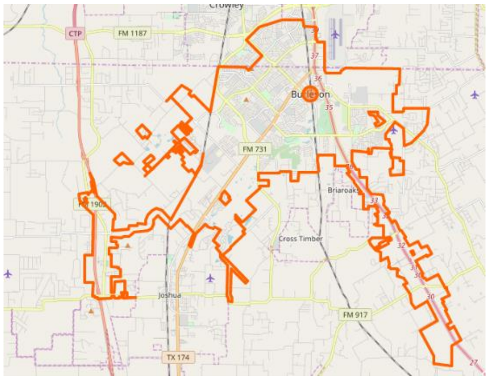
OpenStreet Map data for Burleson, Texas

https://www.openstreetmap.org/relation/115241

The map area for Burleson, Texas has been chosen from the OpenStreetMap Data in order to perform data auditing and data wrangling. I chose Burleson because I live there. The implementation was done in Python and the results then were then uploaded into a database using SQL. The city limits of Burleson are irregular at best.



Initial data analysis of Burleson.osm reveals the following characteristics about the data:

```
'bounds': 1,
 'member': 5248,
 'meta': 1,
 'nd': 390339,
 'node': 328583,
 'note': 1,
 'osm': 1,
 'relation': 110,
 'tag': 96898,
 'way': 58397}
```

With the original .osm file size approx. 74 MB, I created a burlesonsample.osm with a size of 15 MB, which is much easier to work with.

Auditing burlesonsample.osm shows there are few items that need to be cleaned up. I suspect this is because Burleson, itself, is a smaller city.

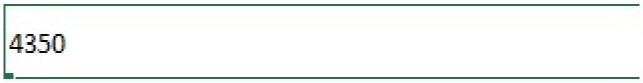
Problems Encountered in the Map

Child Tag Keys

Both individual users and government entities enter information to update this map. As expected, the government data such as, Topologically Integrated Geographic Encoding and Referencing system (TIGER) data, is much more reliable than individual users' data. The tiger tags are only 6.62% of the total tags.

Number of Tiger tags

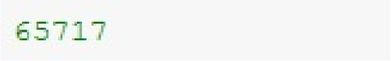
```
sqlite> SELECT COUNT(*)
        FROM (
            SELECT * FROM nodes_tags UNION ALL
            SELECT * FROM ways_tags UNION ALL
            SELECT * FROM relations_tags) e
            WHERE e.type = 'tiger';
```



```
4350
```

Total Number of tags

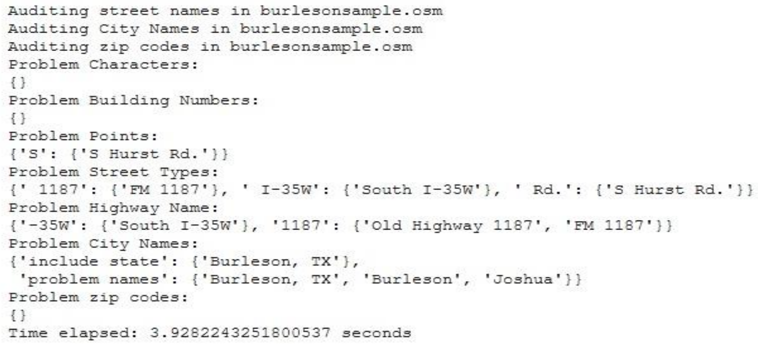
```
sqlite> SELECT COUNT(*)
        FROM (
            SELECT * FROM nodes_tags UNION ALL
            SELECT * FROM ways_tags UNION ALL
            SELECT * FROM relations_tags) e;
```



```
65717
```

Auditing:

After auditing burlesonsample.osm using audit.py, I found the following problems with the following data.



```
Auditing street names in burlesonsample.osm
Auditing City Names in burlesonsample.osm
Auditing zip codes in burlesonsample.osm
Problem Characters:
{}
Problem Building Numbers:
{}
Problem Points:
{'S': {'S Hurst Rd.'}}
Problem Street Types:
{' 1187': {'FM 1187'}, ' I-35W': {'South I-35W'}, ' Rd.': {'S Hurst Rd.'}}
Problem Highway Name:
{'-35W': {'South I-35W'}, '1187': {'Old Highway 1187', 'FM 1187'}}
Problem City Names:
{'include state': {'Burleson, TX'},
 'problem names': {'Burleson, TX', 'Burleson', 'Joshua'}}
Problem zip codes:
{}
Time elapsed: 3.9282243251800537 seconds
```

After auditing burlesonsample.osm using audit.py, I found the following problems with street name:

Problematic Points:
S- This is often used as an abbreviation for the direction of South. The S will be replaced with the word South.

