Lending Club Analysis

Developed by @krilecy and @justinhtn

Company overview

Lending Club is a financial services company headquartered in San Francisco, California. The company enables borrowers to take out unsecured personal loans between $1K and 40K.

Investors are able to search and browse loan listenings and select loans they want to invest in based on a variety of applicant data.

Desired business outcome

In order to reduce the risk of defaulted loans, investors leveraging Lending Club would like to understanding how different applicant data impact two outcomes of interest.

1. What applicat data informs future defaulted loans?
2. What applicant data informs future fully paid loans?

Our process

We took a traditional EDA approach to help provide insight towards Lending Club's business outcome of interest.

Data understanding → **Data cleaning**

What does each column in the data represent? By exploring the data dictionary provided, we were able to understand the features values and their business context to better plan for targeted EDA.

Are there column that have missing data or values unrelated to our desired business outcome? Missing or outlier data can impact conclusions drawn from any data analysis.

Data analysis → **Recommendations**

We evaluated the data and loan outcome across 3 different analytical lenses. Univariate analysis, Bivariate analysis and Derived feature analysis. It is alongside these lenses in which we aimed to understand defaulted loan feature correlation.

In order for Lending Club to take action on reducing the approval rate of high risk loan applications, we provide an analysis of 5 features that are common among defaulted loans.

Data understanding and cleaning

In our exploration, many of the features provided referred to post charge off or postpaid details. Given our business outcome of interest is to understand the influence of features on loan defaults, we quickly removed these features from consideration.

63%

Of the the original 118 columns contained empty data or data that was not relevant for our analyses.

Data irrelevant for this analyses may be informative analysis of different future outcomes of interest. The exclusion of this data from our analysis, does not imply this data is not valuable for other downstream business needs.

Lessons learned

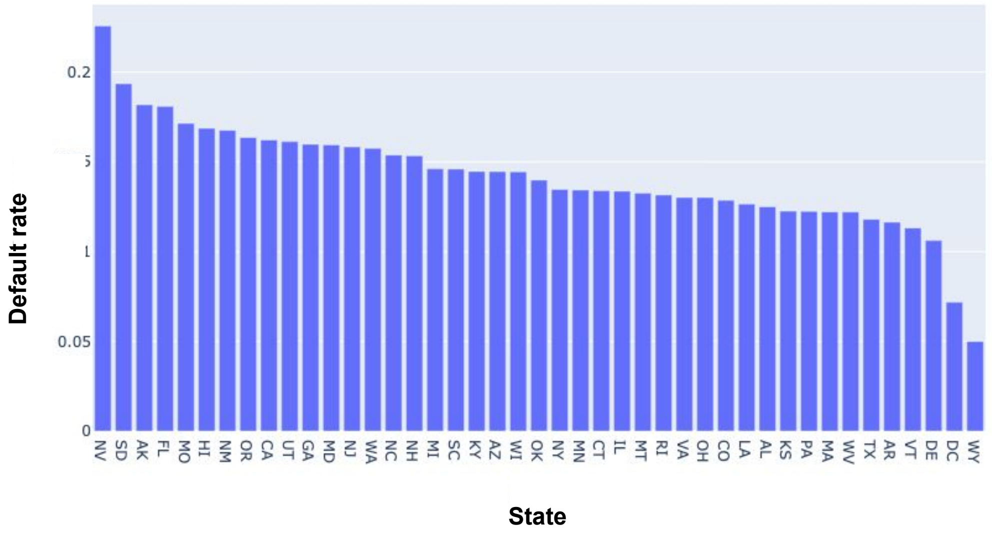
1

Geographic location of the customer is a relatively strong predictor for loan repayment.

