

Public Restrooms in Berlin

Anton Lischewski

2021-01-03

I. Introduction	2
A. Business problem.....	2
1. Analysis of the density of toilets in the two cities.....	2
2. Marking of handicapped accessible restrooms.....	2
3. Comparison of the information provided by the two cities with Foursquare.....	2
B. : Who would be interested in this project.....	2
II. Data	2
A. Description of the data that has been used to solve the problem.....	2
B. Sources of the data.....	3
III. Methodology	3
A. Exploratory Data Analysis	3
B. Inferential statistical analysis	3
C. Machine learnings that were used	3
IV. Results	4
A. Overview	4
B. Grouping and clustering locations	4
C. Foursquare data on public restrooms in Berlin.....	5
D. Foursquare data on public restrooms in Boston, MA.....	6
E. Gold standard or reference	6
V. Discussion.....	7
A. Observations	7
B. Recommendations	7
VI. Conclusion.....	7
VII. Copyrights.....	7

I. Introduction

A. Business problem

The main goal of this project was to create a map showing the public toilets in Berlin, the capital of Germany. A website from Boston, MA in the USA served as a template.

In addition to the main goal, there were other subordinate goals or requirements:

1. Analysis of the density of toilets in the two cities

How many public toilets are there in each city within a 30 km radius?

2. Marking of handicapped accessible restrooms

Can restrooms for the disabled be clearly distinguished from others in their representation on the map?

3. Comparison of the information provided by the two cities with Foursquare

Who provides the “better” data - the respective cities or Foursquare?

B. : Who would be interested in this project

Tourists, families with small children, and especially disabled people are dependent on publicly accessible toilets in tourist centers. Wheelchair users in particular are very interested in finding the nearest handicapped-accessible toilet.

Boston, Massachusetts maintains an excellent service on its website <https://www.boston.gov/departments/311/public-restrooms-city-boston> – this served as a guideline in this project.

II. Data

A. Description of the data that has been used to solve the problem

The main information used in this project is the name of the public toilet, the address, the coordinates, the opening times and the accessibility for the disabled. In addition, further data are available, at least from Berlin and Boston, e.g., the name of the operator, the fee due for the use of the toilet, and the zip code. These additional data could definitely be of interest in a follow-up project.

The data from Berlin comes as an MS Excel list. Data types are string and integer, as well as float. Boolean data are represented as integer values and had to be converted for the purpose of this project.

B. Sources of the data

The main source of the data used was the publicly accessible list of the government of the city of Berlin, which is available at https://www.berlin.de/sen/uvk/_assets/verkehr/infektur/oeffliche-toiletten/berliner-toiletten-standorte.xlsx. In addition to the actual data, there is also a legend with explanations and acronyms on a second sheet in this file.

Unfortunately, I was unable to find the corresponding raw data underlying the Boston map. They are probably in a non-publicly accessible file.

The third data source is the Foursquare database. Here, the data of interest must be requested from the database with an appropriately formulated query.

III. Methodology

A. Exploratory Data Analysis

Berlin's dataset consists of 258 entries with 17 categories. It was found that there are no duplicate entries.

For the *isHandicappedAccessible* category in particular, it was necessary to change the data type from integer to bool. Otherwise, all the data used have already had a usable data type.

The column labeling of the opening times has been shortened from *Number (Time)* to *Time* for convenience.

17 entries are empty or unassigned. These were located exclusively in the *Time* column and were set to the value **Unknown** so that the later output provides a more meaningful statement than just an empty string.

B. Inferential statistical analysis

A statistical analysis of the data was not carried out as only position data, opening times and access for the disabled were considered, for which a statistical analysis would not result in any additionally usefull information.

C. Machine learnings that were used

Only position data, opening times, and access for the disabled were considered. These data are not suitable for prediction. Therefore, no machine learning algorithms were applied.

IV. Results

A. Overview

First, we had to retrieve the location of Berlin's center from Wikipedia, which is (52.518611, 13.408333) and superimpose the locations of the restrooms onto a map that is meant as a first glimpse at the data in the dataset.

In this map we want to distinguish between restrooms that are accessible by handicapped people (outer circle is yellow) versus those that are not (outer circle is blue). Additionally, we want to get an overview on the restrooms that are accessible for free (inner circle is black) compared to those that have to be paid for (inner circle red). The zoom is set to a value that shows Berlin in its entirety - not just the city's center.

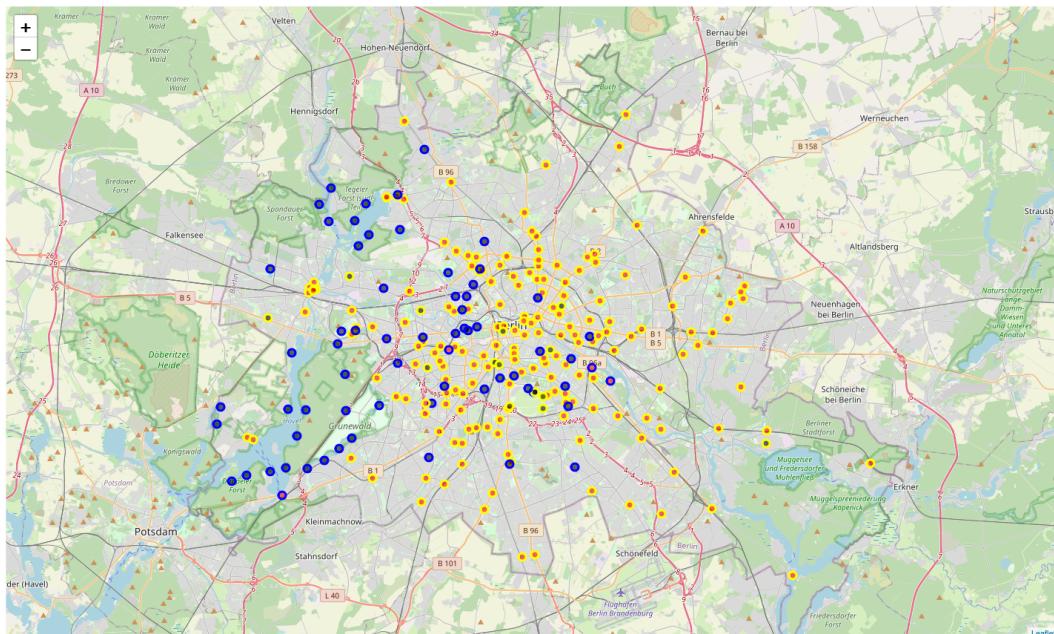


Figure 1 : Overview

B. Grouping and clustering locations

This view shows us a more or less even spread of public restrooms in the city of Berlin. Individual markers are green if the respective restroom can be accessed by somebody who is physically handicapped and red otherwise. Popups show some useful information about the selected restrooms, e.g., opening times, price, and accessibility by wheelchairs.

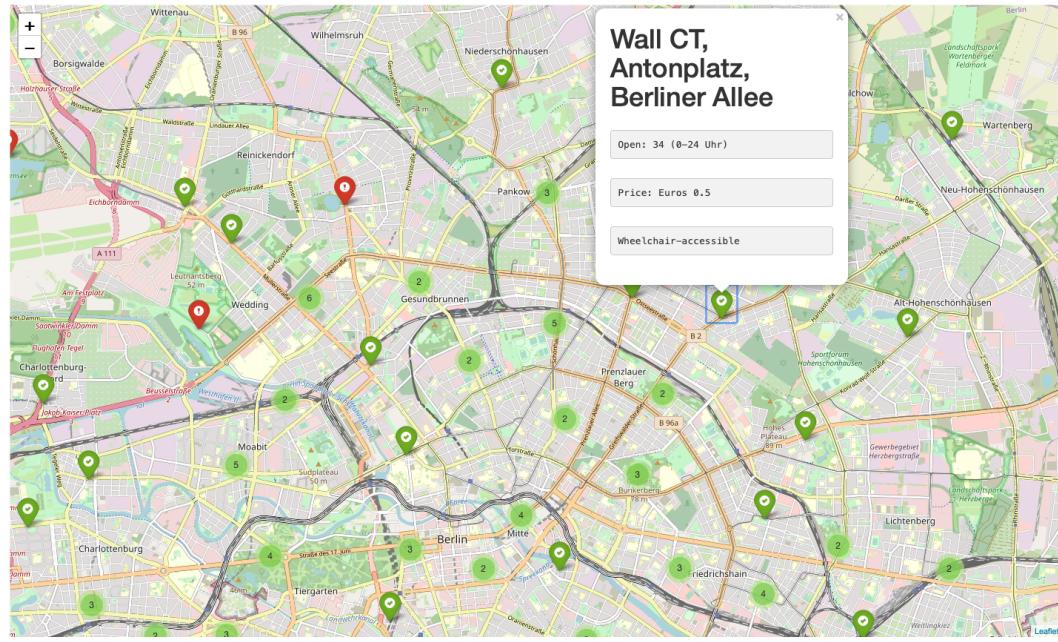


Figure 2 : Grouping and clustering with popup

C. Foursquare data on public restrooms in Berlin

I tried a couple of search queries on Foursquare – English as well as German terms. “Toilette” is the German equivalent for restroom – it leads to more hits than “restroom”, “lavatory”, and “toilet”. In total Foursquare reported 11 venues, which were reduced to 8 after omitting results without an entry for category.

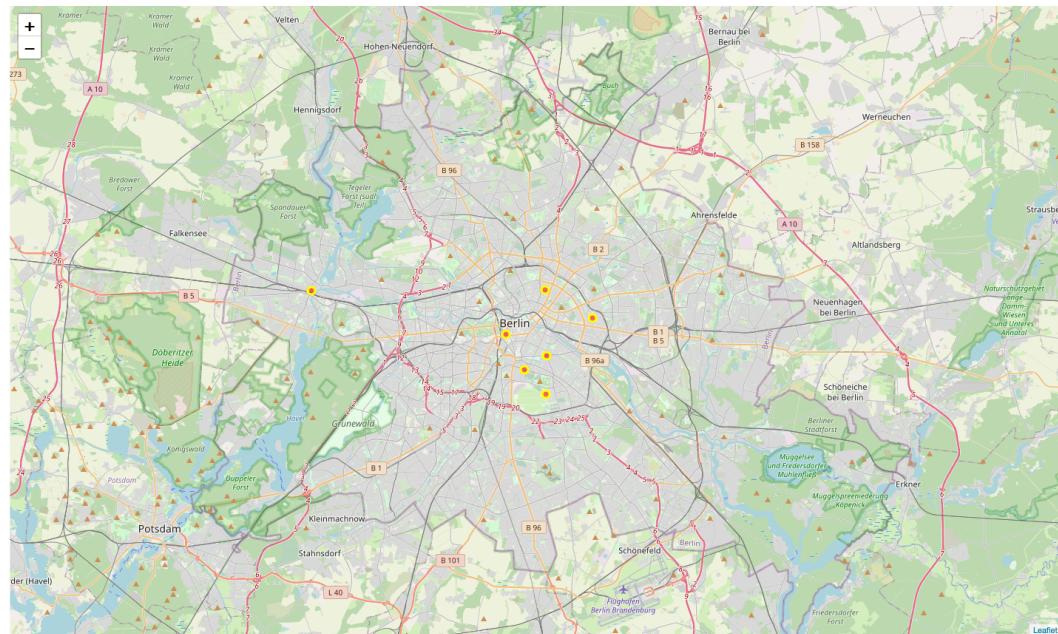


Figure 3 : Foursquare data on public restrooms in Berlin

D. Foursquare data on public restrooms in Boston, MA

Doing the same as in C above, but using “restroom” as search query, resulted in 7 venues or locations. The search radius was set to 30 km, which gives us some locations that are already well outside the city of Boston.

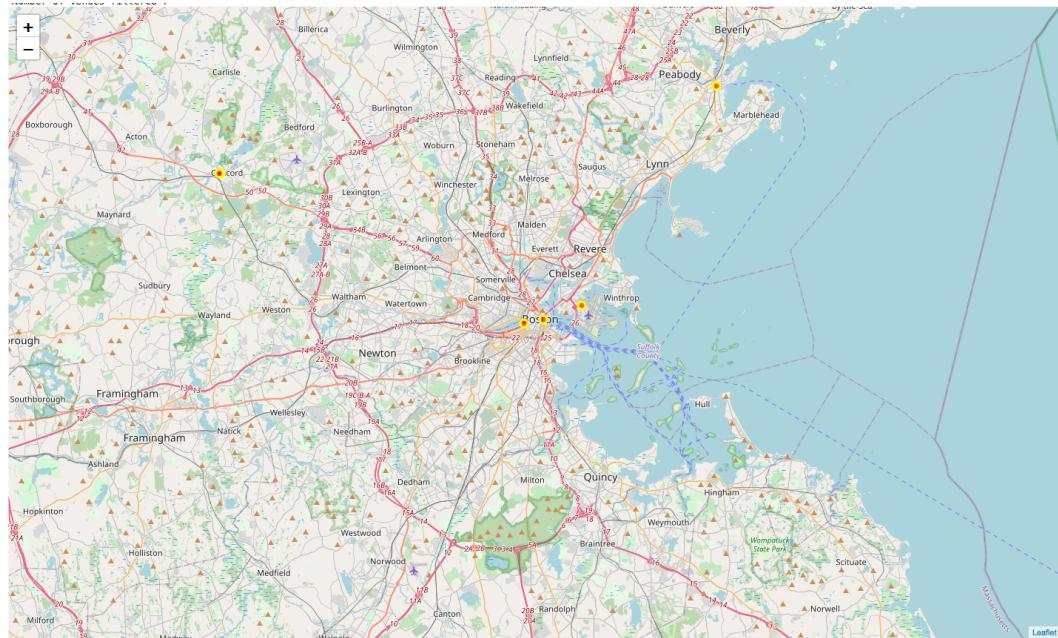


Figure 4 : Foursquare data on public restrooms in Boston, MA

E. Gold standard or reference

Following is a screenshot of the view provided by the city of Boston. This is actually what I was aiming for in the first place.

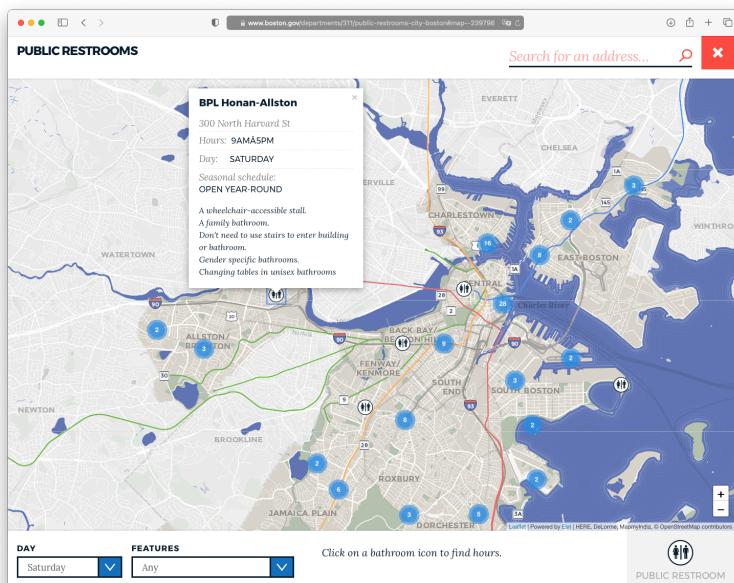


Figure 5 : Public restrooms in the city of Boston by www.boston.gov

V. Discussion

A. Observations

The business problem - showing a map of public restrooms in Berlin - can be realized with Python and publicly available data. The same applies to requirement 2, "Marking of public restrooms", which is arguably even better represented with the markers from the Folium library than the Boston website suggests.

Apparently, Foursquare's database lacks data on public restrooms. From the list provided by the government of Berlin, we know that there are 258 locations – Foursquare's database contains only 8. The same actually applies to the city of Boston.

Unfortunately, raw data from Boston – similar to the dataset from Berlin – is not available publicly. While requirement 3 “Comparison of the information provided by the two cities with Foursquare” is satisfied, the result – 8 for Berlin and 7 for Boston – is useless when taking into account that there must be hundreds more restrooms in each city.

The absence of a dataset from Boston makes requirement 1 “Analysis of the density of toilets in the two cities” unfulfilled. I hope the reader understands that I have refrained from counting the places on the map of Boston by hand.

B. Recommendations

Better than carrying a Jupyter notebook with you is a mobile phone application or an internet site that allows you to find public toilets in Berlin depending on your location. Let's see – maybe that will be one of my next projects ☺.

Foursquare should definitely consider to update their database on public restrooms.

VI. Conclusion

Boston's public restroom map provides the most professional and comprehensive view on public restrooms in this comparison. Next is the map generated in Folium using the data provided by Berlin's government. Foursquare has some room for improvement, as it shows only a fraction of the locations given by the two cities.

It was fun applying my newly acquired knowledge – thanks, IBM.

VII. Copyrights

- © 2010–2019 Vladimir Agafonkin.
- © OpenStreetMap contributors.
- © 1995–2020 Esri.
- © City of Boston.