

LEARNING

PROGRESS REVIEW

Week 6

Diaz Jubairy - Hermulia Hadie
Desi Sulistyowati - Farahul Jannah

Table of Content

KAGGLE

- Kaggle dan Tujuannya
- Kegunaan Kaggle
- Fitur Kaggle



CRITICAL & ANALYTICAL THINKING

- Framework Design Thinking
- Aksi Design Thinking
- Permasalahan Bisnis X Permasalahan Human-centered
- Anda dan User
- Reinvention (Penemuan Kembali)
- Talking About Failure
- Bagaimana Hal-hal Dirangkum?
- Data Science Agile System

PANDAS DATAFRAME - INTERMEDIATE

- Sorting Dataframe
- Filtering Dataframe
- Membuat Kolom tambahan
- Mengelompokkan Data (Grouping)
- Menggabungkan Data antar Dataframe (Merging)



01

Kaggle

DATA SCIENCE COMMUNITY

KAGGLE



- Kaggle adalah wadah untuk komunitas Data Scientist dan Machine Learners, yang dimiliki oleh Google, Inc.
- Kaggle adalah platform untuk berinteraksi, melatih, dan menyelesaikan tantangan terkait Data Science dan Machine Learning
- Awalnya Kaggle menawarkan kompetisi dalam bidang Machine Learning.
- Sekarang, Kaggle juga menawarkan platform untuk public data berbasis cloud

Kegunaan Kaggle



LEARNING

Tempat belajar & praktik bagi Data Scientist pemula untuk mengerjakan kumpulan data aktual



COMPETITION

Tempat untuk bersaing, mempelajari & memenangkan kompetisi dari berbagai perusahaan terkemuka



DATASET

Tempat untuk menemukan & mempublikasikan dataset



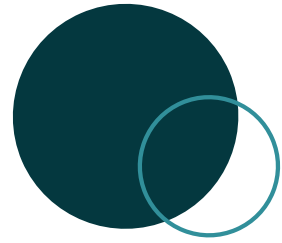
COLABORATION

Tempat untuk berkolaborasi bagi Data Scientist dan Machine Lerner



FITUR KAGGLE

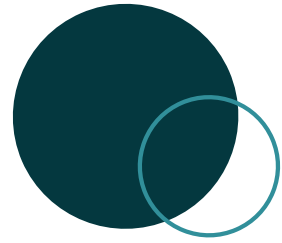
- Dataset
- Code
- Community
- Inspiration
- Competition
- Courses



FITUR KAGGLE

Dataset

- Kaggle memuat beberapa kumpulan data yang bisa digunakan untuk para Data Scientist.
- Sebagian besar kumpulan data dalam format file CSV dan beberapa format file lainnya, seperti JSON, SQLite, dan BigQuery
- Kumpulan data ini bermanfaat untuk para Data Scientist pemula berlatih dan bekerja dengan berbagai format file.
- Dalam dataset, kita juga bisa menemukan penjelasan informasi umum tentang data apa saja yang ada dalam dataset tersebut.



FITUR KAGGLE

❖ Dataset

 Create

 Home

 Competitions


 Datasets

 Code

 Discussions

 Courses

 More

 Your Work

▼ RECENTLY VIEWED

 Netflix subscription...

 Omicron daily case...

 All Songs From ...


 Images from The L...

 Harry Potter Movie...

▼ RECENTLY EDITED


 View Active Events

 Search

 Dataset

Cyberbullying Classification

47k tweets belonging to 6 balanced classes.

 Larxel • updated 24 days ago (Version 1)


[Data](#) [Code \(14\)](#) [Discussion](#) [Activity](#) [Metadata](#)

[Download \(7 MB\)](#)

[New Notebook](#)

 Usability 10.0

 License Attribution 4.0 International (CC BY 4.0)

 Tags text data, social networks, nlp, social issues and advocacy, people and society

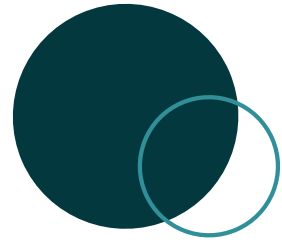
Description

Abstract

With rise of social media coupled with the Covid-19 pandemic, cyberbullying has reached all time highs. We can combat this by creating models to automatically flag potentially harmful tweets as well as break down the patterns of hatred.

About this dataset

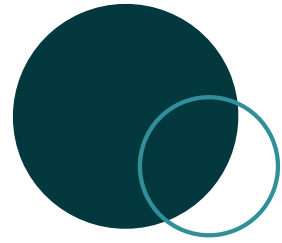
As social media usage becomes increasingly prevalent in every age group, a vast majority of citizens rely on this essential medium for day-to-day communication. Social media's ubiquity means that cyberbullying can effectively impact anyone at any time or anywhere, and the relative anonymity of the internet makes such personal attacks more difficult to stop than traditional bullying.



FITUR KAGGLE

Code

- Di dalam Kaggle kita bisa menemukan banyak codingan dari berbagai pengguna Kaggle di seluruh dunia.
- Untuk melihat codingan dari pengguna Kaggle lainnya, kita bisa mengakses dan mencarinya di dalam Notebook.
- Codingan dari para pengguna Kaggle lain dapat membantu kita untuk belajar, berlatih, dan melihat bagaimana pengguna lain memecahkan permasalahan yang sama dengan berbagai macam cara dan sudut pandang.
- Codingan yang biasanya banyak digunakan oleh pengguna Kaggle adalah bahasa pemrograman Python, selain itu ada juga bahasa pemrograman R, SQLite, dan Julia.



FITUR KAGGLE

■ Code

≡ kaggle

+ Create

🏠 Home

🏆 Competitions

📁 Datasets

<> Code

💬 Discussions

📚 Courses

✓ More

📁 Your Work

▼ RECENTLY VIEWED

📺 Cyberbullying Clas...

📺 Netflix subscription...

📺 Omicron daily case...

🎵 All Songs From ...

📺 Images from The L...

▼ RECENTLY EDITED

📁 View Active Events

🔍 Search

Pizza Price Data

Notebook Data Logs Comments (3)

8

Copy & Edit

Table of Contents

Pizza Price Data

Abstract

In [2]:

```
#Importing the dataset  
  
df=pd.read_csv("../input/pizza-price-prediction-real-data/pizza_data.csv")
```

In [3]:

```
# looking the data set  
df.head()
```

Out[3]:

	Company	Pizza Name	Type	Size	Price
0	Domino's Pizza	Hand Tossed	Cheeses Pizza	Small (10")	\$5.99
1	Domino's Pizza	Hand Tossed	Cheeses Pizza	Medium (12")	\$7.99
2	Domino's Pizza	Hand Tossed	Cheeses Pizza	Large (14")	\$9.99
3	Domino's Pizza	Handmade Pan	Cheeses Pizza	Medium (12")	\$7.99
4	Domino's Pizza	Crunchy Thin Crust	Cheeses Pizza	Small (10")	\$5.99

In [4]:

```
# print the shape dataset  
print("Shape The DataSet ", df.shape )
```

```
Shape The DataSet (371, 5)
```

In [5]:

FITUR KAGGLE

Community

- ❑ Seperti Medium, GitHub, dan LinkedIn, Kaggle berfungsi sebagai komunitas tempat Data Analysts, Data Scientists, dan Machine Learning Engineers untuk belajar, berkembang, dan berjejaring.
- ❑ Pengguna Kaggle dapat membagikan pekerjaannya (misalnya, data, code, dan notebook) mengembangkan komunitas sendiri dan menjangkau komunitas lainnya.

FITUR KAGGLE

Inspiration

Fitur Kaggle, seperti dataset, code, komunitas, kursus dan kompetensi, ini bisa dijadikan pengguna Kaggle sebagai inspirasi. Inspirasi untuk menyelesaikan suatu permasalahan yang rumit atau membuat pekerjaan yang lebih baik.

Competition

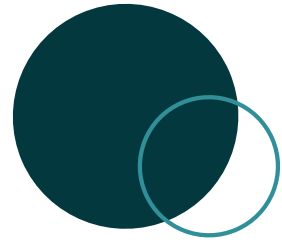
- Fitur kompetisi ini dapat dijadikan sebagai portofolio pada saat akan melamar pekerjaan di suatu organisasi.
- Selain itu, fitur kompetisi ini biasanya memberikan hadiah bagi pemenangnya, sehingga pengguna Kaggle bisa mendapatkan uang ekstra dari kompetisi ini.
- Kompetisi ini juga dapat membantu suatu organisasi untuk menyelesaikan permasalahan yang mungkin mereka tidak bisa menyelesaikannya.

FITUR KAGGLE

■ Competition

The screenshot displays the Kaggle homepage with a sidebar on the left and a main grid of competitions on the right. The sidebar includes navigation links for Home, Competitions, Datasets, Code, Discussions, Courses, and More. It also features a 'Your Work' section with 'RECENTLY VIEWED' items like 'Pizza Price Data' and 'Cyberbullying Clas...', and a 'RECENTLY EDITED' section with 'Omicron daily case...' and 'All Songs From ...'. The main grid shows eight competitions, each with a cover image, title, description, category, team count, prize amount, and time remaining.

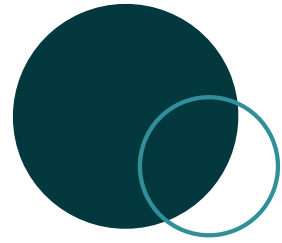
Competition Title	Category	Teams	Prize	Time to Go
Feedback Prize - Evaluating Student Writing	Code Competition	1282 Teams	\$160,000	a month to go
TensorFlow - Help Protect the Great Barrier Reef	Research	2006 Teams	\$150,000	5 days to go
Ubiquant Market Prediction	Code Competition	1191 Teams	\$100,000	2 months to go
NBME - Score Clinical Patient Notes	Code Competition	174 Teams	\$50,000	3 months to go
H&M Personalized Fashion Recommendations	Featured	124 Teams		
Happywhale - Whale and Dolphin Identification	Research	257 Teams		
Tabular Playground Series - Feb 2022	Playground	594 Teams		
Predict Future Sales	Playground	13522 Teams		



FITUR KAGGLE

❑ Courses

- Kaggle juga menyediakan fitur kursus untuk para Data Scientist baru.
- Ada sekitar 14 kursus yang disediakan oleh Kaggle, diantaranya adalah
 - Python
 - Intro to Machine Learning
 - Intermediate Machine Learning
 - Data Visualization
 - Pandas
 - Feature Engineering
 - Deep Learning
 - Intro to SQL
 - Advanced SQL
 - Geospatial Analysis
 - Microchallenges
 - Machine Learning Explainability
 - Natural Language Processing
 - Intro to Game AI and Reinforcement Learning



FITUR KAGGLE

📁 Courses

☰ kaggle

+

 Create

🏠

 Home

🏆

 Competitions

📁

 Datasets

<>

 Code

💬

 Discussions

🎓

 Courses

▼

 More

📋

 Your Work

▼

 RECENTLY VIEWED

👤

 Pizza Price Data

📺

 Cyberbullying Clas...

📺

 Netflix subscription...

📺

 Omicron daily case...

👤

 All Songs From ...

▼

 RECENTLY EDITED

📅

 View Active Events

🔍 Search

👤

📊 Explore Courses

📋 📅

📊

Intro to Machine Learning

Learn the core ideas in machine learning, and build your first models.

📊

Pandas

Solve short hands-on challenges to perfect your data manipulation skills.

