

Lending Club Case study

SUBMITTED BY:-

MANISH KUMAR

NARENDHAR TANGELLA



AGENDA



- Background



- Problem statement



- Approach and Methodologies



- Analysed Data Insights



- Conclusions & Recommendations

Background

Background

- Lending club is a largest online Consumer Finance marketplace specializing in offering a variety of loans to urban customers
- Borrowers can apply for various kinds of loans e.g. Personal, business, auto refinancing, etc.
- Investors can decide whether to finance a loan or not on the basis of loan amount, customer background and etc.
- Using various driving factors, the lending club can identify the applicants/loans that have higher chance of defaulting and reducing the risk to profitability.

Problem Statement

- ❑ Lending Club, an online Consumer Finance marketplace specializing in offering a variety of loans to urban customers, facing challenges in managing its loan approval process. When evaluating loan applications, the company must take decision to minimize financial losses, primarily stemming from loans extended to applicants who are considered “Risky”.
- ❑ The primary objective of this exercise is to assist Lending Club in mitigating credit losses. These challenges arises from two potential scenarios:
 - Correctly identifying applicants with great credit record who could positively contribute to generate revenue for the company.
 - On the other hand, reducing the loan approval for applicants not likely to repay which can lead to credit loss
- ❑ 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 utilize this knowledge for its portfolio and risk assessment.
- ❑ The objective is to identify these risky loan applicants, that helps in reducing the credit losses.
- ❑ This case study aims to identification of such applicants using EDA (Exploratory Data Analysis) on given dataset.

Approach & Methodology

- ❑ Loading dataset- loan.csv
- ❑ Analysis of null values in columns and rows of dataset.
- ❑ Analysis for unique values.
- ❑ Analysis of duplicate records or rows.
- ❑ Splitting of a column into multiple columns (for derived metrics)
- ❑ Dropping Rows & Columns
- ❑ Creating Common Functions
- ❑ Data Conversion / Standardization
- ❑ Treatment of Outliers
- ❑ Imputing values in Columns
- ❑ Binning of few column values (Segmentation).
- ❑ Univariate & segmented univariate analysis across various variables.
- ❑ Bivariate analysis across various continuous & categorical variables.

Data Insights

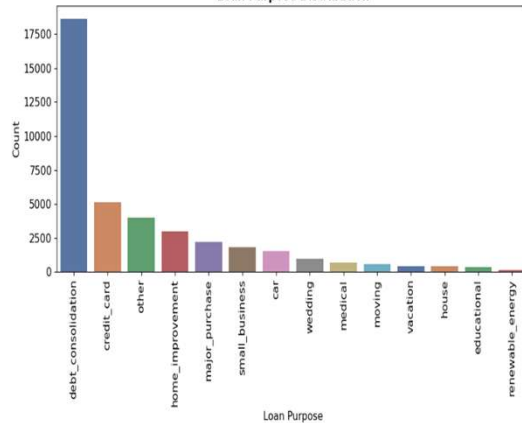
□ Univariate Analysis of loan purpose , loan grade, home ownership, grade

Highest number of customers applied for loan with 'debt_consolidation' as a purpose.

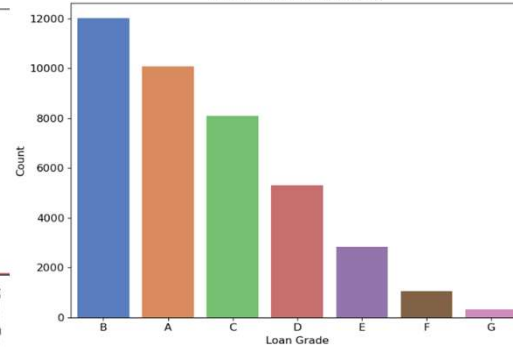
Highest number of customers are staying in 'Rented house' or house with 'Mortgage'.

Highest no of loans are of B grade and subgrade of 'A4'.

