Coursera Capstone Project Full Report

IBM Data Science Certification

By Ian Zhang

Introduction

I just moved to Raleigh North Carolina. I try to find the best neighborhood with most of the restaurant.

Collect Data

- Step1: Find all neighborhoods in the Raleigh NC area.
  - https://en.wikipedia.org/wiki/Raleigh,_North_Carolina_neighborhoods
  - Since it is just simple list in the wiki page, I copy them and save them in a csv file
- Step 2: Find Latitude and Longitude for each neighborhoods.
  - Method 1: Use python geocoder module
    - I use for loop to loop through all neighborhoods and find the coordinates for each one.
  - Method 2: https://www.latlong.net/
    - When I draw the map, I find python geocoder module is not accurate. Some Raleigh neighborhoods are in different city.
    - Therefore, I use https://www.latlong.net/ to correct some coordinates for two neighborhoods.
- Step 3: Find all restaurant venues for each neighborhoods
  - Use Foursquare api to find all categoryId= 4d4b7105d754a06374d81259 (restaurant)

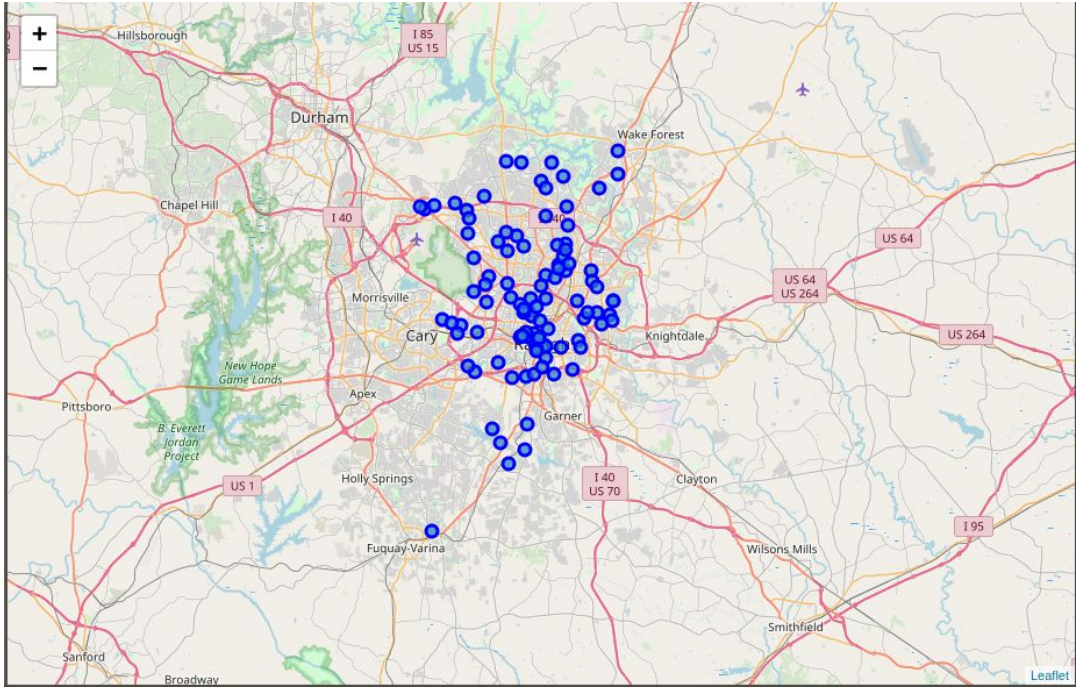
Processing and Analyzing the Data

- Step 1: Read all Raleigh neighborhood data from the saved csv file.

- Step 2: Calculate the coordinates for each neighborhood:
  - For each neighborhood, use python geocoder to find latitude and longitude.
  - Combine Pandas DataFrame to form the “raleigh_neighborhoods” data frame.
  - Save the dataframe to csv file (phase_1 results).
  -

Processing and Analyzing the Data (continued)

- Step 3 Draw map with each neighborhoods on the map.



Processing and Analyzing the Data (continued)

- Step 4: Find all surrounding restaurants for each neighborhoods
  - Use Foursquare api.
  - Foursquare api can find all surrounding venues with
    - Certain categories
    - Latitude
    - Longitude

```
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT,
            cateId)
        #print(url)
        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']
```

Processing and Analyzing the Data (continued)

- Step 5: manipulate the venue results
  - Put all Foursquare results into a Pandas DataFrame
  - Create a “Count” according to the neighborhood.
  - Merge the count DataFrame with Original Neighborhood DataFrame.
- Step 6:
  - Sort the final DataFrame descendingly.
  - Get the top 10 list from the DataFrame.
- Step 7:
  - Print the top 10 list
  - Draw the top 10 neighborhoods (with the most restaurants) on Raleigh map

Final Results

Here is the top 10 neighborhoods with most surrounding restaurants. Some neighborhoods are so closed, therefore they overlapped on the map.