**Risk Evaluation in Financial Institutions**

**Introduction:**

This case study aims to give an idea of applying EDA in a real business scenario. In this case study, we will develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise 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

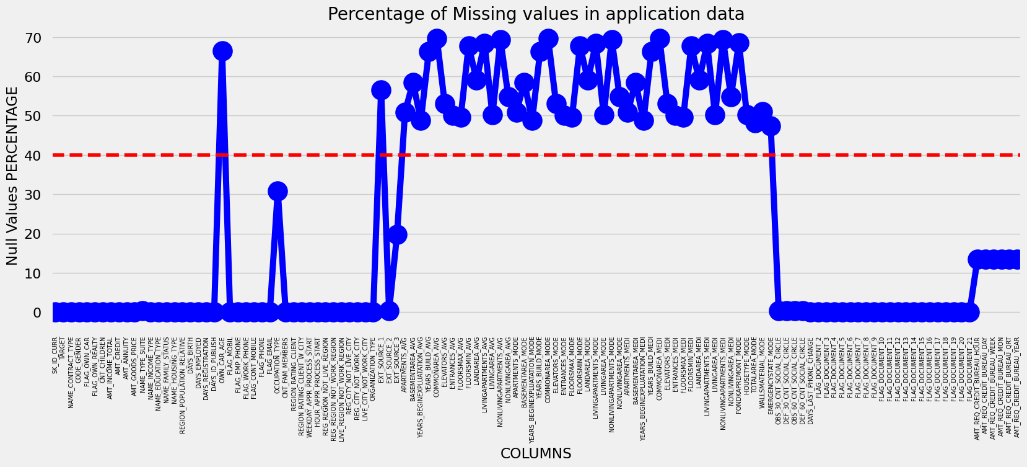
When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

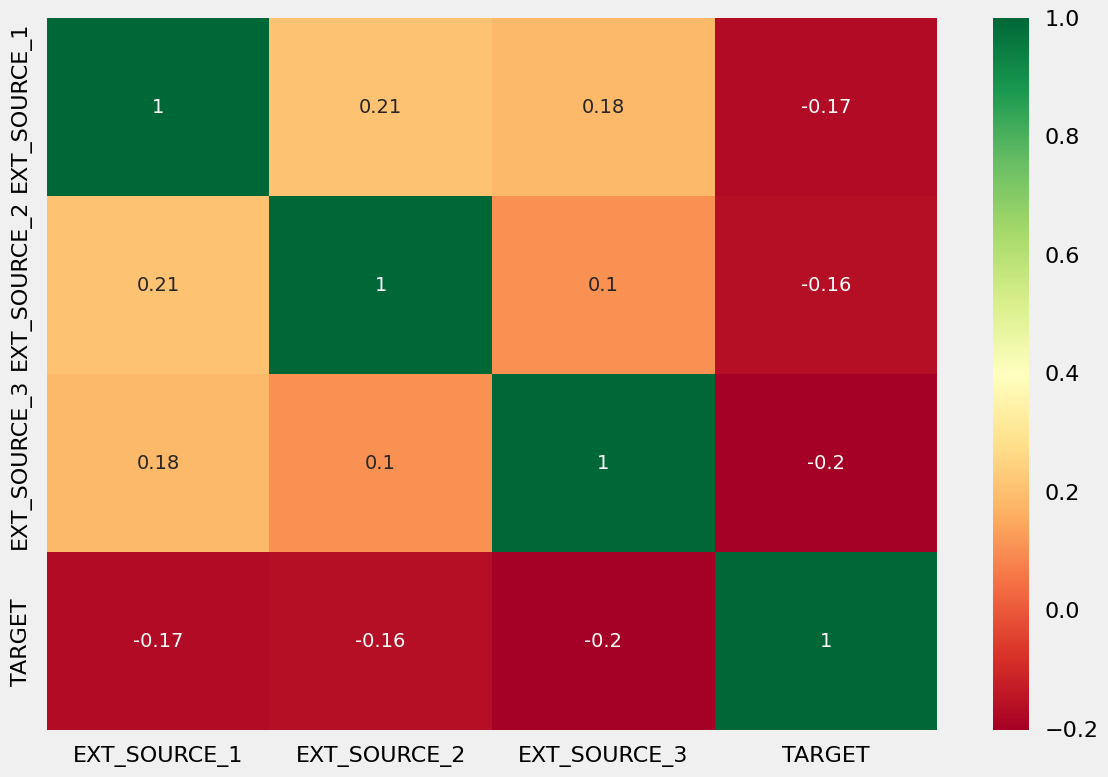
1. **Approved:** The Company has approved loan Application
2. **Cancelled:** The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
3. **Refused:** 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.

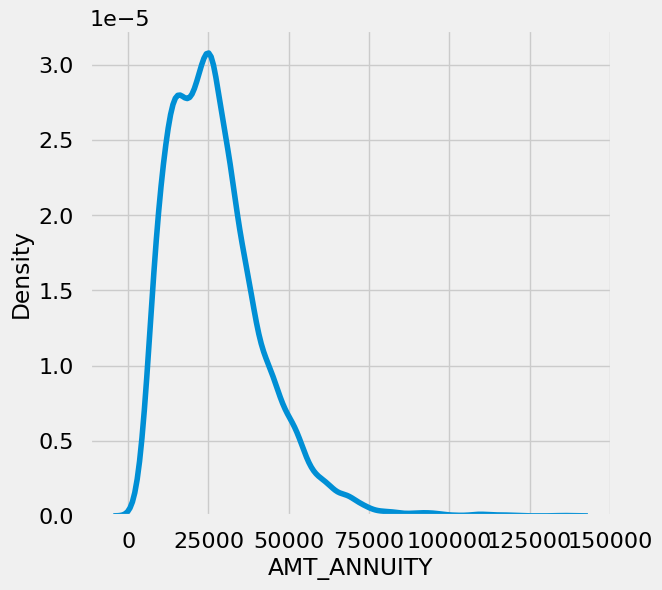
**Business Objective:**

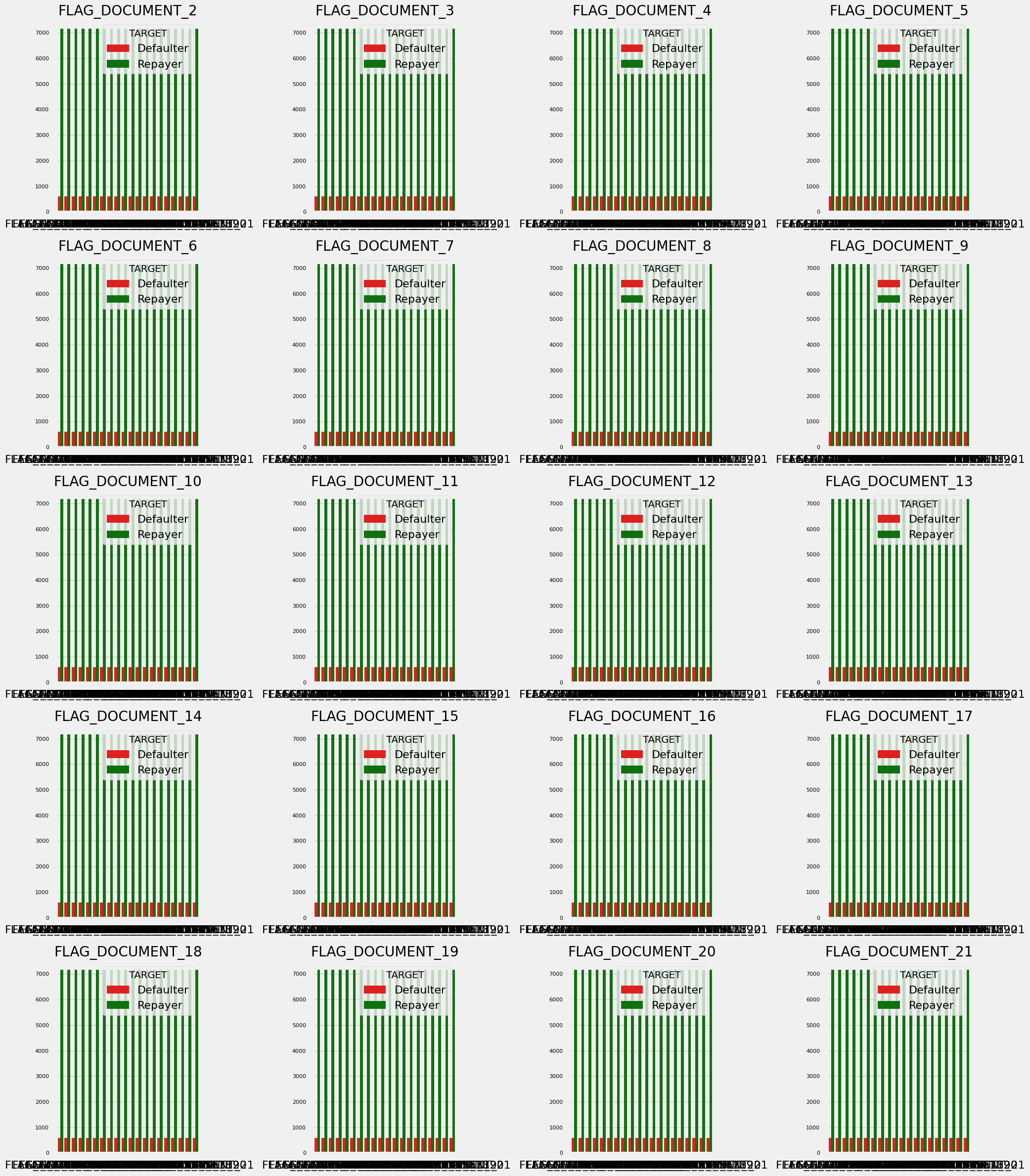
This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA 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 utilise this knowledge for its portfolio and risk assessment.

