Choosing Location for a Jazz Cafe in Zagreb

Natalia Odintsova

May, 2019

1. Introduction - Business Problem

For many years Zagreb has been attracting tourists and businesses from all over the world. The capital of Croatia and its biggest city, Zagreb is known for its interesting architecture, pedestrian zones and parks, and rich cultural life.

> The impressions of many foreigners who live in Zagreb, business people and tourists can be summed up in a single sentence: a large city which managed to stay romantic and safe.
>
> www.europeanbestdestinations.com

This report is targeted to stakeholders who are considering opening a jazz cafe in Zagreb. Zagreb has a vibrant nightlife, with numerous cafes, bars, nightclubs, and lounges. They are mostly concentrated in the city center. Although this seems an ideal location for a jazz cafe, the competition among musical venues in this area may be high. In this project, we will explore the neighborhoods of Zagreb and try to **determine what would be the best location for a jazz cafe**.

We assume that a desirable location for a jazz cafe would be a neighborhood that is similar to the city center, but not necessarily close to it. It would be located in an 'artistic' environment - close to arts galleries, theaters, performance arts centers, etc. It would have some nightlife spots nearby, but preferably no competing music venues, such as jazz cafes or clubs.

2. Data

2.1. Data Sources and Acquisition

2.1.1. Geolocations of neighborhoods

For geolocations of the neighborhoods we will be using zagreb_croatia_places.geojson file that can be found at https://www.nextzen.org/metro-extracts/index.html#zagreb_croatia

The file contains features for 169 locations. Here is an example of a feature:

{'type': 'Feature',
 'properties': {'id': 2.0,
  'osm_id': 331099862.0,
  'name': 'Dugave',
  'type': 'neighbourhood',
  'z_order': 2.0,
  'population': None},
 'geometry': {'type': 'Point',
  'coordinates': [15.997999421868, 45.76503689132497]}}

From each feature we store name, type and coordinates. The data was used to populate a pandas data frame. Below is the first several entries.

	Type	Name	Latitude	Longitude
0	neighbourhood	Dugave	45.765037	15.997999
1	neighbourhood	Botinec	45.754697	15.936869
2	neighbourhood	Travno	45.770753	15.997835
3	neighbourhood	Utrina	45.775019	15.997538
4	neighbourhood	Lanište	45.772431	15.947718

2.1.2 Population Densities

To get the data on population densities of districts of Zagreb we scraped the wiki page https://en.wikipedia.org/wiki/Districts_of_Zagreb. The data was stored in a dataframe.

	District	Size	Population	Density
0	Donji grad(Lower town)	3	37024	12341
1	Gornji Grad–Medveščak(Upper town–Medveščak)	10	30962	3096
2	Trnje	7	42282	6040
3	Maksimir	14	48902	3493
4	Peščenica – Žitnjak	35	56487	1614

2.1.3. Venues Data

We use **Foursquare API** to get venues in each neighborhood. We use **explore** request:

https://api.foursquare.com/v2/venues/explore

with the following **parameters:** radius=500, limit=100, day='any', time='any', openNow=0

The last three parameters are needed to ensure that we get venues independently on whether they are currently open or not. Here is an excerpt from a response for one neighborhood:


```
[{'reasons': {'count': 0,
   'items': [{'summary': 'This spot is popular',
     'type': 'general',
     'reasonName': 'globalInteractionReason'}]},
  'venue': {'id': '4c3c575a980320a16f778ae4',
   'name': 'Buffet Mac',
   'location': {'lat': 45.76398806616097,
    'lng': 16.001122737201435,
    'labeledLatLngs': [{'label': 'display',
      'lat': 45.76398806616097,
      'lng': 16.001122737201435}],
    'distance': 269,
    'cc': 'HR',
    'country': 'Hrvatska',
    'formattedAddress': ['Hrvatska']},
   'categories': [{'id': '4bf58dd8d48988d147941735',
     'name': 'Diner',
     'pluralName': 'Diners',
     'shortName': 'Diner',
     'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/diner_',
      'suffix': '.png'},
     'primary': True}],
   'photos': {'count': 0, 'groups': []}},
  'referralId': 'e-0-4c3c575a980320a16f778ae4-0'},
```


For each venue, we store the following: venue name, venue id, venue location, category, and category id. We combine the venues data with the neighborhood data. Here are several rows from the resulting dataframe:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Category ID
0	Dugave	45.765037	15.997999	Buffet Mac	4c3c575a980320a16f778ae4	45.763988	16.001123	Diner	4bf58dd8d48988d147941735
1	Dugave	45.765037	15.997999	Play Off	4b7a4cb4f964a520e1282fe3	45.766285	16.001196	Café	4bf58dd8d48988d16d941735
2	Dugave	45.765037	15.997999	Goya Bar	4f1b2e23e4b0288a02d11b0f	45.763122	15.992952	Bar	4bf58dd8d48988d116941735
3	Dugave	45.765037	15.997999	Caffe Lupo	4c76752cb474a1cd61f8bbbf	45.762524	16.003188	Café	4bf58dd8d48988d16d941735
4	Dugave	45.765037	15.997999	Konzum	4d59795e56f2b60ce898792f	45.769046	15.999888	Grocery Store	4bf58dd8d48988d118951735

2.1.4. Foursquare Category Hierarchy Tree

We obtained the hierarchical tree of Foursquare's categories by http GET request **https://api.foursquare.com/v2/venues/categories**

The tree was traversed to create dictionaries of categories of our interest: "Arts & Entertainment", "Nightlife Spots" and "Music Venues" categories.

```
Arts and Entertainment Categories
['Amphitheater', 'Aquarium', 'Arcade', 'Art Gallery', 'Bowling Alley', 'Casino', 'Circus', 'Comedy Club', 'Concert
...and more. Total: 65
Nightlife Categories
['Beach Bar', 'Beer Bar', 'Beer Garden', 'Champagne Bar', 'Cocktail Bar', 'Dive Bar', 'Gay Bar', 'Hookah Bar', 'Ho
...and more. Total: 25
Music Venue Categories
['Jazz Club', 'Piano Bar', 'Rock Club', 'Music Venue'] Total: 4
```

Since "Arts & Entertainment" subtree included sports, theme parks, and other categories that we are not interested in, we created a dictionary of arts-only categories.

