BUSINESS REPORT

Data Analysis for Austo Automobile

DECEMBER 13, 2024

SAMRUDDHI POKALE
pokalesw@gmail.com

Table of Contents

List of Tables

List of Figures

Problem Statement

I. Context

Austo Motor Company is a car manufacturer specializing in both economical and luxury cars like SUV, Sedan, and Hatchback models. Some inefficiencies were found in the recent sales and marketing campaigns. We have the data with all the information like customer demographics, financial situations, model preferences, etc. Common fields like Age, Marital Status, Price, and Profession provide a rich context to understand customer needs and patterns.

II. Objective

The main objective is to clean and analyse the raw dataset to enhance the marketing campaigns of Austo Motor Company. Some specific goals are:

- Identify trends in financial behaviour with car models
- Analyse relationships between customer demographics like Age, Sex, Profession and financial preferences.
- Detecting outliers and inconsistencies in the data.
- Provide business recommendations based on observations and insights.

III. Data Description

1. **Age**: The age of the individual in years.
2. **Gender**: The gender of the individual, categorized as male or female.
3. **Profession**: The occupation or profession of the individual.
4. **Marital status**: The marital status of the individual, such as married &, single
5. **Education**: The educational qualification of the individual Graduate and Post Graduate
6. **No_of_Dependents**: The number of dependents (e.g., children, elderly parents) that the individual supports financially.
7. **Personal loan**: A binary variable indicating whether the individual has taken a personal loan "Yes" or "No"
8. **House loan**: A binary variable indicating whether the individual has taken a housing loan "Yes" or "No"
9. **Partner working**: A binary variable indicating whether the individual's partner is employed "Yes" or "No"
10. **Salary**: The individual's salary or income.
11. **Partner salary**: The salary or income of the individual's partner, if applicable.
12. **Total salary**: The total combined salary of the individual and their partner (if applicable).
13. **Price**: The price of a product or service.
14. **Make**: The type of automobile

Data Overview

I. Structure of the data

```
Structure of the data:
   Age  Gender Profession Marital_status      Education  No_of_Dependents  \
0   53    Male   Business        Married  Post Graduate                 4
1   53   Femal   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

  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

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

Column Names:
Index(['Age', 'Gender', 'Profession', 'Marital_status', 'Education',
       'No_of_Dependents', 'Personal_loan', 'House_loan', 'Partner_working',
       'Salary', 'Partner_salary', 'Total_salary', 'Price', 'Make'],
      dtype='object')

Data Shape:
Number of rows =  1581
Number of columns =  14
```

Figure 1 Structure of data

- Here we can see the first five entries in the Austo Automobile dataset.
- There are total 1581 entries in 14 columns.
- There are no duplicates.

II. Types of Data

```
Types of data:
Age                    int64
Gender                object
Profession            object
Marital_status        object
Education             object
No_of_Dependents       int64
Personal_loan         object
House_loan            object
Partner_working       object
Salary                 int64
Partner_salary       float64
Total_salary           int64
Price                  int64
Make                  object
dtype: object
```

Figure 2 Types of data

- There are three different datatypes in this dataset: Integer(int64), Float(float64) and Object

III. Missing Values

```
Missing 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
```

Figure 3 Missing values

- There are 53 missing values in “Gender” column and 106 missing values in “Partner salary” column.
- The null values in “Gender” column are replaced with “Unknown”.
- The null values in “Partner salary” column are replaced in two ways:
    1. For non-working partners, its replaced with “0”.
    2. For working partners, its replaced with the value of “Total salary”- “Salary”.

IV. Statistical Summary

Statistical Summary:

	Age	Gender	Profession	Marital_status	Education	\
count	1581.000000	1581	1581	1581	1581	
unique	NaN	5	2	2	2	
top	NaN	Male	Salaried	Married	Post Graduate	
freq	NaN	1199	896	1443	985	
mean	31.922201	NaN	NaN	NaN	NaN	
std	8.425978	NaN	NaN	NaN	NaN	
min	22.000000	NaN	NaN	NaN	NaN	
25%	25.000000	NaN	NaN	NaN	NaN	
50%	29.000000	NaN	NaN	NaN	NaN	
75%	38.000000	NaN	NaN	NaN	NaN	
max	54.000000	NaN	NaN	NaN	NaN	

	No_of_Dependents	Personal_loan	House_loan	Partner_working	\
count	1581.000000	1581	1581	1581	
unique	NaN	2	2	2	
top	NaN	Yes	No	Yes	
freq	NaN	792	1054	868	
mean	2.457938	NaN	NaN	NaN	
std	0.943483	NaN	NaN	NaN	
min	0.000000	NaN	NaN	NaN	
25%	2.000000	NaN	NaN	NaN	
50%	2.000000	NaN	NaN	NaN	
75%	3.000000	NaN	NaN	NaN	
max	4.000000	NaN	NaN	NaN	

	Salary	Partner_salary	Total_salary	Price	Make
count	1581.000000	1581.000000	1581.000000	1581.000000	1581
unique	NaN	NaN	NaN	NaN	3
top	NaN	NaN	NaN	NaN	Sedan
freq	NaN	NaN	NaN	NaN	702
mean	60392.220114	19062.428115	79625.996205	35597.722960	NaN
std	14674.825044	19477.360938	25545.857768	13633.636545	NaN
min	30000.000000	0.000000	30000.000000	18000.000000	NaN
25%	51900.000000	0.000000	60500.000000	25000.000000	NaN
50%	59500.000000	24900.000000	78000.000000	31000.000000	NaN
75%	71800.000000	38000.000000	95900.000000	47000.000000	NaN

Figure 4 Statistical Summary

- Numerical Columns:
    1. **count**: Number of non-missing values.
    2. **mean**: Average value.
    3. **std**: Standard deviation.
    4. **min**: Minimum value.
    5. **25%, 50%, 75%:** Percentile values (quartiles).
    6. **max**: Maximum value.
- Categorical Columns:
    1. **count**: Number of non-missing values.
    2. **unique**: Number of unique categories.
    3. **top**: Most frequent category.
    4. **freq**: Frequency of the most frequent category.

V. Data Irregularities

```
Unique values in Gender: ['Male' 'Femal' 'Female' 'Unknown' 'Femle']
Unique values in Profession: ['Business' 'Salaried']
Unique values in Marital_status: ['Married' 'Single']
Unique values in Education: ['Post Graduate' 'Graduate']
Unique values in Personal_loan: ['No' 'Yes']
Unique values in House_loan: ['No' 'Yes']
Unique values in Partner_working: ['Yes' 'No']
Unique values in Make: ['SUV' 'Sedan' 'Hatchback']
```

Figure 5 Unique values

- The negative values were replaced with “0” wherever present.
- In “Gender” column, spelling of “Female” is misspelled in some entries which is fixed for further analysis.

