

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

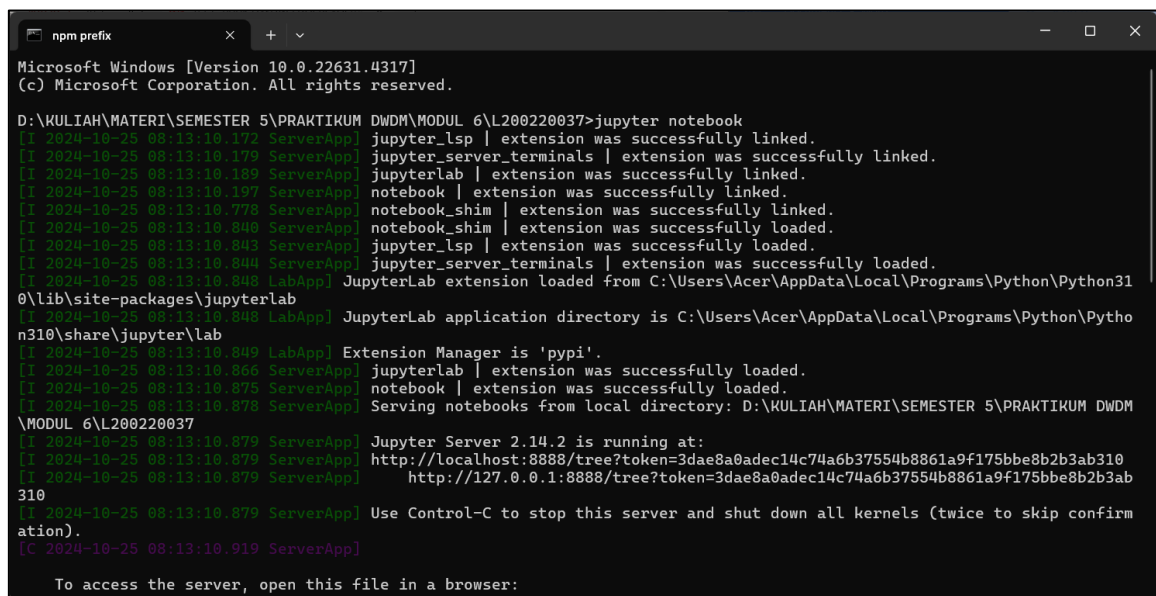
UNIVERSITAS MUHAMMADIYAH SURAKARTA

TAHUN 2024

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



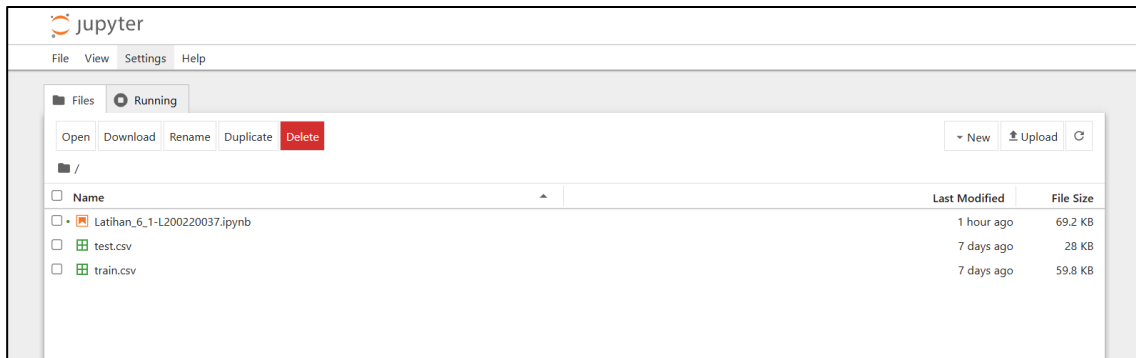
```

Microsoft Windows [Version 10.0.22631.4317]
(c) Microsoft Corporation. All rights reserved.

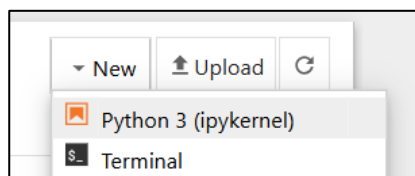
D:\KULIAH\MATERI\SEMESTER 5\PRAKTIKUM DWDW\MODUL 6\L200220037>jupyter notebook
[I 2024-10-25 08:13:10.172 ServerApp] jupyter_lsp | extension was successfully linked.
[I 2024-10-25 08:13:10.179 ServerApp] jupyter_server_terminals | extension was successfully linked.
[I 2024-10-25 08:13:10.189 ServerApp] jupyterlab | extension was successfully linked.
[I 2024-10-25 08:13:10.197 ServerApp] notebook | extension was successfully linked.
[I 2024-10-25 08:13:10.778 ServerApp] notebook_shim | extension was successfully linked.
[I 2024-10-25 08:13:10.848 ServerApp] notebook_shim | extension was successfully loaded.
[I 2024-10-25 08:13:10.843 ServerApp] jupyter_lsp | extension was successfully loaded.
[I 2024-10-25 08:13:10.844 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[I 2024-10-25 08:13:10.848 LabApp] JupyterLab extension loaded from C:\Users\Acer\AppData\Local\Programs\Python\Python310\lib\site-packages\jupyterlab
[I 2024-10-25 08:13:10.848 LabApp] JupyterLab application directory is C:\Users\Acer\AppData\Local\Programs\Python\Python310\share\jupyter\lab
[I 2024-10-25 08:13:10.849 LabApp] Extension Manager is 'pypi'.
