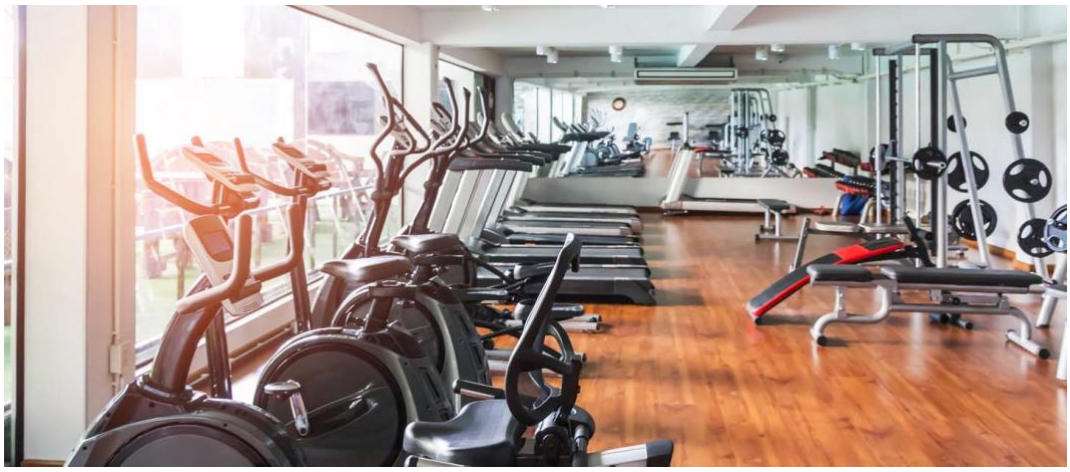
IBM Data Science Capstone Project

--Find the best Gym-business Opportunity in

Melbourne Northern Region



Yige Li

1. Introduction

The purpose of this report is to find the ideal Gym-business opportunity in the north part of Melbourne in Australia by analysing geographic data.

Melbourne northern region is a diverse and vibrant region, featuring Melbourne's major airport, arts and cultural precincts the National Employment and Innovation Cluster in La Trobe and new growth area communities on the northern fringe of the city. The northern region's growing population and economy make more people wants to start Gym-business in this region. This report will focus on analysing the northern region (Especially the government area--City of Banyule), Which includes about 20 suburbs. This report will analyse the data of these suburbs and find the ideal business type for starting a gym-business in one suburb.

2. Problem description

As the development of Melbourne northern region, more and more modern people realize the importance of gym exercise. A positive and energetic work force collaborates better and is more appealing to customers, which means investing in a gym might be an ideal choice to start a business. The following points should be considered for a gym in chosen area:

1) Distance from the Suburb centre
2) Difference between gyms
3) Number of gyms in the near area/suburbs (number of competitors)

3. Limitation

The limitation of this report is that it only based on limited data and some other reason are not considered such as suburb income, social environment and economy.

4. Data

In order to address the above questions and to solve the business problem, this analysis will use following data:

1) **List of Melbourne suburbs and relevant data:**
   https://en.wikipedia.org/wiki/List_of_Melbourne_suburbs

	Suburb	Postcode	Local government area	Location[citation needed]	Distance[3][citation needed]	Area[citation needed]	Population[citation needed]	Population density[citation needed]	Dat
0	Bellfield	3081	City of Banyule	NaN	NaN	0.9 km2	1,793[4]	NaN	
1	Briar Hill	3088	City of Banyule	NaN	NaN	NaN	3,152[4]	NaN	
2	Bundoora	3083	City of Banyule; City of Darebin; City of Whit...	NaN	NaN	15 km2	28653	NaN	
3	Eaglemont	3084	City of Banyule	NaN	NaN	1.9 km2	3873	NaN	
4	Eltham	3095	City of Banyule; Shire of Nillumbik	NaN	NaN	NaN	NaN	NaN	
5	Eltham North	3095	City of Banyule; Shire of Nillumbik	NaN	NaN	NaN	NaN	NaN	
6	Greensborough	3088	City of Banyule; Shire of Nillumbik	NaN	NaN	NaN	NaN	NaN	
7	Heidelberg	3084	City of Banyule	NaN	NaN	NaN	NaN	NaN	
8	Heidelberg Heights	3081	City of Banyule	NaN	NaN	NaN	NaN	NaN	
9	Heidelberg West	3081	City of Banyule	NaN	NaN	NaN	NaN	NaN	
10	Ivanhoe	3079	City of Banyule	NaN	NaN	NaN	NaN	NaN	
11	Ivanhoe East	3079	City of Banyule	NaN	NaN	NaN	NaN	NaN	
12	Lower Plenty	3093	City of Banyule	NaN	NaN	NaN	NaN	NaN	
13	Macleod	3085	City of Banyule; City of Darebin	NaN	NaN	NaN	NaN	NaN	
14	Montmorency	3094	City of Banyule	NaN	NaN	NaN	NaN	NaN	
15	Rosanna	3084	City of Banyule	NaN	NaN	NaN	NaN	NaN	
16	St Helena	3088	City of Banyule	NaN	NaN	NaN	NaN	NaN	
17	Viewbank	3084	City of Banyule	NaN	NaN	NaN	NaN	NaN	
18	Watsonia	3087	City of Banyule	NaN	NaN	NaN	NaN	NaN	

Figure 4.1 List of suburbs in the City of Banyule

2) Geographic/location data of all relevant suburbs

	Suburb	Postcode	Latitude	Longitude
0	Bellfield	3081	-37.742466	145.045725
1	Briar Hill	3088	-37.699765	145.110052
2	Bundoora	3083	-37.702825	145.056074
3	Eaglemont	3084	-37.748655	145.077964
4	Eltham	3095	-37.713485	145.167738
5	Eltham North	3095	-37.713485	145.167738
6	Greensborough	3088	-37.699765	145.110052
7	Heidelberg	3084	-37.748655	145.077964
8	Heidelberg Heights	3081	-37.742466	145.045725
9	Heidelberg West	3081	-37.742466	145.045725
10	Ivanhoe	3079	-37.770370	145.055245
11	Ivanhoe East	3079	-37.770370	145.055245
12	Lower Plenty	3093	-37.737875	145.117955
13	Macleod	3085	-37.721715	145.082647
14	Montmorency	3094	-37.720285	145.123809
15	Rosanna	3084	-37.748655	145.077964
16	St Helena	3088	-37.699765	145.110052
17	Viewbank	3084	-37.748655	145.077964
18	Watsonia	3087	-37.703674	145.085045
19	Watsonia North	3087	-37.703674	145.085045

Figure 4.2 Table of suburbs with location data

3) Venue data and business information of the all neighbourhoods of the target suburbs

Suburb	Suburb Latitude	Suburb Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bellfield	29	29	29	29	29	29
Briar Hill	23	23	23	23	23	23
Bundoora	30	30	30	30	30	30
Eaglemont	30	30	30	30	30	30
Eltham	5	5	5	5	5	5
Eltham North	5	5	5	5	5	5
Greensborough	23	23	23	23	23	23
Heidelberg	30	30	30	30	30	30
Heidelberg Heights	29	29	29	29	29	29
Heidelberg West	29	29	29	29	29	29
Ivanhoe	30	30	30	30	30	30
Ivanhoe East	30	30	30	30	30	30
Lower Plenty	8	8	8	8	8	8
Macleod	17	17	17	17	17	17
Montmorency	14	14	14	14	14	14
Rosanna	30	30	30	30	30	30
St Helena	23	23	23	23	23	23
Viewbank	30	30	30	30	30	30
Watsonia	17	17	17	17	17	17
Watsonia North	17	17	17	17	17	17

Figure 4.3 Count of venue of each suburb

Suburb	Art Gallery	Arts & Crafts Store	Athletics & Sports	Australian Restaurant	Bakery	Basketball Court	Big Box Store	Breakfast Spot	Burger Joint	Café	Candy Store	Chinese Restaurant	Church	Coffee Shop	Convenience Store	Department Store	Dessert Shop	Dumpling Restaurant	Electronics Store	Farm
Bellfield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bellfield	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bellfield	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Bellfield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Bellfield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4.4 Business categories

5. Methodology

After collected the data, we mainly use two methodologies (frequency and K-Mean Clustering) to explore more information.

In order to get a better idea about each suburb what types of venues are popular, we sort the mean of the frequency of occurrence of each category. Result is shown in Figure 4.1 below.

