Summary:

This projects plots 2 different graphs. It shows amount of Prosper loans per state and focus on CALIFORNIA as it appears as highest amount of loans, moreover focus on specific occupation (Engineering) and finds out which engineering field has the highest rate and total loan.

Design:

Exploratory Data Analyst:

I explore a data set about Prospect loans p2p. My goals with the study are understanding how this practice works and find what seems to affect the borrower rate. I have used R for data cleaning and feature extraction, then used the cleaned dataset for creating my tableau story.

Explanatory data analyst R file is attached.

Data Visualization (Tableau):

Visualization Type:

OpenStreetMap: I have decided to use map as we have data based on the states.

Line-plot: I have used line-plot for average credit score vs (median bankcard utilization, borrower rate and loan amount) as I was interested in showing the trends within time. I have used legend colors to separate specific occupations, as it is easier to locate with colors.

Bar-Plot and histogram have been used to show the distribution and stacked values for customer payments and bank utilization.

Scatter-plot: have been used to show available trends and correlation for number of inquiries and monthly loan payments.

I have used tableau to understand the prosper loan for Engineering fields within CALIFORNIA as it is the state with highest amount of loan. Interesting trends which I have found was for Computer programming and Mechanical engineering fields which has the lowest available bankcard credit and also paying highest median borrower rate however they are having similar characteristics in compare to Chemical/Electrical engineering.

Feedbacks:

I have feedback from two person.

Feedback #1:

Your results was not clear, it seems to me you have not done any data wrangling or cleaning, as an example you have multiple groups for Employment status however you can categories them in better way, you also have too many categories for loan status which is not necessary. This was an example I think you need to clean your data first.

Feedback #2:

Too many colors, I didn't understand the correlation within engineering groups, also seems to me dataset is not cleaned.

Post-feedback Design:

After reading the feedbacks I did exploratory data analysis and got several changes as:

- Categories the Employment group to (employed, self-employed, retired and not employed).
- Create an average credit score using low and high credit score for each user.
- Categories the loan status to (Charged off, Completed, Current, Defaulted and Past Due).
- I have used less color and filter and applied my current filter to all story.

Results:

Tableau Public Link: Pre-feedback Design

Tableau Public Link: Post-feedback Design

-

Resources:

- Data visualization Udacity.
- Exploratory Data Analyst Udacity.

Data:

- Prosper_Loan.csv (attached)