

OpenStreetMap Case Study

Introduction

My task is to investigate a data set from a location of my choosing from openstreetmap. I need identify problems, clean it and store the data in SQL. Then, I am to explore the data programmatically and propose ideas on how to improve the data set. This investigation is a practice project needed to complete the Data Analyst Nanodegree from Udacity.

Location

I chose Auckland, New Zealand as my location for my investigation because I have been planning to take a trip here for sometime now. I would like to take the opportunity to get myself familiar with the place by using it as an example for this project.</p>

- <https://www.openstreetmap.org/node/292806332>

Problems Encountered in Your Map

To find and fix the streetnames in the map data, I ran the entire map data through a function that groups all street addresses into a dictionary according to the different variations used in the map. I filter the street names that I expect to be used and review the street addresses that I don't expect. Below are the problems that I will focus on for my audit.</p>

- Misspelled Names. Some street names are spelled incorrectly like Street.
- Incorrect Capitalization. Some street names are not capitalized consistently like road.
- Abbreviated Names. Some of the street names are abbreviated. I would prefer not to use abbreviations of their names. Instead of Hwy, use Highway.
- Problematic Format. Street names with problematic characters will be ignored.

I start fixing the data set by using the function `update_name`, which revises the streetnames according to my specifications that are outlined in `*mapping*`. Once I am satisfied with my data, I process the data in an xml structure according to the example schema. I then turn xml into csv. Once the CSVs are generated and validated, I then import the data into an SQL database to begin my exploration.

Overview of the Data

File Size

- auckland_new-zealand.osm: 658 MB
- project.db: 409 MB
- node_tags.csv: 3.67 MB
- ways.csv: 20 MB
- ways_nodes.csv: 84 MB
- ways_tags.csv: 86 MB
- nodes.csv: 246 MB

Counts

- Number of Unique Users: 953
- Number of Nodes: 2,913,110
- Number of Ways: 328,159
- Number of Hotels: 61
- Number of Attractions: 55
- Number of Museums: 13

Other Ideas about the Dataset

More descriptive Attractions Markers for Tourists

It would be a lot more helpful if tourists attractions are given more detail on their labels. Currently, the tags are labeled as `**attractions**` are not as helpful as much for traveler to plan their trip. If the attractions have more detail such as `**beach**`, `**monuments**` or `**nature**`, then these would be more helpful for would-be travelers planning a visit to Auckland.

Currently, there are 18 types of tourism markers. One of the tourism sites are labeled as *attractions* with 55 markers on Auckland, New Zealand. If these markers are provided with more detail, it could be helpful information to travelers.

I suggest that if when users edit, there should be a snippet of text within text field, where the users would input the data, that would encourage the user to add more detail to their markers or labels. Instead of an empty field, it could instead say, "What can I see here?".

Conclusions

Auckland, New Zealand map is well populated by contributions from active users. But because of the lack of guidelines or motivation to properly fill in the data is affecting the extent of details that each of the users is willing to input. If at the moment of data entry, there would be a descriptive text in a form of a question, users would be more obliged or encourage to place more detail on the data.