

Internsip | Summer 2025

IMPORTANT

You can use this **Jupyter** notebook *symbols_update.ipynb* to complete the assignment.

If you are not familiar with **Jupyter**, please feel free to develop your solution using standard **.py** file instead.

To complete the assignment, you will need to **pip install pandas** library.

Please upload all your work to public **GitHub** repository and share the link with us.

There are three .csv files that we want to upload sequentially to the database: **gas_fr_plants.csv**, **gas_plants.csv**, and **wind_plants.csv**. For the sake of simplicity, the database is represented by another .csv file **database.csv**. These three files contain volume data for certain power plants.

The goal is to complete **PowerPlants** class with three functions:

- **load_new_data_from_file** : loads and processes data from **symbols_update_n.csv** file
- **save_new_data** : saves loaded data to the **database.csv** file
- **get_data_from_database** : returns the most recently updated data for every symbol in **database.csv**
- **aggregate_data_to_monthly** : returns the monthly average, min and max for each wind or Gas plant **database.csv**
- **aggregate_data_to_country** : returns the country total power production by technology type **database.csv**

```
In [ ]: class PowerPlants(object):
        def __init__(self):
            self.database_file = 'database.csv'

        def load_new_data_from_file(self, file_path: str):
            pass

        def save_new_data(self, input_data: pd.DataFrame):
            pass

        def get_data_from_database(self):
            pass

        def aggregate_data_to_monthly(self):
            pass

        def aggregate_data_to_country(self):
            pass
```

```
In [ ]: pp = PowerPlants()
new_data = pp.load_new_data_from_file('wind_plants.csv')
su.save_new_data(new_data)
new_data = pp.load_new_data_from_file('gas_plants.csv')
su.save_new_data(new_data)
new_data = pp.load_new_data_from_file('gas_fr_plants.csv')
su.save_new_data(new_data)
```

Please note:

- *Country* column from **plants.csv** files are stored as full country names in **database.csv**
- There maybe some missing data that needs to be populated (please assume 0 for this)
- The most recent update for each symbol is returned based on the **updatetime** column
- Please see the below for the structure of expected results.

```
In [1]: from IPython.display import Image
Image('get_data_from_database.png')
```

Out[1]:

date	country	SiteName	Technology	updatedby	updatetime	Volume
2024-01-01	France	Blenod-5	Gas	petroineos	2025-05-29 11:06:00.802833	6753.000000
2024-01-02	France	Blenod-5	Gas	petroineos	2025-05-29 11:06:00.802833	3896.000000
2024-01-03	France	Blenod-5	Gas	petroineos	2025-05-29 11:06:00.802833	3636.000000
2024-01-04	France	Blenod-5	Gas	petroineos	2025-05-29 11:06:00.802833	5138.000000
2024-01-05	France	Blenod-5	Gas	petroineos	2025-05-29 11:06:00.802833	5265.000000
...
2025-04-21	Great Britain	Hornsea-2	Wind	petroineos	2025-05-29 11:06:01.001653	711.231619
2025-04-22	Great Britain	Hornsea-2	Wind	petroineos	2025-05-29 11:06:01.001653	808.534585
2025-04-23	Great Britain	Hornsea-2	Wind	petroineos	2025-05-29 11:06:01.001653	142.450340
2025-04-24	Great Britain	Hornsea-2	Wind	petroineos	2025-05-29 11:06:01.001653	392.082184
2025-04-25	Great Britain	Hornsea-2	Wind	petroineos	2025-05-29 11:06:01.001653	920.847877

```
In [2]: from IPython.display import Image
Image('month_average_min_mix.png')
```

Out[2]:

date	Blenod-5 Mean	Blenod-5 Min	Blenod-5 Max	Pembroke-1 Mean	Pembroke-1 Min	Pembroke-1 Max	Pembroke-2 Mean	Pembroke-2 Min	Pembroke-2 Max	Hornsea-1 Mean	Hornsea-1 Min	Hornsea-1 Max	Hornsea-2 Mean	Hornsea-2 Min	Hornsea-2 Max
2024-01-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2024-02-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2024-03-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2024-04-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2024-05-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2024-06-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2024-07-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2024-08-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2024-09-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2024-10-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2024-11-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2024-12-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2025-01-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2025-02-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2025-03-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
2025-04-01	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD

```
In [3]: from IPython.display import Image
Image('country_sum.png')
```

Out[3]:

		Volume	
country	Technology		
France	Gas	TBD	
Great Britain	Gas	TBD	
	Wind	TBD	

Good luck!!!

In []: