SMDM PROJECT REPORT

**Contents**

**1** Austo Motor Company Problem

**A.** What is the important technical information about the dataset that a database administrator would be interested in?.......... 2

**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?................8

**C.** Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business………………………………………... 8

**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………………………………………… 8

**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…………………….. 6

**F.** From the given data, comment on the amount spent on purchasing automobile 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………………………………………………………………………………. 4

**G.** From the current data set comment if having a working partner leads to purchase of a higher priced car……………………………… 2

**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 Gender and Marital\_status - fields to arrive at groups with similar purchase history…………………………………………………………………………... 6

**2** Framing An Analytics Problem……………………………………………………….. 10

Quality of Business Report (Please refer to the Evaluation Guidelines for Business report checklist. Marks in this criteria are at the moderator's discretion)……………………………6

**SMDM Project**

**Analysts are required to explore data and reflect on the insights. Clear writing skill is an integral part of a good report. Note that the explanations must be such that readers with minimum knowledge of analytics is able to grasp the insight.**

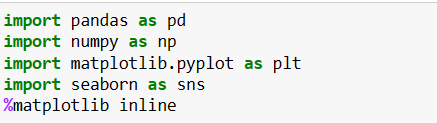
**Austo Motor Company is a leading car manufacturer and specializes in SUV, Sedan and Hatchback models. In its recent board meeting, concerns were raised by the members on the efficiency of the marketing campaign currently being used. The board decides to rope in an analytics professional to improve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

**You as an analyst have been tasked with performing a thorough analysis of the data and to come up with insights to improve the marketing campaign.**

**The instructions below are given to help you complete the project –**

1. **What is the important technical information about the dataset that a database administrator would be interested in? (Hint: Information about the size of the dataset and the nature of the variables)**

# Importing Libraries

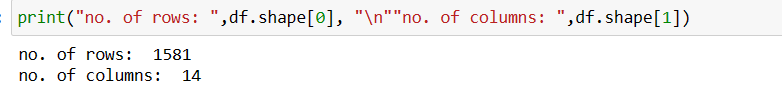


# Loading the data set

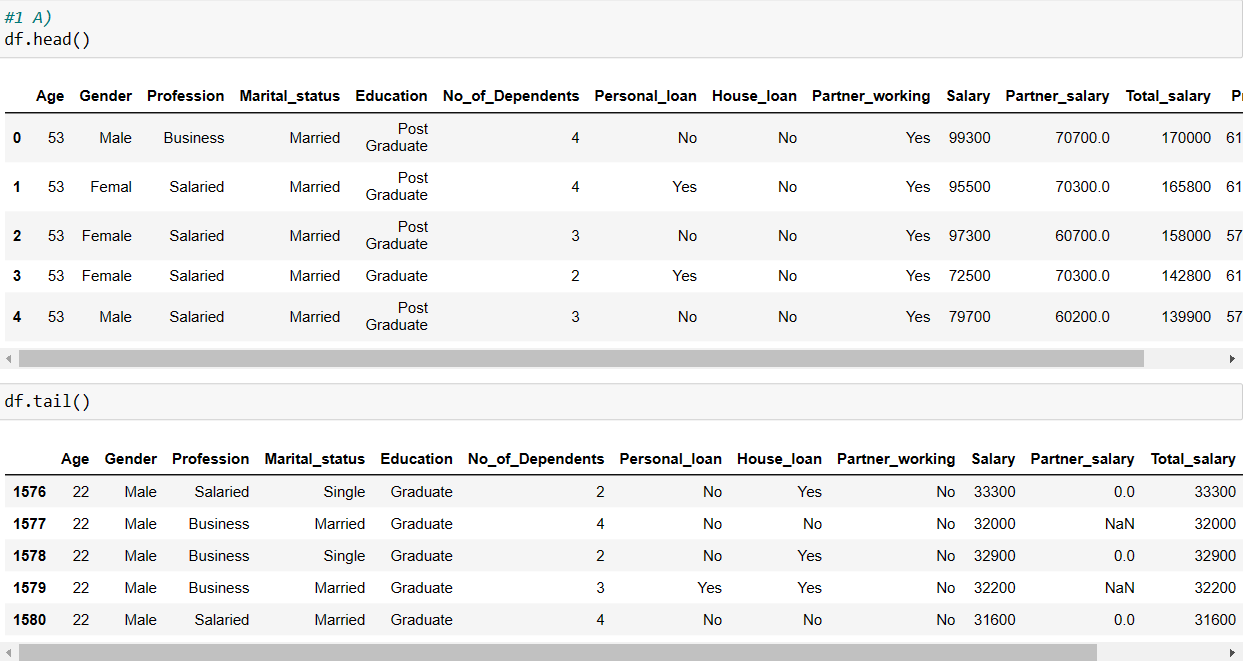
**We will be loading the EDA austo\_automobile using pandas. For this we will be using read\_excel file.**



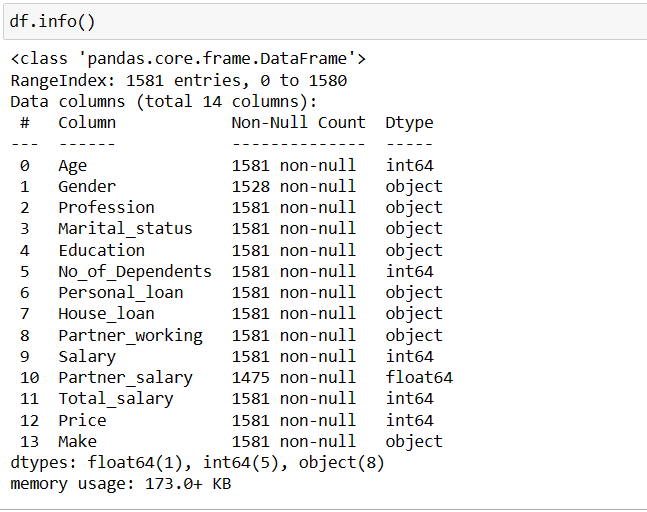
Dataset has 1581 rows and 14 columns.



We can view first 5 rows and last 5 rows by using head() and tail() function respectively.

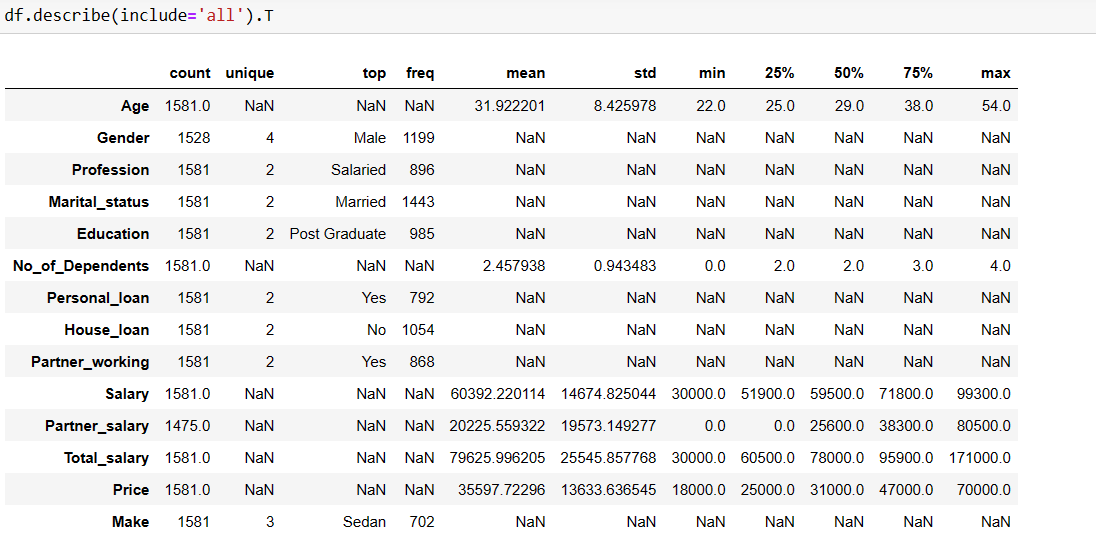


We can view dataset information as below.

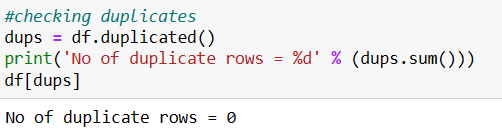


From above dataset information we can see that there are 5 numerical, 8 categorical and 1 float datatypes. There some null values in Gender and Partner\_salary column.

We can describe data with describe() function.



We can find mean,count, minimum value, maximum value etc information of variable.

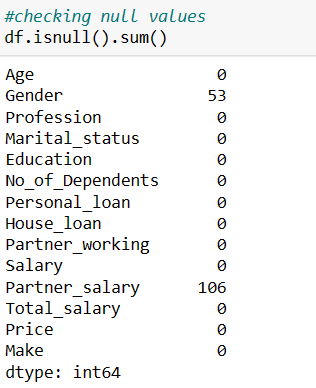


There are no duplicate rows.

