

## **Description of the problem**

So everyone knows that in the last few month we are having a really big struggle with COVID-19. A lot of people got sick, the rest of the world should stay at home on isolation, a lot of businesses shut down, economy is falling down and etc.

This virus has a really big influence on us. So why don't we make a little analysis? Let's say we analyze boroughs of such a big city as Berlin and see which of these boroughs suffered influence with COVID-19 the most and using the 4square will try to see if there any pattern.

Would be better to analyze the trending places with 4square and see the pattern of infections but we can't as due to COVID-19 they are not such places right now and so they cannot be our project topic.

## **Description of the data and how it will be used to solve the problem**

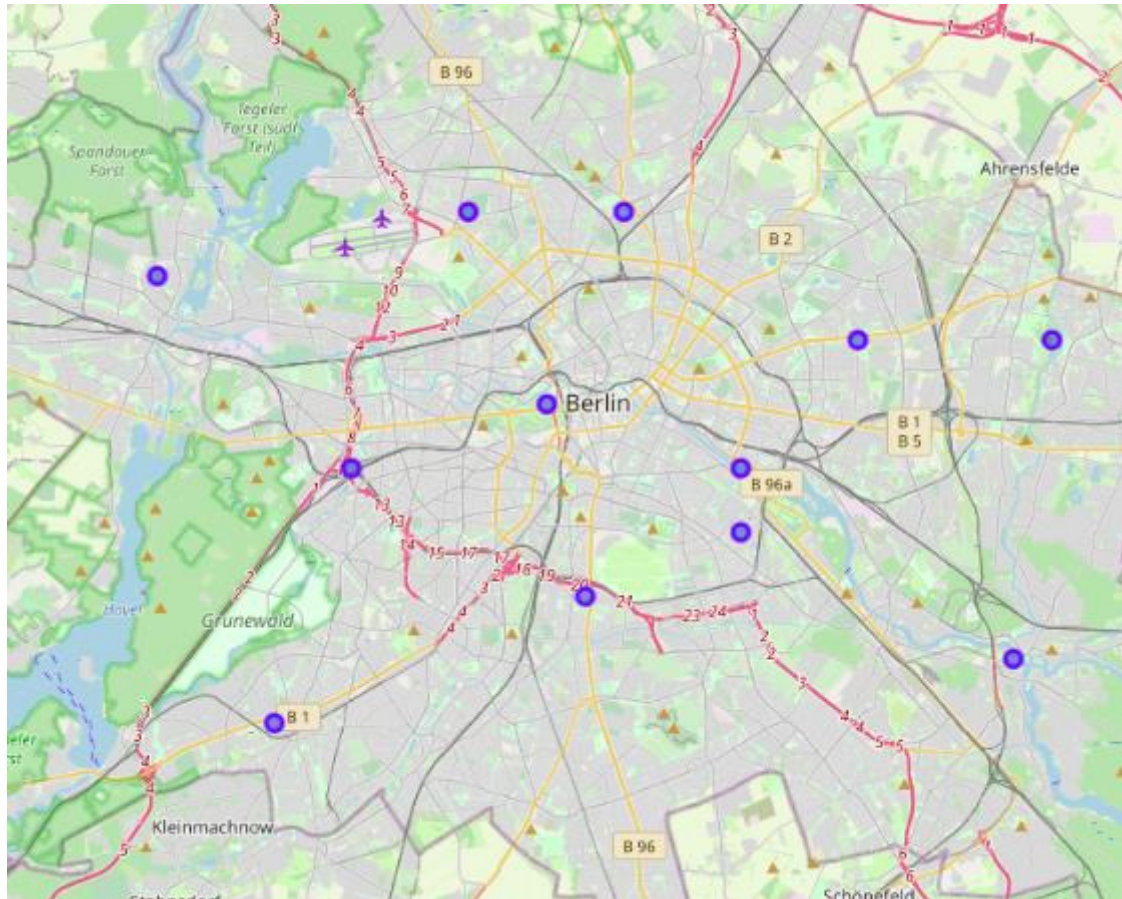
For the problem we are going to research we need the information that can help us to understand where in the city of Berlin we have the high density of points of interest places. We are going to do this with the help of 4square. We will acquire the data of different venues; we cluster them and then we plot it on the map where we can see that "high risk" districts.

