**Documentation including metadata for data created and data quality Insights**

METADATA THAT HAS BEEN CREATED BY ME

1. Zillows
   1. Year: New column heading generated after pivoting columns.
   2. Value: Value of Price of the property year wise.
   3. Predicted\_value\_2018: This value is predicted by the machine learning model by analyzing data of the previous years.
   4. Hike\_in\_one\_year: (Predicted \_value \_2018 – Value)
2. AirBnb
   1. Rent Earned Per year: (Assuming that 75% occupancy) 0.75\*365 \* Price Column.
3. After Merging Airbnb and Zillow’s:
   1. Profit Percentage:

Profit percentage = ("Rent earned per year" + "hike in one year" ) / "initial cost of a property"

This Column will tell us that which apartment Zip code wise is given us the most return in an year.

* The three quality insights from the initial analysis:

Zillow’s data:

1. Most of the property has NULL value for the year 1996-2009. Therefore, I have not used data for given time period for model making.
2. Invalid data: In column “city”, there are so many bad entries which a City could not be filtered on. So, I used Metro and State to filter out NYC data.
3. Data is given for 1996-2017 by each column wise. So, data needs to be pivoted for better analysis.

Airbnb Data:

1. Missing data: “weekly price” and “monthly price columns” has way too much missing data that I prefer not do go for Data Imputation. Instead I preferred to use “price” column which is the price per night of a property.