

BANK LOAN CASE STUDY

Project Description:

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.

Business Understanding:

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

When the company receives a loan application, the company has to decide for 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 to 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.

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.

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

1. **Approved:** The company has approved loan application
2. **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.
3. **Refused:** The company had rejected the loan (because the client does not meet their requirements etc.).
4. **Unused Offer:** Loan has been cancelled by the client but on different stages of the process.

Business Objectives:

It 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.

BANK LOAN 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).

Approach:

Identify the missing data and use appropriate method to deal with it. (Remove columns/or replace it with an appropriate value)

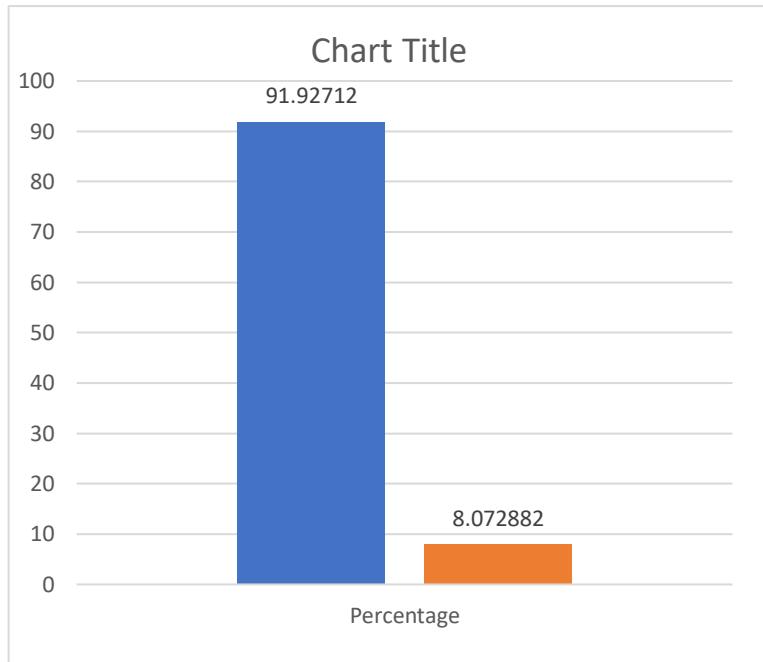
Given that the file contained a large number of columns, it is important to clean and eliminate unnecessary data before analyzing it. The steps involved in cleaning the data include.

- Eliminate columns that do not contribute to the business scenario.
- Remove columns with null values >30%.
- Impute rows with some value for columns with a low percentage of null values.
- Removing outliers from the data set.

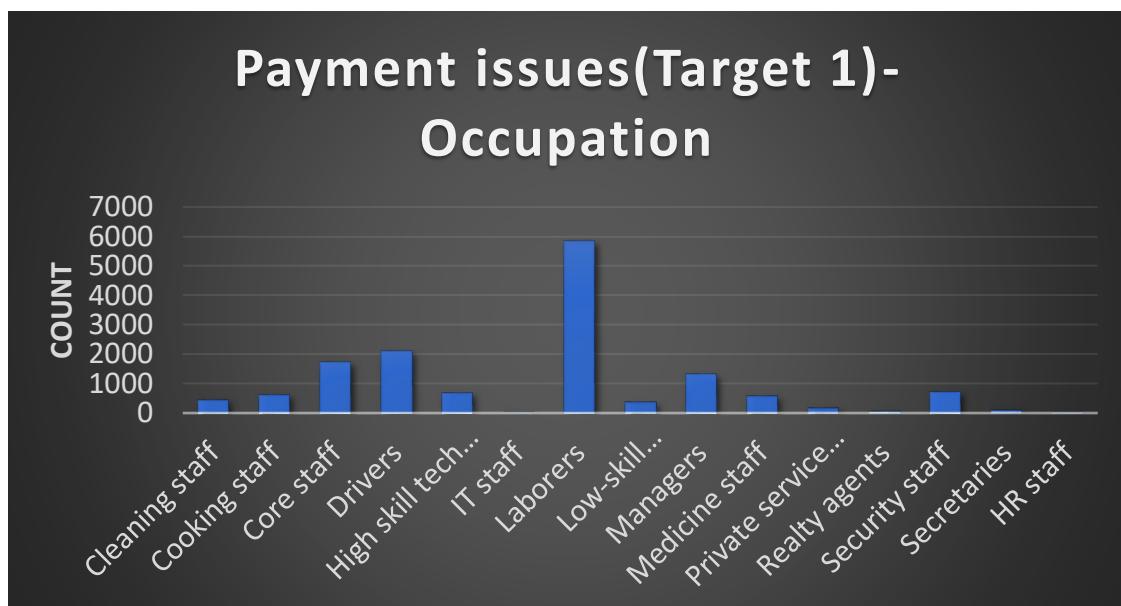
Data Imbalance:

- To identify the imbalance, we must create a client count plot with targets of 1 and 0.
- Target variable (0 - all other situations, which we can name non-default scenarios for ease of use), 1 - client with payment difficulties: he/she had late payment of more than X days on at least one of the first Y installments of the loan in our sample.
- The proportions of 0 and 1 are, respectively, 91.92 and 8.07 This data collection is quite unbalanced.
- Analysis: We will produce several data sets with target values of 0 and 1 and then determine the correlation between various variables for both data frames with target values of 1 and 0.

