



Loan (Payment/Defaulter) Risk Analysis(EDA)

Introduction



When the company receives a loan application, the company has to decide on loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business for the company.
- 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.



Business Understanding - 1



The loan-providing companies find it hard to give loans to people due to their insufficient or non-existent credit history. Because of that, some consumers use it to their advantage by becoming a defaulter. Suppose you work for a consumer finance company specializing in lending various types of loans to urban customers. You have to use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.



Business Understanding - 2



The data given below contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:

- The client with payment difficulties: he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample.
- All other cases: All other cases when the payment is paid on time.



Note –In target column we have taken assumption as-

- (1) TARGET "0" represents Repayers
- (2) TARGET "1" represents Defaulters

Assumptions



When a client applies for a loan, there are four types of decisions that could be taken by the client/company)

Approved	The Company has approved loan Application.
Cancelled	The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
Refused	The company had rejected the loan (because the client does not meet their requirements etc.
Unused offer	Loan has been cancelled by the client but on different stages of the process. In this case study, we will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

Business Objectives



- 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 (too 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.

- In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

- To develop your understanding of the domain, you are advised to independently research a little about risk analytics - understanding the types of variables and their significance should be enough)





This dataset has 3 files as explained below:

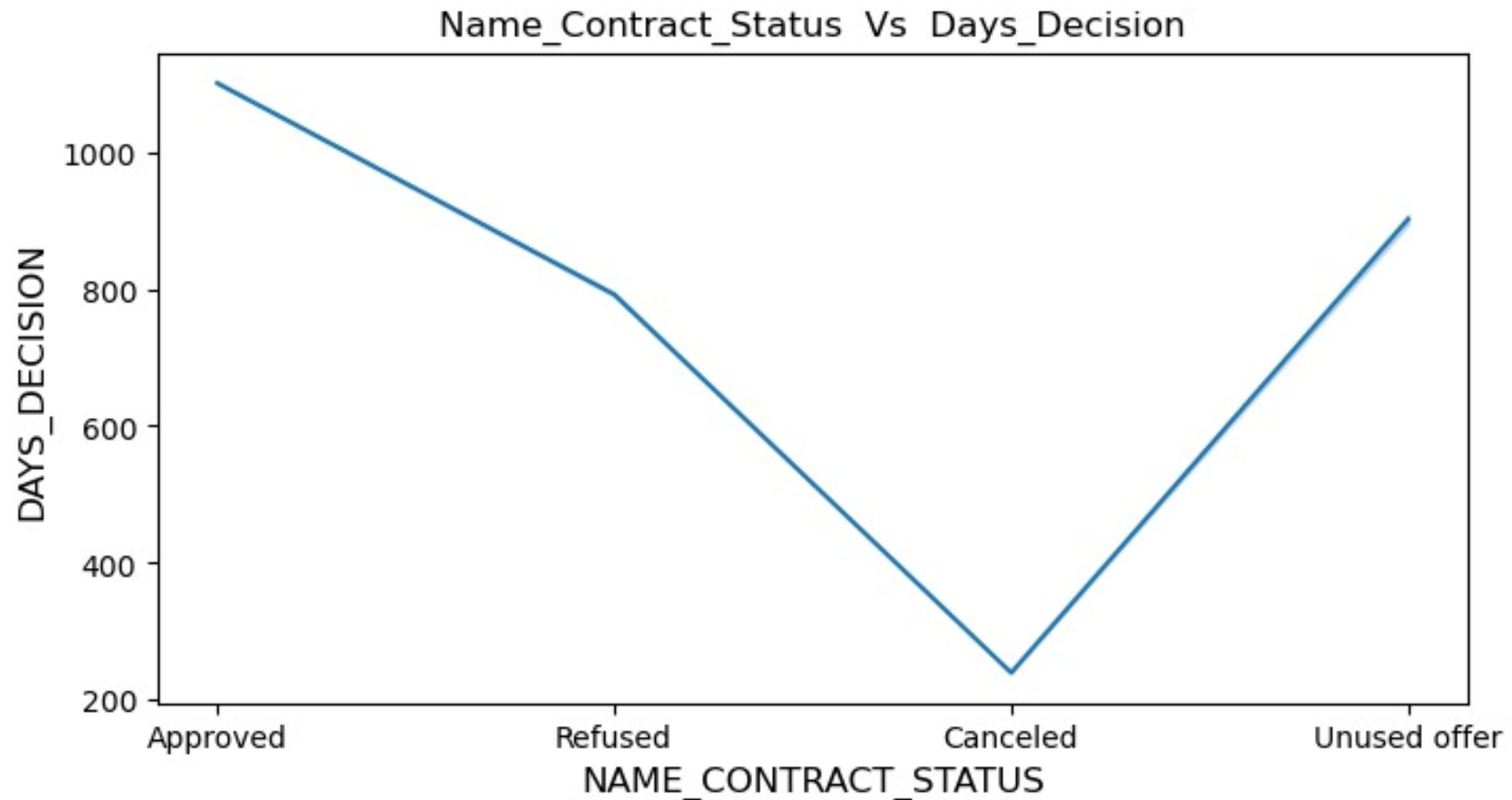
Application_data.csv	It contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.
Previous_application.csv	It contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.
Columns_description.csv	Data dictionary which describes the meaning of the variables.

