Python for Data Science
Business Report

❖ Analysis of car sales for Austo Automobiles

Submitted by

ADDAGUDURU SATVIK

Contents

Problem

Austo Motor Company is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models. In its recent board meeting, concerns were raised by the members on the efficiency of the marketing campaign currently being used. The board decides to rope in analytics professional to improve the existing campaign.

Objective

They want to analyse the data to get a fair idea about the demand of customers which will help them in enhancing their customer experience. Suppose you are a Data Scientist at the company and the Data Science team has shared some of the key questions that need to be answered. Perform the data analysis to find answers to these questions that will help the company to improve the business.

Problem 1: Data Overview

1.1 Import the libraries

We have imported the required libraries from python library, NumPy library for mathematical and scientific computing, Pandas library for analysing, cleaning, exploring, and manipulating data and Matplotlib to create graphs and visualizations and seaborn libraries to create statistical graphics and perform data visualization.

1.2 Load the data

Using Pandas function read_csv we have loaded the austo_automobile1.csv file

1.3 Structure of the data

Using shape method, the number of rows and columns in a Data Frame is accessed and found there are 1581 rows and 14 columns

1.4 Types of data

We have 3 types of datatypes in this dataset, object is for categorical variables and int64 for integer values and float64 for decimal values and We have 8 categorical columns and 6 numerical columns

1.5 Treating missing values

Missing values affects analysis of business so identifying and treating them becomes very important. In this Data set, we have missing values in the columns.

Gender - 53

Partner Salary -106

We have treated these missing values and have replaced missing values in Gender with Mode and Partner salary with mean.

1.6 Statistical summary.

I have used the describe method in Pandas to generate a statistical summary of this Dataset. It has helped in understanding central tendency, dispersion, and shape of the distribution of numerical features.

	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	1475.0	20225.559322	19573.149277	0.0	0.0	25600.0	38300.0	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

*It is observed that minimum Total salary is $30000.0 and maximum Total salary $171000 which is 4.71 times the minimum Total salary

*Most of the buyers are from the Age group of 29 to 38

*Average price of the car is $35597.72 and maximum price of the car is $70000 which is almost twice the average price of the car

*Observed minimum No_of_Dependents is 0 and maximum No_of_Dependents is 4

1.7 Check and treat data irregularities

Upon reviewing and analysing the data, I have noticed there is some spelling mistakes in Gender column. I have corrected the same using Replace function. Observed there is no duplicate values in the dataset.

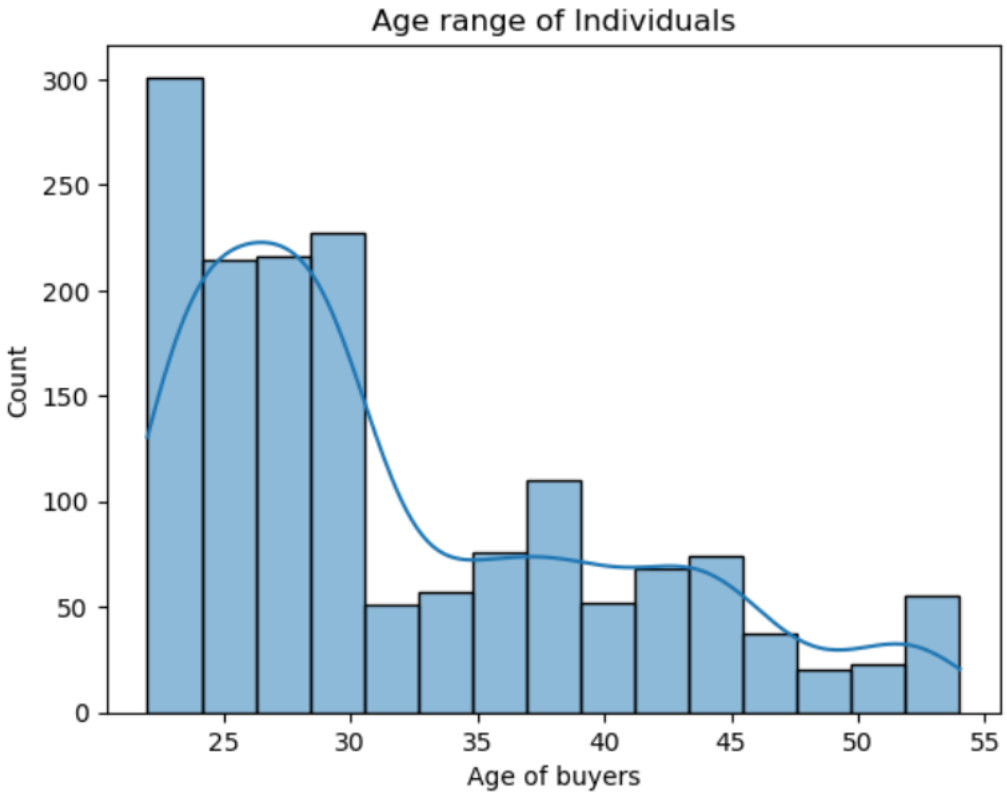
1.8 Observations and Insights

The available Data set looks good and sufficient for the analysing the demand of the customers. Looking into this dataset can provide a good amount of information for increasing sales and customer experience. Data had few anomalies like missing records and multiple texts which has been treated. Data also has inconsistencies like outliers which must be treated.

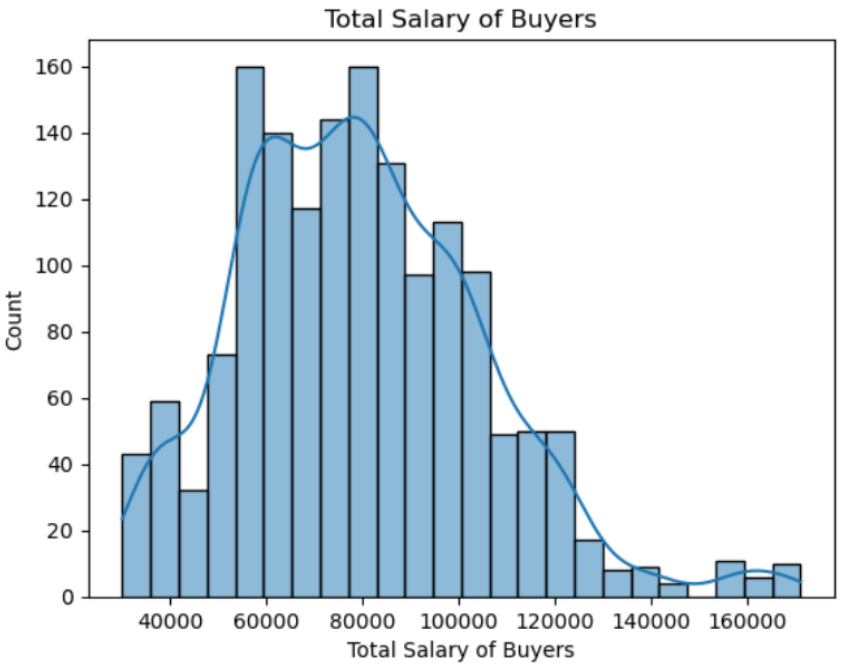
Problem 2: Univariate Analysis

2.1 Exploring all the variables (categorical and numerical) in the data

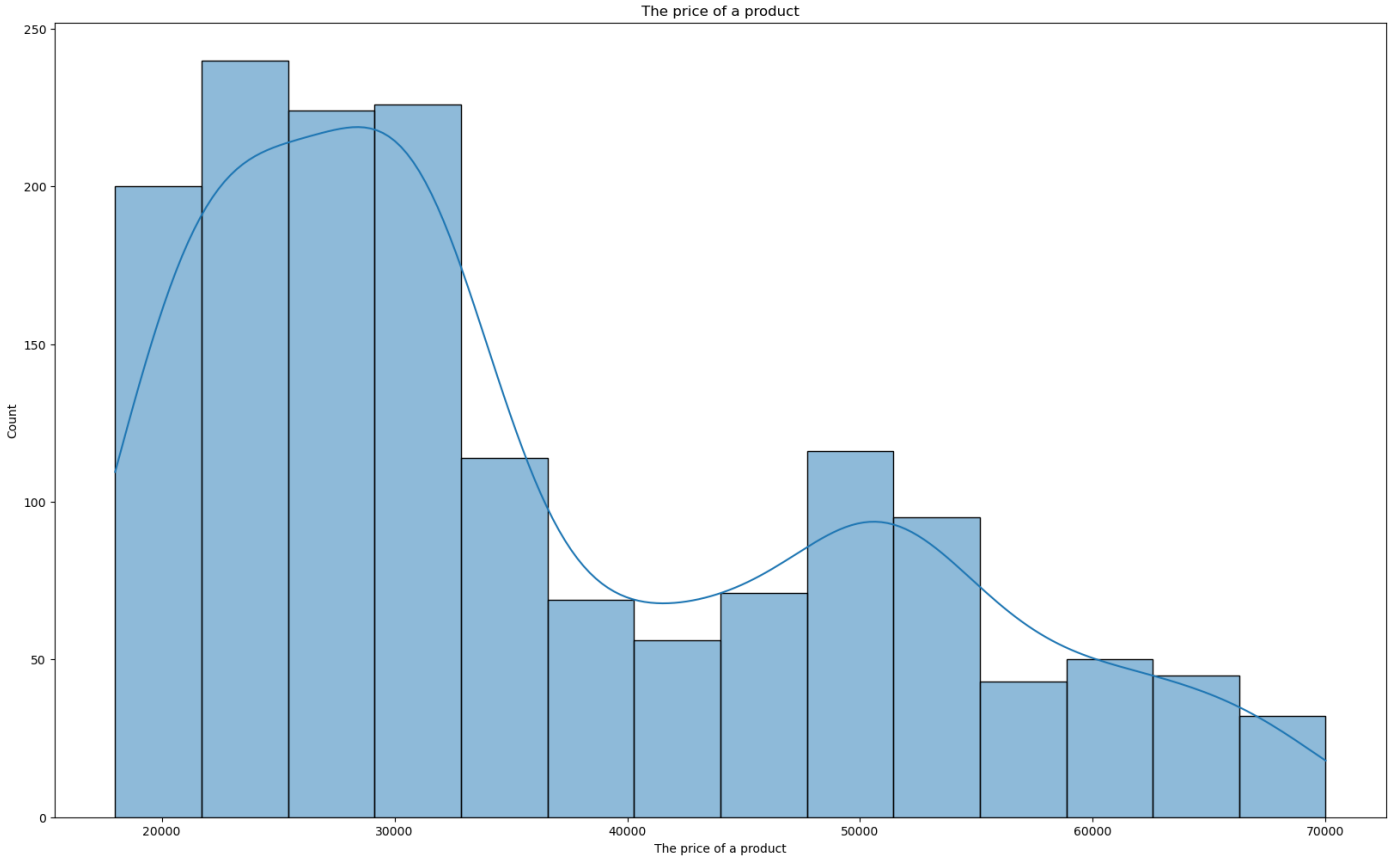
- We have the following categorical columns in the data set. Gender, Profession, Marital Status ,Education ,Personal loan ,House loan, Partner working ,Make.
- We have the following numerical columns in the data set. 'Age','No_of_Dependents','Salary','Partner_salary','Total_salary','Price'



