SMDM PROJECT REPORT

DSBA



Table of Contents

Problem 1

* 1. Describe the Size and Nature of the dataset in hand
  2. Preliminary Analysis of the variables, Quality check on data
  3. Univariate Analysis
  4. Bivariate and Multi-variate Analysis
  5. E1, E2, E3 Statement Questions
  6. F1, F2 statement Questions
  7. Working partner vs Purchase history
  8. Analysis on Gender and Marital Status

Problem 2

* 1. Describe the Size and Nature of the dataset in hand
  2. Preliminary Analysis of the variables, Quality check on data
  3. Feature Engineering
  4. Assumptions on the Dataset
  5. Visualization
  6. Top 5 important variables to increase GoDigit Bank Profits with Justification

Problem 1

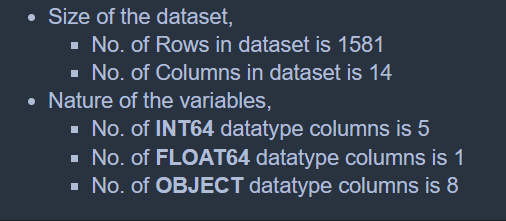
**Analysts are required to explore data and reflect on the insights. Clear writing skill is an integral part of a good report. Note that the explanations must be such that readers with minimum knowledge of analytics is able to grasp the insight.**

**Austo Motor Company is a leading car manufacturer 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.**

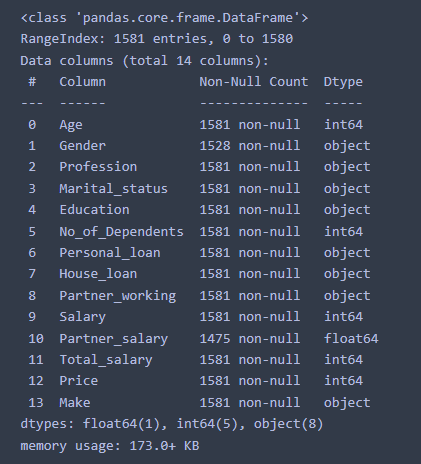
**As an analyst have been tasked with performing a thorough analysis of the data and coming up with insights to improve the marketing campaign.**

**1.1.**

**Describe the Size and Nature of the dataset in hand.**



**Brief understanding of the dataset features and values.**



**1.2. Preliminary Analysis of the variables, Quality check on data**.

**Check Missing Values**



* **Gender** Column has 53 missing values which is **3.3%** of the dataset.
* **Partner salary** column has 106 missing values which is **6.7%** of the dataset.

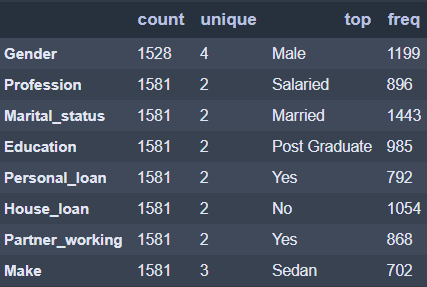
**Describe data for numerical and categorical features**

Numerical Features stats analysis -



* Total Salary feature has maximum value to very high, there can be potential outliers.
* We can find almost bell curve or normal distribution of the data in features like Age, Salary, Total Salary because the Mean and Median are almost equal here.

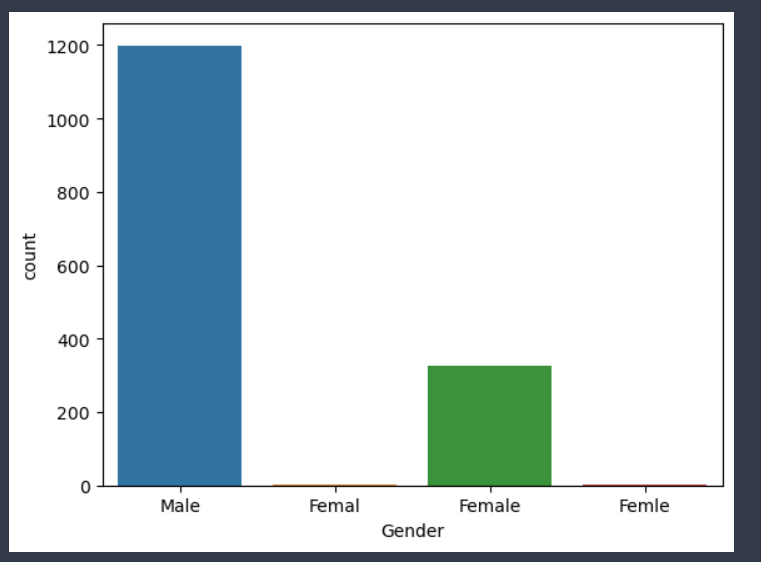
Categorical Features stats analysis -



* Gender has 4 unique values, which is daunting because there could be Male and Female. Bad data might be ingested.
* Around 79% data in Gender are Males, 91% of customers are Married.

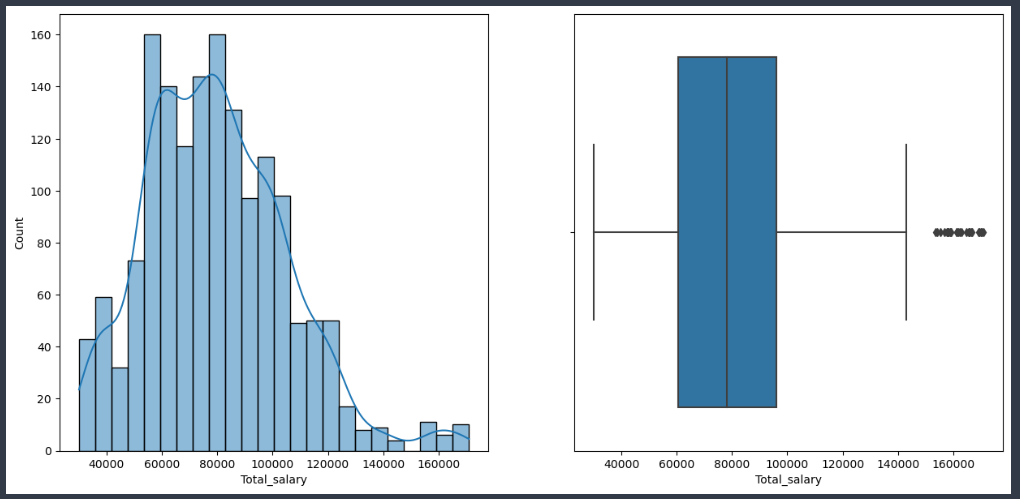
**Preliminary Checks on Variables**

Gender -



* Discrepancy in Gender column
  + Only Male and Female values are valid.
  + But from count-plot, we can see incorrect values like **Femal** and **Femle.**

Total Salary -



**Before Treating Outliers**

* We can confirm that there are outliers in Total Salary feature, we need to treat the outliers using IQR.

**Impute missing values, Treating outliers or bad data**

Total Salary -

* There is total 106 missing values.
* 90 out of 106 rows has Salary equals to Total Salary. Which means Partner salary is zero.
* Remaining 16 rows has variable values. So, imputing as,

**Partner salary = Total salary - Salary**

Basic Sanity check around Partner salary,

* There are in total 138 single users, which means they do not have partner i.e., they will have Partner Working feature as NO in total 138 rows.
* Similarly, all 138 single users should have partner salary as zero (0). There are 122 rows with 0 value but 16 rows having null values, which should be marked as zero only.

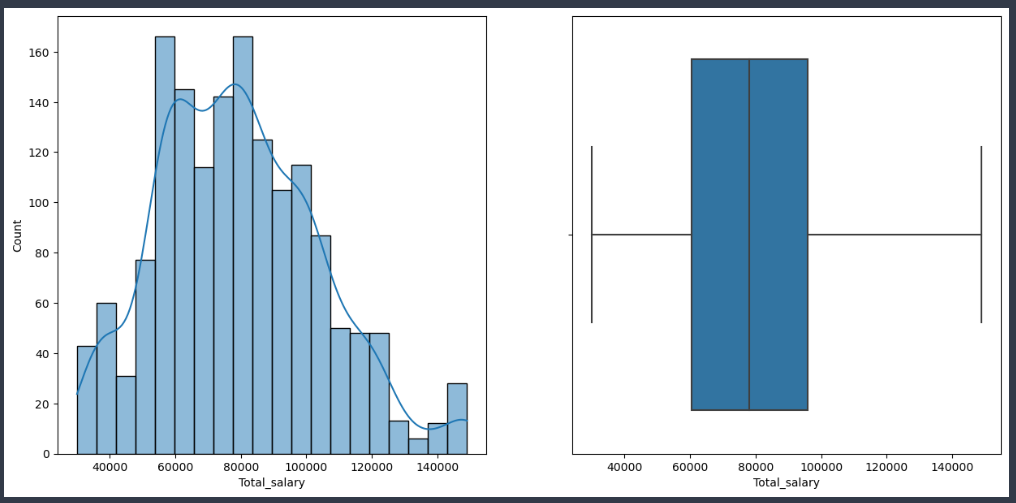
Treat Outliers using Inter-Quartile Range (IQR) in Total Salary,

Formula- **IQR = Q3 - Q1,**

* IQR is Inter-Quartile Range
* Q3 is 75% of the data
* Q1 is 25% of the data

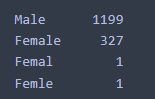
Formula – Upper Limit = Q3 + 1.5\*(IQR)

* Upper Limit is Upper value of Whiskers of box-plot.
* Values beyond Upper limit are called as Outliers.
* From analysis, we have received Upper limit value for Whiskers of box-plot is 149000.
* If Total Salary > 149000 then impute those outliers with upper limit value i.e., 149000.
* There are 27 outliers from the dataset which will be replaced with upper limit value to follow distribution of the dataset.



