Lending Club Case Study

SUBMISSION

Name: Abhishek Singh

Email : aulakh.abhishek@gmail.com

Dated – 19 May 2021

Objective

- **Lending Club**, a consumer finance company which specializes in lending various types of loans to urban customers, wants to understand the Risk associated with Loan applications.

- The company when receives a loan application, It has to decide for loan approval based on the applicant's profile. There are two types of risks that are associated with the bank's decision:
    - If the applicant is **likely to repay** the loan, then not approving the loan results in a loss of business to the company
    - If the applicant is **not likely to repay** the loan, i.e., he/she is likely to default, then approving the loan may lead to a financial loss for the company

- Data for Loans approved through Lending Club between 2007 to 2011 is provided which contains basic information about the loan application details along with confirmation if the applicant has 'Fully Paid' Or 'Charged Off' the loan

- Business objective of this analysis will be to get insights from the data to 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.

- Finally, the Goal will be to identify the driving factors behind a 'Fully Paid' vs 'Charged Off' Loans through EDA to understand how consumer attributes and loan attributes influence the tendency of default. There will be no data modeling.

Approach

- Following workflow chart details the approach of analysis and findings to be followed.
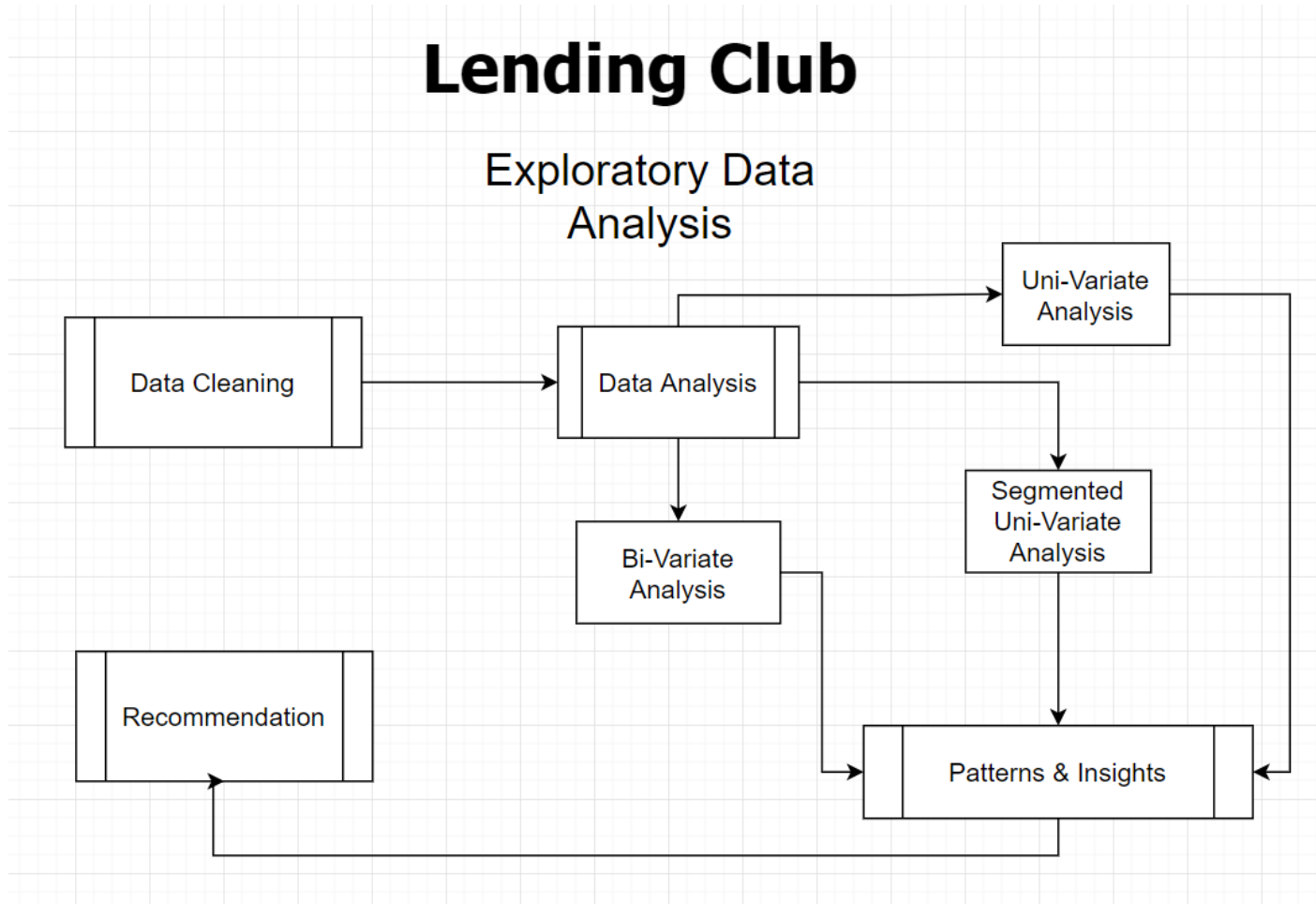
Data Clean-up

