
CAPSTONE PROJECT

MARKET RESEARCH FOR OPENING NEW MEXICAN RESTAURANT IN PARIS

IBM DATA SCIENCE

BAYER KARL

OCT 2020

Summary

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

Introduction

- Paris is the French capital and most interesting city of France.
 - 2140 536 residents in 105 square Kilometers
 - A concentration of latin american
-
- Business Problem : Finding the good districts from 20 districts in Paris to open a Mexican Restaurant
-

Data Description

- Neighborhood data of 20 districts in Paris, Frances
 - Paris Arrondissements & Neighborhoods Map
 - Arrondissements in Paris, France
 - Coordinates (latitude, longitude) of each district
 - Using tool geopy
 - The most famous venues and their relative locations in the raise 500 meters for the center of each
 - Using foursquare API
-

Methodology

- Building Paris neighborhood date
(Postcode, Neighborhood)
- Building the coordinates of all districts
in Paris, France using tool geopy
- Using Foursquare API to generate to
the most famous venues and relative
locations
- Exploring and analyzing the data and
finding the optimal number of clusters
for k-Mean clustering using the Elbow
method
- Using K-Means to cluster the
neighborhoods in the city of Paris
based on 10 venues (top venue) for
each neighborhood district

Methodology

- Visualizing geographic details of Paris and its district neighborhood with are superimposed using tool Folium
- Analizing the clustering result and the proposing some suggestions about the good district to open a Mexican restaurant
- Giving some perspectives to enhance the performance

Results

	postcode	neighbourhood
0	75001	75002,75003,75004,75005,75006,75007,75008,75009
1	75002	75001,75003,75009,75010
2	75003	75001,75002,75004,75010,75011
3	75004	75001,75003,75005,75006,75011,75012
4	75005	75001,75004,75006,75012,75013,75014
5	75006	75001,75004,75005,75007,75014,75015
6	75007	75001,75006,75008,75015,75016
7	75008	75001,75007,75009,75016,75017,75018
8	75009	75001,75002,75008,75010,75017,75018
9	75010	75002,75003,75009,75011,75018,75019,75020
10	75011	75003,75004,75010,75012,75019,75020
11	75012	75004,75005,75011,75013,75020
12	75013	75005,75012,75014
13	75014	75005,75006,75013,75015
14	75015	75006,75007,75014,75016
15	75016	75007,75008,75015,75017
16	75017	75008,75009,75016,75018
17	75018	75008,75009,75010,75017,75019
18	75019	75010,75011,75018,75020
19	75020	75010,75011,75012,75019

➤ **Table 1: Neighborhood Data in Paris , France**

	postcode	latitude	longitude
0	75001	48.863554	2.338856
1	75002	48.867418	2.344256
2	75003	48.862607	2.360211
3	75004	48.856004	2.357028
4	75005	48.852752	2.346343
5	75006	48.853537	2.343370
6	75007	48.855913	2.313839
7	75008	48.872385	2.312707
8	75009	48.877355	2.336856
9	75010	48.879201	2.354391
10	75011	48.855630	2.370806
11	75012	48.839734	2.380054
12	75013	48.826997	2.353396
13	75014	48.828590	2.307541
14	75015	48.838461	2.315728
15	75016	48.855031	2.273958
16	75017	48.883508	2.304923
17	75018	48.893074	2.343881
18	75019	48.878076	2.376198
19	75020	48.857126	2.409257

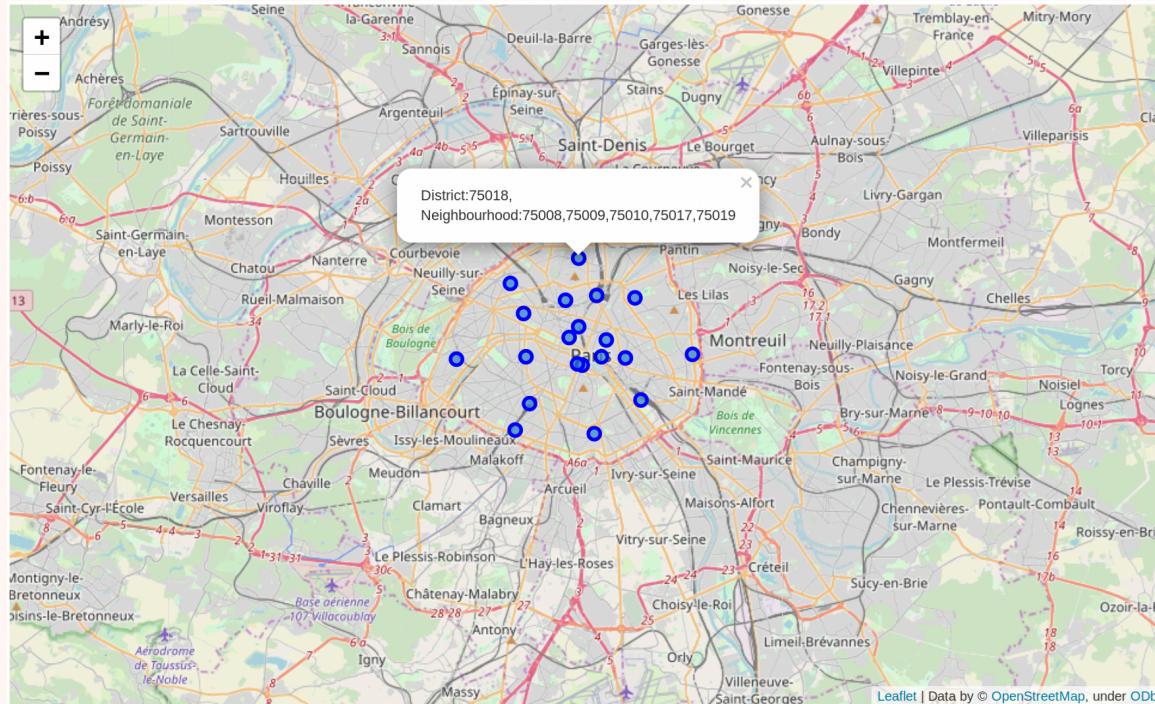
➤ **Table 2: The coordinates of 20 district in Paris, France.**

Results

	neighbourhood	neighbourhood latitude	neighbourhood longitude	Venue	Venue latitude	Venue longitude	Venue Category
0	75002,75003,75004,75005,75006,75007,75008,75009	48.863554	2.338856	Jardin du Palais Royal	48.864941	2.337728	Garden
1	75002,75003,75004,75005,75006,75007,75008,75009	48.863554	2.338856	Palais Royal	48.863236	2.337127	Historic Site
2	75002,75003,75004,75005,75006,75007,75008,75009	48.863554	2.338856	Comédie-Française	48.863088	2.336612	Theater
3	75002,75003,75004,75005,75006,75007,75008,75009	48.863554	2.338856	Place du Palais Royal	48.862523	2.336688	Plaza
4	75002,75003,75004,75005,75006,75007,75008,75009	48.863554	2.338856	Christian Louboutin	48.862697	2.340757	Shoe Store

➤ **Table 3: Example of the top 100 venues in radius 500 metres from the center of each district**

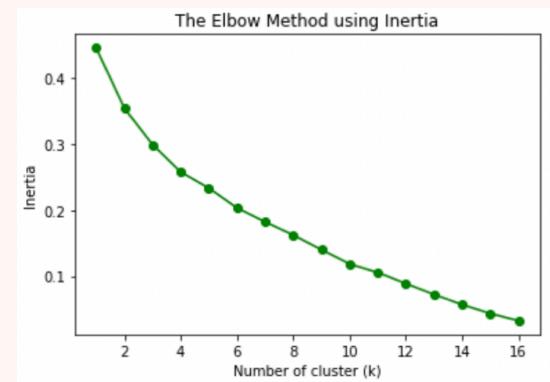
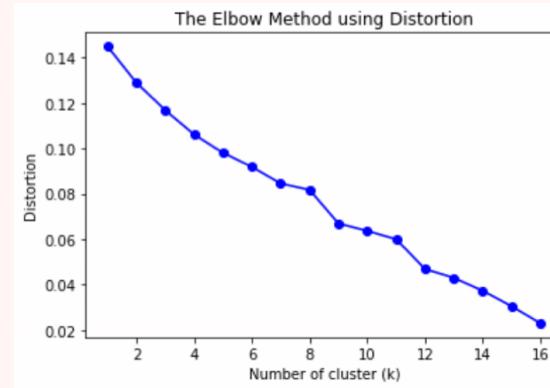
Results



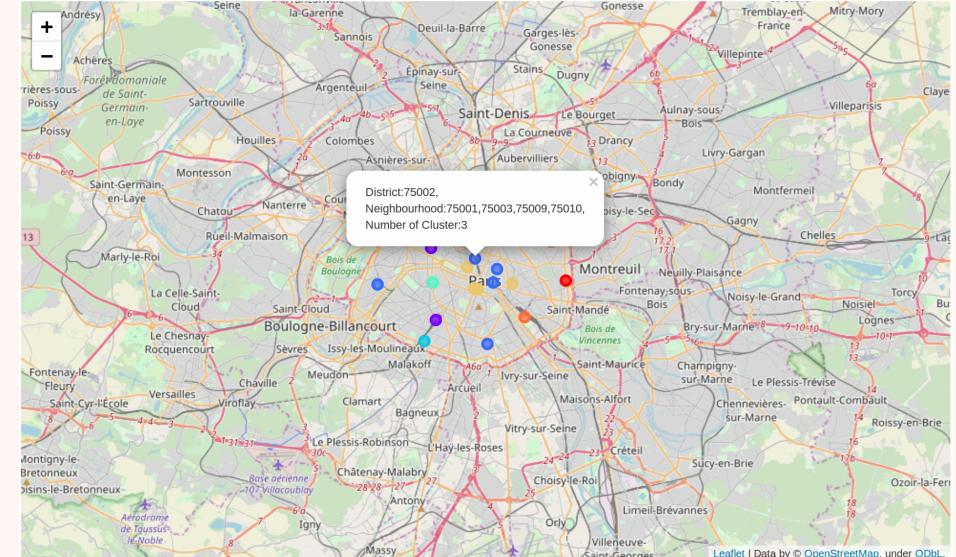
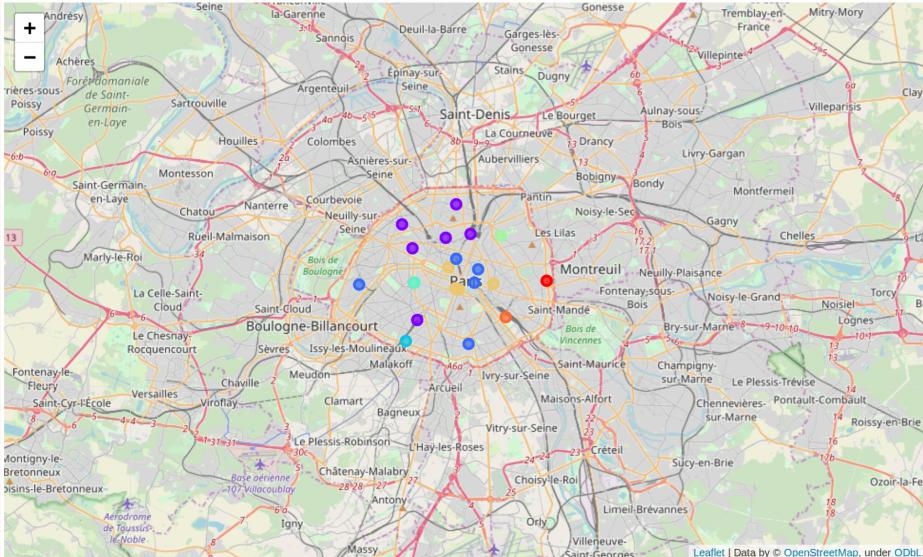
➤ **Visualization of geographic details of Paris and it's district neighborhood with are superimposed on top**

