

Predicting Loan Defaults

...

Springboard Capstone Project 1

Problem Statement

Will a borrower default on their loan and have their loan charged off?

Data

Loan data going back to 2007 taken from the LendingClub website:

<https://www.lendingclub.com/info/download-data.action>

Data Wrangling

Data was converted to numeric format whenever possible

Delete empty columns and rows

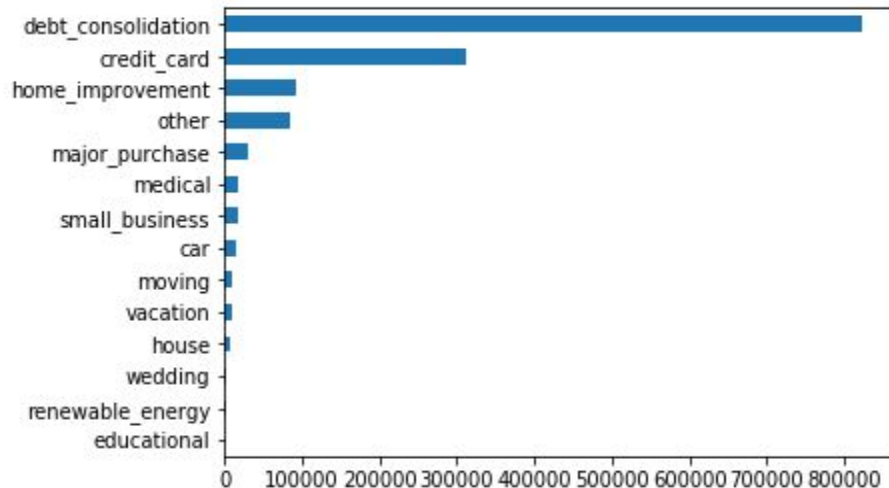
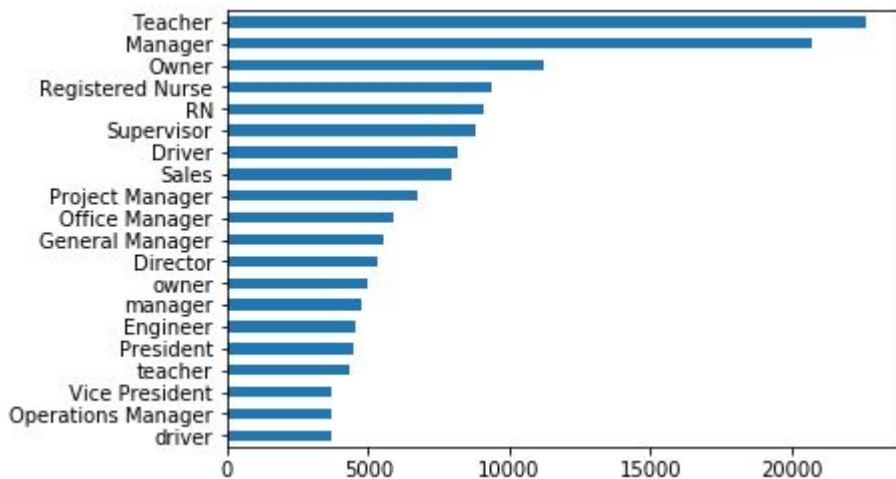
Outliers were corrected if possible and removed otherwise

Data Exploration and Visualizations

We took a look at who was borrowing money, why they were borrowing money, and how much money.

An in-depth statistical analysis was done on the incomes for Fully Paid and Charged Off/Defaulted populations

