Disc | Approch | Tech-Stack | ETL | Outliers | Imbalance | v analyses | Scenarios

Bank Loan Case Study

Report created by - Akash Duwedi

PROJECT DESCRIPTION :

In this Project I am going to analyze how to reduce risk because bank faces a challenge: some customers
who don't have a sufficient credit history take advantage of this and default on their loans. In this project I use Exploratory Data Analysis (EDA) to Analyze patterns in the data and ensure that capable applicants are not rejected.

When a customer applies for a loan, The Bank has faces two type of risks:

1. If the applicant can repay the loan but is not approved, the company loses business.

2. If the applicant cannot repay the loan and is approved, the company faces a financial loss.

In this Project my goal is to use EDA to understand how customer attributes and loan attributes influence the likelihood of default.

PROJECT APPROACH :

In this Project my main aim is to identify patterns that indicate if a customer will have difficulty paying their instalments. This information can be used to make decisions such as denying the loan, reducing the amount of loan, or lending at a higher interest rate to risky applicants. The company wants to understand the key factors behind loan default so it can make better decisions about loan approval.

To the analysis I am going through these steps :

- Use Excel and it's Power Query to perform the(ETL Process) Data cleaning.
- Use Power BI For Analysis The Project and Creating Dashboard and Presentation.

Tech-Stack Used In This Project :

- Excel and it's Power Query for (ETL Process) Data Cleaning.
- Power BI for Analysis and Presentation.

