Bank Loan Analysis

An educational case study

Rajiv Sharma
Vinay Dayal Dadu

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 and Hypothesis is the aim of this case study

Hypotheses

- H0 = Applicant is likely to repay the loan

- H1 = Applicant is NOT likely to repay the loan

|	H0 is True	H0 is False	Description of errors
We decide to reject H0	Alpha (New Business Loss)	Correct Decision	Type 1(Alpha): If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
We fail to reject H0	Correct Decision	Beta (Financial Loss)	Type 2(Beta): If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.
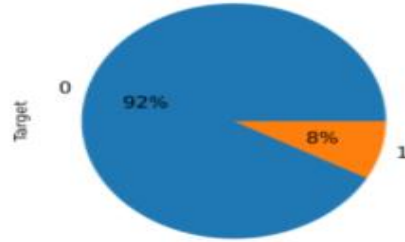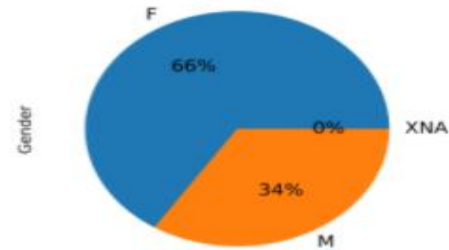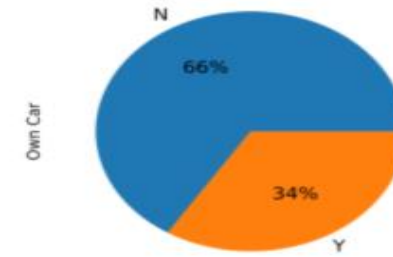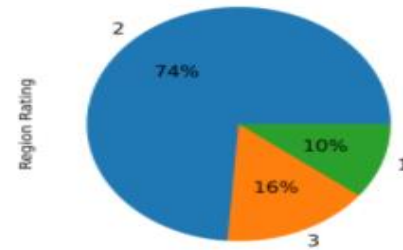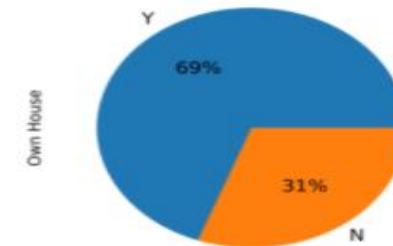
Understanding Variables
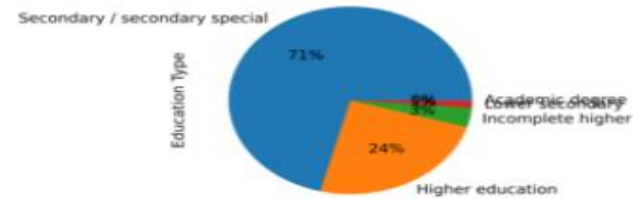
Understanding Variables
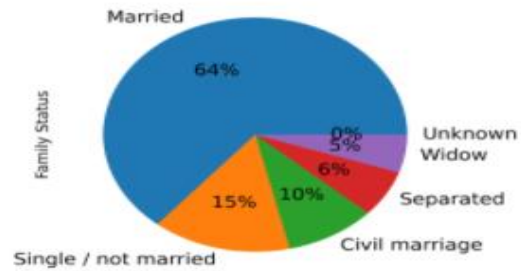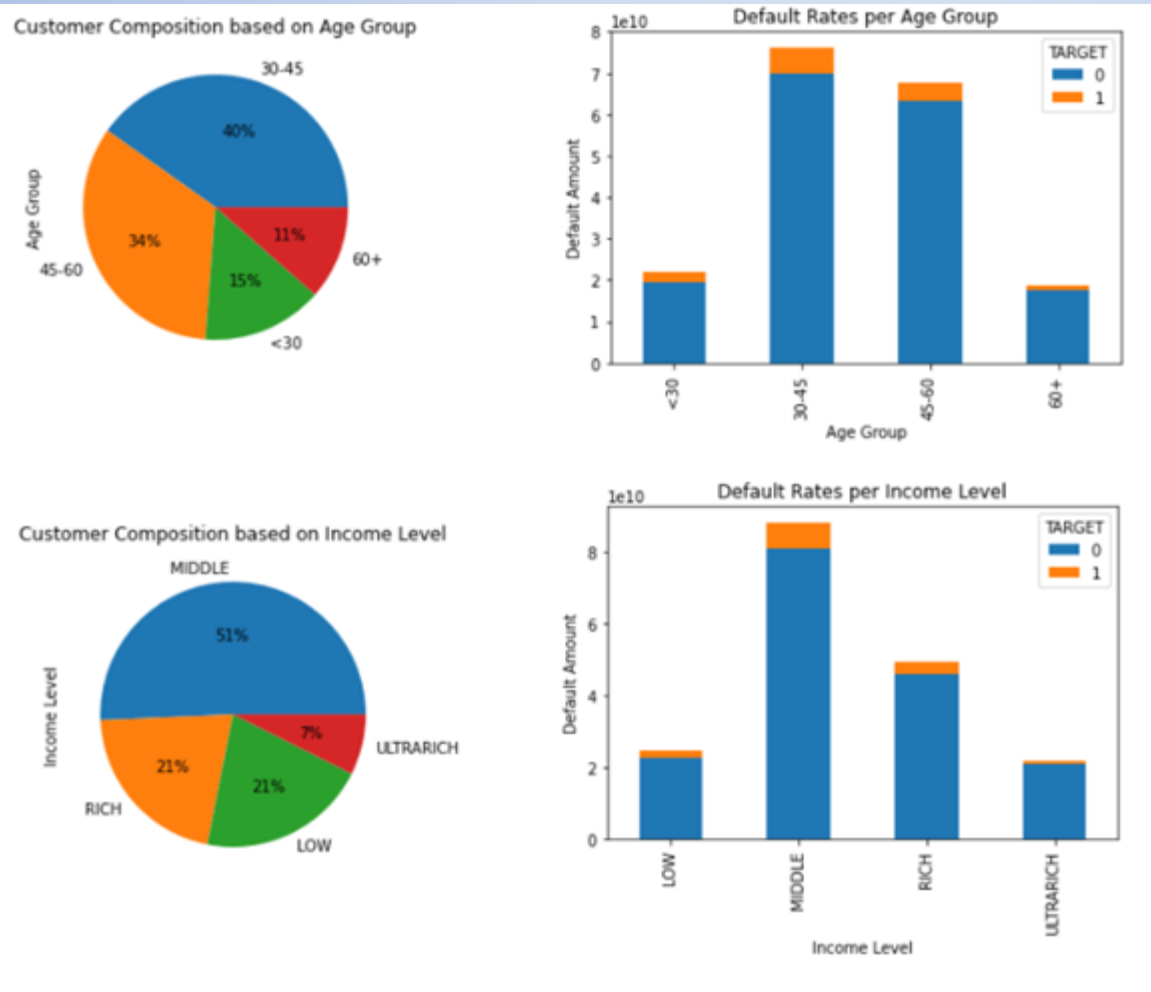

- 85% of the customers are above 30+ years.
- Default rates are higher for <30 years age group.
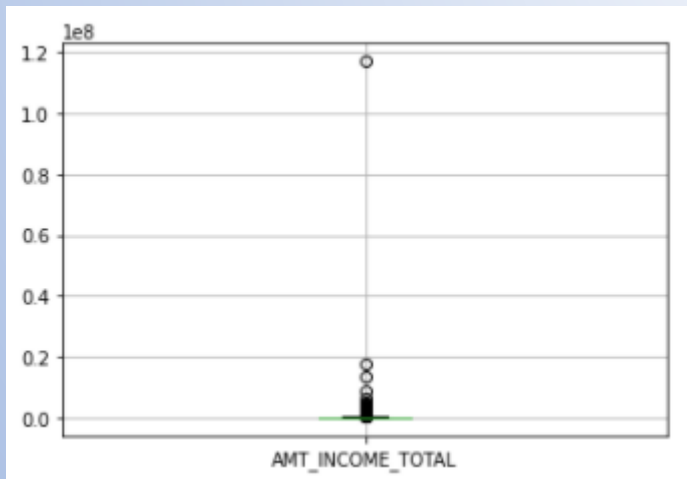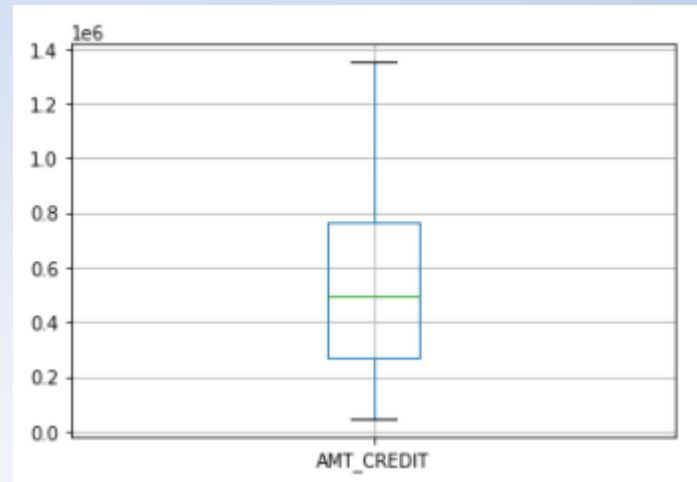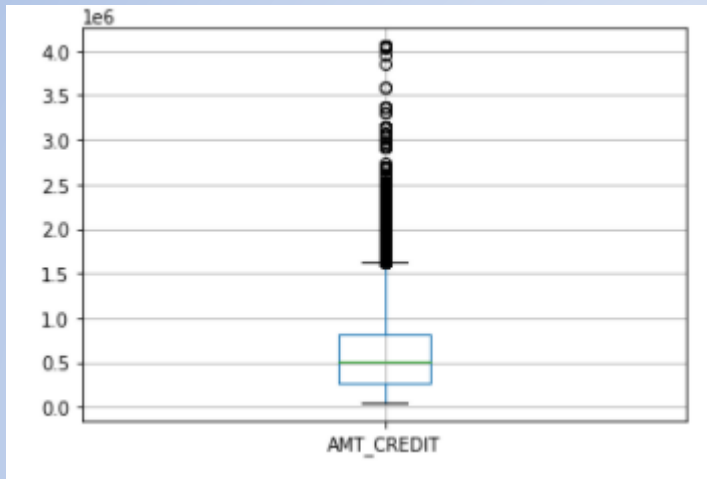- As people become older, the default rates are getting smaller. The default rates (as a %of loan amount within the age group) for <30 age group is close to 11% where as it is reduced to 5.2% for 60+ age group

- Middle Income Group have secured the highest loan amount from the bank.
- These people have also the highest %age default amount causing losses to the bank

Removing Outliers
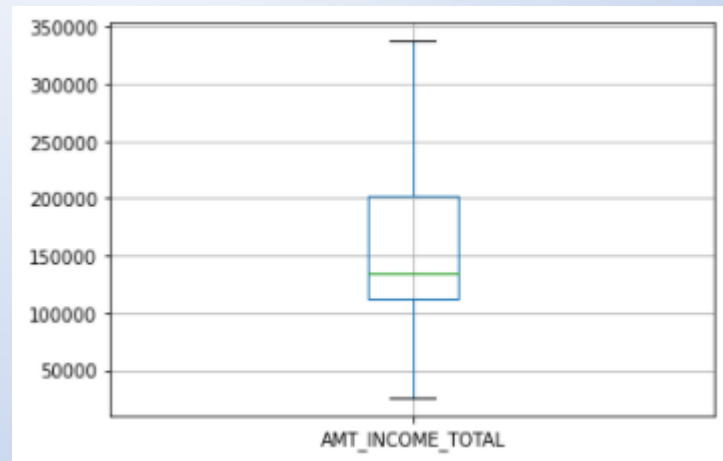
Default Rates
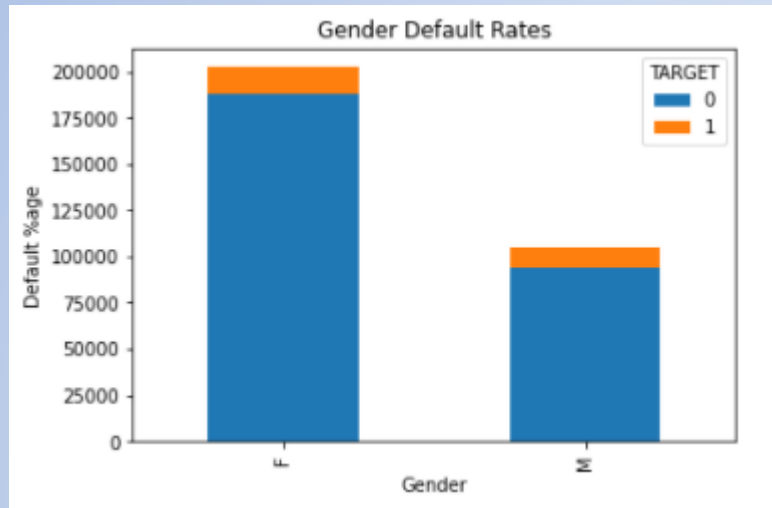