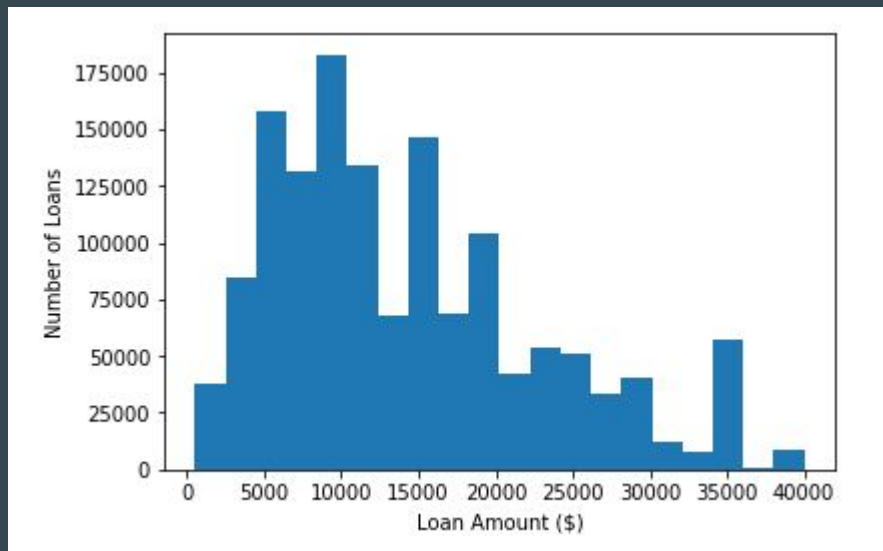
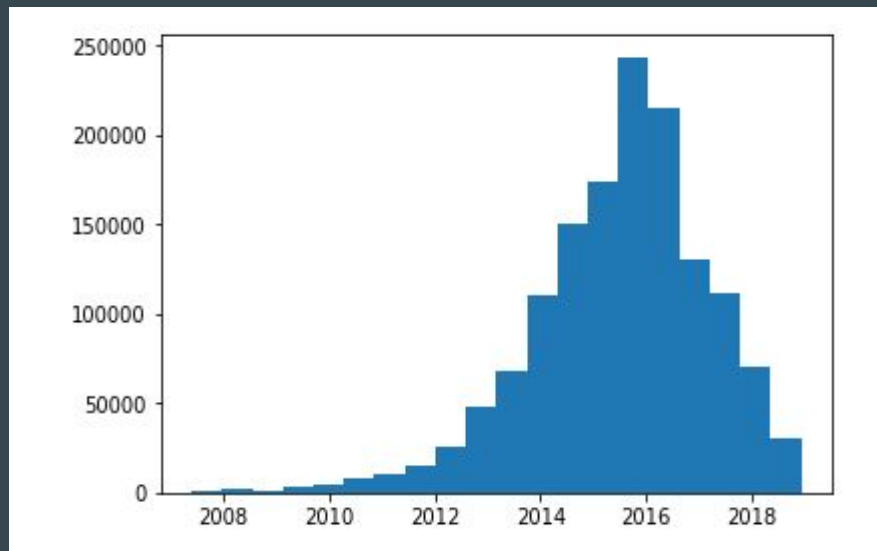
Who and why



It looks like most of the loans are going to Teachers and Managers.

Most of the loans are for Debt Consolidation.

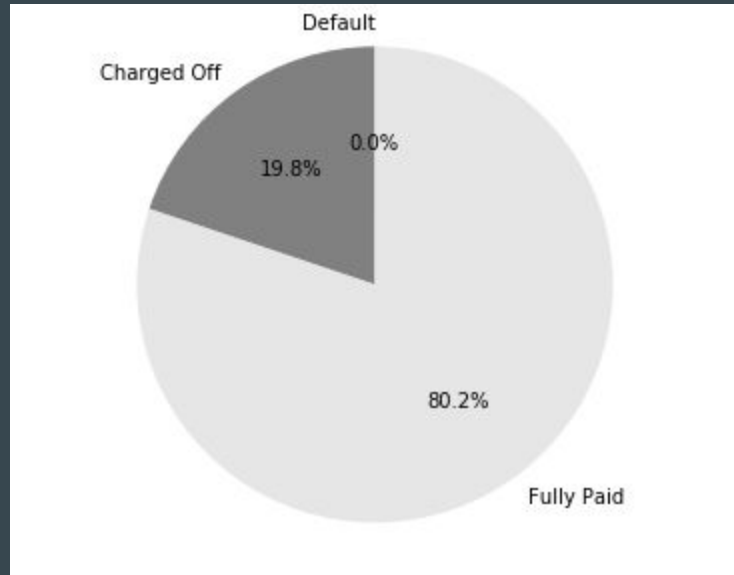
When and how much?



The loan data goes back to 2007, but most of the loan data is from 2016.

