MARKET SEGMENTATION OF ELECTRIC VEHICLE FOR STARTUP

SUKASH L

5 SEPTEMBER 2024

ABSTRACT

This study investigates the electric vehicle (EV) market in India to identify optimal locations and target demographics for introducing new EVs. Utilizing data on EV population, consumer buying behavior, and charging infrastructure, we performed a comprehensive analysis to determine potential market segments. We assessed the distribution of EVs across major cities, analyzed demographic factors influencing vehicle preferences, and examined the availability of charging stations. The study employs various data visualization techniques and clustering methods to segment potential buyers and estimate the profitability of different market segments. Our objective is to develop a strategic plan to effectively enter the Indian EV market by targeting the most promising segments and locations.

1. PROBLEM STATEMENT:

The objective is to identify the most promising locations and customer segments for launching electric vehicles (EVs) in India. This involves analyzing data related to the distribution of EVs in various cities, consumer demographics, and preferences for vehicle types. We seek to understand which cities have the highest potential for EV adoption and which demographic factors are most relevant for targeting potential buyers.

We will also explore the impact of vehicle pricing and income levels on consumer preferences, using clustering techniques to segment the market and identify high-value customer segments. Additionally, we will assess the availability of charging infrastructure and government initiatives supporting EVs.

The ultimate goal is to devise a targeted market entry strategy that aligns with the Innovation Adoption Life Cycle and optimizes pricing and marketing efforts based on detailed data analysis.

2. DATA COLLECTION:

To conduct a comprehensive market segmentation analysis for our electric vehicle (EV) startup in India, we initiated a rigorous data collection process. We sourced data from multiple online platforms and databases, including datasets on EV population, consumer buying behavior, and charging infrastructure.

This process involved:

1. **Acquisition of Relevant Datasets:** We collected data from reputable sources such as Kaggle and various industry reports to ensure the accuracy and relevance of our information.
2. **Dataset Details:**
    - **Electric Vehicle Population Data:** Provided insights into the distribution of EVs across different cities in India.
    - **Consumer Buying Behaviour Data:** Offered detailed demographic information, including age, income, and vehicle preferences.

- **Charging Stations Data:** Highlighted the distribution and availability of EV charging infrastructure across states.

3. **Data Preparation:** We processed and cleaned the datasets to address any inconsistencies and ensure compatibility for analysis. This step involved handling missing values, standardizing formats, and performing exploratory data analysis.

4. **Data Sources:**

   - **Electric Vehicle Population Data:** https://catalog.data.gov/dataset/electric-vehiclepopulation-data

   - **Consumer Buying Behaviour Data:** https://www.kaggle.com/datasets/karivedha/indian-consumers-cars-purchasingbehaviour

   - **Charging Stations Data:** https://www.kaggle.com/datasets/saketpradhan/electricvehicle-charging-stations-in-india

The goal of this data collection and preparation phase is to build a robust foundation for our market analysis. By understanding the distribution of EVs, consumer demographics, and charging infrastructure, we can identify key market segments and develop a targeted strategy for successful market entry.

3. CODE IMPLEMENTATION:

1. Import necessary libraries and load the datasets

```
[105] # Required Libraries
      import pandas as pd
      import matplotlib.pyplot as plt
      import seaborn as sns
      from sklearn.preprocessing import StandardScaler
      from sklearn.cluster import KMeans
      from sklearn.decomposition import PCA

      # Load datasets
      ev_population = pd.read_csv('/content/Electric_Vehicle_Population_Data.csv')
      buying_behaviour = pd.read_csv('/content/Indian_automobile_buying_behaviour.csv')
      charging_stations = pd.read_csv('/content/ev-charging-stations-india.csv')
```

2. Basic information about the datasets

```
[146] # Display columns and basic info
      print("EV Population Data:")
      print(ev_population.info())
```

```
EV Population Data:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 138779 entries, 0 to 138778
Data columns (total 15 columns):
 #   Column                                             Non-Null Count   Dtype
---  ------                                             --------------   -----
 0   VIN (1-10)                                         138779 non-null  object
 1   County                                             138776 non-null  object
 2   City                                               138776 non-null  object
 3   State                                              138779 non-null  object
 4   Postal Code                                        138776 non-null  float64
 5   Model Year                                         138779 non-null  int64
 6   Make                                               138779 non-null  object
 7   Model                                              138493 non-null  object
 8   Electric Vehicle Type                              138779 non-null  object
 9   Clean Alternative Fuel Vehicle (CAFV) Eligibility  138779 non-null  object
 10  Electric Range                                     138779 non-null  int64
 11  Base MSRP                                          138779 non-null  int64
 12  Legislative District                               138464 non-null  float64
 13  DOL Vehicle ID                                     138779 non-null  int64
 14  Vehicle Location                                   138773 non-null  object
dtypes: float64(2), int64(4), object(9)
memory usage: 15.9+ MB
None
```

```
print("\nIndian Buying Behaviour Data:")
print(buying_behaviour.info())
```

```
Indian Buying Behaviour Data:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Age               99 non-null     int64
 1   Profession        99 non-null     object
 2   Marrital Status   99 non-null     object
 3   Education         99 non-null     object
 4   No of Dependents  99 non-null     int64
 5   Personal loan     99 non-null     object
 6   House Loan        99 non-null     object
 7   Wife Working      99 non-null     object
 8   Salary            99 non-null     int64
 9   Wife Salary       99 non-null     int64
 10  Total Salary      99 non-null     int64
 11  Make              99 non-null     object
 12  Price             99 non-null     int64
dtypes: int64(6), object(7)
memory usage: 10.2+ KB
None
```

