Coursera – Capstone Project

Location Identifier – Sport Shop

Problem Statement – WHERE?

Background

- Company ABC is a Multinational Sports equipment and Accessories company with presence across the globe
- Company has decided to enter the Toronto city with its first shop
- Since company does not have any employee in Toronto, company wants to narrow down the locations where it should open a new shop
- Toronto is a large city with area of about 630 sq. km. So manually going through all the neighbourhoods and checking the location is not feasible.
- A great outlet in a wrong place will affect the future of the company in Canada.
- So company decided to follow the following steps for setting up the first outlet
    - Step 1 – Narrowing down the locations suitable for setting up the outlet
    - Step 2 – For the locations identified in step 1, checking the footprint for checking the people visiting the area
    - Step 3 – For the locations identified in step 2, checking the parameters like – rental, space availability etc

Problem Statement – WHERE?

Following problem statement was defined

Which neighbourhoods are best suitable for setting up a new sport shop

This constitutes the first step of finding the location for first outlet in Toronto

Identify Neighbourhoods → Check footprint for 2 months → Check the parameters like rental cost, space availability etc

Data Description

Problem statement – Finding the neighborhoods in Toronto suitable for opening a sports equipment and accessories outlet

Data –

- To solve the above problem, list of neighborhoods in Toronto was collected from the URL – 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
- Location of the existing sports shops and nearby venues were collected using the 'Foursquare API'
  - The search query was used to collect the location of existing sports shops in Toronto. From the resulting set, categories related to sports equipment and accessories were selected while rest were ignored
  - The explore statement was used to find out the nearby venues and their categories for each of the sport shops identified above
- Data on venues located in all the neighbourhoods in Toronto were collected using the 'Foursquare API'
  - List of all the venues in each of the neighbourhood in Toronto was extracted using the explore statement using foursquare API

Methodology – Process Flow

Existing Sport shops were identified using Foursquare API

Locations of the same was plotted on a MAP

Map showed that – Sport shops are not centralized in one neighbourhood but are spread across the city

Venues nearby each of the sport shop identified were extracted using Foursquare API

Distinct categories of the nearby venues were identified

All the identified sport shops showed similar venue categories – shopping centres, restaurants etc

Thus it was decided to divide the neighbourhoods in Toronto with similar venues

Since the requirement was to segment and cluster the neighbourhoods, Kmeans algorithm was decided to use

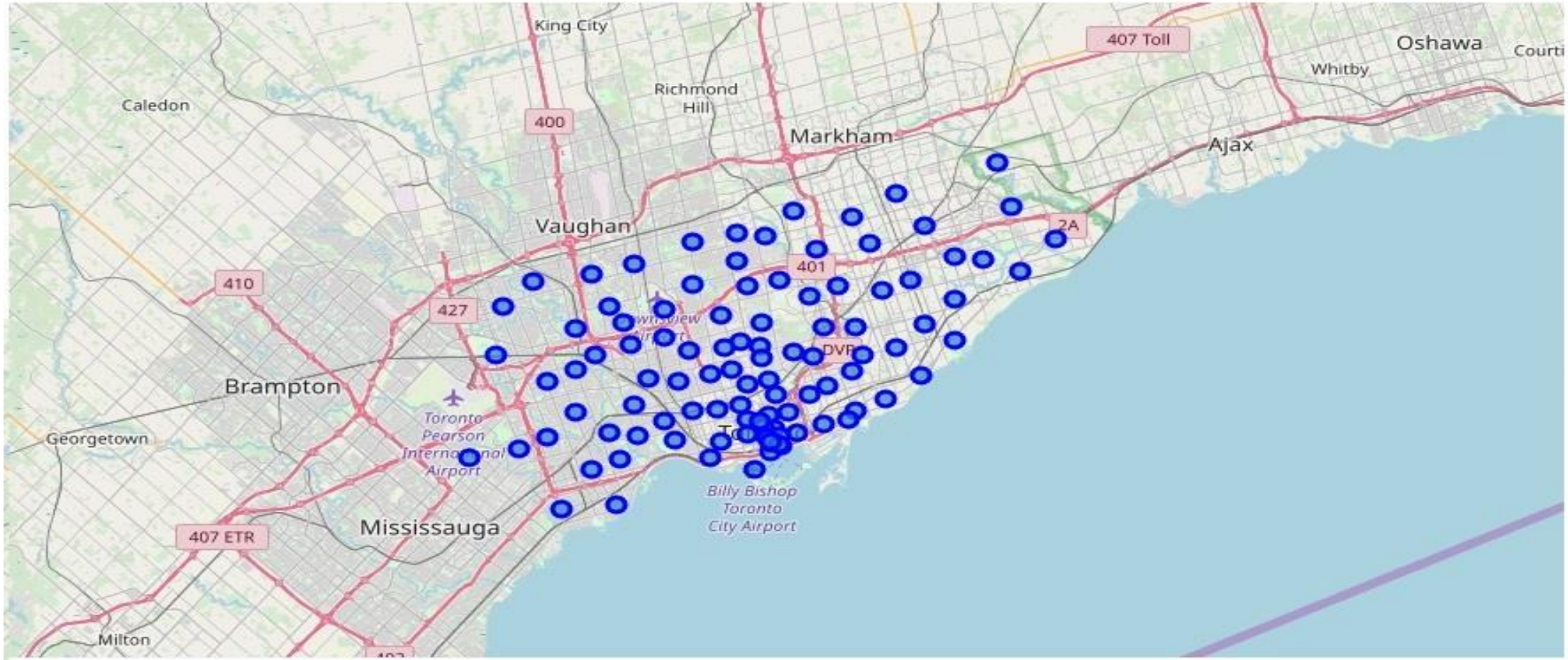
K Means gave 5 clusters out of which characteristics/venues of 1 cluster were suitable for sport shop

