Kaggle

(Data Science Community)

DigitalSkola
Uncover The World Of Digital Skills With Us

Table of Content

What will We Learn Today?

1. Kaggle dan tujuannya
2. Sejarah Kaggle
3. Fitur-fitur Kaggle
4. Mengembangkan skill data melalui Kaggle
5. Kompetisi di Kaggle
6. Eksekusi Ngoding di Kaggle

Profile

Professional

- Senior Data Analyst – Kompas (2021 – Present)
- Data Scientist – Rukita (2020 – 2021)
- Research Assistant Analyst – Ensterna (2017 – 2019)

Educational Background

- Nuclear Engineering – Universitas Gadjah Mada

Connect with me

- https://dataimpact.medium.com/
- https://www.linkedin.com/in/ariprabowo/
- https://github.com/densaiko

Ari Sulistiyo Prabowo

Kaggle

Kaggle adalah sebuah platform untuk berinteraksi, melatih dan menyelesaikan tantangan terkait data science dan machine learning

What is the purpose of Kaggle?

- User dapat menemukan dataset yang sifatnya umum
- Eksplorasi dan membangun model di dalam tampilan web
- Dapat berkoneksi dengan data scientist lain
- Mengikuti kompetisi untuk menyelsaikan suatu permasalahan

History of Kaggle

- Ditemukan pada **April 2010** oleh **Anthony Goldbloom** dan **Jeremy Howard**
- Pada **8 Maret 2017**, Kaggle diakuisisi oleh **Google**
- Pada bulan **Juni 2017**, Kaggle mengumumkan bahwa mereka telah melampaui 1 juta user dan komunitas yang tersebar di 194 negara

Features in Kaggle

Kaggle merupakan platform hebat yang memiliki berbagai fitur untuk melatih diri sendiri. Beberapa fitur di antaranya adalah:

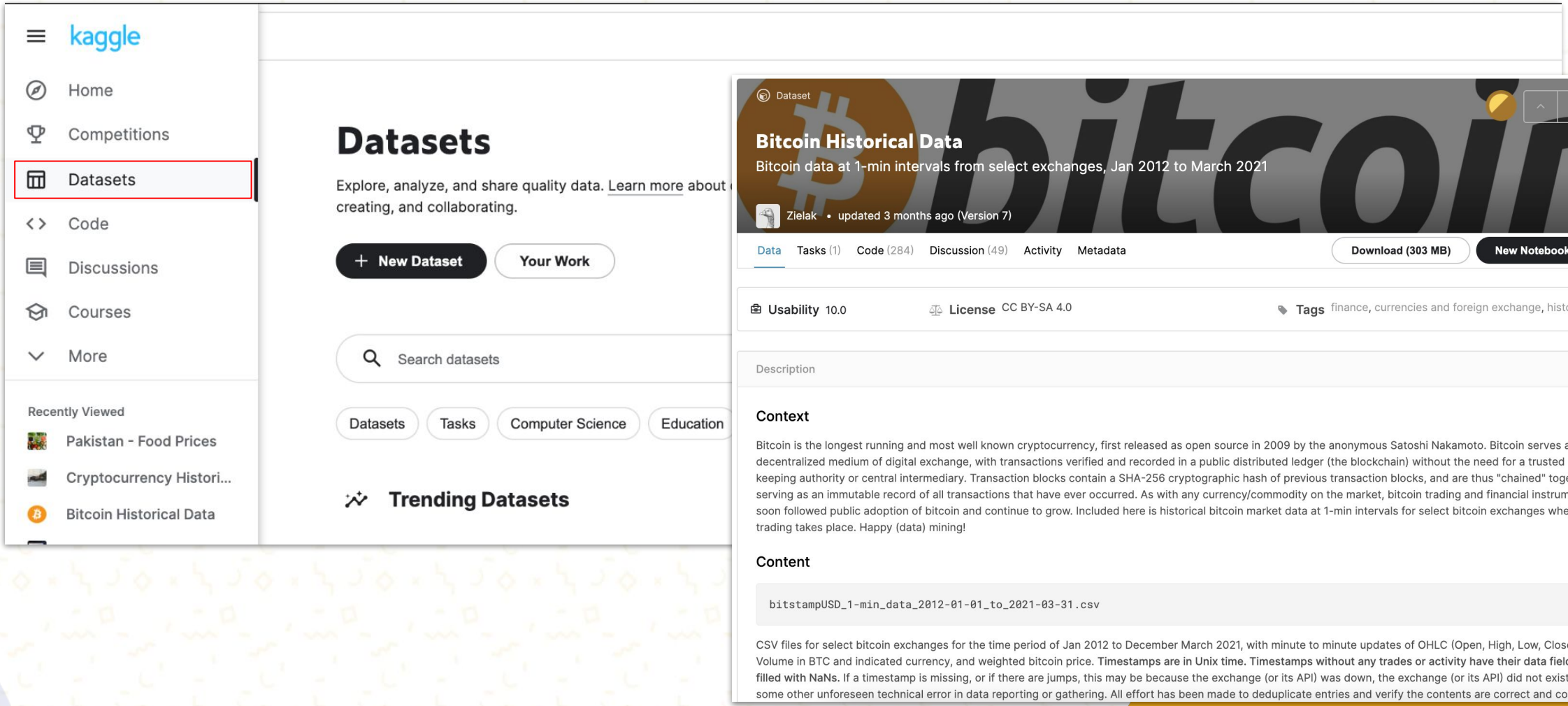
- **Menemukan atau memuat dataset**
- **Eksplore codingan dari data scientist lain**
- **Berkolaborasi dari berbagai negara**
- **Kursus belajar mandiri**
- **Mengikuti kompetisi**

Let’s Create an Account

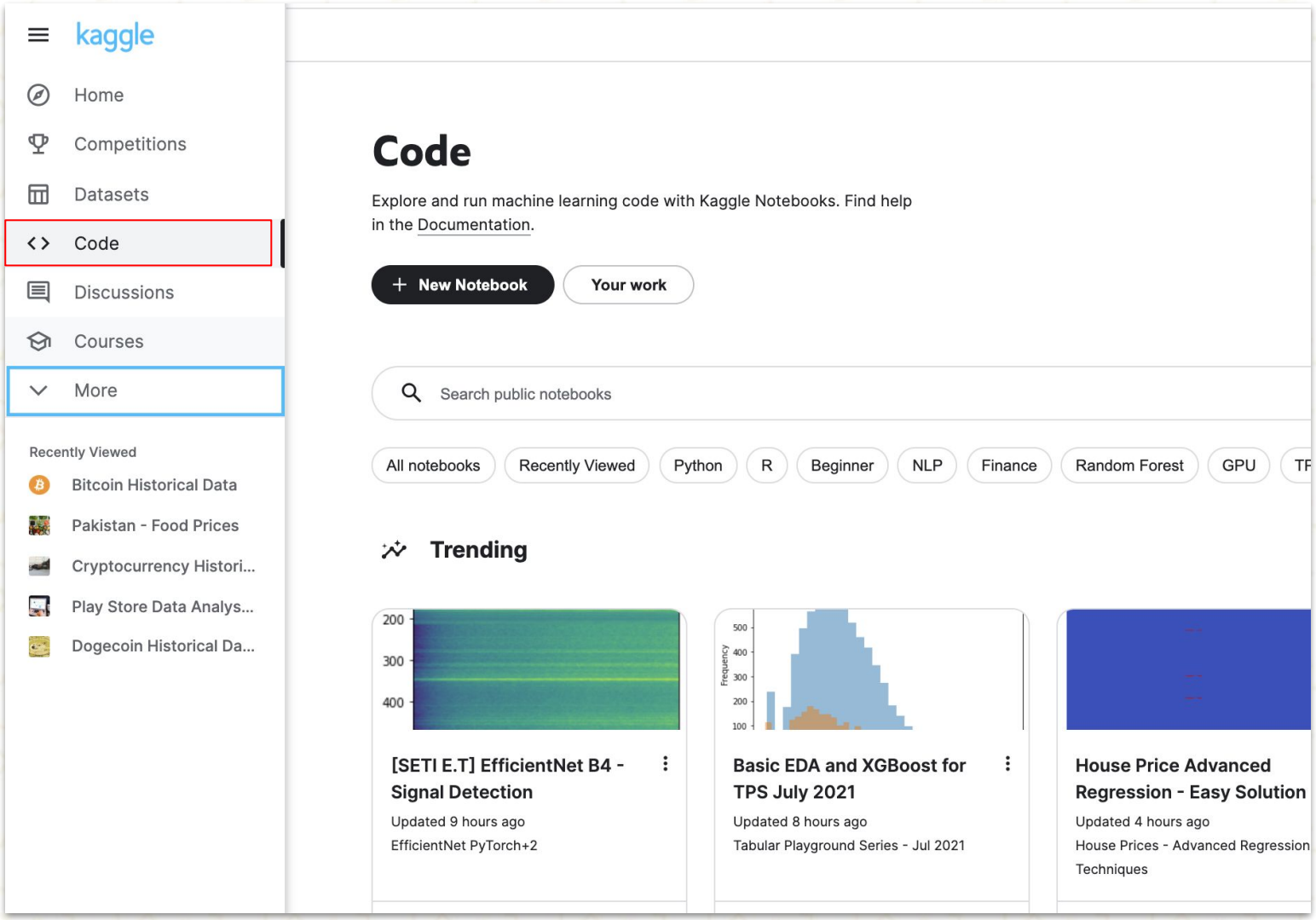
https://www.kaggle.com/account/login

DigitalSkola
Uncover The World Of Digital Skills With Us

Fitur 1: Menemukan dan memuat dataset



Fitur 2: Eksplorasi codingan



