

# IBM DATA SCIENCE CAPSTONE PROJECT

IDENTIFYING THE BEST LOCATION FOR  
YOGEN FRÜZ FRANCHISE

OMER ZIA

# INTRODUCTION

- Yogen Früz has grown to be a world leader in the frozen yogurt category, with over 1400 locations operating in 47 countries around the world. Yogen Früz was recognized by Entrepreneur Magazine for its accomplishments and rated the number one Franchise in the World in 1999 among the Franchise 500.
- Yogen Früz has five franchises/outlets in Pakistan, in three major cities (Islamabad, Lahore & Karachi). Lahore has two franchises. Flagship outlet in DHA and Johar Town, Lahore.
- Yogen Früz provides delicious and nutritious frozen desserts, just what you need to live a healthy and active lifestyle.



# BACKGROUND

In the last 12 months or so, frozen yoghurt has become wildly popular in Pakistan. While ice-cream is still the most popular frozen dessert in the country, frozen yoghurt or 'froyo', as it's commonly referred to, has been able to carve out a niche market for itself.

Therefore, it is advantageous for Yogen Früz to accurately identify areas/towns/UC where the growing demand is and where the new franchise can be opened.



# BUSINESS PROBLEM

- The objective of this capstone project is to analyze and select the best locations in the city of Lahore, Pakistan to open a Yogen Früz - Frozen Yogurt point. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Lahore, Pakistan, if an investor is looking to open a Frozen Yogurt point, where would you recommend that they open it?



# TARGET AUDIENCE

This project is particularly useful to investors looking to open or invest in new Frozen Yogurt joint in the Metropolitan city of Pakistan i.e. Lahore.

This project is timely as the demand of frozen yogurt items is on rise and there are limited places for a big city like Lahore.



# DATA

The Neighborhoods / Towns of Lahore, Pakistan from Wikipedia:  
[https://en.wikipedia.org/wiki/List\\_of\\_towns\\_in\\_Lahore](https://en.wikipedia.org/wiki/List_of_towns_in_Lahore)

The coordinates (latitude, longitude) to these Towns/UC of Lahore from Open Street Map APIs / Postal Codes

From Foursquare we will need following venues data:

- the Frozen Yogurt/Yogurt venues of all Towns/UC
- the office's venues of all Towns/UC
- the high school's venues of all Towns/UC
- the universities venues of all Towns/UC

We will then leverage the data in order to determine which locality is the most appropriate in order to locate the Frozen Yogurt Franchise / Joint venue.



# METHODOLOGY

- For each locality, all office, school, university and venues data have been collected from Foursquare.
- Then for each locality, the sums of the office, school, university and venues were computed.
- For each of this 4 categories, a weight (or penalty) has been defined according to what Yogen Früz considers the most important.
- Frozen Yogurt venues have been weighted with -1, to avoid concurrence.
- Schools have been weighted with 1, since student are good customers.
- Universities have been weighted with 1.5, since students are good customers.
- Offices have been weighted with 2, since employees are even better customers.



