BUSINESS REPORT

AUSTO AUTOMOBILE & Analysis of GO DIGIT BANK

BY SHUBHAM KUMAR DATE :- 8/6/23 SMDA

<u>CONTENT</u>

<u>PROBLEM 1</u>

Problem 1

A. What is the important technical information about the dataset that a database administrator would be interested in?

ANS:- This dataset (AUSTO MOTORS) has 1581 rows and 14 columns.

i.e. this dataset has 14 variables.

Let"s's check the Data type of variables: - **" .info ( )"** will give us the data type of all the variables

```
<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
```

The Above dataset can be divided into two types: -

1. Quantitative Variables: - Salary, Partner_salary, Total_salary, Price

2. Categorical Variables: - Age, Gender, Profession, Marital_status, Education, No_of_Dependents, Personal_loan, House_loan, Partner_working, Make

We can separate these variables in two types: -

1. Independent Variable :- Gender, Age, Profession, No_of_dependents, Martial_status, Education, Personal_loan, House_loan, Partner_working, Salary, Price, Make

2. Dependent Variable :- Total_salary
   Total_salary is the sum of Salary and Partner_salary

Target Variables: -

1. Make
2. Price

CHECKING FOR DATA VARAIBLE TYPE

We can see that No_of_Dependents is a categorical variable but its data type is 'int64' instead of 'object' and also salary should be float but here it is int so we should change its dtype also i.e. Salary, Partner_salary, Total_salary

```
# Lets check if all our datatypes are features/Columns
# We can see that No_of_Dependents are of of int dtype but it should be object so we should change our the dtypes
austo.No_of_Dependents=austo.No_of_Dependents.astype('int64').astype('object')
# Also lets change our salary int64 to object dtype, as it should be float
austo.Salary=austo.Salary.astype('int64').astype('float64')
austo.Total_salary=austo.Total_salary.astype('int64').astype('float64')
austo.info()
```

```
<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   No_of_Dependents  1581 non-null   object
 2   Salary            1581 non-null   float64
 3   Partner_salary    1581 non-null   float64
 4   Total_salary      1581 non-null   float64
 5   Price             1581 non-null   int64
 6   Gender            1581 non-null   object
 7   Profession        1581 non-null   object
 8   Marital_status    1581 non-null   object
 9   Education         1581 non-null   object
 10  Personal_loan     1581 non-null   object
 11  House_loan        1581 non-null   object
 12  Partner_working   1581 non-null   object
 13  Make              1581 non-null   object
dtypes: float64(3), int64(2), object(9)
memory usage: 173.0+ KB
```

In order to know more about our variable such as minimum value, maximum value, unique, frequency, etc. we use .describe( ) function.

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Age	1581.0	NaN	NaN	NaN	31.922201	8.425978	22.0	25.0	29.0	38.0	54.0
No_of_Dependents	1581.0	5.0	3.0	557.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Salary	1581.0	NaN	NaN	NaN	60392.220114	14674.825044	30000.0	51900.0	59500.0	71800.0	99300.0
Partner_salary	1581.0	NaN	NaN	NaN	18869.512966	19570.644035	0.0	0.0	24900.0	38000.0	80500.0
Total_salary	1581.0	NaN	NaN	NaN	79625.996205	25545.857768	30000.0	60500.0	78000.0	95900.0	171000.0
Price	1581.0	NaN	NaN	NaN	35597.72296	13633.636545	18000.0	25000.0	31000.0	47000.0	70000.0
Gender	1581	2	Male	1252	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Profession	1581	2	Salaried	896	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Marital_status	1581	2	Married	1443	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Education	1581	2	Post Graduate	985	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Personal_loan	1581	2	Yes	792	NaN	NaN	NaN	NaN	NaN	NaN	NaN
House_loan	1581	2	No	1054	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Partner_working	1581	2	Yes	868	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Make	1581	3	Sedan	702	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In the above we can see that we get to know about the statistical features of the numerical variables i.e. freq, mean, min, 25%, 50%, 75%, max. It also gives us the count and unique for the categorical variables.

B. Take a critical look at the data and do a preliminary analysis of the variables. Do a quality check of the data so that the variables are consistent. Are there any discrepancies present in the data? If yes, perform preliminary treatment of data.

Ans:- In a Preliminary Analysis of the variables we will first look for the duplicates.

In order to check number of duplicated rows can be checked using " .duplicated ( ) " function

```
# Check for duplicate data
dups=austo.duplicated()
print('Number of duplicate rows = %d' % (dups.sum()))
```

```
Number of duplicate rows = 0
```

In above code we have checked for duplicates. So there are 0 duplicates.

If there had been duplicates we should have dropped them using drop ()

We can also divide our dataset into numerical datatype and categorical datatype which will be useful while doing Preliminary analysis of our dataset of so many variables

In Order to divide the dataset into numerical and categorical dataset we will use **" .select_dtypes() "**

```
num=austo.select_dtypes(['int64','float64'])
cat=austo.select_dtypes('object')
```

The above code will divide the dataset in two types i.e. Numerical dtypes as **"num"** & Categorical dtypes as **"cat"**

In preliminary analysis of the variables is to edit the data to prepare it for further analysis, describe the key features of the data, and summarize the results.

<u># Checking for Outliers</u>

Outliers are only checked for the Numerical Datatypes
In order to check for outliers of a numerical data type we will use boxplot

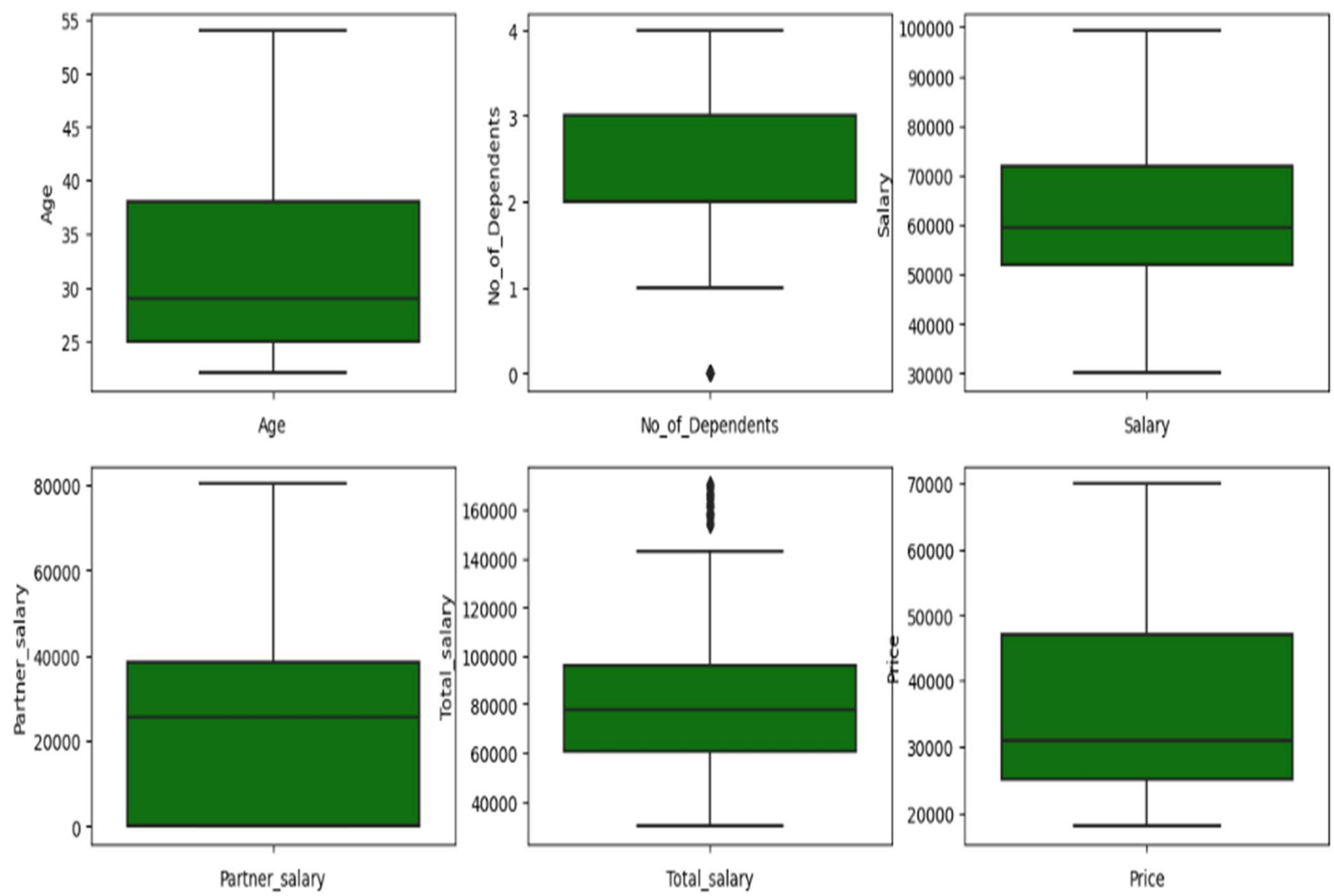
In order to draw a boxplot () of a numerical data type we define a user defined function

```
def drawboxplot():
    col=len(num.columns)
    plt.figure(figsize=(15,7))
    for i in range(0,col):
        plt.subplot(2,3,i+1)
        sns.boxplot(data=num, y=num.columns[i],color='green')
        plt.xlabel(num.columns[i])
```

Every time we call a above user define function, it will draw the boxplot of all the 'num' dataset

From these boxplots we can see if our dataset has the outliers or not.

```
drawboxplot()
```



Looking at the above Boxplot we can see that “Age”, “Salary”, “Partner salary”, “Price” doesn’t have any outliers

There are Outliers in two variables “No of Dependents” and “Total salary”

We need to analyze these two variables: -

No_of_dependents :- We can see from the boxplot and also from the .min( ) function that the outliers is at 0 which is totally fine. As its not mandatory for a person to have a dependent.

```
num.No_of_Dependents.min()
```

```
0
```

Total_salary :- Total_salary is the sum of Salary & Partner Salary so it can have outliers. There are no outliers in either of the two independent variable we can leave it as it is.

TREATING BAD VALUES

Bad values are things which are not useful data points in a dataset or some data points are unusual than the other data points.

We can check unique features of categorical data field using **".unique( )"** as per domain and general knowledge we can check if the unique values entered in categorical data field are correct or there are some incorrect entries

```
In [156]: print('Unique values in Gender are ',cat.Gender.unique())
          print('Unique values in Profession are ',cat.Profession.unique())
          print('Unique values in Marital status are ',cat.Marital_status.unique())
          print('Unique values in Education are ',cat.Education.unique())
          print('Unique values in Gender are ',cat.Personal_loan.unique())
          print('Unique values in Gender are ',cat.House_loan.unique())
          print('Unique values in Gender are ',cat.Partner_working.unique())
          print('Unique values in Gender are ',cat.Make.unique())

          Unique values in Gender are  ['Male' 'Femal' 'Female' nan 'Femle']
          Unique values in Profession are  ['Business' 'Salaried']
          Unique values in Marital status are  ['Married' 'Single']
          Unique values in Education are  ['Post Graduate' 'Graduate']
          Unique values in Gender are  ['No' 'Yes']
          Unique values in Gender are  ['No' 'Yes']
          Unique values in Gender are  ['Yes' 'No']
          Unique values in Gender are  ['SUV' 'Sedan' 'Hatchback']
```

From the above code we can see that Gender column has incorrect values entered which can be due to typing error, so we need to treat this column

In order to treat this column i.e. replace incorrect values with correct one i.e. Female. In order to replace these incorrect values with correct one we use **".replace( )"** function

```
cat.Gender.replace(['Femal','Femle'],'Female',inplace=True)
```

```
cat.Gender.unique()

array(['Male', 'Female', nan], dtype=object)
```

We can also check if the incorrect values have been changed or not.

TREATING MISSING VALUES

Missing values are nothing but the null entries in a dataset so we will check for all the null entries in a data set
Null entries in a dataset can be checked using **".isnull ( ) "**

```
cat.isnull().sum()[cat.isnull().sum()>0]
```

```
Gender    53
dtype: int64
```

```
num.isnull().sum()[num.isnull().sum()>0]
```

```
Partner_salary    106
dtype: int64
```

Above code will show only those columns which has null values in it.

We can see that there are null values in each num dataset aswell as cat dataset i.e. Gender & Partner_Salary

Let's treat these null values: -

1. **Gender of a person cannot be null it could either be Male or Female in general so we will replace these null values with the most frequent occurring element i.e. mode of a categorical column.**

In order to replace these null values we will use fillna( ) function.

```
cat.Gender.fillna(cat.Gender.mode()[0],inplace=True)
```

```
cat.Gender.unique()
```

```
array(['Male', 'Female'], dtype=object)
```

We can also check if the null values of the column have been changed or not.

2. **Partner_salary of a person can be null but we can replace these with 0 to get its statistical features. Also we can check if all the null values have been treated or not.**

```
num.Partner_salary.fillna(0,inplace=True)
```

```
num.Partner_salary.isnull().sum()
```

```
0
```

Important Note: - We are not doing Scaling or Transformation as we don't have to give this dataset to the Model

Before proceeding to next step we need to merge/join our dataset back to its original form. We will use pd.concat( ) function.

```
austo=pd.concat([num,cat],axis=1,join='inner')
```

```
austo.shape
```

```
(1581, 14)
```

```
austo.head()
```

	Age	No_of_Dependents	Salary	Partner_salary	Total_salary	Price	Gender	Profession	Marital_status	Education	Personal_loan	House_loan	Partner_work
0	53	4	99300	70700.0	170000	61000	Male	Business	Married	Post Graduate	No	No	
1	53	4	95500	70300.0	165800	61000	Female	Salaried	Married	Post Graduate	Yes	No	
2	53	3	97300	60700.0	158000	57000	Female	Salaried	Married	Post Graduate	No	No	
3	53	2	72500	70300.0	142800	61000	Female	Salaried	Married	Graduate	Yes	No	
4	53	3	79700	60200.0	139900	57000	Male	Salaried	Married	Post Graduate	No	No	

We can see that our dataset has been merged and all over numerical data type variables comes first followed by the categorical data types variables

We can also check our merged dataset if its same or any different from original dataset

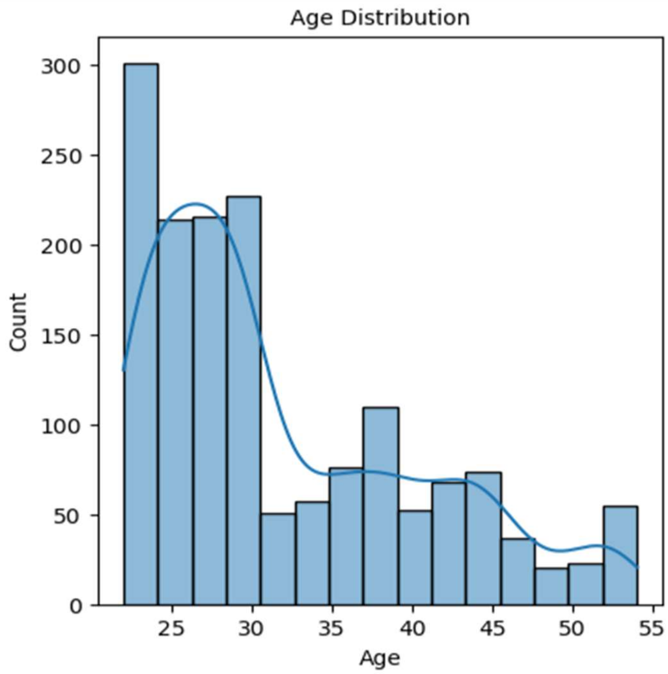
In order to check we used .shape which would give us the size of our dataset which is same as our original dataset

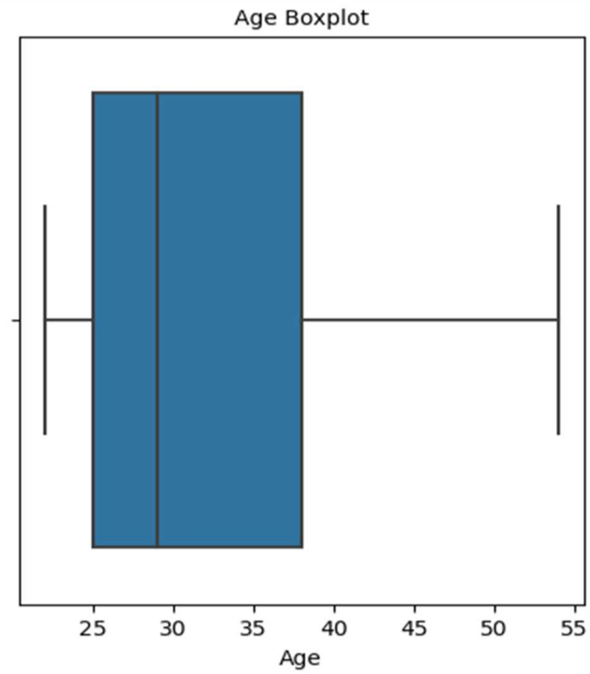
We can also check our first 5 rows of dataset using .head( ) function

