

# Data

[Yoga Australia](#) confirms our earlier assertion that yoga practice is dominated by women. It is also claimed that the majority of people who participate in yoga either hold tertiary educational qualifications or are currently studying at university. Because of this, data on demographics are required in order to confirm - or debunk - the link between educational attainment and participation in yoga and pilates.

In addition, it is well known that yoga and pilates (and, indeed, any exercise outside of the home) are discretionary expenditures. People living on low incomes such as the unemployed, or those on fixed pensions, may not be in a position to act upon an interest in yoga or pilates; these activities may be limited to people on higher incomes. Therefore, data are required to determine the link, if one exists, between income, employment status and participation in yoga and pilates.

Given the above considerations, the following data have been collected.

## **Shapefile of local government areas (LGAs) in Victoria**

Source: [data.gov.au](https://data.gov.au)

This has been obtained from the Australian Government's publicly available data collection, and allows the creation of maps of Victoria that are segmented by LGA. There are also shapefiles available that allow segmentation by suburb, but segmentation by LGA is preferred. There are many hundreds of suburbs in Victoria, and these are grouped into 79 LGAs.

Segmentation by LGA is preferred because some suburbs, particularly in Melbourne, are very small and are almost entirely composed of residential properties. Such suburbs would have very few (perhaps zero) yoga and pilates studios, and people living there would likely travel to another suburb in order to access yoga or pilates. Therefore, if data pertaining to studios are segmented by suburb, then the results may incorrectly indicate that there is no market for yoga or pilates in wholly residential suburbs. Using LGAs rather than suburbs captures those yoga and pilates participants who travel to adjacent suburbs within their own LGA.

## **General community profile (GCP) by LGA**

Source: [ABS census datapacks](#)

The Australian Bureau of Statistics (ABS) has produced a number of 'datapacks' containing statistics that were collected during the last Australian census. The GCP datapack is the one that is most relevant to the problem at hand.

The GCP comprises 59 tables and 15522 columns of data about each LGA. Most of these columns are not required for the current analysis, and have been excluded. A small number

of columns has been selected, because they pertain to sex, educational attainment, income and employment status - these are the demographic characteristics identified earlier as possibly having a deterministic effect on participation in yoga and pilates.

Where possible, statistics specifically pertaining to women are preferred over statistics pertaining to both sexes collectively; this is because the market for yoga and pilates is mostly comprised of women, so the exclusion of data about men yields a dataset that is more closely aligned with the principal participants in yoga and pilates.

The exception to this is the median family weekly income, which incorporates the incomes of both men and women. This has been included to account for the women who are stay-at-home parents living with a partner, as these women would not be included in statistics related to work and income.

The statistics chosen from the 15222 columns in the GCP are shown in the table below.

Column Heading	Meaning	Notes
Tot_P_F	Total number of women in LGA	This is required for the calculation of percentages
Median_tot_fam_inc_weekly	Median total family income per week	
F_650_799_Tot F_800_999_Tot F_1000_1249_Tot F_1250_1499_Tot F_1500_1749_Tot F_1750_1999_Tot F_2000_2999_Tot F_3000_more_Tot	Number of women earning \$650 to \$799 per week / \$800 to \$999 per week / \$1000 to \$1249 per week / \$1250 to \$1499 per week / \$1500 to \$1749 per week / \$1750 to \$1999 per week / \$2000 to \$2999 per week / \$3000 or more per week	These are to be summed and expressed as a percentage of the number of women living in the LGA
Percnt_Employment_to_populn_F	Percentage of women in the LGA who are employed	
F_PGrad_Deg_Total F_GradDip_and_GradCert_Total F_BachDeg_Total F_AdvDip_and_Dip_Total	Number of women with a postgraduate degree / graduate certificate or diploma / bachelor's degree / advanced diploma or diploma	These are to be summed and expressed as a percentage of the number of women living in the LGA

These columns have been extracted from their respective tables and merged into a single table containing 16 columns. A snapshot of this table is shown below.

A	B	C	D	E	F	G
LGA_CODE_2016	Tot_P_F	Median_tot_fam_inc_weekly	F_650_799_Tot	F_800_999_Tot	F_1000_1249_Tot	F_1250_1499_Tot
LGA20110	6265	1322	489	404	294	
LGA20260	5449	1263	438	361	252	
LGA20570	52821	1489	3842	3329	2953	
LGA20660	62644	2033	3942	4152	4290	
LGA20740	16912	1192	1257	934	698	
LGA20830	24938	1462	1818	1526	1192	
LGA20910	50862	2765	2718	3049	3162	
LGA21010	7149	1232	516	438	343	
LGA21110	87085	2652	4630	5219	5859	
LGA21180	97079	1358	6244	5719	4830	
LGA21270	3066	1110	186	148	132	
LGA21370	18765	1349	1497	1192	954	
LGA21450	47773	1700	3575	3303	2799	
LGA21510	150374	1655	10760	10400	8701	

Each row in the table pertains to one LGA, each of which has a unique code (as seen in the first column in the snapshot above).

### Yoga/pilates studios by LGA

Source: [Foursquare Places API](#)

The number of yoga and pilates studios in each LGA has been obtained through the Foursquare API. The 'search' endpoint requires the provision of a location expressed as longitude and latitude; these coordinates are obtained using a geocoder, then used in a call to the API.

The data from Foursquare are used as an additional feature in the table of GCP data. This data is then cleaned and passed into a machine learning model in order to determine which demographics indicate a strong market for yoga and pilates providers.