



PDS Business Report

Presented by :
Sanjay Rajan J

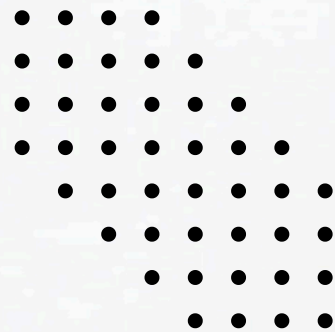


TABLE OF CONTENTS

CHAPTER NO	CONTENT	PAGE NO
	LIST OF FIGURES	3
	LIST OF TABLES	4
1.	DATA OVERVIEW	5
	1.1 About Data	5
	1.2 Data Description	5
	1.3 Data Checks	6
2.	UNIVARIATE ANALYSIS	11
	2.1 Analysis on Numerical Variables	11
	2.2 Analysis on Categorical Variables	13
3.	BIVARIATE ANALYSIS	15
	3.1 Relationship between Numerical Variables	16
	3.2 Relationship between Categorical Variables	17
	3.3 Relationship between Numerical and Categorical Variables	19
4.	KEY QUESTIONS	21
	4.1 Do men tend to prefer SUVs more compared to women?	21
	4.2 What is the likelihood of a salaried person buying a Sedan?	22
	4.3 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?	23

	4.4 How does the the amount spent on purchasing automobiles vary by gender?	24
	4.5 How much money was spent on purchasing automobiles by individuals who took a personal loan?	25
	4.6 How does having a working partner influence the purchase of higher-priced cars?	26
5.	INSIGHTS AND RECOMENDATIONS	27

LIST OF FIGURES

FIG NO.	NAME	PAGE
1.	Unique values	8
2.	Unique values of Gender	8
3.	Analysis on Age	11
4.	Analysis on Partner_salary	11
5.	Analysis on Price	12
6.	Analysis on Salary	12
7.	Analysis on Total_Salary	13
8.	Univariate Analysis on categorical columns 1	13
9.	Univariate Analysis on categorical columns 2	14
10.	Heatmap	15
11.	Age vs Price, Age vs Salary, Salary vs Price	16
12.	Bivariate analysis btw categorical 1	17
13.	Bivariate analysis btw categorical 2	18
14.	Numerical vs Categorical 1	19
15.	Numerical vs Categorical 2	20
16.	Make vs Gender	21
17.	Professioin vs make	22
18.	Profession vs Make vs Gender	23
19.	Gender vs Price	24
20.	Personal_loan vs Price	25
21.	Partner_working vs Price	26
22.	Profession vs Make vs Gender	27

LIST OF TABLES

TABLE NO.	NAME	PAGE
1.	First 10 rows	6
2.	Basic Information	7
3.	Null Value Summary	9
4.	Numerical Statistics	10
5.	Marital Status Count	27
6.	Gender Count	28
7.	Make Count	28

CHAPTER NO. 1

DATA OVERVIEW

1.1 ABOUT DATA:

Austo Motor Company is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models. The primary objective is to explore the given data, derive actionable insights, and recommend improvements to enhance the campaign's efficiency. This report aims to present a comprehensive analysis .

1.2 DATA DESCRIPTION:

Age	The age of the individual in years.
Gender	The gender of the individual, categorized as male or female.
Profession	The occupation or profession of the individual.
Marital_status	The marital status of the individual, such as married &, single
Education	The educational qualification of the individual Graduate and Post Graduate
No_of_Dependents	The number of dependents (e.g., children, elderly parents) that the individual supports financially.
Personal_loan	A binary variable indicating whether the individual has taken a personal loan "Yes" or "No"
House_loan	A binary variable indicating whether the individual has taken a housing loan "Yes" or "No"

Partner_working	A binary variable indicating whether the individual's partner is employed "Yes" or "No"
Salary	The individual's salary or income
Partner_salary	The salary or income of the individual's partner, if applicable
Total_salary	The total combined salary of the individual and their partner (if applicable)
Price	The price of a product or service
Make	The type of automobile

1.3 DATA CHECKS

Preview of the Dataset:

Table 1: First 10 rows

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	170000	61000	SUV
53	Femal	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500	70300	165800	61000	SUV
53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300	60700	158000	57000	SUV
53	Female	Salaried	Married	Graduate	2	Yes	No	Yes	72500	70300	142800	61000	SUV
53	Male	Salaried	Married	Post Graduate	3	No	No	Yes	79700	60200	139900	57000	SUV
53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	79800	60500	140300	57000	SUV
53	Female	Salaried	Married	Post Graduate	3	Yes	No	Yes	80900	50800	131700	51000	SUV
52	Male	Salaried	Married	Graduate	4	No	No	Yes	79100	40400	119500	68000	SUV
52	Male	Salaried	Married	Post Graduate	4	No	No	No	89400	0	89400	61000	SUV
52	Male	Salaried	Married	Post Graduate	3	No	No	No	83700	0	83700	57000	SUV

Shape:

1. There are 1581 rows and 14 columns in this dataset.
2. There are no duplicates in this dataset.

Basic Information of the dataset:

Table 2: Basic Information

```

<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: 173.1+ KB

```

- In “Gender” and “Partner_salary” columns, there are some null values to be treated. The Datatype of “Partner_salary” column should be corrected.
- There are 6 numerical variables and 8 categorcial variables

Numerical Variables

1. Age
2. No_of_Dependents
3. Salary
4. Partner_salary
5. Total_salary
6. Price

Categorical Variables

1. Gender
2. Profession
3. Marital_Status
4. Education
5. Personal_loan
6. House_loan
7. Partner_working
8. Make

Data Corrections:

Fig.1 Unique Values

```
Male      1199
Female    327
Femal     1
Femle     1
Name: Gender, dtype: int64
-----
Salaried   896
Business   685
Name: Profession, dtype: int64
-----
Married    1443
Single     138
Name: Marital_status, dtype: int64
-----
Post Graduate   985
Graduate        596
Name: Education, dtype: int64
-----
Yes      792
No       789
Name: Personal_loan, dtype: int64
-----
No      1054
Yes      527
Name: House_loan, dtype: int64
-----
Yes      868
No       713
Name: Partner_working, dtype: int64
-----
Sedan      702
Hatchback  582
SUV        297
Name: Make, dtype: int64
-----
```

- In “Gender” column, the word Female is spelled incorrectly 2 times as ‘Femal’ and ‘Femle’ and it is corrected.
- After correction,

Fig.2 : Unique values of Gender

```
Male      1199
Female    329
Name: Gender, dtype: int64
```

Null Value Treatments:

Table 3: Null Value Summary

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

- There are 53 null values in the “Gender” column and 106 null values in “Partner_salary”.
- For a categorical column, the null values are treated with the mode of the column. The mode of “Gender” column is “Male” which appears in 1199 rows.
- From the Data Dictionary, we can clearly observe that the “Total_salary” column is the sum of “Salary” and “Partner_Salary”.
- “Total_salary” = “Salary” + “Partner_salary”
- So,
“Partner_salary” = “Total_salary” – “Salary”

Numerical Statistics of data:

Table 4: Numerical Statistics

	count	mean	std	min	25%	50%	75%	max
Age	1581.0	31.922201	8.425978	22.0	25.0	29.0	38.0	54.0
No_of_Dependents	1581.0	2.457938	0.943483	0.0	2.0	2.0	3.0	4.0
Salary	1581.0	60392.220114	14674.825044	30000.0	51900.0	59500.0	71800.0	99300.0
Partner_salary	1581.0	19233.776091	19670.391171	0.0	0.0	25100.0	38100.0	80500.0
Total_salary	1581.0	79625.996205	25545.857768	30000.0	60500.0	78000.0	95900.0	171000.0
Price	1581.0	35597.722960	13633.636545	18000.0	25000.0	31000.0	47000.0	70000.0

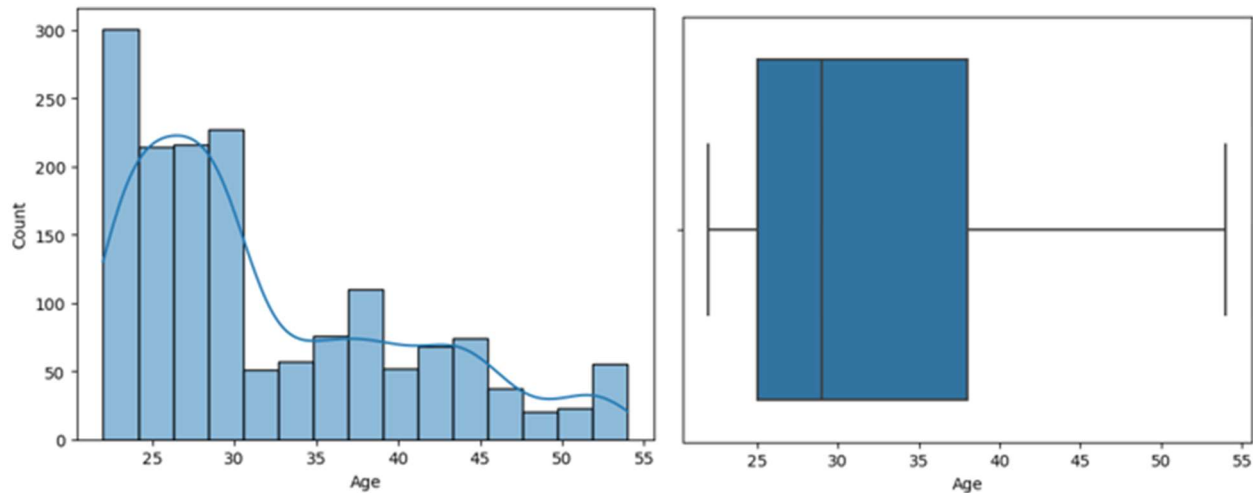
- The Customer's age ranges between 22 and 54 years and the average age of the customer is approximately 32 years.
- The Average No.of.Dependents is 2.
- The Price of the vehicle starts from 18k and ranges upto 70k with an average of 35.5k.
- The Average and the Median of the Salary is approximately same.
- The Total Salary of the customers ranges from 30k to 171k.