**After Treating Outliers**

Gender –

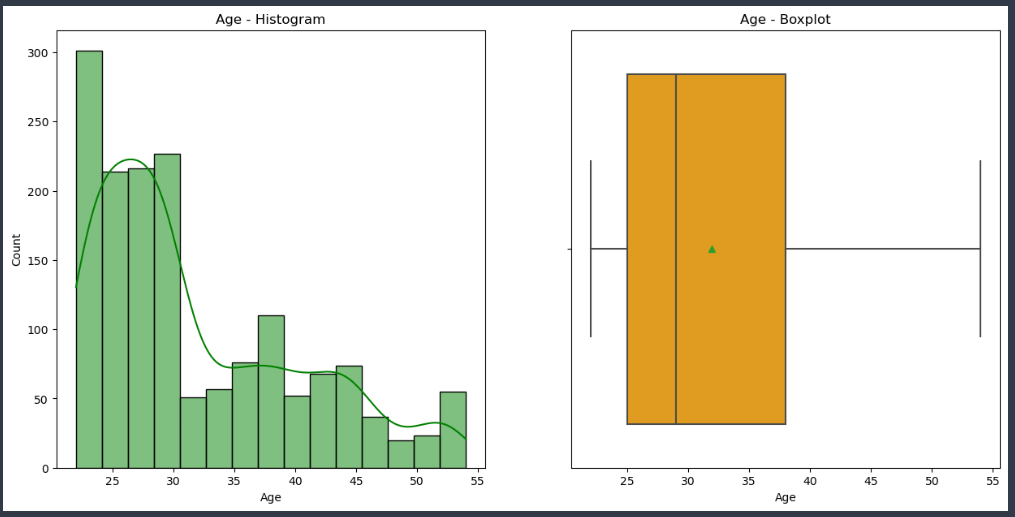


* There are 2 rows which containing bad entries for **Female**. So, we can replace those entries with **Female**.
* Treating Bad data by replacing values,
  + Femal -> Female
  + Femle -> Female



**1.3. Univariate Analysis**

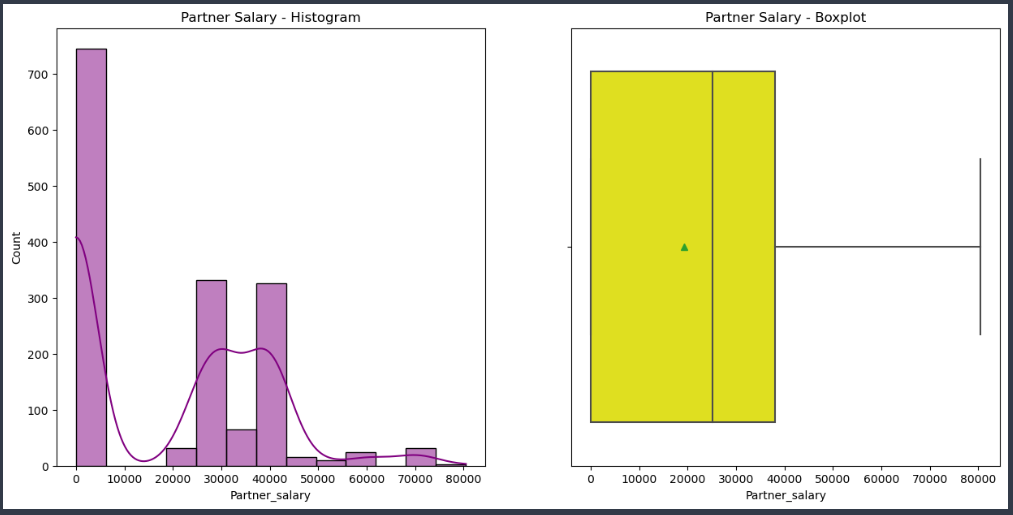
**Age**



**Age – Histogram & Box-plot**

* 75% of the customers under the age < 40 are interested in buying cars. Good insight to pitch special offers for customers in this cluster.
* From Histogram, lots of customers from age between 22 to 30 are shown, assuming young customers cracking first job and planning for a car immediately is a dream/goal. We can revamp our old models and provide better offers/discounts for recent old model cars to target younger generation as they might do not have high salaries.
* Data Interpretation
  + Data is Right/Positively skewed.
  + No outliers. Mean is little high than Median due to right skewed.

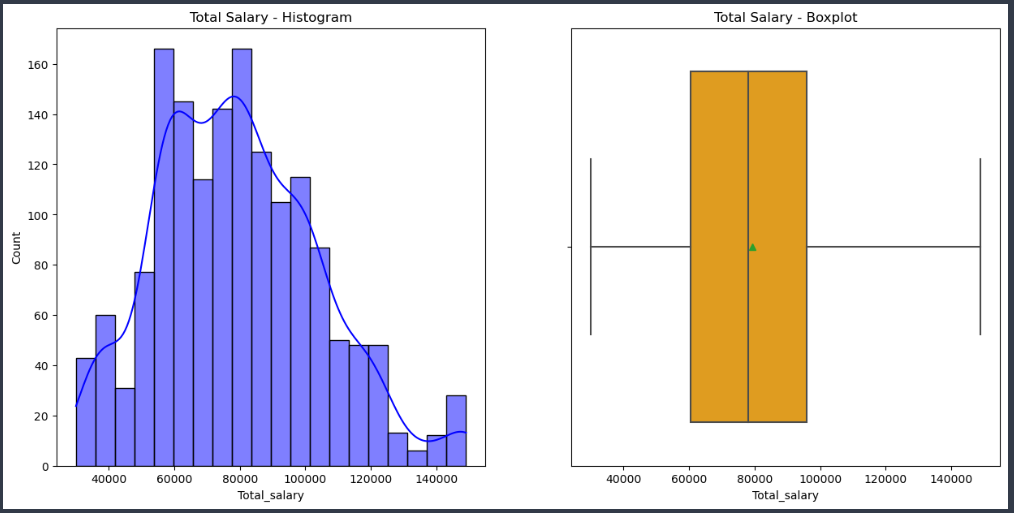
**Partner Salary**



**Partner Salary – Histogram & Box-plot**

* There are around 45% customers do not have a Working Partner to provide financial stability.
* Similarly, there are around 8.7% customers are single, who do not have a Partner.
* From the above 2 points, it is understood that there are customers with No Partner or Has Partner but not working, which led to lots of zero (0) salary for partner in the Histogram.
* On contrast, 55% customers have a Working Partner so it's better to target customers who has financial support from partner which can lead to purchase expensive or High-end cars in the market.
* Data Interpretation,
  + Data is completely scrambled and no proper distribution is observed.

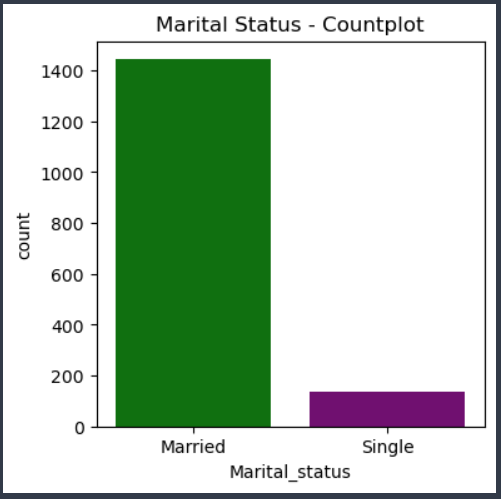
**Total Salary**



**Total Salary – Histogram & Box-plot**

* As per data, the mean(79k) and median(78k) total salary is around 80k(say). But starting 25% of the customers has total salary around 60k. We can target customers whose salary < 60k or falls into 25% of our data with better offers/discounts during festivals to increase sales.
* Data Interpretation,
  + Total Salary Data is very close to Symmetrical data i.e. Bell curve in shape.
  + Which can be proven from Mean around 79k and Median around 78k.

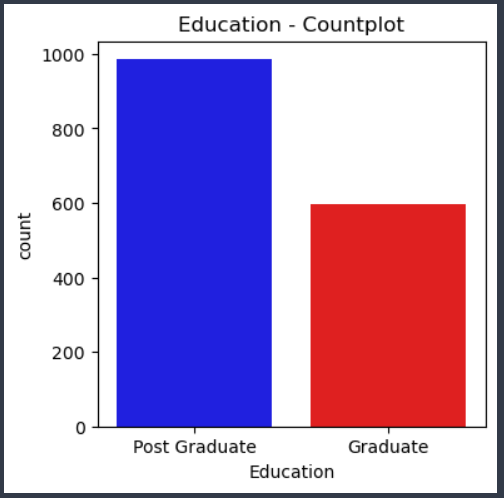
**Marital Status**



**Frequency of Marital Status of Customers**

* Around 91% of the customers are Married from our dataset, which implies Married customers are tend to have Small/Big Family and having a car would be an ideal goal for transportation.
* Targeting Married customers will have good impact on Sales because Customer's partner might also be working and provides financial stability which adds more reason to buy a car.

