Final Project - Battle of Neighborhoods - Paris Restaurants

1. INTRODUCTION

Paris is undoubtedly one of the most important cities in the western world when it comes to tourism, arts, fashion and culinary. As French cuisine expanded all across the globe, the contrary is also true: foreign culinary has set hundreds of restaurants all around Paris.

Since the city is known for many reasons, one of them its culinary, the idea of the project in a nutshell is to explore restaurants' data, like ratings and likes on Foursquare and, since the project is called Battle of Neighborhoods, this project is also going to compare neighborhoods in Paris, examining if french restaurants are the most common ones and, if not, what kind of restaurant is the most common in each of the 20 *arrondissements* of the city.

The idea is to start with Foursquare API in order to obtain information about venues in the city, given the geographic location of each neighborhood center. In order to filter the different types of venues found in the city, some criteria will be used for checking potential data to be treated as french restaurants or similar ones, and to classify bars and foreign cuisines as other types of restaurants.

Later on, the project will provide an overview of the best and most popular restaurants and bars and then focus on investigating the distribution of restaurant types per neighborhood. Tables and maps are going to be used as tools, in order to discover whether each neighborhood has more French Restaurants than bars and cafés, or if international cuisine is more common in a specific area.

The intention, by creating this project, is to provide information for tourists, enthusiasts of the french cuisine and even locals about restaurants available in Paris. Even though there are limitations in using a sample to solve the problems at hand, having an idea of which type of restaurant is the most common per area is an interesting outcome for me personally and treating the data, using foursquare API and adapting the problem to the city of Paris will be quite a challenge.

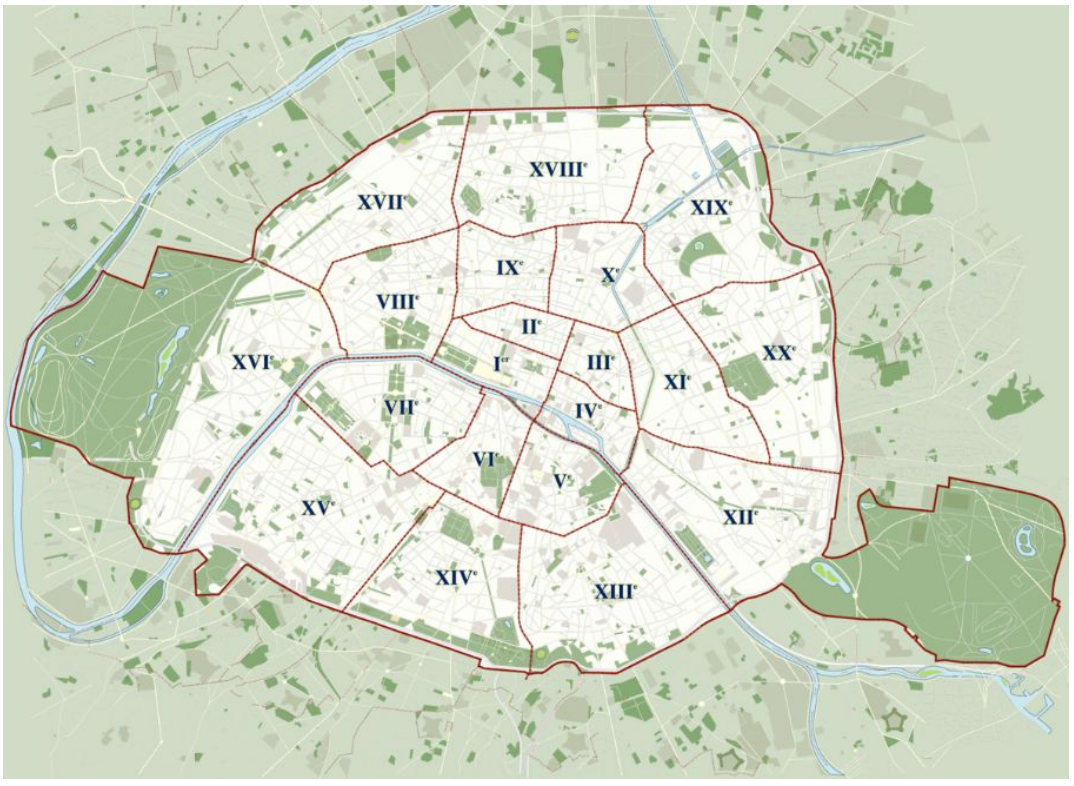
Finally, the idea of the project is mainly to be able to find out good restaurants in Paris and to divide the city by the most common kind of restaurant in each of the 20 neighborhoods.

2. DATA ACQUISITION AND DATA CLEANING

2.1. DATA SOURCES

Since this work aims to identify the distribution of restaurants among specific locations, the Foursquare API is going to be the tool for finding our sample: venues near specific geographic locations.

Working with the map of Paris can be quite challenging, since the city has a 'snail-shell' shape, in which the most central neighborhoods have a much smaller surface than those on the outside. Looking for a solution for finding a great sample of venues per neighborhood without compromising the fact that the area varies a lot between the most central and most peripheral neighborhoods is the challenging task of this part.



Neighborhoods of Paris. Source: Wikipedia

The number of searches per day one can do in foursquare is limited. For that reason, the project may have functions to save dataframes as csv files for later use, avoiding the daily use of foursquare API.

Two functions are to be defined, the first to retrieve the list of venues given the locations and the second one for retrieve Foursquare information about the restaurants to go though further analysis.

The csv file including neighborhoods of Paris and their geographic locations can be found at the French Government website (https://www.data.gouv.fr/). A path to my github account including the table will be used in the project, since the data is required. The data downloaded is yet to be treated for better use.

2.1. DATA CLEANING

The table provided for geographic locations of the neighborhoods comes with several columns not to be used in this project. The names of the remaining columns are also to be changed for better comprehension and use during the project.

One last point to be treated in the data is that the neighborhoods are presented in a random order. So the data will be sorted in ascending order.