# Lending Club Loan Analysis

**Presented by: Prateek Choubey and Tarun Kumar Udar**

# Overview

- About Lending club & Business Problem
- Data Description
- Explanatory Data Analysis
- Univariate Analysis
- Univariate Analysis on Loan amount-Quantitative Variables
- Data Preparation & Processing

# Overview of Lending club and the Business Problem

•This company 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. Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

•If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

•In other words, 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.

•Our business problem is to develop your understanding of the domain, you are advised to independently research a little about risk analytics (understanding the types of variables and their significance should be enough).

# Data Description

The data 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.

In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

When a person applies for a loan, there are two types of decisions that could be taken by the company:

Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:

- Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
- Current: Applicant is in the process of paying the installments, i.e. the tenure of the loan is not yet completed. These candidates are not labeled as 'defaulted'.
- Charged-off: Applicant has not paid the installments in due time for a long period of time, i.e. he/she has defaulted on the loan

Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset
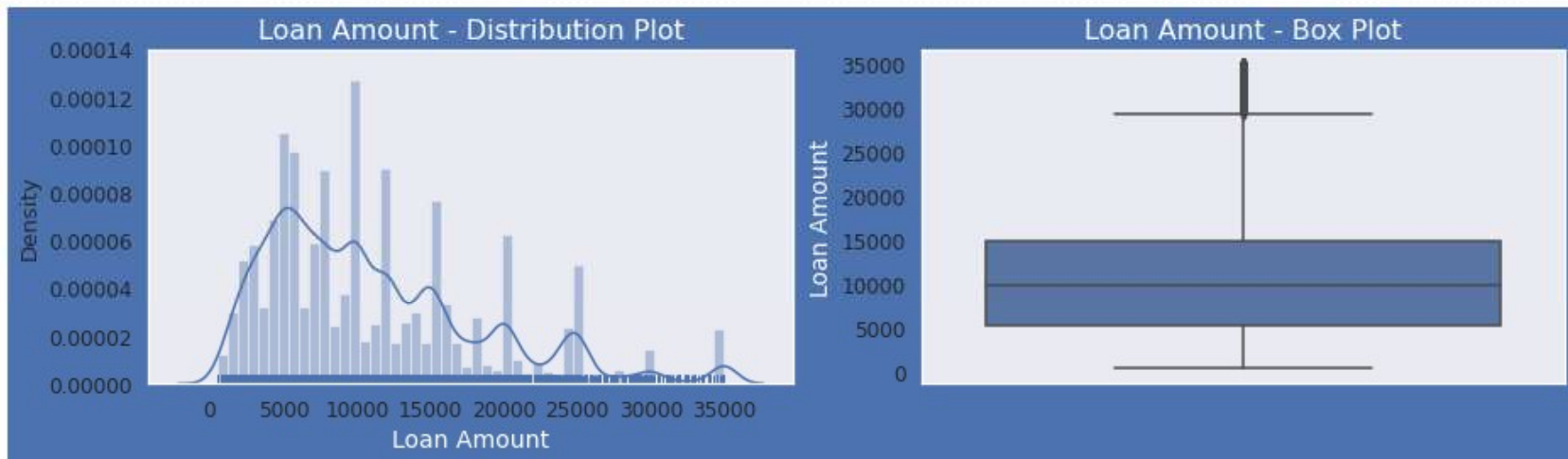
# Univariate Analysis

