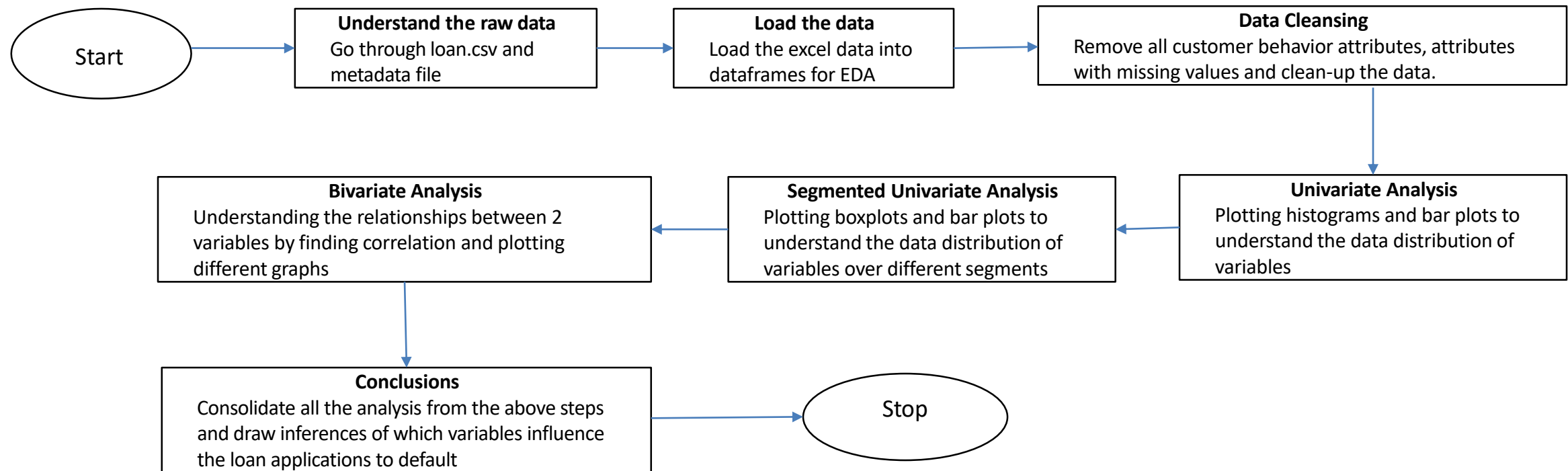# LENDING CLUB CASE STUDY

SUBMISSION

Group:
Naresh Muniganti
Sandipan Sinha

Objective: Lending Club Case Study

- Understanding of risk analytics in banking and financial services using Exploratory Data Analytics and use it to minimise the risk of losing money while lending to customers.
- To identify the driving factors which influences the loan applicants defaulting the loan
- Given dataset Loan.csv has the information about past loan applicants with the loan status whether they 'defaulted' or not.

Analysis steps

```
        ┌─────────┐      ┌──────────────────────────┐      ┌──────────────────────────┐      ┌────────────────────────────────────────┐
        │  Start  │ ───▶ │ Understand the raw data  │ ───▶ │       Load the data      │ ───▶ │          Data Cleansing                │
        │         │      │ Go through loan.csv and  │      │ Load the excel data into │      │ Remove all customer behavior attributes,│
        └─────────┘      │ metadata file            │      │ dataframes for EDA       │      │ attributes with missing values and      │
                         └──────────────────────────┘      └──────────────────────────┘      │ clean-up the data.                      │
                                                                                             └────────────────────────────────────────┘
```

Start

Understand the raw data
Go through loan.csv and metadata file

Load the data
Load the excel data into dataframes for EDA

Data Cleansing
Remove all customer behavior attributes, attributes with missing values and clean-up the data.

Bivariate Analysis
Understanding the relationships between 2 variables by finding correlation and plotting different graphs

Segmented Univariate Analysis
Plotting boxplots and bar plots to understand the data distribution of variables over different segments

Univariate Analysis
Plotting histograms and bar plots to understand the data distribution of variables

Conclusions
Consolidate all the analysis from the above steps and draw inferences of which variables influence the loan applications to default

Stop

Exploratory Data Analysis

- As a part of data cleansing activity
  - All the customer behavior variables are dropped from the dataset.
  - All the columns which have the same value, missing values, null values in all rows are dropped
  - "Current" status rows are also dropped.
  - Some outliers are removed

- The cleansed data set had 38577 rows and 20 columns from 39717 rows and 111 columns.

