# Business Report

# PDS Coded Project

PGPDSBA

Chithira Raj

# Table of Contents

# List of Tables



# List of Figures

# 1. Objective

Austo Motor Company aims to optimize its current marketing campaign effectiveness following discussions in a recent board meeting. The board has resolved to collaborate with analytics experts to enhance the campaign's performance. The objective is to leverage data analysis to understand customer demand better, ultimately improving the overall customer experience. The focus is on addressing key inquiries identified by the Data Science team through comprehensive data analysis. This effort aims to provide actionable insights that will support business improvement initiatives. By uncovering meaningful patterns and trends in customer behavior, we seek to refine our marketing strategies and tailor our offerings to meet customer preferences more effectively. The outcomes of this analysis will inform strategic decisions aimed at elevating customer satisfaction and driving business growth in the competitive automotive market.

# 2. Data Dictionary

| S.No. | Variables | 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 |

Table 1

# 3. Data Overview

## 3.1. Import libraries and load the data

| | 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 | 70700.000 | 170000 | 61000 | SUV |
| 1 | 53 | Femal | Salaried | Married | Post Graduate | 4 | Yes | No | Yes | 95500 | 70300.000 | 165800 | 61000 | SUV |
| 2 | 53 | Female | Salaried | Married | Post Graduate | 3 | No | No | Yes | 97300 | 60700.000 | 158000 | 57000 | SUV |
| 3 | 53 | Female | Salaried | Married | Graduate | 2 | Yes | No | Yes | 72500 | 70300.000 | 142800 | 61000 | SUV |
| 4 | 53 | Male | Salaried | Married | Post Graduate | 3 | No | No | Yes | 79700 | 60200.000 | 139900 | 57000 | SUV |

*Figure 1*

## 3.2. Check the structure of data

Shape of the dataset: 1581 rows and 14 columns

## 3.3. Check the types of the 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
```

*Figure 2*

## 3.4. Check for and treat (if needed) 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*

### *3.4.1. Gender column:*

Null values in Gender column may represent non-binary gender identities, we can replace them with the label "Other".

### *3.4.2. Partner_salary column:*

1. Fill zero in Partner_salary for all the rows with Marital_status as Single.

```
Age                 0
Gender              0
Profession          0
Marital_status      0
Education           0
No_of_Dependents    0
Personal_loan       0
House_loan          0
Partner_working     0
Salary              0
Partner_salary     16
Total_salary        0
Price               0
Make                0
dtype: int64
```

*Figure 4*

2. There are 16 missing values in Partner_salary.

3. We will treat these missing values after understanding the distributions of features in the data, the relationships that exist in the data. This will help us impute these values more effectively.

## 3.5. Check the statistical summary

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| **Age** | 1581.000 | 31.922 | 8.426 | 22.000 | 25.000 | 29.000 | 38.000 | 54.000 |
| **No_of_Dependents** | 1581.000 | 2.458 | 0.943 | 0.000 | 2.000 | 2.000 | 3.000 | 4.000 |
| **Salary** | 1581.000 | 60392.220 | 14674.825 | 30000.000 | 51900.000 | 59500.000 | 71800.000 | 99300.000 |
| **Partner_salary** | 1565.000 | 19062.428 | 19576.736 | 0.000 | 0.000 | 25000.000 | 38100.000 | 80500.000 |
| **Total_salary** | 1581.000 | 79625.996 | 25545.858 | 30000.000 | 60500.000 | 78000.000 | 95900.000 | 171000.000 |
| **Price** | 1581.000 | 35597.723 | 13633.637 | 18000.000 | 25000.000 | 31000.000 | 47000.000 | 70000.000 |

*Figure 5*

## 3.6. Check for and treat (if needed) data irregularities

### *3.6.1. Duplicates*

No Duplicates, since there is no primary key, we are unable to remove rows even if there are duplicates present.

### *3.6.2. Gender:*

```
Gender
Male      1199
Female     327
Other       53
Femal        1
Femle        1
Name: count, dtype: int64
```

*Figure 6*

Replace misspelled words in Gender column

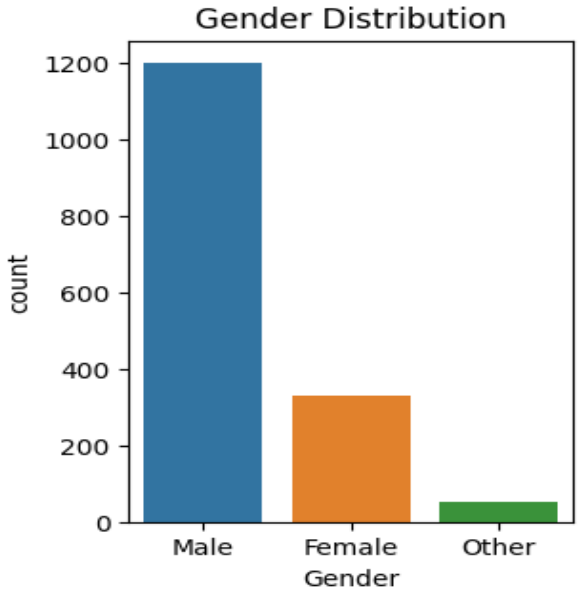



*Figure 7*

## 3.7. Percentage of categorical columns

```
Gender
Male      0.758
Female    0.208
Other     0.034
Name: proportion, dtype: float64
------------------------------------
Profession
Salaried   0.567
Business   0.433
Name: proportion, dtype: float64
```

```
Marital_status
Married   0.913
Single    0.087
Name: proportion, dtype: float64
----------------------------------------
Education
Post Graduate   0.623
Graduate        0.377
Name: proportion, dtype: float64
----------------------------------------
Personal_loan
Yes   0.501
No    0.499
Name: proportion, dtype: float64
----------------------------------------
House_loan
No    0.667
Yes   0.333
Name: proportion, dtype: float64
----------------------------------------
Partner_working
Yes   0.549
No    0.451
Name: proportion, dtype: float64
----------------------------------------
Make
Sedan       0.444
Hatchback   0.368
SUV         0.188
Name: proportion, dtype: float64
----------------------------------------
```

*Figure 8*

## 3.8. Observations and Insights

- The age distribution appears relatively centered around the late 20s to late 30s, with a few individuals in their early 20s and 50s. The standard deviation suggests moderate variability in ages across the sample.
- The average number of dependents per individual is slightly above 2, with relatively low variability (standard deviation of less than 1), suggesting a consistent family size within the sample.
- The salary distribution shows a moderate spread around the mean, with a notable standard deviation indicating variability in income levels among the sample.
- There is significant variation in partner's salaries within the sample, with a sizable portion having no reported income.
- The price distribution indicates a range of expenditures or costs, with a moderate spread around the mean price.
- The majority of individuals in the sample identify as male, with a smaller proportion identifying as female and a small percentage identifying as non-binary or other genders.

- More than half of the sample consists of salaried individuals, while the remaining are engaged in business or entrepreneurial activities.
- The vast majority of individuals in the sample are married, with a relatively small proportion being single.
- A significant portion of the sample has completed post-graduate education, while the remainder have completed undergraduate education.
- A significant portion of the sample has completed post-graduate education, while the remainder have completed undergraduate education.
- A majority of individuals in the sample do not have a house loan, while a third have taken a house loan.
- Slightly more than half of the individuals have a partner who is employed or working.
- Sedans are the most popular type of car among the sample, followed by hatchbacks and SUVs.

# 4. Univariate Analysis

## 4.1. Age

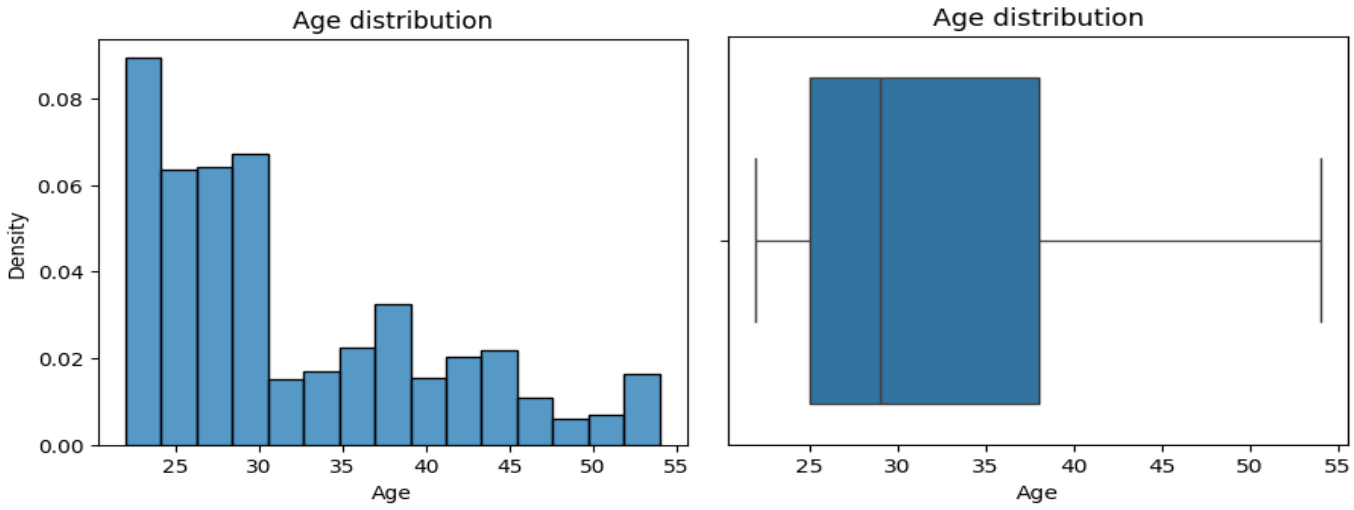
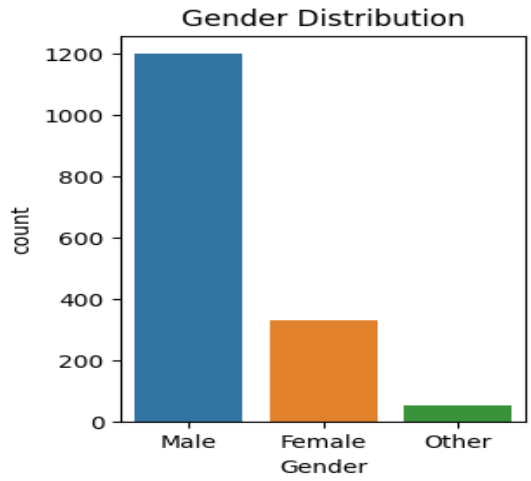



*Figure 9*

Skewed towards right and 75% of the customers are under the age of 40.

## 4.2. Gender




Majority of the customers are male.

*Figure 10*

## 4.3. Profession

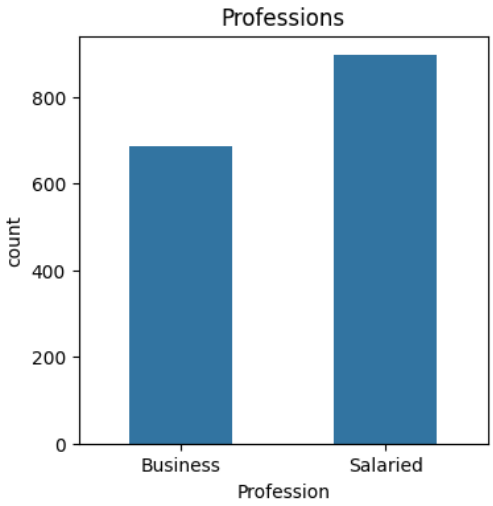
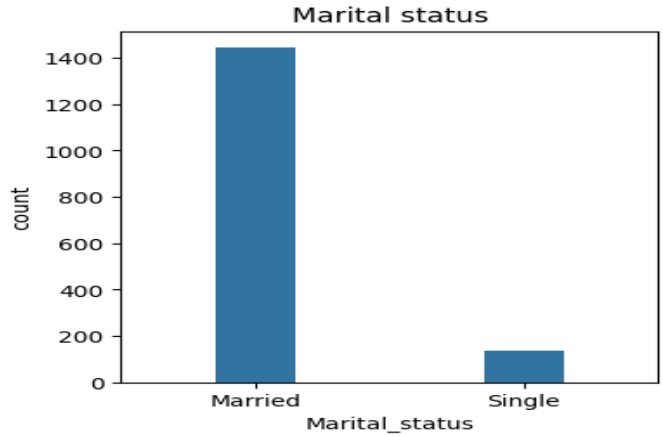



*Figure 11*

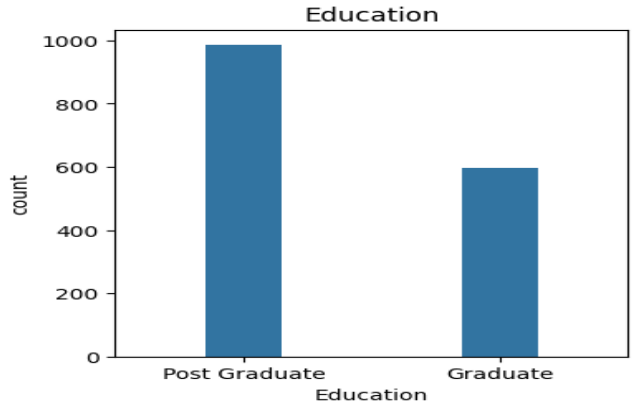
## 4.4. Marital Status




Majority of customers are married.

*Figure 12*

## 4.5. Education




A greater number of individuals who have completed post-graduate studies could suggest higher salaries.