VI. Observation and Insights

- Average salary is $60392.22
- Average partner salary is $19233.77
- Average total salary is $79625.99
- Gender Percentage Distribution:
    1. Male 75.83
    2. Female 20.80
    3. Unknown 3.35
- Marital Status percentage:
    1. Married 91.27
    2. Single 8.72
- Car type percentage:
    1. Sedan 44.40
    2. Hatchback 36.81
    3. SUV 18.78

Univariate Analysis

I. Numerical variables

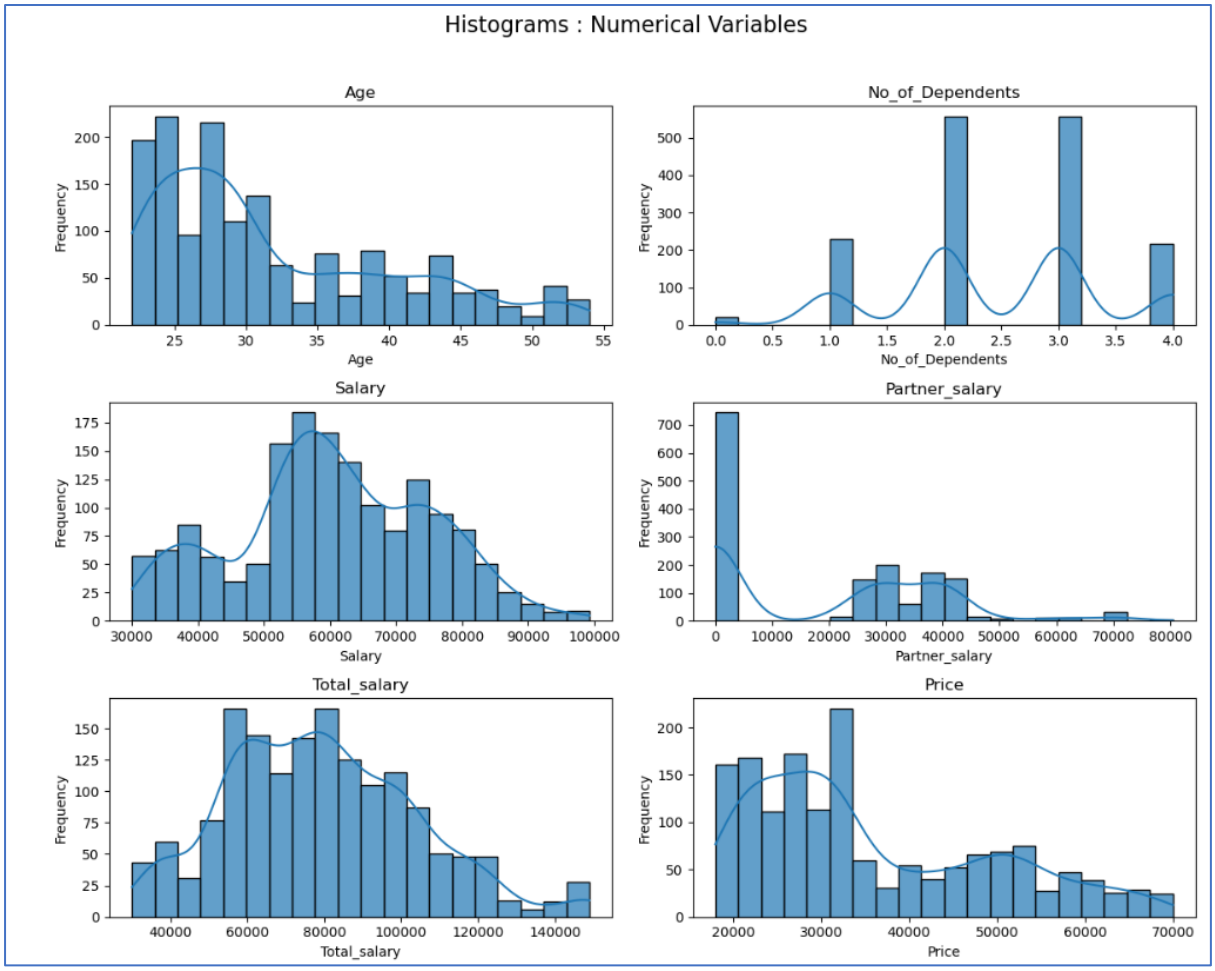

Figure 6 Histograms for numerical variables

- Age has a multimodal distribution and is right skewed with a positive skew of 0.89.
- No of dependents is bi-modal with a negative skew of 0.15.
- Salary has a multimodal distribution with a negative skew of 0.11
- Partner salary seems to be unimodal and right skewed with a positive skew of 0.45.
- Total salary is bimodal with skewness 0.6.
- Price seems to be right skewed with highest skewness of 0.74.

II. Outliers for numerical variables

- To detect the outliers in each column, boxplots are used.
- From the below figure, we can conclude that “No_of_dependents” and “Total salary” column has outliers.
- There are 20 outliers on “No_of_dependents” column and 27 outliers in “Total salary” column.

