Business Report

Contents

Objective:

To analyse the given dataset "**austo_automobile .csv**" consists of various factors such as Age, Gender, Profession, Marital status, Education, No of Dependents, Education, Personal loan, House loan, Partner working, Salary, Partner salary, Total Salary and their impacts on the Price and the Make of the item. This analysis aims to idenfiy the key trends the co-relation between the variables and the target variables leading to the potential strategies and the possible outcomes.

1. Problem : Data Overview

1.1 Import the libraries:

- Import pandas as pd
- Import numpy as np import matplotlib.pyplot as plt
- import seaborn as sns
- import matplotlib.pyplot as plt
- from scipy import stats

1.2 Load the Data:

	Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price	
0	53	Male	Business	Married	Post Graduate	4	No	No	Yes	99300	70700.0	170000	61000	
1	53	Femal	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500	70300.0	165800	61000	
2	53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300	60700.0	158000	57000	
3	53	Female	Salaried	Married	Graduate	2	Yes	No	Yes	72500	70300.0	142800	61000	
4	53	Male	Salaried	Married	Post Graduate	3	No	No	Yes	79700	60200.0	139900	57000	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
1576	22	Male	Salaried	Single	Graduate	2	No	Yes	No	33300	0.0	33300	27000	Hatc
1577	22	Male	Business	Married	Graduate	4	No	No	No	32000	NaN	32000	31000	Hatc
1578	22	Male	Business	Single	Graduate	2	No	Yes	No	32900	0.0	32900	30000	Hatc

1.3 Structure of the data:

- Number of rows = 1581
- Number of columns = 14

1.4 Check the types of data:

- Age -int64
- Gender-object
- Profession-object

- Marital status-object
- Education-object
- No of Dependents-int64
- Education-object
- Personal loan-object
- House loan-object
- Partner working-object
- Salary-int64
- Partner salary-float64
- Total Salary-int64
- Price -int64
- Make-object

There are 1-float64,5int and 8 object datatypes.

1.5 Checking for the missing values:

- Missing values before handling:
- 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

There are missing values in "Gender"(53) and "Partner_salary"(106) columns.

Treating the missing values:

- Treating the missing values in “Partner_salary column using the median.
- Treating the missing values in “ Gender” column using the mode, most number of occurrence is male, the missing values are replaced with the male value.

After treating the missing values:

- 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

1.6 Checking the statistical summary:

	Age	No_of_Dependents	Salary	Partner_salary	Total_salary	Price
count	1581.000000	1581.000000	1581.000000	1475.000000	1581.000000	1581.000000
mean	31.922201	2.457938	60392.220114	20225.559322	79625.996205	35597.722960
std	8.425978	0.943483	14674.825044	19573.149277	25545.857768	13633.636545
min	22.000000	0.000000	30000.000000	0.000000	30000.000000	18000.000000
25%	25.000000	2.000000	51900.000000	0.000000	60500.000000	25000.000000
50%	29.000000	2.000000	59500.000000	25600.000000	78000.000000	31000.000000
75%	38.000000	3.000000	71800.000000	38300.000000	95900.000000	47000.000000
max	54.000000	4.000000	99300.000000	80500.000000	171000.000000	70000.000000

1.7 Checking and Treating the Data irregularities:

- No duplicate rows found.
- Female was misspelled as Femal Femle is changed with correct spelling “Female”

	Age	Gender	Profession	Marital_status	Education	No_of_
0	53	Male	Business	Married	Post Graduate	
1	53	Femal	Salaried	Married	Post Graduate	
2	53	Female	Salaried	Married	Post Graduate	
3	53	Female	Salaried	Married	Graduate	
4	53	Male	Salaried	Married	Post Graduate	
...	...	...	...	...	...	
1576	22	Male	Salaried	Single	Graduate	
1577	22	Male	Business	Married	Graduate	
1578	22	Male	Business	Single	Graduate	

	Age	Gender	Profession	Marital_status
0	53	Male	Business	Married
1	53	Female	Salaried	Married
2	53	Female	Salaried	Married
3	53	Female	Salaried	Married
4	53	Male	Salaried	Married

1.8 Observation and insights:

Introduction:

The dataset includes information about automobile purchases having attributes such as Age, Gender, Profession, Marital status, Education, No of Dependents, Education, Personal loan, House loan, Partner working, Salary, Partner salary, Total Salary, Price and Make.

Data overview:

The dataset consists of 1581 rows and 14 columns Initial exploration reveals that there are 53 missing values in the column "Gender" and 106 missing values in the column "Partner salary"

The missing values in the "Gender" Column is treated with most occurence value.

The missing values in the "Partner_salary" column is treated with the median value.

Female is misspelled as "Femal" is corrected with "Female" in "Gender" column

Data types:

The dataset consists of the data types, *int64(5) *float64(1) *object(8)

Statistical Summary:

Stastistical summary states that the average age is 31 and the average price is 35597.

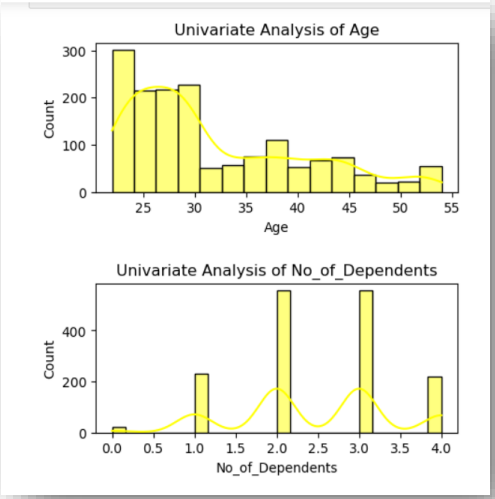
The price ranges of car from a minimum of 18000 to maximum of 70000.

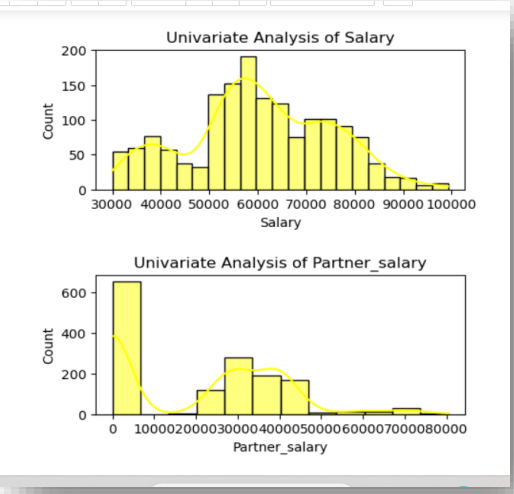
Duplicate rows:

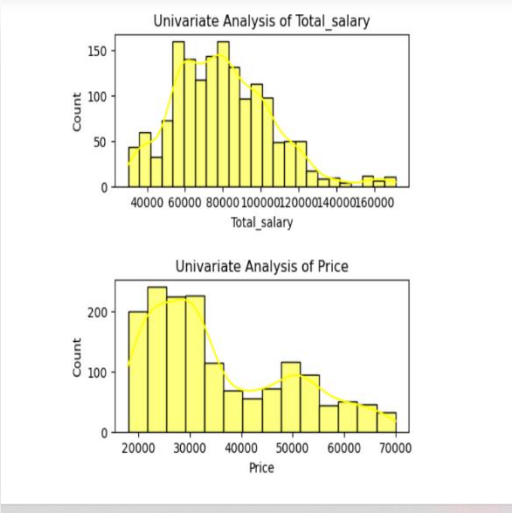
No traces of the duplicate rows.

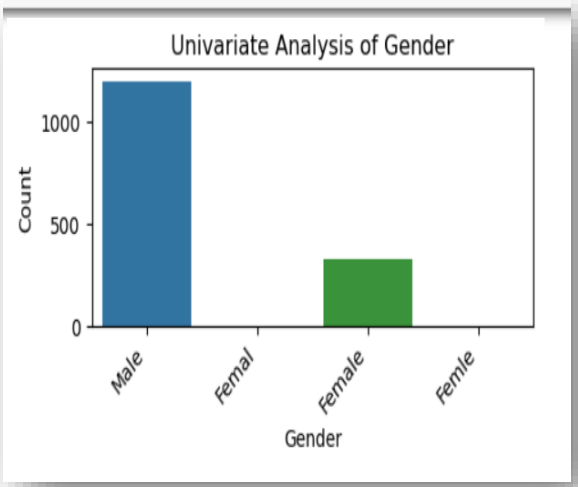
2.Problem : Univariate Analysis:

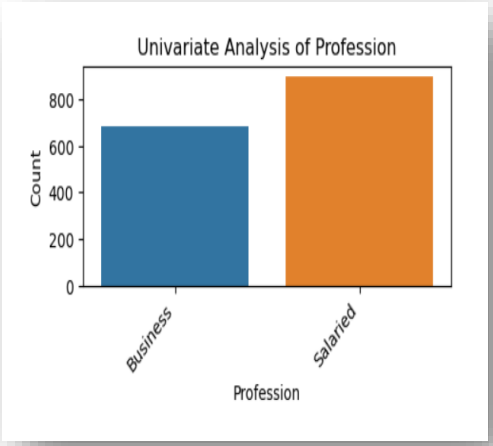
2.1 Explore all numerical and categorical variables:

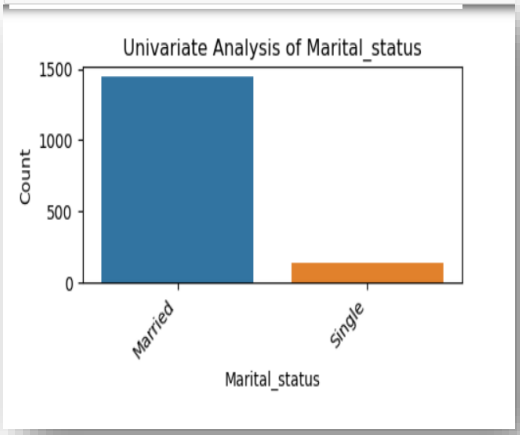


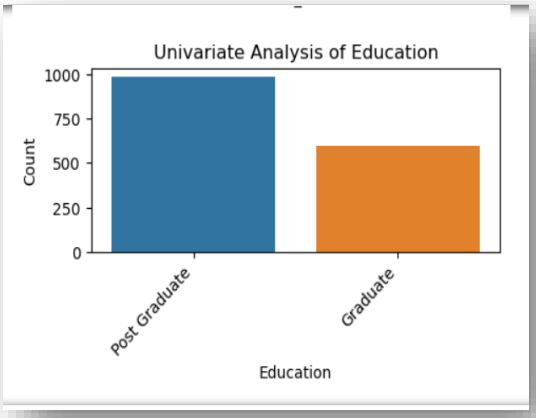


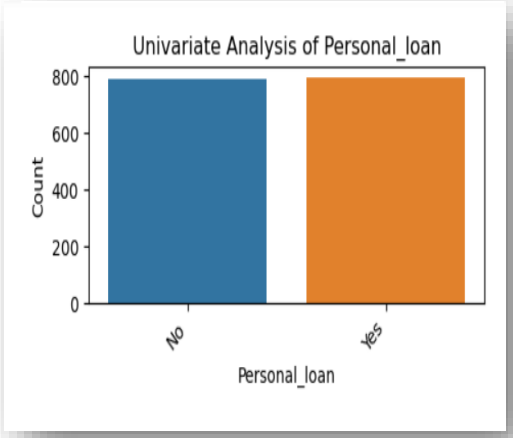


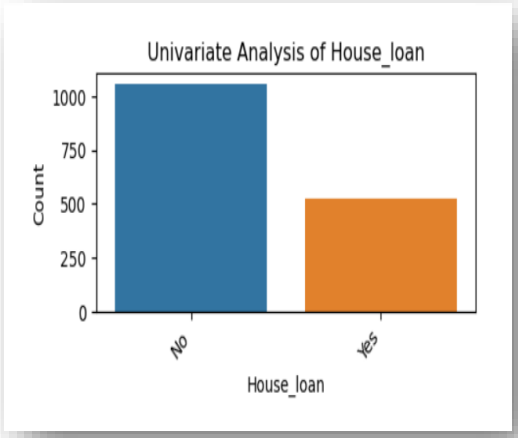


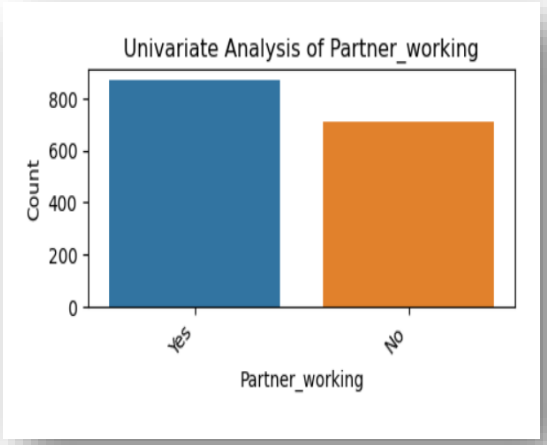


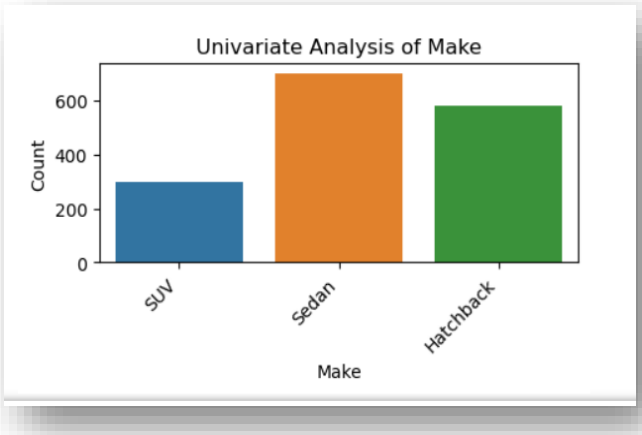








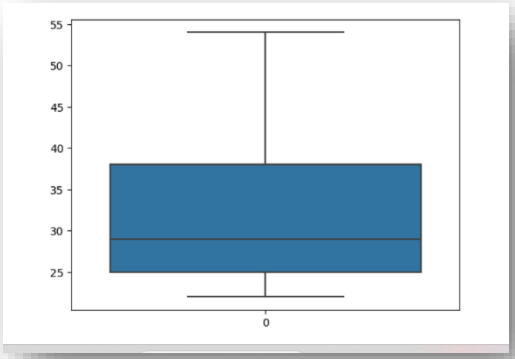




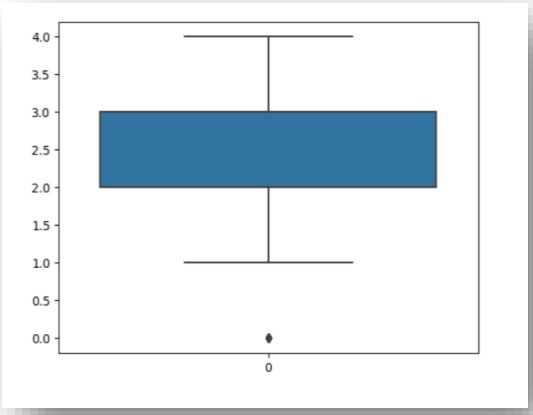
Explored all the variables using Univariate Analysis.

2.2 Checking and Treating the Outliers:

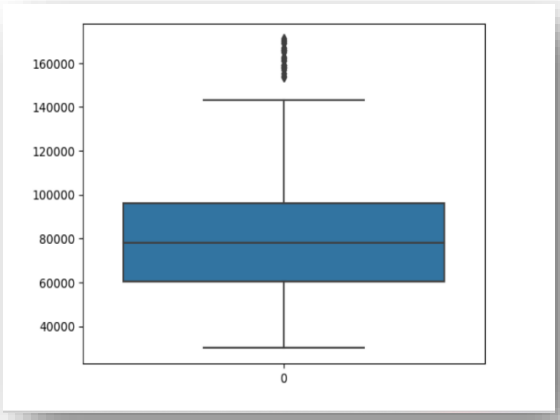
- Checking the outliers in age column.



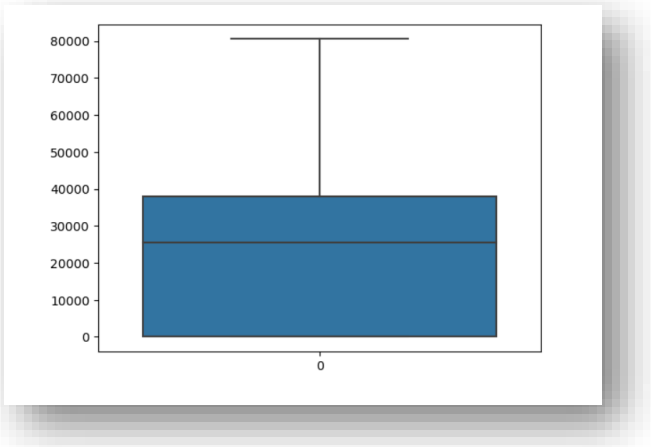
- Checking for Outliers in No of Dependents



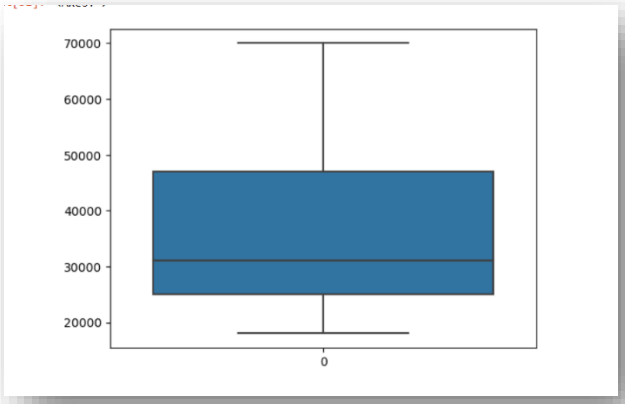
- Checking for Outliers in Total salary column



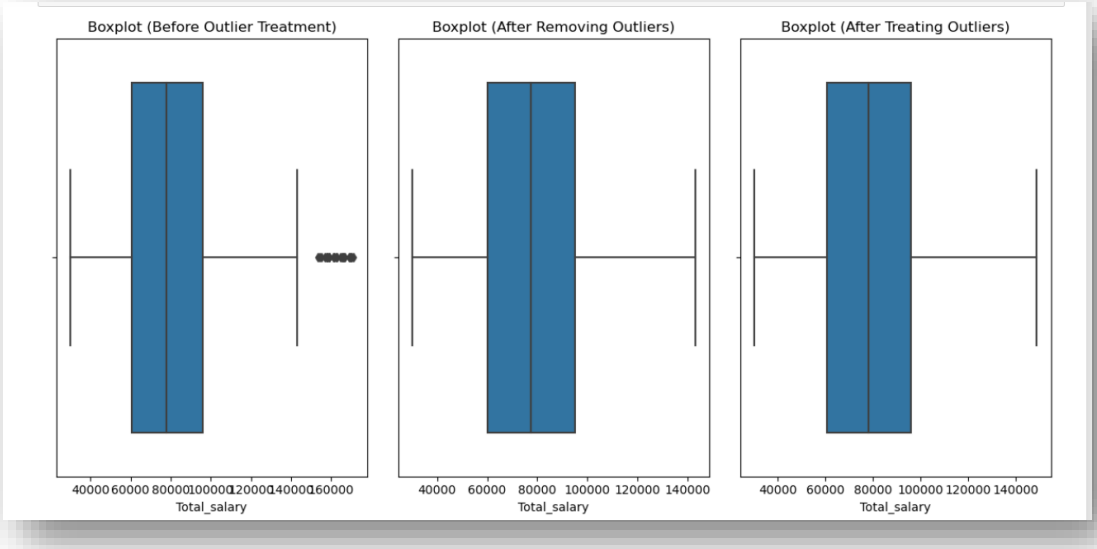
- Checking for outlier in Partner salary column



- Checking for outlier in Price salary column



- Total salary columns having considerable outliers that needed to be treated ,here is the treated output.



2.3 Observation and Insights:

Skewness of the Variables:

- Age-Positive Skewness (distribution of data is stretched to the right)
- No_of_Dependents-Negative Skewness (distribution of data is stretched to the left)
- Salary-Negative Skewness (distribution of data is stretched to the left)
- Partner_salary-Positive Skewness (distribution of data is stretched to the right)
- Total_salary-Positive Skewness (distribution of data is stretched to the right)
- Price-Positive Skewness (distribution of data is stretched to the right)

Age:

- The peak of the histogram, representing the mode, is less than 25 years, indicating a higher frequency of individuals around this age.
- The people between in the age category 45-50 are less in numbers.

No_of_Dependents:

- As per the histogram, more than 500 people are having 2-3 dependents.
- 200-300 of the people are having 1 and 4 dependents.

Salary-column:

- The peak of the histogram, representing that higher number of people are lie in the salary range between 50,000- 60,000.
- The people with the salary range of 90,000-1,00,000 are very less in numbers.

Partner_salary:

- Most of the people partners salary range lies between 0-10,000.
- 400-500 people have their partner who is earning nearly 30,000
- The people having their partners in the salary range of 50,000 and 80,000 are very less in numbers.

