Capstone Project
The Battle of Neighborhoods

Peer-graded Assignment

Lukas Schumi

Final Assignment to finish the
IBM Data Science Professional Certificate

Vienna, Austria
April 2020

Contents

1 Instructions

1.1 Background

I currently work for the company **XXL Sports & Outdoor GMBH**, a Scandinavian sports retailer which is expanding in Austria. The logical starting point was the capital of Austria, Vienna, which inherits almost a quarter of the population of the entire country. It was also a good starting point, for Vienna offers a wide variety for those of us, who want to stay active and enroll in any kind of sportive activity. But there is also a lot of competition on the sporting goods market. And with 5 Stores in 3 major cities and a 6th one just weeks from opening, expansions is always a hot topic. Therefore analyzing locations is key for finding the perfect position for a new store.

1.2 Problem

Austrias national sport is skiing and there are many first-class skiing areas especially in the western regions of the country. In my opinion, when it comes to sport, the skiing market might be the hardest to enter in this country, for there are countless long-established specialized shops. So I think it might be smarter to focus on the eastern regions until the brand of XXL made a name for itself as a reliable retailer. And where better to start a comprehensive branch network than in the capital? So in the following work I focus on the city of Vienna and its immediate surroundings. Searching an optimal location for a new store can be challenging. You have to find the right balance between a place with too much competition and one that's too remote to reach. To narrow the possibilities down, I looked at locations where places are nearby, where sport equipment is needed. The thought behind this is, that if someone is at such a place, say a gym or a beach, and forgot his gear or something breaks during usage, there might be a bigger chance, if a sports good shop is nearby, that said person buys new equipment instead of going home and picking it up or ordering it online. Ideally there should also be some kind of public transport nearby.

1.3 Interest

This project is meant for stakeholders at **XXL Sports & Outdoor GMBH** as aid when searching for the next location to open a new store.

2 Data acquisition

2.1 Data source

For this project I got locations for sport facilities as well as the places of competing stores in Vienna from the **Foursquare API**. The locations of the XXL stores are from the Homepage `https://www.xxlsports.at/store-finder` and for comparisons I used the official city map of Vienna `https://www.wien.gv.at/stadtplan/en/`. From the last one I took the coordinates of the St. Stephen's Cathedral ("Stephansdom") which set as center of the city.

2.2 Feature selection

- I will create two datasets using the **Foursquare API**. One for sporting goods shops in Vienna, containing information about name, location (address as well as geographical coordinates) and associated category. The second dataset will contain the same information about sport facilities.

- The data collected from Foursquare will be saved as **csv files** and then read into **pandas dataframes**.

- To visualize the proximity of the sporting facilities to the shops and compare them to existing XXL stores I will create a **Folium map**.

3 Methodology and Analysis

The goal of this project is to find the perfect location for a new store, where the neighborhood is surrounded by different sport-related venues. Then I will apply **k-Means Clustering**, a simple unsupervised machine learning algorithm, to make segmentations of venues. The new location should cover an area which isn't already covered by another XXL store. There should also be no or hardly any competitors in the same cluster. Finally, when a matching cluster is found, the coordinates of the future store location will be calculated to be the closest point possible to the venues in the cluster. Based on the goal of this project, I will describe below, what is needed to perform this search and which data sources will be used:

- number of sport facilities in Vienna

- number of sport shops in Vienna

- segmentation of sport venues in 4 clusters to match the location of 4 stores

- distance of the final location in relation to the other points in the cluster

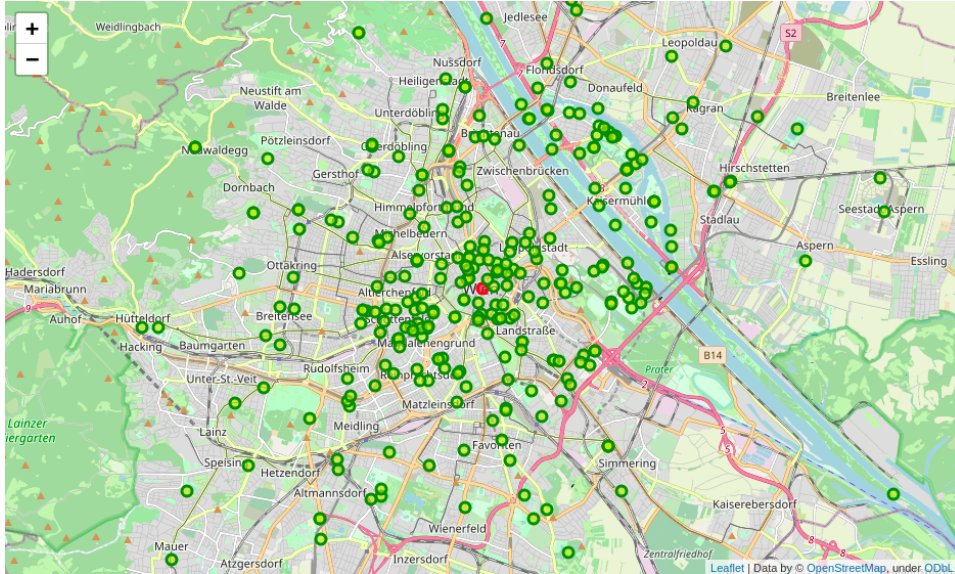- checking for traffic connection and public transport

Figure 1: sport facilities in Vienna

Figure 1 shows a map of all sport facilities in Vienna.

3.1 Clustering sport facilities to check present coverage

Since there are already three XXL Sports & Outdoor Stores in Vienna I used the k-Means machine learning algorithm to divide the venues in three areas (as shown in Fig. 2 and see how good the three stores cover those.

One can see in Fig.3 that the three stores perfectly cover the area of venues created by the k-Means algorithm.

3.2 Clustering sport facilities to check possible future coverage

Since we want to open a fourth store, we now will create 4 clusters with the same algorithm and see if we can find a place for another store.

As one can easily see in Fig.4, the three stores still cover three different areas nicely. A fourth cluster in the north is not yet covered. To find an area that's not too crowded with competitors let's project those on the map.
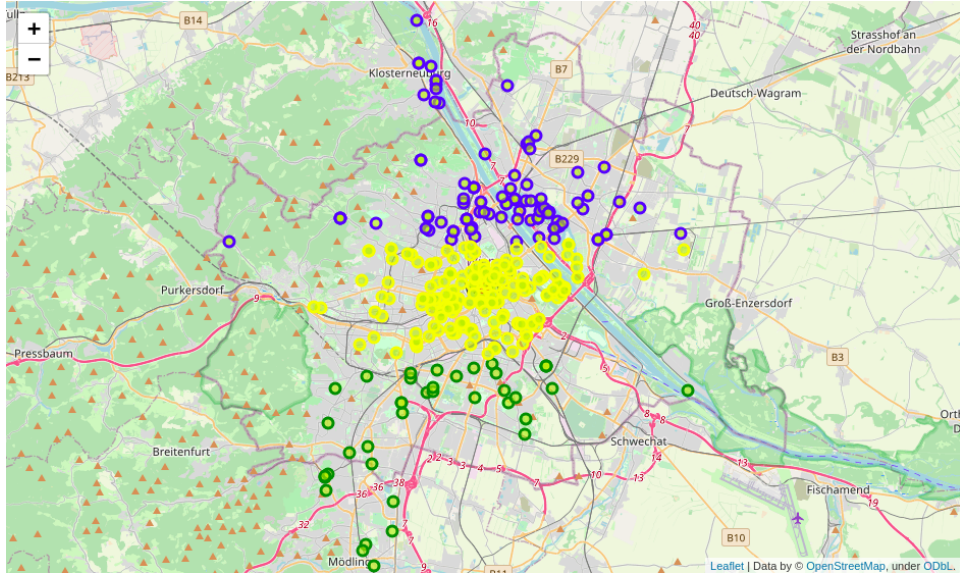
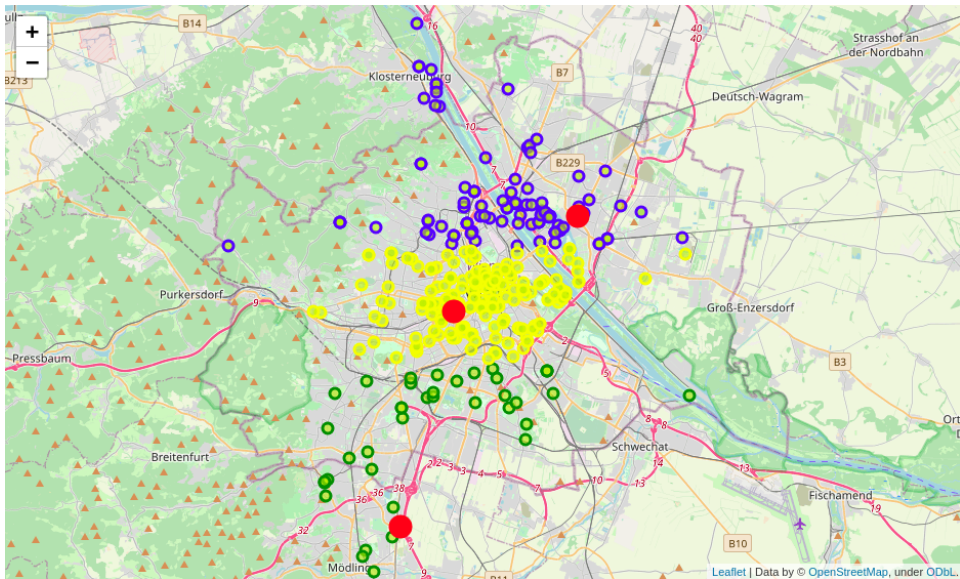Figure 2: k-Means cluster of sport facilities (k = 3)

Figure 3: XXL Sports covering areas created by k-Means

Combining the data to get an overview.

Fig.5 shows that most of the sport facilities as well as the sporting goods shops are in the city center. Another hot spot of sporting venues are along the Danube north-east of the city. Both areas are covered by a XXL Store. What one notices immediately when looking at Fig.5 is, that there are quite some venues in the neighboring city Klosterneuburg, but, as it seems no
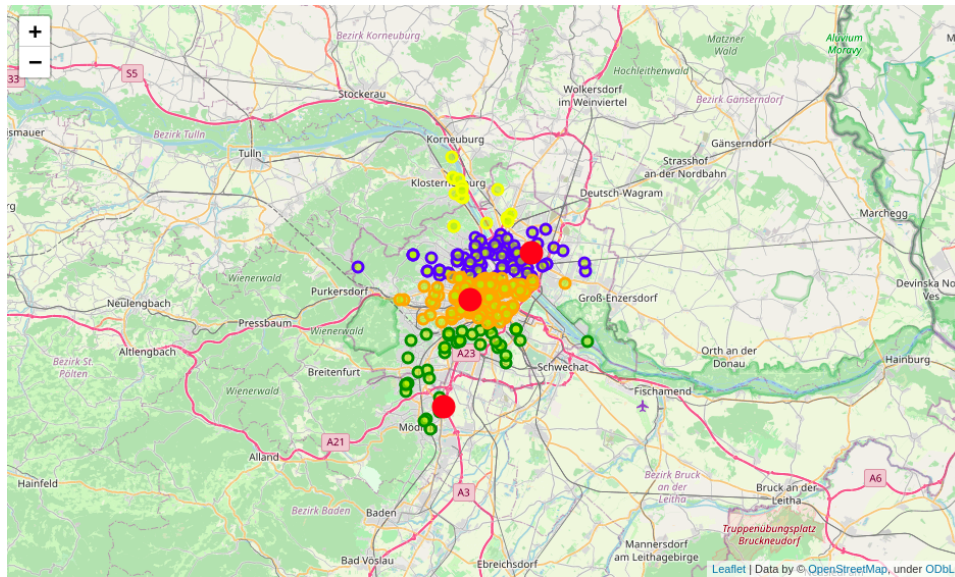
5

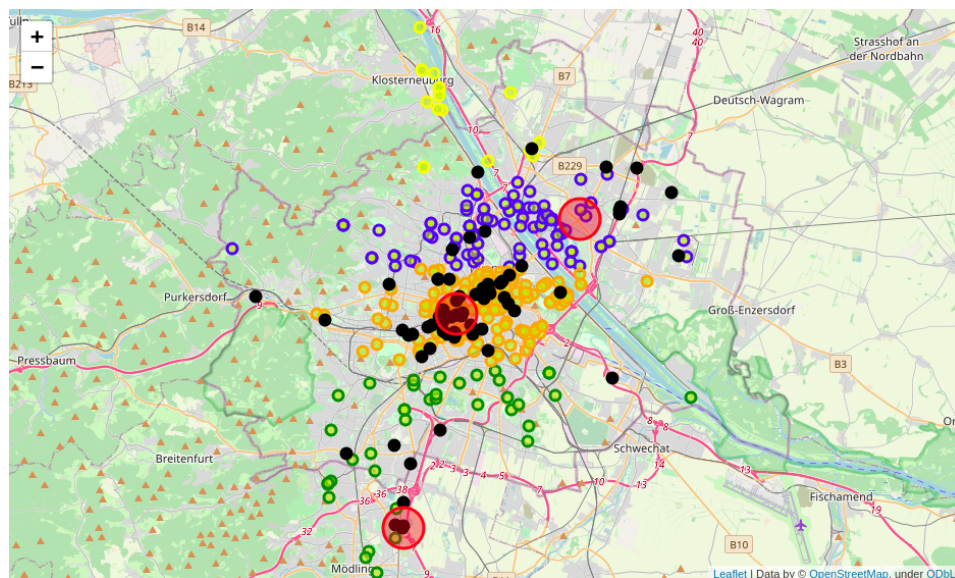Figure 4: k-Means cluster of sport facilities (k = 4)

Figure 5: sport shops superimposed on clusters

sport shops whatsoever. In the next section I will take a closer look at the city of Klosterneuburg.

3.3 Reevaluation: Openeing a store in Klosterneuburg?

Using the **Foursquare API** to examine a potential location in Klosterneuburg

Using a search radius of 2km. A larger radius would on the one hand get too far away from Vienna in the north and on the other hand reach into Vienna in the south.
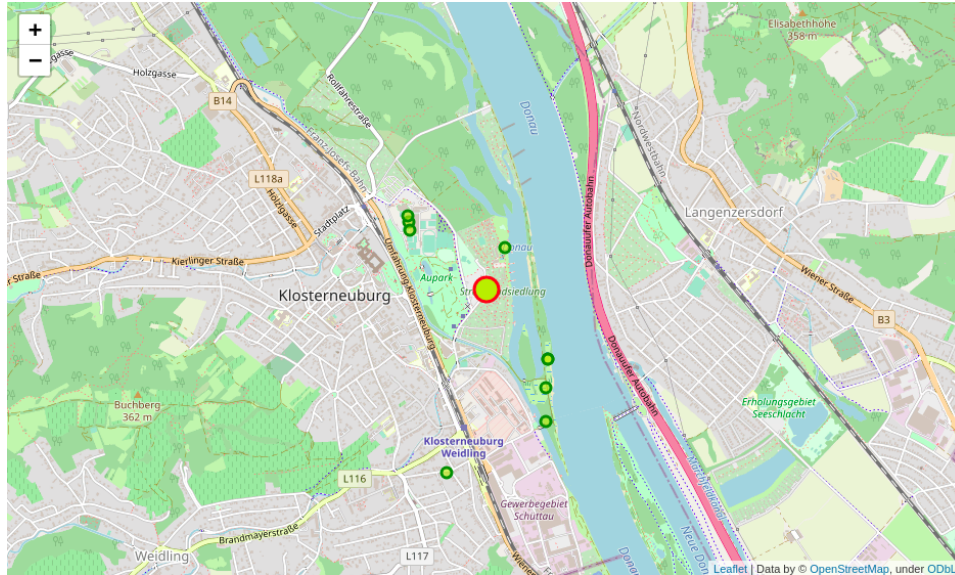

Figure 6: possible new location for a XXL store in Klosterneuburg

4   Results

The so found position satisfies all criteria stated in the first part:
- Sport facilities are nearby

- There is hardly any competition around

- But there are still enough stores which invite to stroll

- The main train station with a park and ride site is just around the corner

5   Conclusions

In this project I set a goal to find a new location for the next **XXL Sports & Outdoor store in Vienna**, considering mainly two prerequisites. The two main conditions for the new location were a close proximity

to any sport-related venue and an area with shopping possibilities but not to much competition.

On **developer.foursquare.com** I found a register of all existing categories of venues. After compiling a catalog of all the categories which are somehow related to sportive activities, I utilized the **Foursquare API** to generate a list of every venue matching at least one entry in the catalog.

In the next step I repeated that process to find all the stores selling sporting goods, once again ordered by category. This categories were selected and recorded, for they might have been important when a final location had been found, to analyze the competition in the immediate surrounding. As it turned out, this step would not have been necessary since there are no shops focus on sports near the found location.

As vantage point for my search I defined the **St. Stephen's Cathedral** ("Stephansdom" in German), one of the most well known landmarks of the city, located in the city center.

Starting from St. Stephen's Cathedral, I drew a radius of 15km and defined that as my search area. I choose that distance for the resulting circle covered the whole city as well as quite a bit of the surrounding areas which is still easy to reach by public transport.

In the first part of the Data Analysis I applied the **k-Means machine learning algorithm** to cluster the previously found sporting venues in 3 areas and checked if the 3 stores already existing in Vienna somehow covered that area. As it turned out every one of those three stores fell exactly into one cluster of venues, which was easily visible on the **Folium map**.

In the next step I used the same algorithm to divide the list of venues into 4 clusters. Once again three of those were still covered by the three existing stores. So for the search of a new location I focused on the fourth cluster.

I superimposed the coordinates of the competitors on the Folium map to see how many other stores were present in the previously mentioned fourth area. As it turned out there were none. Especially not in the neighboring city of **Klosterneuburg** where there seemed to be quite a high density of sporting facilities.

So I now focused my attention on the city of Klosterneuburg and repeated the first steps for this new vantage point. The final location was found by calculating the center of mass of the sporting facilities in an area 2km around the city center. A value of 2km was chosen because a larger radius would on the one hand get too far away from Vienna in the north and on the other hand reach into Vienna in the south. A final step was to check if any form of public transport and traffic connection was nearby, which luckily was the case in form of the main train station and a park and ride parking garage.

The found information is of course by far not enough to make a final decision about expanding. There are countless of other factors to take into

account, like rent prices, available personnel or logistic infrastructure just to name a few. The goal of the above project was not to find a final solution for the problem of expansion, but to give stakeholders a rough idea where they might to look closely for a new branch.