

Updated on 8th March, 2024: changed the data with the new classification

Classfying the postal codes into urban/rural typology

In [92]: *# importing libraries*

```
import pandas as pd
import numpy as np
```

Urban/rural typology of NUTS3 regions

There are two types of classification: RLK and KTU

- **Link for RLK:**

<https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschlanc>

- **Link for KTU:**

<https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschlanc/kreistypen/kreistypen.html?sessionId=3D7A77E9A53718068EDE70ABECEBD732.live11292>

In [146]...

```
# Loading the csv file for Urban/rural typology of NUTS3 regions
df_class = pd.read_csv('RLK_KTU_classification.csv', encoding='cp1252')

df_class.head(9)
```

Out[146]:

	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022
0	DEF01	Flensburg, Stadt	zentral	Dünn besiedelter ländlicher Kreis
1	DEF02	Kiel, Stadt	zentral	kreisfreie Großstadt
2	DEF03	Lübeck, Stadt	sehr zentral	kreisfreie Großstadt
3	DEF04	Neumünster, Stadt	zentral	Ländlicher Kreis mit Verdichtungsansätzen
4	DEF05	Dithmarschen	sehr peripher	Dünn besiedelter ländlicher Kreis
5	DEF06	Herzogtum Lauenburg	zentral	Ländlicher Kreis mit Verdichtungsansätzen
6	DEF07	Nordfriesland	sehr peripher	Dünn besiedelter ländlicher Kreis
7	DEF08	Ostholstein	zentral	Ländlicher Kreis mit Verdichtungsansätzen
8	DEF09	Pinneberg	sehr zentral	Städtischer Kreis

In [147]...

```
df_class['NUTS1_ID'] = df_class['NUTS3_ID'].astype(str).str[:3]
df_class.head(9)
```

Out[147]:

	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID
0	DEF01	Flensburg, Stadt	zentral	Dünn besiedelter ländlicher Kreis	DEF
1	DEF02	Kiel, Stadt	zentral	kreisfreie Großstadt	DEF
2	DEF03	Lübeck, Stadt	sehr zentral	kreisfreie Großstadt	DEF
3	DEF04	Neumünster, Stadt	zentral	Ländlicher Kreis mit Verdichtungsansätzen	DEF
4	DEF05	Dithmarschen	sehr peripher	Dünn besiedelter ländlicher Kreis	DEF
5	DEF06	Herzogtum Lauenburg	zentral	Ländlicher Kreis mit Verdichtungsansätzen	DEF
6	DEF07	Nordfriesland	sehr peripher	Dünn besiedelter ländlicher Kreis	DEF
7	DEF08	Ostholstein	zentral	Ländlicher Kreis mit Verdichtungsansätzen	DEF
8	DEF09	Pinneberg	sehr zentral	Städtischer Kreis	DEF

In [149... df_class['NUTS1_ID'].unique()

Out[149]: array(['DEF', 'DE6', 'DE9', 'DE5', 'DEA', 'DE7', 'DEB', 'DE1', 'DE2', 'DEC', 'DE3', 'DE4', 'DE8', 'DED', 'DEE', 'DEG'], dtype=object)

In [150... df_class['NUTS1_ID'].nunique()

Out[150]: 16

In [151... df_class.dtypes

Out[151]:

```

NUTS3_ID      object
NUTS3_NAME    object
RLK2022       object
KTU2022       object
NUTS1_ID      object
dtype: object

```

In [152... *# this is a shorter list than the previous one*
len(df_class)

Out[152]: 400

In [153... *# Loading the file with NUTS3 ID and postal codes*

```

df_plz = pd.read_csv('pc2020_DE_NUTS-2021_v4.0.csv', sep= ";")

#df_plz[['NUTS3', 'postal_code']] = df_plz['NUTS3;CODE'].str.split(';', expand=True)

df_plz['NUTS3'] = df_plz['NUTS3'].str.replace('\'', '')
df_plz['CODE'] = df_plz['CODE'].str.replace('\'', '')

#df_plz = df_plz[['NUTS3', 'postal_code']]

df_plz.head(9)

```

Out[153]:

	NUTS3	CODE
0	DEA1D	41363
1	DEA1E	41366
2	DEA1E	41748
3	DEA1E	41749
4	DEA1E	41751
5	DEA1E	41747
6	DEA23	50667
7	DEA23	50668
8	DEA23	50670

In [154... df_plz.dtypes

Out[154]:

NUTS3	object
CODE	object
dtype:	object

In [155... df_plz['NUTS1_ID'] = df_plz['NUTS3'].astype(str).str[:3]
df_plz

Out[155]:

	NUTS3	CODE	NUTS1_ID
0	DEA1D	41363	DEA
1	DEA1E	41366	DEA
2	DEA1E	41748	DEA
3	DEA1E	41749	DEA
4	DEA1E	41751	DEA
...
8315	DE722	35641	DE7
8316	DE724	35080	DE7
8317	DE725	35315	DE7
8318	DE914	38543	DE9
8319	DE712	60312	DE7

8320 rows × 3 columns

In [156... df_plz['NUTS1_ID'].unique()

Out[156]:

```
array(['DEA', 'DED', 'DE8', 'DEE', 'DEF', 'DE9', 'DE4', 'DE5', 'DE7',
      'DEB', 'DE1', 'DE2', 'DEG', 'DEC', 'DE3', 'DE6'], dtype=object)
```

In [157... df_plz['NUTS1_ID'].nunique()

Out[157]: 16

```
In [158... len(df_plz['CODE'])
```

Out[158]: 8320

```
In [159... # Create a separate dataframe for DE4
de4_plz = df_plz[df_plz['NUTS1_ID'] == 'DE4']
de4_plz
```

Out[159]:

	NUTS3	CODE	NUTS1_ID
197	DE405	16352	DE4
198	DE40A	16565	DE4
199	DE40D	16918	DE4
320	DE40I	03202	DE4
1360	DE406	15755	DE4
...
8111	DE409	15366	DE4
8162	DE40D	16835	DE4
8174	DE40F	19309	DE4
8209	DE407	03253	DE4
8229	DE40F	19357	DE4

227 rows × 3 columns

```
In [160... de4_class = df_class[df_class['NUTS1_ID'] == 'DE4']
de4_class
```

Out[160]:

	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID
325	DE401	Brandenburg an der Havel, Stadt	peripher	Ländlicher Kreis mit Verdichtungsansätzen	DE4
326	DE402	Cottbus, Stadt	peripher	Ländlicher Kreis mit Verdichtungsansätzen	DE4
327	DE403	Frankfurt (Oder), Stadt	peripher	Dünn besiedelter ländlicher Kreis	DE4
328	DE404	Potsdam, Stadt	sehr zentral	kreisfreie Großstadt	DE4
329	DE405	Barnim	zentral	Ländlicher Kreis mit Verdichtungsansätzen	DE4
330	DE406	Dahme-Spreewald	zentral	Dünn besiedelter ländlicher Kreis	DE4
331	DE407	Elbe-Elster	sehr peripher	Dünn besiedelter ländlicher Kreis	DE4
332	DE408	Havelland	zentral	Dünn besiedelter ländlicher Kreis	DE4
333	DE409	Märkisch-Oderland	zentral	Dünn besiedelter ländlicher Kreis	DE4
334	DE40A	Oberhavel	zentral	Dünn besiedelter ländlicher Kreis	DE4
335	DE40B	Oberspreewald-Lausitz	peripher	Dünn besiedelter ländlicher Kreis	DE4
336	DE40C	Oder-Spree	peripher	Dünn besiedelter ländlicher Kreis	DE4
337	DE40D	Ostprignitz-Ruppin	sehr peripher	Dünn besiedelter ländlicher Kreis	DE4
338	DE40E	Potsdam-Mittelmark	zentral	Ländlicher Kreis mit Verdichtungsansätzen	DE4
339	DE40F	Prignitz	sehr peripher	Dünn besiedelter ländlicher Kreis	DE4
340	DE40G	Spree-Neiße	sehr peripher	Ländlicher Kreis mit Verdichtungsansätzen	DE4
341	DE40H	Teltow-Fläming	zentral	Ländlicher Kreis mit Verdichtungsansätzen	DE4
342	DE40I	Uckermark	sehr peripher	Dünn besiedelter ländlicher Kreis	DE4

Check: NUTS3 ID for DE4 in PLZ and the classification data

In [161]:

```
# NUTS3 in PLZ data: this seems correct
```

```
np.sort(de4_plz['NUTS3'].unique())
```

Out[161]:

```
array(['DE401', 'DE402', 'DE403', 'DE404', 'DE405', 'DE406', 'DE407',
       'DE408', 'DE409', 'DE40A', 'DE40B', 'DE40C', 'DE40D', 'DE40E',
       'DE40F', 'DE40G', 'DE40H', 'DE40I'], dtype=object)
```

