Project Title: Customer Segmentation and Car Preference Analysis for Optimizing Marketing Strategy A Case Study of Austo Motor Company

Domain: Automotive & Marketing Analytics

Tools Used – Python (Numpy, Pandas, Matplotlib Library, Seaborn Library)

Objective:

To analyze the automobile market data of Austo Motor Company to identify customer segments, understand preferences for car models (SUV, Sedan, Hatchback), uncover key behavioral and income-based trends, and provide actionable strategies to improve marketing efficiency, increase sales, and align product offerings with buyer expectations.

Made By – Abhijeet Beura

Austo Motor Company is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models. In its

recent board meeting, concerns were raised by the members on the efficiency of the marketing campaign

currently being used. The board decides to rope in analytics professional to improve the existing campaign.

Imported the libraries for the Data are

Numpy

Pandas

Matplotlib

Seaborn

1. There are some information about the dataset, decision makers should have a look.

The dataset is having 1581 rows and 14 columns.

There is a look on the 5 sample rows to check the data type.

Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price	Make
53	Male	Business	Married	Post Graduate	4	No	No	Yes	99300	70700.0	170000	61000	SUV
53	Femal	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500	70300.0	165800	61000	SUV
53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300	60700.0	158000	57000	SUV
53	Female	Salaried	Married	Graduate	2	Yes	No	Yes	72500	70300.0	142800	61000	SUV
53	Male	Salaried	Married	Post Graduate	3	No	No	Yes	79700	60200.0	139900	57000	SUV

Table 1: Top five rows of the dataset

2. While having a look on the data set information, it is found that there are 6 numerical and 8 categorial variables. The below table contains the same information.

```
 0   Age              1581 non-null   int64
 1   Gender           1528 non-null   object
 2   Profession       1581 non-null   object
 3   Marital_status   1581 non-null   object
 4   Education        1581 non-null   object
 5   No_of_Dependents 1581 non-null   int64
 6   Personal_loan    1581 non-null   object
 7   House_loan       1581 non-null   object
 8   Partner_working  1581 non-null   object
 9   Salary           1581 non-null   int64
 10  Partner_salary   1475 non-null   float64
 11  Total_salary     1581 non-null   int64
 12  Price            1581 non-null   int64
 13  Make             1581 non-null   object
dtypes: float64(1), int64(5), object(8)
memory usage: 173.1+ KB
```

Table 2: Basic information of the data type

3. Checking the data information.

```
: <bound method NDFrame.describe of       Age  Gender Profession Marital_status      Education  No_of_Dependents  \
  0      53    Male   Business        Married  Post Graduate                 4
  1      53  Female   Salaried        Married  Post Graduate                 4
  2      53  Female   Salaried        Married  Post Graduate                 3
  3      53  Female   Salaried        Married       Graduate                 2
  4      53    Male   Salaried        Married  Post Graduate                 3
  ...   ...     ...        ...            ...            ...               ...
  1576   22    Male   Salaried         Single       Graduate                 2
  1577   22    Male   Business        Married       Graduate                 4
  1578   22    Male   Business         Single       Graduate                 2
  1579   22    Male   Business        Married       Graduate                 3
  1580   22    Male   Salaried        Married       Graduate                 4

       Personal_loan House_loan Partner_working  Salary  Partner_salary  \
  0               No         No             Yes   99300         70700.0
  1              Yes         No             Yes   95500         70300.0
  2               No         No             Yes   97300         60700.0
  3              Yes         No             Yes   72500         70300.0
  4               No         No             Yes   79700         60200.0
  ...            ...        ...             ...     ...             ...
  1576            No        Yes              No   33300             0.0
  1577            No         No              No   32000             0.0
  1578            No        Yes              No   32900             0.0
  1579           Yes        Yes              No   32200             0.0
  1580            No         No              No   31600             0.0

        Total_salary  Price       Make
  0           170000  61000        SUV
  1           165800  61000        SUV
  2           158000  57000        SUV
  3           142800  61000        SUV
  4           139900  57000        SUV
  ...            ...    ...        ...
```

Table 3: Basic information of the data

4. Checking the columns of the dataset, to get the name of the variables.

```
Gender
Gender
Male      1252
Female     329
Name: count, dtype: int64

Profession
Profession
Salaried    896
Business    685
Name: count, dtype: int64

Education
Education
Post Graduate    985
Graduate         596
Name: count, dtype: int64

Personal_loan
Personal_loan
Yes    792
No     789
Name: count, dtype: int64

House_loan
House_loan
No     1054
Yes     527
Name: count, dtype: int64
```

```
Partner_working
Partner_working
Yes    868
No     713
Name: count, dtype: int64

Make
Make
Sedan        702
Hatchback    582
SUV          297
Name: count, dtype: int64

Marital_status
Marital_status
Married    1443
Single      138
Name: count, dtype: int64
```

Table 4: Name of the columns present in the dataset

5. Checking the null values:

There are nulls in ‘Gender’ and ‘Partner_salary’ variables.

In ‘Gender’ it is found that there are total 53 null values.

In ‘Partner_salary’ it is found that there are total 106 null values.

```
Age                   0
Gender               53
Profession            0
Marital_status        0
Education             0
No_of_Dependents      0
Personal_loan         0
House_loan            0
Partner_working       0
Salary                0
Partner_salary      106
Total_salary          0
Price                 0
Make                  0
dtype: int64
```

Checking All the Columns & Values

Table 5: Inspecting null values in the dataset

In order to treat the nulls in the ‘Partner_salary’, we have checked where the ‘Total_salary’ is greater than ‘Salary’.

Then we applied condition that,

1.If , 'Total_salary'>'salary' then, 'Partner_salary' = 'Total_salary'-'Salary'

2. If,'Total_salary >/'Salary' then,'Partner_salary' = 0

7. Checking the values count of ‘Gender’, and found that

```
# as we can also see that there are spelling mistake in the gender column which can be rectified as followed
df['Gender'] = df['Gender'].replace('Femal','Female')
df['Gender'] = df['Gender'].replace('Femle','Female')
```

```
# mode of gender
df['Gender'].mode()
```

```
0    Male
Name: Gender, dtype: object
```

```
# impute null values of gender with mode
df['Gender'] = df['Gender'].fillna(df['Gender'].mode()[0])
```

```
print('Gender\n', df['Gender'].value_counts())
```

```
Gender
 Gender
Male      1252
Female     329
Name: count, dtype: int64
```

Table 7: After imputing the Gender

8. Now we are to have a look on statistical summery of the numeric variables of the dataset.

	Age	No_of_Dependents	Salary	Partner_salary	Total_salary	Price
count	1581.000000	1581.000000	1581.000000	1581.000000	1581.000000	1581.000000
mean	31.922201	2.457938	60392.220114	19233.776091	79625.996205	35597.722960
std	8.425978	0.943483	14674.825044	19670.391171	25545.857768	13633.636545
min	22.000000	0.000000	30000.000000	0.000000	30000.000000	18000.000000
25%	25.000000	2.000000	51900.000000	0.000000	60500.000000	25000.000000
50%	29.000000	2.000000	59500.000000	25100.000000	78000.000000	31000.000000
75%	38.000000	3.000000	71800.000000	38100.000000	95900.000000	47000.000000
max	54.000000	4.000000	99300.000000	80500.000000	171000.000000	70000.000000

Table 8: Statistical summary of numeric variables

Table 9: Value counts of categorial variables

Checking for the outliers or extreme values.

```
numerical_cols = ['Salary', 'Age', 'Partner_salary', 'Total_salary', 'Price']
df_melted = df[numerical_cols].melt(var_name='Variable', value_name='Value')

# Setting plot style
plt.figure(figsize=(12, 6))
sns.boxplot(x='Variable', y='Value', data=df_melted, palette='Set1')  # Vibrant color palette

# Adding title and formatting
plt.title('Box Plot for Salary, Age, Partner_Salary, Total_Salary, and Price', fontsize=11)
plt.xlabel('')
plt.ylabel('Value', fontsize=12)
plt.xticks(rotation=30)

# Add figure caption below the plot
plt.figtext(0.5, -0.1, 'Numerical variables', wrap=True,
            horizontalalignment='center', fontsize=12, fontweight='bold')

plt.tight_layout()
plt.show()
```

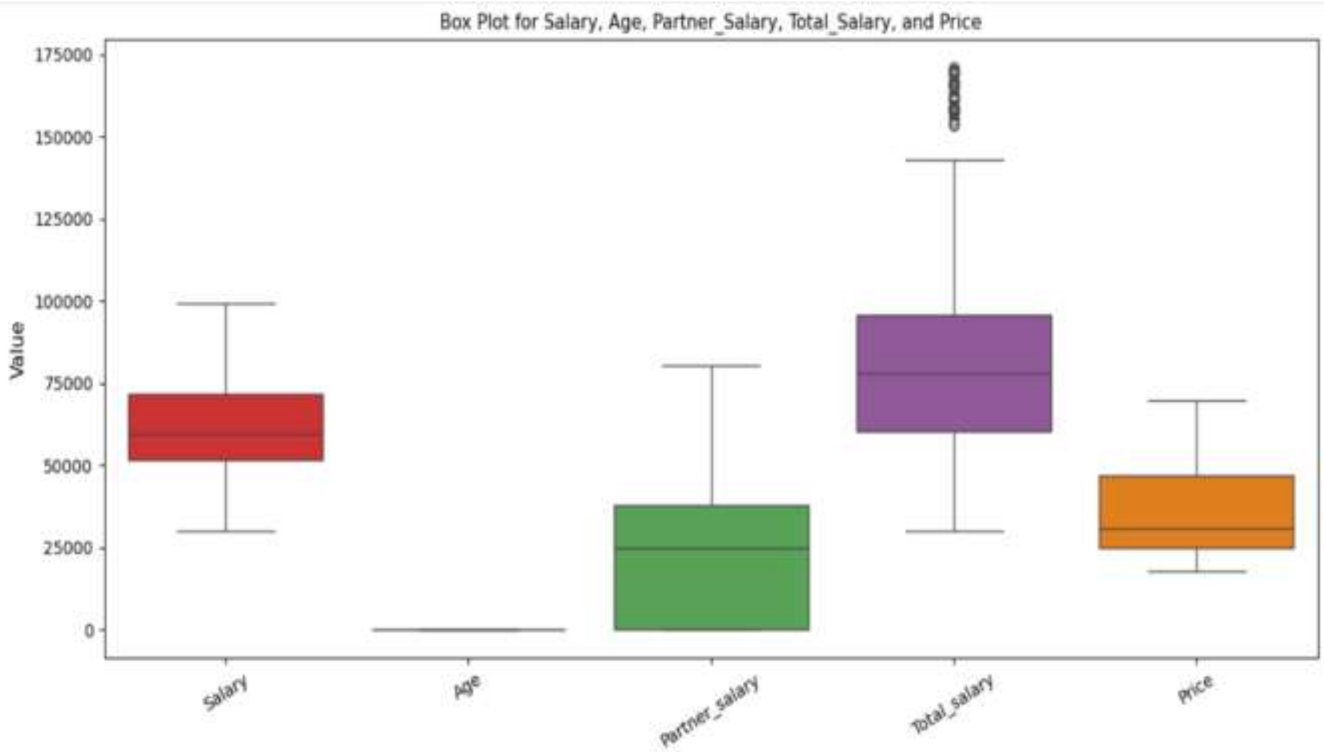
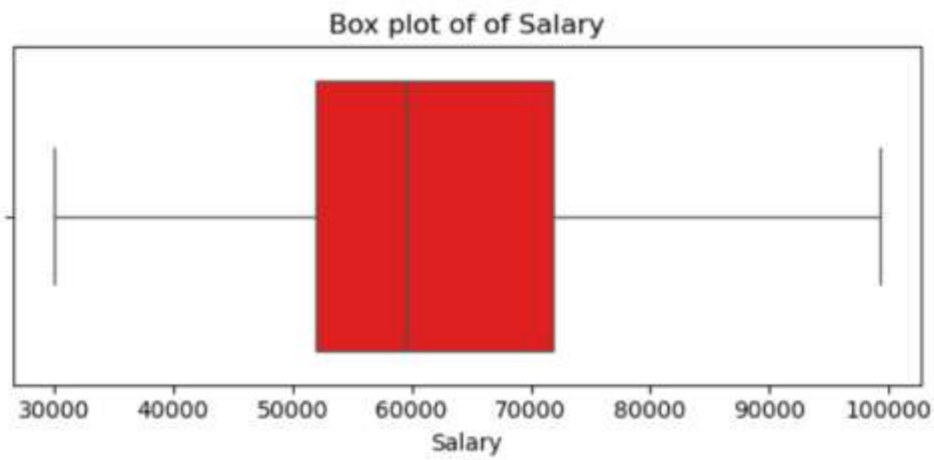
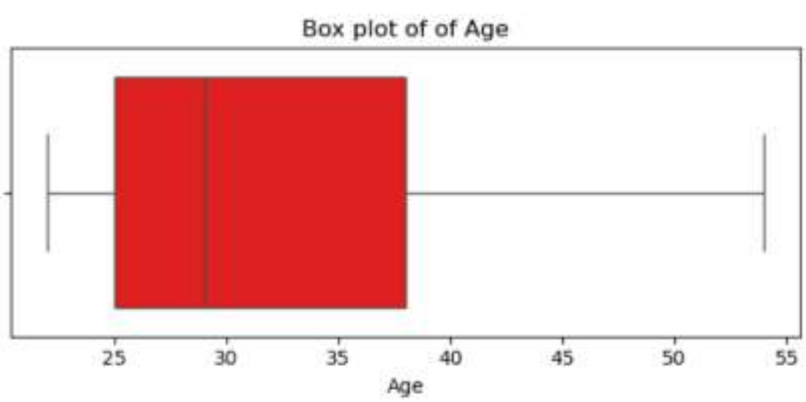


