7/1/2020

Prosper Loan Analysis

- Course: Data Visualization
- Student Name: Anupama Kamepalli
- Student ID: 11410504

Submitted by Anupama Kamepalli
University of North Texas

DSCI 5360 – Data Visualization
Final Project Details Form

Instructions

1. Overwrite the fields in the text below with the information specified
2. Field E may be omitted if your data is not sourced from the web
3. Save this file and upload it to the project website
4. The text below should also be included on the second page (first page after the coversheet) of your final project report.

You are a consultant who has been hired by a Prosper to study Financial peer to peer lending. To do so, you have acquired the prosperLoanData.csv dataset from Kaggle located on the web at https://www.kaggle.com/petersabry84/prosperloandataexploration/data .

A. Organization (e.g., a business, government agency, not-for-profit, etc)
B. Topic (e.g., attrition, e-commerce sales, etc.)
C. Dataset (e.g., AB_NYC_2019.csv)
D. Dataset Source (e.g., Kaggle)
E. Dataset Web Address (e.g., https://www.kaggle.com/ptoscano230382/air-bnb-ny-2019).

Dr. Torres is an executive stakeholder at (A) and has asked for your report.

Introduction: Prosper is peer-to-peer lending platform which aims to connect people who need money and those people who have the money to invest. This platform gaining popularity over years in terms of number of loans across different states in US. This dataset is a great source to analyze financial stability of platform and spread of loans in various geographical locations used to analyze the loan accounts by using various parameters such as ratings, prosper score, occupation, defaults etc., to identify patterns and findings. This data source contains 113937 records with total of 81 variables on each loan. I have explored the Prosper dataset to derive the visualizations on various financial business aspects by using Tableau BI Tool.

Data source: https://www.kaggle.com/petersabry84/prosperloandataexploration/data

References: Used information available in various platforms to enhance skills and to achieve the expected results in project.

Analysis and Findings: I have used Tableau tool for data visualizations:

- Area charts
- Line graphs
- Choropleth maps
- Stacked bars
- Bar graph
- Bar graph with Dual axis
- Scatterplot

For various analysis, I have used above representations with table calculations and filters as well.

- Area chart represented in **Appendix-A** dealt with exploring the prosper lending data and important to check how the business model of Prosper works over a period of time Since 2009, the online personal loan business has increased steadily and climbed up quickly Since 2013 and then dropped down at the beginning of 2014. However, the borrowers credit score has went down over time. If we look closely, we can see a major dip in the Prospers' loan opening around 2008 last quarter. This can be attributed to the mortgage crisis erupted in US around 2008. As we can see in the data it almost took 2009 Q4 for Prosper to return to its prior performance. In this visualization it is clear that, prosper could easily wade off risky accounts from its business models. Maybe it is because of the sophisticated rating / score methodologies. Will find these in visualizations mentioned in below appendix.
- Line graph Represented in **Appendix-B** dealt with exploring the prosper data on the basis of Loan Performance and Prosper score. Colors in each line graph represents the

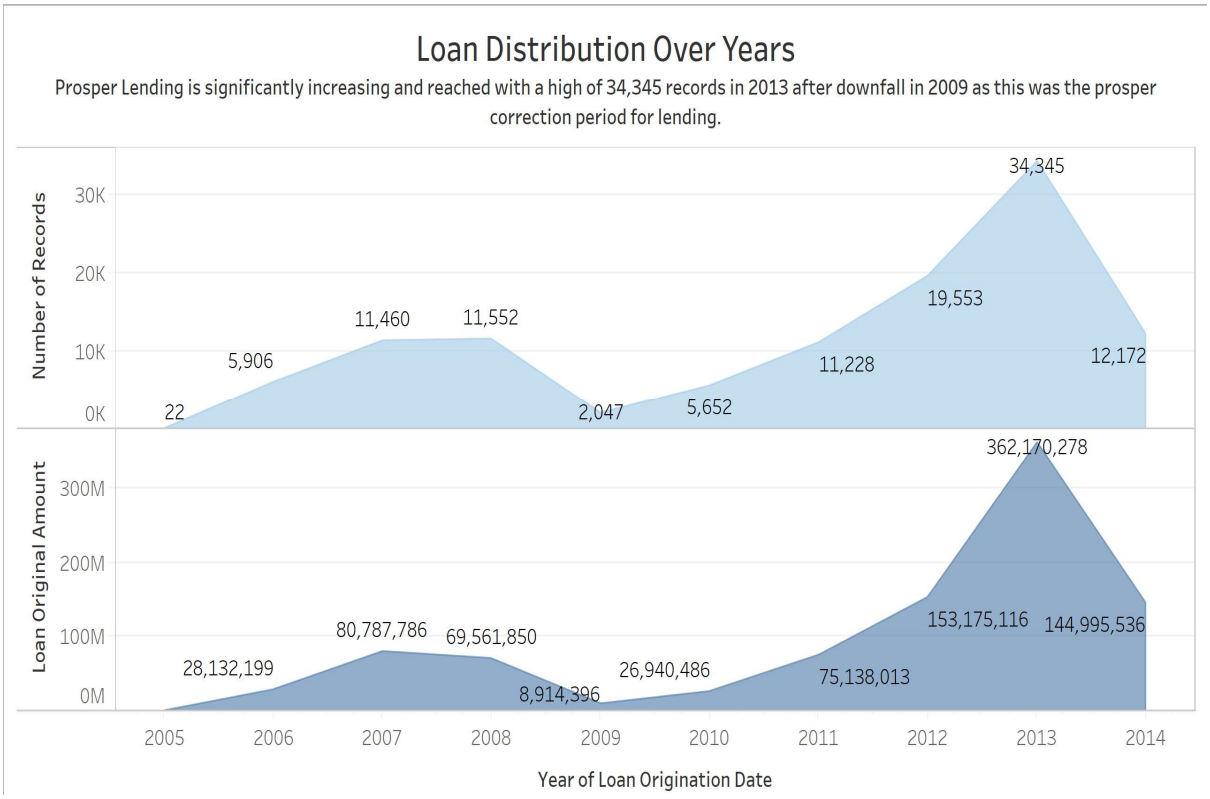
performance of prosper as per each state in US. prosper overall loan performance is drastically increasing from 2013 and there is drop from 2014.

- Map represented in **Appendix-C** dealt with explore the loan default distribution across all the states in US. As per visualization, it is proven that California state has higher default rate among other states.
- Stacked bars represented in **Appendix-D** dealt with explore the 'Loan Default Rate' using prosper rating. As per the visualization, it is proven that the default rate is comparatively low for better rating types like AA than D,E,HR.
- Bar graph represented in **Appendix-E** dealt with explore the 'Loan Default Rate by using 'Income Range'. As it is proven from **the visualization that the income range above $100,000 has lower default rate.** so lower income rages has higher default rate .
- Bar graph represented in **Appendix-F** dealt with verify the expected return and losses across various types of rating. as per the visualization it is proven that the' **Higher return means higher risk'**.
- Bar graph represented in **Appendix-G** dealt with explore the actual loss for types of rating. The rating types C, D have higher principal loss. Hence, as per the visualization analysis , avoid investing money on rating types C, D to avoid risk for default loan.
- Scatter plot represented in **Appendix-H** dealt with explore the Borrow rate on occupation wise. As per the visualization it is proven that lower borrow rate with occupation has lower default rate. An interesting pattern was found among the college student group which shows that higher grade students would have more loans and lower default rate, with exception of sophomore student. students' loans are not a good business, exception for graduate students.
- Bar graph with Dual axis represented in **Appendix-I** dealt with higher grade students will have more loans and lower default rate. With exception of Sophomore students because of variety pattern.
- Map represented in **Appendix-J** dealt with exploring the distribution of Prosper accounts and delinquency across the borrower states. As we can see California is way ahead of the other states .Texas , Florida , New York, Illinois are the other states with significant presence in terms the prosper accounts and delinquency. As per the visualization we say that the state with higher prosper accounts with higher delinquency.
- Bar graph represented in **Appendix-K** identified strong correlation between higher interest rate and difficulty in loan repayment. This also indicates other factors like customers whose Prosper rating / score are less desirable - which indicates risky investments- must put up with higher interest rates may be to offset the risk involved in the transaction