Results

- The chart of the Elbow Method using the distortion or the inertia
- Difficult to determine the Elbow point using inertia
- Applying the distortion with the number of clusters, we choose the optimal number of clusters for k-mean is 8



Results



➤ Using the optimal value K we launch K-Mean technique to cluster the neighborhood into 8 clusters

➤ Using the optimal value K, we launch K-Means technique to cluster the neighborhood into 8 clusters

Discussion

- From above result, we obtain 8 clusters that contain of the carious districts based on top 10 venues for each neighborhood.
 - French Restaurant is the first most common venue in most go districts in Paris
 - When reviewing the clusters we could see that the Mexican restaurant in cluster , Mexican is 2 and 10 most common venues in district 2 and 6
 - District in closet 3 could be a good candidate s
 - Our proposed district could be considered as one of the interesting starting points.
 - In order to analyze more in detail , we need to review the other relevant important factors and conditions.
-

Conclusion

- First we built the neighborhood data of 20 districts in Paris based on the information
 - Second we use tool guppy to convert the addresses that were built in the first step into the latitude values and the longitude values.
 - Third to retrieve the important information of the top 10 famous venues for each district, we use Foursquare API, such as, their coordinates.
 - Fourth, to visualize geographic details of Paris and its district neighborhoods , we use Folium
-

Conclusion

- Fifth, we use Elbow method to determine the optimal value of the number of clusters for K-Means clustering
- Sixth, we laugh I-means technique to cluster the neighborhood into 8 clusters.
- Then, we reuse tool filial to illustrate an information map where the districts in Paris are clustered according to the venue density
- Finally, we proposed some good districts to the investors depending in their requirements and conditions
- District with have had top 1° Mexican restaurants is district 6 and 2
- Other Potential districts: 3 / 4 / 16

Perspectives

- To enhance the features of district, we could add more relevant features for each district such as:
 - The transport information
 - The information of Mexican communities
 - the information of major tourist venues, etc
- To improve the performance of clustering result, we could do with other algorithms, for instance :
 - DBSCAN : Density based clustering
 - Hierarchical K-Means Clustering
 - Fuzzy c-means method
 - HCPC: Hierarchical clustering on principal component
 - Deep learning models

References

- INSEE official estimated population by department and region as of 1 January 2019". 22 January 2019. Archived from the original on 21 April 2017. Retrieved 4 April 2017.
 - Île-de-France, <https://en.wikipedia.org/wiki/%C3%8Ele-de-France>
 - Paris, <https://en.wikipedia.org/wiki/Paris>
 - 5 top Mexicans restaurants in Paris : <https://www.mariefrance.fr/cuisine/mes-bonnes-adresses/paris-restaurants-mexicains-a-tester-absolument-439188.html#item=1> restaurants-paris/
 - Telegraph, 16 March 2018 Archived 30 March 2019 at the Wayback Machine
 - The tutorials in course "Applied Data Science Capstone", <https://www.coursera.org/learn/applied-data-science-capstone/>
 - Paris Arrondissements & Neighborhoods Map, <https://parismap360.com/paris-Housing>
 - Sales Prices & Venues Data Analysis of Istanbul, <https://www.linkedin.com/pulse/housing-sales-prices-venues-data-analysis-ofistanbul-sercan-y%C4%B1ld%C4%B1z/arondissement-map#.XfVpqtEo91I>
 - Arrondissements in Paris, France, <https://francetravelplanner.com/go/paris/areas/arrondismt.html>
 - Elbow Method for optimal value of k in KMeans, <https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/>
-

THANK YOU