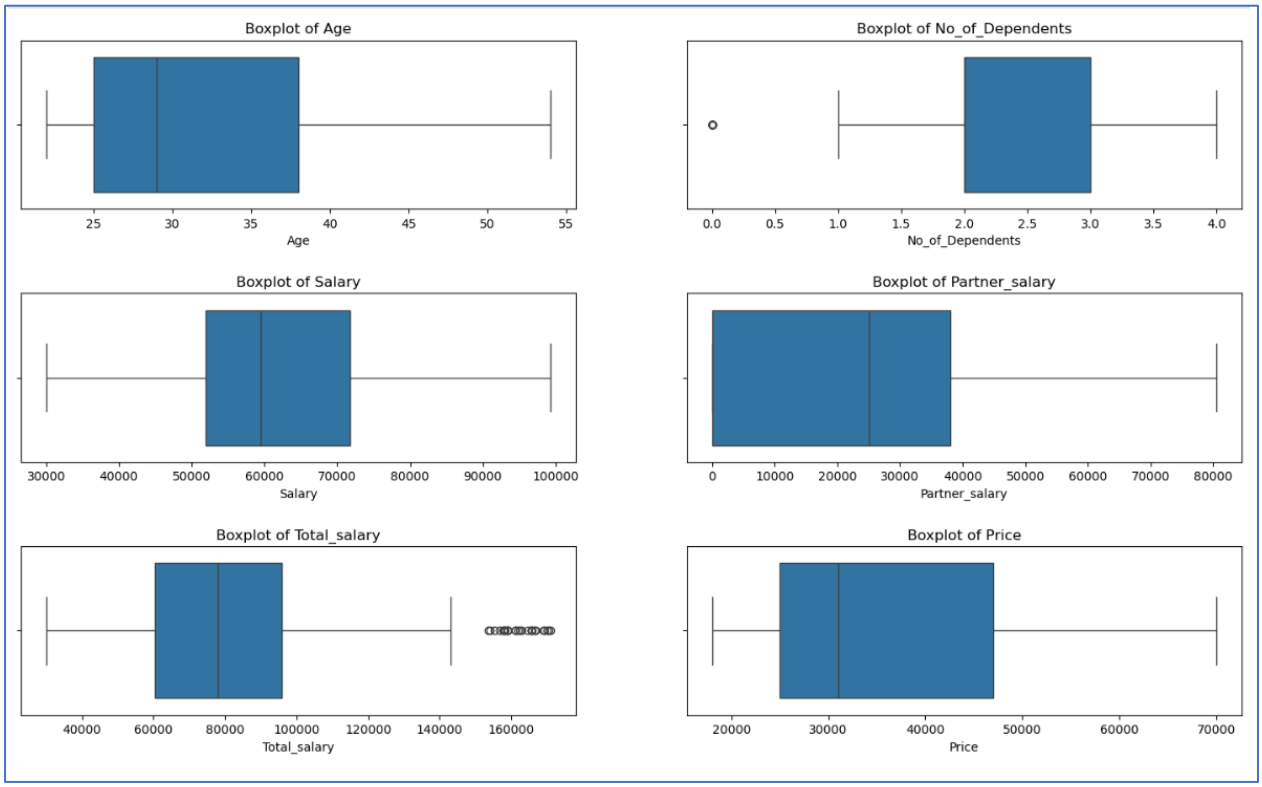

Figure 7 Boxplots for detecting outliers

- Outliers are detected using IQR method and extreme values are replaced with lower or upper bounds in “Total salary” column.
- Formula for calculating lower and upper bounds:

  Q1 = data [column_name]. quantile (0.25)
  Q3 = data [column_name]. quantile (0.75)
  IQR = Q3 - Q1
  lower_bound = Q1 - 1.5 * IQR
  upper_bound = Q3 + 1.5 * IQR
- Lower_bound for Total salary: 7400.0
- Upper_bound for Total salary: 149000.0

III. Categorical variables

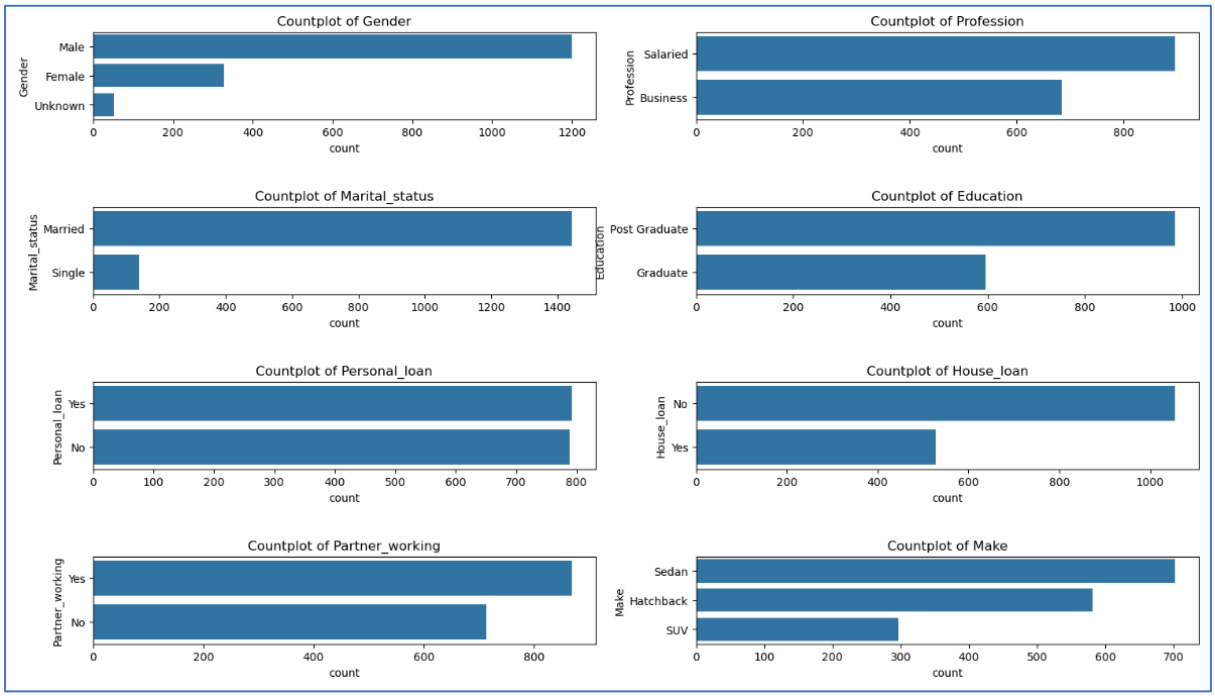

Figure 8 Count plots for categorical variables

Unique value counts for categorical variables:

- Gender:
    1. Male 1199
    2. Female 329
    3. Unknown 53
- Profession
    1. Salaried 896
    2. Business 685
- Marital status
    1. Married 1443
    2. Single 138
- Education
    1. Post Graduate 985
    2. Graduate 596
- Make
    1. Sedan 702
    2. Hatchback 582
    3. SUV 297
- House loan
    1. No 1054
    2. Yes 527
- Partner working
    1. Yes 868
    2. No 713
- Personal loan
    1. Yes 792
    2. No 789

IV. Observations and Insights

- Numerical variables insights:

Columns	Average	Minimum	Maximum
Age	32	22	54
No_of_dependents	2	0	4
Salary	$60392.22	$30000	$ 99300
Partner salary	$ 19233.77	$0	$ 80500
Total salary	$ 79398.54	$ 30000	$ 149000
Price	$ 35597.72	$ 18000	$70000

Table 1 Insights of numerical variables

- Categorical values insights:

Columns	Unique values	Most common value
Gender	3	Male
Profession	2	Salaried
Marital Status	2	Married
Education	2	Postgraduate
Personal loan	2	Yes
House loan	2	No
Partner working	2	Yes
Make	3	Sedan

Table 2 Insights of categorical variables

- Most of the vehicle buyers are male and most preferred vehicle is a Sedan.
- People with 2 or more dependents prefer buying vehicle over 1 or no dependents.
- Single buyers are very less as compared to married buyers.
- Buyers are mostly married and are salaried employees.
- People with a working partner prefer buying a vehicle more as compared to a non-working partner.
- Buyers with no house loan prefer buying vehicles as compared to people with a house loan.

