Chinese Restaurant implementation

Pierre-Alexis Blond

June 2019

I. Introduction

Asian Divine provides a beautiful spot with a well-rounded offering of Thai, Chinese, Japanese, and Malaysian food in an atmosphere that makes you want to stay a while.

With over 150 restaurants over the world, Asian Divine would want to enlarge their network with 5 new restaurants in France.

The sample recommender in this project will provide the following requirements:

- Asian Divine wants to implement new restaurants in cities with at least half a 50 000 inhabitants .
- In order to maximize potential profits, recommendations have to consider at least the mean incomes of inhabitants in each city and a metric showing the interest of imhabitants about asian food.
- The recommendation should not only present the most viable option, but also present a comparison table of all possible town venues.

II. Data requirements

We will use the website salaire-moyen (www.salairemoyen.com) to get statics about inhabitants income in each French cities. The following informations will be collect :

- Quality of life (€)
- Revenue from the reachiest 10%
- Singles Revenues (€)
- Couple without children (€)
- Single parent family (€)
- Couple with 1 children (€)
- Couple with 2 children (€)
- Couple with 3 children (€)

We will be using the FourSquare API to explore neighborhoods in selected towns in France. The Foursquare explore function will be used to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters.

We will also be using the Google map API and the French government API to collect data like latitude, longitude and population of each city.

III. Methodology

The first step will be to gather data from the different API / website and then to prepare and scale data.
Data collection & preparation

- Get all cities with 50K inhabitant at least

Marseille	
Lille	
Strasbourg	
Amiens	
Saint-\u00c9tienne	sa
Montpellier	
N\u00eemes	
Perpignan	
Saint-Denis	

- Request to Google API to get latitude, longitude and population of each city

City	Foursquare url	Latitude	Longitude	Address	Population
Amiens	salaire-ville-80021-Amiens.html	49.894067	2.295753	80000, 80090, 80080	133755
Saint-\u00c9tienne	salaire-ville-42218-Saint_\u00c9tienne.html	45.525587	4.874339	38200	29454
Montpellier	salaire-ville-34172-Montpellier.html	43.610769	3.876716	34000, 34080, 34090, 34070, 34295	281613
Perpignan	salaire-ville-66136-Perpignan.html	42.688659	2.894833	66000, 66100	121875
Saint-Denis	salaire-ville-97411-Saint_Denis.html	48.936181	2.357443	93200, 93210	111354

- Scrap statistical data for each city

City	Population	Latitude	Longitude	Quality of life (€)	Revenue from the reachiest 10%	Singles Revenues (€)	Couple without children (€)	Single parent family (€)	Couple with 1 children (€)	Couple with 2 children (€)	Couple with 3 children (€)
Boulogne-Billancourt	119645	48.839695	2.239912	2821	2.3	2385	4602	3101	5227	6615	6887
Paris (75) [d\u00e9tails]<\/span><\/a><\/td>	2190327	48.856614	2.352222	2321	2.7	2169	4262	2820	4516	5462	3154
Aix-en-Provence	143006	43.529742	5.447427	1978	2.2	1755	3429	2282	3746	4219	3887
Lyon	515695	45.764043	4.835659	1920	2.2	1735	3389	2256	3692	4221	3224
Bordeaux	252040	44.837789	-0.579180	1824	2.3	1523	3125	1980	3496	4225	3939

- Request Foursquare API for venues (Asian restaurant) in each city

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Boulogne-Billancourt	48.839695	2.239912	Mise En Saine	48.832324	2.251850	Asian Restaurant
1	Boulogne-Billancourt	48.839695	2.239912	La Can Tin'h	48.834700	2.251916	Asian Restaurant
2	Boulogne-Billancourt	48.839695	2.239912	Asian Box	48.834464	2.245986	Chinese Restaurant
3	Boulogne-Billancourt	48.839695	2.239912	Tachibana	48.833055	2.245982	Asian Restaurant
4	Boulogne-Billancourt	48.839695	2.239912	Sanki	48.836371	2.251768	Japanese Restaurant

