

In [31]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans

In [41]: buyers=pd.read_csv("Indian automobile buying behaviour study 1.0.csv")
location=pd.read_csv("Location_data.csv")

In [101]: buyers.head()

Out [101]:
Age Profession Marital Status Education No of Dependents Personal loan House Loan Wife Working Salary Wife Salary Total Salary Make Price
0 27 Salafied Married Single Post Graduate 0 Yes No No No 800000 0 800000 i20 800000
1 35 Salafied Married Post Graduate 2 Yes Yes Yes 1400000 600000 2000000 Claz 1000000
2 45 Business Married Graduate 4 Yes Yes No 1800000 0 1800000 Duster 1200000
3 41 Business Married Post Graduate 3 No No No Yes 1600000 600000 2200000 City 1200000
4 31 Salafied Married Post Graduate 2 Yes No No Yes 1800000 800000 2600000 SUV 1600000

In [111]: location.head()

Out [111]:
State Name Total Electric Vehicle Total Non-Electric Vehicle Total No. of Operational PCS
0 Andaman & Nicobar 162 146945 147107 3
1 Arunachal Pradesh 20 252965 252985 9
2 Assam 64766 4877053 4741819 48
3 Bihar 83335 10407078 10409413 83
4 Chandigarh 2812 746881 746953 6

In [121]: buyers.describe()

Out [121]:
Age No of Dependents Salary Wife Salary Total Salary Price
count 99.000000 99.000000 9.800000e+01 9.900000e+01 9.900000e+01 9.900000e+01
mean 36.313131 2.181818 1.793644e+06 5.343434e+05 2.270770e+06 1.194040e+06
std 6.246054 1.335265 6.786217e+05 6.054450e+05 1.050777e+06 4.379555e+05
min 26.000000 0.000000 2.000000e+05 0.000000e+00 2.000000e+05 1.100000e+05
25% 31.000000 2.000000 1.300000e+06 0.000000e+00 1.500000e+06 8.000000e+05
50% 36.000000 2.000000 1.600000e+06 5.000000e+05 2.100000e+06 1.200000e+06
75% 41.000000 3.000000 2.200000e+06 2.700000e+06 5.200000e+06 1.500000e+06
max 51.000000 4.000000 3.800000e+06 2.100000e+06 5.200000e+06 3.000000e+06

In [131]: location.describe()

Out [131]:
Total Electric Vehicle Total Non-Electric Vehicle Total No. of Operational PCS
count 30.000000 3.000000e+01 3.000000e+01 30.000000
mean 4447.933333 9.260523e+06 9.305001e+06 194.133333
std 69664.61668 1.081429e+07 1.096983e+07 364.077269
min 20.000000 9.718900e+04 9.721000e+04 1.000000
25% 1418.500000 6.742398e+05 6.818892e+05 16.500000
50% 18888.500000 5.581995e+05 5.622784e+06 48.000000
75% 60766.250000 1.371388e+07 1.375666e+07 194.250000
max 337180.000000 4.009248e+07 4.042967e+07 1845.000000

In [141]: pd.DataFrame(location.drop(axis=0, columns="State Name").sum())

Out [141]:
Total Electric Vehicle 1334338
Total Non-Electric Vehicle 277815703
Total 279150441
No. of Operational PCS 5824

In [151]: plt.figure(figsize=(12,5))
plt.title('State vs Total Electric Vehicles')
sns.barplot(x='State Name',y='Total Electric Vehicle',data=location)
plt.setp(ax.get_xticklabels(),rotation=75)
plt.show()

State vs Total Electric Vehicles
Bar chart showing the total number of electric vehicles across various Indian states. The y-axis represents the total number of electric vehicles, ranging from 0 to 35,000. The x-axis lists the states. The states are ordered by the total number of electric vehicles in descending order. The states with the highest number of electric vehicles are Uttar Pradesh (approx. 35,000), West Bengal (approx. 15,000), and Tamil Nadu (approx. 12,000). Other states with significant numbers include Karnataka, Kerala, and Maharashtra.

In [161]: plt.figure(figsize=(15,5))
plt.title('State vs Total Electric Vehicles')
sns.barplot(x='State Name',y='Total Non-Electric Vehicle',data=location)
plt.setp(ax.get_xticklabels(),rotation=75)
plt.show()

State vs Total Electric Vehicles
Bar chart showing the total number of non-electric vehicles across various Indian states. The y-axis represents the total number of non-electric vehicles, ranging from 0 to 4e7. The x-axis lists the states. The states are ordered by the total number of non-electric vehicles in descending order. The states with the highest number of non-electric vehicles are Uttar Pradesh (approx. 4e7), West Bengal (approx. 1.5e7), and Tamil Nadu (approx. 1.2e7). Other states with significant numbers include Karnataka, Kerala, and Maharashtra.

