**Address Points to Building Footprints Proof of Concept**

This is a proof of concept taking from the work Jesse Stewart did linking address points to road segments here: <https://github.com/jessestewart1/nrnrrn/blob/development/src/stage_1/segment_addresses.py>

This project takes the address points to road segments code and converts to take address points and match them to the most appropriate building footprint polygon(s). This is done using python 3 and several python modules including:

* Geopandas
* Numpy
* Pandas
* Shapely

**Data Overview**

Below is an overview of how the building footprint and address point data relate to each other.

|  |  |
| --- | --- |
| Query | Count |
| Total number of building footprints | 12,282 |
| Building footprints that intersect an address point | 2.722 |
| Number of building footprints containing multiple address points | 357 |

There 9560 building footprints that do not have an intersecting address point. There are many buildings that we do not want to apply an address to for example 4,566 building footprints with an area of 20m2 or less. In the example below there are several structures in the parking lot that are obviously not addressable structures. An exclusion criterion should be created to filter out these structures.



**Results**

|  |  |  |
| --- | --- | --- |
|  | With Linking Data | With 15m Buffer |
| Total Number of Matches | 6763 | 6124 |
| Total Number Accurate Matches | 5119 | 311 |
| Match Accuracy | 75% | 0.5% |

\* Note that these matches are all done without accounting for point intersections with polygons. The number of accurate matches would be expected to go up significantly if that process was instantiated.

**With Linking Data**

The first method used to find matches used attributes from underlying surveyed parcel data to get the most accurate between building footprints and the address points. The results from this method returned the following results:

Of the 6763 matches output by the tool 5119 (75%) matched exactly. These can be considered good matches

**Without Linking Data**

The second method used was to create 15-meter buffer around every address point and to select all building footprints that were within that buffer zone. This selection was used in place of the linking data that was used to take a smaller subset of the data to use in the distance comparison between the points

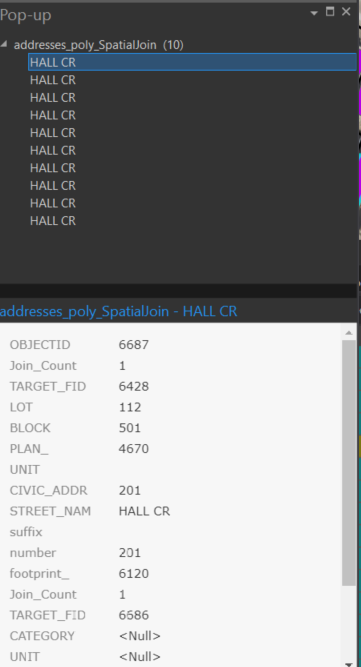
Of the 6124 matches output by the tool 311 (0.5%) were accurate. Although this low percentage of matches is in part down to the lack of intersect recognition in the scripts. The number of matches can be expected to rise significantly once this is added in.

**Mismatches Analysis**

Of the mismatches there were many different scenarios that were prevalent in the data:

1. many were repeats of a building footprint one of which was the correct address. See the example of 411 Hall crescent below



The highlighted building is 411 Hall Crescent however there are several (10) copies of the building underneath each with a different street address.

These copies correspond to the address points in the empty parcels surrounding the building. If these additional copies were filtered out the number of mismatches would be reduced. There are also examples of where a doubling in homes that are attached and share the same parcel but have different address numbers. 

The image above shows a building that has two associated records in the output data. 5112 and 5110 47th Street both share the same polygon meaning two polygons were created in the same place with one being flagged as a mismatch.

1. There are several cases where the matched civic address and the actual civic address are off by >10. These tend to occur in clusters with other addresses
2. There are several cases where a building footprint will be in more than one parcel and the address information for the wrong parcel ends up getting applied to the

Besides mismatches there are also several instances where no match was found when visually it is apparent that a match was possible. There are several different types of these cases.

1. When there are multiple buildings on the same property in the same parcel only one gets an address match.



As seen in the image above the purple building footprints have a match the red buildings are unmatched. While the main buildings on the property got an accurate match, all subordinate buildings are unmatched when they should have at the same address as the main building. Also seen above is a case where the subordinate building was given a match but the main building went unmatched.

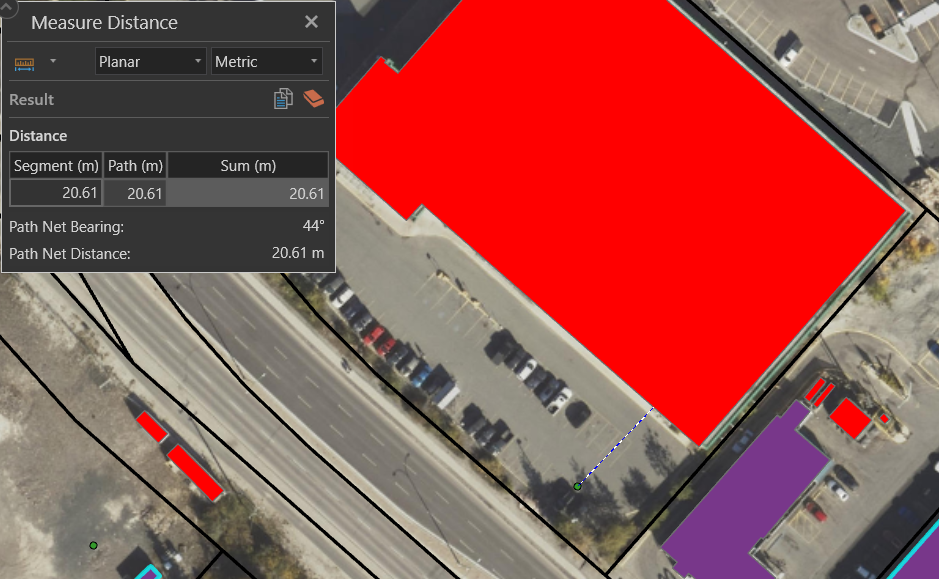
1. There are several cases where large buildings containing many points return no matches.



As seen in the image above there are many points in the building footprint but no match was made this is the case for many large multi points buildings in the city. All the points are listed as units with unit numbers but the same civic address.

**Buffer Methodology Specific Issues**

1. In the case of the non-linking data methodology there were several cases in which the possible point match fell outside of the search area buffer such as in the example below where the correct point is 20m away from the point and therefore outside of the current 15m search radius.



**Methodology Improvement Recommendations**

There are a couple things that can be done to further improve the accuracy of the methodology.

1. Exclude non-addressable structures from the data such as cart returns, bus shelters, and utility sheds. This should be done as part of an initial data cleaning phase. Establish a parameter where such buildings are ignored by the process (ex. Area less than 20m2).
2. In both match finding methods used direct intersections between the address points and the building footprints were not considered. As seen in the issues section above the there are many cases where a point directly intersects a building footprint but no match is made. Adding in this functionality to the methodology should allow for more matches that are overall more accurate.
3. Structure the process to check for linking data first such as surveyed parcel data and use the methodology first as it is far more accurate. All building footprints that are not matched using that process should then be triaged into the buffer methodology where the existence of points within a certain distance of the will be checked for.
4. Add in a way to handle buildings that contain more than one address for example attached townhouses that have different addresses but get counted as the same building