

Applied Data Science Capstone Project

The Battle of the Neighborhoods

City of Choice: Karachi, Pakistan

Introduction

Karachi is the largest city of Pakistan and also its commercial capital as well as the 7th largest city of the world.. Like all major cities around the globe, the food culture in Karachi is top-notch and diverse. There are new restaurants opening up everywhere, but some of them last longer than others. This is because not enough new eateries are able to fulfil their goals and end up shutting down.

Assuming that I want to set up a small restaurant in one of the city's neighbourhoods, I'd have to compare the neighbourhoods in the city to evaluate which would be more suitable. There are multiple factors that can help determine which neighbourhood would be appropriate for setting up the restaurant. In the capstone project, I will be evaluating public places like restaurants nearby, shopping places, movie theatres and the like to see which area is better suited to my business needs.

This project will also be able to help anybody who is looking to set up an office or a restaurant or has any kind of business that is likely to be affected by the crowds coming in from public spaces.



Data

The data will be extracted from the following links and evaluated to solve the problem above:

- Foursquare data for Karachi:**

<https://foursquare.com/v/karachi--%DA%A9%D8%B1%D8%A7%DA%86%DB%8C/4e5ddb7814954da39fe8128c>

The foursquare data from Karachi has details on restaurants, schools, movie theatres and malls. Since it is likely that such places will attract people, it will be worthwhile to visualise and see whether a restaurant is likely to do well here.

- A list of neighbourhoods in Karachi, provided by the Karachi Metropolitan Corporation:**

<http://www.kmc.gos.pk/Contents.aspx?id=13>

Postal Code | Dialing Codes | International dialing codes | Emergency number

POSTAL CODE

POSTAL CODE	
AREA	POSTAL CODE
AIRPORT	72500
BALDIA TOWN	75760
BOARD OF SECONDARY EDUCATION	75150
CANTT	75530
CITY GPO	7100
CLIFTON	75600
COD	75250
DARUL-ULOOM	75180
DEFENCE SOCIETY	75500
EXPORT PROCESSING ZONE	75150
FEDERAL B AREA	75950
GULSHAN-E-IQBAL	75300
HABIB BANK	75650
HOTEL METROPOLE	75520
JINNAH POST GRADUATE MEDICAL CENTER	75510
KARACHI GPO	74200

The data present on KMC website which we will use for our project

Methodology.

Main Libraries/Modules used:

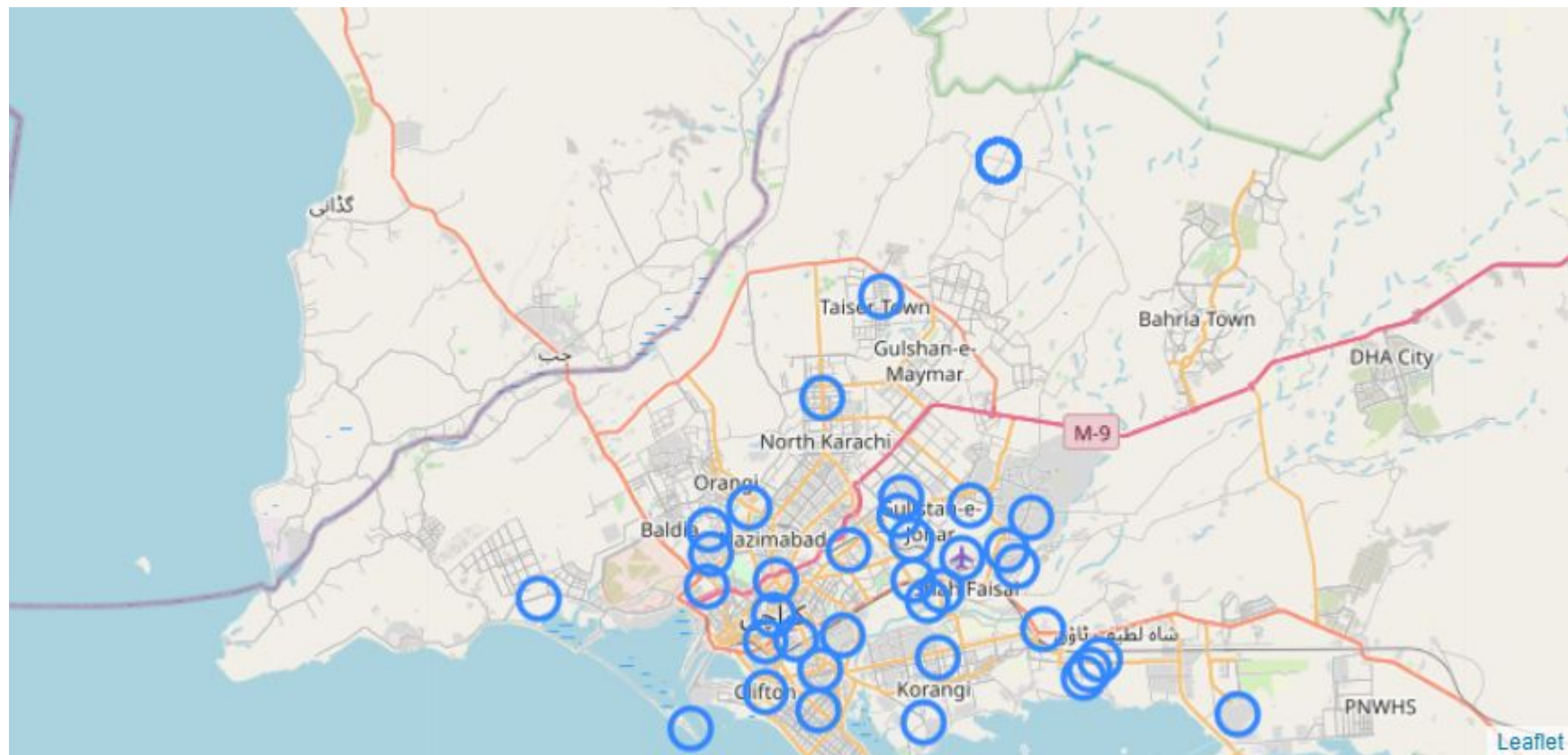
- Pandas
- Numpy
- BeautifulSoup
- Requests
- Folium
- Geocode

The initial steps will be that of data collection and preprocessing from the sources mentioned above. We have used requests and BeautifulSoup to scrape the web page to find the table that we need. We then clean the table and appropriate headings to get a clean table.

The table has the areas of Karachi, from the Karachi Metropolitan Corporation.

We then use the list to find the latitude and longitude of each neighborhood. And then we map Karachi and its neighbourhoods to visualise the distances between them and the way they are clustered.