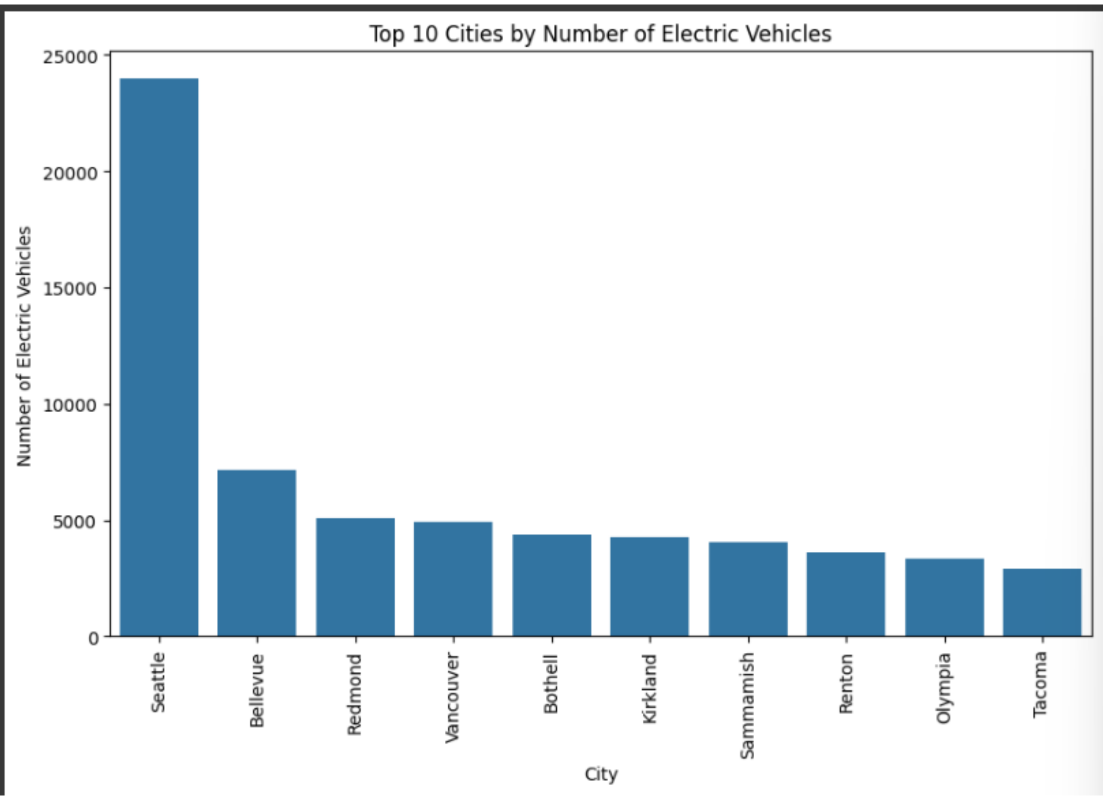
[148]

```
print("\nEV Charging Stations Data:")
print(charging_stations.info())
```

```
EV Charging Stations Data:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1547 entries, 0 to 1546
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   name       1547 non-null   object
 1   state      1547 non-null   object
 2   city       1547 non-null   object
 3   address    1507 non-null   object
 4   lattitude  1541 non-null   object
 5   longitude  1541 non-null   float64
 6   type       1539 non-null   float64
dtypes: float64(2), object(5)
memory usage: 84.7+ KB
None
```

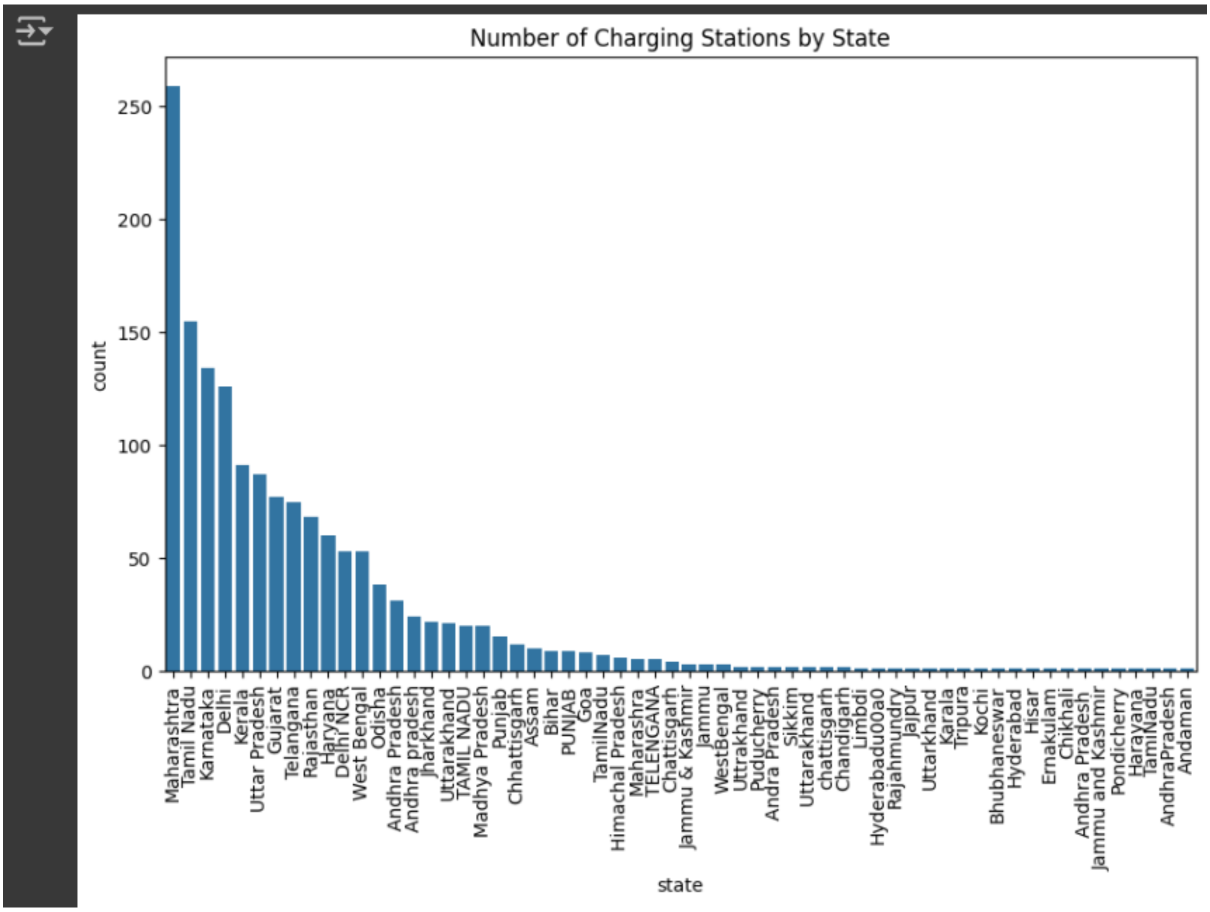
3. The top cities with highest number of electric vehicles

```python
top_cities = ev_population['City'].value_counts().head(10)

# Plot the distribution for only the top cities
plt.figure(figsize=(10, 6))
sns.barplot(x=top_cities.index, y=top_cities.values)
plt.xticks(rotation=90)
plt.title('Top 10 Cities by Number of Electric Vehicles')
plt.ylabel('Number of Electric Vehicles')
plt.xlabel('City')
plt.show()
```



