Tourism Recommender System

1. Introduction/Business Problem

The Idea of the project is to recommend cities to some user to visit, based on previous cities he visited before.

Recommender Systems are data science tools, which attempt to predict items out of large pool a user may highly likely be interested in, and to suggest him the best one. They also support systems helping users to find and/ or to make choices about items that matches their preferences and interests.

In the tourism field, recommender systems aim to match the characteristics of tourism and leisure resources or attractions with the user needs.

By Defining cities characteristics and using user preferences, the recommender system can suggest some new cities for this user, cities that are similar to his needs.

The audience of the project are tourists. That is, by helping them choosing cities they like to visit.

2. Data

There is to types of data sources will be used in this project.

The first one is Foursquare Location Data. The second one is Postal Codes with the names of neighbourhoods of the major cities in the world.

2.1. Foursquare Location Data

Foursquare is a technology company that built a massive dataset of location data. What is interesting about Foursquare is that they were very smart about building their dataset. They actually crowd-sourced their data and had people use their app to build their dataset and add venues and complete any missing information they had in their dataset. Currently its location data is the most comprehensive out there, and quite accurate that it powers location data for many popular services like Apple Maps, Uber, Snapchat, Twitter and many others, and is currently being used by over 100,000 developers, and this number is only growing.

2.2. Postal Codes Data

The Postal Codes Data for most of countries is available at AggData Website (https://www.aggdata.com).

I will download the data of three countries: United States, France and Italy. And then, I will get the major cities for each country.

The Postal Codes from AggData are able to be downloaded as csv files. The data contains also the location of the neighbourhoods which makes it very useful to use it along side with Forusquare API

	Postal Code	Place Name	State	County	City	Latitude	Longitude
0	24000	Périgueux	Aquitaine	Dordogne	Arrondissement de Périgueux	45.1833	0.7167
1	24001 CEDEX	Périgueux	Aquitaine	Dordogne	Arrondissement de Périgueux	45.1833	0.7167
2	24002 CEDEX	Périgueux	Aquitaine	Dordogne	Arrondissement de Périgueux	45.1833	0.7167
3	24003 CEDEX	Périgueux	Aquitaine	Dordogne	Arrondissement de Périgueux	45.1833	0.7167
4	24004 CEDEX	Périgueux	Aquitaine	Dordogne	Arrondissement de Périgueux	45.1833	0.7167

An Example of Postal Codes Data

3. Methodology

The main purpose of the project is to recommend a place to visit for some user based on his previous experience, especially, which cities and places he visited before.

First of all, we will clean and merge the postal code data that is discussed in Data section. Then, using Foursquare API, we will get the venues on each neighborhood. We will do some EDA and Preprocessing on the data to clean data from any noisy data that can affect negatively on the results. After that, an analysis of each neighborhood is required to determine which venue, categories this neighborhood serve. We will use Recommender Systems to recommend the new place to the user. The type of Recommender System used is Content-based Recommendation Systems.
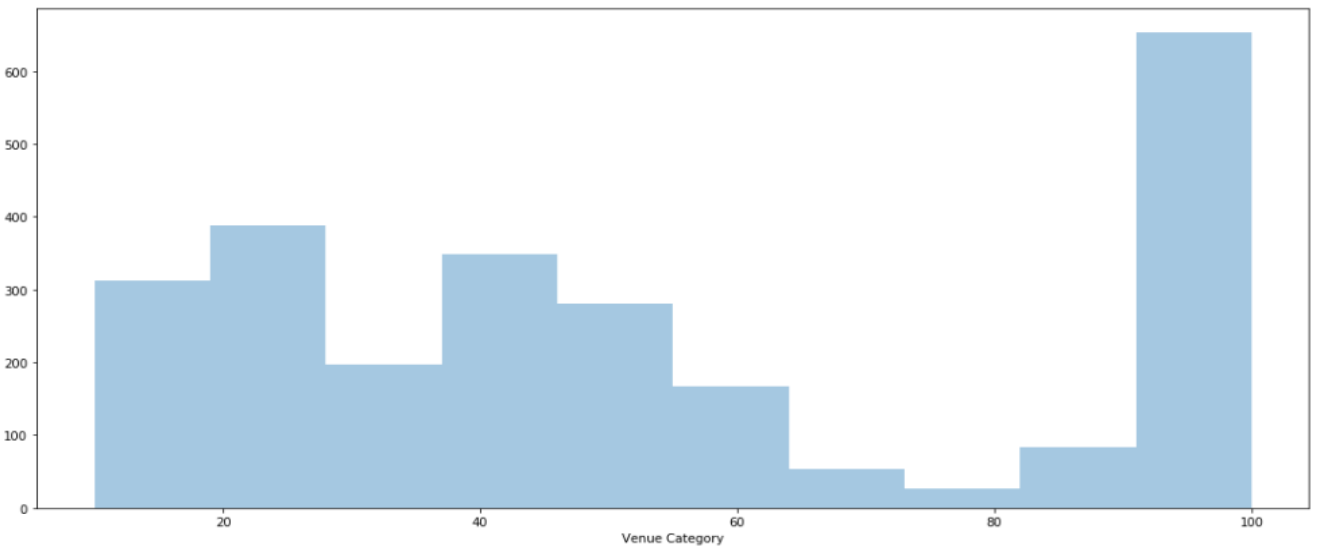
In order to use Content-based Recommendation Systems, we will create genre matrix, which provide info about the categories that are served in each neighborhood.

In Results section, we will create some random user input to test our Recommender System.

3.1. EDA
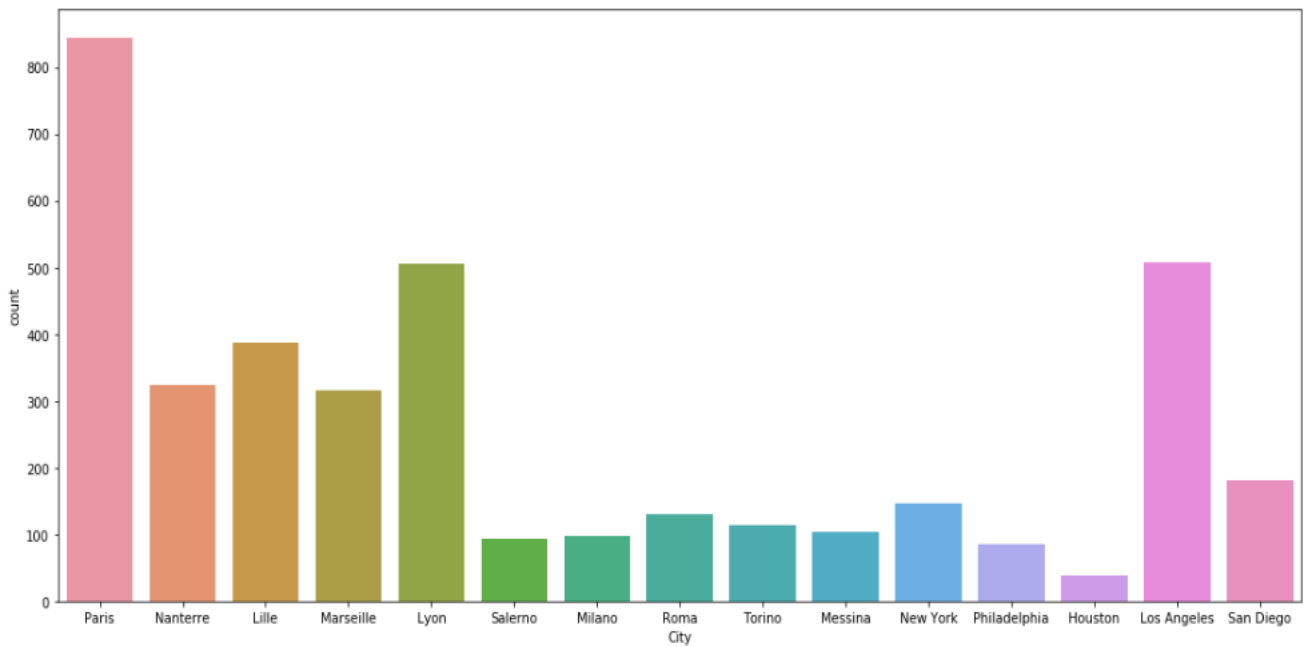
Number of Neighborhoods by City

Venues Categories Histogram

```
French Restaurant      10874
Hotel                   7124
Italian Restaurant      6543
Bar                     4517
Pizza Place             3517
Bakery                  3218
Plaza                   3148
Restaurant              3072
Café                    3037
Japanese Restaurant     2759
```

Top Frequent Venues Categories

3.2. Content-based Recommender Systems

Content-based filtering methods are based on a description of the item and a profile of the user's preferences. These methods are best suited to situations where there is known data on an item (name, location, description, etc.), but not on the user. Content-based recommenders treat recommendation as a user-specific classification problem and learn a classifier for the user's likes and dislikes based on product features.

In this system, keywords are used to describe the items and a user profile is built to indicate the type of item this user likes. In other words, these algorithms try to recommend items that are similar to those that a user liked in the past, or is examining in the present. It does not rely on a user sign-in mechanism to generate this often-temporary profile. In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. This approach has its roots in information retrieval and information filtering research.

To create a user profile, the system mostly focuses on two types of information:

1. A model of the user's preference.

2. A history of the user's interaction with the recommender system.

3.3. Genre Matrix

The genre matrix represents the core kernel of our recommender system.

For any user input we can find the user profile than we can find the recommended item for this user profile.

That means that, we can extract the trend of the user and suggest results that fits his needs and his preferences.

Neighborhood	ATM	Accessories Store	Adult Boutique	Advertising Agency	Afghan Restaurant	African Restaurant	Airport	Airport Terminal	Alsatian Restaurant	American Restaurant	...	Weight Loss Center	Whisky Bar	Wine Bar	Win Sho
Affori - 20161	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.000000	0
Alhambra - 91801	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.033333	0
Alhambra - 91802	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.013889	0
Alhambra - 91803	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.000000	0
Alhambra - 91804	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.013889	0

Genre Matrix

3.4. User Input

The user input is the places (Neighborhoods) that he visited before and for each place how many times he has visited this place.