Total_salary:

- The highest number of people having their total salary ranges between 50,000 - 60,000.
- The people having their total salary ranges between 90,000-1,00,000 are less in numbers.

Price:

- More than 200 people are lies in the price range of 20,000 to 30,000.
- Only less than 50 people are lies in the price range of 60,000-70,000.

Gender:

- More than 1200 people are 'Male' whereas 'Female' are less than 300

Professsion:

- More than 800 people are salaried professionals
- 600-700 people are doing business

Marital status:

- More than 1200 people are married.
- Only less than 200 people are remains single.

Education:

- 800-1000 People are having a POST-GRADUATE degree.
- Nearly 600 people are having GRADUATE degree.

Personal loan:

- 700-800 people are bought personal loan.
- 700 people don't bought personal, nearly half of the people dont't bought personal loan.

Home-loan:

- 400-600 people bought home loan.
- More than 1000 people don't bought house loan.

Partner-working:

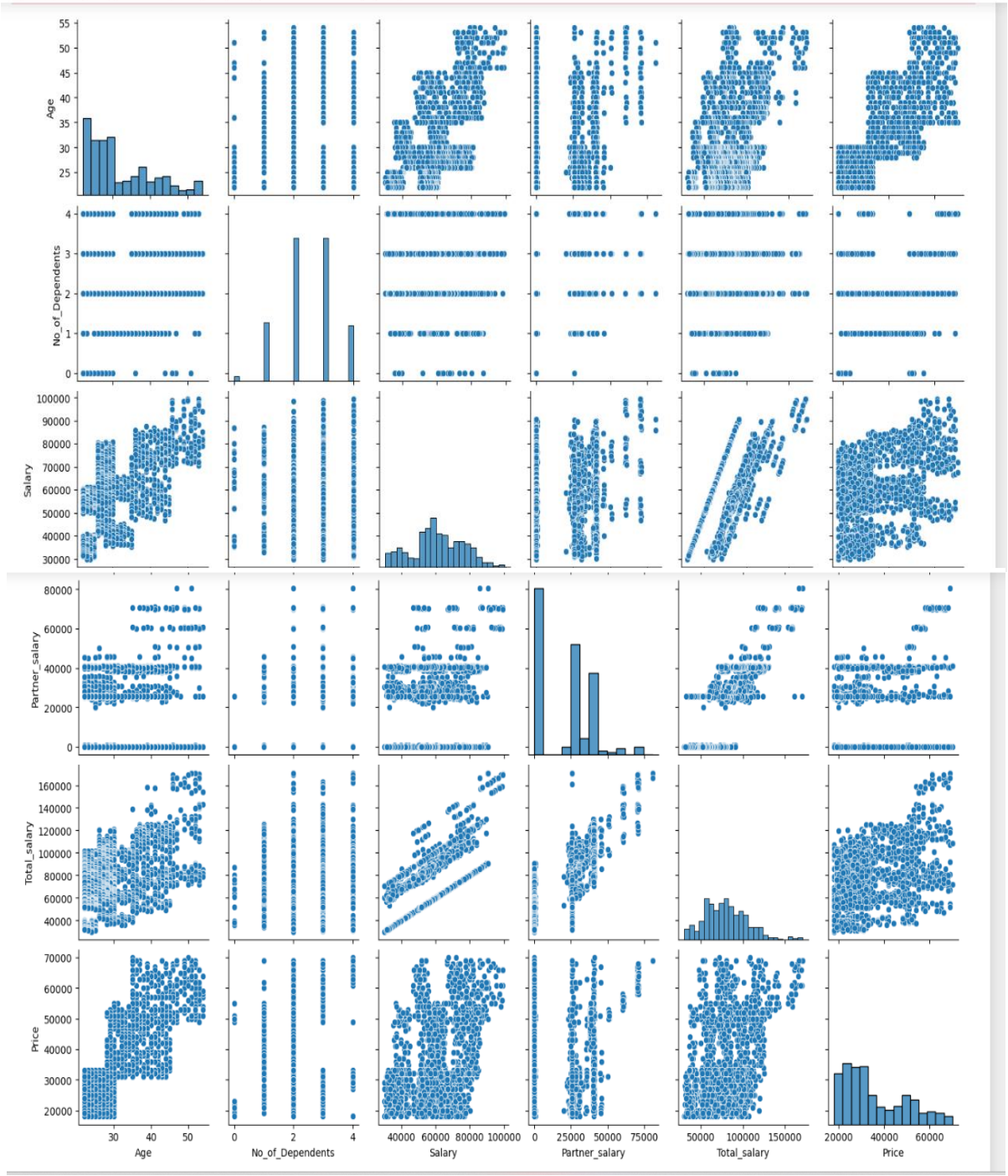
- More than 800 people partner are employed.
- 600-700 people partners are unemployed.

Make:

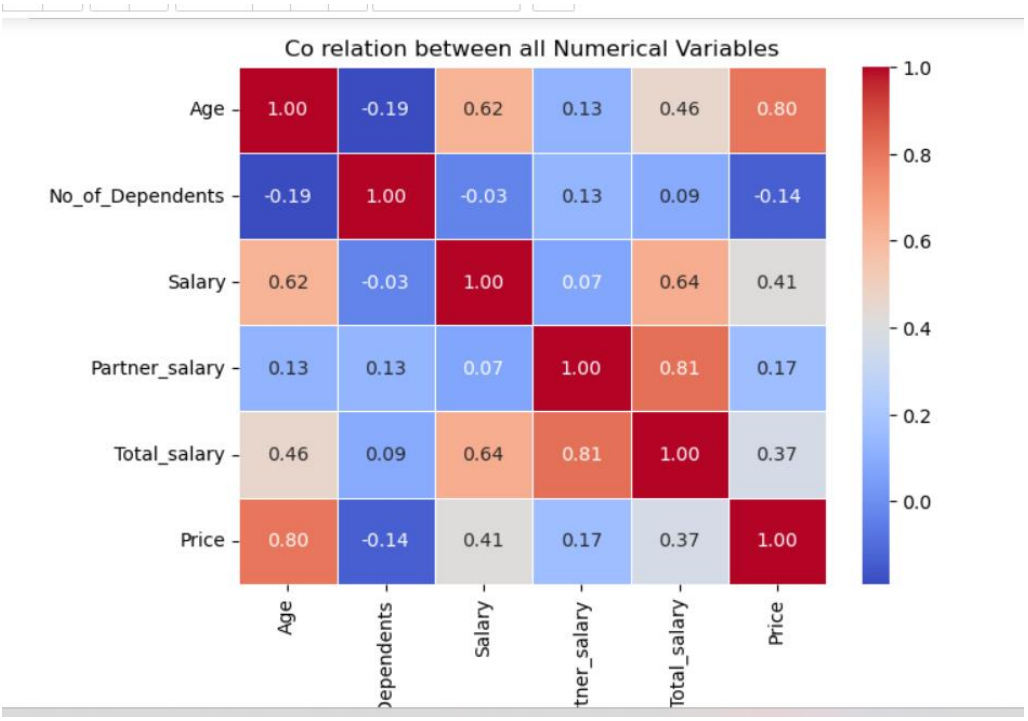
- Nearly 700 people bought SEDAN.
- 500-600 people chosen HATCHBACK.
- Only 250-300 people chosen SUV, SUV performed less in numbers.

