E-Commerce: High Potential Regions

Finding regions within Germany that have a high population density and a low retail store density by utilizing Web Scraping, Geocoding and the Foursquare API

Background & Relevance

Problem:

- COVID-19 has led to a surge in e-commerce and accelerated digital transformation
- BUT: the reopening of physical stores will probably have a slowing effect on the growth

Solution:

- finding regions within Germany that have a high population density as well as a low retail density for uplifting ROI of marketing activities

Data Collection

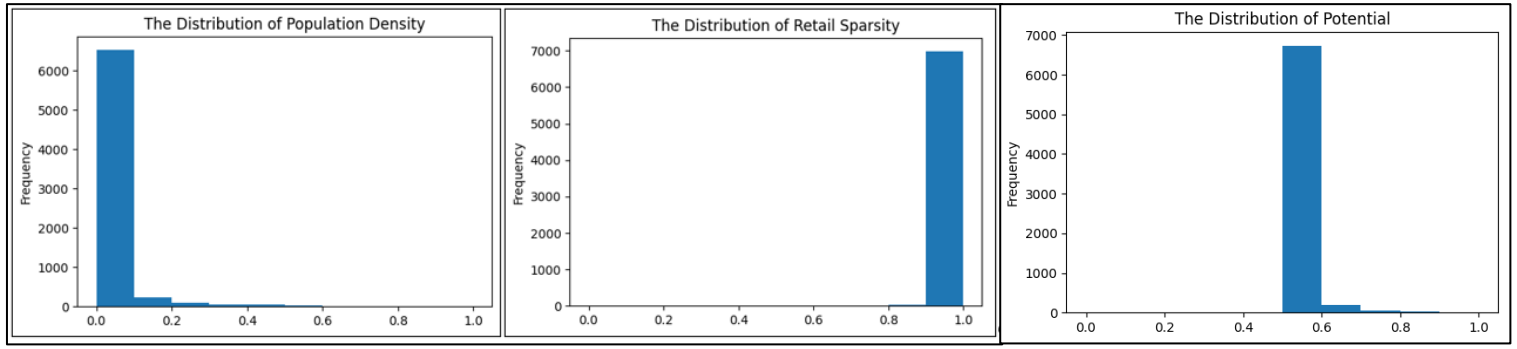
- 4 different sources were used:
  - Web Scraping: Zipcode Locations: https://home.meinestadt.de
  - Web Scraping: Area & Population: https://de.wikipedia.org/
  - Geocoding: Coordinates (lat/lon): Geocoder Python Package
  - API: Venue Names, Categories, Distance: Foursquare

KPI Engineering

- **population density**:
  population (Wikipedia) divided by area (Wikipedia)
- **retail density**:
  shop count (Foursquare) divided by area (Wikipedia). The complement of retail density (**retail sparsity** = 1 – retail density) is used to compute the potential
- **potential:**
  combination of the previous two variables:
  1/2 * (normalized(**population density**) + (normalized(**retail sparsity**))

Results I

- Distribution of the KPIs:



- Population density and retail sparsity are highly skewed!

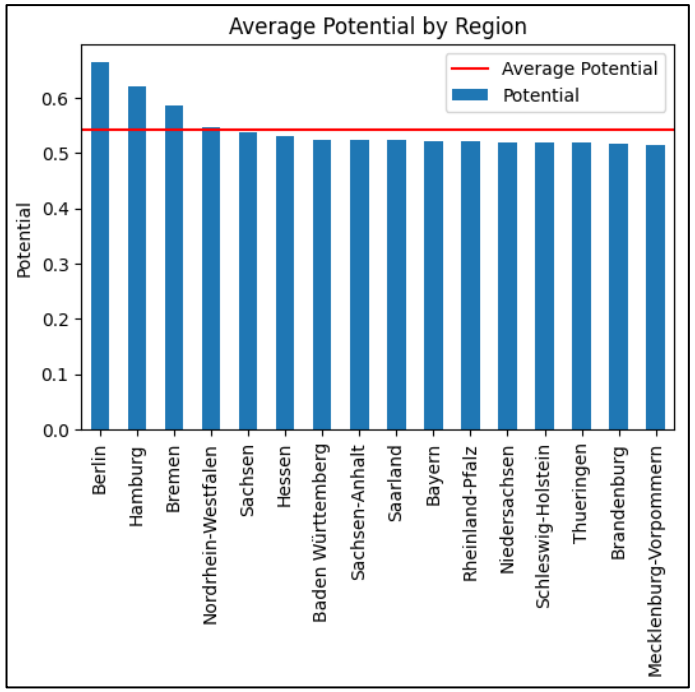
Results II

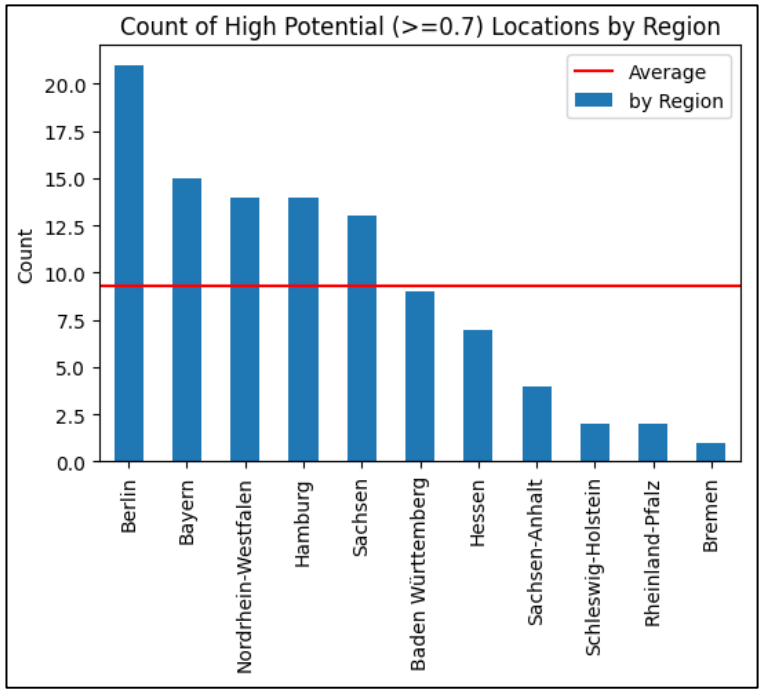
- Top 20 high Potential Locations:

	zipcode	city	quarter	region	learn_more	population_density	retail_sparsity	potential
0	20253	Hamburg	Hamburg-Hoheluft-West	Hamburg	/wiki/Hamburg-Hoheluft-West	1.00	0.86	1.00
1	10369	Berlin	Berlin-Fennpfuhl	Berlin	/wiki/Berlin-Fennpfuhl	0.83	0.99	0.98
2	70182	Stuttgart	Stuttgart-Heusteigviertel	Baden Württemberg	/wiki/Heusteigviertel	0.94	0.87	0.97
3	13357	Berlin	Berlin-Gesundbrunnen	Berlin	/wiki/Berlin-Gesundbrunnen	0.80	0.97	0.95
4	20259	Hamburg	Hamburg-Eimsbüttel	Hamburg	/wiki/Hamburg-Eimsb%C3%BCttel	0.94	0.83	0.95
5	22049	Hamburg	Hamburg-Dulsberg	Hamburg	/wiki/Hamburg-Dulsberg	0.75	0.99	0.93
6	12043	Berlin	Berlin-Neukölln	Berlin	/wiki/Berlin-Neuk%C3%B6lln	0.73	0.99	0.93
7	12353	Berlin	Berlin-Gropiusstadt	Berlin	/wiki/Berlin-Gropiusstadt	0.74	0.98	0.92
8	10969	Berlin	Berlin-Kreuzberg	Berlin	/wiki/Berlin-Kreuzberg	0.77	0.95	0.92
9	81675	München	München-Haidhausen	Bayern	/wiki/Au-Haidhausen	0.76	0.94	0.91
10	81541	München	München-Au	Bayern	/wiki/Au-Haidhausen	0.76	0.94	0.91
11	10405	Berlin	Berlin-Prenzlauer Berg	Berlin	/wiki/Berlin-Prenzlauer_Berg	0.78	0.90	0.90
12	79100	Freiburg im Breisgau	Freiburg im Breisgau-Vauban	Baden Württemberg	/wiki/Vauban_(Freiburg_im_Breisgau)	0.71	0.97	0.90
13	50677	Köln	Köln-Neustadt/Süd	Nordrhein-Westfalen	/wiki/Neustadt-S%C3%BCd_(K%C3%B6ln)	0.71	0.97	0.89
14	22089	Hamburg	Hamburg-Eilbek	Hamburg	/wiki/Hamburg-Eilbek	0.68	0.98	0.89
15	13439	Berlin	Berlin-Märkisches Viertel	Berlin	/wiki/Berlin-M%C3%A4rkisches_Viertel	0.65	0.99	0.88
16	80339	München	München-Schwanthalerhöhe	Bayern	/wiki/Schwanthalerh%C3%B6he	0.75	0.89	0.87
17	10243	Berlin	Berlin-Friedrichshain	Berlin	/wiki/Berlin-Friedrichshain	0.73	0.91	0.87
18	04315	Leipzig	Leipzig-Volkmarsdorf	Sachsen	/wiki/Volkmarsdorf	0.63	1.00	0.87
19	10711	Berlin	Berlin-Halensee	Berlin	/wiki/Berlin-Halensee	0.64	0.98	0.86

Results III

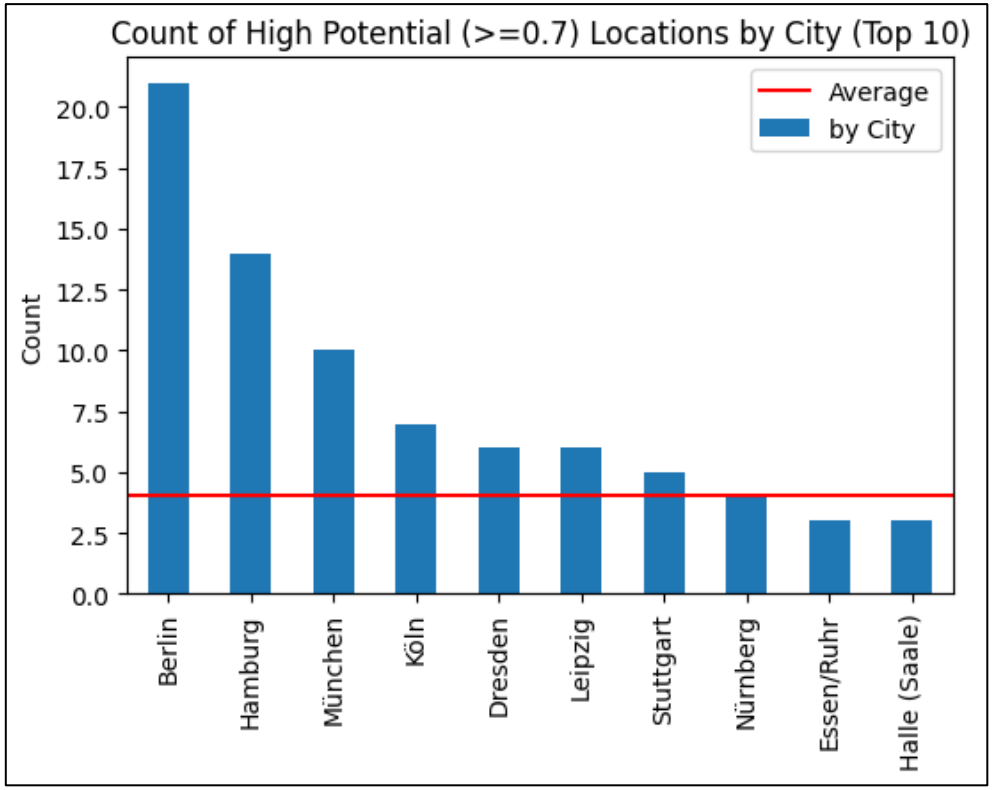
- Deepdive on the Potential of Regions:





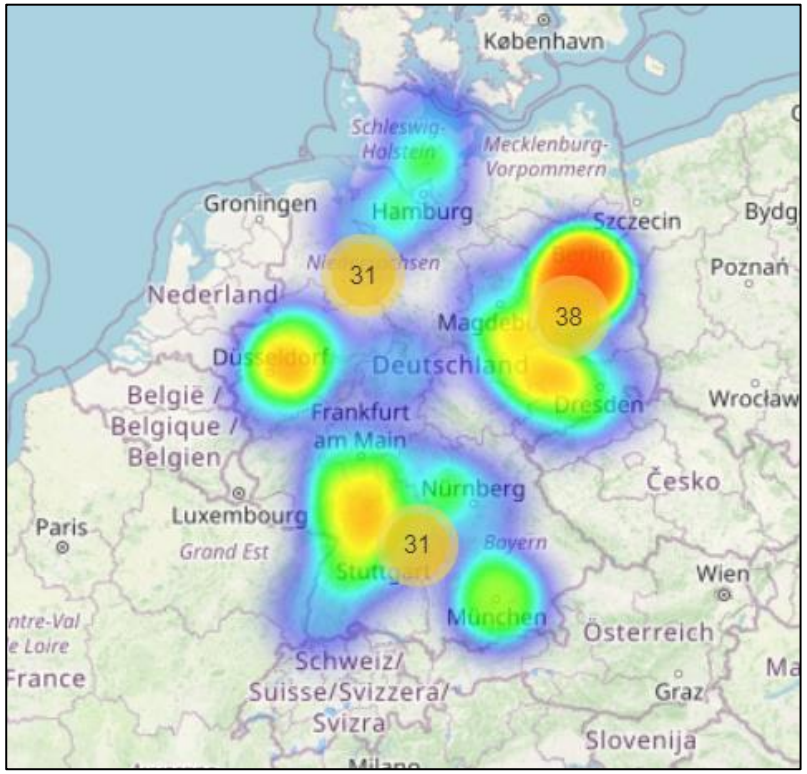
Results IV

- Deepdive on the Potential of Cities:



Results V

- Deepdive on the Geographical Clusters:



- **South (31%)**:
  The south extends from Freiburg and Munich up to Nuremberg and Frankfurt with Munich (10), Stuttgart (5) and Nuremberg (4) in the lead.
- **East (38%)**:
  The east extends from Dresden and Halle (Saale) up to Berlin. Berlin leads the ranking with 21 high potential areas, followed by Dresden (6) and Leipzig (6).
- **West & North (31%)**:
  The West & North extends from Cologne and Kassel up to Kiel. Hamburg (14) has the most high potential locations, followed by Cologne (7).

Discussion I

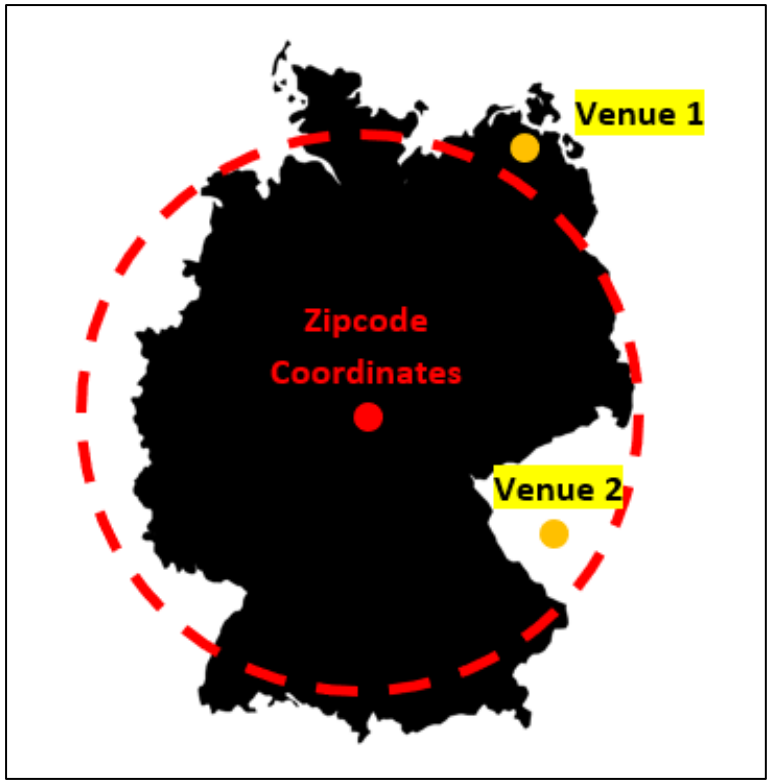
- Scraping of Wikipedia result pages is highly volatile:

search_string	wiki_title	wiki_url	wiki_area_sqkm	wiki_population	latitude	longitude
40210 Düsseldorf-Stadtmitte	Düsseldorf	/wiki/D%C3%BCsseldorf	217.41	620523.0	51.221939	6.784423
40213 Düsseldorf-Carlstadt	Düsseldorf	/wiki/D%C3%BCsseldorf	217.41	620523.0	51.222142	6.773394
40629 Düsseldorf-Ludenberg	Düsseldorf	/wiki/D%C3%BCsseldorf	217.41	620523.0	51.256357	6.866151

- Source of the skewness of retail sparsity:

	wiki_title	wiki_area_sqkm	venue_count	retail_density_per_sqkm	retail_density_per_sqkm_norm
3182	Hamburg-Altstadt	1.2	85	70.833331	1.000000
3192	Hamburg-Sternschanze	0.6	39	64.999997	0.917647

Discussion II



- the zipcode locations / regions are not perfect circles
- The radius is therefore a best guess with the same area size as the location, but not the same form.
- Example on the left:
  Venue 1 is not within the radius but within the location and Venue 2 is within the radius but not within the location

Conclusion & Outlook

- results show clearly that there is regional variance in potential and where E-Commerce companies can concentrate their activities to open up those areas
- Future research could use different venue category codes in the Foursquare API URI and compute the KPIs with a mixed approach (Normalization / Standardization)
- The biggest possibility is the national census 2022 in Germany. This data will include area and population demographically split by age and gender so target groups could be directly addressed and potential refined