The Battle Of Neighborhoods

I. Introduction

1.1. Background

This is the capstone project of IBM Data Science Professional Certificate. The exercise imposes using the Foursquare API in order to cluster a place in the world according to famous venues in this place. I will use the Foursquare location data to explore neighborhoods in Paris, and come up with a problem for which I can use the Foursquare API.

1.2. Problem

I will explore the neighborhoods in Hong Kong and answer the question: "Where is the appropriate place to open an Indian Restaurant in Hong Kong".

1.3. Interest

Paris is one of the biggest international cities in the world, and one of the most touristic cities. Opening a restaurant here is an attractive idea. Nevertheless, Indian restaurant isn't really what tourists are attracted by in the first place when going out in Paris. Are some quarters more suitable for setting an Indian restaurant than others? This is the question we are going to deal with, by exploring different clustering of Paris based on the most famous venues recorded on Foursquare Website.

II. Data

A. Data Sources

Two different kind of data is needed for the comparison.

- **City quarters and respective geographical data:** in order to analyze the cities on a meaningful level, they need to be divided into different areas, in our case in *quartiers (*subdivisions of the 20 *Arrondissements* constituting of Paris. I was able to find the list of the 80 *quartiers* of Paris on Wikipedia[1]. I then web-scrapped the HTML page in order to convert this list into a data frame usable with pandas.
  Using the *Geocoder* python library, I was able to get the geographical coordinates for each quarter.

- **Venue data:** This data, including the Venue name, its category, latitude and longitude, is gathered using the Foursquare API[2].

[1] https://en.wikipedia.org/wiki/Quarters_of_Paris

[2] Foursquare City Guide, commonly known as Foursquare, is a local search-and-discovery mobile app developed by Foursquare Labs Inc. The app provides personalized recommendations of places to go near a user's current location based on users' previous browsing history and check-in history.