Tableau-Project for Prosper Loan Data Set

Introduction:

In this project, I will use Tableau platform to explore some aspects from the Prosper Loan Data set. I will apply the knowledge gained from Udacity Tableau Course that included the best practices for visualizing data using appropriate colors, design and shape.

What is Prosper?

Prosper is the first peer-to-peer lending marketplace in the US, San Francisco. It was founded on 2005. Till now, prosper has helped more than $770,000 people with more than $12 billion loans. Prosper facilitates the borrowing and investing processes for both individuals and institutions. Borrowers can apply for a fixed-rate and fixed-term loan in the range of $20,000 and $40,000. Prosper Co. offers Institutions and individuals attractive returns when they invest in loans.

Prosper Loan Data Set:

This data set contains 113,937 loans with 81 variables on each loan, including loan amount, borrower rate (or interest rate), current loan status, borrower income, and many others.

Prosper Loan Data Set Definition:

Variable	Description
1. ListingKey	Unique key for each listing, same value as the 'key' used in the listing object in the API.
2. ListingNumber	The number that uniquely identifies the listing to the public as displayed on the website.
3. ListingCreationDate	The date the listing was created.
4. CreditGrade	The Credit rating that was assigned at the time the listing went live. Applicable for listings pre-2009 period and will only be populated for those listings.
5. Term	The length of the loan expressed in months.
6. LoanStatus	The current status of the loan: Cancelled, Chargedoff, Completed, Current, Defaulted, FinalPaymentInProgress, PastDue. The PastDue status will be accompanied by a delinquency bucket.
7. ClosedDate	Closed date is applicable for Cancelled, Completed, Chargedoff and Defaulted loan statuses.
8. BorrowerAPR	The Borrower's Annual Percentage Rate (APR) for the loan.
9. BorrowerRate	The Borrower's interest rate for this loan.
10. LenderYield	The Lender yield on the loan. Lender yield is equal to the interest rate on the loan less the servicing fee.
11. EstimatedEffectiveYield	Effective yield is equal to the borrower interest rate (i) minus the servicing fee rate, (ii) minus estimated uncollected interest on charge-offs, (iii) plus estimated collected late fees. Applicable for loans originated after July 2009.
12. EstimatedLoss	Estimated loss is the estimated principal loss on charge-offs. Applicable for loans originated after July 2009.
13. EstimatedReturn	The estimated return assigned to the listing at the time it was created. Estimated return is the difference between the Estimated Effective Yield and the Estimated Loss Rate. Applicable for loans originated after July 2009.
14. ProsperRating (numeric)	The Prosper Rating assigned at the time the listing was created: 0 - N/A, 1 - HR, 2 - E, 3 - D, 4 - C, 5 - B, 6 - A, 7 - AA. Applicable for loans originated after July 2009.
15. ProsperRating (Alpha)	The Prosper Rating assigned at the time the listing was created between AA - HR. Applicable for loans originated after July 2009.
16. ProsperScore	A custom risk score built using historical Prosper data. The score ranges from 1-10, with 10 being the best, or lowest risk score. Applicable for loans originated after July 2009.
17. ListingCategory	The category of the listing that the borrower selected when posting their listing: 0 - Not Available, 1 - Debt Consolidation, 2 - Home Improvement, 3 - Business, 4 - Personal Loan, 5 - Student Use, 6 - Auto, 7- Other, 8 - Baby&Adoption, 9 - Boat, 10 - Cosmetic Procedure, 11 - Engagement Ring, 12 - Green Loans, 13 - Household Expenses, 14 - Large Purchases, 15 - Medical/Dental, 16 - Motorcycle, 17 - RV, 18 - Taxes, 19 - Vacation, 20 - Wedding Loans
18. BorrowerState	The two letter abbreviation of the state of the address of the borrower at the time the Listing was created.
19. Occupation	The Occupation selected by the Borrower at the time they created the listing.