CHAPTER NO. 2

UNIVARIATE ANALYSIS

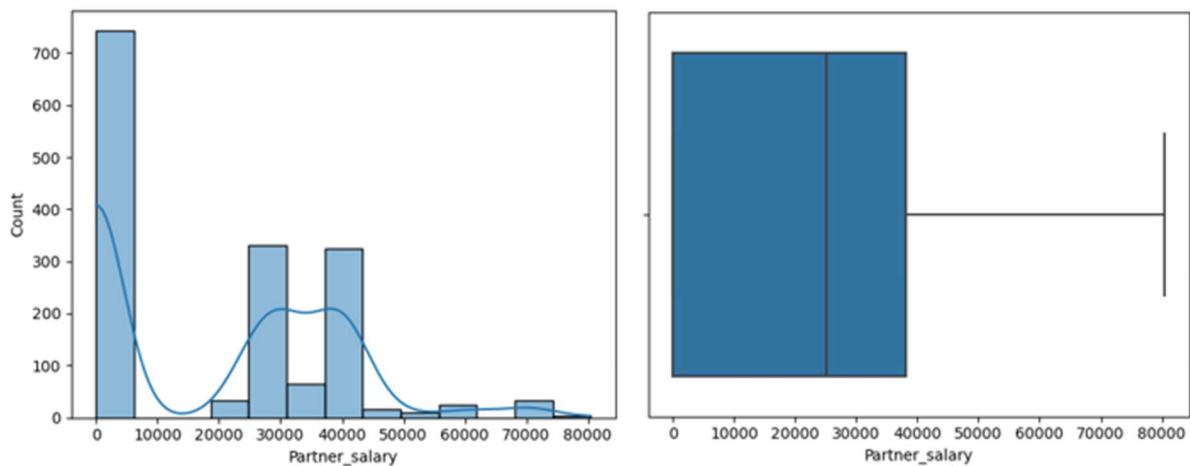
2.1 ANALYSIS ON NUMERICAL VARIABLES

Fig.3 Analysis on Age



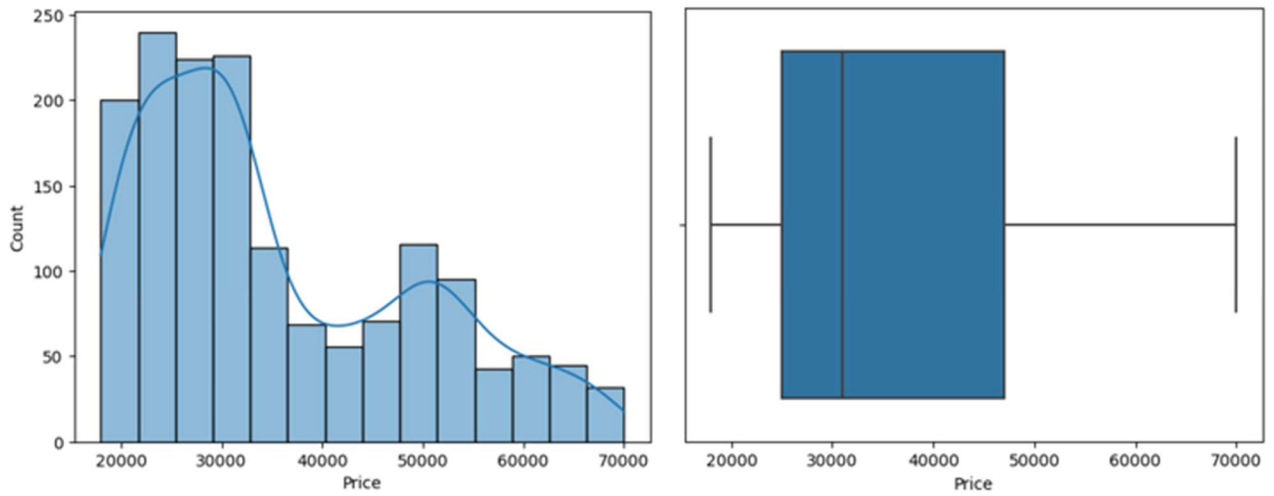
- The Distribution is skewed towards right.

Fig.4 Analysis on Partner_Salary



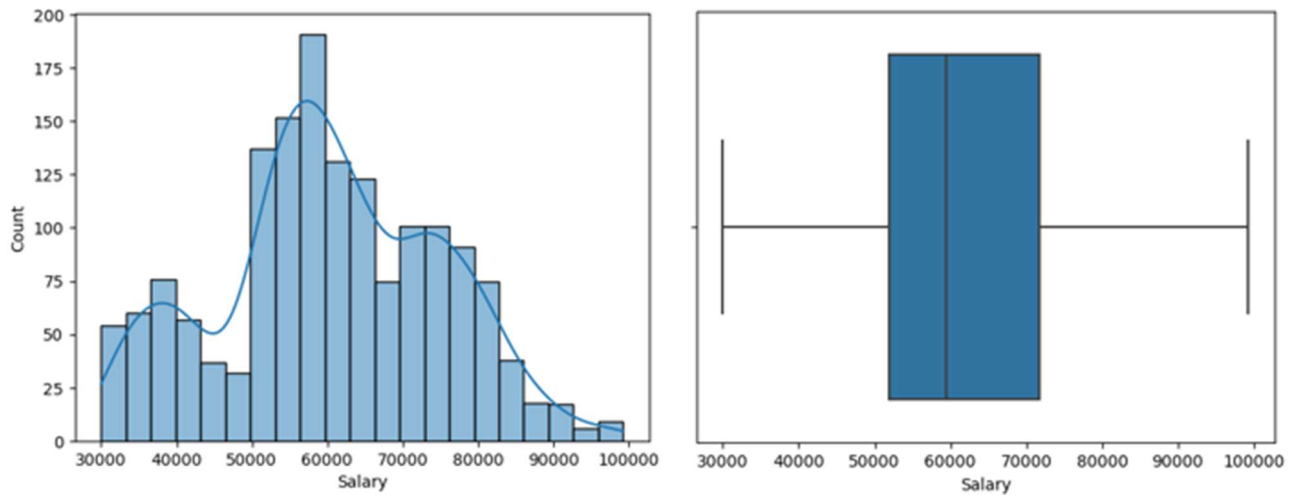
- The Distribution is skewed towards right.
- 25% of the Customers does not have Partner_Salary

Fig.5 Analysis on Price



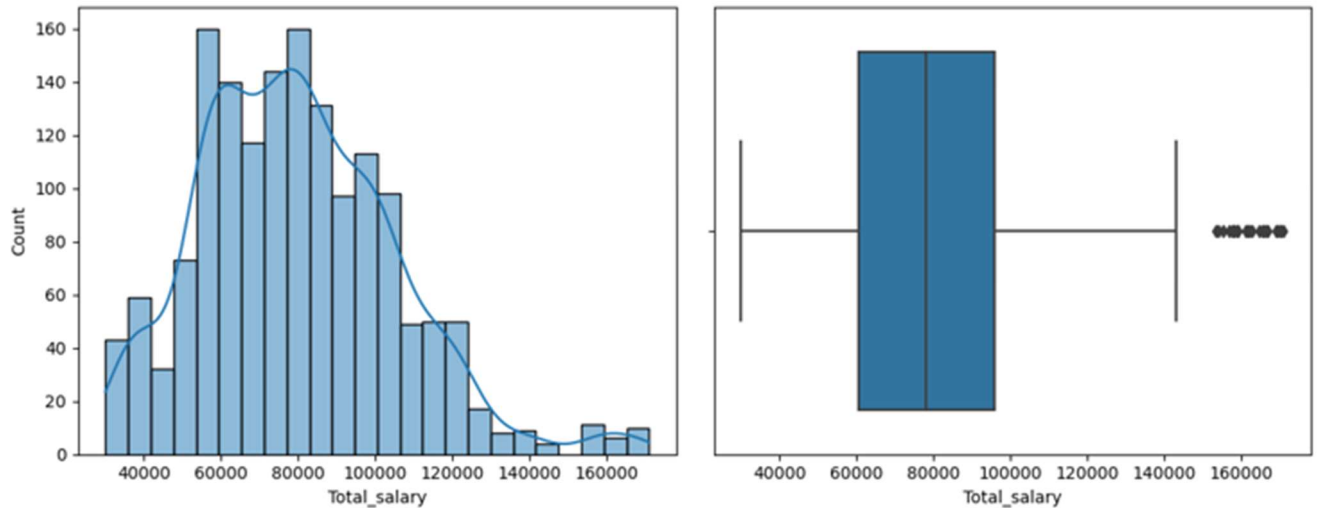
- It has a Bimodal Distribution
- The Distribution is skewed towards right.

Fig.6 Analysis on Salary



- The Distribution is slightly skewed towards right.
- The most common salary range appears to be around 60k.

Fig.7 Analysis on Total_Salary



- There are some outliers present in the Total_salary plot.
- In Business and Financial fields, Prices and Salary can be higher and cannot be treated as outliers.
- If we treat outliers, we might lose some information on trends and accuracy.

