

# Exploratory Analysis of Sydney's Best Fitness Center Location Openings

Capstone Project – Coursera IBM Data Science

By Kristobal Esgana

# Contents

1. Introduction .....	2
a. Background .....	2
b. Problem.....	2
c. Aim.....	2
2. Data .....	2
a. Populations data .....	2
b. Locations data .....	3
c. Foursquare Data .....	3
3. Methodology .....	3
a. Loading and Processing the data.....	3
b. Clustering the Dataset.....	4
c. Analytical explorations.....	4
d. Exploring the Top 10 Populations of Sydney's Suburban Regions.....	4
4. Results.....	4
a. Breakdown of most common fitness center venues.....	4
b. Cluster overview and analysis .....	6
c. Top 10 populations of Sydney's Suburban Regions.....	7
5. Discussion .....	8
6. Conclusion.....	8
7. References .....	8

# 1. Introduction

## a. Background

Sydney is one of the most populous locations of Australia with its city areas much more densely populated brimming with a lively and busy lifestyle. With health and fitness highly marketed in today's society, many fitness centers have drawn attention to many individuals striving to improve their physical outlook or to find sanctuary to relieve their life's tension and gain a sense of empowerment. As such, many types of fitness centers have widely opened across Sydney and in many of its individual suburbs.

## b. Problem

Opening up a certain type of fitness center such as a boxing Fitness center, climbing fitness center or just a general workout fitness center may prove to be challenging due to competition and distance compared to where it is accessible for most people. Insights on which locations are optimal based on existing feedback data from other fitness centers may also act as a guideline on whether certain areas would be optimal or not as explored in this report.

## c. Aim

This report aims to explore the many types of fitness centers available within Sydney and which are the most thriving. This may prove to be useful in identifying which locations and types of fitness center may be optimal to open based on their popularity within that area. The analysis and findings within this report may provide insight to those interested in opening a certain type of fitness center within Sydney's suburbs.

# 2. Data

The following data below will be used for the data analysis within the report. Each of them will be used in combination to provide insights on the fitness center's specific area locations along with the population numbers within that area.

## a. Populations data

The populations data was extracted from the Australia Bureau of Statistics website for the latest census. This will provide some insights with the population density in a certain region. As seen in the image below of an example of the data, the regions will be joined with the location data according to the SA3 (statistical area 3) column which is based on the regional suburban data. This will be used in conjunction with the population data to have an idea of the population within that area.

	S/T code	S/T name	GCCSA code	GCCSA name	SA4 code	SA4 name	SA3 code	SA3 name	SA2 code	SA2 name	Population
0	1	New South Wales	1RNSW	Rest of NSW	101	Capital Region	10102	Queanbeyan	101021007	Braidwood	4039
1	1	New South Wales	1RNSW	Rest of NSW	101	Capital Region	10102	Queanbeyan	101021008	Karabar	8476
2	1	New South Wales	1RNSW	Rest of NSW	101	Capital Region	10102	Queanbeyan	101021009	Queanbeyan	11340
3	1	New South Wales	1RNSW	Rest of NSW	101	Capital Region	10102	Queanbeyan	101021010	Queanbeyan - East	5011
4	1	New South Wales	1RNSW	Rest of NSW	101	Capital Region	10102	Queanbeyan	101021011	Queanbeyan Region	18188

## b. Locations data

The next data to be used is for the locations data of Australia with coordinates derived from Matthew Proctor's open source project as retrieved from the official Australia Post data. This provides a full list of all the suburbs along with their longitude and latitude locations.

	index	id	postcode	locality	state	long	lat	dc	type	status	sa3	sa3name	sa4	sa4name	region
0	420	458	1001	SYDNEY	NSW	151.268071	-33.794883	EAST SUBS MC	LVR	Updated 25-Mar-2020 SA3	11703.0	Sydney Inner City	117.0	Sydney - City and Inner South	R1
1	421	459	1002	SYDNEY	NSW	151.268071	-33.794883	EAST SUBS MC	LVR	Updated 25-Mar-2020 SA3	11703.0	Sydney Inner City	117.0	Sydney - City and Inner South	R1
2	422	460	1003	SYDNEY	NSW	151.268071	-33.794883	EAST SUBS MC	LVR	Updated 25-Mar-2020 SA3	11703.0	Sydney Inner City	117.0	Sydney - City and Inner South	R1
3	423	461	1004	SYDNEY	NSW	151.268071	-33.794883	EAST SUBS MC	LVR	Updated 25-Mar-2020 SA3	11703.0	Sydney Inner City	117.0	Sydney - City and Inner South	R1
4	424	462	1005	SYDNEY	NSW	151.268071	-33.794883	GPO BOX CENTRE	LVR	Updated 25-Mar-2020 SA3	11703.0	Sydney Inner City	117.0	Sydney - City and Inner South	R1

## c. Foursquare Data

The data from Foursquare will be used to compile a list of all the fitness centers within Sydney and their associated popularity along with their location and coordinates. This will be used in combination with the populations and locations data to understand which specific regional suburban area it belongs to and the population for potential customers within that area. The top 5 most common venues will be stored into the dataset.

	Suburban_Region	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Auburn	Gym / Fitness Center	Gym	Climbing Gym	Yoga Studio	Track
1	Bankstown	Gym	Yoga Studio	Track	Pool	Pilates Studio
2	Baulkham Hills	Yoga Studio	Gym	Track	Pool	Pilates Studio
3	Blacktown	Gym	Gym / Fitness Center	Yoga Studio	Pool	Track
4	Blacktown - North	Gym	Pool	Martial Arts Dojo	Yoga Studio	Track

## 3. Methodology

### a. Loading and Processing the data

Firstly, the locations and populations data will be loaded from their csv files and cleaned up to ensure that only the necessary columns are used. Once done, these datasets will then be combined which will be used to feed into a function to generate the foursquare data. Once done, this should provide the full dataset of all the fitness centers within the geographical coordinates, population and the suburban region associated with it.

The foursquare data will then be assigned with a one hot encoding method to assign a value of 1 for each type of fitness center that exists within the suburban location. This is so the numbers can then be used later on to cluster them based on how common the venue is within similar regions.

Suburban_Region	Basketball Court	Climbing Gym	College Rec Center	Coworking Space	Dance Studio	Gym	Gym / Fitness Center	Gym Pool	Martial Arts Dojo	Physical Therapist	Pilates Studio	Pool	Track	Yoga Studio
0	Baulkham Hills	0	0	0	0	1	0	0	0	0	0	0	0	0
1	Baulkham Hills	0	0	0	0	0	0	0	0	0	0	0	0	1
2	Blacktown	0	0	0	0	1	0	0	0	0	0	0	0	0
3	Blacktown	0	0	0	0	0	1	0	0	0	0	0	0	0
4	Blacktown	0	0	0	0	0	1	0	0	0	0	0	0	0

## b. Clustering the Dataset

The foursquare data will then be clustered using the k-means clustering method with up to 5 clusters for optimal clustering classification. This should then produce the clustered dataset for in-depth analysis.

	Suburban_Region	Population	Longitude	Latitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Baulkham Hills	155081	151.030535	-33.735906	1	Yoga Studio	Gym	Track	Pool	Pilates Studio
3	Blacktown	145065	150.966912	-33.760263	0	Gym	Gym / Fitness Center	Yoga Studio	Pool	Track
4	Blacktown - North	113109	150.931074	-33.731523	4	Gym	Pool	Martial Arts Dojo	Yoga Studio	Track
5	Mount Druitt	119106	150.847460	-33.742006	0	Gym	Gym Pool	Gym / Fitness Center	Yoga Studio	Track
6	Botany	54638	151.197644	-33.947087	1	Gym	Yoga Studio	Track	Pool	Pilates Studio

## c. Analytical explorations.

Each of the top 5 most common venues will then be explored to see the ratio of the fitness centers that are within the 1<sup>st</sup> most common venue, 2<sup>nd</sup> most common venues etc. The clusters will then be visualized on the map of Sydney to view how the clusters are placed accordingly. Each cluster will then be explored and discussed with which fitness center venues are the most common within each cluster.

## d. Exploring the Top 10 Populations of Sydney's Suburban Regions

Next the top 10 populations of Sydney's suburban regions will then be explored to see which are the most common venues within these regions. This analysis may provide useful information with which populous regions of Sydney may be much more optimal for opening a fitness center. As seen the 1<sup>st</sup>

# 4. Results

## a. Breakdown of most common fitness center venues

As seen in figure 1 below, within the 1<sup>st</sup> most common venues, the highest occurring are gyms/fitness centers with Both gyms/fitness centers combined having over 50% of the ratio with as gyms usually have a lot more a multi-purpose for use. It also appears that within the 2<sup>nd</sup> most common venues, Gyms are still quite popular however yoga studios may appear to the next most occurring. In the 3<sup>rd</sup> most common venue, it seems that yoga studios, track and martial arts dojos are some next common ones. However, in the 4<sup>th</sup> most common venue, it appears that pool fitness centers are the next to be one of the common fitness centers followed by track and yoga studios. In the 5<sup>th</sup> most common venue, pool and track still seem to be quite common but Pilates studio is now bringing up in to the competition.

