Capstone Project – Indian Food in Melbourne's Suburbs

SHASHANK SHARMA CHITTIMALLA

03, July 2020

1. Introduction

Melbourne is the capital of Victoria state and it is the second most populous city in Australia and Oceania. Melbourne area has been a home for Wurundjeri – Aboriginal Australian for more than 40,000 years before the European settlements. Since its establishment in 1835, Melbourne has been a home to population from Europe, China, Indian sub-continent, and the list goes on to more than 200 countries. In fact, almost 50 per cent of Victorians were either born overseas or have at least one parent who was born overseas.

Melbourne has a vibrant Indian community with Indian-born migrants making up 3 per cent of Melbourne's total population. Deakin University's Prof. Harminder Singh analysed the tax dataset and finds, "the maximum number of taxpaying skilled migrants were born in India, which makes 20 per cent of the taxpaying population of Australia." He adds, "Indian-born migrants have paid $18 billion in tax, second only to those from the UK. Our income as well as tax contribution are going up. It's a proud statistic for our community."

Melbourne is a blend of bustling laneways, world-class restaurants, and captivating museums. The locals love their coffee and serve it up in abundance at cozy cafes. The diversity of our communities contributes to our food, shopping, business, entertainment and more. It is just one of the reasons we are one of the most livable cities in the world.

With successful, young Indian community and a diverse population from other cultures who are open to try and explore food, Melbourne is the right place to start an Indian restaurant. As the main target of an Indian Cuisine Restaurant is Indian community. Here I am going to explore all the suburbs of Melbourne so that an entrepreneur can take a decision on the location of the restaurant business.

2. Problem

With the data available from different sources we are going to answer the following questions –

1. List and visualize all the major parts of Melbourne City which have popular Indian Restaurants
2. Find the top 3 suburbs for starting an Indian Restaurant?
3. Which areas lack Indian Restaurants?
4. What is the best place to stay in Melbourne, if you want to stay close to Indian Food?
5. What is the best location to go if you wish to eat good Indian Food?

3. Data acquisition and cleaning

3.1 Data Sources

To solve the above problems, we will require following data –

- All the suburbs of Melbourne, Australia. The list is spread in following two pages which can be obtained by page scraping -
  Source –
  a)
  https://en.wikipedia.org/w/index.php?title=Category:Suburbs_of_Melbourne&pageuntil=Keilor%2C+Victoria%0AKeilor%2C+Victoria#mw-pages

  b)
  https://en.wikipedia.org/w/index.php?title=Category:Suburbs_of_Melbourne&pagefrom=Keilor%2C+Victoria%0AKeilor%2C+Victoria#mw-pages
  Description – As the list of suburbs is not readily available in any place, it must be scraped out of Wikipedia page. This information has to be refined to be used later on.

- Latitude and Longitude details of all the suburbs.
  Source - https://gist.github.com/randomecho/5020859
  Description - Latitude and Longitude details can be obtained using geocoder. But since it is unreliable, we can use the data from git to obtain it.

- Existing Indian Restaurant details – location, Ratings, Reviews etc. in and around Melbourne Area
  Source – FourSquares API
  Description – Using the above FourSquares APIs and passing the correct inputs, details like location, ratings, reviews can be obtained

3.2 Data Acquiring

The most of the available data sources for the data sources is online web portals, github pages. We do not have a specific data sources in csv files or tabular format available on web. Hence the required data like “list of suburbs”, “geographical location details – latitude, longitude” had been web scraped from the available sources.

For Web Scraping, I used a powerful and user friendly library – Scrapy. Scrapy is a fast high-level web crawling and web scraping framework, used to crawl websites and extract structured data from their pages. It can be used for a wide range of purposes, from data mining to monitoring and automated testing.

You can get more information from - https://docs.scrapy.org/en/latest/

3.3 Data Cleaning

Collecting the data in the raw format using web scraping comes with problems like – unformatted data, duplicate data, missing data, unwanted data etc. All the above issues with the acquired data are dealt with in the code.

Few of the important data cleansing done in the code are –

a) Every suburb name fetched from the Wikipedia had “Victoria” as a suffix, which is not required for our data analysis. So this has been handled using available python-panadas functions
b) Unwanted information like header – “List of Melbourne suburbs” from the Wikipedia page was extracted to the list. This has been deleted
c) Other operations like dropping null values, removing duplicate values, removing wrong information were performed on the web scrapped data

3.4 Feature selection

After performing all the cleaning operations on the web scrapped data. I was able to come to a final list with 379 suburbs in and around Melbourne along with their geolocation data. You can find the list in my code. The below is the head of the final table with 5 / 379 records.

	suburb	latitude	longitude
0	Abbotsford	-37.80	145
1	Aberfeldie	-37.76	144.9
2	Airport West	-37.71	144.89
3	Albanvale	-37.75	144.77
4	Albert Park	-37.84	144.96

FourSquares API was used to fetch the list of Indian Restaurants in different Suburb along with their user likes, which is used as the main data to analyze the Melbourne Suburbs Indian Restaurants. You can find the head of the table with the list of Indian Restaurants with 5/ 235 records in it.

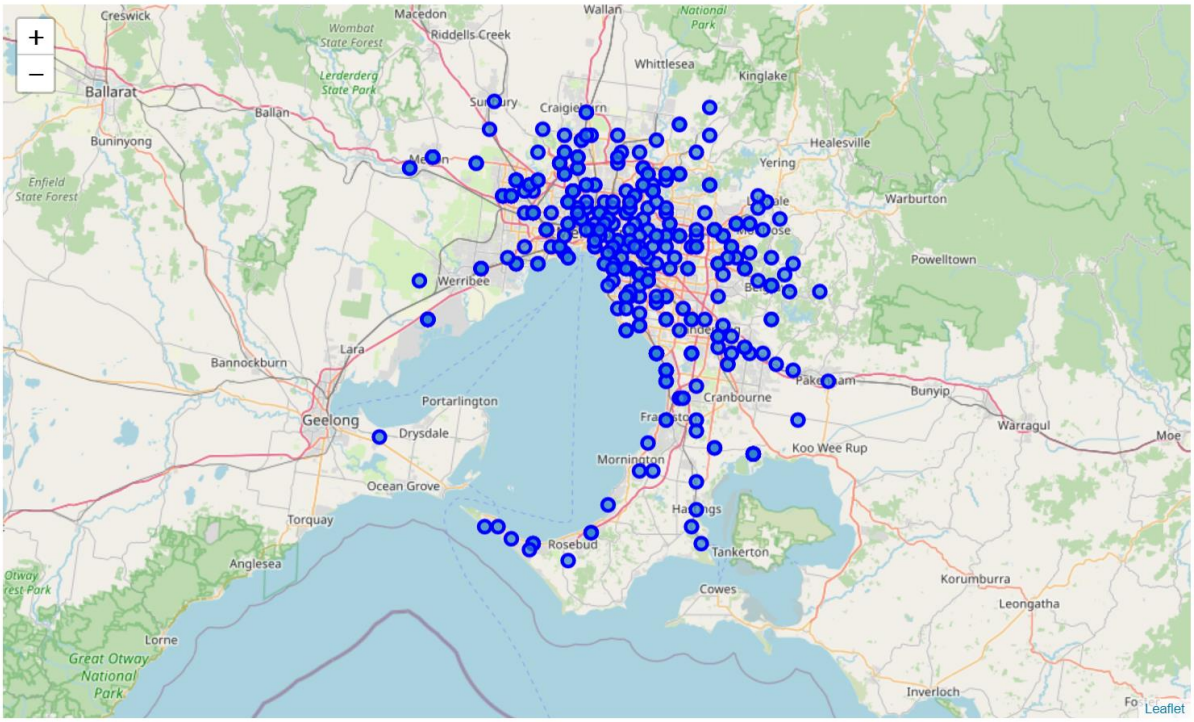
	Suburb	ID	Name	Lat	Lng
0	Abbotsford	5c3c5818bd4009002c5cf256	Dana Pani Indian Restaurant	-37.810279	144.998983
1	Albert Park	4c6369a969a1c9b6bd4b3ba4	indian murchi	-37.833277	144.960784
2	Albert Park	5e4bc8fe3a33460008031e81	AMAN INDIAN RESTAURANT	-37.839276	144.967427
3	Albert Park	565580a7498eec9f81a1a2a1	My Masala Indian Cuisine	-37.831126	144.954201
4	Albert Park	5655800b498eacd755c2045a	My Masala Indian Cuisine	-37.831110	144.953620

4. Exploratory Data Analysis

By performing exploratory data analysis on the acquired data, I was able to find come up with following maps and observations –

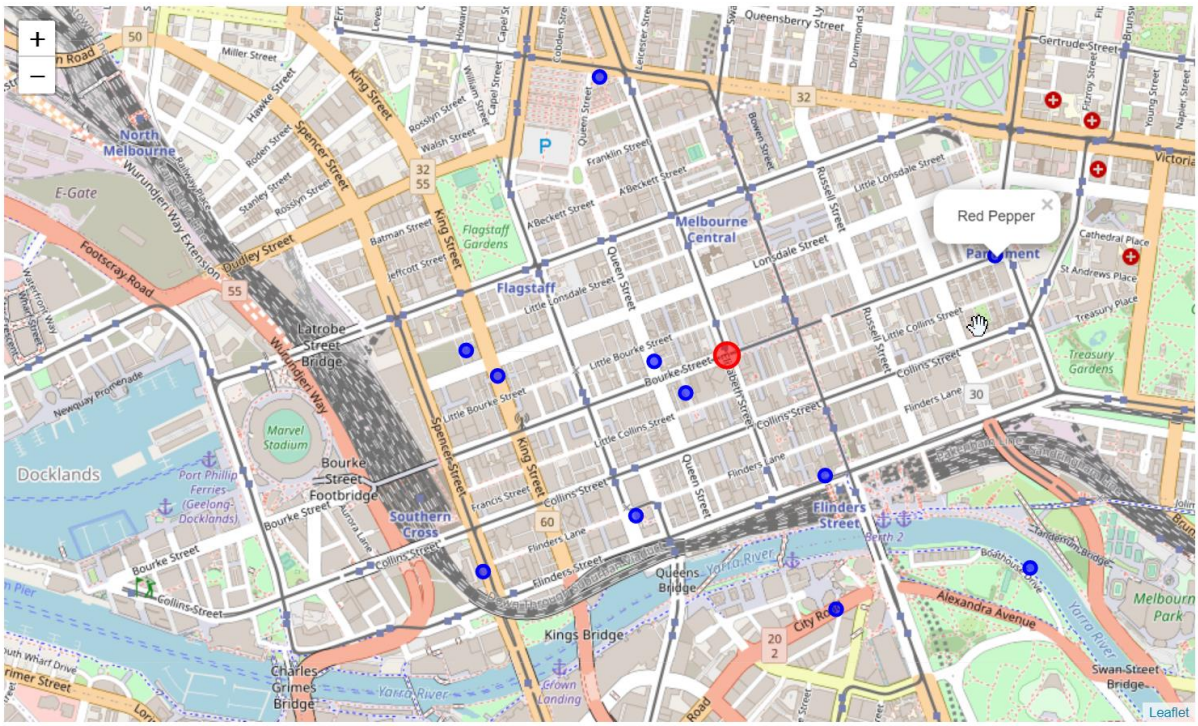
4.1 Mapping Melbourne Suburbs

This is the map of all the Melbourne suburbs. This has an interactive blue bubble on clicking each blue bubble, gives the name of the suburb. This ca even be zoomed in on the code.



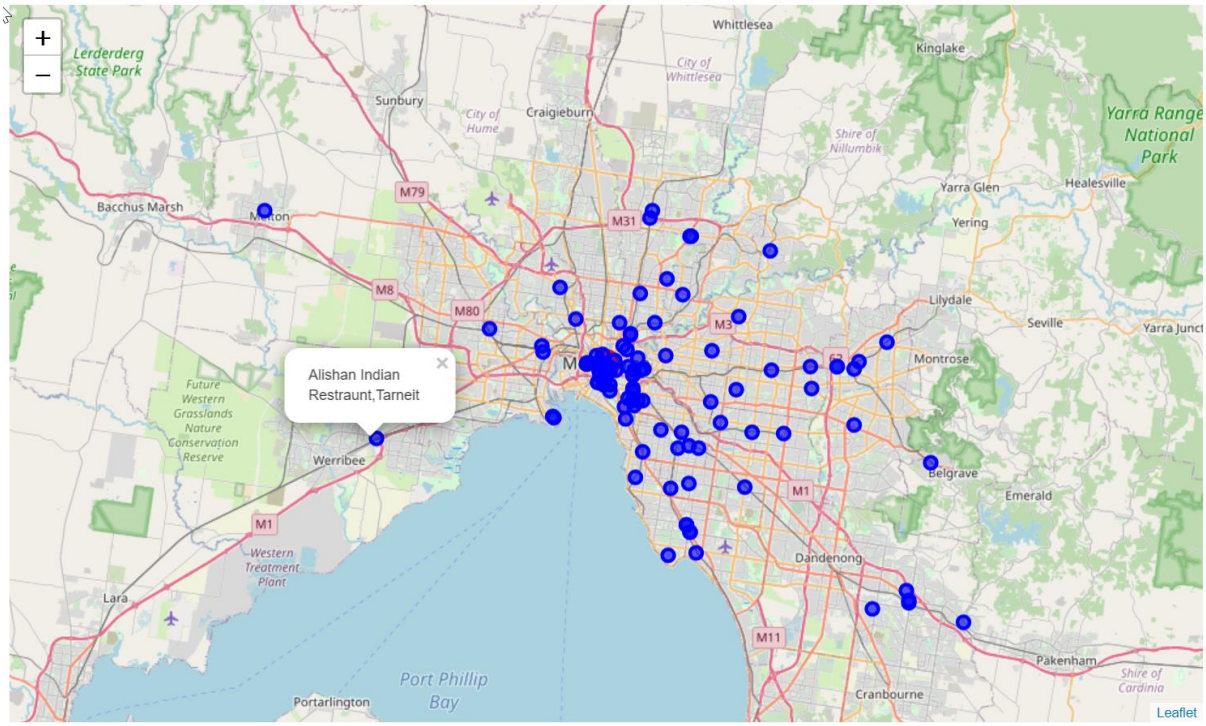
4.2 Mapping Indian Restaurants in Melbourne CBD

All the blue bubbles below show the Indian Restaurants in Melbourne CBD area. One can get the name of the restaurant by clicking the blue bubble.



4.3 Mapping all the Indian Restaurants in and around Melbourne Suburbs

On similar grounds all the Indian Restaurants in Melbourne Suburbs have been mapped. Clicking the blue bubble gives the pop-up with the information – Name of Restaurant, Suburb.



Along with rendering above maps I was able to create 3 clusters of where all the Indian Restaurants in Melbourne are added based on the “likes” / “user rating” on FourSquares platform.

5. K-Means Clustering

There are many models for clustering out there. In this notebook, we will be presenting the model that is considered one of the simplest models amongst them. Despite its simplicity, the K-means is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from unlabeled data. In this notebook, you will learn how to use k-Means for customer segmentation.

Some real-world applications of k-means:

- Customer segmentation
- Understand what the visitors of a website are trying to accomplish
- Pattern recognition
- Machine learning
- Data compression

Here I segmented the Indian restaurants into 3 clusters.

5.1 Cluster#1

This has the Indian restaurants with least number of likes from customers. Below is the head of the list with 5 / 92 restaurants.

	Suburb	ID	Name	Likes	Lat	Lng
0	Abbotsford	5c3c5818bd4009002c5cf256	Dana Pani Indian Restaurant	0	-37.810279	144.998983
1	Albert Park	4c6369a969a1c9b6bd4b3ba4	indian murchi	0	-37.833277	144.960784
2	Albert Park	5e4bc8fe3a33460008031e81	AMAN INDIAN RESTAURANT	0	-37.839276	144.967427
3	Albert Park	565580a7498eec9f81a1a2a1	My Masala Indian Cuisine	0	-37.831126	144.954201
4	Albert Park	5655800b498eacd755c2045a	My Masala Indian Cuisine	1	-37.831110	144.953620

5.2 Cluster#2

This cluster has the list of restaurants which are at medium range in terms of customer satisfaction.

This is the complete list of 9 Indian restaurants with other details.

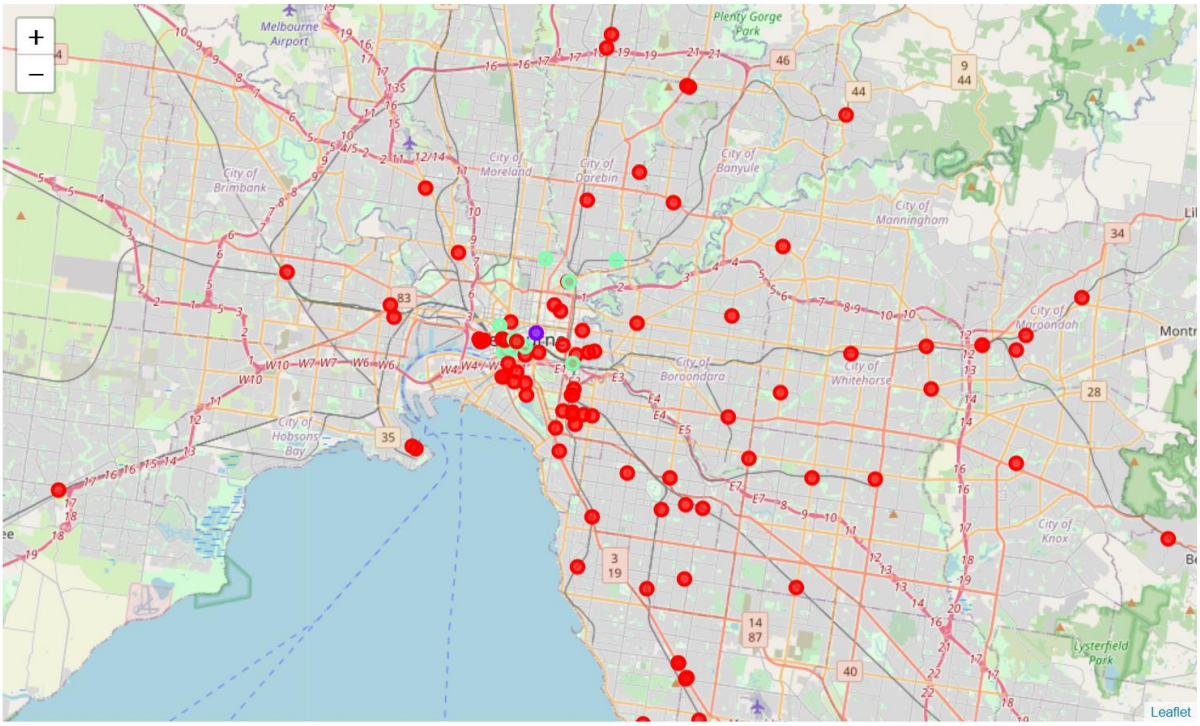
	Suburb	ID	Name	Likes	Lat	Lng
7	Alphington	4b46fcacf964a520342a26e3	Everest Indian Restaurant	5	-37.778806	145.018208
37	Carlton North	4b9961dcf964a520437735e3	Singhs Gourmet Indian Foods	8	-37.778410	144.978337
44	Clifton Hill	4bb708a146d4a593b60cc7c0	Marigold	5	-37.788550	144.991215
47	Port Melbourne	4b05874df964a5205e8a22e3	Gaylord Indian Restaurant	5	-37.820263	144.954499
70	Docklands	4b768359f964a520b84f2ee3	Shiraaz Fine Indian Cuisine	6	-37.818692	144.959940
71	Docklands	4c7a274c97028cfa1a71ddfe	Le Taj Fine Indian Food	7	-37.807956	144.952096
74	Docklands	4b05874ef964a520b28a22e3	Nirankar Indian Restaurant	9	-37.814383	144.960548
83	East Melbourne	4fa10efbe4b07aaa44e6c896	Punjabi Tandoori Corner	6	-37.825000	144.993290
89	South Wharf	4b307875f964a520c0f924e3	Flora Indian Restaurant	21	-37.817592	144.966626

5.3 Cluster#3

This has the Indian restaurant which is most liked by customers. There is only 1 restaurant in this cluster.

	Suburb	ID	Name	Likes	Lat	Lng	Clus_km
91	South Wharf	4b1cbc8ff964a520570924e3	Red Pepper	58	-37.81146	144.972623	1

The same information has been rendered on the map and showed below –



6. Results

I was able to fetch results for all the 5 business problems discussed –

1. List and visualize all the major parts of Melbourne City which have popular Indian Restaurants
2. Find the top 3 suburbs for starting an Indian Restaurant?
3. Which areas lack Indian Restaurants?
4. What is the best place to stay in Melbourne, if you want to stay close to Indian Food?
5. What is the best location to go if you wish to eat good Indian Food?

I can conclude that – **Docklands, Prahran, South Wharf, Albert Park and East Melbourne** are the areas where one can find maximum number of Indian restaurants among all the suburbs.

Few of the above results are discussed as part of observations earlier in the report.

7. Future directions

I was able to collect the list of Indian restaurants from FourSquares Regular API calls which are FREE to use. I came up with the conclusions / results based on one single input from the APIs – "likes". But there are many other fields which can provide better information in FourSquares like – ratings, popularity etc, which are premium calls. Please note that the information fetched through FourSquares is not the latest one. There are many other new Indian restaurants opened in many suburbs and many restaurants are closed, whose information is not available in FourSquares.

To fetch as many restaurants as possible, limit of 1000m is considered from the geolocation of the suburb, which might have fetched the restaurants from other suburb too. This should also be addressed in future by fetching more accurate geolocation of suburbs and even segregating the location of the restaurant using the fetched information rather than the information used to fetch it.

In future I will fetch the information from Google Maps Platform APIs and try to get the latest restaurant list along with latest ratings, users count etc to get the accurate results.

8. References

- Melbourne - Wikipedia
- Suburbs of Melbourne - Wikipedia
- Australian Postcodes - Github
- Forsquare API