# 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.
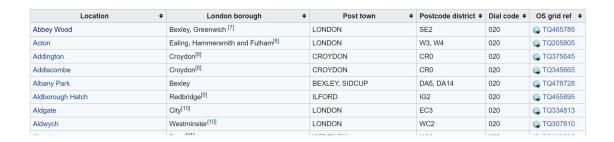
## 2. Data acquisition and cleaning

### 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

Localities in the City of Sydney are:

| | | | | | |
|---|---|---|---|---|---|
| • Broadway | • Chinatown | • The Domain | • Goat Island | • Macdonaldtown | • St James |
| • Central | • Circular Quay | • East Sydney | • Green Square | • Railway Square | • Wynyard |
| • Central Park | • Darling Harbour | • Garden Island | • Kings Cross | • Strawberry Hills | |

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.