Project Report

Electric Vehicle Market Segmentation

Subodh Wasekar

Objective

The objective of this project is to analyze and segment customers in the electric vehicle market based on their vehicle type, battery type, charging type, and battery capacity. The primary goal is to optimize product features and identify a target segment that exhibits specific characteristics, ultimately improving market competitiveness and customer satisfaction.

Data Source

The dataset is taken from the report of Electric vehicle charging Infrastructure and its Grid Integration in India by GILAB@IITB.The dataset includes information about various electric vehicles, such as vehicle type, battery type, charging type, and battery capacity.

Methodology

1. **Data Exploration**
    - The dataset is loaded and examined to understand its structure and basic statistics.
    - Initial exploration involves checking for missing values, data types, and identifying potential features for segmentation.

```
df.head()
```

	Vehicle type	Model	Battery technology	Charging type	Battery capacity (kWh)
0	2W	Electric Photon HX	Li-Ion	AC slow\ncharging	1.872
1	2W	Nyx HX (Dual\nBattery)	Li-Ion	AC slow\ncharging	3.072
2	2W	NYX LX	Li-Ion	AC slow\ncharging	1.536
3	2W	OPTIMA LX\n(VRLA)	VRLA	AC slow\ncharging	0.96
4	2W	Optima LX	Li-Ion	AC slow\ncharging	1.536

2. **Feature Engineering and Preprocessing**
    - Relevant features, including 'Vehicle type,' 'Battery type,' 'Charging type,' and 'Battery capacity (kWh),' are selected for segmentation.
    - Categorical features are encoded using LabelEncoder to prepare the data for clustering.

- Numerical features are standardized using StandardScaler to ensure uniform scaling.

```
df.head()
```

	Vehicle type	Model	Battery technology	Charging type	Battery capacity (kWh)
0	0	Electric Photon HX	0	1	1.872
1	0	Nyx HX (Dual\nBattery)	0	1	3.072
2	0	NYX LX	0	1	1.536
3	0	OPTIMA LX\n(VRLA)	2	1	0.96
4	0	Optima LX	0	1	1.536

3. **Dimensionality Reduction with PCA**
   - Principal Component Analysis (PCA) is applied to reduce the dimensionality of the feature set.
   - This step helps in visualizing and clustering the data effectively.

```
# Extracting features for pca
X=df[['Vehicle type','Battery technology','Charging type','Battery capacity (kWh)']]
```

```
# Standardizing the features
scaler=StandardScaler()
X_scaled=scaler.fit_transform(X)
```

```
X_scaled
```

