

Capstone

Capstone Project - The Battle of Neighborhoods

Mehmet Şimşek

Part 1: Problem Description:

One client wants to start a new business and open a new restaurant in Bronx. Since he does not have much experience, he wants to open his restaurant in a neighborhood that has the least number of restaurants. Because he doesn't want to face a competition. He wants to open an Italian or a Mexican type of restaurant and he wants me to analyze all Bronx to find what kind of restaurants are common. Considering the info that I will provide for him, he will decide to open either Italian or Mexican restaurant after seeing that which of them is uncommon.

Business problem:

Can I make a cluster using Foursquare data and then help the client decide where to open which kind of restaurant?

So my mission is to find the least common type of restaurants and number of restaurants in each neighborhood to help my client decide a location for his restaurant.

Target Audience: Only my client will be interested in this project.

Part 2: Data Description

Since the client wants to build open his restaurant in Bronx, NY, I will need geo-locational information of Bronx and its neighborhood. For this part of the problem I will use postal code information provided by the NYU Spatial Data Repository. By using this information, I will analyze the target area and then with the help of Foursquare I will find the most common restaurant types and total number of restaurants in each neighborhood. I will do this by getting data of location (latitude, longitude), distance to center, category and popularity that is provided by a typical request from Foursquare.

Following **libraries** are used while executing the Capstone Project:

- **Pandas** - Library for Data Analysis
- **NumPy** – Library to handle data in a vectorized manner
- **Requests** – Library to handle http requests
- **Matplotlib** – Python Plotting Module
- **Sklearn** – Python machine learning Library
- **JSON** – Library to handle JSON files
- **Geopy** – To retrieve Location Data

- **Folium** – Map rendering Library

Following **data sources** are used while executing the Capstone Project:

- NYU Spatial Data Repository- Postal codes of NY- File type:JSON
- Foursquare data- typical requests data package- File type:JSON

Foursquare API:

Foursquare is a social location service that allows users to explore the world around them. The Foursquare API allows application developers to interact with the Foursquare platform. The API itself is a RESTful set of addresses to which you can send requests, so there's really nothing to download onto your server. You can currently request output in XML or JSON format, making requests to URLs that look like this: <http://api.foursquare.com/v1/user.json>.

HTTP requests parameters:

Number of neighborhoods: 200

Range: 1000.

Folium:

folium builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the leaflet.js library. folium makes it easy to visualize data that's been manipulated in Python on an interactive leaflet map. I will use folium to visualize best location for our contractor.

K-mean:

It is an unsupervised machine learning calculation and I will use it to making clusters and most common restaurants in each neighborhood.

Part 3: Methodology

- **Step-1: Identifying Neighborhoods inside Bronx**

With the help of JSON file that I got from

https://geo.nyu.edu/catalog/nyu_2451_34572, first I made a dataframe consisting of all neighborhoods of Bronx and then visualized it to understand better.

- **Step-2: Connecting to Foursquare and Retrieving Locational Data for Each Venue in Every Neighborhood**

After the first step, I gathered information about venues in neighborhoods by connecting the Foursquare API. I set the radius parameter to 1000.

- **Step-3: Processing the Retrieved Data and Creating a Data Frame for All the Venues inside the Bronx.**

When the data was completely gathered, I performed processing on that raw data to find my desirable features for each venue. My main feature was the category of that venue (only restaurants/diners are necessary for me). After this stage, the column Venue's Category would be One-hot encoded and different venues would have different feature-columns.

- **Step-4: Applying one of Machine Learning Techniques (K-Means Clustering)**

By using a K-means clustering method, I was able to create five clusters and then find 5 the most common types of restaurants. And as a final step I calculated the number of restaurants in each neighborhood.

Part 4: Results:

We had two criteria:

1. Finding a neighborhood that has the least number of restaurants.
2. Finding the least common restaurant type in this neighborhood.

So according to first criterion Allerton and Port Morris have the least number of restaurants.

And according to second criterion;

In Allerton most common restaurants are: Italian Restaurant, Caribbean Restaurant, Mexican Restaurant, Spanish Restaurant Diner.

In Port Morris most common restaurants are: Italian Restaurant, Greek Restaurant, American Restaurant, Mexican Restaurant Ramen Restaurant.

Based on this analysis, the client should open his Mexican Restaurant in Port Morris.