2.2. Data Cleaning

Examining the neighborhood data obtained from the geojson file showed that 'type' may contain the following values:
['village', 'neighbourhood', 'suburb', 'locality', 'city', 'hamlet', 'town']

We excluded entries with type 'village' and 'hamlet' from our analysis, since we are interested only in Zagreb proper.

Further examining the data showed that some entries had null name. We dropped the rows with null names.
There were several entries with duplicate names:

	Type	Name	Latitude	Longitude
19	neighbourhood	Trnje	45.797163	15.974574
27	neighbourhood	Črnomerec	45.816528	15.936323
73	neighbourhood	Maksimir	45.820726	16.003952
118	suburb	Trnje	45.797038	15.982495
119	suburb	Črnomerec	45.827382	15.933554
120	suburb	Maksimir	45.833312	16.010151

After determining that locations with the duplicate names are located more than 500 m from each other, we decided to keep both entries, renaming the duplicates according to their type.

2.3. Extending the Data

Visualization of neighborhoods on the map showed that we didn't have enough points in the area north, west and east from Zagreb center, so we decided to extend the data using other sources. For the area north of the center - Gornji Grad (Upper Town) - we used data from https://en.wikipedia.org/wiki/Gornji_Grad%E2%80%93Medve%C5%A1%C4%8Dak.
We also computed midpoints to the closest neighborhoods east and west.
The combined dataset contained 151 neighborhoods. Fig.1 shows neighborhood locations on folium map.

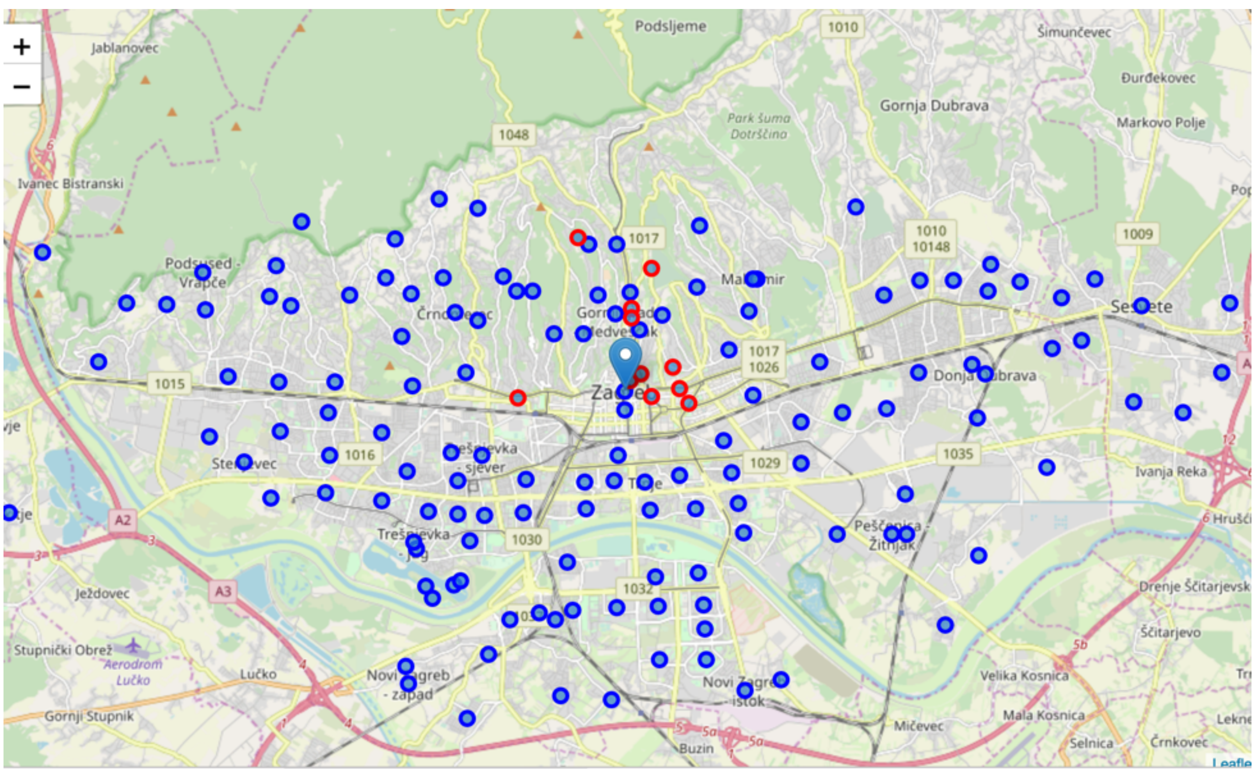

Fig. 1. Neighborhood locations shown as blue circles (geojson data) and red circles (extended data).

3. Methodology

According to the business problem that we described in Introduction section, we define the desirable location for a jazz cafe as follows:

1. It is located in a neighborhood similar to the city center. We define this as neighborhoods that have similar frequencies of venue categories (taking into consideration all venues, not only arts, nightlife and music venues).
2. It is located in an 'artistic' environment, that is, close to arts galleries, theaters, or other arts venues.
3. Have no direct competitors within 500 m radius. By direct competitors we mean music venues like jazz cafes/clubs, rock clubs, piano bars, etc.
4. Have moderate to low density of nightlife venues. We consider nightlife venues (bars, nightclubs, lounges, etc) our second-tier competitors. Some nightlife infrastructure is welcome, but we want to avoid overly hot nightlife areas.