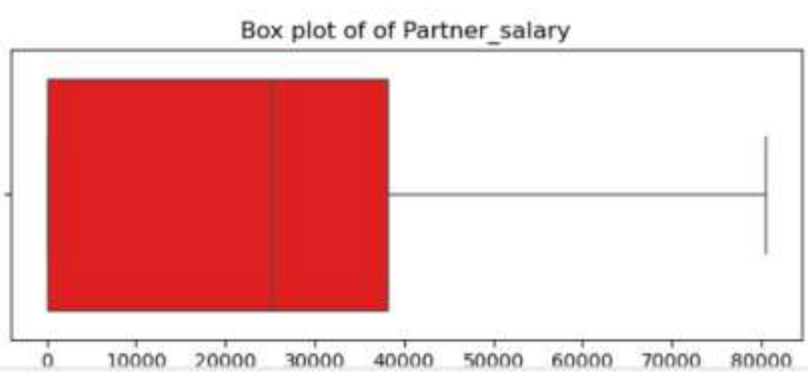
Figure-1 : Box plots of numerical variables

Analyzing box plots of every numerical variables separately:

```
numeric_cols = ['Salary','Age','Partner_salary','Total_salary','Price']
# loop over each numeric column and create a separate box plot
for col  in numeric_cols:
 plt.figure(figsize= (6,3))
 sns.boxplot(x=df[col],color='red')
 plt.title(f'Box plot of of {col}')
 plt.xlabel(col)
 plt.tight_layout()
 plt.show()
```







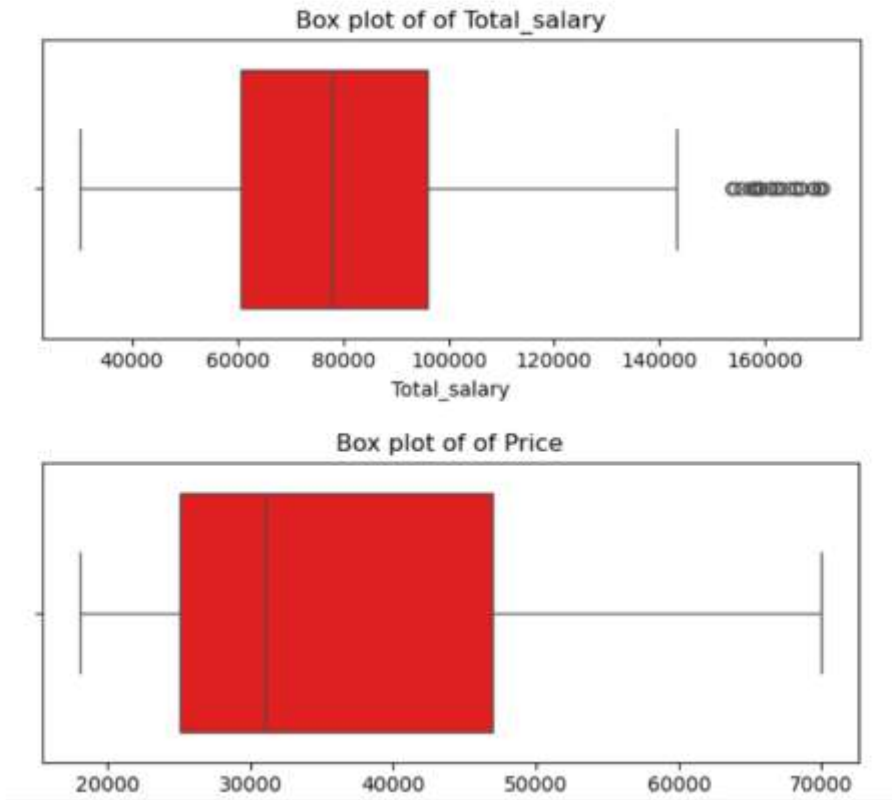
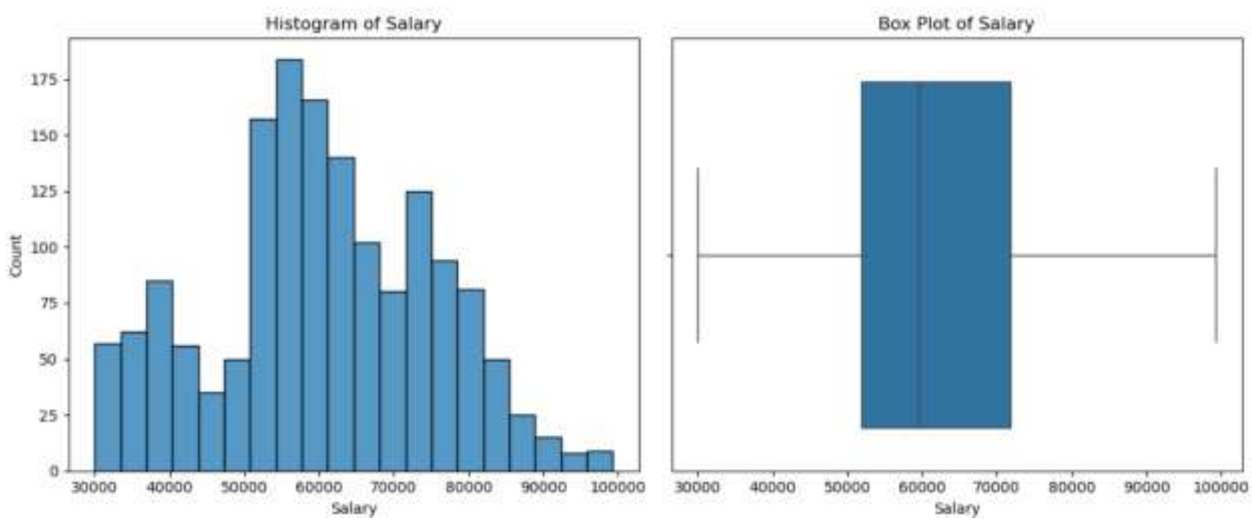
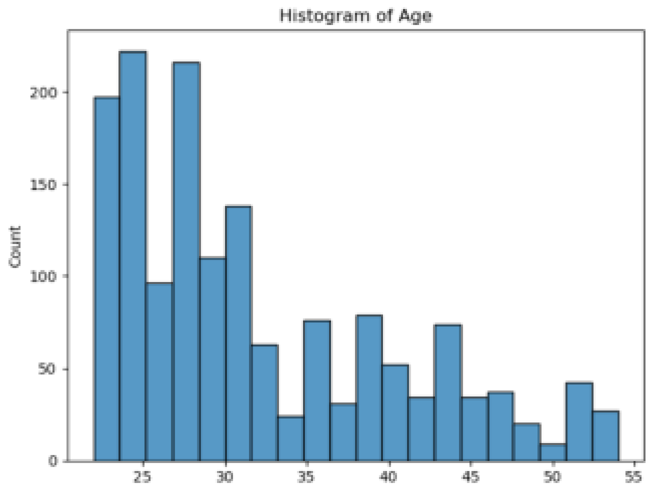
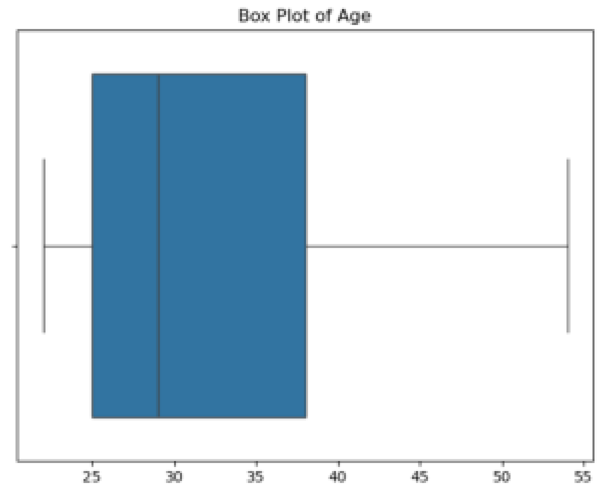


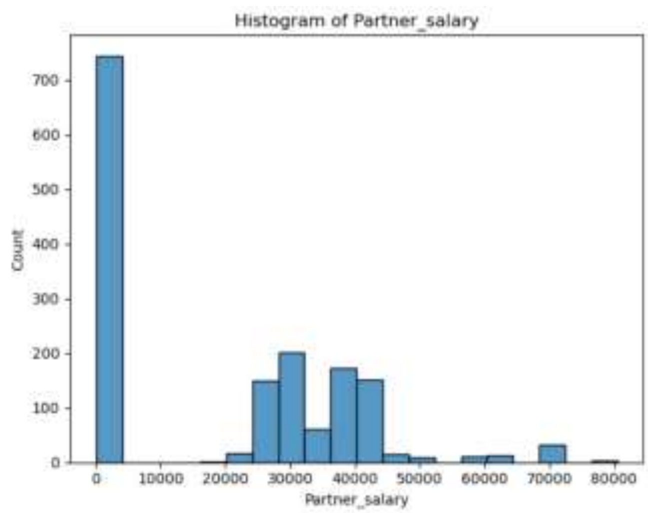
Figure-2 : Box plots of numerical variables individually 1. We can see that there are no negative values present in any numerical category. 2. The 'Total_salary' is having outlier.

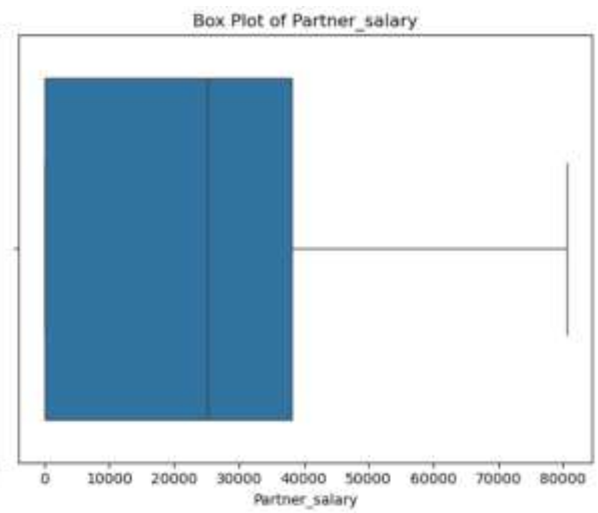
Univariate analysis of numerical variables.

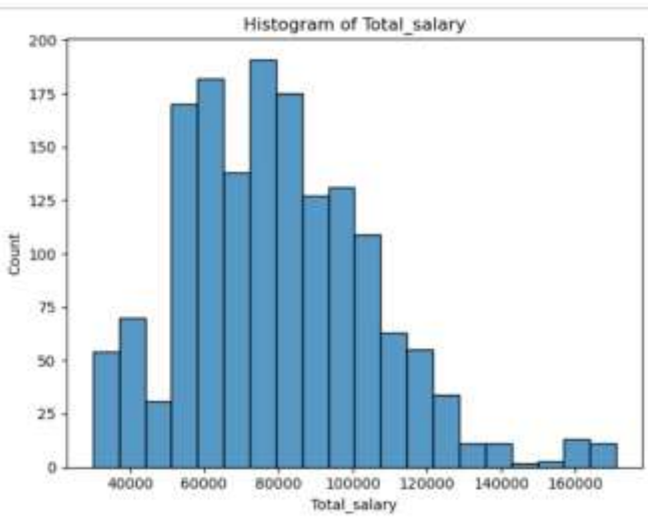


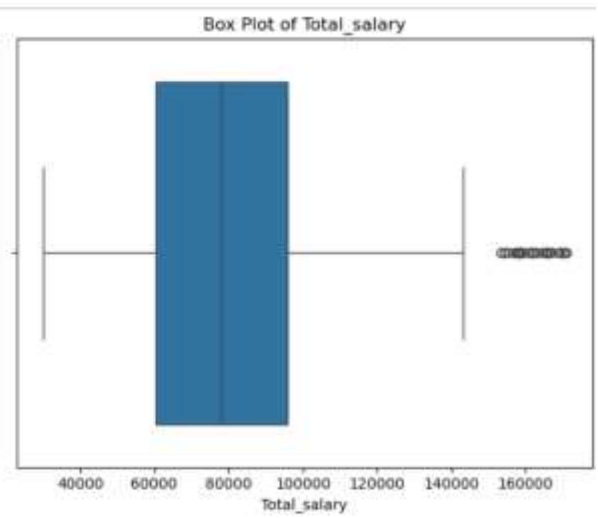

Histogram of Age
Count

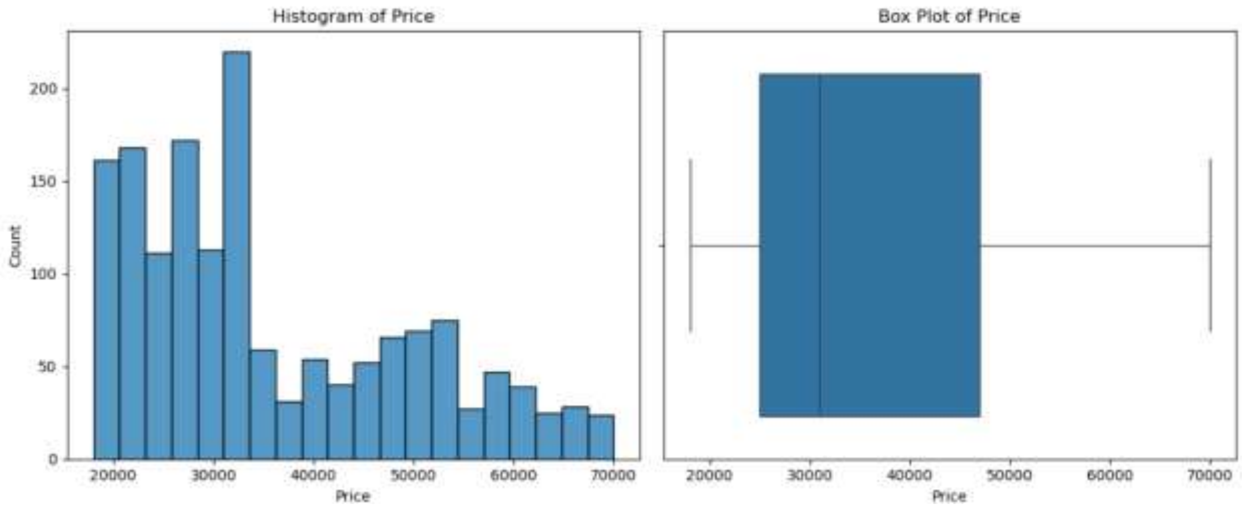

Box Plot of Age


Histogram of Partner_salary
Count
Partner_salary


Box Plot of Partner_salary
Partner_salary

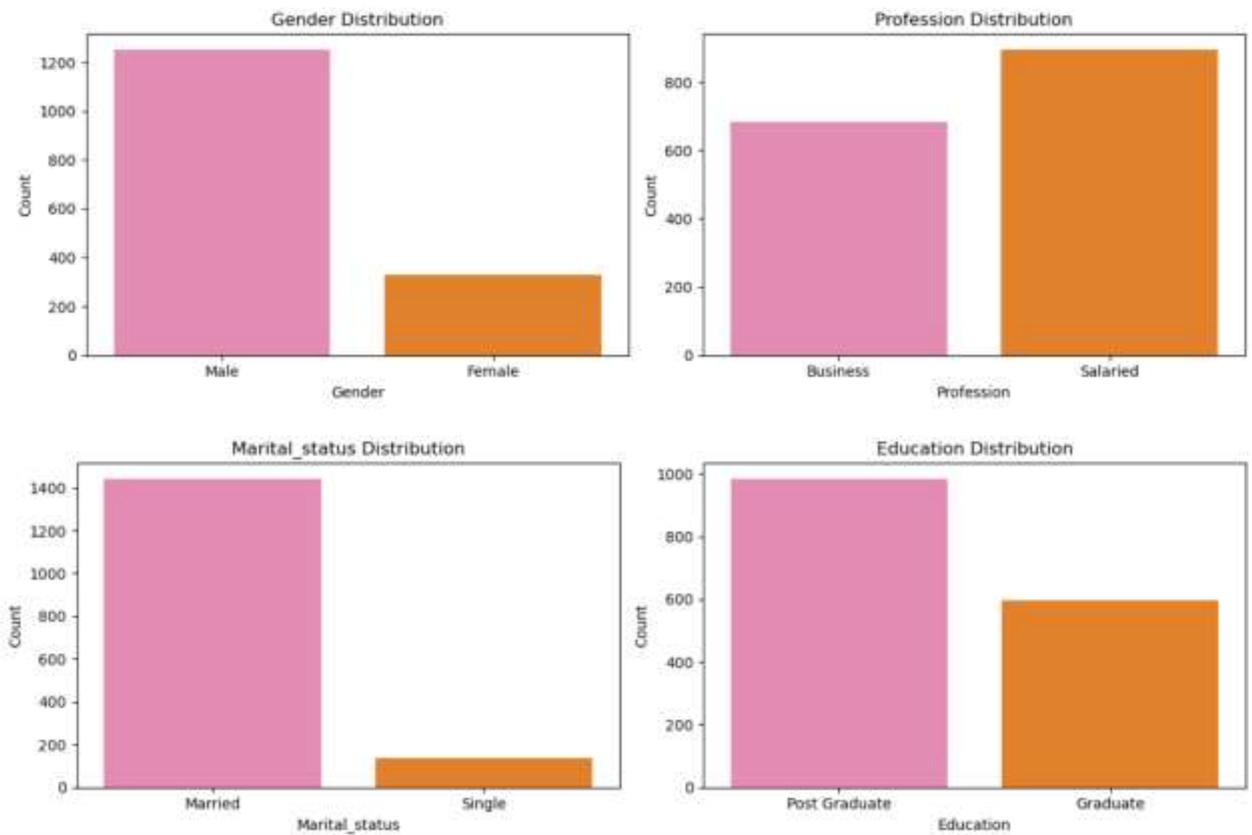

Histogram of Total_salary
Count
Total_salary

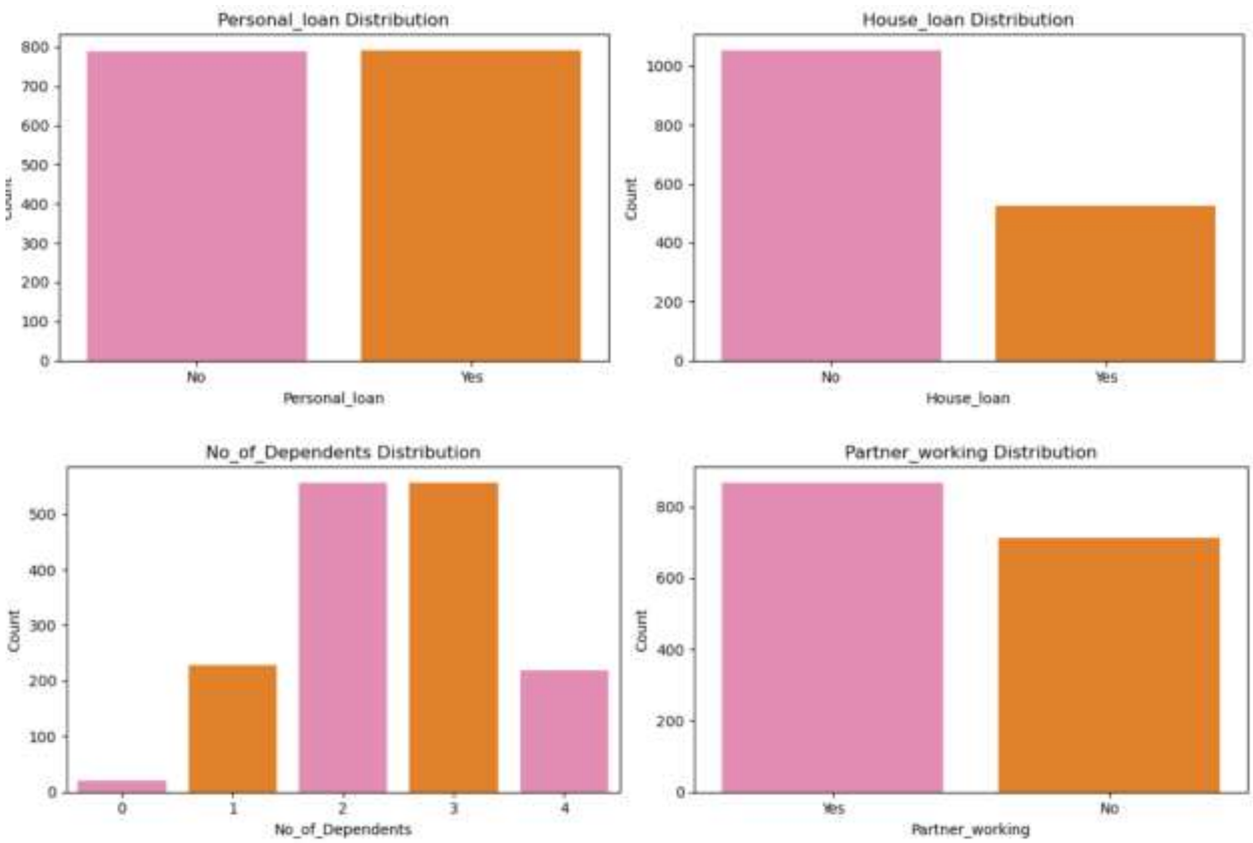

Box Plot of Total_salary
Total_salary

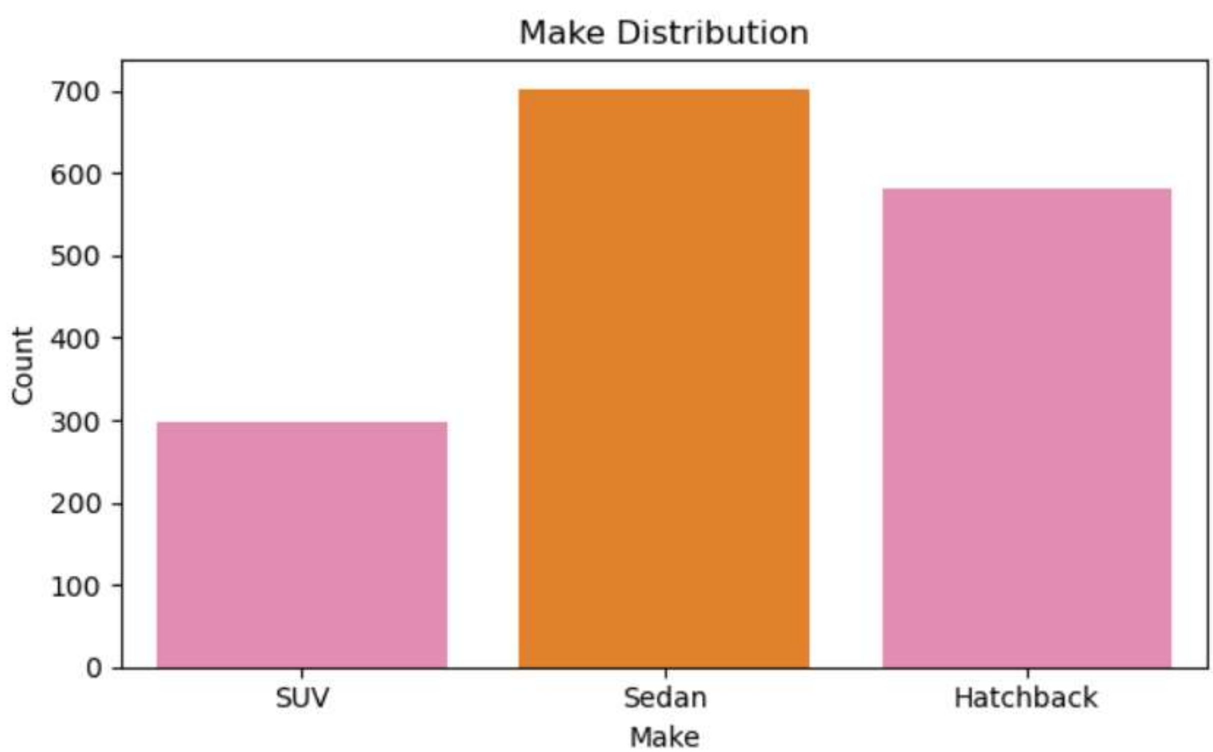


Univariate analysis of numerical variables Inferences: 1. Salary has a range between 50k to 70k. 2. Total salary has a range between 60k to 100k.

Univariate analysis of categorial




Personal_loan Distribution
800
700
600
500
400
300
200
100
0
No
Yes
Personal_loan
House_loan Distribution
1000
800
600
400
200
0
Count
No
Yes
House_loan
