OpenStreetMap Data Case Study

I will preface this with saying that this project and I did not get along. I spent quite a lot of time reading the module documentations and scouring Stack Overflow and GitHub. Unfortunately, in my frustration I did not note which projects I used code from. I've included the websites I did remember to note.

Map Area

[Salt Lake County](#)

I picked this because this is where I grew up and I was familiar with the location.

Problem Data

Salt Lake County uses the grid system instead of normal street names. This means that most of our addresses end in either North, South, East or West. There were several different abbreviations of this throughout the data set. I programmatically cleaned this up and made it uniform, along with what few real street names we have.

Another problem I had was that when I loaded my csvs into Sqlite, it would merge the last column header with the word TEXT. I have yet to figure out why it's doing that, but none of my queries use those last columns so I took that off my list to figure out later.

Data Overview

File Sizes

```
Salt_Lake_County.osm ---------------------- 252 MB
Sample.osm --------------------------------- 2.54 MB
Sale_Lake_County.db ----------------------- 187 MB
nodes.csv --------------------------------- 97.5 MB
nodes_tags.csv ---------------------------- 2.90 MB
ways.csv ---------------------------------- 8.50 MB
ways_nodes.csv ---------------------------- 32.8 MB
ways_tags.csv ----------------------------- 17.60 MB
```

Query Results

Node Count	Ways Count	UID Count
1206794	147098	1157

Top Ten Users			
Chadbunn ------------------------------------	246434	crossing--	78545
Osmjwh ---------------------------------------	153909	parking ---	2320
Butlerm ---------------------------------------	79739	place_of_worship-------------------------------	454
mash84121-----------------------------------	67147	school--	390
mvexel--	62890	restaurant --------------------------------------	154
woodpeck_fixbot ---------------------------	60044	fast_food--	152
MelanieOriet--------------------------------	58971	shelter --	92
Level ---	49020	fuel ---	68
wrk3 --	40727	bank--	67
Ted Percival --------------------------------	23355	toilets --	391

Top Ten Amenities

The first few queries were self-explanatory and done to see if I could do the queries correctly. The queries I was really interested in were Religions and Restaurants (see below). The reason for these is because Utah doesn't have the most diverse population and I was curious about how it would show up with these queries. The Religions query was disappointing because it lumped all the Christianity churches together. I am more interested in the distribution of the different denominations. The Restaurants one was done because I like food and was curious about what was available. One thing I noticed is that there's no standardization for the types of restaurants. I believe it would benefit the dataset to have these standardized upon submission to OpenStreetMap. The only issue with that is when you have those unique restaurants that are regional specific, as in the El Salvadoran restaurant.

Religions
Christian ------------------------------------ 595
Buddhist ------------------------------------ 1
Jewish --------------------------------------- 1
Muslim -------------------------------------- 1
Scientologist -------------------------------- 1
unitarian_universalist ---------------------- 1

Restaurants
American -------------------------------- 134
Mexican --------------------------------- 103
Pizza ------------------------------------- 95
Sandwich -------------------------------- 51
Chinese----------------------------------- 47
Italian------------------------------------ 41
Burger ----------------------------------- 39
Asian------------------------------------- 31
Sushi ------------------------------------- 23
International------------------------------ 16
steak_house ----------------------------- 16
indian------------------------------------ 14
Japanese--------------------------------- 13
Thai -------------------------------------- 12
Greek ------------------------------------ 10
seafood ---------------------------------- 8
ice_cream -------------------------------- 7
regional --------------------------------- 6
American -------------------------------- 5
French ----------------------------------- 4
Barbecue--------------------------------- 3
Chicken----------------------------------- 3
Korean----------------------------------- 3
Vietnamese ------------------------------ 3
Pizza ------------------------------------- 2
Sushi ------------------------------------- 2
Ethiopian -------------------------------- 2
Asian------------------------------------- 1

Diner_(Breakfast_and_Lunch) ------------1
Frozen_Yogurt ----------------------------1
Mexican -----------------------------------1
Regional-----------------------------------1
asian;sushi--------------------------------1
barbecue ----------------------------------1
Brazilian -----------------------------------1
Breakfast-----------------------------------1
burger; ice_cream ------------------------1
burger;33 ----------------------------------1
burgers_and_fries ------------------------1
chinese_&_japanese ----------------------1
coffee_shop -------------------------------1
donuts -------------------------------------1
elsalvadorean -----------------------------1
german -------------------------------------1
gyros ---------------------------------------1
kebab ---------------------------------------1
Lebanese -----------------------------------1
M ---1
Pancake ------------------------------------1
Peruvian------------------------------------1
pizza;Italian-------------------------------1
Spanish -------------------------------------1
steak_&_fish ------------------------------1
vegan ---------------------------------------1
vegetarian ---------------------------------1
yogurt---------------------------------------1

Take Aways

While this project gets an F for frustration, it did pose a few questions that I'm excited about finding the answers to.

1. Why is my schema loading weird?
2. Does this data set break down the different denominations of the religions somewhere?
3. How do we standardize the restaurants?
4. What does Ethiopian food taste like?

Sources Used

http://docs.python-cerberus.org/en/latest/api.html

https://stackoverflow.com/questions/17245415/read-and-write-csv-files-including-unicode-with-python-2-7

https://github.com/jdunck/python-unicodecsv

https://stackoverflow.com/questions/5838605/python-dictwriter-writing-utf-8-encoded-csv-files

https://www.w3schools.com/xml/xml_whatis.asp

https://www.w3schools.com/sql/default.asp

https://conda.io/docs/user-guide/tasks/index.html

https://pandas.pydata.org/pandas-docs/stable/basics.html

https://storage.googleapis.com/supplemental_media/udacityu/5435548805/numpy_pandas_cheatsheet.pdf