AIRBNB ANALYSIS VIENNA

DETECTING SPATIAL DISTRIBUTION OF ARIBNB‘S BY PRICE

JASMIN KOFLER & BENEDIKT MÜLLER

UNIV.-PROF. DR. KRZYSZTOF JANOWICZ
PS Spatial Data Science
University of Vienna

#1 MOTIVATION & RESEARCH GOAL

- Discover **spatial patterns** in the distribution of Airbnb's in Vienna
- Focus on **interrelations** between **price** per night and the **centrality**
- Put theoretical learnings from the course to practice
  - Methods: NN, classification, DBSCAN clustering, convex hulls, distance calculation

#2 PYTHON PACKAGES

- Contextily
- Foilum
- Geopandas
- Matplotlib
- Numpy
- Pandas
- Scikit-learn
- Scipy
- Shapely

#3 INPUT DATA & PRE-PROCESSING

ORIGINAL DATA

- **Source:** Inside Airbnb
  - 14697 data entries
  - 18 columns

	id	neighbourhood	latitude	longitude	price	number_of_reviews
0	38768	Leopoldstadt	48.219240	16.378310	77.0	384
1	40625	Rudolfsheim-F□nfhaus	48.184340	16.327010	150.0	202
2	51287	Leopoldstadt	48.217780	16.378470	73.0	370
3	70637	Leopoldstadt	48.217600	16.380180	50.0	116
4	75500	Brigittenau	48.234930	16.367520	85.0	12

SELECTION AND LOADING

- **Selected columns:** [id, longitude, latitude, price, number_of_reviews, neighbourhood]
- **Filter:** Drop data entries without price and review (outliers, bots)
- Replace neighbourhood names with district codes
- Classification according to price
  - Separation of dataset into 10 quantile datasets
  - Each dataset consists of ~1,000 entries

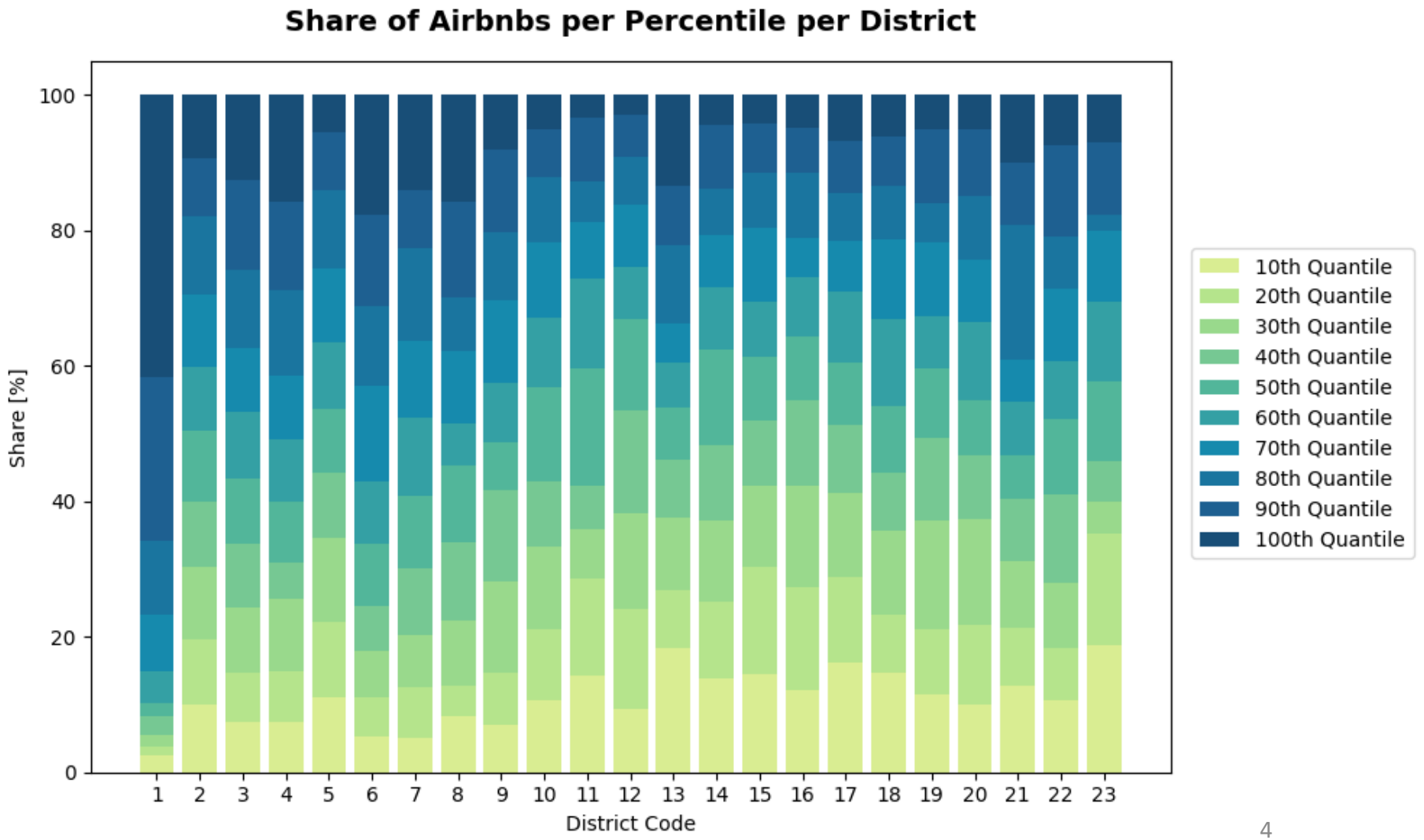
#4 ANALYSIS: SHARE PER QUANTILE PER DISTRICT

