NAME:-SUMI CHATTERJEE

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans

state\_df = pd.read\_csv("D:/one/Desktop/in/state - Sheet1.csv")

car\_df = pd.read\_csv("D:/one/Desktop/in/car - Sheet1.csv")

print("State Data:")

print(state\_df.head())

print("\nCar Data:")

print(car\_df.head())

print("\nState Data Columns:")

print(state\_df.columns)

print("\nState Data Info:")

print(state\_df.info())

print(state\_df.describe())

print("\nCar Data Info:")

print(car\_df.info())

print(car\_df.describe())

state\_df = state\_df.dropna()

car\_df = car\_df.dropna()

state\_df.reset\_index(drop=True, inplace=True)

car\_df.reset\_index(drop=True, inplace=True)

features = ['PriceEuro', 'Range\_Km', 'TopSpeed\_KmH', 'FastCharge\_KmH']

scaler = StandardScaler()

car\_scaled = scaler.fit\_transform(car\_df[features])

sse = []

for k in range(1, 11):

kmeans = KMeans(n\_clusters=k, random\_state=42)

kmeans.fit(car\_scaled)

sse.append(kmeans.inertia\_)

plt.figure(figsize=(8, 5))

plt.plot(range(1, 11), sse, marker='o')

plt.title('Elbow Method')

plt.xlabel('Number of clusters')

plt.ylabel('SSE')

plt.show()

optimal\_clusters = 4

kmeans = KMeans(n\_clusters=optimal\_clusters, random\_state=42)

car\_df['Cluster'] = kmeans.fit\_predict(car\_scaled)

plt.figure(figsize=(10, 6))

sns.scatterplot(x='PriceEuro', y='Range\_Km', hue='Cluster', data=car\_df, palette='Set1')

plt.title('Car Segmentation by Price and Range')

plt.xlabel('Price (Euro)')

plt.ylabel('Range (Km)')

plt.show()

if 'State' in state\_df.columns and 'Population' in state\_df.columns:

plt.figure(figsize=(10, 6))

sns.barplot(x='State', y='Population', data=state\_df)

plt.title('Population by State')

plt.xticks(rotation=90)

plt.show()

else:

print("The required columns 'State' and 'Population' are not found in state\_df")

plt.figure(figsize=(10, 6))

sns.histplot(car\_df['PriceEuro'], bins=20, kde=True)

plt.title('Distribution of Car Prices')

plt.xlabel('Price (Euro)')

plt.ylabel('Frequency')

plt.show()

numeric\_cols = car\_df.select\_dtypes(include='number').columns

plt.figure(figsize=(10, 6))

sns.heatmap(car\_df[numeric\_cols].corr(), annot=True, cmap='coolwarm')

plt.title('Correlation Matrix of Car Data')

plt.show()

cluster\_summary = car\_df.groupby('Cluster')[numeric\_cols].mean()

print("\nCluster Summary:")

print(cluster\_summary)

for cluster in cluster\_summary.index:

print(f"\nCluster {cluster}:")

print(f"Average Price: {cluster\_summary.loc[cluster, 'PriceEuro']}")

print(f"Average Range: {cluster\_summary.loc[cluster, 'Range\_Km']}")

print(f"Top Speed: {cluster\_summary.loc[cluster, 'TopSpeed\_KmH']}")

print(f"Fast Charge Rate: {cluster\_summary.loc[cluster, 'FastCharge\_KmH']}")