1. **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.**

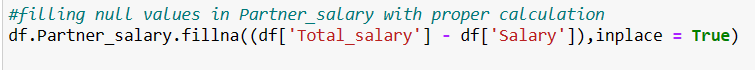
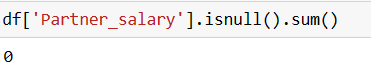
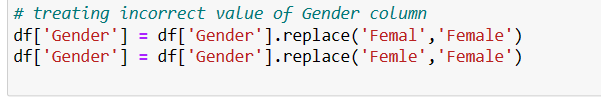
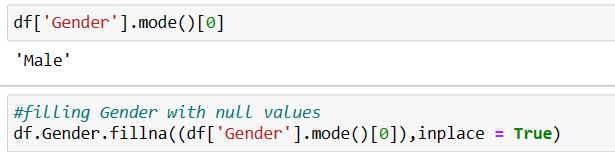
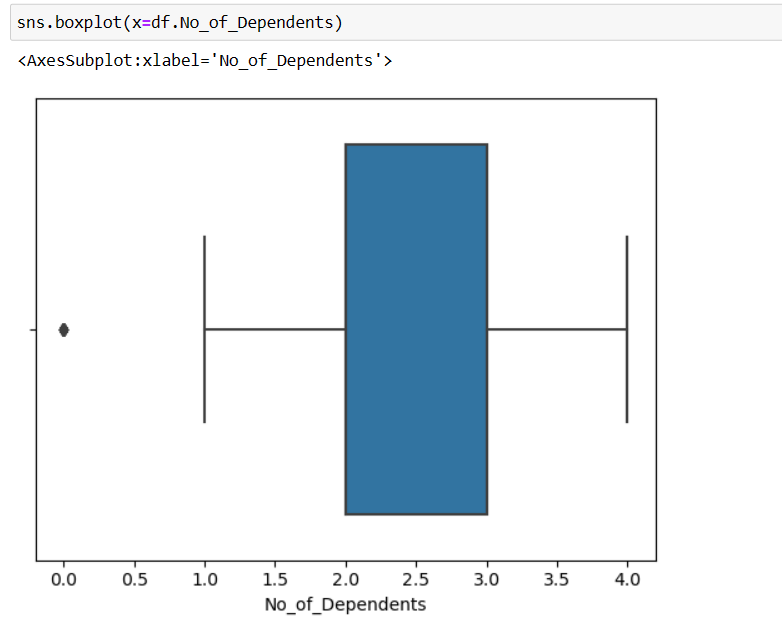
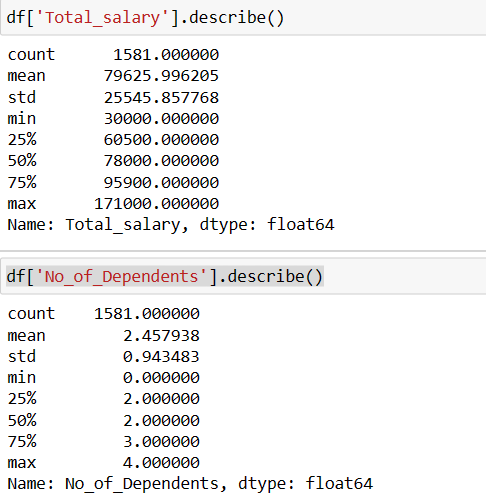
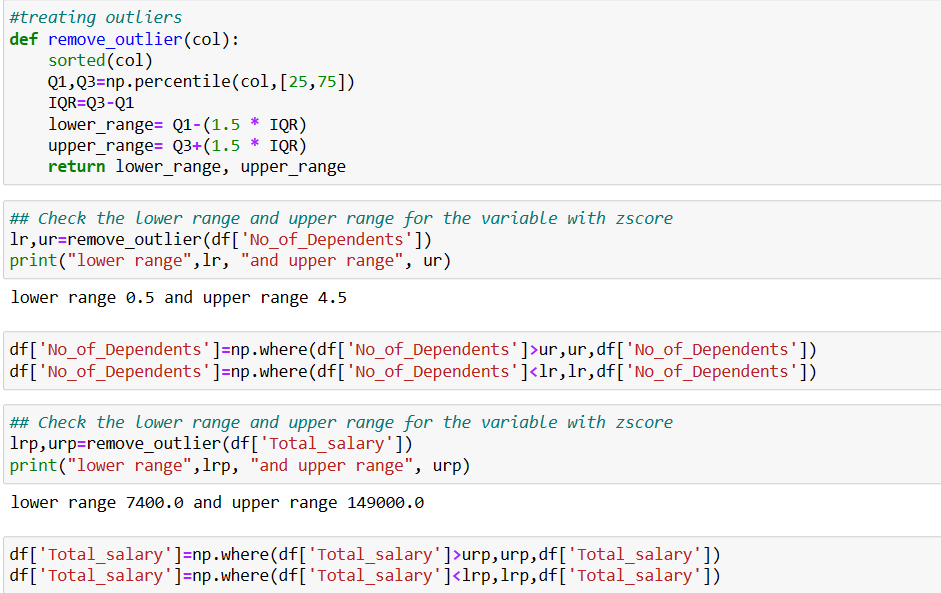
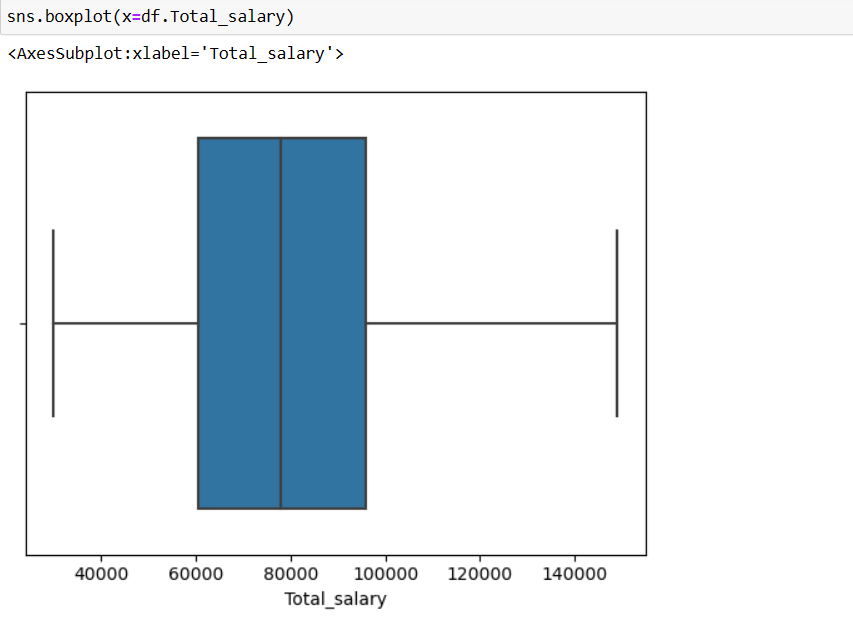
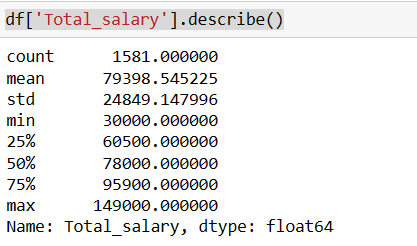
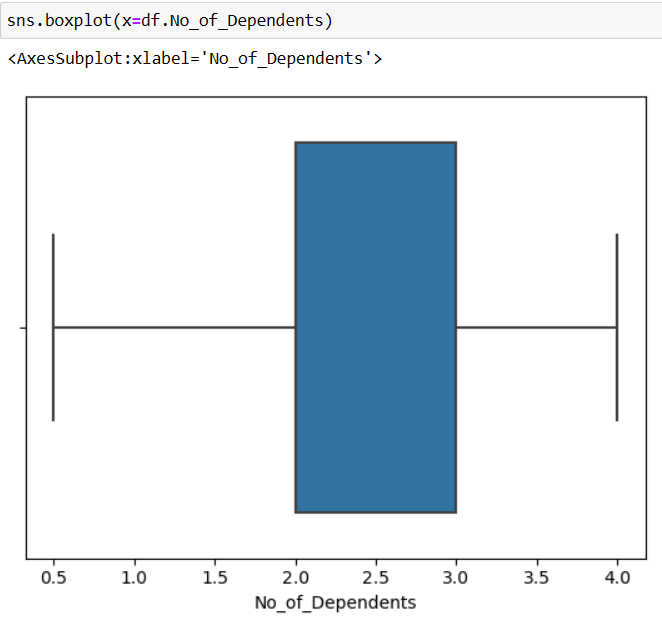
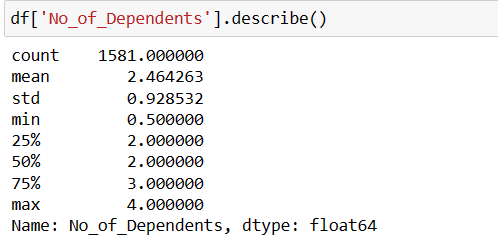
**Preprocessing Data:- Treating wrong values and null values. We have to treat outliers if any.**

1. We can see that there 53 null values in Gender column and 106 null values in Partner\_salary column.



1. We will first treat null values in Partner\_salary column. The three variables on salary are related to one another as:

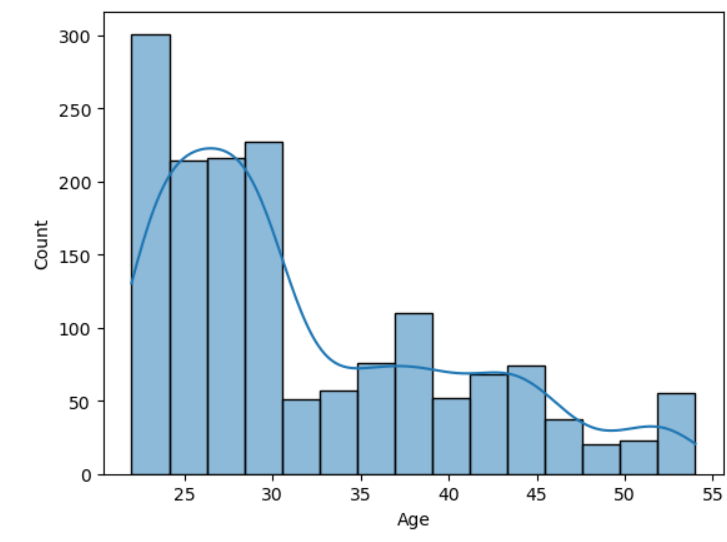


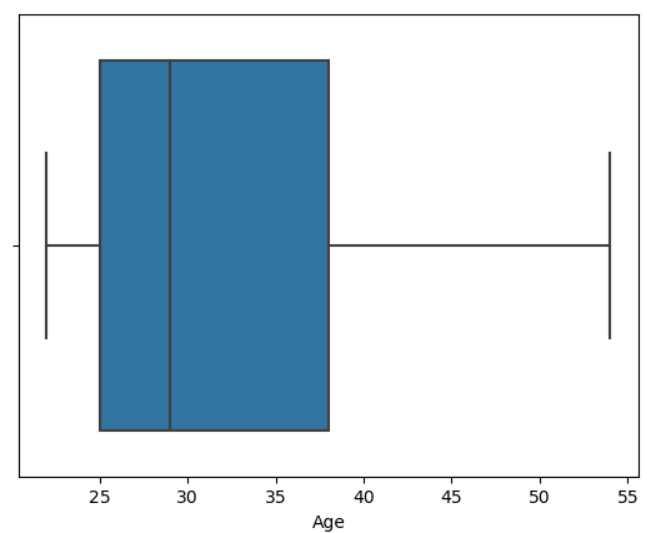
1. So from formula we can say that Partner\_salary = Total\_salary – Salary
2. 
3. Now we can check that there no null values left in Partner\_salary column.
4. 
5. There few wrongs values in Gender column.
6. Female is spelled as Femal and Femle. So we will replace it with Female.
7. 
8. But there still 53 null values left in Gender column. So we will replace null values in Gender column with most frequently occurred value in Gender column.
9. 
10. 
11. Now there are no null values left in Gender column.
12. Now we will plot a graph for all the numeric variable to check for outliers. We can see that there outliers in Total\_salary and No\_of\_Dependents column.
13. 
14. 
15. 
16. There are various techniques to treat Outliers We will use IQR technique as below.
17. We have created a user definded function for finding the lower and upper range for a variable so that outlier can be treated.
18. Now there no Outliers as shown below.
19. 
20. 
21. 
22. 
23. **Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.**

Univariate Analysis of Numerical fields –

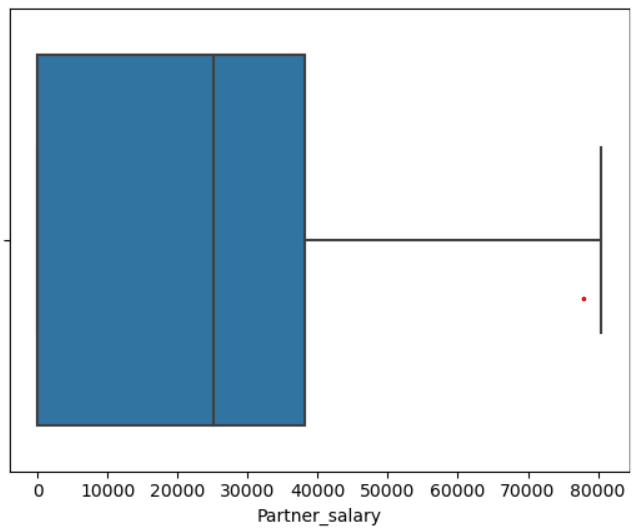
For performing Univariate analysis we will use the Boxplots and Histograms to get better understanding of the distributions. We are already done with data pre-processing.