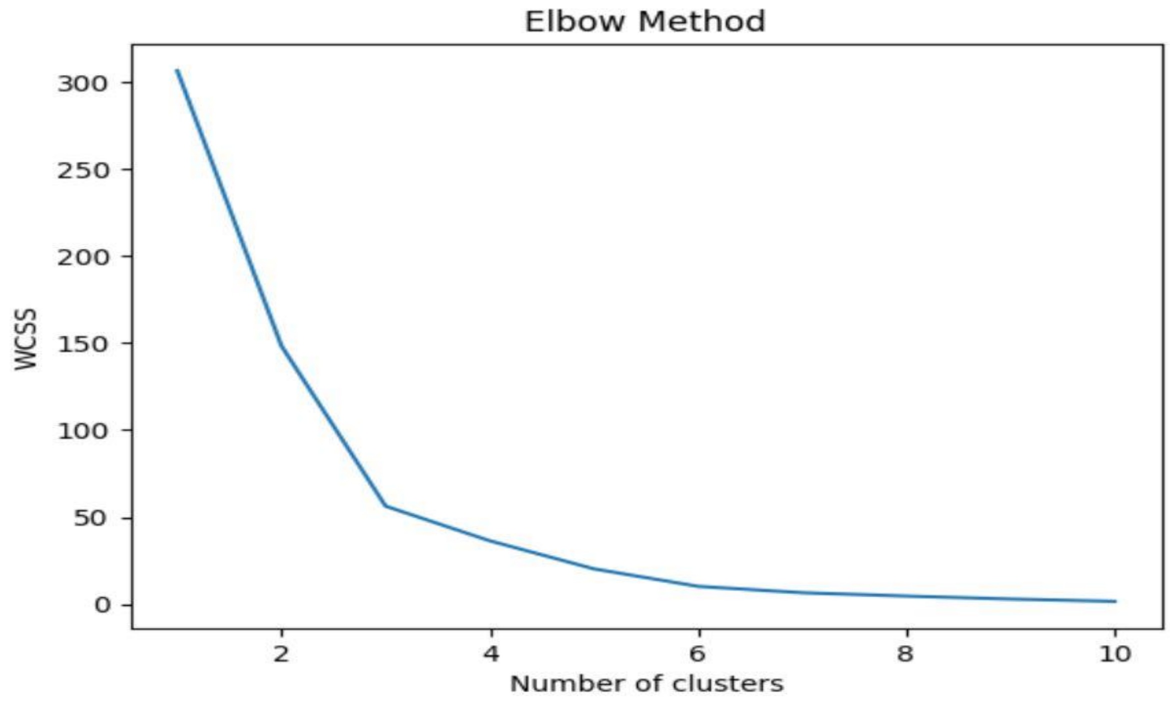
```
array([[-0.92238026, -0.78219274, -0.52530028, -0.38671826],
       [-0.92238026, -0.78219274, -0.52530028, -0.3222505 ],
       [-0.92238026, -0.78219274, -0.52530028, -0.40476924],
```

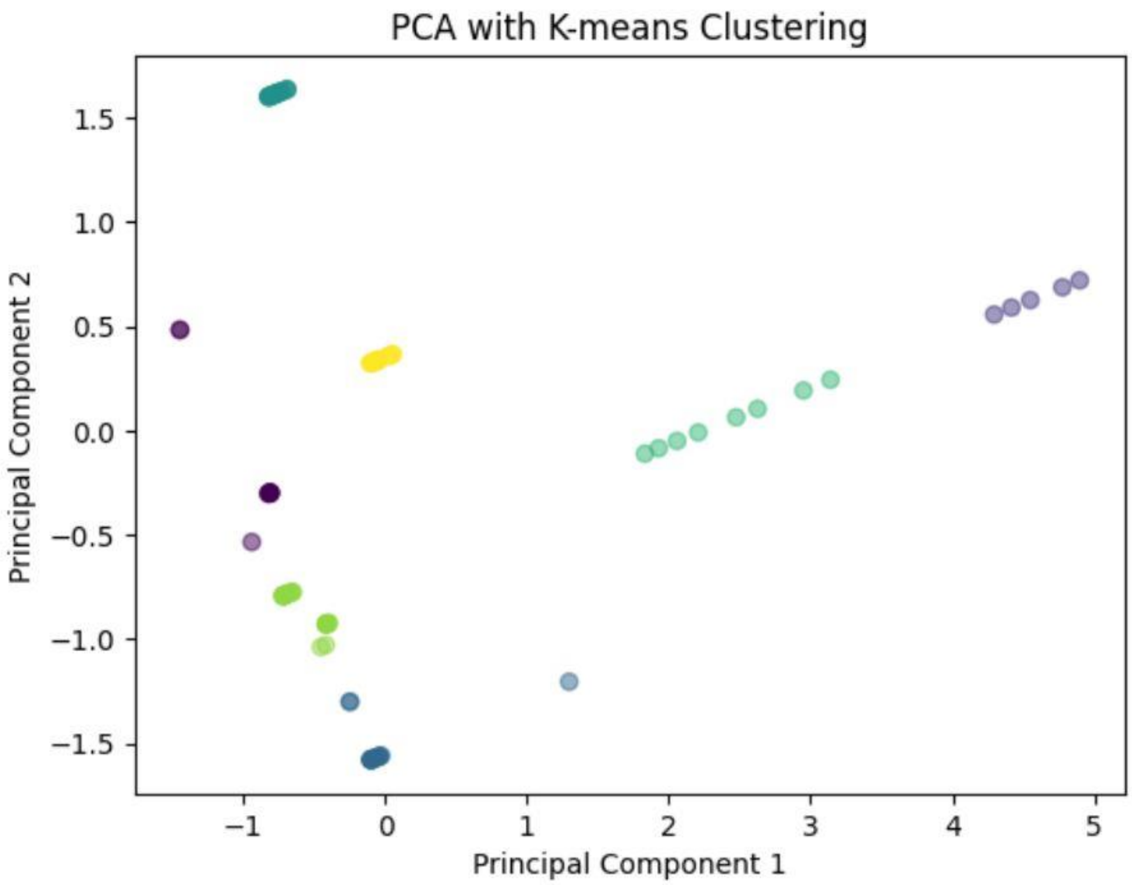
4. **Clustering with K-means**
   - The K-means clustering algorithm is utilized to group customers into segments based on their reduced feature set.
   - The Elbow Method is employed to determine the optimal number of clusters (k).

