

Capstone Project Tunisia : Monuments and Tourism Agencies

01



Author : Seif Eddine Jaballah

i. Introduction

WHY

As part of my Data Science Capstone Project in these sometimes dark and uncertain times I have decided to consider a case of exploring Tunisia venues and help them improve their customer experience by the means of the following:

When I used to live in Tunisia to do my studies I used to have that idea to visit all the cool places in the country.



Who might be interested by this project?



Venues (mainly monuments in our
case) owners

Marketing agencies

City tourism department

Customers

Let's also not forget that this project is a
pilot and in case of our idea viability it
might be scaled up to any city or venue
type. You can find a link to the Project
Description notebook published on my
GitHub repository.

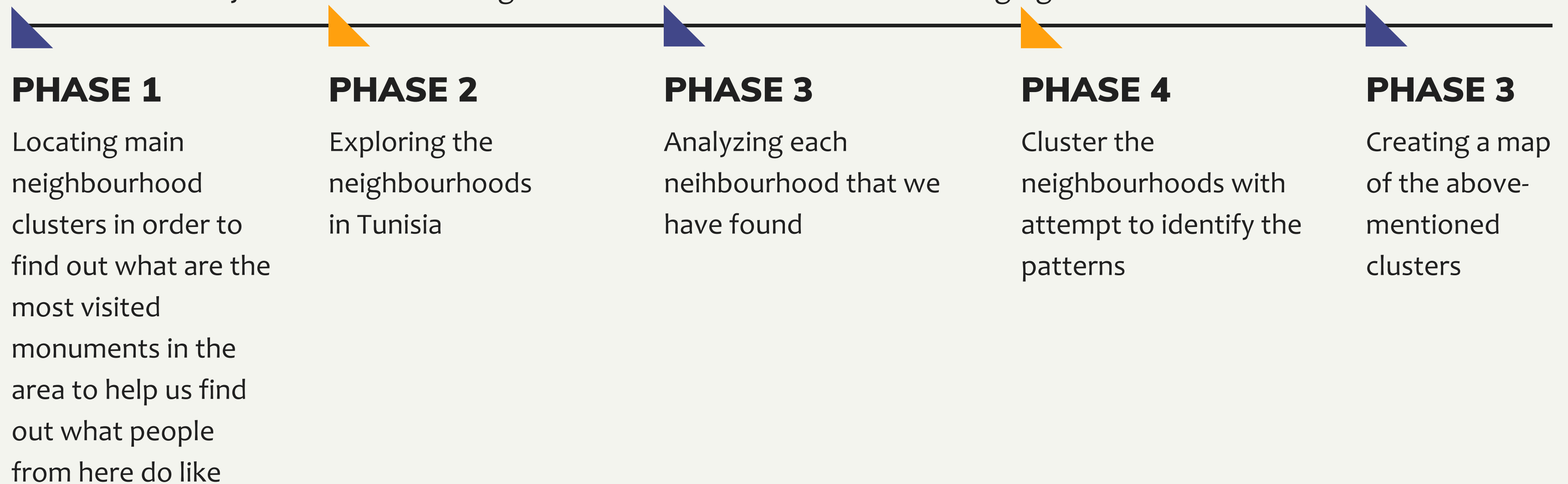
ii. Data Usage

The data we will be using in the Project are:

- Tunisia government open dataset to get neighbourhoods and their locations
- Foursquare open API for fetching the exact location and addresses of the venues
- Additional data from open sources for monuments list extending

