Battle of the Neighborhoods

Cape Town, South Africa

Table of Contents

A description of the problem and a discussion of the background:

When looking to open a restaurant in any city, it is very important to do research on the city or neighborhood in which you are looking to open the restaurant. This will give you a good understanding of what kind of restaurants are in the area, which ones are doing well and why they are/ are not doing so well. Researching the amount of restaurants, the different cuisines and the ratings of these restaurants will be the best indicators of how well the new one will fair in its chosen location.

For the purpose of this project, we will assume that a person is looking to open a Sushi restaurant in the city of Cape Town, South Africa. Home to approximately 3.5 million people, Cape Town, South Africa is a very diverse City, with many restaurants serving food from all over the world. Sushi is a very popular food in Cape Town and there is no shortage of Sushi restaurants.

The person opening the restaurant wants to know how many Sushi restaurants are in each neighborhood as well as Cape Town in general and in which neighborhoods are there Sushi restaurants. From this information they would like a recommendation as to which neighborhood is the best one to open their restaurant in.

A description of the data and how it will be used to solve the problem:

Description of the data:

The data I will be using will be as follows:

- **The Cape Town postal codes:** The table containing all of the postal codes and neighborhood names for the Cape Town area, which are found on the South African Postal Codes website (http://sapostalcodes.info/queryPostal/Cape+Town).
- **The Foursquared API data:** Obtaining the location data of each neighborhood, such as: venue names, geospatial data and categories.
- **Python Libraries/packages:** In order to create the machine learning algorithms, generate maps reflecting the data, encode the data, normalize the data and display the results.
    - **The Python Libraries/packages:**
        - Pandas
        - Numpy
        - Matplotlib
        - JSON
        - Sklearn
        - Folium
        - Requests
        - Geopy

How the data will be used to solve the problem:

Using the Cape Town postal codes as well as the Foursquared API data, I will create data frames in order to normalize the data. These data frames will contain the geospatial data and venue data. The data will be processed, encoded using "One Hot" encoding, run through a K-means clustering algorithm and mapped out visually. The results will then be used to determine which neighborhood will be the best suited to open the new restaurant in.