[pandas] merupakan singkatan dari "panel data", sebuah istilah econometrics untuk sekumpulan data yang diobservasi berulang oleh seseorang. — Wikipedia

Pandas buat apa sih </p>

- Banyak hal bisa dilakukan di Pandas.
- Pandas adalah rumah, kerangka bagi data.
- Pandas akan mengubah tampilan data menjadi Dataframe

Fitur yang banyak dipakai apa saja?

1. Mengkalkulasi statistik
2. Membersihkan data
3. Visualisasi data (dengan matplotlib atau lib lain)
4. Meyimpan data dalam beragam bentuk

Kenapa harus Pandas?

In [3]:
```python
myList=[['Isna', 2020, 'Dosen','Ilmu Komputer'],[12,'Mekanisme import Pandas', 2
019, "65000"],['Garis lurus', 19],['Alfi', 1988, 'Dosen','Ilmu Komputer'],[12,'N
umpym pasti kembali', 2014, "20000"],['Random', 7],['Isna', 2022, 'Guru','Magist
er'],[18,'Apakah visualisasi penting', 1998, "35000"],['Belajar Tuple', 12],['Bu
stoni', 2020, 'Dosen','MIPA'],[56,'teks nggak jelas', 2003, "15000"],['Garis lur
us', 3],['Isna', 2020, 'Dosen','Ilmu Komputer'],[12,'Mekanisme import Pandas', 2
019, "65000"],['Garis lurus', 12],['Alfi', 1988, 'Dosen','Ilmu Komputer'],[12,'N
umpym pasti kembali', 2014, "20000"],['Random', 88],['Isna', 2022, 'Guru','Magis
ter'],[18,'Apakah visualisasi penting', 1998, "35000"],['Belajar Tuple', 4],['Bu
stoni', 2020, 'Dosen','MIPA'],[56,'teks nggak jelas', 2003, "15000"],['Garis lur
us', 6]]
print(myList)
```

```
[['Isna', 2020, 'Dosen', 'Ilmu Komputer'], [12, 'Mekanisme import Pandas', 201
9, '65000'], ['Garis lurus', 19], ['Alfi', 1988, 'Dosen', 'Ilmu Komputer'], [1
2, 'Numpym pasti kembali', 2014, '20000'], ['Random', 7], ['Isna', 2022, 'Guru
', 'Magister'], [18, 'Apakah visualisasi penting', 1998, '35000'], ['Belajar T
uple', 12], ['Bustoni', 2020, 'Dosen', 'MIPA'], [56, 'teks nggak jelas', 2003,
'15000'], ['Garis lurus', 3], ['Isna', 2020, 'Dosen', 'Ilmu Komputer'], [12, '
Mekanisme import Pandas', 2019, '65000'], ['Garis lurus', 12], ['Alfi', 1988,
'Dosen', 'Ilmu Komputer'], [12, 'Numpym pasti kembali', 2014, '20000'], ['Rand
om', 88], ['Isna', 2022, 'Guru', 'Magister'], [18, 'Apakah visualisasi penting
', 1998, '35000'], ['Belajar Tuple', 4], ['Bustoni', 2020, 'Dosen', 'MIPA'],
[56, 'teks nggak jelas', 2003, '15000'], ['Garis lurus', 6]]
```

In [1]:
```python
import pandas as pd
```

```
In [4]: df=pd.DataFrame(myList,columns=['A','B','C','D'])
        df
```

Out[4]:

	A	B	C	D
0	Isna	2020	Dosen	Ilmu Komputer
1	12	Mekanisme import Pandas	2019	65000
2	Garis lurus	19	None	None
3	Alfi	1988	Dosen	Ilmu Komputer
4	12	Numpym pasti kembali	2014	20000
5	Random	7	None	None
6	Isna	2022	Guru	Magister
7	18	Apakah visualisasi penting	1998	35000
8	Belajar Tuple	12	None	None
9	Bustoni	2020	Dosen	MIPA
10	56	teks nggak jelas	2003	15000
11	Garis lurus	3	None	None
12	Isna	2020	Dosen	Ilmu Komputer
13	12	Mekanisme import Pandas	2019	65000
14	Garis lurus	12	None	None
15	Alfi	1988	Dosen	Ilmu Komputer
16	12	Numpym pasti kembali	2014	20000
17	Random	88	None	None
18	Isna	2022	Guru	Magister
19	18	Apakah visualisasi penting	1998	35000
20	Belajar Tuple	4	None	None
21	Bustoni	2020	Dosen	MIPA
22	56	teks nggak jelas	2003	15000
23	Garis lurus	6	None	None

Komponen Pandas </p>

Yuk Praktik! Dasar Pandas </p>

Install dulu Pandas

```
### conda install pandas
### atau
### pip install pandas
```

Berikutnya, import Pandas..

