**Capstone Project - Final Report**

## Introduction:

We want to open a new coffee shop in Melbourne. We want to identify current locations where we have coffee shops. We will use `Four Square` data to determine which district in Melbourne currently have coffee shops. This will tell us which districts have more coffee shops and which have less

Business Problem:

`Melbourne` is the capital and most-populous city of the Australian state of Victoria, and the second-most populous city in Australia and Oceania.

Melbourne is known for its bars, street art and coffee culture. As a coffee lover city, it makes complete sense to open coffee shop. The challenge is to understand which district is the best option to open the new coffee shop.

## Technique we will use:

We will use following techniques to gather information:

- We will extract the list of districts in Melbourne using `request` and `beautiful soup`

- We will use `Geographical Co-ordinates` to get Lat and Long of these districts

- We will use `Four Square API` to get the list of coffee shops in these districts

- We will cluster these locations based on the coffee shops

## Target Audience:

- Entrepreneurs who wants to open a new coffee shop in the city however, are not sure where to open a new shop

- People who already have a shop and want to open a new franchise in different district

## Data Description:

- Brief information about Melbourne:

https://en.wikipedia.org/wiki/Melbourne

- Link of Melbourne Districts:

https://en.wikipedia.org/wiki/Postal\_district\_numbers\_of\_Melbourne

- Four Square API link:

https://developer.foursquare.com/docs/

## Methodology:

* We used request and beautiful soup libraries to extract data from Link of Melbourne Districts.
* This list was then converted into a data frame
* Then using Geographical Coordinates, we got Lat and Long of Melbourne and then we got Lat and Long of all the districts in our data frame
* We visualized our data to see how spread they are in Melbourne
* Using Four Square API we successfully extracted, list of locations where category included Coffee
* We created a word cloud to see the most repeated word in our category column
* We cleaned our data by excluding all the shops which do not belong to Café
* We included top 7 venues to each district
* We visualized each top venue
* We applied one hot encoder on this dataset and grouped them by their mean frequency
* This data was then clustered into 4 different using KMeans Clustering
* We created individual visuals for each cluster using folium

## Results and Discussions:

Our analysis shows that there are many coffee options in Cluster number 2 and 3. There is an opportunity to open a coffee house in Cluster number 1 and 4. Number 4 is interesting as it is nearby Airport and ideally should be good location.

Cluster 2 and 3 are more towards the main suburb, whereas, cluster 1 and 4 are little far off.

From the above analysis, we can also understand in cluster 1 and 4 there are not many dedicated coffee shops, which states that many restaurants/bar may be selling coffee. This gives an opportunity to the new business to open a fancy coffee shop and be successful.

Cluster 2 and 3, may have lot of coffee shops. On further analysis it was observed both clusters are close to Albert Park and Melbourne Station. Albert Part is famous for Formula 1 race track and we have multiple sporting activities near Melbourne station such as Melbourne Cricket Ground, Melbourne Tennis Stadium. This justifies why so many cafes are around these landmark point

## Conclusion:

Purpose of this project was to cluster Melbourne districts based on coffee shops and we were able to do it successfully. If any business wants to open a new coffee shop in Melbourne, we can assist them with the clustering we have done on the district. Cluster 2 and Cluster 3 are more towards down town, whereas, cluster 1 and 4 are more spread out.

Final decision will be made by the stakeholder in which cluster they are looking to open a new cafe.