Loan Fulfillment

A report by Jorge Londono

Problem Statement

- Bank loans that are unsecured can be very risky for lenders. If the borrower does not fully pay the loan, the lender can end up losing money.

- It is very important for lenders to predict whether a loan applicant will be able to fully pay the loan or not.

- What are the best predictors we can find to foresee the if someone will default or charge off the loan?

Data

Origin of Data

- This data was provided by a kaggle user who collected the data from all of the Lending Club loans from 2007 to 2018.

- The data contains over 150 variables for each loan and its corresponding borrower.

Types of Variables

Categorical Variables:

- Loan Terms
- Application Types

Continuous Variables:

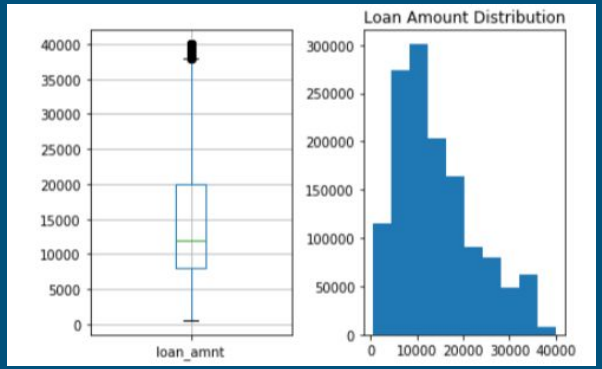
- Loan Amounts
- Annual Income

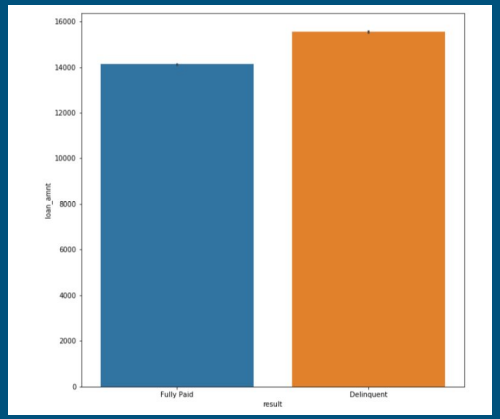
Data Wrangling

- Data is so large that it needs to be loaded in by chunks
- Loans that are not Fully Paid, Default, or Charged Off are removed.
- We will create a new column called result which will have 1 of 2 inputs: Fully Paid, or Delinquent.
- Our Data Frame has a total of 1,346,111 observations

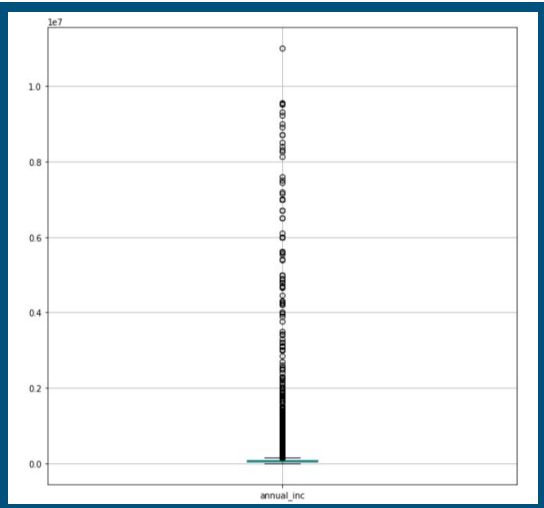
	id
result	
Delinquent	269360
Fully Paid	1076751

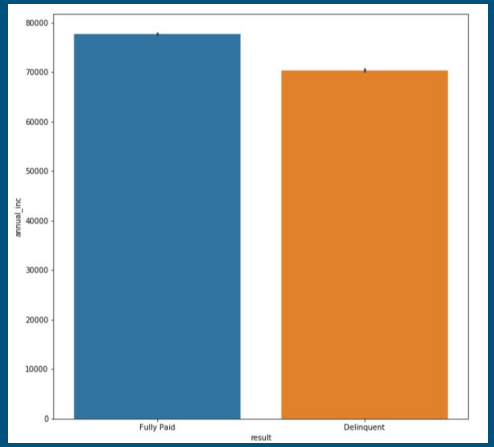
Exploratory Data Analysis - Continuous Variables





