



LEARNING PROGRESS REVIEW

Week 6

Entropy Team

OUR TEAM

Entropy Team



Adhang Muntaha Muhammad

<https://www.linkedin.com/in/adhangmuntaha/>



Aziz Fauzi

<https://www.linkedin.com/in/aziz-fauzi-a6904711b/>



Iwan Wahyu

<https://www.linkedin.com/in/iwan-wahyu-setyawan-506809183>



Marcellina Alvita F

<https://www.linkedin.com/in/marcellina-alvita-faustina-63a284226>



Ramadhan Luthfan

<https://www.linkedin.com/in/luthfan-mahathir-91369b18b>

DAFTAR ISI

1.

Introduction to Kaggle

Pengenalan platform Kaggle

2.

Analytical & Critical Thinking

Kerangka berpikir dalam pemecahan masalah

3.

Intermediate Dataframe

Materi Pandas dataframe (menengah)

01

Introduction to Kaggle

Pengenalan platform
Kaggle

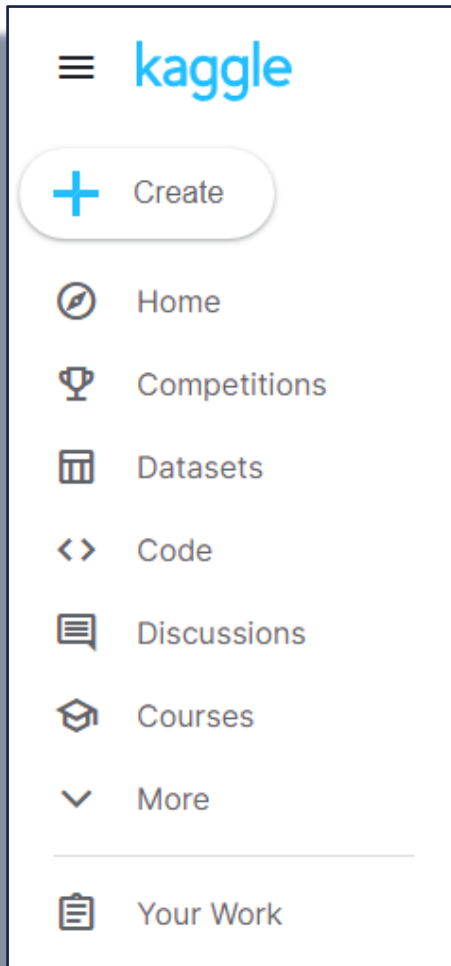
Kaggle

The Kaggle logo, featuring the word "kaggle" in a lowercase, blue, sans-serif font. A small "TM" trademark symbol is positioned to the upper right of the final "e".

kaggle™

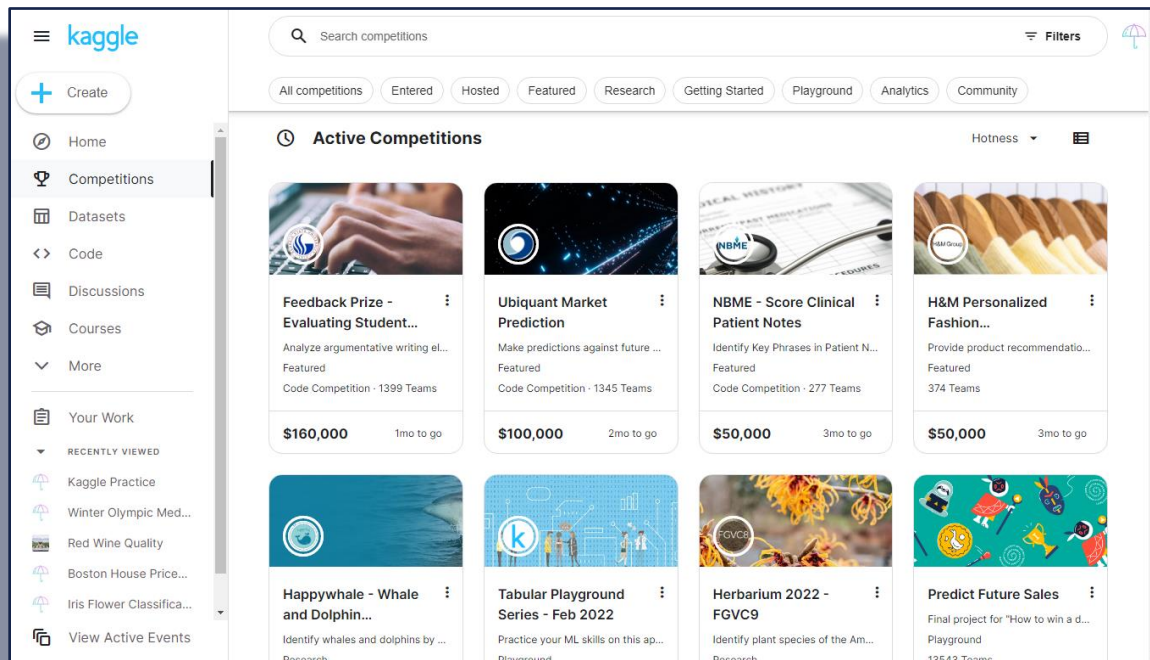
Kaggle merupakan **komunitas**
online bagi ***data scientist*** dan
praktisi ***machine learning***

Fitur dalam Kaggle



- Beberapa fitur dalam Kaggle:
 - **Competitions**
 - **Datasets**
 - **Code**
 - **Discussions**
 - **Courses**
- Kaggle dapat **menjadi portofolio** bagi seorang *data scientist*

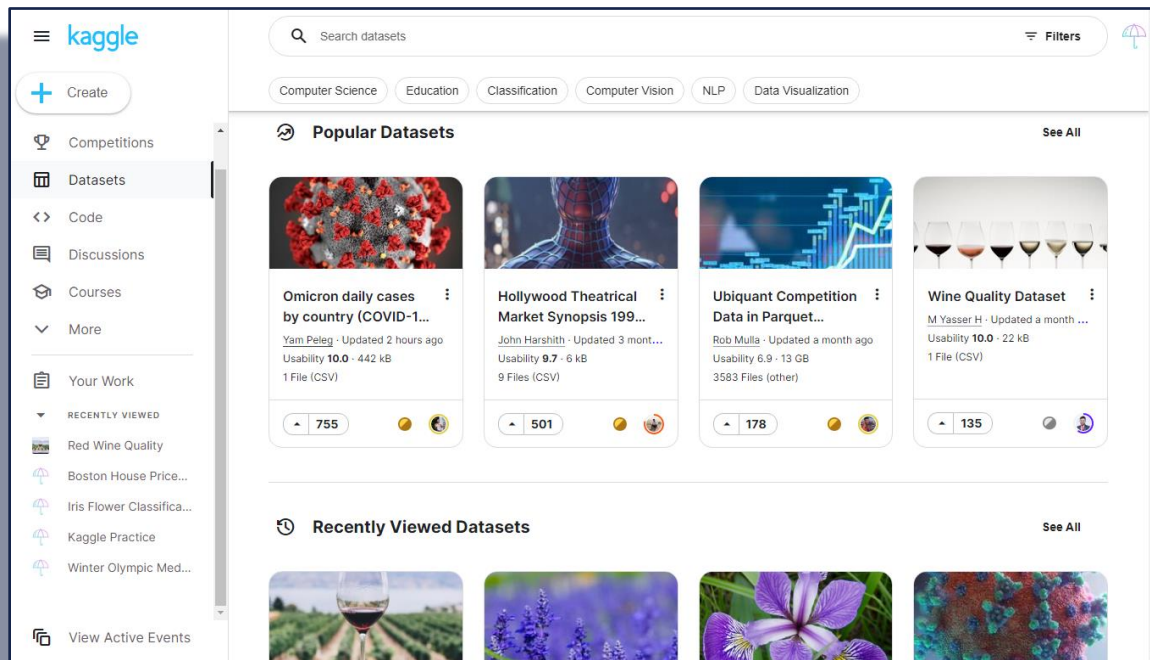
Fitur – Competitions



Setiap *user* dapat **melatih skill** dengan mengikuti kompetisi yang memiliki beragam hadiah, seperti:

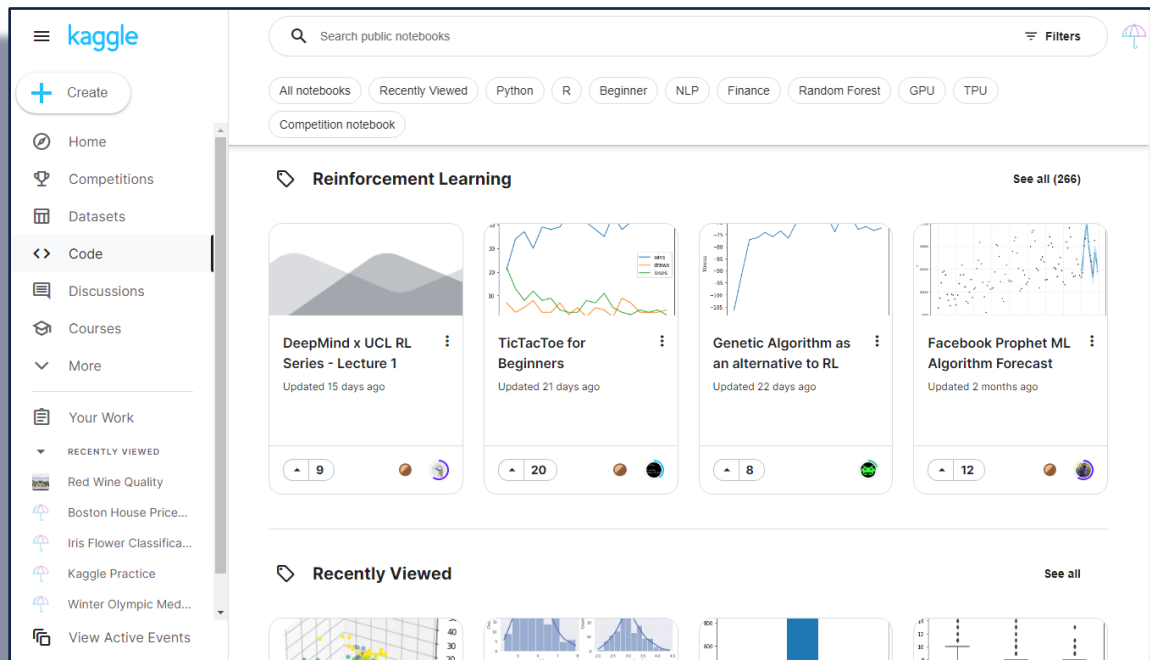
- **Uang**
- *Merchandise (swag)*
- Pengakuan dari penyelenggara (**kudos**)
- Tanpa hadiah, alias hanya untuk **pembelajaran**

Fitur – Datasets



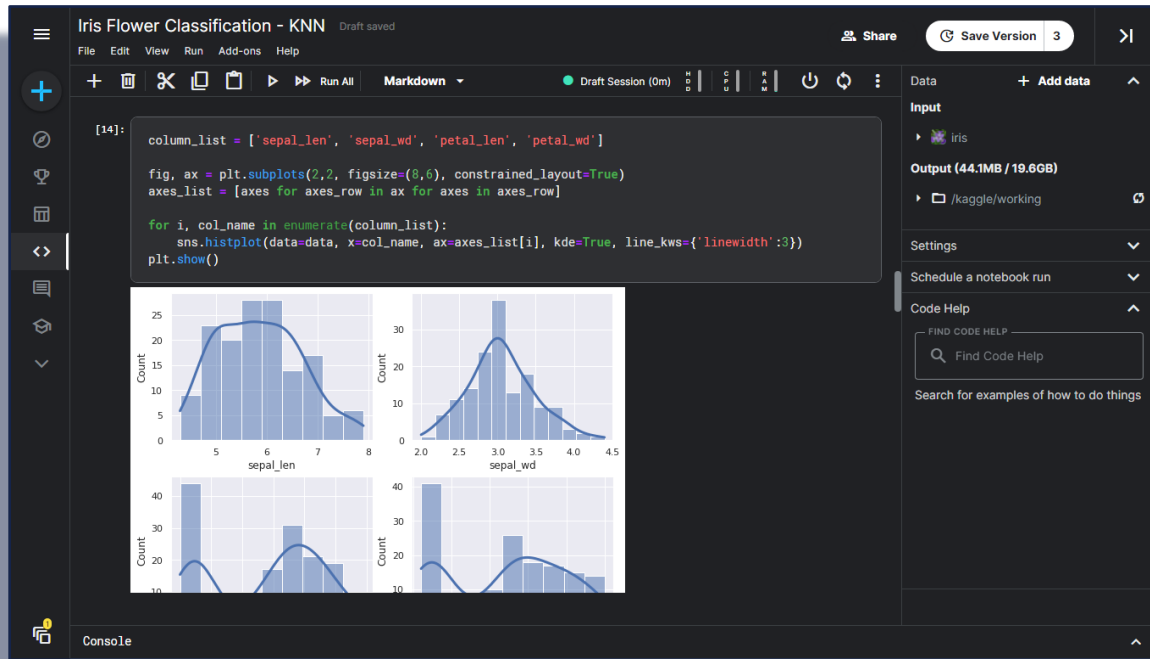
- Setiap *user* dapat **membagikan** (*upload*) **dataset** ke dalam Kaggle
- Setiap *user* dapat **menggunakan** semua **dataset** yang ada pada Kaggle

