IBM Data Science Course Capstone Project

Location Choices for

Chinese Hamburger Chain Stores

in Sydney, Australia

Cheryl Wang

June 21, 2021

# Introductions

This project is part of the completion of the IBM Data Science Course Capstone Project.



Chinese Hamburger, Rou Jia Mo 肉夹馍 is a traditional Chinese fast food. It is semi pre-cooked, easy to grab and go, and it is a delicious fit for breakfast, lunch, or a simplified dinner. The flavor of your choice indulges the nostalgic appetite of Asians, and also easily accepted by other cultures.

A successful businessman wants to open a series of Chinese Hamburger restaurants in Sydney, Australia. The restaurant could be as simple as a food truck in busy areas, or a spacious single business housing with dine in options if real estate is affordable.

The intent is to find a number of locations in Sydney that could fit the Chinese Hamburger restaurants or food trucks.

# Data

The following data was collected to perform this analysis:

1. List of neighborhoods of Sydney.

2. Geo-coordinates of the neighborhoods of Sydney.

3. Top Venues of the neighborhoods.

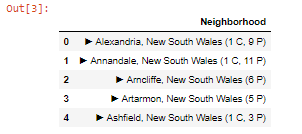
List of the Sydney neighborhoods were obtained from <https://en.wikipedia.org/wiki/Category:Suburbs_of_Sydney>

# Methodology

## 3.1 Obtain the Neighborhood Data

The List of the Sydney neighborhoods were obtained from <https://en.wikipedia.org/wiki/Category:Suburbs_of_Sydney> . Web-scraping was performed using Beautiful Soup, and future parsed into a Pandas Data Frame.