*Figure 13*

## 4.6. Number of dependents

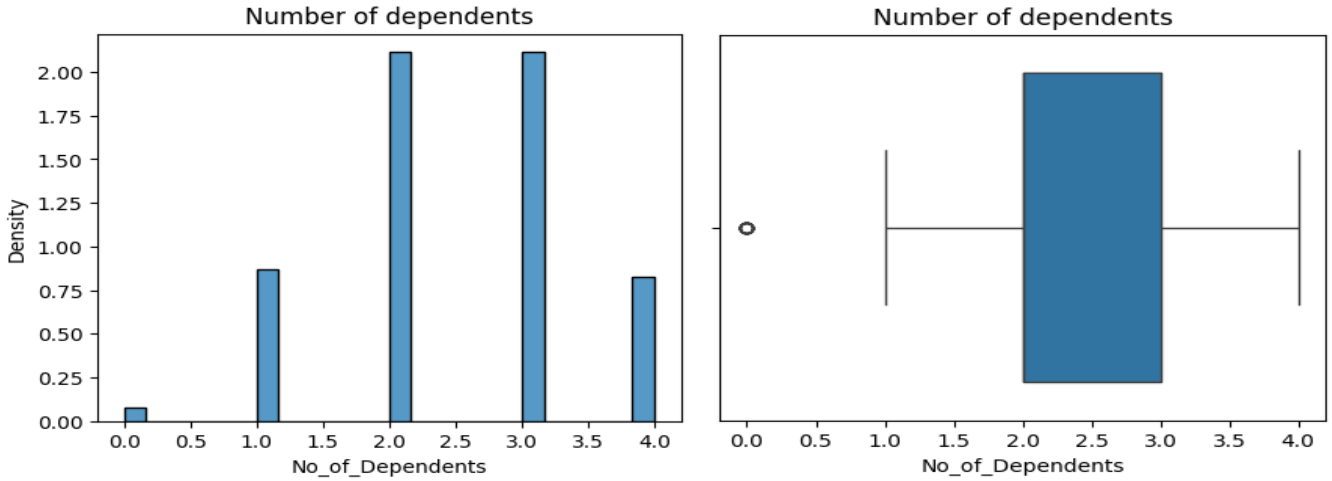



*Figure 14*

Zero dependents are labelled as outliers in the box plot, but we should retain them as such

## 4.7. Personal Loan status

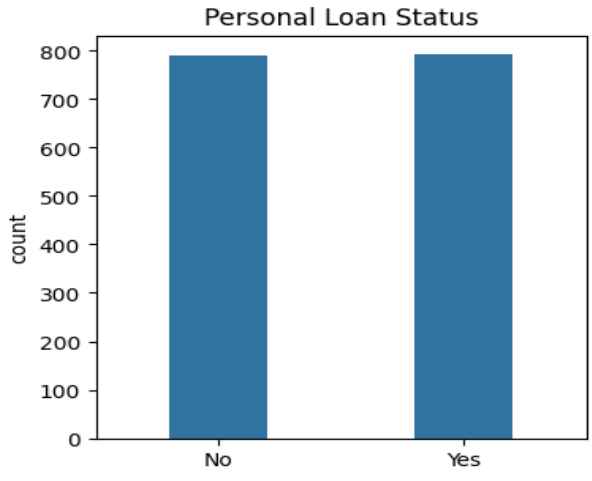
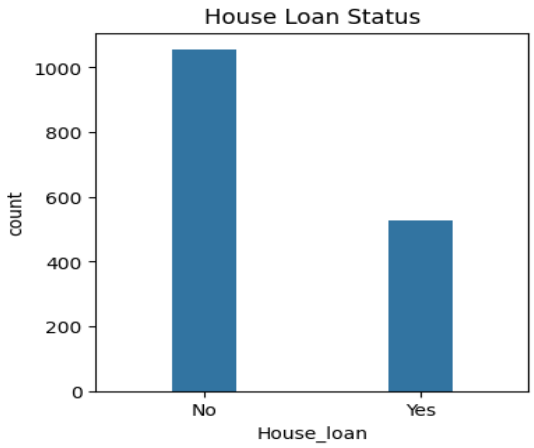



*Figure 15*

Uniformly distributed.

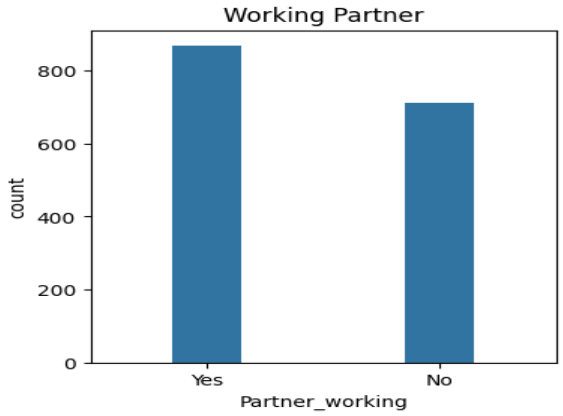
## 4.8. House Loan Status




The majority of customers did not opt for a home loan.

*Figure 16*

## 4.9. Working Partner Status




A higher number of customers have partners who are employed.

*Figure 17*

## 4.10. Individual Salary

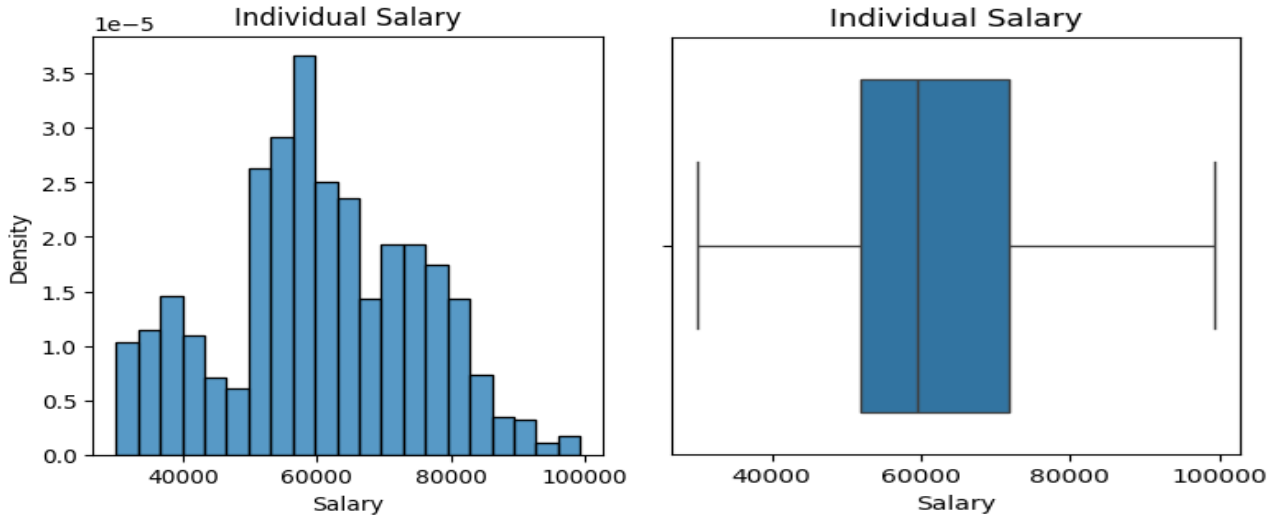



*Figure 18*

Normally distributed with no outliers.

## 4.11. Partner Salary

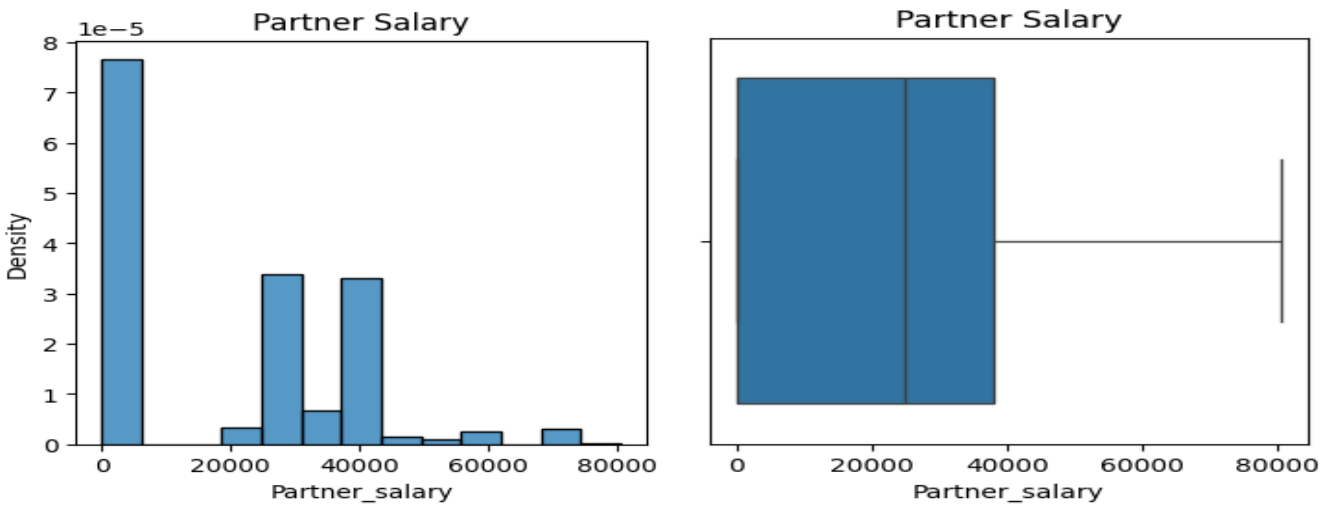



*Figure 19*

- Distribution is right-skewed with no outliers.
- Missing values in the column is replaced with median.

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

*Figure 20*

## 4.12. Total Salary

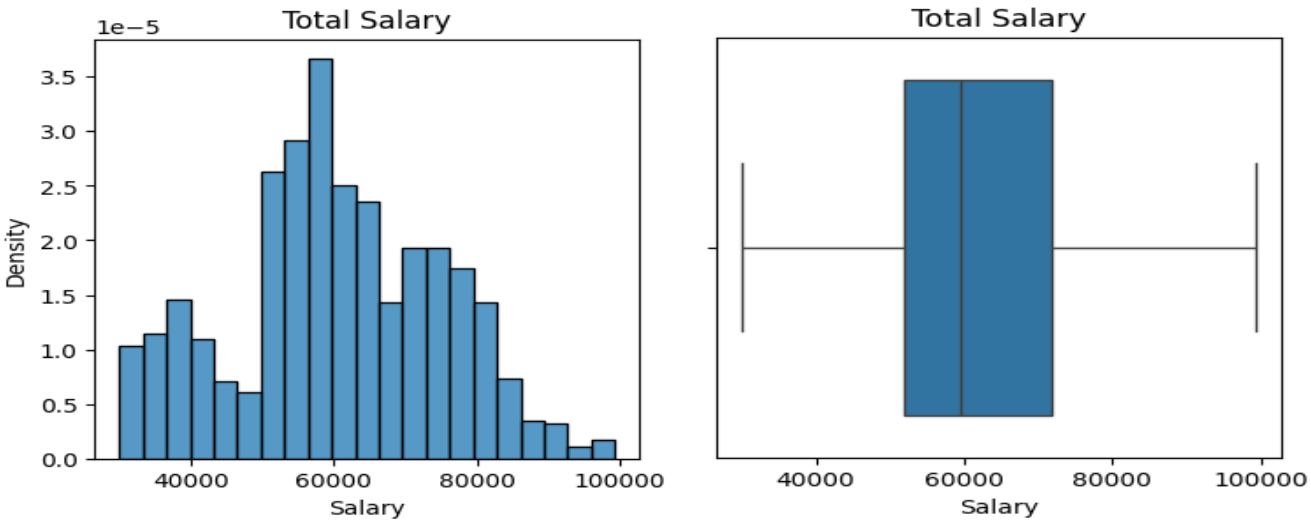



*Figure 21*

Normally distributed with no outliers.

## 4.13. Price

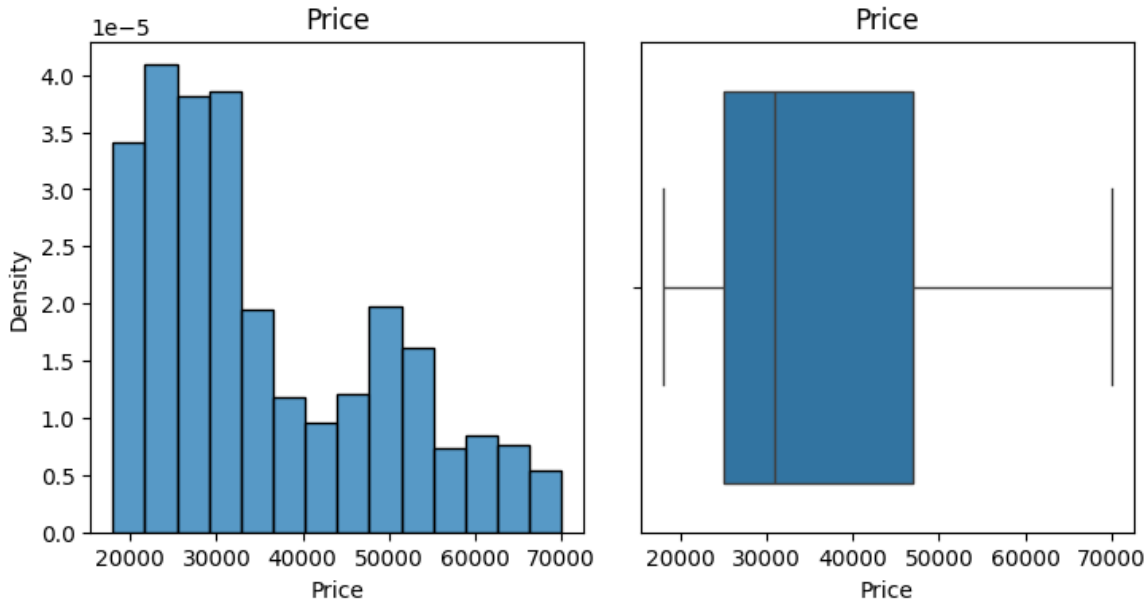



*Figure 22*

The Price column is right-skewed, a larger number of customers purchase cars priced between 20K and 30K.

## 4.14. Car Make

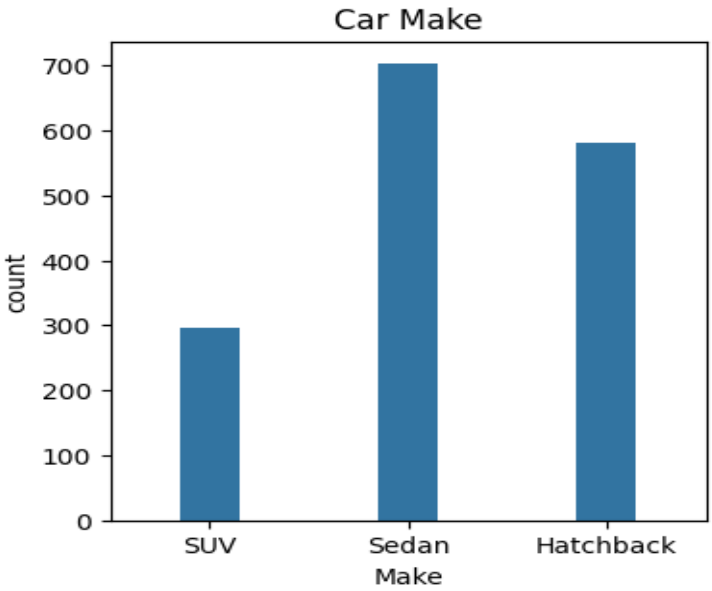



*Figure 23*

Customers show a preference for sedans and hatchbacks over SUVs.

# 5. Bivariate Analysis

## 5.1. Correlation between numerical variables

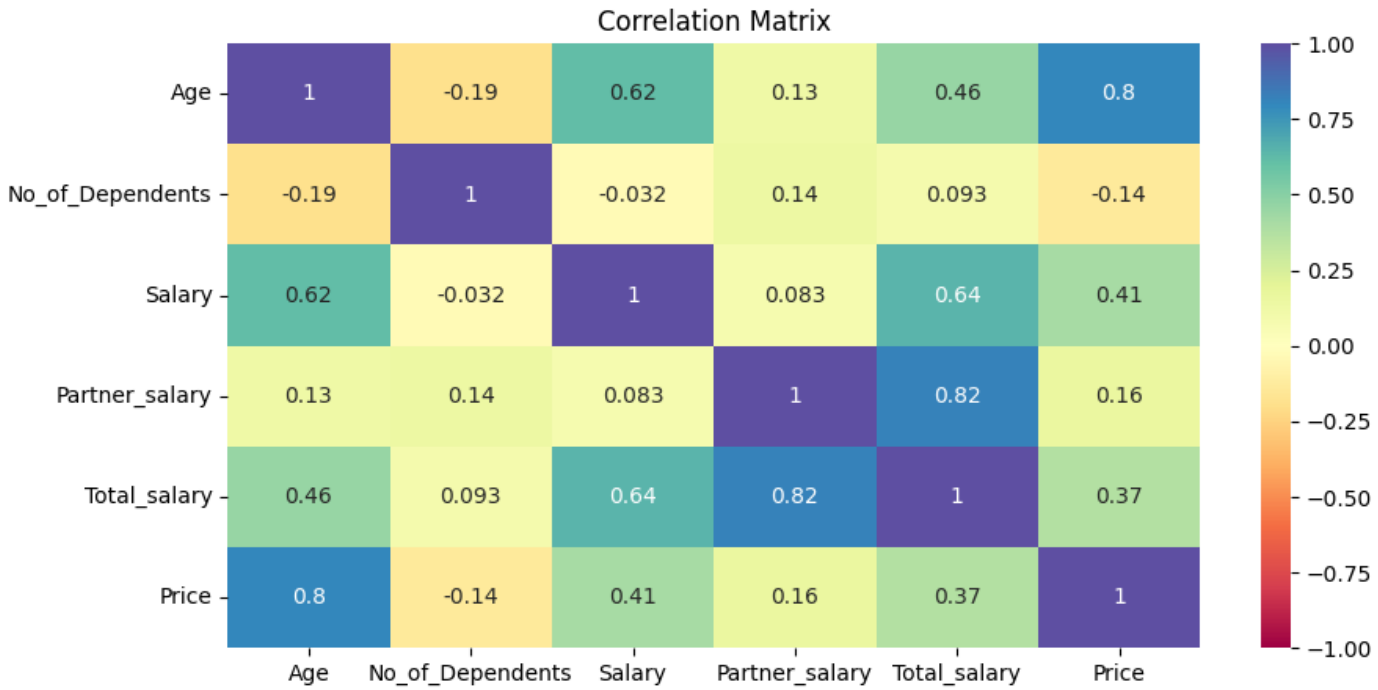



*Figure 24*

### *5.1.1. Insights*

- The price of the car is positively correlated with the age of the person.
- The salary of a partner correlates positively with the total salary in a manner akin to the correlation between individual salary and total salary.
- The age of the person and salary show a positive correlation.
- Slight Negative correlation between salary/age and number of dependents, but this can be regarded as no relationship.
- There is a slight positive correlation between price and total salary, but it's not very high.

## 5.2. Relationship between numerical variables

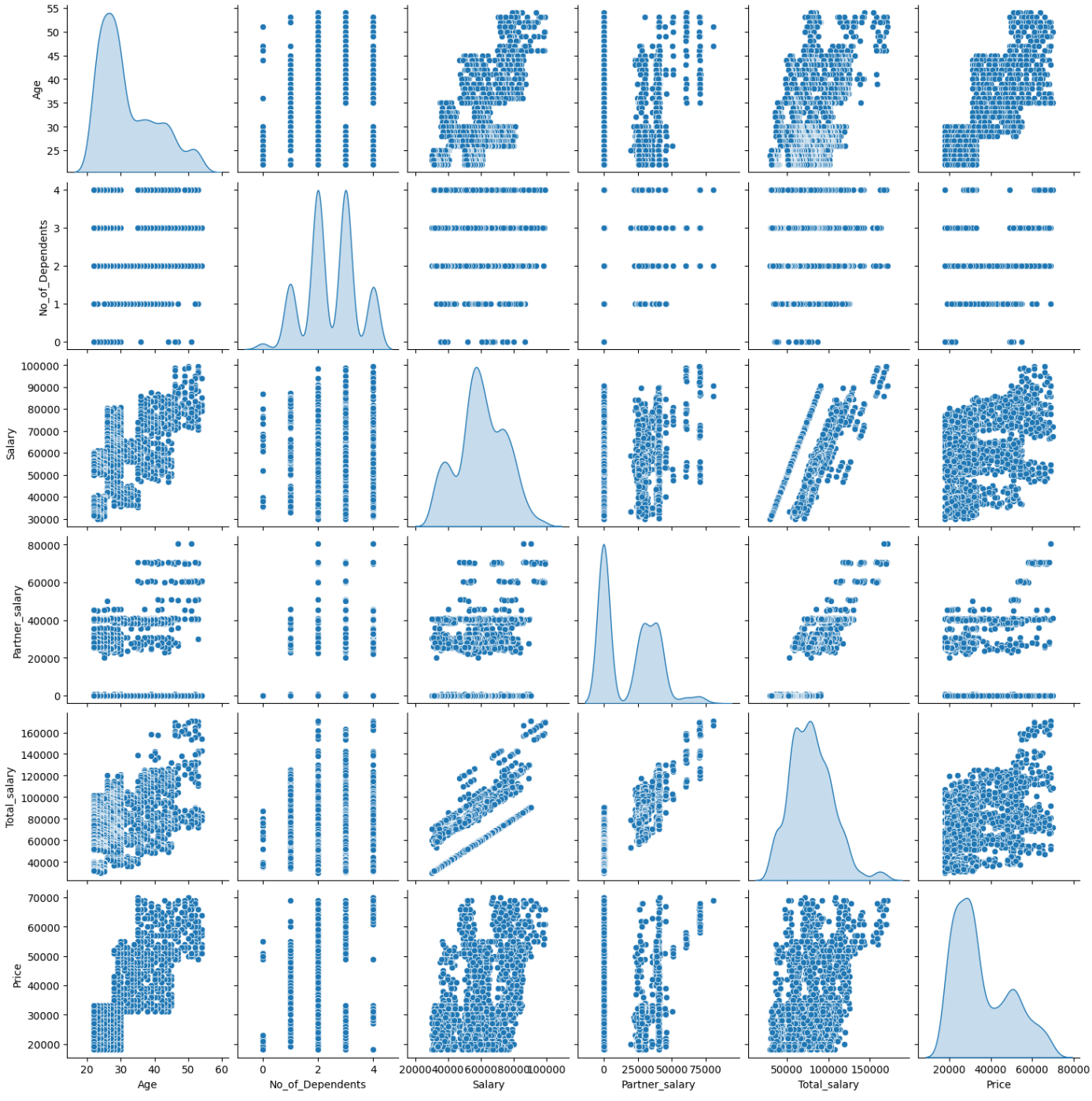



*Figure 25*

### *5.2.1. Insights*

- Age vs Number of dependents: No relationship between Age and number of dependents
- Age vs Salary: There is a positive correlation between Age and Salary
- Age vs Partner salary: No relationship between Age and partner salary
- Age and Total salary: As expected, there is a positive correlation between Age and Total salary but it is not very high.
- Age vs Price: A positive correlation or an increasing trend can be clearly observed between Age and Price
- Number of dependents vs Salary: There is a negative correlation, but it's not high. This can be regarded as no relationship between Number of dependents and salary
- Number of dependents vs Partner Salary: No relationship between Number of dependents and partner salary

- Number of dependents vs Total Salary: No relationship between Number of dependents and total salary
- Number of dependents vs Price: No relationship between Number of dependents and price
- Salary vs Partner salary: No relationship between Salary and Partner salary