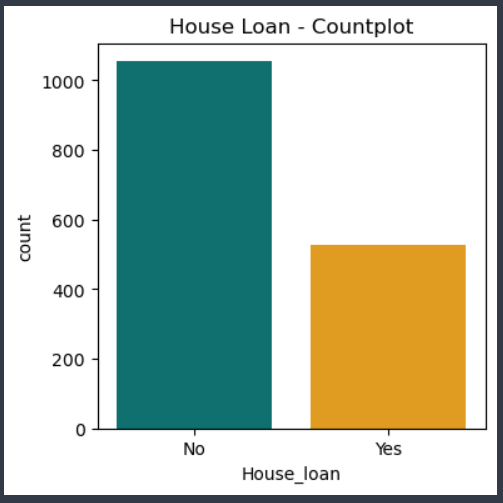
**Education**



**Frequency of type of Education**

* Post Graduates are high in number i.e., around 62% from our dataset. Which implies, Customers pursued Post Graduation tend to have better or high paying Jobs compared to Graduates and hence they tend to invest some amount in buying a car.

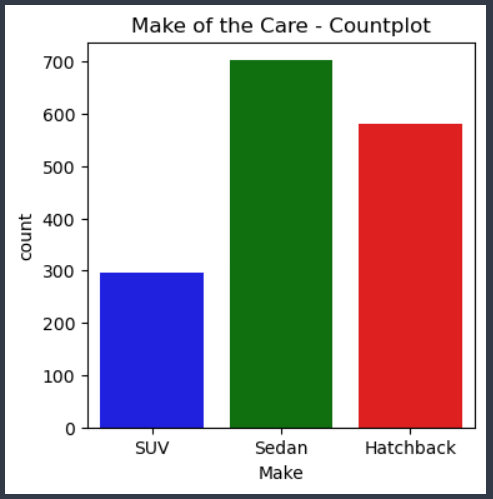
**House Loan**



**Frequency of House Loan opted or Not.**

* There is a chance that customer not having House Loan can be intended to buy a car compared to those having House Loan.
* From the data, we have 67% customers do not have House Loan and 33% has House Loan.
* It can be interpreted that, Customers who did not opt for House Loan might have saved enough money to get a new car.

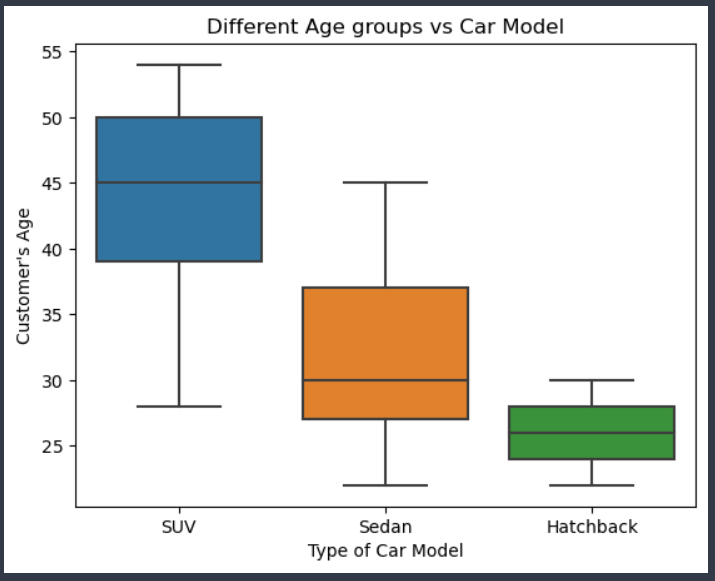
**Make (Model of the Car)**



**Frequency of Make (Type of Car Model)**

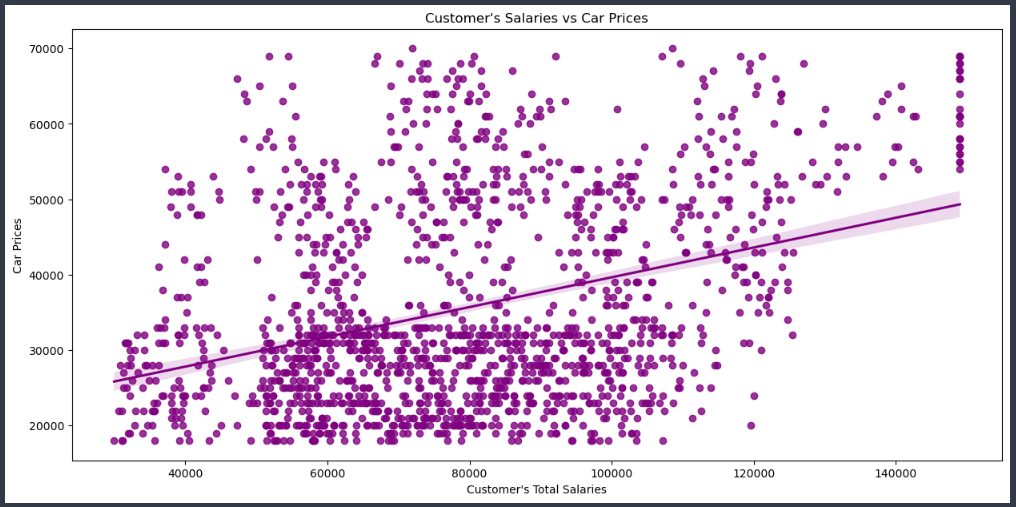
* Customers tend to buy Sedan and Hatchback models more when compared to SUV's.

1.4. Bivariate and Multi-variate Analysis



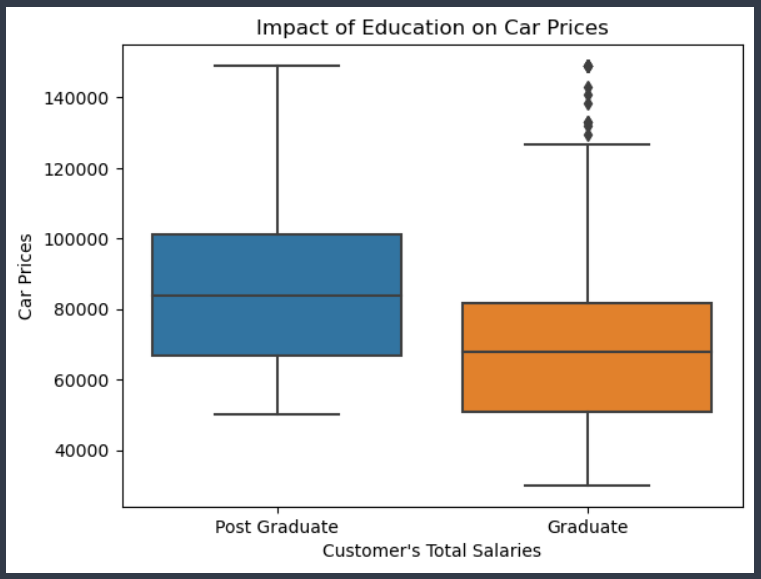
**Age groups vs Car Make**

* We can find out a pattern that,
  + Customers aged around 38+ are interested in buying SUVs.
  + Customers aged between 27 to 40 are interested in buying Sedan.
  + Customers aged less than 30 are interested in buying Hatchback model.
* Age of the customer will tend to describe more about the purchasing habits about the car model.



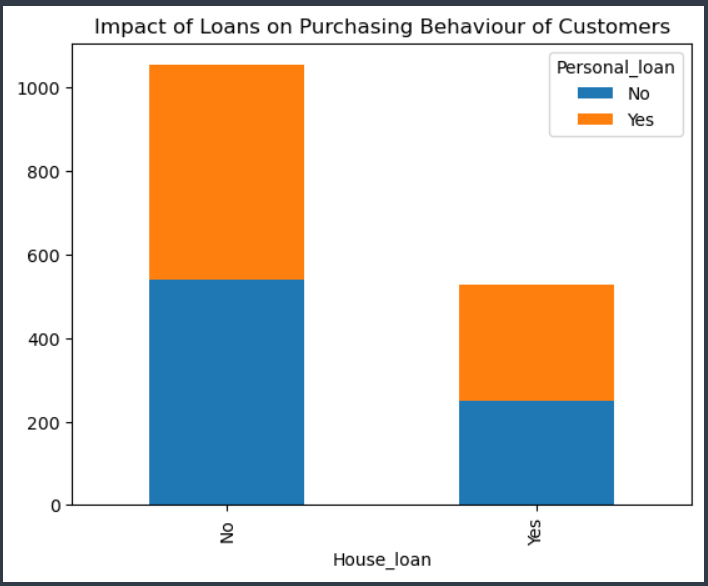
**Salaries vs Car Prices**

* There is a Linear Correlation between the Total Salaries of the Customers and Car Prices.
* Which can be True that if a customer can invest more money to buy a car then they are expecting to buy an expensive car.
* Alternative is True as well, because customer's having salaries less will not go beyond their budget and will opt for a cheaper car.
* We can provide better discounts or sales during festival time for Customer's having less salary will try to opt for an expensive car with better sales. (Important Marketing Campaign point)
* Statistically, there is around 36% positive correlation between Total Salary and Price.