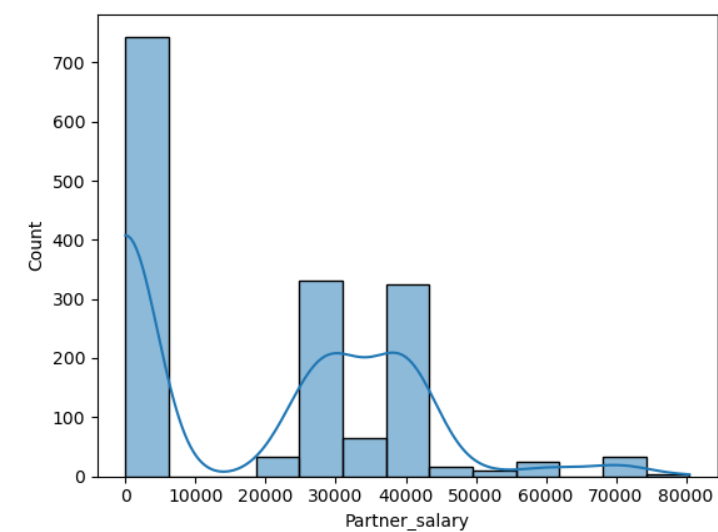
1. Age



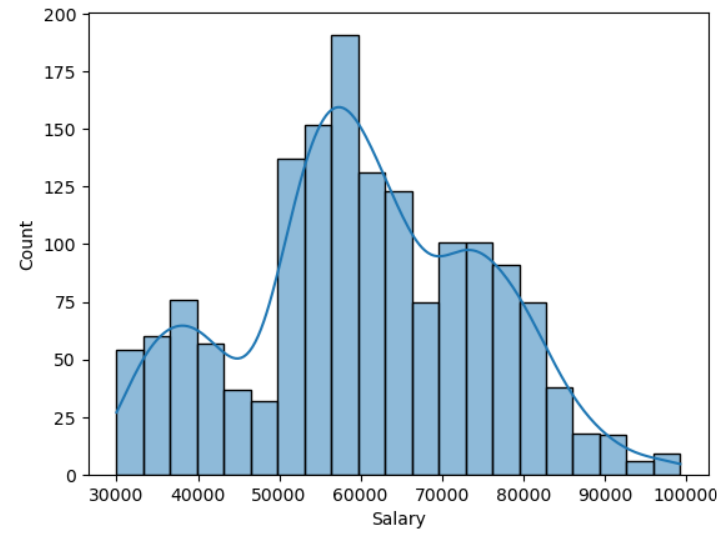


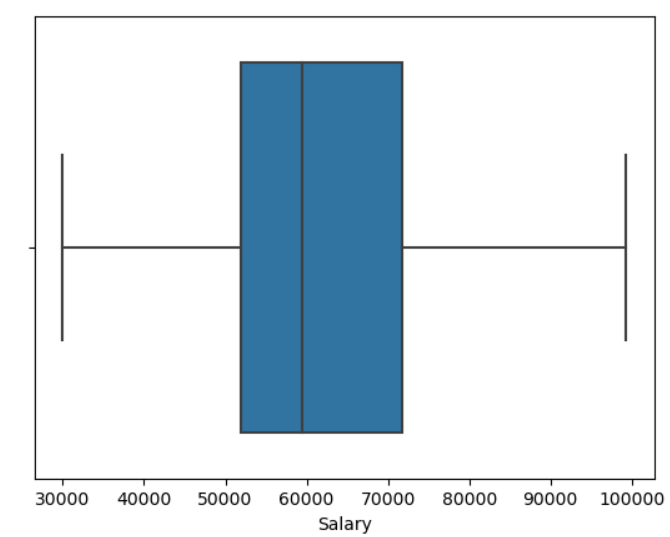
1. Partner\_salary



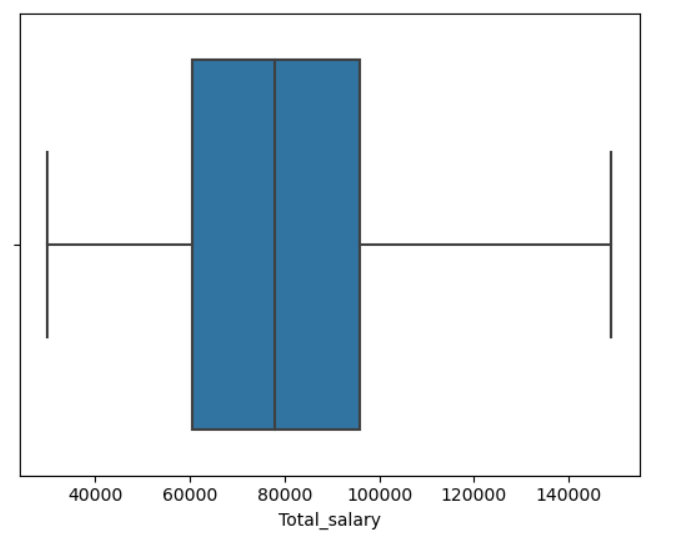


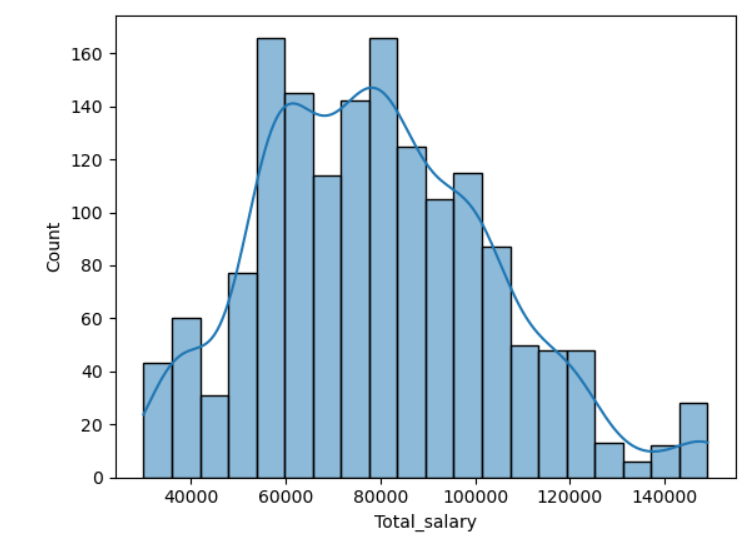
1. Salary



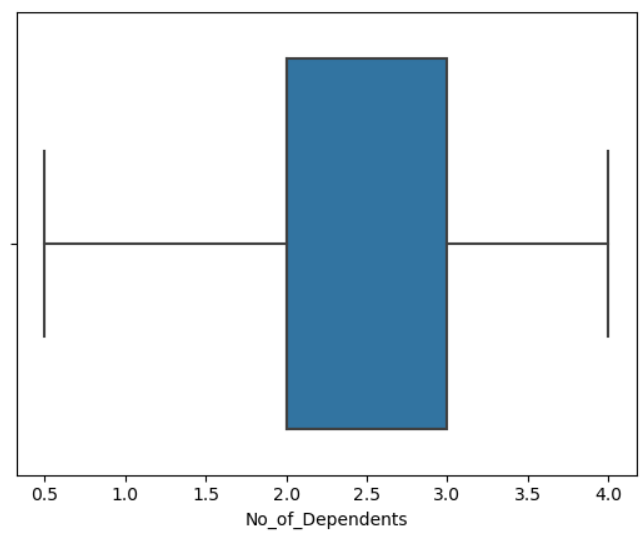


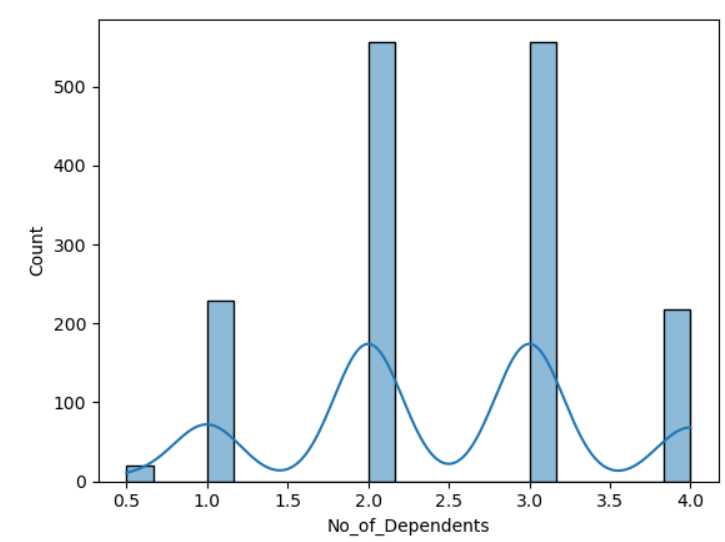
1. Total\_salary



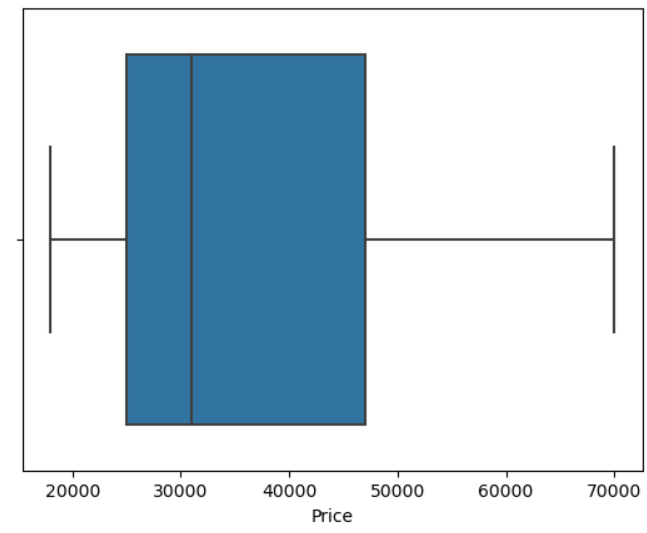


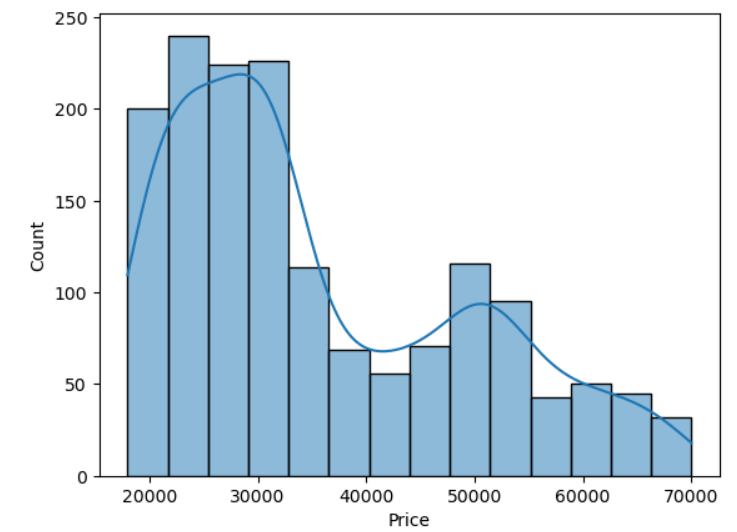
1. No\_of\_Dependents





1. Price

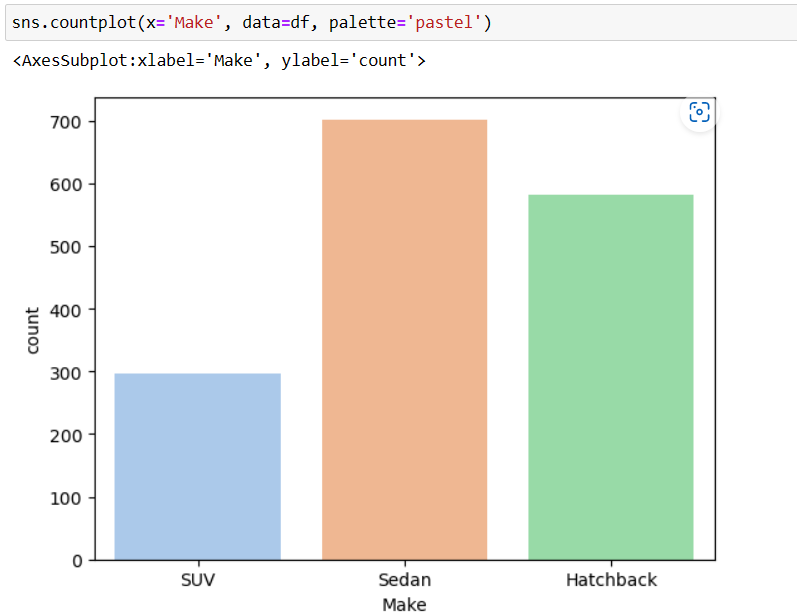
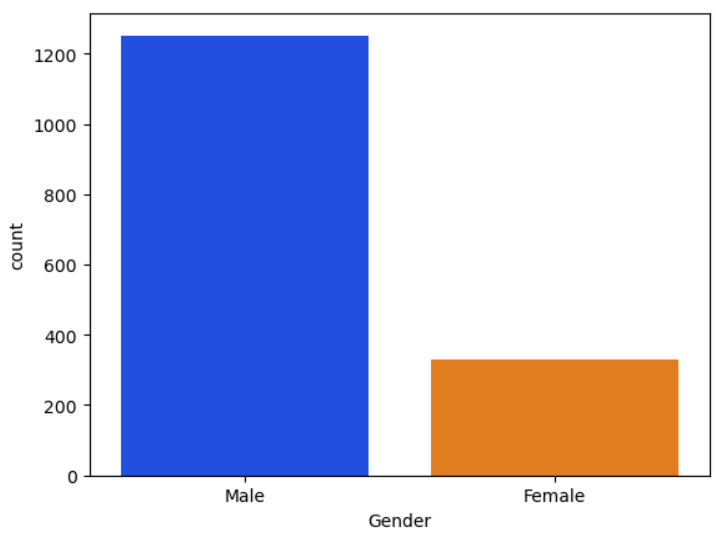
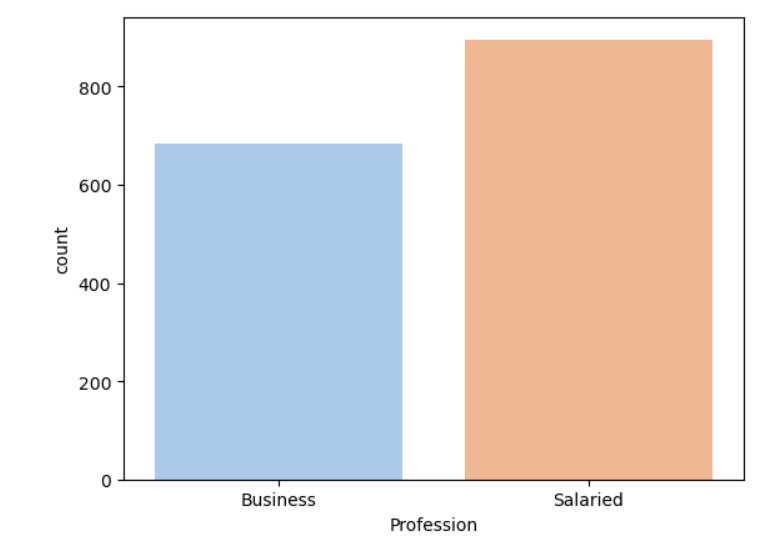
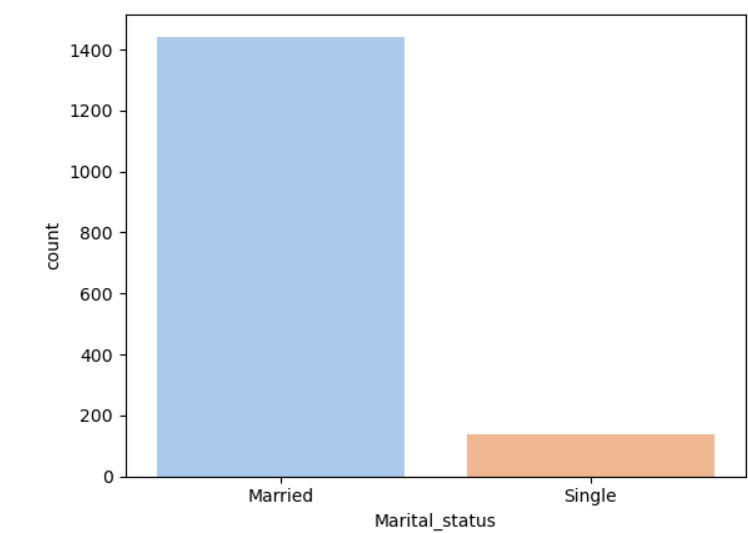
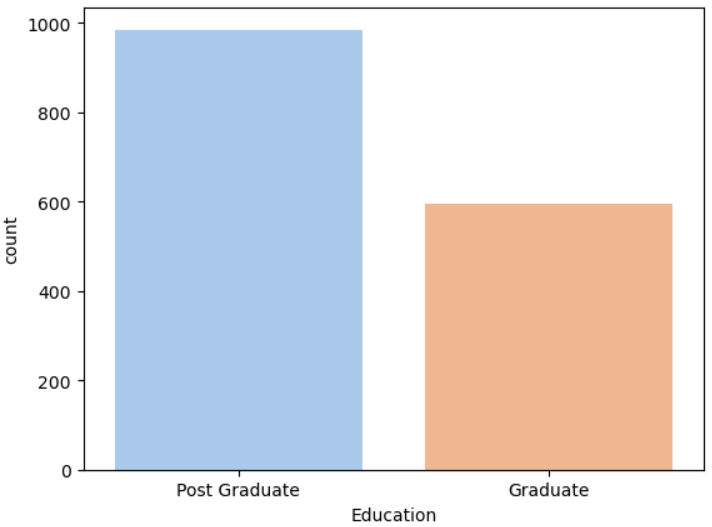
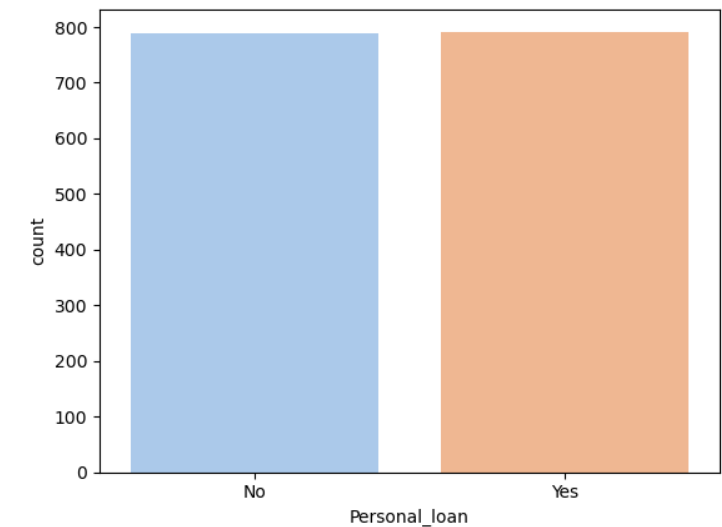
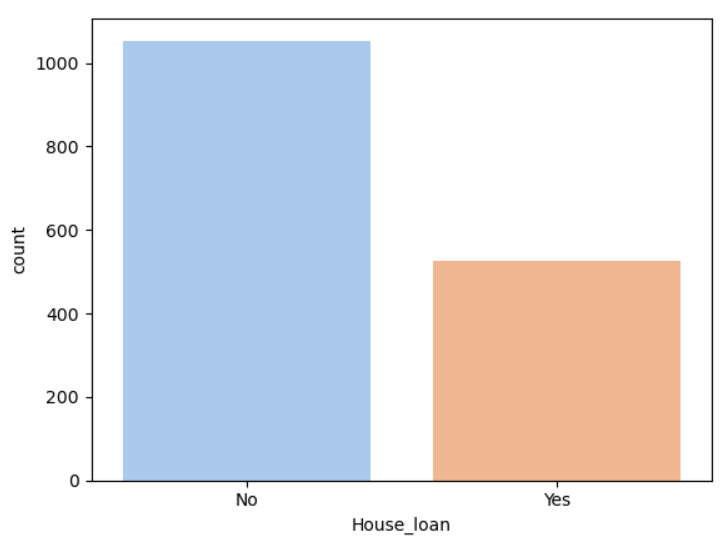
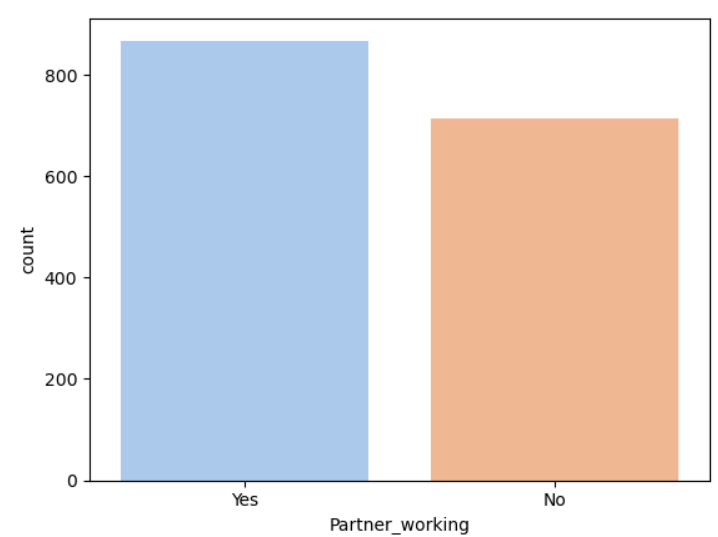




**Insights**

1. Age has multimodal distribution and has a positive skewed graph.