BANK LOAN CASE STUDY

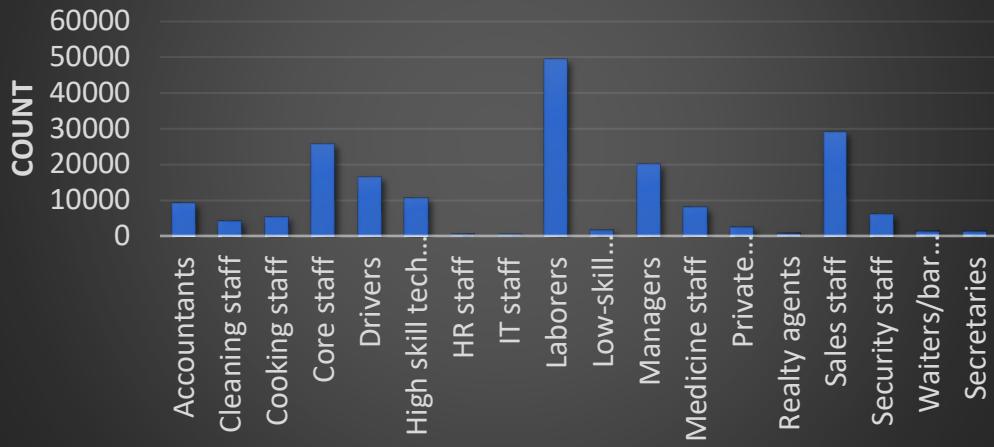


Analysis with the basis of occupation details with respect to Target 0 and Target 1:

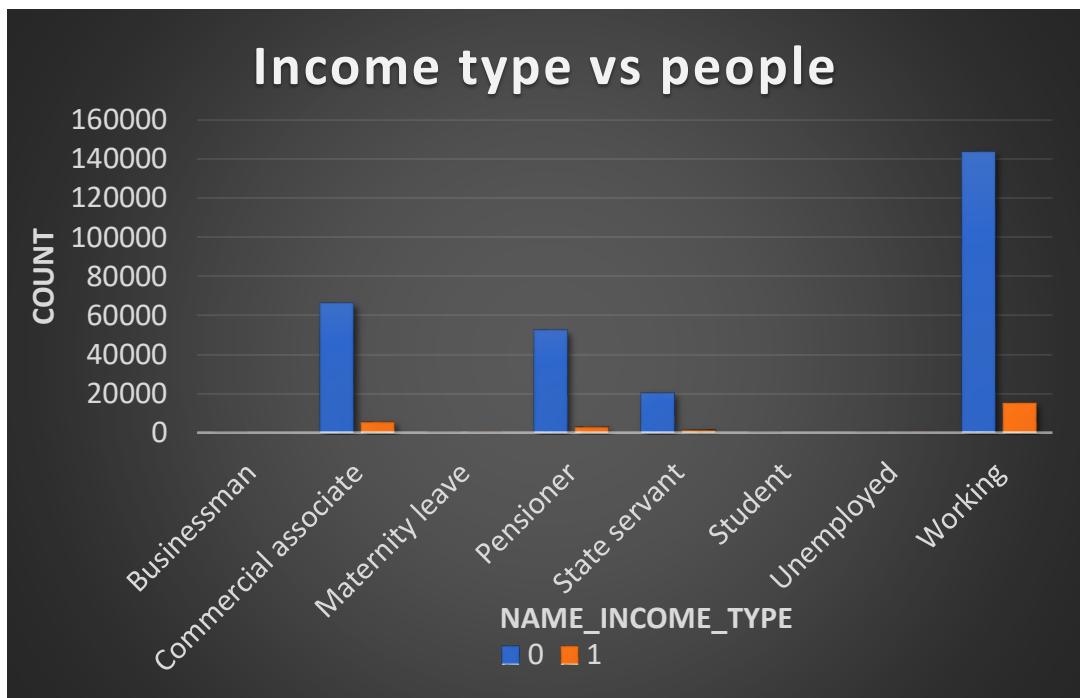


BANK LOAN CASE STUDY

No payment issues(Target 0)-Occupation



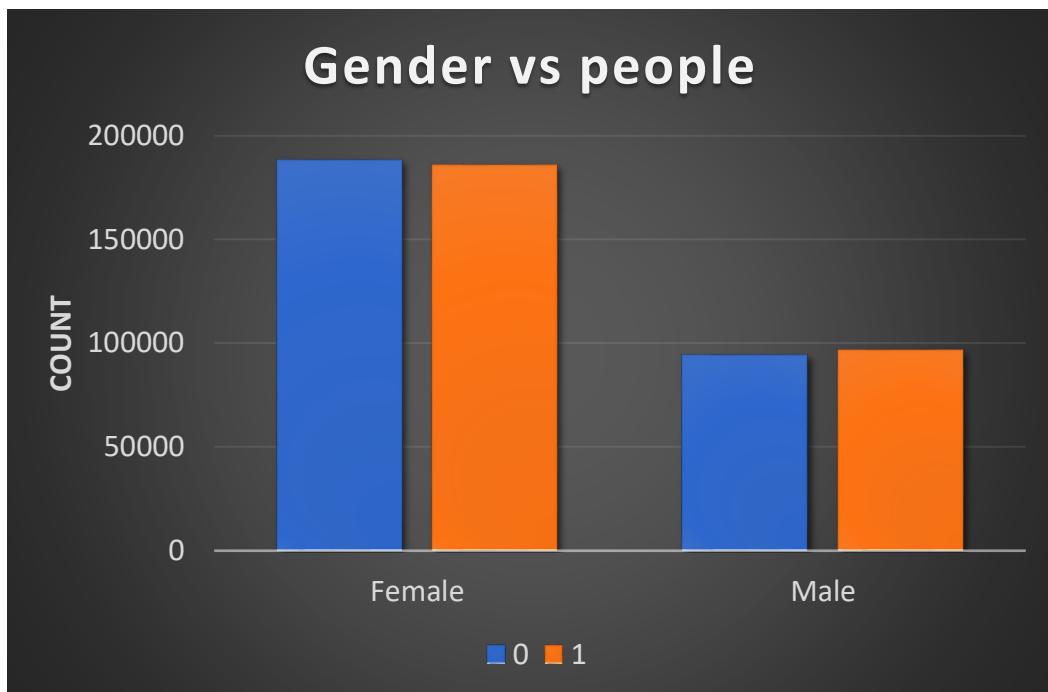
Analysis on the basis of income type of people with respect to Target 0 and Target 1:



The given graph representing people who are working can be targeted to lend loans because they don't have payment issues.

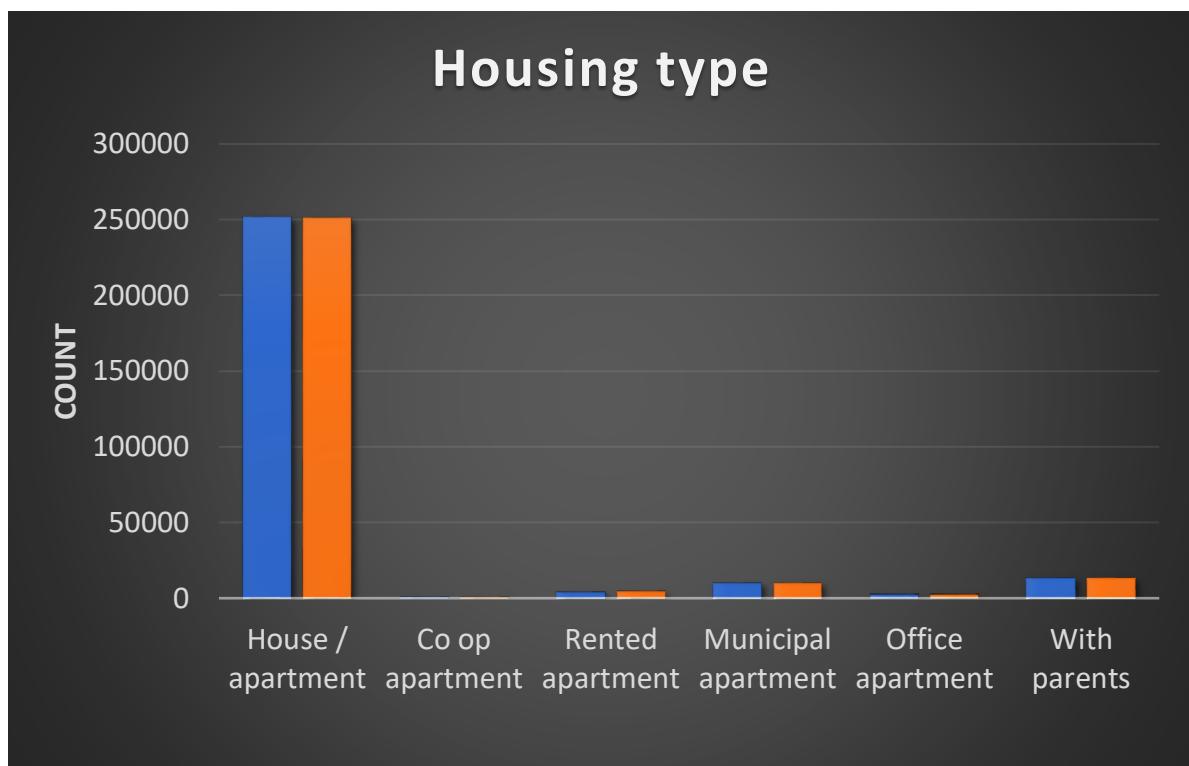
BANK LOAN CASE STUDY

Analysis on the basis of gender type of people with respect to target:



