# Capstone Project - The Battle of Neighborhoods

Restaurant Real-estate in Berlin

# Restaurant real-estate listings in Berlin

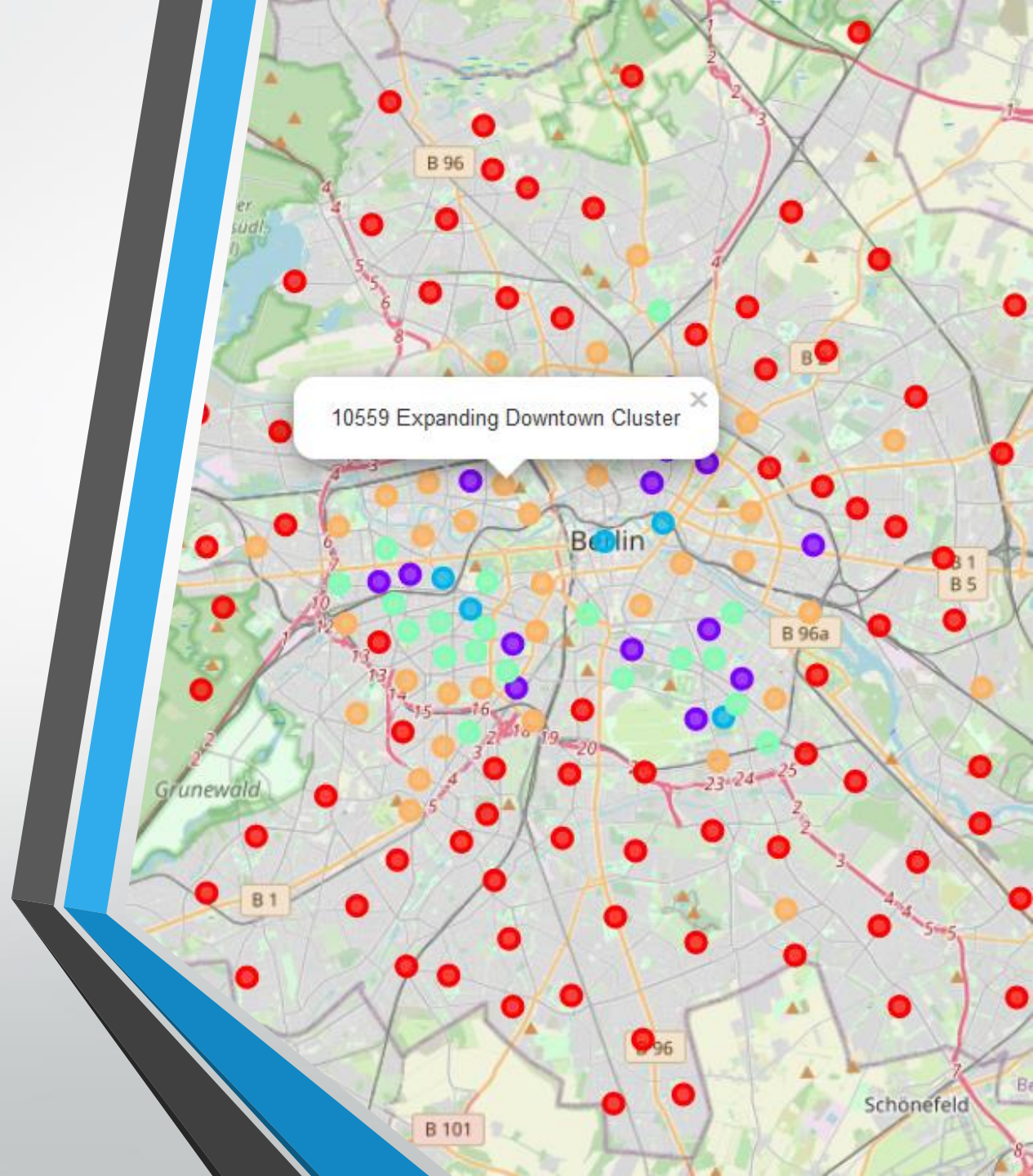
- It could be difficult to quickly assess the real-estate listings in a bigger city, such as Berlin if one isn't familiar with every neighborhood
- A person looking to open a restaurant may have an idea of the neighborhood profile they are looking for
- Age profile of the resident and types of nearby venues should play a role in the decision process
- Another factor could be the amount of competition in the neighborhood