📊

Data Visualization

Make great data visualizations. A great way to see the power of coding!

📊

Feature Engineering

Better features make better models. Discover how to get the most out of your data.

📊

Intro to SQL

Learn SQL for working with databases, using Google BigQuery.

📊

Advanced SQL

Take your SQL skills to the next level.

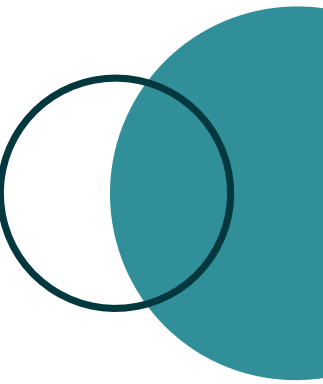
📊

Intro to Deep Learning

Use TensorFlow and Keras to build and train neural networks for structured data.

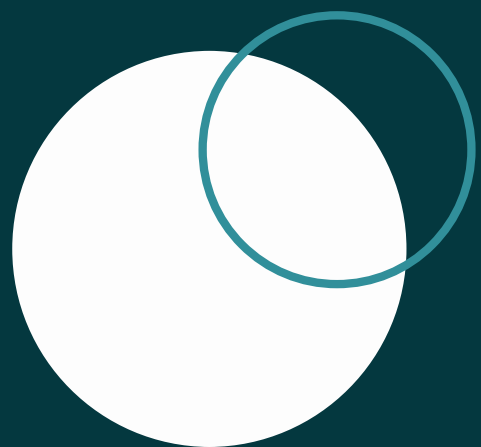
02

Critical & Analytical Thinking



Design Thinking

Design thinking adalah sebuah pemikiran untuk memiliki tujuan untuk meningkatkan skill seseorang berdasarkan pengalaman



Design Thinking Framework



- The Principles:

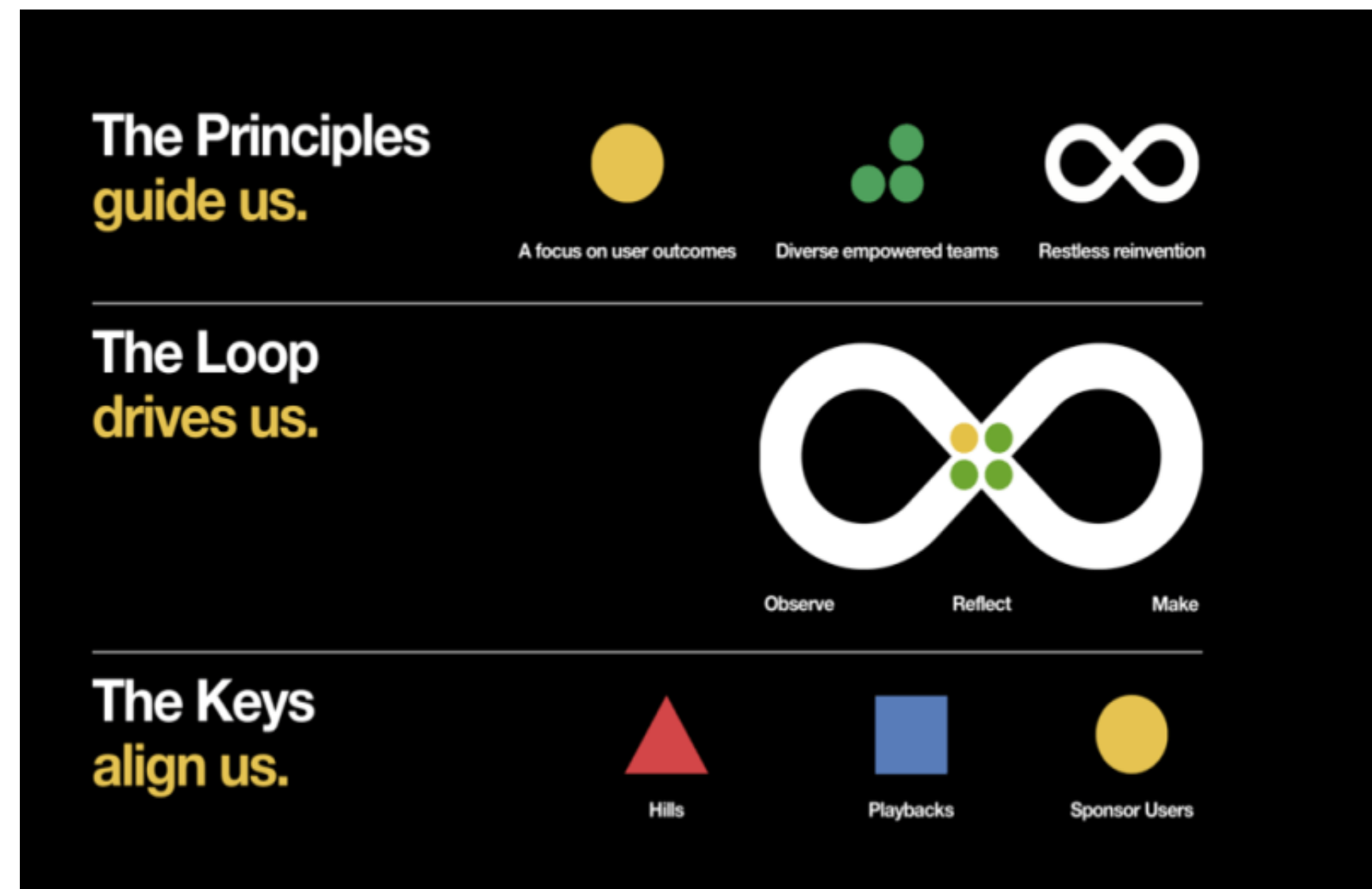
Mendorong bisnis dengan menolong kolega mencapai tujuannya

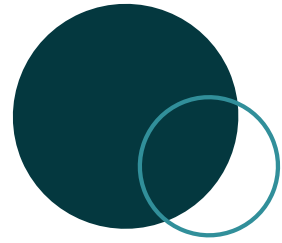
- The Loops:

Selalu memperlakukan tugas atau pekerjaan sebagai sebuah prototype

- The Keys:

Bergerak cepat dengan bekerja bersama dan merangkul keberagaman





The Principle

See problems and solutions from a new point of view



A focus on user
outcomes >

Drive business by helping
users achieve their goals.



