Analyzing Algeria Metro Stations

Background & Problem Statement

- The Algiers Metro is a rapid transit system serving Algiers, the capital of Algeria. The metro has 19 stations and spanning across 18.5 km migrating passengers daily.

- This project, will look at the neighbourhoods surrounding the metro stations of Algeria and classify them for their primary usage.

- This data is useful for city planners to determine where from and where to people are most likely to travel for work and leisure. This can help plan further extension of the network and find places for new development.

Data & Source

- List of metro stations and their geographical coordinates — scraped from this Wikipedia page - https://en.wikipedia.org/wiki/List_of_Algiers_Metro_stations
- Foursquare API to explore venue types surrounding each station. Foursquare outlines these high-level venue categories with more sub-categories.

  - Arts & Entertainment (4d4b7104d754a06370d81259)
  - *College & University (4d4b7105d754a06372d81259)*
  - *Event (4d4b7105d754a06373d81259)*
  - *Food (4d4b7105d754a06374d81259)*
  - *Nightlife Spot (4d4b7105d754a06376d81259)*
  - *Outdoors & Recreation (4d4b7105d754a06377d81259)*
  - *Professional & Other Places (4d4b7105d754a06375d81259)*
  - *Residence (4e67e38e036454776db1fb3a)*
  - *Shop & Service (4d4b7105d754a06378d81259)*
  - *Travel & Transport (4d4b7105d754a06379d81259)*
- Venues will be fetched for each category in a 1000m radius around each station.

Methodology

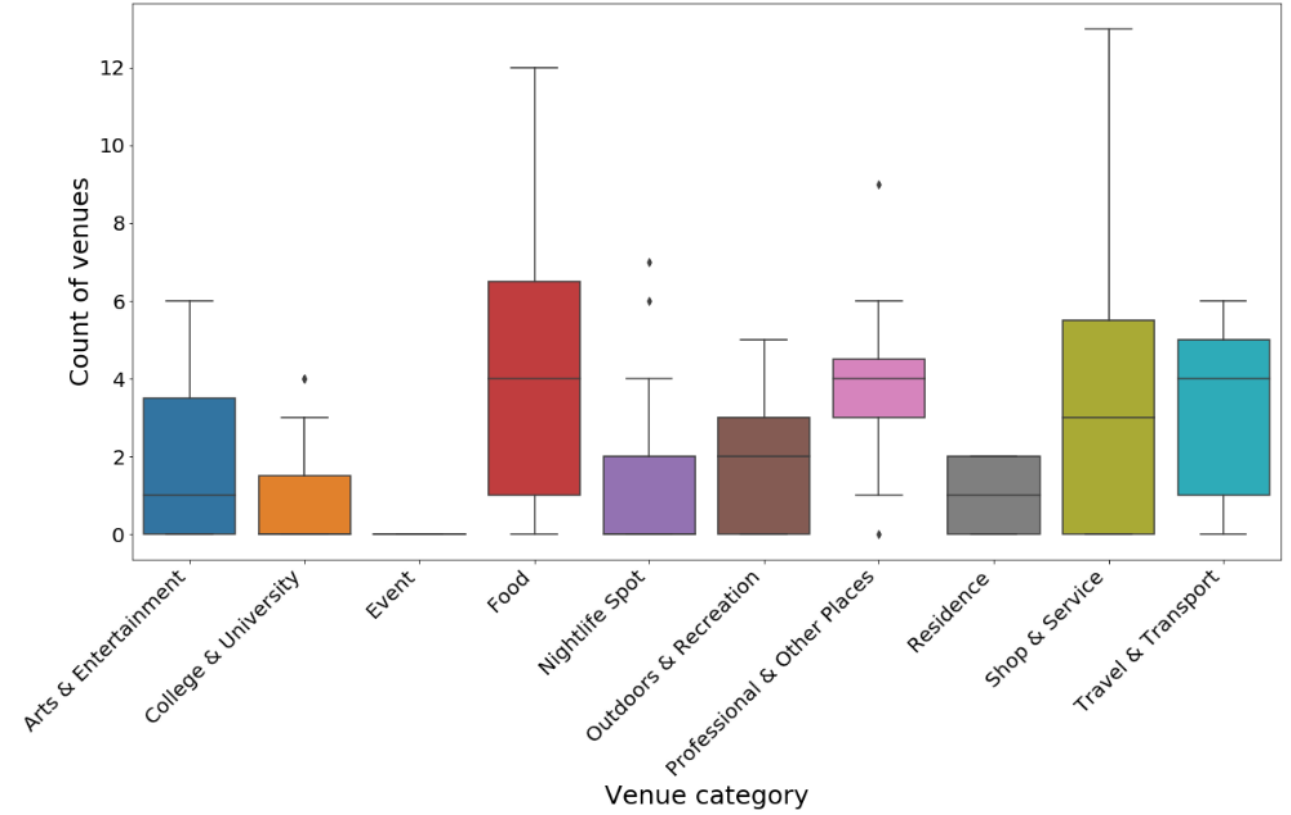
- Using Foursquare explore API returns count of venues of each category in a 1000m radius for each station. Below is a sample.