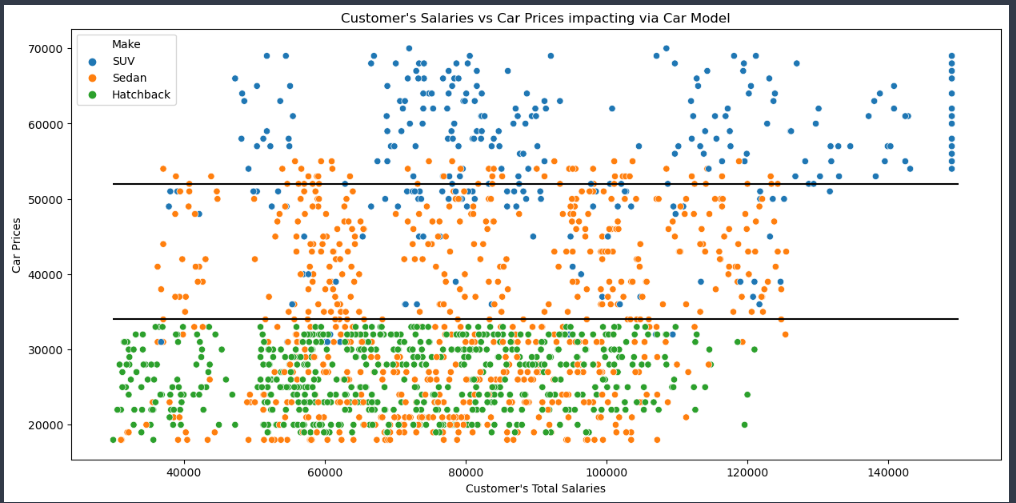
**Salaries vs Prices (via Education)**

* Post Graduates are high in number i.e., around 62% from our dataset.
* Which implies, Customers pursued Post Graduation tend to have better or high paying Salaries compared to Graduates which can be seen from the graph.
* Hence, we can target customers with higher education qualification as we have already seen high salaries tend to purchasing expensive cars.



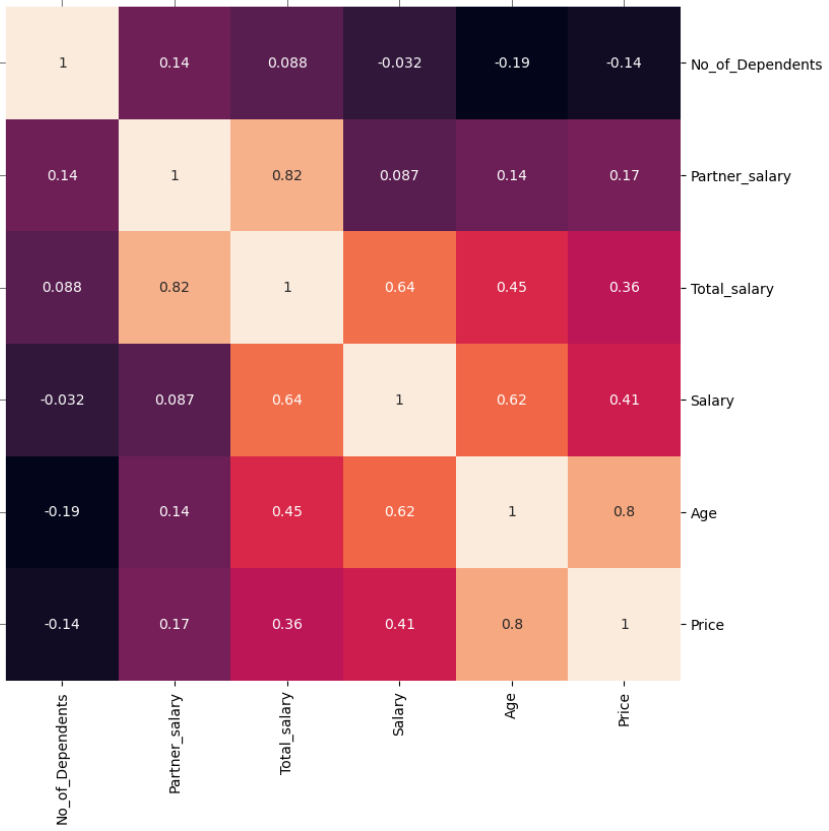
**House Loan vs Personal Loan**

* We can interpret that Customers not having Personal Loan and House Loan are in more in number from Graph. Which implies they are tend to opt for a Car having no loans will plan them to invest on Car.



**Salaries vs Prices (via Car Make)**

* Earlier, we have seen there's a Positive correlation between the Total Salaries of the Customers and Car Prices.
* But if we deep dive by adding Car Model, we can interpret that:
* SUVs are super expensive among the all 3 models. And is affordable by more high salaried customers.
* Hatchback model is quite cheaper and it is attracted by all customers having very low salaries.
* But Sedan model is dispersed a lot in the Graph, assuming customers with low salaries might go for Sedan model after Hatchback if they are intended put to more money. But not as high as SUVs range.



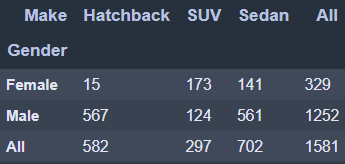
* From heatmap,
  + Customer's Age is 80% Strongly as well as Positively correlated to Price of the Car. Similarly, Age is 45% positively correlated with Total Salary of the customer.
  + With Increase in Age of the Customer, then Salary will also increase. Which can be interpreted as Age is 62% strong and positively correlated to Salary.
  + Total Salary and Salary usually decides the Purchasing Power of the Car. Since, Price is positively correlated at 41% with Salary and 36% with Total Salary respectively.
  + We do not see much negative correlation among any variables. Just few variables showing weak correlation.

1.5.

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.

**Question E1** - Steve Roger says 'Men prefer SUV by a large margin, compared to the woman'.

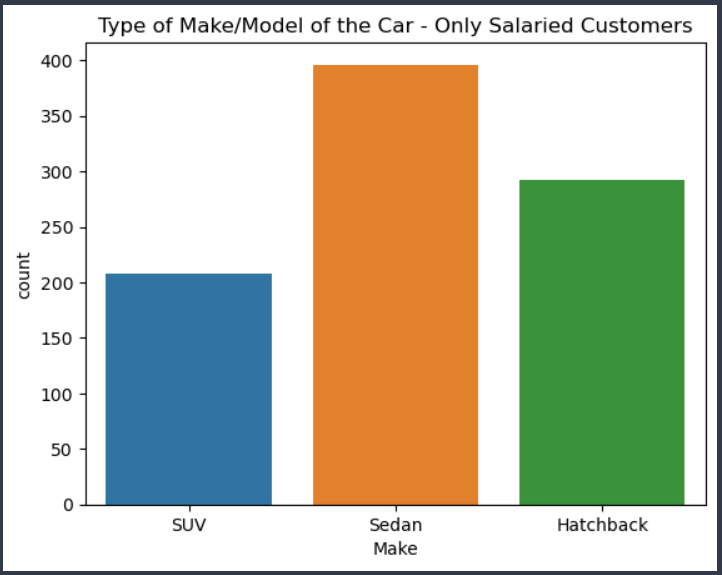
**Solution E1** -



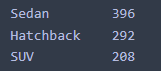
* The above statement is **False**, where Females are using more SUVs compared to Males.
* No. of Males have SUVs is 124, whereas Females have 173 SUVs.

**Question E2** - Ned Stark believes that a salaried person is more likely to buy a Sedan.

**Solution E2** –



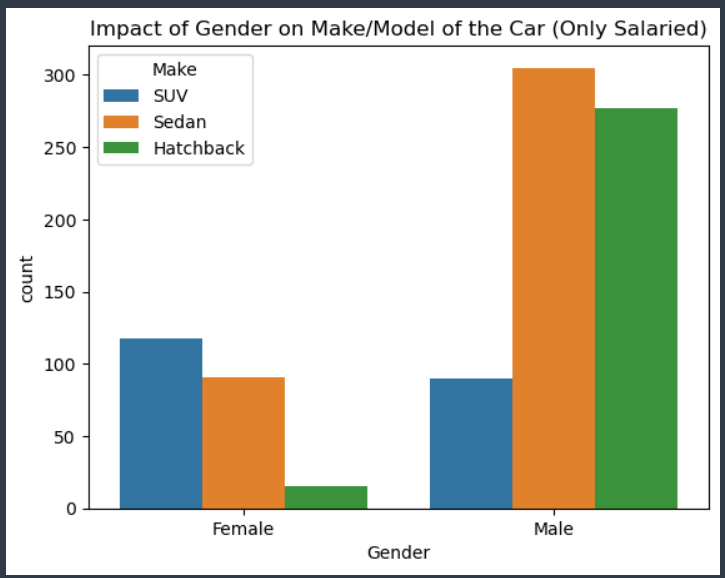
**Type of Car Make (only Salaried)**



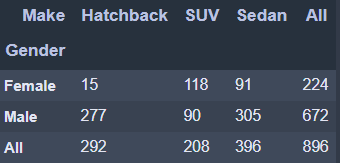
* The above statement is **True**, From the graph we can interpret that Sedan Cars are being sold more. (Only for Salaried Customers)
* There are 396 Sedan Cars sold when compared to 292 Hatchback and 208 SUV models. (Only for Salaried Customers)

**Question 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.

**Answer E3** –



**Gender Impact on Car Make/Model**



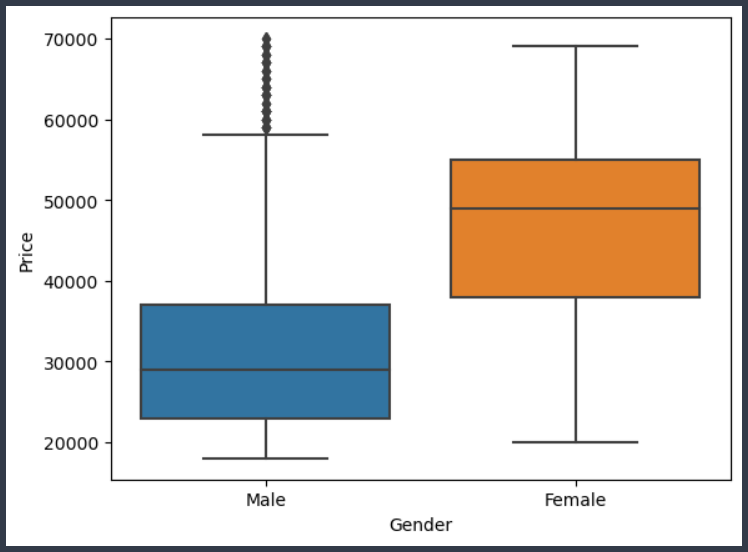
* The above statement is **False** because a Salaried Male is not easier target for a SUV sale.
* From the graph, Salaried Males are easily targeted for Sedan and Hatchback models but least preferred by SUVs.
* Salaried Males bought 305 Sedan, 277 Hatchback, and 90 SUVs models overall.

