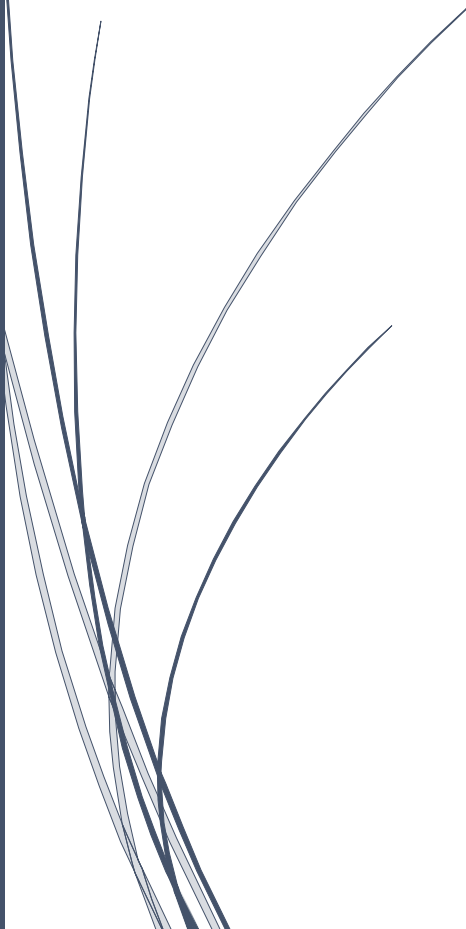


09/18/2022

Statistical Methods for Decision Making Project



Contents

Problem 1 2

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 an analytics professional to improve the existing campaign.

1. A. 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)3
1. 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?.....3
1. C. Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.....5
1. 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.....6
1. 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"
 E2) Ned Stark believes that a salaried person is more likely to buy a Sedan.
 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.....7
1. 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
 F2) Personal loan.....8
1. G. From the current data set comment if having a working partner leads to the purchase of a higher-priced car.....9
1. 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.....10

Problem 2..... 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.....10

Problem 1

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.

1. A. 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)

Ans:

- From the basic analysis we found total 22134 elements are present in the Data Frame.
- We also found the Data type information of the Data Frame and we have 8 Objects, 5 Integers and 1 Float Data Type present in the Data Frame.
- Then we check the total number of rows and columns present in the Data Frame and we found out we have 1581 rows and 14 columns.
- We have checked for 5 number summaries using “describe ()” function to find Max, Min, 1st quartile, 3rd quartile and standard deviation.

1. 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?

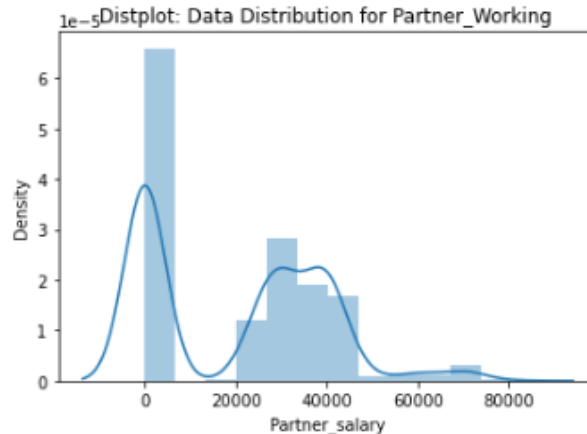
Ans:

- First, we will check for the NULL values and from the analysis we can identify that we have 53 missing values for “Gender” column and 106 missing values for “Partner_Salary” Column.

```
Out[8]:
```

Age	0
Gender	53
Profession	0
Marital_status	0
Education	0
No_of_Dependents	0
Personal_loan	0
House_loan	0
Partner_working	0
Salary	0
Partner_salary	106
Total_salary	0
Price	0
Make	0
dtype:	int64

- Then we will test for the skewness for the “Partner_Salary” Column and from observation we found mentioned column is Skewed, so to impute the missing values we can use the ‘Median’ Value.



- We imputed “Partner_Salary” column with median value to proceed further with the data set.

