

# GRAMENER CASE STUDY

## SUBMISSION

Group Name:

1. Lipsa Satapathy
2. Prabhudatta Praharaj
3. Abinash Panda
4. Ankit Gupta

# Exploratory Data Analysis Case Study Abstract

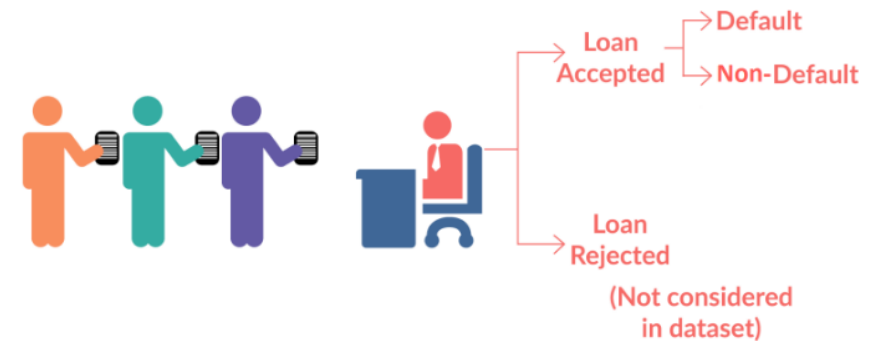
## Business Understanding

A consumer finance company which specializes in lending various types of loans to urban customers.

Wants to analyze the risk for loss of business and financial loss based on existing loan data set .

For given loan there could be 3 status .

- **Fully paid** - Money recovered with interest
- **Current** - Installment in progress and expected to recover full amount
- **Charged-off** - Delay in payment and high risk of Default



## Problem Statement

- Understand and identify the driving factors (consumer & loan attributes ) influence the tendency of default.
- Identify risky loan applicants, thereby cutting down the amount of credit loss.

## Data Understanding

### Customer Demography

- Employment Length
- Zip & State
- Annual Income
- Home Ownership
- Verification Status
- Average Current Balance

### Loan Attributes

- Loan Amount
- Funded Amount
- Interest rate
- Grade
- Installment
- Issue Date
- Principal Amount

### Customer Behavior

- Purpose
- Application Type
- Revolving
- Delinquency
- No of accounts
- Earliest Credit Line
- Public Record

# Exploratory Data Analysis - Data Cleaning and Preparation

We have analysed and have done the following data cleaning activities

## Remove NA

- We have removed 54 columns with only NA values.

## Remove Blank

- Removed “next\_pymnt\_d” column as it had blank values, and only valid values for current loans.

## Remove Duplicate

- Checked duplicates for “id” and “member\_id” columns but didn’t find any duplicated value.

## Data Formatting

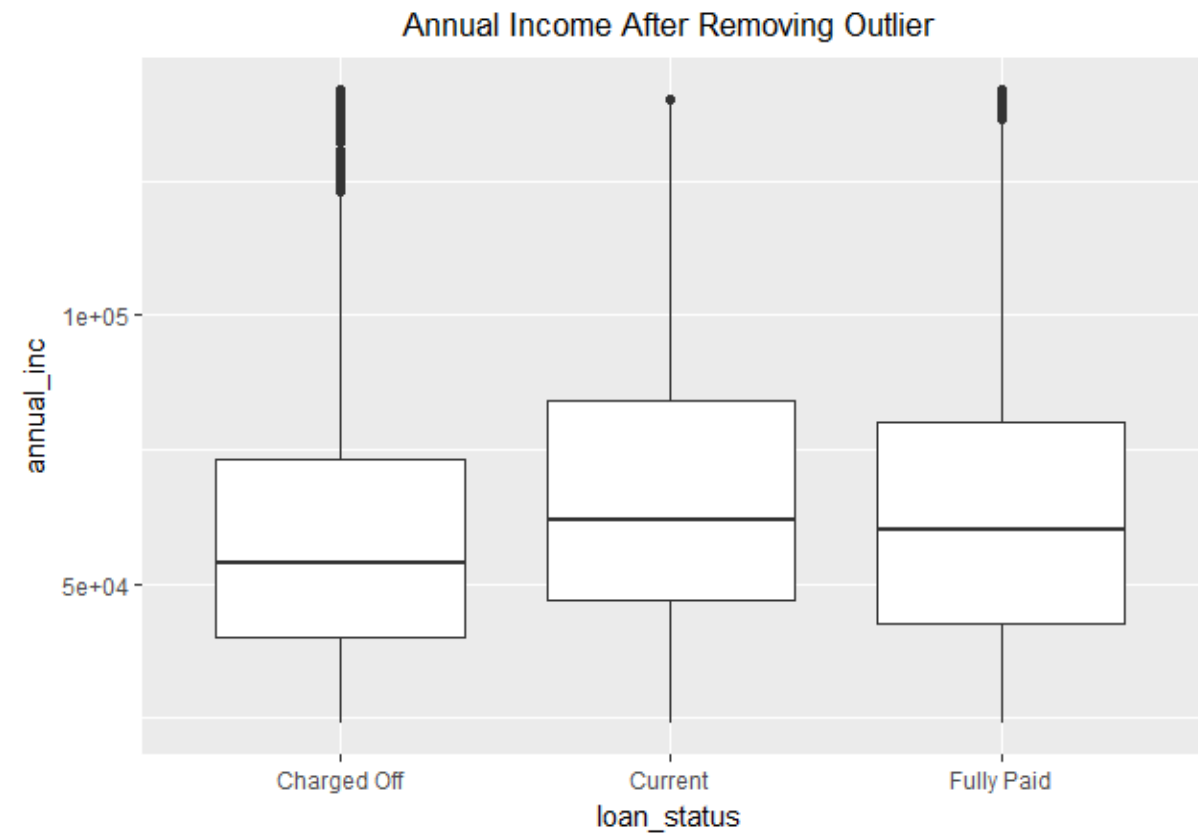
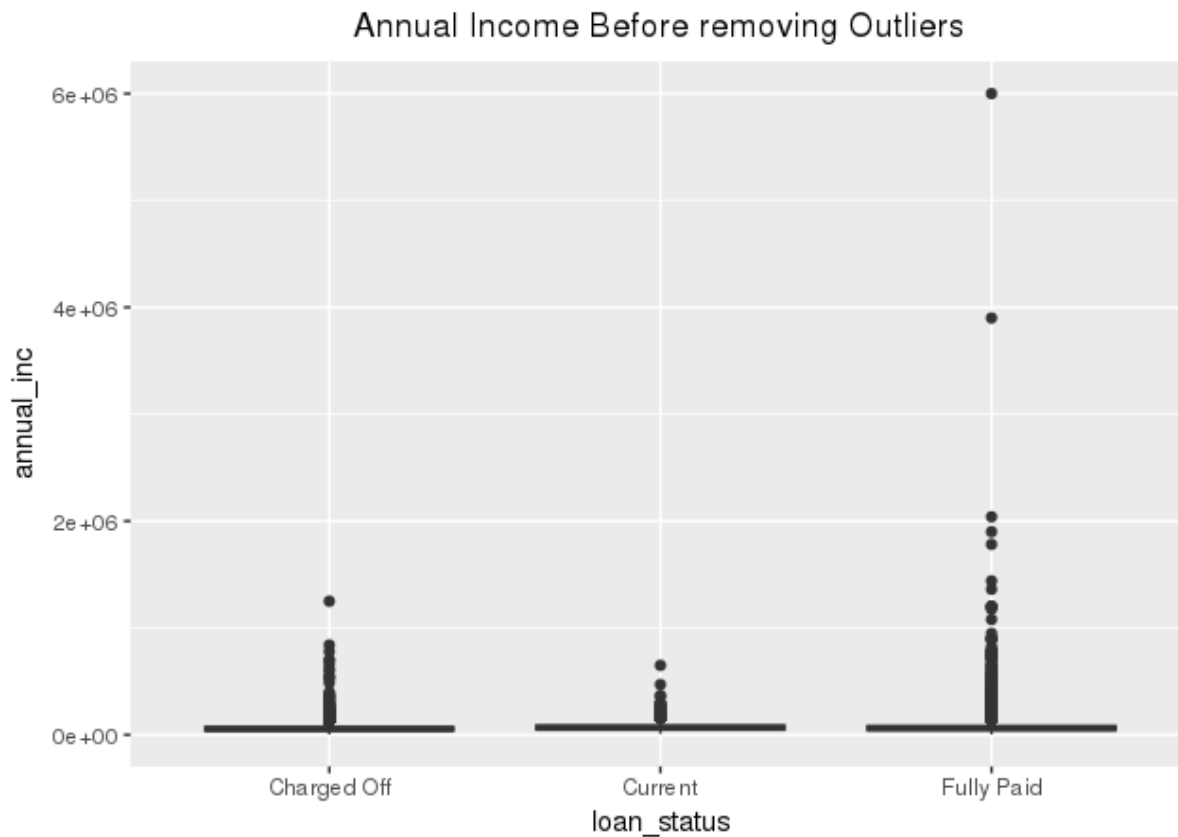
- Formatted date columns.
- Formatted title column with upper case and removed punctuations.
- Removed “%” and converted to number.

