Berlin – The Pulse of a City

Tobias Gerken

April 28, 2019

1 Introduction

According to U.S. Small Business Association [1], approximately 50% of new businesses fail within five years of opening. Given the high risk and challenging retail environment, finding the right target demographic and business environment is therefore key. At the same time, cities are becoming increasingly dynamic and diverse which provides ample opportunities for business segmentation.

Combining data-science approaches with the increasing availability of real time consumer data and open data from city governments can provide valuable insight that can businesses help attract and retain customers.

1.1 Business question

Berlin, the capital of Germany, is home to approximately 3.6 million people and widely recognized to be one of Europe's most dynamic cities. Using the city of Berlin as example, we segment its districts by their population demographics and analyze what type of businesses thrive is the respective neighborhoods. The *trending* feature of the Foursquare API, we analyze locations and neighborhoods in Berlin that are trending within a 24-h window letting us experience the *Pulse of the City*.

In detail we ask:

- Which neighborhoods have the highest density of trending venues?

- Does the age distribution of district residents have a discernible impact on venues density and type?

- Can we segment neighborhoods based on the type of business venues they attract and is there a relationship to the demographics Berlin's districts?

2 Data

Berlin is a city state divided into 12 districts, each with populations between 200,000 and 400,000 (Figure1).

The Berlin state government provides demographics data for Berlin through its Open Data Portal [2]. Geographic shape files of the Berlin districts [3] and postcode [4] areas are available through the *Technologie Stiftung Berlin* (Berlin Technology Foundation).
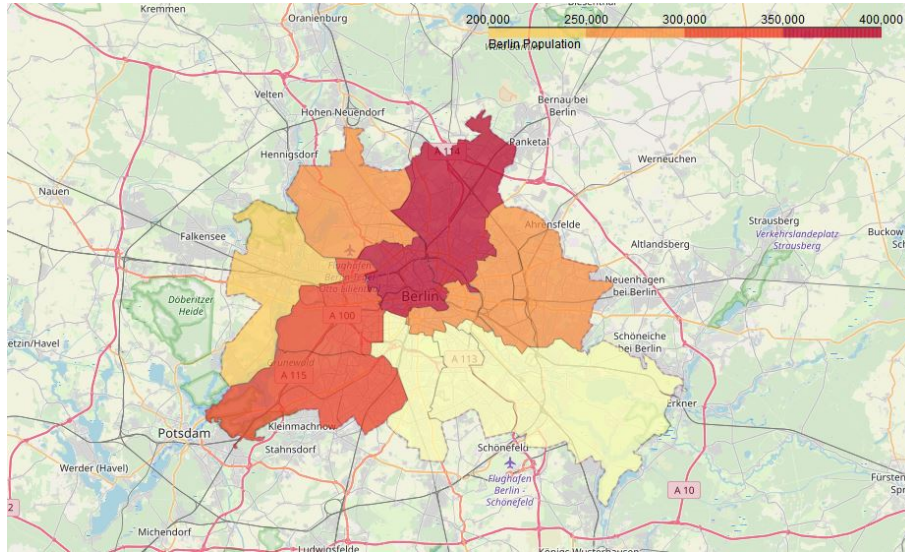

Figure 1: Berlin is home to 3.6 million people, who live in 12 districts with 200,000 to 400,000 inhabitants each.

Foursquare's API [5] allows to query business venue data based on geographical coordinates. Using the *explore* functionality, we receive information about nearby venues, including their postcode and geographic coordinates, which allow them to be linked to Berlin's demographic data. The *trending venues* functionality, allows us to find types of business venues and associated geographic areas that are popular during the course of a day.

2.1 Approach

Demographic data at the city neighborhood-level (*Ortsteil*) were downloaded in CSV-format and aggregated to district level. The demographic data is provided in age groups of 5 years. We suppose that similar age groups have similar consuption patterns and aggregated the data to age groups 0-15, 15-30, 30-50, 60-65, and over 60 years old (Table 1), approximately representing children, young adult, middle age, older, and senior consumers.

To capture variation in business venues, we execute a gridded sampling approach (Figure 2), where Berlin is covered with an equidistant grid of 2 km length for which Foursquare search queries are being executed. The returned venues are aggregated by district and postcode area, retaining information about the nature of the business.

Table 1: There is considerable diversity in age group distribution between Berlin's districts. For example Friedrichshain-Kreuzberg has the larges share of middle age consumers, while Steglitz-Zehlendorf has a much older population.

District	0-15	15-30	30-50	50-65	>65
Charlottenburg-Wilmersdorf	0.11	0.17	0.29	0.23	0.24
Friedrichshain-Kreuzberg	0.14	0.23	0.42	0.16	0.11
Lichtenberg	0.14	0.19	0.3	0.21	0.21
Marzahn-Hellersdorf	0.15	0.17	0.27	0.27	0.2
Mitte	0.14	0.25	0.35	0.18	0.14
Neukölln	0.14	0.21	0.32	0.19	0.19
Pankow	0.16	0.17	0.39	0.19	0.16
Reinickendorf	0.14	0.18	0.26	0.22	0.25
Spandau	0.15	0.19	0.27	0.22	0.23
Steglitz-Zehlendorf	0.13	0.16	0.26	0.22	0.27
Tempelhof-Schöneberg	0.13	0.18	0.29	0.22	0.22
Treptow-Köpenick	0.13	0.16	0.29	0.22	0.24

References

[1] U.S. Small Business Association: Do economic or industry factors affect business survival? 2012
https://www.sba.gov/sites/default/files/Business-Survival.pdf,

[2] Amt für Statistik Berlin-Brandenburg: Einwohnerinnen und Einwohner in den Ortsteilen Berlins am 30.06.2016, 2016
https://daten.berlin.de/datensaetze/einwohnerinnen-und-einwohner-den-ortsteilen-berlins-am-30062016,

[3] Technologiestiftung Berlin: Die Bezirksgrenzen der 12 Berliner Bezirke, 2017
https://data.technologiestiftung-berlin.de/dataset/bezirksgrenzen

[4] Technologiestiftung Berlin: PLZ – Postleitzahlgebiete Berlins, 2015
https://data.technologiestiftung-berlin.de/dataset/plz

[5] Foursquare Labs Inc: Places API, https://https://developer.foursquare.com/docs, Last Access: 4/28/2019
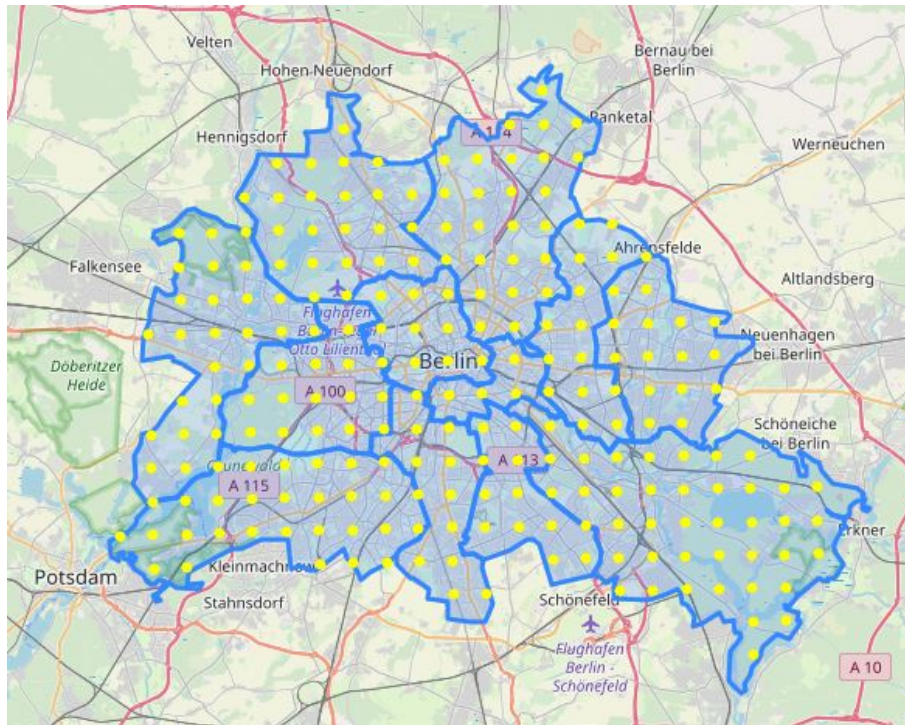
Figure 2: Sampling grid used for execution of the Foursquare venue explore functionality