PROPOSING AN AIRBNB ROOM FOR ASIAN TOURISTS IN PARIS

Emmanuel Folorunsho

October, 2019

1. INTRODUCTION

1.1 Background

In the city of Paris, there are thousands of tourists who have come to see the amazing city for its sheer beauty. These tourists are from different nationals and have cuisines which are specific to them. For example, Asians will prefer to go eat in restaurants that specialize in making Asian cuisines rather than eating in a restaurant that deals with different kinds of cuisines. To this end, the problem i intend to solve is deciding the best AirBnB apartment locations where Asian tourists can reside during a tourist visit to the city of Paris.

By clustering the Asian restaurants using Foursquare API and AirBnB private room listings data for the city of Paris, one can successfully cluster the restaurants into groups and propose the best AirBnB private room apartment based on the average price per night and the number of reviews each room has. By solving this problem, Asian tourists coming to Paris can choose an AirBnB apartment situated in proximity to Asian restaurants so as to get easy access to Asian cuisines.

1.2 The Problem

For this project, the problem that I intend to solve is best formulated as the question:

"Which AirBnB apartment should an Asian tourist choose to reside in during a visit to Paris city in order to enjoy Asian cuisines located nearby"?

1.3   Interest

Obviously, Asian tourists who intend visiting Paris, France will be quite interested in this project as it will help them choose the most preferred AirBnB private room that is situated in proximity to Asian restaurants in the neighborhoods in Paris

2.   DATA ACQUISITION AND CLEANING

2.1   Data Sources

To solve the stated problem, Foursquare's location data which will be accessed using their GET API will be utilized. The kind of data that will be accessing include:

✓ Restaurants in Paris city

✓ Asian restaurants and other interesting places around these restaurants

Also, data from http://public.opendatasoft.com. This data will comprise of all AirBnB private room listings in the city of Paris, their latitude and longitude, price per night, neighborhood of location and some other redundant data. With these set of data, one can determine the best location/AirBnB property in which an Asian tourist should reside in during a visit to the city of Paris by clustering the restaurant data and determining the cluster with the most number of restaurants where Asian cuisines are being sold. Also, with the data, one can explore AirBnB listings around the chosen sites, and propose them as the hotel of choice for Asian tourists that are coming into Paris city.

Fig 1: Sample screenshot of sourced data

2.2   Data Cleaning & Preprocessing

The data sourced from http://public.opendatasoft.com is clean, thus, it required little cleaning.

Although, because of the size of the data set and computation time, this project used only 1000

rows of data which was randomly extracted from the main data set. Redundant columns were

dropped and some columns split into subsequent columns.

	Room ID	Neighborhood	Room Price	Number of reviews	Latitude	Longitude
0	12817090	Entrepôt	30	8	48.873382	2.373455
1	2181753	Reuilly	40	104	48.841314	2.383398
2	29895295	Reuilly	50	7	48.847268	2.400412
3	30728070	Popincourt	139	0	48.860921	2.366604
4	21862193	Entrepôt	40	1	48.879021	2.370081

Fig 2: Sample cleaned data

3. METHODOLOGY

The methodology to adopt for this problem can be briefly described in the following phases:

AirBnB Data collection from http://public.opendatasoft.com

The AirBnB data will be accessed from OpenDataSoft and the data set will be limited to private

rooms in AirBnB's listings in the city of Paris. The collected data set will be stored in a .csv format.

AirBnB Data preprocessing

The Airbnb data will be loaded into a pandas' data frame and data preprocessing techniques such

as cleaning, trimming, shaping etc. will be carried out on it to prepare it for processing.

EDA

Exploratory data analysis will BE carried out on the data to better describe it and get some insight
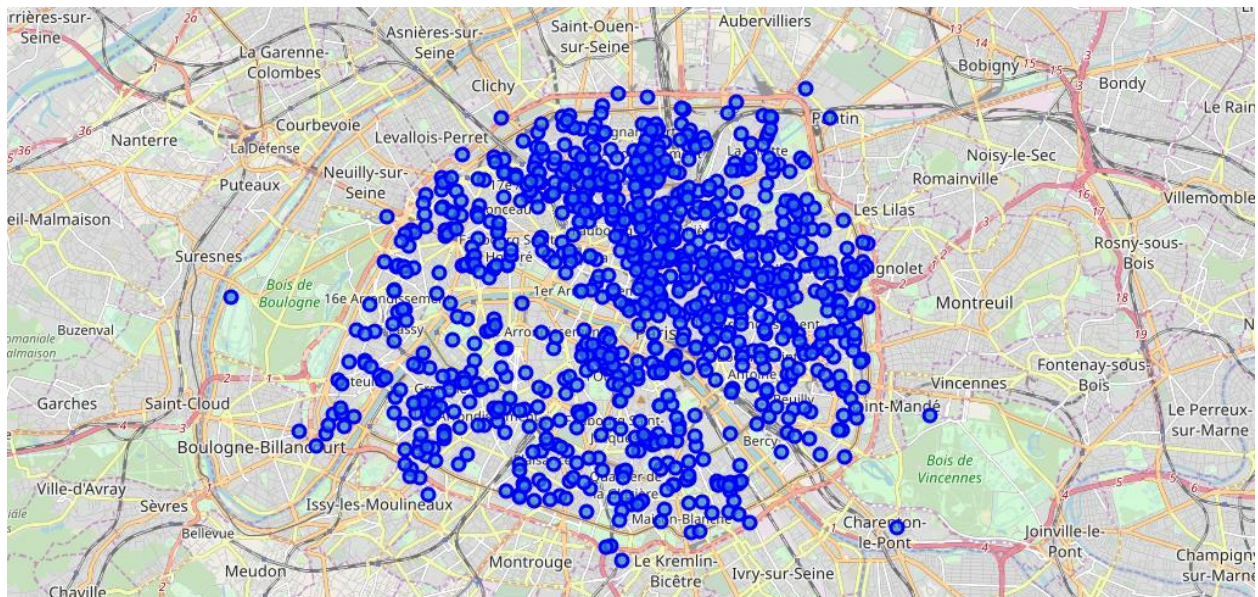
about the data.

Fig 3: Superimposing AirBnB Apartments on the Map of Paris

Utilization of Foursquare API search function

The Foursquare API search function will then be used to find Asian restaurants within 500m of each AirBnB private room in Paris city. The JSON result will be cleaned and made ready for clustering

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Entrepôt	48.873382	2.373455	Asian Soupe	48.873556	2.375092	Vietnamese Restaurant
1	Entrepôt	48.873382	2.373455	Asian Touch	48.872724	2.371047	Thai Restaurant
2	Entrepôt	48.873382	2.373455	Gölbasi Restaurant	48.873567	2.370311	Turkish Restaurant
3	Entrepôt	48.873382	2.373455	Do Brasil	48.873884	2.371596	Brazilian Restaurant
4	Entrepôt	48.873382	2.373455	Barak	48.874663	2.372837	Turkish Restaurant
5	Entrepôt	48.873382	2.373455	Oliva	48.873110	2.375951	Trattoria/Osteria
6	Entrepôt	48.873382	2.373455	Restaurant Les 4 Frères	48.874568	2.373783	Middle Eastern Restaurant
7	Entrepôt	48.873382	2.373455	Rôtisserie Sainte-Marthe	48.873507	2.370876	Restaurant
8	Entrepôt	48.873382	2.373455	Restaurant Tai Yien	48.872311	2.377507	Chinese Restaurant
9	Entrepôt	48.873382	2.373455	Le Firat	48.870307	2.372929	Kebab Restaurant
10	Entrepôt	48.873382	2.373455	Restaurant Le Yun	48.871185	2.377746	Asian Restaurant
11	Entrepôt	48.873382	2.373455	Futur Restaurant Africain	48.876320	2.372068	African Restaurant
12	Entrepôt	48.873382	2.373455	Restaurant Paradis	48.871745	2.377490	Asian Restaurant
13	Entrepôt	48.873382	2.373455	Istanbul Restaurant	48.876798	2.371918	Greek Restaurant
14	Entrepôt	48.873382	2.373455	Leupaleup	48.868500	2.374385	African Restaurant
15	Entrepôt	48.873382	2.373455	Thalassa	48.876332	2.378534	Greek Restaurant
16	Entrepôt	48.873382	2.373455	Restaurant Rapide Ben Long	48.869554	2.370224	Asian Restaurant

Fig 4: Sample Clean Foursquare API Result

Clustering of Asian restaurants

To properly group the Asian restaurants with respect to the locations of AirBnB private rooms, the KNN machine learning technique will be used to cluster the restaurant into ten (10) different clusters.

4. RESULTS & DISCUSSION

The figure below shows the result of the clustering operation of Asian restaurants within 500m distance to each of the AirBnB private rooms. The restaurants were clustered into five (5) different groups, and the clustering result shows the mean price of each cluster and the mean number of reviews of each cluster.
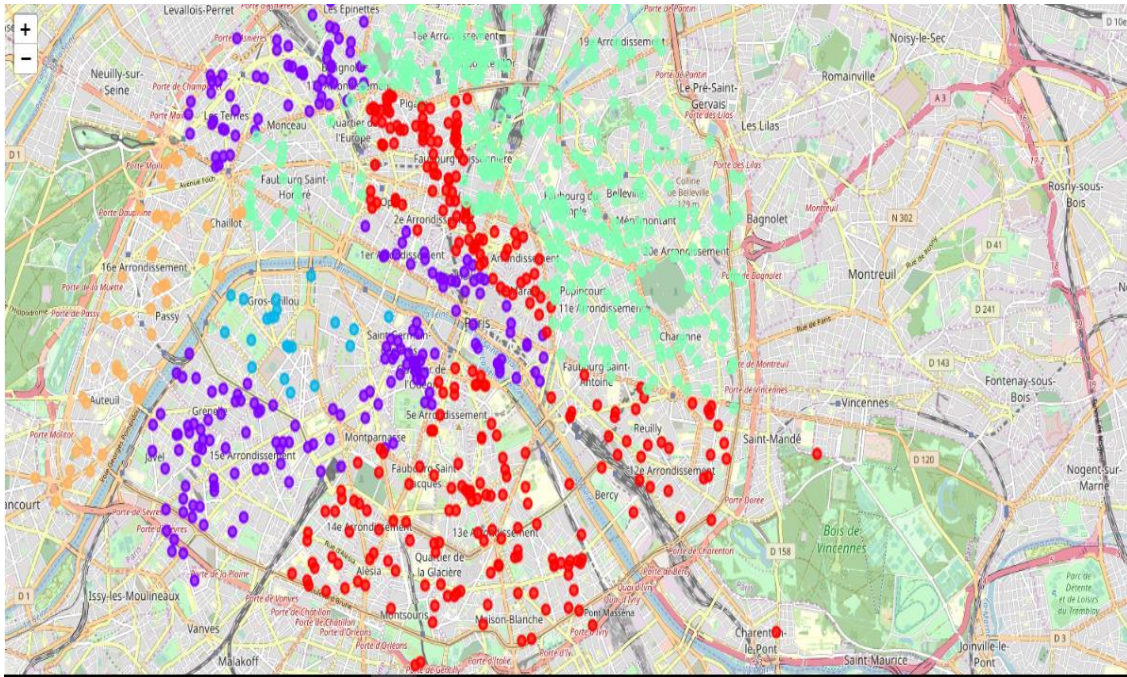
Fig 6: Clustered Asian Restaurants

The table below shows the mean value of the variables in each cluster.

Cluster Labels	Room ID	Room Price	Number of reviews	Latitude	Longitude
0	1.920898e+07	124.703008	27.146617	48.850011	2.351896
1	2.056484e+07	162.981132	28.669811	48.860463	2.316343
2	2.312264e+07	347.761905	24.809524	48.855525	2.308630
3	1.927157e+07	99.283186	21.396018	48.875030	2.366437
4	2.302862e+07	146.040816	18.163265	48.858356	2.273812

From the cluster, one can thus name each cluster based on the mean price of each group.

CLUSTER GROUP	PRICE GROUP
Cluster 1	Very Cheap
Cluster 2	Cheap
Cluster 3	Moderate
Cluster 4	Moderately Expensive
Cluster 5	Expensive

The tables below show sample AirBnB private room listings for each cluster.

CLUSTER 1

	Room ID	Neighborhood	Room Price	Number of reviews	Latitude	Longitude	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	2181753	Reuilly	40	104	48.841314	2.383398	Turkish Restaurant	Kebab Restaurant	Sushi Restaurant	Japanese Restaurant	Vietnamese Restaurant	Lebanese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Mediterranean Restaurant	Falafel Restaurant
2	29895295	Reuilly	50	7	48.847268	2.400412	Turkish Restaurant	Kebab Restaurant	Sushi Restaurant	Japanese Restaurant	Vietnamese Restaurant	Lebanese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Mediterranean Restaurant	Falafel Restaurant
7	12912440	Observatoire	40	1	48.837924	2.335897	Kebab Restaurant	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Cantonese Restaurant	Persian Restaurant	Caribbean Restaurant	Vietnamese Restaurant	Indian Restaurant
8	10902362	Gobelins	40	6	48.828511	2.368842	Chinese Restaurant	Vietnamese Restaurant	Asian Restaurant	Turkish Restaurant	Kebab Restaurant	Thai Restaurant	Middle Eastern Restaurant	Doner Restaurant	Halal Restaurant	Cambodian Restaurant

CLUSTER 2

	Room ID	Neighborhood	Room Price	Number of reviews	Latitude	Longitude	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
16	21585068	Hôtel-de-Ville	128	88	48.854200	2.351054	Burgundian Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Sushi Restaurant	Indian Restaurant	Vietnamese Restaurant	Thai Restaurant	Middle Eastern Restaurant	Falafel Restaurant	Mediterranean Restaurant
19	12318143	Vaugirard	50	13	48.846108	2.293216	Lebanese Restaurant	Korean Restaurant	Persian Restaurant	Indian Restaurant	Thai Restaurant	Mediterranean Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant	Doner Restaurant	Falafel Restaurant
20	6198789	Vaugirard	40	4	48.838193	2.308725	Lebanese Restaurant	Korean Restaurant	Persian Restaurant	Indian Restaurant	Thai Restaurant	Mediterranean Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant	Doner Restaurant	Falafel Restaurant
21	25371291	Luxembourg	100	49	48.856680	2.337738	Lebanese Restaurant	Indian Restaurant	Japanese Restaurant	Vietnamese Restaurant	American Restaurant	Middle Eastern Restaurant	Thai Restaurant	Indonesian Restaurant	Turkish Restaurant	Mediterranean Restaurant

CLUSTER 3

	Room ID	Neighborhood	Room Price	Number of reviews	Latitude	Longitude	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
33	28095621	Palais-Bourbon	69	61	48.848568	2.309096	Korean Restaurant	Thai Restaurant	Indian Restaurant	Spanish Restaurant	Eastern European Restaurant	Tapas Restaurant	Mediterranean Restaurant	Lebanese Restaurant	American Restaurant	Middle Eastern Restaurant
77	15005210	Palais-Bourbon	50	2	48.857104	2.303614	Korean Restaurant	Thai Restaurant	Indian Restaurant	Spanish Restaurant	Eastern European Restaurant	Tapas Restaurant	Mediterranean Restaurant	Lebanese Restaurant	American Restaurant	Middle Eastern Restaurant
99	16624675	Palais-Bourbon	280	4	48.853831	2.309337	Korean Restaurant	Thai Restaurant	Indian Restaurant	Spanish Restaurant	Eastern European Restaurant	Tapas Restaurant	Mediterranean Restaurant	Lebanese Restaurant	American Restaurant	Middle Eastern Restaurant
160	21589739	Palais-Bourbon	42	9	48.856704	2.319156	Korean Restaurant	Thai Restaurant	Indian Restaurant	Spanish Restaurant	Eastern European Restaurant	Tapas Restaurant	Mediterranean Restaurant	Lebanese Restaurant	American Restaurant	Middle Eastern Restaurant
162	34191655	Palais-Bourbon	199	0	48.853945	2.308219	Korean Restaurant	Thai Restaurant	Indian Restaurant	Spanish Restaurant	Eastern European Restaurant	Tapas Restaurant	Mediterranean Restaurant	Lebanese Restaurant	American Restaurant	Middle Eastern Restaurant

CLUSTER 4

	Room ID	Neighborhood	Room Price	Number of reviews	Latitude	Longitude	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	12817090	Entrepôt	30	8	48.873382	2.373455	Turkish Restaurant	Indian Restaurant	Asian Restaurant	Japanese Restaurant	Chinese Restaurant	Thai Restaurant	Middle Eastern Restaurant	Kebab Restaurant	Vegetarian / Vegan Restaurant	Szechuan Restaurant
3	30728070	Popincourt	139	0	48.860921	2.366604	Asian Restaurant	Indian Restaurant	Turkish Restaurant	Kebab Restaurant	Japanese Restaurant	Middle Eastern Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Mediterranean Restaurant	Thai Restaurant
4	21862193	Entrepôt	40	1	48.879021	2.370081	Turkish Restaurant	Indian Restaurant	Asian Restaurant	Japanese Restaurant	Chinese Restaurant	Thai Restaurant	Middle Eastern Restaurant	Kebab Restaurant	Vegetarian / Vegan Restaurant	Szechuan Restaurant
5	26834043	Élysée	1000	0	48.869121	2.309637	Turkish Restaurant	Indian Restaurant	Persian Restaurant	North Indian Restaurant	Thai Restaurant	Chinese Restaurant	Asian Restaurant	Halal Restaurant	Russian Restaurant	Kebab Restaurant

CLUSTER 5

	Room ID	Neighborhood	Room Price	Number of reviews	Latitude	Longitude	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
42	35154311	Passy	75	2	48.872736	2.281648	Japanese Restaurant	Chinese Restaurant	Middle Eastern Restaurant	Korean Restaurant	Thai Restaurant	Australian Restaurant	Lebanese Restaurant	Seafood Restaurant	Asian Restaurant	Persian Restaurant
47	28422062	Passy	119	0	48.848998	2.273493	Japanese Restaurant	Chinese Restaurant	Middle Eastern Restaurant	Korean Restaurant	Thai Restaurant	Australian Restaurant	Lebanese Restaurant	Seafood Restaurant	Asian Restaurant	Persian Restaurant
81	21973329	Passy	40	17	48.868741	2.283000	Japanese Restaurant	Chinese Restaurant	Middle Eastern Restaurant	Korean Restaurant	Thai Restaurant	Australian Restaurant	Lebanese Restaurant	Seafood Restaurant	Asian Restaurant	Persian Restaurant
84	24908135	Passy	30	1	48.838085	2.258432	Japanese Restaurant	Chinese Restaurant	Middle Eastern Restaurant	Korean Restaurant	Thai Restaurant	Australian Restaurant	Lebanese Restaurant	Seafood Restaurant	Asian Restaurant	Persian Restaurant

5. CONCLUSION & RECOMMENDATIONS

From the results, it is seen that the KNN machine learning algorithm is a good tool for clustering Asian restaurants which are in proximity to AirBnB private room listings in the city of Paris. As a recommendation, to make the outcome even much better, one can consider the review of each of these restaurant to help the Asian tourist determine the locations with the most positive reviews. The price of the rooms can be factored into the KNN model. Also, the distance of each restaurant from the AirBnB private apartment can be computed to further help improve the outcome of the model.