## Remove Single Value

- Removed columns with single values.

# Exploratory Data Analysis– Outlier Detection

- We have detected and removed the Outlier for Annual Income.



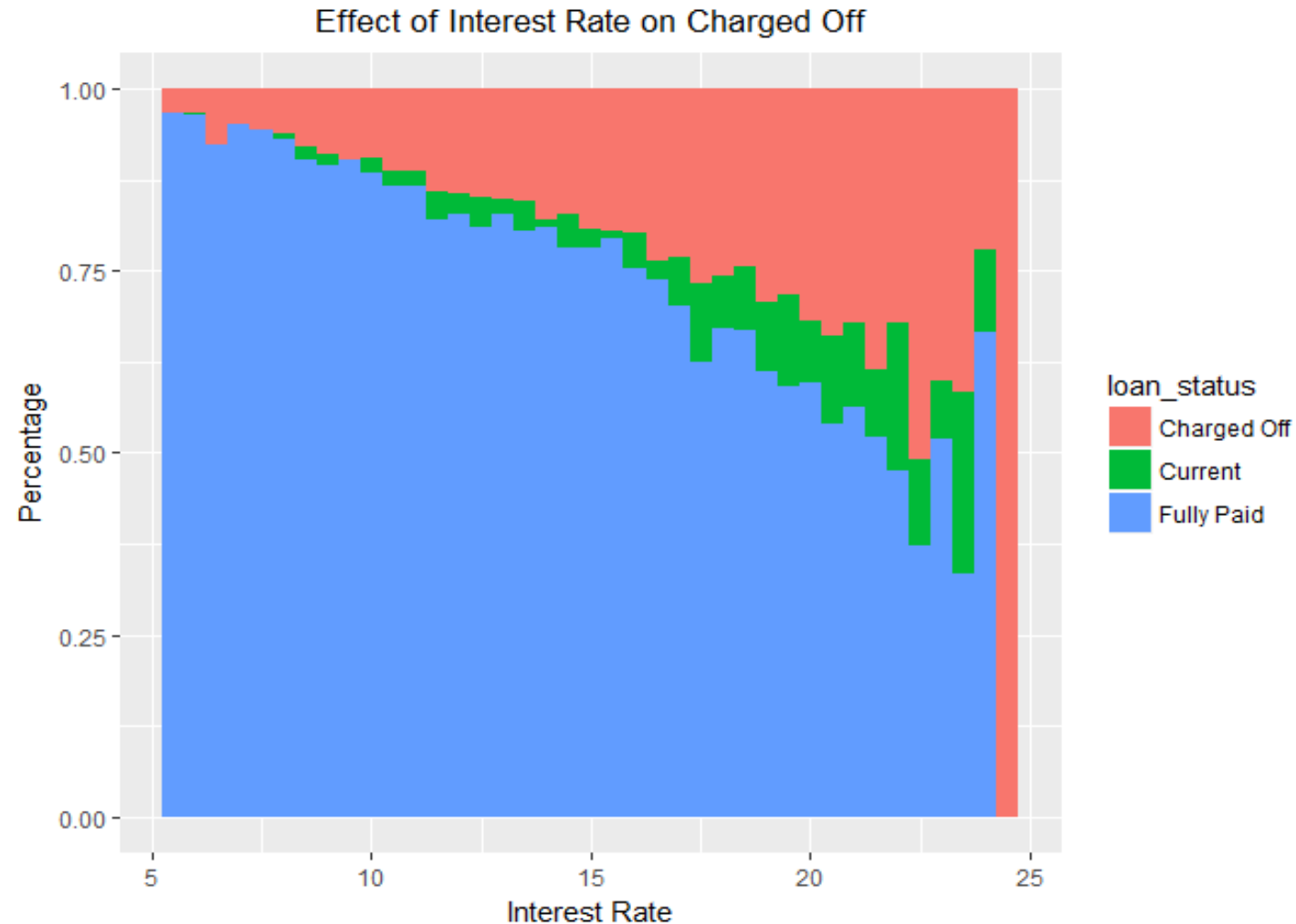
# Exploratory Data Analysis : Univariate Analysis:

## Plot

- We have plotted the relationship between Interest Rate and Loan Status.

## Observation

- As the Interest Rate increases, likelihood of Charged off increases.



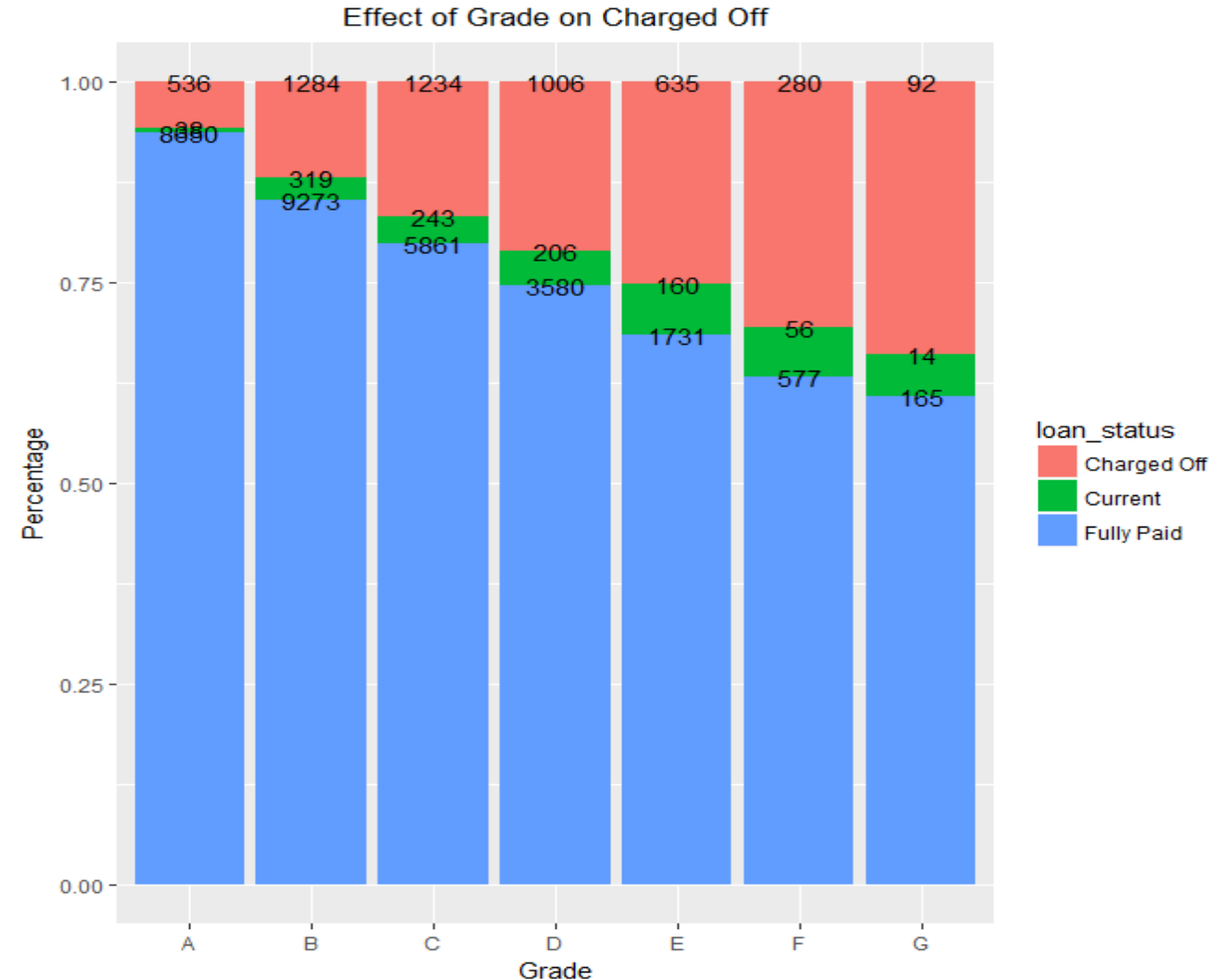
# Exploratory Data Analysis : Univariate Analysis:

## Plot

- We have plotted the relation between Loan Grade and Loan Status

## Observation

- Grade and Charged off are inversely proportional.
- The number of loans granted reduces as the Grade increases.
- Risk of Charged off is high beyond Grade “C”.



