Data Wrangle OpenstreetMaps Data

Does Not Meet Specifications

Student Notes | Code Review 2 | Project Review

Code Functionality

SPECIFICATION

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

MEETS SPECIFICATION

Reviewer Comments

Well done!

Code Readability

SPECIFICATION

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

MEETS SPECIFICATION

Reviewer Comments

Great job!

SPECIFICATION

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

MEETS SPECIFICATION

Reviewer Comments

Great job! The code is well commented. Just remember for the next submission to include the cleaning procedures in the `shape_element()` as specified in the *Code Review* section.

Problems encountered in your map

SPECIFICATION

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.

DOES NOT MEET SPECIFICATION

Reviewer Comments

There is some really good work here. The project performs some deep and thorough cleaning. However, as per the instructions, cleaning must be done to the XML file before making it a JSON file. Then it needs to be put into the database. Basically, the file needs to be cleaned before put into the database and not cleaned after insertion to database.

Reasoning...

- In real world applications, MongoDBs will be held remotely
- OSM is also help remotely
- It is inefficient to import from OSM just to export to MongoDB and then clean. It is not bandwidth efficient nor time efficient
- The logical step is to download raw file, clean, the submit the cleaned file in one push

SPECIFICATION

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

MEETS SPECIFICATION

Reviewer Comments

The cleaning procedure implemented is outstanding! However, the cleaning was done post processing the data into the Database. Please see appropriate area of the evaluation for suggestions.

Overview of the data

SPECIFICATION

The dataset is at least 50 MB.

MEETS SPECIFICATION

Reviewer Comments

61.82MB

SPECIFICATION

Student response provides a statistical overview of a 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

MEETS SPECIFICATION

Reviewer Comments
Well done!

SPECIFICATION
Student response also includes the MongoDB queries used to obtain the statistics.

MEETS SPECIFICATION

Reviewer Comments
Great job!

AWESOME
It is great that you have built a habit of creating pipelines prior to submission. This allows coders to review pipelines for errors that could make unwanted changes to the Database.

Other ideas about the datasets

SPECIFICATION
Student proposes one or more additional ways of improving and analyzing the data and gives thoughtful discussion about the benefits and anticipated problems in implementing the improvement.

DOES NOT MEET SPECIFICATION

Reviewer Comments
This rubric requires an explicit area of the project outlining...

1. Specific methods to improving data quality
2. methods to implement the improvements
3. potential issues that may arise from the implementation

Students are encouraged to think outside the scope of this class on this one. Think of ways to cross-validate, get more users involved. etc. Be sure to cover the numbered guidelines posted above.

Thoroughness and Succinctness of Submission

SPECIFICATION
Student submission is long enough to thoroughly answer the questions asked without giving unnecessary detail. A good general guideline is that your question

responses should take about 3-6 pages.

DOES NOT MEET SPECIFICATION

Reviewer Comments

Student submission is 17 pages. There is a lot of great exploration here, but the amount of added code and subsequent results made it cumbersome to read and distracted from the underlining question trying to be answered. It is ok to use snippets of code and snippets of results. Avoid printing everything from the exploration. Remember, you can submit .py files along with the final report. Reviewers are required to look through these code files so it is not necessary to include everything in the final PDF.

Best practices for your project resubmission

Ben shares 5 helpful tips to get you through revising and resubmitting your project.

Watch Video (3:01)

Have a question about your review? Email us at review-support@udacity.com.