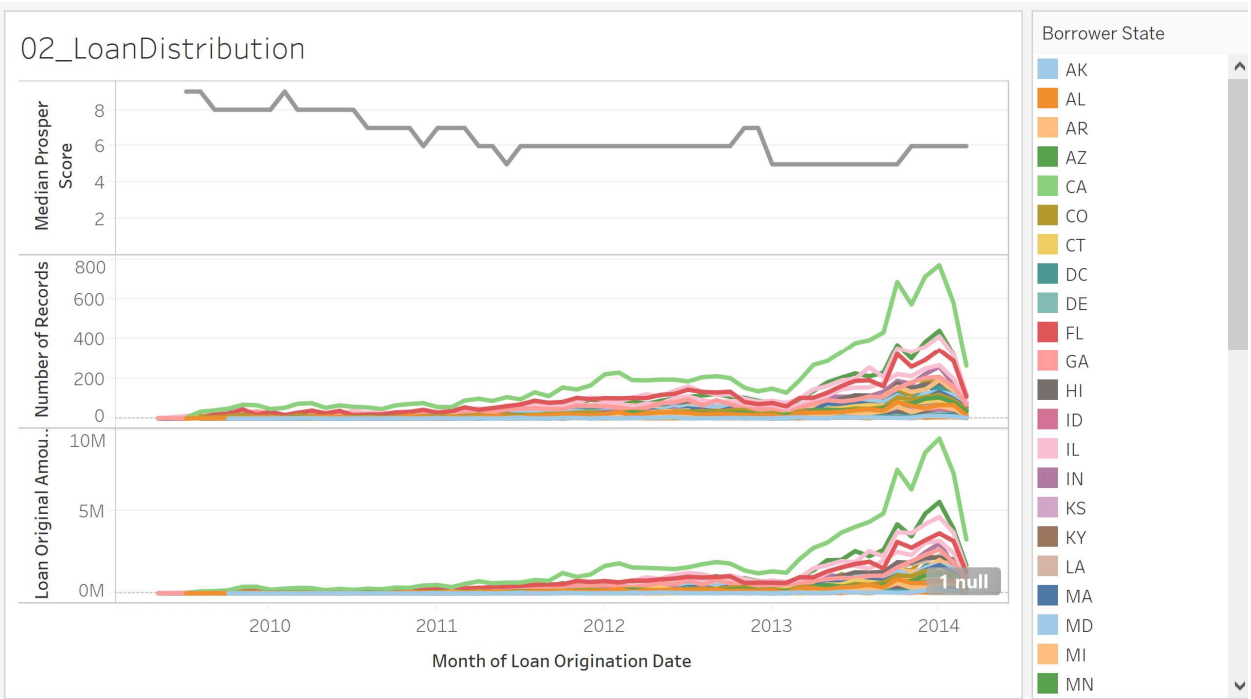
Managerial implications and Recommendations:

- As per the Appendix A, the spread of loans ramped up in 2013, but not withstanding growth in 2014, the decline in number of loans implying the flaw in operational inefficiency of prosper while dealing with new lending's, which results as non-performing loans. This statement be evident on overall analysis of appendix. As per this, prosper need to review the lending strategy to figure out the worthy customers.
- As per the Appendix B, the line graphs representing State wise lending, California is top among the remaining states, we can also see that only 7 states are representing the growth in 2013, most of the states are lying flat in terms of number of loans. Managers should focus on all the states to improve the overall growth of Prosper platform instead of focusing on few states.
- Appendix C, representing state wise loan defaulters, we can observe that the states with more spread of loans which representing 2013 growth are among the top list of defaulters. It is implying that the reckless lending's lead to the more business and increased the risk corner of the Prosper platform. Managers should exercise caution in lending's and device a strategy to resolve the non-performing loans.
- Appendix D is evident that the prosper Ratings are working the right way, The customers with the low rating are turned into non-performing loans. Also Appendix E is proven that the customers with income range below $100000 are turning into non-performing loans. Managers should exercise caution before lending risky loans.
- Appendix F and G are cornering the evidence on interest rate and loan ratings, Managers should avoid lending loans to the 'C and D' rating customers.

