

# Applied Data Science Capstone

## **Business Problem**

A group of medical professionals seek to invest in a rejuvenation clinic that provides infusion-based therapy to clients. Their target clientele are young, professional women in Melbourne, Australia. The investors seek to find a location where the demographics include a large population of young professional women who live in suburbs associated with high-income status. The location must be within 800 metres of a train station. The investors would also like to know the location statistics of current successful clinics, most notably, popular venues according to social media. The venue location statistics from these successful clinics will be looked upon when selecting the final site location.

## **Data**

The data for this problem will come from a variety of sources. Suburbs across Melbourne will be classed as 'high income' according to Australian Census data released from the Australian Taxation Office. The top 10 postcodes who have the highest average taxable income in Victoria will be selected from the 2011/2012, 2012/2013, 2013/2014, 2014/2015, 2015/2016 financial years, respectively. From there, data from the Australian Bureau of Statistics will determine which suburbs have the highest population of young, professional women; that is, women who are 20<34 years of age, and have an income > \$1, 250 per week. This data will be used to construct a cluster map to determine possible site locations.

Using Foursquare, venue location data from current successful anti-aging clinics will then be analysed to determine the most common venues associated with these successful clinics. Each potential site location will then be ranked according to how 'associated' each suburb is to the venue location data. The suburb with the most associated venues will be the site location for the rejuvenation clinic.

## **Methodology**

This report consists of four sections. The first section involved downloading and pre-processing the required datasets. After determining top ten postcodes, only those postcodes with a population of >1, 000 20-34-year-olds were selected for further analysis. Part 2 involved creating a cluster map of Melbourne with potential site locations for the rejuvenation clinic. Part 3 involved attaining venue location data from current successful anti-aging clinics. A google maps query of 'anti-aging clinics' in Melbourne was conducted and only those with a 5-star customer rating were further analysed. Analysis involved a Foursquare API search of most popular venues associated with the successful clinics.

Part 4 involved searching for an ideal location for the clinic. This was done by adding parameters from the previous section to search query. The suburb/postcode with the most associated venues in the parameter search was the suburb/postcode for site location. The parameters were location site must be within 800 metres of a cafe, gym/fitness centre, bookstore, grocery store and train station.

### Results:

Results show a total of 15 postcodes qualified for high-income suburbs based on data provided by the Australian Taxation Office (Table 1). Demographics were obtained from the Australian Bureau of Statistics; Table 2 shows the number of persons aged between 20-34 per postcode. Table 3 shows results after adding latitude and longitude, and selecting for populations >1,000 and women aged >20<34 and earning >\$1, 250 per week. Figure 1 shows the data on a map.