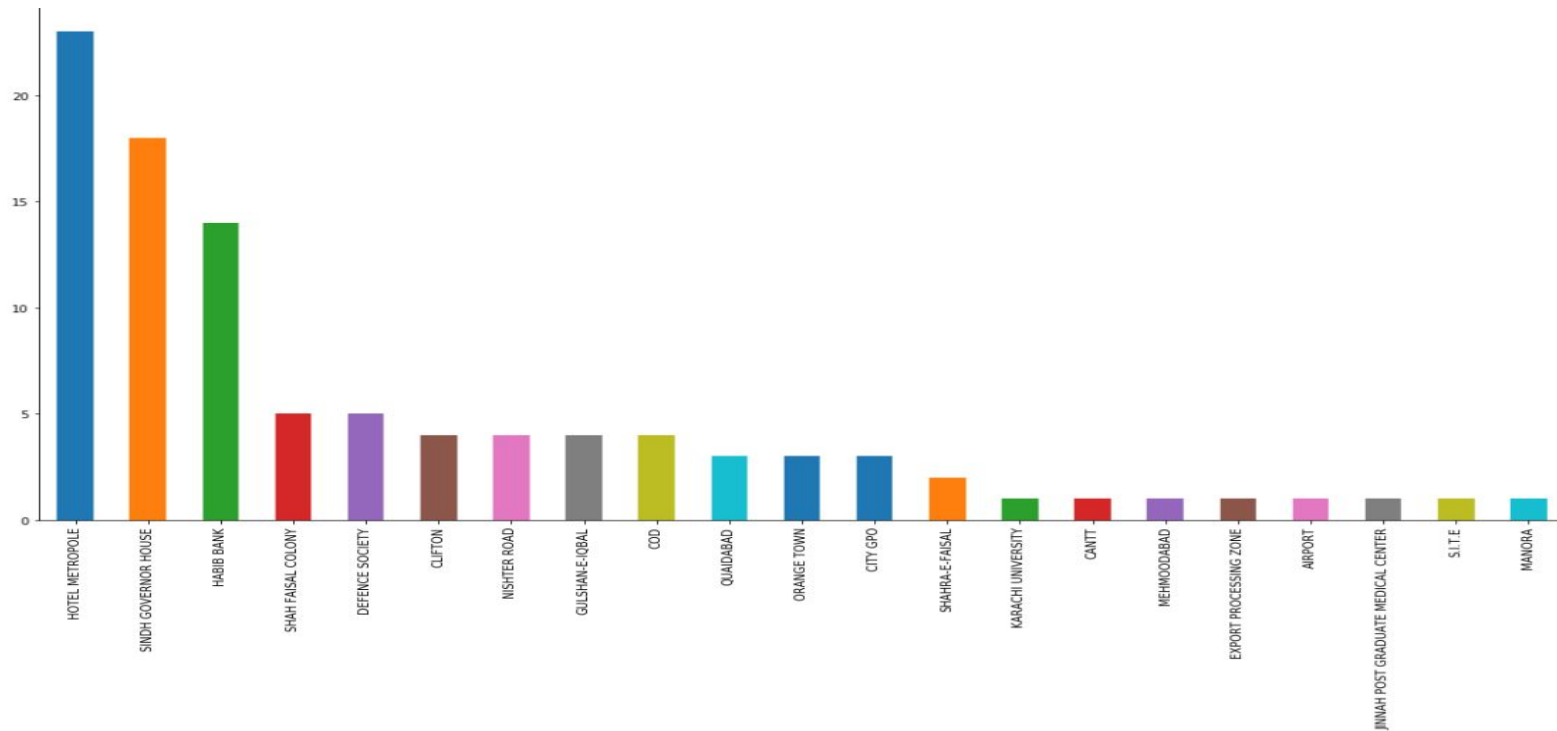
Finally, we use FOURSQUARE API to tabulate the nearby venues of each area present on the foursquare website. This helps us evaluate which areas have the most public places and hence would be a better option for opening a restaurant.



Out[92]:

	Neighborhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	AIRPORT	24.900819	67.159419	Jinnah Airport Cargo Terminal	24.898311	67.155832	Airport Terminal
1	CANTT	24.925758	67.205640	Cantt Bazaar	24.927427	67.204516	Business Service
2	CITY GPO	24.866779	67.031129	Student Biryani	24.864647	67.031615	Diner
3	CITY GPO	24.866779	67.031129	Electronic market	24.867294	67.028642	Electronics Store
4	CITY GPO	24.866779	67.031129	Shahabuddin Market	24.863848	67.034594	Fish Market

The Results



Results

We can see that according to the above data, the top 5 most popular areas in Karachi are Hotel Metropole, Sindh Governor House, Habib Bank, Shah Faisal Colony and Defence Society. Therefore, if one is looking for an area in Karachi to invest in property for a business/restaurant that has the most number of public places (to potentially attract most customers), it should be one of the five areas mentioned in the previous slide.

Discussion

It is to be noted that in this project, we have only used No. of Public places as a criteria for the selection of a neighbourhood. The idea is based on the probability that if an area has more public places, it will have more people passing by and hence will be more likely to attract more customers.

Possible Improvements

However, in the real world, such scenarios are based on many other practical considerations such as distance of the area from a current location, the commute from home, and property prices in the locality. This could also, in the future, be used to evaluate what is the best area for each unique individual, their budget and their business needs.

Conclusion

In Week 5 of the Applied Data Science project, we have implemented a real world scenario with real data (from KMC and Foursquare), preprocessed and cleaned it, applied visualization techniques to show a map of Karachi and its neighbourhoods. We have also used the Foursquare API to get places nearby (or in) each area so we can evaluate which part of the city has the most public places. This information helps us solve the problem we've wanted to solve in this project: which is a suitable area in the city for setting up a restaurant?

End of slide show.