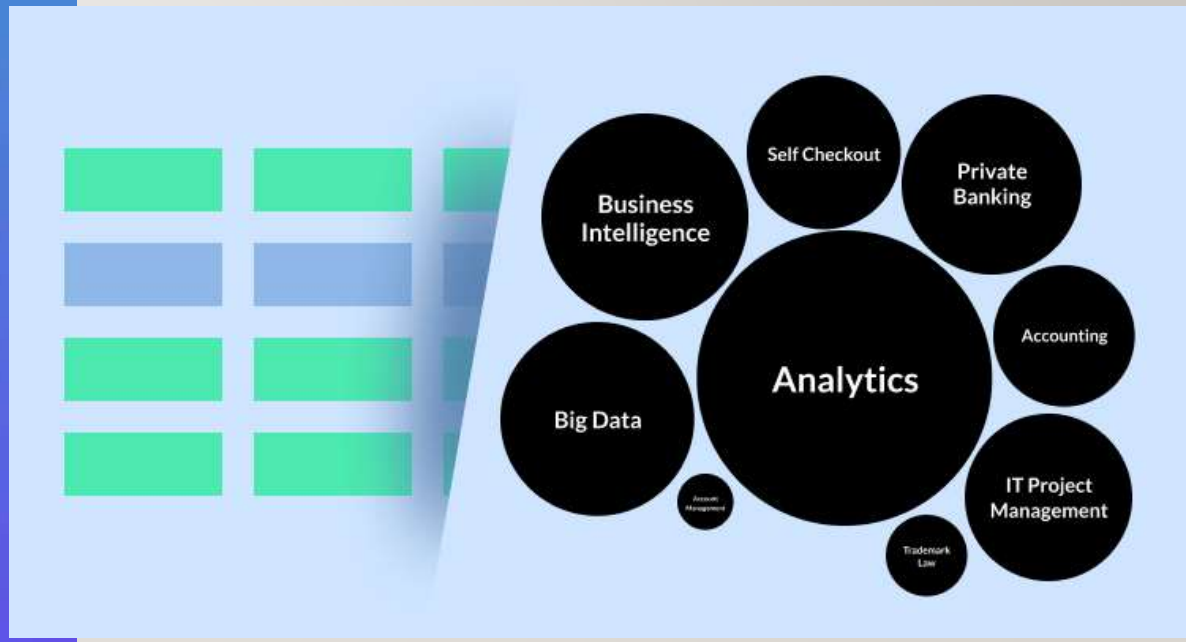
CREDIT EDA CASE STUDY

BY DEEPANSHU

INTRODUCTION

- This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

BANKING BUSINESS STRATEGY VISUALISATION



BUSINESS OBJECTIVE

- This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

DATA ANALYSIS OF EDA CASE STUDY INVOLVES FOLLOWING STEPS

- BUSINESS UNDERSTANDING
- DATA UNDERSTANDING
- DATA PREPARATION
- MODELLING
- EVALUATION
- DEPLOYMENT

DATA UNDERSTANDING

This dataset has 2 files as explained below:

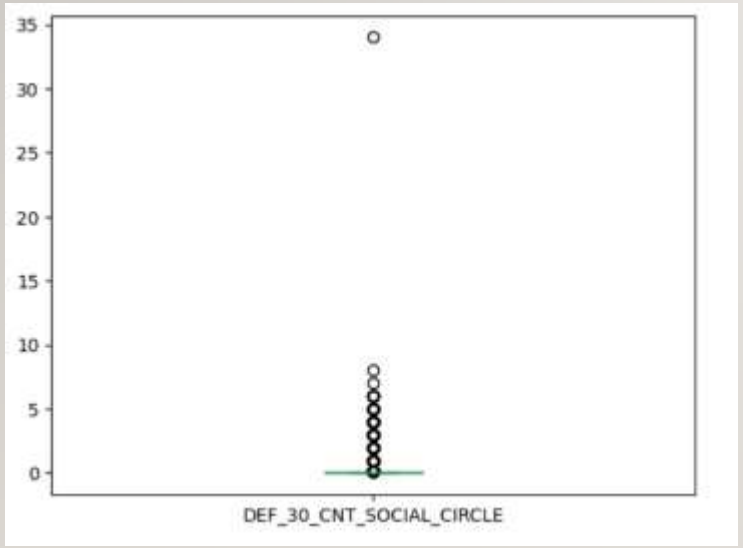
- 'application_data.csv' contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.
- 'previous_application.csv' contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.

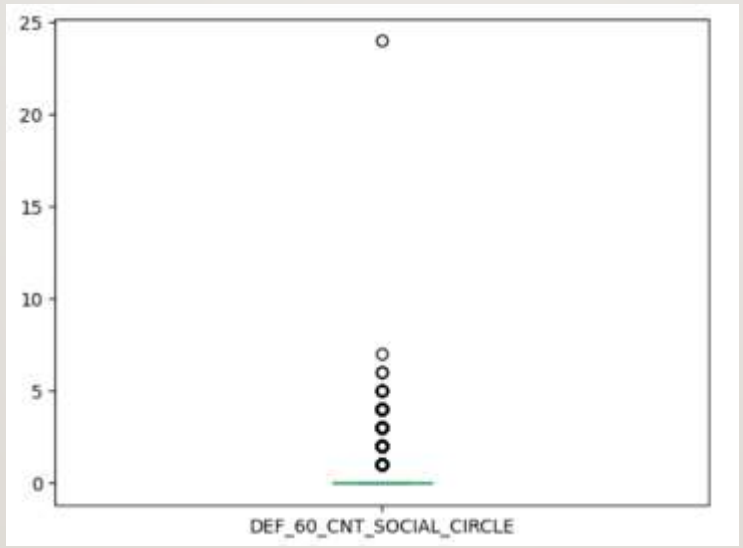
OUTLIERS OCCURRED DURING ANALYSIS

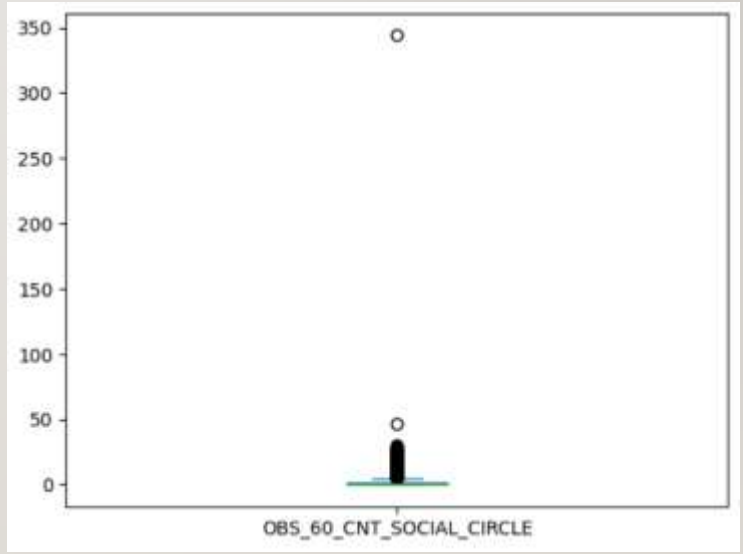
OUTLIER

An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. In a sense, this definition leaves it up to the analyst (or a consensus process) to decide what will be considered abnormal.

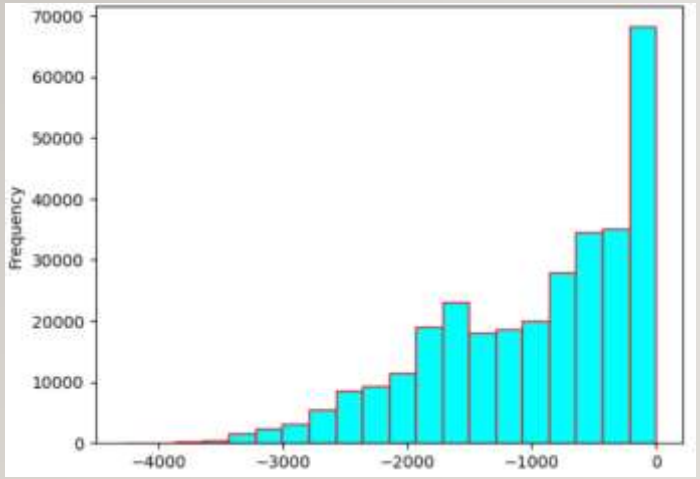
SOME EXAMPLES OF BOXPLOTS FOR OUTLIERS

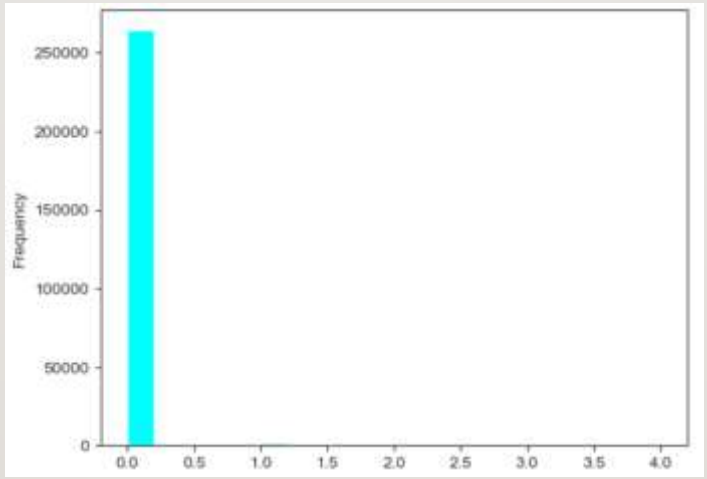


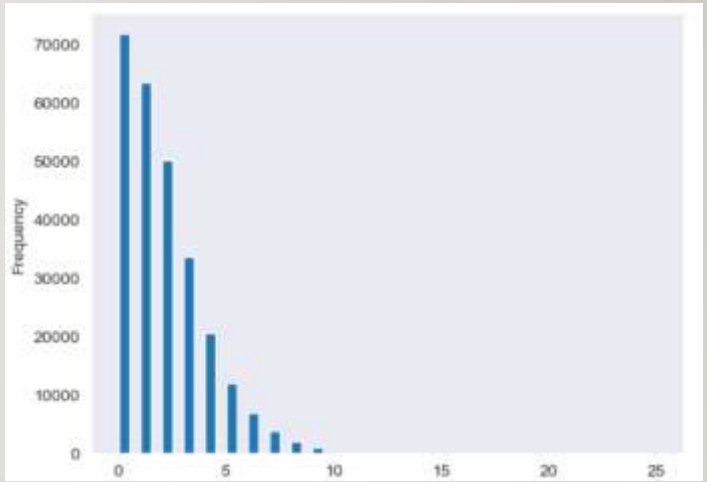




SOME EXAMPLES OF HISTOGRAM DURING CASE STUDY





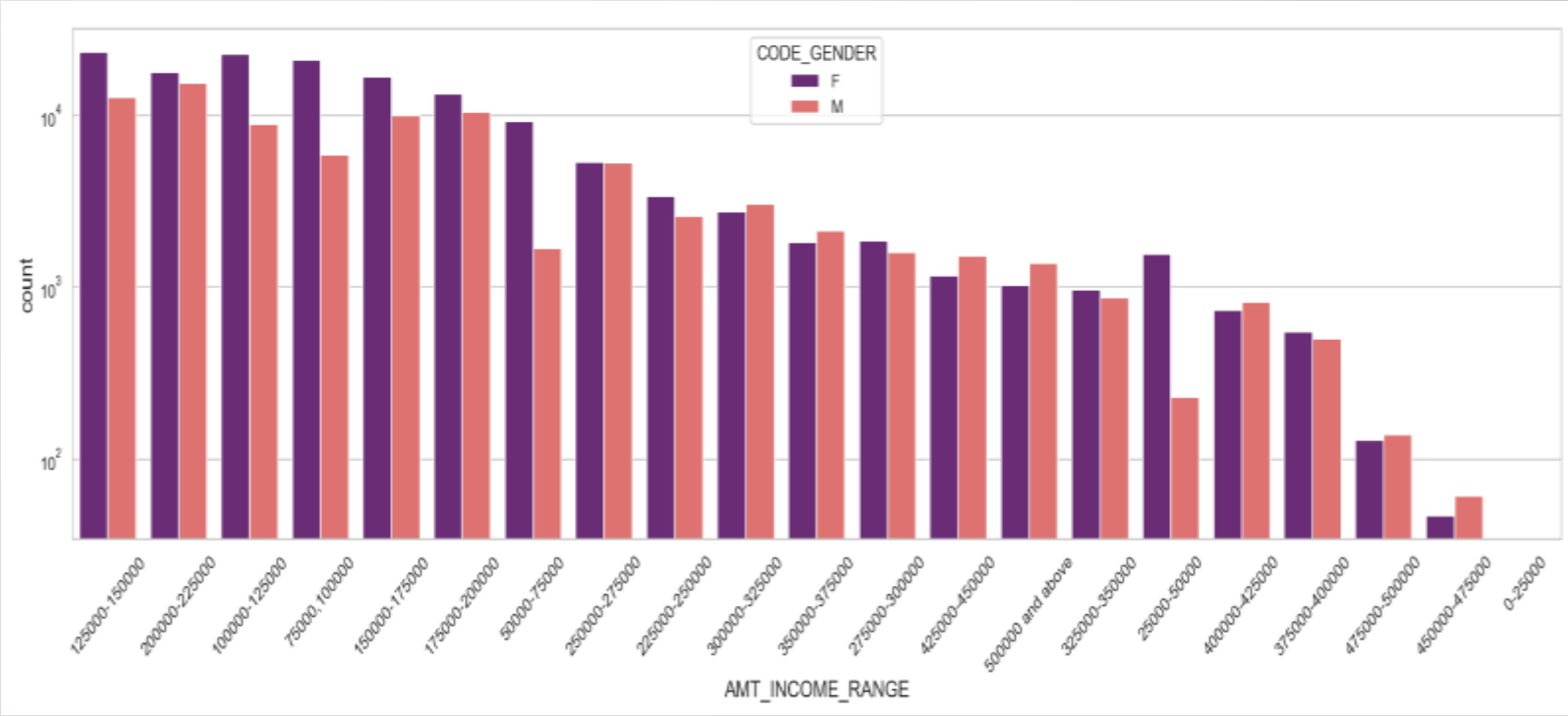


CATEGORICAL UNIVARIATE ANALYSIS

PLOTTING AND DISTRIBUTION OF INCOME RANGE

Key points from given plot:

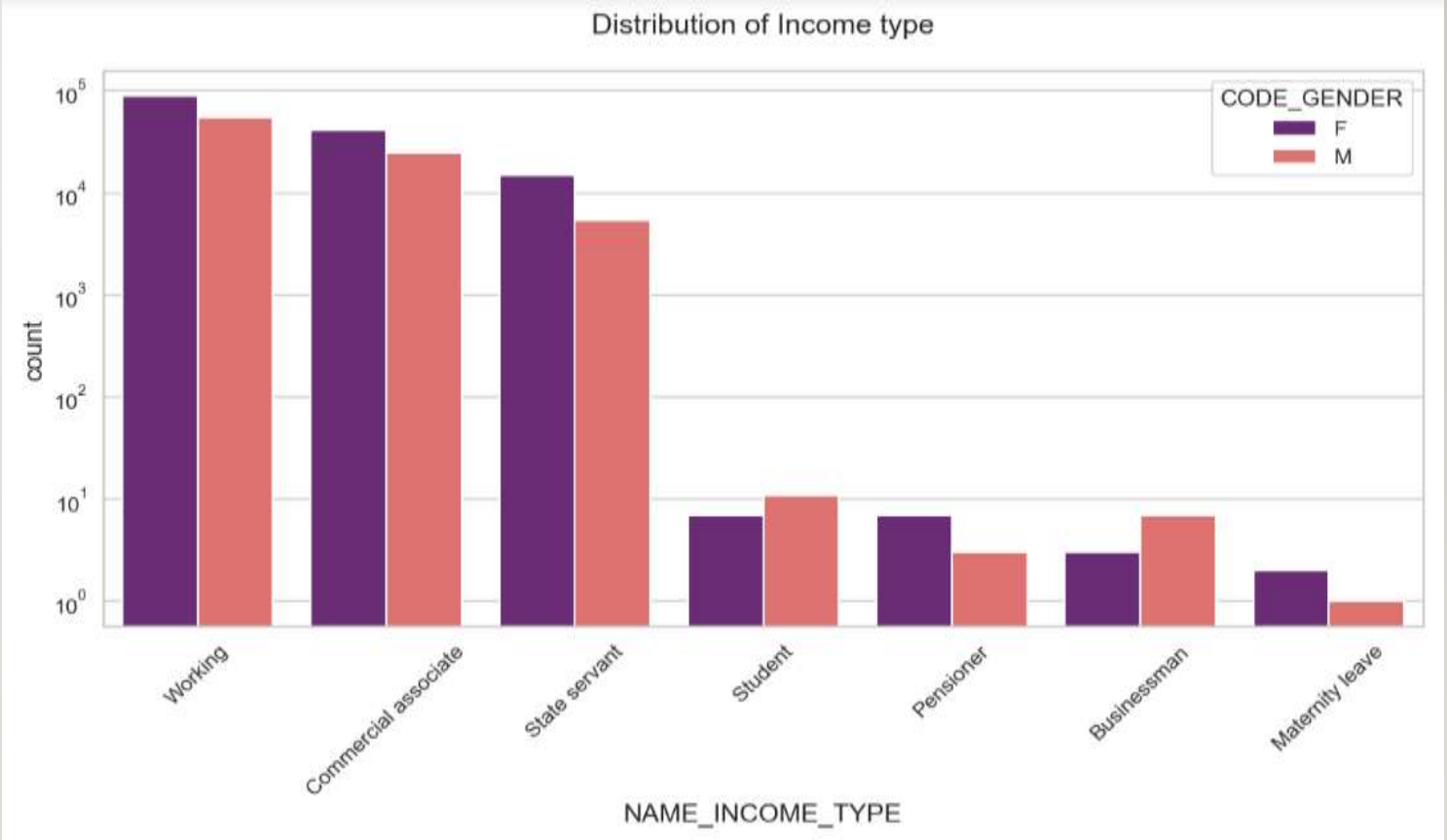
1. According to this plot it clearly shows that female count is more than male count.
2. Very less count for range of 400000 and above.
3. Higher range of credit is between 100000-200000.



PLOTTING AND DISTRIBUTION OF INCOME TYPE

Key points from given plot:

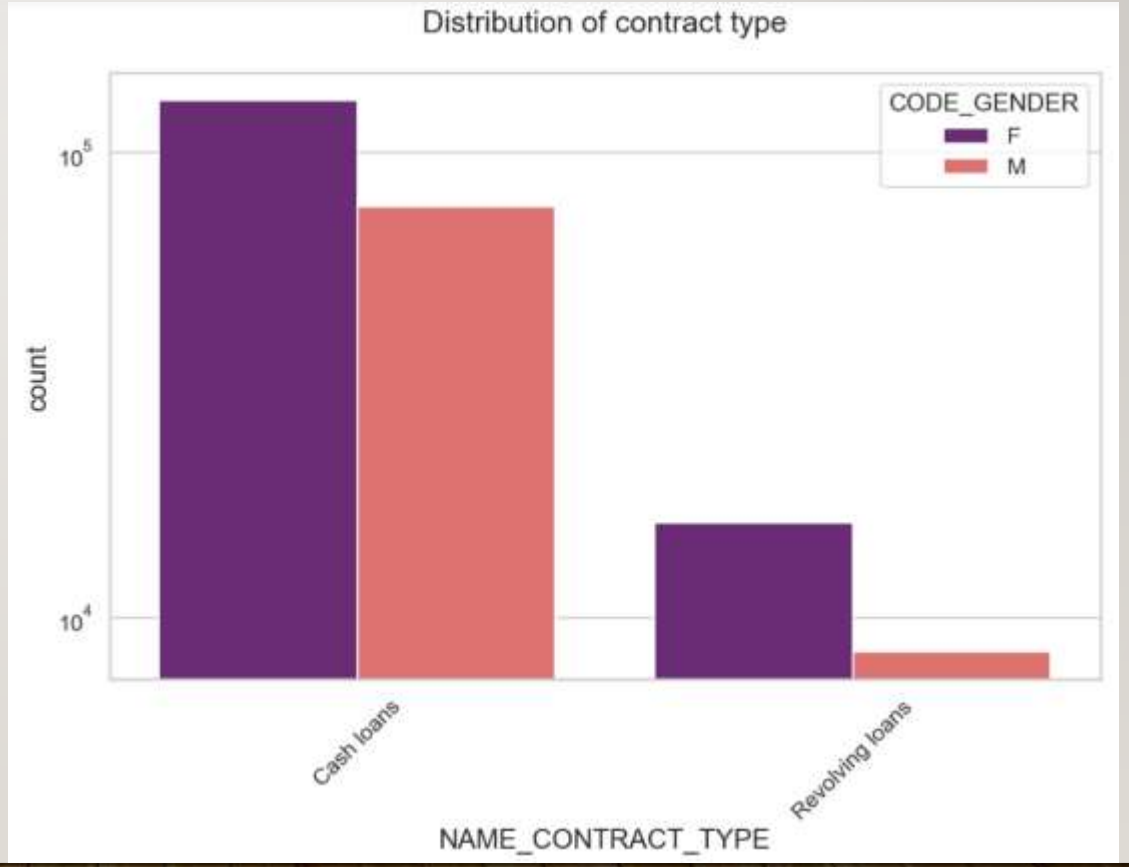
1. According to this plot it clearly shows that ‘working’ has the credit greater than others.
2. In this female credit is more than male.
3. Least number of credit for ‘student’, ’pensioner’, ‘businessman’ and ‘maternity leave’.



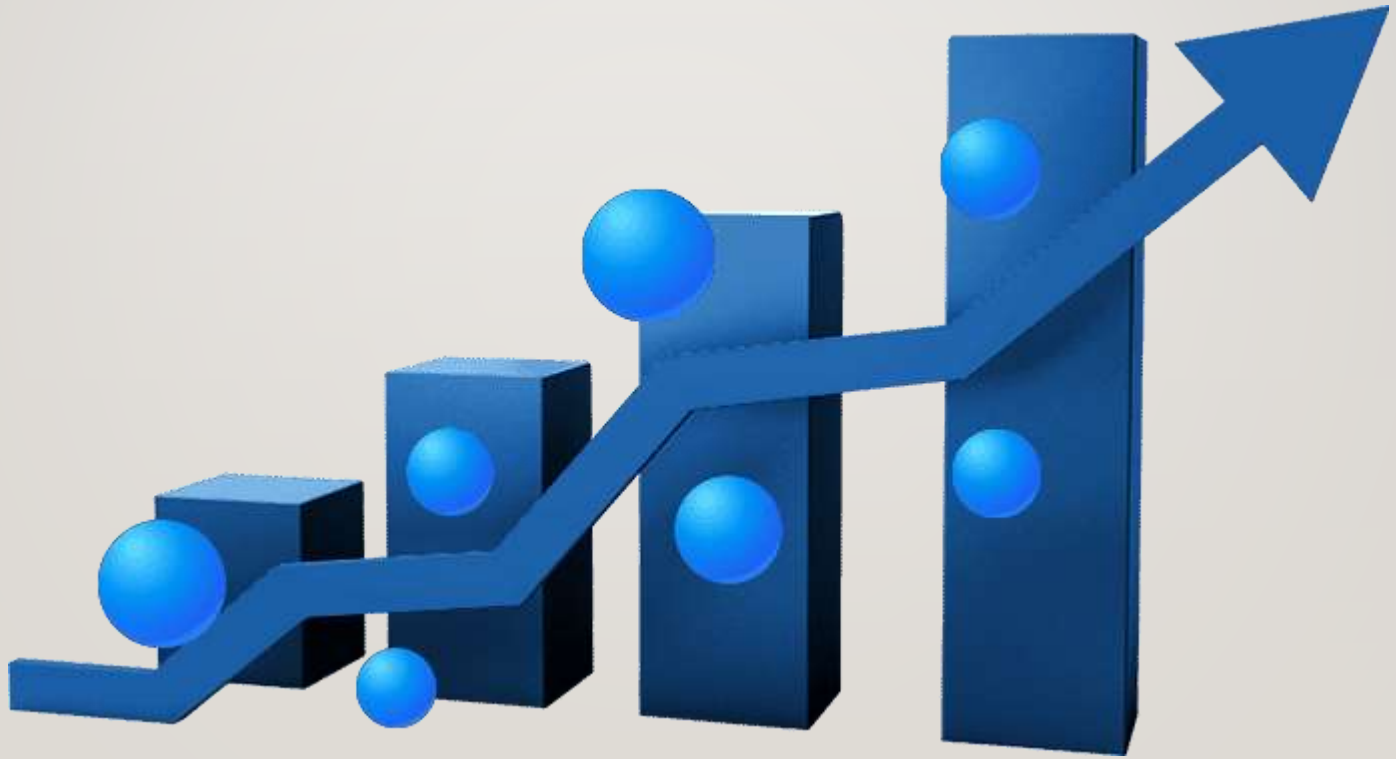
PLOTTING AND DISTRIBUTION OF CONTRACT TYPE

Key points from given plot:

1. According to this plot it clearly shows that “cash loans” has more number of credit than “revolving loans”.
2. In this plot also female credit is more than male.



CORRELATION



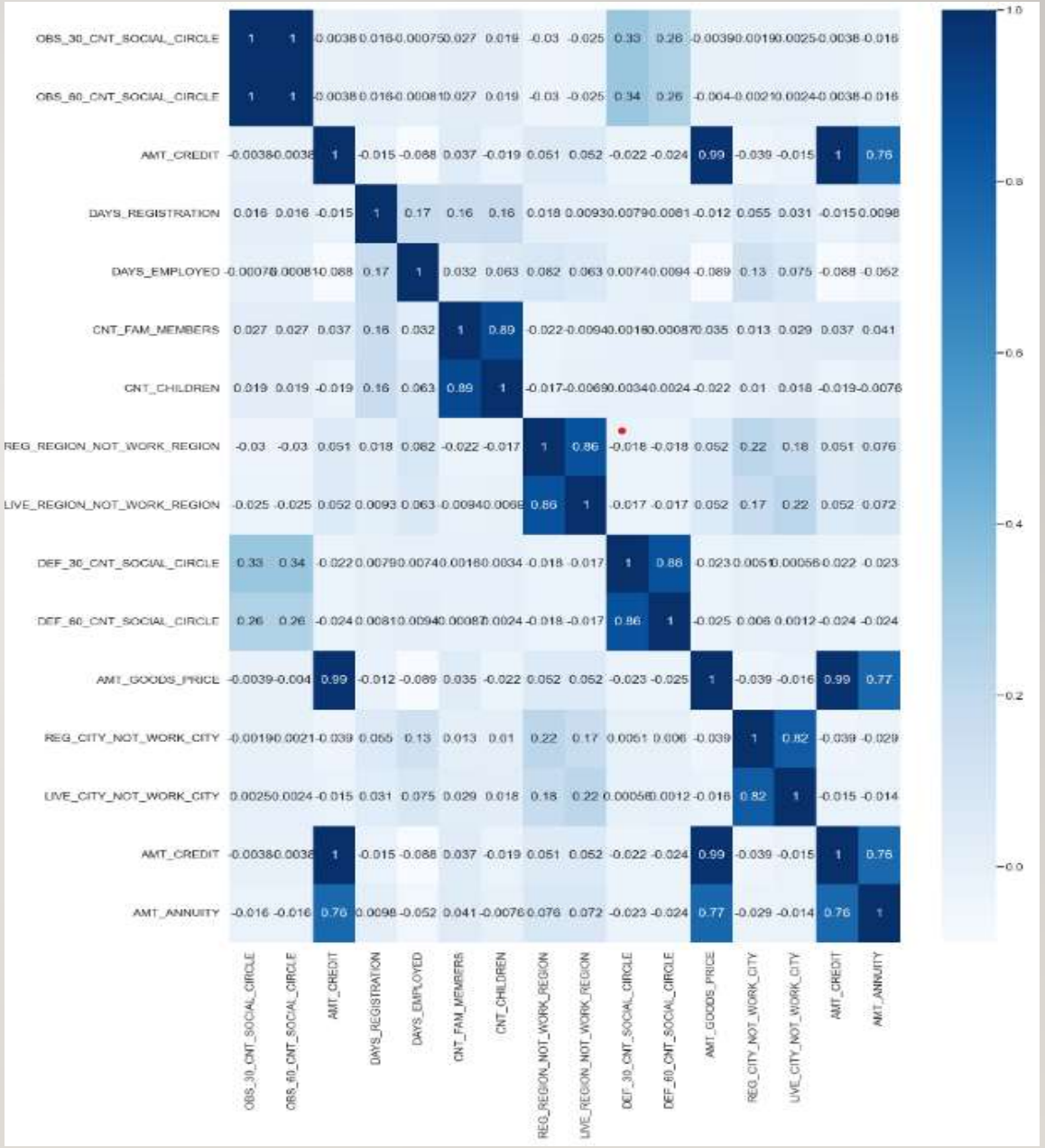
KEY TAKEAWAYS OF CORRELATION

- Correlation is a statistic that measures the degree to which two variables move in relation to each other.
- In finance, the correlation can measure the movement of a stock with that of a benchmark index, such as the S&P 500.
- Correlation is closely tied to diversification, the concept that certain types of risk can be mitigated by investing in assets that are not correlated.
- Correlation measures association, but doesn't show if x causes y or vice versa—or if the association is caused by a third factor.
- Correlation may be easiest to identify using a scatterplot, especially if the variables have a non-linear yet still strong correlation.

TOP 10 CORRELATION OF TARGET 0

- OBS_30_CNT_SOCIAL_CIRCLE OBS_60_CNT_SOCIAL_CIRCLE 1.00
- AMT_CREDIT AMT_APPLICATION 0.97
- DAYS_TERMINATION DAYS_LAST_DUE 0.93
- CNT_FAM_MEMBERS CNT_CHILDREN 0.90
- REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION 0.88
- DEF_30_CNT_SOCIAL_CIRCLE DEF_60_CNT_SOCIAL_CIRCLE 0.87
- AMT_GOODS_PRICE AMT_CREDIT 0.86
- AMT_APPLICATION AMT_GOODS_PRICE 0.85
- REG_CITY_NOT_WORK_CITY LIVE_CITY_NOT_WORK_CITY 0.83
- AMT_CREDIT AMT_ANNUITY 0.81

HEATMAP TO SHOW CORRELATION OF TARGET 0



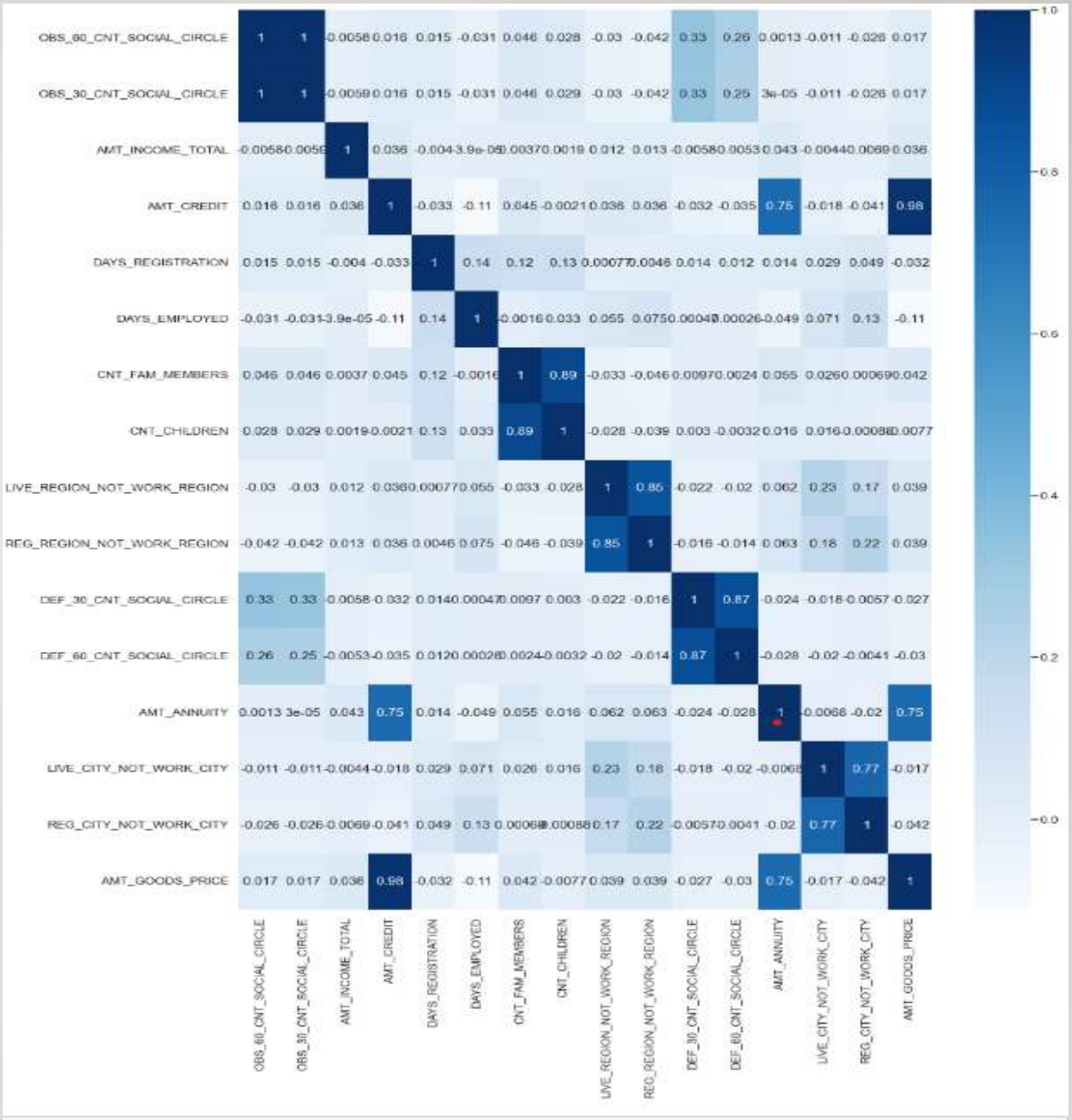
INFERENCES

- AMT_GOODS_PRICE and AMT_APPLICATION have a high correlation, which means the more credit the client asked for previously is proportional to the goods price that the client asked for previously.
- AMT_ANNUITY and AMT_APPLICATION also have a high correlation, which means the higher the loan annuity issued, the higher the goods price that the customer asked for previously.
- If the customer's contact address does not match the work address, then there's a high chance that the customer's permanent address also does not match the work address.
- CNT_CHILDREN and CNT_FAM_MEMBERS are highly correlated which means a customer with children is higly likely to have family members as well.

TOP 10 CORRELATION OF TARGET 1

- OBS_60_CNT_SOCIAL_CIRCLE OBS_30_CNT_SOCIAL_CIRCLE = 1.00
- AMT_APPLICATION AMT_CREDIT = 0.97
- DAYS_TERMINATION DAYS_LAST_DUE = 0.95
- CNT_FAM_MEMBERS CNT_CHILDREN = 0.90
- LIVE_REGION_NOT_WORK_REGION REG_REGION_NOT_WORK_REGION = 0.87
- DEF_30_CNT_SOCIAL_CIRCLE DEF_60_CNT_SOCIAL_CIRCLE = 0.86
- AMT_CREDIT AMT_ANNUITY = 0.83
- LIVE_CITY_NOT_WORK_CITY REG_CITY_NOT_WORK_CITY = 0.78
- AMT_ANNUITY AMT_GOODS_PRICE = 0.76
- AMT_ANNUITY AMT_CREDIT = 0.74

HEATMAP TO SHOW CORRELATION OF TARGET 1



INFERENCES

- In comparison to the Non_defaulter heatmap, AMT_GOODS_PRICE and AMT_APPLICATION have a high correlation here as well.
- In comparison to the Non_defaulter heatmap, AMT_ANNUITY and AMT_APPLICATION also have a high correlation, which means the higher the loan annuity issued, the higher the goods price that the customer asked for previously.
- CNT_CHILDREN and CNT_FAM_MEMBERS are highly correlated which means a customer with children is higly likely to have family members as well (same as with the Non_defaulter heatmap).
- Higher the goods price, higher the credit by the customer.

CLIENT CATEGORIES TO BE TARGETED FOR PROVIDING LOAN

- Clients who are employed for more than 19 years
- Clients in the age range 30-40 and 40-50
- Clients who are Married
- Male clients with Academic degree
- Students and Businessman
- Repeater clients

END OF PRESENTATION

THANK YOU