C750 - Data Wrangling

January 21, 2021

\#

Data Wrangling OpenStreetMap with SQL

By: Kyle Hansen

\#\#

Map Area

\#\#\#\#

Bothell, Washington, USA

https://www.openstreetmap.org/relation/237656

Project Summary: For this project, using data wrangling techniques, i chose Bothell, WA map area from OpenStreetMap.org. The reason for this choice is because i grew up in Bothell. The cleaned, fixed data starts in XML and is converted into CSV format, and is then imported into a SQL database.

0.0.1 Tag Count

When Parsing through the Bothell, WA dataset and couting the unique element types by using the tags function, the following are the number of unique tags:

```
[2]: osm_file = "bothell.osm"
```

```
[3]: import tags
     tags.count_tags(osm_file)
```

```
[3]: defaultdict(int,
                 {'bounds': 1,
                  'member': 34048,
                  'meta': 1,
                  'nd': 700084,
                  'node': 607354,
                  'note': 1,
                  'osm': 1,
                  'relation': 994,
                  'tag': 363047,
                  'way': 79027})
```

0.0.2 Tag Patterns

There are three regular expression, lower is for tags that contain only lowercase letters and are valid. lower_colon is for other valid tags with a colon in the value. problem is for tags with problematic characters. The results are as below:

```
[12]: import tag_type
      tag_type.process_map(osm_file)
```

```
[12]: {'lower': 165323, 'lower_colon': 194710, 'other': 3014, 'problem': 0}
```