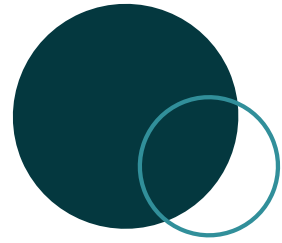
Restless
reinvention >

Stay essential by treating
everything as a prototype.



Diverse empowered
teams >

Move faster by empowering
diverse teams to act.



The Loops

Understand the present and envision the future in a continuous cycle of observing, reflecting, and making



Observe >

Immerse yourself in the real world.

Reflect >

Come together and look within.

Make >

Give concrete form to abstract ideas.

Observe



- Berbicara dengan user,
- melihat mereka bekerja,
- tes ide anda dengan user terkait pemahaman anda

Reflect



- Bersama-sama untuk mensinkronisasikan pergerakan anda,
- implementasikan yang dipelajari
- sharing "aha" moment dgn yang lainnya

Make



Meletakkan ide mentah dan kembangkan sembari berjalannya waktu

The Key



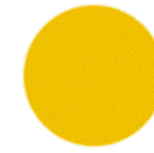
Hills >

Align teams on meaningful user outcomes to achieve.



Playbacks >

Stay aligned by regularly exchanging feedback



Sponsor Users >

Invite users into the work to stay true to real world needs

Hills mendeskripsikan sesuatu yang user dapat lakukan

- Siapa user anda?
- Apa yang akan user dapat lakukan yang sebelumnya tidak dapat lakukan?
- Apa yang anda dapat lakukan untuk menggapai mereka?

Playback : story-based presentations untuk membagikan wawasan, ide dan perkembangan

Sponsor users: client external atau end user yang berkontribusi terkait domain

Aksi Design Thinking



Design thinking diinisiasi dengan "business problem" seperti membuat segmentasi pasar



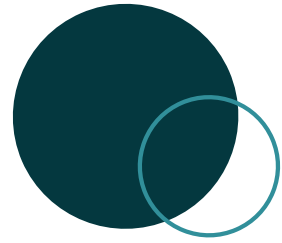
Adopsi design thinking dan pengalamannya untuk mendapatkan outcome yang luar biasa



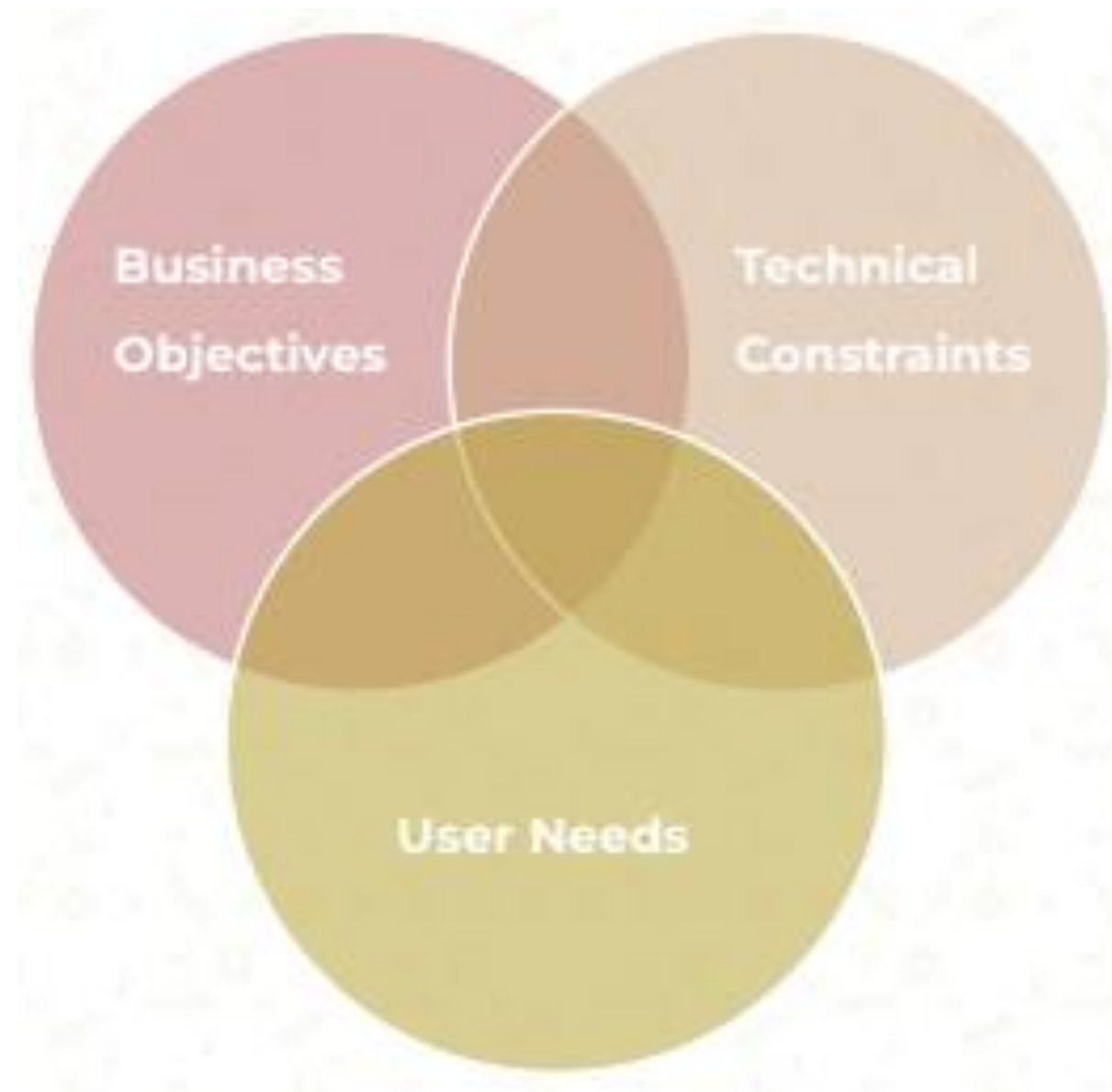
Tools dan pengetahuan kamu perlu membicarakan tantangan dan memanfaatkan kesempatan yang dihadapi sebagai latihan



