BUSINESS REPORT

AMC Customer Data Analysis

23.06.24

Nabankur Ray
PGP-DSBA

Contents

Problem.....3

1. Do men tend to prefer SUVs more compared to women? 13
2. What is the likelihood of a salaried person buying a Sedan? 14
3. 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? 16
4. How does the amount spent on purchasing automobiles vary by gender?.....17
5. How much money was spent on purchasing automobiles by individuals who took a personal loan? 18
6. How does having a working partner influence the purchase of higher-priced cars?.....19

List of figures

Figure 1: Count Plot for categorical variables.....8

Figure 2: Distributions for numerical variables.....9

Figure 3: Boxplot of numerical variables for detecting outliers.....10

Figure 4: Pair plots for numerical variables 11

Figure 5: Correlation of Numerical Variables 12

Figure 6: Gender wise distribution of car type 13

Figure 7: Profession wise distribution of car type 15

Figure 8: Car type distribution among different professions 15

Figure 9: Bar Graph between Gender and Price of the cars 17

Figure 10: Bar Graph between Price of the car and Individuals having Personal loan or not.....18

Figure 11: Bar Graph between Working Partner and Price of the car.....21

List of Tables

Table 1: First 5 rows of the given dataset *5*

Table 2: Last 5 rows of the given dataset *5*

Table 3: Checking the structure and type of data *5*

Table 4: Statistical summary of the Data *6*

Table 5: No of Missing Values in each column *7*

Table 6: Proportion table among type of cars and gender *13*

Table 7: Count table among profession and type of cars *15*

Table 8: Proportion table among profession and type of cars *15*

Table 9: Proportion table among Profession, Gender and Type of car *16*

Table 10: Mean Price based on Gender *17*

Table 11: Mean Price based on Individuals having personal loan *19*

Table 12: Mean Price based on Individuals having working partner *21*

Problem:

Austro Motor Company (AMC), a leading car manufacturer specializing in SUV, Sedan, and Hatchback models, is experiencing concerns about the efficiency of its current marketing campaign. The Board of Directors seeks to leverage data to gain insights into customer demand and enhance the customer experience. This analysis will lead to more targeted marketing efforts, potentially increasing sales and brand loyalty.

Data Availability: A customer dataset(austo_automobile.csv), contains -

- **Demographics information:** Age, Gender, Profession (Business, Salaried), Marital Status, Education (Graduate, Post Graduate), Number of Dependents.
- **Financial Information:** Personal Loan, Housing Loan, Partner Employment Status, Salary, Total Salary
- **Car Purchase Information:** Price, Make (SUV, Sedan, Hatchback)

Executive Summary: This analysis will provide actionable insights regarding customer demand for different car models across various customer segments. These insights will be used to refine Austro Motor Company's marketing campaigns, leading to improved customer targeting and potentially increased sales.

Deliverables:

- Identification of key customer segments with distinct car model preferences.
- Recommendations for targeted marketing campaigns based on customer segment analysis.
- Potential for increased sales and improved customer experience through better targeting.

Understanding the Data

Dataset sample

First 5 rows

	Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price	Make
0	53	Male	Business	Married	Post Graduate	4	No	No	Yes	99300	70700.00	170000	61000	SUV
1	53	Femal	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500	70300.00	165800	61000	SUV
2	53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300	60700.00	158000	57000	SUV
3	53	Female	Salaried	Married	Graduate	2	Yes	No	Yes	72500	70300.00	142800	61000	SUV
4	53	Male	Salaried	Married	Post Graduate	3	No	No	Yes	79700	60200.00	139900	57000	SUV

Table 1: First 5 rows of the given dataset

Last 5 rows

	Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price	Make
1576	22	Male	Salaried	Single	Graduate	2	No	Yes	No	33300	0.00	33300	27000	Hatchback
1577	22	Male	Business	Married	Graduate	4	No	No	No	32000	NaN	32000	31000	Hatchback
1578	22	Male	Business	Single	Graduate	2	No	Yes	No	32900	0.00	32900	30000	Hatchback
1579	22	Male	Business	Married	Graduate	3	Yes	Yes	No	32200	NaN	32200	24000	Hatchback
1580	22	Male	Salaried	Married	Graduate	4	No	No	No	31600	0.00	31600	31000	Hatchback

Table 2: Last 5 rows of the given dataset

Structure and Types of Data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1581 entries, 0 to 1580
Data columns (total 14 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 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: 173.0+ KB
```

Table 3: Checking the structure and type of data

Observations:

- The dataset contains 1581 rows and 14 Columns
- There are 6 numerical columns in the data and 8 object type columns.
- It can be observed that Gender and Partner Salary columns have less entries than other columns (less than 1581 rows) which indicates that there is missing values in the data.

Statistical summary of the Data

	count	mean	std	min	25%	50%	75%	max
Age	1581.00	31.92	8.43	22.00	25.00	29.00	38.00	54.00
No_of_Dependents	1581.00	2.46	0.94	0.00	2.00	2.00	3.00	4.00
Salary	1581.00	60392.22	14674.83	30000.00	51900.00	59500.00	71800.00	99300.00
Partner_salary	1475.00	20225.56	19573.15	0.00	0.00	25600.00	38300.00	80500.00
Total_salary	1581.00	79626.00	25545.86	30000.00	60500.00	78000.00	95900.00	171000.00
Price	1581.00	35597.72	13633.64	18000.00	25000.00	31000.00	47000.00	70000.00

Table 4: Statistical summary of the Data

Observations:

- The age of individuals ranges from 22 to 54 years. 50% of individuals are 29 years or below.
- Maximum No. of Dependent is up to 4 person.
- Salaries range from 30,000 to 99,300, with a mean salary of 60,392.
- Partner salaries range from 0 to 80,500, with a mean of 20,225.
- Total salaries range from 30,000 to 171,000, with a mean of 79,626.
- Prices of automobiles range from 18,000 to 70,000.

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
```

Table 5: No of Missing Values in each column

Observations:

- There are total of 159 missing values.
- Gender column has 53 missing values.
- Partner salary column has 106 missing values.