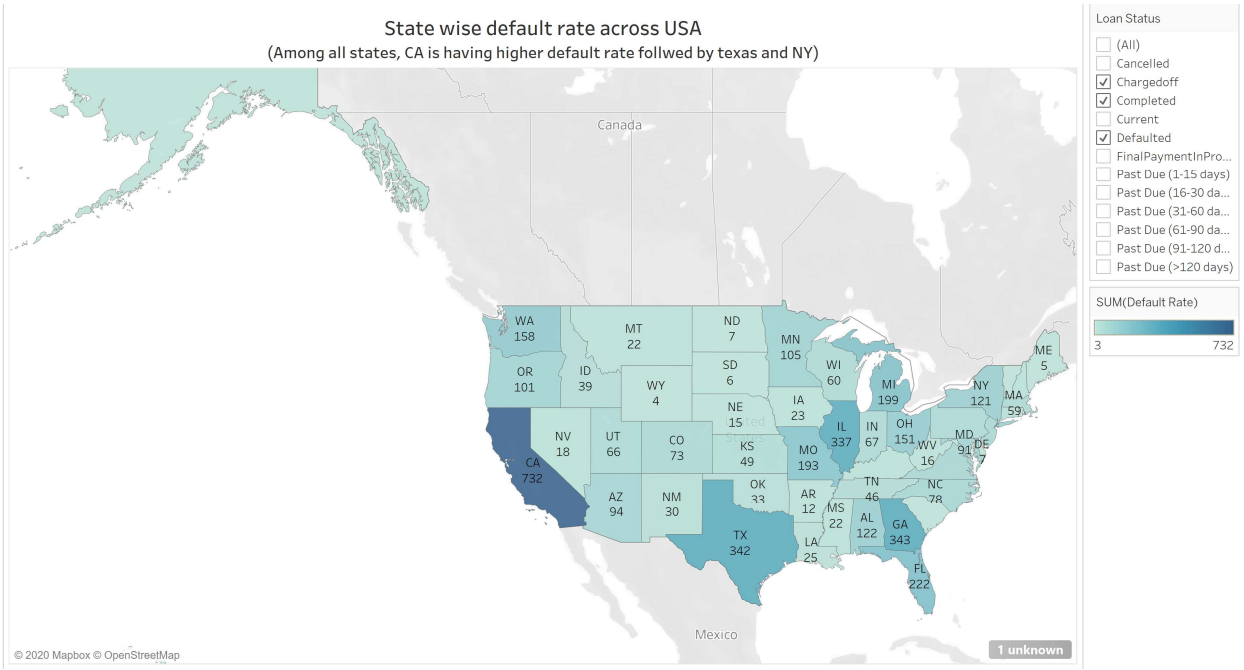
Appendix-A



Appendix-B:

