

## **Summary of Loan Interest Rate Analysis**

### **About the data**

- There are a number of columns that are probably not useful to analysis even at first glance, such as Reason.
- Although Loan Id and Borrower Id are useful from a database standpoint, the values will not contribute much to analysis.
- There is clearly dirty data - dates that are backwards and unclear, dollar and percent signs, null values, etc.

### **The limitations of the data**

- If the interest rate is null, the row will not contribute to the analysis.
- There are too many values in X16 Reason and X18 Loan Title. Both are entered by user freeform and not very clean. Some natural language processing may find something interesting, but the data for those columns currently do not make useful categories.
- X10 Job title or employer is similar to X16 and X18. It also appears to be entered by the user in freeform. This column could be useful if made uniform (for example, title case) and limited to the top counts.
- The data for states ID, IA, NE, and ME is in the single digits. At first glance, it appears that these states have significantly higher or lower interest rates, but if you consider the lack of data, this probably isn't the case.
- For home values, any, other, and none categories were less than 140 records, which isn't a significant number. The other values, mortgage, own, and rent, are clearer in any case.

### **Conclusions and Important Factors**

- There doesn't seem to be a strong correlation between the interest rate and any of the numeric values.
- The strongest correlation appears to be between Number of Payments (36 or 60 months) and interest rate.
- Homeownership and loan categories also seem to have some correlation to interest rate.
- Not verified income also seems to have a lower interest rate, which seems curious.

**Other features that might be helpful to predict interest rates**

- Economic measures for the particular time period (inflation, government, etc.)
- A borrower's credit score
- More demographic data of borrowers.

**Opportunities to consider**

- Adding other factors, such as economic measures and credit scores, may help predict borrower interest rate.
- Adding functionality to the point of record entry to restrict fields, such as job, to certain values.
- Requiring borrowers to give other demographic data at point of record entry.