Methodology – K Means

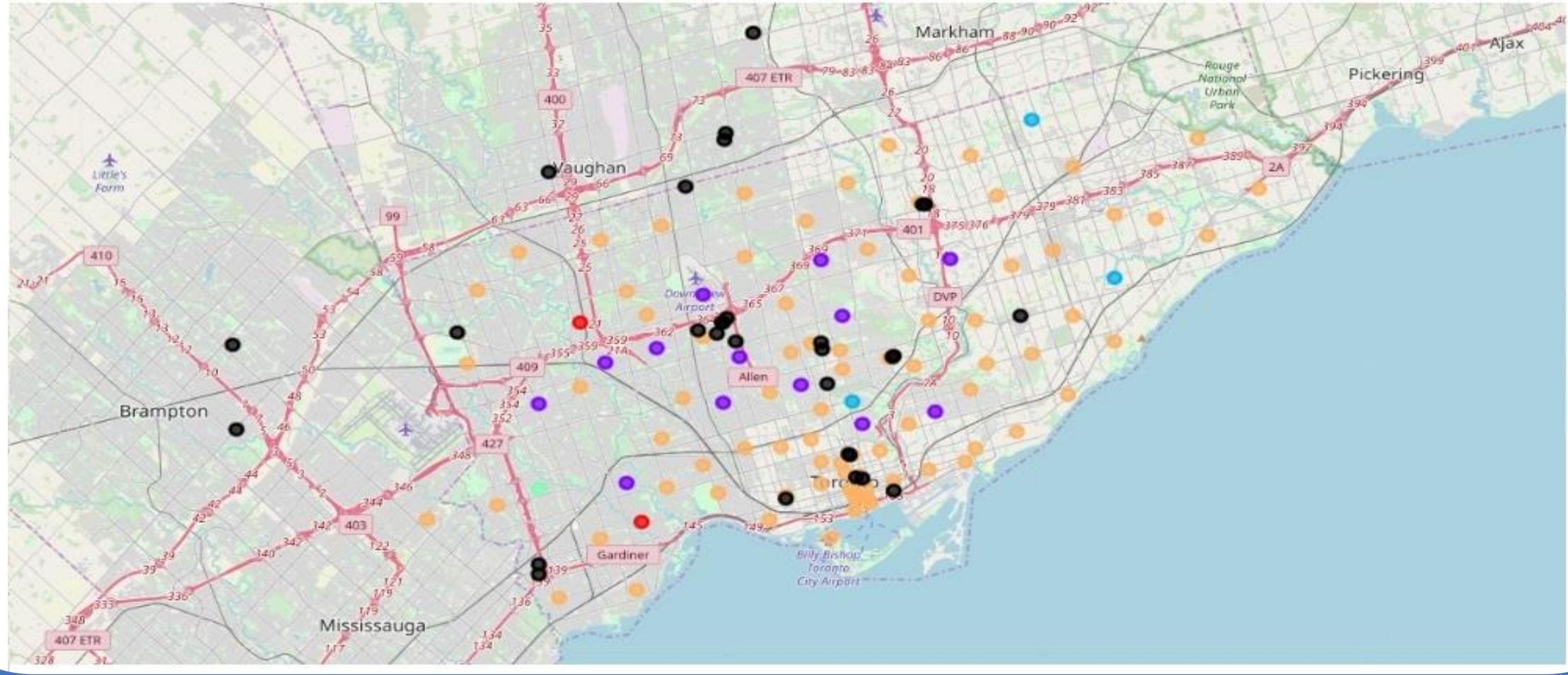
Once the problem statement and data were defined, following methodology was used

- The existing sport shops were plotted to check if they are clustered in one location or are spread across the city. As they were not clustered in one location, we had to carry out further analysis. (Map displayed in next slide)
- Then the venues nearby each of the sport shop were identified using the Foursquare API. These were mainly of similar category like shopping centres, restaurants etc. Thus near by venues can be used to cluster the neighbourhoods into segments. Then find the cluster having characteristics suitable for sport shop.
- As there was requirement of clustering and segmenting the neighbourhoods in Toronto, K Means was used.
- All the venues along with their categories were extracted using Foursquare API. These category values were converted into categorical values and mean of the category venues were calculated for each neighbourhood.
- This data was then analysed using K Means
- K means gave 5 clusters with similar venues in each cluster (Map displayed in next slide)

Methodology – Maps showing distribution of existing sport shops in Toronto



Methodology – Maps showing distribution of clusters and existing sport shops(black) in Toronto



Results

K Means related into 5 cluster of neighbourhoods. Each cluster had similar venues than the other. The following Inferences were drawn

- **Cluster 1, 2,3 –** These clusters did not have many shopping areas or playgrounds. Since the company is planning to open only one outlet for start, it should be in the location either near to playgrounds or shopping centres and restaurants. These areas for this reason was not having existing sports outlet.
- **Cluster 0 –** This cluster has neighborhoods with Baseball fields. However there are no much restaurants or other venues which attracts people. However the baseball players do frequent this area. However given the climateic conditions this area has high footprint in summers only. Company decided to keep this these locations for future Baseball themed outlet.
- **Cluster 4** – These are the neighborhoods where the existing sports shops are spread and the neighborhoods with similar venues like playgrounds, shopping centres, restaurants etc. Thus these neighborhoods attract people who are out for SHOPPING. Hence neighborhoods from this cluster are considered for next step.

Next Step – Further data from neighbourhood of cluster 4 like rental cost, space availability and footprint (using Foursquare Trending for 2 months everyday) the Location of the First Sport Outlet in Toronto will be Finalized.