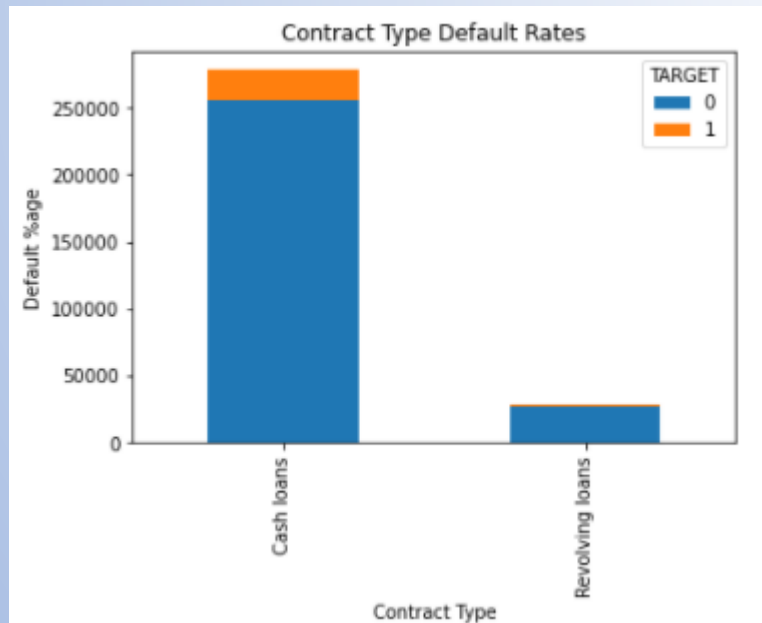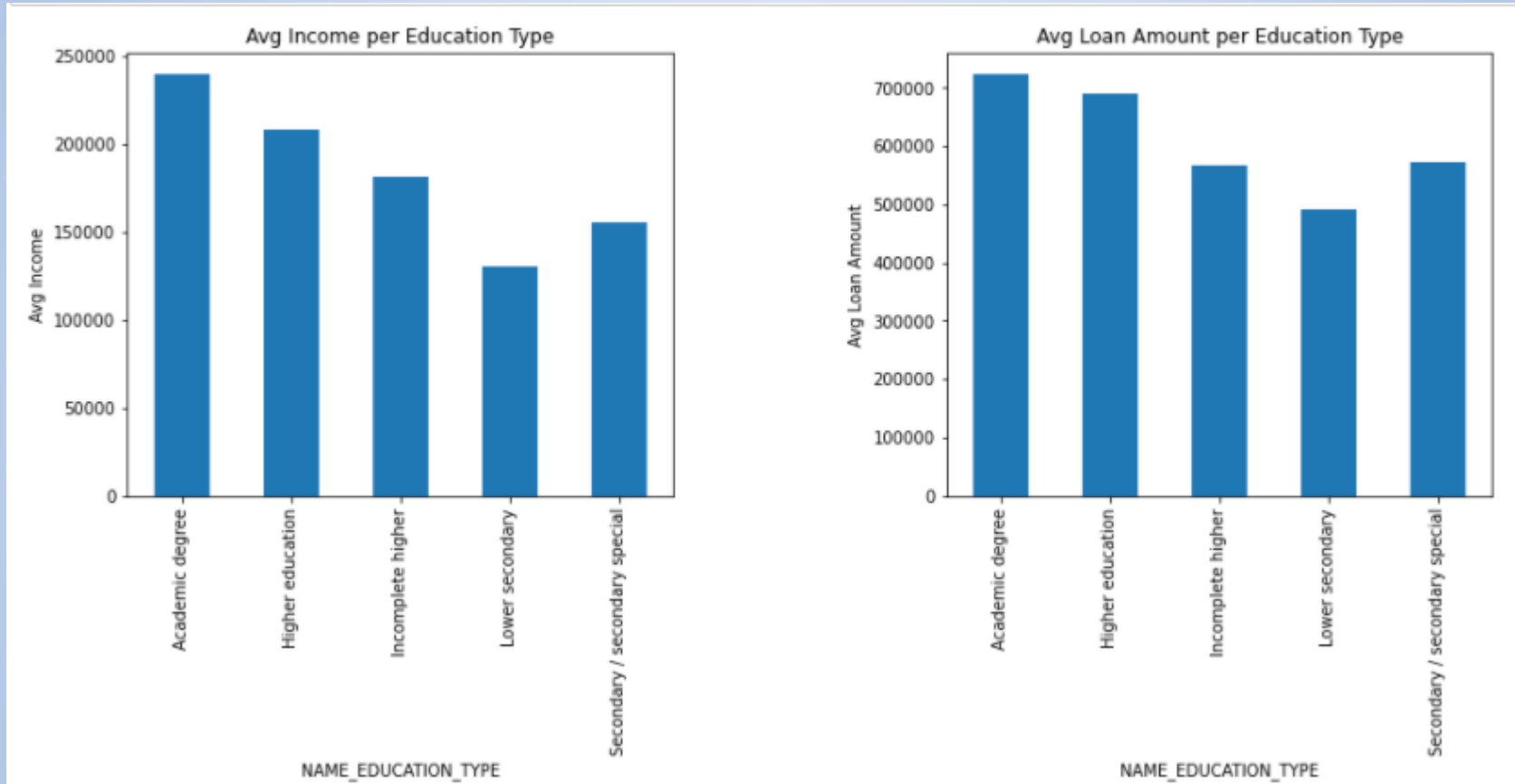- Loans granted to female population are higher than loans granted to male population as seen in slide #4
- The default rate of female population is lesser at 7% compared to male population at 10%
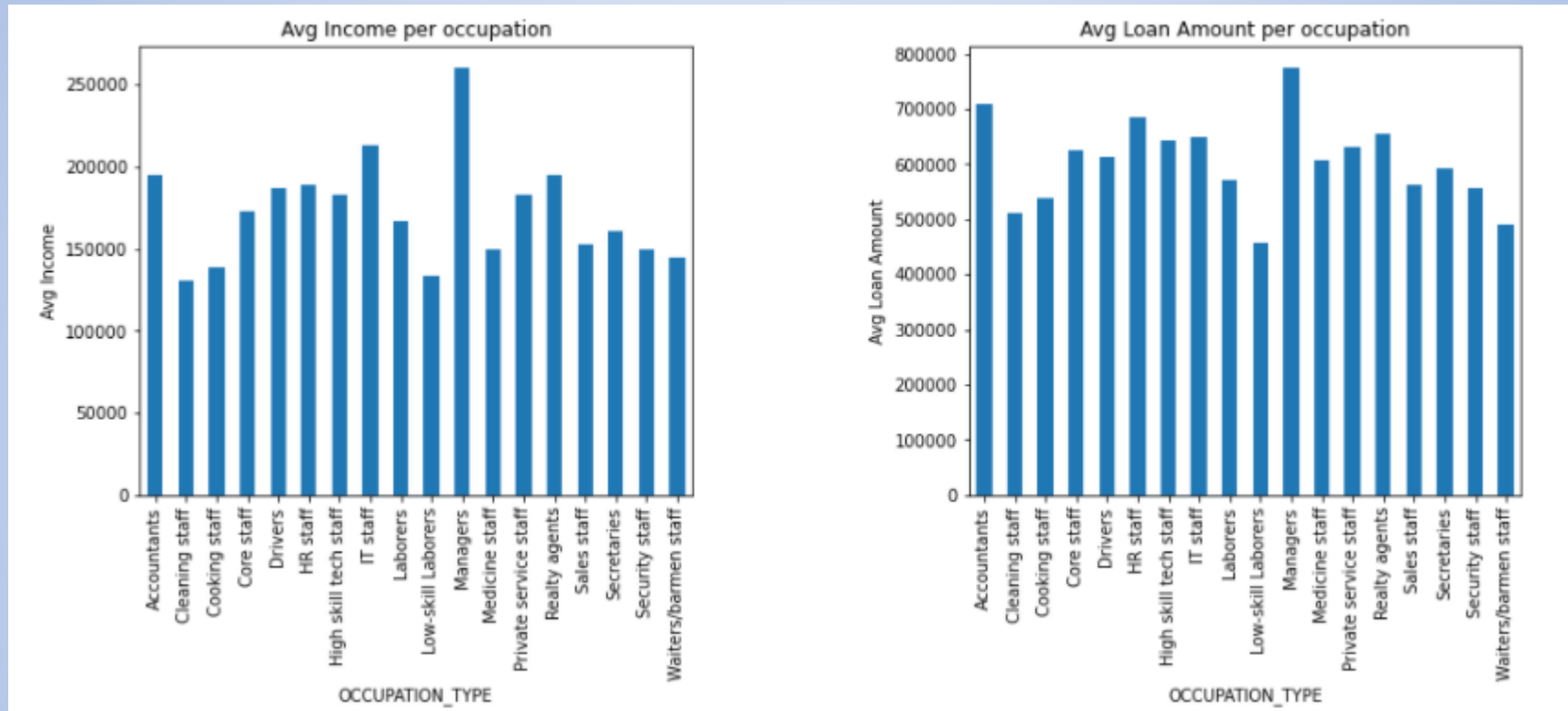