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

Membuat dataframe dari dasar

Ada banyak cara membuat dataframe, tapi mari coba cara paling gampang, yakni menggunakan dictionary menggunakan `dict` Misalkan, saya akan membuat dict transaksi perhari pada toko Indoapril. Ada 3 item yang dijual, yakni:

- Roti
- Cola
- Tisu.

```
In [6]:  import pandas as pd
         data = {
             'roti': [3, 2, 0, 1],
             'cola': [0, 3, 7, 2],
             'tisu': [5, 1, 1, 8]
         }
         #pip install pandas
         #conda install pandas
```

```
In [26]: transaksi = pd.DataFrame(data)
         transaksi
```

Out[26]:

	roti	cola	tisu
0	3	0	5
1	2	3	1
2	0	7	1
3	1	2	8

Index

Secara default index dimulai dari 0, namun kita juga bisa mengubah index sesuai keinginan kita. Misal, kita ubah jadi nama customer.

```
In [28]: transaksi = pd.DataFrame(data, index=['Ani', 'Budi', 'Citra', 'David'])
         transaksi
```

Out[28]:

	roti	cola	tisu
Ani	3	0	5
Budi	2	3	1
Citra	0	7	1
David	1	2	8

```
In [8]: customer=['Ani', 'Budi', 'Citra', 'David']
        transaksi = pd.DataFrame(data, index=customer)
        transaksi
```

Out[8]:

	roti	cola	tisu
Ani	3	0	5
Budi	2	3	1
Citra	0	7	1
David	1	2	8

Melihat transaksi untuk masing-masing customer

```
In [30]: transaksi.loc['Citra']
```

```
Out[30]: roti    0
         cola    7
         tisu    1
         Name: Citra, dtype: int64
```

Melihat transaksi dari index nya

```
In [32]: transaksi.iloc[1] #Budi
```

```
Out[32]: roti    2
         cola    3
         tisu    1
         Name: Budi, dtype: int64
```

Mengimport File </p>

[https://github.com/isnaalfi/DigitalTalent/tree/master/2020 (https://github.com/isnaalfi/DigitalTalent/tree/master/2020)](https://github.com/isnaalfi/DigitalTalent/tree/master/2020)

```
In [45]: # github.com/isnaalfi --> Digital Talent --> 2020 transaksi.csv, .xls, IMDB movi
         es
         df = pd.read_csv('transaksi.csv')

         df
```

Out[45]:

	Unnamed: 0	roti	cola	tisu
0	Ani	3	0	5
1	Budi	2	3	1
2	Citra	0	7	1
3	David	1	2	8

```
In [14]: df = pd.read_csv(r'C:\Users\Bunda Freya\01 TUTORIAL ISNA\transaksi.csv')

         df
```

Out[14]:

	Unnamed: 0	roti	cola	tisu
0	Ani	3	0	5
1	Budi	2	3	1
2	Citra	0	7	1
3	David	1	2	8

```
In [16]: df.columns.values
         #Rename dulu
         new_columns = df.columns.values
         new_columns[0] = 'nama'
         df.columns = new_columns
```

```
In [17]: df = df.set_index('nama')
         df
```

Out[17]:

	roti	cola	tisu
nama			
Ani	3	0	5
Budi	2	3	1
Citra	0	7	1
David	1	2	8

Secara default, data dari csv menggunakan index dari 0. Namun kita bisa mengubah index menggunakan salah satu column..

```
In [18]: df = pd.read_csv('transaksi.csv', index_col=0)

         df
```

Out[18]:

	roti	cola	tisu
Ani	3	0	5
Budi	2	3	1
Citra	0	7	1
David	1	2	8

Membaca data Excel

```
In [19]: import pandas as pd
         from pandas import ExcelWriter
         from pandas import ExcelFile

         df_excel = pd.read_excel('transaksi.xlsx', index_col=0)

         print("Column headings:")
         print(df_excel.columns)
         df_excel
```

```
Column headings:
Index(['roti', 'cola', 'susu'], dtype='object')
```

Out[19]:

	roti	cola	susu
Ani	3	0	5
Budi	2	3	1
Citra	0	7	1
David	1	2	8

Menulis data dalam bentu CSV dan Excel

```
In [20]: df_excel.to_csv('new_transaksi.csv')

         with ExcelWriter('new_transaksi.xlsx') as writer:
             df_excel.to_excel(writer)
```

Operasi-operasi umum dalam Pandas </p>

Menggunakan dataset movies

```
In [21]: movies_df = pd.read_csv("IMDB-Movie-Data.csv", index_col="Title")
```

