# Moving to Prague?

**Alina Sergeyeva**

**October 29, 2020**


## 1. Introduction

One of the most welcoming destinations for expats in the center of Europe is Prague. And not only because of its world famous beer, but also for it's vivid social life. If one decides to move to Prague for these reasons, they would need to analyze which district they should choose for living, that would be the most convenient for their needs. And with this comes a difficult decision to make. Therefore, an analysis performed with Python would be of great help.

The analysis presented in this article is based on the Foursquare location data (venues in Prague). And its main aim is to show which districts in Prague have the most active live in terms of availability of cafes, restaurants, night clubs and so on.

## 2. Data Sources

To perform this analysis, the following data is required:

1. List Prague's districts including the coordinates (was manually acquired based on GPS)

2. Venues of districts (Foursquare location data)

## 3. Analysis

Once we clean up and prepare the data, we will use *k-means* clustering method to partition our district into clusters to help us identify the most similar districts. The amount of clusters will be chosen based on silhouette score that provides an optimal number of clusters.

**Step A. Upload the Prague's districts** including the coordinates (.csv file) and remove any duplicated items, as we need only main districts data

| | postalCode | District | Latitude | Longitude |
|---|---|---|---|---|
| 0 | 100 00 | Prague 10 | 50.06763 | 14.46016 |
| 1 | 110 00 | Prague 1 | 50.08728 | 14.41742 |
| 2 | 120 00 | Prague 2 | 50.07395 | 14.43957 |
| 3 | 130 00 | Prague 3 | 50.08261 | 14.45538 |
| 4 | 140 00 | Prague 4 | 50.04233 | 14.44812 |
| 5 | 150 00 | Prague 5 | 50.07167 | 14.40099 |
| 6 | 160 00 | Prague 6 | 50.10106 | 14.39981 |
| 7 | 170 00 | Prague 7 | 50.10801 | 14.43797 |
| 8 | 180 00 | Prague 8 | 50.12692 | 14.45672 |
| 9 | 190 00 | Prague 9 | 50.11059 | 14.50034 |

We can see that there are 10 major districts in Prague.

## Step B. Exploring districts in Prague

We retrieve data for all the districts in Prague using Foursquare and combine with the districts by Postal code:

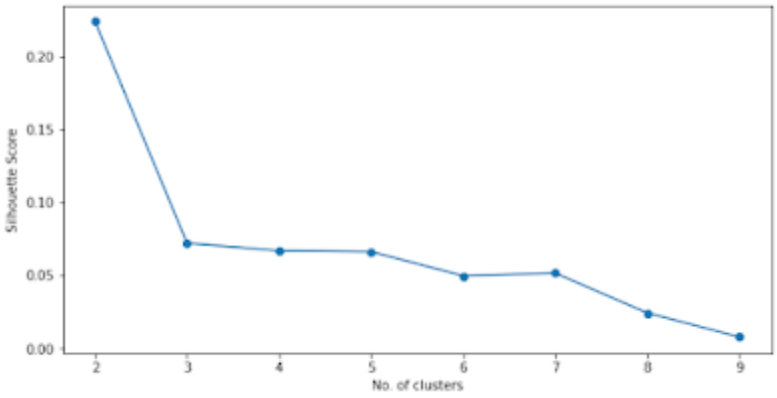
| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Prague 10 | 50.06763 | 14.46016 | Barriqáda | 50.068373 | 14.458013 | Wine Bar |
| 1 | Prague 10 | 50.06763 | 14.46016 | Tvoje máma | 50.069106 | 14.458015 | Café |
| 2 | Prague 10 | 50.06763 | 14.46016 | Milá tchýně | 50.067318 | 14.457746 | Cocktail Bar |
| 3 | Prague 10 | 50.06763 | 14.46016 | Pivní zastávka | 50.070652 | 14.459943 | Beer Bar |
| 4 | Prague 10 | 50.06763 | 14.46016 | Antonínovo pekařství | 50.068972 | 14.454778 | Bakery |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 563 | Prague 9 | 50.11059 | 14.50034 | PC Plus CZ | 50.113029 | 14.501279 | Electronics Store |
| 564 | Prague 9 | 50.11059 | 14.50034 | Restaurace Pragovka | 50.108589 | 14.496557 | Sports Bar |
| 565 | Prague 9 | 50.11059 | 14.50034 | Skateplaza Na Jetelce | 50.113008 | 14.496351 | Plaza |
| 566 | Prague 9 | 50.11059 | 14.50034 | Lanové centrum Proud | 50.111068 | 14.494197 | Playground |
| 567 | Prague 9 | 50.11059 | 14.50034 | Jetelka | 50.114441 | 14.497073 | Park |

At this point we can see that 568 venues were returned for the districts. And after additional manipulation we reorganize the data into the following view:

| Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|
| Prague 1 | Café | Italian Restaurant | Cocktail Bar | Pub | Plaza | Restaurant | Hotel | Vegetarian / Vegan Restaurant | Art Gallery | Theater |
| Prague 10 | Wine Bar | Beer Garden | Vegetarian / Vegan Restaurant | Plaza | Bakery | Café | Chinese Restaurant | Pub | Soccer Stadium | Leather Goods Store |
| Prague 2 | Café | Wine Bar | Bar | Pub | Yoga Studio | Beer Bar | Bistro | Burger Joint | Cocktail Bar | Wine Shop |
| Prague 7 | Gym | Breakfast | Tech Startup | Restaurant | Cafe | Theater | Spa | Laser Tag | Music Venue | Port |
| Prague 8 | Restaurant | Gastropub | Café | Wine Bar | Bar | Sporting Goods Shop | Park | Chinese Restaurant | Doner Restaurant | Soccer Field |
| Prague 9 | Restaurant | Gastropub | Czech Restaurant | Pharmacy | Clothing Store | Hotel | Coffee Shop | Park | Shoe Store | Electronics Store |

