**Assignment**

**Predicting Loan Default**

**Exploratory Data Analysis (EDA)**

**Business Scenario:**

Loan default prediction is a common problem for various financial companies and a well-defined and known problem in data science. This is the type of problem banks, credit card companies (e.g., Visa), micro-credit banks or FinTech companies face whenever customers ask for a loan. The data from Lending Club about lenders is available on their website and we will be using it to make predictions about loan default and whether or not we should lend to a customer. We will be using the data from 2007-2011 because most of the loans have been repaid or defaulted on.

**Problem Statement:**

1. To find which category customers availing loan are more defaulters.

2. To reduce the financial losses

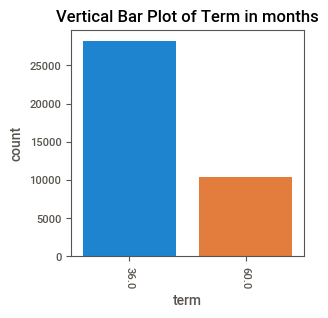
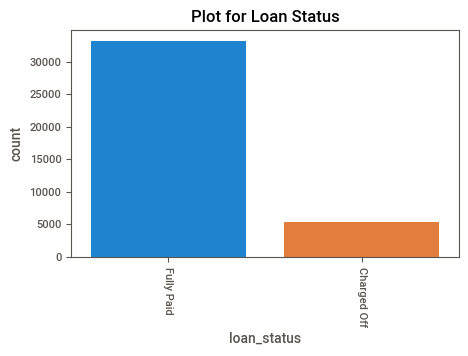
**Source:**

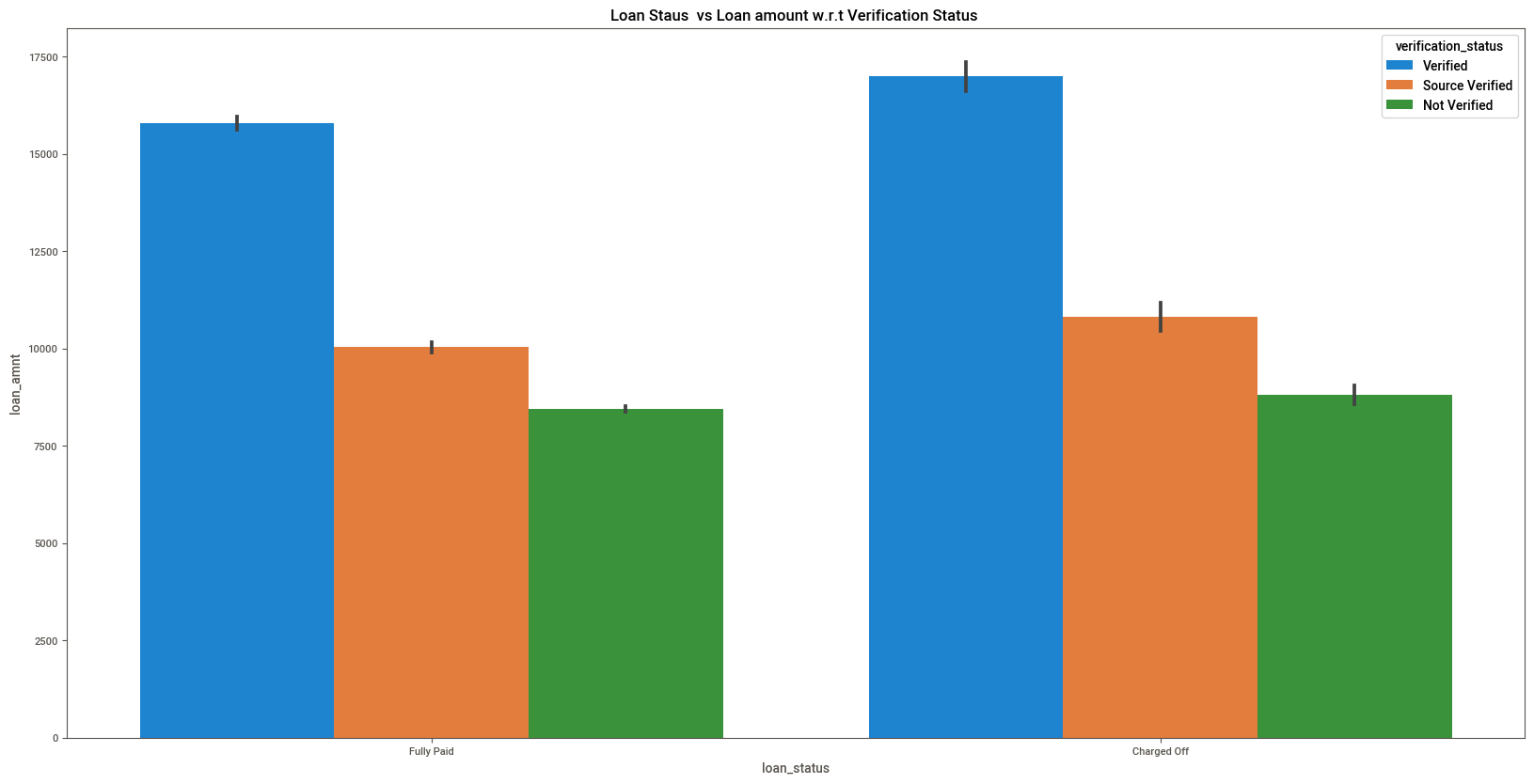
The sample consist of rows & columns. The data was collected to understand the defaulter’s rate and reason of loan defaulters

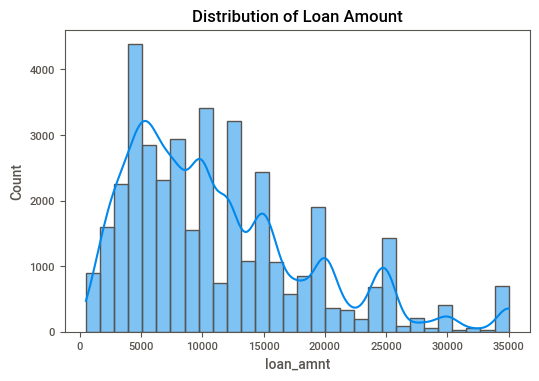
Data has (rows 42538, columns 52)

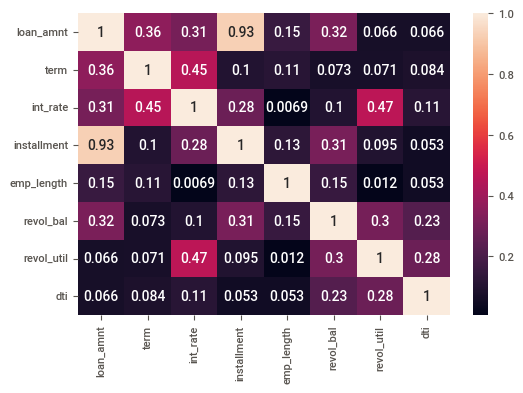
Numerical & Categorical columns

**Exploratory Data Analysis**

****

****

****

· 

**Questions:**

1. Distribution of Numerical Attributes (distplot and histogram)

a) Distribution of Loan Value /Loan Amount

What is distribution of Loan Amount

b) Interest Rate

c) Instalment

d) Annual Income

e) Revol Balance

f) Revol Utilization

2. Unique and counts of Categorical attributes (Count plot, pie-chart)

a) Term

b) Employment Length

c) Loan Status

d) Home Ownership

e) Verification Status

f) Purpose/Title

3. Relation between two numerical attributes (scatter Plot

a) 6C2

a) What is the instalment with interest rate?

b) How Interest rate is applied wrt Annual Income

c) What are the instalments wrt to interest rates

d) What is the loan amount wrt to interest rate

e) What is the Loan amount sanctioned wrt Annual income

4. Correlation Matrix is between numerical values

Heat Map

5. Relation between numerical and categorical values

a. All Numerical attributes with respect to loan status.

b. All Numerical attributes with respect to loan status add the hue all categorical values one at a time.

**Insights:**

1. Maximum loan was given at the interest rate of 10.99 % total count was 958 (9.58%)

2nd & 3rd int rate 13.49 % count 831 (8.31%) & 11.49 % count is 826 (8.26%). 3. lowest rate of interest was 14.67 with 1 count (1%).

1. Loan status is as below Fully Paid 85.71 %, Charged Off 14.23 % & Current 0.047780%
2. Customers who stay on Rent & Has Mortgage there home has taken loan Value goes as Below

RENT 47.442287 % MORTGAGE 44.569561 % OWN 7.642578 % OTHER 0.319714 % NONE 0.018807 %

1. Maximum customers Data seems not verified before disbursing loan with contributes approx. 44% over total data
2. Not Verified 44.097043 % Verified 31.668156 % Source Verified 24.227749 %
3. Percentage of customer taking loan to repay debt is max which is 46.49 %.

|  |  |
| --- | --- |
| Purpose | Percentage |
| debt\_consolidation | 47.24 |
| Credit\_card | 12.95 |
| other | 9.90 |
| home\_improvement | 7.47 |
| major purchase | 5.46 |
| small\_business | 4.61 |
| car | 3.88 |

1. Default rate Late payment issues are more for the customers who has taken loan for debt\_consolidation & small\_business
2. Customers take loan to repay bills for debt\_consolidation 46 % & credit\_card 12% wrt to total loan data
3. Data with Source Verified has maximum defaulter cases
4. Below details of the Loan Status vs Loan amount

|  |  |
| --- | --- |
| loan\_status | loan\_amnt |
| Fully Paid | 369961250.0 |
| Charged Off | 66607975 |

1. Customers who fall in max employee length which is 10yrs has max sum of Loan also customers who falls in 0-2 yrs. as below

|  |  |
| --- | --- |
| emp\_length | loan\_amnt |
| **10** | 116370725 |
| **2** | 45068400 |
| **0** | 44231800 |

1. Default rate Late payment issues are more for the customers who has taken loan for debt\_consolidation & small\_business & Data with Source Verified has maximum defaulter cases so proper verification should be done before providing loan on the given categories.

Also as Source Verified issued loan has maximum default & late payment issues so agencies who verifies customer details should be monitored properly