Opening an Indonesian restaurant in Strasbourg (France)

Coursera / IBM Applied Data Science Capstone Project

Prepared by: Mike Helde
Date: August 5th, 2020

1. Introduction and business problem

1.1. Background

I'm a French citizen and my hometown is Strasbourg, a mid-sized city in the North-East of France. I've been living in Indonesia for the past eight years and thoroughly enjoy Indonesian cuisine. In the past years, during my home trips to France I noticed there is a growing number of Asian restaurants, with new ones opening every year. However I also noticed there is no Indonesian restaurant in my hometown whereas Chinese, Japanese, Vietnamese and Thai restaurants are blooming.

So I decided to apply the skills learned during the Coursera Applied Data Science Specialization to find what would be the optimal area to open an Indonesian restaurant in my hometown of Strasbourg.

1.2. Problem to solve

The challenge is to find an area that fits following characteristics:

- High income area so more likely to eat out in restaurant
- Populated area to ensure potential customers are living closeby
- Sufficient restaurant density to ensure customer traffic in the area

1.3. Interest

This can be of interest for anyone looking to open a new restaurant in Strasbourg, to support the decision making. Also this a good practical case for those interested in using the Foursquare API, looking for statistics in France as well as Data Science students.

2. Data

In France detailed data regarding population is being compiled by the INSEE, which is the French National Institute of Statistics, based on their census data.

2.1. Required data

Here I will list all the data that is relevant to perform the analysis:

- Proper breakdown of the city in relevant areas: for midsize cities in France, unfortunately there is no area breakdown by postocode (as for example in New York or Toronto). The INSEE has broken down all the French territory into "IRIS" areas which are clusters sized for relevant statistical analysis, and are the basis of most statistical analysis in France
- Mean income per household by area
- Population of the area
- Geographical coordinates of each area
- Restaurant data for each area

2.2. Data sources

As mentioned, the INSEE is collecting all the census data such as population or income, however they are not very good at making this detailed data available to the public. Their website doesn't have many free datasets that are relevant for our analysis. After some research I found that all the INSEE data is available and accessible in user-friendly manner on the Opendatasoft website, which offers lots of public data at following URL : https://public.opendatasoft.com/ This data can be easily exported as csv file, which is what we will do later on.

For restaurant data the Foursquare API will be used.

2.3. Data Preparation

2.3.1. Prepare the statistical area data

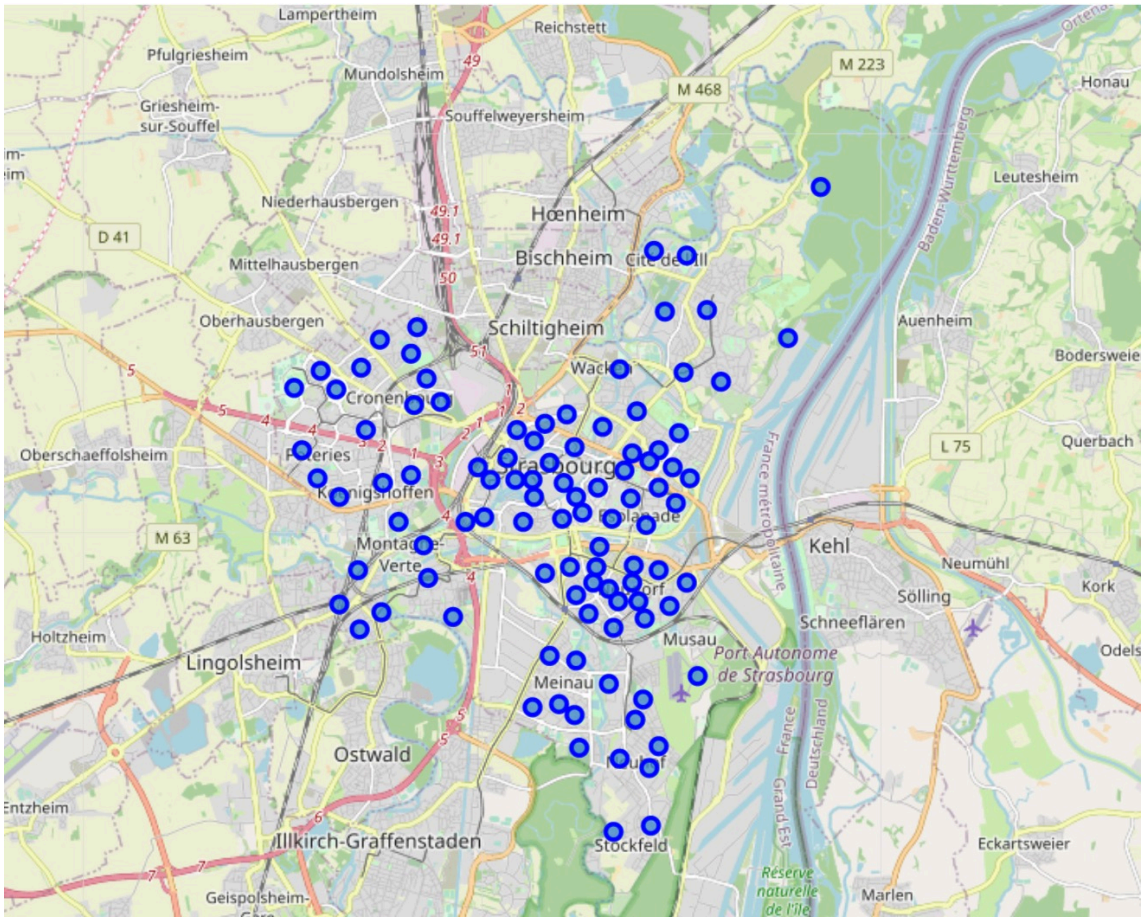
Clean and prepare the following statistical data from INSEE / Opendatsoft:

- Area codes and names using the IRIS breakdown (as not postcode breakdown is possible in our case)
- Population per area
- Mean income per area
- latitude and longitude data per area

We obtain following dataframe (here I only show the first five lines):

Area_code	Mean_income	Area_name	Population	lat	lon
674822104	28728.0	Robertsau Centre	2146.842963	48.607300	7.783869
674821001	16056.0	Esplanade Sud Est	3365.658019	48.575999	7.770276
674823001	18924.0	Stockfeld Est	2267.729980	48.531956	7.771555
674822106	29956.0	Robertsau Est	2638.509955	48.603239	7.801769
674820202	22341.0	Petite France Nord Ouest	2292.148315	48.582471	7.741456

Let's map all these areas:



2.3.2. Prepare the restaurant data by area

Using the Foursquare API, we get the food venues of each area within a radius of 500 meters (limiting to maximum 100 venues per area), by creating a function that:

- extracts name, latitude and longitude of an areas from our dataset
- inputs the client credentials, name, latitude, longitude to the Foursquare API to explore restaurants within 500m radius of the neighbourhood
- restrict the API call to Food venues category, which according to Foursquare's API documentation has the category id of '4d4b7105d754a06374d81259' (note: there is no specific category for restaurants only)
- retrieves restaurants data from the Json code that Foursquare API outputs
- creates a dataframe of all the relevant information required to analyze the restaurants

Running the function on our dataset we get following dataframe (extract of the first five lines):

Area_name	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Robertsau Centre	48.6073	7.783869	La Vignette	48.603874	7.785202	French Restaurant
Robertsau Centre	48.6073	7.783869	Restaurant Au Coq Blanc	48.603501	7.783838	French Restaurant
Robertsau Centre	48.6073	7.783869	Le Joyeux Pécheur	48.605472	7.778098	French Restaurant
Robertsau Centre	48.6073	7.783869	Le Violon d'Ingres	48.603929	7.782247	Diner
Robertsau Centre	48.6073	7.783869	Boulangerie Patisserie Materne	48.604399	7.782040	Bakery

By checking the resulting dataframe we find there are 1,322 food venues in the areas from our dataset. Let's see what different types of foods Strasbourg has to offer:

```
array(['French Restaurant', 'Diner', 'Bakery', 'Pizza Place',
       'Sandwich Place', 'Middle Eastern Restaurant', 'Taco Place',
       'Fast Food Restaurant', 'Gastropub', 'Doner Restaurant',
       'Brasserie', 'Steakhouse', 'Restaurant', 'Brazilian Restaurant',
       'Asian Restaurant', 'Belgian Restaurant', 'Café',
       'Italian Restaurant', 'Spanish Restaurant',
       'Vietnamese Restaurant', 'Alsatian Restaurant', 'Bistro',
       'Vegetarian / Vegan Restaurant', 'Mediterranean Restaurant',
       'Sushi Restaurant', 'Cigkofte Place', 'Bagel Shop',
       'Thai Restaurant', 'Japanese Restaurant', 'Tapas Restaurant',
       'German Restaurant', 'Trattoria/Osteria', 'Burger Joint',
       'Chinese Restaurant', 'Mexican Restaurant', 'Indian Restaurant',
       'Cafeteria', 'Food Truck', 'Comfort Food Restaurant',
       'Snack Place', 'Fried Chicken Joint', 'Dim Sum Restaurant',
       'Korean Restaurant', 'Lebanese Restaurant', 'Mac & Cheese Joint',
       'Deli / Bodega', 'Paella Restaurant', 'Kebab Restaurant',
       'BBQ Joint', 'Food Court', 'Falafel Restaurant',
       'Fish & Chips Shop', 'Turkish Restaurant', 'Portuguese Restaurant',
       'American Restaurant', 'Moroccan Restaurant', 'Creperie', 'Food',
       'Greek Restaurant'], dtype=object)
```

Interestingly we can see that there is no Indonesian restaurant so we would be the first one! That’s promising!