2.2 ANALYSIS ON CATEGORICAL VARIABLES

Fig.8 Univariate Analysis on categorical columns 1

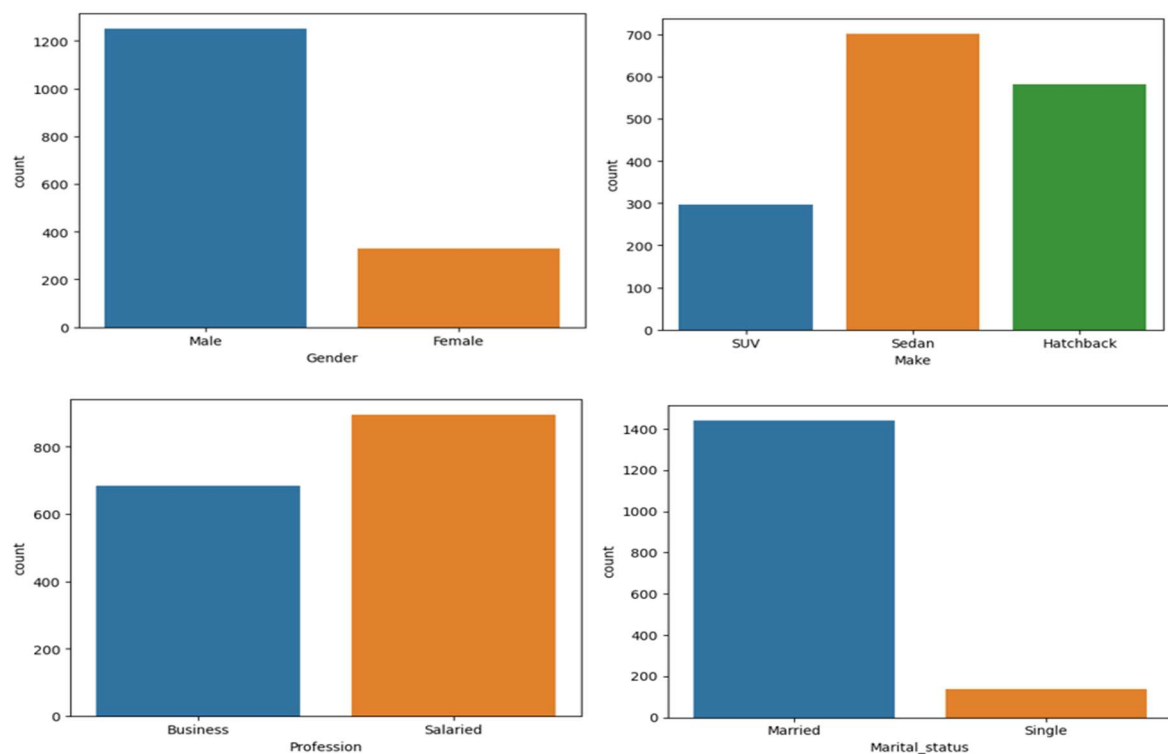
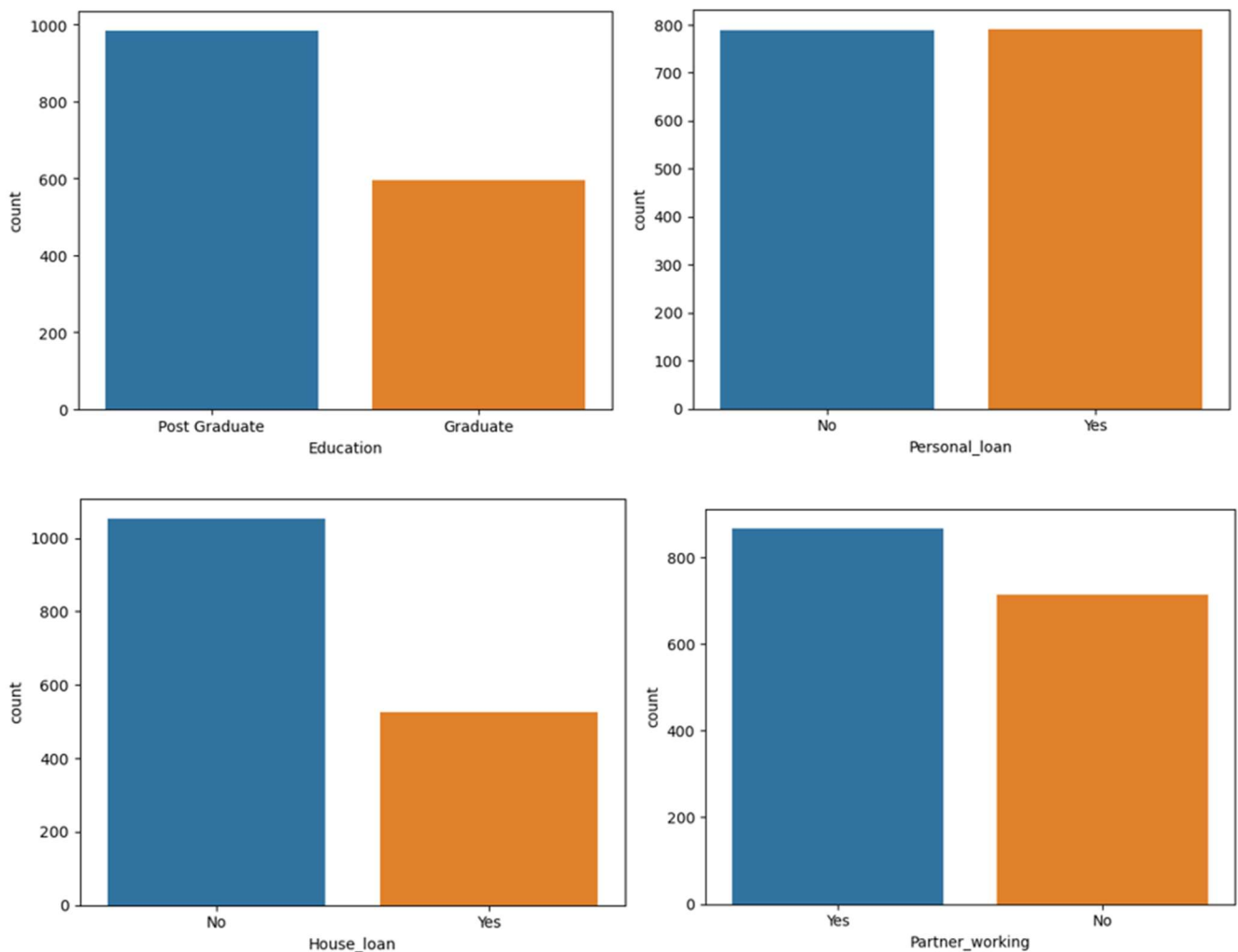


Fig. 9 Univariate on categorical columns 2



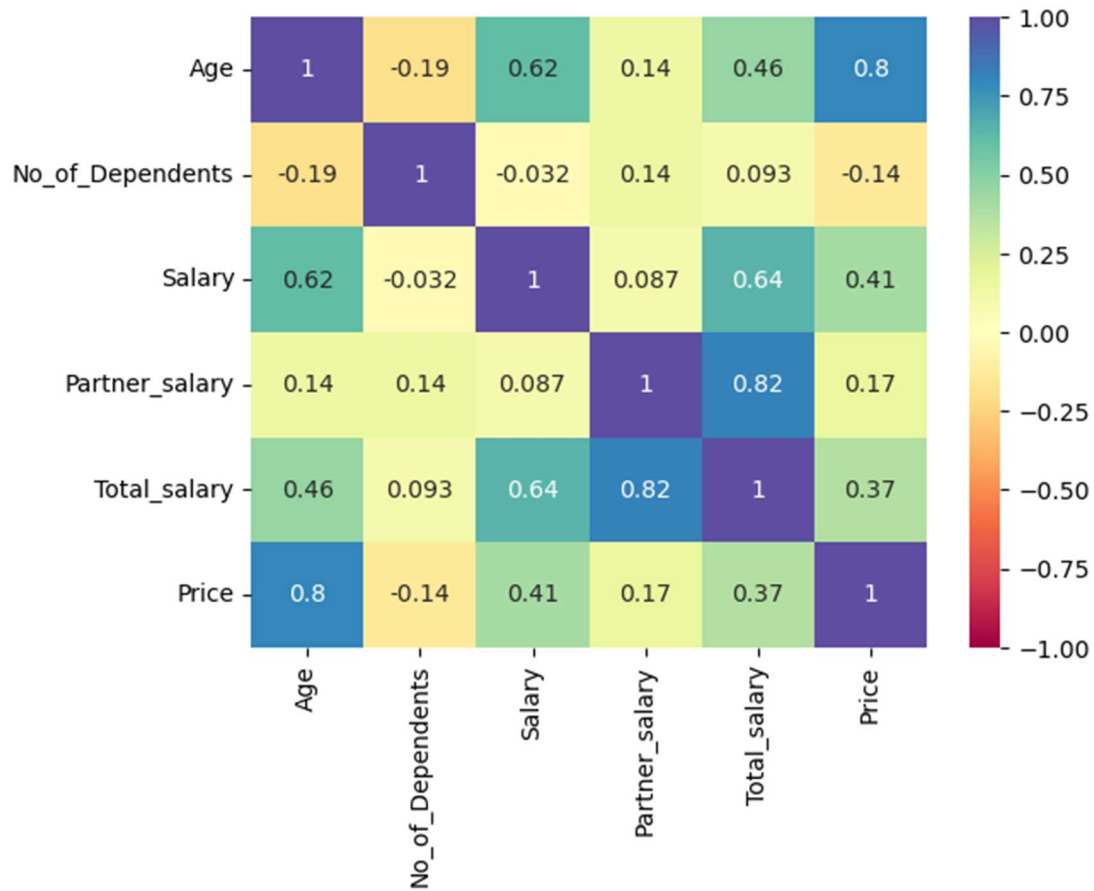
Observations:

1. Majority of the customers are Male.
2. The Sedan make is the most purchased type of automobile after Hatchback and SUV.
3. Salaried individuals are higher than the Business individuals.
4. 91% of the clients are married, compared to just eight percent who are single. This suggests that married people make up the majority.
5. The proportion of consumers who have taken out personal loans is roughly equal to that of those who have not.
6. Majority of customers do not have house loans.

