Data Science Capstone Project

Most Liveable Places in Melbourne

by Christian Themin

7 July 2020



[1]

Abstract

The purpose of this project is to implement Data Science methodology and Machine Learning algorithms on the geospatial data of Victoria map to identify Melbourne busiest areas and what businesses are popular in the neighbourhoods. The result can be used to determine what business can have a potential impact to grow in the areas.

1. Introduction

1.1 Background

Melbourne has become one of the fastest-growing cities in the developed world, racing away from the rest of Australia and doubling the rate of growth of most cities in advanced economies.

At 2.7 per cent in 2017, the city is growing at a rate usually associated with rapidly-expanding cities in China and South America - putting pressure on infrastructure and creating a political minefield for the Victorian and federal governments.

The city's surge past 5 million people has fuelled productivity-sapping transport and road bottlenecks in the CBD, frustrating commuters and hopeful homeowners struggling to get into the market.[2]

1.2 Problem

Keeping the city alive is important and it is considered as major role for every business owner to attract their local community and maintain their shops busy. The more people commute, the busier and liveable the city will become.

Cafes/Coffee shops, Restaurants, Shopping Centres, and Retailers are the major business for a city. This project will investigate the city of Melbourne's area to find the top 10 Most common venues visited by the local community.

---

[1] Image source : www.pixabay.com

[2] https://www.theage.com.au/national/victoria/melbourne-is-one-of-the-fastest-growing-cities-in-the-developed-world-20180920-p504zn.html#:~:text=New%20figures%20released%20by%20the,1.8%20per%20cent%20for%20Sydney.&text=ABS%20data%20shows%20its%20five,just%20above%20Melbourne's%20current%20rate.

2. Data Collection and Preparation

As Melbourne is part of Victoria states, the geospatial data used in this report is Victoria Map. The data was collected from the Australia GeoNames website and it is an open source data. It contains 7 variables such as Index number, Place with latitude/longitude, Postcode, Country, Admin 1, Admin 2, and Admin 3. The following is the example of the raw data from the website:

GeoNames Home | Postal Codes | Download / Webservice | About search login

Victoria - postal codes

Australia Victoria search

Either enter a postal code (eg. "9011", "AB1", "9980-999") or a city (eg. "London")

	Place	Code	Country	Admin1	Admin2	Admin3
1	Melbourne -37.813/144.961	3000	Australia	Victoria	MELBOURNE CITY	
2	East Melbourne -37.813/144.984	3002	Australia	Victoria	MELBOURNE CITY	
3	West Melbourne -37.809/144.947	3003	Australia	Victoria	MELBOURNE CITY	
4	Melbourne -37.842/144.976	3004	Australia	Victoria	MELBOURNE CITY	
5	Docklands -37.818/144.944	3008	Australia	Victoria	MELBOURNE CITY	
6	Footscray -37.798/144.895	3011	Australia	Victoria	FOOTSCRAY	
7	West Footscray -37.799/144.871	3012	Australia	Victoria	FOOTSCRAY	
8	Yarraville -37.817/144.884	3013	Australia	Victoria	FOOTSCRAY	
9	Newport -37.843/144.878	3015	Australia	Victoria	FOOTSCRAY	
10	Williamstown -37.86/144.892	3016	Australia	Victoria	FOOTSCRAY	

The data contains 200 rows and is required for data correction and cleansing.

The processing method is done using Jupyterlab. The data is retrieved from the url website into pandas dataframe and is shown as below:

	Unnamed: 0	Place	Code	Country	Admin1	Admin2	Admin3
0	1.0	Melbourne	3000	Australia	Victoria	MELBOURNE CITY	NaN
1	NaN	-37.813/144.961	-37.813/144.961	-37.813/144.961	-37.813/144.961	-37.813/144.961	-37.813/144.961
2	2.0	East Melbourne	3002	Australia	Victoria	MELBOURNE CITY	NaN
3	NaN	-37.813/144.984	-37.813/144.984	-37.813/144.984	-37.813/144.984	-37.813/144.984	-37.813/144.984
4	3.0	West Melbourne	3003	Australia	Victoria	MELBOURNE CITY	NaN

As can be seen from the table, the attribute “Place” has different values in each row. The first line is the city name and the second line is combined of latitude and longitude value. It needs to be separated into its own column: Place, Latitude, and Longitude.

The first step performed in data preparation was to split the rows into odd and even numbers; the odd row numbers have the value of latitude/longitude while the even row numbers contain the city name. The other unnecessary columns such as Unnamed:0, Country, Admin 1 to 3 are being dropped and the corrected and cleansed data is shown as below:

	Latitude	Longitude	Suburb	Postcode
0	-37.813	144.961	Melbourne	3000
1	-37.813	144.984	East Melbourne	3002
2	-37.809	144.947	West Melbourne	3003
3	-37.842	144.976	Melbourne	3004
4	-37.818	144.944	Docklands	3008
...	...	...	...	...
195	-37.783	144.937	Travancore	3032
196	-37.761	144.862	Avondale Heights	3034
197	-37.7	144.766	Sydenham	3037
198	-37.691	144.743	Hillside	3037
199	-37.715	144.753	Taylors Hill	3037

200 rows × 4 columns

The next step is to use Geopy library to read the latitude and longitude of Victoria, then convert it to a Map as below:

For this report, the focus is to check the activity of Melbourne's main suburbs, therefore the other suburbs that are out of Melbourne city will be excluded.

3. Data Exploration

To explore what activities that have been carried out in Melbourne area, Foursquare Places API will come in handy to obtain the recommended places at the time of execution. The Foursquare application offers real-time access to its global database of rich venue data and user content.

The following table shows the top 5 nearby venues in Melbourne that matches the latitude and longitude of the Foursquare Places API:

	Suburb	Suburb Latitude	Suburb Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Melbourne	-37.813	144.961	Shortstop Coffee & Donuts	-37.811240	144.961000	Donut Shop
1	Melbourne	-37.813	144.961	Brother Baba Budan	-37.813445	144.962137	Coffee Shop
2	Melbourne	-37.813	144.961	Tipo 00	-37.813527	144.961978	Italian Restaurant
3	Melbourne	-37.813	144.961	Brick Lane	-37.811341	144.959816	Café
4	Melbourne	-37.813	144.961	Kirk's Wine Bar	-37.813661	144.961351	Wine Bar

Below is the table that display the top 10 most common venues in Melbourne hotspot neighbourhoods:

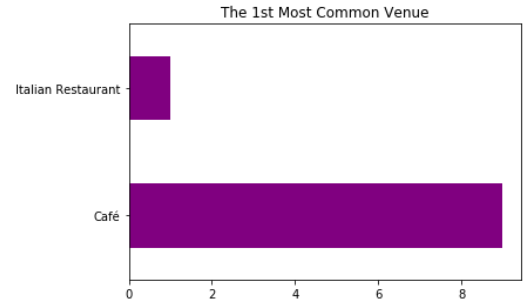
Suburb	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Carlton	Italian Restaurant	Ice Cream Shop	Deli / Bodega	Pizza Place	Café	Gourmet Shop	Dessert Shop	Egyptian Restaurant	Coffee Shop	Cheese Shop
Docklands	Coffee Shop	Café	Italian Restaurant	Vietnamese Restaurant	Bar	Fish & Chips Shop	Indian Restaurant	Middle Eastern Restaurant	Pizza Place	Pub
East Melbourne	Café	Hotel	Park	Australian Restaurant	Grocery Store	Convenience Store	Coffee Shop	Fish & Chips Shop	Garden	Sushi Restaurant
Kensington	Café	Park	Pizza Place	Japanese Restaurant	Fish & Chips Shop	Pub	Gym	Ice Cream Shop	Burger Joint	Bar
Melbourne	Café	Coffee Shop	Hotel	Dessert Shop	Korean Restaurant	Clothing Store	Indonesian Restaurant	Italian Restaurant	Argentinian Restaurant	Pizza Place
North Melbourne	Café	Light Rail Station	Pub	Park	Hotel	Grocery Store	Garden	Bar	Italian Restaurant	Thai Restaurant
Port Melbourne	Café	Bakery	Gym	Pub	Italian Restaurant	Fish & Chips Shop	Breakfast Spot	Vietnamese Restaurant	Pizza Place	Juice Bar
Richmond	Café	Gym	Japanese Restaurant	Dumpling Restaurant	Pub	Burmese Restaurant	Frozen Yogurt Shop	Sports Bar	Sandwich Place	Fast Food Restaurant
South Melbourne	Café	Breakfast Spot	Thai Restaurant	Mexican Restaurant	Fish & Chips Shop	Sandwich Place	Fried Chicken Joint	Bookstore	Salad Place	Furniture / Home Store
St Kilda	Café	Pub	Hotel	Thai Restaurant	Hostel	History Museum	Gym / Fitness Center	Japanese Restaurant	Light Rail Station	Fried Chicken Joint

The 1st most common venue in Melbourne is Café and then followed by various restaurants.