Suburb	Art Gallery	Arts & Crafts Store	Asian Restaurant	Australian Restaurant	Bakery	Basketball Court	Big Box Store	Burger Joint	Café	Candy Store	Chinese Restaurant	Church	Coffee Shop	Construction & Landscaping	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store
Bellfield	0.000000	0.034483	0.000000	0.0	0.034483	0.000000	0.034483	0.068966	0.103448	0.000000	0.034483	0.0	0.034483	0.0	0.000000	0.000000	0.000000	0.034483
Briar Hill	0.000000	0.000000	0.000000	0.0	0.000000	0.041667	0.041667	0.000000	0.125000	0.041667	0.000000	0.0	0.000000	0.0	0.000000	0.041667	0.041667	0.041667
Bundoora	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.000000
Eaglemont	0.033333	0.033333	0.033333	0.0	0.000000	0.000000	0.000000	0.000000	0.233333	0.000000	0.000000	0.0	0.000000	0.0	0.033333	0.000000	0.000000	0.033333
Eltham	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.250000	0.000000	0.000000	0.000000
Eltham North	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.250000	0.000000	0.000000	0.000000
Greensborough	0.000000	0.000000	0.000000	0.0	0.000000	0.041667	0.041667	0.000000	0.125000	0.041667	0.000000	0.0	0.000000	0.0	0.000000	0.041667	0.041667	0.041667
Heidelberg	0.033333	0.033333	0.033333	0.0	0.000000	0.000000	0.000000	0.000000	0.233333	0.000000	0.000000	0.0	0.000000	0.0	0.033333	0.000000	0.000000	0.033333
Heidelberg Heights	0.000000	0.034483	0.000000	0.0	0.034483	0.000000	0.034483	0.068966	0.103448	0.000000	0.034483	0.0	0.034483	0.0	0.000000	0.000000	0.000000	0.034483
Heidelberg West	0.000000	0.034483	0.000000	0.0	0.034483	0.000000	0.034483	0.068966	0.103448	0.000000	0.034483	0.0	0.034483	0.0	0.000000	0.000000	0.000000	0.034483

Figure 5.1 The mean of the frequency of occurrence of each category

After we know the mean of the frequency of occurrence of each category, K-Mean Clustering is used, and the 19 suburbs are clustered into 3 groups. Every point is assigned to a cluster. The result of the 3 groups are shown in the figure below:

	Suburb	Gym	Cluster Labels	Latitude	Longitude
0	Bellfield	0.034483	1	-37.742466	145.045725
1	Briar Hill	0.000000	0	-37.699765	145.110052
2	Bundoora	0.000000	0	-37.702825	145.056074
3	Eaglemont	0.000000	0	-37.748655	145.077964
4	Eltham	0.000000	0	-37.713485	145.167738
5	Eltham North	0.000000	0	-37.713485	145.167738
6	Greensborough	0.000000	0	-37.699765	145.110052
7	Heidelberg	0.000000	0	-37.748655	145.077964
8	Heidelberg Heights	0.034483	1	-37.742466	145.045725
9	Heidelberg West	0.034483	1	-37.742466	145.045725
10	Ivanhoe	0.033333	1	-37.770370	145.055245
11	Ivanhoe East	0.033333	1	-37.770370	145.055245
12	Lower Plenty	0.000000	0	-37.737875	145.117955
13	Macleod	0.058824	2	-37.721715	145.082647
14	Montmorency	0.000000	0	-37.720285	145.123809
15	Rosanna	0.000000	0	-37.748655	145.077964
16	St Helena	0.000000	0	-37.699765	145.110052
17	Viewbank	0.000000	0	-37.748655	145.077964
18	Watsonia	0.000000	0	-37.703674	145.085045
19	Watsonia North	0.000000	0	-37.703674	145.085045

Figure 5.2 Cluster group of suburbs

	Suburb	Gym	Cluster Labels	Latitude	Longitude
13	Macleod	0.058824	2	-37.721715	145.082647

	Suburb	Gym	Cluster Labels
0	Bellfield	0.034483	1
8	Heidelberg Heights	0.034483	1
9	Heidelberg West	0.034483	1
10	Ivanhoe	0.033333	1
11	Ivanhoe East	0.033333	1

Figure 5.3 Cluster group 1,2

6. Results

The following map shows the map of 10 prefered suburbs of City of Banyule assigned to 3 clusters. Green point is the suburb (Macleod) which has the most number of gyms, Purple points are less, Red points are least.

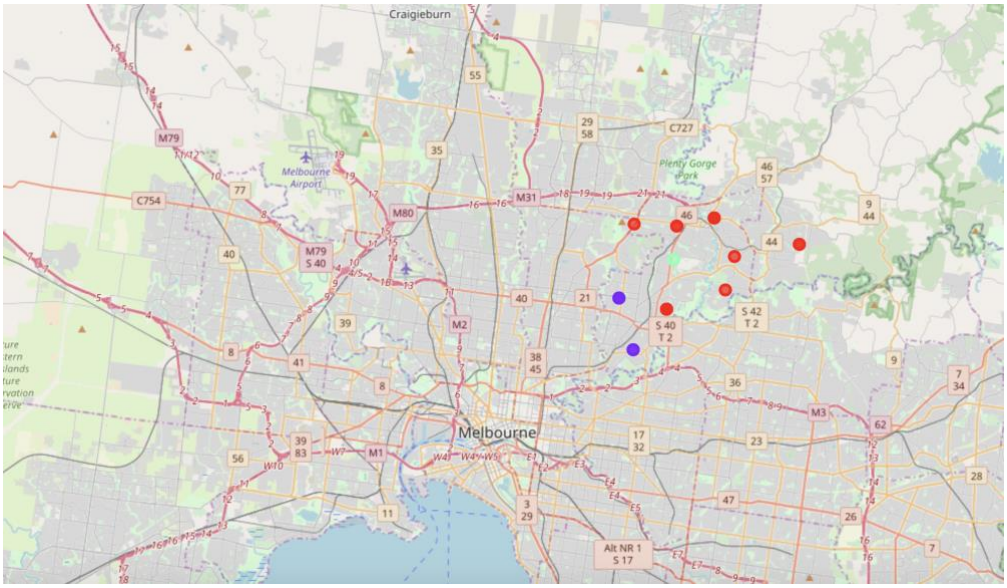


Figure 6.1 Clustered suburbs in the map

We also get the most 8 common venue of each suburb to understand the business environments of each suburb. The result is shown as below.

	Suburb	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Bellfield	Café	Sandwich Place	Burger Joint	Department Store	Coffee Shop	Dessert Shop	Pet Store	Park
1	Briar Hill	Café	Fast Food Restaurant	Supermarket	Gym / Fitness Center	Park	Mexican Restaurant	Pharmacy	Department Store
2	Bundoora	Supermarket	Pharmacy	Park	Pizza Place	Shopping Mall	Café	Vietnamese Restaurant	Paper / Office Supplies Store
3	Eaglemont	Café	Park	Pizza Place	Italian Restaurant	Grocery Store	Theater	Pub	Hotel Bar
4	Eltham	Hotel	Playground	Convenience Store	Café	Pizza Place	Golf Course	Gas Station	Football Stadium
5	Eltham North	Hotel	Playground	Convenience Store	Café	Pizza Place	Golf Course	Gas Station	Football Stadium

Figure 6.2 Most common venue of each suburb

7. Discussion

As can be seen from Figure 5.3 and Figure 6.1, 5 suburbs have the highest mean of the frequency of occurrence of each category. However, Gym is the 4th most common venue in Briar hill. If we take population of each suburb into consideration, we could recommend that the suburb within the 5 and with the largest population is the ideal one to invest. However, it would also be a good choice if we want to invest a gym in a suburb that with the relative high population in group 0, because it would have less competitor. The final decision will depend on how much risk the investor would tolerant when taking other factors into consideration.

8. Conclusion

This report first introduces the current situation of Melbourne northern region and the problem of investing a gym. Then it gives a more specific description of problem. Mainly three types of data are collected for analysis—suburbs, location data and information from Foursquare. Based on the data, it gets the frequency of venue and use k-mean cluster to further virtualize

the suburbs with gym data. In conclusion, we could recommend that the suburb within the 5 and with the largest population is the ideal one to invest. However, it would also be a good choice if we want to invest a gym in a suburb that with the relative high population but in group 0. Although, the final decision will depend on many other factors, the result from the data analysis is a helpful reference for investor to make the decision.