**PRAKTIKUM DATA WAREHOUSING DAN DATA MINING**

**MODUL 6**

**PENGENALAN APLIKASI DATA MINING**

****

**Disusun oleh:**

**Adinda Aulia Hapsari**

**L200220037**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS KOMUNIKASI DAN INFORMATIKA**

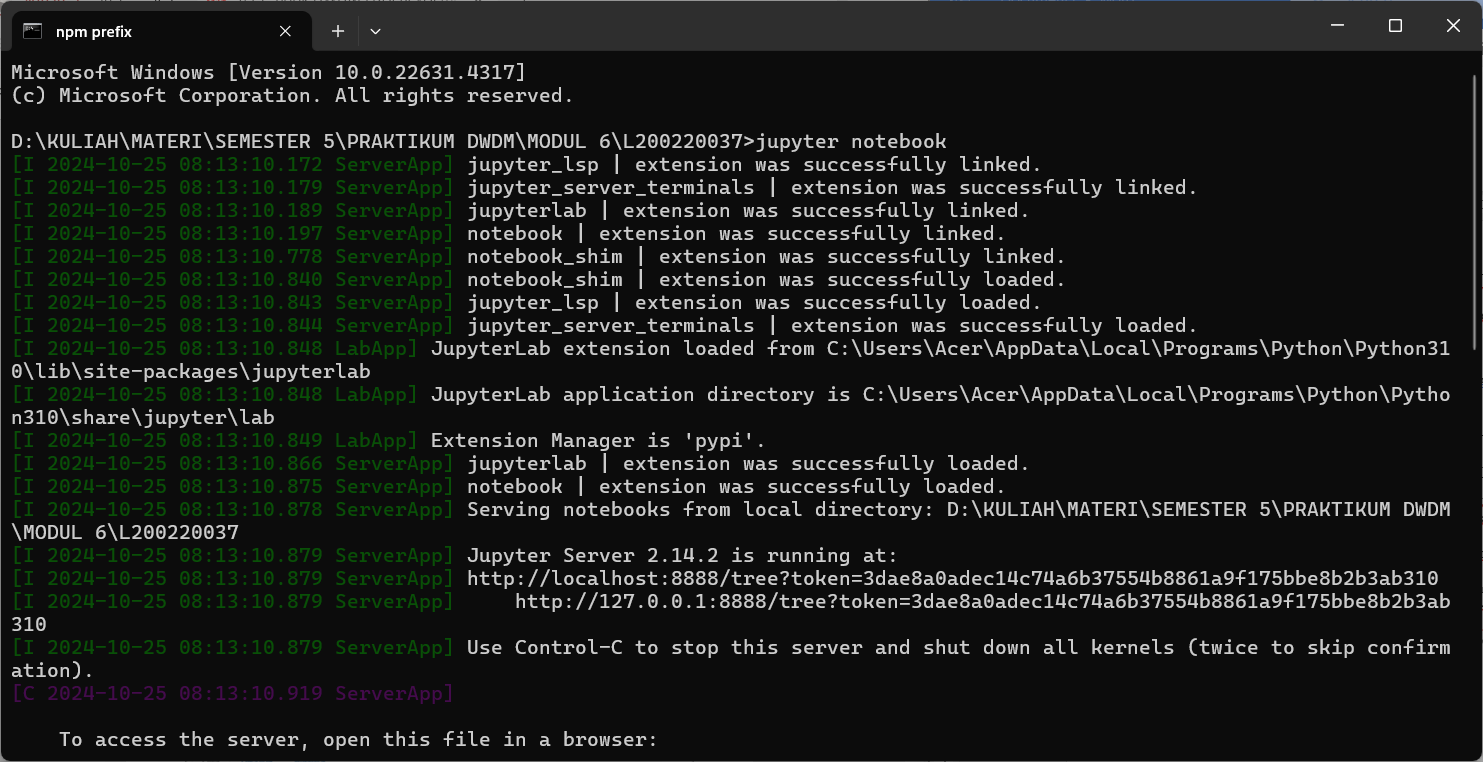
**UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**TAHUN 2024**