4. The number of charging stations in every state

```python
# Plot distribution of charging stations by State
plt.figure(figsize=(10, 6))
sns.countplot(data=charging_stations, x='state', order=charging_stations['state'].value_counts().index)
plt.xticks(rotation=90)
plt.title('Number of Charging Stations by State')
plt.show()
```



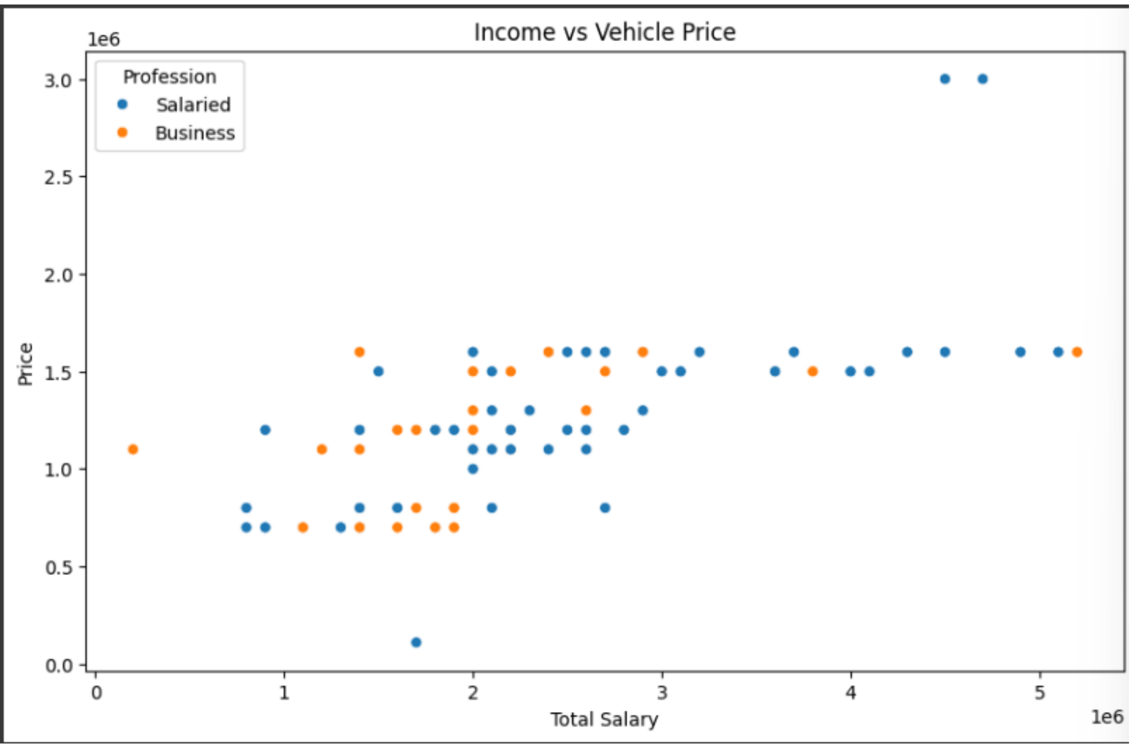
5. Statistics of the demographics present in the dataset

```
[151] # Descriptive statistics of demographics
      print(buying_behaviour[['Age', 'Salary', 'Total Salary', 'Price']].describe())
```

	Age	Salary	Total Salary	Price
count	99.000000	9.900000e+01	9.900000e+01	9.900000e+01
mean	36.313131	1.736364e+06	2.270707e+06	1.194040e+06
std	6.246054	6.736217e+05	1.050777e+06	4.376955e+05
min	26.000000	2.000000e+05	2.000000e+05	1.100000e+05
25%	31.000000	1.300000e+06	1.550000e+06	8.000000e+05
50%	36.000000	1.600000e+06	2.100000e+06	1.200000e+06
75%	41.000000	2.200000e+06	2.700000e+06	1.500000e+06
max	51.000000	3.800000e+06	5.200000e+06	3.000000e+06

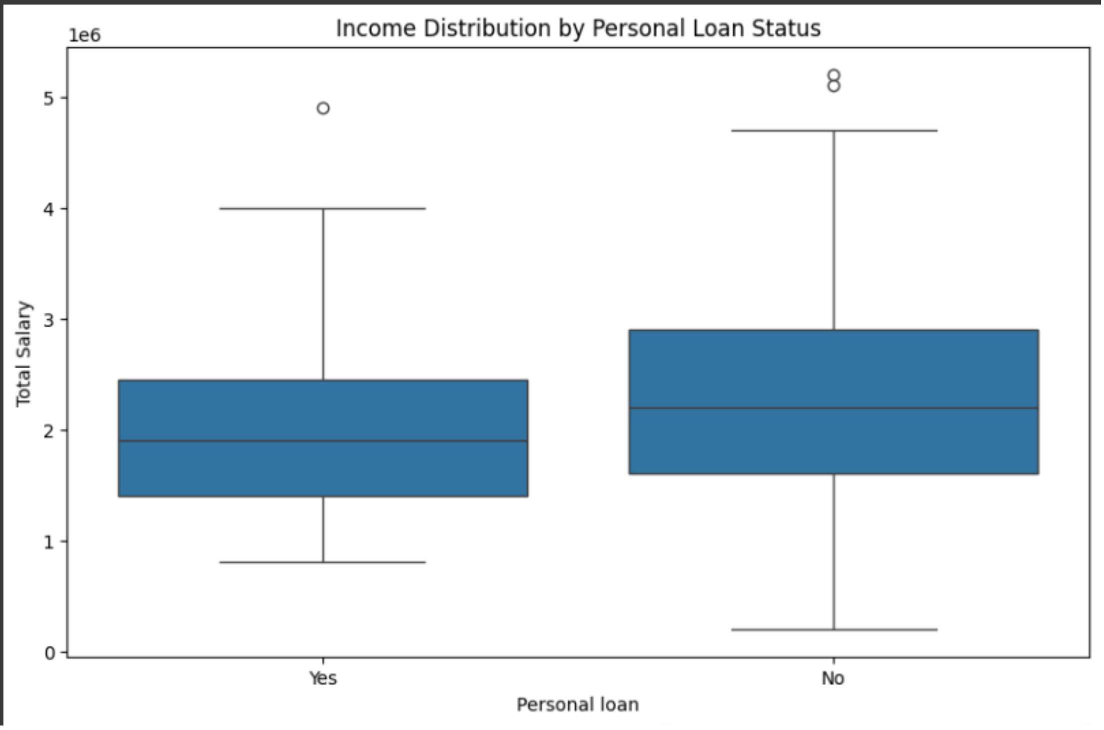
6. Plot between the income and vehicle price

```python
# Income vs Vehicle Price Analysis
plt.figure(figsize=(10, 6))
sns.scatterplot(data=buying_behaviour, x='Total Salary', y='Price', hue='Profession')
plt.title('Income vs Vehicle Price')
plt.show()
```



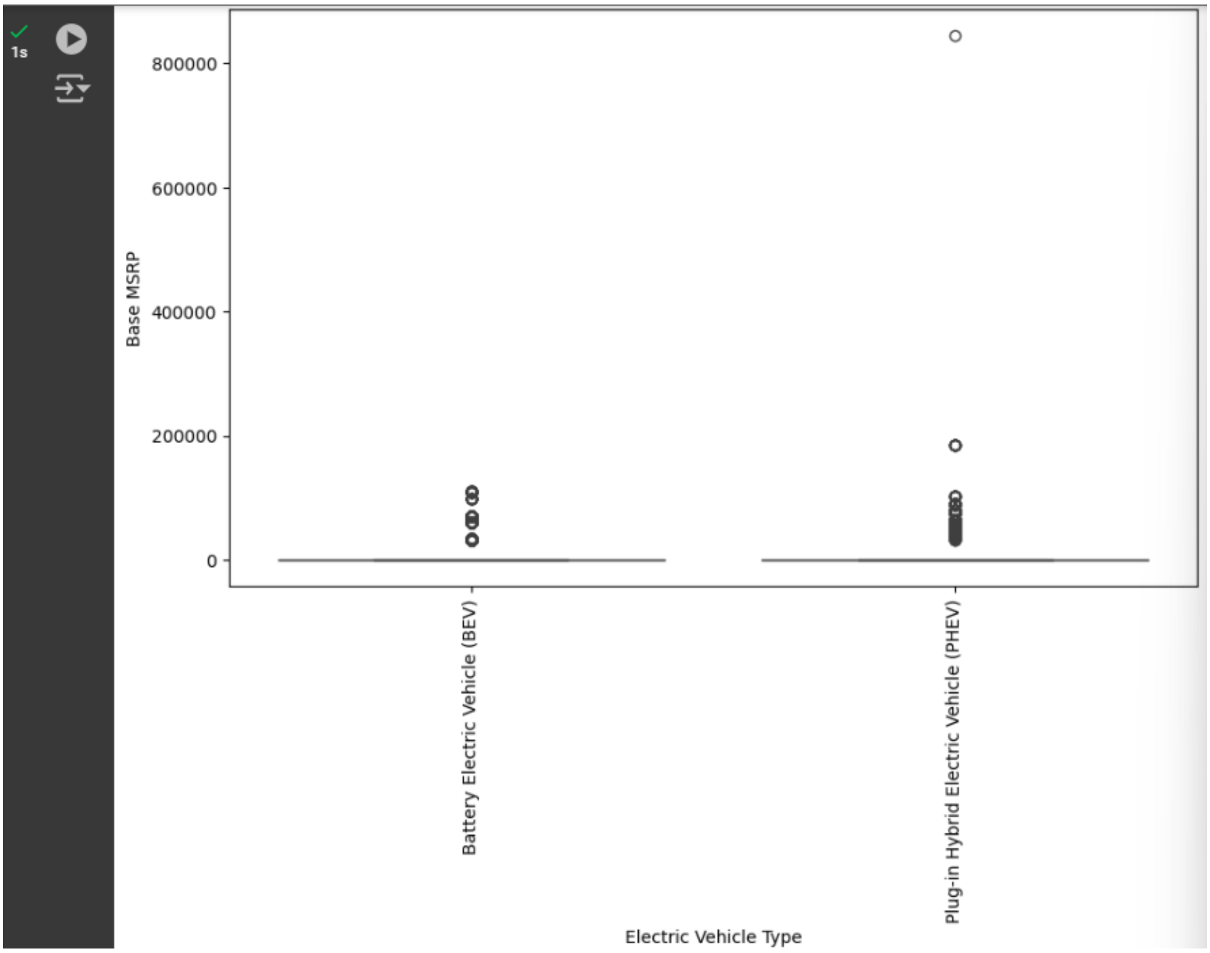
7. Plot between income and loan status

```python
# Boxplot: Income vs Loan Status
plt.figure(figsize=(10, 6))
sns.boxplot(data=buying_behaviour, x='Personal loan', y='Total Salary')
plt.title('Income Distribution by Personal Loan Status')
plt.show()
```



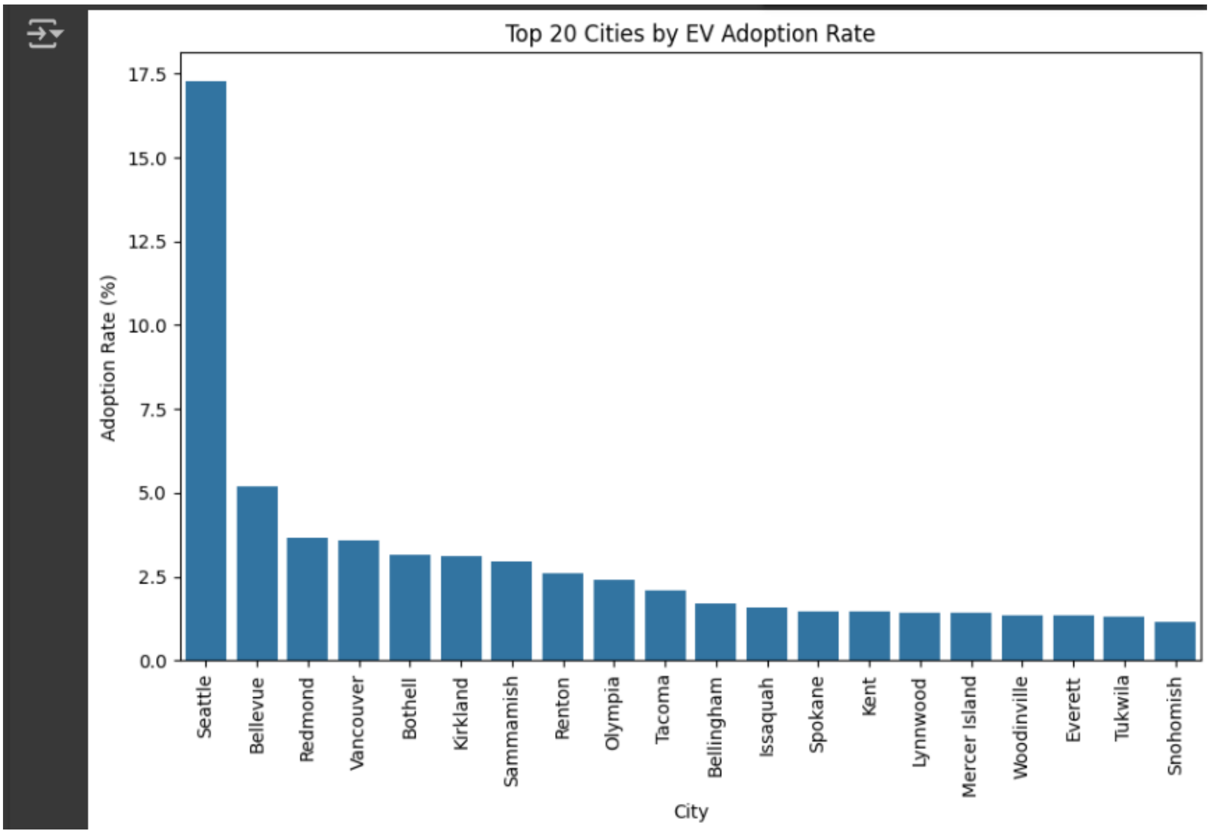
8. Base MSRP for electric vehicle type

```python
# Base MSRP by Electric Vehicle Type
plt.figure(figsize=(10, 6))
sns.boxplot(data=ev_population, x='Electric Vehicle Type', y='Base MSRP')
plt.xticks(rotation=90)
plt.title('Base MSRP by Electric Vehicle Type')
plt.show()
```



