LEARNING PROGRESS REVIEW WEEK 6

BY OMICRON (GROUP 3)

TEAM OMICRON MEMBERS

ANUGRAHYAZID GHANI

https://www.linkedin.com/in/anugrah-yazid-7253bb221/

EDO MOHAMMAD HADAD GIBRAN

https://www.linkedin.com/in/edo-gibran-38505a142/

FAJAR ACHMAD

https://www.linkedin.com/in/fajar-achmad-755945111/

MUHAMMAD FIKRI FADILA

https://www.linkedin.com/in/muhammad-fikri-fadila-a551161a6/

OUTLINE

INTRODUCTION TO KAGGLE

ANALYTICAL & CRITICAL THINKING

PANDAS INTERMEDIATE DATAFRAME

DigitalSkola

DigitalSkola

INTRODUCTION TO KAGGLE

BY OMICRON (GROUP 3)

APA ITU KAGGLE?

Online community of data scientists and machine learning practitioners

Platform untuk **BERINTERAKSI**, **MELATIH** dan **MENYELESAIKAN TANTANGAN** terkait data science dan machine learning

DigitalSkola

HISTORY

- Ditemukan April 2010 oleh **Anthony Goldbloom** dan **Jeremy Howard**.
- Tahun 2010 menawarkan **kompetisi** machine learning dan sekarang juga menawarkan platform *public data*, *workbench* berbasis cloud untuk data science, and Artificial Intelligence education.
- Pada 8 Maret 2017, Kaggle diakuisisi oleh **Google**
- Pada bulan Juni 2017, Kaggle mengumumkan bahwa ia melewati 1 juta user dan komunitas tersebar di 194 negara.
- Pada tahun 2021 memiliki lebih dari 8 juta user.

PURPOSE

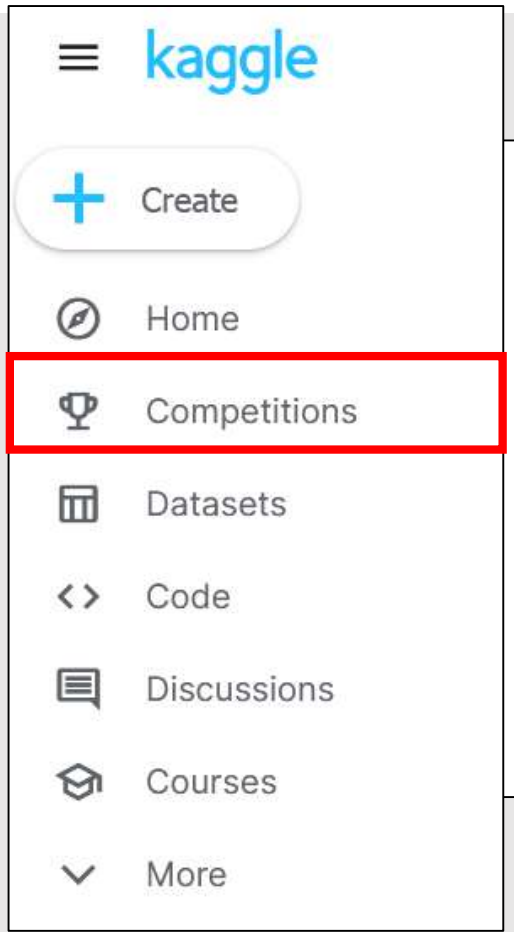
- Menemukan dan publikasikan dataset yang sifatnya umum.
- Menjelajahi dan membangun model dalam data science environment berbasis web.
- Connect dan Bekerja dengan data scientists dan machine learning engineers lain.
- Mengikuti kompetisi untuk memecahkan data science challenges.

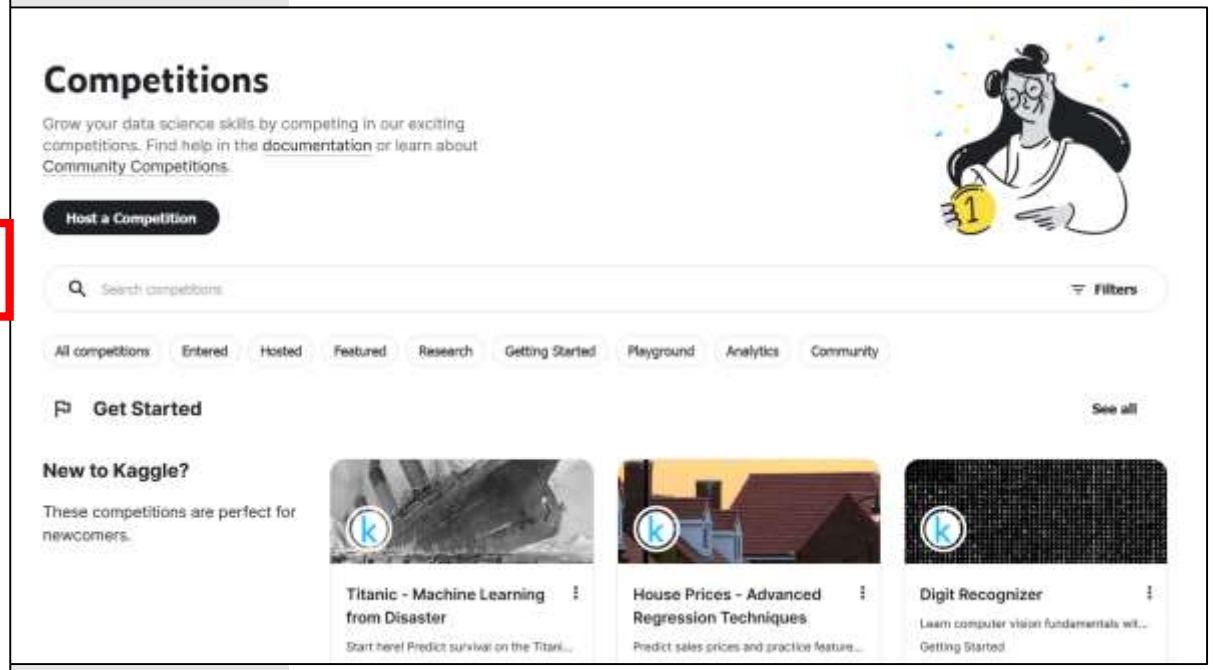
FEATURES

Kaggle merupakan platform hebat yang memiliki berbagai fitur untuk melatih diri sendiri, diantaranya adalah:

- Competitions
- Datasets
- Code
- Discussions
- Courses
- User Rankings

FEATURE 01 : COMPETITIONS





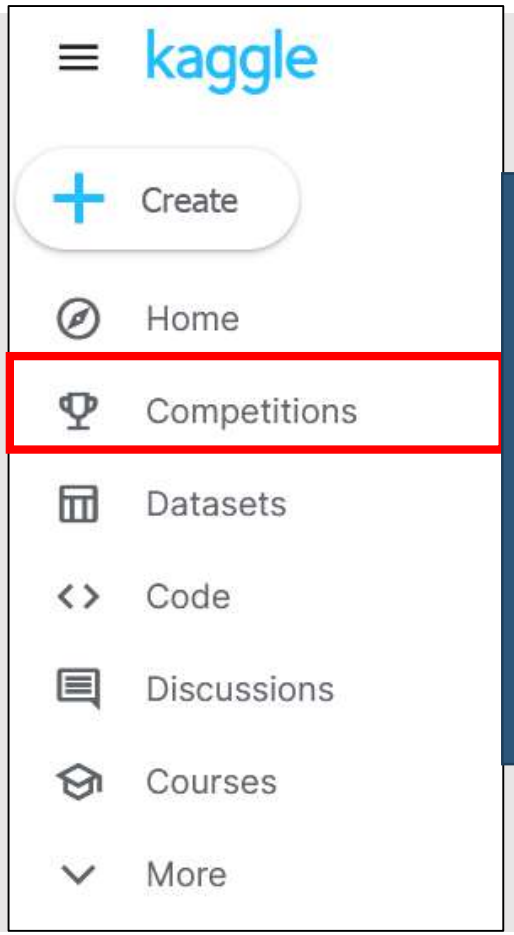
ABOUT

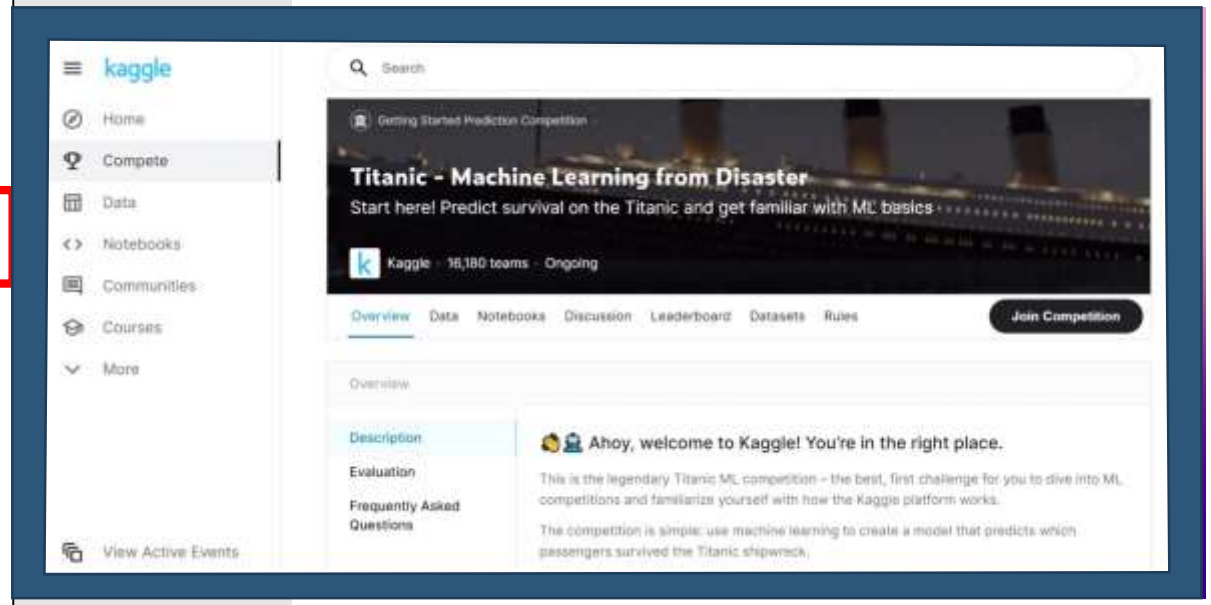
Kompetisi Kaggle secara teratur menarik lebih dari seribu tim dan individu.

Komunitas Kaggle memiliki ribuan kumpulan data publik dan cuplikan kode (disebut "kernel" di Kaggle).

Banyak dari peneliti ini menerbitkan makalah di jurnal peer-review berdasarkan kinerja mereka dalam kompetisi Kaggle.

