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, such as ongoing rental costs and structural integrity of the building of which the potential café may be located, 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).

	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

4. Data Analysis Methodology

(Part 2)