BUSINESS REPORT
AUSTO MOTOR

October 2022

Table of Contents:

Problem Statement 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 an analytics professional to improve the existing campaign.

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

Revenue Generated	5.62 Cr	
Model	**Revenue per Make**	**Count of models sold**
Sedan	**2.43 cr**	**702**
SUV	**1.65 cr**	**297**
Hatchback	**1.54 cr**	**582**

- The total number of customer base is: 1581
- Total characteristics available in the base: 14

Features	Nature of the features	Type
Age	Age of the customers	Numeric
Gender	Gender of the customers	Categoric
Profession	Customers with different professions	Categoric
Marital_status	Marital_status of the customers	Categoric
Education	Education Level of the customers	Categoric
No_of_Dependents	No_of_Dependents in the family	Categoric
Personal_loan	Personal_loan accquired by the car owner(Y/N)	Categoric
House_loan	House_loan accquired by the car owner(Y/N)	Categoric
Partner_working	Partner working status(Y/N)	Categoric
Salary	Salary of the customer	Numeric
Partner_salary	Customers dependant's salary	Numeric
Total_salary	Addition of Partner Salary and Customer Salary	Numeric
Price	Revenue generated by selling each model	Numeric
Make	Categories of the sold cars	Categoric

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?**

- Below table gives us an idea about the unique information that each feature holds.

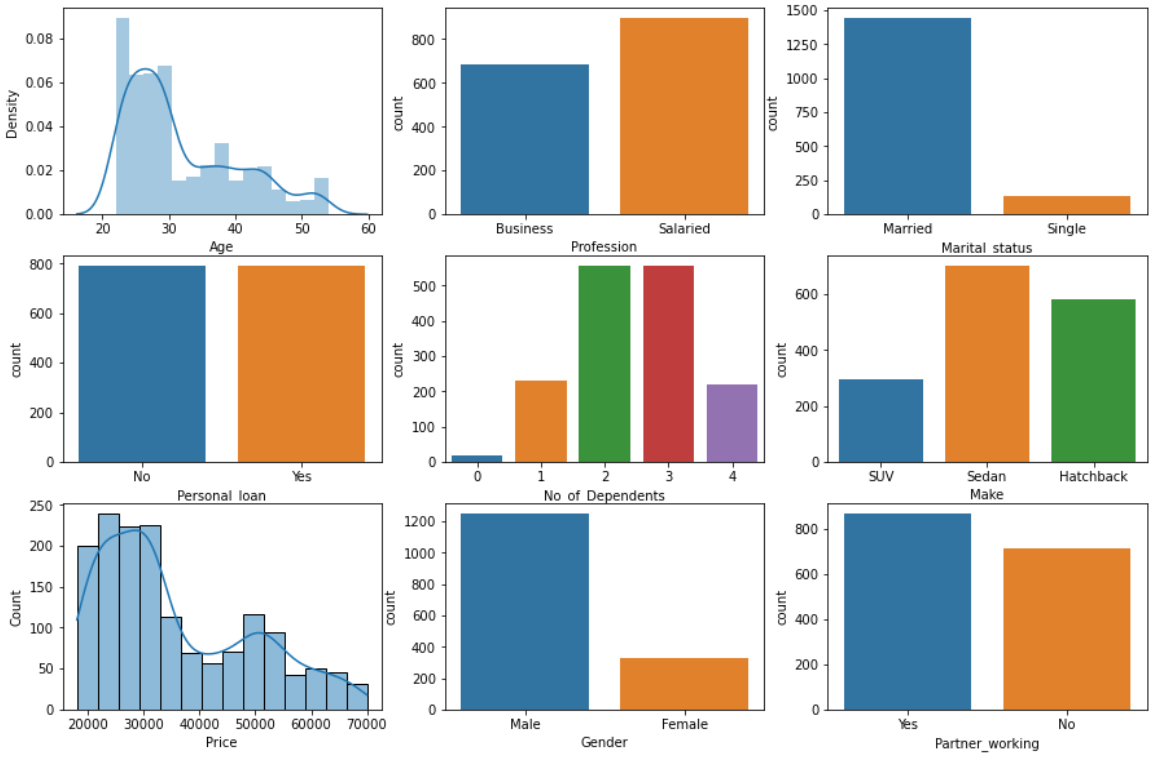
Features	Count of Individual features
Age	33
Gender	4
Profession	2
Marital_status	2
Education	2
No_of_Dependents	5
Personal_loan	2
House_loan	2
Partner_working	2
Salary	538
Partner_salary	149
Total_salary	754
Price	53
Make	3

