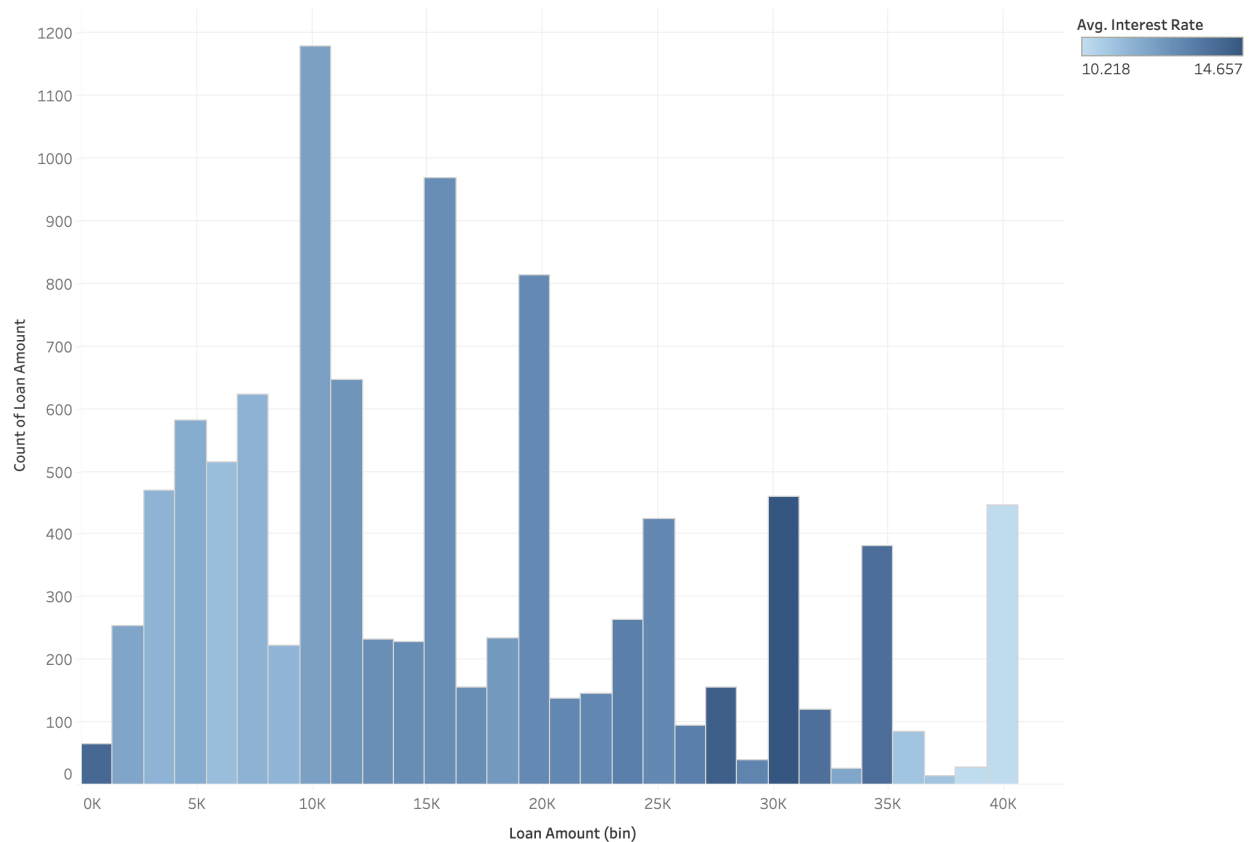


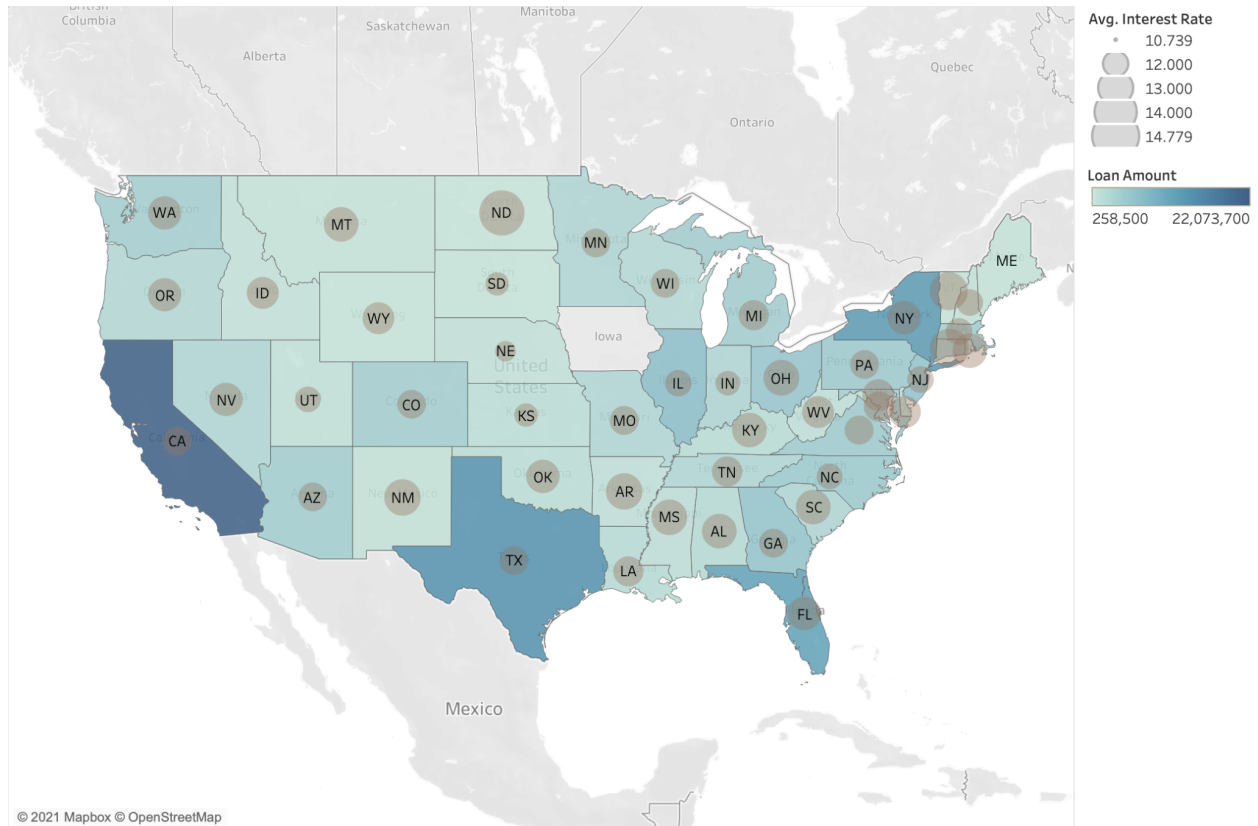
## Case Study #1

- Describe the dataset and any issues with it.
- Generate a minimum of 5 unique visualizations using the data and write a brief description of your observations. Additionally, all attempts should be made to make the visualizations visually appealing
- Create a feature set and create a model which predicts *interest rate* using at least 2 algorithms. Describe any data cleansing that must be performed and analysis when examining the data.
- Visualize the test results and propose enhancements to the model, what would you do if you had more time. Also describe assumptions you made and your approach.

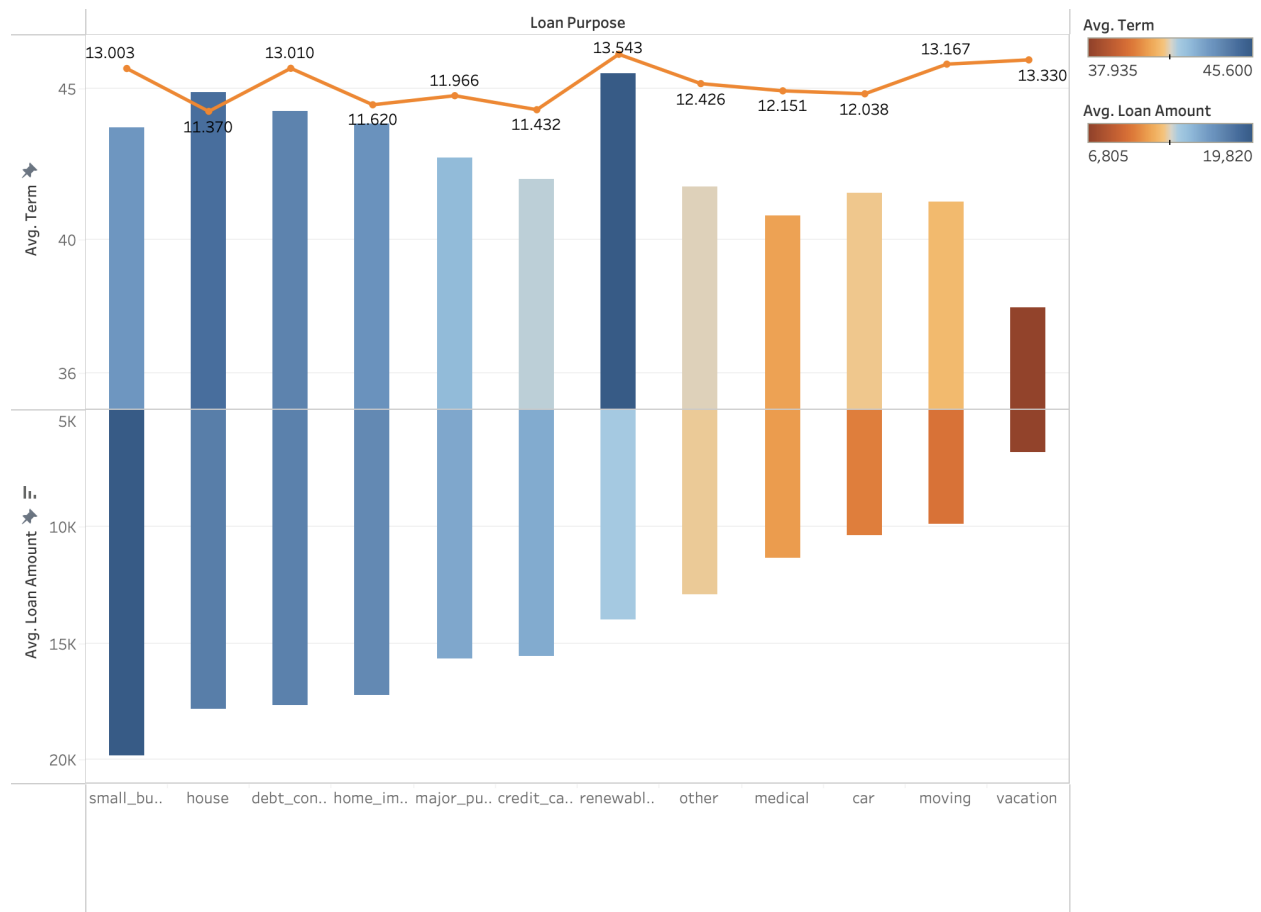
1. We first plot the histogram of the average loan amount. Darker blue means higher average interest rate and lighter means lower interest rate. We notice that on the far left and far right bins, the bin with the smallest amount has a very high interest rate, and the bin with the largest amount has an extremely low interest rate. Other than the two ends, loan amount and interest rate seem to have a positive relationship.



2. California, Texas, New York and Florida are the top 4 states with the largest total amount of loans issued, but they have relatively low levels of interest rate.



- Business and housing loans are larger in amount and are issued with longer terms on average, while more occasional loans for medical, car, moving and vacation purposes are short term and much lower in amount.  
Interest rates are in general lower for loans concerning the well-being of families to ensure better living qualities, like housing, home improvement and credit card, and higher for business and recreational purposes.



4. Bad Debt rate is not very different across individual / joint applications. Individual applicants have a slightly higher bad debt rate.