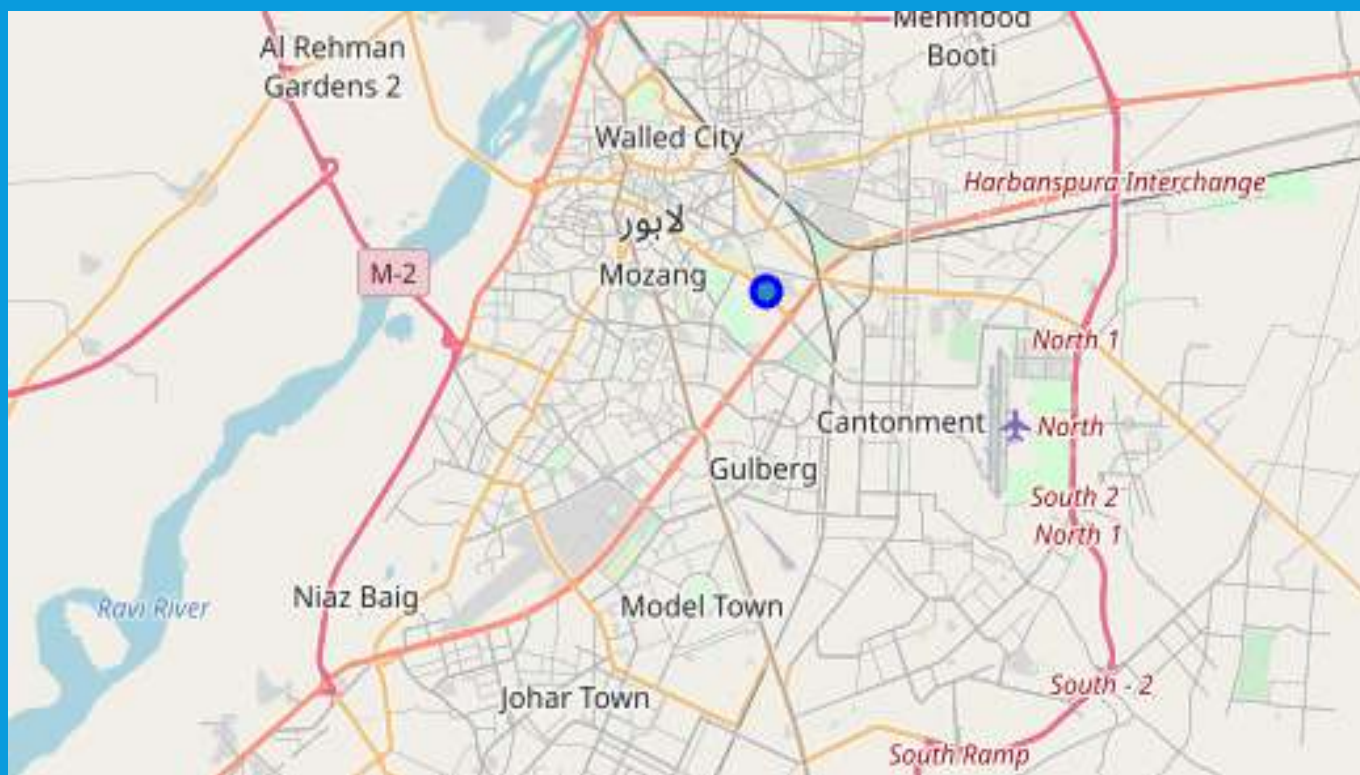
# TOP 5 CITIES OF PAKISTAN

	Province	Capital	City	Country	Code	Latitude	Longitude	Population	Total Population
0	Sindh	admin	Karachi	Pakistan	PK	24.905600	67.082200	12130000	11624219
1	Punjab	admin	Lahore	Pakistan	PK	31.549722	74.343611	6577000	6310888
2	Punjab	minor	Faisalābād	Pakistan	PK	31.416667	73.083333	2617000	2506595
3	Khyber Pakhtunkhwa		Serai	Pakistan	PK	34.739330	72.335655	1860310	1860310
4	Punjab	minor	Rāwalpindi	Pakistan	PK	33.597331	73.047904	1858000	1743101

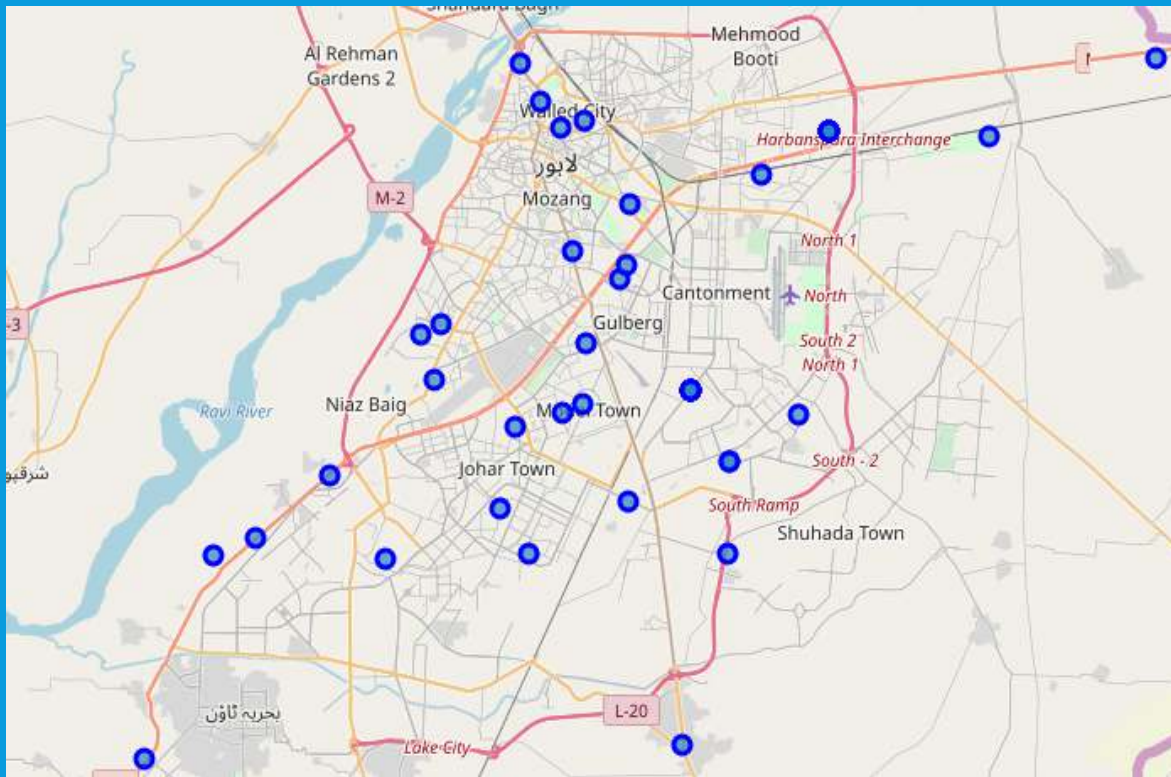




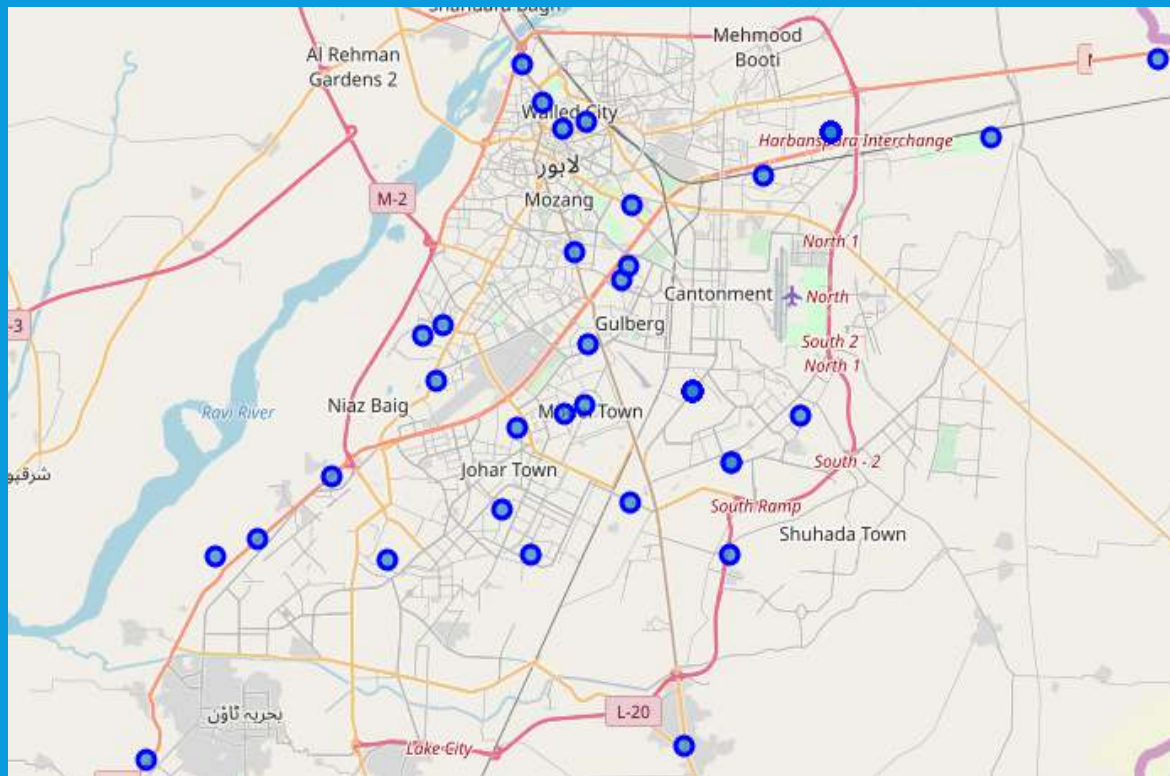
# MAP OF LAHORE



# MAP OF NEIGHBORHOOD/TOWNS/UC'S LAHORE, PAKISTAN



# MAP OF NEIGHBORHOOD/TOWNS/UC'S LAHORE, PAKISTAN





# ANALYSIS OF EACH NEIGHBORHOOD IN LAHORE

Foursquare API to explore the Neighborhood Clusters

Venues in Neighborhood/Towns/UC's Lahore, Pakistan

```
lahore_venues = getNearbyVenues(names=df_data_0['Neighborhood'],  
                                latitudes=df_data_0['Latitude'],  
                                longitudes=df_data_0['Longitude']  
                                )
```

```
print(lahore_venues.shape)  
lahore_venues.head(127)
```

```
(4353, 7)
```

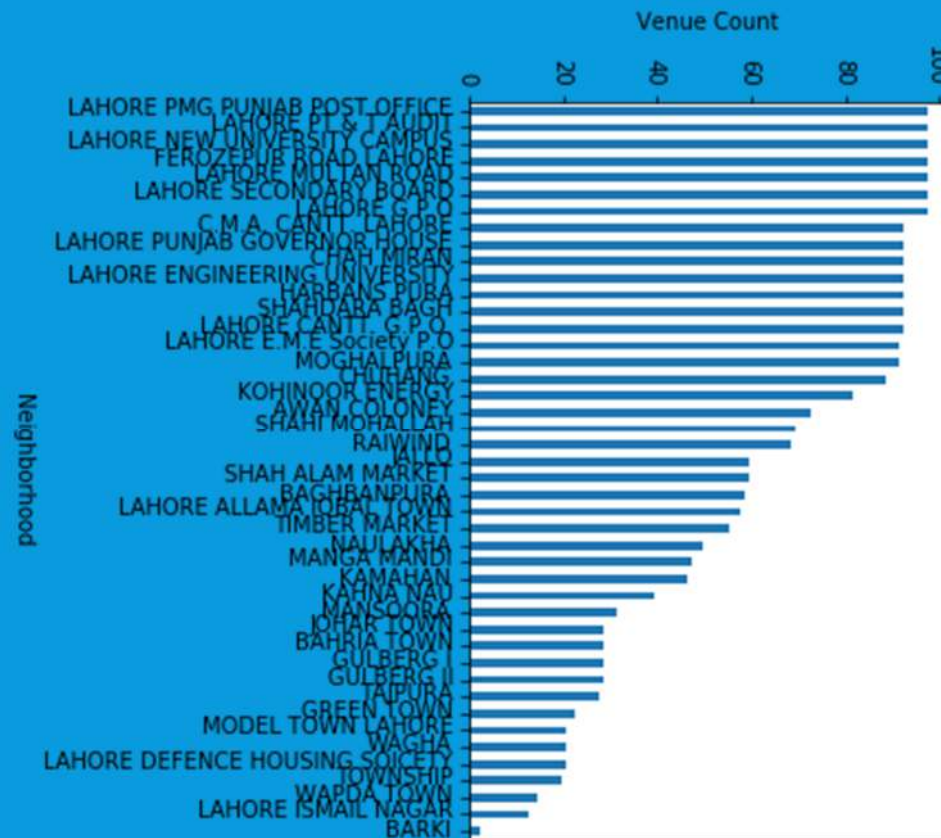


# TOP 10 MOST COMMON VENUES OF LAHORE

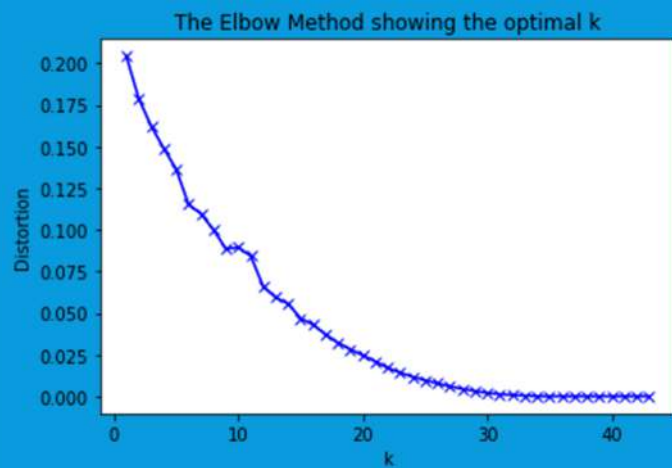
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	AWAN COLONEY	Housing Development	BBQ Joint	Bakery	Gym	Fast Food Restaurant	Clothing Store	Hospital	Food Truck	Gas Station	Bank
1	BAGHBANPURA	Factory	Farm	Office	Industrial Estate	Gas Station	Warehouse	Hospital	Fast Food Restaurant	Housing Development	Restaurant
2	BAHRIA TOWN	Restaurant	Bakery	BBQ Joint	Housing Development	Factory	Yoga Studio	Music Store	Recording Studio	Pizza Place	Coffee Shop
3	BARKI	Gas Station	Train Station	Yoga Studio	Dentist's Office	Event Service	Electronics Store	Donut Shop	Dog Run	Doctor's Office	Distribution Center
4	C.M.A. CANTT. LAHORE	Housing Development	Event Space	Mosque	Building	Hospital	University	Factory	Café	Hookah Bar	Office

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	AWAN COLONEY	Housing Development	BBQ Joint	Bakery	Clothing Store	Fast Food Restaurant	Gym	Food Truck	Dentist's Office	Hospital	Mosque
1	BAGHBANPURA	Factory	Farm	Office	Industrial Estate	Gas Station	Warehouse	Hospital	Fast Food Restaurant	Housing Development	Restaurant
2	BAHRIA TOWN	Bakery	BBQ Joint	Housing Development	Factory	Restaurant	Yoga Studio	Office	Recording Studio	Pizza Place	Coffee Shop
3	BARKI	Gas Station	Train Station	Yoga Studio	Dentist's Office	Event Service	Electronics Store	Donut Shop	Dog Run	Doctor's Office	Distribution Center
4	C.M.A. CANTT. LAHORE	Housing Development	Mosque	Event Space	Hospital	Factory	University	Building	Café	Office	Bakery

# TOP 10 MOST COMMON VENUES OF LAHORE



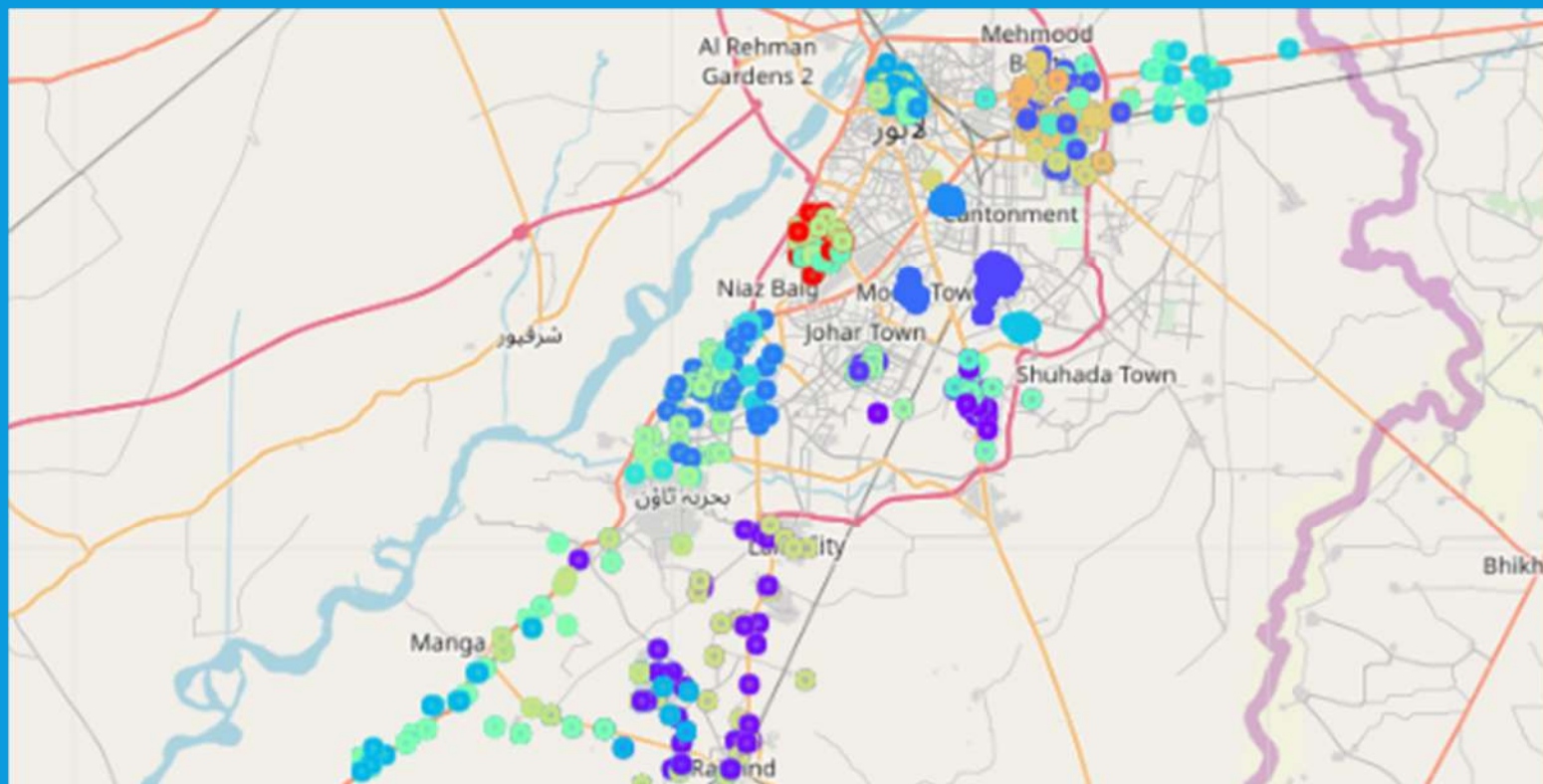
# CLUSTERING



To determine the optimal number of clusters, we have to select the value of  $k$  at the “elbow” ie the point after which the distortion/inertia start decreasing in a linear fashion. Thus for the given data, we conclude that the optimal number of clusters for the data is **10**.

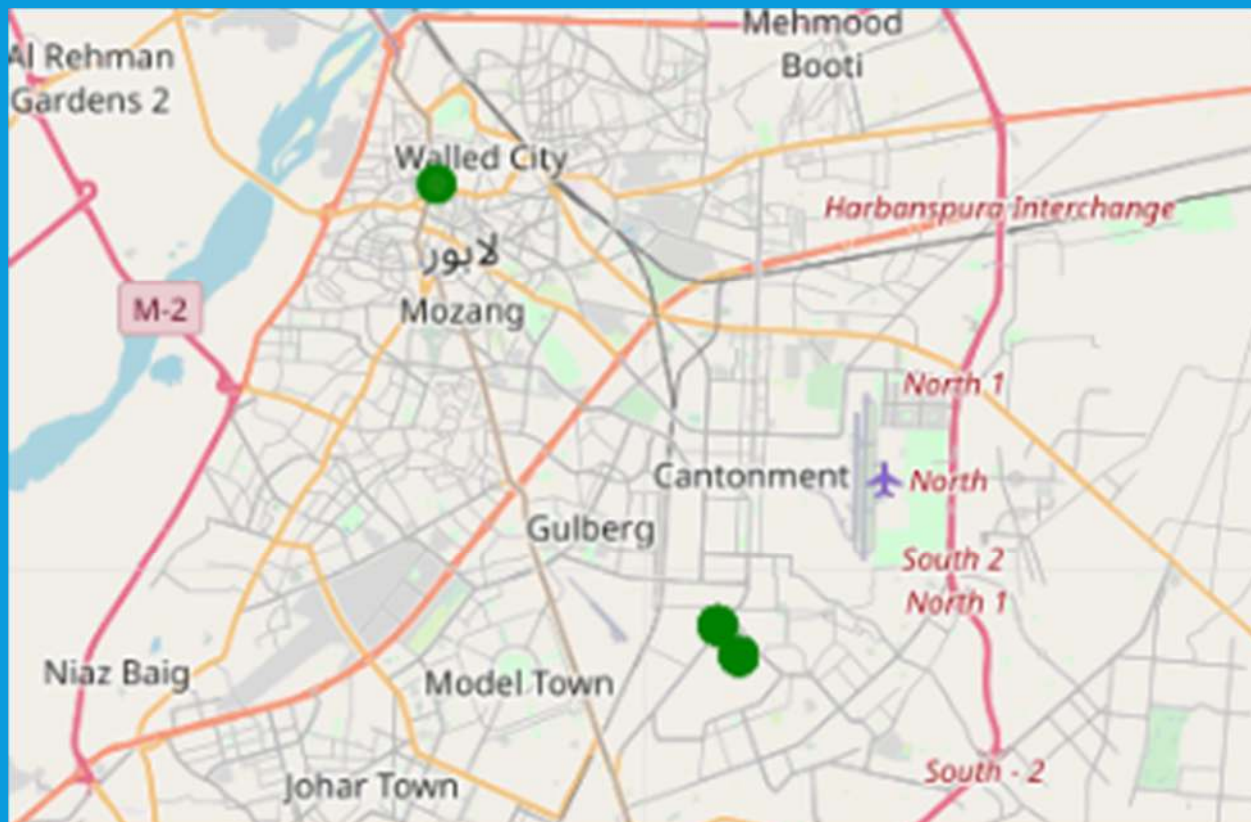


# MAP OF CLUSTERS





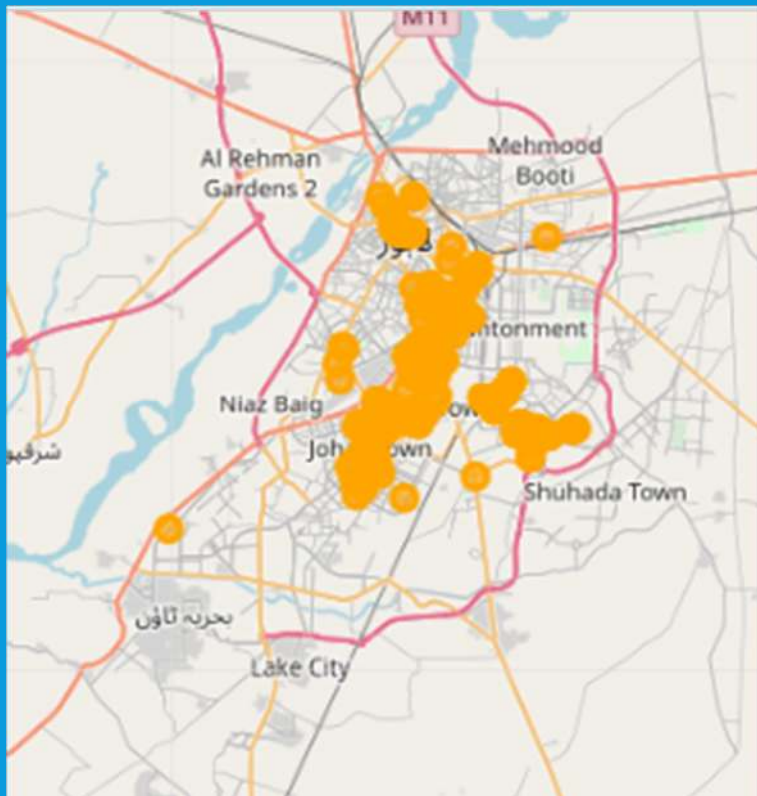
# FROZEN YOGURT VENUE IN LAHORE



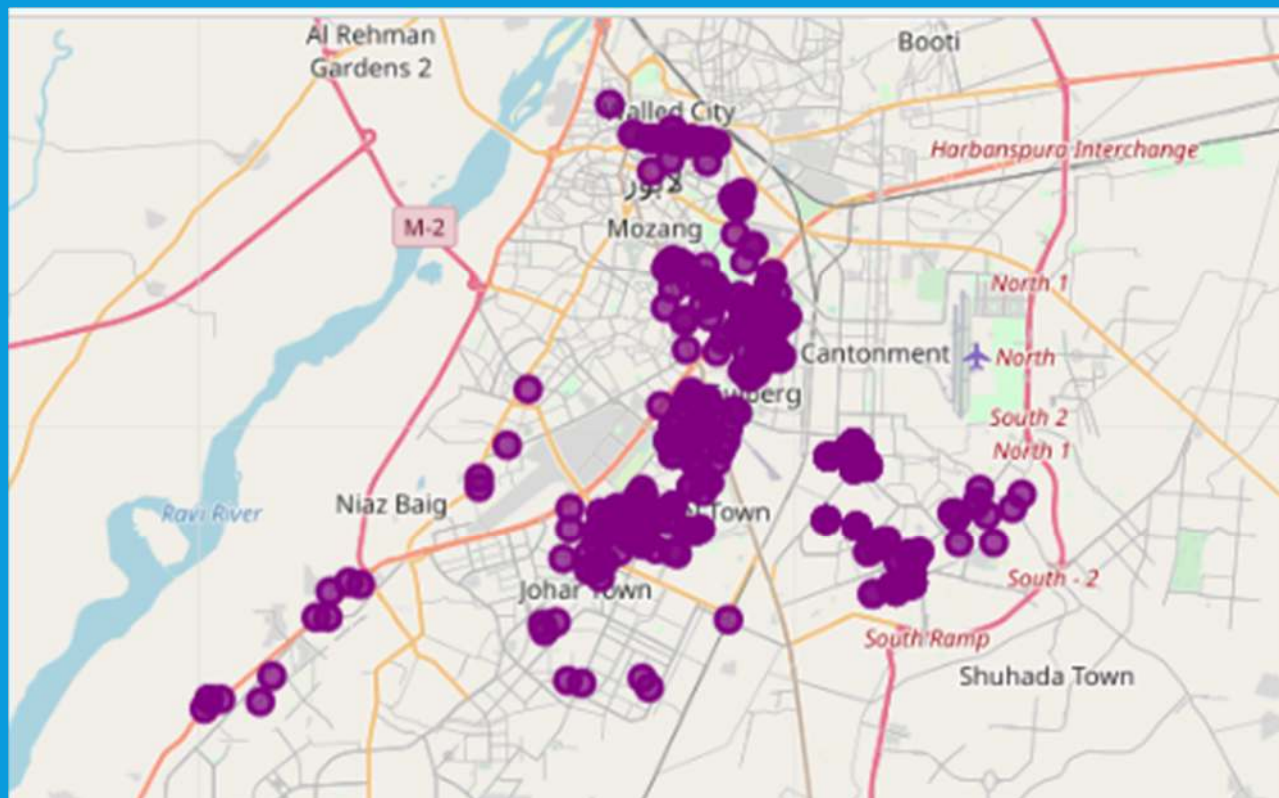
# SCHOOLS & HIGH SCHOOLS VENUE IN LAHORE



# COLLEGES & UNIVERSITIES VENUES IN LAHORE



# OFFICES VENUES IN LAHORE





# WEIGHTED SCORE OF TOP NEIGHBORHOODS /TOWNS / UC'S

Neighborhood	Score
FAISAL TOWN	121.5
GULBERG II	119.5
GULBERG III	116
GULBERG I	112
LAHORE DEFENCE HOUSING SOCIETY	109.5
LAHORE SHADMAN WOMEN MODEL P.O	80
SHAH DARA BAGH	75
CHAH MIRAN	75
TOWNSHIP	57.5
AITCHESON COLLEGE	49.5
SHAH ALAM MARKET	47
MODEL TOWN LAHORE	44.5
LAHORE SECONDARY BOARD	37.5
LAHORE PT & T AUDIT	37.5



# WEIGHTED SCORE OF TOP NEIGHBORHOODS /TOWNS / UC'S

The 5 localities with the best score are

1. “Faisal Town”
2. “Gulberg II”
3. “Gulberg III”,
4. “Gulberg I”
5. “Lahore Defense Housing Authority”

These options maximizes the number of potential customers from offices, schools, and universities and at the same time have not too large competence.



# RECOMMENDATIONS

This analysis can be improved with following extensions:

- Consider more venue categories. For example like “Restaurants” which is also good source for customers. But also like “Ice Cream”, which even if not Frozen Yogurt may be some concurrency if too many.
- In the locality itself, it can also be computer the distance between all the venues in order to find a place with the most number of potential customers.
- Using smaller geographical areas like Neighborhoods could improve the accuracy for the scores.

