Lending Club Case Study

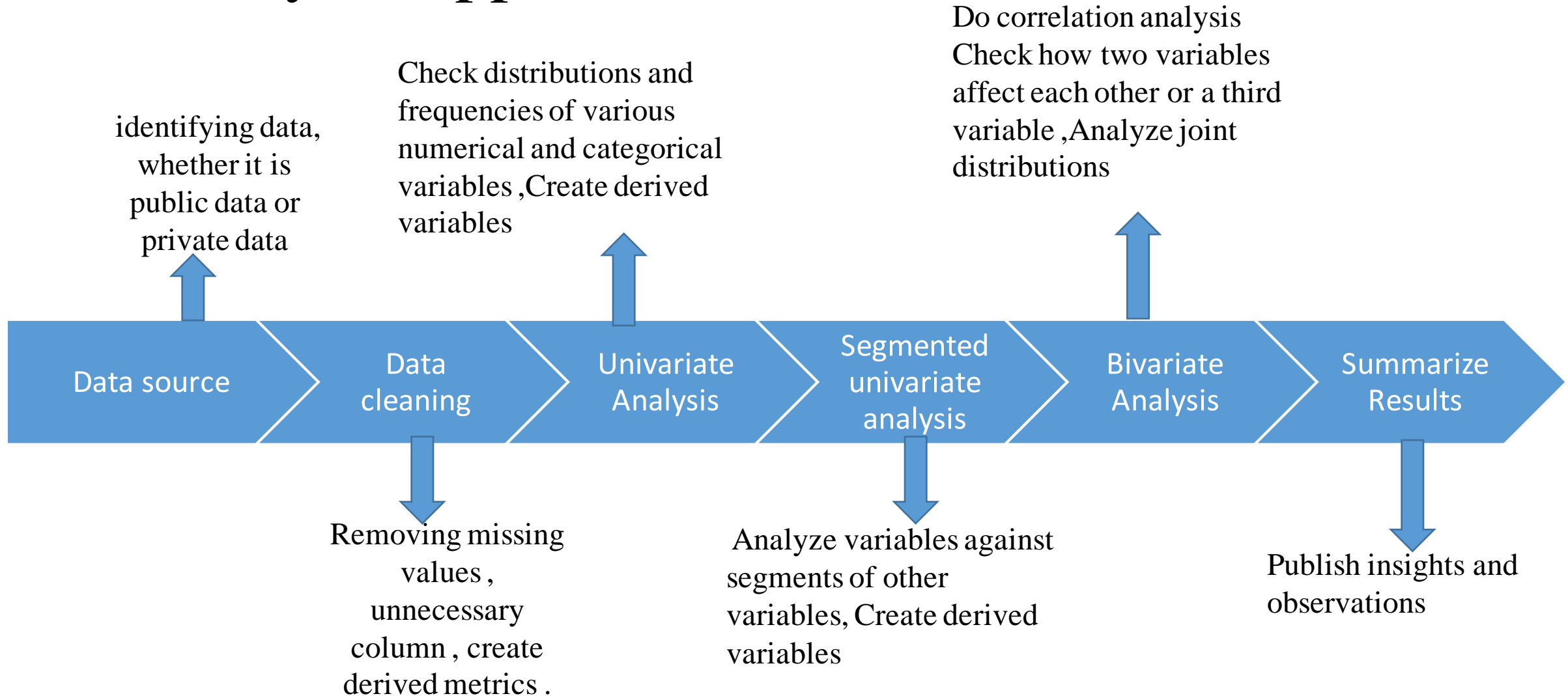Group Members:

1. Yasoda Nagella

2. Amrapali Das

Domain Background:

- Lending Club is the largest online loan marketplace , facilitating personal loans , business loans, and financing of medical procedures.

- Borrowers can easily access lower interest rate loans through a fast online interface.