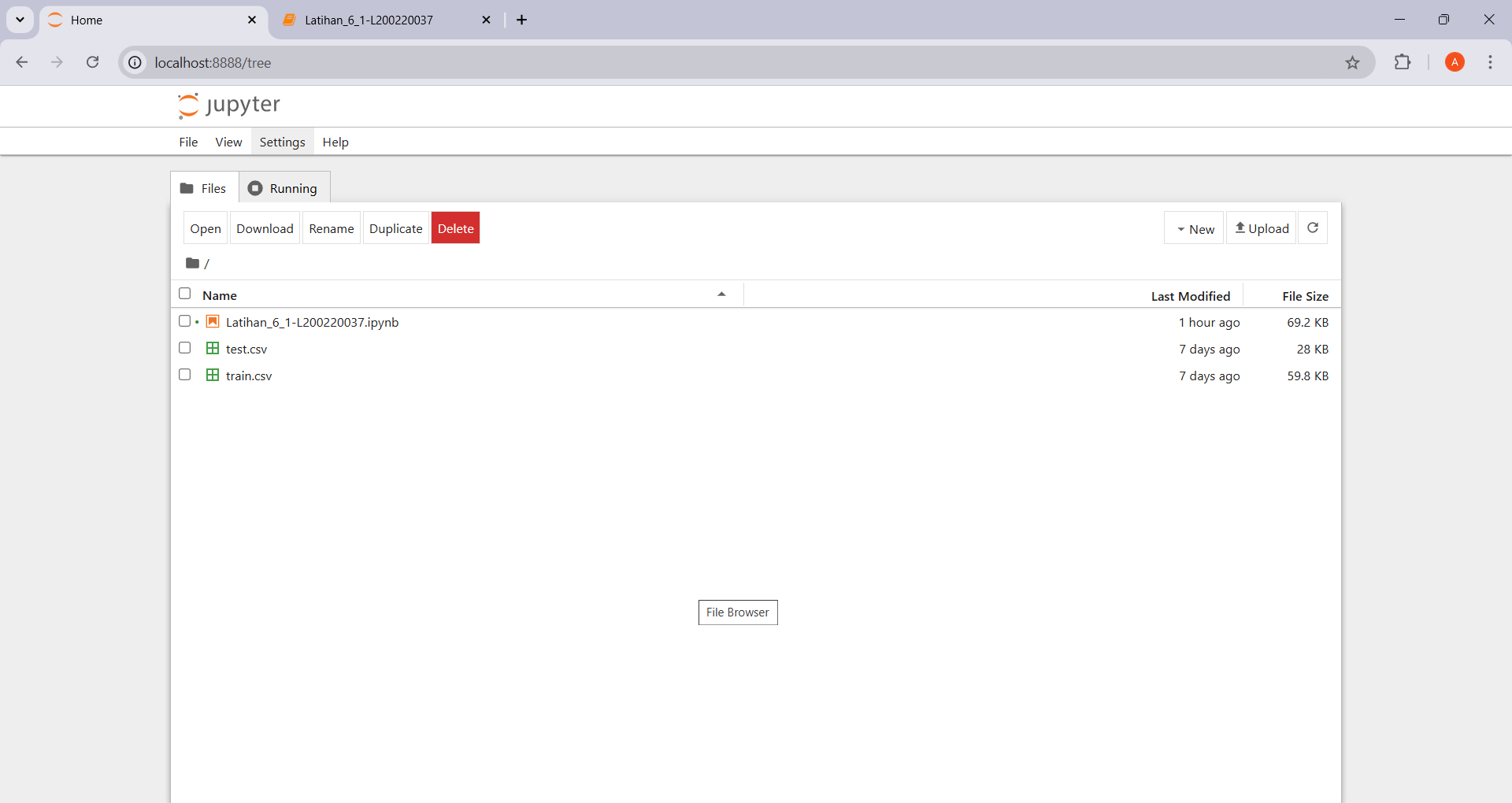
|  |  |
| --- | --- |
| Setelah kegiatan selesai, lembar kerja ini dicetak (di-print) dan dikumpulkan ke asisten.  NIM : L200220037  Nama : Adinda Aulia Hapsari  Nama Asisten : Diva Halimah  Tanggal Praktikum : 25 Oktober 2024 | (Diisi oleh Asisten)  Nilai Praktek :  Tanda Tangan : |

**KEGIATAN PRAKTIKUM**

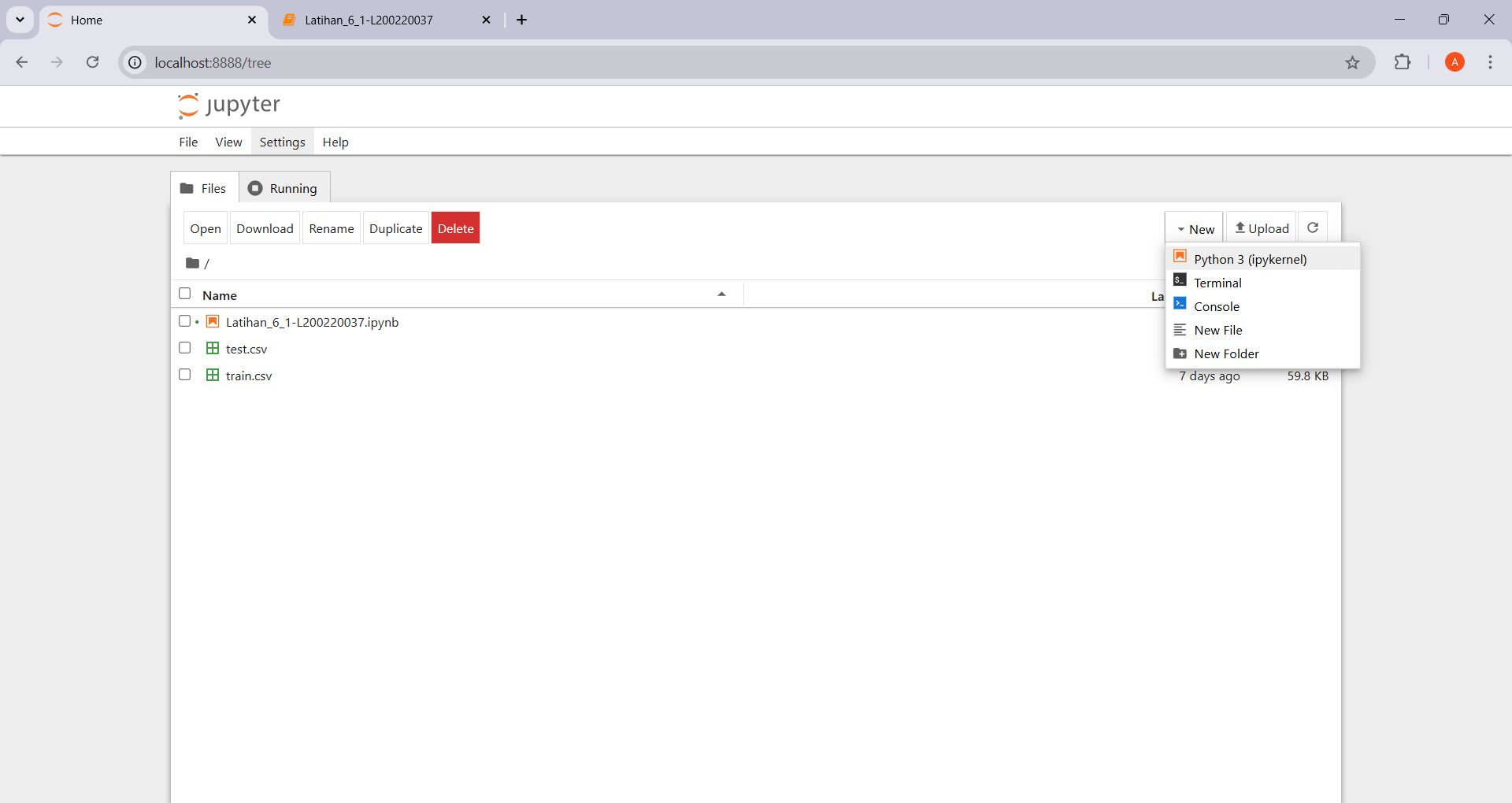
1. Buka Windows Explorer dan arahkan pada folder Praktikum Data Warehousing dan Data Mining. Buat sebuah folder sesuai dengan NIM mahasiswa di dalam folder Praktikum Data Warehousing dan Data Mining.
2. Gunakan data pelatihan dari dataset Titanic dengan nama train.csv yang telah diunduh dari repository gitea.
3. Simpan file data pelatihan tersebut pada folder NIM yang telah dibuat pada langkah 1.
4. Buka aplikasi Anaconda Navigator untuk menjalankan Jupyter Notebook. Atau jika menggunakan “command prompt”, mula-mula arahkan drive pada folder yang telah dibuat sesuai dengan NIM. Ketikkan perintah jupyter notebook kemudian tekan Enter.