	Station Name	Coordinates	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
0	Place des Martyrs	36.78556,3.06222	5	1	0	3	0	4	6	1	3	4
1	Ali Boumendjel	36.77917,3.05806	3	3	0	10	2	3	4	1	6	5
2	Tafourah - Grande Poste	36.77194,3.05806	4	3	0	12	2	3	3	0	11	6
3	Khelifa Boukhalfa	36.76639,3.05361	4	4	0	9	7	2	4	2	13	6
4	1er Mai	36.76056,3.05528	4	4	0	6	6	4	4	2	10	5

- **Amirouche** station has the highest number of Professional & Other Places (9) while **El Harrach Centre** station has the lowest number of Professional & Other Places (0).

Exploratory Data Analysis & Data Cleaning

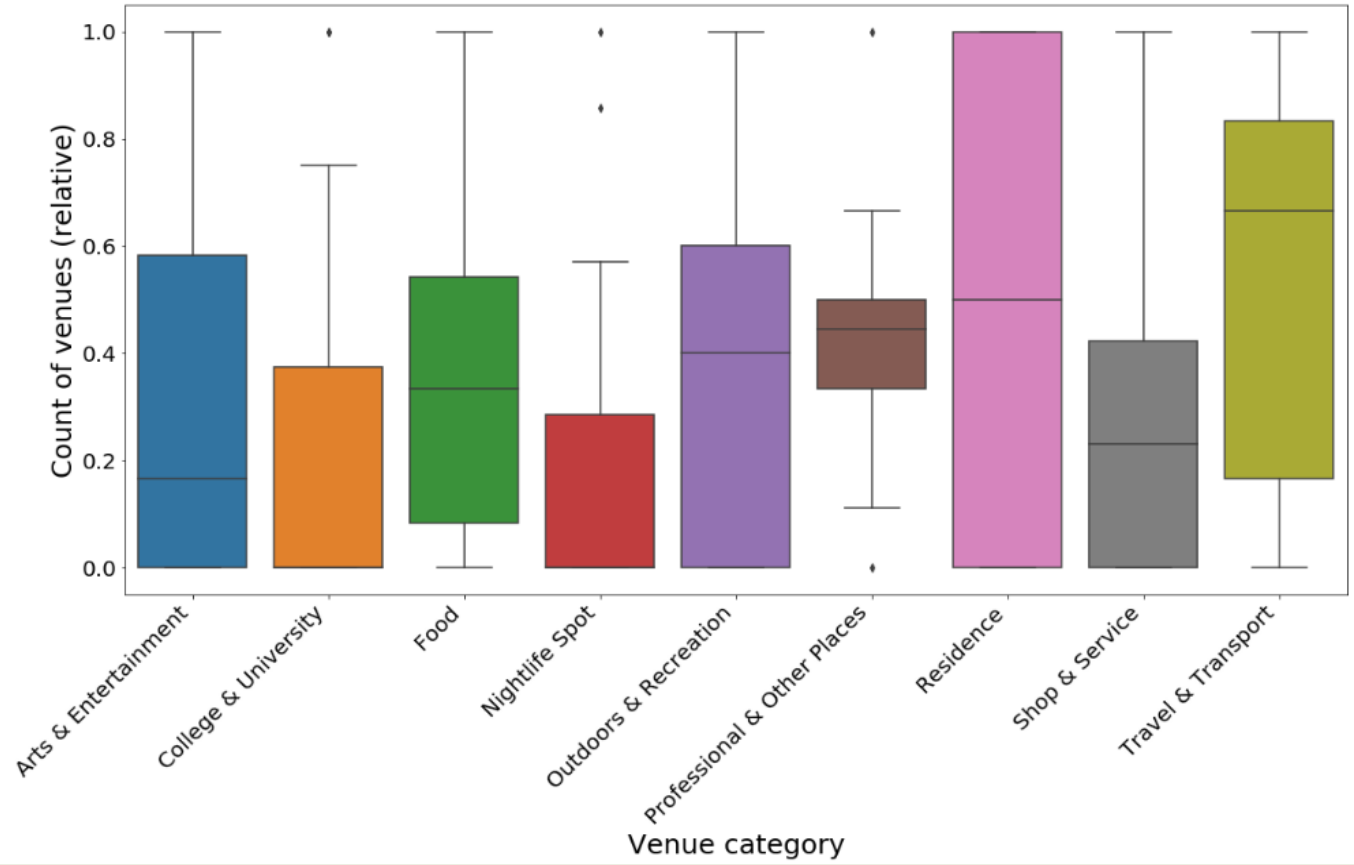
- Boxplots of number of venues in each category



- We can see that the most frequent venue categories are **Food** and **Shop & Service.** Event has very little data, so will discard it.

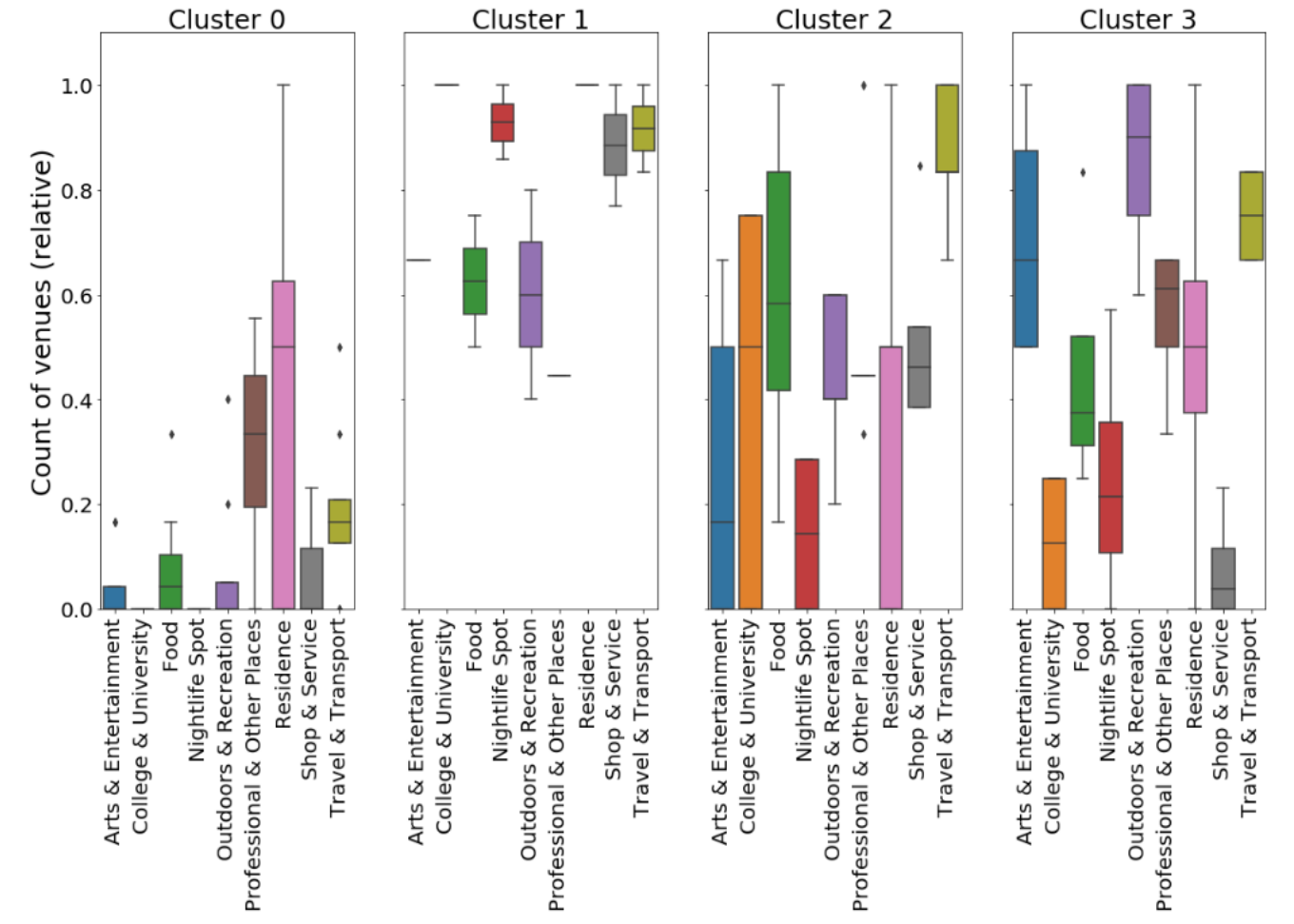
Data Preparation

- Normalize the data using min-max scaling. Scale count of venues from 0 to 1 where 0 is the lowest value in a set and 1 is highest.
- Boxplots of number of venues in each category (scaled)



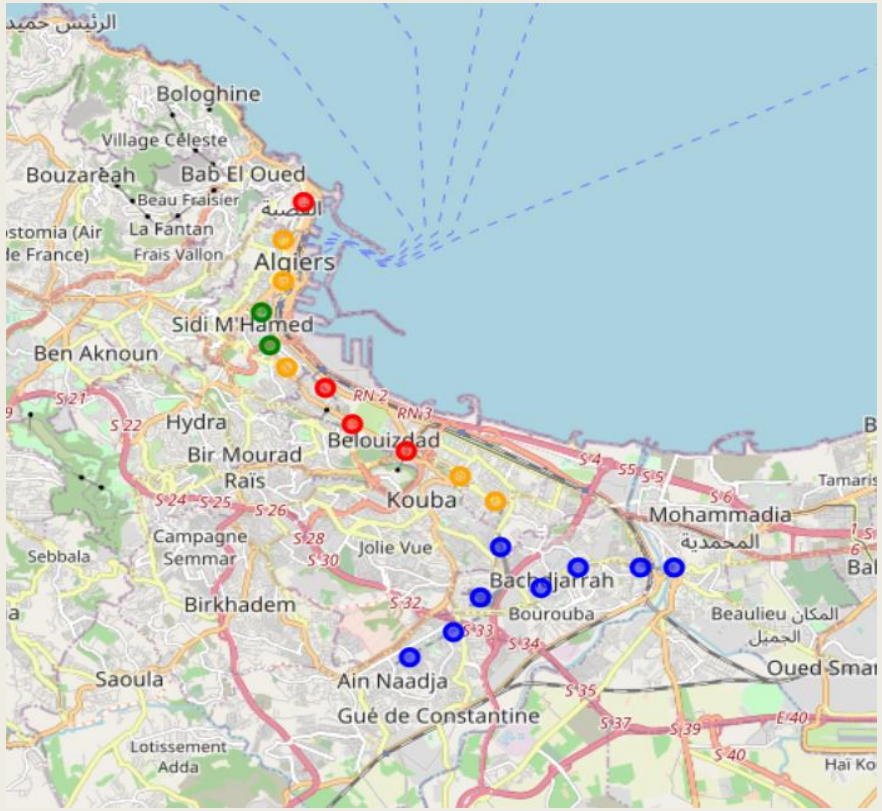
Clustering

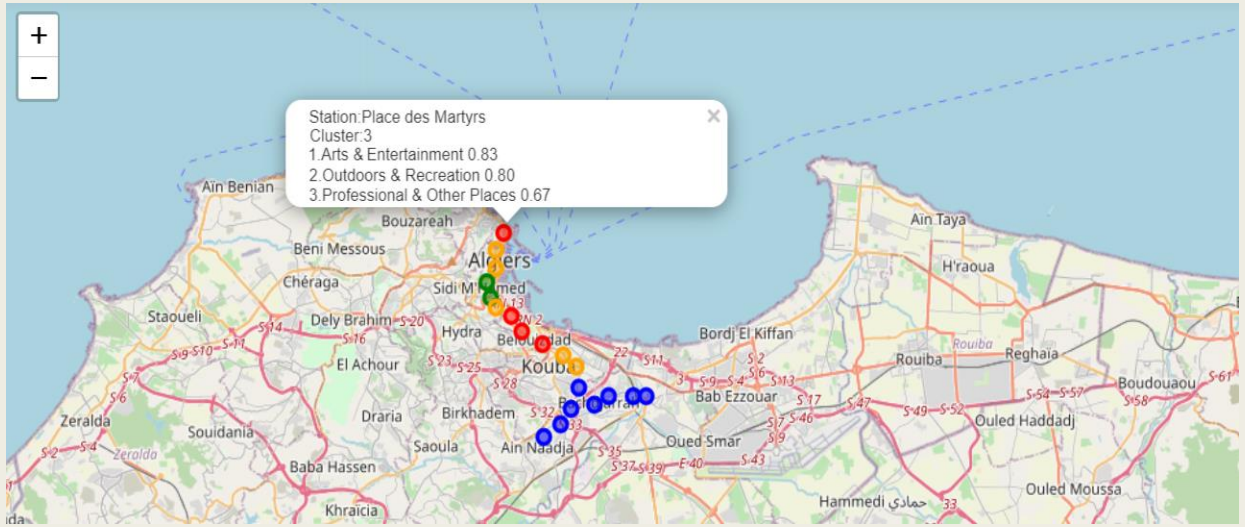
- Using k-means clustering.
- With k=4, visualizing the clusters profiles using boxplots



Plotting Clusters on Map

- Cluster 0 is Blue, 1 is Green, 2 is Yellow and 3 is Red.
- For each station displaying top 3 venue categories and their 0 to 1 scores for that category.





Results

- Here is how we can characterize the clusters by looking at venue scores:
  - *Cluster 0 (Blue) scores high on Residence, Professional & other places. This appears to be residential area as well as business part of the city.*
  - *Cluster 1 (Green) scores high on nightlife, shops & services and travel & transport.*
  - *Cluster 2 (Orange) scores high on college & university, food and travel & transport. This appears to be the student hub of the city.*
  - *Cluster 3 (Red) scores high on outdoor & recreational and art & entertainment.*

- Plotting the clusters on a map shows us that:
  - *Most of the stations are located along the coastal line which appears to be old and developed part of the city.*
  - *Cluster 0 is the most inside of the coastal line than other clusters indicating it as primary upcoming areas for residential and business development.*
  - *The highest number of venues are in the Food and Shop & Service categories.*

Discussion & Conclusion

- The clustering in this analysis has been done basis venues obtained from Foursquare data which is not all encompassing. It does not consider aspects such as footfall in venues, venue's size which can considerably impact the results. Thus, this analysis is far from being conclusory. Furthermore, the results could also potentially vary if we use some other clustering techniques like DBSCAN.

- Finally, to conclude this project we can say that Foursquare data has certainly given some preliminary information and brief insight into city's development. This data can be combined with other sources to provide more accurate results.