Bekerja dengan sebuah **business problem** dengan menyusun hal tersebut dengan **human experiences**.

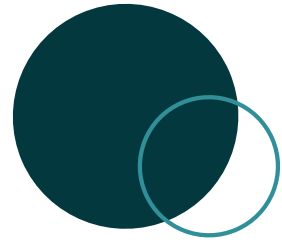


Business Problems to Human-Centered Problems



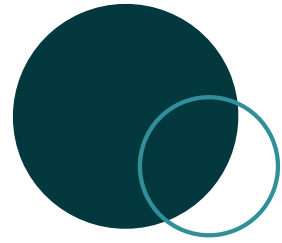
Business problem yang anda coba untuk selesaikan secara day-to-day fokus pada sesuatu yang bukan “human”, seperti segmentasi market

Dalam hal untuk memulai fokus pada hal tersebut, anda bisa mulai dengan fokus pada **permasalahan user-nya** yang mendasari business problem



Anda & Users

- You are not users
- Walaupun memiliki pengalaman yang sama, namun kita tidak tahu dari dekat seperti apa kehidupannya
- Dalam hal untuk dekat dengan user tersebut, kita harus memiliki asumsi dengan mempelajari background user, observasi dan latihan



Observasi Membangun Empati

- Kita disini tidak membangun “vas bunga”, tapi kita disini untuk memahami **kebutuhan** dan **keinginan** seseorang
- Membangun **empati** terhadap user dan mengungkapkan kebutuhan mereka akan memudahkan kita untuk menyelesaikan “business problem” daripada hanya berfokus pada bisnis prosesnya





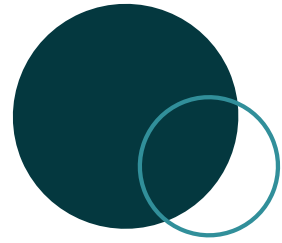
Saat membangun design thinking Anda, penting untuk mengasah rasa ingin tahu Anda

Ingat! Ketika kamu kecil, **seringkali kamu bertanya tentang hal-hal tentang dunia karena kamu penasaran, kan?**

Restless Reinvention



- **Restless Reinvention:** Pekerjaan kita tidak akan pernah selesai
- Prinsip dari design thinking yang mewakili **pengujian** dan **pembelajaran berkelanjutan yang aktif** yang bertujuan untuk meningkatkan solus**pengujian**ani



View everything as prototype

- Kita bisa mulai mengubah pola pikir dari menunggu hingga sesuatu sempurna menjadi menunggu hingga siap
- Mode kerja ini memungkinkan pengiriman nilai yang lebih teratur kepada pengguna
- Pertimbangan: kami masih memperbaiki cara kami berpindah dari titik A ke titik B, Contoh:
 - Hari ini kami membuat prototipe untuk menggantikan kereta kuda ke mobil
 - Mobil hari ini hanyalah prototipe lain untuk transportasi masa depan

Prototype



Tidak ada solusi yang sempurna dan kita tidak akan pernah selesai

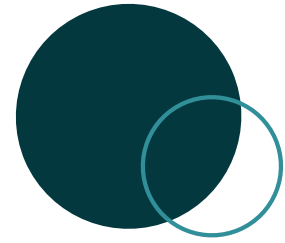
Bahkan pengguna selalu mencari pengalaman yang lebih baik untuk menyelesaikan masalah mereka.

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

Ketika Anda memiliki ide baru untuk dibagikan dengan tim Anda. Anda harus menggambarkan terlebih dahulu, lalu bagikan ke tim Anda, .



A picture is worth a thousand words...



Talking about Failure

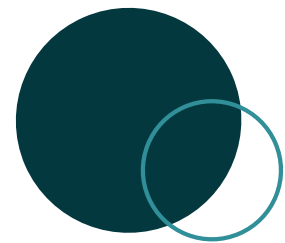
Apakah kamu takut jika task yang kamu kerjakan gagal?

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



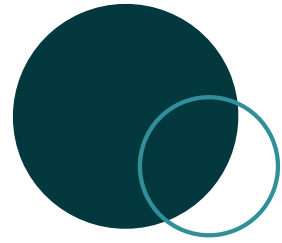
How things stick

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:



Starting a new project or initiative. Answer questions like: *Who will be the users and stakeholders? What experience are we trying to improve and why?*



Deciding as a team on a future experience for your users. Answer questions like: *What do we think our users need to be successful? How are we going to serve those needs?*



Reviewing progress as you deliver. Answer questions like: *Do we successfully deliver value to our users? Are we still aligned as a team?*

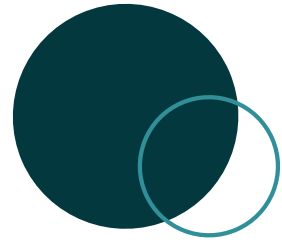


Teams who share their goals by talking about a user and their needs, and invite feedback along the way, are more likely to understand and deliver on those goals together over time.

