**AAS Exam Team 7 (Roll no: 11, 43, 55, 60, 86)**

**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 (info), 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\_app**

2) prervious\_application.csv alias **df\_prev**

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 35% of null values were dropped.
* There were some irrelevant columns that do not affect our analysis (like documents, flags, contacts) so we dropped these columns too.

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.

1. **Visualizations :-**

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

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

1) The AMT\_CREDIT has outliers.

2) With the help of Inter Quartile Range, we can find out the statistical dispersion.

3) With the help of Bar chart we understand that more number of people are applying for Cash Loans as compared to Revolving loans.

4)The clients who are taking the more number of loans are in the "working" income type, followed by "commercial asscociate" income type.

5) There are outliers when we check Income Amount of Client

6)No.of people defaulting in the first Y installments are less compared to people defaulting for other reasons.

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

1. Banks should focus more on contract type ‘Working’ income type clients.