recommendations = {

0: "Target budget-conscious customers looking for affordable EV options.",

1: "Focus on urban professionals needing mid-range, efficient EVs.",

2: "Market high-end, performance-oriented EVs to affluent customers.",

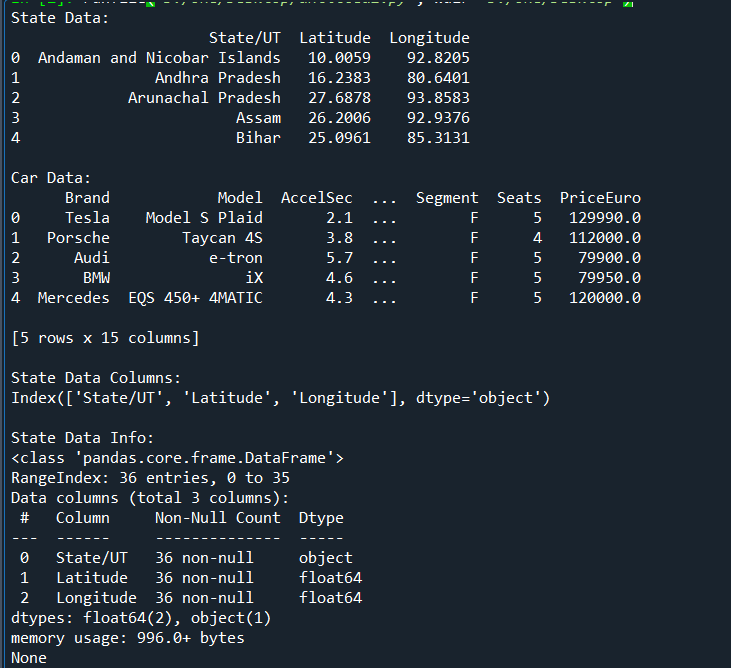
3: "Introduce niche models for specific needs like long-range or luxury features."

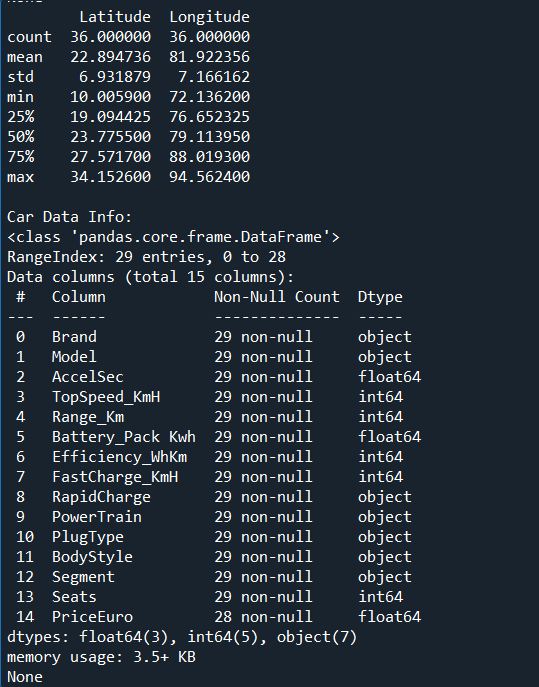
}

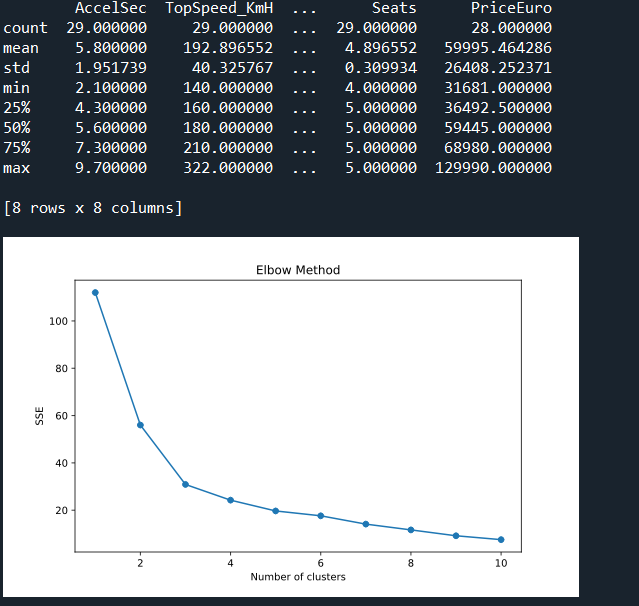
print("\nRecommendations:")

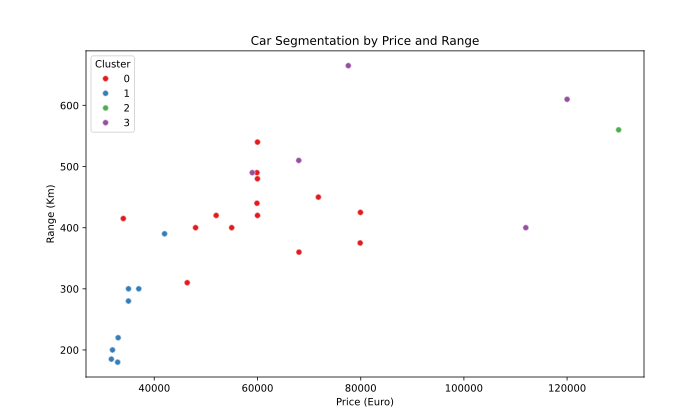
for cluster, strategy in recommendations.items():