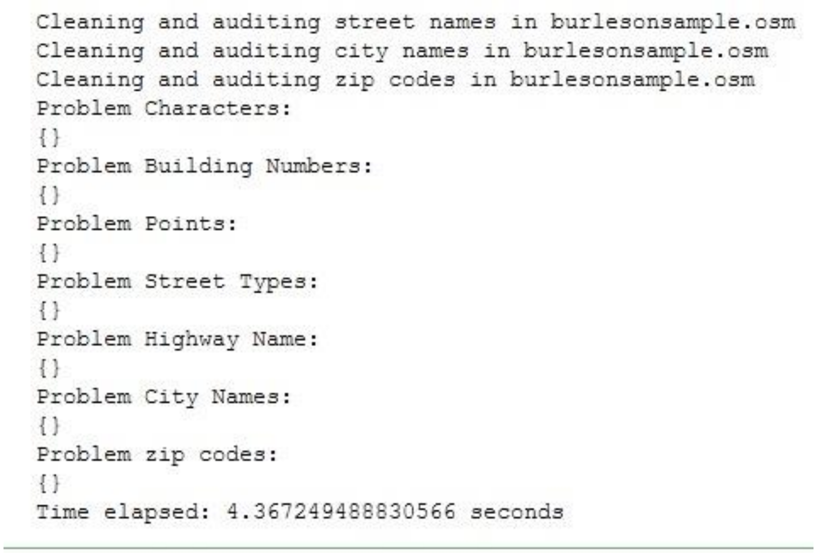
Problematic Street Types:
The streets lack consistency and some are not complete. The solution is set a standard naming solution and apply it to each street type.

Problematic Highway Names:
Similar to the street types, the highway names lack consistency and some are not complete. The solution is set a standard naming solution and apply it to each street type.

Problematic City Names:
The city names are to include the state, some do, some don’t. The solution is to add the same state abbreviation to each city name that is missing the state.

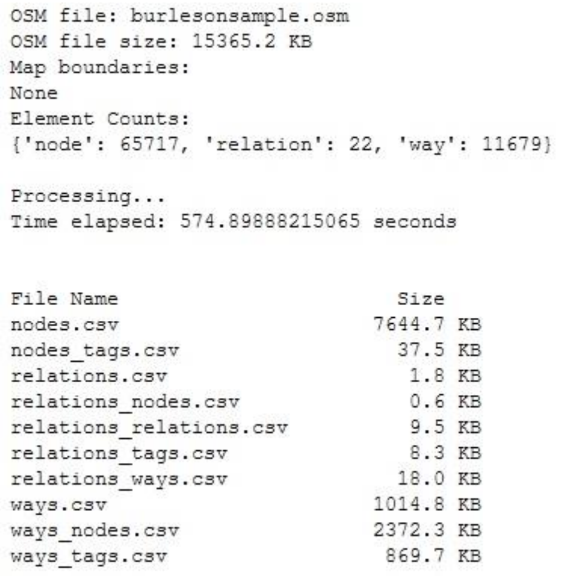
Cleaning:

Cleaning burlesonsample.osm takes just a little more time than auditing the data.



```
Cleaning and auditing street names in burlesonsample.osm
Cleaning and auditing city names in burlesonsample.osm
Cleaning and auditing zip codes in burlesonsample.osm
Problem Characters:
{}
Problem Building Numbers:
{}
Problem Points:
{}
Problem Street Types:
{}
Problem Highway Name:
{}
Problem City Names:
{}
Problem zip codes:
{}
Time elapsed: 4.367249488830566 seconds
```

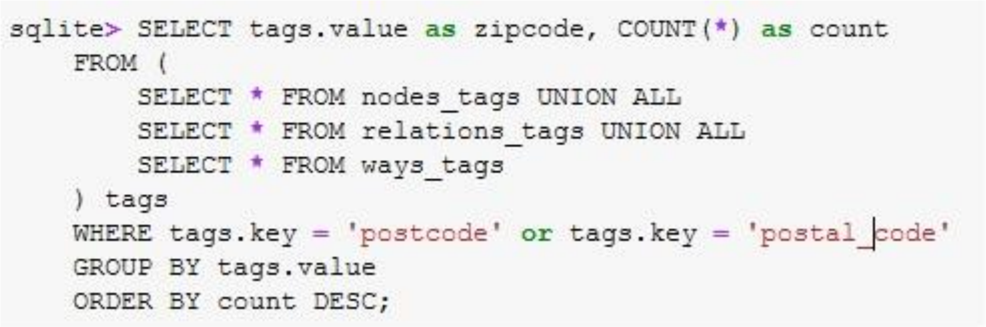
After cleaning the burlesonsample.osm file, I was able to create data files and gather their files sizes:



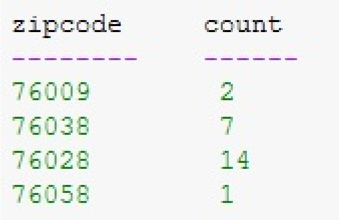
```
OSM file: burlesonsample.osm
OSM file size: 15365.2 KB
Map boundaries:
None
Element Counts:
{'node': 65717, 'relation': 22, 'way': 11679}

Processing...
Time elapsed: 574.89888215065 seconds


File Name                        Size
nodes.csv                      7644.7 KB
nodes_tags.csv                   37.5 KB
relations.csv                     1.8 KB
relations_nodes.csv               0.6 KB
relations_relations.csv           9.5 KB
relations_tags.csv                8.3 KB
relations_ways.csv               18.0 KB
ways.csv                       1014.8 KB
ways_nodes.csv                 2372.3 KB
ways_tags.csv                   869.7 KB
```

Once the cleaned Burleson data is loaded into the database we can gather and consolidate data into useful information.

The count of postal codes within the map area:

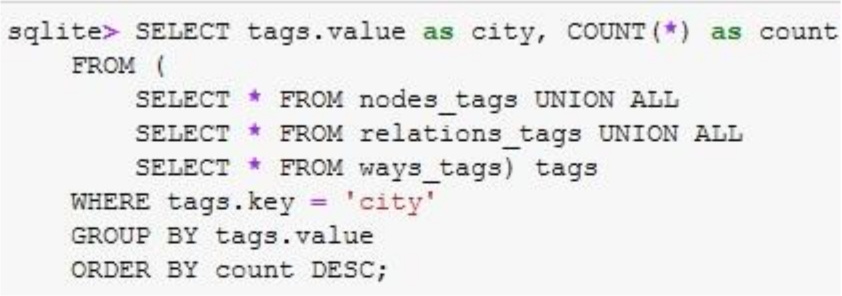


```
sqlite> SELECT tags.value as zipcode, COUNT(*) as count
    FROM (
        SELECT * FROM nodes_tags UNION ALL
        SELECT * FROM relations_tags UNION ALL
        SELECT * FROM ways_tags
    ) tags
    WHERE tags.key = 'postcode' or tags.key = 'postal_code'
    GROUP BY tags.value
    ORDER BY count DESC;
```

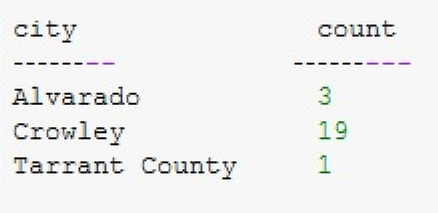


```
zipcode     count
--------    ------
76009        2
76038        7
76028        14
76058        1
```

The count of cities within the map area:



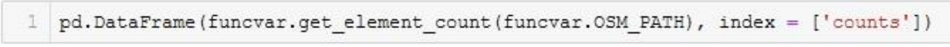
```
sqlite> SELECT tags.value as city, COUNT(*) as count
    FROM (
        SELECT * FROM nodes_tags UNION ALL
        SELECT * FROM relations_tags UNION ALL
        SELECT * FROM ways_tags) tags
    WHERE tags.key = 'city'
    GROUP BY tags.value
    ORDER BY count DESC;
```



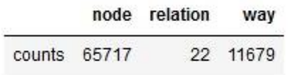
```
city                count
--------          ---------
Alvarado            3
Crowley             19
Tarrant County      1
```

We can count the tags as well:

Burlesonsample.osm element counts:

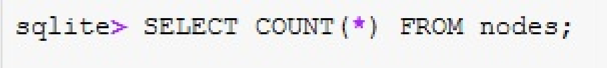


```
pd.DataFrame(funcvar.get_element_count(funcvar.OSM_PATH), index = ['counts'])
```

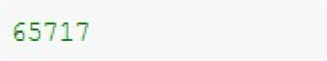


	node	relation	way
counts	65717	22	11679

Total Node tags:

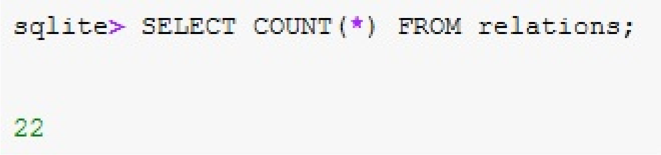


```
sqlite> SELECT COUNT(*) FROM nodes;
```



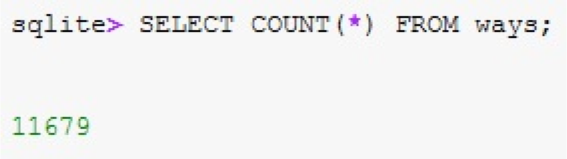
```
65717
```

Total Relations tags:



```
sqlite> SELECT COUNT(*) FROM relations;


22
```

Total Way tags:



```
sqlite> SELECT COUNT(*) FROM ways;


11679
```

The number or elements in the .osm file matches the database elements which indicates that all elements were imported into the database.

Other Ideas About the Dataset

Datasets are only as good as the data within them. Have multiple sources of data without any standard makes the data hard to turn into information that can be used by the masses. If the programmers for Openstreet could design a template that forces the use of a standard data format, the data would be consistent and more reliable.

Conclusion

In this project the dataset obtained for Burleson from openmaps was audited and the result was imported to a databse using SQL. The data set has a lot of potential given that it is created by users but this also introduces imperfections as each user is not following a specific guideline in uploading their data. If there is a template to update data, consistency would greatly increase. If Openstreet could simplify the method to update maps, more users would participate. Burleson is a smaller community but is growing quickly. The rate of growth can make the Openstreet data obsolete in a matter of months.

Appendix

Output for auditing the street names

```
Memorial Plaza -> Memorial Plaza

Bear Plaza -> Bear Plaza

S. Freeway Suite 150 -> S. Freeway Suite 150
```

```
Farm-to-Market Road 1902 -> Farm-to-Market Road 1902

John Jones -> John Jones

NW John Jones Dr -> NW John Jones Drive

County Road 806 -> County Road 806

W Bufford St -> W Bufford Street

SW Wilshire Blvd -> SW Wilshire Boulevard

Burleson Blvd -> Burleson Boulevard

South Freeway -> South Freeway

FM 1187 -> FM 1187

Business 1187 -> Business 1187

Old Highway 1187 -> Old Highway 1187

Oak Grove Road South -> Oak Grove Road South

S Hurst Rd. -> S Hurst Rd.

Stillglen Trace -> Stillglen Trace

Hunters Mill Trace -> Hunters Mill Trace

Meadow Rose Trace -> Meadow Rose Trace

Silvermist Trace -> Silvermist Trace

Shadow Wood Trace -> Shadow Wood Trace

Wildcat Way Street North -> Wildcat Way Street North

FM731 -> FM731

County Road 707 -> County Road 707

South I-35W -> South I-35W

{'S': {'S Hurst Rd.'}, 'S.': {'S. Freeway Suite 150'}, 'W': {'W Bufford St'}}

S. Freeway Suite 150 -> S. Freeway Suite 150 -> South Freeway Suite 150

W Bufford St -> W Bufford Street -> West Bufford Street

S Hurst Rd. -> S Hurst Rd. -> South Hurst Rd.
```