IBM Data Science Capstone Project

I- Introduction

1) Description and discussion of the business problem

In this capstone project we will explore the location data of Paris and will try to answer to a specific question regarding opening a vegetarian restaurant.

Paris multi-cultural french city, it is a interesting type of restaurant because it is trendy right now.

The target of this report are stakeholders who want to invest in opening a vegetarian restaurant in Paris without knowing the right borough to do it.

The purpose of trying to have a look at Paris's boroughs is to find locations that don't have this type of restaurant, therefore it will be an opportunity for investors.

By categorizing the boroughs into clusters of venues we will try to answer the business problem of opening a vegetarian restaurant in Paris

I- Data description

For the data retrieving we will set up some criterias :

- Locations that are close to the city center are privileged
- List of restaurants by borough
- list of boroughs with no vegetarian restaurant
- Trending venues in the area

for those criterias we will need to retrieve multiple data sets :

- Dataset of Paris Arrondissements
- Make a list of those Arrondissements to search the latitude and longitude of those districts using geolocator from geopy
- Data set of venues from the Foursquare API
- Count the number of vegeterian restaurants
- Try to identify Arrondissements with low density vegetarian restaurants
- Vizualise clusters of venues in each borough

Data retrieving : Arrondissements of Paris

	Arrondissement (R for Right Bank, L for Left Bank)	Name	Area (km2)	Population(2017 estimate)	Density (2017) (inhabitants per km2)	Peak of population	Mayor (2020-2026)
0	1st(Ier) RAdministratively part of Paris Centre	Louvre	5.59 km2 (2.16 sq mi)	100196	17924	before 1861	Ariel Weil (PS)
1	2nd (IIe) RAdministratively part of Paris Centre	Bourse	5.59 km2 (2.16 sq mi)	100196	17924	before 1861	Ariel Weil (PS)
2	3rd (IIIe) RAdministratively part of Paris Centre	Temple	5.59 km2 (2.16 sq mi)	100196	17924	before 1861	Ariel Weil (PS)
3	4th (IVe) RAdministratively part of Paris Centre	Hôtel-de-Ville	5.59 km2 (2.16 sq mi)	100196	17924	before 1861	Ariel Weil (PS)
4	5th (Ve) L	Panthéon	2.541 km2 (0.981 sq mi)	59631	23477	1911	Florence Berthout (DVD)
5	6th (VIe) L	Luxembourg	2.154 km2 (0.832 sq mi)	41976	19524	1911	Jean-Pierre Lecoq (LR)

We will in this section transform this dataset to obtain a list of arrondissements of paris to retrieve the longitude and latitude of each one

	Name
0	Louvre
1	Bourse
2	Temple
3	Hôtel-de-Ville
4	Panthéon
5	Luxembourg

	Arrondissements	Latitude	Longitude
0	Bourse	48.869100	2.340220
1	Temple	48.864660	2.364260
2	Hôtel-de-Ville	48.857101	2.353064
3	Panthéon	48.846210	2.346110
4	Luxembourg	48.847580	2.340940
5	Palais-Bourbon	48.860830	2.318590

With the foursquare API we can retrieve 100 venue by district at best so I managed to filter the API request to obtain a dataframe with the venue, categorie, distance from the center.

	Subdistrict	Subdistrict Latitude	Subdistrict Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bourse	48.8691	2.34022	Karaage-Ya Bourse	48.870150	2.341826	Japanese Restaurant
1	Bourse	48.8691	2.34022	Galerie Vivienne	48.866731	2.339800	Historic Site
2	Bourse	48.8691	2.34022	Passage des Panoramas	48.870992	2.341676	Pedestrian Plaza
3	Bourse	48.8691	2.34022	Place des Victoires	48.865759	2.341219	Plaza
4	Bourse	48.8691	2.34022	ZenZoo – Bar à Thé	48.867021	2.336738	Bubble Tea Shop
5	Bourse	48.8691	2.34022	Max Linder Panorama	48.871384	2.344809	Indie Movie Theater

Methodology

K-means clustering algorithm to categorize each of the subdistricts within Paris

A one-hot encoding will be done on the venue dataframe and it will be grouped by subdistrict. The encoding will return venue categories as column per subdistrict, which will then be grouped to provide weighting of venue type occurence on each subdistrict.

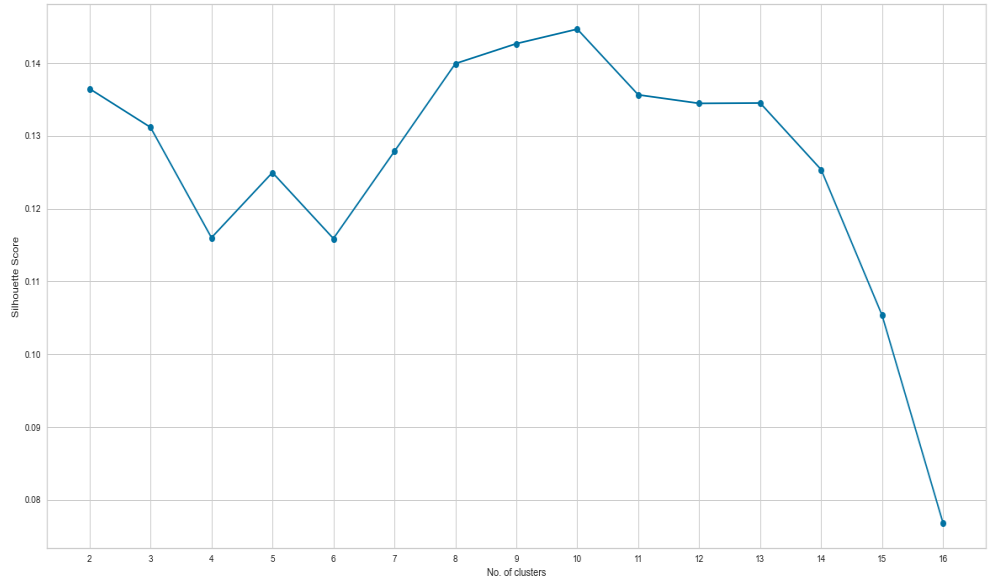
The encoded dataframe will be further filtered into top venues before the K-means clustering algorithm will be run over it. This will return cluster labels over the subdistricts. The clusters will be observed one by one manually to determine its content.

Recommendation will be made based on the clustering.

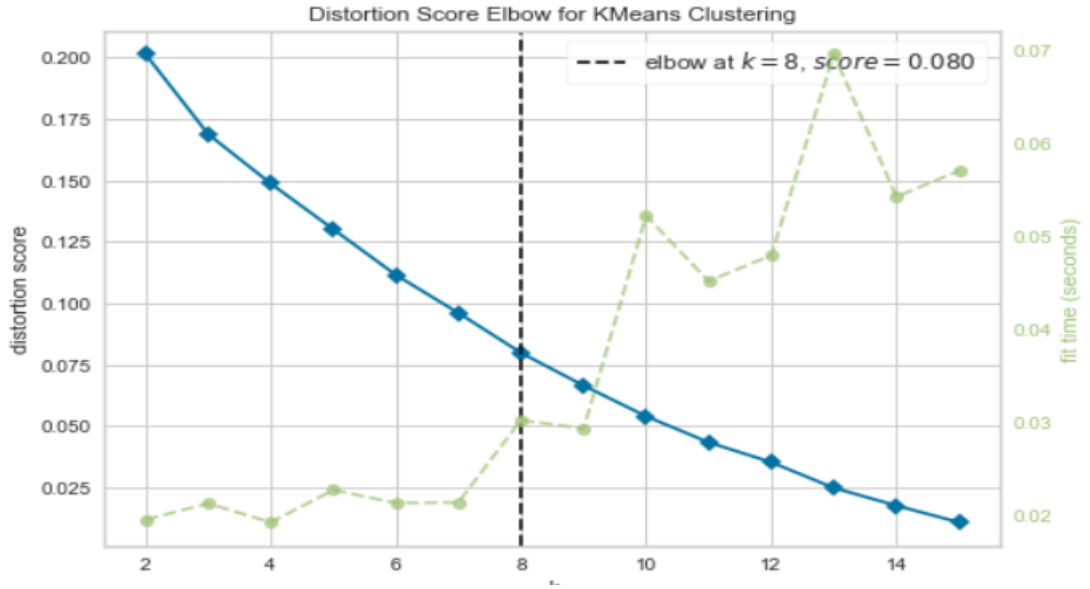
To determine the right number of k clusters needed to explain the dataset that we made I used elbow technique and silhouette

Results

The silhouette technique resulted in a 10 clusters as optimal number of k clusters



Elbow technique on the other hand demonstrate that 8 clusters are sufficient to minimize the sum of squared errors in our dataset.