Bivariate Analysis

I. Relationship between all numerical variables

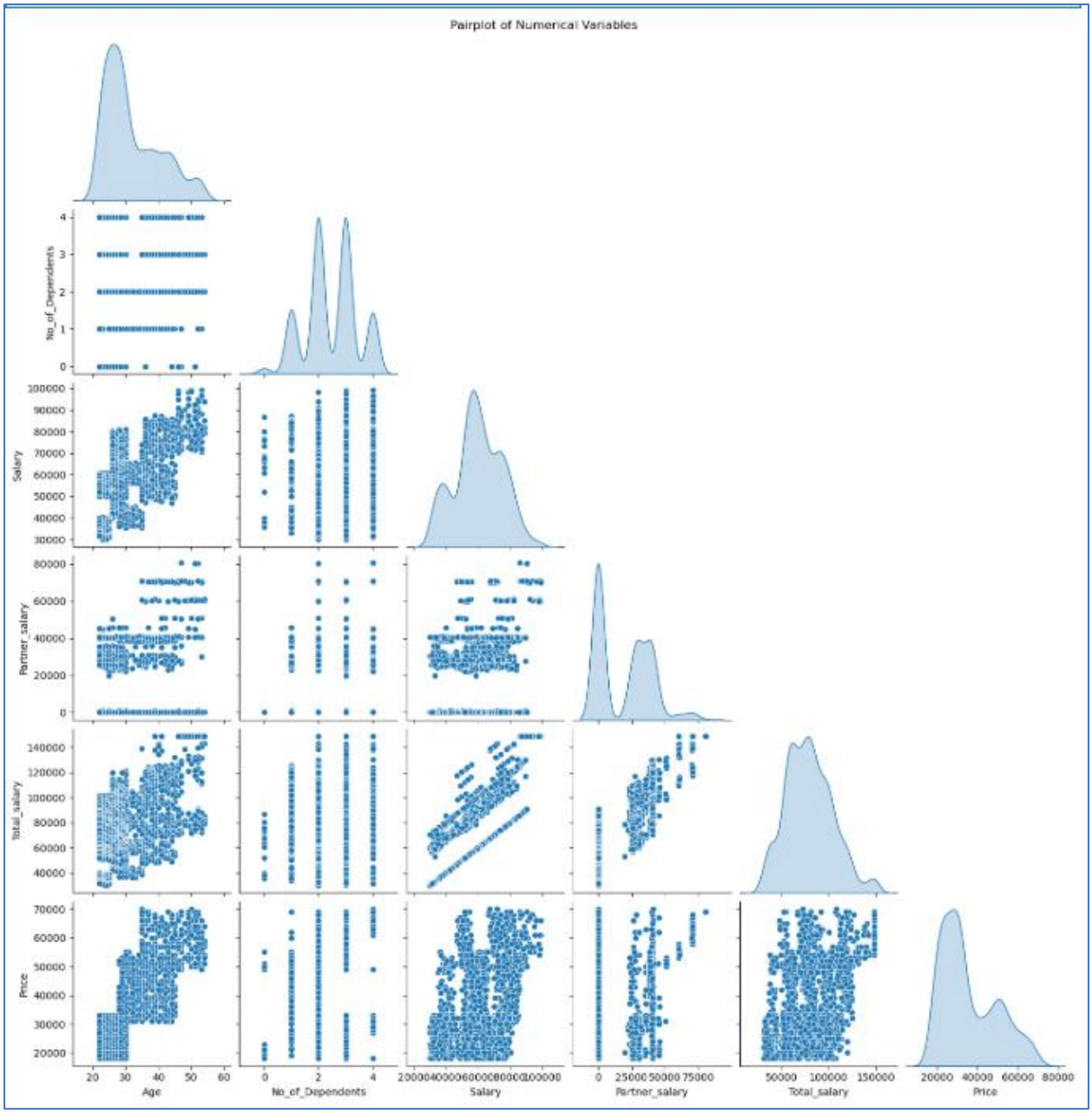

Figure 9 Pair plots for numerical variables

II. Correlation between numerical values

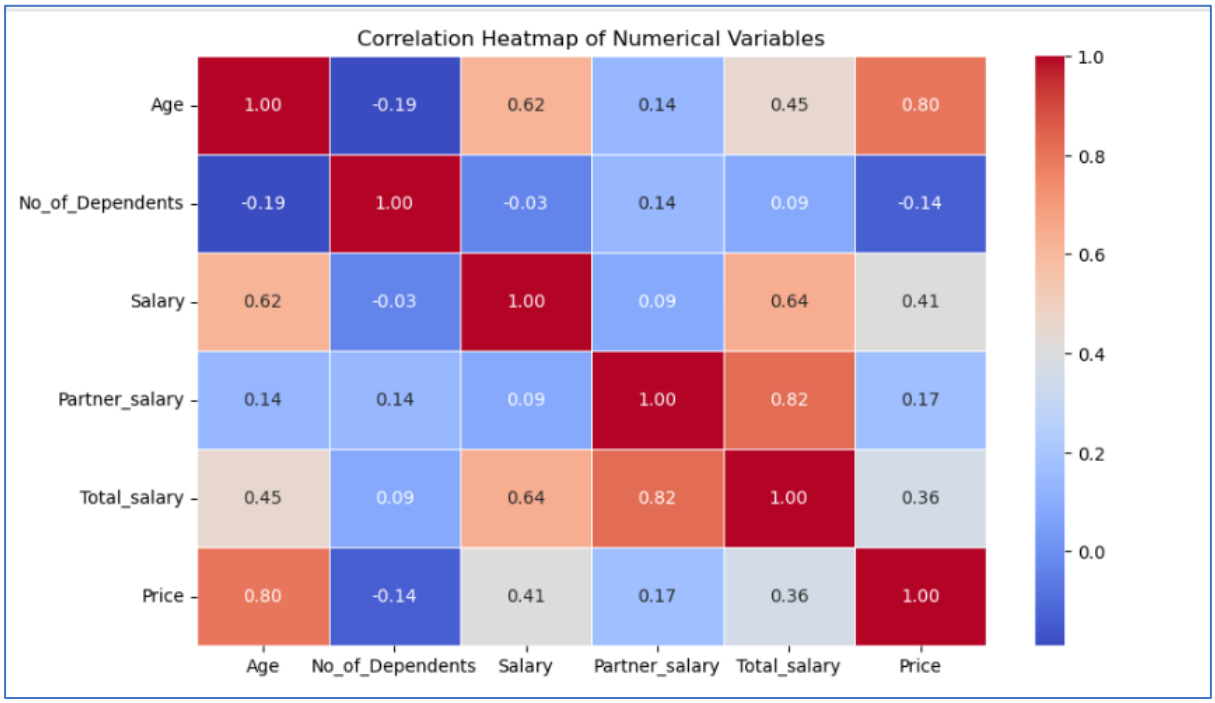

Figure 10 Correlation heatmap of Numerical variables

- **Strong Positive Correlations:**
    1. Age and Price (**0.8**): Elder buyers tend to buy expensive vehicles.
    2. Partner salary and Total salary (**0.82**): A strong relationship indicates that partner salaries contribute significantly to total salary.
- **Moderate Positive Correlations**:
    1. Age and Salary (**0.62**): Elder buyers earn more salaries.
    2. Salary and Total salary (**0.62**): Salary contributes to the total salary.
    3. Age and Total salary (**0.45**): Age contributes to higher total income, though the relationship is weaker compared to salary alone.
    4. Salary and Price (**0.41**): buyers earning more can afford more expensive items.
- **Weak Positive Correlations**:
    1. Partner salary and Price (**0.17**): Partner salary doesn't play much role in the price of vehicle.
    2. No_of_Dependents and Partner salary (**0.14**): A weak correlation between dependents and partner income.
- **Weak Negative Correlations**:
    1. No_of_Dependents and Age (**-0.19**): Younger individuals tend to have more dependents.
    2. No_of_Dependents and Price (**-0.14**): Households with more dependents tend to buy lower-priced vehicles.