FEATURE 01 : COMPETITIONS



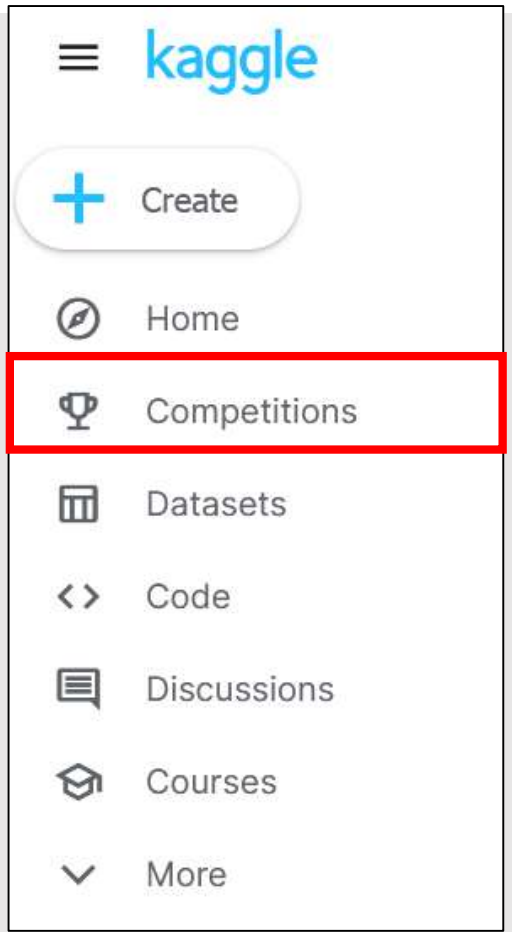


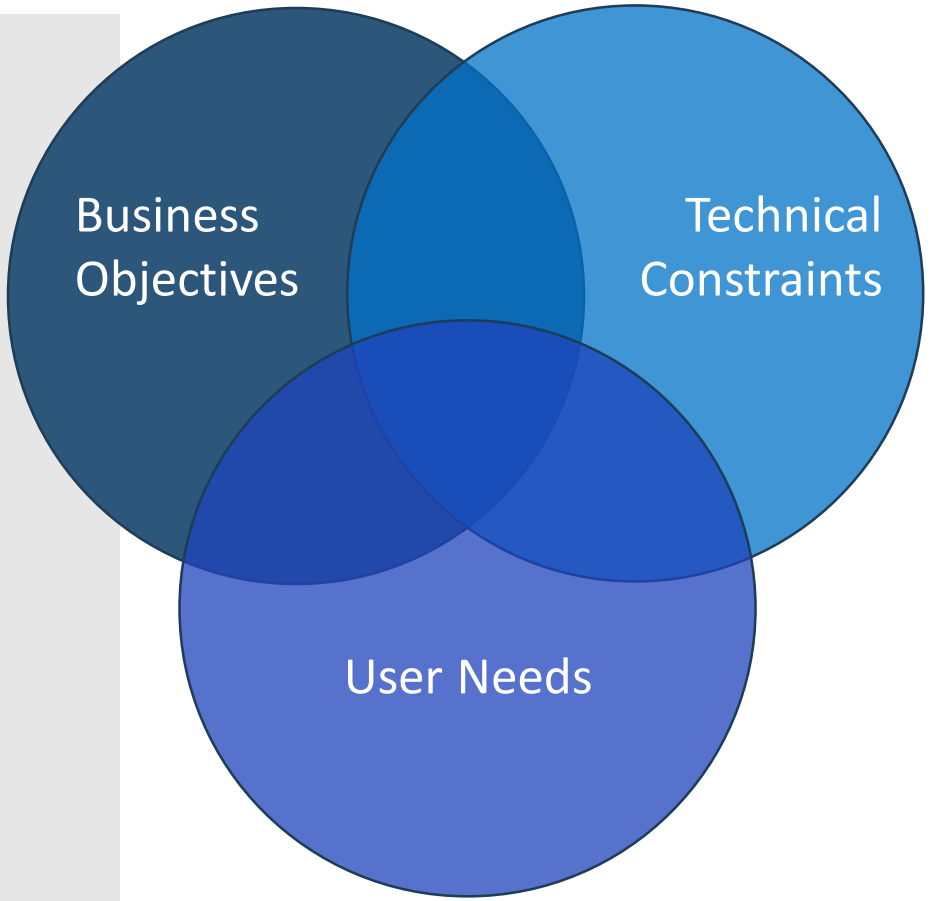
STEPS

1. Click Join Competition
2. Download the data
3. Read the rules
4. Study the open kernels
5. Create your kernel
6. Make a submission
7. Check the leaderboard
8. Improve your score

DigitalSkola

FEATURE 01 : COMPETITIONS





BENEFIT

Problem Solving:

- Mempertajam skill untuk menyelesaikan suatu permasalahan dengan *analytical thinking.*

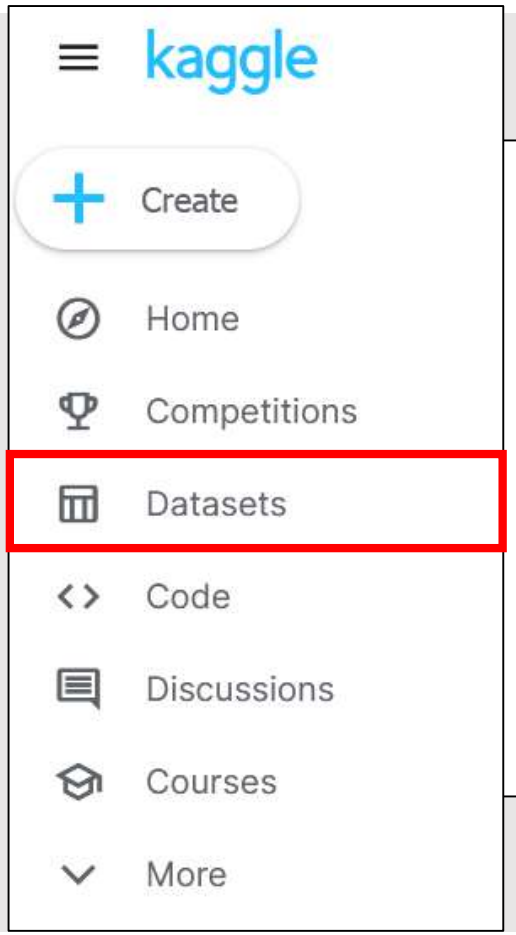
Domain Bisnis:

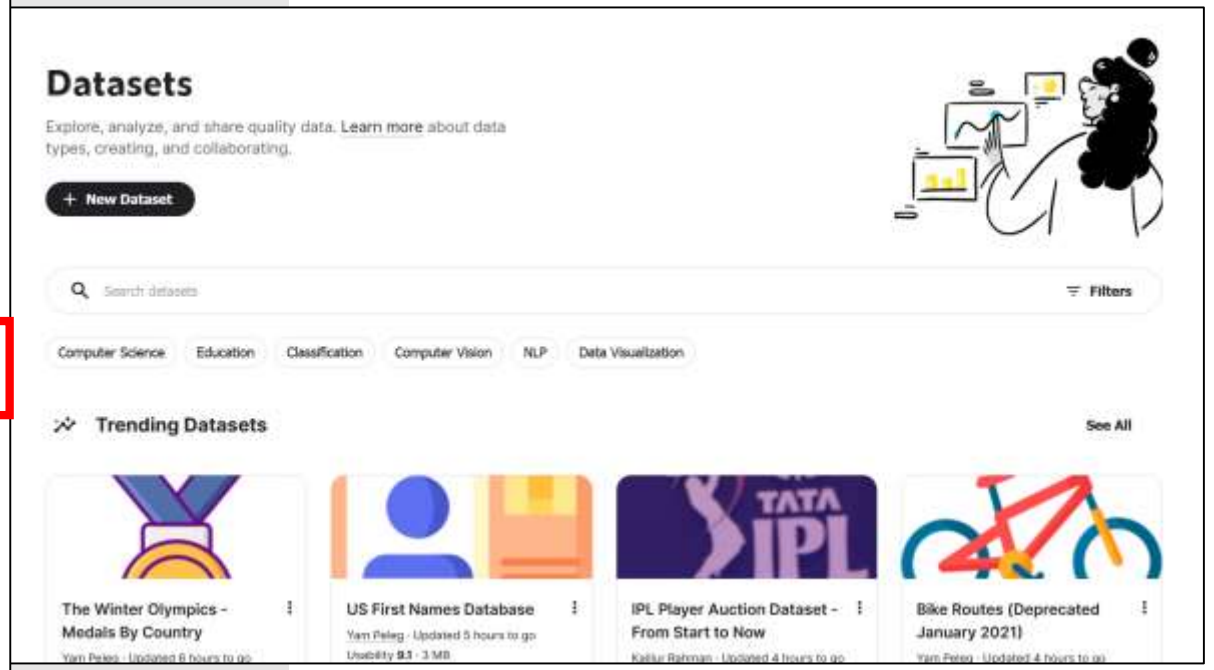
- Menyelesaikan masalah dengan domain bisnis tertentu akan memperkaya wawasan.

Users:

- Menjadikan kompetisi tersebut sebagai bukti yang dapat ditampilkan di portopolio.

