Find the best neighborhood in Paris to setup offices for a startup

1. Introduction

1.1 Background

As a founder of a new startup, a friend of mine is looking to setup a new office in Paris. The company is growing up very fast and the need of hiring new people is critical. He would like to hire talented people who have been newly graduated. He's looking as well to offer internship for students. Therefore, it's important for the company to be attractive to students and young peoples. One way to attract talents is to have nice offices located in a neighborhood with a good quality of life.

1.2 Problem

The problem is to find the best neighborhood in Paris to setup the offices. The neighborhood must be attractive for young people in order to give them some interest on the company. This project aims to find the best places by selecting neighborhood on criteria important for students and newly graduated people. I focused on the following criteria as I supposed they were important for the targeted audience. A neighborhood is supposed to be attractive to young people when there is a lot of:

- Restaurants and Food venues. Students like to go outside to grab their meals
- Nightlife Spots and Bars to chill out with friends and meet new people
- Outdoors and Recreational venues to relax and playing sports

2. Data acquisition and cleaning

2.1 Data sources

In order to have the best dataset for running analysis, data is collected from different sources. The first source to get Paris neighborhood information is from the online platform http://opendata.paris.fr. This service provides public and general information on the city of Paris.

To expand the first data set I used Wikipedia to get population of the different neighborhoods. https://fr.wikipedia.org/wiki/Liste_des_quartiers_administratifs_de_Paris

Finally, I used Foursquare to get information on the venues.

2.2 Data cleaning and feature selection

Globally the data quality from the different sources was good and there was no need to spend a lot effort on cleaning the data. There was not too much missing data. When it was the case I handled missing data by filling empty values with the mean of the columns.