How to use data to find the best spot for a gastronomy
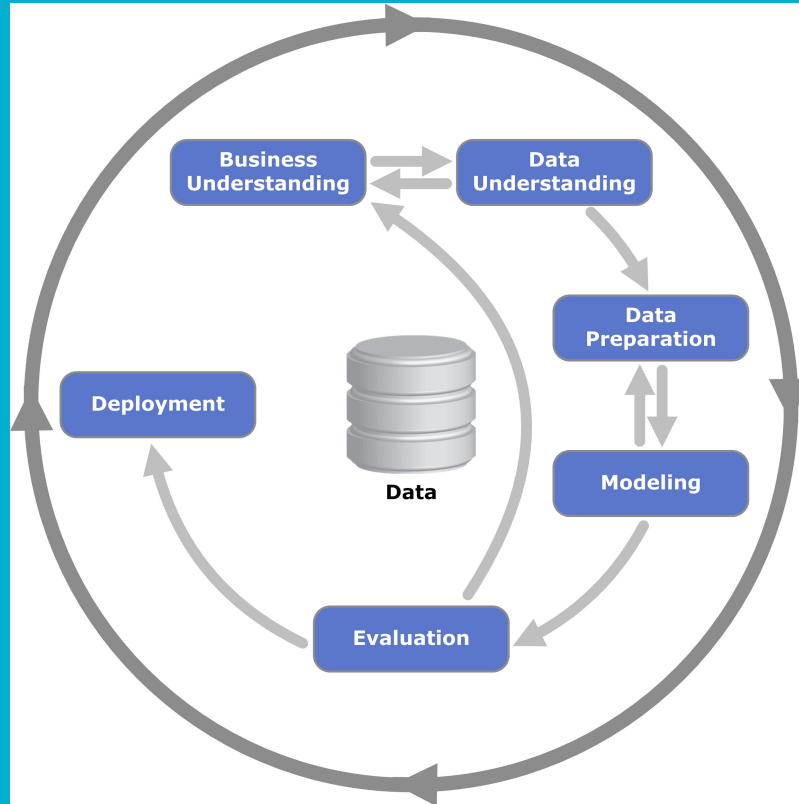
—

Manuel Strigli

Capstone Project

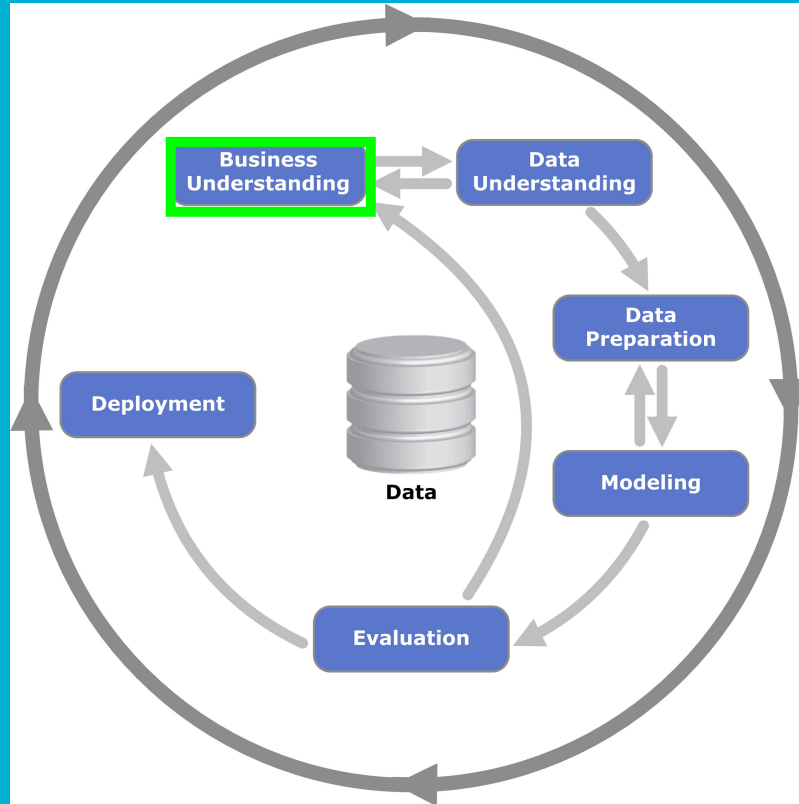Coursera:

IBM Data Science Specialization

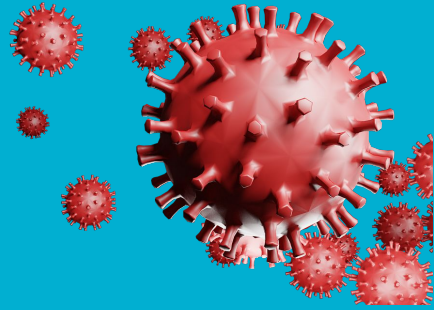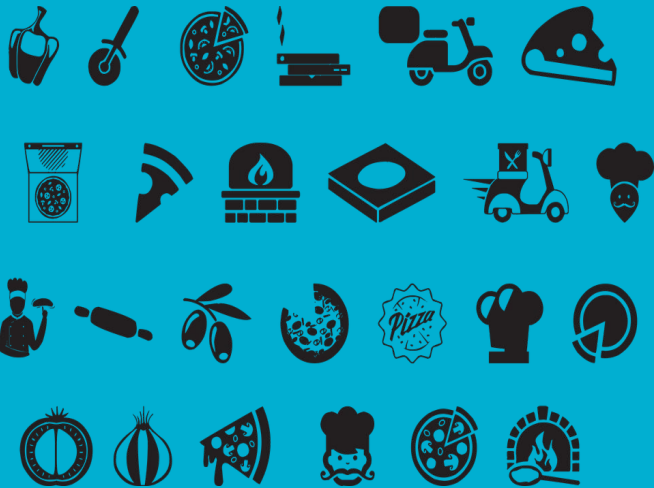# Use the CRISP-Model to solve the Problem