C. **Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.**

Ans :- Let's check each features separately using proper visualizations

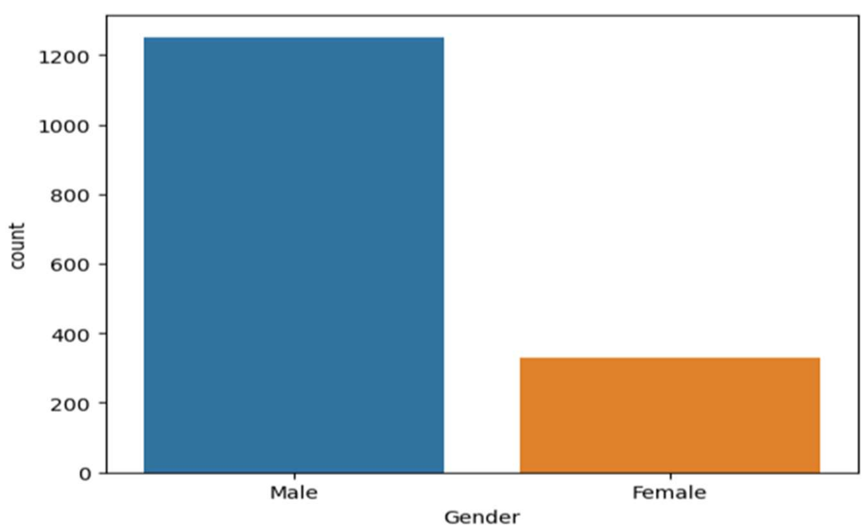
1. Age :-





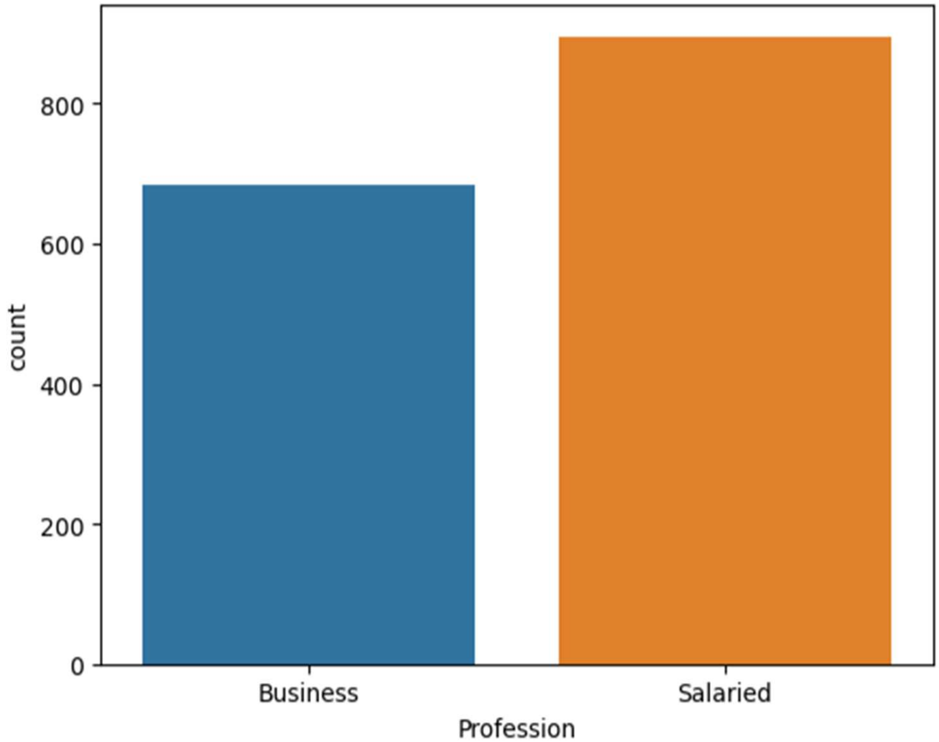
- **From above Plot we can see that the large number of people having cars are between age group of (22-30) years.**
- **There are average number of people having cars are between age group of (31-45) years**
- **There are few number of people having cars between age group of (46-54) years.**
- **We can further analyze that what type of cars are being most preferred by each age group.**

2. Gender: -



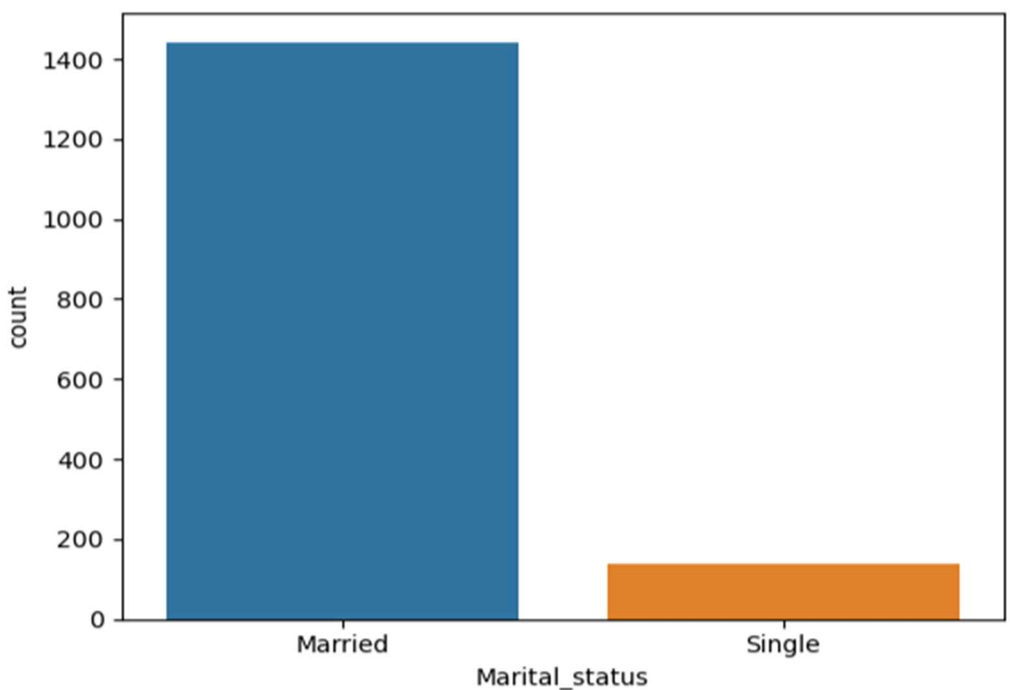
- **We can see that car owned by Males are more in number compared to Females.**
- **We can look further for which cars are more owned by Male or Female.**
- **We can also look for what does Female prefer while purchasing a car so that we could work on it and increase number of car owned by Female**

3. **Profession**



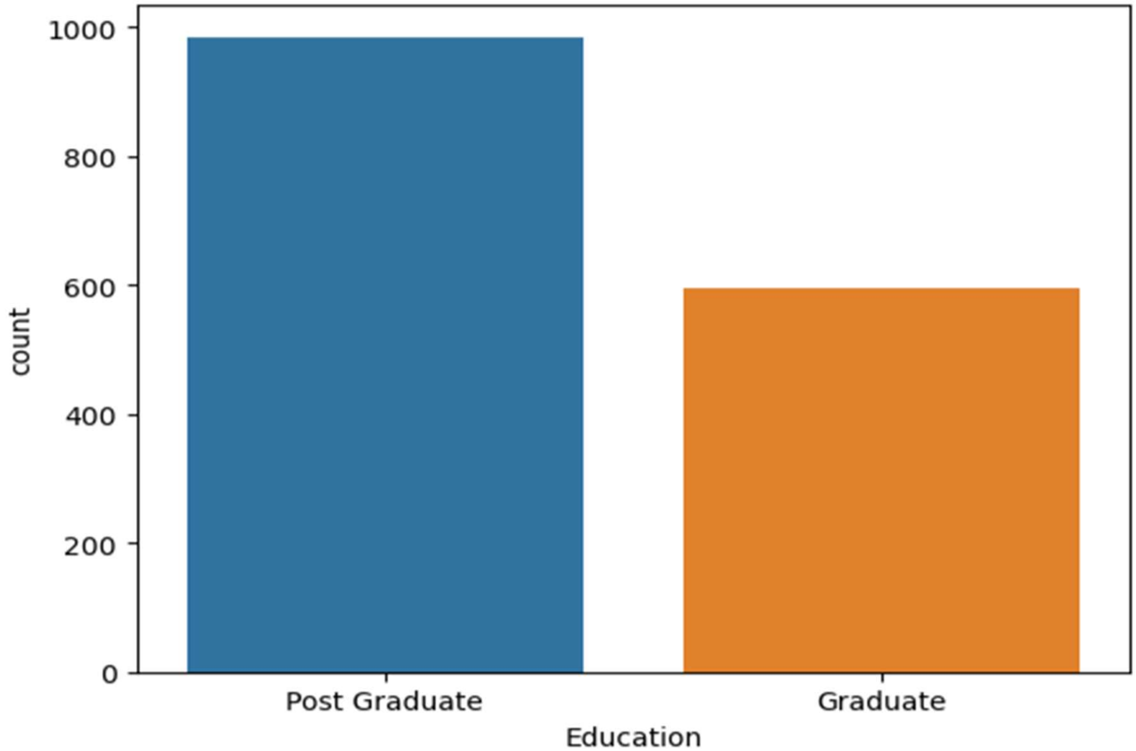
- **We can see that car owned by Salaried People are more than Car Owned by Business People**
- **We can look for what does the people doing Business prefers while buying a car**

4. **Marital Status**



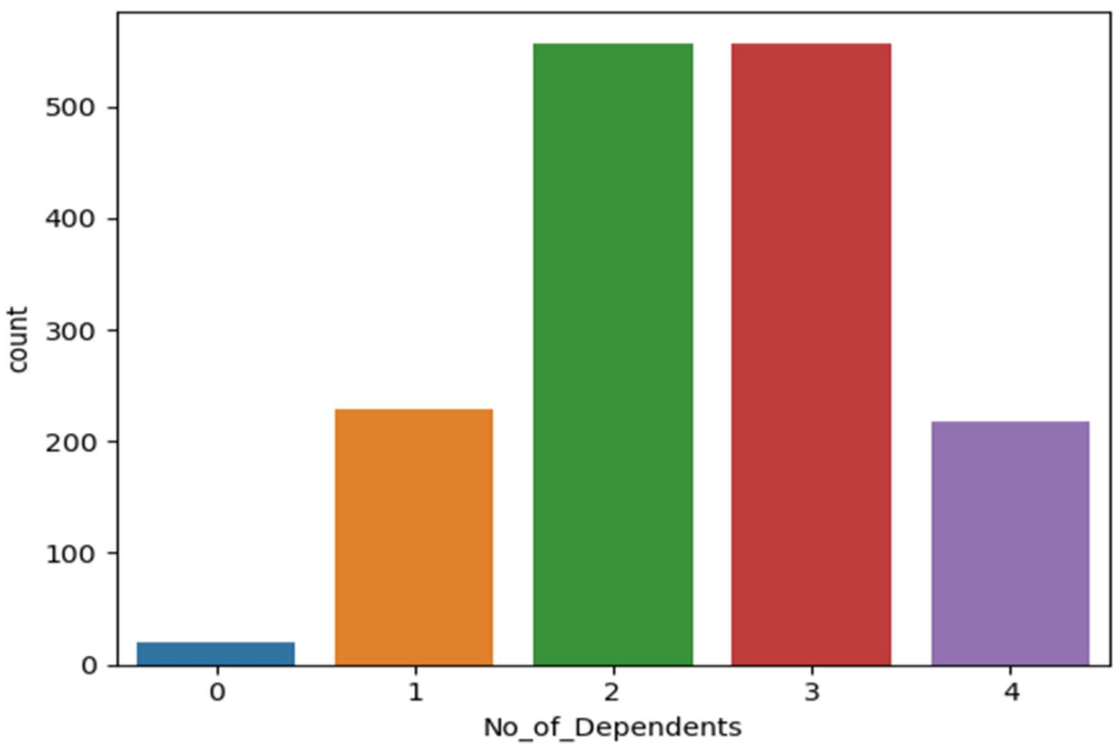
- **We can see that car owned by Married People are more than Single People**

5. **Education**



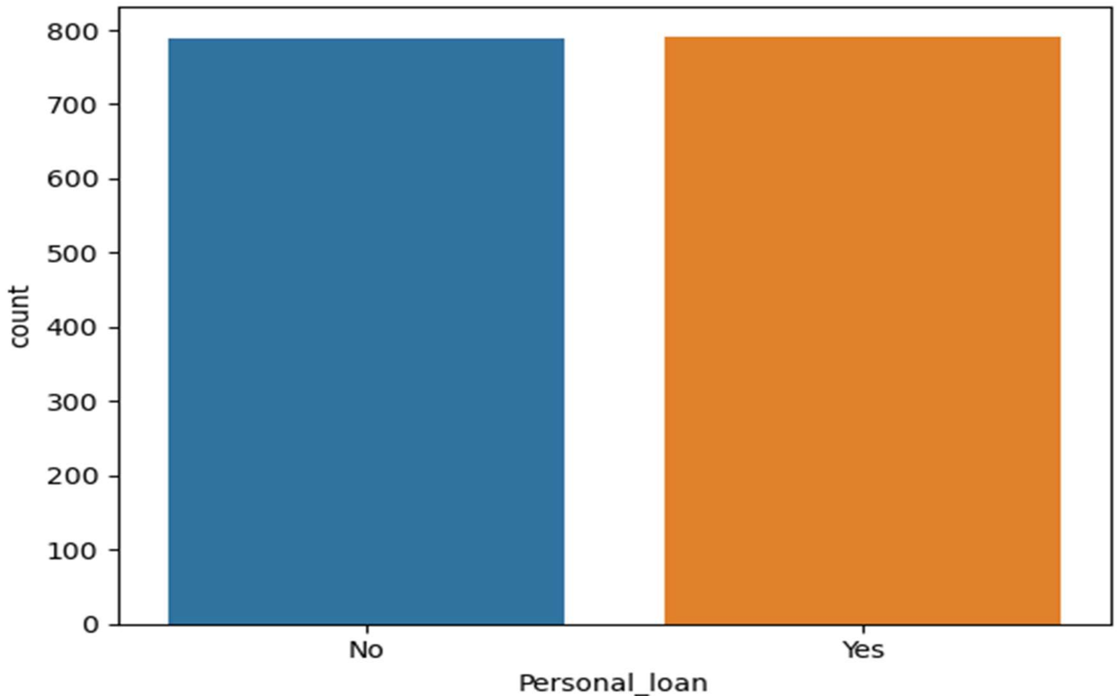
- **We can see that car owned by people completed their Post Graduate are more than Graduate**

6. **Number of Dependents**



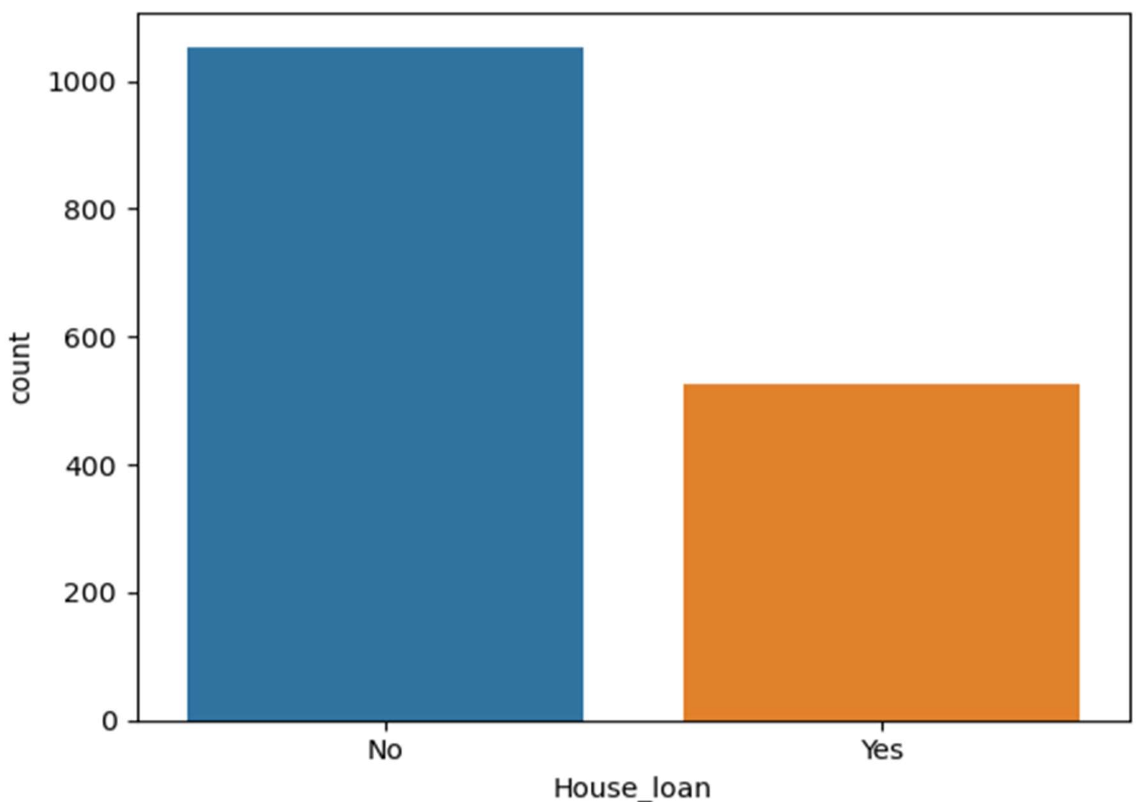
- **We can see that car owned by people who having 2 or 3 dependents are most**
- **There are very average number of people with cars who are having 1 or 4 dependents**

7. **Personal loan**



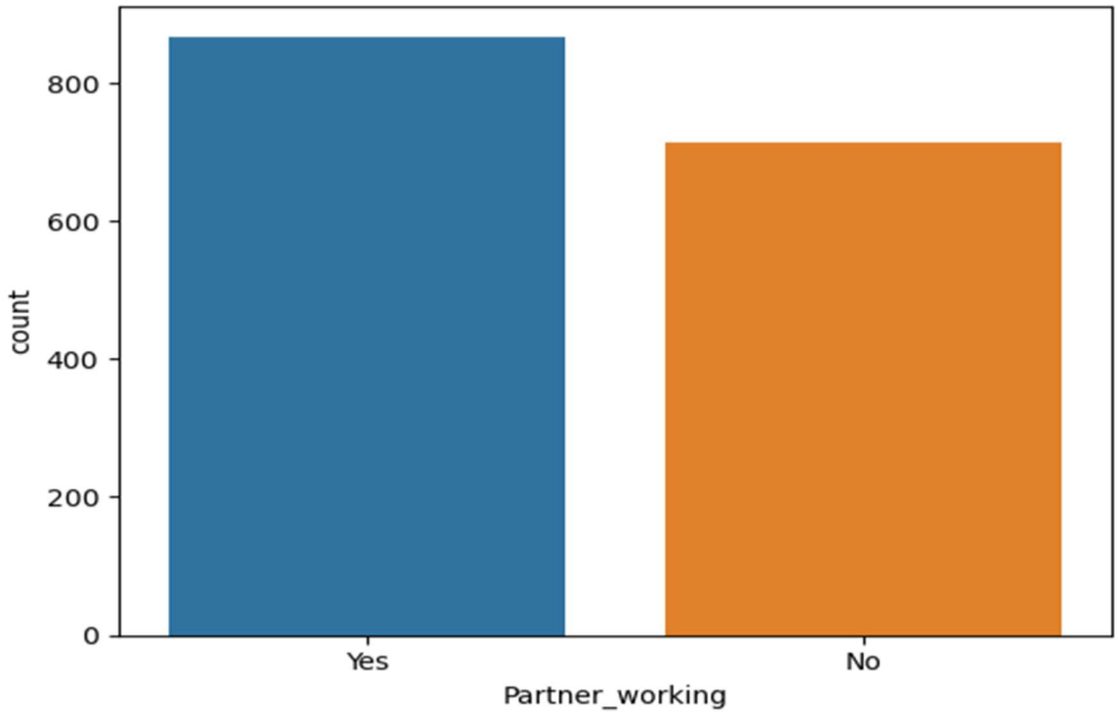
- **We can see that car owned by people who having personal loan or not are almost equal.**