20. EmploymentStatus	The employment status of the borrower at the time they posted the listing.
21. EmploymentStatusDuration	The length in months of the employment status at the time the listing was created.
22. IsBorrowerHomeowner	A Borrower will be classified as a homowner if they have a mortgage on their credit profile or provide documentation confirming they are a homeowner.
23. CurrentlyInGroup	Specifies whether or not the Borrower was in a group at the time the listing was created.
24. GroupKey	The Key of the group in which the Borrower is a member of. Value will be null if the borrower does not have a group affiliation.
25. DateCreditPulled	The date the credit profile was pulled.
26. CreditScoreRangeLower	The lower value representing the range of the borrower's credit score as provided by a consumer credit rating agency.
27. CreditScoreRangeUpper	The upper value representing the range of the borrower's credit score as provided by a consumer credit rating agency.
28. FirstRecordedCreditLine	The date the first credit line was opened.
29. CurrentCreditLines	Number of current credit lines at the time the credit profile was pulled.
30. OpenCreditLines	Number of open credit lines at the time the credit profile was pulled.
31. TotalCreditLinespast7years	Number of credit lines in the past seven years at the time the credit profile was pulled.
32. OpenRevolvingAccounts	Number of open revolving accounts at the time the credit profile was pulled.
33. OpenRevolvingMonthlyPayment	Monthly payment on revolving accounts at the time the credit profile was pulled.
34. InquiriesLast6Months	Number of inquiries in the past six months at the time the credit profile was pulled.
35. TotalInquiries	Total number of inquiries at the time the credit profile was pulled.
36. CurrentDelinquencies	Number of accounts delinquent at the time the credit profile was pulled.
37. AmountDelinquent	Dollars delinquent at the time the credit profile was pulled.
38. DelinquenciesLast7Years	Number of delinquencies in the past 7 years at the time the credit profile was pulled.
39. PublicRecordsLast10Years	Number of public records in the past 10 years at the time the credit profile was pulled.
40. PublicRecordsLast12Months	Number of public records in the past 12 months at the time the credit profile was pulled.
41. RevolvingCreditBalance	Dollars of revolving credit at the time the credit profile was pulled.
42. BankcardUtilization	The percentage of available revolving credit that is utilized at the time the credit profile was pulled.
43. AvailableBankcardCredit	The total available credit via bank card at the time the credit profile was pulled.
44. TotalTrades	Number of trade lines ever opened at the time the credit profile was pulled.
45. TradesNeverDelinquent	Number of trades that have never been delinquent at the time the credit profile was pulled.
46. TradesOpenedLast6Months	Number of trades opened in the last 6 months at the time the credit profile was pulled.
47. DebtToIncomeRatio	The debt to income ratio of the borrower at the time the credit profile was pulled. This value is Null if the debt to income ratio is

	not available. This value is capped at 10.01 (any debt to income ratio larger than 1000% will be returned as 1001%).
48. IncomeRange	The income range of the borrower at the time the listing was created.
49. IncomeVerifiable	The borrower indicated they have the required documentation to support their income.
50. StatedMonthlyIncome	The monthly income the borrower stated at the time the listing was created.
51. LoanKey	Unique key for each loan. This is the same key that is used in the API.
52. TotalProsperLoans	Number of Prosper loans the borrower at the time they created this listing. This value will be null if the borrower had no prior loans.
53. TotalProsperPaymentsBilled	Number of on time payments the borrower made on Prosper loans at the time they created this listing. This value will be null if the borrower had no prior loans.
54. OnTimeProsperPayments	Number of on time payments the borrower had made on Prosper loans at the time they created this listing. This value will be null if the borrower has no prior loans.
55. ProsperPaymentsLessThanOneMonthLate	Number of payments the borrower made on Prosper loans that were less than one month late at the time they created this listing. This value will be null if the borrower had no prior loans.
56. ProsperPaymentsOneMonthPlusLate	Number of payments the borrower made on Prosper loans that were greater than one month late at the time they created this listing. This value will be null if the borrower had no prior loans.
57. ProsperPrincipalBorrowed	Total principal borrowed on Prosper loans at the time the listing was created. This value will be null if the borrower had no prior loans.
58. ProsperPrincipalOutstanding	Principal outstanding on Prosper loans at the time the listing was created. This value will be null if the borrower had no prior loans.
59. ScorexChangeAtTimeOfListing	Borrower's credit score change at the time the credit profile was pulled. This will be the change relative to the borrower's last Prosper loan. This value will be null if the borrower had no prior loans.
60. LoanCurrentDaysDelinquent	The number of days delinquent.
61. LoanFirstDefaultedCycleNumber	The cycle the loan was charged off. If the loan has not charged off the value will be null.
62. LoanMonthsSinceOrigination	Number of months since the loan originated.
63. LoanNumber	Unique numeric value associated with the loan.
64. LoanOriginalAmount	The origination amount of the loan.
65. LoanOriginationDate	The date the loan was originated.
66. LoanOriginationQuarter	The quarter in which the loan was originated.
67. MemberKey	The unique key that is associated with the borrower. This is the same identifier that is used in the API member object.
68. MonthlyLoanPayment	The scheduled monthly loan payment.
69. LP_CustomerPayments	Pre charge-off cumulative gross payments made by the borrower on the loan. If the loan has charged off, this value will exclude any recoveries.
70. LP_CustomerPrincipalPayments	Pre charge-off cumulative principal payments made by the borrower on the loan. If the loan has charged off, this value will exclude any recoveries.
71. LP_InterestandFees	Pre charge-off cumulative interest and fees paid by the borrower. If the loan has charged off, this value will exclude any recoveries.