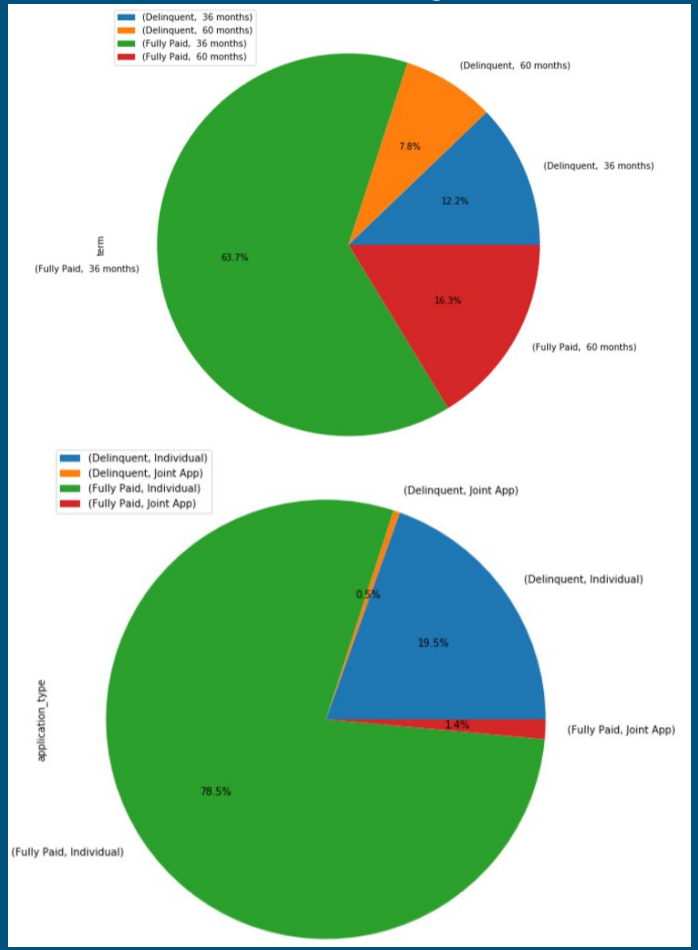
- The first row of graphs show the total loan amount.
- The data skews to higher loan amounts.
- The average loan amount is greater for delinquent accounts.





- The second row of graphs show the annual income of borrower
- The data skews to higher annual income
- The average annual income is greater for fully paid accounts accounts.

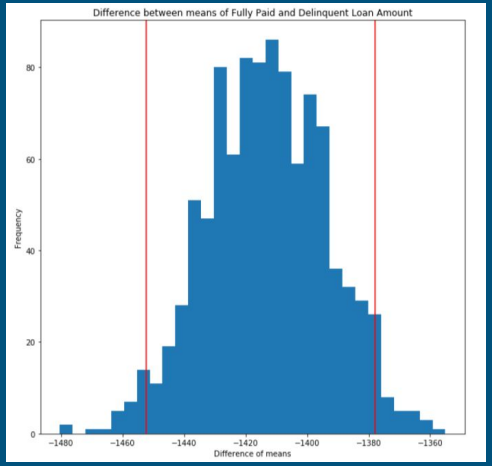
Exploratory Data Analysis - Categorical Variables



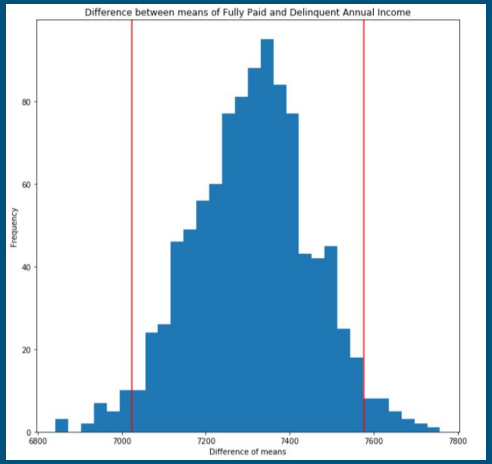
- From this pie chart we can see that the proportion of delinquent loans versus fully paid loans is ls larger when there are 60 month payment terms than 36 month payment terms.

- This pie chart shows us that around ⅓ joint applicants end up with delinquent loans while about ¼ individual applicants end up fully paying off the loans. This shows us that joint applicants are more likely to not fully pay their loan.

Bootstrapping



- Null Hypothesis: There is no difference in the mean loan amount of fully paid and delinquent groups
- Alternate Hypothesis: There is a difference in the 2 means
- Because this interval does not include 0, at 95% significance we can reject the null hypothesis



- Null Hypothesis: There is no difference in the mean annual income of fully paid and delinquent groups.
- Alternate Hypothesis: There is a difference in the 2 means
- Because this interval does not include 0, at 95% significance we can reject the null hypothesis