Up-and-coming: Venue location scouting in Berlin

A Data Science Project Presentation by Frederik Laubisch

Problem

Berlin has many well known popular neighbourhoods with many cafes, bars and restaurants. Opening a new business here can be problematic. It would be beneficial to know which less popular areas make a good location for such business.

The goal is to identify these up-and-coming neighbourhoods.

Data

A list of Berlin neighbourhoods was pulled from Wikipedia

The FourSquare API was then utilized to get a list of venues for each of the neighborhoods

Only service business venues were retained in dataset

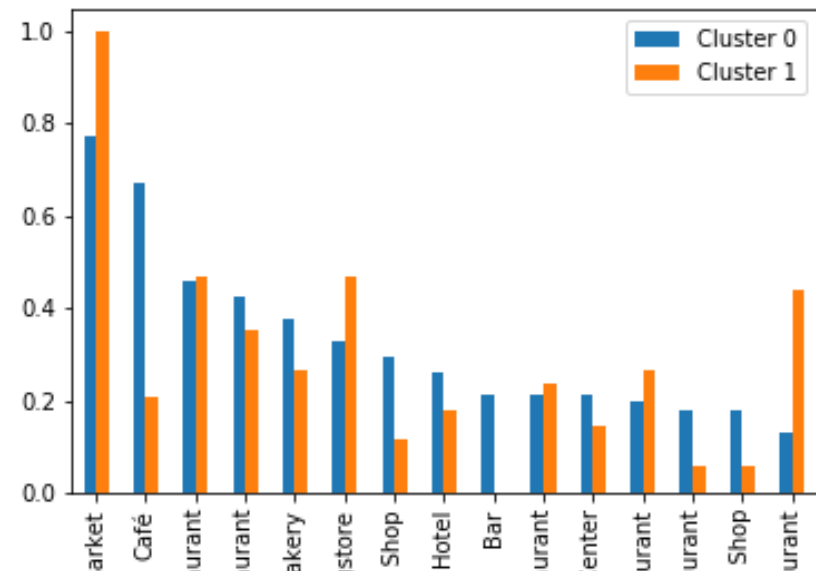Relative frequency of each category was calculated for each neighborhood

Density of venues per 1000 inhabitants was calculated

Methods: Clustering

Data was clustered using k-means algorithm with k=3 determined to be optimal number of clusters.
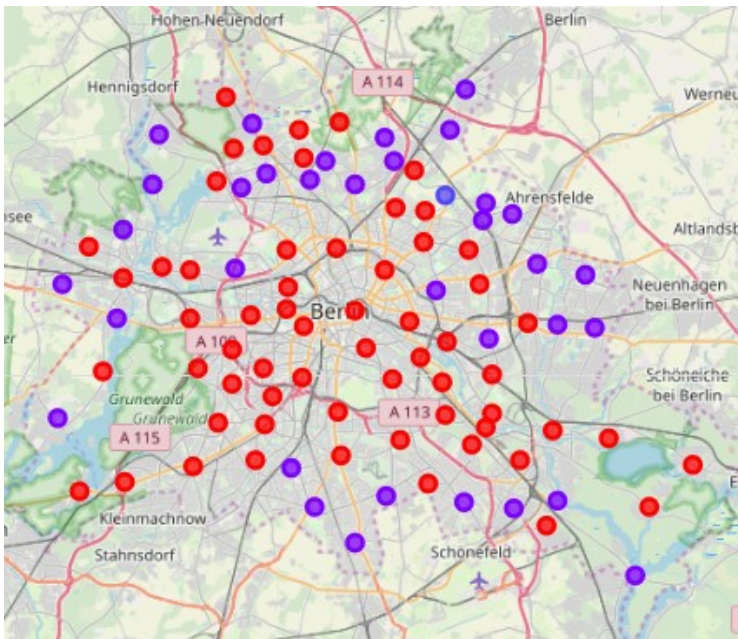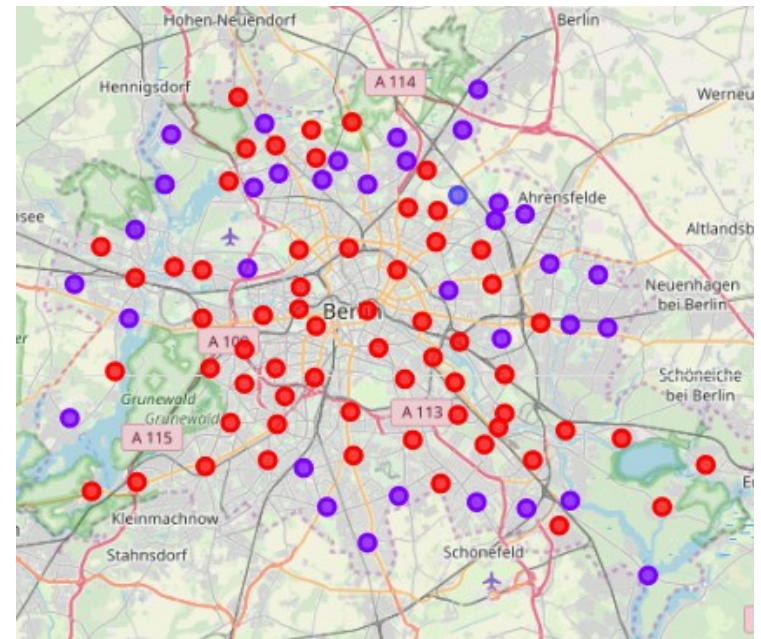

Map of Clusters

Methods: Clustering

Data was clustered using k-means algorithm with k=3 determined to be optimal number of clusters.

One cluster only contained 1 Neighborhood
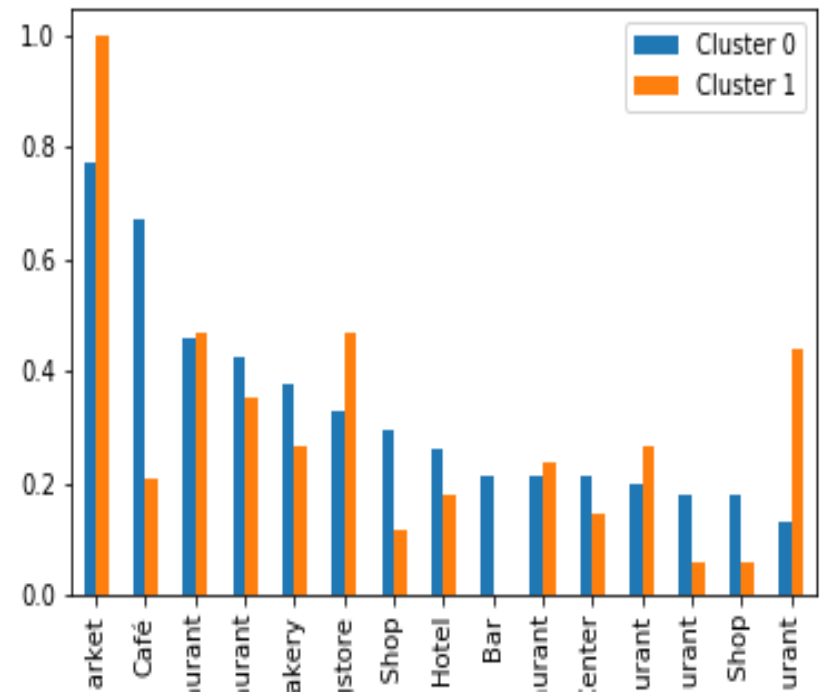
Two distinct clusters were cal-culated.

Map of Clusters

Methods: Clustering

One cluster consists of outskirt residential areas with many supermarkets, drug-stores and fast-food restaurants.

Other cluster has many bars, Cafes and coffee shops. These are of interest, and contain the popular areas.

Results

A list of neighborhoods that are similar to the popular neighbour-hoods in their venue makeup, but less saturated (lower venue density) can be considered recommended locations for new venues.

	Neighbourhood	Population	N_venues	Venue Density
85	Reinickendorf	83.447	25	0.299591
94	Märkisches Viertel	40.258	23	0.571315
54	Gropiusstadt	37.450	24	0.640854
71	Biesdorf	27.723	18	0.649280
77	Lichtenberg	41.112	36	0.875657
47	Mariendorf	52.954	48	0.906447
82	Alt-Hohenschönhausen	48.728	45	0.923494
9	Weißensee	53.737	50	0.930458
76	Karlshorst	27.522	27	0.981033
51	Britz	42.640	42	0.984991

Discussion

In order to get a better understanding of which of these areas are really up-and-coming, it would be a good idea to look at historical data.

A predictive model could be constructed to look at the growth of these neighborhoods and predict which one are likely to be the next big thing.

Conclusion

We were able to provide a list of potential neighborhoods that could be up-and-coming and are likely candidates for investment.

However more data is needed to construct more complex models to get a better idea of which neighborhoods are really going to be the next big thing.