In [171]: location['Percentage 1']=location['Total Electric Vehicle']/location['Total']

In [181]: plt.figure(figsize=(15,5))
plt.title('State vs Percentage of Electric Vehicles')
sns.barplot(x='State Name',y='Percentage 1',data=location)
plt.setp(ax1.get_xticklabels(),rotation=90)
plt.show()

State vs Percentage of Electric Vehicles
Bar chart showing the percentage of electric vehicles across various Indian states. The y-axis represents the percentage, ranging from 0.0000 to 0.0200. The x-axis lists the states. The states are ordered by the percentage of electric vehicles in descending order. The states with the highest percentage of electric vehicles are Uttar Pradesh (approx. 0.0200), West Bengal (approx. 0.0150), and Tamil Nadu (approx. 0.0120). Other states with significant percentages include Karnataka, Kerala, and Maharashtra.

In [191]: location['Percentage 2']=location['Total Non-Electric Vehicle']/location['Total']

In [201]: plt.figure(figsize=(15,5))
plt.title('State vs Percentage of Electric Vehicles')
sns.barplot(x='State Name',y='Percentage 2',data=location)
plt.setp(ax2.get_xticklabels(),rotation=90)
plt.show()

State vs Percentage of Electric Vehicles
Bar chart showing the percentage of non-electric vehicles across various Indian states. The y-axis represents the percentage, ranging from 0.0 to 1.0. The x-axis lists the states. The states are ordered by the percentage of non-electric vehicles in descending order. The states with the highest percentage of non-electric vehicles are Uttar Pradesh (approx. 0.9800), West Bengal (approx. 0.9700), and Tamil Nadu (approx. 0.9600). Other states with significant percentages include Karnataka, Kerala, and Maharashtra.

In [211]: charging_stations = pd.read_csv('charging_stations_state_wise.csv')

In [221]: charging_stations.head()

Out [221]:
Sl. No. StateUT No. of Operational PCS
0 1 Andaman and Nicobar 3
1 2 Andhra Pradesh 222
2 3 Arunachal Pradesh 9
3 4 Assam 48
4 5 Bihar 83

In [231]: plt.figure(figsize=(15,5))
plt.title('State vs Total Charging Stations')
sns.barplot(x='State/UT',y='No. of Operational PCS',data=charging_stations)
plt.setp(ax.get_xticklabels(),rotation=90)
plt.grid(axis='y', linestyle='--')
plt.show()

State vs Total Charging Stations
Bar chart showing the total number of charging stations across various Indian states. The y-axis represents the number of operational PCS, ranging from 0 to 6000. The x-axis lists the states. The states are ordered by the total number of charging stations in descending order. The states with the highest number of charging stations are Uttar Pradesh (approx. 6000), West Bengal (approx. 1500), and Tamil Nadu (approx. 1000). Other states with significant numbers include Karnataka, Kerala, and Maharashtra.

In [241]: Sold_Evs_India = pd.read_csv('sold_EVs.csv')

In [251]: Sold_Evs_India.head()

Out [251]:
Sl. No. Wheeler Type Total No. of Vehicle
0 1 2 Wheeler 77573
1 2 3 Wheeler 87570
2 3 4 Wheeler 8997
3 Total Total 872920

In [261]: wheeler_types = Sold_Evs_India['Wheeler Type']
total_vehicles = Sold_Evs_India['Total No. of Vehicle']

plt.figure(figsize=(5, 6))
bars = plt.bar(wheeler_types, total_vehicles, color='red')

for bar in bars:
 yval = bar.get_height()
 plt.text(bar.get_x() + bar.get_width()/2, yval, round(yval), va='bottom', ha='center', color='black', fontsize=9)

plt.title('Total Number of Electric Vehicles by Wheeler Type')
plt.xlabel('Wheeler Type')
plt.ylabel('Total Number of Vehicles')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--')
plt.tight_layout()
plt.show()

Total Number of Electric Vehicles by Wheeler Type
Bar chart showing the total number of electric vehicles by wheeler type. The y-axis represents the total number of vehicles, ranging from 0 to 80,000. The x-axis lists the wheeler types. The states with the highest number of electric vehicles are 3 Wheeler (approx. 77,573), 2 Wheeler (approx. 87,570), and 4 Wheeler (approx. 8,997).

In [271]: statewide_vehicles = pd.read_csv('state_wise_EVs.csv')

In [281]: statewide_vehicles.head()

Out [281]:
Sl. No. StateUT Till date State wise - Total Number of Vehicles Registered Till date State wise - Total Vehicle Registered as Electric Proportion of Electric Vehicle
0 1 Andaman and Nicobar Islands 160375 186 0.12
1 2 Andhra Pradesh 16517516 66500 0.40
2 3 Arunachal Pradesh 299371 25 0.01
3 4 Assam 5393842 116605 2.16
4 5 Bihar 11728184 155457 1.33

In [291]: states = statewide_vehicles['State/UT']
total_registered = statewide_vehicles['Tilll date State wise - Total Number of Vehicles Registered']
total_ev_registered = statewide_vehicles['Tilll date State wise - Total Vehicle Registered as Electric']

plt.figure(figsize=(12,8))
plt.bar(states, total_registered, color='lightblue', label='Total Registered Vehicles')
plt.bar(states, total_ev_registered, color='darkblue', label='Total Electric Vehicles')
plt.yscale('log')
plt.title('Total Registered Vehicles vs Total Electric Vehicles by State/UT')
plt.xlabel('State/UT')
plt.ylabel('Number of Vehicles')
plt.xticks(rotation=90)
plt.grid(axis='y', linestyle='--')
plt.legend()
plt.grid(axis='y', linestyle='--')
plt.tight_layout()
plt.show()

Total Registered Vehicles vs Total Electric Vehicles by State/UT
Bar chart showing the total registered vehicles and total electric vehicles by state/UT. The y-axis represents the number of vehicles on a logarithmic scale, ranging from 10^0 to 10^6. The x-axis lists the states. The states with the highest number of registered vehicles are Uttar Pradesh, West Bengal, and Tamil Nadu.

In [346]: State_sales = pd.read_csv('Vehicles_Data.csv')

In [347]: State_sales = State_sales.drop(State_sales.index[33],axis=0)

In [348]: State_sales.head(n=40)

Out [348]:
Sl. No. State/UT 2020 2021 2022 2023 Grand Total
0 1 Andaman and Nicobar Islands 1650 92 23 18 169
1 2 Andhra Pradesh 36540 8738 29450 29546 70388
2 3 Arunachal Pradesh 50 2 2 17 26
3 4 Assam 83570 15634 40719 56448 121156
5 6 Bihar 124470 23062 55793 79469 190749
6 7 Chhattisgarh 14890 4215 22364 35115 63183
7 8 Delhi 123770 25816 62264 64382 164831
8 9 Goa 820 9765 5688 8637 15503
9 10 Gujarat 11200 8765 68998 83965 163848
10 11 Haryana 29820 8660 25864 27465 64971
12 12 Himachal Pradesh 1810 327 1008 1003 2519
13 13 Jammu and Kashmir 740 1146 4689 8823 13805
14 14 Jharkhand 15150 3741 13862 19096 38732
15 15 Karnataka 97070 33307 95900 141057 279971
16 16 Kerala 13620 8743 39602 69624 119351
17 17 Ladakh NaN NaN 6084 218 74
18 18 Madhya Pradesh 33580 10427 36804 61678 112267
19 19 Maharashtra 71340 29914 130501 178488 351587
20 20 Manipur 1040 114 341 386 945
21 21 Mizoram 10 1 36 144 162
22 22 Nagaland 110 2 3 6 22
23 23 Odisha 9040 5626 28446 39839 74815
24 25 Puducherry 880 405 1481 2405 4379
26 26 Punjab 8320 4643 14053 23378 42906
27 27 Rajasthan 56040 23464 78238 86376 193822
28 28 Tamil Nadu 56970 30300 66953 84252 186822
29 29 Dadra and Nagar Haveli and Daman and Diu 240 29 141 142 336
30 30 Uttar Pradesh 312680 66705 162862 248223 510509
31 31 Uttarakhand 23950 5324 15561 15330 39610
32 33 West Bengal 100790 6408 11150 17699 45336

In [349]: df1 = State_sales['2020'].sum()

In [350]: df1

Out [350]:
124681.0

In [351]: df2 = State_sales['2021'].sum()

In [352]: df2

Out [352]:
331635

In [353]: df3 = State_sales['2022'].sum()

In [354]: df3

Out [354]:
1025118

In [355]: df4 = State_sales['2023'].sum()

In [356]: df4

Out [356]:
1395701

In [357]: plt.figure(figsize=(15,5))
plt.title('State vs 2023 Vehcles Number')
sns.barplot(x='State/UT',y='2023',data=State_sales)
plt.setp(ax.get_xticklabels(),rotation=90)
plt.show()

State vs 2023 Vehcles Number
Bar chart showing the number of vehicles in 2023 across various Indian states. The y-axis represents the number of vehicles, ranging from 0 to 25,000. The x-axis lists the states. The states with the highest number of vehicles in 2023 are Uttar Pradesh (approx. 25,000), West Bengal (approx. 15,000), and Tamil Nadu (approx. 12,000). Other states with significant numbers include Karnataka, Kerala, and Maharashtra.

In [366]: buyers.head(n=10)

Out [366]:
Age Profession Marital Status Education No of Dependents Personal loan House Loan Wife Working Salary Wife Salary Total Salary Make Price
0 27 Salafied Married Single Post Graduate 0 Yes No No No 800000 0 800000 i20 800000
1 35 Salafied Married Post Graduate 2 Yes Yes Yes 1400000 600000 2000000 Claz 1000000
2 45 Business Married Graduate 4 Yes Yes No 1800000 0 1800000 Duster 1200000
3 41 Business Married Post Graduate 3 No No No Yes 1600000 600000 2200000 SUV 1200000
4 31 Salafied Married Post Graduate 2 Yes No No Yes 1800000 800000 2600000 SUV 1600000
5 28 Salafied Married Graduate 3 Yes Yes No 900000 0 900000 Baleno 700000
6 31 Salafied Married Graduate 4 No No No Yes 1200000 600000 1800000 City 1200000
7 33 Business Married Post Graduate 4 No No No No 1400000 0 1400000 Baleno 700000
8 34 Business Married Post Graduate 4 No No No No 2000000 0 2000000 Verna 1100000
9 34 Salafied Married Graduate 3 Yes Yes Yes Yes 1200000 700000 1900000 i20 800000

In [359]: plt.figure(figsize=(10,5))
sns.countplot(buyers['Make'])

Out [359]:
C:\Users\pca\pytho\lib\site-packages\seaborn\decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be "data", and passing other arguments without an explicit keyword will result in an error or misinterpretation.
sns.countplot(buyers['Make'])

Out [359]:
<AxesSubplot: xlabel='Make', ylabel='count'>

Make vs Count
Bar chart showing the count of vehicles by make. The y-axis represents the count, ranging from 0.0 to 17.5. The x-axis lists the makes. The makes with the highest count are i20, Claz, Duster, City, SUV, Make, Baleno, Verna, and Luxury.

In [364]: plt.figure(figsize=(10,5))
sns.boxplot(data=buyers,x='Price',y='Make')

Out [364]:
<AxesSubplot: xlabel='Price', ylabel='Make'>

Price vs Make
Box plot showing the distribution of vehicle prices by make. The x-axis represents the price, ranging from 0.0 to 3.0. The y-axis lists the makes. The makes are i20, Claz, Duster, City, SUV, Make, Baleno, Verna, and Luxury.

In [371]: buyers.describe()

Out [371]:
Age No of Dependents Salary Wife Salary Total Salary Price
count 99.000000 99.000000 9.800000e+01 9.900000e+01 9.900000e+01 9.900000e+01
mean 36.313131 2.181818 1.793644e+06 5.343434e+05 2.270770e+06 1.194040e+06
std 6.246054 1.335265 6.786217e+05 6.054450e+05 1.050777e+06 4.379555e+05
min 26.000000 0.000000 2.000000e+05 0.000000e+00 2.000000e+05 1.100000e+05
25% 31.000000 2.000000 1.300000e+06 0.000000e+00 1.500000e+06 8.000000e+05
50% 36.000000 2.000000 1.600000e+06 5.000000e+05 2.100000e+06 1.200000e+06
75% 41.000000 3.000000 2.200000e+06 2.700000e+06 5.200000e+06 1.500000e+06
max 51.000000 4.000000 3.800000e+06 2.100000e+06 5.200000e+06 3.000000e+06

In [373]: x= plt.figure(figsize=(15,8))
sns.heatmap(buyers.corr(),linewidths=1,linecolor='white',annot=True)

Out [373]:
<AxesSubplot: >

Correlation Heatmap
Heatmap showing the correlation between variables. The x-axis and y-axis list the variables: Age, No of Dependents, Salary, Wife Salary, Total Salary, and Price. The color scale ranges from -0.2 (dark blue) to 1.0 (dark red).

In [374]: sns.pairplot(buyers)

Out [374]:
<seaborn.axisgrid.PairGrid at 0x167870>

Pair Plot
Pair plot showing the distribution and relationships between variables. The variables are Age, No of Dependents, Salary, Wife Salary, Total Salary, and Price. The diagonal shows histograms, and the off-diagonal shows scatter plots.

In []: