PIZZA DI MARIO

By:
Muzamal Azam

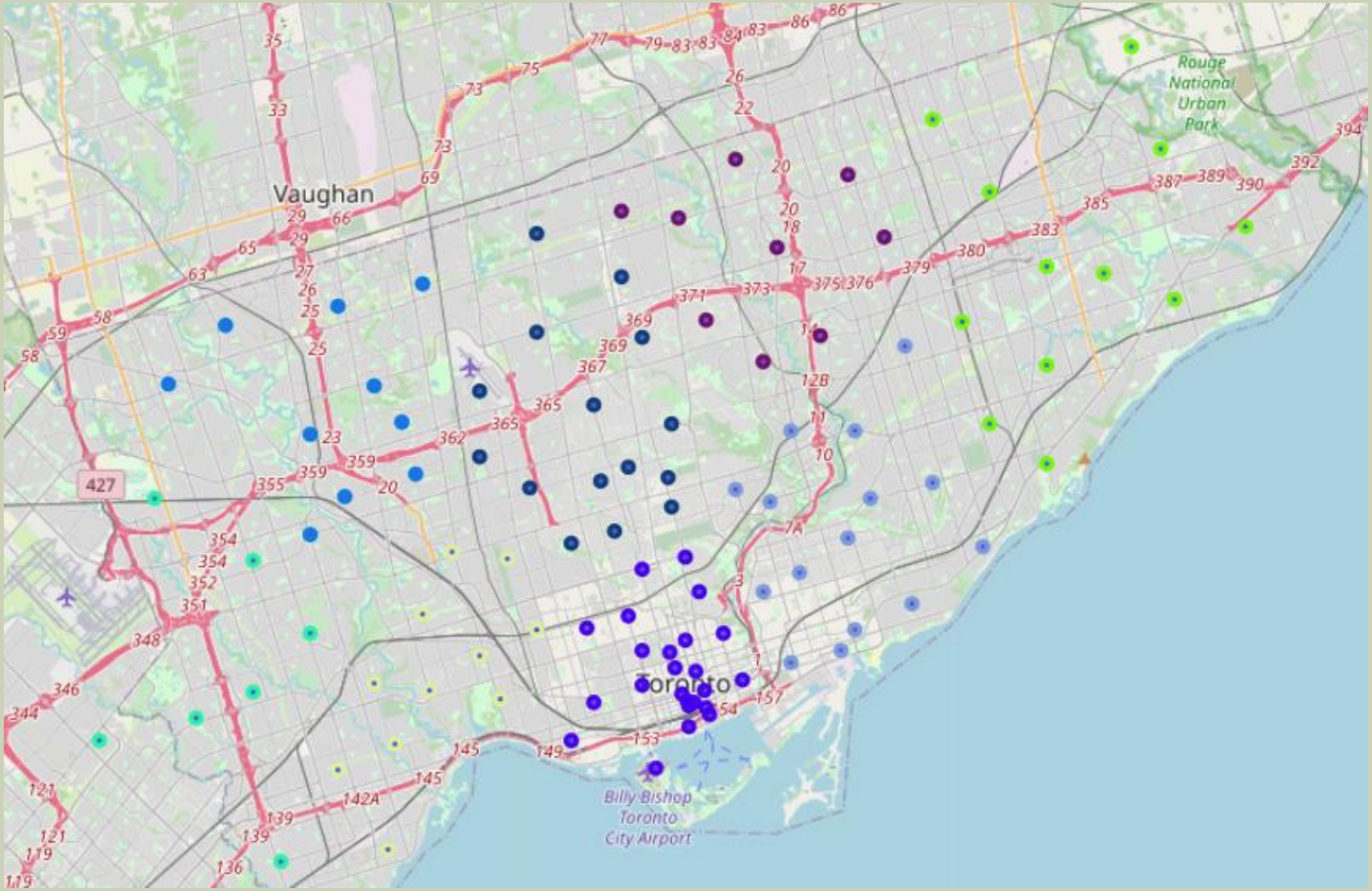
OBJECTIVES

- FIND THE RIGHT LOCATION FOR A NEW BUSINESS THAT WANTS TO SELL THE BEST PIZZA FROM NAPLES
- KNOW MORE ABOUT THE NEIGHBOURHOOD WHERE THE BUSINESS WOULD LIKE TO OPEN
- FIND THE PLACE WITH LESS BUSINESS COMPETITION

DATA ACQUISITION AND CLEANING

- The DATASETS USED FOR THIS PROJECTS ARE:
  - TORONTO POSTAL CODES (WEBPAGE FROM WIKIPEDIA THAT HAS BEEN WEBSCRAPPED)
  - DINESAFE, WHICH INCLUDES A SET OF SHOPS AND RESTAURANTS (THE DATA HAS BEEN TRANSFORMED FROM XML TO CSV)
  - TORONTO POINTS OF INTERESTS, WHICH INCLUDES DATA REGARDING MONUMENTS, MUSEUMS, COMMUNITY PLACES, SCHOOLS, AND HISTORICAL SITES. (PRIOR THE DATA HAS BEEN TRANSFORMED FROM SHAPE FILE TO CVS)

INITIAL CLUSTER OF 8 POINTS



ASSIGNING CLUSTERS TO RESTAURAUNTS BY THEIR DISTANCE

assign the cluster to the pizza shops to the closest cluster

In [19]:
```
locs_df['Labels'] = locs_df.apply(lambda row: closestCluster(row['LATITUDE'],row['LONGITUDE']), axis=1)
```

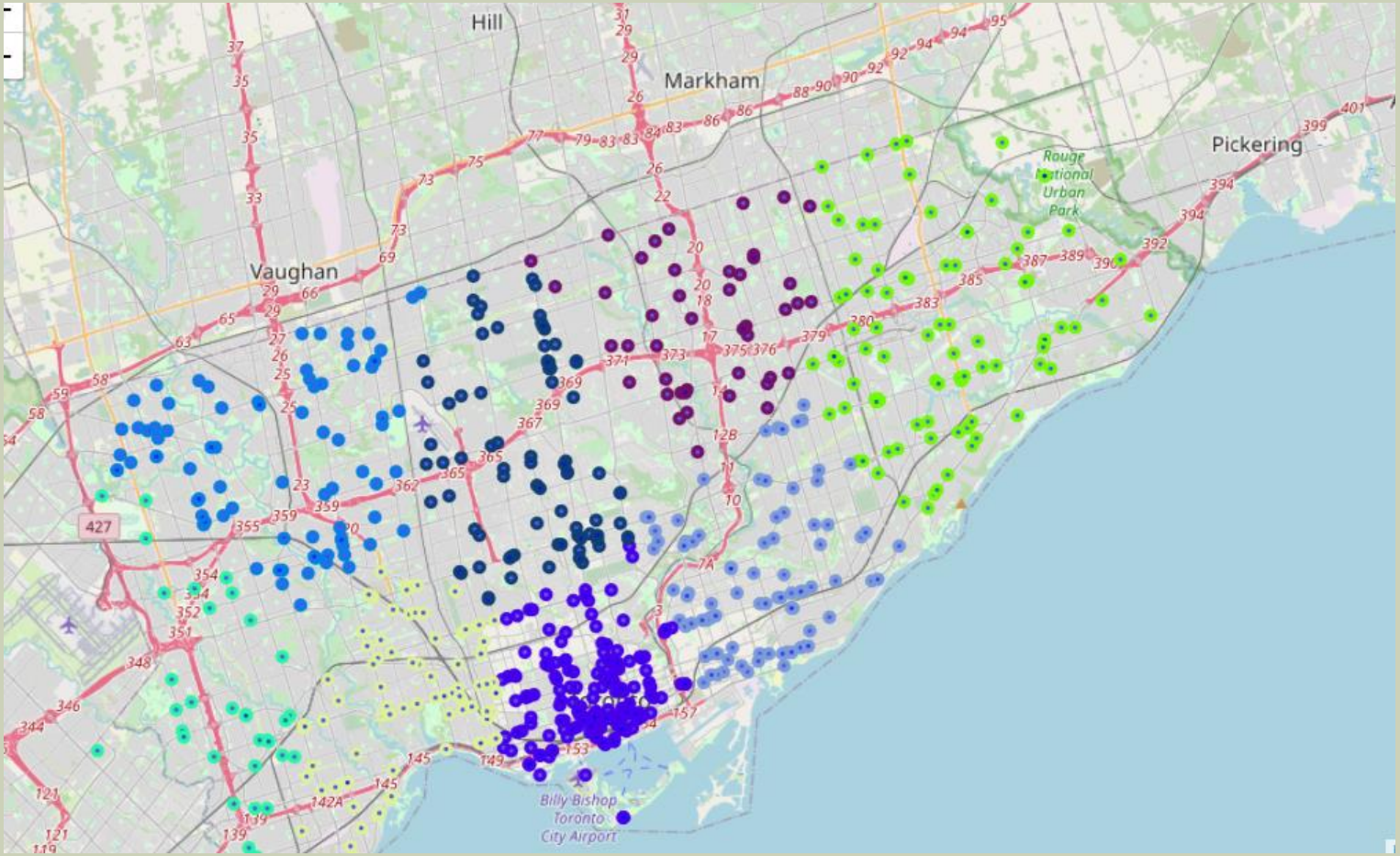
In [20]:
```
locs_df.head()
```

Out[20]:

	ID	NAME	TYPE	ADDRESS	LATITUDE	LONGITUDE	Labels
0	10500438	1 PLUS 1 PIZZA	Food Take Out	361 OAKWOOD AVE	43.68725	-79.43842	1
1	10300086	115-PIZZA NOVA	Food Take Out	1 BLUE JAYS WAY	43.64168	-79.39012	5
2	10495816	2 4 1 Pizza	Food Take Out	1383 DAVENPORT RD	43.67276	-79.44105	3
3	9337469	2 OR 3 PIZZA & WINGS	Food Take Out	2382 LAKE SHORE BLVD W	43.61570	-79.48860	3
4	9000026	2-4-1 PIZZA	Food Take Out	2372 EGLINTON AVE E	43.73204	-79.27102	4

draw a map of pizza shops in toronto

PLOT THE PIZZA SHOPS ON THE



PIZZA SHOPS CLUSTER DATA

```
In [22]: locs_df['Labels'].value_counts()

Out[22]: 5    144
         3     80
         4     77
         2     71
         0     70
         1     65
         7     35
         6     27
         Name: Labels, dtype: int64
```

POI DATASET

[35]:

	WKT	PNT_OF_INT	DESCRPTION	SOCIAL_MED	WEBSITE	CATEGORY	LOCATION
28	POINT (-79.134593 43.774895)	Chesterton Shores Park	Chesterton Shores Park is located east of Port...	NaN	http://www1.toronto.ca/parks/prd/facilities/co...	Park	-79.134593,43.774895
29	POINT (-79.185406 43.780004)	Miller Lash House	The Miller Lash Estate is nestled in the pictu...	NaN	http://www.millerlashhouse.ca	Heritage	-79.185406,43.780004
36	POINT (-79.254642 43.757527)	McCowan Log House	This cabin was built about 1830 in the northea...	NaN	NaN	Heritage	-79.254642,43.757527
37	POINT (-79.254768371582 43.7581710815429)	Cornell House	Built in 1858 in Scarborough Village, this was...	NaN	NaN	Heritage	-79.254768371582,43.7581710815429
38	POINT (-79.2549821 43.7592247)	Thomson Memorial Park	A 41.8 heactare park at Lawrence Ave East and ...	NaN	http://www1.toronto.ca/parks/prd/facilities/co...	Park	-79.2549821,43.7592247
...	...	...	...	...	...	...	...
236	POINT (-79.491232 43.651455)	Old Mill Bridge	The Old Mill Bridge was erected in 1916 during...	NaN	NaN	Architecture	-79.491232,43.651455
237	POINT (-79.49164 43.648877)	Old Mill Subway Station Bridge	The Old Mill Subway Station Bridge was complet...	NaN	NaN	Architecture	-79.491640,43.648877
238	POINT (-79.491983 43.648474)	Bloor Street Bridge	Built in 1924 in an Art Deco style, the Bloor...	NaN	NaN	Architecture	-79.491983,43.648474

CLUSTERS FOR POI

```
7]: exclude_cat = (poi_df['CATEGORY'] != 'Public Art') & (poi_df['CATEGORY'] != 'Creative')  & (poi_df['CATEGORY'] != 'Business')
    poi_df = poi_df[exclude_cat]
    poi_df['Labels'].value_counts()
```

```
7]: 3    102
    4     64
    6     54
    2     25
    0      8
    7      1
    Name: Labels, dtype: int64
```

CONCLUSIONS

- BUILT A K-MEAN MODEL IN ORDER TO MAKE CLUSTERS
- DATA CLEANING
- ASSIGN DATA TO EXISTING CLUSTERS WITH THEIR DISTANCE
- MAKING A MAP OF BUSINESSES LIKE CUSTOMER'S
- IDEAS INCLUDE
  - USE BUSINESS GROWTH DATA IN CORRELATION TO THE CLUSTERS
  - HAVE A RENT SHEET