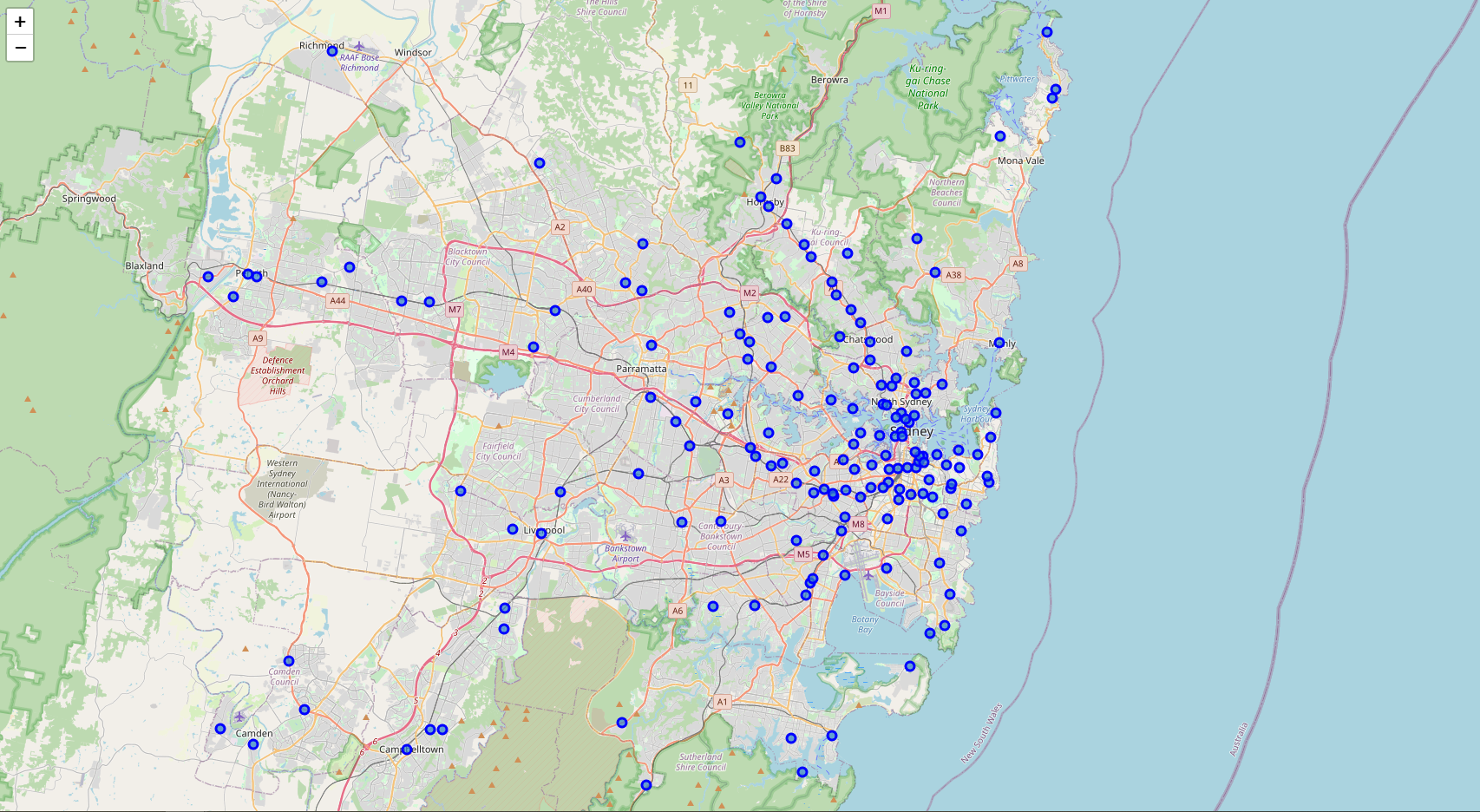




## 3.2 Map the Neighborhoods

Geocoder is used to obtain the Latitude and Longitude of the neighborhoods, and mapped the locations through *folium*.



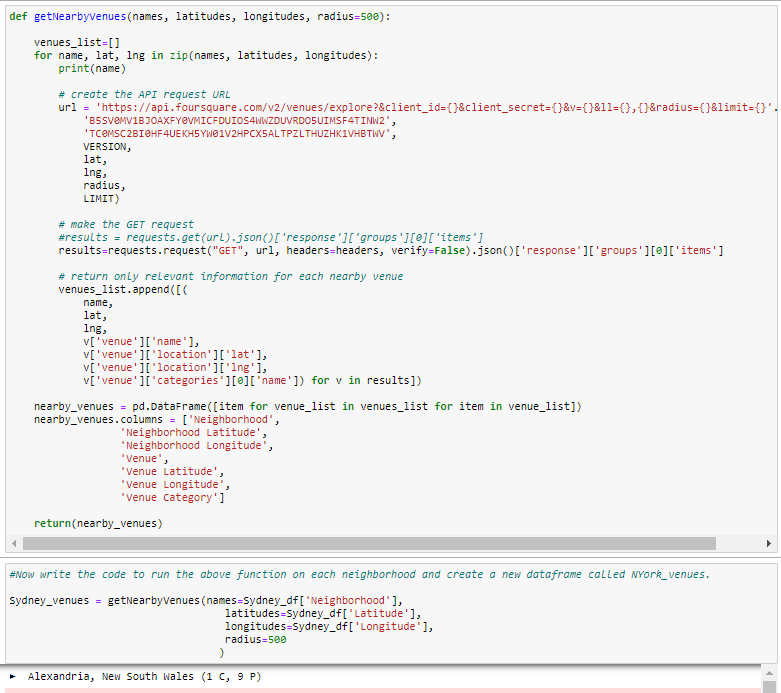


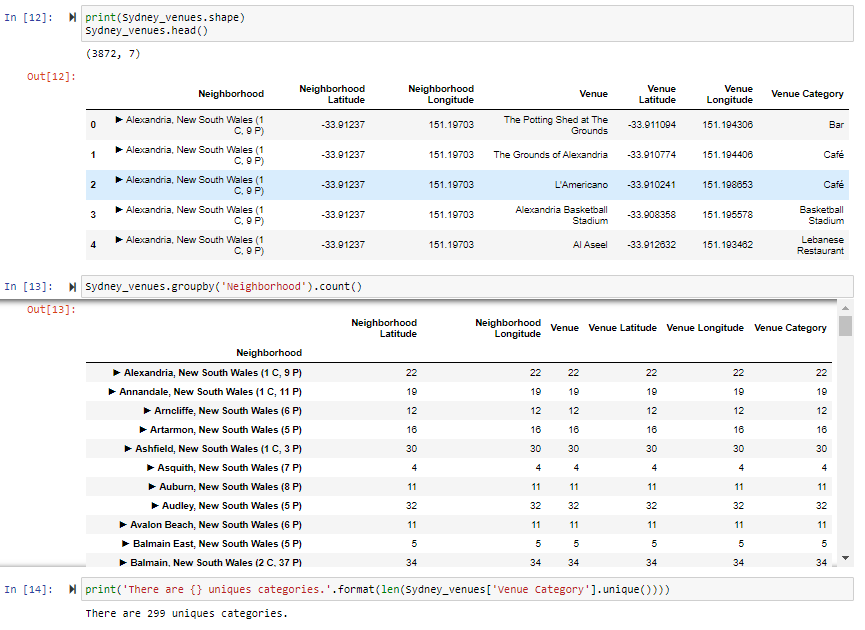
# 

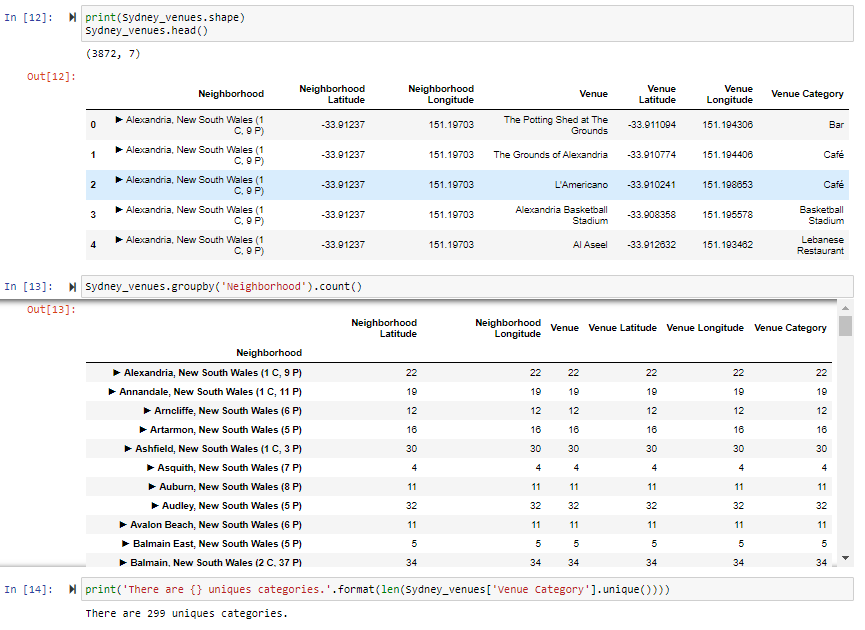
## 3.3 Explore the neighborhoods

The Foursquare API was used to explore the neighborhoods and obtain the top venues of each neighborhood.