Based on this, we performed the following analysis:

1. Used k-means clustering to determine the neighborhoods that are similar to the center. Zoom-in on those neighborhoods.
2. Using the locations of music venues, determined the areas to stay clear of.
3. Using the heat map of arts venues, determined the areas of high densities of arts venues. These are areas we want to be in.
4. Using the heat map of nightlife venues, determined areas of moderate to low densities. These are areas we want to be in.

3.1. Data Exploration and Analysis

With Foursquare API we got venues for 142 neighborhoods, 1350 venues. Histogram of number of venues returned for a neighborhood is shown on Fig. 2.

We see that for more than 80 neighborhoods we got less than 10 venues. This is more than a half of all the neighborhoods.

This can be explained by the fact that Zagreb is a big city where the population density is distributed very unevenly, with a large amount of districts with low density. Fig. 3 shows the histogram of population density for districts of Zagreb, from the population data we described in section 2.1.2.

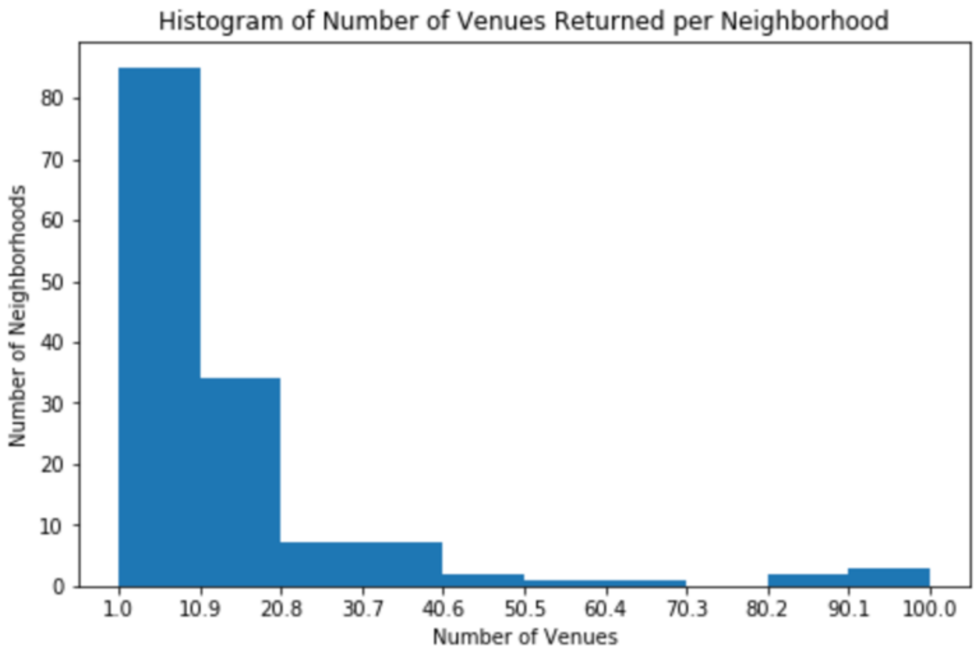

Fig. 2. Histogram of number of venues returned for a neighborhood by Foursquare API.

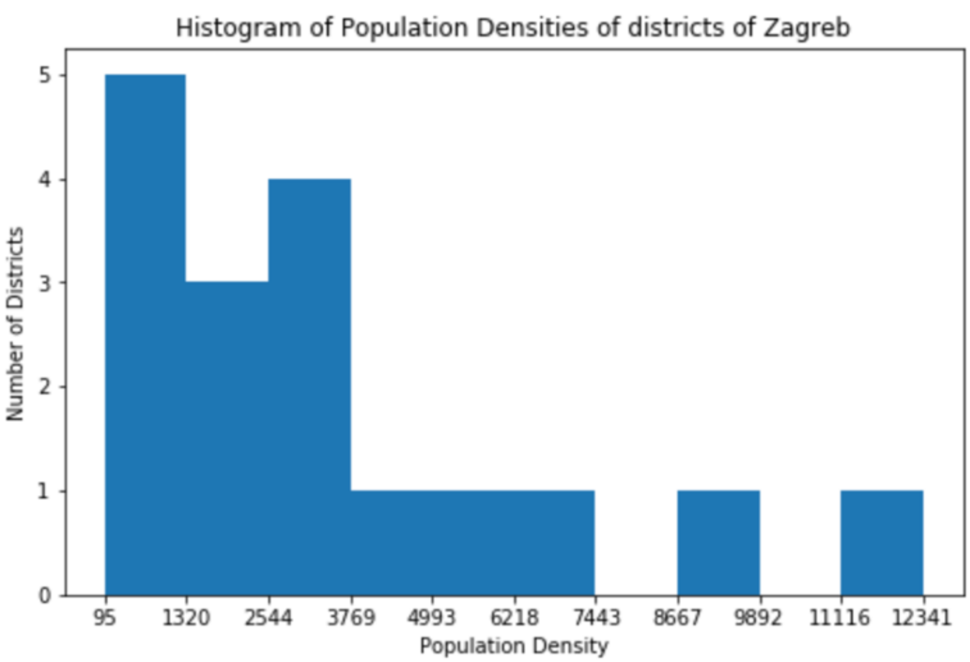

Fig. 3. Histogram of population densities of districts of Zagreb.

Comparing the two histograms, we can see that there is definite correlation between them.

We assumed that the neighborhoods for which we got low number of venues are residential districts with low population (and venue) density. We dropped the neighborhoods for which we got less than 10 venues from our analysis.

This left us with 60 neighborhoods and 1109 venues of 150 unique categories.

3.2. Clustering the Neighborhoods

We used k-means clustering to segment the neighborhoods according to the frequencies of occurrence of each category in the neighborhood. There was 150 unique categories in the venues dataset, so we had 150 features for the clustering algorithm.

To select the optimal number of clusters we used the ‘elbow method’. We ran k-means clustering for number of clusters in the range from 1 to 10, computing the mean distance of cluster datapoints to cluster centroids. Fig. 4 shows the resulting graph. The elbow is at k=6.

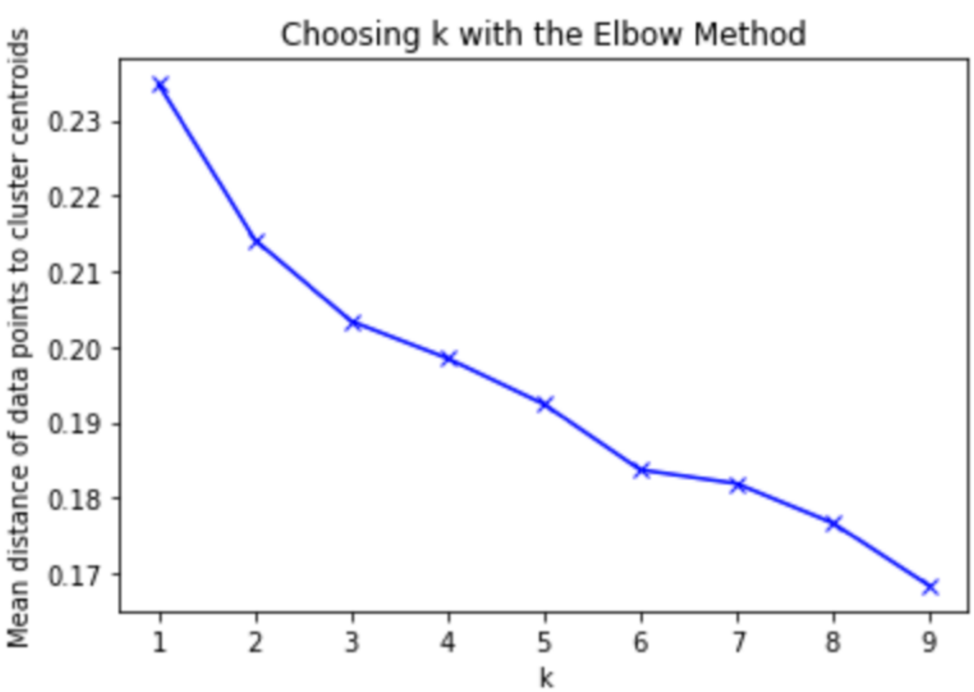

Fig. 4 Choosing number of clusters k with the elbow method.

We ran k-means clustering with k=6. Fig. 5 shows the resulting clusters on the map. **We are interested in the neighborhoods belonging to the same cluster as central neighborhoods**. They are shown in orange. We will refer to this cluster as **'orange cluster' or 'central cluster'**.

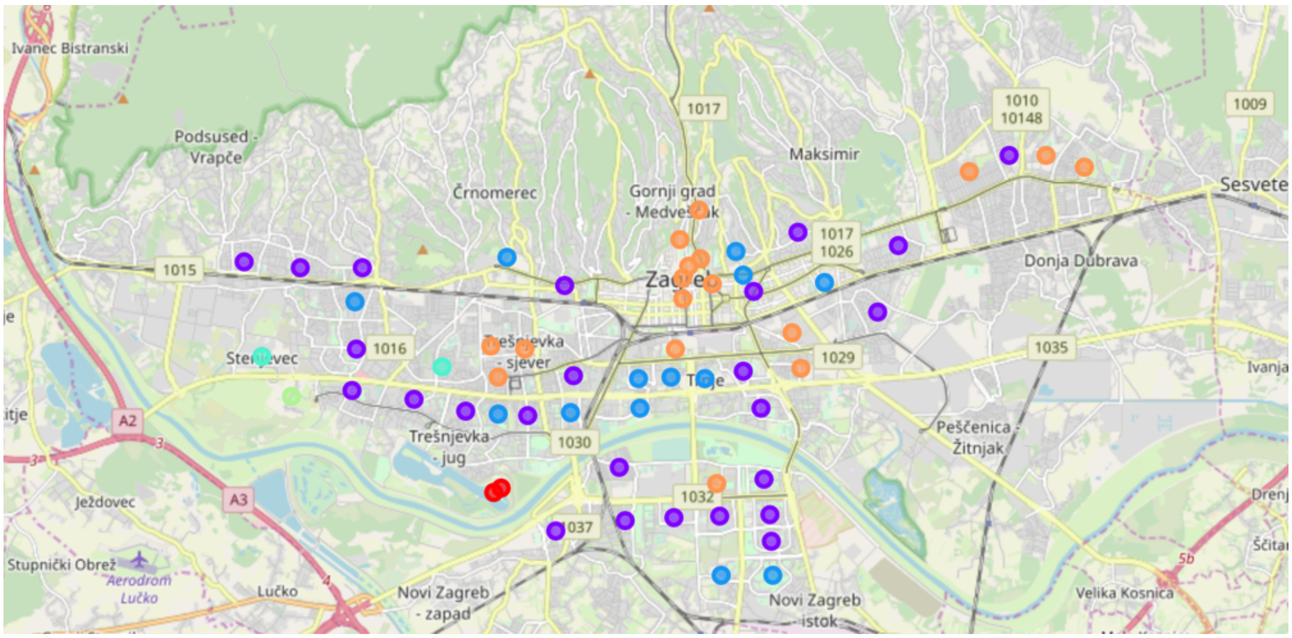

Fig. 5 Neighborhood clusters. Neighborhoods shown in orange belong to 'Central Cluster'. These are neighborhoods we are interested in.

3.3. Selecting Locations

We used the category dictionaries that we described in section 2.1.4 to extract sets of Arts venues, Nightlife Spot venues, and Music venues from the venues dataset. Fig. 6 shows the heatmap of Nightlife Spot venues, superimposed with the neighborhood clusters.