CHAPTER NO. 3

BIVARIATE ANALYSIS

Fig 10. Heatmap

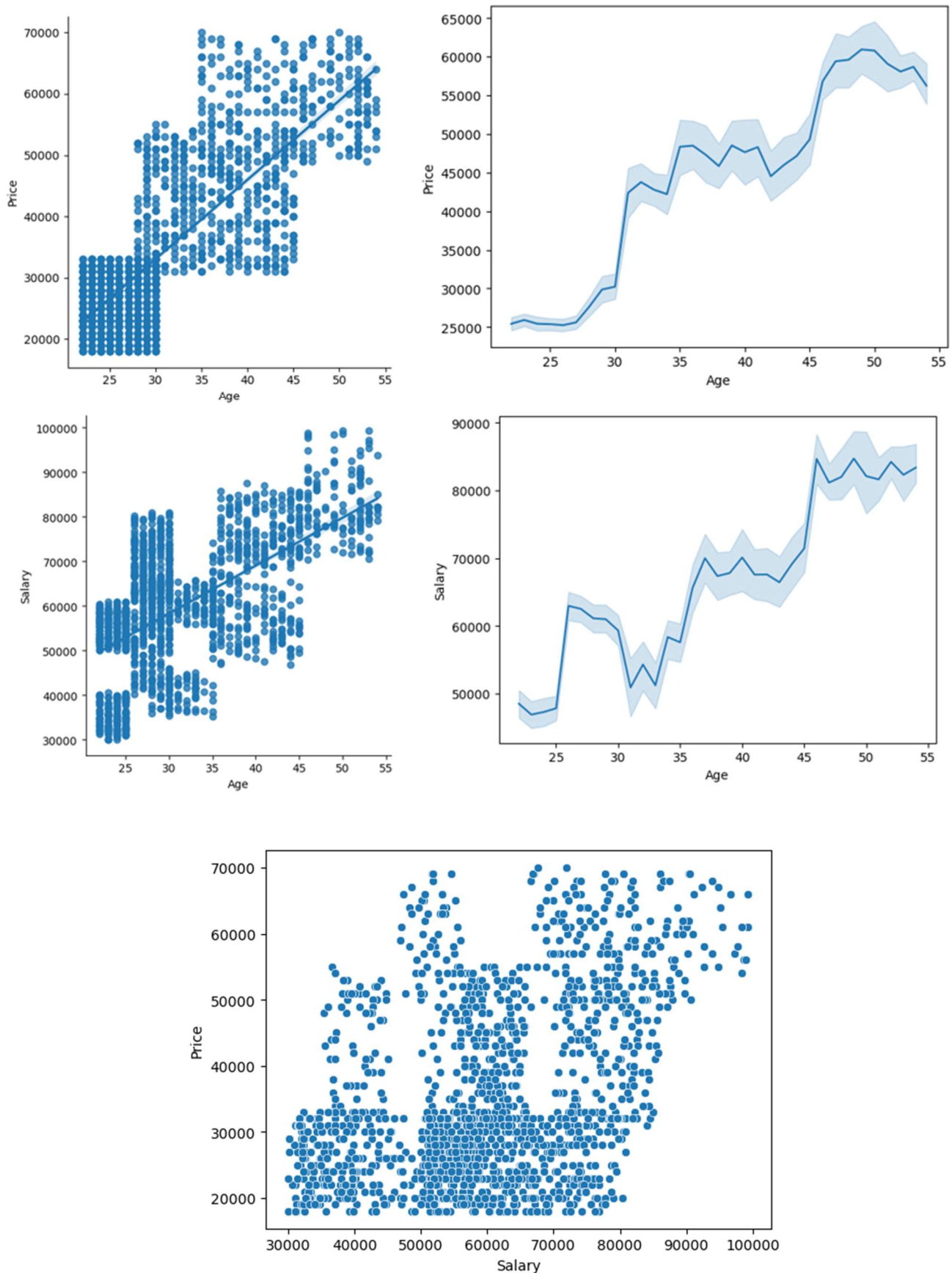


Observations:

- Age column shows a high correlation with Price column which indicates that as age increases, price of the automobile increases.
- A positive correlation showing that as salary increases, the total salary also increases.
- Salary column shows a positive correlation with Age column.
- A weaker correlation exists between Price and No.of.Dependents.
- A weaker correlation exists between Age and No.of.Dependents.
- A weaker correlation exists between Salary and No.of.Dependents

3.1 Relationships Between Numerical (Continuous) variables:

Fig.11 Age vs Price, Age vs Salary, Salary vs Price



Observations:

- Age vs Price: A positive correlation is observed between age and price which indicates that if amount spent increases as age increases.
- Age vs Salary: A weak positive correlation is observed between age and salary which indicates that salary increases with age.
- Salary vs Price: The scatterplot shows the relationship between salary and price, with data points spread across the entire range. A weak positive correlation is observed.

3.2 Relationships Between Categorical variables:

Fig.12 Bivariate analysis btw categorical 1

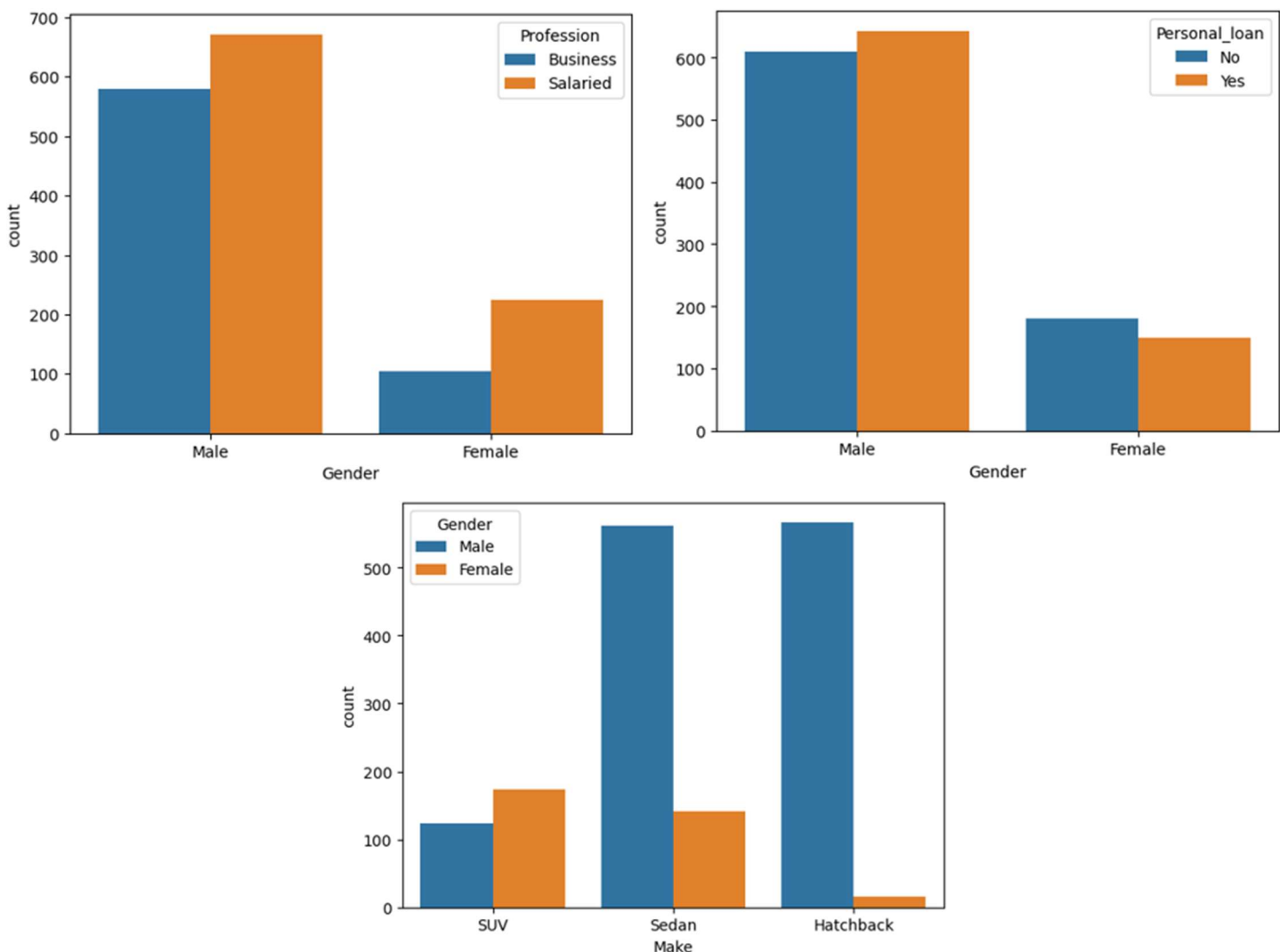
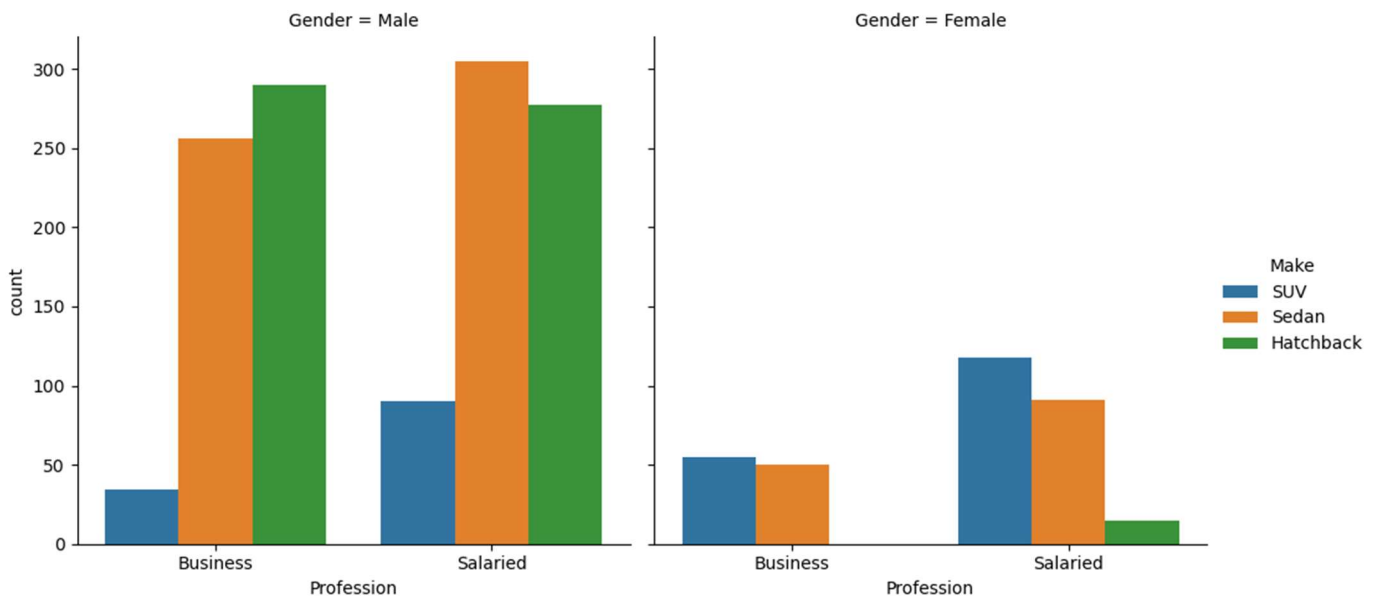


