Venue Recommendation for a new Indian Restaurant in Adelaide

Leena Javkar
12 December, 2019

1. Introduction

1.1 Background

Adelaide, the capital of South Australia, is a vibrant city with plenty of attractions and a rich cultural heritage. This city boasts a lot of multicultural restaurants and is particularly known for the world-famous wine regions only minutes away from the main city. It is also the largest city in South Australia with a history of international immigration. They come from many parts of the world.

The census data of 2016 revealed that the Indian population in Australia has grown to 455,389, up from 295,362 recorded in Census 2011. Of Australia's 24.4 million people, today Indians comprise of 1.9% of the total population. With its diverse culture, comes the diverse cuisine.

1.2 Problem

Here I am considering a business proposal to invest in or open an Indian restaurant in Adelaide. Hence, as part of this project we will be exploring the best locations for opening an Indian restaurant in Adelaide, Australia. Since there are lots of restaurants in Adelaide, I will try to look for locations with not many restaurants. We are also particularly interested in areas which are popular with people to ensure good influx of customers.

2. Data acquisition and cleaning

2.1 Data sources

Scraping Data from Wikipedia:

I will use the Adelaide City data from Wikipedia which includes all the Suburbs, Local Government Areas (LGAs) and Postal Codes.
Data Source: https://en.wikipedia.org/wiki/List_of_Adelaide_suburbs.

Getting Coordinates for the Postal Codes:

I will retrieve the geographical coordinates based on the postal codes of the Suburbs using Arcgis geocoder.

Venue Details using Four Square API:

The Four Square API (www.foursquare.com) will be used to explore suburbs in City of Adelaide. The Foursquare search function will be used to get Indian restaurants in each suburb and the suburbs will also be grouped into clusters based on the popular nearby venues. The following information is retrieved:

- Venue ID
- Venue Name
- Venue Coordinates i.e. Latitude and Longitude
- Venue Category

2.2 Data Cleanup and Re-grouping:

I first parse the Wikipedia page using the BeautifulSoup python library and transform the data from the table into a pandas dataframe. We can see in the dataframe that there is more than one Suburb for some postal codes. Such rows are combined into one row with the suburbs separated with a comma.

2.3 Retrieve coordinates:

I will be using arcgis geocoder to retrieve the coordinates (latitude and longitude) of each LGA. These coordinates will be used in retrieval of Foursquare API location data. The coordinates are then merged with the data frame with the postal codes and suburbs.

3. Exploratory Data Analysis

3.1 Exploring Neighborhoods in Adelaide using Four Square API

Using the below Four-Square API query URL, I have explored venues in all suburbs in Adelaide.

https://api.foursquare.com/v2/venues/explore?client_id=CLIENT_ID&client_secret=CLIENT_SECRET&ll=LATITUDE,LONGITUDE&v=VERSION&radius=RADIUS&limit=LIMIT

The function getNearbyVenues extracts the following information for the data frame it generates—

- Venue ID
- Venue Name
- Venue Coordinates (Latitude and Longitude)
- Category Name
- Category Id

The function getVenues performs a category-based venue search to simulate user venue searches based on certain places of interest. This search extracts the following information—

- Venue ID
- Venue Name
- Venue Coordinates (Latitude and Longitude)
- Category Name

Using the getVenues function I have determined that there are a total of 51 Indian restaurants in Adelaide.

