**AAS Exam Team 6 (Roll no: 26, 27, 29, 54, 84)**

**Problem Statement:**

In this analysis, 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 to their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specializes 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.

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 installments 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):

**Approved:-** The Company has approved loan Application

**Canceled:-** The client canceled 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 canceled by the client but on different stages of the process.

In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

**Solution:**

1. **Understanding the DataSet :-**

We have checked the dimensions of dataset (using shape), data types (dtypes), and column-names (columns), statistical information about dataset (using the describe(): This information includes statistics that summarize the central tendency of the variable, their dispersion, the presence of empty values and their shape) and other information about the data (using info(): it provides purely descriptive information about the dataset.)

There are two data sets

1) application\_data.csv alias **df**

2) prervious\_application.csv alias **pa**

1. **Cleaning the Data :-**

* The data is cleaned where null values and columns with single values and duplicate

values were handled.

* Many columns had null values so The columns with more than 40% of null values were dropped.
* Then data below **40%** null values was handled by replacing some the null values with **mode** in columns and other rows were dropped as it would have created bias.
* The bar graph shows column having null values greater than 40%. After that we checked the values which is equal to 0 and applied mode.
* There were some irrelevant columns that do not affect our analysis (like documents, flags, contacts) so we dropped these columns too.
* Also, for further analysis datatype of **14 columns** is converted into numeric type.

1. **EDA :-**

* Exploratory Data Analysis was done to check univariate analysis, categorical variables and bivariate analysis. We can find that some numerical variables consisted of very high values as compared to their respective means.
* That's why we have created charts using boxplot to understand the patterns.

From the boxplot we were able to determine the outliers and now we are going to treat them.

1. **Visualizations :-**

For visualization of the data, we have used different types of graphs like bar graph, pie chart, histogram, scatter plot, box plot etc.

1. **Observations from data analysis :-**

1) Total income is higher for client without payment difficulties as compared to client with payment difficulties.

2) By observing the boxplot, we can infer that Client with payment difficulties are in range of 31-49, whereas client without payment difficulties are in range 34-54

3) Here we observe that Females have more Loan payment difficulties as compared to Male's.

4) Pensioner and Govt Employees have **better** on-time payments.

5) Maternity category is significantly higher problem in repayment

6) AMT\_INCOME\_TOTAL, AMT\_CREDIT, AMT\_ANNUITY, AGE, AMT\_GOODS\_PRICE

**Observations**:

=AMT\_GOODS\_PRICE and AMT\_CREDIT show **positive** correlation

=AMT\_GOODS\_PRICE and AMT\_ANNUITY show **positive** correlation

7) Female applicants have more difficulties in payment as compared to male applicant's.

8) Applicant’s who are businessman and student's pay their loan on time although there count is low.

9) Applicant's who have higher education have less difficulty in paying loan as compared to Secondary/secondary special.

**6. Suggestions for the Bank / Conclusion :-**

1. Banks should focus more on contract type ‘Student’, ’pensioner’ and ‘Businessman’ with housing type other than ‘Co-op apartment’ for successful payments.

2. Banks should focus less on income type ‘Working’ as they are having most number of unsuccessful payments.

3. Also with loan purpose ‘Repair’ is having higher number of unsuccessful payments on time.

4. Get as much as clients from housing type ‘With parents’ as they are having least number of unsuccessful payments.

5. Client categories to be targeted for providing loan

* Clients who are employed for more than 19 years
* Clients in the age range 30-40 and 40-50
* Clients who are Married
* Male clients with Academic degree
* Repeater clients