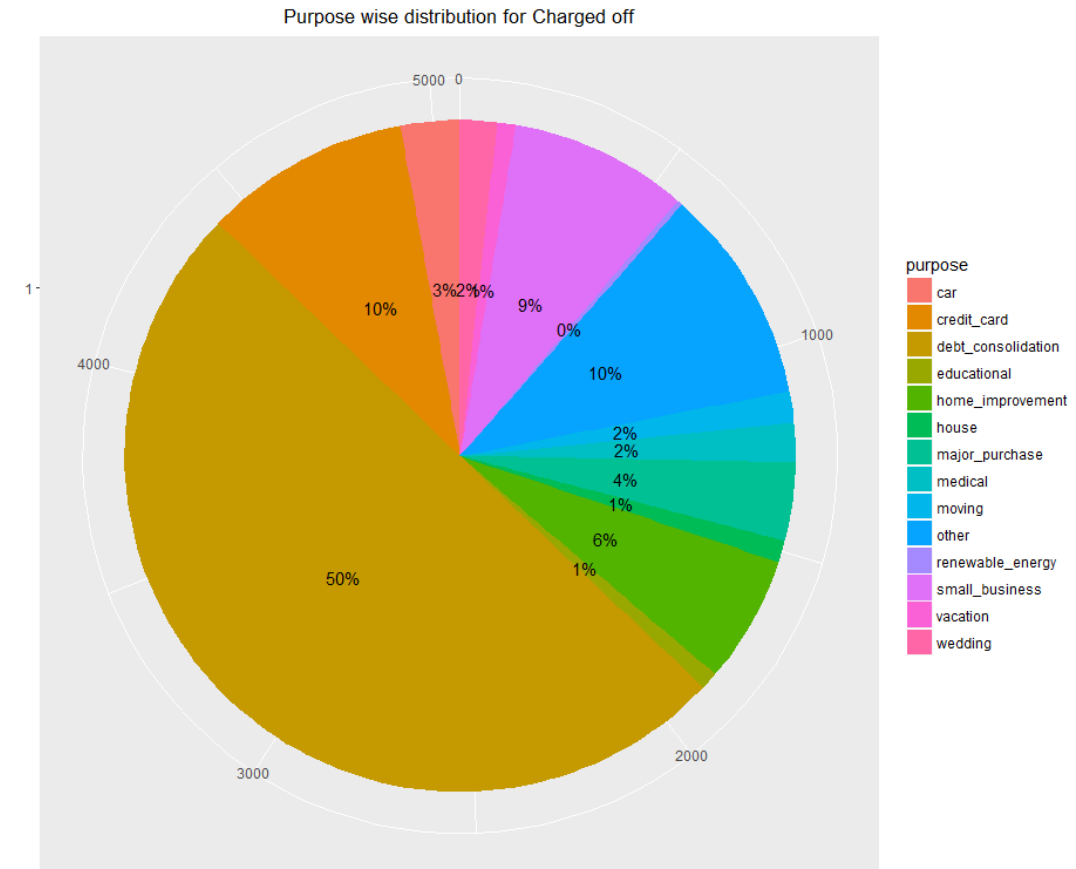
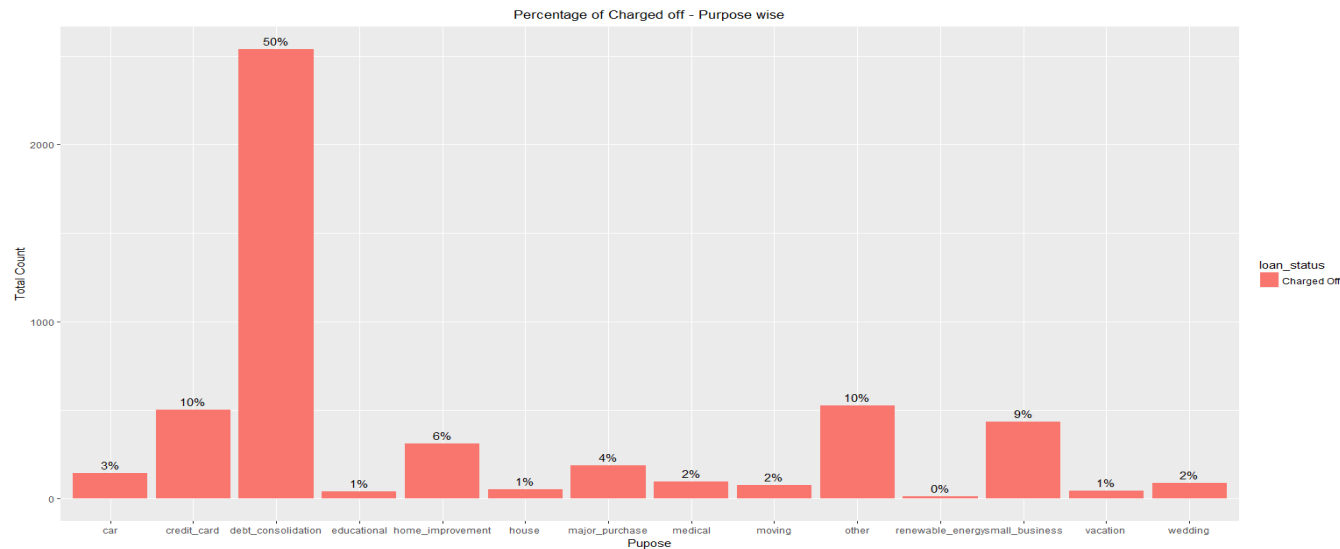
# Exploratory Data Analysis : Univariate Analysis:

## Plot

- We have plotted relation between Loan purpose and Charged off.

## Observation

- Purpose of loan is significantly related to charged off percentage.
- Following 4 sectors to be observed critically before granting loan:
  - Debt Consolidation
  - Small Business
  - Credit Card
  - Others





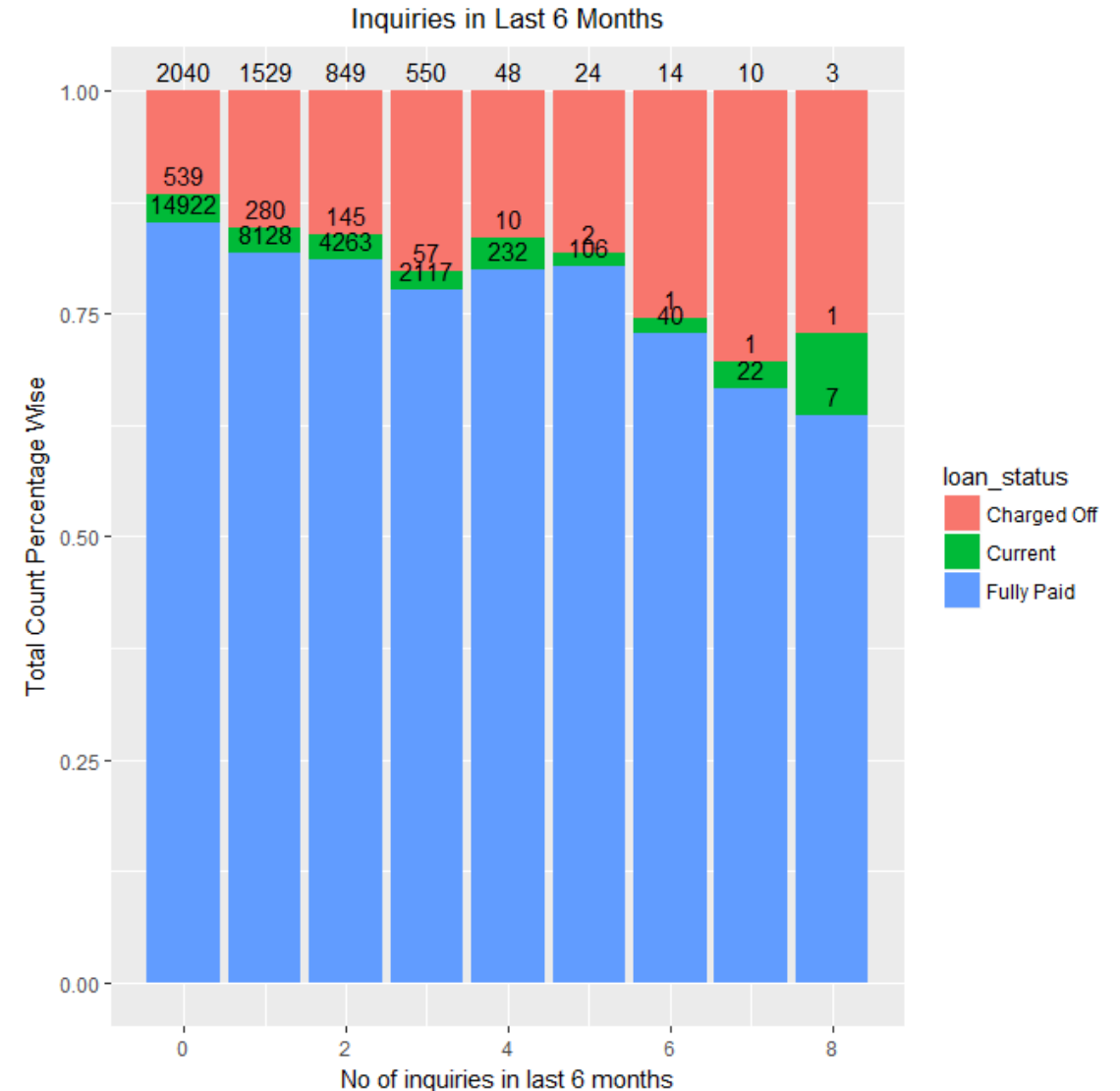
# Exploratory Data Analysis : Univariate Analysis:

## Plot

- We have plotted the relationship between No of inquiries in last 6 Months and Loan Status.

## Observation

- As the number of inq\_last\_6mnths increase the likelihood of loan getting defaulted also increases.



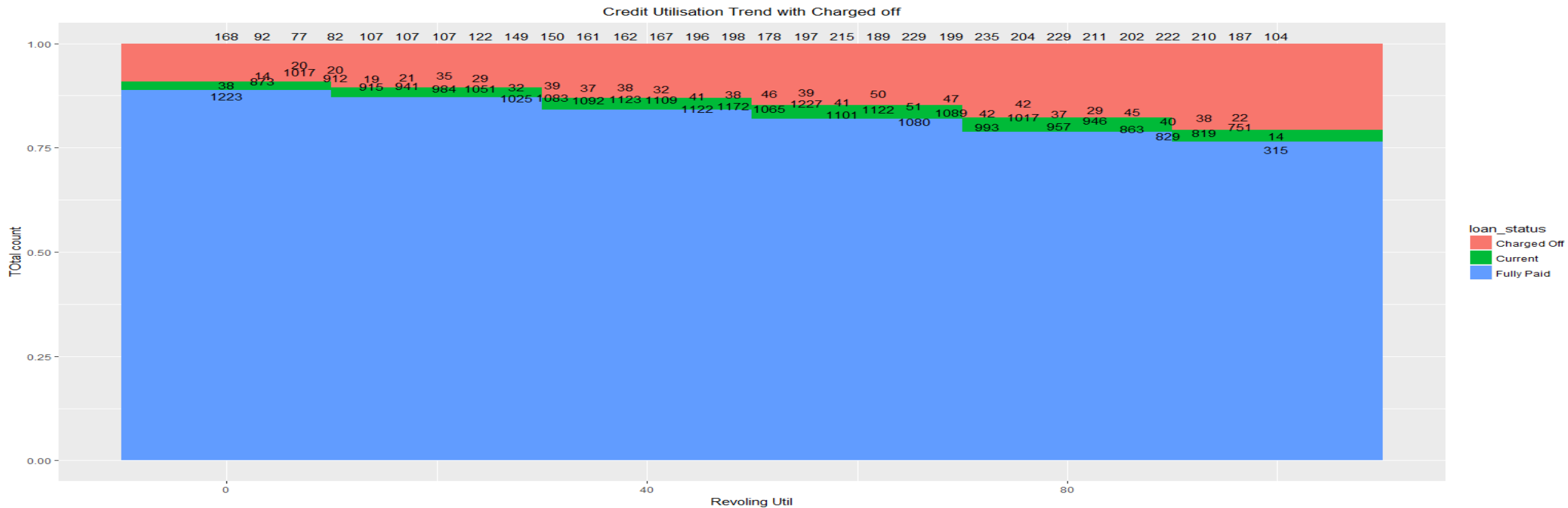
# Exploratory Data Analysis : Univariate Analysis:

## Plot

- We have plotted the relationship between Revolving Utilization and Loan Status.

## Observation

- Credit Utilization is correlated with the Charged off Status.
- As the revolving utilization increases, the risk of charged off increases along with that.



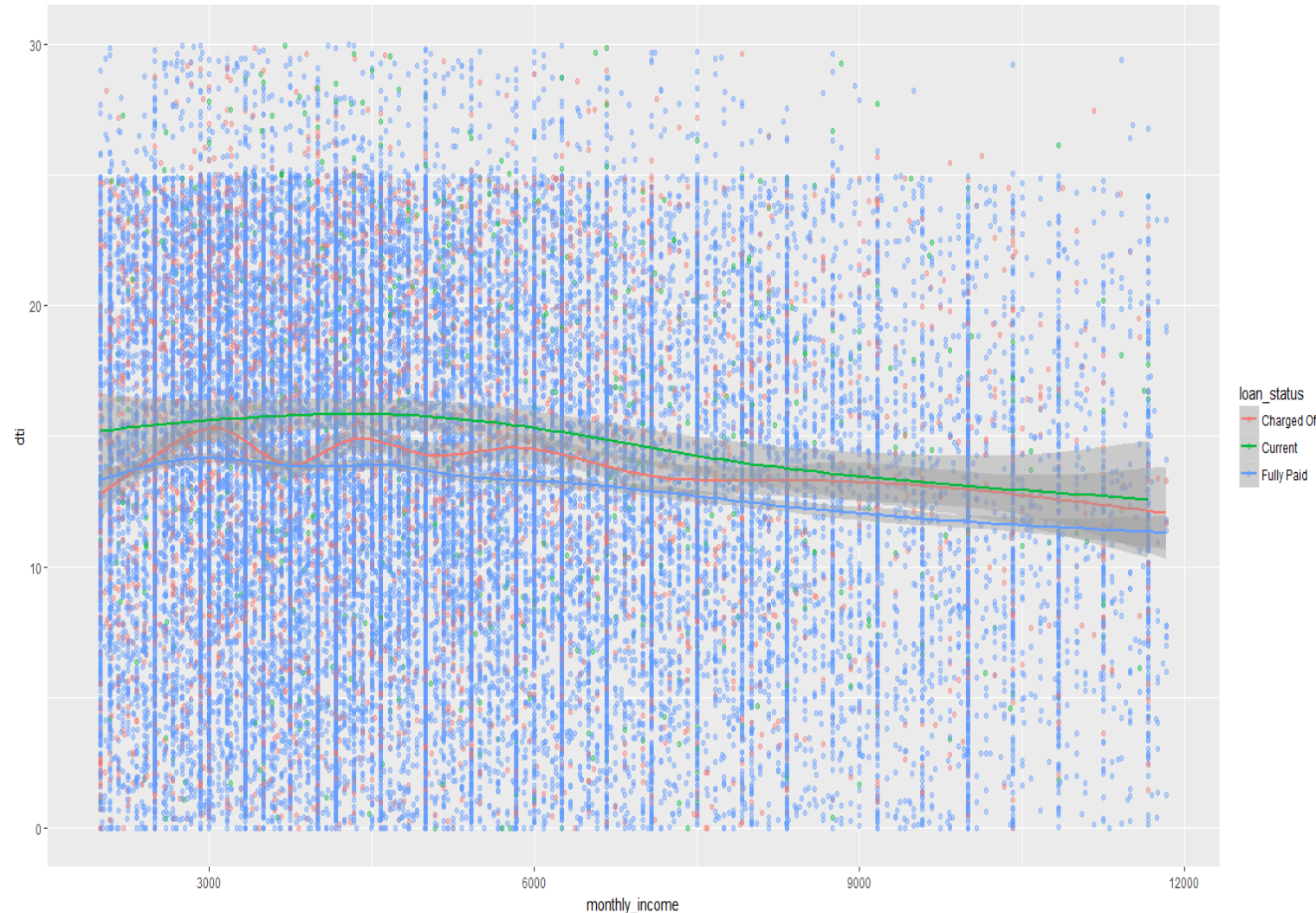
# Exploratory Data Analysis: Derived Variable

## Plot

- We have derived the monthly income from annual income and plotted a correlation between “**dti**”, “**Monthly Income**” and “Loan Status”.

## Observation

- Above 25% dti, no of loans granted are less.
- No loan was given above 30% dti.
- In the lower middle Income Group there is a spike in defaulters with above 14% DTI.



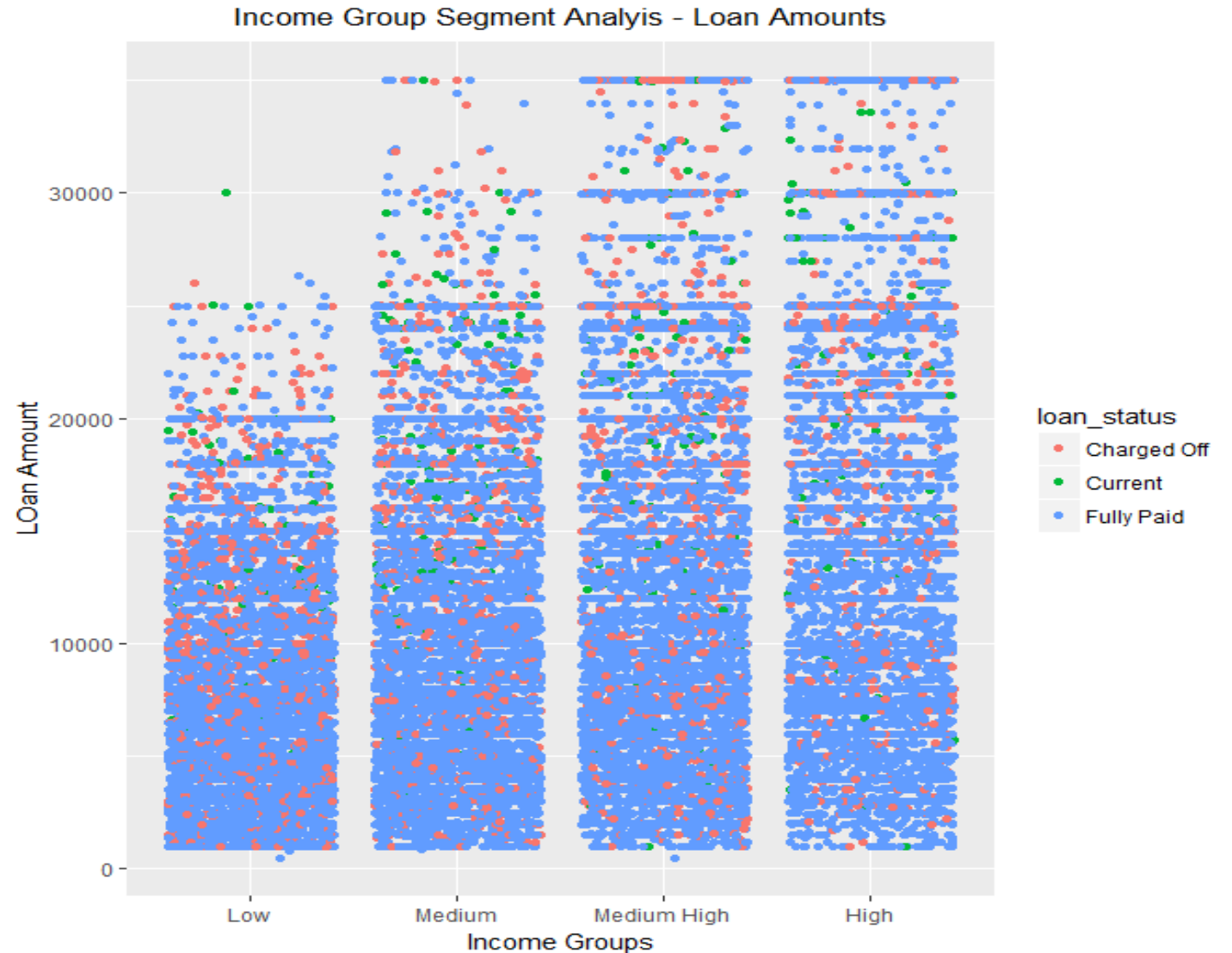
# Exploratory Data Analysis: Segmented Analysis

## Plot

- We have created segments of income groups based on annual income.
- We have plotted a graph between income group and Loan Amount.

## Observation

- Individuals with low and medium income with large loan amount tend to default.



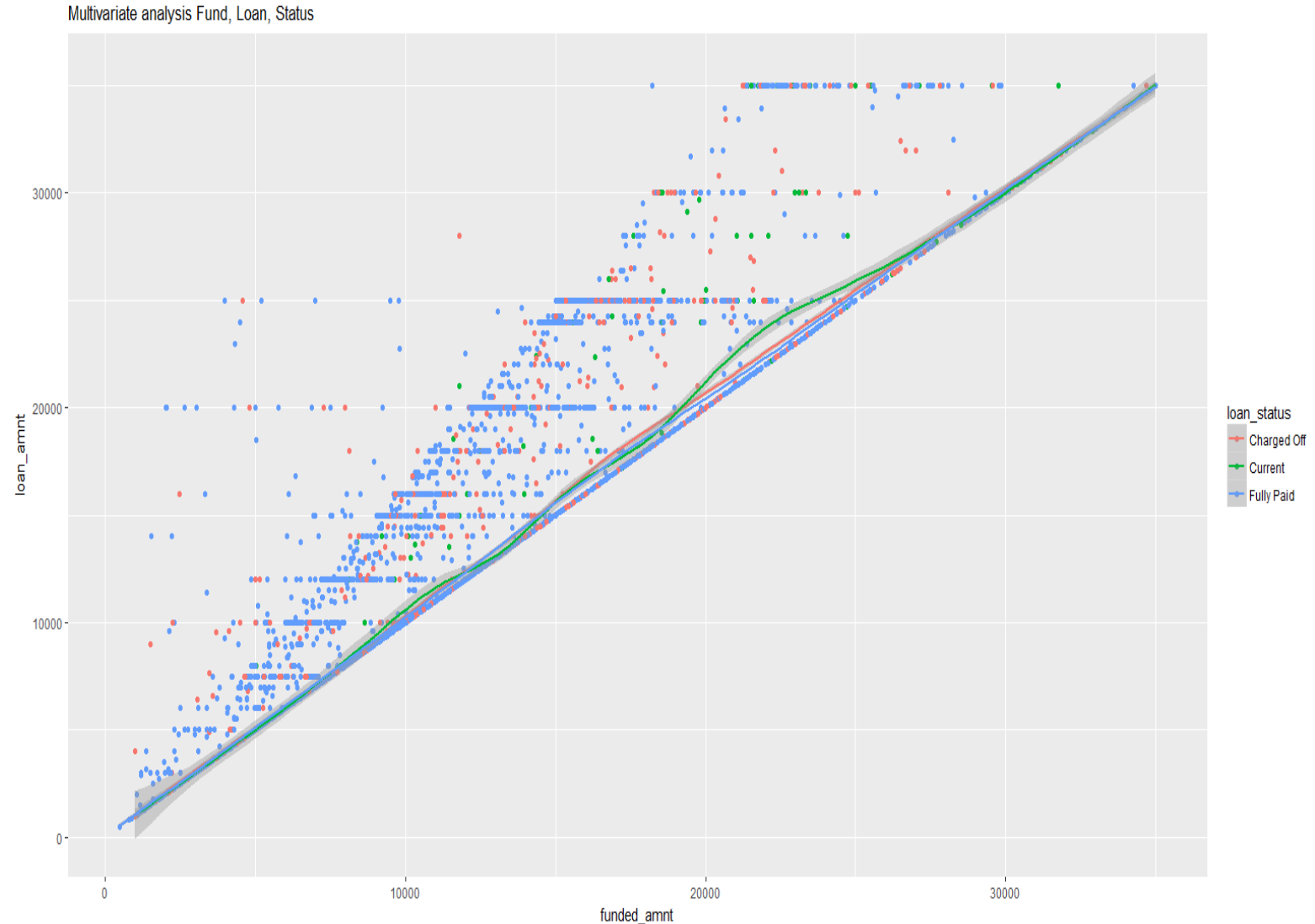
# Exploratory Data Analysis: Multivariate Analysis

## Plot

- We have plotted relation between Fund amount and Loan Amount.

## Observation

- There is a linear progression between Fund Amount and Loan Amount.
- Fully paid and Charged off are equally distributed across the graph.



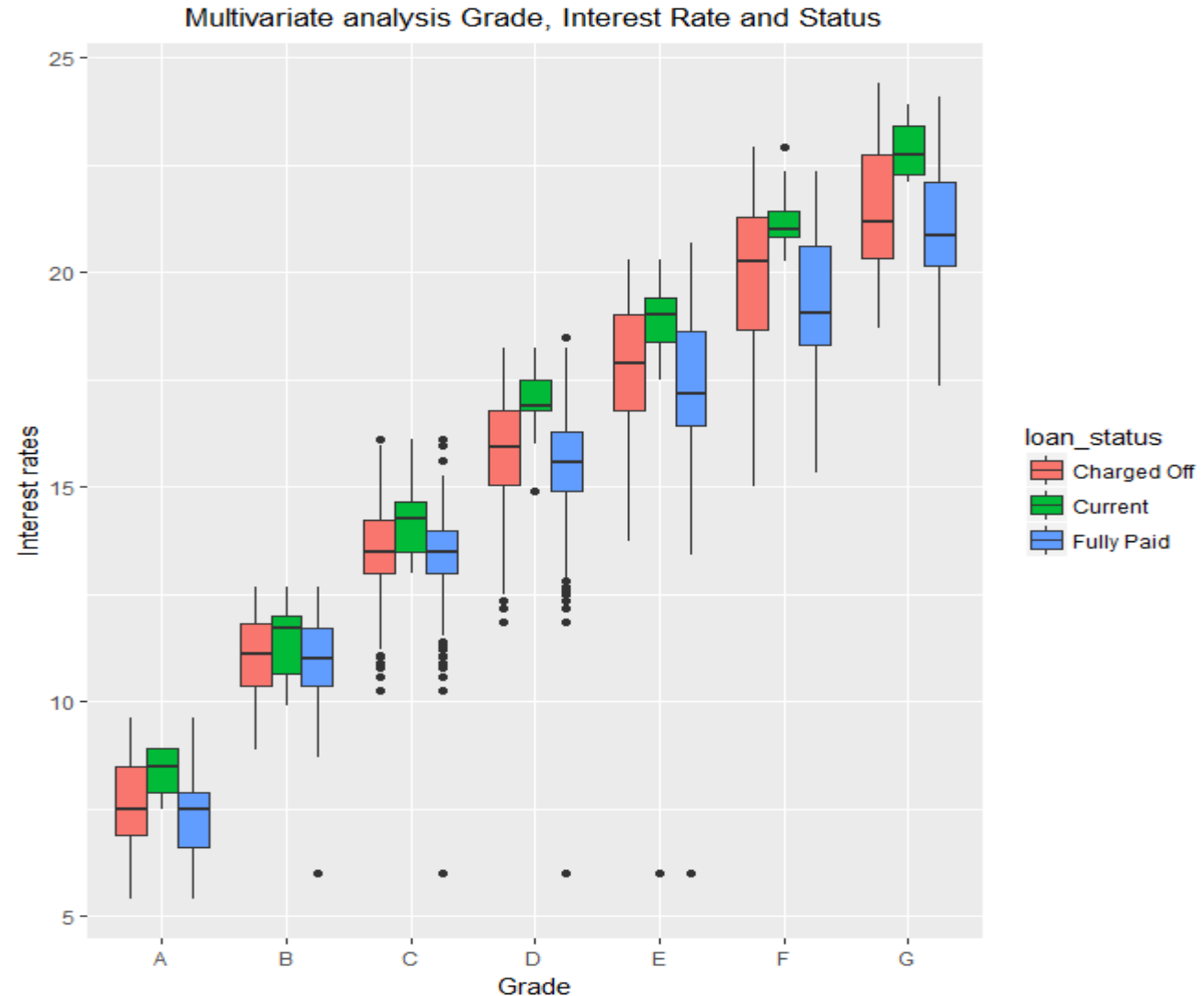
# Exploratory Data Analysis: Multivariate Analysis

## Plot

- We have plotted a box plot across Grade and interest rate for various Loan Status.

## Observations

- Interest Rate for Charged off Accounts are higher than Fully Paid Loans.
- Under the same Grade, individuals with higher interest rate tend to default their loan.



## Conclusion:

Based on the observations we are concluding the following variables are playing significant role leading to charged of scenario, hence recommend to do further verification for the following criteria's:

- **Loan Purpose** –
  - For the purpose of Debt Consolidation, Credit Card, Small Business and Others, more cautious approach needs to be taken before granting loan.
- **Grade** –
  - For Loans given for Grade beyond “C” are more prone to being charged off.
- **Inquiries in Last 6 months** –
  - Higher the number of inquiries in past 6 months indicates possibilities of delay in payment which leads to charged off.
- **Grade and Interest Rate** –
  - Combination of High Grade and High Interest Rate are more prone to charged off.
- **Revolving Utilization** –
  - Higher is the revolving utilisation, higher is the chance of charged off.