[I 2024-10-25 08:13:10.866 ServerApp] jupyterlab | extension was successfully loaded.
[I 2024-10-25 08:13:10.875 ServerApp] notebook | extension was successfully loaded.
[I 2024-10-25 08:13:10.878 ServerApp] Serving notebooks from local directory: D:\KULIAH\MATERI\SEMESTER 5\PRAKTIKUM DWDW\MODUL 6\L200220037
[I 2024-10-25 08:13:10.879 ServerApp] Jupyter Server 2.14.2 is running at:
[I 2024-10-25 08:13:10.879 ServerApp] http://localhost:8888/tree?token=3dae8a0adec14c74a6b37554b8861a9f175bbe8b2b3ab310
[I 2024-10-25 08:13:10.879 ServerApp] http://127.0.0.1:8888/tree?token=3dae8a0adec14c74a6b37554b8861a9f175bbe8b2b3ab310
[I 2024-10-25 08:13:10.879 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 2024-10-25 08:13:10.919 ServerApp]

To access the server, open this file in a browser:
  
```

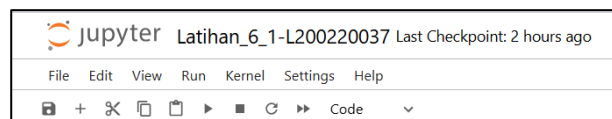
5. Aplikasi Jupyter Notebook akan dijalankan pada sebuah browser yang terinstal di komputer.



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



7. 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:
 - a. Nama file: digunakan untuk menentukan nama file yang akan memiliki ekstensi .ipynb.
 - b. 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.
 - c. Code cell: digunakan untuk menuliskan kode-kode program.
8. 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.



9. Pada percobaan pertama dalam Bab 6 ini, kita akan belajar bagaimana menampilkan data dari file CSV. Untuk mengeksekusi dengan Python, kita membutuhkan library pandas.
10. 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.

```
[5]: import pandas as pd
```

