Coffee Shop Business in Sydney CBD and surrounding Suburbs

Andrew Au
IBM Data Science Capstone

Introduction

According to statista, the Australia coffee market is among the largest in the world, reaching a revenue greater than 1.4 Billion US dollars in 2017. Furthermore, Australians consumed around 1.91 kilograms of coffee per person in 2019 on average. In particular, residents in Sydney are more inclined to purchase fresh coffee based on a research published by Roy Morgan Single Source. As a result, this represents lucrative opportunities for entrepreneurs who would like to establish their coffee shop business in Sydney.

The purpose of this report is to **help potential entrepreneurs in Sydney who wish to set up their own coffee shop**. This report will provide them with insights on the **optimal suburb that they can consider in establishing their coffee business**.

The rationales behind this analysis are twofold, (1) **entrepreneurs should establish their coffee shop in suburb with venues that are able to drawn large amount of foot traffic and (2) Avoid suburbs with large numbers of coffee shop.**

Data Description

In order to approach the problem, I have extracted relevant data by the following ways:

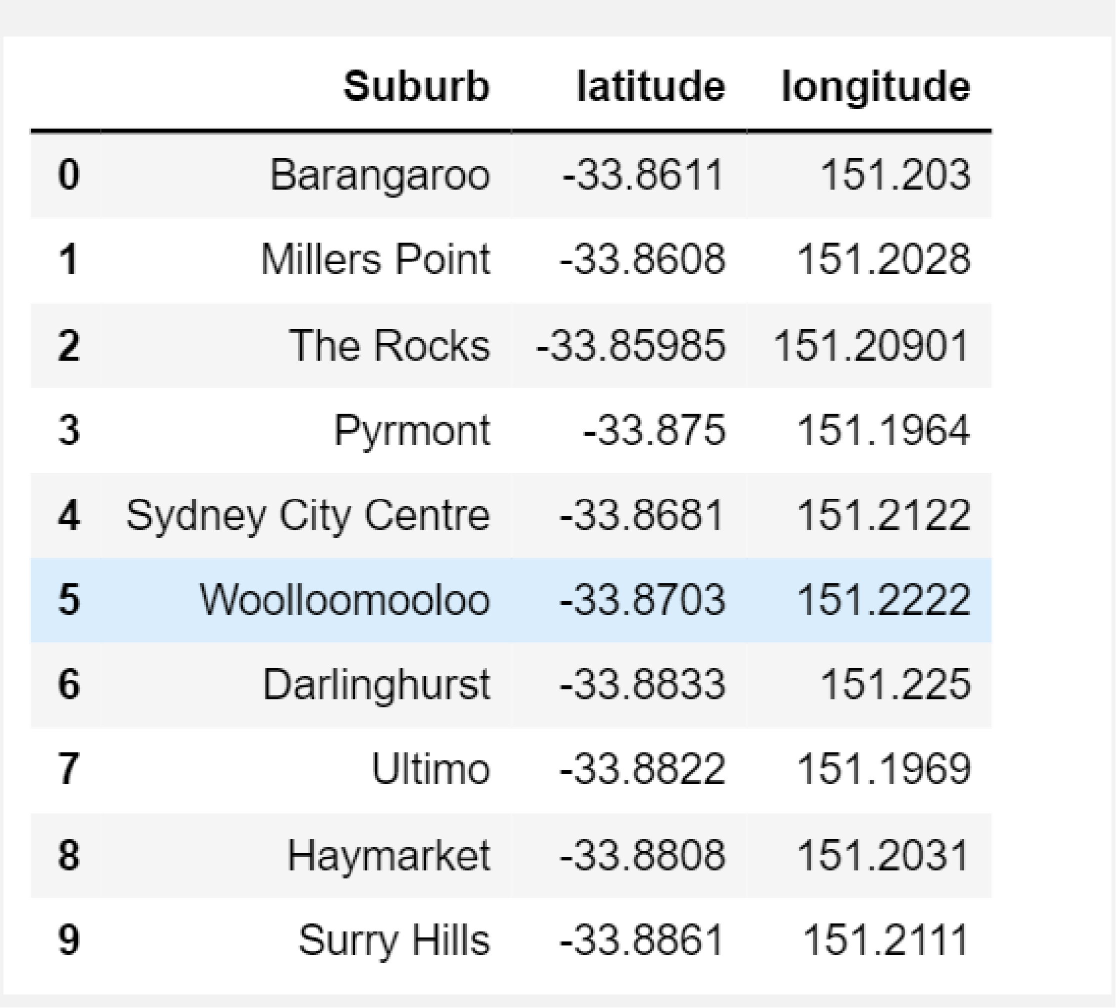
1. I have utilized requests and Beautiful Soup package to scrape the Wikipedia page of Sydney CBD (Sydney City Centre) and its surrounding suburbs to

obtain the latitude and longitude of each location respectively ([Sydney central business district — Wikipedia]())

