

The perfect location for Berlins next climbing hall

Data Science specialization capstone project

Fabian Frieese

18.10.2019

Background

- Climbing sport quite popular in Germany
 - More than 1.000.000 members in national climbing association
 - High increase in memberships shows need for additional climbing halls
 - Many climbers mainly climb in indoor halls
 - Especially northern part of Germany with flat geography
- Focus on where to open a climbing hall in Berlin

Business problem

- Opening and maintaining a climbing hall is quite expensive
 - High entrance fees
 - Finding appropriate locations for a new climbing hall is a crucial task

Interest

- Investors thinking of investing in a new climbing hall
- Very passionate climbers thinking about relocation

Foursquare as data source

Vorschläge für kletterhallen in der Nähe von Berlin

Filter: [Specials](#) [Nicht hier gewesen](#) [Following](#) [Preis](#) [Jetzt geöffnet](#) [Gespeichert](#) [gefällt das](#)

Entdecke Orte, die deine Freunde und Experten lieben

Wenn du ein Konto erstellst, kannst du Freunden und Experten folgen, denen du vertraust, und sehen, welche Orte sie empfehlen.

[Mit Facebook anmelden](#) or [Mit E-Mail registrieren](#)



1. Berta Block

Kletterhalle • €€
Mühlenstr. 62, Berlin

[Speichern](#)



Rebecca T. • Juni 25, 2015

Die Boulder-Hallen sprießen überall aus dem Boden, aber Berta Block hat es raus: Routen für jeden Schwierigkeitsgrad, nette Mitarbeiter und obendrauf Café-Flair: guter Kaffee, selbstgebackener Kuchen!

9.3



2. Boulder Garten

Kletterhalle
Thiemannstr 1, Tor 4, 2.Hinterhof, Berlin

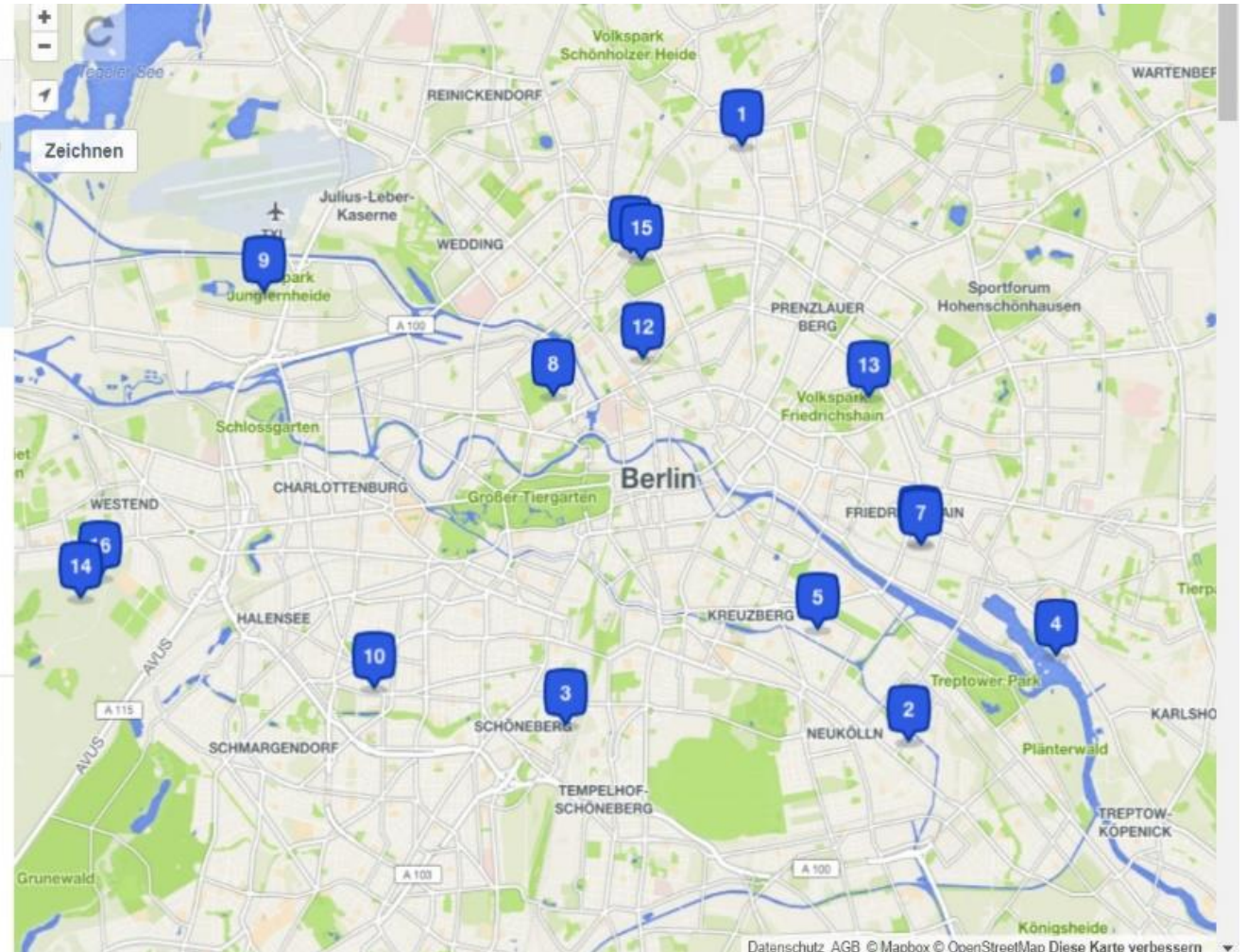
[Speichern](#)



www.Ricardo.sexy23.pw B. • July 17

Excellent bouldering gym, big and with many features, has a slackline and a Sauna!

9.1



Additional data sources

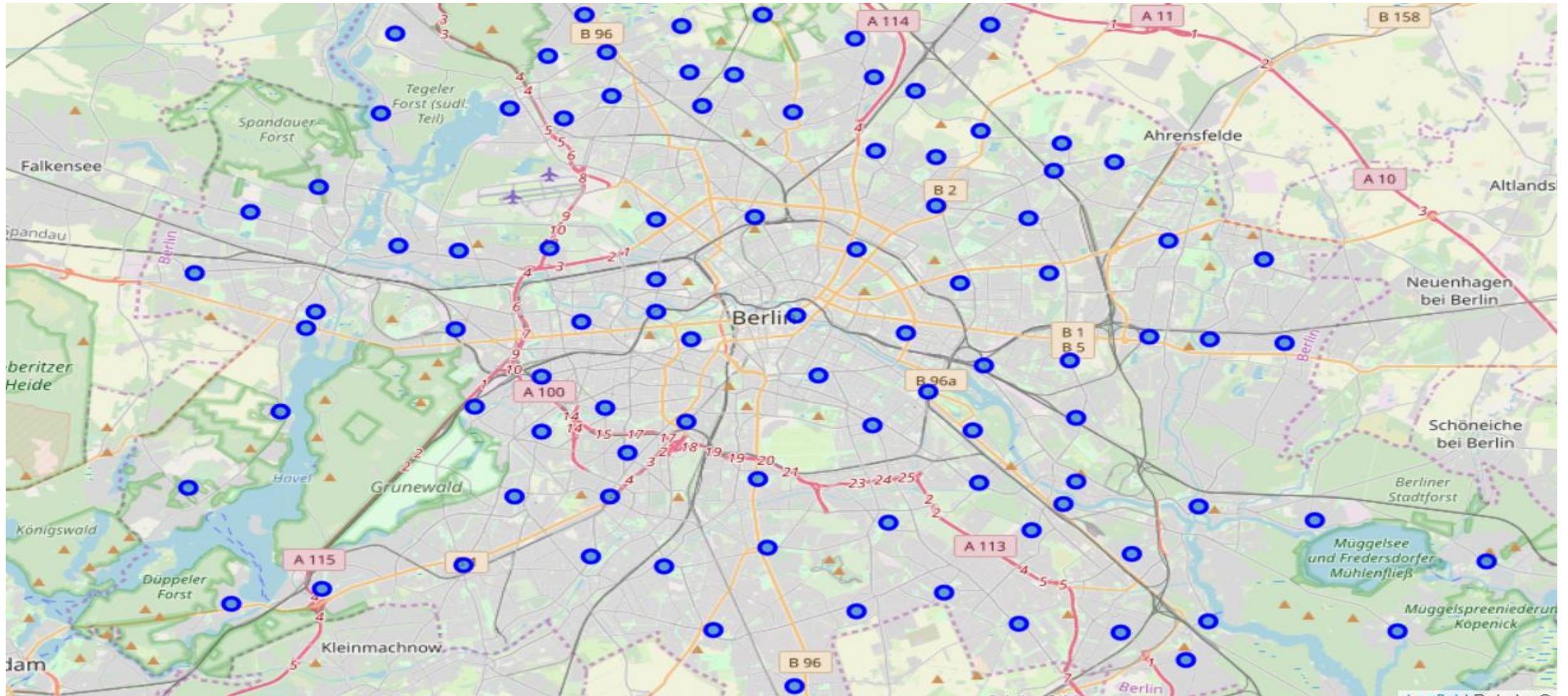
- Information about the neighbourhoods of Berlin and the people living there
- Statistics about personal profile of climbers and their interest