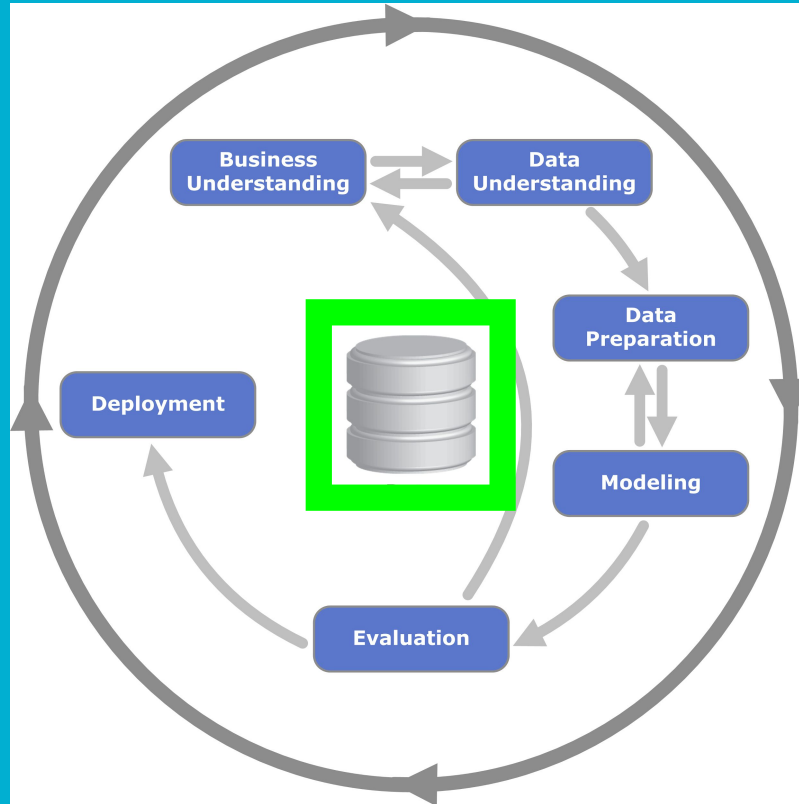# How gastronomy business in Berlin works?

In Berlin there is a hard competition in gastronomy market.
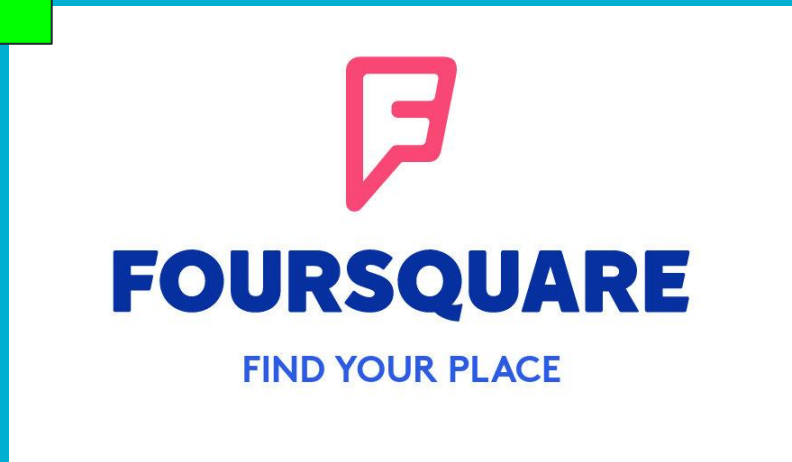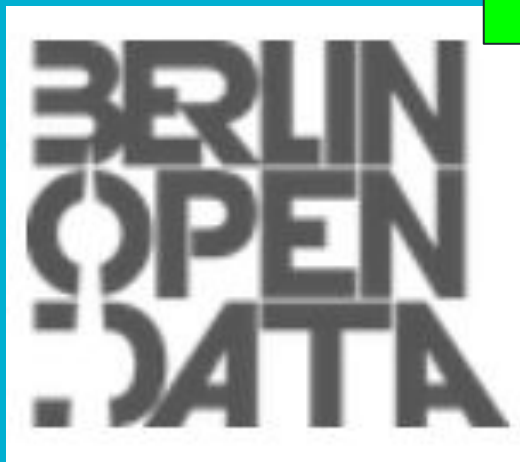
Gastronomies with delivery service have a business advantage because of the lockdown
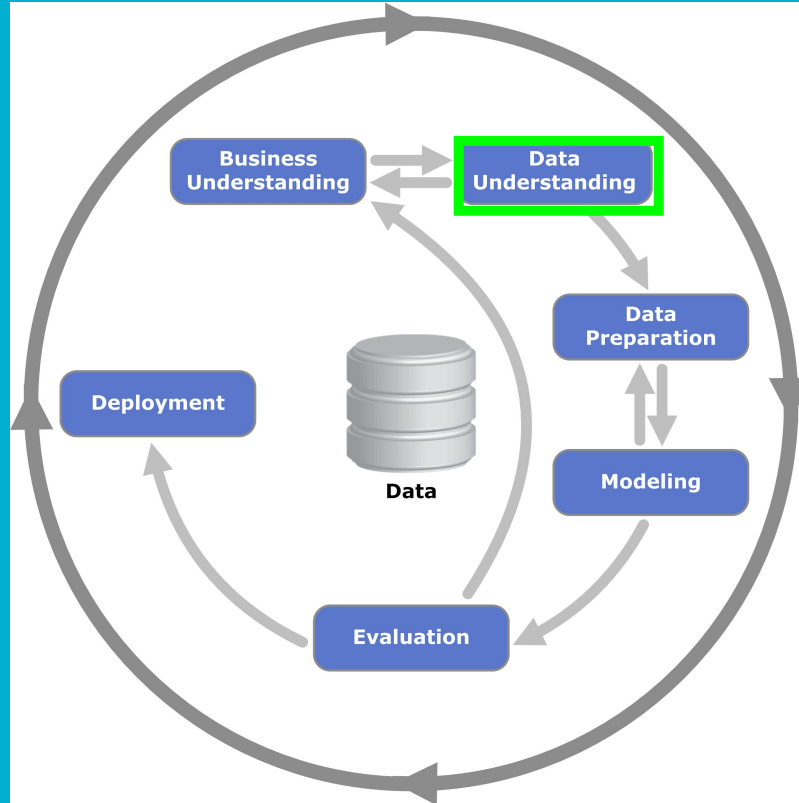
Where comes the necessary data from?

# Foursquare and Berlin government are our data sources.

How does our data looks like?

