
Finding the Best Location to Open a Burger Joint in Berlin
IBM Data Science Course Capstone Project
2020/01/08
by Jonas Folesky
© OpenStreetMap contributors

Introduction

- Objective: Findig the best location for a burger joint in Berlin Germany
- Target audience: entrepreneurs, investors

Introduction

- What factors are important to define a good or bad location ?
  - Demand - > Customers
    - Customers can be residents, tourists, students, labourers, office employees and many more
  - Supply - > Competition
    - Competition is given by already existing food facilities i.e. other burger places

Introduction

- What data is availlable to answer the target question ?
  - Public data from the City of Berlin will provide information on demography
  - FoursquareAPI will provide data on selected stores/resaurants/instances to include in the analysis

Introduction

- What methodology will be used ?
  - Data will be sorted, cleaned and prepared for analysis first
  - Data will be visualized to find ituitive solutions to our question and to identify data dependencies
  - Based on the viszualization the machine lernaing algorithm kmeans++ is identified as suitable to solve our problem

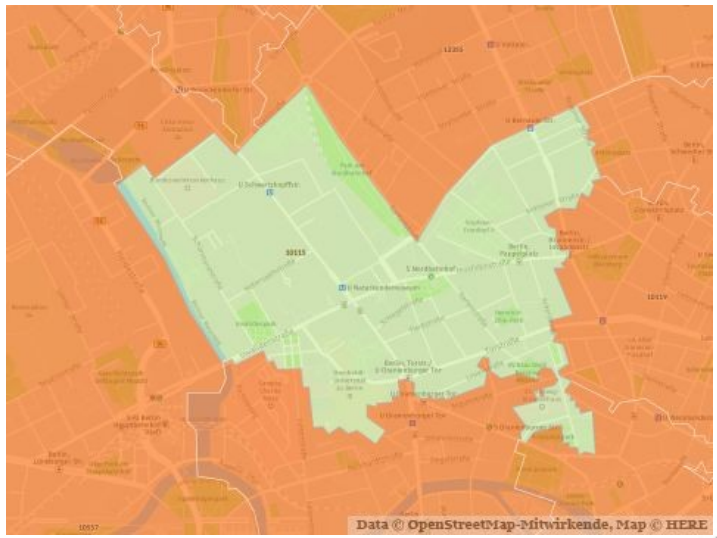
Data

- Demographic data is obtained from German census 2011 provided by the web page www.suche-postleitzahl.org
- Same page provides zip code boundary polygons

Citizens per zip code

zip	citizens
10115	20313
10117	12217
10119	16363
10178	12167
10179	18664
10243	...

Zip code district Berlin 10115



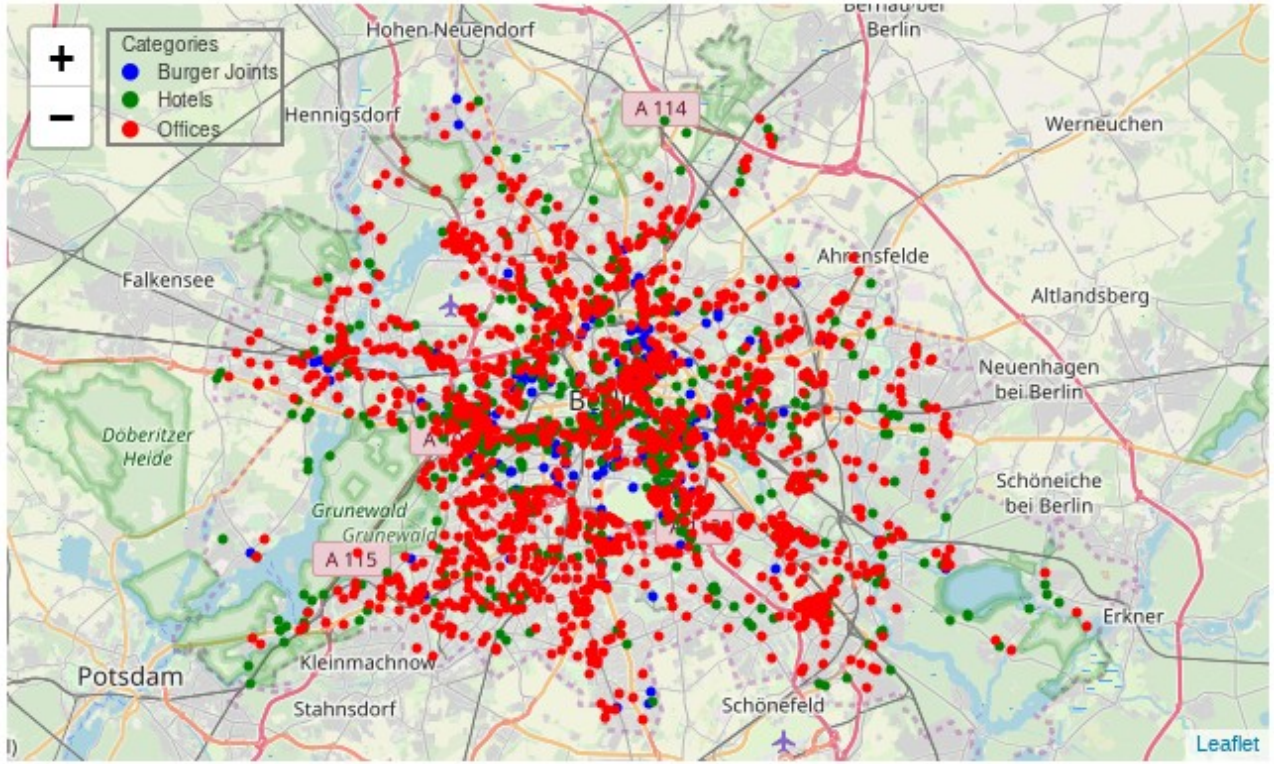
Data

- FoursquareAPI is used to gather data on : competitors and costumers

burger places hotels and offices

	name	lat	lon	zip
0	Kladow Grill Burger Pizza	52.453026	13.141852	14089
1	Volcano Burger	52.525032	13.196861	13595
2	Cruise-In	52.532584	13.178774	13581
3	Bastis Currys, Burgers & Fries	52.533400	13.190330	13581
4	Burger Route	52.530632	13.196140	13581

