

The Battle of Neighbourhoods - Week 1

Introduction & Business Problem:

Problem Background:

Being able to express yourself to an audience is important for aspiring musicians. While uploading your tracks to web audience lets you gain publicity, it's just as important to perform live. It might be hard to do if you live in a small city or one without any live music clubs. Analysing foursquare data allows determining what city has the best distribution of music clubs.

Problem Description:

An up-and-coming indie music band is looking for a city to move. They decided they will choose one of major Polish cities: Cracow, Poznan, Warsaw or Wroclaw. It's important for them to live in a city where they will have opportunities to play their songs and listen to other fresh music groups. The city with most music venues and music stores with best spatial distribution and ratings of facilities will win.

Target Audience:

Finding the best location for musicians will interest both artists and their audience.

Success Criteria:

The success criteria of this project is recommending the best of selected cities based on availability of music venues.

Data

This project is based on Foursquare API. The coordinates of each city were specified using *geopy library*. Venues matching defined categories - Jazz and rock clubs, music stores and their details were found using Foursquare look ups.

Data examples

geopy

In [3]:

```
from geopy.geocoders import Nominatim #import the function
geolocator = Nominatim(user_agent="capstone") #define geolocator
location = geolocator.geocode('New York City') #Look up the city
location
```

Out[3]:

```
Location(New York, United States of America, (40.7127281, -74.0060152, 0.0))
```

In [4]:

```
latitude = location.latitude
longitude = location.longitude
"{}, {}".format((latitude),(longitude))
```

Out[4]:

```
'40.7127281, -74.0060152'
```

Thanks to geopy we can easily find needed coordinates.

Foursquare

In [7]:

```

import numpy as np # Library to handle data in a vectorized manner
import pandas as pd # Library for data analysis
import requests # Library to handle requests
from pandas.io.json import json_normalize # transform JSON file into a pandas dataframe
###
CLIENT_ID = 'JQYNPSOQXI12NK311020I2K4FAKMKEV25CMLKTGSEUEVMUR' # your Foursquare ID
CLIENT_SECRET = 'XRECAKXLEOK324BN2Z3KDYNK1NBJ3UDYYS45PQAJ2LIM0BFX' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version

print('Your credentials:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET: ' + CLIENT_SECRET)
###
LIMIT = 100
cities = ['Cracow']
results = {}
for city in cities:
    url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&near={}&limit={}&categoryId={}'.format(
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        city,
        LIMIT,
        "4bf58dd8d48988d1e7931735,4bf58dd8d48988d1e9931735")
    results[city] = requests.get(url).json()
df_venues={}
for city in cities:
    venues = json_normalize(results[city]['response']['groups'][0]['items'])
    df_venues[city] = venues[['venue.name', 'venue.location.address', 'venue.id', 'venue.location.lat', 'venue.location.lng']]
    df_venues[city].columns = ['Name', 'Address', 'id', 'Lat', 'Lng']

```

Your credentials:

CLIENT_ID: JQYNPSOQXI12NK311020I2K4FAKMKEV25CMLKTGSEUEVMUR
 CLIENT_SECRET: XRECAKXLEOK324BN2Z3KDYNK1NBJ3UDYYS45PQAJ2LIM0BFX

In [18]:

```

df_venues['Cracow'].head(2)
df=df_venues['Cracow']
df.head(2)

```

Out[18]:

	Name	Address		id	Lat	Lng
0	Harris Piano Jazz	Rynek Główny 28		4bbf90a2b492d13a4fdfa260	50.061703	19.935660
1	Jazz Rock Cafe	Sławkowska 12		4befc97724f19c747926f983	50.063866	19.937492

First Foursquare request returns music clubs in defined cities. With ids of the venues we can look up more details about them:

In [20]:

```
lookup_id = df['id'][0]
URL='https://api.foursquare.com/v2/venues/{}/tips'.format(lookup_id)
venue = requests.get(URL).json()
```

In [21]:

```
URL = 'https://api.foursquare.com/v2/venues/{}?&client_id={}&client_secret={}&v={}'.format(
    lookup_id,
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
)
grab = requests.get(URL).json()
```

In [22]:

```
grab['response']['venue']['rating']
```

Out[22]:

7.9

Location and rating data will be used to determine the best city for young artists.

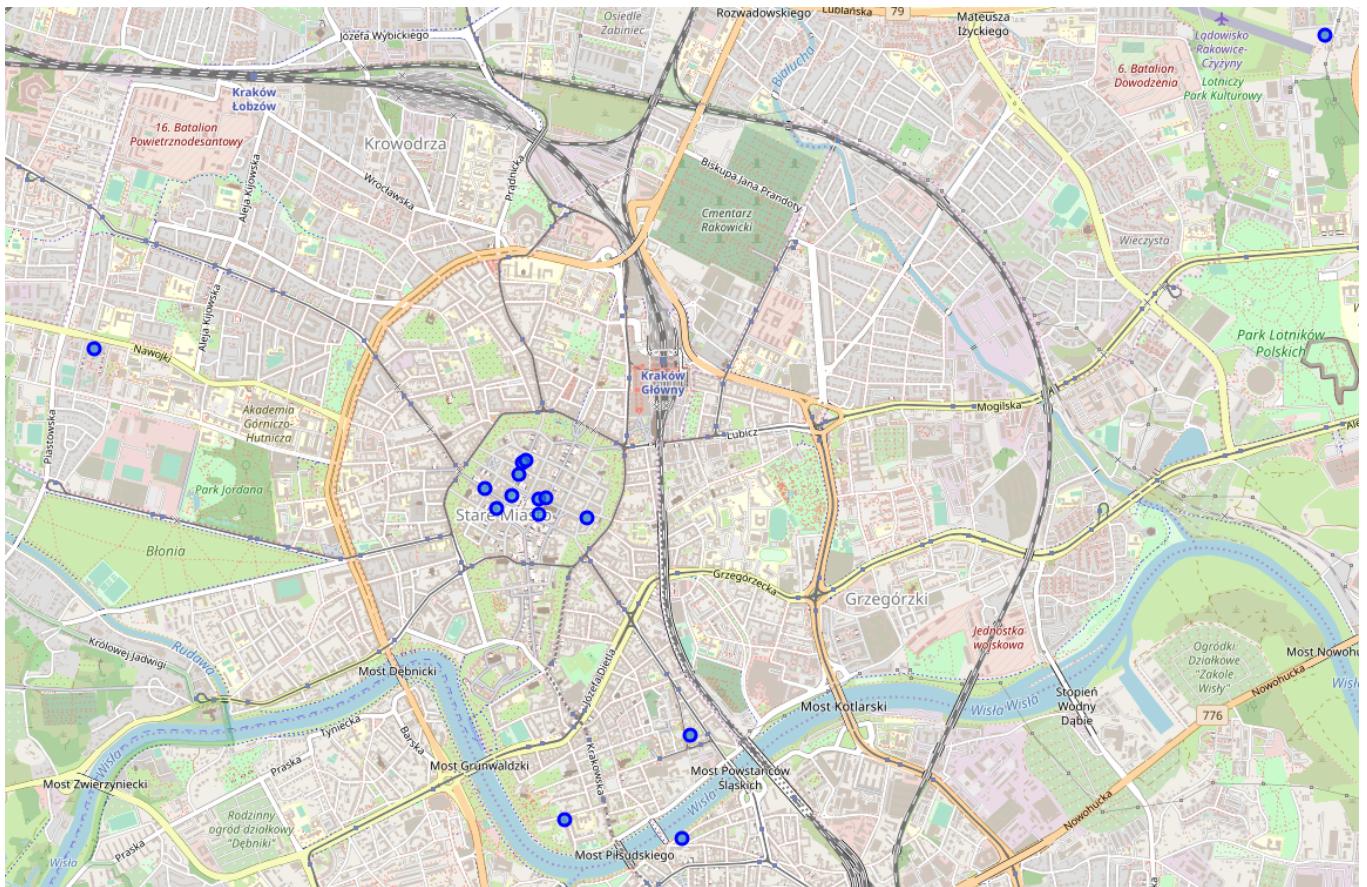
Methodology:

