**Coursera Capstone**

IBM Applied Data Science Capstone

# Cities Divided by Iron Curtain new Opportunities

By: Miroslav Danek

April 2021

**Prague**



vs

**Munich**



**Introduction**

Two cities with formerly very similar history and only 3,5 hours apart from each other, were divided with iron curtain for over 40 years. This long separation has changed the evolution of the isolated city Prague dramatically. But the iron curtain is history for almost 32 years. Prague is economically -measured by GDP one of the richest cities in EU. But are between these two cities still some differences, which could be caused by the 40 years lasting division?

**Business Problem**

The comparison of these two cities using the methods of data science, could reveal some new business opportunities. Probably in is some fields the convergence process of both cities not completed or in Munich are missing some venues which will come from the east in the future, as the relevance of the eastern economies will rise.

**Target audience**

This project useful for city administration or for entrepreneurs who have business in the areas descripted by the used data service four square. The all can benefit from knowing the differences of these two cities, which are both similar populated capitals of similar big states Czech lRepublic and Bavaria.

**Data**

Following data will be used:

* ZIP Lists of both cities found on the internet.
* Coordinates of all neighborhoods automatically assigned by an API.
* Venue Data obtained from the Four Square API

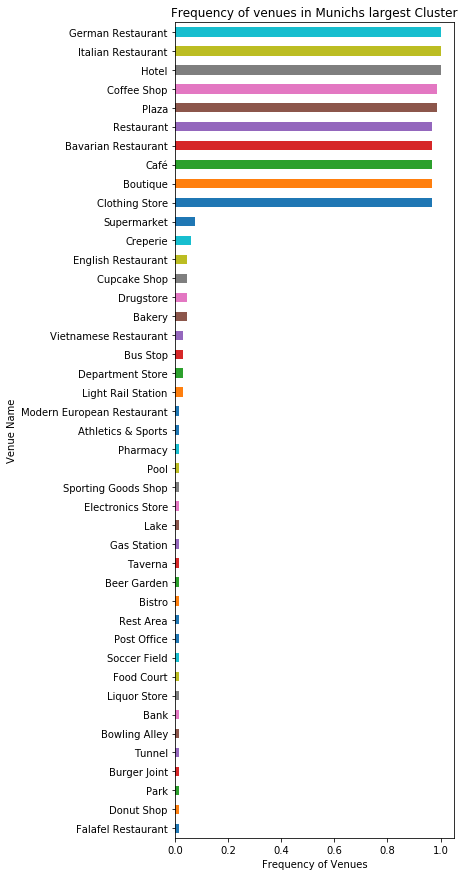
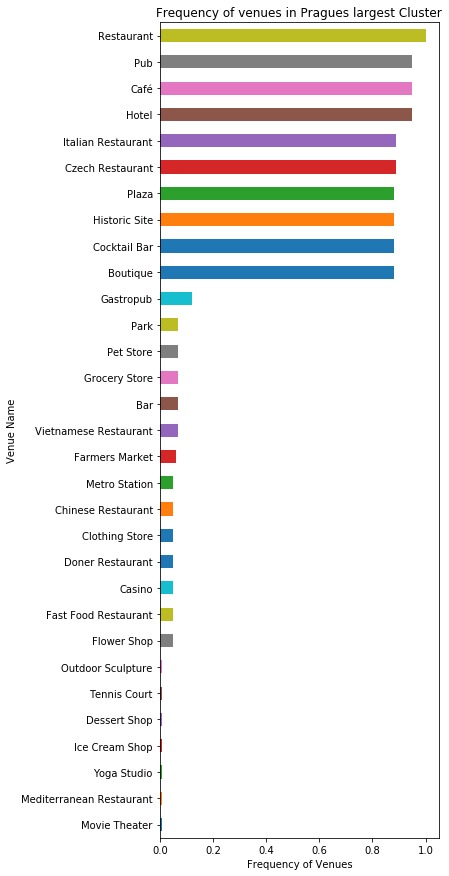
**Methodology**

The methods are the same for both cities. All must be performed twice.

