# BANK LOAN CASE STUDY

# DESCRIPTION

- In this project we are going to identify patterns which indicate if a client has difficulty in paying their installments which may be used for taking actions such as denying the loan , reducing the amount of loan ,lending at higher rate of interest.
- We are going to analyse the data by EDA(Exploratory Data Analysis) which is used to summarize ,visualize and understand the characteristics of the given dataset.

# APPROACH

Approach towards the project means Overall strategy that outlines how the project will be executed or done to achieve its Objectives.

In this project first we need to import the dataset that provided and load that into Excel. After loading the data use specific functions and formulae in excel in order to achieve the required tasks in the project. If needed use charts to visualise the data in effective way.

# TECH USED:

* MS EXCEL ( FOR ANALYSIS )

*JUPYTER ( FOR CLEANING PREVIOUS APPLICATION )

*MS POWERPOINT ( FOR MAKING PPT OF PROJECT )

# CLEANING DATA

**Cleaning** the data is an crucial step in any data analysis as it ensures the data is accurate, reliable and consistent .

Without data cleaning ,Our data analysis will be inaccu**RATE** ,incomplete and inconsistent which can lead to serious consequences in decision making.

There are some steps involved in Cleaning the data .Those are discussed below in detail

STEP 1 : **Removing duplicates**

First we need to check if there are any duplicates present in the primary key of the given data set. Here I took SK_ID_CURR as primary key because it's the unique value of the given data set.I removed duplicate by using "Remove Duplicates " feature in Excel by selecting category as SK_ID_CURR

## STEP 2: **Finding Missing Values**

Missing values are those which are not available in our dataset due to varIOus reasons such as human errors or system issues.We should handle those values. First I found the count of null or missing values in each column using COUNTBLANK function in Excel and also calculated percentage of null values in each column.

## STEP 3: **Handling Missing Values**

After finding the percentage of Missing values in each column and I deleted the columns which are having blanks percentage more than 50 %. Because the analysis wouldn’t be fair if we are having majority values in columns are missing. So I used this Criteria to delete columns.

## STEP 4:**Removing Irrelevant Columns**

There are many unnecessary columns present in our dataset .It is better to delete those columns as it helps to reduce the size of dataset ,improve efficiency of our analysis and eliminate irrelevant information.In the given data set columns giving normalised information about building where the client lives,apartment size ,living area, common area, number of elevators , number of entrances, number of floors etc..I deleted these columns and also majority of these columns having missing values.

STEP 5: **Imputing with Mean/Median**

We can replace the values in columns with less than 50 % with Mean or Median depending upon the category of the column.

Here , For continuous variable columns such as Amount_Annuity , Ext_Source_3 , I found mean of that columns in the Excel using “**MEAN**” function and then replaced null values using “Find and Replace” option  by keeping null in “Find” and mean in “Replace”

For categorical columns such as AMT_REQ_CREDIT_BUREAU(Number of Enquiries), I found most repeated category by using “**MODE**” function in Excel.Then I replaced all missing values with Mode of that column

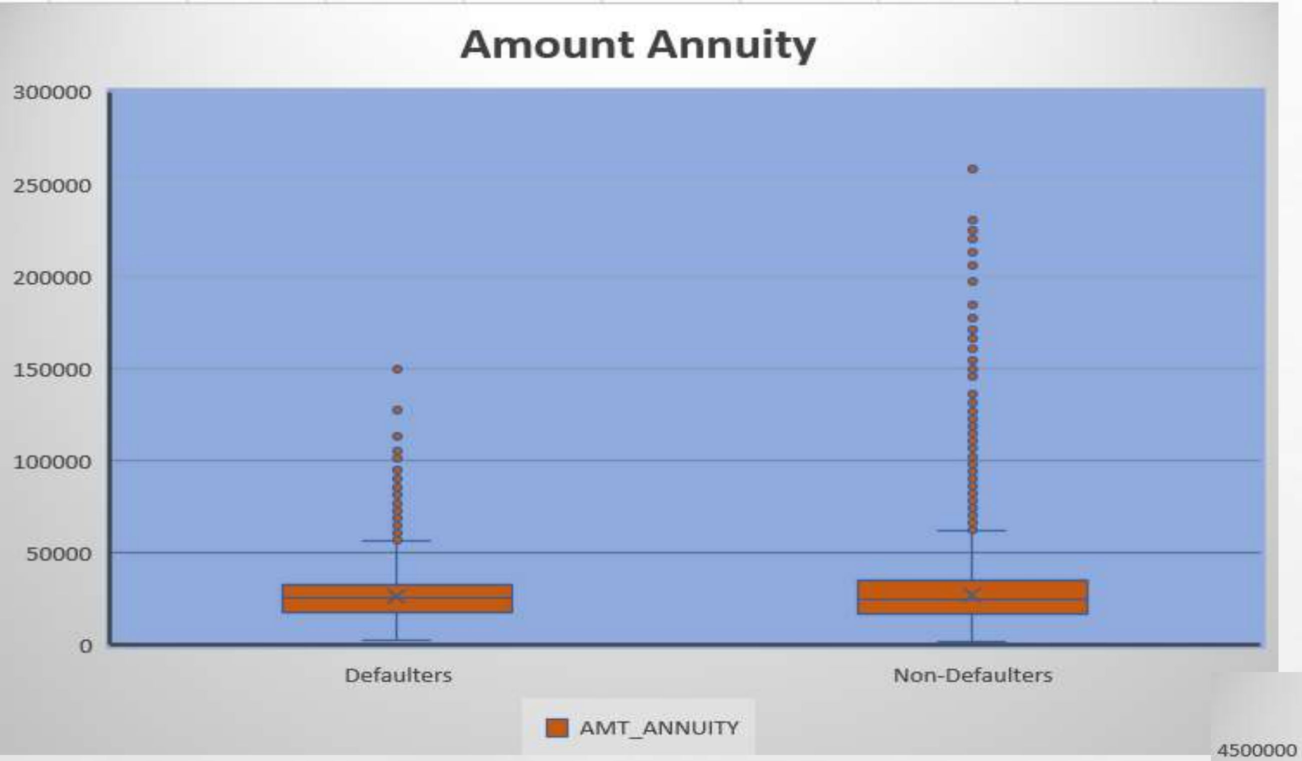
# IDENTIFYING OUTLINERS

- We know that the **Outliers** are the data points that lie far away from the rest of the data points due to some measurement errors , data entry errors and some other natural errors.
- These can affect the statistical analysis of the data as they reduce accuracy of models.I found those outliers using Box and Whisker plot .
- We can also find outliers using Tukey's method.

Steps to find Outliers using Tukey's method:

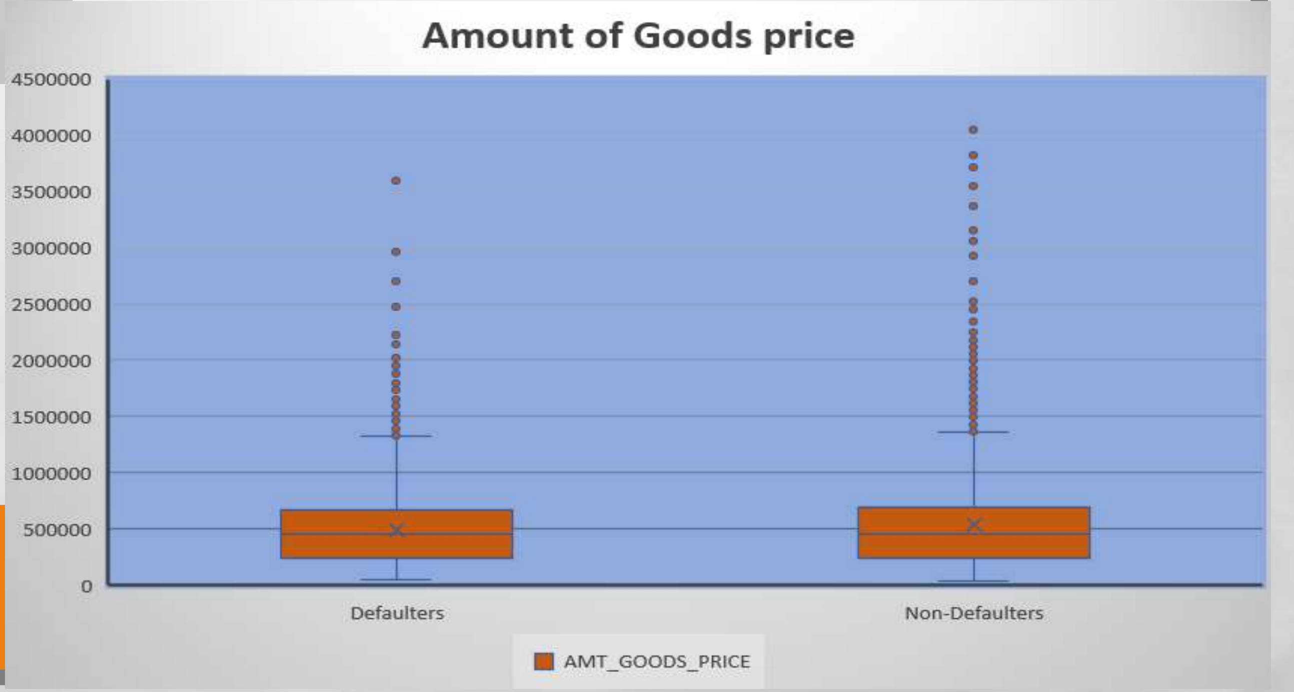
1.Finding 1st Quartile Q1 and 3rd Quartile Q3
2.Finding Inner Quarter Range(IQR)
3.Finding Upper Bound(Q3+(1.5*IQR)
4.Finding Lower Bound(Q1-(1.5*IQR)
5.Now, any data point above Upper Bound or Below Lower Bound Considered as Outlier

Below , I plotted some Box-Whisker Plots to find there are any outliers present in the data.

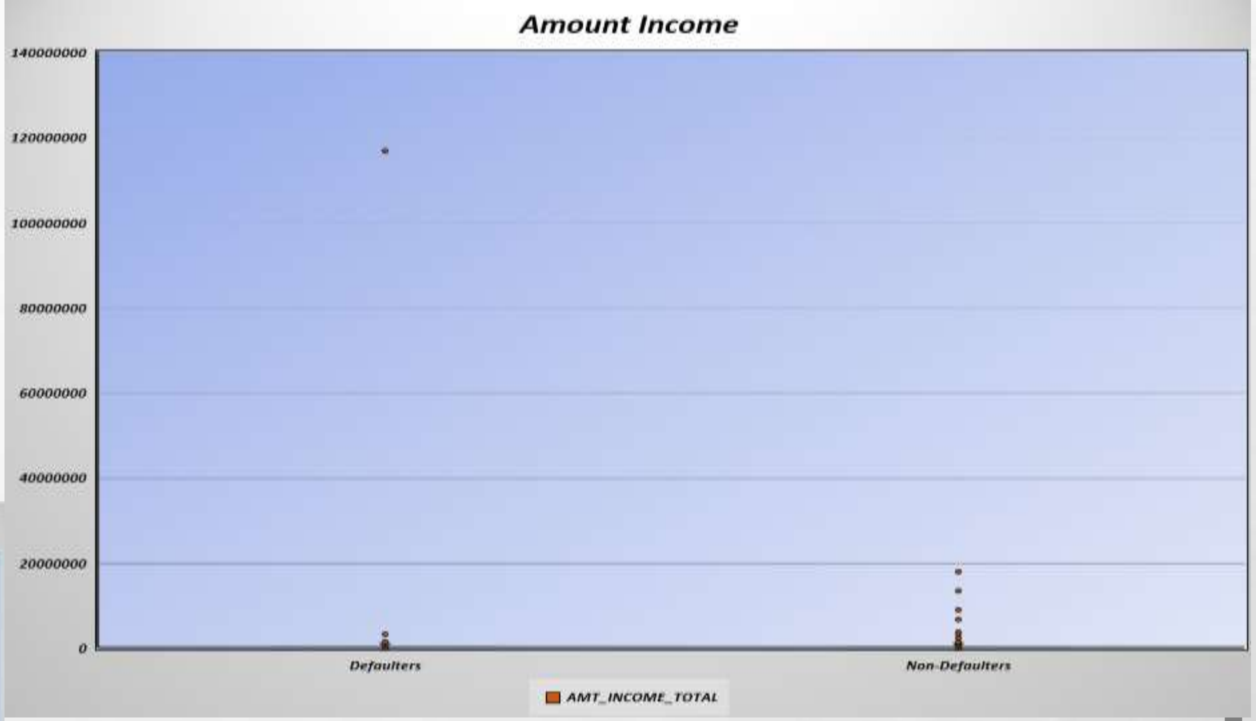


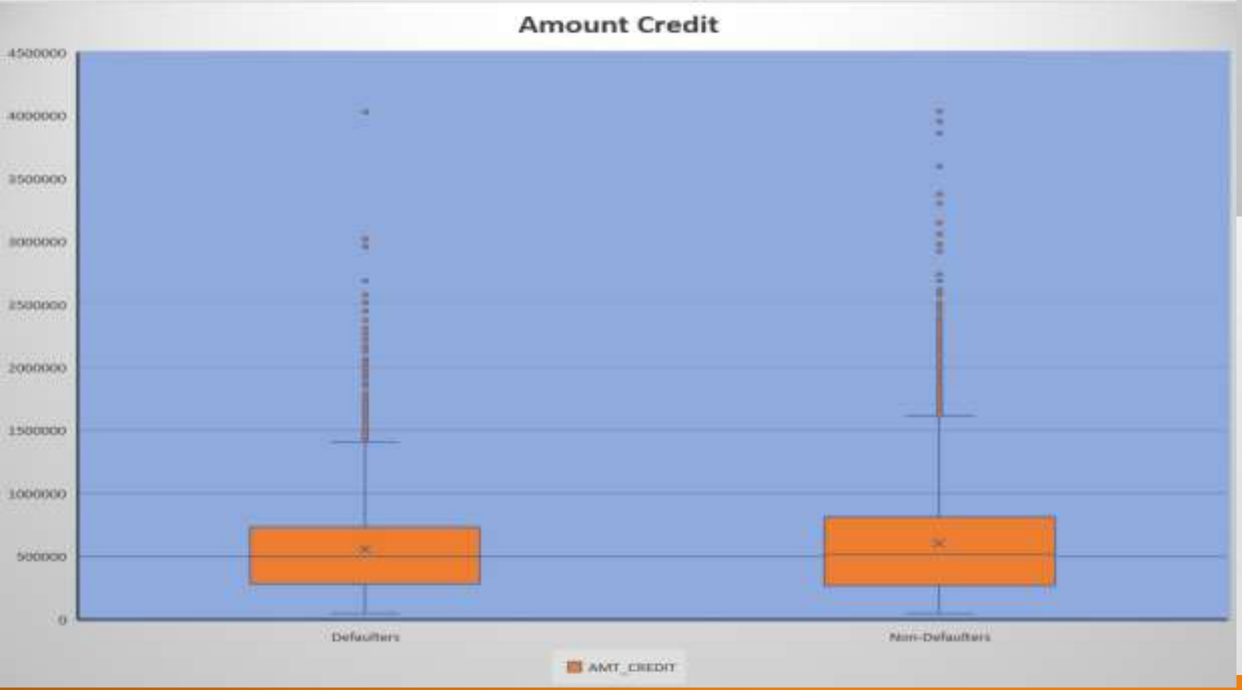
In Box-Whisker plot. Box represents the middle 50% of the data i.e. inner Quarter Range ,while the whisker represents range of data outside of the box. Outliers are the data points that lie outside the whiskers

We can see that there are outliers in “Amount Annuity” and “Amount of Goods Price” as there are many data points lying outside the Whiskers
Non-Defaulters category having more number of Outliers when compared to Defaulters



In “Amount Income” column ,there are Outliers present in both Defaulters and Non-Defaulters column. We can clearly see that the data point touches 120 million mark. These values can affect the statistical analysis of the data



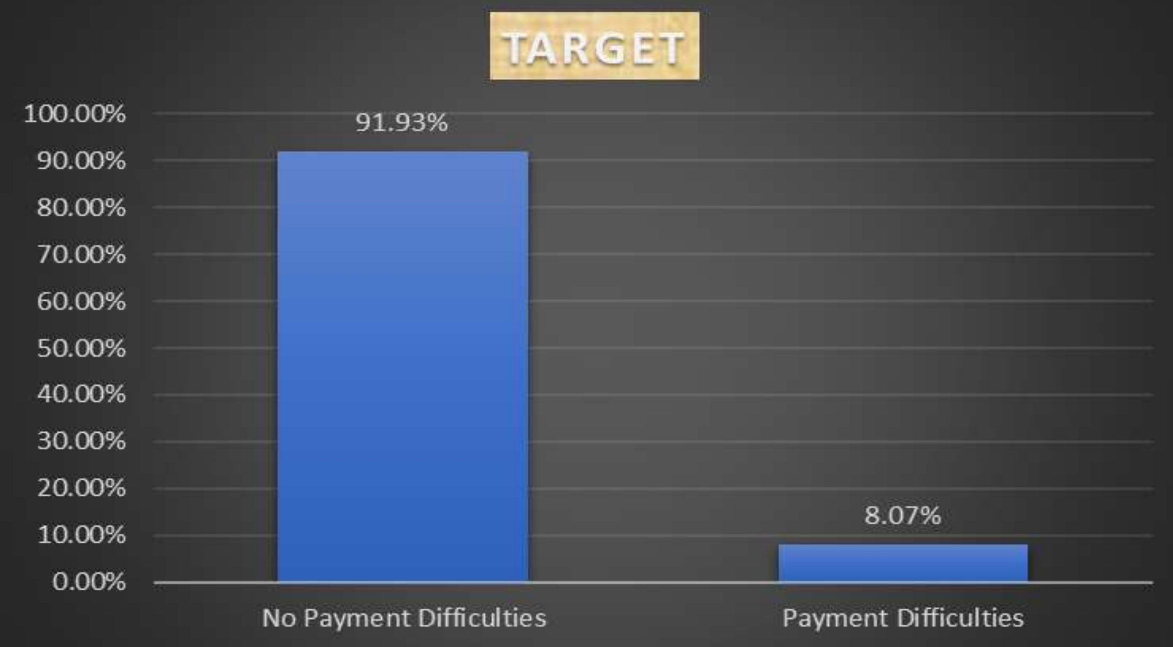


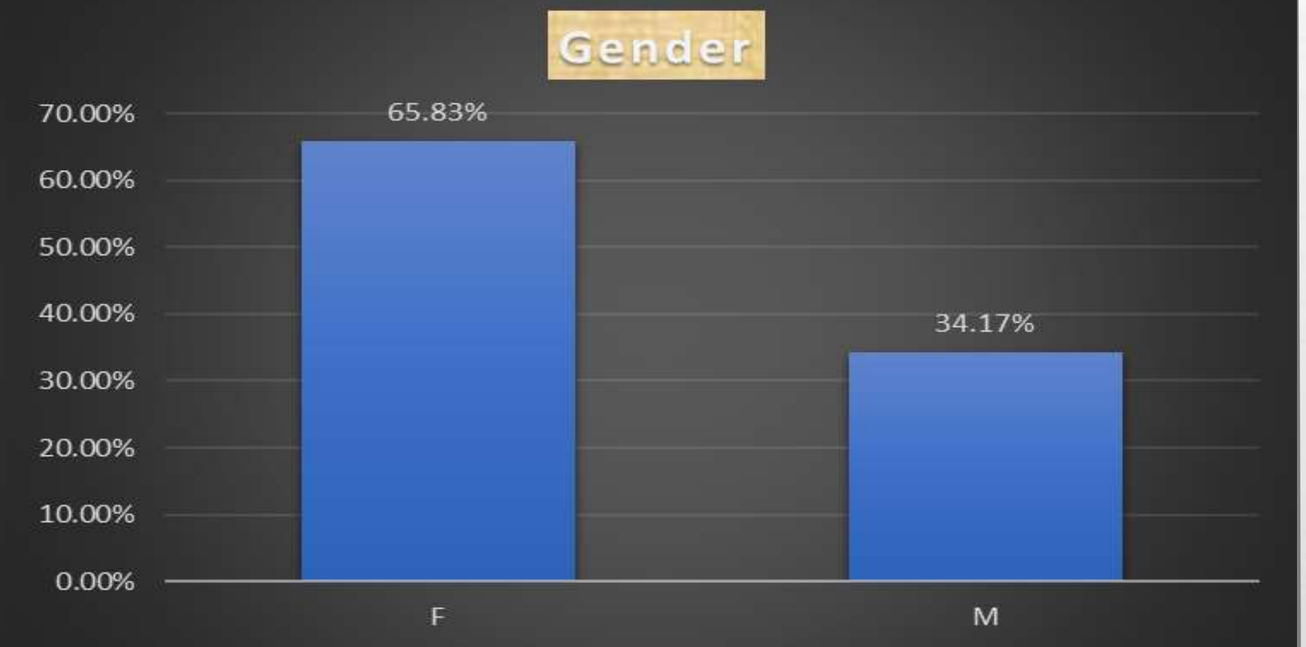
In “Amount Credit” Column ,we can observe that both Defaulters and Non-Defaulters ,Large number of Outliers present in the data. In EDA [Exploratory Data Analysis], It’s not necessary to remove data points.