FEATURE 02 : DATASETS



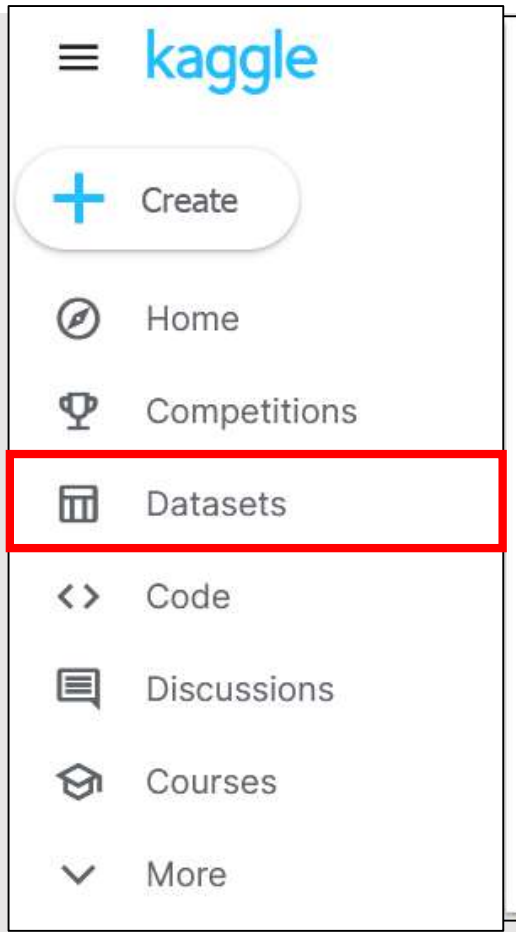


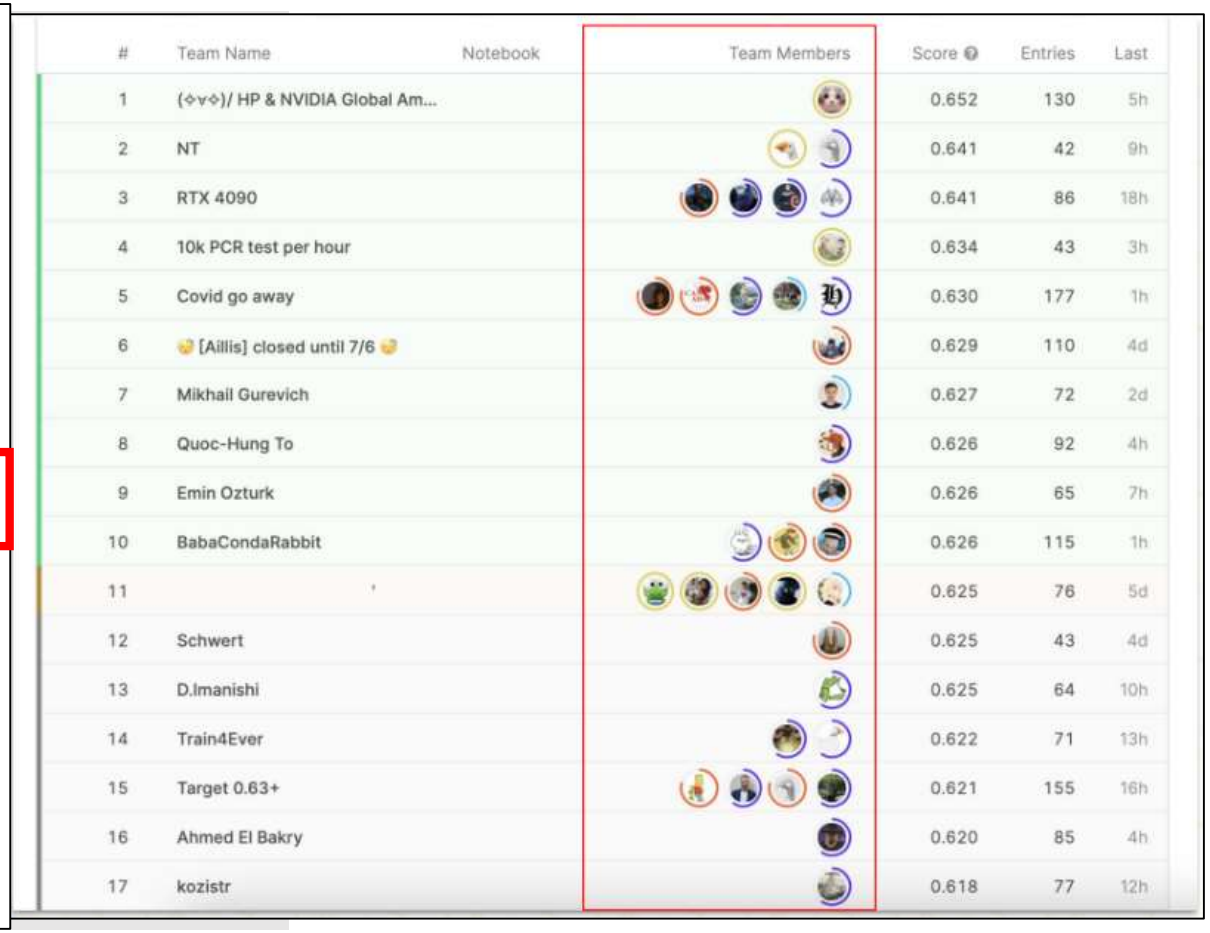
ABOUT

Explore, analyze, and share quality data

Creating and collaborating dengan beragam data scientists dan machine learning engineer denagn beragam tipe data.

FEATURE 02 : DATASETS





#	Team Name	Notebook	Team Members	Score	Entries	Last
1	(✧v✧)/ HP & NVIDIA Global Am...			0.652	130	5h
2	NT			0.641	42	9h
3	RTX 4090			0.641	86	18h
4	10k PCR test per hour			0.634	43	3h
5	Covid go away			0.630	177	1h
6	[Aillis] closed until 7/6			0.629	110	4d
7	Mikhail Gurevich			0.627	72	2d
8	Quoc-Hung To			0.626	92	4h
9	Emin Ozturk			0.626	65	7h
10	BabaCondaRabbit			0.626	115	1h
11				0.625	76	5d
12	Schwert			0.625	43	4d
13	D.Imanishi			0.625	64	10h
14	Train4Ever			0.622	71	13h
15	Target 0.63+			0.621	155	16h
16	Ahmed El Bakry			0.620	85	4h
17	kozistr			0.618	77	12h

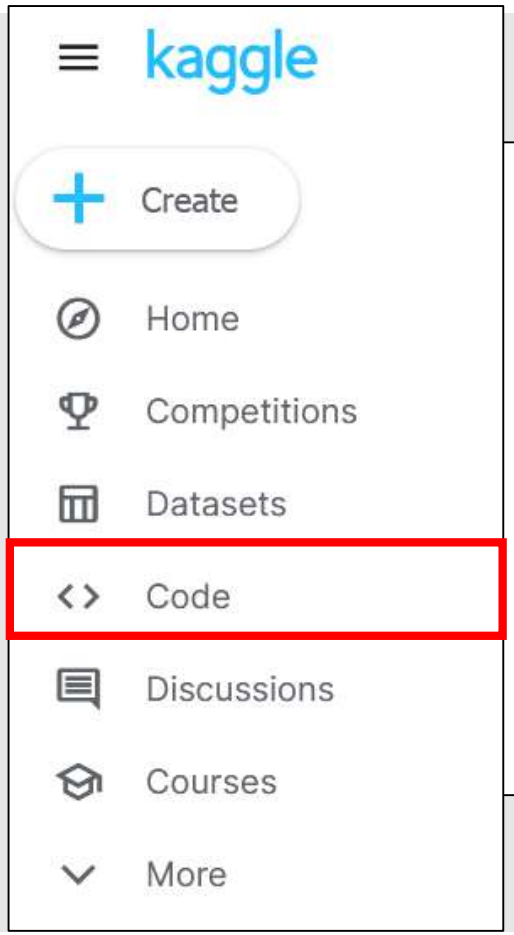
COLLABORATING ON NOTEBOOKS

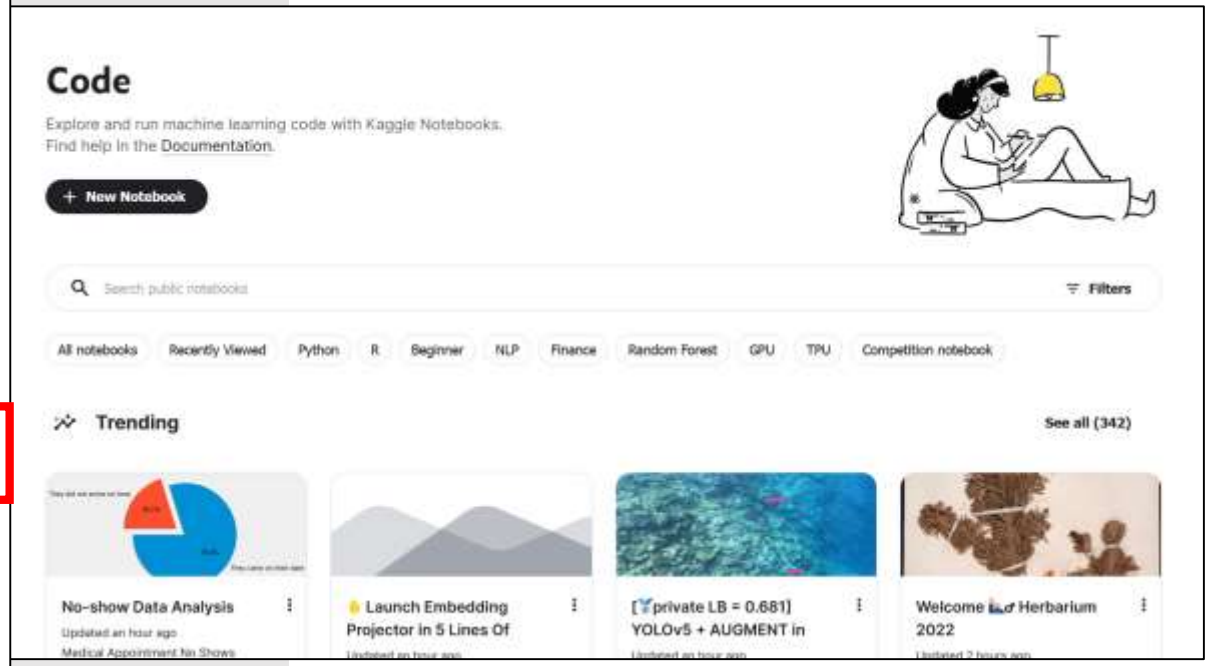
Notebooks collaboration is a powerful feature.

Memungkinkan banyak pengguna untuk memiliki dan mengedit bersama Notebook.

Misalnya, Anda dapat bekerja dengan rekan tim Kompetisi untuk iterasi model atau berkolaborasi dengan teman sekelas dalam data science project.

FEATURE 03 : CODE



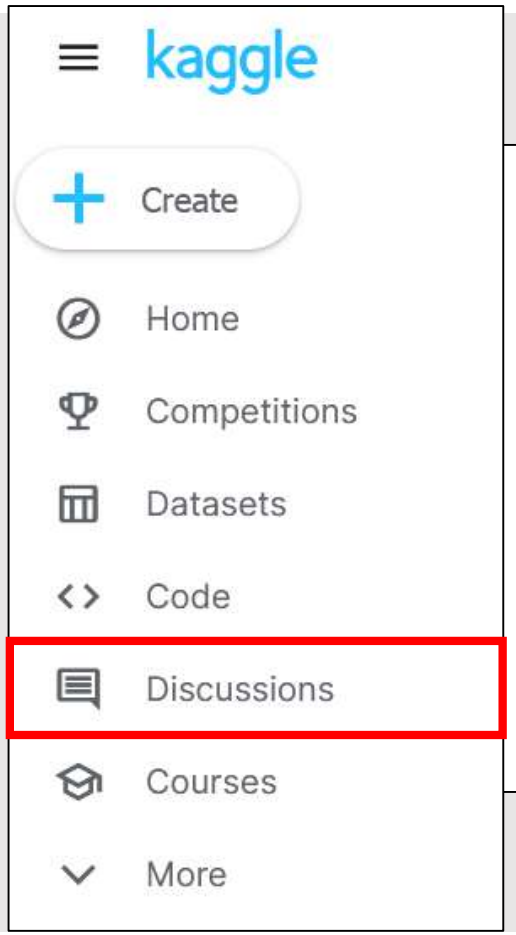


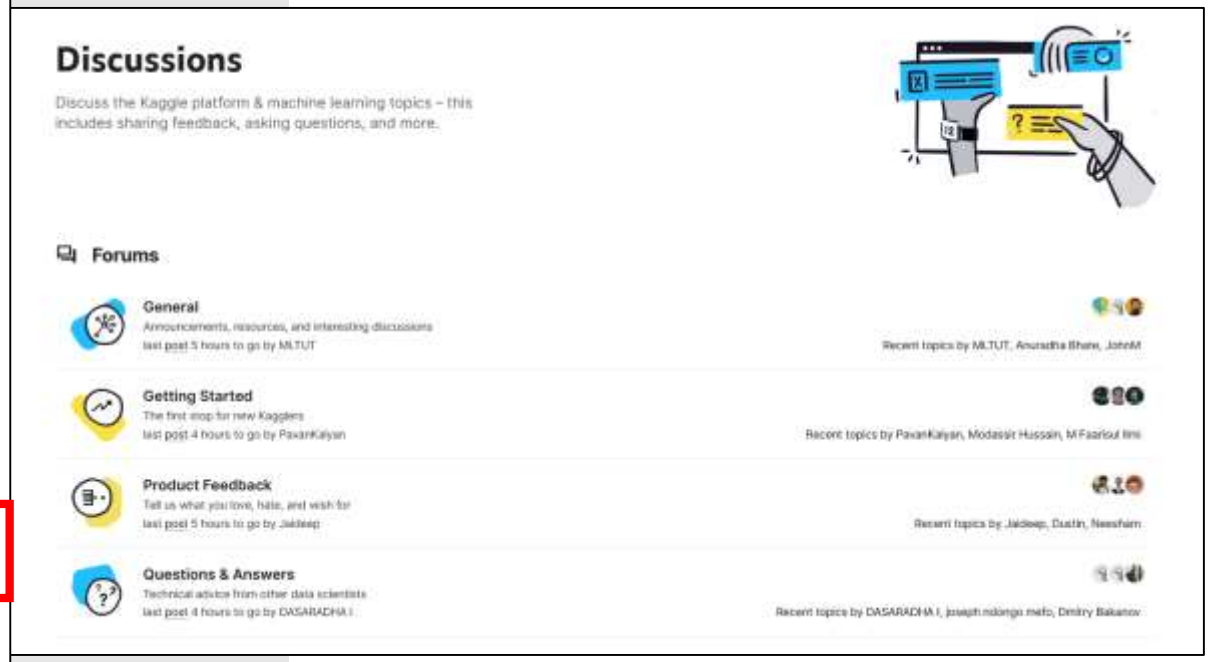
ABOUT

Explore, run code, share koding dan analisis machine learning code dengan Kaggle Notebooks (**Python & R**).

Tempat yang baik untuk **meningkatkan** skill koding dan analisis.

FEATURE 04 : DISCUSSIONS





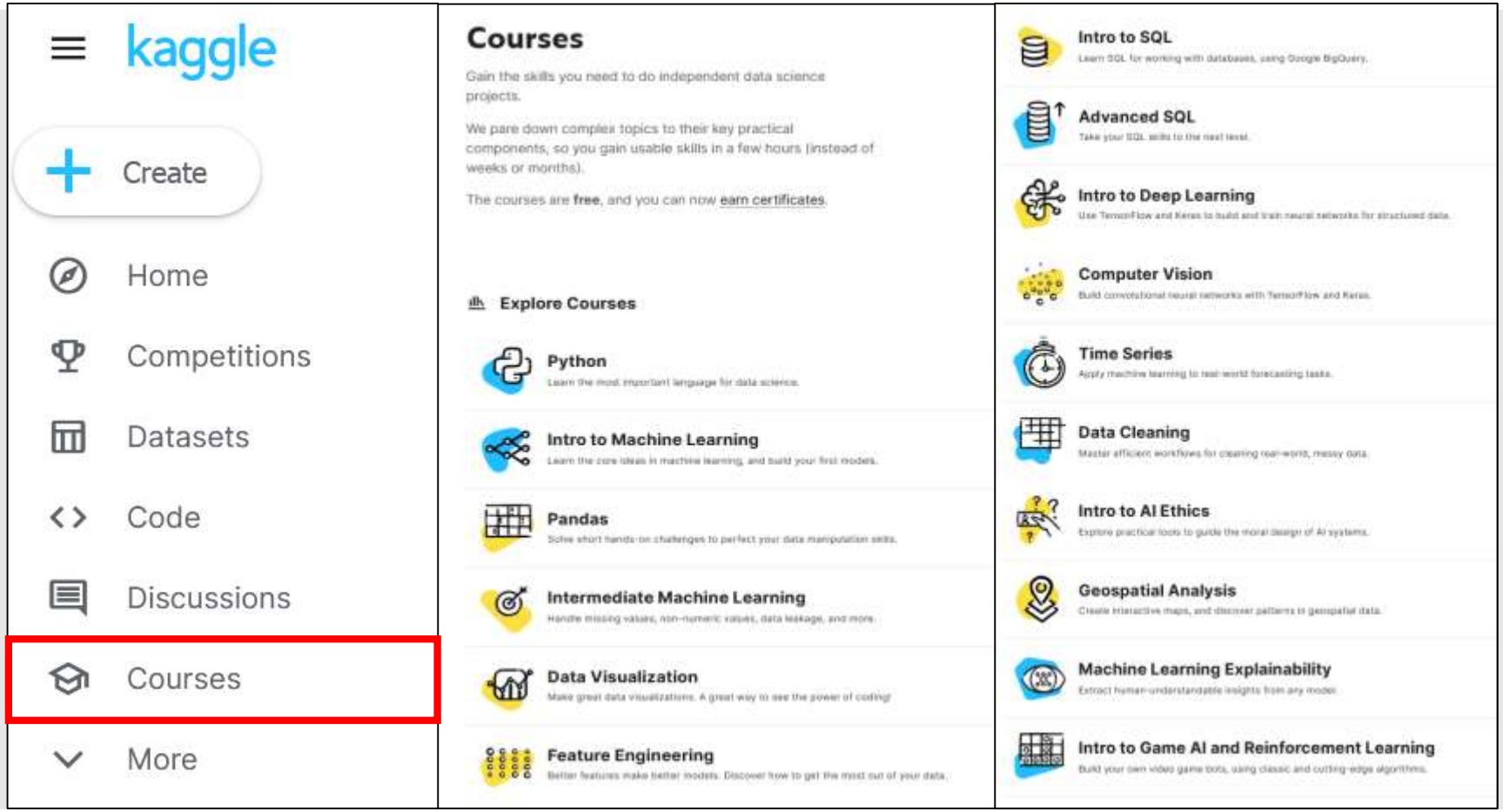
ABOUT

Mendiskusikan tentang:

- Platform Kaggle
- Machine Learning
- Data Science

Berbagi *feedback*, mengajukan pertanyaan, dll.

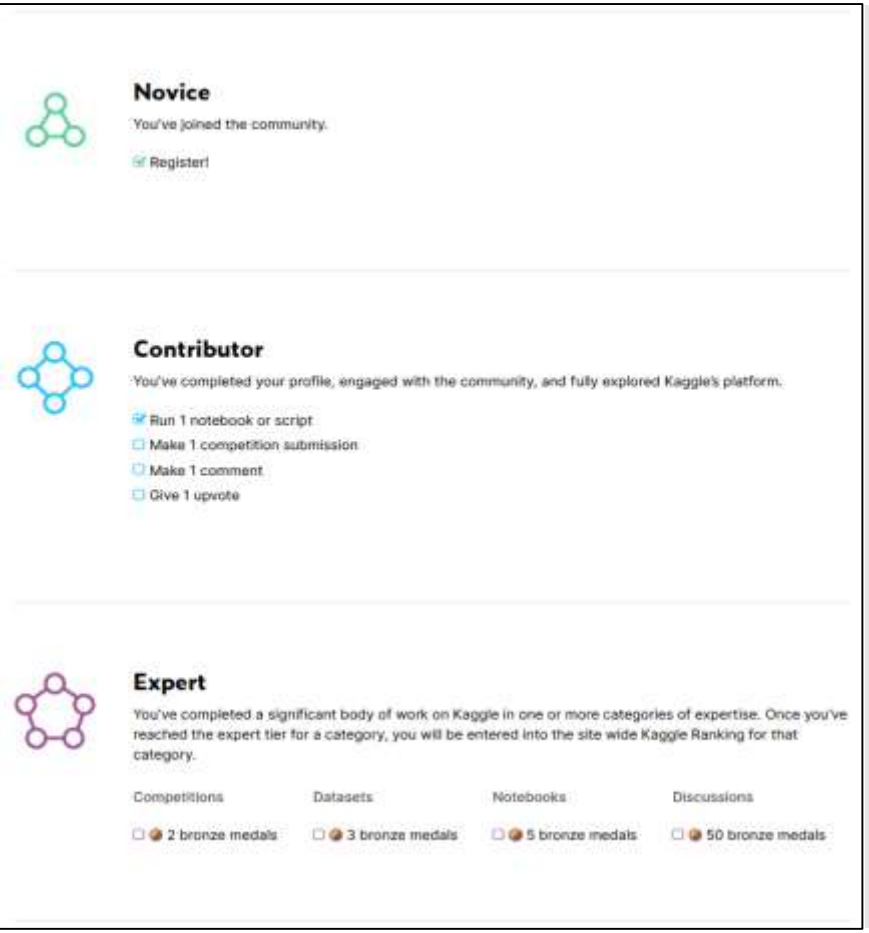
FEATURE 05 : COURSES



ABOUT

- **Mempersiapkan** kandidat data scientist dengan kursus yang terpercaya.
- **Mendapatkan keterampilan** yang dibutuhkan untuk melakukan data science project secara independen.
- **Merangkum** topik kompleks menjadi komponen praktis utama, sehingga user memperoleh keterampilan yang dapat digunakan dengan cepat.
- Coursenya **gratis**, dan user bisa mendapatkan **sertifikat**.

OTHER FEATURE : USER RANKINGS

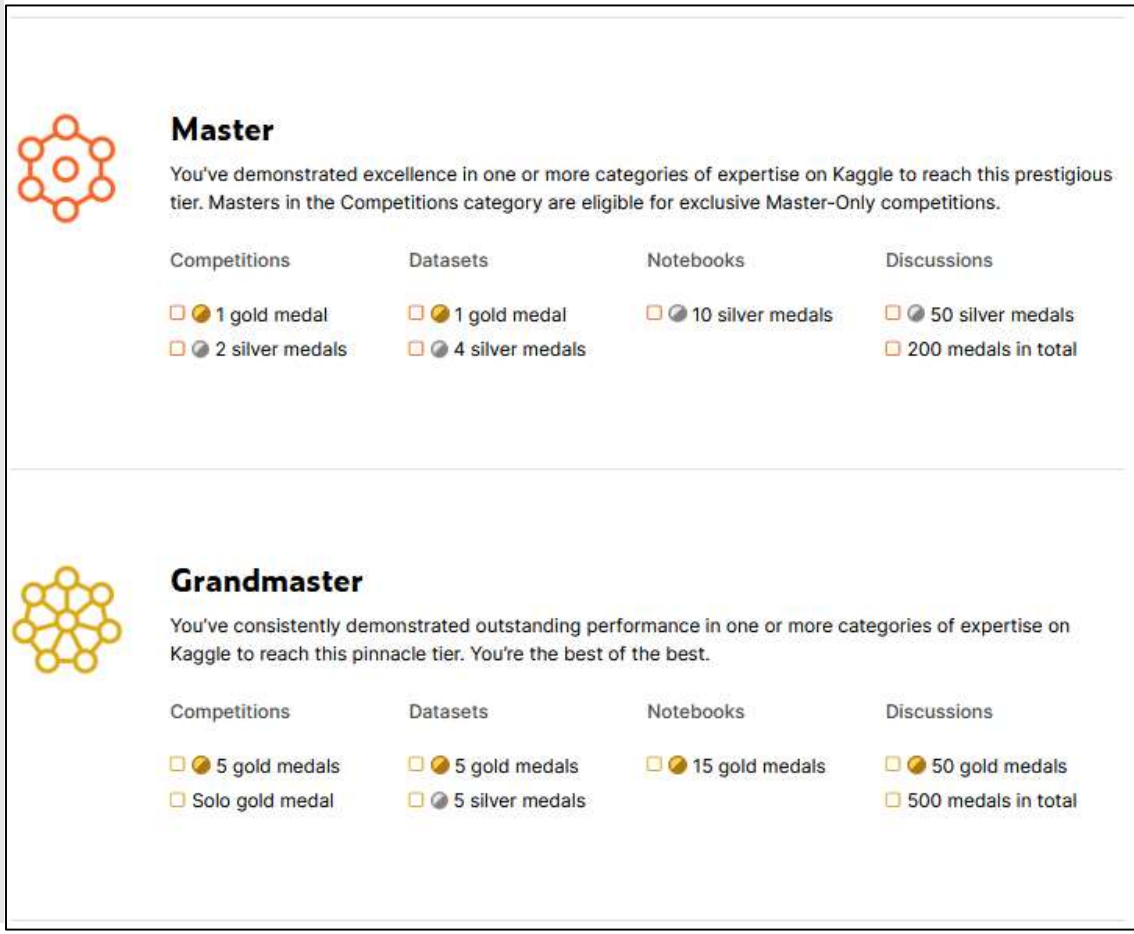


ABOUT

Kaggle's Progression System menggunakan tingkat kinerja untuk melacak pertumbuhan user sebagai ilmuwan data di Kaggle.

User akan mendapatkan medali untuk pencapaian user dan bersaing untuk kejayaan ilmu data di papan peringkat langsung.

Novice -> Contributor -> Expert -> Master -> Grandmaster



DigitalSkola

DigitalSkola

ANALYTICAL & CRITICAL THINKING

BY OMICRON (GROUP 3)

DESIGN THINKING

Sebuah Pemikiran untuk memiliki tujuan guna meningkatkan kemampuan seseorang berdasarkan pengalaman yang dimilikinya.

DigitalSkola

DESIGN THINKING FRAMEWORK

The Principles

- Mendorong bisnis dengan membantu kolega mencapai tujuannya.

The Loops

- Selalu menganggap tugas/pekerjaan sebagai sebuah prototipe.

The Keys

- Bergerak cepat dengan bekerja sama dan merangkul keberagaman.

THE PRINCIPLES

FOKUS PADA TUJUAN USER

Seberapa banyak waktu anda berfokus pada tujuan user ?

REINVENTION

Melakukan pengembangan terhadap metode baru yang mungkin dapat mempermudah proses pencapaian tujuan.

KEBERAGAMAN TIM

Menjadikan setiap perbedaan perspektif/pendapat di tim menjadi suatu kumpulan ide – ide yang dapat dimanfaatkan untuk mencapai tujuan.

THE LOOPS

Observe

- Berbicara dengan user
- Melihat mereka bekerja
- Selaraskan pemahaman ide anda dengan user

Reflect

- Bersama mengsinkronisasikan pergerakan anda
- Implementasikan hal – hal yang telah dipelajari
- Membagikan momen “aha” dengan anggota tim

Make

- Meletakkan ide mentah dan mengembangkan sembari berjalannya waktu

THE KEYS

HILLS

Mendeskripsikan sesuatu yang user dapa lakukan, seperti :

1. Siapa user anda?
2. Apa yang user dapat lakukan yang sebelumnya tidak dapat dilakukan?

PLAYBACK

Presentasi berdasar pada *story-based presentation* untuk membagikan wawasan, ide, dan perkembangan *project* ke *stakeholder*

SPONSOR USERS

Klien eksternal atau *end user* yang berkontribusi terkait domain pekerjaan

AKSI DESIGN THINKING

Design Thinking diinisiasi dengan *'business problem'*

Adopsi design thinking dan pengalamannya untuk mendapatkan hasil yang baik

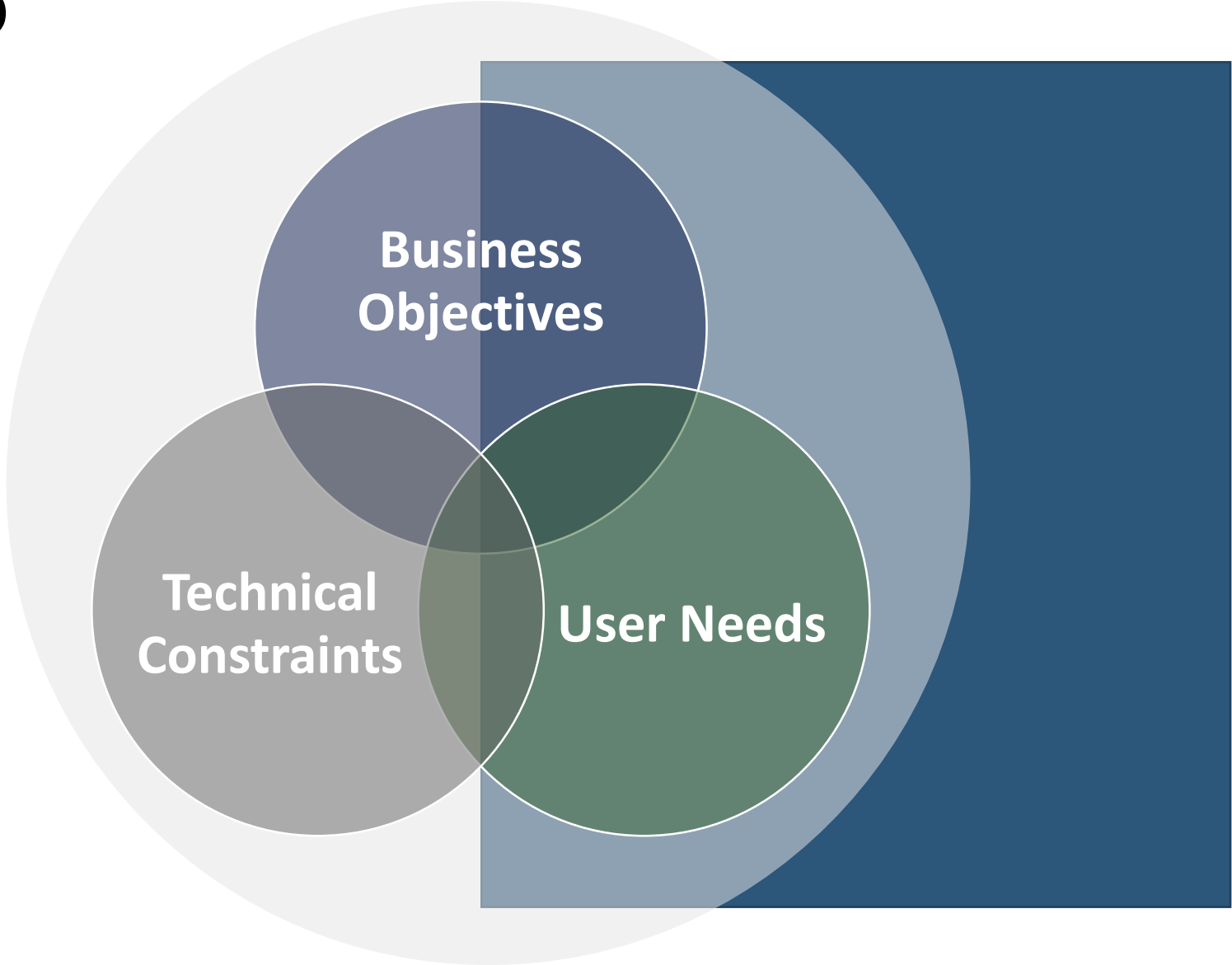
Tools dan pengetahuan perlu membicarakan terkait tantangan dan memanfaatkan kesempatan yang ada

Bekerja pada sebuah ***business problem*** dengan menyusun hal tersebut berdasarkan **pengalaman**

BUSINESS PROBLEMS TO HUMAN-CENTERED PROBLEMS

Business problem yang coba diselesaikan secara *day-to-day* difokuskan terhadap permasalahan yang dihadapi.

Dalam hal untuk fokus terhadap hal itu, maka bias dimulai dengan focus terhadap permasalahan yang dihadapi oleh ***user*** yang mendasari *business problem* yang dihadapi.



STAY CURIOUS

RASA KEINGINTAHUAN TINGGI

Rasa ingin tahu senantiasa akan memotivasi diri untuk terus mencari dan mengetahui hal-hal yang baru sehingga akan memperbanyak ilmu pengetahuan dan pengalaman dalam kegiatan belajar.

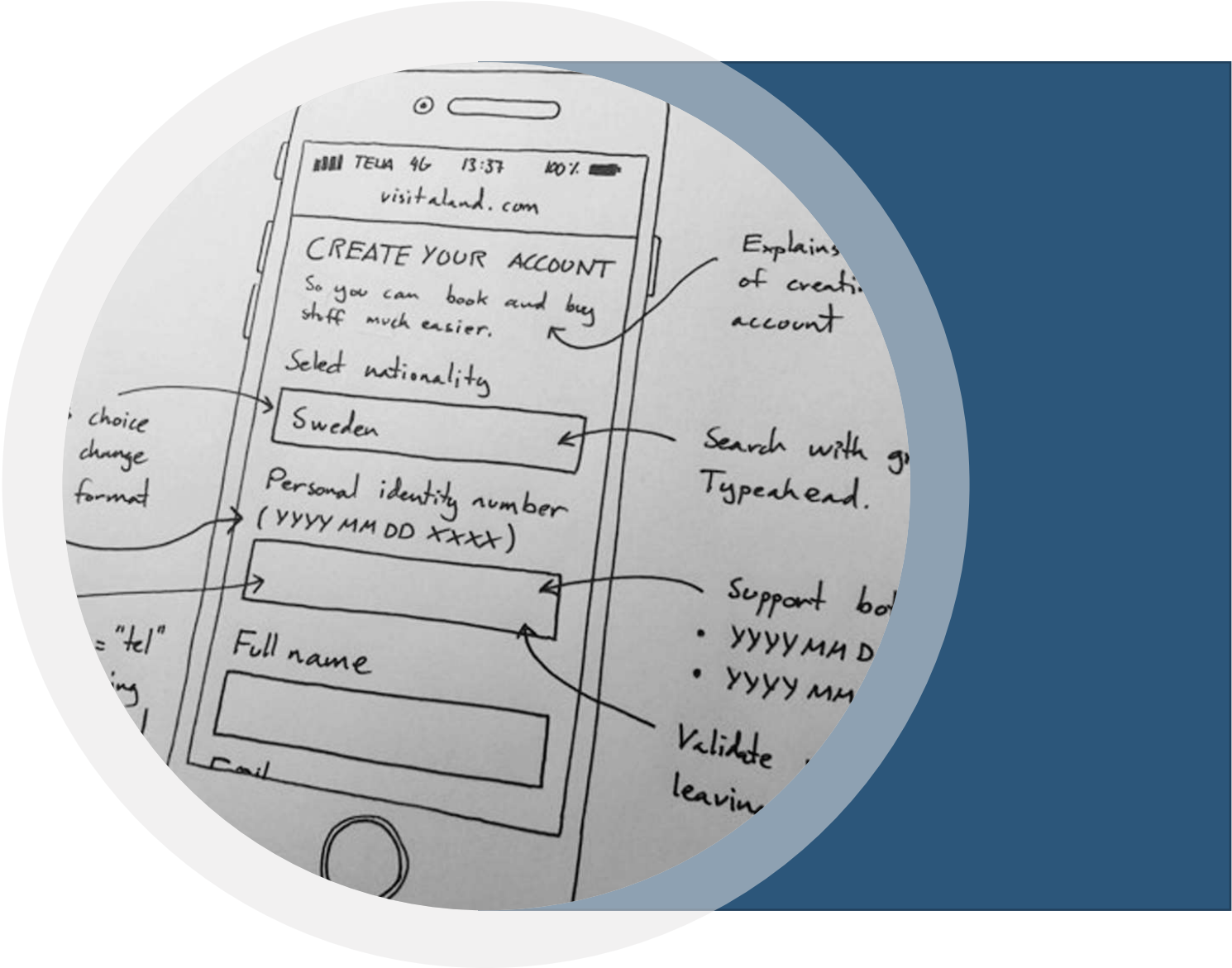
Restless Reinvention

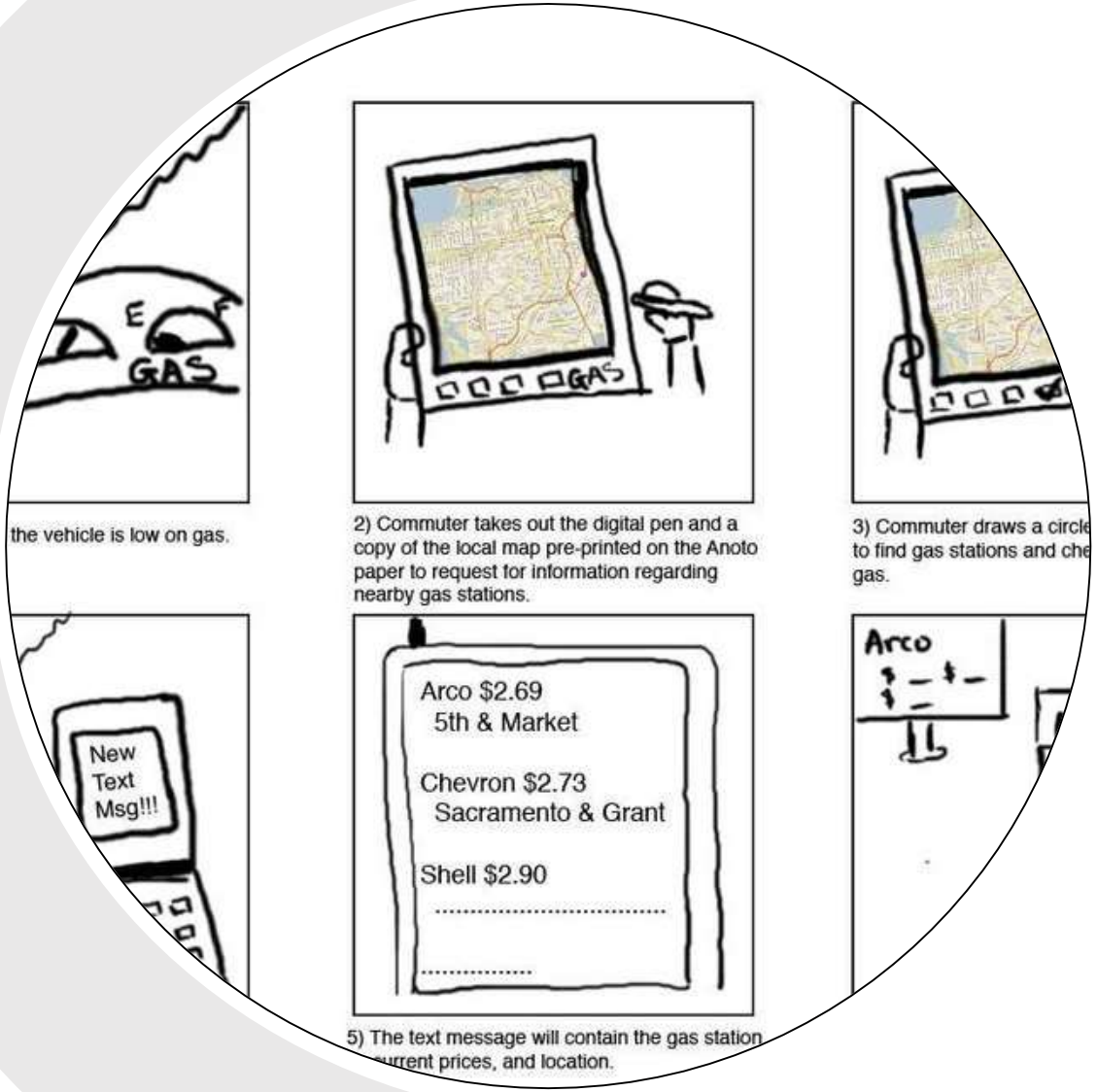
Hal ini berarti pekerjaan kita tidak akan pernah selesai

Prinsip dari design thinking yang mewakili pengujian dan pembelajaran berkelanjutan yang aktif yang bertujuan untuk untuk meningkatkan solusi

VIEW EVERYTHING AS PROTOTYPE

Dalam perancangan produk tahap prototype sangatlah penting. Karena memengaruhi keunggulan produk serta keberhasilan pengembangannya. Setelah dibuat, hasil prototype akan diuji oleh tim khusus untuk menentukan apakah sudah sesuai atau belum.





PROTOTYPE

Tidak ada solusi yang sempurna dan kita tidak akan pernah selesai

Bahkan pengguna selalu mencari pengalaman yang lebih baik untuk menyelesaikan masalah mereka.

Prototype adalah contoh awal yang paling pertama digunakan sebagai model untuk dikembangkan di masa depan

TALKING ABOUT FAILURE

Apakah kamu takut jika task yang kamu kerjakan

gagal?

Kenyataannya, kebanyakan orang sangat takut gagal sehingga

mereka tidak pernah mengambil risiko.

Akan tetapi!!! Untuk membuat sebuah pemikiran yang hebat, kita harus mulai memikirkan hal dari yang terkecil, risiko yang lebih mudah dikelola (yang menghasilkan kegagalan yang kecil)

Pikirkan cara Anda berpikir seperti seorang Scientist. Karena kamu adalah seorang…..

Data Scientist

HOW THINGS STICK

Untuk mendorong kolaborasi yang sehat di antara tim Anda yang beragam adalah melalui storytelling

Storytelling yang baik adalah alasan kenapa banyak orang menyukai cerita tentang Harry Potter di Hogwarts. Aturan yang sama berlaku untuk pekerjaan Anda

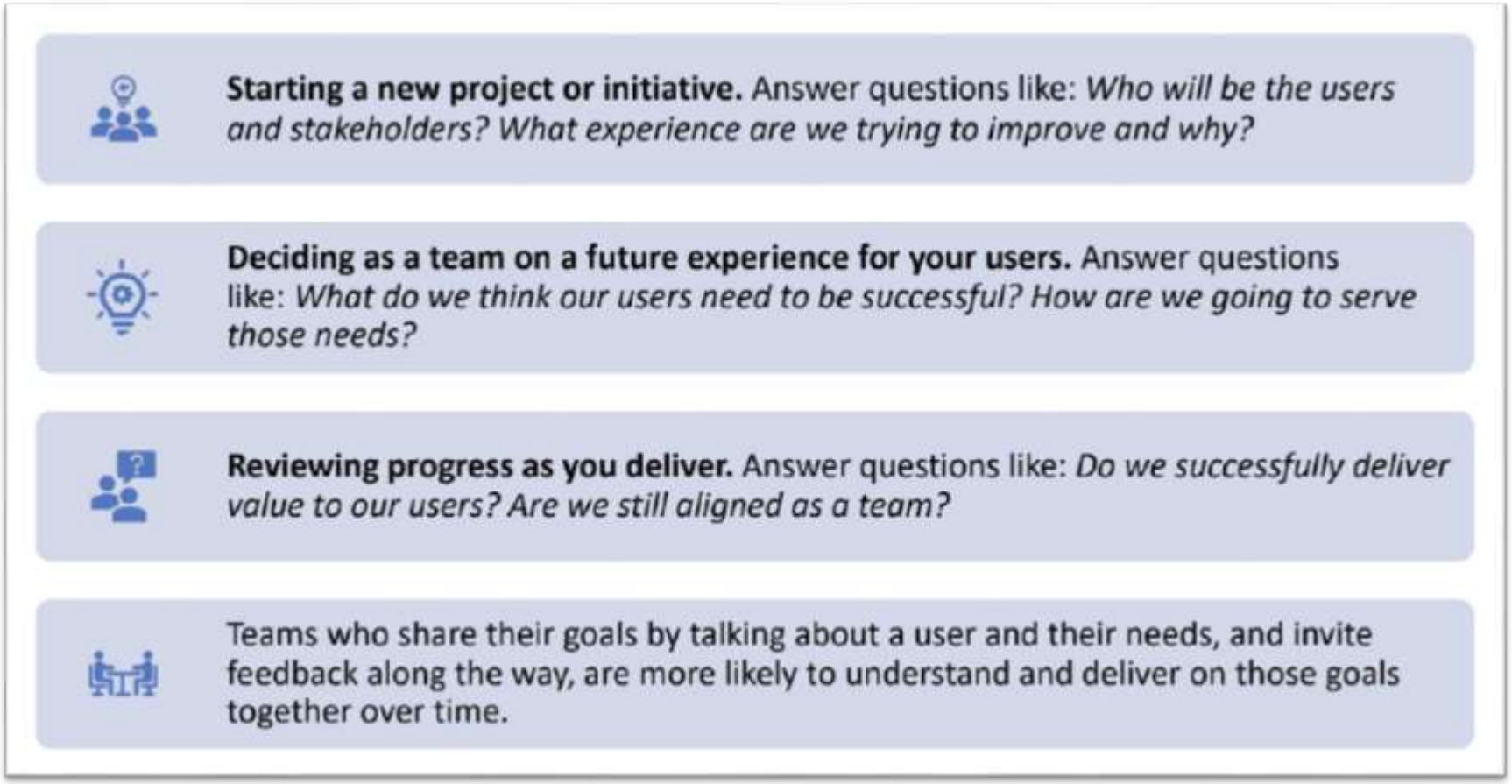
Playbacks adalah kesempatan untuk menceritakan

kisah yang berkesan dan berpusat pada manusia untuk

berbagi ide, prototipe, strategi, dan banyak lagi

Build Playbacks into your workflow

Ada beberapa momen spesifik di mana setiap orang dalam tim perlu diselaraskan:



KASUS

Tim pengadaan ingin tahu pengeluaran dari barang-barang yang dibelanjakan setiap bulannya.
Tim tersebut membutuhkan pertolongan dari tim data

Apakah ada dokumen?

dokumen membantu data scientist/data analyst untuk tahu situasi sekarang dari tim pengadaan

Meminta user membuat rangkuman (ppt, doc)

ppt/doc membantu baik tim data dan tim pengadaan untuk melihat spesifik kebutuhannya

Laporkan hasilnya

mengemas seluruh hasil ke dalam ppt dan mengingatkan “pain point” telah dikelola dengan baik

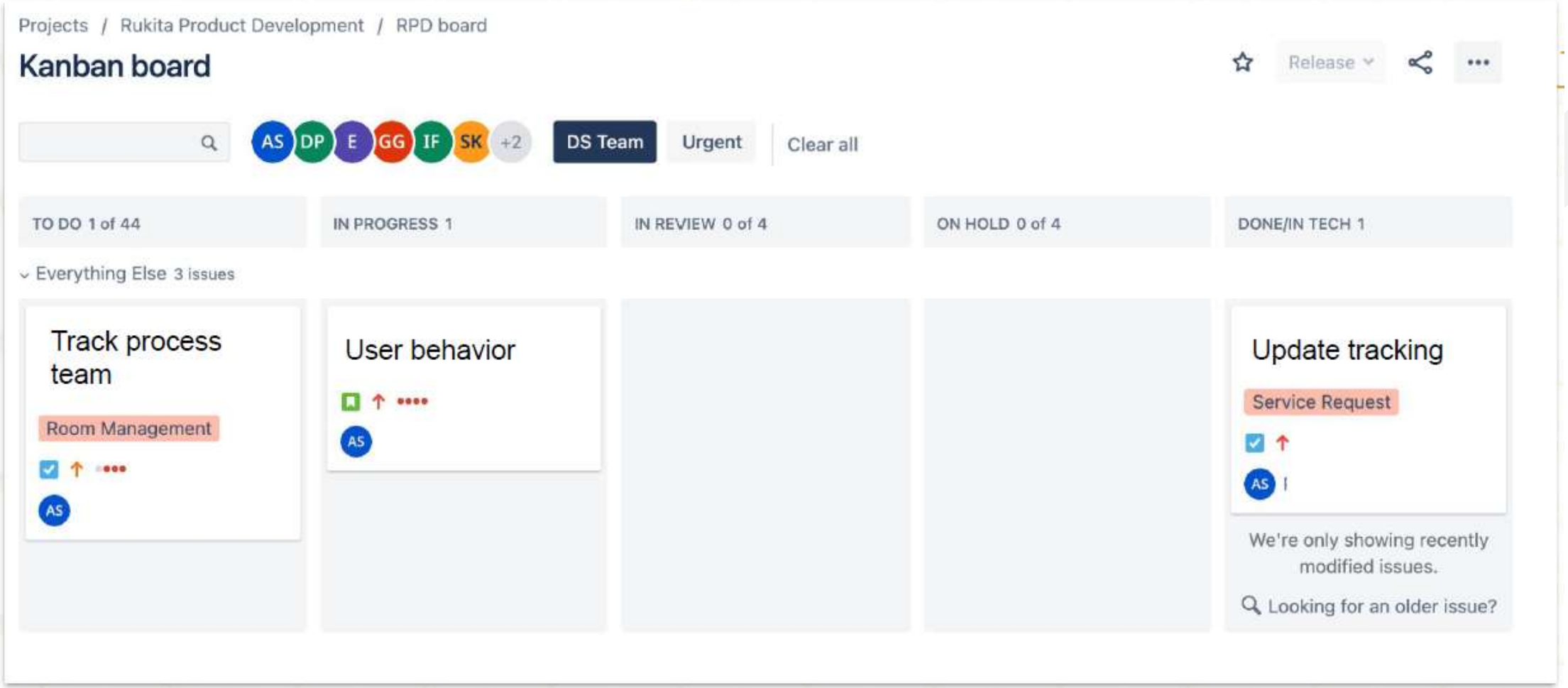
Meeting

Di dalam meeting, lontarkan pertanyaan yang telah dibuat

Membuat pertanyaan

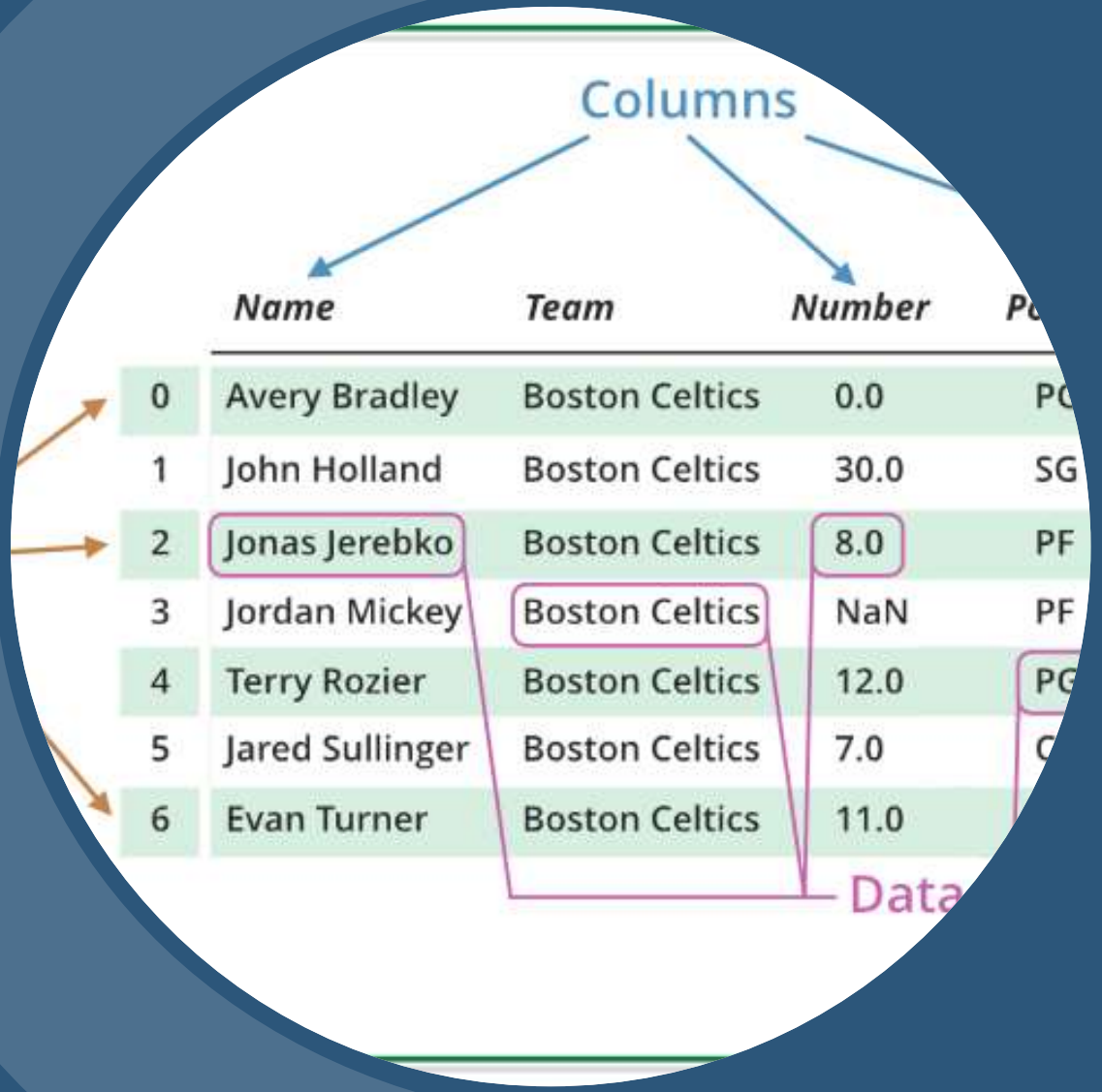
Membuat pertanyaan-pertanyaan seperti hipotesa singkat untuk meyakinkan bahwa permasalahannya sesuai dengan yang dihadapi

DATA SCIENCE AGILE SYSTEM



DigitalSkola

DigitalSkola



PANDAS INTERMEDIATE DATAFRAME

BY OMICRON (GROUP 3)

DigitalSkola

Sorting Data Frame

PART 1: Alphabetic Sorting

PART 2: Numeric Sorting

PART 3: 2 Columns Sorting

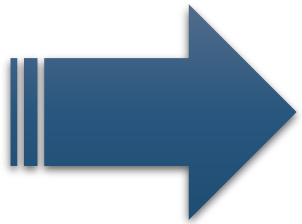
PART 4: Resetting Index

SORTING DATA FRAME

1. Alphabetic Sorting — Sorting data frame alphabetically (*A - Z*) using column with ‘object’ data type.

```
dataframe.sort_values(by = ['department'])
```

index	Unnamed: 0	employee_id	department	region	education
0	33498	32835	Operations	region_10	Bachelor's
1	31110	72340	Operations	region_2	Below Secondary
2	47412	52336	Sales & Marketing	region_31	Bachelor's
3	9900	75679	Finance	region_26	Bachelor's
4	39275	72476	Operations	region_27	Bachelor's
5	48061	52430	Technology	region_13	Bachelor's
6	37126	14998	Technology	region_2	Bachelor's
7	5069	72210	HR	region_2	Bachelor's
8	45239	20938	Finance	region_13	NaN
9	39767	30660	Analytics	region_22	Master's & above



index	Unnamed: 0	employee_id	department	region	education
9	39767	30660	Analytics	region_22	Master's & above
3	9900	75679	Finance	region_26	Bachelor's
8	45239	20938	Finance	region_13	NaN
7	5069	72210	HR	region_2	Bachelor's
0	33498	32835	Operations	region_10	Bachelor's
1	31110	72340	Operations	region_2	Below Secondary
4	39275	72476	Operations	region_27	Bachelor's
2	47412	52336	Sales & Marketing	region_31	Bachelor's
5	48061	52430	Technology	region_13	Bachelor's
6	37126	14998	Technology	region_2	Bachelor's