iii. Methodology

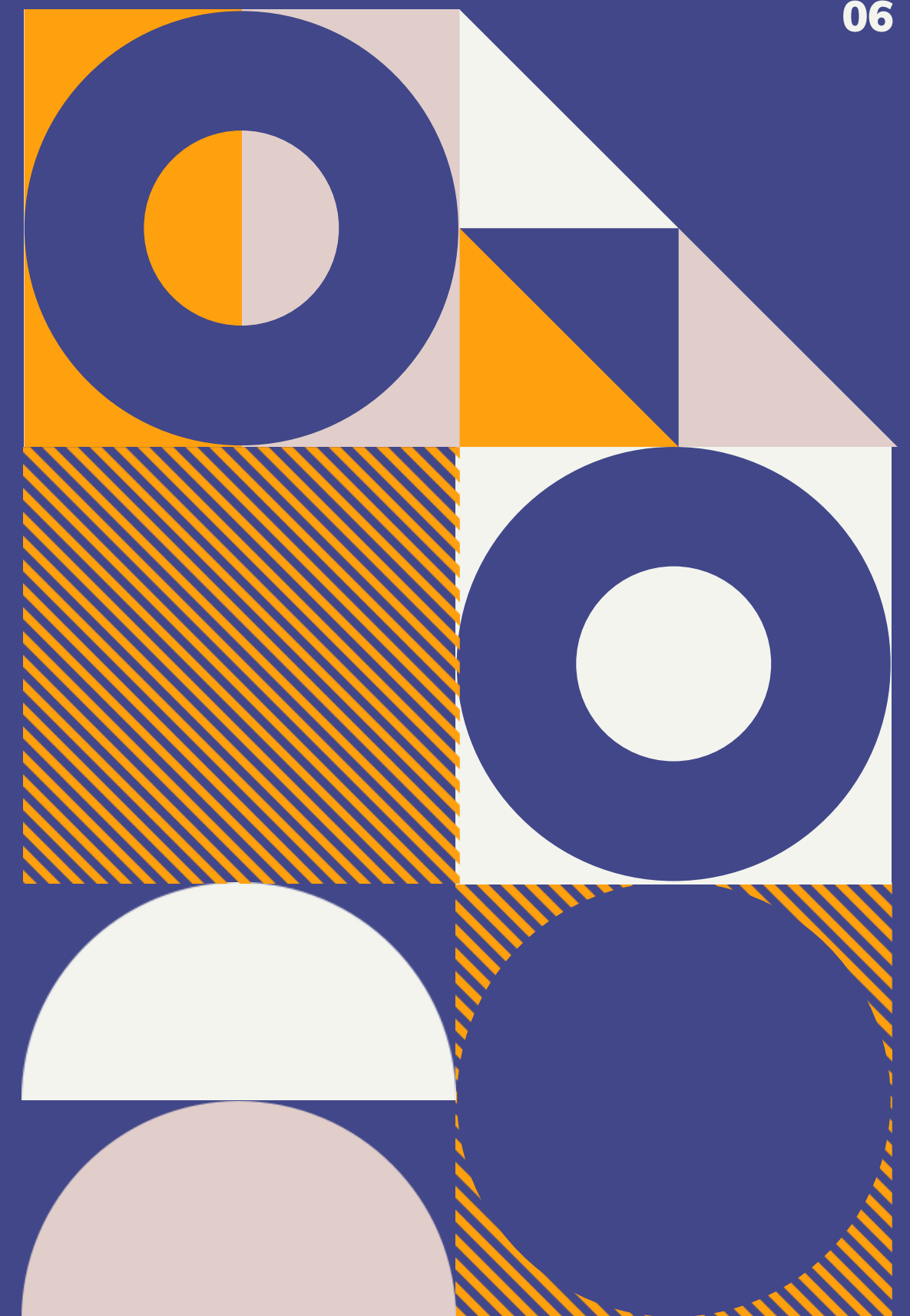
In this project, we will use the Foursquare API to explore neighborhoods in Tunisia. We will use the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. We will use the k-means clustering algorithm to complete this task. Also, we will use the Folium library to visualize the neighborhoods in Tunisia and their emerging clusters.



iv. Results

GENERAL OVERVIEW

- We have fetched the Open Data dataset for tunis
- We have created superimposed map of Tunisia with neighborhoods marked on it
- We have used Foursquare API to categorize the venue for each neighborhood on the basis of 100 venues within the radius of 500 meters
- We used one-hot encoding to explore the categories of the venue by calculating the mean of the frequency of occurrence of each category
- We have also calcuated the frequency for each neighborshood's venue category
- We have obtained 5 clusters for our neighborhoods and top 10 venues using k-means
- We have examined each cluster
- We have used monuments dataframe to create a map and joined the map layer to our existing Tunisia clusters map



iv. Results

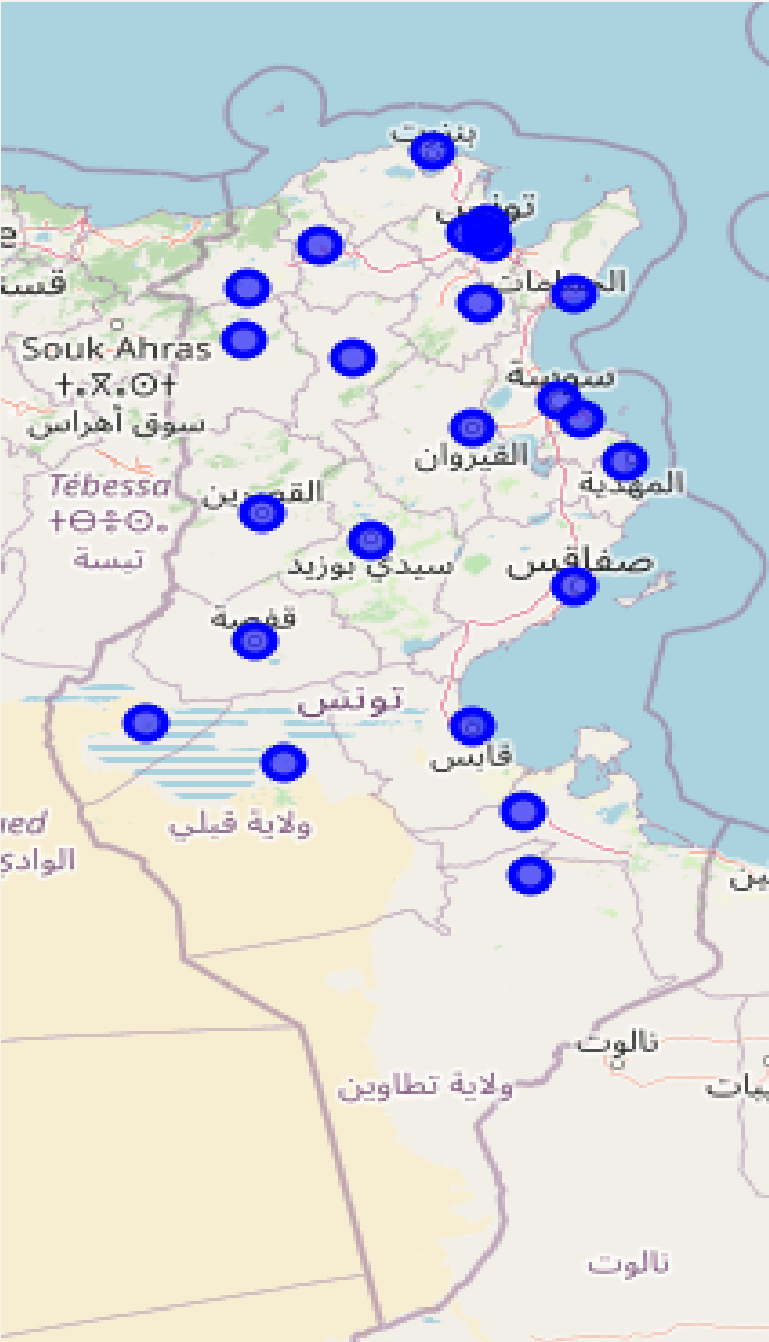


City

We have found all the neighborhoods coordinates and names

	city	lat	lng	country	iso2	admin_name	capital	population	population_proper
0	Tunis	36.8008	10.1800	Tunisia	TN	Tunis	primary	1200000.0	1200000.0
1	Sfax	34.7500	10.7200	Tunisia	TN	Sfax	admin	453050.0	277278.0
2	Sousse	35.8300	10.6250	Tunisia	TN	Sousse	admin	327004.0	164123.0
3	Gabès	33.9004	10.1000	Tunisia	TN	Gabès	admin	219517.0	110075.0
4	Kairouan	35.6804	10.1000	Tunisia	TN	Kairouan	admin	144522.0	119794.0
5	Bizerte	37.2904	9.8550	Tunisia	TN	Bizerte	admin	139843.0	115268.0
6	Gafsa	34.4204	8.7800	Tunisia	TN	Gafsa	admin	126803.0	81232.0
7	Nabeul	36.4603	10.7300	Tunisia	TN	Nabeul	admin	115149.0	115149.0
8	Ariana	36.8667	10.2000	Tunisia	TN	L'Ariana	admin	97687.0	97687.0
9	Kasserine	35.1804	8.8300	Tunisia	TN	Kasserine	admin	81987.0	78158.0
10	Monastir	35.7307	10.7673	Tunisia	TN	Monastir	admin	71546.0	41400.0
11	Tataouine	33.0000	10.4667	Tunisia	TN	Tataouine	admin	62577.0	62577.0
12	Medenine	33.4000	10.4167	Tunisia	TN	Médenine	admin	61705.0	61705.0
13	Béja	36.7304	9.1900	Tunisia	TN	Béja	admin	59567.0	57233.0
14	Jendouba	36.5000	8.7500	Tunisia	TN	Jendouba	admin	51408.0	51408.0
15	El Kef	36.1826	8.7148	Tunisia	TN	Kef	admin	47979.0	36628.0
16	Mahdia	35.4839	11.0409	Tunisia	TN	Mahdia	admin	45977.0	45977.0

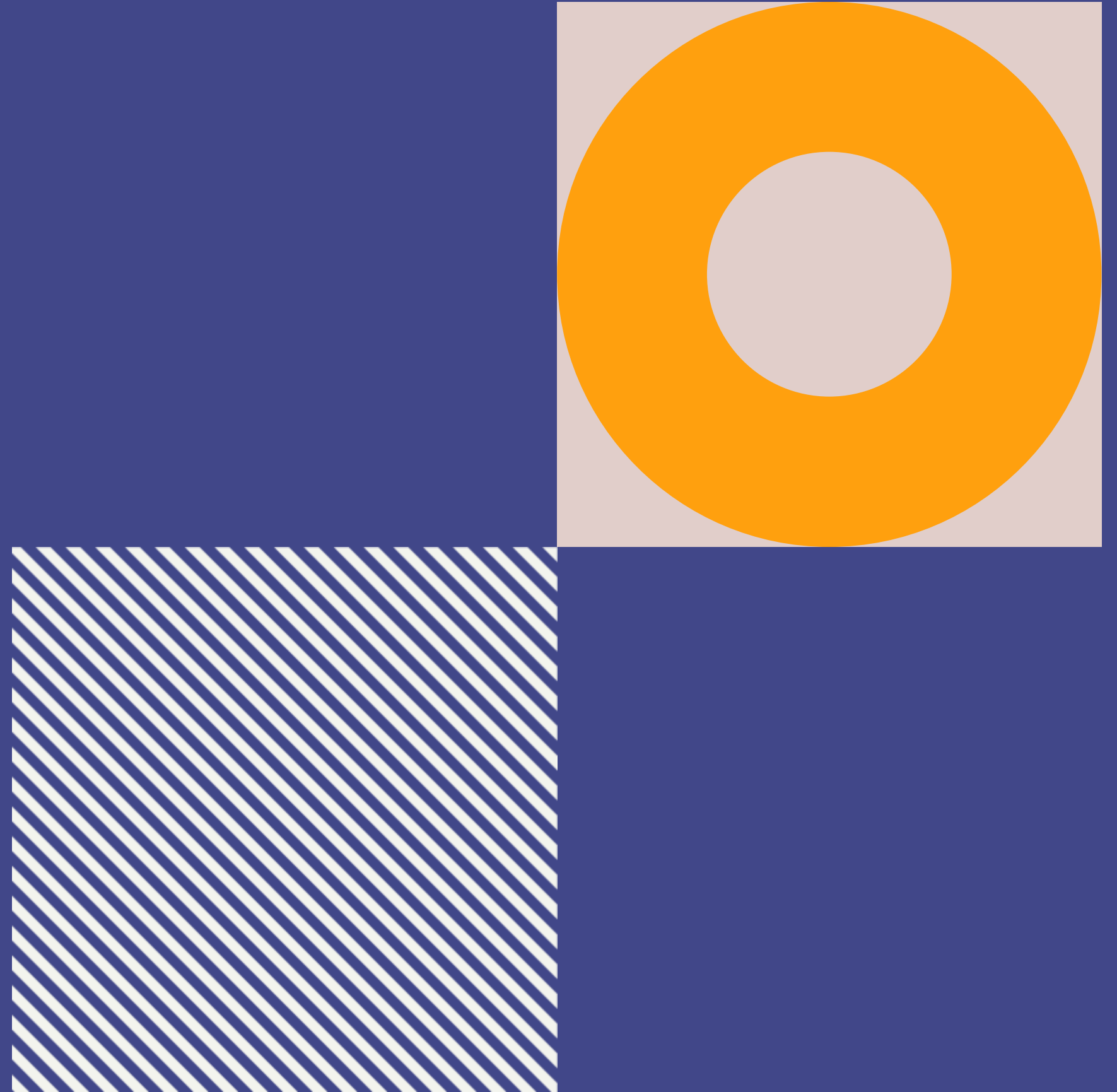
Also we created a map of the neighborhoods



iv. Results

USING FOURSQUARE

Foursquare API helped us to fetch 1246 different venues within the limit of 100 entries with a radius of 500 meters. Grouping venues by neighborhood helps us to better understand how many venues do we have in every arrondissement.

