UNIVERSIDADE LUSÓFONA

“Watt” am I doing?

João Caldeira

Assistant Professor

Email. joao.caldeira@ulusofona.pt

March 19, 2025

“Watt” am I doing?

Current Situation

- **Inexperienced Data Scientists**
  - Ignore resource consumption
  - Lack skills to optimize software
  - Don’t have tools to easily help them improve

“Watt” am I doing?

Problem

```
supporters = pd.read_csv("supporters.csv")
```

✓ 0.0s

```
supporters.head()
```

✓ 0.0s Open 'supporters' in Data Wrangler

	First Name	Last Name	Age	Gender	Years_of_Support	Power_Voting
0	Alice	Brown	66	Male	5	10.0
1	Jack	Jackson	56	Male	52	7.0
2	Charlie	Johnson	61	Female	55	16.0
3	Alice	Anderson	70	Female	69	5.0
4	Alice	Johnson	75	Male	0	16.0

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 6 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   First Name        1000 non-null   object
 1   Last Name         1000 non-null   object
 2   Age               1000 non-null   int64
 3   Gender            1000 non-null   object
 4   Years_of_Support  1000 non-null   int64
 5   Power_Voting      1000 non-null   float64
dtypes: float64(1), int64(2), object(3)
memory usage: 47.0+ KB
```

```
# Get total memory usage in bytes
total_memory = supporters.memory_usage(deep=True).sum()
print(f"Total memory usage: {total_memory / 1024:.2f} KB")
```

✓ 0.0s

```
Total memory usage: 206.08 KB
```

“Watt” am I doing?

Problem

```
dtypes = { "First Name": "category", "Last Name": "category", "Age": "int8", "Gender": "category", "Years_of_Support": "int8", "Power_Voting": "int8" }
```
✓ 0.0s

```
supporters = pd.read_csv("supporters.csv", dtype=dtypes)
```
✓ 0.0s

```
supporters.head()
```
✓ 0.0s Open 'supporters' in Data Wrangler

	First Name	Last Name	Age	Gender	Years_of_Support	Power_Voting
0	Alice	Brown	66	Male	5	10
1	Jack	Jackson	56	Male	52	7
2	Charlie	Johnson	61	Female	55	16
3	Alice	Anderson	70	Female	69	5
4	Alice	Johnson	75	Male	0	16

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 6 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   First Name        1000 non-null   category
 1   Last Name         1000 non-null   category
 2   Age               1000 non-null   int8
 3   Gender            1000 non-null   category
 4   Years_of_Support  1000 non-null   int8
 5   Power_Voting      1000 non-null   int8
dtypes: category(3), int8(3)
memory usage: 6.8 KB
```

```
# Get total memory usage in bytes
total_memory = supporters.memory_usage(deep=True).sum()
print(f"Total memory usage: {total_memory / 1024:.2f} KB")
```
✓ 0.0s

```
Total memory usage: 8.02 KB
```

UNIVERSIDADE LUSÓFONA

“Watt” am I doing?

Solution

Benchmarking Energy Consumption

Assessing Energy in Python Tasks

Gonçalo Cabeleira & João Caldeira @ 2025

This Notebook runs multiple benchmarks on Data Science Code examples.

Code to Benchmark Python code

```python
from ul.core.tracker import Tracker
```

Python

```python
tracker = Tracker()
tracker.start()
```

Python

```python
# Python code here
```

Python

```python
tracker.stop()
```

Python

“Watt” am I doing?

Solution

- **Collection**
  - Hardware, Software & Data metrics
  - Product & Process metrics

```
timestamp,project_name,run_id,experiment_id,duration,emissions,emissions_rate,cpu_power,gpu_power,ram_power,cpu_energy,gpu_energy,ram_energy,
2025-03-02T19:55:48,codecarbon,b4681a7d-227e-46f4-99b9-c93047d8ab7e,5b0fa12a-3dd7-45bb-9766-cc326314d9f1,0.0596987079479731,1.308117143353209
2025-03-02T19:57:45,codecarbon,50459a0a-4e35-4076-b64f-65d7aa5688c4,5b0fa12a-3dd7-45bb-9766-cc326314d9f1,0.0525198749965056,1.150141044711092
2025-03-02T19:59:25,codecarbon,9c8df34c-f2f6-4daa-8d85-f95fe1b09aab,5b0fa12a-3dd7-45bb-9766-cc326314d9f1,0.10721020802157,2.370631580945304e-
2025-03-02T20:01:13,codecarbon,88ca660d-89cf-494d-a6c0-8bf1144d9e01,5b0fa12a-3dd7-45bb-9766-cc326314d9f1,0.0490317500079982,1.075388675196038
2025-03-02T20:03:40,codecarbon,52ee16ee-39ae-4114-ba8b-f0ecc6f6a365,5b0fa12a-3dd7-45bb-9766-cc326314d9f1,0.0667147919884882,1.468381633937921
2025-03-02T20:05:12,codecarbon,17bd6726-4235-42ee-86fd-afb0a1ef4410,5b0fa12a-3dd7-45bb-9766-cc326314d9f1,0.0545112910331226,1.193856384826657
2025-03-02T20:08:05,codecarbon,d9bf6503-0276-42ab-92eb-0a59445e2b7a,5b0fa12a-3dd7-45bb-9766-cc326314d9f1,0.0295651669730432,6.282353562739643
2025-03-02T20:09:15,codecarbon,0705078c-c118-45d3-a7e9-e01bd0b0fd52,5b0fa12a-3dd7-45bb-9766-cc326314d9f1,0.03220904094632715,7.00185056980518
```

“Watt” am I doing?

Solution

- **Contributions**
  - Design of novel metrics to assess notebooks quality.
  - Development of an artifact to collect metric data about resource consumption and execution times.
  - Publish a real dataset with notebooks execution data allowing for the creation of ML models able to predict resource consumption.

“Watt” am I doing?

Benefits

- **Valued Added**
  - Community can build ML models for prediction of Notebooks execution times and resource consumption.
  - Each Data Scientist can use it to assess the time and resource consumption of their notebooks without/before running them.

UNIVERSIDADE LUSÓFONA

UNIVERSIDADE LUSÓFONA

“Watt” am I doing?

João Caldeira

Assistant Professor

Email. joao.caldeira@ulusofona.pt

March 19, 2025