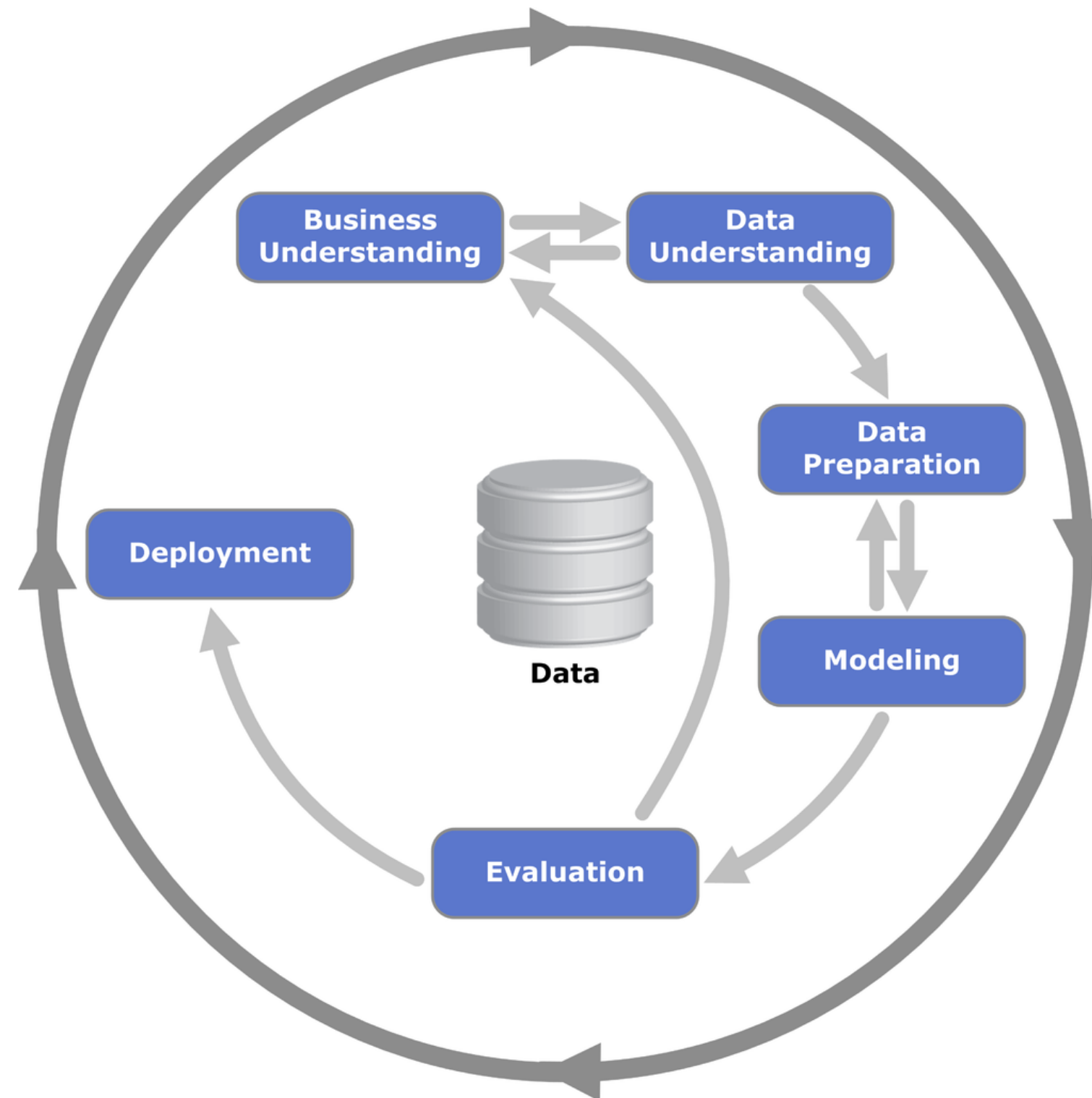
Day To Day Data Science Cycle

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





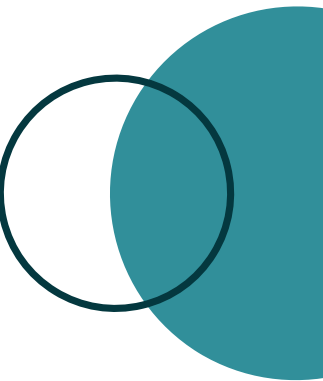
Data Science Agile System



03

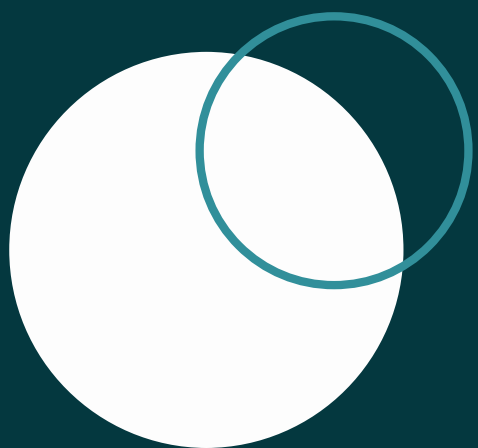
Pandas Dataframe

Intermediate



Content

- Sorting Dataframe
- Filtering Dataframe
- Membuat Kolom tambahan
- Mengelompokkan Data (Grouping)
- Menggabungkan Data antar Dataframe (Merging)



Sorting

- Pandas dapat melakukan pengurutan suatu data berdasarkan perintah dan output yang diinginkan.
- Mengurutkan dataframe pandas dapat berdasarkan nilai baris/nilai kolom atau indeks baris/indeks kolom.