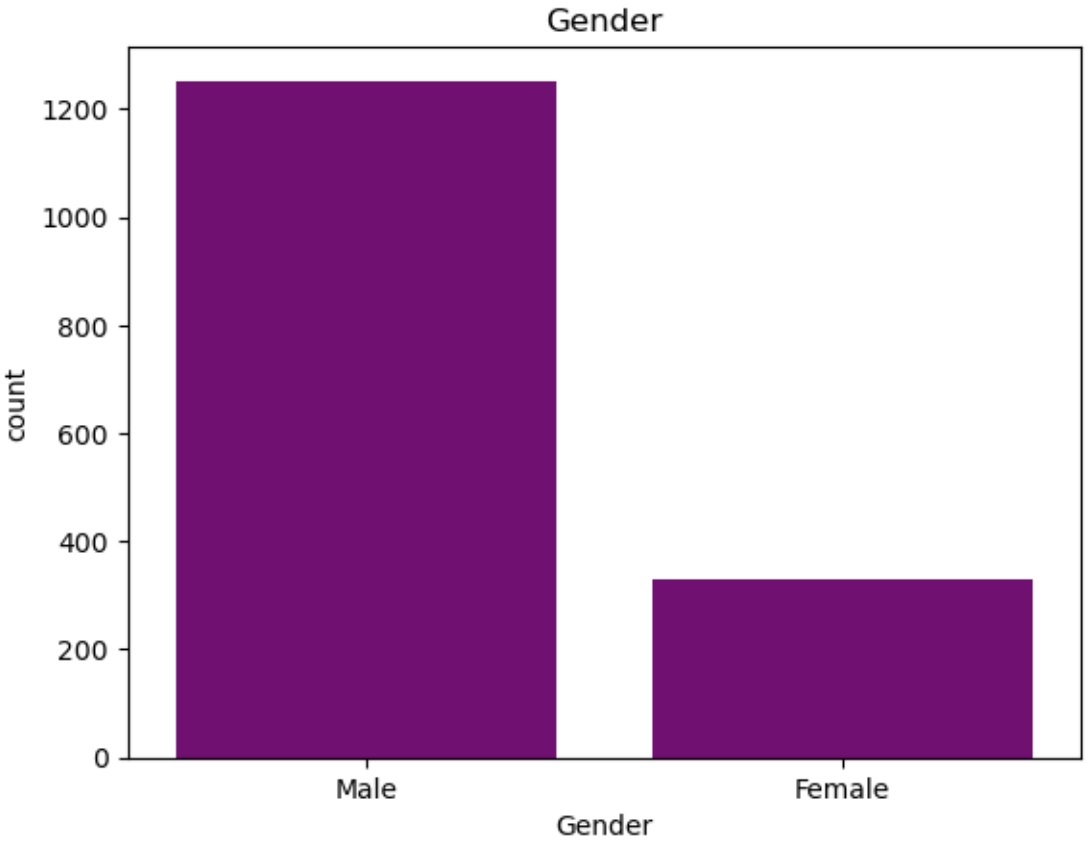
- Most of the buyers are from the Age group of 29 to 38



- Most of the people have salary in between $50000 to $85000
- Only few People have salary more than $130000