2. Salary has multimodal distribution and has zero skewness.
3. Total\_salary has no outliers after outlier treatment and has normal distribution.
4. Price has bimodal distribution with positive skewness
5. No\_of\_Dependents has multimodal distribution.

Univariate Analysis of Categorical fields –

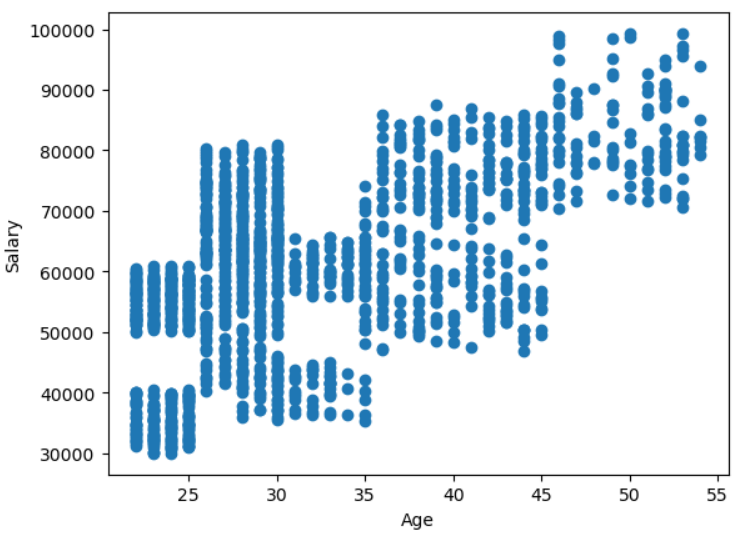
1. 
2. 
3. 
4. 
5. 
6. 
7. 
8. 

**INSIGHTS**

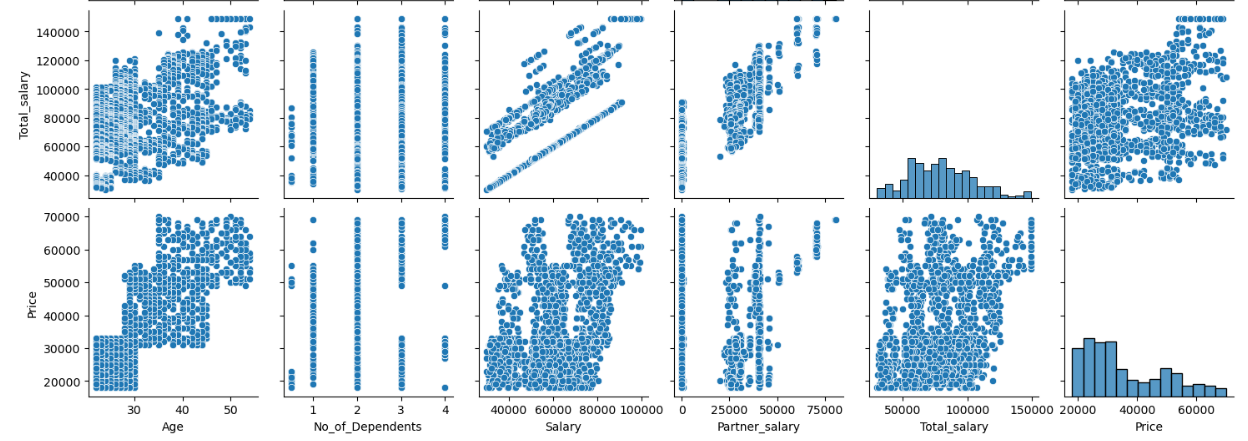
1. SEDAN is preferred more followed by SUV and then Hatchback.
2. Males are more than Females.
3. Count of Salaried people is more than Business people.
4. The data consists of very small proportion of Single customers when compared to married customers.
5. Most of the people are Post Graduate.
6. Few people have taken personal loan.
7. Working people are more than non working people.
8. **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.**