PROCESS DATA

1. Calculate absolute numbers of Airbnb's per district
2. Calculate percentual share per district

PROCESS DATA

- Most expensive Airbnb's in 1st and 6th
- Cheapest Airbnb's in 13th and 23rd



#5.1 ANALYSIS: CLUSTERING

PROCESS DATA

1. Transform quantile datasets to GeoJSON
2. Reproject GeoJSONs to EPSG: 32633 (units: m)
3. Load GeoJSON into GeoDataFrame
4. Transform ‘geometry’ column to separate ‘x’ and ‘y’ columns
5. Extract columns ‘x’ and ‘y’ into own DataFrames for each quantile

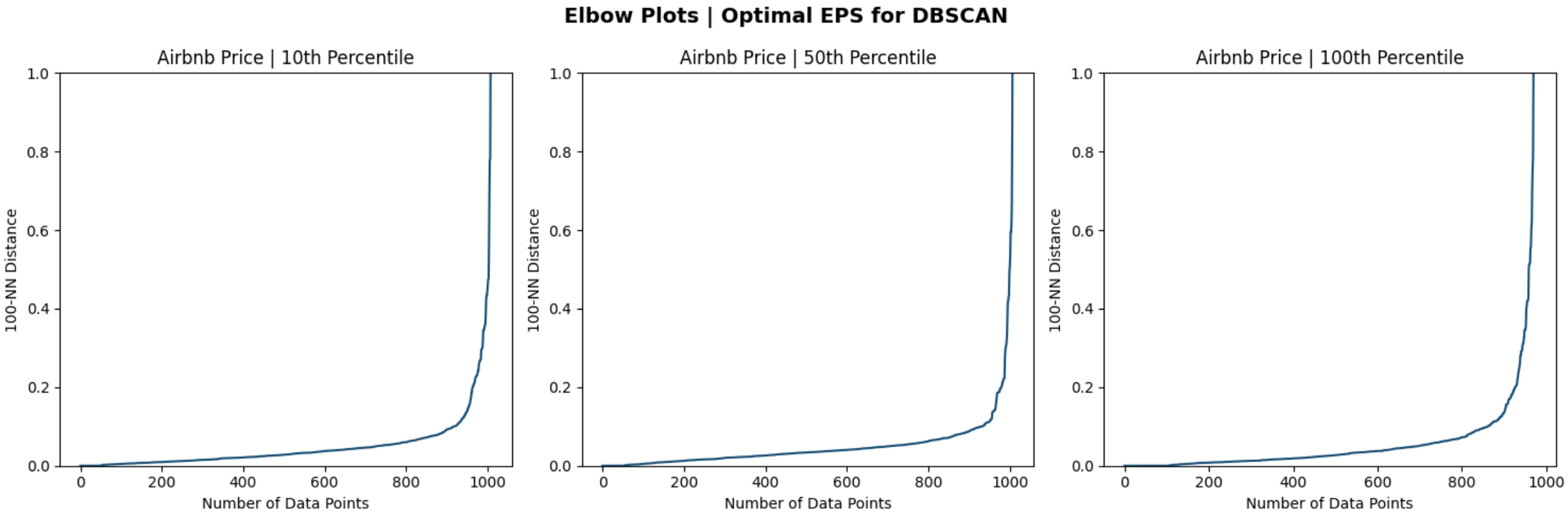
ELBOW PLOTS

- Identify best eps value for DBSCAN clustering

#5.2 ANALYSIS: CLUSTERING

ELBOW PLOTS

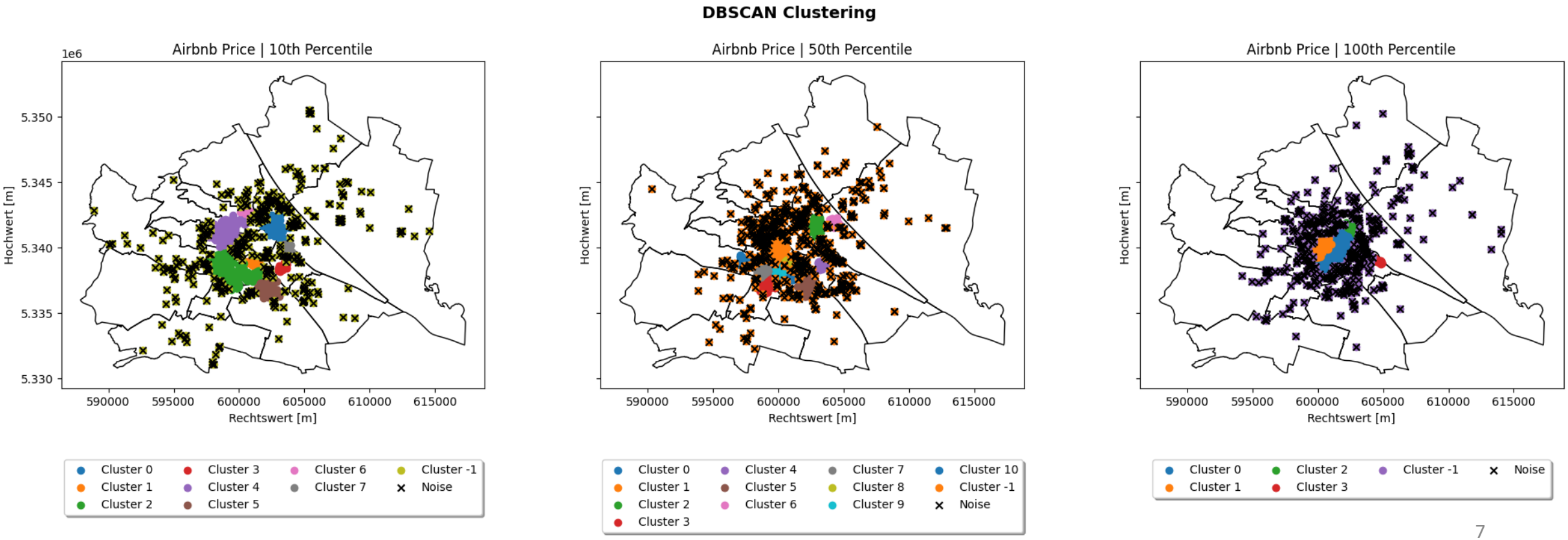
- Identify best eps value for DBSCAN clustering -> eps: 0.17, min_samples: 20