- Mengizinkan data scientist untuk share koding dan analisis menggunakan Python dan R
- Tempat yang baik untuk meningkatkan skill koding dan analisis

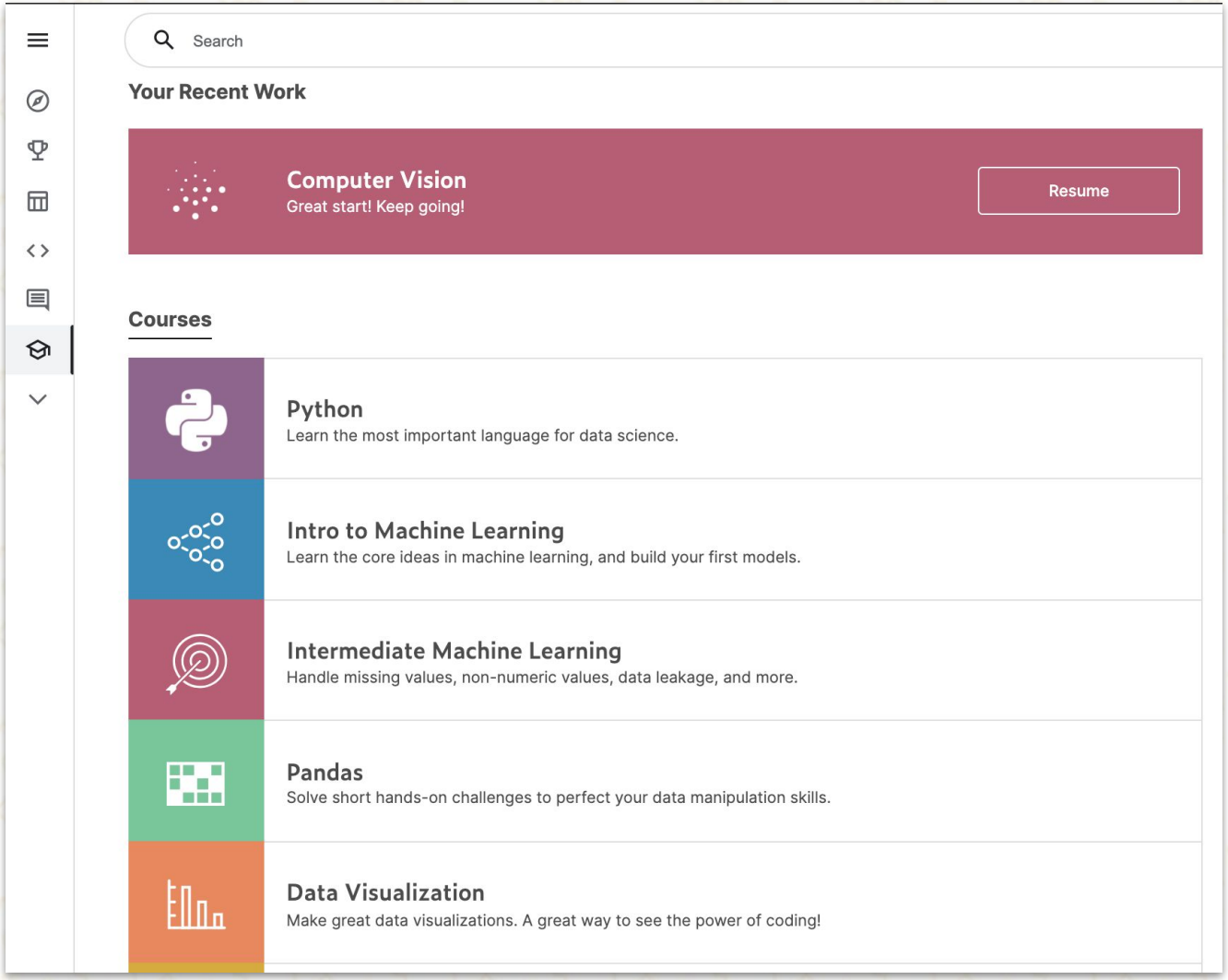
DigitalSkola
Uncover The World Of Digital Skills With Us