# DATA IMBALANCE

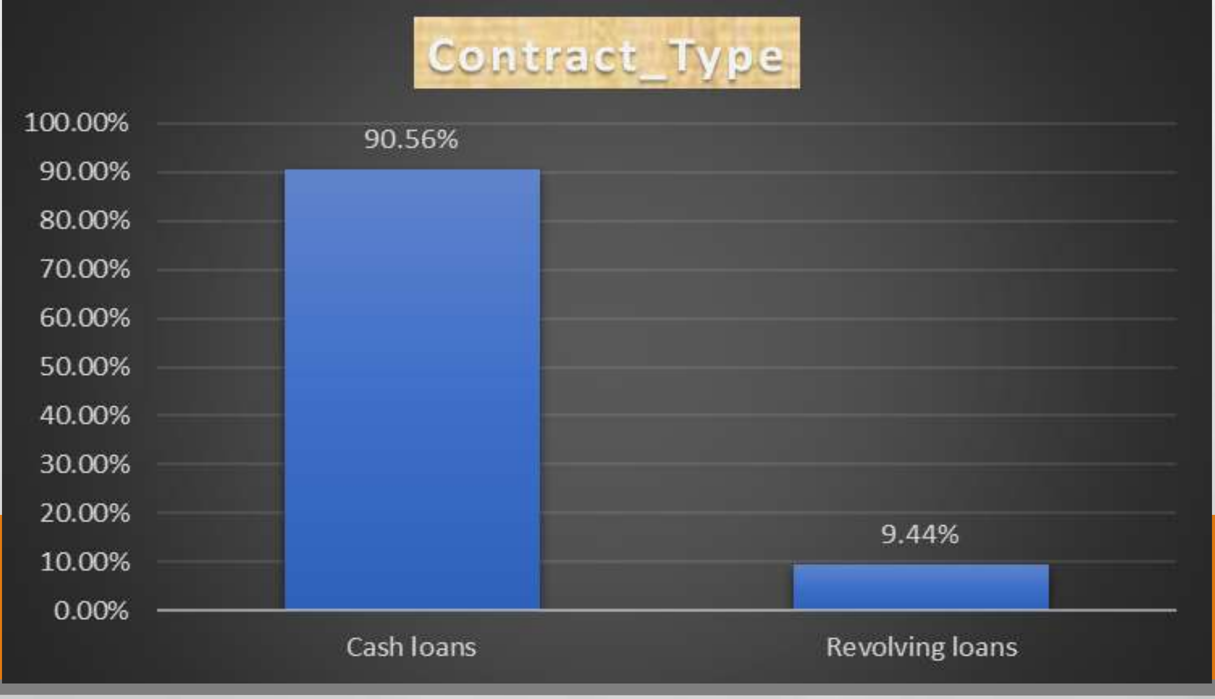
- Data Imbalance refers to situation where the classes in classification problem are not equally represented in the dataset.
- Imbalanced datasets can lead to poor performance in data analysis because data model may be biased towards majority of classes and have difficulty in identifying minority classes.

- We can find the data imbalances by creating the pivot chart of Histogram in excel.
- First we need to insert cleaned data into Pivot Table and then select the column you need to find if there is any data balance in “rows” field and again inti “columns” field by changing summarize value by “Count”
- Then we can change “show value as” into “%of grand total” and create a pivot chart by selecting Histogram or any other charts to visualise data imbalance.



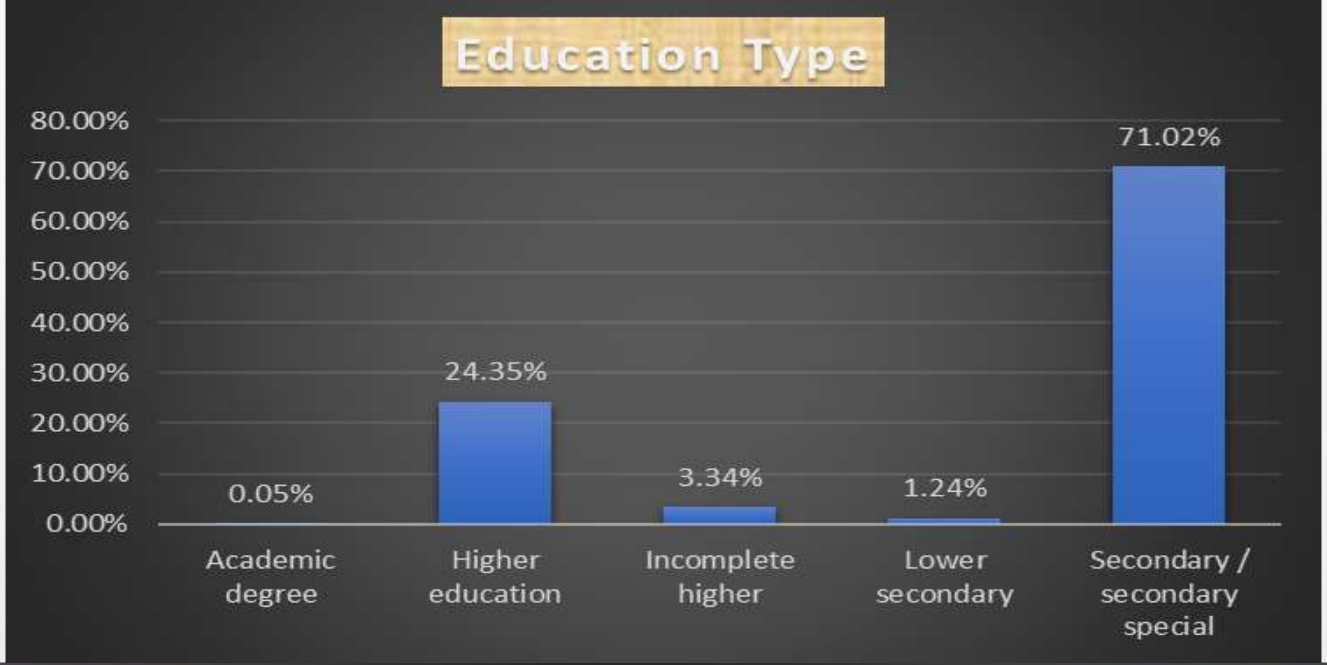


- In Target variable, Clients with No Payment Difficulties are 91% where as Clients with Payment Difficulties are only 9%.Clearly there is imbalance in data. In Gender Category Males(34%) are nearly half of the percentage of Females(66%).

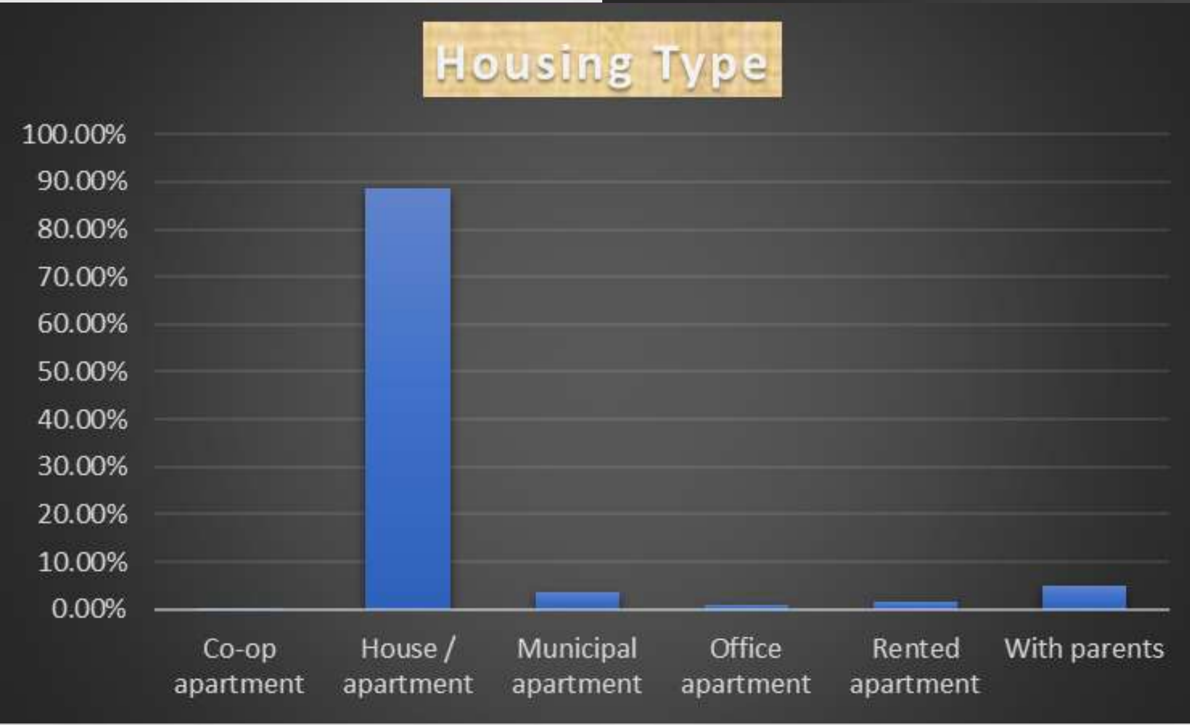


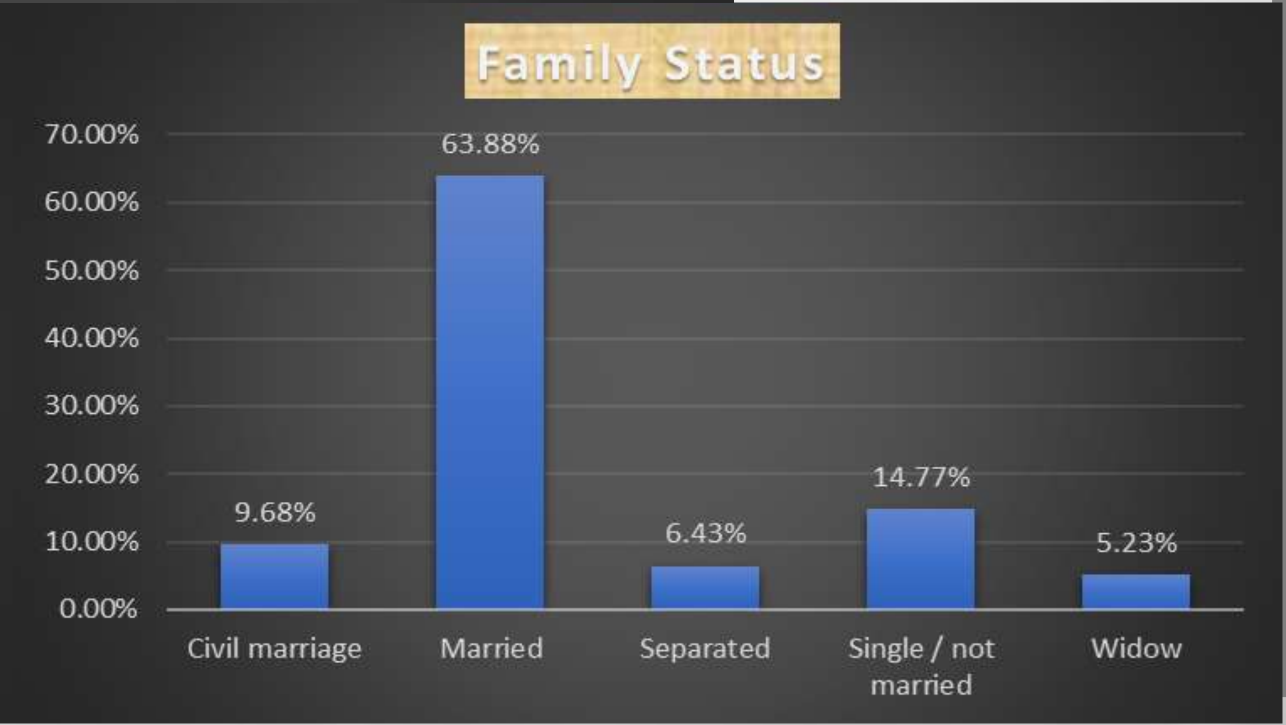
- In Contract_Type , Cash loans are nearly 91 % where as Revolving loans are just 9 %

- ➢ Many of the Clients Education is Secondary/Secondary Special (71%)and there are least clients (.05%) having academic Degree



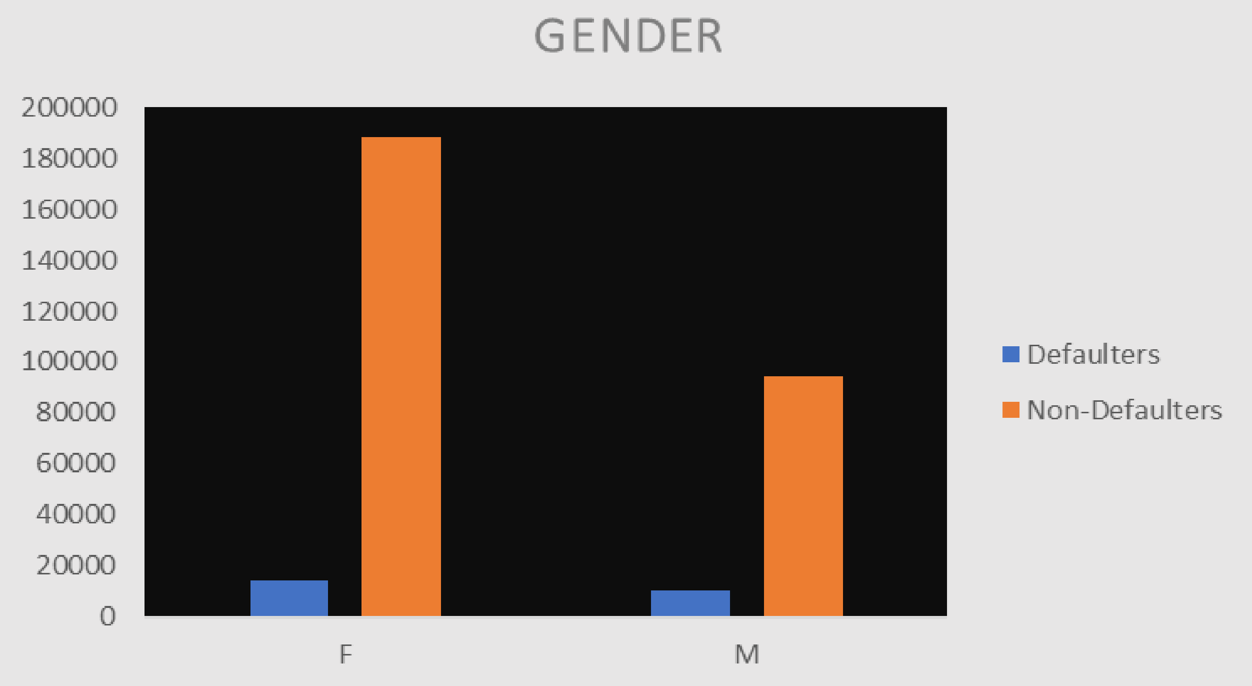
- ➢ Many of the Clients Family status is Married(64%)  and Most of the clients (Housing type) live in House/Apartment(90%)