**Bivariate analysis of Numerical variables:**

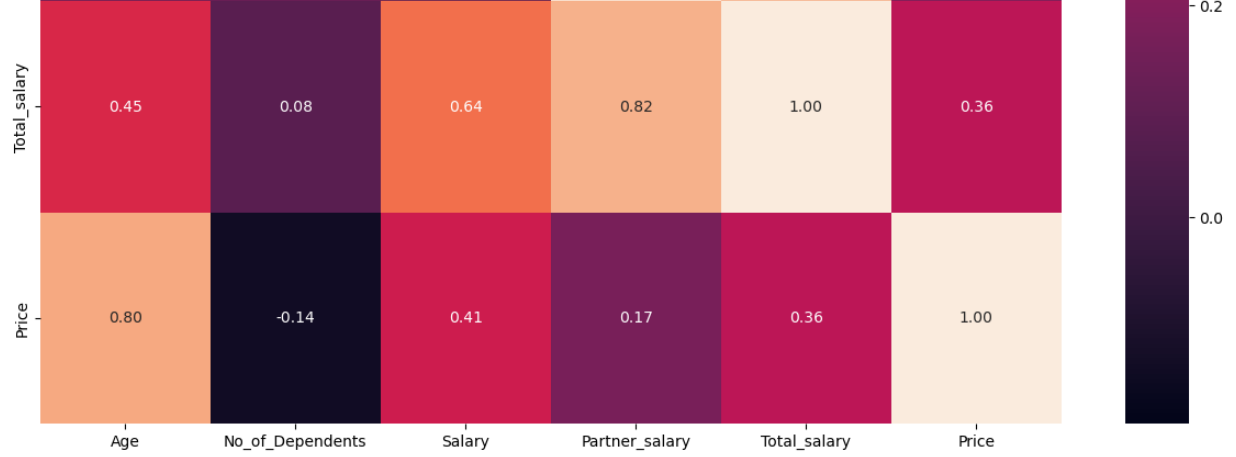
1. Numeric Vs Numeric

****

****

****

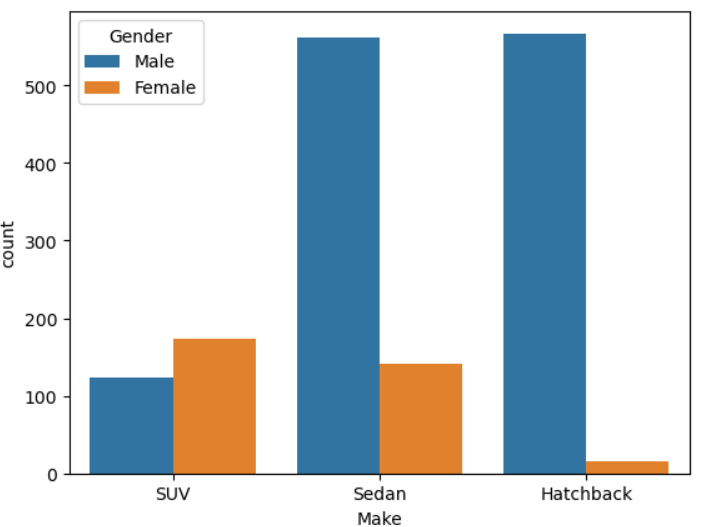
****

****

**Insight**

1. There is hardly linear relationship between Age and Salary. But we can say that as Age increases Salary for some of the customer also increases.
2. There is positive correlation between Price and Age.

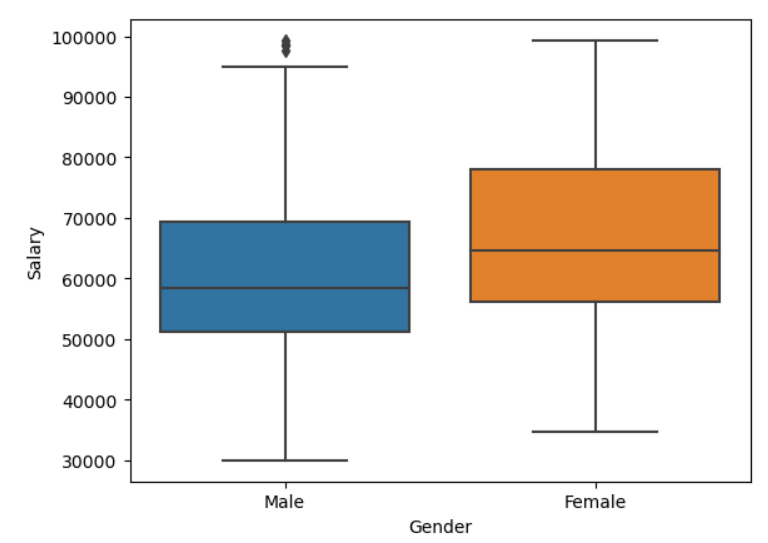
**Categorical vs Categorical**

****

**Insights**

1. We can see that Males prefer both Sedan and Hatchback equally. Females prefer SUV over other cars.

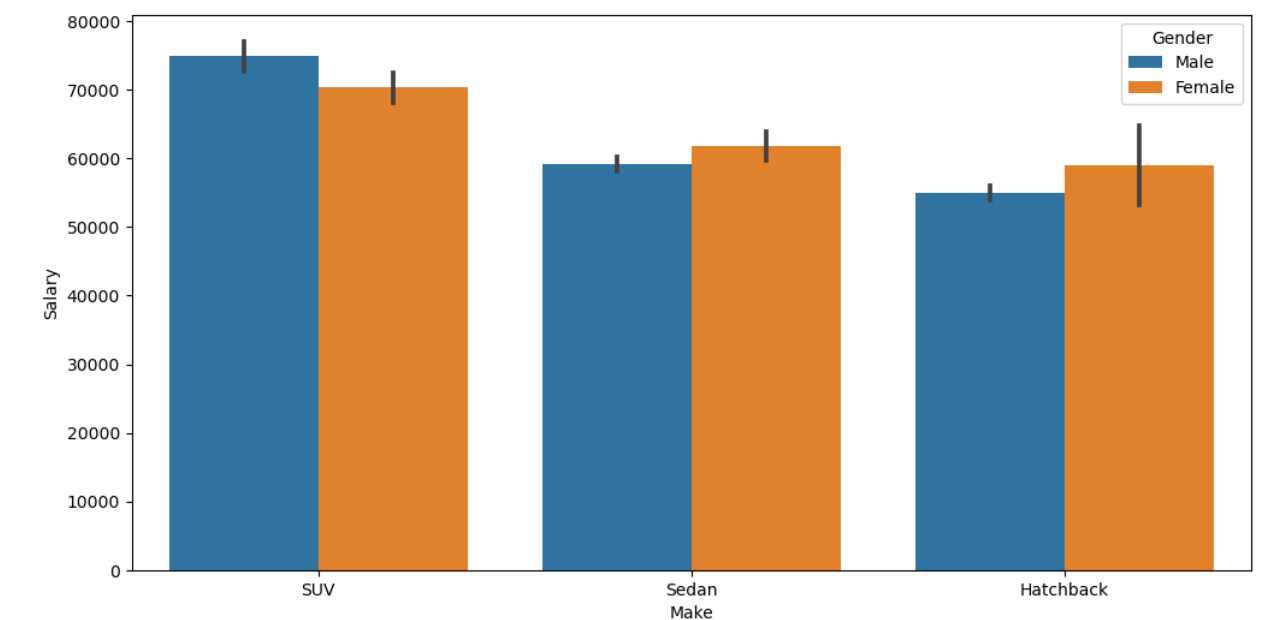
**Categorical vs Numeric**

****

**Insights**

**1)**Females have higher salary than males.

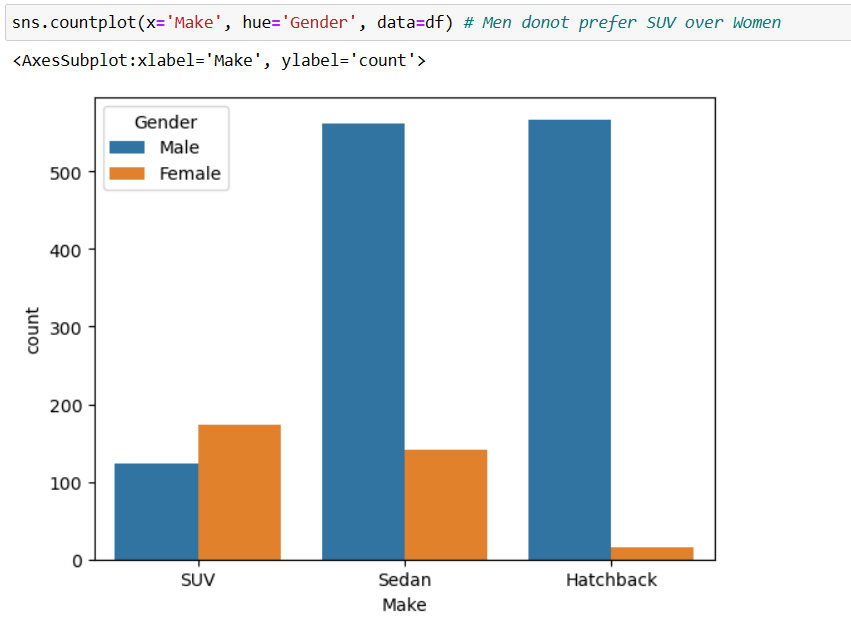
**#Multivariate Analysis**

****

We can see that Males who have higher salary prefer SUV more than any other car and Females with higher salary also prefer SUV.

1. **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”**

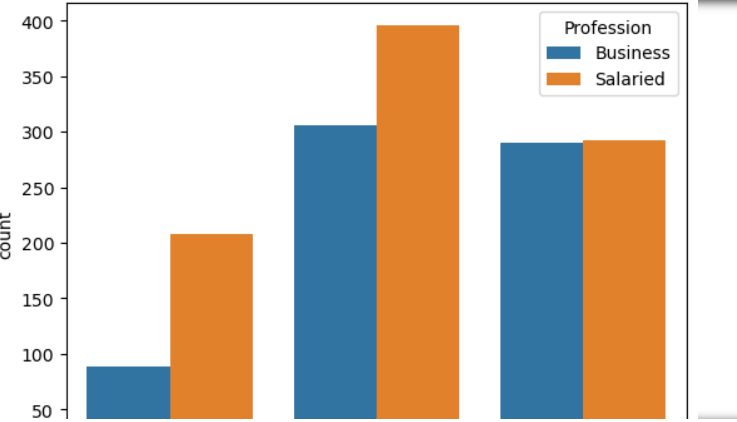


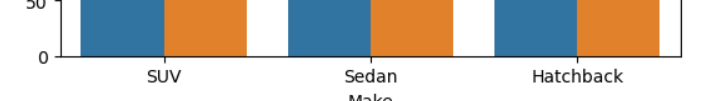
Analysing the ratio of SUV purchases for both the Genders

Men do not prefer SUV by larger margin. Infact mostly prefer Sedan and Hatchback.

