

## Data acquisition and cleaning:

A database is acquired from the URL

[https://www.matthewproctor.com/australian\\_postcodes](https://www.matthewproctor.com/australian_postcodes). This database contains geographic information about all the suburbs in all the states of Australia.

**Table 1: Geographic information of Australian states.**

	id	postcode	locality	state	long	lat	dc	type	status	sa3	sa3name	sa4	sa4name	region
0	230	200		ANU	ACT	0.000000	0.000000	NaN	NaN	NaN	NaN	NaN	NaN	R1
1	21820	200	Australian National University	ACT	149.118900	-35.277700	NaN	NaN	Added 19-Jan-2020	NaN	NaN	NaN	NaN	R1
2	232	800	DARWIN	NT	130.836680	-12.458684	NaN	NaN	Updated 6-Feb-2020	70101.0	Darwin City	701.0	Darwin	R1
3	233	801	DARWIN	NT	130.836680	-12.458684	NaN	NaN	Updated 25-Mar-2020 SA3	70101.0	Darwin City	701.0	Darwin	R1
4	234	804	PARAP	NT	130.873315	-12.428017	NaN	NaN	Updated 25-Mar-2020 SA3	70102.0	Darwin Suburbs	701.0	Darwin	R1

The city of Melbourne exists in Victoria state, the above table is filtered for the state of Victoria and the information about the other states is dropped. The only information required from the above table after filtering is postcode, locality, longitude, latitude, and council area. Dropped all the unnecessary columns, renamed the sa3name column to the council, and reset the index. The cleaned data looks like the table below.

**Table 2: Geographic data of Victoria.**

	postcode	locality	long	lat	council
0	3000	MELBOURNE	144.956776	-37.817403	Melbourne City
1	3001	MELBOURNE	144.956776	-37.817403	Port Phillip
2	3002	EAST MELBOURNE	144.982207	-37.818517	Melbourne City
3	3003	WEST MELBOURNE	144.949592	-37.810871	Melbourne City
4	3004	MELBOURNE	144.970161	-37.844246	Port Phillip

In Victoria, some big suburbs have more than one postcode, which made the suburb to be duplicated in the table. The duplicated suburbs are dropped to keep one entry of each suburb.

Also, the above table contains information about all the suburbs in the state of Victoria. However, for this project, only the information about the suburbs in the greater Melbourne metropolitan area is required. To extract the information of Melbourne's suburbs, another table is built which contains the names of all the suburbs in Melbourne.

**Table 3: List of Melbourne suburbs.**

Suburb	
0	Abbotsford
1	Aberfeldie
2	Aintree
3	Airport West
4	Albanvale

Both tables are inner merged and finally, get a data-frame containing the geographic information about all the suburbs in Melbourne city.

**Table 4: Geographic information about Melbourne suburbs.**

	postcode	suburb	council	lat	long
0	3000	MELBOURNE	Melbourne City	-37.817403	144.956776
1	3002	EAST MELBOURNE	Melbourne City	-37.818517	144.982207
2	3003	WEST MELBOURNE	Melbourne City	-37.810871	144.949592
3	3006	SOUTHBANK	Port Phillip	-37.823258	144.965926
4	3008	DOCKLANDS	Melbourne City	-37.814719	144.948039

### **How the above data will be used to solve the problem?**

Many small suburbs in Melbourne city have a similar postcode. All such suburbs will be combined to condense the data. Then all the suburbs of Melbourne will be displayed on a map using the Folium library of Python. The purpose of displaying the suburbs on the map is to see the vicinity of the suburbs.

In the next step, for each suburb, the ten most famous venues will be fetched using Foursquare API. The newly created data will be clustered using the scikit-learn library of Python. The optimized number of clusters will be created by running the clustering function for different values of the number of clusters.

These clusters will be displayed on the map to visually observe, how these clusters are covering the Melbourne region. Each cluster will be individually inspected for a variety of venues. Also, each cluster will be inspected for having Thai restaurants.

The cluster having most Thai restaurants and other kinds of dining places in different suburbs will be finalized for further evaluation. The selected cluster will be further summarized to the suburbs having a Thai restaurant in five topmost famous venues. All these suburbs will be displayed on the map to see the distance of these suburbs from the busiest destinations in Melbourne city.

In the last step, these suburbs will be ranked based on the vicinity and the presence of fine dining places in the suburb. The top three suburbs will be suggested to James to open his new Thai restaurant.