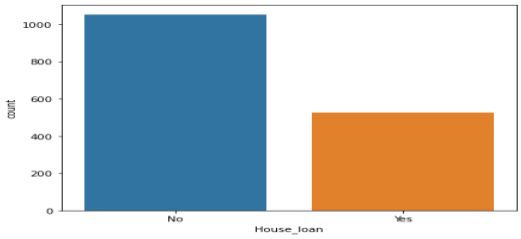
- After a primary quality check on the customer base, we analyzed that 2 features in the data have missing values:
1. 'GENDER': 53 missing values
2. 'PARTNER_SALARY': 106 missing values

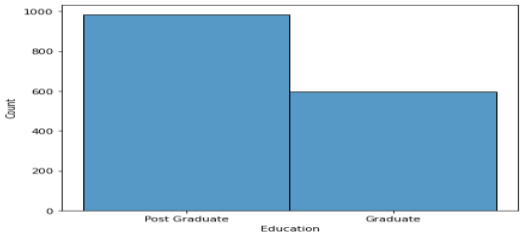
- Multiple typo/data entry errors observed in 'Gender'

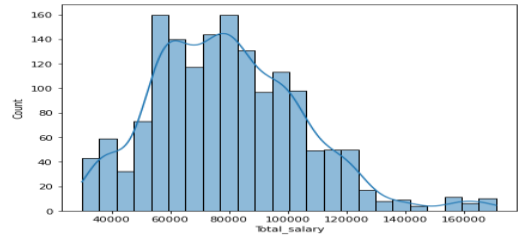
```
array(['Male', 'Femal', 'Female', nan, 'Femle'], dtype=object)
```

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







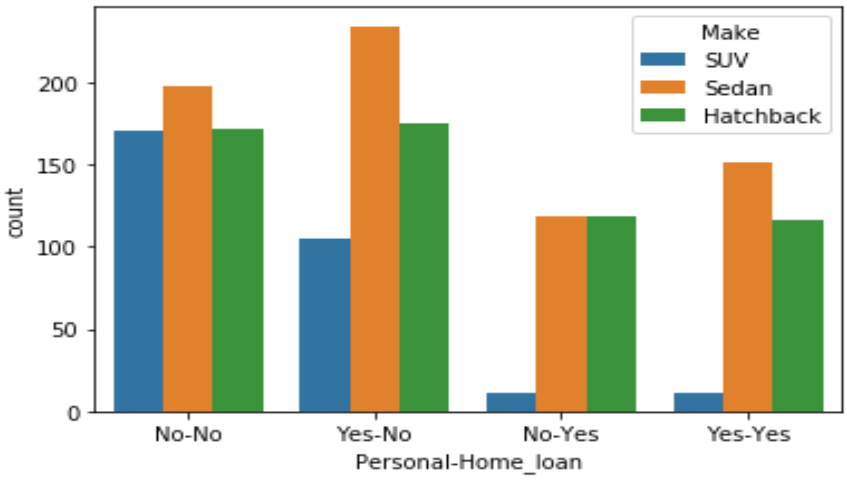


Insights:

- Customers with age group 30-60 have less contribution in revenue generation.
- SUV has the least selling count.
- Contribution of customers is very low where there are no dependents, or 1 or 4 dependents.
- Count of customers who are single and do not prefer a car is very low.
- We have a higher count of male customers than female.
- Graduates purchasing a car is low compared to Post Graduates.
- Dip is seen in count of cars sold after the price increase from 30000.
- Trend shows that low customer count is dipping after total salary of 80000.
- Less population observed of people with home loan, whereas personal loan is not affecting the count of cars sold.
- Less sale is observed with customers whose partners are not working.

