#### 1) explore.ipynb

Input: output\_1850\_bk\_add\_std.csv / output\_1850\_mn\_add\_std.csv / output\_1880\_add\_std.csv

Output: doubt.xlsx, not so confident.xlsx, confident\_new.xlsx

Role: Splits input into 3 categories depending on whether or not they have a clean address (with proper house number and street name)

#### 5) combine-new3.ipynb

Input: doubt-mn1880-v7-s.csv / not so confident-mn1880-v7-s.csv (could also be 1850 bk/1850 mn)

Output: mn\_1880\_combinedoubt4.csv / mn\_1880\_combine-not so confident4.csv

Role: Combines the address standardization output with the output from doubt.ipynb (hence, leading to near and corner cases being incorporated).

### 2) doubt.ipynb

Input: doubt.xlsx / not so confident.xlsx

Output: doubt-mn1880-v7.xlsx, not so confident-mn1880-v7.xlsx (could also be 1850 bk/1850 mn)

Role: Performs cleaning on the addresses by using regex. Also adds columns with corner and near street names

### 4) address\_standarization\_code-Copy1.ipynb

Input: doubt-mn1880-v7.csv, not so confident-mn1880-v7.csv (could also be 1850 bk/1850 mn)

Output: doubt-mn1880-v7-s.csv, not so confident-mn1880-v7-s.csv (could also be 1850 bk/1850 mn)

Role: Adds more columns like new h house number, new h street name

## 2) 02\_CRF\_1880-Copy2.ipynb

Input: doubt-mn1880-v7.xlsx, not so confident-mn1880-v7.xlsx (could also be 1850 bk/1850 mn)

Output: doubt-mn1880-v7.pkl, not so confident-mn1880-v7.pkl (could also be 1850 bk/1850 mn)

Role: Creates POS (parts of speech) tags for the addresses

# 3) 03\_Final\_Output\_1880.ipynb

Input: doubt-mn1880-v7.pkl, not so confident-mn1880-v7.pkl (could also be 1850 bk/1850 mn)

Output: doubt-mn1880-v7.csv, not so confident-mn1880-v7.csv (could also be 1850 bk/1850 mn)

Role: Adds numerous columns like H House Number, H Street Name, H City

# 6) Street cleaning modified.R

Input:

mn\_1880\_combine-doubt4.csv / mn\_1880\_combine-not so confident4.csv

#### Output:

confident4.csv

mn\_1880\_after\_cleaning\_doubt4. csv / mn\_1880\_after\_cleaning\_not so

Role: Replaces street names with matched counterparts

# 7) before address builder.ipynb

Input:

mn\_1880\_after\_cleaning\_doubt4.
csv
/ mn\_1880\_after\_cleaning\_not so
confident4.csv

#### Output:

mn\_1880\_doubt\_ready4.csv / mn\_1880\_not so confident\_ready4.csv

Role: Combines all columns into hn 1 and best match

### 8) Address builder.Rmd

Input:

mn\_1880\_doubt\_ready4.csv / mn\_1880\_not so confident\_ready4.csv

Output: mn\_1880\_doubt\_built.csv

/ mn\_1880\_not so confident built.csv

Role: Creates base\_function, spell\_ave\_true, suffix\_true and both\_true columns for geocoding