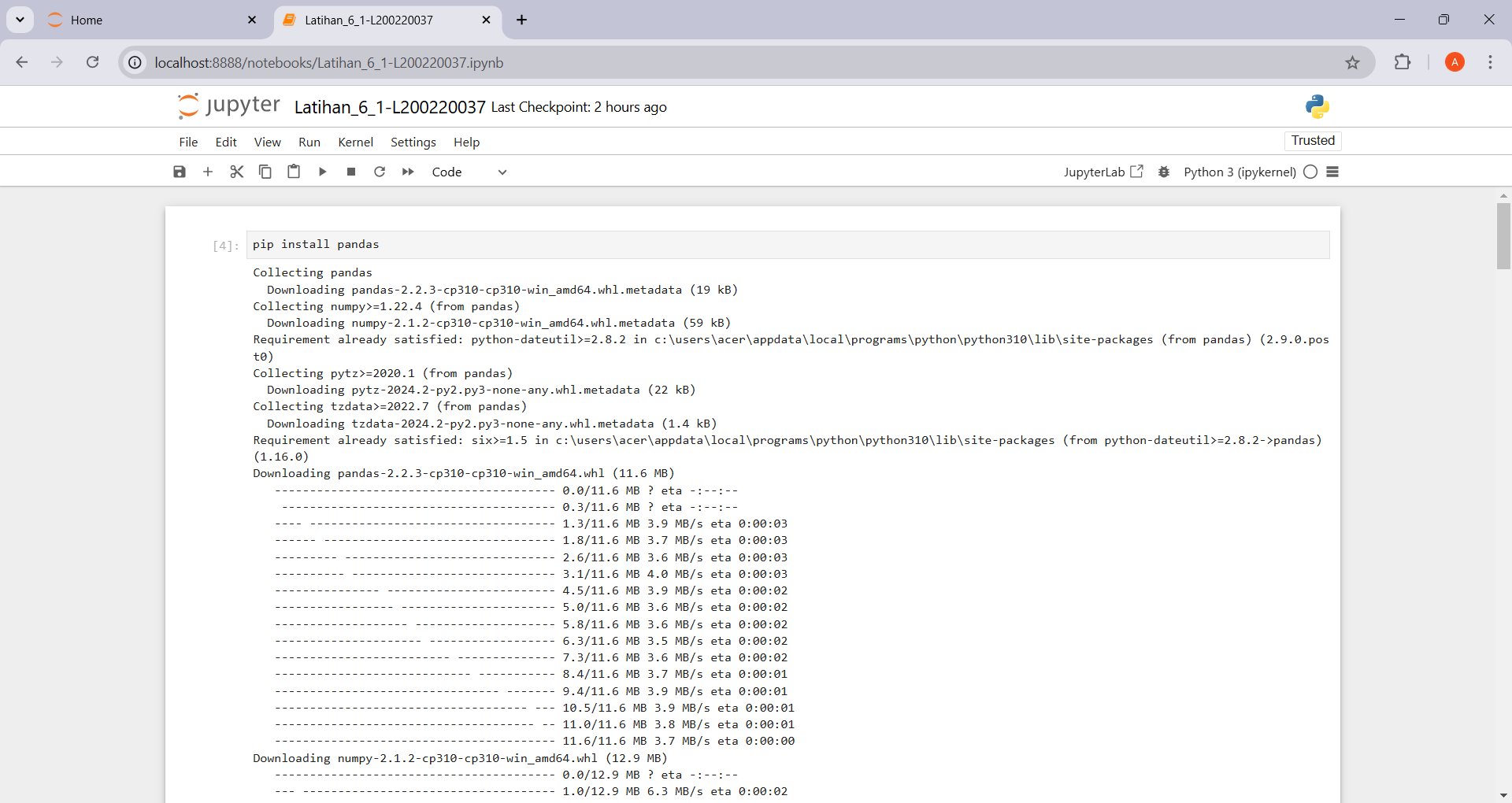
1. Aplikasi Jupyter Notebook akan dijalankan pada sebuah browser yang terinstal di komputer.