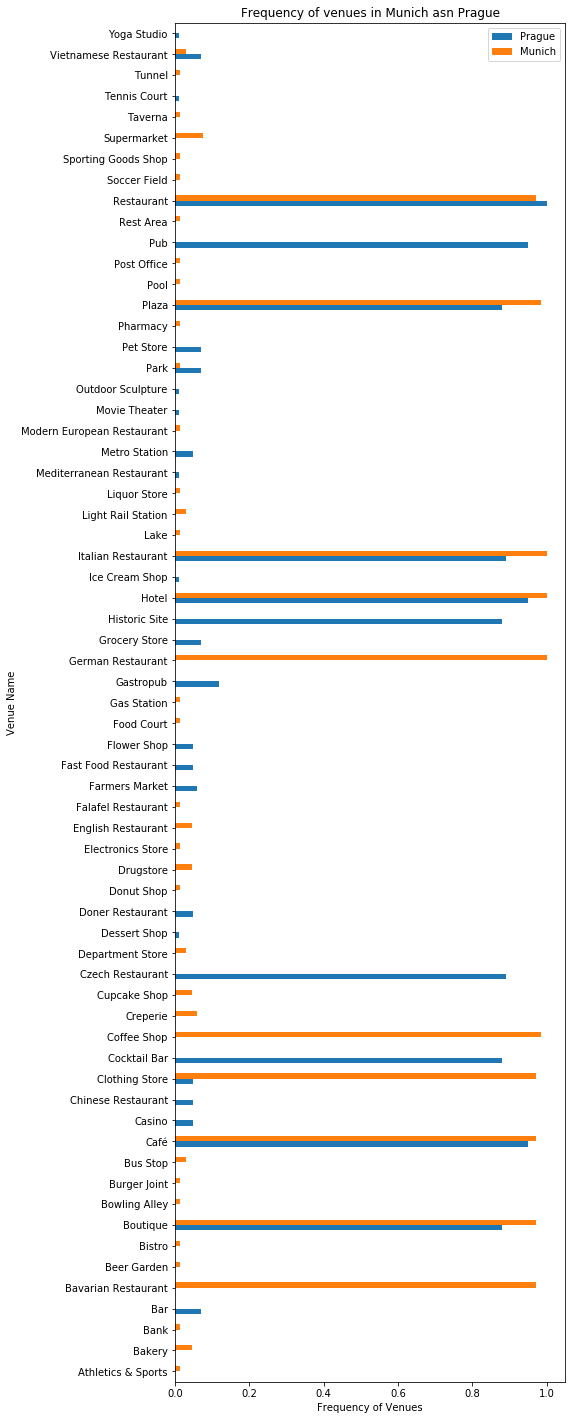
1. A public chart with neighborhoods names and ZIPs will be scraped, cleaned and saved in to the pandas
2. Coordinates will be added to the dataset using geocoder API
3. The neighborhoods will be visualized using the folium maps
4. Venues will be assigned to the coordinates from the Four-square API
5. 5 Clusters of neighborhoods will be created using the k-means method.
6. The largest cluster will be automatically selected and converted in to a one-dimensional dataset. The largest cluster changes
7. A bar chart will be generated from the obtained data.
8. The same method will be repeated for the other city.
9. On the end the prepared datasets with most common venues will be joined in to one dataset
10. The grouped bar charts will be created.

