Battle of the neighbourhoods

Introduction

This analysis is the final project in the Course "*Applied Data Science*" by IBM. All recently learned new tools that are useful to solve the business problem will be applied here, like data acquiring, cleaning, sorting and visualization on a map.

Business problem

Pub crawls exist in a lot of major cities by now. They are not the best way to get to know locals because mostly they are done by tourists, but they are an excellent way of getting to know the nightlife. If one wants to get to know the city by day, all tourist attraction are available, but the nightlife in a city makes up a significant part of its character, so not getting to know local venues and bars would be missing out a major part of a city's character. Pub crawls are mostly done by foot, so an important part is, that all venues are in walk-able distances. Everyone new to a city needs a map, so why not integrate the map with the best venues for a pub crawl?

The question to solve is: which city is best suited for a pub crawl, where are good venues in a close radius around city center? For this example the cities of Manhattan (New York), Toronto, Munich and Berlin are chosen for analysis, but the code shall be written universally.

Target audience

There are multiple possible target audiences: first there are Hostels chains that can see in which city it is worth to invest. Second, activity providers like tripadvisor can leverage this information to advertise a professional pub crawl. Lastly, tourists may pay a small amount of money for a convenient map of a city with all important venues for nightlife. Most cities give out free maps with touristy attractions, but what's missing there are the good pubs to visit in the evening.

Data sources

The code is planned to be modular, to be able to be applied to any city, and will apply multiple data sources. Given the city name, the city center will be found with pythons Geolocator. A list of all venues will be gathered by requests to the Foursquare API. When all data is processed, the results will be shown in a map with folium.

Data processing

The dataframe of venues of a city will be acquired with a request to the Foursquare API. The radius of the request will be 1km around city center. Then, the data will be cleaned up, the location and rating included (with another request to the API) and sorted. The header of the final dataframe will consist of location, name and rating of all given venues. With the cleaned up data, a score for this city will be calculated. This will simply be the sum of all ratings of the seven best venues. When the city with the seven best venues is found, a map for the inner city with all venues will be created, which can be printed for use.