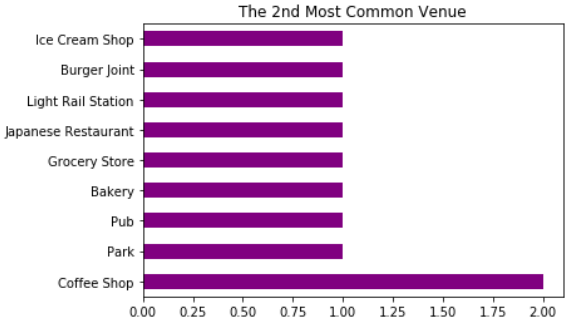
Using the Foursquare Places API, we can learn more about any specific venue or store or shop, like their full address, their working hours, and their menu if they have one. We can also explore a given location by finding what popular sports exist in the vicinity of the location.

From the result, we can also visualize the few top common venues into barcharts as below:

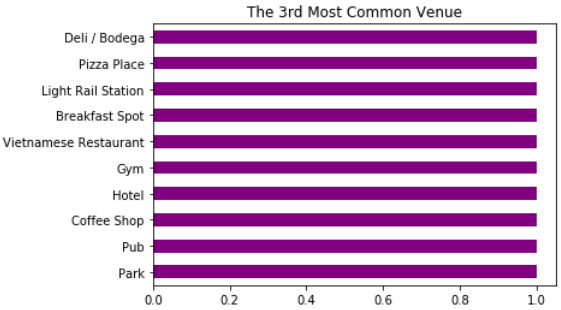
The 1st Most Common Venue:



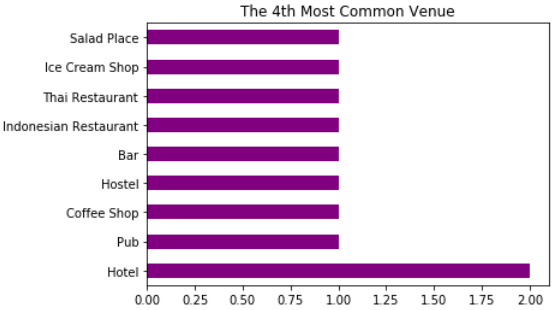
The 2nd Most Common Venue:



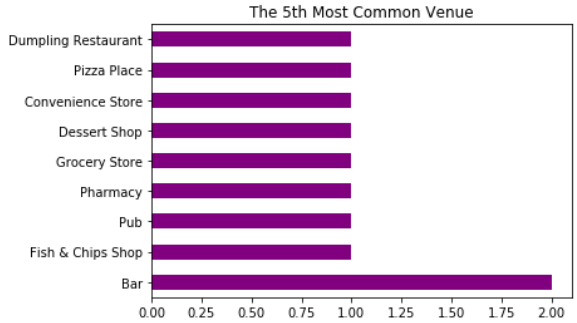
The 3rd Most Common Venue:



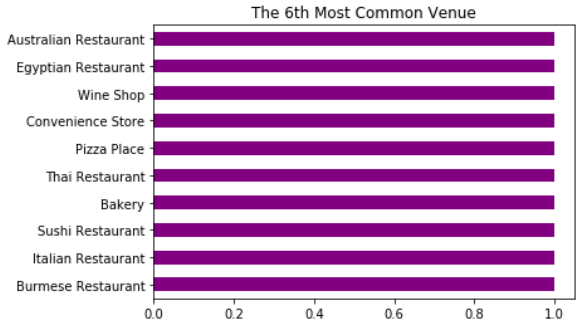
The 4th Most Common Venue:



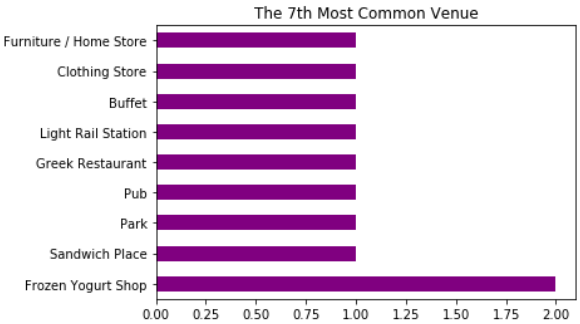
The 5th Most Common Venue:



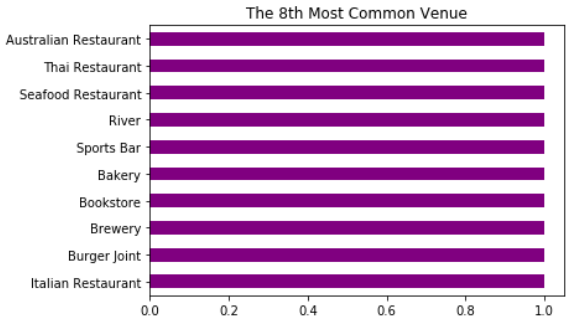
The 6th Most Common Venue:



The 7th Most Common Venue:



The 8th Most Common Venue:



4. Cluster Suburbs

In the clustering section, a Machine Learning technique is involved in grouping the data points. K-Means algorithm was used to compute the distances between suburbs and venue categories. The best number of clusters for the data set is 3 clusters.

After examined each cluster and determine the venue, the following is the results of the different clusters:

Cluster 1 turns to be 0 results.

Cluster 2 with 238 observations:

	Suburb	Venue Longitude	Venue Category	Cluster Labels	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	9th Most Common Venue	10th Most Common Venue
0	Melbourne	144.961000	Donut Shop	1	Café	Coffee Shop	Hotel	Dessert Shop	Clothing Store	Italian Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Bar	Scandinavian Restaurant
1	Melbourne	144.962137	Coffee Shop	1	Café	Coffee Shop	Hotel	Dessert Shop	Clothing Store	Italian Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Bar	Scandinavian Restaurant
2	Melbourne	144.961978	Italian Restaurant	1	Café	Coffee Shop	Hotel	Dessert Shop	Clothing Store	Italian Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Bar	Scandinavian Restaurant
3	Melbourne	144.959816	Café	1	Café	Coffee Shop	Hotel	Dessert Shop	Clothing Store	Italian Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Bar	Scandinavian Restaurant
4	Melbourne	144.963170	Cocktail Bar	1	Café	Coffee Shop	Hotel	Dessert Shop	Clothing Store	Italian Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Bar	Scandinavian Restaurant
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1426	Port Melbourne	144.939440	Grocery Store	1	Café	Bakery	Breakfast Spot	Fish & Chips Shop	Pub	Italian Restaurant	Thai Restaurant	Mexican Restaurant	Sandwich Place	Frozen Yogurt Shop
1427	Port Melbourne	144.939387	Breakfast Spot	1	Café	Bakery	Breakfast Spot	Fish & Chips Shop	Pub	Italian Restaurant	Thai Restaurant	Mexican Restaurant	Sandwich Place	Frozen Yogurt Shop
1428	Port Melbourne	144.936760	Café	1	Café	Bakery	Breakfast Spot	Fish & Chips Shop	Pub	Italian Restaurant	Thai Restaurant	Mexican Restaurant	Sandwich Place	Frozen Yogurt Shop
1429	Port Melbourne	144.940029	Café	1	Café	Bakery	Breakfast Spot	Fish & Chips Shop	Pub	Italian Restaurant	Thai Restaurant	Mexican Restaurant	Sandwich Place	Frozen Yogurt Shop
1430	Port Melbourne	144.941564	Supermarket	1	Café	Bakery	Breakfast Spot	Fish & Chips Shop	Pub	Italian Restaurant	Thai Restaurant	Mexican Restaurant	Sandwich Place	Frozen Yogurt Shop

238 rows × 14 columns

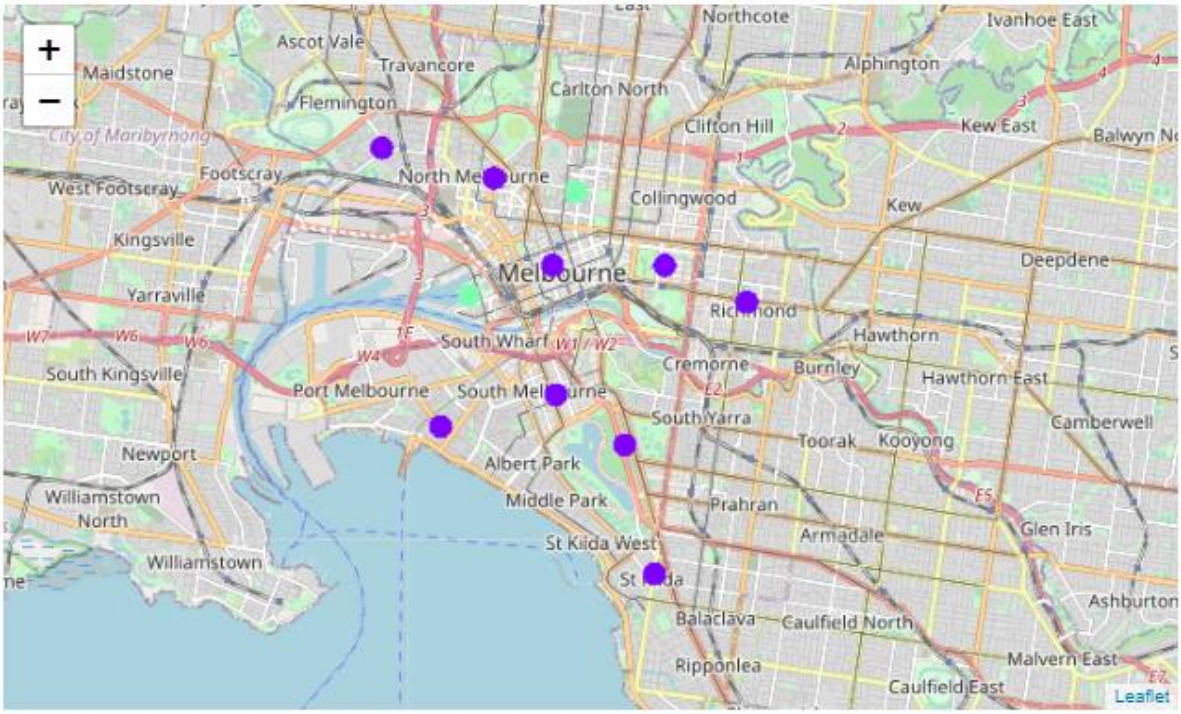
Cluster 3:

	Suburb	Venue Longitude	Venue Category	Cluster Labels	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	9th Most Common Venue	10th Most Common Venue
101	Docklands	144.944355	Coffee Shop	2	Coffee Shop	Café	Italian Restaurant	Vietnamese Restaurant	Bar	Fish & Chips Shop	Indian Restaurant	Middle Eastern Restaurant	Pizza Place	Pub
102	Docklands	144.945172	Vietnamese Restaurant	2	Coffee Shop	Café	Italian Restaurant	Vietnamese Restaurant	Bar	Fish & Chips Shop	Indian Restaurant	Middle Eastern Restaurant	Pizza Place	Pub
103	Docklands	144.943757	Middle Eastern Restaurant	2	Coffee Shop	Café	Italian Restaurant	Vietnamese Restaurant	Bar	Fish & Chips Shop	Indian Restaurant	Middle Eastern Restaurant	Pizza Place	Pub
104	Docklands	144.945275	Thai Restaurant	2	Coffee Shop	Café	Italian Restaurant	Vietnamese Restaurant	Bar	Fish & Chips Shop	Indian Restaurant	Middle Eastern Restaurant	Pizza Place	Pub
105	Docklands	144.944797	Fish & Chips Shop	2	Coffee Shop	Café	Italian Restaurant	Vietnamese Restaurant	Bar	Fish & Chips Shop	Indian Restaurant	Middle Eastern Restaurant	Pizza Place	Pub
106	Docklands	144.944072	Bakery	2	Coffee Shop	Café	Italian Restaurant	Vietnamese Restaurant	Bar	Fish & Chips Shop	Indian Restaurant	Middle Eastern Restaurant	Pizza Place	Pub

5. Results

In summary, Café has been the topmost business that runs the city.

However, depending on the location and the culture in the community, various restaurants can also be considered as succeeding in running the city; provided what services that is in demand. The selected venues can also be clustered as below:



6. Discussion

The limitation of the analysis includes only to find the business that operates in the city area. There are many aspects that we should investigate further in the location such as safety of the area, population growth, future plan, community gathering activities and so on. Each city has its own population story and is varied over time.

The same method of clustering can also be applied into different areas with different purposes of analysis.

Keeping the city alive is important for the country's economy. The more people commute & busier it is, the liveable the city will become.

7. Conclusion

As a result, with geospatial data combined with the community survey data, we can get an insight about what is trending in the area that can help the city to grow. People can achieve better outcomes through their interest in opening a business.

Investors may start considering which location has more prospect in success and helping the community grows.