- Salary vs Total Salary: There is positive correlation between salary and total salary, it implies that higher individual salaries are associated with higher combined (total) salaries when considering both the individual's salary and their partner's salary.
- Salary vs Price: There is positive correlation between Price and Salary, but it's not very high.
- Partner Salary vs Total Salary: There is positive correlation between Partner_salary and total salary, it implies that higher partner salaries are associated with higher combined (total) salaries when considering both the individual's salary and their partner's salary.
- Partner salary vs Price: No relationship between Price and partner salary
- Total Salary vs Price: There is a slight positive correlation between price and total salary, but it's not very high.

## 5.3. Explore the relationship between categorical vs numerical variables

### *5.3.1. Gender vs Price of cars*

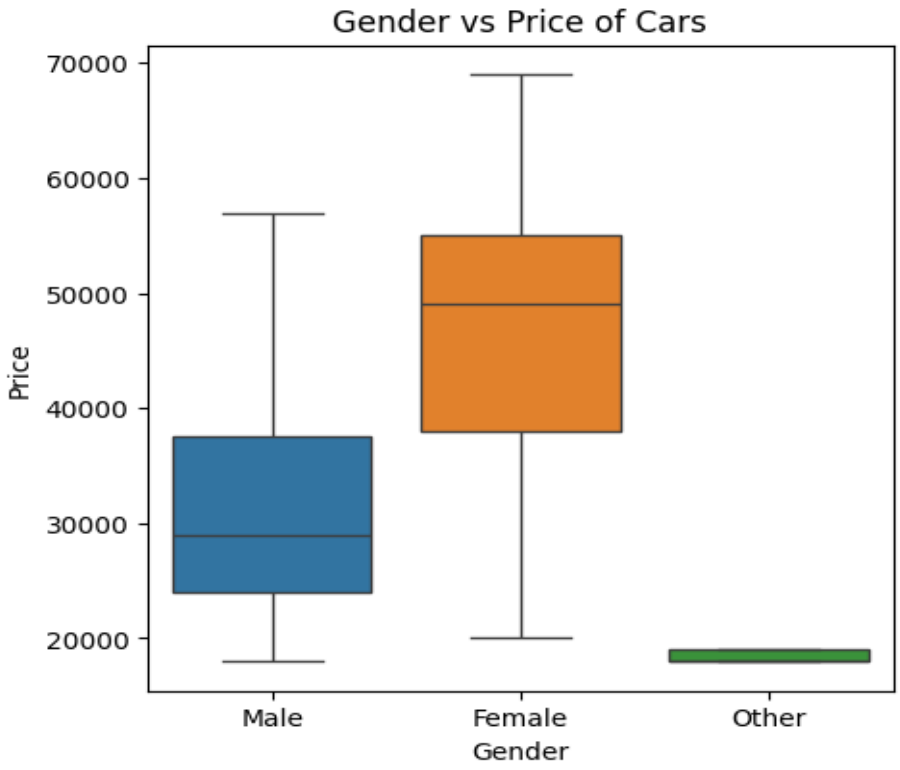
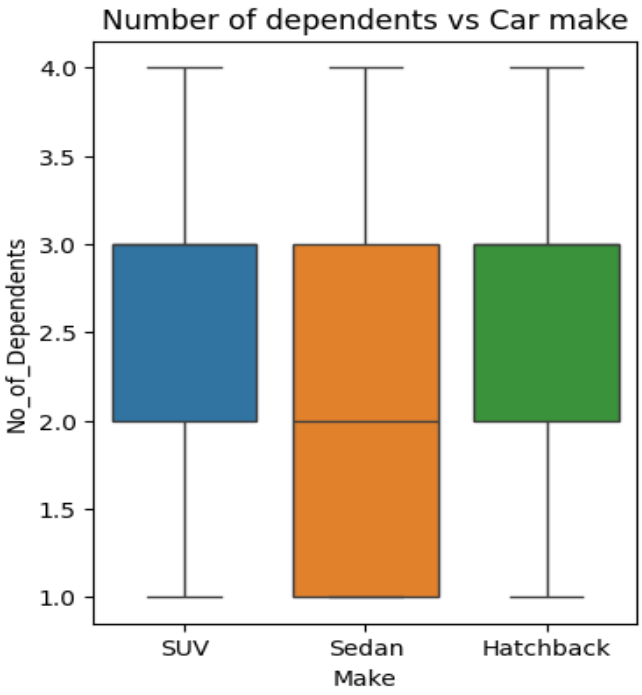



*Figure 26*

There is a significant disparity between the median amount spent on cars by men and women. The median car price for men is around 30,000, whereas for women, it is approximately 50,000, indicating a substantial difference.

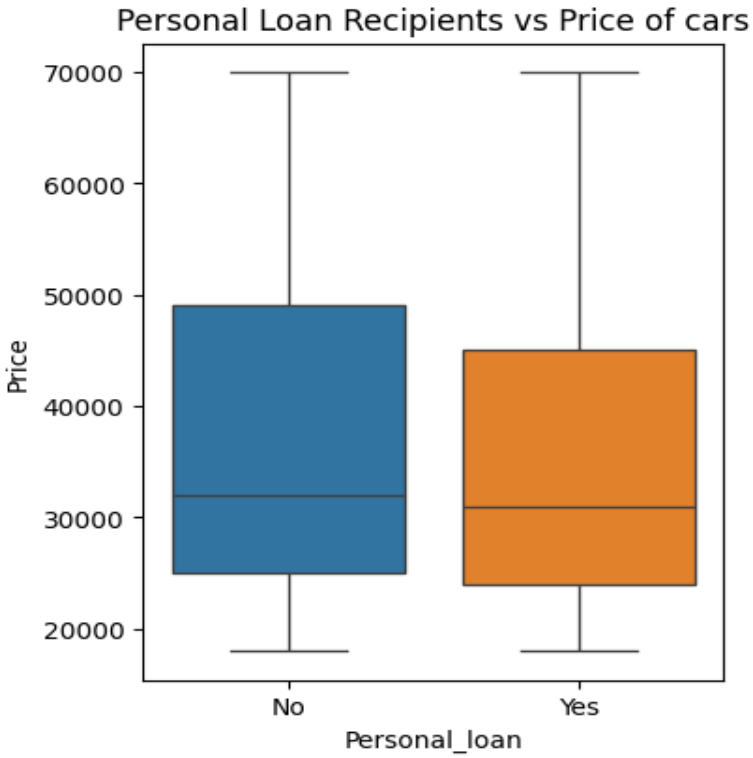
### 5.3.2. Number of dependents vs Car make




Customers may lean towards SUVs or hatchbacks when they have more dependents to accommodate.

*Figure 27*

### 5.3.3. Personal Loan Recipients vs Price of cars

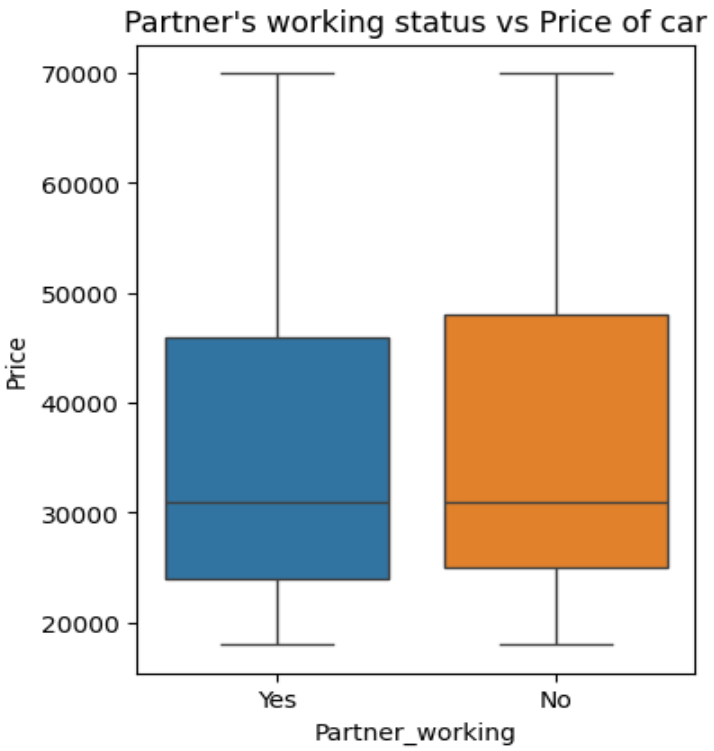



Distribution is comparable in terms of central tendency and median.

Personal Loan do not influence negatively on car expenditure.

*Figure 28*

### *5.3.4. Partner's working status vs Price of car*




Distribution is comparable in terms of central tendency and median.

Unemployment of Partner Does Not Negatively Impact Car Expenditure.

*Figure 29*

## 5.4. Explore the relationship between categorical variables

### *5.4.1. Gender vs Make*

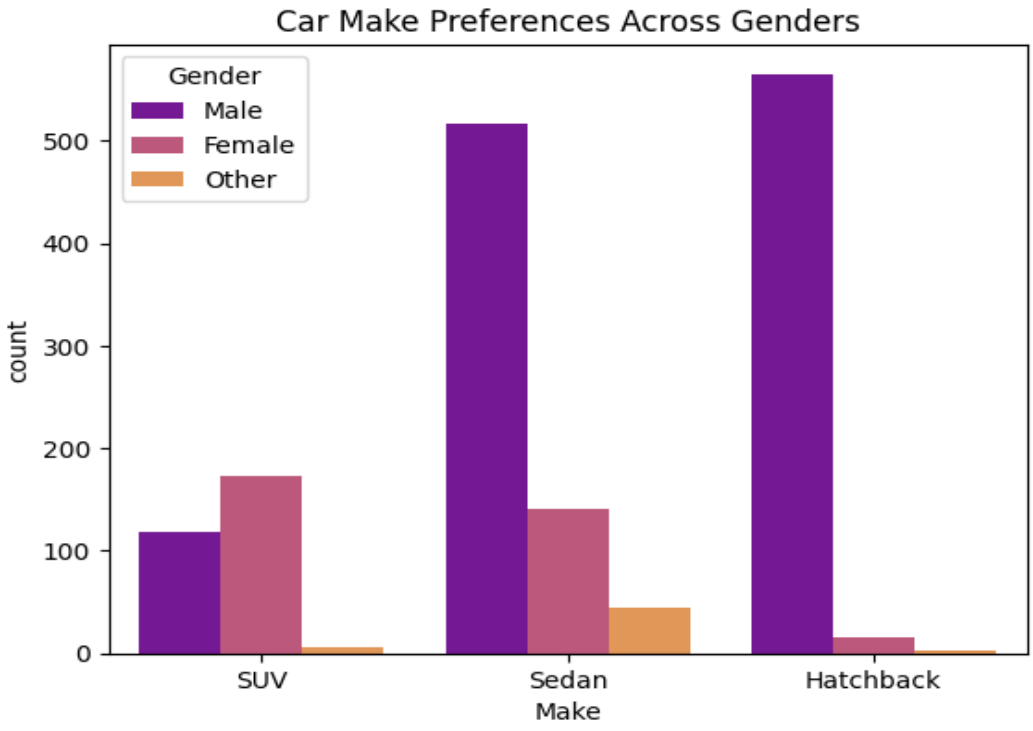
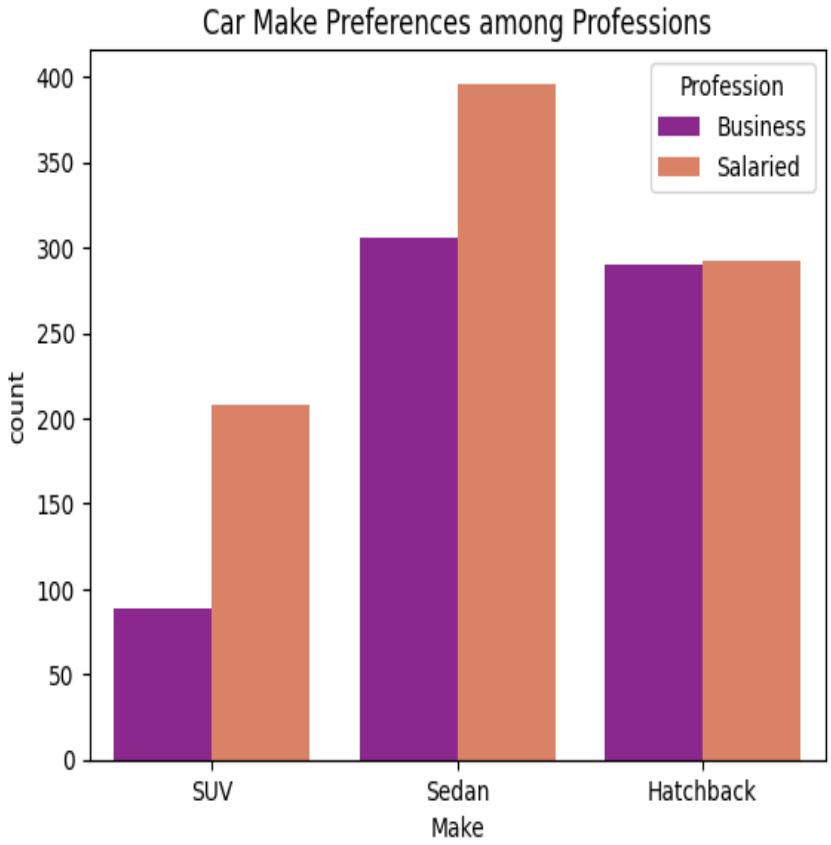



*Figure 30*

Women Favor SUVs, While Men Prefer Hatchbacks and Sedans.

### 5.4.2. Profession vs Make




Sedans Preferred by Both Salaried Employees and Business Owners

*Figure 31*

### 5.4.3. Profession vs Gender vs Make of car

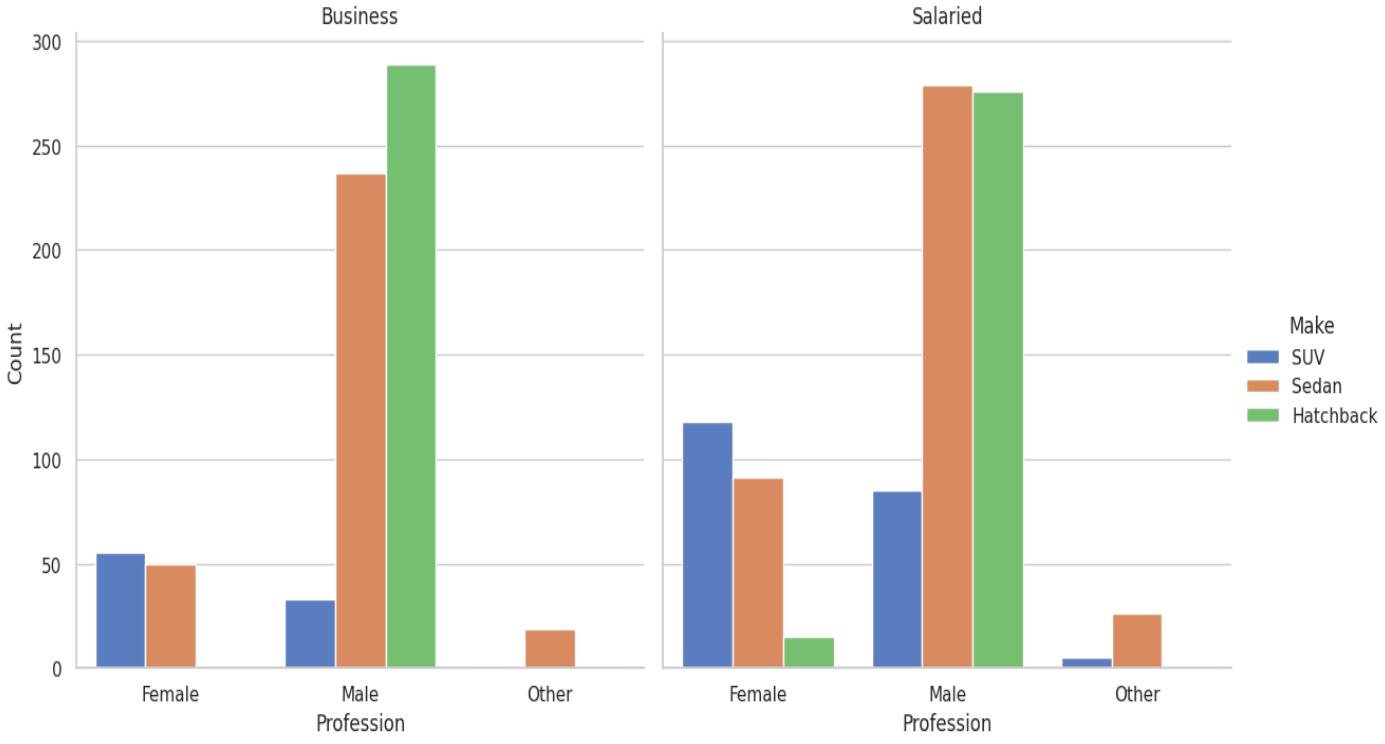



*Figure 32*

Salaried men prefer Sedan and Hatchback over SUV.

# 6. Key Questions

## 6.1. Question 1: Do men tend to prefer SUVs more compared to women?

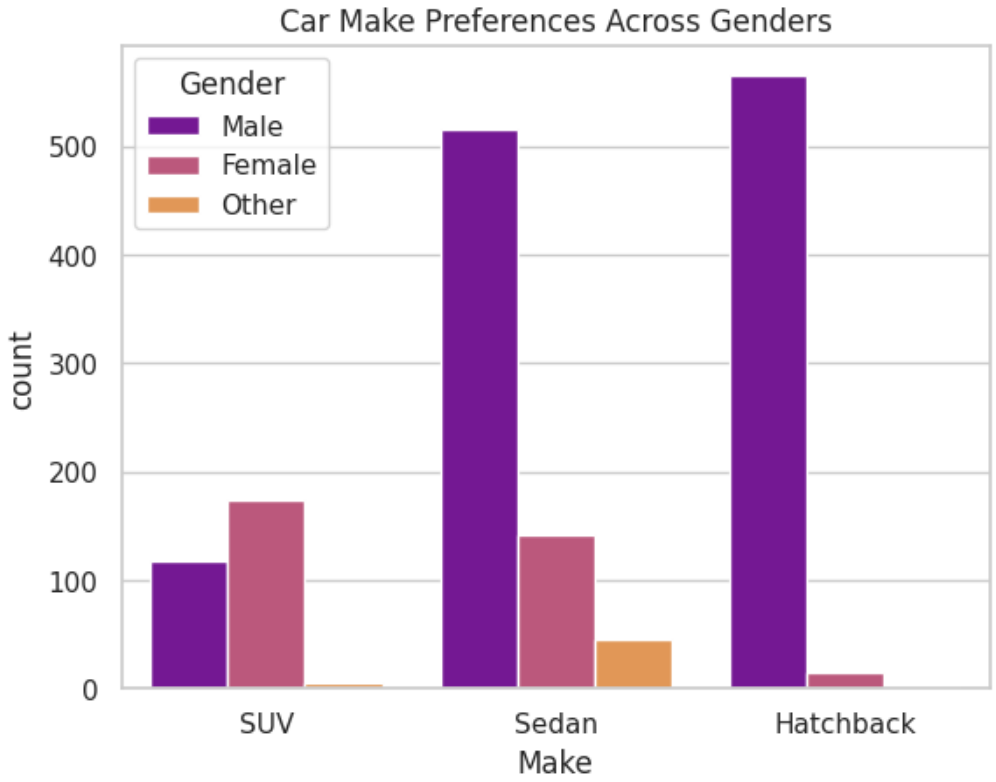



*Figure 33*

| | Gender | Make | Count |
|---|---|---|---|
| **1** | Female | SUV | 173 |
| **4** | Male | SUV | 118 |
| **7** | Other | SUV | 6 |

*Figure 34*

No, Women show a greater preference for SUVs compared to men.

## 6.2. Question 2: What is the likelihood of a salaried person buying a Sedan?

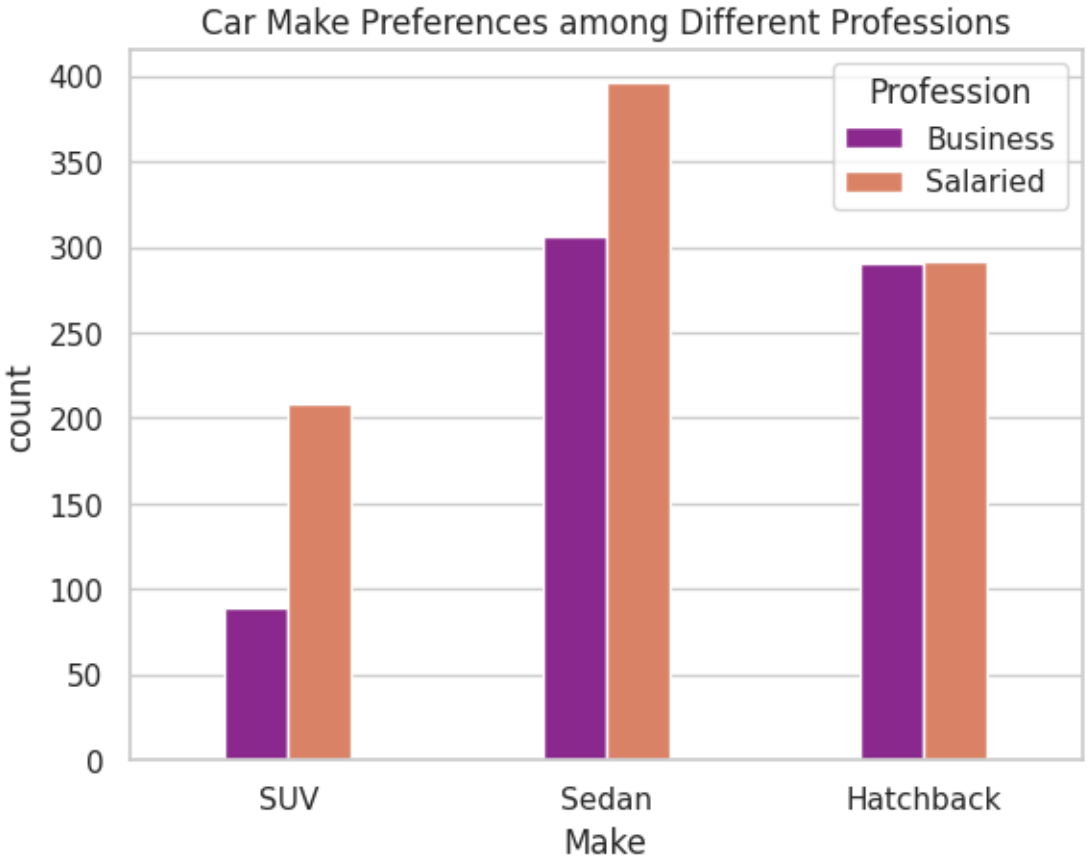



*Figure 35*

| | Profession | Make | Count |
|---|---|---|---|
| **0** | Business | Hatchback | 290 |
| **1** | Business | SUV | 89 |
| **2** | Business | Sedan | 306 |
| **3** | Salaried | Hatchback | 292 |
| **4** | Salaried | SUV | 208 |
| **5** | Salaried | Sedan | 396 |

*Figure 36*

```
Number of Salaried People: 896
Number of People who buys Sedan: 396
Percentage of Salaried person buying a Sedan: 44.2 %
```

*Figure 37*

A salaried person has a 44% probability of purchasing a sedan.

6.3. Question 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?

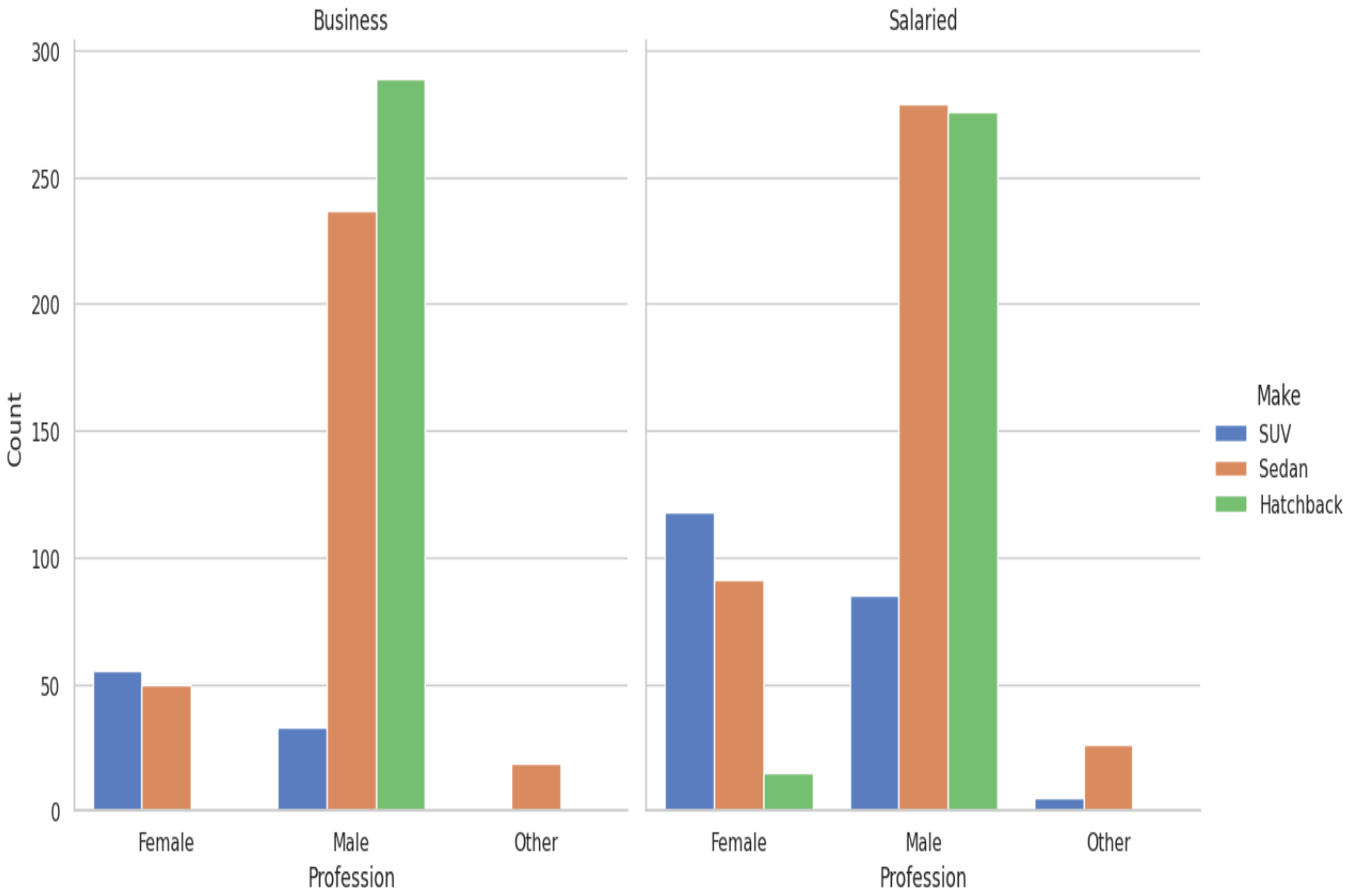



*Figure 38*

According to the data visualization, salaried men tend to prefer sedans and hatchbacks over SUVs. Targeting men for SUV sales could be effective, given their preference for sedans and hatchbacks according to the plot.

## 6.4. Question 4: How does the amount spent on purchasing automobiles vary by gender?

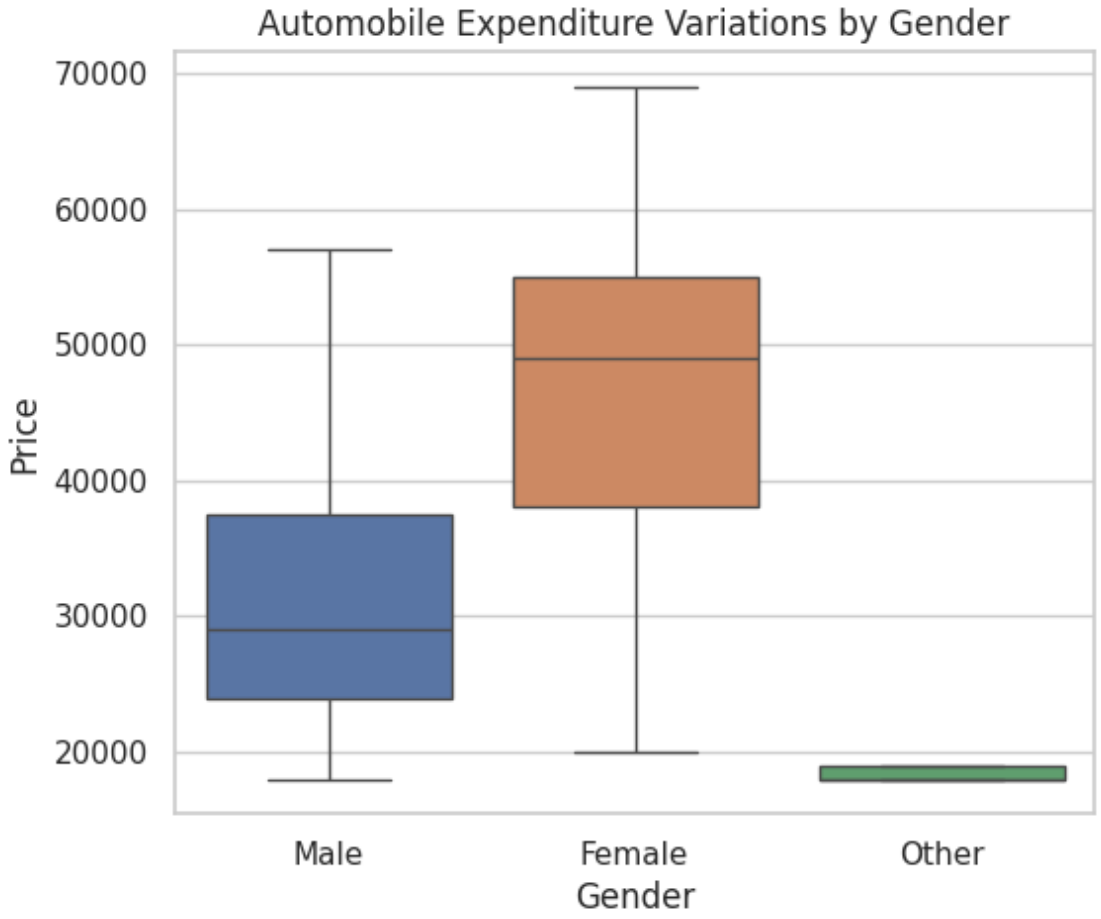



*Figure 39*