- Target variable is loan_status and analysis is done to identify the patterns for 2 categories

  1. Charged off

  2. Fully paid

- After data cleansing, with the help of univariate, segmented univariate and bivariate analysis some of the observations were made and using all this info, we arrived at some conclusions.

Exploratory Data Analysis

Charged off customer segment Analysis:

Data analysis shows that loan amount, interest rate, Home ownership, Installment, bankruptcies records are prominent factor which impact this segment of loan seeker.

- There is clear relation ship which shows that customer tends to default the loan when interest rate is higher

- customer tends to default the loan when annual income is low

- Customer default loan when loan term is 60 months

- Loan applicant default the loan when verification status is unknown to bank.

Fully paid customer segment Analysis:

This segments of customer is having common behavior like
- High Annual income
- Less interest rate
- Loan term is 36 months
- No bankruptcies records
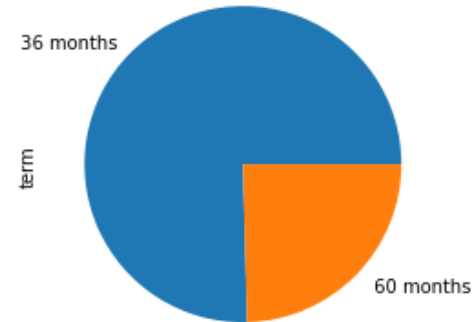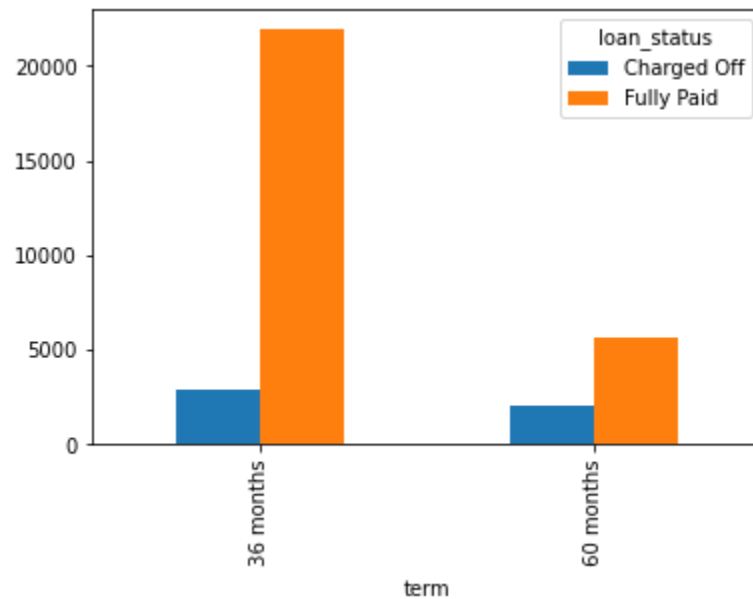- Source verified

Exploratory Data Analysis
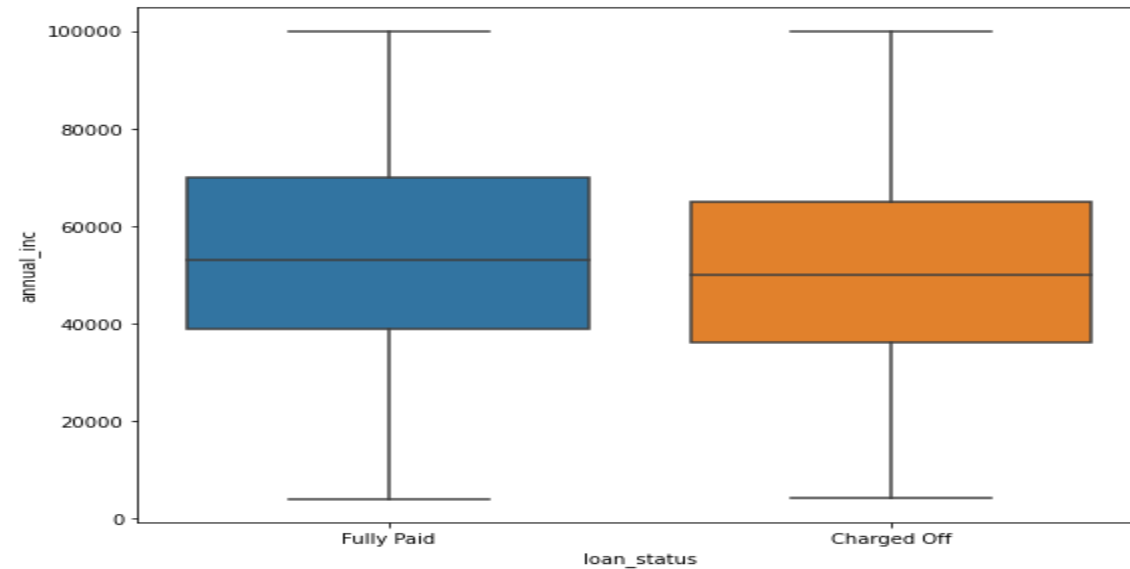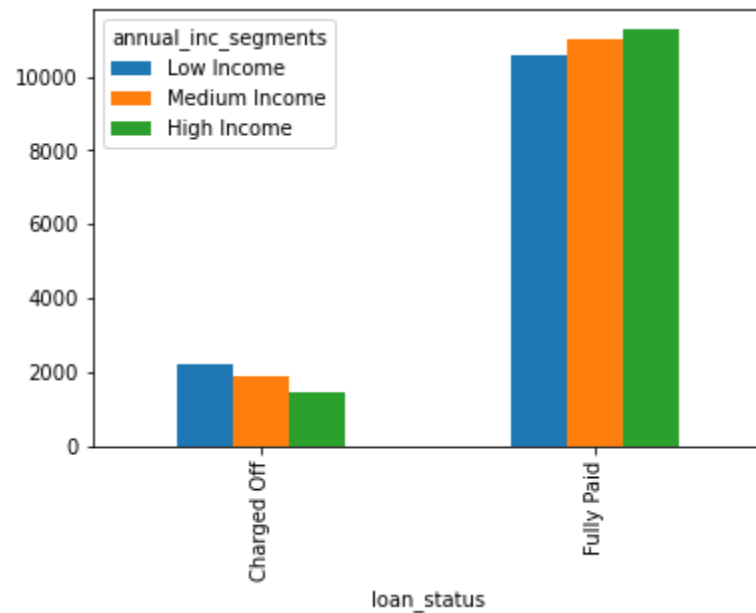
Looking at the graph, we can say that people with **lower incomes** tend to default the loans and higher income customer generally pay off the loans.
