Looking for a property in Sydney

Aug 2019

Background

- A client came from Hong Kong is looking for a property in Sydney close to a train station within a radius of 20km from the Sydney central business district (CBD). To decide which location to look at, the clients wants to have a sense of community and neighborhood of all train stations within 20km from the Sydney CBD. The client approaches our property agency for the information.

Problem

- Data analyst in the property agency
- Provide a neighborhood analysis of all train stations within a radius of 20km from the Sydney CBD.
- This project is timely and we have very limited recourse
- Online available data and Foursquare API
- Apply clustering techniques to group the stations into different clusters

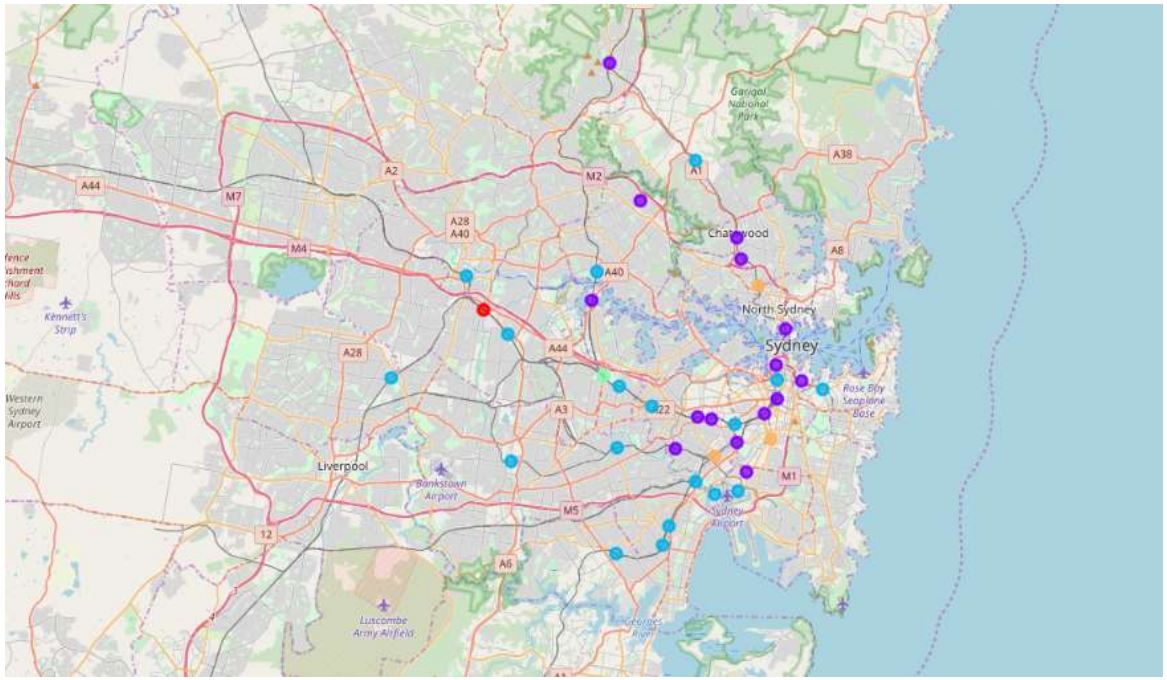
Data

- Geographical coordination (i.e. Latitude and longitude coordination) of Sydney CBD
- List of train stations within a radius of 20km from the Sydney CBD
- List of neighborhoods (i.e. venues data) for each train stations

Methodology

- Unsupervised machine learning method
- k-means clustering algorithm
- One of the simplest and popular unsupervised machine learning algorithms
- It is easy to understand and delivers results quickly
- Group similar data points together and discover underlying similarity
- Number of clusters = 5

Visualize the train stations on map



Result

- Cluster 1: Clyde Station
- The most common venues in this cluster are gym, Asian restaurant, rental car location. Compared to other clusters, there are not many food options nearby.

Result (cont'd)

- Cluster 2: Wynyard Station, King's Cross Station, Chatswood Station, Milsons Point Station, Artamon Station, Rhodes Station, St Peters Station, Redfern Station, Mascot Station, Petersham Station, Central Station, Stanmore Station, Hornsby Station, Dulwich Hill Station and Macquarie University Station
- There are heaps of good cafés, coffee shops, pubs in this cluster. Compared to other clusters, this cluster has a wide range of food options including Italian, Thai, Japanese and Asian restaurant.

Result (cont'd)

- Cluster 3: Town Hall Station, Hurstville Station, Ashfield Station, Bankstone Station, Auburn Station, International Airport Station, Wolli Creek Station, Kogarah Station, Campsie Station, Parrrmatta Station, Fairfield Station, Domestic Airport Station, Burwood Station, Edgecliff Station, Newtown Station, Gordon Station, Meadowbank Station and Rockdale Station
- This cluster has a high concentration of Asian food, including Chinese, Thai, Japanese, Vietnamese and Korean restaurant. This is easy to find daily necessities nearby as we can see supermarket and grocery store are some of the most common venues.

Result (cont'd)

- Cluster 4: Strathfield Station
- This cluster has high concentration of Korean restaurant and café. Strathfield is also known as "Little Korea" in Sydney.

Result (cont'd)

- Cluster 5: Green Square Station, Sydenham Station, St. Leonards Station
- The most common venues in this cluster is café. We can access to gym easily. We can also find farmers market and fish market in some of the places.

Conclusion

- This analysis gives the client a brief overview on the characteristics of different communities within 20km radius of Sydney city Centre
- The API we choose will impact what type of data we obtain
- The foursquare API has details information on restaurants and shops while it does not have much information on public facilities
- Explore new data from other APIs so that we can perform a more comprehensive analysis