Market Analysis of Electric Vehicles in India

[Project link]

Dataset-1 ([Link](#)): Analysing the monthly sales data of Electric Vehicles

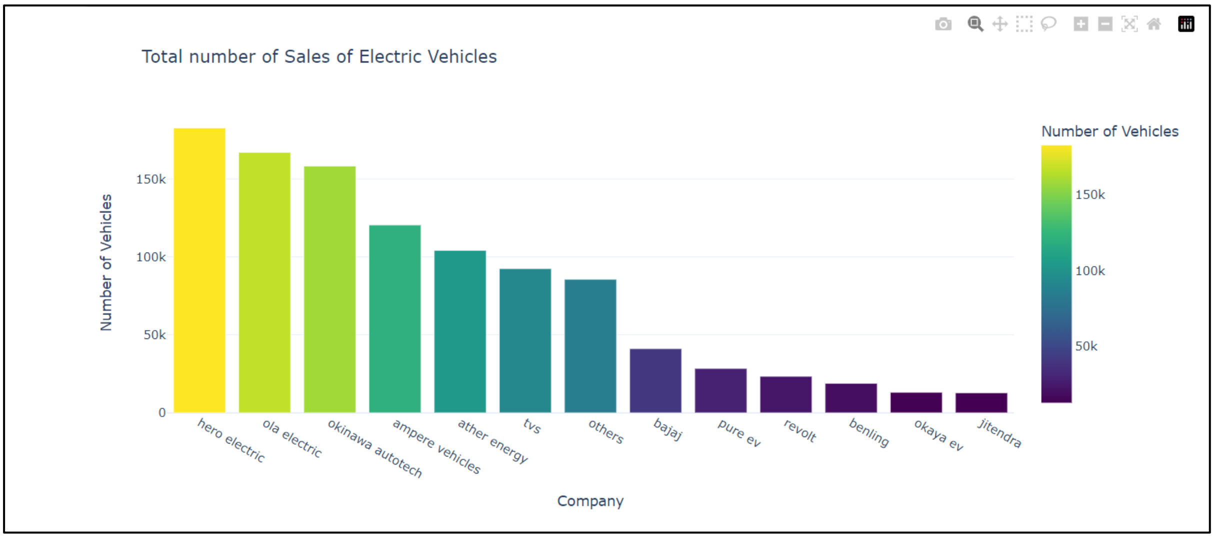
```
df4 = pd.read_excel('smev_data.xlsx', sheet_name='EV 2W FY 22-23')
df4['maker'] = df4['maker'].str.lower()
# Standardizing the name format of vehicles in df4 correspoding to other dataframes.
df4['maker'] = df4['maker'].replace({'ampere':'ampere vehicles','ather':'ather energy','okinawa':'okinawa autotech','jitendra new

monthly_ev = pd.concat([df1, df2, df3,df4], ignore_index=True)
monthly_ev.sample(10)
```

114]:

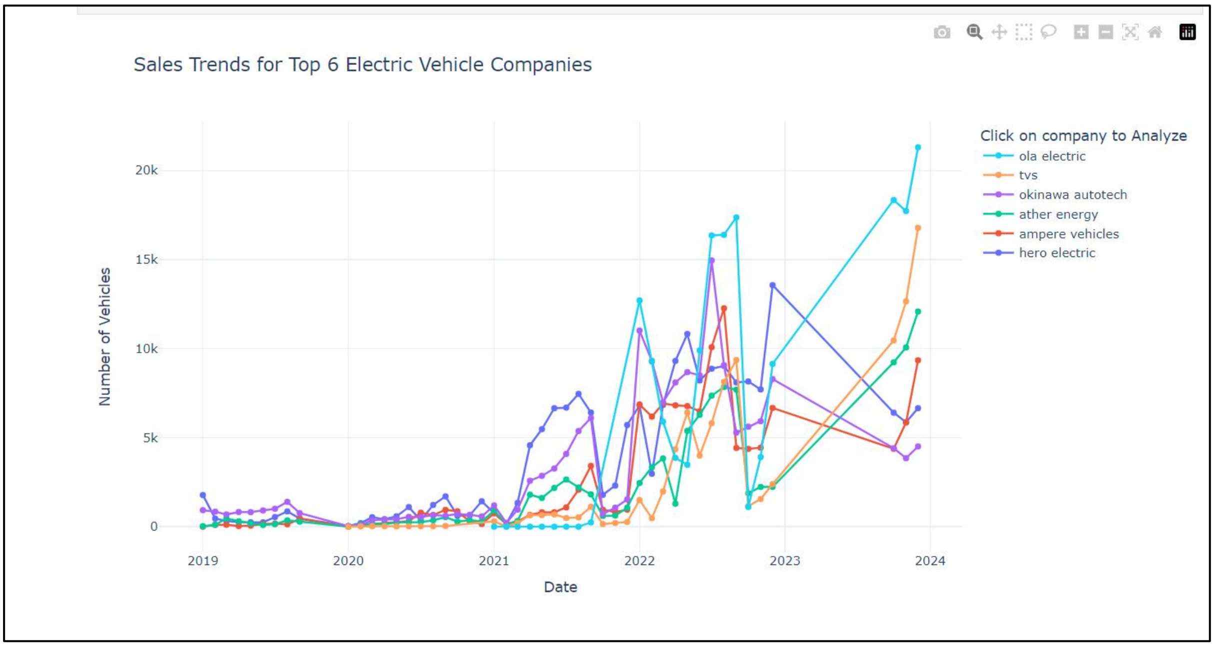
	financial_year	maker	market_share	month	num_vehicles
2	2019-20	hero electric	0.30	Jun	329
393	2022-23	ampere vehicles	0.12	Jan	4371
79	2020-21	hero electric	0.36	Nov	1219
142	2020-21	bajaj	0.03	Feb	111
78	2020-21	hero electric	0.36	Oct	355
178	2020-21	jitendra	0.01	Feb	124
281	2021-22	tvs	0.04	Sep	688
57	2019-20	revolt	0.04	Jan	314
18	2019-20	okinawa autotech	0.38	Oct	1006
465	2022-23	pure ev	0.02	Jan	716

- **Total Sales of Electric Vehicle by maker**-



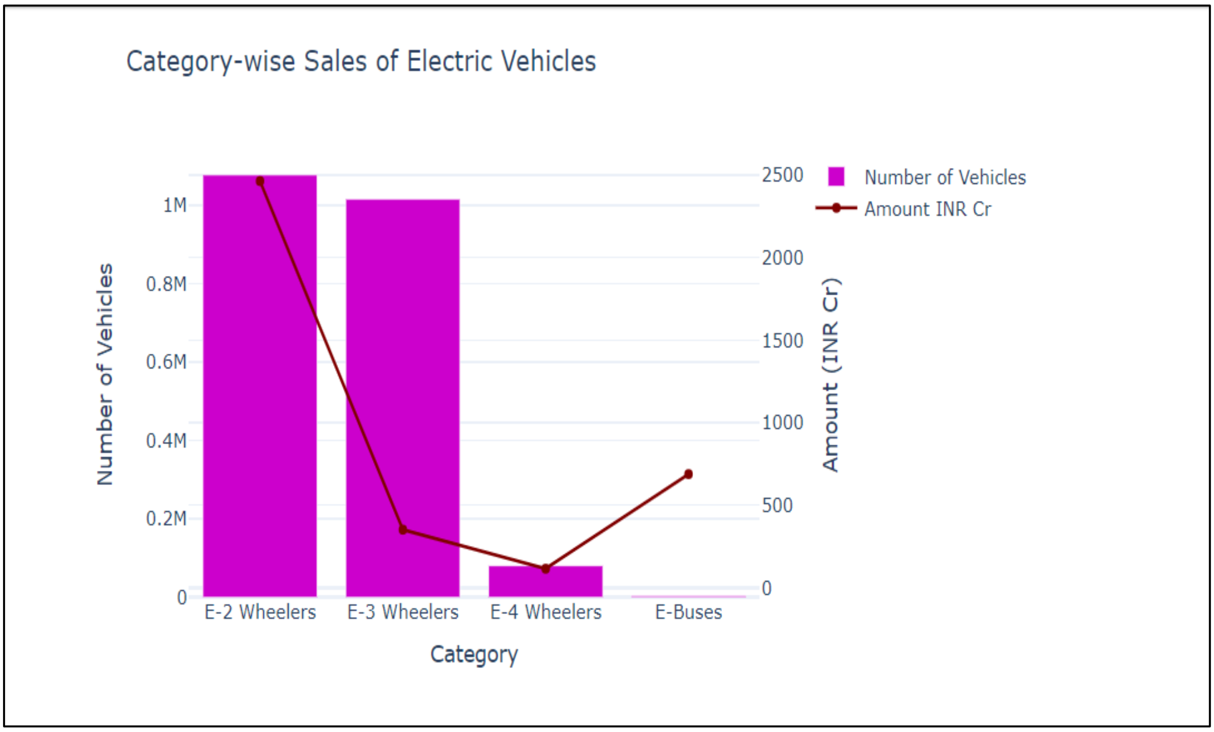
Hero-electric is the most sold electric vehicle followed by Ola-electric till March, 2023