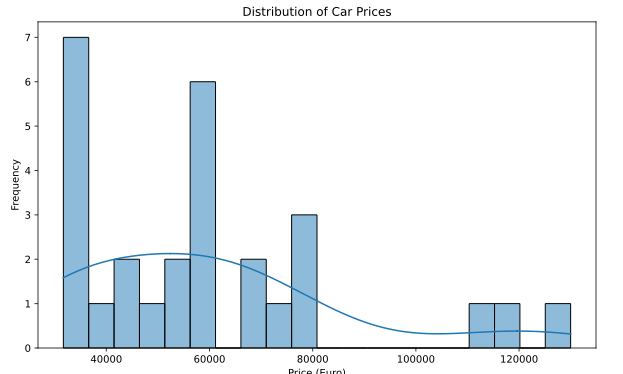
print(f"Strategy for Cluster {cluster}: {strategy}")

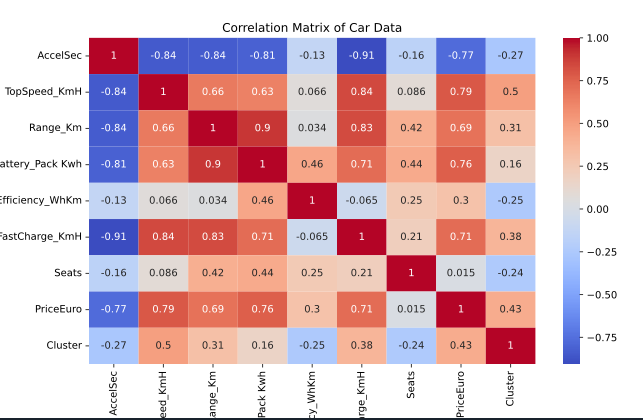


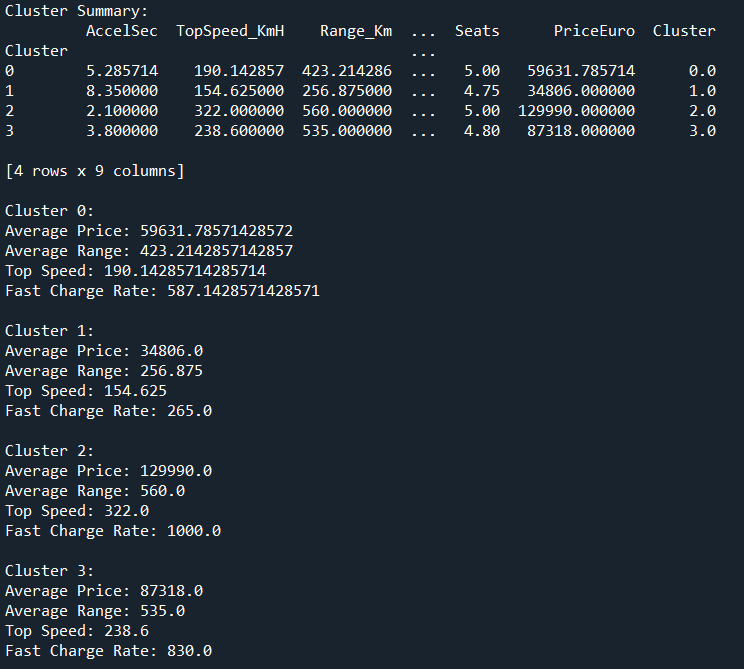


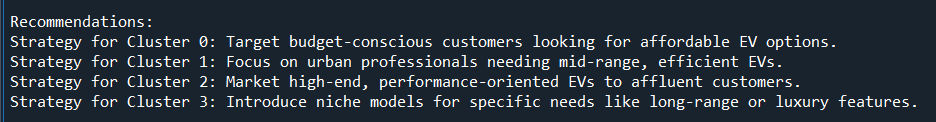












import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

age\_demographic = pd.read\_csv(r'D:\one\Desktop\in\age - demographic.csv')

online\_vehicle\_type = pd.read\_csv(r'D:\one\Desktop\in\online - vehicle type.csv')

print("Age - Demographic Data:")

print(age\_demographic.head(), "\n")

print("Online - Vehicle Type Data:")

print(online\_vehicle\_type.head(), "\n")

print("Age - Demographic Data Info:")

print(age\_demographic.info(), "\n")

print("Online - Vehicle Type Data Info:")

