Lending Club Case Study

GROUP COORDINATOR : THAHURA NAAZ SHAIK

AIML BATCH : C37

Data Analysis :

▶ **Problem statement** : Lending club is a finance company which offers loans to applicants. In order to make profitable business, they need to reduce the number of default or charged-off loans. **Charged-off** loans are those loans where the applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan.

▶ **Analysis Approach :** The data is systematically analyzed in with different methods and conclusions are drawn from the analysis. The steps include :

1. **Data Cleaning**

2. **Derived data**

3. **Univariate Analysis on :**

    A. **Categorical data**

    B. **Numeric data**

4. **Bivariate Analysis on :**

    1. **Two numeric data**

    2. **Two categorical data**

    3. **Two categorical with one numeric data**

5. **Multivariate Analysis**

Data Cleaning

- Remove current loan status records and keep only fully paid and Charged-off loan status records

- Drop not needed columns

- Drop similar value columns

- Check for null values

- Imputation of data :

  - Missing value replacement

  - Dropping invalid records

- Remove symbols like %

- Split time into day/month/year

Univariate Data Analysis

▶ Charged off records make 14.2% of total loan records

▶ Loan amount is in range of 5-15K, Outliers in loan amount are >= 30K

▶ On average, accounts hold active credit records between 7-12

▶ Majority of charged-off accounts are with longer loan

term of 60 months (56.7%) . In fully paid accounts there are

lesser number of long term loans.
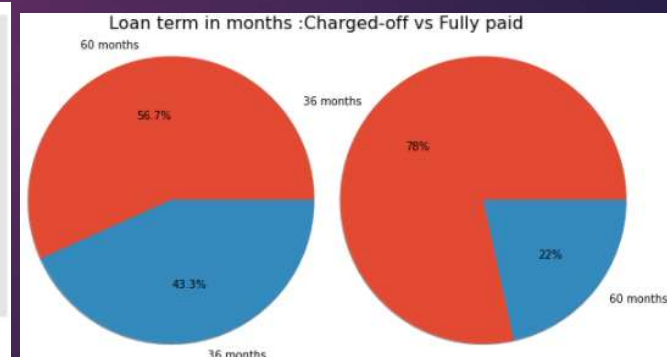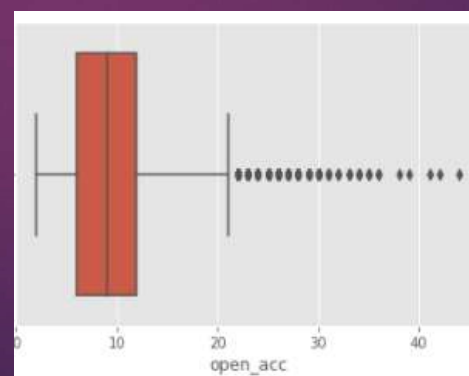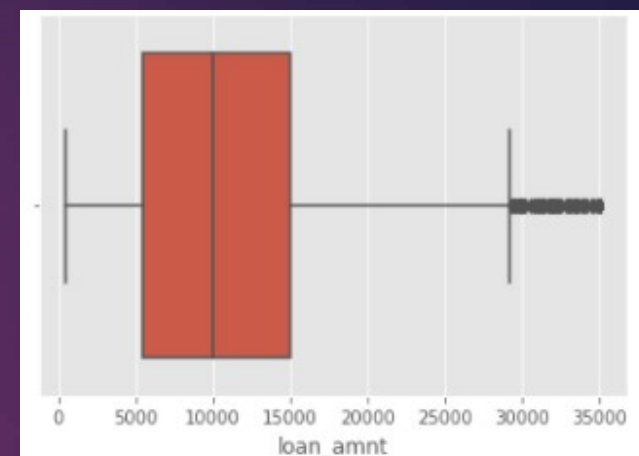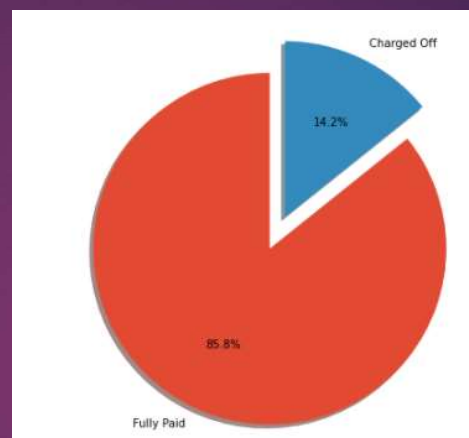
Univariate Data Analysis

