

Recommendations on starting a new business in the City of London

Table of contents

- Introduction
- Data
- Methodology
- Analysis
- Results
- Discussion
- Conclusion

Introduction: Business Problem

London, the capital and financial hub of England has always been one of the most popular expat destinations in the world. Some of the factors that make London a great place to live in include its living standards, rich culture, diversity, benefits, education, career prospects and healthcare, to name a few. There is plenty to see and do for people of all ages and backgrounds but there is a lot more to this unique city than its attractions. Entrepreneurs and budding business people from around the world head to the Big Smoke when in the process of establishing themselves. London has now also become one of the leading hubs for start-ups in Europe.

In this notebook, we will be exploring the different neighbourhoods of London using Foursquare location data to get reviews, ratings, tips and counts, which would give us an insight into the different businesses within London such as bars, restaurants, grocery stores, hotels etc. Once we have enough data available on different businesses, we can then decide what kind of business to establish and where we would make the most returns on investment. The recommendations will be based on the results returned by Foursquare API on major businesses.

Data

The recommendations will be based on the data returned by four square such as business name, category, review and location. This will help us understand what kind of business will have a great impact and in which location within London.

Following data sources will be needed to extract/generate the required information:

- **Google Maps API using Geocoding** to get the coordinates/latitudes and longitudes of London.
- **Four Square API** to get the name, address, category, street, latitude and longitude of different businesses.
- We will use pandas json_normalize to transform our json data into a data frame.

- **Folium Maps** to generate maps for the businesses in London and which areas there are located.

Methodology

In this project we will be focusing on manipulating data in the categories of hotels, bars, restaurants and shops within London. We will be analysing businesses that are in 100000 metres radius from London. This will help us get all the places that have been rated by Foursquare users within and around London.

We will first make an api request to Four square and return our data. The dataset will include hotels, bars, restaurants etc. that have been shared and rated. Our dataset attributes will consist of the place name, address, category, street, latitude and longitude. Our dataset will be in json format when first returned from Foursquare.

Secondly, we will use pandas json_normalize to transform our json data into a data frame. We will pass the appropriate columns when processing.

Finally, we will replace the appropriate address names to businesses returning NaN. Once our data has been processed properly, we will use Folium to map the businesses using their latitude and longitude values. This will give us a visibility of where these businesses are located within London. We will also be exploring the highest rated business and the kinds of businesses that are situated around it.

Our recommendation will be based on what category of businesses people prefer to visit when they come to London. Breaking it down by category and giving the count will help potential business owners on what category of businesses to make their investment.

Analysis

To perform our analysis, we will first convert London to its latitudes and longitudes. We will use this to pass in our lat and lng to the API request.

The Api request using Foursquare will help us to extract the categories of the businesses. This will help us understand which business category has the highest investment and attraction in London.

In our resulting data frame, some of the columns return NAN such as cross street column. We will replace them with Unknown to keep it simple. The important thing is that we don't have any NaN values in the address column which is a more relevant data for us.

We will then use folium library to map all the businesses we have retrieved from Foursquare.

Based on the data returned from Foursquare, we see that St James's Park has the highest ratings. This is so because Foursquare presents data based on the highest to the lowest ratings. We will manipulate other businesses around St James's park by passing its latitude and longitude to the API request. This will let the potential business owners have an idea of the different types of businesses that are situated around St James's Park.

The resulting dataset depicts that over 80% of businesses, hotels, art galleries, museums and stores are situated in London. Comparing the first dataset where we

used London coordinates with the second one which we used St James's Park with the highest ratings, we can see similar establishments. This suggests that the same group of customers are moving between these businesses. Establishing a business with similar categories in mind, we have an 80% assurance that these same customers will come around.

Results

Grouping the dataset by categories, we see that the category Park (theme Park and attractions within the park) has the highest amount of entry according to our Foursquare data, followed by hotels. We can advise our new investor to make their investment in the order of Theme Parks, Hotels, Theatre and grocery store. These will have a good return on investments in London based on our database.

We will then further explore neighbourhoods around St James' Park. In this sense, the neighbourhood refers to businesses and not the physical location.

We will check how many venues were returned for each neighbourhood and also the unique categories. We will group rows by neighbourhood and by taking the mean of the frequency of occurrence of each category. We will then print each neighbourhood along with the top 5 most common venues and put the results in a pandas data frame.

Conclusion

In this notebook, we identified popular businesses and business categories in London in order to aid potential business owners on what type of businesses to venture with. We acquired our data from Foursquare and processed it, which included the business names, categories, address, street, latitude and longitude. We further mapped the businesses using Folium.

After exploring the top venues (categories), we see that businesses in the categories of coffee shops, Pubs ,Café, Hotel restaurants are the top most attractive businesses situated in London. We can conclude that venturing in any of these categories will yield returns in investment.