UDACITY

PROJECT

Wrangle OpenStreetMap Data

A part of the Data Analyst Nanodegree Program

PROJECT REVIEW

CODE REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Meets Specifications

Next, Using Data Visualization for Analytics

Congratulations on completing the course. Next we will be going into the world of R Statistical Language - a very powerful deep analytical and visualization tool that is gaining popularity in the Data Science Field in an exponential rate. It is one of my favorite languages and I hope you enjoy it as well. Good luck!

Code Functionality

Final project code functionality reflects the description in the project document.

Code Readability

Final project code follows an intuitive, easy-to-follow logical structure.

The cleaning steps are a little odd. First the file needs to be converted and then it is cleaned. This was not intuitively obvious. It took me a little while to figure it out. Ideally, a single script should do the entire process and if not that, there should be instructions for the reviewer on exactly how to get the desired output. Please remember this for the future projects.

Final project code that is not intuitively readable is well-documented with comments.

Problems encountered in your map

Student response shows understanding of the process of auditing, and ways to correct or standardize the data, including dealing with problems specific to the location, e.g. related to language or traditional ways of formatting.

An outstanding work demonstrating a variety of auditing and cleaning procedures. I'm glad you went beyond the demonstrations of Lesson 6 and found additional cleaning issues other than street related issues. I especially appreciate the level of thoroughness in documenting the methods.

Some of the problems encountered during data audit are cleaned programmatically.

You have done a great job of demonstrating programmatic cleaning procedures. This is one of the most important advantages computers can give us.

Overview of the data

The OSM XML file is at least 50 MB uncompressed.

Database queries are used to provide a statistical overview of the dataset, like:

- **size of the file**
- **number of unique users**
- **number of nodes and ways**
- **number of chosen type of nodes, like cafes, shops etc.**

Additional statistics not in the list above are computed. For SQL submissions some queries make use of more than one table.

The submission document includes the database queries and statistics from above.

Other ideas about the dataset

Submission document includes one or more additional suggestions for improving the data or its analysis. The suggestions are backed up by at least one investigative query.

Submission document includes thoughtful discussion about the benefits as well as some anticipated problems in implementing the improvement.

Thoroughness and Succinctness of Submission

Submission document is long enough to thoroughly answer the questions asked without giving unnecessary detail. A good general guideline is that the question responses should take about 3-6 pages.

I must say, terrific work on revising the formatting and presentation of this project. The information is concisely provided and presented in an aesthetically pleasing manner without overwhelming the reader with large blocks of code and unnecessary detail. This is a very important step in creating a great Job-Ready portfolio. Well done.

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review

Student FAQ