2. I have relied on Foursqaure API to obtain the most common venues of each suburb

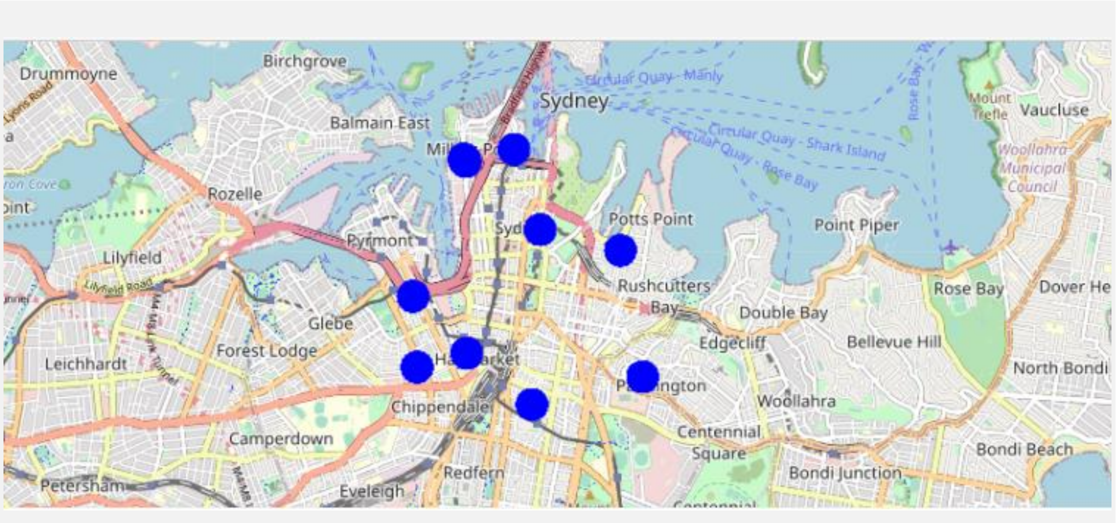
Methodology

First of all, I had extract the the relevant suburbs and its geo coordinates by scrapping the Wikipedia page of Sydney Business Centre to compile a dataframe as fellows

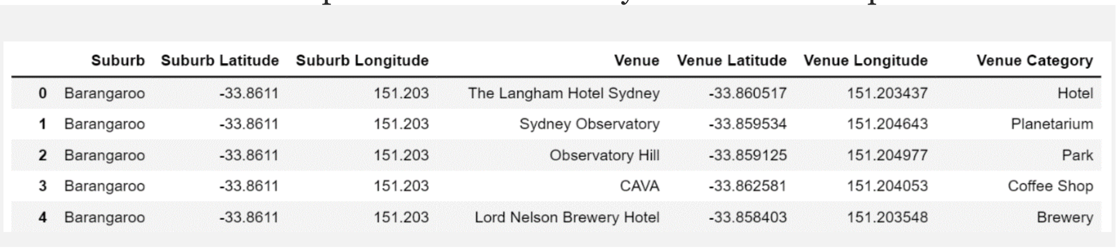


	Suburb	latitude	longitude
0	Barangaroo	-33.8611	151.203
1	Millers Point	-33.8608	151.2028
2	The Rocks	-33.85985	151.20901
3	Pyrmont	-33.875	151.1964
4	Sydney City Centre	-33.8681	151.2122
5	Woolloomooloo	-33.8703	151.2222
6	Darlinghurst	-33.8833	151.225
7	Ultimo	-33.8822	151.1969
8	Haymarket	-33.8808	151.2031
9	Surry Hills	-33.8861	151.2111

I had also illustrated the relevant suburbs through utilizing the **folium** library. These suburbs are shown as below with each blue dot represented one suburb

