

# Collecting weather data from an API

## About the data

In this notebook, we will be collecting daily weather data from the National Centers for Environmental Information (NCEI) API. We will use the Global Historical Climatology Network - Daily (GHCND) data set; see the documentation [here](#).

Note: The NCEI is part of the National Oceanic and Atmospheric Administration (NOAA) and, as you can see from the URL for the API, this resource was created when the NCEI was called the NCDC. Should the URL for this resource change in the future, you can search for the NCEI weather API to find the updated one.

Paste your token below.

```
In [7]: import requests

def make_req(endpoint, payload=None):
    return requests.get(
        f'https://www.ncdc.noaa.gov/cdo-web/api/v2/{endpoint}',
        headers={
            'token' : 'COGqDwKIRDvdlxMHVhxgqTsoJqtCoAfz'
        }, params=payload
    )
```

```
In [8]: #Let's check if the API is working properly

response = make_req('datasets', {'startdate': '2018-10-01'})
response.status_code
```

Out[8]: 200

## Collect All Data Points for 2018 In NYC (Various Stations)

We can make a loop to query for all the data points one day at a time. Here we create a list of all the results:

```
In [9]: import datetime
from IPython import import display as dis

current = datetime.date(2018,1,1)
end = datetime.date(2019,1,1)

results = []

while current < end :
    dis.clear_output(wait=True)
```

```

dis.display(f'gathering data for {str(current)}')

response = make_req(
    'data',{
        'datasetid':'GHCND',
        'locationid':'CITY:US360019',
        'startdate':current,
        'enddate':current,
        'units':'metric',
        'limit':1000
    }
)
if response.ok:
    results.extend(response.json()['results'])

current += datetime.timedelta(days=1)

```

'gathering data for 2018-12-31'

Now, we can create a dataframe with all this data. Notice there are multiple stations with values for each datatype on a given day. We don't know what the stations are, but we can look them up and add them to the data:

```

In [27]: import pandas as pd

df = pd.DataFrame(results)
df.head()

```

```

Out[27]:

```

	date	datatype	station	attributes	value
0	2018-01-01T00:00:00	PRCP	GHCND:US1CTFR0039	„N,0800	0.0
1	2018-01-01T00:00:00	PRCP	GHCND:US1NJBG0015	„N,1050	0.0
2	2018-01-01T00:00:00	SNOW	GHCND:US1NJBG0015	„N,1050	0.0
3	2018-01-01T00:00:00	PRCP	GHCND:US1NJBG0017	„N,0920	0.0
4	2018-01-01T00:00:00	SNOW	GHCND:US1NJBG0017	„N,0920	0.0

Save this data to a file:

```

In [28]: path = df.to_csv('nyc_weather_2018.csv', index=False)

```

and write it to the database:

```

In [29]: import sqlite3 as sq3

with sq3.connect('weather.db') as connection:
    df.to_sql(
        'weather',connection, index=False, if_exists='replace'
    )

```

For learning about merging dataframes, we will also get the data mapping station IDs to information about the station:

```
In [30]: response = make_req(
    'stations',{
        'datasetid':'GHCND',
        'locationid':'CITY:US360019',
        'limit':1000
    }
)
# searched using the database(SQLite3) connection from previous cell ↑↑↑
stations = pd.DataFrame(response.json()['results'])[['id','name','latitude','longit
stations.to_csv('weather_stations.csv', index=False) # create a csv of the weather

with sq3.connect('weather.db') as connection:
    stations.to_sql('stations',connection,index=False, if_exists='replace')
```

```
In [26]: ws = pd.read_csv('weather_stations.csv')
ws
```

Out[26]:

	id	name	latitude	longitude	elevation
0	GHCND:US1CTFR0022	STAMFORD 2.6 SSW, CT US	41.064100	-73.577000	36.6
1	GHCND:US1CTFR0039	STAMFORD 4.2 S, CT US	41.037788	-73.568176	6.4
2	GHCND:US1NJBG0001	BERGENFIELD 0.3 SW, NJ US	40.921298	-74.001983	20.1
3	GHCND:US1NJBG0002	SADDLE BROOK TWP 0.6 E, NJ US	40.902694	-74.083358	16.8
4	GHCND:US1NJBG0003	TENAFLY 1.3 W, NJ US	40.914670	-73.977500	21.6
...	...	...	...	...	...
325	GHCND:USW00054787	FARMINGDALE REPUBLIC AIRPORT, NY US	40.734430	-73.416370	22.8
326	GHCND:USW00094728	NY CITY CENTRAL PARK, NY US	40.778980	-73.969250	42.7
327	GHCND:USW00094741	TETERBORO AIRPORT, NJ US	40.858980	-74.056160	0.8
328	GHCND:USW00094745	WESTCHESTER CO AIRPORT, NY US	41.062360	-73.704540	112.9
329	GHCND:USW00094789	JFK INTERNATIONAL AIRPORT, NY US	40.639150	-73.763900	2.7

330 rows × 5 columns