

CAPSTONE PROJECT

MOROCCAN CITIES CLUSTERING

Ridouane OULHIQ, Data Scientist

Introduction

Cluster Moroccan cities based on venues
Detect the cities with a low competition level in terms of coffees

Data Description

Wikipedia + Beautiful Soup

	City	Population	Region
0	Casablanca	3359818	Casablanca-Settat
1	Fez	1112072	Fès-Meknès
2	Tangier	947952	Tanger-Tétouan-Al Hoceima
3	Marrakesh	928850	Marrakesh-Safi
4	Salé	890403	Rabat-Salé-Kénitra

geopy

	City	Population	name	categories	lat	lng
0	Casablanca	3359818	Casa Jose	Tapas Restaurant	33.597823	-7.615341
1	Casablanca	3359818	Six PM	Hotel Bar	33.595940	-7.618684
2	Casablanca	3359818	Sofitel Casablanca Tour Blanche	Hotel	33.598251	-7.613960
3	Casablanca	3359818	Le Rouget de l'Isle	French Restaurant	33.592591	-7.622857
4	Casablanca	3359818	Hvatt Regency Casablanca	Hotel	33.596195	-7.618708

Foursquare API

Methodology

- Get the list of Moroccan cities
- Convert the data to numeric
- Get latitude and longitude from address
- Define Foursquare Credentials and Version
- Get the venues of each city
- Finalize the dataset by regrouping the population the venues and the total venues of each city
- Cluster cities based on all the venues
- Cluster the cities based on coffees

Results and Discussion



Conclusion