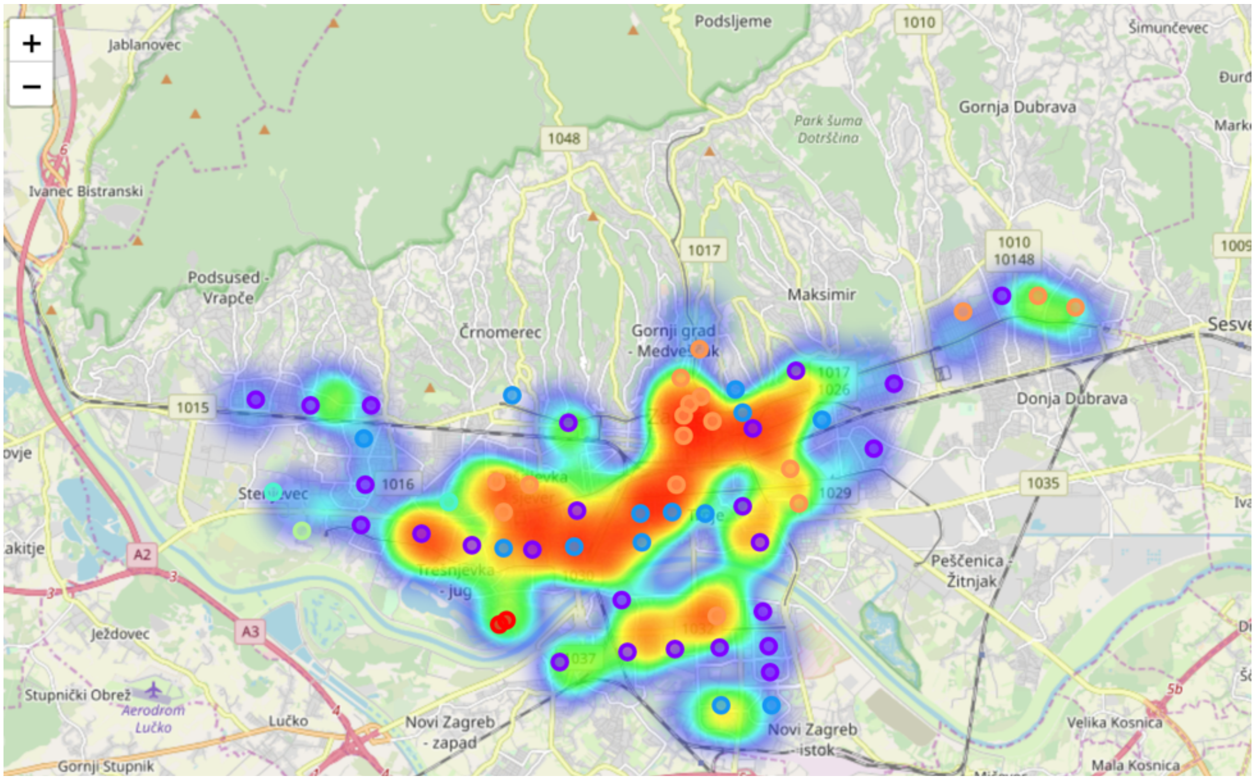

Fig.6. Heatmap of nightlife spots superimposed with neighborhood clusters.

We see that the majority of the neighborhoods of 'orange cluster' are located in hot nightlife areas. Exceptions are far north-east (3 orange neighborhoods there), and several neighborhoods south of the center. Since the requirement for the location was low-to-moderate density of nightlife spots, we need to move out of central area.

Fig. 7 shows the heatmap of arts venues. This map is much less dense. The majority of arts venues are concentrated in the central districts. Outside that central part, we have several 'orange' neighborhoods in low-to-moderate arts density areas.