#5.3 ANALYSIS: CLUSTERING

DBSCAN

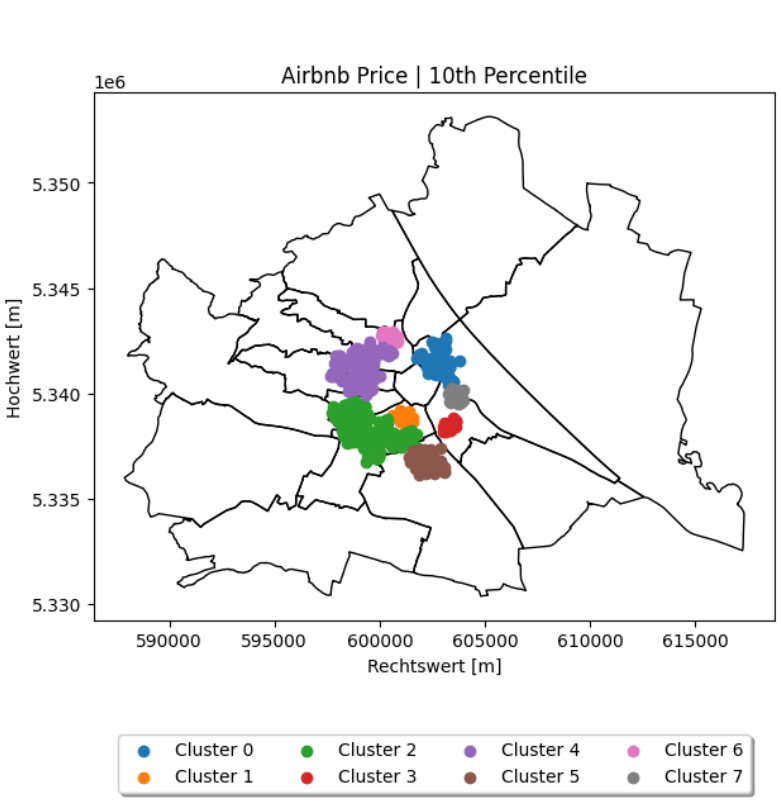
- DBSCAN-algorithm based on eps: 0.17 and min_samples: 20 for 10th, 50th, and 100th quantile

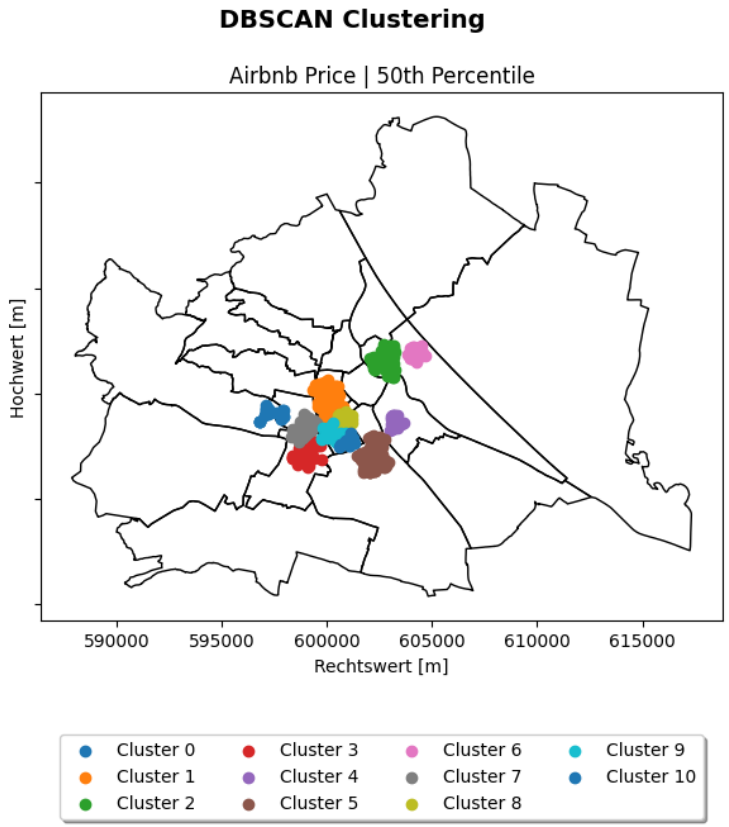


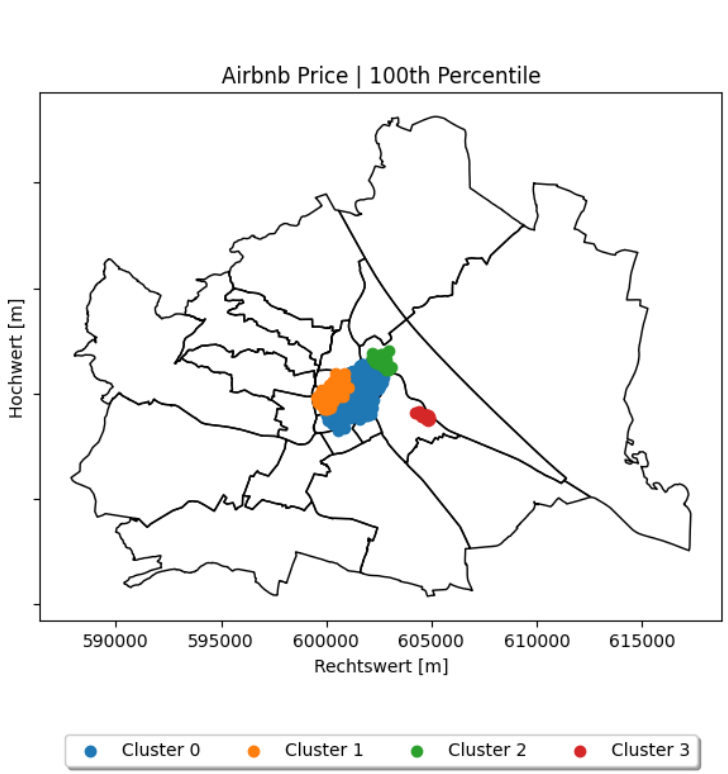
#5.4 ANALYSIS: CLUSTERING

DBSCAN

- DBSCAN-algorithm based on eps: 0.17 and min_samples: 20 for 10th, 50th, and 100th quantile







#6.1 ANALYSIS: HULL & DISTANCE CALUCLATION

PROCESS DATA

1. Extract clusters numbers per quantile dataset and write into list
2. Remove cluster == -1 (noise)
3. Extract points for each cluster and for each quantile
4. Calculate convex hull for cluster points
5. Transform hull into polygon
6. Calculate centroid of polygon
7. Calculate distance from centroid to Vienna center
8. Calculate mean distance across clusters for each quantile
9. Create multipolygons from polygons (for visualization purposes)

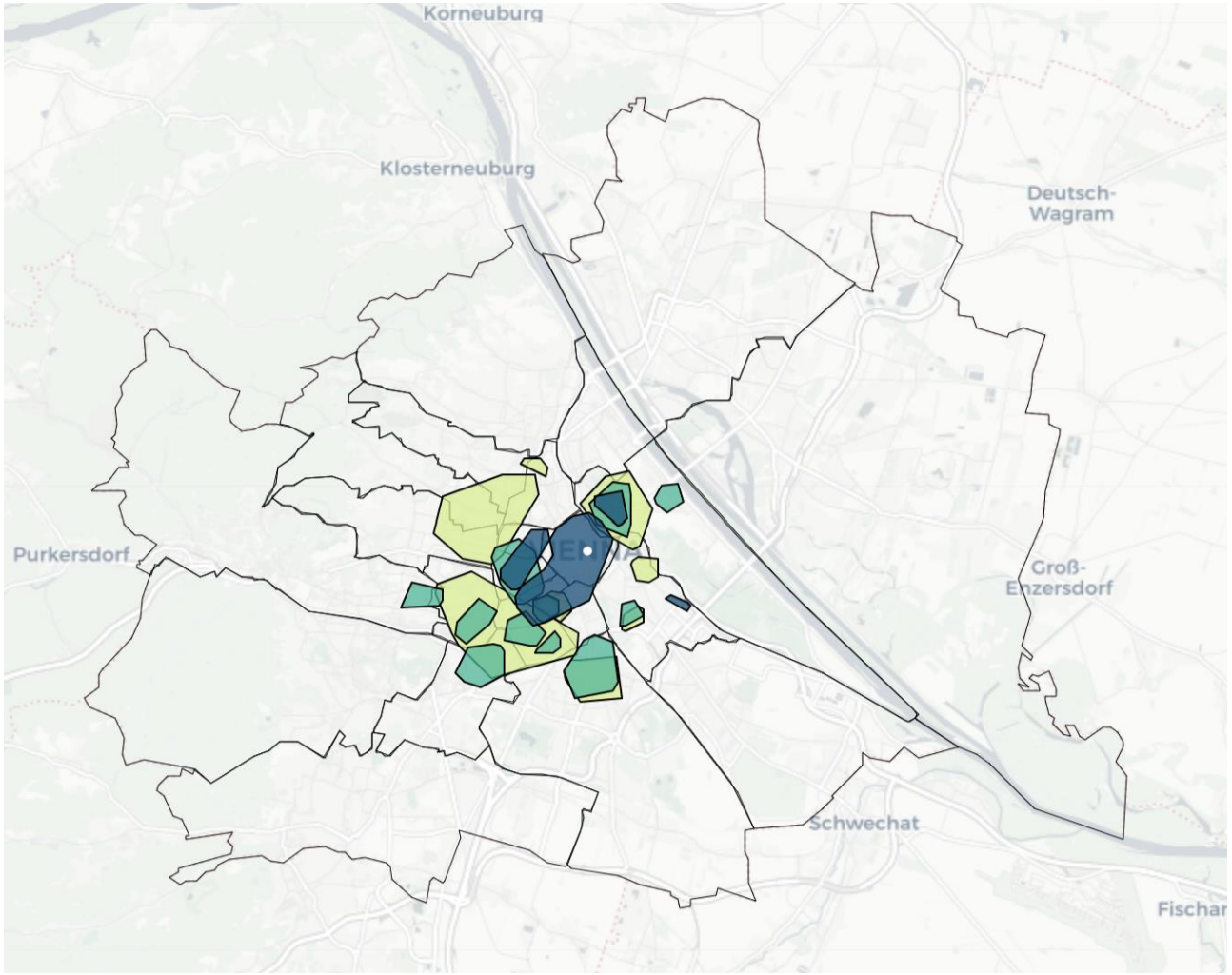
#6.2 ANALYSIS: HULL & DISTANCE CALUCLATION

RESULTS

- 10th quantile: 2.48 km to center
- 50th quantile: 2.92 km to center
- 100th quantile: 1.72 km to center

LEGEND

- Vienna center
- Vienna district boundaries
- 10th quantile
- 50th quantile
- 100th quantile

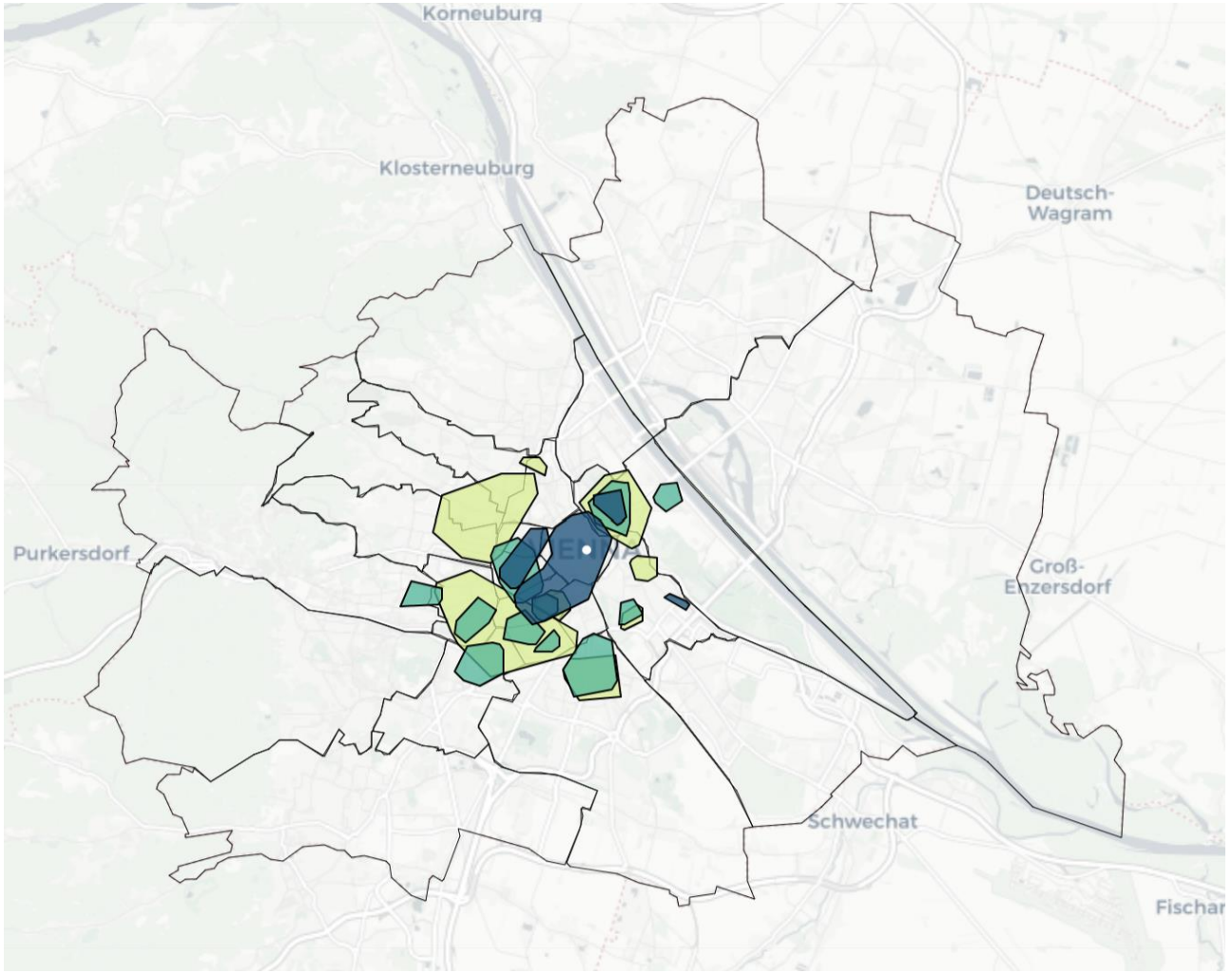


#7 SHORTCOMINGS & FUTURE WORK

- Carry out analysis for all quantiles
- Write code more flexible to accommodate for different datasets
- Discover effects of different clustering input parameters on results

LEGEND

- Vienna center
- Vienna district boundaries
- 10th quantile
- 50th quantile
- 100th quantile



#7 REFERENCES

- **Inside Airbnb** (2023). Listings.csv. Summary information and metrics for listing in Vienna (good for visualizations). Web. Available online at: http://data.insideairbnb.com/austria/vienna/vienna/2023-12-15/visualisations/listings.csv
- **Stadt Wien** – Magistratsabteilung 41, Stadtvermessung (2023). Bezirksgrenzen Wien. Web. Available online at: https://data.wien.gv.at/daten/geo?version=1.1.0&service=WFS&request=GetCapabilities

THANK YOU FOR YOUR ATTENTION

ANY QUESTIONS?

JASMIN KOFLER & BENEDIKT MÜLLER

UNIV.-PROF. DR. KRZYSZTOF JANOWICZ
PS Spatial Data Science
University of Vienna