

# Coursera Capstone

## IBM Applied Data Science Capstone

Opening a New Cafe in Mumbai, India

BY: NISHCHAY NAGPAL

JULY, 2020

# Business Problem

- ▶ Objective: To analyze and select the best locations of Cafes in the city of Mumbai, India to open a new cafe
- ▶ Location of the cafes is one of the most important decisions that will determine whether the shop will be a success or a failure
- ▶ Business question
  - ▶ In the city of Mumbai, India, if a person is looking to open a new cafe, where would you recommend that they open it?

# Data

- ▶ Data required
  - ▶ List of neighborhoods in Mumbai
  - ▶ Latitude and longitude coordinates of the neighborhoods
  - ▶ Venue data, particularly data related to cafes
- ▶ Sources of data
  - ▶ Wikipedia page for neighborhoods  
([https://en.wikipedia.org/wiki/Category:Suburbs\\_in\\_Mumbai](https://en.wikipedia.org/wiki/Category:Suburbs_in_Mumbai))
  - ▶ 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 into 5 neighborhood clusters:
  - ▶ Cluster 0: Neighborhoods with low number of Cafe
  - ▶ Cluster 1: Neighborhoods with moderate number to no existence of Cafe
  - ▶ Cluster 2: Neighborhoods with high concentration of Cafe
  - ▶ Cluster 3: Neighborhoods with highest concentration of Cafe
  - ▶ Cluster 4: Neighborhoods with moderate concentration of Cafe

# Conclusion

- ▶ Cafes are concentrated in the southern and eastern area of Mumbai
- ▶ The highest number in cluster 2 and moderate number in cluster 4, cluster 0 has very low number to totally no Cafes in the neighborhoods.
- ▶ This represents a great opportunity and high potential areas to open new Cafes as there is very little to no competition from existing shops.
- ▶ Meanwhile, cafes in cluster 3 are likely suffering from intense competition due to oversupply and high concentration of cafes.
- ▶ This project recommends developers to capitalize on these findings to open new cafes in neighborhoods in cluster 0 with little to no competition.
- ▶ Developers with unique selling propositions to stand out from the competition can also open new cafes in neighborhoods in cluster 2 and cluster 4 with moderate competition.
- ▶ Lastly, developers are advised to avoid neighborhoods in cluster 3 which already has high concentration of cafes and is suffering from intense competition.



Thank You