INFERENCE

- All missing values have been addressed. The Gender and Partner salary columns are now filled as below -
    - For Gender Column: Missing value is filled with 'Unknown'
    - For Partner_salary, missing values is filled with 0 if Partner_working is NO and if Partner_working is YES then missing value is filled with the formulae "Total_Salary-salary=Partner_Salary"
- Gender has incorrect entries: 'Femal' and 'Femle'., which is corrected to 'Female'.

Univariate Analysis

To understand the distribution of individual variables. We will use summary statistics and visualization tools like histograms or box plots for numerical variables, and bar charts for categorical variables.

Observations on Categorical Variables

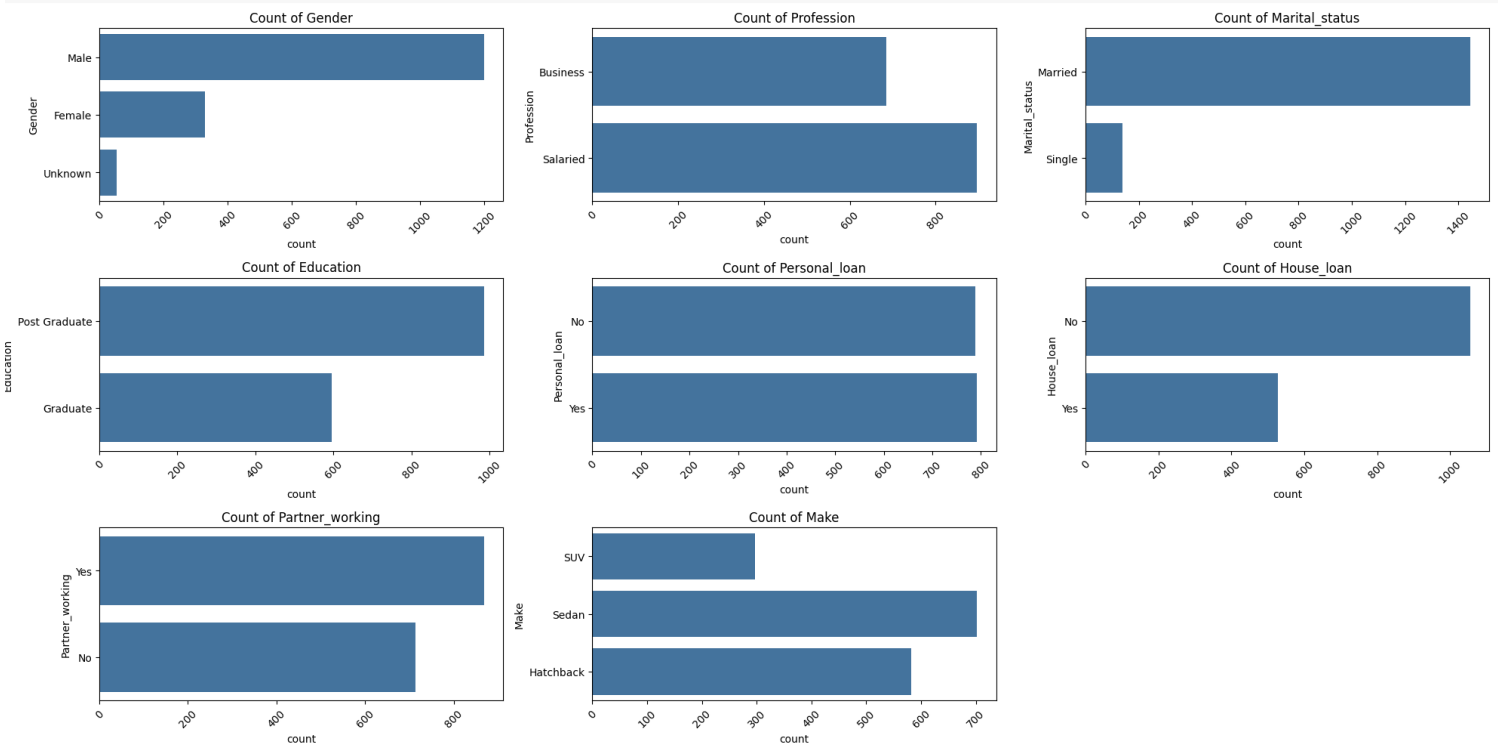

Figure 1: Count Plot for categorical variables

- The count of Male customer is more than the female customer.
- Salaried Person Prefer to buy cars more than Businessperson.
- Married individuals are more potential buyers of cars than those who are single.
- The count of individuals with a post-graduate education is higher than those with a graduate education, which implies that a post-graduate background individual has high chances of buying a car.
- Person without house loan have brought more cars than Person with house loan.

- The individuals With or without Personal loan have very little significant as the no. of person buying car who has personal loan and who does not have personal loan are same.
- Customers purchase sedan cars more than the other two types of cars.

Observation on Numerical Variables

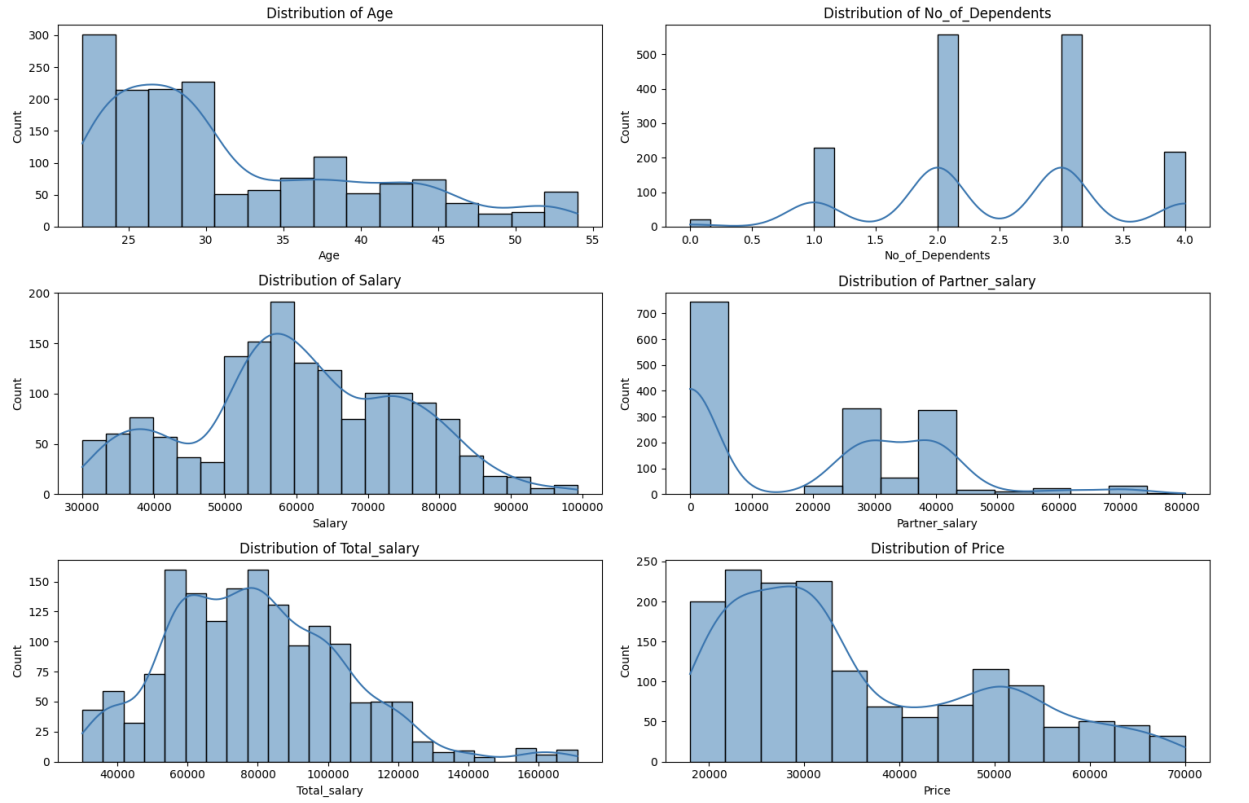

Figure 2: Distributions for numerical variables

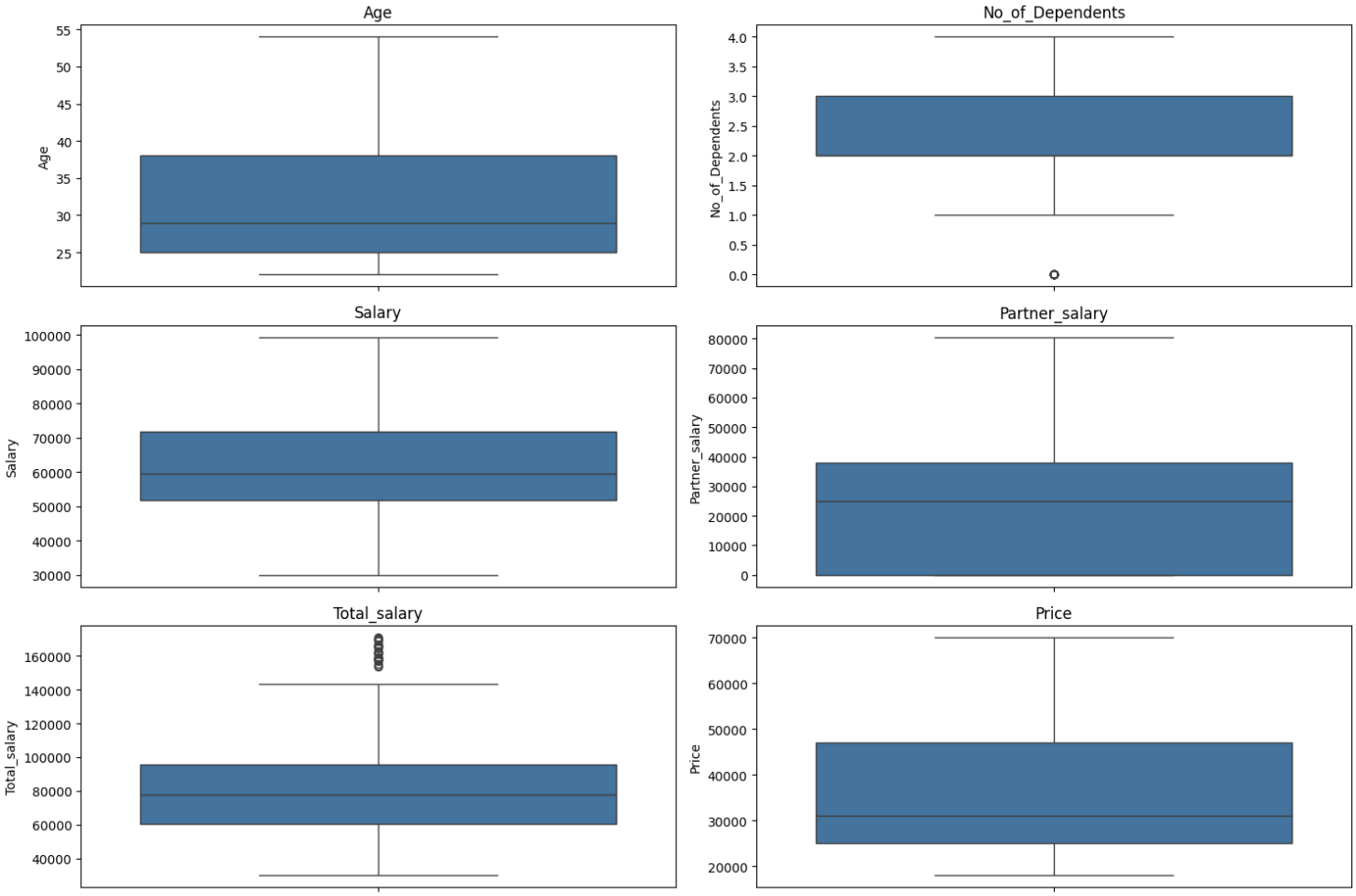

Figure 3: Boxplot of numerical variables for detecting outliers

- Distribution of Age, Salary, Partner salary, Total Salary, Price are right skewed.
- It can be seen that Total Salary and No. of Dependent Column have outliers
- The outliers in this column can be considered as genuine values because No. of Dependent Person can be 0 and Total salary of an individual can be more than 1,40,000. In this case we will not consider these values as outliers.

Bivariate Analysis

To explore the relationships between pairs of variables. Use scatter plots and correlation matrices

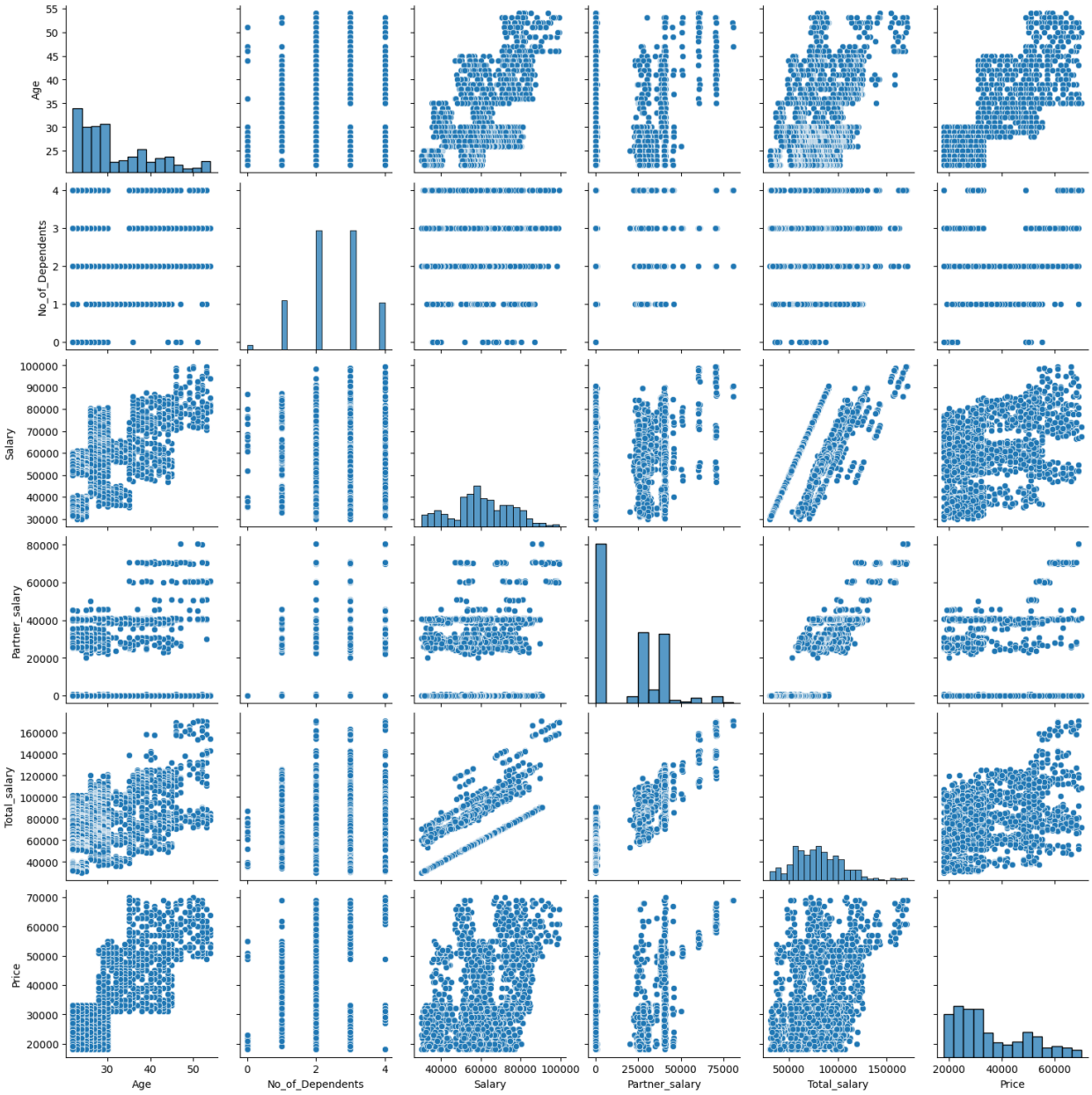

Figure 4: Pair plots for numerical variables

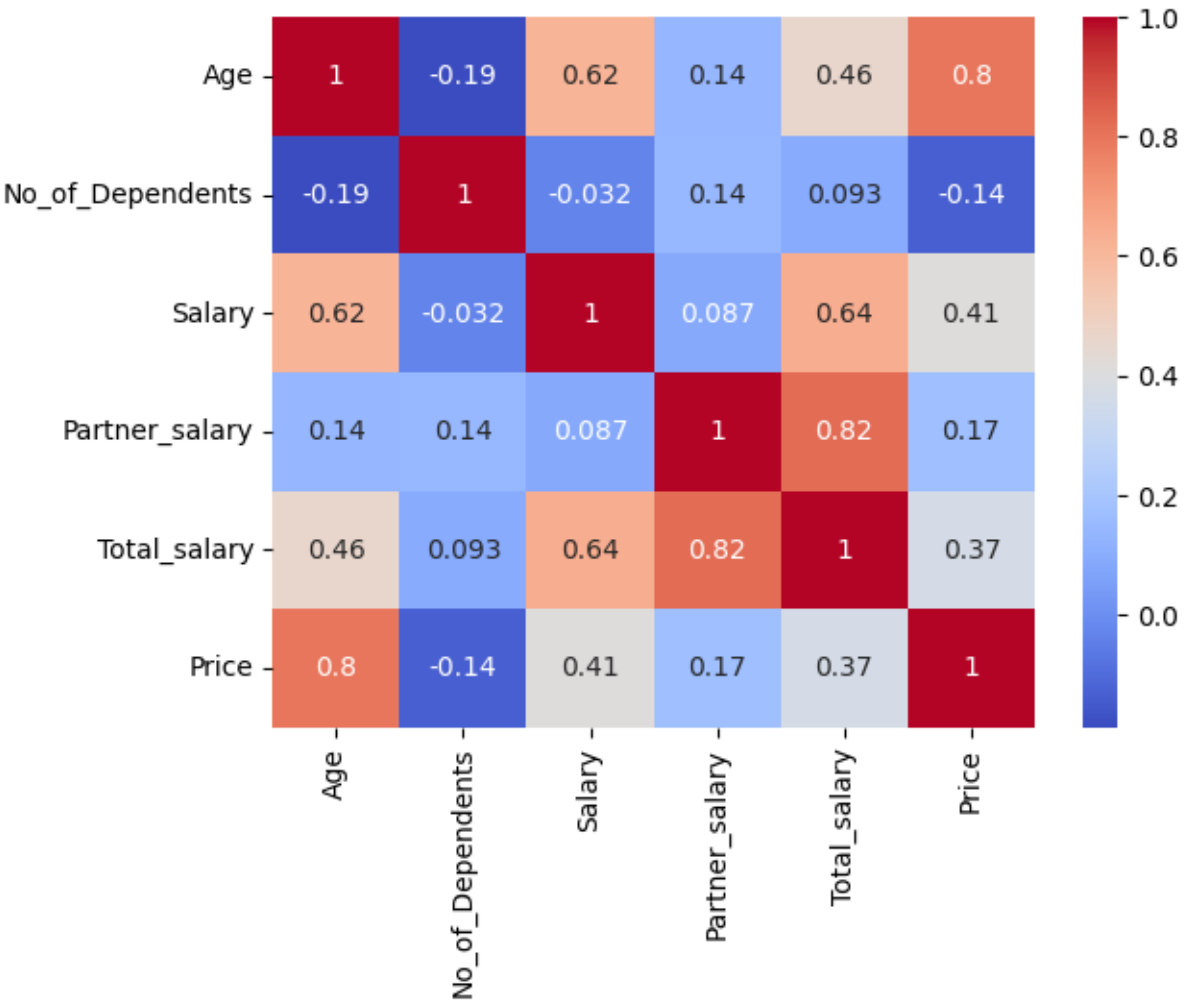

Figure 5: Correlation Heatmap of Numerical Variables

Observations:

- The Price of cars & Age of the buyers is positively related.
- There is very strong correlation between Age of the individual and Price of the car which implies that as age increase the chances of a customer buying a higher priced car increase.
- There is a negative correlation between No. of dependent and Age/ Price of the car/ Salary.

Questions -

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

Answer: No, as per data female customers prefer SUVs compared to male

Make	Gender	
Hatchback	Male	97.08
	Female	2.58
	Unknown	0.34
SUV	Female	58.25
	Male	39.73
	Unknown	2.02
Sedan	Male	73.50
	Female	20.09
	Unknown	6.41

Name: proportion, dtype: float64

Table 6: Proportion table among type of cars and gender

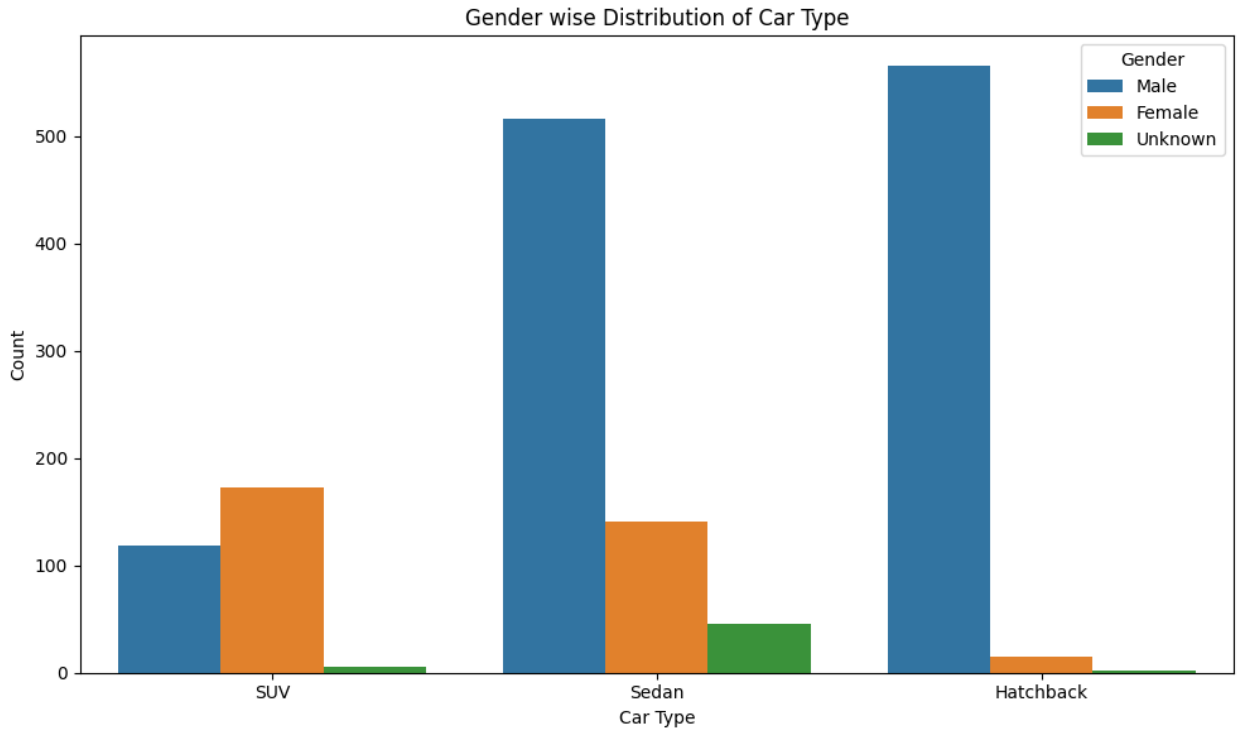

Figure 6: Gender wise distribution of car type

Observations/ Insights:

- 97.08% of Hatchback purchases made by men and only a small fraction, 2.58%, of Hatchback purchases are made by women.
- Most SUV buyers are female, with 58.25% of SUV purchases made by women. And Men contribute to 39.73% of SUV purchases.
- 73.50% of Sedan purchases made by men. Females make up 20.09% of the Sedan purchases.

Inference:

- Sedan and Hatchback are predominantly purchased by male customer.
- Female customers preferred to buy sedan as per data.

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

Answer: There is a very high chance that a salaried person will buy a sedan car.

```
Profession  Make
Business    Sedan        306
            Hatchback    290
            SUV           89
Salaried    Sedan        396
            Hatchback    292
            SUV          208
Name: count, dtype: int64
```

Table 7: Count table among profession and type of cars

```
Profession  Make
Business    Sedan       44.67
            Hatchback   42.34
            SUV         12.99
Salaried    Sedan       44.20
            Hatchback   32.59
            SUV         23.21
Name: proportion, dtype: float64
```

Table 8: Proportion table among profession and type of cars

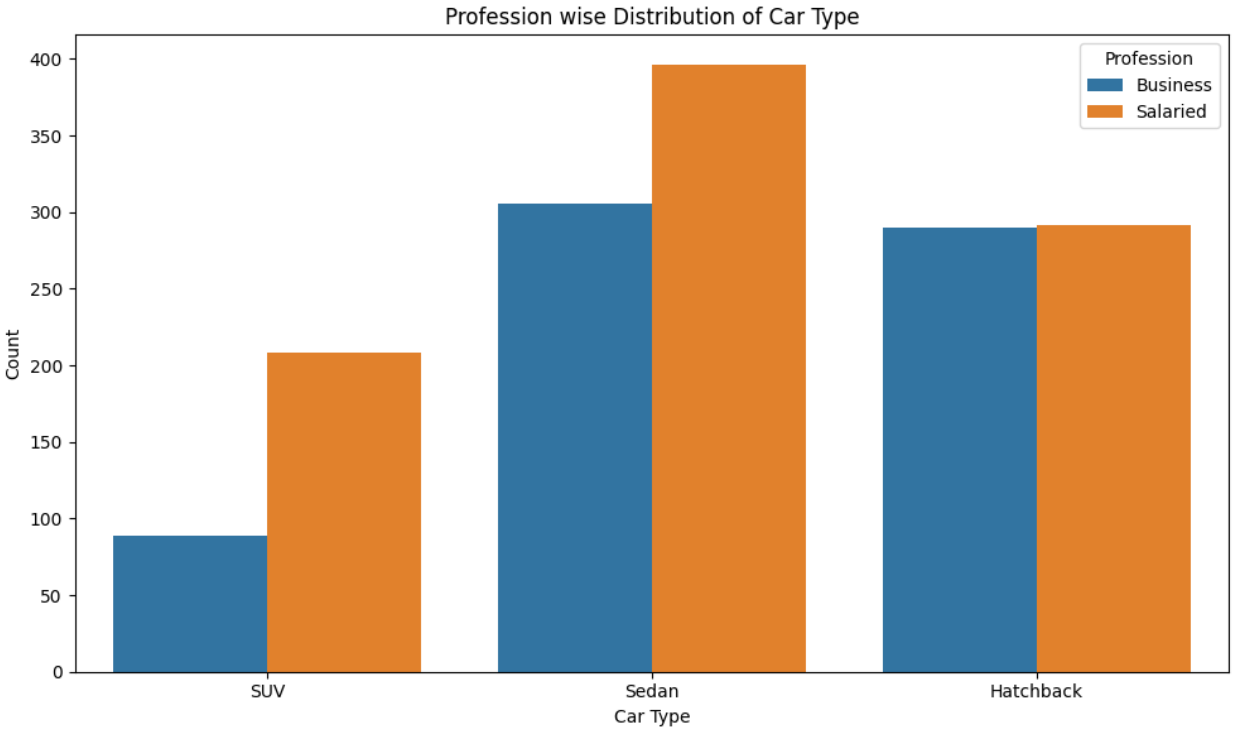

Figure 7: Profession wise distribution of car type

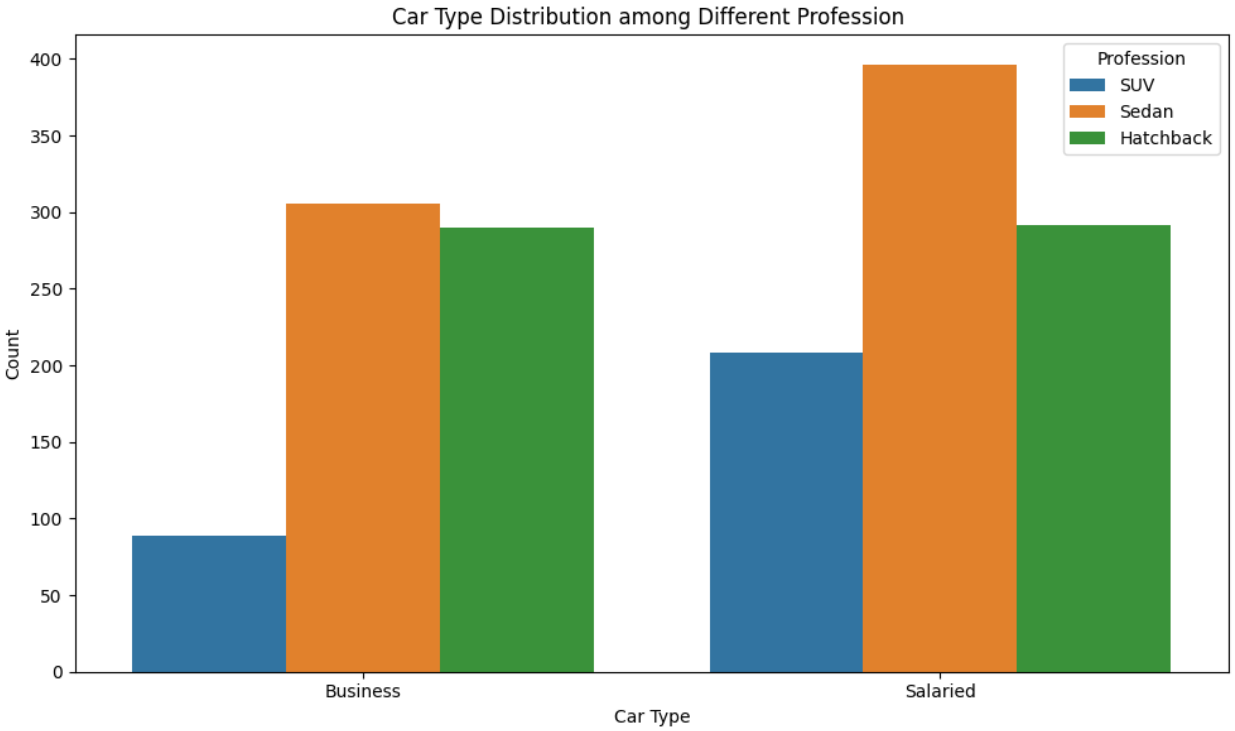

Figure 8: Car type distribution among different profession

Observations/ Insights:

- 44.67% Business Individuals purchases sedan, 42.34 Business Individuals purchases hatchback
- 44.2 % Salaried person bought sedan and 32.59% of salaried people brought hatchback.
- 56.41% of Sedan purchases are made by Salaried Person. Businesspeople make up 43.59% of Sedan purchases.

Inference:

- All types of cars are predominantly purchased by salaried customers.
- Female customers preferred to buy sedan as per data.

3. 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?

Answer: There is no evidence that can claim that a salaried male is an easier target for a SUV sale over a Sedan sale.

```
Profession  Gender   Make
Business    Female   SUV         52.38
                     Sedan       47.62
            Male     Hatchback   51.70
                     Sedan       42.40
                     SUV          5.90
            Unknown  Sedan       90.48
                     Hatchback    4.76
                     SUV          4.76
Salaried    Female   SUV         52.68
                     Sedan       40.62
                     Hatchback    6.70
            Male     Sedan       43.59
                     Hatchback   43.12
                     SUV         13.28
            Unknown  Sedan       81.25
                     SUV         15.62
                     Hatchback    3.12
Name: proportion, dtype: float64
```

Table 9: Proportion table among Profession, Gender and Type of car

Observations/ Insights:

- Salaried Male who prefers to buy Sedan is 43.59%.
- Business Male who prefers to buy SUV is 5.9%, which is least among all.

Inference:

- Salaried Male is not an easier target for a SUV sale.
- Salaried Female is easier to target for SUV Sale over a sedan sale.
- Business Female have a slight preference for SUVs over Sedans.
- Salaried Males prefer Sedan or hatchback over SUVs.
- Hatchback Type car is most preferred by Business Male.

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

Answer: Based on the data Females are more likely to purchase vehicle with higher cost compared to male customers.

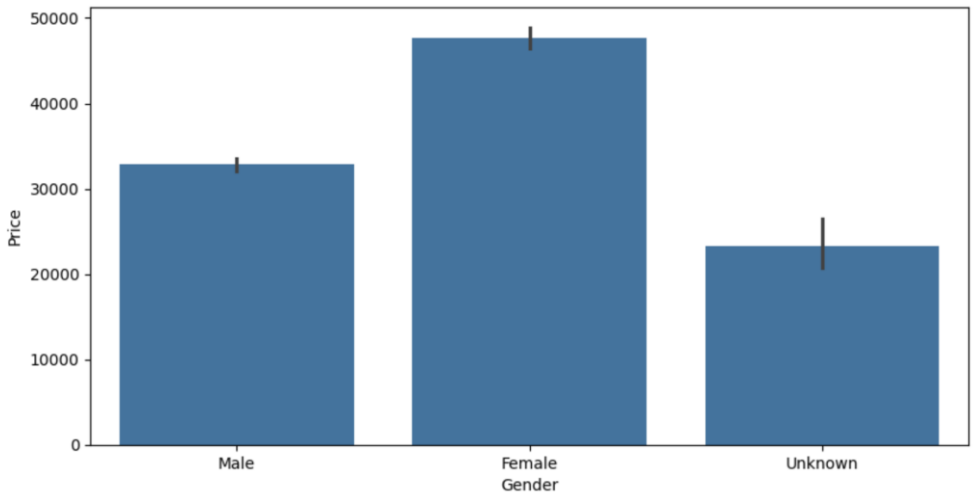

Figure 9: Bar Graph between Gender and Price of the cars

```
Gender
Female    47705.17
Male      32817.35
Unknown   23339.62
Name: Price, dtype: float64
```

Table 10: Mean Price based on Gender

Observations/ Insights:

Avg. Purchasing amount spent by female on buy a automobile is $ 47705

Avg. Purchasing amount spent by male on buy a automobile is $ 32817

Inference:

Females are more flexible in purchasing high-cost vehicle compared to male customers

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

Answer: Based on the data Females are more likely to purchase vehicle with higher cost compared to male customers.

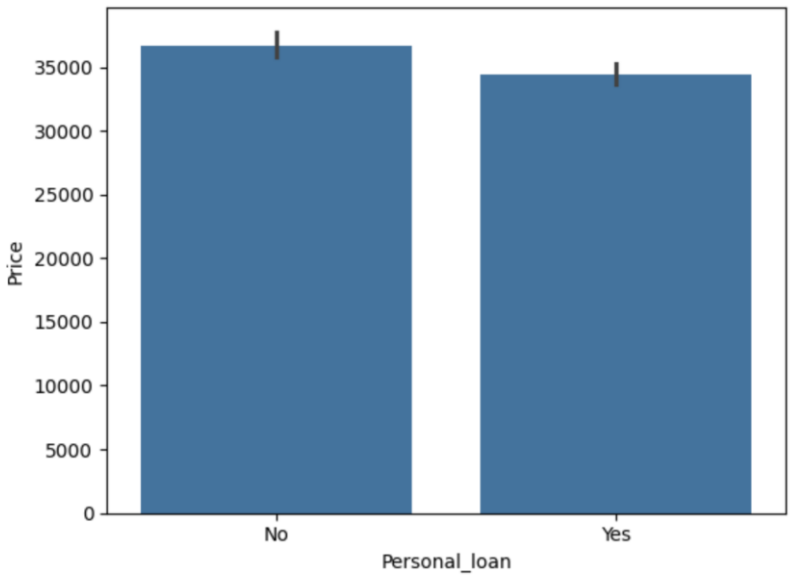

Figure 10: Bar Graph between Price of the car and Individuals having Personal loan or not

```
Personal_loan
No    36742.71
Yes   34457.07
Name: Price, dtype: float64
```

Table 11: Mean Price based on Individuals having personal loan

Observations/ Insights:

- Average spent on purchasing automobile by individuals who took a personal loan is 34,457.
- Average spent on purchasing automobile by individuals who do not take a personal loan, is 36,742.
- There is very less difference between the Average amount spent on purchasing automobile by individuals who took personal loan and who do not take personal loan.

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

Answer: Working partner have little influence over the price of the car.

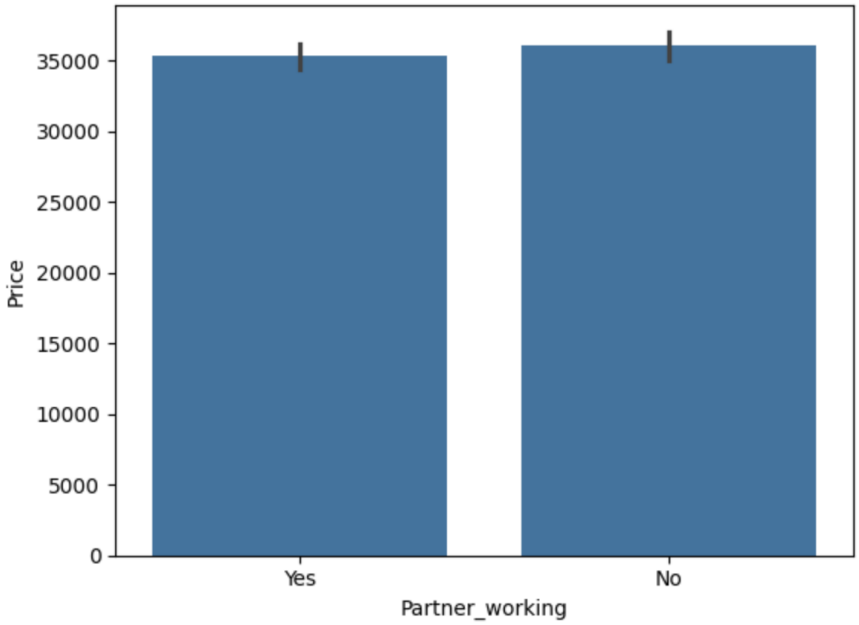

Figure 11: Bar Graph between Working Partner and Price of the car

```
Partner_working
No    36000.00
Yes   35267.28
Name: Price, dtype: float64
```

Table 12: Mean Price based on Individuals having working partner

Observations/ Insights:

- If the partner is not working, the average price of the car is slightly greater than if partner is working, which implies that Working partner little influences over the price of the car.

Actionable Insights:

- The count of Married customers who brought vehicles is very high compared to the count of Single customers.
- The count of customers without a house loan is also significantly higher compared to the count of customers with a loan.
- Sedans are the most selling car type and SUVs are the least selling car type.

Business Recommendations:

- Married and Without house loan customers are key customer segment to increase the sale.
- Male customers prefer sedans and sedan is most selling car type, which make it good for targeted marketing campaigns and boost sales

This data-driven analysis will empower Austro Motor Company to make informed marketing decisions, which may lead to a competitive advantage in the automotive industry.

THE END