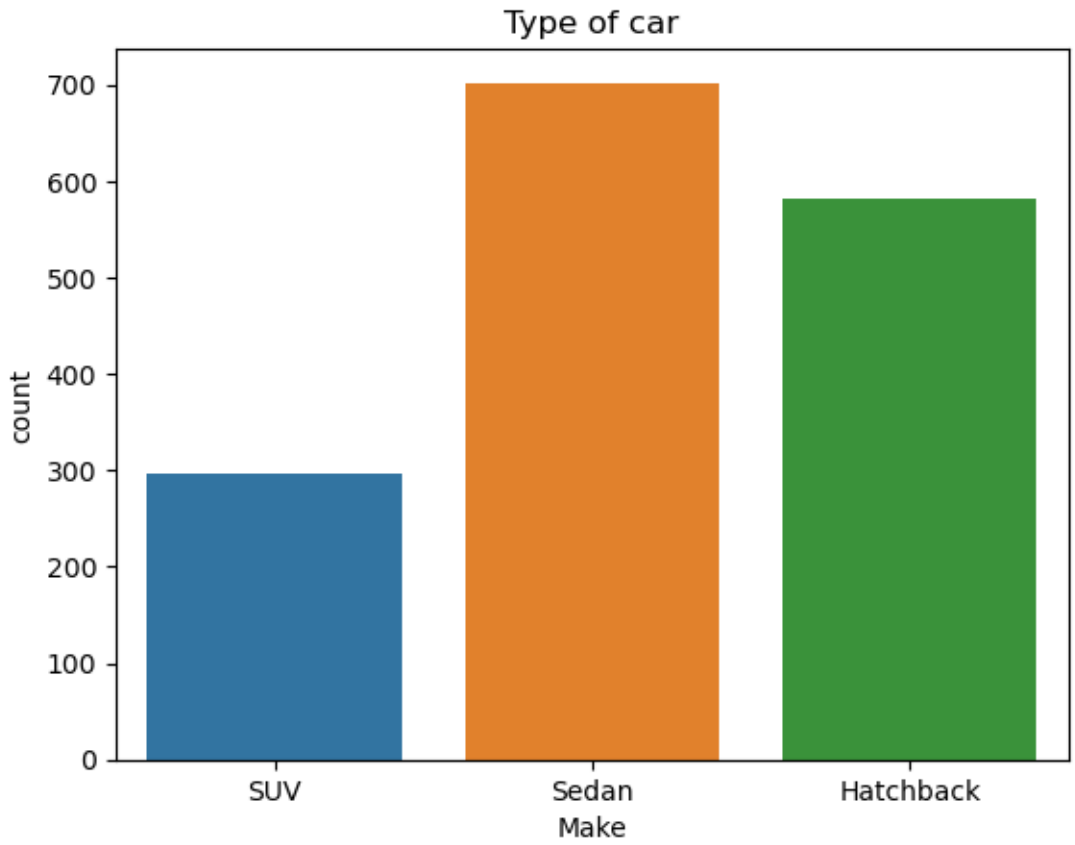


- Maximum no. of cars is priced between $20000 to $35000

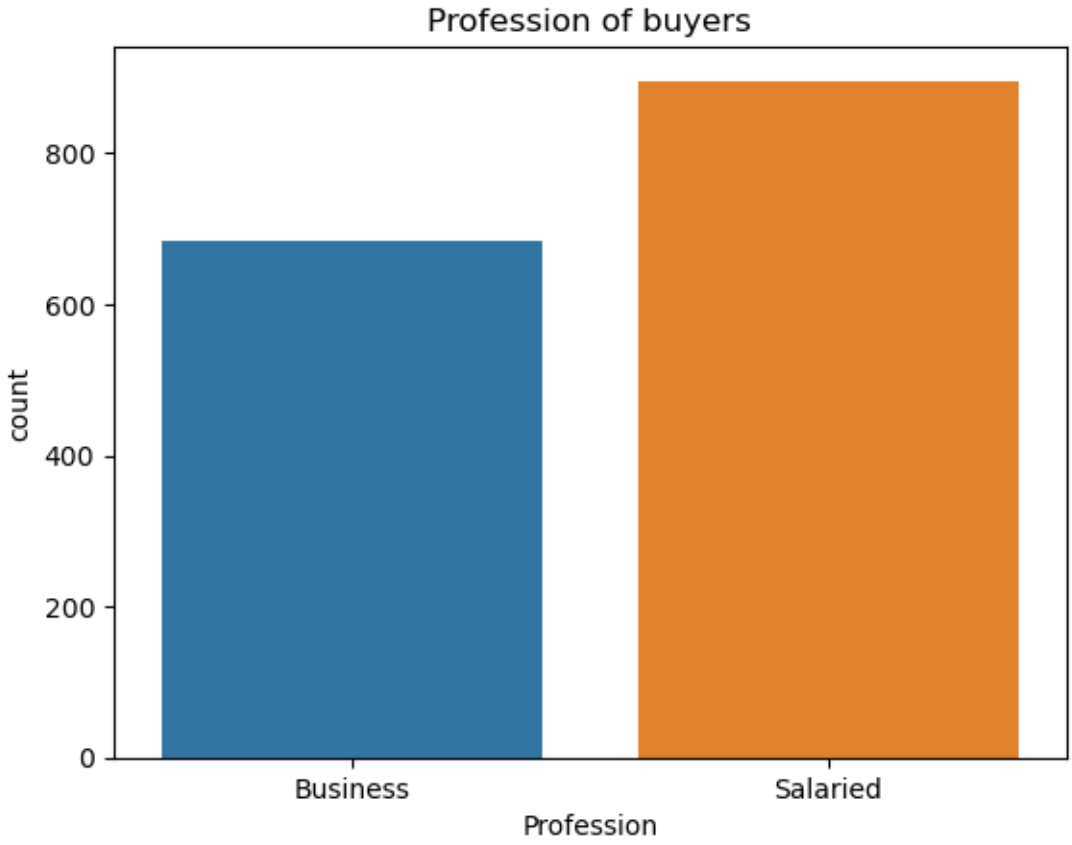


- Maximum no. of buyers is Male

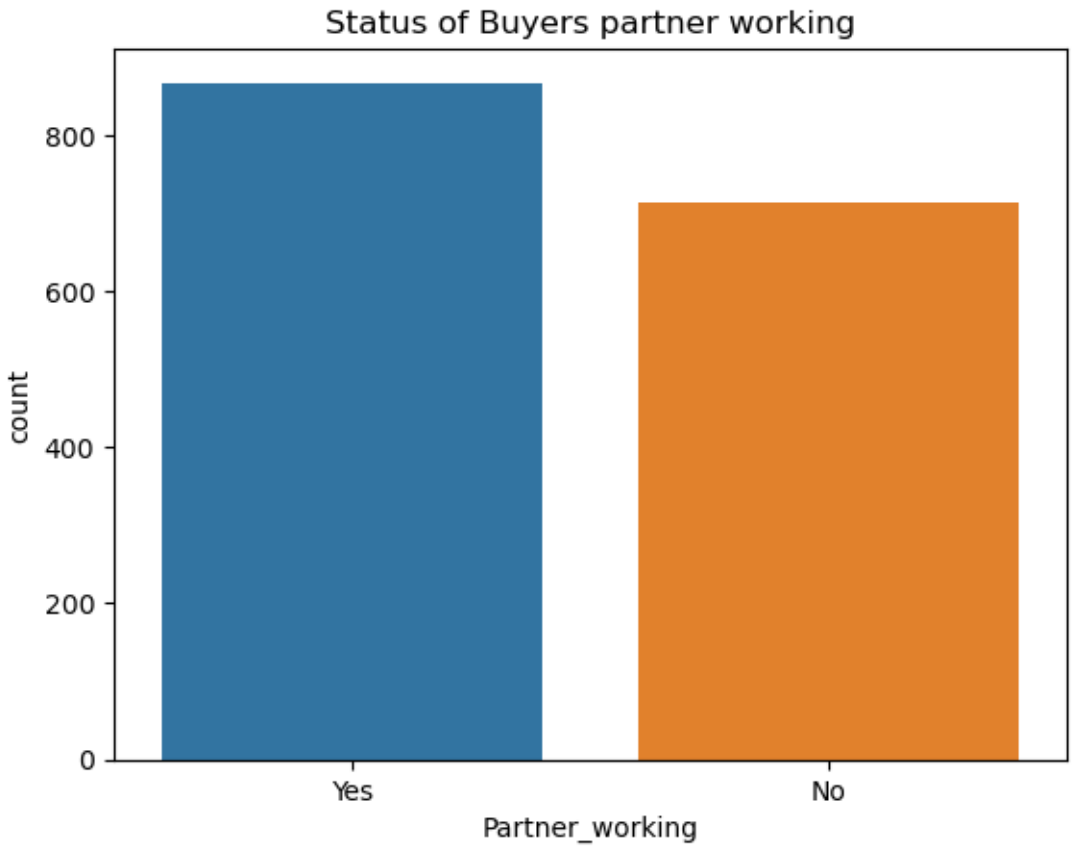
- No. of buyer's opted for House loan are half of the number of people not opted for house loan



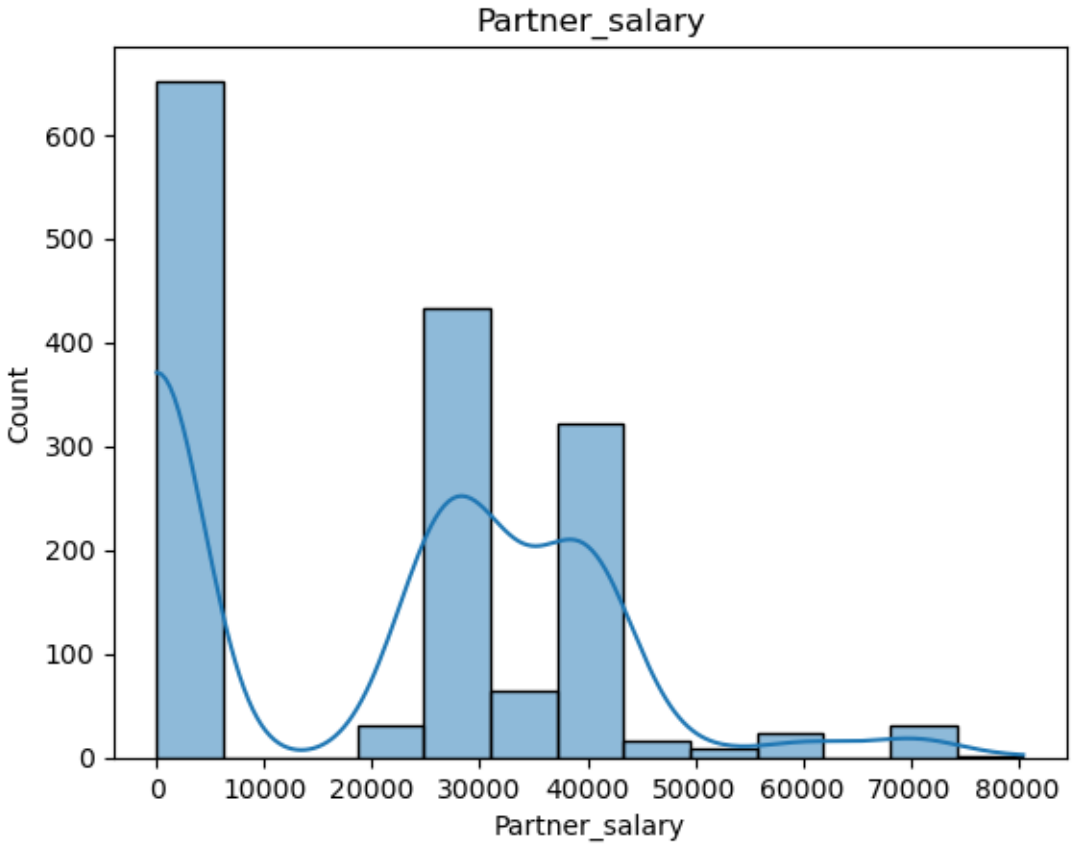
- The highest number of people prefer sedans over SUV and Hatchback



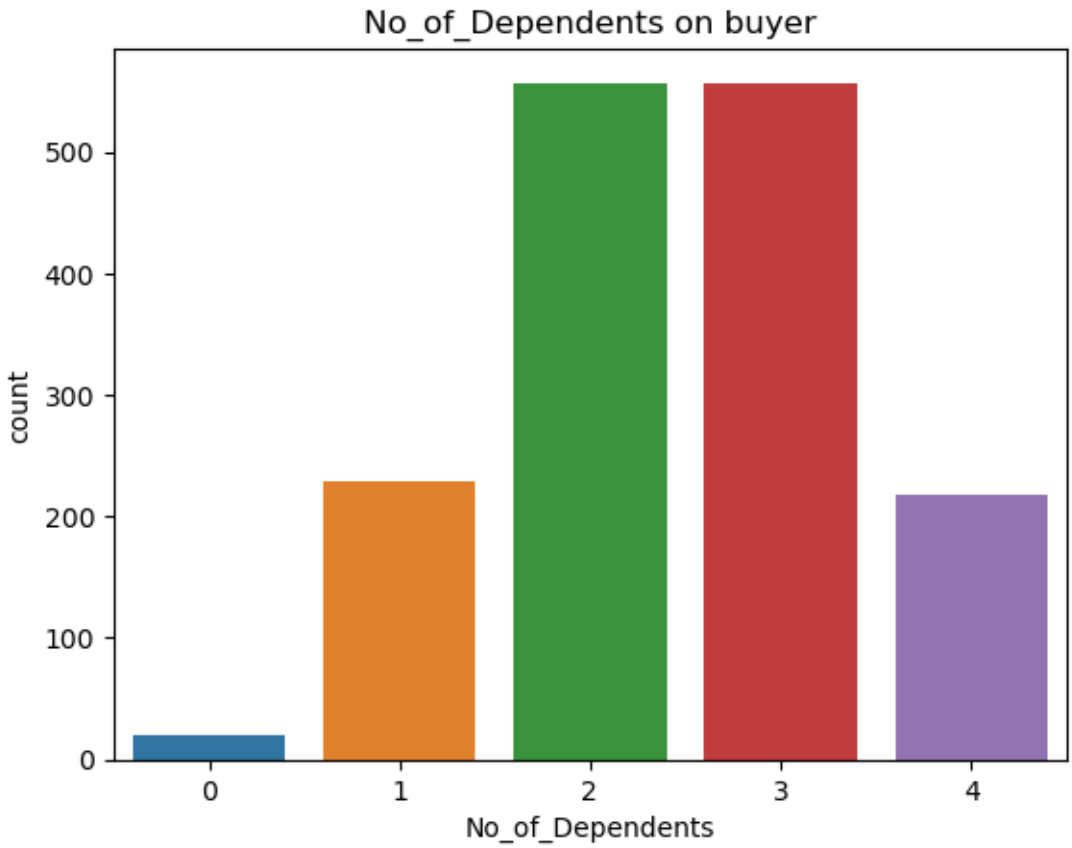
- Salaried individuals are more than business individuals



- There is little difference between buyer's having partner working or not.