Fig.13 Bivariate analysis btw categorical 2



Observations:

- Salaried men purchased more number of automobiles than Business Men.
- Salaried women purchased more number of automobiles than Business women.
- Men purchased more number of automobiles than Business Men.
- The number of males who have not taken a personal loan is lower than the number of males who have taken personal loans.
- Women prefer SUV make more than Men.
- Men prefer Sedan make more than women.
- Low number of women prefer Hatchback make compared to other types.
- Salaried male prefer Sedan make whereas Business Men prefer Hatchback.
- Business Women does not prefer Hatchback Make.
- Salaried women prefer SUV make.

3.3 Relationships Between Numerical and Categorical variables:

Fig.14 Numerical vs Categorical 1

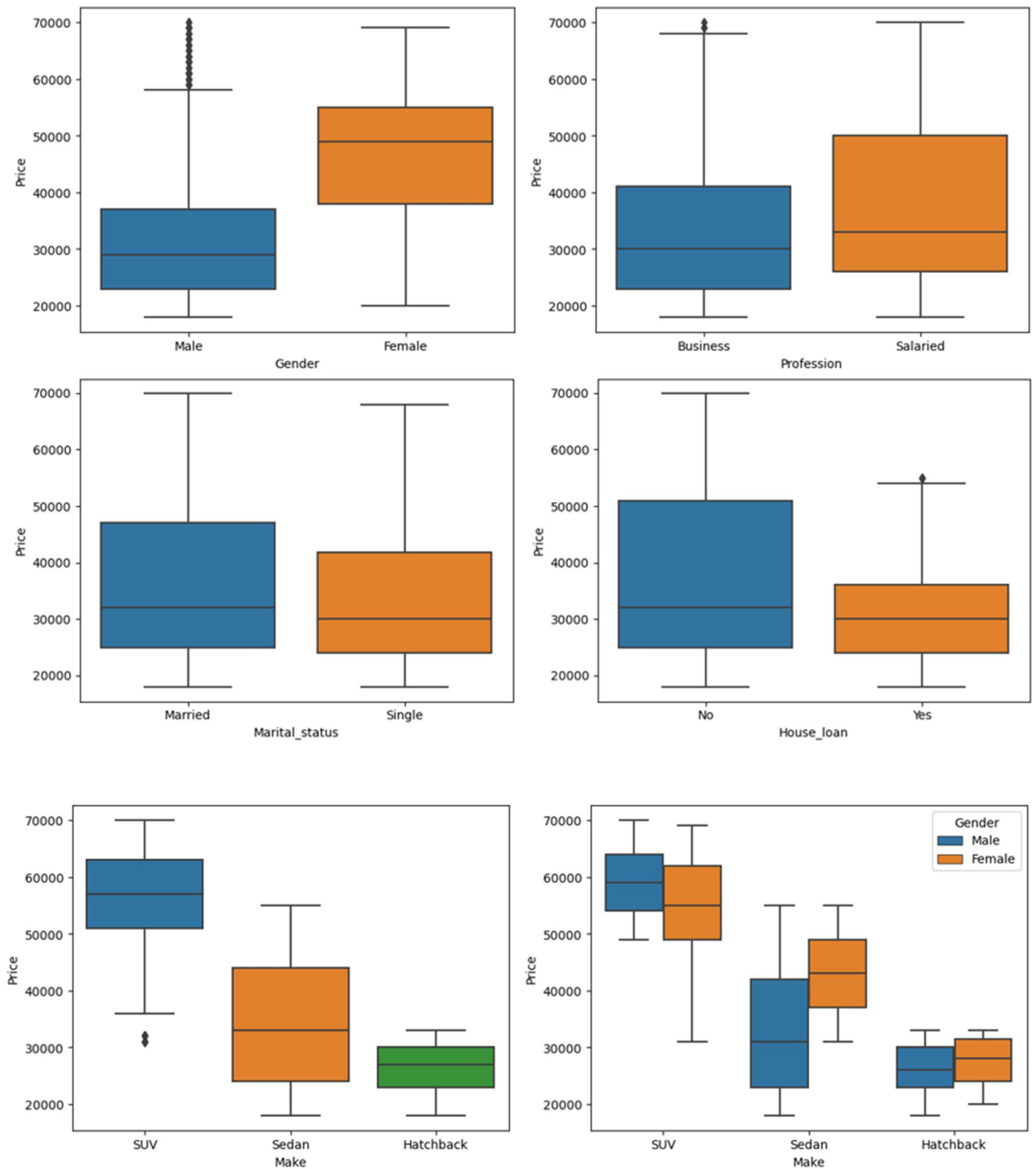
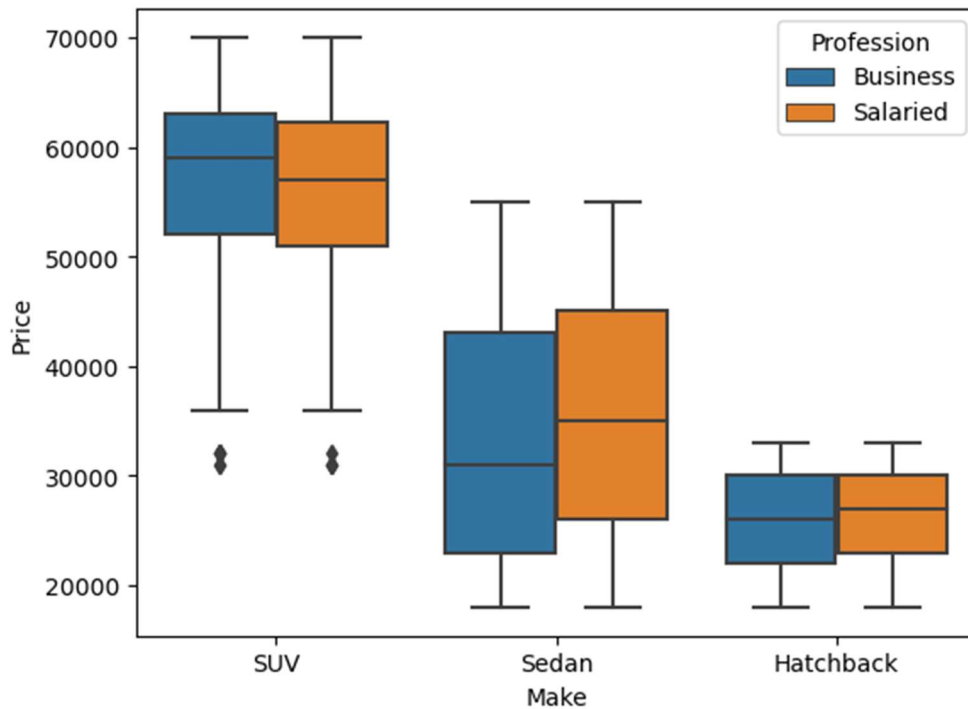


Fig.15 Numerical vs Categorical 2



Observations:

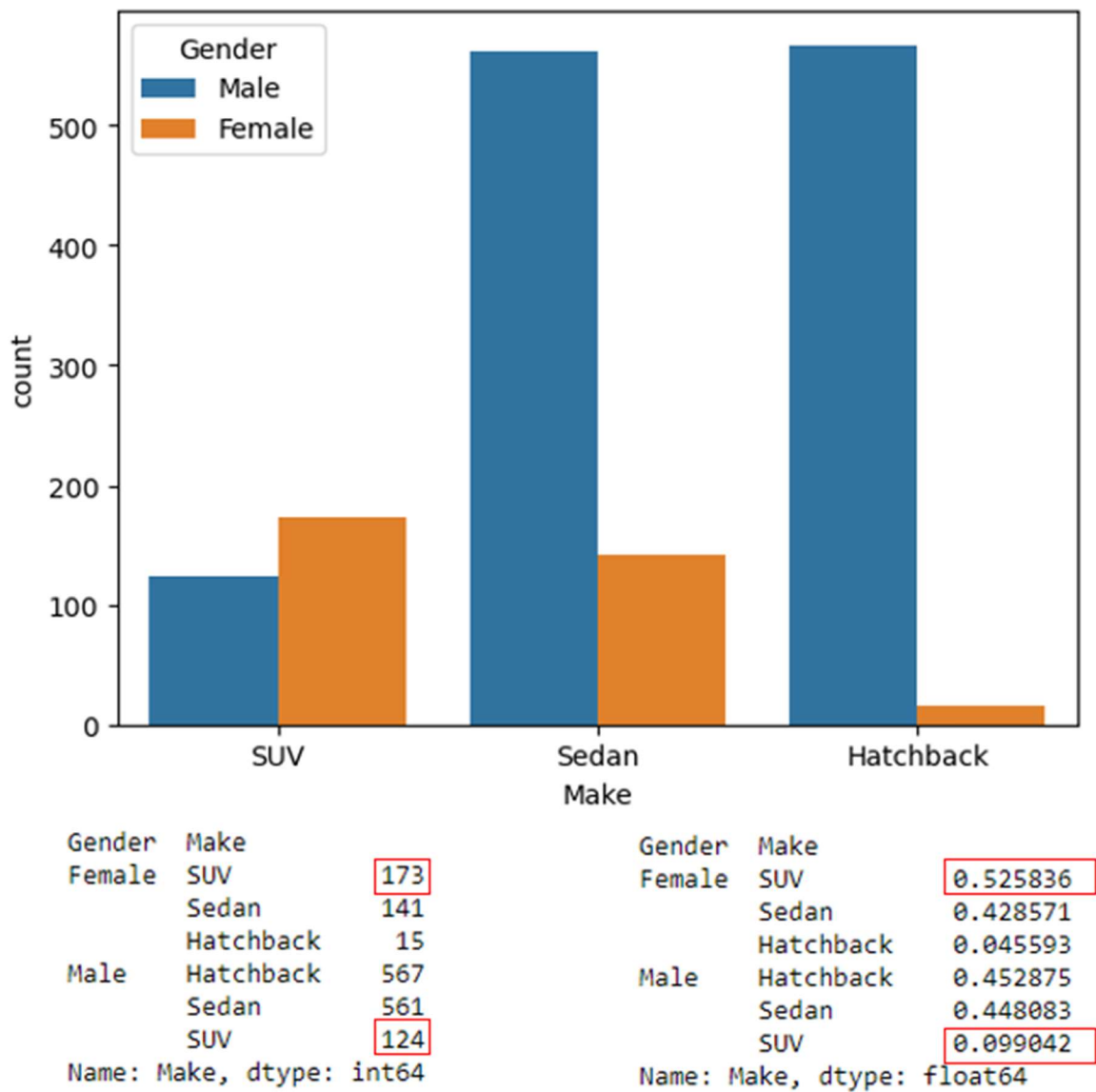
- Despite being fewer in number, women spend significantly more on purchasing automobiles than men.
- Salaried people spend higher than the business people in purchasing automobiles.
- People without house loans spend higher in purchasing than the people with house loans.
- SUV make has the higher price range followed by Sedan and Hatchback make.
- The median price for Sedan and Hatchback is lower in the Business category than in the Salary category.
- Sedan show greater variability in prices for both profession category, indicating a wider range of prices.
- Hatchback has the lowest price range in all categories.

CHAPTER NO. 4

KEY QUESTIONS

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

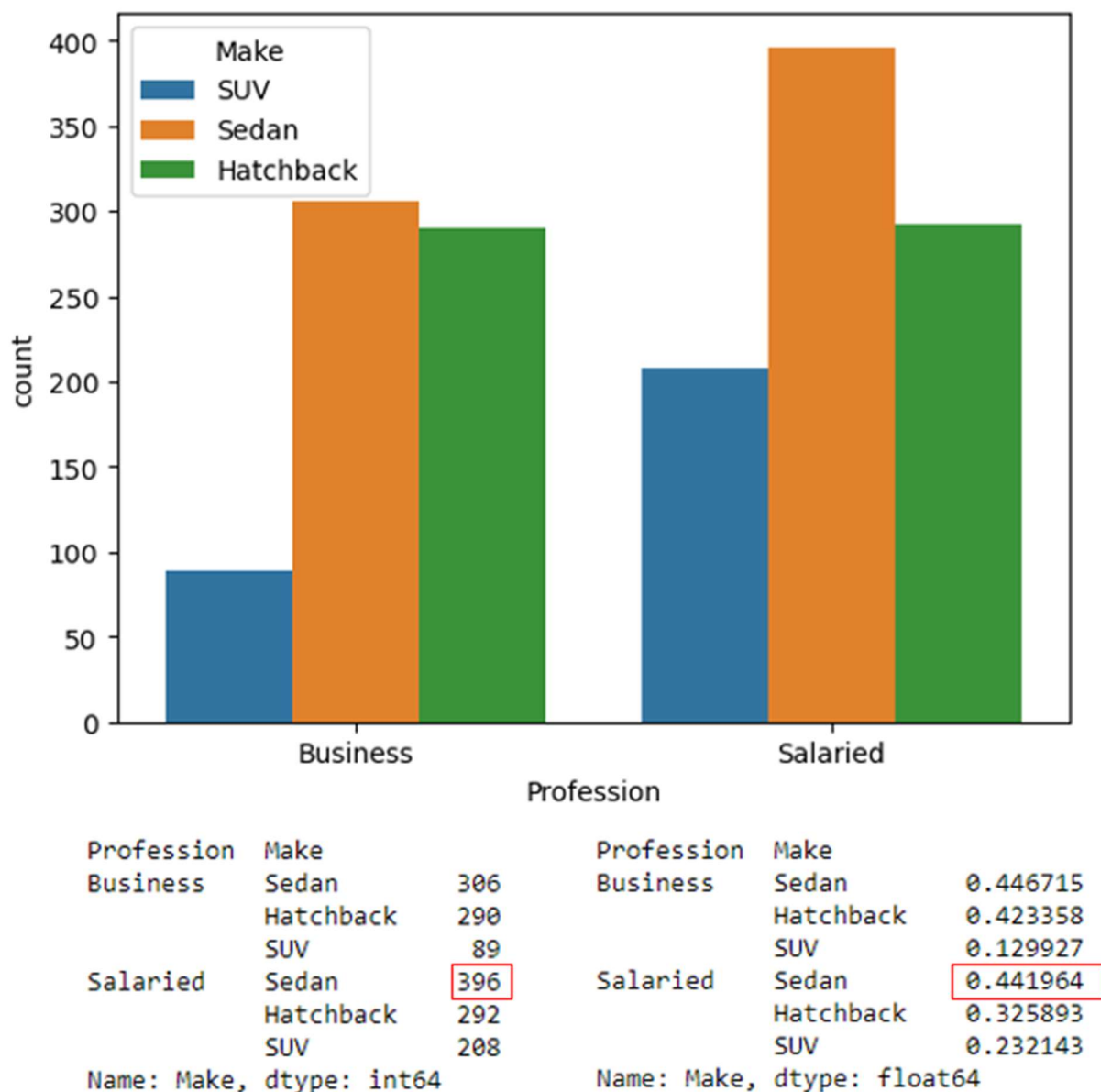
Fig.16 Make vs Gender



- From the tables and plots, we can clearly observe that, 52% of the women prefer SUV make, whereas only 9% of the men prefer SUV.
- Hatchback was the most preferred automobile type of men with 45%.
- SUV make was mostly preferred by women.

