Clearly define a problem or an idea of your choice. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.

This submission will eventually become your **Introduction/Business Problem** section in your final report. So I recommend that you push the report (having your Introduction/Business Problem section only for now) to your Github repository and submit a link to it.

Credit card companies and peer to peer lenders are keen on lending money to people who will not default. Invariably, some will default. The challenge to lenders is identifying who will default while also not discriminating along race, gender, or other factors. Without discriminating there are some factors that one could consider that will lead to determining whether or not the one receiving a loan will default. Namely, lenders could look at the age, previous defaults, levels of current debt, incomes, assets and liabilities.

Describe the data that you will be using to solve the problem or execute your idea. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using.

This submission will eventually become your **Data** section in your final report. So I recommend that you push the report (having your **Data** section) to your Github repository and submit a link to it.

In this notebook I utilize a data set of a group of Americans from 2005-2015. It represents three attributes: age, income measured in USD, and loan amount also measured in USD. I drop the “Identification” at the beginning as this attribute is not helpful.

First, I employ a confusion matrix to better understand toward which class the model is biased. Next I use the K Nearest Neighbor Classifier to determine the nearest training examples in the feature space. This is a non-parametric methodology for classification and regression.