- Lending Club wants to understand the **driving factors behind loan default,** i.e. the **driver variables** which are strong indicators of default.
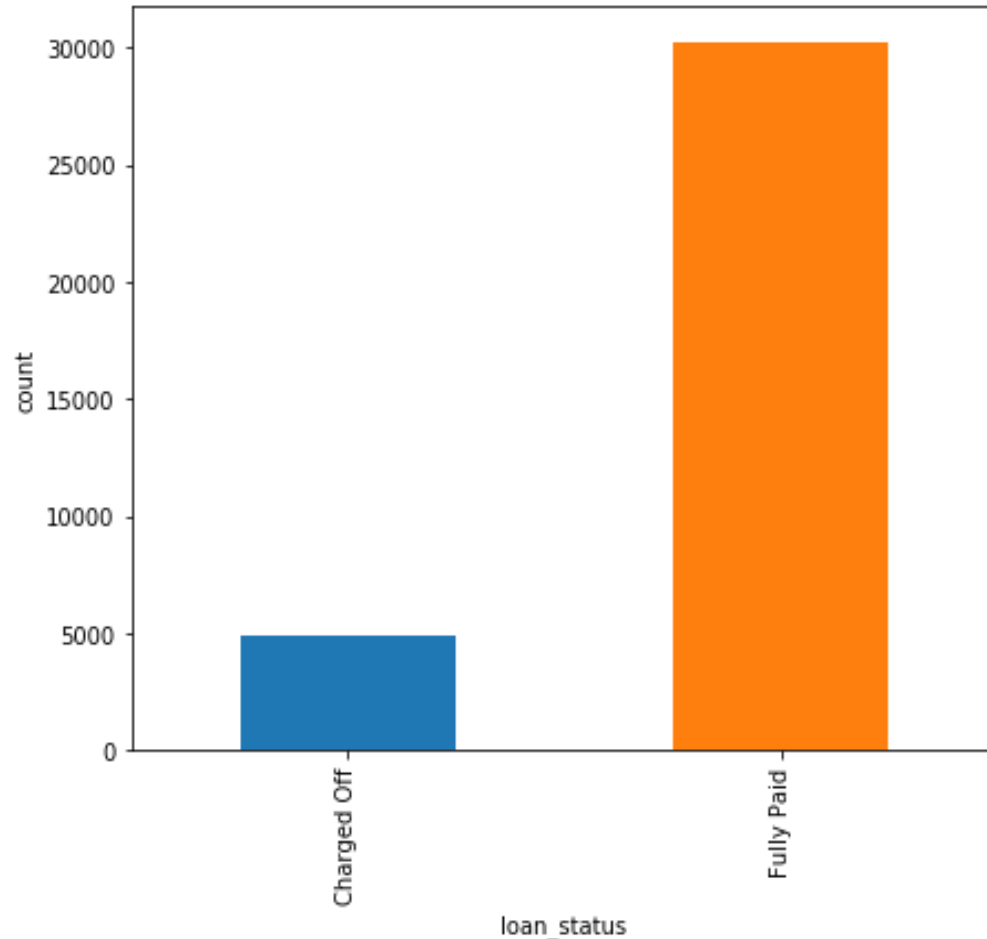
Business objective:

- 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 utilise this knowledge for its portfolio and risk assessment.
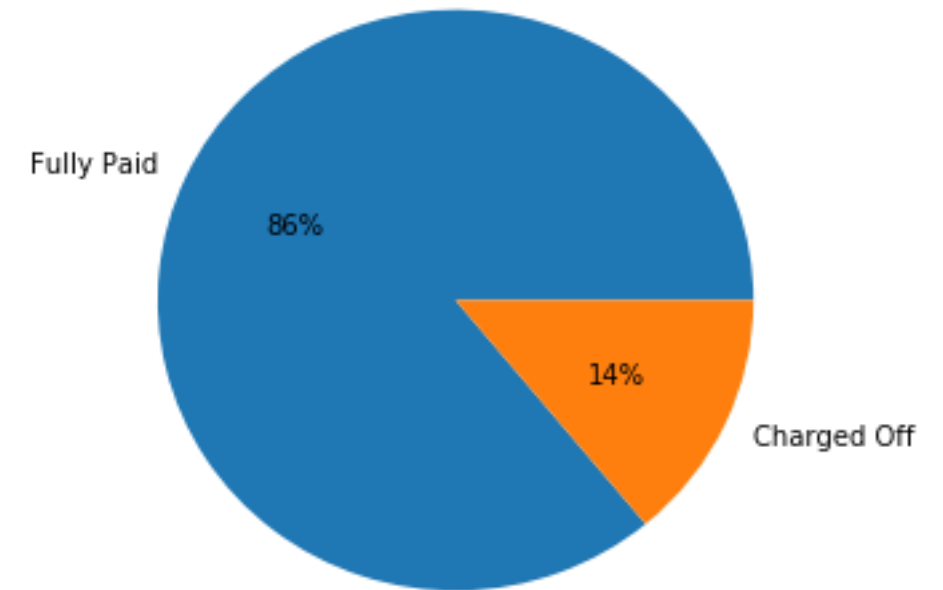
Analysis Approach

identifying data, whether it is public data or private data

Check distributions and frequencies of various numerical and categorical variables ,Create derived variables

Do correlation analysis Check how two variables affect each other or a third variable ,Analyze joint distributions

Data source → **Data cleaning** → **Univariate Analysis** → **Segmented univariate analysis** → **Bivariate Analysis** → **Summarize Results**

Removing missing values , unnecessary column , create derived metrics .

Analyze variables against segments of other variables, Create derived variables

Publish insights and observations

Analysis: Loan Status

Percentage of loans fully paid, charged off

Approximately 14% of loans are defaulted
Any variable that increases percentage of default to higher than
16.5% should be considered a business risk.

