Data Wrangling Project

Udacity Nanodegree

By: Ada Zamora

In this folder I present the code I used to work on the project, each .py file starts with a letter that indicates the order that the files should be run, as variables and data from some files is later used in other files.

The order of the files is as follows:

- a_write_samples.py
- b_top_level_tags.py
- c_audit_streets.py
- d_audit_coordinates.py
- e_audit_postal_codes.py
- f_clean_osm_data.py
- g_write_csv.py
- h_create_db.py
- i_db_queries.py

There's also a schema.py file that contains the schema used to create the tables for the database.

All of this code was created in Python 3

List of all the other files in this folder:

- BA_postalcodes: the dataset used to verify the accuracy and consistency of the postal codes in the osm file, downloaded from https://yadi.sk/d/WIc5FNVEtk9U8
- BuenosAires: the database created with the Buenos Aires data
- data_link: containing the link where I downloaded the data
- small_sample: a file of approximately 1/50 of the size of buenos-aires_argentina data