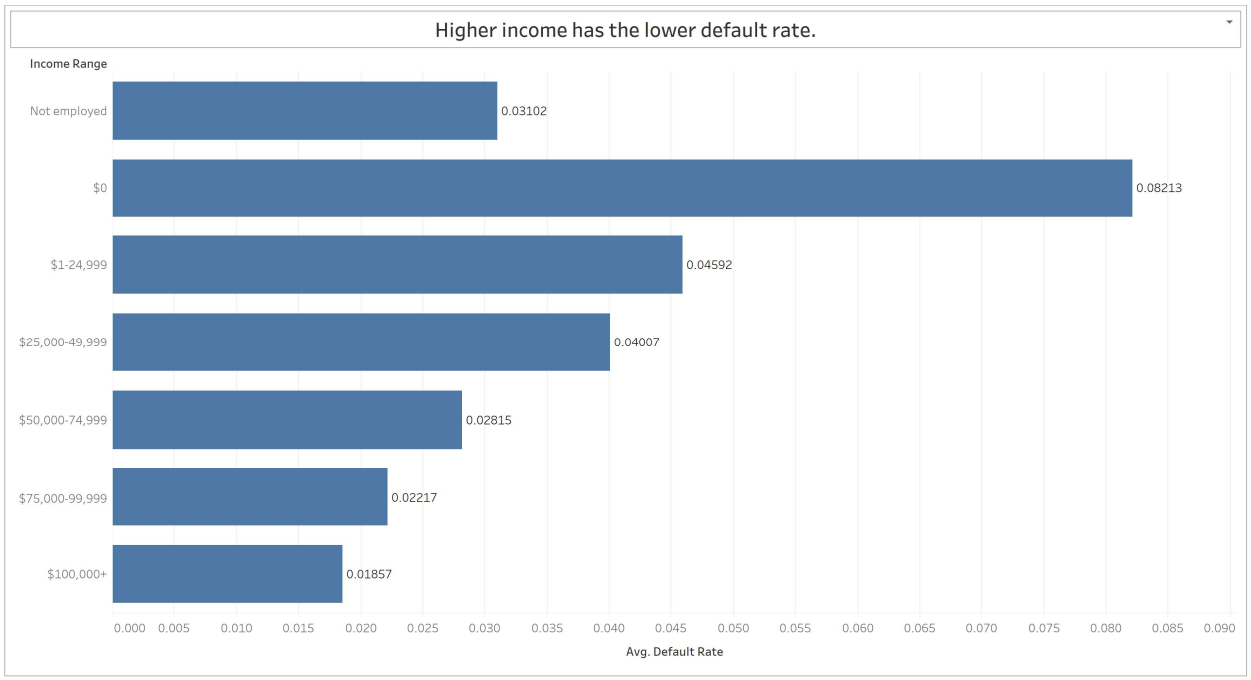


Appendix-C



Appendix-D

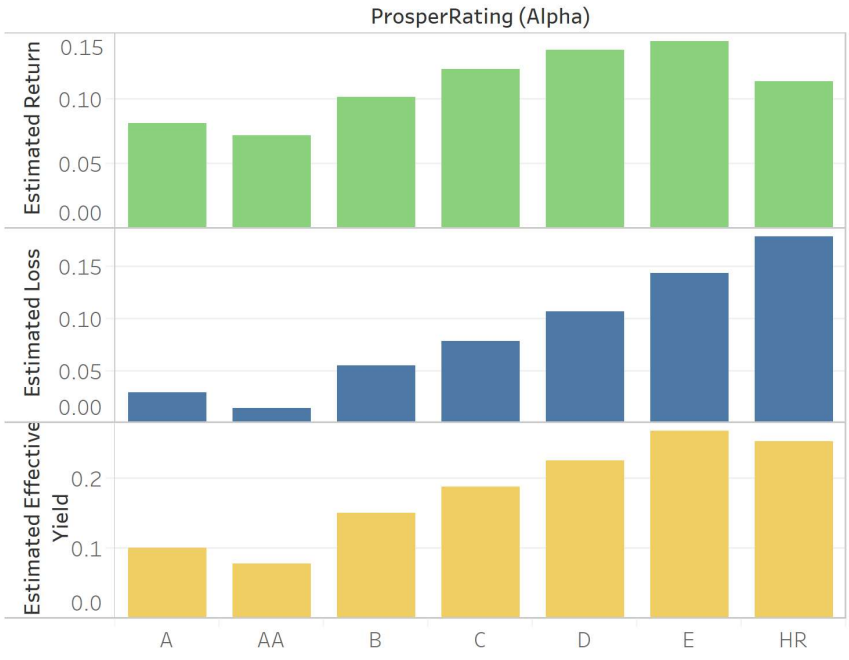
Appendix-E



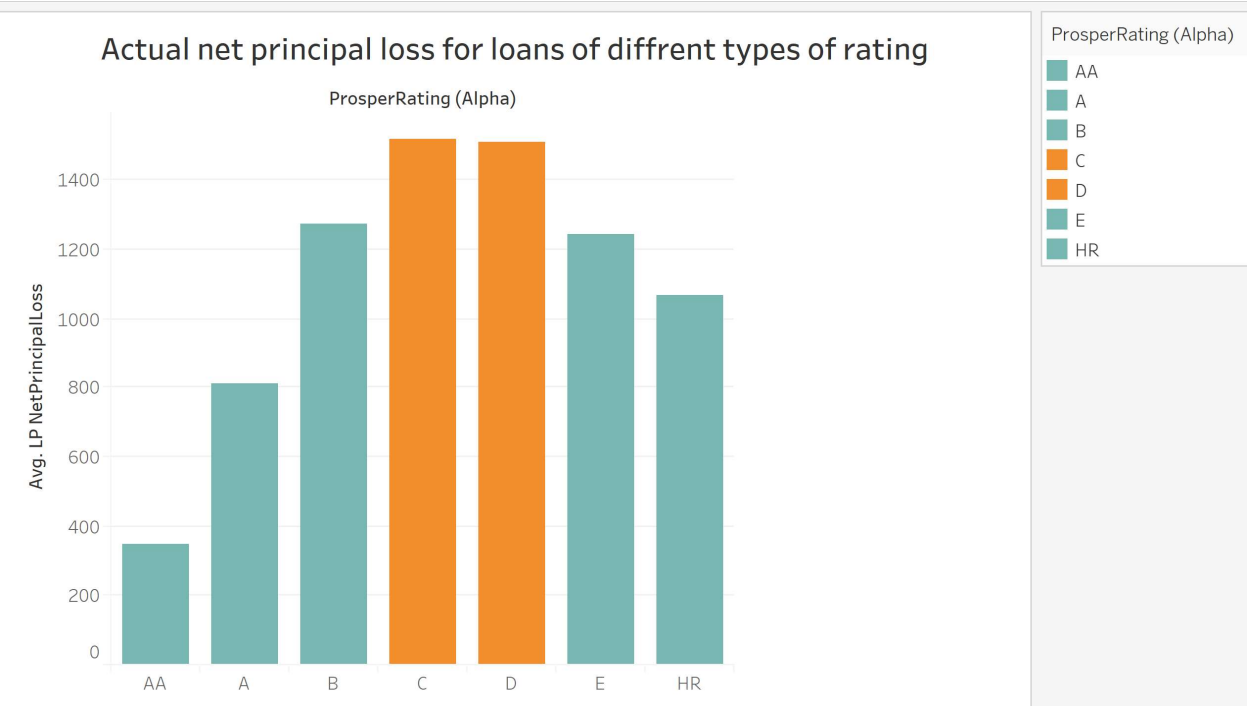
Appendix-F

Expected returns,loss for the loans of various types of ratings

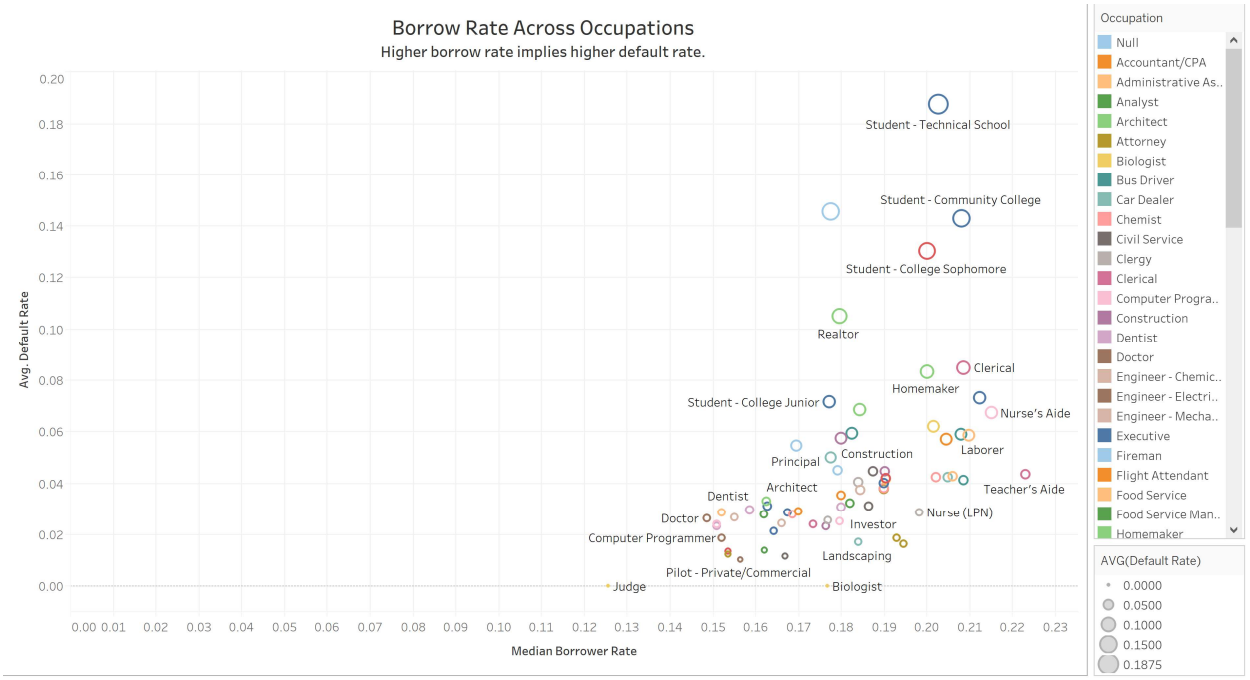
AA- Highly rated,less risky
HR-Least Rated,High risky



