Data analytics for opening a new Moroccan restaurant in Paris, France

Author: Naoufal NAIM
Course: Applied Data Science Capstone – IBM

- Introduction
- Data Description
- Methodology
- Results
- Discussion
- Conclusion
- Perspectives
- References

Introduction

- Paris is the French capital and most interesting city of France
- 2,138,551 residents in 105 square kilometers [1]
- Paris was the fifth most expensive city in the world, after Singapore, Osaka, Hong Kong, and New York in 2019 [2].
- **Business problem:** Finding the ideal districts from 20 districts in Paris to open a Moroccan restaurant.

Data Description

- District data for Paris including names and location data [3]

- The most famous venues and their relative locations in the radius 500 meters from the center of each neighborhood -> Using **Fousquare API**
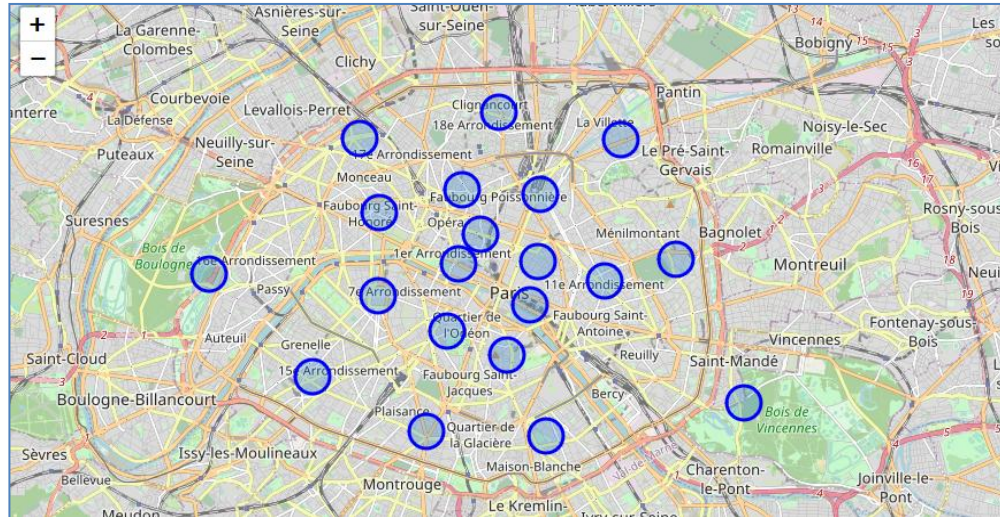
Methodology

- Initial data wrangling and cleaning.
- Result: Neighborhood data in Paris, France

	Arrondissement_Num	Neighborhood	Arrondissement_Name	Latitude	Longitude
0	3	Temple	3eme Ardt	48.862872	2.360001
1	19	Buttes-Chaumont	19eme Ardt	48.887076	2.384821
2	14	Observatoire	14eme Ardt	48.829245	2.326542
3	10	Entrepot	10eme Ardt	48.876130	2.360728
4	12	Reuilly	12eme Ardt	48.834974	2.421325
5	16	Passy	16eme Ardt	48.860392	2.261971
6	11	Popincourt	11eme Ardt	48.859059	2.380058

Methodology

- Get the latitude and longitude values of Paris using **geopy** tool.

- Plot the map of neighborhoods in Paris

Methodology

- Using **Foursquare API** to generate to the most famous venues and the relative locations.

- Result : The top-100 venues in radius 500 meters from the centre of each district

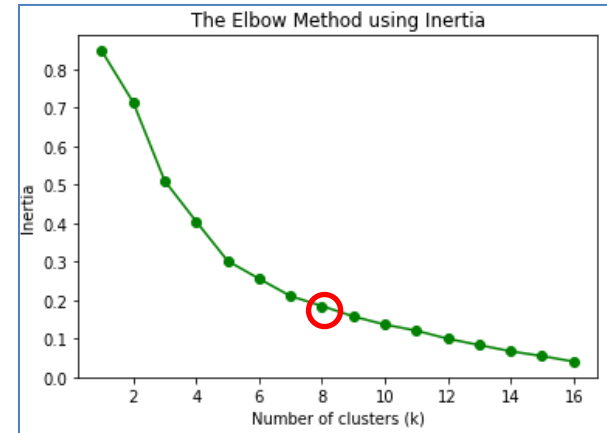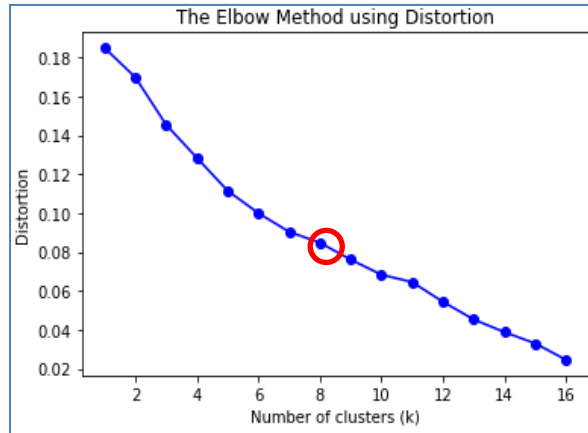|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Temple	48.862872	2.360001	Mmmozza	48.863910	2.360591	Sandwich Place
1	Temple	48.862872	2.360001	Chez Alain Miam Miam	48.862369	2.361950	Sandwich Place
2	Temple	48.862872	2.360001	Fromagerie Jouannault	48.862947	2.362530	Cheese Shop
3	Temple	48.862872	2.360001	Square du Temple	48.864475	2.360816	Park
4	Temple	48.862872	2.360001	Marché des Enfants Rouges	48.862806	2.361996	Farmers Market

Methodology

- Analyzing, exploring the data and finding the optimal number of clusters for K-Means clustering using the **Elbow method**.

The Elbow Method using Distortion

The Elbow Method using Inertia
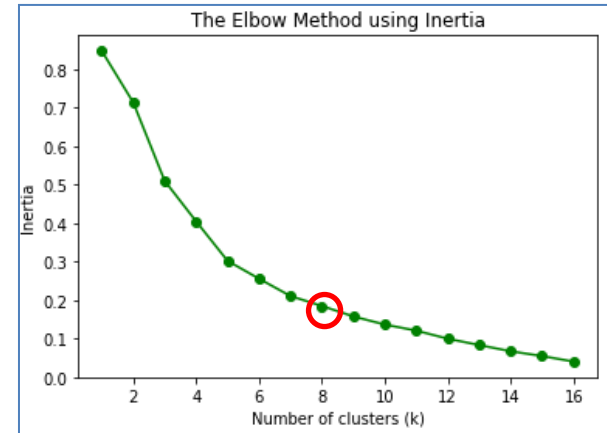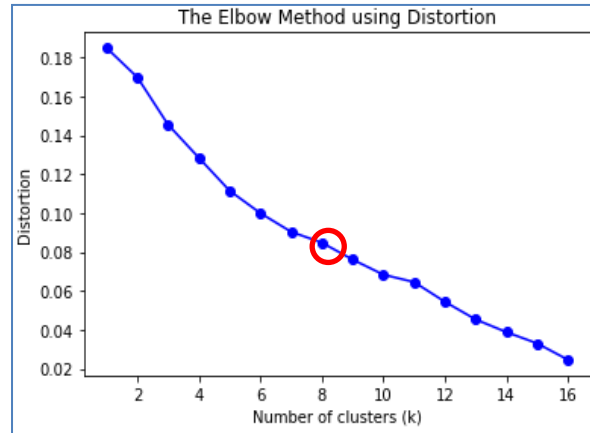
Methodology

- The charts of the Elbow Method display the distortion or the inertia with the number of clusters.

- The optimal number of clusters for K-Means is 8.