## Step C. Districts' Analysis

We use one-hot encoding technique to converts the categorical values into form that could be provided to machine learning algorithm.
And for the *k-means* clustering we use the silhouette score to identify an optimal amount of clusters:

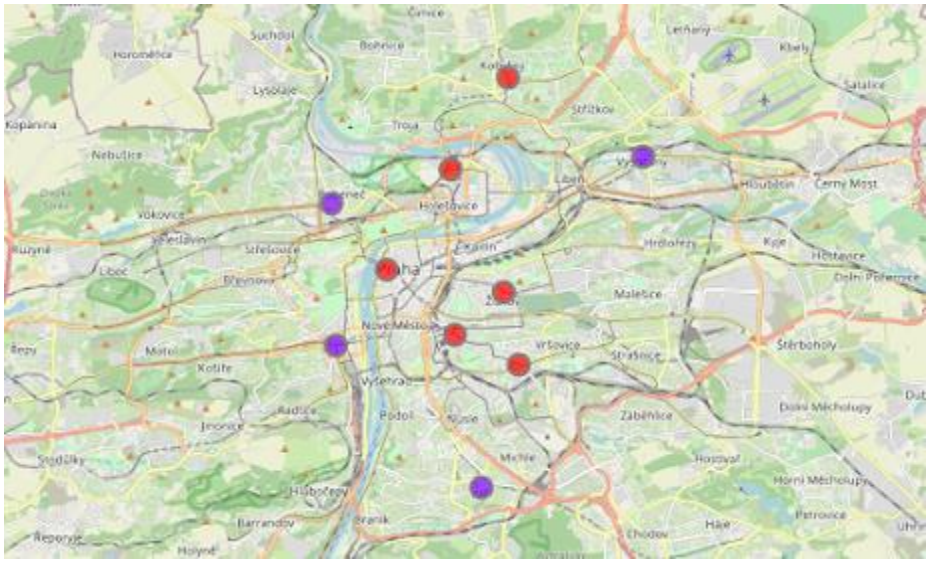



The plot above suggest that the optimal amount of clusters is 2, as it has the biggest score.

The clusters are assigned to the districts by using *k-means* algorithm:

| District | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Prague 10 | 50.06763 | 14.46016 | 0 | Wine Bar | Beer Garden | Vegetarian / Vegan Restaurant | Plaza | Bakery | Café | Chinese Restaurant | Pub | Soccer Stadium |
| Prague 1 | 50.08728 | 14.41742 | 0 | Café | Italian Restaurant | Cocktail Bar | Pub | Plaza | Restaurant | Hotel | Vegetarian / Vegan Restaurant | Art Gallery |
| Prague 2 | 50.07395 | 14.43957 | 0 | Café | Wine Bar | Bar | Pub | Yoga Studio | Beer Bar | Bistro | Burger Joint | Cocktail Bar |
| Prague 3 | 50.08261 | 14.45538 | 0 | Restaurant | Bistro | Cocktail Bar | Café | Asian Restaurant | Czech Restaurant | Pub | Hotel | Gym / Fitness Center |
| Prague 4 | 50.04233 | 14.44812 | 1 | Electronics Store | Clothing Store | Hotel | Café | Drugstore | Bakery | Coffee Shop | Mediterranean Restaurant | Fast Food Restaurant |
| Prague 5 | 50.07167 | 14.40099 | 1 | Hotel | Cosmetics Shop | Pub | Café | Coffee Shop | Clothing Store | Bistro | Park | Vegetarian / Vegan Restaurant |
| Prague 6 | 50.10106 | 14.39981 | 1 | Café | Italian Restaurant | Pub | Bakery | Vietnamese Restaurant | Coffee Shop | Tea Room | Pizza Place | Health Food Store |
| Prague 7 | 50.10801 | 14.43797 | 2 | Climbing Gym | Bed & Breakfast | Tech Startup | Restaurant | Café | Movie Theater | Spa | Laser Tag | Music Venue |
| Prague 8 | 50.12092 | 14.45672 | 0 | Restaurant | Gastropub | Café | Wine Bar | Bar | Sporting Goods Shop | Park | Chinese Restaurant | Doner Restaurant |
| Prague 9 | 50.11059 | 14.50934 | 1 | Restaurant | Gastropub | Czech Restaurant | Pharmacy | Clothing Store | Hotel | Coffee Shop | Park | Shoe Store |

**Step D. Visualize the result on the map**




## 4. Result

The 10 Prague's districts were split into two main clusters (red and purple).

The red cluster represents districts with the most venues in their vicinity. This also can be explained that the most of activities usually happen in the center of the city.

And the purple cluster represents districts that have less venues and are mostly residential areas.

## 5. Conclusion

The districts that are marked red can be considered as potential locations  for those who about to move to Prague and is interested to live in an active district with a lot of social events.