- The data provided for Loans disbursed between 2007 to 2011 contains basic information details such as Member ID, Loan Amount, Term of Loan, Interest Rate, Grade, Annual Income, Purpose, Issued Date and lot of other details along with whether the Loan was 'Fully Paid' or 'Charged Off'
- The most important metric for EDA in dataset will be to identify and draw insights for finding driving factors which generally leads to Defaulting of Loan or Loan being paid in Full
- Dataset is cleaned up for fields which were not required, relevant to analysis, contains more than accepted level of Null values.
- The date, Interest Rate fields were manipulated into datetime and float data-type.
- Further, Records where important fields were missing values were also removed to avoid introducing bias in the results.
- Finally, The Numerical columns were binned or segmented into categorical variable to plot useful charts for insights
- The Analysis was focused on 'Fully Paid' vs 'Charged Off' loans therefore the data for 'Current' loans was discarded
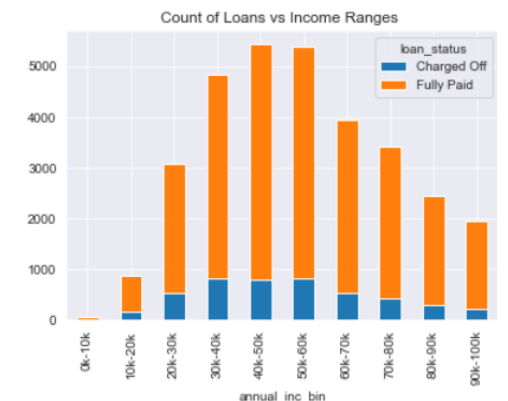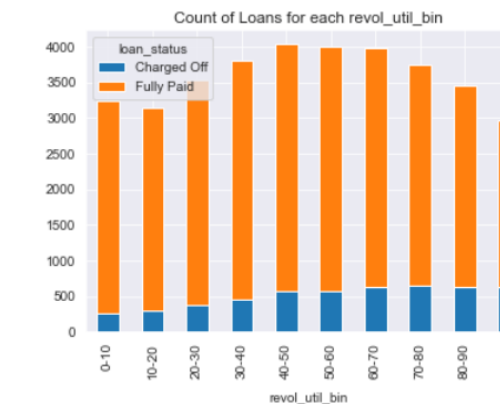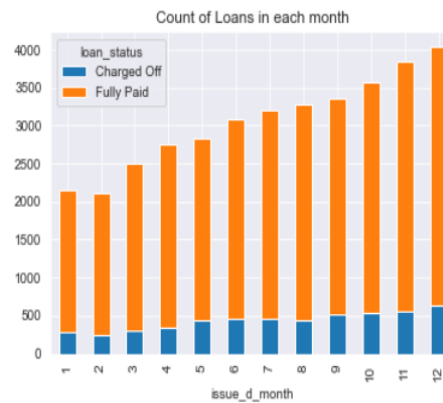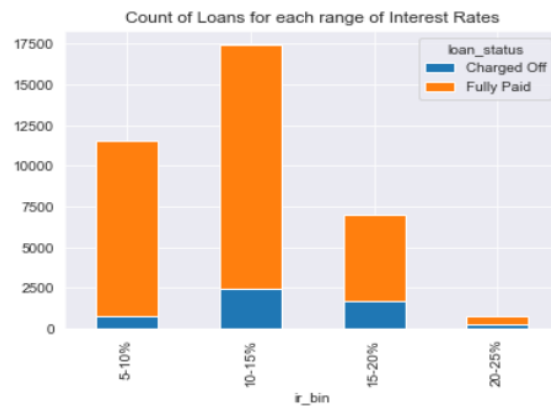

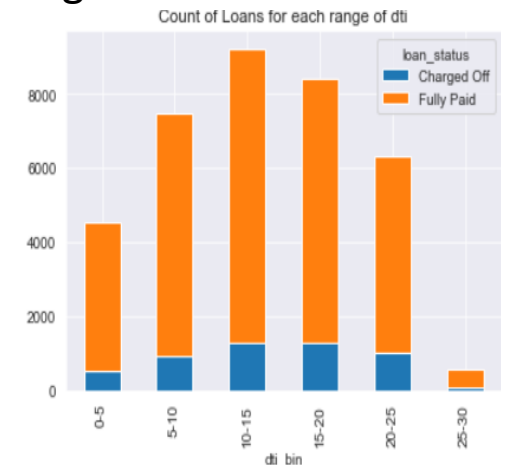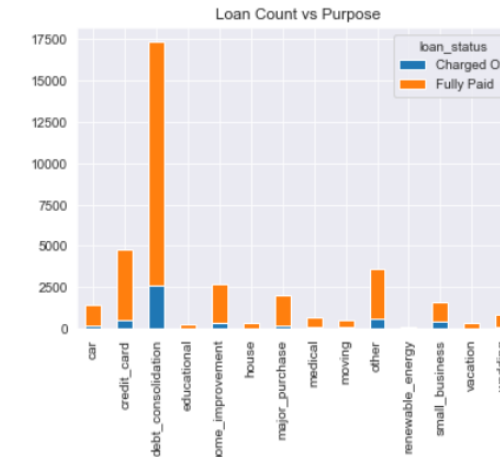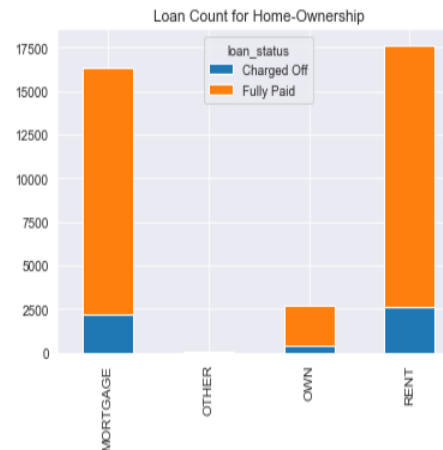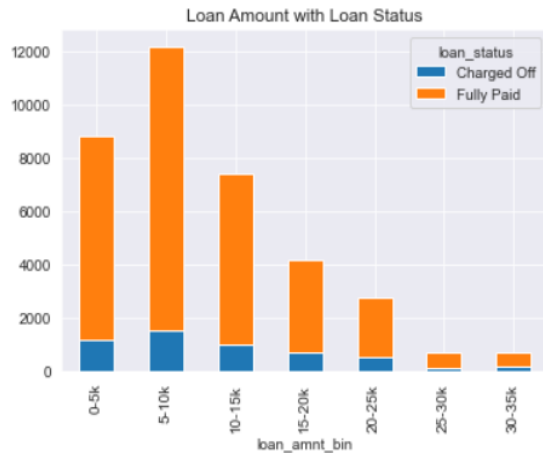Fully Paid vs Charged Off Loans
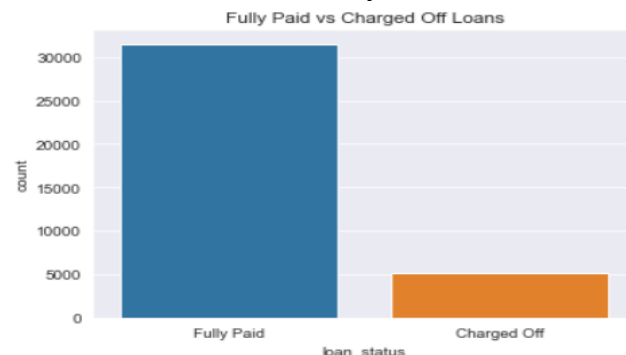
Data Analysis – Uni-variate