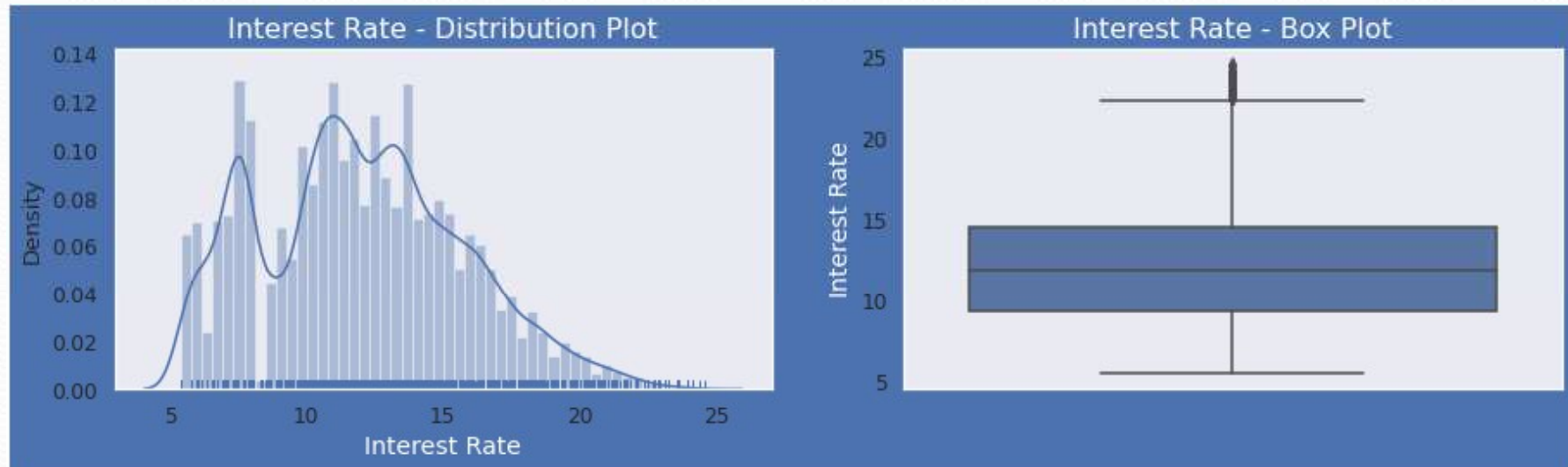Distribution of three loan amount fields using distribution plot.
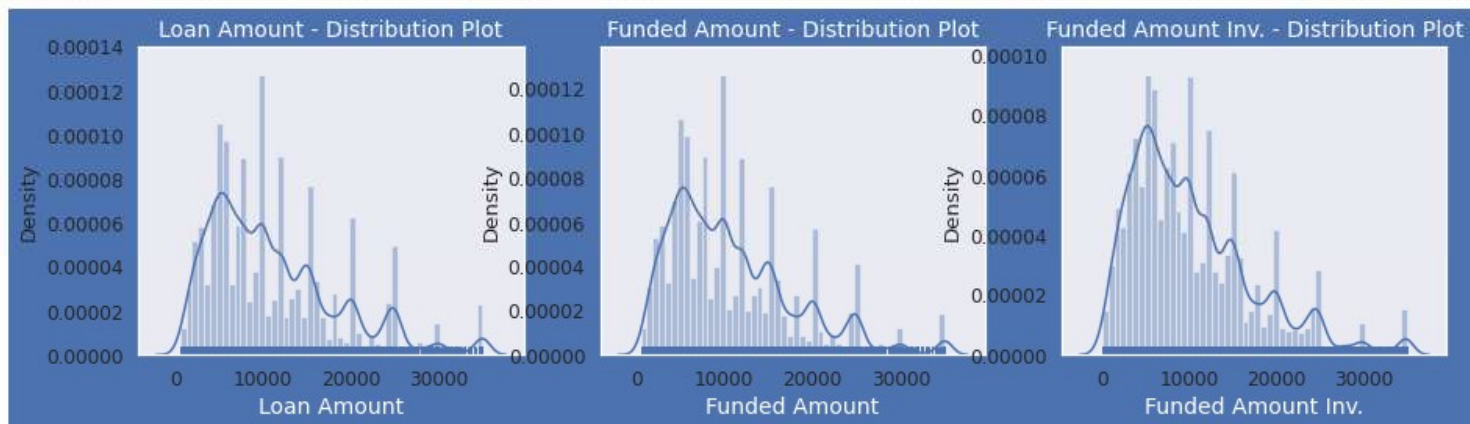


**Distribution Plot of loan amount**

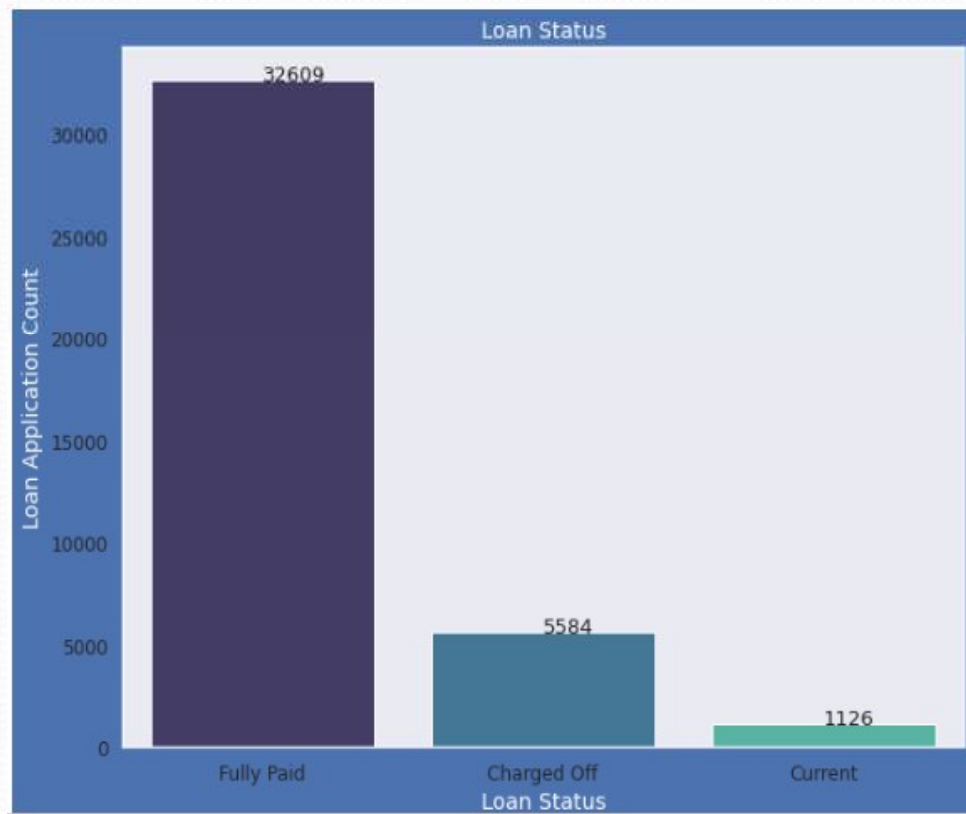# Univariate Analysis on Loan amount-Quantitative Variables



**Distribution & Box Plot of Loan Amount**

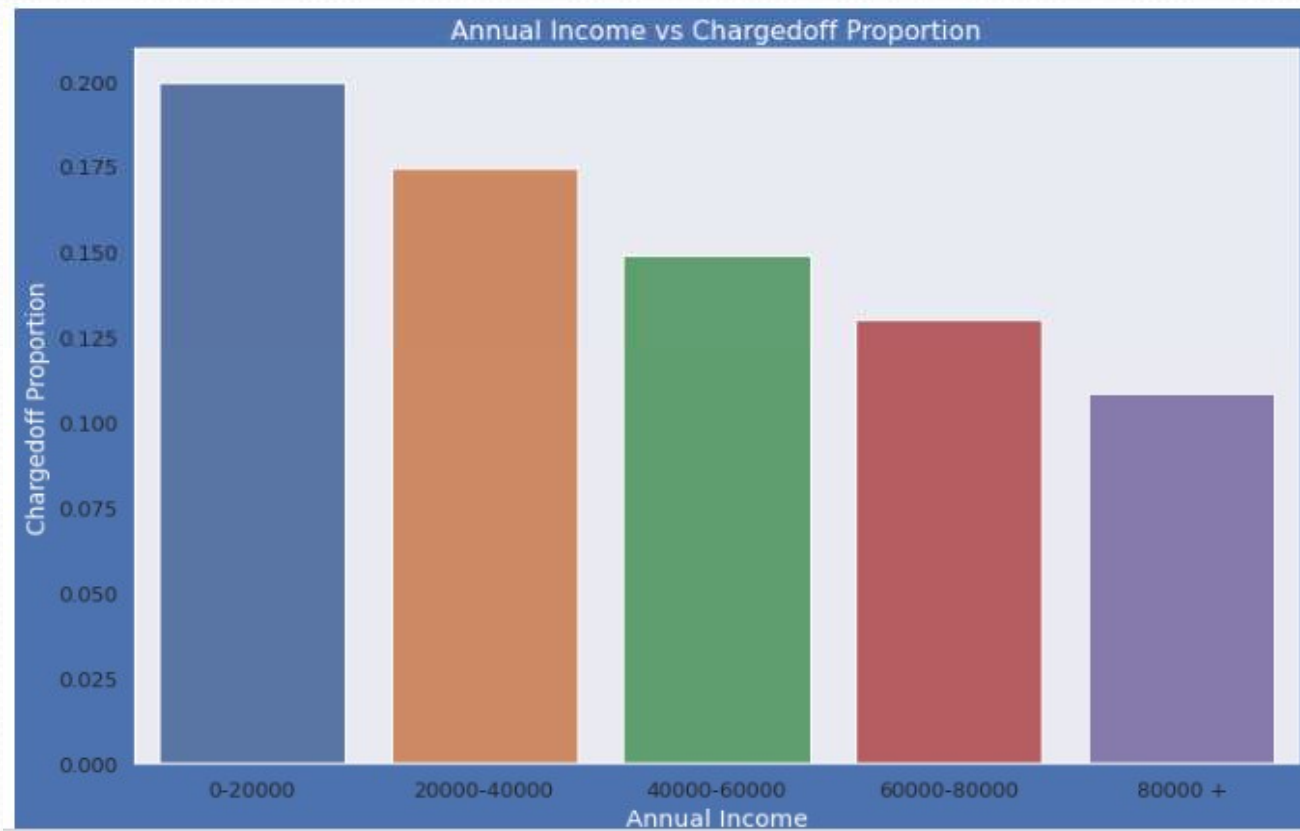# Univariate Analysis on Intrest Rate-Quantitative Variables



**Distribution & Box plot of Interest Rate**

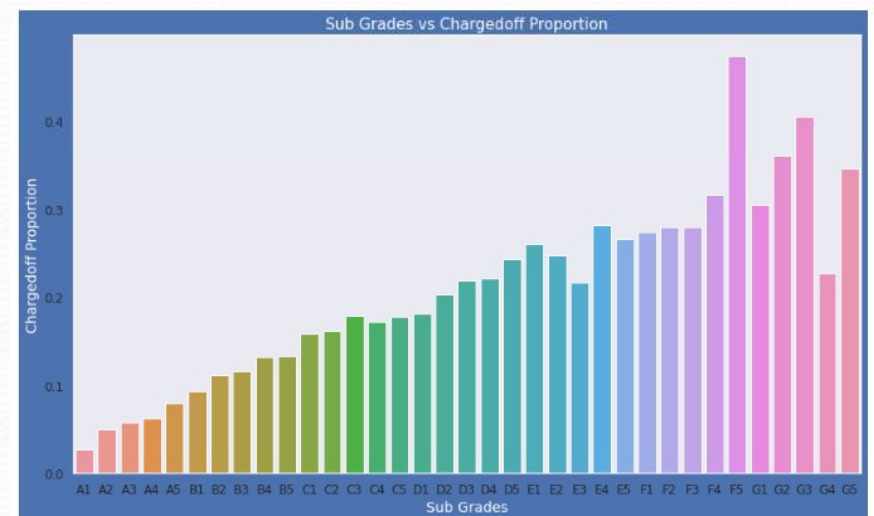# Univariate Analysis - Unordered Categorical Variables - Loan Status
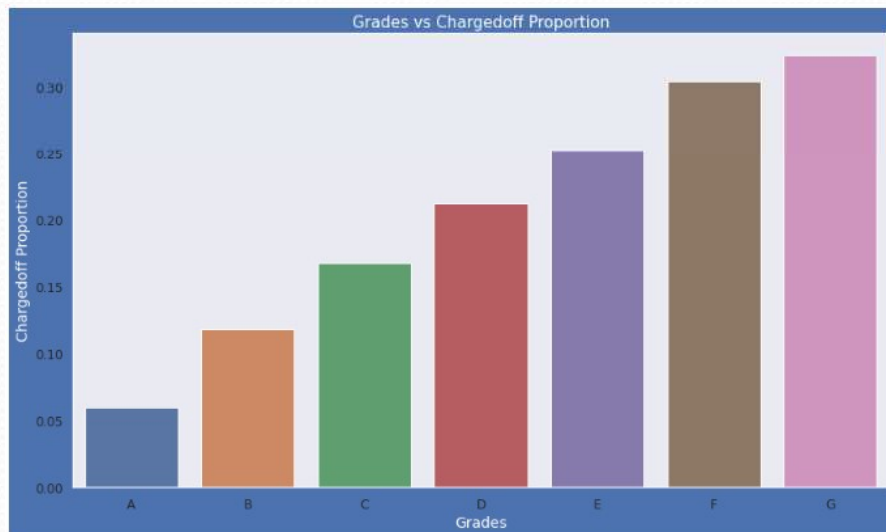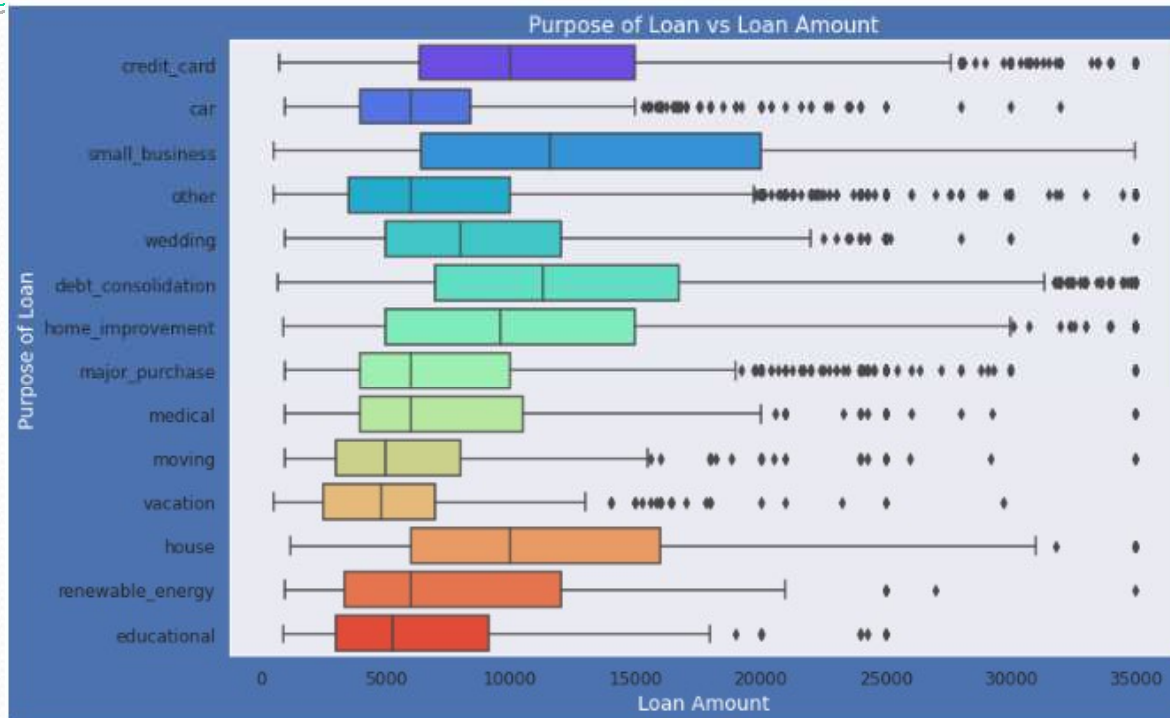
# Annual Income vs Chargedoff Proportion

# Bivariate Analysis on grade against Chargedoff_Proportion

*Analyzing Chargedoff Proportion for each grades and sub grades.*

**Purpose of loan vs Loan amount**

# Data Preparation & Processing

- Removed the columns which are having null values.

- Removed the variables which has only one category such as member_id, url, emp title, zip_code, tax_liens.

- Three types of categorical features: ordinal, nominal and binary.

- Will do Univariate, Bivariate & Multivariate Analysis.

- Derived Column Analysis.