5. **Elbow Method**
   - The Elbow Method involves running K-means clustering for a range of values of k and plotting the Within-Cluster Sum of Squares (WCSS) against k.
   - The "elbow" point in the graph helps identify the optimal number of clusters, providing a balance between model complexity and segmentation effectiveness.

```
# Finding the  optimal number of clusters using the Elbow Method
wcss = [] #wcss: within cluster sum of squares
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state=42)
    kmeans.fit(X_pca)
    wcss.append(kmeans.inertia_)
```

- Below plots show the optimal number of clusters and PCA with k-means clustering.





6. **Segment Profiling**
    - After clustering, the mean of each feature within each cluster is calculated to create segment profiles.
    - Segment profiling allows for a deeper understanding of the characteristics and preferences of different customer groups.

```
# Calculating mean of each features within each clusters
segment_profile = df.groupby('Segment')[numeric_columns].mean()
print(segment_profile)
```

```
         Vehicle type  Battery technology  Charging type  \
Segment
0              0.0000            1.818182       3.090909
1              2.0000            0.000000       5.000000
2              0.0625            0.000000       4.000000
3              1.0000            2.000000       0.000000
4              2.0000            0.000000       5.000000
5              0.0000            0.304348       2.000000
6              1.0000            0.000000       0.000000

         Battery capacity (kWh)  Segment
Segment
0                      1.131636      0.0
1                     84.740000      1.0
2                      2.418250      2.0
3                      4.983478      3.0
4                     23.968750      4.0
5                      2.281130      5.0
6                      5.495333      6.0
```

7. **Target Segment Selection**
    - The target segment is selected based on specific criteria, such as the highest average 'Battery capacity (kWh)’, ‘Vehicle type’ and ‘Battery technology’ within the segment profiles.
    - This step aligns with the project objective of optimizing product features to meet the preferences of the identified target customer group.

```
# Based on  highest battery capacity
target_segment = segment_profile['Battery capacity (kWh)'].idxmax()
print("Target Segment:", target_segment)
```

```
Target Segment: 1
```

```python
# Based on vehicle type and battery technology
segment_profile[['Vehicle type', 'Battery technology']]
selected_segment = segment_profile[(segment_profile['Vehicle type'] == 1) & (segment_profile['Battery technology'] == 0)]
print("Selected Segment:\n", selected_segment)
```

```
Selected Segment:
          Vehicle type  Battery technology  Charging type  \
Segment
6                 1.0                 0.0            0.0

         Battery capacity (kWh)  Segment
Segment
6                      5.495333      6.0
```

8. Conclusion

By implementing K-means clustering and utilizing the Elbow Method for cluster number selection, this project successfully segments customers in the electric vehicle market. The resulting target segment, identified through objective criteria, provides actionable insights for optimizing product features and tailoring offerings to specific customer preferences.

[**Github Link**]()