

```
In [1]: import pandas as pd
import geopandas as gpd
import numpy as np
import time
from tqdm import tqdm
```

```
/Users/lorenzo/opt/anaconda3/lib/python3.9/site-packages/pandas/core/
computation/expressions.py:21: UserWarning: Pandas requires version '
2.8.4' or newer of 'numexpr' (version '2.8.3' currently installed).
  from pandas.core.computation.check import NUMEXPR_INSTALLED
/Users/lorenzo/opt/anaconda3/lib/python3.9/site-packages/pandas/core/
arrays/masked.py:60: UserWarning: Pandas requires version '1.3.6' or
newer of 'bottleneck' (version '1.3.5' currently installed).
  from pandas.core import (
```

```
In [20]: HST = pd.read_csv("/Volumes/T7/Water Project/npdes_downloads/NPDES_QNC
dtype = {'YEARQTR' : str})
```

```
In [27]: HST
```

```
Out[27]:
```

	NPDES_ID	YEARQTR	HLRNC	NUME90Q	NUMCVDT	NUMSVCD	NUMPSCH
0	AK0000086	19932		1	0	0	0
1	AK0000086	19951		0	0	0	2
2	AK0000086	19952		0	0	0	2
3	AK0000086	19953		0	0	0	2
4	AK0000086	19954		0	0	0	2
...
6895200	WVR106895	20182	C	0	0	0	0
6895201	WVR106895	20183	C	0	0	0	0
6895202	WVR106895	20184	C	0	0	0	0
6895203	WVR106895	20191	C	0	0	0	0
6895204	WVR106895	20192	C	0	0	0	0

6895205 rows × 7 columns

```
In [23]: HST["HLRNC"].value_counts()
```

```
Out[23]: HLRNC
U      2148017
N      1150327
W       776577
C       774225
D      434064
V      403096
        370081
R      311400
E      165503
S      160047
P      129139
X       55285
T       17025
Q         419
Name: count, dtype: int64
```

```
In [4]: twtyears = HST[HST['YEARQTR'].str[:4] > "2003"]
```

```
In [5]: IDS = twtyears["NPDES_ID"].unique().tolist()
```

```
In [6]: len(IDS)
```

```
Out[6]: 552531
```

```
In [7]: #for i in IDS:
        #   SNC = ["S", "E", "X", "T", "D", "W"]
        #   data = twtyears[twtyears["NPDES_ID"] == i]
        #   print(i)
        #   print(len(data))
        #   print(len(data[data["HLRNC"].isin(SNC)]))
```

```
In [8]: TRIAL1 = {"NPDES_ID" : list(),
                  "TOT_QTR_REP": list(),
                  "QTR_IN_SNC" : list()}

SNC = ["S", "E", "X", "T", "D", "W"]
```

```
In [9]: grouped = twtyears.groupby("NPDES_ID")

# Use tqdm to display progress bar
for i, data in tqdm(grouped):
    table = data["HLRNC"].value_counts()
    TRIAL1["NPDES_ID"].append(i)
    TRIAL1["TOT_QTR_REP"].append(table.sum())
    TRIAL1["QTR_IN_SNC"].append(table[table.index.isin(SNC)].sum())

100%|████████████████████████████████████████| 552531/552531 [01:39<00:00, 5572.45it/s]
```

```
In [10]: inf_QTR = pd.DataFrame.from_dict(TRIAL1)
```

```
In [11]: inf_QTR["PERCENT"] = 100 * inf_QTR["QTR_IN_SNC"] / inf_QTR["TOT_QTR_REP"]
```

```
In [25]: hundred = inf_QTR[inf_QTR["PERCENT"] == 100]
```

```
In [26]: np.mean(hundred["TOT_QTR_REP"])
```

```
Out[26]: 14.050573325286663
```

```
In [13]: Geos = pd.read_csv("/Volumes/T7/Water Project/npdes_downloads/ICIS_FAC")
```

```
In [14]: Geos.columns
```

```
Out[14]: Index(['ICIS_FACILITY_INTEREST_ID', 'NPDES_ID', 'FACILITY_UIN',
                'FACILITY_TYPE_CODE', 'FACILITY_NAME', 'LOCATION_ADDRESS',
                'SUPPLEMENTAL_ADDRESS_TEXT', 'CITY', 'COUNTY_CODE', 'STATE_CODE',
                'ZIP',
                'GEOCODE_LATITUDE', 'GEOCODE_LONGITUDE', 'IMPAIRED_WATERS'],
              dtype='object')
```

```
In [15]: Geos = Geos[["NPDES_ID", "FACILITY_NAME", "CITY",
                      "COUNTY_CODE", "STATE_CODE", "ZIP",
                      "GEOCODE_LATITUDE", "GEOCODE_LONGITUDE"]]
```

```
In [16]: merged = inf_QTR.merge(Geos, on = "NPDES_ID", how = "right")
```

In [17]: merged[merged]

Out[17]:

DT_QTR_REP	QTR_IN_SNC	PERCENT	FACILITY_NAME	CITY	COUNTY_CODE	STAT
2.0	0.0	0.00	HAWKINS FILL SITE	SHEPHERDSVILLE	KY029	
16.0	1.0	6.25	E BATON ROUGE CITY-PAR (SOUTH)	BATON ROUGE	LA033	
NaN	NaN	NaN	DEVELOPMENT AT 65 SUNNYSIDE AVENUE	HOLDEN	MA027	
NaN	NaN	NaN	DUARTA WASTE REPAIR SHOP	BAYAMON	PR021	
NaN	NaN	NaN	BURDICK PAVING	ROOSEVELT	UT013	
...
NaN	NaN	NaN	RETREAT AT MORSE	NOBLESVILLE	NaN	
NaN	NaN	NaN	AIRPORT EXPY & ARDMORE AVE STORMWATER IMPROVEM...	FORT WAYNE	NaN	
NaN	NaN	NaN	ADDITIONS AND RENOVATIONS TO CARROLL HIGH SCHOOL	FORT WAYNE	NaN	
NaN	NaN	NaN	NIVER CREEK TRIBUTARY M IMPROVEMENTS	THORNTON	NaN	
1.0	0.0	0.00	DOLLAR GENERAL - WHITWELL	WHITWELL	TN115	

ans

In [18]: merged.to_csv("/Volumes/T7/Water Project/NPDES_FINAL.csv")