Melihat Dataframe

In [25]: `movies_df.head(22)`

Out[25]:

Title	Rank	Genre	Description	Director	Actors	Year	Runtime (Minutes)
Guardians of the Galaxy	1	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	12
Prometheus	2	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	12
Split	3	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	11
Sing	4	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	10
Suicide Squad	5	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	12
The Great Wall	6	Action,Adventure,Fantasy	European mercenaries searching for black powde...	Yimou Zhang	Matt Damon, Tian Jing, Willem Dafoe, Andy Lau	2016	10
La La Land	7	Comedy,Drama,Music	A jazz pianist falls for an aspiring actress i...	Damien Chazelle	Ryan Gosling, Emma Stone, Rosemarie DeWitt, J....	2016	12
Mindhorn	8	Comedy	A has-been actor best known for playing the ti...	Sean Foley	Essie Davis, Andrea Riseborough, Julian Barrat...	2016	8
The Lost City of Z	9	Action,Adventure,Biography	A true-life drama, centering on British explor...	James Gray	Charlie Hunnam, Robert Pattinson, Sienna Mille...	2016	14
Passengers	10	Adventure,Drama,Romance	A spacecraft traveling to a distant colony pla...	Morten Tyldum	Jennifer Lawrence, Chris Pratt, Michael Sheen,...	2016	11
Fantastic Beasts and Where to Find Them	11	Adventure,Family,Fantasy	The adventures of writer Newt Scamander in New...	David Yates	Eddie Redmayne, Katherine Waterston, Alison Su...	2016	13
Hidden Figures	12	Biography,Drama,History	The story of a team of female African-American...	Theodore Melfi	Taraji P. Henson, Octavia Spencer, Janelle Mon...	2016	12
Rogue One	13	Action,Adventure,Sci-Fi	The Rebel Alliance makes a risky move to steal...	Gareth Edwards	Felicity Jones, Diego Luna, Alan Tudyk, Donnie...	2016	13

In [62]: `movies_df.head(10)`

Out[62]:

Title	Rank	Genre	Description	Director	Actors	Year	Runtime (Minutes)
Guardians of the Galaxy	1	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121
Prometheus	2	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124
Split	3	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	117
Sing	4	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108
Suicide Squad	5	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	123
The Great Wall	6	Action,Adventure,Fantasy	European mercenaries searching for black powde...	Yimou Zhang	Matt Damon, Tian Jing, Willem Dafoe, Andy Lau	2016	103
La La Land	7	Comedy,Drama,Music	A jazz pianist falls for an aspiring actress i...	Damien Chazelle	Ryan Gosling, Emma Stone, Rosemarie DeWitt, J....	2016	128
Mindhorn	8	Comedy	A has-been actor best known for playing the ti...	Sean Foley	Essie Davis, Andrea Riseborough, Julian Barrat...	2016	89
The Lost City of Z	9	Action,Adventure,Biography	A true-life drama, centering on British explor...	James Gray	Charlie Hunnam, Robert Pattinson, Sienna Mille...	2016	141
Passengers	10	Adventure,Drama,Romance	A spacecraft traveling to a distant colony pla...	Morten Tyldum	Jennifer Lawrence, Chris Pratt, Michael Sheen,...	2016	116

In [26]: `movies_df.tail(3)`

Out[26]:

Title	Rank	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Vo
Step Up 2: The Streets	998	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robert Hoffman, Briana Evigan, Cassie Ventura,...	2008	98	6.2	70(
Search Party	999	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong	Adam Pally, T.J. Miller, Thomas Middleditch,Sh...	2014	93	5.6	4;
Nine Lives	1000	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	2016	87	5.3	12-

Melihat informasi data

In [66]: `movies_df.info()`

```
<class 'pandas.core.frame.DataFrame'>
Index: 1000 entries, Guardians of the Galaxy to Nine Lives
Data columns (total 11 columns):
Rank                1000 non-null int64
Genre               1000 non-null object
Description         1000 non-null object
Director            1000 non-null object
Actors              1000 non-null object
Year                1000 non-null int64
Runtime (Minutes)   1000 non-null int64
Rating              1000 non-null float64
Votes               1000 non-null int64
Revenue (Millions)  872 non-null float64
Metascore           936 non-null float64
dtypes: float64(3), int64(4), object(4)
memory usage: 93.8+ KB
```

In [28]: `movies_df.shape`

Out[28]: `(1000, 11)`

Mengatasi duplikasi data </p>

Sampel

Dataset kita buat duplikat dulu

In [29]: `temp_df = movies_df.append(movies_df)` *#bikin duplikasi*

`temp_df.shape`

Out[29]: `(2000, 11)`

Lakukan drop baris yang sama

```
In [32]: temp_df=temp_df.drop_duplicates()
         #akan mengeluarka return value data diremove duplikatnya, lalu ditangkap temp_df

         temp_df.shape
```

```
Out[32]: (1000, 11)
```

```
In [33]: temp_df.drop_duplicates(inplace=True)
```

Argument penting lain dalam drop_duplicates() adalah `keep` , yang bisa memiliki nilai:

```
first: (default) Drop duplikat kecuali temuan pertama
last: Drop duplikat kecuali temuan pterakhir
False: Drop semua isi, sehingga kosong
```

```
In [34]: temp_df2 = movies_df.append(movies_df)   # make a new copy

         temp_df2.drop_duplicates(inplace=True, keep=False)

         temp_df2.shape
```

```
Out[34]: (0, 11)
```

Membersihkan Kolom </p>

```
In [35]: movies_df.columns
```

```
Out[35]: Index(['Rank', 'Genre', 'Description', 'Director', 'Actors', 'Year',
                'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
                'Metascore'],
               dtype='object')
```

Rename Kolom

```
In [36]: movies_df.rename(columns={
                 'Runtime (Minutes)': 'Runtime',
                 'Revenue (Millions)': 'Revenue_millions'
             }, inplace=True)
         movies_df.columns
```

```
Out[36]: Index(['Rank', 'Genre', 'Description', 'Director', 'Actors', 'Year', 'Runtime
         ',
                 'Rating', 'Votes', 'Revenue_millions', 'Metascore'],
               dtype='object')
```

Kalau mau ubah jadi lowercase?

```
In [37]: movies_df.columns = ['rank', 'genre', 'description', 'director', 'actors', 'year
         ', 'runtime',
                               'rating', 'votes', 'revenue_millions', 'metascore']

         movies_df.columns
```

```
Out[37]: Index(['rank', 'genre', 'description', 'director', 'actors', 'year', 'runtime
         ',
                'rating', 'votes', 'revenue_millions', 'metascore'],
               dtype='object')
```

```
In [38]: movies_df.columns = [col.lower() for col in movies_df]

         movies_df.columns
```

```
Out[38]: Index(['rank', 'genre', 'description', 'director', 'actors', 'year', 'runtime
         ',
                'rating', 'votes', 'revenue_millions', 'metascore'],
               dtype='object')
```

Mengatasi Missing Values </p>

Ada dua cara mengatasi missing values:

```
- Hilangkan baris atau kolom dengan null
- Ganti dengan nilai non-null (imputation)
```

Cek apakah dataframe mnegandung Null

In [39]: movies_df.isnull()

Out[39]:

Title	rank	genre	description	director	actors	year	runtime	rating	votes	revenue_millions
Guardians of the Galaxy	False	False	False	False	False	False	False	False	False	False
Prometheus	False	False	False	False	False	False	False	False	False	False
Split	False	False	False	False	False	False	False	False	False	False
Sing	False	False	False	False	False	False	False	False	False	False
Suicide Squad	False	False	False	False	False	False	False	False	False	False
The Great Wall	False	False	False	False	False	False	False	False	False	False
La La Land	False	False	False	False	False	False	False	False	False	False
Mindhorn	False	False	False	False	False	False	False	False	False	True
The Lost City of Z	False	False	False	False	False	False	False	False	False	False
Passengers	False	False	False	False	False	False	False	False	False	False
Fantastic Beasts and Where to Find Them	False	False	False	False	False	False	False	False	False	False
Hidden Figures	False	False	False	False	False	False	False	False	False	False
Rogue One	False	False	False	False	False	False	False	False	False	False
Moana	False	False	False	False	False	False	False	False	False	False
Colossal	False	False	False	False	False	False	False	False	False	False
The Secret Life of Pets	False	False	False	False	False	False	False	False	False	False
Hacksaw Ridge	False	False	False	False	False	False	False	False	False	False
Jason Bourne	False	False	False	False	False	False	False	False	False	False
Lion	False	False	False	False	False	False	False	False	False	False
Arrival	False	False	False	False	False	False	False	False	False	False
Gold	False	False	False	False	False	False	False	False	False	False
Manchester by the Sea	False	False	False	False	False	False	False	False	False	False
Hounds of Love	False	False	False	False	False	False	False	False	False	True
Trolls	False	False	False	False	False	False	False	False	False	False
Independence Day: Resurgence	False	False	False	False	False	False	False	False	False	False
Paris pieds nus	False	False	False	False	False	False	False	False	False	True
Bahubali: The Beginning	False	False	False	False	False	False	False	False	False	False
Dead Awake	False	False	False	False	False	False	False	False	False	False
Bad Moms	False	False	False	False	False	False	False	False	False	False
Assassin's Creed	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...
Texas Chainsaw 3D	False	False	False	False	False	False	False	False	False	False
Disturbia	False	False	False	False	False	False	False	False	False	False

Menghitung jumlah yang Null

```
In [40]:   movies_df.isnull().sum()
```

```
Out[40]:   rank                  0
           genre                 0
           description           0
           director              0
           actors                0
           year                  0
           runtime               0
           rating                0
           votes                 0
           revenue_millions    128
           metascore            64
           dtype: int64
```

Dropna

Digunakan untuk membuang data Null revenue memiliki 128 NaN/Null sedangkan metascore 64. Total ada 192 (catatan tidak overlap) Maka kurang lebih akan berkurang 192, menyisakan 808 baris data.

```
In [41]:   test_drop=movies_df.dropna()
           test_drop.shape   #terdapat baris overlap NULL
```

```
Out[41]:   (838, 11)
```

```
In [42]:   #Mengapus Kolom
           test_drop=movies_df.dropna(axis=1)
           test_drop.shape
```

```
Out[42]:   (1000, 9)
```

Imputation

Mengganti nilai Null

```
In [46]:   #ekstract dulu revenue, agar lebih jelas
           revenue = movies_df['revenue_millions']
           revenue.head()
```

```
Out[46]:   Title
           Guardians of the Galaxy    333.13
           Prometheus                 126.46
           Split                      138.12
           Sing                       270.32
           Suicide Squad              325.02
           Name: revenue_millions, dtype: float64
```

```
In [47]:   #ganti dengan mean/ median
           revenue_mean = revenue.mean()
           revenue_mean
```

```
Out[47]:   82.95637614678897
```

Isi Null dengan mean

```
In [48]:   revenue.fillna(revenue_mean, inplace=True)
```

```
In [49]: movies_df.isnull().sum()
```

```
Out[49]: rank                  0
         genre                 0
         description           0
         director              0
         actors                0
         year                  0
         runtime               0
         rating                0
         votes                 0
         revenue_millions      0
         metascore            64
         dtype: int64
```

Melihat statistik deskriptif

```
In [93]: movies_df.describe()
```

Out[93]:

	rank	year	runtime	rating	votes	revenue_millions	metascore
count	1000.000000	1000.000000	1000.000000	1000.000000	1.000000e+03	1000.000000	936.000000
mean	500.500000	2012.783000	113.172000	6.723200	1.698083e+05	82.956376	58.985043
std	288.819436	3.205962	18.810908	0.945429	1.887626e+05	96.412043	17.194757
min	1.000000	2006.000000	66.000000	1.900000	6.100000e+01	0.000000	11.000000
25%	250.750000	2010.000000	100.000000	6.200000	3.630900e+04	17.442500	47.000000
50%	500.500000	2014.000000	111.000000	6.800000	1.107990e+05	60.375000	59.500000
75%	750.250000	2016.000000	123.000000	7.400000	2.399098e+05	99.177500	72.000000
max	1000.000000	2016.000000	191.000000	9.000000	1.791916e+06	936.630000	100.000000

```
In [51]: movies_df.shape?
```

```
In [53]: movies_df['genre'].describe()
         #ada 1000 value, dimana ada 205 uniq value. Untuk genre. Top genre "action adv,
         scifi", dengan frekuensi hingga 50x.
```

```
Out[53]: count                          1000
         unique                          207
         top        Action,Adventure,Sci-Fi
         freq                             50
         Name: genre, dtype: object
```

```
In [56]: movies_df['genre'].value_counts().head(100) #memunculkan jumlah per genre
```

```
Out[56]: Action,Adventure,Sci-Fi       50
         Drama                          48
         Comedy,Drama,Romance           35
         Comedy                         32
         Drama,Romance                  31
         Comedy,Drama                   27
         Action,Adventure,Fantasy       27
         Animation,Adventure,Comedy     27
         Comedy,Romance                 26
         Crime,Drama,Thriller           24
         Crime,Drama,Mystery            23
         Action,Adventure,Drama         18
         Action,Crime,Drama             17
         Horror,Thriller                16
         Drama,Thriller                 15
         Adventure,Family,Fantasy       14
         Biography,Drama                14
         Action,Adventure,Comedy        14
         Biography,Drama,History        14
         Action,Comedy,Crime            12
         Action,Crime,Thriller          12
         Horror                         11
         Action,Adventure,Thriller      11
         Crime,Drama                    10
         Thriller                        9
         Animation,Action,Adventure      9
         Horror,Mystery,Thriller         9
         Action,Thriller                 9
         Biography,Crime,Drama           9
         Drama,Mystery,Thriller          8
                                        ..
         Horror,Sci-Fi,Thriller          3
         Action,Fantasy,Horror           3
         Action,Crime,Mystery            3
         Drama,War                       3
         Comedy,Drama,Family             3
         Adventure,Sci-Fi,Thriller       3
         Comedy,Fantasy,Horror           3
         Action,Mystery,Sci-Fi           3
         Biography,Drama,Thriller        3
         Action,Sci-Fi                   3
         Adventure,Drama,Thriller        3
         Action,Mystery,Thriller         3
         Action,Drama,Fantasy            3
         Action,Comedy,Fantasy           3
         Adventure,Fantasy               3
         Action,Adventure                3
         Crime,Drama,Horror              3
         Drama,Romance,Thriller          3
         Fantasy,Horror                  2
         Comedy,Drama,Horror             2
         Comedy,Horror,Thriller          2
         Action,Drama,Mystery            2
         Action,Adventure,Western        2
         Adventure,Comedy                2
         Adventure,Comedy,Romance        2
         Adventure,Drama,Romance         2
         Comedy,Drama,Fantasy            2
         Action,Drama,Sport              2
         Action,Adventure,Romance        2
         Comedy,Horror                   2
         Name: genre, Length: 100, dtype: int64
```