11. 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.

```
[4]: pip install pandas

Collecting pandas
  Downloading pandas-2.2.3-cp310-cp310-win_amd64.whl.metadata (19 kB)
Collecting numpy>=1.22.4 (from pandas)
  Downloading numpy-2.1.2-cp310-cp310-win_amd64.whl.metadata (59 kB)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\acer\appdata\local\programs\python\python310\lib\site-packages (from pandas) (2.9.0.post0)
Collecting pytz>=2020.1 (from pandas)
  Downloading pytz-2024.2-py2.py3-none-any.whl.metadata (22 kB)
Collecting tzdata>=2022.7 (from pandas)
  Downloading tzdata-2024.2-py2.py3-none-any.whl.metadata (1.4 kB)
Requirement already satisfied: six>=1.5 in c:\users\acer\appdata\local\programs\python\python310\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Downloading pandas-2.2.3-cp310-cp310-win_amd64.whl (11.6 MB)
----- 0.0/11.6 MB ? eta -:--
----- 0.3/11.6 MB ? eta -:--
----- 1.3/11.6 MB 3.9 MB/s eta 0:00:03
----- 1.8/11.6 MB 3.7 MB/s eta 0:00:03
----- 2.6/11.6 MB 3.6 MB/s eta 0:00:03
----- 3.1/11.6 MB 4.0 MB/s eta 0:00:03
----- 4.5/11.6 MB 3.9 MB/s eta 0:00:02
----- 5.0/11.6 MB 3.6 MB/s eta 0:00:02
----- 5.8/11.6 MB 3.6 MB/s eta 0:00:02
----- 6.3/11.6 MB 3.5 MB/s eta 0:00:02
----- 7.3/11.6 MB 3.6 MB/s eta 0:00:02
----- 8.4/11.6 MB 3.7 MB/s eta 0:00:01
----- 9.4/11.6 MB 3.9 MB/s eta 0:00:01
----- 10.5/11.6 MB 3.9 MB/s eta 0:00:01
----- 11.0/11.6 MB 3.8 MB/s eta 0:00:01
----- 11.6/11.6 MB 3.7 MB/s eta 0:00:00
Downloading numpy-2.1.2-cp310-cp310-win_amd64.whl (12.9 MB)
----- 0.0/12.9 MB ? eta -:--
----- 1.0/12.9 MB 6.3 MB/s eta 0:00:02
```

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

12. 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.

```
[6]: train_data = pd.read_csv("train.csv")
     train_data.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Jika ingin melihat 10 data pertama, tambahkan angka 10 dalam tanda kurung, misalnya `train_data.head(10)`.

```
[7]: train_data = pd.read_csv("train.csv")
     train_data.head(10)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C

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 detail dari dataset terkait tipe data masing-masing atribut adalah dapat menggunakan perintah `train_data.info()`.

```
[9]: train_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column             Non-Null Count  Dtype  
---  -
0   PassengerId         891 non-null    int64  
1   Survived            891 non-null    int64  
2   Pclass              891 non-null    int64  
3   Name                891 non-null    object  
4   Sex                 891 non-null    object  
5   Age                 714 non-null    float64 
6   SibSp               891 non-null    int64  
7   Parch               891 non-null    int64  
8   Ticket              891 non-null    object  
9   Fare                891 non-null    float64 
10  Cabin               204 non-null    object  
11  Embarked            889 non-null    object  
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

13. Untuk mengetahui jumlah data yang terdapat pada dataset `train.csv`, ketikkan kode `print(len(train_data))`.

```
[10]: print("Jumlah data train:", len(train_data))

Jumlah data train: 891
```

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

14. Untuk melihat salah satu data pada baris tertentu dalam dataset, maka bisa menggunakan nomor index untuk menentukan barisnya, dimana nomor index diawali dari angka 0 (nol) dengan menggunakan method `iloc[]`. Misalnya untuk melihat detail 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.

```
[13]: train_data.iloc[0]

[13]: PassengerId         1
      Survived            0
      Pclass              3
      Name                Braund, Mr. Owen Harris
      Sex                 male
      Age                 22.0
      SibSp               1
      Parch               0
      Ticket              A/5 21171
      Fare                7.25
      Cabin              NaN
      Embarked            S
      Name: 0, dtype: object
```

15. 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).

