

# LENDING CLUB CASE STUDY

## GROUP - SUBMISSION

Name:

- 1) Kaliraj Balakrishnan
- 2) Tapas Kundu

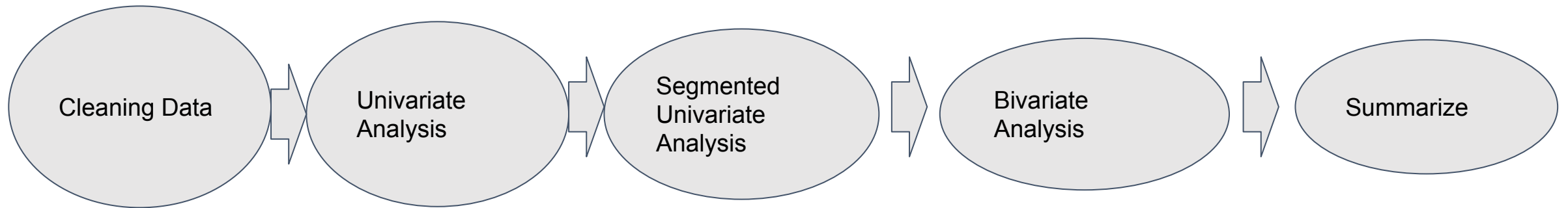
# Problem Statement:

**Lending Club** is the largest online loan marketplace facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

As a data scientist, we need to analyse the existing data set from Lending club and understand how customer or loan attributes can contribute to loan defaults.

# Methodology



# Methodology - Cleaning Data & Filling missing values

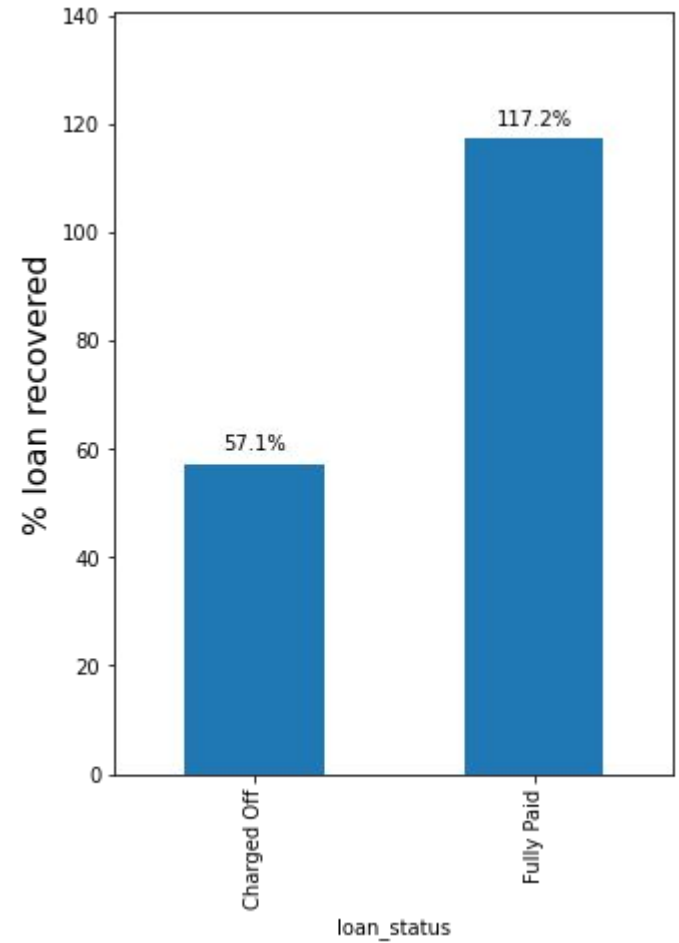
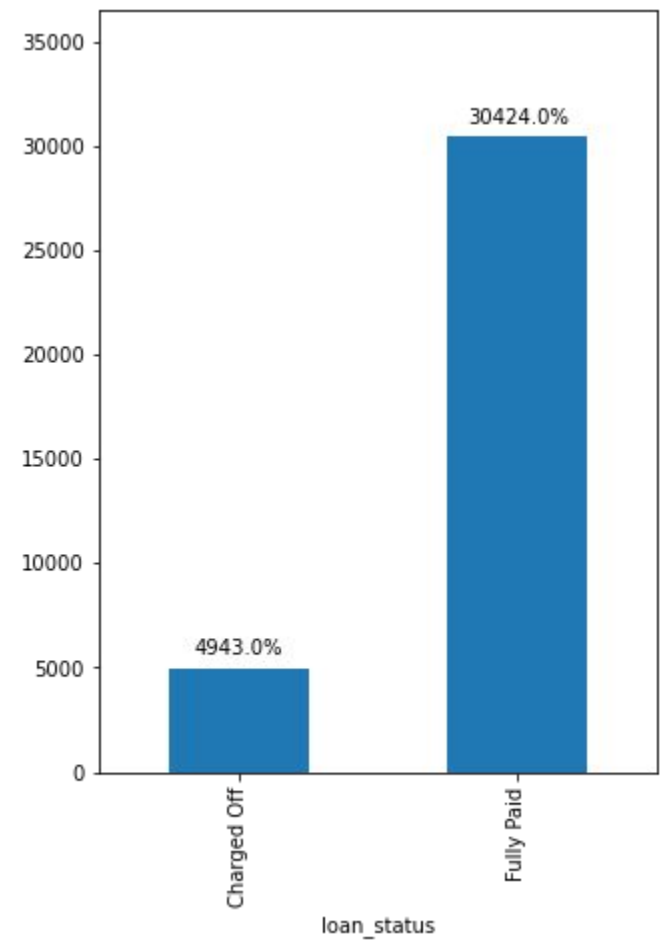
- Drop columns having followings:
  - Null Values.
  - Random values like id, etc doesn't contribute to analysis.
  - Single category variable.
- Convert values to right type for comparisons:
  - Strings to int (8 years to 8)
  - Removed extra characters like %.
  - Broken date column and derived month and year column for better analysis.

## Analysis - Understanding Loans

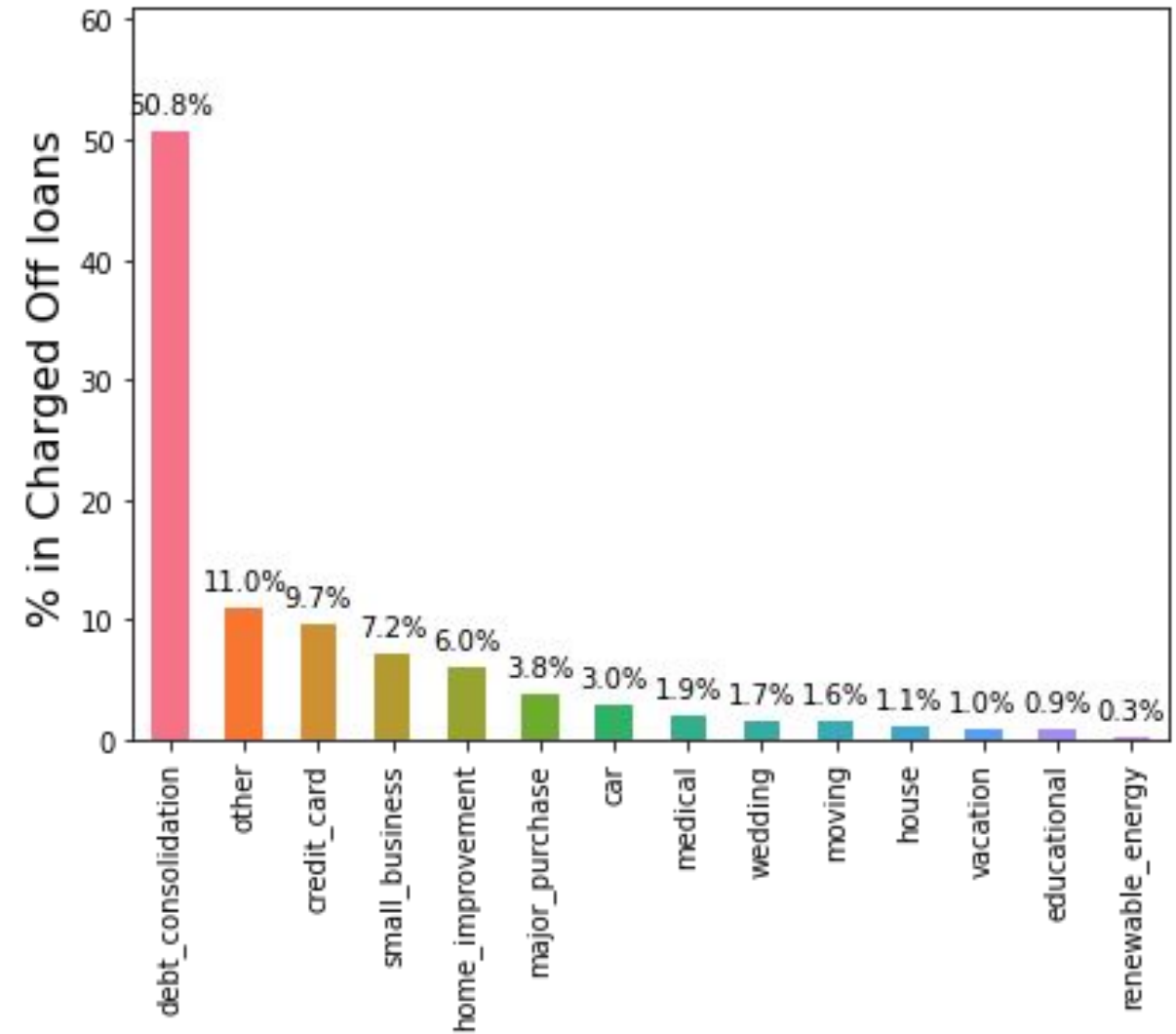
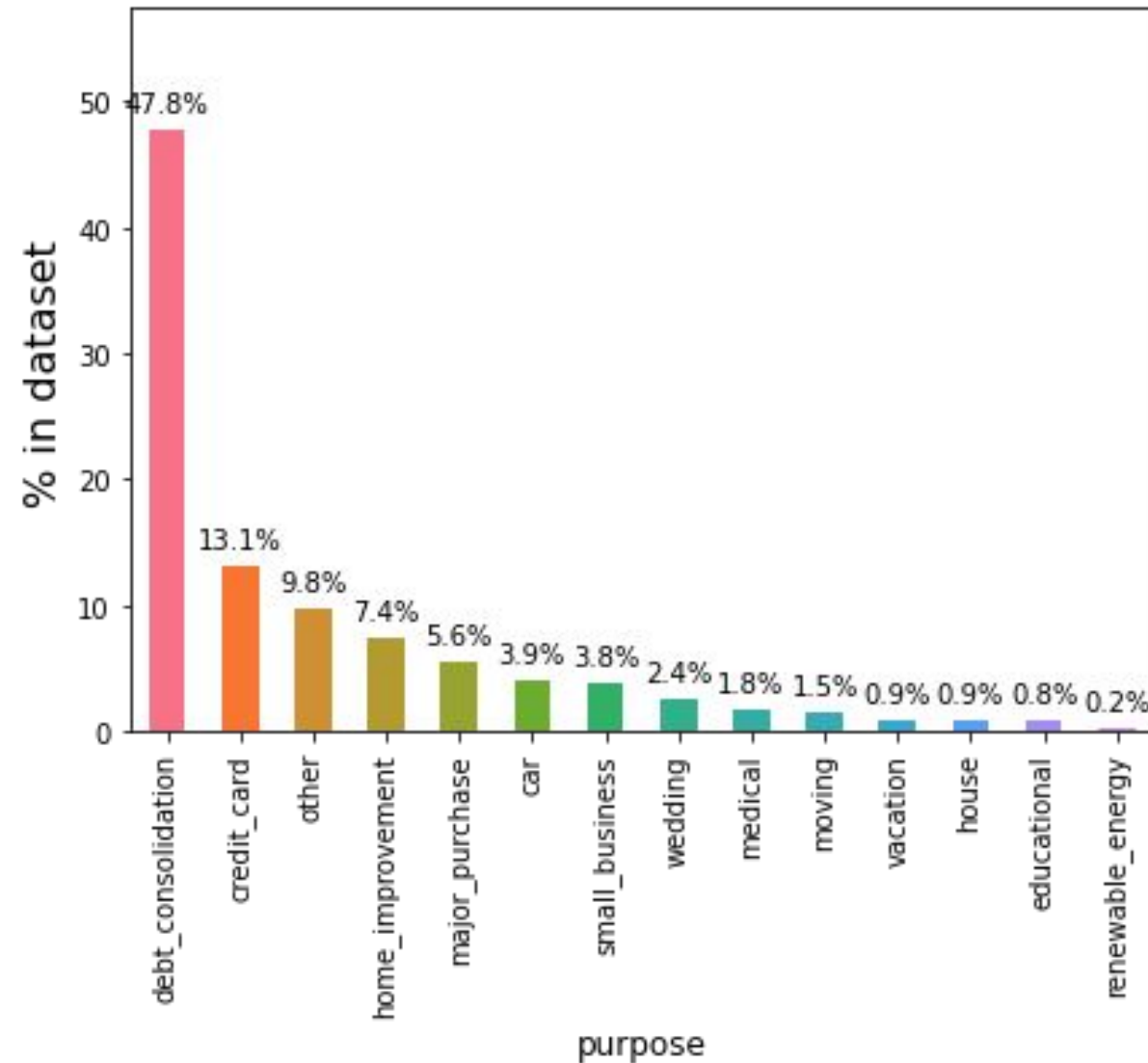
- Check distribution and frequencies of numerical and categorical values.
- Did univariate, bivariate analysis against segment of other variables.
- Maximum number of loans are for debt consolidation, followed by credit card.
- Lower grades have higher incidence of defaults on loans.
- Lending club charges higher interest rate as the grade of loan become worse.
- Lending club has extended high amount loans with low income.
- Included plots for visualisations and summarise the most important results

# Analysis - Overall Loan Status

- Approximately **14%** of loans are defaulted
- Lending Club only recovers **57%** of the loan amount when loans are defaulted. On fully paid up loans, the company makes **17%** profit.

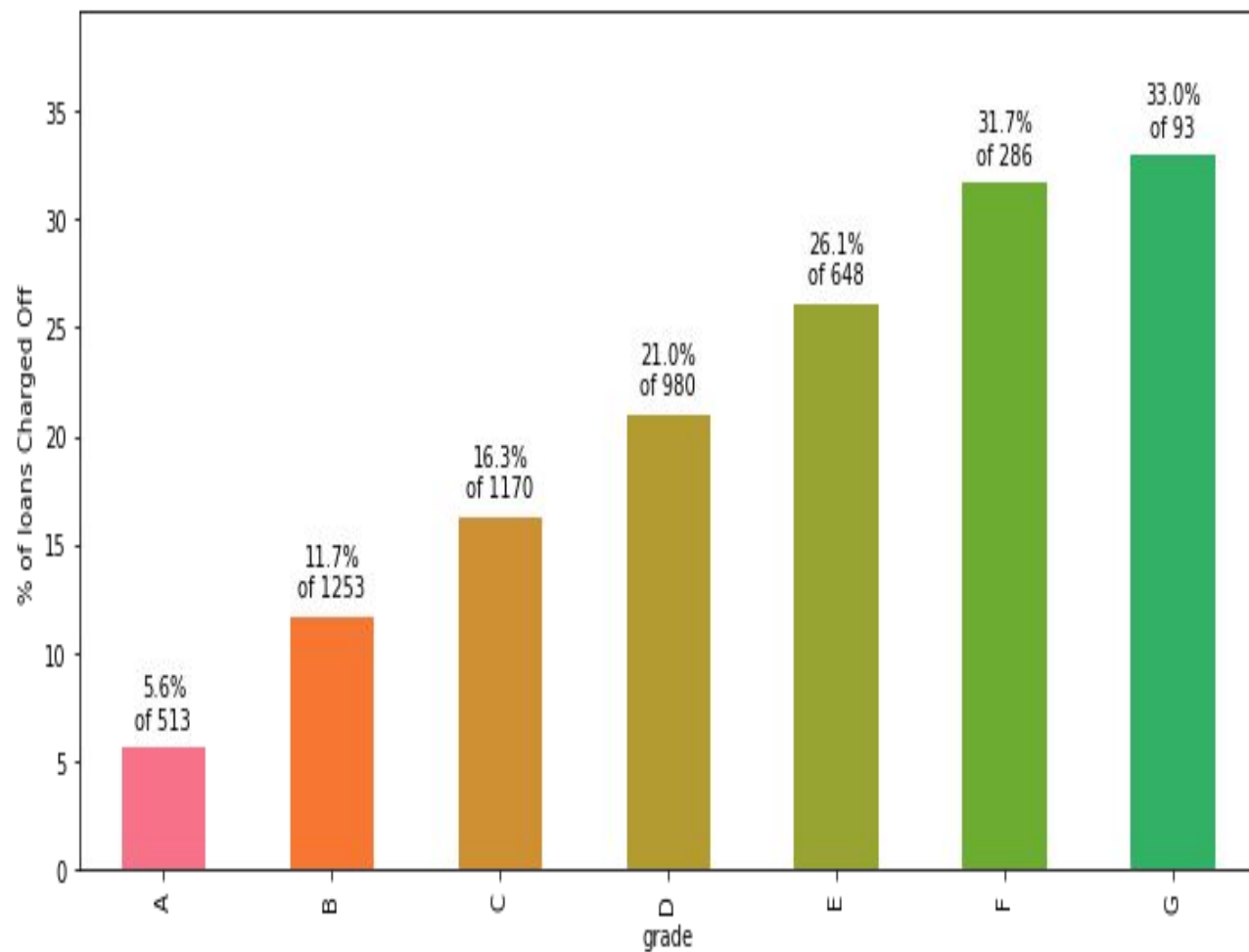


## Result : Plots for purpose

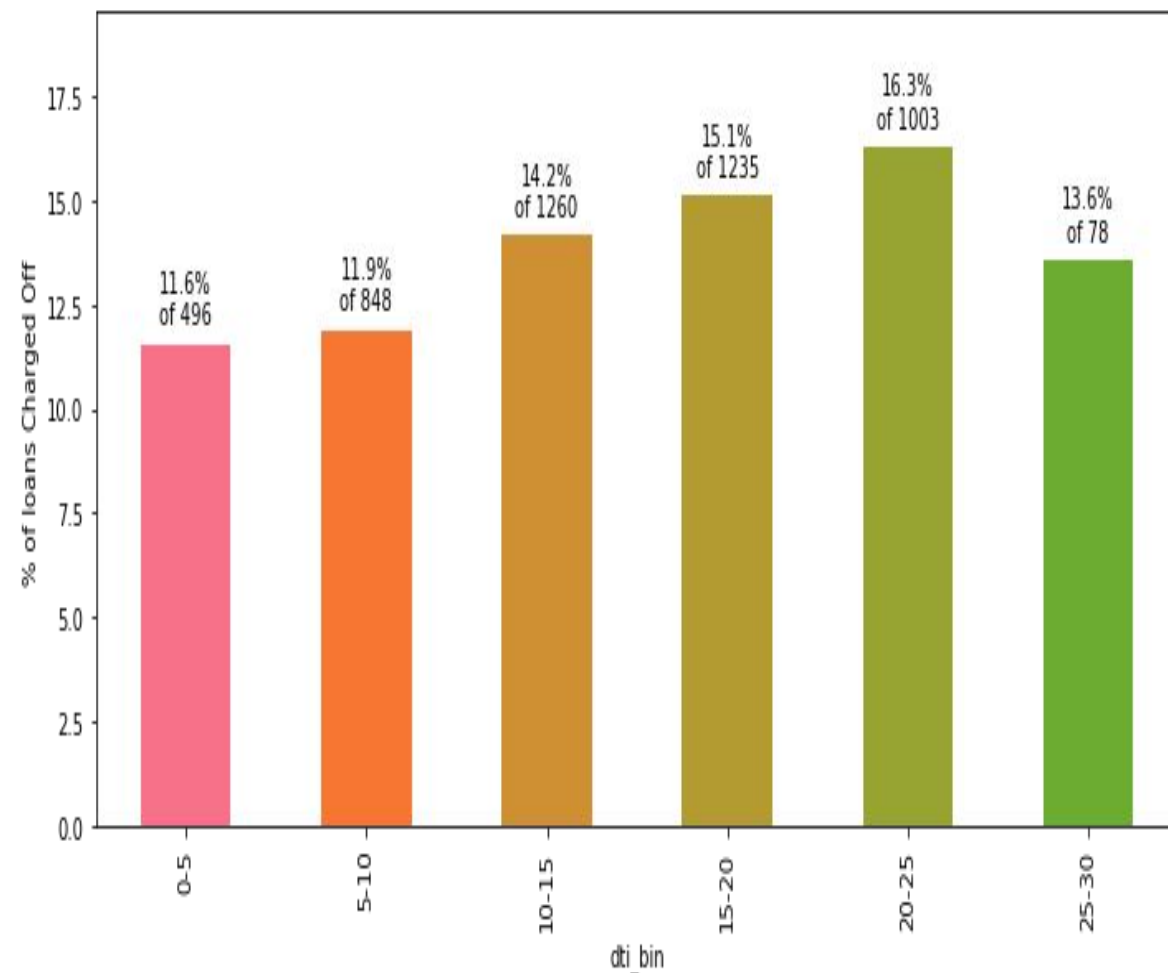


## Result : Plots

**grade vs Charged Off**



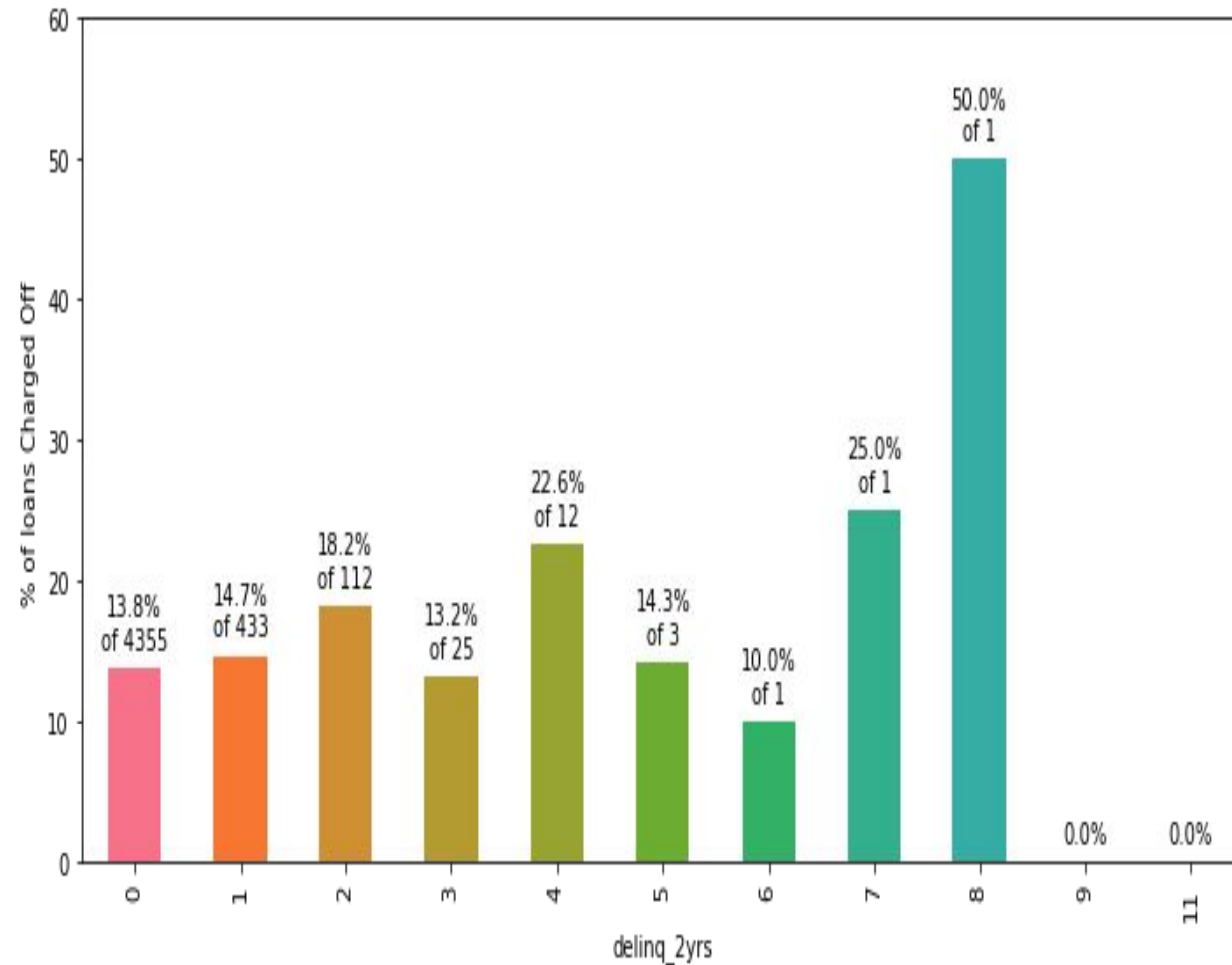
**dti bins vs Charged Off**



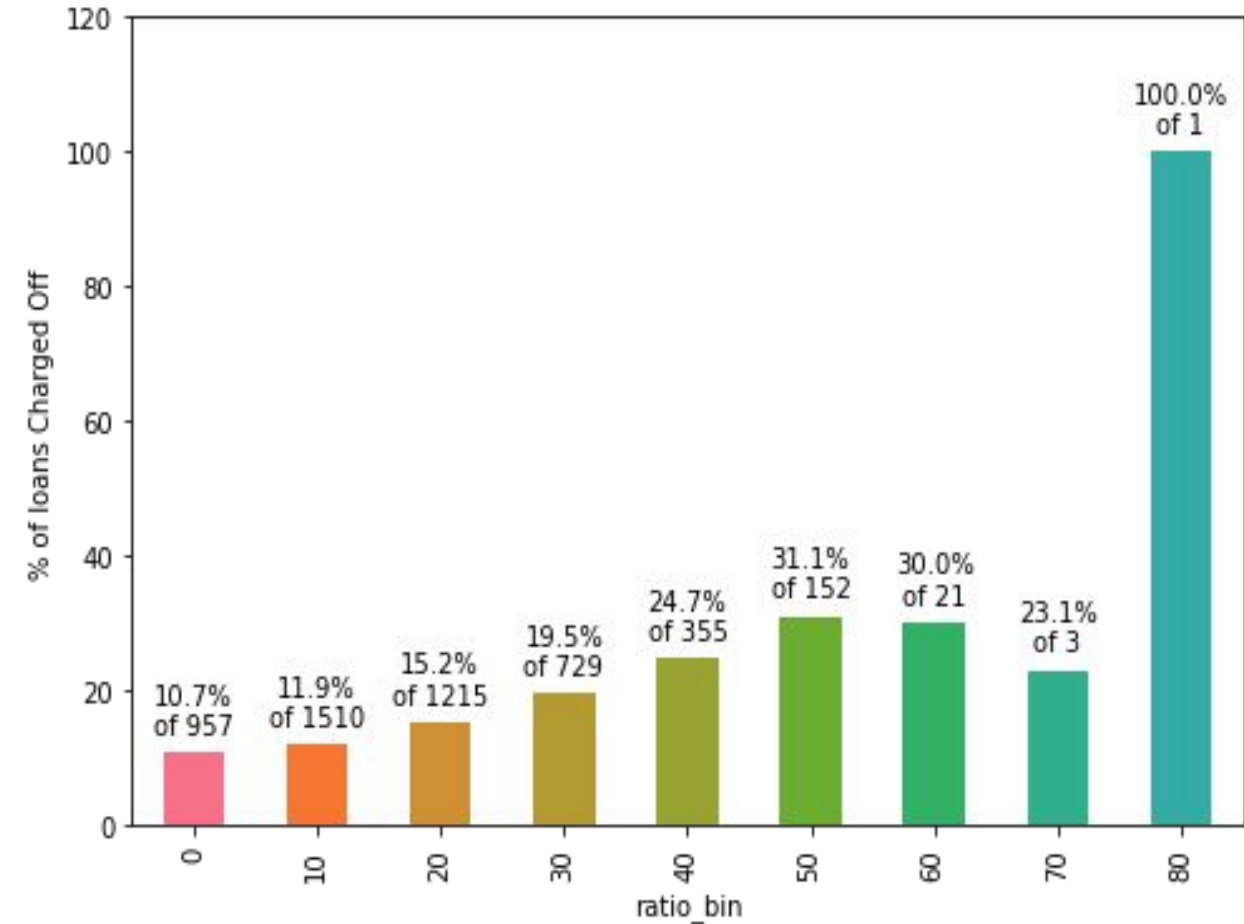


## Result : Plots

**delinq for 2 yrs**



**annual income ratio**



# Results & Conclusions

- Approx, **14%** of loan is defaulted.
- **Lending Club** only recovers **57%** of the loan amount when loans are **defaulted**. On **Fully Paid** up loans, the company makes **17%** profit.
- Most loans are high quality, with a **grade** of **A** or **B**. Lower grades have higher incidence of defaults on loans and the grading system helps in analysis.
- Maximum number of loans are for **debt consolidation**, followed by **credit card**
- Lending Club charges **higher interest rates** as the **grade** of loan becomes **worse**.
- More than a quarter of loans taken for the purpose of running a small business see defaults.
- Loan default decreases with higher annual income.
- Loan amounts of **30%** of annual income or higher see a **high rate of default**.

# Recommendations:

- **Stop :**
  1. Approving loans if loan amount is higher than **30%** of **annual income**
  2. Approving loans to people with prior bad record. Or at least stop approving high-value loans
  3. Approving high amount loan to lower grade.
- **Reduce :**
  1. Number of approvals where loan purpose is small business
- **Start :**
  1. Charging higher interest rates for loans with **dti** greater than **20**