**IBM Certification for Professional Data Scientist (Coursera)**

**Topic chosen: BATTLE OF NEIGHBOURHOODS IN MELBOURNE**

1. Introduction
   1. About Melbourne
   2. Exploring Neighbourhoods
   3. Business problem
2. Data Acquisition and Data Cleaning
   1. Data Sources
   2. Data Cleaning
3. Visualization
4. Methodology
5. Results
6. Discussion
7. Conclusion
8. **Introduction**
   1. About Melbourne

Melbourne is the Capital city of Victoria State, Australia. It is the global centre for street art, live music, theatre and also hosts many International events.

This city is sweet home to many migrants in recent days and is always the most Liveable city as it is an International culture centre celebrating a wide variety of annual cultural events and festivals of all types.

1.2 Exploring Neighbourhoods

With my data science skills I acquired from IBM course in Coursera and also with the help of Foursquare API I am going to analyse different cuisines available in Melbourne city and also wanted to know the top 10 cuisines for each of the places in and around Melbourne. Since, I recently migrated to Melbourne from India and I love trying different cuisines made me choose this for my Capstone project.

1.3 Business Problem

For this project, I am going to put on my entrepreneur hat and create a simple guide as which will be the best place for a particular cuisine based on the Foursquare’s restaurant category, geographical location of restaurants, likes, comments and tips. After that I will cluster those restaurants based on the similarities so that anyone can determine what type of cuisines are best to eat based on feedback of Foursquare.

1. **Data Acquisition and Data Cleaning**
   1. Data sources

For the above said business problem data is collected from below link and Web scraping is done

<https://www.geonames.org/postal-codes/AU/VIC/victoria.html>

Based on Latitude and Longitude values obtained for Neighbourhoods, we use Foursquare API to get

1. Venue Name

2. Venue ID

3. Venue Location

4. Venue Category

5. Number of Likes

6. Top 10 Venues

2.2. Data Cleaning

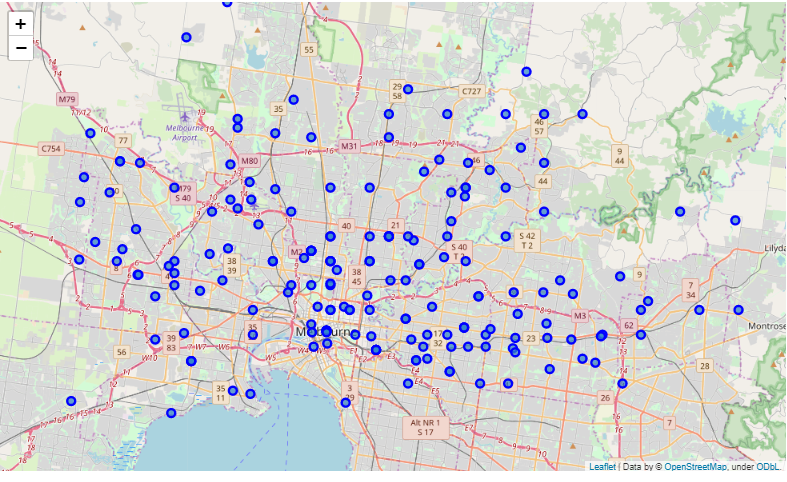
Using Beautiful Soup and Requests library, I scraped the data from the above link and converted to json file. Once it is done, I took table from the website having Postal code of Melbourne, Latitude, Longitude, Neighbourhoods, State and Country.

Since we do not require Country for our data as we already have other features to determine venues, I dropped that column and converted to a data frame having meaningful information.

I created account in [www.foursquare.com](http://www.foursquare.com) and got client ID and client secret to connect API to get the venues and clustered those venues based on similarities.

1. **Visualization**

The whole data is visualized using Folium Maps



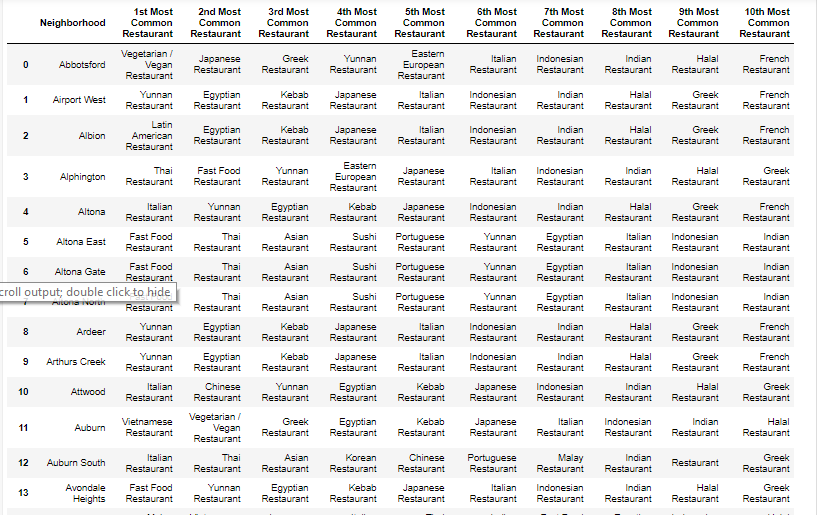
1. **Methodology**

After scraping the data from the above link, since we want to analyse only for the Neighbourhoods of Melbourne, condition is applied to remove rows having ‘Vic’ as it implies Regional Victoria which we don’t need for our analysis.

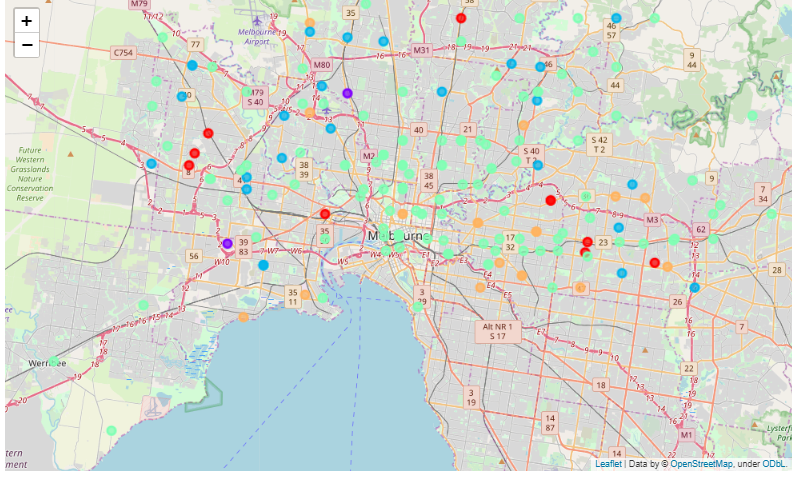


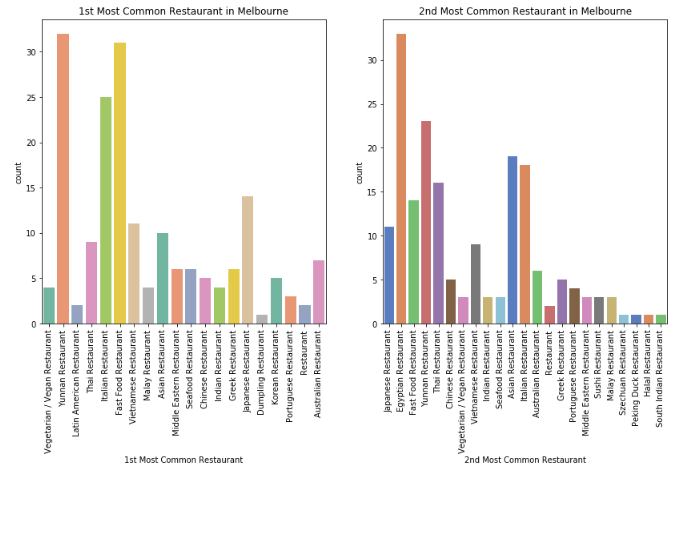
Foursquare API is used to get the nearby venues based on Likes, tips, Frequency and from those venues, we filter venues limiting the categories only to 100 and radius of each Neighbourhood to 1000.

Onehot Encoder is applied as we have categorical features. Since our aim is to look for restaurants, we apply condition and get only categories ending with ‘Restaurant’. Based on the frequency of the venues in each Neighbourhoods, we take only 10 most common restaurants for each Neighbourhoods. We obtain a table having all Neighbourhoods and their top most common restaurants. Since it is difficult to analyse, I visualized based on the counts of the Most common Restaurants and Conclusions are drawn.



K-Means Clustering Algorithm is applied to Neighbourhoods and clustered based on the similarities. Using Folium Maps, the clusters are visualized for better understanding of each Neighbourhoods in Melbourne.





1. **Results**

Based on the Visualization we can arrive to a following Conclusion

1. In the 1st Most common Restaurants in Neighbourhoods, it is clearly seen that ‘Yunnan Restaurant’ is the most common cuisines in Melbourne and most of the Restaurants has it followed by Fast food and Italian cuisines
2. In the 2nd Most common Restaurants in Neighbourhoods, it is seen that ‘Egyptian Restaurant’ is more in number followed by Yunnan and Asian cuisines.

In the plot having the 1st most common restaurants, there is cuisine called Restaurant and since we do not know what it serves, we assume those Restaurants have all types of cuisines and they commonly mentioned as Restaurant.

1. **Discussion**

It is really interesting to see these results and after applying K-means clustering, one of the clusters with many Neighbourhoods having Restaurants are situated close to Melbourne Central Business District.

The results obtained are purely based on the Foursquare API likes, tips and most visited restaurants. The radius for each Neighbourhoods used for this analysis is 1000m and Categories are limited to 100. Hence the results can still be modified by using lower or higher radius for each Neighbourhoods.

The real challenge for this project includes

1. Web scraping and extracting table from the website.
2. For data obtained through API calls, it is likely to get different results with different set of variables and at different point of time. Multiple trial and error runs are required to get optimal results.
3. It was initially difficult to choose Neighbourhoods which are based only in Melbourne as there is always a chance to miss out some Neighbourhoods having meaningful information on Venues.
4. **Conclusion**

Interesting results are obtained based on the analysis of Neighbourhoods. From all the graphs we can clearly mention that Melbourne has wide variety Restaurants and Cuisines.

By taking all the findings from this analysis and considering Environment, Financial factors we can arrive at a conclusion as where to open a restaurant that can give us high Revenue.