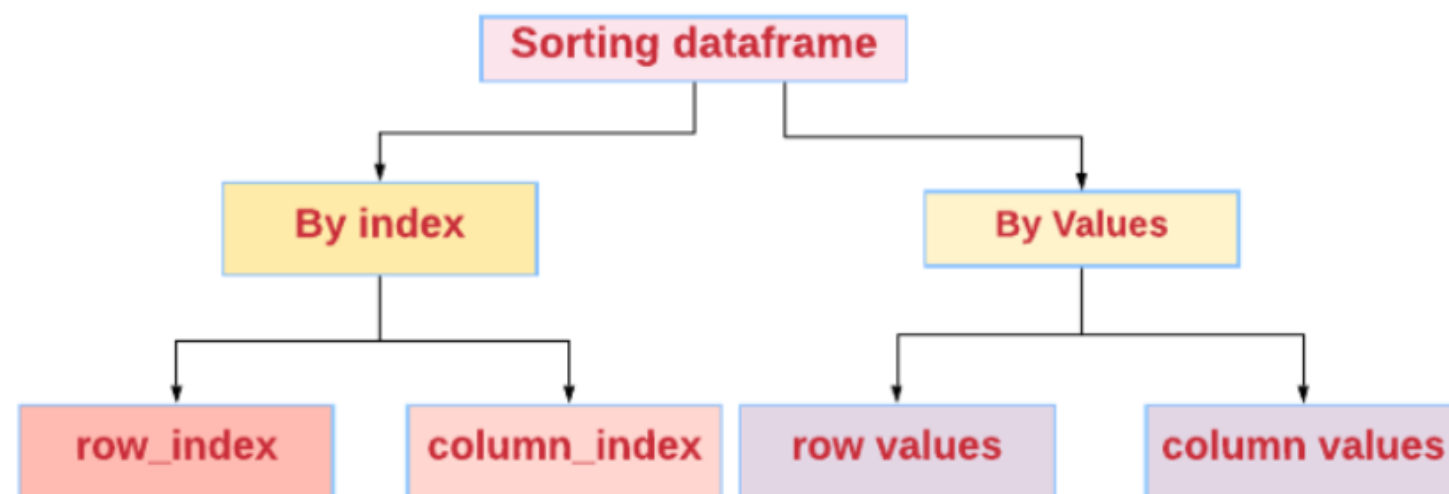


Diagram illustrating a pandas dataframe with sorting annotations:

	EmpID	Salary	Skill
0	21	15000.0	Python
1	12	NaN	python
2	15	5000.0	JavaScript
3	7	12000.0	JavaScript
4	10	10000.0	PYTHON

Annotations:

- Column_index**: Points to the header row (EmpID, Salary, Skill).
- Row_index**: Points to the index column (0, 1, 2, 3, 4).
- Column_values**: Points to the data columns (EmpID, Salary, Skill).

Diagram illustrating the sorted dataframe (sorted by index):

	EmpID	Salary	Skill
0	21	15000.0	Python
1	12	NaN	python
2	15	5000.0	JavaScript
3	7	12000.0	JavaScript
4	10	10000.0	PYTHON

Annotation:

- Row_values**: Points to the data rows (0, 1, 2, 3, 4).

Sorting in Dataframe

Alphabetic Descending

```
## Alphabetic sorting descending  
data.sort_values(by=[ 'sex' ], ascending=False)
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520



	age	sex	bmi	children	smoker	region	charges
446	60	male	29.64	0	no	northeast	12730.9996
1052	49	male	29.83	1	no	northeast	9288.0267
1070	37	male	37.07	1	yes	southeast	39871.7043
550	63	male	30.80	0	no	southwest	13390.5590
1068	63	male	21.66	1	no	northwest	14349.8544

Sorting in Dataframe

Numerical

```
## Numerical sorting ascending  
data.sort_values(by=[ 'age' ])
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520



	age	sex	bmi	children	smoker	region	charges
1248	18	female	39.82	0	no	southeast	1633.96180
482	18	female	31.35	0	no	southeast	1622.18850
492	18	female	25.08	0	no	northeast	2196.47320
525	18	female	33.88	0	no	southeast	11482.63485
529	18	male	25.46	0	no	northeast	1708.00140

Sorting in Dataframe

Sorting two columns

```
## Sorting more than two columns  
data.sort_values(by=[ 'age', 'bmi' ])
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520



	age	sex	bmi	children	smoker	region	charges
172	18	male	15.960	0	no	northeast	1694.79640
250	18	male	17.290	2	yes	northeast	12829.45510
359	18	female	20.790	0	no	southeast	1607.51010
1212	18	male	21.470	0	no	northeast	1702.45530
1033	18	male	21.565	0	yes	northeast	13747.87235

Filtering



Pandas dapat menyeleksi beberapa data berdasarkan perintah dan output yang diinginkan. Contoh:

- Seleksi kolom tertentu
- Seleksi baris dan kolom
- Seleksi data tertentu pada suatu kolom

Filtering in Dataframe

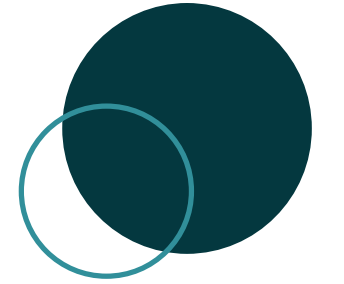
Filtering some columns