In [162]:

```
# NUTS3 in classification data
```

```
np.sort(de4_class['NUTS3_ID'].unique())
```

```
Out[162]: array(['DE401', 'DE402', 'DE403', 'DE404', 'DE405', 'DE406', 'DE407',
      'DE408', 'DE409', 'DE40A', 'DE40B', 'DE40C', 'DE40D', 'DE40E',
      'DE40F', 'DE40G', 'DE40H', 'DE40I'], dtype=object)
```

Merge the classification data with the PLZ data

```
In [163... # Creating NUTS1 ID column
```

```
df_plz.head()
```

```
Out[163]:
```

	NUTS3	CODE	NUTS1_ID
0	DEA1D	41363	DEA
1	DEA1E	41366	DEA
2	DEA1E	41748	DEA
3	DEA1E	41749	DEA
4	DEA1E	41751	DEA

```
In [164... df_class.head()
```

```
Out[164]:
```

	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID
0	DEF01	Flensburg, Stadt	zentral	Dünn besiedelter ländlicher Kreis	DEF
1	DEF02	Kiel, Stadt	zentral	kreisfreie Großstadt	DEF
2	DEF03	Lübeck, Stadt	sehr zentral	kreisfreie Großstadt	DEF
3	DEF04	Neumünster, Stadt	zentral	Ländlicher Kreis mit Verdichtungsansätzen	DEF
4	DEF05	Dithmarschen	sehr peripher	Dünn besiedelter ländlicher Kreis	DEF

```
In [165... ## Merging the PLZ and classification data after fixing the discrepancies in DE3 NUTS3
```

```
df_with_null = pd.merge(df_plz, df_class, left_on = ['NUTS3'], right_on = ['NUTS3_ID'])
df_with_null = df_with_null[['CODE', 'NUTS3_ID', 'NUTS3_NAME', 'RLK2022', 'KTU2022']]
```

```
df_with_null.head(10)
```

Out[165]:

	CODE	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022
0	41363	DEA1D	Rhein-Kreis Neuss	sehr zentral	Städtischer Kreis
1	41366	DEA1E	Viersen	sehr zentral	Städtischer Kreis
2	41748	DEA1E	Viersen	sehr zentral	Städtischer Kreis
3	41749	DEA1E	Viersen	sehr zentral	Städtischer Kreis
4	41751	DEA1E	Viersen	sehr zentral	Städtischer Kreis
5	41747	DEA1E	Viersen	sehr zentral	Städtischer Kreis
6	50667	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt
7	50668	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt
8	50670	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt
9	41836	DEA29	Heinsberg	sehr zentral	Städtischer Kreis

In [166... `len(df_with_null)`

Out[166]: 8320

In [167... `df_with_null['NUTS1_ID'] = df_with_null['NUTS3_ID'].astype(str).str[:3]`
`df_with_null.head(10)`

Out[167]:

	CODE	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID
0	41363	DEA1D	Rhein-Kreis Neuss	sehr zentral	Städtischer Kreis	DEA
1	41366	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA
2	41748	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA
3	41749	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA
4	41751	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA
5	41747	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA
6	50667	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA
7	50668	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA
8	50670	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA
9	41836	DEA29	Heinsberg	sehr zentral	Städtischer Kreis	DEA

In [168... `df_with_null['NUTS1_ID'].unique()`

Out[168]: `array(['DEA', 'DED', 'DE8', 'DEE', 'DEF', 'DE9', 'DE4', 'DE5', 'DE7',
'DEB', 'DE1', 'DE2', 'DEG', 'DEC', 'DE3', 'DE6', 'nan'],
dtype=object)`

In [169... `df_with_null['NUTS1_ID'].nunique()`

Out[169]: 17

In [170]...

```
df_with_null['NUTS2_ID'] = df_with_null['NUTS3_ID'].astype(str).str[:4]
df_with_null.head(10)
```

Out[170]:

	CODE	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID	NUTS2_ID
0	41363	DEA1D	Rhein-Kreis Neuss	sehr zentral	Städtischer Kreis	DEA	DEA1
1	41366	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1
2	41748	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1
3	41749	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1
4	41751	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1
5	41747	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1
6	50667	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2
7	50668	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2
8	50670	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2
9	41836	DEA29	Heinsberg	sehr zentral	Städtischer Kreis	DEA	DEA2

In [171]...

```
df_with_null['NUTS2_ID'].unique()
```

Out[171]:

```
array(['DEA1', 'DEA2', 'DEA3', 'DEA4', 'DEA5', 'DED4', 'DE80', 'DEE0',
      'DEF0', 'DE93', 'DE40', 'DE91', 'DE92', 'DE94', 'DE50', 'DE71',
      'DEB1', 'DEB2', 'DEB3', 'DE11', 'DE12', 'DE13', 'DE14', 'DE21',
      'DE22', 'DE27', 'DE23', 'DE24', 'DE25', 'DED2', 'DED5', 'DEG0',
      'DE26', 'DE72', 'DE73', 'DEC0', 'DE30', 'DE60', 'nan'],
      dtype=object)
```

In [172]...

```
df_with_null['NUTS2_ID'].nunique()
```

Out[172]:

39

In [173]...

```
df_with_null[df_with_null['NUTS1_ID'] == 'DE4']
```


Out[173]:

	CODE	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID	NUTS2_ID
197	16352	DE405	Barnim	zentral	Ländlicher Kreis mit Verdichtungsansätzen	DE4	DE40
198	16565	DE40A	Oberhavel	zentral	Dünn besiedelter ländlicher Kreis	DE4	DE40
199	16918	DE40D	Ostprignitz-Ruppin	sehr peripher	Dünn besiedelter ländlicher Kreis	DE4	DE40
320	03202	DE40I	Uckermark	sehr peripher	Dünn besiedelter ländlicher Kreis	DE4	DE40
1360	15755	DE406	Dahme-Spreewald	zentral	Dünn besiedelter ländlicher Kreis	DE4	DE40
...
8111	15366	DE409	Märkisch-Oderland	zentral	Dünn besiedelter ländlicher Kreis	DE4	DE40
8162	16835	DE40D	Ostprignitz-Ruppin	sehr peripher	Dünn besiedelter ländlicher Kreis	DE4	DE40
8174	19309	DE40F	Prignitz	sehr peripher	Dünn besiedelter ländlicher Kreis	DE4	DE40
8209	03253	DE407	Elbe-Elster	sehr peripher	Dünn besiedelter ländlicher Kreis	DE4	DE40
8229	19357	DE40F	Prignitz	sehr peripher	Dünn besiedelter ländlicher Kreis	DE4	DE40

227 rows × 7 columns

In [174... df_plz[df_plz['CODE'] == '66629']

Out[174]:

	NUTS3	CODE	NUTS1_ID
1695	DEC06	66629	DEC

In [175... df_class[df_class['NUTS1_ID'] == 'DEC']

Out[175]:

	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID
318	DEC01	Regionalverband Saarbrücken	sehr zentral	Städtischer Kreis	DEC
319	DEC02	Merzig-Wadern	zentral	Städtischer Kreis	DEC
320	DEC03	Neunkirchen	zentral	Städtischer Kreis	DEC
321	DEC04	Saarlouis	zentral	Städtischer Kreis	DEC
322	DEC05	Saarpfalz-Kreis	zentral	Städtischer Kreis	DEC
323	DEC06	St. Wendel	zentral	Städtischer Kreis	DEC

In [176... # missing: DEC01, DEC04, DEC05, DEC06

```
df_class_null = df_with_null[df_with_null['RLK2022'].isnull()]
df_class_null.head()
```

Out[176]:

	CODE	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID	NUTS2_ID
5500	99817	NaN	NaN	NaN	NaN	nan	nan

Adding other NUTS level

used file: NUTS2021

source: <https://ec.europa.eu/eurostat/de/web/nuts/background>

In [177... *## Adding NUTS1*

```
NUTS1 = pd.read_csv('NUTS1.csv', encoding='cp1252')
NUTS1
```

Out[177]:

	Code 2021	NUTS level 1
0	DE1	Baden-Württemberg
1	DE2	Bayern
2	DE3	Berlin
3	DE4	Brandenburg
4	DE5	Bremen
5	DE6	Hamburg
6	DE7	Hessen
7	DE8	Mecklenburg-Vorpommern
8	DE9	Niedersachsen
9	DEA	Nordrhein-Westfalen
10	DEB	Rheinland-Pfalz
11	DEC	Saarland
12	DED	Sachsen
13	DEE	Sachsen-Anhalt
14	DEF	Schleswig-Holstein
15	DEG	Thüringen
16	DEZ	Extra-Regio NUTS 1

In [178... *## Adding NUTS2*

```
NUTS2 = pd.read_csv('NUTS2.csv', encoding='cp1252')
NUTS2
```

Out[178]:

	Code 2021	NUTS level 2
0	DE11	Stuttgart
1	DE12	Karlsruhe
2	DE13	Freiburg
3	DE14	Tübingen
4	DE21	Oberbayern
5	DE22	Niederbayern
6	DE23	Oberpfalz
7	DE24	Oberfranken
8	DE25	Mittelfranken
9	DE26	Unterfranken
10	DE27	Schwaben
11	DE30	Berlin
12	DE40	Brandenburg
13	DE50	Bremen
14	DE60	Hamburg
15	DE71	Darmstadt
16	DE72	Gießen
17	DE73	Kassel
18	DE80	Mecklenburg-Vorpommern
19	DE91	Braunschweig
20	DE92	Hannover
21	DE93	Lüneburg
22	DE94	Weser-Ems
23	DEA1	Düsseldorf
24	DEA2	Köln
25	DEA3	Münster
26	DEA4	Detmold
27	DEA5	Arnsberg
28	DEB1	Koblenz
29	DEB2	Trier
30	DEB3	Rheinhessen-Pfalz
31	DEC0	Saarland
32	DED2	Dresden
33	DED4	Chemnitz

	Code 2021	NUTS level 2
34	DED5	Leipzig
35	DEE0	Sachsen-Anhalt
36	DEF0	Schleswig-Holstein
37	DEG0	Thüringen
38	DEZZ	Extra-Regio NUTS 2

In [179... `df_with_null.head()`

Out[179]:

	CODE	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID	NUTS2_ID
0	41363	DEA1D	Rhein-Kreis Neuss	sehr zentral	Städtischer Kreis	DEA	DEA1
1	41366	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1
2	41748	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1
3	41749	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1
4	41751	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1

In [180... *# Left join to add the region names*

```
df_with_null = pd.merge(df_with_null, NUTS1, left_on = ['NUTS1_ID'], right_on = ['Code
df_with_null.head(10)
```

Out[180]:

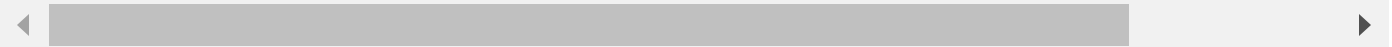
	CODE	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID	NUTS2_ID	Code 2021	NUTS level 1
0	41363	DEA1D	Rhein-Kreis Neuss	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein-Westfalen
1	41366	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein-Westfalen
2	41748	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein-Westfalen
3	41749	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein-Westfalen
4	41751	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein-Westfalen
5	41747	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein-Westfalen
6	50667	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2	DEA	Nordrhein-Westfalen
7	50668	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2	DEA	Nordrhein-Westfalen
8	50670	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2	DEA	Nordrhein-Westfalen
9	41836	DEA29	Heinsberg	sehr zentral	Städtischer Kreis	DEA	DEA2	DEA	Nordrhein-Westfalen

In [181...

```
df_with_null = pd.merge(df_with_null, NUTS2, left_on = ['NUTS2_ID'], right_on = ['Code
df_with_null.head(10)
```

Out[181]:

	CODE	NUTS3_ID	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID	NUTS2_ID	Code 2021_x	NUTS level 1
0	41363	DEA1D	Rhein-Kreis Neuss	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein- Westfalen
1	41366	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein- Westfalen
2	41748	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein- Westfalen
3	41749	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein- Westfalen
4	41751	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein- Westfalen
5	41747	DEA1E	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	DEA	Nordrhein- Westfalen
6	50667	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2	DEA	Nordrhein- Westfalen
7	50668	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2	DEA	Nordrhein- Westfalen
8	50670	DEA23	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2	DEA	Nordrhein- Westfalen
9	41836	DEA29	Heinsberg	sehr zentral	Städtischer Kreis	DEA	DEA2	DEA	Nordrhein- Westfalen



In [182...

```
df_with_null = df_with_null[['NUTS3_ID', 'CODE', 'NUTS3_NAME', 'RLK2022', 'KTU2022' ],
df_with_null.head(10)
```

Out[182]:

