IBM Data Science Capstone Project

Using k-means clustering to find the best area in Toronto to expand to for a food delivery company.

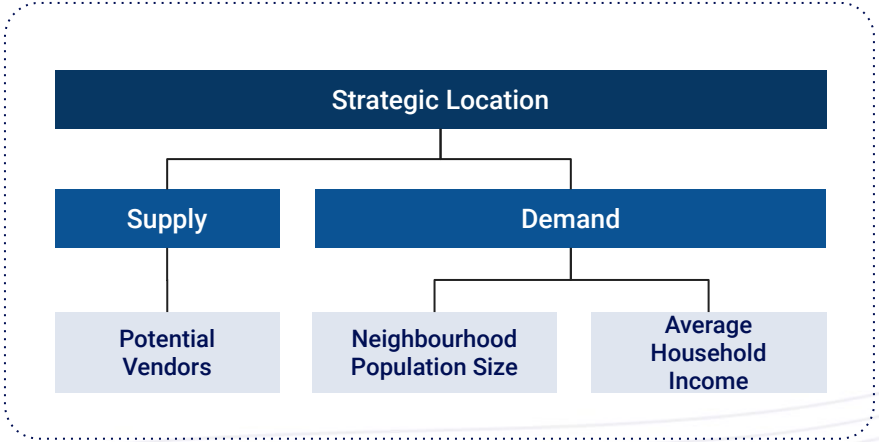
Table of Contents

Food Delivery Company: Strategic Expansion

Context

In this project, we will hypothesise that a new food delivery company, like that of Uber Eats decides to expand to the city of Toronto in the midst of a COVID-19 lockdown. To ensure that every dollar is strategically spent, this food delivery company must selectively choose the most optimal regions for expansion in order to maintain a high gross profit margin.

Issue Tree



Problem Statement

This project seeks determine the most strategic region in Toronto for a food delivery company to expand to.

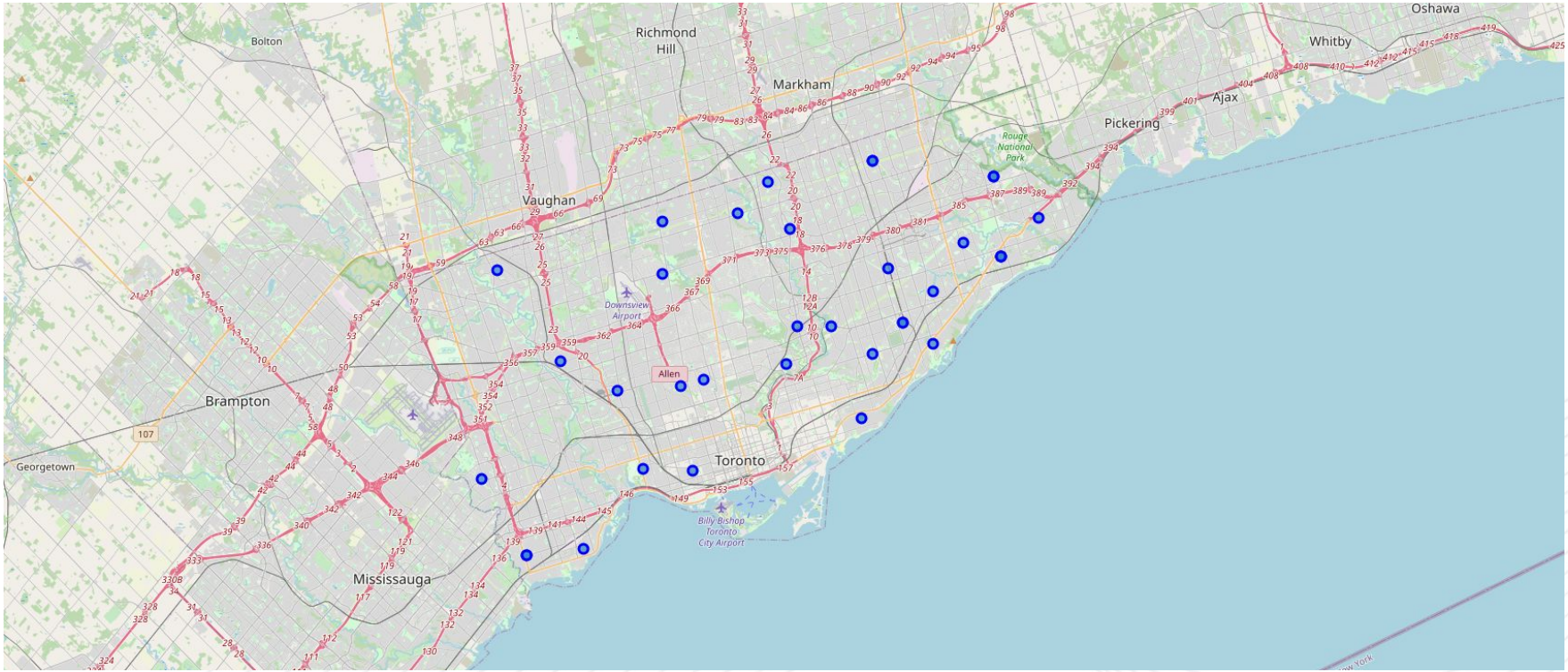
Table of Contents

Data Acquisition and Cleaning

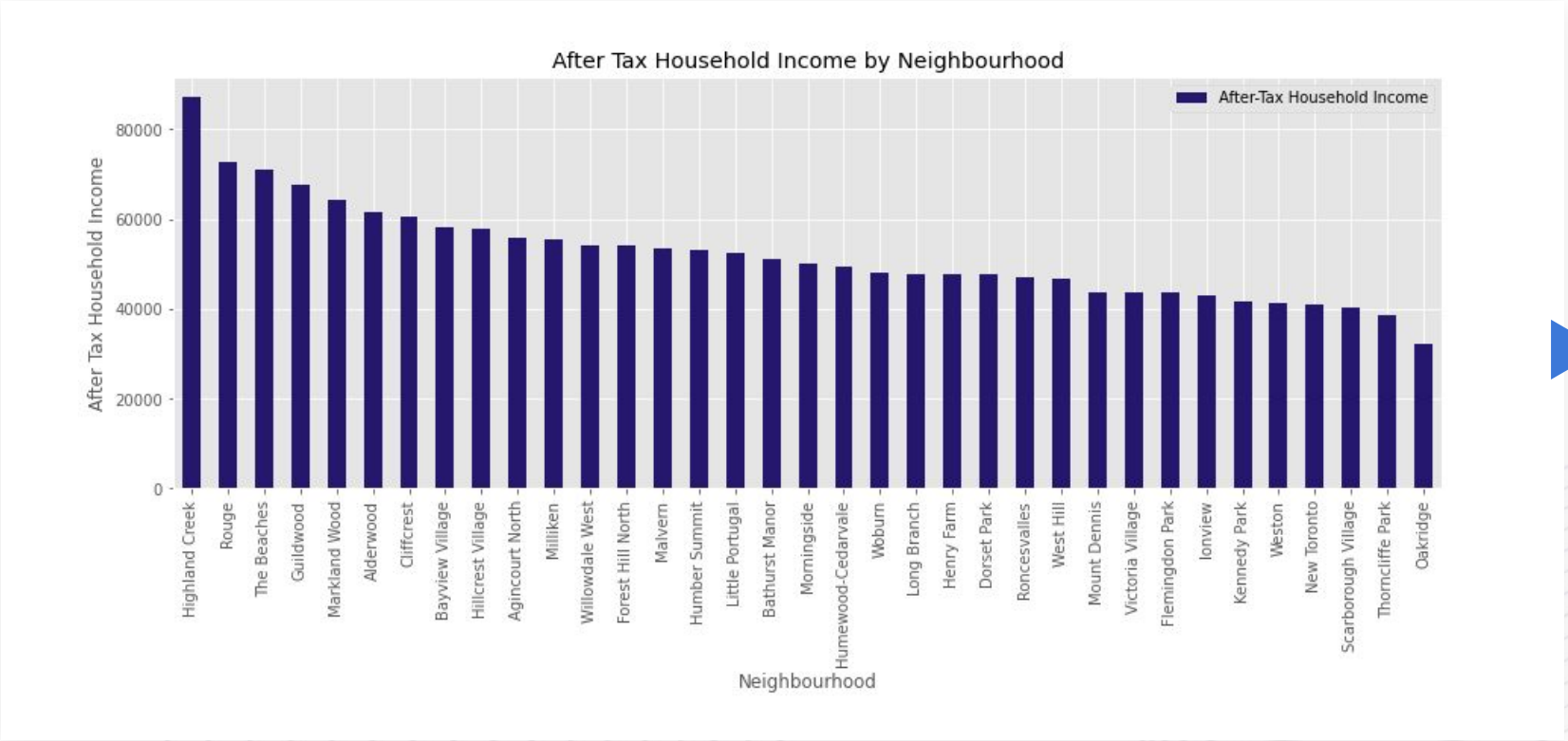
1. Toronto neighborhood data was web scraped from Wikipedia
2. Number of nearby restaurants from each neighbourhood was extracted via Foursquare API
3. Household income and neighbourhood population size data was extracted from Toronto Wellbeing
4. Household income, neighbourhood population size and nearby restaurants are all merged into a single data frame
5. The clean data contains 3 features (household income, population size and restaurant count)