##

Problems Seen in the Data

0.0.3 Street Types:

Because the data is inputed by users, there is no standard format for names. With a list of expected street types inputed, the following dictionary of streets have unexpected types. The majority of the issues involve abbreviations such as St for Street.

```
[14]: import audit
      audit.audit(osm_file)
```

```
[14]: defaultdict(set,
                  {'99': {'Highway 99', 'Hwy 99', 'State Highway 99'},
                   'Ave': {'44th Ave'},
                   'Blvd': {'Alderwood Mall Blvd'},
                   'C': {'228th Street SE Suite C'},
                   'E': {'Martin Way E'},
                   'East': {'88th Avenue East'},
                   'H': {'Hwy 99, Ste H'},
                   'Hwy': {'Bothell-Everett Hwy'},
                   'NE': {'104th Ave NE',
                    '120th Avenue NE',
                    '127th Ave NE',
                    '138th Way NE',
                    '141st Pl NE',
                    '19115 112th Ave NE',
                    '25th Ave NE',
                    '5th Ave NE',
                    '68th Ave NE',
                    '94th Pl NE',
                    'Ballinger Way NE',
                    'Bothell Way NE',
                    'Juanita-Woodinville Way NE'},
                   'North': {'Ashworth Avenue North',
                    'Bagley Avenue North',
                    'Burke Avenue North',
                    'Corliss Avenue North',
```

```
      'Courtland Place North',
      'Densmore Avenue North',
      'Interlake Avenue North',
      'Meridian Avenue North',
      'North Creek Parkway North',
      'Roosevelt Way North',
      'Roslyn Place North',
      'Stone Avenue North',
      'Sunnyside Avenue North',
      'Wallingford Avenue North',
      'Wayne Place North'},
 'Northeast': {'100th Avenue Northeast',
      '100th Place Northeast',
      '101st Avenue Northeast',
      '101st Place Northeast',
      '102nd Avenue Northeast',
      '102nd Court Northeast',
      '102nd Place Northeast',
      '103rd Avenue Northeast',
      '103rd Court Northeast',
      '103rd Place Northeast',
      '104th Avenue Northeast',
      '104th Place Northeast',
      '105th Avenue Northeast',
      '105th Court Northeast',
      '105th Place Northeast',
      '106th Avenue Northeast',
      '106th Place Northeast',
      '107th Avenue Northeast',
      '107th Place Northeast',
      '108th Avenue Northeast',
      '108th Place Northeast',
      '109th Avenue Northeast',
      '109th Court Northeast',
      '109th Place Northeast',
      '10th Avenue Northeast',
      '110th Avenue Northeast',
      '110th Place Northeast',
      '111th Avenue Northeast',
      '111th Place Northeast',
      '112th Avenue Northeast',
      '112th Lane Northeast',
      '112th Place Northeast',
      '113th Avenue Northeast',
      '113th Place Northeast',
      '114th Avenue Northeast',
      '115th Avenue Northeast',
```

```
'116th Avenue Northeast',
'116th Place Northeast',
'117th Avenue Northeast',
'117th Court Northeast',
'117th Place Northeast',
'118th Avenue Northeast',
'118th Court Northeast',
'118th Lane Northeast',
'118th Place Northeast',
'119th Avenue Northeast',
'119th Court Northeast',
'119th Lane Northeast',
'119th Place Northeast',
'119th Terrace Northeast',
'11th Avenue Northeast',
'120th Avenue Northeast',
'120th Court Northeast',
'120th Lane Northeast',
'120th Place Northeast',
'121st Avenue Northeast',
'121st Court Northeast',
'121st Place Northeast',
'122nd Avenue Northeast',
'122nd Court Northeast',
'122nd Place Northeast',
'123rd Avenue Northeast',
'123rd Lane Northeast',
'123rd Place Northeast',
'124th Avenue Northeast',
'124th Court Northeast',
'124th Place Northeast',
'125th Avenue Northeast',
'125th Lane Northeast',
'125th Place Northeast',
'126th Avenue Northeast',
'126th Place Northeast',
'127th Avenue Northeast',
'127th Court Northeast',
'127th Lane Northeast',
'127th Place Northeast',
'128th Avenue Northeast',
'128th Court Northeast',
'128th Place Northeast',
'129th Avenue Northeast',
'129th Court Northeast',
'129th Lane Northeast',
'129th Place Northeast',
```

```
'12th Avenue Northeast',
'130th Avenue Northeast',
'130th Court Northeast',
'130th Place Northeast',
'131st Avenue Northeast',
'131st Court Northeast',
'131st Lane Northeast',
'131st Place Northeast',
'132nd Avenue Northeast',
'133rd Avenue Northeast',
'133rd Place Northeast',
'134th Avenue Northeast',
'134th Court Northeast',
'135th Avenue Northeast',
'136th Avenue Northeast',
'139th Avenue Northeast',
'140th Avenue Northeast',
'140th Place Northeast',
'144th Avenue Northeast',
'148th Avenue Northeast',
'14th Place Northeast',
'155th Avenue Northeast',
'156th Avenue Northeast',
'156th Place Northeast',
'157th Avenue Northeast',
'157th Place Northeast',
'159th Avenue Northeast',
'15th Avenue Northeast',
'15th Place Northeast',
'160th Avenue Northeast',
'160th Place Northeast',
'162nd Avenue Northeast',
'164th Avenue Northeast',
'165th Avenue Northeast',
'166th Avenue Northeast',
'168th Avenue Northeast',
'16th Avenue Northeast',
'170th Lane Northeast',
'171st Avenue Northeast',
'171st Lane Northeast',
'172nd Avenue Northeast',
'173rd Place Northeast',
'174th Avenue Northeast',
'176th Avenue Northeast',
'177th Avenue Northeast',
'17th Avenue Northeast',
'194th Place Northeast',
```

```
'19th Avenue Northeast',
'1st Avenue Northeast',
'20th Avenue Northeast',
'22nd Avenue Northeast',
'23rd Avenue Northeast',
'23rd Place Northeast',
'24th Avenue Northeast',
'24th Place Northeast',
'25th Avenue Northeast',
'26th Avenue Northeast',
'27th Avenue Northeast',
'28th Avenue Northeast',
'2nd Avenue Northeast',
'30th Avenue Northeast',
'31st Avenue Northeast',
'31st Drive Northeast',
'32nd Avenue Northeast',
'35th Avenue Northeast',
'35th Court Northeast',
'36th Avenue Northeast',
'36th Court Northeast',
'37th Avenue Northeast',
'38th Avenue Northeast',
'39th Avenue Northeast',
'3rd Avenue Northeast',
'40th Avenue Northeast',
'41st Avenue Northeast',
'42nd Avenue Northeast',
'42nd Place Northeast',
'44th Avenue Northeast',
'45th Avenue Northeast',
'47th Avenue Northeast',
'4th Avenue Northeast',
'55th Avenue Northeast',
'56th Avenue Northeast',
'56th Lane Northeast',
'57th Avenue Northeast',
'57th Place Northeast',
'58th Avenue Northeast',
'58th Lane Northeast',
'58th Place Northeast',
'59th Place Northeast',
'5th Avenue Northeast',
'60th Avenue Northeast',
'60th Place Northeast',
'61st Avenue Northeast',
'61st Court Northeast',
```

'61st Lane Northeast',
'61st Place Northeast',
'62nd Avenue Northeast',
'62nd Court Northeast',
'62nd Place Northeast',
'62nd Way Northeast',
'63rd Avenue Northeast',
'63rd Lane Northeast',
'63rd Place Northeast',
'64th Avenue Northeast',
'64th Court Northeast',
'64th Lane Northeast',
'64th Place Northeast',
'65th Avenue Northeast',
'65th Court Northeast',
'65th Lane Northeast',
'65th Place Northeast',
'66th Avenue Northeast',
'66th Court Northeast',
'66th Place Northeast',
'67th Avenue Northeast',
'67th Lane Northeast',
'67th Place Northeast',
'68th Avenue Northeast',
'69th Avenue Northeast',
'69th Lane Northeast',
'69th Place Northeast',
'70th Avenue Northeast',
'70th Court Northeast',
'70th Lane Northeast',
'71st Avenue Northeast',
'71st Place Northeast',
'72nd Avenue Northeast',
'72nd Court Northeast',
'72nd Place Northeast',
'73rd Avenue Northeast',
'73rd Court Northeast',
'73rd Place Northeast',
'74th Avenue Northeast',
'74th Court Northeast',
'74th Place Northeast',
'75th Avenue Northeast',
'76th Avenue Northeast',
'76th Court Northeast',
'76th Place Northeast',
'77th Avenue Northeast',
'77th Place Northeast',

'78th Avenue Northeast',
'78th Place Northeast',
'79th Avenue Northeast',
'79th Court Northeast',
'79th Place Northeast',
'80th Avenue Northeast',
'80th Court Northeast',
'80th Place Northeast',
'81st Avenue Northeast',
'81st Court Northeast',
'81st Lane Northeast',
'81st Place Northeast',
'82nd Avenue Northeast',
'82nd Court Northeast',
'82nd Lane Northeast',
'82nd Place Northeast',
'83rd Avenue Northeast',
'83rd Court Northeast',
'83rd Place Northeast',
'84th Avenue Northeast',
'84th Court Northeast',
'84th Lane Northeast',
'84th Place Northeast',
'85th Avenue Northeast',
'85th Place Northeast',
'86th Avenue Northeast',
'86th Place Northeast',
'87th Avenue Northeast',
'87th Court Northeast',
'87th Place Northeast',
'88th Avenue Northeast',
'88th Court Northeast',
'88th Place Northeast',
'89th Avenue Northeast',
'89th Court Northeast',
'89th Place Northeast',
'8th Avenue Northeast',
'8th Court Northeast',
'90th Avenue Northeast',
'90th Court Northeast',
'90th Place Northeast',
'91st Avenue Northeast',
'91st Court Northeast',
'91st Place Northeast',
'92nd Avenue Northeast',
'92nd Court Northeast',
'92nd Place Northeast',

```
'92th Avenue Northeast',
'93rd Avenue Northeast',
'93rd Boulevard Northeast',
'93rd Court Northeast',
'93rd Place Northeast',
'94th Avenue Northeast',
'94th Drive Northeast',
'94th Place Northeast',
'95th Avenue Northeast',
'95th Place Northeast',
'96th Avenue Northeast',
'96th Court Northeast',
'96th Drive Northeast',
'97th Avenue Northeast',
'97th Court Northeast',
'98th Avenue Northeast',
'98th Court Northeast',
'99th Avenue Northeast',
'99th Court Northeast',
'99th Place Northeast',
'Avondale Road Northeast',
'Ballinger Way Northeast',
'Bear Creek Road Northeast',
'Bothell Way Northeast',
'Brickyard Road Northeast',
'Brookside Boulevard Northeast',
'Campus Way Northeast',
'Erickson Place Northeast',
'Garden Way Northeast',
'Hollyhills Drive Northeast',
'Holmes Point Drive Northeast',
'Inglewood Lane Northeast',
'Inglewood Place Northeast',
'Inglewood Road Northeast',
'Inglewood Terrace Northeast',
'Juanita Drive Northeast',
'Juanita Woodinville Way Northeast',
'Juanita-Woodinville Way Northeast',
'Kenlake Place Northeast',
'Lake City Way Northeast',
'Par Place Northeast',
'Redmond-Woodinville Road Northeast',
'Riviera Place Northeast',
'Roosevelt Way Northeast',
'Shore Drive Northeast',
'Simonds Road Northeast',
'Snohomish Woodinville Road Northeast',
```

```
                'Waynita Way Northeast',
                'Westwood Place Northeast',
                'Woodcrest Drive Northeast',
                'Woodinville Redmond Road Northeast',
                'Woodinville Snohomish Road Northeast',
                'Woodinville-Redmond Road Northeast',
                'Woodinville-Redmond Road Northeast;Woodinville Redmond Road
Northeast'},
                'Northwest': {'Juanita Drive Northwest'},
                'Pl': {'N 202nd Pl'},
                'SE': {'17th Ave SE',
                 '208th St SE',
                 '27th Ave SE',
                 '37th DR SE',
                 '39th Dr. SE'},
                'SW': {'196th St SW', '196th Street SW', '236th St SW'},
                'South': {'Meridian Avenue South', 'North Creek Parkway South'},
                'Southeast': {'10th Avenue Southeast',
                 '10th Court Southeast',
                 '10th Drive Southeast',
                 '11th Court Southeast',
                 '11th Drive Southeast',
                 '121st Avenue Southeast',
                 '12th Drive Southeast',
                 '13th Avenue Southeast',
                 '13th Drive Southeast',
                 '14th Drive Southeast',
                 '15th Avenue Southeast',
                 '15th Place Southeast',
                 '16th Avenue Southeast',
                 '16th Drive Southeast',
                 '16th Place Southeast',
                 '17th Avenue Southeast',
                 '18th Avenue Southeast',
                 '191st Place Southeast',
                 '192nd Street Southeast',
                 '195th Street Southeast',
                 '196th Street Southeast',
                 '197th Street Southeast',
                 '19th Avenue Southeast',
                 '19th Drive Southeast',
                 '1st Avenue Southeast',
                 '1st Drive Southeast',
                 '201st Place Southeast',
                 '206th Place Southeast',
                 '206th Street Southeast',
                 '207th Place Southeast',
```

```
'208th Place Southeast',
'208th Street Southeast',
'209th Street Southeast',
'20th Avenue Southeast',
'20th Drive Southeast',
'210th Street Southeast',
'211th Place Southeast',
'211th Street Southeast',
'212th Street Southeast',
'213th Street Southeast',
'214th Street Southeast',
'215th Place Southeast',
'215th Street Southeast',
'216th Place Southeast',
'216th Street Southeast',
'217th Place Southeast',
'217th Street Southeast',
'218th Place Southeast',
'218th Street Southeast',
'219th Place Southeast',
'21st Avenue Southeast',
'21st Drive Southeast',
'220th Street Southeast',
'221st Place Southeast',
'221st Street Southeast',
'222nd Street Southeast',
'223rd Place Southeast',
'223rd Street Southeast',
'224th Place Southeast',
'224th Street Southeast',
'225th Place Southeast',
'225th Street Southeast',
'226th Place Southeast',
'226th Street Southeast',
'227th Street Southeast',
'228th Place Southeast',
'228th Street Southeast',
'228th Street Southeast and 45th Avenue Southeast',
'22nd Avenue Southeast',
'22nd Drive Southeast',
'231st Street Southeast',
'232nd Place Southeast',
'232nd Street Southeast',
'233rd Place Southeast',
'233rd Street Southeast',
'234th Place Southeast',
'234th Street Southeast',
```

'235th Place Southeast',
'235th Street Southeast',
'236 Street Southeast',
'236th Place Southeast',
'236th Street Southeast',
'237th Place Southeast',
'237th Street Southeast',
'238th Place Southeast',
'238th Street Southeast',
'239th Place Southeast',
'239th Street Southeast',
'23rd Avenue Southeast',
'23rd Drive Southeast',
'240th Place Southeast',
'240th Street Southeast',
'241st Place Southeast',
'241st Street Southeast',
'242nd Place Southeast',
'242nd Street Southeast',
'243rd Avenue Southeast',
'243rd Place Southeast',
'244th Street Southeast',
'24th Avenue Southeast',
'24th Drive Southeast',
'25th Avenue Southeast',
'25th Drive Southeast',
'26th Avenue Southeast',
'26th Drive Southeast',
'27th Avenue Southeast',
'27th Drive Southeast',
'28th Avenue Southeast',
'29th Avenue Southeast',
'29th Drive Southeast',
'2nd Avenue Southeast',
'2nd Drive Southeast',
'2nd Place Southeast',
'30th Avenue Southeast',
'30th Court Southeast',
'30th Drive Southeast',
'31st Avenue Southeast',
'31st Drive Southeast',
'32nd Avenue Southeast',
'32nd Drive Southeast',
'33rd Avenue Southeast',
'33rd Drive Southeast',
'34th Avenue Southeast',
'34th Drive Southeast',

```
 '35th Avenue Southeast',
 '38th Drive Southeast',
 '39th Avenue Southeast',
 '39th Court Southeast',
 '39th Lane Southeast',
 '3rd Avenue Southeast',
 '3rd Drive Southeast',
 '40th Avenue Southeast',
 '40th Drive Southeast',
 '40th Lane Southeast',
 '41st Avenue Southeast',
 '41st Drive Southeast',
 '42nd Avenue Southeast',
 '44th Avenue Southeast',
 '44th Drive Southeast',
 '45th Avenue Southeast',
 '47th Avenue Southeast',
 '49th Drive Southeast',
 '4th Avenue Southeast',
 '4th Drive Southeast',
 '51st Avenue Southeast',
 '53rd Avenue Southeast',
 '57th Avenue Southeast',
 '5th Avenue Southeast',
 '5th Drive Southeast',
 '6th Avenue Southeast',
 '6th Drive Southeast',
 '76th Avenue Southeast',
 '7th Avenue Southeast',
 '7th Drive Southeast',
 '86th Avenue Southeast',
 '87th Avenue Southeast',
 '8th Avenue Southeast',
 '8th Drive Southeast',
 '9th Avenue Southeast',
 '9th Court Southeast',
 '9th Drive Southeast',
 '9th Lane Southeast',
 'Bothell-Everett Highway Southeast',
 'Meridian Drive Southeast',
 'State Highway 9 Southeast',
 'state Highway 9 Southeast'},
'Southwest': {'192nd Place Southwest',
 '193rd Place Southwest',
 '194th Place Southwest',
 '194th Street Southwest',
 '195th Street Southwest',
```

```
'196th Place Southwest',
'196th Street Southwest',
'198th Street Southwest',
'199th Street Southwest',
'200th Street Southwest',
'201st Street Southwest',
'202nd Place Southwest',
'202nd Street Southwest',
'203rd Street Southwest',
'204th Place Southwest',
'205th Place Southwest',
'206th Place Southwest',
'206th Street Southwest',
'207th Place Southwest',
'207th Street Southwest',
'208th Place Southwest',
'208th Street Southwest',
'209th Street Southwest',
'210th Street Southwest',
'211th Place Southwest',
'211th Street Southwest',
'212th Street Southwest',
'213th Place Southwest',
'214th Ave Southwest',
'215th Street Southwest',
'216th Street Southwest',
'217th Place Southwest',
'217th Street Southwest',
'218th Place Southwest',
'218th Street Southwest',
'219th Place Southwest',
'220th Street Southwest',
'221st Place Southwest',
'221st Street Southwest',
'222nd Street Southwest',
'223rd Street Southwest',
'224th Place Southwest',
'224th Street Southwest',
'225th Place Southwest',
'225th Street Southwest',
'226th Place Southwest',
'226th Street Southwest',
'227th Place Southwest',
'228th Street Southwest',
'229th Place Southwest',
'229th Street Southwest',
'230th Street Southwest',
```

```
    '232nd Place Southwest',
    '232nd Street Southwest',
    '233rd Place Southwest',
    '233rd Street Southwest',
    '234th Place Southwest',
    '234th Street Southwest',
    '235th Place Southwest',
    '236th Place Southwest',
    '236th Street Southwest',
    '237th Place Southwest',
    '237th Street Southwest',
    '238th Place Southwest',
    '238th Street Southwest',
    '239th Place Southwest',
    '239th Street Southwest',
    '240th Place Southwest',
    '240th Street Southwest',
    '241st Street Southwest',
    '242nd Place Southwest',
    '242nd Street Southwest',
    '243rd Place Southwest',
    '244th Street Southwest',
    'North 205th Street - 244th Street Southwest',
    'Street Southwest'},
 'Sport': {'Woodinville Ski and Sport'},
 'St': {'190th St',
    'Main St',
    'NE 145th St',
    'NE 158th St',
    'NE 185th St',
    'NE 186th St',
    'NE 205th St'},
 'St.': {'NE 145th St.'},
 'W': {'68th Ave W'},
 'WA-99': {'WA-99'},
 'West': {'12th Avenue West',
    '1st Avenue West',
    '1st Place West',
    '2nd Avenue West',
    '2nd Place West',
    '2st Avenue West',
    '31st Place West',
    '36th Avenue West',
    '3rd Place West',
    '40th Avenue West',
    '44th Avenue West',
    '48th Avenue West',
```

```
'4th Avenue West',
'4th Place West',
'52nd Avenue West',
'53rd Avenue West',
'54th Avenue West',
'55th Avenue West',
'56th Avenue West',
'57th Avenue West',
'57th Place West',
'58th Avenue West',
'58th Place West',
'59th Avenue West',
'59th Place West',
'5th Avenue West',
'5th Place West',
'60th Avenue West',
'61st Avenue West',
'61st Place West',
'63rd Avenue West',
'64th Avenue West',
'66th Avenue West',
'67th Avenue West',
'68th Avenue West',
'68th Place West',
'6th Avenue West',
'6th Place West',
'70th Avenue West',
'72nd Avenue West',
'72nd Place West',
'73rd Place West',
'74th Avenue West',
'74th Place West',
'75th Avenue West',
'75th Place West',
'76th Avenue West',
'76th Place West',
'77th Avenue West',
'77th Place West',
'78th Avenue West',
'78th Place West',
'79th Avenue West',
'79th Lane West',
'79th Place West',
'7th Avenue West',
'7th Place West',
'80th Avenue West',
'80th Court West',
```

```
        '80th Lane West',
        '80th Place West',
        '8th Place West',
        '9th Avenue West',
        '9th Place West',
        'Meridian Place West'},
    'sw': {'196th st sw'}})
```

To fix the abbreviations in the street names, to update with full names, a mapping dictionary can be used:

```
[ ]: mapping = { "Ave": "Avenue",
                 "Ave.": "Avenue",
                 "avenue": "Avenue",
                 "ave": "Avenue",
                 "Blvd": "Boulevard",
                 "Blvd.": "Boulevard",
                 "Blvd,": "Boulevard",
                 "Boulavard": "Boulevard",
                 "Boulvard": "Boulevard",
                 "Ct": "Court",
                 "Dr": "Drive",
                 "Dr.": "Drive",
                 "E": "East",
                 "Hwy": "Highway",
                 "Ln": "Lane",
                 "Ln.": "Lane",
                 "N": "North",
                 "Pl": "Place",
                 "Plz": "Plaza",
                 "Rd": "Road",
                 "Rd.": "Road",
                 "S:": "South",
                 "St": "Street",
                 "St.": "Street",
                 "st": "Street",
                 "street": "Street",
                 "square": "Square",
                 "parkway": "Parkway",
                 "W": "West",
                 "NW": "Northwest",
                 "NE": "Northeast",
                 "SW": "Southwest",
                 "sw": "Southwest",
                 "SE": "Southeast",
                 "state": "State",
                 "99": "Highway 99",
```

```
                "WA-99":"Highway 99"
                }
```

```
[33]: def fix_street(osmfile):
          st_types = audit.audit(osmfile)
          for st_type, ways in st_types.iteritems():
              for name in ways:
                  if st_type in audit.mapping:
                      better_name = name.replace(st_type, audit.mapping[st_type])
                      print (name, "=>", better_name)

      fix_street(osm_file)
```

```
('NE 145th St.', '=>', 'NE 145th Street')
('N 202nd Pl', '=>', 'N 202nd Place')
('25th Ave NE', '=>', '25th Ave Northeast')
('Ballinger Way NE', '=>', 'Ballinger Way Northeast')
('Juanita-Woodinville Way NE', '=>', 'Juanita-Woodinville Way Northeast')
('19115 112th Ave NE', '=>', '19115 112th Ave Northeast')
('5th Ave NE', '=>', '5th Ave Northeast')
('94th Pl NE', '=>', '94th Pl Northeast')
('104th Ave NE', '=>', '104th Ave Northeast')
('141st Pl NE', '=>', '141st Pl Northeast')
('68th Ave NE', '=>', '68th Ave Northeast')
('138th Way NE', '=>', '138th Way Northeast')
('Bothell Way NE', '=>', 'Bothell Way Northeast')
('120th Avenue NE', '=>', '120th Avenue Northeast')
('127th Ave NE', '=>', '127th Ave Northeast')
('Bothell-Everett Hwy', '=>', 'Bothell-Everett Highway')
('196th St SW', '=>', '196th St Southwest')
('236th St SW', '=>', '236th St Southwest')
('196th Street SW', '=>', '196th Street Southwest')
('Martin Way E', '=>', 'Martin Way East')
('NE 185th St', '=>', 'NE 185th Street')
('NE 205th St', '=>', 'NE 205th Street')
('190th St', '=>', '190th Street')
('NE 158th St', '=>', 'NE 158th Street')
('NE 186th St', '=>', 'NE 186th Street')
('NE 145th St', '=>', 'NE 145th Street')
('Main St', '=>', 'Main Street')
('68th Ave W', '=>', '68th Ave West')
('196th st sw', '=>', '196th st Southwest')
('Alderwood Mall Blvd', '=>', 'Alderwood Mall Boulevard')
('44th Ave', '=>', '44th Avenue')
('208th St SE', '=>', '208th St Southeast')
('37th DR SE', '=>', '37th DR Southeast')
('17th Ave SE', '=>', '17th Ave Southeast')
('27th Ave SE', '=>', '27th Ave Southeast')
```

```
('39th Dr. SE', '=>', '39th Dr. Southeast')
```

##

Data Overview

Basic Statistics of the Data: Bothell.osm 141,444 KB Bothell.db 75,838 KB nodes.csv 54,103 KB nodes_tags.csv 2,134 KB ways.csv 5,083 KB ways_nodes.csv 17,096 KB ways_tags.csv 10,487 KB

0.0.4 Number of Nodes

```
[40]: query =  """SELECT COUNT(*) FROM nodes; """
```

output: 607354

0.0.5 Number of Ways

```
[41]: query = '''SELECT COUNT(*) FROM ways; '''
```

output: 79027

0.0.6 Most Common Way

```
[43]: query = '''SELECT key, count(*)
             FROM ways_tags
             GROUP BY 1
             ORDER BY count(*) DESC
             LIMIT 1;
          '''
```

output: building

0.0.7 Number of Unique Users

```
[44]: query = '''SELECT COUNT(distinct(uid))
             FROM (SELECT uid FROM nodes UNION ALL SELECT uid FROM ways);
          '''
```

output: 1276

0.0.8 Number of Unique Bars

```
[45]: query = '''SELECT COUNT(distinct id) FROM nodes_tags WHERE value="bar"; '''
```

output: 8

0.0.9 Number of Unique Schools

```
[46]: query = '''SELECT COUNT(*) FROM nodes_tags WHERE value="school"; '''
```

output = 37

0.0.10 Top Contributer

```
[48]: query = '''SELECT user, COUNT(*) as num
            FROM (SELECT user FROM nodes UNION ALL SELECT user FROM ways) user
            GROUP BY user
            ORDER BY num DESC
            LIMIT 1)
        '''
```

output: 'patricknoll_import', 238998

0.0.11 Most Popular Religion

```
[49]: query = '''SELECT nodes_tags.value, COUNT(*) as num
            FROM nodes_tags
            JOIN (SELECT DISTINCT(id) FROM nodes_tags WHERE␣
    ↪value="place_of_worship") i
            ON nodes_tags.id=i.id
            WHERE nodes_tags.key="religion"
            GROUP BY nodes_tags.value
            ORDER BY num DESC
            LIMIT 1
        '''
```

output: 'christian', 14

0.0.12 Most Popular Amenity

```
[51]: query = '''SELECT value, COUNT(distinct id) as num
            FROM nodes_tags
            WHERE key="amenity"
            GROUP BY value
            ORDER BY num DESC
            LIMIT 1
        '''
```

output: 'restaurant', 149

##

Additional Ideas

Create a standard format for street addresses by autocorrecting abbreviations or allowing the user to choose from a drop down list of options when submitting address

The issue for creating a stricter format for inputing contributions would be that it would cost time and money for developement of an updated website interface for doing so.

Allow data from outside sources to be imported to expand the information on amentities, attractions, etc.

One issue that would occur from allowing outside sources would be matching up the data and making sure the format is inline with the current data