- Individual Variables were analyzed to plot the Rank-Frequency plots various important variables

- Observation: Loan Amount, Grade, Home-Ownership and Purpose showed good characteristics of Power Law. Large number of loans were disbursed for Grade-B, DTI range 10-15, Term-30m, Int-Rate b/w 10-15%, during the fall of the year, with Experience of more than 10 yrs.
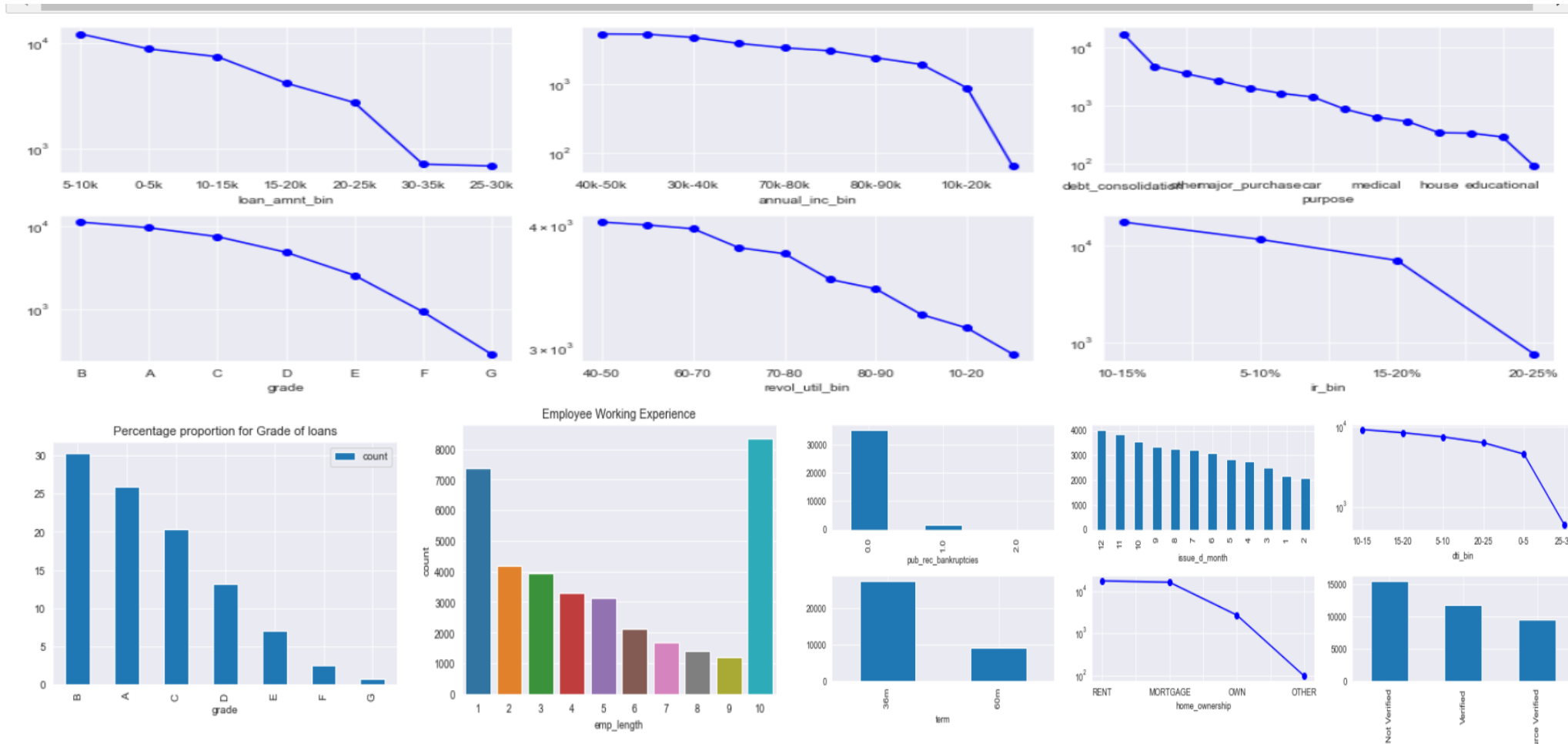
# Data Analysis – Segmented Uni-variate

- The variables were analyzed by categorizing them and distinguishing the behavior for 'Fully Paid' and 'Charged Off'

- Observation: Proportionally, Loan amount 5-10k, Home-Ownership of RENT, Purpose of 'Debt-Recons', DTI – 10-20, Interest Rate 10-15%, End of Year, Revol_Util 60+, Income range – 30k-60k is higher for Charged Off Loans.
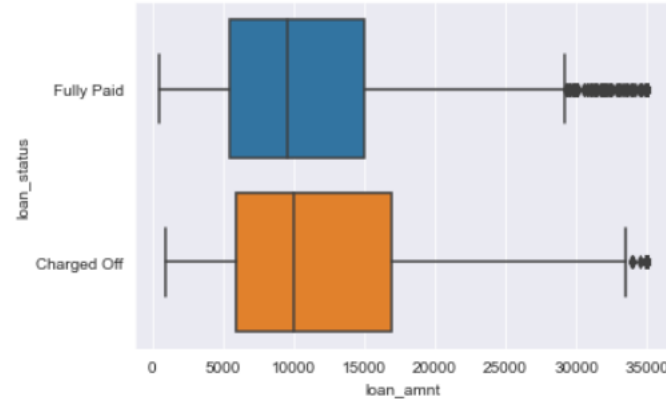
Data Analysis- Bi-variate

- Two set of variables are compared with each other against 'Fully Paid' and 'Charged Off' Loans

- Observation-1. Median value of Interest Rates & Loan Amount for Charged Off Loan is higher than Fully paid.

- Observation-2. Median value of Int-Rates is higher for Grade G, Debt-Consolidation and term of 60 months for Charged off Loans
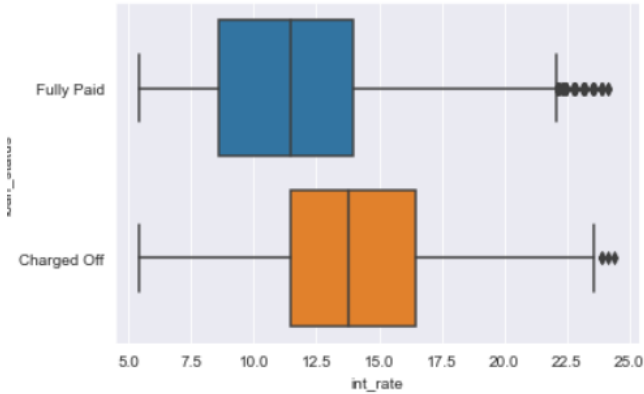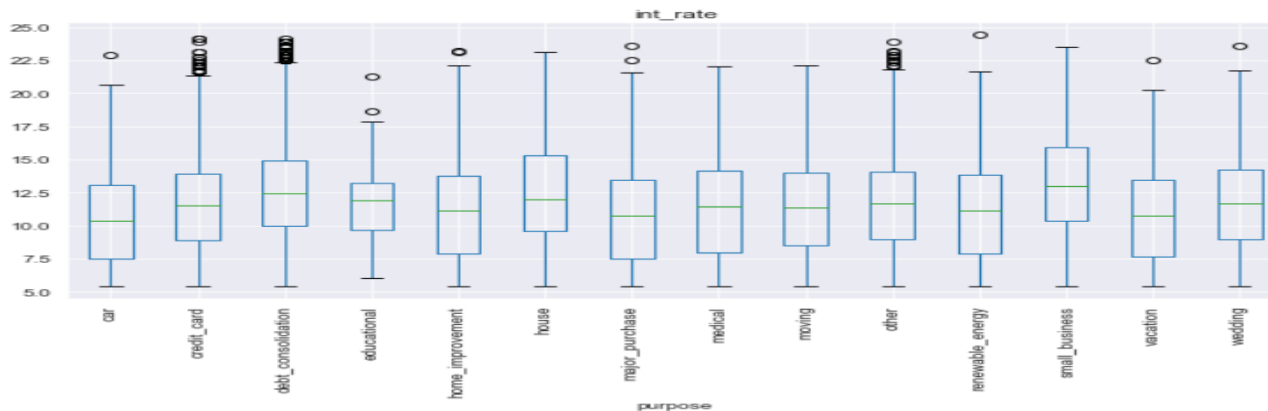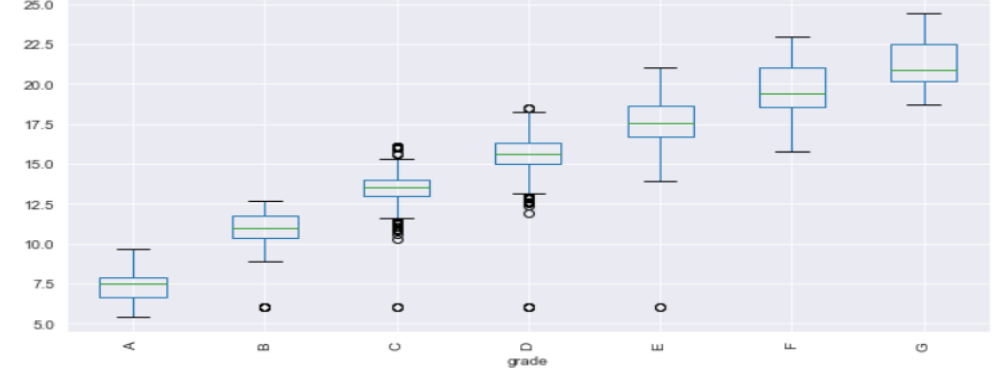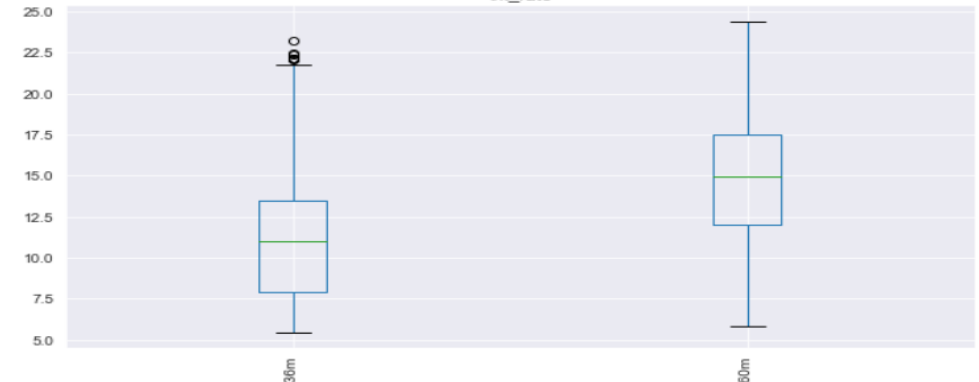
Data Analysis- Bi-variate Cont.