1.6.

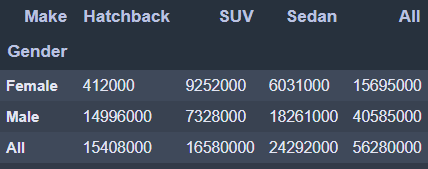
From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions.

1. Gender
2. Personal Loan

Gender –

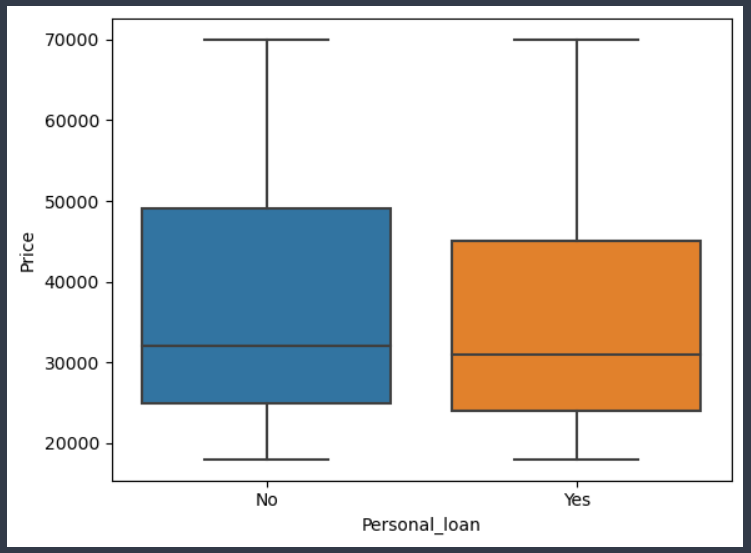


**Gender vs Prices**

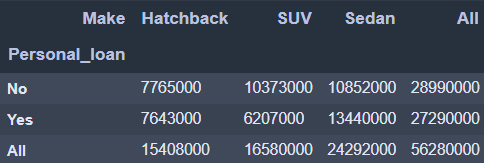


* The Total Amount spent on purchasing automobiles by Gender is,
  + Males - 4,05,85,000 $
  + Females - 1,56,95,000 $
* But we have very biased data w.r.t Gender i.e.,
  + Males are 80% (1252 Males)
  + Females are 20% (329 Females)
* So, let us calculate total amount spent per taking individual customer in Gender,
  + Total Amount spent on Automobile purchasing for Males / No. of Males.
    - Amount spent by Each male on Automobile = (40585000/1252) = **32416.14 $**
* Total Amount spent on Automobile purchasing for Females / No. of Females.
  + Amount spent by Each male on Automobile = (15695000/329) = **47705.17 $**
* From the above calculations, we can suggest that targeting **Female customers will have better sales compared to Males**. There is a significant difference of **15,290 $** value.

Personal Loan –



**Personal Loan vs Prices**

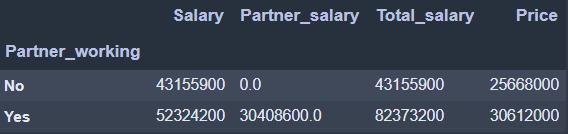


* The Total Amount spent on purchasing automobiles based on Personal Loan is,
  + Yes, Customers has Personal Loan - 2,72,90,000 $
  + No, Customers don't have Personal Loan - 2,89,90,000 $
* We don't have biased data w.r.t Personal Loan i.e.,
  + Customers having Personal Loan are ~50% (792)
  + Customers not having Personal Loan are ~50% (789)
* So, let's calculate total amount spent per taking individual customer on the basis of Personal Loan,
  + Total Amount spent on Automobile purchasing for customers opted Loan / No. of customers opted Personal Loan.
    - Amount spent by Each customer opted Personal Loan = (27290000/792) = **34457.08 $**
* Total Amount spent on Automobile purchasing for customers not opted Loan / No. of customers not opted Personal Loan.
  + Amount spent by Each customer not opted Personal Loan = (28990000/789) = **36742.72 $**
* From the above calculations, we can suggest that targeting customers having Personal loan or not having Personal Loan is not going to impact a lot. But Yes, Total Amount spent by **customers not having Personal Loan has spent little high** with difference of **2,285.7 $** value.

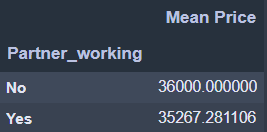
1.7.

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

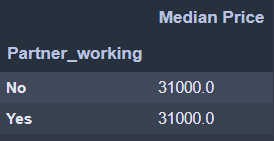
**Solution –**



* The Total Amount spent on purchasing automobiles based on Working Partner is,
  + Yes, Customers has Working Partner - 3,06,12,000 $
  + No, Customers don't have Working Partner - 2,56,68,000 $
* We don't have biased data w.r.t Working Partner column i.e.,
  + Customers having Working Partner are ~55% (868)
  + Customers not having Working Partner are ~45% (713)
* So, let us calculate total amount spent per taking individual customer based on Working Partner,
  + Total Amount spent on Automobile purchasing for customers having Working Partner / No. of customers having Working Partner.
    - Amount spent by Each customer having Working Partner = (30612000/868) = **35267.28 $**
* Total Amount spent on Automobile purchasing for customers not having Working Partner / No. of customers not having Working Partner.
  + Amount spent by Each customer not having Working Partner = (25668000/713) = **36000 $**
* From the above calculations, we can suggest that targeting customers having Working Partner or not having Working Partner is not going to impact a lot. But Yes, **Total Amount spent by customers having Working Partner has spent little high with difference of 732.7 $** value.



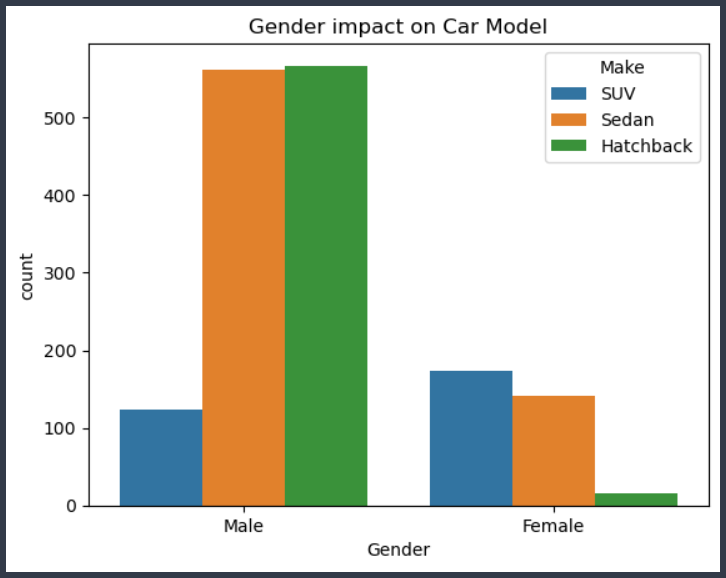
* From Mean, Having Working Partner or not having is not going to impact much with sales. But there's slightly less chances for buying higher priced car when having working partner.
  + Mean Price,
    - Having Working Partner - 35267.29 $
    - Not Having Working Partner - 36000 $
    - **Difference in Mean Price is 732.71 $**



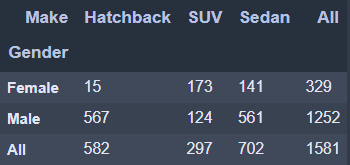
* From Median, Having Working Partner or not having is not going to impact, Because the median value is same in both cases.
  + Median Price,
    - Having Working Partner - 31000 $
    - Not Having Working Partner - 31000 $
    - **No difference in Median Price value.**

1.8.

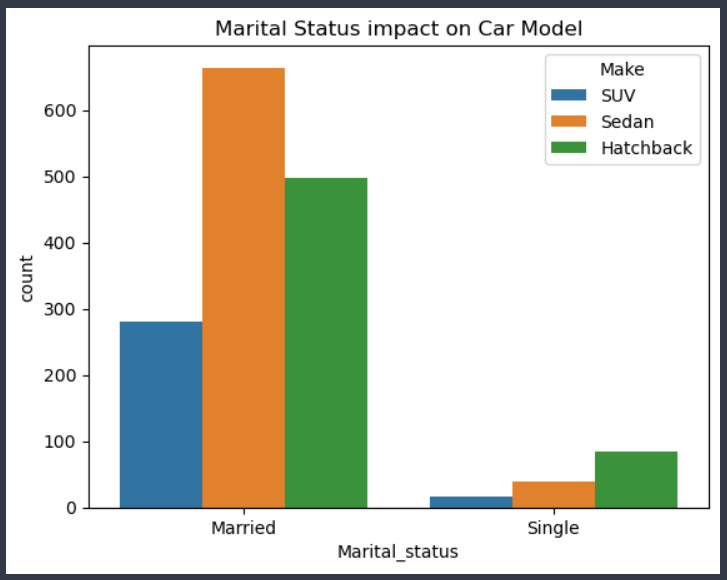
The main objective of this analysis is to devise an improved marketing strategy to send targeted information to different groups of potential buyers present in the data. For the current analysis use Gender and Marital status - fields to arrive at groups with similar purchase history.