9. EV adoption rate by city

```
# EV Adoption Rate by City
adoption_rate = ev_population['City'].value_counts() / len(ev_population) * 100

# Select the top 10 cities by number of EVs for plotting
top_cities_adoption_rate = adoption_rate.head(20)

# Plot the EV adoption rate for the top 10 cities
plt.figure(figsize=(10, 6))
sns.barplot(x=top_cities_adoption_rate.index, y=top_cities_adoption_rate.values)
plt.xticks(rotation=90)
plt.title('Top 20 Cities by EV Adoption Rate')
plt.ylabel('Adoption Rate (%)')
plt.xlabel('City')
plt.show()
```



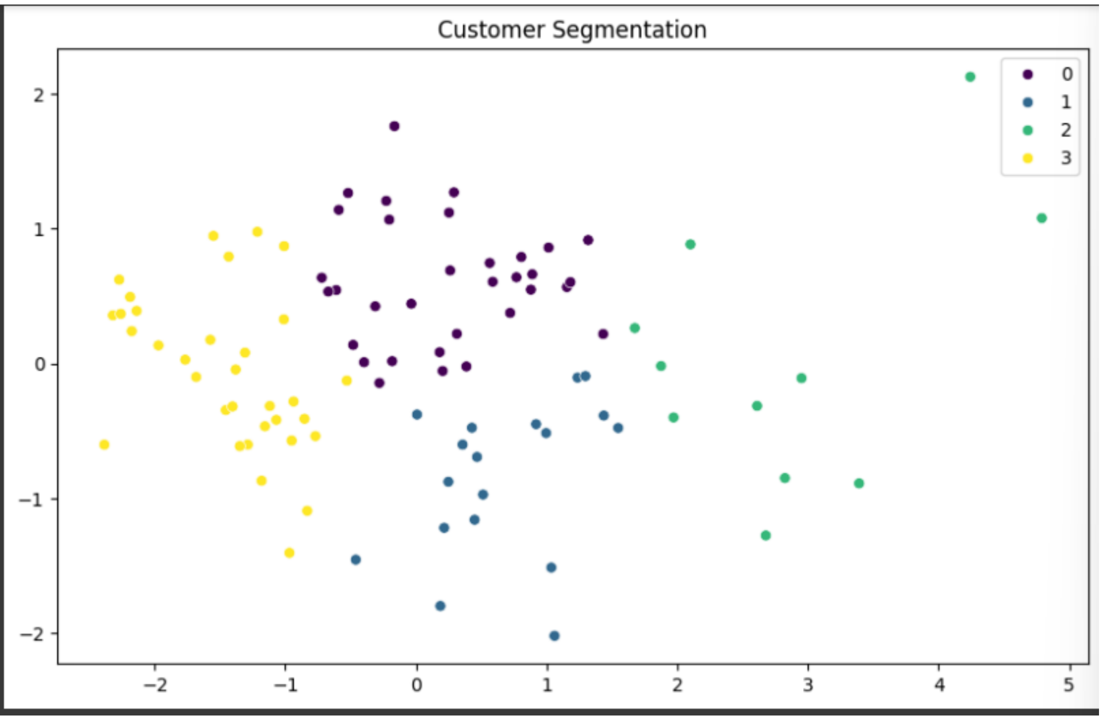
10. Correlation with the demographics

```
[157] # Correlation with demographics
      print("Correlation between Salary and Price:")
      print(buying_behaviour[['Total Salary', 'Price']].corr())
```

```
Correlation between Salary and Price:
              Total Salary     Price
Total Salary      1.000000  0.717442
Price             0.717442  1.000000
```

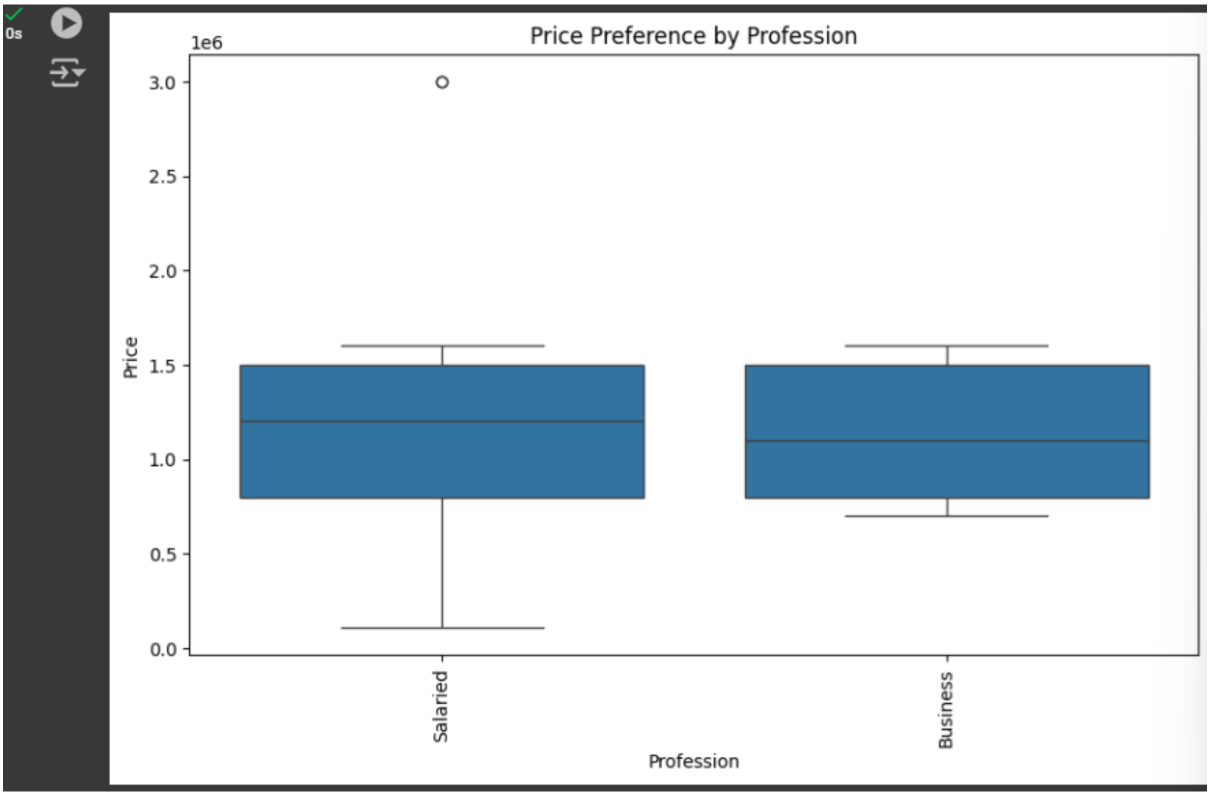
11. Customer Segmentation using clustering

```
scaler = StandardScaler()
scaled_data = scaler.fit_transform(buying_behaviour[['Age', 'Total Salary', 'Price']].dropna())

# PCA for visualization (reduce to 2 dimensions)
pca = PCA(n_components=2)
pca_data = pca.fit_transform(scaled_data)

# KMeans Clustering
kmeans = KMeans(n_clusters=4, random_state=42)
clusters = kmeans.fit_predict(pca_data)

# Add cluster labels to DataFrame
buying_behaviour['Cluster'] = clusters

# Plot clusters
plt.figure(figsize=(10, 6))
sns.scatterplot(x=pca_data[:, 0], y=pca_data[:, 1], hue=clusters, palette='viridis')
plt.title('Customer Segmentation')
plt.show()
```



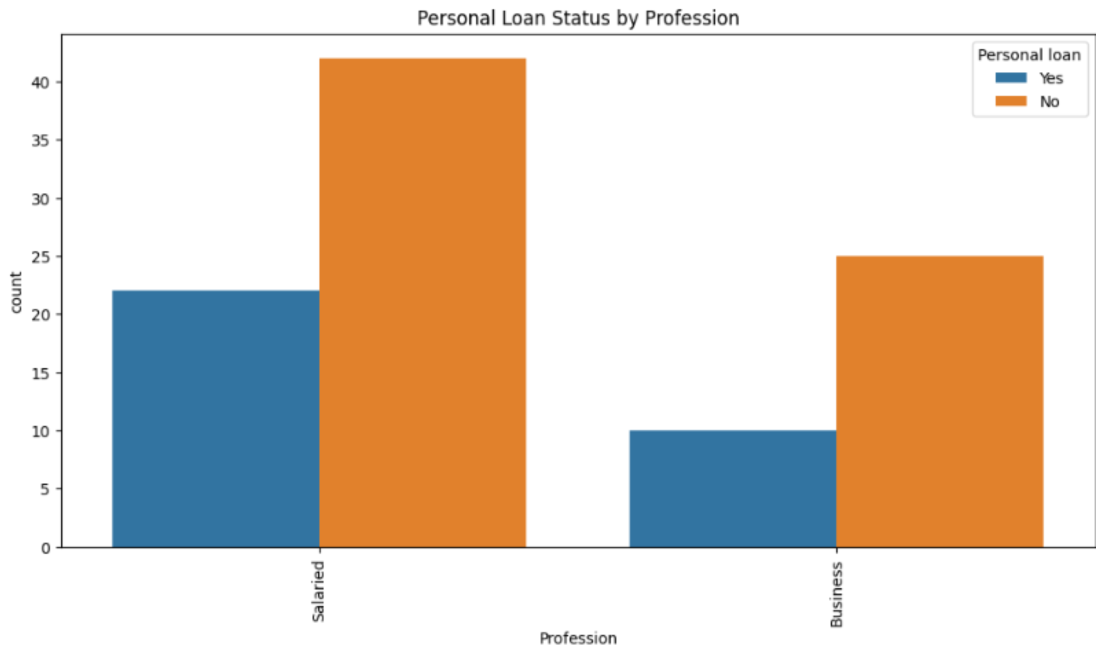
12. Psychographic segmentation through income and price

```
# Psychographic Segmentation: Income and Price Preferences
plt.figure(figsize=(10, 6))
sns.boxplot(data=buying_behaviour, x='Profession', y='Price')
plt.xticks(rotation=90)
plt.title('Price Preference by Profession')
plt.show()
```


0s
Price Preference by Profession
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Price
Salaried
Business
Profession

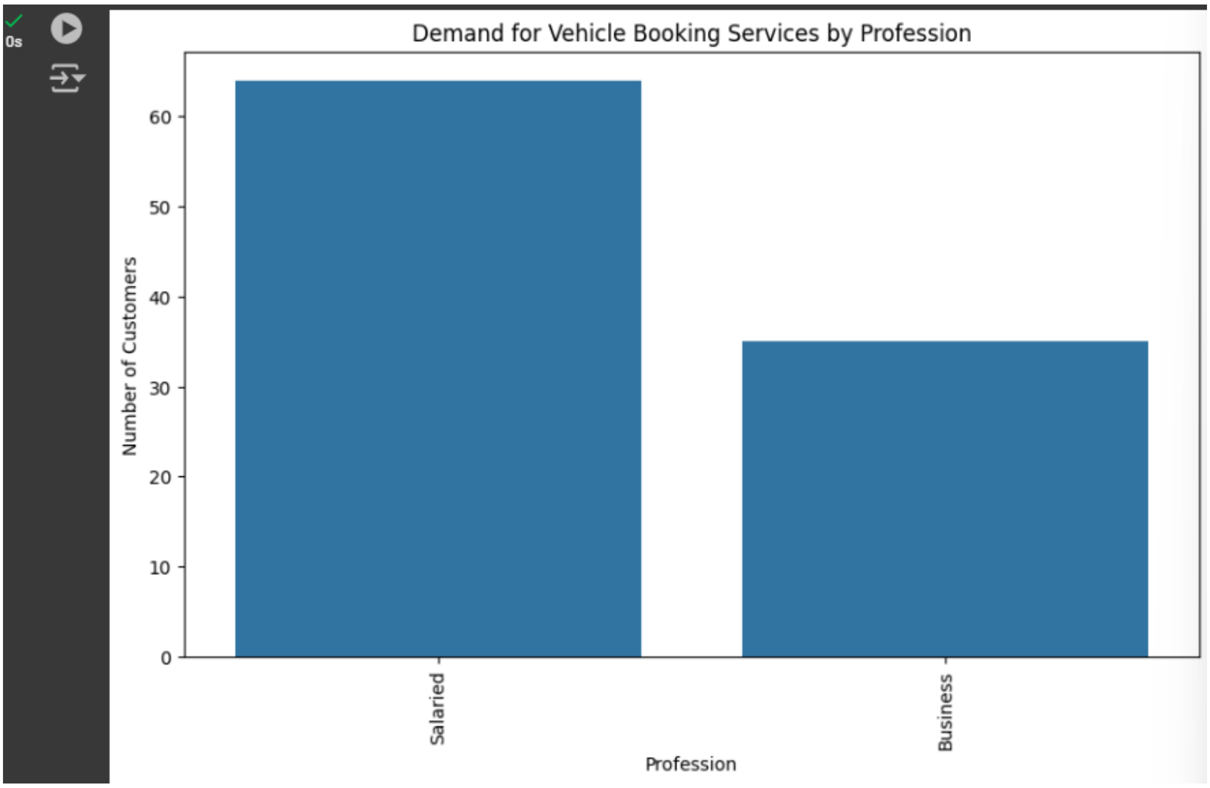
13. Behavioural segmentation through profession and personal loan

```
[160] # Behavioural Segmentation: Profession vs. Personal Loan Status
      plt.figure(figsize=(12, 6))
      sns.countplot(data=buying_behaviour, x='Profession', hue='Personal loan')
      plt.xticks(rotation=90)
      plt.title('Personal Loan Status by Profession')
      plt.show()
```



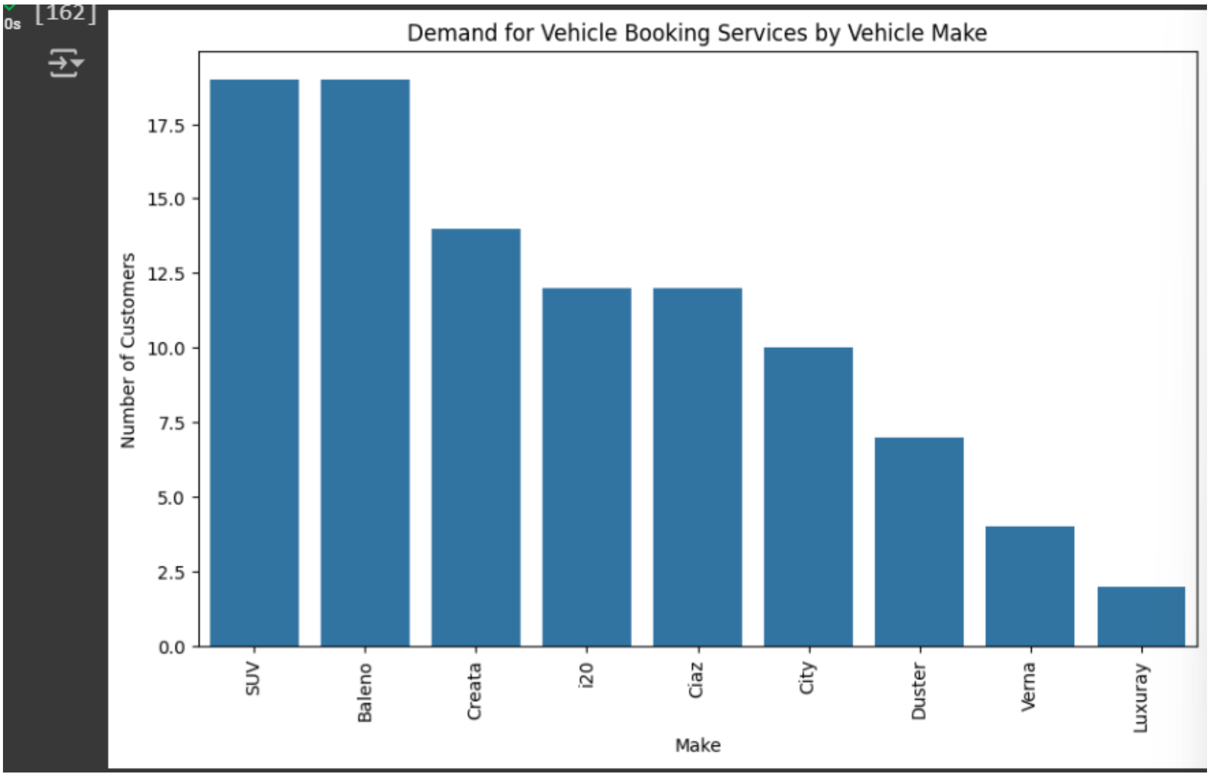
14. Analyse the demand for vehicle booking services by profession

```
# Count the number of people in each profession and vehicle make category
profession_counts = buying_behaviour['Profession'].value_counts()
make_counts = buying_behaviour['Make'].value_counts()

# Plot the demand by Profession
plt.figure(figsize=(10, 6))
sns.barplot(x=profession_counts.index, y=profession_counts.values)
plt.xticks(rotation=90)
plt.title('Demand for Vehicle Booking Services by Profession')
plt.ylabel('Number of Customers')
plt.show()
```



15. Analyse the demand for vehicle booking services by vehicle make

```
[162] # Plot the demand by Vehicle Make
      plt.figure(figsize=(10, 6))
      sns.barplot(x=make_counts.index, y=make_counts.values)
      plt.xticks(rotation=90)
      plt.title('Demand for Vehicle Booking Services by Vehicle Make')
      plt.ylabel('Number of Customers')
      plt.show()
```