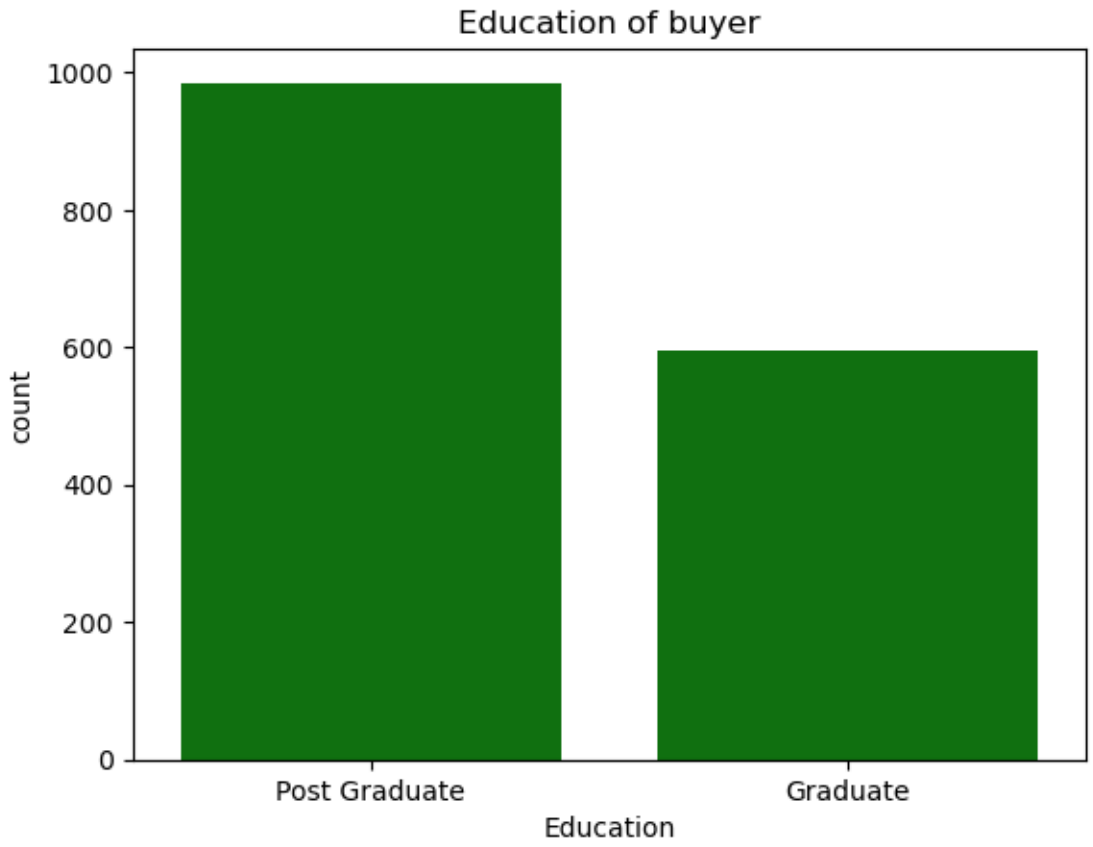


- Most of the Partner salary is in the range of $30000 to $40000

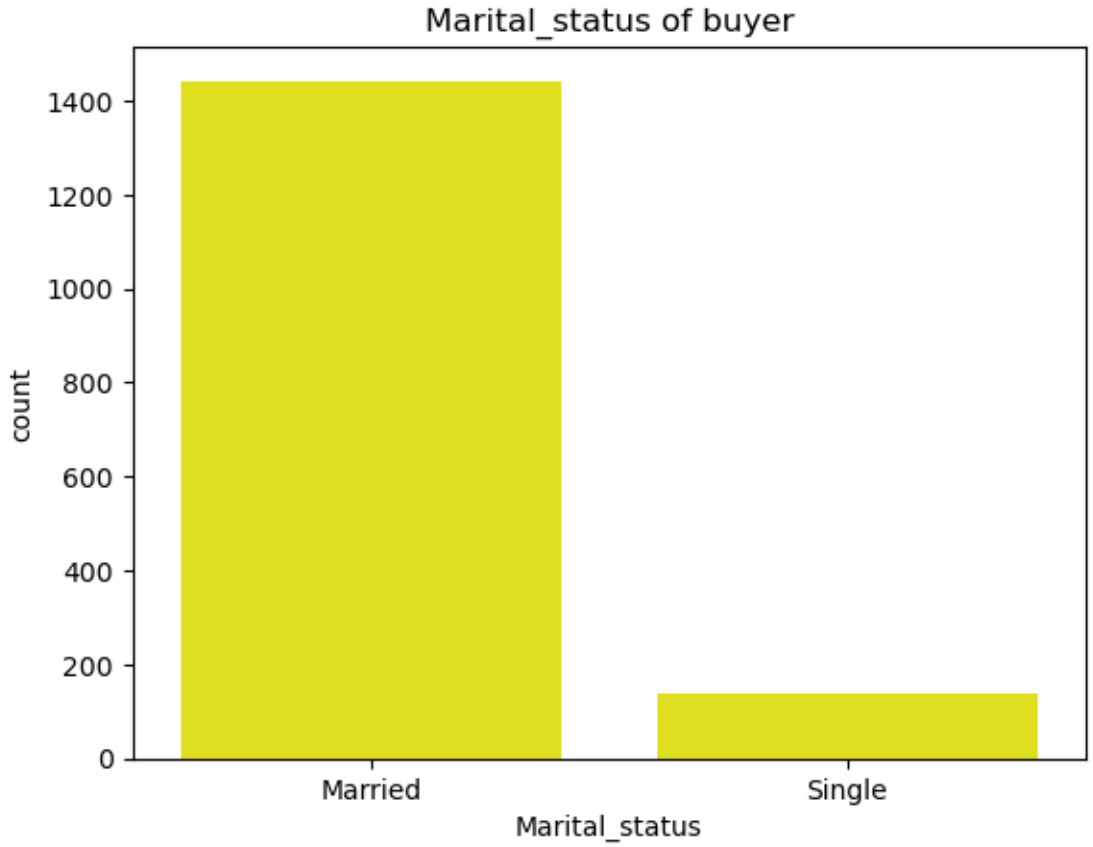


- Over 500 people have 2 or 3 dependents and nearly 200 people have 1 or 4 dependents

- The no. of people having the Personal loan is equal to the number of people not having the personal loan

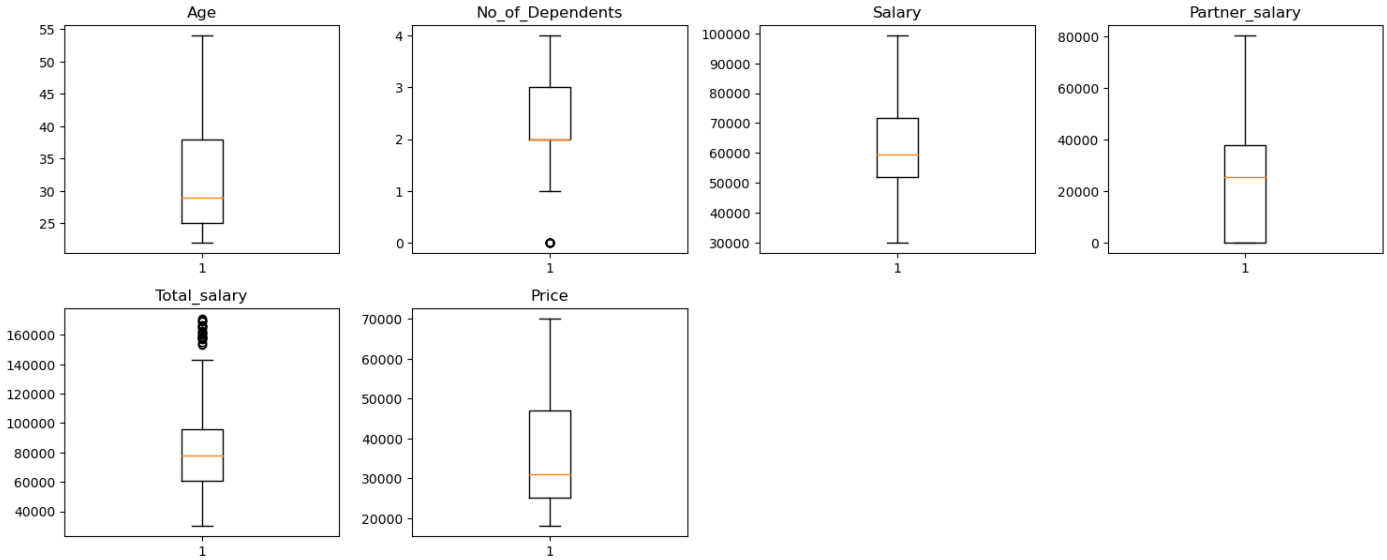


- No. of Post Graduate people are twice that of no. of graduate people

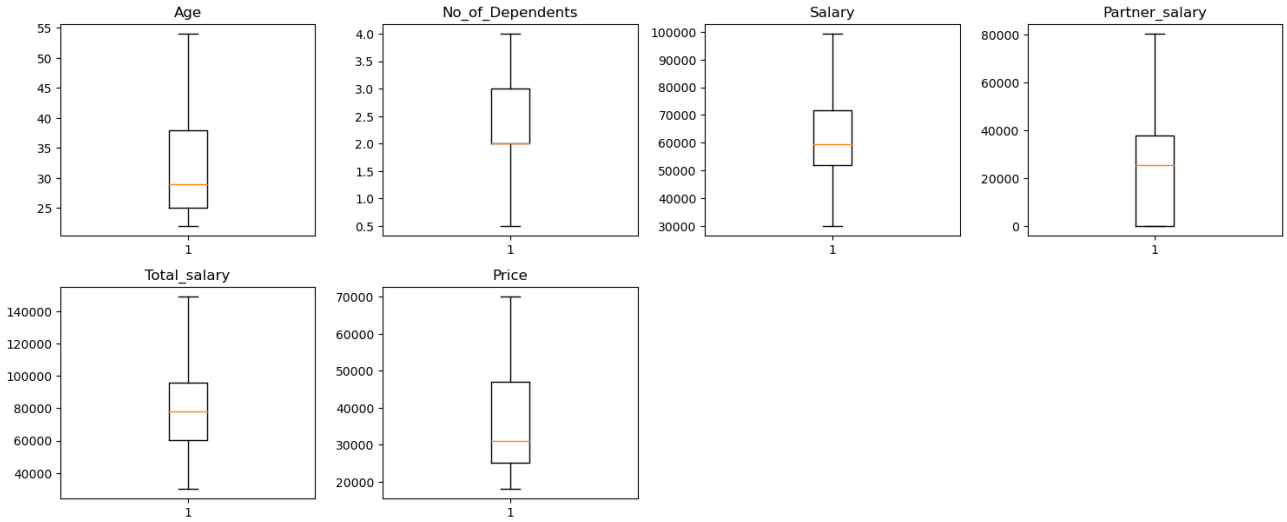


- Married people are significantly more than single people

2.2Check and treat outliers



- Observed Outliers in No_of_Dependents columns and Total_salary
- Treated the Outliers by using clip function from NumPy library
- The outliers bellow the lower whisker is clipped to Q1 and outliers above the upper whisker is clipped to Q3.

