EDA Analysis for Lending Club

Purpose

Consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision 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 contains information 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, 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.

EDA Analysis Steps

Step 1: Data Cleaning

Step 2: Univariate Analysis

Step 3: Univariate Analysis

Step 4: Bivariate/Multivariate Analysis

Step 5: Results

Data Cleaning

Step 1: Identified and dropped null columns and columns with only one unique values

Step 2: Removed duplicate rows. Removed rows with loan status “Current” as those represent ongoing loan.

Step 3: Standardized data by replacing required values like “Source Verified” with “Verified”, by changing data type.

Step 4: Removed columns not required for analysis id, member_id etc

Step 5: Identified outliers and removed required outliers from “annual_inc” column.

Univariate Analysis

Max no. of loans are charged off have verification status is Verified.

Loan with purpose "small business, renewable energy, debt_consolidation, educational" are charged off more.

E,F,G grade are charged off more.

Loans for Term 36 months are less charged off than Loans for 60 months.

Year 2011 had higher number of charged off loan percentage.

Person with 10years of employment length have higher number of charged off loans.

Loans with home ownership status "OWN" are less risky.

Most Charged off loans have annual income between 3K-31K, loans with higher income are less risky.

Loans at the rate of 13-17% have charged off more.

Bivariate Analysis

Higher Grade likely to get higher loan_amount and have high chances for defaulted loan.

Loan for 60 Months term have higher interest rate. Loan with Higher interest rate is risky.

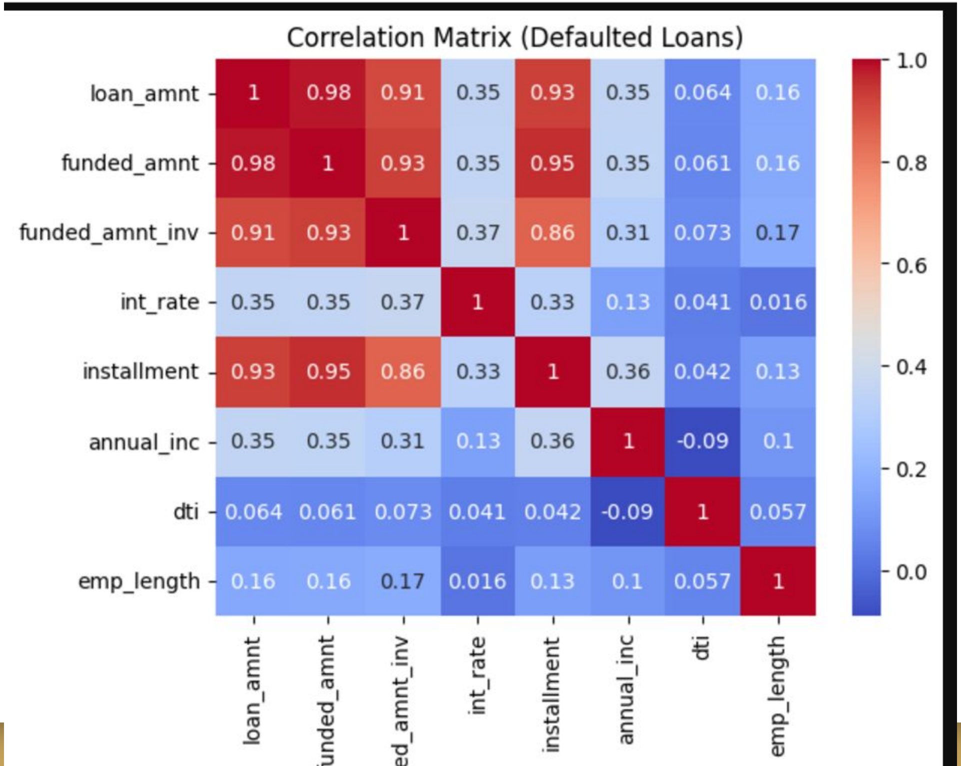
More annual income between 112K-145K showed more DTI and more charged off loans.

Loans with 10+ Employment Length are likely to get charged off.

Loans got defaulted more in year 2011, loans increased from 2007 to 2011.

Most Loans defaulted have annual income less than 40K.

Correlation



Correlation Analysis

Loan_amount,funded_amt_inv,fundend_amt are corelated.

dti and annual_inc negatively coreleated.

Summary

Verified loans can get defaulted.

60-month loans are more likely to default.

Loans High DTI shows more number of default loans.

Annual incomes less than $40,000 have a higher risk of defaulting.