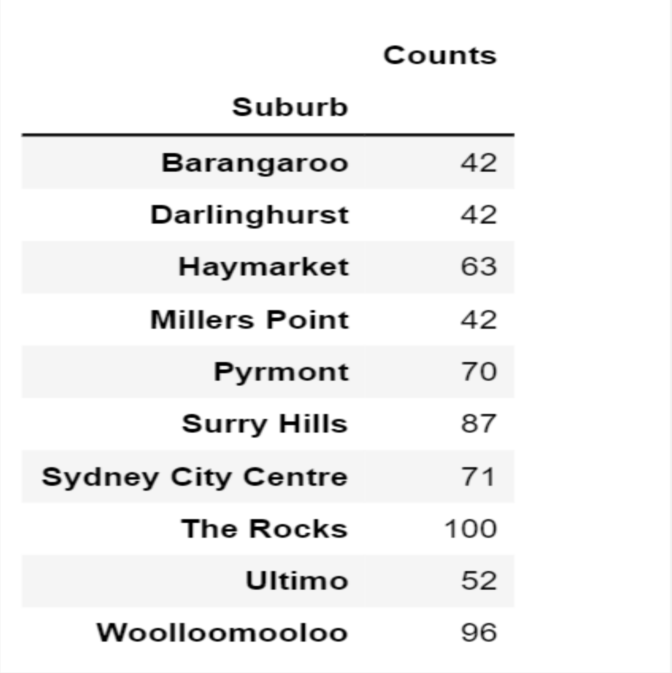


Upon obtaining the relevant suburbs and geo data of each suburb, I had retrieved data that are pertained to this study from the FourSquare API.

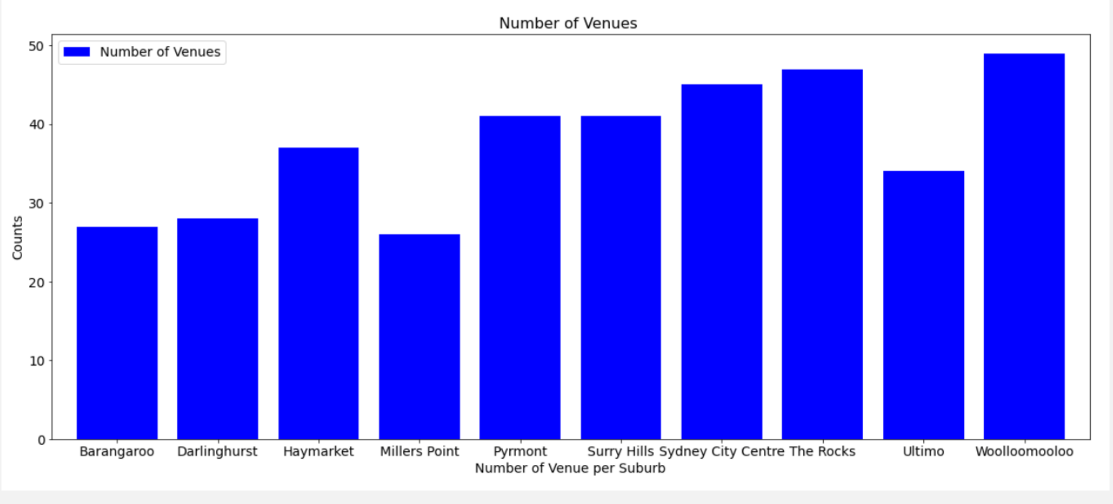


	Suburb	Suburb Latitude	Suburb Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barangaroo	-33.8611	151.203	The Langham Hotel Sydney	-33.860517	151.203437	Hotel
1	Barangaroo	-33.8611	151.203	Sydney Observatory	-33.859534	151.204643	Planetarium
2	Barangaroo	-33.8611	151.203	Observatory Hill	-33.859125	151.204977	Park
3	Barangaroo	-33.8611	151.203	CAVA	-33.862581	151.204053	Coffee Shop
4	Barangaroo	-33.8611	151.203	Lord Nelson Brewery Hotel	-33.858403	151.203548	Brewery

