NumFloors Methodology

Goal: To review discrepancies between PLUTO's NUMFLOORS field and the number of stories implied by building footprints' HEIGHTROOF field, and make suggestions for repairs.

Result: I am suggesting that 25 PLUTO records be fixed via the manual corrections file.

Query Development Process:

Create three temporary tables:

1. **PLUTO**: Select records from PLUTO where the NUMBLDGS field is 1 or less than 1. Additionally, NUMFLOORS should be greater than zero. When doing the selection, I evaluated the LANDUSE field for residential values: values '1', '2', or '3' will be considered residential. The query should return a floor height value of 10 feet for residential structures, and 12 for non-residential structures.
2. **Building Footprints:** Select building footprints records, and limit the result set to single occurrences of a BBL. (Note that selecting only those footprints with one record per BBL omits the 205-floor building in Queens.) Return MPLUTO_BBL and HEIGHTROOF, grouping by MPLUTO_BBL. HEIGHTROOF must be greater than zero.
3. **Housing**: Select records from the housing table, where the job type is 'New Building', and the completion date is greater than 2013.

Calculate the approximate number of floors for building footprints, and the difference between PLUTO's NUMFLOORS and the approximate number from footprints. Join PLUTO to footprints by BBL (using MPLUTO_BBL on footprints), omitting those records having a match on the temporary housing table with new building permits.

1. **Footprints Approximate Floors:** Divide HEIGHTROOF by the floor height value established earlier (10 or 12 feet depending upon the building class). Round the result to the nearest digit.
2. **PLUTO / Footprints Difference:** Subtract the approximate number of floors from PLUTO's NUMFLOORS. Convert the result to its absolute value.

Limiting the size of the output list:

1. I limited the results to situations where the difference between PLUTO and building footprints is large: the PLUTO / Footprints difference was more than twice the NUMFLOORS value or the calculated number of floors from footprints.
2. Additionally, the number of floors must be greater than 10 on either PLUTO or footprints. (I did not want to evaluate situations where one was twice the other, but where the numbers were quite small to begin with, say 2 floors on one side, and 4 on the other.)

3. Finally, I omitted records where the first position of the building class is 'M' (churches, synagogues, etc.) or 'Q' (monuments such as Grant's Tomb). Traditional churches often have one floor only, with that floor being greatly in excess of 12 feet high (although I realize there are non-traditional situations). I did not want to pick up monuments since I didn't really think that was the goal of the project.

This query returned 92 rows for manual review. There is a handful of PLUTO records with high NUMFLOORS values that have no corresponding match on the footprints table, a 100-floor Staten Island building among them. I am recommending that this lot be fixed but otherwise am leaving these alone.

Follow-Up with Another Threshold:

I continued to experiment with thresholds and ran another query where the PLUTO / Footprints difference was more than one times the NUMFLOORS value or the calculated number of floors from footprints. The effect of this was to expand the result set from 92 to 136 rows, but in reviewing these results manually I only identified one additional record for repair.

Manual Review Process:

I used Cyclomedia for this part of the process. I downloaded the shapefile for the latest version of MapPLUTO from DCP's website, loaded it into ArcMap, selected the BBL, zoomed into the highlighted lot, and clicked on recordings to bring up the relevant Cyclorama images. The idea was to review at least one image and count the floors, if possible.

It wasn't long before I noticed that my list contained a fair number of new buildings, which had not been filtered out with the query. Often in these cases PLUTO's NUMFLOORS was more accurate than the number suggested by the algorithm. I left these cases alone.

Sometimes I could not count the floors. This might have happened because there were trees blocking the building, or because the building had a large tower and it was difficult to determine how many floors were in that tower. Sometimes I saw towers and wasn't sure if the tower was part of the building in question, or part of a neighboring building. I left these cases alone.

With this process I determined that there were 23 records requiring repair. All of the extreme cases that had caught our eye earlier (the 205 story building in Queens, for instance) are included on the list.

Follow-Up using BLDGCLASS check in lieu of LANDUSE:

I ran an additional query to identify possible discrepancies using the BLDGCLASS code instead of LANDUSE. If a lot's BLDGCLASS code equals S3, S4, S5, R1, R2, R3, R4, R6, or its first position equals 'A', 'B', 'C', or 'D', I assign a floor height of 10 feet. All other buildings are assigned a floor height of 12 feet.

Like the previous query, this query determines the difference between the NUMFLOORS value and the derived floors, converts the result to absolute value, and then compares the difference with the NUMFLOORS value and the derived value to make sure the difference is at least the same size as one of them.

This query identified all the same lots as before, but eleven additional ones as well. Having reviewed these manually, I am recommending changes to two lots, bringing the total to 25.

Supporting Documents:

Jupyter Notebook containing SQL queries

Excel Spreadsheet of Potential Changes for Evaluation (may be downloaded from Github)

Excel Spreadsheet of Additional Potential Change, Threshold of 1

CSV of Manual Corrections