Memunculkan korelasi

```
In [99]: movies_df.corr()
```

Out[99]:

	rank	year	runtime	rating	votes	revenue_millions	metascore
rank	1.000000	-0.261605	-0.221739	-0.219555	-0.283876	-0.252996	-0.191869
year	-0.261605	1.000000	-0.164900	-0.211219	-0.411904	-0.117562	-0.079305
runtime	-0.221739	-0.164900	1.000000	0.392214	0.407062	0.247834	0.211978
rating	-0.219555	-0.211219	0.392214	1.000000	0.511537	0.189527	0.631897
votes	-0.283876	-0.411904	0.407062	0.511537	1.000000	0.607941	0.325684
revenue_millions	-0.252996	-0.117562	0.247834	0.189527	0.607941	1.000000	0.133328
metascore	-0.191869	-0.079305	0.211978	0.631897	0.325684	0.133328	1.000000

DataFrame slicing, selecting, extracting </p>

Mengembalikan data dalam bentu series

```
In [100]: genre_col = movies_df['genre']

          type(genre_col)
```

Out[100]: pandas.core.series.Series

Mengembalikan data dalam bentu Dataframe

```
In [ ]: genre_col = movies_df[['genre']]

        type(genre_col)
```

Menambah kolom Dataframe

```
In [106]: subset = movies_df[['genre', 'rating']]

          subset.head()
```

Out[106]:

	genre	rating
Title		
Guardians of the Galaxy	Action,Adventure,Sci-Fi	8.1
Prometheus	Adventure,Mystery,Sci-Fi	7.0
Split	Horror,Thriller	7.3
Sing	Animation,Comedy,Family	7.2
Suicide Squad	Action,Adventure,Fantasy	6.2

Menacari data

```
In [107]: prom = movies_df.loc["Prometheus"]

          prom
```

```
Out[107]: rank                                              2
          genre                      Adventure,Mystery,Sci-Fi
          description        Following clues to the origin of mankind, a te...
          director                               Ridley Scott
          actors             Noomi Rapace, Logan Marshall-Green, Michael Fa...
          year                                           2012
          runtime                                         124
          rating                                            7
          votes                                        485820
          revenue_millions                             126.46
          metascore                                        65
          Name: Prometheus, dtype: object
```

Menggunakan Slice

```
In [108]: movie_subset = movies_df.loc['Prometheus':'Sing'] #mengambil data dari Promethe
          us ke sing

          #movie_subset = movies_df.iloc[1:4]

          movie_subset
```

Out[108]:

	rank	genre	description	director	actors	year	runtime	rati
Title								
Prometheus	2	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	.
Split	3	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	117	.
Sing	4	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108	.

Melakukan seleksi data dengan kondisi tertentu

Misal, ingin melihat film yang disutradarai Ridley dengan rating > 8.

```
In [110]: condition = (movies_df['director'] == "Ridley Scott") #Memunculkan series film,
          dan niai boolean apakah disutradarai atau tidak

          condition.head()
```

```
Out[110]: Title
          Guardians of the Galaxy     False
          Prometheus                   True
          Split                       False
          Sing                        False
          Suicide Squad               False
          Name: director, dtype: bool
```

Memunculkan film yang true di sutradari Ridley

Agar bisa direturn jika true, kita retirn juga bentuk dataframe

Select `movies_df` where `movies_df --> director` equals Ridley Scott.

