Report

Coursera Capstone Project

IBM Data Science

Exploration of Popular Nightlife Areas

Assignment by Jennifer B.

Date: 13th of January, 2021

Outline

The assignment's task of the Capstone Project is to come up with an idea to leverage Foursquare location data to explore or compare neighborhoods or cities of choice or to come up with a problem that can be solved by using the Foursquare location data. The report consists of 6 parts:

1. Introduction
2. Data
3. Methodology
4. Results
5. Discussion
6. Conclusion

In the introduction, the business problem I came up with is described. Afterwards, the data and methodology used to solve the problem are explained. In the results' section, the outcome is presented. Finally, the results are discussed and the report is concluded in the conclusion's section.

1. Introduction

When traveling or moving to a foreign city, people are often interested in exploring the cities nightlife. Though, often it is hard to decide which places to visit as usually there is a high amount of bars, clubs and pubs. This notebook makes it easy to find areas with a high density of nightlife venues in a city someone is interested in. Perfect for pub crawls or nightclub hopping! Thus, a tourist can explore as many different bars, clubs and pubs as possible within a short time and without walking to far from one venue to another. Tourist, travel guides and people who just moved to a new town (e.g. students) might profit from this project. If the outcome of this project is successful, code in the notebook could be used to extend popular tourist apps or websites (e.g. *tripadvisor)* by this function.

2. Data

The data used for this purpose are geolocation data from the online service Foursquare City-Guide (https://de.foursquare.com/city-guide). The following venue categories are searched within a radius of 10km of the city/address of choice by using the category ID:

Category	**Category ID**
Nightclub	4bf58dd8d48988d11f941735
Irish Pub	52e81612bcbc57f1066b7a06
Hotel Bar	4bf58dd8d48988d1d5941735
Lounge	4bf58dd8d48988d121941735
Cocktail Bar	4bf58dd8d48988d11e941735
Bar	4bf58dd8d48988d116941735
Pub	4bf58dd8d48988d11b941735

With a developer account, the data is obtained as requests with the Foursquare API and stored in a pandas dataframe for further processing. Among others, the dataframe contains the name of the venue, its category, category id, address and its geographical coordinates.

3. Methodology

The business problem is solved by creating a jupyter notebook by using the python programming language. The code in the notebook consists of four parts. In a first step, the location is defined. The user is required to type in the name of a city or any address of choice, which the user wants to explore for nightlife venues. With the geocoder package, the geographical coordinates to this city or address are found. In the second step, the nightlife venues within a 10km radius around the location are searched with the Foursquare API. To find only nightlife venues, a list of category IDs is defined to only search for the categories listed in the above table in the data section. An algorithm loops through the categories and sends a request to Foursquare. The results are appended to a dataframe called df_night. In the third step, the data in df_night is normalized with a standardscaler and clustered with the density based clustering algorithm DBSCAN from sklearn.cluster. The clustering is based

on the geographical coordinates latitude and longitude. This results in each venue being assigned to a cluster. An example of results for Miami is shown in the table below. All venues with a cluster ID higher or equal 0 belong to a cluster. Venues with the same cluster ID are close to each other and thus, belong to the same cluster. Venues with ClusterID = -1 do not belong to a cluster, as they are to far away.

venue name	categories	lat	lng	ClusterID
Mr. Jones Lounge	Lounge	25.79053	-80.1312	0
DayGlow	Nightclub	25.79502	-80.1348	0
Voodoo Rooftop	Lounge	25.77975	-80.1311	0
INK	Nightclub	25.77744	-80.1332	0
La Otra	Nightclub	25.76655	-80.1999	1
Wynnwood Factory	Nightclub	25.80023	-80.1937	2
Congas Nightclub & Restaurant	Nightclub	25.77713	-80.1848	3
La v Brickell	Nightclub	25.76242	-80.1932	1
The Hangar	Nightclub	25.78458	-80.1928	4
Basecamp	Bar	25.83226	-80.19	-1
La Divina Gastro Club	Nightclub	25.76263	-80.1931	1
SoHo Lounge	Nightclub	25.81074	-80.1919	5
Studio 23	Nightclub	25.79876	-80.1288	0
Souz Ultra VIP Lounge	Nightclub	25.77774	-80.1331	0
The 5Th Base	Beer Bar	25.7782	-80.221	-1
Brickell Blush Nightclub	Nightclub	25.76823	-80.1947	1
Stiletto Gentlemen´s Club Miami	Nightclub	25.80292	-80.2551	-1
Space 34	Nightclub	25.78457	-80.1914	4
La Cueva Bar	Nightclub	25.76275	-80.195	1
Studio 60	Nightclub	25.80901	-80.2343	-1
12345 West Dixie	Nightclub	25.88867	-80.1882	-1
Casablanca Cigar Lounge	Nightclub	25.76777	-80.135	0
90° DEGREES	Nightclub	25.78447	-80.1925	4
Mynt Lounge	Nightclub	25.79532	-80.1291	0
Cameo Nightclub	Nightclub	25.78699	-80.1315	0
Centro Wynwood	Nightclub	25.79963	-80.2011	2
private key club	Nightclub	25.80187	-80.2045	2
Ocean's Ten	Nightclub	25.78018	-80.1308	0
Twist	Gay Bar	25.78171	-80.1324	0
Treehouse Miami	Nightclub	25.79904	-80.1295	0

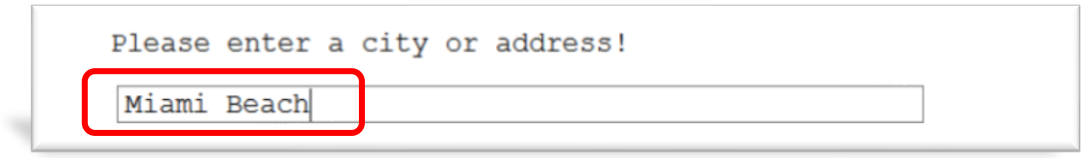
In the last step, the results as shown in the table are visualized on a folium map, that clearly shows the clusters with a high density of nightlife venues in different colors. Each venue is marked by a circle and shows a label with the name of the venue and its category, when clicking on it.

4. Results

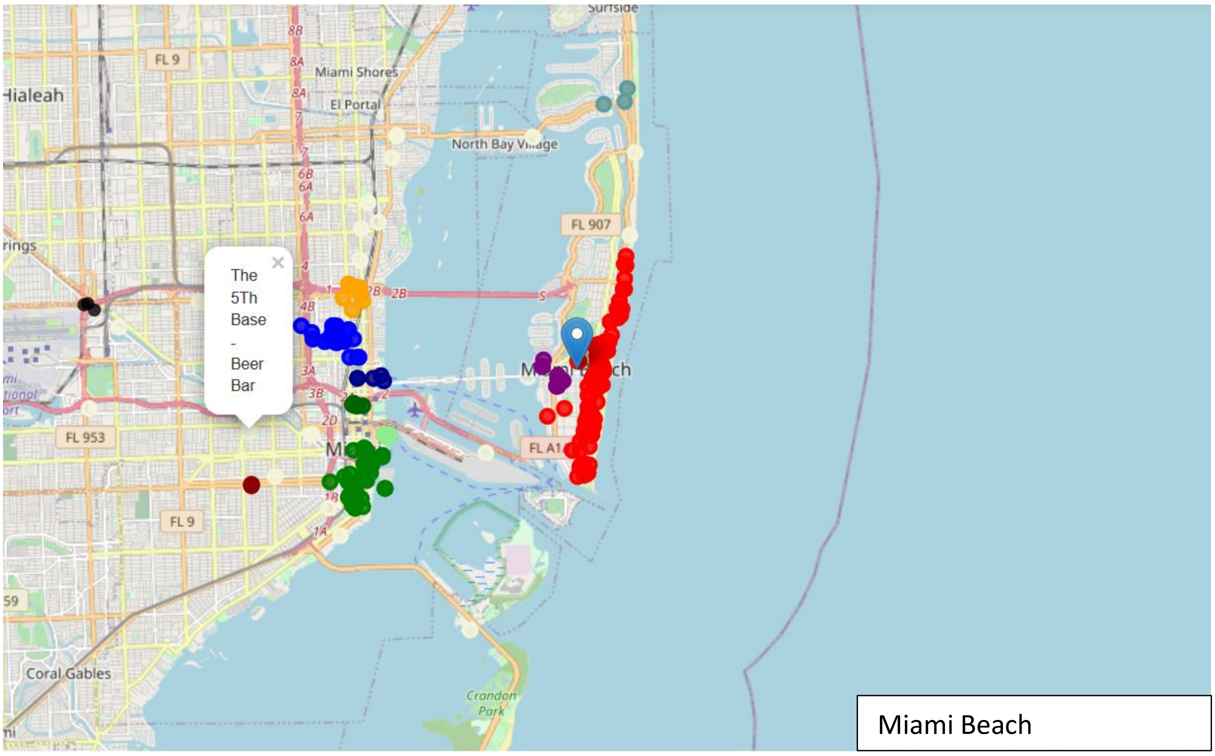
The notebook was tested for the following locations:

Location of interest	latitude	longitude
Miami Beach	25.7929198	-80.1353006
Östermalm, Stockholm	59.336723	18.0859178
Riga, Latvia	56.9493977	24.1051846
Paris	48.8566969	2.3514616

The user of the notebook can type the locations into an input line in the notebook as shown below:

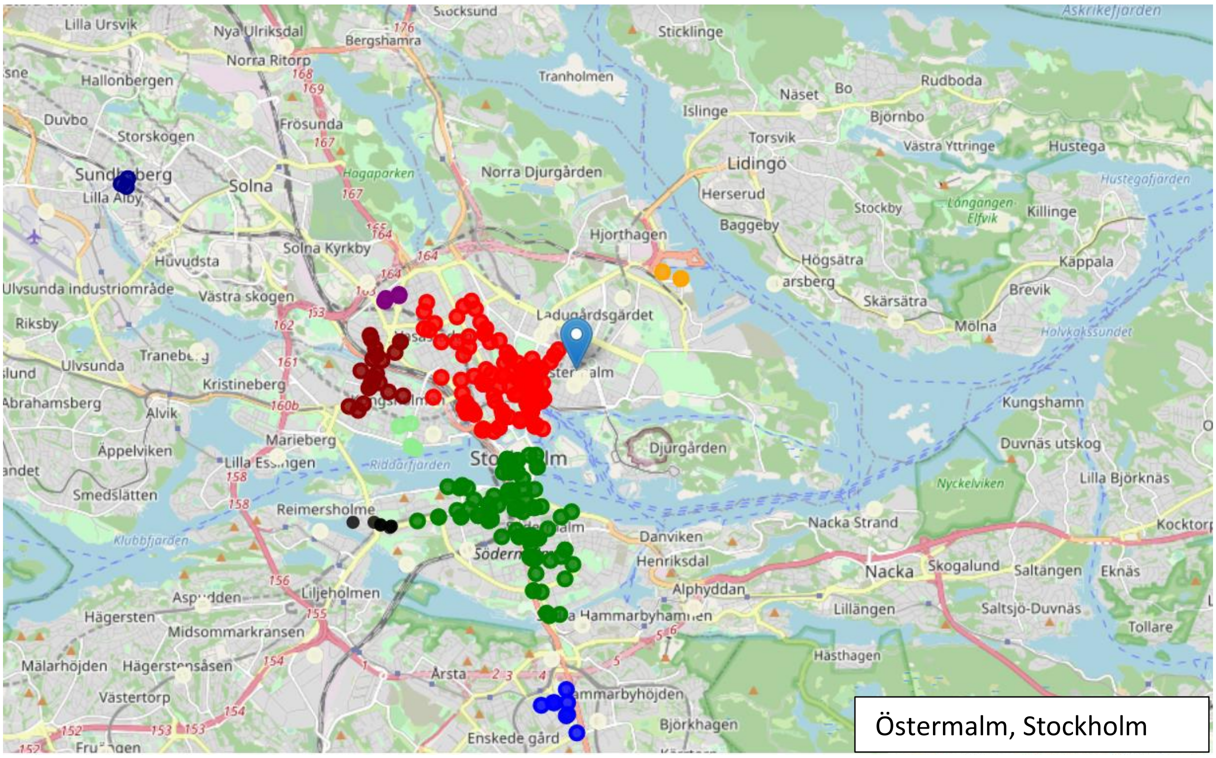


The image on the next page shows the resulting clusters for Miami Beach on the map. The light blue marker shows the geographical location. The eastern coastline of Miami Beach has a very high density of clubs, bars and pubs (red cluster). There are also areas with many nightlife venues near the center of Miami (e.g. green, blue, darkblue and orange clusters). In the image, you can also see the label of the venue 'The 5th Base' which does not belong to a cluster, as there are not many other nightlife venues nearby. These venues, that do not belong to a cluster (ClusterID = -1), are marked beige and are less visible.

