The Battle of Neighborhoods

Jason Kenner

Data

Data sources. Data was scraped from several sources, and also obtained from Foursquare. Firstly, postcode data including latitude and longitude was obtained from https://www.matthewproctor.com/australian_postcodes as a csv file. It appeared as such in a data frame:

	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	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

Also, demographic data was obtained from the Australian Bureau of Statistics http://stat.data.abs.gov.au/index.aspx# also as a csv file. It was a much more complex set of data including many rows with the same postcode:

	POA	Postal Area Code	SEIFAINDEXTYPE	Index Type	SEIFA_MEASURE	Measure	TIME	Time	Value	Flag Codes	Flags
0	800	800	IEO	Index of Education and Occupation	SCORE	Score	2016.0	2016.0	1089.0	NaN	NaN
1	800	800	IEO	Index of Education and Occupation	RWAR	Rank within Australia	2016.0	2016.0	2287.0	NaN	NaN
2	800	800	IEO	Index of Education and Occupation	RWAD	Rank within Australia - Decile	2016.0	2016.0	9.0	NaN	NaN
3	800	800	IEO	Index of Education and Occupation	RWAP	Rank within Australia - Percentile	2016.0	2016.0	87.0	NaN	NaN
4	800	800	IEO	Index of Education and Occupation	RWSR	Rank within State or Territory	2016.0	2016.0	33.0	NaN	NaN

Finally, location data was obtained from Foursquare https://developer.foursquare.com as a json file.

	Locality	Locality lat	Locality long	Bar	Bar lat	Bar long	Bar category
0	MELBOURNE	-37.817403	144.956776	Dikstein's Corner Bar	-37.816189	144.960353	Bar
1	MELBOURNE	-37.817403	144.956776	The Irish Times	-37.816135	144.960563	Bar
2	MELBOURNE	-37.817403	144.956776	MoVida Terraza	-37.814688	144.958567	Bar
3	MELBOURNE	-37.817403	144.956776	Bonnie Coffee Brewers	-37.818153	144.957636	Coffee Shop
4	MELBOURNE	-37.817403	144.956776	Patricia Coffee Brewers	-37.814598	144.958350	Coffee Shop
5	MELBOURNE	-37.817403	144.956776	The Lui Bar	-37.819067	144.957739	Cocktail Bar
6	MELBOURNE	-37.817403	144.956776	Shamble Coffee Brewers	-37.816056	144.960779	Café
7	MELBOURNE	-37.817403	144.956776	The Deck	-37.820254	144.957515	Bar
8	MELBOURNE	-37.817403	144.956776	Syracuse	-37.816207	144.960253	Restaurant
9	MELBOURNE	-37.817403	144.956776	Saint & Rogue	-37.817512	144.955491	Bar

Data cleaning. The postcode and demographic data required cleaning, including removal of unwanted columns and rows, renaming headings and merging. The demographic data was firstly made into three new data frames, one for each demographic data point, then those data frames combined, and finally merged with the postcode data. The result included postcode, latitude, longitude and three demographic measures: Index of Education and Occupation, Index of Economic Resources, and Index of Social Advantage and Disadvantage for each suburb. The dataframe was then reduced to only include suburbs within a 5km radius of the Melbourne City Centre using a csv file obtained from FreeMapTools: https://www.freemaptools.com/find-australian-postcodes-inside-radius.htm. The final data frame of inner suburb data was created by calling the postcode subset data frame using .isin:

	postcode	locality	state	long	lat	ieo rank	ier rank	irsad rank
0	3000	MELBOURNE	VIC	144.956776	-37.817403	10.0	1.0	8.0
1	3002	EAST MELBOURNE	VIC	144.982207	-37.818517	10.0	4.0	10.0
2	3003	WEST MELBOURNE	VIC	144.949592	-37.810871	10.0	1.0	10.0
3	3004	MELBOURNE	VIC	144.970161	-37.844246	10.0	2.0	10.0
4	3004	ST KILDA ROAD CENTRAL	VIC	144.970161	-37.844246	10.0	2.0	10.0

The Foursquare data was also grouped by suburb into a new data frame and combined with the demographic data for each suburb for the clustering analysis allowing a single data frame that also included total number of bars for each suburb:

	postcode	locality	long	lat	ieo rank	ier rank	irsad rank	Bars
0	3000	MELBOURNE	144.956776	-37.817403	10.0	1.0	8.0	54.0
1	3002	EAST MELBOURNE	144.982207	-37.818517	10.0	4.0	10.0	6.0
2	3003	WEST MELBOURNE	144.949592	-37.810871	10.0	1.0	10.0	6.0
3	3004	MELBOURNE	144.970161	-37.844246	10.0	2.0	10.0	54.0
4	3005	WORLD TRADE CENTRE	144.950858	-37.824608	10.0	1.0	10.0	11.0