Choosing Location for a Jazz Cafe in Zagreb

Natalia Odintsova

May, 20019

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 the 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 has the same vibe as the city center, but not necessarily close to it. It would have a developed nightlife infrastructure, but preferably no competing musical venues, such as jazz cafes or clubs.

Data

1. Data Sources and Acquisition

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

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'},
```

From each value, 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	4bf58dd8d48968d147941735
1	Dugave	45.765037	15.997999	Play Off	4b7a4cb4f964a520e12821e3	45.766285	16.001196	Café	4bf58dd8d48968d16d941735
2	Dugave	45.765037	15.997999	Goya Bar	4f1b2a23e4b0288a02d11b0f	45.763122	15.992952	Bar	4bf58dd8d48968d116941735
3	Dugave	45.765037	15.997999	Caffe Lupo	4c76752cb474a1od61f8bbbf	45.762524	16.003188	Café	4bf58dd8d48968d16d941735
4	Dugave	45.765037	15.997999	Konzum	4d59795e56f2b60ce898792f	45.769046	15.999688	Grocery Store	4bf58dd8d48968d118951735

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 exclude 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.

3. How the Data Will Be Used to Solve the Problem

We will use k-means clustering to form clusters of neighborhoods based on the frequencies of the categories of their venues. We will then zoom in on the neighborhoods that fall into the same cluster as the city center. For each such neighborhood, we will see how many venues fall into "Arts/Entertainment" category (using category ids) and who many of those fall into "Music Venues" category. Per Foursquare's Venues Category Hierarchy, "Music Venues" include Jazz Club, Piano Bar, and Rock Club subcategories - these are the venues that we consider our competitors. We then select neighborhoods with the lowest ratio of "Music Venues" venues to "Arts/Entertainment" venues.