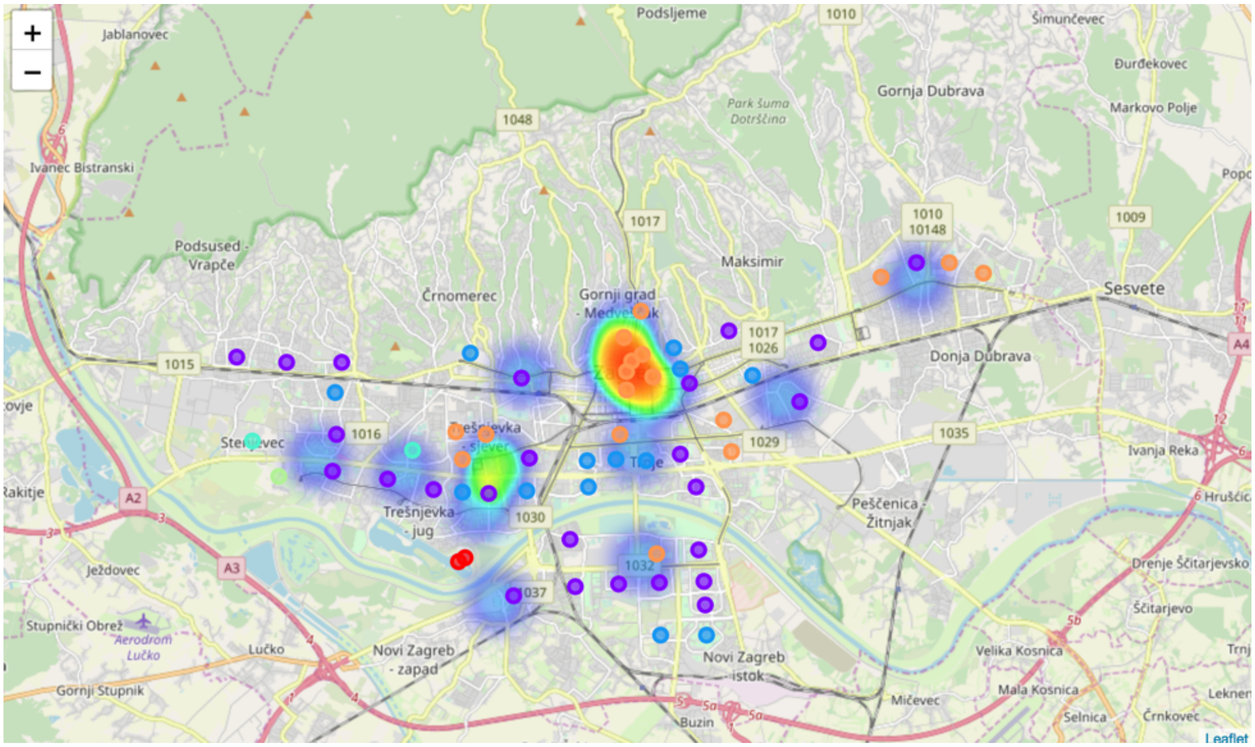

Fig. 7. Heatmap of arts venues, superimposed with neighborhood clusters.

According to the stakeholder's requirements, we want locations in the 'orange' neighborhoods, in high arts venues density area, and in low-to-moderate nightlife density area, with no competing music venues within 500m.

Fig. 8 zooms in on central districts. It shows music venues (by markers with music icons) and arts venues (green circles) combined with neighborhood clusters and nightlife heatmap. White circles of radius 500m surrounding music venues show the areas to stay clear of.

We selected the locations manually, using the neighborhood centers (orange) and arts venues (green) coordinates as reference points.

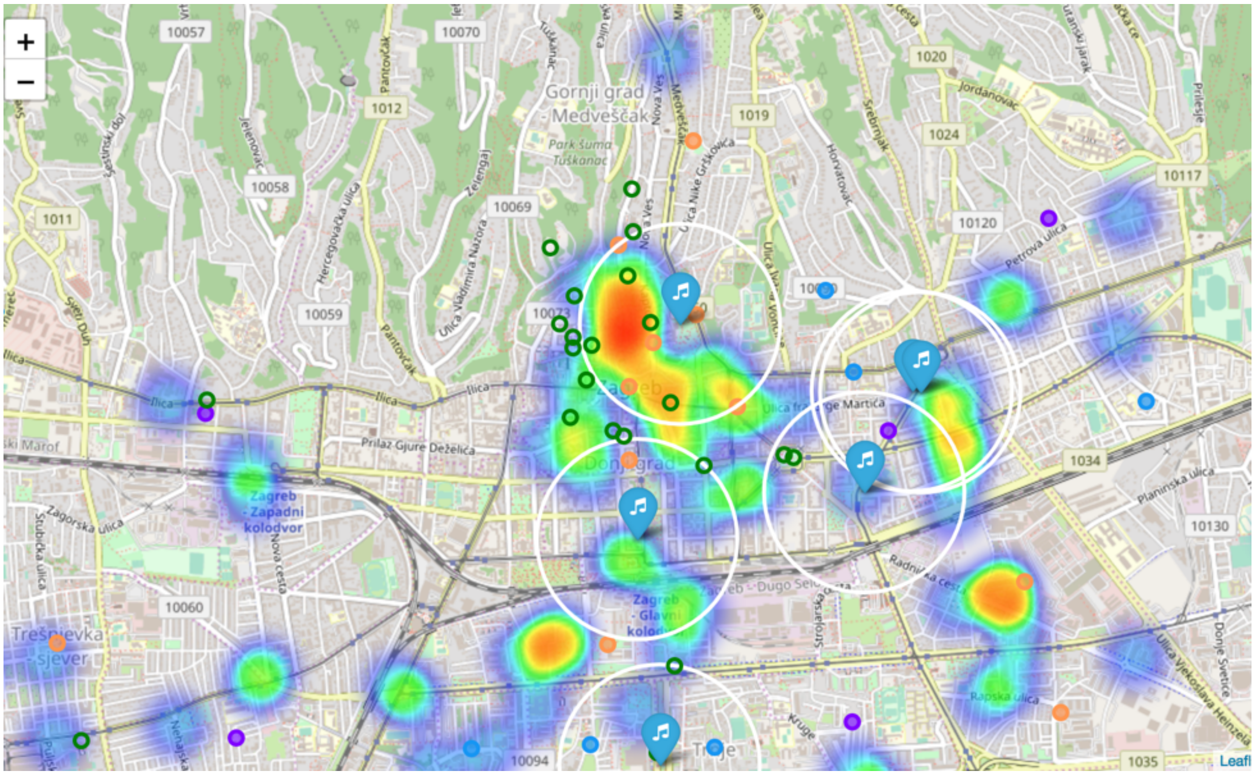

Fig. 8. Competing music venues (markers with music icons) and arts venues location (green circles) combined with neighborhood clusters and nightlife heatmap. White circles of radius 500m surrounding music venues show the areas to stay clear of.

The following locations were selected:

1. Midpoint between Hrvatsko dizajnersko drustvo, Art Gallery, Gornji grad, and Gornji grad neighborhood center
2. Around Kino Europa, Indie Movie Theater, Donji grad
3. Around Hrvatski Povijesni Muzej, History Museum, Tkalčićeva Street
4. Midpoint between Gliptoteka, Art Gallery, Medveščak and Tkalčićeva Street neighborhood center
5. Midpoint between Koncertna dvorana Vatroslava Lisinskog, Concert Hall, Martinovka, and Martinovka neighborhood center

Fig. 9 shows the selected locations.