3. Methodology

3.1. Exploratory data analysis

In this section we will explore our data. First I want to identify whether a venue is Asian or Non Asian. From the category list above we can define all Asian restaurant categories in Strasbourg as being following:

- Asian Restaurant
- Vietnamese Restaurant
- Sushi Restaurant
- Thai Restaurant
- Japanese Restaurant
- Chinese Restaurant
- Korean Restaurant

After merging the statistical / population data with the restaurants data, we can create a summary dataframe showing the following (extract of the first five lines):

Area_name	Asian	Non Asian	Area_code	Mean_income	Population	lat	lon
Canardiere Est Est	1	3	674822702	7052.0	3044.921472	48.548277	7.754681
Canardiere Ouest Est	1	4	674822602	13714.0	2167.715969	48.549848	7.751284
Canardiere Ouest Ouest	1	3	674822601	19106.0	3843.497615	48.549290	7.745445
Cite de l'Ill Est	0	3	674822002	12136.0	1845.866862	48.615211	7.779522
Cite de l'Ill Ouest	0	1	674822001	10358.0	2746.380994	48.615941	7.772018

Let's check the statistics of this dataframe:

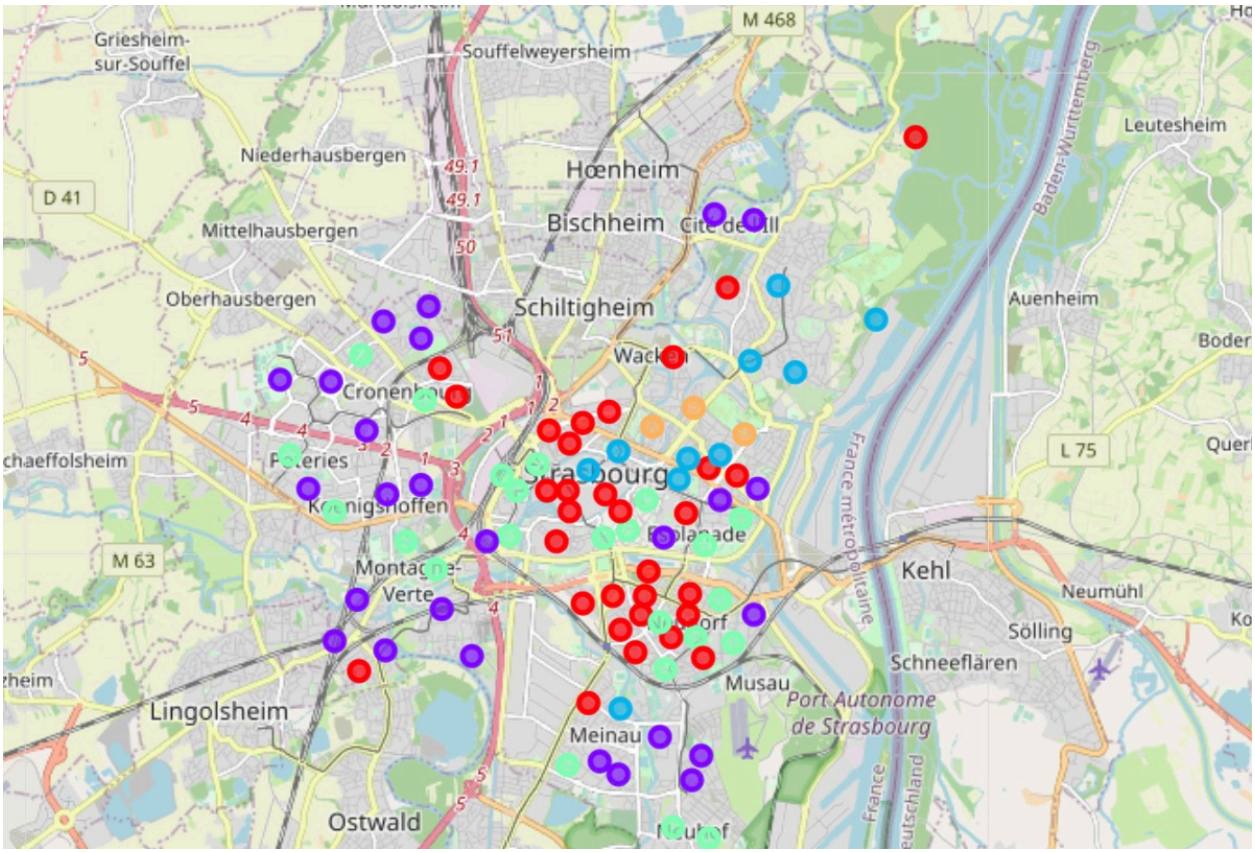
	Asian	Non Asian	Area_code	Mean_income	Population	lat	lon
count	94.000000	94.000000	9.400000e+01	94.000000	94.000000	94.000000	94.000000
mean	1.265957	12.797872	6.748215e+08	19456.585106	2670.143222	48.578593	7.749775
std	2.166090	17.994372	8.071956e+02	6979.877591	708.482747	0.017467	0.025801
min	0.000000	1.000000	6.748201e+08	5584.000000	1365.319345	48.531956	7.692865
25%	0.000000	4.000000	6.748208e+08	13898.000000	2141.375251	48.567420	7.731384
50%	0.000000	5.000000	6.748214e+08	19896.000000	2562.127392	48.580072	7.754920
75%	1.000000	12.000000	6.748221e+08	23033.000000	3161.780208	48.589633	7.768567
max	8.000000	93.000000	6.748230e+08	40438.000000	5214.369785	48.625189	7.808822

From this dataset we can see that overall in Strasbourg for each area:

- Mean income per person in an area is 19,456 EURO yearly income
- Mean population of an area is 2,670 inhabitants
- Mean number of Asian restaurants per area is 1.27
- Mean number of Non Asian restaurants per area is 12.8

3.2. Clustering of the areas

We will run k-means clustering with 5 clusters after normalizing the data. Let's see the resulting clusters:



Let's make a summary of the data to see the characteristics of each cluster:

Cluster Labels	Area_name	Asian	Non Asian	Mean_income	Population
0	31	61	636	22713.935484	76350.318941
1	26	9	120	10723.730769	71045.257264
2	10	8	110	29032.500000	24785.413978
3	24	40	317	18434.958333	69242.979112
4	3	1	20	37735.333333	9569.493565

What we can say about the clusters:

- **Cluster 0:** the most populated cluster with 31 areas and 76,350 inhabitants, income of 22,713 Euros is above the city's mean income 19,456 Euros, and the highest number of restaurants → **Good potential cluster**
- **Cluster 1:** the mean income is too low for our case
- **Cluster 2:** a smaller cluster with less than 25,000 people however there is a very high income of 29,032 Euros, almsot 10,000 Euros higher than the mean income in the city → **Good potential cluster**
- **Cluster 3:** the mean income is too low as we are looking for lcuster above the mean income of the city
- **Cluster 4:** a very small cluster with less than 10,000 inhabitants but a the highest mean income in the city, also there is only one Asian restaurant in this cluster → **Good potential cluster**

4. Results

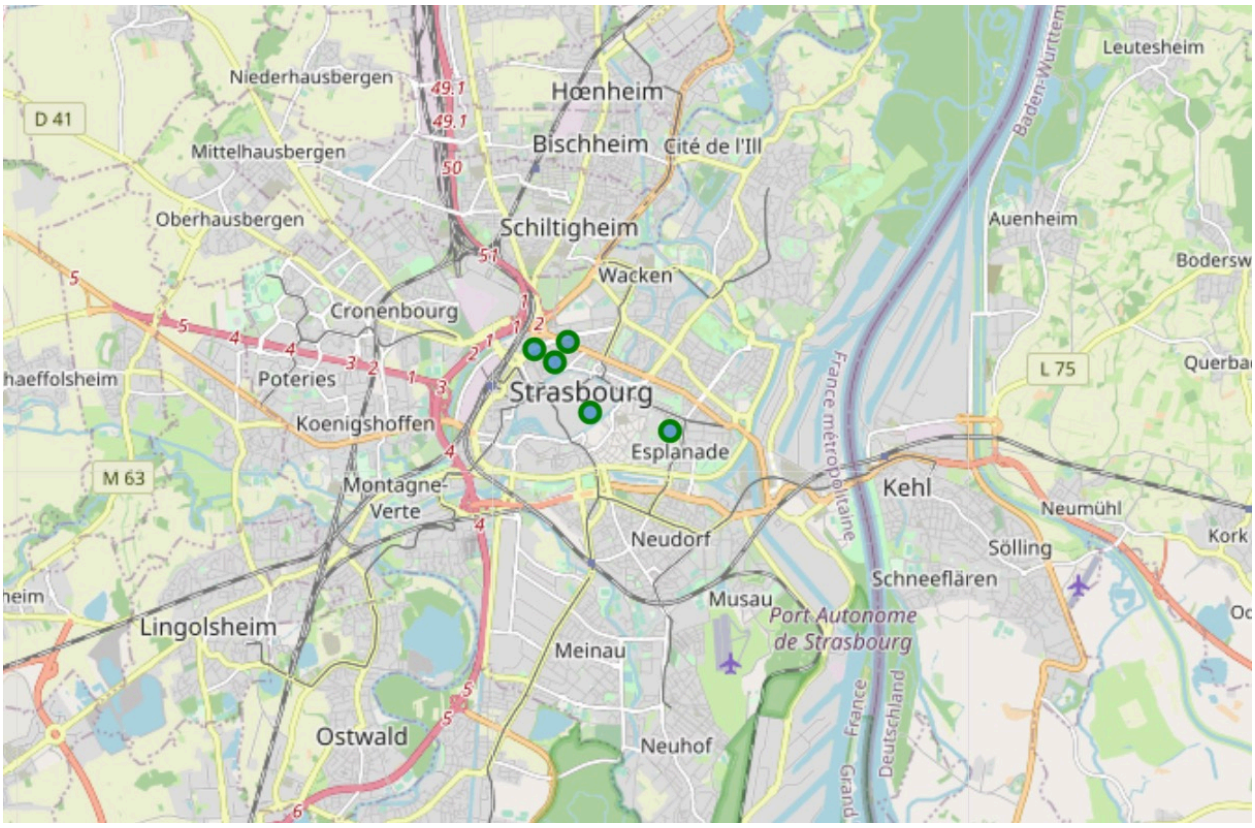
Based on the performed exploratory analysis and clustering, we can tell the best areas in which to open an Indonesian restaurant in Strasbourg have following characteristics:

- Mean income bigger than the city's mean income
- Population of the area bigger than the mean area population
- Density of Non Asian restaurants in the area bigger than the city's mean density (to ensure there is enough customer traffic and validate the area population is eating out)
- Area is located in clusters 0, 2 or 4

So let's list the areas that fulfill these criterias:

Cluster Labels	Area_name	Asian	Non Asian	Area_code	Mean_income	Population	lat	lon
0	Esplanade Nord Ouest	1	16	674821005	22360.0	2746.041336	48.579749	7.766852
0	Kable Sud Ouest	2	19	674820503	23270.0	2962.160086	48.590608	7.748234
0	Mairie Sud	5	69	674820102	24812.0	3741.745583	48.582155	7.752335
0	Poincare Est	4	15	674820402	22446.0	2975.528259	48.588116	7.745589
0	Poincare Ouest	6	26	674820401	23068.0	3754.317640	48.589735	7.741896

We found 5 areas fulfilling our requirements. Let's map them out on the city map:



The 5 selected areas are pretty much downtown and in the nicer neighbourhoods of Strasbourg.

5. Discussion

Our analysis led us to define what are the best potential areas to open a new type of restaurant (in our case Indonesian) in a mid-size city (in our case Strasbourg in France, approximately 277,000 inhabitants).

Unlike bigger cities the city can not be analyzed by simply using postal code breakdown, as this data isn't granular enough, so we had to find an alternative segmentation that is more appropriate. In our case this happened to be the IRIS areas breakdown defined by the INSEE, French Institute of Statistics, which ahd the advantage to have lots of interesting data from census (such as population and income) that are accessible using thord party website.

Based on the available data we first explored some basic characteristics of each area and ran a k-clustering to see some pattern that can't be identified by looking at the data manually. Based on the different clusters we then identified which areas are fulfilling all

our criterias and as a result ended up with a shortlist of the 5 best areas where to open an Indonesian restaurant in Strasbourg (France):

- Esplanade Nord Ouest
- Kable Sud Ouest
- Mairie Sud
- Poincare Est
- Poincare Ouest

6. Conclusion

Purpose of this project was to identify the best area in Strasbourg (France) to open an Indonesian restaurant, which would be the first one in the city. Based on our analysis we identified 5 potential areas fulfilling our criterias defined in the methodology part. This analysis is very helful to pre-screen the potential sites and allows us to speed up the process of identifying the perfect location, as we already identified in which areas it's bet to open an Indonesian restaurant.

This is a good starting point for exploring further these pre-selected areas. The decision of the precise location needs to be made also based on other factors such as real estate availability, rental prices, accessibility of the location.