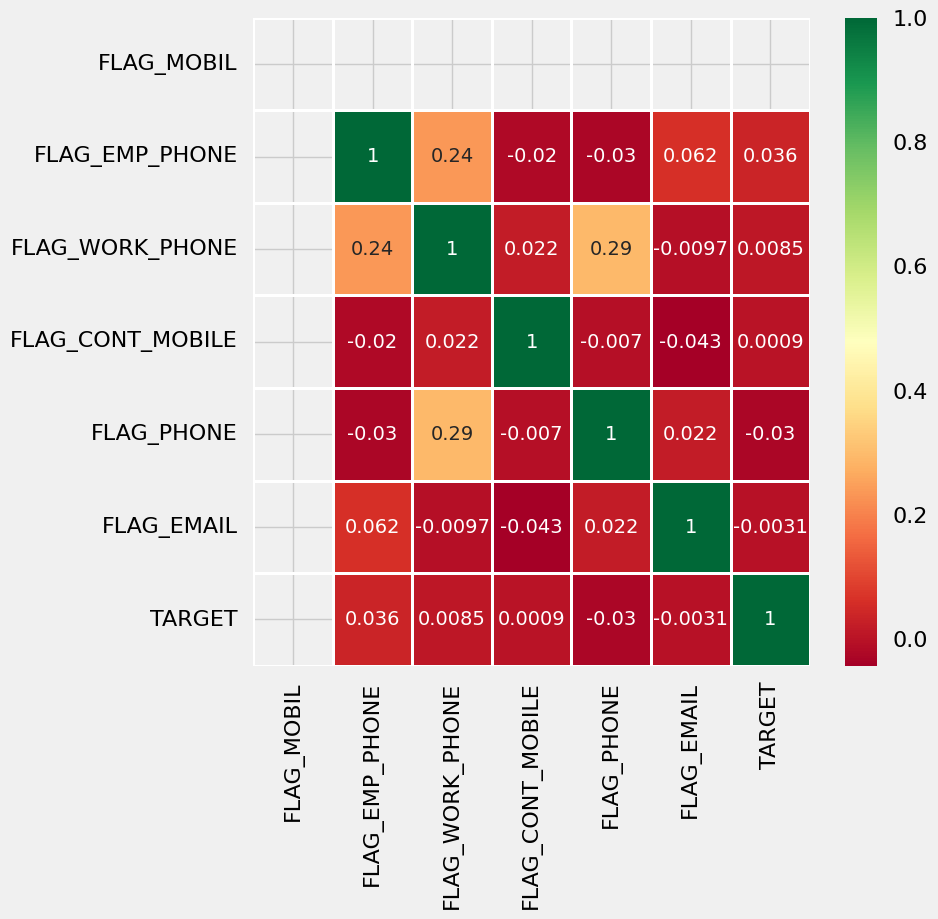
**Data Visualizations**

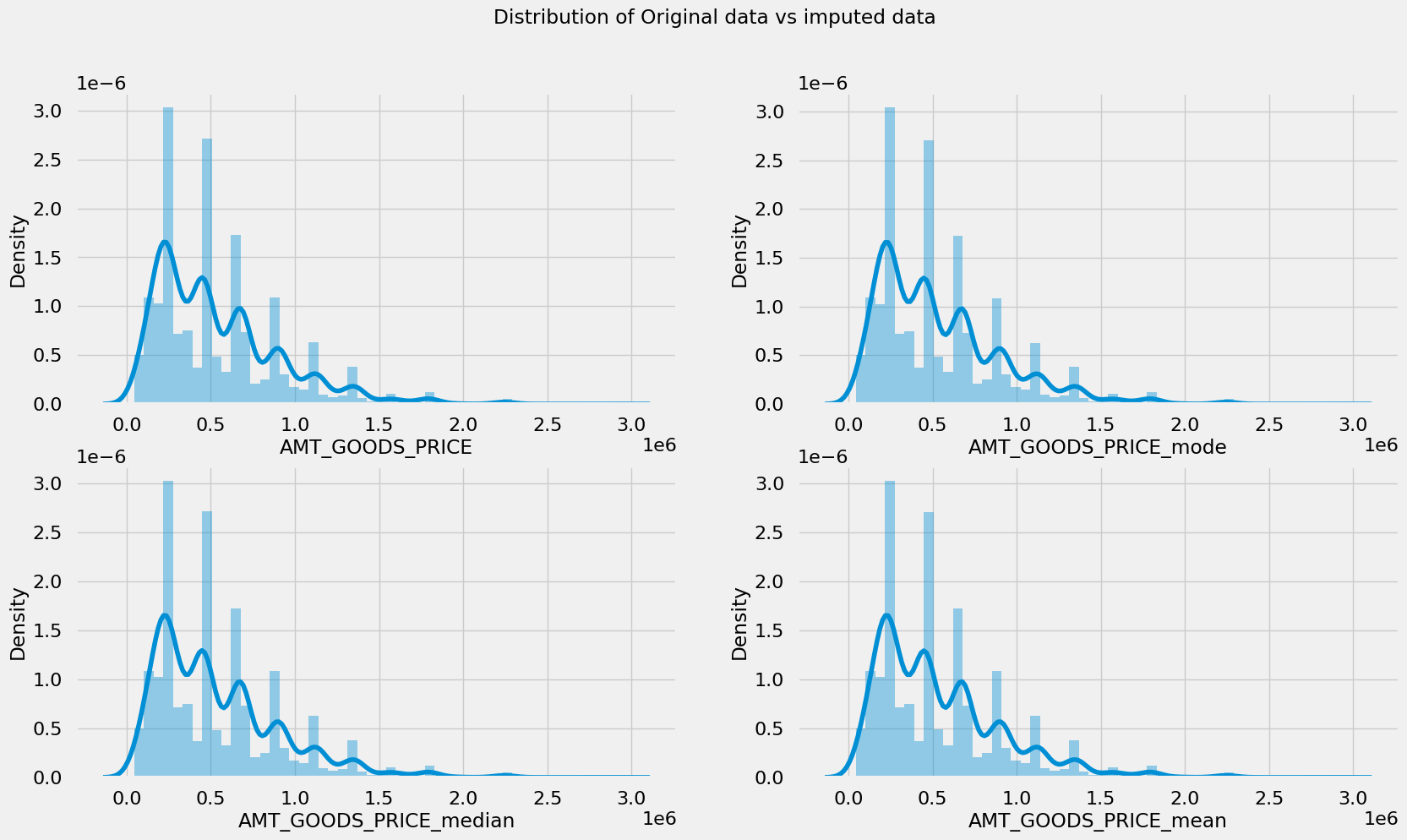




