# Analyzing tech hub real estate investments

A Coursera capstone project

# Background

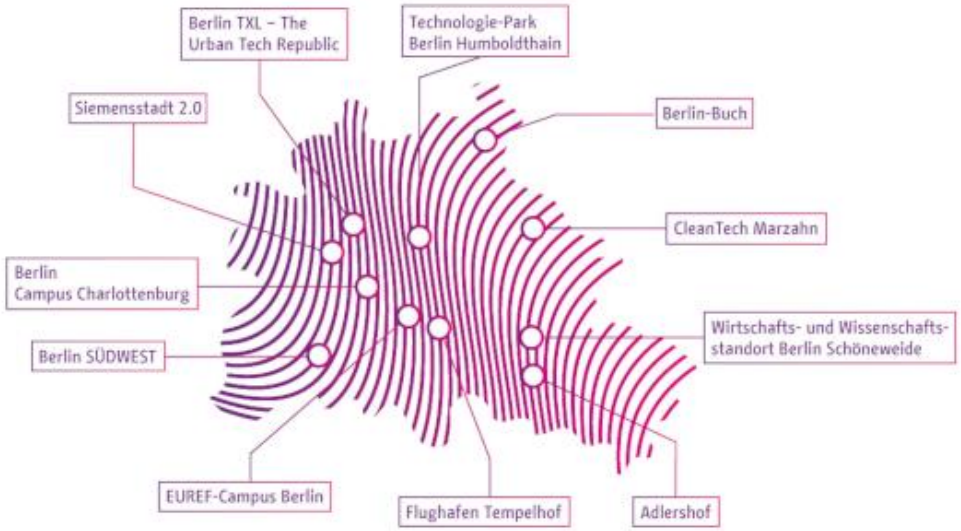
- The city of **Berlin, Germany** plans to strengthen its position as a startup hub
- To attract deep tech medical companies they’re considering to **reconstruct the “Mäusebunker”** and invest in laboratory equipment
- The “Mäusebunker” (which literaly translates to mice bunker) is a former animal testing lab
- As this is going to be a costly investment of approx. 90 Million € a lot of different aspects need to be considered to **ensure the likelihood of the success** of this investment
- Part of the initial analysis of a new real estate venture is to investigate if the **surrounding infrastructure** to evaluate if it fits the needs of the object
- This work will quantitative analyze this part of the analysis

# Problem & Interest

- Considering various data points prior to an investment is essential **to maximize the outcome of you investment**
- **Analyzing the surrounding infrastructure** is an essential part of the development of real estate.
- This is supports the decision whether the investment in the conversion of the building is worth the financial risk
- Supporting the **analysis with data improves the previous qualitative work** and gives a more objective indication
- The surrounding infrastructure will be compared to already established tech hubs within Berlin
- A **greater similarity** with existing hubs is considered a positive indicator for a possible fruitful investment.

# Data -Sources

- Postal codes and corresponding latitudes and longitudes by Git Hub user Zauberware [Link]
- Foursquare’s Places API – explore endpoint [Link]
- Location data of Berlin’s future list [Link]

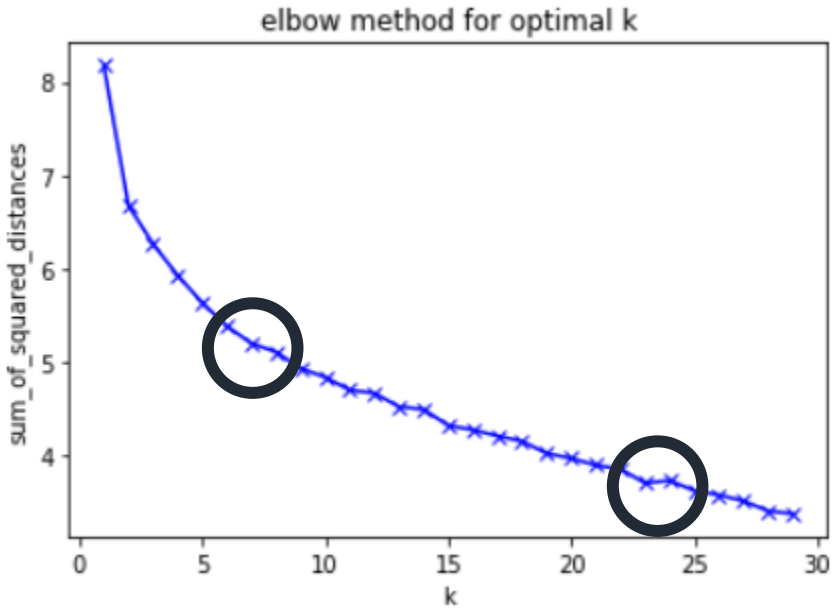


# Data -Cleaning

- Cleaning the data set containing the location data
  - dropping irrelevant columns (only postal code, place, latitude and longitude are needed)
  - dropping irrelevant rows (the data set contains data for all of Germany)
- Averaging the size of postal codes within Berlin to approx. 4 km²
- Requesting the max amount of results per call to 50 venues
- Grouping the venues of each postal code by category and sort the venue category from most to least common