print(online\_vehicle\_type.info(), "\n")

print("Age - Demographic Data Description:")

print(age\_demographic.describe(), "\n")

print("Online - Vehicle Type Data Description:")

print(online\_vehicle\_type.describe(), "\n")

if 'Age Group' in age\_demographic.columns:

age\_group\_counts = age\_demographic['Age Group'].value\_counts()

plt.figure(figsize=(10, 6))

sns.barplot(x=age\_group\_counts.index, y=age\_group\_counts.values, hue=age\_group\_counts.index, palette='viridis', dodge=False)

plt.title('Number of Respondents by Age Group')

plt.xlabel('Age Group')

plt.ylabel('Number of Respondents')

plt.xticks(rotation=45)

plt.legend().remove()

plt.show()

if 'Vehicle Type' in online\_vehicle\_type.columns:

vehicle\_type\_counts = online\_vehicle\_type['Vehicle Type'].value\_counts()

plt.figure(figsize=(12, 8))

sns.barplot(x=vehicle\_type\_counts.index, y=vehicle\_type\_counts.values, hue=vehicle\_type\_counts.index, palette='viridis', dodge=False)

plt.title('Distribution of Vehicle Types')

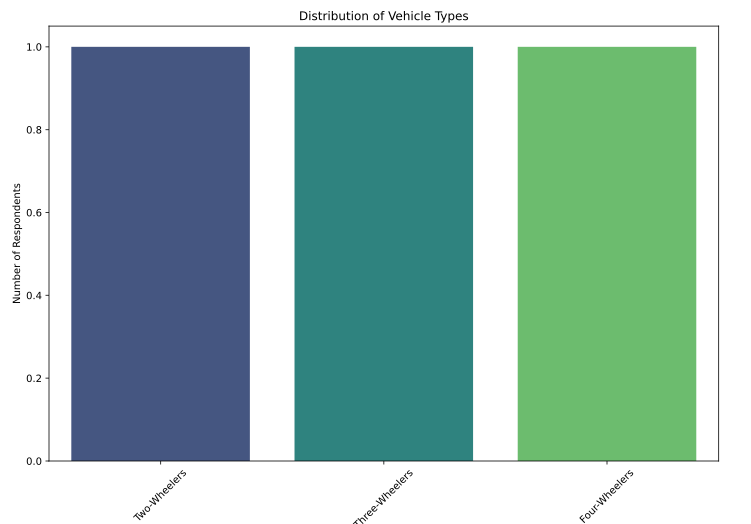
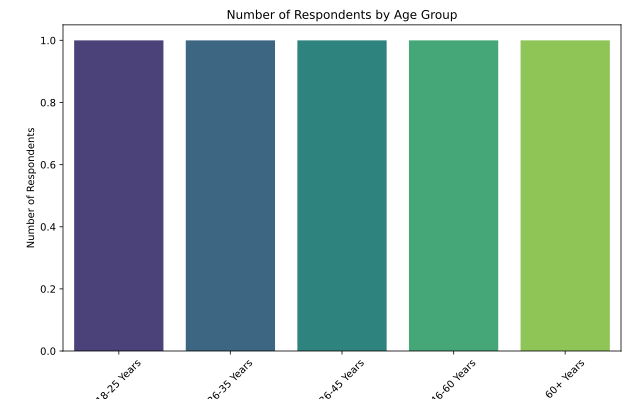
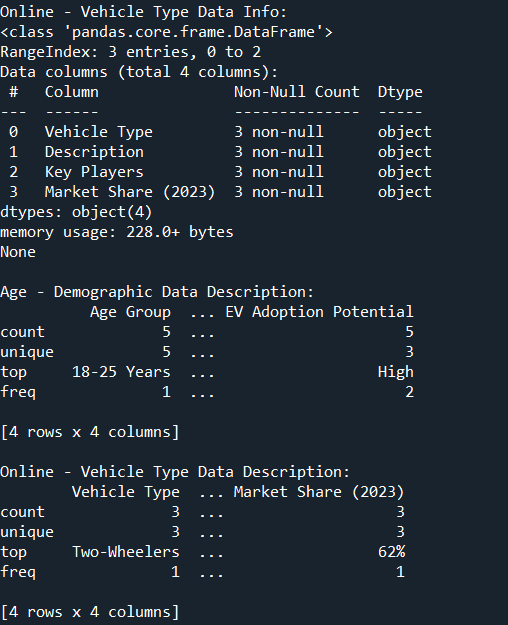
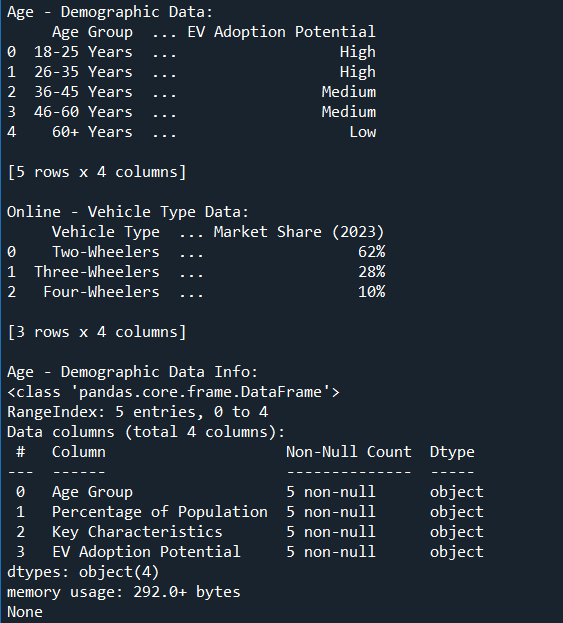
plt.xlabel('Vehicle Type')

