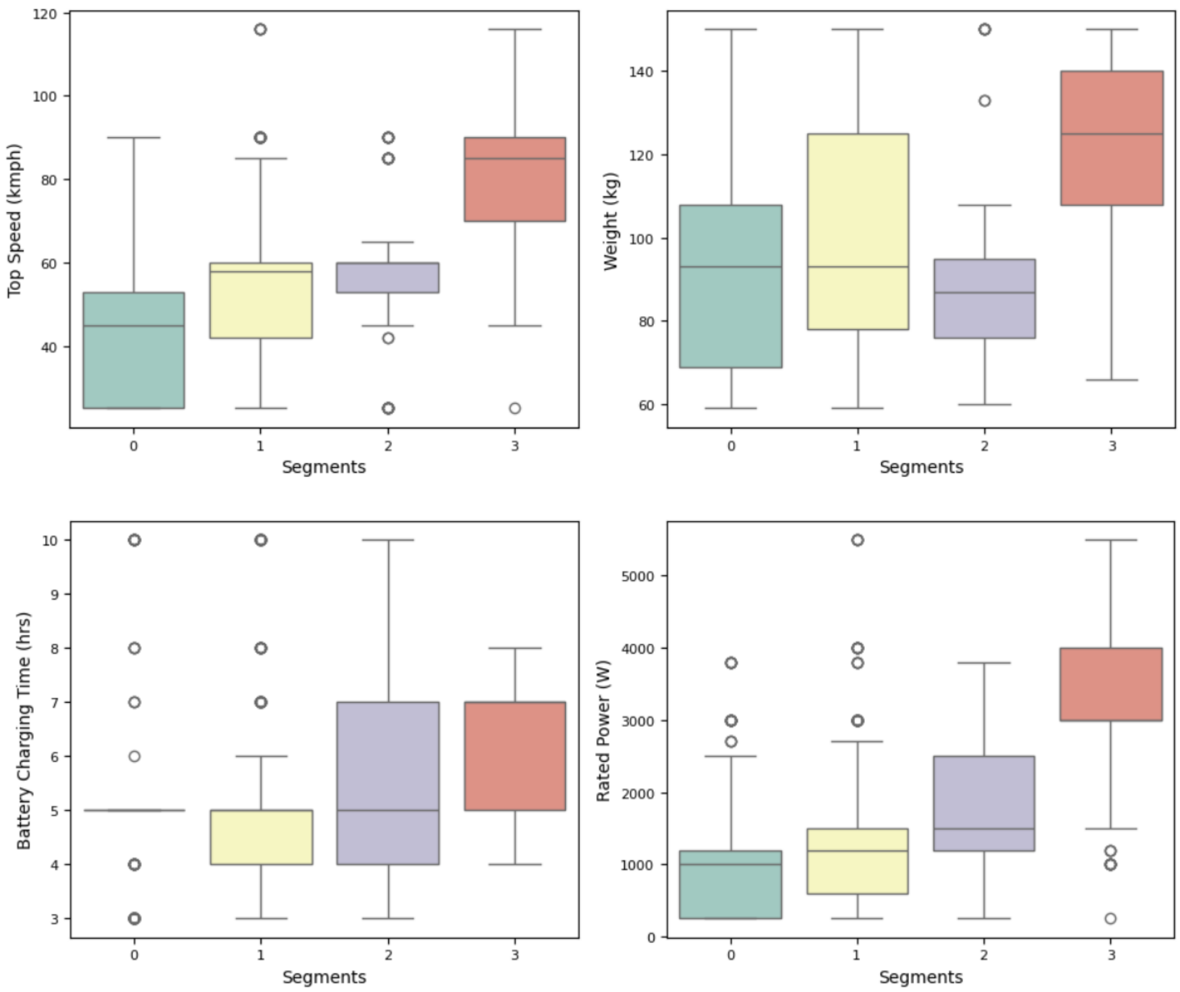
Ridden for	10000-15000 kms	5000-10000 kms	< 5000 kms	> 15000 kms	Never owned
cluster					
0	11	44	158	14	0
1	15	61	185	18	1
2	5	18	62	5	0
3	1	15	54	1	0

The distances covered by consumers, indicating that all segments predominantly use electric vehicles for commuting, with most users covering distances below 5000 kms. A small portion falls in the 5000 to 10000 kms range, aligning with their commuting needs.



Above parallel box and whisker plot, emphasizes significant differences in average ratings among segments. Specifically, Segment 1 consumers express dissatisfaction across all perceptions, leading to lower overall ratings.





In analysing technical specification of electric vehicles across different segments, distinct patterns emerge.

Segment 0 prefers premium EVs with a higher price range and extended riding range, emphasizing consumer preference for luxury and long-distance travel.

Segment 1 focuses on budget-friendly options with lower prices and moderate riding ranges, suitable for daily commuting.

Segment 2 and Segment 3 prioritize affordability, with slight differences in riding range and speed preferences. Weight preferences vary, with Segment 0 and Segment 1 favouring heavier vehicles, while Segment 2 and Segment 3 prefer lighter options.

Charging time also differs, with Segment 0 and Segment 3 opting for longer durations for overnight charging, while Segment 1 and Segment 2 prioritize faster charging for quick turnaround times.

These preferences shape the electric vehicle market in India.

Selection of Target Segment

The strategic target segments for the electric vehicle market are identified as Segment 1 (39% of consumers) and Segment 2 (33% of consumers). Segment 1's diverse preferences and dissatisfaction points present an opportunity for improving customer satisfaction and loyalty by directly addressing their specific demands. Segment 2 values visual appeal, reliability, service experience, and comfort, offering a chance to customize electric vehicles to meet these expectations and emphasize value for money. The strategy involves addressing dissatisfaction points in Segment 1 and enhancing positive elements in Segment 2, aligning electric vehicles with the distinct expectations of each segment to ensure competitive advantage and sustained market growth.

Conclusion

In conclusion, our analysis of India's electric vehicle market identified Segment 1, which constitutes 39% of the consumer base, as the most promising target. By aligning our electric two-wheeler specifications with the preferences of this segment, we position our products to meet substantial market demand.

This strategic choice is grounded in a detailed market segmentation analysis, reflecting a deep understanding of consumer behaviour and technical needs. By focusing on this segment, we ensure our market entry is precise and relevant, laying a strong foundation for success in India's evolving electric vehicle landscape.