- **Sales Trends for Top 6 Electric Vehicle Companies-**



Ola Electric shows drastic change in sales even though it started its market in 2021; its sales are more in the latest trends.

- **Category-wise Sales of Electric Vehicles-**



The dual y-axis Bar-chart represents the **Number of sales** and their corresponding total **Amount INR Cr** in the market respectively.

It shows **Electric 2-wheelers** are sold the most followed by **Electric 3-wheelers**. But we can see a drastic difference in their Amounts.

Dataset-2 ([Link](#)): Analysing the EV-Bike Dataset

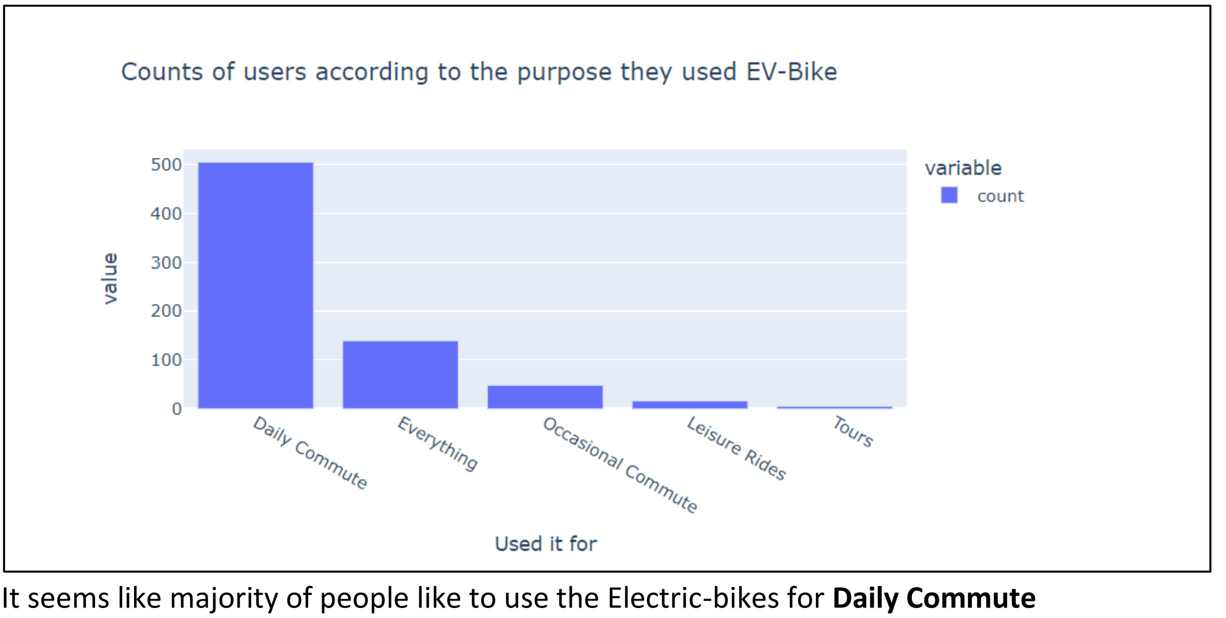
```
[119]: ev_bike = pd.read_csv('ev2_bikewale.csv',usecols=['Used it for' ,'Owned for','rating','Visual Appeal','Reliability','Model Name'])
       print(ev_bike.shape)
       print(ev_bike.isnull().sum())
       ev_bike.head()

       (844, 6)
       Used it for       0
       Owned for         0
       rating            0
       Visual Appeal   105
       Reliability     128
       Model Name        0
       dtype: int64
```