16. Estimation of profit potential based on salary and vehicle price

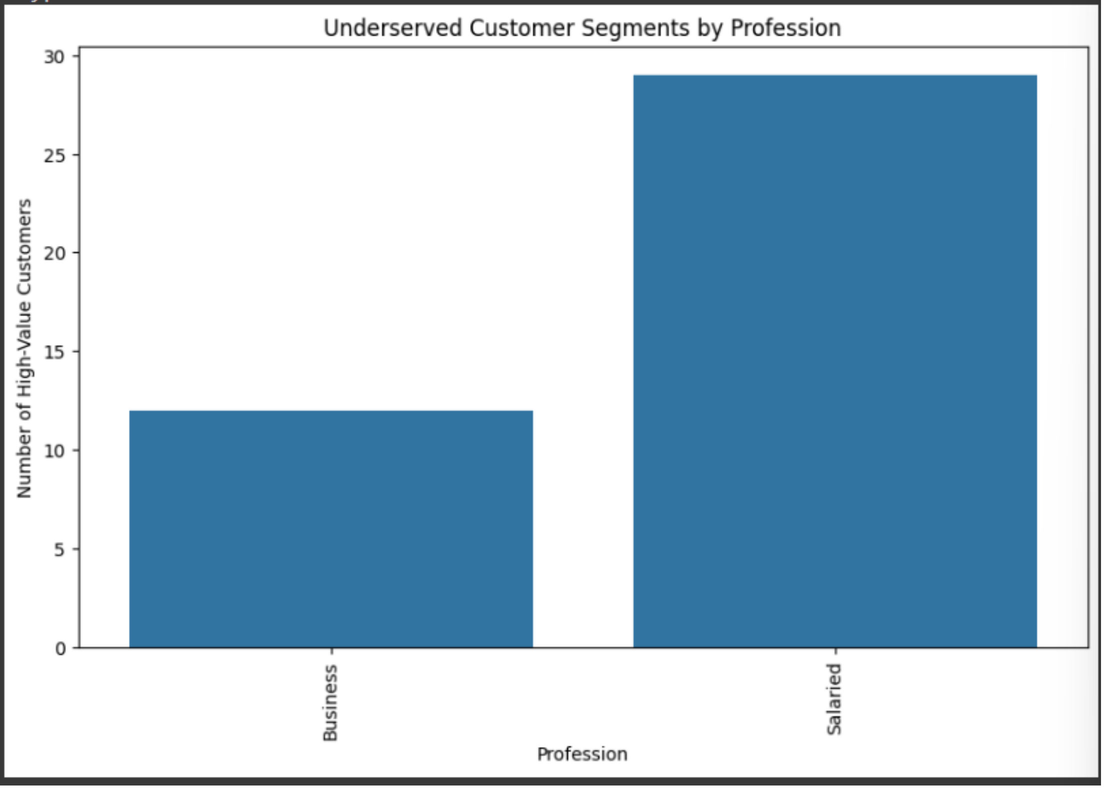
```
# Calculate potential profit for each customer
buying_behaviour['Potential Profit'] = buying_behaviour['Price'] * 0.1  # Assume 10% of vehicle price as profit margin

# Sum up the potential profit for each segment
profit_by_profession = buying_behaviour.groupby('Profession')['Potential Profit'].sum()
profit_by_make = buying_behaviour.groupby('Make')['Potential Profit'].sum()

# Display profit potential by profession
print("Profit potential by Profession:")
print(profit_by_profession)

# Display profit potential by vehicle make
print("\nProfit potential by Vehicle Make:")
print(profit_by_make)
```

```
Profit potential by Profession:
Profession
Business    3920000.0
Salaried    7901000.0
Name: Potential Profit, dtype: float64

Profit potential by Vehicle Make:
Make
Baleno     1340000.0
Ciaz       1310000.0
City       1200000.0
Creata     2100000.0
Duster      920000.0
Luxuray     600000.0
SUV        3030000.0
Verna       361000.0
i20         960000.0
Name: Potential Profit, dtype: float64
```

17. Identify underserved segments based on salary and vehicle price

```python
# Target underserved customer segments where price is high and number of customers is low

high_value_customers = buying_behaviour[buying_behaviour['Price'] > buying_behaviour['Price'].median()]
underserved_segments = high_value_customers.groupby('Profession').size().sort_values(ascending=True)

print("\nUnderserved segments based on Profession:")
print(underserved_segments)

# Plot underserved segments
plt.figure(figsize=(10, 6))
sns.barplot(x=underserved_segments.index, y=underserved_segments.values)
plt.xticks(rotation=90)
plt.title('Underserved Customer Segments by Profession')
plt.ylabel('Number of High-Value Customers')
plt.show()
```

```
Underserved segments based on Profession:
Profession
Business     12
Salaried     29
dtype: int64
```



CONCLUSION

Based on the comprehensive market segmentation analysis conducted for the electric vehicle (EV) startup in India, some of the key insights are as follows:

1. **Market Potential**:
   - The analysis of EV population data reveals significant variations in EV adoption across different Indian cities, with some showing higher concentrations of EVs. This information can guide targeted market entry strategies.
   - There's a clear correlation between income levels and EV preferences, suggesting a need for tailored product offerings for different income segments.
2. **Infrastructure Considerations**:
   - The uneven distribution of charging stations across states highlights potential opportunities and challenges. Areas with lower station density might represent untapped markets but may require infrastructure investments.
3. **Consumer Segmentation**:
   - Demographic factors such as age, income, and profession play crucial roles in EV adoption. The clustering analysis revealed distinct customer segments with varying preferences and purchasing power.

    - The psychographic segmentation based on income and vehicle price preferences provides valuable insights for product positioning and pricing strategies.
4. **Behavioral Insights**:
    - The analysis of vehicle booking services demand across different professions and vehicle makes offers guidance for marketing and partnership strategies.
    - Loan status correlations with income levels suggest potential for targeted financing options to boost adoption.
5. **Profit Potential and Underserved Segments**:
    - The estimation of profit potential based on salary and vehicle price identifies lucrative market segments.
    - The analysis also revealed underserved segments, presenting opportunities for tailored products or services to capture these markets.
6. **Strategic Implications**:
    - The startup should consider a phased market entry, focusing initially on cities with high EV populations and adequate charging infrastructure.
    - Product development and pricing strategies should align with the identified customer segments, particularly considering the income-price relationships observed.

- Partnerships with charging infrastructure providers could be crucial in underserved areas.
- Tailored marketing and financing options for different professional groups could drive adoption.

In conclusion, this market analysis provides a solid foundation for the EV startup to develop a targeted, data-driven strategy for entering and expanding in the Indian market. By leveraging these insights, the company can position itself effectively in this rapidly evolving sector.