Determining a Suitably Located Cafe in Sydney, Australia

1. Introduction

Sydney is the most populous city in the state of New South Wales, as well as in Australia. According to the 2016 Census, over 4.8 million people call Sydney home. Sydney is a global city of commerce, being the home location of ASX, one of the world's foremost stock exchanges, as well as being a regional hub for many companies in business, including investment banks, consultancies and accounting firms. In addition, Sydney is the most visited location in Australia, with no shortage of tourist attractions such as Sydney Opera House, Sydney Harbour Bridge and Bondi Beach.

As a bustling global city with a raft of economic activity, both white-collar and blue-collar, throughout the CBD (Central Business District), whether it is a multinational chain or a local 'mum-and-dad' business, it can be an attractive place to set up business to tap into the many thousands of workers who commute to the CBD every working day and the tourists who come along to experience life in Sydney. Conversely, it can often mean having extreme competition in competing for the many potential customers, and unsurprisingly, businesses must prepare well to survive in cutthroat competition – even popular eateries and cafes have closed down.

Coffee shops, and by extension, cafes, are big business across all Australian metropolitan areas. There appears to be a distinct 'café culture' around Sydney, with a distinctive range of coffee styles not found elsewhere around the world.

2. Business Problem

2.1 Problem

Given its population and its vibrant business culture, many cafes have opened up around Sydney. However, with strong competition and variety of different kinds of cuisines and offering arounds, opening up a new café in Sydney is no easy task. Location is a prime factor in the success or failure of an establishment. Thereby, for this project, the problem is defined to be "Where should a café be opened in the Sydney CBD and surrounds?"

2.2 Interest

The key stakeholders in this problem are potential café owners, as well as existing owners who could be reconsidering their business strategy, or are looking to relocate to their business to tap in on new customers. Landlords and developers of multi-use properties may also have interest when considering utilization of their properties, namely having a café on the ground level with offices above.

2.3 Rationale for Consideration of Problem

Location is paramount to success in order to tap into the most possible number of customers, whether it during the morning rush to work, a coffee break in the morning, for lunch hour, or if opening up late to cater for the tourists. It also goes without saying that a café in the middle of the city is a very expensive investment. The stakes may be high but with tens, if not hundreds, of thousands of people living and working around Sydney,

there is huge potential to attract customers should the business is executed well. It is therefore imperative for owners to carefully consider location and show care for this problem to maximise their potential in taking a slice of this billion-dollar industry.

Other factors that may affect suitability of location of where the potential café may be located, such as ongoing rental costs and structural integrity of the building, are not considered in this project.

3. **Data**

3.1 Source of Data

Data for this project was conveniently obtained from an open-source database on a website maintained by Matthew Proctor (link: https://www.matthewproctor.com/australian_postcodes). Here, a .csv file was downloaded that includes a listing of postcodes across all of Australia – some 3200 unique postcodes. The format was conveniently organized, such that it was easily scraped and converted into a pandas dataframe, as evident in Figure 1. Clearly the required data was evident – 'postcode', 'locality' (equivalent to neighbourhood in North American English terms), 'latitude' and 'longitude'.

Foursquare API and other relevant Python libraries such as 'geocoder' were also used, which tapped into latitude and longitude to enable Foursquare data to be leveraged.

No other datasets are used in this project.

```
[2]: # Loads the html into a pandas dataframe
     australia = pd.read_csv('australian_postcodes.csv') # Loads the .csv into a pandas df

[3]: australia.head()
```

[3]:

	id	postcode	locality	state	long	lat	dc	type	status	sa3	sa3name	sa4	sa4name	region	Lat_precise	Long_precise
0	230	200	ANU	ACT	149.119000	-35.277700	NaN	NaN	NaN	NaN	NaN	NaN	NaN	R1	-35.277700	149.119000
1	21820	200	Australian National University	ACT	149.118900	-35.277700	NaN	NaN	Added 19-Jan-2020	NaN	NaN	NaN	NaN	R1	-35.277700	149.118527
2	232	800	DARWIN	NT	130.836680	-12.458684	NaN	NaN	Updated 6-Feb-2020	70101.0	Darwin City	701.0	Darwin	R1	-12.393279	130.776661
3	233	801	DARWIN	NT	130.836680	-12.458684	NaN	NaN	Updated 25-Mar-2020 SA3	70101.0	Darwin City	701.0	Darwin	R1	-12.463440	130.845642
4	234	804	PARAP	NT	130.873315	-12.428017	NaN	NaN	Updated 25-Mar-2020 SA3	70102.0	Darwin Suburbs	701.0	Darwin	R1	-12.432480	130.846254

Figure 1: Dataframe head of the Australian postcodes file.

3.2 Data Cleaning

With only one source of data, the data cleaning process becomes rather straightforward.

A dataframe consisting of 'postcode', 'locality', 'latitude' and 'longitude' suffices. Other columns of data, such as 'region' and 'Lat_precise', while being invaluable to an app developer, were a surplus to this project, and therefore omitted, as per Figure 2.

	postcode	locality	long	lat
0	2000	BARANGAROO	151.201580	-33.860520
1	2000	DARLING HARBOUR	151.256649	-33.859953
2	2000	DAWES POINT	151.256649	-33.859953
3	2000	HAYMARKET	151.256649	-33.859953
4	2000	MILLERS POINT	151.256649	-33.859953
5	2000	PARLIAMENT HOUSE	151.256649	-33.859953
6	2000	SYDNEY	151.256649	-33.859953
7	2000	SYDNEY SOUTH	151.256649	-33.859953
8	2000	THE ROCKS	151.256649	-33.859953
9	2001	SYDNEY	151.268071	-33.794883
10	2002	WORLD SQUARE	151.206924	-33.877121
11	2004	ALEXANDRIA MC	151.190000	-33.908000
12	2004	EASTERN SUBURBS MC	151.210000	-33.950800
13	2006	THE UNIVERSITY OF SYDNEY	151.186507	-33.889219
14	2007	BROADWAY	151.196650	-33.883189
15	2007	ULTIMO	151.196650	-33.883189

Figure 2: A part of the dataframe of the postcodes relevant for this project

Obviously we only need to consider data around Sydney CBD. This involves restricting data based on postcode. Sydney CBD itself has postcode 2000, and neighbourhoods immediately surrounding the CBD have postcodes from 2000 to 2010. However, there are some neighbourhoods that are clearly much further away from the CBD than others, but have a postcode within the aforementioned range, whereas there are some neighbourhoods that are of closer range but with a postcode not within the range. For simplicity, the project will consider only postcodes 2000 to 2025. From these 23 postcodes alone (Figure 3), there are over 1200 venues to be analysed in this project.

```
print('The dataframe has {} postcodes and {} localities.'.format(
        len(nsw3['postcode'].unique()),
        nsw3.shape[0]
    )
)
```

```
The dataframe has 23 postcodes and 57 localities.
```

Figure 3: Number of postcodes in this project

At this point, the data is ready to be plotted on a map using folium.Map, having established the geographical coordinates of Sydney. Using Foursquare credentials, a list of venues in those neighbourhoods was returned by Foursquare API. Figure 4 shows the venues within postcode 2000 alone.

	name	categories	lat	lng
0	The Langham Hotel Sydney	Hotel	-33.860517	151.203437
1	Palisade Hotel	Pub	-33.857979	151.202264
2	Lord Nelson Brewery Hotel	Brewery	-33.858403	151.203548
3	Fish at the Rocks	Seafood Restaurant	-33.858673	151.203500
4	Sydney Observatory	Planetarium	-33.859534	151.204643
5	Observatory Hill	Park	-33.859125	151.204977
6	Barangaroo Reserve	Park	-33.857052	151.201100
7	CAVA	Coffee Shop	-33.862581	151.204053
8	Bourke Street Bakery	Bakery	-33.864570	151.201480
9	Blu Bar On 36	Hotel Bar	-33.861067	151.206361
10	Harts Pub	Pub	-33.861870	151.206314
11	Shangri-La Hotel	Hotel	-33.861141	151.206460
12	Sydney Theatre Company	Performing Arts Venue	-33.857028	151.204938
13	Sydney Harbour YHA	Hostel	-33.860128	151.206844
14	Roslyn Packer Theatre	Theater	-33.857019	151.204947
15	Shirt Bar	Bar	-33.864302	151.202609

Figure 4: Dataframe of venues given by Foursquare API within postcode 2000

4. Data Analysis Methodology

Having obtained the venues given by Foursquare API for one postcode, it follows that the same analysis is to be applied to the rest of the postcodes relevant to this project. Venues were obtained within a 750 metre radius from each locality's data point for reasonable coverage of the areas covered by the postcodes.

In all, 1902 venues altogether are involved in this project. However, it is clearly visible from Figure 5 that 'venue category' can be anything from a cafe to a park and hotel.

```
print(syd_venues.shape)
syd_venues.head()
```

```
(1902, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BARANGAROO	-33.86052	151.20158	The Langham Hotel Sydney	-33.860517	151.203437	Hotel
1	BARANGAROO	-33.86052	151.20158	Palisade Hotel	-33.857979	151.202264	Pub
2	BARANGAROO	-33.86052	151.20158	Lord Nelson Brewery Hotel	-33.858403	151.203548	Brewery
3	BARANGAROO	-33.86052	151.20158	Fish at the Rocks	-33.858673	151.203500	Seafood Restaurant
4	BARANGAROO	-33.86052	151.20158	Observatory Hill	-33.859125	151.204977	Park

Figure 5: Dataframe of venues given by Foursquare API for all postcodes.

The venues are then grouped by locality (Figure 6):

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
ALEXANDRIA	48	48	48	48	48	48
ALEXANDRIA MC	36	36	36	36	36	36
BANKSMEADOW	6	6	6	6	6	6
BARANGAROO	50	50	50	50	50	50
BEACONSFIELD	48	48	48	48	48	48
BELLEVUE HILL	8	8	8	8	8	8
BONDI JUNCTION	41	41	41	41	41	41
BONDI JUNCTION PLAZA	41	41	41	41	41	41
BOTANY	6	6	6	6	6	6
BROADWAY	50	50	50	50	50	50
BRONTE	32	32	32	32	32	32
CENTENNIAL PARK	46	46	46	46	46	46
CHARING CROSS	32	32	32	32	32	32

Figure 6: Dataframe of venues being grouped by locality

Now each locality is ready to be analysed. For each locality, the top 10 venues are obtained (Figure 7).

	Neighborhood	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 Most Common Venue	10th Most Common Venue
0	ALEXANDRIA	Café	Brewery	Pet Store	Coffee Shop	Miscellaneous Shop	Italian Restaurant	Furniture / Home Store	Electronics Store	Playground	Basketball Stadium
1	ALEXANDRIA MC	Café	Playground	Sandwich Place	Bar	Supermarket	Basketball Stadium	Bowling Green	Shipping Store	Seafood Restaurant	Brewery
2	BANKSMEADOW	Café	Coffee Shop	Antique Shop	Bakery	Badminton Court	Park	Dive Bar	Fast Food Restaurant	Farmers Market	Event Space
3	BARANGAROO	Café	Pub	Hotel	Seafood Restaurant	Coffee Shop	Bar	Bakery	Park	Chinese Restaurant	Flea Market
4	BEACONSFIELD	Café	Brewery	Pet Store	Coffee Shop	Miscellaneous Shop	Italian Restaurant	Furniture / Home Store	Electronics Store	Playground	Basketball Stadium

Figure 7: Top 10 venues for each locality

K-means clustering model is then applied to the data. 5 clusters are used in this project. Cluster labels now group localities based on venue commonality, and are combined with the postcode of the locality along with its latitude and longitude (Figure 8):

	postcode	locality	long	lat	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
0	2000	BARANGAROO	151.201580	-33.860520	3.0	Café	Pub	Hotel	Seafood Restaurant	Coffee Shop	Bar	Bakery	Park
1	2000	DARLING HARBOUR	151.256649	-33.859953	1.0	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
2	2000	DAWES POINT	151.256649	-33.859953	1.0	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
3	2000	HAYMARKET	151.256649	-33.859953	1.0	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
4	2000	MILLERS POINT	151.256649	-33.859953	1.0	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store

Figure 8: K-means clustering is applied

There is a problem, however. Cluster labels are not in integers, which makes further analysis too difficult. Upon closer analysis, there appears NaN on certain lines. Here, rows that have NaN are removed, which are taken to be localities that either have no venues or

of which no data is available. After removal of rows in question, cluster labels are now integers, allowing further processing (Figure 9):

	postcode	locality	long	lat	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
0	2000	BARANGAROO	151.201580	-33.860520	3	Café	Pub	Hotel	Seafood Restaurant	Coffee Shop	Bar	Bakery	Park
1	2000	DARLING HARBOUR	151.256649	-33.859953	1	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
2	2000	DAWES POINT	151.256649	-33.859953	1	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
3	2000	HAYMARKET	151.256649	-33.859953	1	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
4	2000	MILLERS POINT	151.256649	-33.859953	1	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store

Figure 9: Data is adjusted to ensure cluster label columns are integers

A map of clusters is now produced (Figure 10):

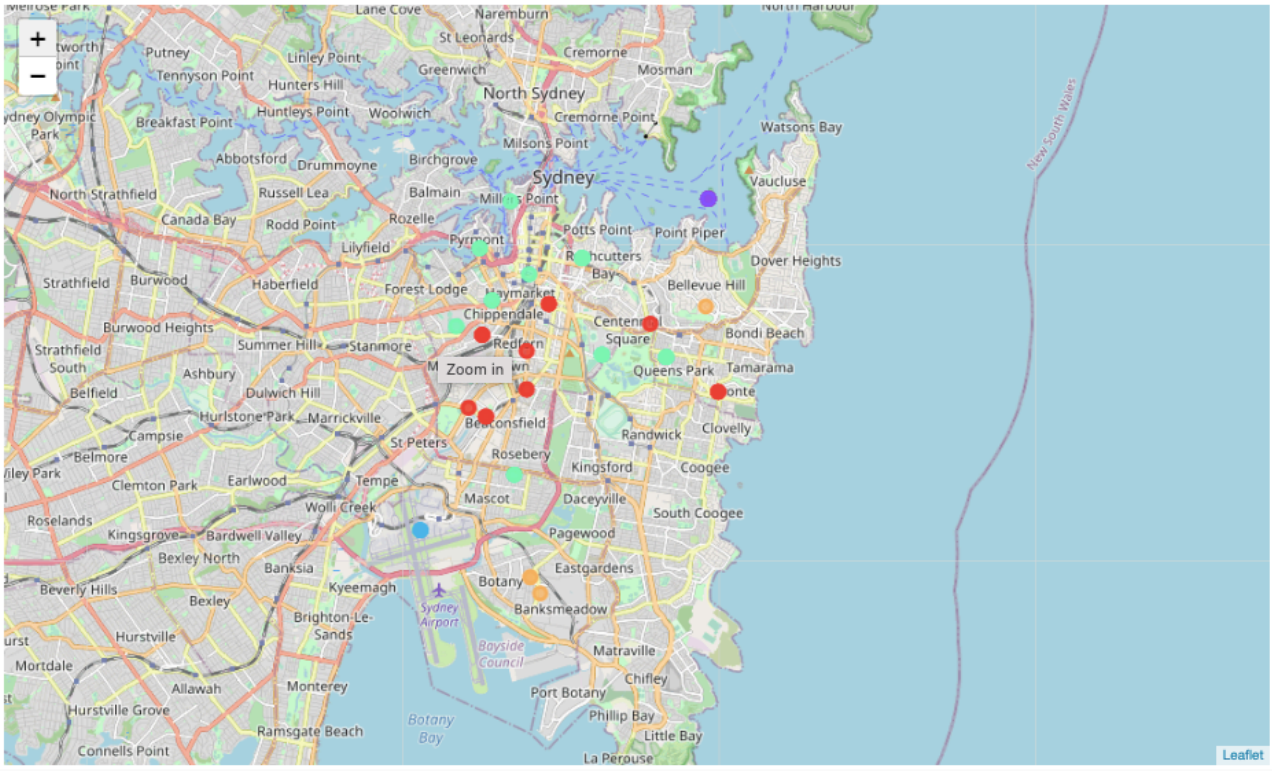

Figure 10: Map of the 5 clusters

5. Results

5 clusters of venues were obtained.

Cluster 1 consists of Cafe, with pubs/bars/brewery as 2nd (Figure 11).

	locality	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 Most Common Venue	10th Most Common Venue
6	SYDNEY	Café	Park	Harbor / Marina	Japanese Restaurant	Hardware Store	Furniture / Home Store	Shopping Mall	Farmers Market	Supermarket	Liquor Store
9	SYDNEY	Café	Park	Harbor / Marina	Japanese Restaurant	Hardware Store	Furniture / Home Store	Shopping Mall	Farmers Market	Supermarket	Liquor Store
11	ALEXANDRIA MC	Café	Playground	Sandwich Place	Bar	Supermarket	Basketball Stadium	Bowling Green	Shipping Store	Seafood Restaurant	Brewery
16	CHIPPENDALE	Café	Bar	Pub	Coffee Shop	Bakery	Farmers Market	Beer Bar	Market	Dive Bar	Ramen Restaurant
17	DARLINGTON	Café	Bar	Pub	Coffee Shop	Bakery	Farmers Market	Beer Bar	Market	Dive Bar	Ramen Restaurant
18	GOLDEN GROVE	Café	Bar	Pub	Coffee Shop	Bakery	Farmers Market	Beer Bar	Market	Dive Bar	Ramen Restaurant
21	DARLINGHURST	Café	Coffee Shop	Pub	Yoga Studio	Sandwich Place	Japanese Restaurant	Vegetarian / Vegan Restaurant	Bar	Pizza Place	Gym
22	SURRY HILLS	Café	Coffee Shop	Pub	Yoga Studio	Sandwich Place	Japanese Restaurant	Vegetarian / Vegan Restaurant	Bar	Pizza Place	Gym
23	TAYLOR SQUARE	Café	Coffee Shop	Pub	Yoga Studio	Sandwich Place	Japanese Restaurant	Vegetarian / Vegan Restaurant	Bar	Pizza Place	Gym
32	ALEXANDRIA	Café	Brewery	Pet Store	Coffee Shop	Miscellaneous Shop	Italian Restaurant	Furniture / Home Store	Electronics Store	Playground	Basketball Stadium
33	BEACONSFIELD	Café	Brewery	Pet Store	Coffee Shop	Miscellaneous Shop	Italian Restaurant	Furniture / Home Store	Electronics Store	Playground	Basketball Stadium
34	EVELEIGH	Café	Brewery	Pet Store	Coffee Shop	Miscellaneous Shop	Italian Restaurant	Furniture / Home Store	Electronics Store	Playground	Basketball Stadium
35	REDFERN	Café	Bar	Bakery	Pub	Park	Thai Restaurant	Pizza Place	Vietnamese Restaurant	Mediterranean Restaurant	Ice Cream Shop
36	WATERLOO	Café	Park	Pizza Place	Sandwich Place	Bar	Thai Restaurant	Sporting Goods Shop	Grocery Store	Gym	Coffee Shop
38	ZETLAND	Café	Park	Pizza Place	Sandwich Place	Bar	Thai Restaurant	Sporting Goods Shop	Grocery Store	Gym	Coffee Shop
53	BRONTE	Café	Juice Bar	Pub	Pizza Place	Park	Breakfast Spot	Fish & Chips Shop	French Restaurant	Australian Restaurant	Bus Stop
54	CHARING CROSS	Café	Juice Bar	Pub	Pizza Place	Park	Breakfast Spot	Fish & Chips Shop	French Restaurant	Australian Restaurant	Bus Stop
55	WAVERLEY	Café	Juice Bar	Pub	Pizza Place	Park	Breakfast Spot	Fish & Chips Shop	French Restaurant	Australian Restaurant	Bus Stop
56	WOOLLAHRA	Café	Pub	Pizza Place	Italian Restaurant	Japanese Restaurant	Gastropub	Bar	Bakery	Cheese Shop	French Restaurant

Figure 11: Dataframe of Cluster 1

Cluster 2 consists of harbour/marina (Figure 12).

	locality	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 Most Common Venue	10th Most Common Venue
1	DARLING HARBOUR	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Dumpling Restaurant	Donut Shop
2	DAWES POINT	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Dumpling Restaurant	Donut Shop
3	HAYMARKET	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Dumpling Restaurant	Donut Shop
4	MILLERS POINT	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Dumpling Restaurant	Donut Shop
5	PARLIAMENT HOUSE	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Dumpling Restaurant	Donut Shop
7	SYDNEY SOUTH	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Dumpling Restaurant	Donut Shop
8	THE ROCKS	Harbor / Marina	Park	Beach	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Dumpling Restaurant	Donut Shop

Figure 12: Dataframe of Cluster 2

Cluster 3 consists of airport lounge (Figure 13).

	locality	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 Most Common Venue	10th Most Common Venue
43	MASCOT	Airport Lounge	Airport Service	Fast Food Restaurant	Café	Coffee Shop	Donut Shop	Beer Bar	Tapas Restaurant	Women's Store	Juice Bar
44	SYDNEY DOMESTIC AIRPORT	Airport Lounge	Airport Service	Fast Food Restaurant	Café	Coffee Shop	Donut Shop	Beer Bar	Tapas Restaurant	Women's Store	Juice Bar
45	SYDNEY INTERNATIONAL AIRPORT	Airport Lounge	Airport Service	Fast Food Restaurant	Café	Coffee Shop	Donut Shop	Beer Bar	Tapas Restaurant	Women's Store	Juice Bar

Figure 13: Dataframe of Cluster 3

Cluster 4 consists of a mixture of cafe with Japanese Restaurant, Thai Restaurant and Italian Restaurant (Cricket Ground somehow being clustered in this) (Figure 14):

	locality	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 Most Common Venue	10th Most Common Venue
0	BARANGAROO	Café	Pub	Hotel	Seafood Restaurant	Coffee Shop	Bar	Bakery	Park	Chinese Restaurant	Flea Market
10	WORLD SQUARE	Thai Restaurant	Japanese Restaurant	Hotel	Café	Korean BBQ Restaurant	Malay Restaurant	Chinese Restaurant	Coffee Shop	Burger Joint	Breakfast Spot
13	THE UNIVERSITY OF SYDNEY	Café	Thai Restaurant	Pub	Vietnamese Restaurant	Pizza Place	Italian Restaurant	Coffee Shop	Cosmetics Shop	Sandwich Place	Fast Food Restaurant
14	BROADWAY	Café	Coffee Shop	Thai Restaurant	Bar	Pub	Dessert Shop	Hotel	Supermarket	Shopping Mall	Bookstore
15	ULTIMO	Café	Coffee Shop	Thai Restaurant	Bar	Pub	Dessert Shop	Hotel	Supermarket	Shopping Mall	Bookstore
19	DARLING ISLAND	Café	Pub	Fish Market	Bar	Seafood Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	French Restaurant	Australian Restaurant
20	PYRMONT	Café	Pub	Fish Market	Bar	Seafood Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	French Restaurant	Australian Restaurant
24	ELIZABETH BAY	Café	Italian Restaurant	Australian Restaurant	Pub	Coffee Shop	Wine Bar	Gym	Lounge	Japanese Restaurant	Speakeasy
25	HMAS KUTTABUL	Café	Italian Restaurant	Australian Restaurant	Pub	Coffee Shop	Wine Bar	Gym	Lounge	Japanese Restaurant	Speakeasy
26	KINGS CROSS	Café	Italian Restaurant	Australian Restaurant	Pub	Coffee Shop	Wine Bar	Gym	Lounge	Japanese Restaurant	Speakeasy
27	POTTS POINT	Café	Italian Restaurant	Australian Restaurant	Pub	Coffee Shop	Wine Bar	Gym	Lounge	Japanese Restaurant	Speakeasy
28	RUSHCUTTERS BAY	Café	Italian Restaurant	Australian Restaurant	Pub	Coffee Shop	Wine Bar	Gym	Lounge	Japanese Restaurant	Speakeasy
29	WOOLLOOMOOLOO	Café	Italian Restaurant	Australian Restaurant	Pub	Coffee Shop	Wine Bar	Gym	Lounge	Japanese Restaurant	Speakeasy
39	EASTLAKES	Bakery	Café	Bus Stop	Australian Restaurant	Souvlaki Shop	Shopping Mall	Pet Store	Sports Bar	Seafood Restaurant	Dog Run
40	ROSEBERY	Bakery	Café	Bus Stop	Australian Restaurant	Souvlaki Shop	Shopping Mall	Pet Store	Sports Bar	Seafood Restaurant	Dog Run
46	CENTENNIAL PARK	Cricket Ground	Café	Movie Theater	Park	Food Truck	Event Space	Multiplex	Comedy Club	Bowling Alley	Sandwich Place
47	MOORE PARK	Cricket Ground	Café	Movie Theater	Park	Food Truck	Event Space	Multiplex	Comedy Club	Bowling Alley	Sandwich Place
48	PADDINGTON	Cricket Ground	Café	Movie Theater	Park	Food Truck	Event Space	Multiplex	Comedy Club	Bowling Alley	Sandwich Place
49	BONDI JUNCTION	Japanese Restaurant	Café	Coffee Shop	Pub	Electronics Store	Sushi Restaurant	Gym	Noodle House	Burger Joint	Farmers Market
50	BONDI JUNCTION PLAZA	Japanese Restaurant	Café	Coffee Shop	Pub	Electronics Store	Sushi Restaurant	Gym	Noodle House	Burger Joint	Farmers Market
51	QUEENS PARK	Japanese Restaurant	Café	Coffee Shop	Pub	Electronics Store	Sushi Restaurant	Gym	Noodle House	Burger Joint	Farmers Market

Figure 14: Dataframe of Cluster 4

Cluster 5 appears to be some random mix of park, bakery, cafe and coffee shop (Figure 5).

	locality	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 Most Common Venue	10th Most Common Venue
12	EASTERN SUBURBS MC	Bakery	Park	Sandwich Place	Coffee Shop	Badminton Court	Café	Discount Store	Event Space	Electronics Store	Dumpling Restaurant
41	BANKSMEADOW	Café	Coffee Shop	Antique Shop	Bakery	Badminton Court	Park	Dive Bar	Fast Food Restaurant	Farmers Market	Event Space
42	BOTANY	Café	Coffee Shop	Antique Shop	Bakery	Badminton Court	Park	Dive Bar	Fast Food Restaurant	Farmers Market	Event Space
52	BELLEVUE HILL	Park	Café	Pizza Place	Italian Restaurant	Bagel Shop	Discount Store	Farmers Market	Event Space	Electronics Store	Dumpling Restaurant

Figure 15: Dataframe of Cluster 5

6. Discussion

Having clustered the data, the relevance of each cluster to the initial question of “where should a café be opened in the Sydney CBD and surrounds?” is made.

Cluster 1 contains the bulk of the localities. It is clustered based on the fact the most common venue in all of those is a cafe of some sort, with the next common being a pub, brewery or coffee shop (the latter being essentially the same as a cafe). From personal experience, these localities are those that are slightly farther from the CBD and have less foot traffic as there are less business around there (white collar, office-type ones at least). However, this does not mean there is no significant business around there at all - a major global university covers several of the localities there, and the surrounding localities are known for its ‘hipster’ vibe with its industries being more technological and design-based.