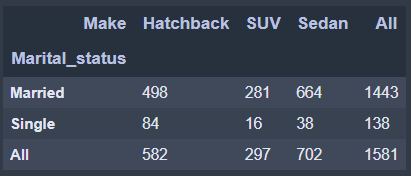
**Frequency of Gender via type of Car Make**



* From the above graph, it is suggested that targeting Males with Hatchback and Sedan will have a better impact on sales.
* A campaign can be conducted for these 2 models to attract lots of Males and provide better discounts.
* SUVs are least preferred by Male, finding out reasons and change marketing strategies for SUVs at different locations with set of targeting customers might increase sales.
* As we do have very biased with Gender, Female customers are very less in percentage as compared to Males. We want to implement a campaign to attract to Female and provide better offers for any woman purchasing automobile.
* Women are interested in purchasing SUVs which is completely contrast to Male. So, if we can gain more Female customers then eventually sales for SUVs would go up.



**Frequency of Marital Status via Car Make**



* We have very biased data for Marital Status, where married customers are ~91% and Singles are just ~9%.
* From Married customers, we can analyse that sedan is top most car bought by many married customers then followed by Hatchback and then finally SUVs.
* We can do a campaign for increasing the sales for Hatchback to a little extent and a find strong reasons for SUVs campaign especially for married couples.
* SUVs are being very expensive so no one is showing more interest to buy, Maybe Automobile organisation can solute to provide better offers or discounts for celebrating anniversaries for married couples to increase sales.

**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.

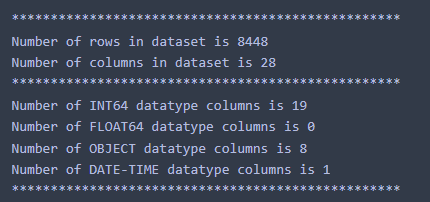
GODIGT Bank also has a set of customers who were given credit cards based on risk policy and customer category class but due to huge competition in the credit card market, the bank is observing high attrition in credit card spending. The bank makes money only if customers spend more on credit cards.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

**Objective**: **Framing an Analytics Problem**

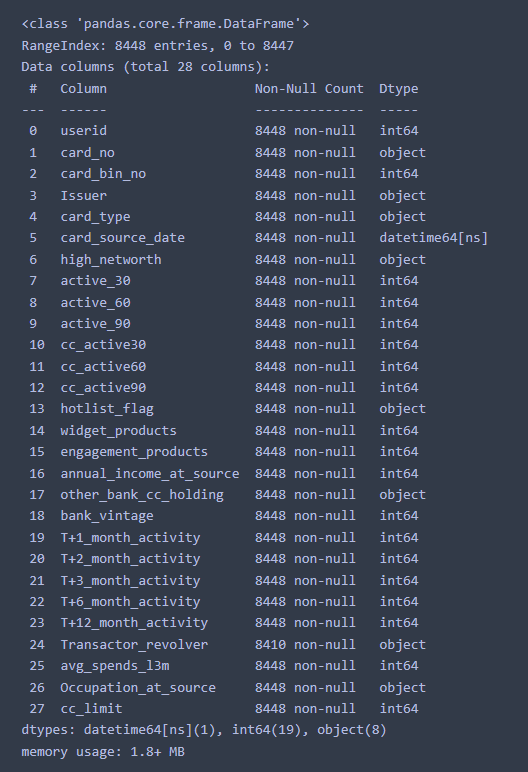
* Analyse the dataset and list down the top 5 important variables, along with the business justifications.

**2.1.**

**Describe the Size and Nature of the dataset in hand.**



**Brief understanding of the dataset features and values.**



**2.2.** Preliminary Analysis of the variables, Quality check on data.

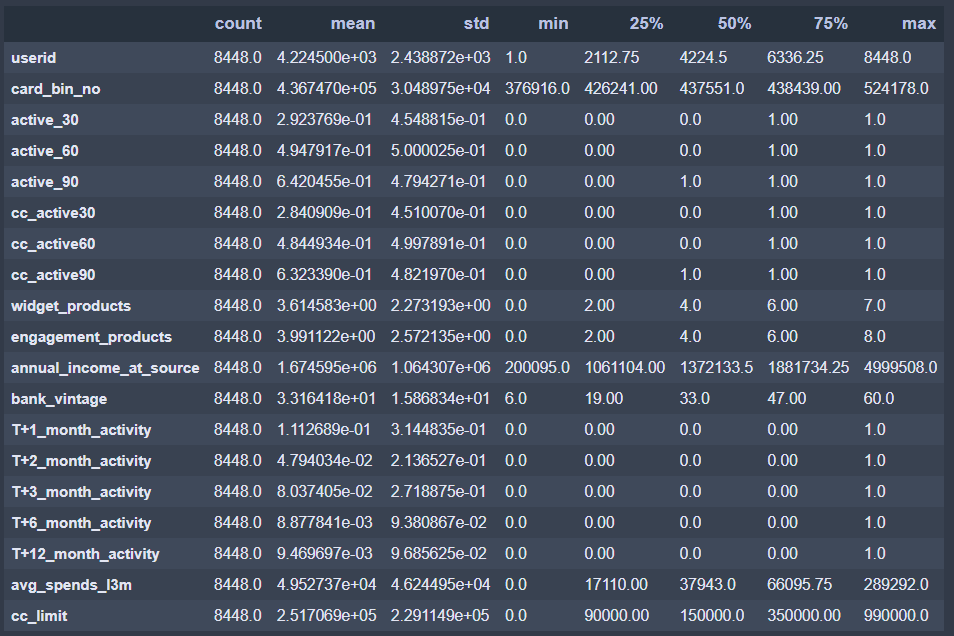
**Checking Null values**



* There are 38 null values in Transactor Revolver column.

**Describe data for numerical and categorical features**

Numerical Features stats analysis -



* Annual Income at source looks like having an outlier in the data because 75% data obeys 18.8Lakhs but maximum value stays at ~50Lakhs, which can be a very expensive customer with very High salary.
* Average spends in last 3 months variable also records outliers where 75% data shows spend ~66k but maximum spending value is ~2.9Lakhs.
* Credit Card Limit has few insights as well,
  + There are customers with NO credit card limit i.e., 0 cc\_limit
  + Highest credit card limit issued to customer/s is 9.9Lakhs

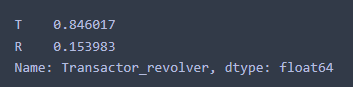
Categorical Features stats analysis -



* Visa Card seems to be issued to lots of customers from GoDigit Bank.
* Majority of the customers did not hotlist their card and most of them follow Transactor approach by clearing all dues in single payment.

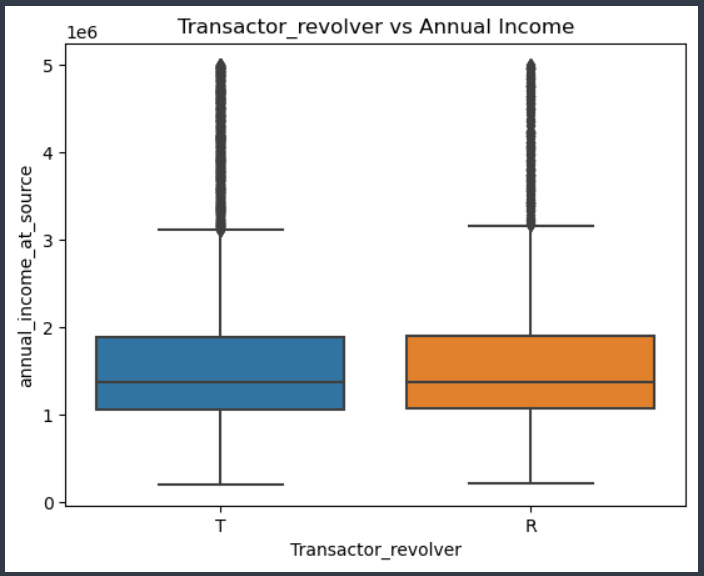
**Impute missing values, Treating outliers or bad data**

* Transactor\_revolver column has 38 null values from the dataset which yields ~0.45% of overall data.
* Either we can drop these 38 rows given very low percent of weightage or impute the values by doing analysis. (Let’s do analysis and impute)



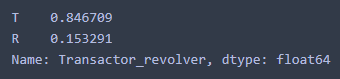
**Before Imputing null values**

* From the Transactor\_revolver column, it is observed that ~85% customers come under Transactor(T) i.e., most of the customers are willing to clear their balances in Full in every month and just ~15% come under Revolver(R) where customers are revolving their balances to the next months.

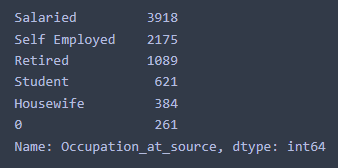


**Transactor\_revolver vs Annual Income**

* As clearing balances will depend on the Annual Income of the customers. But from above graph, we don't find any significant difference in Annual Incomes of Customers with Transactor or Revolver categories.
* From this analysis, we can strongly impute this column with Transactor(T) as most of the customers fall under this category.