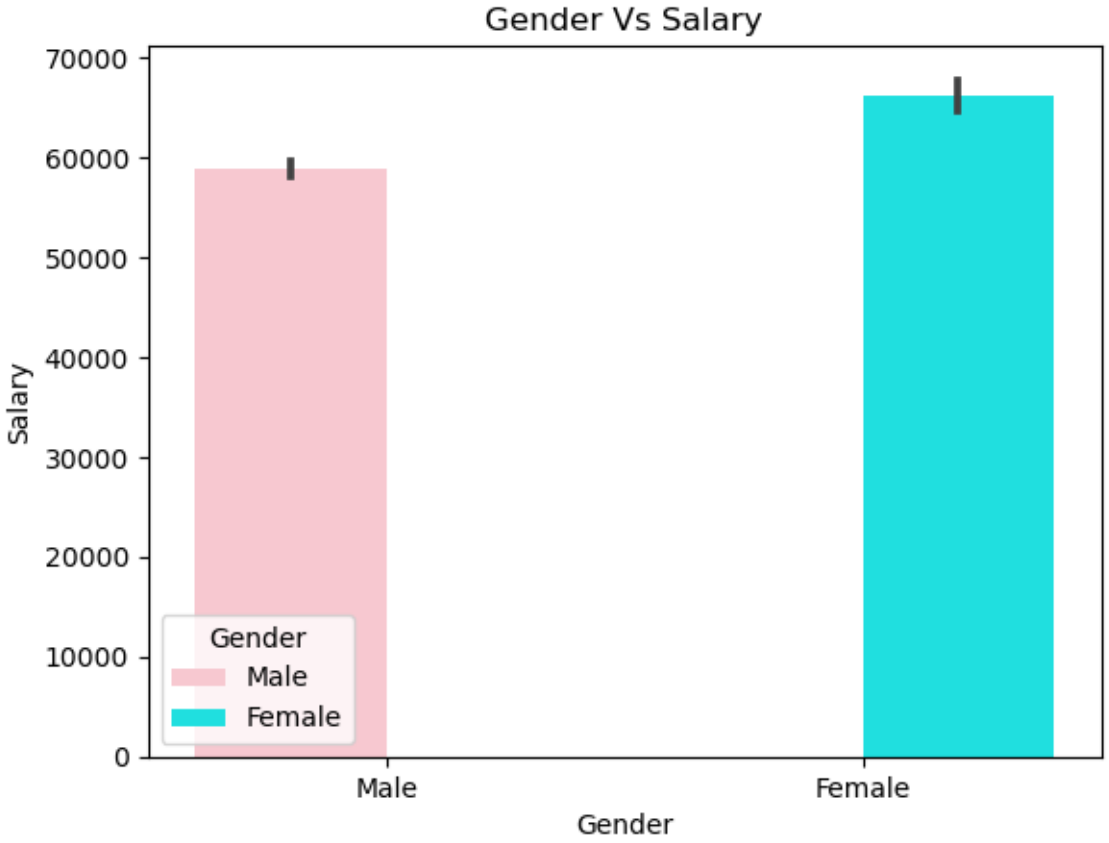


2.3 Observations and Insights

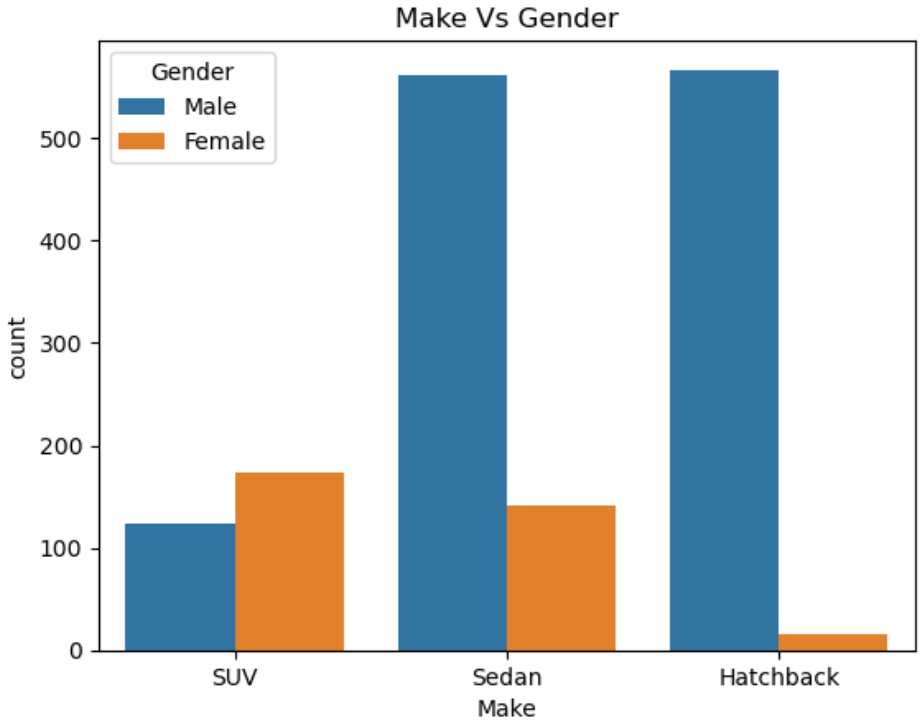
- Observed the relationship between the numerical datatypes and categorical datatypes
- checked outliers present in the dataset and treated the outliers
- Graphical representation of numerical and categorical datatype

Problem 3: Bivariate Analysis

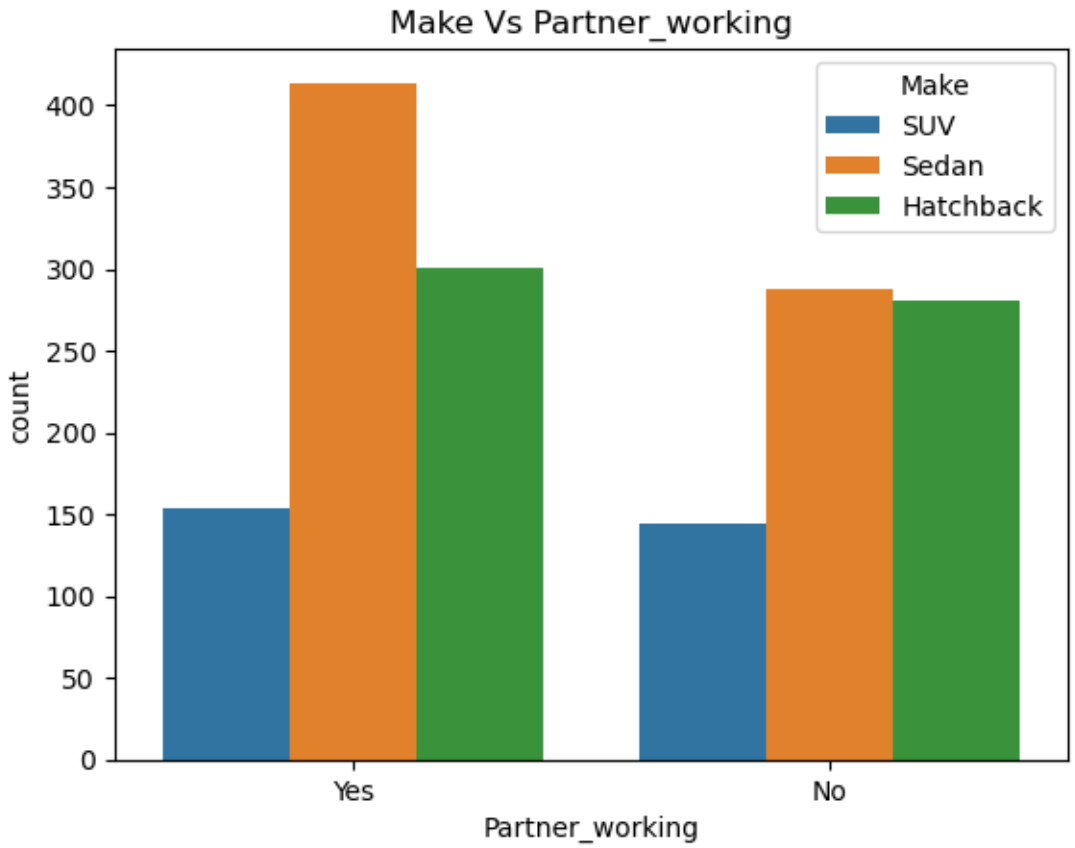
3.1 Explore the relationship between all numerical variables



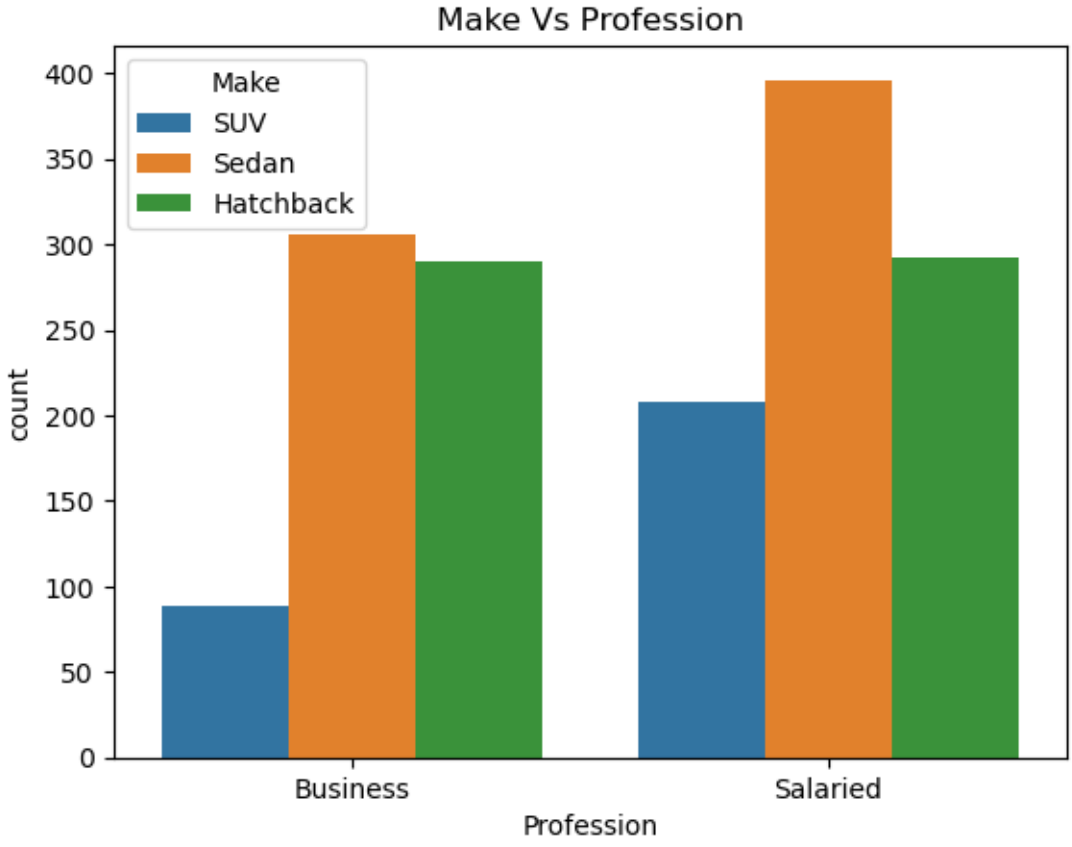
- Maximum amount of salary is earned by Female buyer's when compared to male buyer's



