

IBM Capstone Project

Identify Good Locations To Open A New Flower Shop

Chi Ho Man

15 July 2019

1. Background

In this capstone project, I have used geographic data from **Foursquare API** (<https://developer.foursquare.com/docs/api>) to identify good locations for opening a flower shop in the Sydney city area. By analysing the surrounding neighbourhoods of Sydney's train stations and flower shops, we will find other suitable locations which are similar to those with flower shops operating, but having no or less competition.

2. Problem Description

When we consider opening a new business, we usually have a good understanding on the business we want to start. For example, a florist who wants to start her own business might had already worked in a flower shop for a number of years, she might already have good knowledge about flowers and suppliers of running a flower shop business.

However, when it comes to location, it is always difficult to judge whether it is good or bad to start the same business in a location which we are not familiar with. Characteristics of another location, such as client type, taste, or even the primary reasons for buying flowers may be completely different.

While it is almost impossible to spend time researching all different locations, using data analysis would provide a good starting point to identify locations which have higher chance of success, as we can start by identifying locations which share higher similarities with those already running.

3. Data

3.1. Data Description

To set this project in a more manageable scope, I have limited the target area to be the surrounding areas (within 500m) of all train stations within 50km of the Sydney CBD (*latitude: -33.86785, longitude: 151.20732*). The rationale of choosing surrounding areas of train stations is because higher pedestrian traffic are in general more favourable for operating flower shop businesses. In this project, we will be collecting the following data:

1. Train stations geographic data in the target area.
2. All flower shop locations in the target area, and identify those within 500m of any train stations.
3. Other surrounding venue information i.e. types of venues within 500m of all train stations (to cluster stations by similarities surrounding neighbourhoods)

3.2. Data Collection

All data sources will be sourced from the Foursquare API. We will obtain the geographic data and other venue data e.g. type of venues.

3.3. Data Cleaning

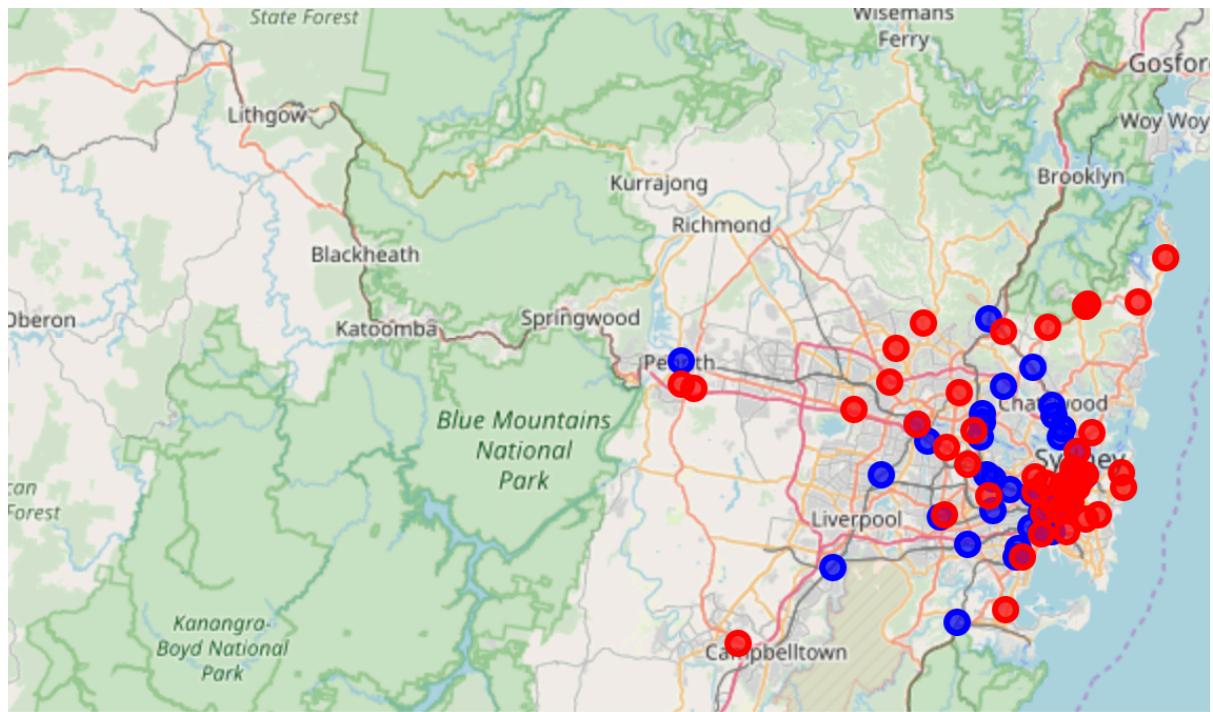
User generated data from Foursquare API are not completely consistent. For example, A venue categorised as “train station” can sometimes be referred to a platform, a interchange or actually a train station.