Females prefer SUV more as compared to Men.Hence the Steve Roger Statement is not correct

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

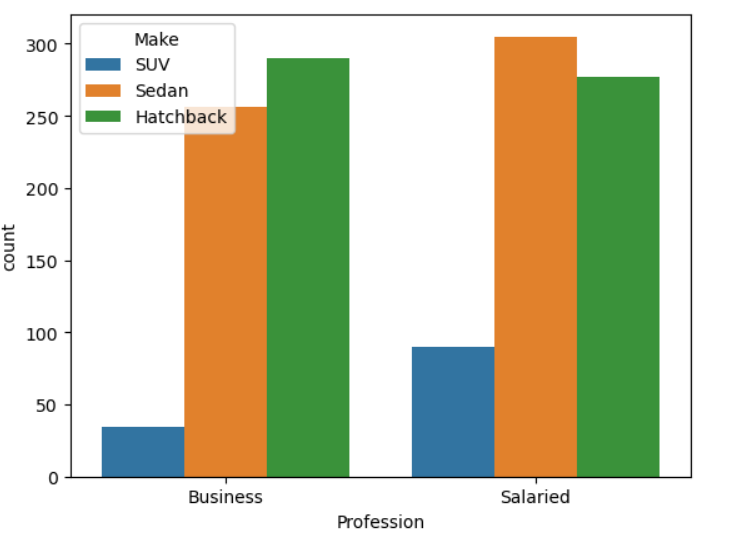




We can see that Salaried person prefer Sedan followed by Hatchback and then SUV same as business person. So Ned Stark is absolutely correct.

**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.**

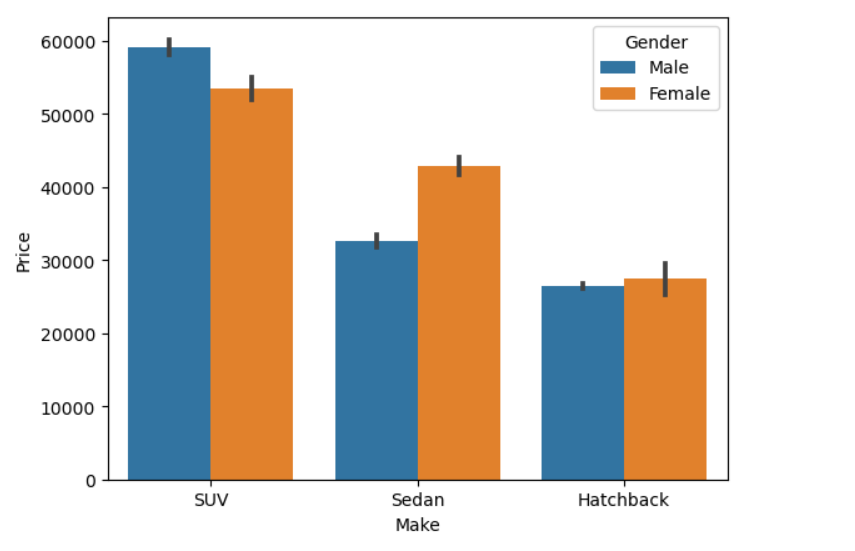
1. For Males



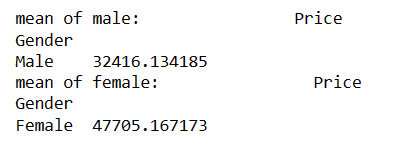
From above plot visualization it is clear that the salaried Male prefer Sedan over other cars. X axis has Profession and Make is a hue parameter.So Sheldon cooper is wrong.

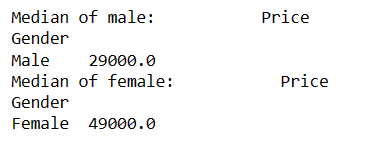
1. **From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a Business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions. Give justification along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

**\*\*\*F1) Gender**

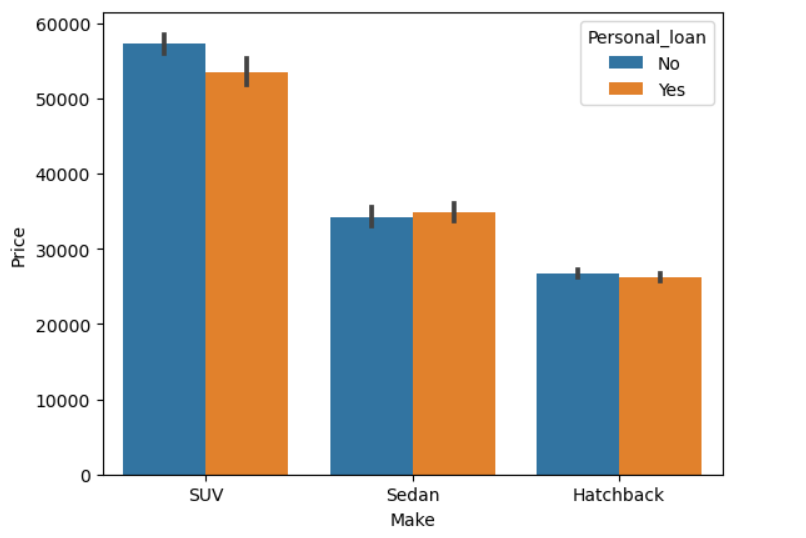


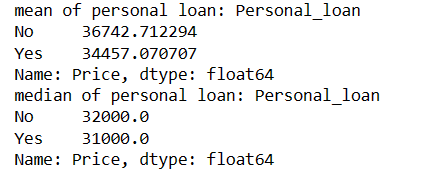
Females are likely to buy SUV and spends more on car buying.





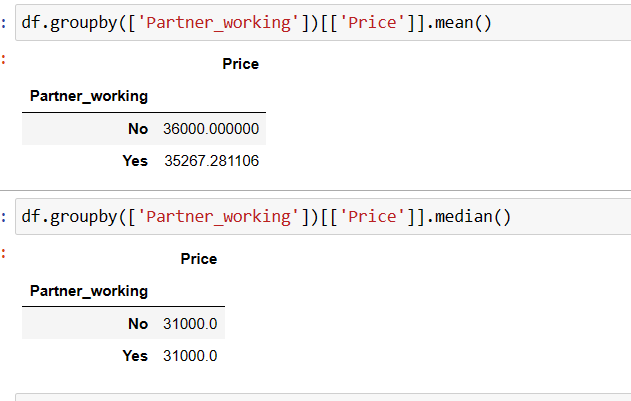
**F2) Personal\_loan**

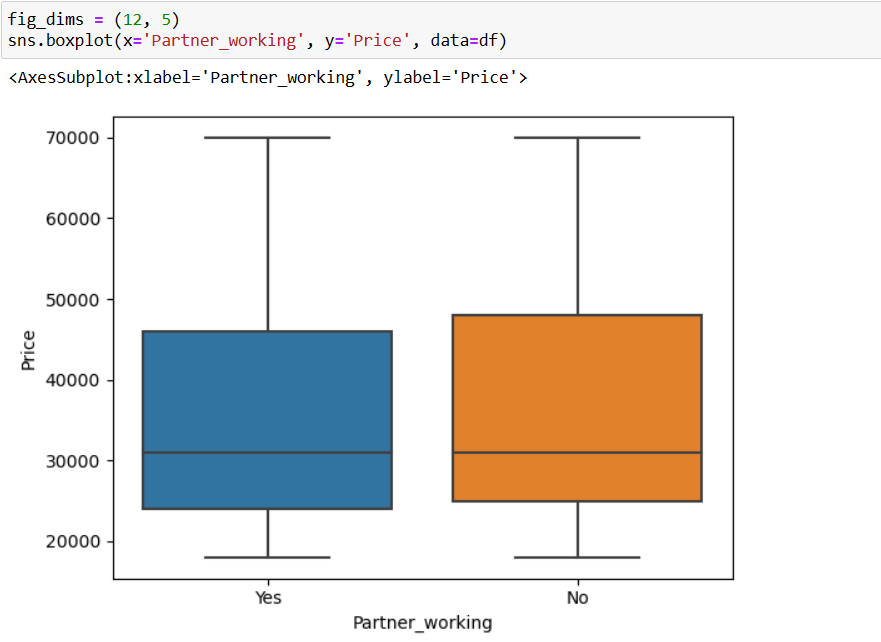




Customer who did not take Personal loan prefer SUV and spends higher money on it.

