Table of Contents

Introduction ***1***

Problem and a discussion of the background **1**

Data and how it is used to solve the problem ***2***

Data acquisition and cleaning ***2***

Methodology ***2***

Results ***4***

Examining Clusters **4**

Best Suburbs to open a new Gym **4**

Map of best Suburbs for new Gym **5**

Discussion ***5***

Conclusion and future improvements ***5***

Introduction

Problem and a discussion of the background

This project is to help fitness professionals in Melbourne to find the best suburbs for opening a new Gym

Opening your own Gym has a lot of perks – you no longer have to work in someone else's gym, which allows you to be in control of your schedule, design everything the way you want it to be, and be your own boss.

Most fitness professionals have a dream to open their own facility.

You need to put your facility in a suburb where you can acquire the best clients to keep it busy. Because without clients, there's no point in having a facility at all!

The right suburbs are where people love to spend their time. Important common venues are:

- Cafes
- Restaurants
- Grocery Shops
- Pubs and bars

You also would like to see a low number of already existing gyms.

With the metropolitan area of Melbourne having more than 400 suburbs choosing a suburb is not easy.

Opening a new facility is a big investment, which means it's up to the aspiring fitness professional to research and use data – not just emotions – to make decisions.

Data and how it is used to solve the problem

- Webpage with all Suburbs and their respective postcodes. https://www.citypostcodes.com.au/Melbourne
- Geolocator is used to get the latitude and longitude values for the list of suburbs. The Foursquare API is used to get the top 100 venues per suburb.
- Download and prepare the Suburb dataset.
- Clean the data
    - Use Geolocator to get the latitude and longitude values for each suburb.
    - Create a map of Melbourne with suburbs superimposed on top
- Explore the first Suburb
    - Get the top 5 venues
    - Explore the result
- Explore all Suburbs in Melbourne
    - Work out the number of venues per suburb
    - Analysze each suburb
    - Group rows by suburb and by taking the mean of the frequency of occurrence of each category
    - Display the top 5 venues for each suburb
- Cluster Suburbs
    - Create a new dataframe that includes the cluster as well as the top 10 venues for each suburb
    - Visualize the resulting clusters
- Examine Clusters
    - The best cluster for opening a new Gym has a high number of: Cafes, Restaurants, Grocery Shops, Pubs and Bars.
    - Additional it has a low number of existing Gyms.
    - Visualize the Suburbs within the best cluster to open a new Gym on a map.

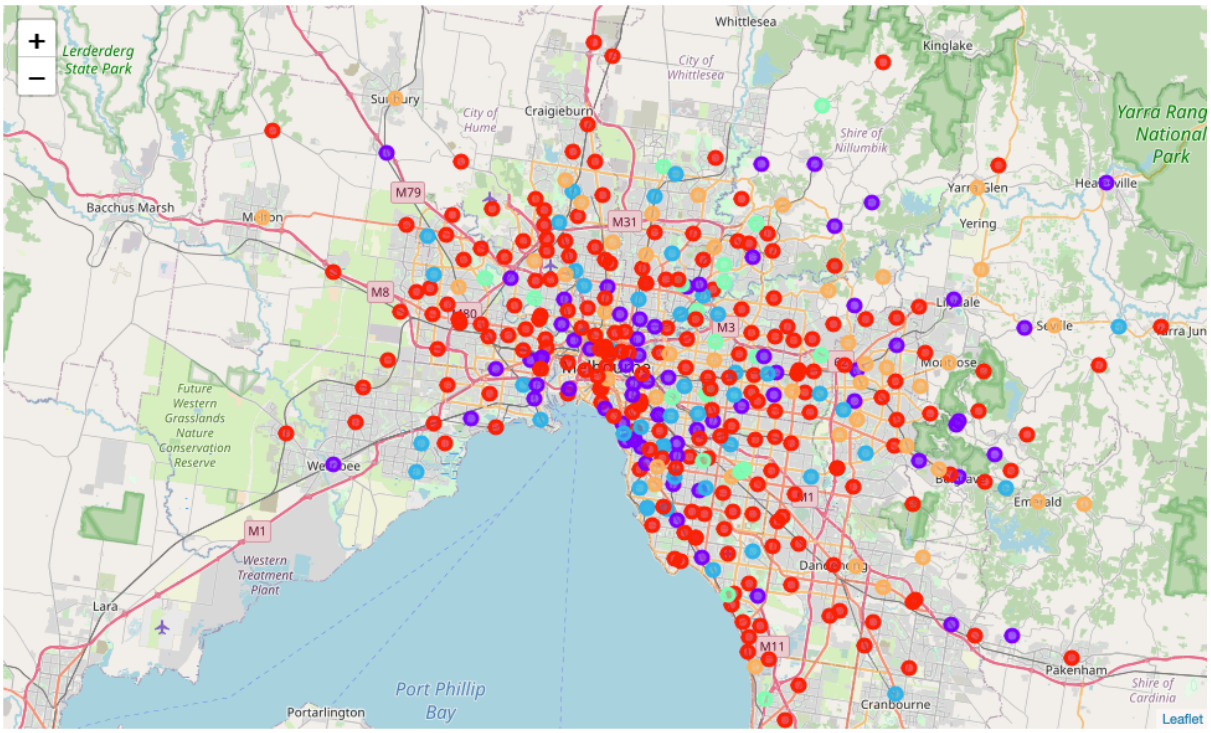
Data acquisition and cleaning

- Data scraped from https://www.citypostcodes.com.au/Melbourne
- In total 486 rows and 3 features in the raw dataset
- PO Boxes, City Column were dropped
- Each row is enriched by Latitude, Longitude obtained from Geolocator
- The cleaned dataset contains 477 rows with 4 features

Methodology

- Foursquare API is used to obtain nearby venues for each suburb
- K-means clustering method is used with k=5 to cluster the suburbs
- Cluster are manually analysed reg Cafes, Restaurants, Grocery Shops, Pubs and bars and already existing gyms.

Map of clustered Suburbs


Whittlesea
Kinglake
Lerderderg State Park
City of Whittlesea
Craigieburn
City of Hume
Shire of Nillumbik
Yarra Ranges National Park
Bacchus Marsh
M79
M31
Yarra Glen
Yering
M8
M3
Future Western Grasslands Nature Conservation Reserve
M1
Western Treatment Plant
Lara
M11
Port Phillip Bay
Portarlington
Cranbourne
Pakenham
Shire of Cardinia
Leaflet

Results

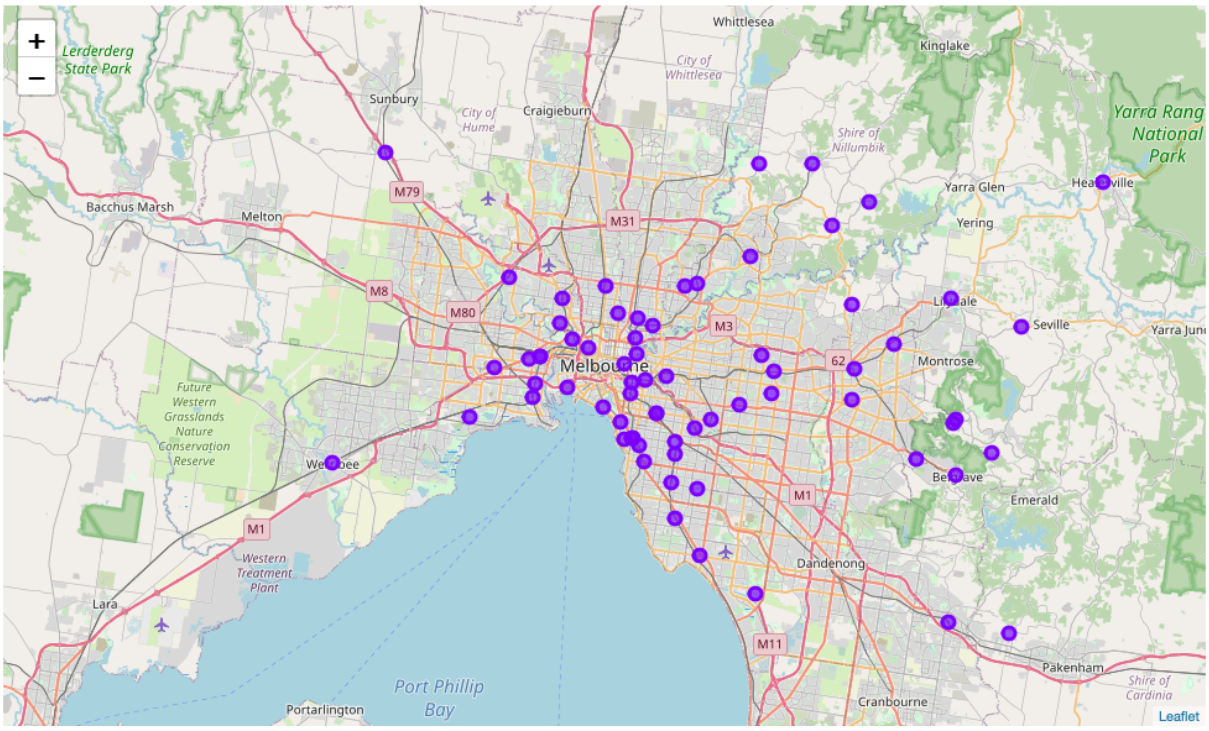
Examining Clusters

- Cluster 1: has the suburbs with the most Cafes, Restaurants, Grocery Shops, Pubs and Bars. It also has a low number of existing Gyms. The suburbs of this cluster are best suited to open a new Gym.**
- Cluster 0: most common venues are supermarkets followed by stortes and sports clubs. There is a lower number of coffee shops.
- Cluster 2: has a high number of restaurants, pubs and parks.
- Cluster 3: most common venues are parks
- Cluster 4: most common venues are restaurants followed by stores and a lower number of coffee shops
-

Best Suburbs to open a new Gym

Abbotsford, Altona, Armadale, Armadale North, Ascot Vale, Ashburton, Bedford Road, Belgrave, Bennettswood, Bentleigh, Bentleigh East, Berwick, Blackburn, Blackburn South, Brighton Road, Brooklyn, Brunswick East, Burnley, Caulfield East, Clifton Hill, Coburg, Cremorne, Croydon, Darling, Diggers Rest, East Melbourne, Elsternwick, Elwood, Essendon, Fairfield, Flinders, Gardenvale, Glen Huntly, Hawthorn, Healesville, Heidelberg Heights, Highett, Hurstbridge, Kangaroo Ground, Karingal, Keilor East, Kensington, Kerrimuir, Kingsville, Knox City Centre, Lilydale, Mentone, Middle Park, Montmorency, Newport, North Melbourne, Northcote, Officer, Port Melbourne, Ripponlea, Rosanna, Sassafras, Sassafras Gully, Seddon, Seddon West, South Yarra, Spotswood, St Kilda, The Patch, Upper Ferntree Gully, Wandin North, Warrandyte South, Waterways, Watsons Creek, Werribee, Yarrambat

Map of best Suburbs for new Gym



Discussion

The APIs used for the submission (Geolocator, Foursquare) are both prone to errors and exceptions. A caching mechanism preventing multiple calls for the same dataset would be nice to have.
Reg the k-means algorithm there is a significant effort to evaluate the results for large datasets. A clear design for result evaluation is key for estimating project effort.

Conclusion and future improvements

- Build automated evaluation function
- Optimize number of clusters
- Rank suburbs in best cluster
- Capture more data, like number of residents per suburb and age structure