1. Untuk membuat file kerja baru untuk menulis kode-kode program, klik tombol New -> Python 2 atau 3. Dalam contoh ini, kita menggunakan Python 3.



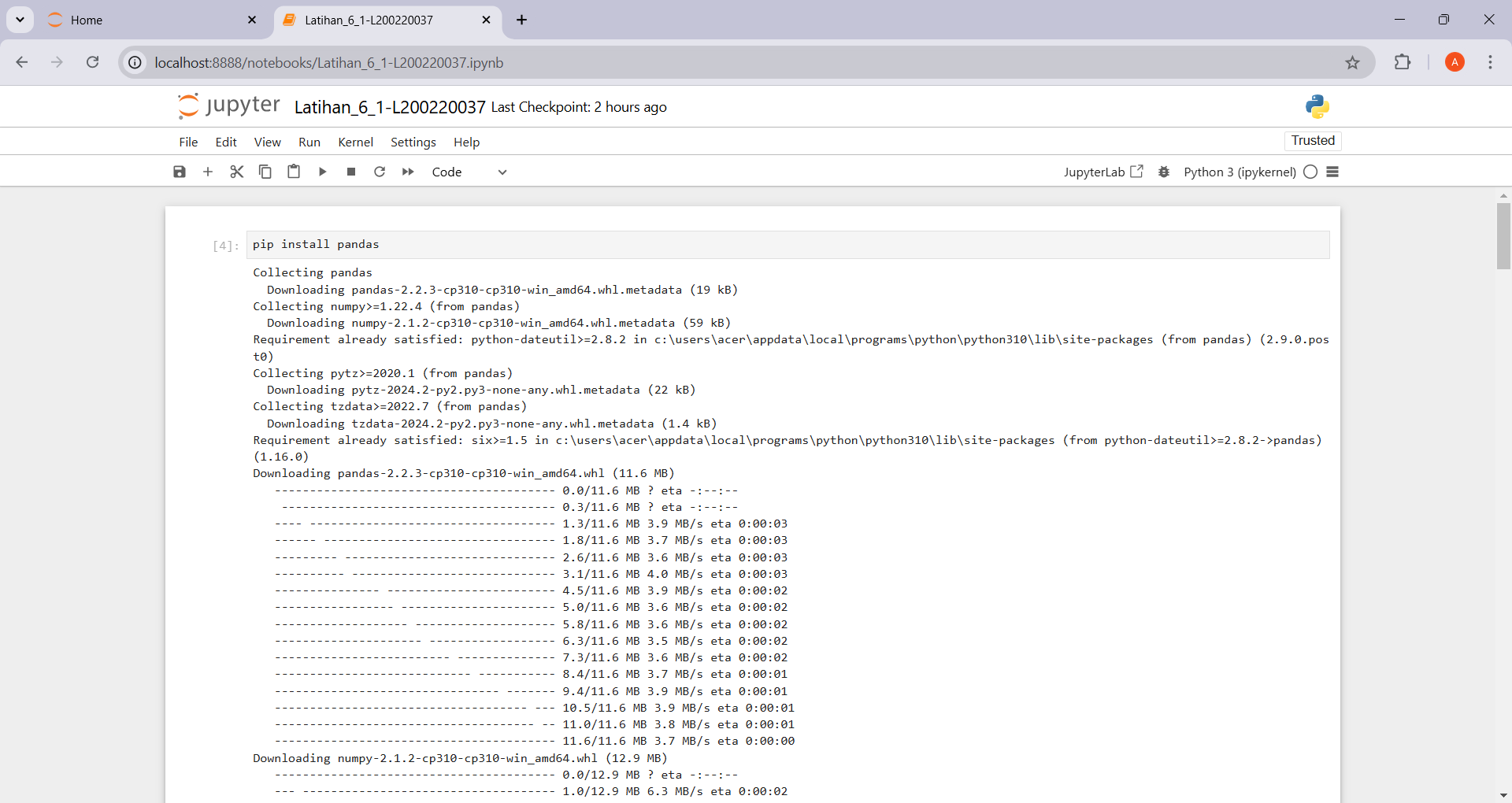
1. Sebuah tab baru pada browser akan ditampilkan. Tab baru ini menampilkan halaman code editor yang berfungsi sebagai tempat menulis kode-kode Python. Beberapa komponen penting dalam halaman ini antara lain:
   1. Nama file: digunakan untuk menentukan nama file yang akan memiliki ekstensi .ipynb.
   2. Toolbar: berisi beragam tombol untuk melakukan tugas-tugas tertentu, misalnya save, insert cell, cut cell, copy, paste, move cell up, move cell down, run cell, interrupt kernel, restart kernel, restart and rerun kernel, and style.
   3. Code cell: digunakan untuk menuliskan kode-kode program.
2. Dengan mengklik nama file “Untitled”, ubahlah nama file dengan format “Latihan 6\_1-NIM”, misalnya “Latihan\_6\_1-L2002201000”. File ini akan tersimpan di komputer dalam folder C:/Praktikum Data Warehousing dan Data Mining/NIM/Latihan 6\_1-L2002201000.ipynb. Setelah klik Rename, maka nama file akan berubah sesuai dengan yang diinputkan.



1. Pada percobaan pertama dalam Bab 6 ini, kita akan belajar bagaimana menampilkan data dari file CSV. Untuk mengeksekusi dengan Python, kita membutuhkan library pandas.
2. Pada sel pertama, ketikkan baris perintah berikut untuk mengimpor library pandas untuk mengolah dataframe. Klik tombol atau bisa dengan menekan tombol SHIFT+ENTER pada keyboard untuk mengeksekusi kode.