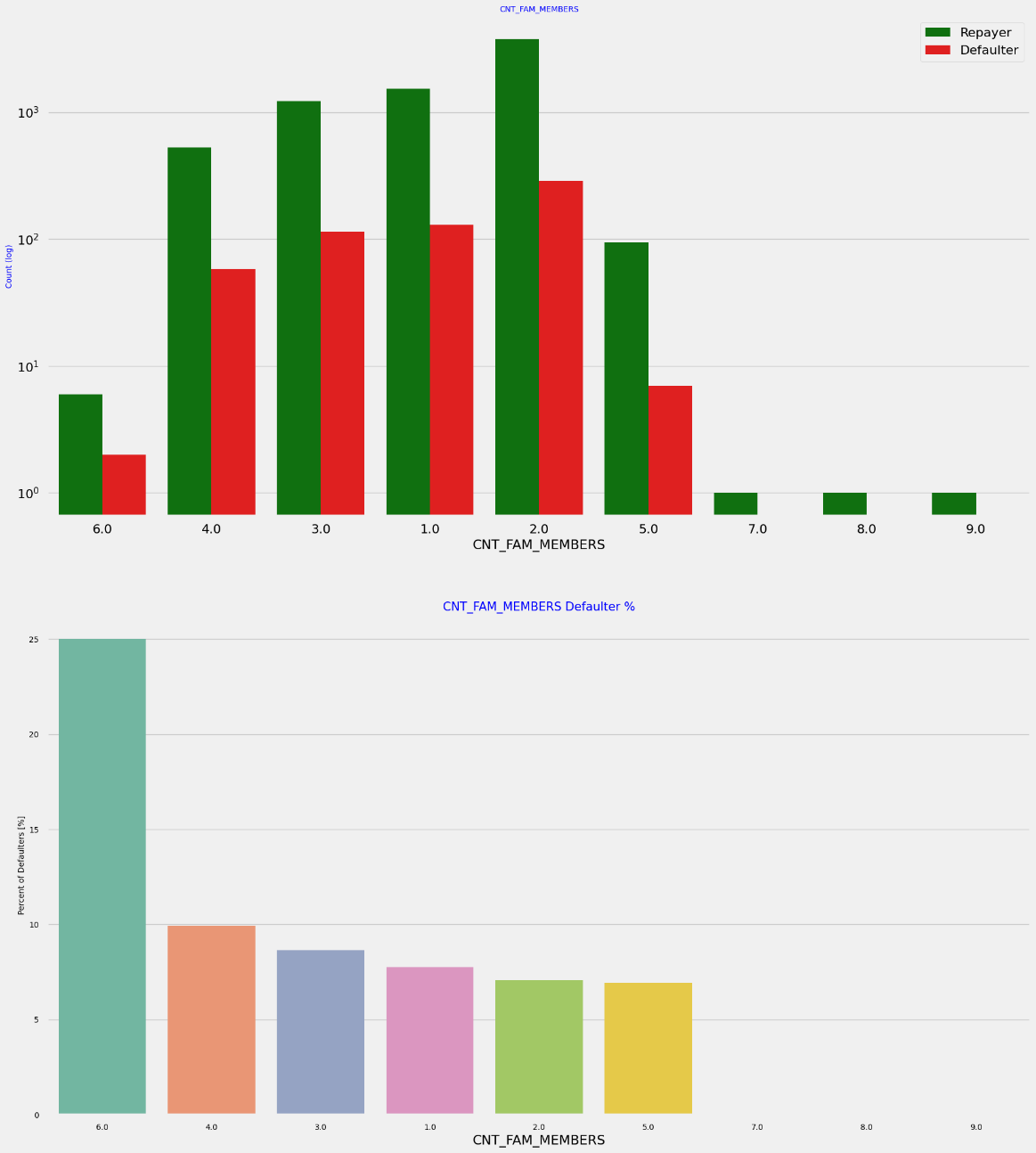




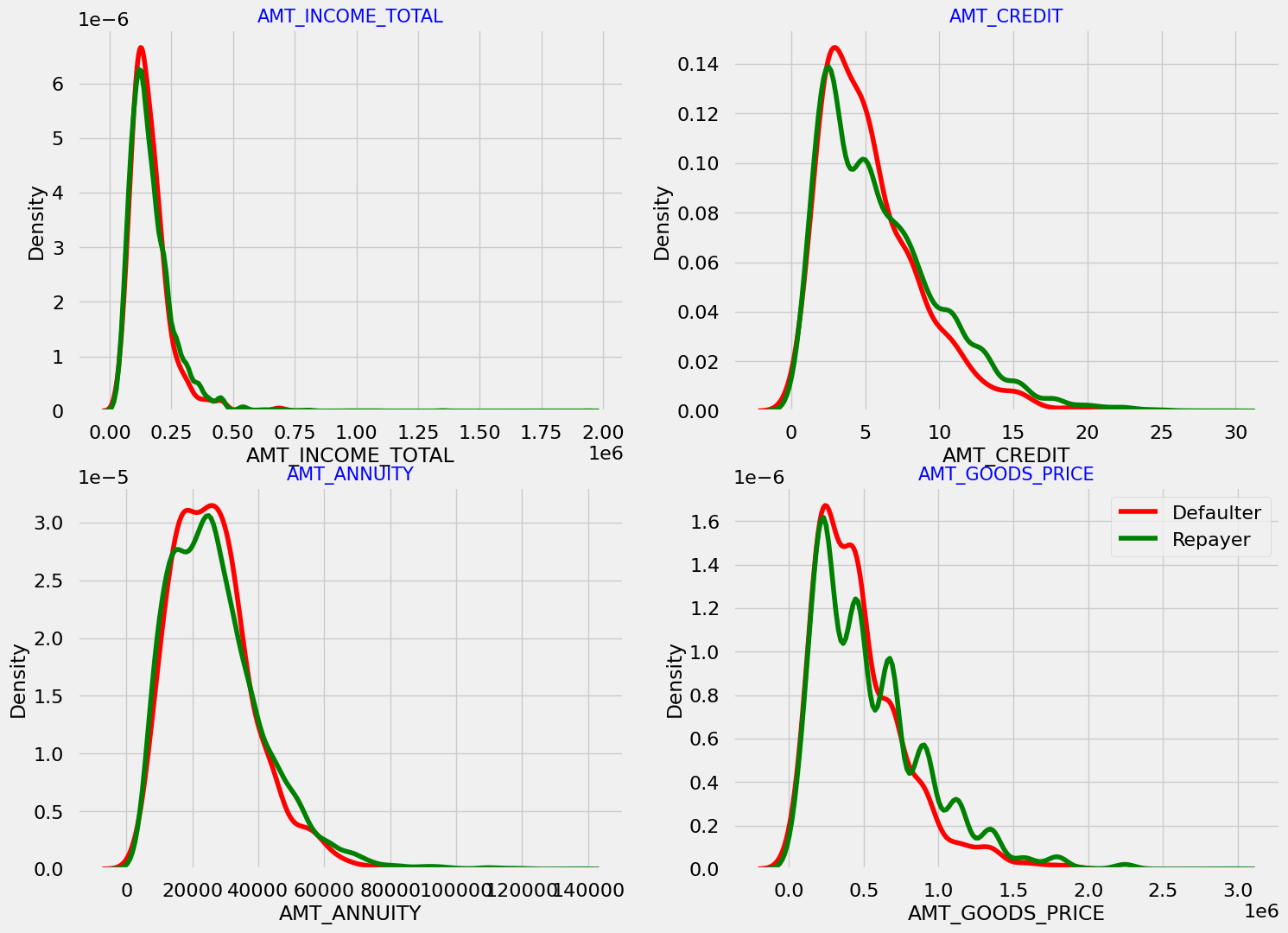


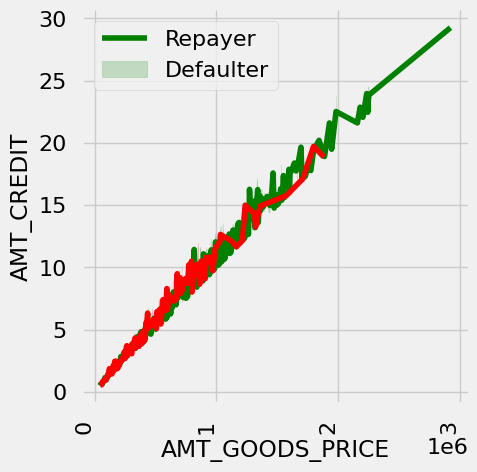


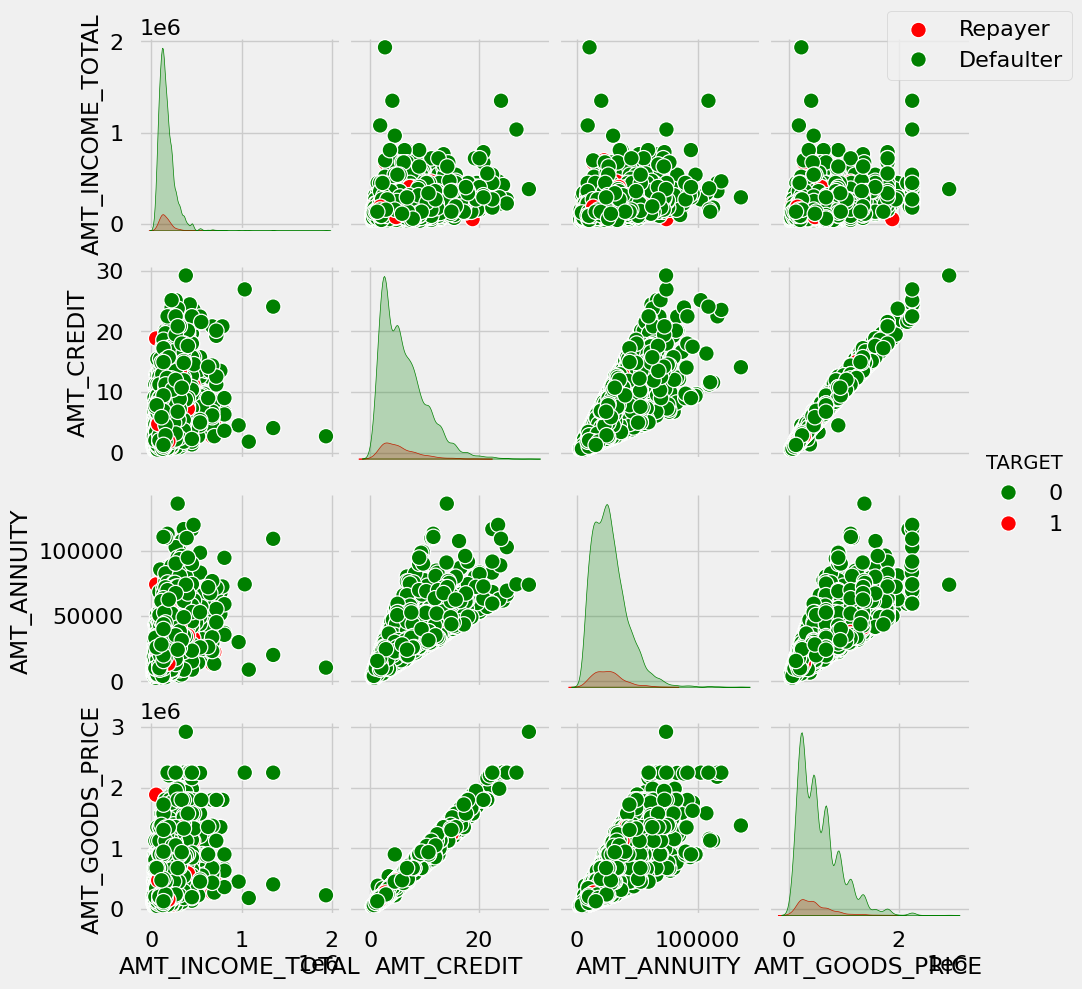












**Conclusion:**

The analysis identifies key attributes that can help the bank determine whether a loan applicant is likely to repay their loan or default.

