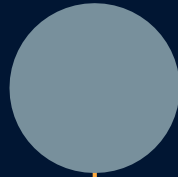


# Learning Progress Week-6

Group 4, Optimistic Team



## TEAM MEMBER



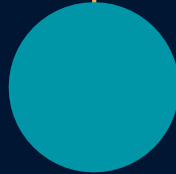
**Asprizal Rizky**



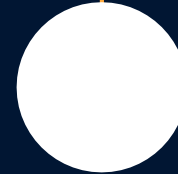
**Lutfia Humairosi**



**Aldiva Wibowo**



**Gilang Rahmat**



**Millenia Winadya**

# Introduction & Dataframe

## **Dataframe**

Dataframe merupakan objek utama ketika akan melakukan proses analisis data menggunakan Python, tampilan Dataframe berupa kolom dan baris yang dimana datanya sudah terstruktur.

## **Pandas**

Pandas adalah library pada Python yang dibangun diatas Numpy yang berguna untuk melakukan analisis data dan manipulasi data. Pandas digunakan penulisan serta untuk membaca Data CSV, Excel, dll.

# Introduction & Dataframe

## Dataframe

	Name	Age	Fav Fruit
1	Pia	21	Pear
2	Milen	21	Apple
3	Aldiva	21	Mango
4	Rizky	21	Mango
5	Gilang	21	Orange



# Creating Dataframe

## Method 1, List of lists

Cara membuat dataframe dengan pandas dengan list adalah sebagai berikut:

```
# method 1 : list of lists
data_1 = [['Pia', 21, 'Pear'], ['Milen', 21, 'Apple'], ['Aldiva', 21, 'Mango'], ['Rizky', 21, 'Mango'], ['Gilang', 21, 'Orange']]

df_1 = pd.DataFrame(data_1, columns = ['Name', 'Age', 'Fav Fruit'])
df_1
```

	Name	Age	Fav Fruit
0	Pia	21	Pear
1	Milen	21	Apple
2	Aldiva	21	Mango
3	Rizky	21	Mango
4	Gilang	21	Orange

```
# ganti index
df_1.index += 1
df_1
```

	Name	Age	Fav Fruit
1	Pia	21	Pear
2	Milen	21	Apple
3	Aldiva	21	Mango
4	Rizky	21	Mango
5	Gilang	21	Orange

# Creating Dataframe

## Method 2, Dict of lists

Cara membuat dataframe dengan pandas dengan list dalam dict adalah sebagai berikut:

```
# method 2 : dict of lists
data_2 = {'Name' : ['Pia', 'Milen', 'Aldiva', 'Rizky', 'Gilang'],
          'Age' : [21, 21, 21, 21, 21],
          'Fav Fruit' : ['Pear', 'Apple', 'Mango', 'Mango', 'Orange']}

df_2 = pd.DataFrame (data_2)
df_2
```

	Name	Age	Fav Fruit
0	Pia	21	Pear
1	Milen	21	Apple
2	Aldiva	21	Mango
3	Rizky	21	Mango
4	Gilang	21	Orange

# Creating Dataframe

## Method 3, List of dicts

Cara membuat dataframe dengan pandas dengan dict dalam list adalah sebagai berikut:

```
# method 3 : list of dicts
data_3 = [ {'Name':'Pia','Age':21,'Fav Fruit':'Pear'},
            {'Name':'Milen','Age':21,'Fav Fruit':'Apple'},
            {'Name':'Aldiva','Age':21,'Fav Fruit':'Mango'},
            {'Name':'Rizky','Age':21,'Fav Fruit':'Mango'},
            {'Name':'Gilang','Age':21,'Fav Fruit':'Orange'}]

df_3 = pd.DataFrame(data_3)
df_3
```

	Name	Age	Fav Fruit
0	Pia	21	Pear
1	Milen	21	Apple
2	Aldiva	21	Mango
3	Rizky	21	Mango
4	Gilang	21	Orange

# Read Dataframe csv

```
In [26]: article_read = pd.read_csv('pandas_tutorial_read.csv', delimiter=';',
                                     names = ['my_datetime', 'event', 'country', 'user_id', 'source', 'topic'])
```

```
In [27]: article_read
```

```
Out[27]:
```

	my_datetime	event	country	user_id	source	topic
0	2018-01-01 00:01:01	read	country_7	2458151261	SEO	North America
1	2018-01-01 00:03:20	read	country_7	2458151262	SEO	South America
2	2018-01-01 00:04:01	read	country_7	2458151263	AdWords	Africa
3	2018-01-01 00:04:02	read	country_7	2458151264	AdWords	Europe
4	2018-01-01 00:05:03	read	country_8	2458151265	Reddit	North America
5	2018-01-01 00:05:42	read	country_6	2458151266	Reddit	North America
6	2018-01-01 00:06:06	read	country_2	2458151267	Reddit	Europe
7	2018-01-01 00:06:15	read	country_6	2458151268	AdWords	Europe
8	2018-01-01 00:07:21	read	country_7	2458151269	AdWords	North America
9	2018-01-01 00:07:29	read	country_5	2458151270	Reddit	North America
10	2018-01-01 00:07:57	read	country_5	2458151271	AdWords	Asia