There are no 'neighborhoods' but everything else needed

```
{'type': 'FeatureCollection',
 'features': [{'type': 'Feature',
   'geometry': {'type': 'Point', 'coordinates': [13.38212, 52.53116]},
   'properties': {'title': '771',
    'href': '262',
    'description': ' <a href="262">Mehr...</a>',
    'id': '/sen/web/service/liefer-und-abholdienste/index.php/detail/771',
    'data': {'id': '771',
     'unique_id': '262',
     'name': 'Brasserie la bonne franquette',
     'strasse_nr': 'Chausseestraße 110',
     'plz': '10115',
     'art': 'Gastronomie (Café, Restaurant, Imbiss, Lebensmittelhandlung, usw.)',
     'angebot': 'Klassische Französische Küche',
     'lieferung': 'FALSCH',
     'beschreibung_lieferangebot': '',
     'selbstabholung': 'WAHR',
     'angebot_selbstabholung': 'Bestellungen werden jederzeit entgegengenommen. Abholung (später auch Lieferung) immer dienstags
bis samstags in der Zeit von 17 bis 21 Uhr',
     'fon': '+493094405363',
     'w3': 'https://labonnefranquette.de',
     'mail': 'essen@labonnefranquette.de',
     'montag': '',
     'dienstag': '17:00-21:00',
     'mittwoch': '17:00-21:00',
     'donnerstag': '17:00-21:00',
     'freitag': '17:00-21:00',
     'samstag': '17:00-21:00',
     'sonntag': ''}}},
```

How to prepare the data?
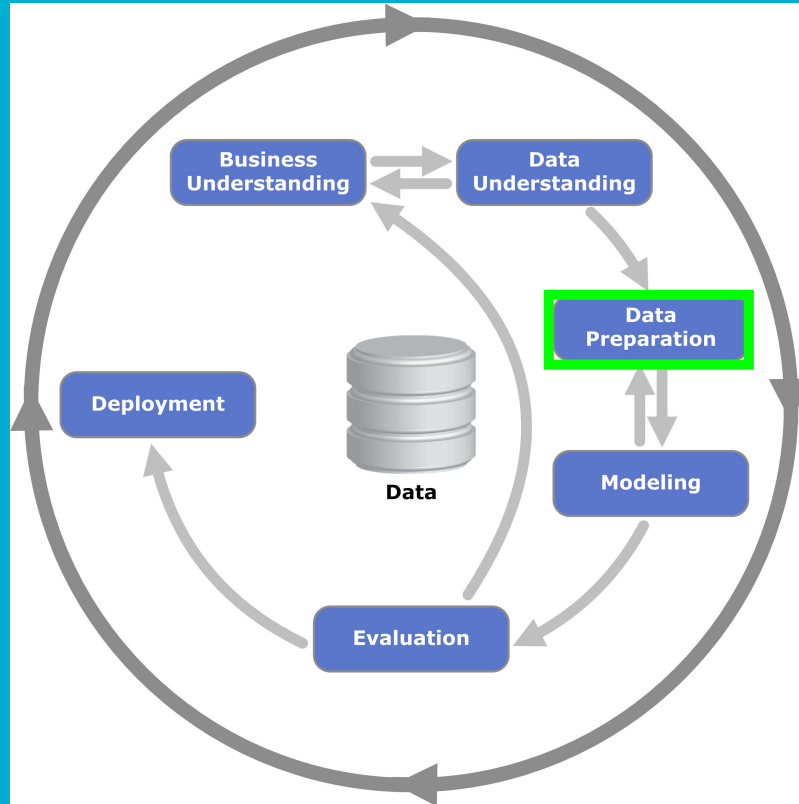
Preparing the data as a clean dataframe make modeling possible

	Name	Speciality	Delivery	Pickup	Postal_code	Latitude	Longitude
0	Brasserie la bonne franquette	Klassische Französische Küche	FALSCH	WAHR	10115	52.53116	13.38212
1	Alpenstück	Süddeutsche Spezialitäten wie Maultaschen Köni...	FALSCH	WAHR	10115	52.53033	13.39184
2	sagrantino 136	Italian fusion kitchen 3 -course menus by Mat...	FALSCH	WAHR	10115	52.52633	13.38897
3	Risorante Bonfini	Italienische Küche, Pasta und Pizza	FALSCH	WAHR	10115	52.52953	13.38480
4	CAFE RIBO	Maultaschen, schwäbische küche, cafe	FALSCH	WAHR	10115	52.53089	13.39654