The main objective is to find the city with best music clubs and stores while considering their distribution. In first step information about music-related venues is fetched from FourSquare in each city. Then the venues were clustered by their location - the biggest cluster was considered a representation of a city - the distance within cluster as simplified distribution of the venues and their rating.

Results:

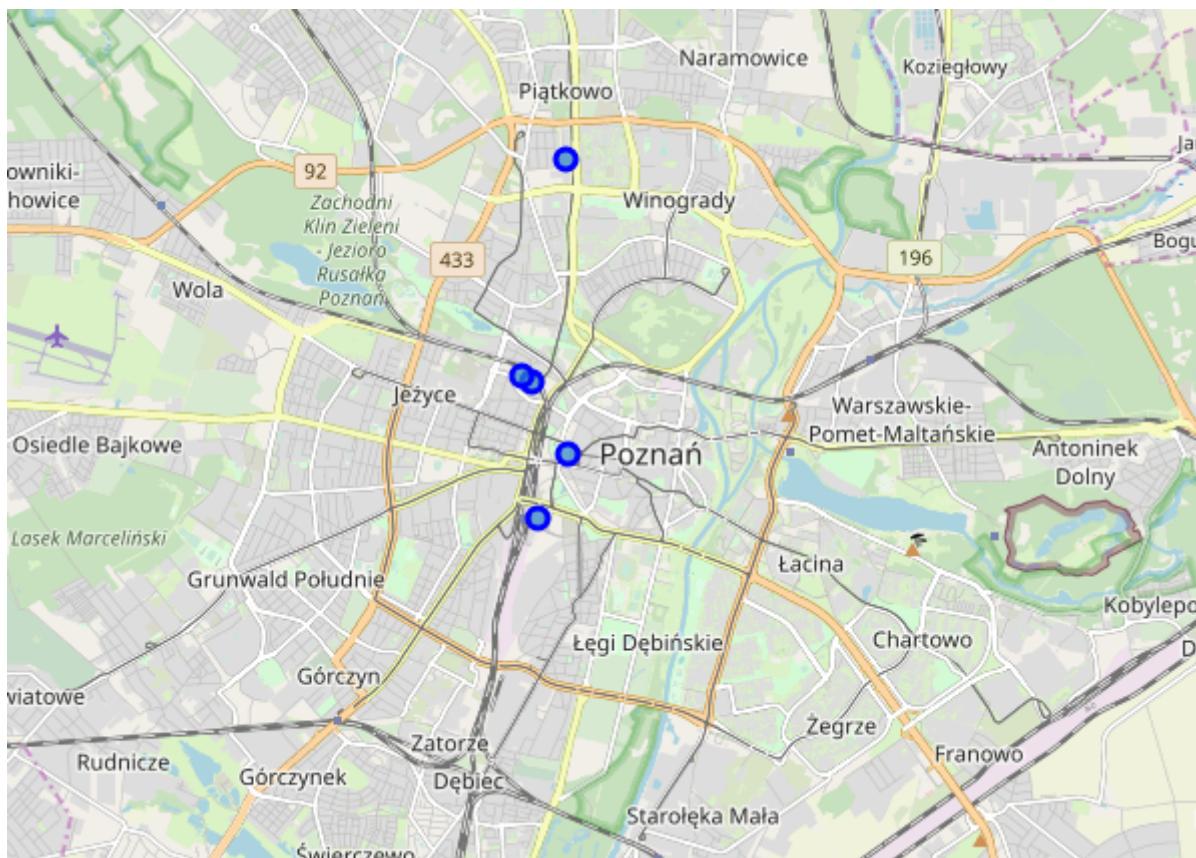
Finding the venues

Cracow:



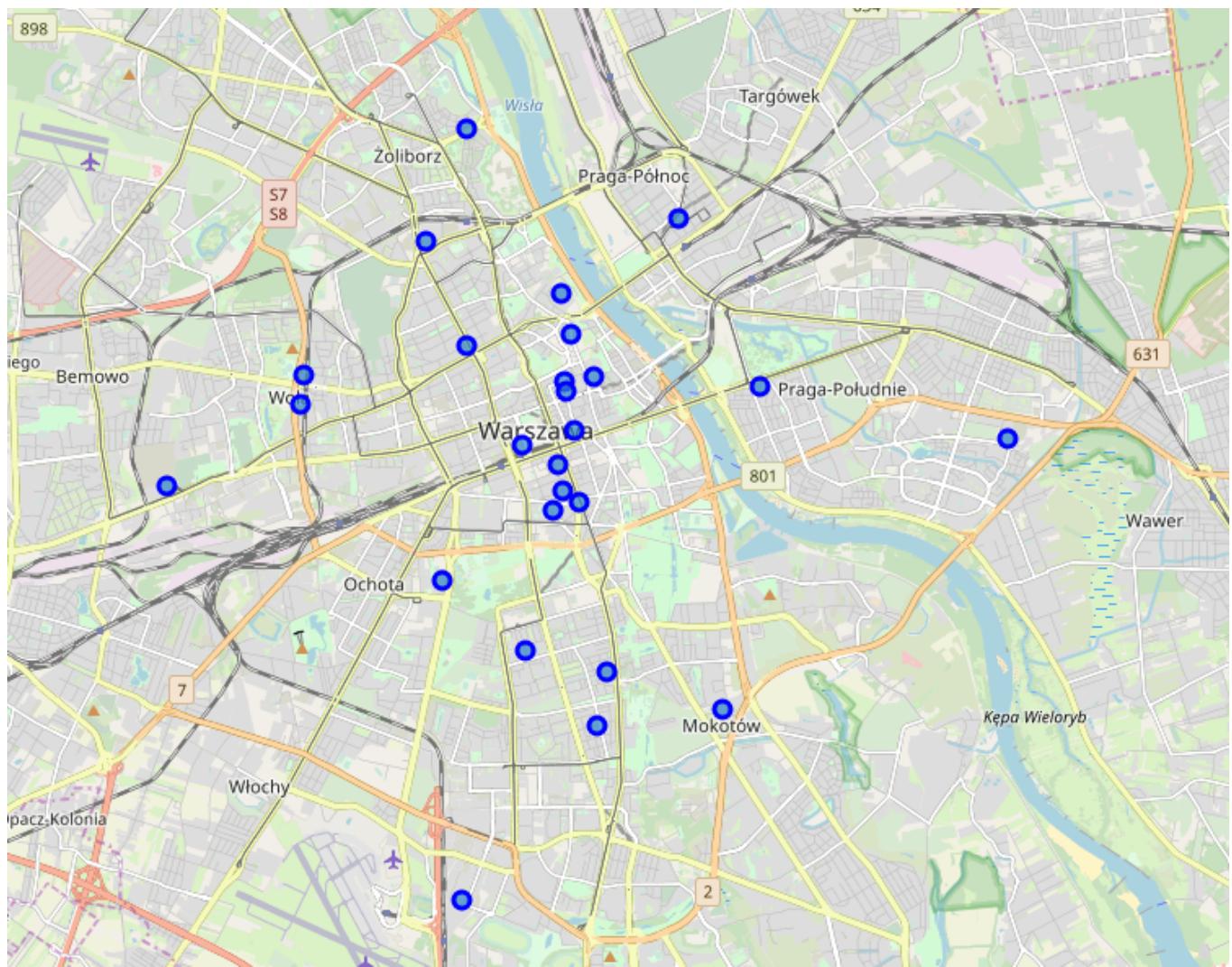
There are 15 music clubs and stores in Cracow. Most of them are concentrated old town. There are two single outliers - one far east and one east and a group of three venues Podgorze district.