72. LP_ServiceFees	Cumulative service fees paid by the investors who have invested in the loan.
73. LP_CollectionFees	Cumulative collection fees paid by the investors who have invested in the loan.
74. LP_GrossPrincipalLoss	The gross charged off amount of the loan.
75. LP_NetPrincipalLoss	The principal that remains uncollected after any recoveries.
76. LP_NonPrincipalRecoverypayments	The interest and fee component of any recovery payments. The current payment policy applies payments in the following order: Fees, interest, principal.
77. PercentFunded	Percent the listing was funded.
78. Recommendations	Number of recommendations the borrower had at the time the listing was created.
79. InvestmentFromFriendsCount	Number of friends that made an investment in the loan.
80. InvestmentFromFriendsAmount	Dollar amount of investments that were made by friends.
81. Investors	The number of investors that funded the loan.

Summary:

Using Tableau, I was trying to answer two main questions about the prosper loan data set.

The first question was what would affect the payment of the loan? In the first chart, I observed completed and defaulted loans over the years between 2005-2014. I noticed that the percentage of defaulted loans is decreasing after the year of 2008. In the second chart, I displayed the percentage of completed loans and defaulted loans over the US states plotted all in a map. Next, I plotted the different income ranges with respect to completed and defaulted loans. I observed that with higher income range, the less likely that the loan will be defaulted. Then I studied the loan different purposes with respect to the completed and defaulted loans. Loans under Baby and Adoption, Boat, and Engagement Ring categories don't have defaulted rate, and Personal loans have the highest defaulted rate.

The second question was what would affect the estimated return for an investor? In the first chart, I studied the relationship between the credit score and the estimated return. I found that loans with low credit score have less estimated return on loans than loans with high credit score. Then I was interested to explore the relationship between prosper rating and the

estimated return. I found also that with riskier prosper ratings, the higher the estimated return is. Finally, I explored the relationship between the loan amount and the estimated return on it. It appeared from the graph that there is no apparent or significance relationship between them.

Link to the Project First Version:

https://public.tableau.com/profile/reem.alashhab#!/vizhome/ProsperLoanTableauProjectUdacityVersion1/Story1

Design Choices and Explanation:

- **First Version:**

1. **Colors:**
   To be sensitive to those with colorblindness, I used blue and orange consistently over the project. This approach also supported the graphs to be simple and consistent in a way that don't distract viewers and let them remember these two colors across the project. I also used these two colors to distinguish the completed and defaulted loans as they are the main categorial variables in this project. For other categorical variables such as the credit score range, I used extra colors from the automatic color palette that are comfortable to the normal and sensitive eyes.

2. **Labels:**
   Across the project, I added needed labels and titles to make it easier for the viewers to interpret data and focus on the overall conclusion.

3. **Size:**
   I tried here to make visuals and symbols appropriate to the overall visualizations. For example, in the map graph I made the pie chart fit with states sizes to not cover the states or make the map look overloaded.

4. **Chart Types:**

   In this project, I used different types of graphical charts to display data. I used the map to display information related to the US states. The map helps the viewer to find data easily by the location of the state for example. I also used the bar chart for continuance data like years combined with one categorial data such as the loan completion status. In addition, I used it for a discrete variable like income range to show the loan completion status for each range. This chart helped to show the difference between each category separately in an easy way for the eye to capture. In addition, I used it to show the loan status for each loan purpose and the average estimated return per prosper rating. I preferred here to use the circles to make the chart less condense and overloaded because with long bars and many values the graph will look so overloaded and hard to capture data from. Moreover, I used the boxplot that helped to display the different information about the estimated return in relative to credit score range such as the min, max, and mean. For the last chart, I used the scatter plot to study the relationship between the estimated return and the loan amount. This chart is best to use when studying the relationship between two numerical values.

Project Feedback:

For the first worksheet, it is better if you changed the states filter to be a drop-down list to prevent null choices, and that for all worksheets using this filter also.

For the second worksheet, it is not a good thing to chart the percentage of the income, you should chart the value to get a clear visualization that show the difference per each bar.

For the third chart, what if you changed the axes and circle size as a mark.

For the Fifth sheet, I see that it is better if you remove the filter and the color maybe better if it was discrete.

For the Sixth sheet, I think you should take the average for the estimated return and then draw a chart between the loan and on the x-axis and the estimated return on the y-axis and the take line type chart.

(Feedback by one of my colleagues from the Data Analysis Track from the slack group).

Link to the Project Second Version:

https://public.tableau.com/profile/reem.alashhab#!/vizhome/ProsperLoanTableauProjectUdacityVersion2/Story2?publish=yes

Design Choices and Explanation:

- **Second Version:**

1. **Filter Choice:**
   I kept the filter for the borrower state and the loan purposes as a drop-down list with multi options because there is no option that prevent users from un clicking all values. I need the drop-down list with multi options to help the user choose several options and compare between them easily. However, I will add an explanation label, so the user can understand that the map can appear again when selecting values from the list.

2. **Chart of Income Ranges Displayed in Percentage:**
   I modified the percentage under each income range to the number of records to distinguish between completed and defaulted loans. This choice is perceived as a fair choice as per the viewer feedback as the percentage can deceive the final conclusion of the graph.

3. **Income Range x - axis values Arrangement:**
   While reviewing the income range graph, I noticed that the 'not employed' value is positioned after the income ranges. So, I rearranged it to become the first value. This will help the viewers to view data organized with fair comparison.

4. **Chart Direction for Loan Purposes Vs Loan Completion Graph:**
   I swapped the chart direction here to make it easier for the viewer to view the data clearly and interpret results quickly. I also increased the circle size to make it more readable and easier to interpret by the viewer. These two approached were recommended by the viewer.

5. **The Filter and the Color in the Average Estimated Return Per Prosper Rating Graph:**
   I removed the filter as per the viewer feedback because it appears as an extra thing to the overall visualization. For the color, I preferred to keep it as displaying different colors will not help to show that with bad prosper rating, the higher the estimated return on the loan will be. The light and dark color range is best to show the difference here.

6. **The Average Loan and chart type in the Estimated Return Vs Loan Original Amount Vs Loan Completion Graph:**
   I changed the value for the estimated return to the average as per the viewer feedback. This approach helped me to make the graph smoother and less condense to view and interpret. However, I preferred to keep it as a scatter plot as the line chart will make the data look very condense and hard to view and read.

References:

https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_tablecalculations_custom.htm

https://onlinehelp.tableau.com/current/pro/desktop/en-us/maps_howto_filledpiechart.htm

https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_aggregation.htm

https://community.tableau.com/thread/207654

https://community.tableau.com/thread/120950

https://www.youtube.com/watch?v=_NzllWQDuBY

https://www.credera.com/blog/business-intelligence/tableau-workaround-part-2-create-an-include-exclude-filter-on-the-same-field/

https://www.50states.com/abbreviations.htm

https://community.tableau.com/thread/154495

https://onlinehelp.tableau.com/current/pro/desktop/en-us/publish_workbooks_tableaupublic.htm