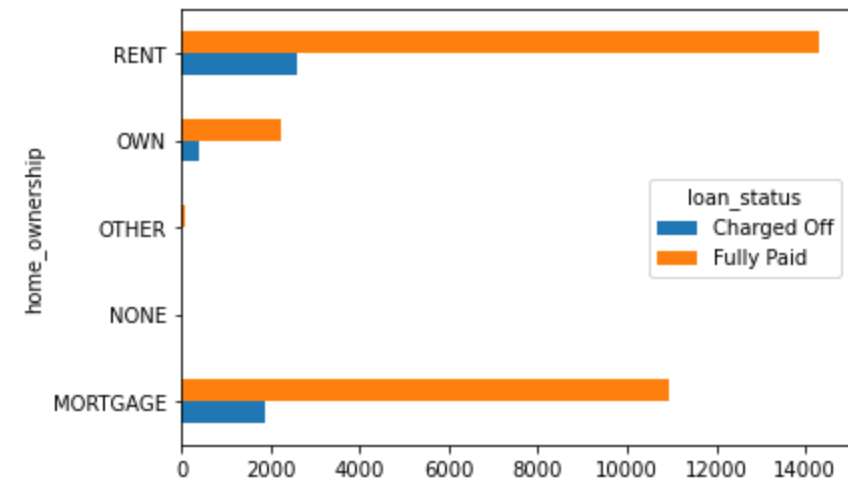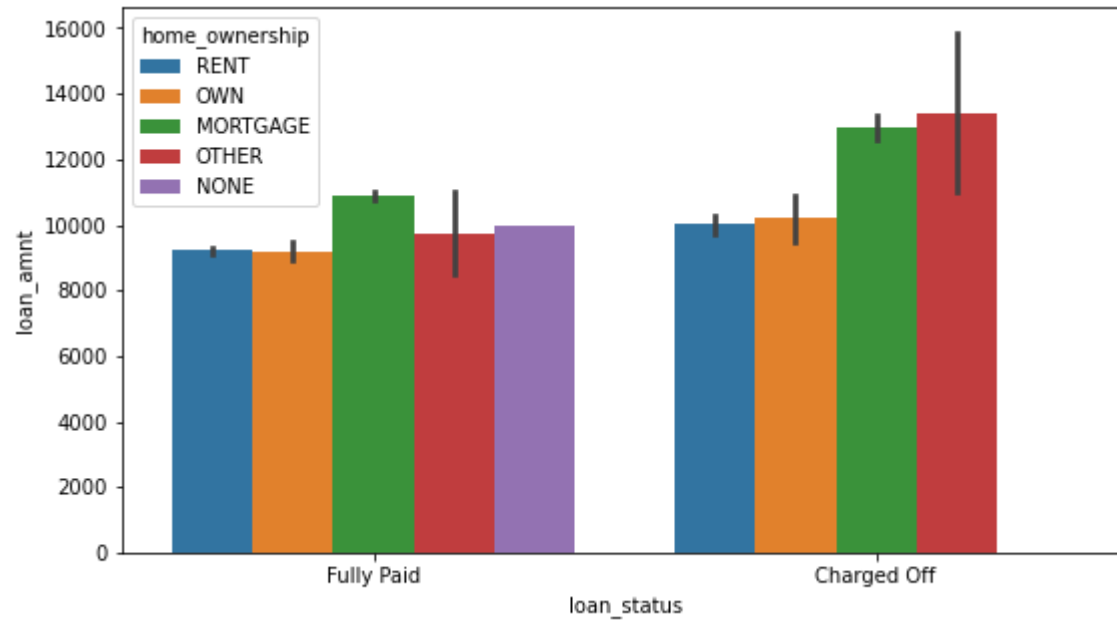
Exploratory Data Analysis
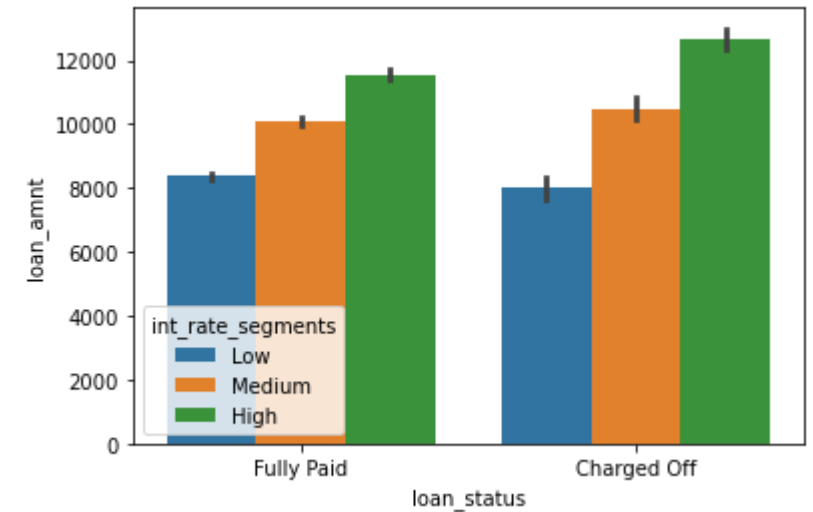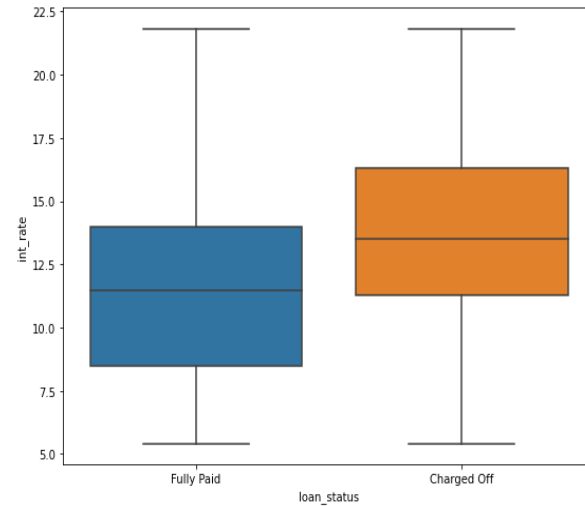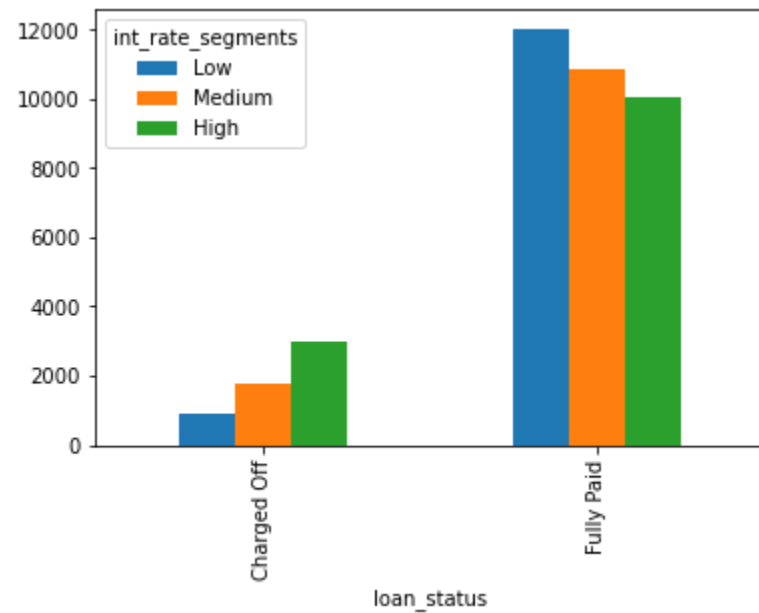
Customer tend to default the loans whenever the **interest rate** is higher comparing the above plots

Exploratory Data Analysis
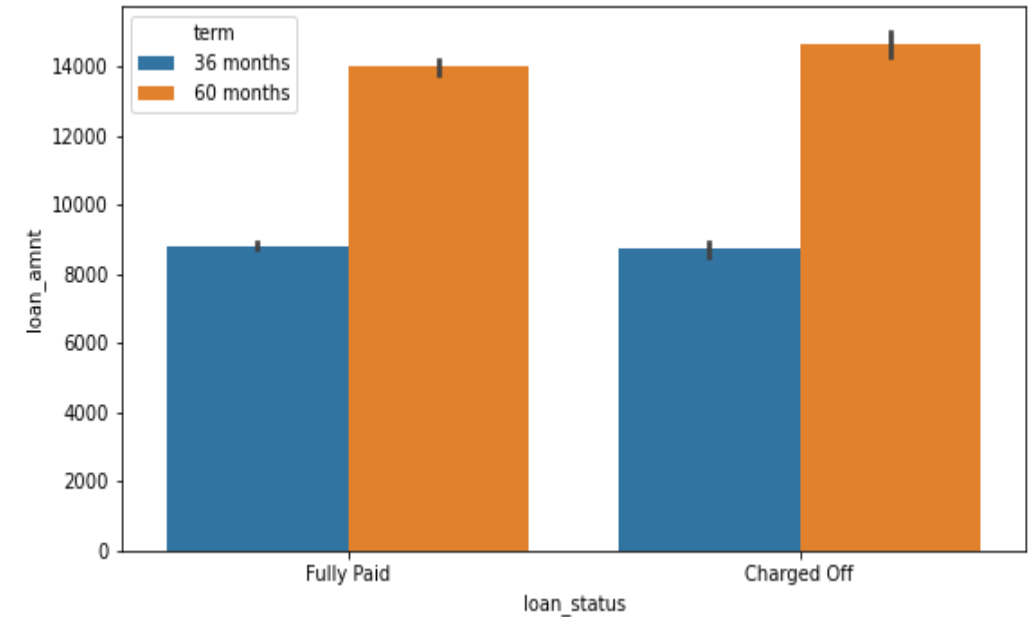
Loan applications are more for 36 months compared to **60 months**. 36 months loans are more likely to get paid off compared to the 60 month loans.

Exploratory Data Analysis

Home_ownership variable analysis

- Loan amount, funded amount, funded amount investor seems to have high **positive** correlation (**0.98, 0.93, 0.95**) indicating that when one variable value increases then the other variables values also increase.
- Loan amount and interest rate also seems to have high positive correlation of **0.93**
- Loan amount and installment also seems to have high **positive** correlation of **0.93**
- Loan amount seems to have slightly **positive** correlation with interest rate, annual income, DTI
- installment and interest rate also seems to have **positive** correlation of **0.24**
- installment and annual income seems to have positive correlation of **0.37**
- Annual income to dti seems to have slightly negative correlation
- Annual income seems to have slightly positive correlation with loan amount, interest rate, installment
- Public bankruptcies record and loan amount is **Negatively** Correlated
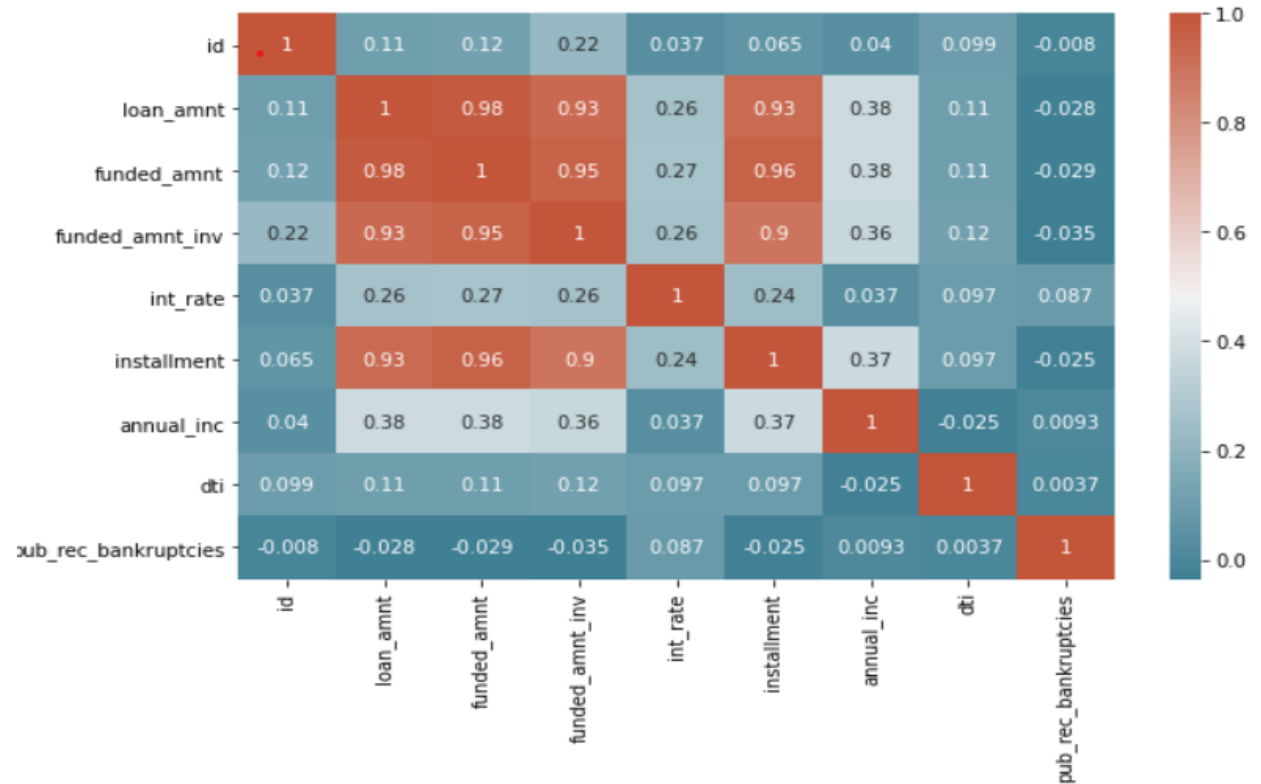
Conclusions

Based on our analysis of the LendingClub data we can conclude the following:

➢ Variables which influence the loan_status are:
- Loan_amount
- Int_rate
- Installment
- Term
- Purpose
- Emp_length
- DTI
- Grade
- Verification_status

➢ Some of the customer trends observed when the load is defaulted:
1. Interest rate is higher around 10-14%
2. Installments are higher
3. Customers with 10+ years of employment
4. DTI is around 10 to 18
5. belong to "B" Grade and "A4" Sub-Grade
6. loan applicants stay in rented properties
7. Debt consolidation and credit card card repayment is major purpose of Loan
8. Customer tends to default the loan when annual income is low
9. Customer default loan when loan term is 60 months
10. Loan applicant default the loan when verification status is "Not Verified".
11. higher loan amount and purpose seems to be 'small business'
12. belong to grade "F" when their loan amount is high
13. high loan and more bankruptcies records on their name
14. higher loan amount and their homeownership details are specified as "Other"
15. higher DTI having mortgage