LENDING CLUB CASE STUDY

SUBMITTED BY:

PRAVEEN PENDEKANTI

AMIT KUMAR

AGENDA

- Problem Statement

- Data Understanding

- Data Cleaning and Manipulation

- Data Analysis

- Plotting

- Insights and Recommendations

PROBLEM STATEMENT

- Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision

    1) If the applicant is likely to repay the loan, then not approving the loan results in a loss of business for the company

    2) 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

- The aim is to identify patterns indicating if a person is likely to default, which may be used for actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

DATA UNDERSTANDING

- Before diving into understanding the dataset; went through the lendingclub.com website to understand their business

- Read through the data dictionary, understood how many columns the dataset has, and comprehend what each column means.

- Understood the data high-level, and read through the product to gain as much context that will help for EDA

- Imported the dataset and explored the data by looking at sample data – by using head and tail functions

- Identified and reported the data quality issues in the dataset – null values, outliers, and special characters that will block from transforming or plotting (added comments where ever required)

DATA CLEANING AND MANIPULATION

- Here are the steps, that are performed for Data Cleaning and Manipulation:
 - Checked the number of null values in each of the columns and dropped the column which has all NULL values
  - The dataset has three different types of data (loan status as Defaulted, Not Defaulted, and Current (neither fully paid nor defaulted). Hence this data has dropped before proceeding with the analysis
  - Checked the % of missing values and removed rows with very high missing percentages
  - Possible data quality issues are addressed. Ex – missing values, outlier, and data redundancies. Etc…
  - Manipulated/ Formatted the data based on need. Ex - string to int conversion, rounding off the decimal points, Etc…

DATA ANALYSIS

- Started the analysis by looking at the types of variables – Demographic, Loan Characteristics and customer behavior variables

- As Customer behavior variables are not available at the time of loan application and thus they are excluded from the dataset

- Once the dataset is ready; started with a Univariate analysis for various categorical and continuous variables – Strategized the box plot on each of the columns and identify any null values and outliers. Fixed the data anomalies and drop the outliers and validate/review of the data after fixing the issues (**Note** – Couple plots are added in the last slide for reference)

- Used different python and pandas functions to understand the metadata of attributes

- On each of the steps the insights and comments are added for ease of understanding the code

- Fabricated box plots, count plots, pair plots, scatter plots, and heat map to perform the Multi-variate analysis and detected interactions among attributes

- Data has been grouped for ease of understanding and created a derived column wherever required for analysis