Takeway

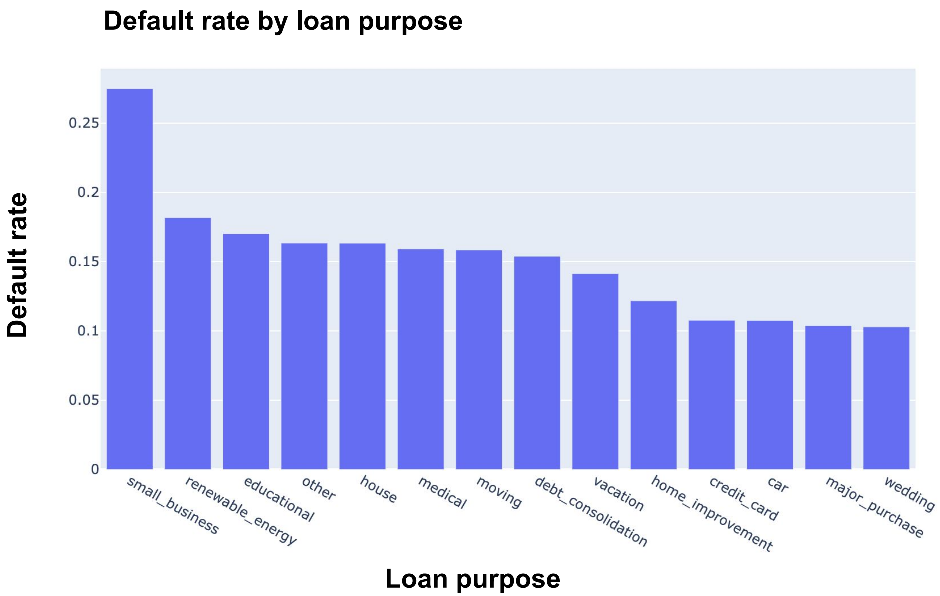
We find that the default rate varies from state to state and spans a range of factor >4.

Default rate by state



Lessons learned

2 Specific loan ‘purposes’ are correlated with loan status.



Takeway

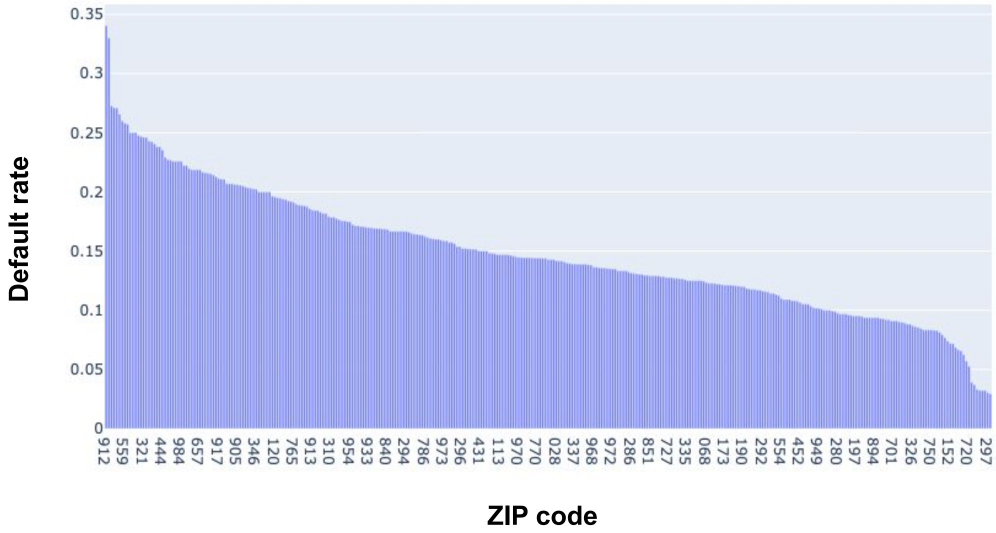
Loan applicants provide a ‘purpose’ when filing for a loan application. We see a significant higher default rate for small businesses.

As we know that interest rates were higher for accounts that ultimately had a status of 'Charged Off', this would present as a feature of interest for lending companies looking to be more cautious with their loan approvals.