Scraping neighbourhood information

- Web scraping from Wikipedia with beautiful soup
- Enhancing dataframe with coordinates with geocoder

| | Number | Neighbourhood | Borough | Area | Inhabitants | Inhabitants per sqkm | Longitude | Latitude |
|---|--------|----------------|--------------------------|------|-------------|----------------------|-----------|-----------|
| 0 | 0101 | Mitte | Mitte | 10,7 | 99.998 | 9346 | 13.402376 | 52.517690 |
| 1 | 0102 | Moabit | Mitte | 7,72 | 78.491 | 10.167 | 13.342542 | 52.530102 |
| 2 | 0103 | Hansaviertel | Mitte | 0,53 | 5.831 | 11.002 | 13.341872 | 52.519123 |
| 3 | 0104 | Tiergarten | Mitte | 5,17 | 14.529 | 2810 | 13.357260 | 52.509778 |
| 4 | 0105 | Wedding | Mitte | 9,23 | 86.468 | 9368 | 13.341970 | 52.550123 |
| 5 | 0106 | Gesundbrunnen | Mitte | 6,13 | 94.293 | 15.382 | 13.384846 | 52.550920 |
| 6 | 0201 | Friedrichshain | Friedrichshain-Kreuzberg | 9,78 | 131.953 | 13.492 | 13.450290 | 52.512215 |
| 7 | 0202 | Kreuzberg | Friedrichshain-Kreuzberg | 10,4 | 154.010 | 14.809 | 13.411914 | 52.497644 |

Visualizing the neighbourhoods of Berlin with folium



Getting venue information from Foursquare: parameter choice: Radius=10000m -> too big

| Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---------------------------|-----------------------|------------------------|----------------------------------|----------------|-----------------|--------------------|
| Wedding | 52.550123 | 13.341970 | Magic Mountain Kletterhalle | 52.548613 | 13.381902 | Climbing Gym |
| Wedding | 52.550123 | 13.341970 | Waldhochseilgarten Jungfernheide | 52.543001 | 13.290644 | Rock Climbing Spot |
| Gesundbrunnen | 52.550920 | 13.384846 | Berta Block | 52.563548 | 13.409754 | Climbing Gym |
| Weißensee | 52.554619 | 13.463002 | Berta Block | 52.563548 | 13.409754 | Climbing Gym |
| Blankenburg | 52.593211 | 13.454182 | Berta Block | 52.563548 | 13.409754 | Climbing Gym |
| Heinersdorf | 52.572825 | 13.437015 | Berta Block | 52.563548 | 13.409754 | Climbing Gym |
| Stadtrandsiedlung Malchow | 52.571019 | 13.463285 | Berta Block | 52.563548 | 13.409754 | Climbing Gym |
| Pankow | 52.597663 | 13.436351 | Berta Block | 52.563548 | 13.409754 | Climbing Gym |
| Niederschönhausen | 52.585806 | 13.401397 | Berta Block | 52.563548 | 13.409754 | Climbing Gym |
| Niederschönhausen | 52.585806 | 13.401397 | Magic Mountain Kletterhalle | 52.548613 | 13.381902 | Climbing Gym |
| Rosenthal | 52.598319 | 13.375519 | Berta Block | 52.563548 | 13.409754 | Climbing Gym |
| Wilhelmsruh | 52.588012 | 13.362206 | Berta Block | 52.563548 | 13.409754 | Climbing Gym |

Getting venue information from Foursquare: parameter choice: Radius=500m -> too small

| Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---------------------|-----------------------|------------------------|----------------------------------|----------------|-----------------|--------------------|
| Gesundbrunnen | 52.550920 | 13.384846 | Magic Mountain Kletterhalle | 52.548613 | 13.381902 | Climbing Gym |
| Wilmerdorf | 52.487115 | 13.320330 | Boulderworx | 52.486881 | 13.318215 | Climbing Gym |
| Charlottenburg-Nord | 52.540525 | 13.296266 | Waldhochseilgarten Jungfernheide | 52.543001 | 13.290644 | Rock Climbing Spot |
| Hellersdorf | 52.536854 | 13.604774 | BergWerk | 52.537724 | 13.603606 | Rock Climbing Spot |

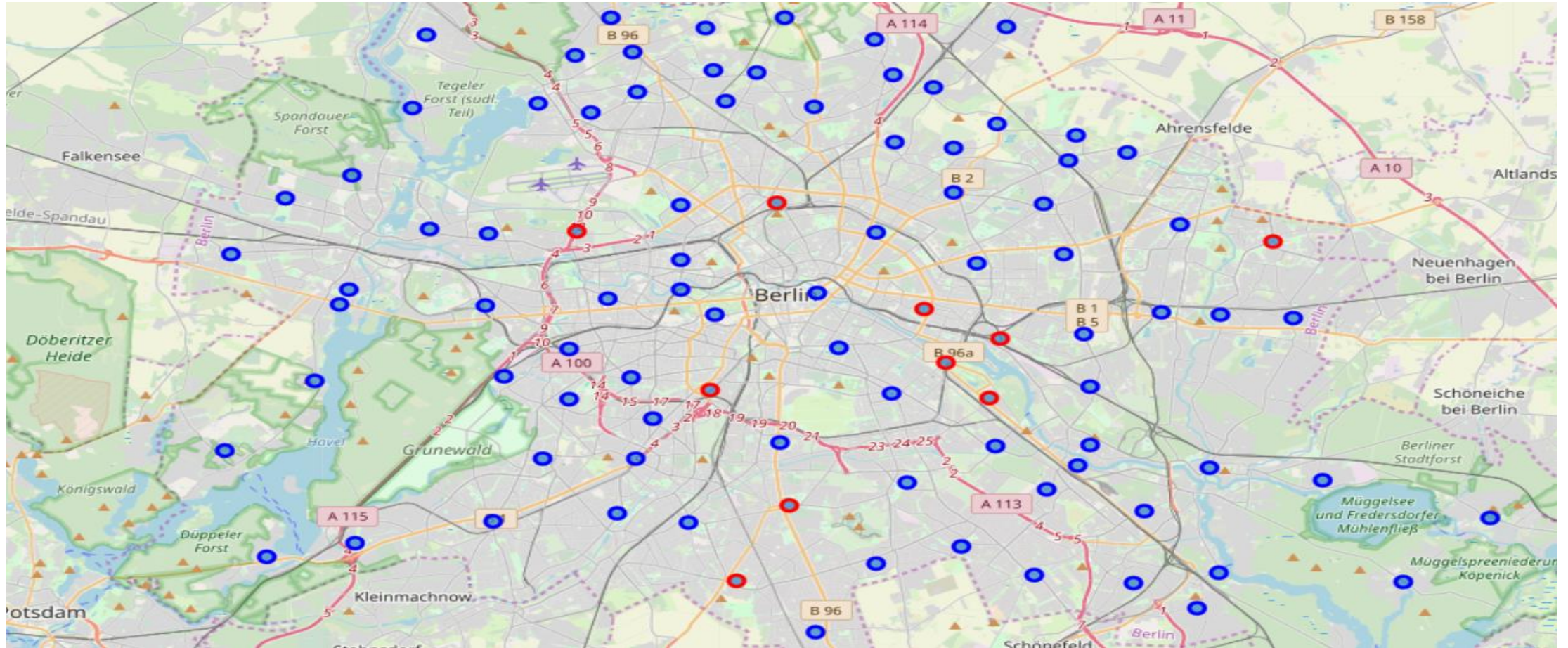
Getting venue information from Foursquare: parameter choice: Radius=1750m -> good :)

| Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---------------------|-----------------------|------------------------|----------------------------------|----------------|-----------------|--------------------|
| Moabit | 52.530102 | 13.342542 | DAV Kletterzentrum Berlin | 52.528419 | 13.362783 | Climbing Gym |
| Gesundbrunnen | 52.550920 | 13.384846 | Magic Mountain Kletterhalle | 52.548613 | 13.381902 | Climbing Gym |
| Friedrichshain | 52.512215 | 13.450290 | Der Kegel | 52.507287 | 13.454390 | Climbing Gym |
| Charlottenburg-Nord | 52.540525 | 13.296266 | Waldhochseilgarten Jungfernheide | 52.543001 | 13.290644 | Rock Climbing Spot |
| Schöneberg | 52.482157 | 13.355190 | Bright Site | 52.481732 | 13.365768 | Climbing Gym |
| Mariendorf | 52.440080 | 13.390028 | Südbloc Boulderhalle | 52.439969 | 13.379566 | Climbing Gym |
| Alt-Treptow | 52.492563 | 13.459874 | Bouldergarten | 52.479470 | 13.451368 | Climbing Gym |
| Plänterwald | 52.479544 | 13.478808 | Ostbloc Boulderhalle | 52.491496 | 13.487964 | Climbing Gym |
| Hellersdorf | 52.536854 | 13.604774 | BergWerk | 52.537724 | 13.603606 | Rock Climbing Spot |
| Rummelsburg | 52.501370 | 13.483514 | Ostbloc Boulderhalle | 52.491496 | 13.487964 | Climbing Gym |