PLOTTING

- Here are the sample plots that are created for the analysis and understanding the data:
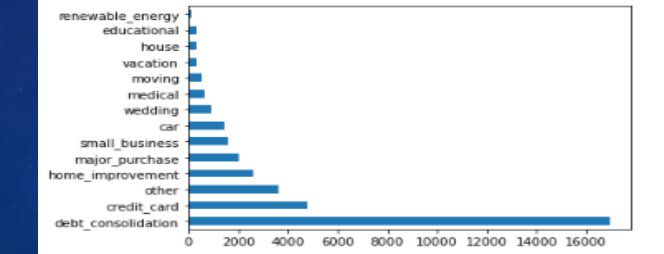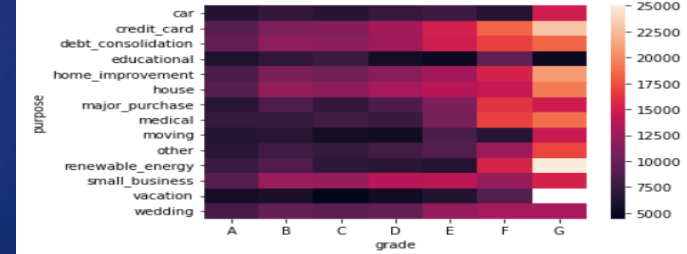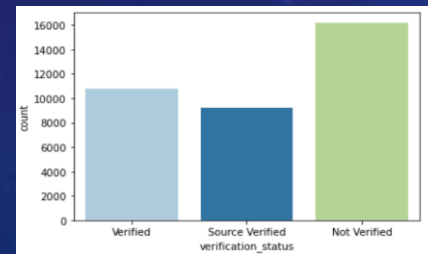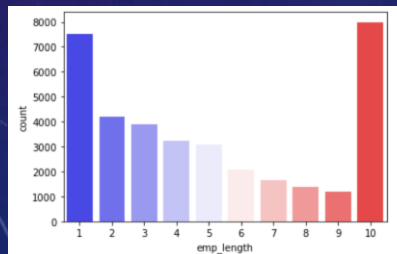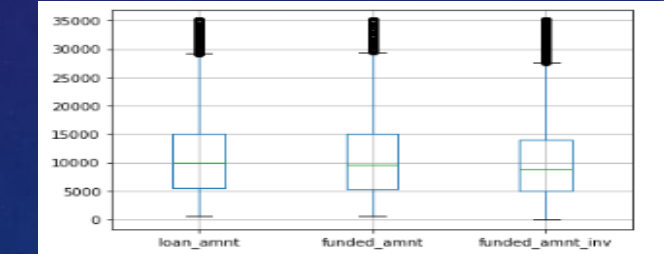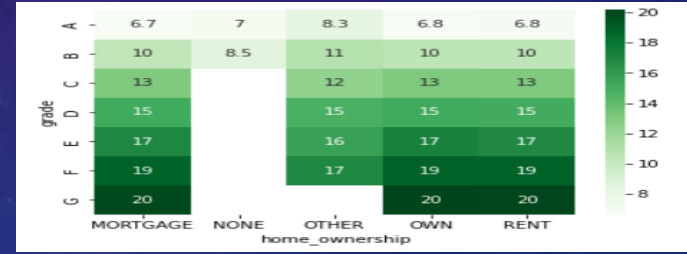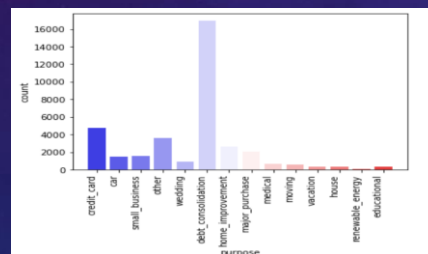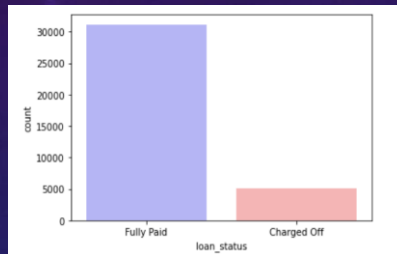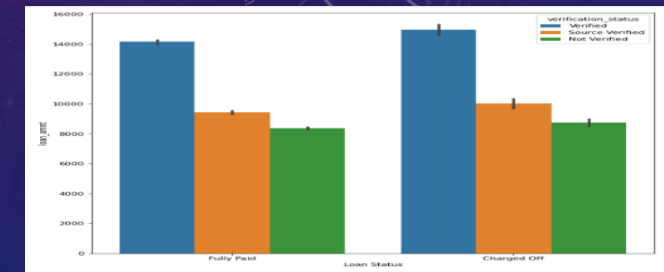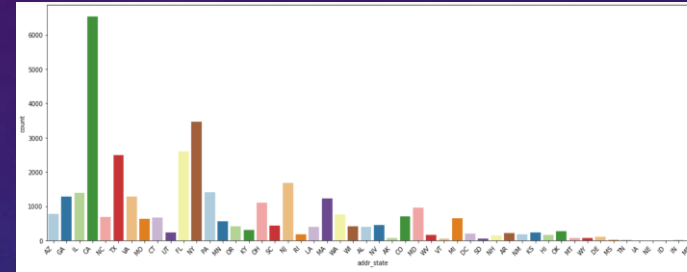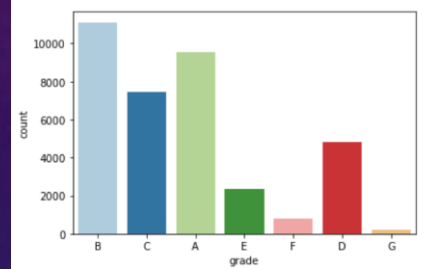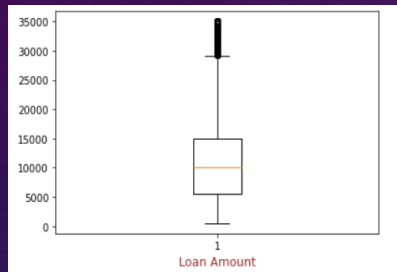