```
data[["age", "sex"]]
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

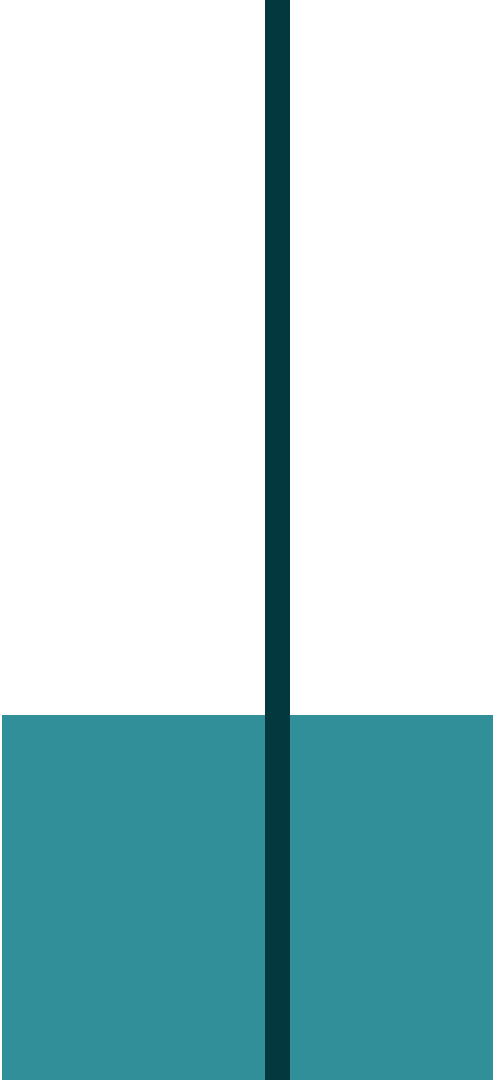


	age	sex
0	19	female
1	18	male
2	28	male
3	33	male
4	32	male



Filtering in Dataframe

Pada Dataframe, terdapat 2 fungsi untuk melakukan seleksi yaitu loc dan iloc.

- loc untuk melakukan seleksi kolom dan baris yang berdasar pada nama baris dan kolom.
 - iloc untuk melakukan seleksi kolom dan baris yang berdasar indeks baris dan kolom.
- 

Filtering in Dataframe

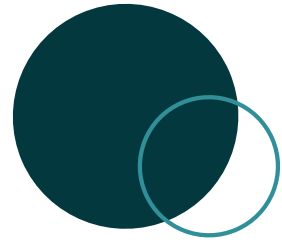
loc

```
# loc  
data.loc[(data["sex"] == "male")]
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520



	age	sex	bmi	children	smoker	region	charges
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
8	37	male	29.830	2	no	northeast	6406.41070



Filtering in Dataframe

iloc

```
# iloc with index 10 to 20  
data.iloc[10:21,:]
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520



	age	sex	bmi	children	smoker	region	charges
10	25	male	26.22	0	no	northeast	2721.3208
11	62	female	26.29	0	yes	southeast	27808.7251
12	23	male	34.40	0	no	southwest	1826.8430
13	56	female	39.82	0	no	southeast	11090.7178
14	27	male	42.13	0	yes	southeast	39611.7577

Creating new variable



Pandas juga dapat melakukan penambahan variable dari dataset yang ada. Contoh:

- Melakukan standar kalkulasi (+, -, x, :)
- Melakukan pengelompokan suatu data

Creating new variable

Menambah variable (kolom diskon)

```
data['discount_charges'] = data['charges']*0.15  
data.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520



	age	sex	bmi	children	smoker	region	charges	discount_charges
0	19	female	27.900	0	yes	southwest	16884.92400	2532.738600
1	18	male	33.770	1	no	southeast	1725.55230	258.832845
2	28	male	33.000	3	no	southeast	4449.46200	667.419300
3	33	male	22.705	0	no	northwest	21984.47061	3297.670591
4	32	male	28.880	0	no	northwest	3866.85520	580.028280

Grouping

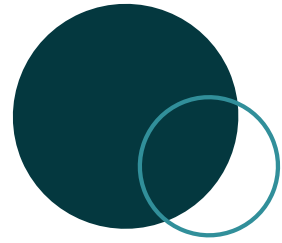
Pandas juga memfasilitasi grouping data, kita dapat melakukan pengelompokan data berdasarkan nama tertentu dan menghitung statistiknya.

```
# grouping average age and median charges in sex variable
data.groupby('sex').agg(average_age = ('age', 'mean'),
                        median_charges = ('charges', 'median'))
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520



	average_age	median_charges
sex		
female	39.503021	9412.96250
male	38.917160	9369.61575



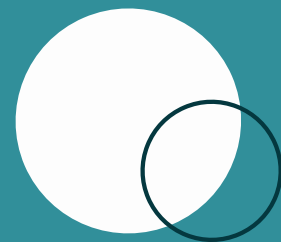
Merging Dataframe

Pandas memiliki fungsi untuk menggabungkan dataset berdasarkan kemiripan suatu fitur. Fungsi bawaan Pandas . *Merge()* menyediakan metode yang ampuh untuk menggabungkan dua DataFrame menggunakan merge gaya database

	age	sex	bmi	children	smoker	region	charges	discount_charges	Name		Type of House	Price House	Name
0	19	female	27.900	0	yes	southwest	16884.92400	2532.738600	doni	0	House 1	10000	doni
1	18	male	33.770	1	no	southeast	1725.55230	258.832845	jojo	1	House 2	12300	jojo
2	28	male	33.000	3	no	southeast	4449.46200	667.419300	beki	2	House 3	11000	beki
3	33	male	22.705	0	no	northwest	21984.47061	3297.670591	madrid	3	House 4	14000	madrid
4	32	male	28.880	0	no	northwest	3866.85520	580.028280	milan	4	House 5	12100	milan

```
data_5.merge(data2, how='inner', on='Name')
```

	age	sex	bmi	children	smoker	region	charges	discount_charges	Name	Type of House	Price House
0	19	female	27.900	0	yes	southwest	16884.92400	2532.738600	doni	House 1	10000
1	18	male	33.770	1	no	southeast	1725.55230	258.832845	jojo	House 2	12300
2	28	male	33.000	3	no	southeast	4449.46200	667.419300	beki	House 3	11000
3	33	male	22.705	0	no	northwest	21984.47061	3297.670591	madrid	House 4	14000
4	32	male	28.880	0	no	northwest	3866.85520	580.028280	milan	House 5	12100



”

THANKYOU



“