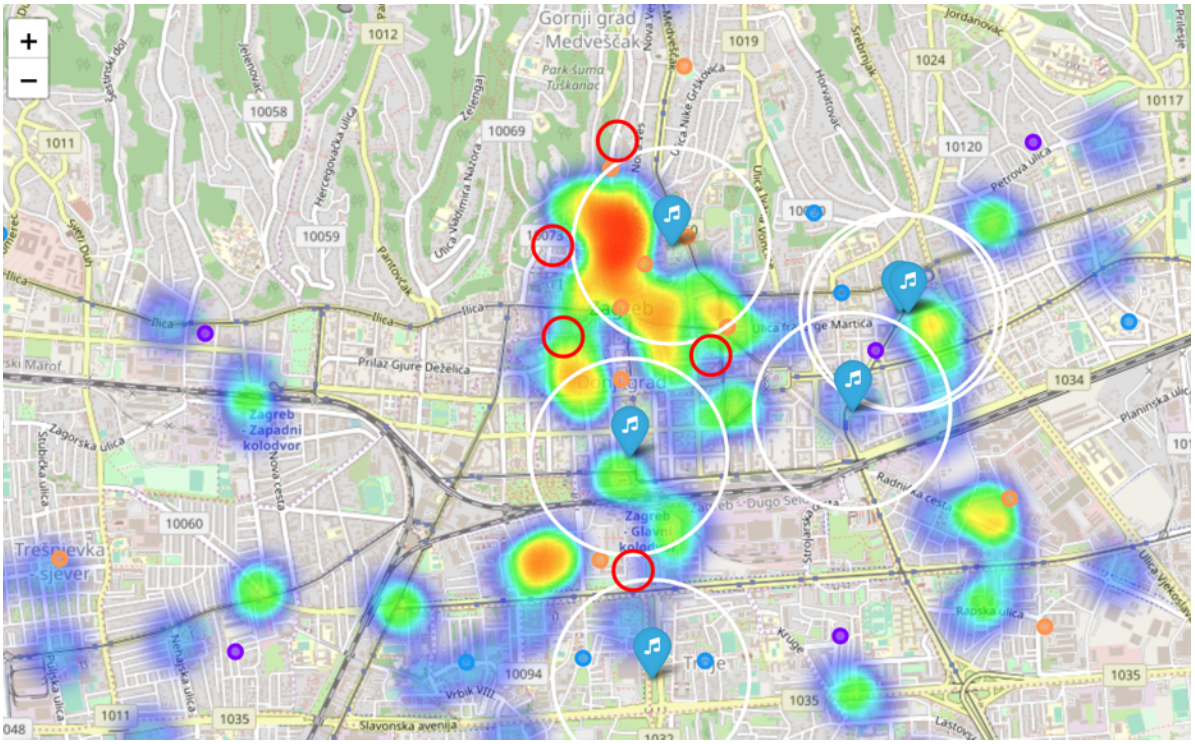

Fig. 9. Selected locations (red circles) for central districts.

There was one more 'orange' neighborhood, further south from center, across the river, that we wanted to consider - Središće. Although it is a bit far from center, it is located in a nice area.There is one arts venue nearby - Muzej Suvremene Umjetnosti of Zagreb (Museum of Modern Arts). The neighborhood is located in a moderate nightlife area. Fig. 10 shows the suggested location on the map.

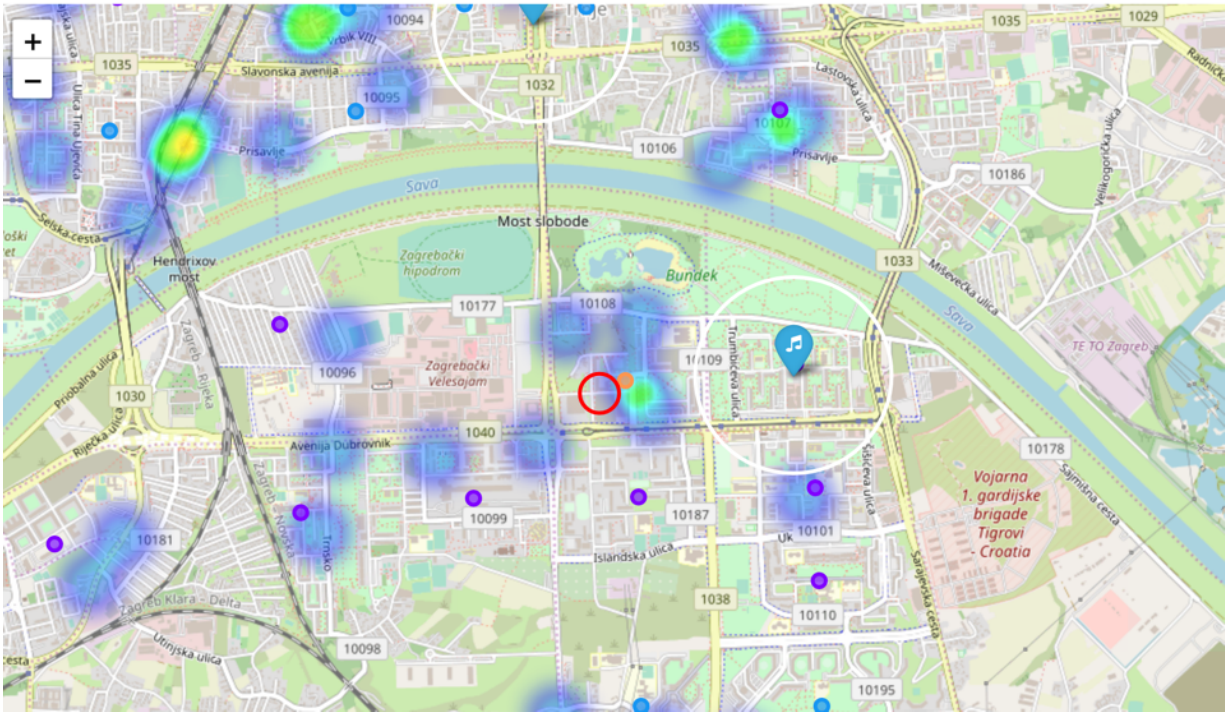

Fig. 10. Suggested location (red circle) in Središće neighborhood.

4. Results

We present the summary of the suggested locations. For each location, we give its distance from the center, the number of arts venues within 250m, the number of nightlife venues within 250m, and the number of music venues (competitors) within 500m.

	Location	Latitude	Longitude	Distance from Center,km	Num Arts Venues within 250m	Num Night Venues within 250m	Num Competing Music Venues within 500m
1	Donji grad	45.811830	15.973250	0.330929	3	4	0
2	Tkalčićeva	45.816000	15.972560	0.469197	5	3	0
0	Gornji grad	45.810987	15.982906	0.516291	1	4	0
3	Medveščak	45.820780	15.976793	0.845270	2	0	0
4	Martinovka	45.801159	15.977822	1.337139	1	1	0
5	Središće	45.779329	15.983404	3.794393	1	1	0

Fig. 11 shows all suggested locations on the map.

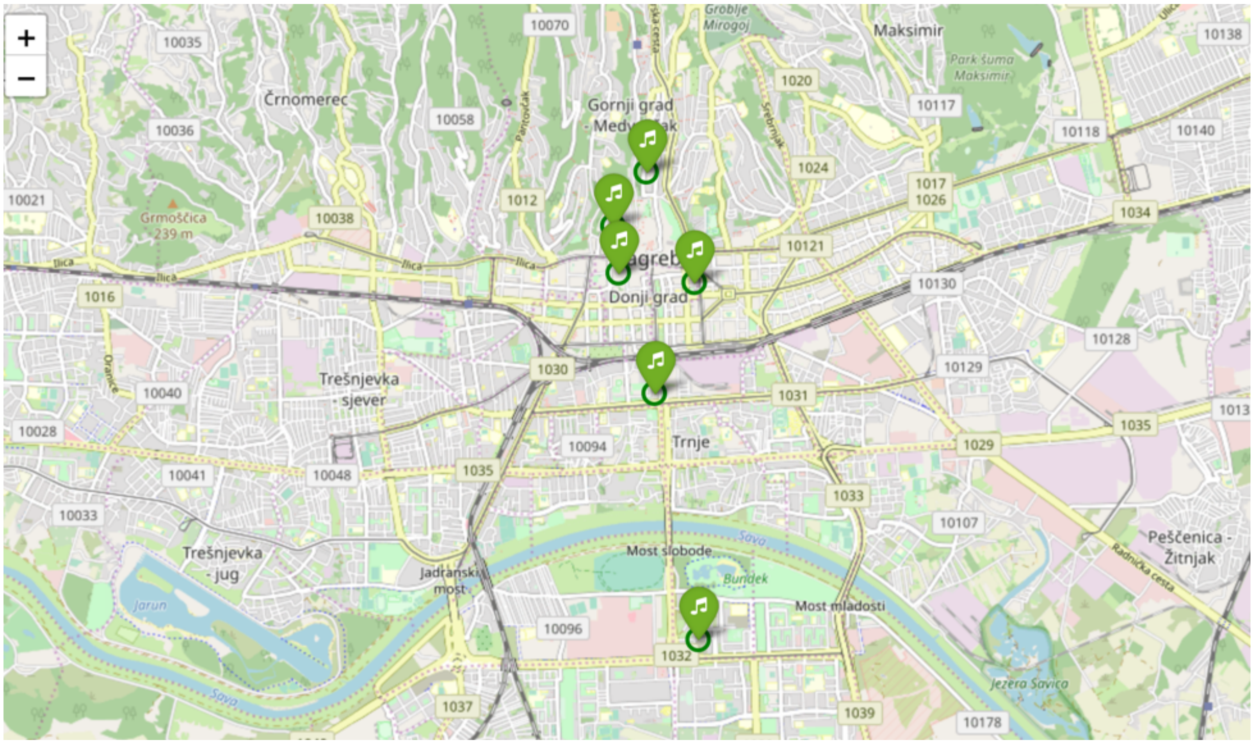

Fig. 11. Suggested locations for jazz cafe in Zagreb.

5. Discussion

Analyzing the venue data returned by Foursquare API, we saw that for more than half of the neighborhoods we got less than 10 venues per neighborhood. As we mentioned in the Data Analysis section, this is on par with the population density distribution of Zagreb.

We dropped the neighborhoods with less than 10 venues from the analysis, but still venue frequencies were unevenly distributed, with the prevailing number of neighborhoods with 20 venues or less. This may have affected the outcome of the clustering algorithm.

We saw that 'arts' venues are mostly concentrated in the central parts of the city. This forced us to look for the desired locations in the central neighborhoods, where both, the music venue density (our first-tier competitors) and nightlife spots density (our second-tier competitors) were high, so we had to select locations carefully. Also, this prevented us from considering more remote neighborhoods from the same cluster. Still, we found two non-central locations, with one being almost 4 km from the city center.

There are other factors that the stakeholder would need to consider - location availability, commercial real estate prices in the area, and others. The locations we suggest fall into wide range of distancies from the city center - the closest being 300m from the center, and the most remote almost 4km from the center. This wide range gives the stakeholder more flexibility, since real estate prices may be significantly higher in the central neighborhoods than in more remote areas.

6. Conclusion

This concludes our exploration of Zagreb neighborhoods and selection of the best location for a jazz cafe. We ran k-means clustering to segment the neighborhoods according to the frequencies of occurrences of venue categories in each neighborhood, and determined which neighborhoods were similar to Zagreb center. We determined the densities of arts and nightlife venues, and competing music venues. We then suggested several locations for jazz cafe that satisfy the stakeholder's requirements: they are located in the neighborhoods that are similar to the center, don't have competing music venues in the area, located near arts venues, and have low to moderate nightlife spots density.

We have provided the summary of the locations, including distance from center and the number of music, arts and nightlife venues in the vicinity. The stakeholder can choose a location and explore it further.