**Predict the value of a property in New York City**

The value of a property in NYC is very dynamic and is based on multiple factors like location, year built, etc. An issue faced by both buyers and sellers is to find out the correct value of a property due to differences between the value of the property presented by different brokers. This project aims to solve that problem and provide the buyer/seller with a range for the property value.

For this project we are using the property sales data from the NYC gov website below:

<https://www1.nyc.gov/site/finance/taxes/property-rolling-sales-data.page>

The data here is present in the form of 5 excel spreadsheets (one for each borough).

In order to be able to work with all the data the excel spreadsheets were all ingested and concatenated together into a single dataframe.

The next step was to check all the columns and make sure that there is no missing data in any of the columns. The following columns had some discrepancies:

* Sales Price: The sales price was found to be $0 for 24,707 rows. Since the Sales Price is a major factor in the analysis we are going to ignore those records. This brings or sample set down to 58,707 records.
* Gross Square Feet: It was found that the Gross Square Feet value was 0sqft for 29,221 rows. This is almost 50% of the records. So, we are not going to take this column into consideration for our analysis.
* Zip Code: The Zip Code was found to be missing for 299 records. A mapping was made between the neighborhood and zip code to map a single neighborhood to the zip code. During the process it was found that a single neighborhood had more than 1 zip code related with it. So, the zip code that had a higher count for the neighborhood was considered for the mapping. This mapping was then used to fill in the zip code for the missing zip codes.
* Year Built: The Year Built was also found to be missing for 3,698 records which were removed from our dataset since the year built is vital to our analysis.

All the other columns had no data missing and have been taken into consideration for further analysis to successfully predict the value of a property in NYC.