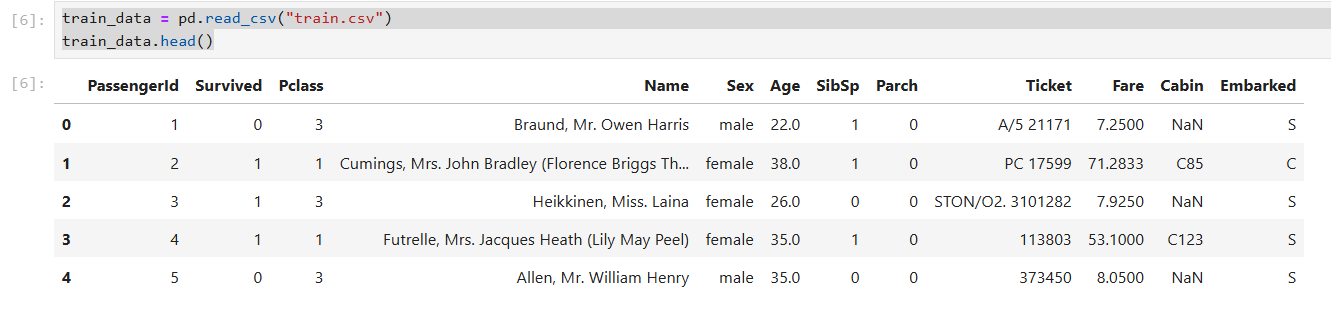


1. Tunggu proses import library selesai. Jika library tidak dikenali oleh Python, maka library tersebut perlu diinstal terlebih dahulu menggunakan perintah pip, selama terhubung dengan internet. Ketikkan baris kode berikut untuk menginstal library yang dibutuhkan.

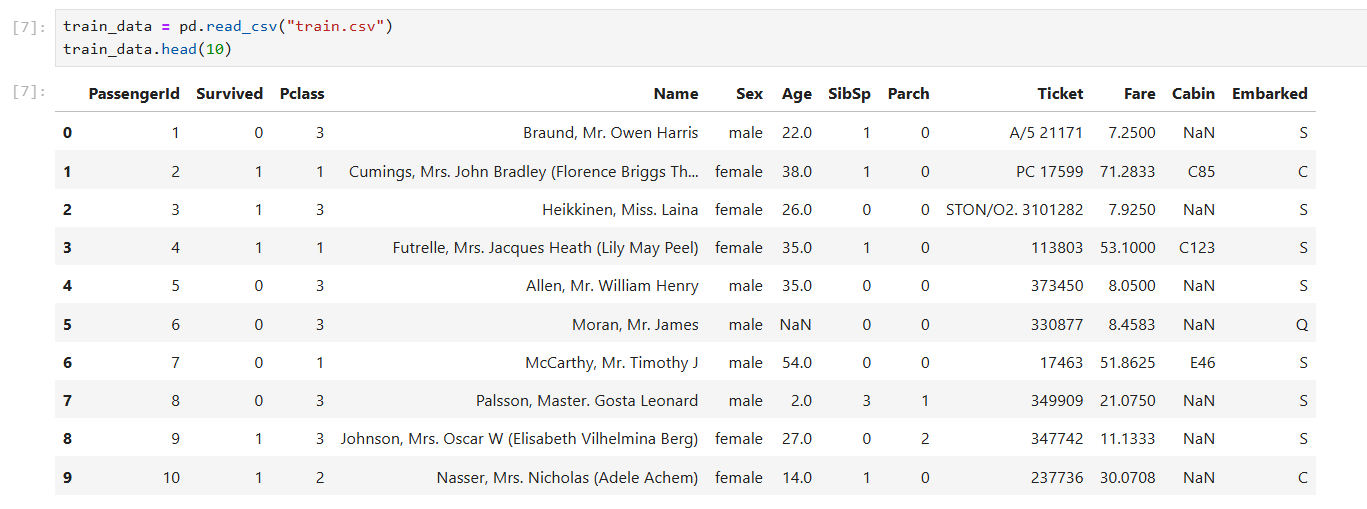


Setelah proses instal library selesai, ulangi langkah 10 kemudian dilanjutkan ke langkah 12.

1. Ketik method read\_csv() pada pandas untuk menampilkan file dataset Titanic dengan nama train.csv. Untuk menampilkan 5 data pertama, tambahkan method head() pada dataframe. Setelah dieksekusi, maka ditampilkan 5 data pertama dataset train Titanic yang terdiri dari 12 kolom.

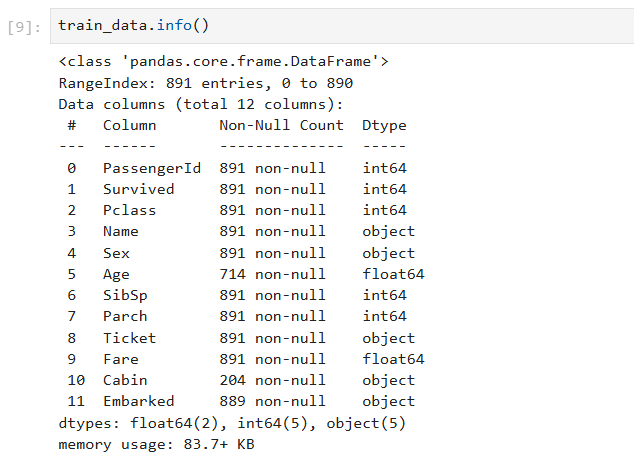


Jika ingin melihat 10 data pertama, tambahkan angka 10 dalam tanda kurung, misalnya train\_data.head(10).

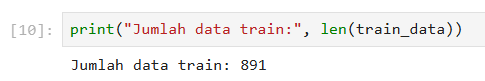


Atribut atau kolom tabel merupakan komponen penting dalam sebuah dataset. Untuk melihat nama-nama atribut yang digunakan dalam dataset, ketikkan perintah train\_data.columns, kemudian eksekusi. Sehingga akan ditampilkan nama dari semua atribut dalam dataset.

Sedangkan untuk mengetahui informasi detil dari dataset terkait tipe data masing-masing atribut adalah dapat menggunakan perintah train\_data.info().

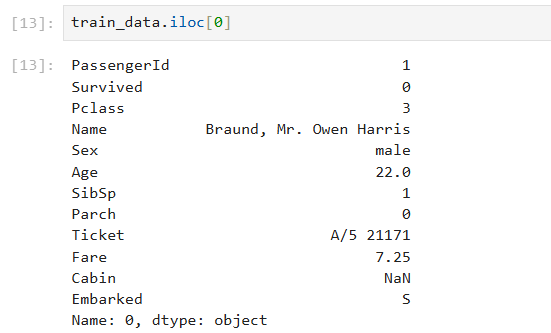


1. Untuk mengetahui jumlah data yang terdapat pada dataset train.csv, ketikkan kode print(len(train\_data)).



Berdasarkan hasil eksekusi, dapat diketahui bahwa jumlah data yang terdapat pada dataset train.csv adalah sebanyak 891 baris.

1. Untuk melihat salah satu data pada baris tertentu dalam dataset, maka bisa mengunakan nomor index untuk menentukan barisnya, dimana nomor index diawali dari angka 0 (nol) dengan menggunakan method iloc[]. Misalnya untuk melihat detil dari data pada baris pertama, maka nomor index yang digunakan adalah 0. Ketikkan perintah berikut pada sel train\_data.iloc[0], kemudian eksekusi kode tersebut.

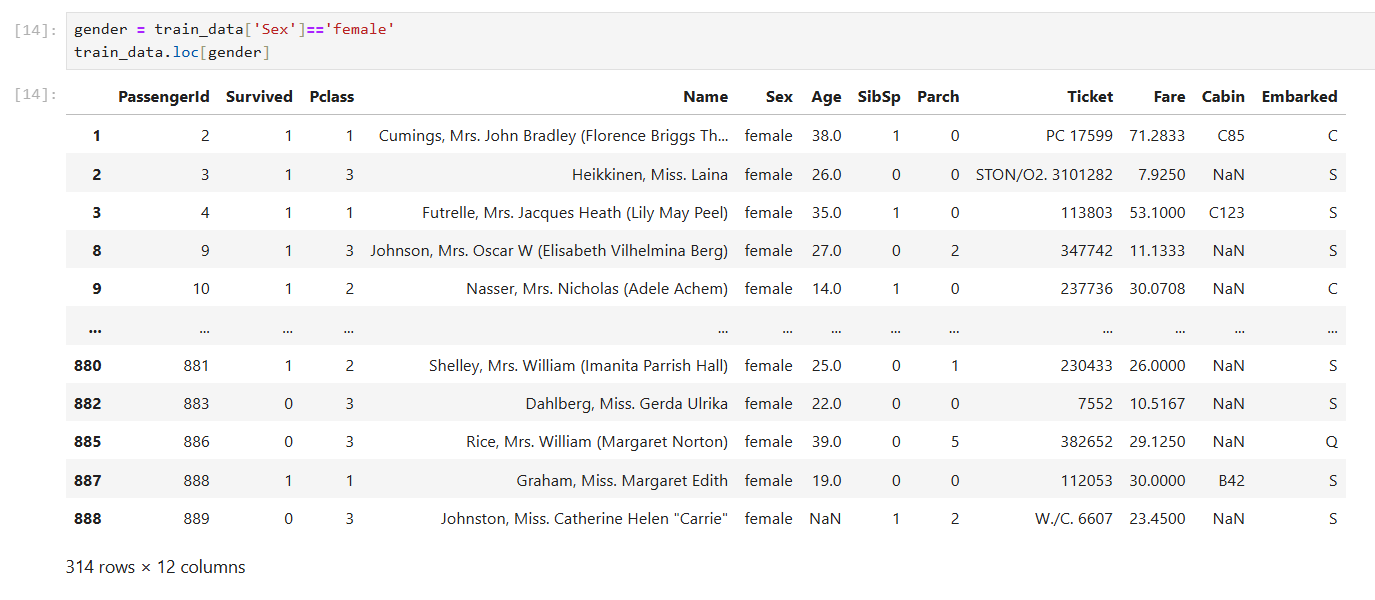


1. Jika ingin melakukan filtering, yaitu hanya untuk menampilkan data berdasarkan nilai tertentu pada kolom tertentu, maka bisa menggunakan perintah loc[] yang ditambahkan pada dataframe. Contohnya jika hanya ingin menampilkan penumpang Titanic yang berjenis kelamin wanita (female), maka bisa menuliskan kode berikut:

1| gender = train\_data[‘Sex’]==’female’

2| train\_data.loc[gender] 121

Sehingga jika dieksekusi akan menghasilkan 314 data berdasarkan jenis kelamin wanita (female).

