(Communicate-Data-Finding)
by (Chidozie Patrick Ahamefule)

Prosper Loan Dataset

> During the course of this project, you would have to immerse yourself in the world of Prosper Loans company in order to have a full grasp of the transactional/historical data. Each of the data has two parts that could show the importance of data visualizations and its uniqueness and technical requirements in the data analysis process. Almost throughout this project, you will use Python visualization libraries like the ever flexible Matplotlib, the ubiquitous Seaborn, the ever relaible Pandas to systematically explore & visualize a selected dataset(s).

Here , we begin by knowing the composition of our inherited dataset. Also, the datatypes that come with the columns by default to see if it will have any negative impact in the behavior of the dataset in the nearest future. We started by making Univariate plots using only single variables then gradually, we proceeded into orchestrating Bivariate plots and eventually Multivariate plots.

In the other part of this project, we are required to produce a short presentation/slideshow that clearly describes interesting findings, discoveries, and relationships/correlations that were discovered while wrangling our Prosper Loan dataset. The major way to convey all findings, observations, trends etc would be by transforming your exploratory visualizations from the first part into polished, explanatory visualizations..

Summary of Findings

>. During the course of visually exploring the Prosper Loan dataset, the following points could be stated to have been found/discovered/uncovered. The original dataset comprised of 113,937 rows of data & 81 columns According to the ratings as stipulated in the Prosper company data dictionary, (0 - N/A, 1 - HR, 2 - E, 3 - D, 4 - C, 5 - B, 6 - A, 7 - AA), 7 - AA is the highest possible rating. From the univariate barchat above, ONLY about 5,000 Prosper customers fall into this category out of a possible 90,000. A huge majority of Prosper loans are current/ongoing loans. This is good.
Loan Defaulters constitute only a little percentage of total loans disbursed.
Close to twenty thousand(20,000) loanees have completed their repayment transactions.
A vast majority of Prosper Loanees are 'Employed'. This is a good sign.
The 'StatedMonthlyIncome' is highly correlated with the 'TotalProsperLoan'.
Also, the 'LoanOriginalAmount' is correlated to the 'StatedMonthlyIncome'
Defaulted loanees seem to be higher than expected and many such loanees seem to possess high ProsperRatings.
Many defaulters have a low ProsperRating (Alpha).

Key Insights for Presentation

>I delved deeper to understand the influence of the EmploymentStatus,
StatedMonthlyIncome, and ProsperRating (Alpha) on the probability of
loans to be repayed completly without the risk of defaulting or been
charged off. Most of the loanees who had an employment had a high
probability of repaying their loans.

Also, high earners tend to complete their loans while those with a good
Prosper rating had a very high probability of completing their loan
repayments/obligations with Prosper Loans.