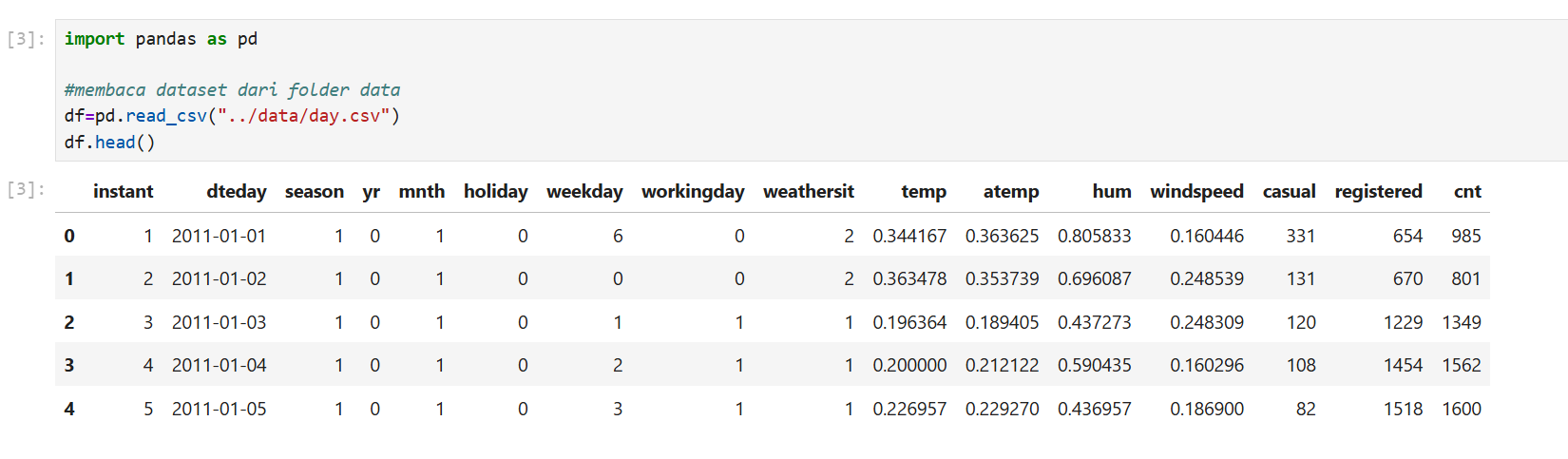
**EKSPLORASI AWAL BIKE SHARING DATASET DENGAN PANDAS DI JUPYTER NOTEBOOK**

Nama : Sulthan Nabil Al Hakim

TI : 02

NIM : 0110224234

1. **Membaca Dataset Dari Folder Data**



**PENJELASAN CODE**

**Input =**

Import pandas as pd 🡪 mengimpor library pandas untuk mengolah data tabular.

df=pd.read\_csv(“../data/day.csv”) 🡪 membaca file CSV bernama day.csv dari folder data.

df.head() 🡪 menampilkan 5 baris pertama dari dataset untuk melihat struktur datanya.

**Output =**

instant 🡪 indeks urutan hari.

dteday 🡪 tanggal.

season 🡪 musim (1=semi, 2=panas, 3=gugur, 4=dingin).

yr 🡪 tahun (0=2011, 1=2012).

mnth 🡪 bulan (1-12).

holiday 🡪 1 jika hari libur, 0 jika bukan.

weekday 🡪 hari dalam seminggu (0=minggu, 1=senin, dst).

workingday 🡪 1 jika hari kerja (bukan libur/weekend), 0 jika tidak.

weathersit 🡪 kondisi cuaca (1=cerah, 2=berawan, 3=hujan ringan, 4=hujan lebat)

temp, atemp 🡪 suhu aktual & terasa (dinormalisasi 0–1)

hum 🡪 kelembapan (0–1)

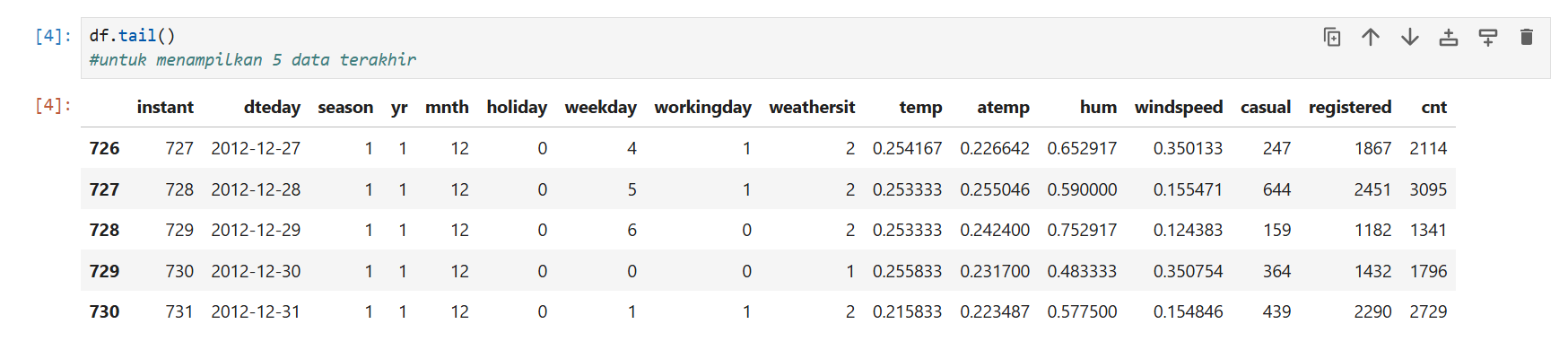
windspeed: kecepatan angin (0–1)

casual 🡪 jumlah peminjam sepeda non-member

registered 🡪 jumlah peminjam sepeda yang terdaftar

cnt 🡪 total peminjam sepeda (casual + registered)

1. **Melihat 5 Baris Terakhir Data**

****

**PENJELASAN CODE**

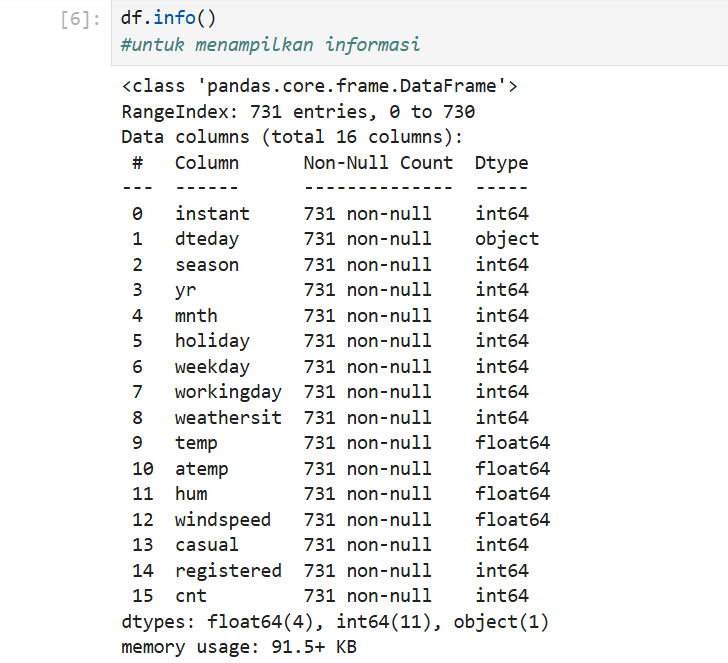
**Input =**

df.tail() 🡪 menampilkan 5 baris terakhir dari dataset. (untuk mengecek akhir periode data).

**Output =**

Data ditampilkan dari 27-31 Desember 2012 (totak ada 731 hari = 2 tahun data).

1. **Informasi Lengkap Dataset**

****

**PENJELASAN CODE**

**Input =**

df.info() 🡪 menampilkan ringkasan dataset :

* Jumlah baris (731)
* Jumlah kolom (16)
* Nama kolom, jumlah nilai non-null, dan tipe data (int, float, object).

**Output =**

Datasetnya punya 731 baris x 16 kolom.

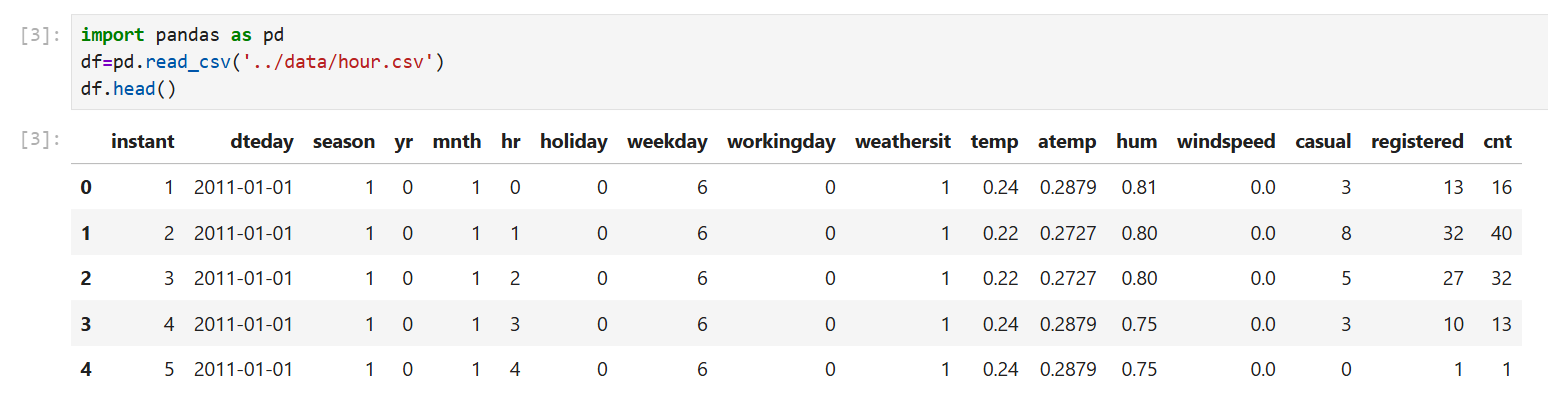
Tipe data :

11 kolom integer ( seperti season, yr, holiday, cnt ).

4 kolom float ( temp, atemp, hum, windspeed ).

1 kolom object ( dtday ).

1. **Membaca Dataset hour.csv**

****

**PENJELASAN CODE**

**Input =**

hour.csv() 🡪 membaca dataset kedua (data/jam).

df.head() 🡪 menampilkan 5 baris pertama

**Output =**

hr 🡪 jam (0-23). Total ada 8760 baris (24 jam x 365 hari).

|  |  |  |
| --- | --- | --- |
| instant | int | 🡪 Nomor indeks (ID baris data). |

|  |  |  |
| --- | --- | --- |
| dteday | object 🡪 | Tanggal (format YYYY-MM-DD). |

|  |  |  |
| --- | --- | --- |
| season | int | 🡪 Musim: 1=Semi, 2=Panas, 3=Gugur, 4=Dingin. |

|  |  |  |
| --- | --- | --- |
| yr | int | 🡪 Tahun: 0=2011, 1=2012. |

|  |  |  |
| --- | --- | --- |
| mnth | int | 🡪 Bulan (1–12). |

|  |  |  |
| --- | --- | --- |
| hr | int | 🡪 Jam (0–23). |

|  |  |  |
| --- | --- | --- |
| holiday | int 🡪 | 1 jika hari libur nasional, 0 jika bukan. |

|  |  |  |
| --- | --- | --- |
| weekday | int 🡪 | Hari dalam minggu (0=Minggu, 1=Senin, dst.). |

|  |  |  |
| --- | --- | --- |
| workingday | int 🡪 | 1 jika hari kerja (bukan libur/weekend), 0 jika tidak. |

|  |  |  |
| --- | --- | --- |
| Weathersit | int 🡪 | Kondisi cuaca: 1=Cerah, 2=Berawan, 3=Hujan ringan, 4=Hujan lebat. |

|  |  |  |
| --- | --- | --- |
| temp | float 🡪 | Suhu normalisasi (0–1), nilai asli sekitar 41°C. |

|  |  |  |
| --- | --- | --- |
| atemp | float 🡪 | Suhu yang terasa (apparent temperature), normalisasi (0–1). |

|  |  |  |
| --- | --- | --- |
| hum | float | 🡪 Kelembapan udara (0–1). |

|  |  |  |
| --- | --- | --- |
| Windspeed | float 🡪 | Kecepatan angin (0–1). |

|  |  |  |
| --- | --- | --- |
| casual | int | 🡪 Jumlah pengguna sepeda non-member (kasual). |

|  |  |  |
| --- | --- | --- |
| registered | int 🡪 | Jumlah pengguna sepeda yang terdaftar (member). |

|  |  |  |
| --- | --- | --- |
| cnt | int | 🡪 Total pengguna sepeda (casual + registered). |

**KESIMPULAN**

1. Dataset berhasil dibaca dengan benar menggunakan pandas.read\_csv(). Ada dua dataset yang dipakai, yaitu:

day.csv 🡪 data peminjaman sepeda per hari (731 baris, 16 kolom).

hour.csv 🡪 data peminjaman sepeda per jam (8760 baris, 17 kolom).

1. Struktur data sudah dipahami melalui:

df.head() 🡪 melihat 5 baris awal untuk mengenali pola data.

df.tail() 🡪 melihat 5 baris akhir untuk memastikan periode data berakhir di 31 Desember 2012.

df.info() 🡪 mengetahui jumlah baris, jumlah kolom, tipe data, serta memastikan tidak ada missing value.

1. Karakteristik dataset:

Dataset lengkap, tidak ada nilai kosong (non-null semua).

kolom berisi informasi waktu (tanggal, bulan, tahun, jam), kondisi cuaca, serta jumlah peminjaman sepeda (casual, registered, total).

Data siap digunakan untuk tahap berikutnya seperti visualisasi dan implementasi algoritma machine learning (misalnya regresi untuk prediksi jumlah peminjaman).

Link Github :  
<https://github.com/Mrhankim/Praktikum1_MachineLearning>