Capstone Project 1 Proposal
Predict the value of a Property in New York City

1. What is the problem you want to solve?
   The value of a property in NYC is very dynamic and is based on many factors like the location, age of the building, etc. An issue that both buyers and sellers face today is to find out the correct value of a property due to differences between the values presented by different brokers. The aim of this project would be to build a predictive model that will provide the buyers and sellers with a range for the value based on data from past sales.

2. Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?
   The primary client here would be the people who are buying or selling a property in NYC. They could have access to this predictive model through a website where they enter the details of the property they are planning to buy/sell and get the results. This way they would get the real picture will be aware of the actual value instead of having to depend on brokers.

3. What data are you using? How will you acquire the data?
   The data being used to solve this problem will be the NYC sales rolling data that is available as an excel sheet here:
   https://www1.nyc.gov/site/finance/taxes/property-rolling-sales-data.page

   Since the data is available as an excel sheet, it can be easily read and predictive models can be created based on the data present in it.

4. Briefly outline how you will solve this problem.
   In order to solve this problem many criteria (Borough, Zip Code, past sale price, Year Built, etc.) that are present in the dataset can be considered.
   - The first step would be to perform Data Wrangling that will be used to clean the data and check for any inconsistencies in the dataset.
   - Exploratory Data Analysis (EDA) would be the next step to check for possible trend and/or correlations in the data. Preliminary prediction of the value of the property using less criteria should be possible at this stage.
   - The third step would be to obtain statistical inferences of the data based on the results obtained in the previous step.
   - Finally, it is time to work on the model and refine it to get the range of the value of any given property in New York City based on the previous sales.

5. What are your deliverables?
   Deliverables for this project would include the following:
   - Code in iPython Notebook
   - Final Report with statistical analysis results and sample data runs.