1 Part 3 . . . brief assessment of the collected data

Following the database being populated, an initial evaluation of the collected data and an assessment of its usefulness is now being carried out.

The diagram at the end shows quite well the geographical distribution of the sampled locations on a map of Germany.

connect to the db

1.1 Some figures about the collected data

Now I will check how many data sets are available at the end of data fetching and which characteristics they have. The results are to be used to assess whether the data are suitable in principle and which preselections are to be made for later analysis.

To make the visualisation easier, I use a simple class for the creation of a data frame to display the values nicely.

1.1.1 Records

Using almost the entire number of 1,000 free API calls, around 9,800 data sets were collected. Of the locations collected, around 7,200 have values at peak hours (around 73%). Almost all locations are designated as operational, except for a few sites (approx. 0.05%).

```
                            description value
0                         recorded places  9815
1  business status is signed as "operational"  9772
2                   "peak hours" provided by  7264
```

1.1.2 Gas stations and more

Of approximately 14,000 petrol stations in Germany, about 8,600 locations are recorded as gas stations in the database (type "gas station"). This corresponds to about 60% of all German gas stations. The entire data set includes about 9,800 locations, which in addition to gas stations also includes locations searched for as parking areas and service areas. In addition, there are certainly some false matches amongst them.

The 10 largest brands of gas stations in Germany can be found in the database with shares between 24% and 66%. With a total of about 5,100 locations, they represent about 52% of all gas stations found in the database.

The coverage of data on gas stations seems to be comprehensive. However, in subsequent evaluations, it is important to ensure that the attributes 'Gas station' and others are correctly taken into account.

```
                            description  \
0   "gas station" is in the place types list
1
2                                       Aral
```

```
3                              Shell
4                              Total
5                               Esso
6                               Avia
7                                Jet
8                         Raiffeisen
9                               Star
10                              Agip
11                               HEM
12                                 *


                                value
0                                8612
1
2                   1436 (von 2300*)
3                   1076 (von 1955*)
4                    644 (von 1157*)
5                     644 (von 961*)
6                     299 (von 818*)
7                     276 (von 843*)
8                     172 (von 715*)
9                     192 (von 582*)
10                    177 (von 468*)
11                    165 (von 411*)
12  Zahlen von ADAC - Markenverteilung der Tankste...

array([<AxesSubplot:title={'center':'value'}, xlabel='description'>],
      dtype=object)
```
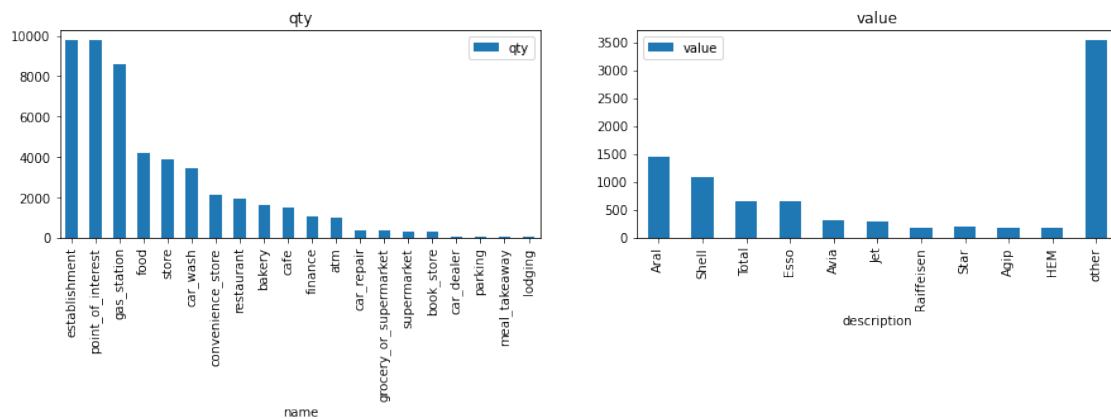
1.1.3 Autobahnen

An important place for fast-charging stations will probably be on or near an Autobahn (motorway). I therefore take a look at the data sets where the proximity to an Autobahn can be inferred from the name or address. Here is the number of records where the strings "Autobahn", "BAB" (federal motorway) or "Autohof" (Truck stop near the Autobahn) are found.

In total, these are just under 600 entries. Also here, it must be taken care of later that the locations are provided with the attribute "gas station" if one only wants to consider the gas stations. In 2020, there were apparently 358 petrol stations on German motorways (source: statista.de)

```
                          description  value
0   "Autobahn" is in the name or address     68
1        "BAB" is in the name or address    279
2    "Autohof" is in the name or address    251
```

1.1.4 Raststätten

Raststätten (service areas) along and near motorways can theoretically also become locations for fast charging stations. An initial query shows that just over 300 Raststätten have been recorded in the database. These can be locations that also include a gas station, or a separately operated gas station.

```
                          description  value
0   "Raststätte" is in the name or address    333
```

1.1.5 Geographical distribution

Germany is organised into slightly more than 400 administrative districts. These can be actual counties, but also entire cities. The database contains locations from slightly more than half of the German districts. This can be explained by the fact that in addition to the automatic search, a targeted search for locations near motorways was also carried out. The background for this procedure was the limitation of API accesses already mentioned. The situation is similar for the zip codes. Of the 8,181 zip codes (source: dpdhl.com/ press release 2018), one or more entries are available for around 43% of all zip codes.

With regard to the concentration in metropolitan areas and motorways, the overall geographical distribution appears to be relatively balanced.

```
              description  value
0   different zip codes   3559
1   different districts    214
```

Here I show how the search results are distributed across the postcodes, or how many hits are found per zip code. This time I have set up a VIEW, as I thought I would be able to use this more often later.

First, the number of database entries per zip code is recorded and transferred to a DataFrame.

Next, I create a dataframe with all the zip codes, in order to be able to use this . . .

... with the other data frame.

```
   records_per_zip  count_all_items
0              0.0             4637
1              1.0             1529
2              2.0              778
```

The two diagrams below show the frequency of occurrence of search results in individual post-codes. There are no search results for over 4,600 postcodes. In approx. 1,500 postcodes there is exactly one search result in the database, in approx. 780 postcodes there are exactly 2 hits, and so on. The left diagram shows the frequency with 0 to 3 occurrences. The right diagram shows the frequency with occurrences of 3 to 20. In sum, more than 3 search results have been recorded in the database for 1,252 postcodes.

```
1252
```

```
array([<AxesSubplot:title={'center':'count_all_items'},
xlabel='records_per_zip'>],
       dtype=object)
```
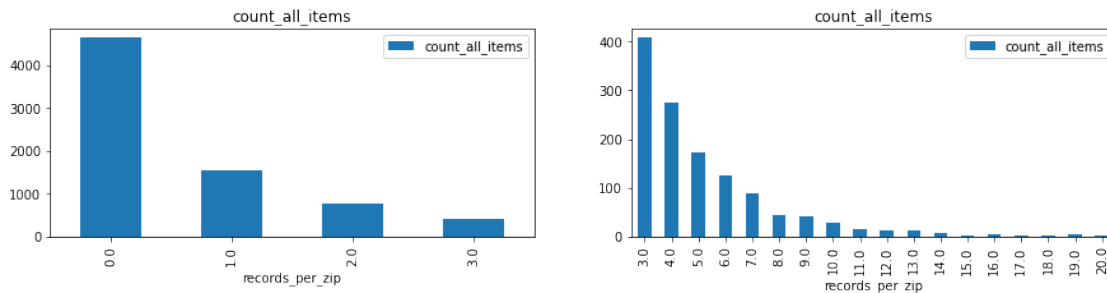

1.2 Global overview of the geographical distribution

First of all, a few technical remarks on the creation of a geographical overview. The diagram itself is displayed at the end of the section.

1.2.1 load the geodata (longitude and latitude) of all places

```
    location_lng    location_lat
0  13.6938440000  51.0612678000
1  13.7636915000  51.0643382000
```

1.2.2 Load the path of the Autobahnen

To see the location of the places in relation to the motorways in Germany, the paths motorways are to be superimposed. For this purpose, a geojson file is read in from opendata esri. The file is very large because it contains the German motorways as well as those of other European countries.

In order to increase the speed when trying out the different views (and because not all the data is necessary for this), only the relevant lines (country: Germany and the lines corresponding to motorways) are extracted and saved locally as a DataFrame in a pickle file,

source of geojson file: Autobahnen shapefile: https://opendata-esri-de.opendata.arcgis.com/search?categories=infrastructure&q=shapefile%20autobahn

```
file = open('xn--DLM1000_Straenachsen-mwb.geojson')
df = gpd.read_file(file)
autobahnen = df[(df['LAND'] == 'DE') & ((df['BEZ'].str.match('^E[0-9]+#A[0-9]+'))|(df['BEZ'].str
autobahnen.to_pickle("autobahnen_shapes.pkl")
```

1.2.3 Load the previously locally saved dataframe

```
     FID LAND MODELLART OBJART      OBJART_TXT             OBJID  HDU_X  \
188  189   DE   DLM1000  42003  AX_Strassenachse  DEBKGDL100001WHS      0

                         BEGINN ENDE OBJART_Z  ...    BEZ FSZ   FTR   IBD  \
188  2018-12-31T08:00:00+00:00          42002  ...  E35#A5   4  2000  2001

      WDM ZUS BEMERKUNG   SYMBOLNR SHAPE_Length  \
188  1301            42003_201      0.146848

                                          geometry
188  MULTILINESTRING ((7.58983 47.90849, 7.58778 47...

[1 rows x 21 columns]
```
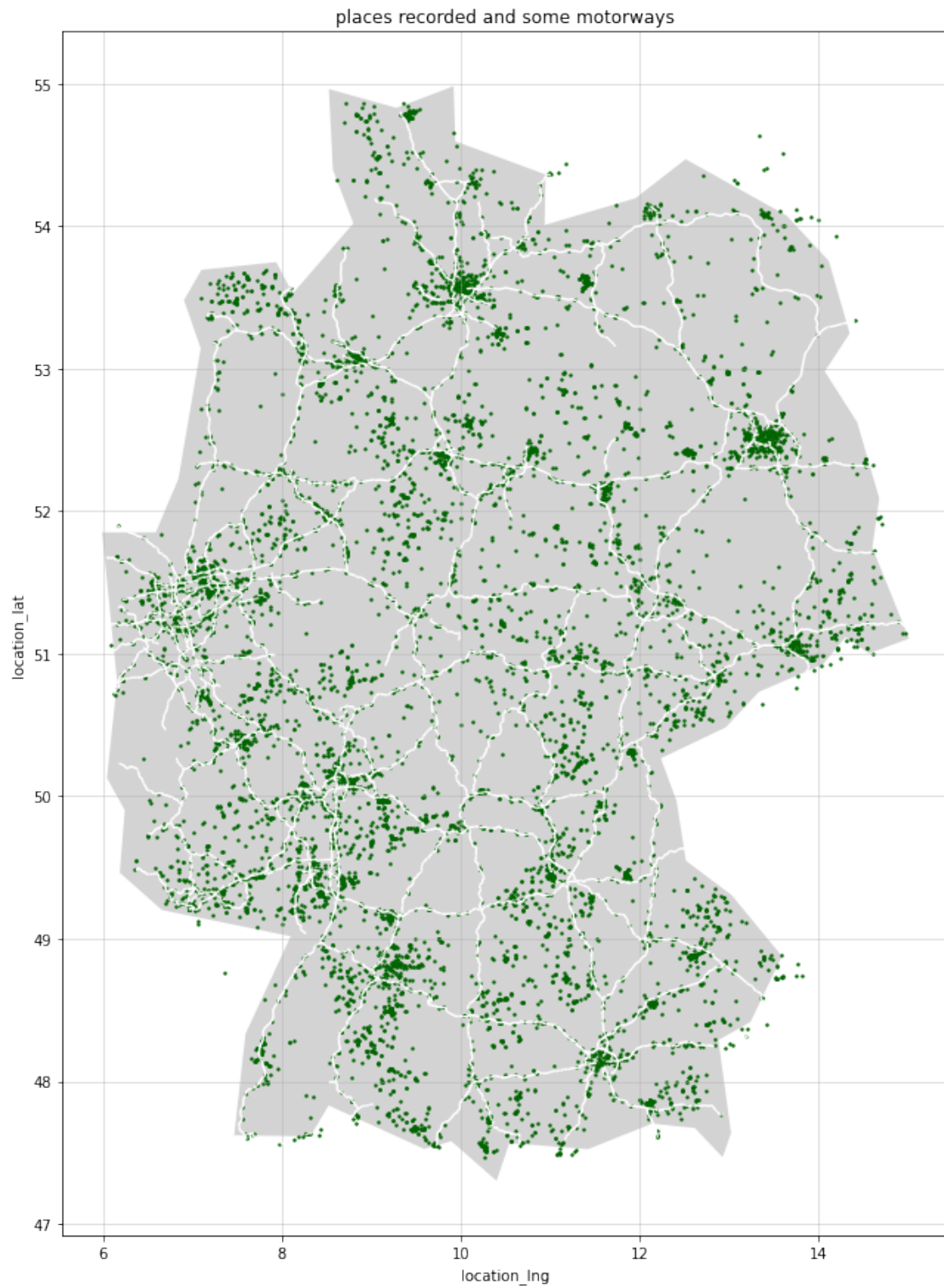
1.2.4 load a simple shape of Germany

```
      pop_est continent     name iso_a3  gdp_md_est  \
121  80594017    Europe  Germany    DEU   3979000.0

                                          geometry
121  POLYGON ((14.11969 53.75703, 14.35332 53.24817...
```

1.2.5 plot all three layers into one

places recorded and some motorways

1.2.6 Conclusion

The locations of the saved places seem to be relatively well disseminated across Germany. A few concentrations can be identified, which broadly correspond to metropolitan areas. There also appear to be sufficient locations along the motorways.

All in all, I believe that the available data should provide a representative basis for the intended analyses. Both the amount of available data and the assignable characteristics (e.g. proximity to a motorway, gas station, service area, etc.) seem to be sufficient to be able to draw valid conclusions.

Finally close the db connection