- Cash Loans have higher default at 8.3% compared to Revolving loans at 5.4%

Income and Loan Amount basis Education type

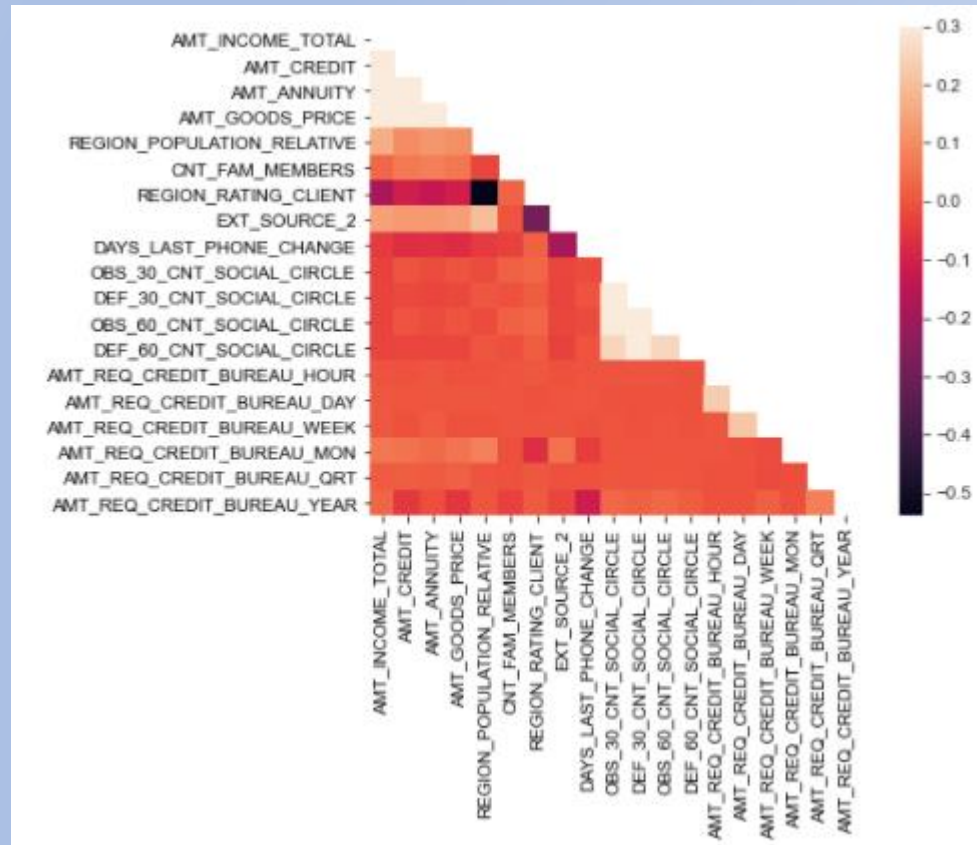# Income and Loan Amount basis occupation type

Income and Loan Amount basis income type
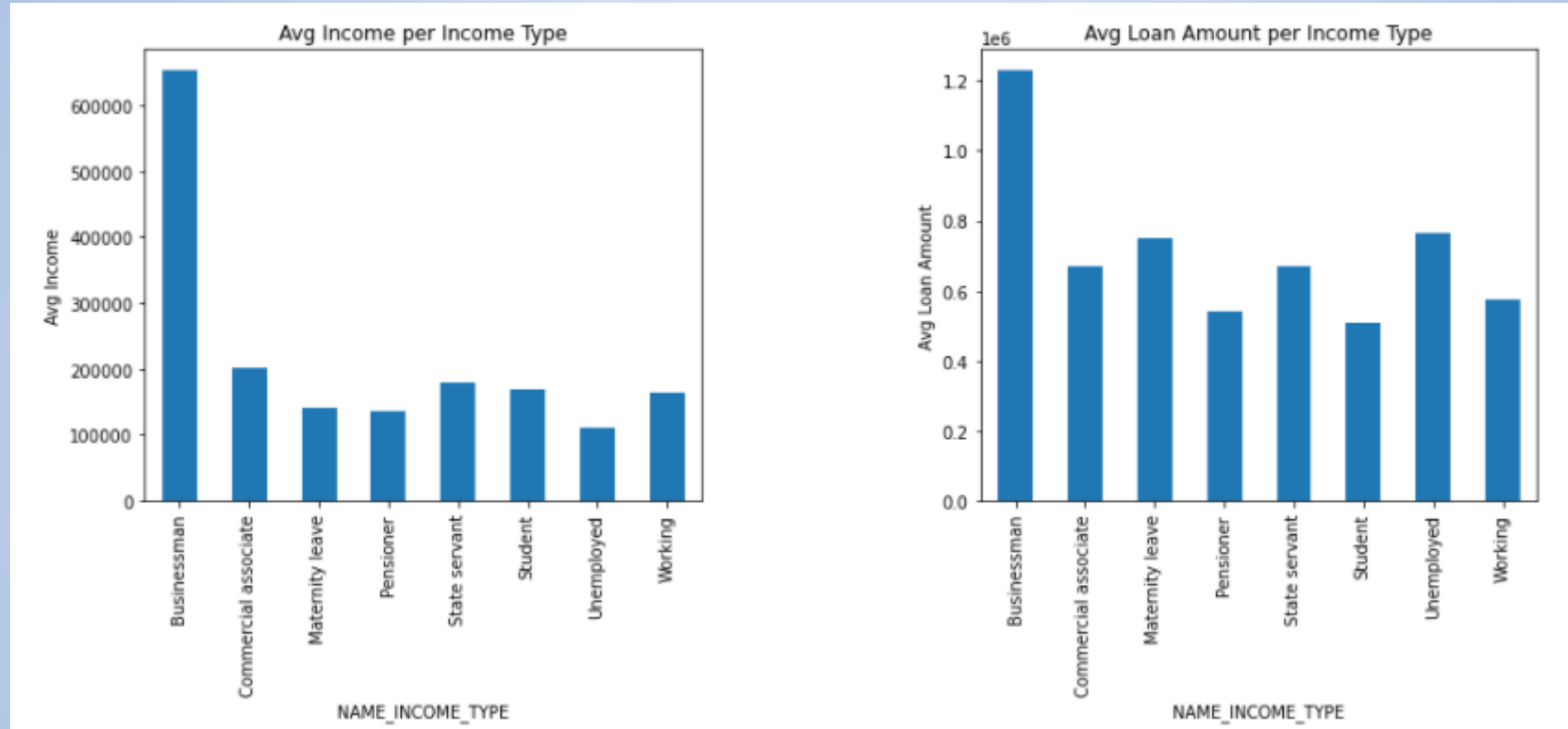
# Correlation for customers who pay on time

- Region Rating has a high correlation with population density prompting the analyst to dig deeper
- Days phone change seems to have correlation with credit bureau enquiry for the client over the one year periods, although this correlation does not seem to be the causation

Correlation for customers who default
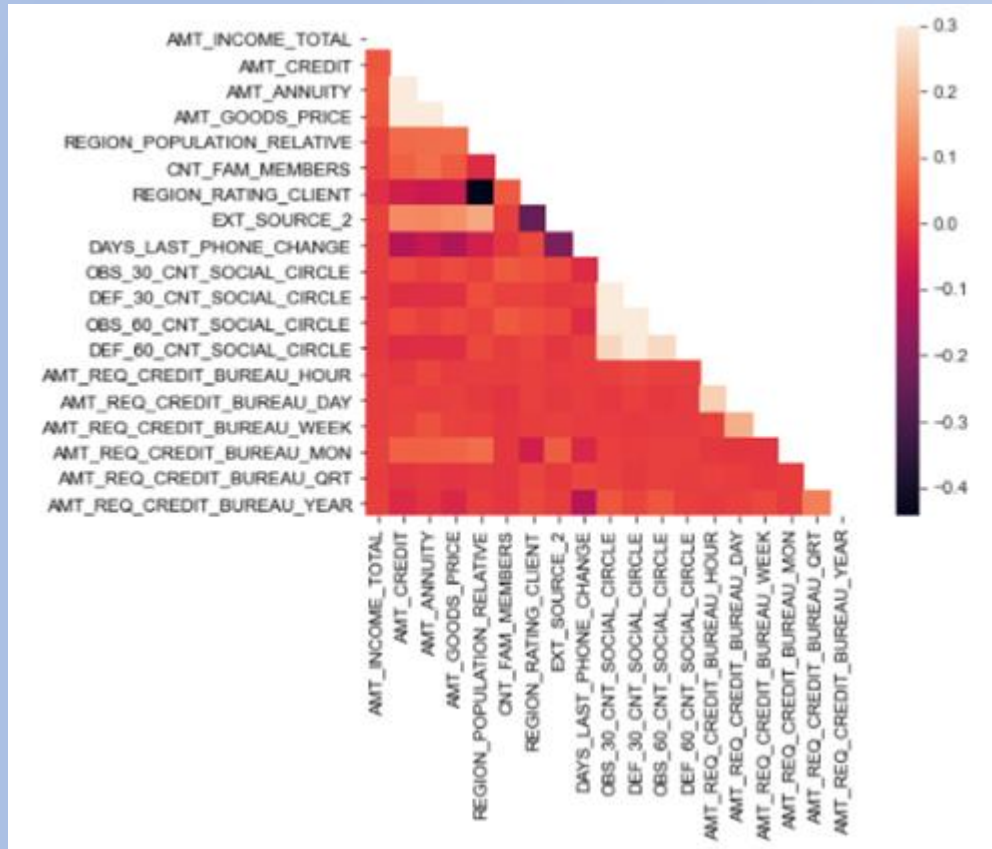
- There is a clear correlation between amount of loan and last phone change, amount annuity and last phone change, amount goods price and last phone change. Although, this does not indicate good or bad, the analyst need to dig deeper to find how and at what cut-off values these are correlated
- There is clear correlation between defaulters and their social surroundings prompting the loan analyst to understand more about this observation
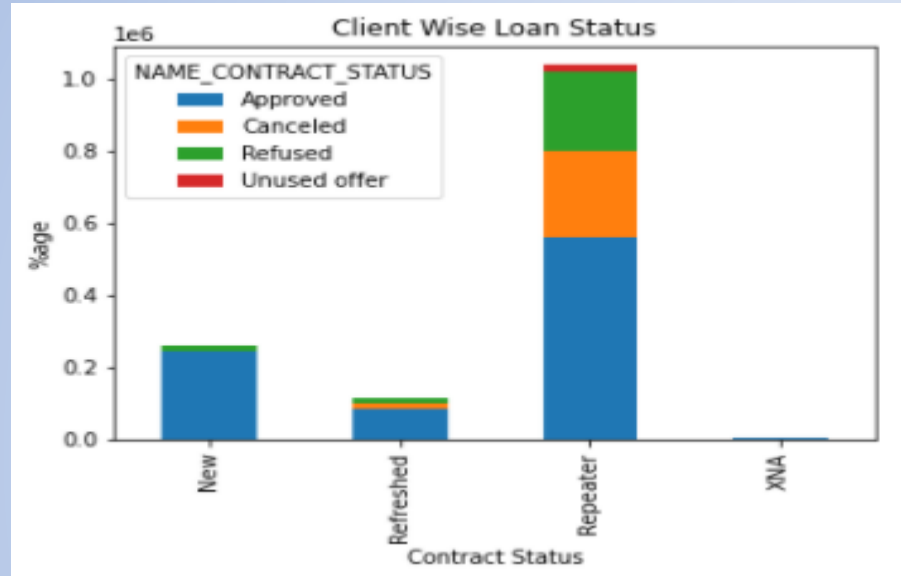
Industry wise Loan application
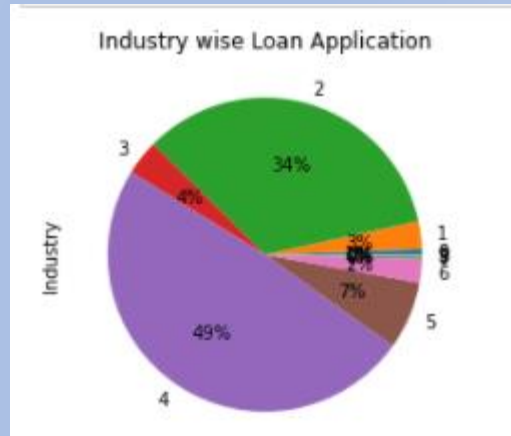
|	NAME_SELLER_INDUSTRY	SK_ID_PREV
4	Consumer electronics	341696
2	Connectivity	238341
5	Furniture	48963
3	Construction	25417
1	Clothing	19896
6	Industry	16567
0	Auto technology	4080
7	Jewelry	2107
8	MLM partners	921
9	Tourism	374

- As can be seen from the above, it is very clear that "Consumer electronics" and "Connectivity" industries are the ones where most of the loans applications are made. Hence, these industries will give most loan business to the bank

Client wise Loan application and status

- Repeater clients have a good number of canceled and Unused offer loans, so the banks loan marketing departments can follow-up with these customers

Client wise Loan application and status
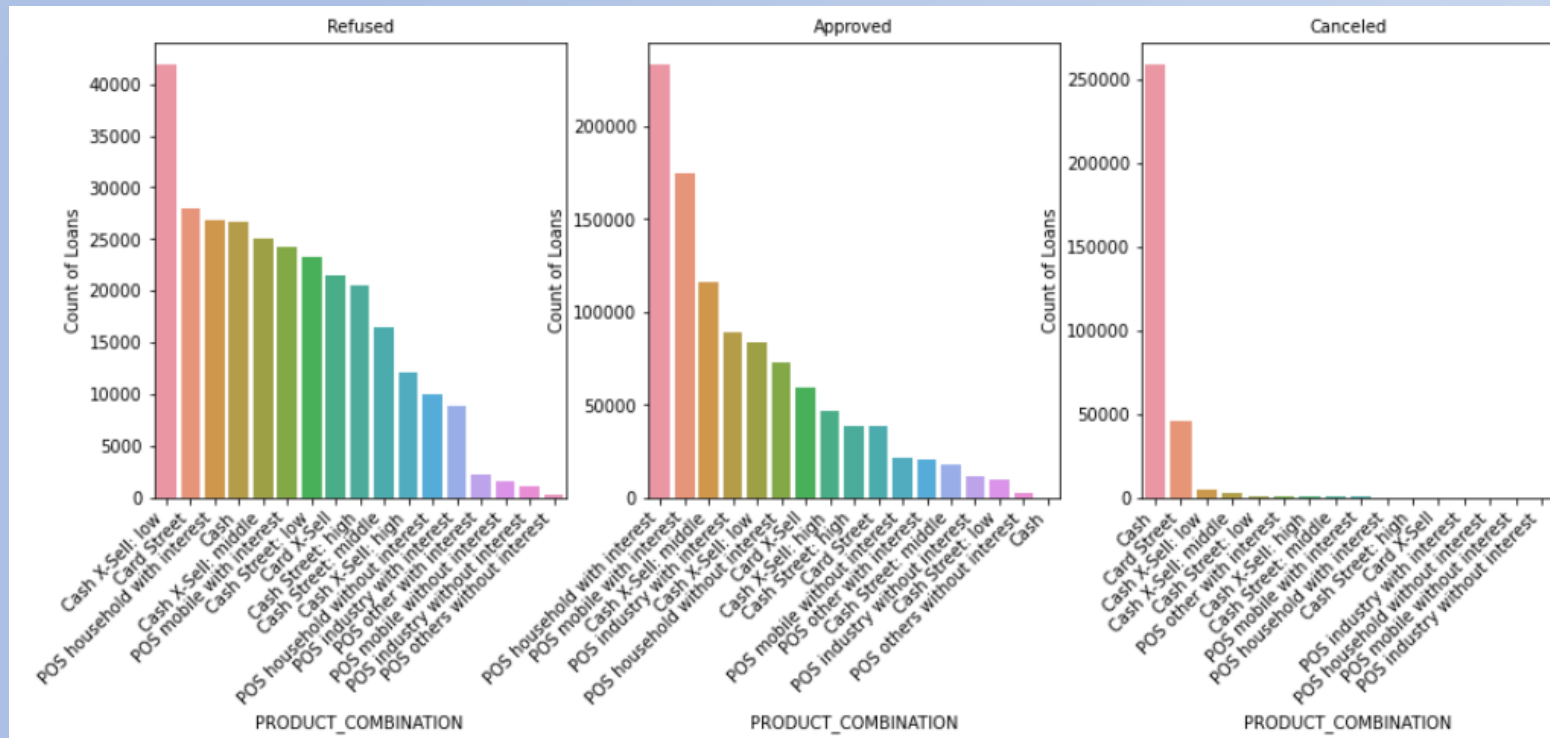
- Most of refused loans were for Product combination of 'Cash X-Sell:low'. It is also clear that most of loans were approved for 'POS household with interest'. And most Canceled loans were for 'Cash loans'