▶ Majority of charged off accounts are in grade B

▶ Majority of charged off accounts are in sub grade B5

▶ Majority defaults are 10+ yrs , 2nd hightest <1 yr

▶ Majority of charged-off accounts are for 36 months (3yrs)
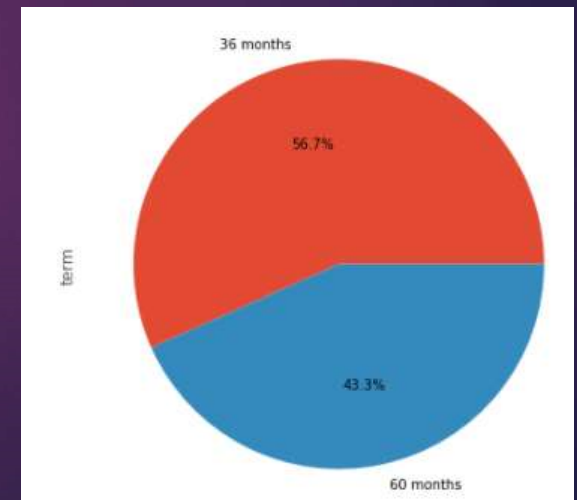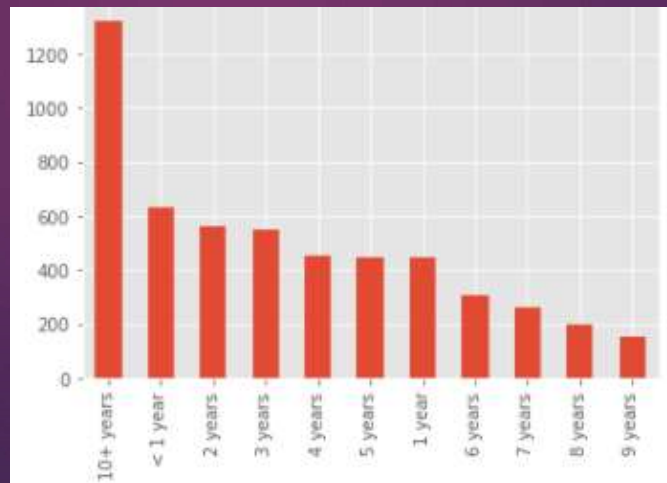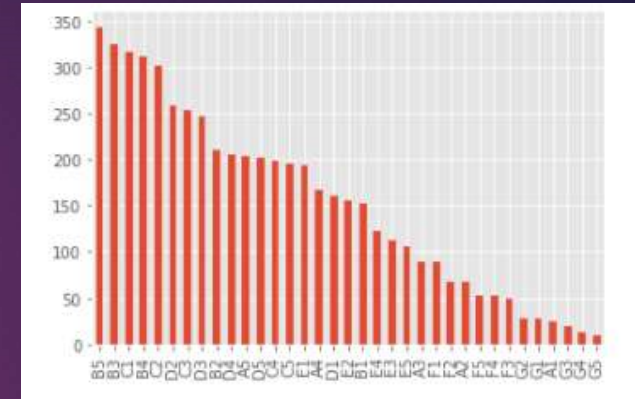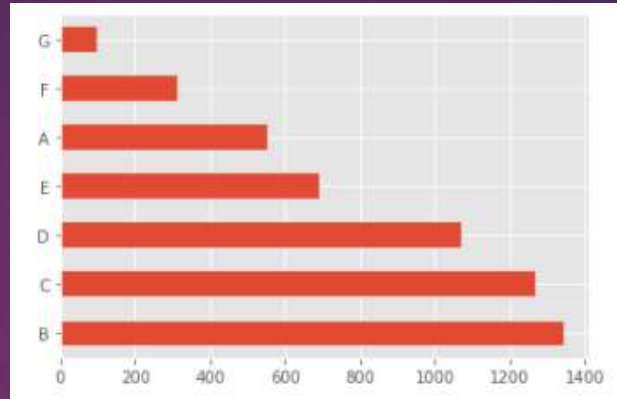
Univariate Data Analysis

