data_understanding_and_preparation

December 9, 2021

```python
[262]: import os
       import matplotlib.pyplot as plt
       import pandas as pd
       import geopandas as gpd
       import earthpy as et
```

```python
[263]: os.chdir(os.path.join(et.io.HOME, 'mahbub-dev/
       →beuth_hochschule_for_technik_berlin/berlin_datathon_2021/data'))
```

```python
[264]: count_parking_spots_estimated_kfz_path = os.path.
       →join("Bezirksamt-20211120T113159Z-001", "Bezirksamt", "parking",␣
       →"estimated_parking_spots_kfz.geojson")
       count_parking_spots_estimated_kfz = gpd.
       →read_file(count_parking_spots_estimated_kfz_path)
       count_parking_spots_estimated_kfz.dtypes
```

```
[264]: highway                object
       name                   object
       parking                object
       orientation            object
       position               object
       capacity               object
       source_orientation     object
       source_position        object
       source_capacity        object
       geometry             geometry
       dtype: object
```

```python
[265]: count_parking_spots_estimated_kfz.head(6)
```

```
[265]:      highway                        name parking orientation   position  \
       0  secondary  Platz der Vereinten Nationen    lane    parallel  on_street
       1  secondary  Platz der Vereinten Nationen    lane    parallel  on_street
       2  secondary  Platz der Vereinten Nationen    lane    parallel  on_street
       3  secondary  Platz der Vereinten Nationen    lane    parallel  on_street
       4  secondary  Platz der Vereinten Nationen    lane    parallel  on_street
       5   tertiary                  Thaerstraße    lane    parallel  on_street
```

```
    capacity source_orientation source_position source_capacity  \
0          2             estimated        estimated        estimated
1          2             estimated        estimated        estimated
2          3             estimated        estimated        estimated
3          3             estimated        estimated        estimated
4          3             estimated        estimated        estimated
5         19             estimated        estimated        estimated

                          geometry
0  POINT (393464.636 5820319.350)
1  POINT (393465.659 5820324.449)
2  POINT (393467.819 5820340.710)
3  POINT (393468.221 5820345.894)
4  POINT (393468.623 5820351.079)
5  POINT (395184.661 5820162.413)
```

[266]:
```python
registered_cars_path = os.path.join("Bezirksamt-20211120T113159Z-001",
  "Bezirksamt", "demographics", "registered_cars.geojson")
registered_cars = gpd.read_file(registered_cars_path)
registered_cars.dtypes
```

[266]:
```
district                         object
lor                              object
Bezirksregion                    object
Prognoseraum                     object
lor size in m²                  float64
inhabitants_total                 int64
of_those_inhabitants_18+          int64
vehicles_overall                  int64
cars_only                         int64
vehicles_per_1000_inhabitants     int64
cars_per_1000_inhabitants         int64
geometry                       geometry
dtype: object
```

[267]:
```python
registered_cars.head(6)
```

[267]:
```
                   district              lor            Bezirksregion  \
0  Friedrichshain-Kreuzberg       Wrangelkiez        Südliche Luisenstadt
1  Friedrichshain-Kreuzberg    Stralauer Kiez   Frankfurter Allee Süd FK
2  Friedrichshain-Kreuzberg    Wassertorplatz    Südliche Friedrichstadt
3  Friedrichshain-Kreuzberg      Oranienplatz       Nördliche Luisenstadt
4  Friedrichshain-Kreuzberg      Mehringplatz    Südliche Friedrichstadt
5  Friedrichshain-Kreuzberg  Askanischer Platz   Südliche Friedrichstadt

         Prognoseraum  lor size in m²  inhabitants_total  \
```

```
0        Kreuzberg Ost        526529.00              10979
1  Friedrichshain Ost        1354726.79              7752
2      Kreuzberg Nord         388873.74              5065
3       Kreuzberg Ost         480676.24              8847
4      Kreuzberg Nord        1036267.66             13833
5      Kreuzberg Nord         707419.20              6428


   of_those_inhabitants_18+  vehicles_overall  cars_only  \
0                      9433              2375       1838
1                      6739              4904       3621
2                      4086              1113        908
3                      7537              2230       1694
4                     11037              3904       3499
5                      5395              2309       2113


   vehicles_per_1000_inhabitants  cars_per_1000_inhabitants  \
0                            216                        167
1                            632                        467
2                            219                        179
3                            252                        191
4                            282                        252
5                            359                        328


                                             geometry
0  POLYGON ((393564.961 5817715.429, 393969.009 5…
1  POLYGON ((394458.458 5817914.213, 394513.006 5…
2  POLYGON ((391133.382 5817632.388, 391157.070 5…
3  POLYGON ((392393.802 5818353.854, 392395.806 5…
4  POLYGON ((390442.484 5818691.844, 390468.602 5…
5  POLYGON ((389687.563 5818256.442, 389681.790 5…
```

```python
[268]: kfz_lor_planungsraum_path = os.path.join("neukoelln", "kfz_lor_planungsraum.
       ↪geojson") # parking_way, parking_area
       df_kfz_lor_planungsraum = gpd.read_file(kfz_lor_planungsraum_path)
       #Planungsraum = lor
       df_kfz_lor_planungsraum.dtypes
```

```
[268]: Schlüssel                       object
       Bezirk                          object
       Planungsraum                    object
       Bezirksregion                   object
       Prognoseraum                    object
       Flächengröße in m²             float64
       Einwohner insgesamt              int64
       darunter 18 Jahre und älter      int64
       Kraftfahrzeuge insgesamt         int64
       darunter Pkw                     int64
```

```
Kfz pro 1000 EW                     int64
Pkw pro 1000 EW                     int64
geometry                         geometry
dtype: object
```

[269]: `df_kfz_lor_planungsraum.head(6)`

[269]:
```
   Schlüssel                  Bezirk      Planungsraum  Bezirksregion Prognoseraum  \
0  10010310  Marzahn-Hellersdorf       Alt-Marzahn    Marzahn Süd      Marzahn
1  10010101  Marzahn-Hellersdorf      Marzahn West   Marzahn Nord      Marzahn
2  10010206  Marzahn-Hellersdorf     Ringkolonnaden  Marzahn Mitte     Marzahn
3  10010204  Marzahn-Hellersdorf     Wuhletalstraße  Marzahn Mitte     Marzahn
4  10010102  Marzahn-Hellersdorf     Havemannstraße  Marzahn Nord      Marzahn
5  10010311  Marzahn-Hellersdorf     Landsberger Tor  Marzahn Süd      Marzahn

   Flächengröße in m²  Einwohner insgesamt  darunter 18 Jahre und älter  \
0          2238368.26                18290                        15454
1           694985.47                 5676                         4528
2          1486009.41                17341                        14565
3           845563.37                10622                         8488
4          1898742.52                18508                        14659
5          1608201.44                 3694                         3028

   Kraftfahrzeuge insgesamt  darunter Pkw  Kfz pro 1000 EW  Pkw pro 1000 EW  \
0                      7132          6457              389              353
1                      1959          1783              345              314
2                      5947          5387              342              310
3                      3439          3174              323              298
4                      5514          5090              297              275
5                      1634          1469              442              397

                                            geometry
0  POLYGON ((13.53692 52.53470, 13.53739 52.53467…
1  POLYGON ((13.54462 52.56693, 13.54471 52.56694…
2  POLYGON ((13.54802 52.55098, 13.54883 52.55074…
3  POLYGON ((13.55390 52.55776, 13.55476 52.55748…
4  POLYGON ((13.55825 52.56300, 13.55882 52.56279…
5  POLYGON ((13.58610 52.54937, 13.58616 52.54943…
```

[270]:
```python
parking_way_path = os.path.join("neukoelln", "parking_way.geojson") #␣
 ↪parking_way, parking_area
df_parking_way = gpd.read_file(parking_way_path)
df_parking_way.dtypes
```

[270]:
```
condition           object
parking             object
orientation         object
```

```
condition:other        object
position               object
condition:other:time   object
source:capacity        object
oneway_direction       object
highway                object
highway:name           object
maxstay                object
capacity                int64
osm-location           object
osm_id                 object
id                      int64
length                float64
geometry              geometry
dtype: object
```

[271]: `df_parking_way.head(6)`

[271]:

	condition	parking	orientation	condition:other	position
0	free	street_side	perpendicular	None	street_side
1	free	street_side	parallel	None	street_side
2	free	street_side	diagonal	None	street_side
3	free	street_side	diagonal	None	street_side
4	free	street_side	perpendicular	no	street_side
5	free	street_side	perpendicular	no	street_side

	condition:other:time	source:capacity	oneway_direction	highway
0	None	OSM	None	residential
1	None	estimated	None	residential
2	None	OSM	None	tertiary
3	None	OSM	None	tertiary
4	Apr-Sep: We 14:00-18:00	OSM	None	living_street
5	Apr-Sep: We 14:00-18:00	OSM	None	living_street

	highway:name	maxstay	capacity	osm-location	osm_id	id
0	Schaffhausener Straße	None	38	separate	way/37722270	1
1	Körtestraße	None	6	separate	way/851610744	2
2	Wiener Straße	None	40	separate	way/554762162	3
3	Wiener Straße	None	6	separate	way/554762163	4
4	Böckhstraße	None	3	separate	way/867318940	5
5	Böckhstraße	None	4	separate	way/867318940	6

	length	geometry
0	90.389	LINESTRING (13.41520 52.46092, 13.41389 52.46107)
1	31.143	LINESTRING (13.41283 52.49142, 13.41240 52.49152)
2	111.645	LINESTRING (13.43235 52.49734, 13.43379 52.49685)
3	20.422	LINESTRING (13.43490 52.49647, 13.43516 52.49638)

```
4     6.762  LINESTRING (13.41570 52.49460, 13.41579 52.49458)
5     8.970  LINESTRING (13.41611 52.49450, 13.41623 52.49447)
```

[272]:
```
print(count_parking_spots_estimated_kfz.shape)
print(registered_cars.shape)
# Make a spatial join
df_district_friedrichshain_kreuzberg = gpd.
 ↪sjoin(count_parking_spots_estimated_kfz, registered_cars, how="inner",␣
 ↪predicate="within")
df_district_friedrichshain_kreuzberg.shape
```

```
(70086, 10)
(26, 12)
```

[272]: (66868, 22)

[273]:
```
df_district_neukölln = df_kfz_lor_planungsraum.query("Bezirk == 'Neukölln'")
print(df_parking_way.shape)
print(df_district_neukölln.shape)
# Make a spatial join
df_district_neukölln = gpd.sjoin(df_parking_way, df_district_neukölln,␣
 ↪how="inner", predicate="intersects")
df_district_neukölln.shape
```

```
(5210, 17)
(40, 13)
```

[273]: (3396, 30)