3.2 Analyzing the LGAs and suburbs in Adelaide

We can see from the below plot that **City of Onkaparinga** and **City of Port Adelaide Enfield** are the top 2 LGAs when it comes to the number of Suburbs. However this does not mean that these are the best locations in terms of popularity.
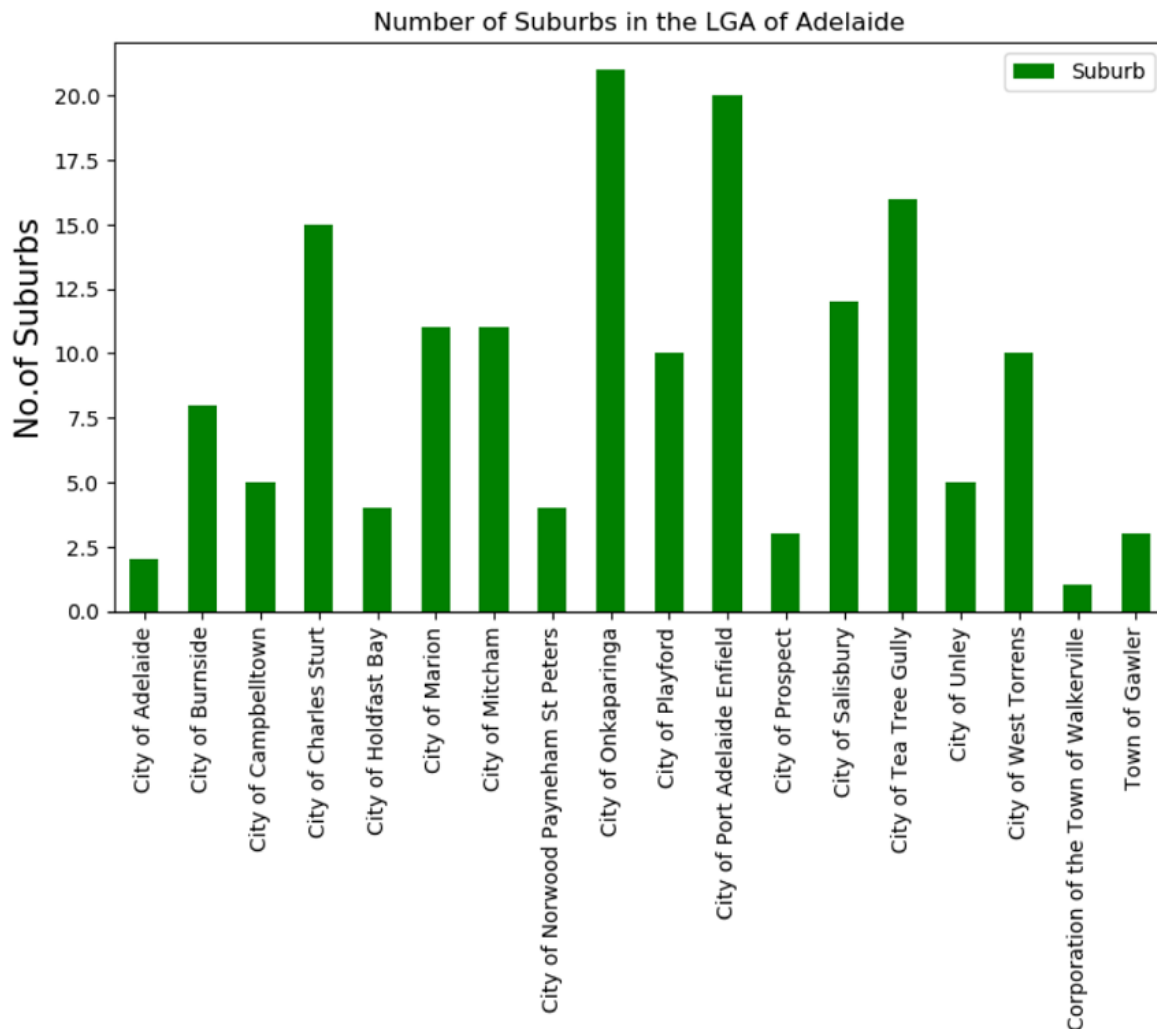
Now I will determine the number of Indian Restaurants in each LGA and we can see that City of Charles Sturt and City of Adelaide are top two with respect to number of Indian Restaurants.

Number of Indian Restaurants in Adelaide per LGA
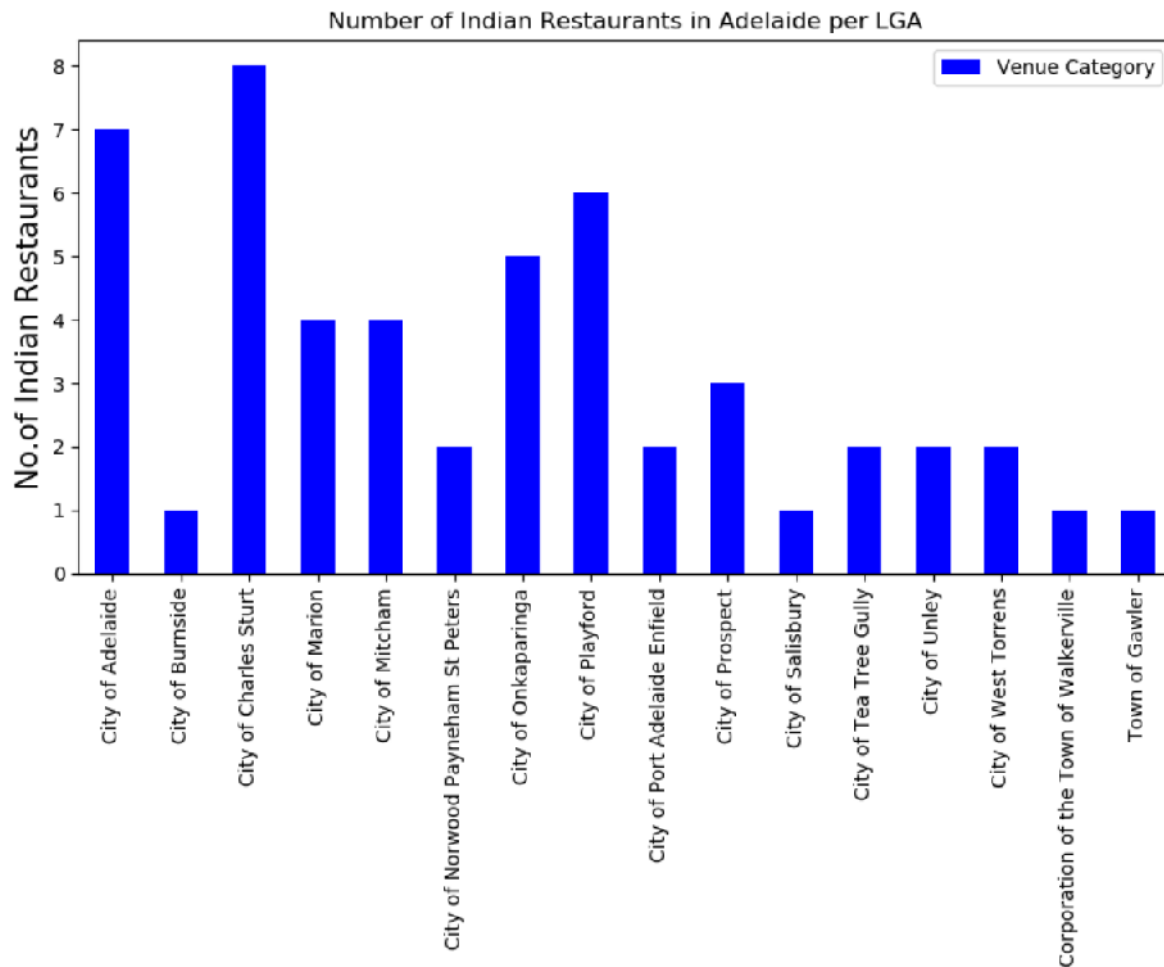
Now I will determine the top 100 venues for the suburbs of Adelaide and the below plot shows the number of venues per LGA.

From this we can see that City of Adelaide, City of Onkaparinga, City of Playford, City of Charles Sturt and City of Port Adelaide Enfield are top 5 LGAs with the most venues. So we will consider these 5 LGAs for our further analysis to determine the best locations for an Indian Restaurant.

Number of Venues in Adelaide per LGA
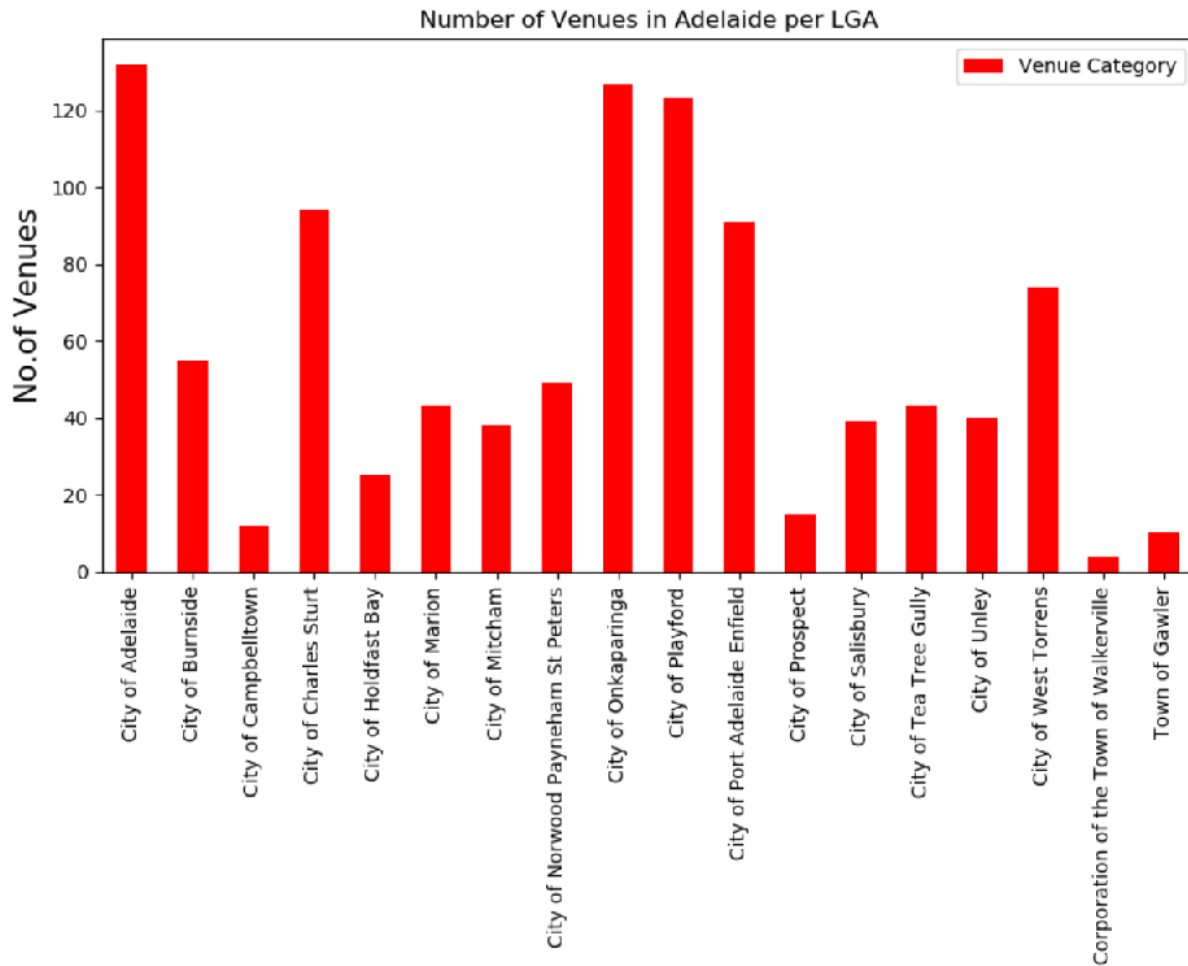
4. Visualization and Data Exploration

Based on above analysis I created a new data frame for these 5 LGAs. Now, exploring the data some more. To know about the top 5 venues of each suburb we proceed as follows -

- Create a data-frame with pandas one hot encoding for the venue categories.
- Use pandas groupby on the Neighborhood (i.e. Suburb) column and obtain the mean of the one-hot encoded venue categories.
- Transpose the data-frame from above step and arrange in descending order.

