**Problem Statement:**

**This analysis is done for basic understanding of risk analytics in banking and financial services and to understand how data is used to minimise the risk of losing money while lending to customers. We have created visualisations of various attributes like age, occupation, family status etc. which correlate with customers who are vulnerable to default on payments.**

* **To Achieve the model following methods were followed:**

**1. Cleaning The Data:**

* **The data is cleaned where null values and columns with single values and duplicate values were handled. We calculated the percentage of null values. The columns with more than 30% of null values were dropped.**
* **We also checked for the columns which can have the possibility of duplicates and fortunately there was no duplicate data for ID’s.**
* **The values which were negative were converted into positive variables according to their interpretation like employed days,birth days and registration days were changed to positive .**
* **Some columns which had irrelevant data were also dropped.**

**2. EDA:**

* **Exploratory Data Analysis was done to check categorical variables for ex, TARGET where we converted the possibility of problem to repay amount and easy repayment of loan to 0 and 1 respectively.**
* **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.**
* **We have observed the outliers of distribution of Annuity and Income Total.**
* **We also performed a Binning operation to segregate data into various batches in order to convert some continuous fields like age,income to categorical values for better interpretation of their correlation with defaulters.**

**3. Analysis:**

* **In univariate analysis we did analysis of various categories like credit amount range and payment issues, age related payment issues.**
* **We concluded that medium income group tends comparatively lesser to borrow credit**
* **92% of clients have no difficulty with payment whereas 8% of clients are facing difficulties in payment.**
* **In Bivariate Analysis we compared is their any dependency of factors like credit amount and education status, Income range and Genders are dependent or not.**
* **For further checking dependency and correlation we made Heatmap of all factors related to each other.**