Prosper Loan Data Review

Assessment and Cleaning Process

for the prosperLoanData.csv

Wrangle :

Data was provided via Udacity, no further data was sought aside from the single .csv

Assessment :

81 Columns, 113937 Rows. Single file will be split into multiple dataframes based on the required Observational Units sought. Each group of related columns were divided into color groups, Yellow, Olive, Orange, Flesh, dark Blue, Red, Pink, Grey and White. These were then assigned to different dataframes for observation. Some of these will be further refined in order to produce desired tables. Initial observation unit columns assigned to each are as follows:

* Yellow : PercentFunded
* Olive : CreditScoreRangeLower/CreditScoreRangeUpper
* Orange : BorrowerRate
* Flesh : ListingCategory
* Blue : CurrentlyInGroup
* Red : OpenRevolvingMonthlyPayment
* Pink : CurrentDelinquencies
* Grey : LoanOriginalAmount
* White : LoanStatus
* Brown : IncomeRange
* Green : ?

Note these are the first pass observational units we’re interested in, each of these first wave dataframes may need to be further cut as we locate other points of interest. At this point, we can ask a few questions:

* What has the largest impact on PercentFunded
* How does delinquency impact CreditScoreRange(Lower/Upper)
* What impacts BorrowerRate
* How does ListingCategory effect rate/percentfunded/loan amount
* What is the effect of being CurrentlyInGroup
* How are RevolvingMonthlyPayments related to delinquencies
* How does Currentdelinquency impact scorechanges
* What relationships does loan original amount have
* What impacts loanstatus

Further analysis may find additional questions worth asking. Yellow contains many possible values to breakdown, by design, at this stage.