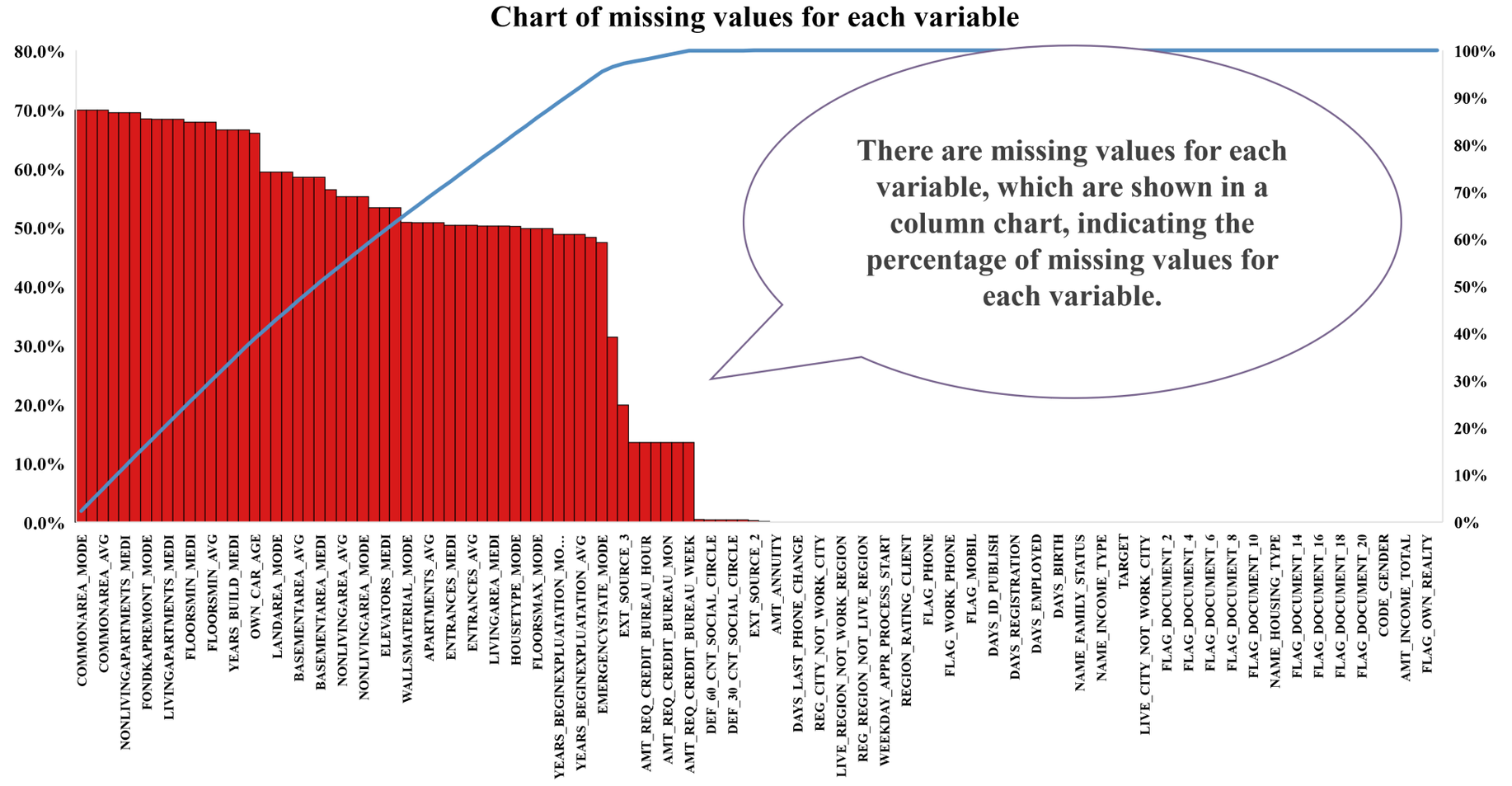
Total Records
49960

Loan Status
All

TARGET
All

Home | Disc | Approch | Tech-Stack | ETL | Outliers | Imbalance | v analyses | Scenarios

Identify Missing Data and Deal with it Appropriately



Total Records
49960

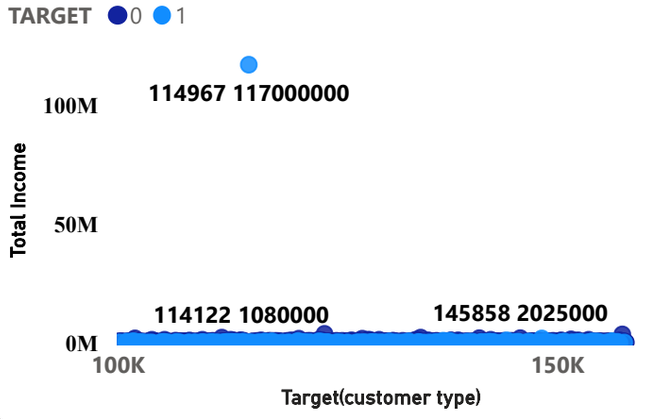
Loan Status
All

TARGET
All

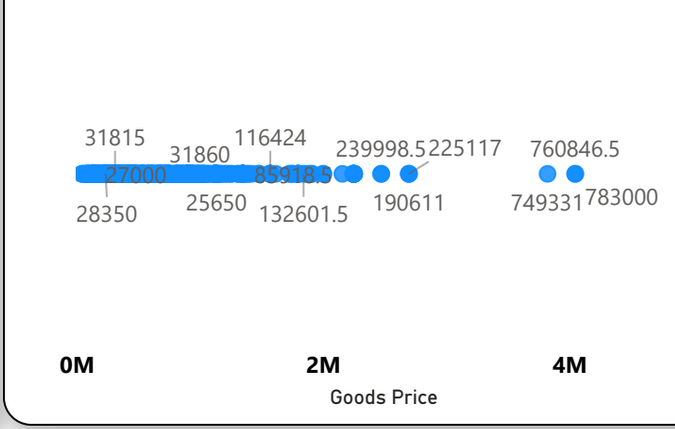
Home | Disc | Approch | Tech-Stack | ETL | Outliers | Imbalance | v analyses | Scenarios