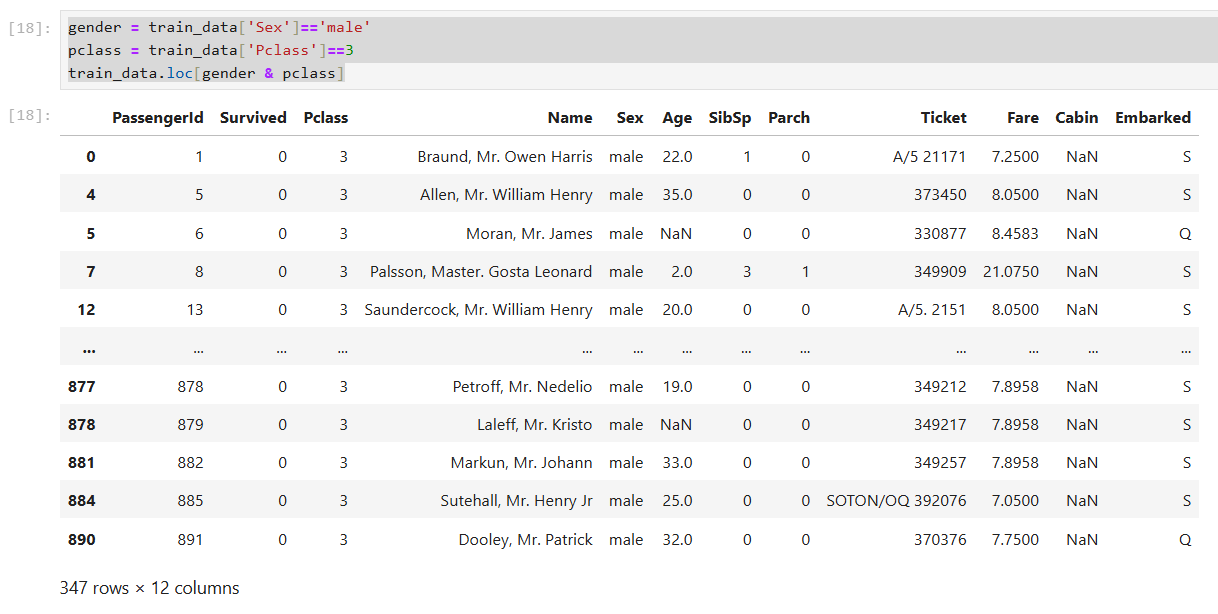


1. Jika filtering berdasarkan nilai tertentu dari 2 kolom atau lebih, maka bisa menggunakan operator AND (&) dan/atau OR (|). Misalnya akan menampilkan data penumpang dengan jenis kelamin pria (male) dan yang menempati kelas penumpang 3.

1| gender = train\_data[‘Sex’]==’male’

2| pclass = train\_data[‘Pclass’]==3

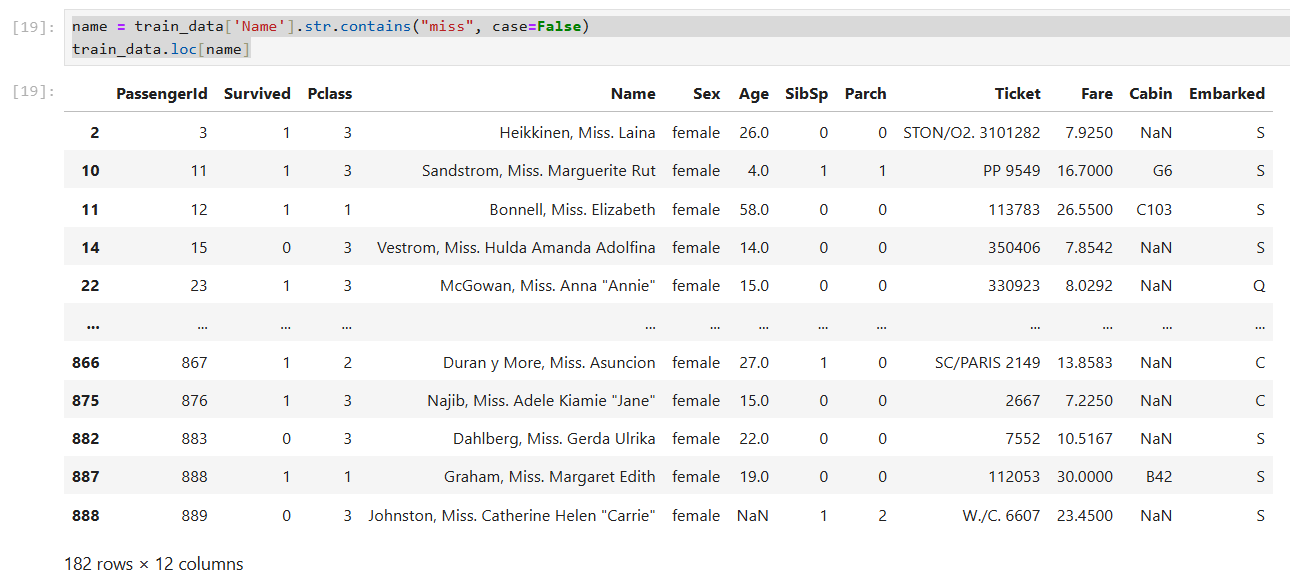
3| train\_data.loc[gender & pclass]



1. Sedangkan jika ingin menyaring data berdasarkan string yang terkandung dalam suatu nilai kolom, maka bisa menggunakan method str.contains(). Misalnya data yang ditampilkan adalah penumpang wanita yang namanya mengandung kata “Miss”. Jika hasil filtering mengabaikan huruf besar kecil, maka tambahkan case=False pada perintah contains().