Fitur 3: Berkolaborasi

#	Team Name	Notebook	Team Members	Score	Entries	Last
1	(✧∀✧)/ 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

DigitalSkola
Uncover The World Of Digital Skills With Us

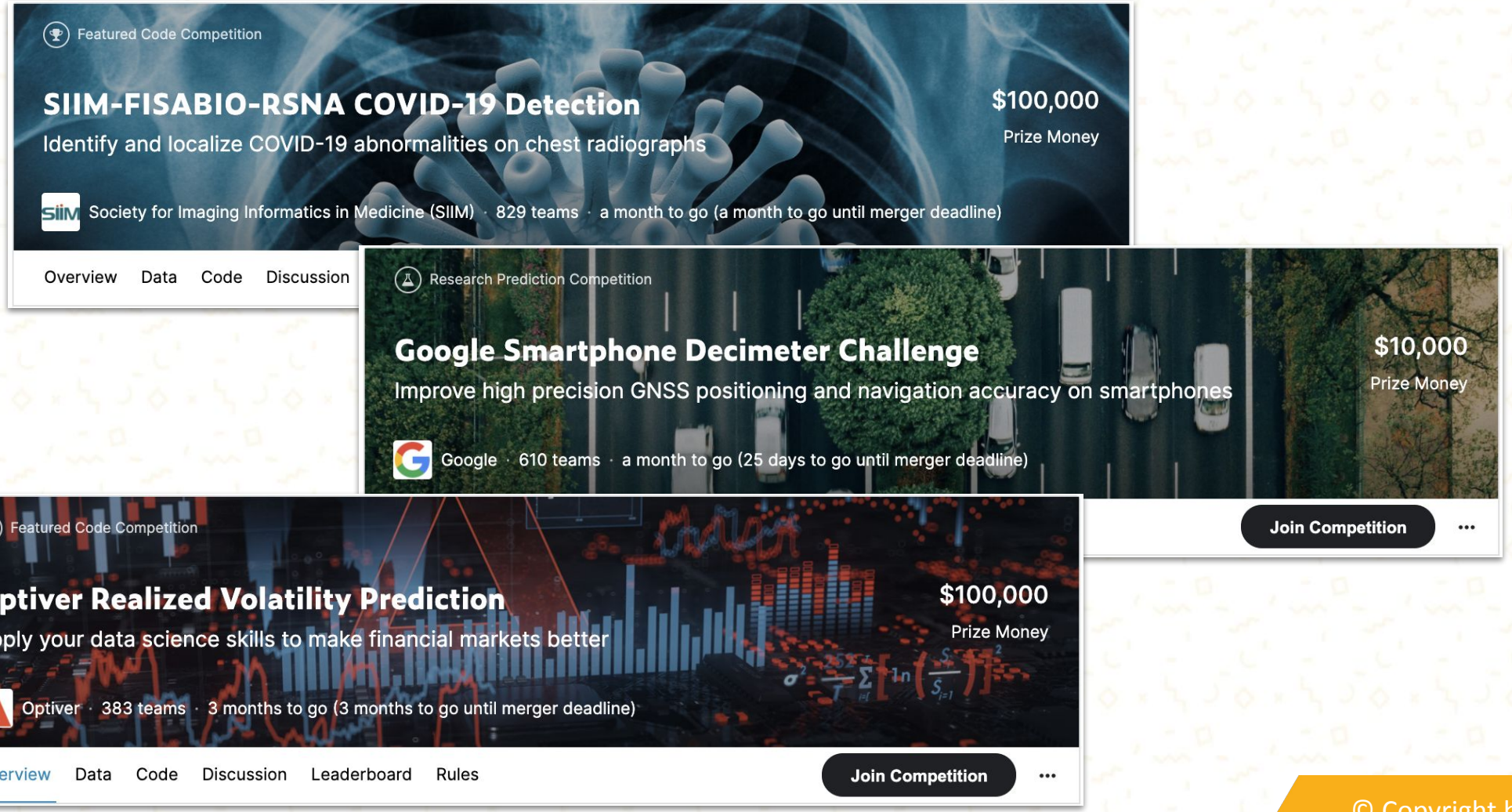
Fitur 4: Kursus belajar gratis



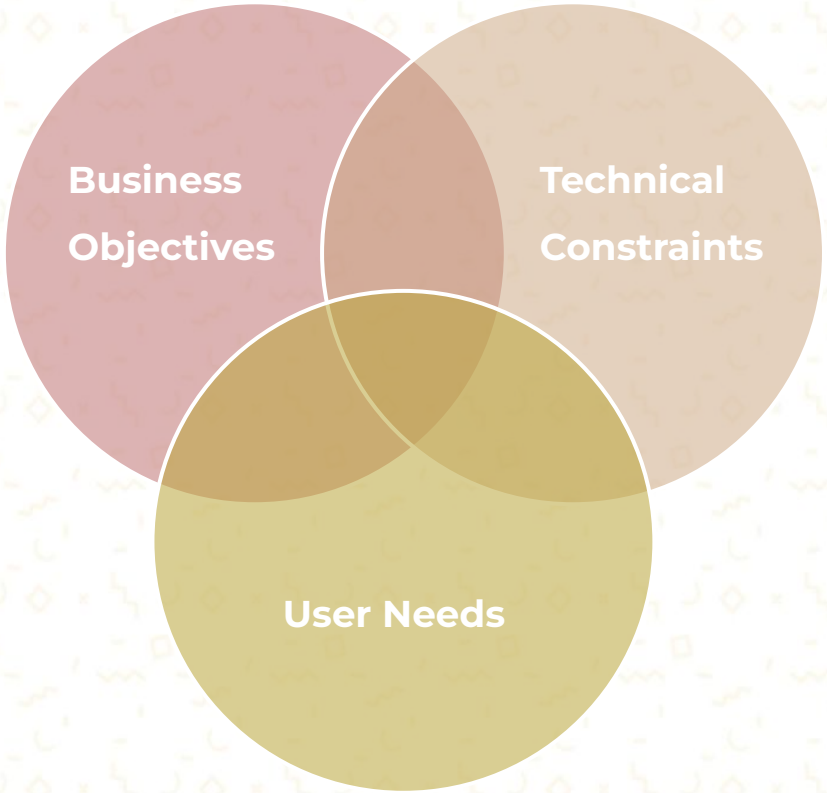
- Mempersiapkan kandidat data scientist dengan kursus yang terpercaya
- Mendapatkan sertifikat jika kamu menyelesaikan kursusnya

DigitalSkola
Uncover The World Of Digital Skills With Us

Fitur 5: Mengikuti Kompetisi



Manfaat mengikuti kompetisi



Problem Solving:

- Mempertajam skill untuk menyelesaikan suatu permasalahan dengan (analytical thinking)

Domain Bisnis:

- Menyelesaikan masalah dengan domain bisnis tertentu akan memperkaya wawasan

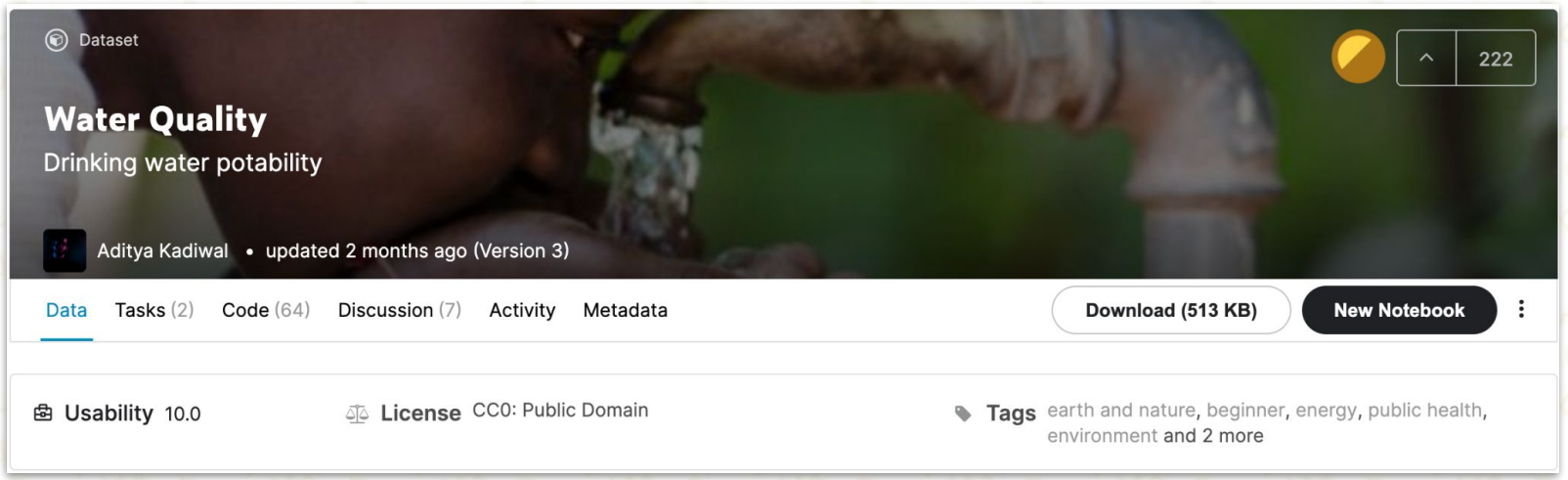
Users:

- Menjadikan kompetisi tersebut sebagai bukti yang dapat ditampilkan di portofolio

Let’s Practice

https://www.kaggle.com/adityakadiwal/water-potability

Water Potability Dataset



Baca terlebih dahulu tugasnya:

- Pahami dulu permasalahannya seperti apa dan data-nya bagaimana

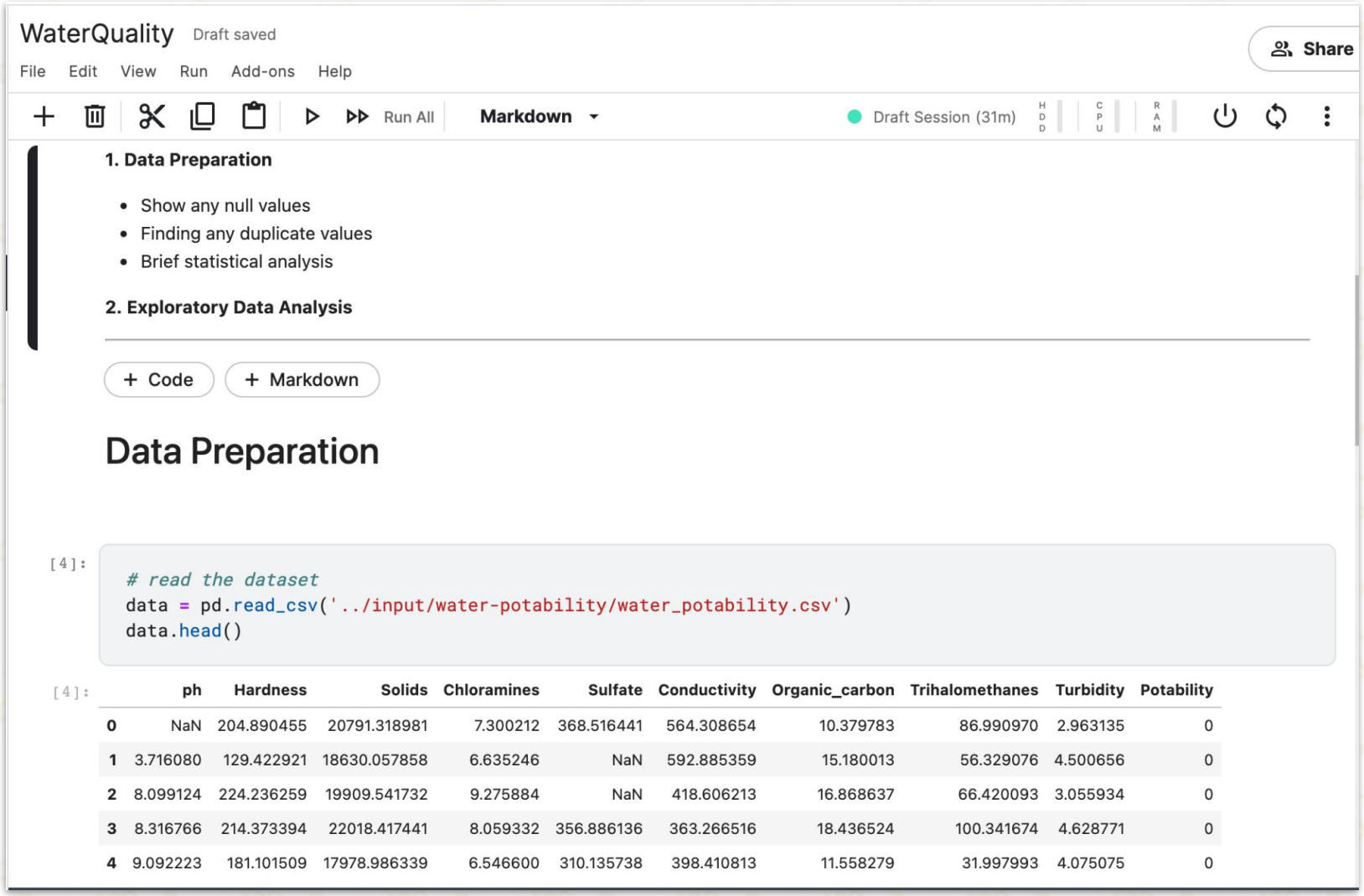
Lihat codingan yang dibuat

- Melakukan benchmark sebelum memulai analisis data

Mulai koding

- Mulai notebook

Water Potability Dataset



	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
0	NaN	204.890455	20791.318981	7.300212	368.516441	564.308654	10.379783	86.990970	2.963135	0
1	3.716080	129.422921	18630.057858	6.635246	NaN	592.885359	15.180013	56.329076	4.500656	0
2	8.099124	224.236259	19909.541732	9.275884	NaN	418.606213	16.868637	66.420093	3.055934	0
3	8.316766	214.373394	22018.417441	8.059332	356.886136	363.266516	18.436524	100.341674	4.628771	0
4	9.092223	181.101509	17978.986339	6.546600	310.135738	398.410813	11.558279	31.997993	4.075075	0

Tugas

https://www.kaggle.com/ramontanoeiro/winter-olympic-medals-1924-2018

Panduan tugas

1. Berikan deskripsi terkait data ini
2. Membuat notebook langsung di kaggle
3. Eksekusi data di dalam notebook kaggle dengan memunculkan
    a. Load data
    b. Informasi singkat terkait tipe datanya
    c. Informasi singkat terkait statistik dasar

DigitalSkola

Uncover The World Of Digital Skills With Us

Thank YOU