Neighborhood	visits
Paris 16 Passy - 75768 CEDEX 16	6
Burbank - 91526	1
Milano - 20136	2
Paris 12 Reuilly - 75574 CEDEX 12	9
Cernusco Sul Naviglio - 20063	5

Example of user input

We will get user genre by inner-join the user input with the genre matrix. Then we will get the recommended place by manipulating the user genre with the visits count vectors.

4. Results

The resulted place to visit based on previous data is "Paris 16 Passy - 75216 CEDEX 16"
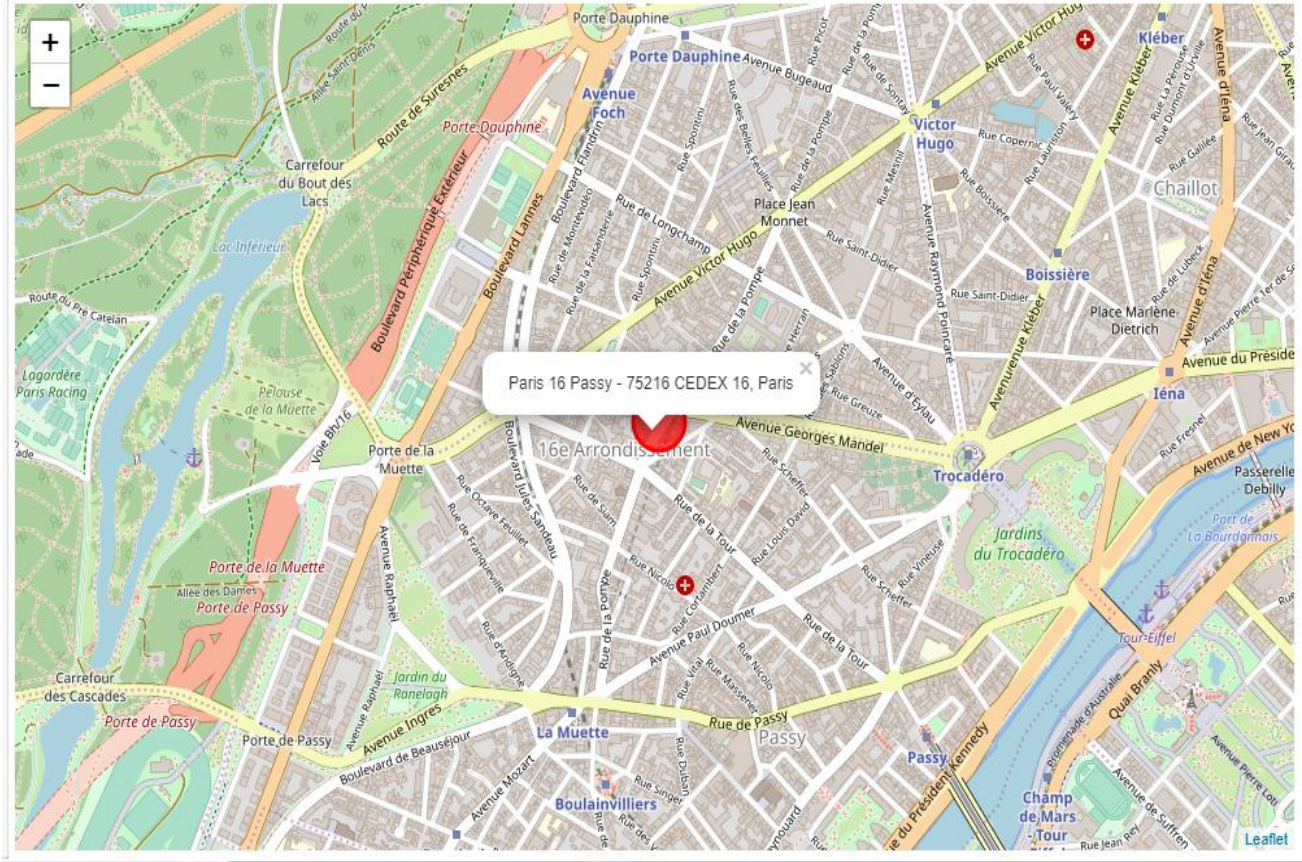
5. Discussion

We see that our recommender system did the target of the project. However, we can we can improve the results by using hybrid recommender systems to combine the advantages of different types of Recommender Systems.

6. Conclusion

In this project, we create a content-based recommender system. This recommender system works based on user experiences and a genre matrix that is extracted by using AggData Postal Codes of 15 City from 3 countries, Foursquare API which provide us with the venues of each neighborhoods of those cities.

We intend to extend the project to cover whole the world by getting more and more data. Also, we can improve the project by using hybrid recommender systems to combine the advantages of different types of Recommender Systems.