



The Battle of Neighbourhoods Week 5

PRESENTATION



Business problem

- ▶ We want to figure out in which area in Toronto we can open a new "real italian restaurant.
- ▶ We want to accomplish this goal by finding out where in toronto are already italian restaurant situated so we can deterime a good spot.

Data Selection

Following data sources will be used to get the required information:

- ▶ Wikipedia will be used scrap Toronto neighbourhoods.
- ▶ Geospatial_Coordinates.csv will be used to get Latitude and Longitude information.
- ▶ Foursquare API will be used to get restaurants data related to these 2 cities.

Data Flow

Following data sources will be used to get the required information:

- ▶ First, it is used data from get city open data to get city information as well as latitude and longitude coordinates
- ▶ Then, we created a data frame with borough and neighbourhood information. For Toronto, it is used Wikipedia to get the list of Postal Code of all Neighbourhoods in Toronto
- ▶ And. for the neighbourhood in this data frame, it will be gathered the list of restaurants from Foursquare. With this information it is possible to come up with a total as well as draw the maps with Italian restaurants locations.

Methodology

- ▶ The goal of this project is to come up with a simple study to identify area's in the city of Toronto, where Italian Restaurants are located. So we can define areas of opportunities to invest / start an new Italian Restaurant.
- ▶ After that, it will be presented some number to justify the decision about which area has the most restaurant which helps us determine other area's where we could start our restaurant.
- ▶ And finally, in the last part of this study, it is showed a map showing the spots where these Italian restaurants are located, and helps us to visualize the areas of opportunity for our restaurant.

Analysis

► Get geo information from Toronto from Wikipedia

```
In [2]: #We will use BeautifulSoup to get the zip code information of Canada from Wikipedia
page = requests.get("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M")
soup = BeautifulSoup(page.content, 'html.parser')

In [3]: table_contents=[]
table=soup.find('table')
for row in table.findAll('tr'):
    cell = {}
    if row.span.text=='Not assigned':
        pass
    else:
        cell['PostalCode'] = row.p.text[:3]
        cell['Borough'] = (row.span.text).split('(')[0]
        cell['Neighborhood'] = (((((row.span.text).split('(')[1]).strip(' '))).replace(' /',',')).replace(')',',').strip(' ')
        table_contents.append(cell)

In [4]: #We save this to dataframe (df)
df=pd.DataFrame(table_contents)
df['Borough']=df['Borough'].replace({'Downtown TorontoStn A PO Boxes25 The Esplanade':'Downtown Toronto Stn A',
                                     'East TorontoBusiness reply mail Processing Centre969 Eastern':'East Toronto Business',
                                     'EtobicokeNorthwest':'Etobicoke Northwest','East YorkEast Toronto':'East York/East Toronto',
                                     'MississaugaCanada Post Gateway Processing Centre':'Mississauga'})
```

Analysis

- We extracted some venue data from foursquare.

```
In [10]: # Lets get the venue data from foursquare
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]["groups"][0]["items"]

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                            'Neighborhood Latitude',
                            'Neighborhood Longitude',
                            'Venue',
                            'Venue Latitude',
                            'Venue Longitude',
                            'Venue Category']

    return(nearby_venues)
```

Analysis

- We filtered it to only Italian restaurants.

```
In [19]: #create a new data frame with only the italian restaurants (df4)
df4 = to_grouped[["Neighborhoods", "Italian Restaurant"]]

#show the first 5 rows
df4.head ()
```

Out[19]:

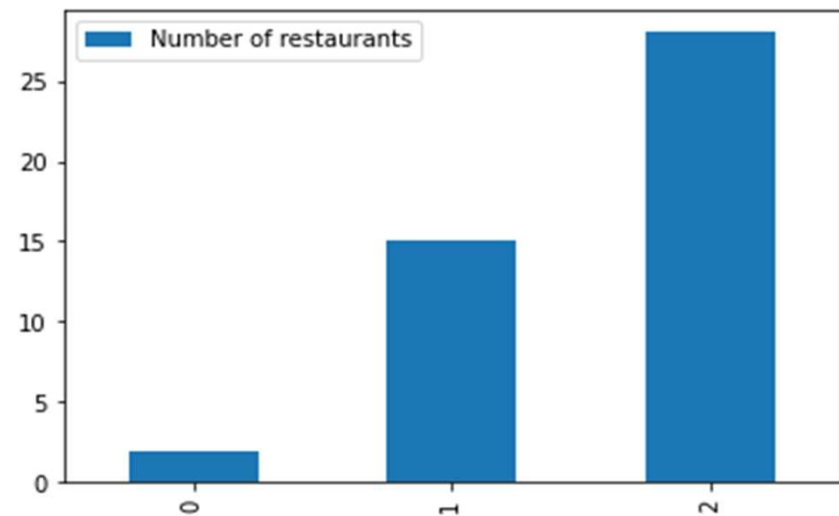
	Neighborhoods	Italian Restaurant
0	Agincourt	0.000000
1	Alderwood, Long Branch	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.000000
3	Bayview Village	0.000000
4	Bedford Park, Lawrence Manor East	0.083333

K-Mean

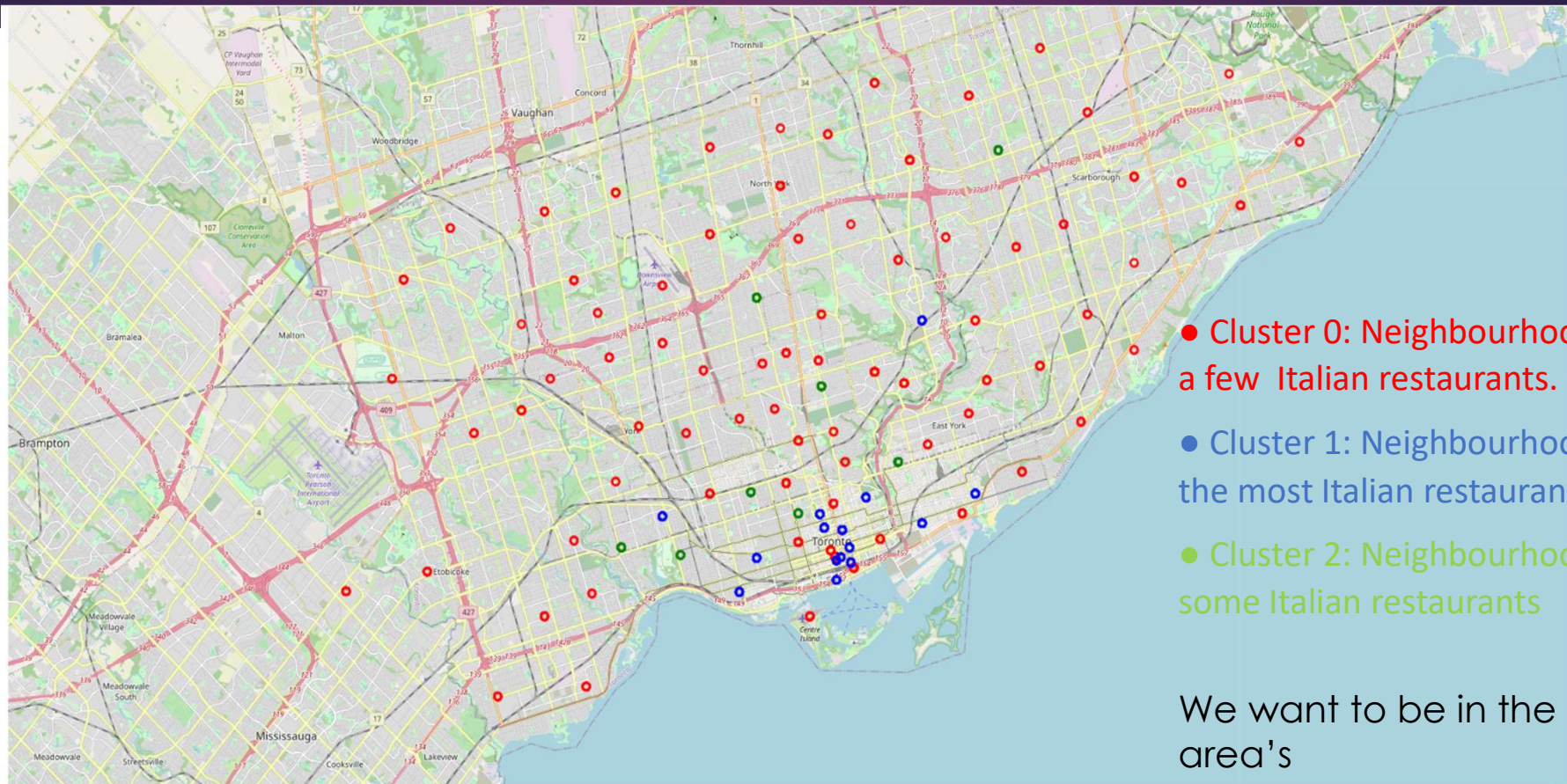
- ▶ With the K-mean we determined 3 clusters to see how many restaurants are in each cluster.
- ▶ The results of this you can find on the next 3 slides.

Overview restaurants in the area's

Segment	Number of Italian Restaurants
0	2
1	15
2	28



Map of the segments,



- Cluster 0: Neighbourhoods with a few Italian restaurants.

- Cluster 1: Neighbourhoods with the most Italian restaurants.

- Cluster 2: Neighbourhoods with some Italian restaurants

We want to be in the red area's



Recommendations

- ▶ Most of the Italian restaurants are in cluster 1 lowest in Cluster 0.
- ▶ Looking at nearby venues it seems cluster 1 might be a good location as there are not a lot of Italian restaurants in these areas. We therefore recommend the entrepreneur to open an authentic Italian restaurant in these locations.
- ▶ We might want to explore the areas close to the blue and green areas first because there are likely to be more downtown.

Map of the area with cluster 0,