Purger places obtained from FoursquareAPI

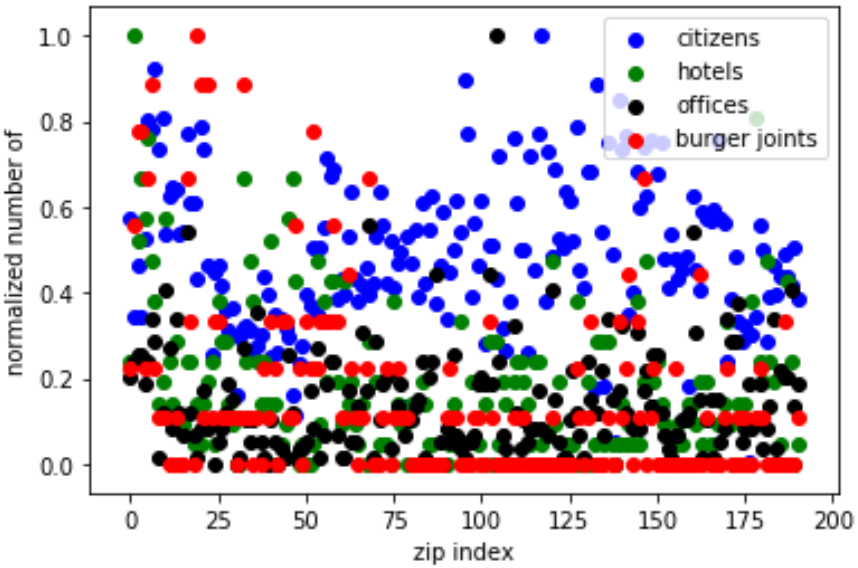


Map of Berlin for all 3 venue types obtained from FoursquareAPI

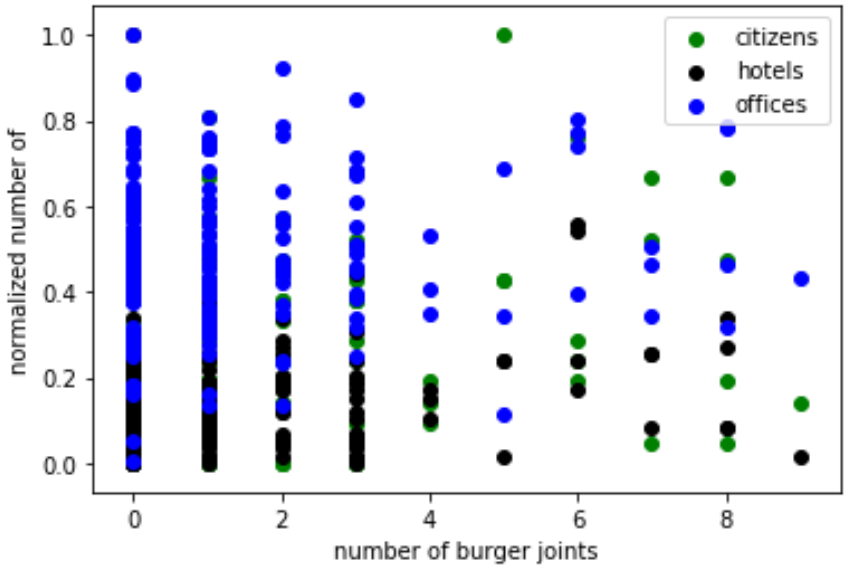
Methodology

- The data is examined to find trends and dependencies

Citizens, Hotels, Offices, Burger joints vs. Zip code index



Citizens, Hotels, Offices vs. Burger joints

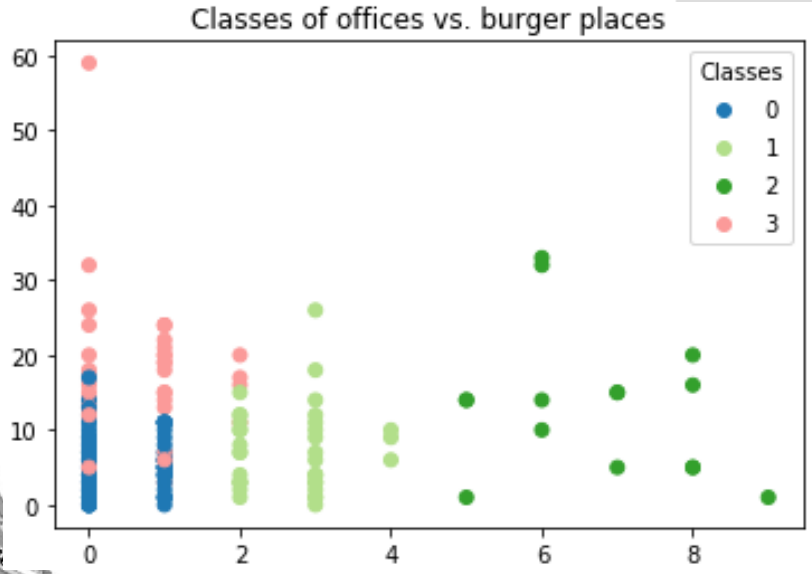


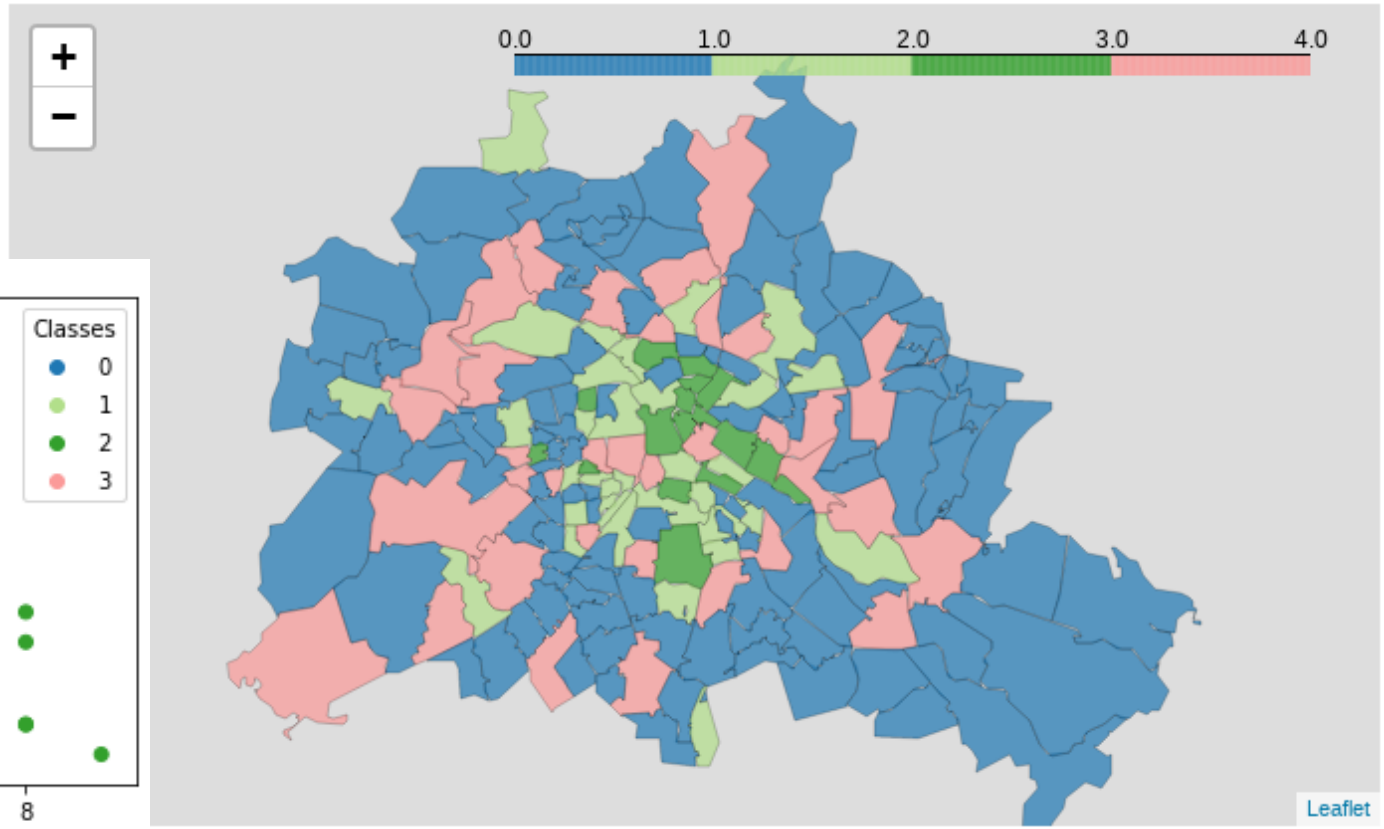
- No obvious trend identified
- kmeans++ clustering will provide good data separation

Results

- Data is separated into 4 clusters based on 4 data categories

Clustering results



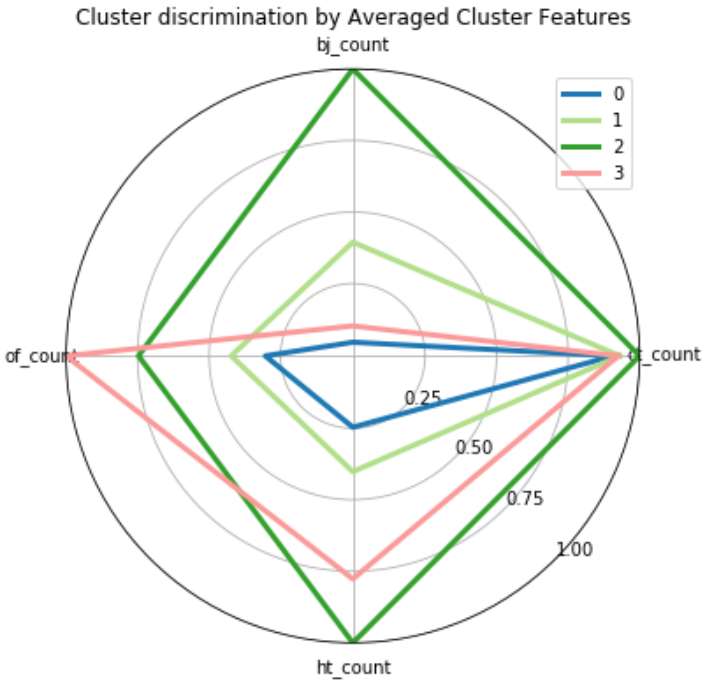


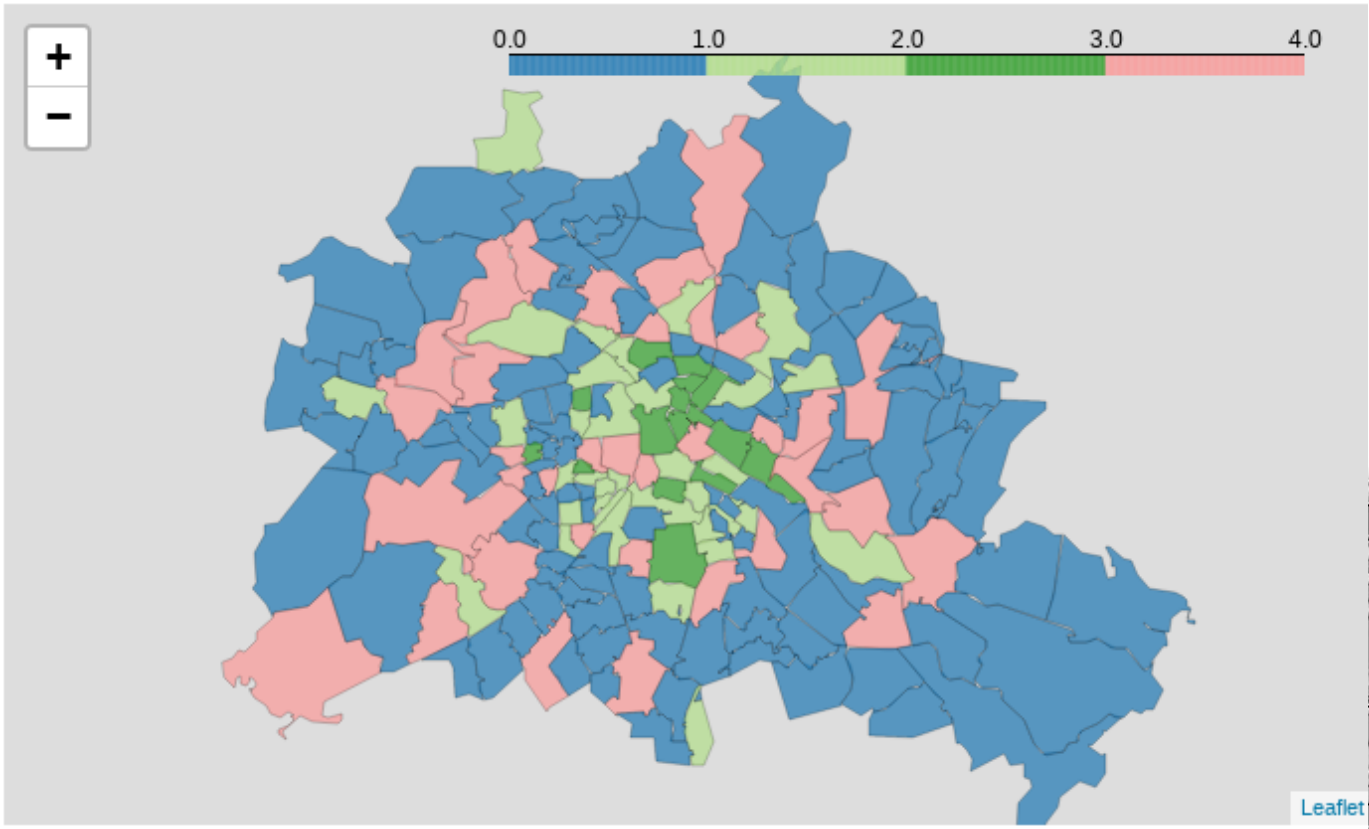
Clustering results projected onto Berlin map

Results

- Data clusters must be ranked to give the clusterization some meaning
- It is desireable to have many customers availlable and little competition
- Based on the category plot the ranking of clusters is [3,0,1/2]

Average Clusters Features





Clustering results projected onto Berlin map

by Jonas Folesky

Conclusion

- We identified a cluster of zip code areas with best features for opening a new burger place -> Cluster 3. It shows little competition but many propable customers
- The analysis was based on simple proxies for demand and supply -> cutomers and cokpetition -> citizens, hotels, offices and burger joints
- The analysis can easily be expandet with mire categories and weighting between categories can be varied