8. **House loan**



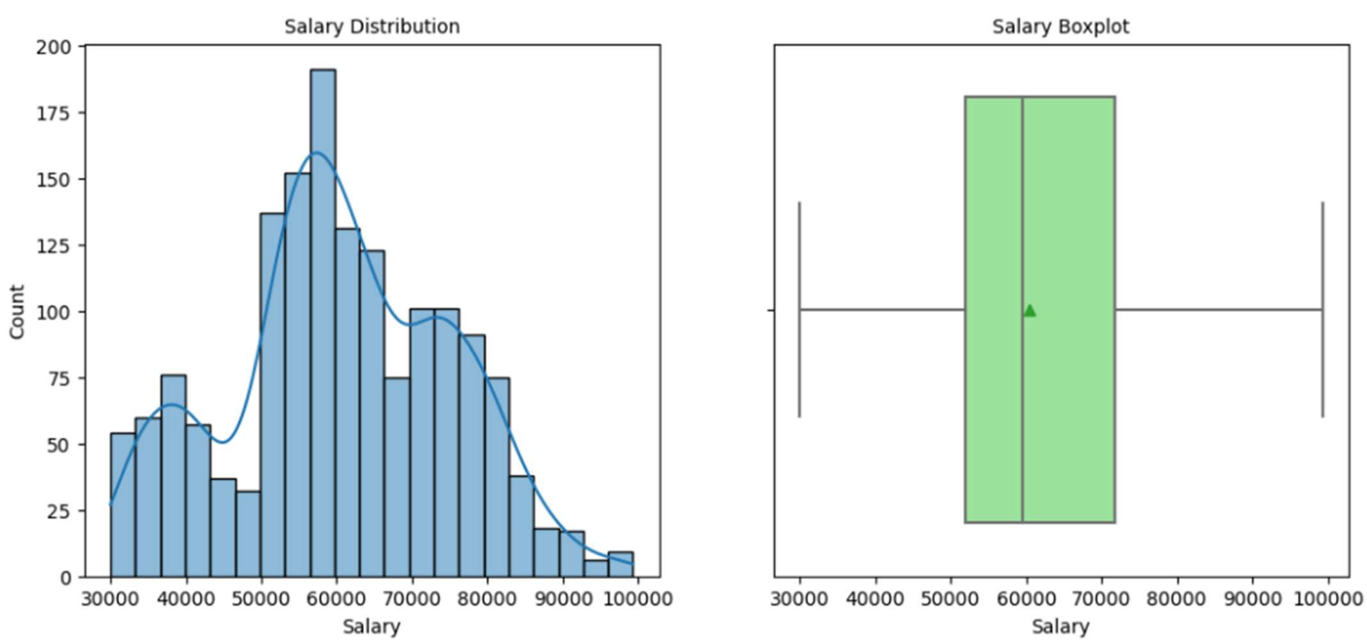
- **We can see that car owned by people who are not having House loan are more than people having house loan**

9. **Partner Working**



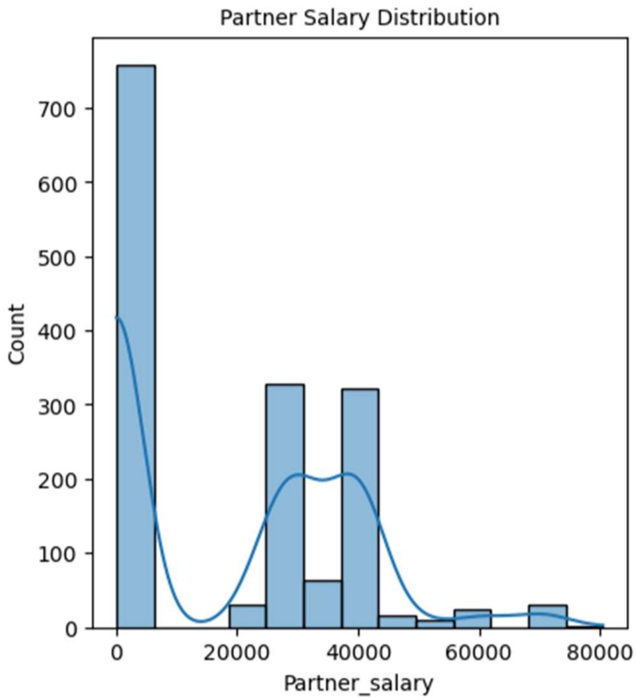
- **We can see that car owned by people whose partner are working are more than whose partner are not working**

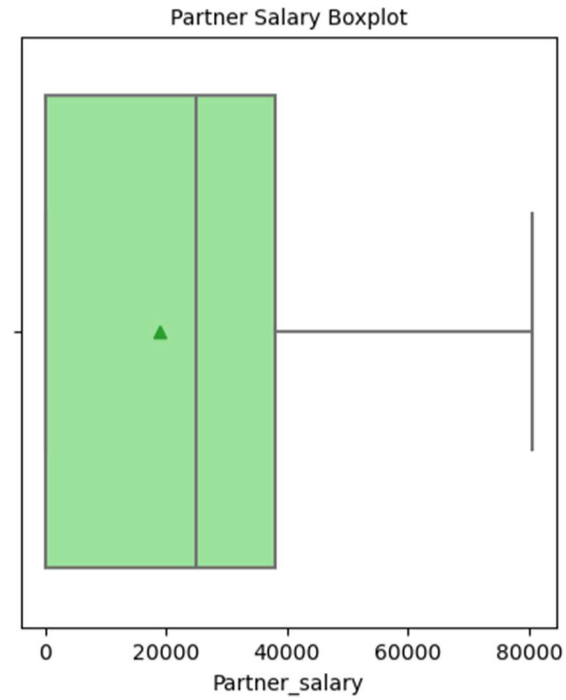
10. **Salary**



- **We can see that car owned by people whose Salary is in between (50000-72000) is most i.e. 50%**
- **From the boxplot we can see that our mean is greater than median which means Salary is normally Right/Positive skewed.**

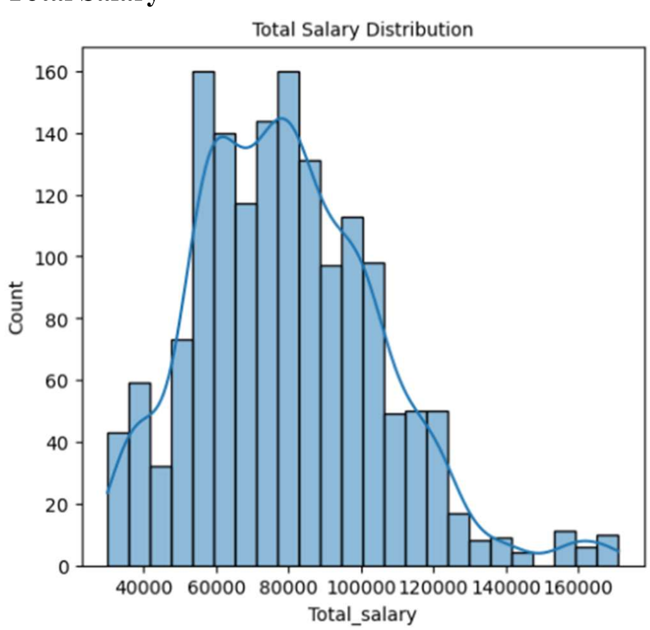
11. Partner Salary

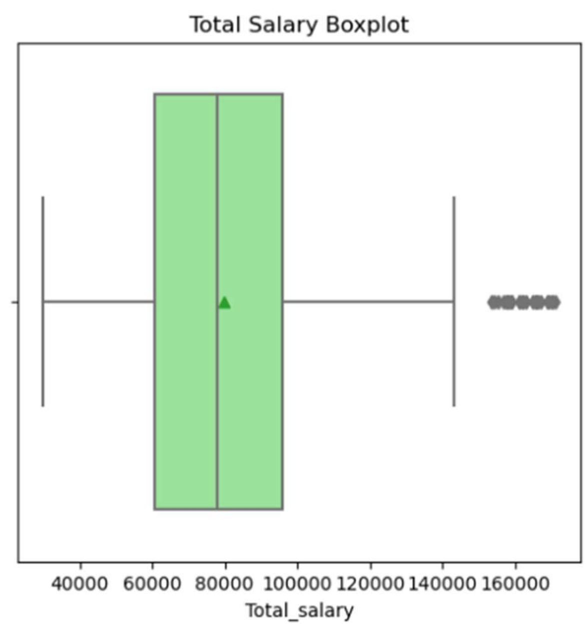




- **We can see that car owned by people whose Partner salary is in between (0-38000) is most i.e. 50%**
- **From the boxplot we can see that our mean is less than median which means Salary is slightly left/Negative skewed.**

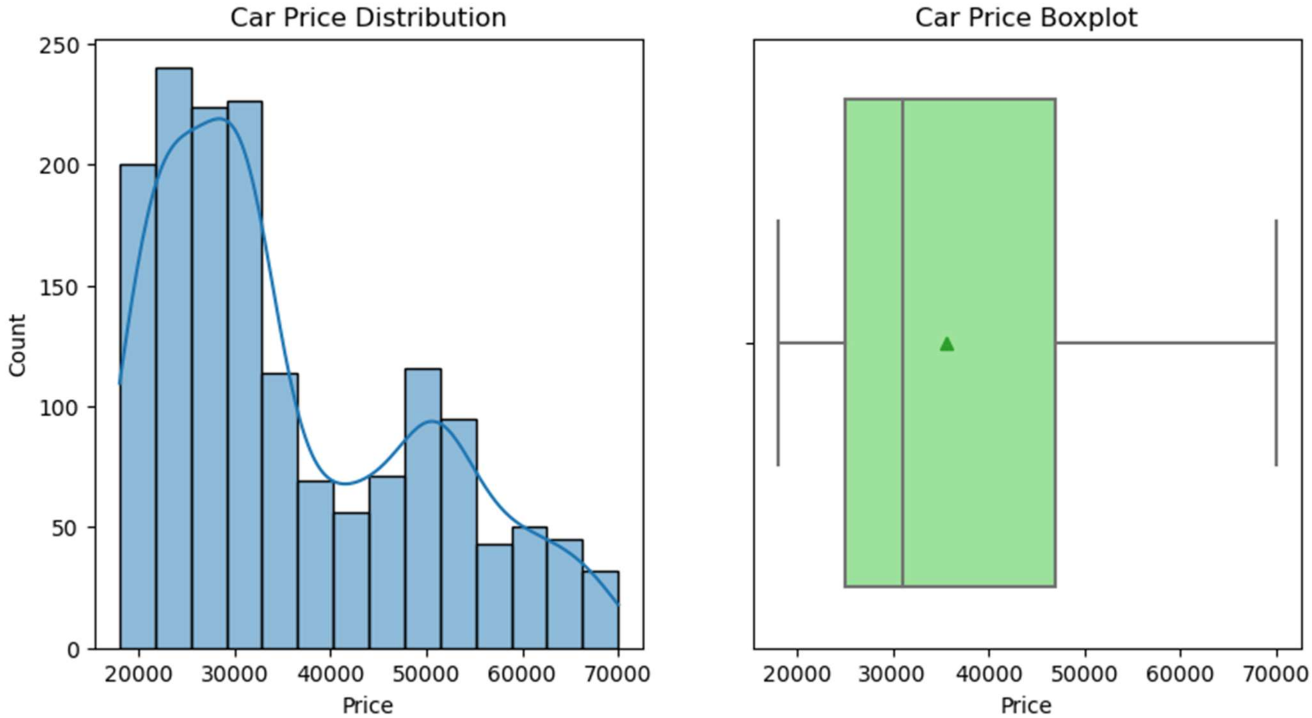
12. Total Salary





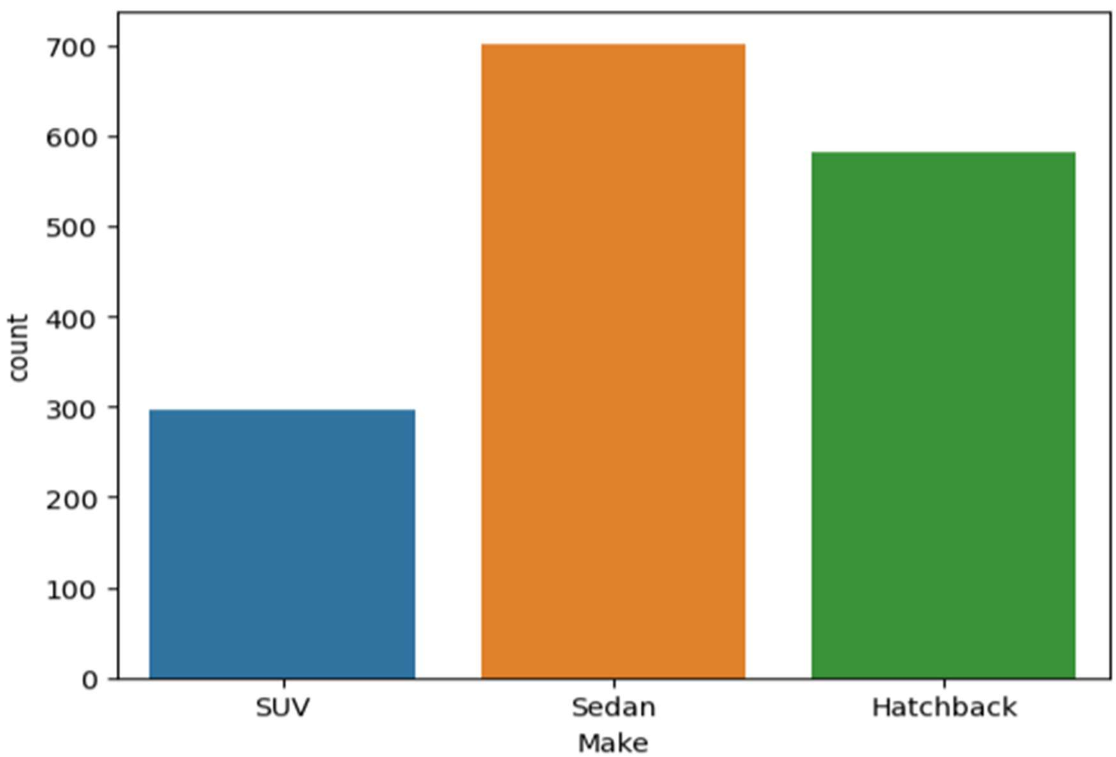
- **We can see that car owned by people whose Total salary is in between (60000-95000) is most which is 50%.**
- **From the boxplot we can see that our mean is greater than median which means Salary is slightly Right/Positive skewed.**

13. Price



- **We can see that car owned by people whose price is in between (25000-47000) are most which is 50%**

14. Make

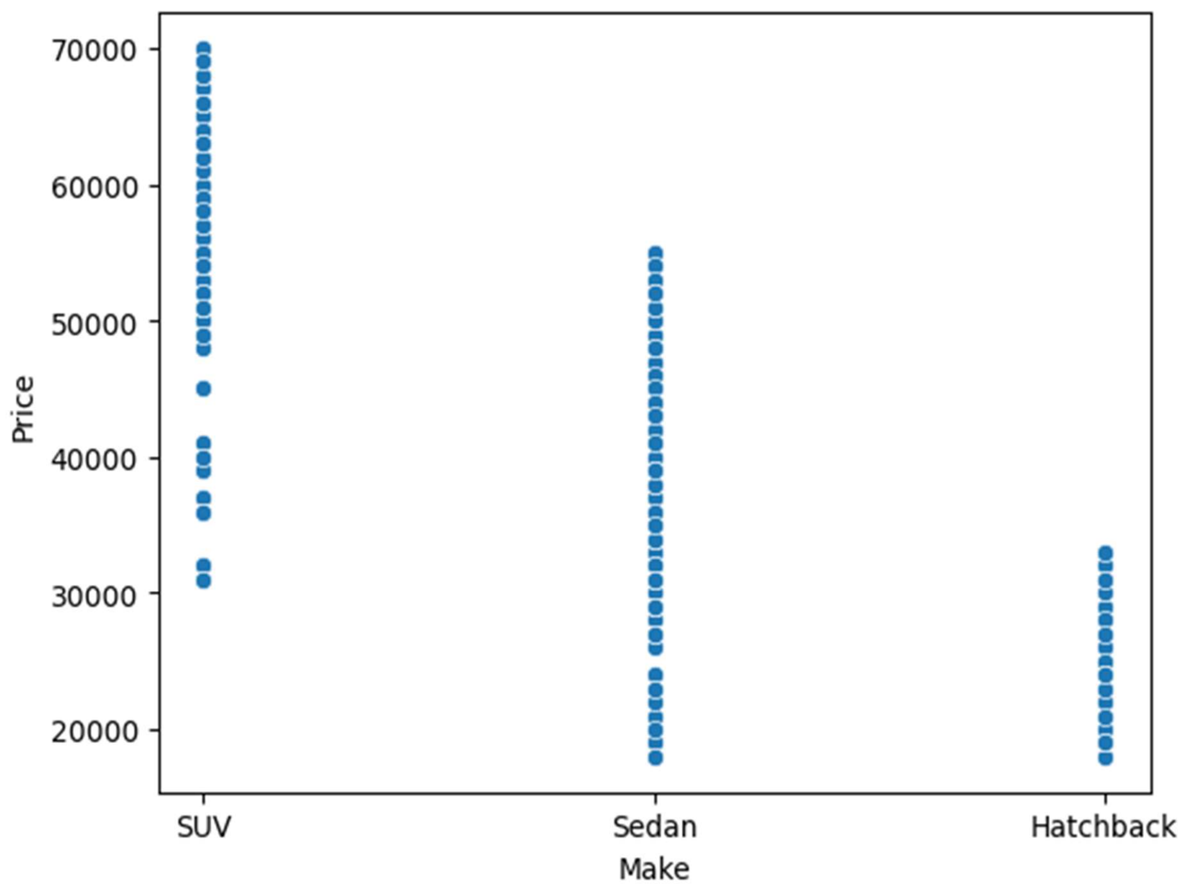


- **We can see that car owned by people are most Sedan followed by Hatchback**
- **SUV cars are least Owned**

D. Understanding the relationships among the variables in the dataset is crucial for every analytical project. Perform analysis on the data fields to gain deeper insights. Comment on your understanding of the data.

