



Coursera Capstone Project, IBM Data science Certification

**G HARIKRISHNAN** 



# PROBLEM STATEMENT

- To find the location for Theatre and restaurant business in Trivandrum, India which would be profitable.
- To utilize data science approach to find optimum localities in the city for opening the sad businesses.

### TARGET AUDIENCE

- This project is oriented for the people who are looking to start off a business in the city
- Since the global recession is happening, it would be beneficial for people or group of people who are in need to start a business investment.
- Individuals looking to expand the business to new areas.

#### DATA



- The list of locations In the city is available through the Wikipedia page (<a href="https://en.wikipedia.org/wiki/Category:Suburbs\_of\_Thiruvananthapuram">https://en.wikipedia.org/wiki/Category:Suburbs\_of\_Thiruvananthapuram</a>).
- Using these position vectors and foursquare api we will get the location data of all these places and their corresponding business.
- This acquired data will first be used for location of theatre and then for restaurants.
- For both these cases we will be dividing the data into useable columns and then clustering them.
- These clusters will then be analysed to find the best spots. The best spots can be chosen by the data in conjunction with the nearness of the location to certain point of interests

### **METHODOLOGY**



- First of all, we import all the python run time libraries like Numpy and pandas. We also import scikit learn for the clustering and folium and geocoder for the visual analysis.
- Using Beautifulsoup we scrap a Wikipedia
  page(<a href="https://en.wikipedia.org/wiki/Category:Suburbs\_of\_Thiruvananthapuram">https://en.wikipedia.org/wiki/Category:Suburbs\_of\_Thiruvananthapuram</a>) for the locations in the city.
- Using geocoder, a utility we find the latitudes and longitudes of all these localities. We use an ArcGIS instead of Google for acquiring the Coordinates.
- Using map visualisation, we remove those places which are wrongly pointed and those which are overlapping. These points are dropped from the data frame.
- Using foursquare API, we then find the venues nearby the localities, Here I have placed a limit of 3000 (3 KM).
- This is a random assumption that a person won't be willing to look for food beyond 3 KM radius of his house.

#### METHODOLOGY CONT.

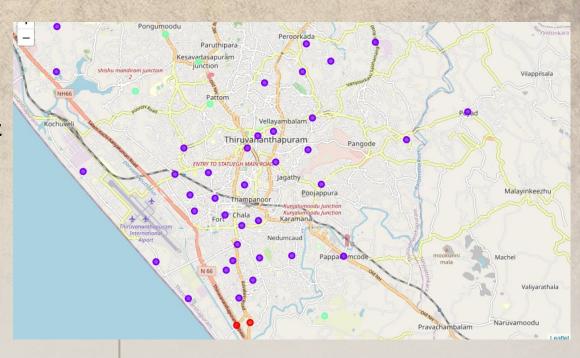
- While passing all these credentials we will get all the area details with foursquare. Using dummy variable can see what all are available.
- We the take a frequency of these in the areas and take a group it by each neighbourhood
- For the theatre, from the frequency data frame we choose the columns that are the shows the frequency of theatres.
- Here chose different columns as all those will indirectly mean that a theatre is present in that location. These are then made into a separate data frame.
- Using K means method of unsupervised clustering, we then split this into 4 frames depending on the values of these columns.
- Using the different colour in the map we can observe what is the cluster of each points in the map.

- Also, the map allows us to get insight of the locality where the theatres are more, the geo economic condition of that are.
- The finally we separate the data frames individually to know the underlying information of all cluster.
- Just like in the case of theatres, we made a separate data frame from the frequency data frame.
- Using K means clustering we divided the area into 3 individual locations.
- The clusters along with the location of each clusters are then utilised to forma a map using folium.
- The observations are done here also of the locality. The clusters are then individually analysed to come to a conclusion.

#### **RESULT - RESTAURANT**

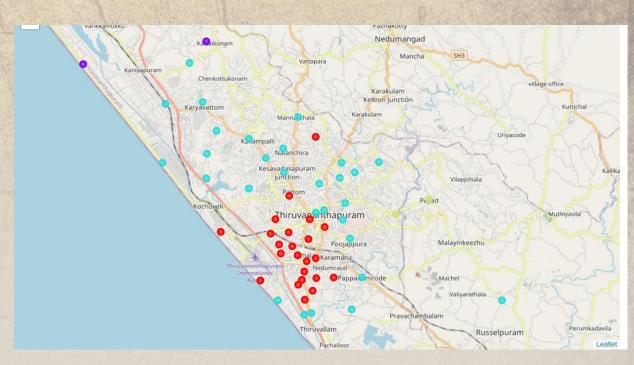


- Cluster I has large concentration of restaurants hence in cluster I we should not open a restaurant,
- Cluster 2 we see that neither restaurant nor fast food restaurants are common, means they are rare, these even include the city centre.
- Cluster 3 has good number of restaurants and fast food chains but lower than cluster 1.
- So best option will be to open restaurants in areas in cluster 2.



# **RESULT - THEATRE**

- Cluster I has high concentration of movie theatres, this can be seen in the map also that these are the sites in the heart of the city, hence this also points out that the city centre is having sufficient theatres.
- Cluster 2, it lies far from city centre but the cluster has seen that here are theatres but low in number.
- Cluster 3 lies throughout the city, has low to none of the amount available theatres.
- Cluster 4 is similar to cluster 2 has multiplexes.
- From a business point of view, it is better to open a theatre in the places marked in cluster 3



#### CONCLUSION

- This analysis will prove good for any business man who is interested in starting or acquiring business in the area.
- To directly answer the problem statement, I would like to suggest <u>cluster 3 in the analysis for starting a Theatre business</u> and <u>cluster 2 of the analysis of restaurants for starting a restaurant business</u>.

# REFERENCES

- <a href="https://en.wikipedia.org/wiki/Category:Suburbs\_of\_Thiruvananthapuram">https://en.wikipedia.org/wiki/Category:Suburbs\_of\_Thiruvananthapuram</a>
- <a href="https://foursquare.com/developers/apps">https://foursquare.com/developers/apps</a>