Also we need the list of boroughs of Berlin, the number of people who has COVID-19 in Berlin up to this date.

In result we should be able to see how different boroughs have different number of infected and why. The results actually can be used for people to be informed which boroughs are more dangerous in the matter of the covid-19 and why. Of course such topics require really deep investigation but as it is our first project we can try our best :)

Our methodology supposes analytical approach. We are working the data that we gained from the open sources and then we with that data.

At first let's determine the main boroughs of the Berlin. They are 12 boroughs in Berlin. They all allocated with blue points on the map below. They all have different structure, population etc.

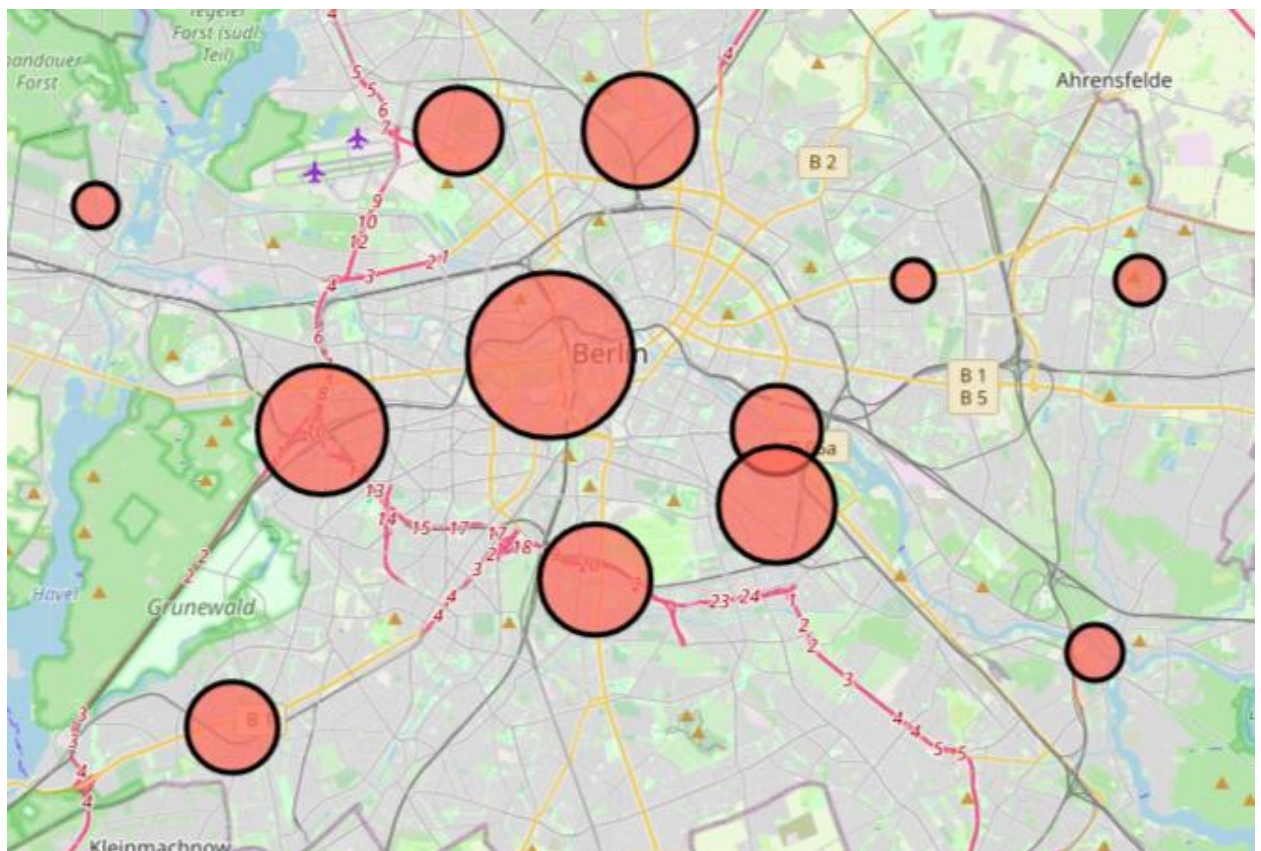


Next we are getting the database with boroughs, coordinates and total number of covid-19 infected up to 23-04-2020. It look like that:

	Borough	Lat	Long	Total
0	Charlottenburg-Wilmersdorf	52.500000	13.283333	627
1	Friedrichshain-Kreuzberg	52.500000	13.450000	434
2	Lichtenberg	52.533333	13.500000	202
3	Marzahn-Hellersdorf	52.533333	13.583333	238
4	Mitte	52.516667	13.366667	806
5	Neukölln	52.483333	13.450000	568
6	Pankow	52.566667	13.400000	552
7	Reinickendorf	52.566667	13.333333	423
8	Spandau	52.550000	13.200000	218
9	Steglitz-Zehlendorf	52.433333	13.250000	443
10	Tempelhof-Schöneberg	52.466667	13.383333	540
11	Treptow-Köpenick	52.450000	13.566667	273

As we can see the most infected is in Mitte, Charlottenburg, Tempelhof districts. For our convenience let put the markers on the map that shows us districts in relation of infected of covid-19.

Using the folium tool we can build a map where we can show the quantity of infected depending on borough.



As we can see the number of infected people is different from borough to borough. The biggest circle is Mitte.

So we can hope that we can find any pattern during our analysis.

Now as we already have the information about the latitude and longitude of the boroughs let's use the 4square API to analyze the closest to the determined boroughs venues.

At first we are defining the radius and the limit of the venues that we get after the request to 4square.

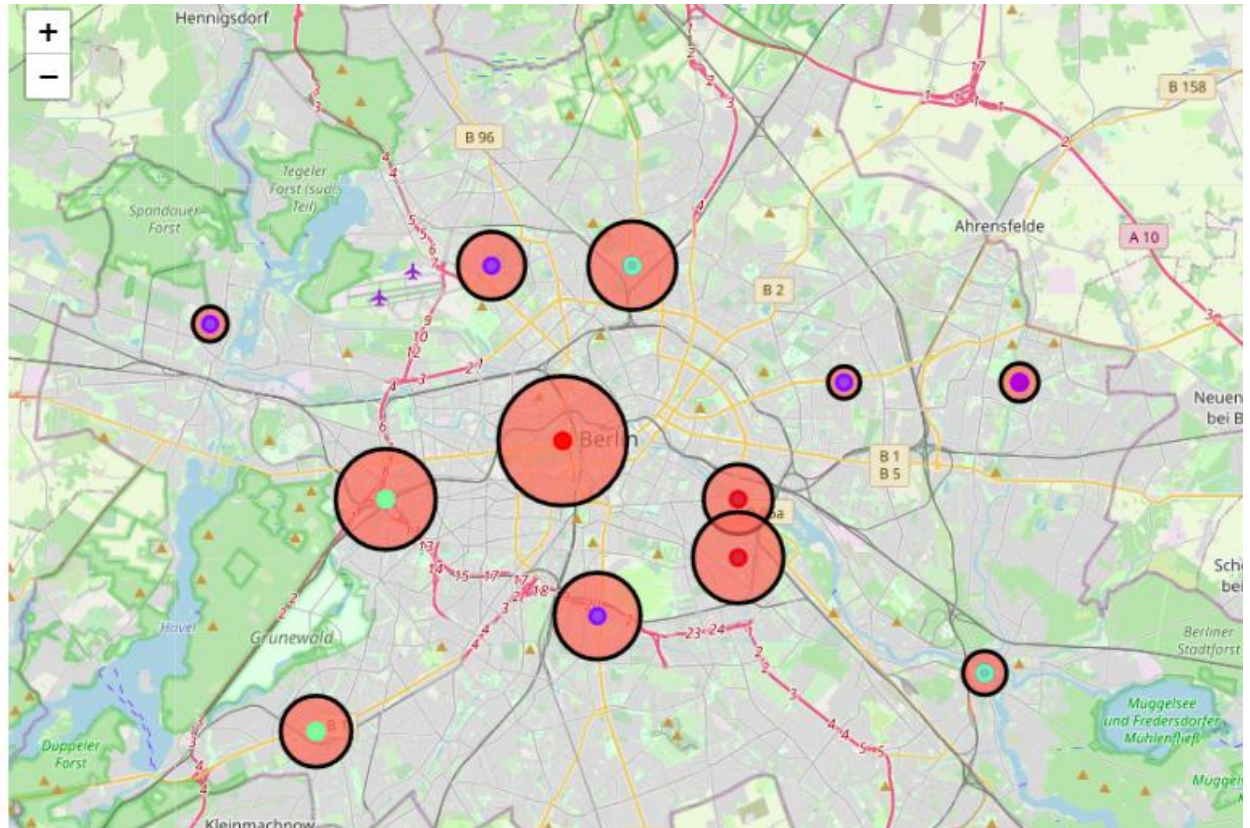
We set it on 200 limit and 2000 radius.

Now we are using our 4square account to get the info. We are getting the venues that a close to our boroughs in Berlin and see what kind of venues are located there. As we can see the most infected borough is in the center of Berlin and common venues there are mostly connected with tourism such as Hotels, Museums etc. On the opposite such boroughs like Spandau or Marzahn-Hellersdorf that a fat from the city center showing us lower number of infected people and this due to the fact that they are mostly 'industrial' boroughs and the most common venues there is supermarket, transport stops and so on.

	Borough	Lat	Long	Total	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Charlottenburg-Wilmersdorf	52.500000	13.283333	627	2	Italian Restaurant	Café	Trattoria/Osteria	Hotel
1	Friedrichshain-Kreuzberg	52.500000	13.450000	434	0	Café	Ice Cream Shop	Coffee Shop	Bar
2	Lichtenberg	52.533333	13.500000	202	1	Supermarket	Tram Station	Park	Furniture / Home Store
3	Marzahn-Hellersdorf	52.533333	13.583333	238	1	Supermarket	Garden	Tram Station	Drugstore
4	Mitte	52.516667	13.366667	806	0	Hotel	Concert Hall	Plaza	Science Museum
5	Neukölln	52.483333	13.450000	568	0	Coffee Shop	Café	Bar	Vegetarian / Vegan Restaurant
6	Pankow	52.566667	13.400000	552	2	Café	Supermarket	Italian Restaurant	Drugstore
7	Reinickendorf	52.566667	13.333333	423	1	Supermarket	Park	Bakery	Metro Station
8	Spandau	52.550000	13.200000	218	1	Supermarket	Bus Stop	Drugstore	German Restaurant
9	Steglitz-Zehlendorf	52.433333	13.250000	443	2	Italian Restaurant	Café	Bus Stop	Supermarket
10	Tempelhof-Schöneberg	52.466667	13.383333	540	1	Park	Supermarket	Italian Restaurant	Café
11	Treptow-Köpenick	52.450000	13.566667	273	2	Café	German Restaurant	Supermarket	Drugstore



We clustered our information into 3 main clusters and we can actually see the pattern. That the red ones showing us the biggest numbers. Teal is a smaller than a red one. And the smallest ones are blue.



## Conclusion

The main problem during the COVID-19 pandemic is not the disease itself but the gatherings of people. Because where the large number of people there are big numbers of infection rates. So, our little project proved that. In times of big pandemic big cities should react quickly especially towards the districts with high density of tourist's places, cafés, restaurants.

Data science will certainly help in the matter of understanding the COVID-19 as it has patterns that can be researched. But the main work will start when the pandemic will be over and data scientist will have the big data set to work with.

BE HEALTHY! 😊

## **Resources that were used during the project**

[https://www.citypopulation.de/en/germany/covid/11\\_\\_berlin/](https://www.citypopulation.de/en/germany/covid/11__berlin/)

<https://en.wikipedia.org/wiki/Berlin>

[https://en.wikipedia.org/wiki/Boroughs\\_and\\_neighborhoods\\_of\\_Berlin](https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin)

<https://simplemaps.com/data/de-cities>

<https://tools.wmflabs.org/>