First Check the relation between our target variable/feature i.e. Make & Price

Here we will get to know about the Price of each Type of Cars



- **We can see that the SUV cars are costliest among all three cars**
- **Hatchbacks are least expensive car among all three cars**
- **Sedan cars are of average price.**
- **Car Price Ranges from 20000 – 70000**
- **Sedan Cars are most in number**
- **SUV Cars are least in number**

Now let check the relationship between Target variable and all other variables

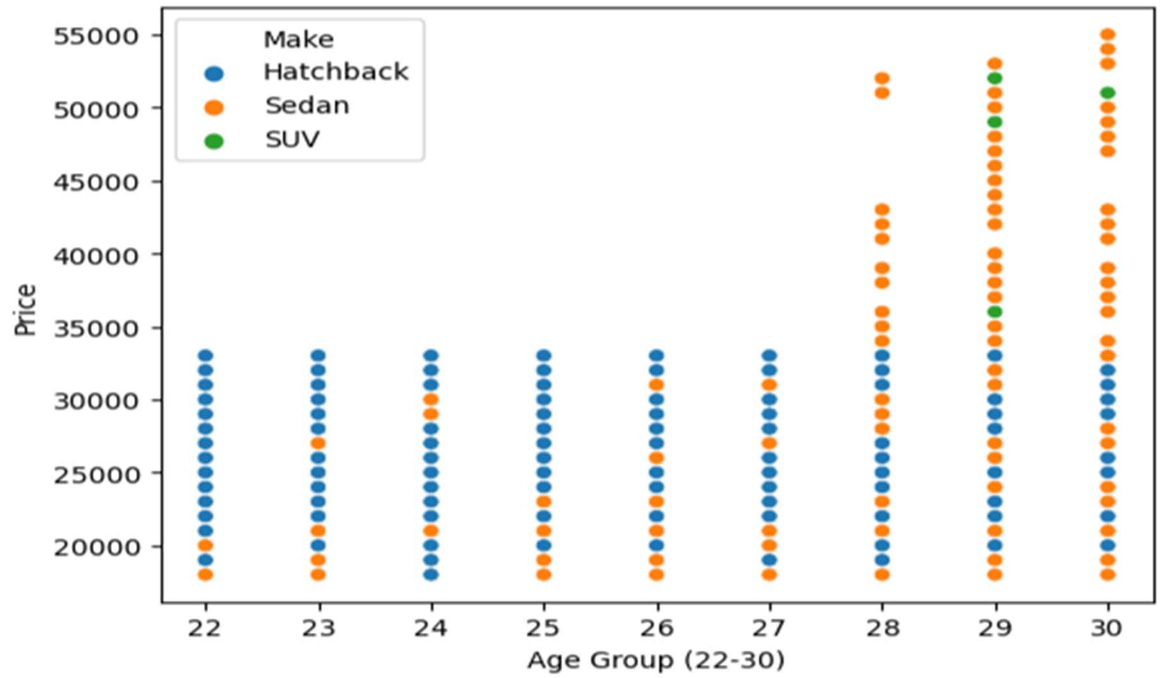
1. **AGE**

We can divide Age group into 3 categories as (22-30) (31-45) & (40-54) as it would be convenient to study this way.

Age (22-30)

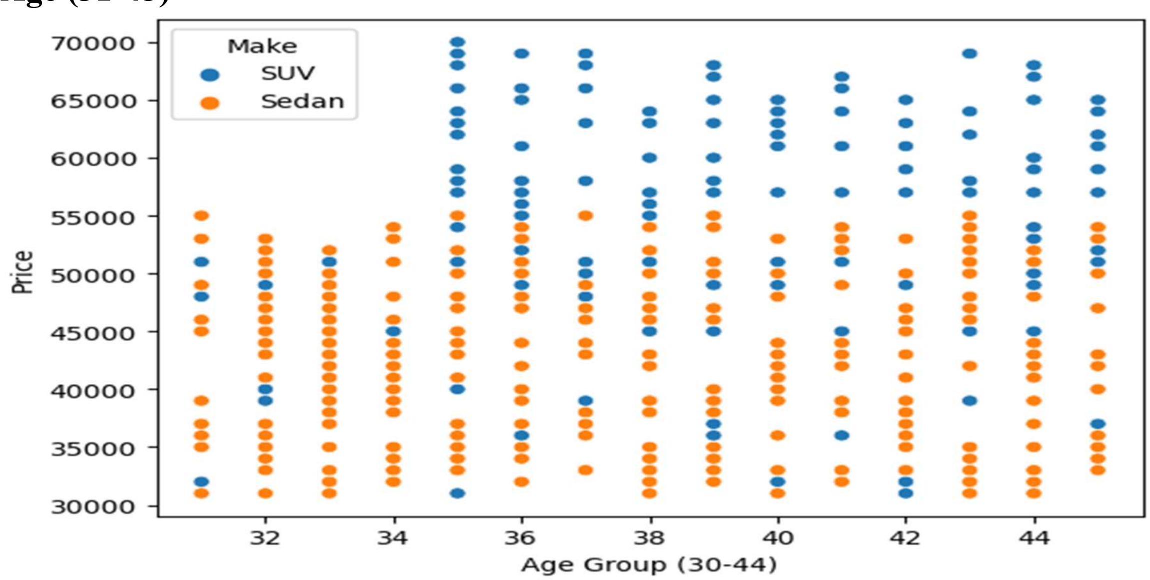
```
austo_age_upto30.Make.value_counts()

Hatchback     582
Sedan         367
SUV             9
Name: Make, dtype: int64
```



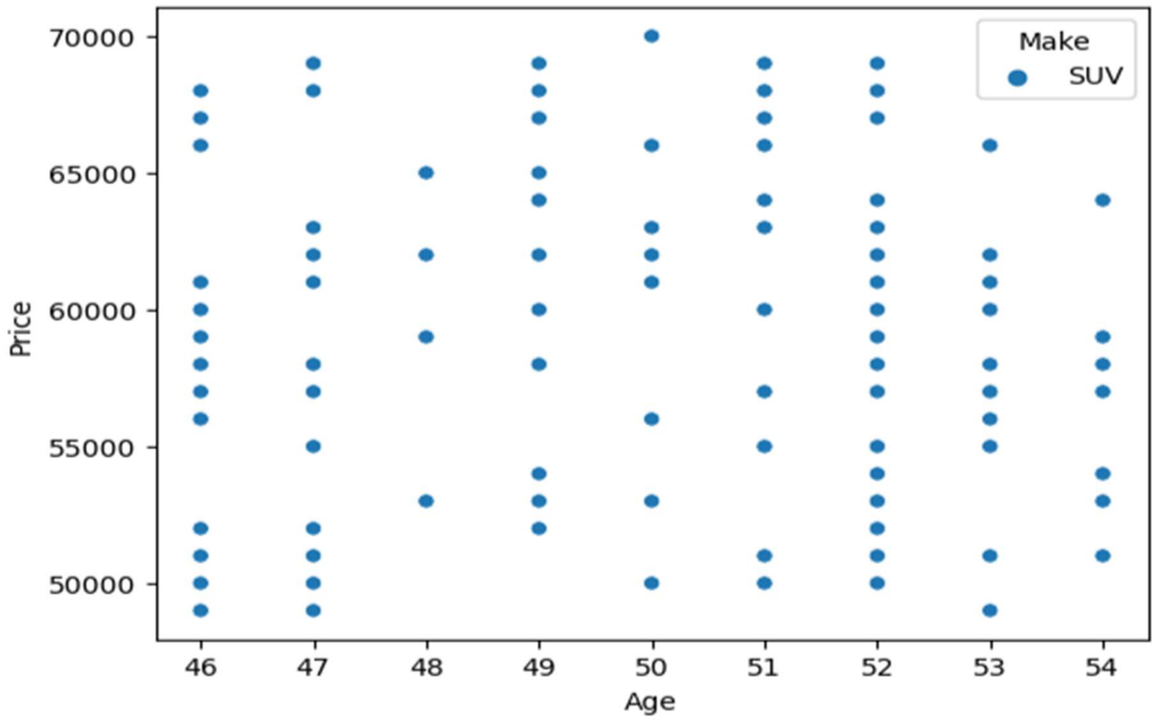
- **We can see from above plot that people of Age group (22-30) prefers Hatchback most**
- **Sedan cars more compared to SUV which can be due to low Total salary**
- **Also people from Age (22-28) only prefers Hatchback & Sedan**
- **There are very SUV cars in Age (29-30) i.e. 9 numbers**

Age (31-45)



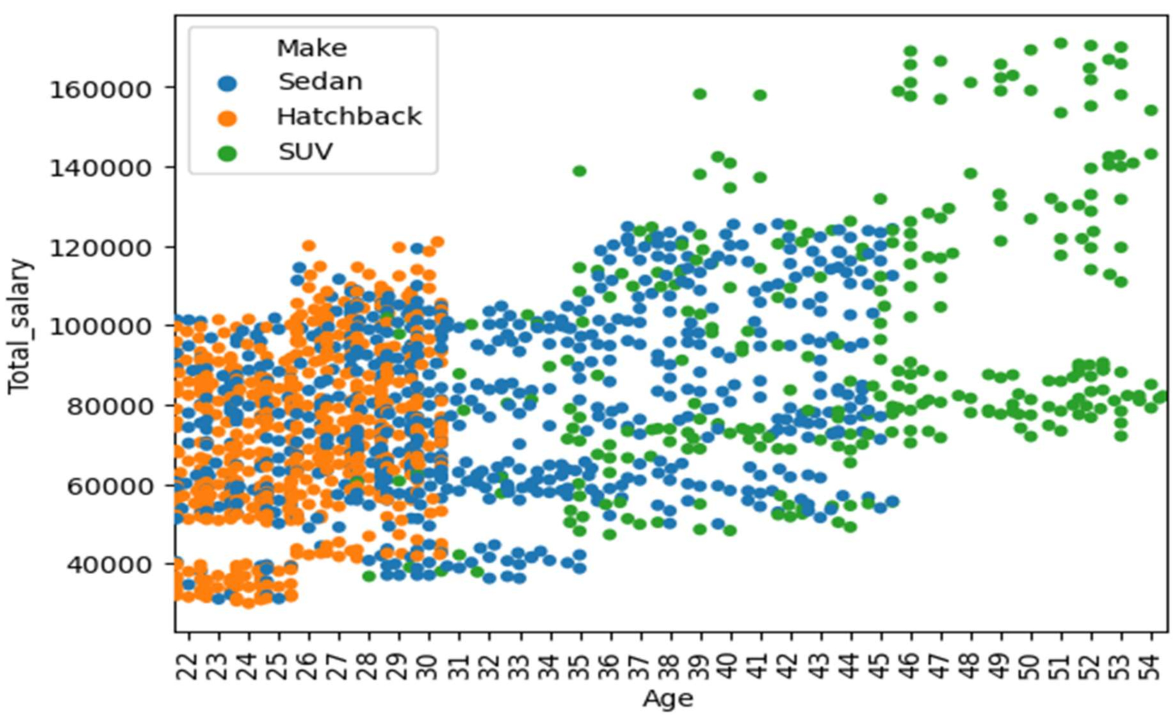
- **There are No Hatchback cars in Age (31-45)**
- **In Age (31-45) people only prefers Sedan & SUV Cars**

Age (46-54)



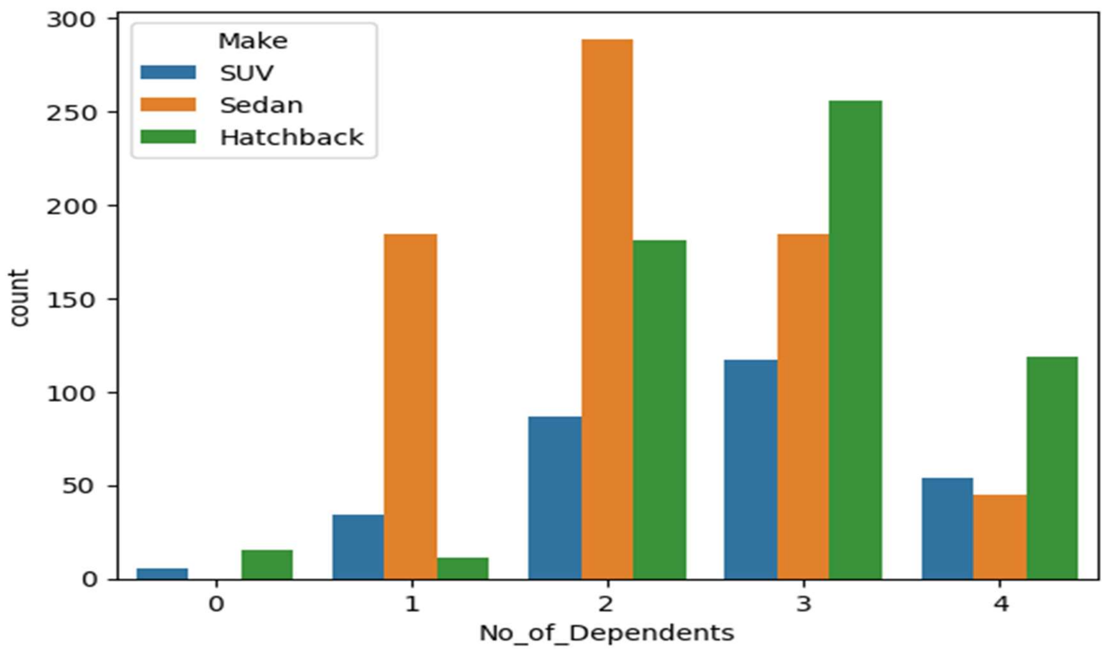
- **In Age (46-54) people only buys SUV Cars**

Let's compare Age & Total Salary with Make



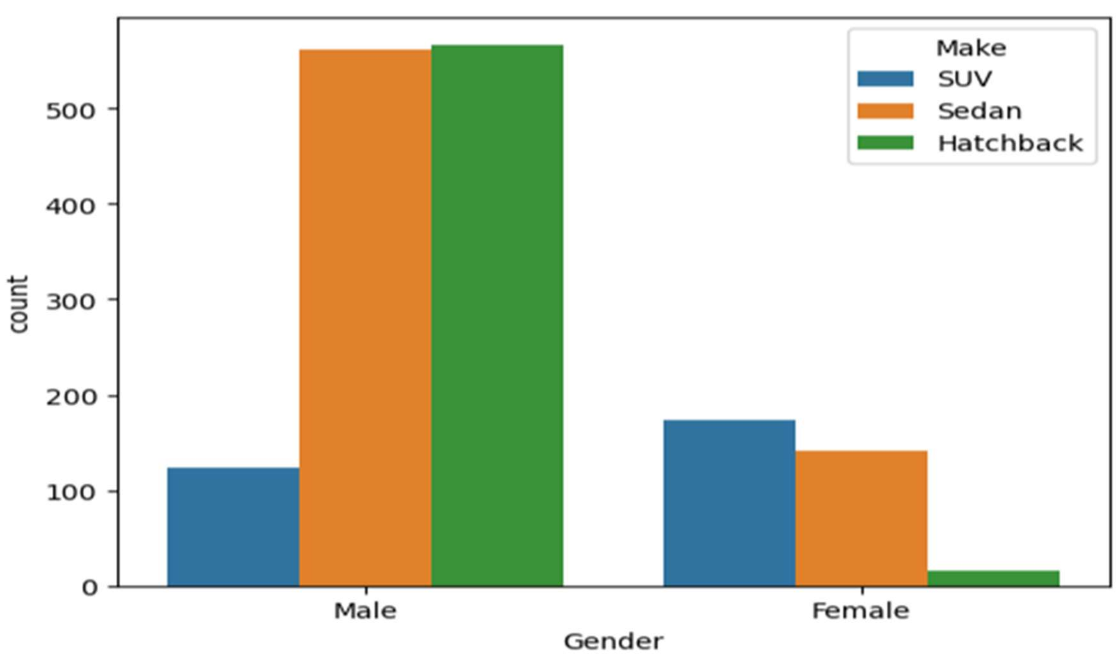
- **We can see that with increase in Persons Age , Total Salary also increase and they prefers SUV cars which is most costly among all Cars**

1. **Number of Dependents**



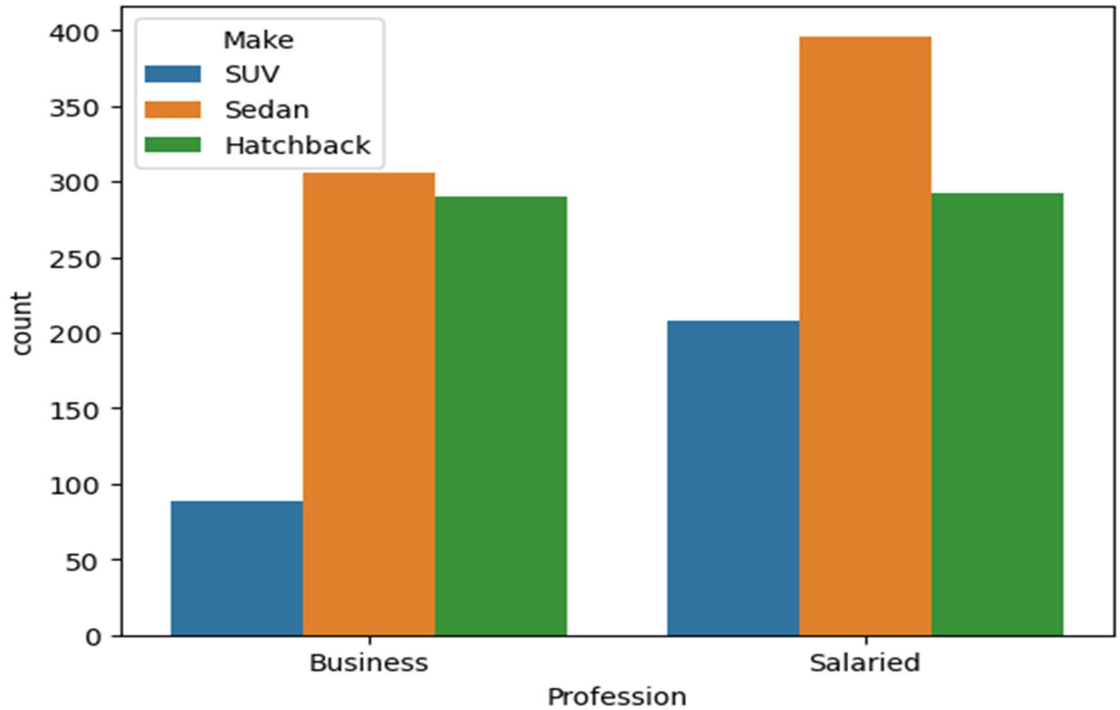
- **We can see that people having 0 dependents buys SUV & Hatchback cars only**
- **People having 1 & 2 dependents prefers Sedan Car Over SUV & Hatchback Cars.**
- **People having 3 & 4 dependents prefers Hatchback Cars over SUV & Sedan Cars.**
- **SUV Cars are mostly preferred by people having 3 dependents.**
- **Sedan Cars are mostly preferred by people having 2 dependents.**
- **Hatchback Cars are mostly preferred by people having 3 dependents.**

2. **Gender**



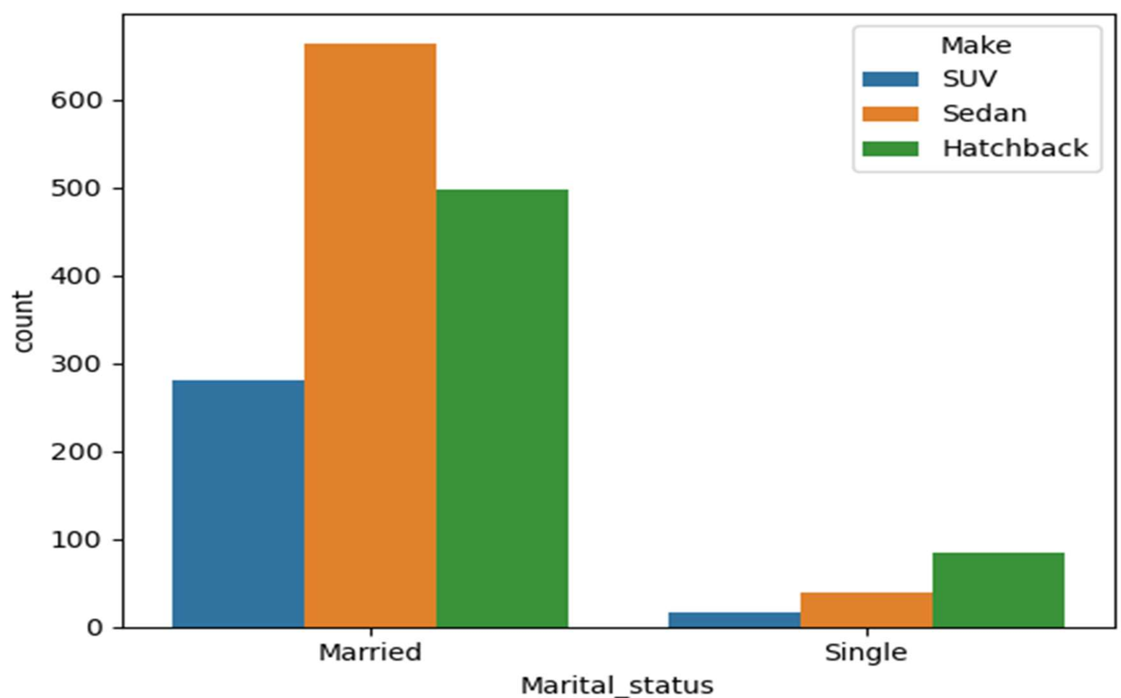
- **Male prefers Sedan & Hatchback Cars over SUV Cars**
- **Female prefers SUV cars over Sedan & Hatchback Cars**

3. Profession



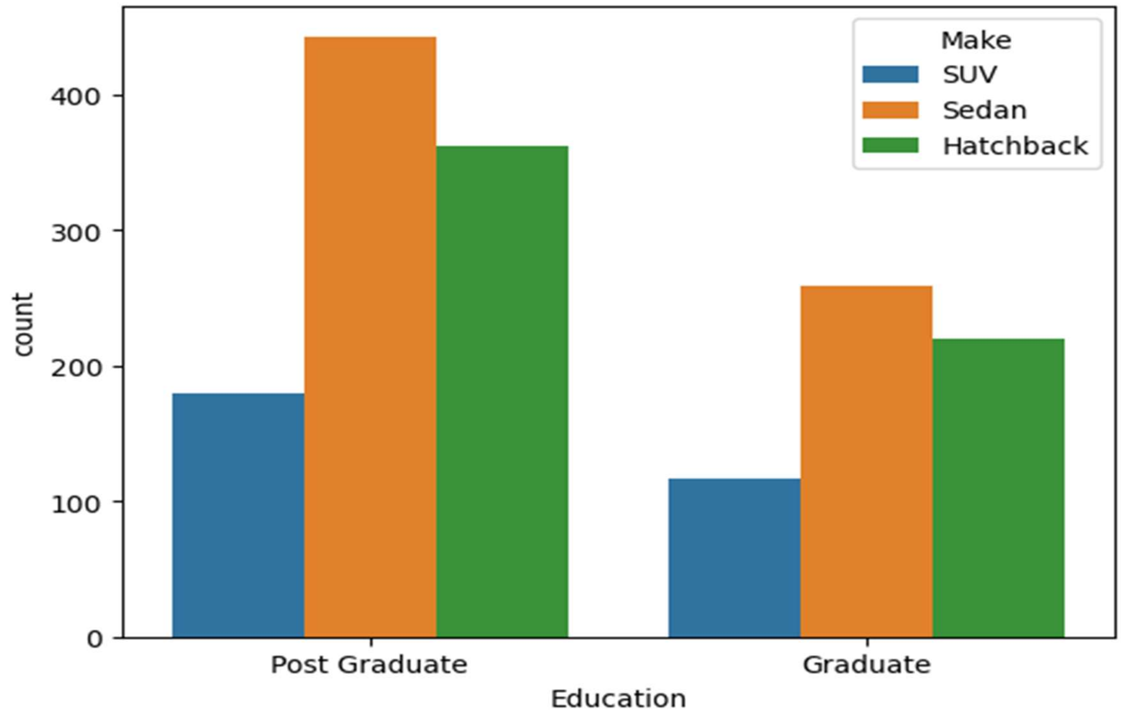
- **People based on Business or Salary prefers Sedan & Hatchback Cars Over SUV Cars**
- **SUV Cars are most preferred by Salaried people**
- **Sedan cars are most preferred by Salaried people**
- **Hatchback cars are equally preferred by Salaried and Business people**

4. Marital Status



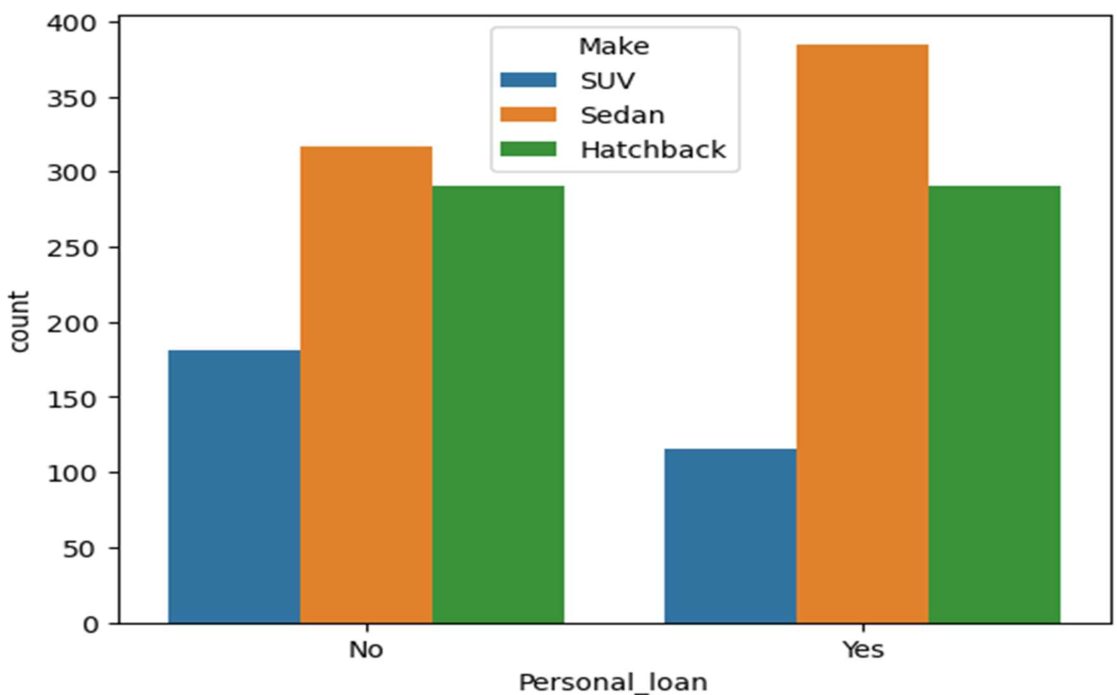
- **Single people mostly prefers Hatchback cars over SUV & Sedan cars**
- **Married People mostly prefers Sedan cars over Hatchback and SUV**

5. Education



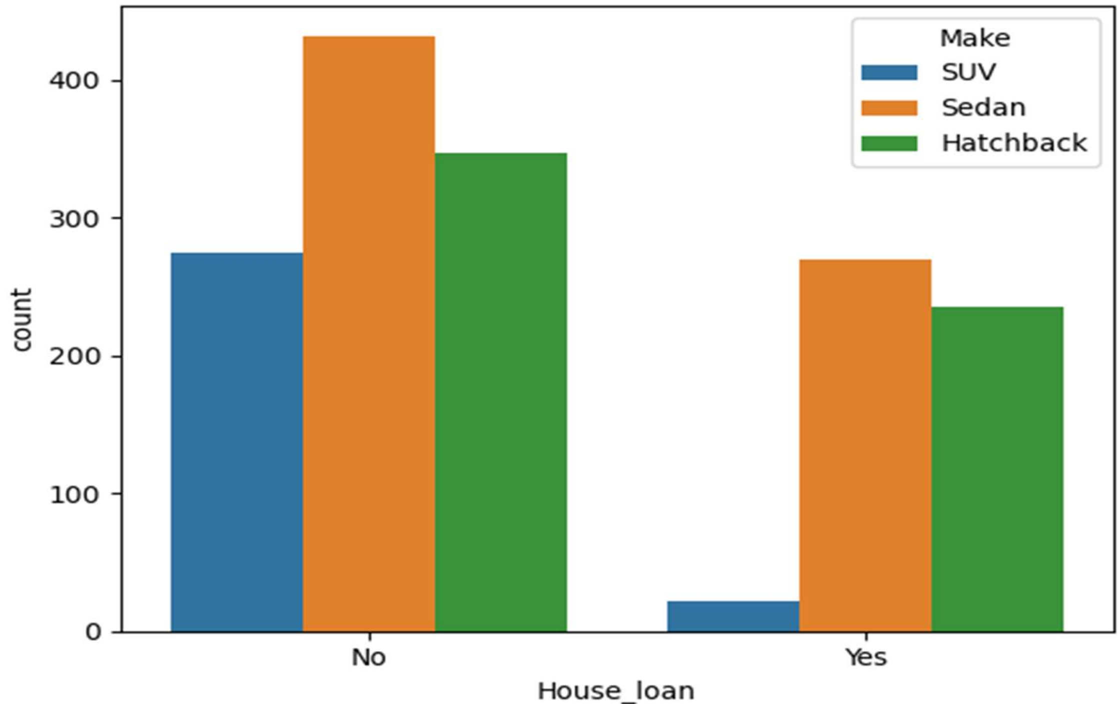
- **Post Graduate & Graduate person prefers Sedan cars & Hatchback cars over SUV**
- **SUV cars are more liked by Post Graduate person than Graduate person**

6. Personal Loan



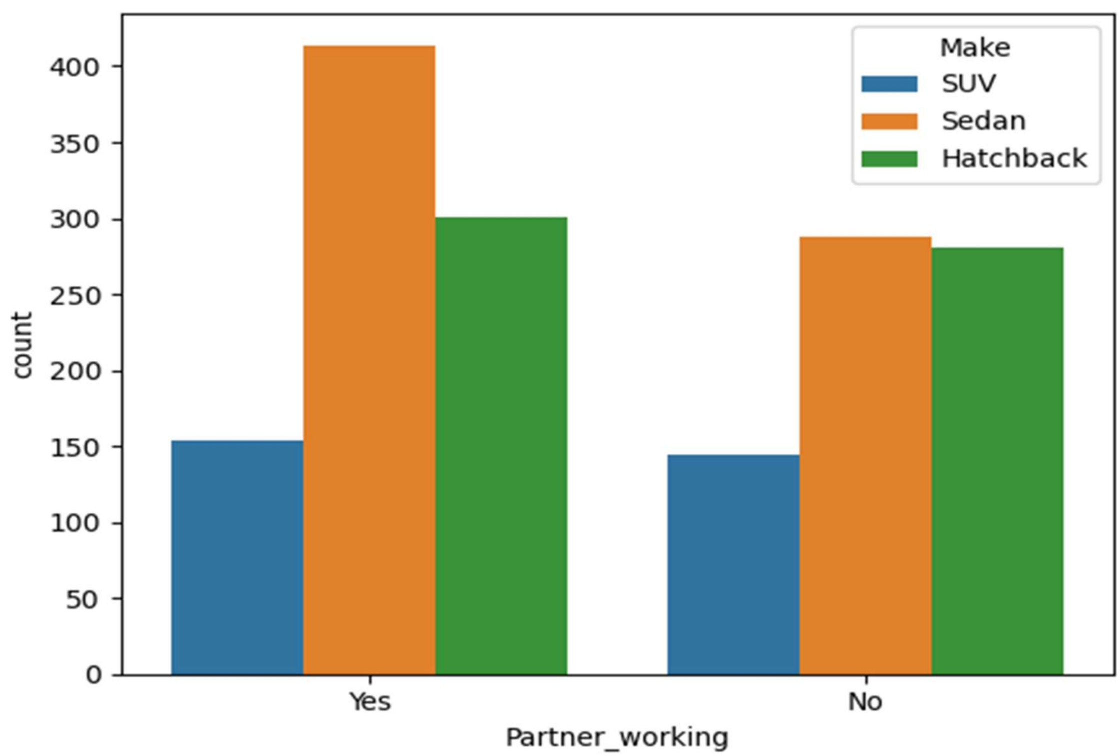
- **People having Personal loan or not prefers sedan car over SUV & Hatchback cars**
- **SUV are more preferred by people who doesn't have Personal loan**

7. House Loan



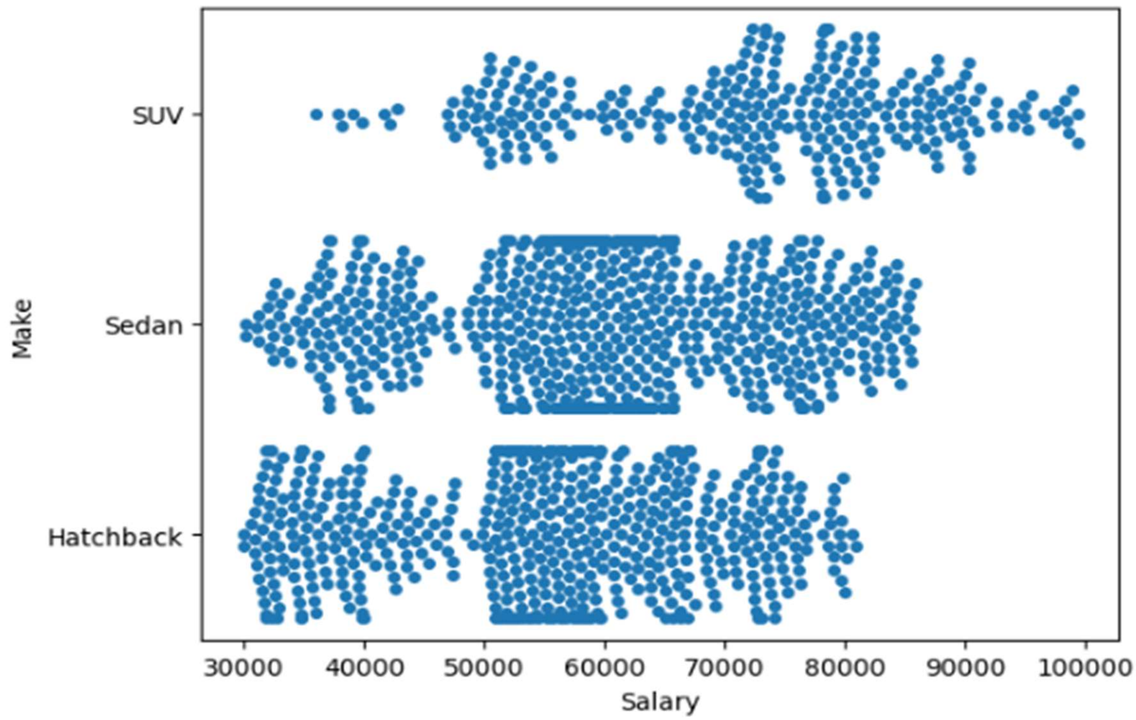
- **People having House loan or not prefers sedan car over SUV & Hatchback cars**
- **SUV & Hatchback are more preferred by people who doesn't have House loan**

8. Partner Working



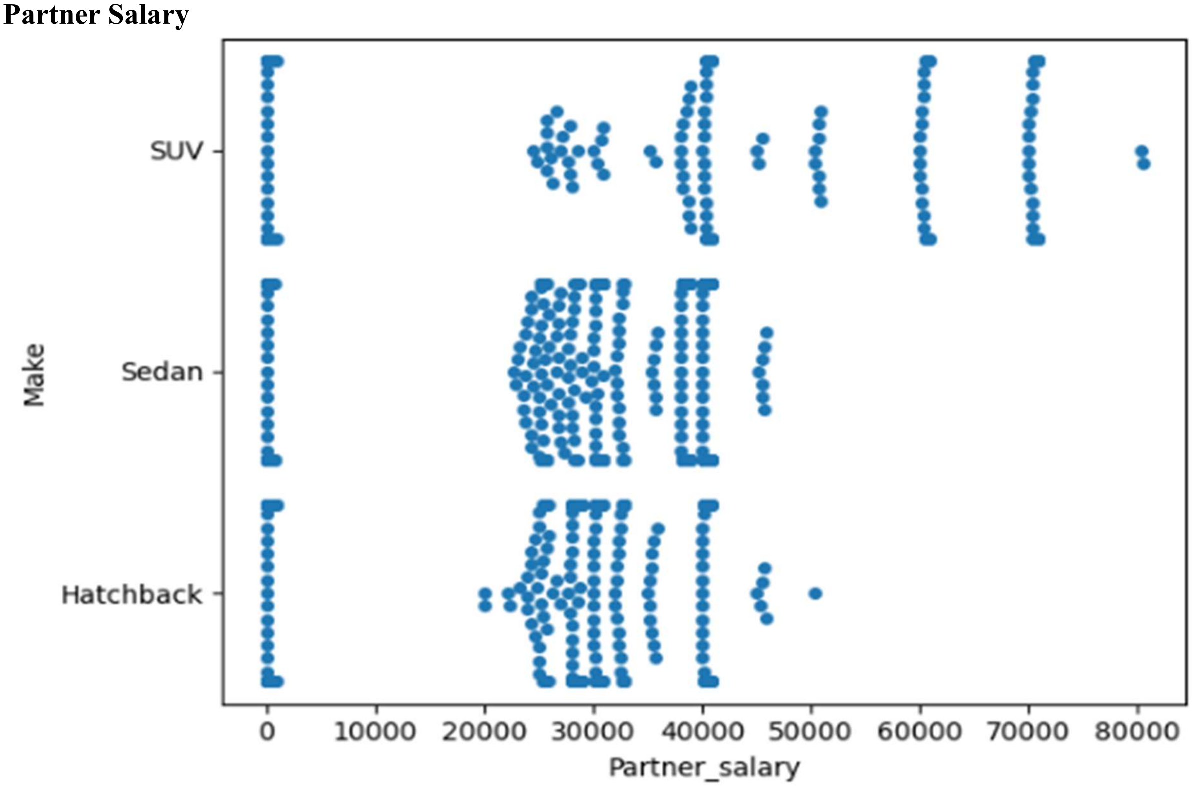
- **People whose partner are working prefers all make of cars over people whose partner are not working**