- There are more Male buyers for sedan and Hatchback while there are more Female buyers for SUV

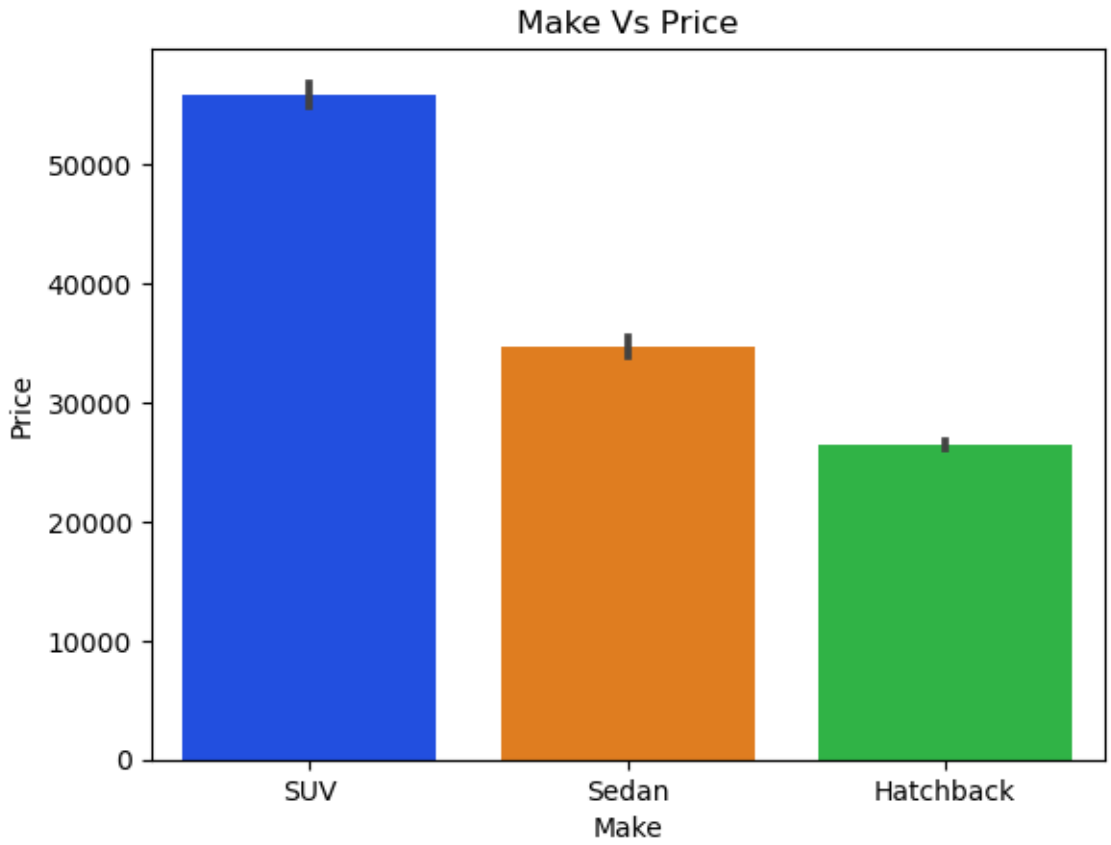


- Most of the buyers having working partner prefer sedan over SUV and Hatchback

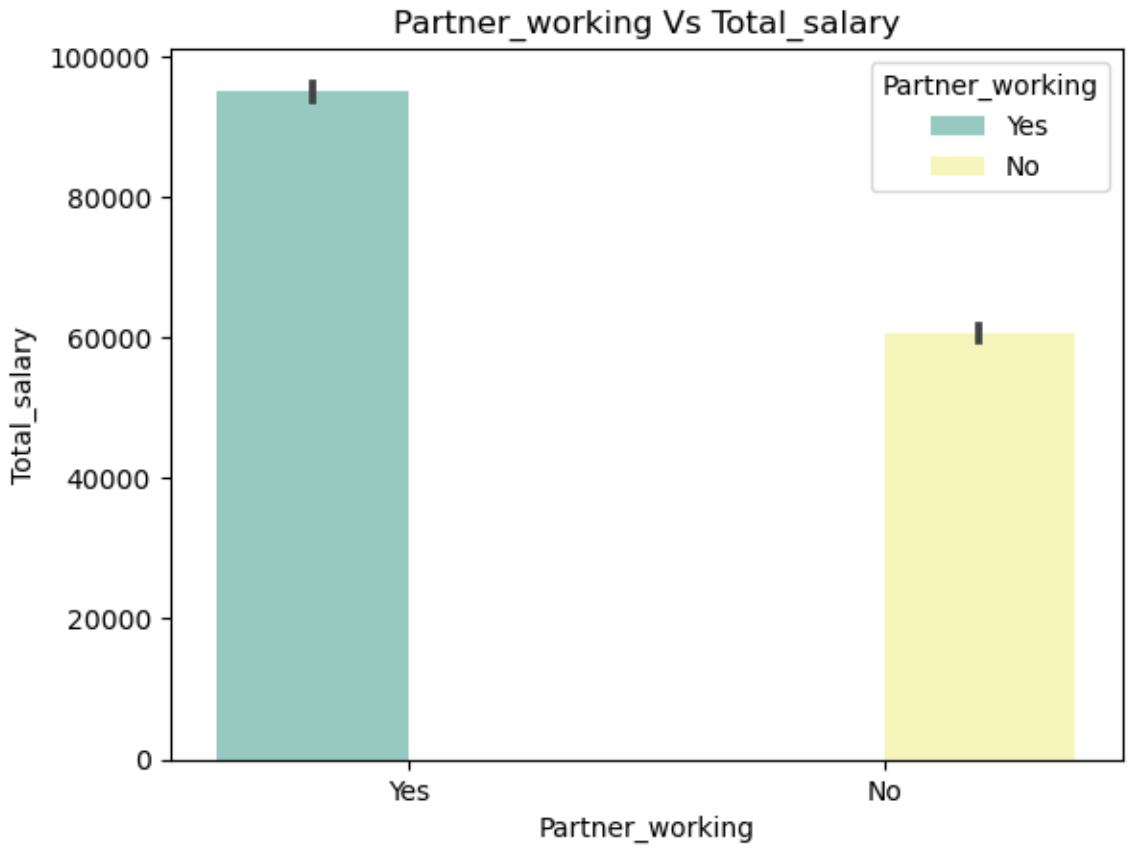


- Salaried professionals prefer Sedan and SUV over Business professionals whereas Hatchbacks are equally preferred

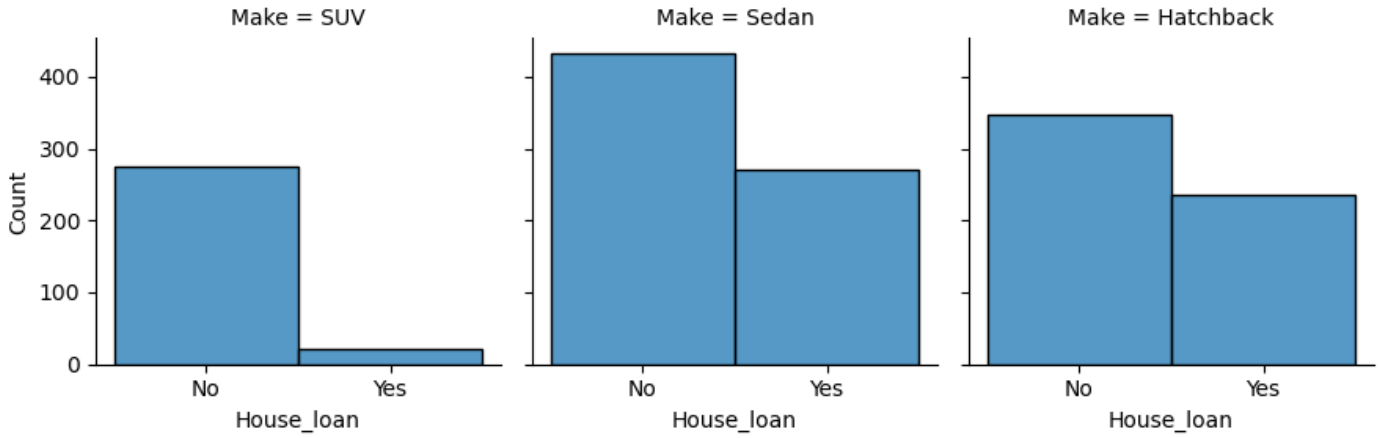
- Buyers prefer SUVs over sedan and Hatchback after combining their partners salary



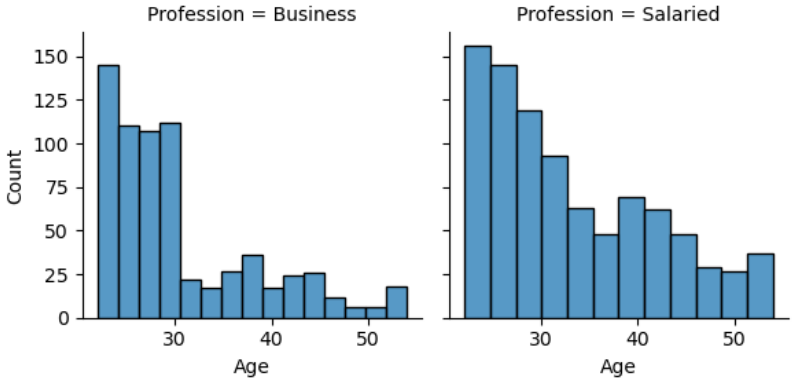
- SUVs are priced higher than sedan and Hatchback



- Buyers with working partners have greater total salary than buyers not having working partners

