

November 17, 2020



# House Sales Prices and Venues Data Analysis of Prague with Python

Published on November 17, 2020



Ufuk Taner Ceyhanli | [Follow](#)  
Engineer



## Introduction

Prague is one of the most popular city in Europe with many outstanding views and historic, exciting structures. My journey of Prague has started with the hiring of my wife by a global company there. In this study, I give a brief insight about the population, house prices and types by comparing municipal parte of the city for anyone to buy a property in Prague or interested in real estate industry. I'll go over these steps;

- Observe the municipal parts of Prague
- Population distribution by municipal parts

- House prices by municipal parts
- Number of different types of venues inside 1km radius of the center of municipal parts
- Cluster municipal parts of Prague

## **Municipal Parts of Prague**

If you visited Prague, probably you might have been to Prague 1, 2 and 3, however, the city has fifty-seven municipal parts according to the official resources, while first twenty-second region is called by a number, others by names.

Here is a map of Prague regions which I created using Folium library of Python.

## **Population of Prague by Municipal Parts**

I downloaded the population data for 57 Municipal Parts of Prague from the web page below and created a pandas data frame. Resource:

<https://www.citypopulation.de/en/czechrep/praguecity/>

Let's visualize the population of Prague, with the help of choropleth library.

***There are more people living in  
Praha 4 and 10 than other municipal  
regions.***

## **House Sale Prices**

**Sreality.com** is the most used webpage for advertisement of houses in Czech Republic. I scraped all available house advertisements in the website for Prague and refined the data. In the data frame below, there are 4390 advertising with its features.

By using latitude and longitude and Prague municipal part geojson file, I received region info for each object and drop title column. First 5 items of the data frame is below.



Let's do some Exploratory Data Analysis excluding the items belonging to the municipal parts having advertisements less than 30.

*Average price for a house in Praha 1 and 6  
are higher than others.*

*The houses in Praha 1 and Praha 6 are  
larger than others.*

Price per square meter is one of most popular metric in real estates industry.

*The price per square meter is high for  
Praha 1 and 2, compared to others.*

Praha 6 is not in the first ranks for the price per square meter, however it was in the first ranks for the average price for a house, it seems that the reason of this, the size of the houses in Praha 6 is larger.

Here I would like to describe type of the houses in Czech Republic. The naming could be unfamiliar to you but basically,

1+1: one room and one separated kitchen 1+kk: one room and the kitchen is inside room 2+1: two rooms and one separated kitchen 2+kk: two rooms and the kitchen is inside one of the room

*2+kk is most prevalent house type in Prague.*

We may wonder where the most expensive houses are located in Prague.

*The most expensive houses are mainly  
located in Praha 1.*

## **Clustering Municipal Parts by Venues**

In this section, I will cluster municipal parts based on the number of venues in different categories around the center of the municipal parts. I defined the latitude and longitude of the center of the municipal parts and obtain the available venues in 1 km radius of the center point by using **Google Places API**.

I created a data frame for the first ten municipal part and its surrounding supermarkest, bus stations, museums, subway stations and pharmacies.

### **K-means Clustering**

After applying k-means clustering from scikit-learn library, the municipal parts are divided to 3 cluster.

It seems that **Praha 1 and 2** have many venues around the center so these two are distinctive than others. **Praha 5** can be separated because it comes forward with bus station availability. **Other regions** have similar features.

## Results and Discussions

- There are more people living in Praha 4 and 10 than others municipal regions.
- Average price for an house in Praha 1 and 6 are higher than others.
- The houses in Praha 1 and Praha 6 are larger than others.
- The price per square meter is high for Praha 1 and 2, compared to others.
- Praha 6 is not in the first ranks for the price per square meter, however it was in the first ranks for the average price for a house, it seems that the reason of this, the size of the houses in Praha 6 is larger.
- 2+kk is most prevelant house type in Prague.
- The most expensive houses are mainly located in Praha 1.



- Based on the venues available, Prague municipal parts can be divided in three clusters.

Do the venues around municipal parts contribute the price of the house?

Yes. We may divide the municipal parts to 3 groups based on price per square meter. Praha 1 and 2 are more expensive than other municipal parts, following Praha 5 and others. This fits the cluster which is created only by the numbers of different types of venues around the municipal parts' center.

## Conclusion

Prague has 57 municipal parts and each of them may have different features. However some of them has more venues around and this contributes the price of the house.

*Ufuk Taner CEYHANLI*

*11/17/2020*

## Published By



Ufuk Taner Ceyhanlı  
Engineer

[Follow](#)

Today I accomplished "IBM Data Science Professional Certificate", 9 courses in total, with my "Housing Sales Prices and Venues Data Analysis of Prague with Python" Capstone Project. Please do not hesitate to share your valuable comments with me for future studies.

0 comments



[Sign in](#) to leave your comment

## More from Ufuk Taner Ceyhanlı [1 article](#)

© 2020

[Accessibility](#)

[Privacy Policy](#)

[Copyright Policy](#)

[Guest Controls](#)

[Language](#)

[About](#)

[User Agreement](#)

[Cookie Policy](#)

[Brand Policy](#)

[Community Guidelines](#)