No_of_Dependents Distribution
500
400
300
200
100
0
Count
0
1
2
3
4
No_of_Dependents
Partner_working Distribution
800
600
400
200
0
Count
Yes
No
Partner_working


Make Distribution
700
600
500
400
300
200
100
0
Count
SUV
Sedan
Hatchback
Make

Univariate analysis of categorial variables

Inferences:

1. Sedan is most preferred, after Hatchback and SUV respectively.

2. The buyers with working partner are higher than the buyers with non-working partners or single status.

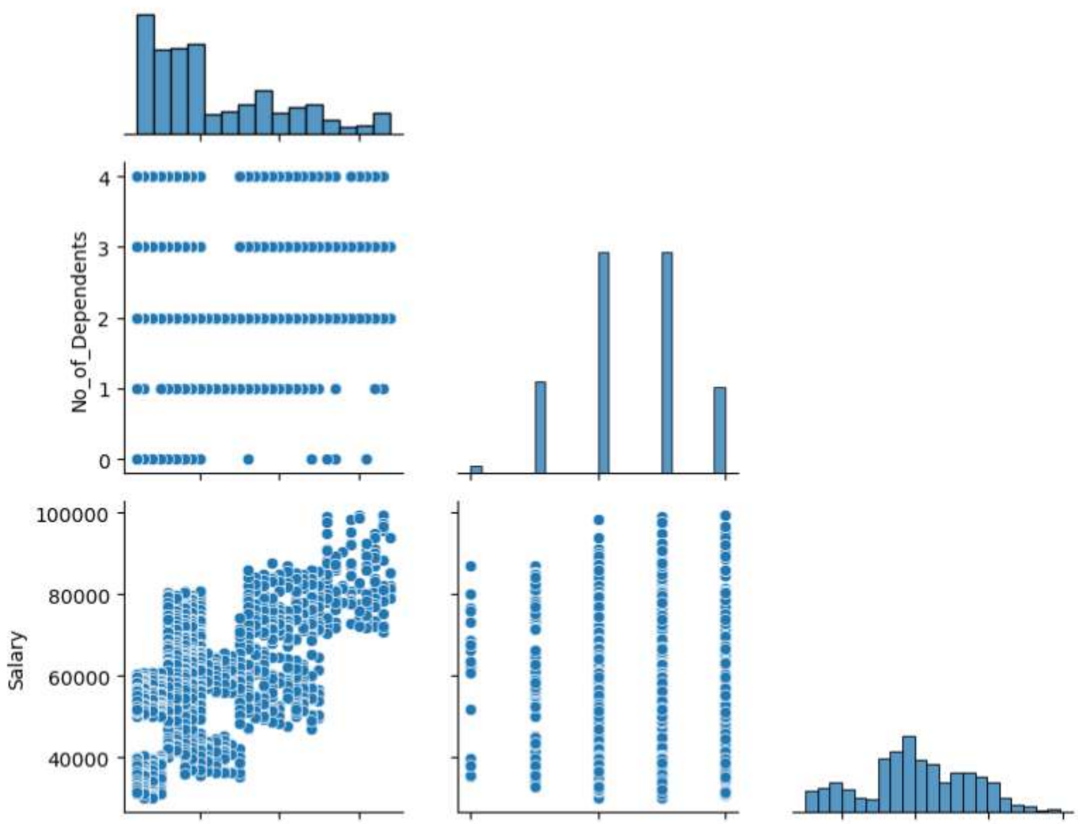
3. The married buyers are very higher than the single status.

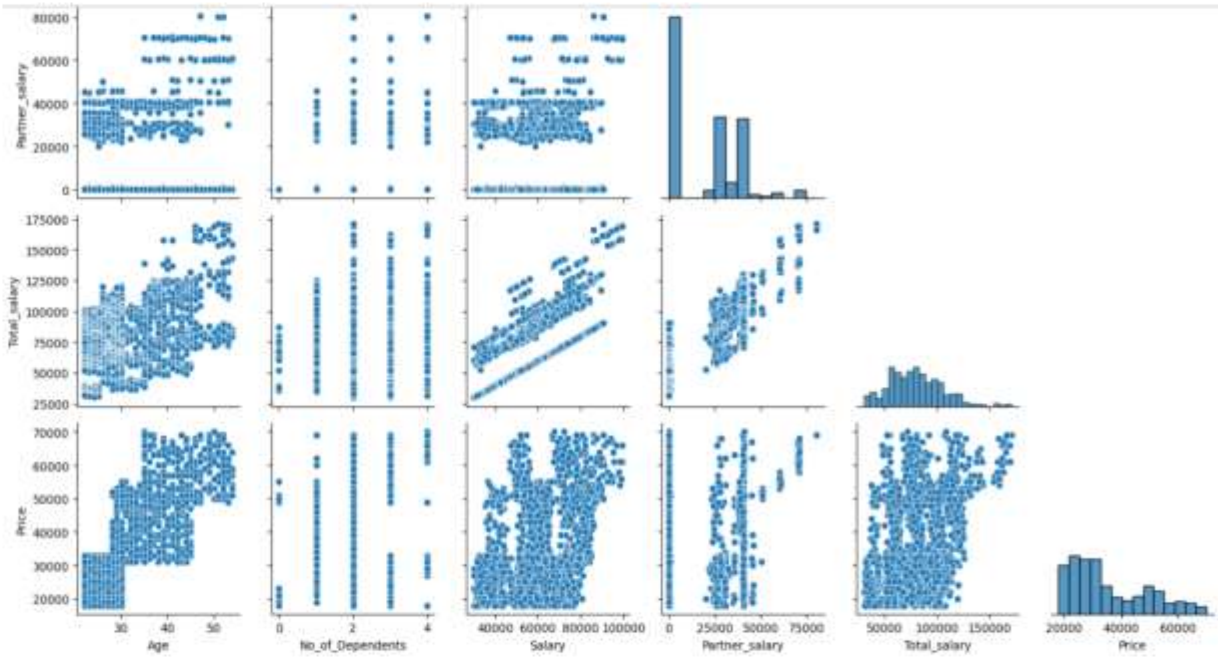
4. Major of the buyers are having postgraduate.

5. Buyers having business are little less than the number of buyers being salaried.

6. The buyers with having 2-3 dependents are higher in the dataset. Then comes the buyers with 1 & 4 dependents and the buyers having 0 dependents are very less.