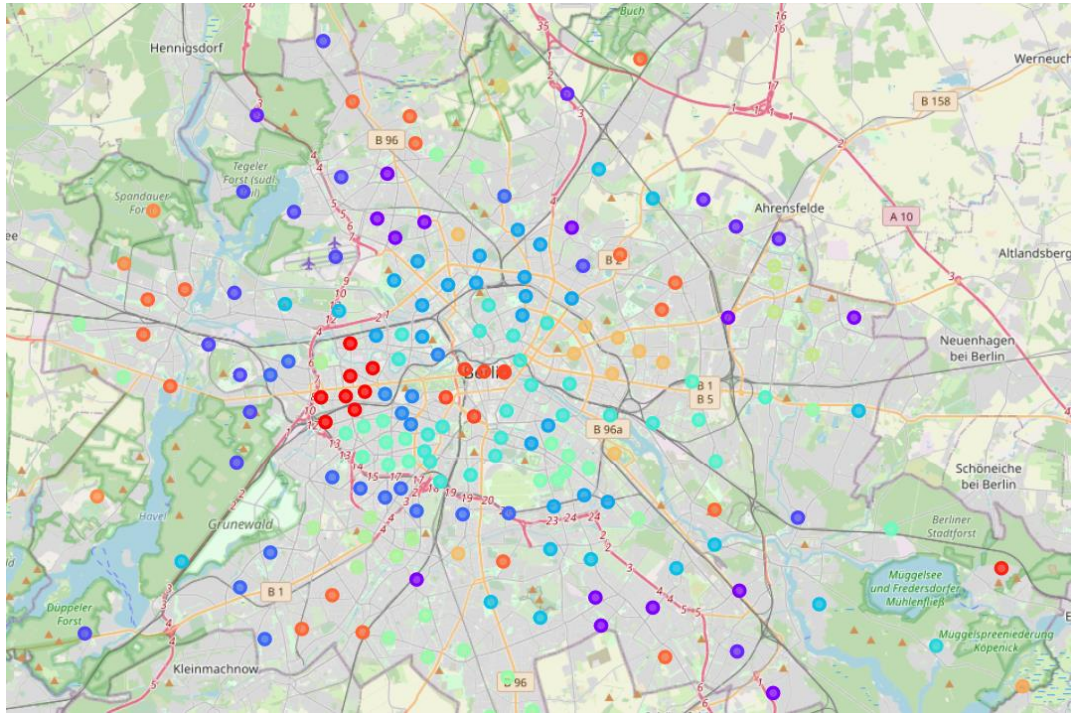
# Methodology

- K-means clustering to cluster the postal codes by similarity
- Using an elbow plot to evaluate the optimal value of k – we're looking for the point after which the distortion/inertia start decreasing
- Of the two possible options 7 and 23, the later seems to be more reasonable as 7 doesn't provide enough differentiation
- Using the 12 existing tech hubs as references for a positive infrastructure surrounding
- Comparing the 12 reference cluster to the cluster containing the Mäusebunker
- Visualizing all 23 clusters on a map of Berlin



| Object | Postal code | Cluster |
|---|---|---|
| Charlottenburger Innovations-Centrum – CHIC | 10625 | 0 |
| Existenzgründerzentrum Technische Dienstleistungen an der HTW Berlin | 10318 | 10 |
| Gründerinnenzentrum WeiberWirtschaft | 10115 | 9 |
| Gründerzentrum der HWR Berlin | 13629 | 7 |
| Gründungszentrum „GründerScout“ der Beuth Hochschule für Technik Berlin | 13409 | 17 |
| Innovations- und GründerZentrum Berlin-Adlershof – IGZ | 12489 | 1 |
| Innovations- und Gründerzentrum – IGZ – Campus Berlin-Buch | 13125 | 20 |
| Innovationspark Wuhlheide | 12555 | 3 |
| PHÖNIX Gründerzentrum Am Borsigturm | 13507 | 3 |
| Technologie- und Gründerzentrum Spreeknie – TGS | 12459 | 20 |
| Unternehmerinnen und Gründerinnen Zentrum in Charlottenburg-Wilmersdorf – UCW | 10713 | 11 |
| Wissenschafts- und Technologiepark Berlin-Adlershof | 12489 | 1 |
| Mäusebunker | 12203 | 14 |

# Results




- The cluster containing the “Mäusebunker” is number 14
- None of the other tech hub we’re using as reference for an supporting infrastructure is in the same cluster as the “Mäusebunker”
- Thereby we can see that the surrounding venues of the “Mäusebunker” and the other objects are not similar to each other

# Discussion and Conclusion

- Using “only” 50 venues per 4 km² could be not enough data points in a highly dense city like Berlin.
- The surrounding venues are only part of the indication whether a real estate should be developed.
- The used reference objects could have a suboptimal surrounding infrastructure and be successful regardless of that fact.
- A clear recommendation cannot be given based on the taken analysis. It can only be part of a wider evaluation.
- It could be more reasonable using an expert opinion to define which venues are most supportive for a venture like the researched one.
- The results will be taken into consideration and will be followed by a qualitative research to support the findings.
- The results will be part of a Master’s thesis in architecture