Lessons learned

3 ZIP codes show a similar & more granular distribution.

Default rate by ZIP code



Takeway

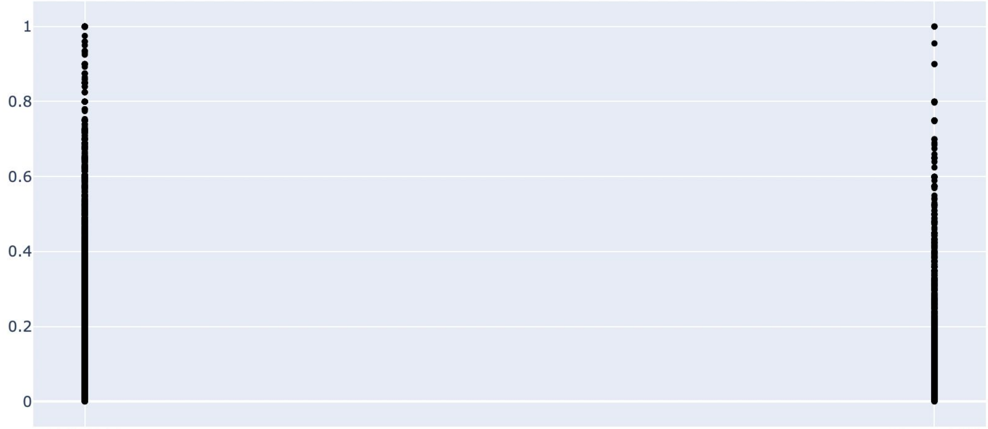
The authors would like to note that it is socially problematic to use factors like the zip code to determine somebody's loan eligibility.

Facilitators like Lending Club should strive to employ better measures.

Lessons learned

4

Derived ‘expertise’ metric that takes into account job length / income highlights the value of those variables on the concentration of charge off accounts.



Fully paid accounts

Charged off accounts

Takeway

Leveraging a derived metric like ‘quality of expertise’ helps uncover patterns in the data that Lending Club can use to more easily extract insight from. In this case, we can see that quality of expertise is distributed more equally across Fully paid accounts than charged off accounts.

The ‘larger’ value of expertise, the lower the concentration of charge off accounts.

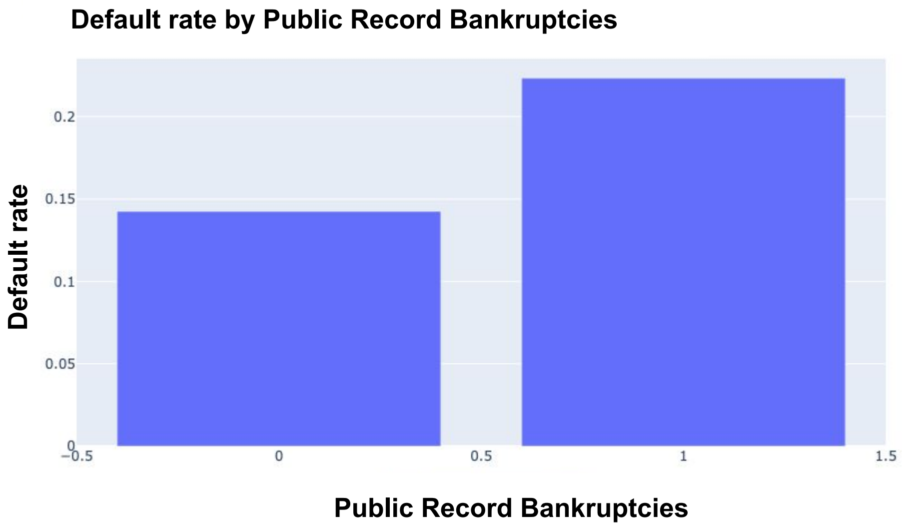
Lessons learned

5 A customer with a history of bankruptcy is 56% more likely to default on their debt (again).

Takeway

The 1634 customers in the database who had a bankruptcy on record they were more likely to default on their debt again. As these are public records we can advise Lending Club to use them as a useful third party data source with valuable insights.

Customers who default on their debt ask for a longer loan term in advance.



Recommendations

Whether high risk loans might are worth investing in depends on the interest rate Lending Club investors receive and Lending Club's own arbitrage.

When assessing the risk of whether or not an account will default or not, we advise Lending Club to take location, employment (income and duration), purpose of the loan previous bankruptcies and the loans term into account. The later has an impact on the overall payed interest.