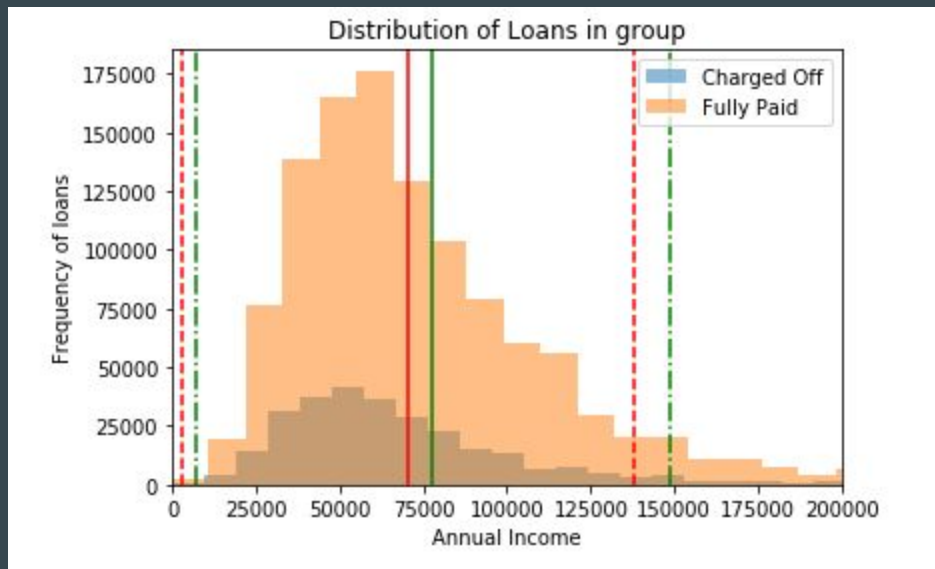
The loans range in size from \$500 up to \$40,000

How many people default?



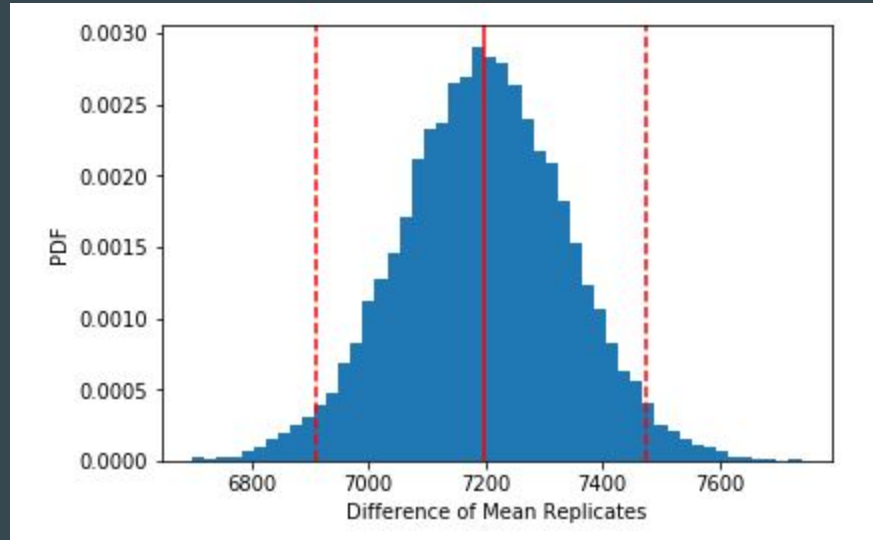
About 20% of loans have been charged off.

Distribution of Charged Off and Fully Paid loans by Income



Income does seem to be higher for the fully paid loans.

Statistical Difference in Income



We can say with 95% confidence that the difference in income between the charged off and fully paid groups is between \$6,912 and \$7,477.

Data Pre-processing

Removed Null values with average or a constant value as appropriate

Used a heatmap to determine what columns were highly correlated and remove the data

Change categorical data to “One Hot” columns

Scaled the data

Machine Learning

A number of Machine Learning algorithms were tried.

Results were compared using AUROC.

Results

Here are the results of the scoring:

1. 0.806 Catboost
2. 0.776 Random Forest
3. 0.742 Logistic Regression
4. 0.738 Ridge
5. 0.661 K-Nearest Neighbors
6. 0.500 SVC

Interpretation

The results from CatBoost were quite good.

Discuss results with business partners and tailor the results as required.