Intro to Data Analysis Final Project

The main task of the project is to analyze a dataset and then communicate the findings about it. and I should use the Python libraries NumPy, Pandas, and Matplotlib to make the analysis easier.

Udacity's Introduction to the project

> For the final project, you will conduct your own data analysis and create a file to share that documents your findings. You should start by taking a look at your dataset and brainstorming what questions you could answer using it. Then you should use Pandas and NumPy to answer the questions you are most interested in, and create a report sharing the answers. You will not be required to use statistics or machine learning to complete this project, but you should make it clear in your communications that your findings are tentative. This project is open-ended in that we are not looking for one right answer.

Choosing a dataset

There are two available data sets:

1. Titanic Data (https://www.udacity.com/api/nodes/5454512672/supplemental_media/titanic-datacsv/download) - Contains demographics and passenger information from 891 of the 2224 passengers and crew on board the Titanic. You can view a description of this dataset on the Kaggle website (https://www.kaggle.com/c/titanic/data), where the data was obtained.
2. Baseball Data (http://www.seanlahman.com/baseball-archive/statistics/) - A data set containing complete batting and pitching statistics from 1871 to 2014, plus fielding statistics, standings, team stats, managerial records, post-season data, and more. This dataset contains many files, but you can choose to analyze only the one(s) you are most interested in.

*I don't know any thing about Baseball so I chose **the Titanic Data***

Getting started

to be organized I created a new folder for the project which contains this Ipython notebook and the dataset "titanic_data.csv"

Analyzing The Data

In [1]:
```
# importing the python libraries that will be used for analysis and visualization
%matplotlib inline
import numpy as np
import pandas as pd
```