- High Interest Rate, Grade G and F, High Loan Amount, Longer Loan term, High Revol Utilization seemed to be main driving parameters which were observed for Charged Off Loan.
- Interest-Rate, Loan Amount, Annual Income, Revol Utilization and Grade are compared in below plots
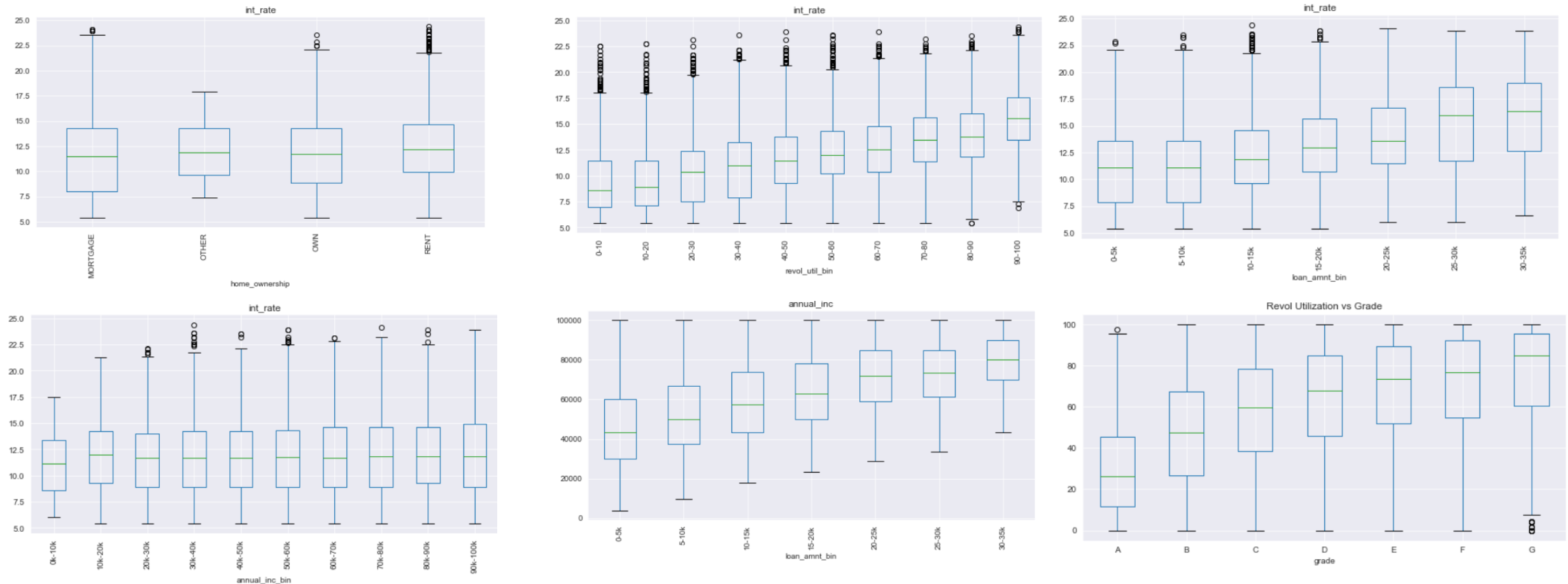
Data Analysis- Multi-variate