Analysis: Purpose

highest number of borrowers have taken debt consolidation loan, followed by credit_ card oan, followed by other loan, home_improvement loan, etc
debt consolidation is by far the largest 'purpose' for seeking a loan means Approx most of the applicants applied loan for paying their other loans(Debt Consolidation), default rate is high in debt_consolidation .
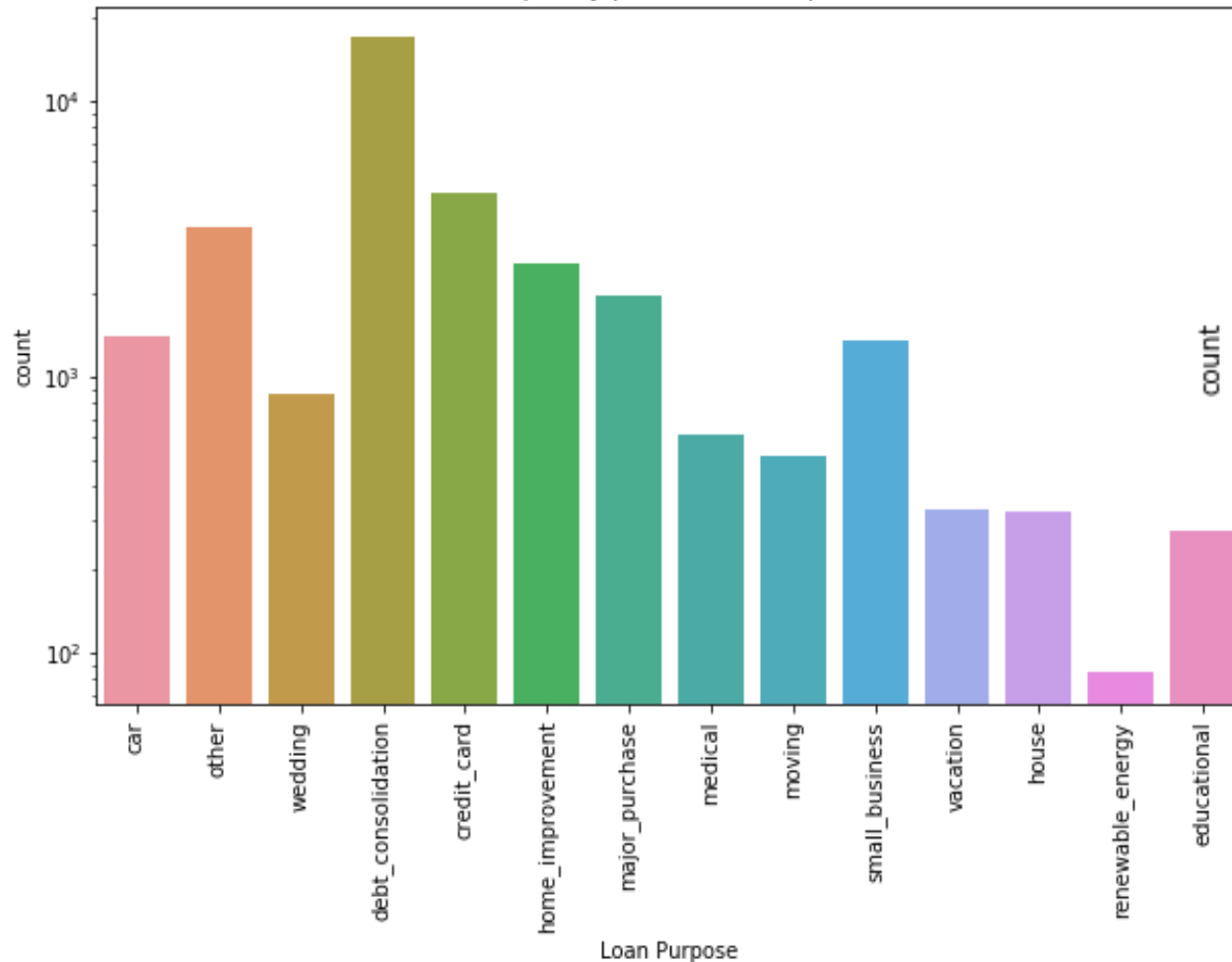