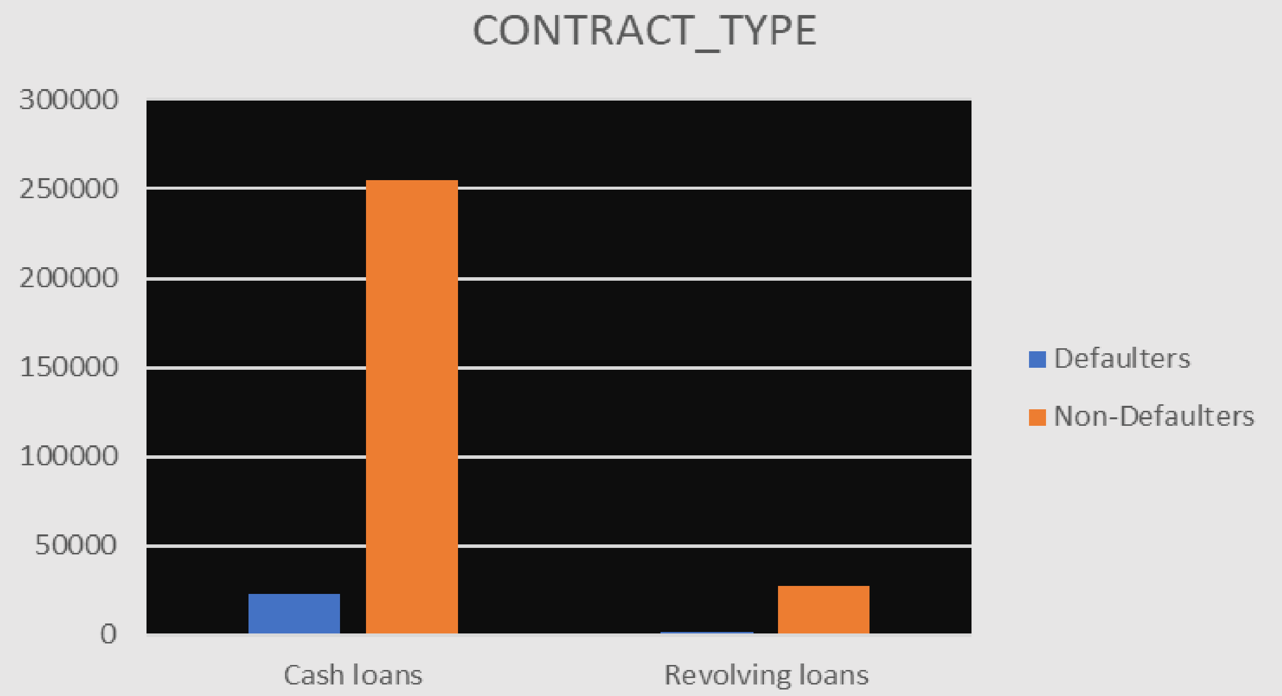


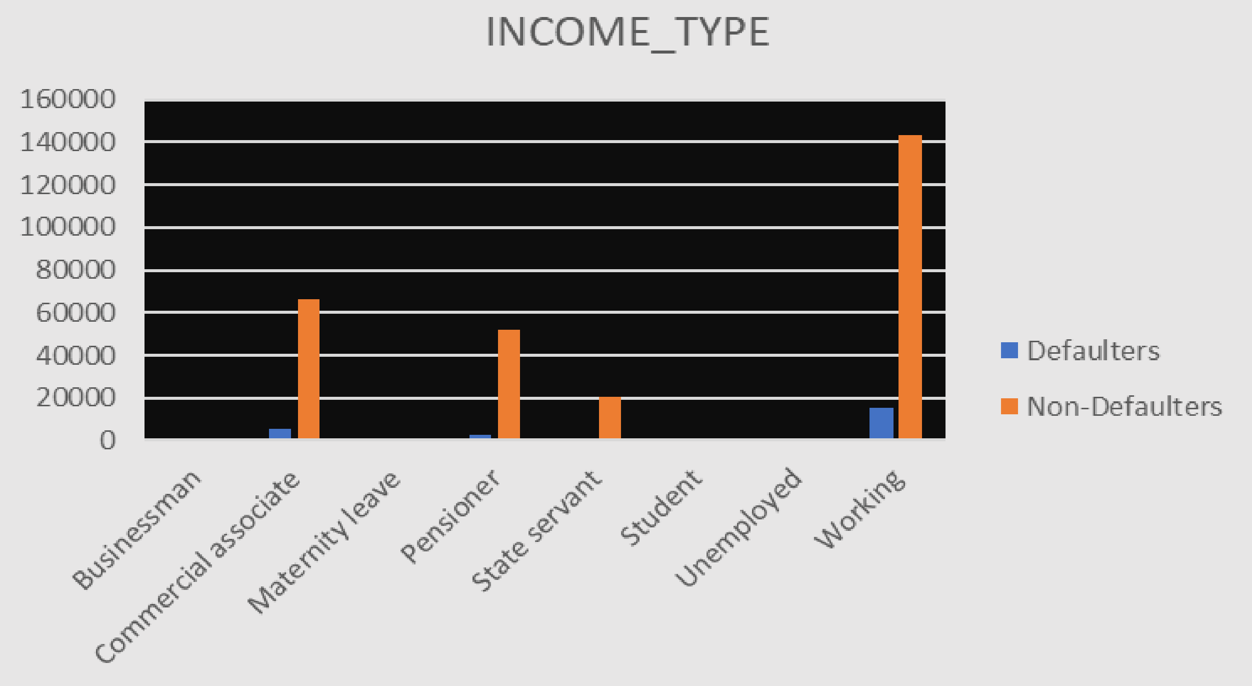


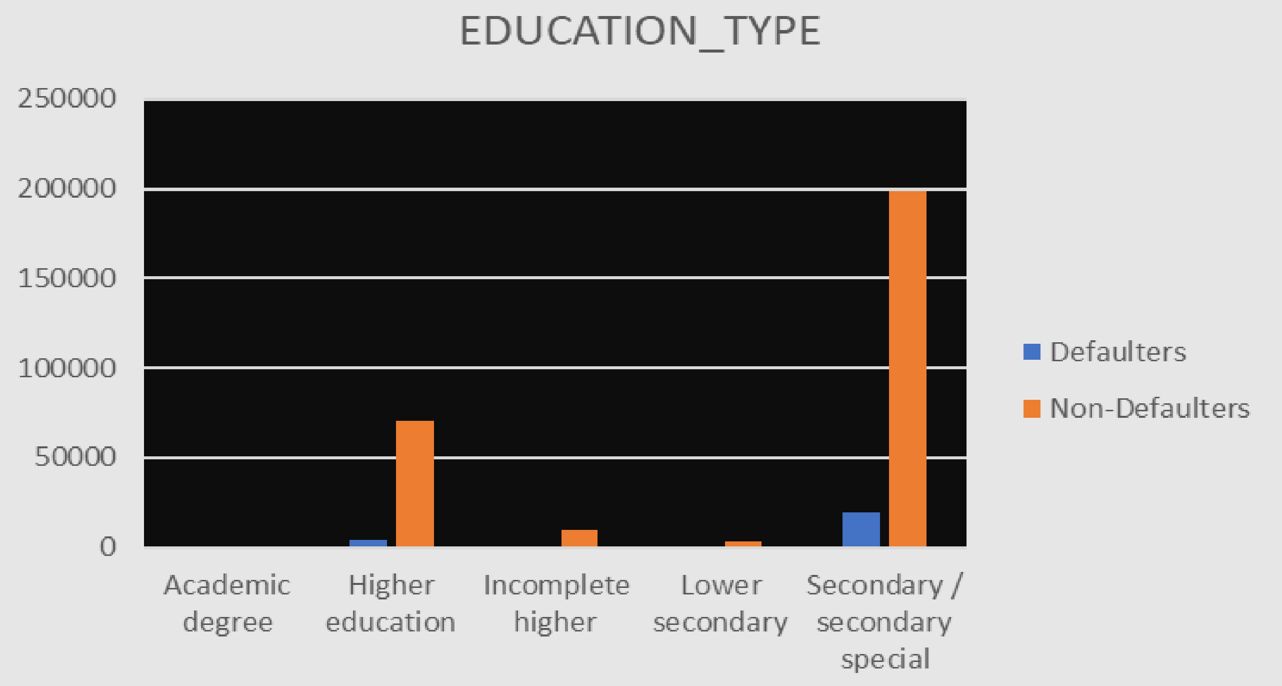
# UNIVARIATE ANALYSIS ON CAREGORICAL VARIABLE

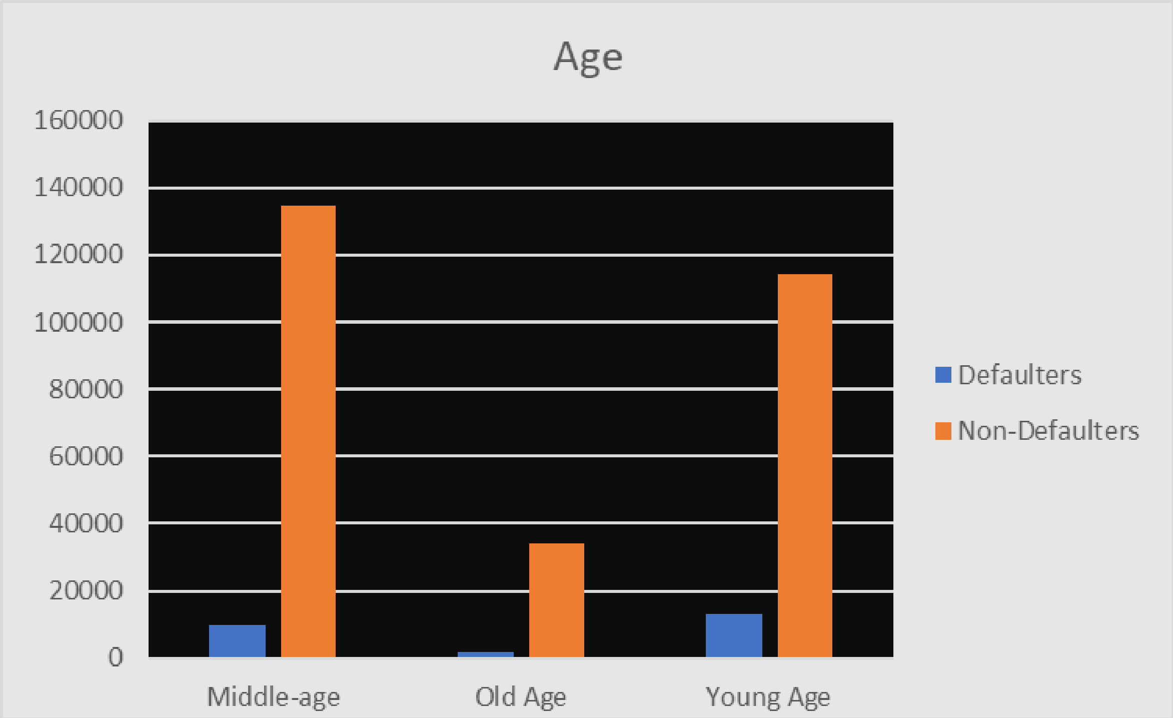
- Univariate analysis is statistical method to analyze the data with one variable.
- It involves the examining the distribution of single variable and deriving insights from it.
- Univariate analysis of categorical variables involves summarizing and examining the frequency or proportion of each category to gain better understanding of distribution and relationship between variables


GENDER
200000
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
F
M
Defaulters
Non-Defaulters

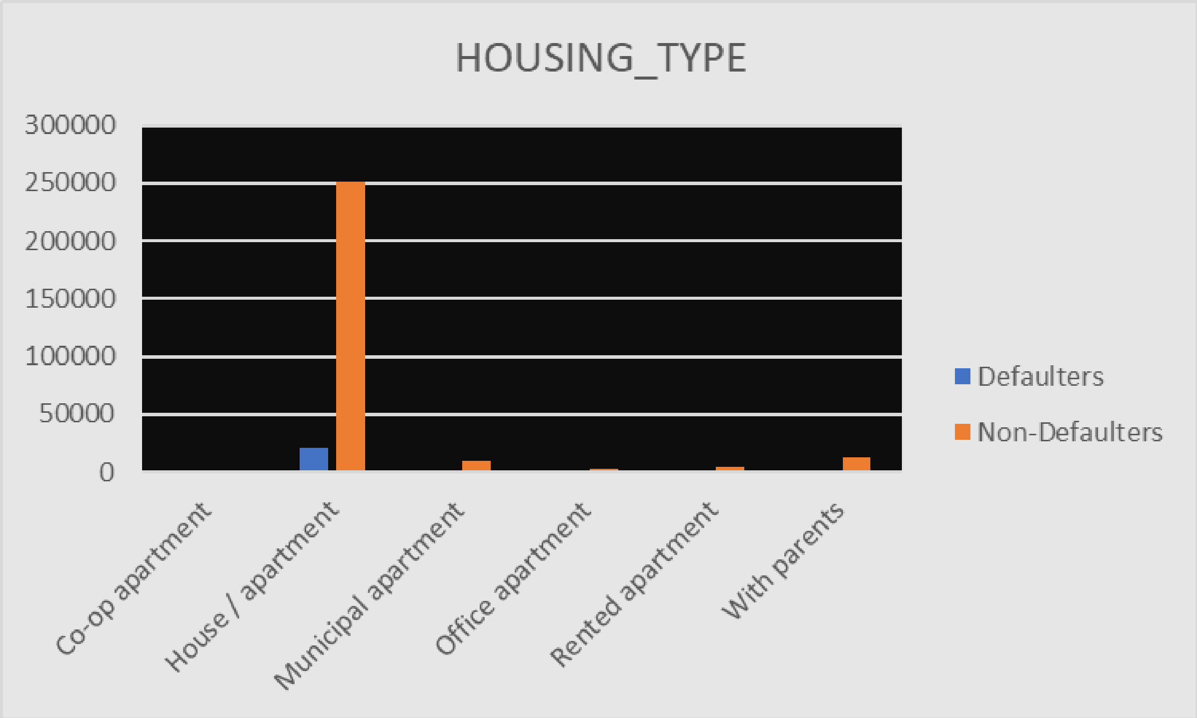

CONTRACT_TYPE
300000
250000
200000
150000
100000
50000
0
Cash loans
Revolving loans
Defaulters
Non-Defaulters


INCOME_TYPE
160000
140000
120000
100000
80000
60000
40000
20000
0
Businessman
Commercial associate
Maternity leave
Pensioner
State servant
Student
Unemployed
Working
Defaulters
Non-Defaulters

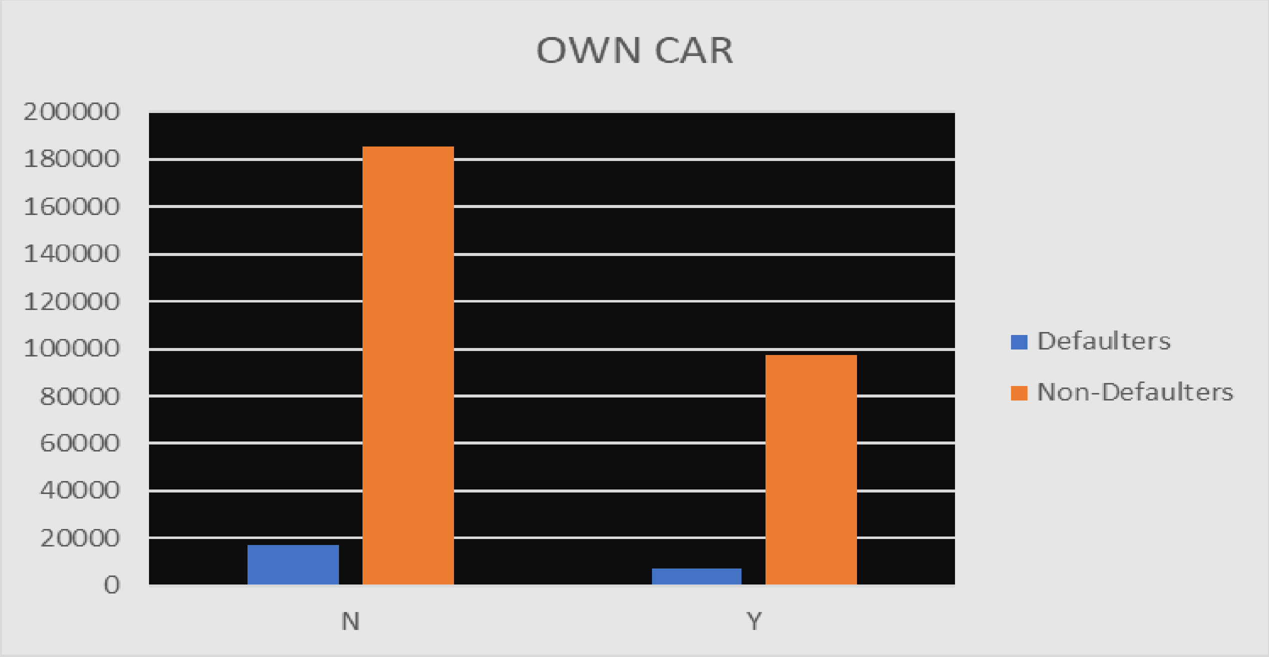

EDUCATION_TYPE
250000
200000
150000
100000
50000
0
Academic degree
Higher education
Incomplete higher
Lower secondary
Secondary / secondary special
Defaulters
Non-Defaulters



Age



HOUSING_TYPE

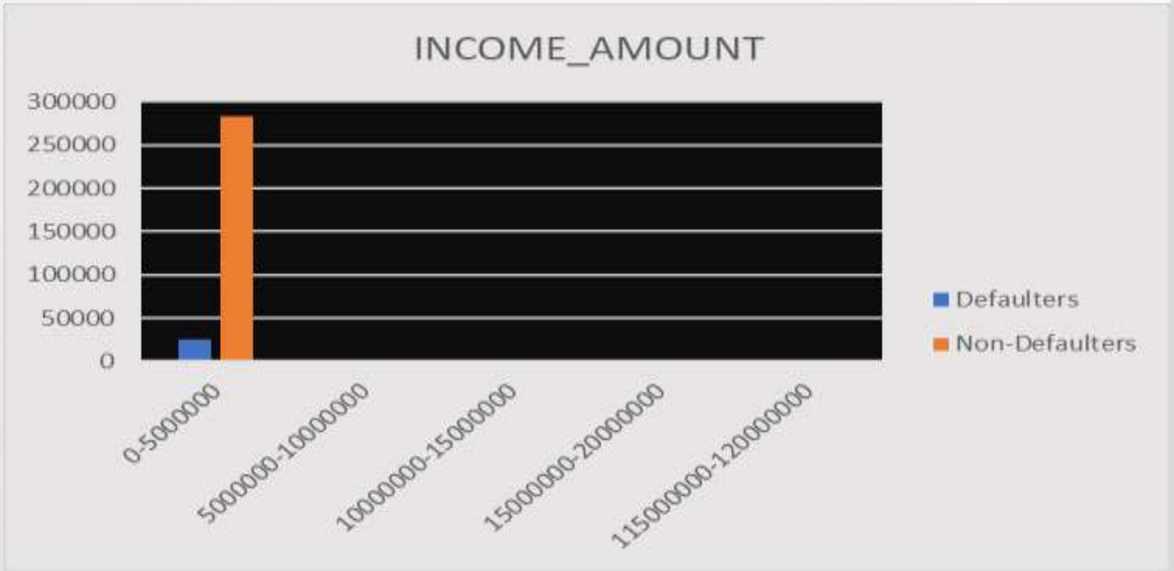


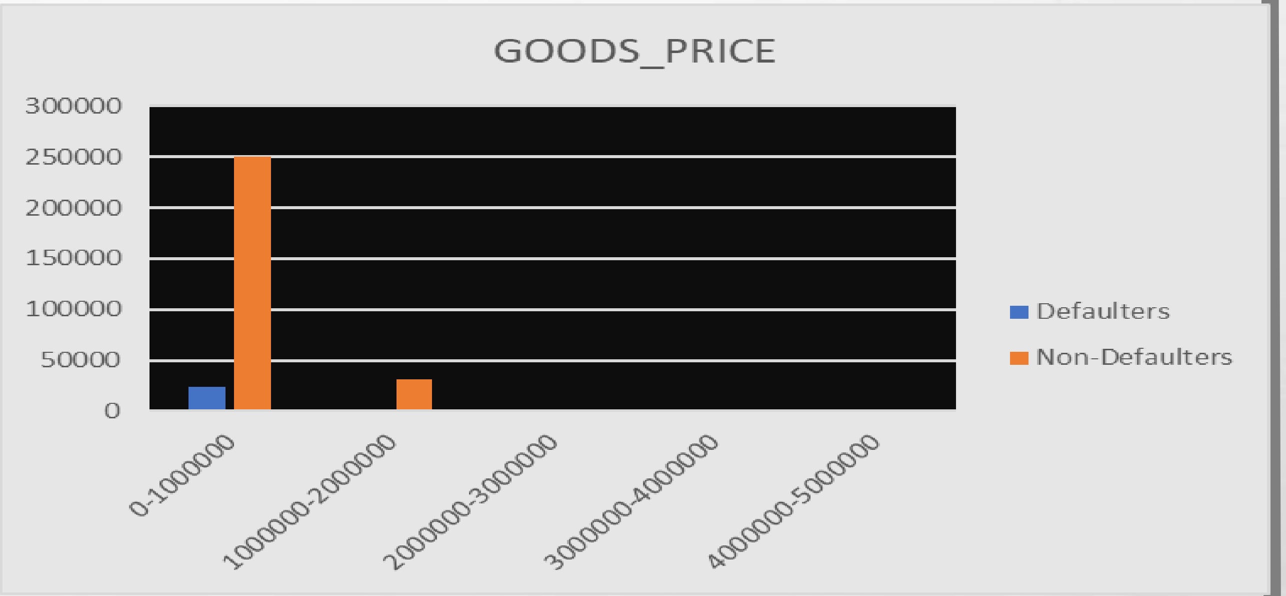
OWN CAR

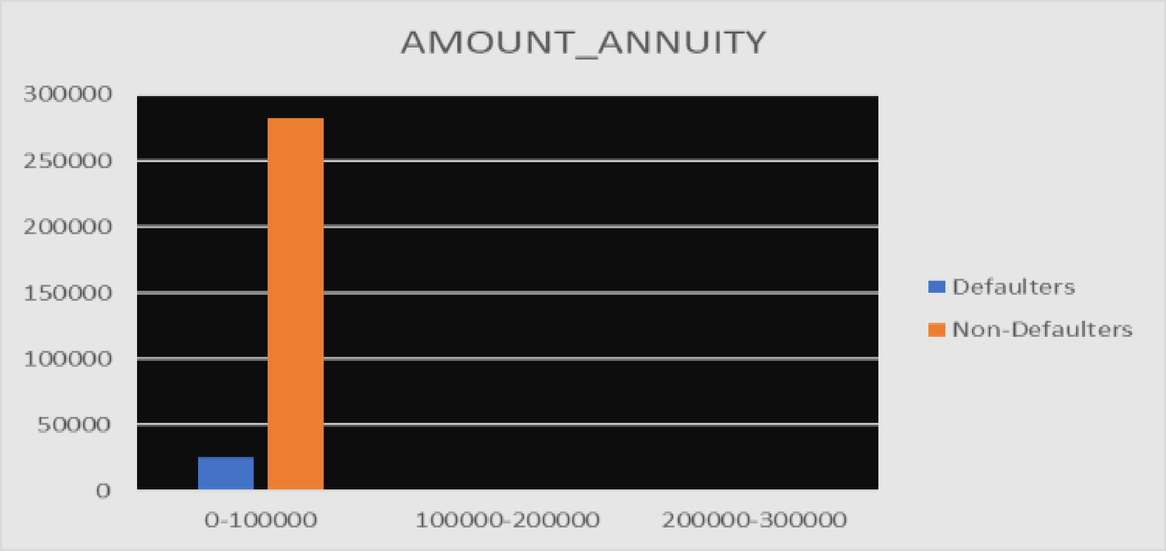
# INSIGHTS

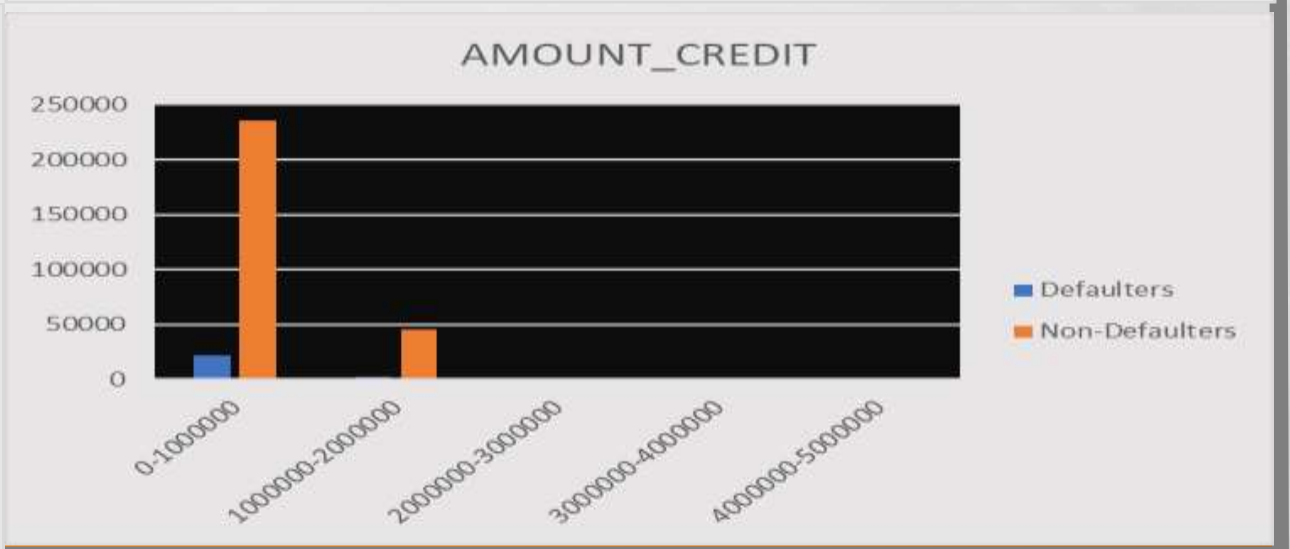
- **Gender :** We can infer that many of the Clients are Females when compared to males.Females clients having more payment difficulties when compared to Male Clients
- **Contract_Type :** Clients prefer Cash loans when compared to Revolving loans and also Client gets Payment difficulties in it .
- **Income_Type :** Income of the many clients who took loans are from Working profession and unemployed ,student took least number of loans
- **Education_Type:** Clients who done Secondary Education took maximum number of loans and some of them having payment difficulties
- **Age_Bin :** Middle Aged and Young Age people took high number of loans Meanwhile Old age people have taken less number of loans.Young Age people having more Payment Difficulties when compared to others
- **Housing_Type:** Many of the Clients are living in House/Apartment .Clients who are having rented apartments finding difficulty in paying loans.
- **Own_Car:** We can observe that there is not much variation in Clients having car or not. Both data distribution is much similar

# UNIVARIATE ANALYSIS ON CONTINUOUS VARIABLE

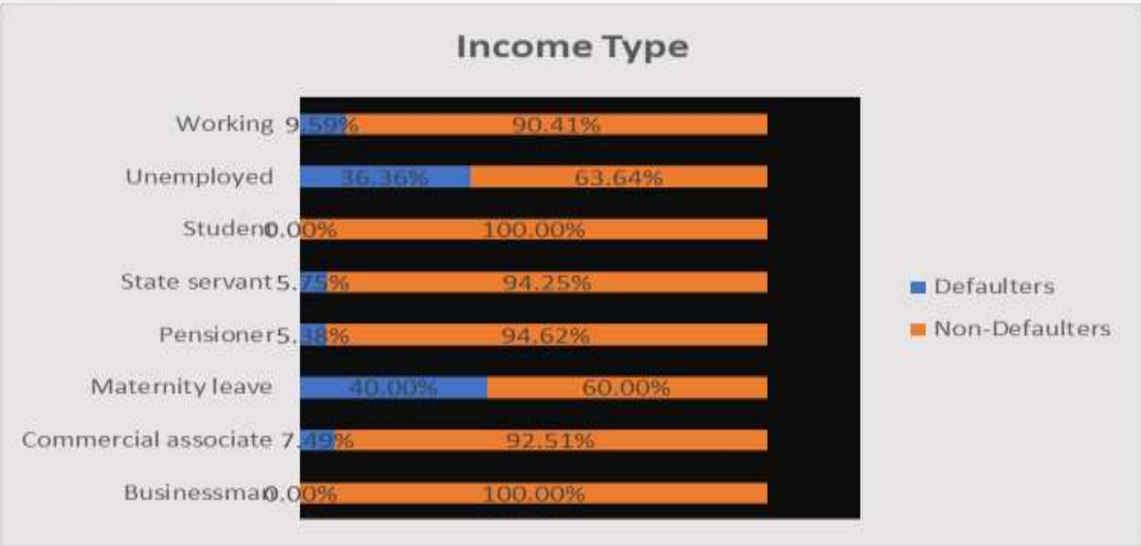


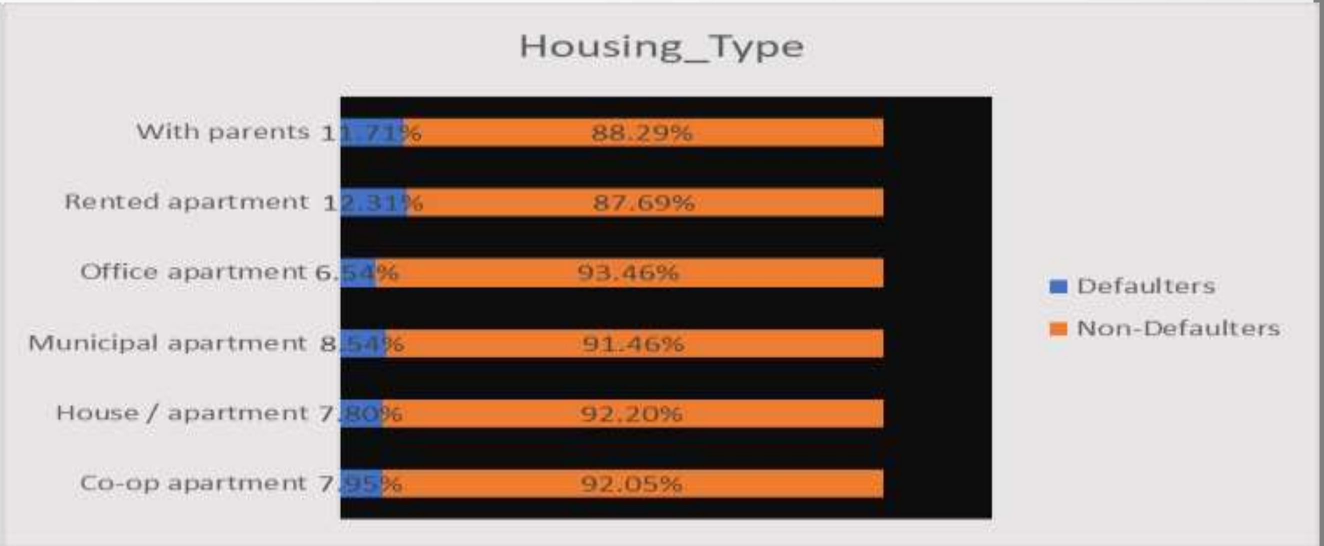


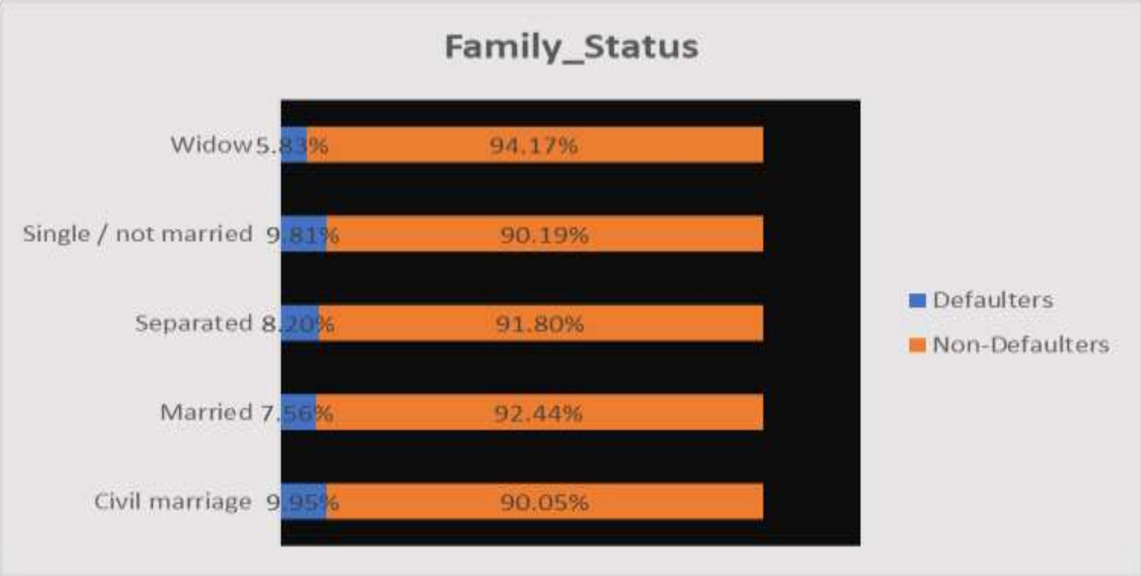


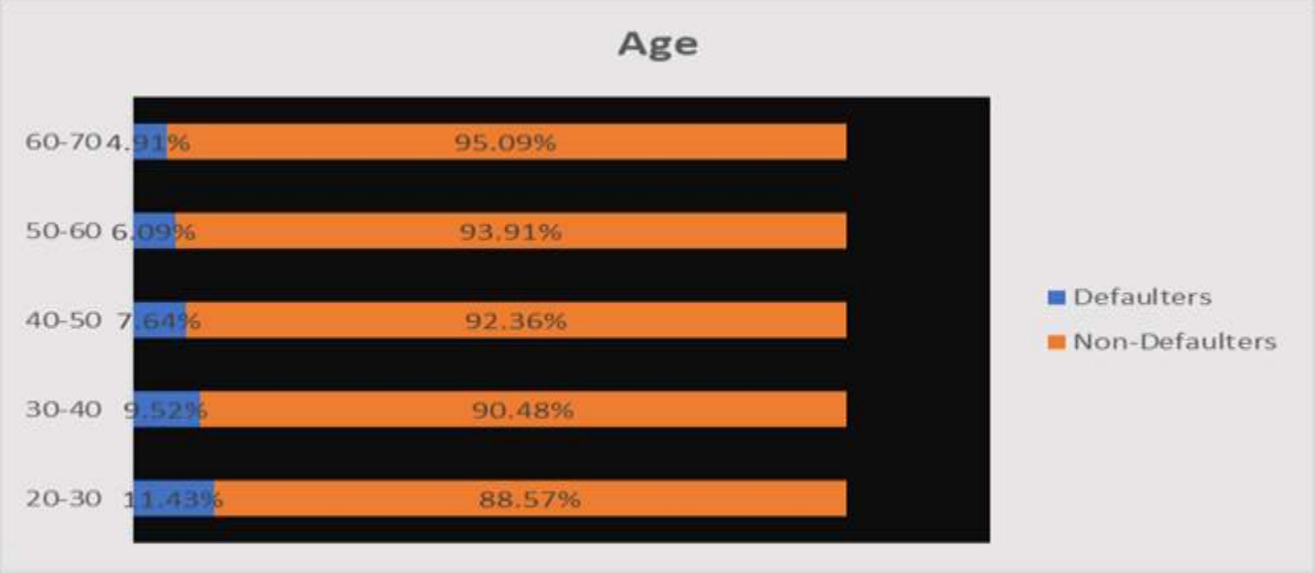


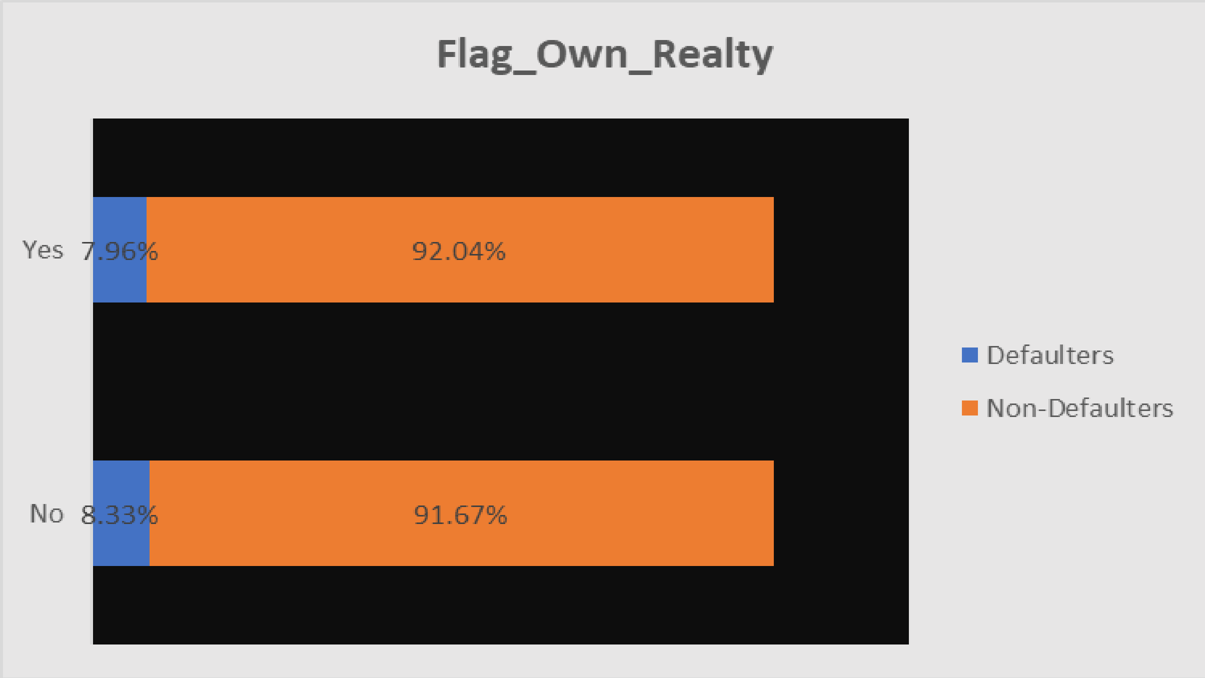
# UNIVARIATE SEGMENTED ANALYSIS


Income Type

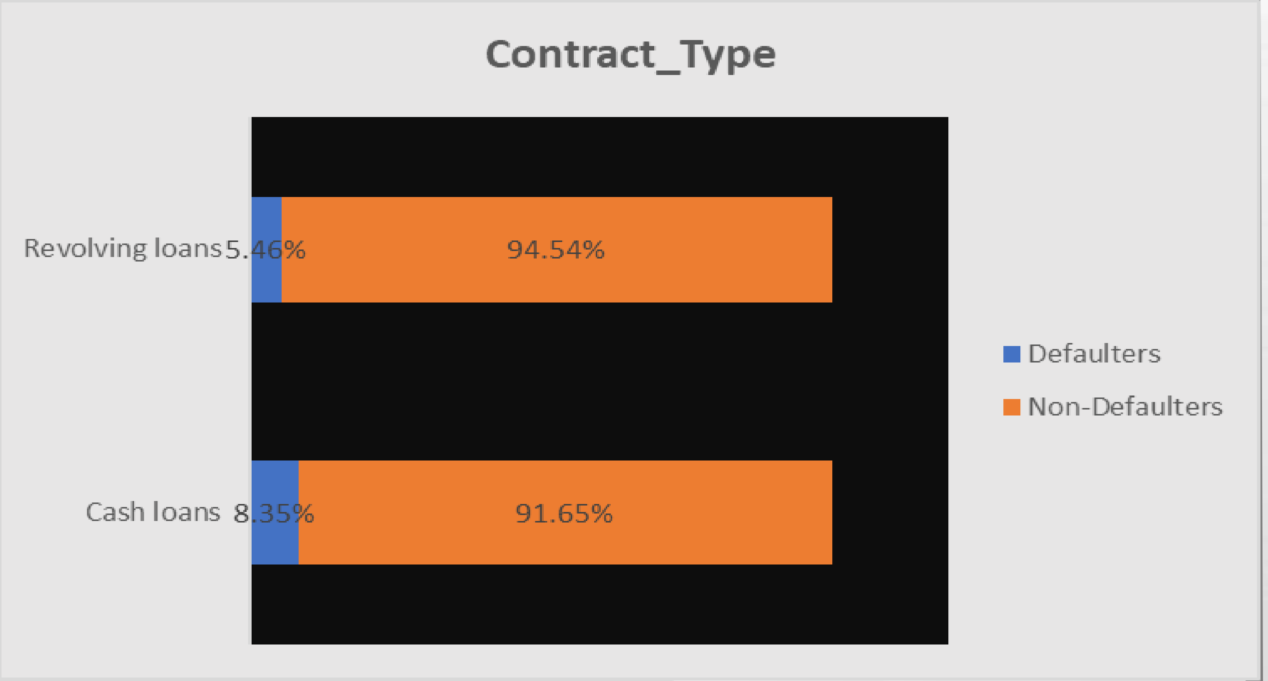

Housing_Type


Family_Status

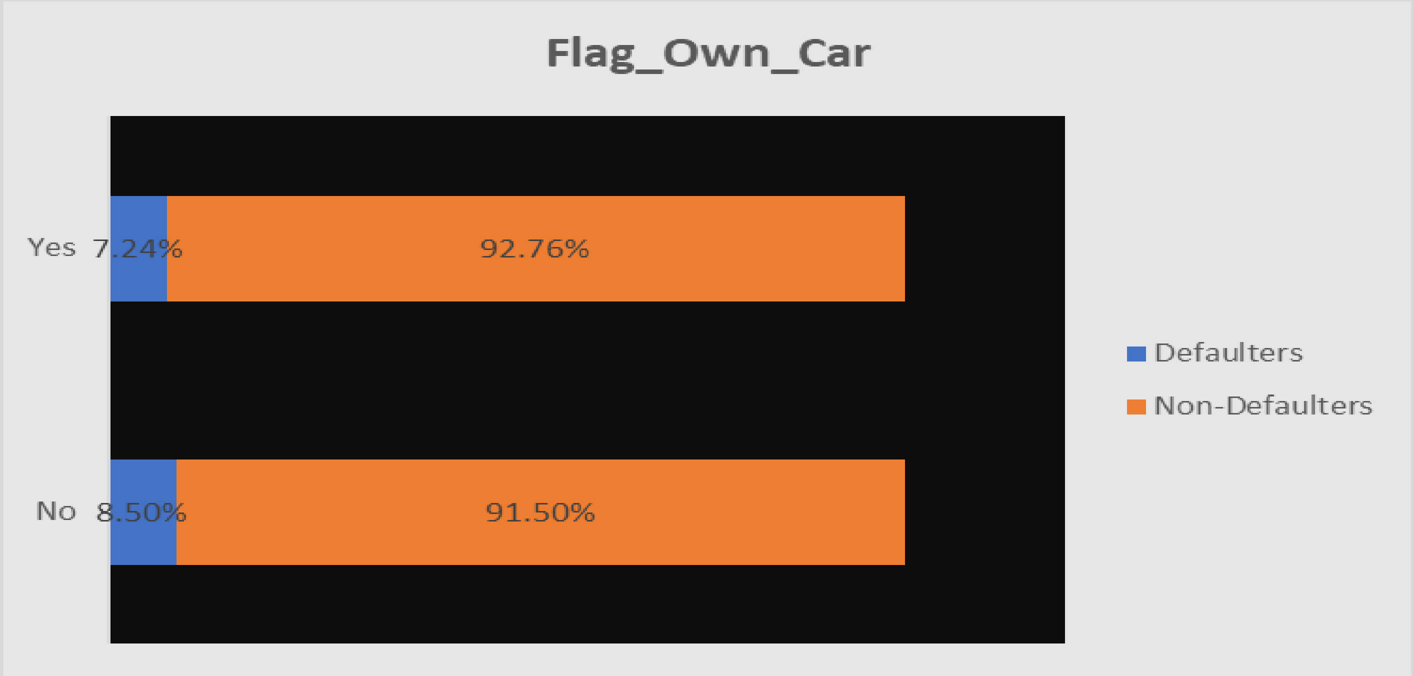

Age



Flag_Own_Realty



Contract_Type

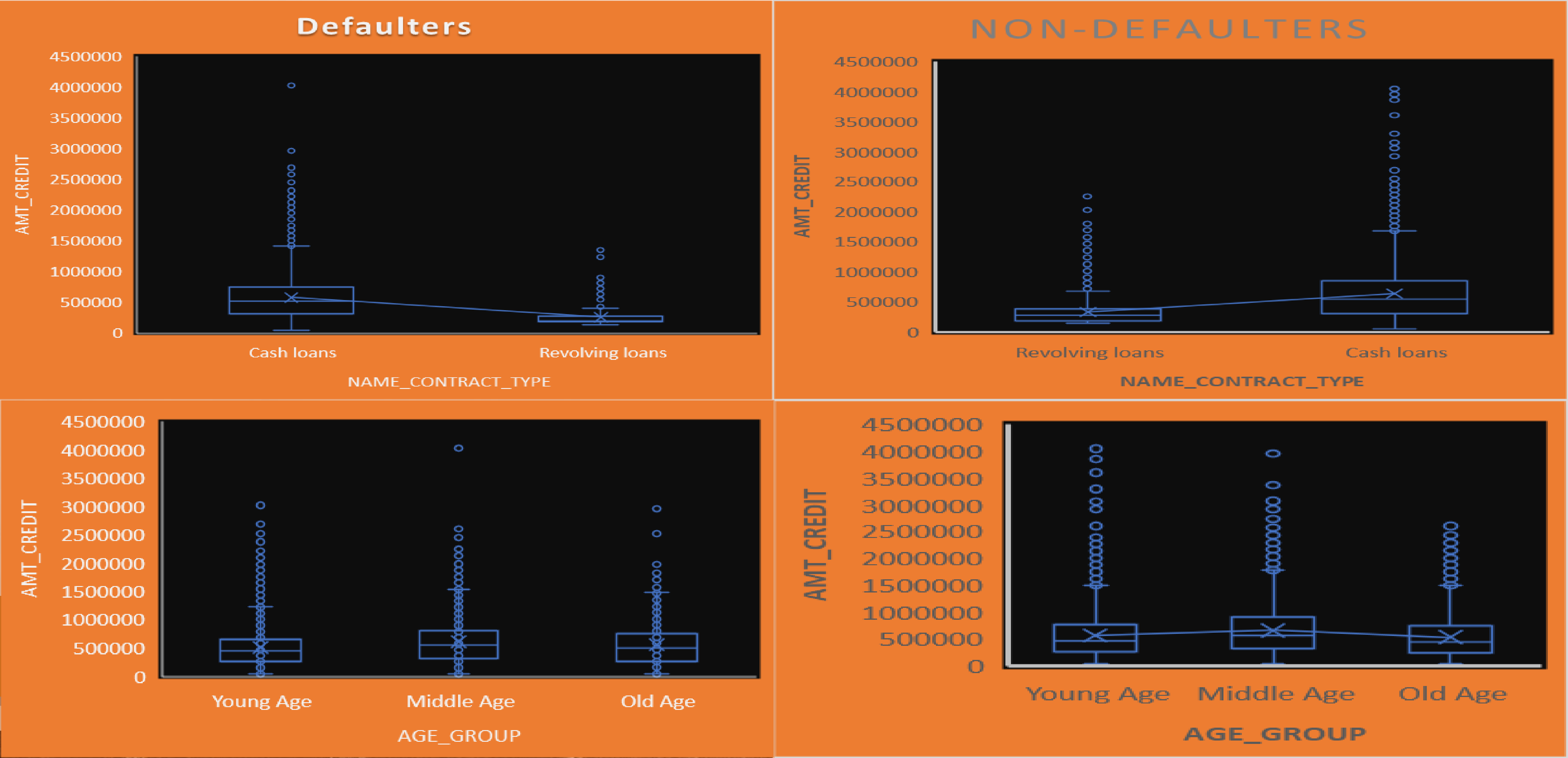


Flag_Own_Car

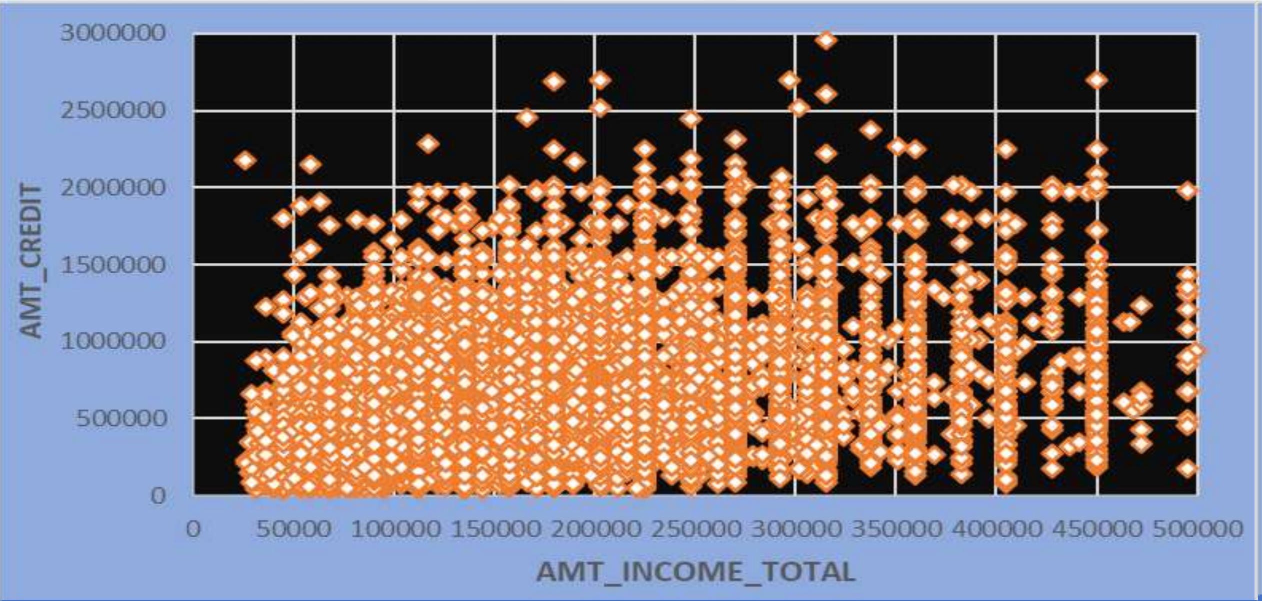
# INSIGHTS

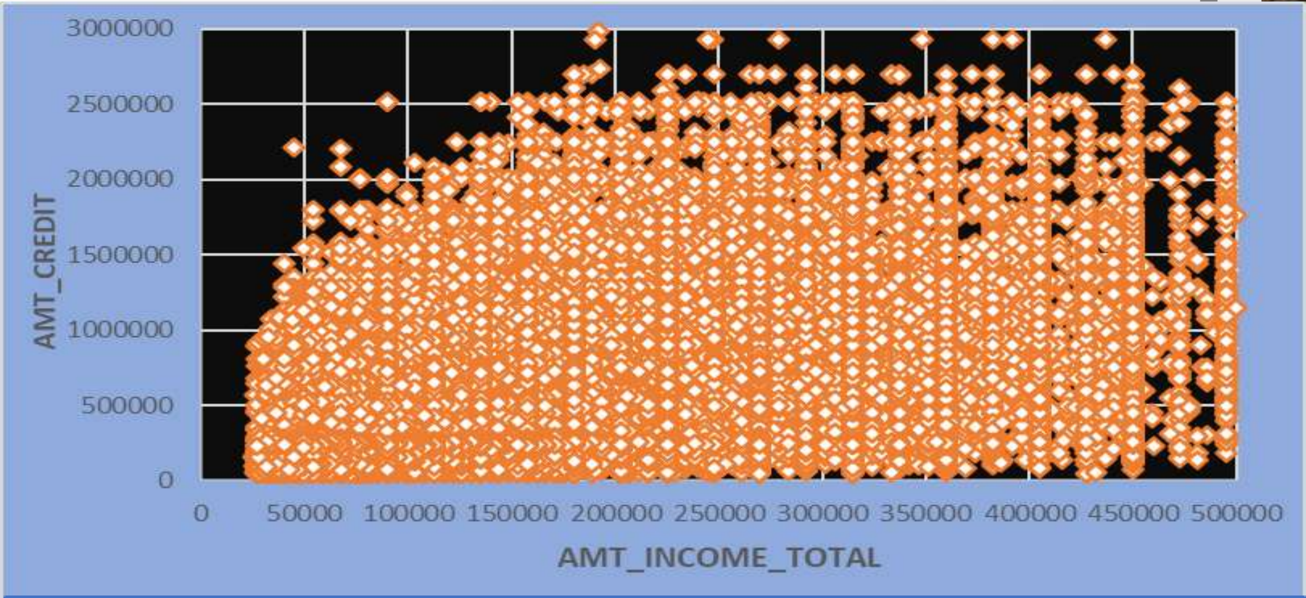
- **Income Type :** We can see that Students and Business man capable of Paying back the loans without any due. While those are Maternity leave are facing difficulties in paying back their loans(42%)
- **Housing Type :** Clients living with parents about 11% having difficulty in paying loans and those are living in apartments with least about 6 %
- **Family Status :** About 10 % of Single/Not Married Clients facing difficulties in paying back loans and Widows face least difficulty in paying the loan (5%)
- **Age :** Most of the Clients, About 12 %,whose age in between 20-30 are considered as defaulters as they don’t pay loan in time. While majority of older people paid loan in time.
- **Own Realty :** There is no significance difference in Clients which own property or not
- **Contract Type :** Clients about(9%) who took cash loans, having difficulties in paying loans when compared to Clients who took Revolving Loans

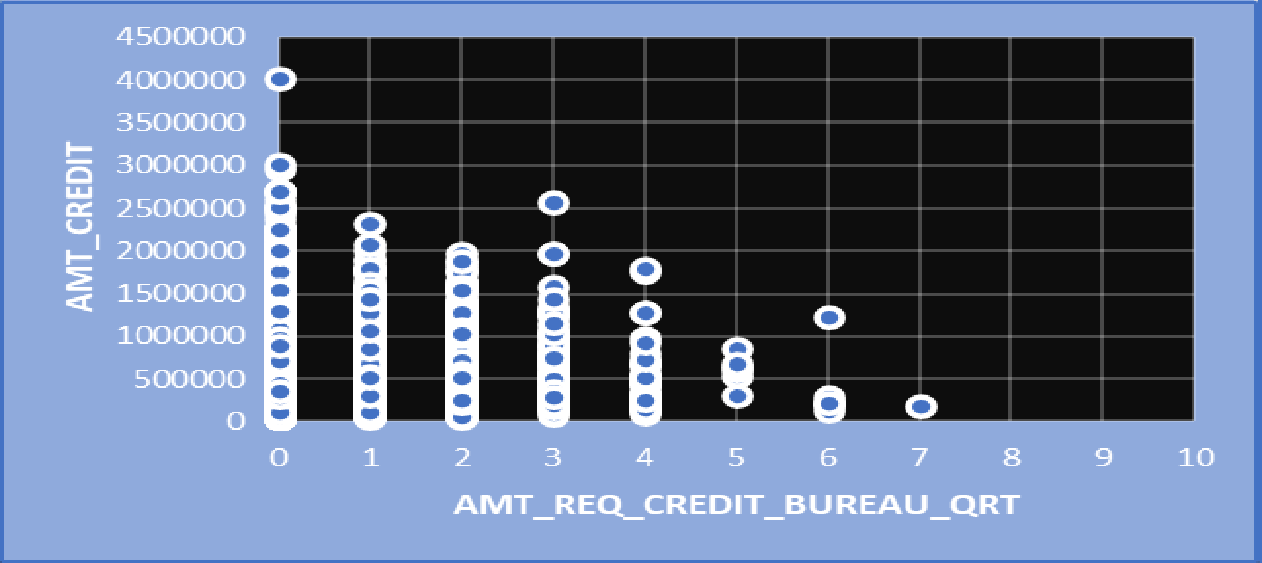
# BI- VARIATE ANALYSIS ON CATEGORY VARIABLE

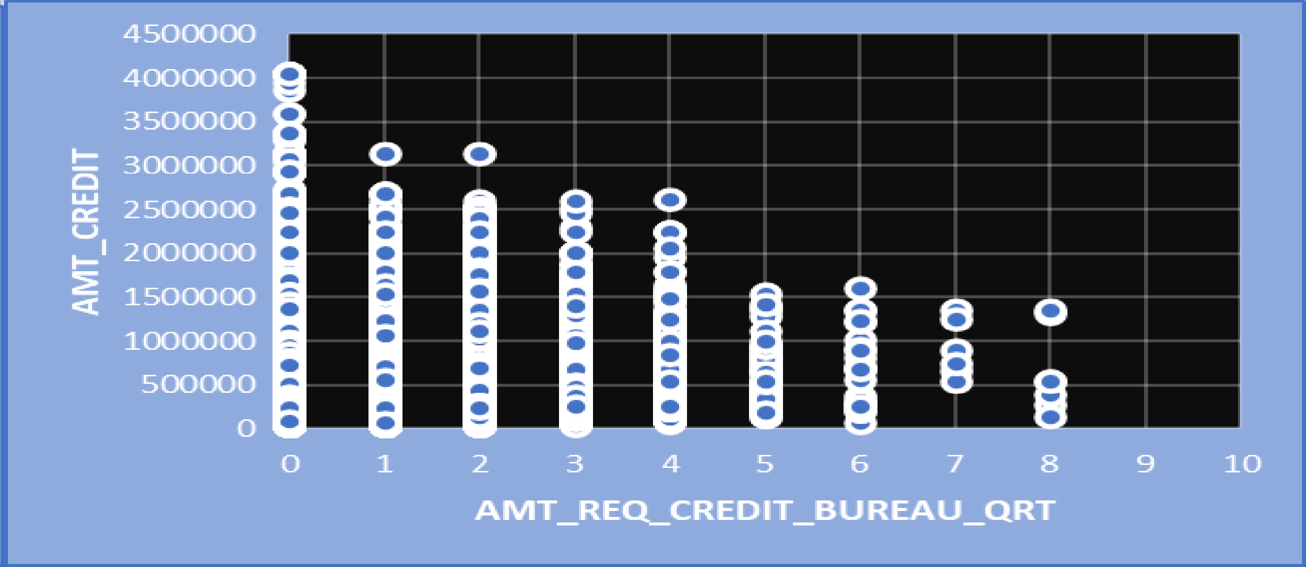


# BI- VARIATE ANALYSIS ON CONTINUOUS VARIABLE









# INSIGHTS

- **Amount Credit and Contract Type** : We can find more outliers in Credit amount for the Cash loans when compared to the Revolving loans. Both Defaulters and Non-Defaulters follows same pattern. The Box and Whisker plot shows the relation between them.
- **Amount Credit and Age Group:** Both Defaulters and Non-Defaulters have credit below 4.5 Million. We can see there are more outliers as the credit amount increase especially in younger and middle aged people
- **Amount Credit and Amount Income :** As the Income increases Credit also increases much in Non-Defaulters when compared to Defaulters. We can observe this from the Scatter Plot. Defaulters having more Outliers when compared to others.
- **Amount credit and credit Number of enquiries in Quarter :**We are having less number of enquiries if the credit amount is more. This pattern is strong in Defaulters case when compared to Non-Defaulters .

# CORRELATION ANALYSIS

- Correlation analysis is Statistical method used to measure strength and direction of the relationship between two variables.
- Correlation Coefficient is the measure of linear relationship between two variables. The value of correlation coefficient ranges from -1 to +1
- Positive value indicates positive relationship and Negative Value indicates negative relationship. Zero value indicates there is no relationship between those variables
- To get correlation matrix ,First we need to select the variables data of which we want to get the correlation matrix then go to “Data analysis” feature in Excel and then Select “Correlation” to get Correlation Matrix
- The cell value In the Correlation matrix is Correlation coefficient between corresponding row variable and column variable.
- The Diagonal values always equal to 1 in Correlation Matrix as they represent correlation of a variable itself.

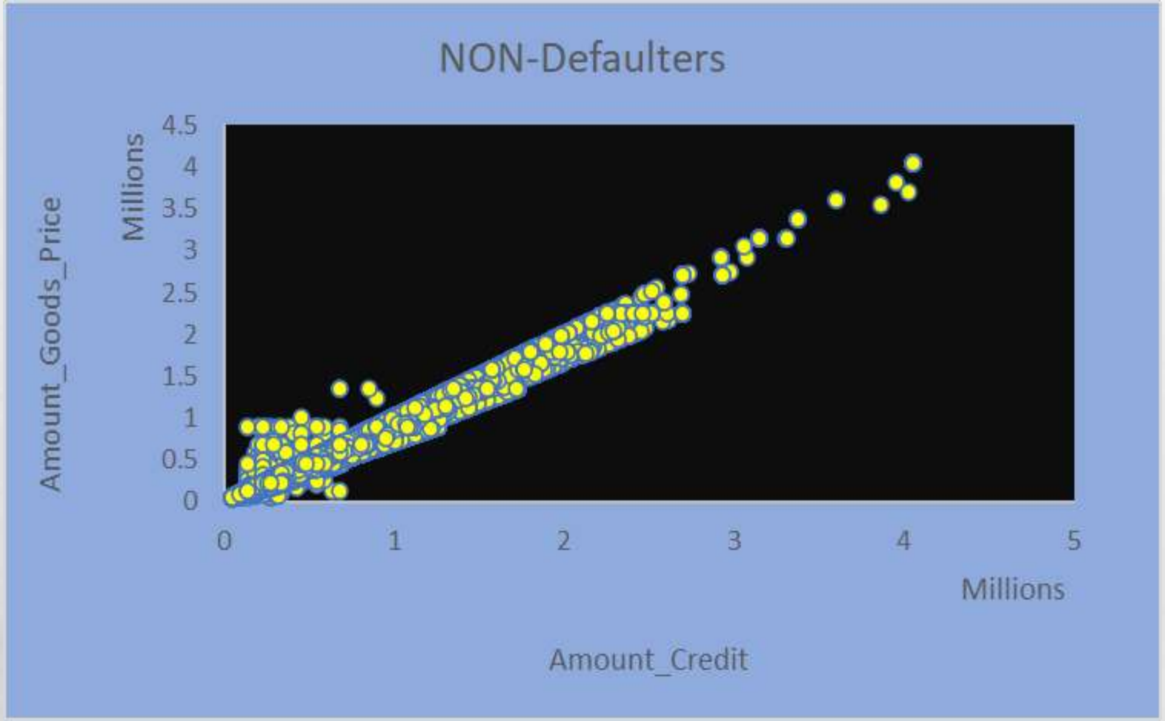
# CORRELATION MATRIX FOR NON-DEFAULTER

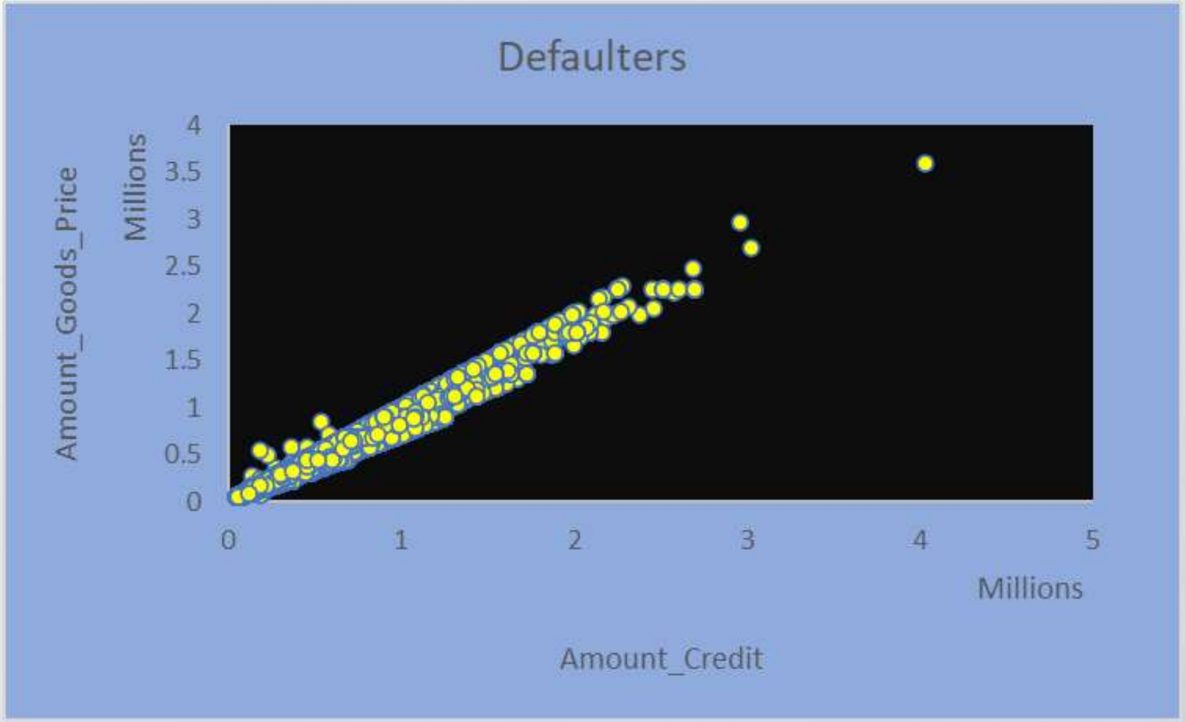
| | CNT_CHILDREN | AMT_INCOME_TOTAL | AMT_CREDIT | AMT_ANNUITY | AMT_GOODS_PRICE | REGION_POPULATION_RELATIVE | Age | Years Employed | CNT_FAM_MEMBERS | REGION_RATING_CLIENT | EXT_SOURCE_2 | EXT_SOURCE_3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CNT_CHILDREN | 1 | 0.028 | 0.003 | 0.021 | -0.001 | -0.024 | -0.337 | -0.041 | 0.879 | 0.023 | -0.015 | -0.037 |
| AMT_INCOME_TOTAL | 0.028 | 1 | 0.343 | 0.419 | 0.349 | 0.168 | -0.063 | 0.035 | 0.034 | -0.187 | 0.140 | -0.059 |
| AMT_CREDIT | 0.003 | 0.343 | 1 | 0.771 | 0.987 | 0.100 | 0.047 | 0.084 | 0.065 | -0.103 | 0.129 | 0.033 |
| AMT_ANNUITY | 0.021 | 0.419 | 0.771 | 1 | 0.777 | 0.121 | -0.013 | 0.053 | 0.076 | -0.132 | 0.127 | 0.025 |
| AMT_GOODS_PRICE | -0.001 | 0.349 | 0.987 | 0.777 | 1 | 0.104 | 0.045 | 0.085 | 0.063 | -0.104 | 0.136 | 0.036 |
| REGION_POPULATION_RELATIVE | -0.024 | 0.168 | 0.100 | 0.121 | 0.104 | 1 | 0.025 | -0.009 | -0.023 | -0.539 | 0.198 | -0.011 |
| Age | -0.337 | -0.063 | 0.047 | -0.013 | 0.045 | 0.025 | 1 | 0.226 | -0.286 | -0.002 | 0.078 | 0.175 |
| Years Employed | -0.041 | 0.035 | 0.084 | 0.053 | 0.085 | -0.009 | 0.226 | 1 | -0.010 | 0.016 | 0.074 | 0.097 |
| CNT_FAM_MEMBERS | 0.879 | 0.034 | 0.065 | 0.076 | 0.063 | -0.023 | -0.286 | -0.010 | 1 | 0.028 | -0.001 | -0.023 |
| REGION_RATING_CLIENT | 0.023 | -0.187 | -0.103 | -0.132 | -0.104 | -0.539 | -0.002 | 0.016 | 0.028 | 1 | -0.291 | -0.004 |
| EXT_SOURCE_2 | -0.015 | 0.140 | 0.129 | 0.127 | 0.136 | 0.198 | 0.078 | 0.074 | -0.001 | -0.291 | 1 | 0.076 |
| EXT_SOURCE_3 | -0.037 | -0.059 | 0.033 | 0.025 | 0.036 | -0.011 | 0.175 | 0.097 | -0.023 | -0.004 | 0.076 | 1 |

# FOR DEFAULTERS

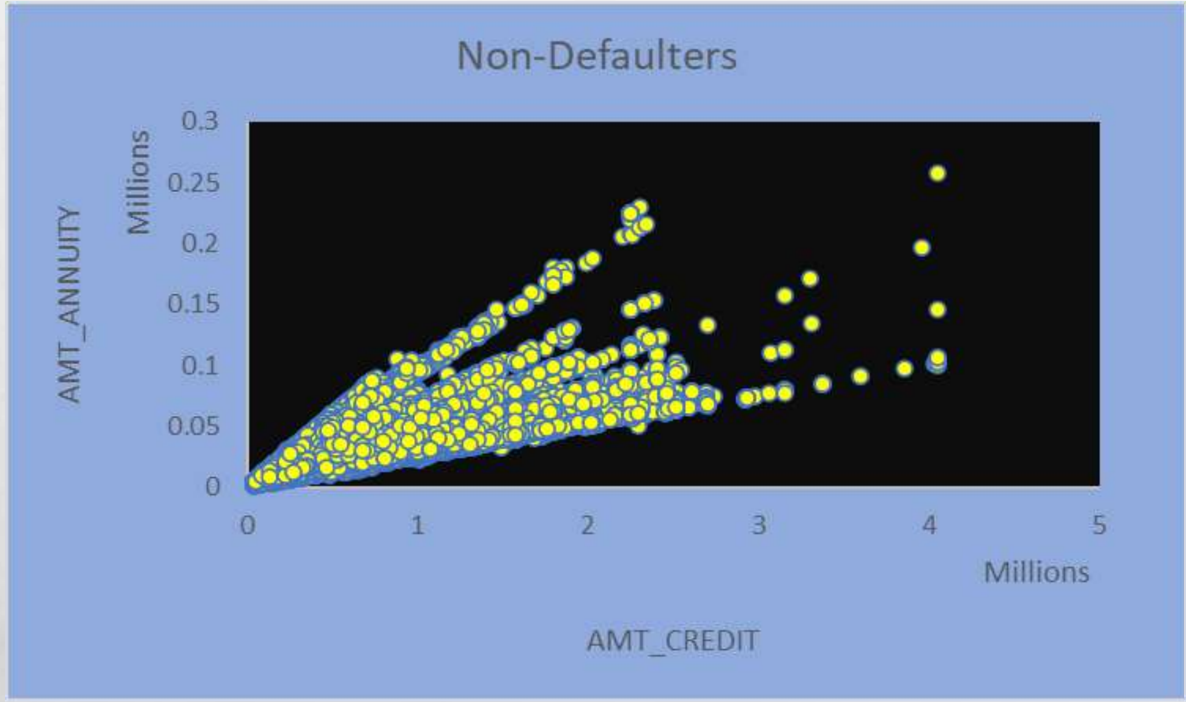
| | CNT_CHILDREN | AMT_INCOME_TOTAL | AMT_CREDIT | AMT_ANNUITY | AMT_GOODS_PRICE | REGION_POPULATION_RELATIVE | Age | Years Employed | CNT_FAM_MEMBERS | REGION_RATING_CLIENT | EXT_SOURCE_2 | EXT_SOURCE_3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CNT_CHILDREN | 1 | 0.005 | -0.002 | 0.031 | -0.008 | -0.032 | -0.259 | -0.044 | 0.886 | 0.041 | -0.012 | -0.024 |
| AMT_INCOME_TOTAL | 0.005 | 1 | 0.038 | 0.046 | 0.038 | 0.009 | -0.003 | -0.001 | 0.007 | -0.021 | 0.007 | -0.018 |
| AMT_CREDIT | -0.002 | 0.0381 | 1 | 0.752 | 0.983 | 0.069 | 0.135 | 0.098 | 0.051 | -0.059 | 0.121 | 0.052 |
| AMT_ANNUITY | 0.031 | 0.0464 | 0.752 | 1 | 0.753 | 0.072 | 0.014 | 0.041 | 0.076 | -0.074 | 0.116 | 0.032 |
| AMT_GOODS_PRICE | -0.008 | 0.0376 | 0.983 | 0.753 | 1 | 0.076 | 0.136 | 0.104 | 0.047 | -0.066 | 0.131 | 0.053 |
| REGION_POPULATION_RELATIVE | -0.032 | 0.0091 | 0.069 | 0.072 | 0.076 | 1 | 0.048 | 0.016 | -0.030 | -0.443 | 0.170 | -0.010 |
| Age | -0.259 | -0.0031 | 0.135 | 0.014 | 0.136 | 0.048 | 1 | 0.281 | -0.203 | -0.034 | 0.108 | 0.134 |
| Years Employed | -0.044 | -0.0010 | 0.098 | 0.041 | 0.104 | 0.016 | 0.281 | 1 | -0.010 | -0.005 | 0.089 | 0.056 |
| CNT_FAM_MEMBERS | 0.886 | 0.0067 | 0.051 | 0.076 | 0.047 | -0.030 | -0.203 | -0.010 | 1 | 0.044 | 0.002 | -0.029 |
| REGION_RATING_CLIENT | 0.041 | -0.0215 | -0.059 | -0.074 | -0.066 | -0.443 | -0.034 | -0.005 | 0.044 | 1 | -0.250 | 0.014 |
| EXT_SOURCE_2 | -0.012 | 0.0071 | 0.121 | 0.116 | 0.131 | 0.170 | 0.108 | 0.089 | 0.002 | -0.250 | 1 | 0.049 |
| EXT_SOURCE_3 | -0.024 | -0.0182 | 0.052 | 0.032 | 0.053 | -0.010 | 0.134 | 0.056 | -0.029 | 0.014 | 0.049 | 1 |

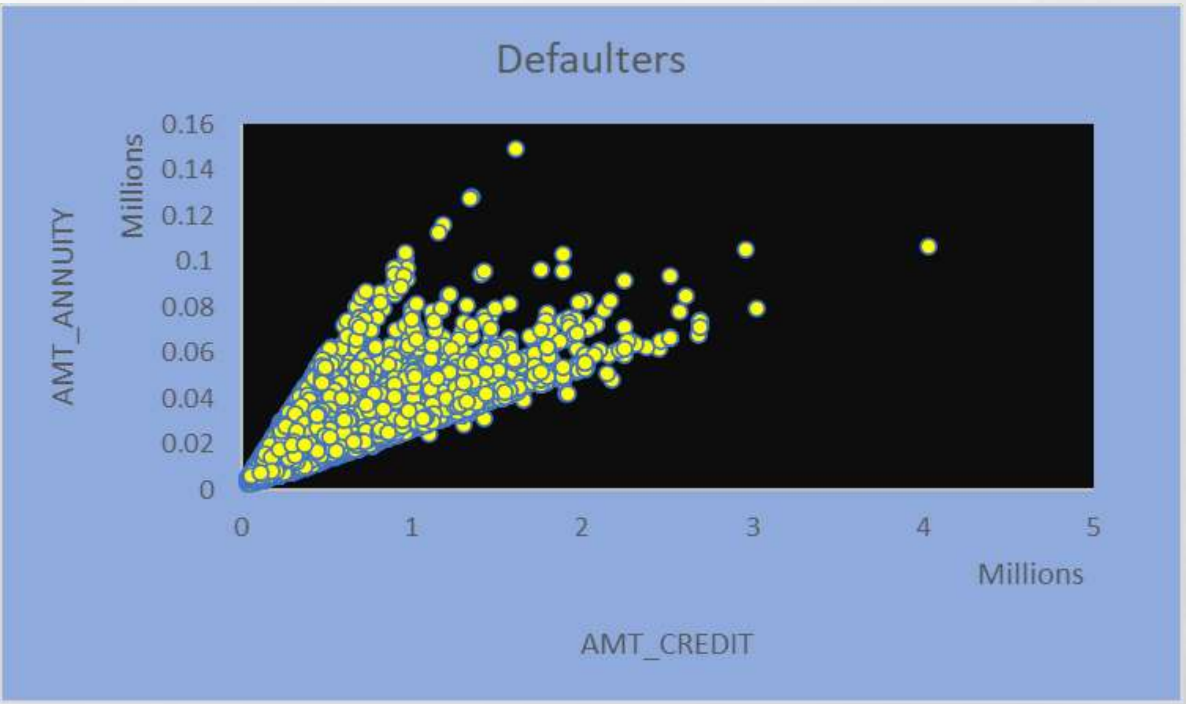
- **Amount Credit and Amount Goods Price**: The Correlation coefficient between these two variables is 0.9 which is very close to 1.
- I have plotted Scatter Plot to show Visually. If the amount of credit increases then Amount of goods price also increases for both Defaulters and Non-Defaulters.
- There is a strong linear relationship between these two variables.



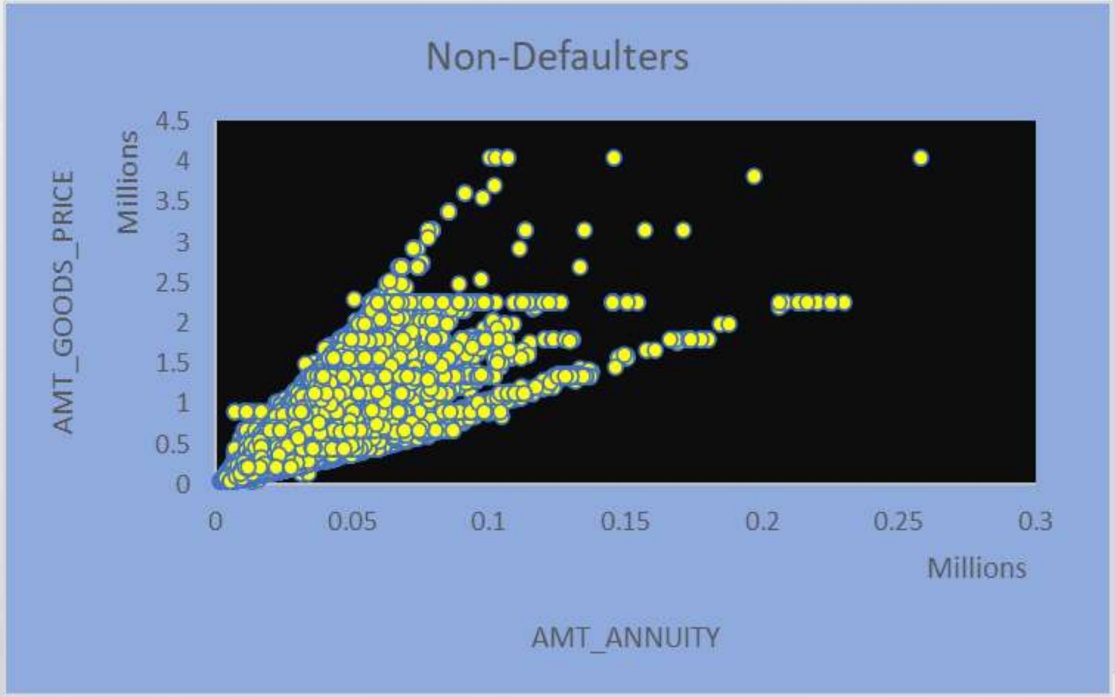


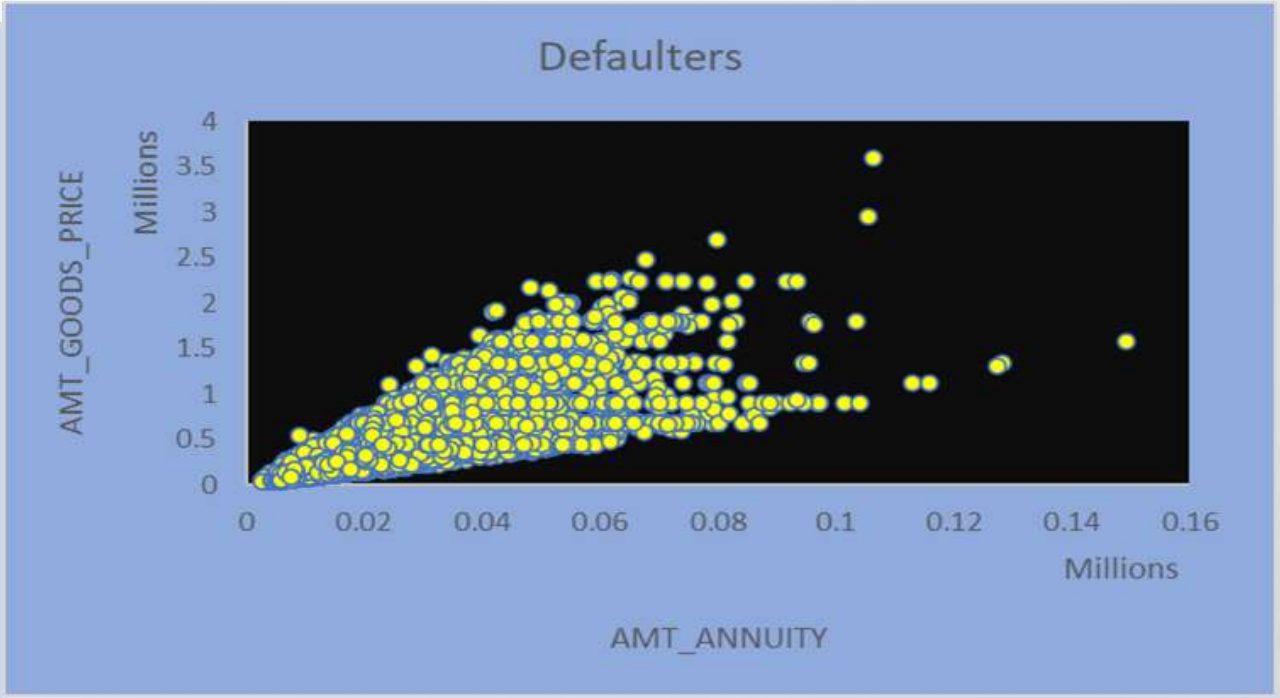
- **Amount Annuity and Amount Credit** :The Correlation coefficient of these two variables is around 0.7 which is somewhat strong linear relationship between those variables.
- If the credit amount increases amount annuity also increases in both Defaulters and Non-Defaulters case.
- The Scatter plot visualizes the data efficiently.



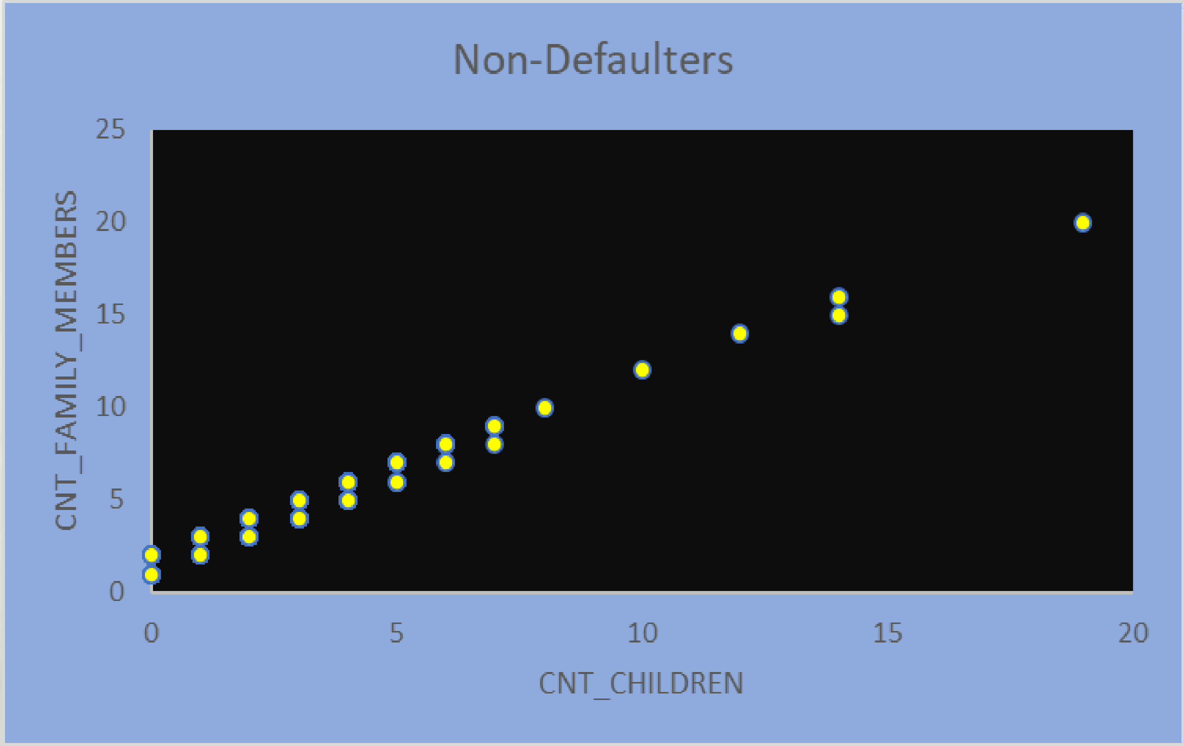


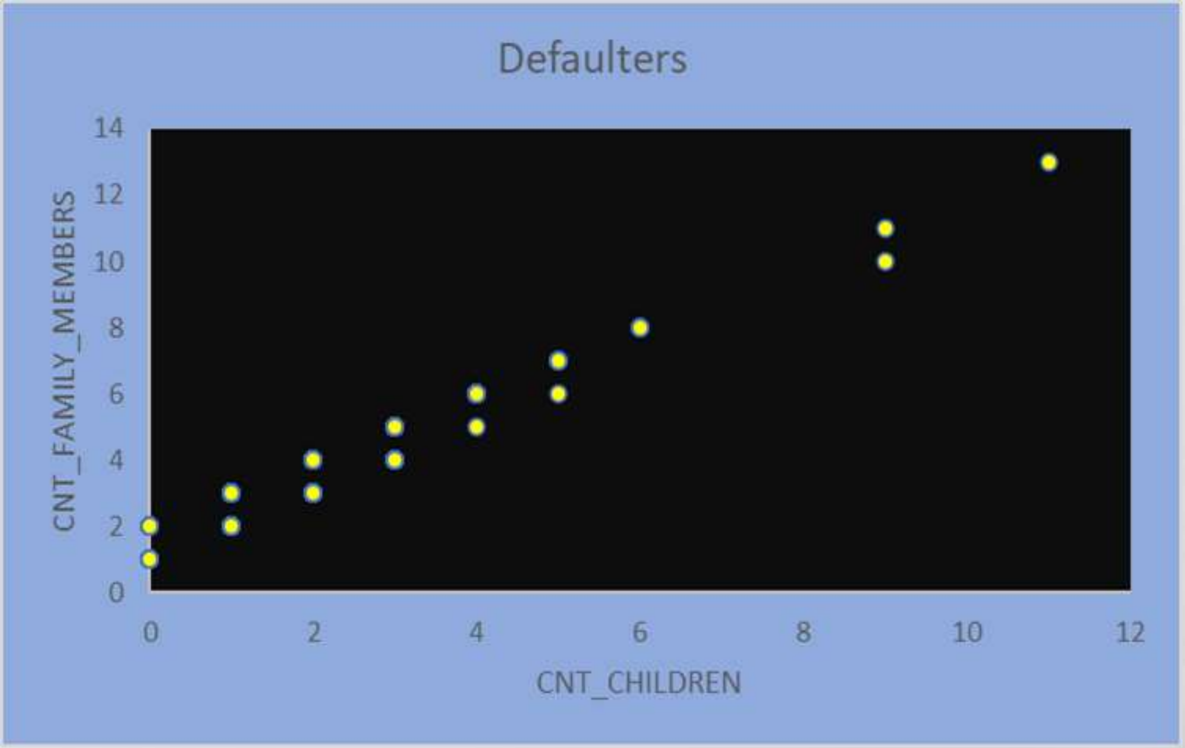
- **Amount Annuity and Goods Price:** The Correlation coefficient between Amount Annuity and Goods price is very close to .75 which indicates the Strong linear relationship between both of them.
- More the amount of annuity more will be the Amount of Goods price





- **Family members and Children:** The Correlation coefficient between Family members and Children is too close to 0.9 which indicates Strong relationship between these two variable.
- It’s obvious that if the number of children in a family is more then the total number of family members also more.
- I plotted a Scatter plot to visualize the data in effective way.