III. Relationship between categorical vs numerical variables

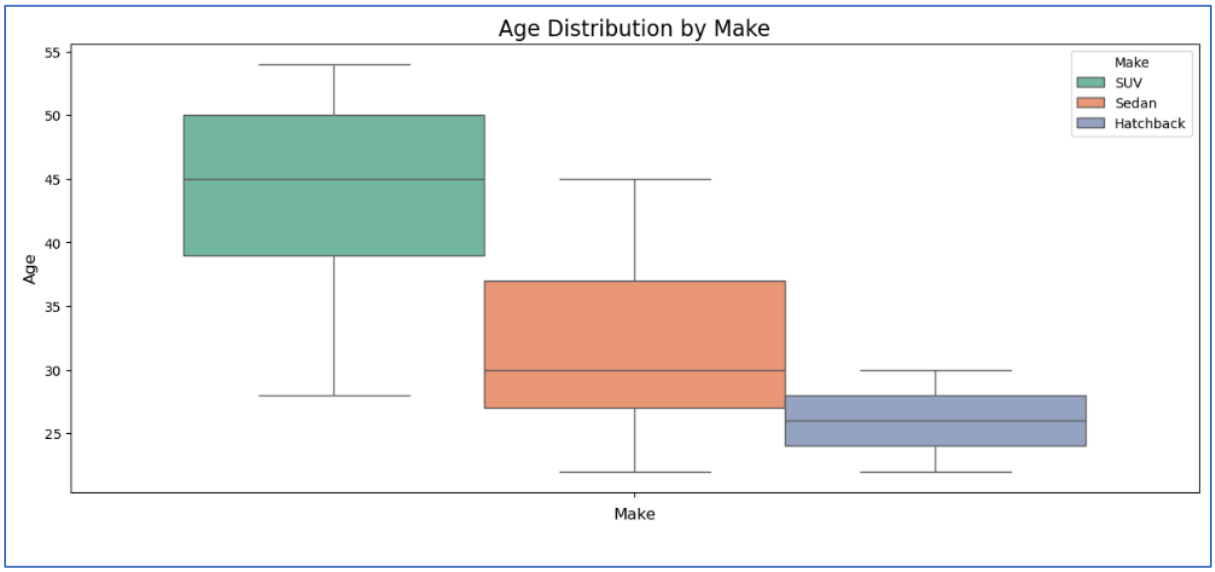

Figure 11 Age vs Make

- Elder buyers prefer SUV while younger buyers prefer Sedan and Hatchback.

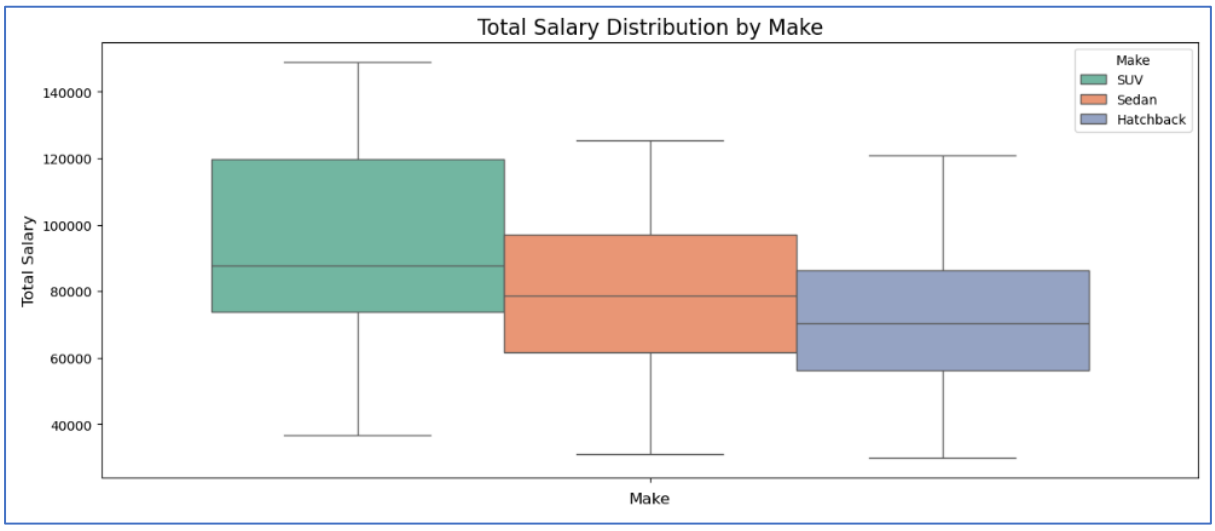

Figure 12 Total Salary vs Make

- Buyers with higher salary prefer SUV.
- As the salary decreases buyers prefer Sedan and then Hatchback.

Figure 13 Price vs Gender

- Female buyers prefer expensive vehicles as compared to male buyers.

Figure 14 Price vs Make

- SUVs are most expensive cars followed by Sedan and the Hatchback.

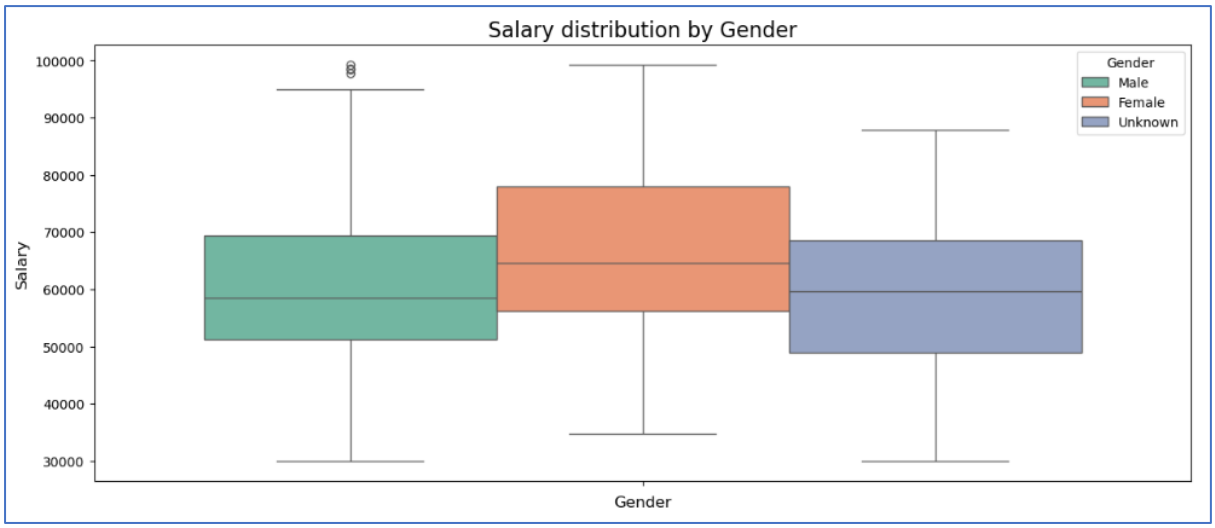

Figure 15 Salary vs Gender

- Female buyers tend to have more salary as compared to male buyers.

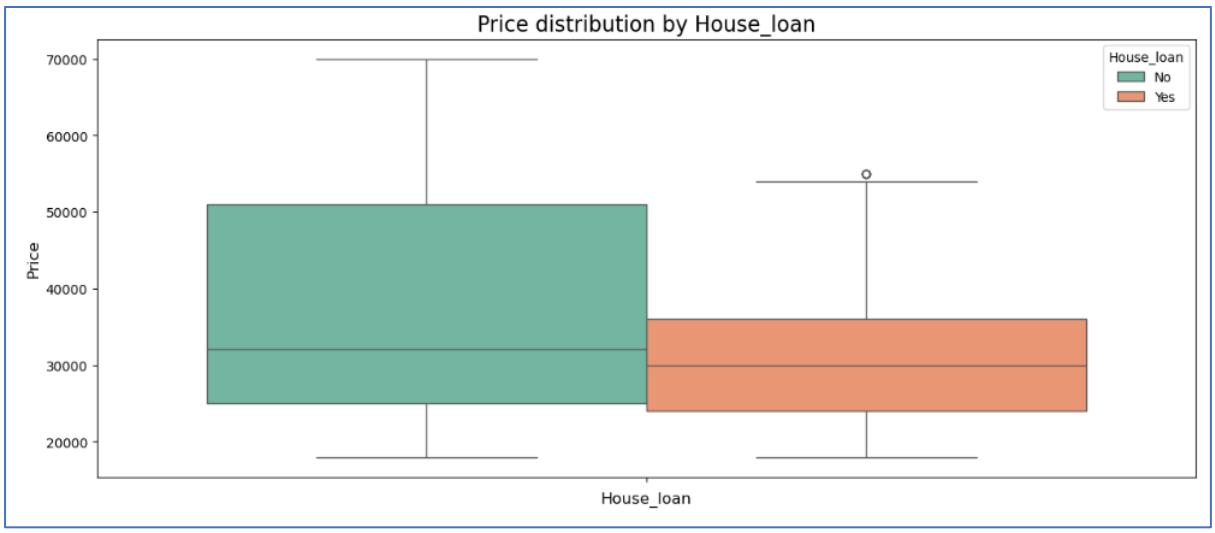

Figure 16 Price vs House Loan

- Buyers without house loan buy more expensive vehicles.

IV. Miscellaneous plots

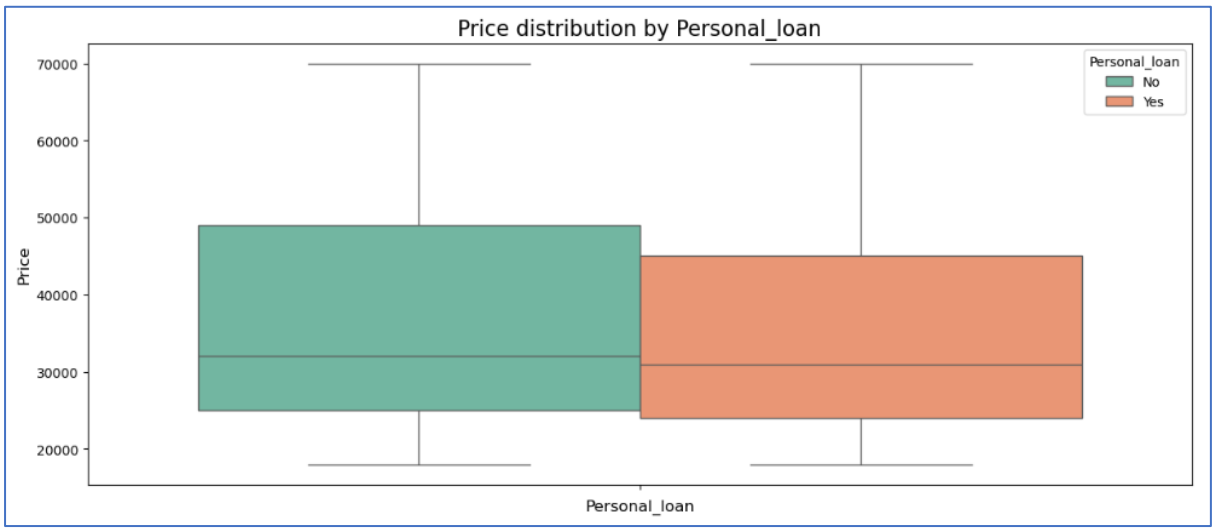

Figure 17 Price vs Personal loan

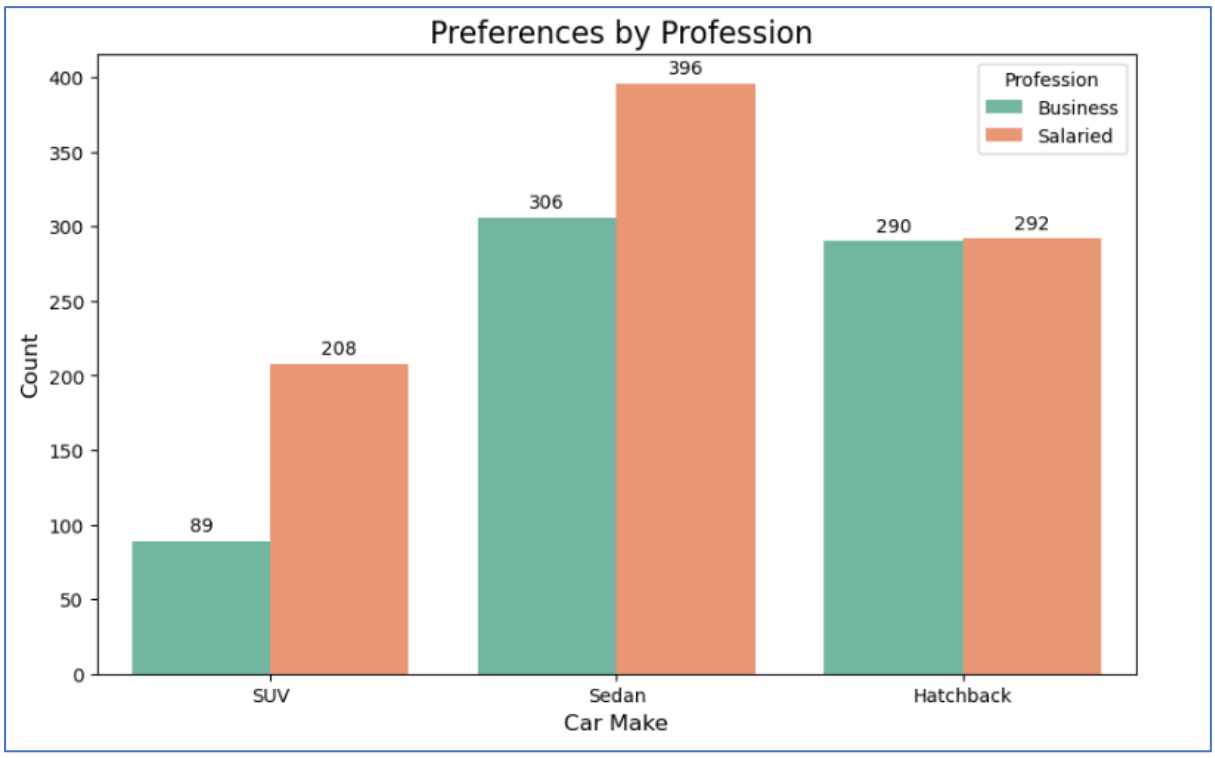

Figure 18 Preferences by Profession

V. Observations and Insights

- Partner salary and Total salary have the largest connection (0.82), indicating that both the buyer and the partner have an impact on buying options.
- Both salary and price are significantly impacted by age, suggesting a possible maturity-related preference for pricey cars.
- Poor correlations, like those between No_of_Dependents and other variables, imply that dependents affect consumers' purchasing decisions but with a lesser impact.
- Sedans and hatchbacks draw younger purchasers, but SUVs get older buyers.
- Due to their high cost, SUV are preferred by purchasers with larger salaries overall.
- Those without home loans Favor high-end automobiles.
- It appears that personal loans are utilized for more costly purchases because buyers with them typically acquire SUV.
- Sedan is the top option for both business and salaried professionals.

Key Questions

I. Do men tend to prefer SUVs more compared to women?

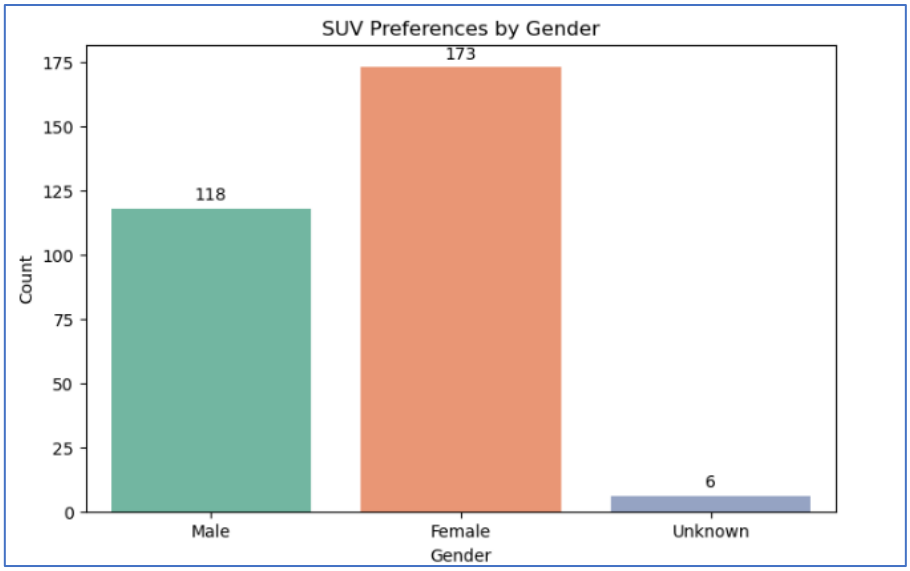

Figure 19 SUV preferences by Gender

- Women (173) tend to prefer SUVs as compared to men (118).

II. What is the likelihood of a salaried person buying a Sedan?

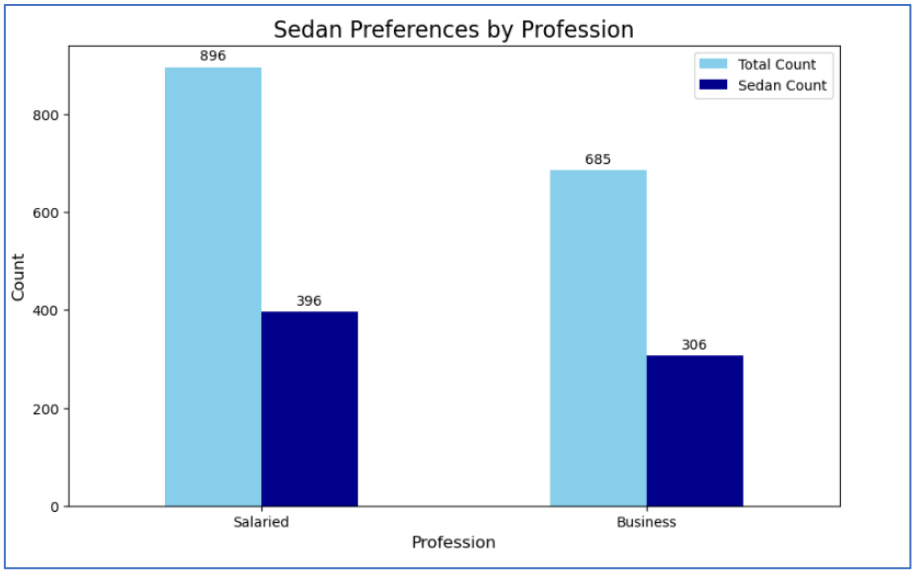

Figure 20 Sedan preferences by Profession

- Salaried professionals are more likely to buy Sedans as compared to businessmen which is approximately 56% of total salaried buyers.

III. What evidence or data supports Sheldon Cooper's claim that a salaried male is an easier target for a SUV sale over a Sedan sale?

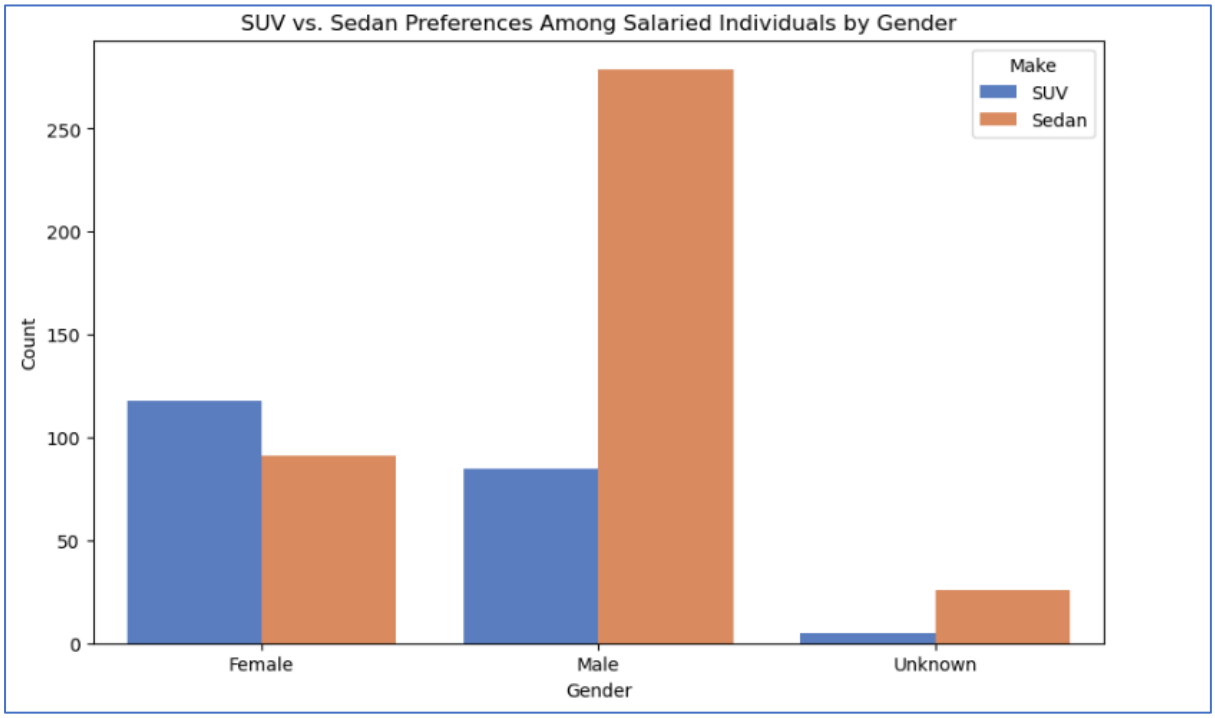

Figure 21 SUV vs Sedan preferences in salaried individuals

- There is no data that supports Sheldon Cooper's claim that a salaried male is an easier target for a SUV sale over a Sedan sale.
- On the contrary, salaried males prefer Sedan over SUV.

IV. How does the amount spent on purchasing automobiles vary by gender?

Refer: Figure 13 Price vs Gender

- Female buyers spend more money on vehicles as compared to male counterparts.
- Female buyers tend to spend between $40000 to $60000 on vehicles.

V. How much money was spent on purchasing automobiles by individuals who took a personal loan?

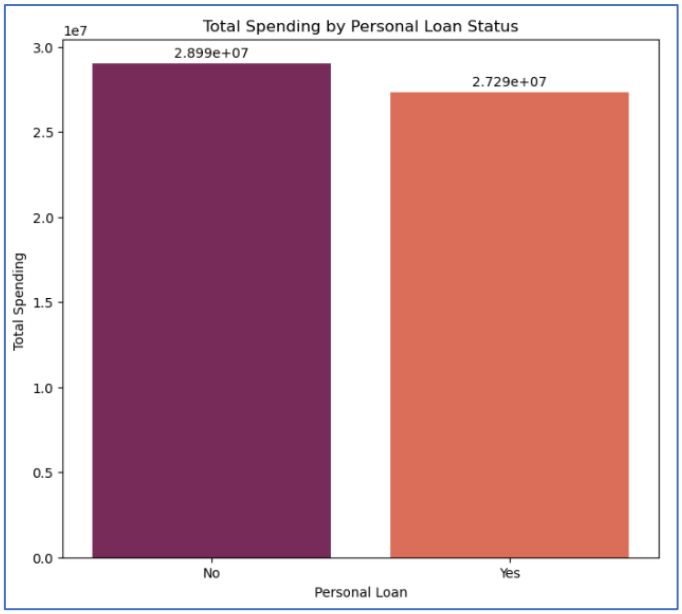

Figure 22 Total spending vs Personal Loan

- Total amount spent on purchasing automobiles by individuals who took personal loan is almost $3 crores.
- Buyer with personal loan spends around $25000 to $50000. (Refer: Figure 17 Price vs Personal loan)
- On average, individual with personal loan spends about $34,500 on car.

VI. How does having a working partner influence the purchase of higher-priced cars?

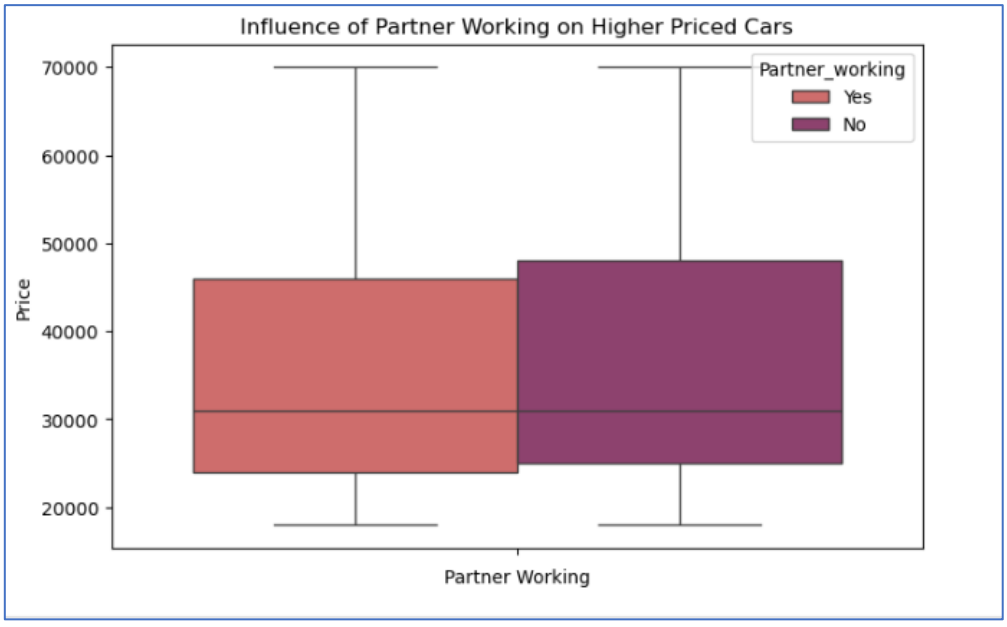

Figure 23 Prices vs Working Partner

- There is no such influence of working partner on the buying of expensive vehicles.

Actionable Insights and Business Recommendations

Sr. No.	Actionable Insights	Business Recommendations
1	Female buyers prefer expensive vehicles as compared to male buyers.	• Create gender specific campaigns. • For male customers, concentrate on performance, power and price. • For female customers, emphasize on usefulness, fuel-efficiency. • Offer customization like interiors, colours and other add-ons.
2	Salaried professionals prefer affordable cars like Sedans	• Offer discounts and EMIs on semi-luxury and luxury cars for salaried individuals. • Advertise premium models with luxury interiors for business professionals. • Offer cashback deals.
3	Buyers with bigger families prefer affordable cars.	• Create campaigns for dual-income households. • Emphasize on features like child-security and larger space in cars.
4	Young buyers prefer economical while elder buyers prefer expensive vehicles.	• Introduce different pricing strategies for different age groups. • To attract young buyers, introduce online ads and social media campaigns. • Launch websites with online test drive and 360° view of cars.
5	Customers tend to buy expensive vehicles with the help of personal loans.	• Collaborate with banks to offer loans at low interest rates and other schemes.
6	Customer loyalty programs	• Create programs to retain old customers like discounts or free first servicing or free interiors, etc
7	Regional Expansion	• Explore rural markets and introduce offers according to regional preferences and budgets. • Collaborate with regional dealers to increase company's presence.
8	After sales services	• Introduce free pickups and drops for servicing. • Update customers about servicing through apps or emails

Table 3 Actionable insights and business recommendations