	NUTS3_ID	CODE	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID	NUTS2_ID	NUTS level 1	NUT level
0	DEA1D	41363	Rhein-Kreis Neuss	sehr zentral	Städtischer Kreis	DEA	DEA1	Nordrhein-Westfalen	Düsseldorf
1	DEA1E	41366	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	Nordrhein-Westfalen	Düsseldorf
2	DEA1E	41748	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	Nordrhein-Westfalen	Düsseldorf
3	DEA1E	41749	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	Nordrhein-Westfalen	Düsseldorf
4	DEA1E	41751	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	Nordrhein-Westfalen	Düsseldorf
5	DEA1E	41747	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	Nordrhein-Westfalen	Düsseldorf
6	DEA23	50667	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2	Nordrhein-Westfalen	Köln
7	DEA23	50668	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2	Nordrhein-Westfalen	Köln
8	DEA23	50670	Köln, Stadt	sehr zentral	kreisfreie Großstadt	DEA	DEA2	Nordrhein-Westfalen	Köln
9	DEA29	41836	Heinsberg	sehr zentral	Städtischer Kreis	DEA	DEA2	Nordrhein-Westfalen	Köln

In [183...

df_with_null = df_with_null.rename(columns={'CODE': 'POSTAL_CODE', 'NUTS level 1': 'NUTS1_NAME'})
df_with_null.head()

Out[183]:

	NUTS3_ID	POSTAL_CODE	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID	NUTS2_ID	NUTS1_NAME
0	DEA1D	41363	Rhein-Kreis Neuss	sehr zentral	Städtischer Kreis	DEA	DEA1	Nordrhein-Westfalen
1	DEA1E	41366	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	Nordrhein-Westfalen
2	DEA1E	41748	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	Nordrhein-Westfalen
3	DEA1E	41749	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	Nordrhein-Westfalen
4	DEA1E	41751	Viersen	sehr zentral	Städtischer Kreis	DEA	DEA1	Nordrhein-Westfalen

In [184...

len(df_with_null)

Out[184]:

8320

missing classification

```
In [185... df_class_null = df_with_null[df_with_null['RLK2022'].isnull()]
df_class_null.head()
```

```
Out[185]:
```

	NUTS3_ID	POSTAL_CODE	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID	NUTS2_ID	NUTS1_NAME
5500	NaN	99817	NaN	NaN	NaN	nan	nan	...

```
In [186... len(df_class_null)
```

```
Out[186]: 1
```

```
In [187... df_class_null['NUTS1_ID'].unique()
```

```
Out[187]: array(['nan'], dtype=object)
```

```
In [188... df_not_null = df_with_null[df_with_null['RLK2022'].notnull()]
len(df_not_null)
```

```
Out[188]: 8319
```

```
In [189... df_class_null2 = df_with_null[df_with_null['KTU2022'].isnull()]
df_class_null2.head()
```

```
Out[189]:
```

	NUTS3_ID	POSTAL_CODE	NUTS3_NAME	RLK2022	KTU2022	NUTS1_ID	NUTS2_ID	NUTS1_NAME
5500	NaN	99817	NaN	NaN	NaN	nan	nan	...

```
In [190... len(df_class_null2)
```

```
Out[190]: 1
```

```
In [194... df_not_null_final = df_not_null[['POSTAL_CODE', 'NUTS3_ID', 'NUTS1_NAME', 'NUTS2_NAME',
df_not_null_final.head()
```


Out[194]:

	POSTAL_CODE	NUTS3_ID	NUTS1_NAME	NUTS2_NAME	NUTS3_NAME	RLK2022	KTU2022
0	41363	DEA1D	Nordrhein-Westfalen	Düsseldorf	Rhein-Kreis Neuss	sehr zentral	Städtischer Kreis
1	41366	DEA1E	Nordrhein-Westfalen	Düsseldorf	Viersen	sehr zentral	Städtischer Kreis
2	41748	DEA1E	Nordrhein-Westfalen	Düsseldorf	Viersen	sehr zentral	Städtischer Kreis
3	41749	DEA1E	Nordrhein-Westfalen	Düsseldorf	Viersen	sehr zentral	Städtischer Kreis
4	41751	DEA1E	Nordrhein-Westfalen	Düsseldorf	Viersen	sehr zentral	Städtischer Kreis

In [195...

```
len(df_not_null_final[df_not_null_final['KTU2022'].isnull()])
```

Out[195]:

0

In [196...

```
df_not_null_final.to_csv('new_classification_urban_by_postal_code.csv', encoding='cp12
```

In []: