**Similarities Between Paris And New York**

Made By: Mohammad Ahsan

**1.Problem**

Imagine we have a friend who lives in Paris and who plans to travel over to US for a business trip and stop-over in New York for a long weekend and enjoy the city.

When in Paris, he likes to hang out in Saint Germain, a district where you can find plethora of cafes, brasseries, bars, and restaurants. You also find art galleries, antique stores, theatres and cinemas, street markets, parks, clothing retailers and other shops. Saint Germain in quite central of Paris and located in the 6th arrondissement.

Now, he would like to find the same in New York but doesn't know much about that city. Where is the Saint Germain of New York? What neighbourhood of New York is similar to Saint Germain?

To answer questions like the one asked by our friend, we analyse here data from Foursquare, a popular location-based social network. Foursquare enables users to share their current location with friends, rate and comment on venues they visit (places such as restaurants, hotels, cafeterias, bookshops, and museums) and read reviews of venues that other users have left.

For the purposes of this project, we use geographical datasets of the two cities New York, more specifically Manhattan, and Paris. We take the neighbourhoods to be areas on a city's geographic map. We then inquire the venues recommendations from Foursquare through their API. Each area is associated with the set of venues within predefined boundaries, based on our manually collected ground-truth data. In turn, each venue is associated with a name, a category (e.g. restaurant, coffee shop, bar, museum, art gallery, etc.), and a geographic location (latitude and longitude). We got about 1,800 venues listed for Paris and 3,300 venues for Manhattan.

## **2. Data**

#### Paris geographical data

Paris is divided into 20 administrative districts, referred to as arrondissements. The 20 arrondissements are arranged in the form of a clockwise spiral (often likened to a snail shell), starting from the centre of the city.

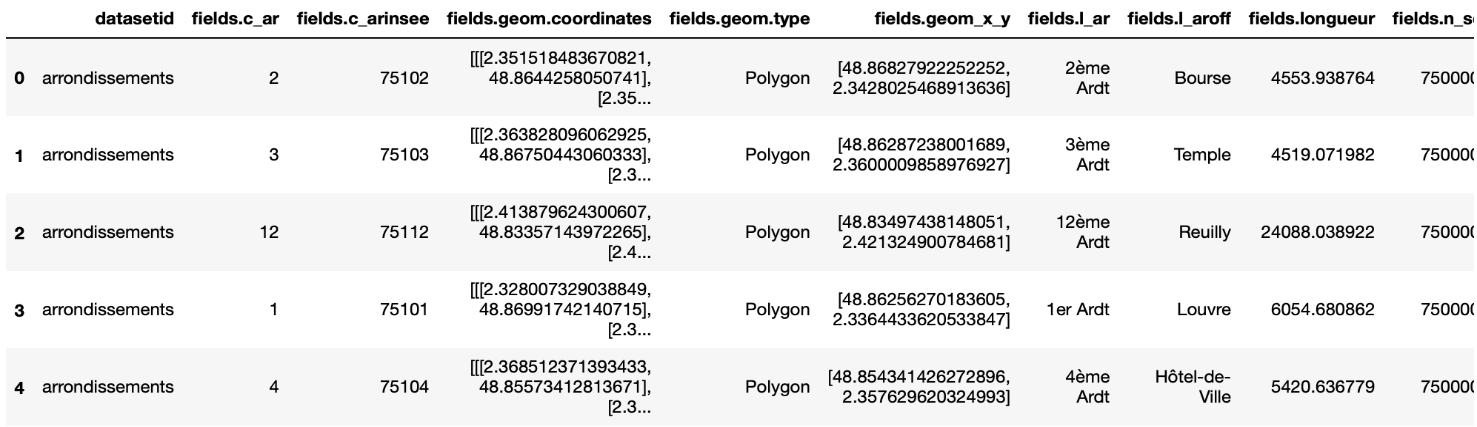
Among landmarks of Paris, we can find the **Louvre Museum** in the 1st arrondissement, the **Museum of Modern Art (Georges Pompidou)** in the 3rd arrondissement, **Notre Dame Cathedral**in the 4th arrondissement, the **Eiffel Tower** in the 7th arrondissement, the **Champs-Elysées Avenue** in the 8th arrondissement, the **Arc of Triumph** at the limits of the 8th, 16th, and 17th arrondissements, and so on.

One may find open datasets about Paris at <https://opendata.paris.fr/page/home/>. Here below are the links to the arrondissements geographical data.

csv file: <https://opendata.paris.fr/explore/dataset/arrondissements/download?format=csv&timezone=Europe/Berlin&use_labels_for_header=true>

json file: <https://opendata.paris.fr/explore/dataset/arrondissements/download?format=json&timezone=Europe/Berlin&use_labels_for_header=true>

Here is a sample of the raw Paris dataset, once it is converted into a data frame.



We use the location, i.e. **latitude**and **longitude** information, as well as the **arrondissement number** and **name,** associated to each arrondissement from the table above.

#### New York geographical data

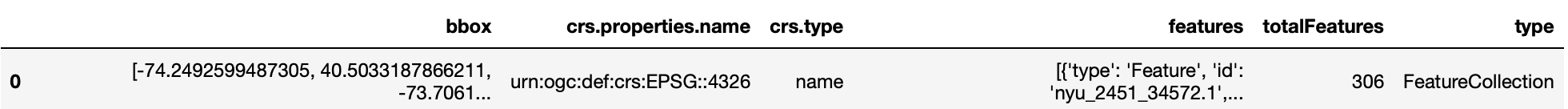
Geographically, New York is a city with 5 boroughs, which are the Bronx, Brooklyn, Manhattan, Queens, and Staten Island, and hundreds of neighbourhoods.

When we think of New York City, Manhattan is often the first place we picture. The borough is home to well-known attractions, such as **Central Park**, the **Empire State Building**, world-class museums such as **MOMA**, the bright lights of **Times Square**and **Broadway**. Manhattan contains big-name neighbourhood’s, international restaurants, classy boutiques, trendy bars and more.

There’s no shortage of information about New York City. For the sake of simplicity, we use the one provided by IBM in the segmenting and clustering lab of this course.

json file: <https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json>

Here is a sample of the raw New York dataset, once it is converted into a data frame.



We use the features subset/key information, from where we can extract the borough name, neighbourhood name, and latitude and longitude, associated to each borough.

Venues recommendations API by Foursquare

Foursquare features a developer API that lets developers’ applications make use of Four square’s location data. Their API powers various geo-enabled searches of venues with sophisticated details (e.g. tips, hours, menus, stats over time), searches of users, check-ins, etc. In the scope of the resources allocated to this project, we just use the API feature of exploring top recommended venues nearby a location that returns basic venue data (name, location, etc.), category, and ID.

Combining the geographical cities data with Foursquare venues

The approach is to gather venues data from Foursquare based on our two cities neighbourhood’s. We then use the venues categories feature for clustering the neighbourhood’s and look for basic similarities between these neighbourhoods of Paris and Manhattan.

## **3. Methodology**

### 3.1 Datasets Exploration and Setting

In this section, we perform **data wrangling**, as below:

1. Geographical data of Paris arrondissements and New York boroughs/neighbourhoods.
   * Extracting neighbourhood geographical information for the cities of Paris and New York from json files (refer to the links provided in the Data section above).
   * Creating Pandas data frames that include:
     + names of the boroughs/arrondissements,
     + names of the neighborhoods,
     + locations of the neighborhoods in terms of latitude and longitude.
   * Visualizing the neighbourhood’s in on a Leaflet map via the Folium library.
2. Foursquare recommended venues data based on neighbourhood’s delocalisation.
   * Gathering venues recommendations through the explore endpoint API.
   * Creating Pandas data frames that include:
     + names of the boroughs/arrondissements,
     + names of the neighborhoods,
     + locations of the neighborhoods in terms of latitude and longitude,
     + the names of the venues,
     + locations of the venues in terms of latitude and longitude,
     + categories of the venues.