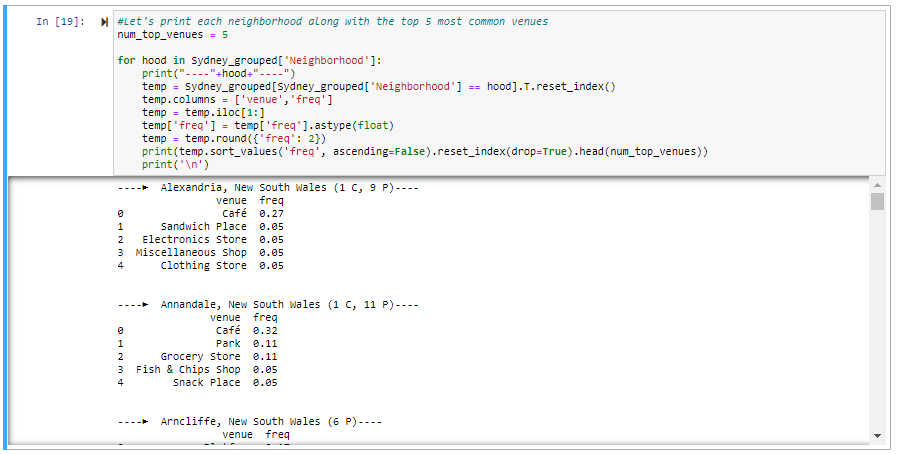


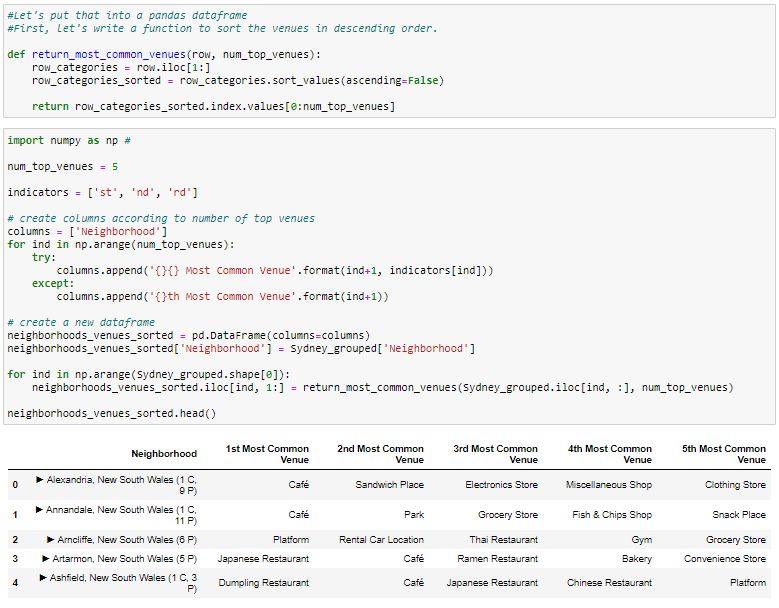




## 3.4 Analyze each neighborhoods

The top 5 venues of each neighborhood is listed, and put in order.

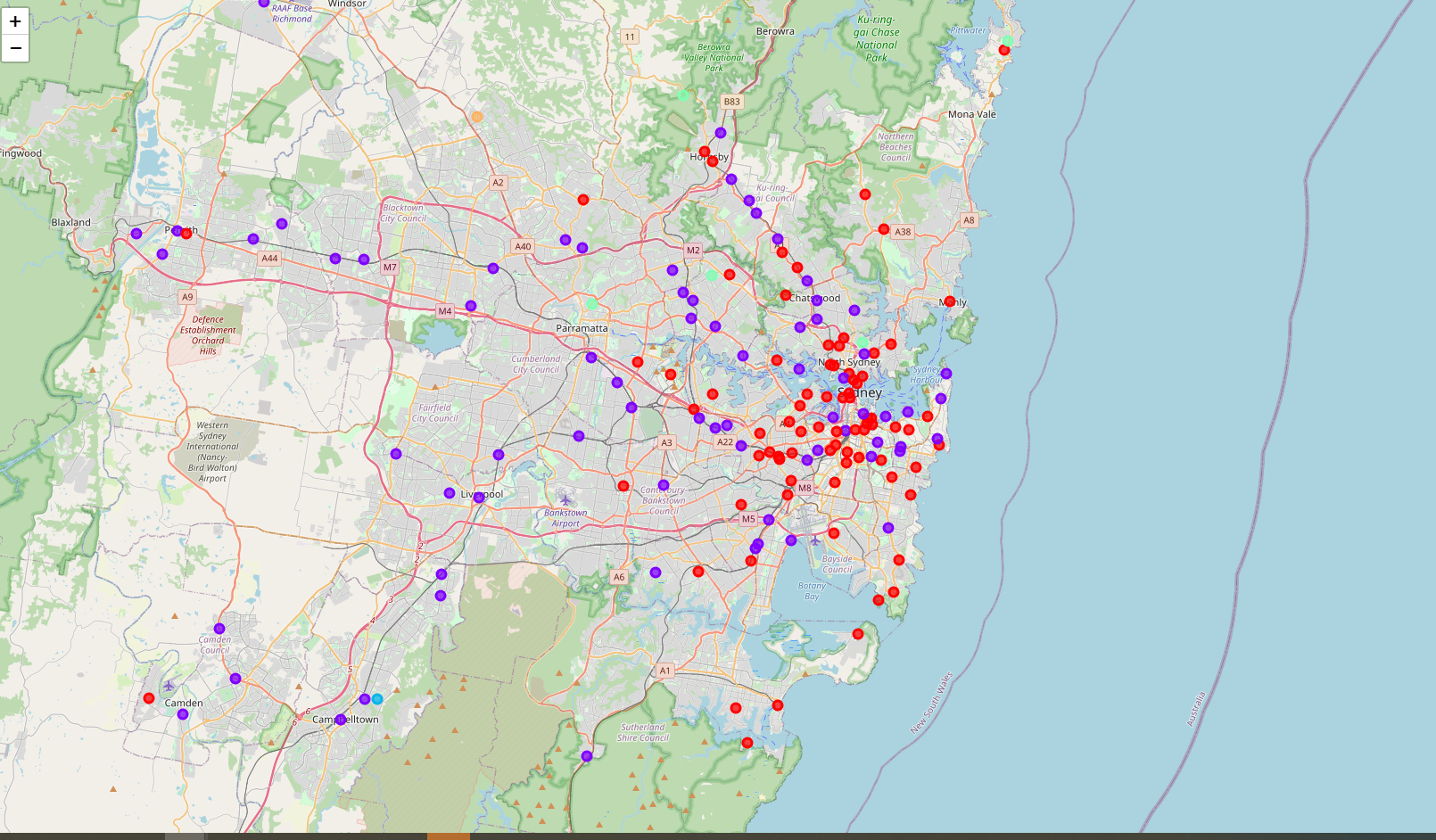




## 3.5 Cluster neighborhoods

Kmeans is used to clusters the neighborhoods and grouped the 163 neighborhoods into 5 clusters.

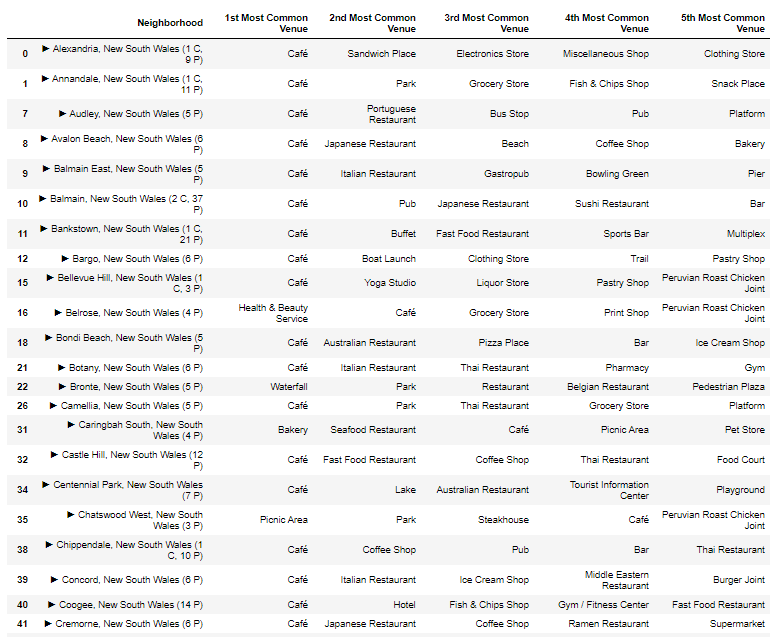




# Results

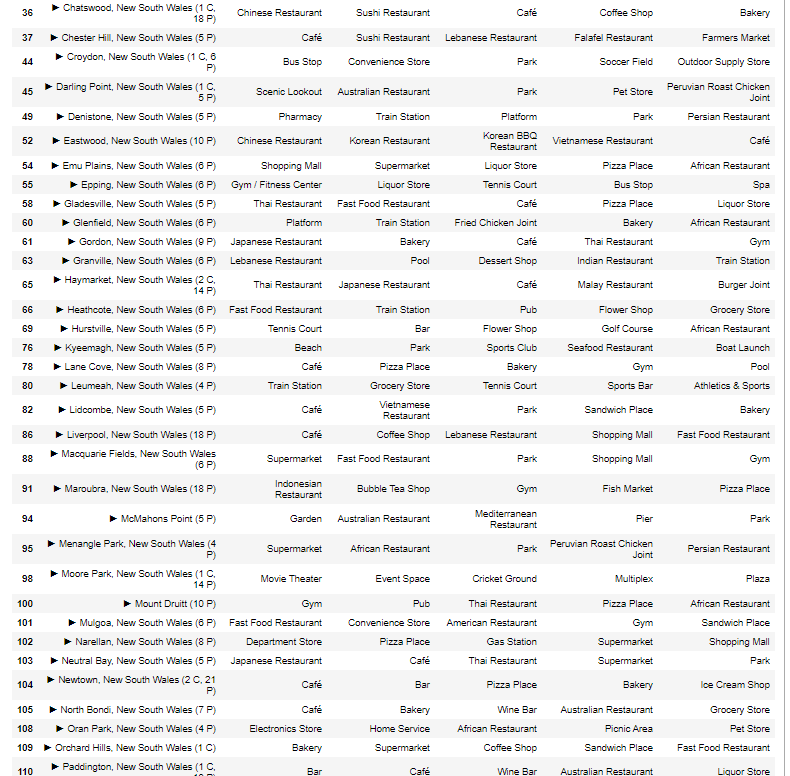
## 4.1 Neighborhoods Cluster 1

Cluster 1 has 79 neighborhoods. Overwhelming trend is that the most common venue is Café, from which we can assume these are the areas people mainly go there for food, and it has a plethora of good restaurants already.



## 4.2 Neighborhoods Cluster 2

Cluster 2 has 73 neighborhoods. The neighborhoods in this cluster has a variety of top venues, but most of them have food/drink as one of their top 5 venues. From here, we can assume these are the areas people go for activities and grab a meal as convenience. Therefore it is a good choice to have fast food restaurants.



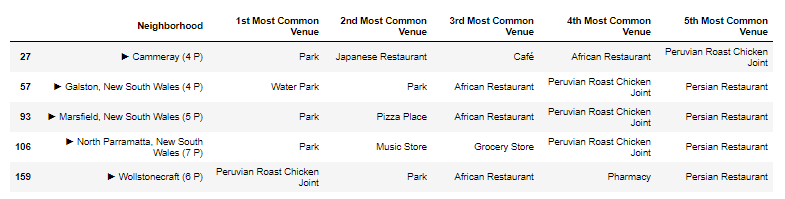
## 4.3 Neighborhoods Cluster 3.

Clusters 3 have 1 neighborhoods, and the top venue is a Food and Drink Shop.

## 

## 4.4 Neighborhoods Cluster 4

Clusters 4 have 5 neighborhoods, and the top venues are park.



## 4.5 Neighborhoods Cluster 5

Clusters 5 have 1 neighborhood, and the top venue is a bus stop.



# Discussions and Recommendations

From the clustering above, Neighborhoods in cluster No. 2 are good choices for a Chinese Hamburger restaurants. It is where people go for errands/leisure and with options for a meal. It has restaurants from all cultures but not focusing on a specific one. So Chinese fast food could be a good addition.

6. Conclusion

The Python libraries and Kmeans tool is really helpful in performing this neighborhoods analysis. Room for improvement is to better understand how the neighborhoods are clustered, ways to optimize the number of clusters. Also the categorization of venues has confusion as well. How are café and restaurants differentiated; why does food and drink shop get singled out, etc.