Poznan:



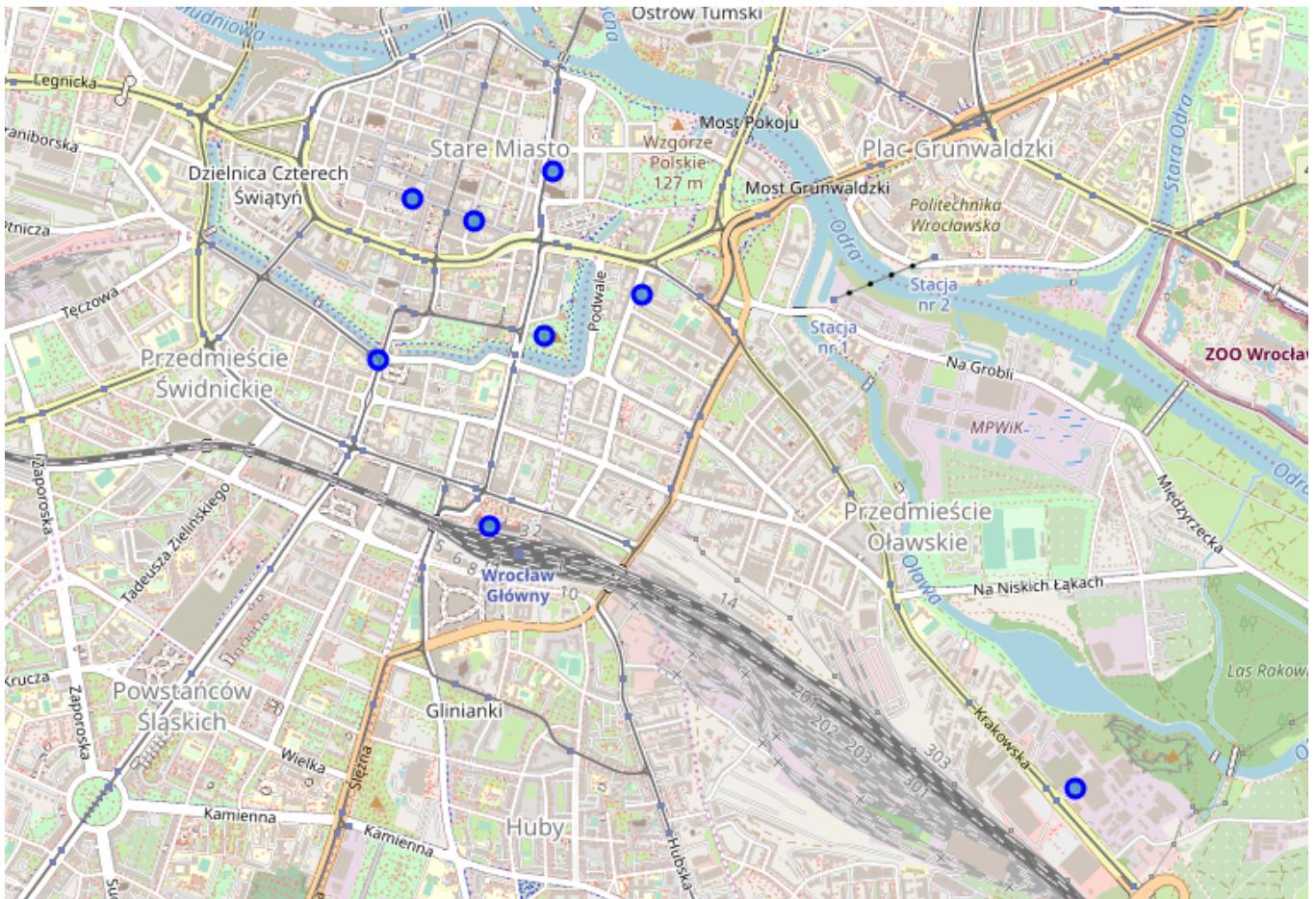
In Poznań there are 5 venues, distributed roughly around straight line.

Warsaw:



There are 26 venues in Warsaw, spread out all over the city in different districts, with majority of them in the city centre.

Wroclaw:

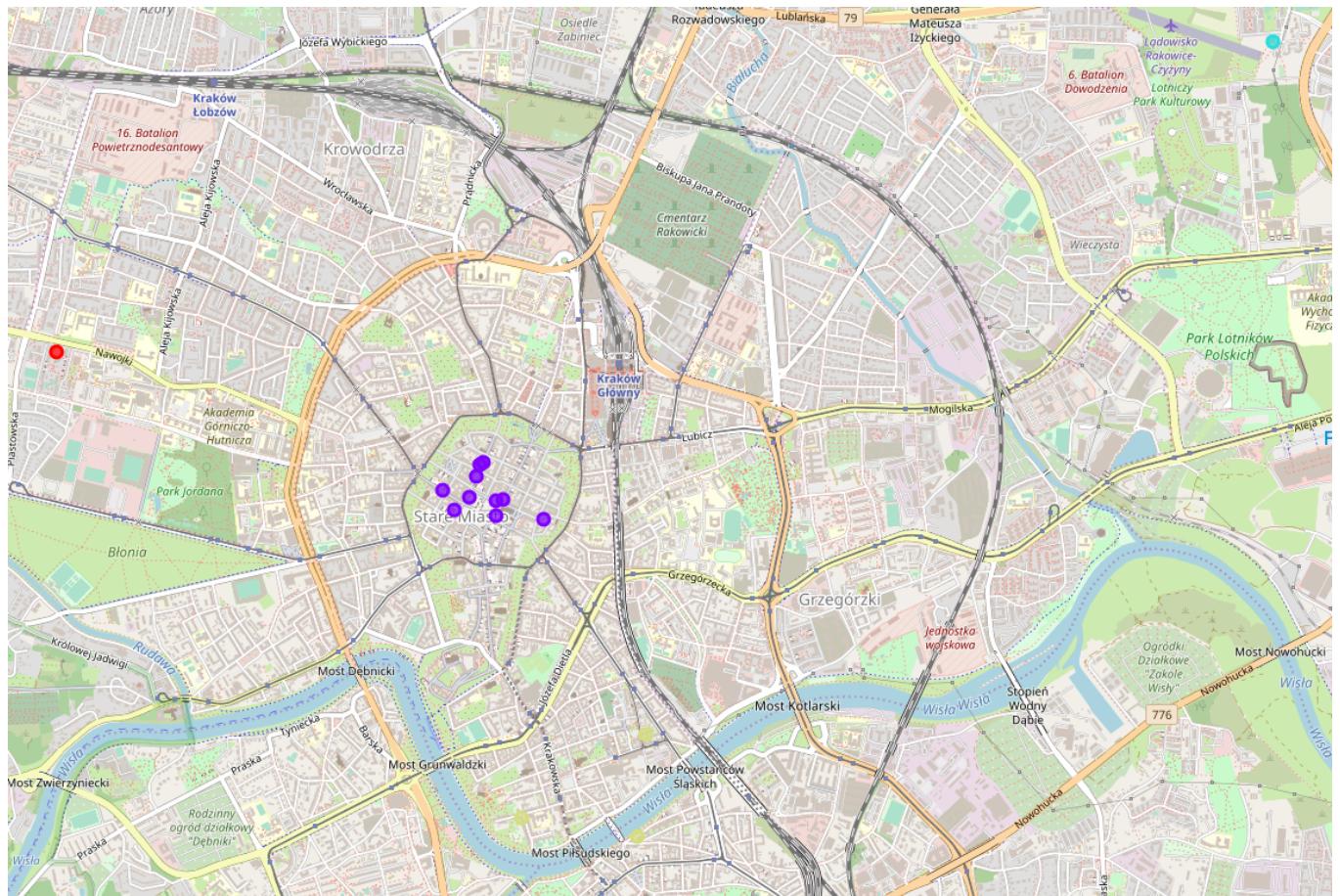


There are 8 venues of interest in Wrocław. Their distribution is similar to Cracow's, most of the venues are located close to Old Town district.

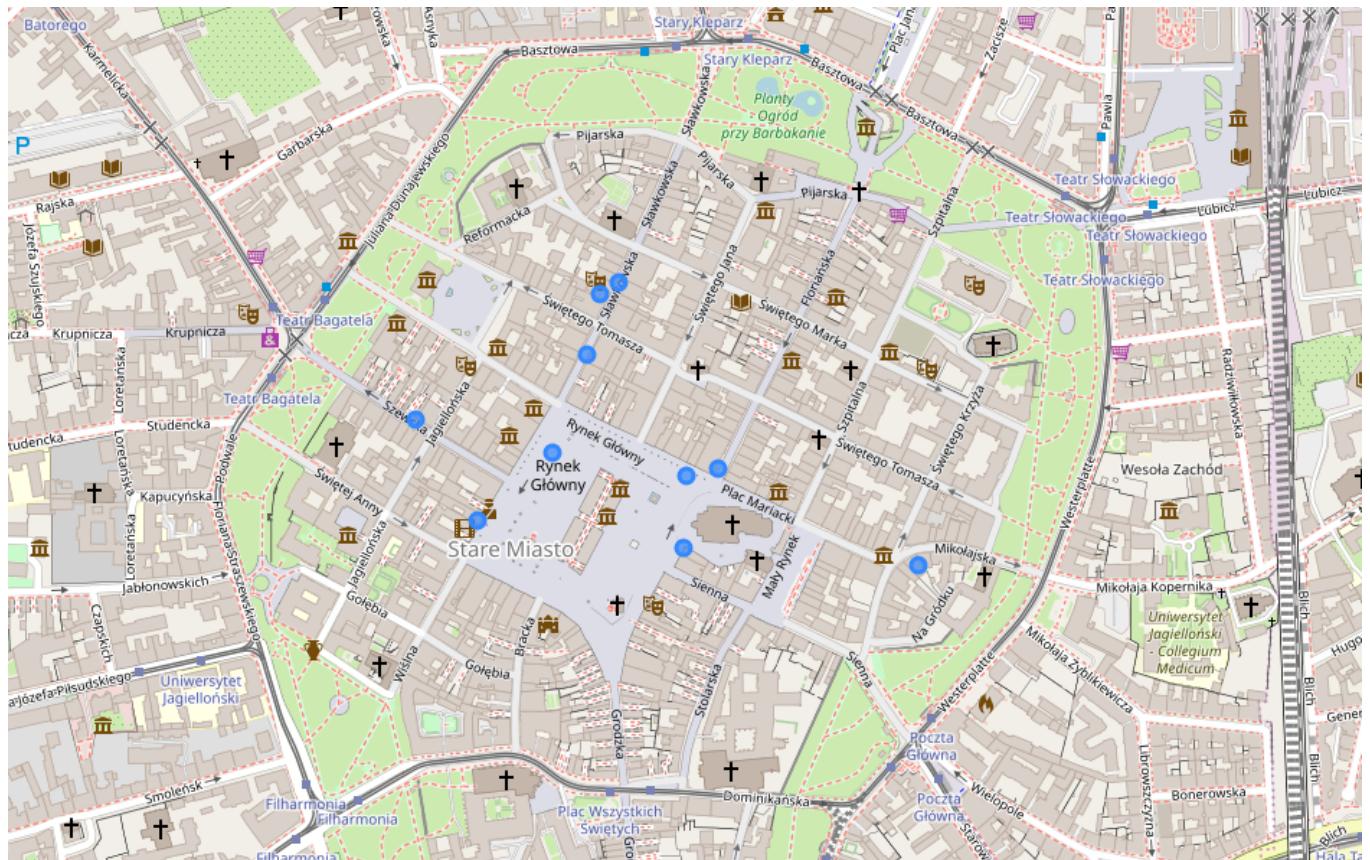
Clustering

In order to better compare the cities, the venues were clustered using k-means clustering by their location. This allows finding the city with best spatial distribution of venues.

