Applied Data Science Capstone – Week 4

Peer-graded Assignment: Capstone Project – The Battle of Neighborhoods

Business Problem & Data Selection

The Battle of Neighborhoods | Business Problem & Introduction

Introduction:

I am currently living in the Long Island City neighborhood in New York, which is located in the Queens borough. I like my current neighborhood very much because of its multiculturalism, it's venues and also because of the density of venues. I have been living there for several years and really enjoy the surrounding.

Recently. I have been given a job offer in Toronto, Canada. To work in this place, I have to move to St. James Town in the Downtown of Toronto. I have never been in Canada before so it's a huge step for me to move there. I also have family. So, I want to make sure that there are enough high rated high schools in St. James Town. In addition to that, my family really like to eat Italian food. So I kind off have to make sure, that there will be enough high rated Italian restaurants.

I once lived in the Allapattah neighborhood in Miami. I didn't like it at all. So before I move to Toronto, I really want to be sure, that the St. James Town neighborhood is more like the Long Island City neighborhood and not like Allapattah.

If St. James Town is more like Long Island City and not like Allapattah, I will move to Toronto. The other way around: If St. James Town turns out to be more like Allapattah, which I did not like at all, I will decline the job offer and stay in Long Island City.

Tools:

This project will make use of the programming language Python. Therefore, the following API's and libraries will be used:

Foursquare API:
This API will be the primal source of data. The Foursquare compony gathers data of a huge variety of places and stores them in a database. The Foursquare API will be used to connect to the Foursquare database and get the needed information and data.

Clustering techniques:
To compare different neighborhoods, this project will use different clustering approaches. By building models we can therefore categorize the St. James Town neighborhood and compare it to other neighborhoods. K-Means will be one of the used algorithms.

Libraries:
Pandas: For creating DataFrames to store/ manipulate data and information.
NumPy: As fundamental package for scientific computing.
Folium: Python's visualization library to generate interactive leaflet maps.
Scikit Learn: For developing different clustering / classification models.
Matplotlib: For visualization of data.
Geocoder: To retrieve locational Coordinates.
JSCON: To handle JSON files.

The Battle of Neighborhoods | Data Description

Datalinks:

- Neighborhoods of Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Neighborhoods of Miami:
- https://www.zipdatamaps.com/nh-miami-neighborhood-allapattah
- Neighborhoods of New York City:
- https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm

Data handling:

I will use the pandas library to scrap the websites for their tables. After I've gotten all the needed postal/zip codes I will use geocoder to get the relevant coordinates. After that, I can visualize the Area by using the folium package.

By making use of the Foursquare API, this project will then characterize Long Island City and Allapattah. Afterwards already learned techniques will be used to classify the St. James Town neighborhood to compare it to the aforementioned neighborhoods.

I will use different endpoints to explore different locations, get tips and evaluate them and compare the overall density of venues in each area.Moreover, I will search for high rated high schools, since it is a huge and important factor. Since Italian restaurants seem to be desired, the Foursquare API will also be used to check for high rated fitting locations.

So, by using different endpoints, I will get and use the following data as features:

- Neighborhood
- Postal Code of Neighborhood
- Neighborhood Latitude
- Neighborhood Longitude
- Categories of venues
- Density of venues
- Name of venues
- Tips from users to check the ratings of different places
- Categories of schools
- Rating of high schools
- Amount of high rated Italian restaurants