

# Coursera Capstone

## IBM Applied Data Science Capstone

### Finding a potential venue for opening new Hotel in Mumbai, India

By: Hardik Dave

# Introduction/Business Problem

- ▶ In a metro city like Mumbai, the business tourism has been developing quite rapidly.
- ▶ The main motivation of doing this capstone project is to develop a plan to find best suitable venue or location for opening new hotels in Mumbai, India.
- ▶ The business problem can be defined as “If we have data of neighborhoods and venue details, can we suggest a potential venue or area for starting new hotel project?”
- ▶ The target audience for this project is property developers and hotel business professional builders who are looking to invest or venture by opening new hotel

# Data Description

- ▶ Following data are required for developing this project :
- ▶ List of suburbs in Mumbai, India
- ▶ Coordinates for each suburb (Neighborhood) in terms of Longitude and Latitude
- ▶ The data related to Hotels (Venue Data)
- ▶ The suburbs data can be retrieved from [https://en.wikipedia.org/wiki/Category:Suburbs\\_of\\_Mumbai](https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai)
- ▶ We can get coordinates in terms of Longitude and Latitude from Geocoder package of python for each suburbs of Mumbai.
- ▶ Foursquare API can be used for collecting venue data for the neighborhoods.

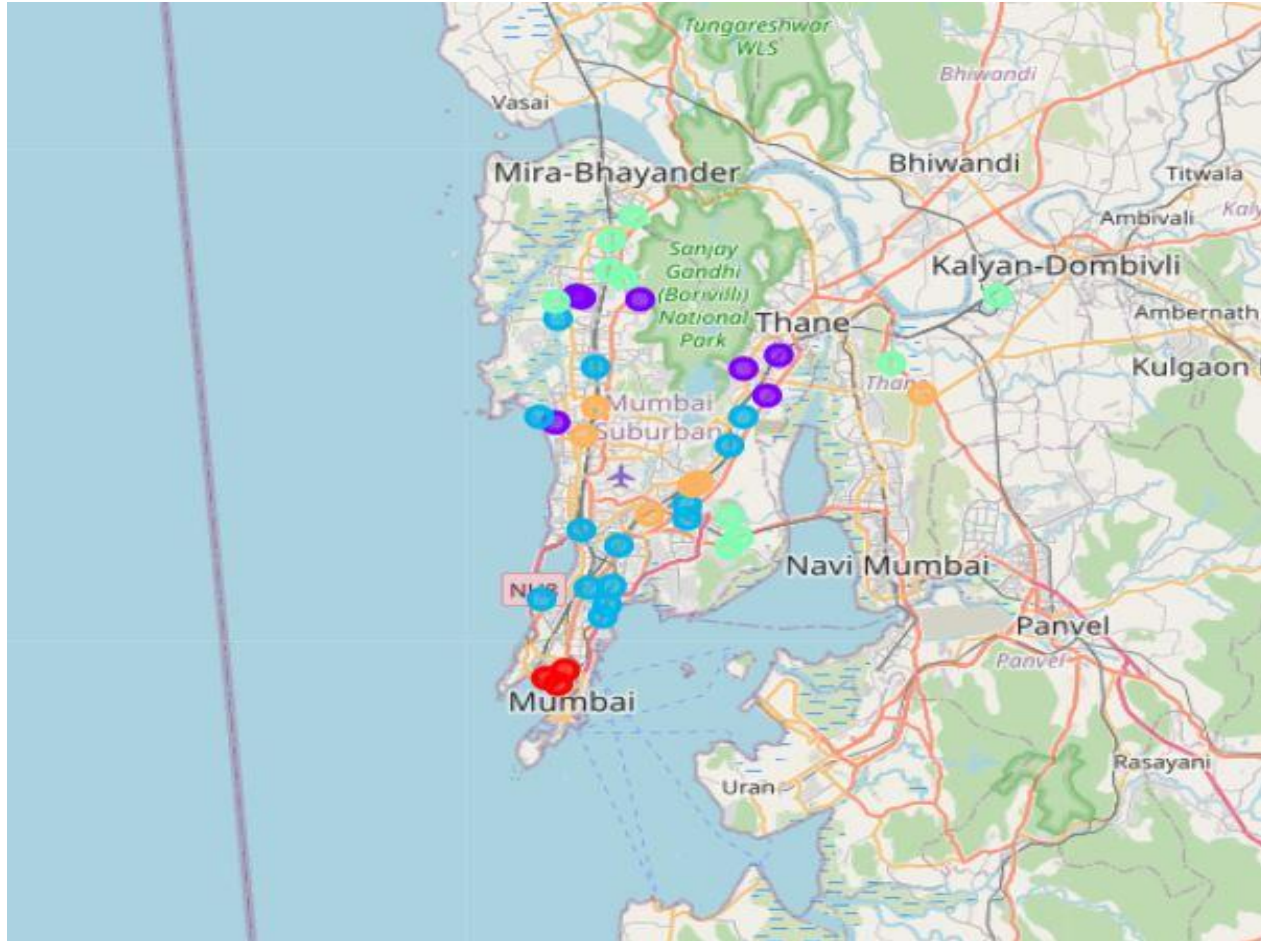
# Methodology

- ▶ We can gather suburbs data from [https://en.wikipedia.org/wiki/Category:Suburbs\\_of\\_Mumbai](https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai)
- ▶ We can use Web Scrapping to collect this data which can be further used for getting coordinates (Longitude and Latitude) of each suburb.
- ▶ Foursquare API can be used to collect information of surrounding and other information by querying with coordinates.
- ▶ We can pass this information to Foursquare API calls to get venue data in JSON format.
- ▶ For this project we have used top 100 venues within the radius of 5000 meters. From the collected data, we extract venue name, category, longitude value and latitude value

# Methodology (Cont.)

- ▶ We get 3805 values for different venues associated with different suburbs. There are 159 unique categories found for venues.
- ▶ Then we extract information which contains “Hotel” as venue category. So we can have new dataframe consisting only neighborhood and hotel information.
- ▶ We can now use clustering algorithm K-Means Clustering to cluster these information in 5 different clusters.
- ▶ After clustering we can visualize the results. Also we can view details of suburbs falling in each cluster individually to get an idea of hotel density among these suburbs.

# Results



## Results (Cont.)

- ▶ The map indicates that there are different density of hotels in different areas. So that we can visualize this in a way to determine potential areas for developing new hotel project.
- ▶ By results of clustering and seeing the suburbs and their respective hotel venue data, we can draw an important insight as follows: Mostly the hotels are scattered in cluster 3 and 5.
- ▶ Because these areas have higher financial capabilities and nearer to Railway Stations and Airport. We can see that cluster 2 and 4 have almost no hotels in the neighborhood.
- ▶ So that we can see these areas as potential areas for opening new hotels.

# Discussion

- ▶ From the results and observations, we can say that Cluster-2 and Cluster-4 are most suitable for development of new hotel project.
- ▶ The suburbs in the areas falling under Cluster-2 and Cluster-4 are can be thoughtful suggestions locations for building new hotel.
- ▶ In contrast, the Cluster-1 and Cluster-5 has excessive amount of hotels as compared to other areas.
- ▶ So we can say that in the areas falling under the Cluster-2 and Cluster-4 has great potential and provide great opportunity for building a new hotel project



# Conclusion

- ▶ From this project, we can conclude that we can find potential areas or locations which can be beneficial for developing and starting new hotel business.
- ▶ We can use data science concepts along with the appropriate data and usage of data to make suggestions which can not be given without looking at and understanding data.
- ▶ We have used Wikipedia data for collecting suburbs information of Mumbai, India which is fed to Geocoder package for obtaining coordinates.
- ▶ Then we have employed Foursquare API for extracting the venues information based on the longitude and latitude information of each suburb.
- ▶ Then the clusters are created based on the data using K-Means clustering algorithm and from the results we can suggest or recommend potential and beneficial areas for developing new hotel project.

# References

- ▶ Suburbs in Mumbai, India. Source :  
[https://en.wikipedia.org/wiki/Category:Suburbs\\_of\\_Mumbai](https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai)
- ▶ Foursquare API Documentation. Source :  
<https://developer.foursquare.com/docs/places-api/>

Thank you...