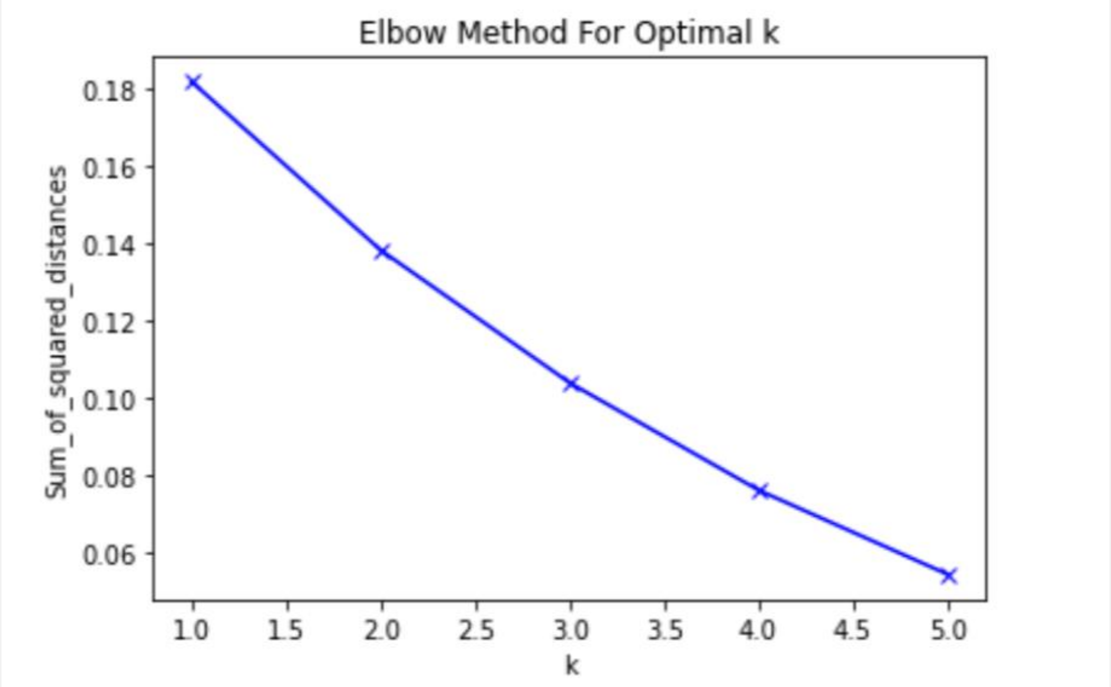
There are a total of **665** venues and the number of venues of each suburb is illustrated as below



	Counts
Suburb	
Barangaroo	42
Darlinghurst	42
Haymarket	63
Millers Point	42
Pyrmont	70
Surry Hills	87
Sydney City Centre	71
The Rocks	100
Ultimo	52
Woolloomooloo	96

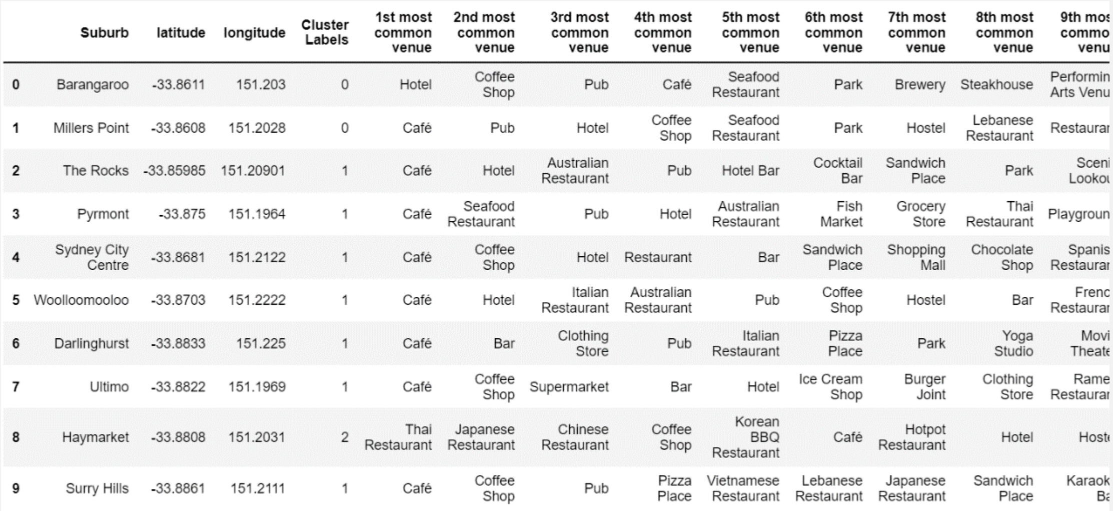


Next, the categorical data was converted into dummies using the **one-hot encoding** in order to implement the K Means algorithm to the dataset. Also, I had identify the **optimal number of clusters** to conduct this analysis based on the **elbow method**.

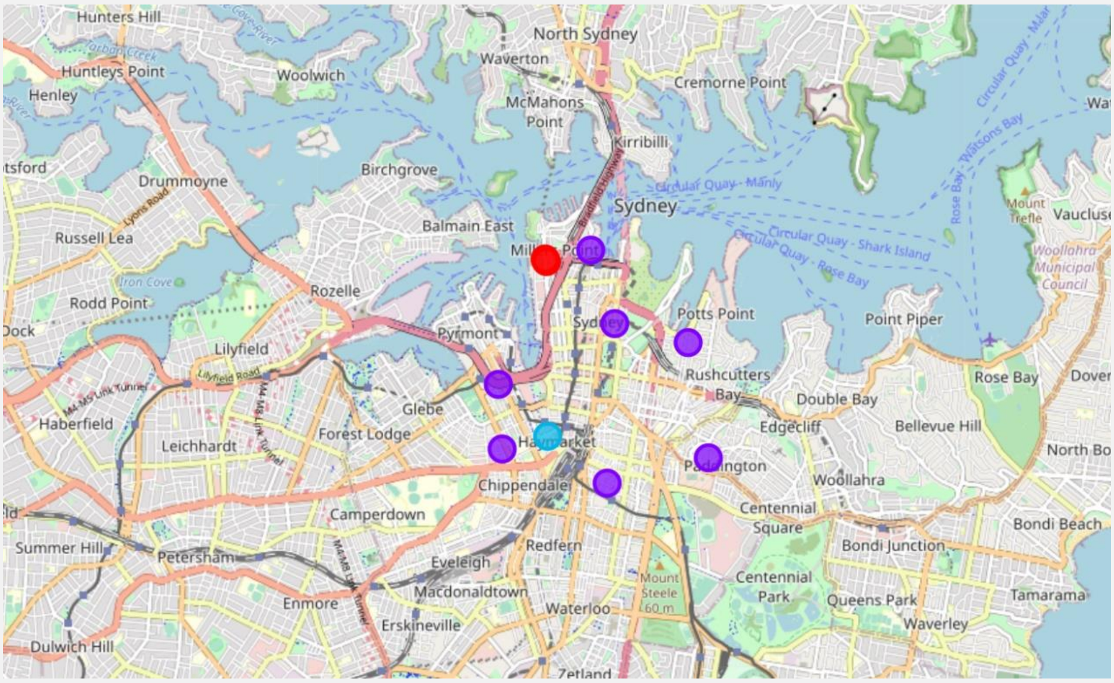


According to the Elbow Method, I decided to adopt K=3 as the number of cluster for this analysis.

Following the implementation of K Means algorithm, each suburb is assigned to a specific cluster, resulting in the following dataframe and map.



	Suburb	latitude	longitude	Cluster Labels	1st most common venue	2nd most common venue	3rd most common venue	4th most common venue	5th most common venue	6th most common venue	7th most common venue	8th most common venue	9th mos commo venu
0	Barangaroo	-33.8611	151.203	0	Hotel	Coffee Shop	Pub	Café	Seafood Restaurant	Park	Brewery	Steakhouse	Performin Arts Venu
1	Millers Point	-33.8608	151.2028	0	Café	Pub	Hotel	Coffee Shop	Seafood Restaurant	Park	Hostel	Lebanese Restaurant	Restaurar
2	The Rocks	-33.85985	151.20901	1	Café	Hotel	Australian Restaurant	Pub	Hotel Bar	Cocktail Bar	Sandwich Place	Park	Sceni Lookou
3	Pyrmont	-33.875	151.1964	1	Café	Seafood Restaurant	Pub	Hotel	Australian Restaurant	Fish Market	Grocery Store	Thai Restaurant	Playgroun
4	Sydney City Centre	-33.8681	151.2122	1	Café	Coffee Shop	Hotel	Restaurant	Bar	Sandwich Place	Shopping Mall	Chocolate Shop	Spanis Restaurar
5	Woolloomooloo	-33.8703	151.2222	1	Café	Hotel	Italian Restaurant	Australian Restaurant	Pub	Coffee Shop	Hostel	Bar	Frenc Restaurar
6	Darlinghurst	-33.8833	151.225	1	Café	Bar	Clothing Store	Pub	Italian Restaurant	Pizza Place	Park	Yoga Studio	Movi Theate
7	Ultimo	-33.8822	151.1969	1	Café	Coffee Shop	Supermarket	Bar	Hotel	Ice Cream Shop	Burger Joint	Clothing Store	Rame Restaurar
8	Haymarket	-33.8808	151.2031	2	Thai Restaurant	Japanese Restaurant	Chinese Restaurant	Coffee Shop	Korean BBQ Restaurant	Café	Hotpot Restaurant	Hotel	Hoste
9	Surry Hills	-33.8861	151.2111	1	Café	Coffee Shop	Pub	Pizza Place	Vietnamese Restaurant	Lebanese Restaurant	Japanese Restaurant	Sandwich Place	Karaok Ba



We can interrupt the above map as follows:

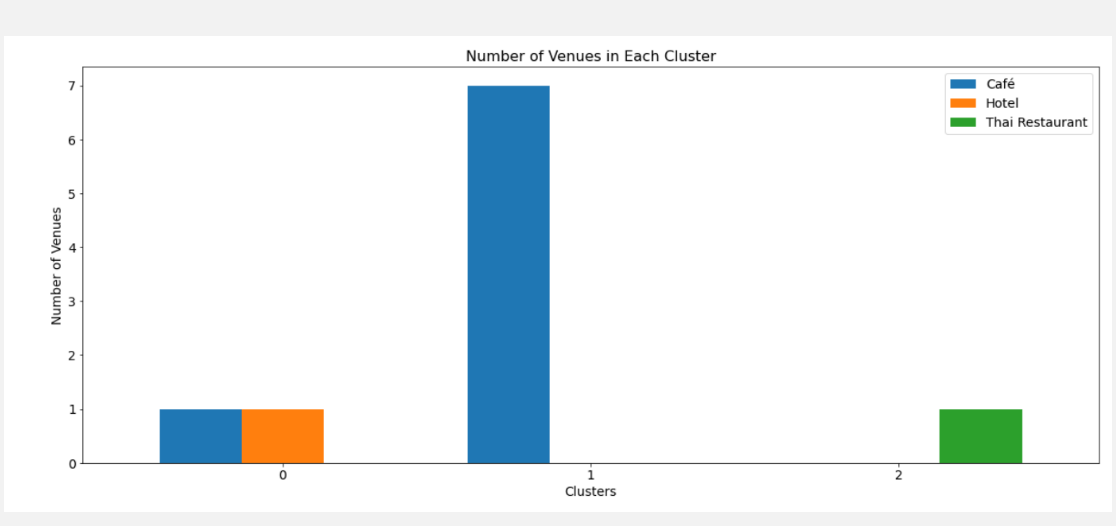
Cluster Red (cluster label = 0) : **Barangaroo** and **Millers Point**

Cluster Purple (cluster label = 1): **The Rocks**, **Pyrmont**, **Sydney City Centre**, **Woolloomooloo**, **Darlinghurst**, **Ultimo** and **Surry Hills**

Cluster Blue (cluster label = 2) : **Haymarket**

Results

Based on the results of the K Means algorithm, we are able to categorize the suburb based on **frequency of each venues** within each suburb. To investigate which cluster is optimal to set up the new coffee shop, we have to examine **1st Most Common Venue in each cluster** to understand the dominant venues in each cluster**.**



From the bar chart above, we can see that cluster 1 is populated with Cafe and there are 7 cafes within this cluster, while cluster 0 has 1 cafe and followed by cluster 2 which has no cafe at all.

In conclusion, I would suggest to **potential entrepreneurs to set up their own coffee shop in cluster 2 (Haymarket)** because **coffee shop is not the most common venue in this cluster** and this represents less intense competition than cluster 0 and cluster 1. In addition, **cluster 2 (Haymarket)**

is populated with venues that are able to drawn large amount of foot traffic , including Thai restaurant, Japanese restaurant and Chinese restaurant.

Discussion

Although the results suggest Cluster 2 the optimal location to set up the coffee shop, I believe there a more in-depth analysis is needed to be done to derive a sound and sophisticated decision. There are **2 important criteria** that a potential entrepreneur can drill into. The first being the **rent distribution** and the second being the **price range of restaurants and especially coffee shops** among the 10 suburbs that are concerned within this analysis. I believe this two criteria are important for potential entrepreneur to effectively maintain a good **cost control** and setting their **pricing strategy**.