NumBldgs - PLUTO and Building Footprints Discrepancies

Goal: To determine how often the number of buildings on a lot differs between PLUTO (derived from GeoSupport) and building footprints, and how often one or the other is correct.

Result: Building Footprints is often the more correct source, but perhaps not 100% of the time.

The steps I took were the following:

Find Lots with NumBldgs Discrepancies

I obtained the number of building footprints per lot by grouping on MPLUTO_BBL. Then I joined the the result to PLUTO on BBL, and selected only those lots where NumBldgs wasn't equal to the count of footprints.

This gave me a count of 20,797 lots.

Determine the Impact by Borough

Were we to substitute the building footprint count for the GeoSupport number we store now, the impact by borough would be as follows:

QN	7798
BK	5942
SI	4082
BX	2294
MN	681

Distribution of Differences

I subtracted the building footprints count from NumBldgs for a difference for each lot. I then created a data frame of difference values and counts using Panda's value_counts function. The results are on one of the tabs in the spreadsheet of results. Unsurprisingly, the most common differences are 1 and -1, but occasionally there are large differences such as those found in such areas as Breezy Point (an area largely destroyed by Hurricane Sandy).

Which Source is More Correct?

In order to determine which source would be more correct, I took every 100th lot of the 20,797 lots with differences, for a total of 207 lots, and checked them manually using ZOLA, Cyclomedia, and Google Street View. First I used ZOLA to get a sense of how the lot was situated, and then used the aerial view of Google Street view to determine how many buildings are really standing on the lot. If Google did not give me satisfactory results, I used Cyclomedia for more up to date information.

Thursday, March 19, 2020

In 75% of the cases, Building Footprints appeared to have correct information, but in 17% of cases, I thought PLUTO was correct. In another 6% of cases, I wasn't sure which was right, and for one lot, I didn't believe either was correct. A list of the lots I checked with the results I found is included on the spreadsheet of results.

Supporting Documents

Jupyter Notebook

Spreadsheet of Research Results

Thursday, March 19, 2020