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

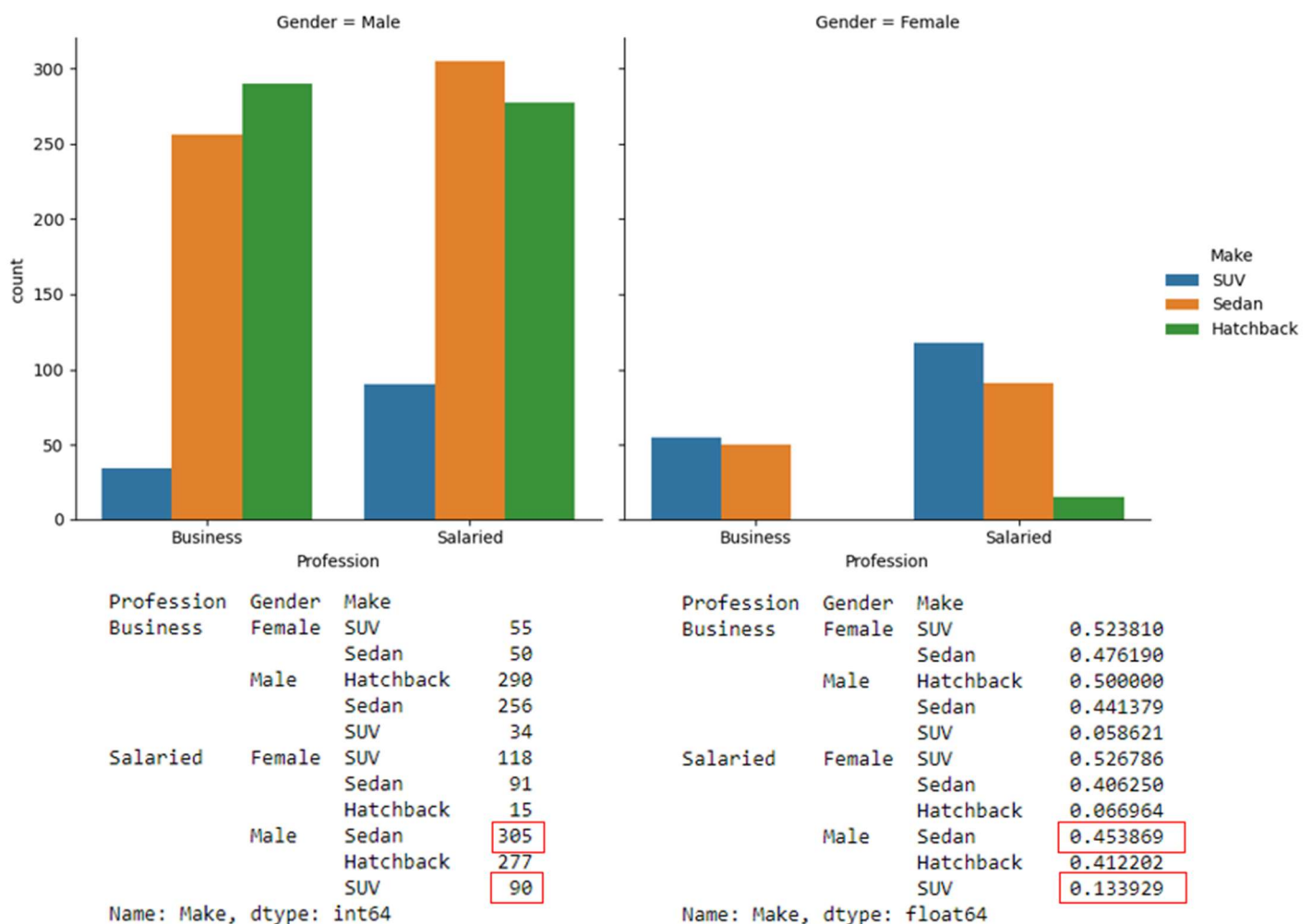
Fig.17 Profession vs Make



- By analysing the data, we get:
 - 44% of the Salaried people prefer Sedan make,
 - 32% of the Salaried people prefer Hatchback make,
 - 23% of the salaried people prefer SUV make.
- 396 out of 896 salaried people prefer Sedan.
- From the above plot, it is clearly evident that majority of the Salaried people prefer Sedan make.

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?

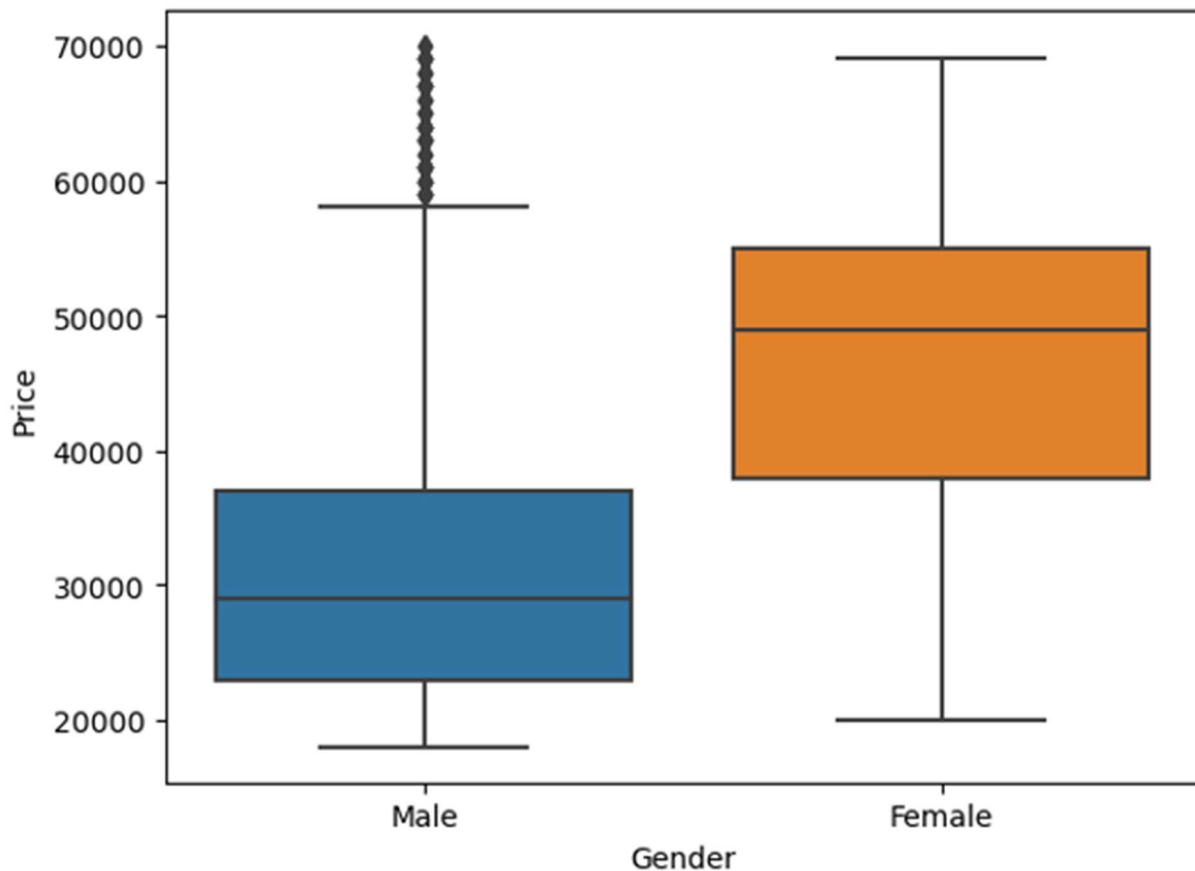
Fig.18 Profession vs Make vs Gender



- By Analysing the data, we get:
 - 45% of the Salaried men prefer Sedan make,
 - 41% of the Salaried men prefer Hatchback make,
 - 13% of the Salaried men prefer SUV make.
- 305 salaried men prefer Sedan make, whereas only 90 men prefer SUV make.
- From the above plot, it is clearly evident that salaried male is an easier target for Sedan sale over SUV sale.
- Hence, Mr. Sheldon Cooper's claim is incorrect.

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

Fig.19 Gender vs Price

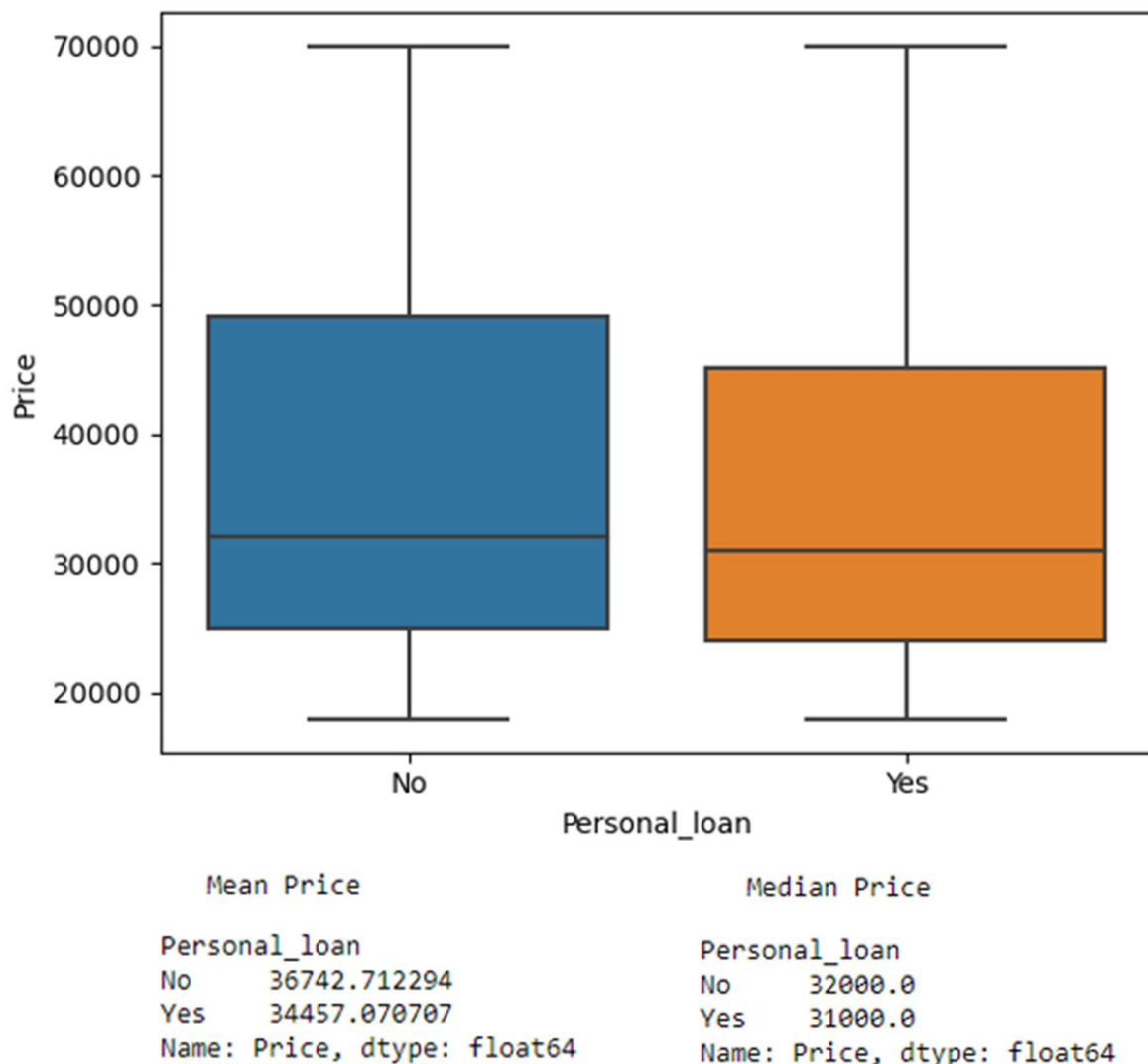


		Price		Price	
		Gender	Mean	Gender	Median
Male	1252	Female	47705.167173	Female	49000.0
Female	329	Male	32416.134185	Male	29000.0
Name: Gender, dtype: int64					