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

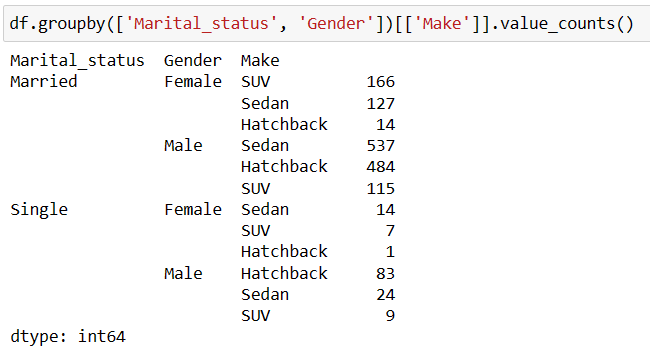


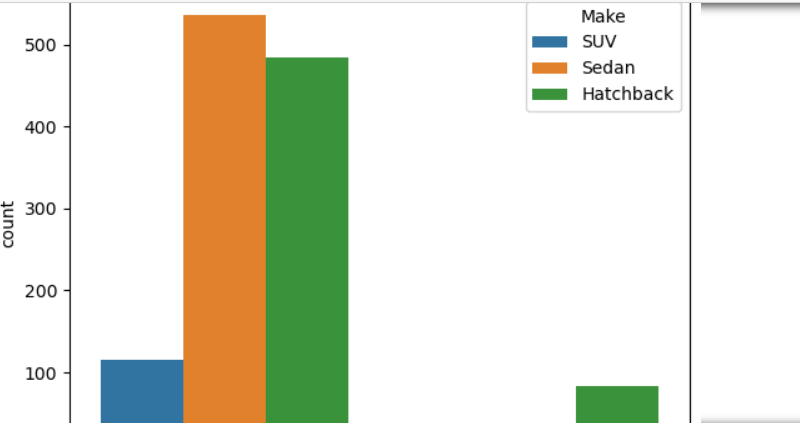


The Mean and Median price of the purchased automobile is almost similar across the Partner\_working category, thus indicating that partner working or not has no effect on the Purchase made by the customer. With visualization we can see that there is slight difference in working and non working partner which is negligible.

1. **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 Gender and Marital\_status - fields to arrive at groups with similar purchase history.**

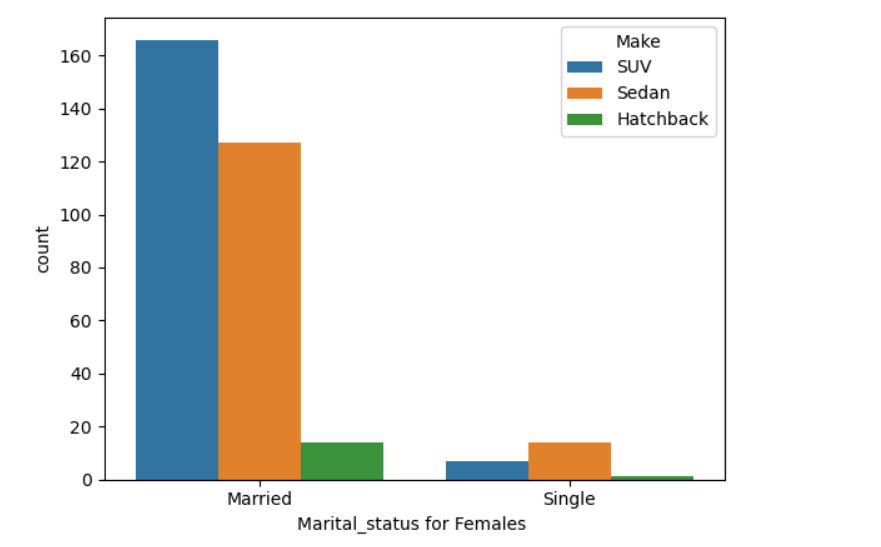
For marketing strategy, we can use the historic data and determine the type of Car that was most frequently purchased by the group of buyers based on the two variables Gender & Marital\_status.



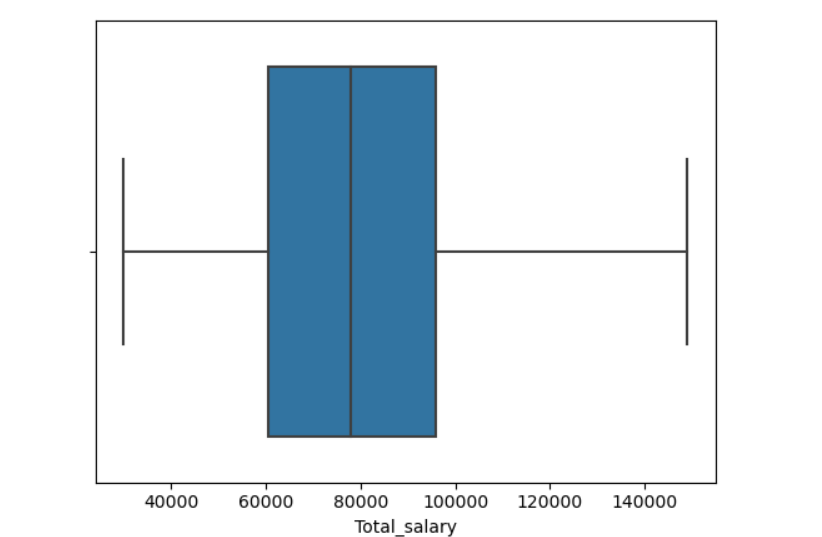


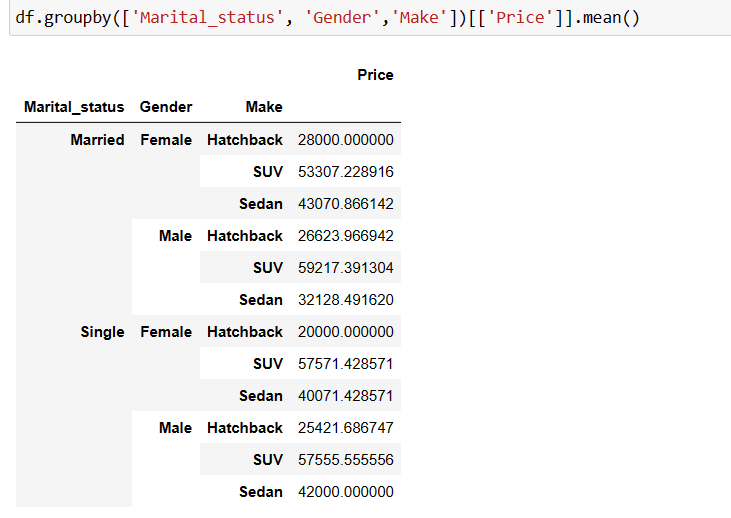


.



1. We can say that Maried Females prefer SUV over any other car.
2. Single Female prefer Sedan over any other car.
3. Married Men prefer Sedan.
4. Single Men prefer Hatchback over any other car.



1. Mean Price for purchases made by Married Females for Hatchback is 28000,for SUV is 53307and for Sedan is 43070.
2. Mean Price for purchases made by Married Males for Hatchback = 26623,for SUV is 59217 and for Sedan is 32128.
3. 
4. From plot above it is very clear that Mode of the Car make for Gender and Marital\_status fields,we get that Male Married & Female Married prefer SUV.

Problem 2

A bank can generate revenue in a variety of ways, such as charging interest, transaction fees and financial advice. Interest charged on the capital that the bank lends out to customers has historically been the most significant method of revenue generation. The bank earns profits from the difference between the interest rates it pays on deposits and other sources of funds, and the interest rates it charges on the loans it gives out.

GODIGT Bank is a mid-sized private bank that deals in all kinds of banking products, such as savings accounts, current accounts, investment products, etc. among other offerings. The bank also cross-sells asset products to its existing customers through personal loans, auto loans, business loans, etc., and to do so they use various communication methods including cold calling, e-mails, recommendations on the net banking, mobile banking, etc.

GODIGT Bank also has a set of customers who were given credit cards based on risk policy and customer category class but due to huge competition in the credit card market, the bank is observing high attrition in credit card spending. The bank makes money only if customers spend more on credit cards. Given the attrition, the Bank wants to revisit its credit card policy and make sure that the card given to the customer is the right credit card. The bank will make a profit only through the customers that show higher intent towards a recommended credit card. (Higher intent means consumers would want to use the card and hence not be attrite.)

**Problem 2 Question: ( Analyze the**[**dataset**](https://olympus.mygreatlearning.com/courses/87947/files/7448235/download?verifier=UAWiCLhr9BLURkbucshg9bYAql537cF4zNOHF9uI&wrap=1)**and list down the top 5 important variables, along with the business justifications. (10 Points) Data Dictionary -**[**Link**](https://olympus.mygreatlearning.com/courses/87947/files/7448234/download?verifier=wRlRiHoMsVsd5ahcjhHoTCCDOBqXgmS3mmmm2hR5&wrap=1)**)**

bank\_vintage: Vintage with the bank (in months) as on Tth month

T+1\_month\_activity: Whether customer uses credit card in T+1 month

T+2\_month\_activity: Whether customer uses credit card in T+2 month

T+3\_month\_activity: Whether customer uses credit card in T+3 month

T+6\_month\_activity: Whether customer uses credit card in T+6 month

T+12\_month\_activity: Whether customer uses credit card in T+12 month

Transactor\_Revolver: Customer who carries balances over from one month to the next.Transactor: Customer who pays off their balances in full every month.

avg\_spends\_l3m: Average credit card spends in last 3 months

Occupation\_at\_source Occupation recorded at the time of credit card application

cc\_limit Current credit card limit16

**Top 5 important variables are**

1. **cc\_limit:-** Credit card limit is decided by many factors like Income,payment history,expenses,debt. These are the factors used by bank to decide how much limit credit card should be given to customer. So this is considered important factor for Risk management and to sort out the list of defaulters.
2. **cc\_active30:-**Credit card account shows the usage of credit card by customer.It help bank decide the limit of the credit card. If customer is not using credit card then bank can cancel the card or reduce limit of credit card.
3. **annual\_income\_at\_source:-** Annaul income of a customer plays important role at Risk Management to the Bank. It can help bank study customer’s behaviour. Based on annual income bank decides how much limit should be given at credit card.
4. **T+1\_month\_activity :–** T+1\_month\_activity can be used to plan out campaigns and promotional offers so as to increase activity in the credit card.
5. **avg\_spends\_l3m:-**This variable is Average Credit card spends in last 3 months.This factor helps the bank to decide the limit of next credit card to a customer based on the average spends using existing credit card.This factor can help us identify if the card is primary or secondary card of a customer. If the customer spends more then it is his primary card and less than secondary card.We can also add good customers to campaign.

**Least used variable**

1. **userid**
2. **Card\_no**