Outliers

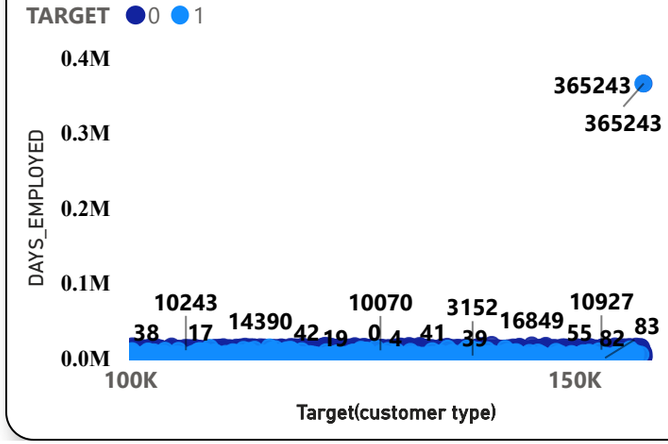
Outliers By Total Income



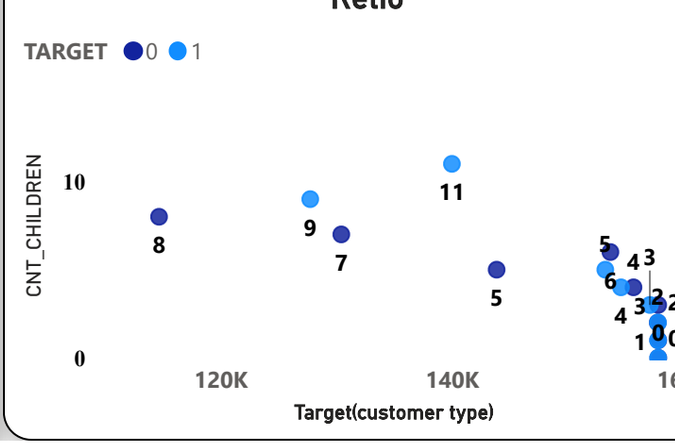
Outliers By Goods And Income Retio



Outlier By Applicant And Days Employed Retio



Outlier By Loan Applicants Having Children Retio



Insights: While analyzing this project, I found several applicants with unique and exceptional characteristics that set them apart from the majority of the applicants.

1. Applicant with SKN_ID_CURR 114967:
- This applicant has an extraordinarily high income of 117,000,000 and is seeking a loan.
- The high income level makes this applicant an outlier in terms of total income.

2. Applicant with SKN_ID 157830:
- Interestingly, this applicant's "days_employed" value is 365,243, which is approximately 1000 years. Such an extreme value for "days_employed" is likely an error or an outlier in the data.

3. Applicant with SKN_ID 110403:
- Interestingly, a total of 5853 clients are going to take goods beyond their own limits.

4. Applicants with SKN_ID 140032 and 127782:

Both of these applicants have an unusually large number of children, with more than 8 children each. This large number of children can have implications for their financial situation and ability to repay a loan.

These applicants, along with their unique characteristics, are grouped as targeted customers who have difficulty paying back the loan amount. In banking terms, they are considered delinquent customers due to the challenges they may face in meeting their loan obligations.

It is crucial to treat these cases with extra attention during the analysis and modeling process. Outliers and extreme data points can significantly impact the results and may require special handling to avoid biases in the final outcomes. Additionally, reviewing and validating the data for accuracy is essential, as some extreme values like the "days_employed" mentioned above might indicate data entry errors or other anomalies. Properly addressing and understanding these outliers will lead to a more accurate and reliable analysis for the project.