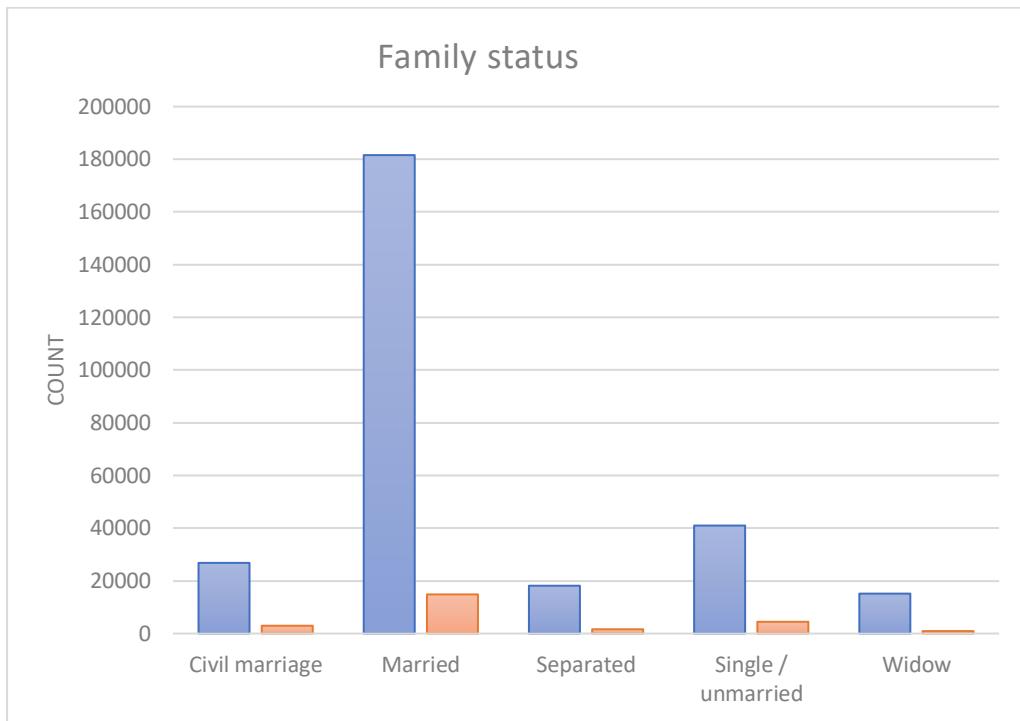
From the above graph, we can see female borrowers make timely loan payments therefore, the bank should consider females first for lending money.

Analysis on the basis of housing type with respect to target:

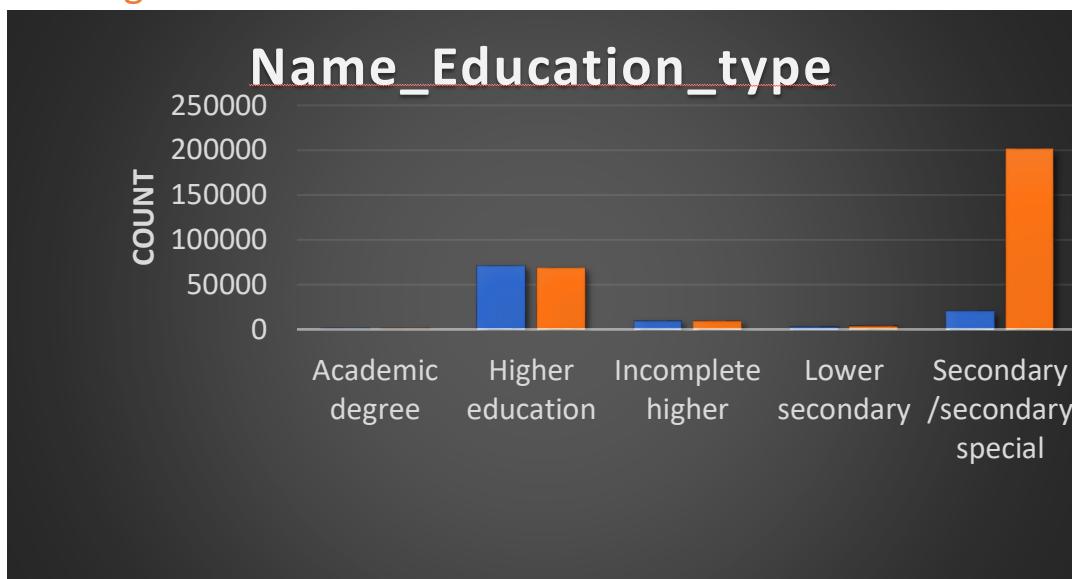


BANK LOAN CASE STUDY

Analysis on the basis of family status with respect to target:



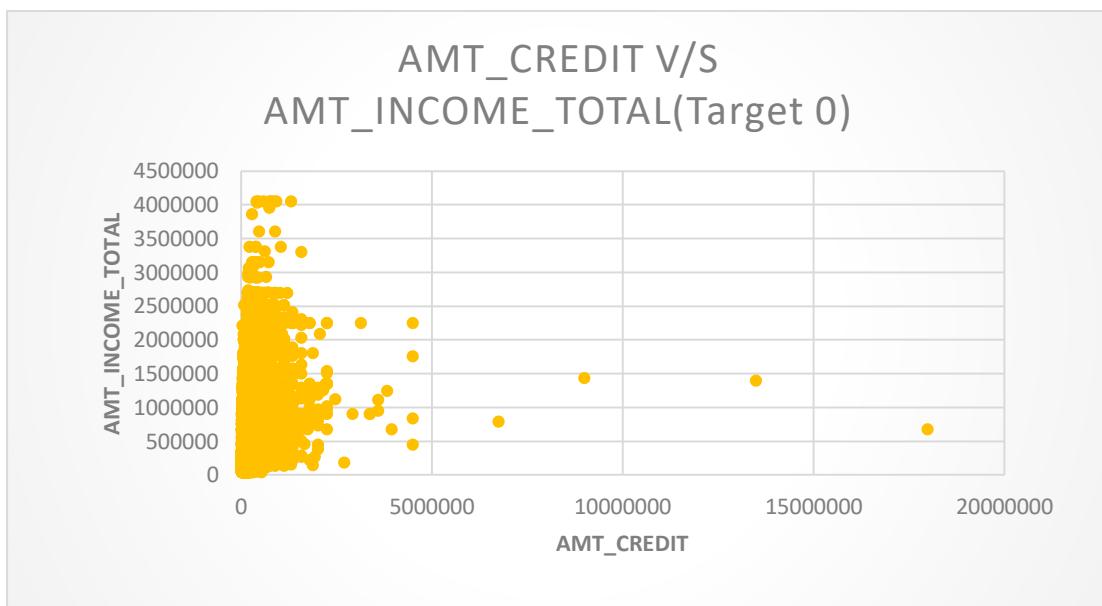
Analysis on the basis of the education type of people with respect to the target:



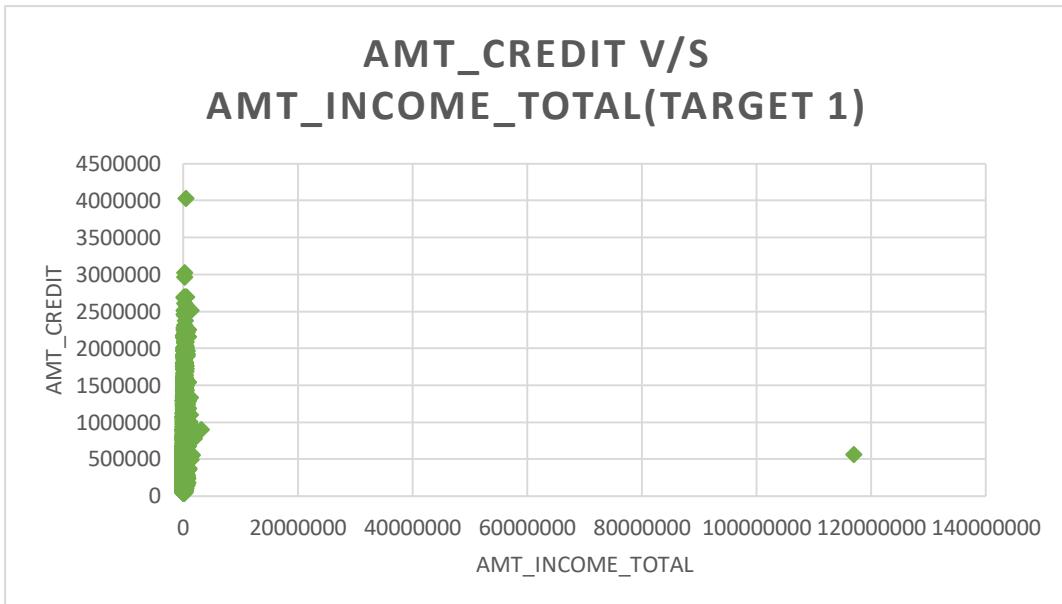
BANK LOAN CASE STUDY

"We can analyze that the bank should target married people as they have a higher likelihood of paying back loans on time, in addition, clients with houses or apartments also tend to pay back loans on time, therefore, the bank should also target them for lending opportunities. Furthermore, our analysis of the graph shows that people with secondary education also have a higher likelihood of timely loan repayment, making them a valuable target group for the bank to lend money to."

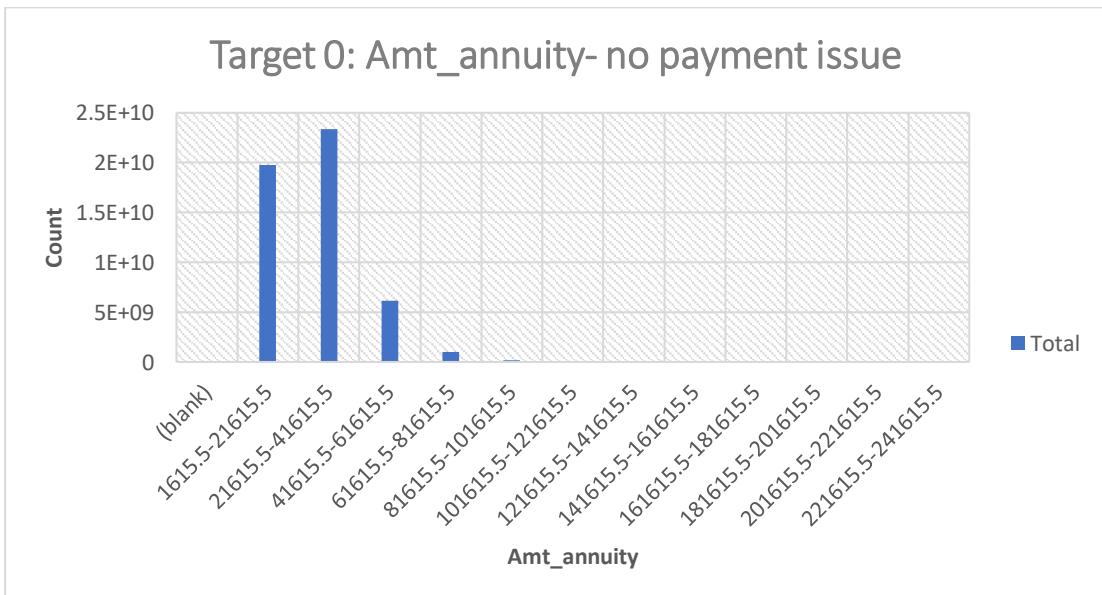
AMT_CREDIT V/S AMT_INCOME_TOTAL



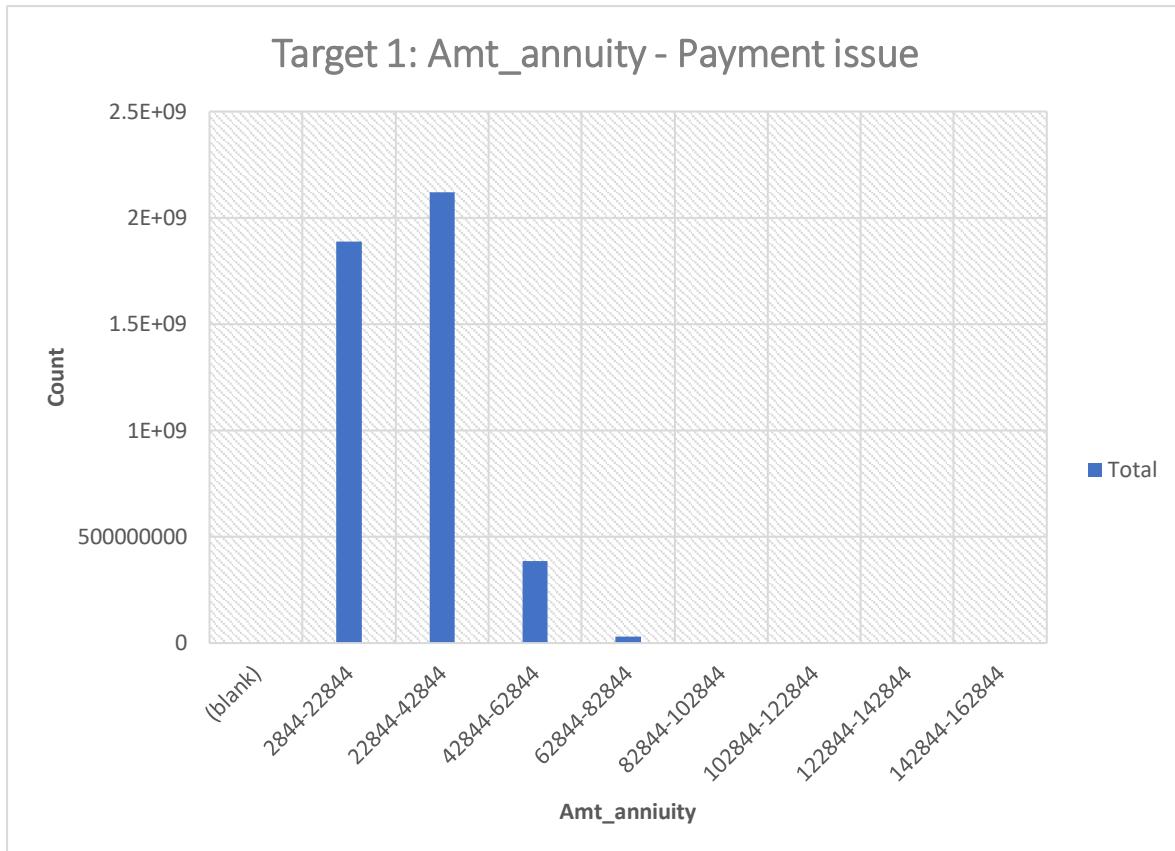
BANK LOAN CASE STUDY



"From our observations, we can see that individuals who consistently pay back their loans on time are offered higher credit limits by the bank, as opposed to those who have a history of payment issues."



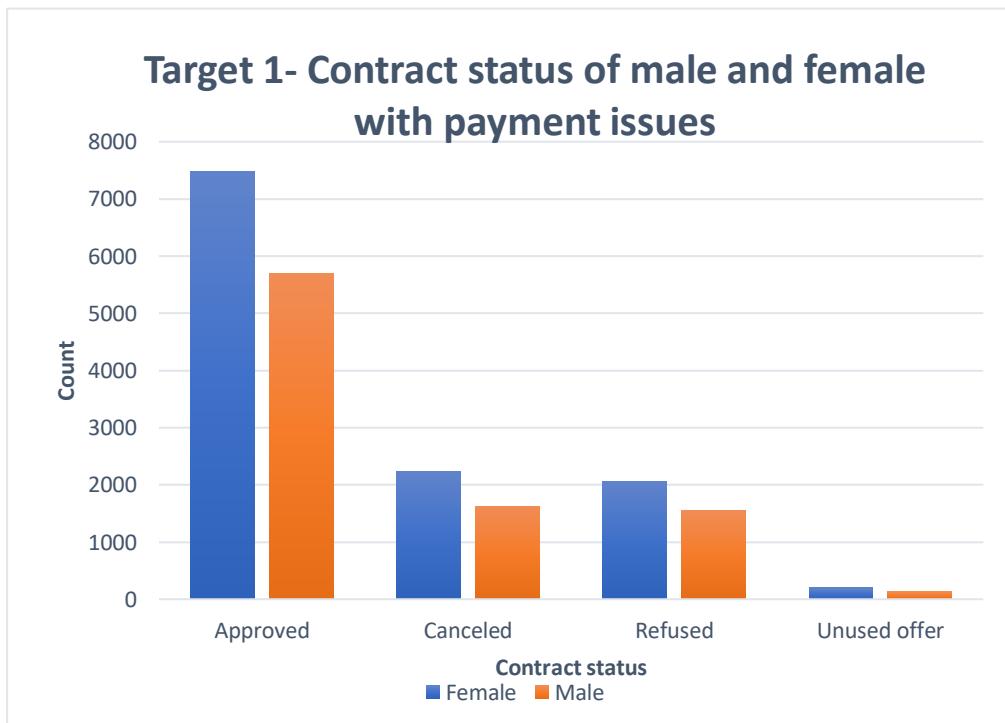
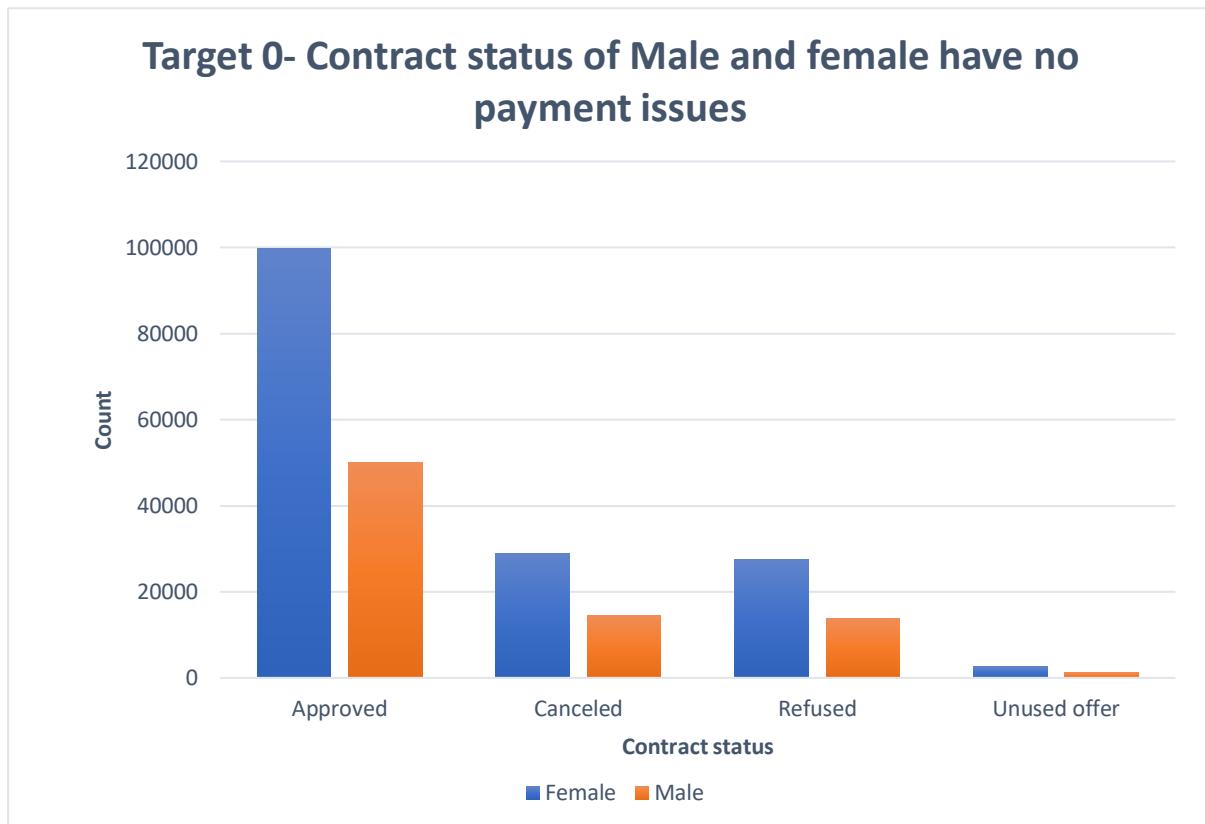
BANK LOAN CASE STUDY



"The graph above indicates that the group of individuals with loan amounts between 22844 and 42844 have a higher likelihood of experiencing difficulty making timely payments."

BANK LOAN CASE STUDY

The contract status of males and females with respect to the target:



BANK LOAN CASE STUDY

"By examining the graphs above, we can see that historically, a majority of female clients with no payment issues had their loan contracts approved. Therefore, the bank should focus on targeting female clients with approved contract status as a prime group for lending loans currently."

Top 10 correlation:

	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10	Column 11
Column 1	1										
Column 2	0.038131	1									
Column 3	0.046421	0.752195	1								
Column 4	0.037583	0.983103	0.752699	1							
Column 5	0.003096	-0.13532	-0.0143	-0.13581	1						
Column 6	-0.01498	-0.00097	-0.08255	0.003587	-0.5751	1					
Column 7	-0.12686	-0.06594	-0.08152	-0.07806	-0.00174	0.023443	1				
Column 8	0.006654	0.051224	0.075711	0.047388	0.203267	-0.18656	-0.02736	1			
Column 9	0.004796	-0.00167	0.031257	-0.00811	0.259109	-0.19194	0.006231	0.885484	1		
Column 10	-0.00421	-0.05233	-0.01677	-0.05609	0.252863	-0.22647	0.017019	-0.03178	-0.0323	1	
Column 11	0.000158	-0.02585	0.034279	-0.02568	0.289114	-0.18893	-0.02401	0.145828	0.149154	0.096833	1

Column1:AMT_INCOME_TOTAL

Column 2: AMT_CREDIT

Column 3: AMT_ANNUITY

Column 4:AMT_GOODS_PRICE

Column 5: DAYS_BIRTH

Column 6: DAYS_EMPLOYED

Column 7: OWN_CAR_AGE

Column 8:CNT_FAM_MEMBERS

Column 9: CNT_CHILDREN

Column 10: DAYS_ID_PUBLISH

BANK LOAN CASE STUDY

RESULT:

- ✓ *From the analysis of the data, we can see that the majority of clients opt to repay their loans using the 'Cash through the bank' option, while the 'non-Cash from your account' & 'Cashless from the account of the employee options are not widely used. Additionally, most loan applications are for 'Cash loans and 'Consumer loans', but cash loans have a higher rate of being rejected.*
- ✓ *A large portion of loan applications come from repeat customers, with 70% being from repeat clients, however, these customers are also more likely to have their loans rejected.*
- ✓ *The loan applications for individuals with lower AMT_ANNUITY are more likely to be canceled or unused, and similarly, loan applications with a high AMT_ANNUITY also have a higher rate of being rejected. Additionally, loan applications with a low credit amount are also more likely to be canceled or unused.*
- ✓ *The clients who have been approved for a loan in the past are less likely to have issues with loan repayment as compared to those who were previously denied.*