**Data Analyst Nanodegree – DAND P3 Project: OpenStreetMap Data Case Study Read Me First File**

By Teresa Aysan – Submitted April 29, 2017

**Read Me First**

1. **Copy** all of the files that I am submitting with this project into the same directory.
2. **Run** the program called "**data.py**".
    a. It calls:
        i. **Sample_100.osm** <in my report, the results are based on the full Nashville_tennessee.osm file>
        ii. **schema**
        iii. **update_name_street** which calls
            1. **myDictionaries**
                a. which were built with the assistance of the **audit** code
    b. It creates the csv files that are necessary for doing the SQL queries:
        i. nodes
        ii. nodes_tags
        iii. ways
        iv. ways_tags
        v. ways_nodes
3. **Copy** the files that you want sized into a sub directory and make note of that sub directory name. You will need the path name later when you run the "P3 SQL Queries.py" program. If you want to find out the sizes of the csv files, then it is important to do this step **after** you run "data.py". The files I wanted sized and that I copied into my sub directory were:
    a. myNashvilleOSMdb.db
    b. nashville_tennessee.osm
    c. nodes.csv
    d. nodes_tags.csv
    e. ways.csv
    f. ways_nodes.csv
    g. ways_tags.csv
4. **Run** the program called "**P3 SQL Queries.py**". Follow the prompts to run all of the queries that I have created for this project.
5. Run the other program files as you see fit:
    a. audit – in addition to assisting to build the myDictionaries, this code was used to help figure out which data needed cleaning, but it is not used in the final code
    b. mapparser – this code is used to show what tags are in the database, how many, and to get the feeling on how much of which data you can expect to have in the map.
    c. tags – this code provides 3 regular expressions to check for certain patterns in the tags.
    d. users – this code outputs the unique users' IDs that have contributed to the map in this particular area.
6. The following files are also attached – just fyi unless otherwise noted. I chose not to use the sample_xxx.osm files because they were too small for my purposes. I chose to use the full nashville_tennessee.osm file for my report:

    a. Read me First DAND P3 Project Teresa Aysan – this pdf file
    b. data_wrangling_schema_raw.sql – the sql version of the schema
    c. sample_100.osm
7. Open the "list of references" file in a text editor t see how much research I did. ☺
8. Open "map information" to see which map I chose and why.
9. **Open** the "**DAND P3 Project Report Teresa Aysan**" pdf file and read my report. ☺
10. Phone me or email me with any questions that you have.
    a. 416-617-4844
    b. My personal email is: teresa@aysan.ca
    c. But you can email me at whatever email you have on file / on my profile
11. The project took me a long time, because I wanted to learn a lot. I enjoyed doing it. I enjoy the rush getting a program to work after figuring it out.
12. I hope you enjoy my report. And thank you for your time.

Teresa Aysan

**APPENDIX: Summary of the collection of the files – per the Submission instructions:**

1. A pdf document containing your answers to the rubric questions. This file documents my data wrangling process:
    a. DAND P3 Project Report Teresa Aysan
2. My Python code I used in auditing and cleaning my dataset for the final project. I did not include lessons quizzes code, although some of my code uses information from the lessons quizzes code:
    a. Used in auditing and cleaning:
        i. audit.py
        ii. update_name_street.py
        iii. myDictionaries.py
        iv. sample_100.osm
        v. mapparser
        vi. tags
        vii. users
    b. Used in / generated by processing the map & used in the queries:
        i. myNashvilleOSMdb.db
        ii. sample_100.osm
        iii. data.py
        iv. schema.py
        v. nodes.csv
        vi. nodes_tags.csv
        vii. ways.csv
        viii. ways_tags.csv
        ix. ways_nodes.csv
        x. P3 SQL Queries.py
    c. A text file containing a link to the map position I wrangled in my project, a short description of the area and a reason for my choice:

    i. map information
d. An .osm file containing a sample part of the map region I used (around 1-10Mb in size).
    i. sample_100.osm (almost 3Mb in size)
e. A text file containing a list of Web sites, books, forums, blog posts, github repositories etc that I referred to or used in this submission.
    i. List of references
f. Other text files:
    i. Read me First DAND P3 Project Teresa Aysan – this file ☺