PLOTTING
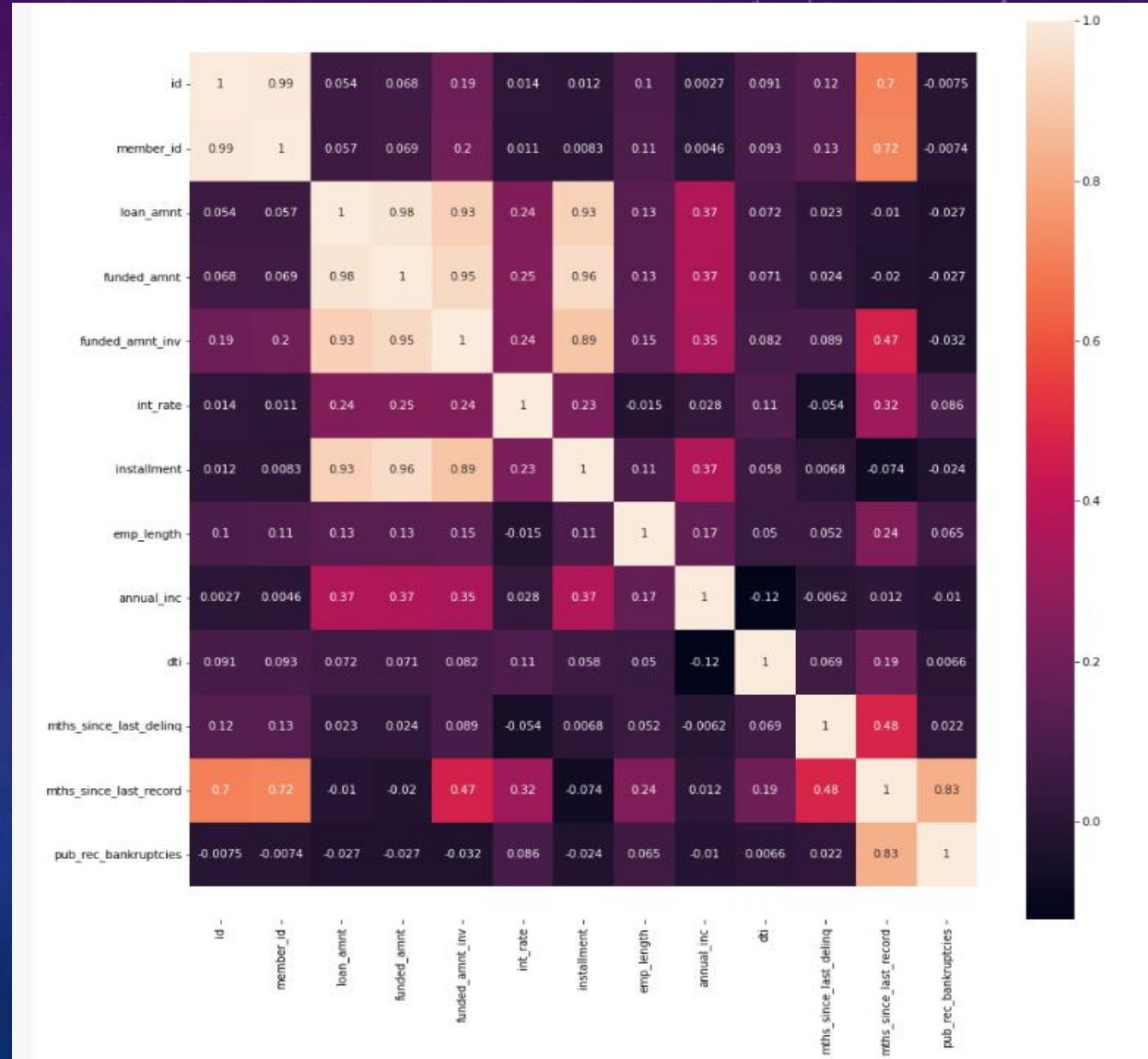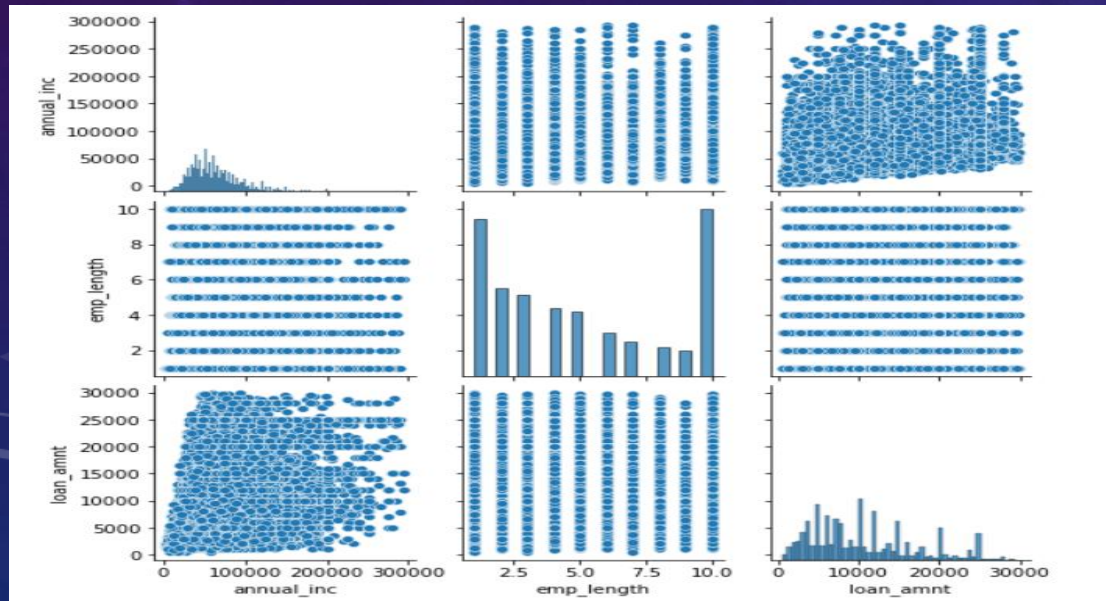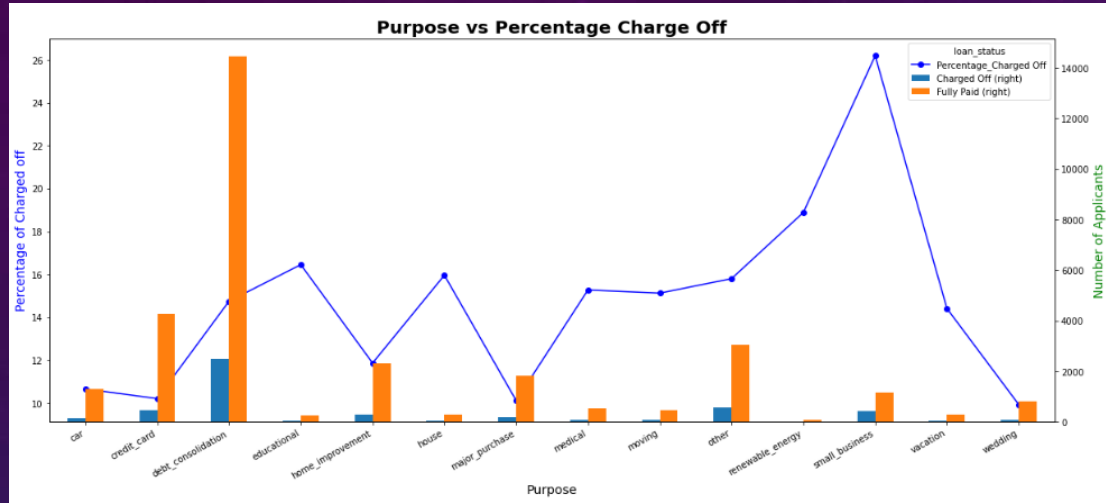
INSIGHTS AND RECOMMENDATIONS

- Lending Club should consider accepting more loans for grades A and B. It should be cautious of loans falling in grades E, F & G

- Lending Club should be wary of the loans where the purpose is Small Business as the percentage of a loan being charged off is more. Accepting loans for the purpose of Weddings, major purchases, car, and credit cards are highly recommended

- Lending Club should consider accepting more loans from people who own a house

- Lending Club should consider accepting more loans from applicants whose annual income is > 100000 as their probability of charge off is minimal

- Lending Club should accept more loans for the term of 36 months as the % of charged-off loans is less and the no. of loan applicants is more

- Driver Variables: purpose, grade, term, int_rate, emp_length, annual_inc, loan_amnt, dti, home_ownership, addr_state, verification status;

THANK YOU