Coursera Capstone

IBM Applied Data Science Capstone

The Battle of Neighborhoods

Opening a New Hotel in Paris, France

By Racha Khairallah
September 2019

Contents

1- Business Problem

For many tourists, visiting Paris is a great way to relax and enjoy themselves during weekends and holidays. They can do shopping, dine at restaurants, shop at the various fashion outlets, watch movies , visit many historical sites and organize many others activities.

Paris is one of the busiest cities in Europe, appear as the most visited city around the world with 17.95 million tourists visitors in 2018.

Building hotels allow property developers to earn consistent rental income. Of course, as with any business decision, opening a new hotel requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the hotel is one of the most important decisions that will determine whether the hotel will be a success or a failure.

By using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the following business question:

In the city of Paris, France, if a property developer is looking to open a new hotel, where would you recommend that they open it?

This project is particularly useful to property developers and investors looking to open or invest in new hotel in the capital city of France, Paris.

3- Methodology

To solve the problem, we will need the following data:

• List of neighborhoods in Paris.
• Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
• Venue data, particularly data related to hotels. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them

1- We will use web scraping techniques to extract the data from the Wikipedia page, (https://en.wikipedia.org/wiki/Category:Districts_of_Paris) with the help of Python requests and beautifulsoup packages.

2- We will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

3- We will use Foursquare API to get the venue data for those neighborhoods**.**

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). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

4- Results

4- Discussion

5- Conclusion