OpenStreetMap Case Study

Cassandra Lamendola

Portland, Oregon, U.S.

- openstreetmap.org
- mapzen.com

Portland, Oregon is one of my favorite cities. I visit there at least once a year because I love the culture of the city. That's why I chose to use the Portland OpenStreetMap data for this case study.

Problems Encountered in the Map

- Abbreviated street types ('Ave.', 'St.', 'Blvd.', etc.)
- Abbreviated street directions (S, NE, W)
- Inconsistency in zip codes (some include 4 digit secondary zip code, some zip codes are outside of Portland)
- Unexpected keys in way tag attributes (brand:wikidata, contact:twitter, tiger:source, tiger:upload_uid, tiger:name_type, fixme, FIXME:access)

Some solutions

Abbreviated street types

To correct the abbreviated street types, I used the regular expression from an example in the SQL lessons. This is the function from my PortlandCaseStudy.py file that I wrote to do the correction:

```
def update_name(name, mapping):
    name = name.split(" ")
    old_street_type = name.pop()
    if old_street_type in mapping:
        name.append(mapping[old_street_type])
    else:
        name.append(old_street_type)
    name = " ".join(name)
    return name
```

Abbreviated street directions

In order to correct the abbreviated street directions, I decided to update the function I had written for correcting street types so that it could take care of both jobs. I also updated the mapping dictionary to include possible directions. Here is the updated function from PortlandCaseStudy.py:

```
def update_name(name, mapping):
    name = name.split(" ")
    for word in range(len(name)):
        if name[word] in mapping.keys():
            name[word] = mapping[name[word]]
    name = " ".join(name)
    return name
```

Overview of the data

Number of unique users

```
SELECT COUNT(DISTINCT(subq.uid))
FROM (
    SELECT uid
    FROM nodes
    UNION
    SELECT uid
    FROM ways)
AS subq;
```

Number of nodes

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

Number of ways

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

Number of cafes

```
SELECT COUNT(*)
FROM nodes_tags
WHERE value = 'cafe'
OR value = 'coffee_shop';
```

Number of bus stops

```
SELECT COUNT(*)
FROM nodes_tags
WHERE value = 'bus_stop';
```

Find the top 5 amenities in Portland

```
SELECT value, COUNT(*) AS num
FROM nodes_tags
WHERE key = 'amenity'
GROUP BY value
ORDER BY num DESC
LIMIT 5;
```

Other ideas about the database

Improving the database further

One way the data could be improved would be to have more consistency in naming amenities. I noticed, for example, that 'cafe' and 'coffee_shop' were both used. I'm sure this problem occurred in other categories as well.

Number of streets with a dedicated bike path

I've heard that Portland is a very bicycle-friendly city. This makes me wonder how many streets in Portland have a dedicated bike path.

```
SELECT COUNT(*)
FROM (
    SELECT *
    FROM (ways_tags JOIN ways
        ON ways_tags.id = ways.id)
    WHERE ways_tags.key = 'bicycle'
    AND ways_tags.value = 'yes'
    OR ways_tags.value = 'designated');
```

I would like to know what percentage of non-highway streets have bike paths.

```
SELECT COUNT(DISTINCT(subq.id))
FROM (
    SELECT *
    FROM ways JOIN ways_tags
    ON ways.id = ways_tags.id
    WHERE ways_tags.key != 'highway')
AS subq;
```

That means there are 39,323 ways containting bike paths out of 807,563 total ways. So almost 5% of streets in Portland have a designated bike path. I would like to see how this compares with other cities in the US.