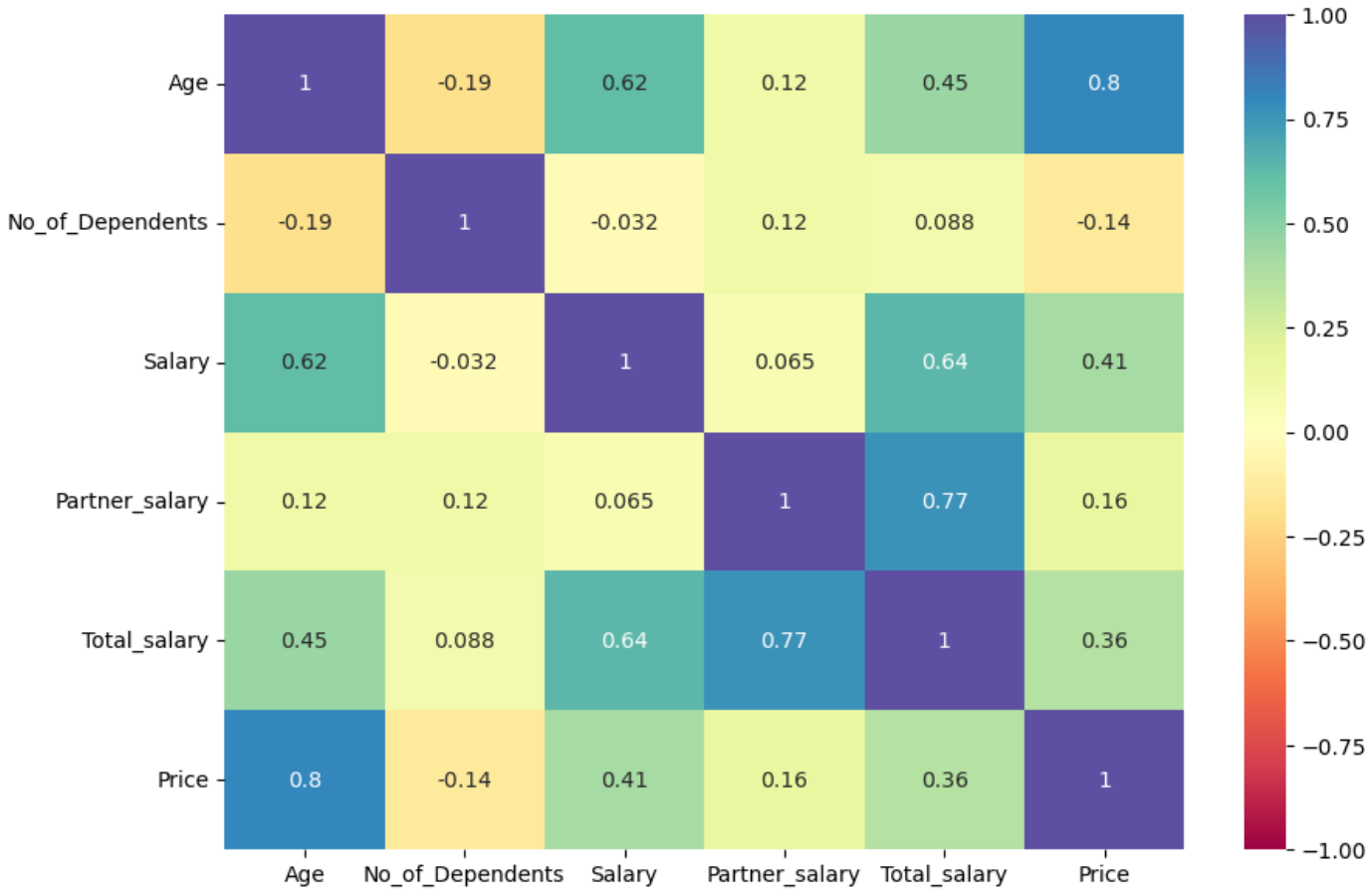


- Large number of Sedan have house loan



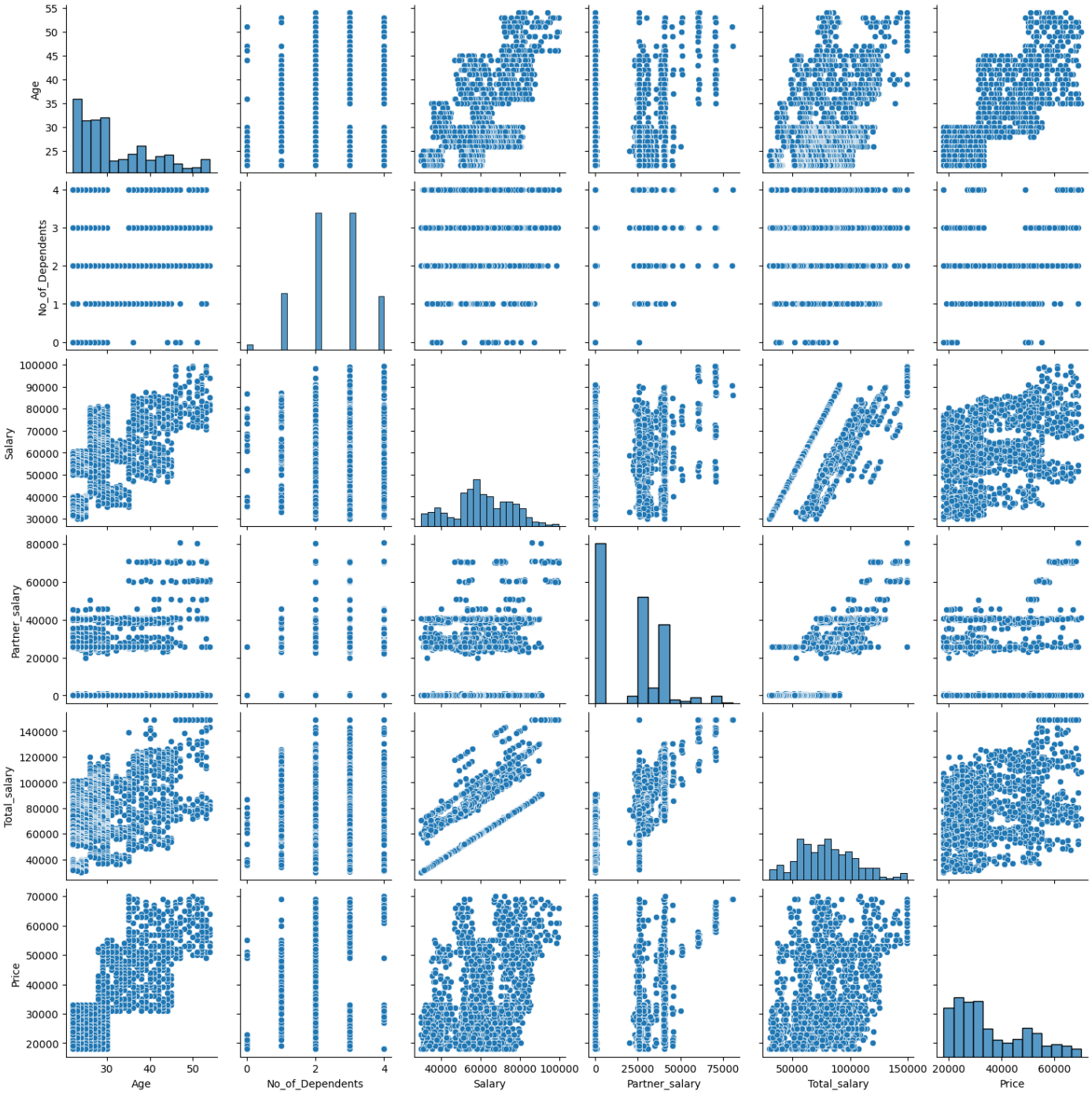
- There is more professional bellow the Age 30

3.2 Explore the correlation between all numerical variables.



- The price and age of buyers are positively corelated with 0.8 , this means with the increase in age tend to buy higher priced cars
- Total Salary and partners salary are positively corelated with 0.77
- No. of dependents and price are negatively corelated with -0.14 this means with increase in no. dependents the price is less
- No. of dependents and salary of the buyers are also negative corelated with -0.032

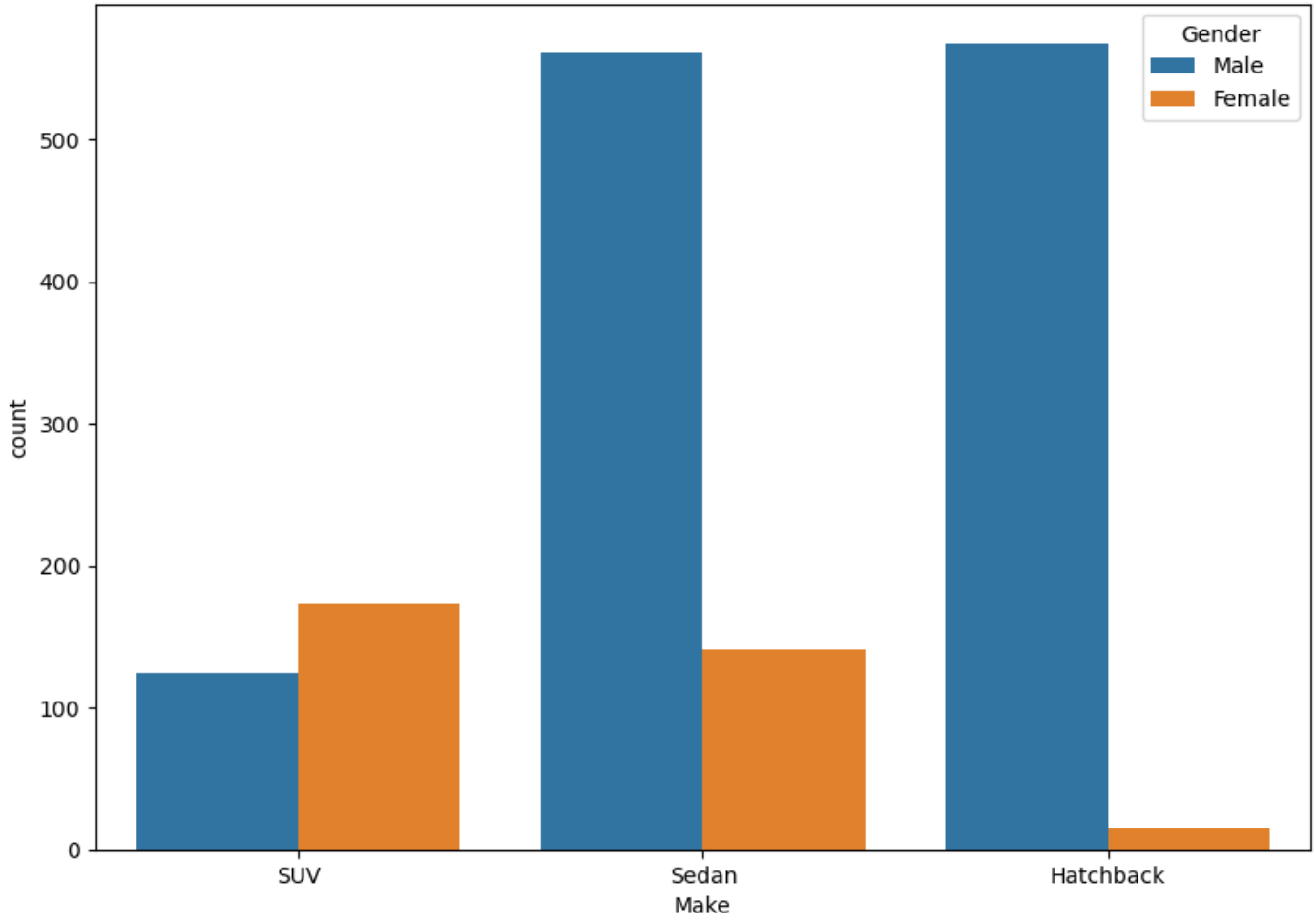
3.3 Explore the relationship between categorical vs numerical variables



