

# Coursera Capstone

## IBM Applied Data Science Capstone

***Opening a New Café in Kochi, Kerala, India***

By: Job Thomas T

June 2020

# Business Problem

- Location of the Café is one of the most important decisions that will determine whether the mall will be a success or a failure
- Objective: To analyze and select the best locations in the city of Kochi, Kerala, India to open a new Café
- This project is timely as the city is currently becoming more popular among tourists and visitors
- Business question
  - In the city of Kochi, India, if an entrepreneur is looking to open a new Café, where would you recommend that they open it?

# Data

- Data required
  - List of neighborhoods in Kochi
  - Latitude and longitude coordinates of the neighborhoods
  - Venue data, particularly data related to Cafés
- Sources of data
  - Wikipedia page for neighborhoods  
([https://en.wikipedia.org/wiki/Category:Suburbs\\_of\\_Kochi](https://en.wikipedia.org/wiki/Category:Suburbs_of_Kochi))

○Geocoder package for latitude and longitude coordinates

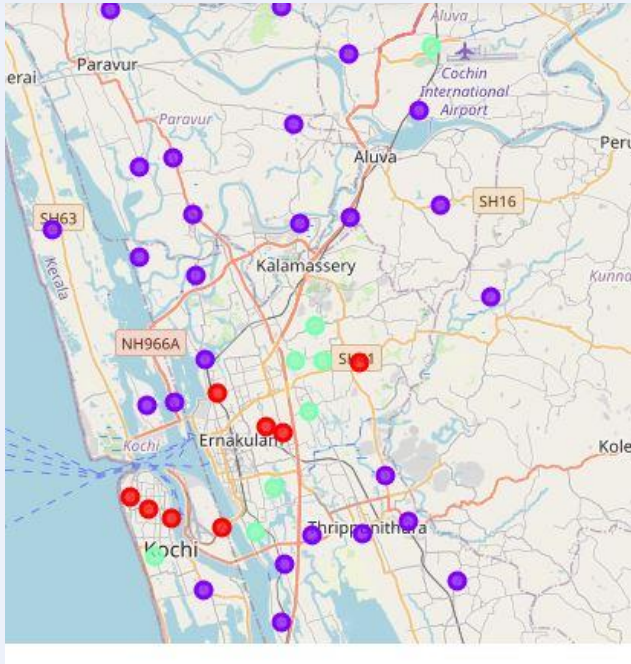
○Foursquare API for venue data

# Methodology

- Web scraping Wikipedia page for neighborhoods list
- Get latitude and longitude coordinates using Geocoder
- Use Foursquare API to get venue data
- Group data by neighborhood and taking the mean of the frequency of occurrence of each venue category
- Filter venue category by Café

- Perform clustering on the data by using k-means clustering
- Visualize the clusters in a map using Folium

# Results



- Categorized the neighborhoods into 3 clusters :



- Cluster 0: Neighborhoods with high number of Cafés (Red )
- Cluster 1: Neighborhoods with low number to no existence of Cafés (Purple)
- Cluster 2: Neighborhoods with moderate concentration of Cafés (Light Blue)

## Discussion

- Most of the Cafés are concentrated in cluster 0 and 2 which is the central area of Kochi in the Ernakulam district.
- Highest number in cluster 0 and moderate number in cluster 2
- Cluster 1 has very low number to no Café in the neighborhoods
- Cluster 1 is in the outskirts of city and have less chance of becoming successful.

- Cluster 2 has area is within the city, and face only moderate competition. Hence they have a good prospects of becoming successful.

## Recommendations

- Open new Cafés in neighborhoods in cluster 2 with moderate competition
- Avoid neighborhoods in cluster 0, as already high concentration of Cafés and intense competition



# Conclusion

- Answer to business question: The neighborhoods in cluster 2 are the most preferred locations to open a new Café
- Findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new Café

Thank you!

[illegible]