Cluster 2 appears purely geographical - grouped together with harbour/marina as the most common venue. This cluster is only of interest if one is to open a cafe to tap into the commuters who take the ferries in morning and evening peak hours, as well as the tourists who use the ferries at any time of the day.

Cluster 3 is also purely geographical - Sydney Airport. Like Cluster 2, opening a cafe is recommended only if interested in serving travellers which come any all times during the day, but are not recurring like other forms of transport.

Cluster 4 has either a restaurant or cafe as the most common venue. With further clustering, the localities with cricket ground as most common venue would be another cluster, as well as the ones with bakery and others as Japanese Restaurant, so this cluster really is a mixed bag, which just happens to be clustered the same based on Foursquare. Given all are in the same cluster, it implies that the geographical characteristics of these localities are similar, hence opening a cafe anywhere in this cluster, whether right in the middle of the city or further away, has similar sort of risks and rewards.

Cluster 5 also appears to be geographical, in the southern part of Sydney, close to the airport but significantly further away from the CBD. While there are cafes which serve those who work in the industrial complexes, particularly logistics, transport and manufacturing, the market here is very different to that of the other clusters.

Overall, where a cafe should be opened around Sydney will depend on the interest parties' interests, but as shown, can be narrowed down to at least 5 distinct areas. It will depend on their business strategy as to which segment of society they want to tap in-office workers in the CBD? Tourists in and around the airport? Commuters using the ferries? Blue-collar workers in industrial areas?

As mentioned in Section 2, there are other factors besides location that will influence these business decisions, but are beyond the scope of the project. If the project was to be developed further, the localities could be broken up into smaller areas, or even streets, which will take much more time and require further datasets on street coordinates which may or may not be available.

On a personal level, Cluster 1 is recommended to the interested parties to this project. It seems like a safer choice than Cluster 4 when considering second most common venue and so forth. While Cluster 1 has all of its localities having cafes being most common, next most common tend to be pubs/bars, which do not usually attract the same type of customers at a given point of time, especially during weekday business hours. Opening in Cluster 4 will consider competition with restaurants for the same customers. Both clusters do cater to primarily white collar business people and commuters.

7. Conclusion

This project has looked into where a cafe should be opened around Sydney CBD and surrounds. It can be narrowed down into 5 distinct clusters of localities, each with its unique qualities that will be of interest depending on the cafe owner or interested parties who may be looking to open a cafe. Cluster 1 was recommended. For future direction, one should perform further analysis on this data, as well as own research around localities of interest to maximise their chances of success.