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.

Different combinations of variables helped us conclude that:

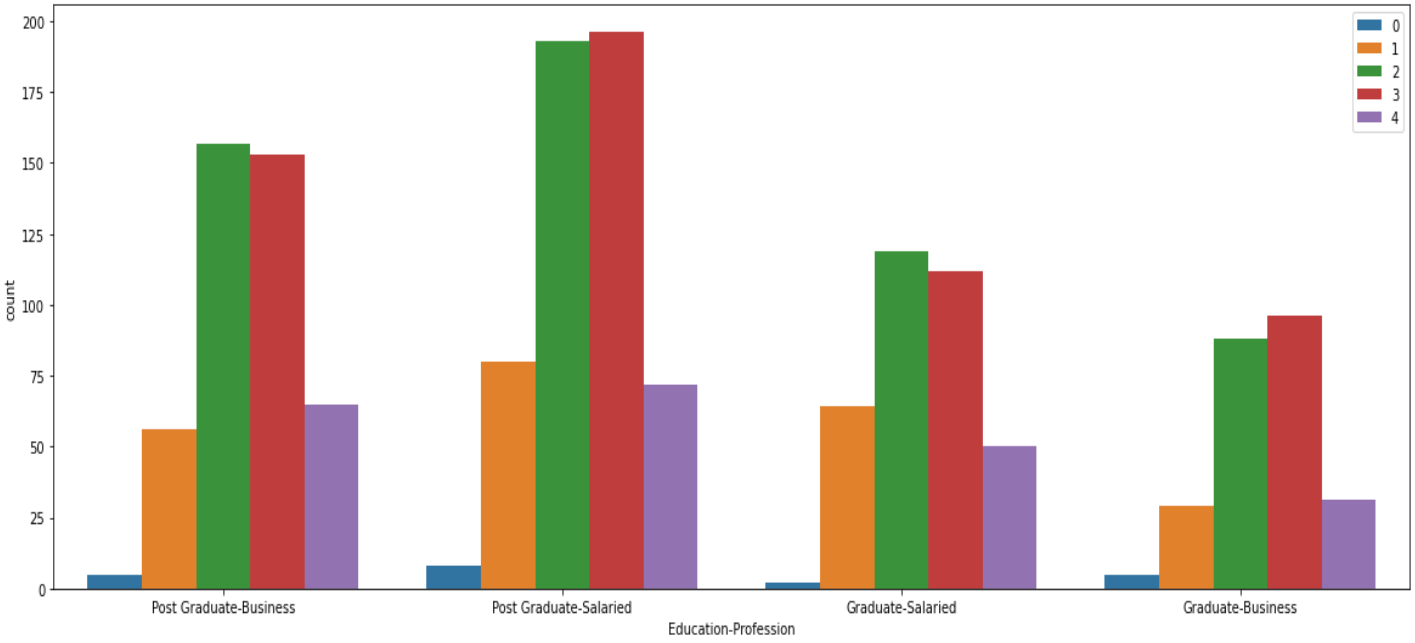


- Least sale is observed in SUV Model where customers have taken loan(both or any single loan)

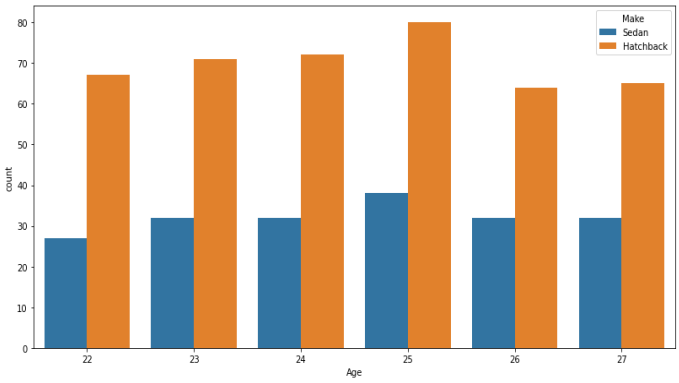
- Focus can be more selling Hatchbacks to the customers where we observe loan.