3.Problem: Bivariate Analysis:

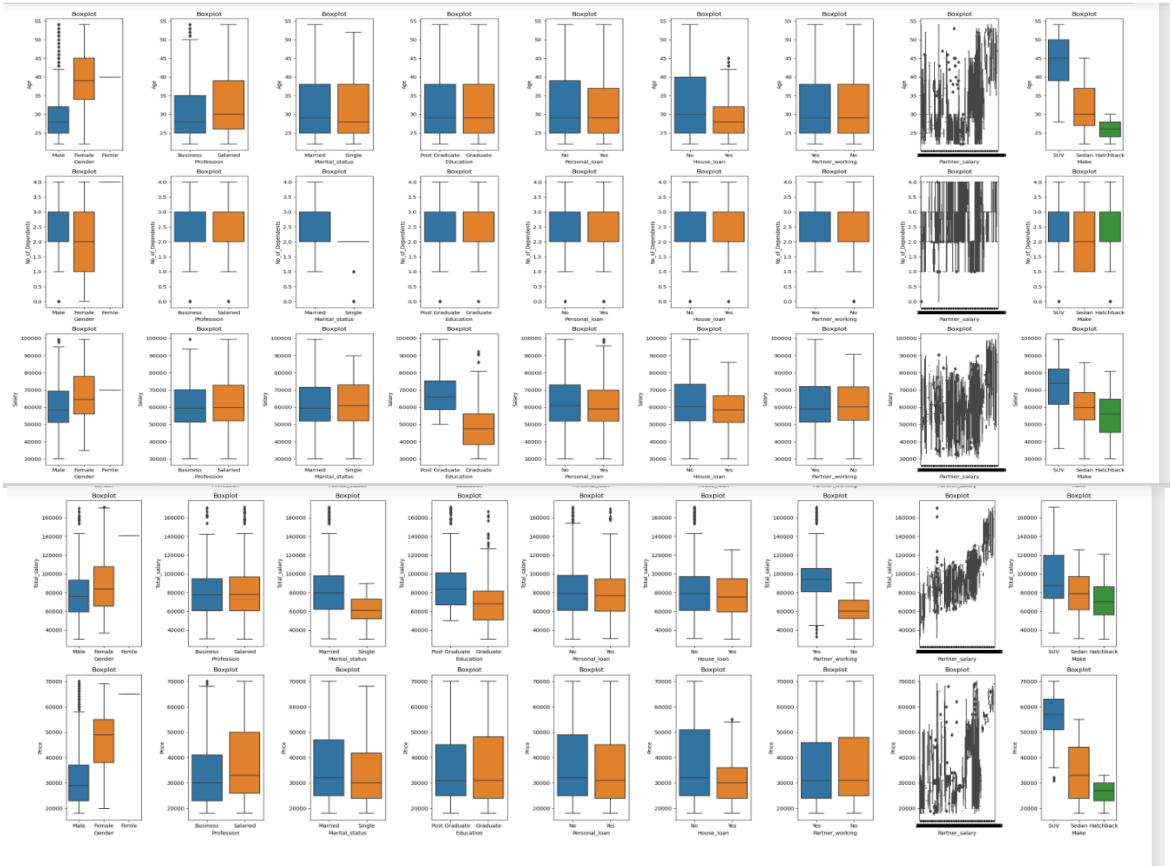
3.1 The Relationship between all numerical variables:



3.2 The Correlation between all numerical variables:

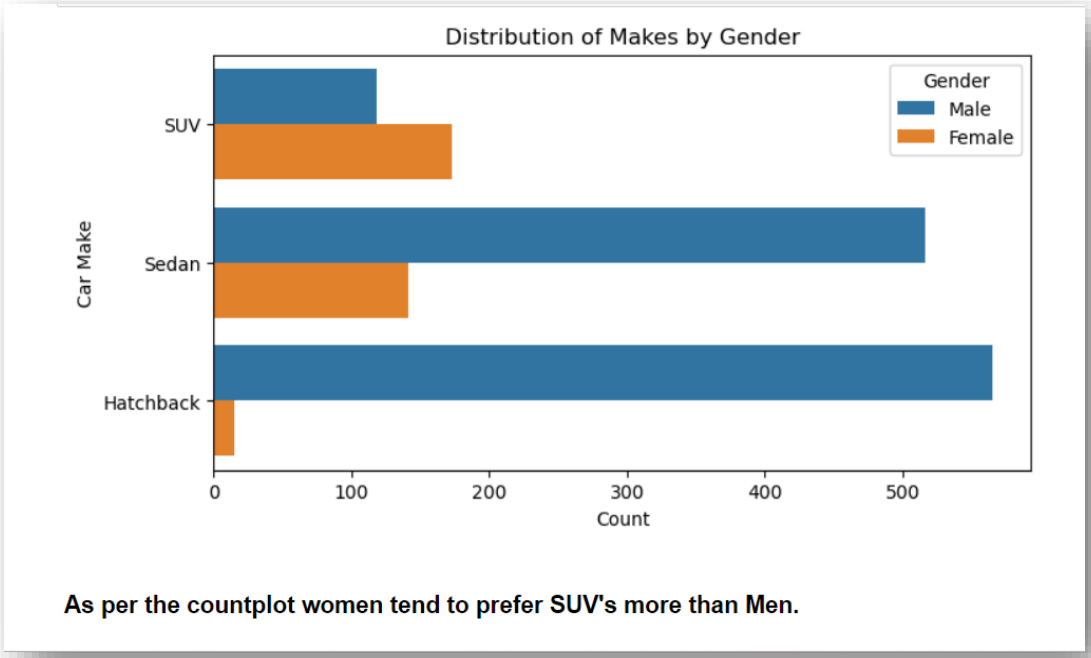


3.3 The Relationship between categorical vs numerical variables:



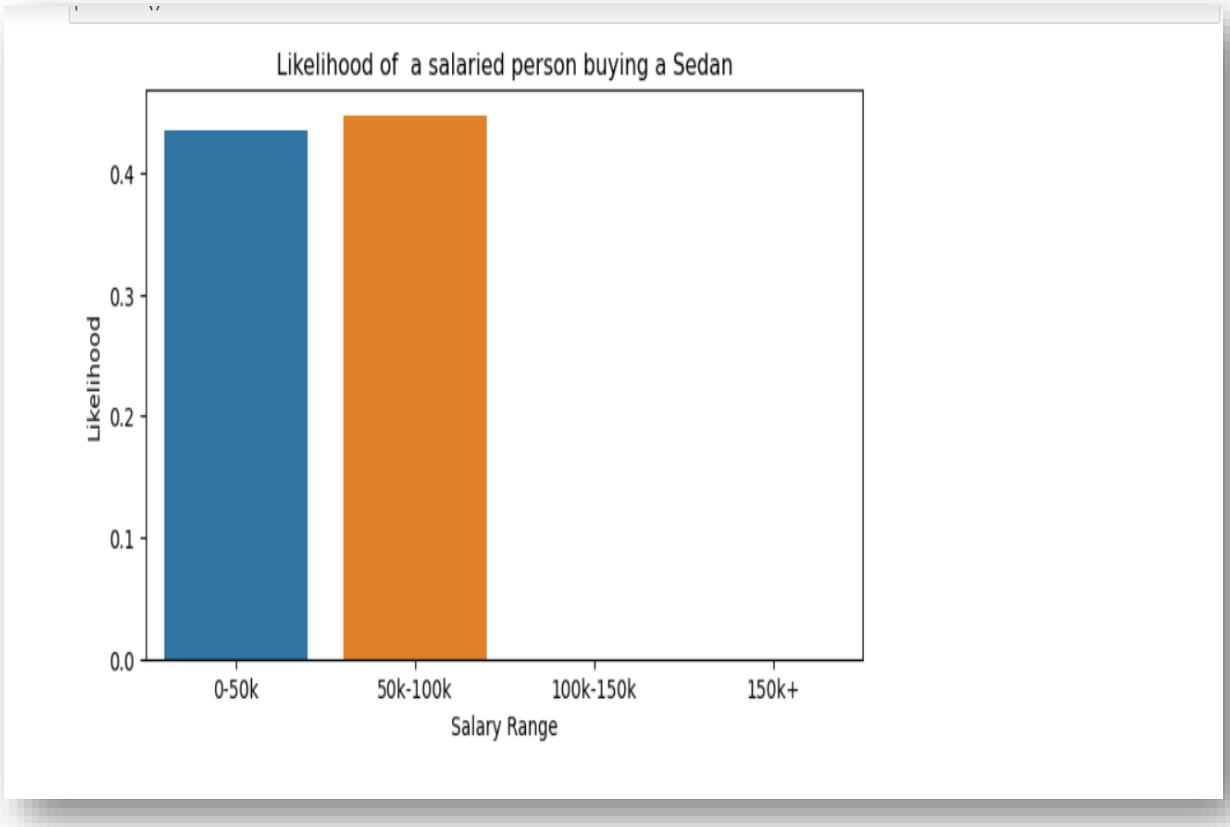
4 Key Questions :

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

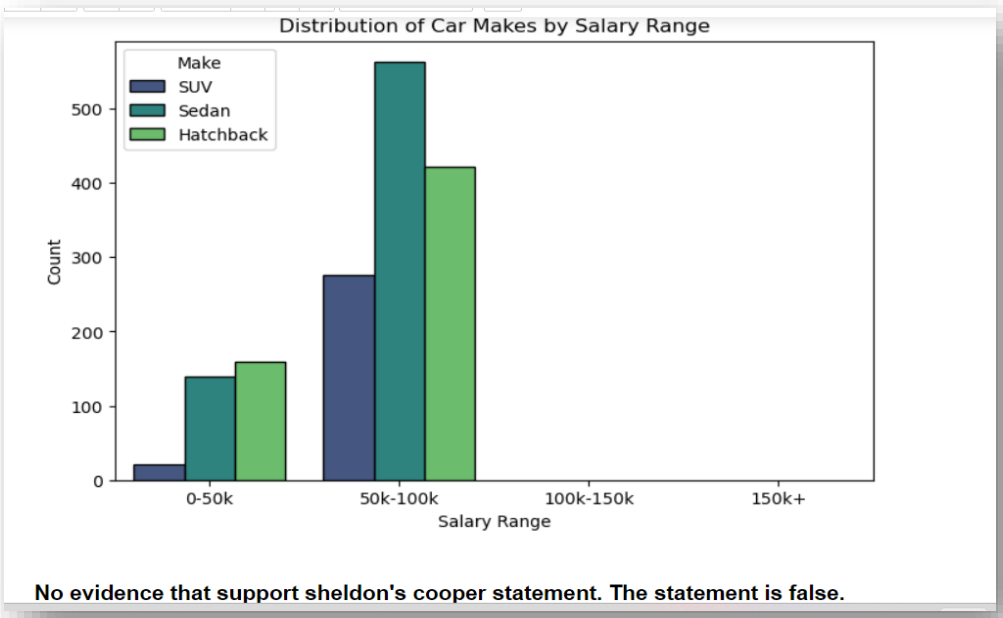


As per the countplot women tend to prefer SUV's more than Men.

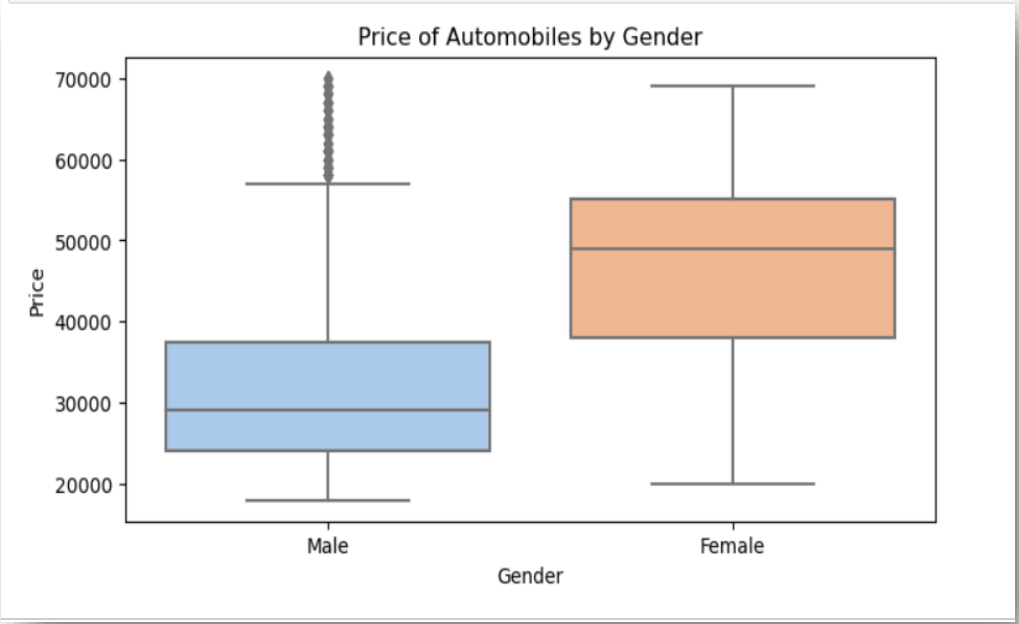
4.2 What is the likelihood of a salaried person buying a Sedan?:



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?:



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

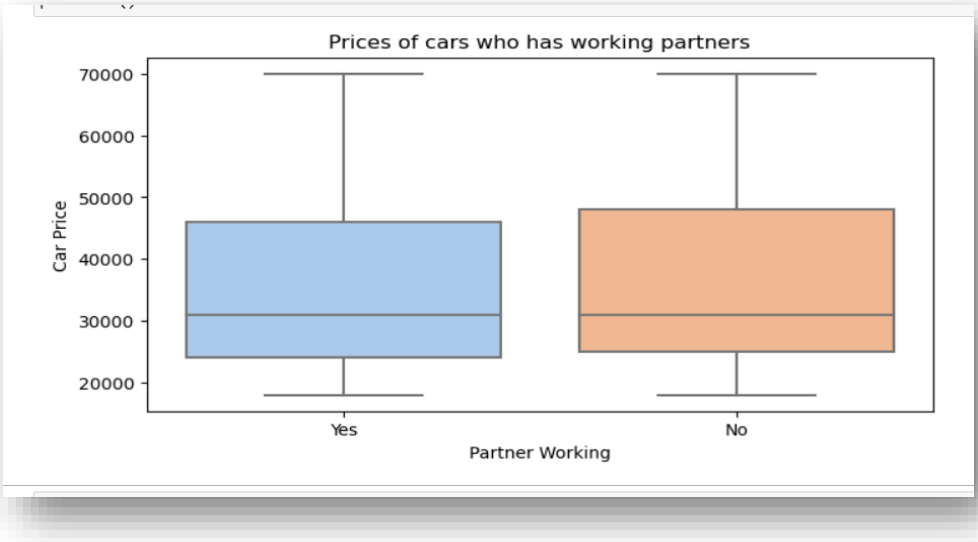


	count	mean	std	min	25%	50%	75%	max
Gender								
Female	329.0	47705.167173	11244.836378	20000.0	38000.0	49000.0	55000.0	69000.0
Male	1252.0	32416.134185	12366.253107	18000.0	23000.0	29000.0	37000.0	70000.0

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

```
Personal_loan
No      28990000
Yes     27290000
```

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



Partner_working	count	mean	std	min	25%	50%	75%	max
No	713.0	36000.000000	13817.734086	18000.0	25000.0	31000.0	48000.0	70000.0
Yes	868.0	35267.281106	13479.532555	18000.0	24000.0	31000.0	46000.0	70000.0

5. Actionable Insights & Recommendations:

AGE:

While seeing the data with regarding to age groups SUV's are preferred by the peoples in the age group between 40-50,and sedan between 25-40, Hatchback performed little bit well in the age category less than 25 so we can improve the marketing strategies to attract the audience above 25 age also.

Considering the financial preferences by different age groups.

Gender:

Analysing the choice of make between genders ,SUV was mostly preferred by women than men,Sedan was preferred mostly by men than women,Hatchback performs was poor as very less number of people only choosing hatchback.

So,look for the reason why hatchback is performing lower and derive the solution for that.

Profession:

Analysing the preferences of price ranges of Salaried and business peoples,and providing specific offerings for people on various industries

Marital-status:

Family related marketing campaigns,emphasizing safety,comfort can attaract the family audience.

Personal Loan and House Loan:

There is no big difference in choice of price ranges by people having loans,counselling on how car purchasing can fit in their financial planning makes a difference.

People having house loans chooses lower price ranges, giving them the plan to fit in their financial need makes a difference

Partner Working:

People who have dual income can be driven to purchase of higher price cars,as per now the price range of people having partner working dont make any big differnece.

Price and Make:

Can attract peoples by providing accessories,and offers on high price cars,as female tend to spend higher amount on p purchsing attract them more by giving offers,and can attract male customer by the attributes of car.

Salary range:

People who have salary ranges between 50k-100k preferred high price cars,to the people who have low salry can be guided w with the financial plans and can be driven to buy high priced cars.

As Hatchback and Sedan performing well on people with high salaries also try to attract them with accessories and royalty.

SUV can be customized and intorducing new markeitng startigies may increase the sales flow.

6. Framing Analytics Problem2:

6.1 Analyse the dataset and list down the top 5 important variables, along with the business justifications:

	Issuer	T+1_month_activity	high_networth	Transactor_revolver	avg_spends_l3m
0	Visa	0	B	T	27729
1	Visa	0	A	R	280854
2	Visa	1	C	R	70587
3	Visa	0	E	T	9156
4	Visa	1	B	T	38108
...	...	...	...	...	...
8443	Visa	0	A	T	108713
8444	Amex	0	A	T	42369
8445	Visa	0	D	T	8459
8446	Visa	1	C	T	57100
8447	Visa	0	D	T	45244

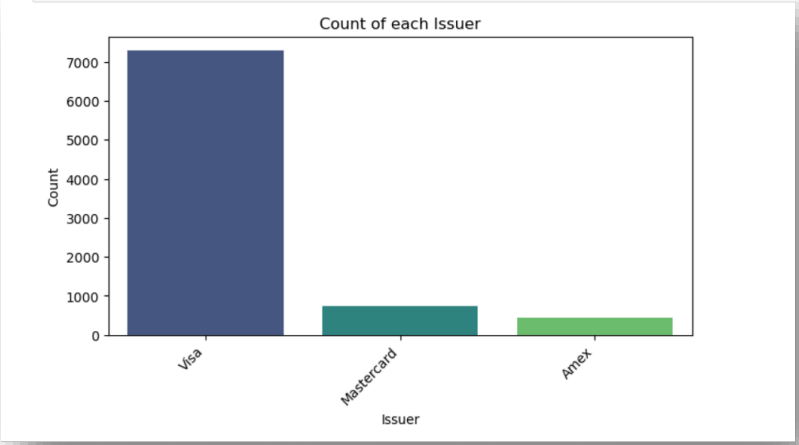
8448 rows × 5 columns

1.Card Issuer: The type of card plays a vital role in business context.

Business justification:

Analysing and Understanding the various card distributed among the customers help us to provide the suitable plans,offerings,and the marketing ideas for various card types, additional card benefits and the rewards will be additional drive for customers to be the long-lasting one.

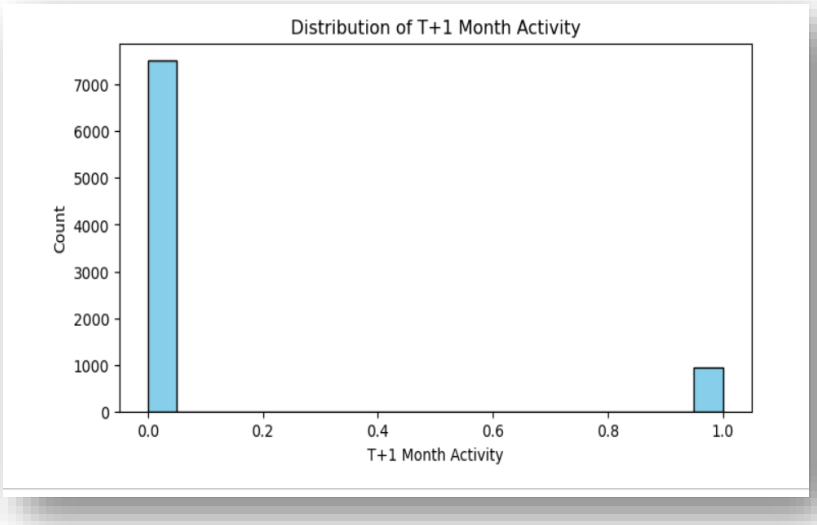
```
Issuer
Visa          7279
Mastercard     728
Amex           441
Name: count, dtype: int64
```



2.T+1_month_activity :

Business justification:

The activity at the initial stages of the issue of card is crucial,the reason i am picking this one is at the initial stages the rapport should be flawless with the customers in term of onboarding ,user ineterface,plans and involvement of user right after the card purchase will gives us the information to proceed further.Customer service and the customer loyalty are important factors during the intial stages.

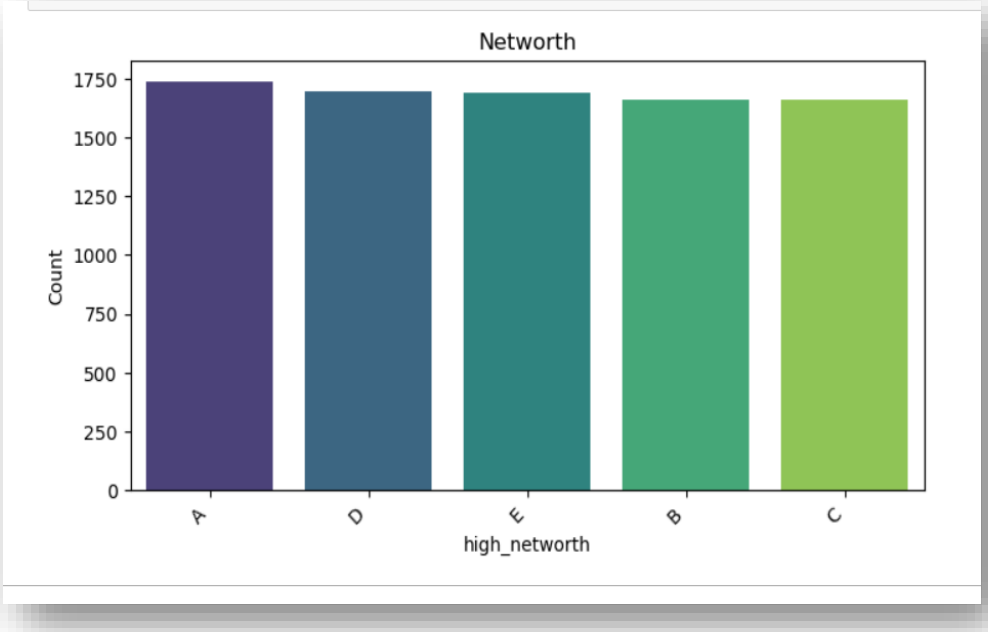


3.High-networth:

Business justification:

The networth will defines the customer about what are their potential and what are plans that can work well for them and the risk analysis in the plans,the people will higher net worth can be driven to business ideas and the growth of their welfare side by side with influence a lot in this category.

```
high_networth
A    1740
D    1696
E    1693
B    1660
C    1659
```

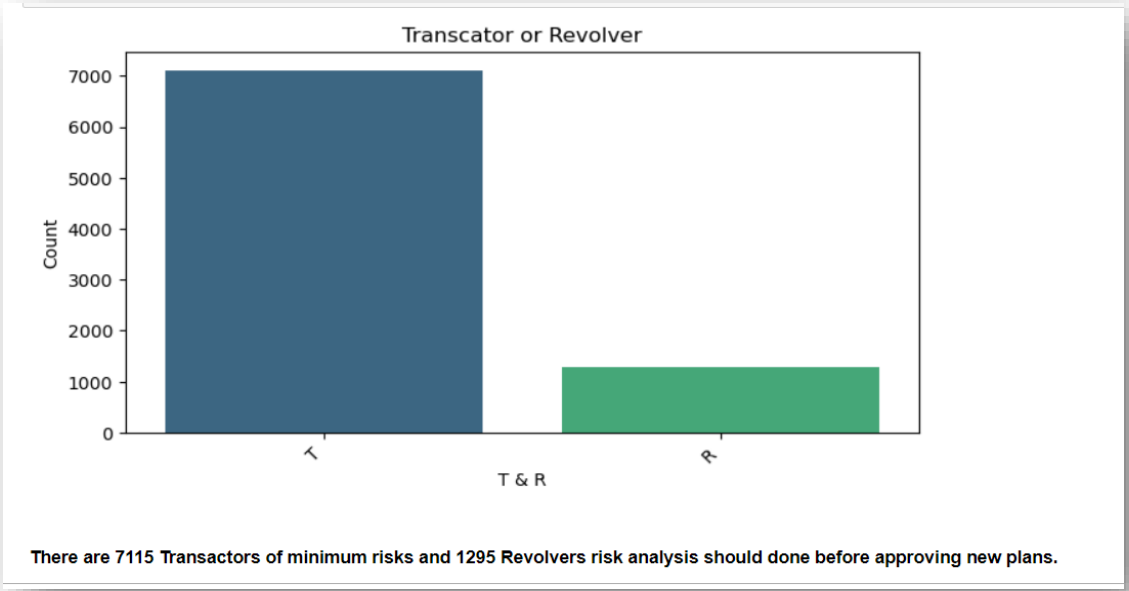


4.Transactor_revolver

Business justification:

The transactor and the revolver is important in risk analysis and management by identifying the type we can decide the suitable palns they may work and the risks in the revolver category can also be minimized.

```
Transactor_revolver
T    7115
R    1295
```



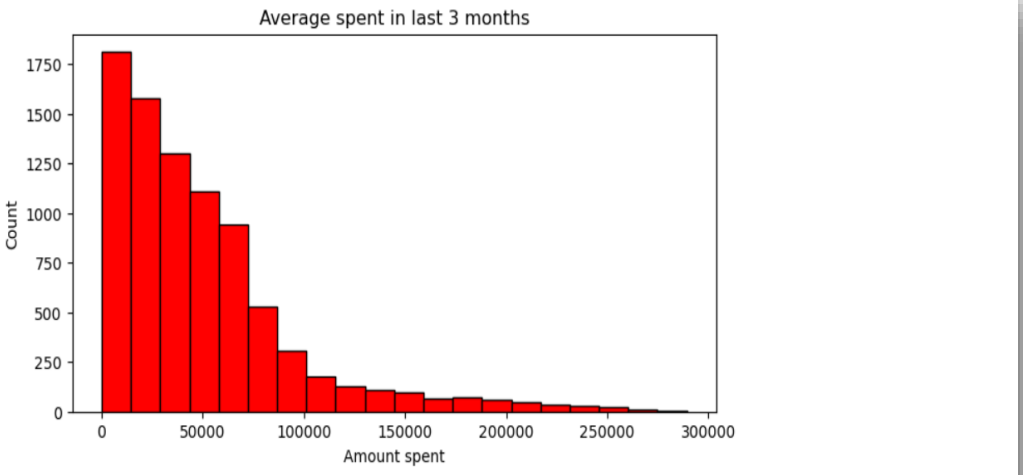
There are 7115 Transactors of minimum risks and 1295 Revolvers risk analysis should done before approving new plans.

5. Avg_spends_l3m :

Business justification:

The average amount spend in the last three months will gives us a brief overview of the potential of the customer , risk involved in the process,ensuring they having the better experience and what will be plans that suit their needs everything can be derived from it.

```
count      8448.000000
mean      49527.365530
std       46244.954836
min           0.000000
25%       17110.000000
50%       37943.000000
75%       66095.750000
max      289292.000000
```



The average amount spend by the customers in the last three months is 49527, so that we can suggest or introduce new plans according to the capabalities.The maximum amount spent in last three months is 289292.