How to model our data?
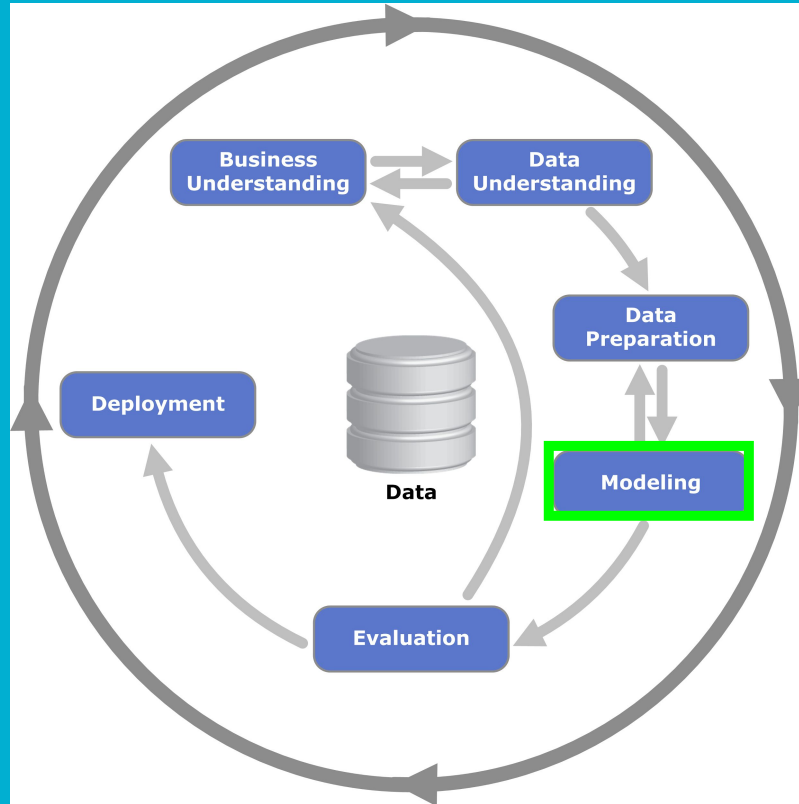
Folium Map with Delivery Classification
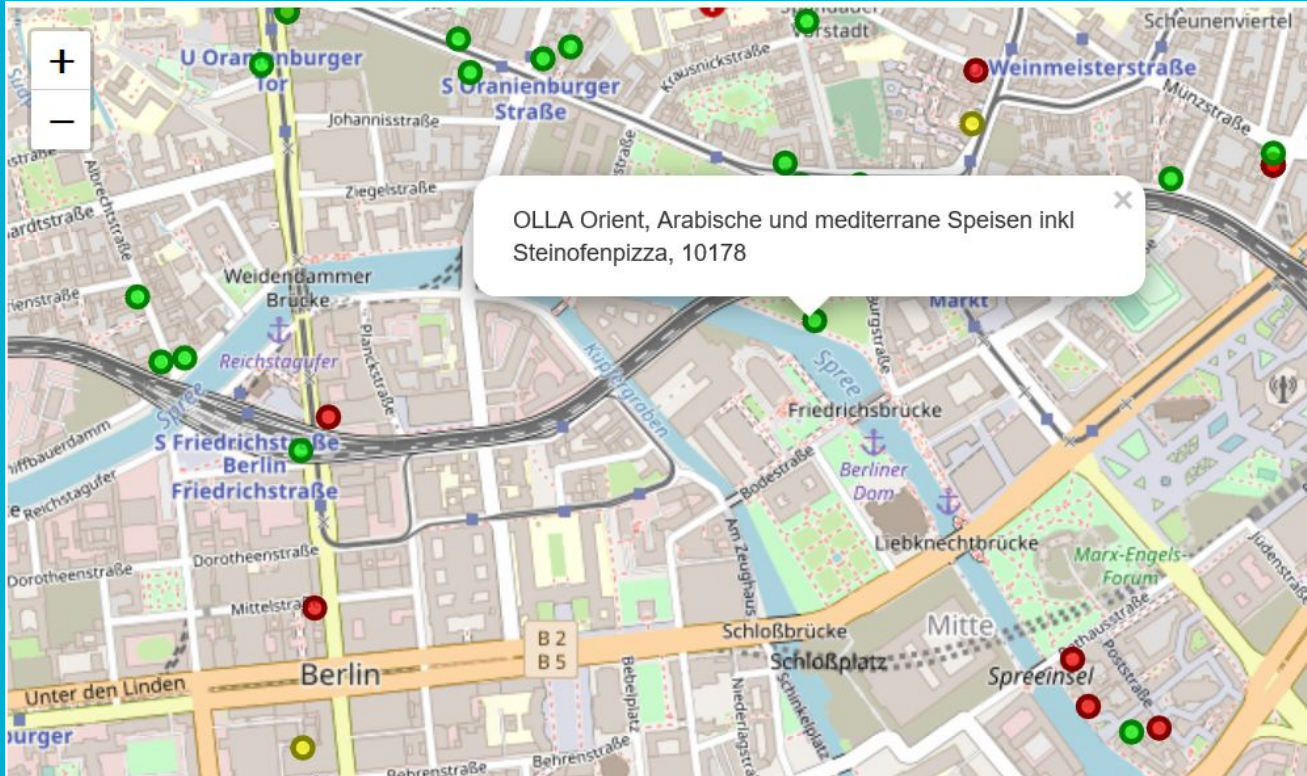
How useful is the map?
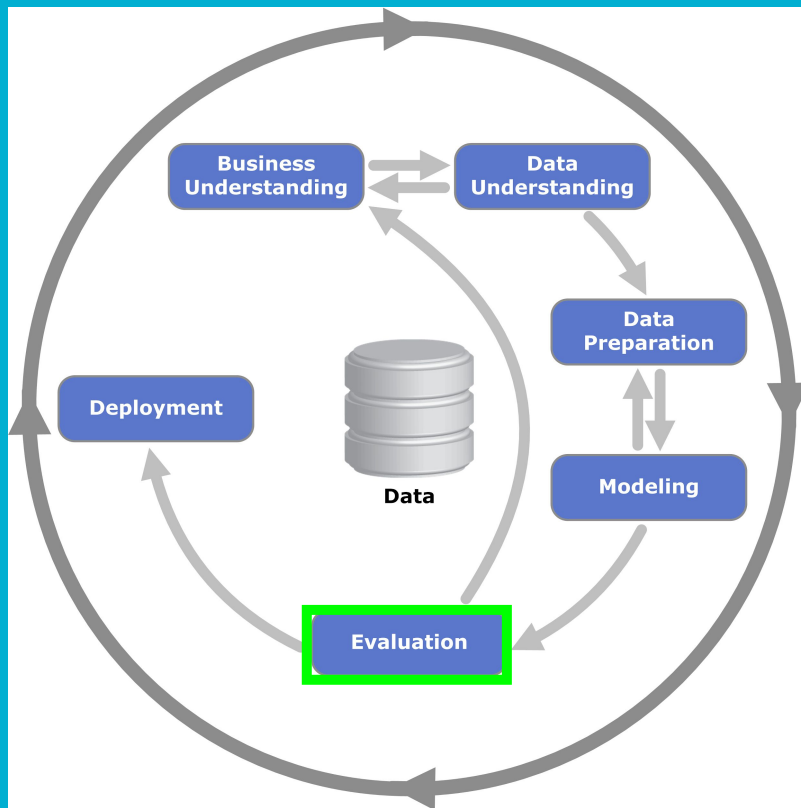
There are gaps from e.g. parks and missing values for the postal codes too

Espresso Bar, Wraps, shaking salads, pasta, sweets
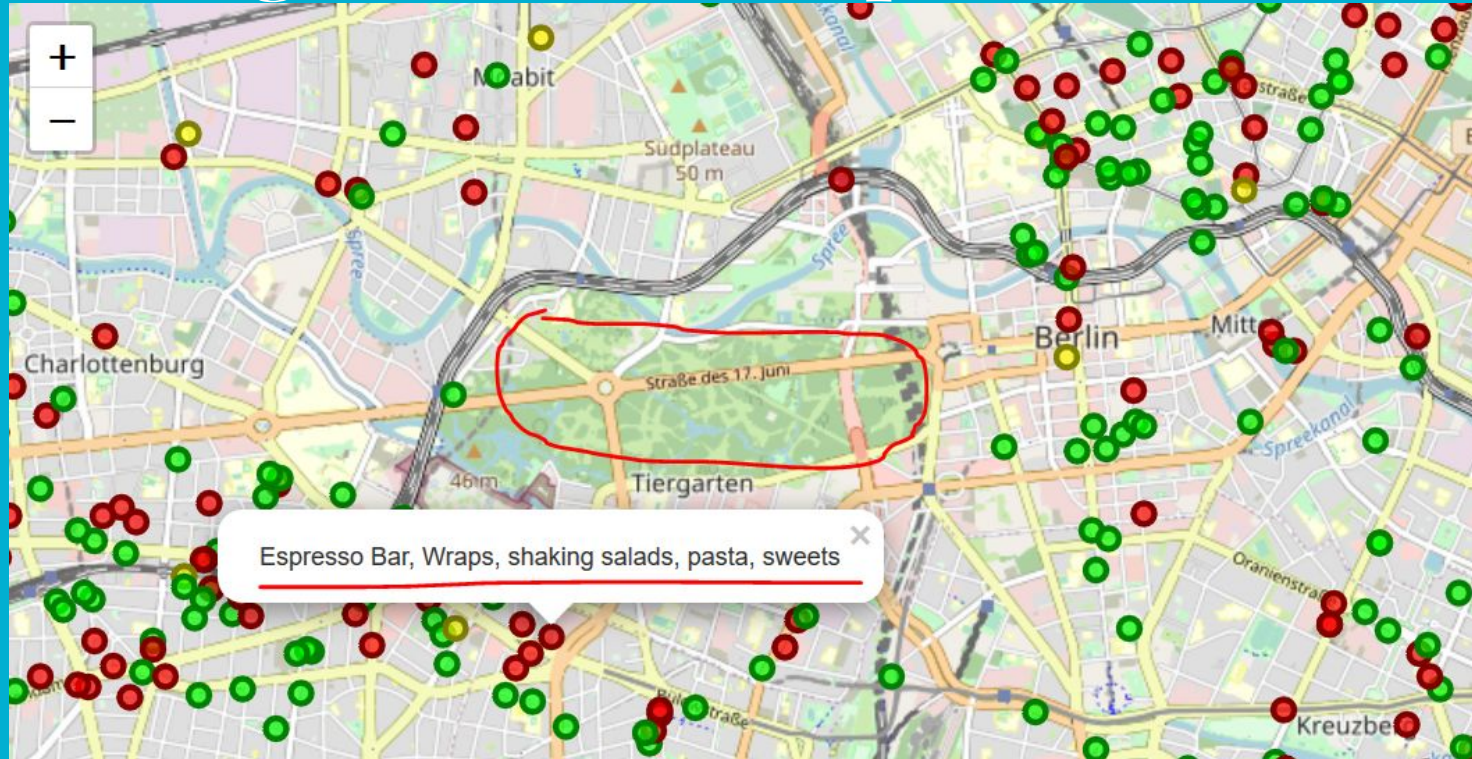
Jumping back to Business Understanding
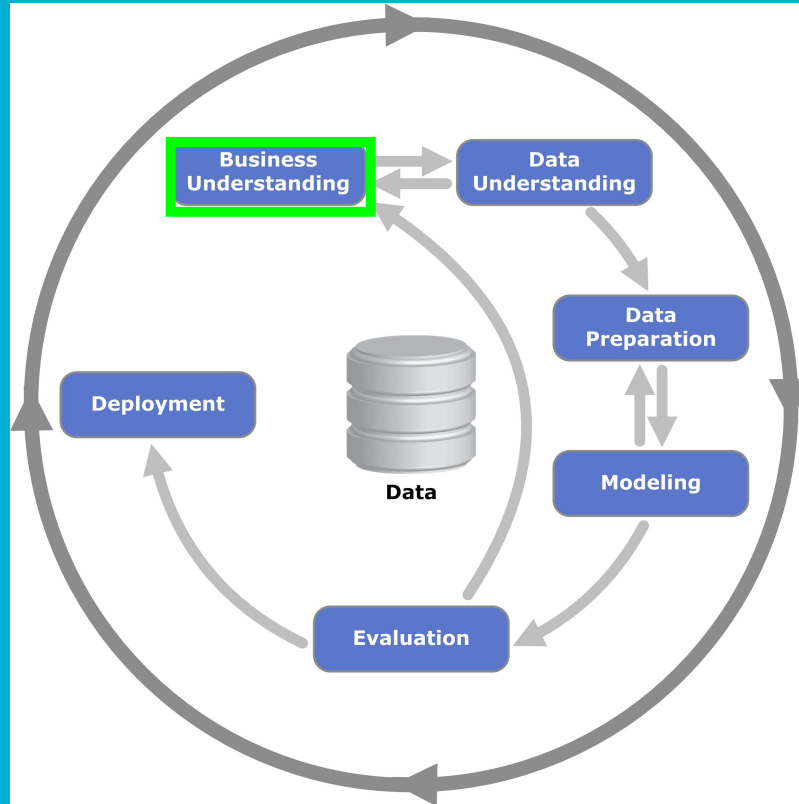
Map is useful for orientation

dataset is incomplete

gastronomy place depends on people density

Focus on Foursquare as Data Source

How does our data looks like?

Working through the data by districts

Note that the districts are different big and there is no consideration of that for the search

Add a radius for each district where to look for venues

How to prepare the data?

Preparing the data as a clean dataframe make modeling possible

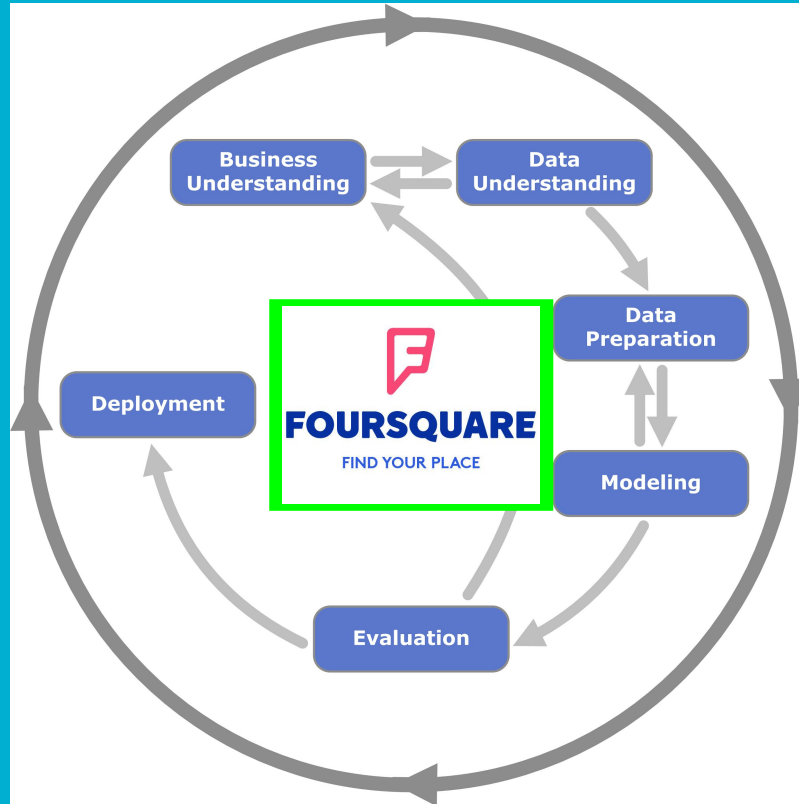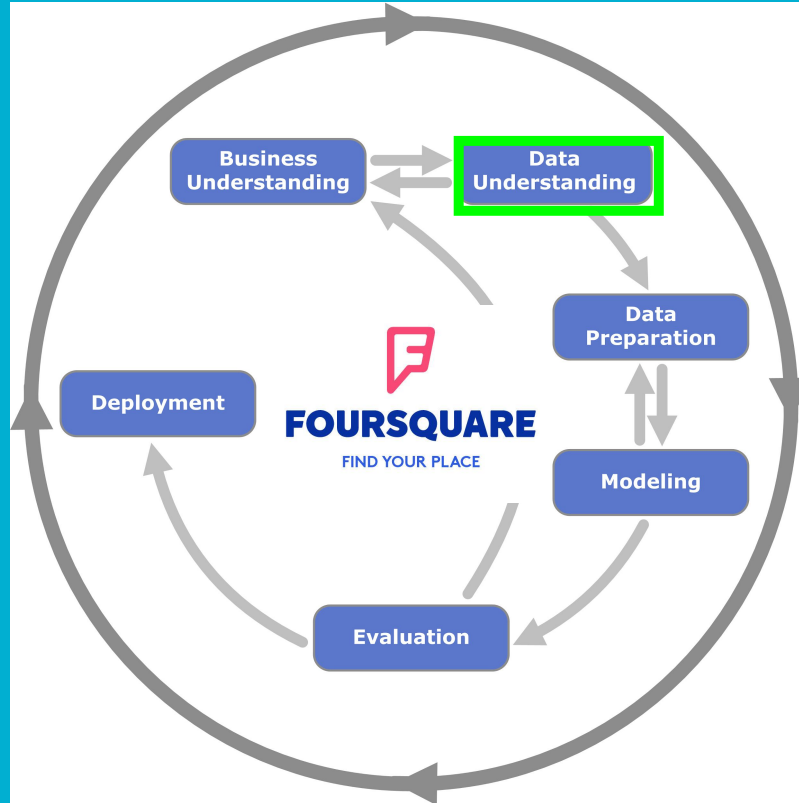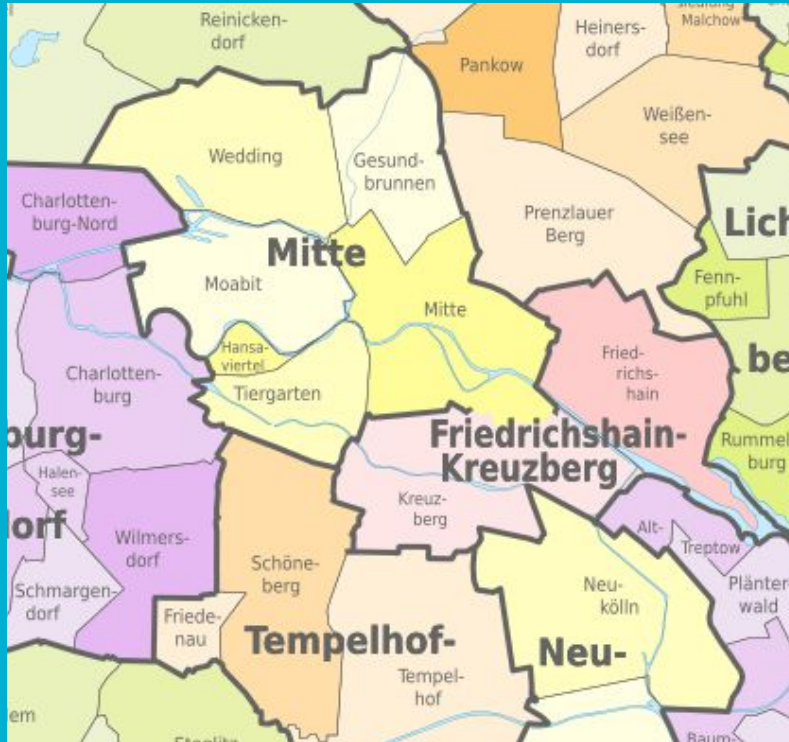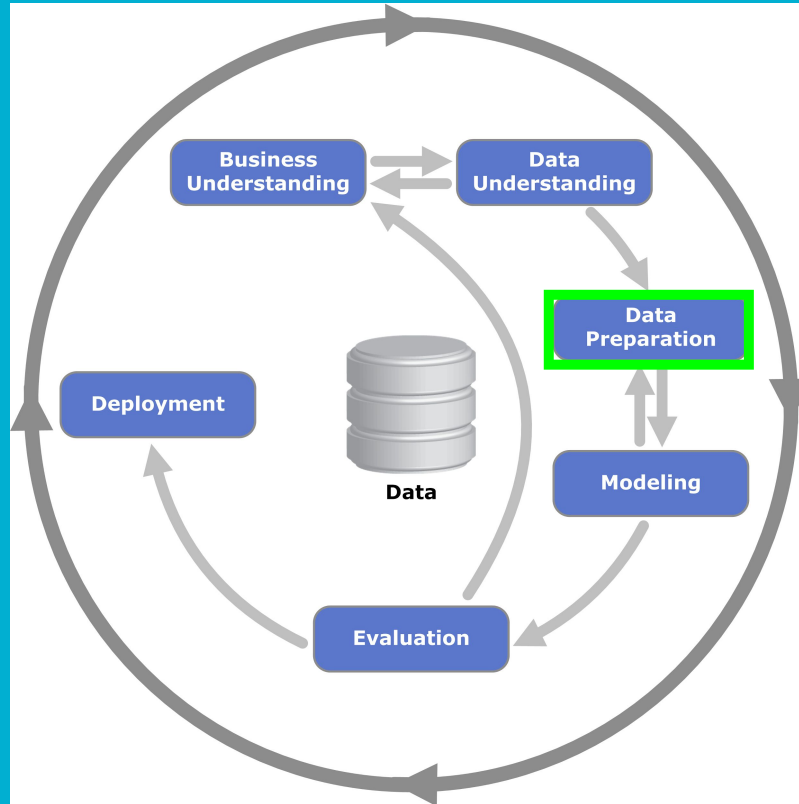|	District	Latitude	Longitude	Radius
0	Mitte	52.522295	13.362587	2000
1	Tiergarten	52.510807	13.336607	1500
2	Hansaviertel	52.518247	13.333539	500
3	Moabit	52.524654	13.322123	1500
4	Charlottenburg	52.517217	13.345421	3000
5	Halensee	52.495719	13.275483	500
6	Wilmersdorf	52.485248	13.289401	2000
7	Schöneberg	52.479915	13.321338	2500
8	Tempelhof	52.465193	13.356444	2000
9	Neukölln	52.477365	13.407260	2000
10	Kreuzberg	52.496636	13.375818	2000
11	Alt-Treptow	52.492591	13.399446	500
12	Friedrichshain	52.508585	13.420751	2000
13	Prenzlauer Berg	52.539299	13.399509	2000
14	Gesundbrunnen	52.550445	13.349664	1500
15	Wedding	52.550123	13.302204	1500

getting the venues for each district
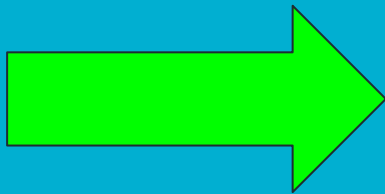
use one hot encoding

	District	African Restaurant	Airport Lounge	Airport Service	Aquarium	Art Gallery
0	Alt-Treptow	0.017857	0.00	0.000	0.00	0.00
1	Charlottenburg	0.000000	0.00	0.000	0.00	0.01
2	Friedrichshain	0.010000	0.00	0.000	0.00	0.03
3	Gesundbrunnen	0.010000	0.00	0.000	0.00	0.00
4	Halensee	0.000000	0.00	0.000	0.00	0.00
5	Hansaviertel	0.000000	0.00	0.000	0.00	0.01
6	Kreuzberg	0.000000	0.00	0.000	0.00	0.01
7	Mitte	0.000000	0.00	0.000	0.00	0.01
8	Moabit	0.000000	0.00	0.000	0.00	0.01
9	Neukölln	0.010000	0.00	0.000	0.00	0.02
10	Prenzlauer Berg	0.000000	0.00	0.000	0.00	0.00
11	Schöneberg	0.000000	0.00	0.000	0.00	0.00
12	Tempelhof	0.010000	0.00	0.000	0.00	0.00
13	Tiergarten	0.000000	0.00	0.000	0.01	0.01
14	Wedding	0.000000	0.05	0.175	0.00	0.00
15	Wilmersdorf	0.000000	0.00	0.000	0.00	0.00

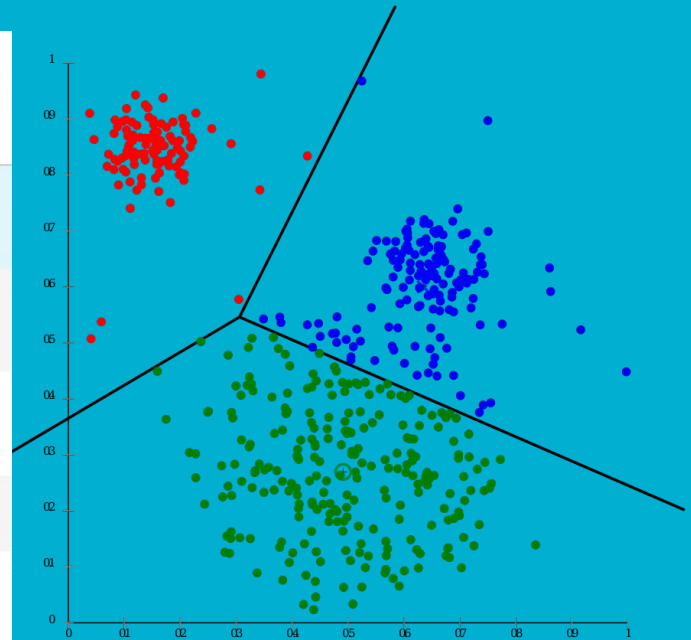# Preparing the data as a clean dataframe make modeling possible

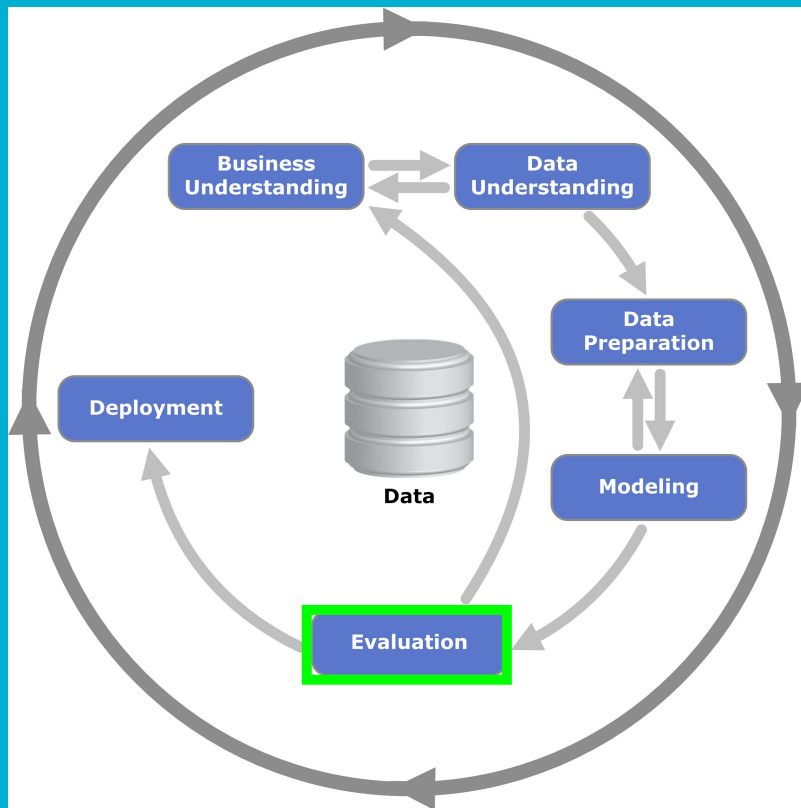| District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Alt-Treptow	Coffee Shop	Café	Italian Restaurant	Bakery	Cocktail Bar	Mediterranean Restaurant	Bookstore	Gift Shop	Kebab Restaurant	Breakfast Spot
Charlottenburg	Hotel	Café	Zoo Exhibit	Coffee Shop	Beer Garden	Art Museum	Concert Hall	Cocktail Bar	Monument / Landmark	Gym / Fitness Center
Friedrichshain	Bar	Coffee Shop	Café	Turkish Restaurant	Italian Restaurant	Bakery	Hotel	Art Gallery	Plaza	Organic Grocery
Gesundbrunnen	Café	Bar	Park	Coffee Shop	Ice Cream Shop	Supermarket	Drugstore	Turkish Restaurant	Chinese Restaurant	Fast Food Restaurant
Halensee	Historic Site	Italian Restaurant	Light Rail Station	Lake	Automotive Shop	Beach	Farmers Market	Falafel Restaurant	Fair	Fabric Shop

How to model our data?
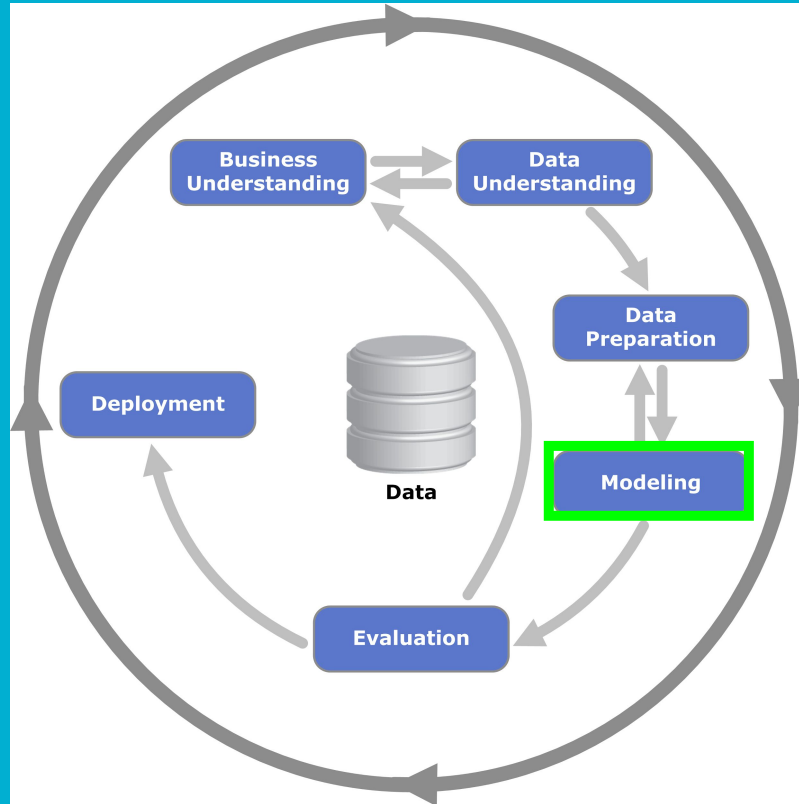
Using k-means algorithm for clustering

	District	Latitude	Longitude	Radius	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Mitte	52.522295	13.362587	2000	2	Hotel	Monument / Landmark	Spa	Park
1	Tiergarten	52.510807	13.336607	1500	2	Zoo Exhibit	Hotel	Art Museum	Jazz Club
2	Hansaviertel	52.518247	13.333539	500	3	Hotel	Pub	Hotel Bar	Gastropub
3	Moabit	52.524654	13.322123	1500	0	Hotel	Bakery	Supermarket	Asian Restaurant
4	Charlottenburg	52.517217	13.345421	3000	2	Hotel	Café	Zoo Exhibit	Coffee Shop

How useful is the clustering?

Cluster shows what kind of gastronomy is missing and the map where to place it in each district
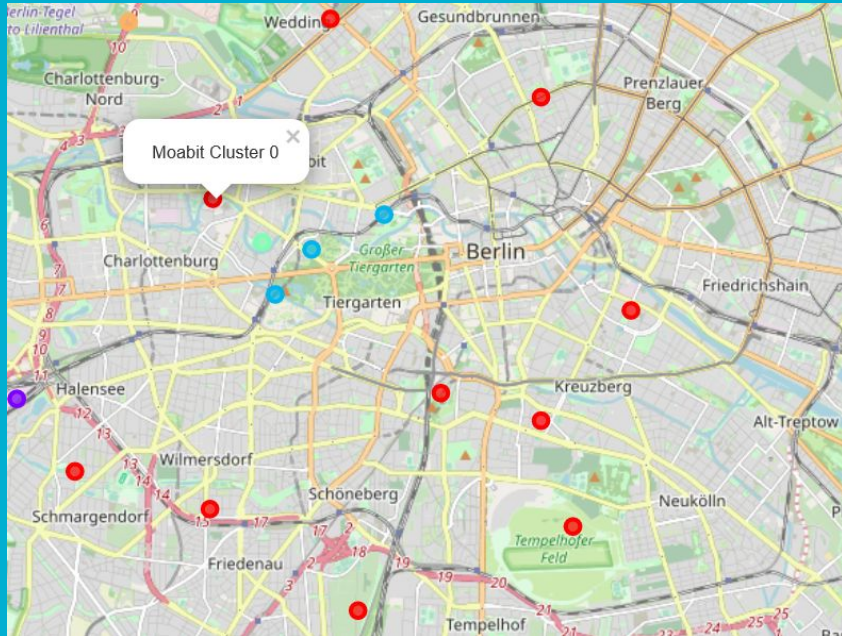
	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	Moabit	Bakery	Hotel	Supermarket	Bar	Asian Restaurant

in my district Moabit there is a e.g. a vegan restaurant missing and an entrepreneur should look for a place somewhere around the spot shown on the map

If you are a brave entrepreneur, the deployment lies in your hands!