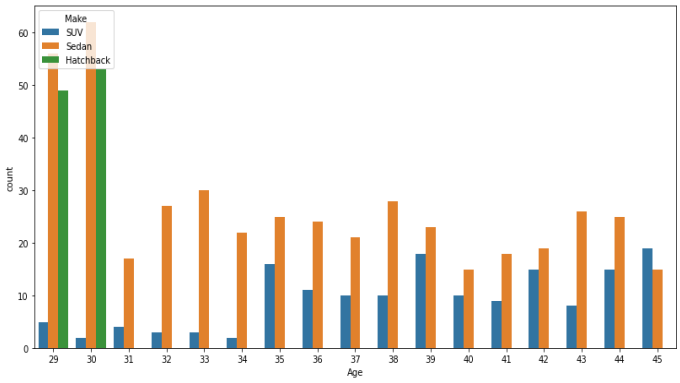
Make No_of_Dependents	Hatchback	SUV	Sedan
0	15	5	0
1	11	34	184
2	181	87	289
3	256	117	184
4	119	54	45

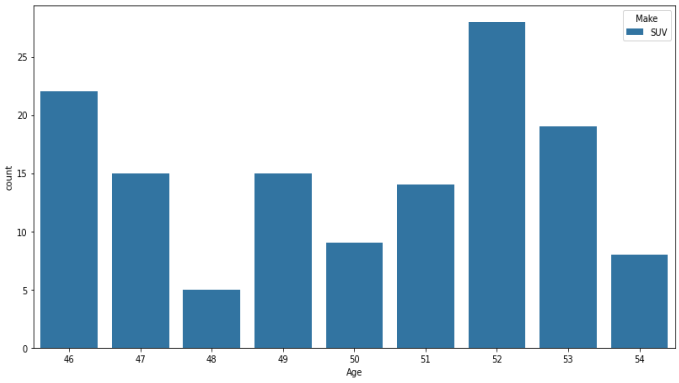
- Sales should be more of SUV's & Sedan's where customer's have 3 or more dependents.



- Sales is least amongst graduates, who have no dependents.
- Graduates owning a business can contribute in SUV Sales.







- Above graph infers that, people within the age bucket of 22-27 do not purchase SUV.
- No Hatchbacks are purchased by customers with age range 31-45.
- Higher models of SUV can be pitched to customers above 46 as they prefer buying a high-end car and 20 % of customers have salary above 1 lac.

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”

Below table infers, count of SUV’s sold to female is high.

Gender	Female	Male
Make		
Hatchback	15	567
SUV	173	124
Sedan	141	561

Steve Roger’s statement is FALSE.

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

Below table infers, a salaried person prefers SEDAN over other make models.

Profession	Business	Salaried
Make		
Hatchback	290	292
SUV	89	208
Sedan	306	396

Ned Stark's statement is TRUE.

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.
Table infers, sales of SUV cars is less to salaried male as compared to Sedan.
Sheldon Cooper's statement is FALSE.