Fitur – Code



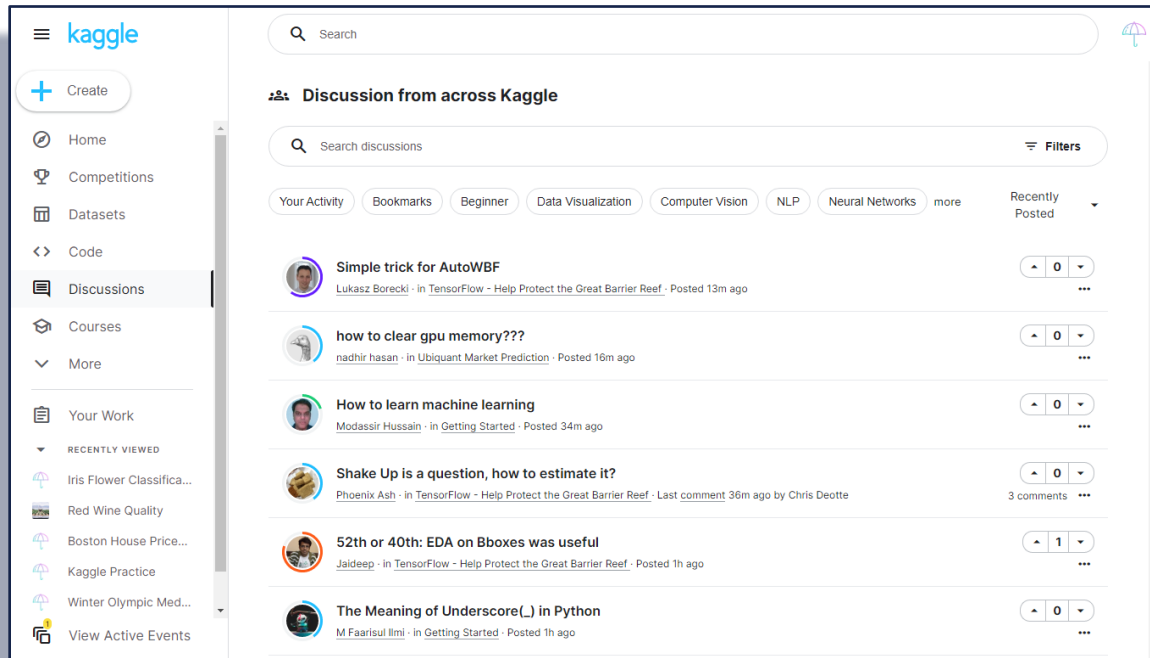
- Setiap *user* dapat **melihat code** dari *user* lain yang bersifat **publik**
- Setiap *user* juga dapat **membagikan code** secara **publik**

Fitur – Code (Notebook)



