Applied Data Science Capstone Project

For new kids Swimming school centre

CAPSTONE PROJECT - REPORT

IBM- Coursera

Upendra Pamireddy – 2019

Table of Content :

1). Introduction Section:

Report is for course of IBM Data Science and main goal of this report is to Identify “Recommendation of new kids swimming school centre in Richmond City Virginia State(USA)?"

Problem to be solved through Data Science:

"ABC swimming firm hired data scientist to analyse possible new swimming centre for expanding business in Richmond City".

According to firm "Swimming lessons are a must for any homeowner with children and a swimming pool in Marietta. Pool safety should always be the number one priority of parents and it's never too early to teach your kids how to swim. In addition to learning a life-saving technique, swimming lessons have several great benefits for both children and adults"

Richmond city : is the capital of the Commonwealth of Virginia in the United States. It is the center of the Richmond Metropolitan Statistical Area (MSA) and the Greater Richmond Region. As of the 2010 census, the city's population was 204,214;[4] in 2016, the population was estimated to be 223,170,[4] making Richmond the fourth-most populous city in Virginia. The Richmond Metropolitan Area has a population of 1,260,029, the third-most populous metro in the state. Source: https://en.wikipedia.org/wiki/Richmond,_Virginia

Need to consider nearby locations like schools,residential houses and other swimming schools to decide possible locations for opening a new swimming school.

2) Data Section :

Description of data that will solve the problem and the sources.

2.1 Need to fetch following data points to solve the request:

a) List of Elementary schools list in Richmond city. Refrence from Wiki link: https://en.wikipedia.org/wiki/Richmond_Public_Schools#Elementary_schools

We will concentrate on Elementary schools to identify kids population pattern near a neighbourhood.

Downloaded content from both wiki links into CSV files.

b) Extract list of existing swimming schools in Richmond city with corresponding address by using Foursquare data for segmentation and clustering with the string search

2.2 Uses of Data for finalizing a centre

With the help of Foursquare and geopy data :

a) Map the neighbouhoods/sub cities of Richmond City. Foursquare map of Richmond displays this information.
b) Map the Elementary schools of Richmond City to identify the density of kids population in the neighbourhood.
c) Map the current list of swimming schools in Richmond City.

3) Methodology Section :

```
# # Importing all necessary libraries for executing project.
!conda install -c conda-forge geopy --yes
from geopy.geocoders import Nominatim # module to convert an address into latitude and longitude values
import requests # library to handle requests
import pandas as pd # library for data analsysis
import numpy as np # library to handle data in a vectorized manner
import random # library for random number generation

# libraries for displaying images
from IPython.display import Image
from IPython.core.display import HTML

# tranforming json file into a pandas dataframe library
from pandas.io.json import json_normalize

!conda install -c conda-forge folium=0.5.0 --yes
import folium # plotting library

print('Folium installed')
print('Libraries imported.')
```

To fetch latitude and Longitude of Richmond City to build map and map swim schools

```
# Fetch Latitude and Longitude of Richmond using Nominatim.
address = 'Richmond, VA'

geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Richmond City are {}, {}.'.format(latitude, longitude))
```

Extract list of existing swimming schools in Richmond city with corresponding address by using Foursquare data for segmentation and clustering¶

```
# search list of Swim schools with the key word.
search_query = 'Aquatics Swim school'
print(search_query + ' .... OK!')
```

```
Aquatics Swim school .... OK!
```

```
url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}'.format(C

results = requests.get(url).json()
```

After series of format on the data , below list of Swimming schools were identified.

List of current Swim schools in RICHMOND CITY

Display DATAFRAME containing list of Swim School centres

```
# Total 8 centres that were found by using Foursquare data for segmentation and clustering.
# Out of 8 centres, one of them is Theme Park, drop the theme park catgory.

swim_centres = swim_centres[swim_centres.categories != 'Theme Park']
swim_centres
```

	name	categories	lat	lng
0	Goldfish Swim School - Richmond West End	Swim School	37.606600	-77.593630
1	Aqua-Tots Swim Schools Richmond	Pool	37.619497	-77.521620
2	Aqua-tots Swim School	Pool	37.508178	-77.602666
3	Aqua-Tots Swim Schools Midlothian	Swim School	37.508333	-77.602675
4	Collegiate School Aquatics Center	Pool	37.449543	-77.492450
5	Morgan Swim School - Munsey	Gym Pool	37.698833	-77.563333

RICHMOND Elementary Schools Data and Mapping

School list from WIKI page is converted into .CSV file. Around 15 Elementary schools are listed for analysis using FOURSQUARE.

[10]:
```
Elem_School = pd.read_csv('Rich_Elem_School_List.csv')
Elem_School["Lat"]=''
Elem_School["Long"]=''
Elem_School.head()
```

[10]:

	Elementary schools	School Address	Lat	Long
0	Barack Obama Elementary School	Barack Obama Elementary School , Richmond City,VA		
1	Bellevue Elementary School	Bellevue Elementary School, Richmond City,VA		
2	Blackwell Elementary School	Blackwell Elementary School , Richmond City,VA		
3	Broad Rock Elementary School	Broad Rock Elementary School , Richmond City,VA		
4	Fairfield Court Elementary School	Fairfield Court Elementary School , Richmond C...		

```
## Fetching latitude and longitude for list of schools using Nominatim
## Importing Time parameter.
import time
for i in range(len(Elem_School)):
    address= Elem_School['School Address'][i]
    geolocator = Nominatim(user_agent="my-application")
    location = geolocator.geocode(address)
    latitude = location.latitude
    longitude = location.longitude
    Elem_School['Lat'][i]=latitude
    Elem_School['Long'][i]=longitude
    #print(n,latitude,longitude)
    time.sleep(2)

print('completed')
Elem_School
```

```
completed
```

	Elementary schools	School Address	Lat	Long
0	Barack Obama Elementary School	Barack Obama Elementary School , Richmond City,VA	37.5738	-77.4355
1	Bellevue Elementary School	Bellevue Elementary School, Richmond City,VA	37.5314	-77.4212
2	Blackwell Elementary School	Blackwell Elementary School , Richmond City,VA	37.5167	-77.4434
3	Broad Rock Elementary School	Broad Rock Elementary School , Richmond City,VA	37.4764	-77.4787

3.1 MAP of Richmond showing the location of current Swimming school centres

Popups indicate the name of the SWIM centre

```
# create map of Richmond place  using latitude and longitude values

latitude = 37.5385087
longitude =  -77.43428
map_richmond = folium.Map(location=[latitude, longitude], zoom_start=10)
# add markers to map
for lat, lng, label in zip(swim_centres['lat'], swim_centres['lng'], swim_centres['name'].astype(str)):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        number_of_sides=30,
        radius=7,
        popup=label,
        color='blue',
        fill_color='#0f0f0f',
        fill_opacity=0.6,
     parse_html=False).add_to(map_richmond)
## Display Richmond Map
map_richmond
```

3.2 MAP of Richmond showing the location of Elementary schools

Popups Indicate name of the school

```
# create map of Richmond place  using latitude and longitude values

latitude = 37.5385087
longitude =  -77.43428
map_richmond = folium.Map(location=[latitude, longitude], zoom_start=11)
# Display elementary schools in Richmond Map.

for lat, lng, label in zip(Elem_School['Lat'], Elem_School['Long'], Elem_School['Elementary schools'].astype(str)
    label = folium.Popup(label, parse_html=True)
    folium.RegularPolygonMarker(
        [lat, lng],
        number_of_sides=6,
        radius=6,
        popup=label,
        color='red',
        fill_color='red',
        fill_opacity=2.5,
     ).add_to(map_richmond)
# Display MAP
map_richmond
```

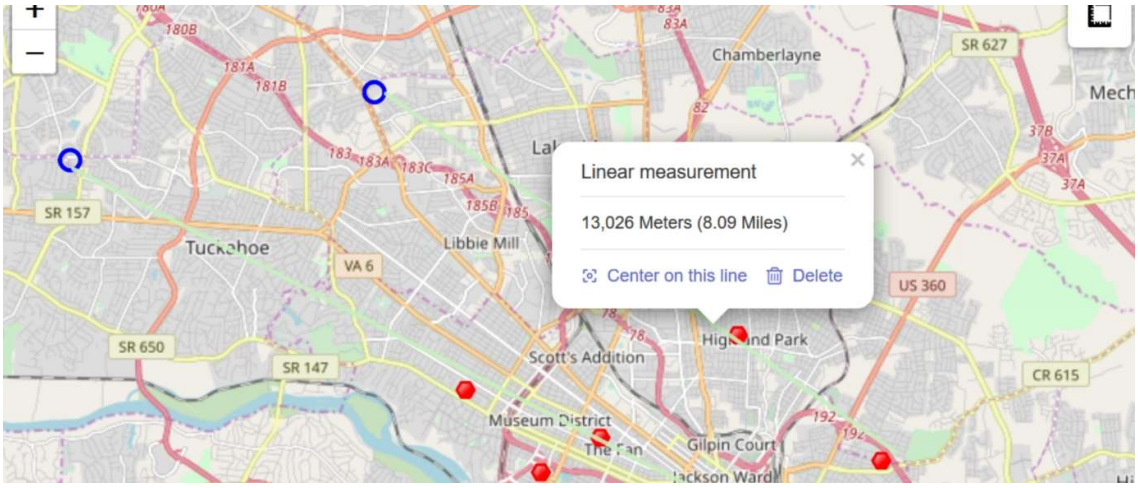
4) Results

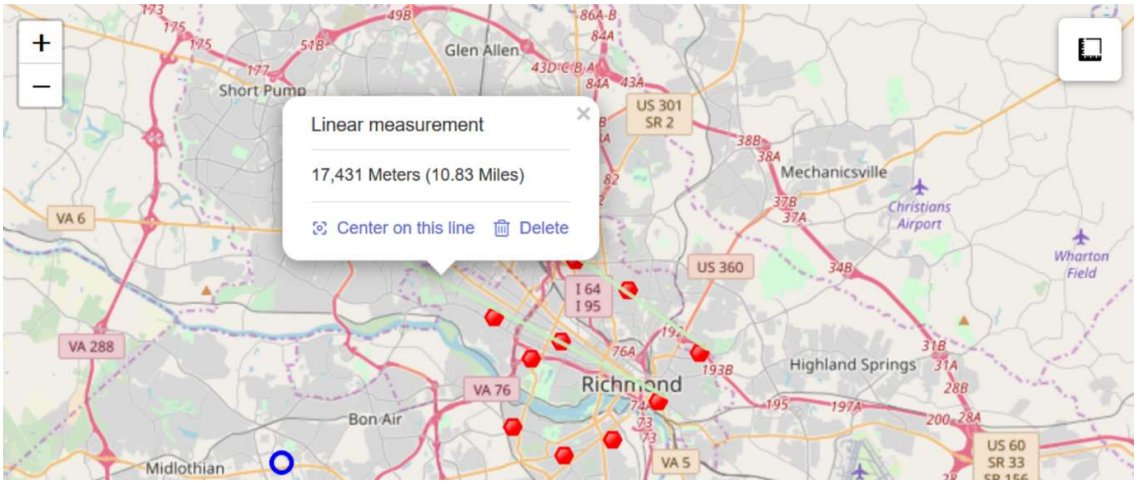
Consolidating two maps to identify possible new swim school centre

a) Map of Richmond city with current list of Swim school centres
b) Map of Richmond City with list of Elementary schools

Blue circles represent Swim School Centres and red circle represent elementary schools , click on circles for viewing name of school

There is a "ruler" Icon in the top right corner to measure distance between perticular neighbourhood or school to any Swim School centre.





5.0 Conclusion

After reviewing consolidated map and comparing the distance with ruler, we have 3 possible locations for opening new Swim centre

All the 7 Swim schools are concentrated in the North,West,central and sounth of Richmond .

There is no Swim school in the eastside of richmond.

a) ChurchHill North (Near by school Bellevue Elementary School), distance to near by swim school is 10.8 miles

b) Highland Park (Near by school Barack Obama Elementary School) , distance to near by swim school is 5.6 miles

c) Washington park (Near by Elementary School Linwood Holton Elementary School) , distance to near by Swim School is 4.91 miles

All these three neighbourhoods from eastside of Richmond are surronded by Schools and have good density of pupils. Best option is to open a centre in Church Hill North , if there are any other Business or Cost constrains Highland Park or Washington park can be considered.

5 Discussion

a) This course helped to demonstrate and analytical think about good project(Capstone Lab). Thank You IBM and Coursera Team.

b) Course content helped for a good start to Data Science Carrier

6 Conclusions

Based on the Data Science analysis report submitted to ABC swimming firm , three neighbourhoods are best options,but if the firm wants to understand further details like availability of commercial space cost and monthly price at which students prefer Swim school..etc

Data Scientist needs to add more parameters in the criteria and analyse further to narrow down detailed options.