1| name = train\_data[‘Name’].str.contains(“miss”, case=False)

2| train\_data.loc[name]

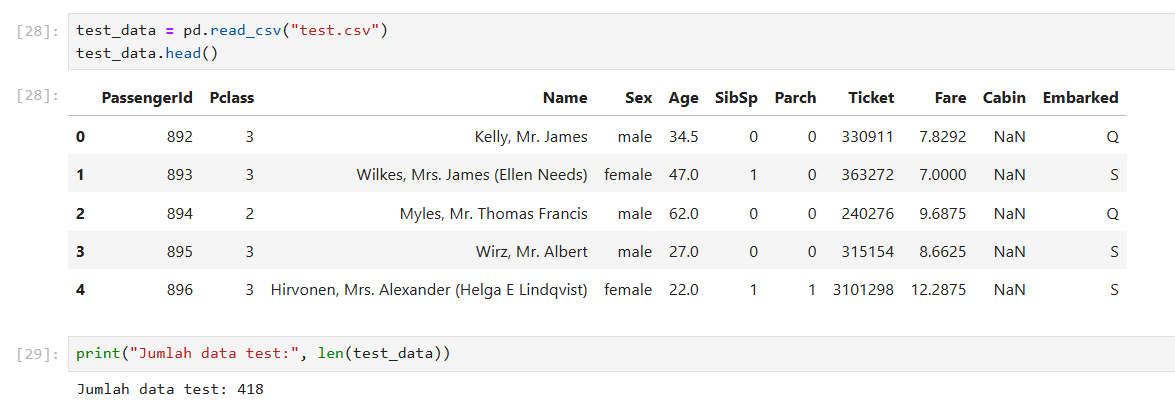


**TUGAS**

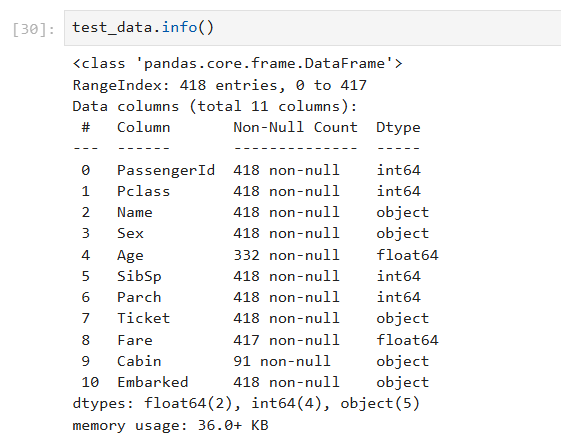
Dengan menggunakan dataset test.csv yang sudah diunduh dari repository, kerjakan perintah dan jawablah pertanyaan berikut ini.

1. Bukalah data test.csv dengan kode python. Ada berapa data yang terdapat dalam dataset tersebut?

418 data.

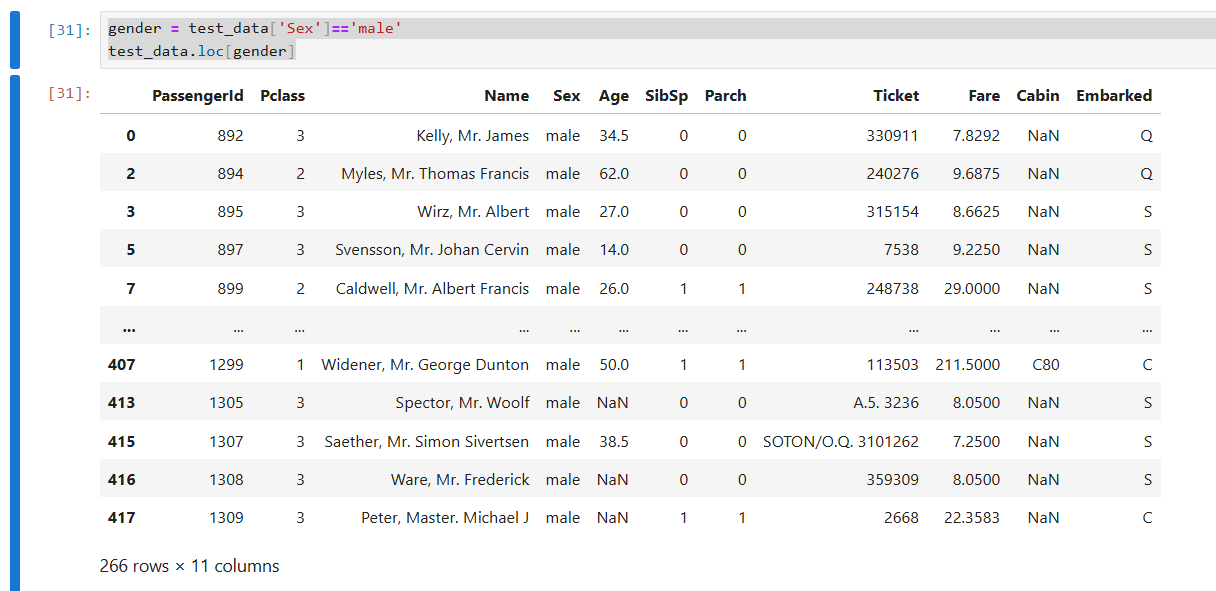


1. Dengan menggunakan kode python, tampilkan nama-nama kolom dan tipe datanya pada dataset tersebut, serta tampilkan jumlah kolomnya!



1. Tampilkan dan tentukan jumlah data penumpang yang berjenis kelamin pria (male)!

Jumlah = 266 data.



1. Dalam satu tabel, tampilkan datanya dan tentukan jumlah data penumpang wanita (female) DAN namanya mengandung kata ”James” ATAU penumpang pria (male) DAN namanya mengandung kata ”Samuel”!

Jumlah male = 6 data.

Jumlah female = 4 data.