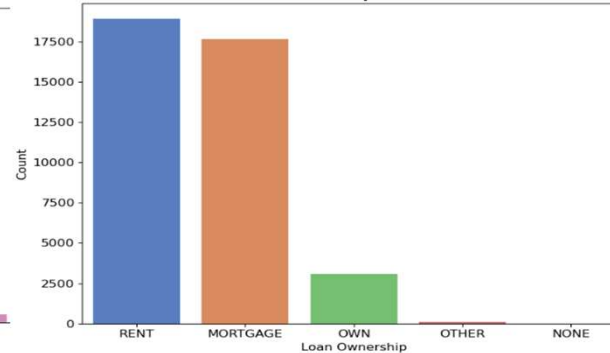
Loan Purpose Distribution



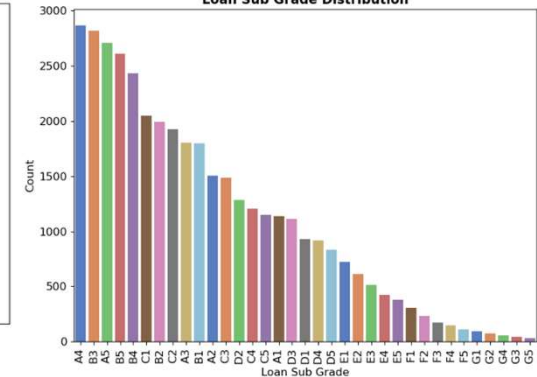
Loan Grade Distribution



Home Ownership Distribution



Loan Sub Grade Distribution

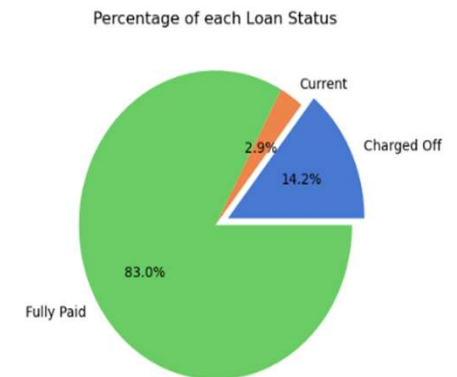
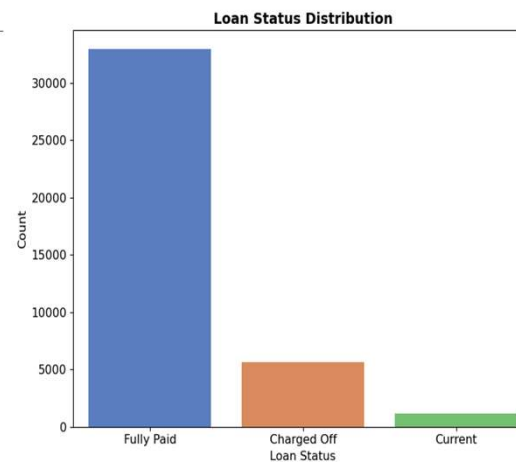
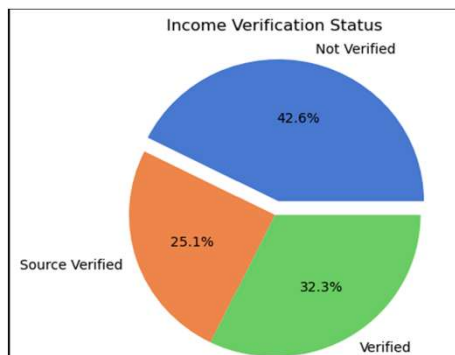
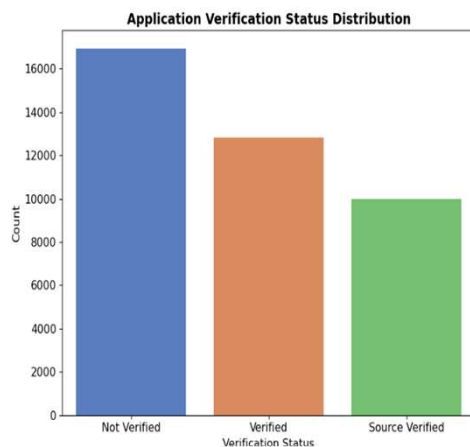


Data Insights(contd...)

- Categorical variables 'Verification Status', 'Loan status' are good indicators and we will use this for further analysis :-

Income source for 42.6% of successful loan applications is not verified.

14.2% of total approved loans are 'Charged Off' meaning they are delinquent.



