Tourism promotion in NYC

Report

Problem and background

New York is one of the most visited cities in the world every year. Over 60 million tourists visit the city, and they spend at least USD 670 per tourist. However, in some of the webpages in which these people give their opinions, there are several critiques, such as the lack of diverse sightseeings and information of new and different spots that can be visited.

For this need, the use of data is central. By using machine learning algorithms, we could create a platform in which there could be suggestions for not-that-well-known spots that can be visited in the city. In this way, two main needs could be fulfilled: first, that of tourists willing to get to know the "real" city; second, the economic growth of places, beyond Manhattan.

This is what we are going to prove in this project.

Data

We have worked with geographic information from NYC, which can be found in the repository of NYU. In that file (converted into a JSON file), we can find not just each one of the neighborhoods of the city, but also its coordinates and the borough they belong to.

With that information, we have created a dataframe that looks like this:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

We found that we are dealing with a dataframe with 6 boroughs and 306 neighbourhoods.

After graphically seeing the location of each neighbourhood, with the aid of the Folium library, we start looking for popular spots in each one of the neighbourhoods, with the tools provided by Foursquare. With that information, we create another dataframe. The dataframe, for the first neighborhood looks like this:

	name	categories	lat	lng
0	The Bar Room at Temple Court	Hotel Bar	40.711448	-74.006802

1	The Beekman - A Thompson Hotel	Hotel	40.711173	-74.006702
2	Alba Dry Cleaner & Tailor	Laundry Service	40.711434	-74.006272
3	City Hall Park	Park	40.712415	-74.006724
4	Gibney Dance Center Downtown	Dance Studio	40.713923	-74.005661

In the case of the first neighborhood, we have found 100 popular venues.

Now, we repeat the process for all the neighborhoods. With this, we will apply a Machine Learning algorithm in order to obtain the main spots and create a tourist guide that can attend the changing needs and desires of the tourists visiting NYC.