Capstone Project – Open a coffee shop in Paris

1. Introduction

1.1 Background

I've decided to study the coffee shops in Paris. Why Paris and why the coffee shops?
First of all, I'm french and when I go in Paris, I always see the coffee shops crowded.
Paris is the center of French economy, politics, traffic and culture. Paris is one of the most multicultural cities. It's the capital and the largest city of France. Indeed, Parisian coffees serve as a center of social and culinary life in Paris. They have existed since the 17th century, and serve as the meeting place, neighborhood hub, conversation matrix, meeting spot, and networking source, a place to relax or to refuel - the social and political pulse of the city. Parisian coffees show the Parisian way of sitting undisturbed for a couple of hours, watching things happening and people going by.

According to the SOFRES survey 13% of the French population visit several times a week a coffee shop (which 4% visit almost daily). Also, the survey shows the kind of people who frequent coffee shops. They are primarily men, young people and workers. So, we can see that all sections of the population go to the coffee shops and especially active people.

1.2 **Problem**

If someone is looking to open a coffee shop in Paris, it is interesting to study many features very important. Before creating any business, we have to look at it three different standpoints: marketing, financial and operational.
One of the most important feature is the location. Where will coffee shop be? So, the mainly problem is: *Where can I open a coffee shop in Paris?*

Here our target are the active people. We can study three important points:

- the rental cost according to the place
- the place activity
- the potential customers attracted by coffee shops
- the number of coffee shops in each borough to study the potential competition in each borough

1.3 **Interest**

This study in very interesting for all people who looking to open a coffee shop in Paris. Also, I think it can be interesting for people who just looking for open a business in Paris because in this study I'll give many information about boroughs in Paris. These information are interesting for someone who looking to open other businesses like restaurant, store and other in Paris.

2. Data acquisition and description

2.1 Data sources

I need the average commercial rental price in Paris.
I've got the data in this website: https://www.localcommercial.net/estimerloyer/38080/paris-20-arrondissement.html
Then, I created a csv file that contains postal code and average commercial rental price.

With foursquare API, I've got:
- The coffee shop with their address, postal code, name, longitude and latitude.
- The nearby venues in Paris to see many information:
Where are the active places in Paris?
What are the most common venues in terms of borough?
What are the boroughs where coffee is a most common venue?
When coffee is a most common venue in a borough, what are the other most common venues?
What is the kind of people who frequent each borough?

I've got the features to create a map with the borough of Paris with a geoJSON file:
arrondissements.geojson downloaded in this page :
https://opendata.paris.fr/explore/dataset/arrondissements/export/?disjunctive.c_ar&disjunctive.c_arinsee&disjunctive.l_ar&location=13,48.85156,2.32327

2.2 Data creation

I created a dataframe with information about coffee shops in Paris with foursquare API. I've got one hundred coffee shops with the “get venue recommendations endpoint”. I think it's interesting to get the coffees with recommendations to take example of business that work. So, I didn't take coffee without recommendations. I think that to open a business that work, we have to take example of good coffee shops. With the csv with average commercial rental price, I created a dataframe which contains the average commercial rental price depending on the borough in Paris.
I merged this two dataframes to one dataframe named *merged_inner.* I created another dataframe named *dfcount* which contains the number of coffee shops for each borough in Paris. I added to the *dfcount* the column with the average commercial rental price. I did this because I was looking to see if there had any relationship between the rental price and the number of coffee shop in the place.
I also created a df with the geographic information about the borough in Paris with the geojson file.
I got these informations to have a good vision of each borough in a choropleth map.

To answer to the problem *Where can I open a coffee shop in Paris*? an essential thing is to study each borough. To open a business you have to know many things about each place, like people who go in this place to study the potential clients of my business. Who is the target, what kind of clients can be attracted to my business according to the place ? To study each borough, I created a dataframe named *paris_venues.* I got this data with foursquare API, I got all venues in each borough. Then I got the most common venues in each borough to have an idea about people who haunting each place and see where the business could work.

I created a dataframe named *venues_sorted* with the ten most common venues in each

borough. I join this dataframe with the *merged_inner* dataframe where we have all information about each coffee to see what are the top ten of most common places where we have as much coffee shops as possible.

2.3 Feature selection

With foursquare API I got many features but I selected only the interesting features for my analyze. So, when I've got the hundred coffee shops with recommendations in Paris, I selected the name of a coffee, the latitude, the longitude, the postal code and the address. I also selected the city if I have another city by error. The geographic information about these coffees allows us to view there in the map and to link with other dataframe when I've got features about each borough. I added the average commercial rental price by m2 to this df for each borough. I divided by twelve the rent to get the rent by month because I think that it's more significant than by year.

Coffee shop	Name of coffee shop
Coffee Latitude	Geographical information about the coffee shop
Coffee longitude	Geographical information about the coffee shop
Postal Code	The postal code of the coffee shop
City	The city of the coffee shop
Address	The address of the coffee shop
average commercial rental price by m2	The average commercial rental price by month by m2 of the coffee shop

When I created the *paris_venues* dataframe I selected the name, the latitude, the longitude, the postal code, the city, the address and the venue category.

Venues	The name of venue
Venues Latitude	The geographic location of venue
Venues Longitude	The geographic location of venue
Venues Postal code	The borough of venue
Venues City	The city of venue
Venues Address	The address of venue
Venues Category	The category name of venues (example: coffee shop)

I've selected the geographic information about the venues to see what is the most activity in each borough, what kind of people go in each borough, what are the boroughs where coffee shops work well, where are the active place?