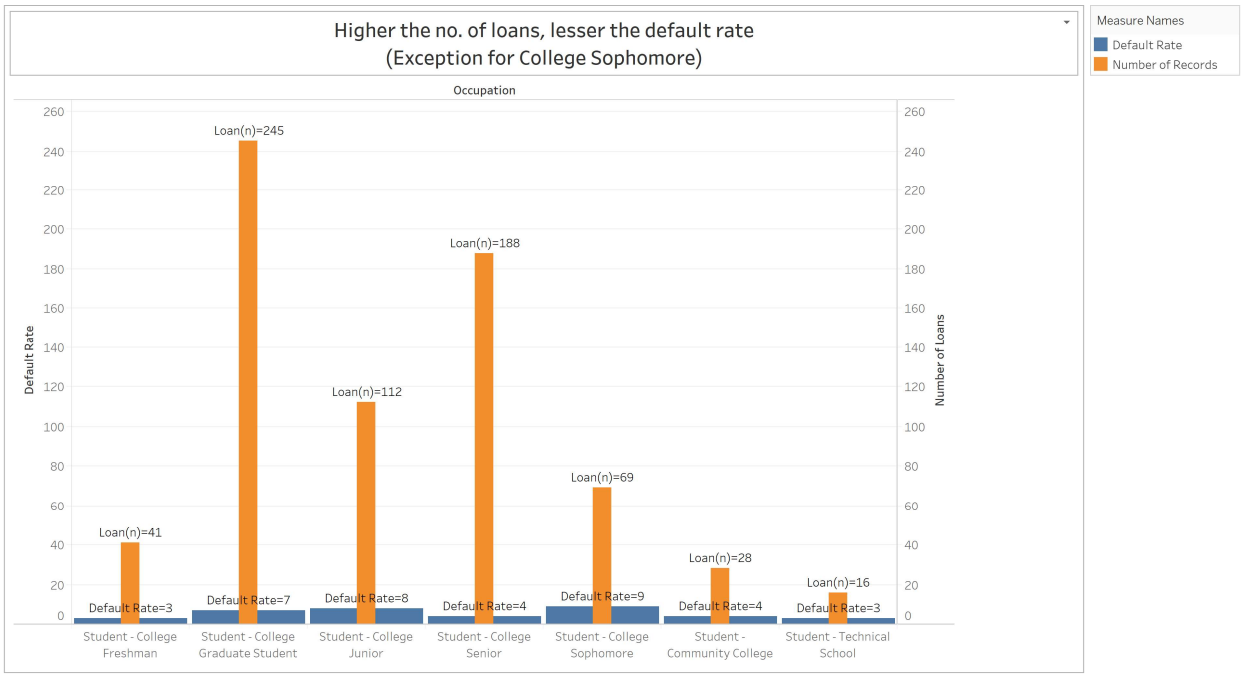
Appendix-G



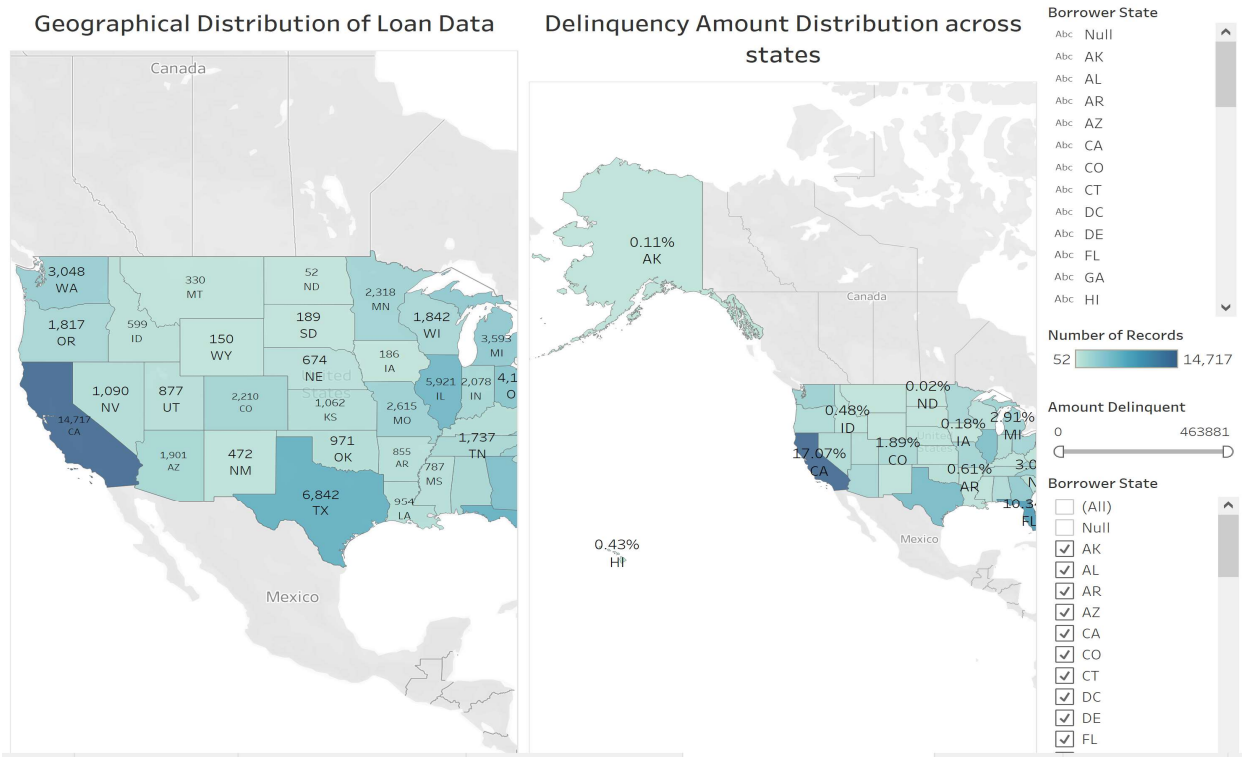
Appendix-H



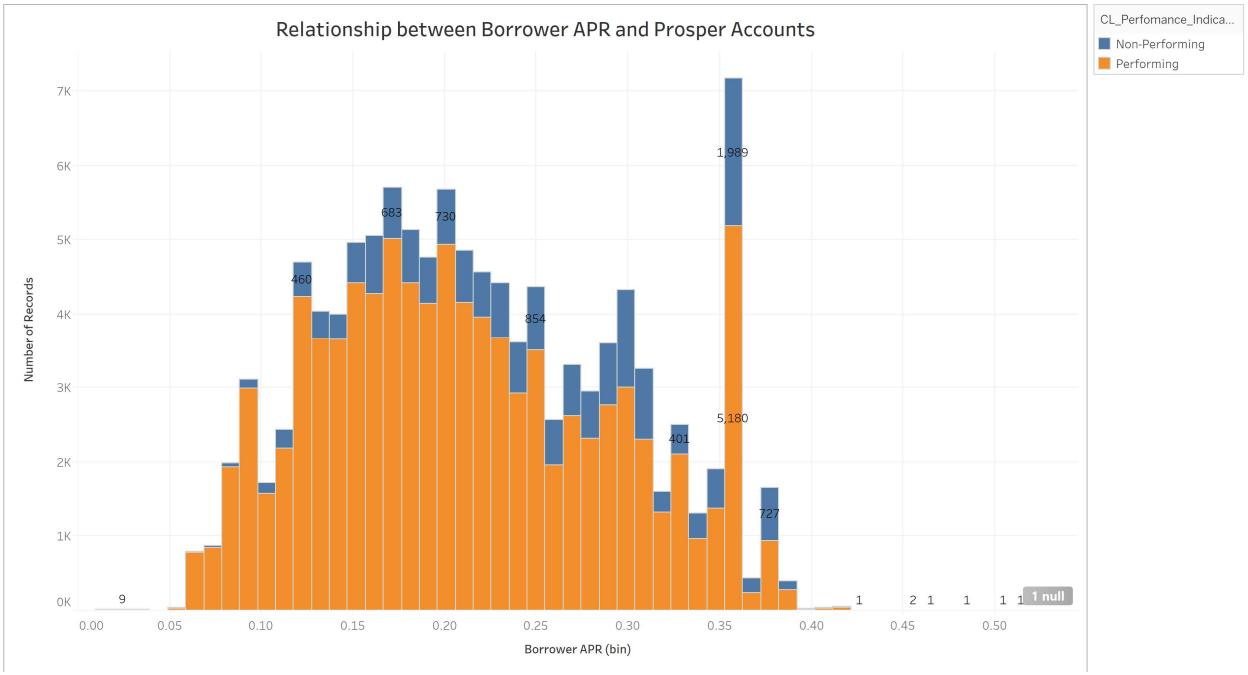
Appendix-I



Appendix-J



Appendix-K



Calculated Field: CL_performance_Indicator

IF [Loan Status]='Chargedoff' or [Loan Status]='Defaulted' or [Loan Status]='Past Due (91-120 days)' or [Loan Status]='Past Due (>120 days)'

THEN 'Non-Performing'

ELSEIF [Loan Status]='Past Due (1-15 days)' or [Loan Status]='Past Due (16-30 days)' or [Loan Status]='Past Due (31-60 days)' or [Loan Status]='Past Due (61-90 days)'

THEN 'Non-Performing'

ELSEIF [Loan Status]='Completed' or [Loan Status]='Current' or [Loan Status]='FinalPaymentInProgress'

THEN 'Performing'

END

Conclusion: In prosper lending has borrowers, lenders and company. I have done analysis on how different stake holders have connected each other, what affects borrowers Prosper Score and who defaults the foremost.

First, I even have done a statistic trend analysis starting from year 2007 – 2014 about the amount of loans taken by borrowers, the quantity of their loans and the way their Prosper Score got

affected during this duration. I noticed that since 2009, the loan business increased and climbed up quickly since 2013 then dropped down at beginning of 2014 while the borrower credit scores constantly dropping over this point and a few states having default rates quite 30%.

I have explored the defaulters, reason for defaults, reason for borrowers to require loan, which occupation requires loans I acknowledged that the people with lower income have highest default rates and most defaulters invest within the loan type 'D'. High income will have lower default rate.

Breaking right down to occupation-wise, a stimulating pattern was found that the school student group which are enrolled in higher grade studies have more loans, higher borrower and default rates. As per the visualization, the higher borrower rate will have higher default rate.

The unique patter found that the sophomore students were the highest defaulters and having lower number of loans are not having income as find the default rate with respect to occupation.

Explored the incomes and losses on different loan ratings the 'HR' loan rating had the very best loss albeit this sort of loan is given to most credit-worthy borrowers. However, I noticed that loans C&D rating types had highest losses than other loans and are most risky. So it is advisable to avoid investing money for rating types C&D as this considered 'Actual Loss'.

Analyzing the relationship between the borrower APR and loan accounts. there is a positive correlation between loan performance and APR. Ratio of nonperformance loans are very less compared to the total number of loans if you across the distribution plotted on the higher APR the ratio of nonperforming loans varies from 1.76% to 46% . This suggests a strong correlation between higher interest rate and difficulty in loan repayment. This also indicates other factors like customers whose Prosper rating / score are less desirable which indicates risky investments has to put up with higher interest rates may be to offset the risk involved in the transaction.