# Lending Club Case Study

GROUP MEMBERS:

Hamza Tanveer

Yashraj Pathak

BUSINESS OBJECTIVE

Using the techniques of EDA, develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers. Identify driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

ANALYSIS APPROACH

| Data Cleansing and Preparation | → | Univariate Analysis | → | Bivariate Analysis | → | Conclusion |
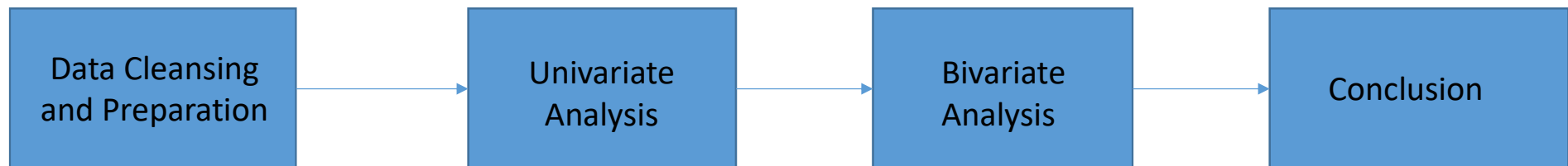
Data Cleansing and Preparation

- **Removing columns for**
    - which 50% of the values are null
    - which have distinct values for each row
    - which one known value for all the rows
    - which have high co-linearity with other columns
    - which gets available after loan is issued

- **Standardizing the columns**

- **Removing rows for**
    - CURRENT loan status as we are not aware if those will be defaulted or not

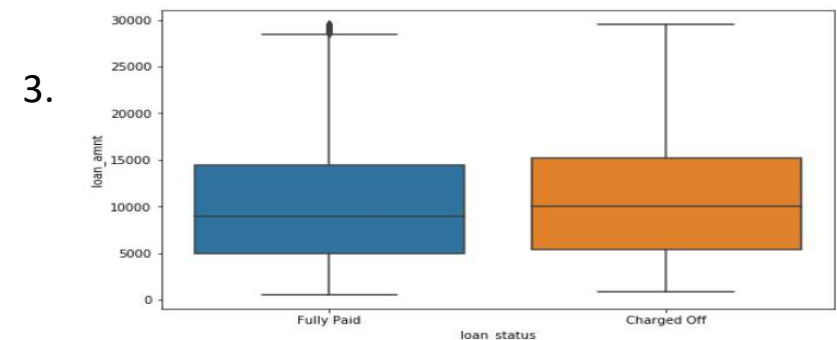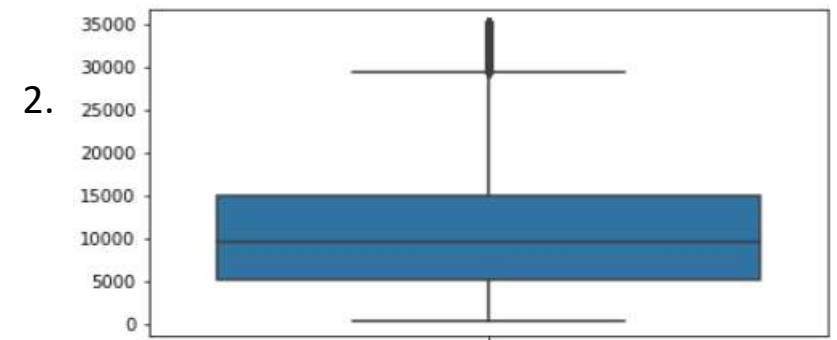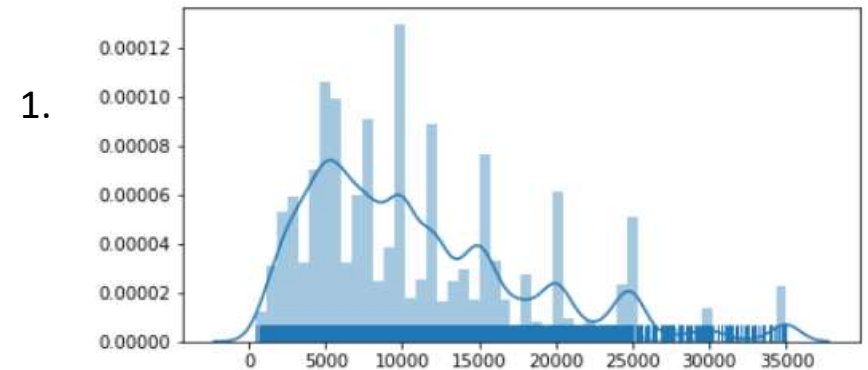- **Created column for issued year**

Univariate Analysis

- **Quantitative Variables (Loan Amount)** Identify the outliers for the loan_amount and removing the outlier's

 Using Formula: Q3 + (Q3-Q1)*1.5

Figures:

1. Distribution plot of loan_amount
2. Box pot of loan_amount
3. Segmented box plot of loan_amount after removing outliers

Univariate Analysis

- **Categorical columns (grade, sub_grade, home_ownership, verification_status, loan_status, purpose, addr_state):** Categorical Variables Count for segmented into Fully Paid or Charged Off
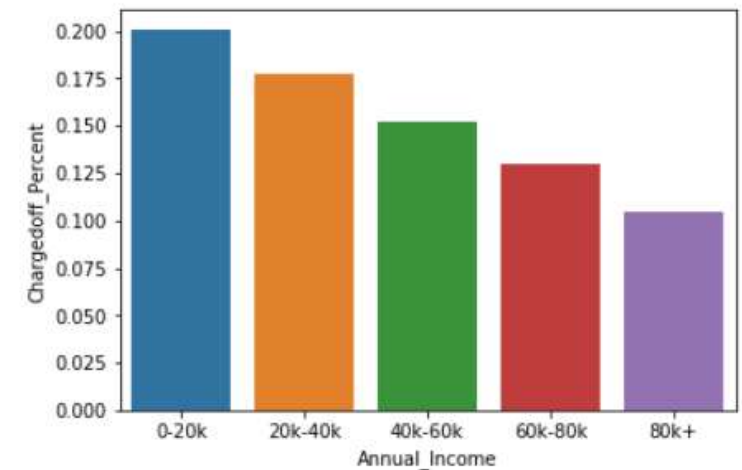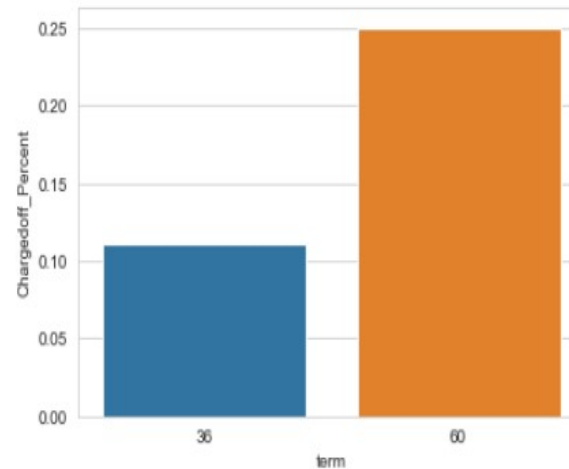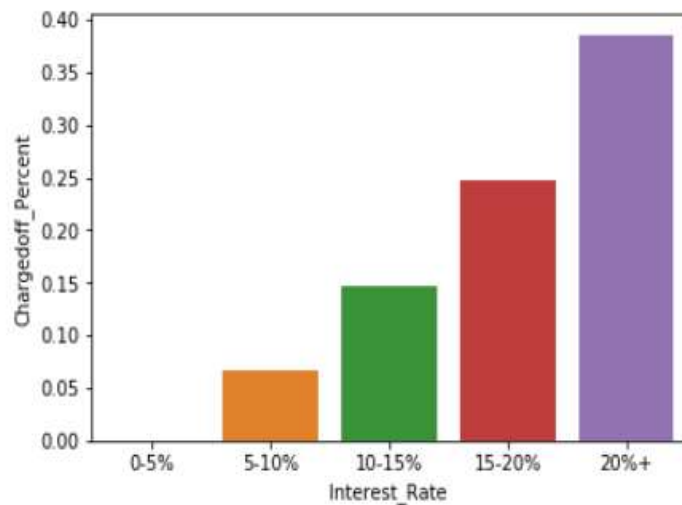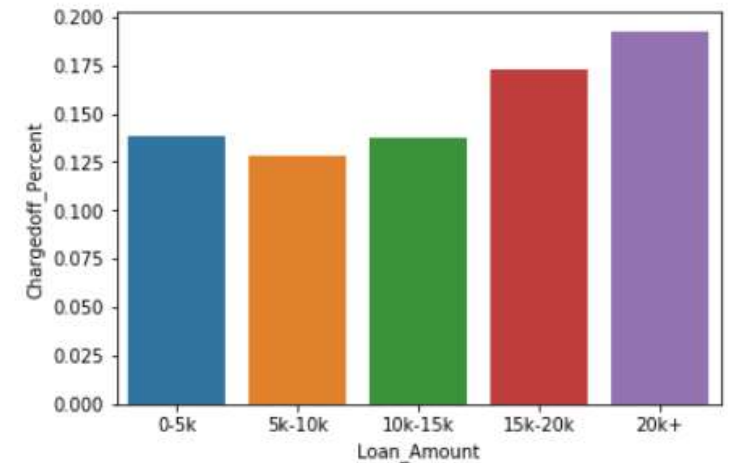
NOTE: Using log scale for better readability of graph
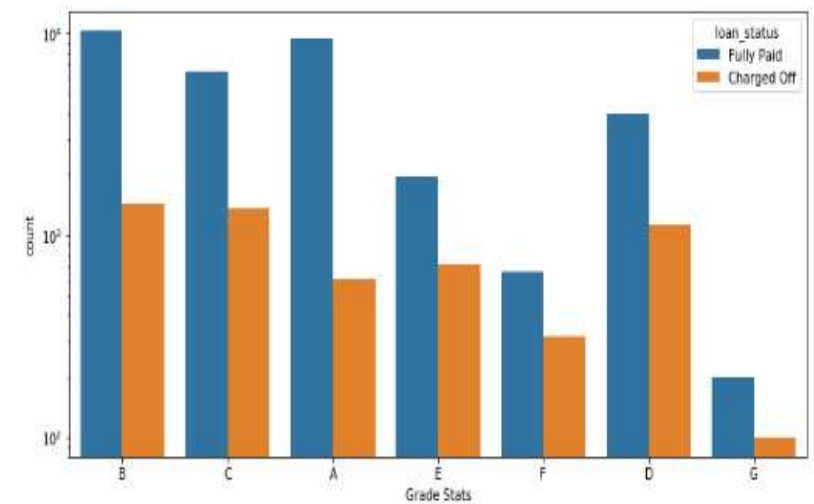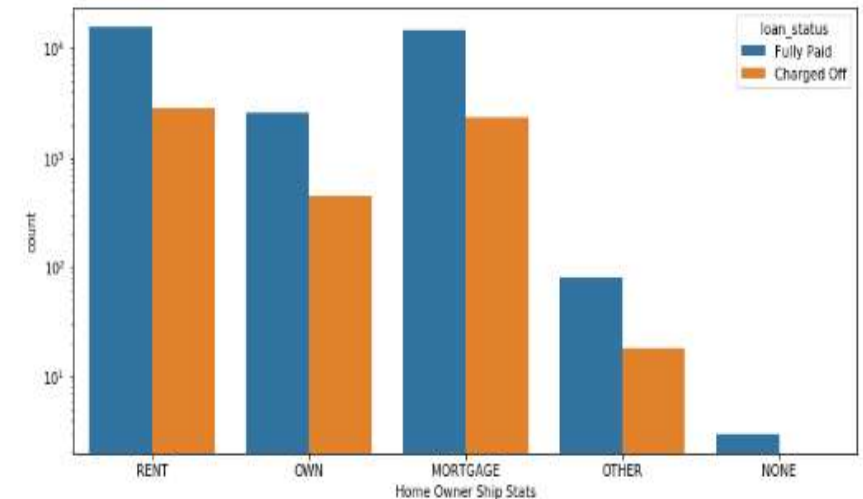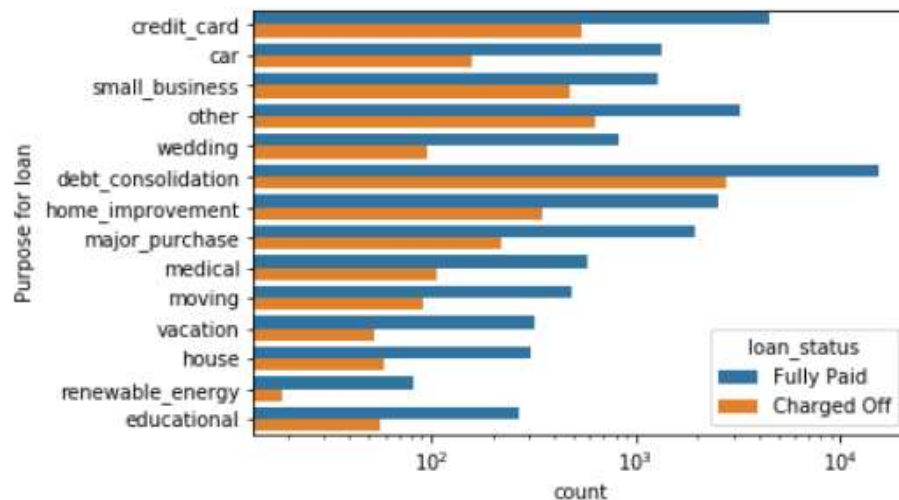
Bivariate Analysis