In [115]: `movies_df[movies_df['director'] == "Ridley Scott"]`

Out[115]:

Title	rank	genre	description	director	actors	year	runtime	rating	votes
Prometheus	2	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	7.0	485820
The Martian	103	Adventure,Drama,Sci-Fi	An astronaut becomes stranded on Mars after hi...	Ridley Scott	Matt Damon, Jessica Chastain, Kristen Wiig, Ka...	2015	144	8.0	556097
Robin Hood	388	Action,Adventure,Drama	In 12th century England, Robin and his band of...	Ridley Scott	Russell Crowe, Cate Blanchett, Matthew Macfady...	2010	140	6.7	221117
American Gangster	471	Biography,Crime,Drama	In 1970s America, a detective works to bring d...	Ridley Scott	Denzel Washington, Russell Crowe, Chiwetel Eji...	2007	157	7.8	337835
Exodus: Gods and Kings	517	Action,Adventure,Drama	The defiant leader Moses rises up against the ...	Ridley Scott	Christian Bale, Joel Edgerton, Ben Kingsley, S...	2014	150	6.0	137299
The Counselor	522	Crime,Drama,Thriller	A lawyer finds himself in over his head when h...	Ridley Scott	Michael Fassbender, Penélope Cruz, Cameron Dia...	2013	117	5.3	84927
A Good Year	531	Comedy,Drama,Romance	A British investment broker inherits his uncle...	Ridley Scott	Russell Crowe, Abbie Cornish, Albert Finney, M...	2006	117	6.9	74674
Body of Lies	738	Action,Drama,Romance	A CIA agent on the ground in Jordan hunts down...	Ridley Scott	Leonardo DiCaprio, Russell Crowe, Mark Strong,...	2008	128	7.1	182305

In [116]: `movies_df[movies_df['rating'] >= 8.6].head(3)`

Out[116]:

Title	rank	genre	description	director	actors	year	runtime	rating	vot
Interstellar	37	Adventure,Drama,Sci-Fi	A team of explorers travel through a wormhole ...	Christopher Nolan	Matthew McConaughey, Anne Hathaway, Jessica Ch...	2014	169	8.6	104774
The Dark Knight	55	Action,Crime,Drama	When the menace known as the Joker wreaks havo...	Christopher Nolan	Christian Bale, Heath Ledger, Aaron Eckhart,Mi...	2008	152	9.0	17919
Inception	81	Action,Adventure,Sci-Fi	A thief, who steals corporate secrets through ...	Christopher Nolan	Leonardo DiCaprio, Joseph Gordon-Levitt, Ellen...	2010	148	8.8	158363

Kombinasi Kondisi

Film apa yang disutradarai oleh 'Christopher Nolan' atau 'Ridley Scott' dan memiliki rating > 8.5

In [119]:
```
movies_df[(movies_df['director'] == 'Christopher Nolan')
          | (movies_df['director'] == 'Ridley Scott')
          & (movies_df['rating']>8.5)].head()
```

Out[119]:

Title	rank	genre	description	director	actors	year	runtime	rating	vot
Interstellar	37	Adventure,Drama,Sci-Fi	A team of explorers travel through a wormhole ...	Christopher Nolan	Matthew McConaughey, Anne Hathaway, Jessica Ch...	2014	169	8.6	10477
The Dark Knight	55	Action,Crime,Drama	When the menace known as the Joker wreaks havo...	Christopher Nolan	Christian Bale, Heath Ledger, Aaron Eckhart,Mi...	2008	152	9.0	17919
The Prestige	65	Drama,Mystery,Sci-Fi	Two stage magicians engage in competitive one-...	Christopher Nolan	Christian Bale, Hugh Jackman, Scarlett Johanss...	2006	130	8.5	9131
Inception	81	Action,Adventure,Sci-Fi	A thief, who steals corporate secrets through ...	Christopher Nolan	Leonardo DiCaprio, Joseph Gordon-Levitt, Ellen...	2010	148	8.8	15836
The Dark Knight Rises	125	Action,Thriller	Eight years after the Joker's reign of anarchy...	Christopher Nolan	Christian Bale, Tom Hardy, Anne Hathaway,Gary ...	2012	164	8.5	12226

Kalau ada >2 sutradara yang mau di cek gimana?

pakai `isin` untuk mengecek apakah ada dalam list

```
In [120]: movies_df[movies_df['director'].isin(['Christopher Nolan', 'Ridley Scott'])].head()
```

Out[120]:

Title	rank	genre	description	director	actors	year	runtime	rating	v
Prometheus	2	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	7.0	48!
Interstellar	37	Adventure,Drama,Sci-Fi	A team of explorers travel through a wormhole ...	Christopher Nolan	Matthew McConaughey, Anne Hathaway, Jessica Ch...	2014	169	8.6	104'
The Dark Knight	55	Action,Crime,Drama	When the menace known as the Joker wreaks havo...	Christopher Nolan	Christian Bale, Heath Ledger, Aaron Eckhart,Mi...	2008	152	9.0	179
The Prestige	65	Drama,Mystery,Sci-Fi	Two stage magicians engage in competitive one-...	Christopher Nolan	Christian Bale, Hugh Jackman, Scarlett Johanss...	2006	130	8.5	91:
Inception	81	Action,Adventure,Sci-Fi	A thief, who steals corporate secrets through ...	Christopher Nolan	Leonardo DiCaprio, Joseph Gordon-Levitt, Ellen...	2010	148	8.8	158:

Cek film dari tahun 2005 --> 2010 dengan rating > 8 tapi revenue ada di quartil pertama

```
In [121]: movies_df[
              ((movies_df['year'] >= 2005) & (movies_df['year'] <= 2010))
              & (movies_df['rating'] > 8.0)
              & (movies_df['revenue_millions'] < movies_df['revenue_millions'].quantile
          (0.25))
          ]
```

Out[121]:

Title	rank	genre	description	director	actors	year	runtime	rating	v
3 Idiots	431	Comedy,Drama	Two friends are searching for their long lost ...	Rajkumar Hirani	Aamir Khan, Madhavan, Mona Singh, Sharman Joshi	2009	170	8.4	238
The Lives of Others	477	Drama,Thriller	In 1984 East Berlin, an agent of the secret po...	Florian Henckel von Donnersmarck	Ulrich Mühe, Martina Gedeck,Sebastian Koch, Ul...	2006	137	8.5	278
Incendies	714	Drama,Mystery,War	Twins journey to the Middle East to discover t...	Denis Villeneuve	Lubna Azabal, Mélissa Désormeaux-Poulin, Maxim...	2010	131	8.2	92
Taare Zameen Par	992	Drama,Family,Music	An eight-year-old boy is thought to be a lazy ...	Aamir Khan	Darsheel Safary, Aamir Khan, Tanay Chheda, Sac...	2007	165	8.5	102

```
In [123]: movies_df.describe()
```

Out[123]:

	rank	year	runtime	rating	votes	revenue_millions	metascore
count	1000.000000	1000.000000	1000.000000	1000.000000	1.000000e+03	1000.000000	936.000000
mean	500.500000	2012.783000	113.172000	6.723200	1.698083e+05	82.956376	58.985043
std	288.819436	3.205962	18.810908	0.945429	1.887626e+05	96.412043	17.194757
min	1.000000	2006.000000	66.000000	1.900000	6.100000e+01	0.000000	11.000000
25%	250.750000	2010.000000	100.000000	6.200000	3.630900e+04	17.442500	47.000000
50%	500.500000	2014.000000	111.000000	6.800000	1.107990e+05	60.375000	59.500000
75%	750.250000	2016.000000	123.000000	7.400000	2.399098e+05	99.177500	72.000000
max	1000.000000	2016.000000	191.000000	9.000000	1.791916e+06	936.630000	100.000000

Menggunakan Fungsi </p>

```
In [124]: # Membuat fungsi
          def rating_function(x):
              if x >= 8.0:
                  return "good"
              else:
                  return "bad"
```

Mengaplikasikan fungsi

In [128]:
```python
movies_df["rating_category"] = movies_df["rating"].apply(rating_function)
#menambah kolom baru bernama "rating_category" dengan data rating yang dikenai
fungsi

movies_df.head(2)
```

Out[128]:

Title	rank	genre	description	director	actors	year	runtime	rating	votes	rev
Guardians of the Galaxy	1	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121	8.1	757074	
Prometheus	2	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	7.0	485820	

Drop kolom

In [129]:
```python
del movies_df['rating_category']
movies_df.head()
```

Out[129]:

Title	rank	genre	description	director	actors	year	runtime	rati
Guardians of the Galaxy	1	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121	ε
Prometheus	2	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	·
Split	3	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	117	·
Sing	4	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108	·
Suicide Squad	5	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	123	(

Membuat fungsi anonim menggunakan lambda

In [130]:
```python
movies_df["rating_category"] = movies_df["rating"].apply(lambda x: 'good' if x >= 8.0 else 'bad')

movies_df.head(2)
```

Out[130]:

Title	rank	genre	description	director	actors	year	runtime	rating	votes	rev
Guardians of the Galaxy	1	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121	8.1	757074	
Prometheus	2	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	7.0	485820	

read.. plot
series

pandas

youtube.com/isnaalfi

In [ ]: