Battle of Neighborhood

A data science project

Introduction:

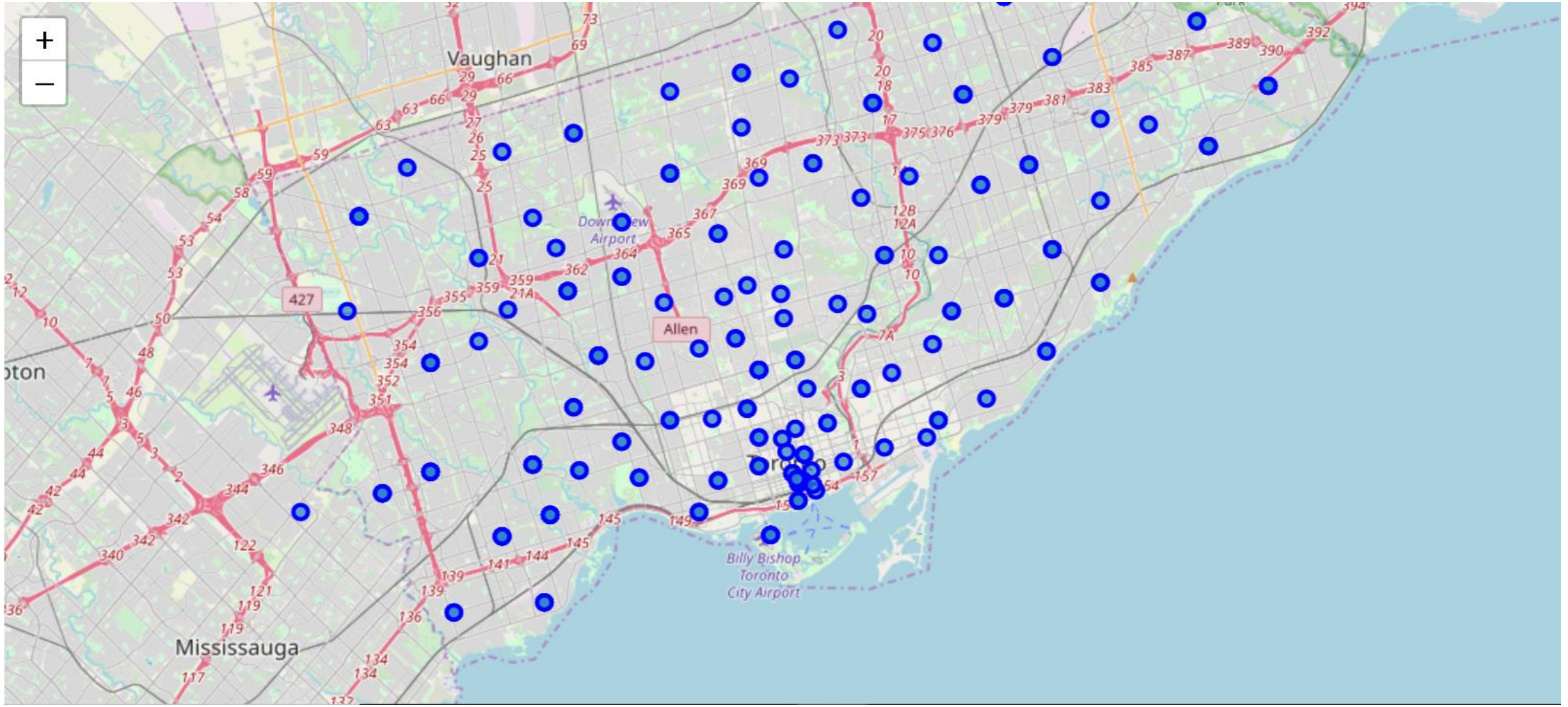
Toronto being financial capital of Canada, it attracts new immigrants as well as travellers. Indians are among top immigrants to Canada. As Indian population is rising in Toronto, want of Indian food & grocery is increasing. A restaurant chain name XYZ, which is popular in India is interested to open an Indian restaurant in Toronto. They want to analyse how the existing Indian restaurants in Toronto are distributed among the neighborhoods.

With the help of data science, we will classify Toronto neighborhoods with high, moderate & low density of Indian restaurants.

Data acquisition & cleaning:

- details of Toronto’s neighbourhood from Wikipedia.
- restaurant details, their type & location in every neighbourhood from **Foursquare API**
- duplicate, blank data rows dropped.
- data filtered for Indian restaurants only.
- after grouping the data contain 201 rows.

Neighborhood considered

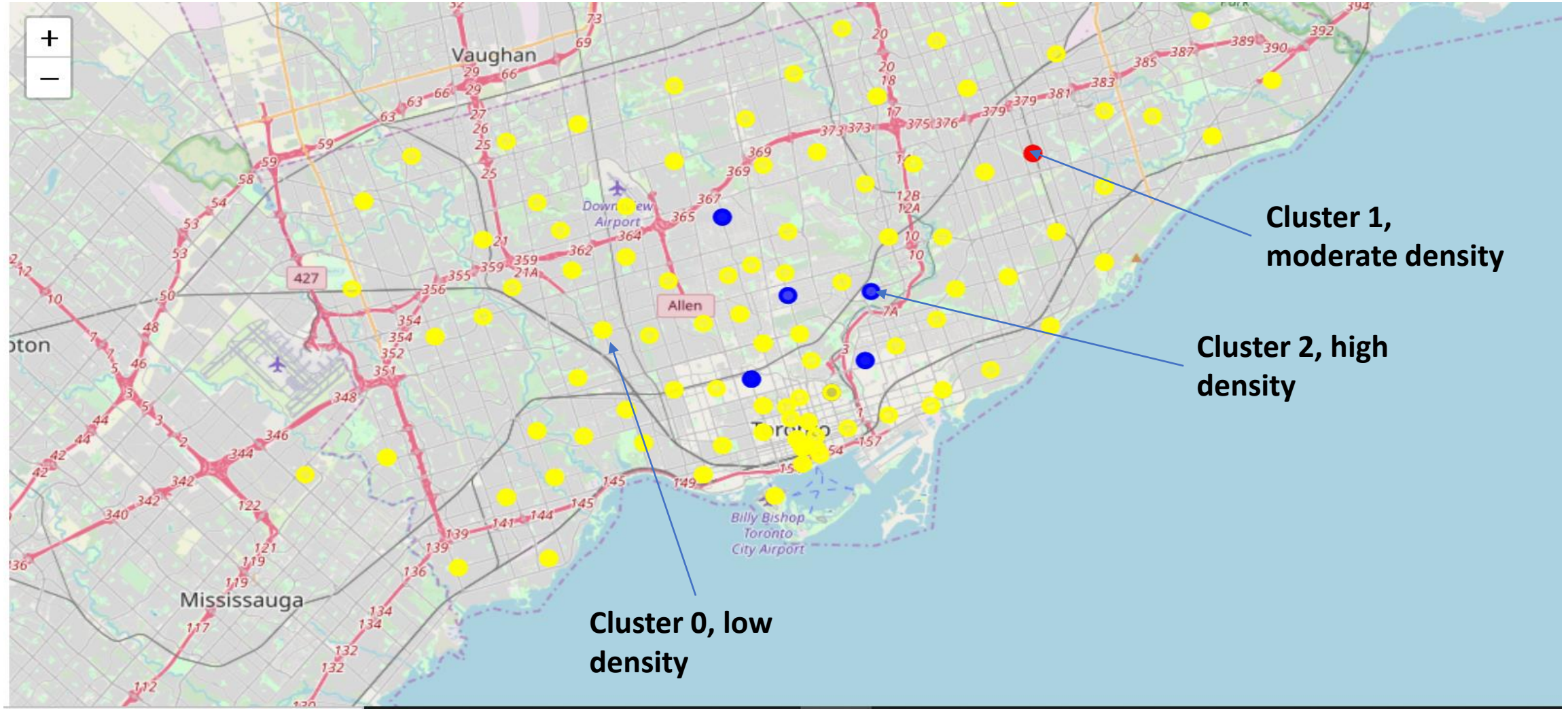


Clustering method:

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into **clusters**.

We use **k-means** method for clustering the data with 3 clusters. **k-means clustering** aims to partition n observations into **k clusters** in which each observation belongs to the **cluster** with the nearest **mean**, serving as a prototype of the **cluster**.

Result:


Vaughan
427
Allen
Mississauga
Billy Bishop
Toronto
City Airport
Cluster 1,
moderate density
Cluster 2, high
density
Cluster 0, low
density

Conclusion and future directions:

There are few neighborhoods like Bedford Park, North Midtown, Yorkville etc with good number of Indian Restaurants. Whereas, other areas have very low numbers. To open a specific cousin restaurant mainly depends on the demand for that particular food. Opening a restaurant where existing Indian Restaurants are very low or none in numbers means low competition but it also means low demand. As the company wants to identify areas where there are high number of Indian restaurants compare to others, we have identified Cluster 2 for them.