**Table 1 – High-income Suburbs.** Top 10 Postcodes in Victoria with high-income from 2011-2016.

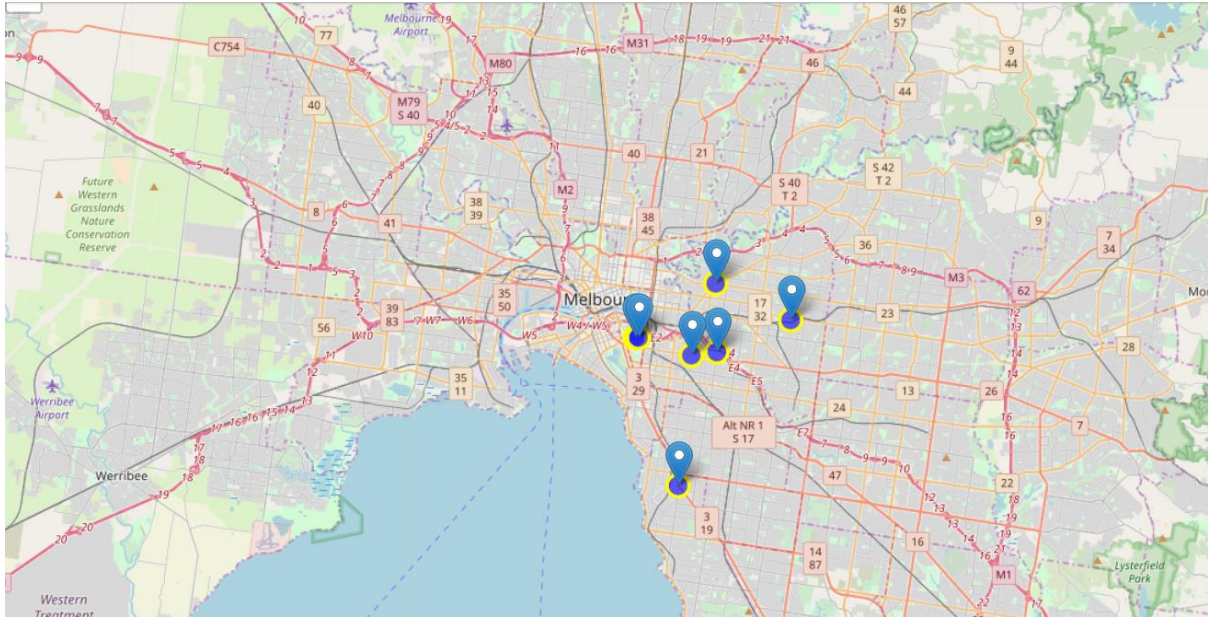
	Postcode	Suburb
0	3142	HAWKS BURN, TOORAK
1	3944	PORTSEA
2	3186	BRIGHTON, BRIGHTON NORTH, DENDY
3	3126	CAMBERWELL EAST, CANTERBURY
4	3206	ALBERT PARK, MIDDLE PARK
5	3144	KOOYONG, MALVERN, MALVERN NORTH
6	3002	EAST MELBOURNE
7	3929	FLINDERS
8	3143	ARMADALE, ARMADALE NORTH
9	3101	COTHAM, KEW
10	3761	ST ANDREWS
11	3874	MCLOUGHLINS BEACH, WOODSIDE, WOODSIDE BEACH, W...
12	3928	MAIN RIDGE
13	3141	SOUTH YARRA
14	3813	TYNONG, TYNONG NORTH

**Table 2 – Number of Persons Aged >20<34 per Postcode.** Results obtained from Australian Census data.

	Postcode	Suburb	Latitude	Longitude	n20_34
0	3142	HAWKSBUrn, TOORAK	-37.8416	145.016	2997
1	3944	PORTSEA	-38.3167	144.717	16
2	3186	BRIGHTON, BRIGHTON NORTH, DENDY	-37.9056	145.008	3196
3	3126	CAMBERWELL EAST, CANTERBURY	-37.8247	145.078	1272
4	3206	ALBERT PARK, MIDDLE PARK	-37.8333	144.981	1663
5	3144	KOOYONG, MALVERN, MALVERN NORTH	-37.8399	145.032	2139
6	3002	EAST MELBOURNE	-38.4739	145.019	1750
7	3929	FLINDERS	-37.85	145.027	59
8	3143	ARMADALE, ARMADALE NORTH	-37.8064	145.031	2545
9	3101	COTHAM, KEW	-37.5782	145.307	5319
10	3761	ST ANDREWS	-38.5238	146.804	200
11	3874	MCLOUGHLINS BEACH, WOODSIDE, WOODSIDE BEACH, W...	-38.3973	144.972	44
12	3928	MAIN RIDGE	-38.3973	144.972	29
13	3141	SOUTH YARRA	-37.8333	144.983	11723
14	3813	TYNONG, TYNONG NORTH	-38.0834	145.621	119

**Table 3 – Final demographics for ideal clientele.**

	Postcode	Suburb	Latitude	Longitude	n20_34	nwomen_high_income
0	3142	HAWKSBUrn, TOORAK	-37.8416	145.016	2997	447
2	3186	BRIGHTON, BRIGHTON NORTH, DENDY	-37.9056	145.008	3196	427
3	3126	CAMBERWELL EAST, CANTERBURY	-37.8247	145.078	1272	128
4	3206	ALBERT PARK, MIDDLE PARK	-37.8333	144.981	1663	298
5	3144	KOOYONG, MALVERN, MALVERN NORTH	-37.8399	145.032	2139	294
6	3002	EAST MELBOURNE	-38.4739	145.019	1750	371
8	3143	ARMADALE, ARMADALE NORTH	-37.8064	145.031	2545	387
9	3101	COTHAM, KEW	-37.5782	145.307	5319	639
13	3141	SOUTH YARRA	-37.8333	144.983	11723	2092



**Figure 1 – Map of all possible site locations for the rejuvenation clinic.**

Location data results showed a total of five 5-star anti-aging clinics were available for further analysis (Table 4). Table 5 shows the most common venues associated with each postcode/suburb. These venues will then used as parameters for selecting potential site location.

**Table 4 – Query results for most popular venues at successful clinics.**

	Venue Number	Postcode	Latitude	Longitude
0	1	3149	-37.885026	145.129849
1	2	3127	-37.823632	145.095292
2	3	3144	-37.852266	145.034585
3	4	3126	-37.824140	145.078873
4	5	3123	-37.824079	145.048312

**Table 5 – Sum of all unique venues found within an 800-metre radius of the successful clinics.** The top 4 venues are Bar, Yoga Studio, Café, and Beer Garden. The parameters for location site search are Gym/fitness centre, Beer Garden

Bar	4
Yoga Studio	4
Café	4
Beer Garden	3
Grocery Store	3
Bookstore	3
Food	2
Thai Restaurant	2
Pub	2
Train Station	2
Shopping Mall	2
Burger Joint	2
Dance Studio	2



**Markovska81**

Parameters: Café, Grocery Store, Bookstore, Gym/Fitness Centre, Train Station



**Markovska81**

Bar and Beer Garden not in site location search



**Markovska81**

Yoga replaced by Gym/Fitness Centre in site location search

The location site search results (Table 6) show none of the suburbs/postcodes fulfilled all search criteria. Postcodes 3142, 3144 and 3206 returned only 1 search criteria each. Postcodes 3141 and 3002 returned 2 search criteria each, whilst 3186 and 3143 returned 3 search criteria and 3126 returned 4. Since the investors were specific only about train station proximity, postcodes 3126, 3144 and 3186 were the only ones to fulfil this criterion (Table 7). Camberwell East and Canterbury had the most associated venues.

**Table 6 – Results from location site search.** No suburb/postcode captured all parameters.

	Postcode	Café	Bookstore	Train Station	Gym / Fitness Center	Grocery Store
0	3002	3	0	0	0	1
1	3126	4	1	1	0	1
2	3141	1	0	0	0	1
3	3142	0	0	0	0	1
4	3143	6	1	0	0	2
5	3144	0	0	1	0	0
6	3186	3	0	1	1	0
7	3206	1	0	0	0	0

**Table 7 – Final Search Results for Location Site**

	Postcode	Café	Bookstore	Gym / Fitness Center	Grocery Store	Suburb	nwomen_high_income
1	3126	4	1	0	1	CAMBERWELL EAST, CANTERBURY	128
5	3144	0	0	0	0	KOOYONG, MALVERN, MALVERN NORTH	294
6	3186	3	0	1	0	BRIGHTON, BRIGHTON NORTH, DENDY	427

## **Discussion:**

The results showed Camberwell East and Canterbury to be the best location site for the new clinic. The downside to this is the rather low number of young, professional women compared to other postcodes/suburbs. Brighton, Brighton North and Dendy (3186) are ranked in second place for location site as they have a total of 5 associated venues, but they do have a rather large population of young, professional women compared to Camberwell East and Canterbury.

Yoga studios were associated with 4/5 successful clinics; however query results returned no yoga studios when searching for ideal location. This was later changed to gym/fitness centre as most gyms/fitness centres contain yoga classes/groups. Only 3186 returned a positive result for gym/fitness centre. The next step would be for the investors to experience firsthand the 3186 and 3126 business districts to find an ideal location site.

## **Conclusion**

The purpose of this study was to select a site location for a group of investors specific to their clientele. The location site had to be within 800 metres of a train station and had to have associated venues with current successful clinics. An associated venue search for current successful clinics found cafe, bakery and yoga studios to be most associated with these clinics. Camberwell East and Canterbury were the suburbs with the most associated venues from the successful clinics. A downside to Camberwell East and Canterbury is the number of young, professional women; the lowest of all suburbs.

Limitations of this study include geographic isolation. These results are based in Melbourne, Australia and it is hypothesized different search results will be produced for different geographic locations. Also, these search results are not certain, given sporadic opening and closing of venues. If a site location were to be chosen based on associated venues, these venues could change location/venue anytime.