The Elbow Method using Distortion

The Elbow Method using Inertia

Methodology

- Using K-Means to cluster the neighborhoods in the city of Paris based on the top 25 venues for each neighborhood district.
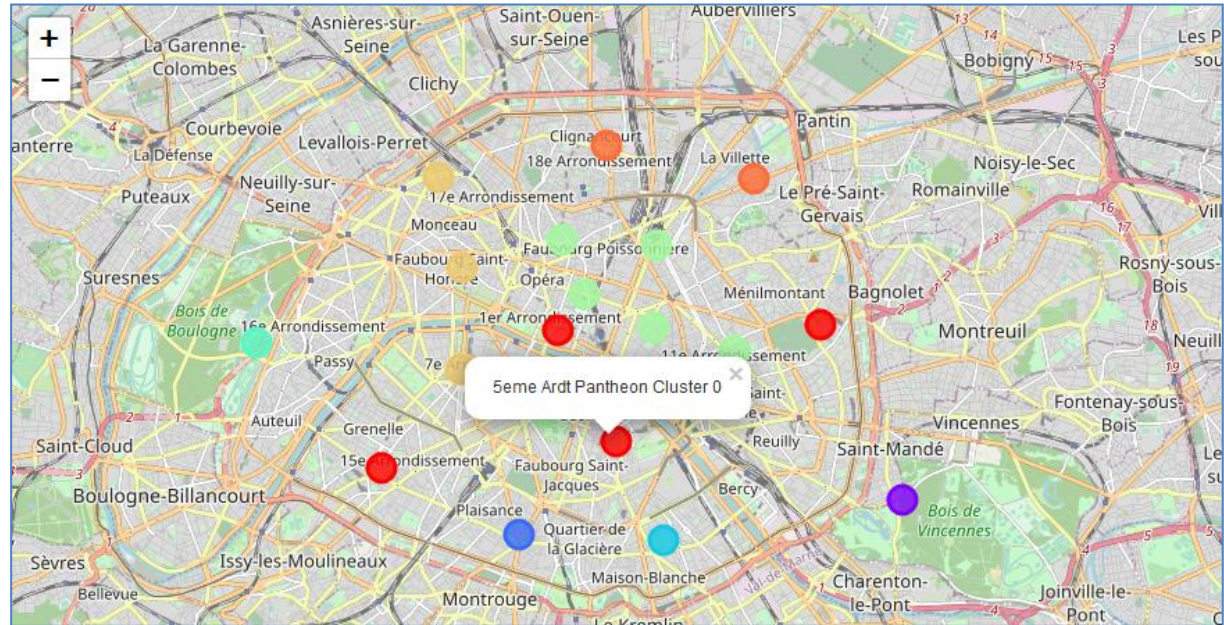
	Arrondissement_Num	Neighborhood	Arrondissement_Name	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	3	Temple	3eme Ardt	48.862872	2.360001	5	French Restaurant	Japanese Restaurant	Gourmet Shop	Cocktail Bar
1	19	Buttes-Chaumont	19eme Ardt	48.887076	2.384821	7	French Restaurant	Bar	Pizza Place	Restaurant
2	14	Observatoire	14eme Ardt	48.829245	2.326542	2	French Restaurant	Hotel	Bakery	Brasserie
3	10	Entrepot	10eme Ardt	48.876130	2.360728	5	French Restaurant	Hotel	Bistro	Coffee Shop
4	12	Reuilly	12eme Ardt	48.834974	2.421325	1	Zoo Exhibit	Bike Rental / Bike Share	Monument / Landmark	Supermarket
5	16	Passy	16eme Ardt	48.860392	2.261971	4	Plaza	Park	Lake	Pool

Results

- Visualizing geographic details of Paris and its district neighborhoods which are superimposed on top using **folium.**
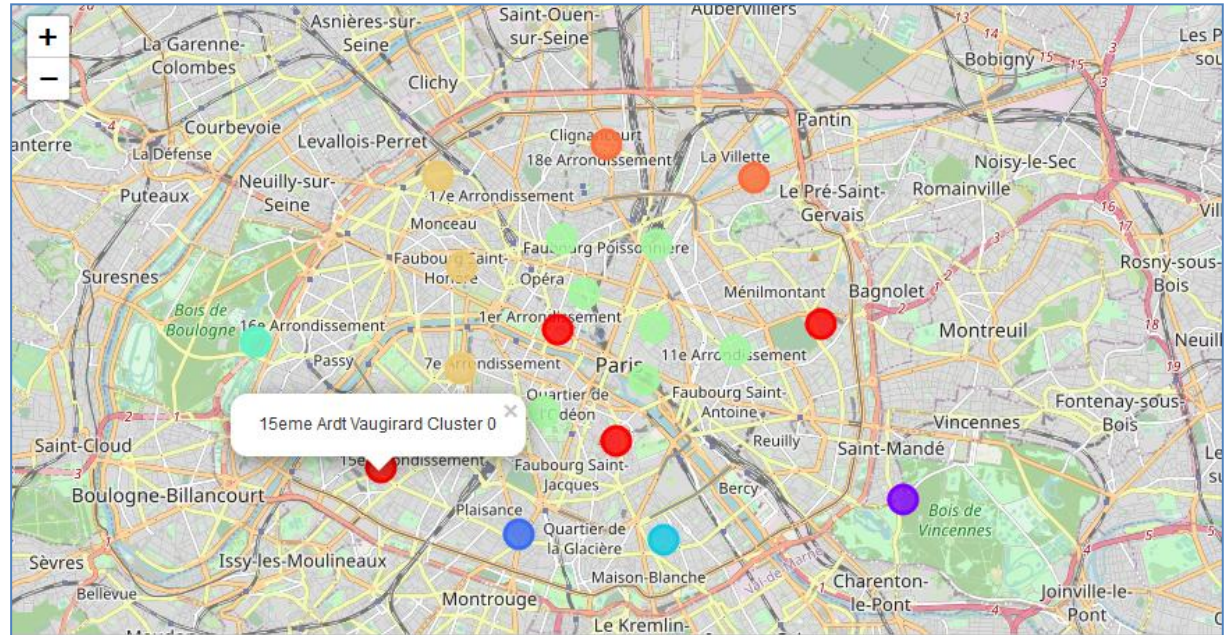
Results

- The 5th arrondissement where there is Moroccan restaurants

Results

- The 15th arrondissement where there is Moroccan restaurants

Results

- Moroccan restaurants exists in cluster 1

11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue	15th Most Common Venue	16th Most Common Venue	17th Most Common Venue	18th Most Common Venue	19th Most Common Venue	20th Most Common Venue	21th Most Common Venue	22th Most Common Venue	23th Most Common Venue
Coffee Shop	Thai Restaurant	Udon Restaurant	Garden	Dessert Shop	Exhibit	Boutique	Museum	Cosmetics Shop	Pedestrian Plaza	Perfume Shop	Furniture / Home Store	General College & University
Greek Restaurant	Garden	Pizza Place	Bistro	Historic Site	Bar	Wine Bar	Creperie	Museum	Hostel	Moroccan Restaurant	Portuguese Restaurant	Chocolate Shop
Bookstore	Latin American Restaurant	Fast Food Restaurant	Street Art	Market	Mexican Restaurant	Roof Deck	Restaurant	Diner	Movie Theater	Pizza Place	Korean Restaurant	Brasserie
Brasserie	Lebanese Restaurant	Cocktail Bar	Sushi Restaurant	Bistro	Persian Restaurant	Sports Bar	Moroccan Restaurant	Shoe Store	Scandinavian Restaurant	Irish Pub	Plaza	Basque Restaurant

Discussion

- In the above results, we have the clustering of the various districts based on top venues for each neighborhood.
- The French Restaurant is the first most common venue in most of districts in Paris.
- When reviewing the clusters, we could see that the **Moroccan restaurant** is in cluster 1.

Discussion

- The districts, which are clustered in **cluster 1**, could be potential candidates.

- Our proposed districts could be considered as one of the interesting starting points.

- In order to analyze more in details, we need to review other important factors and conditions.

Conclusion

- Firstly, we imported the neighborhood data of 20 districts in Paris.

- Secondly, we used python package geopy to convert the addresses into the latitude and the longitude values. And, to retrieve the important information of the top 25 famous venues for each district using Foursquare API.

Conclusion

- Thridly, we used python package folium to generate a map in which the district neighborhoods are superimposed on top.
- Finally, we used Elbow method to determine the optimal value of the number of clusters for K-Means clustering. And, we launched K-Means technique to cluster the neighborhood into 8 clusters.
- Then, we reused tool folium to illustrate the data in a map where the districts of Paris are clustered according to the venue density.

Conclusion

- We proposed some good districts to the investors depending on their requirements and conditions.

- ✓ District where there are Moroccan restaurants **District 15 or District 5.**

- ✓ Other good potential districts: **District 1 (Louvre) and District 20 (Menilmontant) .**

Perspectives

- To go further in the analysis, we could enhance the features of district by adding more relevant features for each district such as:

✓ the transport info (public transport, parking, etc.),

✓ the information of Moroccan and Arab communities,

✓ the information of major tourist venues.

Perspectives

- To improve the performance of clustering result, we could do with other algorithms, for instance:

- ✓ FCM : Fuzzy c-means method
- ✓ DBSCAN: Density-Based Spatial Clustering of Applications with Noise
- ✓ Hierarchical K-Means Clustering
- ✓ HCPC: Hierarchical clustering on principal components
- ✓ Deep Learning Models. To see more details, please check "Unsupervised deep embedding for clustering analysis"

- [1] Paris population, https://www.worldometers.info/world-population/france-population/

- [2] The Economist Intelligence Unit, https://www.eiu.com/n/campaigns/worldwide-cost-of-living-2020

- [3] The tutorials in course "Applied Data Science Capstone", https://www.coursera.org/learn/applied-data-science-capstone/

- [4] Elbow Method for optimal value of k in KMeans, https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/

Thank you for your attention