[119]:

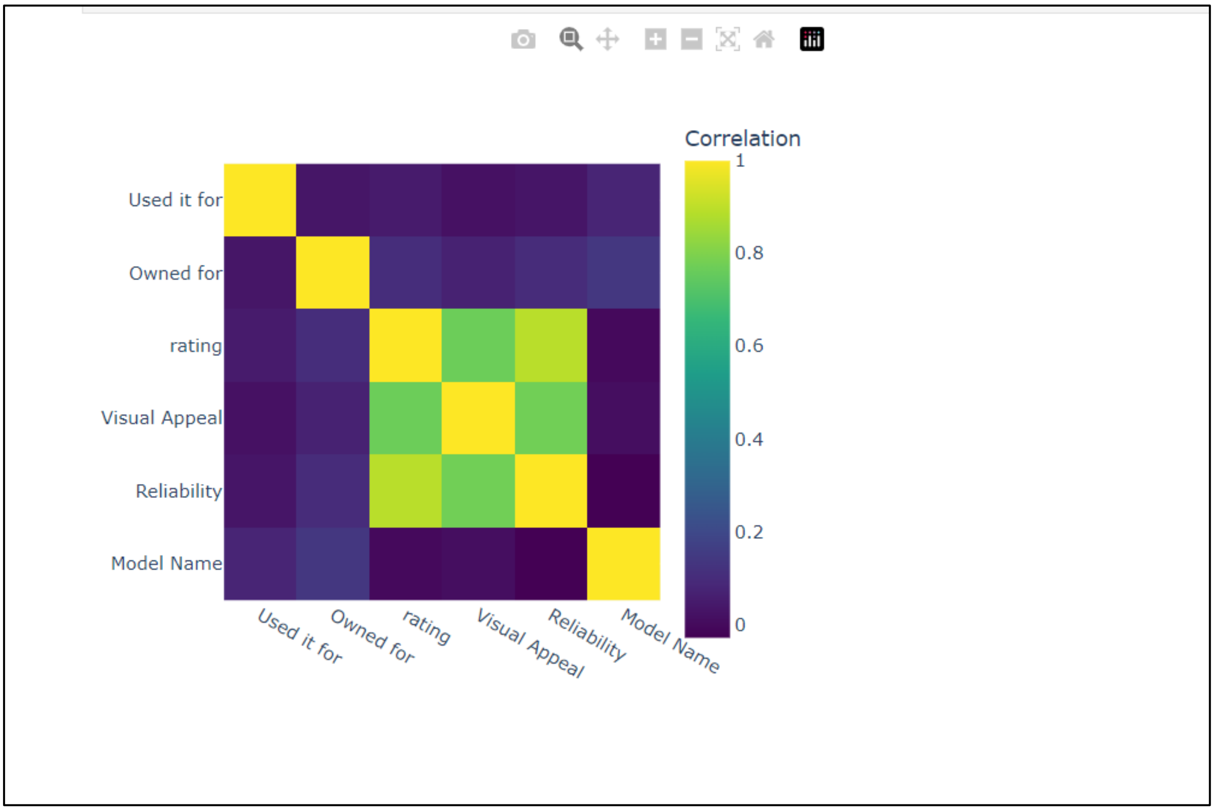
	Used it for	Owned for	rating	Visual Appeal	Reliability	Model Name
0	Daily Commute	Never owned	1	3.0	4.0	TVS iQube
1	Everything	> 1 yr	1	3.0	1.0	TVS iQube
2	Daily Commute	< 3 months	3	4.0	4.0	TVS iQube
3	Daily Commute	6 months-1 yr	1	1.0	1.0	TVS iQube
4	Daily Commute	6 months-1 yr	1	3.0	4.0	TVS iQube

- **Number of users according to the purpose they used EV-Bike-**



It seems like majority of people like to use the Electric-bikes for **Daily Commute**

- **Correlation Heatmap:**



The Correlation between the rating-based features like **rating, Visual Appeal** and **Reliability** are having strong positive correlations among themself.

- **Principal Component Analysis:**

PCA

```
[128]: from sklearn.decomposition import PCA

segmentation_variables = ev_bike.iloc[:, 0:6]

pca = PCA()
pca_model = pca.fit(segmentation_variables)

# standard deviations (square root of explained variance)
std_devs = np.sqrt(pca.explained_variance_)

prop_var = pca.explained_variance_ratio_

cum_prop_var = np.cumsum(prop_var)

print("Importance of components:\n")
print("Standard deviation:", np.round(std_devs, 4),'\n')
print("Proportion of Variance:", np.round(prop_var, 4,),'\n')
print("Cumulative Proportion:", np.round(cum_prop_var, 4))
```

```
Importance of components:

Standard deviation: [10.1176  2.5386  1.2815  0.8826  0.7369  0.5366]

Proportion of Variance: [0.9135 0.0575 0.0147 0.007  0.0048 0.0026]

Cumulative Proportion: [0.9135 0.971  0.9856 0.9926 0.9974 1.    ]
```

1. Standard Deviation

- **Values:** [10.1176, 2.5386, 1.2815, 0.8826, 0.7369, 0.5366]
- **Interpretation:** These values represent the standard deviations of each principal component. A higher standard deviation indicates that the principal component explains a larger amount of variance in the dataset.
  - PC1 has the highest standard deviation, suggesting it captures the most variance.
  - Subsequent principal components have progressively lower standard deviations.

2. Proportion of Variance

- **Values:** [0.9135, 0.0575, 0.0147, 0.007, 0.0048, 0.0026]
- **Interpretation**: This array shows the proportion of the total variance explained by each principal component.
  - PC1 explains approximately 91.35% of the variance, indicating it is the most important component.
  - PC2 explains 5.75% of the variance, making it the second most important.
  - The remaining components explain very little variance individually.

3. Cumulative Proportion

- **Values**: [0.9135, 0.971, 0.9856, 0.9926, 0.9974, 1.]
- **Interpretation:** This array represents the cumulative proportion of variance explained by the first k principal components.
    - By including PC1 and PC2, you capture 97.1% of the total variance.
    - By including PC1, PC2, and PC3, you capture approximately 98.56%.
    - The remaining components only add a small additional amount of variance (1.74% in total), suggesting that most of the variability in the data is captured by the first few principal components.

- ## PCA - Rotation Matrix

```
[129]: rotation_matrix = pd.DataFrame(pca.components_.T,
                                      columns=['PC{}'.format(i+1) for i in range(pca.components_.shape[0])],
                                      index= ev_bike.iloc[:, 0:11].columns)
       rotation_matrix = -rotation_matrix

       print("Rotation (n x k) = ({} x {}):".format(rotation_matrix.shape[0], rotation_matrix.shape[1]))
       print(rotation_matrix)

       Rotation (n x k) = (6 x 6):
                          PC1       PC2       PC3       PC4       PC5       PC6
       Used it for   -0.006823  0.015030  0.023487 -0.995130 -0.090661  0.025948
       Owned for     -0.017862  0.069819  0.996173  0.028231 -0.040297 -0.004923
       rating         0.000718  0.645634 -0.031027 -0.044294  0.399509 -0.648556
       Visual Appeal -0.001468  0.469413 -0.070961  0.083074 -0.873159 -0.072845
       Reliability    0.004487  0.598085 -0.027659  0.004315  0.261012  0.757207
       Model Name    -0.999806  0.001109 -0.017999  0.006152  0.004080  0.002950
```

Matrix Overview

- **PC1 to PC6**: These columns represent the principal components (PCs) that were derived from the PCA.
- **Used it for, Owned for, rating, Visual Appeal, Reliability, Model Name**: These rows represent the original features of the dataset.
- The values in the matrix indicate how much each feature contributes to each principal component.

Insights

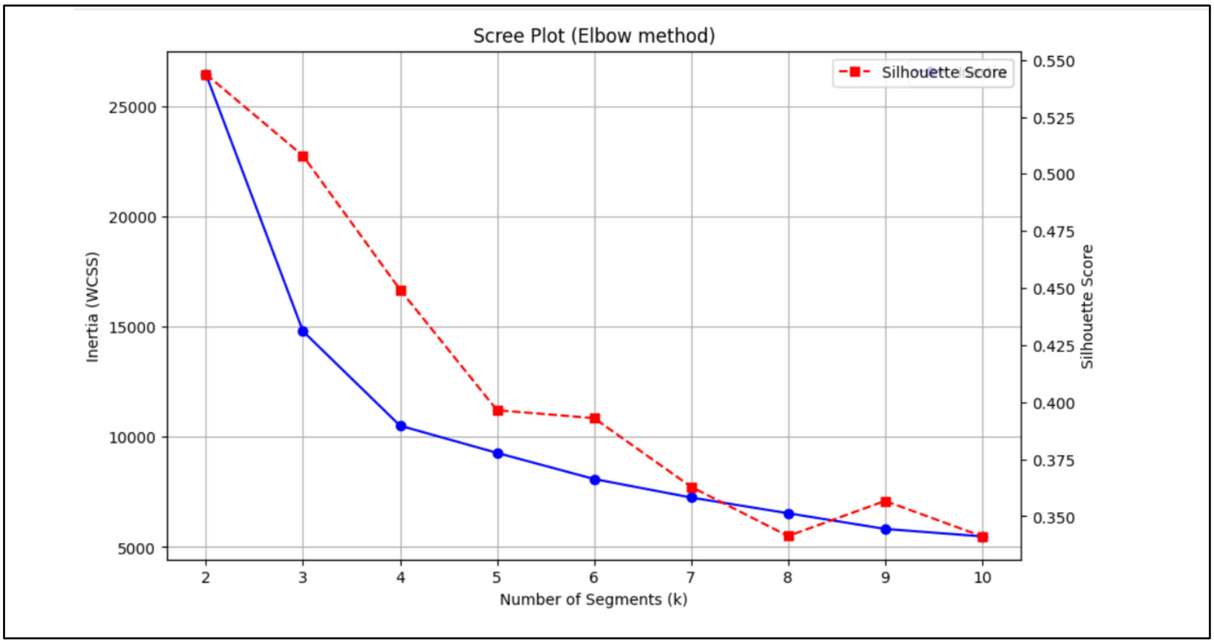
1. **PC1**:
    - **Model Name (-0.999806)** has a very high negative loading on PC1. This indicates that PC1 is almost entirely driven by the variations in the "Model Name" feature.
    - **Interpretation**: The first principal component is primarily associated with the "Model Name" of the electric vehicles.
2. **PC2**:

    - **rating (0.645634)** and **Reliability (0.598085)** have high positive loadings.
    - **Visual Appeal (0.469413)** also has a notable positive loading.
    - **Interpretation**: The second principal component captures the combined variance of "rating," "Reliability," and "Visual Appeal," suggesting these features are correlated and represent a dimension of overall user satisfaction or quality perception.
3. **PC3**:
    - **Owned for (0.996173)** has a very high positive loading.
    - **Interpretation**: The third principal component is almost entirely driven by the "Owned for" feature, indicating it captures the variance in the duration of ownership of the electric vehicles.
4. **PC4**:
    - **Used it for (-0.995130)** has a very high negative loading.
    - **Interpretation**: The fourth principal component is predominantly influenced by the "Used it for" feature, representing the purpose or usage of the electric vehicles.
5. **PC5**:
    - **Visual Appeal (-0.873159)** has a high negative loading.
    - **rating (0.399509)** has a moderate positive loading.
    - **Interpretation**: The fifth principal component captures the variance between "Visual Appeal" and "rating," indicating a possible trade-off or difference in how these two features are perceived.
6. **PC6**:
    - **Reliability (0.757207)** has a high positive loading.
    - **rating (-0.648556)** has a high negative loading.
    - **Interpretation**: The sixth principal component shows an inverse relationship between "Reliability" and "rating," suggesting that as one increases, the other tends to decrease, highlighting a possible divergence in user ratings and perceived reliability.

Overall Analysis

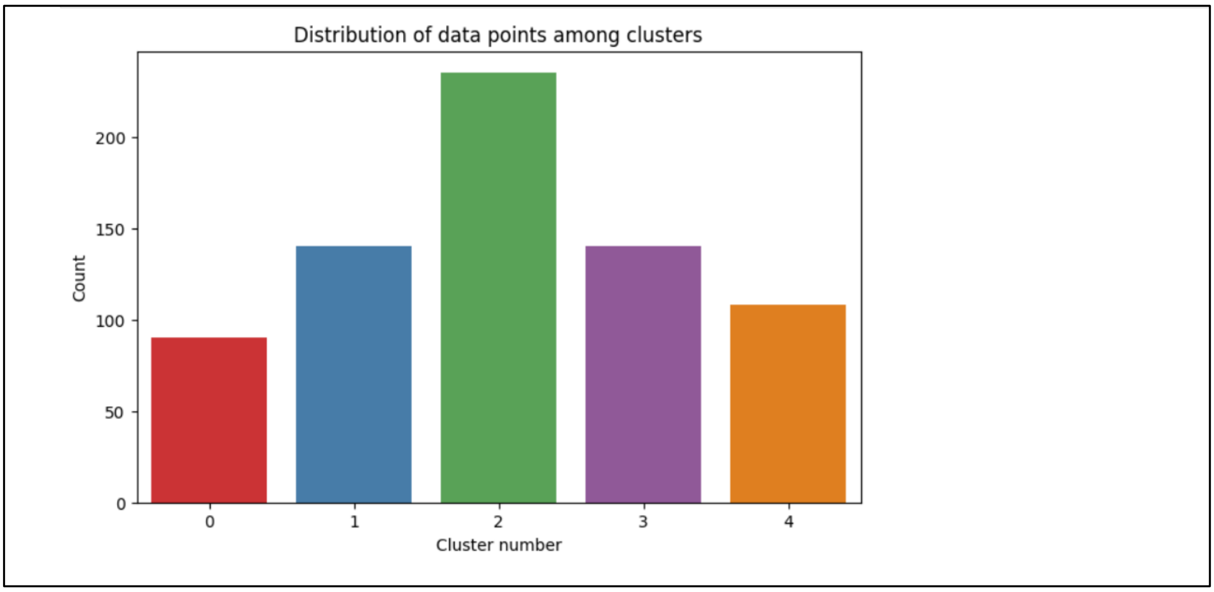
- **Dominant Features**: The "Model Name" is the dominant feature in PC1, while "rating," "Reliability," and "Visual Appeal" dominate PC2. "Owned for" and "Used it for" are the primary drivers for PC3 and PC4, respectively.
- **Correlations**: The PCA indicates correlations among the features. For example, "rating," "Reliability," and "Visual Appeal" are correlated, forming a combined dimension in PC2.
- **Variability Representation**: Each principal component represents a different dimension of variability in the data, capturing specific aspects of the electric vehicle market, such as model differences, user satisfaction, ownership duration, and usage purpose.

- **K-Means Clustering-**



- By analysing the **Elbow Curve,** we can make a decision to make **5** clusters of our small dataset, as the further curve is kind of constant.
- Also, the **Silhouette Score** from 5th to 6th cluster in not much deviating, rather it is kind of decreasing further as the number of clusters increases. A higher silhouette score indicates better-defined clusters.

- **Distribution of data points based on Clusters**



We can observe that **Cluster-2** grouped the highest number of data points. This indicates that the majority of the electric vehicles in the market share similar characteristics captured by the features used in our analysis.

Manufacturers, marketers, and stakeholders in the electric vehicle industry can leverage this insight to tailor their products, services, and marketing efforts to cater to the preferences and needs of this dominant segment. By focusing on the attributes that define cluster-2, businesses can enhance customer satisfaction, improve market penetration, and drive growth in the electric vehicle market.

Overall, the clustering results highlight the diversity within the EV market while underscoring the significant potential of the largest segment, cluster-2, for targeted business strategies and informed decision-making.