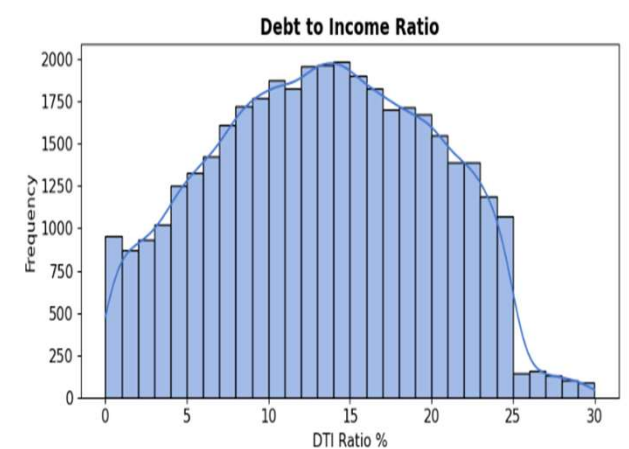
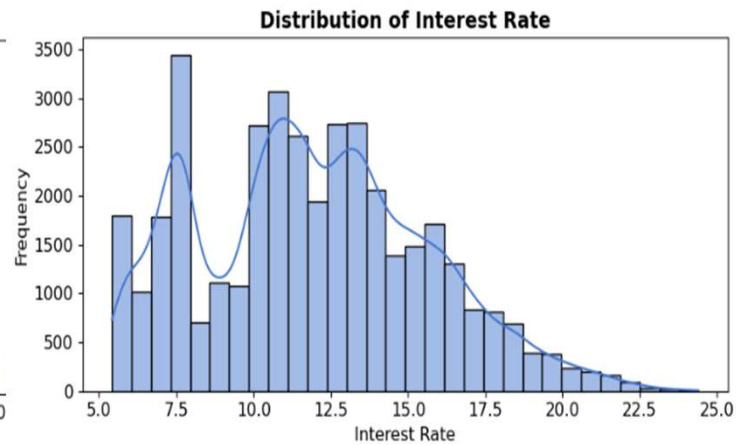
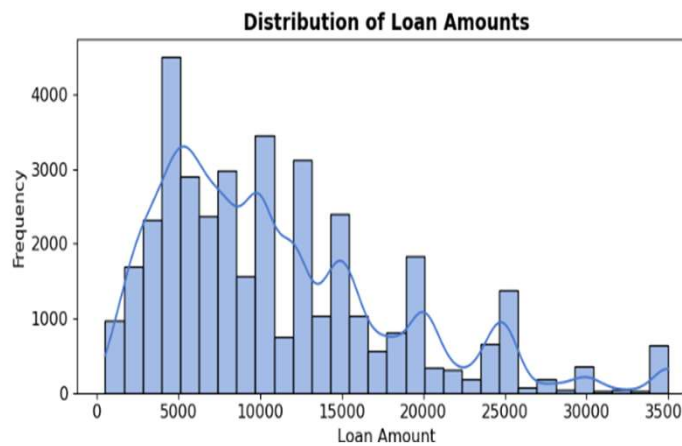
Data Insights(contd...)

❑ Distribution of Loan Amount, Interest Rate, Debt to Income Ratio (dti)

Most of the loans have loan amounts between 5000 and 16000.

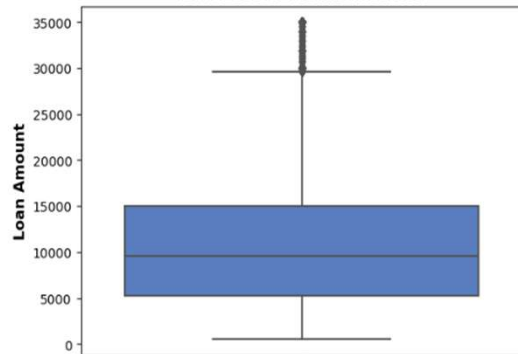
Interest rates offered on majority of the loans are between 7% and 16%.

Debt to Income ratio of the customer on most loans is between 5% and 25%

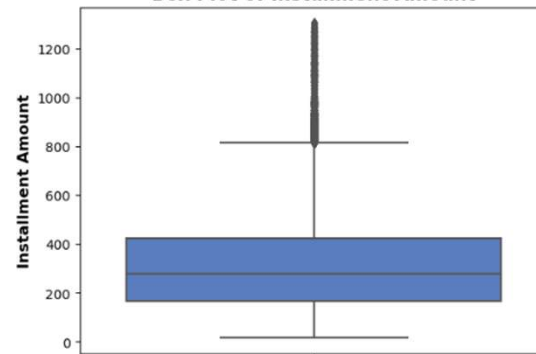


Data Insights(contd...)