Median value for "Partner_salary" : 25600.0

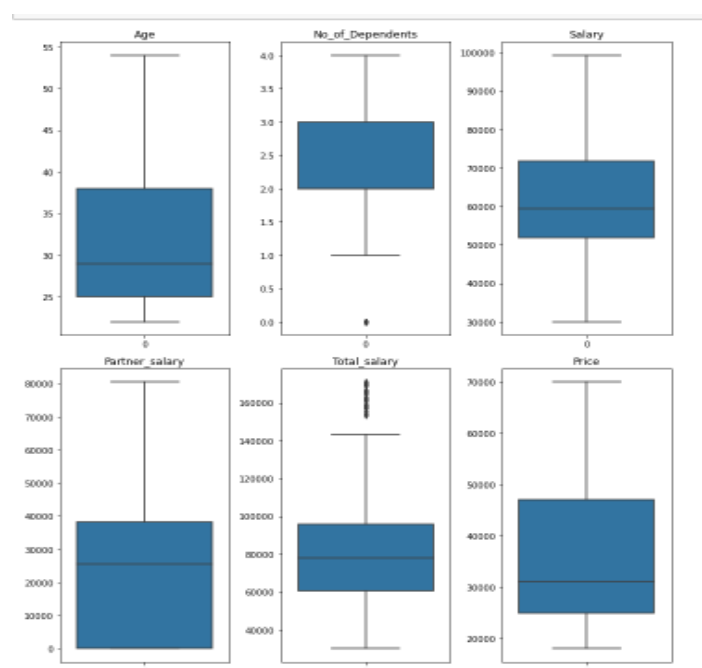
- We found no duplicate values in the Data-set.

Total number of duplicate values : 0

- We have checked the unique values for each and every column present in the data set to check for the bad values and discrepancies. We found some of the columns having some bad data.

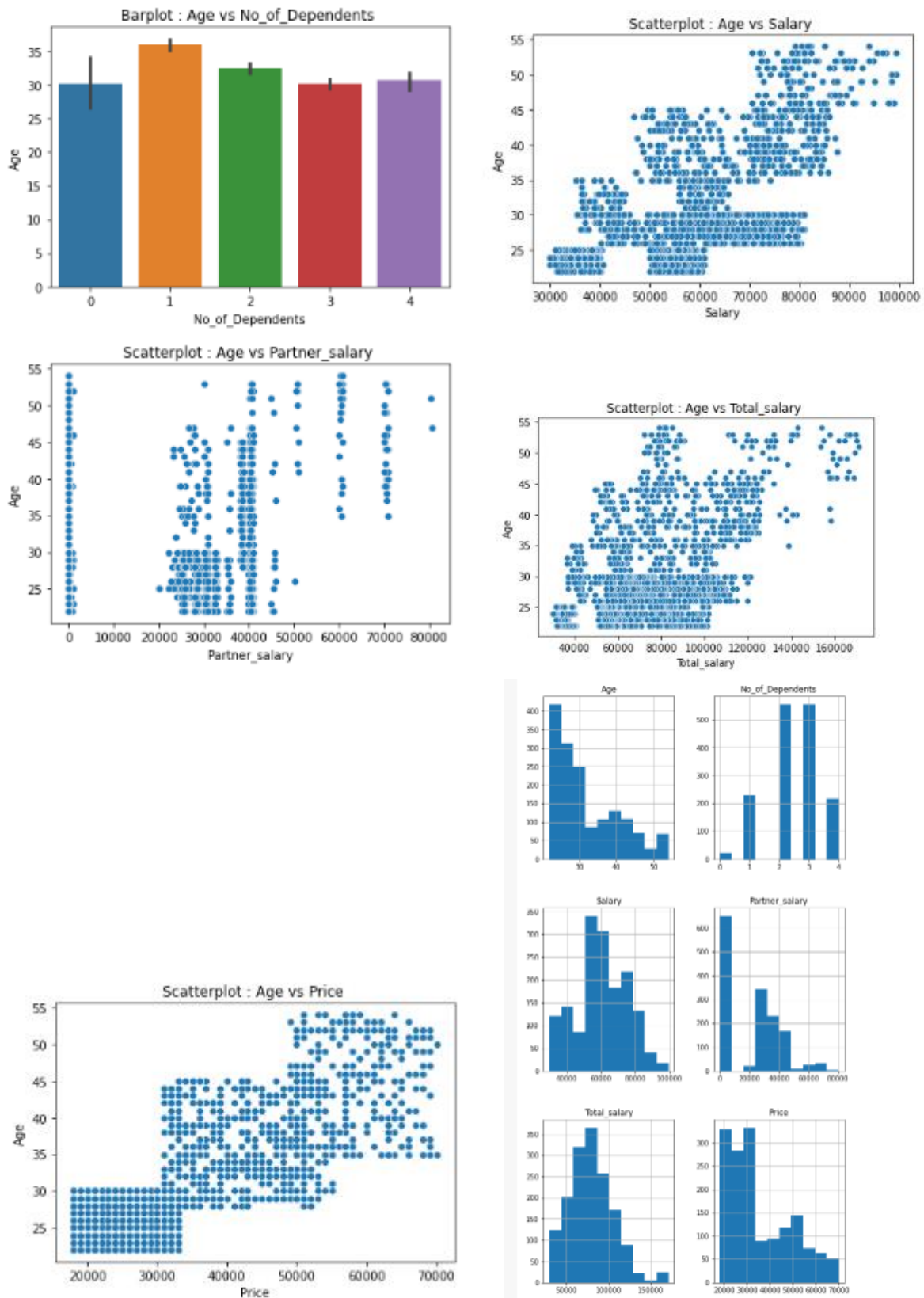
Unique values for Gender : ['Male' 'Femal' 'Female' nan 'Femle']