Overall, these graphs show that that primarily gym/fitness centers are one of the most common fitness center venues throughout each of the regions followed by Tracks, Yoga Studios and Martial Studios. Although not as common, the next common ones will be pools and pilates studios.

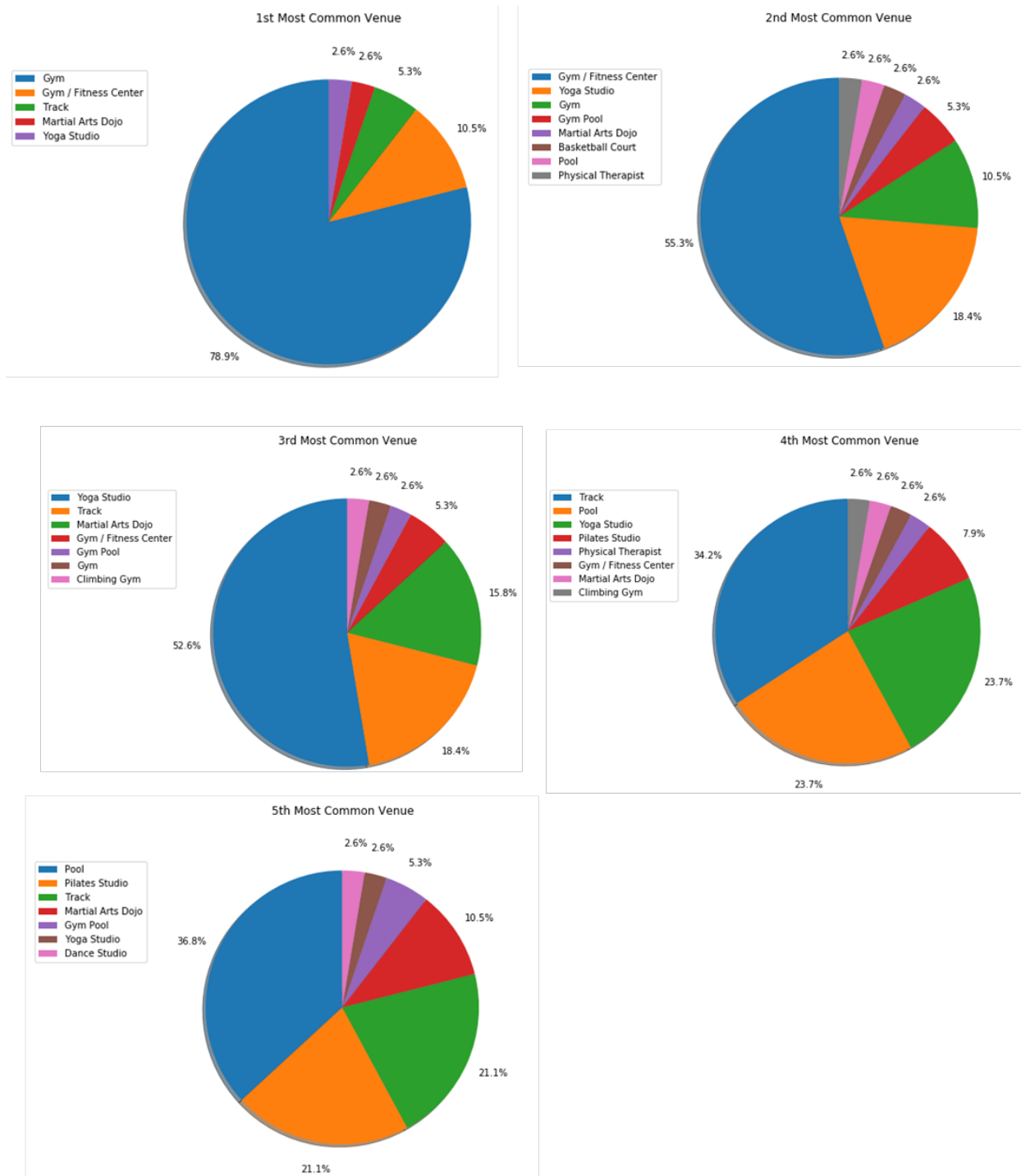


FIGURE 1: BREAKDOWN OF RATION WITHIN MOST COMMON VENUES

## b. Cluster overview and analysis

In figure 2, a map of Sydney's suburban areas is displayed along with which cluster they are based on their color. The results show that in cluster 1 colored as purple, primarily these regions will have Gyms/Fitness centers and Yoga studios followed by pools, track and dojos. Cluster 2 analysis shows that primarily Gyms, Yoga Studios and Track fitness centers are among the most common fitness center venue within these locations. Cluster 3 now shows that within these locations the primary common venues are Track, Martial Arts Dojo and Yoga Studios and secondary are pools and Pilates Studio. Cluster 4 this time shows gym/fitness centers and yoga studios as most common venues followed by track, pools and pilates studios. Cluster 5 this time shows Gyms, yoga studios, pools and tracks as the primary common venues with physical therapists, martial arts studios and basketball courts as the secondary venues.

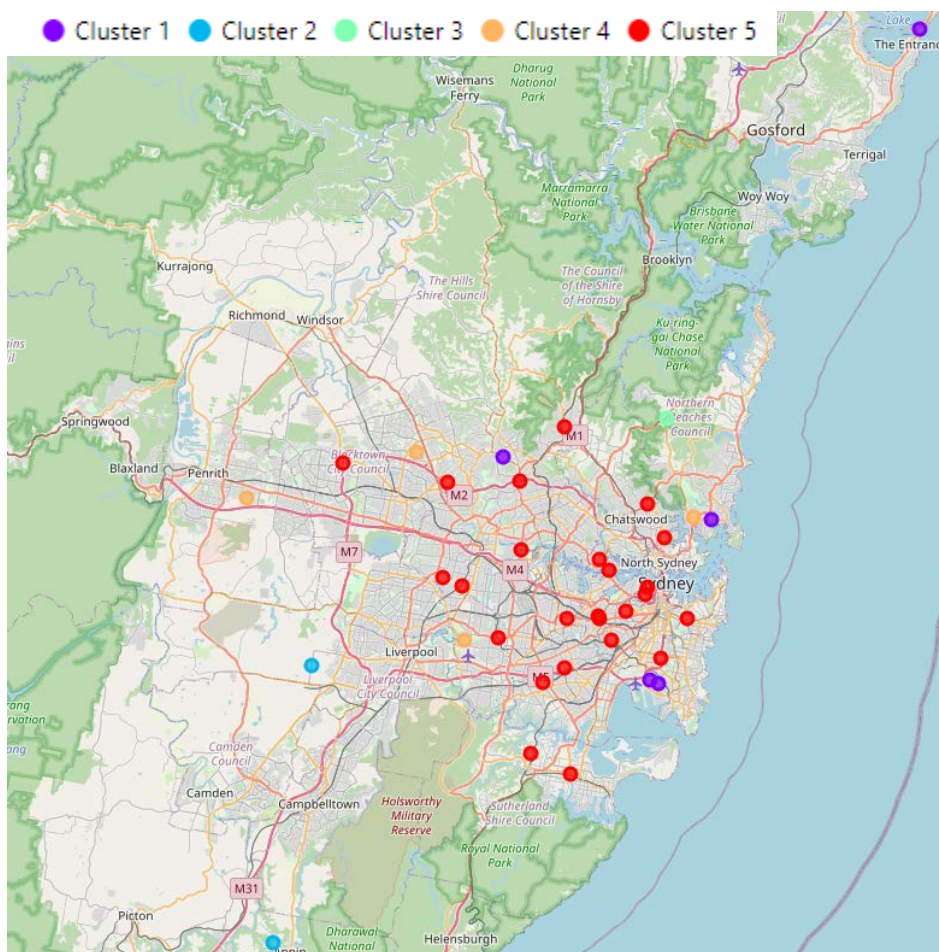
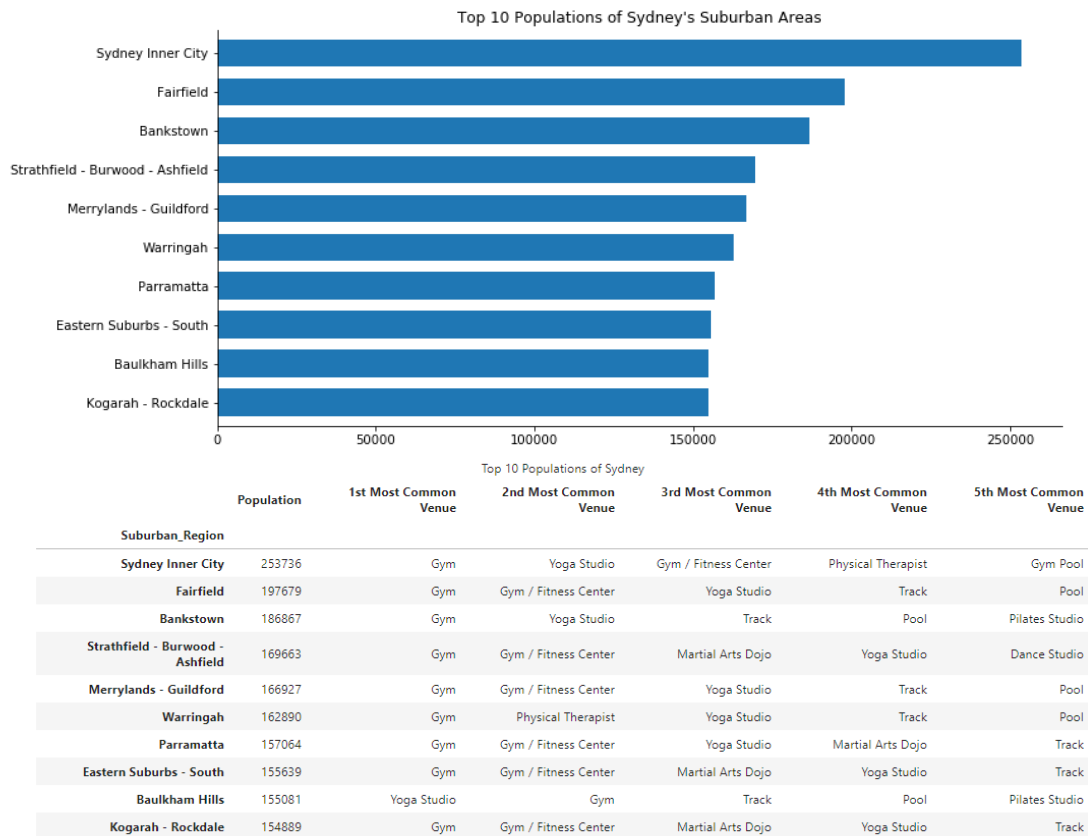


FIGURE 2: SYDNEY CLUSTER MAP OF MOST COMMON FITNESS CENTERS

### c. Top 10 populations of Sydney's Suburban Regions

In figure 3, the top 10 populations of Sydney are displayed along with their top 5 common venues within those regions. As shown in the analysis, Gyms/Fitness centers are still one of the most popular, followed by yoga studios. Tracks, martial arts dojo and pools also seem to be next as the next most common.



**FIGURE 3: TOP 10 POPULATIONS OF SYDNEY'S SUBRBAN REGIONS**



## 5. Discussion

Based on the results seen, some of the most common fitness centers to exist are primarily gyms/fitness centers followed by yoga studios, martial arts dojos and tracks. Opening up a fitness center revolving around one of these may have some competition but at the same time these prove to be the most popular type of fitness center venues to open. To increase highly increase the chances of customer attraction, it is recommended to open within one of Sydney's top 10 populous suburban regions such as Sydney City, Fairfield or Bankstown. If competition is not sought after, then opening a fitness center other than what's been displayed in the results will be optimal, however it may also prove to not be as popular compared to these venues.

## 6. Conclusion

In conclusion, as people progress to strive for physical development, gyms prove to be one of the most common and versatile fitness centers to open as shown as being the most common and popular venue in the results. Yoga studios and martial arts dojos also provide people with a great way to relieve themselves from the stresses of life and learn practical skills such as self-defense. Tracks also prove to be the next most common and popular for training in athletic events as well as marathons. Overall opening up one of these fitness center venues especially in Sydney's top 10 populous regions may ensure that there will be great customer attraction but also with competition.

## 7. References

- <https://www.abs.gov.au/>
- [https://www.matthewproctor.com/australian\\_postcodes](https://www.matthewproctor.com/australian_postcodes)