▶ Majority of charged off accounts have home ownership as Rent (50.4%)

▶ Majority of charged-off accounts are not verified
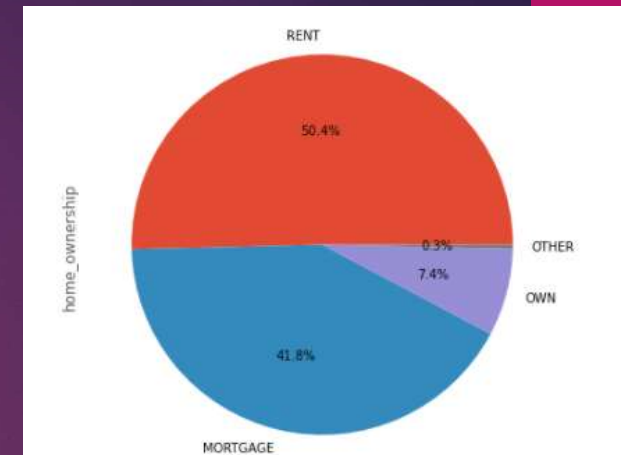
Loan status vs important data

▶ **loan_status vs Loan amount :** 15.7 % of total amount is charged off. These are the problematic accounts which needs to be analysed to ensure newer sactioned loans will reduce the % of charged off amount

▶ **loan_status vs dti :** The average (median) of DTI for Charged off loans is 14.32 which is higher compared to Fully paid loans

▶ **loan_status vs inq_last_6mths :** The number of inquiries in past 6 months (excluding auto and mortgage inquiries) . Average number of enquiries done by charged off records is more than fully paid records
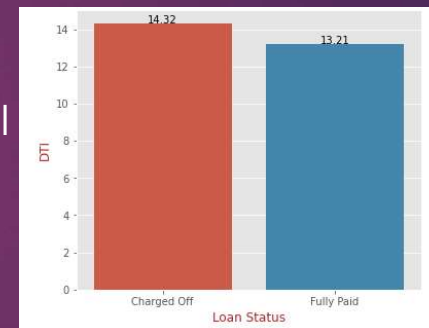
Loan status vs important data

▶ **loan_status vs open_acc :** Ratio of open_acc to total_acc is more in charged off records compared to Fully paid records, but the difference is not very big, cant be a differentiating factor

▶ **loan_status vs pub_rec :** Average number of derogatory public records are higher in Charged off records compared to fully paid records