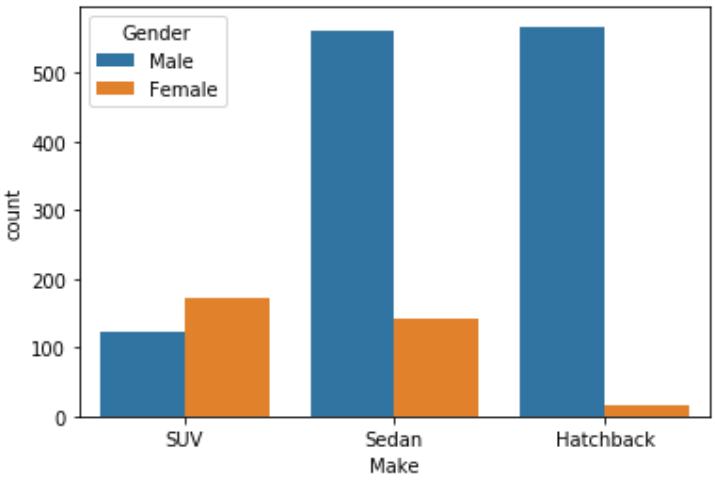
Make	Profession	Gender	Count
Hatchback	Business	Male	290
	Salaried	Male	277
		Female	15
SUV	Business	Female	55
		Male	34
	Salaried	Female	118
		Male	90
Sedan	Business	Male	256
		Female	50
	Salaried	Male	305
		Female	91

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.**

Table below infers the revenue generated by male and female across the models.

Gender	Make	Revenue
Female	Hatchback	4,12,000
	SUV	92,52,000
	Sedan	60,31,000
Male	Hatchback	1,49,59,000
	SUV	70,31,000
	Sedan	1,73,58,000

F1) Gender:
As per the graph, females have purchased less Hatchbacks and males have purchased least count of SUVs.



F2) Personal_loan: To check the cars taken by customers with loan we have considered below graphs which states:

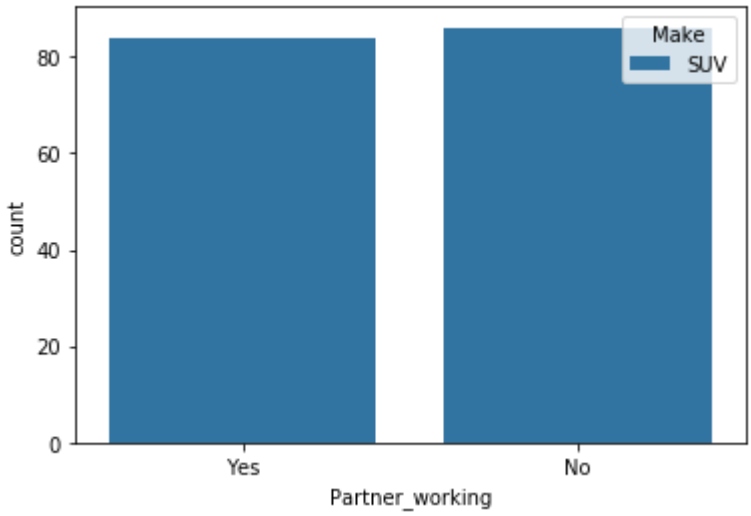
Personal_loan	Make	Count
No	Hatchback	291
	SUV	181
	Sedan	317
Yes	Hatchback	291
	SUV	116
	Sedan	385

Insights

• With Personal loans, people are buying Sedan Models.
• Sales of Hatchback has no difference with the loan status.
•On the overall SUV sales, population is more where people have not taken personal loan.

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

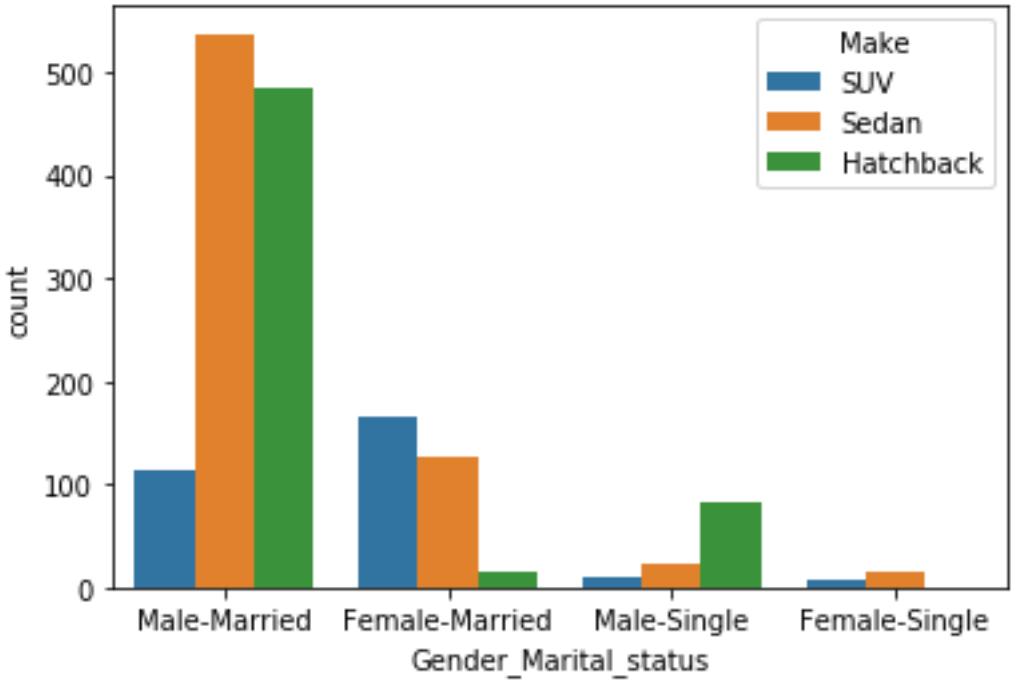
As per the given data, SUVs are the high-priced.



- Not much difference is observed in the sale of cars if the partner is working or not.

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.

For better insights, features gender and married are grouped in the below graph which states:



- Focus should be on married male for SUV sale campaign, considering the increase in dependency.
- Female single group can help improving the purchase of Hatchbacks.

Problem 2:

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

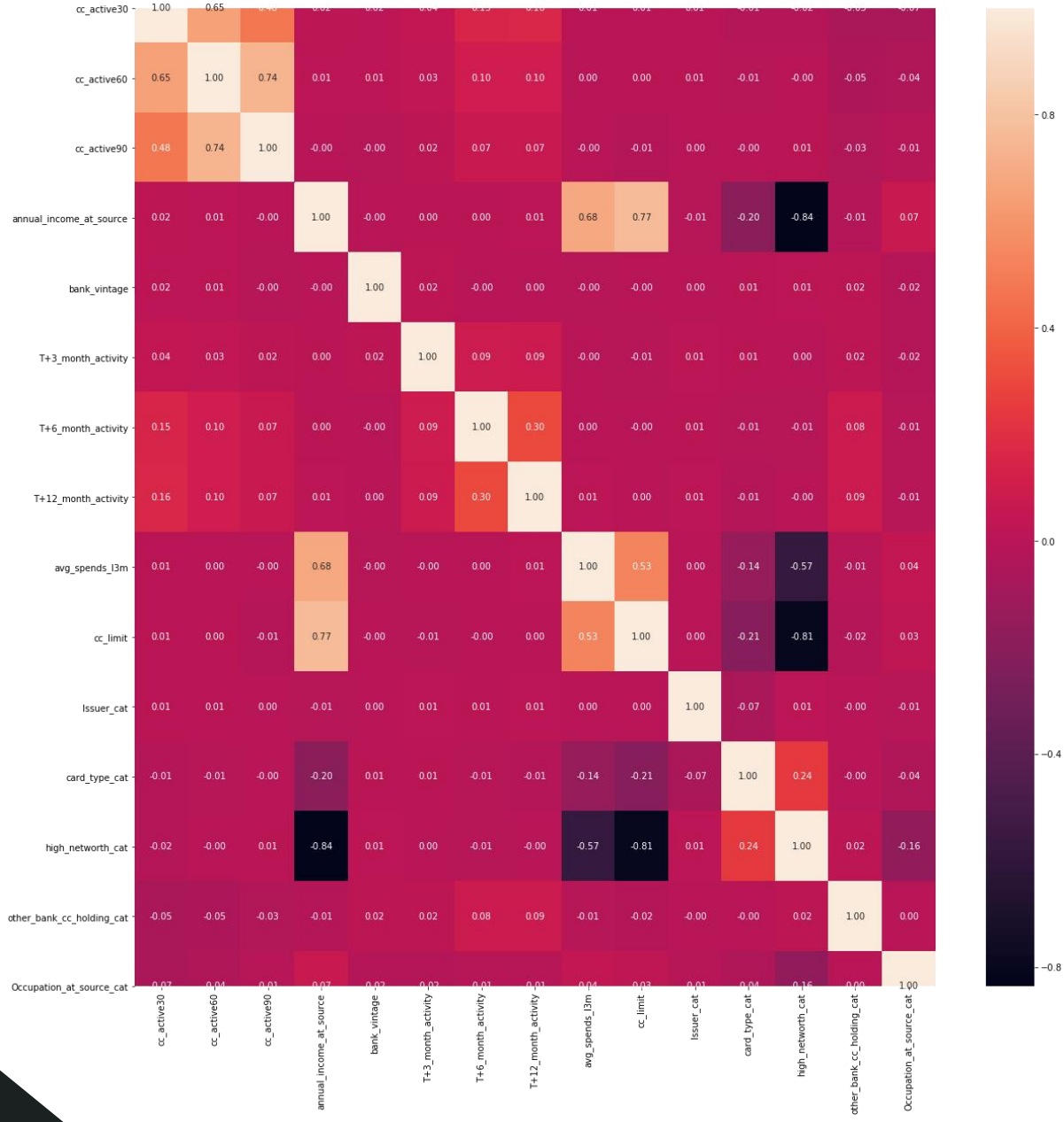
GODIGT Bank has a set of customers who were given credit cards based on risk policy and customer category class but due to huge competition in the credit card market, the bank is observing high attrition in credit card spending. The bank makes money only if customers spend more on credit cards. Given 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.)

- Customers in the Data base are **8448**
- Feature defining their activities: **28**

- To find the 5 most significant variables, have dropped multiple features which show no affect in Customer attrition.

Dropped features	Reason
userid	Each User id is unique with no relevance to attrition.
card No	Card Number with no
card source date	No relevance in attrition cause
active 30_60_90	Customer activity will not help to stop attrition
hotlist flag	No relevance in attrition cause
Widget_products	No relevance
T+1 & T+2	Considering the avg spent in 3 months
Transactor_revolver	No relevance

The below graph infers that:


cc_active30
cc_active60
cc_active90
annual_income_at_source
bank_vintage
T+3_month_activity
T+6_month_activity
T+12_month_activity
avg_spends_l3m
cc_limit
Issuer_cat
card_type_cat
high_networth_cat
other_bank_cc_holding_cat
Occupation_at_source_cat
0.8
0.4
0.0
−0.4
−0.8

- Features with negative correlation made more significance than features with positive correlation.
- Set of variables which negative correlation and high relevance to attrition are:
    a.high net worth- CC limit.
    b.high net worth- annual income.
    c.high networth- average spent in last 3 months.
    d.cc limit and cc type.

With above features we can draw following insights:

1. High Net worth: With high net worth the customer will expect a premium card with multiple benefits and good cc limit.
2. CC Type: The customers are not satisfied with the type of credit card offered to them.
3. CC Limit: The credit card limit should satisfy the customers' expectations else he will look for other competitors.
4. Annual Income: More the annual income more expectations on cc limit.
5. Average spent in last 3 months: Checking history will help us identify the activity of customers to draw conclusion on customer satisfaction.

To conclude this, we can say that the 5 most significant features are:

1. **high_networth**
2. **card_type**
3. **cc_limit**
4. **annual_income_at_source**
5. **avg_spends_l3m**

THANK YOU