Exploratory Analysis of Sydney's Optimal Gym Location Openings

Capstone Project – Coursera IBM Data Science

By Kristobal Esgana

Contents

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 gyms 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 gyms have widely opened across Sydney and in many if its individual suburbs.

b. Problem

Opening up a certain type of gym such as a Boxing Gym, Climbing Gym or just a general workout Gym 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 gyms 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 gyms available within Sydney and which are the most thriving. This may prove to be useful in identifying which locations and types of gym may be optimal to open based on their ratings within that area. The analysis and findings within this report may provide insight to those interested in opening a certain type of gym 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 gym'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 form 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 the all the gyms within Sydney and their associated ratings 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.

3. Methodology

4. Results

5. Discussion

6. Conclusion

7. References

- https://www.abs.gov.au/
- https://www.matthewproctor.com/australian_postcodes
- https://simplemaps.com/data/au-cities