- We have checked the continuous variables for outliers, and from analysis we can say we have outliers in our data-set. “Total_salary” column and “No_of_Dependents” has outliers present in it. We can consider "No_of_Dependents" with no outliers as 0 value indicates having no dependent, for column "Total_salary" we need not to do outlier treatment as some of them can have a higher salary if their partner is also earning.



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

Ans:



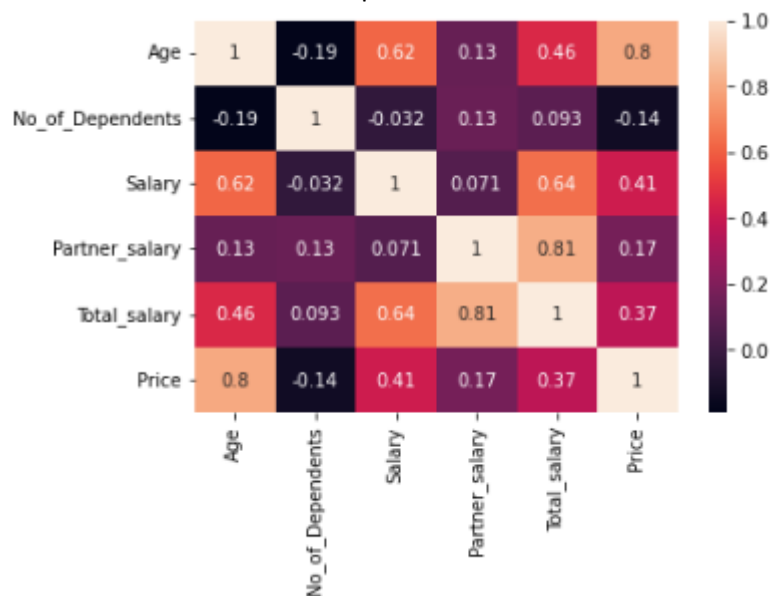
- From the analysis we can conclude Age and Price has positive co-relation, Age and Total Salary has some co-relation. Age and salary have a positive co-relation.

1. 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.

Ans:



- We also checked for heatmap to check the co-relation:



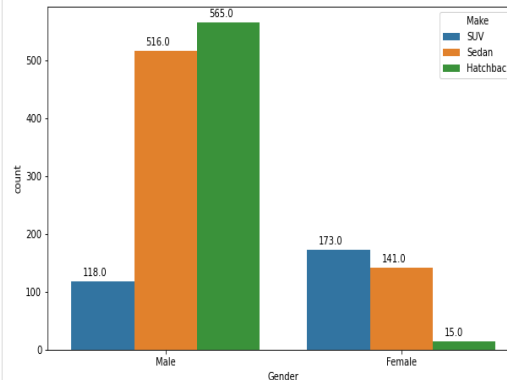
- From the analysis we can conclude that Age and Salary has positive co-relation.
- From the analysis we can conclude that Age and Price has positive co-relation.
- From the analysis we can conclude that Salary and Total_salary has positive co-relation.
- From the analysis we can conclude that Partner_Salary and Total_salary has positive co-relation.

1. 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:” From the projection we can clearly see Female count for SUV is more than men, so we can strongly disagree with Steve’s statement.”

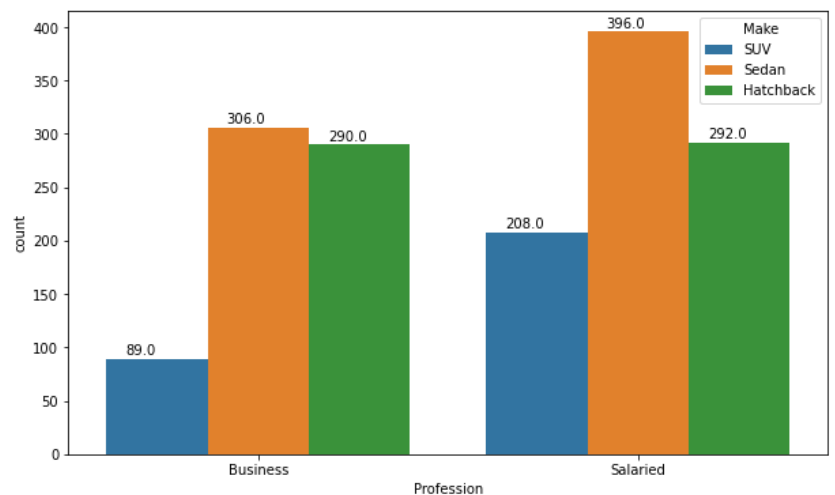
```
Gender  Make
Female  Hatchback    15
        SUV         173
        Sedan        141
Male    Hatchback    565
        SUV         118
        Sedan        516
dtype: int64
```



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

Ans:” From the analysis we can conclude that a salaried person is more likely to buy a Sedan. Ned is correct about his believe.”

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



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: "From the projection we can see male count for "Sedan" is more than that of "SUV", so we strongly agree with the statement."

```

Profession  Gender  Make
Business   Female  SUV      55
           Female  Sedan    50
           Male   Hatchback 289
           Male   SUV      33
           Male   Sedan    237
Salaried    Female  Hatchback 15
           Female  SUV      118
           Female  Sedan    91
           Male   Hatchback 276
           Male   SUV      85
           Male   Sedan    279

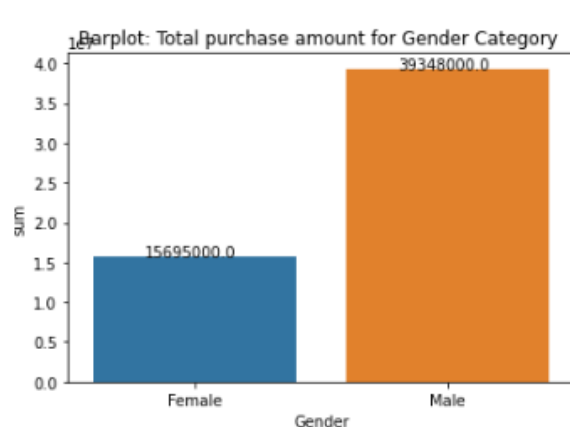
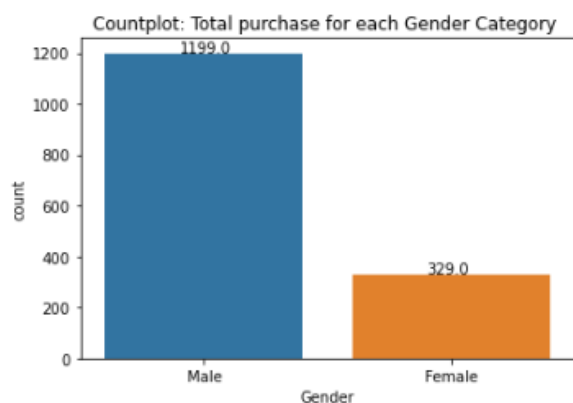
dtype: int64

```

1. 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.

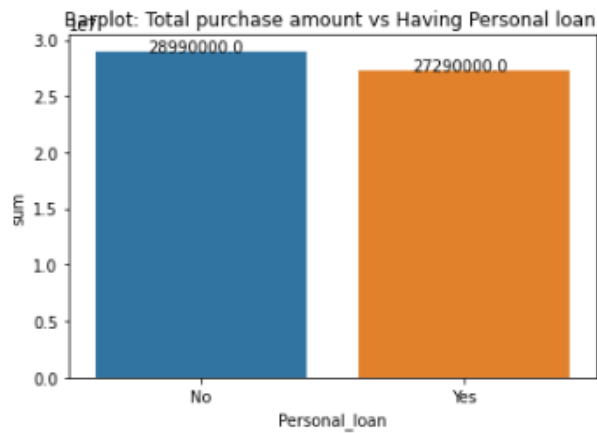
F1) Gender

Ans: "From the above analysis we can say the purchasing capability of Men is higher."



F2) Personal loan

Ans: “From the above analysis we can conclude that not having a personal loan lead to the higher purchase capability of an auto-mobile cars and having a personal loan lead to lesser purchase capability.”



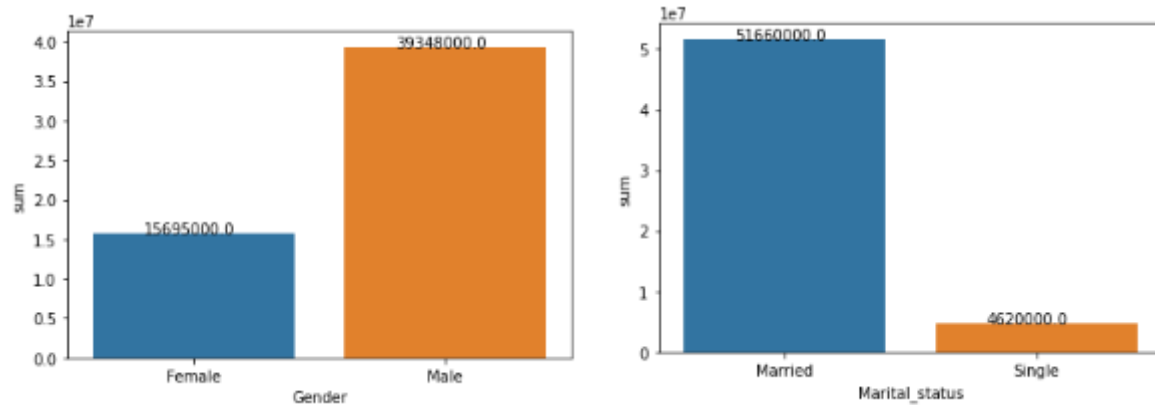
1. G. From the current data set comment if having a working partner leads to the purchase of a higher-priced car.

Ans: “From the above analysis we can conclude having a working partner leads to the purchase of a higher-priced car.”



1. 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 : So, from the analysis we can say married people buys most cars, for single people they are less interested in cars compared to married once. The Females who are single are least interested about cars.



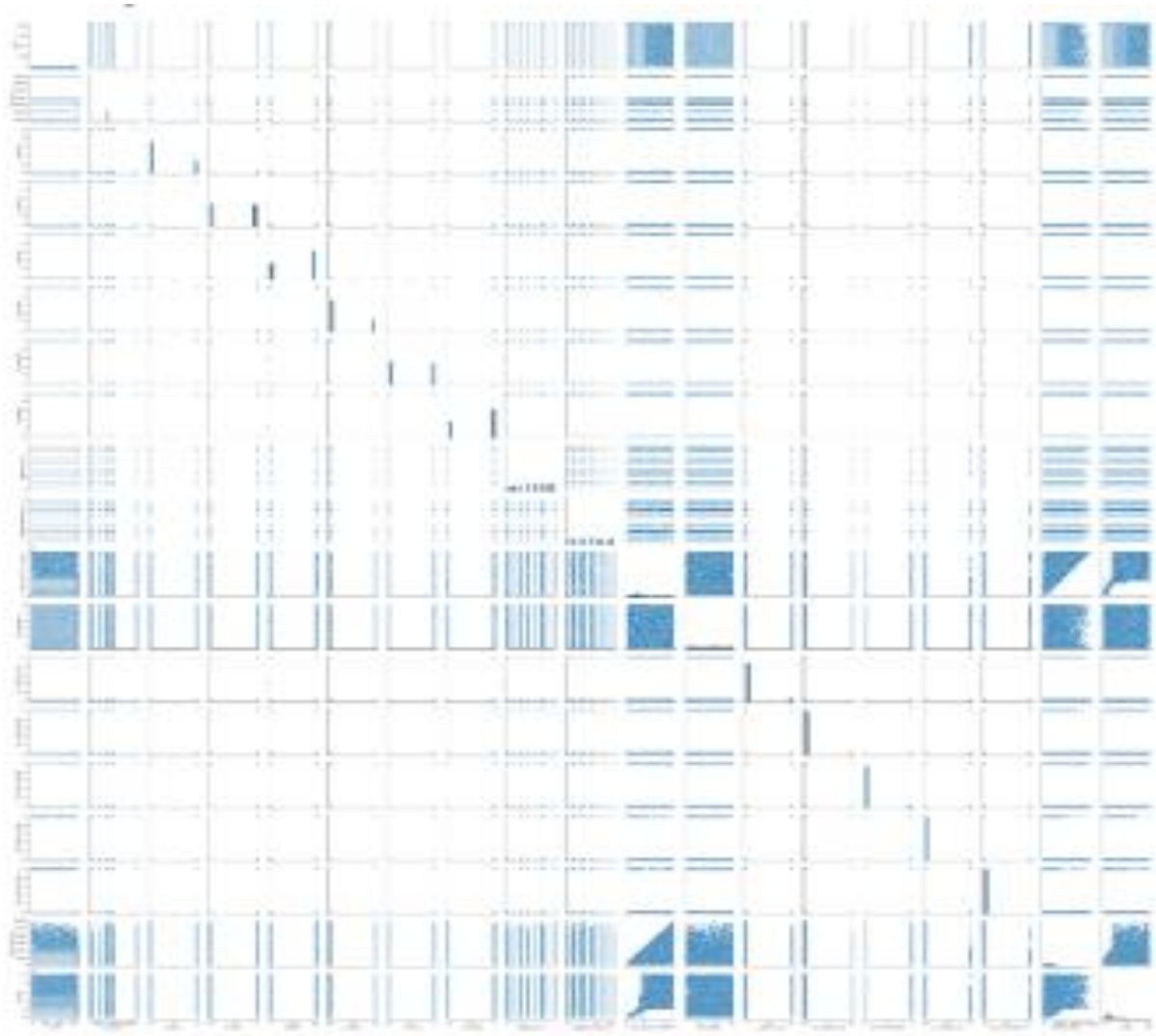
```
Gender  Marital_status
Female  Married         307
        Single          22
Male    Married       1088
        Single         111
dtype: int64
```

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.



Co-relation plot for all the variables to get an idea about their relation.

Our objective is to find the list down the top 5 important variables which could help us to get more valuable insights.

The top 5 important variables which could help us to get more valuable insights are : cc_limit, Occupation_at_source, avg_spends_l3m, other_bank_cc_holding, annual_income_at_source

#cc_limit: For customers who are holding credit card could be interested bying more if he/she gets more limit on purchase. We also keep in mind that we need to upgrade the limit to such limit where the customer could repay their spending.