# Data sets and sources

- Demographic data from December 31st 2021 split by postal codes and age ranges : https://www.statistik-berlin-brandenburg.de/publikationen/stat_berichte/2021/SB_A01-05-00_2020h02_BE.xlsx
- Coordinates f the areas based on postal code retrieved using the **pgeocode** library
- Nearby venue data (<500m) and venue categories using the **Foursqare** service and the groupings from their website for developers https://developer.foursquare.com/docs/build-with-foursquare/categories/
- In total, 180 rows and 12 features in the dataset

# Using k-means to cluster the neighborhoods

- Using the K-means algorithm we've arrived at 5 distinct clusters:
  - Tourist Traps - high in culture, recreation and shopping, mid-level in competition, not many children living in the area.
  - Wider Downtown - second highest concentration of places of culture, these places are just off the main tourist traps and could be a great place for a restaurant aimed at locals.
  - Food Central - largest concentration of competition, also highest in nightlife venues
  - Expanding Downtown - highest concentration of young adults, not too much competition, possible areas for downtown expansion
  - Residential - largest cluster with low sores in all venue types, highest concentration of retired residents and kids



| | Female | Kids | Young Adults | Adults | Older Adults | Retirees | Culture | Food | Nightlife | Recreation | Shopping | Travel | Category |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cluster Labels** | | | | | | | | | | | | | |
| **0** | 0.511118 | 0.165760 | 0.090057 | 0.247539 | 0.270032 | 0.226612 | 0.271845 | 1.699029 | 0.116505 | 1.203883 | 2.067961 | 1.000000 | Residential |
| **1** | 0.490344 | 0.149513 | 0.098121 | 0.388262 | 0.257453 | 0.106653 | 1.933333 | 46.466667 | 9.600000 | 3.000000 | 13.266667 | 1.933333 | Food Central |
| **2** | 0.485346 | 0.127782 | 0.098617 | 0.336357 | 0.256839 | 0.180404 | 6.600000 | 29.600000 | 6.400000 | 4.600000 | 19.400000 | 7.600000 | Tourist Traps |
| **3** | 0.499760 | 0.139919 | 0.095536 | 0.317828 | 0.271832 | 0.174885 | 2.526316 | 23.842105 | 3.105263 | 2.947368 | 7.000000 | 3.368421 | Wider Downtown |
| **4** | 0.498277 | 0.152255 | 0.110557 | 0.324728 | 0.246626 | 0.165835 | 1.394737 | 10.157895 | 1.736842 | 2.552632 | 4.921053 | 1.921053 | Expanding Downtown |

# Feature averages for each cluster

Using these we've made observations about each cluster and labeled their categories

| | Female | Kids | Young Adults | Adults | Older Adults | Retirees | Culture | Food | Nightlife | Recreation | Shopping | Travel | Cluster Labels | Category |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **10115** | 0.492448 | 0.168237 | 0.092082 | 0.421340 | 0.252281 | 0.066061 | 5 | 13 | 6 | 4 | 6 | 6 | 4 | Expanding Downtown |
| **10179** | 0.495657 | 0.126057 | 0.113029 | 0.346819 | 0.220343 | 0.193752 | 4 | 9 | 9 | 0 | 1 | 2 | 4 | Expanding Downtown |
| **10243** | 0.488476 | 0.130933 | 0.118787 | 0.395478 | 0.205338 | 0.149464 | 2 | 8 | 7 | 4 | 8 | 3 | 4 | Expanding Downtown |
| **10245** | 0.476699 | 0.145493 | 0.095713 | 0.511121 | 0.194660 | 0.053013 | 3 | 8 | 6 | 6 | 2 | 0 | 4 | Expanding Downtown |
| **10249** | 0.495305 | 0.149195 | 0.084943 | 0.426693 | 0.205265 | 0.133903 | 2 | 8 | 0 | 3 | 3 | 2 | 4 | Expanding Downtown |

# Detailed information for each postal code

The detailed data set can be used for quick analysis of any interesting venue listing.

# Results and Conclusion

- The neighborhood clusters serve to identify neighborhoods with a similar profile
- This in turn can serve to quickly identify which neighborhoods have the characteristics desired by the user
- The user can use the map to pick which neighborhoods would interest them in their search for a venue and then use the provided postal codes to filter the venue listings
- Alternatively, they can use the postal codes of the available listings to cross-check the neighborhood profile in the detailed data set