**Project Motivation**

This is an Udacity Nanodegree project.I was interested in using Prosper Loan Data to better understand :

* Univariate Exploration:
  + Loan status
  + Employment Status
  + Stated Monthly Income
* Bivariate Exploration:
  + Status and Prosper Rating
  + Credit Start with Listing Category
  + Loan Status and Loan Amount
  + Prosper Rating and Employment Status
* Multivariate Exploration:
  + Rating, Loan Amount and Loan Status
  + Credit category, Credit rating and outcome of Credit
  + Amount, Listing Category Loan and Loan Status Interact

### **Project Overview**

This project has two parts that demonstrate the importance and value of data visualization techniques in the data analysis process. In the first part, you will use Python visualization libraries to systematically explore a selected dataset, starting from plots of single variables and building up to plots of multiple variables. In the second part, you will produce a short presentation that illustrates interesting properties, trends, and relationships that you discovered in your selected dataset. The primary method of conveying your findings will be through transforming your exploratory visualizations from the first part into polished, explanatory visualizations.

### **Project Details**

This project is divided into two major parts. In the first part, you will conduct an exploratory data analysis on a dataset of your choosing. You will use Python data science and data visualization libraries to explore the dataset’s variables and understand the data’s structure, oddities, patterns and relationships. The analysis in this part should be structured, going from simple univariate relationships up through multivariate relationships, but it does not need to be clean or perfect. There is no one single answer that needs to come out of a given dataset. This part of the project is your opportunity to ask questions of the data and make your own discoveries. It’s important to keep in mind that sometimes exploration can lead to dead ends, and that it can take multiple steps to dig down to what you’re truly looking for.

In the second part, you will take your main findings from your exploration and convey them to others through an explanatory analysis. To this end, you will create a slide deck that leverages polished, explanatory visualizations to communicate your results. This part of the project should make heavy use of the first part of the project.

#### **Choose your Dataset**

First, you will choose a dataset from the Dataset Options.

Download the Dataset Options file for full details & descriptions from the Resources Tab.

* Click on Resources in the leftmost panel of your classroom
* Click the File Name to start download

**Quick List Below:**

Dataset Options

* [Ford GoBike System Data](https://www.google.com/url?q=https://www.fordgobike.com/system-data&sa=D&ust=1554486256012000)
* [Flights](https://www.google.com/url?q=http://stat-computing.org/dataexpo/2009/the-data.html&sa=D&ust=1554486256017000)
* [Loan Data from Prosper](https://www.google.com/url?q=https://s3.amazonaws.com/udacity-hosted-downloads/ud651/prosperLoanData.csv&sa=D&ust=1554486256021000) with [Prosper Data Dictionary to Explain Dataset's Variables](https://www.google.com/url?q=https://docs.google.com/spreadsheet/ccc?key%3D0AllIqIyvWZdadDd5NTlqZ1pBMHlsUjdrOTZHaVBuSlE%26usp%3Dsharing&sa=D&ust=1554486256024000)
* [PISA Data](https://www.google.com/url?q=https://s3.amazonaws.com/udacity-hosted-downloads/ud507/pisa2012.csv.zip&sa=D&ust=1554482573645000) with [PISA Data Dictionary to Explain Dataset's Variables](https://www.google.com/url?q=https://s3.amazonaws.com/udacity-hosted-downloads/ud507/pisadict2012.csv&sa=D&ust=1554482573645000)

#### **Step 1.2: Explore Your Data**

It’s time to get to the interesting bits. Explore your data and document your findings in a report. The report should briefly introduce the dataset, then systematically walk through the points of exploration that you conducted. You should have headers and text that organize your thoughts and findings. Visualizations in this part of the project need not be completely polished: this is just your own exploration at this point. However, you should still make sure that you adhere to principles of using appropriate plot types and encodings so that accurate conclusions can be drawn, and that you have enough comments and labeling so that when you return to your work, you can quickly grasp your analysis steps.

#### **Step 2.1: Document your Story**

At the end of your exploration, you probably have a bunch of things that you’ve discovered. Now it’s time to organize your findings and select a story that you will convey to others. In your readme document, you should summarize your main findings and reflect on the steps you took in your data exploration. You should also lay out the key insights that you want to convey in your explanatory report as well as any changes to visualizations, or note new visualizations that will be created to bridge between your insights.

#### **Step 2.2: Create your Slide Deck**

Follow the plans you laid out in the previous step and create a slide deck with explanatory data visualizations to tell a story about the data you explored. You can start with code that you used in your exploration, but you should make sure that the code is revised so that your plots are polished. Make sure that you also pay attention to aspects of design integrity in your revisions.

#### **Step 2.3: Review and Submit the Project**

There’s one last thing to do before you submit your project. You should closely read the [project rubric](https://review.udacity.com/#!/projects/8ff9475b-3d6b-4c5b-9593-96794db62987/rubric).

### **Dataset**

This data set contains information on peer to peer loans facilitated by credit company Prosper. There are 113,937 loans with 81 variables. For the purpose of this investigation I've taken the following variables: Term, LoanStatus, BorrowerRate, ProsperRating (Alpha), ListingCategory (numeric), EmploymentStatus, DelinquenciesLast7Years, StatedMonthlyIncome, TotalProsperLoans, LoanOriginalAmount, LoanOriginationDate, Recommendations and Investors.

### **Results**

* Most of the loans in the data set are actually current loans. Past due loans are split in several groups based on the length of payment delay. Other big part is completed loans, defaulted loans compromise a minority, however chargedoff loans also comporomise a substanial amount.

[Chart, histogram

Description automatically generated](https://raw.githubusercontent.com/Abhishek20182/Communicate-Data-Findings/master/Result/Result-1.png)

- The majority of borrowers are employed and all other categories as small part of borrowers. In small Group full time has highest, after that self empolyed are there and so on.

[Chart, histogram

Description automatically generated](https://raw.githubusercontent.com/Abhishek20182/Communicate-Data-Findings/master/Result/Result-2.png)

- With a boundary of mean and 3 times standard deviations distribution of monthly income still has noticeable right skew but now we can see that mode is about 5000.

[Chart, histogram

Description automatically generated](https://raw.githubusercontent.com/Abhishek20182/Communicate-Data-Findings/master/Result/Result-3.png)

- The most frequent rating among defaulted loans is actually D. And the most frequent rating among Completed is alsoDand second highest is A and so on in Status and Prosper Rating.

[Chart, histogram

Description automatically generated](https://raw.githubusercontent.com/Abhishek20182/Communicate-Data-Findings/master/Result/Result-4.png)

- In both of them Credit Start with Listing Category Graphs of the debt Consolidation have most frequency among all of them

[Chart, histogram

Description automatically generated](https://raw.githubusercontent.com/Abhishek20182/Communicate-Data-Findings/master/Result/Result-5.png) [Chart, box and whisker chart

Description automatically generated](https://raw.githubusercontent.com/Abhishek20182/Communicate-Data-Findings/master/Result/Result-6.png)

- Lower ratings seem to have greater proportions of individuals with employment status Not Employed, Self-employed, Retired and Part-Time.

[Chart, histogram

Description automatically generated](https://raw.githubusercontent.com/Abhishek20182/Communicate-Data-Findings/master/Result/Result-7.png)

- Except for the lowest ratings defaulted credits tend to be larger than completed. Most of the defaulted credits comes from individuals with low Prosper rating.

[Chart, box and whisker chart

Description automatically generated](https://raw.githubusercontent.com/Abhishek20182/Communicate-Data-Findings/master/Result/Result-8.png)

- Except for Auto, Business and Home Improvemrnt dont have nearly equal mean amoong all of them. Business category tend to have larger amount.

[Chart

Description automatically generated](https://raw.githubusercontent.com/Abhishek20182/Communicate-Data-Findings/master/Result/Result-9.png) [Shape

Description automatically generated](https://raw.githubusercontent.com/Abhishek20182/Communicate-Data-Findings/master/Result/Result-10.png)

## Insights for Presentation

I've chosen key plots with high data-to-ink ratio for the presentation. The plots I've chosen shows distribution of main variables, Loan status, monthly income, Prosper rating and I've tried to tell a story what are major predictors for loan status and Prosper rating variables.

### Why this project?

Data visualization is an important skill that is used in many parts of the data analysis process.

**Exploratory** data visualization generally occurs during and after the data wrangling process, and is the main method that you will use to understand the patterns and relationships present in your data. This understanding will help you approach any statistical analyses and will help you build conclusions and findings. This process might also illuminate additional data cleaning tasks to be performed.

**Explanatory** data visualization techniques are used after generating your findings, and are used to help communicate your results to others. Understanding design considerations will make sure that your message is clear and effective. In addition to being a good producer of visualizations, going through this project will also help you be a good consumer of visualizations that are presented to you by others.