SORTING DATA FRAME

2. Numeric Sorting

Sorting dataframe numerically ascending (from less to higher value) using column with ‘integer’ datatype.

```
dataframe.sort_values(by = ['age'])
```

index	employee_id	department	region	gender	age
0	32835	Operations	region_10	f	30
1	72340	Operations	region_2	m	23
2	52336	Sales & Marketing	region_31	f	34
3	75679	Finance	region_26	f	25
4	72476	Operations	region_27	m	33
5	52430	Technology	region_13	m	34
6	14998	Technology	region_2	f	28
7	72210	HR	region_2	f	35
8	20938	Finance	region_13	m	29
9	30660	Analytics	region_22	m	38

index	employee_id	department	region	gender	age
1	72340	Operations	region_2	m	23
3	75679	Finance	region_26	f	25
6	14998	Technology	region_2	f	28
8	20938	Finance	region_13	m	29
0	32835	Operations	region_10	f	30
4	72476	Operations	region_27	m	33
2	52336	Sales & Marketing	region_31	f	34
5	52430	Technology	region_13	m	34
7	72210	HR	region_2	f	35
9	30660	Analytics	region_22	m	38

SORTING DATA FRAME

APPLYING DESCENDING SORT

Sorting data frame from higher to less value or (Z - A) for 'object'

```
dataframe.sort_values(by = ['age'], ascending = False)
```

index	employee_id	department	region	gender	age
1	72340	Operations	region_2	m	23
3	75679	Finance	region_26	f	25
6	14998	Technology	region_2	f	28
8	20938	Finance	region_13	m	29
0	32835	Operations	region_10	f	30
4	72476	Operations	region_27	m	33
2	52336	Sales & Marketing	region_31	f	34
5	52430	Technology	region_13	m	34
7	72210	HR	region_2	f	35
9	30660	Analytics	region_22	m	38

index	employee_id	department	region	gender	age
9	30660	Analytics	region_22	m	38
7	72210	HR	region_2	f	35
2	52336	Sales & Marketing	region_31	f	34
5	52430	Technology	region_13	m	34
4	72476	Operations	region_27	m	33
0	32835	Operations	region_10	f	30
8	20938	Finance	region_13	m	29
6	14998	Technology	region_2	f	28
3	75679	Finance	region_26	f	25
1	72340	Operations	region_2	m	23

DigitalSkola

SORTING DATA FRAME

3. 2 Columns Sorting

Sorting data frame using 2 columns variable (double sorting)

```
dataframe.sort_values(by = ['gender', 'age'])
```

index	employee_id	department	region	gender	age
3	75679	Finance	region_26	f	25
6	14998	Technology	region_2	f	28
0	32835	Operations	region_10	f	30
2	52336	Sales & Marketing	region_31	f	34
7	72210	HR	region_2	f	35
1	72340	Operations	region_2	m	23
8	20938	Finance	region_13	m	29
4	72476	Operations	region_27	m	33
5	52430	Technology	region_13	m	34
9	30660	Analytics	region_22	m	38

SORTING DATA FRAME

4. Resetting Index

Used to rearrange the index after sorting process, so the index column still look neat.

```
dataframe.sort_values(by = ['gender', 'age'], ascending = False).reset_index(drop = True)
```

index	employee_id	department	region	gender	age
0	30660	Analytics	region_22	m	38
1	52430	Technology	region_13	m	34
2	72476	Operations	region_27	m	33
3	20938	Finance	region_13	m	29
4	72340	Operations	region_2	m	23
5	72210	HR	region_2	f	35
6	52336	Sales & Marketing	region_31	f	34
7	32835	Operations	region_10	f	30
8	14998	Technology	region_2	f	28
9	75679	Finance	region_26	f	25

used to drop the original index that disordered after the sorting process

DigitalSkola

Filtering Data Frame

PART 1: Filtering Several Columns

PART 2: Using ***loc*** and ***iloc*** Function

FILTERING DATA FRAME

1. Filtering Several Columns

Used to retrieve data points from a specific column or row. There are several ways to filtering several columns.

index	employee_id	department	region	education	gender	recruitment_channel	no_of_trainings	age
0	59183	Operations	region_27	Bachelor's	m	sourcing	1	45
1	47281	Sales & Marketing	region_4	Master's & above	m	sourcing	1	57
2	60776	Operations	region_19	Master's & above	m	sourcing	1	37
3	28476	Technology	region_31	Master's & above	m	sourcing	1	46
4	56627	Sales & Marketing	region_7	Master's & above	f	other	1	34
5	22933	Analytics	region_26	Bachelor's	m	sourcing	2	29
6	39431	Operations	region_13	Bachelor's	m	other	1	28
7	36147	Operations	region_15	Master's & above	f	other	1	40
8	26502	Procurement	region_2	Bachelor's	m	sourcing	2	29
9	61121	Analytics	region_22	Bachelor's	m	sourcing	2	27

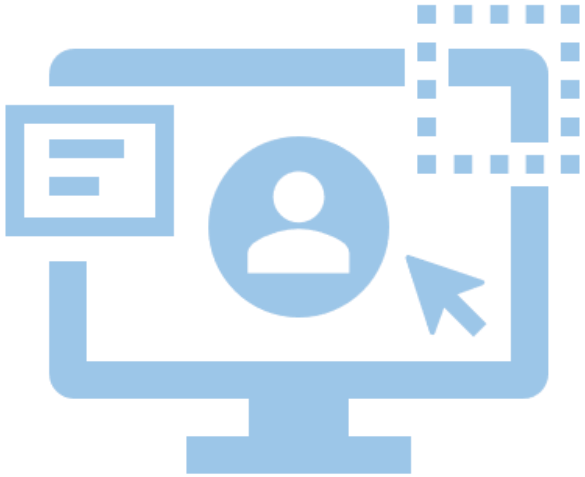
```
data_baru.filter(items = ['employee_id', 'department'])
```

```
data_baru[['employee_id', 'department']]
```

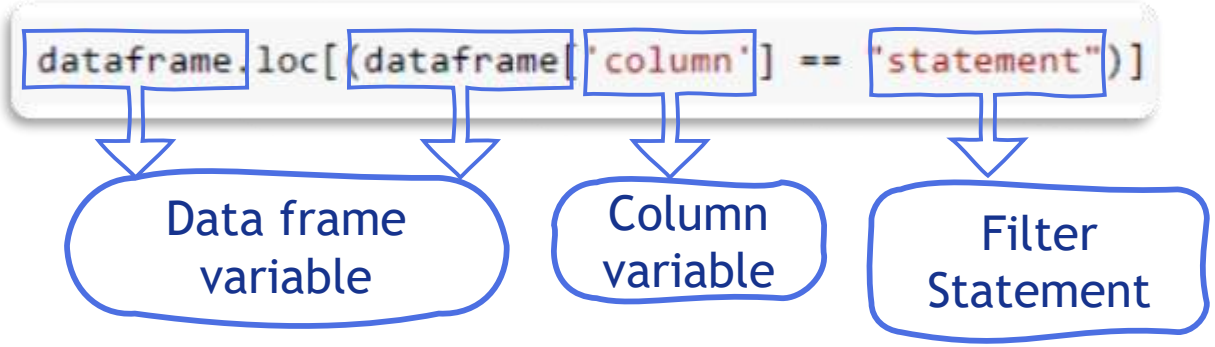
	employee_id	department
0	59183	Operations
1	47281	Sales & Marketing
2	60776	Operations
3	28476	Technology
4	56627	Sales & Marketing
5	22933	Analytics
6	39431	Operations
7	36147	Operations
8	26502	Procurement
9	61121	Analytics

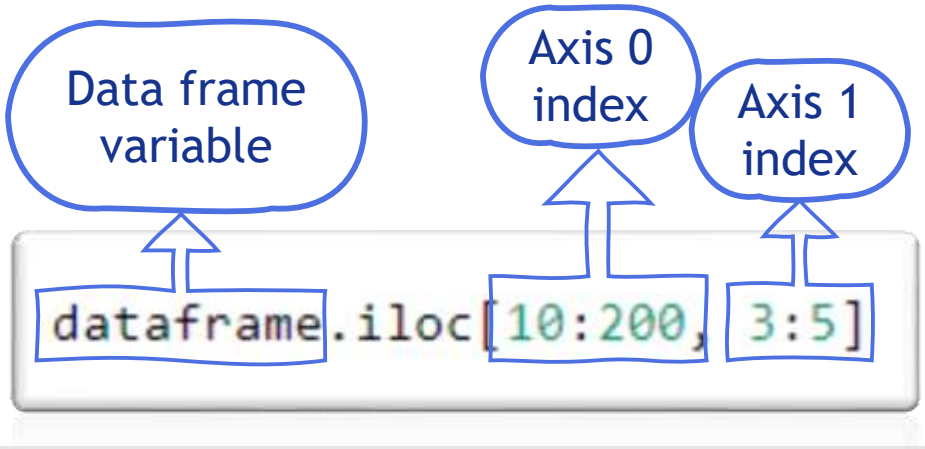
FILTERING DATA FRAME

2. Using ***loc*** and ***iloc*** Function



to filtering in the name of variable **loc**





iloc to filtering using index

DigitalSkola

FILTERING DATA FRAME

Using *loc*

index	employee_id	department	region	education	gender	recruitment_channel	no_of_trainings	age
0	59183	Operations	region_27	Bachelor's	m	sourcing	1	45
1	47281	Sales & Marketing	region_4	Master's & above	m	sourcing	1	57
2	60776	Operations	region_19	Master's & above	m	sourcing	1	37
3	28476	Technology	region_31	Master's & above	m	sourcing	1	46
4	56627	Sales & Marketing	region_7	Master's & above	f	other	1	34
5	22933	Analytics	region_26	Bachelor's	m	sourcing	2	29
6	39431	Operations	region_13	Bachelor's	m	other	1	28
7	36147	Operations	region_15	Master's & above	f	other	1	40
8	26502	Procurement	region_2	Bachelor's	m	sourcing	2	29
9	61121	Analytics	region_22	Bachelor's	m	sourcing	2	27

```
dataframe.loc[3:5,['employee_id', 'gender']]
```

	employee_id	gender
3	28476	m
4	56627	f
5	22933	m

DigitalSkola

FILTERING DATA FRAME

Using *iloc*

index	employee_id	department	region	education	gender	recruitment_channel	no_of_trainings	age
0	59183	Operations	region_27	Bachelor's	m	sourcing	1	45
1	47281	Sales & Marketing	region_4	Master's & above	m	sourcing	1	57
2	60776	Operations	region_19	Master's & above	m	sourcing	1	37
3	28476	Technology	region_31	Master's & above	m	sourcing	1	46
4	56627	Sales & Marketing	region_7	Master's & above	f	other	1	34
5	22933	Analytics	region_26	Bachelor's	m	sourcing	2	29
6	39431	Operations	region_13	Bachelor's	m	other	1	28
7	36147	Operations	region_15	Master's & above	f	other	1	40
8	26502	Procurement	region_2	Bachelor's	m	sourcing	2	29
9	61121	Analytics	region_22	Bachelor's	m	sourcing	2	27

```
dataframe.iloc[-10:, 3:6]
```

	education	gender	recruitment_channel
10	Bachelor's	m	sourcing
11	Master's & above	m	other
12	Bachelor's	m	other
13	Master's & above	f	other
14	Master's & above	f	sourcing
15	Bachelor's	m	other
16	Bachelor's	f	other
17	Bachelor's	m	other
18	Master's & above	m	sourcing
19	Bachelor's	m	sourcing

DigitalSkola

THANK YOU!

BY OMICRON (GROUP 3)