```
[14]: gender = train_data['Sex']=='female'
train_data.loc[gender]
```

```
[14]:
```

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	2	1	Cummings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
8	9	1	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C
...
880	881	1	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26.0000	NaN	S
882	883	0	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	NaN	S
885	886	0	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN	Q
887	888	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S

314 rows × 12 columns

16. 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]`

```
[18]: gender = train_data['Sex']=='male'
pclass = train_data['Pclass']==3
train_data.loc[gender & pclass]
```

```
[18]:
```

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
4	5	0	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
7	8	0	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
12	13	0	Saunderscock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.0500	NaN	S
...
877	878	0	Petroff, Mr. Nedelio	male	19.0	0	0	349212	7.8958	NaN	S
878	879	0	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8958	NaN	S
881	882	0	Markun, Mr. Johann	male	33.0	0	0	349257	7.8958	NaN	S
884	885	0	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0500	NaN	S
890	891	0	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

347 rows × 12 columns

17. 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]`

[19]: name = train_data['Name'].str.contains("miss", case=False)
train_data.loc[name]

[19]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
	2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S	
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S	
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.8542	NaN	S	
22	23	1	3	McGowan, Miss. Anna "Annie"	female	15.0	0	0	330923	8.0292	NaN	Q	
...	
866	867	1	2	Duran y More, Miss. Asuncion	female	27.0	1	0	SC/PARIS 2149	13.8583	NaN	C	
875	876	1	3	Najib, Miss. Adele Kiamie "Jane"	female	15.0	0	0	2667	7.2250	NaN	C	
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	NaN	S	
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S	

182 rows x 12 columns

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.

```
[28]: test_data = pd.read_csv("test.csv")
      test_data.head()

[28]:
```

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

```
[29]: print("Jumlah data test:", len(test_data))

      Jumlah data test: 418
```

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

```
[30]: test_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype  
---  --
 0   PassengerId    418 non-null    int64  
 1   Pclass         418 non-null    int64  
 2   Name           418 non-null    object  
 3   Sex            418 non-null    object  
 4   Age            332 non-null    float64 
 5   SibSp          418 non-null    int64  
 6   Parch          418 non-null    int64  
 7   Ticket         418 non-null    object  
 8   Fare           417 non-null    float64 
 9   Cabin          91 non-null     object  
10  Embarked       418 non-null    object  
dtypes: float64(2), int64(4), object(5)
memory usage: 36.0+ KB
```

3. Tampilkan dan tentukan jumlah data penumpang yang berjenis kelamin pria (male)!
Jumlah = 266 data.


```
[31]: gender = test_data['Sex']=='male'
test_data.loc[gender]
```

```
[31]:
```

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
5	897	3	Svensson, Mr. Johan Cervin	male	14.0	0	0	7538	9.2250	NaN	S
7	899	2	Caldwell, Mr. Albert Francis	male	26.0	1	1	248738	29.0000	NaN	S
...
407	1299	1	Widener, Mr. George Dunton	male	50.0	1	1	113503	211.5000	C80	C
413	1305	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500	NaN	S
415	1307	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500	NaN	S
416	1308	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500	NaN	S
417	1309	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583	NaN	C

266 rows × 11 columns

4. 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.

```
[38]: gender1 = test_data['Sex']=='female'
name1 = test_data['Name'].str.contains("James", case=False)
gender2 = test_data['Sex']=='male'
name2 = test_data['Name'].str.contains("Samuel", case=False)
test_data.loc[gender1&name1|gender2&name2]
```

```
[38]:
```

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
9	901	3	Davies, Mr. John Samuel	male	21.0	2	0	A/4 48871	24.1500	NaN	S
110	1002	2	Stanton, Mr. Samuel Ward	male	41.0	0	0	237734	15.0458	NaN	C
229	1121	2	Hocking, Mr. Samuel James Metcalfe	male	36.0	0	0	242963	13.0000	NaN	S
257	1149	3	Niklasson, Mr. Samuel	male	28.0	0	0	363611	8.0500	NaN	S
320	1212	3	Andersson, Mr. Johan Samuel	male	26.0	0	0	347075	7.7750	NaN	S
343	1235	1	Cardeza, Mrs. James Warburton Martinez (Charlo...	female	58.0	0	1	PC 17755	512.3292	B51 B53 B55	C
353	1245	2	Herman, Mr. Samuel	male	49.0	1	2	220845	65.0000	NaN	S
362	1254	2	Ware, Mrs. John James (Florence Louise Long)	female	31.0	0	0	CA 31352	21.0000	NaN	S
391	1283	1	Lines, Mrs. Ernest H (Elizabeth Lindsey James)	female	51.0	0	1	PC 17592	39.4000	D28	S