Opening shopping mall in Paris

1. Introduction

For many shoppers, visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. Property developers are also taking advantage of this trend to build more shopping malls to cater to the demand. As a result, there are many shopping malls in the city of Paris and many more are being built. Opening shopping malls allow property developers to earn consistent rental income. As with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

2. Business Problem

The objective of this project is to analyze and select the best locations in the city of Paris, France, to open a new shopping mall. This project is mainly focused on geospatial analysis of the Paris City to understand which would be the best place to open a new mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Paris, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

3. Sources of Data and methods to extract the Data.

This [Wikipedia page](#) is a list of neighborhoods in Paris, with 29 neighborhoods. I have used web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautiful soup packages. Then we can get the latitude and longitude coordinates of the neighborhoods using Python Geocoder package. After that, I have used the Foursquare API to get the venue data for those neighborhoods. Foursquare API will provide many categories of the venue data, and we are particularly interested in the Shopping Mall category to help us solve the business problem. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

3.1 Foursquare API Data

In order to get the location data for each neighborhood, we will be using Foursquare to get information about the venues within an area of interest.

After scraping the data from the Wikipedia pages, we then use Foursquare to gather information about venues within a certain distance from each neighborhood. We will also use Foursquare to get the number of shopping malls in each neighborhood. For this project, we have chosen the radius to be 2000 meters.

Based on the information retrieved for Paris, we will cluster neighborhoods based on similar venue categories.

4. Methodology

First, we get the list of neighborhoods in the City of Paris from the [Wikipedia page](#). Second, we scrape the data using the requests library, Beautiful Soup packages, and pandas. We use python requests library to get HTML content from the website, Beautiful Soup packages to parse HTML documents, and pandas to read HTML tables on the website into a data frame.

Third, we use the geocoder package to retrieve the location coordinates of each neighborhood. Next, we use the Foursquare API to get the top 100 venues within 2000 meters. We pass the neighborhood's latitude and longitude to the API and extract the venue's name, latitude and longitude, and category. Then we analyze each neighborhood and group them based on the frequency of each venue category. In our project, we will group by the number of shopping malls in each neighborhood. Hence our interest of venue category will be "Shopping Plaza."

Lastly, we will cluster the data using k-means clustering. We will cluster the neighborhoods into 2 clusters based on the frequency of shopping malls in each neighborhood. The clusters would show which neighborhoods have higher number and which neighborhoods have lower number of shopping malls. By looking at the results, we can determine which would be the best location to open a shopping mall.

5. Results

By using the k-means clustering, we can classify our list of neighborhoods into 3 clusters. From our results, we can see that:

cluster 0 contains all the places which do not have a shopping mall

[23]:		Neighborhood	Shopping Plaza	Cluster Labels	Latitude	Longitude
	0	Batignolles	0	0	48.883330	2.316670
	25	The Marais	0	0	48.858680	2.360980
	24	Saint-Germain-des-Prés	0	0	48.853770	2.333310
	23	Quartier des Grandes-Carrières	0	0	48.890880	2.331020
	22	Quartier de La Chapelle	0	0	48.884149	2.357046
	21	Quartier Pigalle	0	0	48.882026	2.337575
	20	Quarters of Paris	0	0	48.857170	2.341400
	19	Petit-Montrouge	0	0	48.826420	2.325200
	18	Passy	0	0	46.540550	4.540730
	17	Paris Rive Gauche	0	0	48.831880	2.339530
	16	Ménilmontant	0	0	48.863171	2.387277
	15	Montparnasse	0	0	48.843130	2.321290
	14	Montmartre	0	0	48.886170	2.337970
	13	Les Halles	0	0	48.863190	2.342010
	12	Latin Quarter, Paris	0	0	48.847640	2.351730
	11	Javel, France	0	0	48.843910	2.286180
	10	Jardin de Tivoli, Paris	0	0	48.873540	2.334590
	9	Grenelle	0	0	48.848420	2.291640
	8	Goutte d'Or	0	0	48.885040	2.353950
	7	Front de Seine	0	0	48.849350	2.285730
	6	Faubourg Saint-Germain	0	0	48.857815	2.323802
	5	Faubourg Saint-Antoine	0	0	48.850940	2.375670
	4	Cour des miracles	0	0	46.100403	4.323412
	3	Cité des Fleurs	0	0	48.892615	2.320325
	1	Belleville, Paris	0	0	48.870180	2.384230
	26	Trocadéro	0	0	48.862340	2.288070
	27	Épinettes	0	0	48.842963	2.325298

Cluster 1 contains a one places and all of them contain exactly 1 shopping

	Neighborhood	Shopping Plaza	Cluster Labels	Latitude	Longitude
2	Bercy	1	1	48.83488	2.38459

Below is the map showing the three clusters with cluster 0 in red, cluster 1 in purple
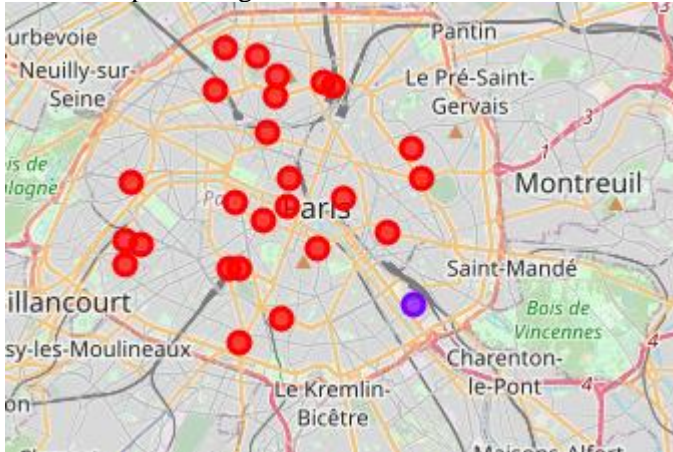
6. Discussion

By looking at the results from k-means clustering, we can see that no found any shopping malls are in the central area of Paris and a one shopping mall are located in the Bercy neighborhood. Cluster 0 has no shopping malls, hence neighborhoods in these areas would be a great location to open a shopping mall as there is less competition.

7. Conclusion

Cluster 0 has no shopping malls. This represents a great opportunity and high potential areas to open new shopping malls, as there is very little to no competition from existing malls.
But for setting up a shopping mall we need to consider other factors such as the cost of rent, the surroundings around the shopping mall, the kind of people in the locality-if it's a luxurious area many people prefer going out, their lifestyle will be different from others and therefore spend a lot. If we decide a place where the competition is less, then we need to consider the people living in that locality as well. If the people in that area spend a lot and love going out then it'll be a success. If the people staying near the mall don't prefer going out, then it's better to consider some other place with less competition and a good crowd.