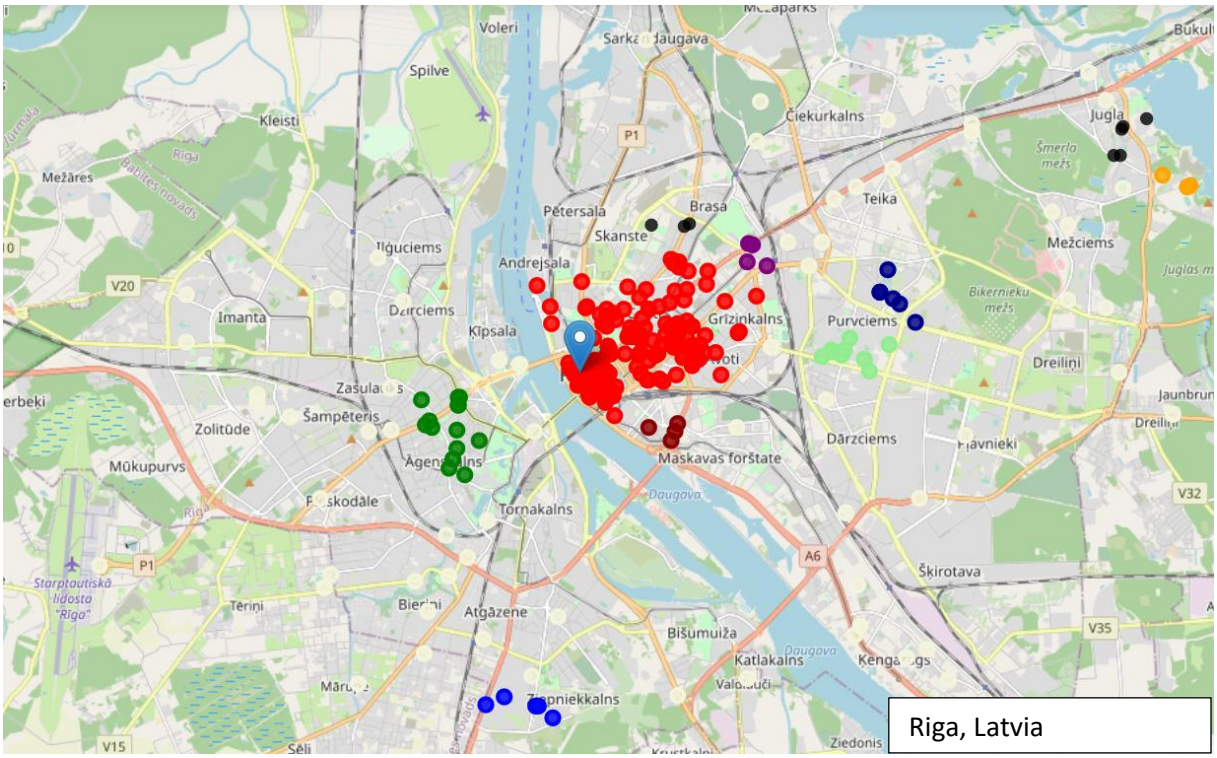


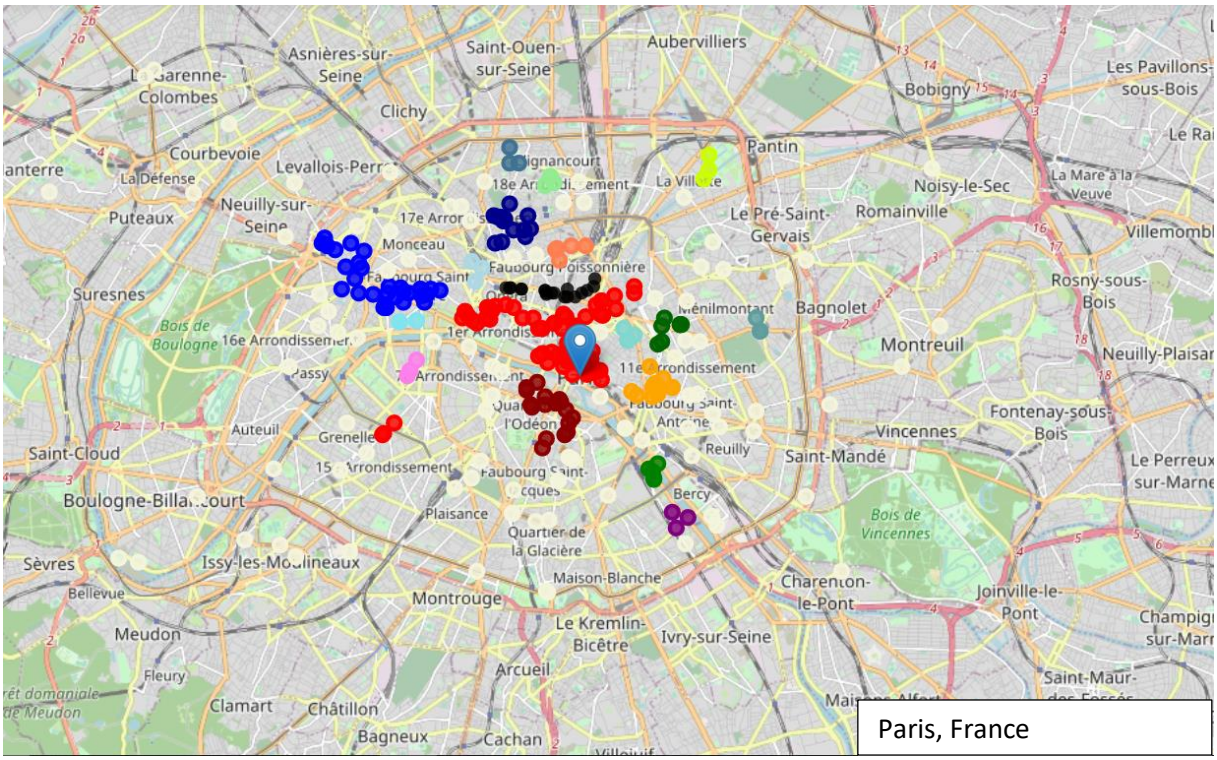
Miami Beach

Similar maps are created for the other three tested locations as shown in the images below:



Östermalm, Stockholm


Voleri
Spilve
Kleisti
Mežāres
Ķīšezers
Pētersala
Brasa
Skanste
Andrejsala
Teika
Ciekurkalns
Jugla
Mežciems
Imanta
Dzirciems
Ķipsala
Grīziņkalns
Purvciems
Dreiliņi
Zasulauks
Šampēteris
Zolitūde
Mūkupurvs
Maskavas forštate
Dārzciems
Tornakalns
Daugava
Škirotava
Atgāzene
Bišumuiža
Katlakalns
Ziepniekkalns
Ziedonis
Starptautiskā lidosta "Rīga"
Rīga, Latvia


Asnières-sur-Seine
Saint-Ouen-sur-Seine
Aubervilliers
Clichy
Bobigny
Les Pavillons-sous-Bois
Courbevoie
La Défense
Levallois-Perret
Pantin
Puteaux
Neuilly-sur-Seine
Noisy-le-Sec
Romainville
Le Pré-Saint-Gervais
Monceau
Suresnes
Bois de Boulogne
Bagnolet
Montreuil
Rosny-sous-Bois
Saint-Cloud
Auteuil
Grenelle
Vincennes
Fontenay-sous-Bois
Saint-Mandé
Reuilly
Boulogne-Billancourt
Plaisance
Bercy
Quartier de la Glacière
Sèvres
Issy-les-Moulineaux
Maison-Blanche
Charenton-le-Pont
Joinville-le-Pont
Montrouge
Le Kremlin-Bicêtre
Ivry-sur-Seine
Meudon
Arcueil
Fleury
Clamart
Châtillon
Bagneux
Cachan
Saint-Maur-des-Fossés
Paris, France

5. Discussion

The results show that the code worked out properly and clusters could be found. The business problem could be solved, as the resulting maps clearly show areas with a high density of nightlife venues. It has to be noted, that with the free developer account of Foursquare the amount of venues someone can get from one request is limited to 50. This means, that for each category (nightclub, pub, lounge,...) only 50 venues are obtained. If the limit is higher, the clusters would be even more informative and dense. Optimization of the machine learning algorithm's parameters and output information can make the program valuable to be integrated into tourist apps or websites.

6. Conclusion

Using the code for different cities and addresses, it could be shown that the project successfully solved the business problem despite limitations within the free developer account of Foursquare. Density based clustering was the rightly chosen machine learning algorithm for this purpose. The notebook is sufficient to give tourists or travel guides an overview of popular areas to go with interesting nightlife in a city.