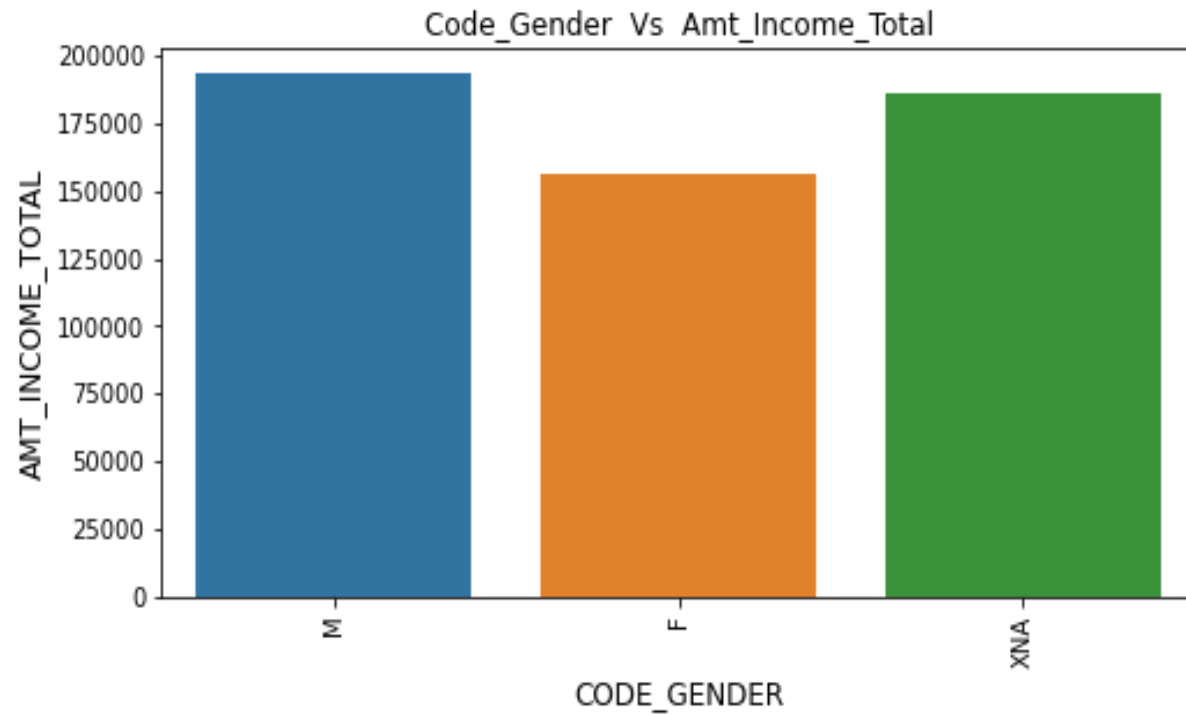


BIVARIATE ANALYSIS



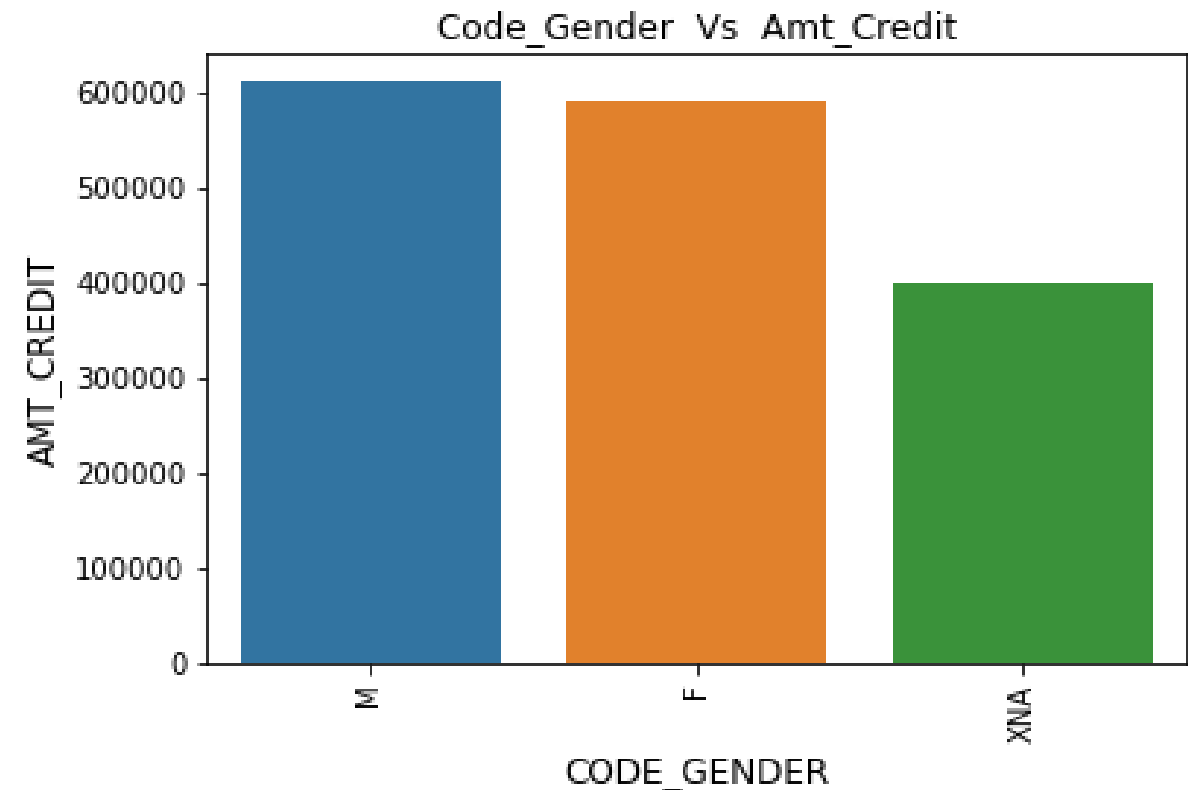
- Bank has taken total more than 1000 days to approve the loan.
- It take almost 900 days to take decision for unused offer.

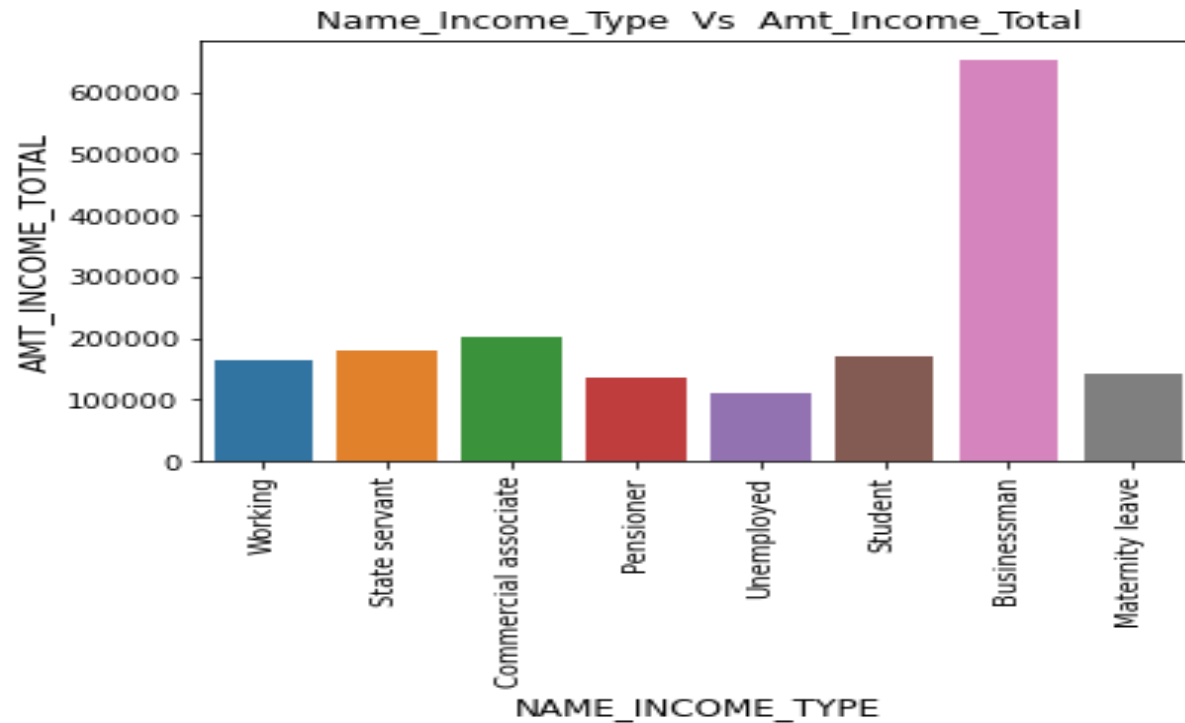




- Female applicant has credited loan almost equal to male applicant.

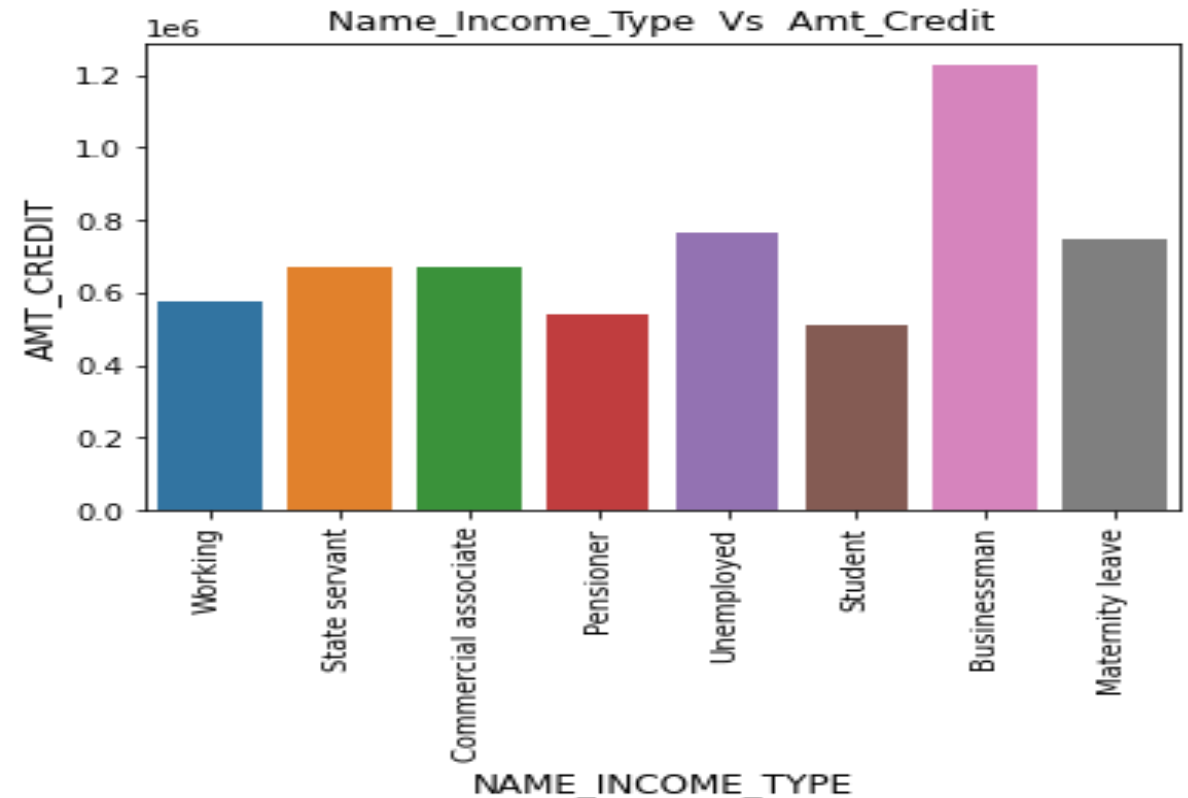
- Male applicant and XNA applicant have higher income than Female

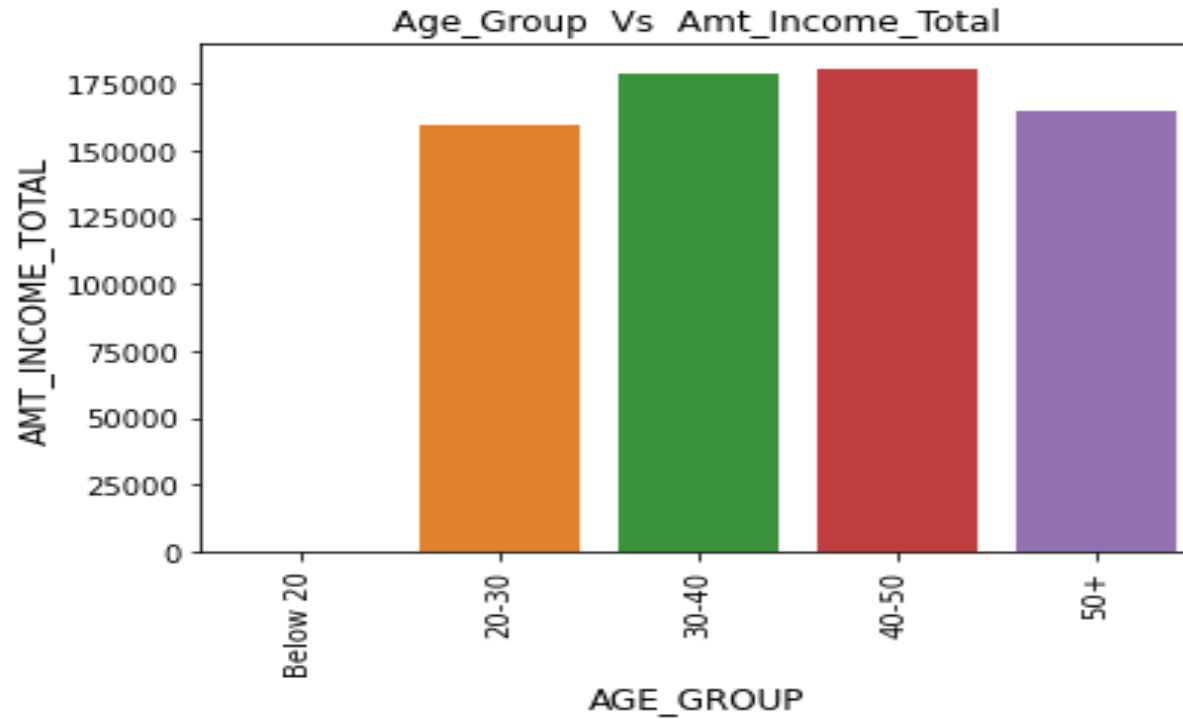




- Businessmen are taking loan 2 times higher than the others.
- Maternity leave and unemployed person have credited loan of almost 0.8m.
- State servant and Commercial associate have credited 0.7m loan amount.

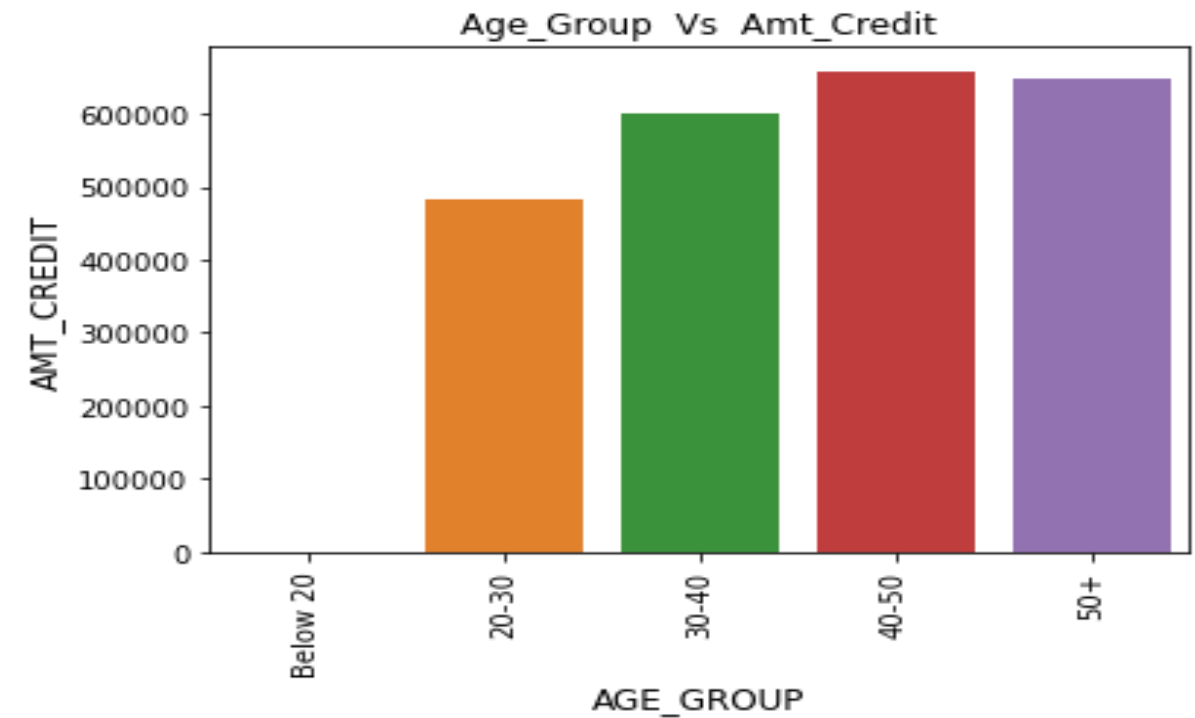
- Income chart depicts that businessman has 3 times higher Income than others.





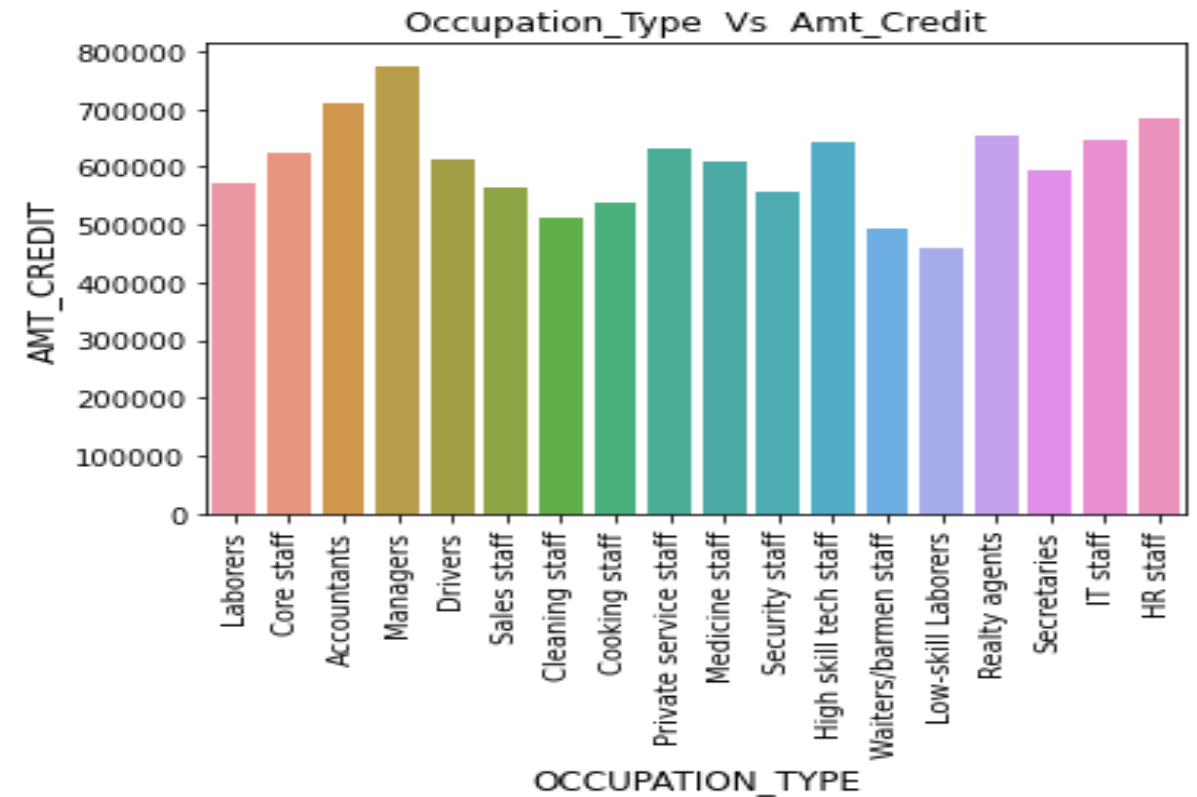
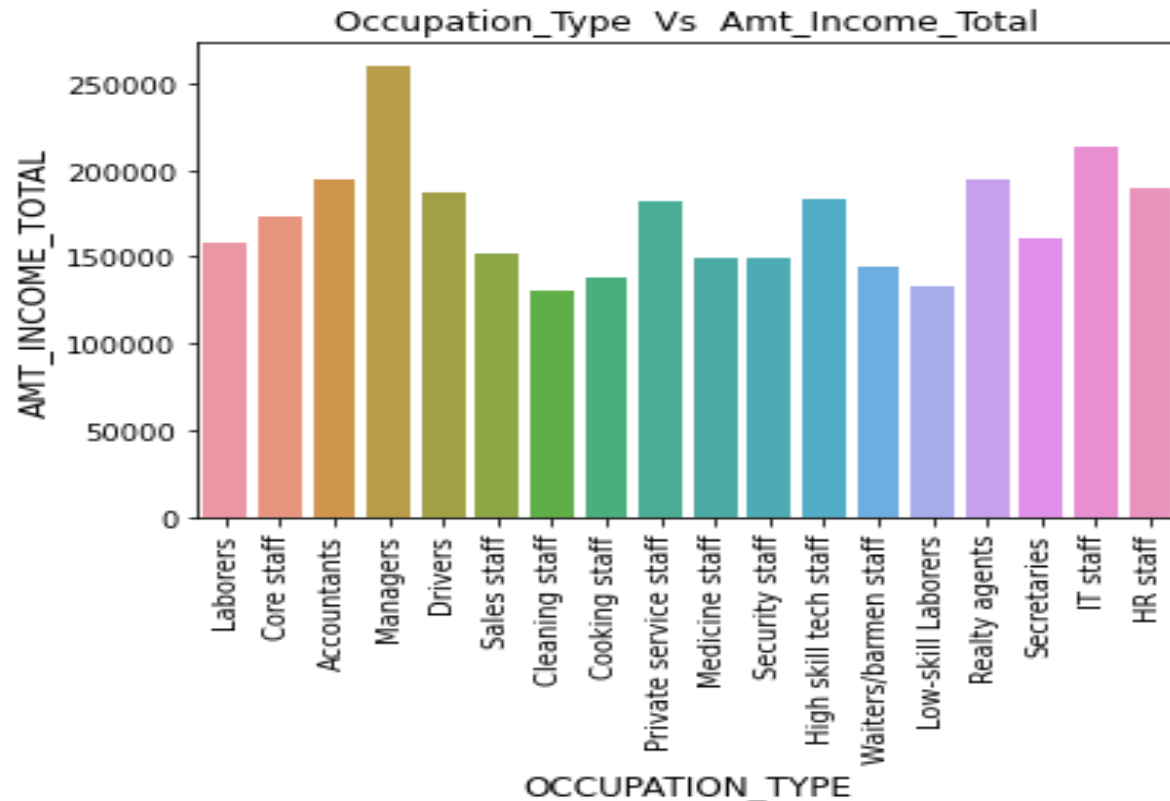
- The 40-50 age and 50+ age group applicant have credited loan more than 6Lacs.