We obtain 8 clusters that represents 2 categories of venues as a most common venue :

These clusters contains mostly french restaurants in their most common venue

	Arrondissements	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
1	Temple	Coffee Shop	French Restaurant	Bakery	Restaurant	Bookstore	Clothing Store	Sandwich Place	Italian Restaurant	Café	Seafood Restaurant
7	Opéra	French Restaurant	Bakery	Italian Restaurant	Coffee Shop	Bar	Wine Bar	Hotel	Gym / Fitness Center	Bistro	Lebanese Restaurant
16	Butte-Montmartre	French Restaurant	Bakery	Italian Restaurant	Bar	Plaza	Pizza Place	Café	Wine Bar	Sandwich Place	Bistro

	Arrondissements	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	Bourse	Plaza	Hotel	French Restaurant	Bakery	Vegetarian / Vegan Restaurant	Pedestrian Plaza	Garden	Historic Site	Café	Indie Movie Theater
2	Hôtel-de-Ville	French Restaurant	Plaza	Coffee Shop	Park	Art Gallery	Bakery	Art Museum	Pedestrian Plaza	Garden	Wine Bar
3	Panthéon	Plaza	French Restaurant	Wine Bar	Ice Cream Shop	Bakery	Garden	Bookstore	Greek Restaurant	Miscellaneous Shop	Café
4	Luxembourg	French Restaurant	Plaza	Bakery	Café	Wine Bar	Hotel	Pastry Shop	Bookstore	Ice Cream Shop	Garden

	Arrondissements	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
12	Observatoire	French Restaurant	Hotel	Bar	Italian Restaurant	Bistro	Creperie	Pizza Place	Coffee Shop	Plaza	Japanese Restaurant
13	Vaugirard	French Restaurant	Hotel	Bakery	Creperie	Japanese Restaurant	Italian Restaurant	Dessert Shop	Pastry Shop	Park	Pizza Place
15	Batignolles-Monceau	French Restaurant	Wine Bar	Hotel	Italian Restaurant	Bar	Park	Bistro	Bakery	Farmers Market	Thai Restaurant

	Arrondissements	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
9	Popincourt	French Restaurant	Cocktail Bar	Italian Restaurant	Art Gallery	Bistro	Bookstore	Coffee Shop	Bar	Bakery	Beer Store
17	Buttes-Chaumont	French Restaurant	Bar	Restaurant	Bistro	Café	Park	Pizza Place	Italian Restaurant	Cocktail Bar	Chinese Restaurant
18	Ménilmontant	French Restaurant	Bar	Bakery	Wine Bar	Restaurant	Bistro	Bookstore	Italian Restaurant	Beer Store	Moroccan Restaurant

	Arrondissements	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
8	Entrepôt	French Restaurant	Cocktail Bar	Wine Bar	Museum	Pedestrian Plaza	Italian Restaurant	Bar	Hotel	Bookstore	Sandwich Place
10	Reuilly	French Restaurant	Bar	Wine Bar	Creperie	Italian Restaurant	Farmers Market	Bookstore	Plaza	Hotel	Seafood Restaurant

	Arrondissements	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
5	Palais-Bourbon	French Restaurant	Hotel	Plaza	Historic Site	Art Museum	Garden	Boutique	Coffee Shop	Dessert Shop	Tailor Shop

On the other hand we have clusters who represents the influence of other cultures in the type of cuisine :

The most common venue in Elysée's district is Italian restaurant

	Arrondissements	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
6	Élysée	Italian Restaurant	Bar	French Restaurant	Indian Restaurant	Café	Hotel	Coffee Shop	Park	Plaza	Dessert Shop

The most common venue in Gobelins district is Vietnamese restaurant

	Arrondissements	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
11	Gobelins	Vietnamese Restaurant	French Restaurant	Thai Restaurant	Hotel	Asian Restaurant	Bakery	Bistro	Italian Restaurant	Bar	Plaza

Results

Acording to our clusters formation we can deduce the importance of restaurants in the french culture and in the urban landscape of venues,

This is due to the spread of the art of cooking after guilds were abolished after the french revolution,alowing french cooks to operate with fewer restrictions.

The influence of other culture is clearly visible amongst the clusters which contain for 2 of them Vietnamese and Italian as a most common venue.

The French have enjoyed an excellent cultural culinary exchange with many Asian countries. Vietnam, for instance, was a former French colony, which allowed the two countries to exchange culinary traditions.

Discussion

There are definite limitation with using the FourSquare API as the 100 venues limit might skew the result of the more densely populated subdistrict. Also, some subdistricts have low count of venues that it might be considered to be insufficient in determining its characteristics.

It might also be the case that FourSquare user base are skewed to the foodie type, which might explain the limited trending venues on the Elysée arrondissement

Conclusion

To conclude by answering to the bussiness problem we can say that opening a vegetarian restaurant is a niche market segment that is flourishing all arround the globe.

The under representative number of this type of restaurant make it easier to have a comparative advantage regarding the growing demand for those type of restaurants.

The vegetarian restaurant can be a good investment in all the arrondissements.