Which Variables are likely correlated?

Choose : You should explore between 10-15 variables in your analysis.

**Step Four - Document your Analysis**

You will want to document your exploration and analysis in an **RMD file** which you will submit. That file should be formatted in markdown and should contain (in order):

1. **A stream-of-consciousness analysis and exploration of the data.**

a. Headings and text should organize your thoughts and reflect your analysis as you explored the data.

b. Plots in this analysis do not need to be polished with labels, units, and titles; these plots are exploratory (quick and dirty). They should, however, be of the appropriate type and effectively convey the information you glean from them.

c. You can iterate on a plot in the same R chunk, but you don’t need to show every plot iteration in your analysis.

1. **A section at the end called “Final Plots and Summary”**

**You will select three plots from your analysis to polish and share in this section. The three plots should show different trends and should be polished with appropriate labels, units, and titles** (see the [**Project Rubric**](https://review.udacity.com/#!/projects/3165188753/rubric) for more information).

1. **A final section called “Reflection”**

This should contain a few sentences about your struggles, successes, and ideas for future exploration on the data set (see the [**Project Rubric**](https://review.udacity.com/#!/projects/3165188753/rubric) for more information).

Project Rubric

**The project contains at least 20 visualizations.** The visualizations are varied and show multiple comparisons and trends.

Relevant statistics (e.g. mean, median, confidence intervals, correlations) are computed throughout the analysis when an inference is made about the data.

**The project includes a Final Plots and Summary section containing three plots and commentary**. All plots in this section reflect what has been explored in the main body of the analysis.

Plot1 - These plots show that the BorrowerAPR depends on the CreditScore/CreditGrade. In the first plot, we plot the BorrowerAPR vs the CreditGrade for the year 2007, 2014. One can see that the APR's were lowest in the year 2006 and continued to increase until

* Draw comparisons.
* Identify trends.
* Engage a wide audience.
* Explain a complicated finding.
* Clarify a gap between perception and reality.
* Enable the reader to digest large amounts of information.

**Are the plots polished?**

* are axes labeled?
* are units labeled on each axis?
* are plots titled?
* are all labels, titles, and units readable?

|  |  |
| --- | --- |
| **Has a Reflection section been included in the project?** | The project includes a Reflection section discussing the analysis performed. |
| **Does the section provide a written reflection of the analysis?** Consider the following in your reflections:   * Where did I run into difficulties in the analysis? * Where did I find successes? * How could the analysis be enriched in future work (e.g. additional data and analyses)? | The section reflects on how the analysis was conducted and reports on the struggles and successes throughout the analysis. The section provides at least one idea or question for future work. The section explains any important decisions in the analysis and how those decisions affected the analysis. |

**Step Five - Knit your RMD file**

Your knitted RMD file should not be one long chunk of R code. It should contain text and plots interspersed throughout. The goal is to give the person reading the file insight into what you were thinking as you explored your data.

**Step Six - Document your Data (if you chose your own data set)**

The data set you submit (only if you chose your own) should include a text file, like those in the R documentation (e.g. ?diamonds) that describes the source of your data and an explanation of the variables in the data set (definition of any variables, units, levels of categorical variables, and the data generating process, such as how data was collected if possible).

#Understanding The Variables:

The variables below correspond to the Loan Payments. There is also the amount of fee which is uncollected if the loan is charged off (I understand that this means closed)

|  |
| --- |
| LP\_CustomerPayments |
| LP\_CustomerPrincipalPayments |
| LP\_InterestandFees |
| LP\_ServiceFees |
| LP\_CollectionFees |
| LP\_GrossPrincipalLoss |
|  |

Monthly Loan Payment, Term of the loan, BorrowerRate (Interest rate of the borrower), BorrowerAPR, BorrowerState, CreditGrade, IncomeRanger, IncomeVerifiable,IsBorrowerHomeowner, "LoanOriginalAmount","OpenRevolvingAccounts" "OpenRevolvingMonthlyPayment", “ListingCategory”

“Listing Category” : The reason for the loan

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

The following Variables are associated with a loan that was Charged off:

|  |  |
| --- | --- |
| 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. |
| 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. |
| 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. |
| LP\_ServiceFees | Cumulative service fees paid by the investors who have invested in the loan. |
| LP\_CollectionFees | Cumulative collection fees paid by the investors who have invested in the loan. |
| LP\_GrossPrincipalLoss | The gross charged off amount of the loan. (Amount that was charged off) |
| LP\_NetPrincipalLoss | The principal that remains uncollected after any recoveries. (Principal that remains Uncollected) |
| LP\_NonPrincipalRecoverypayments | The interest and fee component of any recovery payments. The current payment policy applies payments in the following order: Fees, interest, principal. |
| EstimatedLoss | Estimated loss is the estimated principal loss on charge-offs. Applicable for loans originated after July 2009.  (Estimated Loss of the Principal) |

* 1. **Investigation of the Loan Status**

How do people who paid off their loans compare with people whose loans had to be charged off? How do they compare with delinquent, **and other types of LoanStatus**

Q How is the estimated loss of the principal connected with the Net Principal loss? How is the principal loss “Estimated”?

What fraction of the gross payment is principal and what fraction is interest? This will pertain to loans that were charged off.

Compare variables like CreditScore, CurrentCreditLines (are they too many), EmploymentStatus,DelinquenciesLast7Years,OpenRevolvingMonthlyPayment, AmountDelinquent,BankcardUtilization,DebtToIncomeRatio , AvailableBankcardCredit (7 variables)

* 1. **Which factors affect your CreditGrade /CreditScore**

**(Boxplots, barcharts, Trends, Correlations)**

CurrentCreditLines (are they too many), EmploymentStatus,DelinquenciesLast7Years,OpenRevolvingMonthlyPayment, AmountDelinquent,BankcardUtilization,DebtToIncomeRatio , AvailableBankcardCredit

Q Which type of loans are most likely to be paid off? (LisitngCategory). Which categories have high/low APR?

Whats the difference between CurrentCreditLines and OpenCreditLines? What are OpenRevolvingAccounts?

Q : Do people who have lots of OpenCreditLines also have OpenRevolvingAccounts?

Q: How does InquiriesLast6Months and TotalInquiries affet your CreditScore.

Q: What fraction of people asking for loans had prior Prosper Loans? Do people who have Prosper Loans get a lower APR compared to people who don’t? Does it matter if they have made payments on time or not (OnTimeProsperPayments, NumberofDatsDelinquent), What about ProsperPrincipalBorrowed?

Q: Does ProsperRating and CreditGrade have a correlation?

**Variation of BorrowerAPR with certain numerical variables** like OpenCreditLines, OpenRevolvingAccounts, TotalProsperLoans, EmploymentStatusDuration, AmountDelinquent,RevolvingCreditBalance,DebtToIncomeRatio, IncomeRange, IncomeVerifiable

Which variables have a high correlation?

Q: How many people who have income have a verifiable Income? How does it affect their BorrowerAPR

Notes:

Everyone has OpenRevolvingAccounts (No NA’s)

In the data 591 people have no CreditCards -> Remove these.

5 people do not have any APR’s

29084 have no Prosper Loans or Ratings. EstimatedEffectiveYield EstimatedLoss

Seem to depend on how people who have prosper loans have worked in the past.

LenderYield has no Na’s.

Everyone has a loanstatus assigned. (No NA’s).

Everyone has a CreditGrade.