Bivariate analysis of all the numerical variables

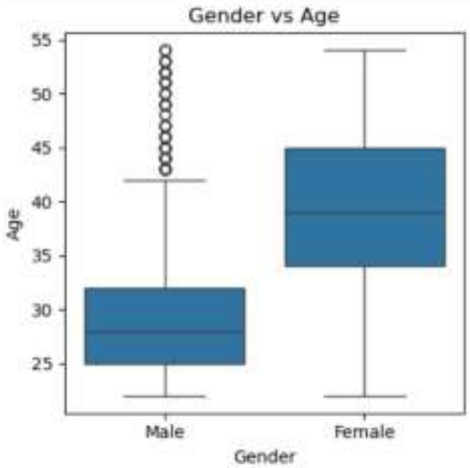


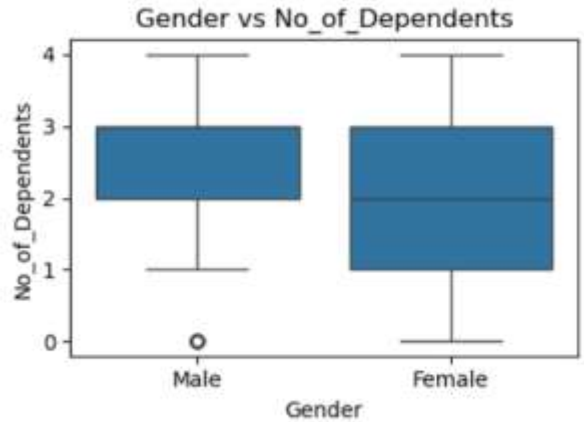


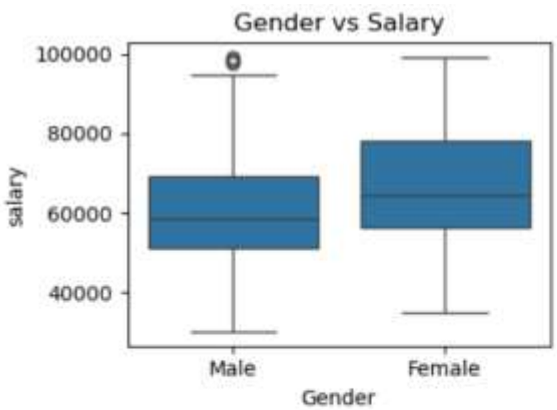
Pair plot of the dataset numerical variables

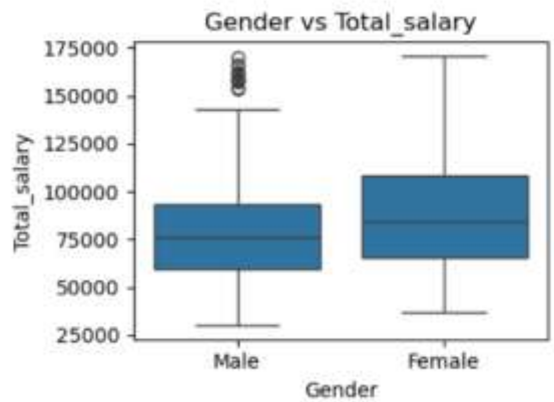
Correlation heatmap of numerical variables

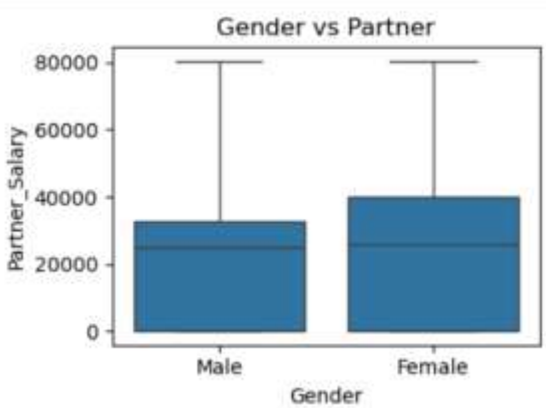
Bivariate analysis of all the categorial vs numerical variables.

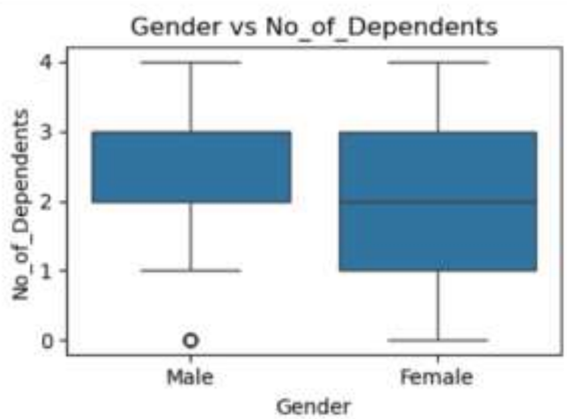


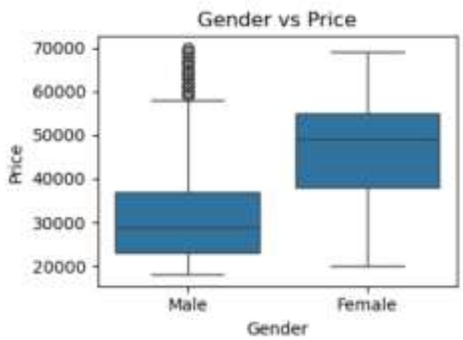


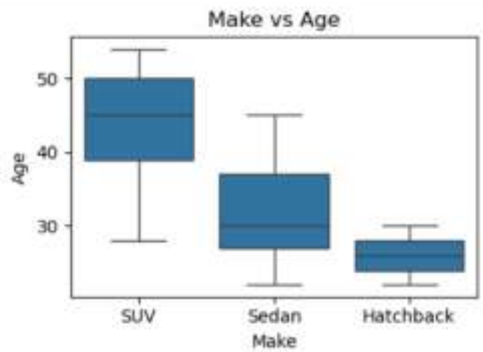


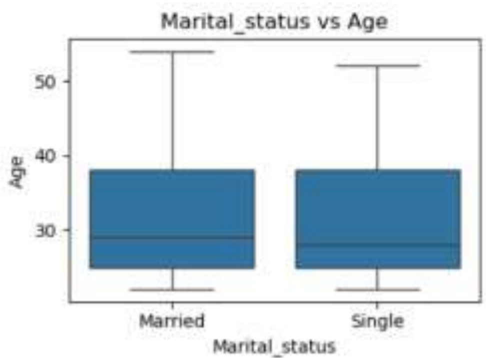


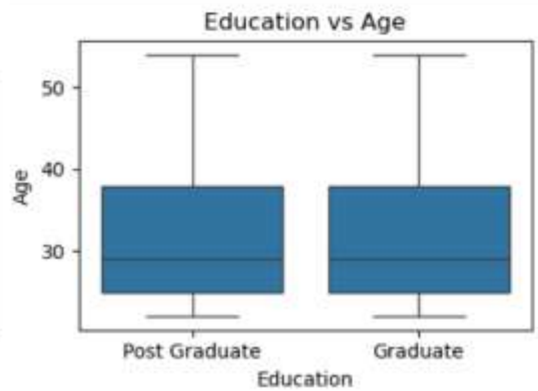


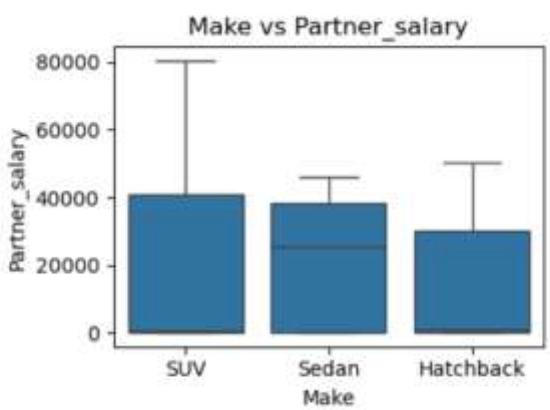


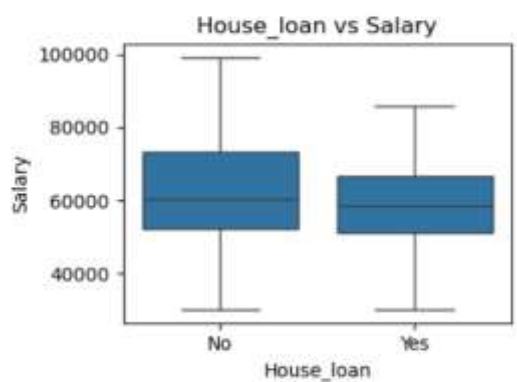

Gender vs Price
70000
60000
50000
40000
30000
20000
Price
Male
Female
Gender

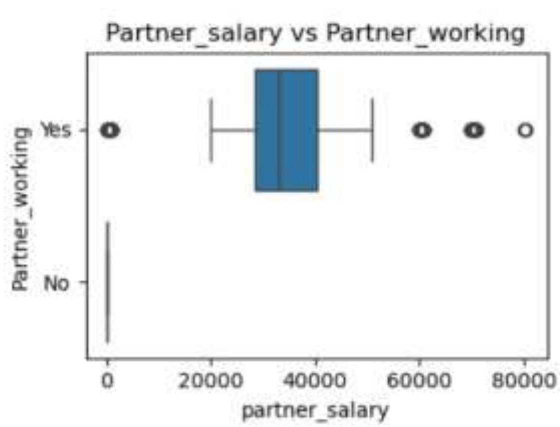

Make vs Age
50
40
30
Age
SUV
Sedan
Hatchback
Make

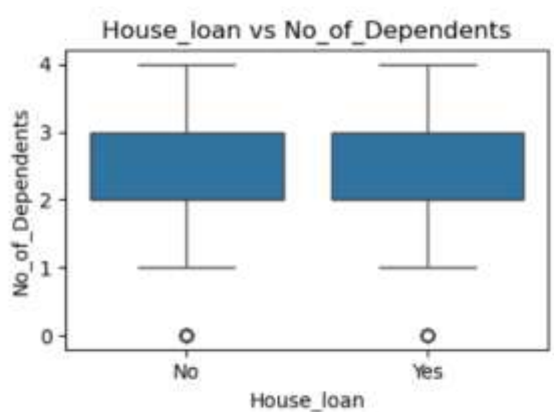

Marital_status vs Age
50
40
30
Age
Married
Single
Marital_status

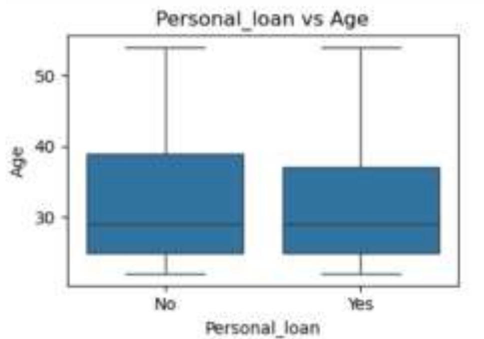

Education vs Age
50
40
30
Age
Post Graduate
Graduate
Education

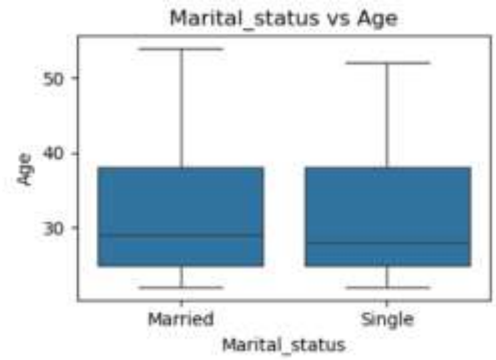

Make vs Partner_salary
80000
60000
40000
20000
0
Partner_salary
SUV
Sedan
Hatchback
Make

