Problem Statement

- The problem statement is about a consumer finance company that specializes in lending various types of loans to urban customers.
- The company wants to minimize the risk of financial loss due to loan defaults.
- The company has data about past loan applicants and whether they defaulted or not.
- The aim is to identify patterns which indicate if a person is likely to default.
- 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 driving factors behind loan default, i.e., the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.
- The problem is to use Exploratory Data Analysis (EDA) to understand how consumer attributes and loan attributes influence the tendency of default.

Solution Approach

Data Understanding:

Start by understanding the data provided. This includes understanding each variable and its significance.
Following some of the data dictionary provided:

- loan_amnt,int_rate, grade, verification_status, loan_status, total_acc, term, funded_amnt_inv etc

Data Cleaning:

Clean the data by handling missing values, outliers, and incorrect data entries.
Provided data set do not have any missing values.

Exploratory Data Analysis (EDA):

Perform EDA to understand the patterns and relationships in the data.
This can include univariate analysis, bivariate analysis, correlation analysis, etc.

Solution Approach…

Univariate Analysis

Analysis with Single variable
Mean, Median, max, min, std, count etc
Distribution - Histogram, Distplot, Countplot, boxplot

- Loan Amount should be less when borrower have more than one credit lines account.
- In the histplot shows when borrower has 10 credit lines account they have more loans that's leads to defaulter.
- When customer is not verified, loan should not be given.
- Maximum loan is given to the customer who have Rented or Mortgage house.

Solution Approach…

Bivariate Analysis

Analysis with two variables
Relationship with two variables - Scatterplot, boxplot, barplot etc

- Loan should be given only for verified customer
- Maximum loan given to the Rented or Mortgage customer
- Avoid to give loan, who has lower grade

Solution Approach…

Multivariate Analysis

Analysis with more than two variables
Relationship between more than two variables
Heatmap

- Loan amount and open account has Correlation 0.18, That means customer has more credit account likely to be a defaulter.