```
Expenditure by Male: 39348000
Expenditure by Female: 15695000
Expenditure by Other: 1237000
```

*Figure 40*

Women tend to spend more money on cars compared to men and individuals of other genders.

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

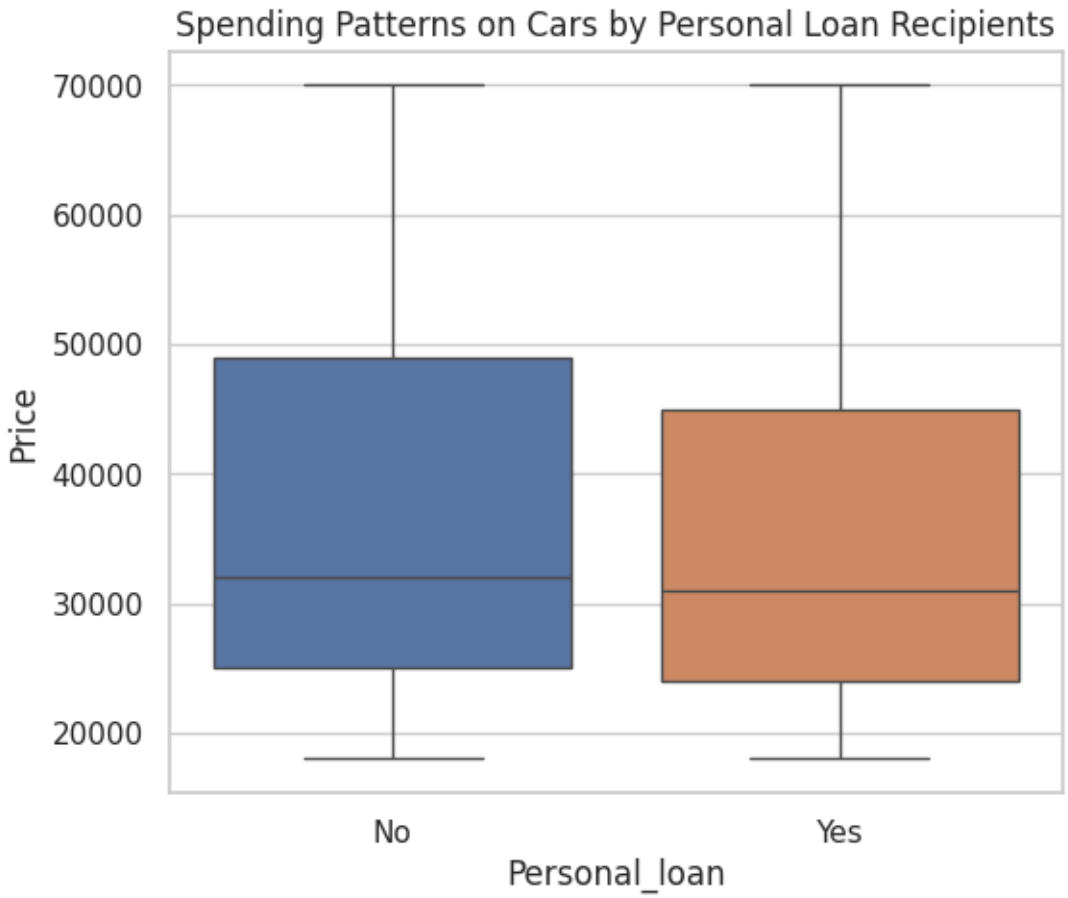



*Figure 41*

Sum of money spent on cars by Personal loan recipient

```
Personal_loan
No     28990000
Yes    27290000
Name: Price, dtype: int64
```

*Figure 42*

Personal loan recipients spend 27290000 on cars.

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

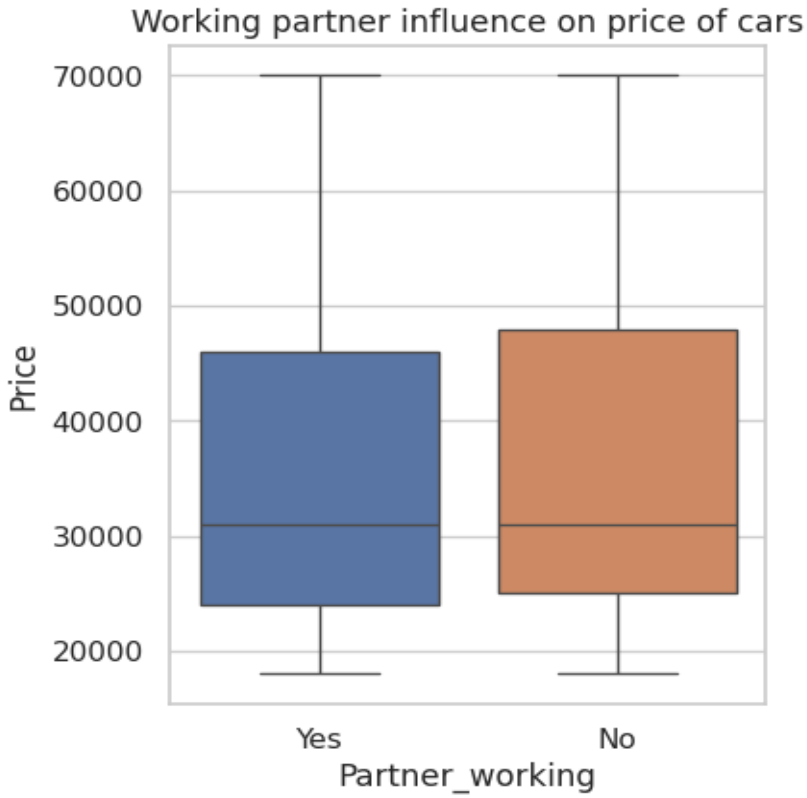



*Figure 43*

### *6.6.1. t-test*

- Null Hypothesis (H0): The mean price of cars bought by individuals with working partners is equal to the mean price of cars bought by individuals without working partners.
- Alternative Hypothesis (H1): The mean price of cars bought by individuals with working partners is different from the mean price of cars bought by individuals without working partners.

The study examined the spending on cars among individuals with a working partner versus those with a non-working partner. It calculated the t-statistic and p-value to substantiate the hypothesis.

