Capstone Project – The Battle of Neighborhoods

Applied Data Science Capstone

by IBM

on coursera

Introduction

- **Business problem**: find the most profitable location to open a pizzeria in Scarborough, Toronto, Canada
- **Interested parties**: Luigi and his family, who would like to open a pizzeria near their house in Scarborough.

Data

Raw data present postal codes, neighborhoods and geographical location of Scarborough.

0	M1B	Scarborough	Rouge, Malvern	43.811650	-79.195561
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.785605	-79.158701
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.765690	-79.175299
3	M1G	Scarborough	Woburn	43.768216	-79.217610
4	M1H	Scarborough	Cedarbrae	43.769608	-79.239440
5	M1J	Scarborough	Scarborough Village	43.743085	-79.232172
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.726260	-79.263670
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.713213	-79.284910
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.723575	-79.234976
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.696690	-79.260069
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.759975	-79.268974
11	M1R	Scarborough	Maryvale, Wexford	43.750803	-79.300560
12	M1S	Scarborough	Agincourt	43.793940	-79.267976
13	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.784725	-79.299244
14	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	43.817595	-79.280147
15	M1W	Scarborough	L'Amoreaux West, Steeles West	43.800698	-79.320740

Methodology

Following libraries were installed for data analysis

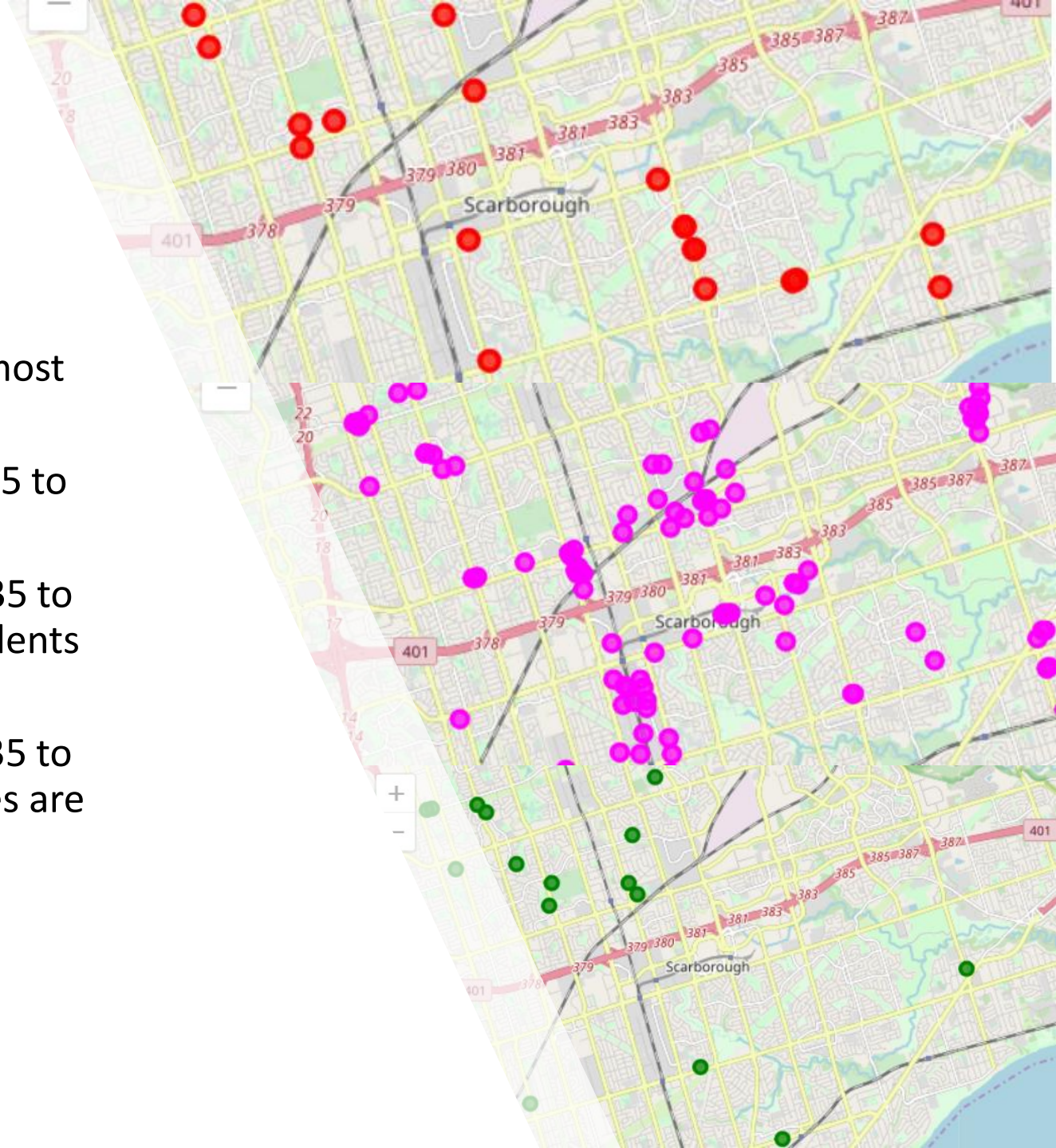
```
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analsysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

!conda install -c conda-forge geopy --yes
# convert an address into latitude and longitude values
from geopy.geocoders import Nominatim

import requests # library to handle requests
# tranform JSON file into a pandas dataframe
from pandas.io.json import json_normalize

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

!conda install -c conda-forge folium=0.5.0 --yes
import folium # map rendering library
```

Methodology

- Neighborhoods in Scarborough where explored using Foursquare API with following URL and Foursquare API version:
- url: 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&ll={},{}&v={}&radius={}&limit={}'.format (CLIENT_ID, CLIENT_SECRET, latitude, longitude, VERSION, radius, LIMIT)
- version: '20180605'

Methodology

- Following approaches were done to find the most profitable location of a pizzeria:
- # Use category id 4bf58dd8d48988d1ca941735 to only find pizzerias in Scarborough
- # Use category id 4bf58dd8d48988d13d941735 to only find high schools in Scarborough, since students are good customers
- # Use category id 4bf58dd8d48988d124941735 to only find offices in Scarborough, since employees are good customers.



Methodology

- Following approaches were done to find the most profitable location of a pizzeria:
- # Number of pizzerias, high schools and offices were summed up per neighborhood and weighted according to the following sheme:

```
# negative weight, because Luigi wants to open a pizzeria
#and thus wants to avoid concurrence as much as possible
weight_pizza = -1

# positive weight, because high school students are good custor
weight_schools = 1

# positive weight because employees are even better customers
weight_offices = 2
```

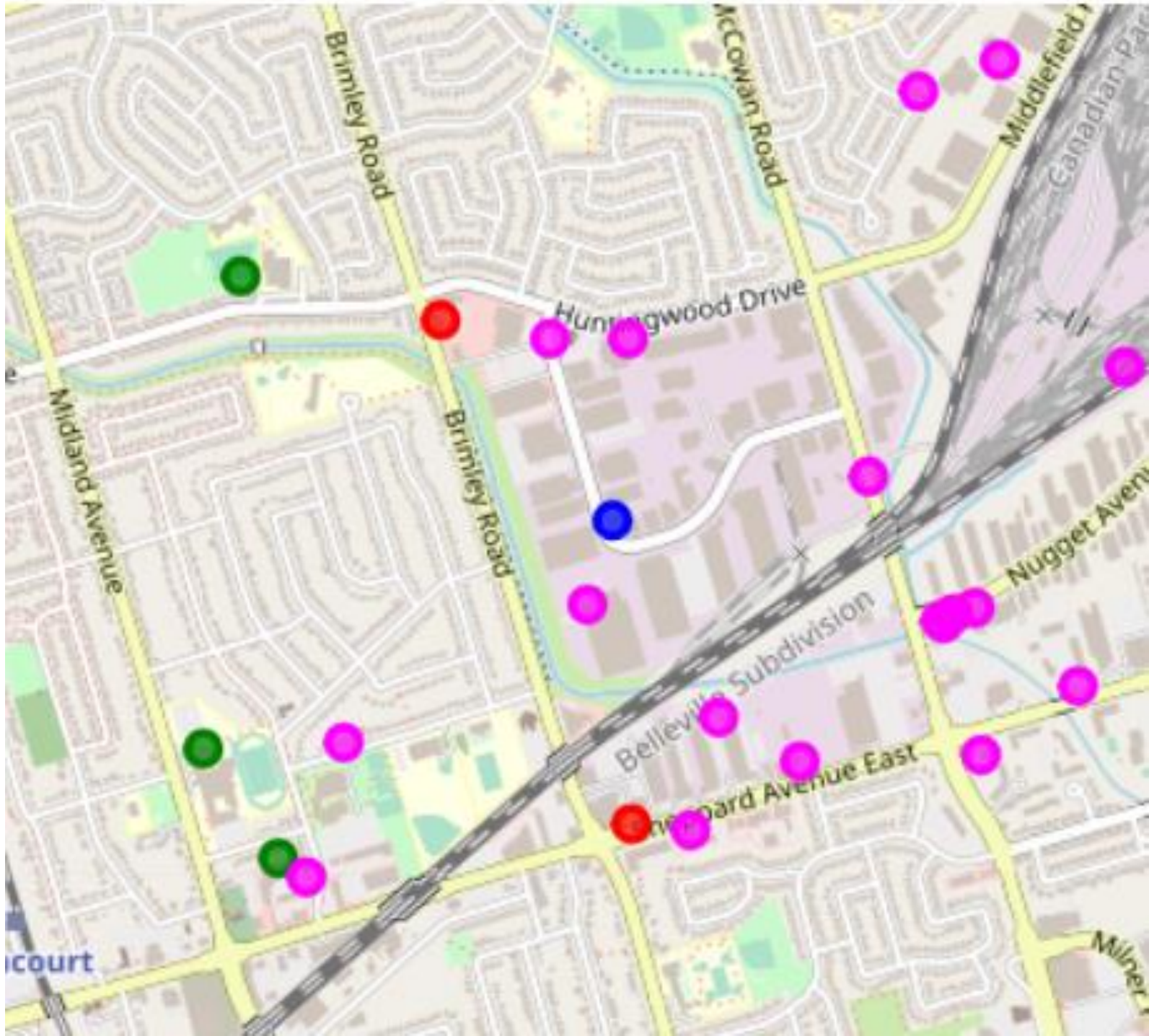
Results

- Sums of each category were multiplied with weights to generate following table, which summarizes the best locations in Scarborough to open a pizzeria. Agincourt presents the highest score of neighborhoods in Scarborough.

	Neighborhood	Score
12	Agincourt	39.0
15	L'Amoreaux West, Steeles West	36.0
10	Dorset Park, Scarborough Town Centre, Wexford ...	35.0
6	East Birchmount Park, Ionview, Kennedy Park	30.0
13	Clarks Corners, Sullivan, Tam O'Shanter	28.0
2	Guildwood, Morningside, West Hill	21.0
4	Cedarbrae	17.0
0	Rouge, Malvern	15.0
7	Clairlea, Golden Mile, Oakridge	13.0
11	Maryvale, Wexford	13.0
8	Cliffcrest, Cliffside, Scarborough Village West	12.0
14	Agincourt North, L'Amoreaux East, Milliken, St...	9.0
9	Birch Cliff, Cliffside West	8.0
1	Highland Creek, Rouge Hill, Port Union	8.0
3	Woburn	3.0
16	Upper Rouge	2.0
5	Scarborough Village	-2.0

Results

- Following maps shows the neighborhood of Agincourt (blue) with existing pizzerias (red), high schools (green) and offices (fuchsia).



Discussion

- Further analysis can be done on the second and third ranking neighborhood in Scarborough to find alternative spots that might be lucrative but also more close to Luigis home.
- The data analysis can be improved with following extensions:
- Consider more categories, e.g. Nightlife Spots (4d4b7105d754a06376d81259) or Food/Restaurants (4d4b7105d754a06374d81259), which can be a source of customers but also present some concurrence.