

EXPLORATORY DATA ANALYSIS ON CREDIT RISK ANALYSIS

PREPARED BY
SHAIL SHAH



WHAT IS CREDIT RISK

Credit risk is the credit risk or potential loss that may occur due to failure of borrower/ counterparty to meet the obligation on agreed terms and conditions of financial contract

AIM OF THE CASE STUDY

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.



WE HAVE USED TWO DATASET

- 1)Application_data.csv
- 2)Previous_data.csv

STEPS FOR EDA

- 1) Data sourcing
- 2) Data cleaning
- 3) Data pr-processing
- 4) Data Visualiszation
- 5) Conclusion



Data collection is the process of gathering information in an established systematic way that enables one to test hypothesis and evaluate outcomes easily.

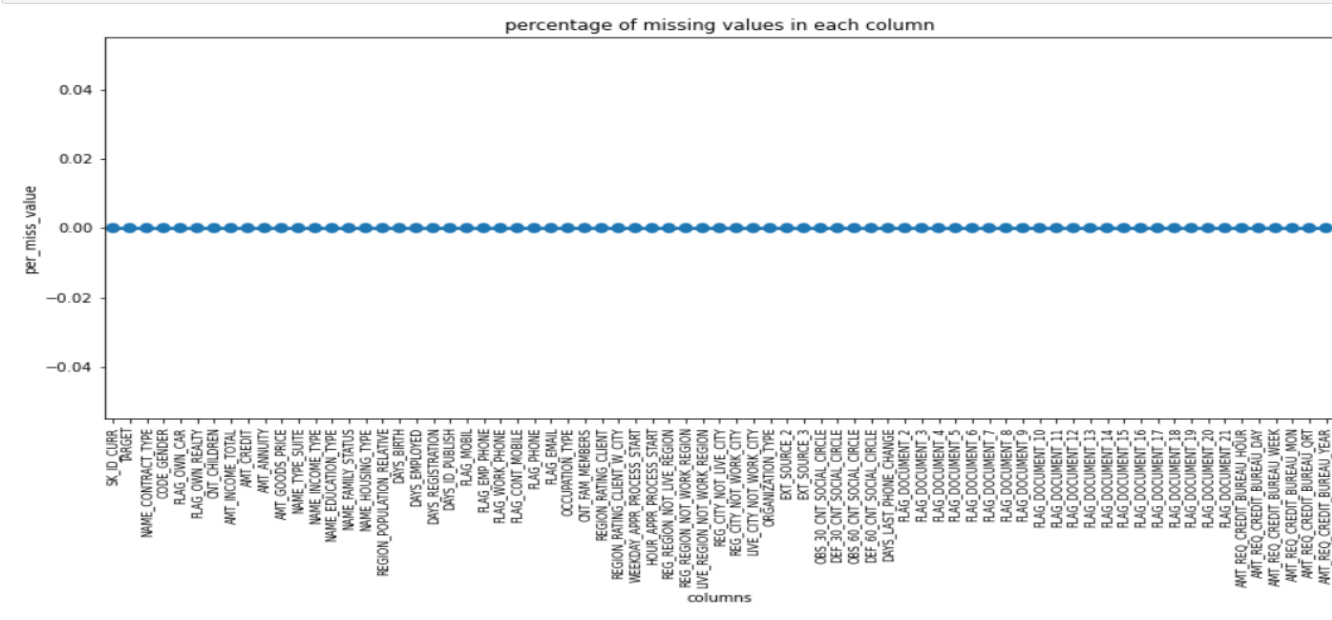
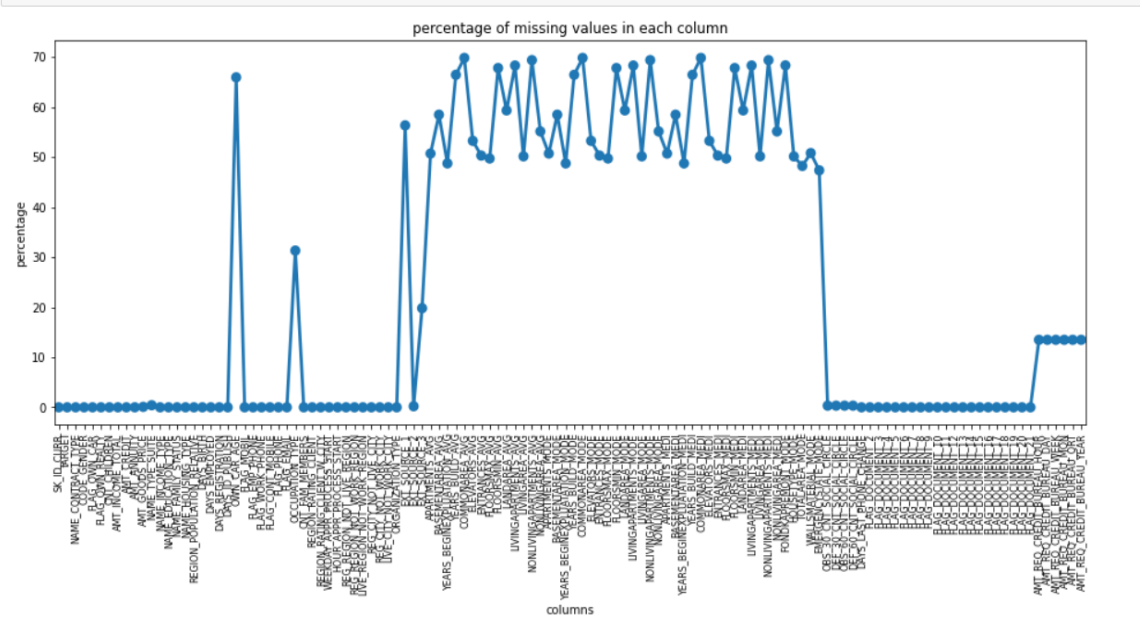
Data cleaning is the process of ensuring that your data is correct and useable by identifying any errors in the data, or missing data by correcting or deleting them.

Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. It includes normalisation and standardisation, transformation, feature extraction and selection, etc. The product of data pre-processing is the final training dataset.

Data visualisation is the graphical representation of information and data. It uses statistical graphics, plots, information graphics and other tools to communicate information clearly and efficiently.



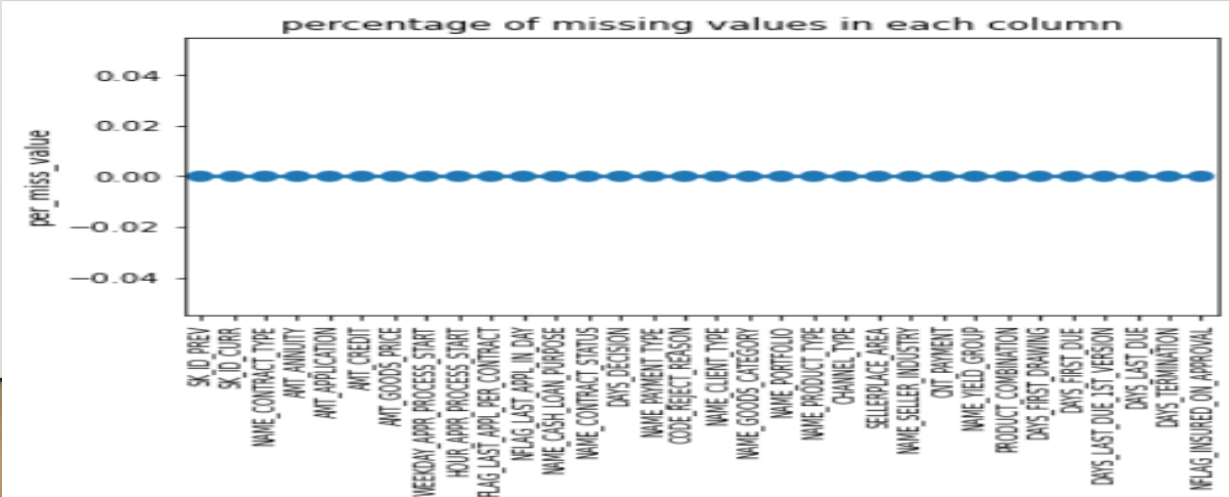
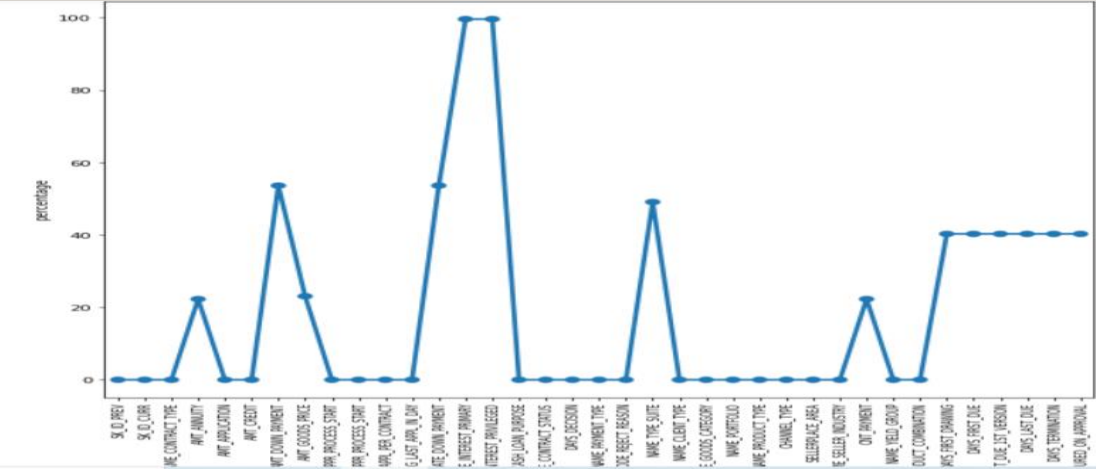
From application dataset



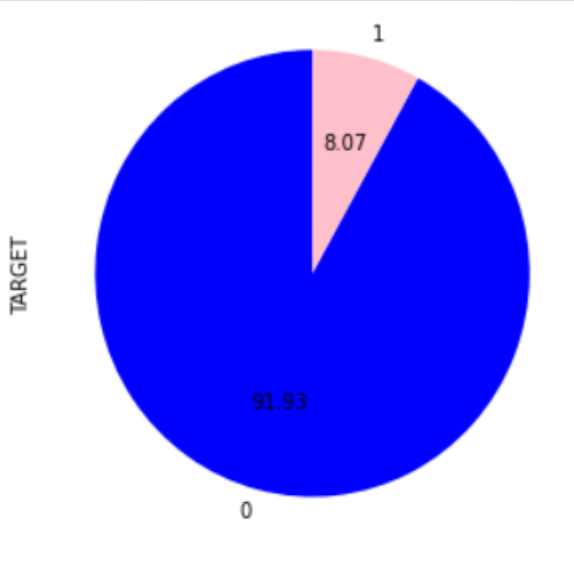
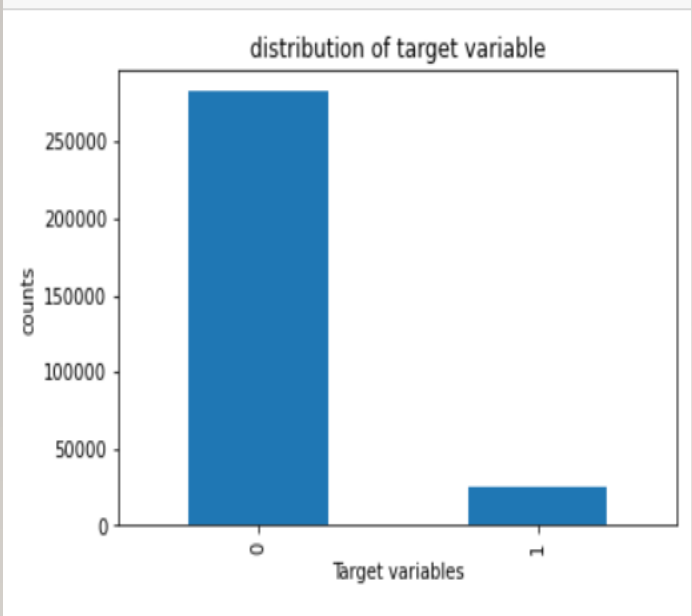
Before cleaning dataset having null values

After cleaning

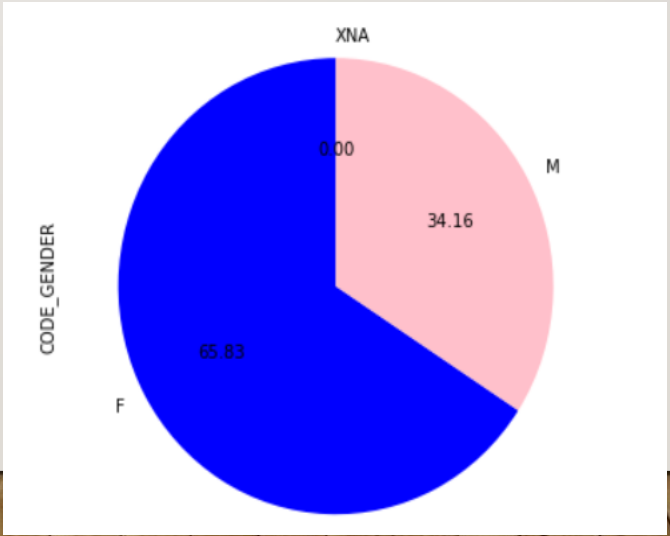
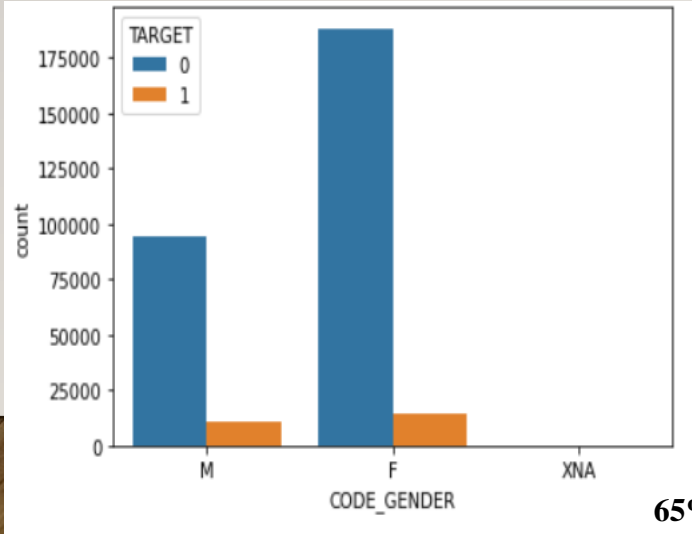
From previous dataset



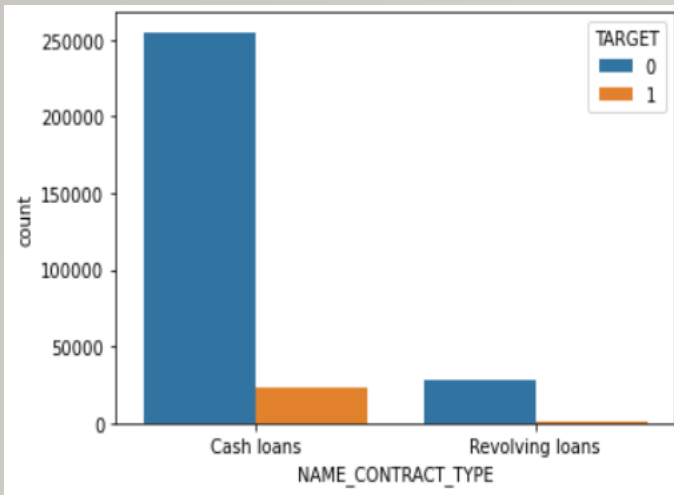
Data exploration and visualization



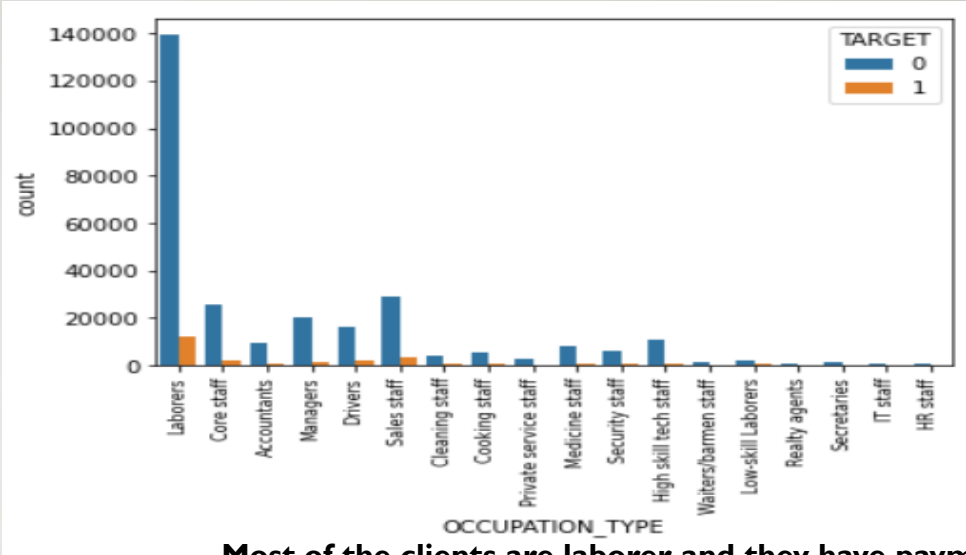
Target variables are variables which shows the defaulters, out of the total clients who has applied for loans , 8% are likely to fail to repay the loan



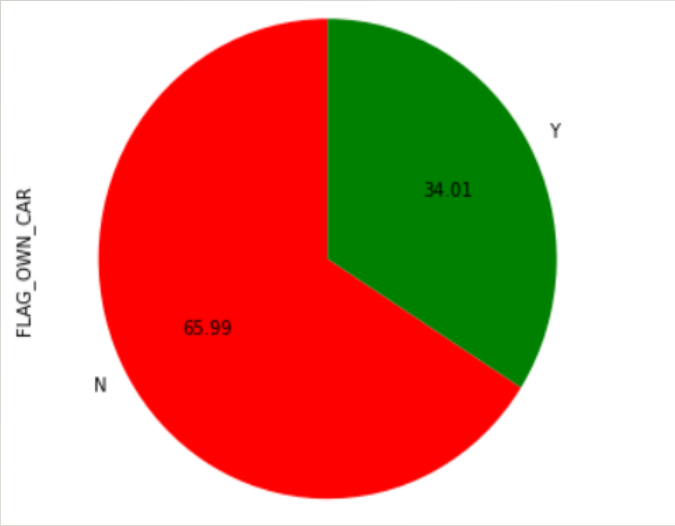
65% of the clients are female and they have more difficulties in paying loan comparison to male clients.



Most of the clients applied for cash loan and among those clients are the defaulters



Most of the clients are laborer and they have payment difficulty than IT or HR staff

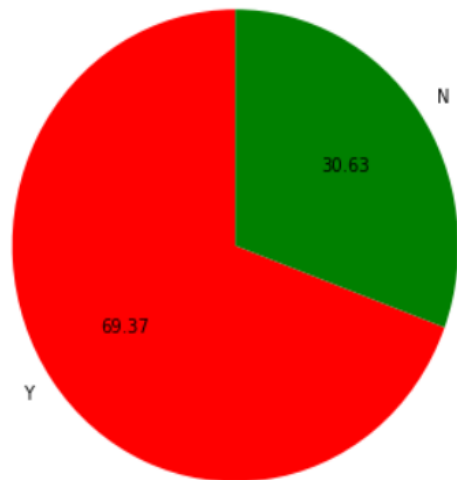


66% of the clients don't have car

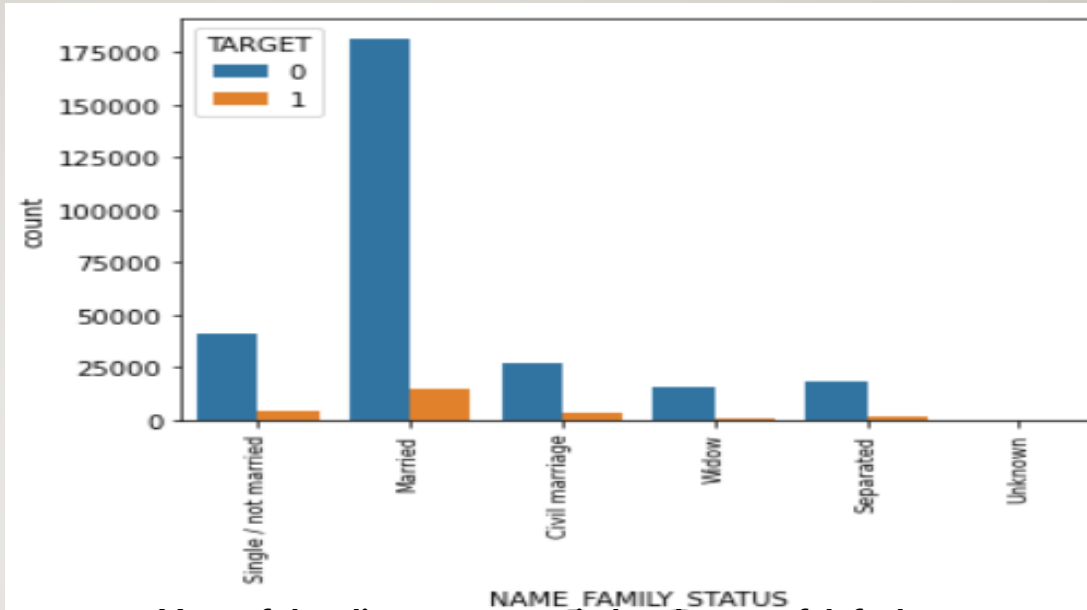


57% of male clients and 43% of female clients own car

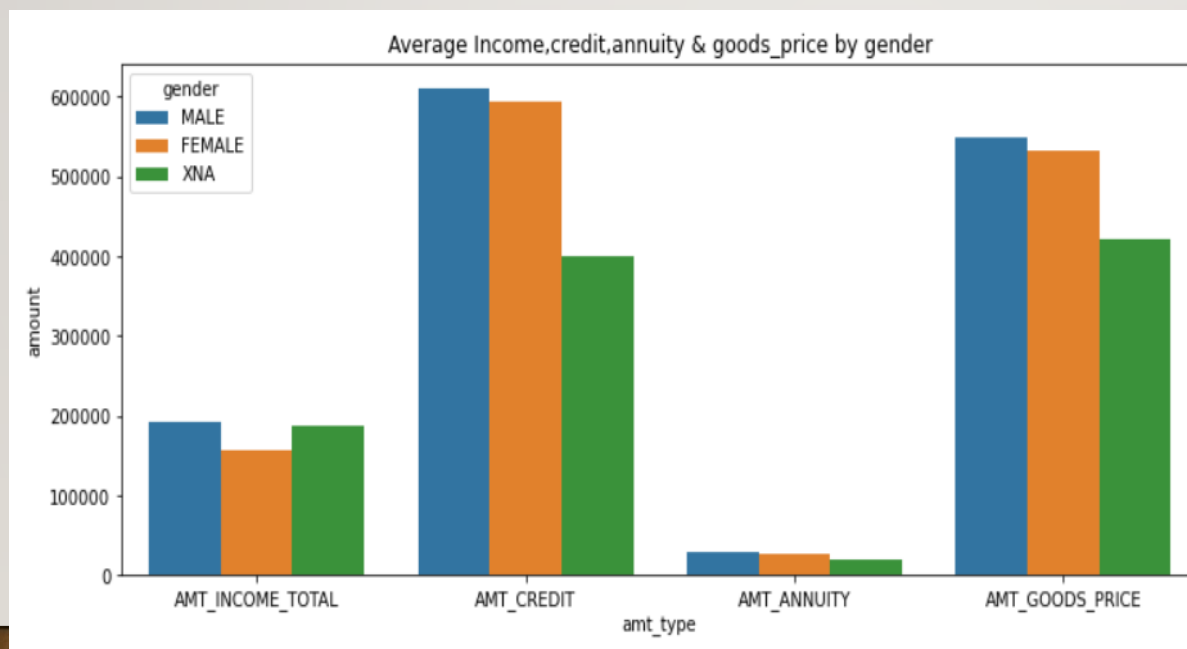
FLAG_OWN_REALTY



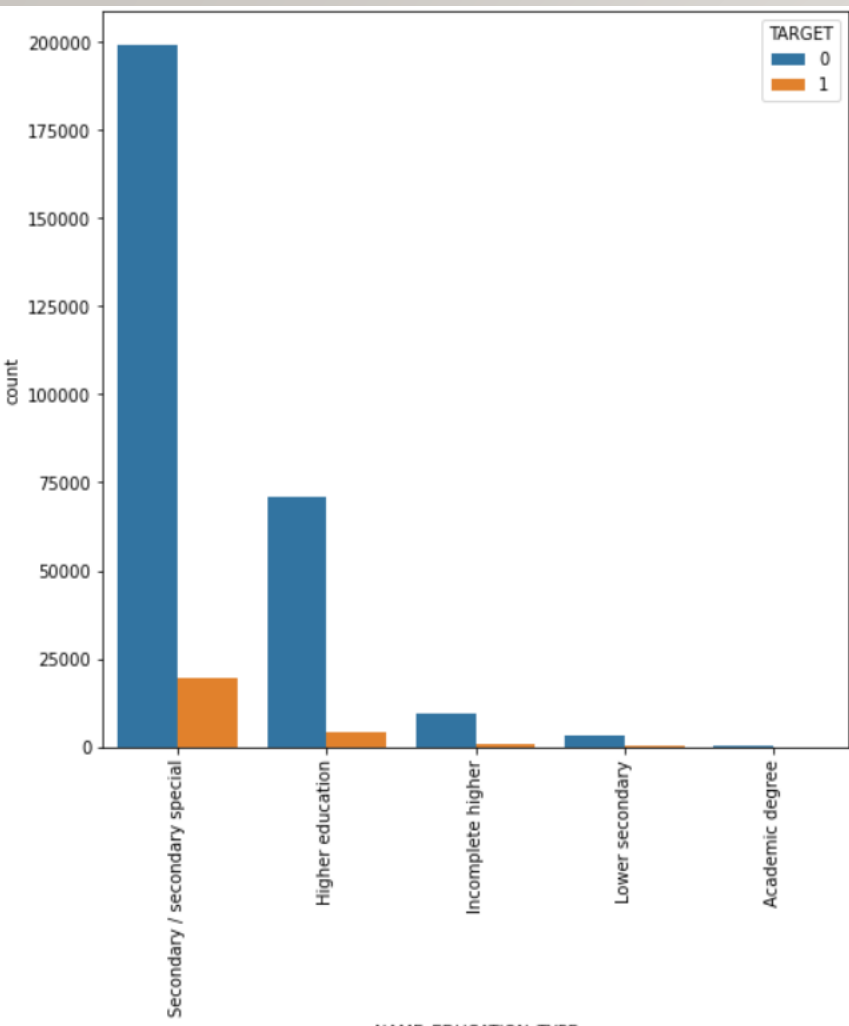
69% of the clients have their property



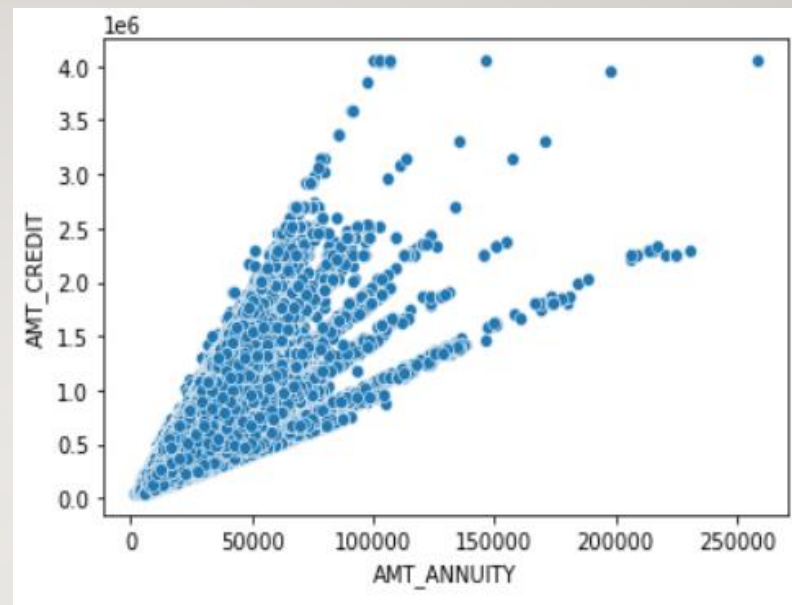
Most of the clients are married and most of defaulters are among them



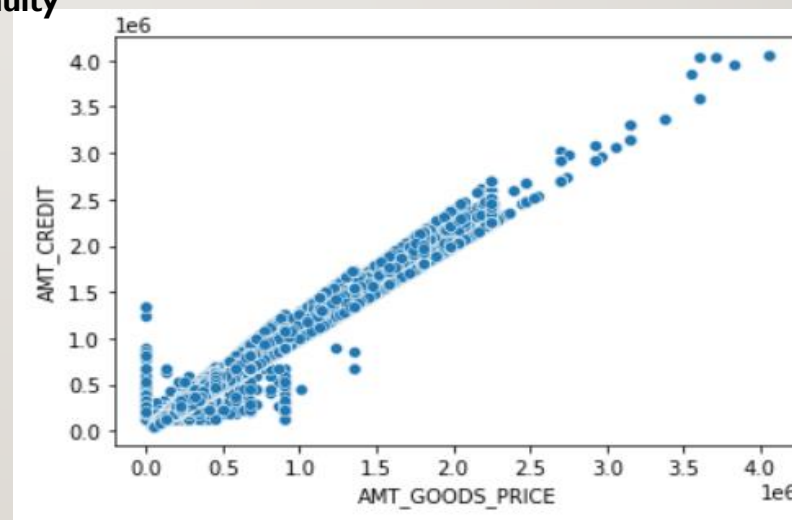
average income and amount annuity paid is same for both male and female.



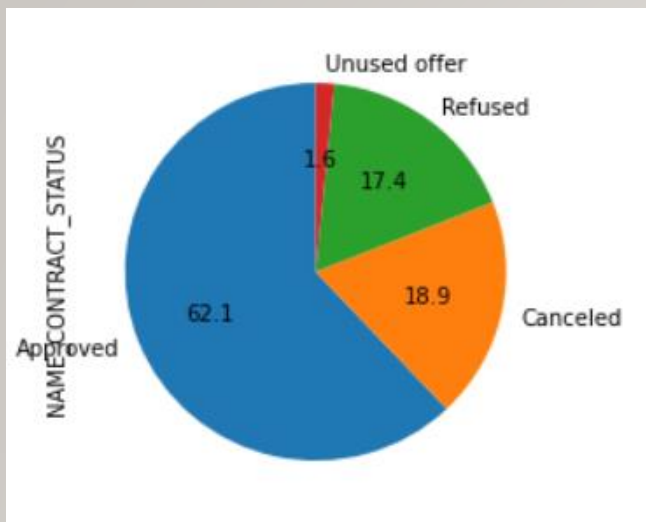
most of the clients with less than higher education are defaulters



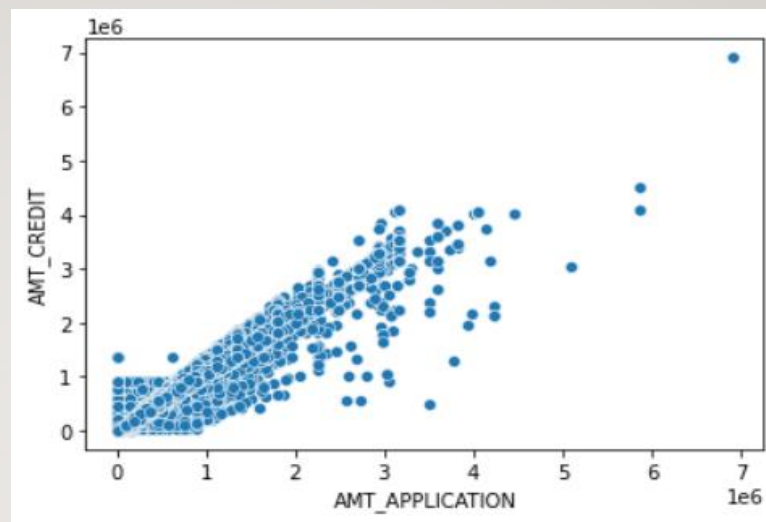
Its positively correlated. Higher the loan amount loan, higher will be the annuity



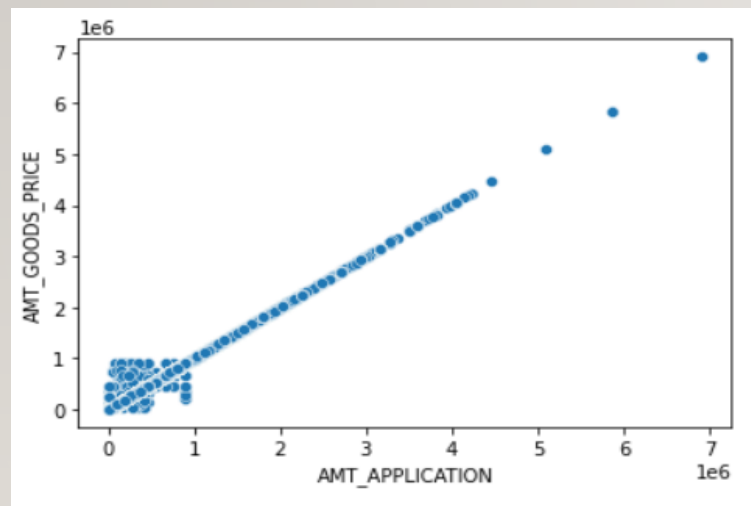
Its positively correlated with 0.9 correlation. Higher the price of goods, higher the loan amount respectively



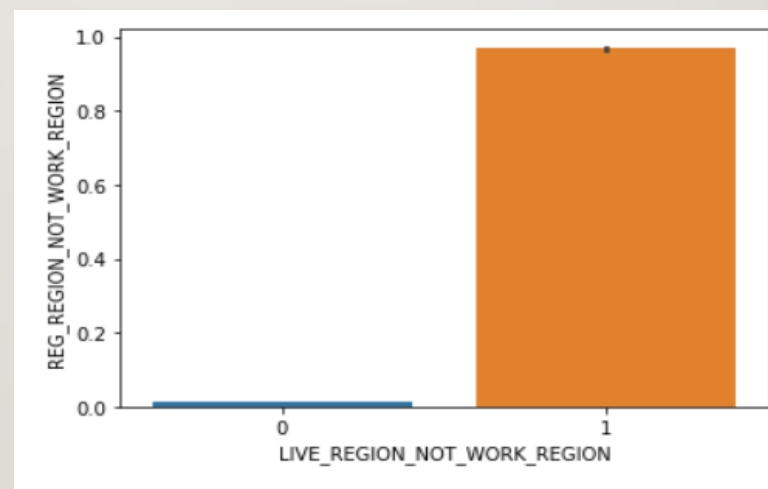
62% of the loan has been approved



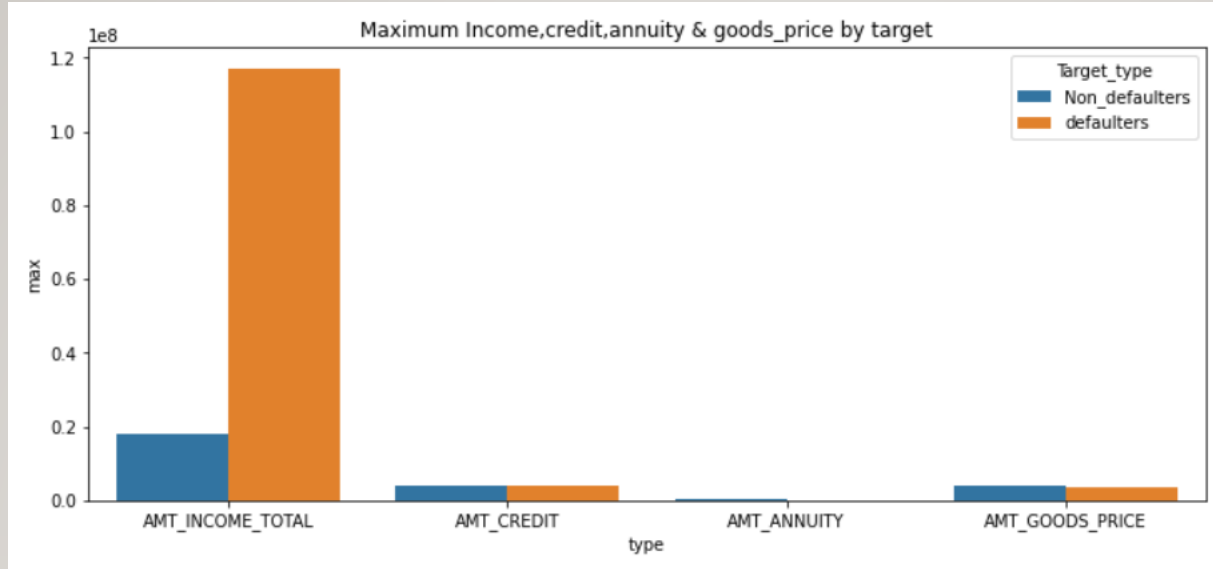
correlation is 0.9. means the amount has been credited as applied for.



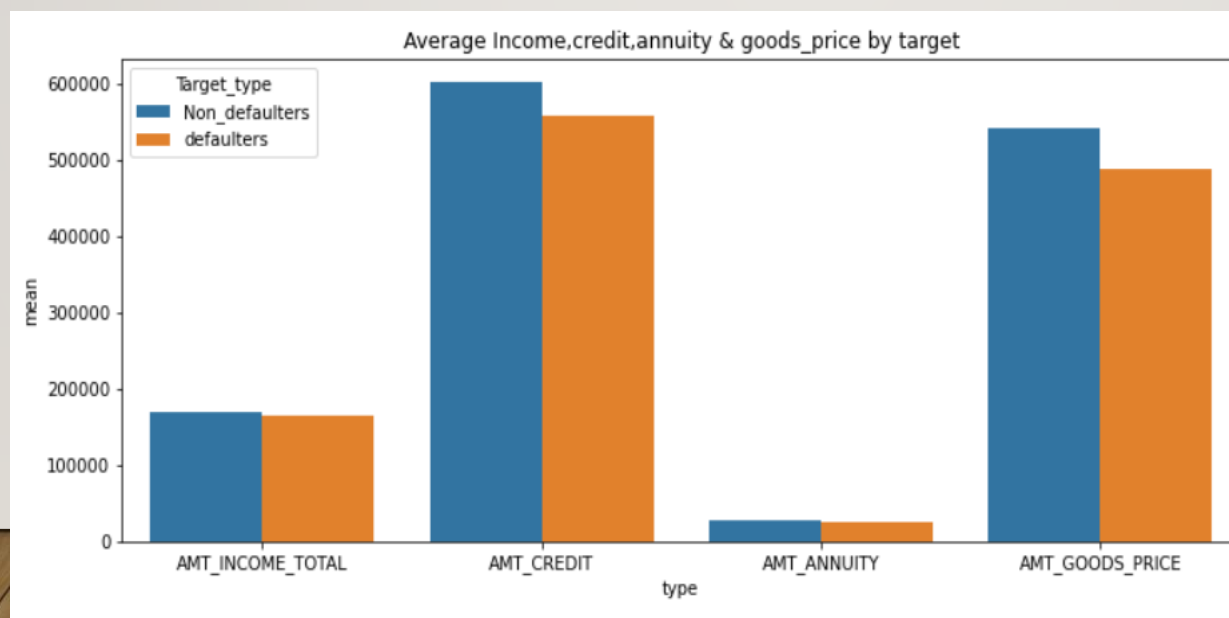
positive correlation.



most of the client's permanent address differs from work address



Defaulters are among those who has maximum income



average income and annuity is almost same for defaulters and non defaulters

CONCLUSION

- **Only 8% of the clients are defaulters**
- **Out of these 8% defaulters most of them are female clients**
- **66% of the clients don't have car and among 66% female clients don't have car**
- **Married clients are more defaulters.**
- **Most of the defaulters have less than higher education**
- **Defaulters are among those who has maximum income**