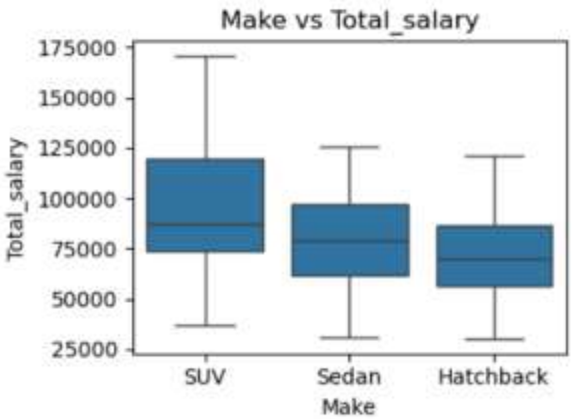

House_loan vs Salary
100000
80000
60000
40000
Salary
No
Yes
House_loan

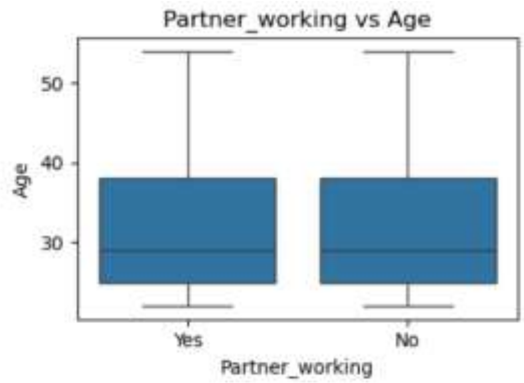

Partner_salary vs Partner_working
Yes
No
Partner_working
0
20000
40000
60000
80000
partner_salary


House_loan vs No_of_Dependents
4
3
2
1
0
No_of_Dependents
No
Yes
House_loan

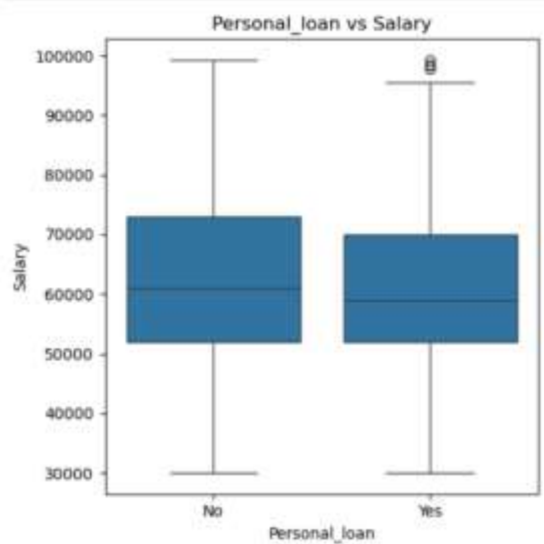

Personal_loan vs Age
Age
30
40
50
No
Yes
Personal_loan

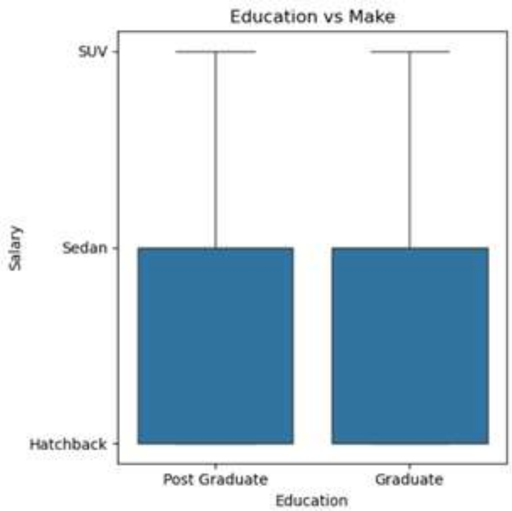

Marital_status vs Age
Age
30
40
50
Married
Single
Marital_status

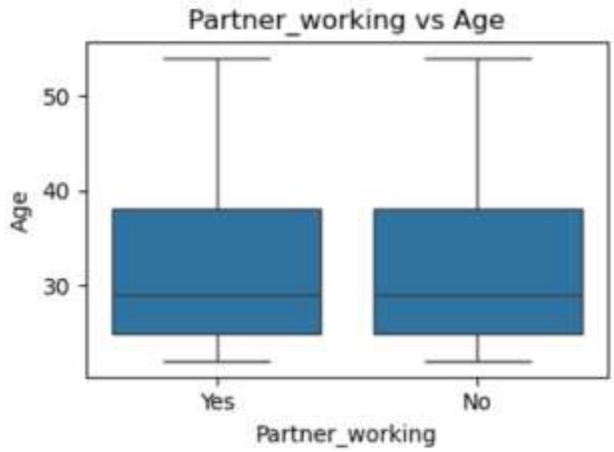

Make vs Total_salary
Total_salary
25000
50000
75000
100000
125000
150000
175000
SUV
Sedan
Hatchback
Make


Partner_working vs Age
Age
30
40
50
Yes
No
Partner_working

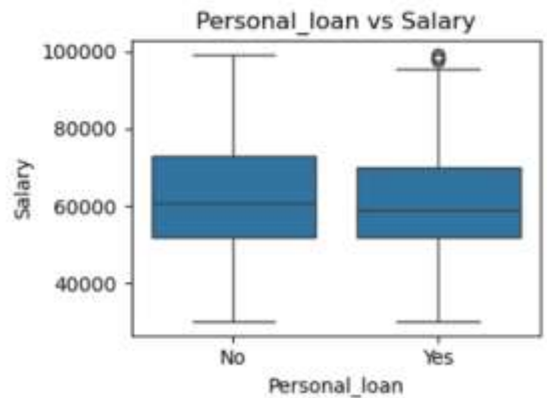

Partner_working vs Price
Price
20000
30000
40000
50000
60000
70000
Yes
No
Partner_working

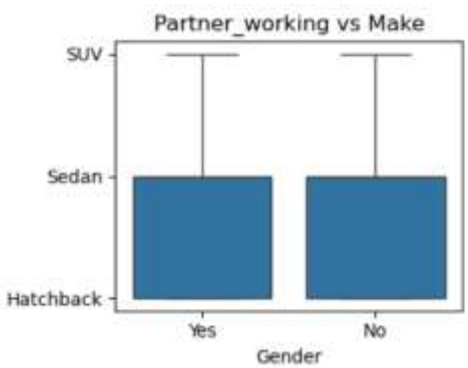

Personal_loan vs Salary
Salary
30000
40000
50000
60000
70000
80000
90000
100000
No
Yes
Personal_loan

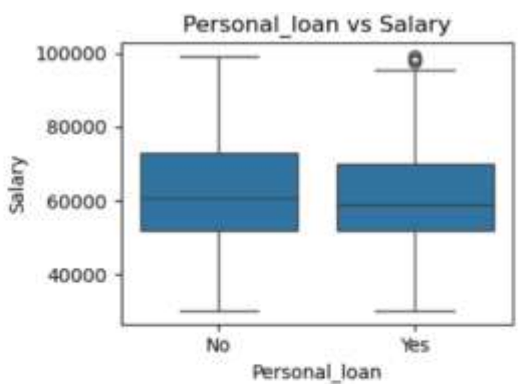

Education vs Make
SUV
Sedan
Hatchback
Salary
Post Graduate
Graduate
Education

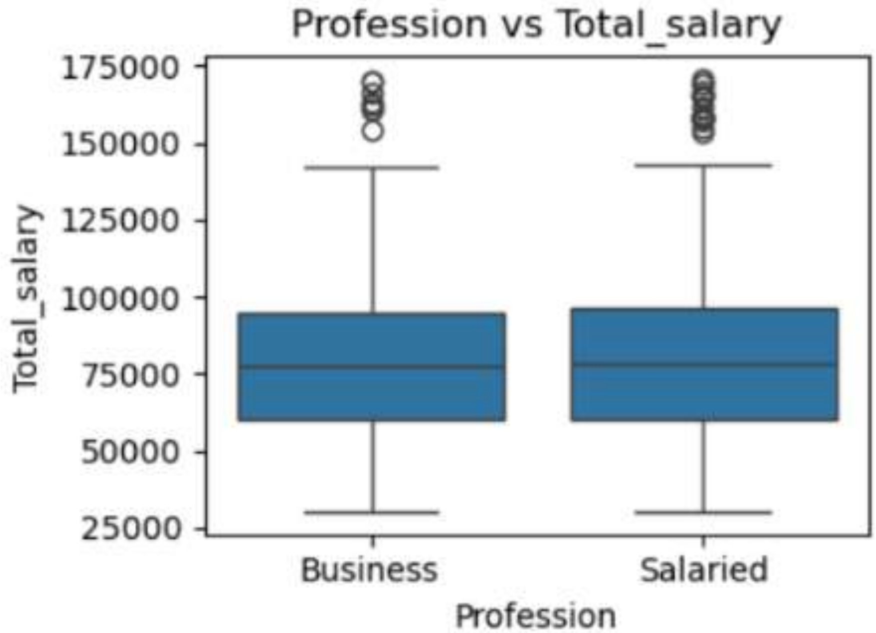

Partner_working vs Age
50
40
30
Age
Yes
No
Partner_working

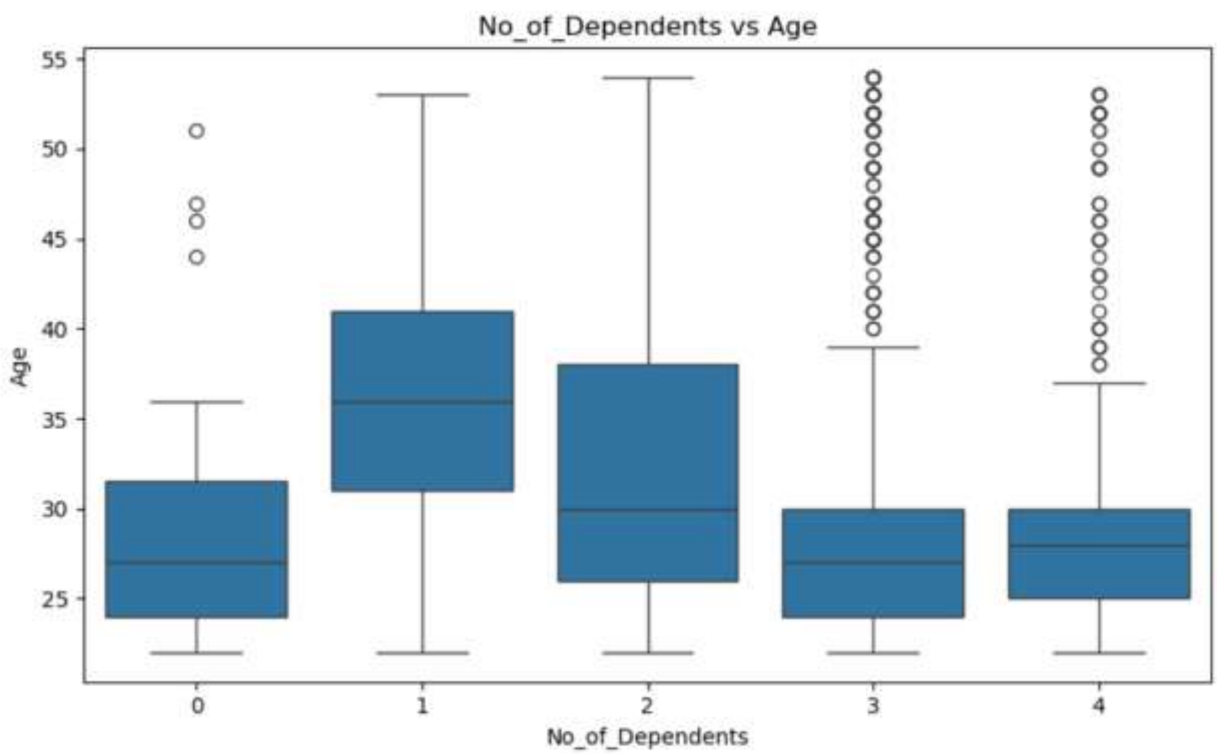

Partner_working vs Price
70000
60000
50000
40000
30000
20000
Price
Yes
No
Partner_working

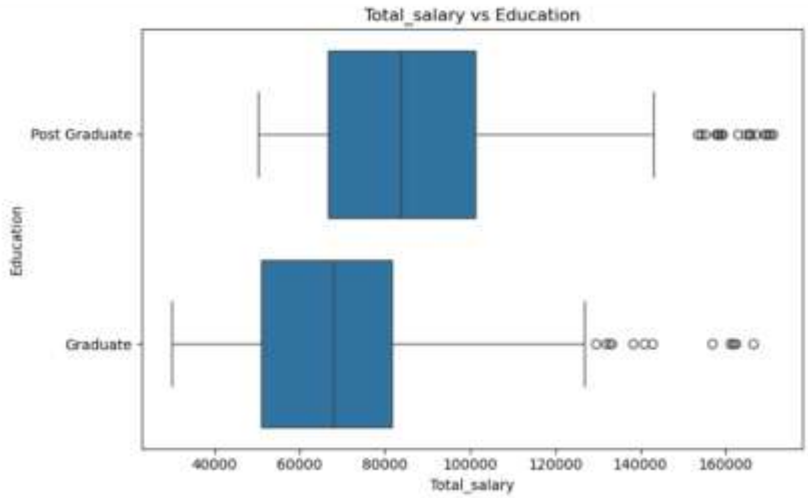

Personal_loan vs Salary
100000
80000
60000
40000
Salary
No
Yes
Personal_loan

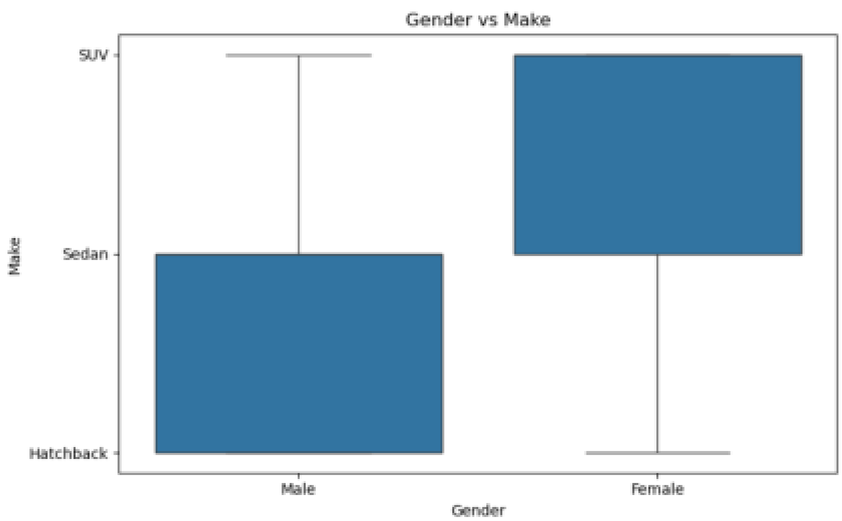

Partner_working vs Make
SUV
Sedan
Hatchback
Make
Yes
No
Gender

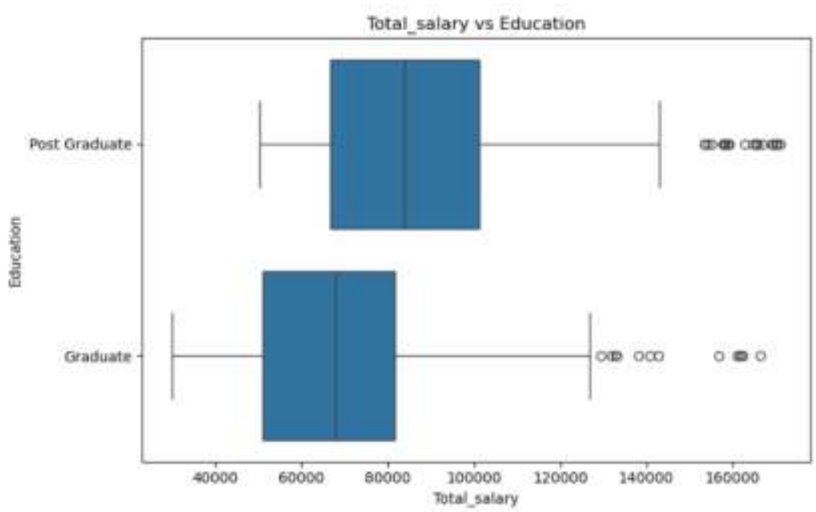

Personal_loan vs Salary
100000
80000
60000
40000
Salary
No
Yes
Personal_loan

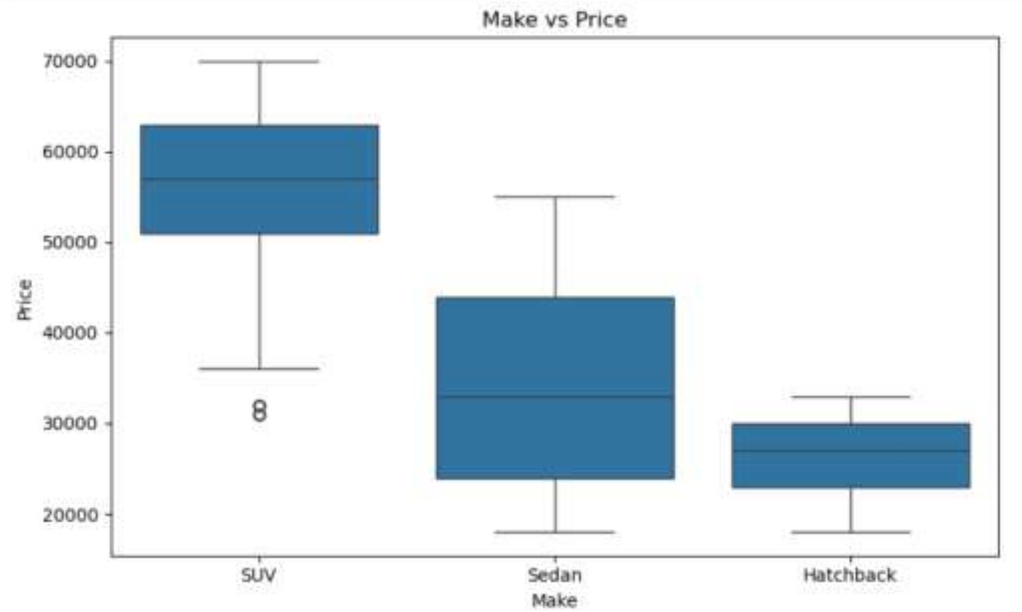

Profession vs Total_salary
175000
150000
125000
100000
75000
50000
25000
Total_salary
Business
Salaried
Profession

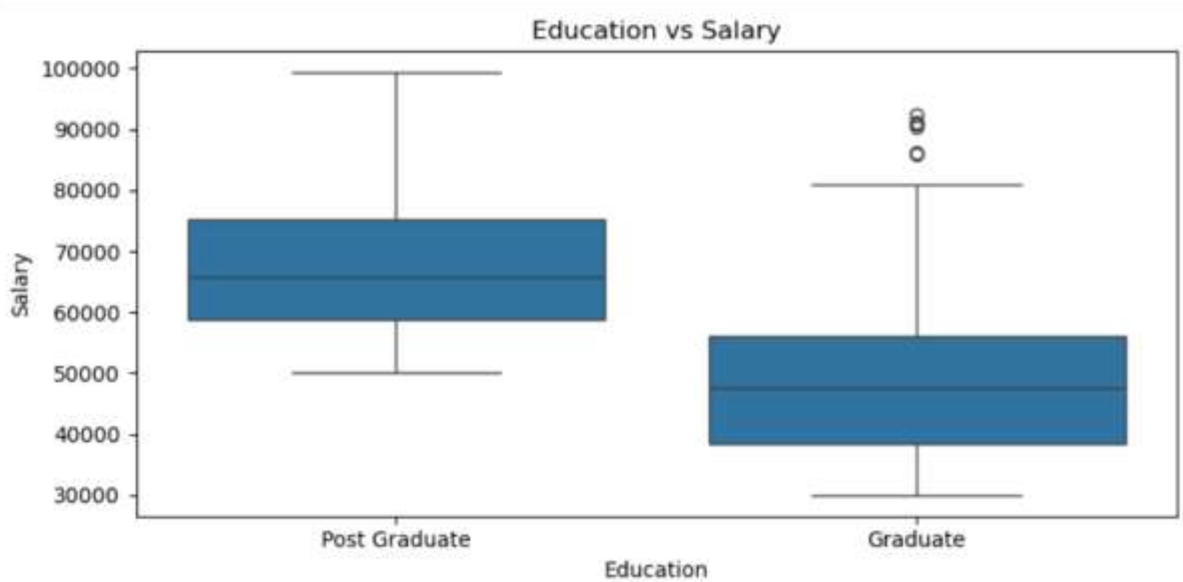

No_of_Dependents vs Age
55
50
45
40
35
30
25
Age
0
1
2
3
4
No_of_Dependents


Total_salary vs Education
Post Graduate
Graduate
Education
40000
60000
80000
100000
120000
140000
160000
Total_salary


Gender vs Make
SUV
Sedan
Hatchback
Make
Male
Female
Gender


Total_salary vs Education
Post Graduate
Graduate
Education
40000
60000
80000
100000
120000
140000
160000
Total_salary





Bivariate relationship of categorial vs numerical variables

Gender-Based Insights

- Age: Females are generally older than males.
- Dependents:
  - Females: Mostly have 1–3 dependents
  - Males: Commonly have 2–3 dependents

- Salary Range:
    - Females: ₹34,800 – ₹99,300
    - Males: ₹30,000 – ₹99,300
      Females have a higher minimum salary.
- Partner Salary: Slightly higher among females.
- Total Salary: Higher for females than males.
- Automobile Price: Females tend to purchase higher-priced cars.

Profession-Based Insights

- Salaried buyers are generally older than business buyers.
- Dependents: Both salaried and business buyers mostly have 2–3 dependents.
- Own Salary: Slightly higher for salaried buyers.
- Partner Salary: Slightly higher among salaried buyers (not significant).
- Total Salary: More outliers seen in salaried buyers indicates income diversity.
- Automobile Spending: Salaried buyers spend more on cars than business buyers.

Marital Status Insights

- Age: Married and single buyers fall in a similar age group.
- Dependents: Some singles have 1 dependent.
- Salary: Roughly similar between married and single buyers.
- Total Salary: Higher for married buyers.
- Automobile Spending: Married buyers spend more on cars.

Education-Based Insights

- Age: Graduates and postgraduates are in a similar age bracket.
- Dependents: Both mostly have 2–3 dependents.
- Salary: Higher for postgraduates.
- Partner Salary: Slightly higher for graduates.
- Total Salary: Higher for postgraduates.
- Car Spending: Surprisingly, graduates spent more on cars.

Loan & Salary-Based Insights

- Buyers with salary > ₹70,000 have not taken personal loans.

- Buyers with 2–3 dependents are more likely to have house loans.
- Buyers with a working partner have a higher total salary.

Car Type Preferences

By Age Group:

- SUV: Preferred by 38–50 years
- Sedan: Preferred by 27–37 years
- Hatchback: Preferred by 25–28 years

By Dependents:

- SUV & Hatchback: Buyers with 2–3 dependents
- Sedan: Buyers with 1–3 dependents

By Salary Range:

- SUV: ₹62,000 – ₹82,000
- Sedan: ₹52,000 – ₹68,000
- Hatchback: ₹44,000 – ₹66,000

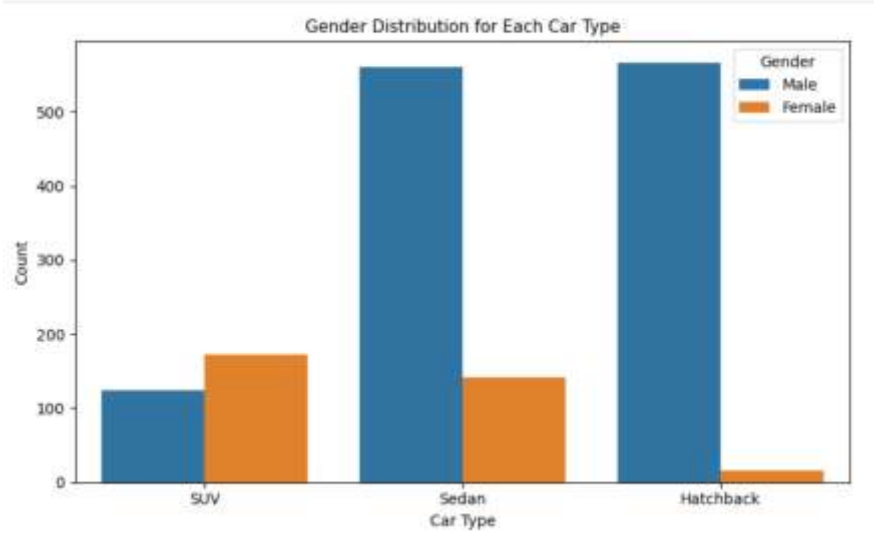
By Total Salary:

- SUV: ₹66,000 – ₹1,20,000
- Sedan: ₹64,000 – ₹88,000
- Hatchback: ₹60,000 – ₹84,000

By Car Price:

- SUVs: ₹52,000 – ₹64,000
- Sedans: ₹26,000 – ₹44,000
- Hatchbacks: ₹24,000 – ₹30,000

Do men tend to prefer SUVs more compared to women?

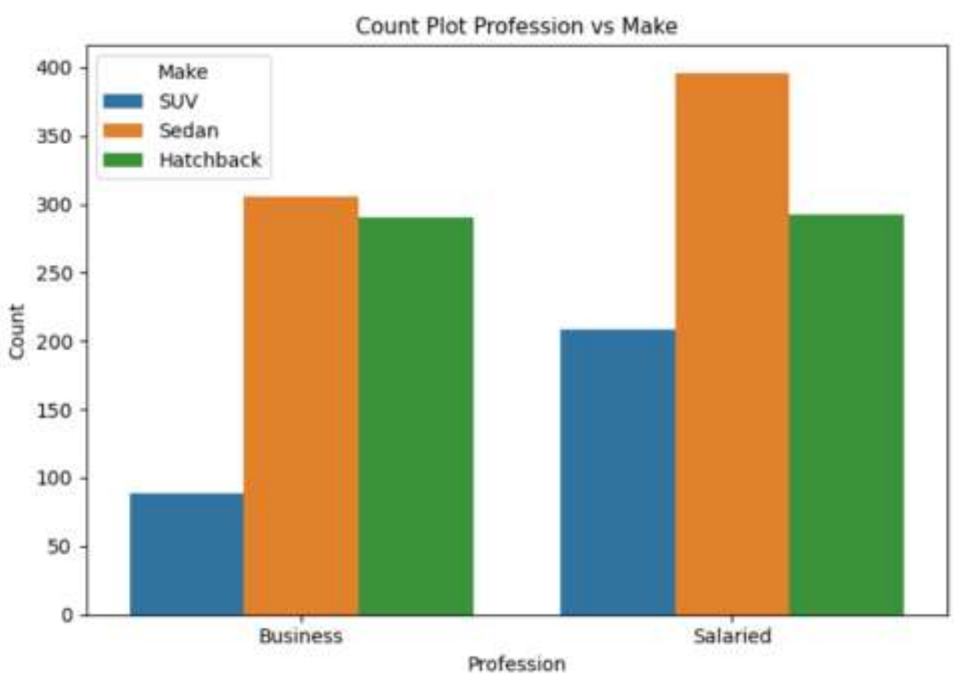


Count plot of Gender vs Make

We can properly see that the women more likely prefer SUVs compared to men.

So, the answer for the Question is 'No'.

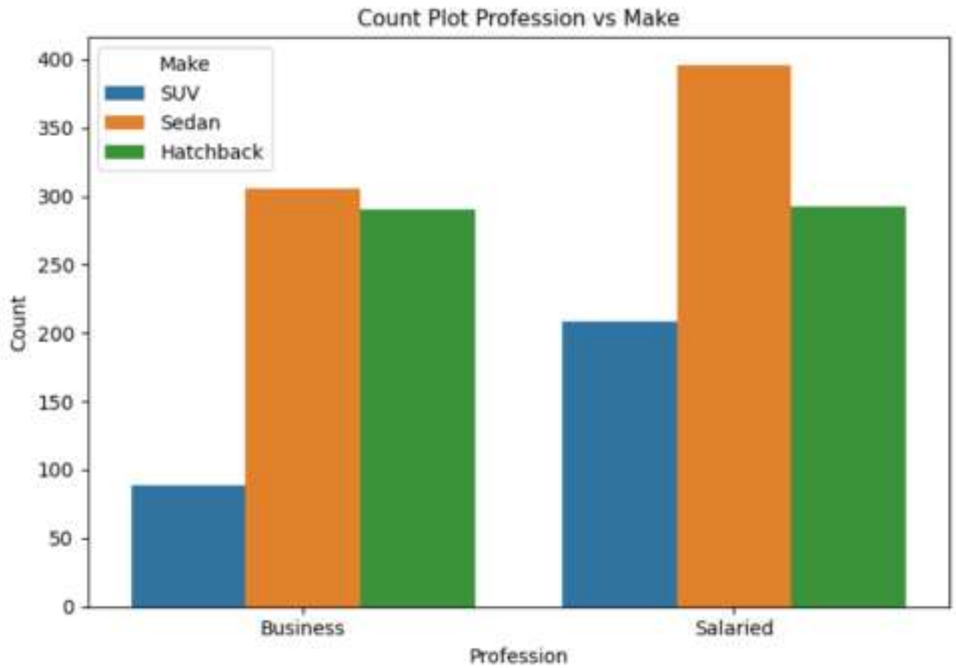
What is the likelihood of a salaried person buying a Hatchback?



Count plot of Profession vs Make

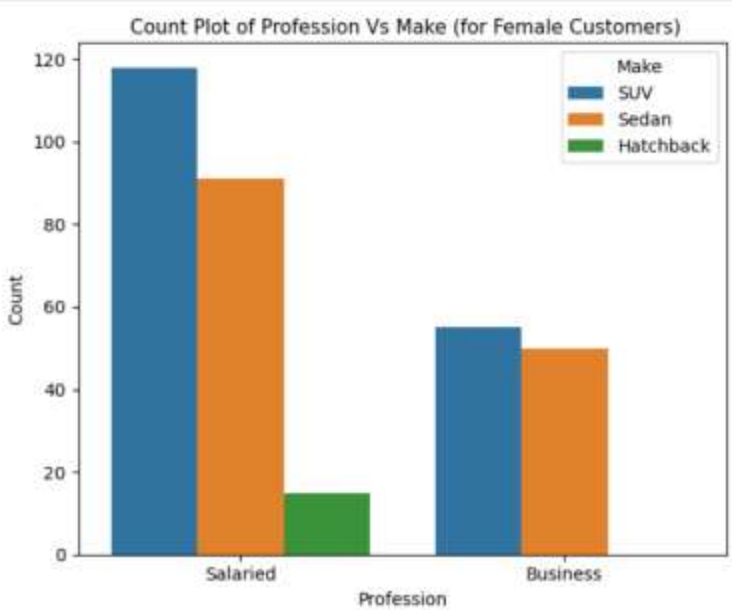
Salaried individuals are more likely to purchase Hatchbacks as their second preference after Sedans. This suggests a balance between comfort and cost-effectiveness in their purchasing decisions.

Which car is least choice of salary person & Business ?



SUV is the least choice Car by both salary person & Business Person because SUV Car cost is more as compared to Other Cars that main Reason People Like to Buy other Cars.

Which Car Is most Liked my Female in Salary Category ?



SUV is the most loved Car in Working Female Category & Hatchback is the least liked car in working women category.

Sedan is the second most loved car in Business Category & SUV also tops here also That means overall Female Likes SUV more than Males As there In income is more. So they like the most Costly car in the car Category.

Insights and Recommendations

Gender-Based Insights

- Females are generally older and tend to have higher base and total salaries.
- They are more likely to purchase higher-priced cars, especially SUVs, which are also the most preferred among salaried females.
- Recommendation: Design marketing campaigns for premium car models (like SUVs) targeted at working females, highlighting features like comfort, safety, and luxury.

Profession-Based Insights

- Salaried individuals prefer Sedans first, then Hatchbacks, while SUVs are the least preferred.
- Business owners show more interest in SUVs, likely due to higher disposable income and status-driven choices.
- Recommendation:
  - Promote economy variants of Sedans and Hatchbacks for salaried professionals.
  - Highlight prestige, spaciousness, and durability of SUVs for business buyers.

Marital Status Insights

- Married buyers have higher total salaries and spend more on car purchases.
- Single buyers have similar salary levels but lower total income.
- Recommendation: Bundle family-friendly features or EMI-based offers for married buyers, and emphasize affordability and style for singles.

Education-Based Insights

- Postgraduates earn more and have higher total salaries.
- However, graduates spend more on cars despite earning slightly less.
- Recommendation: Focus marketing on value-driven messaging for postgraduates and offer aspirational branding for graduates who are willing to spend more.

Loan & Salary Patterns

- High salary earners (above ₹70,000) avoid personal loans.

- House loans are common among those with 2–3 dependents.
- Buyers with a working partner have higher total income.
- Recommendation: Offer cash discounts or exchange offers to high-earning customers; target working couples with dual-income family packages.

Car Type Preferences

- SUV: Preferred by buyers aged 38–50 and earning ₹62k–₹82k.
- Sedan: Popular among 27–37-year-olds, earning ₹52k–₹68k.
- Hatchback: Chosen by 25–28-year-olds, with salaries between ₹44k–₹66k.
- SUVs have the highest total salary bracket (₹66k–₹1.2L) and cost more.
- Recommendation:
    - SUV ads should target older, affluent buyers.
    - Sedan and Hatchback marketing should appeal to young professionals.

Behavioral Insight

- SUVs are more popular among females than males.
- Salaried individuals prefer Sedans, but also show a second preference for Hatchbacks.
- SUV is the least chosen by both salaried and business buyers due to its high cost.
- Recommendation: For Hatchbacks and Sedans, focus on cost-effectiveness, fuel efficiency, and EMI plans.

Summary

The analysis of customer data for Austo Motor Company reveals distinct buyer profiles based on gender, profession, marital status, and education. Females and postgraduates lean toward premium purchases like SUVs, while salaried individuals and younger buyers prefer economical options like Sedans and Hatchbacks.

To optimize the current marketing strategy:

- Target working women and business owners for premium SUVs.
- Highlight affordability and efficiency for salaried youth.
- Use demographic segmentation to align car promotions with income, age, and lifestyle preferences.