DANCE STUDIO IN BUDAPEST

COURSERA DATA SCIENCE CAPSTONE PROJECT – GABOR NEMETH

TABLE OF CONTENTS

BUSINESS PROBLEM AND BACKGROUND

- This project intends to identify the optimal place for a dance studio in Budapest.
- Stakeholder is my brother, but the results of my analysis could be interesting to anyone having similar plans in the capital of Hungary .
- I will be searching for areas that have no or just a few dance studios nearby, but which are easily accessible for a number of people without cars.
- Finding the best spot is not in my intention, I will attempt to locate the top 3 neighborhoods to help narrow down the list of potential areas to consider for a new dance studio.

DATA 1

What data would be useful/needed?

- The number and location of existing dance studios in Budapest as of today.
- The distance of the studios from the center of the city.
- The average ratings the studios received from the users.
- Crime data of the neighborhoods/districts.
- Entry fees or hourly rates of the studios.

DATA 2

What data are available?

- Foursquare page has various but limited information about the places. Useful for my purpose from these are the geo ones and the ratings, so I will concentrate on retrieving those primarily, via the Foursquare API.
- Understanding the crime heatmap about the districts would be an asset to find a rather secure location. Unfortunately, I did not find any available dataset about the criminal records nor on Foursquare neither on any other website, so I put this view aside for the current analysis.
- As for the prices, I have no information about the fees and rates for those studios the Foursquare site lists. Hence, I postpone this aspect too for a later study.

DATA 3

	name	lat	lng	dist	addr	id	cat
0	MIRAVOS Dance Studio	47.510802	19.033793	1470	Bajcsy-Zsilinszky út 66	50c2397ce4b0dbacdb37117f	Dance Studio
1	Aerialarts Pole dance studio	47.494974	19.021891	1447	NaN	53bd64d7498e6200782fb06c	Sports Club
2	Professional Pole Dance Stúdió	47.507802	19.058340	1704	1066. Budapest, Jókai utca 26.	52d25d9511d2066ed9545fc3	Gym

I will first capture the latitude and longitude coordinates of Budapest via an API call,

Then pull the list of venues in the city, transform the json file into a dataframe and clean it
Pre-processing

Slice the venue part from the JSON file and transform it into a pandas dataframe,
Tidy up the category so that it displays the relevant part, the name only,

Filter out the needed columns and rename them

After these steps the head of the dataframe takes the above shape:

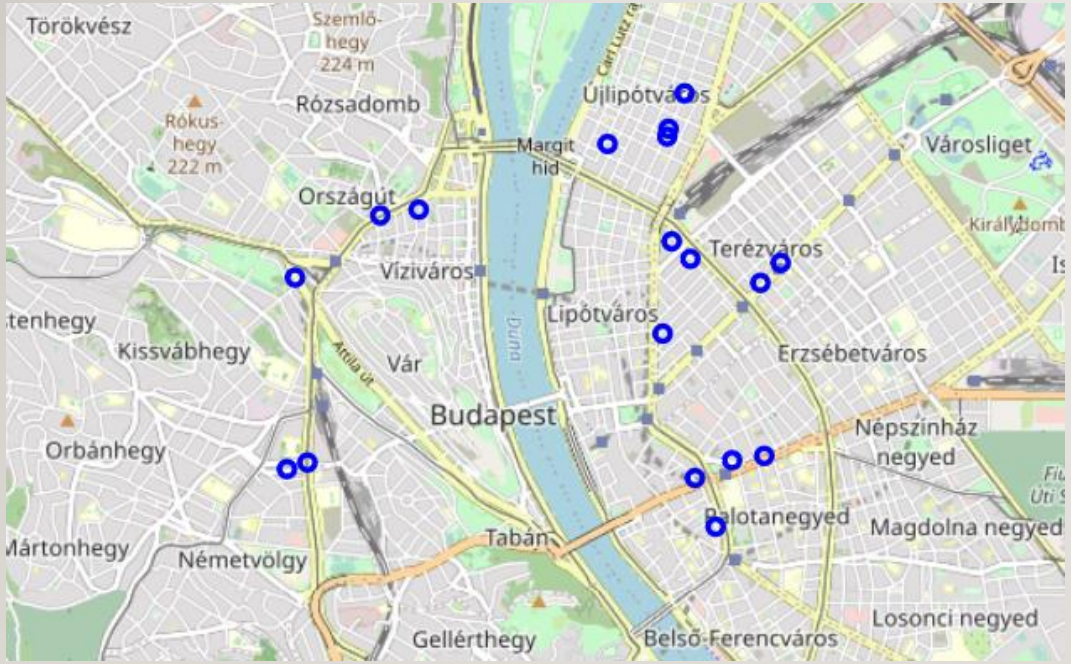
METHODOLOGY AND ANALYSIS 1

This project focuses on analyzing the dance studio density. I will use folium to visualize their spread with a heatmap first, looking for the low numbered areas and then use k-means clustering machine learning technique (as we are talking about location data) to group the studios based on their location and see if further patterns, insights might become visible enabling better exploration for the optimal location.

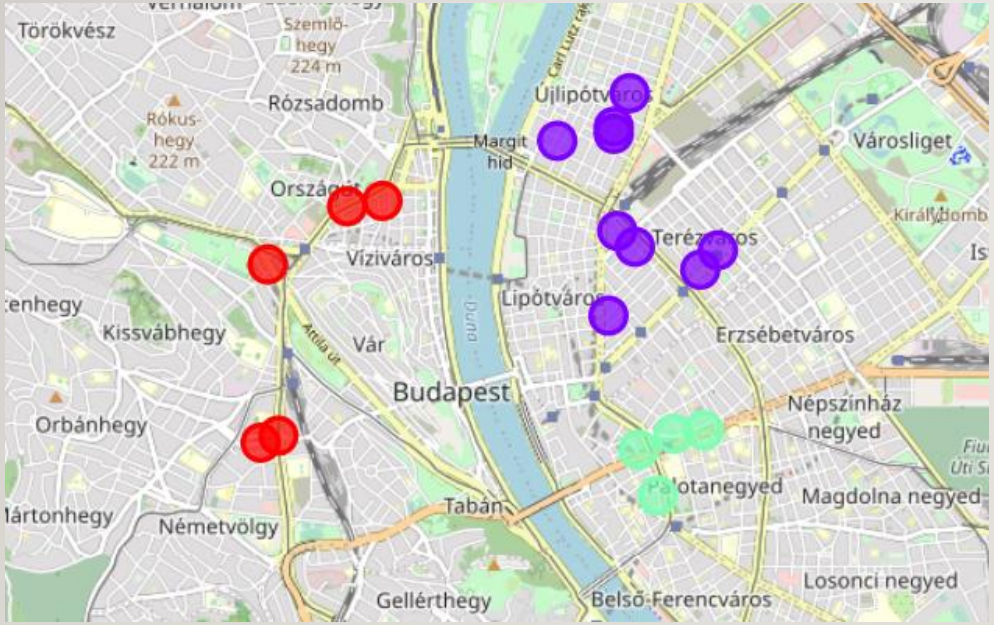
METHODOLOGY AND ANALYSIS 2

- There is an average of 1776 meters of distance from city center and a standard deviation of 361 meters. This indicates the dance studios are not straight in the so-called center, but a bit further away. I'll use the folium map to visualize them and see if there is any noticeable pattern about their location.
- Seeing the max 2534 meters gives me a hint on what radius I could have chosen earlier on but seem with the 5000 I made no mistake of closing out any places.

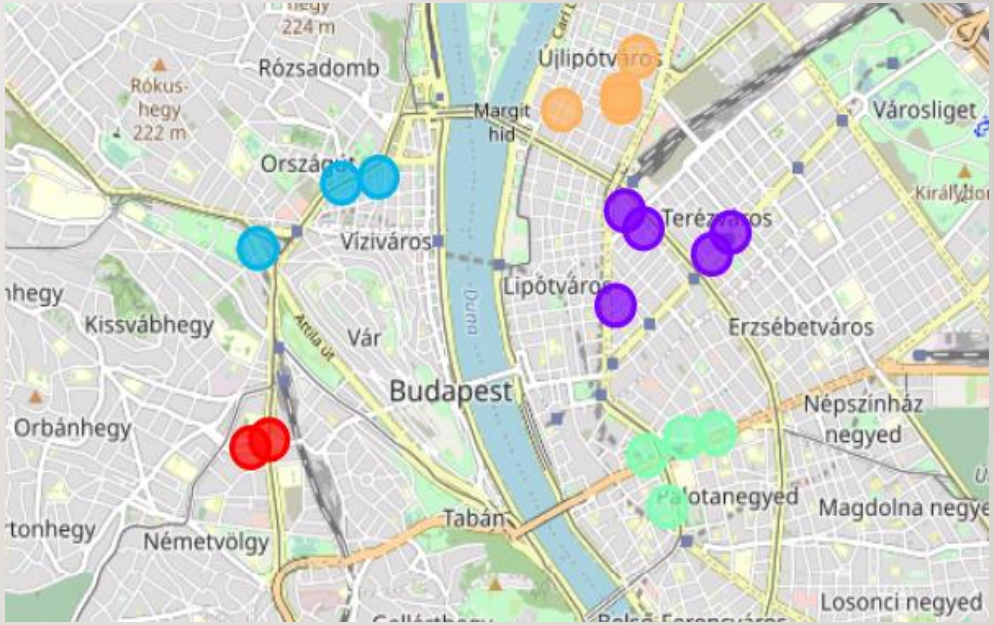
METHODOLOGY AND ANALYSIS 3 - DENSITY



METHODOLOGY AND ANALYSIS 4 - CLUSTERS



3 clusters



5 clusters

RESULTS AND DISCUSSION

- The slightly large distance from the city center seems confirmed.
- Existing studios tend to be around, or very close to the inner and middle boulevard of the city.
- The Buda side (left to river Danube) is less "crowded" of dance schools.
- Studios are relatively close to railway stations.

CONCLUSION

- This project had the purpose to find an ideal location of a new dance studio in Budapest. Under ideal I understood, being in an attracting neighborhood but also being away from the rest of the schools. I pulled and wrangled the Foursquare data about the currently existing (and listed) dance studios, their names, location, categories, etc.
- I used the clustering technique and complemented my analysis with a heatmap for density visualization.
- I wanted to offer a top 3 options about the location based on other schools locations.
- Final decision is to be made by the stakeholder(s).