9. Salary



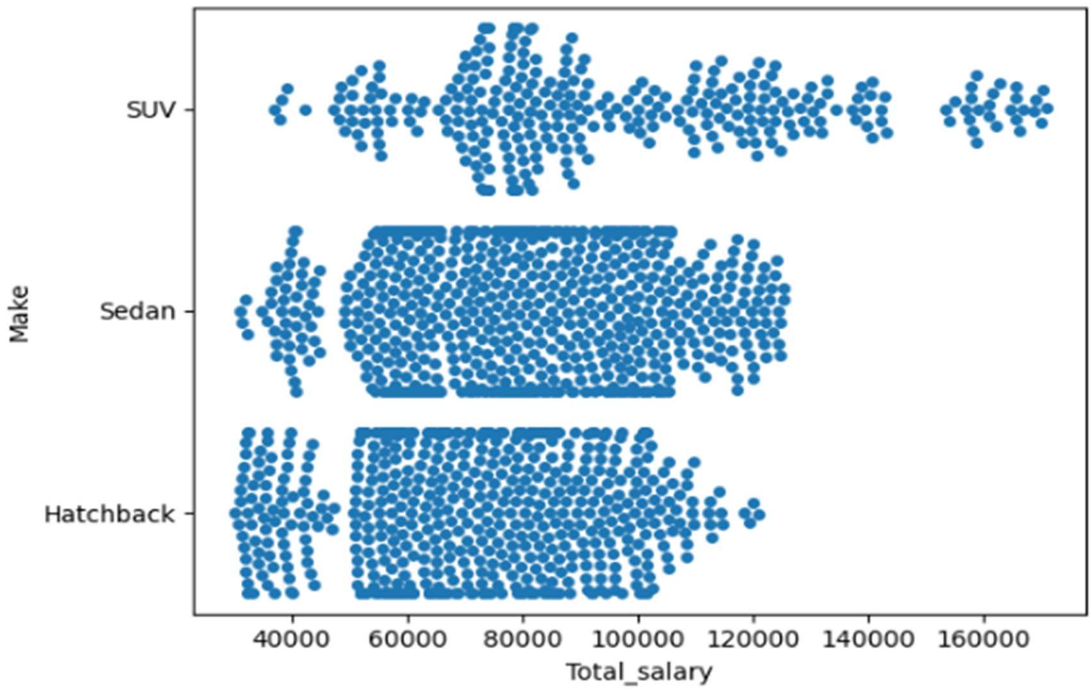
- **We can see that with increase in salary person prefers higher Price car as we know that Hatchback is least expensive and SUV are most expensive cars**

10. Partner Salary



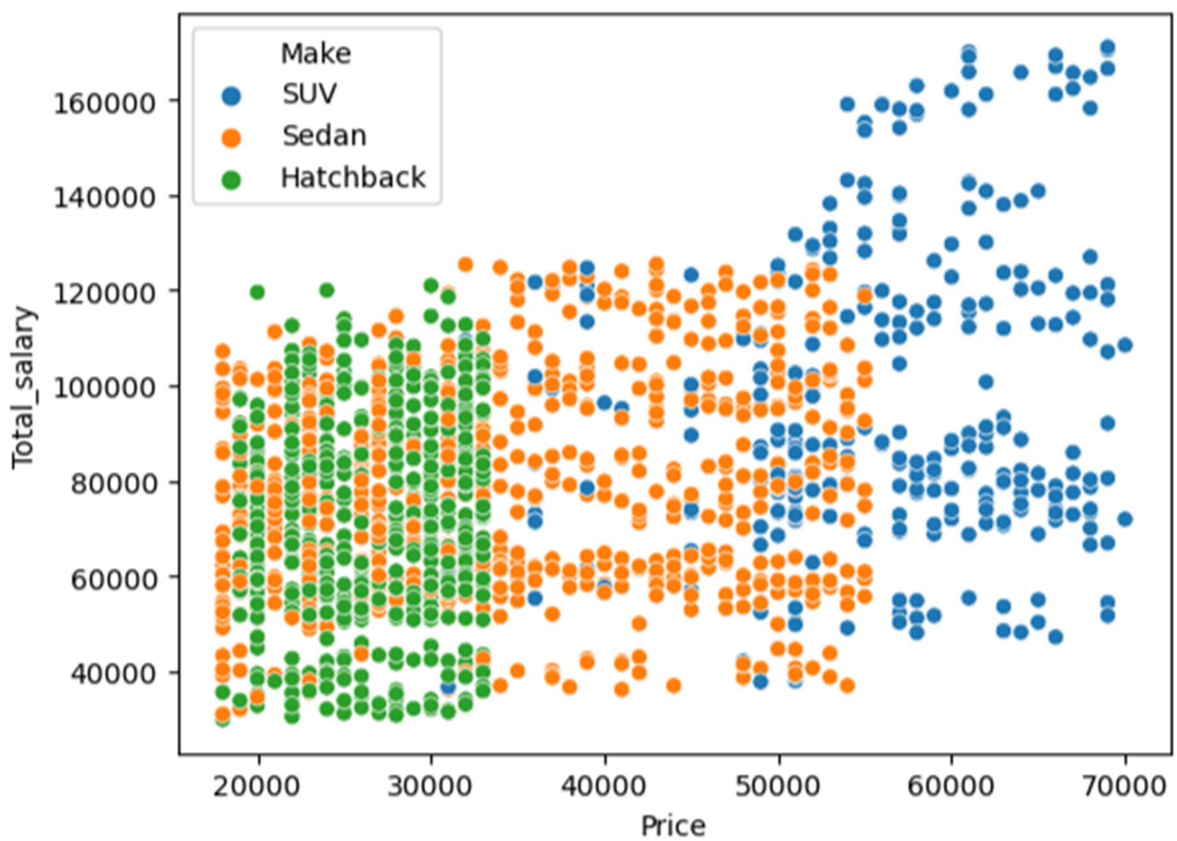
- **We can see that the person whose partner salary is more prefers SUV cars**

11. Total Salary



- **We can see that the person whose Total salary is more prefers SUV**

Let's compare Make – Price – Total Salary



- **We can see in the above plot that Person having high Total salary prefers high price cars**
- **Also person having high Total Salary prefers SUV cars which we have seen above also**

HEAT MAP of Numerical Variables

From the above heat map, we can see that the: -

- **Price and Age of the car are strongly related which means with increase in person's Age the car price of car owned by person also increases**
- **Age and Salary is moderate related**
- **Partner Salary and Total Salary are also strongly related. Also Total Salary is the sum of Salary and Partner Salary so it will obviously be highly related**
- **Partner Salary and Salary is Slightly related**

E. Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available.

E1) Steve Roger says "Men prefer SUV by a large margin, compared to the women"

Ans:-

```
print('total number of Male',austo.Gender.value_counts().max())
print('total number of Female',austo.Gender.value_counts().min())
```

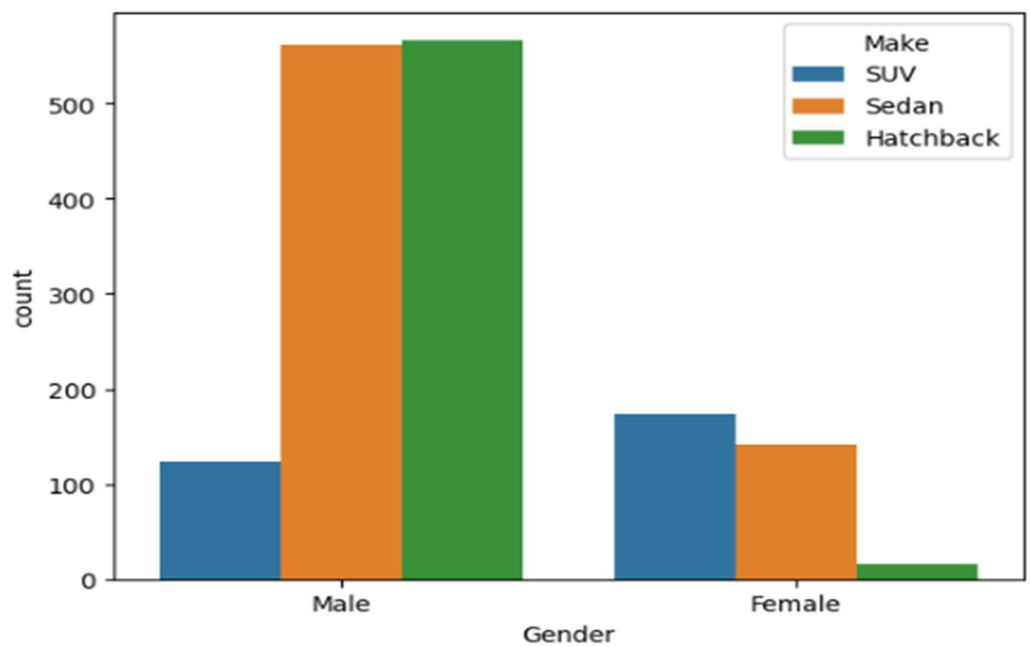
```
total number of Male 1252
total number of Female 329
```

```
austo.groupby('Gender')['Make'].value_counts()
```

```
Gender  Make
Female  SUV          173
        Sedan        141
        Hatchback     15
Male    Hatchback    567
        Sedan        561
        SUV          124
Name: Make, dtype: int64
```

Out of total Female i.e. 329, female having SUV are 173 i.e. 54% of total woman

Out of total Male i.e. 1252, Male having SUV are 124 i.e. 10% approx. of total woman



Also in above plot we can see that the Female owner of SUV are more compared to Male.

So, I totally disagree with Steve Rogers. Men's does not prefer SUV compared to Female

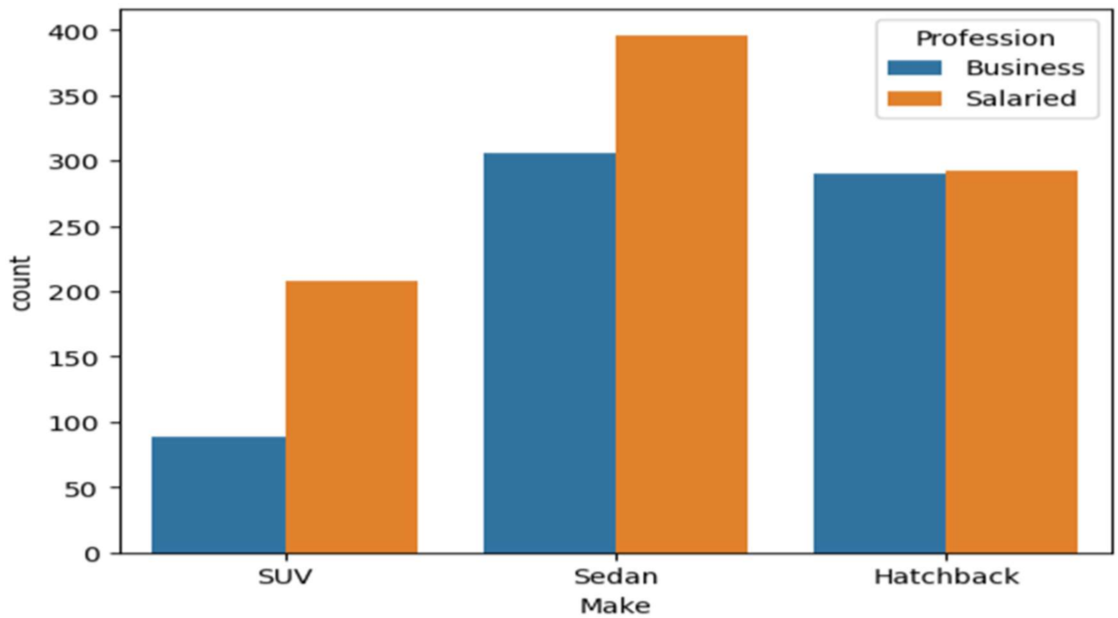
E2) Ned Stark believes that a salaried person is more likely to buy a Sedan

Ans:-

```
austo.groupby('Profession')['Make'].value_counts()
```

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

We can see from above chart that there are more numbers of salaried person who bought Sedan.



Also from the above chart we can see that the Large number of salaried people owns Sedan cars

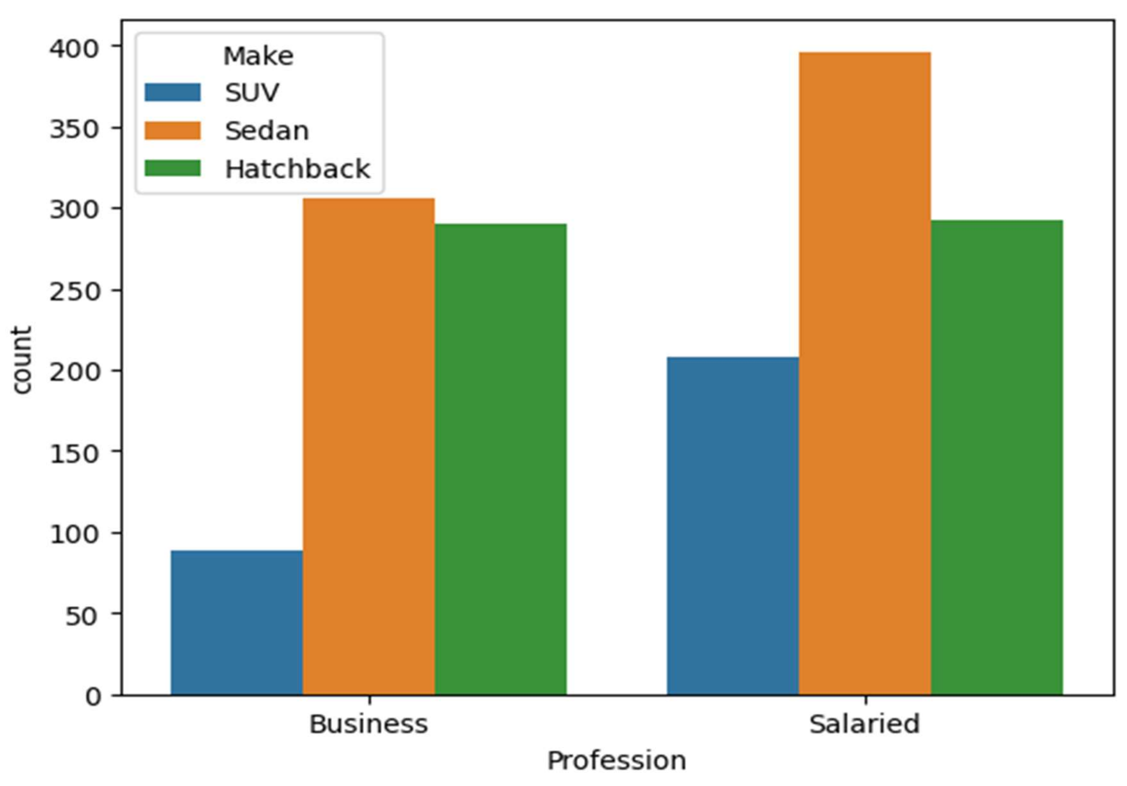
So, I agree with Ned Stark. Large number of Salaried person are more likely to buy Sedan Cars

E3) Sheldon Cooper does not believe any of them; he claims that a salaried male is an easier target for a SUV sale over a Sedan Sale.

Ans :-

```
austo.groupby(['Gender','Profession'])['Make'].value_counts()
```

```
Gender  Profession  Make
Female  Business    SUV           55
                    Sedan         50
        Salaried    SUV          118
                    Sedan         91
                    Hatchback     15
Male    Business    Hatchback    290
                    Sedan        256
                    SUV           34
        Salaried    Sedan        305
                    Hatchback    277
                    SUV           90
```



From the above chart and graph we can see that maximum number Male Salaried person owns Sedan and least number of Male Salaried person owns SUV.

So I totally disagree with Sheldon Cooper in his statement: - that a salaried male is an easier target for a SUV sale over a Sedan Sale.

F. From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a Business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions.

Give justification along with presenting metrics/charts used for arriving at the conclusions.

F1) Gender

Ans:-

```
austo.groupby('Gender')['Total_salary','Price'].sum()
```

Gender	Total_salary	Price
Female	29178800	15695000
Male	96709900	40585000

We can see that Female spent 15695000 from their total salary 29178800 i.e. 53.8% of their total salary

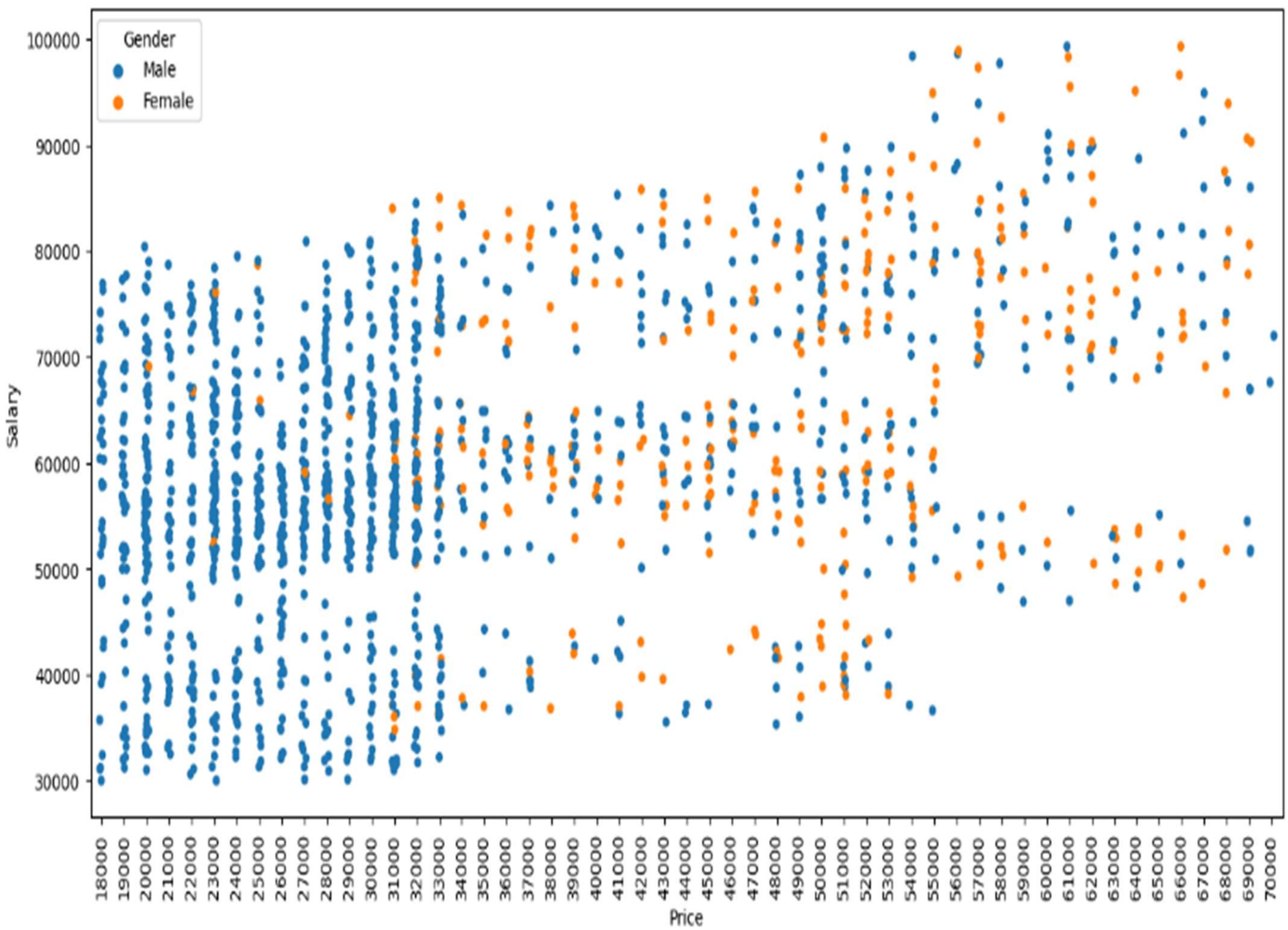
Whereas Male spent 40585000 from their total salary 96709900 i.e. 42% of their total salary

```
austo.groupby('Gender')['Salary','Price'].sum()
```

Gender	Salary	Price
Female	21778000	15695000
Male	73702100	40585000

We can see that Female spent 15695000 from their Salary 21778800 i.e. 73% of their Salary

Whereas Male spent 40585000 from their total salary 73702100 i.e. 55% of their Salary

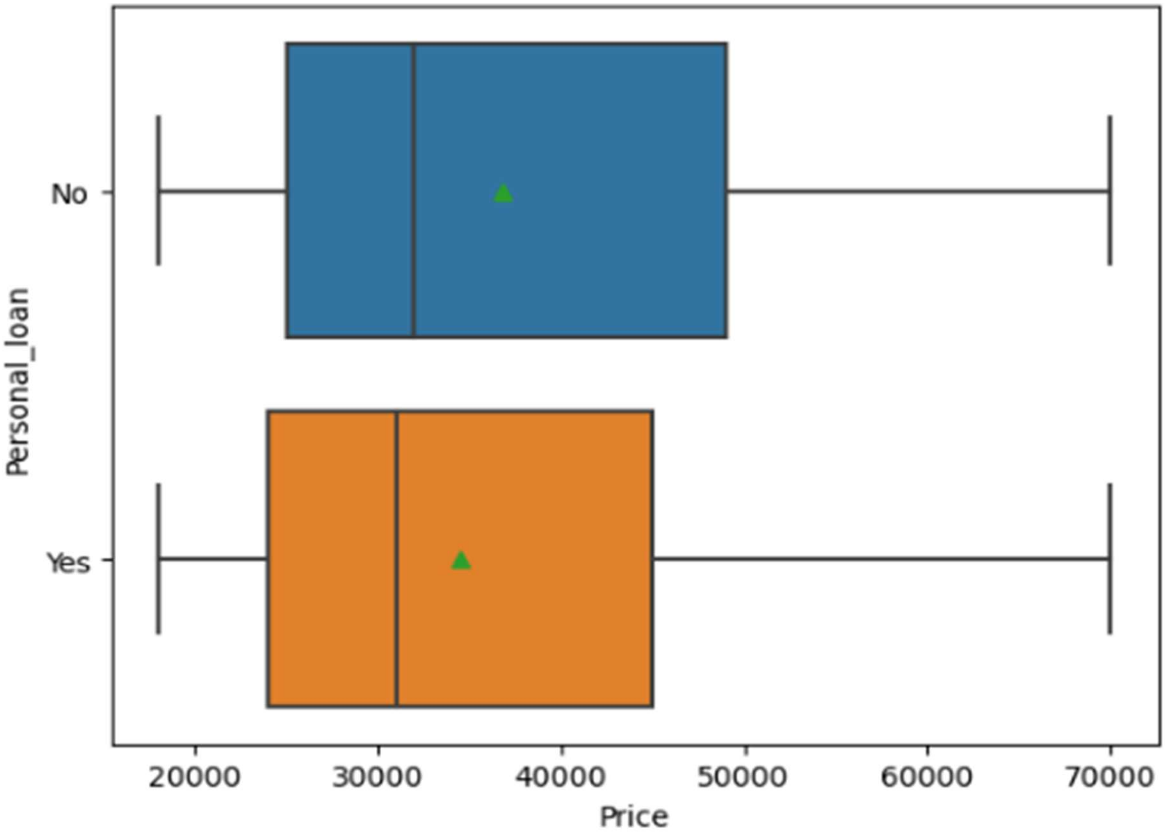


In Above Plot we can see that female generally prefers high priced car compared to male.

From Above we can conclude that: -

1. Female are easier Target for high priced car than Male
2. Marketing Campaign should be increased focusing Males
3. Number of cars owned by Males are higher than the Female

F2) Personal Loan



From Above boxplot we can see that the person who had not taken Personal Loan have higher mean as well as median compared to person who has taken Personal Loan. Which also conclude that the Person who had not taken Personal loan prefers to buy more expensive car compared to person who have taken Personal Loan.

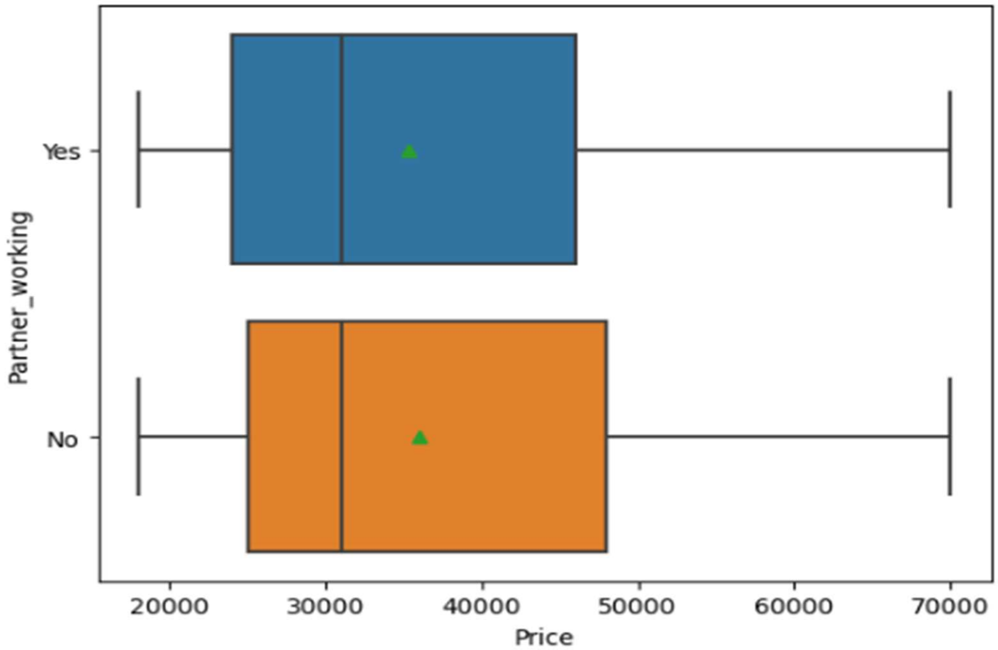
```
austo.groupby('Personal_loan')['Price'].sum()
```

```
Personal_loan
No      28990000
Yes     27290000
```

Also Person who had not taken personal loan has spent more money on Cars than person who had taken loan

G. From the current data set comment if having a working partner leads to the purchase of a higher-priced car

Ans :-



In Above boxplot we can see that the person whose partner be working or not working have nearly equal Mean & Median so this doesn't give us enough information about our objective

Let's separate partner working & partner not working.

```
Part_work=austo[austo.Partner_working=='Yes']
Part_nwork=austo[austo.Partner_working=='No']
```

```
Part_work.shape
```

```
(868, 14)
```

```
Part_nwork.shape
```

```
(713, 14)
```

Total no of person whose partner are working are 868

Total number of people whose partner are not working is 713

Let's assume car price above 60000 to be an expensive car

```
print('When Partner is working and number of car having price more than 60000 is  ',len(Part_nwork.loc[Part_nwork.Price>60000]))
print('When Partner is not working and number of car having price more than 60000 is ',len(Part_work.loc[Part_work.Price>60000]))
```

```
When Partner is working and number of car having price more than 60000 is   52
When Partner is not working and number of car having price more than 60000 is  55
```

Total number of cars above 60000 when partner is working: - 55 i.e. 6.3%

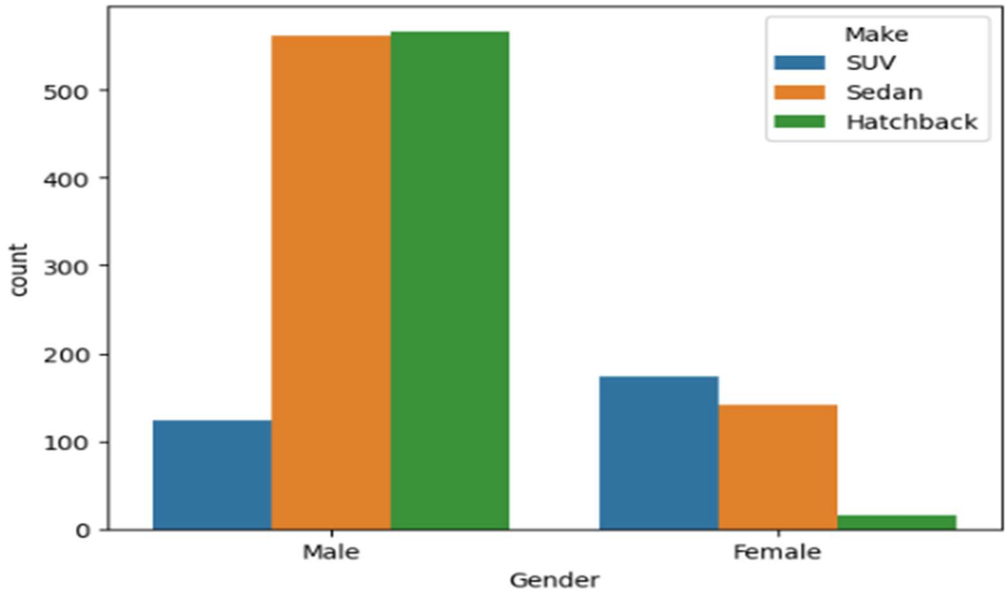
Total number of cars above 60000 when partner is not working: - 52 i.e. 7.3%

So we don't get any evident proof to get to conclusion that having working partner leads to purchase of high priced cars.

H. The main objective of this analysis is to devise an improved marketing strategy to send targeted information to different groups of potential buyers present in the data. For the current analysis use the Gender and Marital_status - fields to arrive at groups with similar purchase history

Ans :-

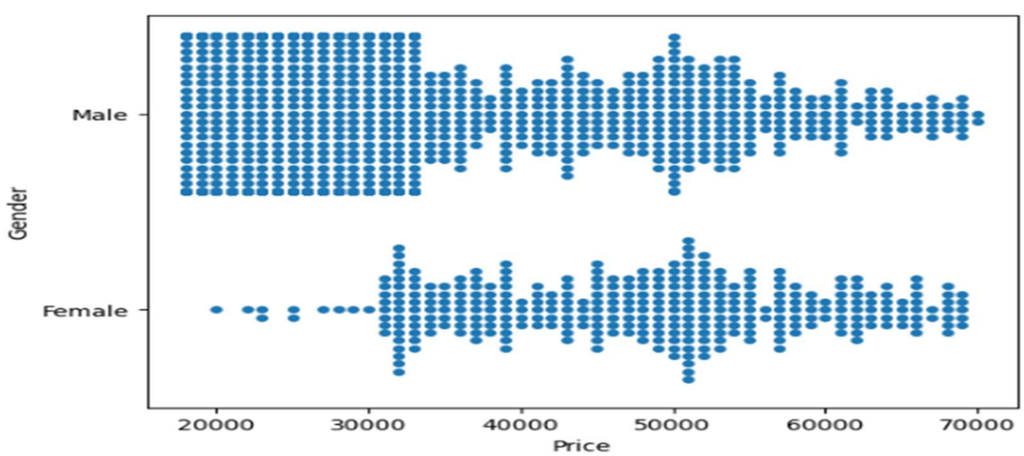
1. Let's see how does Gender affect the selection of Car Make



From above Chart/Plot we can get following conclusion: -

- Male prefers Hatchback & Sedan Over SUV cars: - We improve our Marketing Strategy based on Men for SUV cars
- Female prefers SUV cars over Sedan & Hatchback: - We improve our Marketing Strategy based on Female for Sedan and SUV cars

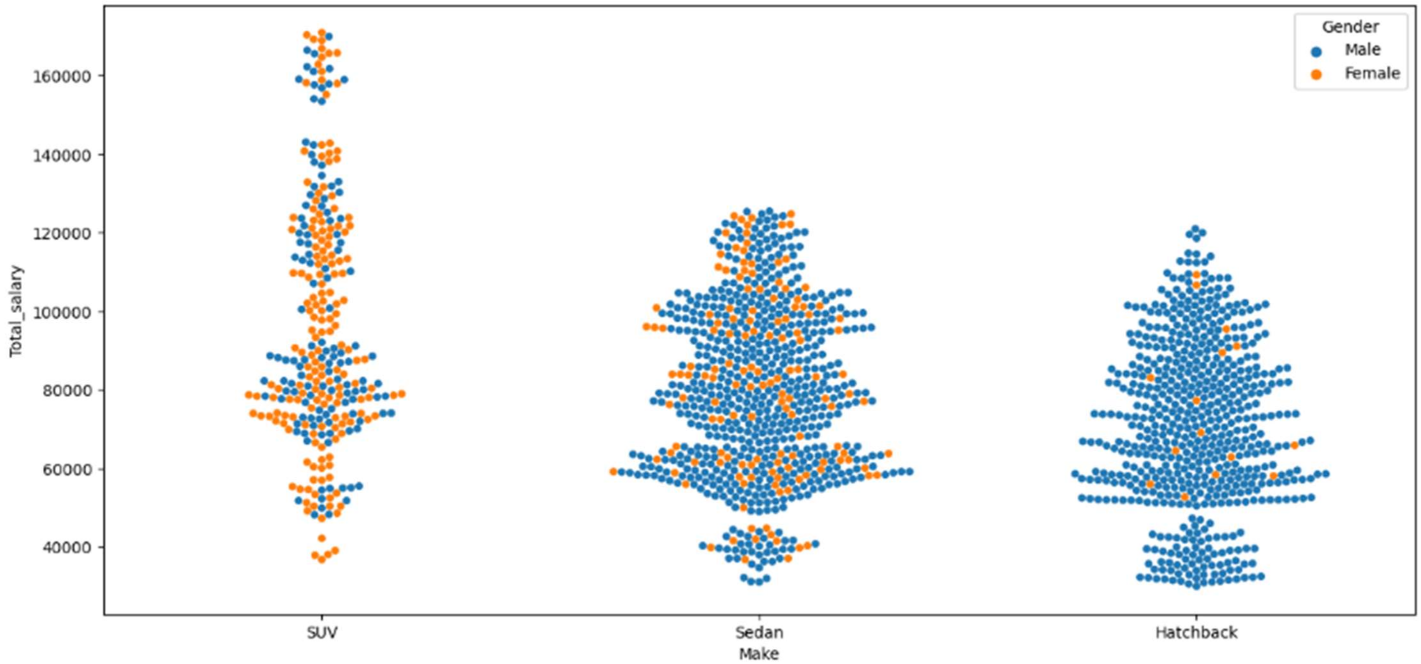
Let's Compare Gender and Price of a Car



From above Chart/Plot we can get following conclusion: -

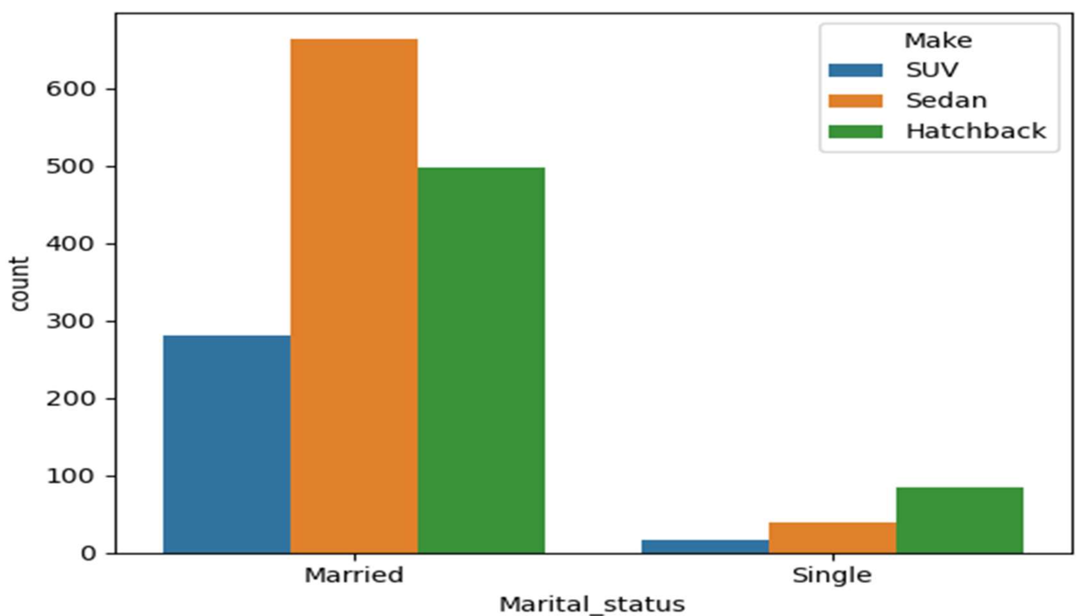
- Male Generally prefers low price cars compared to high price cars
- Female prefers medium and high price cars compared to low price cars

Comparing Gender , Total Salary & Make



- We can see that large number of woman who has high or average salary prefers SUV cars compared to Male
- Male having average total salary prefers Sedan and hatchback cars over SUV
- There are very few females buying Hatchback cars which can be of great interests

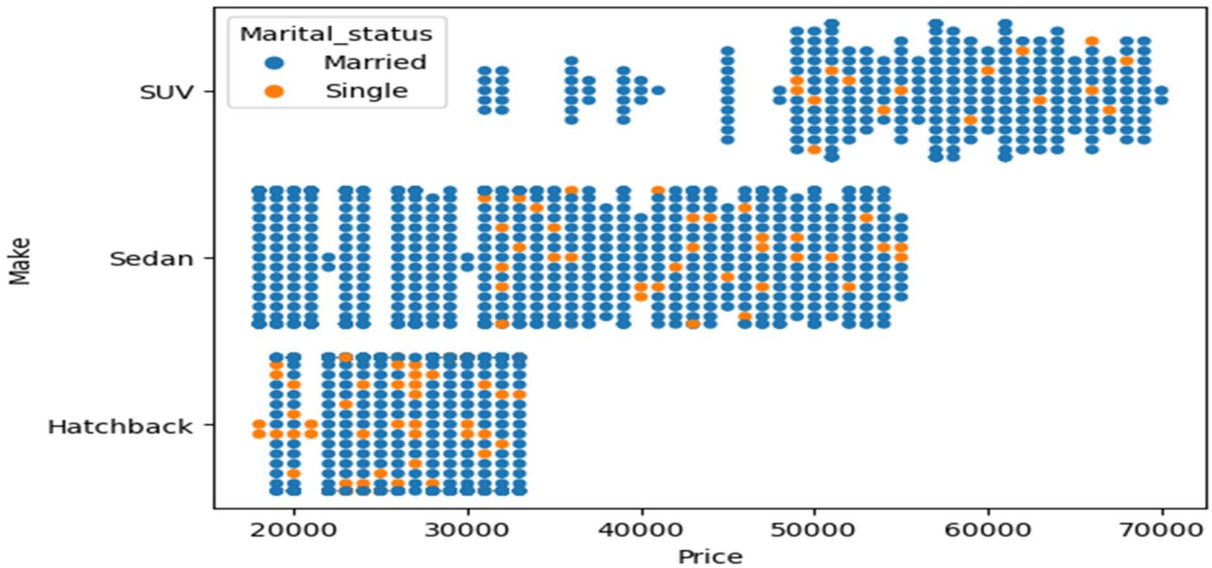
2. Let's see how does Marital Status affect the selection of Make and Price of a Car Marital Status and Make: -



From above Chart/Plot we can get following conclusion: -

- Married person prefers Sedan Cars
- Single person prefers Hatchback Cars

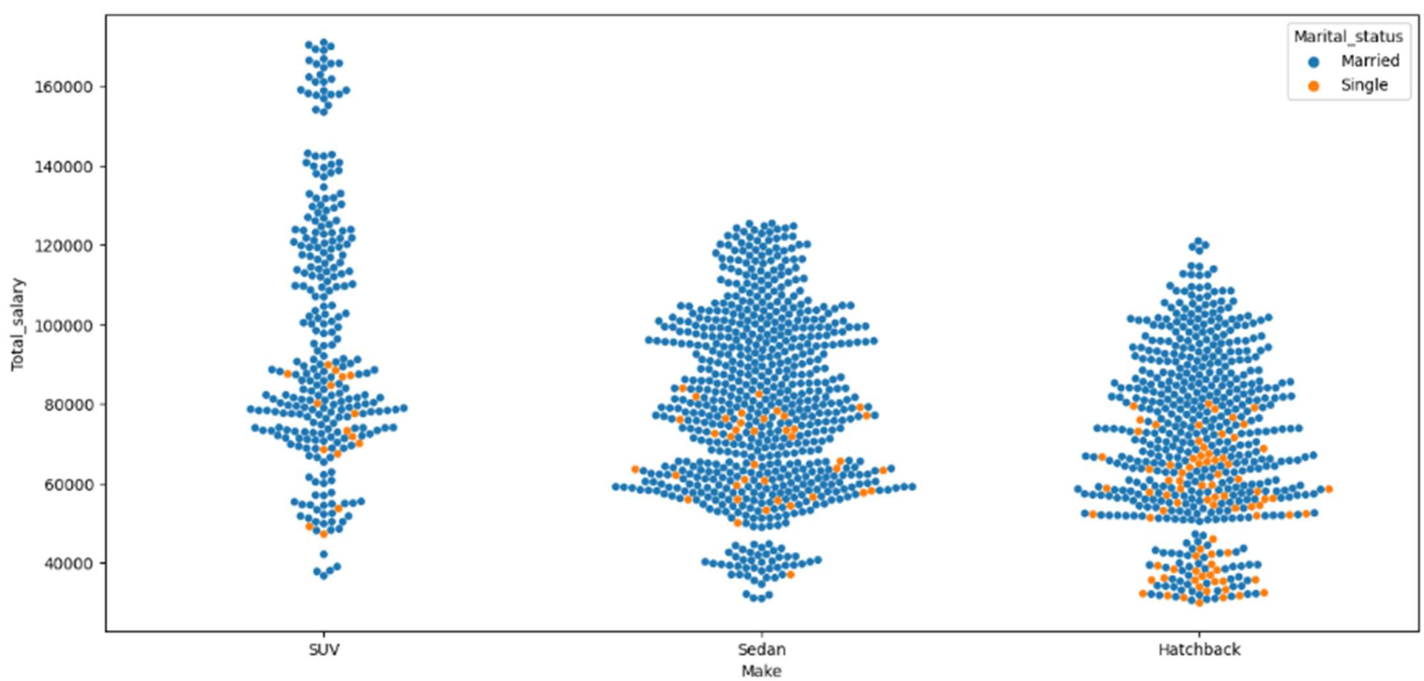
Marital Status, Price & Make: -



From above Chart/Plot we can get following conclusion: -

- Single Person prefers cheap cars compared to costly cars
- Married Person also prefers cheap cars compared to costly cars

Marital Status, Total Salary & Make: -



- **There are very few are none single person who has high salary and buy cars , so we can look for those individual who are single and what they look to buy cras**
- **There are few Married persons who has Average salary but buys SUV Cars in compared to Sedan and Hatchback Cars**

*****Framing An Analytics Problem*** Analyse the dataset and list down the top 5 important variables, along with the business justifications.**

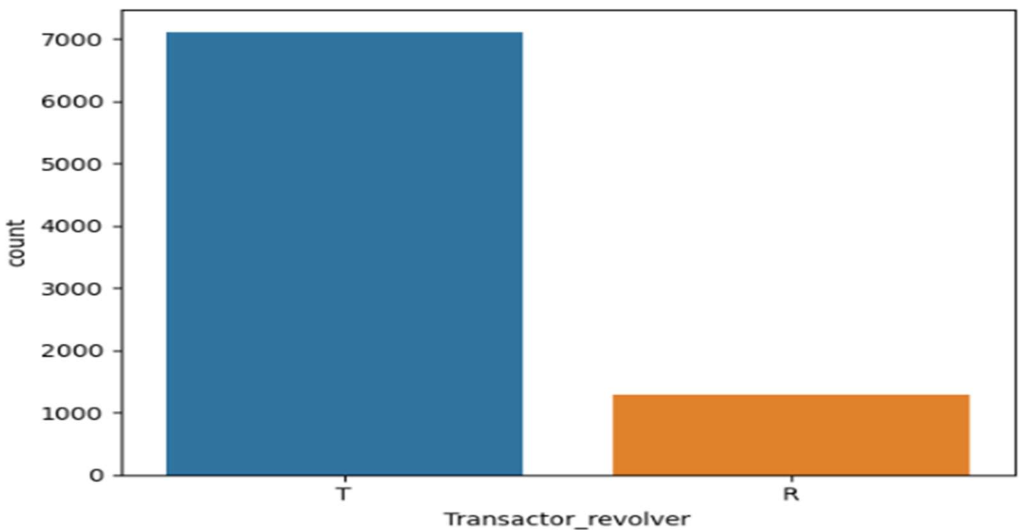
Ans :-

The cards which gives us the maximum Credit Card Limit are:

1. **Amex: - centurion, gold, platinum**
2. **MasterCard: - elite, prime, pulse**
3. **Visa: - C=chartered, edge, prosperity**

Most Important Variables are: -

1. **Transactor_revolver**
   Bank only makes profit if the credit card holder revolver the credit card bill month amount to next month as they had to pay interest for the due amount in next bill. As we can see there are very few Revolvers.

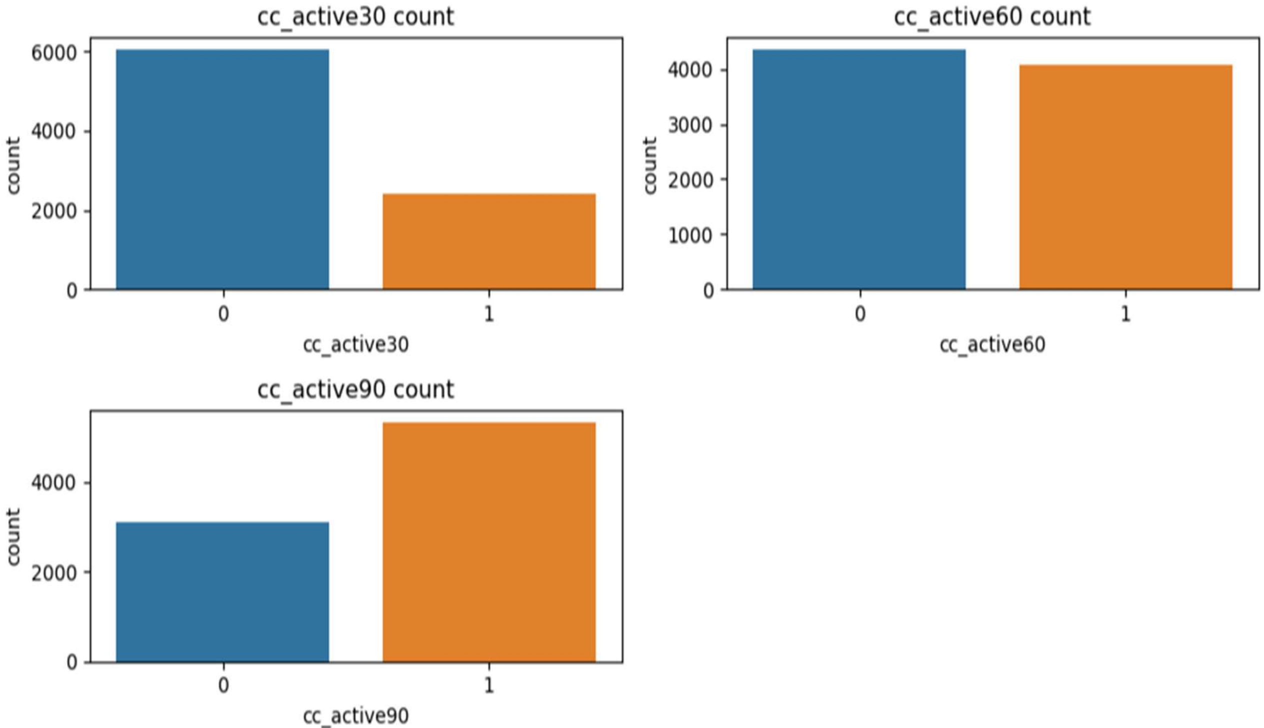



```
go.groupby('Transactor_revolver')[['cc_active30','cc_active60','cc_active90']].value_counts()
```

```
Transactor_revolver  cc_active30  cc_active60  cc_active90
R                    0            0            0               475
                     1            1            1               359
                     0            1            1               264
                                  0            1               197
T                    0            0            0              2621
                     1            1            1              2029
                     0            1            1              1419
                                  0            1              1046
```

We can see that there are lots of person who are paying the credit card bill in full (Transactor) so we can increase their card limit or the card type as they can have more money to spend and increase the chance of Revolver.

2. Credit Card Activity (cc_active)



- **We Can analyze the activity of credit card in past three month and as per the activity we can change the card type**
- **As in Above we can see that the activity of credit card has been decreasing so should analyze the reason**
- **As if the activity of the credit card is more we can enhance the limit or change the card to premium card**
- **If the activity is less we can give them some points or offers in order to attract them to use the credit card**

3. <u>Occupation at Source and High Net worth</u>

The person Occupation when combined with its Net worth can be a very important feature while providing a credit card. As in this we can review his monetary value and provide him the suitable card with some

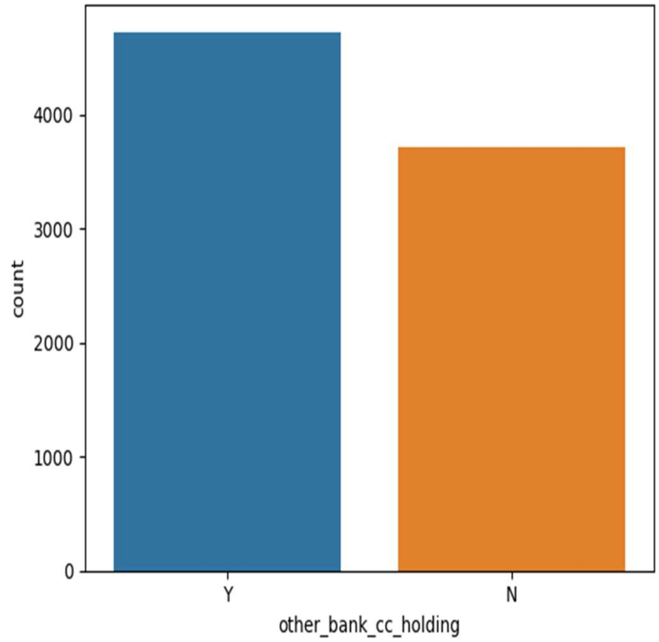
issuer	card_type	high_networth	cc_limit
Amex	centurion	B	980000
	platinum	B	920000
Visa	prosperity	B	990000
	edge	B	990000
Mastercard	elite	B	960000
Visa	chartered	B	990000
Mastercard	prime	B	980000
	pulse	B	990000
Visa	rewards	E	210000
	prime	E	40000
	indianoil	E	210000
	cashback	E	210000
	aura	E	210000
Mastercard	rewards	E	210000
	shoprite	E	210000
Visa	shoprite	E	210000
Mastercard	indianoil	E	210000
	cashback	E	200000
	aura	E	210000
Amex	smartearn	E	210000
	rewards	E	210000
	gold	E	980000
Mastercard	smartearn	E	200000
Visa	smartearn	E	210000

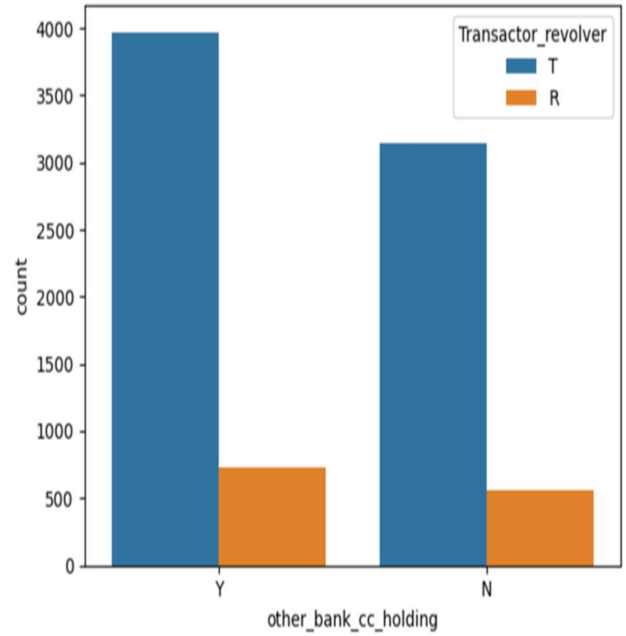
We can see that there are none of High Net Worth i.e. whose Net worth is A has been given premium cards which is Amex: - centurion, gold, platinum, MasterCard: - elite, prime, pulse or Visa: - chartered, edge, prosperity

Also there are some Low Net Worth person who has been given Premium cards .

In this we can further analyze the persons Net worth in order to know if the person is suitable for that card or not

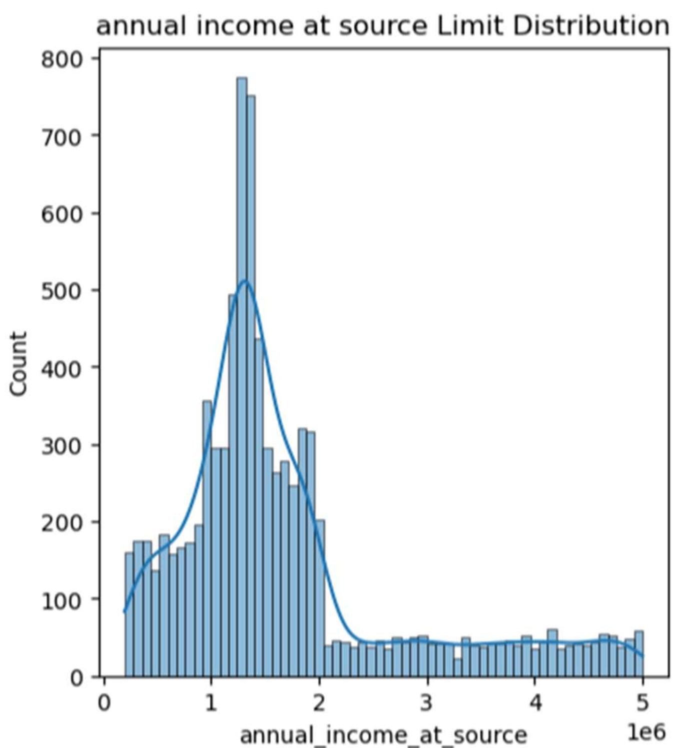
4. **<u>Other Bank Credit Card Holdding</u>**

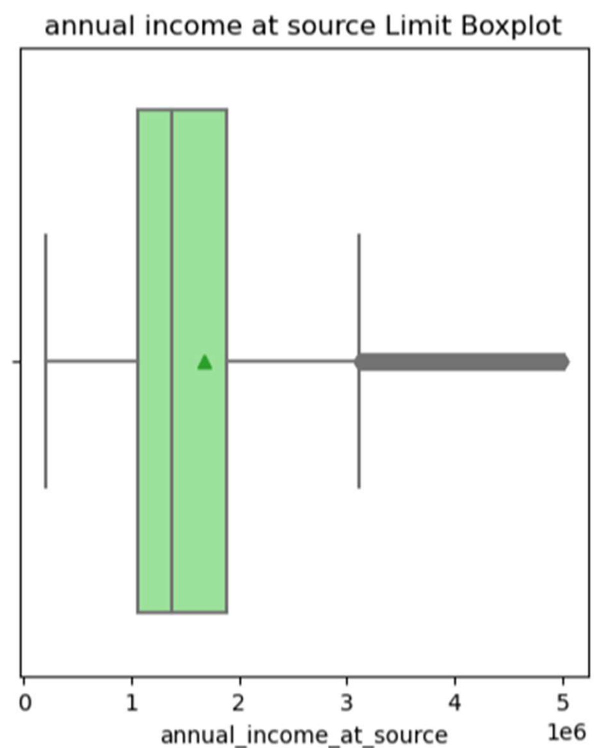




- **We can see that there are large number of customer who are holding a credit card of other bank account too so we need to look into the need of the customer more precisely who are holding credit card of other bank account**
- **Also there are large no of customer who are paying their bills in full which may not benefit to the bank. So we can look to increase the limit or change the card type of these customer to the premium**

5. **<u>Average Income at Source</u>**





- **Average income at source is one of the most important factor when it comes to paying the bills**
- **So we should always keep in mind that the person average income and credit card limit should not differ by large**