**Results**

* Venues in Prague and Munich

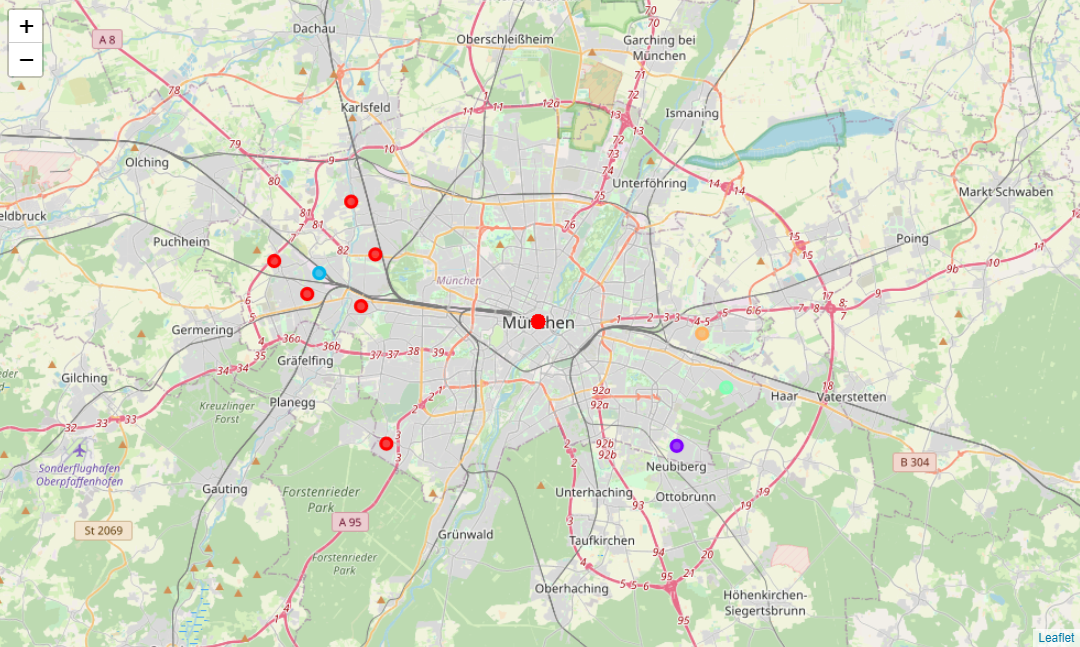
 

* Comparison of venues in both cities

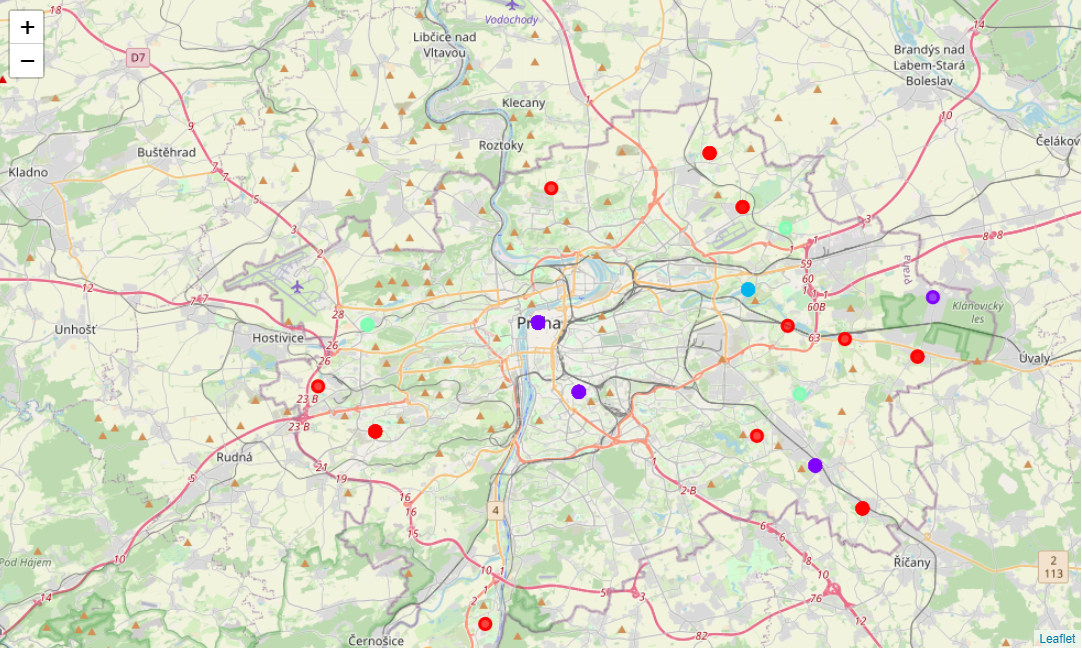


|  |  |  |
| --- | --- | --- |
|  | **Prague** | **Munich** |
| **Athletics & Sports** | NaN | 0.015152 |
| **Bakery** | NaN | 0.045455 |
| **Bank** | NaN | 0.015152 |
| **Bar** | 0.07 | NaN |
| **Bavarian Restaurant** | NaN | 0.969697 |
| **Beer Garden** | NaN | 0.015152 |
| **Bistro** | NaN | 0.015152 |
| **Boutique** | 0.88 | 0.969697 |
| **Bowling Alley** | NaN | 0.015152 |
| **Burger Joint** | NaN | 0.015152 |
| **Bus Stop** | NaN | 0.030303 |
| **Café** | 0.95 | 0.969697 |
| **Casino** | 0.05 | NaN |
| **Chinese Restaurant** | 0.05 | NaN |
| **Clothing Store** | 0.05 | 0.969697 |
| **Cocktail Bar** | 0.88 | NaN |
| **Coffee Shop** | NaN | 0.984848 |
| **Creperie** | NaN | 0.060606 |
| **Cupcake Shop** | NaN | 0.045455 |
| **Czech Restaurant** | 0.89 | NaN |
| **Department Store** | NaN | 0.030303 |
| **Dessert Shop** | 0.01 | NaN |
| **Doner Restaurant** | 0.05 | NaN |
| **Donut Shop** | NaN | 0.015152 |
| **Drugstore** | NaN | 0.045455 |
| **Electronics Store** | NaN | 0.015152 |
| **English Restaurant** | NaN | 0.045455 |
| **Falafel Restaurant** | NaN | 0.015152 |
| **Farmers Market** | 0.06 | NaN |
| **Fast Food Restaurant** | 0.05 | NaN |
| **...** | ... | ... |
| **Grocery Store** | 0.07 | NaN |
| **Historic Site** | 0.88 | NaN |
| **Hotel** | 0.95 | 1.000000 |
| **Ice Cream Shop** | 0.01 | NaN |
| **Italian Restaurant** | 0.89 | 1.000000 |
| **Lake** | NaN | 0.015152 |
| **Light Rail Station** | NaN | 0.030303 |
| **Liquor Store** | NaN | 0.015152 |
| **Mediterranean Restaurant** | 0.01 | NaN |
| **Metro Station** | 0.05 | NaN |
| **Modern European Restaurant** | NaN | 0.015152 |
| **Movie Theater** | 0.01 | NaN |
| **Outdoor Sculpture** | 0.01 | NaN |
| **Park** | 0.07 | 0.015152 |
| **Pet Store** | 0.07 | NaN |
| **Pharmacy** | NaN | 0.015152 |
| **Plaza** | 0.88 | 0.984848 |
| **Pool** | NaN | 0.015152 |
| **Post Office** | NaN | 0.015152 |
| **Pub** | 0.95 | NaN |
| **Rest Area** | NaN | 0.015152 |
| **Restaurant** | 1.00 | 0.969697 |
| **Soccer Field** | NaN | 0.015152 |
| **Sporting Goods Shop** | NaN | 0.015152 |
| **Supermarket** | NaN | 0.075758 |
| **Taverna** | NaN | 0.015152 |
| **Tennis Court** | 0.01 | NaN |
| **Tunnel** | NaN | 0.015152 |
| **Vietnamese Restaurant** | 0.07 | 0.030303 |
| **Yoga Studio** | 0.01 | NaN |

**Clusters Munich**



**Clusters Prague**



**Discussion Section and limitation of the Model**

In this project we rely on the data from the four-square service. This API has found more unique venues in Munich than in Prague. It seems also, that for some in the reality identical venues in Prague and Munich were different names been used. Only 9 from 65 venues are the same in both cities. For example, a Subway is present in both cities, but only in Prague it is called Metro station as in English. In Munich it is called U-Bahn and no such a category was detected by the four-square API. Which can mediate an incorrect assumption, that in Munich is no Subway.

I know both cities very well an in general I can admit, that the result seems to be realistic (with exception for the Transportation). It is also important to remember, that only the 10 most frequent venues in each neighborhood where considered.

**Conclusion**

The project has shown, that a data research is much easier than many people expect.

Munich has a broader variety of venues. Predominately a broader variety of different cuisines restaurants from different countries. This make sense, because Munich is more multicultural. It can also be expected, that in Prague will emerge demand on more luxurious “unnecessary” services, as Prague’s wealth level will converge to Munich’s. These are listed in the results section.

Prague is on the other side stronger in services determined for tourists, but it cannot be expected that this will change in Munich. Prague is also stronger in Farmers Markets which could also be interesting for Munich.