

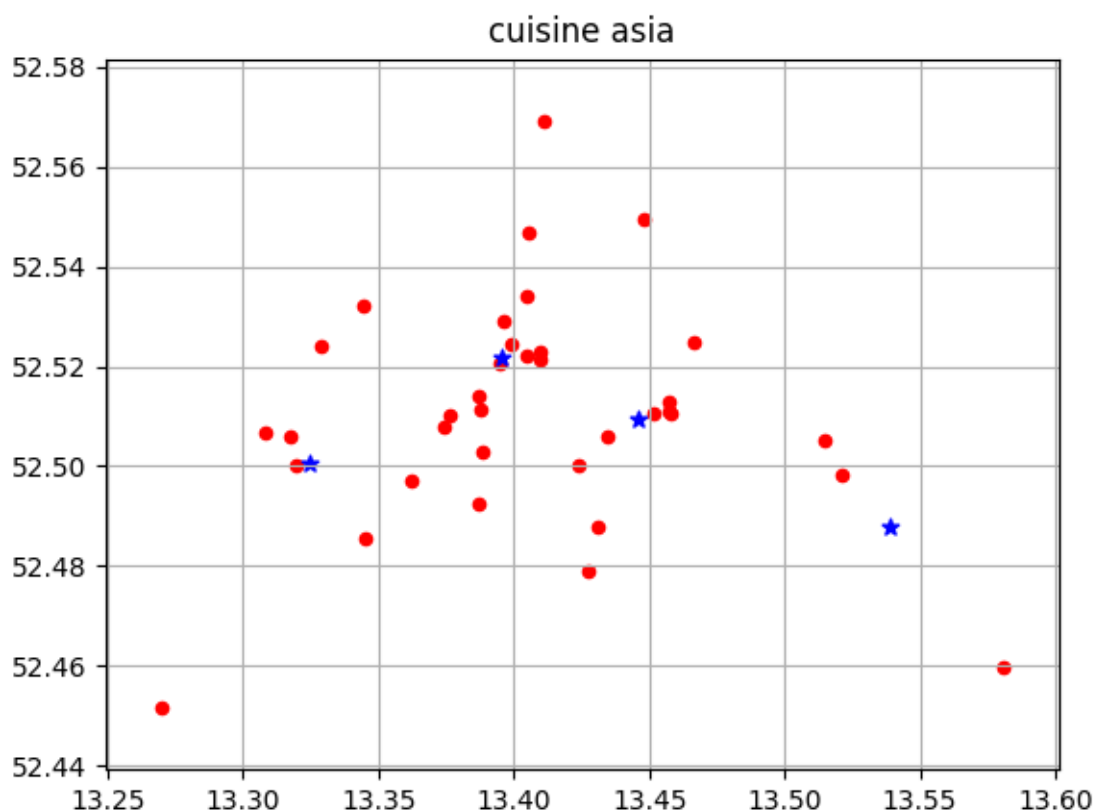
Test_1 - Report

In the first time I tried to organise the information related to the restaurants type. Dataset has incomplete fields and another ones which are non well written, typo errors.

I tried to use an NLP learning process but some tags are written in english and another ones in german so, in order to gain some time, I created an json file with all the synonyms. Information is not hard coded in the script, it's written in an external file, so it can be modified and improved in the future.

Once data about restaurant was organised I took a look over transportation data. Longitude and latitude columns were present but the information related to disability access was not available. The tram and bus system in Berlin is ready for wheelchairs so the only relevant information was which U-Bahn and S-Bahn stations have access for disabled people.

Trying to find where the restaurants are located I created a plot for all the restaurants type with more than fifteen occurrences in the dataset. I applied a Kmeans algorithm to find the centroid of the clusters.



The right number of clusters was determined by the elbow method.

For example, red circle represent restaurants location while blue stars are the centroid of the four clusters. Through the geocoder module we can determine the best location if we are planning to eat some kind of food. For this example, we should go to:

- Georgenstraße/Am Kupfergraben, Georgenstraße, Spandauer Vorstadt, Mitte
- Metro, An der Ostbahn, Fhain, Friedrichshain-Kreuzberg

- 19, Zwieseler Straße, Karlshorst, Lichtenberg
- 23, Bleibtreustraße, Charlottenburg, Charlottenburg-Wilmersdorf

Running this script, correlation.py, it can be check how popular restaurants for tourists are closer to the city centre rather than minorities cuisines. Also the number of restaurants for tourists is clearly higher than local food establishments.

	Number of restaurants
Arabic	22
Asian	185
Austria	23
Burger	50
Chinese	112
Coffee Shop	15
Croatian	24
France	47
Germany	396
Greece	101
Indian	186
International	81
Italian	752
Japan	162
Korean	46
Mediterranean	31
Mexican	61
Spanish	45
Steak	52
Thai	78
Turkish	75
Veggie food	40

There are almost eight hundreds places of Italian food and only four hundreds dedicated to local food.

We have more information, information related to welfare society. Availability of supermarkets, stations, banks, restaurants and gas stations can be a good factor of how good life is.

In density_map.py script there are several utilities for plot the distribution of these services along the city. But it could be interesting to know which places in the city have more services. And each citizen will have different priorities. If someone is thinking about moving to the city center the availability of stations could be more priority than the number of gas stations.

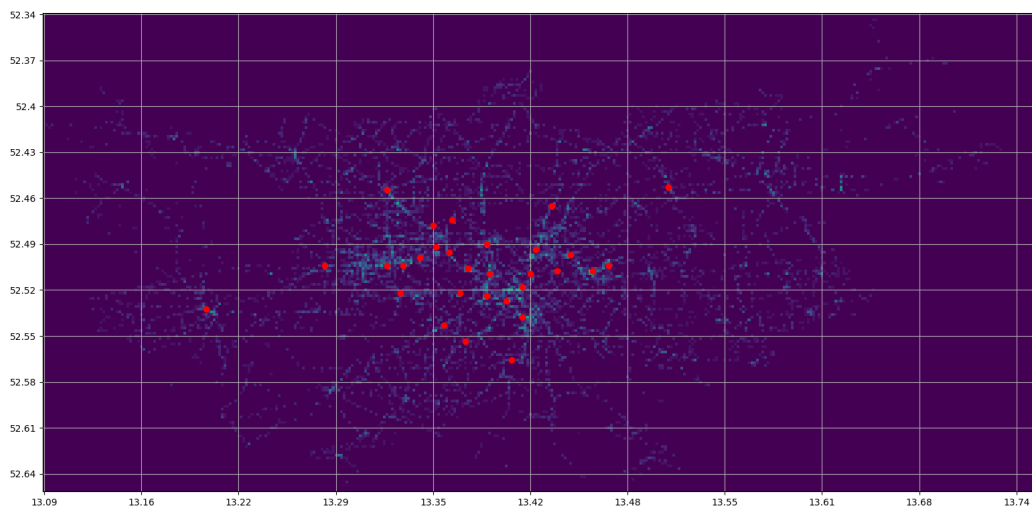
In `get_peaks.py` a map will merge all the previous maps. In this situation I thought about someone who will be use public transportation.

The total distance from west to east and from south to north is divide in squares with a side length of two hundreds meters. The locations available inside each square are multiply for an “interest factor”.

In our case:

$$2 * \text{restaurants number} + 2 * \text{stations} + 2 * \text{supermarkets} + 2 * \text{ATMs} + 1 * \text{gas_stations}$$

If we compute this equation for each square in the map we will get a new map. Local maximums will be the more interesting locations to live in Berlin.



The address of these points are extracted from the latitude/longitude values.