Deciding where to open vegetables store in Paris

Introduction

For my Capstone project I will help an investor who is about to start a business of importing and selling BIO: vegetables and fruits.

Description and background

In the past few years my client has tried to start his dream business but he failed many times. So this time he consulted me for help and as a data scientist I decided to start by describing my approach.

NB: our client wants to open his Bio store in Paris city but no matter which neighbourhood.
- First: we will search a list of all paris neighbourhoods.
- Second: we will sort all neighbourhoods by the number of restaurants and fast foods
- Then the Third step: is to inverse sort neighbourhoods by the number of store and supermarkets
- The last step is to reverse order our data by the population

Data of our project

This project will analyze neighborhoods of Paris city. The data below will be used for this analysis.

Boroughs and neighborhoods in Paris:

Paris has in total 20 neighborhoods (called arrondissements in French). To explore, analyze and segment neighborhoods, longitude and latitude of each neighborhood will be added.

This dataset exists for free on the web.

Links to the dataset are:

Paris population:

Number of populations in each borough will be analyzed to find the most populous borough in Paris (more populous Borough = Higher gain).

The dataset that exists in the same link above

Restaurants and markets in Paris:

This data will cover different restaurants that already exist in Paris and different food markets.

The JSON file has data about all the neighbourhoods in France; we will limit it to Paris using department code = 75. columns we will be using are:
1. postal_code: Postal codes for France
2. nom_comm: Name of Neighborhoods in France
3. nom_dept: Name of the boroughs, equivalent to towns in France
4. geo_point_2d: Tuple containing the latitude and longitude of the Neighborhoods.

Foursquare API

In this next step we will be using Foursquare data api to search for the number of restaurants in every Paris area. And to have better idea on our concurrents we will look for the number of stores or supermarkets in all those areas

Methodology

Business understanding:

In this project we are helping a businessman opening his vegetables store in paris, but he cannot decide where is the best location or neighborhood.
NB: Paris has 20 Boroughs

Exploratory Data Analysis:

population of Paris:

This data is contained in a JSON file that is downloaded from link:

https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259
114e

This data is also cleaned to extract the necessary of information and processed like this:

- Transform the data of nested python dictionaries into a data frame.
- This data frame contains geographical localization and population by 1000 of each borough in Paris.

	code_comm	nom_dept	statut	nom_region	code_reg	code_dept	geo_point_2d	postal_code	nom_comm	population
0	645	ESSONNE	Commune simple	ILE-DE-FRANCE	11	91	[48.750443119964764, 2.251712972144151]	91370	VERRIERES-LE-BUISSON	15.5
1	133	SEINE-ET-MARNE	Commune simple	ILE-DE-FRANCE	11	77	[48.41256065214989, 3.052940505560729]	77126	COURCELLES-EN-BASSEE	0.2
2	378	ESSONNE	Commune simple	ILE-DE-FRANCE	11	91	[48.52726809075556, 2.19718165044305]	91730	MAUCHAMPS	0.3
3	243	SEINE-ET-MARNE	Chef-lieu canton	ILE-DE-FRANCE	11	77	[48.87307018579678, 2.7097808131278462]	77400	LAGNY-SUR-MARNE	20.2
4	414	SEINE-ET-MARNE	Commune simple	ILE-DE-FRANCE	11	77	[48.62891464105825, 3.2582355268439223]	77160	SAINT-HILLIERS	0.4

This table shows a dataframe that contains different elements for each borough in Paris. Each borough in Paris is determined by a number (from 1 to 20). You will find in the notebook the definition of each column in the table.

- This data will be used to get venues data from Foursquare.
- Folium is the library that we will use to create a map of Paris with marks for each borough.

Results

Count restaurants:

We call Foursquare Explore API and look for venues inside the food category (Restaurants, steakhouse, pizzeria …)

```python
food_category_id = '4d4b7105d754a06374d81259,4bf58dd8d48988d1ca941735,4bf58dd8d48988d1cc941735'
count_restaurants = {"Neighborhood":[], "count_restaurants":[]}
for lat, lng, Neighborhood in zip(paris_df['Latitude'], paris_df['Longitude'], paris_df["Neighborhood"]):
    url='https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}&ca
        CLIENT_ID,CLIENT_SECRET,VERSION,
        lat,lng,
        radius, LIMIT, food_category_id)
    results = requests.get(url).json()
    venues = results["response"]["groups"][0]["items"]
    count_restaurants["Neighborhood"].append(Neighborhood)
    count_restaurants["count_restaurants"].append(len(venues))
```

Then we get the count of restaurants in each Paris borough and we add new field to our dataframe

	Latitude	Longitude	City	Neighborhood	postal_code	population	count_restaurants
0	48.876896	2.337460	PARIS	PARIS-9E-ARRONDISSEMENT	75009	60.3	100
1	48.867903	2.344107	PARIS	PARIS-2E-ARRONDISSEMENT	75002	22.4	100
2	48.859415	2.378741	PARIS	PARIS-11E-ARRONDISSEMENT	75011	152.7	45
3	48.872527	2.312583	PARIS	PARIS-8E-ARRONDISSEMENT	75008	40.3	41
4	48.828718	2.362468	PARIS	PARIS-13E-ARRONDISSEMENT	75013	182.0	62
5	48.835156	2.419807	PARIS	PARIS-12E-ARRONDISSEMENT	75012	142.9	5
6	48.863054	2.359361	PARIS	PARIS-3E-ARRONDISSEMENT	75003	35.7	86
7	48.848968	2.332671	PARIS	PARIS-6E-ARRONDISSEMENT	75006	43.1	33
8	48.854228	2.357362	PARIS	PARIS-4E-ARRONDISSEMENT	75004	28.2	100
9	48.876029	2.361113	PARIS	PARIS-10E-ARRONDISSEMENT	75010	95.9	100
10	48.860399	2.262100	PARIS	PARIS-16E-ARRONDISSEMENT	75016	169.4	2
11	48.844509	2.349859	PARIS	PARIS-5E-ARRONDISSEMENT	75005	61.5	79
12	48.886869	2.384694	PARIS	PARIS-19E-ARRONDISSEMENT	75019	184.8	31
13	48.863187	2.400820	PARIS	PARIS-20E-ARRONDISSEMENT	75020	197.1	48
14	48.856083	2.312439	PARIS	PARIS-7E-ARRONDISSEMENT	75007	57.4	73
15	48.892735	2.348712	PARIS	PARIS-18E-ARRONDISSEMENT	75018	200.6	51
16	48.887337	2.307486	PARIS	PARIS-17E-ARRONDISSEMENT	75017	168.5	64
17	48.840155	2.293559	PARIS	PARIS-15E-ARRONDISSEMENT	75015	236.5	61
18	48.862630	2.336293	PARIS	PARIS-1ER-ARRONDISSEMENT	75001	17.6	82
19	48.828993	2.327101	PARIS	PARIS-14E-ARRONDISSEMENT	75014	137.2	24

Count stores

In this step we will use Foursquare to count the number of Fruits and vegetables stores for each borough in Paris city

```
food_category_id = '52f2ab2ebcbc57f1066b8b1c'
counts_stores = {"Neighborhood":[], "count_stores":[]}
for lat, lng, Neighborhood in zip(paris_df['Latitude'], paris_df['Longitude'], paris_df["Neighborhood"]):
    url='https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}&ca
        CLIENT_ID,CLIENT_SECRET,VERSION,
        lat,lng,
        radius, LIMIT, food_category_id)
    results = requests.get(url).json()
    venues = results["response"]["groups"][0]["items"]
    counts_stores["Neighborhood"].append(Neighborhood)
    counts_stores["count_stores"].append(len(venues))
```

and as you can see in the next table, our final dataframe sorted in descending order of population and ascending order of restaurants and stores

	Latitude	Longitude	City	Neighborhood	postal_code	population	count_restaurants	count_stores
17	48.840155	2.293559	PARIS	PARIS-15E-ARRONDISSEMENT	75015	236.5	61	3
15	48.892735	2.348712	PARIS	PARIS-18E-ARRONDISSEMENT	75018	200.6	51	2
13	48.863187	2.400820	PARIS	PARIS-20E-ARRONDISSEMENT	75020	197.1	48	4
12	48.886869	2.384694	PARIS	PARIS-19E-ARRONDISSEMENT	75019	184.8	31	2
4	48.828718	2.362468	PARIS	PARIS-13E-ARRONDISSEMENT	75013	182.0	62	1
10	48.860399	2.262100	PARIS	PARIS-16E-ARRONDISSEMENT	75016	169.4	2	0
16	48.887337	2.307486	PARIS	PARIS-17E-ARRONDISSEMENT	75017	168.5	64	1
2	48.859415	2.378741	PARIS	PARIS-11E-ARRONDISSEMENT	75011	152.7	45	6
5	48.835156	2.419807	PARIS	PARIS-12E-ARRONDISSEMENT	75012	142.9	5	0
19	48.828993	2.327101	PARIS	PARIS-14E-ARRONDISSEMENT	75014	137.2	24	1
9	48.876029	2.361113	PARIS	PARIS-10E-ARRONDISSEMENT	75010	95.9	100	3
11	48.844509	2.349859	PARIS	PARIS-5E-ARRONDISSEMENT	75005	61.5	79	4
0	48.876896	2.337460	PARIS	PARIS-9E-ARRONDISSEMENT	75009	60.3	100	3
14	48.856083	2.312439	PARIS	PARIS-7E-ARRONDISSEMENT	75007	57.4	73	0
7	48.848968	2.332671	PARIS	PARIS-6E-ARRONDISSEMENT	75006	43.1	33	0
3	48.872527	2.312583	PARIS	PARIS-8E-ARRONDISSEMENT	75008	40.3	41	0
6	48.863054	2.359361	PARIS	PARIS-3E-ARRONDISSEMENT	75003	35.7	86	2
8	48.854228	2.357362	PARIS	PARIS-4E-ARRONDISSEMENT	75004	28.2	100	3
1	48.867903	2.344107	PARIS	PARIS-2E-ARRONDISSEMENT	75002	22.4	100	3
18	48.862630	2.336293	PARIS	PARIS-1ER-ARRONDISSEMENT	75001	17.6	82	0

Conclusion

Final Step

After all those researches we will provide to our client a list of all neighbourhoods with more restaurants and less concurrency and of course the results will be sorted and we let our customer choose where he will be opening his business.

Décision:

From all the above results we can see that the best places to open the new business are:
● PARIS-15E-ARRONDISSEMENT
● PARIS-18E-ARRONDISSEMENT
● PARIS-20E-ARRONDISSEMENT

Due to the huge number of population and restaurants compared to the number of stores.

Or we can also choose neighborhood where the number of stores is 0 or 1 sow there will be less or no competition Like:

- PARIS-16E-ARRONDISSEMENT
- PARIS-12E-ARRONDISSEMENT
- PARIS-13E-ARRONDISSEMENT