

# Berlin venues and the spread of the coronavirus

Are there certain types of urban structures that increase the risk of infection?

Kristin Lehnert

December 2020

## Introduction

The corona virus has kept the world on edge all the past year. Many are currently hoping that the recent approval of the first corona vaccine in the EU will bring a rapid improvement. However, at this moment is still unclear how quickly it can be delivered, who will be supplied first and how many Germans will actually be willing to be vaccinated in the end. The persistent uncertainty has made many people cautious<sup>1</sup>. In the meantime, the numbers continue to rise in Berlin and everywhere else.

The city of Berlin in particular has been the subject of much criticism from the German government in recent months<sup>2</sup>. The beginning of the Christmas season was doused in some districts on the streets with mulled wine. Even if stores and restaurants had closed, people met at their old usual places – especially places that are known to be busy from “normal” times and that are popular with young people for their rich leisure offer. The districts of Neukölln and Mitte were the first to see case numbers exceed the critical range<sup>3</sup>.

But is there a particular reason why the virus is spreading faster in certain districts than in others? Is there even such a thing as a “risky infrastructure”? And if so, are these just a few or various places in the city? In this investigation, Berlin districts were clustered based on their venue infrastructure and placed in the context of Covid case numbers. The aim was to analyze whether a certain constellation of places where people come together may have an influence on the infection figures in this region.

The problem was studied as part of the capstone project for the IBM Professional Certificate in Data Science.

---

<sup>1</sup> <https://www.aerzteblatt.de/nachrichten/118522/Laut-Umfrage-wollen-weniger-Menschen-definitiv-Impfung-gegen-Corona>

<sup>2</sup> <https://www.welt.de/vermishtes/article217032028/Berlin-voller-Corona-Hotspots-was-ist-nur-los-in-der-Hauptstadt.html>

<sup>3</sup> <https://www.tagesspiegel.de/berlin/deutsche-corona-hotspots-vier-berliner-bezirke-gelten-inzwischen-als-risikogebiete/26239218.html>

## Data

To approach the problem, the following data sources were consulted, pre-processed and analyzed.

*Table 1 overview of data and sources*

List of Berlin districts and boroughs	Wikipedia	The table was scraped from <a href="https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins">https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins</a> using an html-parser for conversion into a pandas data frame
Geospatial data of all districts	geopy	This data was obtained from the free Open Streetmap-API <a href="https://geocoder.readthedocs.io/api.html#forward-geocoding">https://geocoder.readthedocs.io/api.html#forward-geocoding</a>
List of venues in each district	Foursquare	A json file with all district's venues was taken from Foursquare, a proven independent location data platform <a href="https://developer.foursquare.com/">https://developer.foursquare.com/</a>
Current covid infection figures	Berlin Open Data	The current Corona infection figures were retrieved from the official capital portal of the city of Berlin <a href="https://daten.berlin.de/datensaetze/covid-19-berlin-verteilung-den-bezirken">https://daten.berlin.de/datensaetze/covid-19-berlin-verteilung-den-bezirken</a>