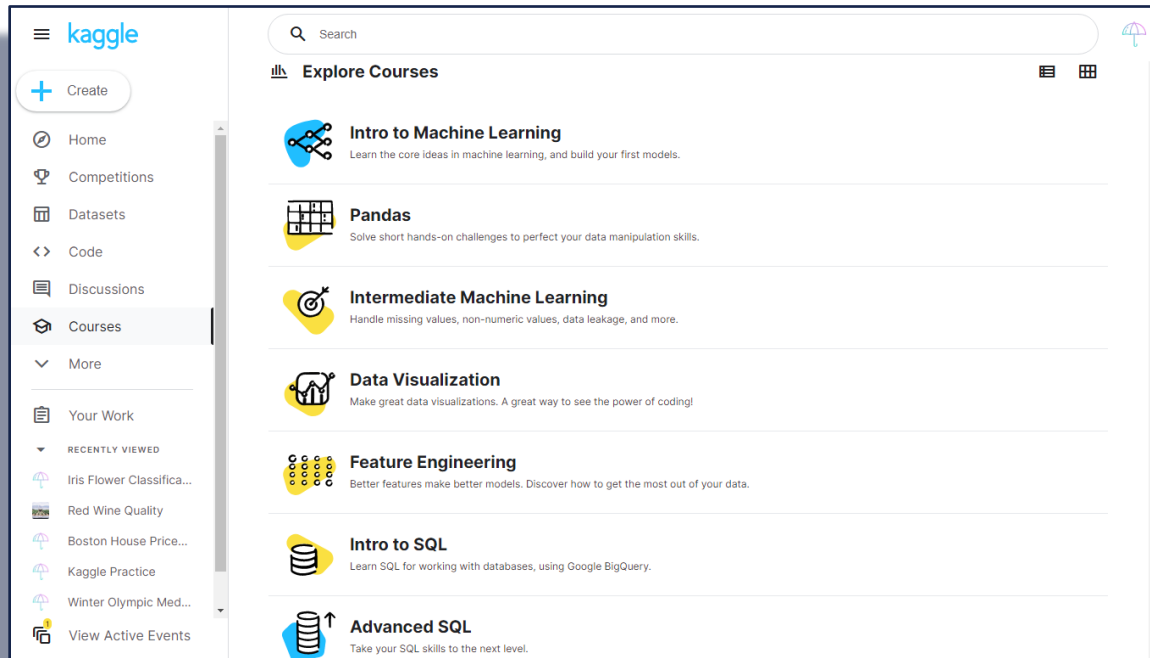
- Kaggle menyediakan **web-based notebook** yang dapat digunakan untuk **menjalankan code secara online**
- *Notebook* tersebut juga dapat digunakan untuk **kolaborasi** bersama *user* lain

Fitur – Discussions



- Setiap *user* dapat **berdiskusi tentang segala hal** yang berkaitan dengan *data science* dan *machine learning*
- Setiap *user* juga dapat berdiskusi bahkan **berkolaborasi** terkait suatu *project*

Fitur – Courses



- Kaggle menyediakan **materi pembelajaran** yang dapat diakses secara **gratis**

Kaggle Progression System



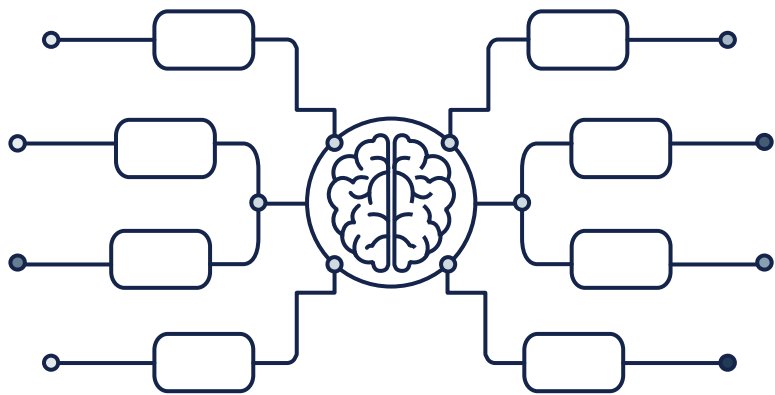
- Setiap *user* dapat memiliki beragam tingkatan (*tier*)
- *Tier* tersebut digunakan untuk melihat **performa** dan **pencapaian** dari tiap *user*

02

Analytical & Critical Thinking

Kerangka berpikir
dalam pemecahan masalah

Design Thinking

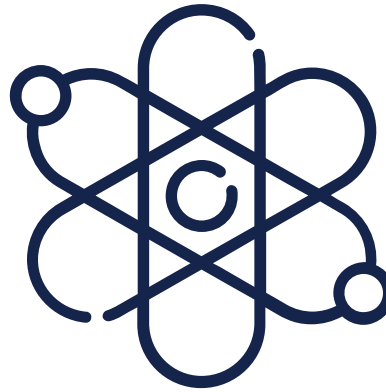


Design thinking adalah istilah yang digunakan untuk mewakili suatu **strategi** atau rencana dalam **pemecahan masalah** dan berfokus pada *user*.

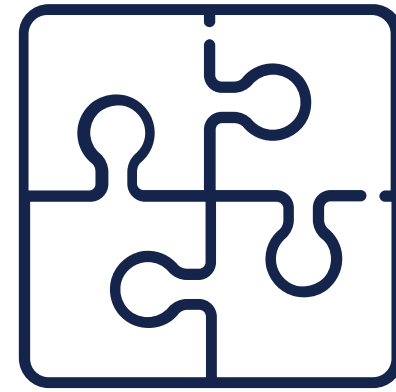
IBM Design Thinking Framework



*The Principle
guide us*



*The Loop
drives us*



*The Keys
align us*

Sumber: <https://www.ibm.com/design/thinking/page/framework>

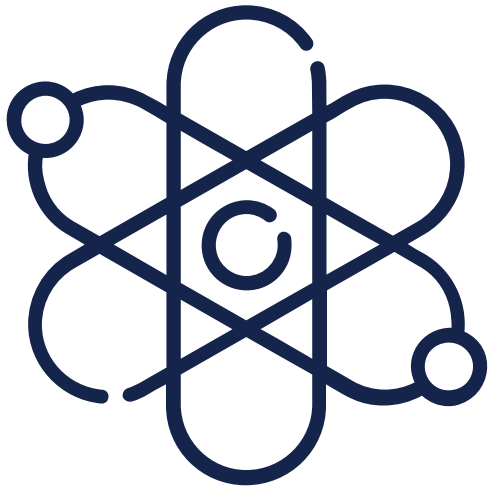
The Principles



Melihat masalah dan solusi sebagai percakapan yang berkelanjutan.

- **Fokus pada *user***
Dorong bisnis dengan membantu *user* mencapai tujuannya.
- ***Reinvention***
Selalu melakukan pembaruan dan perlakukan semua hal sebagai prototipe.
- **Pemberdayaan keberagaman**
Melihat sesuatu dari beberapa sudut pandang dapat menghasilkan ide terobosan yang beragam.

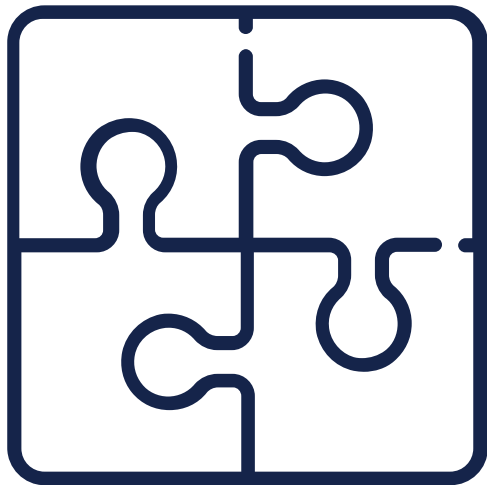
The Loop



Memahami masa kini dan membayangkan masa depan dalam siklus observasi, refleksi, dan pembuatan yang berkelanjutan.

- **Observasi**
Suatu ide terobosan dapat tercipta dari pemahaman tentang masalah di dunia nyata.
- **Refleksi**
Menyinkronkan langkah dengan tim dan menyatukan apa yang telah dipelajari.
- **Pembuatan**
Satu-satunya cara untuk melihat hasil dari suatu gagasan adalah dengan membuatnya.

The Keys



Membantu menjaga tim tetap fokus dan selaras pada hasil yang penting bagi *user*.

- ***Hills***
Menyelaraskan kita sebagai tim.
- ***Playback***
Menyelaraskan kita sepanjang waktu.
- ***Sponsor users***
Menyelaraskan kita dengan kebutuhan *user*.

03

Intermediate Dataframe

Materi Pandas dataframe
(menengah)

Sorting

Data awal

	name	age	score
0	Entropy	25	80
1	Team	24	90
2	Digital	24	70
3	Skola	25	85

Digunakan untuk **mengurutkan** dataframe **berdasarkan** suatu **kolom** atau beberapa kolom

Sorting – Satu Kolom

```
in | # ascending  
   | data.sort_values(by='score')  
  
in | # descending  
   | data.sort_values(by='score', ascending=False)
```

Diurutkan secara *ascending*

	name	age	score
2	Digital	24	70
0	Entropy	25	80
3	Skola	25	85
1	Team	24	90

Diurutkan secara *descending*

	name	age	score
1	Team	24	90
3	Skola	25	85
0	Entropy	25	80
2	Digital	24	70

Sorting – Beberapa Kolom

```
in | data.sort_values(by=['age', 'score'])
```

- **Sorting dimulai dari kolom pertama.** Jika terdapat data yang sama, maka diurutkan lagi berdasarkan **kolom kedua, dan seterusnya**
- Baris ke-1 dan ke-2 memiliki 'age' yang sama, sehingga urutan baris tersebut ditentukan oleh 'score'

	name	age	score
2	Digital	24	70
1	Team	24	90
0	Entropy	25	80
3	Skola	25	85

Filtering

Data awal

	name	age	score
0	Entropy	25	80
1	Team	24	90
2	Digital	24	70
3	Skola	25	85

Digunakan untuk **mengekstrak suatu bagian** dari dataframe

Filtering – Kolom Tertentu

```
in data[['name', 'age']]

in data.filter(items=['name', 'age'])
```

	name	age
0	Entropy	25
1	Team	24
2	Digital	24
3	Skola	25

- Dapat menggunakan **double square brackets** `[[nama kolom]]` ataupun fungsi **filter()**
- Penggunaan **double square brackets** akan menghasilkan **dataframe** baru, sedangkan **single square bracket** akan menghasilkan **series**

Filtering – Data Tertentu

```
in data.loc[:2, ['name', 'age']]
```

```
in data.iloc[:3, [1, 2]]
```

	name	age
0	Entropy	25
1	Team	24
2	Digital	24

- ***loc*** digunakan untuk memilih baris dan kolom **berdasarkan namanya**
- ***iloc*** digunakan untuk memilih baris dan kolom **berdasarkan indeksnya**
- *Syntax* penulisannya yaitu **[baris, kolom]**

Filtering – Berdasarkan Kondisi

```
in | data.loc[data['score'] > 80]
```

	name	age	score
1	Team	24	90
3	Skola	25	85

- Digunakan untuk memilih data berdasarkan **suatu kondisi**
- Kondisi dapat berjumlah lebih dari satu
- **Beberapa kondisi** dapat dihubungkan dengan **operator logika**
 - & – operator ‘dan’
 - | – operator ‘atau’

Membuat Kolom Baru

```
in | data['score_10'] = data['score']/10
```

Sebelum

	name	age	score
0	Entropy	25	80
1	Team	24	90
2	Digital	24	70
3	Skola	25	85



Sesudah

	name	age	score	score_10
0	Entropy	25	80	8.0
1	Team	24	90	9.0
2	Digital	24	70	7.0
3	Skola	25	85	8.5

Grouping

Data awal

	name	city	age
0	Entropy	Jogja	25
1	Team	Jakarta	26
2	Digital	Jogja	27
3	Skola	Jakarta	28

Digunakan untuk **mengelompokkan data** berdasarkan suatu kolom atau beberapa kolom, sekaligus **melakukan agregasi**

Grouping – Satu Agregasi

```
in | data[['city', 'age']].groupby('city').mean()
```

- Misal mencari rata-rata usia (kolom 'age') berdasarkan nama kota (kolom 'city')
- Setelah *grouping* berdasarkan nama kota, nama kota akan menjadi indeks

city	age
Jakarta	27.0
Jogja	26.0

Grouping – Beberapa Agregasi

```
in | data[['city', 'age']].groupby('city').agg(avg_age = ('age', 'mean'),  
                                              max_age = ('age', 'max'))
```

