Map Area:

Quincy, Massachusetts, United States

<https://www.openstreetmap.org/export#map=12/42.2623/-71.0159>

The reason that I chose this place is that I’ve always admired the founding fathers like Washington, Adams, Jefferson, and especially Jefferson. Seeing that there is no chance that I can get a dataset representing Monticello, instead I chose Quincy, the birthplace of the Adamses.

Problems encountered in the map

1. Invalid tag type and value

As further query shows, there are multiple tags with unmeaningful types and values, like for example, there is a tag with type ‘noref’ and value ‘4’. Without any gold standard data, there is no appropriate method to remedy this problem

1. Several street names begin with ‘s’ instead of ‘south’, with ‘n’ instead of ‘north’

As running several select statements show, there are multiple street names that begin with ‘s’ instead of ‘south’, like ‘S 22nd Street’. The function that I wrote to remedy this problem is presented below:

**def** update(string, mapping):  
 words = string.split()  
 **for** w **in** range(len(words)):  
 **if** words[w] **in** mapping:  
 words[w] = mapping[words[w]]  
 name = **" "**.join(words)  
 **return** name

The function is checking for any inconsistency in the tag values, and replace them with the corresponding correct values before they go into the database.

1. There are several data entries with inconsistent responses, like an office tag with a value of yes, and a religion tag with a value of Asian. I am guessing that it might not be actual data failure, might be because lower level tags overriding the higher level ones, as mentioned by the sample project. But in my case, instead of running python code to clean the data, I chose to run SQL DELETE statements to delete the inconsistent data, since I think it is more succinct in the way.

Overview

1. Running ls -al on the files

drwxr-xr-x 18 peizhizhang staff 576 Jun 26 16:53 .

drwxr-xr-x 12 peizhizhang staff 384 Jun 12 17:58 ..

-rw-r--r--@ 1 peizhizhang staff 6148 Jun 25 16:27 .DS\_Store

drwxr-xr-x 7 peizhizhang staff 224 Jun 25 22:23 .idea

-rw-r--r-- 1 peizhizhang staff 0 Jun 12 15:52 README.md

drwxr-xr-x 3 peizhizhang staff 96 Jun 22 18:17 \_\_pycache\_\_

-rw-r--r--@ 1 peizhizhang staff 64174633 Jun 21 15:38 data.osm

-rw-r--r-- 1 peizhizhang staff 6301 Jun 24 15:10 data.py

-rw-r--r-- 1 peizhizhang staff 0 Jun 26 16:53 file\_sizes.txt

-rw-r--r--@ 1 peizhizhang staff 20156519 Jun 24 15:41 nodes.csv

-rw-r--r--@ 1 peizhizhang staff 879470 Jun 24 15:41 nodes\_tags.csv

-rw-r--r-- 1 peizhizhang staff 33996800 Jun 24 20:51 openstreetmap.db

-rw-r--r--@ 1 peizhizhang staff 12149 Jun 25 22:30 p3.docx

-rw-r--r-- 1 peizhizhang staff 2313 Jun 22 16:54 schema.py

-rw-r--r--@ 1 peizhizhang staff 3254491 Jun 24 15:41 ways.csv

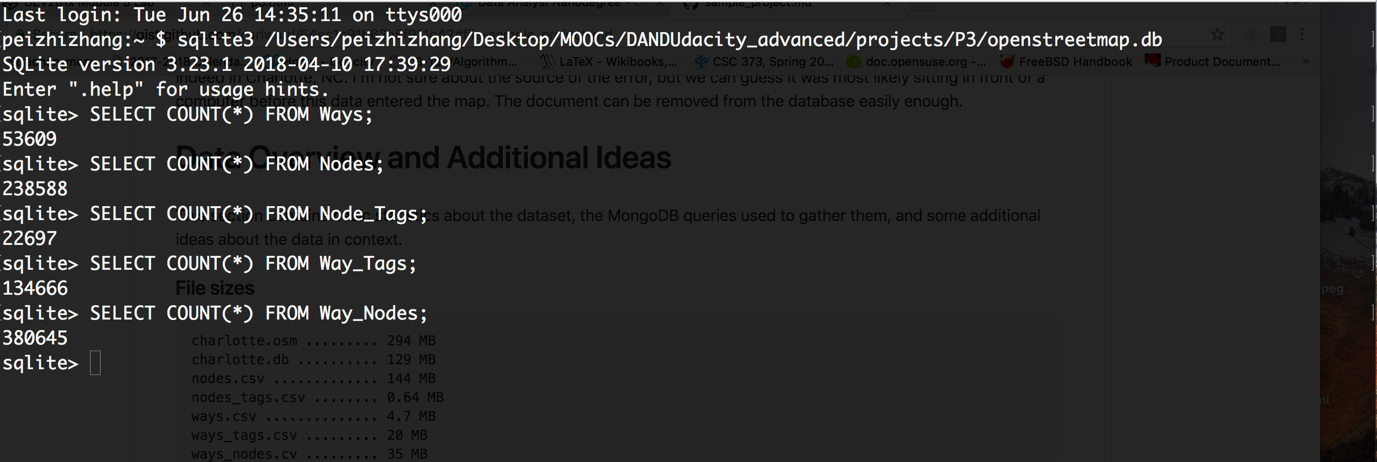
-rw-r--r--@ 1 peizhizhang staff 9101978 Jun 24 15:41 ways\_nodes.csv

-rw-r--r-- 1 peizhizhang staff 4552938 Jun 24 15:41 ways\_tags.csv

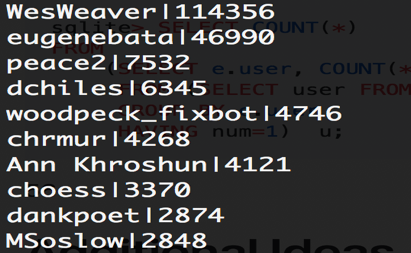
-rw-r--r--@ 1 peizhizhang staff 162 Jun 26 14:27 ~$p3.docx

As shown above, the database file (the one called ‘openstreetmap.db’) is of size 33996800B, or 34.7MB.

2. metadata for the dataset



Top 10 contributing users

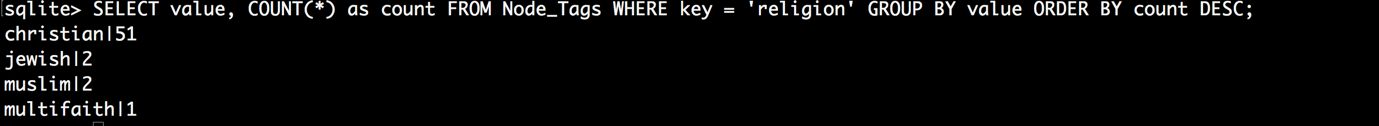


Popular cuisines



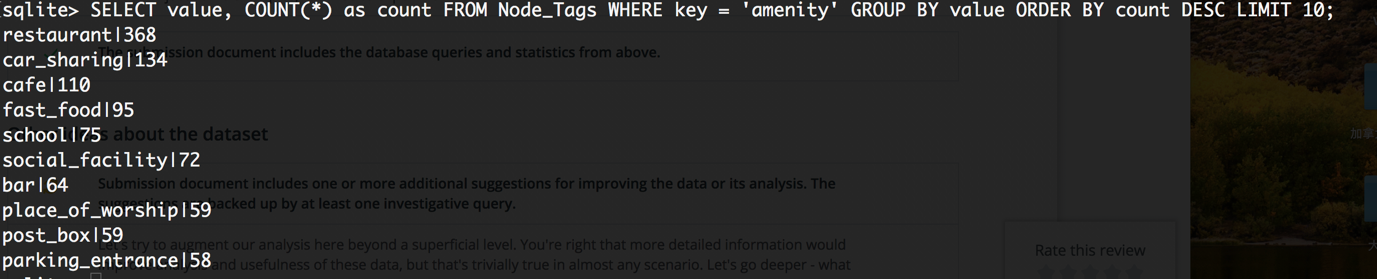
As per the diagram, we know that pizza may be the number one choice for food.

Most popular religion



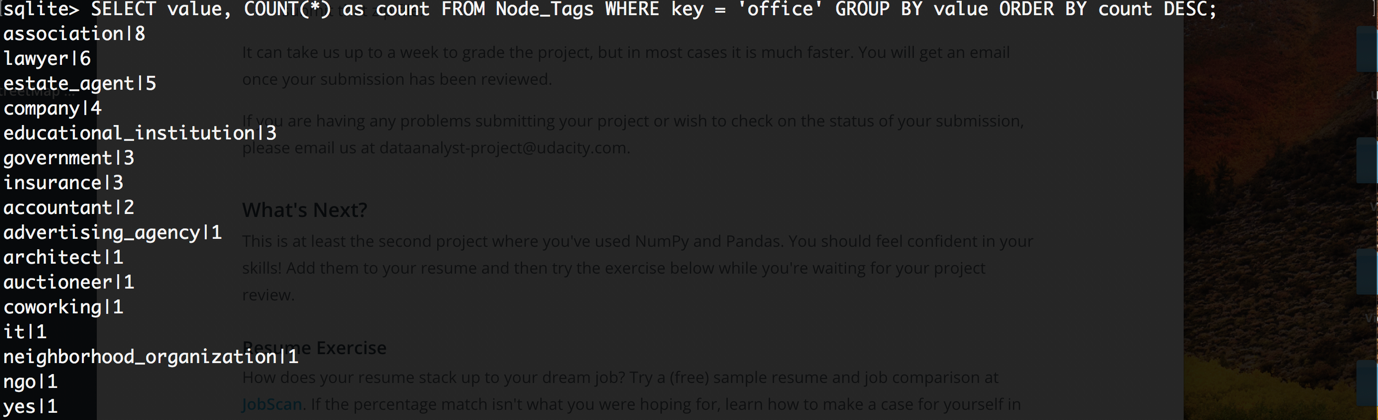
As per the query, the most popular religion, as would be expected, is Christian.

Most popular amenity



As per the diagram, the most popular amenity is restaurant, having 368 occurrences.

Most popular office type



As per the query, we know that the most popular office type is office for associations.

Additional Ideas

1. More detailed info about the data may be required, like for restaurant nodes, a kind of tree-like data structure detailing the land area, the menu, the available space may be constructed in order to better understand and display the information about the node. And on the top level, more detailed user info may be required, instead of a simple username and uid. Thus, the single biggest deficiency of the data, in my opinion, is user demographics.
2. After careful data exploration, I find that top ten contributing users make up about 82.8% of the dataset. Therefore I suggest that some inciting measure must be undertaken in order for more users to contribute. For example, the vendor can place ads on Facebook or Google Maps in order to boost popular familiarity of the API service, and the user experience, as may be noted by many others, can be greatly improved, for example, changing the design of the web page to match Google’s design guidelines can be a huge bonus to the web service. 