Table of Contents

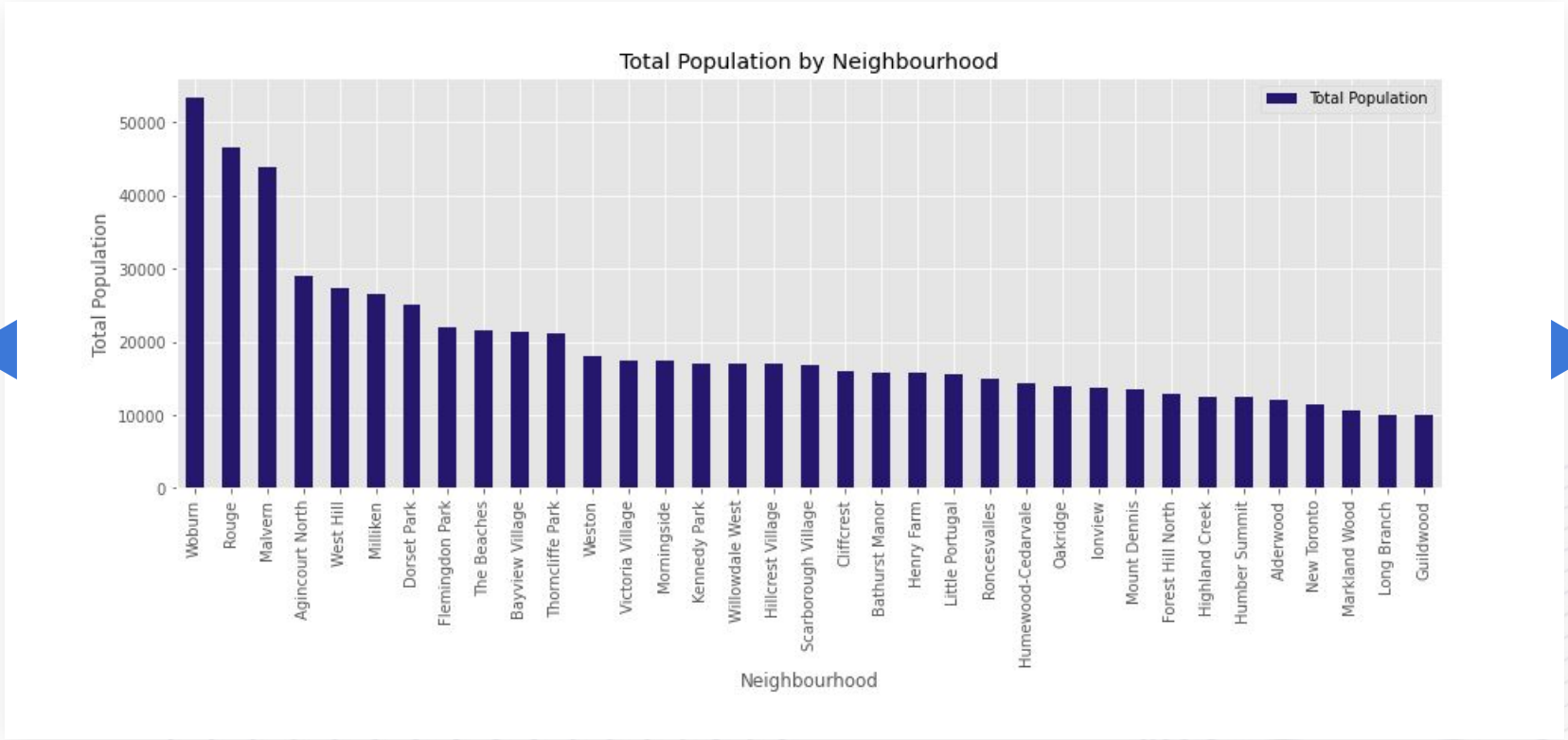
Exploratory Data Analysis: Folium Mapping



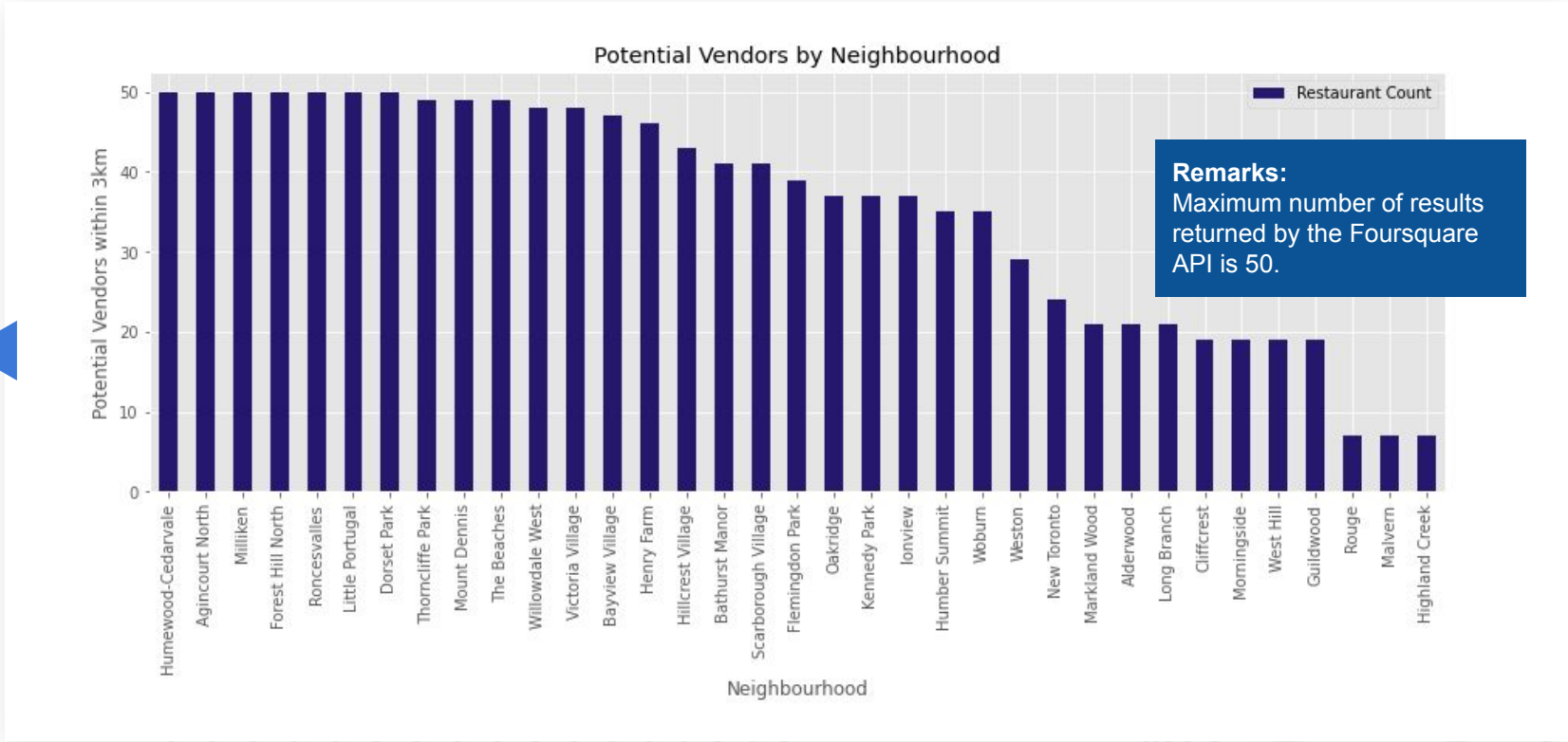
Exploratory Data Analysis: Household Income by Neighbourhood



Exploratory Data Analysis: Total Population by Neighbourhood



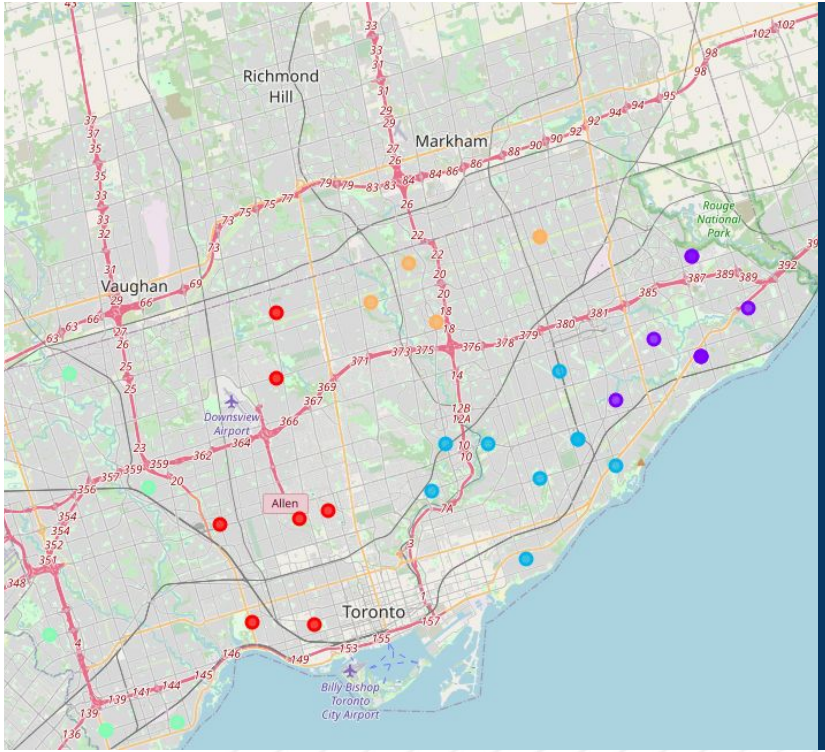
Exploratory Data Analysis: Potential Vendors by Neighbourhood



Remarks:
Maximum number of results returned by the Foursquare API is 50.

Table of Contents

K-Means Clustering



Remarks:
5 clusters returned based on Elbow method

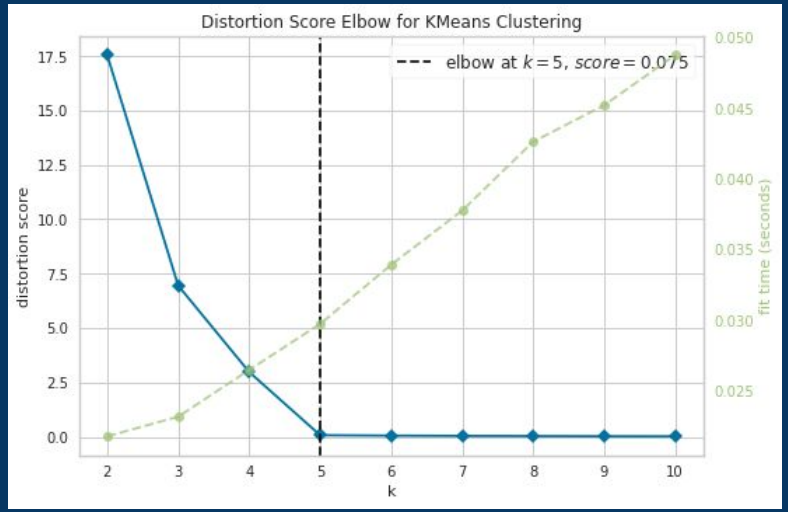
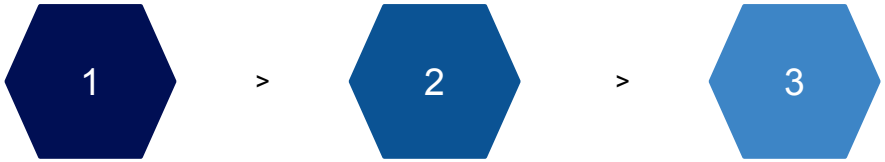


Table of Contents

Analysing Results



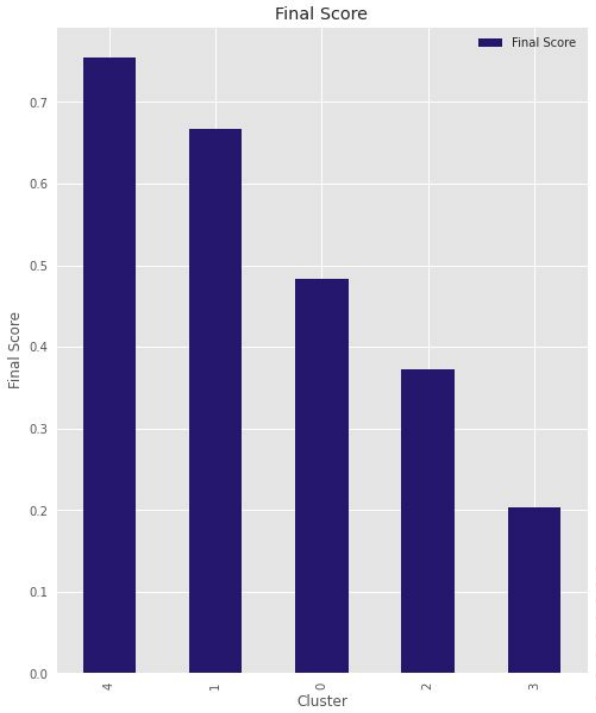
Normalising feature data (population size, household income, restaurant count)

Calculating the mean score of each cluster (equal weightage for each feature)

Visualising results with a bar chart

Table of Contents

Conclusion and Future Directions



Results

Cluster 4 is the best region to expand to, followed by cluster 1. Both clusters have a final score of more than 0.5, therefore are considered most optimal.

Future Directions

- The model assumes that only residents of the neighbourhood will issue delivery orders. While this is true during a pandemic lockdown, the model will become obsolete once the lockdown is over since people moves around throughout the day. (i.e. working at the office). Therefore it might be wise to consider the distribution of people in Toronto throughout the day during normal circumstances.
- Population size and average household income alone produces limited results. Realistically speaking, demand can also be driven by factors such as popularity of available restaurants, age and delivery fees.
- A larger radius can be used to locate more venues. A radius of 10km would produce results that are more realistic.