- Applicants between 30-50 age group have income more than 1.75lacs.
- 50+ age group applicant have income more than 1.5lacs.



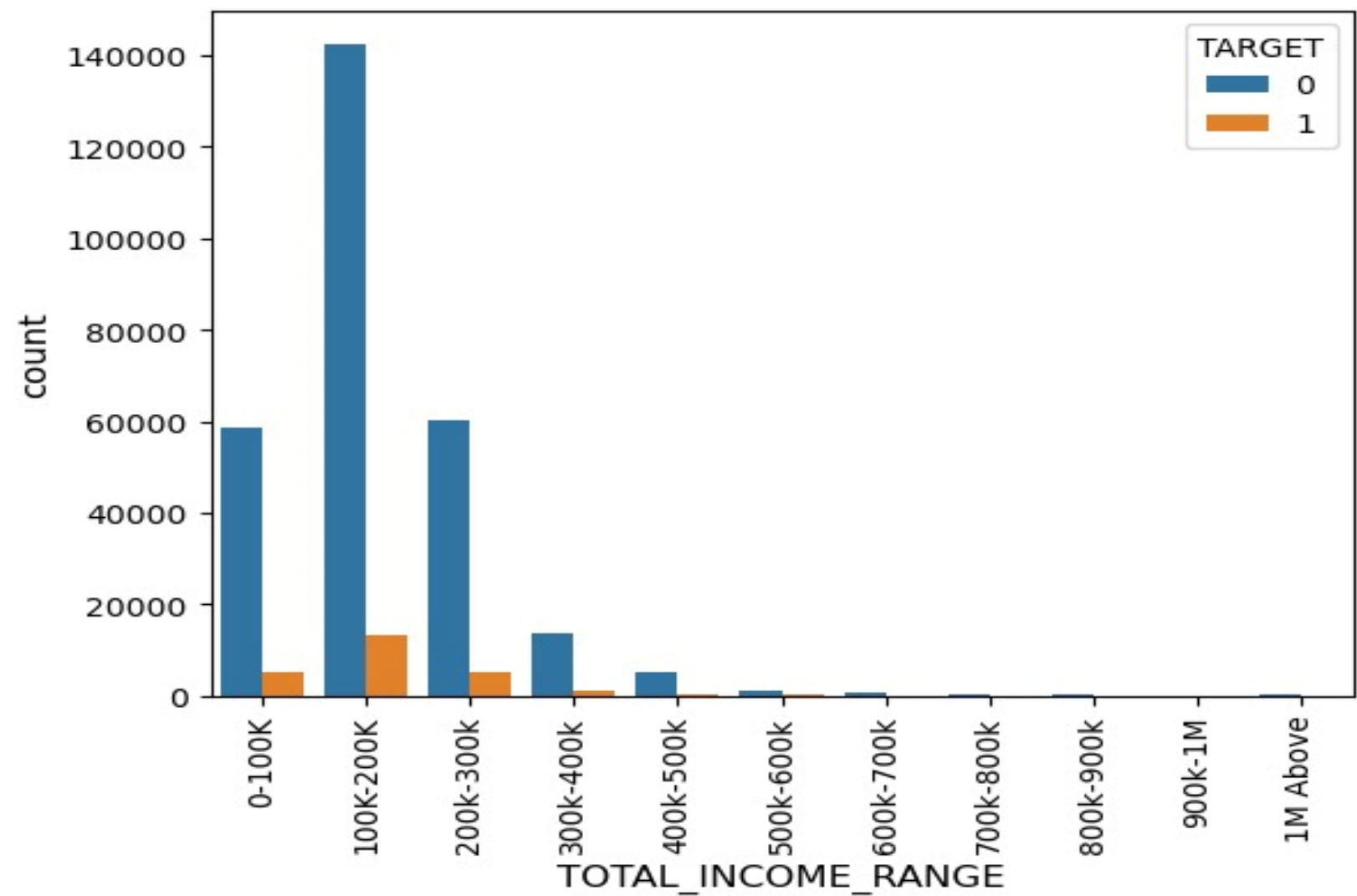


- Managers have highest income of (2.5lacs) followed by IT staff(2.2lakhs), Reality agents and accountants (1.9Lakhs) each.
- As Managers have highest income, they have credited highest amount of loan (almost 8Lakhs) followed by accountants with 7lakhs and HR staff 6.9 lakhs.





- Applicants who have income ranging between 100K-300K are paying EMI on time w.r.t highest income applicants.

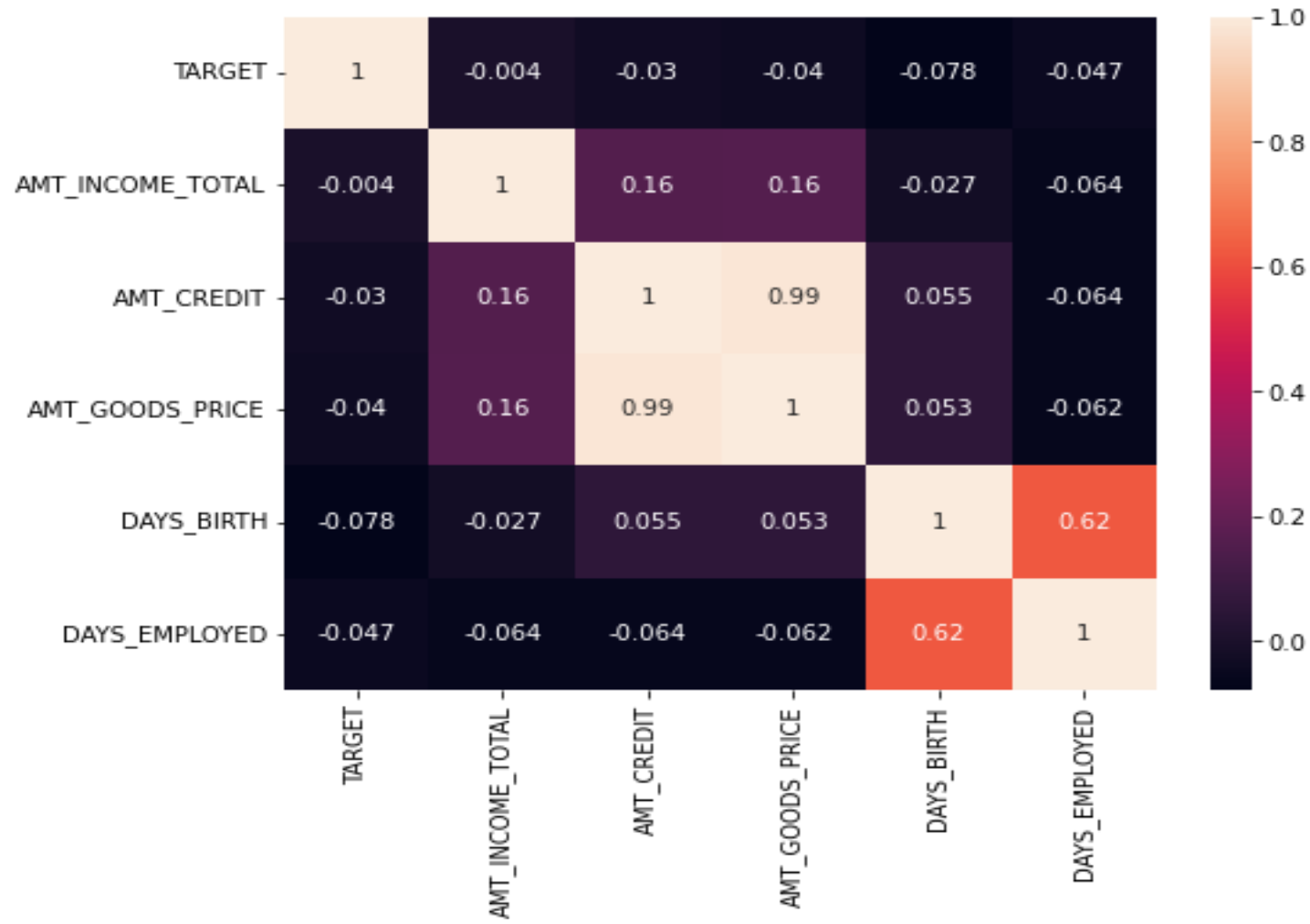




MULTIVARIATE ANALYSIS

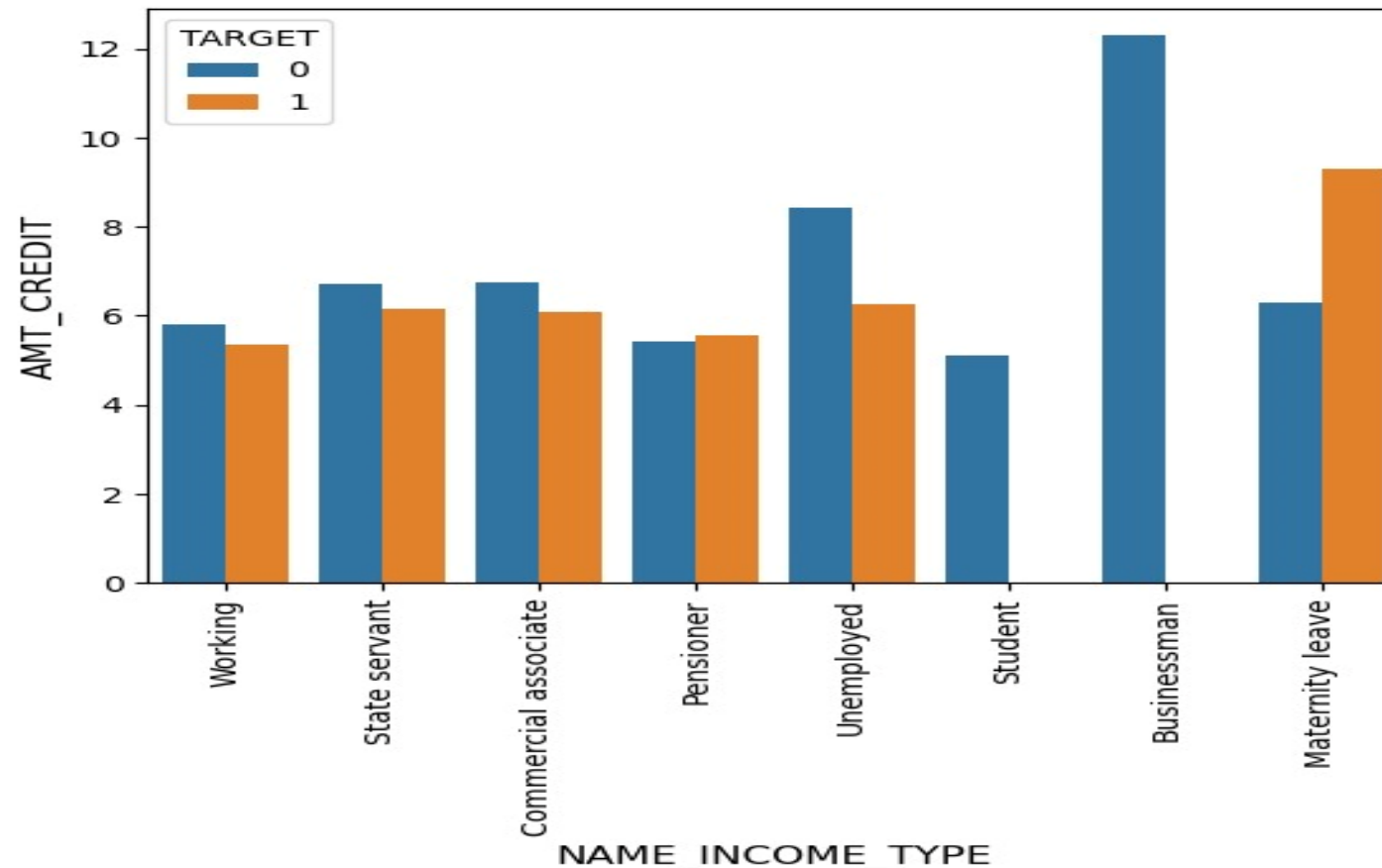


- AMT_CREDIT is highly(0.99) correlated with the AMT_GOODS_PRICE.
- DAYS_BIRTH is directly(0.62) correlated with the DAYS_EMPLOYED.
- AMT_INCOME_TOTAL is positively(0.16) correlated with the AMT_CREDIT.



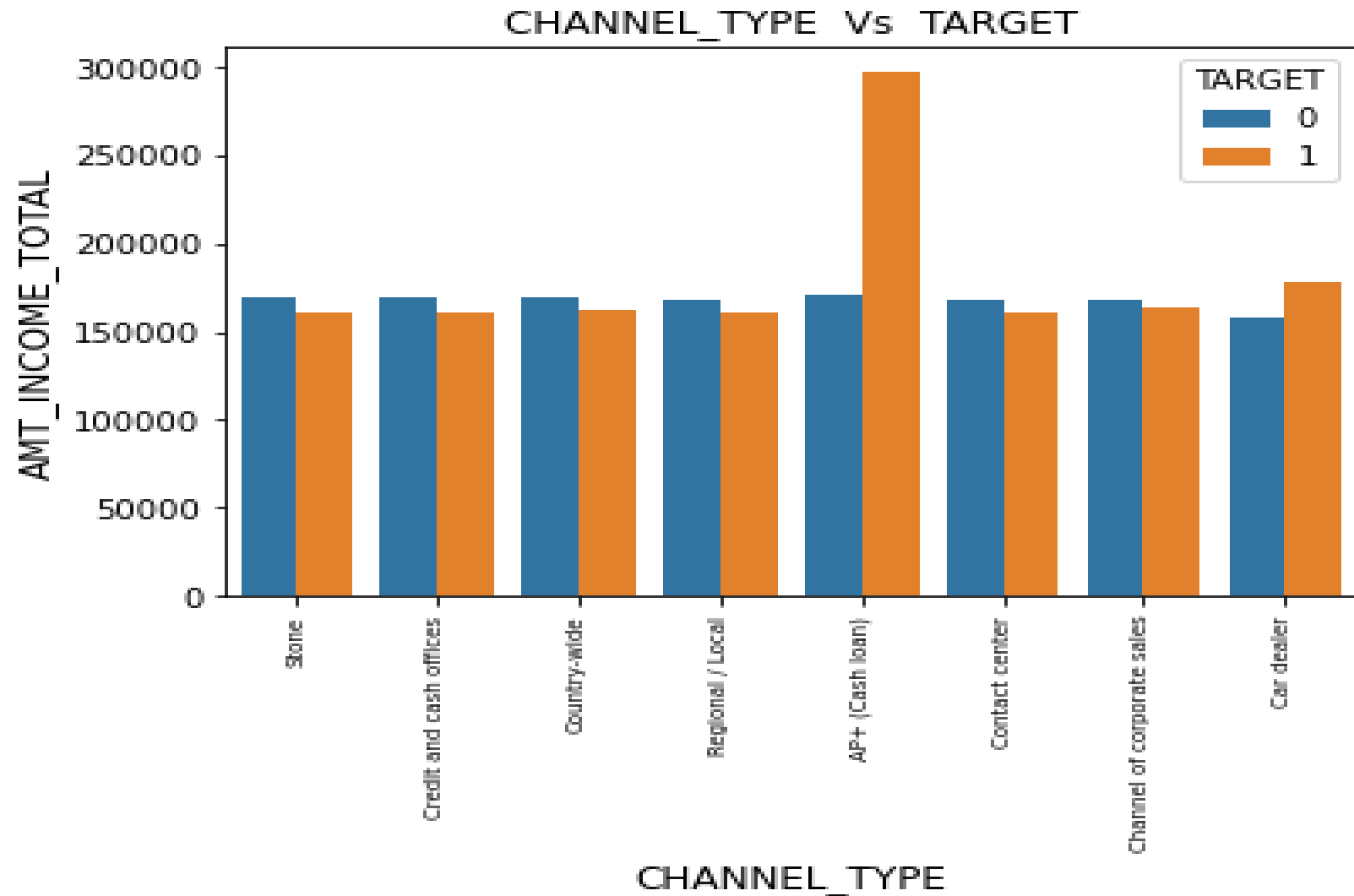


- Maternity leave income type applicants are major defaulter in repayment of loans followed by pensioners.
- Commercial associate, State servants and unemployed applicants have positive payment history (i.e. they are good in repayment).
- Businessman category and students have clear history in repayment which might be possible due to minimal data points.





- AP+(Cash loan) channel witnessed major defaulter followed by Car dealer.
- The stone and Credit and cash offices channel have paid EMIs on time.

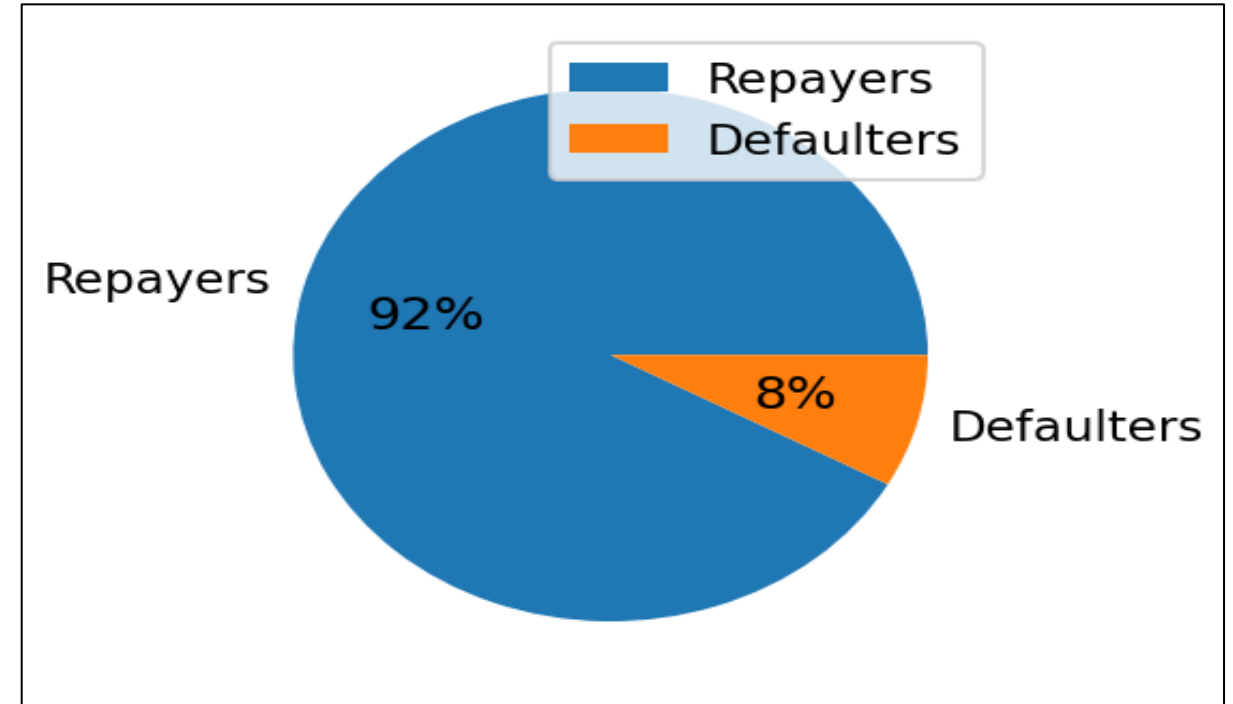
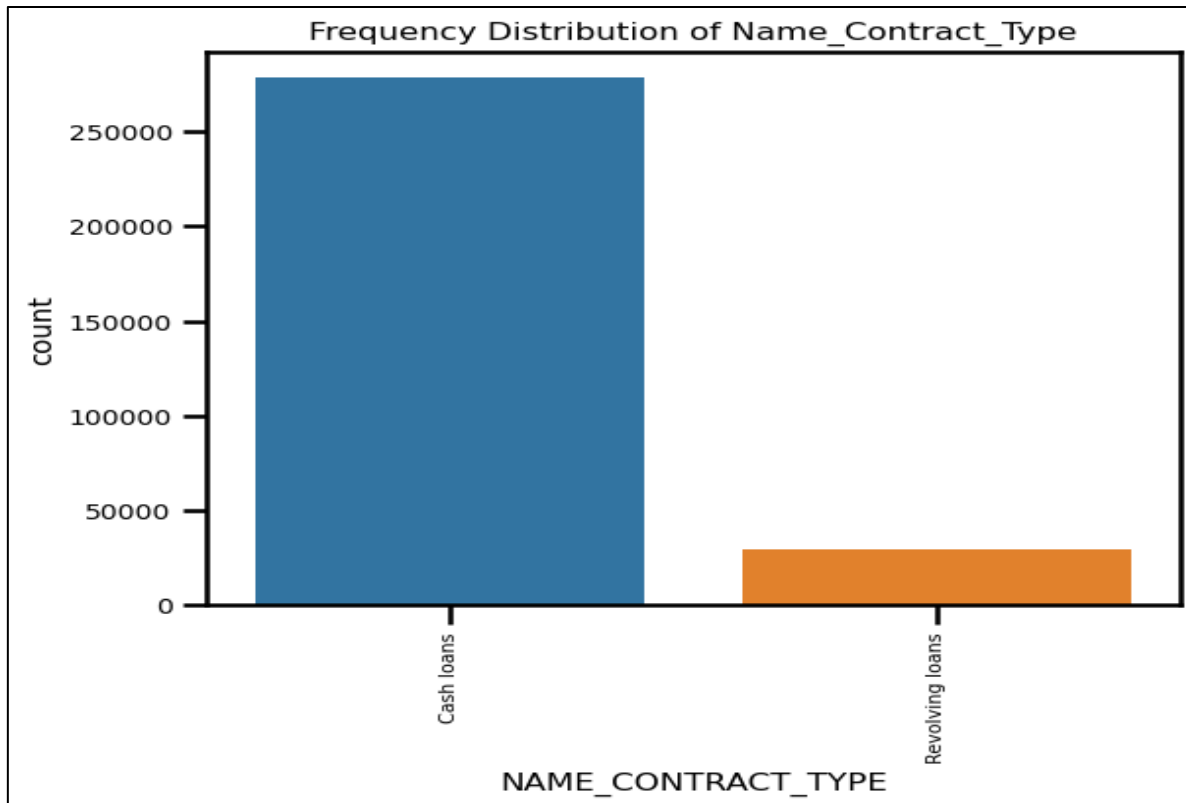




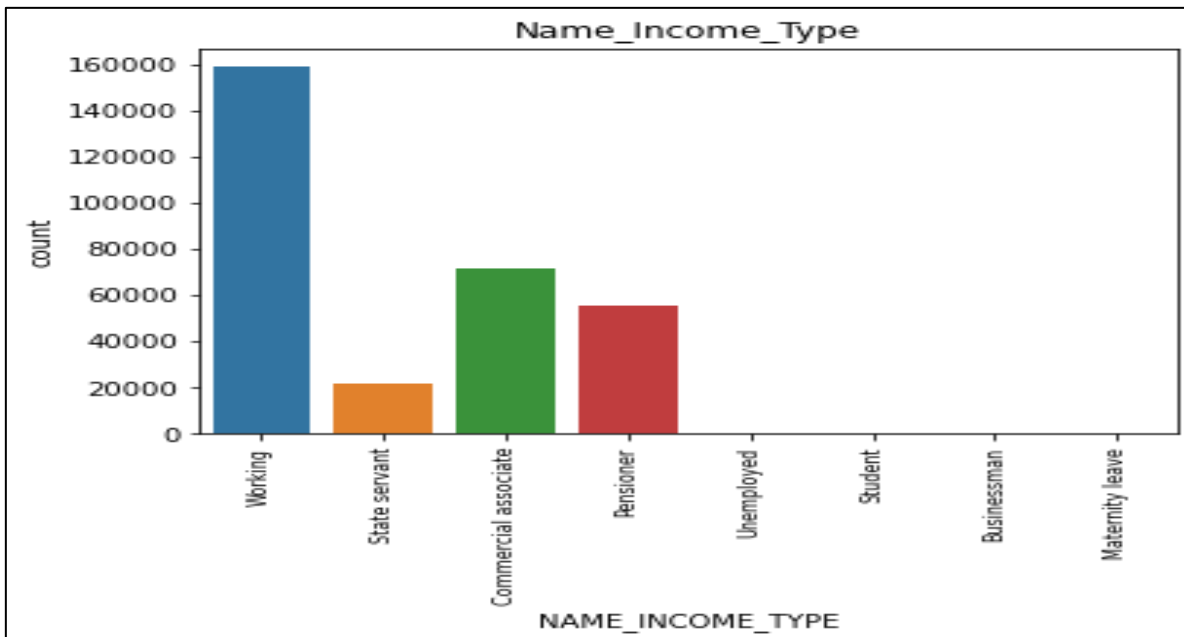
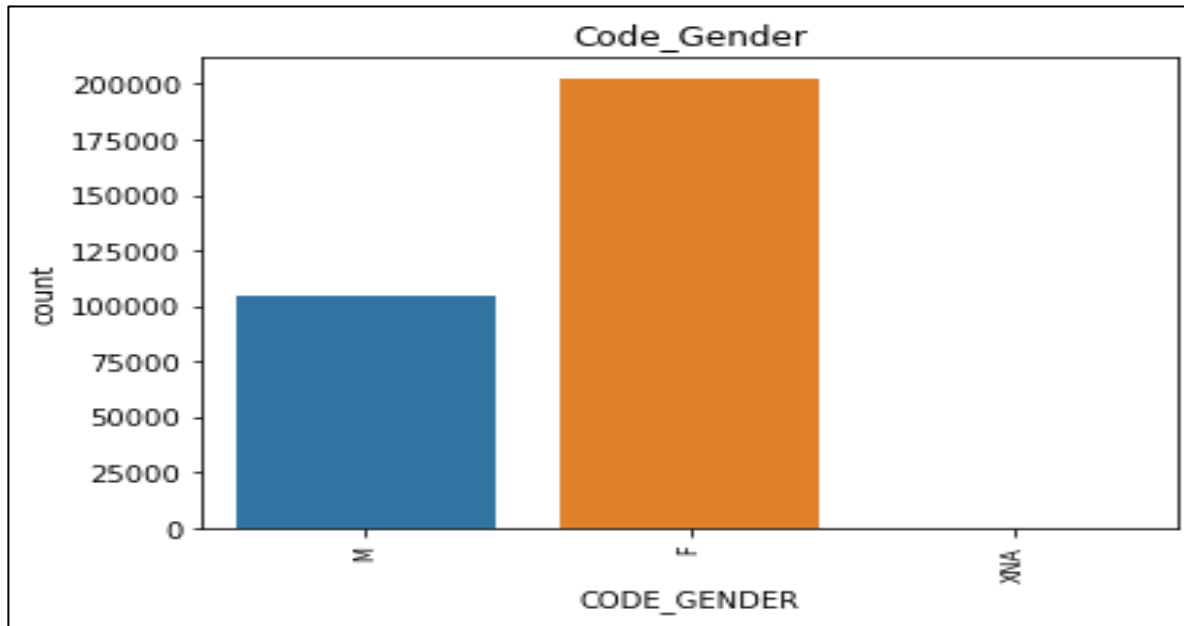
UNIVARIATE ANALYSIS



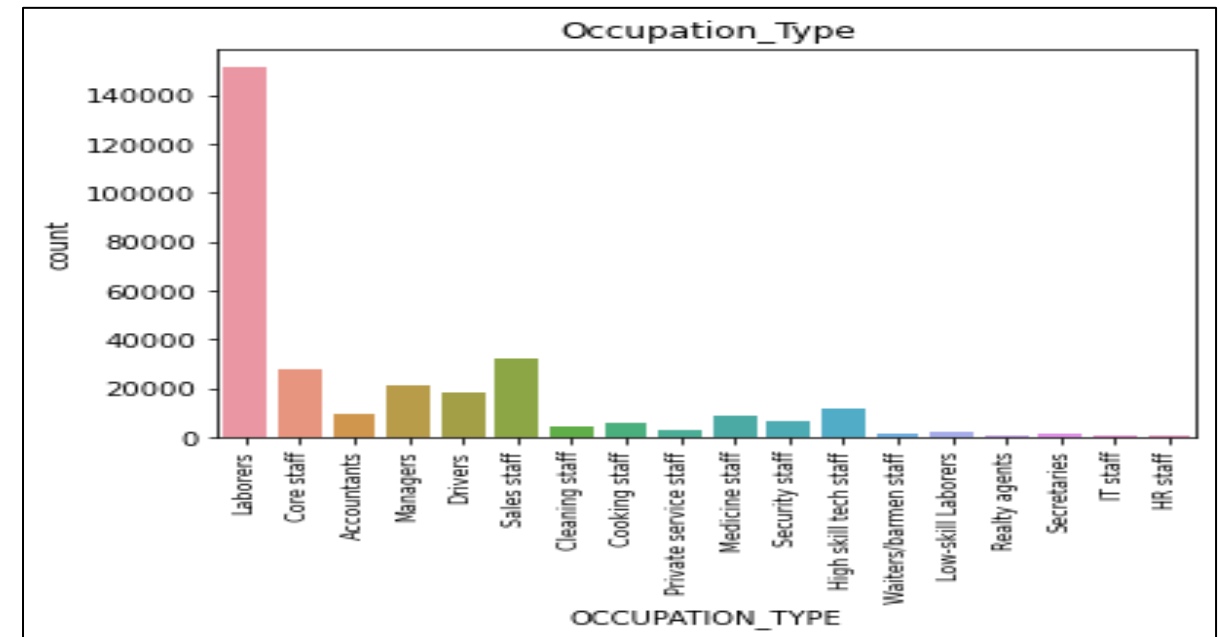
- The number of loan repayers (92%) is higher compared to defaulters(8%).

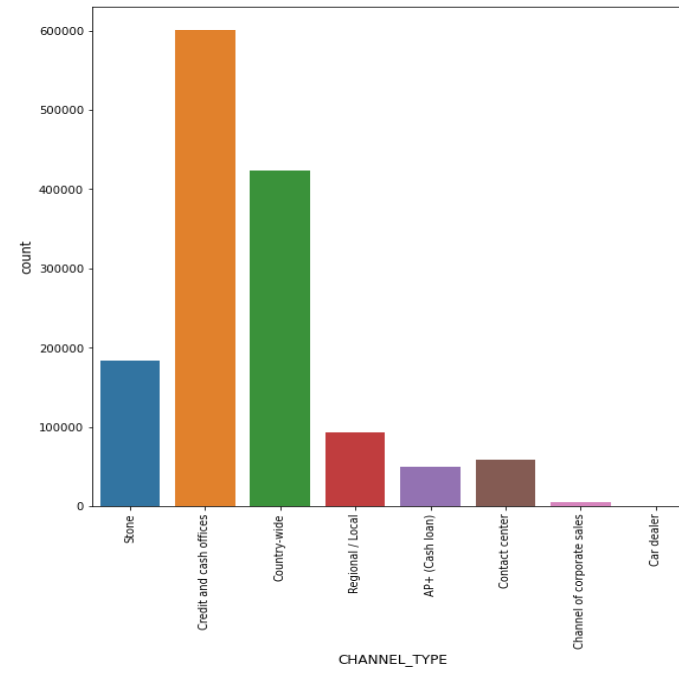
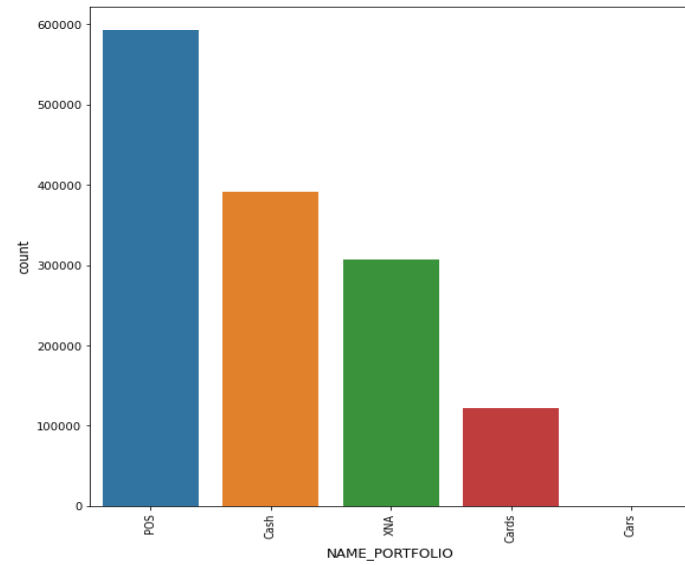
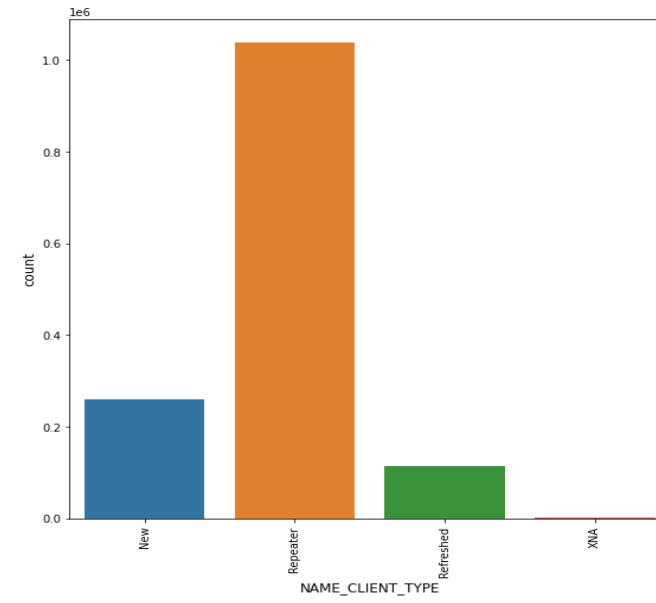
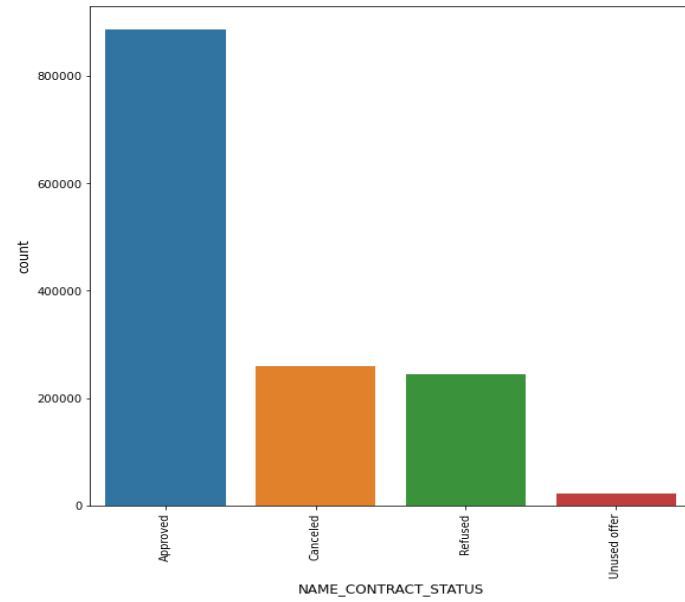


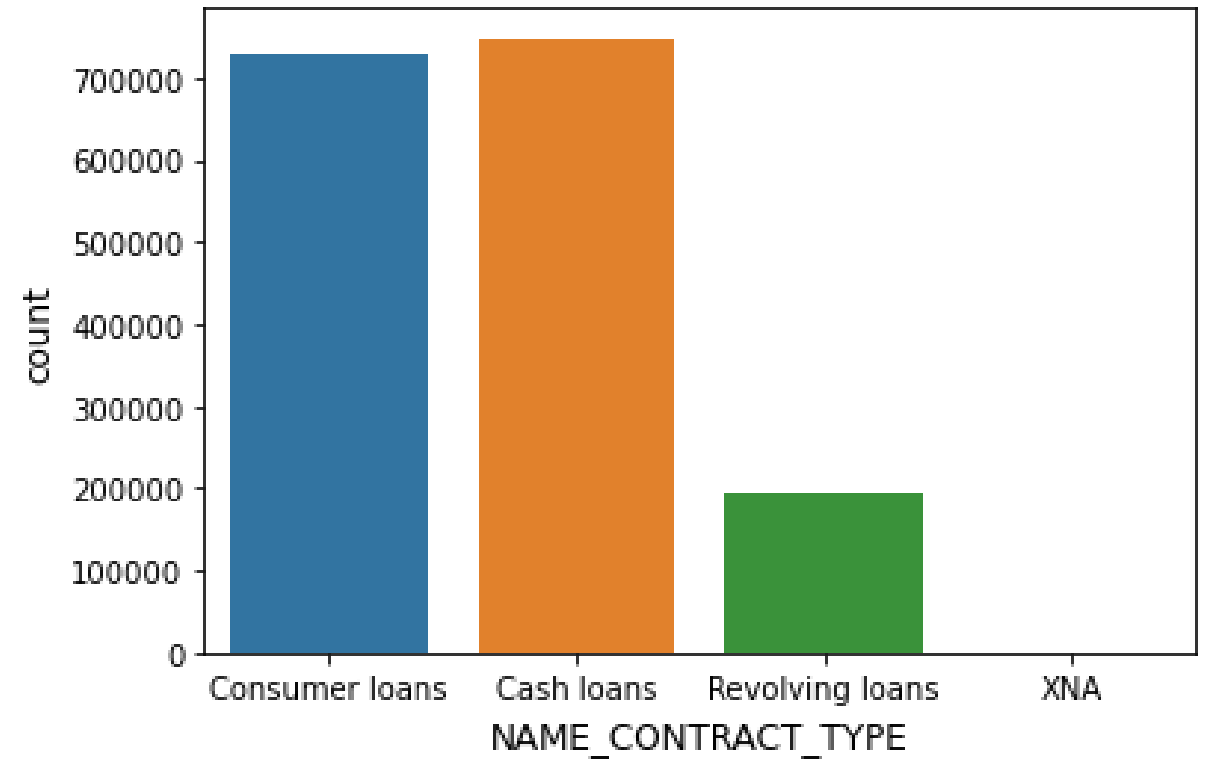
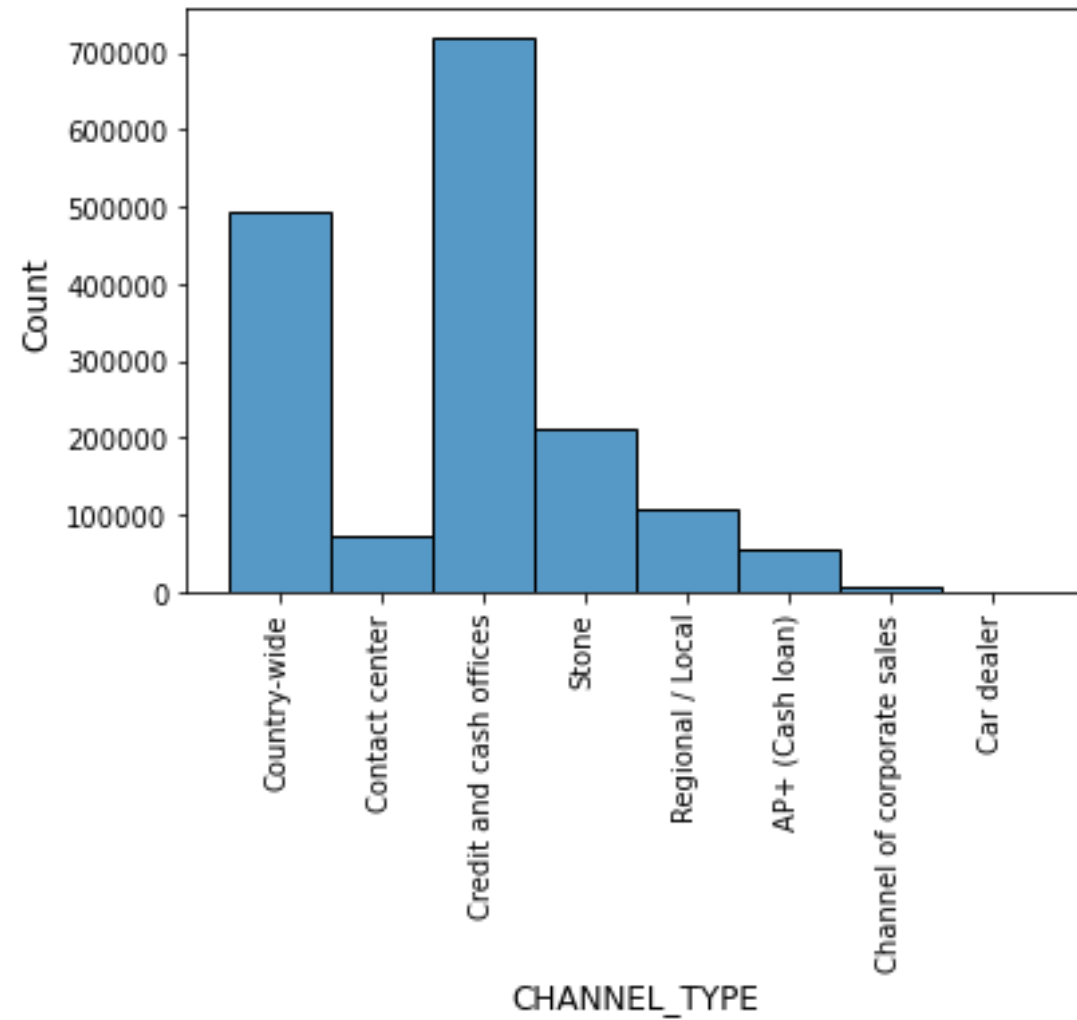
- The demand for cash loan is more than 2.5laks and for revolving loan it is lesser than 50K.



- The number of Female applicants are 2 times more than the number of Male applicants.
- The demand for Loan is higher by Working, followed by Commercial Associate and Pensioners.
- The demand for Loan is higher by Laborers, followed by Sales staff and Core staff.







Conclusion



- Bank should provide loans to the applicant who have income ranging between 100k - 300k as they have Good Payment history.
- Businessman, Commercial associates, State servant and Students have done loan annuity on time, so have good opportunity to provide loan.
- AP+(Cash loan) channel witnessed major defaulter followed by Car dealer, any loan should provide with caution.
- Maternity leaves Applicants have higher defaulter history compare to on time Repayment so considering the defaulter history any loan should provide with caution.



Thank
You