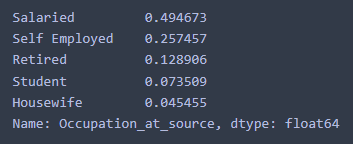


**After Imputing null values**



**Bad data in Occupation column**

* There are 261 bad values (0 value) in Occupation column. We need to treat it before doing any analysis.
* A good indication that almost half of the data i.e., ~46% Customers are Salaried. We can replace all the ‘0’ values with ‘Salaried’ to address this bad data.
* Occupation column can be a good predictor for improving bank profits by targeting right set of customers.



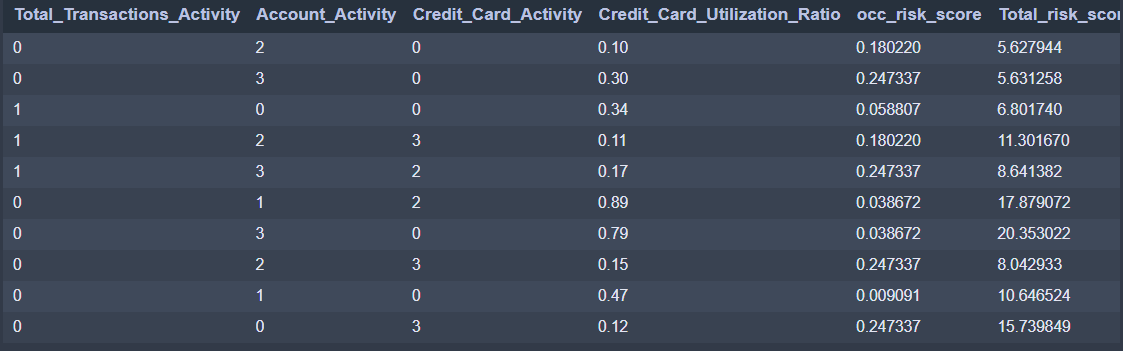
**Data after treating bad values.**

* After replacing, we can see that Salaried weightage increased from ~46% to ~49%.

**2.3. Feature Engineering**

* **Aggregate columns**,
  + Total Transactions Activity (T+1, T+2, T+3, T+6, T+12)
  + Account activity (active\_30, active\_60, active\_90)
  + Credit Card activity (cc\_active30, cc\_active60, cc\_active90)
* **Credit card utilization ratio** - average spends over last 3 months by credit card limit will give us the ratio about utilization of the credit card. (High ratio is best)
* **Risk values based on Occupation** - (Below risk scores are assigned as per General research from Banking history data, these values might vary in different situations or aspects. But for Generalisation, these are credit risk scores.
  + Salaried: 5 (Medium risk)
  + Self Employed: 7 (High risk)
  + Retired: 3 (Low risk)
  + Student: 8 (High risk)
  + Housewife: 2 (Low risk)
* **Total Risk score** = (Occupation risk value) + (Annual income/credit card limit) + (average spends/credit card limit)

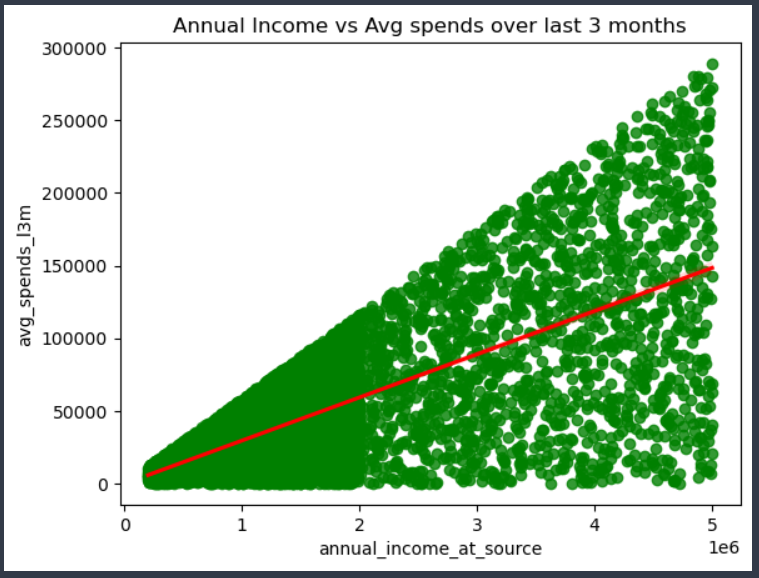
A glimpse of the newly generated data from feature engineering



**2.4. Assumptions on the Dataset**

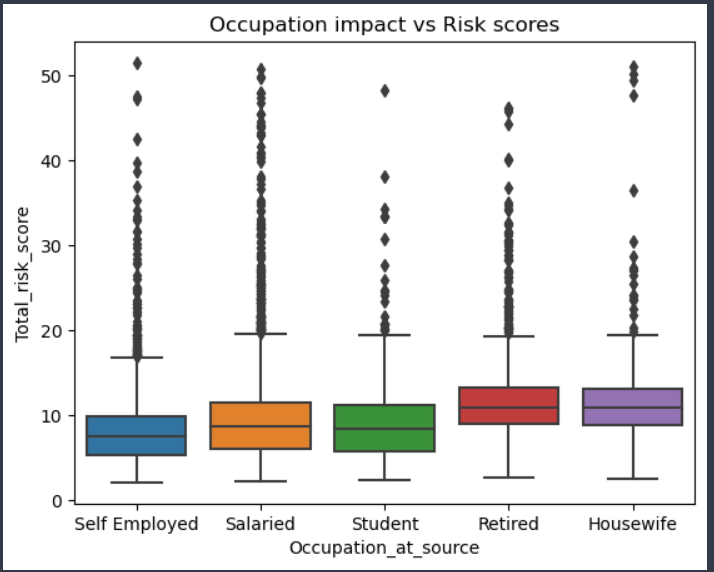
* **Objective** - To increase profits for GoDigit Bank but decrease attrition rate on Credit Cards.
* Customer needs to spend more via Credit Card,
  + It can only happen when Annual Income of Customer is bearable or decent in market.
  + Customers occupation decides the ability of them to the purchasing capacity of the credit card.
  + On contrast, Customers not using credit card since last 60 or 90 days, will have to decrease the credit limit as well as update policies. Similarly, customers who are using card regularly needs to be taken care and provide special benefits such as cash backs, offers, coupons, etc.
  + Average spending on last 3 months variable decides the purchasing history of the customers.
  + Credit limit will enforce customers to spend till the limit provided but not over exceeding. So, this will be a crucial parameter to estimate.
  + If customers are already having other banks credit card, then it influences customers to less utilize GoDigit Credit cards, Hence Bank should come up with an idea like waiving off charges on minimum amount spent or other offers to attract customers.
  + If credit card is hot-listed for some reason then it hinders user's usage on the card, which is a bad impact.
  + Average spends over last 3 months in relation with,
    - Annual Income of the customer
    - Credit card limit

**2.5. Visualization**



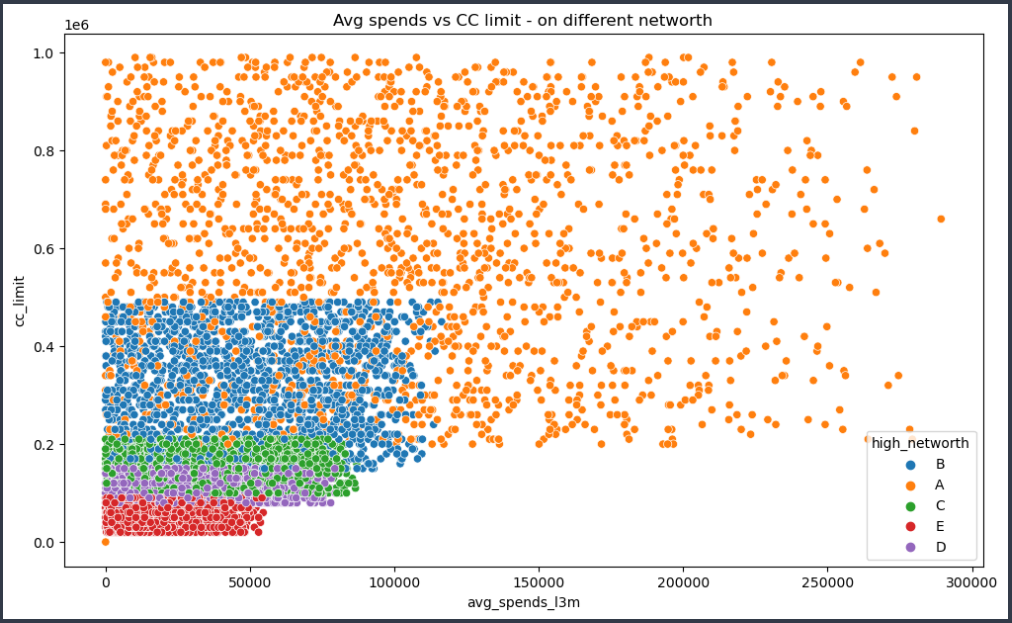
**Annual Income vs Average spends**

* There is a Strong and Positive correlation of 68% between Annual Income of the Customers and Average spends over last 3 months.
* Which is technically valid, if a customer's income is high then spending over the credit card will be eventually high as customer can clear dues without any efforts.



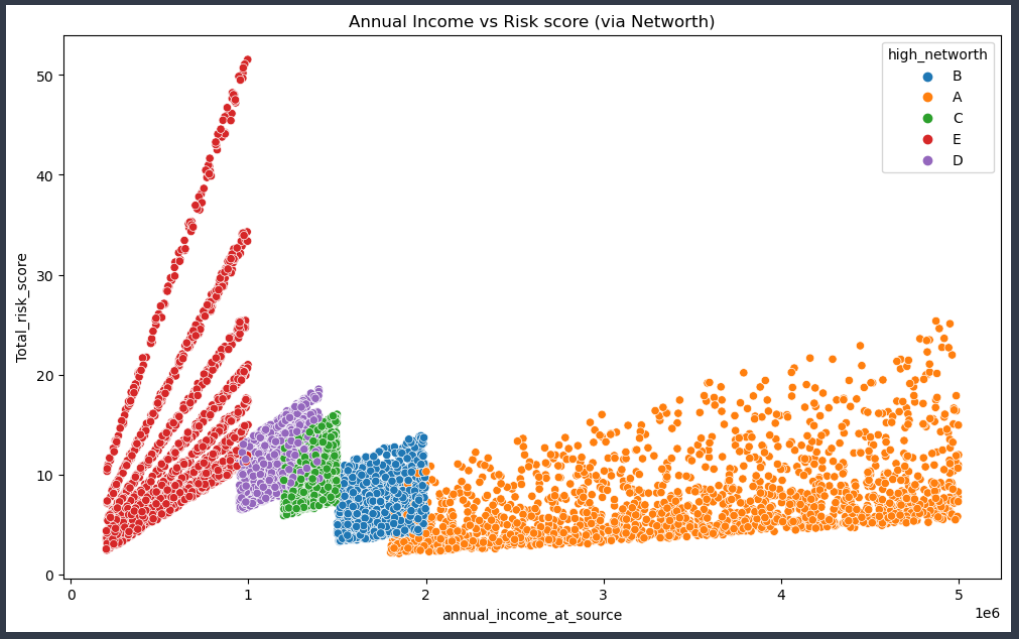
**Occupation vs Total risk score**

* Self Employed and salaried are having less risk score because they have source of income when compared to Housewives who might have or might not have proper income source to clear credit card dues.
* We can use Risk scores to target different occupation customers based on their demo-graphics data and risk score before issuing credit cards to customers.



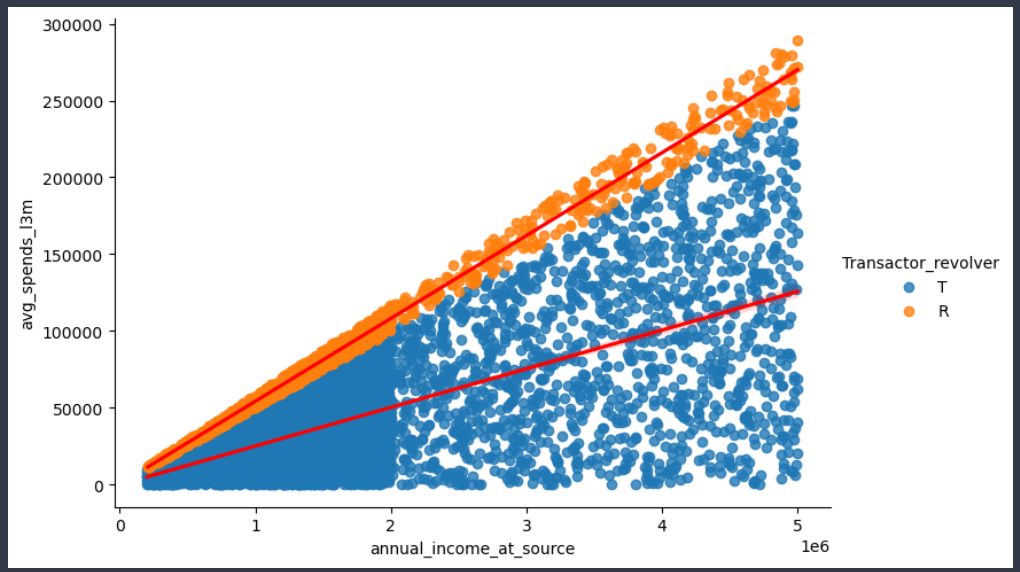
**Average Spends vs Credit Card limit**

* Average spends over last 3 months has very keen impact on the credit card limit.
* Higher the credit card limit will enforce customers to utilize more via credit card.
* By adding net-worth parameter to the graph, we can identify High net-worth customers are having high credit limit and spending more on credit card. Followed by other categories from data.
* From a business perspective, Banks can target customers that fall under this category and provide better benefits to the customers.



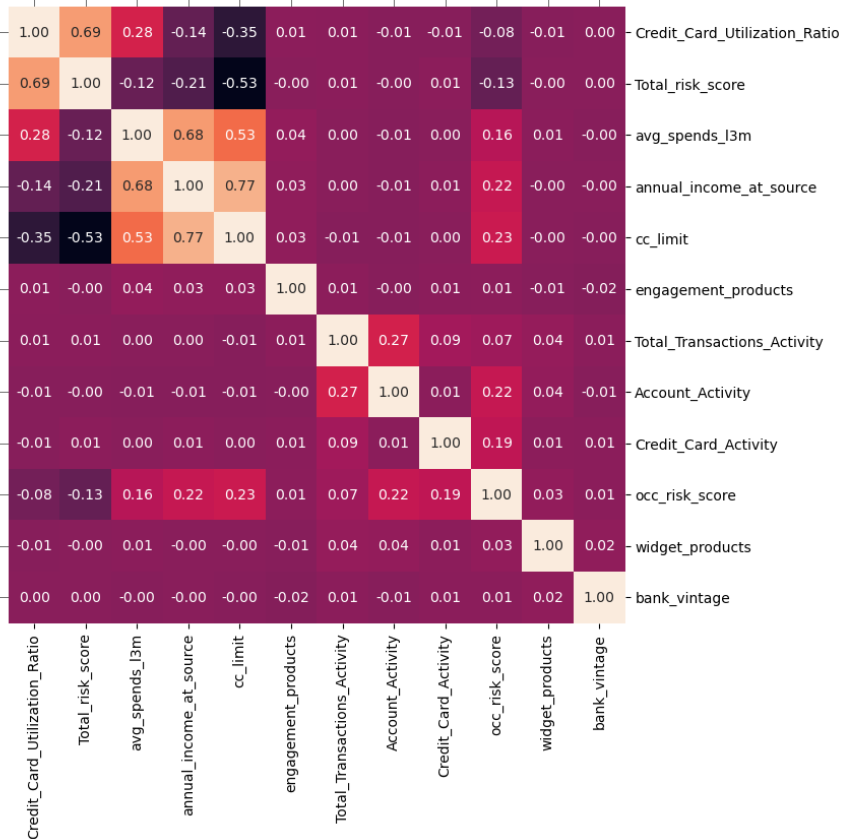
**Annual Income vs Total risk score**

* Annual Income is playing a vital role versus Total risk score and clustering customers via Net-worth parameter.
* Customers with low annual income cannot clear dues on time and hence are given very high-risk scores. On contrast customers with very low risk score are the one who's annual income is very high.



**Annual Income vs Average spends**

* From the above graph, which depicts relation between Annual Income and Average spends over last 3 months has very strong and positive correlation.
* By adding, Transactor\_revolver as 3rd variable we can identify those customers who are spending high are those with high annual income. But high spending customers are the one who are opting for revolver option i.e., pushing their dues to next month’s instead of paying it in full in same month.



**Heatmap**

* From the heatmap, we can identify the variables which are good predictors in providing conclusion for deciding credit card policies.
* **Correlation values** -
  + Total Income vs Credit card limit - 77% (Strong Positive)
  + Total risk score vs Credit card Utilization ratio - 69% (Strong Positive)
  + Total Income vs Average spends overs last 3 months - 68% (Strong Positive)
  + Credit Card limit vs Average spends overs last 3 months - 53% (Positive)
  + Total risk score vs Credit card limit - 53% (Negative)

**2.6. Top 5 important variables for GoDigit Bank to make profits with justification**.

* **Total Income of the Customer (annual\_income\_at\_source)** 
  + Income of the customer will give an idea to banks to whether give a credit card to a customer or not.
  + A high-income customer will utilize the credit card to it’s fullest and clear the dues within time and there’ll be no high risk. A valid point for banks to make profit and rely on customers with good credit history.
  + Usually, high salaried customer will be provided with credit card easily compared to the customer with low income.
* **Credit card limit (cc\_limit)** 
  + Credit card limit will let customer to know till how much to utilize and repay on time without any issues.
  + A high credit limit will let customer to spend more using card and hence banks can make profit only if customers spend more on card.
* **Average spends on credit card (avg\_spends\_l3m)**
  + If customer’s spendings on the credit card are high then it is a valid indication that banks can make profit from those customers.
  + Average spends on last few months of the customers data will give an idea to cluster customers those spending high, medium, and low. Then provide different benefits and offers for different clustered customers.
* **Total Risk Score (Total\_risk\_score)**
  + A risk score will be deciding factor whether to give a credit card to a particular customer or not.
  + High risk score always leads to bad impact on the customer and banks need to check those with high-risk score customers and try to lesser the credit card limits and update credit card policies accordingly.
  + Customers with low-risk score must be evaluated and loop them in with better offers and discounts which will eventually increases the revenue of the banks.
* **Credit card utilization ratio (Credit\_Card\_Utilization\_Ratio)**
  + A better usage of credit card will have positive impact on the bank’s revenue.
  + Customers with high ratio should be taken care and eventually those with low ratio needs to be addressed
  + A ratio can give us better understanding of usage of the credit card for different set of customers.
  + We can define a threshold ratio for customers who are falling low ratio and revert the policies.