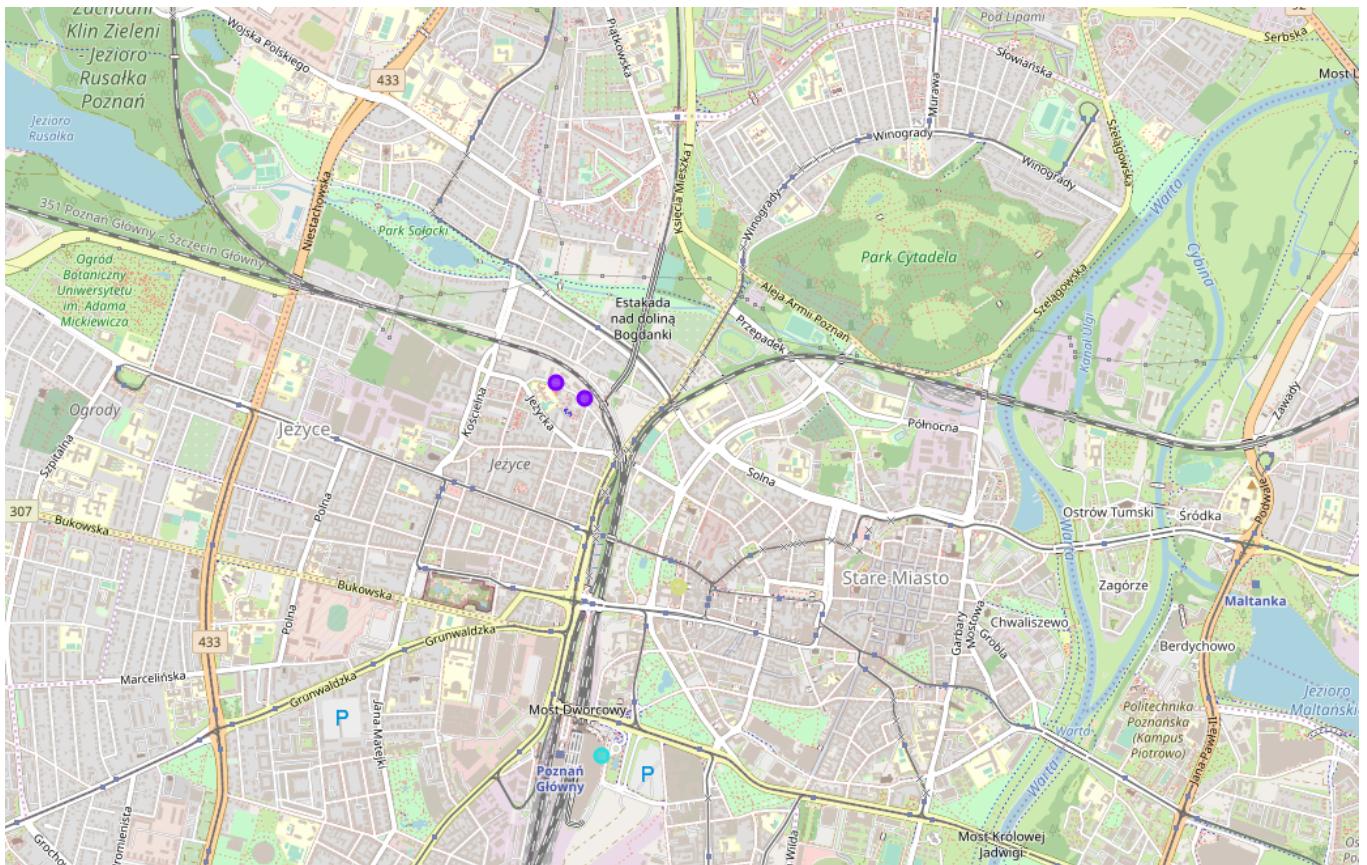
Cracow:



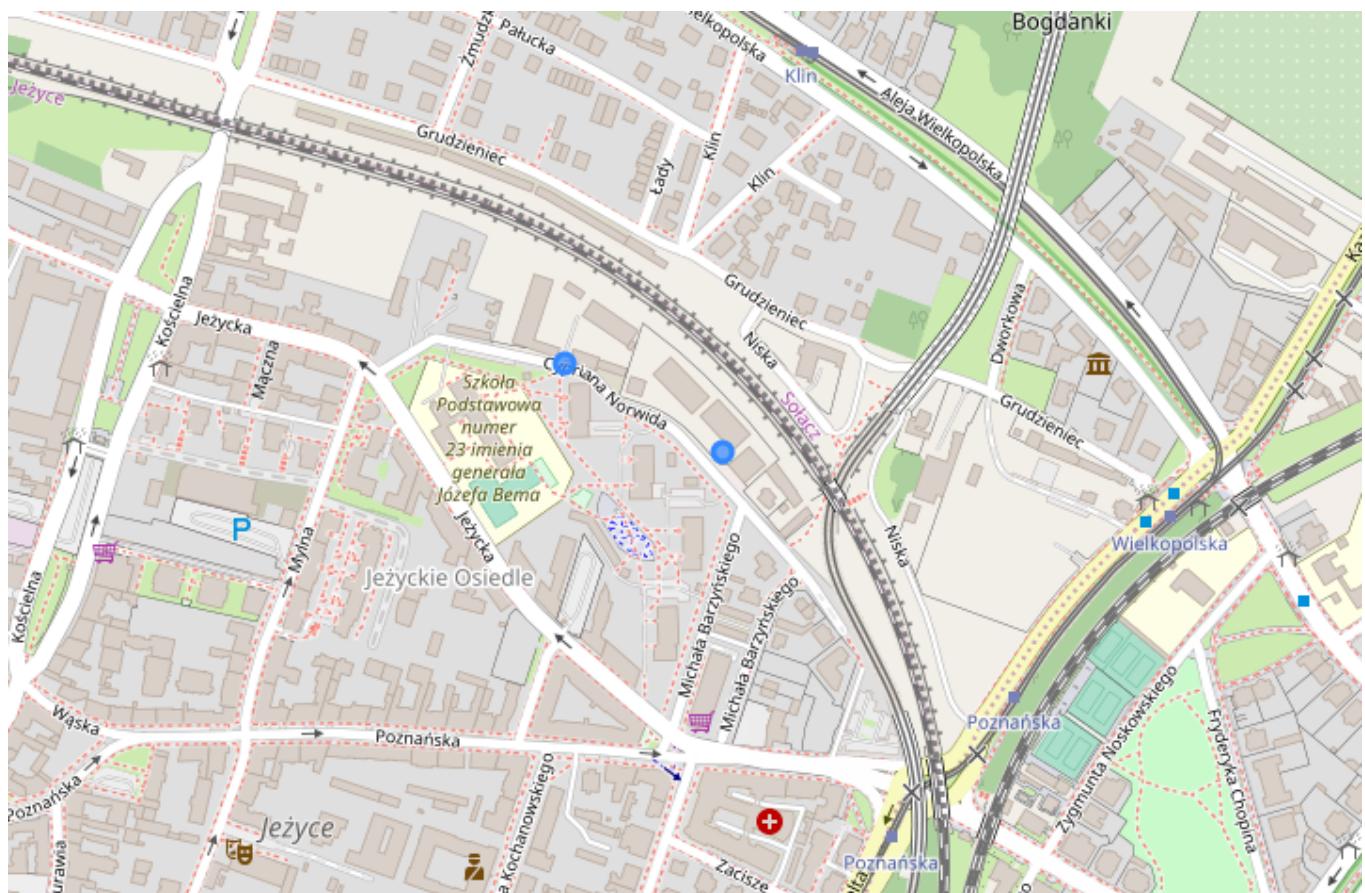
Cracow was divided into 4 clusters, with the most populated containing 10 venues.