- There are 1252 men and 329 women.
- Since majority of the persons are Men, we cannot compare with the sum of the amount spent.
- The average amount spent by Women is greater than of Men,
 - Avg. price spent by Women = 47k (approx.)
 - Avg. price spent by Men = 32k (approx.)
- Despite being fewer in number, women spend significantly more on purchasing automobiles than men, with an average expenditure of approximately 47k compared to 32k for men.

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

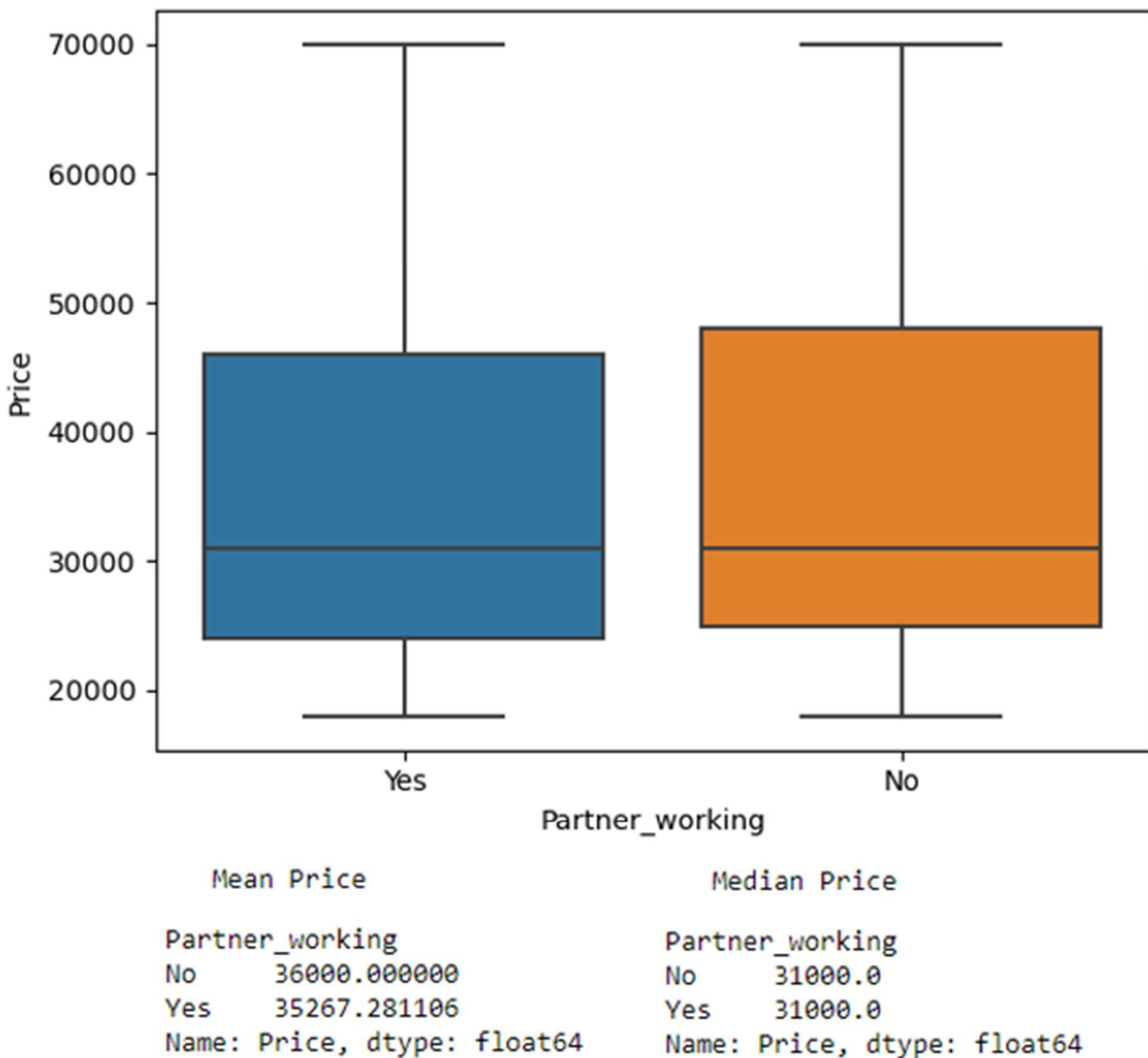
Fig.20 Personal_loan vs Price



- Avg. price spent by customers without personal loan = 36k (approx.)
- Avg. price spent by customers with personal loan = 34k (approx.)
- The Average and the Median price spent on purchasing by customers without personal loans are greater than the customers with personal loans.
- To conclude, Customers without personal loans tend to spend more on purchasing, with an average expenditure of approximately 36k compared to 34k for those with personal loans.

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

Fig.21 Partner_working vs Price



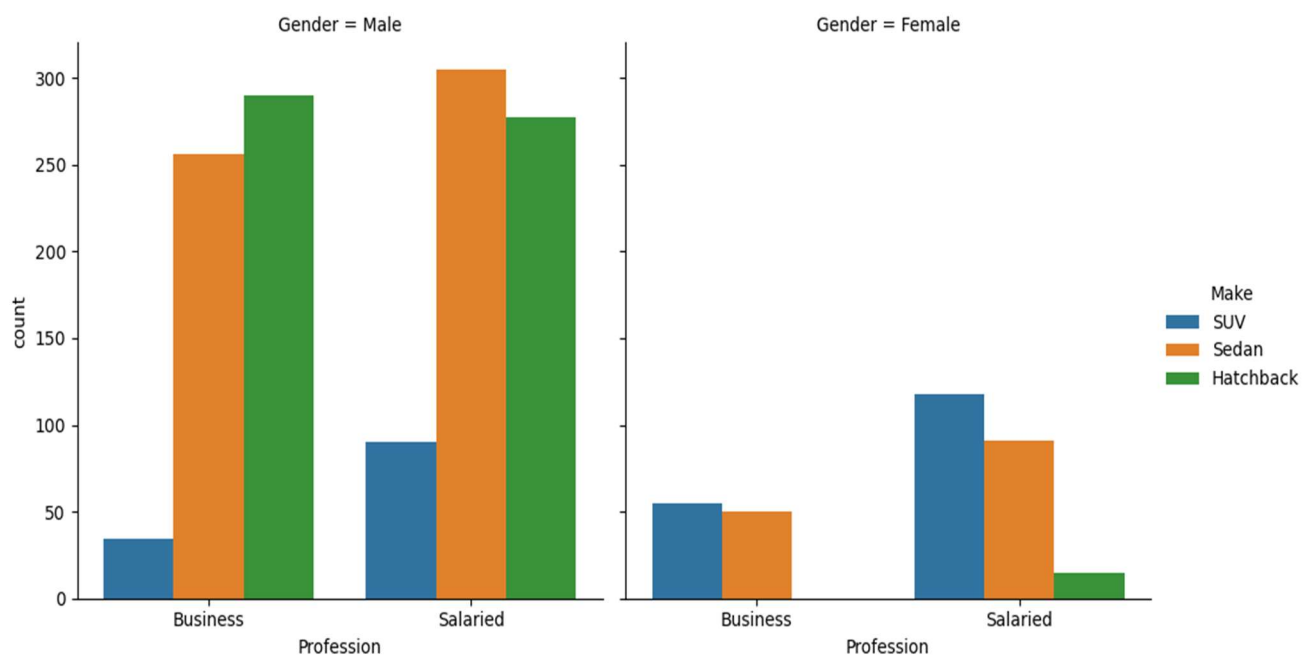
- Avg. price spent by customers with Working partner: 35k (approx.)
- Avg. price spent by customers without Working partner: 36k
- The Mean and the Median price spent by customers having working partner is similar to the customers without having working partner.
- So, having a working partner doesn't influence the purchase of higher-priced cars

CHAPTER NO. 5

ACTIONABLE INSIGHTS AND RECOMMENDATIONS

1. Business Women had not purchased any Hatchback make. We can target them and employ business strategies to improve Hatchback sales.

Fig.22 Profession vs Make vs Gender



2. Since the majority of the individuals are married, targeting them will improve the business.

Table 5: Marital Status Count

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

3. Individuals whose marital status is Single should be targeted and explore why the segment is lower and attract them.

4. Since the count of Female Customers is less, special attention to given on marketing more expensive cars to female consumers by emphasizing features and perks that are appealing to them.

Table 6: Gender Count

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

5. Employ Marketing strategies to improve the sales of SUV make.

Table 7: Make Count

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