Step 1: Data Cleanup and Attribute Selection

I chose to explore:

1. Avg Revolving Balance
2. Avg Revolving Utilization
3. Fraction paid/not paid by grade
4. Avg DTI by state
5. Loan amount by state
6. Funded amount
7. Interest rate

1. How do the attributes differ between borrowers who pay back their loans versus those who don't?
    a. **Bonus Question**: If there are any differences between certain attributes, are the attributes statistically significant? A [t-test](#) or [mann-whitney-u](#) test could help answer this question.

Reference: '1 - Differences in paid-not paid lendingclub loans.twbx ' workbook; 'Differences between paid/not paid' dashboard.

The most profound difference I found between those who pay back loans and those who don't is the revolving credit utilization is typically higher for those who don't pay back. Strangely, Montana for some reason has a much higher revolving balance for those who don't pay. Lower grades tend to default more often, but the majority of low grades still tend to pay back. I'd say lending club has a problem with their rating system.

2. How do accepted and rejected loan data differ among different locations?

Reference: '2 - Accepted-rejected lendingclub loans stats by state.twbx' workbook; 'Accepted/rejected stats by state' dashboard.
The debt to income ratio is typically higher for defaulters, and Nebraska had the highest median DTI for accepted loans. Maine had the lowest. Nebraska and Maine had the lowest average loan amounts for accepted loans as well, along with Mississippi.

3. How have issued loans changed over time?

Reference: '3 - Trends in funded lendingclub loans over time.twbx' workbook; '' story.

The total amount funded has dramatically increased. The average funded amount was jumpy at first, but started increasing in 2011. DTI is slightly higher for defaulters, and

increased steadily in 2011. Average revolving balance was higher for defaulters in 2008/2009, but was roughly the same in 2010/2011. Balance utilization has been steady, and always higher for defaulters.

4. What is your additional question that you proposed? What is the answer? How did you come up with this question?

Reference: '4 - Customer and loan properties relationship to avg interest rate.twbx' workbook; 'Avg Int Rate relations' dashboard.

I asked: how is interest rate related to properties of the loan and the loaner? I was wondering if lower grade loans tend to have higher interest rates, and if other properties of the customer correlate to interest rate. I basically found that higher revolving debt utilization correlates very well with loan sub-grade and interest rate. The sub-grade correlates well with interest rate as well. Basically, higher revolving debt utilization correlates to a lower grade, and a higher interest rate. Also, Revolving Balance peaks for the mid-grades (B/C), and the average number of open accounts tends to increase with decreasing grades.