Cleaning the venue data

| Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---------------------|-----------------------|------------------------|----------------------------------|----------------|-----------------|--------------------|
| Moabit | 52.530102 | 13.342542 | DAV Kletterzentrum Berlin | 52.528419 | 13.362783 | Climbing Gym |
| Gesundbrunnen | 52.550920 | 13.384846 | Magic Mountain Kletterhalle | 52.548613 | 13.381902 | Climbing Gym |
| Friedrichshain | 52.512215 | 13.450290 | Der Kegel | 52.507287 | 13.454390 | Climbing Gym |
| Charlottenburg-Nord | 52.540525 | 13.296266 | Waldhochseilgarten Jungfernheide | 52.543001 | 13.290644 | Rock Climbing Spot |
| Schöneberg | 52.482157 | 13.355190 | Bright Site | 52.481732 | 13.365768 | Climbing Gym |
| Mariendorf | 52.440080 | 13.390028 | Südbloc Boulderhalle | 52.439969 | 13.379566 | Climbing Gym |
| Alt-Treptow | 52.492563 | 13.459874 | Bouldegarten | 52.479470 | 13.451368 | Climbing Gym |
| Plänterwald | 52.479544 | 13.478808 | Ostbloc Boulderhalle | 52.491496 | 13.487964 | Climbing Gym |
| Hellersdorf | 52.536854 | 13.604774 | BergWerk | 52.537724 | 13.603606 | Rock Climbing Spot |
| Rummelsburg | 52.501370 | 13.483514 | Ostbloc Boulderhalle | 52.491496 | 13.487964 | Climbing Gym |

Visualization of neighbourhoods with climbing halls



Method 1: Finding the neighbourhoods with the most venues

| Neighborhood | |
|-----------------|-----|
| Gesundbrunnen | 100 |
| Tiergarten | 100 |
| Halensee | 100 |
| Hansaviertel | 100 |
| Rummelsburg | 100 |
| Prenzlauer Berg | 100 |
| Friedrichshain | 100 |
| Friedenau | 100 |
| Steglitz | 100 |
| Fennpfuhl | 100 |
| Charlottenburg | 100 |

| | |
|-------------|-----|
| Neukölln | 100 |
| Wedding | 100 |
| Moabit | 100 |
| Alt-Treptow | 100 |
| Mitte | 100 |
| Wilmerdorf | 100 |
| Kreuzberg | 100 |
| Schöneberg | 100 |
| Tempelhof | 98 |
| Borsigwalde | 82 |
| Westend | 79 |

Method 2: Venues close to climbing halls – Frequent venues

Frequent near climbing halls:

| | index | 0 |
|-----|--------------------|----------|
| 36 | Café | 0.059355 |
| 192 | Supermarket | 0.046452 |
| 14 | Bar | 0.037419 |
| 45 | Coffee Shop | 0.029677 |
| 141 | Park | 0.024516 |
| 136 | Nightclub | 0.023226 |
| 12 | Bakery | 0.023226 |
| 150 | Pizza Place | 0.020645 |
| 112 | Italian Restaurant | 0.020645 |
| 106 | Ice Cream Shop | 0.018065 |

Frequent in all neighbourhoods:

| | index | 0 |
|-----|--------------------|----------|
| 310 | Supermarket | 0.086616 |
| 59 | Café | 0.047226 |
| 176 | Italian Restaurant | 0.036213 |
| 163 | Hotel | 0.025413 |
| 231 | Park | 0.025413 |
| 140 | German Restaurant | 0.023931 |
| 24 | Bakery | 0.023719 |
| 105 | Drugstore | 0.022660 |
| 54 | Bus Stop | 0.020330 |
| 166 | Ice Cream Shop | 0.020119 |

Method 2: Venues close to climbing halls – biggest differences close to climbing halls

Top 10:

| | | | |
|-------------------------------|----------|----------|----------|
| Bar | 0.017577 | 0.037419 | 0.019842 |
| Nightclub | 0.005294 | 0.023226 | 0.017931 |
| Café | 0.047226 | 0.059355 | 0.012129 |
| Coffee Shop | 0.019060 | 0.029677 | 0.010618 |
| Climbing Gym | 0.001694 | 0.010323 | 0.008628 |
| Rock Club | 0.001694 | 0.010323 | 0.008628 |
| Vegetarian / Vegan Restaurant | 0.005930 | 0.012903 | 0.006974 |
| Pizza Place | 0.013765 | 0.020645 | 0.006880 |
| Airport Service | 0.001271 | 0.007742 | 0.006471 |
| Wine Bar | 0.004024 | 0.010323 | 0.006299 |