**Predictors of Loan Repayment:**

* Education Level: Applicants with an academic degree are less likely to default.
* Income Type: Students and businessmen show no history of defaults.
* Region Rating: Clients from regions with a rating of 1 are safer.
* Organization Type: Those working in Trade Type 4 and 5, as well as Industry Type 8, have a default rate of less than 3%.
* Age (DAYS\_BIRTH): Individuals over 50 years old have a low probability of defaulting.
* Employment Duration (DAYS\_EMPLOYED): Applicants with more than 40 years of experience have a default rate of less than 1%.
* Total Income (AMT\_INCOME\_TOTAL): Applicants with an income exceeding 700,000 are less likely to default.
* Loan Purpose (NAME\_CASH\_LOAN\_PURPOSE): Loans taken out for hobbies or buying a garage are mostly repaid.
* Number of Children (CNT\_CHILDREN): Applicants with zero to two children tend to repay their loans.

**Predictors of Loan Default:**

* Gender (CODE\_GENDER): Men have a relatively higher default rate.
* Family Status (NAME\_FAMILY\_STATUS): Individuals who are in civil marriages or are single have a higher likelihood of defaulting.
* Education Level: Those with lower secondary or secondary education default more frequently.
* Income Type: Clients on maternity leave or unemployed have a higher default rate.
* Region Rating: Clients living in regions with a rating of 3 have the highest default rates.
* Occupation Type: Low-skill laborers, drivers, waiters, security staff, laborers, and cooking staff have high default rates.
* Organization Type: Organizations with high default rates include Transport Type 3 (16%), Industry Type 13 (13.5%), Industry Type 8 (12.5%), and restaurants (less than 12%). Self-employed individuals also have a relatively high default rate, suggesting the need for higher interest rates to mitigate risk.
* Age (DAYS\_BIRTH): Younger individuals, particularly those aged 20-40, have a higher probability of defaulting.
* Employment Duration (DAYS\_EMPLOYED): Applicants with less than 5 years of employment have a high default rate.
* Number of Children (CNT\_CHILDREN) and Family Members (CNT\_FAM\_MEMBERS): Clients with nine or more children have a 100% default rate, necessitating loan rejection.
* Credit Amount (AMT\_GOODS\_PRICE): Default rates increase when the credit amount exceeds 3 million.

**Additional Considerations:**

* Housing Type: A high number of loan applications come from individuals living in rented apartments or with parents. Offering loans with higher interest rates to these groups could mitigate potential losses.
* Loan Amount (AMT\_CREDIT): Applicants seeking loans between 300,000 and 600,000 tend to default more, warranting higher interest rates for this credit range.
* Total Income (AMT\_INCOME): Since 90% of applicants have an income total of less than 300,000 and a higher probability of defaulting, higher interest rates should be considered for this income category.
* Number of Children (CNT\_CHILDREN) and Family Members (CNT\_FAM\_MEMBERS): Applicants with four to eight children have a very high default rate, suggesting the need for higher interest rates.
* Loan Purpose (NAME\_CASH\_LOAN\_PURPOSE): Loans taken out for repairs have the highest default rate. This risk could be mitigated by continuing to either reject these applications or offer loans with high-interest rates, as has been done in the past.

**Strategic Opportunities:**

* Cancelled Clients: 90% of previously canceled clients have repaid their loans. Documenting the reasons for cancellation could help the bank negotiate terms with these clients in the future, potentially increasing business opportunities.
* Rejected Clients: 88% of clients who were previously refused loans by the bank have since become repaying customers. Recording the reasons for rejection could prevent business loss and allow the bank to reach out to these clients for future loans.