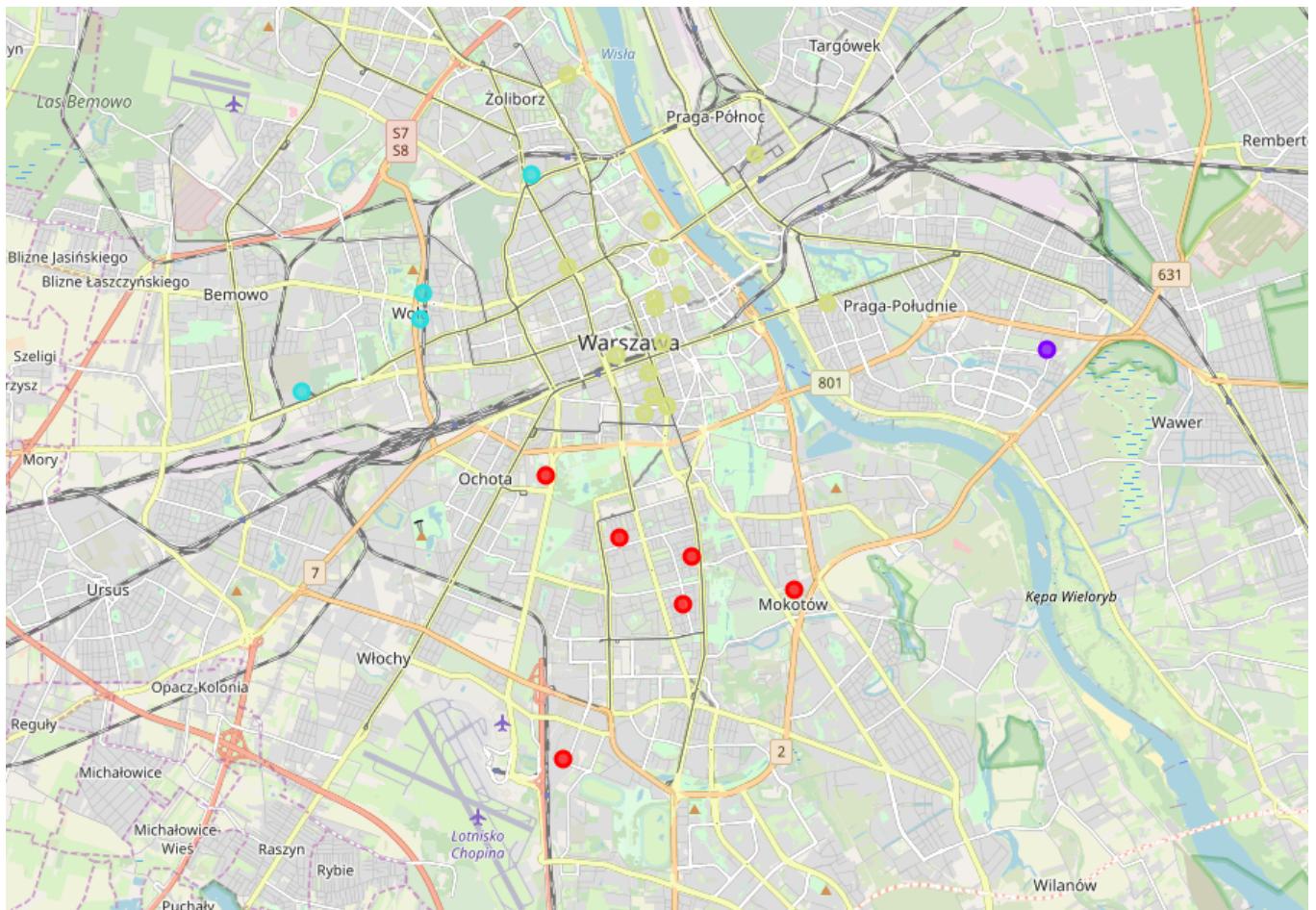
Poznań:



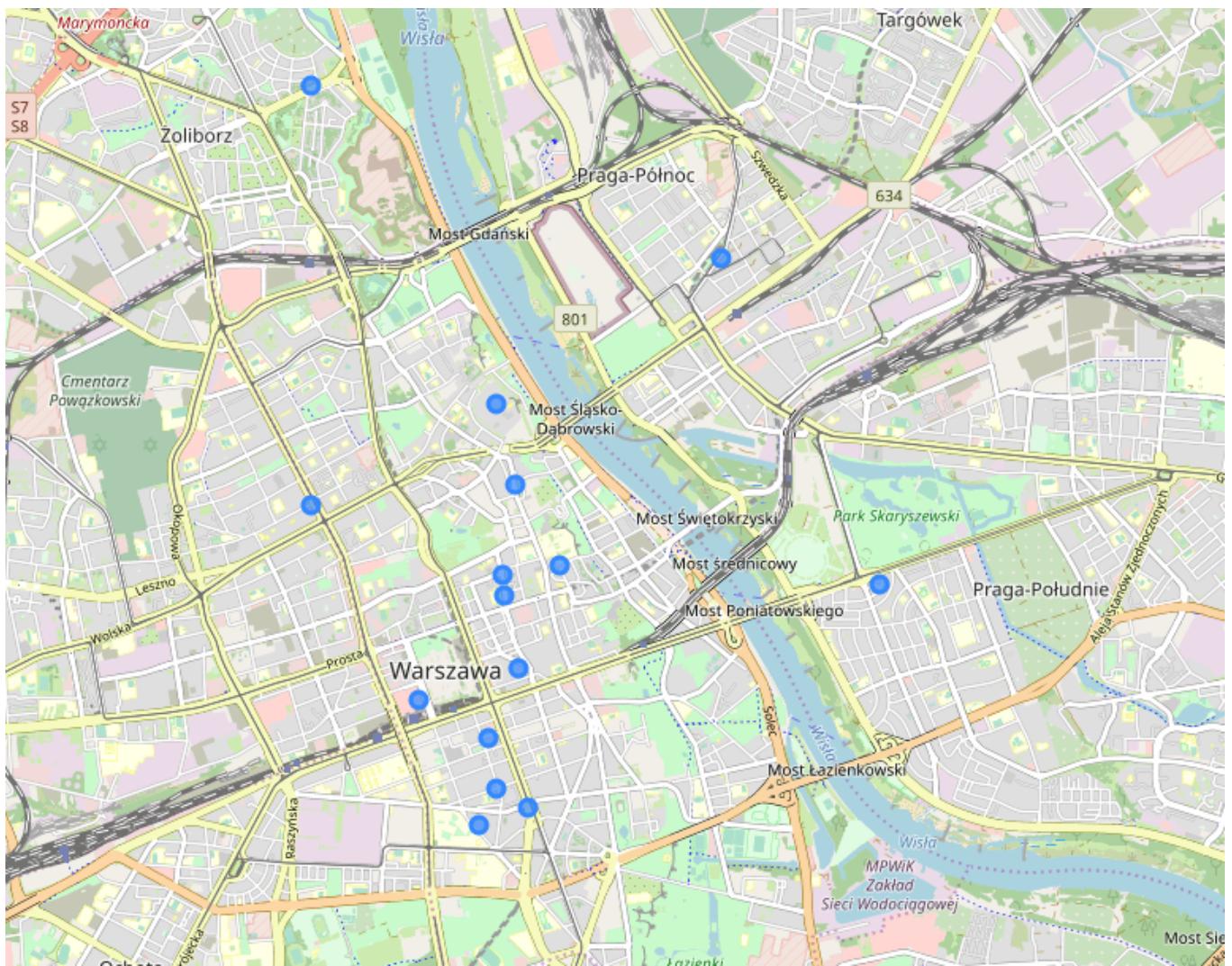
The winning cluster in Poznan has only 2 venues. At this point it's obvious that Poznan is the worst city for the objective of this project.



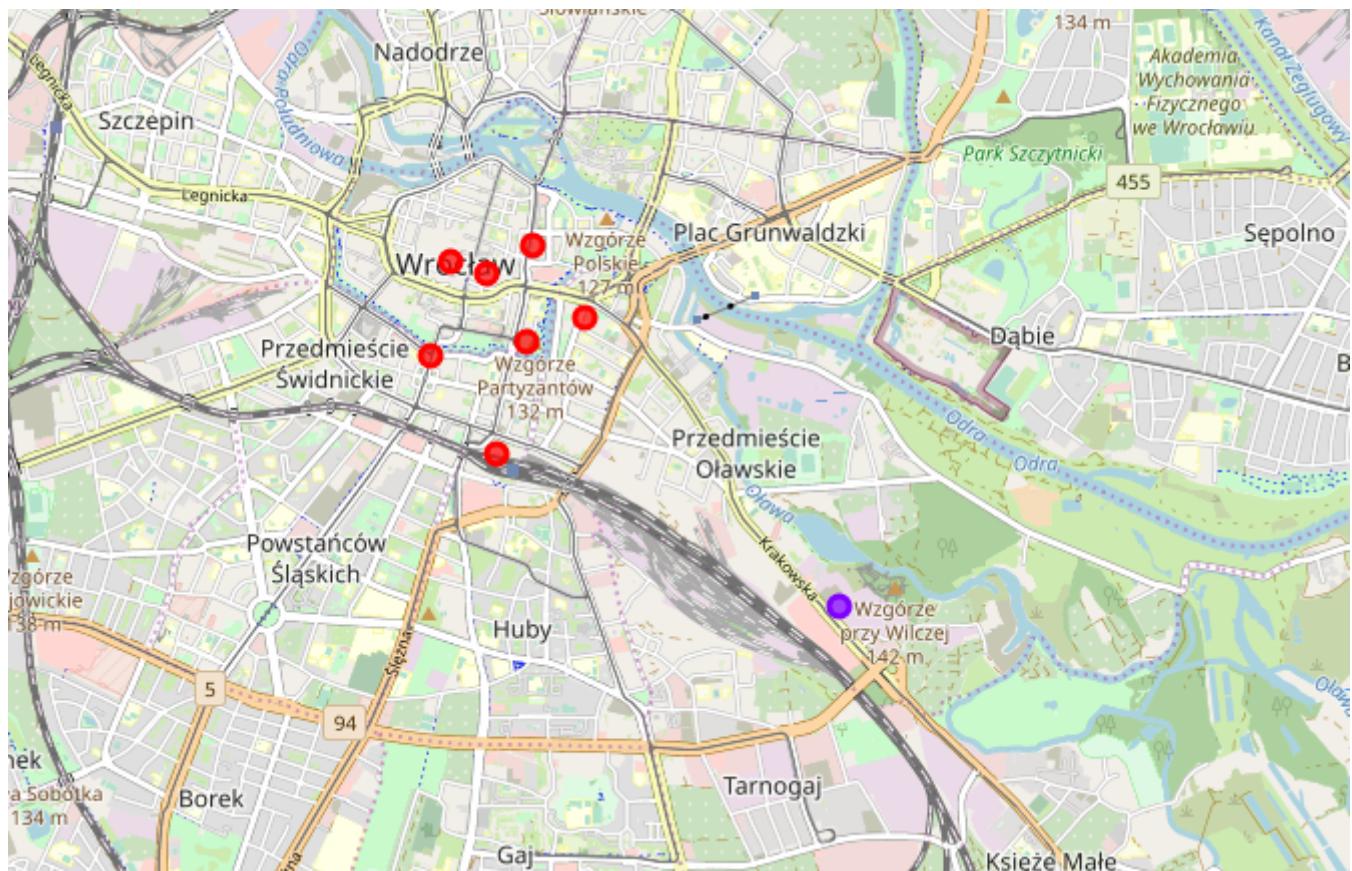
Warsaw:



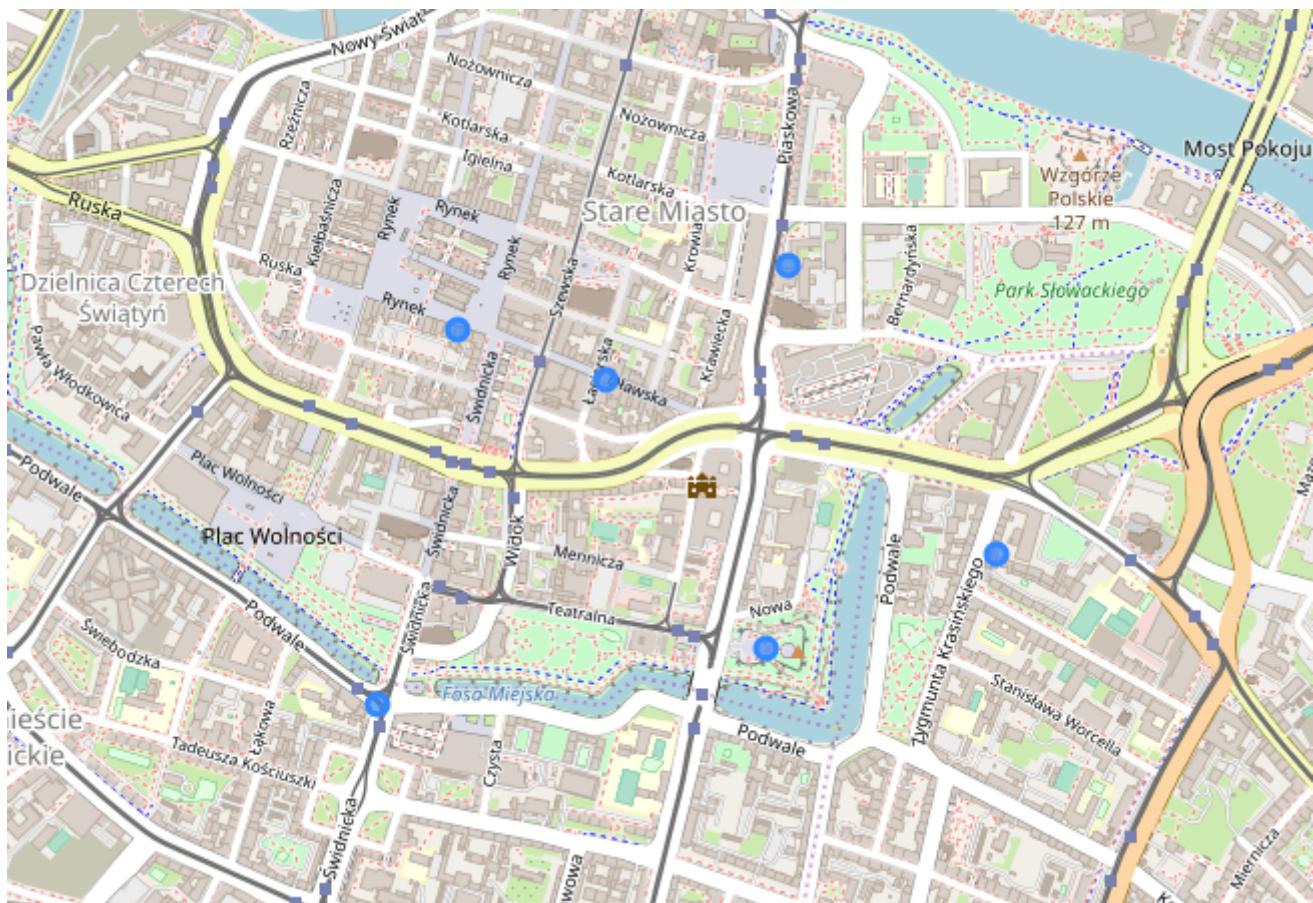
The biggest cluster in Warsaw is populated by 15 venues. Warsaw has both most venues overall and when considering the biggest cluster.



Wrocław:



Biggest cluster in Wrocław contains 7 venues.



Distance within clusters

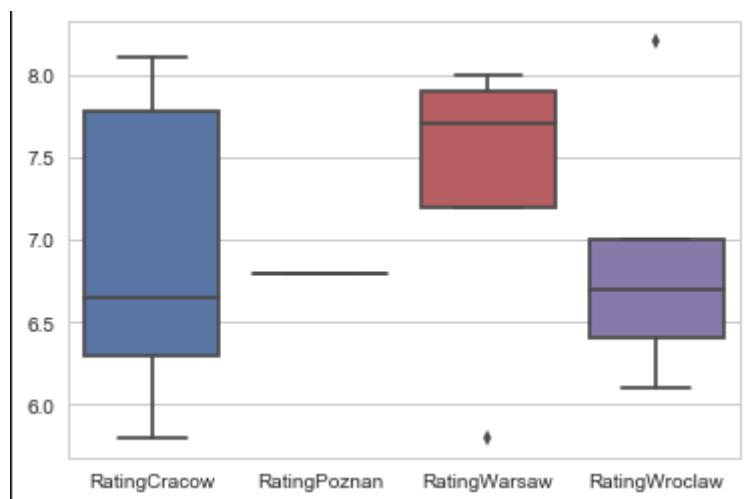
The distance within clusters was calculated using city block distance, as it's a more sensible metric than Euclid distance.

City	Number of venues in biggest cluster	Mean distance
Cracow	10	0.024
Poznan	2	0.021
Warsaw	15	0.059
Wroclaw	7	0.018

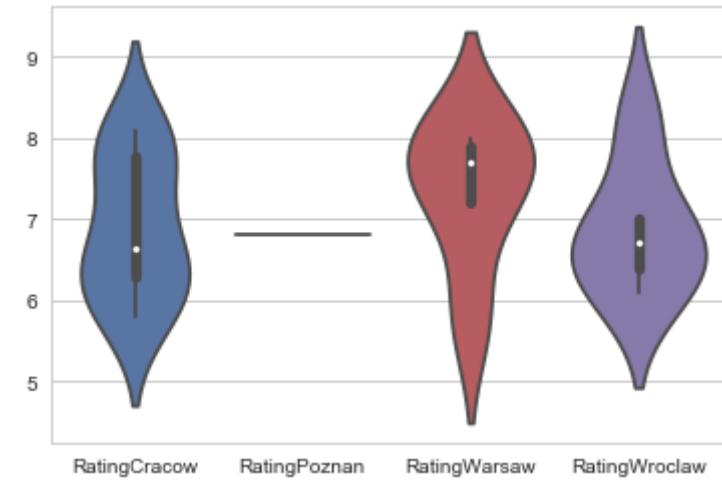
Venue rating in winning clusters

	Cracow	Poznan	Warsaw	Wroclaw
count	10	1	5	5
mean	6.89	6.8	7.32	6.88
std	0.86	NaN	0.90	0.81
min	5.8	6.8	5.8	6.1
25%	6.3	6.8	7.2	6.4
50%	6.65	6.8	7.7	6.7
75%	7.775	6.8	7.9	7.0
max	8.1	6.8	8.0	8.2

A limited number of venues were rated - for example in Warsaw only 5 of venues were rated compared to 15 venues in cluster. In Cracow all venues are rated, meaning there are more City Guide users there than other cities, possibly meaning wider audience.



Cracow and Wroclaw are similar in both spatial distribution of venues and their ratings. Cracow has twice as much rated venues as Warsaw and Wroclaw. The median rating is highest in Warsaw, but it may be skewed by low number of rated venues – results could change when considering rest of the venues in cluster.



Violin plot shows flatter distributions of venues in Cracow than in Warsaw or Wroclaw. The distribution in Warsaw is top heavy, while in Wroclaw its bottom heavy.

Discussion

Poznan lacks music facilities. Cracow and Wroclaw are similar when considering venues' location and ratings.

While the biggest cluster in Warsaw had 15 venues, only 5 of them were rated, which may indicate that Cracow is more popular among City Guide users. Further characterization of an average City Guide user is needed for more complex conclusions.

Based on this analysis, Cracow is the best city for up-and-coming music band. There are many music venues – both clubs and stores with good ratings. Secondary choice is Warsaw – it's the biggest city in Poland with most venues, but due to limited information rating data is uncertain. While free Foursquare API access gives access to vast amounts of data, query limit can be an issue.

Conclusions

The best city for up-and-coming indie music band is Cracow - it has good amount of quality music clubs and stores and is a popular tourist destination. Close second is Warsaw, thanks to being the biggest city in Poland it has the most venues overall, but not many of them popular among City Guide users.

Wroclaw is a smaller city with less venues, but when considering city centre the results were similar to Cracow.

Poznan is the worst choice - it simply lacks venues.