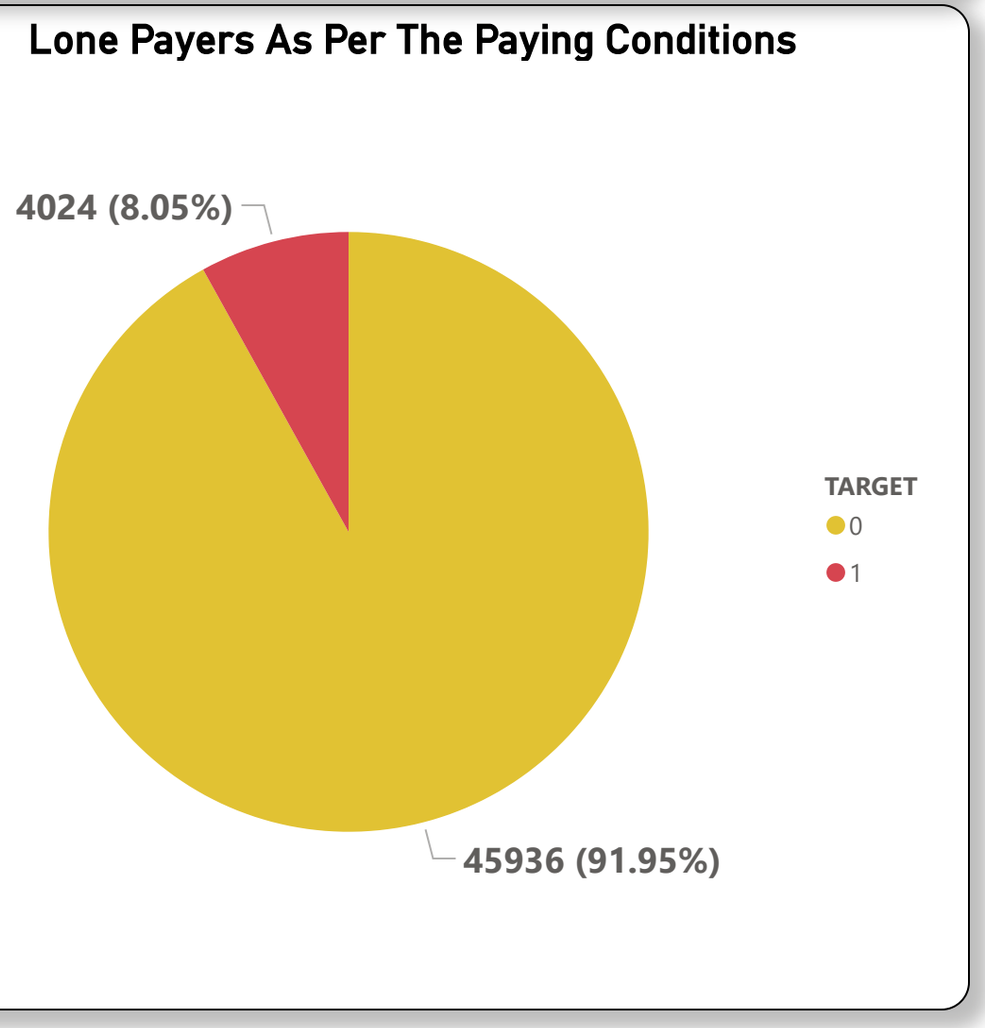
Data Imbalance

Insights:

After conducting the analysis on the database, it was observed that the target variable, which indicates whether a client has payment difficulties or not, exhibits a significant data imbalance. The data is divided into two classes:

1. **Class 0 (Client with No Payment Difficulties):** This class represents cases where clients have not experienced late payments (more than X days) on any of the first Y instalments of the loan. The analysis revealed that this class constitutes the majority, accounting for approximately 91.95% of the data.
2. **Class 1 (Client with Payment Difficulties):** This class includes cases where clients have faced late payments (more than X days) on at least one of the first Y instalments of the loan. The analysis found that this class is a minority, making up only around 8.05% of the data.

Hence Class 1 clint are defaulters in the given database it's risky.



Total Records
49960

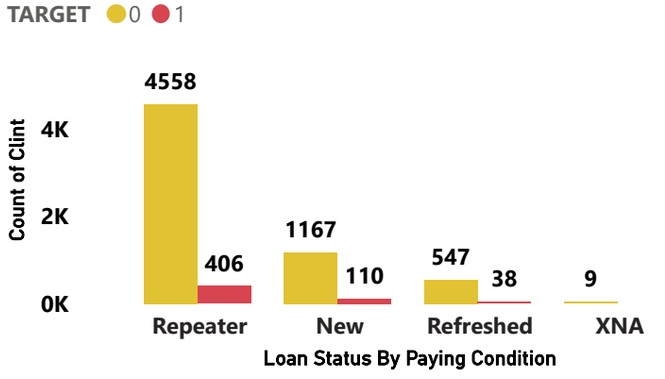
Loan Status
All

TARGET
All

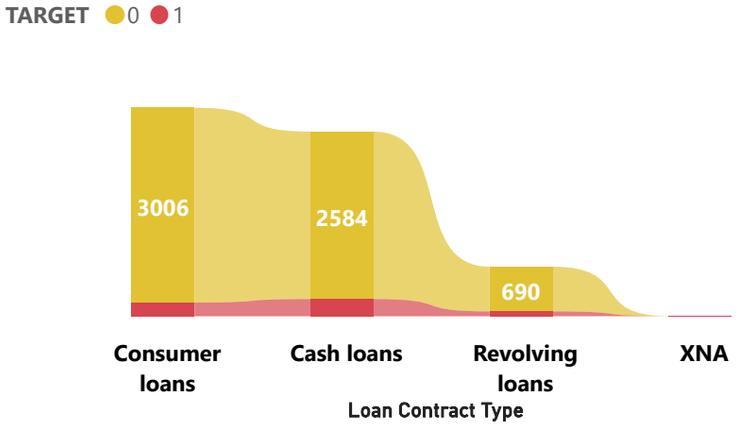
Home | Disc | Approch | Tech-Stack | ETL | Outliers | Imbalance | v analyses | Scenarios

Univariate Segmented Univariate, and Bivariate Analysis

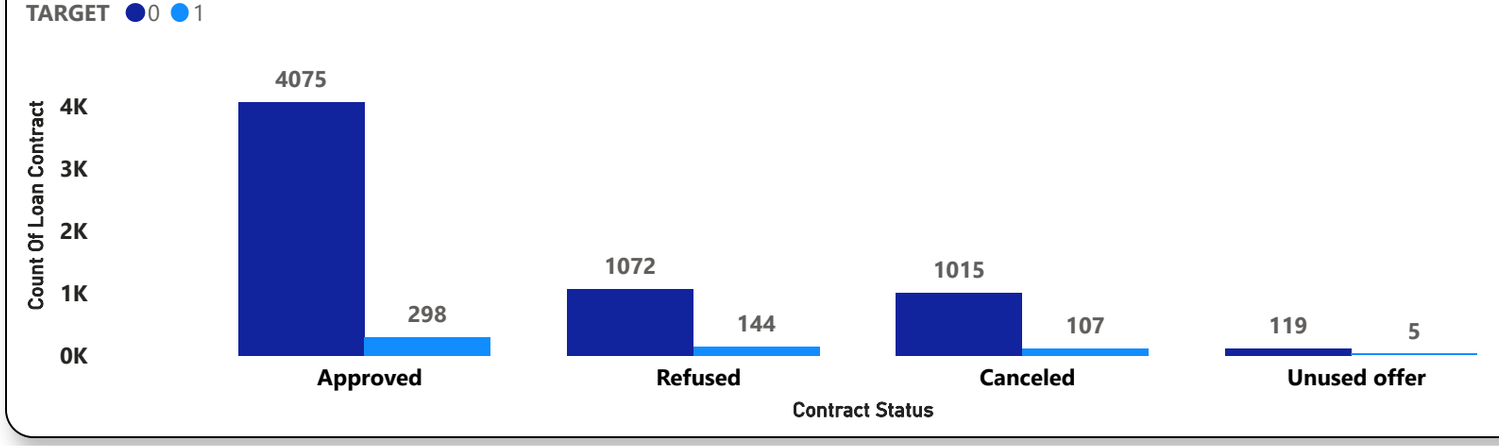
Clint By His Paying Condition And Loan Status



Loan Contract Type and Target Retio



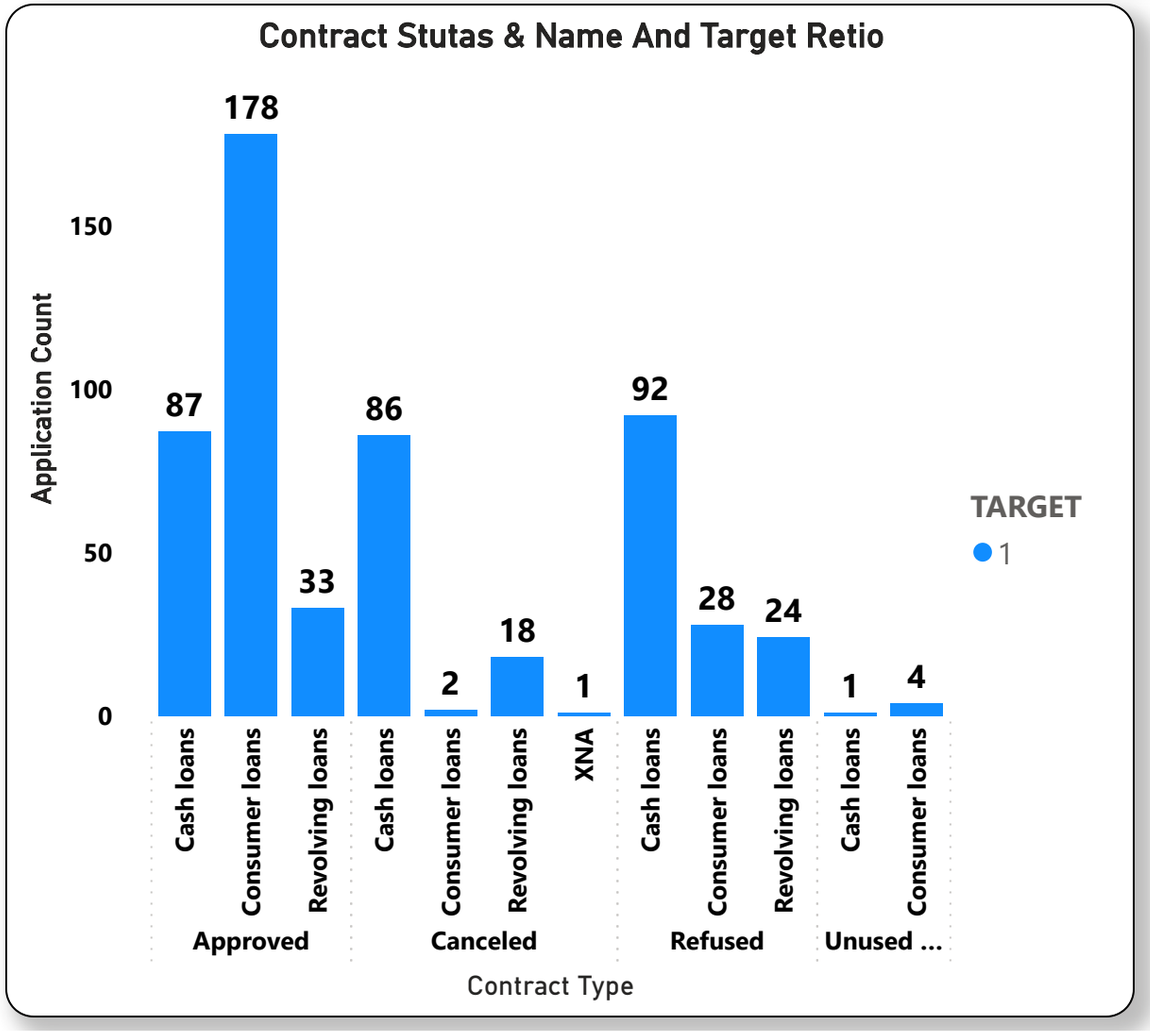
Contract Status By Targeted Clint



Insights:

After conducting the analysis on the database, it was observed that the clients under target category 1 are taking a huge amount of loan and are unable to reimburse it back. In the chart, it is easy to find repeaters under category 0 and 1, as well as the type of contract and status. We have to reduce the number of clients in category 1.

Identify Top Correlations for Different Scenarios



Insights:

After conducting the analysis on the database, it was observed that the target variable, which indicates whether a client has payment difficulties or not, was used to filter and focus on clients with payment difficulties (target 1). It was found that cash and consumer type of loan categories were prevalent across all cases, whether they were approved, canceled, refused, or unused offers. Additionally, it was noticed that these types of clients were taking a high amount of credit relative to their income.