- Loan Amount, Term, Interest Rate, Grade and Purpose were analyzed together to understand the behavior on how they affect the Status of Loan between 'Fully Paid' and 'Charged Off'
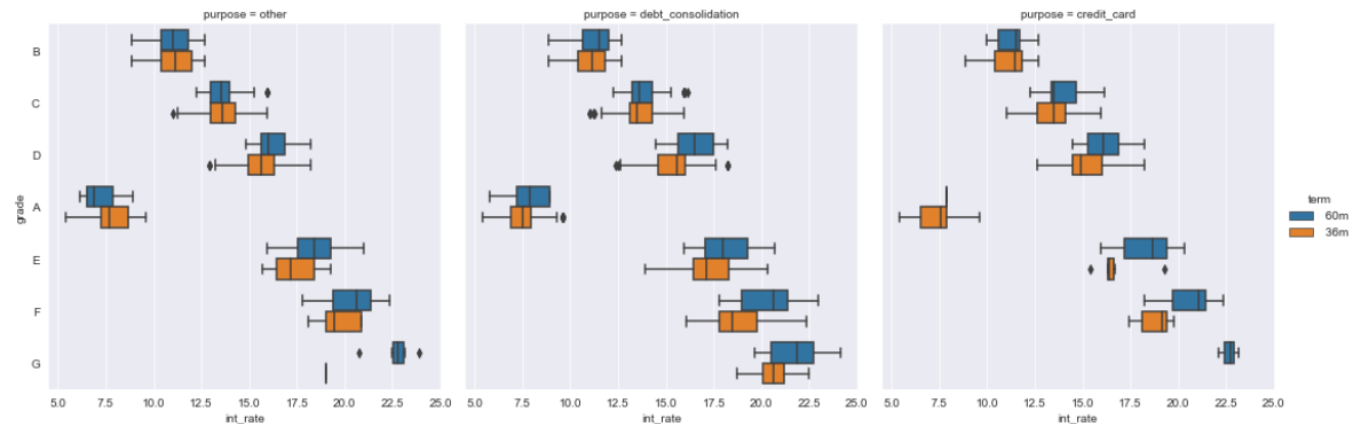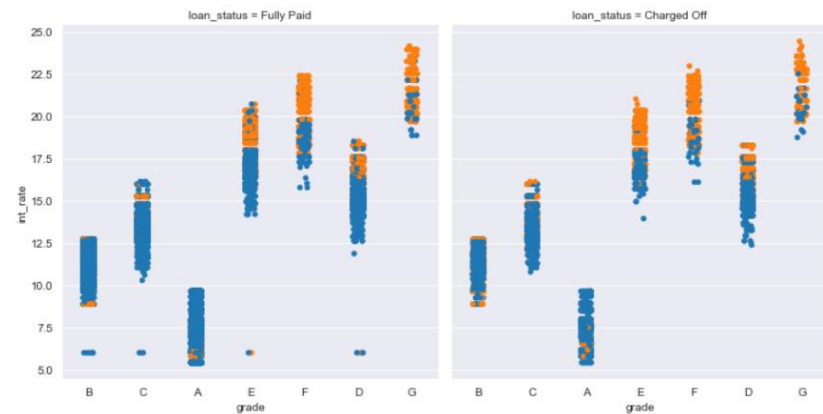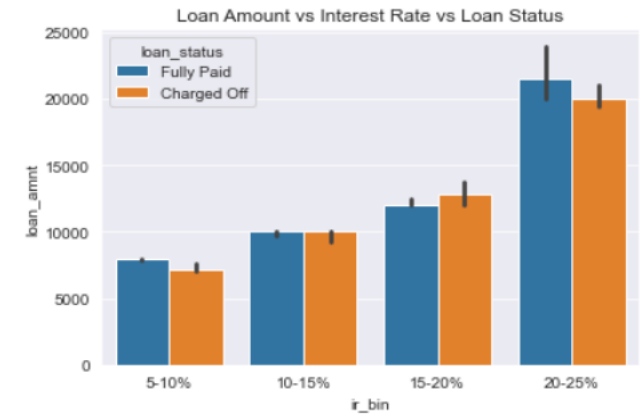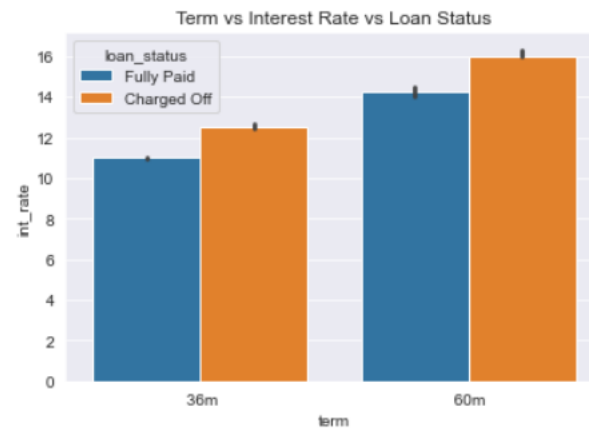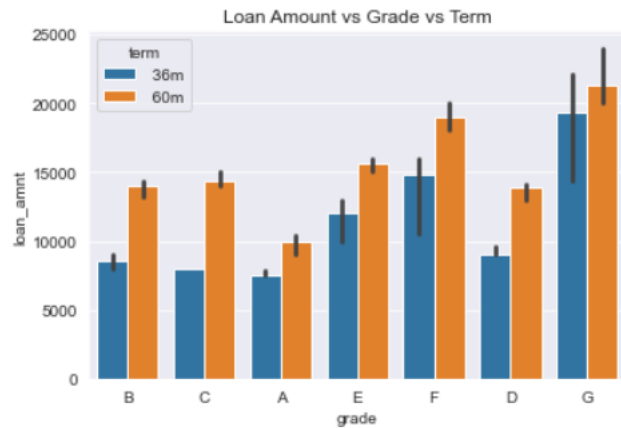
Recommendation.

- Based on the Analysis, whereby identifying the main driving parameters from the data which can cause a Loan application to become Risky, following are the recommendations

  - Lending Club should avoid giving High Loan amounts at High Interest Rate.
  - Lending Club should avoid giving High Loan amounts at High interest Rate for Grade G & F Loan applications.
  - Lending Club should avoid giving High Loan amounts at High interest Rate for longer Term of 60 months.
  - Lending Club should check for low values of Revol Utilization, Debt to Income Ratio(DTI)

  - Finally, Lending Club should check further for loan disbursed in the state of Wyoming (WY) as the median count of loans issued in that state have highest value for Charged Off Loans