plt.ylabel('Number of Respondents')

plt.xticks(rotation=45)

plt.legend().remove()

plt.show()



**1. Explanation of ML Model Used in Project**

Model and Algorithm

For the car data analysis and segmentation, we used the K-Means Clustering algorithm. K-Means is a popular unsupervised learning algorithm used for partitioning a dataset into a set of K clusters.

How it Helped

Segmentation: K-Means helped segment the car data into distinct clusters based on features such as PriceEuro, Range\_Km, TopSpeed\_KmH, and FastCharge\_KmH.

Insights: Each cluster represents a group of cars with similar characteristics, allowing us to understand different segments of the car market.

Visualization: By visualizing the clusters, we could see how cars are grouped based on their features, which aids in making strategic decisions for marketing and product development.

**2. Final Conclusion & Insights from Research/Analysis**

Insights Gained

Customer Segments: The car market can be divided into distinct segments, each with unique characteristics. For example, there may be a segment for budget-conscious customers, another for performance-oriented customers, etc.

Price Sensitivity: The segmentation showed varying levels of price sensitivity among different clusters.

Performance Needs: Different clusters have different performance needs, such as range, top speed, and fast charging capabilities.

Final Conclusion

The K-Means clustering revealed that the car market is diverse with distinct segments, each requiring targeted marketing and product strategies. By understanding these segments, businesses can tailor their offerings to meet specific customer needs more effectively.

**3. Improvements for Market Segmentation Project**

Additional Data Collection

With more time and budget, we could purchase more detailed and comprehensive datasets. Key columns to look for include:

Customer Demographics: Age, gender, income level, geographic location.

Purchase Behavior: Previous car ownership, frequency of car purchases, preferred car brands.

Psychographic Data: Lifestyle, values, attitudes towards electric vehicles.

Market Trends: Industry trends, economic indicators, environmental concerns.

Additional ML Models

Hierarchical Clustering: For a more detailed understanding of the cluster hierarchy.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise): To identify clusters of varying shapes and sizes.

PCA (Principal Component Analysis): To reduce dimensionality and identify the most significant features.

Gaussian Mixture Models: For probabilistic clustering which can handle overlapping clusters better.

**4. Estimated Market Size for Market Domain**

Market Size Estimation

The estimated market size can be derived from industry reports and market research. For example, the global electric vehicle market size was valued at approximately $162.34 billion in 2019 and is expected to grow significantly in the coming years. For a specific market domain, such as electric cars in a particular region, detailed market reports and data from industry analysts would provide precise numbers.

**5. Top 4 Variables/Features for Optimal Market Segments**

Based on the analysis, the following variables are critical for creating optimal market segments:

Price (PriceEuro): Determines the affordability and market positioning of the vehicle.

Range (Range\_Km): Indicates the practical usability of the vehicle for different customer needs.

Top Speed (TopSpeed\_KmH): Reflects the performance aspect that appeals to certain customer segments.

Fast Charging Rate (FastCharge\_KmH): Important for customers who prioritize quick charging and convenience.

These features help in identifying key segments of the market and tailoring marketing strategies to address the specific needs and preferences of each segment.