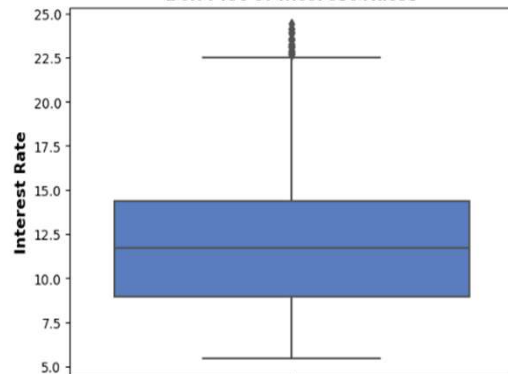
Box Plot of Loan Amounts



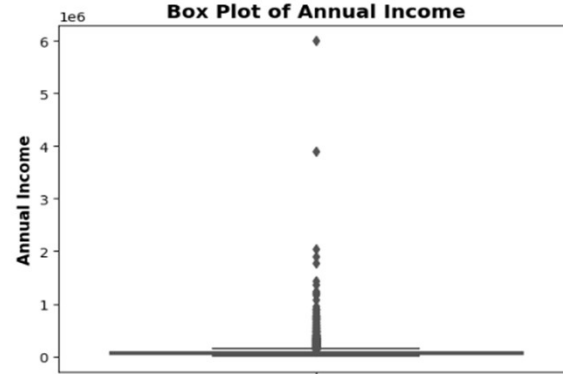
Box Plot of Installment Amount



Box Plot of Interest Rates



Box Plot of Annual Income

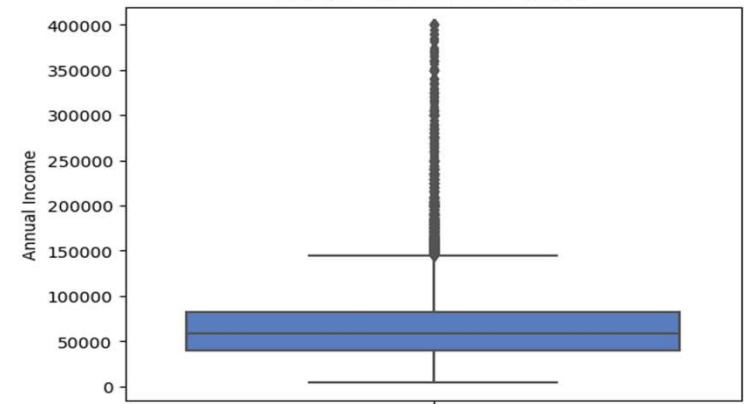


Outlier Analysis

Annual Income has some outliers

Outliers have been addressed

Annual Income Outlier Check



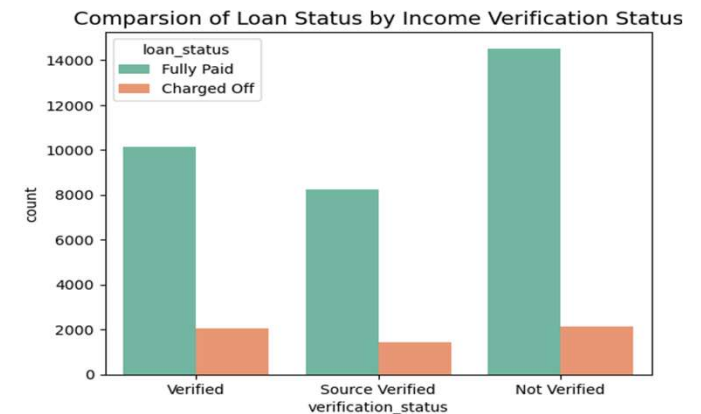
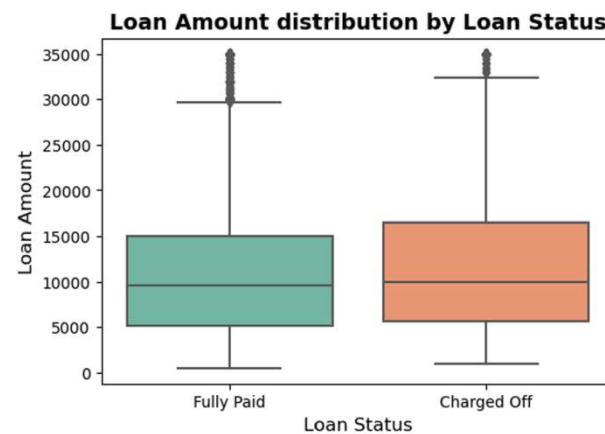
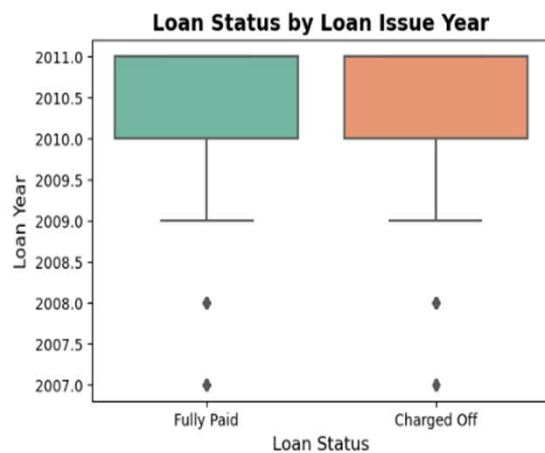
Data Insights(contd...)

❑ BIVARIATE ANALYSIS OF LOAN STATUS

Most Loans issued in 2010 and 2011 have high defaults ('Charged Off')

Most loans of size between 6000 and 16000 have high defaults ('Charged Off')

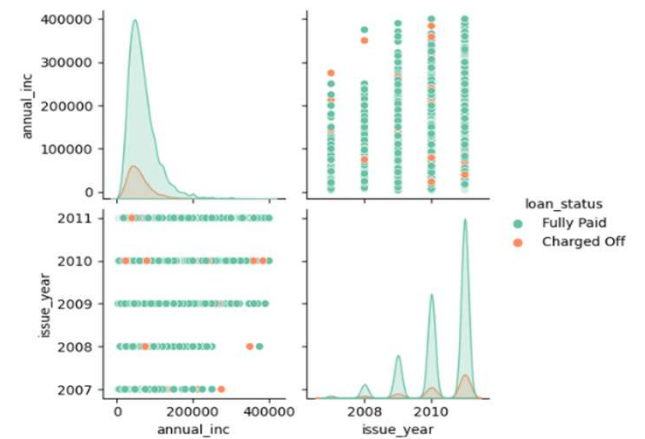
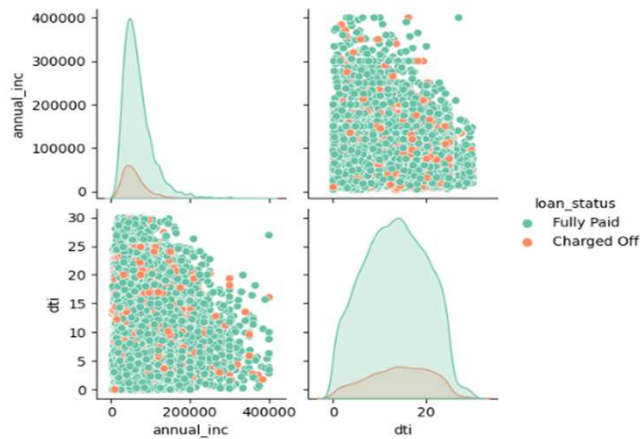
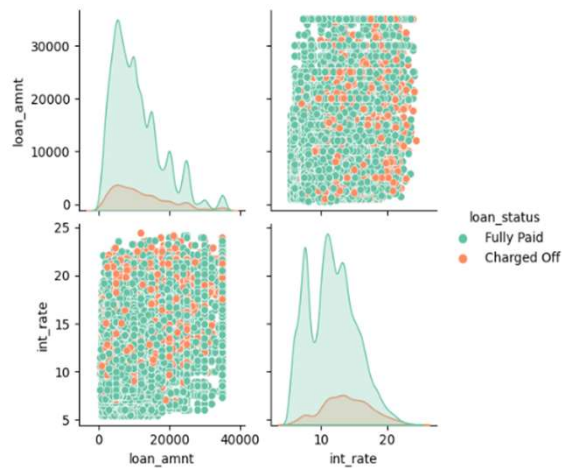
Surprisingly the income verification status has no major impact on Loan Status



Data Insights(contd...)

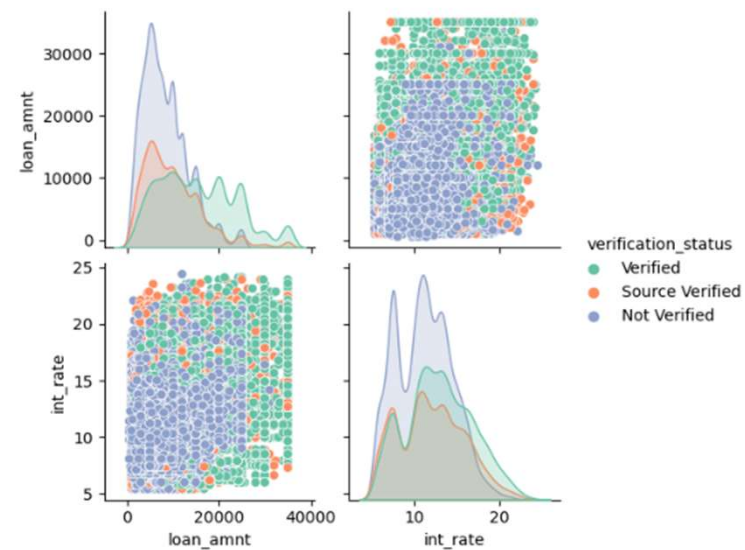
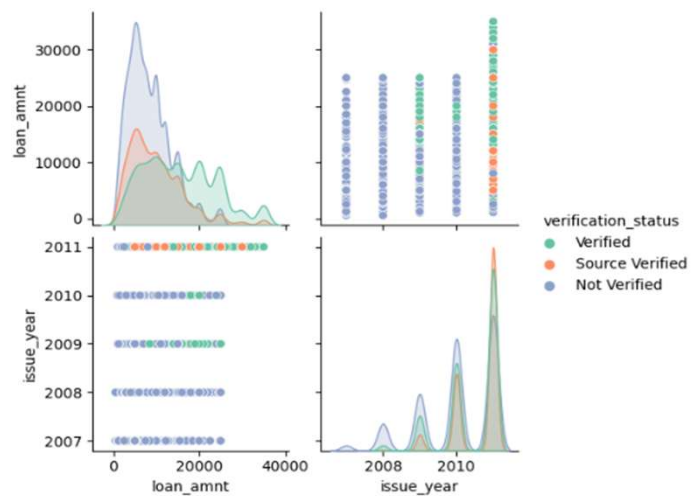
Most 'Charged Off' loans have higher interest rate between 10% and 20%