# PREVIOUS APPLICATION DATA

# DATA CLEANING

The data set Previous_application contains nearly 1.6 Million records.
We know that Excel is a powerful tool for data cleaning and analysis but it's ability handle large datasets is limited.
So, I cleaned this dataset using Python and analysed data in Excel.
Steps followed to clean data in Python:

1.Import "Numpy" and "Pandas" libraries for data analysis and manipulation

2.Create dataframe and read the csv file using "read_csv" function

3.Find out the Blanks percentage of each column in the dataframe  by using

```
blanks_percentage=df.isnull().sum()/(len(df))*100
```

4.Finding the columns whose blanks percentage greater than 30 and store them in a variable drop_columns

```
drop_columns=list(blanks_percentage[blanks_percentage>30].index)
```

5.Dropping the Columns above (blanks_percentage >30)

```
df=df.drop(columns=drop_columns)
```

6.Replacing the unknown values "XNA" and "XAP" with the null values

```
df.replace('XNA',np.nan,inplace=True)
df.replace('XAP',np.nan,inplace=True)
```

7-The data set Previous_application contains nearly 1.6 Million records.
We know that Excel is a powerful tool for data cleaning and analysis but it’s ability handle large datasets is limited.
So I cleaned this dataset using Python and analysed data in Excel.
Steps followed to clean data in Python:

1.Import “Numpy” and “Pandas” libraries for data analysis and manipulation

2.Create dataframe and read the csv file using “read_csv” function

3.Find out the Blanks percentage of each column in the dataframe  by using

```
blanks_percentage=df.isnull().sum()/(len(df))*100
```

4.Finding the columns whose blanks percentage greater than 30 and store them in a variable drop_columns

```
drop_columns=list(blanks_percentage[blanks_percentage>30].index)
```

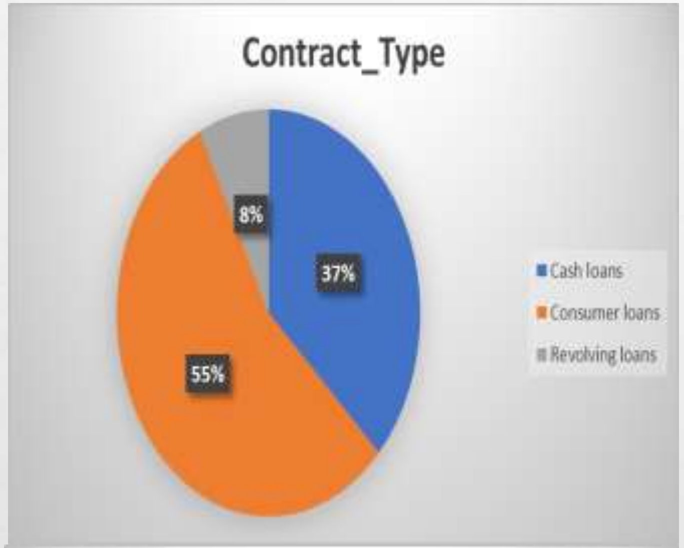
5.Dropping the Columns above (blanks_percentage >30)

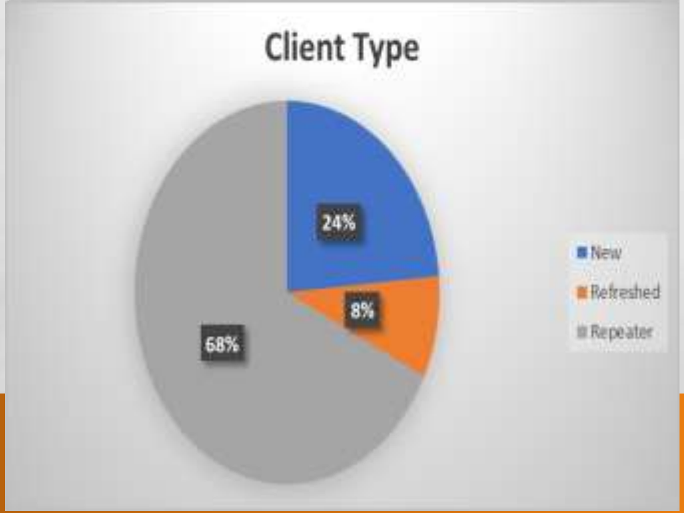
```
df=df.drop(columns=drop_columns)
```

6.Replacing the unknown values “XNA” and “XAP” with the null values

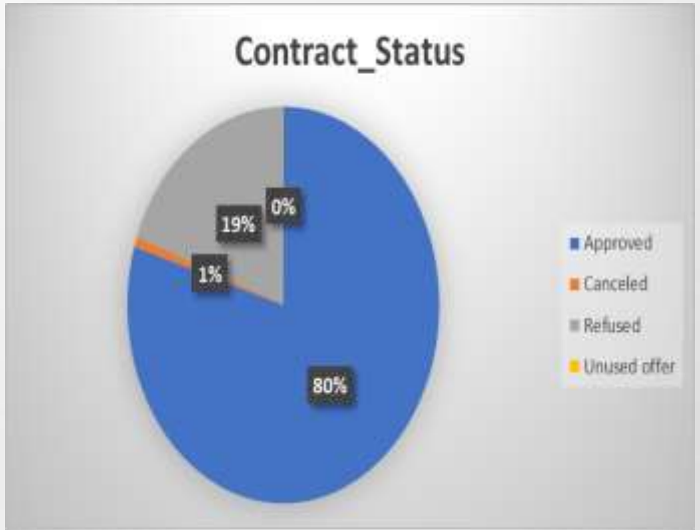
```
df.replace('XNA',np.nan,inplace=True)
df.replace('XAP',np.nan,inplace=True)
```

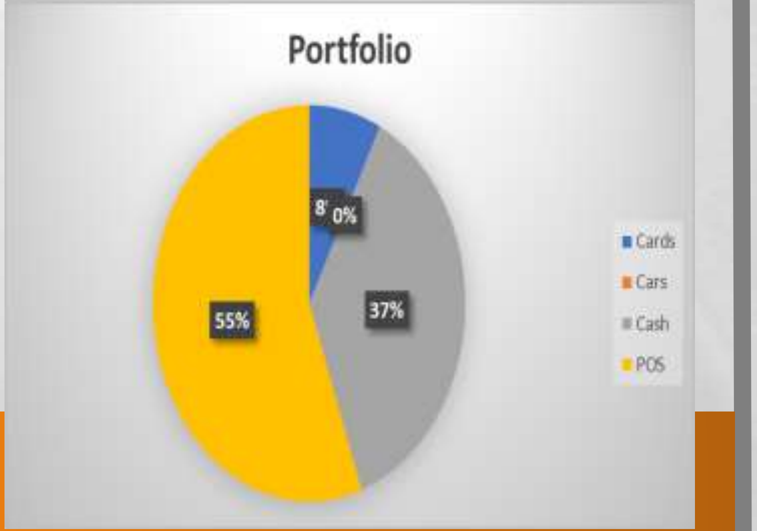
# DATA IMBALANCE



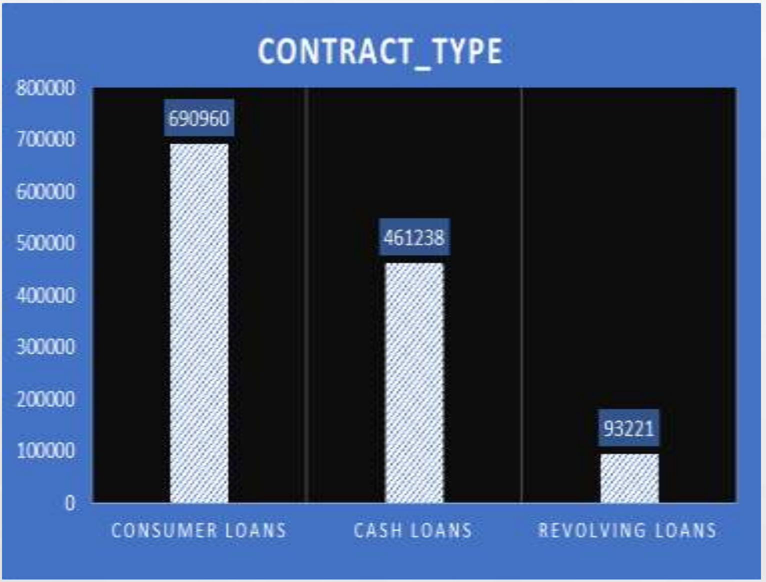


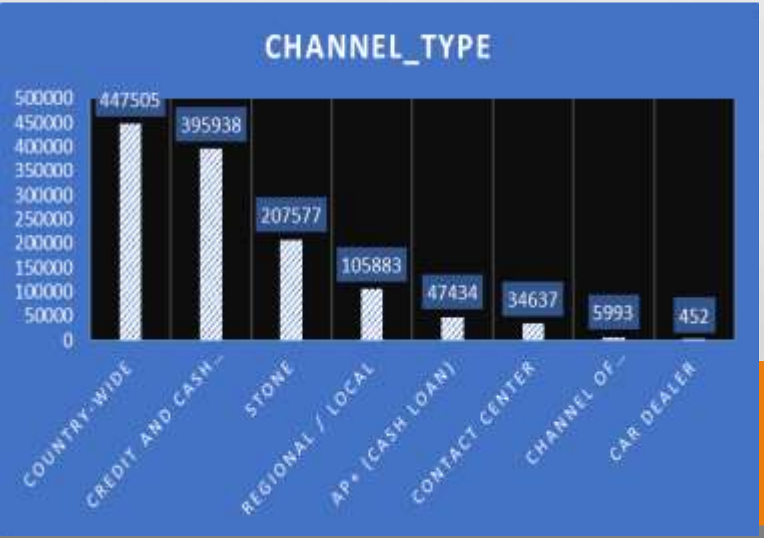
- In Contract_Type, 55% loans are Consumer loans and Revolving loans are 8% only
- In Contract_Status, Approved loans are 80 % and Refused loans are 19%
- In Client_Type, 68% of the clients are “repeaters” only and “New” clients are 24 % only
- In Portfolio, 55 % are of “Cards” and 37 % are of “Cards”

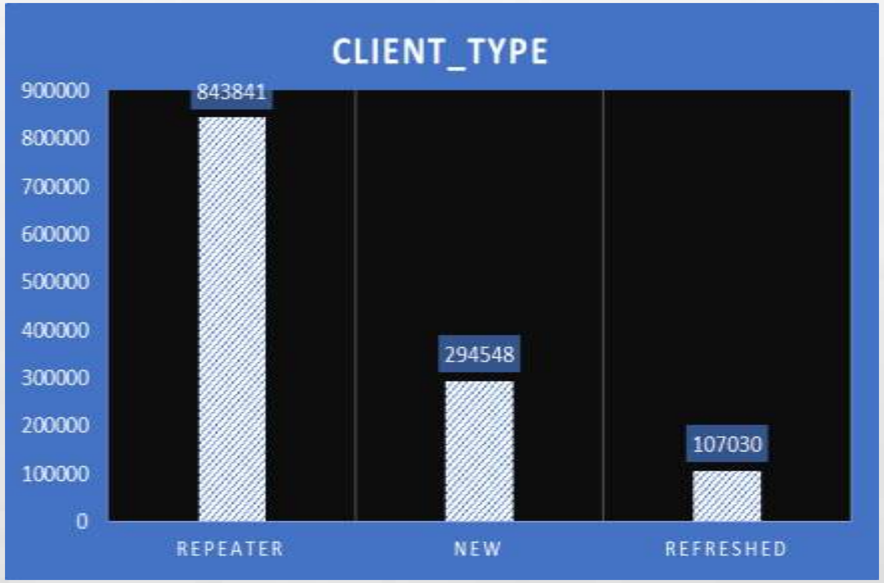


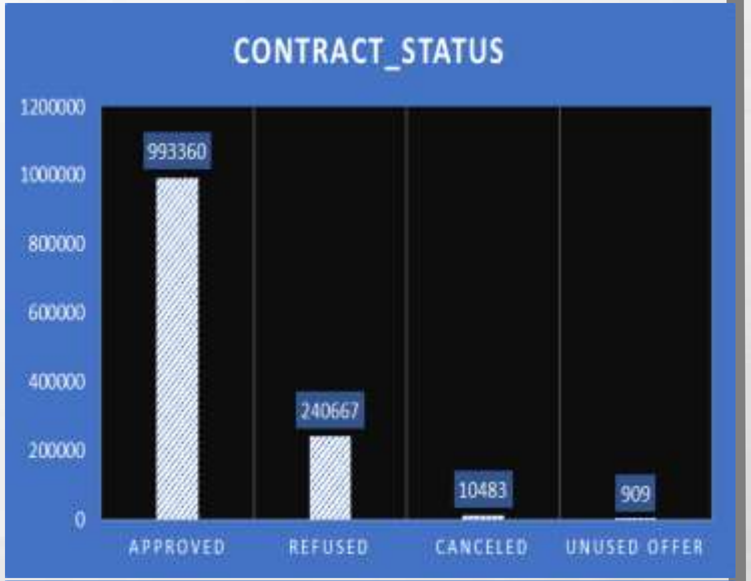


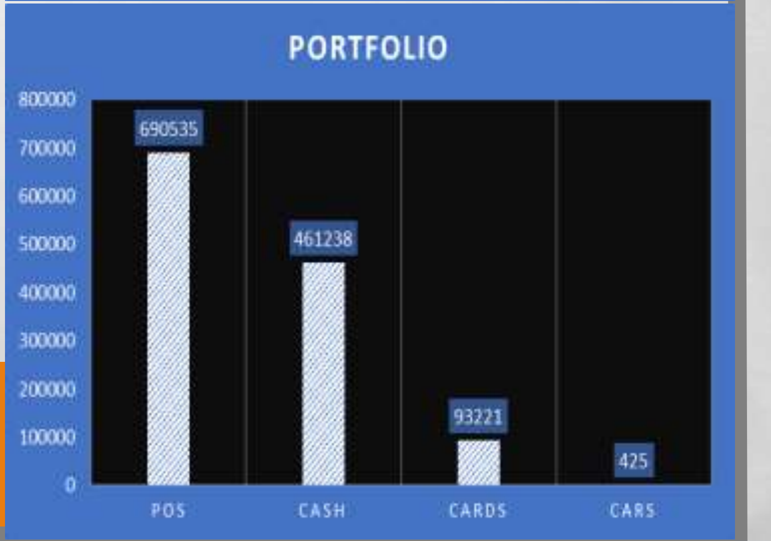
# UNIVARIATE ANALYSIS







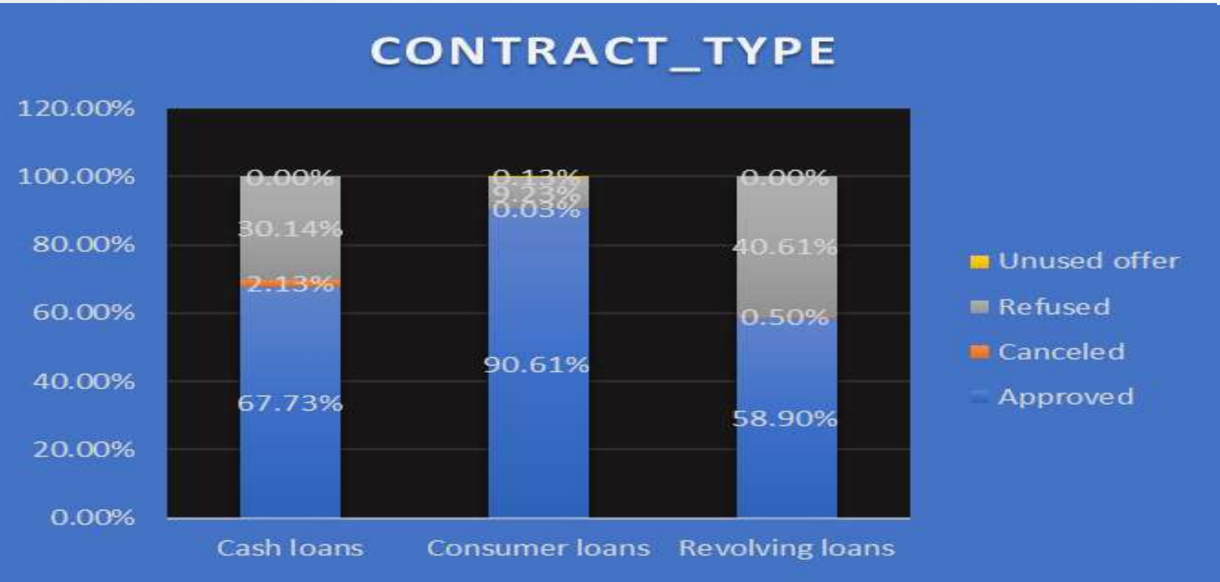


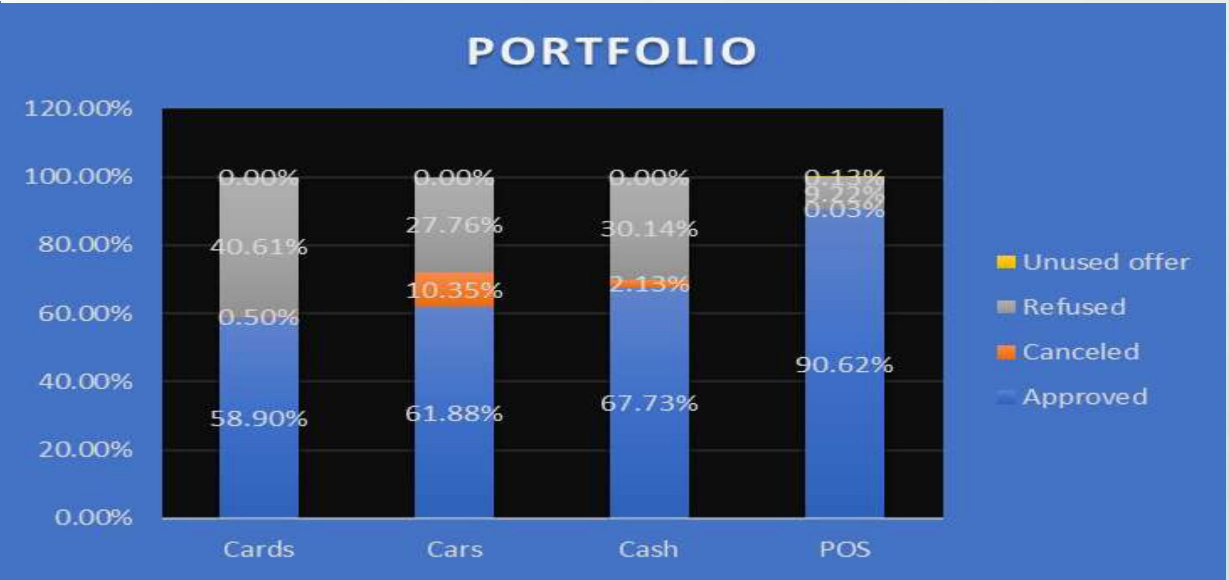


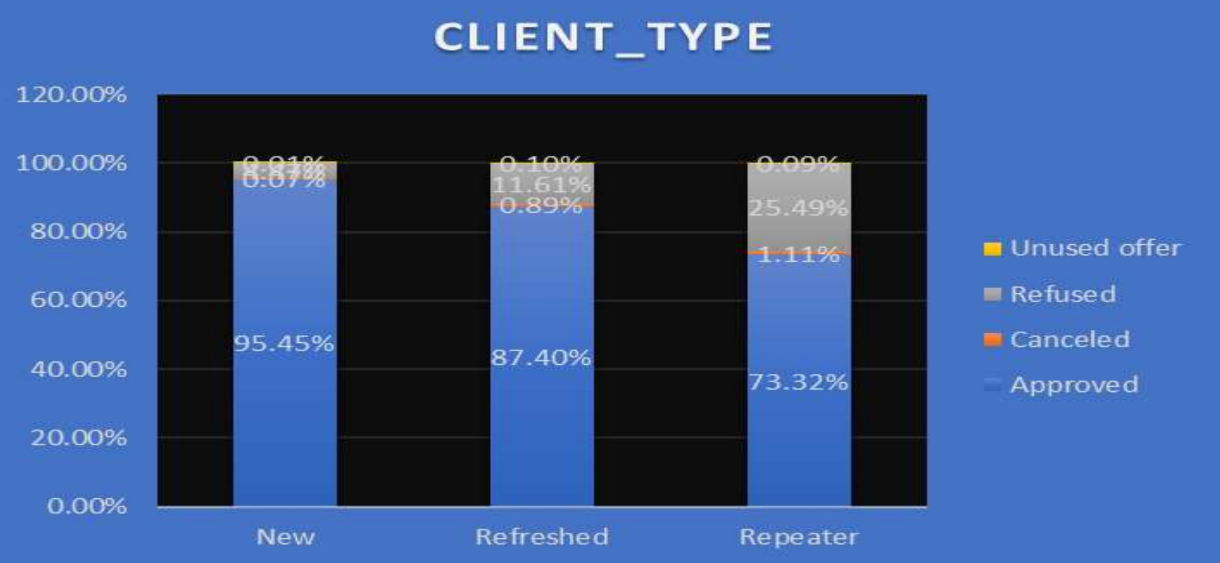
# INSIGHTS

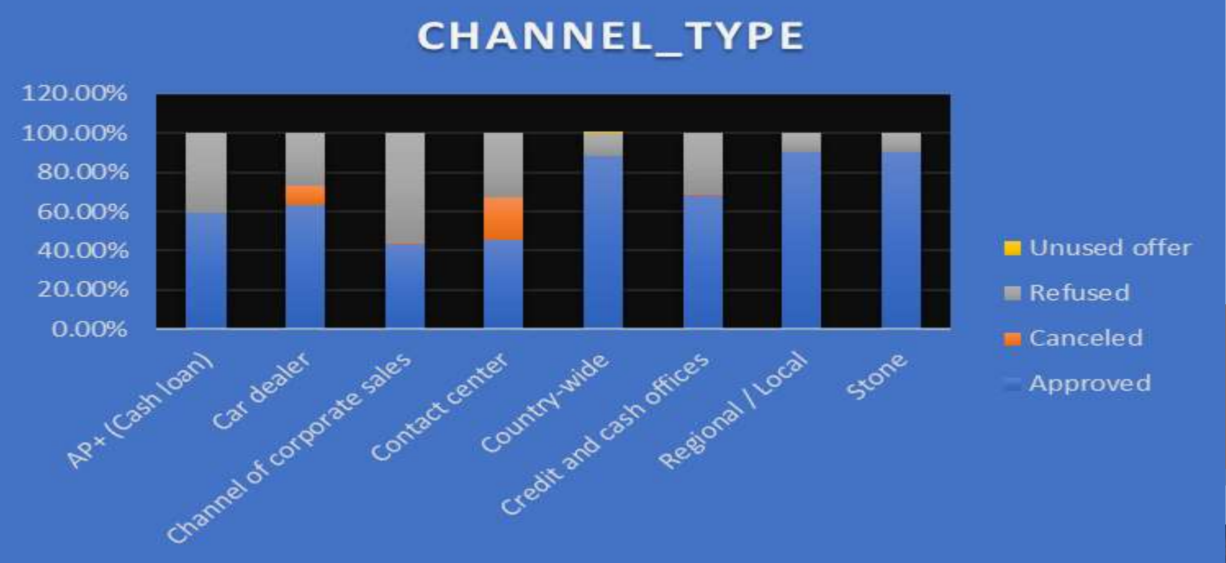
- **Contract Type:** loans are most taken loans by clients and cash loans also good number .Revolving loans are the least taken loans
- **Contract Status :** Most the status of the contract got Approved and only few of them rejected
- **Client Type :** We can observe that Most of the clients who kept applied are “Repeaters” only. New client applications are less percent when compared to repeaters
- **Channel Type:** Through “Country Wide” and “Credit and cash offices” are most acquired channel type for the clients on the previous application
- **Portfolio:** Many of the previous application portfolio was for “POS” and “Cash”

# UNIVARATED SEGMENTED ANALYSIS


CONTRACT_TYPE


PORTFOLIO


CLIENT_TYPE


CHANNEL_TYPE

# INSIGHTS

- **Contract Type :**In consumer loans, approval rate is more than 90% and rejected rate is close to 9% .In Cash loans 67% is the approval rate and  30% is the rejected rate

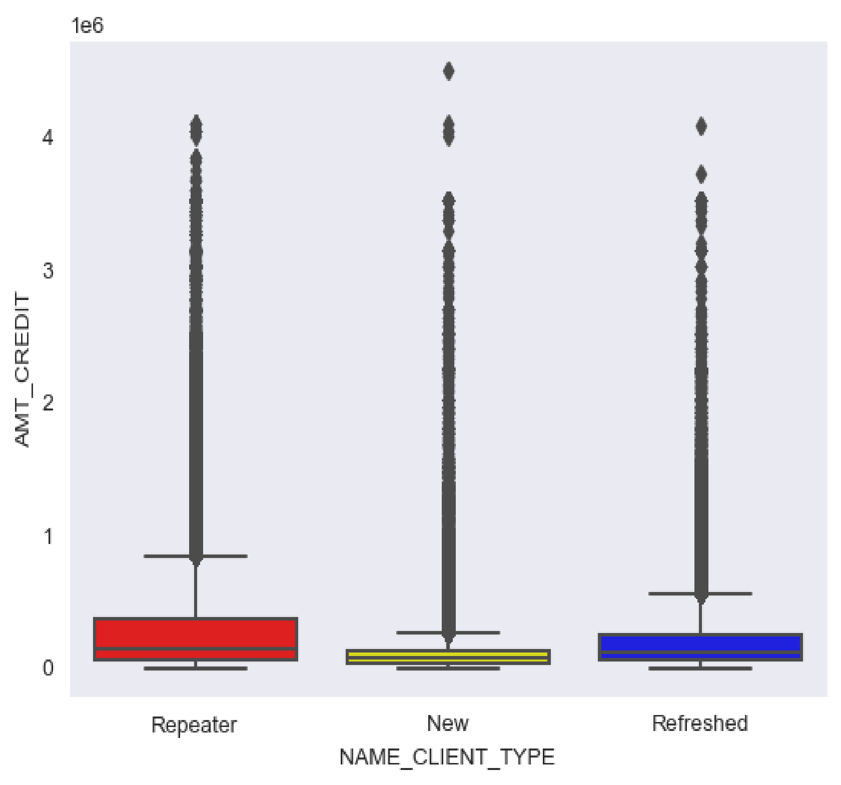
- **Portfolio :** In portfolio “POS” approval rate is 90 % and “Cash” approval rate is  68 %.The “Cards” are having the most rejected rate i.e., upto 40 %

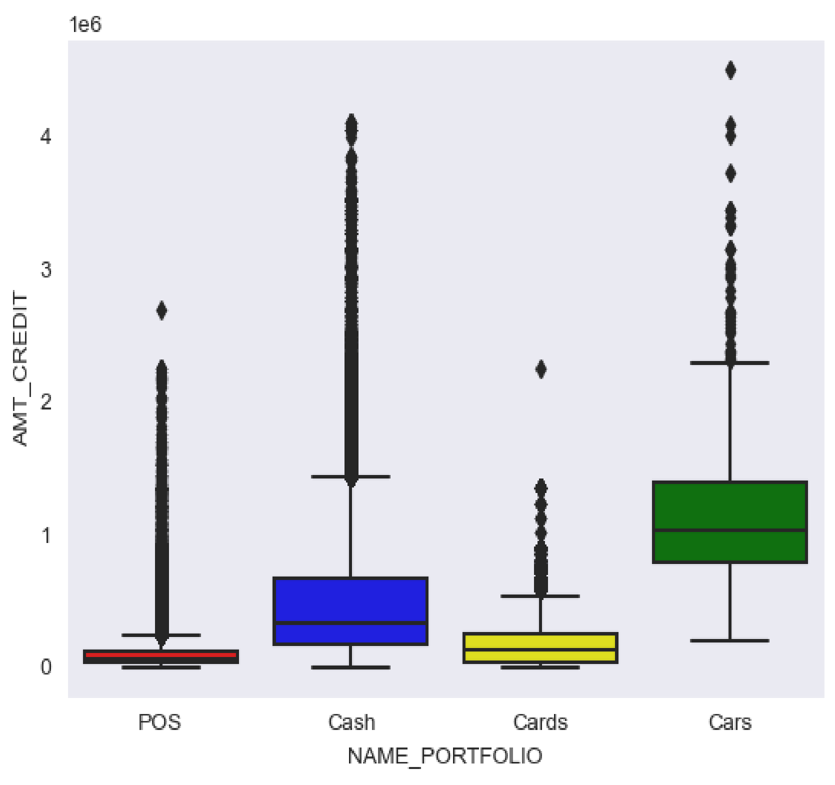
- **Client Type :**  “New” loans approval rate is close to 95 % and “Refreshed Loans” approval rate is close to 87 %.This shows that new loans can be easily approved

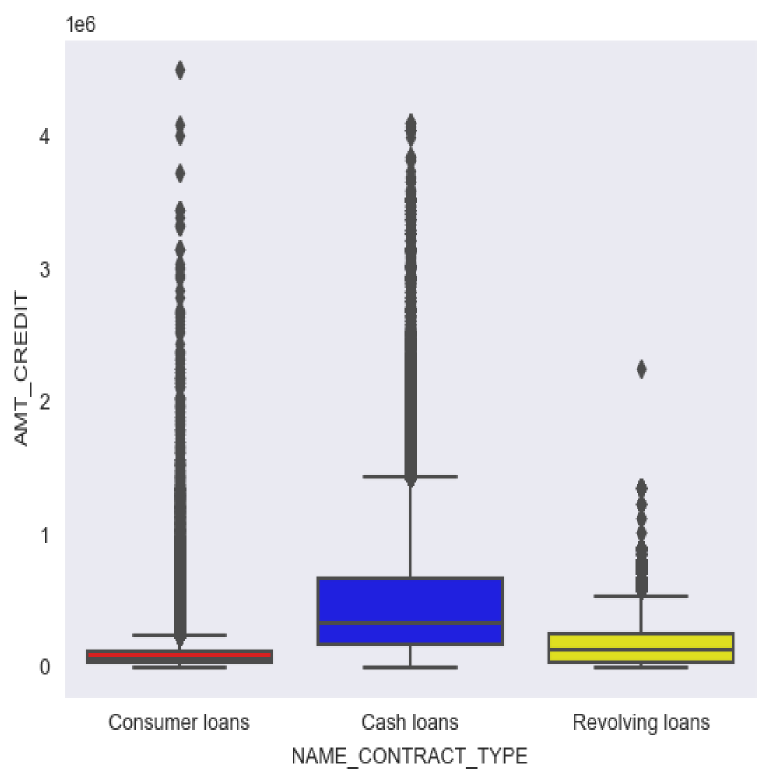
- **Channel Type :** Channels like “Regional”, “Stone” having the most approval rate when compared to other channels. Corporate sales is having the most rejections i.e., upto 55 %

# BI- VARIATE ANALYSIS

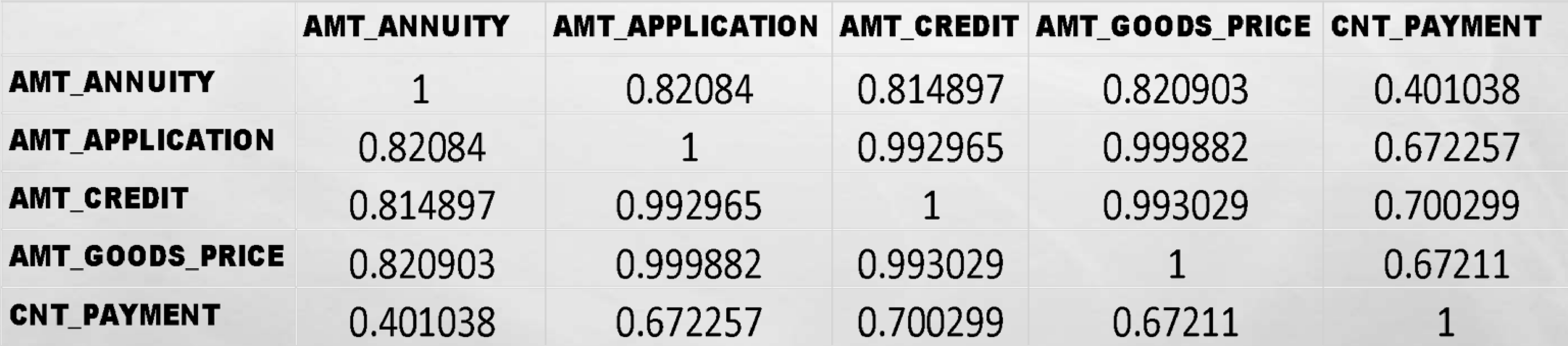
- ➢ I plotted the box and whisker between various variables for Bivariate analysis
- ➢ We can see that Credit amount for the “Repeater” is high for among various types of Clients
- ➢ Credit Amount for the “Cash Loans” is high when compared to other contract types like Consumer loans and Revolving Loans





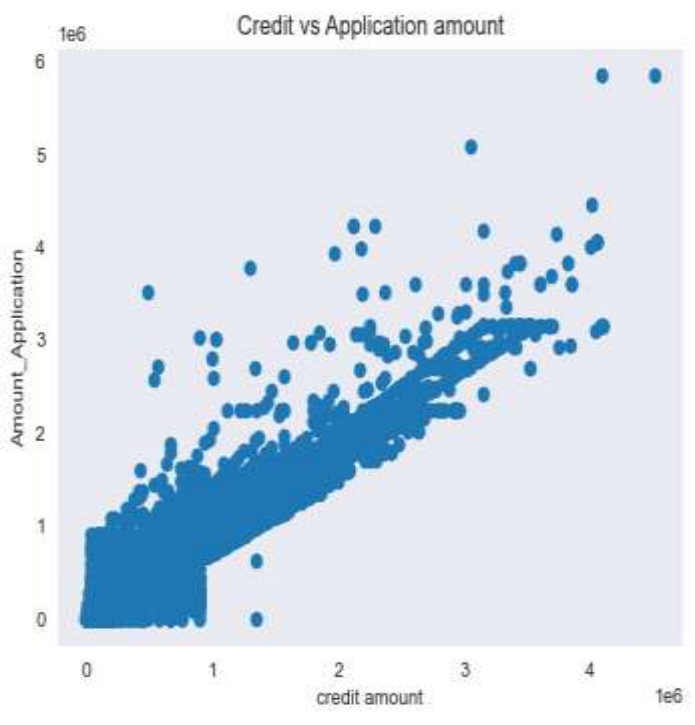


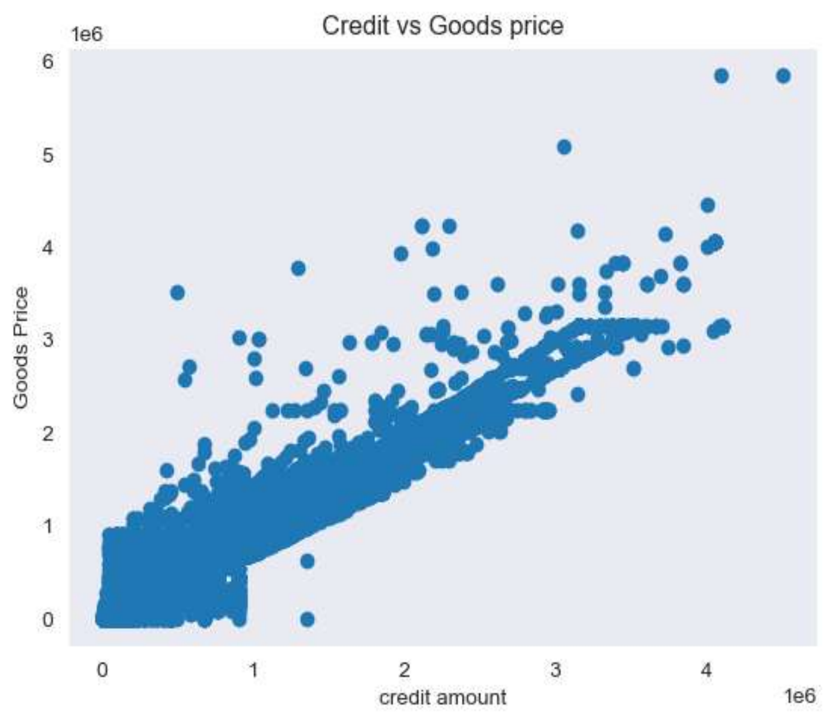
# CORRELATION MATRIX BETWEEN CONTINUOUS ANALYSIS

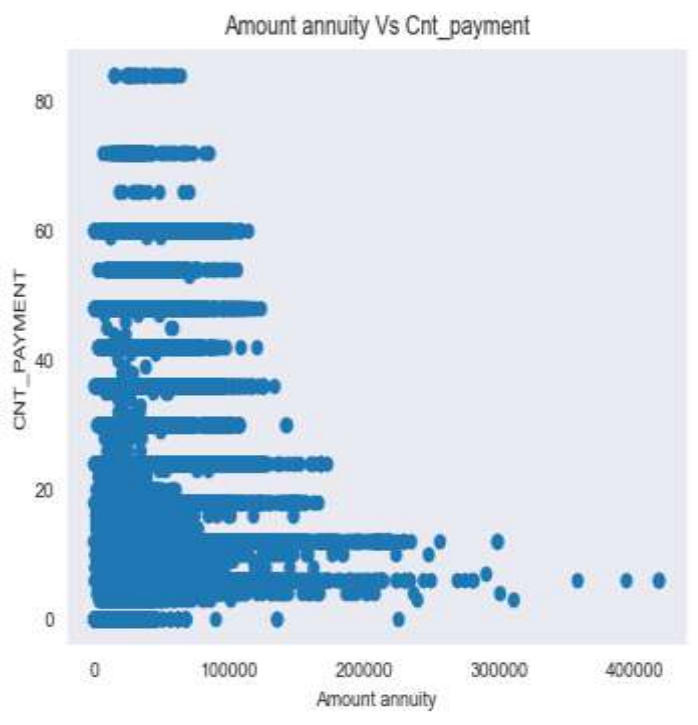


| | AMT_ANNUITY | AMT_APPLICATION | AMT_CREDIT | AMT_GOODS_PRICE | CNT_PAYMENT |
|---|---|---|---|---|---|
| AMT_ANNUITY | 1 | 0.82084 | 0.814897 | 0.820903 | 0.401038 |
| AMT_APPLICATION | 0.82084 | 1 | 0.992965 | 0.999882 | 0.672257 |
| AMT_CREDIT | 0.814897 | 0.992965 | 1 | 0.993029 | 0.700299 |
| AMT_GOODS_PRICE | 0.820903 | 0.999882 | 0.993029 | 1 | 0.67211 |
| CNT_PAYMENT | 0.401038 | 0.672257 | 0.700299 | 0.67211 | 1 |

- The correlation coefficient between Credit and Application amount is 0.9 suggesting that there is strong relationship between them
- Similarly Correlation Coefficient between Credit amount and Goods price is also cl**o**se to 1 suggesting strong linear relationship between them.
- But the Correlation Coefficient between the Amount annuity and Cnt payment is somewhat low i.e 0.4 That’s why we did not get linear Scatter Plot







# CONCLUSION

## APPLICATION_DATA

- Clients who hold Academic degree are highly capable of paying loans in time
- Female Clients are more capable than Male Clients in repaying the loans
- Clients who taken Revolving loans are facing less difficulty in repaying the loan when compared to clients who took other type of loans
- Old Aged People i.e who aged above 60 are more capable of repaying the loans
- Students and Businessmen are getting no difficulties in re paying the loans

## PREVIOUS_APPLICATION

- The approval rate for Client who applied for Consumer loans are high
- If the Credit amount is high then chances of getting approved is high
- New Clients are approved most of time i.e 90%
- Corporate sales is having the most rejections i.e upto 55 %

# RESULTS

From this BANK LOAN CASE STUDY , I learned how the data is being analyzed in Banking sector to give loans and asses the risks in them.I gained valuble insights from this Case Study by analyzing historical loan data and customer characteristics and I may be able to develop models to predict the likelihood of loan defaults.I also knew about the “Exploratory Data Analysis” and it’s importance in data analysis.

# THANK YOU