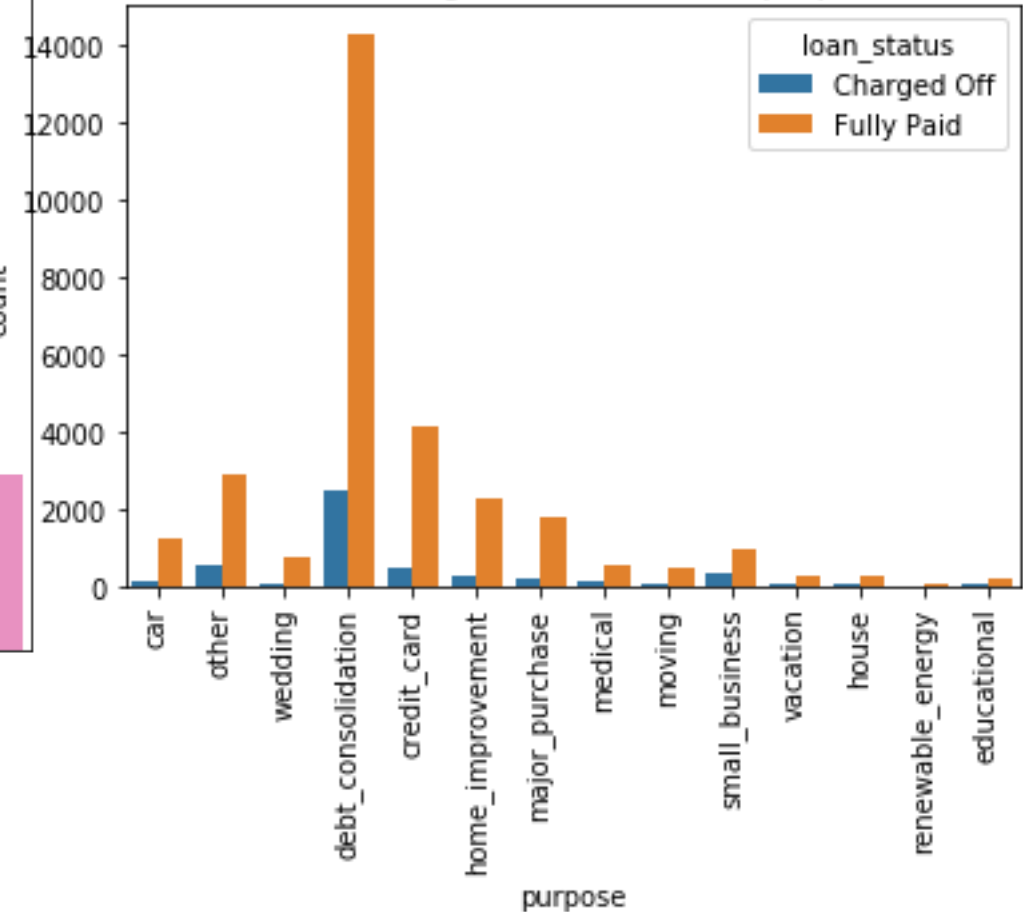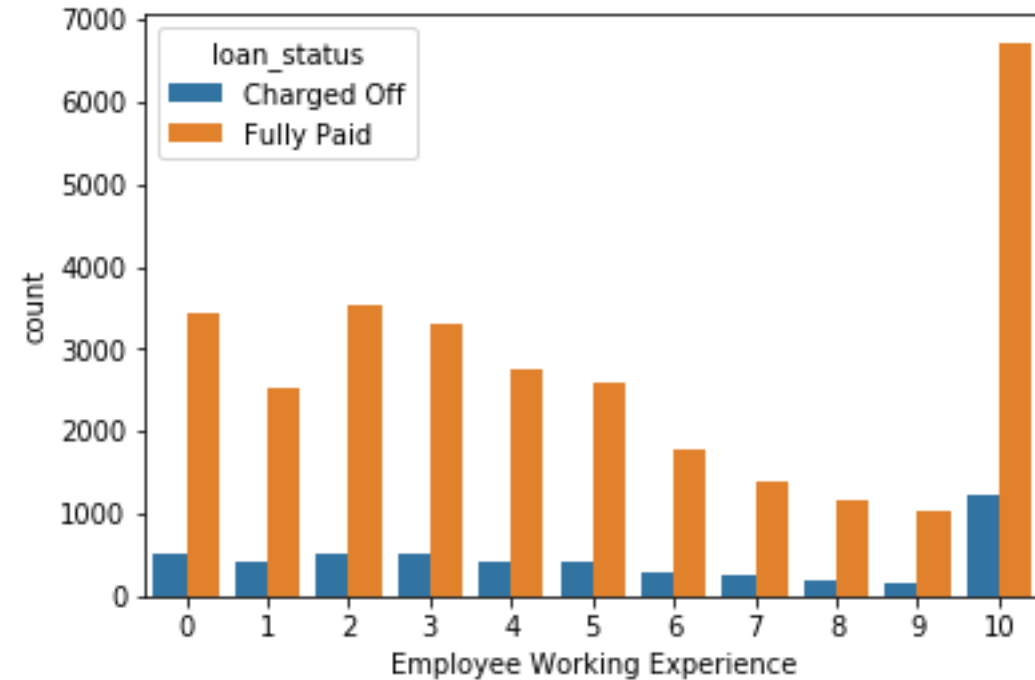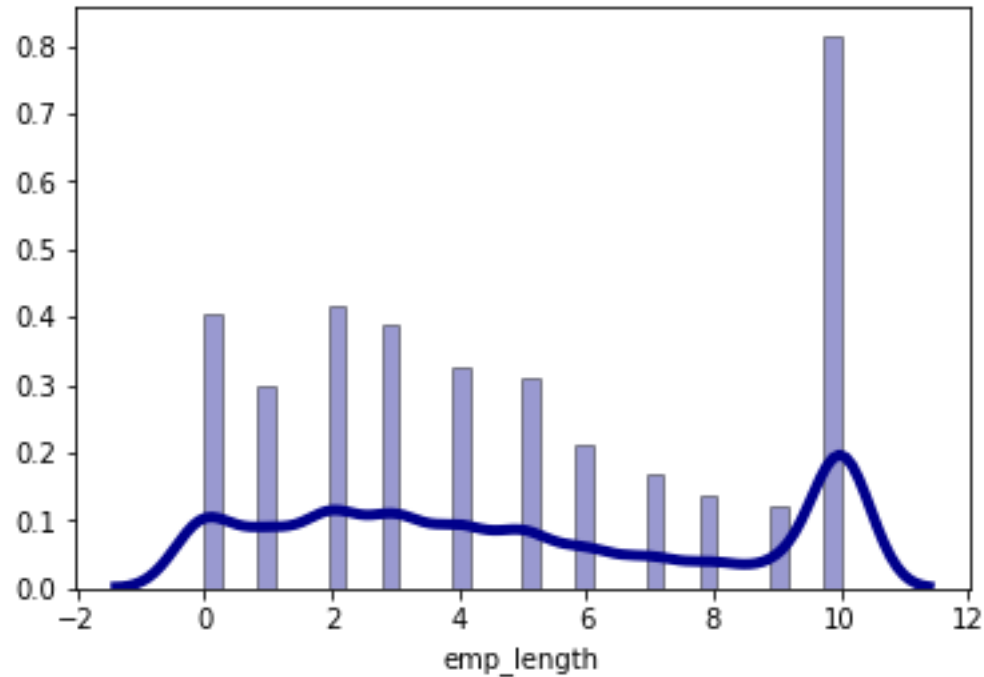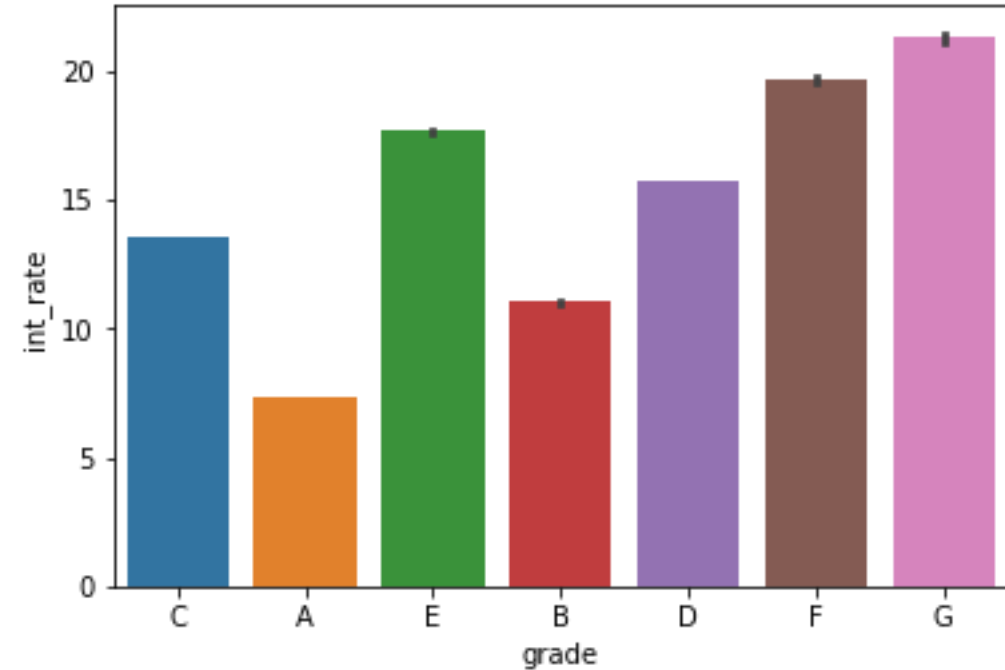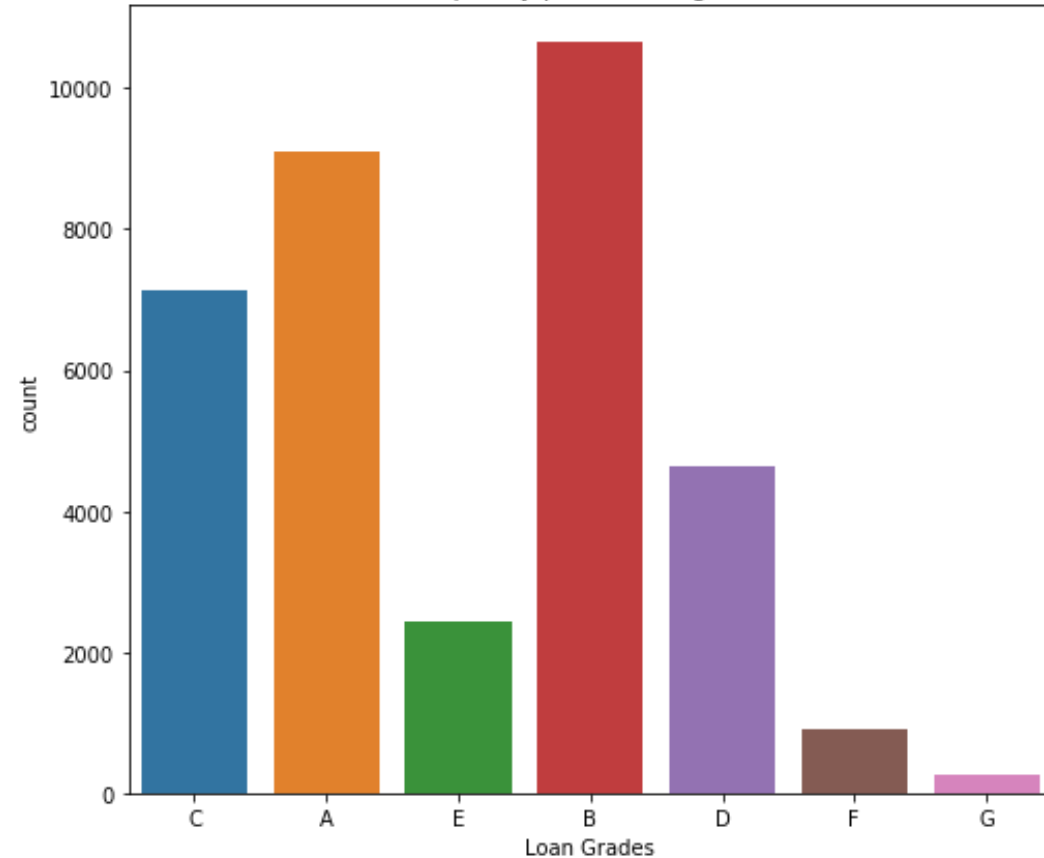
Analysis: Employment length

➤ barrowers with >10 years of employment period have took most loans and also has defaulted the most.

Analysis: Grade

Frequency plot of loan grades

➤ Loan quality increases, interest rate decreases. A being highest quality loan and G being lowest.

➤Finding from performing Univariate analysis on Grade column. Grade is a ordered categorical variable. A < B < C < D < E < F < G

➤Grade A means lower risk for lending club and Grade G means higher risk for lending club.
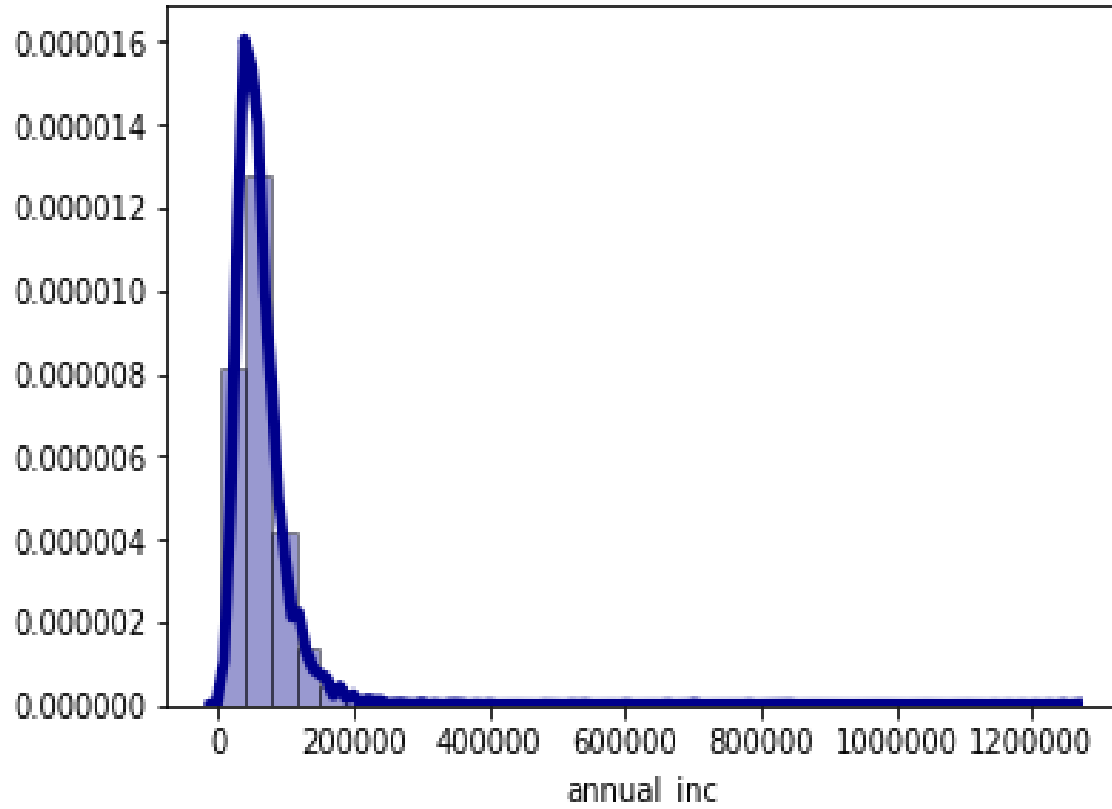
Analysis: Interest rate
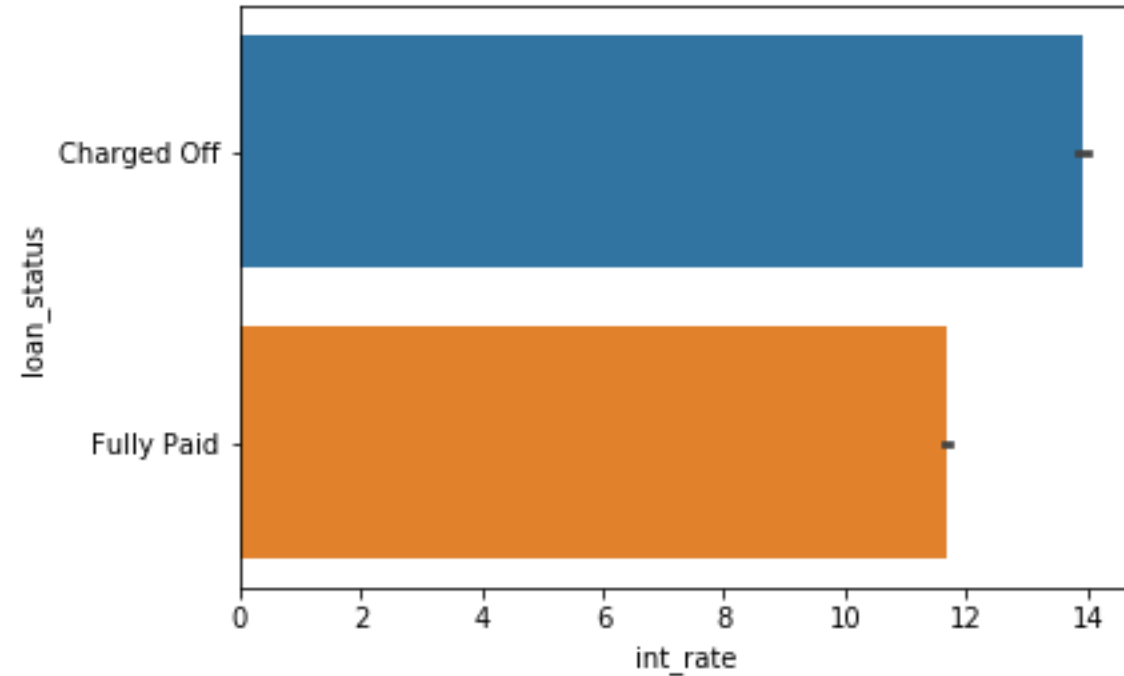
➢Defaults increases monotonically with higher interest rates
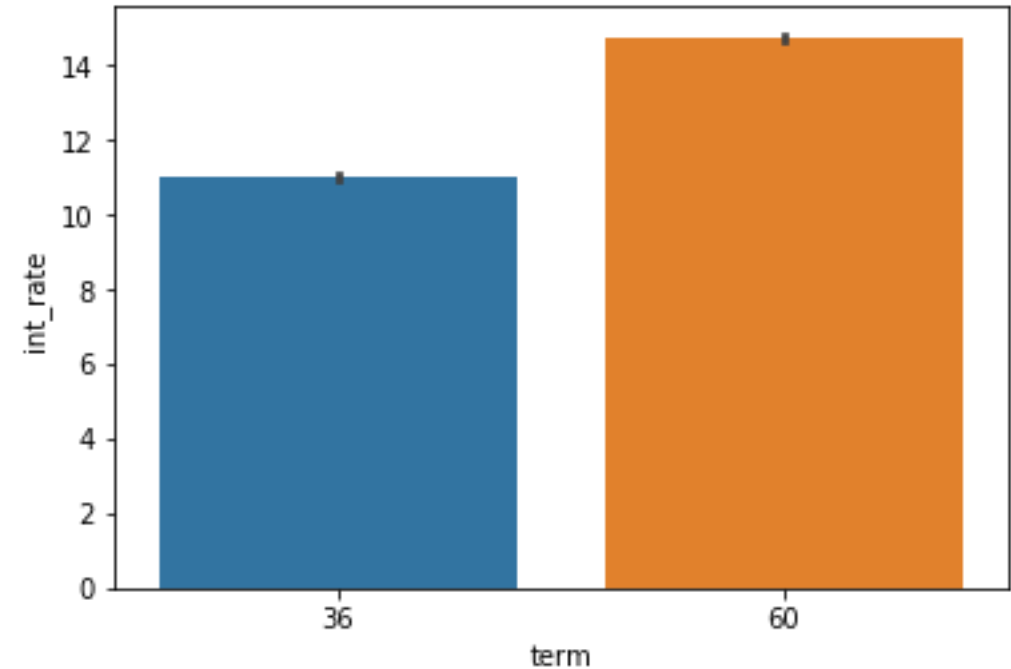
➢Intrest rates are besed on term. Larger amounts were seen to be given for higher term. Ther rate of intrest associated with them is also high.
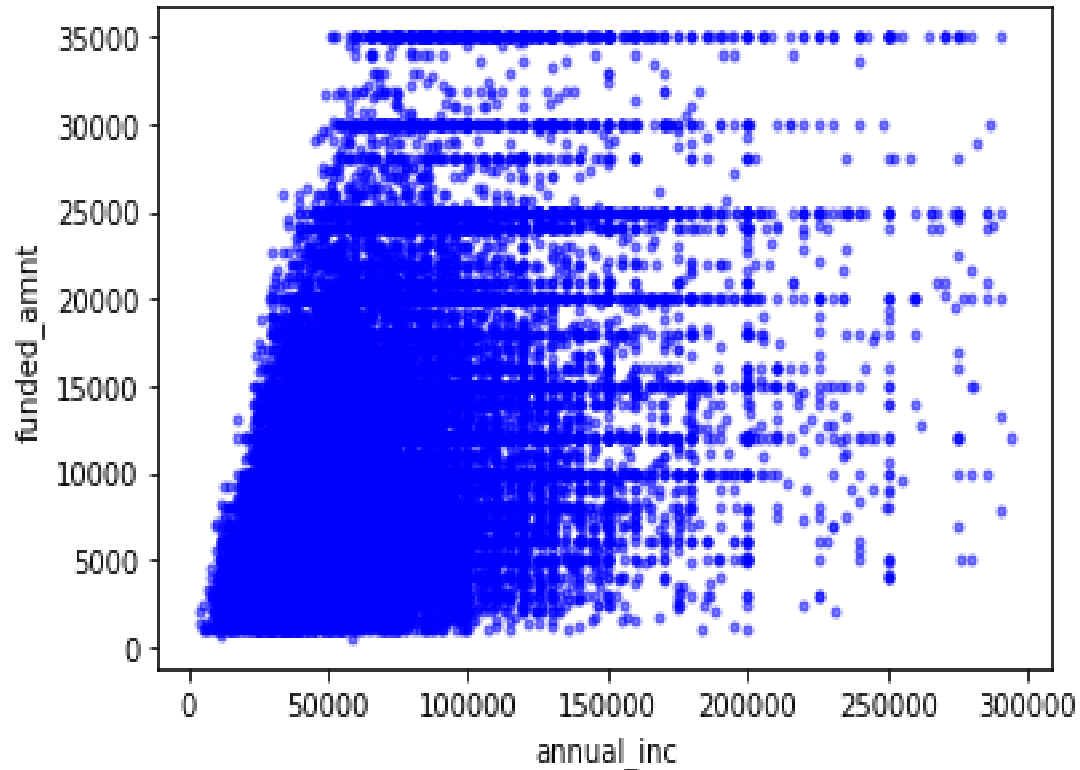
Analysis: Annual income

➤ from the plots it seems the majority of defaulters belong to income levels between 0-200000.

Analysis: Annual income

Scatter plot indicating relation between annual income and funded amount

➢We see here that Lending Club has extended high-value loans to people with low income.

➢There are many cases of people with income 50000 or less getting loans of 25000 or more.

➢This practice should be curtailed.

Analysis: dept to income ratio

Scatter plot indicating relation between interest rate and dept to income ratio
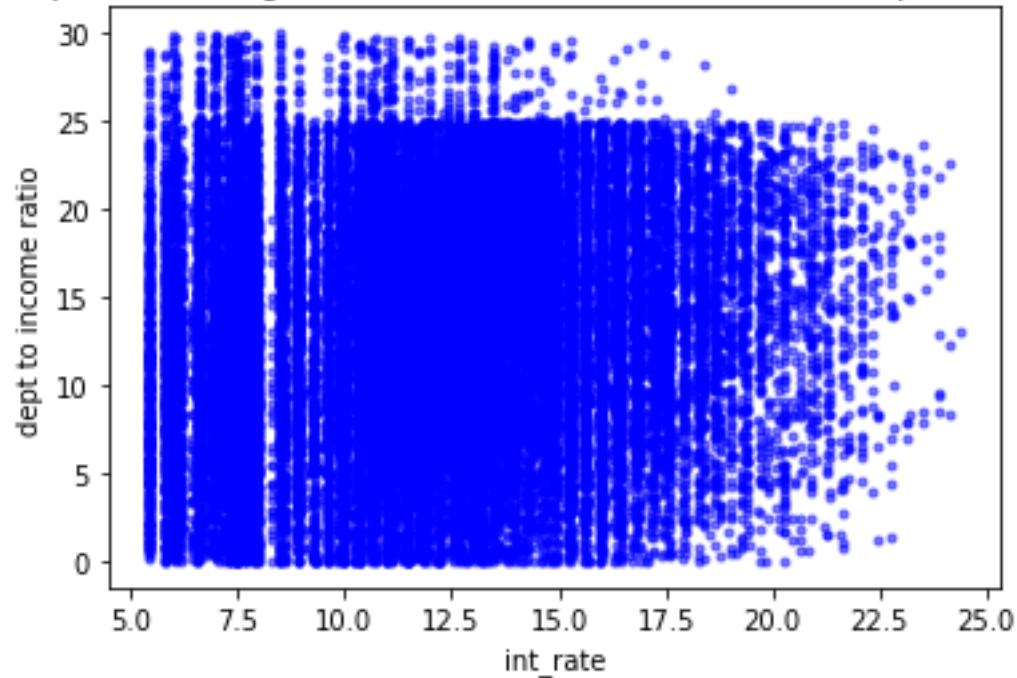
➢Percentage of default rises with dti ratio.

➢As the dti ratio rises above 20, the loans become risky.

➢Higher interest rates should be charged for higher dti

Conclusion

Target Variable

1. Loan Status

Top-5 Major Variables to consider for loan prediction: i.e. variables which are strong indicators of default.

1. Purpose of Loan

2. Employment Length

3. Grade

4. Interest Rate

5. Term

Conclusion

➢Loan amounts of 30% of annual income or higher see a high rate of default.

➢loans taken for the purpose of running a dept consolidation are defaulters

➢charging lower interest rates for loans with dti greater than 20 will be a defaulter.

➢As the loan amount increases, the chance of default also increases.