Once I have the broad overview of the different types of venues around 5 major LGAs of Adelaide, it is time to use clustering the suburbs using K-Means.

4.1 Clustering the Districts

Finally, we try to cluster these 5 LGAs based on the venue categories and use K-Means clustering. So our expectation would be based on the similarities of venue categories, these LGAs will be clustered. I have used the code snippet below—

Now we use *k*-means to cluster the neighborhoods into 5 clusters

```
# set number of clusters
kclusters = 3
adelaide_grouped_clustering = adelaide_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(adelaide_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0])

We create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster', kmeans.labels_)
adelaide_merged = city_adelaide

# merge adelaide_grouped with city_adelaide to add latitude/longitude for each neighborhood
adelaide_merged = adelaide_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Suburb')
adelaide_merged.dropna(inplace=True)
adelaide_merged.head()
```

	PostalCode	LGA	Suburb	Latitude	Longitude	Cluster	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	5000	City of Adelaide	Adelaide	-34.925650	138.599907	0.0	Bar	Café	Coffee Shop	Hotel	Sculpture Garden	Asian Restaurant	Vietnamese Restaurant	Italian Restaurant	Japanese Restaurant	Wine Bar
1	5006	City of Adelaide	North Adelaide	-34.909165	138.594803	0.0	Pub	Burger Joint	Italian Restaurant	Asian Restaurant	Thai Restaurant	Café	Wine Bar	Coffee Shop	Restaurant	Park
2	5007	City of Charles Sturt	Bowden,Brompton,Hindmarsh,Welland,West Hindmarsh	-34.901890	138.570164	0.0	Gym	Afghan Restaurant	Pub	Liquor Store	Bar	Bakery	Indian Restaurant	Burger Joint	Dog Run	Electronics Store

4.2 Folium Library and Map

Folium is a python library that can create interactive leaflet map using coordinate data. Since I am interested in knowing the locations for opening Indian restaurants at popular spots, I put the clusters on the map and then superimpose the current Indian restaurants on them.
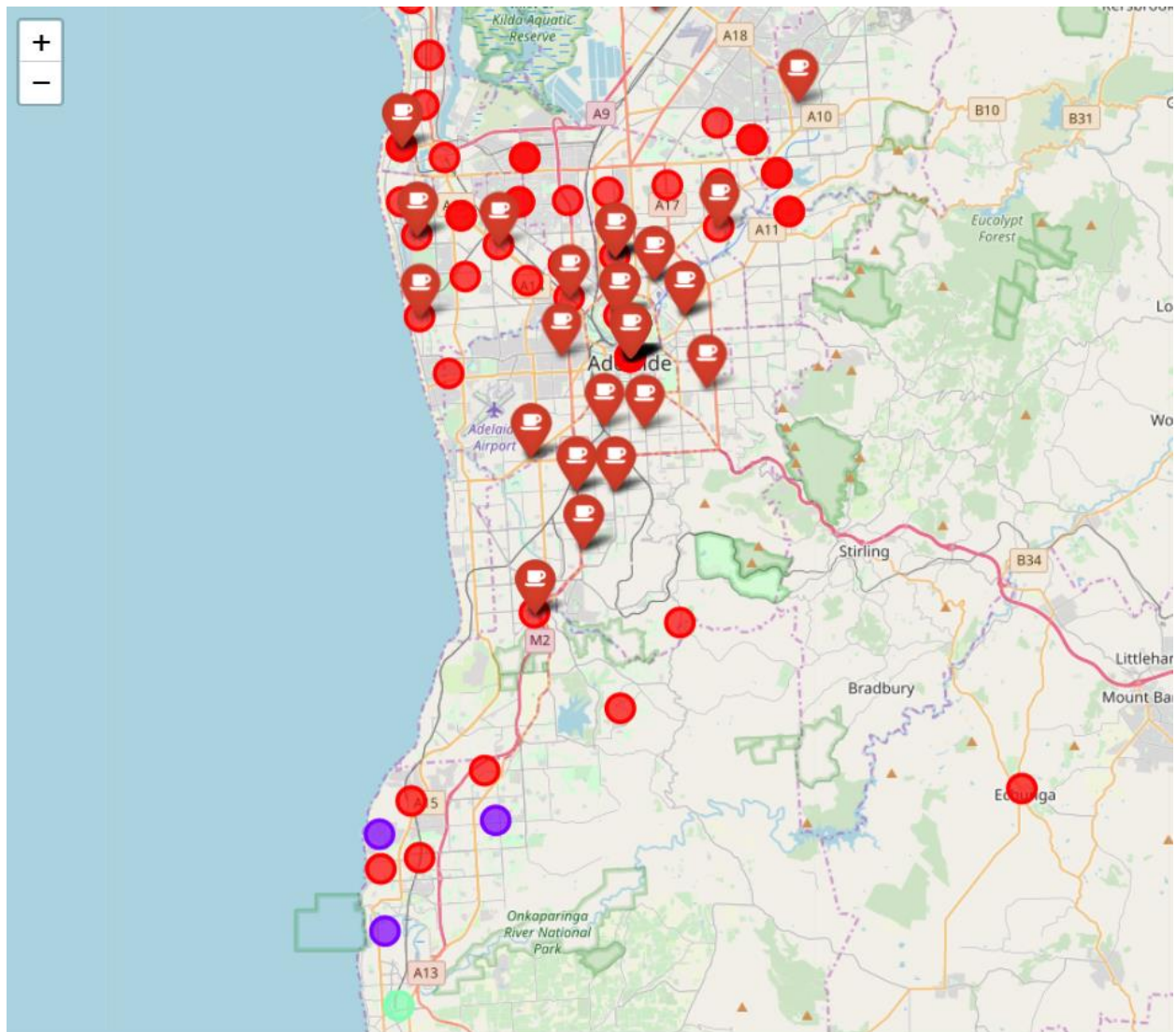
5. Results and Discussion

We have now reached at the end of the analysis, where we got a sneak peak of the 5 popular LGAs of Adelaide. As the business problem started with the best location for opening an Indian restaurant in one of the busiest districts, the data exploration was mostly concentrated on the LGAs with popular locations. I have used data from web resources like Wikipedia, python libraries like BeautifulSoup, Arcgis Geocoder and Foursquare API, to set up a very realistic data-analysis scenario. I have found out that City of Adelaide and City of Port Adelaide Enfield are dominated by restaurants as the the most common venue whereas City of Playford is dominated by bars and sport related venues as most common venues.

According to this analysis, City of Onkaparinga LGA will provide least competition for an upcoming Indian restaurant as the most common venue in this area does not include any Asian or Indian Restaurant. Also, the frequency of restaurants as common venue are very low compared

to the remaining LGAs. So, definitely this region could potentially be a target for starting quality Indian restaurants.

Limitations:

- The clustering is completely based on the most common venues obtained from Foursquare data.
- This analysis takes into consideration of only the 5 major LGAs of Adelaide, taking into account of all the LGAs would give us an even more realistic picture.

6. Conclusion

Finally, to conclude this project, I have used some popular python libraries to scrap the web-data, used Foursquare API to explore the major LGAs of Adelaide and determined the results of clustering and segmentation of LGAs using Folium leaflet map. Based on the analysis with the help of these methods I have made recommendations for probable locations for an Indian Restaurant. This is because in these areas, we have the highest number of popular venues visited by the people. Hence, opening an Indian restaurant here would definitely be a good business idea.

I have also listed some of the drawbacks and limitations and a chance for improvements to represent even more realistic results.

You can find the complete code at **Github**.