```
t_statistic: -1.063364944753348
p-value: 0.28777907676113346
```

*Figure 44*

p-value is greater than the significance level i.e., 0.05. So, we can't reject the null hypothesis. Distribution is comparable in terms of central tendency and median. No negative or positive impact on expenditure on cars based on partner's working status.

# 7. Actionable Insights and Recommendations

## 7.1. Insights

We analysed a dataset containing information from 1,500 customers regarding their car types and expenditures. Several factors influence car purchases, such as age, gender, profession, and loan status. Additionally, individual and partner salaries, as well as the number of dependents, can also play significant roles in these purchasing decisions. Therefore, we identified the factors that can have a positive impact on the marketing campaign.

1. The preferred car type among customers is sedan, followed by hatchback and then SUV.
2. The majority of customers purchase cars priced under 50,000.
3. Most customers are married and customers may lean towards SUVs or hatchbacks when they have more dependents to accommodate.
4. There is a positive correlation between customer age and car price.
5. Women tend to spend more on cars compared to men.
6. The working status of the partner or personal loan status does not impact car purchases.
7. Women prefer SUVs while men prefer hatchbacks.
8. Customers who own businesses prefer hatchbacks and sedans.
9. Salaried males prefer sedans followed by hatchbacks.
10. Men who own businesses prefer hatchbacks over sedans while salaried or businesswomen prefer SUVs.

## 7.2. Recommendations

1. Given the demand for sedans, it would be beneficial to focus more marketing efforts on promoting this car type.
2. Since women tend to spend more on cars, a targeted marketing campaign highlighting specific car features that appeal to women could be effective in attracting this demographic.
3. The marketing campaign could focus on targeting married individuals, who make up the majority of the customer base. However, the campaign can also include strategies to attract the minority of single individuals.
4. Given that customers with more dependents tend to prefer SUVs or hatchbacks, it would be strategic to concentrate marketing efforts on this demographic.
5. Additional data on the timeline of purchases should be acquired to facilitate trend analysis and the development of a predictive model.