- Count venues in each city and merging it with the main dataframe

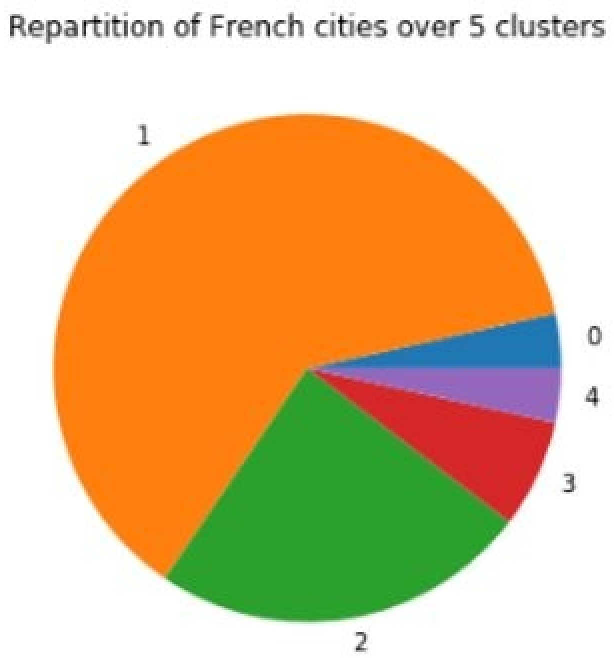
City	Population	Latitude	Longitude	Quality of life (€)	Revenue from the reachiest 10%	Singles Revenues (€)	Couple without children (€)	Single parent family (€)	Couple with 1 children (€)	Couple with 2 children (€)	Couple with 3 children (€)	Venues Count
Boulogne-Billancourt	119645	48.839695	2.239912	2821	2.3	2385	4602	3101	5227	6615	6887	11.0
Paris (75) [d\u00e9tails] <\/span><\/a><\/td>	2190327	48.856614	2.352222	2321	2.7	2169	4262	2820	4516	5462	3154	11.0
Aix-en-Provence	143006	43.529742	5.447427	1978	2.2	1755	3429	2282	3746	4219	3887	11.0
Lyon	515695	45.764043	4.835659	1920	2.2	1735	3389	2256	3692	4221	3224	11.0
Bordeaux	252040	44.837789	-0.579180	1824	2.3	1523	3125	1980	3496	4225	3939	12.0

Modeling

We will try to cluster and segment data using different unsupervised algorithms:

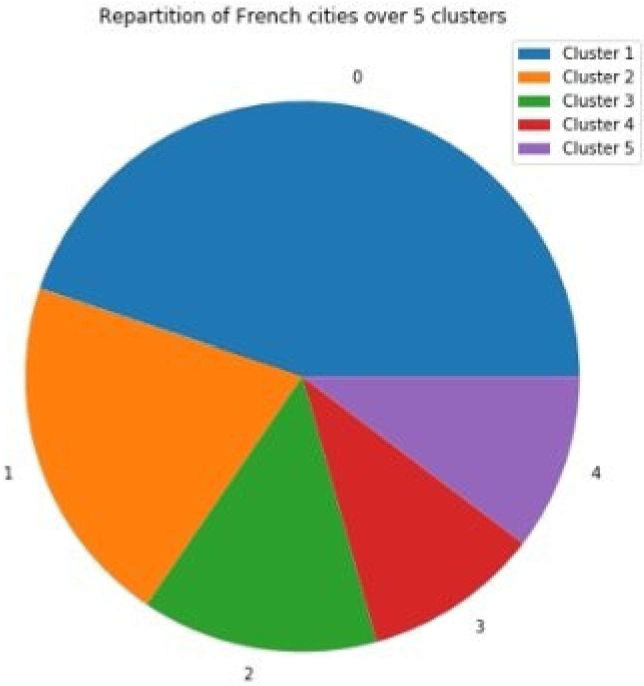
- - Kmeans
- - Agglomerative Clustering which recursively merges the pair of clusters that minimally increases a given linkage distance.
- - DBSCAN Clustering (Density-Based Spatial Clustering of Applications with Noise)

  - Kmeans method



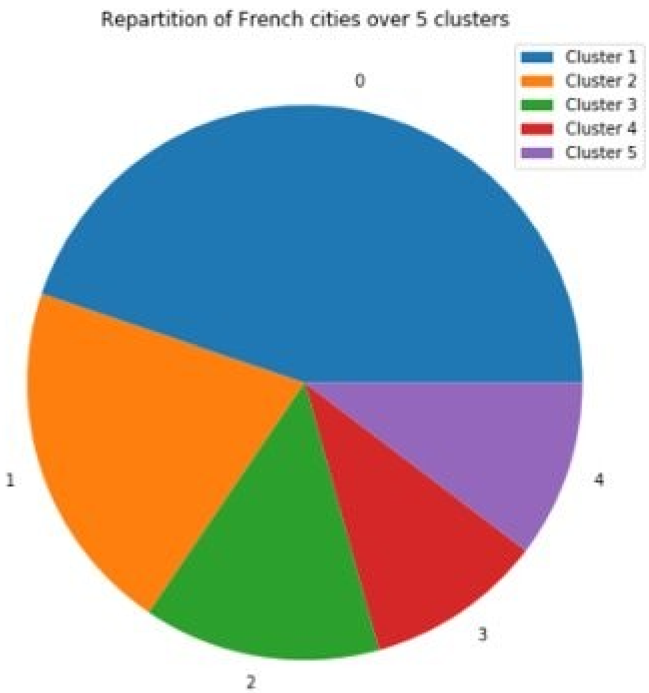
City	Population	Latitude	Longitude	Quality of life (€)	Revenue from the reachiest 10%	Singles Revenues (€)	Couple without children (€)	Single parent family (€)	Couple with 1 children (€)	Couple with 2 children (€)	Couple with 3 children (€)	Venues Count	Cluster
Boulogne-Billancourt	119645	48.839695	2.239912	2821	2.3	2385	4602	3101	5227	6615	6887	11.0	1
Paris (75) [d\u00e9tails] <\/span><\/a> <\/td>	2190327	48.856614	2.352222	2321	2.7	2169	4262	2820	4516	5462	3154	11.0	4
Aix-en-Provence	143006	43.529742	5.447427	1978	2.2	1755	3429	2282	3746	4219	3887	11.0	1
Lyon	515695	45.764043	4.835659	1920	2.2	1735	3389	2256	3692	4221	3224	11.0	3
Bordeaux	252040	44.837789	-0.579180	1824	2.3	1523	3125	1980	3496	4225	3939	12.0	2

- Agglomerative Clustering



	Latitude	Population	Longitude	Quality of life (€)	Revenue from the reachiest 10%	Singles Revenues (€)	Couple without children (€)	Single parent family (€)	Couple with 1 children (€)	Couple with 2 children (€)	Couple with 3 children (€)	Venues Count	Cluster
0	119645.0	48.839695	2.239912	2821.0	2.3	2385.0	4602.0	3101.0	5227.0	6615.0	6887.0	11.0	1
1	2190327.0	48.856614	2.352222	2321.0	2.7	2169.0	4262.0	2820.0	4516.0	5462.0	3154.0	11.0	1
2	143006.0	43.529742	5.447427	1978.0	2.2	1755.0	3429.0	2282.0	3746.0	4219.0	3887.0	11.0	1
3	515695.0	45.764043	4.835659	1920.0	2.2	1735.0	3389.0	2256.0	3692.0	4221.0	3224.0	11.0	1
4	252040.0	44.837789	-0.579180	1824.0	2.3	1523.0	3125.0	1980.0	3496.0	4225.0	3939.0	12.0	1

- DBSCAN Clustering



City	Latitude	Population	Longitude	Quality of life (€)	Revenue from the reachiest 10%	Singles Revenues (€)	Couple without children (€)	Single parent family (€)	Couple with 1 children (€)	Couple with 2 children (€)	Couple with 3 children (€)	Venues Count	Cluster
Amiens	133755.0	49.894067	2.295753	1414.0	2.2	1473.0	2627.0	1915.0	2669.0	2976.0	1849.0	27.0	3
Saint-\u00c9tienne	29454.0	45.525587	4.874339	1409.0	2.0	1448.0	2588.0	1882.0	2713.0	2982.0	1955.0	11.0	0
Montpellier	281613.0	43.610769	3.876716	1404.0	2.3	1479.0	2730.0	1923.0	2695.0	2888.0	1721.0	19.0	0
Perpignan	121875.0	42.688659	2.894833	1205.0	2.4	1307.0	2352.0	1699.0	2164.0	2369.0	1412.0	14.0	0
Saint-Denis	111354.0	48.936181	2.357443	1178.0	3.1	1186.0	2196.0	1542.0	2376.0	2885.0	1976.0	4.0	0

IV – Results and recommendations

Now that we clustered data using different model. We want to understand what each cluster represent and which cluster is suited for Asian Divine

A fundamental requirement of this project was "to consider at least the mean incomes of inhabitants in each city and a metric showing the
interest of imhabitants about asian food".

For this pupose we will use a function that return a score for each cluster calculated as follow:

$$\overline{\text{score}} = \overline{\text{income}} \text{ x } \overline{\text{venues count}}$$

Using this function we will recommand the "best potential profit" cluster.

- ## Results using Kmeans method

Using Kmeans method we can now recommend the following cities:

City	Population	Latitude	Longitude	Quality of life (€)	Revenue from the reachiest 10%	Singles Revenues (€)	Couple without children (€)	Single parent family (€)	Couple with 1 children (€)	Couple with 2 children (€)	Couple with 3 children (€)	Venues Count	Cluster
Lyon	515695	45.764043	4.835659	1920	2.2	1735	3389	2256	3692	4221	3224	11.0	3
Toulouse	475438	43.604652	1.444209	1700	2.2	1571	3014	2042	3253	3767	2421	26.0	3

- ## Results using Agglomerative Clustering

Using Agglomerative clustering, we can now recommend the following cities:

City	Latitude	Population	Longitude	Quality of life (€)	Revenue from the reachiest 10%	Singles Revenues (€)	Couple without children (€)	Single parent family (€)	Couple with 1 children (€)	Couple with 2 children (€)	Couple with 3 children (€)	Venues Count	Cluster
Toulouse	475438.0	43.604652	1.444209	1700.0	2.2	1571.0	3014.0	2042.0	3253.0	3767.0	2421.0	26.0	3
Tours	136565.0	47.394144	0.684840	1580.0	2.1	1495.0	2792.0	1944.0	2891.0	3316.0	2483.0	27.0	3
Amiens	133755.0	49.894067	2.295753	1414.0	2.2	1473.0	2627.0	1915.0	2669.0	2976.0	1849.0	27.0	3

- **Results using DBSCAN Clustering**

Using Agglomerative clustering, we can now recommend the following cities:

City	Latitude	Population	Longitude	Quality of life (€)	Revenue from the reachiest 10%	Singles Revenues (€)	Couple without children (€)	Single parent family (€)	Couple with 1 children (€)	Couple with 2 children (€)	Couple with 3 children (€)	Venues Count	Cluster
Toulouse	475438.0	43.604652	1.444209	1700.0	2.2	1571.0	3014.0	2042.0	3253.0	3767.0	2421.0	26.0	3
Tours	136565.0	47.394144	0.684840	1580.0	2.1	1495.0	2792.0	1944.0	2891.0	3316.0	2483.0	27.0	3
Amiens	133755.0	49.894067	2.295753	1414.0	2.2	1473.0	2627.0	1915.0	2669.0	2976.0	1849.0	27.0	3

V – Discussion and Conclusion

On this project, analysis of best town, recommendations based on venues count and on income data has been presented.

Using Foursquare API, we do not have collected a good amount of venues (asian restaurant) recommnedations. It means that this project is mainly for education purposes and it is not "production ready" at all.

The generated clusters from our results shows that there are very good way to represent interesting cities based on population income, position ... This kind of results may be very interesting and can help any business which want to enlarge their position and invest in foreign areas