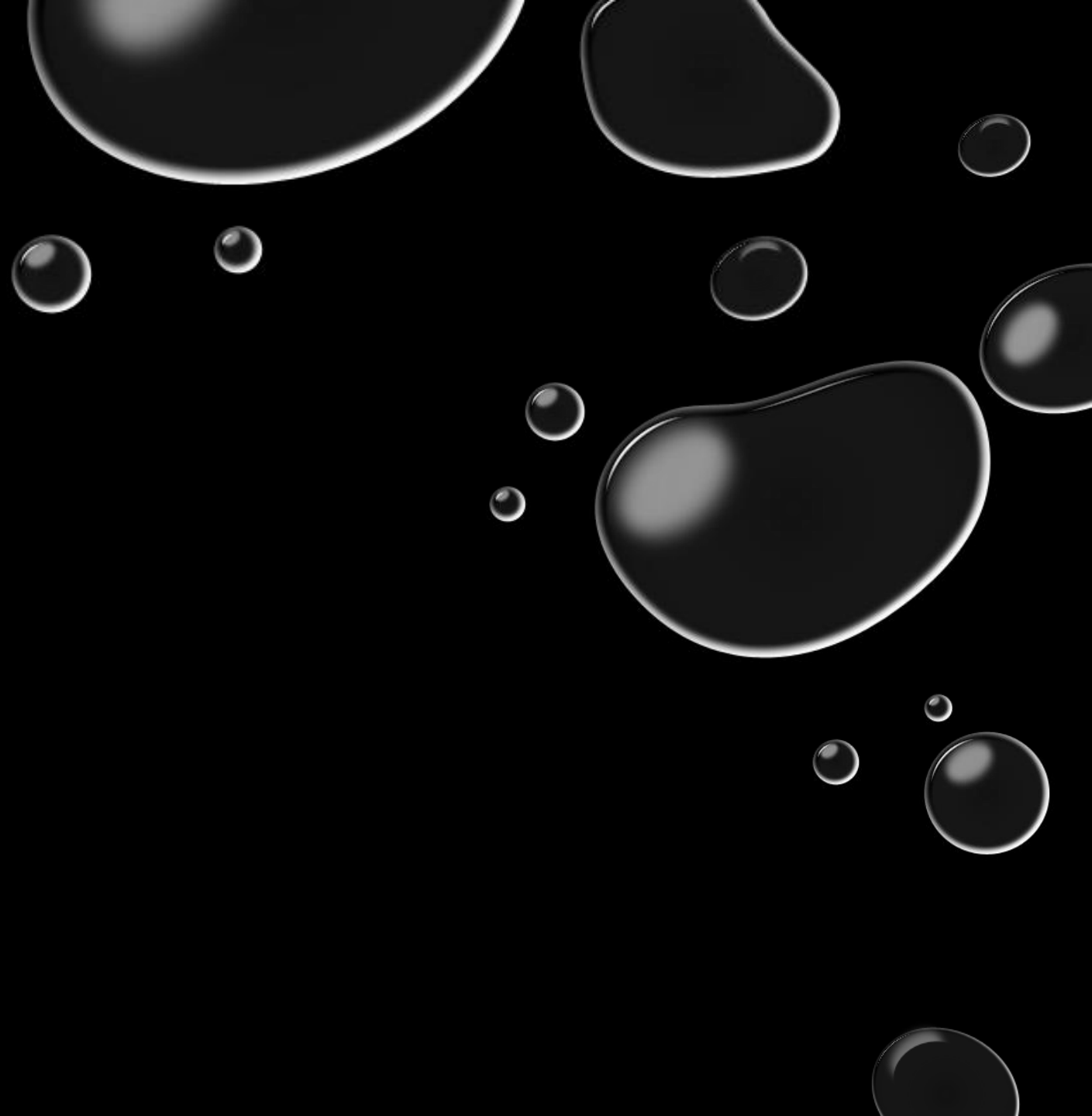


# DATA SCIENCE CAPSTONE

PRIYANK SHAH



# INTRODUCTION

- THIS PROJECT IS TO ANALYZE THE DISTRICTS IN HOUSTON AND FIND ONE FOR A YOUNG COUPLE IN THEIR 20'S.
- THEY ENJOY GOING TO WINE BARS AND YOGA STUDIOS AND WOULD PREFER A LOCATION WHERE THESE ARE IN PLENTY.
- THE OBJECTIVE OF THIS PROJECT IS TO ANALYZE THE DISTRICTS IN HOUSTON AND FIND ONE THAT FITS THIS COUPLES NEEDS.

# DATA SETS

- **DISTRICT DATA SET**

- THIS DATASET IS EXTRACTED FROM THE WIKIPEDIA PAGE AND WE USE BEAUTIFUL SOUP TO SCRAPE THE WEBPAGE.

- **GEOCODING DATA:**

- THIS DATASET IS OBTAINED USING THE GEOCODING API THAT ARE THEN USED IN THE DATAFRAME

- **VENUE:**

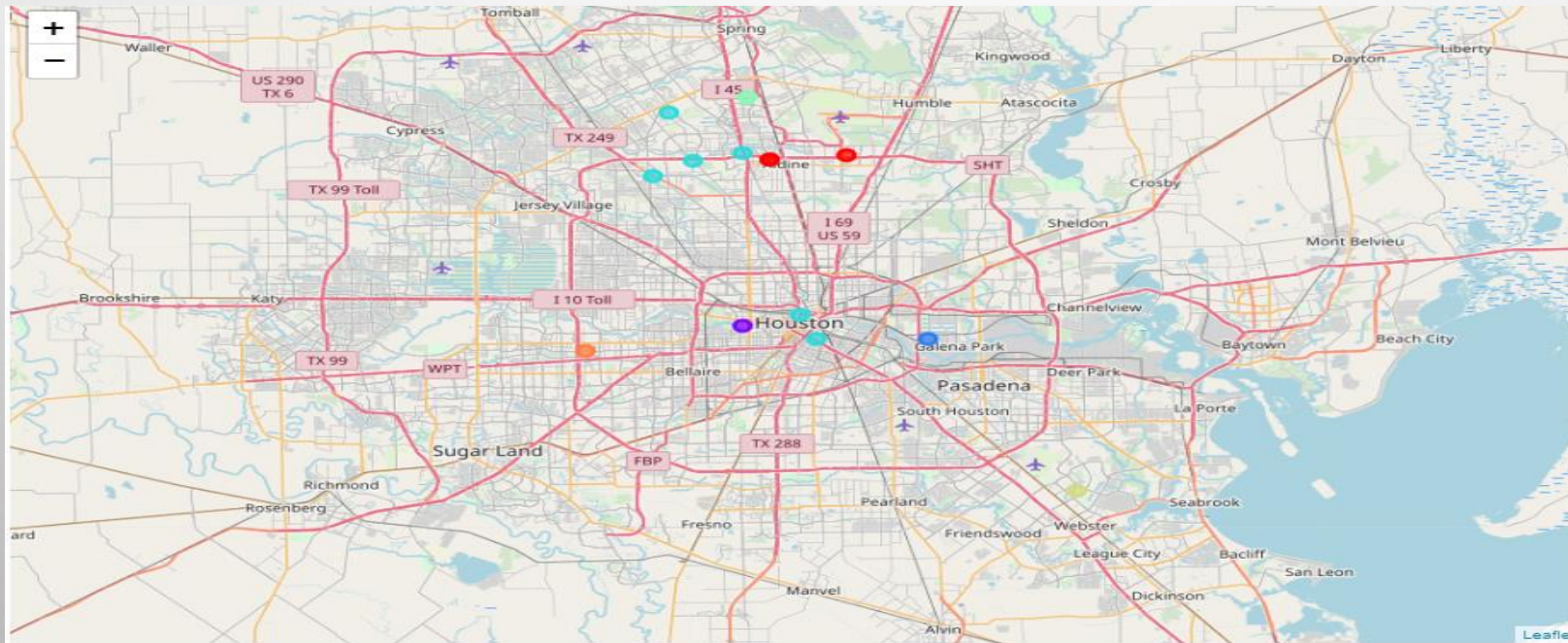
- THIS DATASET IS OBTAINED FROM THE FOURSQUARE API. WE GET INFORMATION ABOUT VENUES AROUND THE HOUSTON DISTRICTS.

# METHODOLOGY

- WEBSITE SCRAPING:
  - THIS INVOLVES USING BEAUTIFUL SOUP TO GET DATA OF THE WIKIPEDIA PAGE FOR HOUSTON:  
'HTTPS://EN.WIKIPEDIA.ORG/WIKI/LIST\_OF\_HOUSTON\_NEIGHBORHOODS'
- GEOCODING:
  - THIS INVOLVES USING THE GEOCODING API TO GET THE LOCATION COORDINATES FOR EACH DISTRICT
- FOURSQUARE API:
  - THIS API WAS USED TO GET ALL THE VENUES IN EACH DISTRICT
- CLUSTERING:
  - K-MEANS CLUSTERING ALGORITHM WAS USED TO CREATE CLUSTERS OF NEIGHBORHOODS. SINCE THERE ARE SO MANY VENUES, K-MEANS WAS THE BEST OPTION
- FOLIUM:
  - THE FOLIUM LIBRARY WAS USED TO DISPLAY DATA ON MAPS.

# RESULTS

- THE DISTRICTS ARE DIVIDED INTO CLUSTERS USING THE K MEANS CLUSTERING ALGORITHM. THEY ARE GROUPED INTO CLUSTERS BASED ON THE VENUES THAT THEY HAVE.



# DISCUSSION

- CLUSTER 2 HAS PLENTY OF WINE BARS AND YOGA STUDIOS
- THE COUPLE SHOULD LOOK TO MOVE INTO THIS AREA

Out[65]:

	Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	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
5	Greater East End Management District	East End, Magnolia Park, and Harrisburg	East of the East Downtown Management District, north of Interstate 45, south of Clinton Drive, and east of Interstate 610	29.741647	-95.256811	2	Wine Bar	Yoga Studio	Fried Chicken Joint	Discount Store	Donut Shop	Fast Food Restaurant	Flower Shop	Food

# CONCLUSION

- THE COUPLE SHOULD LOOK TO MOVE TO THE GREATER EAST END MANAGEMENT DISTRICT BASED ON THEIR LOVE FOR WINE AND YOGA. THIS AREA WOULD FIT THEIR NEEDS THE BEST.