▶ **loan_status vs pub_rec_bankruptcies :** Average number of public record bankruptcies is higher in Charged of records
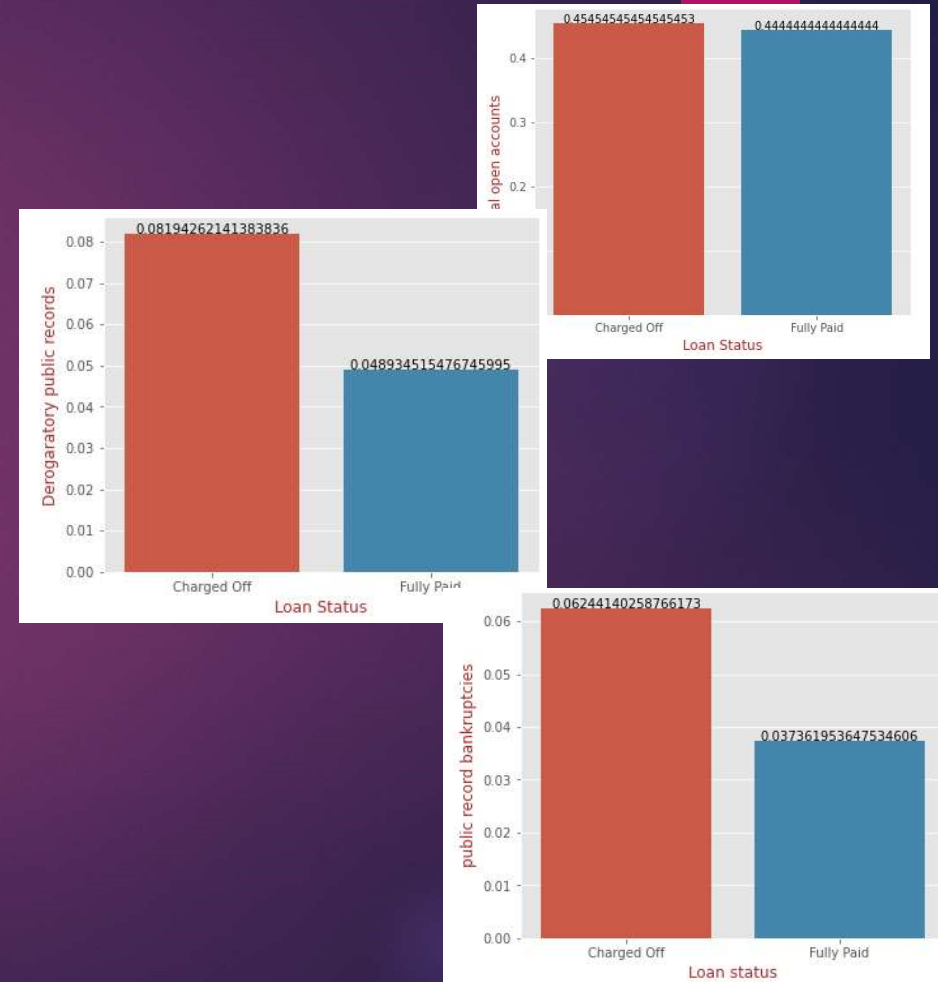
Binned Annual income vs important data

▶ **<u>Binned annual Income vs dti:</u>** Low annual income group have more dti , hence they are more likely to faulty

▶ **<u>Binned Annual Income bin vs pub_rec:</u>** Low and medium annual income group has many public records

▶ **<u>Binned Annual Income bin vs pub_rec_bankruptcies :</u>** Low and medium annual income group has many number of public record bankruptcies

Binning of Annual Income :