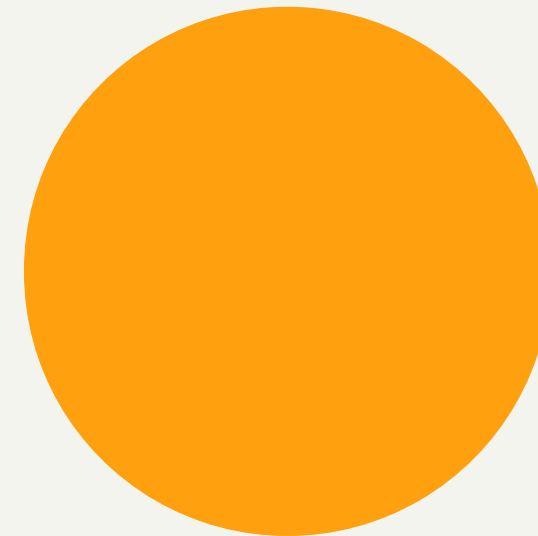


iv. Results



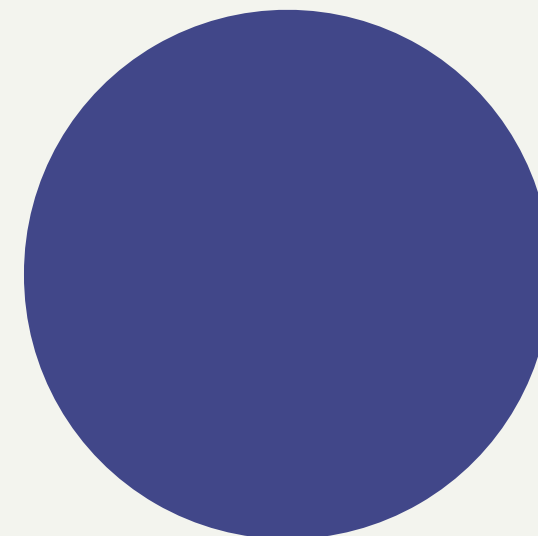
ONE-HOT ENCODING

Using *get_dummies* helps us to group our neighborhoods and explore each one based on the categories of the venues.



MEANS

We also transformed the dataframe to calculate the frequency of occurrence of each category.



CATEGORIES

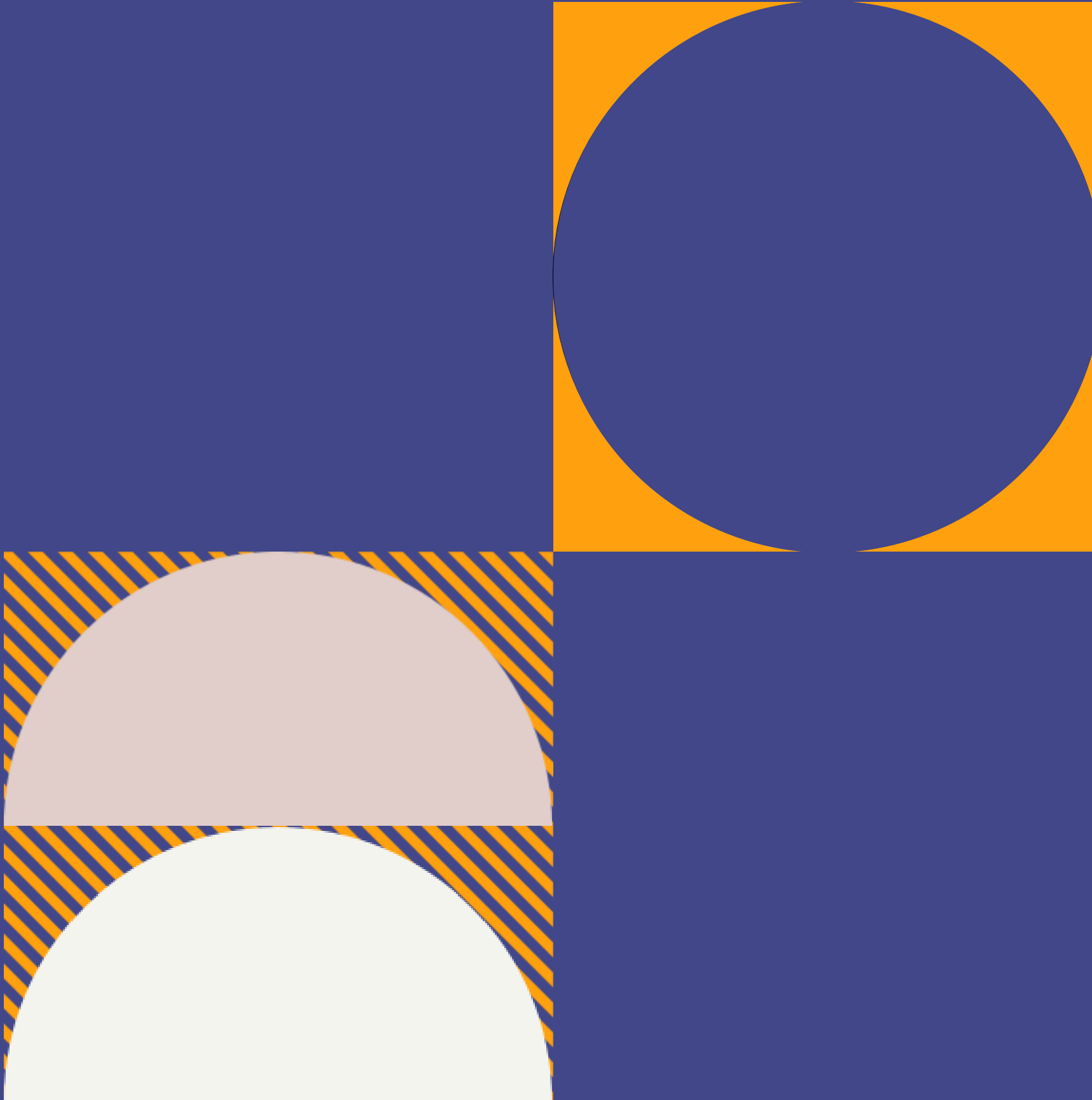
We obtain 198 columns containing each venue category



iv. Results

Top 5

Using grouping we obtain each neighborhood table with its top 5 venues (example left).





TOP 10

Next, we are creating a dataframe with 10 most common venues for the neighborhoods.

	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	Ariana	Café	Coffee Shop	Supermarket	Restaurant	Electronics Store	Cultural Center	Department Store	Dessert Shop	Diner	Women's Store
1	Ben Arous	Café	Coffee Shop	Gym / Fitness Center	Mobile Phone Shop	Big Box Store	Food Truck	Flower Shop	Fish & Chips Shop	Cultural Center	Fast Food Restaurant
2	Bizerte	Grocery Store	Beach	Flower Shop	Fish & Chips Shop	Diner	Women's Store	Cosmetics Shop	Gastropub	Furniture / Home Store	French Restaurant
3	Béja	Soccer Stadium	Women's Store	Coffee Shop	Gastropub	Furniture / Home Store	French Restaurant	Food Truck	Flower Shop	Fish & Chips Shop	Fast Food Restaurant
4	El Kef	Café	Historic Site	Gastropub	Pizza Place	Plaza	Electronics Store	Department Store	Dessert Shop	Diner	Fast Food Restaurant
5	Gafsa	Café	Women's Store	Cosmetics Shop	Grocery Store	Gastropub	Furniture / Home Store	French Restaurant	Food Truck	Flower Shop	Fish & Chips Shop
6	Kairouan	Historic Site	Hotel	Ice Cream Shop	Restaurant	Cosmetics Shop	Furniture / Home Store	French Restaurant	Food Truck	Flower Shop	Fish & Chips Shop

iv. Results

iv. Results

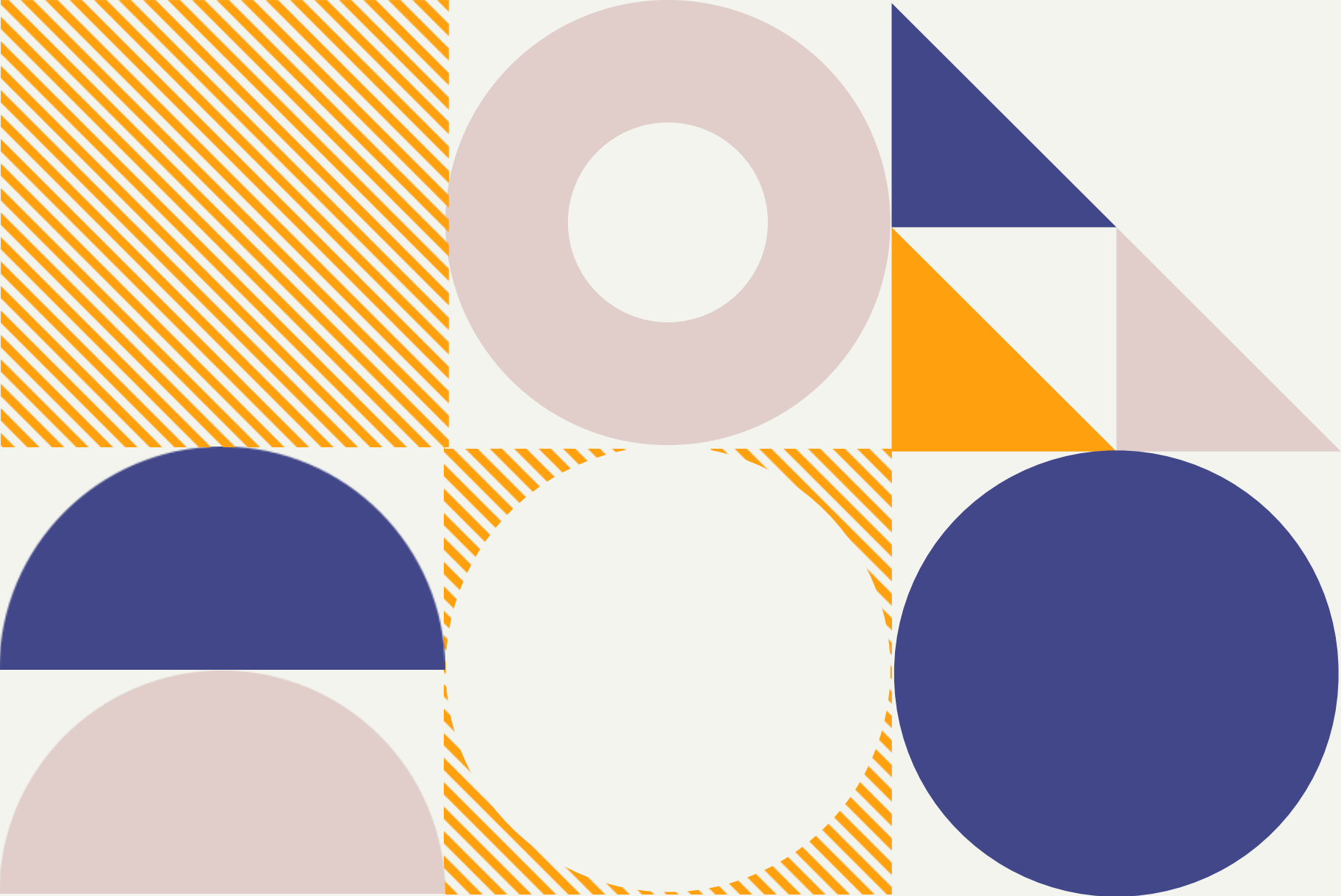
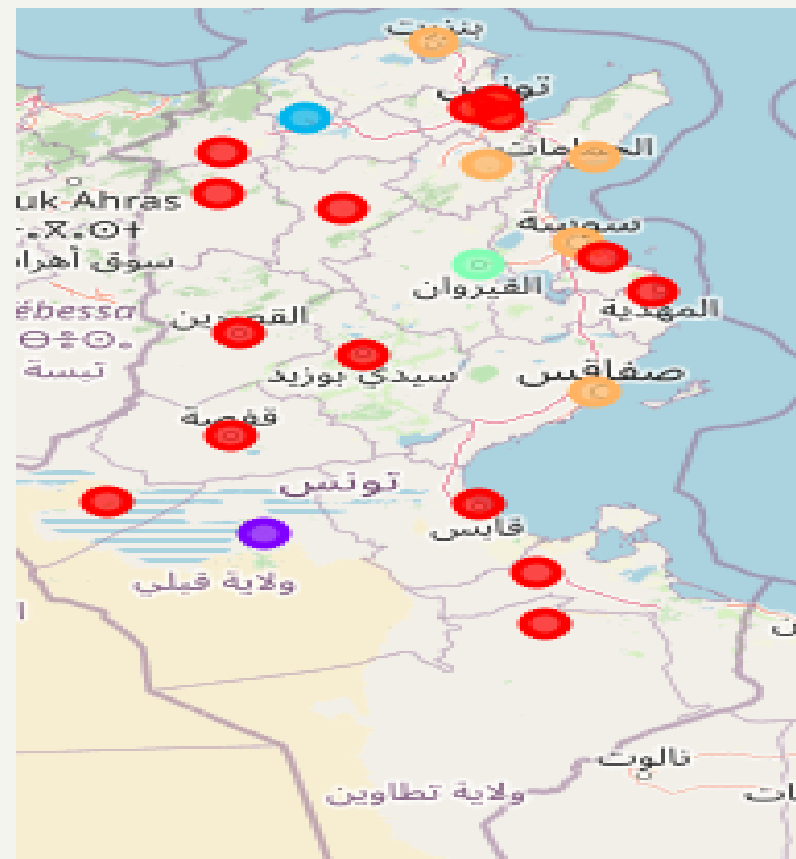
K-Means

Finally, we are running k-means clustering to divide our data to 5 clusters.

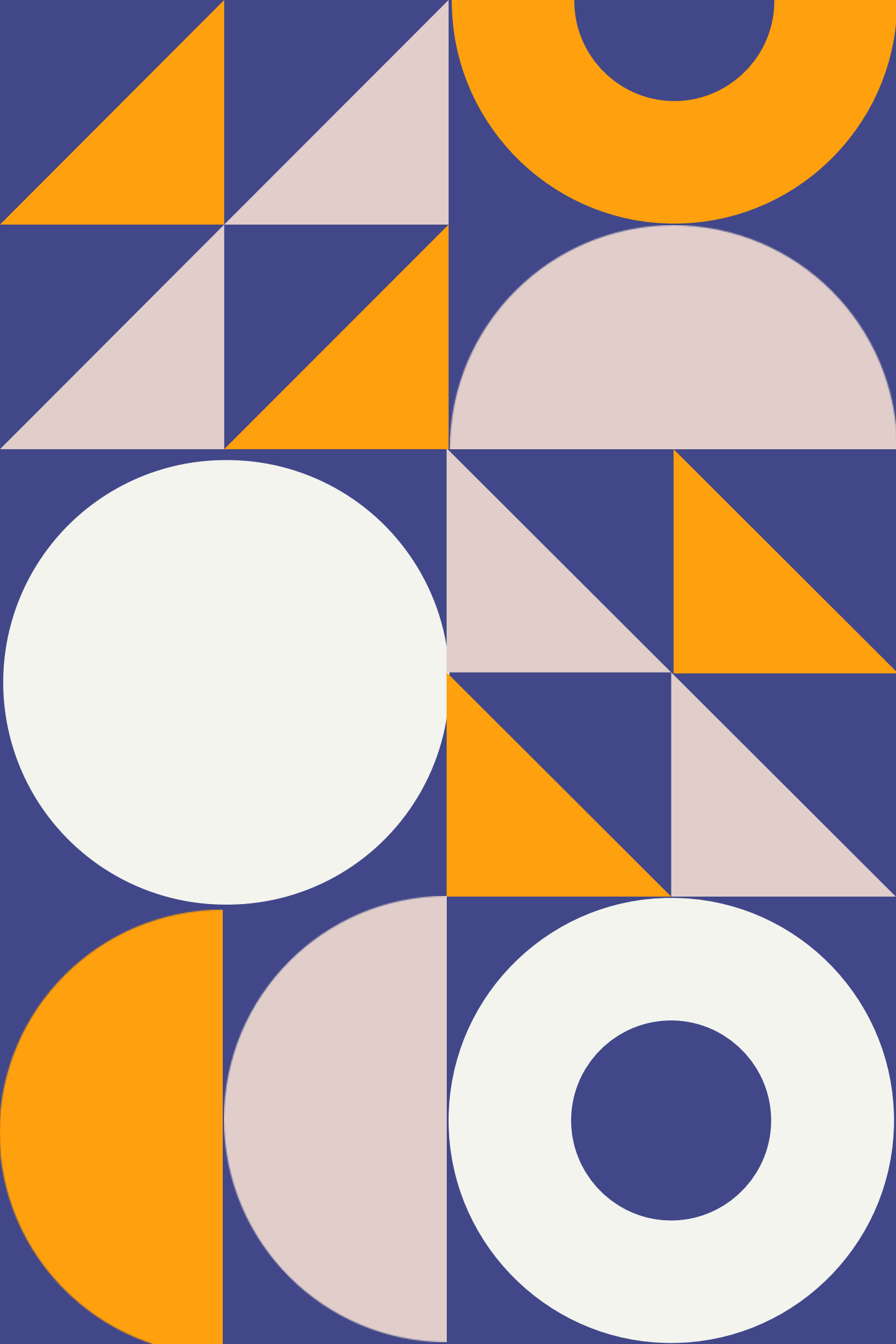


	city	lat	lng	country	iso2	admin_name	capital	population	population_proper	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Tunis	36.8008	10.180	Tunisia	TN	Tunis	primary	1200000.0	1200000.0	4	Café	Fast Food Restaurant	Coffee Shop	Hotel	Bar
1	Sfax	34.7500	10.720	Tunisia	TN	Sfax	admin	453050.0	277278.0	4	Coffee Shop	Furniture / Home Store	Café	Shopping Mall	French Restaurant
2	Sousse	35.8300	10.625	Tunisia	TN	Sousse	admin	327004.0	164123.0	4	Diner	Coffee Shop	Café	Athletics & Sports	Snack Place

iv. Results



We have created cluster map of Tunisia



Clusters

We have examined each of the clusters

CLUSTER 0

CLUSTER 1

CLUSTER 2

CLUSTER 3

CLUSTER 4

5 Neighborhoods

1 Neighborhoods

1 Neighborhoods

1 Neighborhoods

16 Neighborhoods

GENERAL OVERVIEW



v+vi. Observations and Conclusion

WHAT

Whether you are deciding to open a tourist agency our analysis helps us to know what place will be more popular where. We can see that many of the clusters have Monuments as their 1st most popular venue category, except for the 4th cluster.



This project demonstrates the capabilities of combining any dataframe with geographical data using Python. We have used folium to build our maps, and Foursquare API enabled using venues data for our analysis. As the data might not be always precise, I was considering this project as an opportunity to enhance my skills and apply them directly via this practical task. When extending new skills further (which I hope I will be able to do) I will continue to create notebooks using more advanced techniques and statistical methods.