Applicants with lower income have higher chance of defaulting(Charged Off)



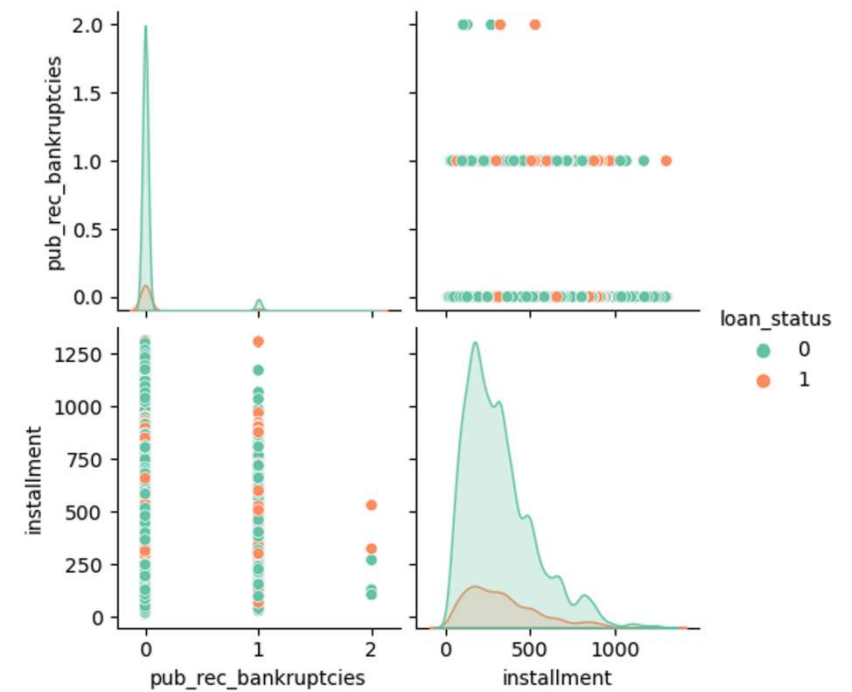
Data Insights(contd...)

- ❑ Income verification was low in 2010 which could have contributed to 'Charged Off' loans in 2010.
- ❑ Applicants income was not verified for small loan amounts



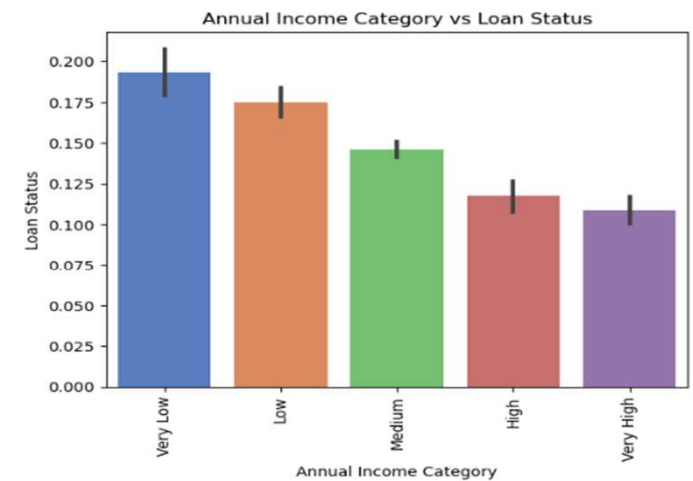
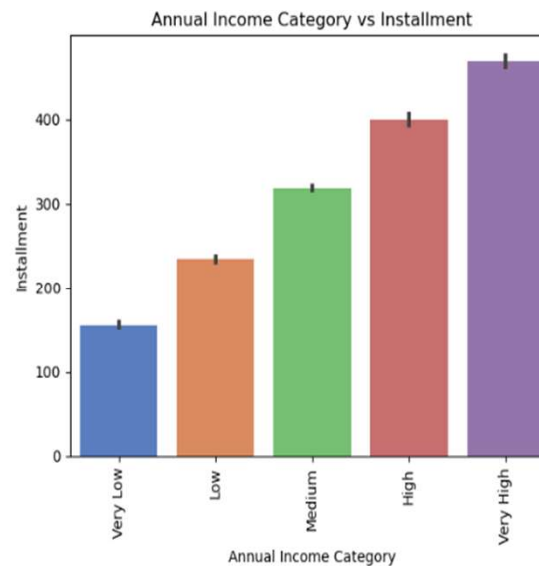
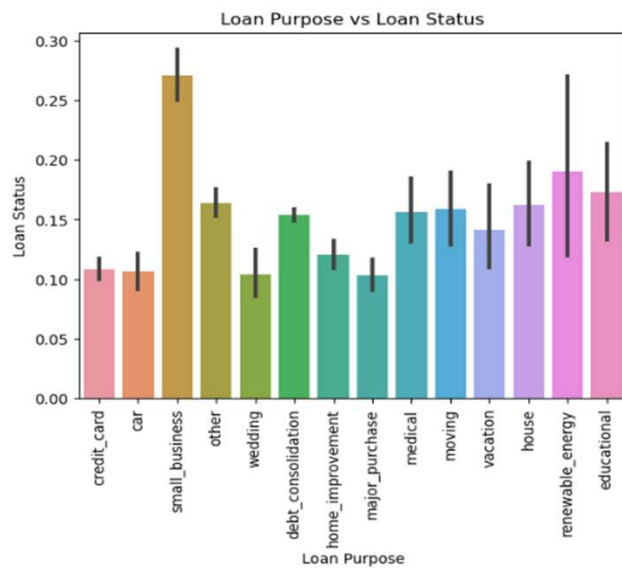
Data Insights(Contd..)

- ▶ Customers with one or more bankruptcies have higher chance of default



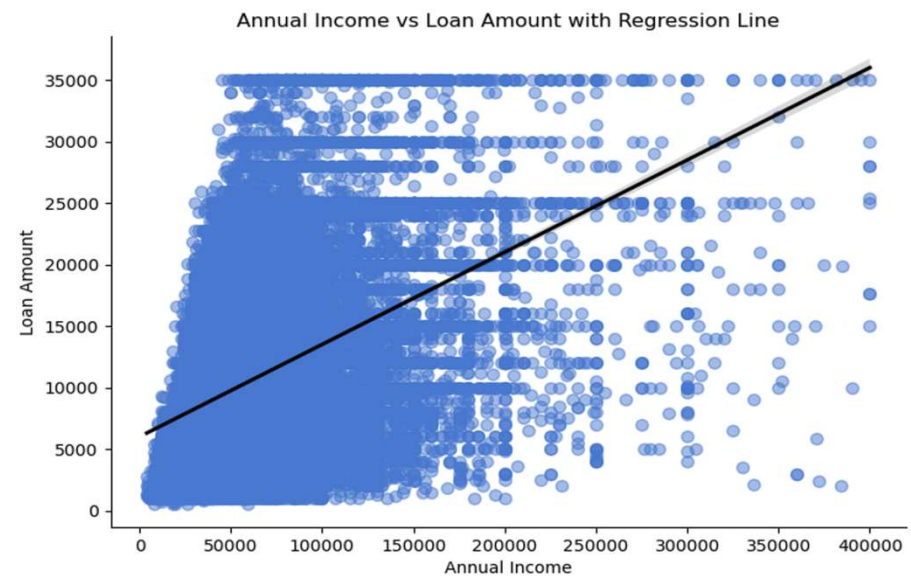
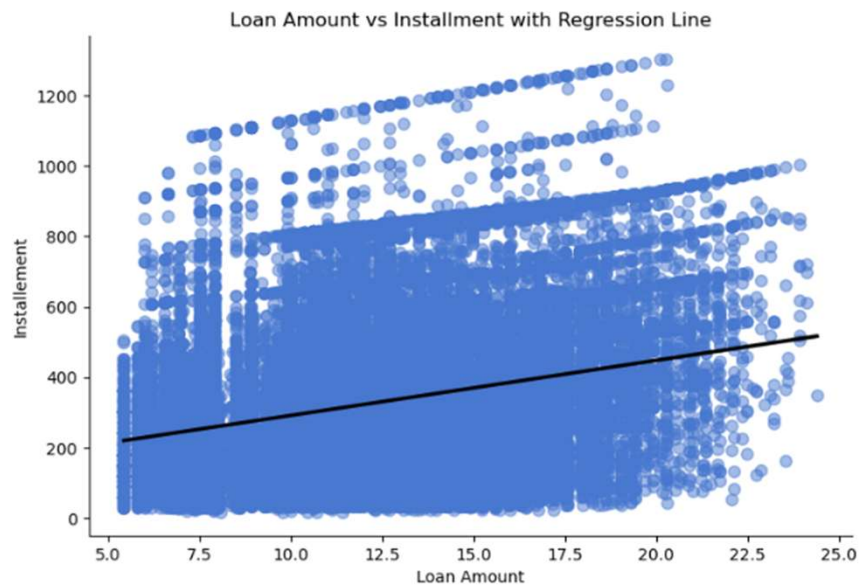
Data Insights(contd...)

- ❑ Majority of small business loans have high default (1- 'Charged Off') rate.
- ❑ Customers with very low and low income have higher default (1-'Charged Off') rate.



Data Insights(contd...)

- ❑ Instalment amount increases linearly to Loan Amount
- ❑ Loan Amount increases linearly to Annual Income



Data Insights (contd...)

Correlation Matrix for continuous variables

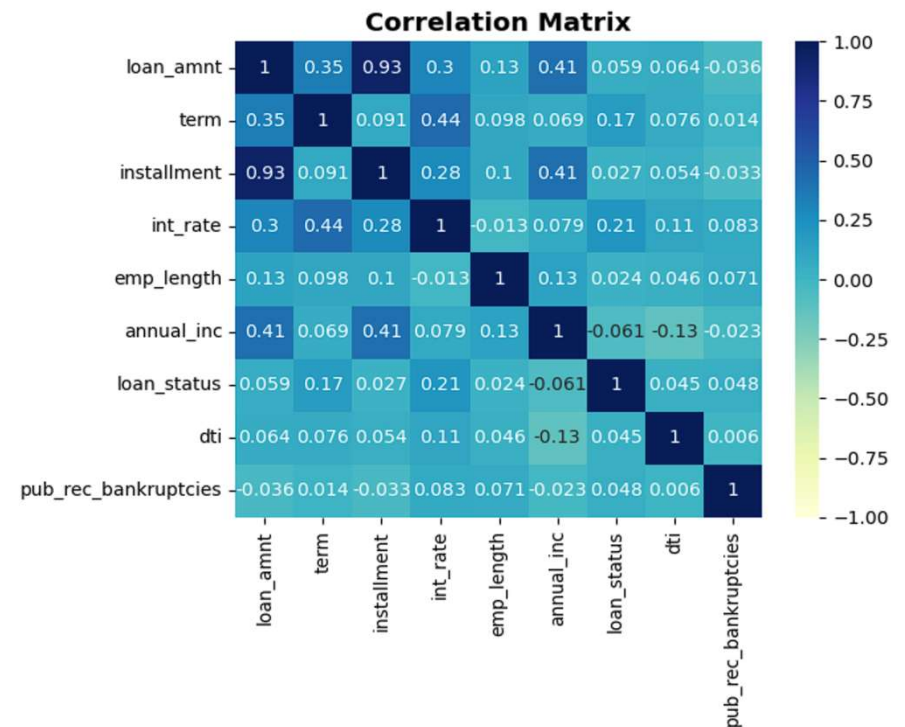
Instalment has high positive correlation to Loan Amount

Annual Income has positive correlation to Loan Amount

Interest rate has positive correlation to Term

Interest rate has positive correlation to Loan Amount

Annual income has negative correlation to Debt to Interest ratio



Recommendations

Recommendations

- ▶ Since the small loan amounts have higher default rate and lower income verification, Lending club could improve the income verification process
- ▶ Applicants who do not own a house or have mortgage have higher risk of default.
- ▶ Small business loans have higher default rate. Lending club could increase the income verification process to reduce the default rate.
- ▶ Customers with one or more declared bankruptcies have higher chance of default. Lending Club should carry out more income verification checks for such customers
- ▶ Lending Club could offer lower interest rates to low income customers if income is verified which contributes to reduction of default rate.

References & Useful Links

Technologies & Packages Used:

- ❑ Python
- ❑ Matplotlib
- ❑ NumPy
- ❑ Pandas
- ❑ Seaborn



THANK YOU