Data points collected which are not train stations are dropped from the dataset.

4. Exploratory Data Analysis

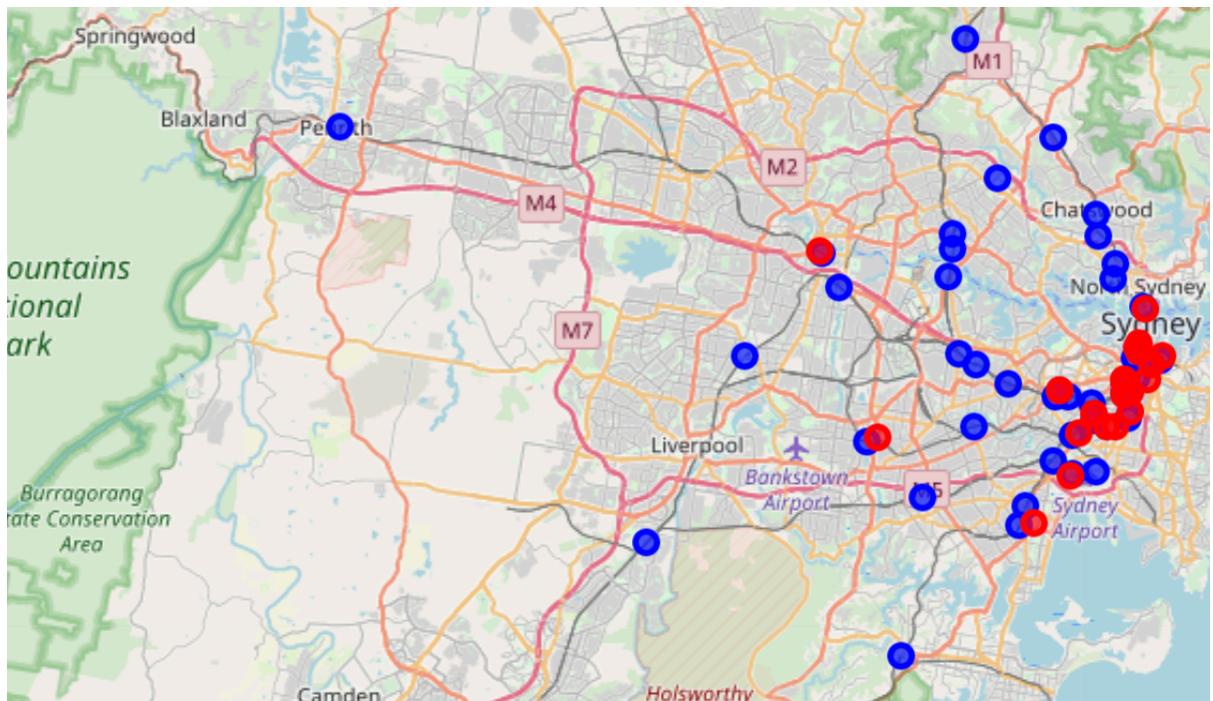
4.1 Visualising Locations

To understand the distribution of train stations and flower shops in the target area, I have plotted them into a map using Folium. One general observation we can draw from the map is, most flower shops are either located nearby CBD, heavily populated areas, which usually along train lines or major roads



Sydney train stations (blue) and flower shops (red)

To find out how many flower shops are opened along the train lines, we have plotted those within 500m of any train stations into another map. We found that out of 50 flower shops, 21 (i.e. 42%) of them are operating along train lines.

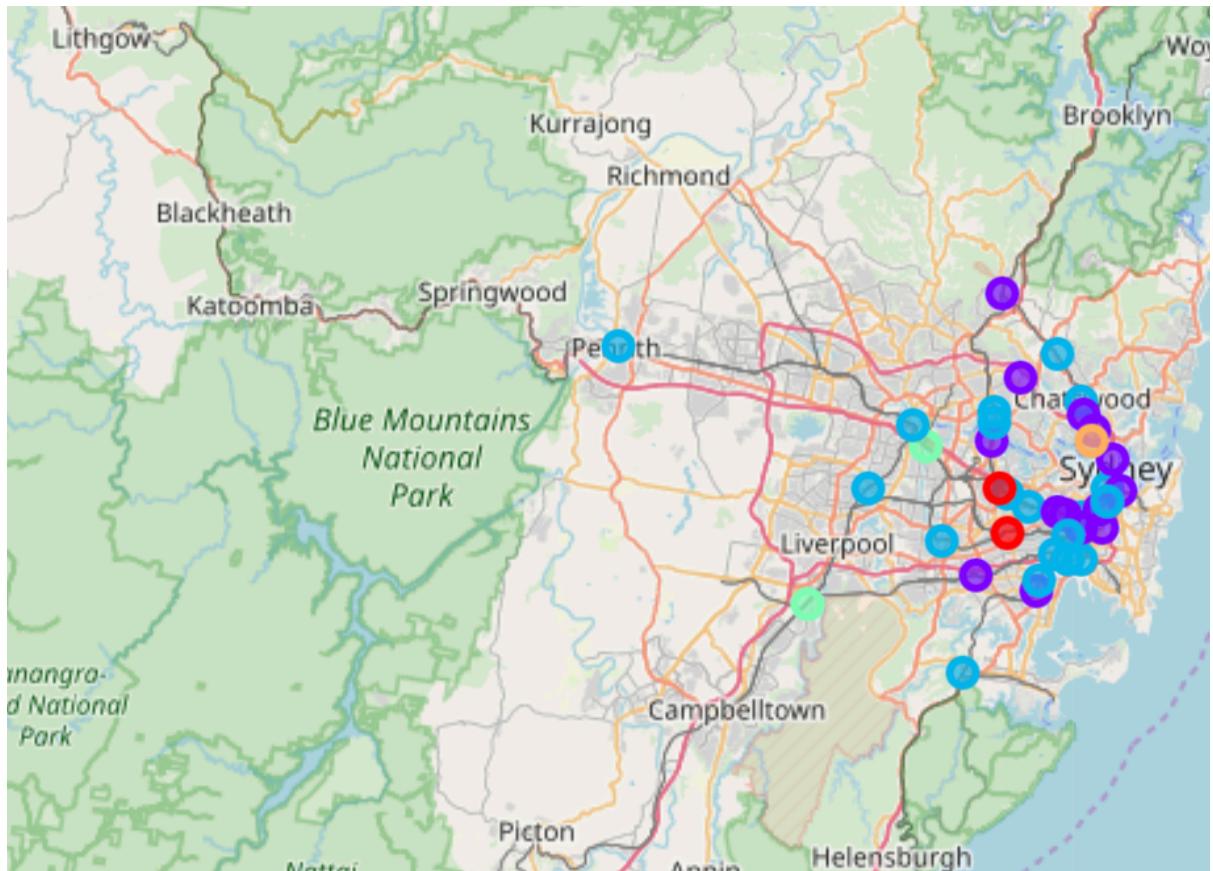


Flower shops (red) that are within 500m of one or more train stations

5. Problem Solving Approach

To solve the identifying shop location problem, the approach of using the data would be:

1. Collecting the types of venues of different train stations, perform one-hot encoding to encode the top 20 "surrounding neighbourhoods" of each train station.
2. Using k -mean clustering to put each train stations in a cluster. Stations should have high similarity with others within it.



Stations are classified into different clusters (one colour per cluster)

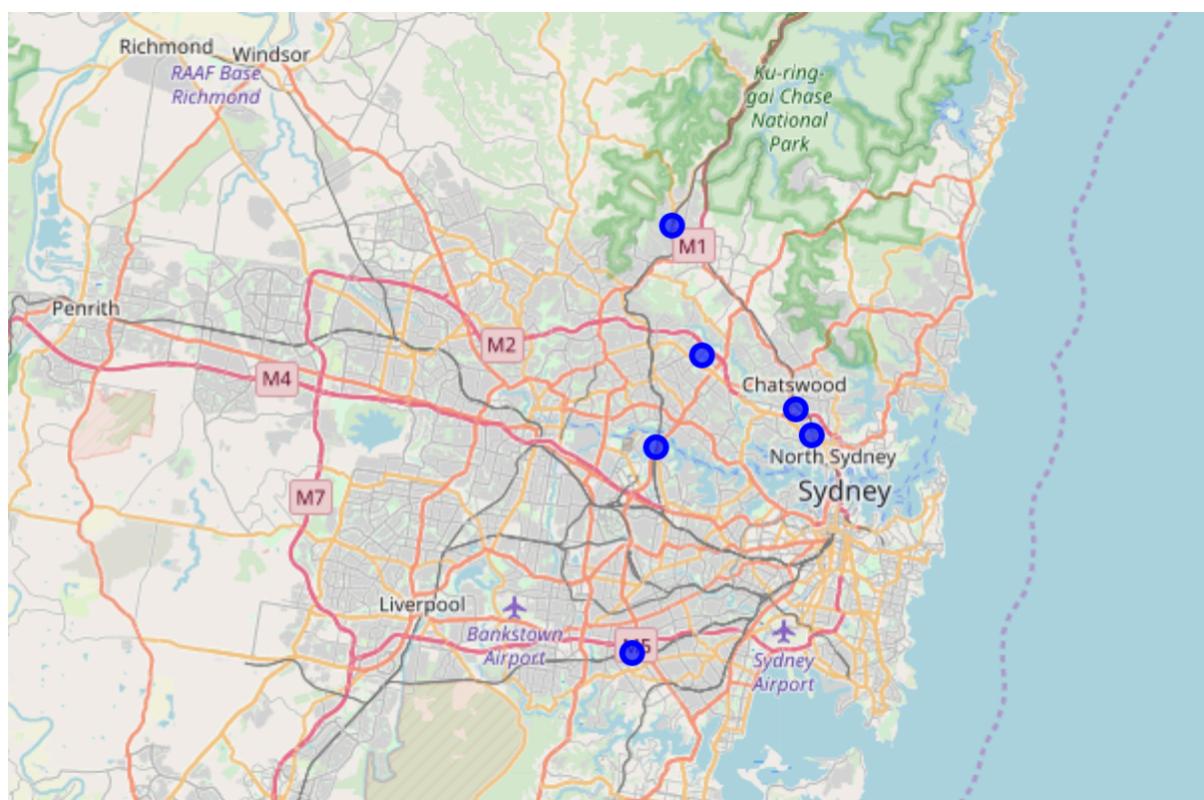
3. Calculating the average concentration of flower shops in each cluster and identify the cluster with the highest density of flower shops (best cluster). From the table below, cluster 1 is the best cluster for operating flower shops as it has the highest average number of flower shops per average, there are 14 stations in this cluster.

Table 1. Cluster, Train Stations, Flower Shops

Cluster Number	Number of Train Station	Average Number of Flower Shops
Cluster 0	2	0.0000
Cluster 1	14	1.0000
Cluster 2	19	0.5789
Cluster 3	2	0.0000
Cluster 4	1	0.0000

4. Within the best cluster, find out the list of stations (locations suitable for opening flower shop) with no flower shop within 500m of the stations.

- Rhodes Stations
- Artarmon Station
- Macquarie University Station
- St Leonards Station
- Narwee Station
- Hornsby Station



Locations of stations suitable for opening a flower shop nearby

6. Conclusion

In the project, by using the geographical and informational data from Foursquare, we have identified 6 stations with similar surrounding as those stations with high concentration of flower shop but no having one nearby these stations. This would provide a good starting point for someone who want to open a business in locations that he/she is not familiar with.