Problem 4: Key Questions

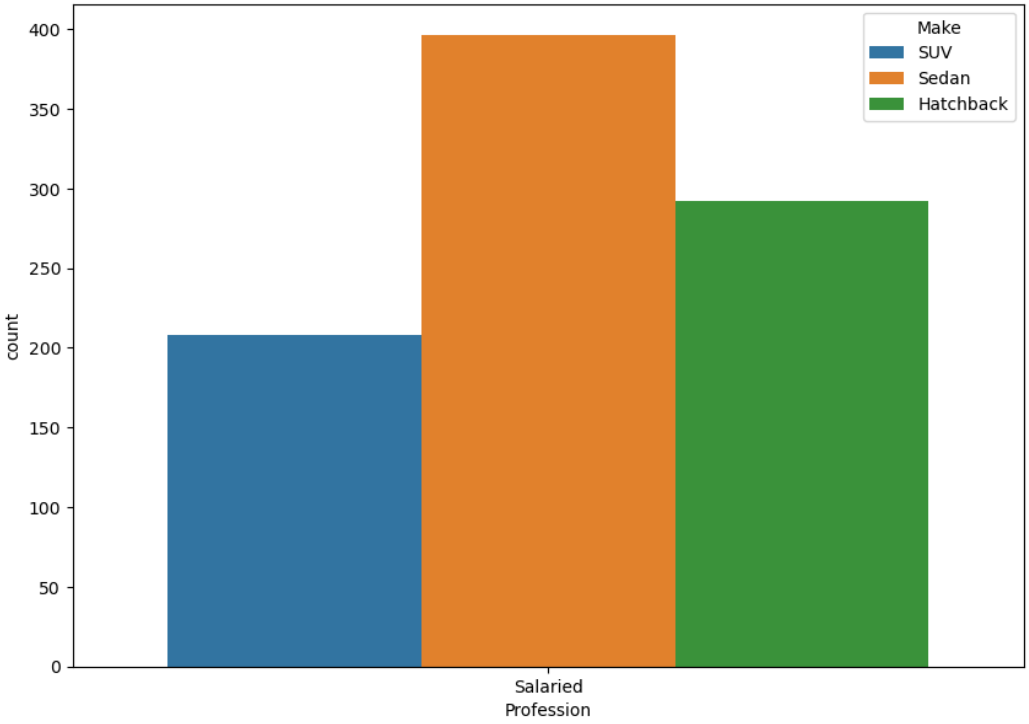
4.1 Do men tend to prefer SUVs more compared to women?

Ans) No, Women tend to prefer SUVs more compared to men.



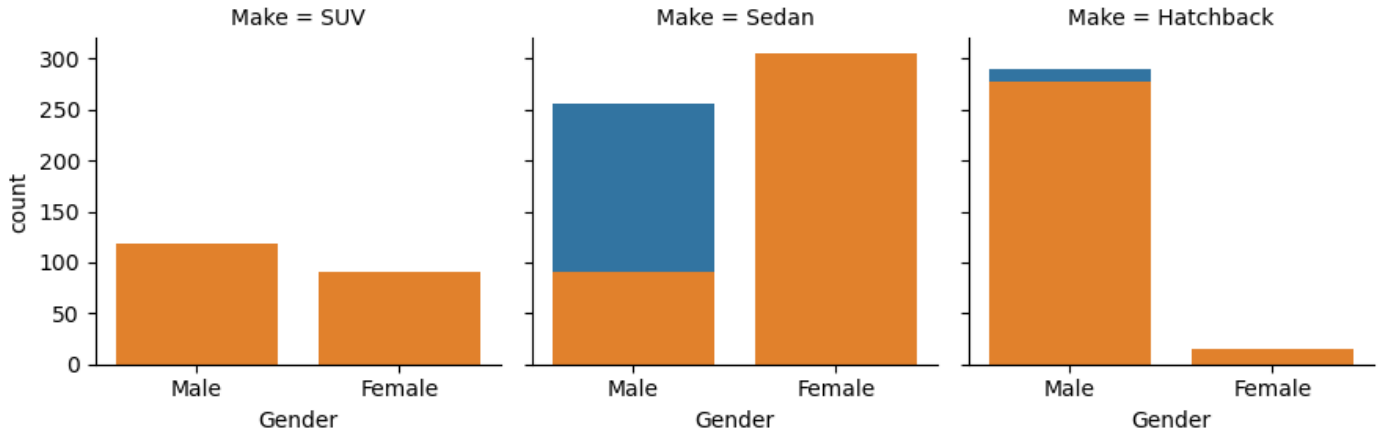
4.2 What is the likelihood of a salaried person buying a Sedan?

Ans) A salaried person is more likely to by a Sedan and compared to SUV and Hatchback



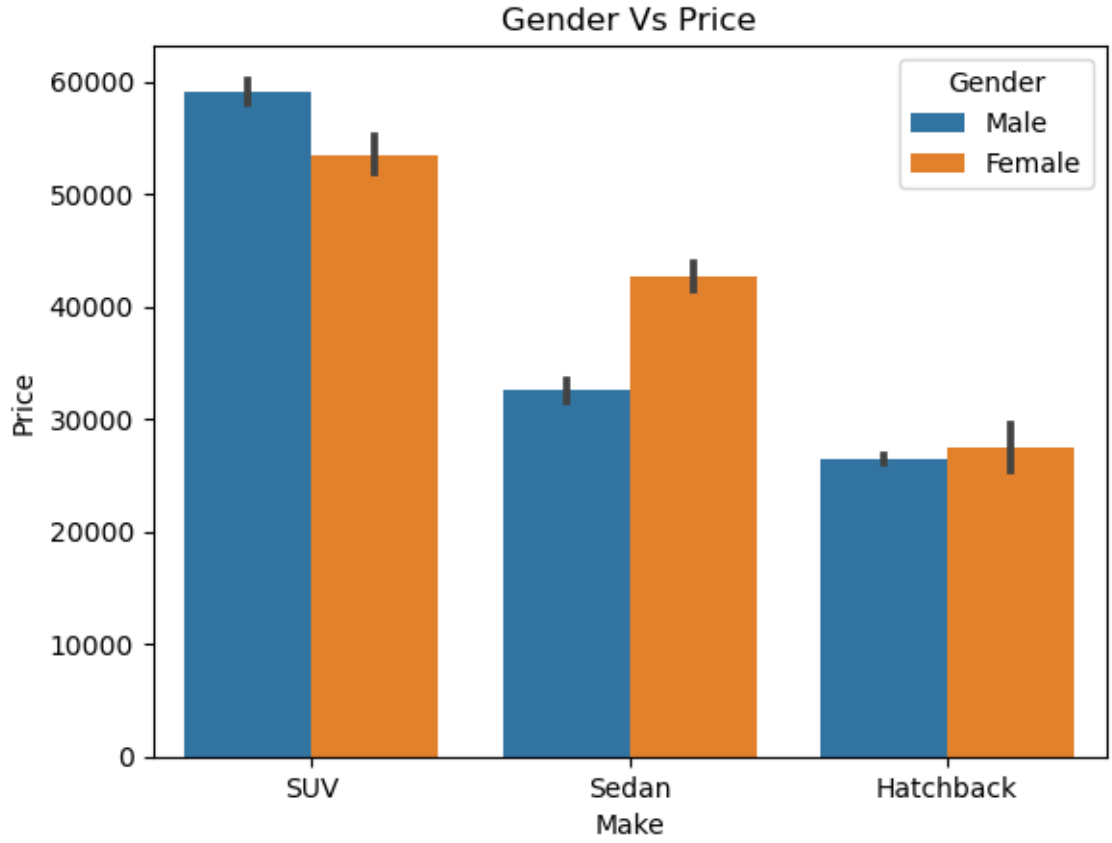
4.3 What evidence or data supports Sheldon Cooper's claim that a salaried male is an easier target for a SUV sale over a Sedan sale?

Ans) From the Graphs bellow evidence doesn't hold the Sheldon Cooper's claim



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

Ans) Female spend more than Male



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

Ans) The approximate amount spend on purchasing automobile by individual who tooks personal loan is $34k

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

Ans) Having a working partner doesn't influence the purchase of higher-priced cars

Problem 5: Actionable Insights & Recommendations

5.1 Actionable Insights & Recommendations

- Most of the buyers are from the Age group of 29 to 38 so campaign more frequently to this age group Buyers
- Most of the people have salary in between $50000 to $85000 and Only few People have salary more than $130000
- Maximum no. of cars is priced between $20000 to $35000
- Maximum no. of buyers is Male so campaign more for Male customers
- The highest number of people prefer sedans over SUV and Hatchback
- Salaried individuals are more Buyers than business individuals
- There are more Married people buying cars than single people
- Maximum amount of salary is earned by Female buyer's when compared to male buyer's
- There are more Male buyers for sedan and Hatchback while there are more Female buyers for SUV
- Most of the buyers having working partner prefer sedan over SUV and Hatchback
- Salaried professionals prefer Sedan and SUV over Business professionals whereas Hatchbacks are equally preferred
- Buyers prefer SUVs over sedan and Hatchback after combining their partners salary
- SUVs are priced higher than sedan and Hatchback
- Buyers with working partners have greater total salary than buyers not having working partners
- The price and age of buyers are positively corelated with 0.8, this means with the increase in age tend to buy higher priced cars
- Total Salary and partners salary are positively corelated with 0.77 so they have more potential to buy
- No. of dependents and price are negatively corelated with -0.14 this means with increase in no. dependents the price is less
- No. of dependents and salary of the buyers are also negative corelated with -0.032 so with an increase in no. of dependents the chance of buying is less.