VL = 0 - 44872

L = 44873 - 58829

M = 58830 - 77560

H = 77561 - 138264

VH >= 138265
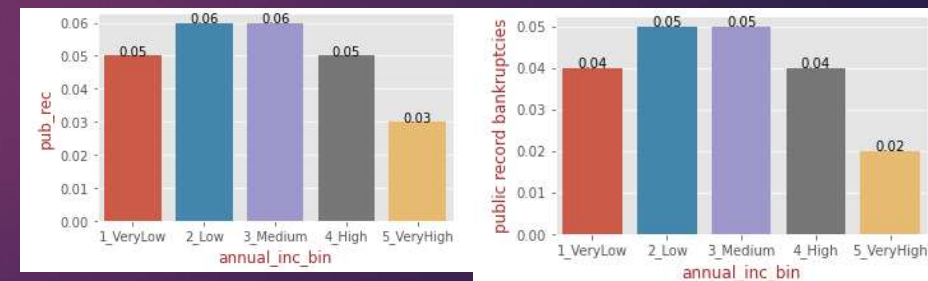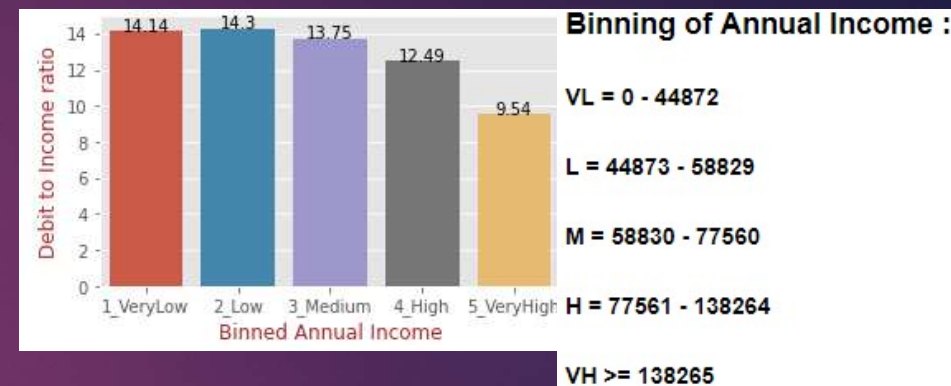
Grade , Home ownership vs important data

▶ **Grade vs number of months since the borrower's last delinquency**: Grade A shows more frequent delinquency(error) and in Grade A, more delinquency is in lower sub-grades

▶ **Grade vs total_rec_late_fee :** By looking at the ratio of late fees received vs total fees recieved in the charged-off dataset, Group A has more number of late fees paid

▶ **home_ownership vs loan_amnt :** Mortgage home owerships take more loan

▶ **home_ownership vs mths_since_last_delinq :** Mortgage has lower mths_since_last_delinq , ie, more frequent defaults have occurred in Mortgage home ownerships
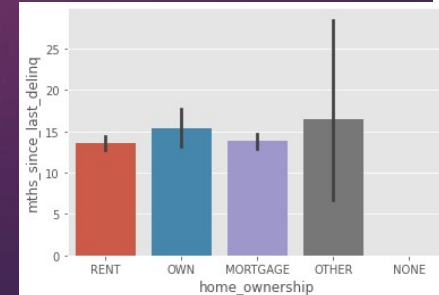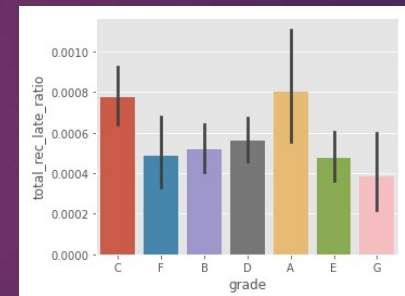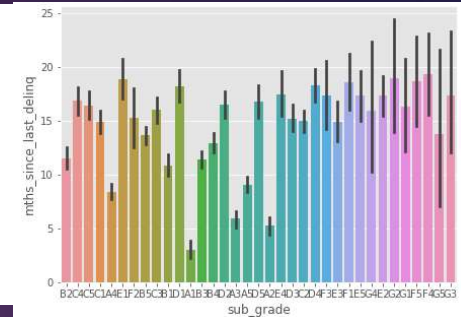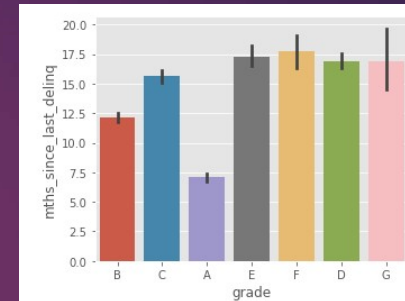
verification status vs important data

▶ **verification_status vs pub_rec:** Not verified records have many public records

▶ **Purpose vs loan amount:** small business higher reason for loan taking pupose