Bar graph of (Interest Rate, Loan Amount, Annual Income) columns against the ChargedOff_Percent
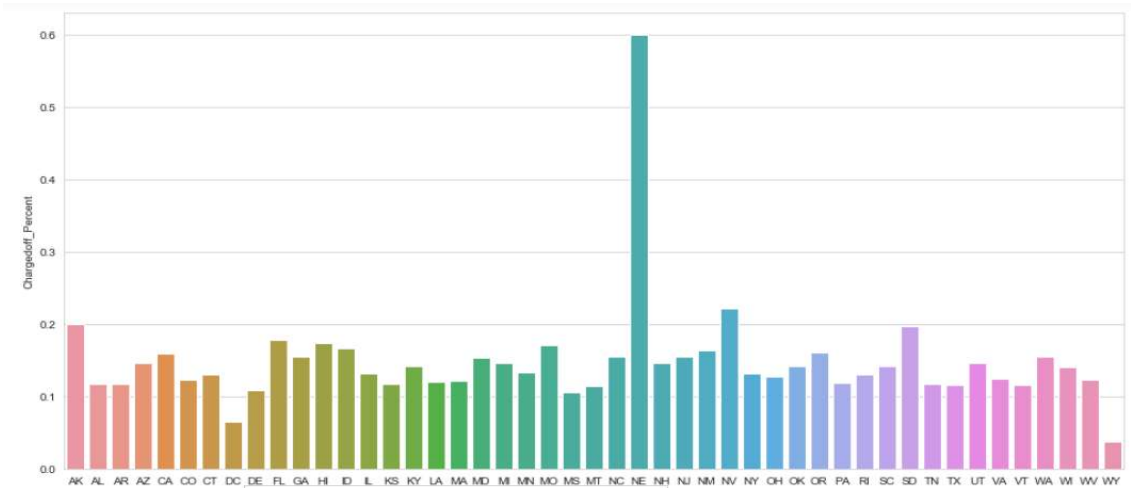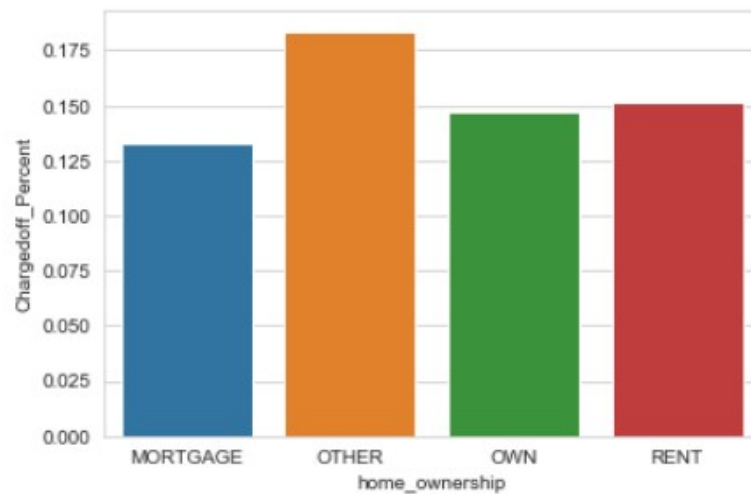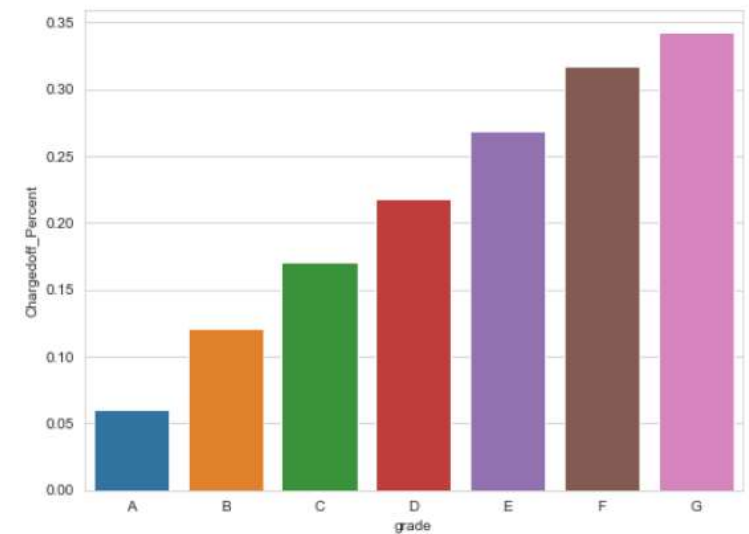
$$\text{ChargedOff Percent} = \frac{\text{No of Charged Loans}}{\text{Total number of Loans}}$$

Bivariate Analysis

Bar graph of (Home_Ownership, State, Grade) columns against the ChargedOff_Percent

$$\text{ChargedOff Percent} = \frac{\text{No of Charged Loans}}{\text{Total number of Loans}}$$

Conclusion

Driving Factors (variables)

- Loan Factors:
  - Lending Club should check before disbursing higher amount of loans as they are risky of getting defaulted
  - Interest Rates should be reduced for customers for longer tenures like 60 months.
  - Lending Club should check before issuing loans for small scale businesses as they are more likely for getting defaulted
  - Lending Club should examine financial crisis for the year as mostly loans are defaulted in the year of financial crisis.
  - Grades being an important metrics to check for loan defaulters, should do proper check for other parameters before issuing loans.

Conclusion

Driving Factors (variables) continued...

- Customers Factors:
  - Lending Club should entertain more customers with highly annual income are less prone to Loan Default
  - High amount should not be sanctioned for borrowers with unknown home status
  - Lending Club should stop issuing Loans for customers from NE
  - Revolving Balance is an important key factor so higher the revolving balance higher the chances for defaulters.
  - People with more numbers open credit line are risky for issuing loans.
  - Probability of higher charged off is with customers enquired the most or declared bankruptcy even once..