- Misal mencari nilai terbesar dan rata-rata usia (kolom 'age') berdasarkan nama kota (kolom 'city')

city	avg_age	max_age
Jakarta	27.0	28
Jogja	26.0	27

Grouping – Beberapa Kolom

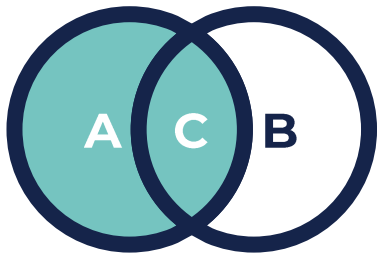
```
in | data[['city', 'class', 'age']].groupby(['city', 'class']).mean()
```

	name	city	class	age
0	Entropy	Jogja	DS	25
1	Team	Jakarta	DS	26
2	Digital	Jogja	DE	27
3	Skola	Jakarta	DS	28

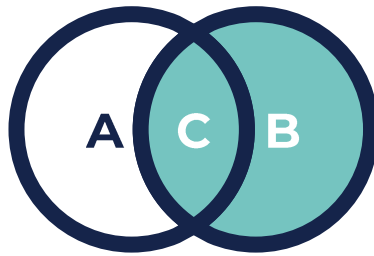


city	class	age
Jakarta	DS	27.0
Jogja	DE	27.0
	DS	25.0

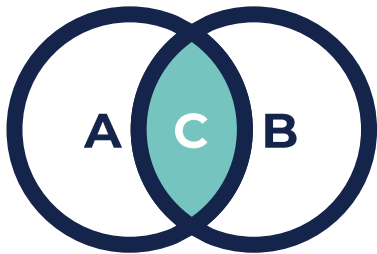
Merging



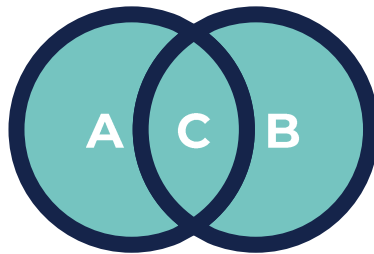
LEFT JOIN
(A + C)



RIGHT JOIN
(C + B)



INNER JOIN
(C)



FULL OUTER JOIN
(A + C + B)

Digunakan untuk
menggabungkan beberapa
dataframe dari segi **kolom**

Merging – Inner Join

```
in | data_1.merge(data_2, how='inner', on='id')
```

- Hasilnya hanyalah data yang beririsan

data_1

	id	name
0	1	Entropy
1	2	Team
2	3	Digital



data_2

	id	class
0	2	DS
1	3	DE
2	4	DA



	id	name	class
0	2	Team	DS
1	3	Digital	DE

Merging – Left Join

```
in | data_1.merge(data_2, how='left', on='id')
```

- Hasilnya adalah data yang beririsan dan semua data pada dataframe pertama
- Data yang tidak beririsan akan bernilai null (NaN)

data_1

	id	name
0	1	Entropy
1	2	Team
2	3	Digital



data_2

	id	class
0	2	DS
1	3	DE
2	4	DA



	id	name	class
0	1	Entropy	NaN
1	2	Team	DS
3	3	Digital	DE

Merging – Right Join

```
in | data_1.merge(data_2, how='right', on='id')
```

- Hasilnya adalah data yang beririsan dan semua data pada dataframe kedua
- Data yang tidak beririsan akan bernilai null (NaN)

data_1

	id	name
0	1	Entropy
1	2	Team
2	3	Digital



data_2

	id	class
0	2	DS
1	3	DE
2	4	DA



	id	name	class
0	2	Team	DS
1	3	Digital	DE
3	4	NaN	DA

Merging – Full Outer Join

```
in | data_1.merge(data_2, how='outer', on='id')
```

- Hasilnya adalah semua data pada dataframe pertama dan kedua
- Data yang tidak beririsan akan bernilai null (NaN)

data_1

	id	name
0	1	Entropy
1	2	Team
2	3	Digital



data_2

	id	class
0	2	DS
1	3	DE
2	4	DA



	id	name	class
0	1	Entropy	NaN
1	2	Team	DS
3	3	Digital	DE
4	4	NaN	DA

THANKS

Entropy Team

CREDITS: This presentation template was originally created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**