Bottom 10:

| | | | |
|--------------------|----------|----------|-----------|
| Gas Station | 0.007624 | 0.002581 | -0.005043 |
| Plaza | 0.011436 | 0.005161 | -0.006275 |
| Trattoria/Osteria | 0.012071 | 0.005161 | -0.006910 |
| Greek Restaurant | 0.010801 | 0.003871 | -0.006930 |
| Tram Station | 0.012706 | 0.005161 | -0.007545 |
| German Restaurant | 0.023931 | 0.012903 | -0.011027 |
| Drugstore | 0.022660 | 0.010323 | -0.012337 |
| Hotel | 0.025413 | 0.010323 | -0.015090 |
| Italian Restaurant | 0.036213 | 0.020645 | -0.015568 |
| Supermarket | 0.086616 | 0.046452 | -0.040164 |

Method 2: Venues close to climbing halls –
calculating the score

Score

$$\begin{aligned} &= \sum \text{Frequencies of Top 10 venues} \quad - \\ &\quad \sum \text{Frequencies of Bottom 10 venues} \\ &= \text{Top Score} - \text{Neg Score} \end{aligned}$$

Method 2: Venues close to climbing halls – Top 20 neighbourhoods with similar venues

| Neighbourhood | Pos Score | Neg Score | Tot Score |
|---------------------|-----------|-----------|-----------|
| Friedrichshain | 0.330000 | 0.050000 | 0.280000 |
| Gesundbrunnen | 0.320000 | 0.040000 | 0.280000 |
| Neukölln | 0.350000 | 0.070000 | 0.280000 |
| Alt-Treptow | 0.280000 | 0.040000 | 0.240000 |
| Rummelsburg | 0.300000 | 0.080000 | 0.220000 |
| Kreuzberg | 0.280000 | 0.060000 | 0.220000 |
| Wedding | 0.280000 | 0.110000 | 0.170000 |
| Blankenfelde | 0.142857 | 0.000000 | 0.142857 |
| Schöneberg | 0.230000 | 0.100000 | 0.130000 |
| Prenzlauer Berg | 0.190000 | 0.070000 | 0.120000 |
| Moabit | 0.190000 | 0.110000 | 0.080000 |
| Fennpfuhl | 0.250000 | 0.180000 | 0.070000 |
| Steglitz | 0.140000 | 0.160000 | -0.020000 |
| Charlottenburg-Nord | 0.157143 | 0.185714 | -0.028571 |
| Friedenau | 0.160000 | 0.210000 | -0.050000 |
| Plänterwald | 0.134328 | 0.223881 | -0.089552 |
| Mitte | 0.070000 | 0.160000 | -0.090000 |
| Hansaviertel | 0.100000 | 0.190000 | -0.090000 |
| Dahlem | 0.127273 | 0.218182 | -0.090909 |
| Tiergarten | 0.080000 | 0.190000 | -0.110000 |

Method 3: Logistic regression – Difficulties in applying logistic regression

- Knowledge where climbing halls are (suitable neighbourhoods)
 - NO knowledge which neighbourhoods are inherently NOT suitable
- No pristine training data set

Method 3: Logistic regression – Solution for the difficulties

- Assumption that the market for climbing halls is nearly saturated
 - Assuming all neighbourhoods without climbing halls are not suitable halls only lead to small errors
- Fitting the model using all data for training
- Classifying all neighbourhoods with fitted logistic regression
- Looking for neighbourhoods without existing halls now classified as suitable

Method 3: Logistic regression – accuracy metrics

| | |
|---------------------|--------|
| Jaccard similiarity | 0,9375 |
| F1-Score | 0,9436 |
| Log Loss | 0,6273 |

- All neighbourhoods with existing climbing halls classified as suitable
- Six additional neighbourhoods classified as suitable
- Deviations from perfect fit of regression only due to neighbourhoods newly defined as suitable

➤ Very good results!

Method 3: Logistic regression – suitable neighbourhoods

```
In [32]: ▶ new = yhat - Y_train  
new_neighbourhoods = berlin_grouped["Neighbourhood"].loc[new == 1]  
new_neighbourhoods
```

```
Out[32]: 18    Fennpfuhl  
         44    Kreuzberg  
         56      Mitte  
         61    Neukölln  
         83    Tempelhof  
         88    Wedding
```

Summarizing the results

| Neighbourhood | Biggest neighbourhoods | Top 10 / Bottom 10 Venues | Logistic regression | Exist climbinghall |
|---------------------|------------------------|---------------------------|---------------------|--------------------|
| Alt-Treptow | X | X | X | X |
| Blankenfelde | | X | | |
| Charlottenburg | X | | | |
| Charlottenburg-Nord | | X | X | X |
| Dahlem | | X | | |
| Fennpfuhl | X | X | X | |
| Friedenau | X | X | | |
| Friedrichshain | X | X | X | X |
| Gesundbrunnen | X | X | X | X |
| Halensee | X | | | |
| Hansaviertel | X | X | | |
| Hellersdorf | | | X | X |
| Kreuzberg | X | X | X | |
| Mariendorf | | | X | X |
| Marienfelde | | | X | X |
| Mitte | X | X | X | |
| Moabit | X | X | | |
| Neukölln | X | X | X | |
| Plänterwald | | X | X | |
| Prenzlauer Berg | X | X | | |
| Rummelsburg | X | X | X | X |
| Schöneberg | X | X | X | X |
| Steglitz | X | X | | |
| Tempelhof | X | | X | |
| Tiergarten | X | X | | |
| Wedding | X | X | X | |
| Wilmersdorf | X | | | |

- Blue: hall existing
- Red: only one method positive
- Orange: two methods positive, but not logistic regression
- Light green: logistic regression and one other method positive
- Dark green: All methods positive

Potential neighbourhoods for new halls

| Neighbourhood | Biggest neighbourhoods | Top 10 / Bottom 10 Venues | Logistic regression | Exist climbinghall |
|---------------------|------------------------|---------------------------|---------------------|--------------------|
| Alt-Treptow | X | X | X | X |
| Blankenfelde | | X | | |
| Charlottenburg | X | | | |
| Charlottenburg-Nord | | X | X | X |
| Dahlem | | X | | |
| Fennpfuhl | X | X | X | |
| Friedenau | X | X | | |
| Friedrichshain | X | X | X | X |
| Gesundbrunnen | X | X | X | X |
| Halensee | X | | | |
| Hansaviertel | X | X | | |
| Hellersdorf | | | X | X |
| Kreuzberg | X | X | X | |
| Mariendorf | | | X | X |
| Marienfelde | | | X | X |
| Mitte | X | X | X | |
| Moabit | X | X | | |
| Neukölln | X | X | X | |
| Plänterwald | | X | X | |
| Prenzlauer Berg | X | X | | |
| Rummelsburg | X | X | X | X |
| Schöneberg | X | X | X | X |
| Steglitz | X | X | | |
| Tempelhof | X | | X | |
| Tiergarten | X | X | | |
| Wedding | X | X | X | |
| Wilmerisdorf | X | | | |

- Fennpfuhl

- Kreuzberg

- Mitte

- Neukölln

- Wedding

Conclusion

- Simple study with only few lines of code
 - Still plausible and helpful results
 - Enormous power of quantitative studies
- Simplifying assumption: potential success of climbing hall depends on surrounding venues
 - Plausible for many real applications
 - Practical examples show that climbing halls can be successful even in remote places
- Method helps to narrow down list of candidate neighbourhoods

Future directions

- All methods can be developed further
 - Method 1: increase venue limit
 - Method 2: consider more categories und develop weighting methodology
 - Method 3: try other machine learning approaches like random forests or KNNs
- Take in to account expertise of marketing expert

Thank you very much
for your attention :)