London travel recommendations

Srikant Balan

March 20, 2021

1. **Introduction –**

The tourism and travel industry has seen near exponential growth over the years. The modern-day traveller is savvier than ever before and has a multitude of options in terms of destinations based on his/her interests.

In this project, we will use a data science-based approach, using datasets from the city of Sydney, Australia and London, United Kingdom.
We will also use location data from Foursquare.

Using these data points, we will try to

- Recommend to a traveller from Sydney a visit to London given the diversity of London
- Which area in London is likely to be most similar to ones he finds back home, based on type of venues.
- For a traveller interested in trying new restaurants and cuisines, make recommendations on which area he/she is likely to find more options

This would be very useful to a first-time traveller considering London, who might not know much about the layout of the city, beyond the usual tourist spots.

It would also be of interest to an intrepid travel and/or a foodie looking and willing to try new things.

2. Data acquisition and cleaning

2.1 London

London, United Kingdom is the capital and largest city in England and the UK. It is administered by City of London and has 32 boroughs. Each borough consists of many post codes.
The details of these boroughs and post codes are conveniently located on Wikipedia, from where we can source the data.

https://en.wikipedia.org/wiki/List_of_areas_of_London

This page was scraped for tables, and the primary table contains areas, boroughs as well as their post codes.

Location	London borough	Post town	Postcode district	Dial code	OS grid ref
Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
Aldborough Hatch	Redbridge[9]	ILFORD	IG2	020	TQ455895
Aldgate	City[10]	LONDON	EC3	020	TQ334813
Aldwych	Westminster[10]	LONDON	WC2	020	TQ307810

This was isolated for London-only areas and boroughs (removing greater London areas for simplicity)

Only requisite columns were kept and we ended up with a dataset like this

Out[7]:

	Borough	Postcode
0	City	EC3
1	Westminster	WC2
2	Bromley	SE20
3	Islington	N19
4	Wandsworth	SW12

2.2 Sydney

Sydney, Australia or rather the city of Sydney, is the largest city in Australia in terms of population (according to census 2019). It is closely followed by Melbourne.
Sydney is roughly divided into a number of localities (17). This was obtained from Wikipedia.

https://en.wikipedia.org/wiki/City_of_Sydney

Since it is more of an informal sub-division, the Latitude and Longitude of these localities were obtained using search engine searches. (sources provided in the notebook and here)

https://raw.githubusercontent.com/sbalanchickoo/datasets/master/coursera/SydneyLocalities.csv

2.3 Foursquare

Foursquare is a technology company, that uses user-provided and scraped details, and provides information about venues / points of interests, and related metadata in a given location. These can be queried using geographical Latitude and Longitude of a location, via API calls.

3. Methodology

3.1 Geographical details

Since Foursquare was used to gather details about venues in a locality, and since the API calls for Foursquare require geographical coordinates (Latitude and Longitude), the first step for London, was to use a python library (geopy) to derive this using address (post code) details.

We used another library of Folium to visualize the locations. As can be seen, some of them are quite clustered.

For Sydney, since post codes weren't readily available, search engine searches were used to obtain Latitude and Longitude for a point with the locality.

London

```
df_london.head()
```

Out[8]:

	Borough	Postcode	Latitude	Longitude
0	City	EC3	51.511333	-0.081960
1	Westminster	WC2	51.511740	-0.122472
2	Bromley	SE20	51.410011	-0.058192
3	Islington	N19	51.567824	-0.138827
4	Wandsworth	SW12	51.443800	-0.152629

```
map_london
```

Out[9]:

Sydney

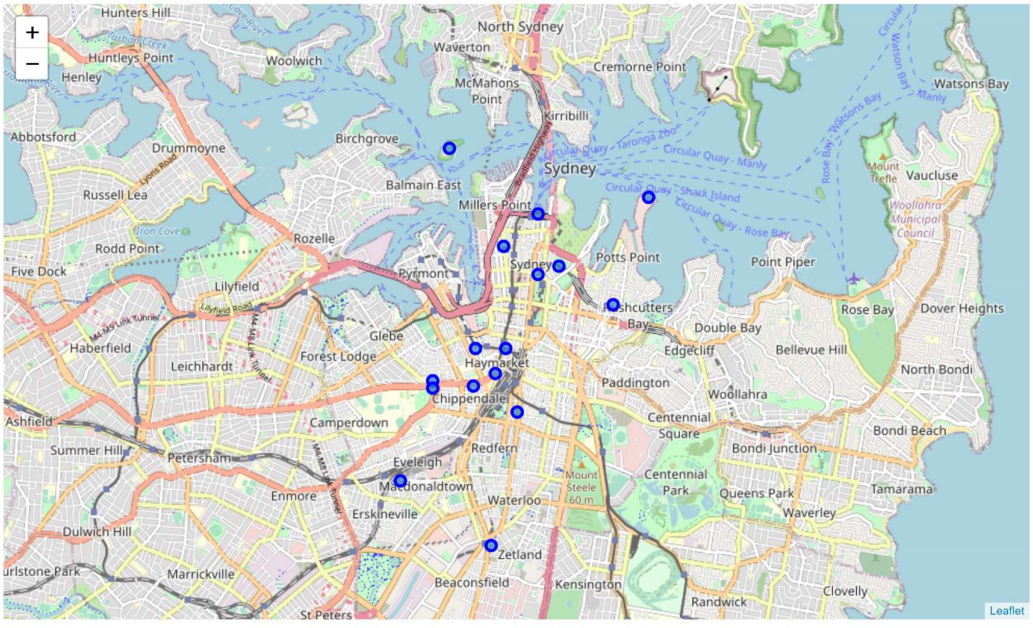
```
df_sydney
```

Out[20]:

	Locality	Latitude	Longitude
0	Broadway	-33.8839	151.1939
1	Central	-33.8849	151.1939
2	Central Park	-33.8846	151.2006
3	Chinatown	-33.8796	151.2059
4	Circular Quay	-33.8611	151.2111
5	Darling Harbour	-33.8796	151.2009
6	The Domain	-33.8683	151.2147
7	Garden Island	-33.8589	151.2294
8	Goat Island	-33.8522	151.1966
9	Green Square	-33.9065	151.2035
10	Kings Cross	-33.8736	151.2236
11	Macdonaldtown	-33.8977	151.1885
12	Railway Square	-33.8829	151.2042
13	Strawberry Hills	-33.8883	151.2078
14	St James	-33.8694	151.2111
15	Wynyard	-33.8656	151.2055

```
map_sydney
```

Out[21]:



3.2 Foursquare for venue details

Foursquare provides a number of APIs, with various tiered pricing plans.

One of these APIs takes in Latitude and Longitude and returns venues and other metadata for those venues within a certain radius. We used a radius of 200 meters and got venue details

Out[14]:

	Borough	Postcode	Postcode Latitude	Postcode Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	City	EC3	51.511333	-0.08196	St Dunstan in the East Garden	51.509716	-0.082354	Garden
1	City	EC3	51.511333	-0.08196	Sky Garden	51.511168	-0.083625	Scenic Lookout
2	City	EC3	51.511333	-0.08196	BrewDog Tower Hill	51.509948	-0.080977	Beer Bar
3	City	EC3	51.511333	-0.08196	The Garden at 120	51.512101	-0.080799	Garden
4	City	EC3	51.511333	-0.08196	Curators Coffee Studio	51.512085	-0.082568	Coffee Shop

3.3 Venue analysis

Once venue details were obtained, we then categorized, grouped and one-hot encoded them to be able to do statistical analysis.

One-hot encoding transforms a set of categorical data type columns into separate columns for each distinct value and populates with 1s and 0s depending on existence of value. This will allow us to do more analysis and get types and percentage of venue category within each locality.

Out[16]:

	Borough	Airport Service	American Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Austrian Restaurant	BBQ Joint	...	Train Station	Turkish Restaurant	Vegetarian / Vegan Restaurant
0	Barnet	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.013889	0.000000
1	Bexley	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
2	Brent	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.137255	0.000000	0.000000
3	Bromley	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
4	Camden	0.00463	0.009259	0.000000	0.000000	0.000000	0.000000	0.000000	0.004630	0.000000	...	0.000000	0.000000	0.000000
5	City	0.00000	0.000000	0.000000	0.000000	0.014085	0.028169	0.000000	0.000000	0.000000	...	0.000000	0.014085	0.000000
6	Ealing	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
7	Enfield	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.074074	0.037037
8	Greenwich	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.090909	0.000000	0.000000
9	Hackney	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.135135	0.000000	0.000000
10	Hammersmith and Fulham	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.016393	0.000000	0.000000
11	Haringey	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
12	Hounslow	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
13	Islington	0.00000	0.000000	0.000000	0.000000	0.133333	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
14	Kensington and Chelsea	0.00000	0.000000	0.016807	0.012605	0.000000	0.000000	0.000000	0.000000	0.008403	...	0.000000	0.000000	0.000000

3.4 Repeat for Sydney

The same process of 3.2 and 3.3 was repeated for Sydney

Out[22]:

	Borough	Postcode	Postcode Latitude	Postcode Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Broadway	Broadway	-33.8839	151.1939	Staves Brewery	-33.884055	151.194281	Brewery
1	Broadway	Broadway	-33.8839	151.1939	Harris Farm Markets	-33.883650	151.193760	Fruit & Vegetable Store
2	Broadway	Broadway	-33.8839	151.1939	Badde Manors	-33.883651	151.191968	Vegetarian / Vegan Restaurant
3	Broadway	Broadway	-33.8839	151.1939	At Home Thai	-33.883883	151.192314	Thai Restaurant
4	Broadway	Broadway	-33.8839	151.1939	Passiontree Velvet	-33.883649	151.193859	Café

Out[24]:

	Locality	Arcade	Art Gallery	Asian Restaurant	Australian Restaurant	Bakery	Bar	Basketball Court	Big Box Store	Bistro	...	Thai Restaurant	Theater	Trail	Trai Statio
0	Broadway	0.022222	0.000000	0.000000	0.000000	0.000000	0.044444	0.022222	0.022222	0.000000	...	0.044444	0.000000	0.000000	0.00000
1	Central	0.024390	0.000000	0.000000	0.000000	0.000000	0.073171	0.000000	0.024390	0.000000	...	0.048780	0.000000	0.000000	0.00000
2	Central Park	0.000000	0.032258	0.000000	0.032258	0.000000	0.032258	0.000000	0.000000	0.000000	...	0.032258	0.000000	0.000000	0.00000
3	Chinatown	0.000000	0.034483	0.000000	0.000000	0.000000	0.034483	0.000000	0.000000	0.000000	...	0.172414	0.034483	0.000000	0.00000
4	Circular Quay	0.000000	0.000000	0.000000	0.027027	0.000000	0.081081	0.000000	0.000000	0.027027	...	0.027027	0.000000	0.027027	0.02702
5	Darling Harbour	0.000000	0.000000	0.000000	0.071429	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.00000
6	Garden Island	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.00000
7	Goat Island	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.00000
8	Green Square	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.25000

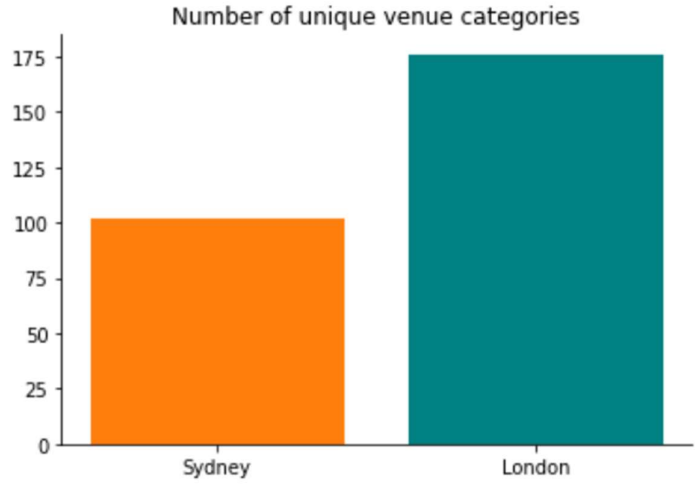
4. Results

For ease of comparison, we only considered the top 10 venue categories within each borough/locality and considered boroughs and localities with at 5 venues to prevent outlier areas from skewing the data.

The following observations and results were found.

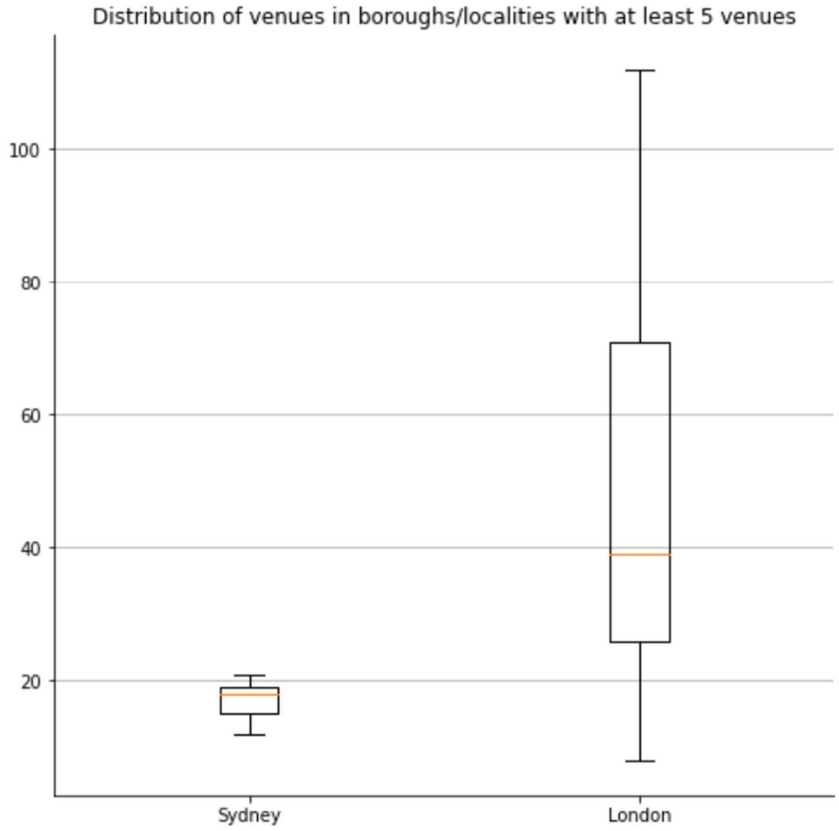
4.1 Venue categories

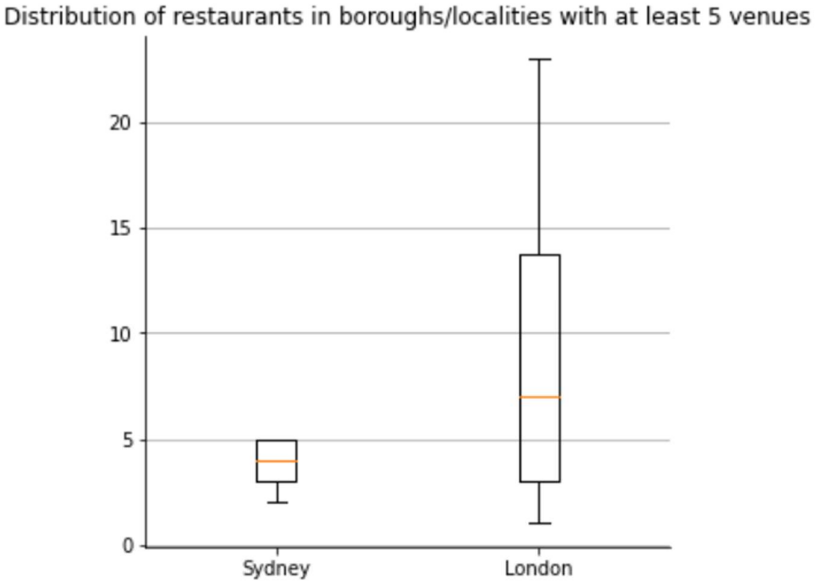
The number of unique venue categories and restaurants is as below.



4.2 Distribution of venues

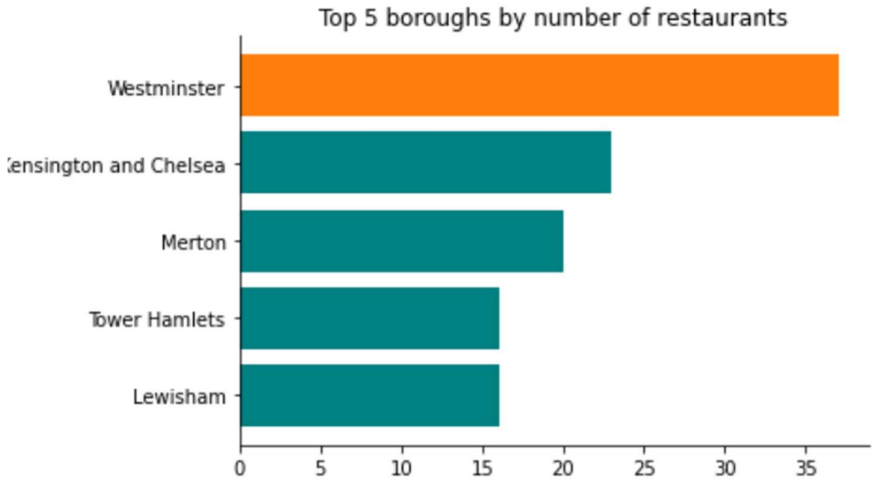
The distribution of venues and restaurants is as follows





4.3 London restaurant distribution

The distribution of restaurants among boroughs in London is like so



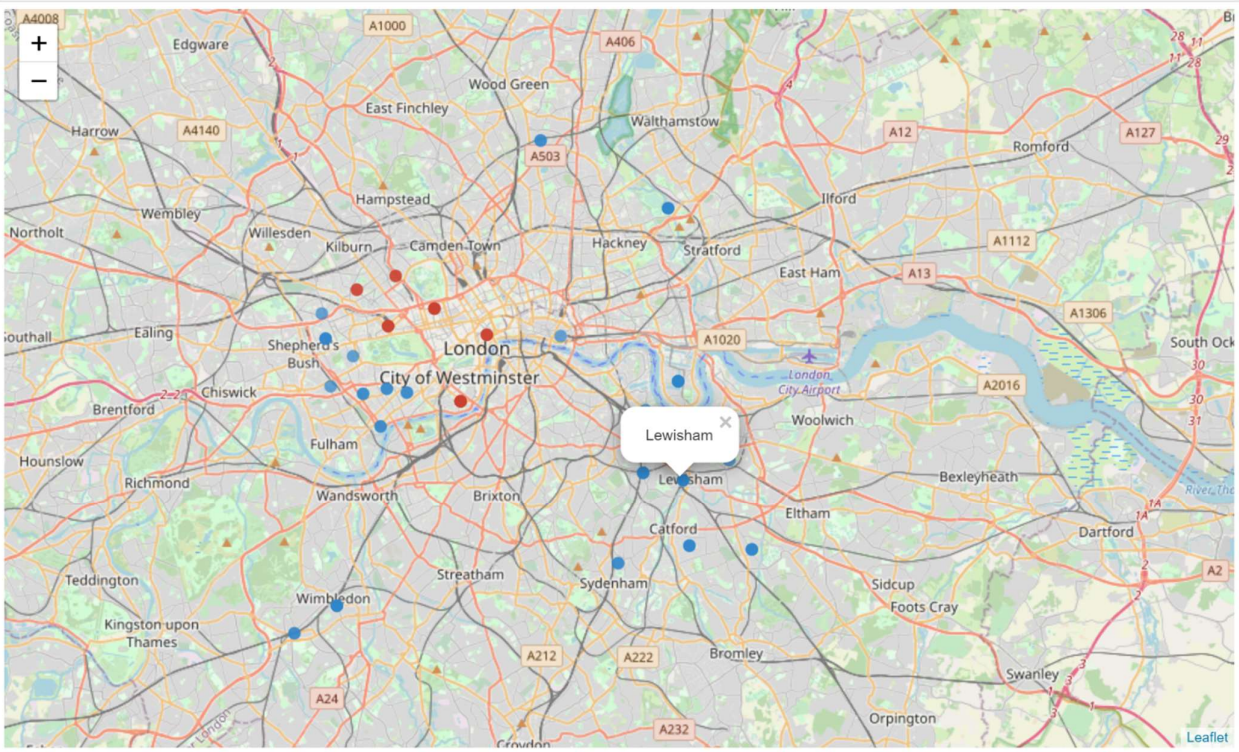
5. Discussion

From the results we can clearly see that London has more unique venue categories and more restaurants than Sydney.

Furthermore, from the box and whisker plots, we can see that the distribution of venues in general, and restaurants in particular is more spread out in London than Sydney. This implies that each borough has something unique to offer and there is plenty to discover for a first time as well as a repeat visitor.

Lastly, from the restaurant distribution in London, we can make two observations:

a. Unsurprisingly Westminster has a large number of restaurants since it is in prime central London
b. However, outside of central London, there are restaurants in other boroughs, which general tend to be lower priced and might provide more variety
   This can be seen in the map below



6. Conclusion

We can conclude from this analysis that there is plenty to offer in terms of points of interest and venues, in order to safely recommend London as a travel destination for a traveller even from a diverse city such as Sydney.

Furthermore, for a traveller interested in trying new restaurants or cuisines, there are lots of options not just in the city centre but spread out across the city of London.

Last but not least, a budget-conscious traveller and equally someone venturing into some of the outer boroughs, the restaurant scene is equally spread so you wouldn't be limited in options regardless of where your journey takes you in London.