# Read Dataframe Excel

```
In [17]: # Reading an Excel file and viewing the first 5 rows.
results = pd.read_excel("2016_results.xlsx")
results.head()
```

Out[17]:

	state	state_abbreviation	county	fips	party	candidate	votes	fraction_votes
0	Oklahoma	OK	Major	40093	Republican	John Kasich	45	0.024000
1	Iowa	IA	Wayne	19185	Democrat	Uncommitted	0	0.000000
2	Ohio	OH	Adams	39001	Republican	Donald Trump	2594	0.482000
3	Nebraska	NE	Stanton	31167	Democrat	Hillary Clinton	19	0.352000
4	New Hampshire	NH	Cheshire	33005	Republican	Marco Rubio	1237	0.093926

# Descriptive Statistics

```
In [4]: df.describe()
```

```
Out[4]:
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

# Descriptive Statistics Object include

```
In [7]: df.describe(include="object")
```

```
Out[7]:
```

	Name	Sex	Ticket	Cabin	Embarked
count	891	891	891	204	889
unique	891	2	681	147	3
top	Panula, Mr. Ernesti Arvid	male	347082	C23 C25 C27	S
freq	1	577	7	4	644

# Descriptive Statistics for All Include Object

```
df.describe(include="all")
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
count	891.000000	891.000000	891.000000	891	891	714.000000	891.000000	891.000000	891	891.000000	204	889
unique	NaN	NaN	NaN	891	2	NaN	NaN	NaN	681	NaN	147	3
top	NaN	NaN	NaN	Panula, Mr. Ernesti Arvid	male	NaN	NaN	NaN	347082	NaN	C23 C25 C27	S
freq	NaN	NaN	NaN	1	577	NaN	NaN	NaN	7	NaN	4	644
mean	446.000000	0.383838	2.308642	NaN	NaN	29.699118	0.523008	0.381594	NaN	32.204208	NaN	NaN
std	257.353842	0.486592	0.836071	NaN	NaN	14.526497	1.102743	0.806057	NaN	49.693429	NaN	NaN
min	1.000000	0.000000	1.000000	NaN	NaN	0.420000	0.000000	0.000000	NaN	0.000000	NaN	NaN
25%	223.500000	0.000000	2.000000	NaN	NaN	20.125000	0.000000	0.000000	NaN	7.910400	NaN	NaN
50%	446.000000	0.000000	3.000000	NaN	NaN	28.000000	0.000000	0.000000	NaN	14.454200	NaN	NaN
75%	668.500000	1.000000	3.000000	NaN	NaN	38.000000	1.000000	0.000000	NaN	31.000000	NaN	NaN
max	891.000000	1.000000	3.000000	NaN	NaN	80.000000	8.000000	6.000000	NaN	512.329200	NaN	NaN

# Export Dataframe

Dalam export dataframe bisa menjadi ekstensi file seperti CSV, EXCEL, SQL, HTML, dll

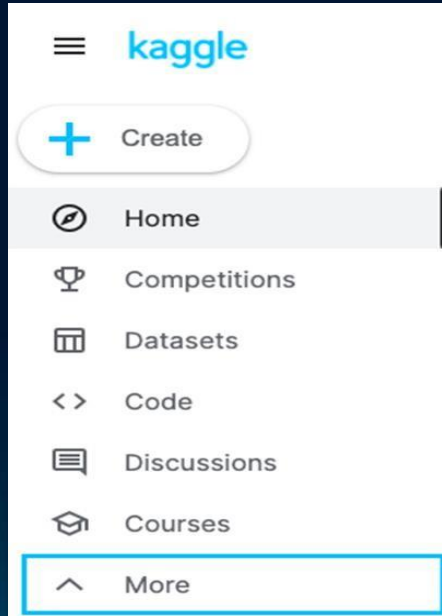
```
# Export to CSV
df.to_csv('filepath.csv', index=False, delimiter=';')

#Export to Excel
df.to_excel('filepath.xlsx', sheet_name='Sheet1', index=False)
```

# INTRODUCTION TO KAGGLE

Apa itu Kaggle?

Komunitas online data science terbesar di dunia dengan fitur lengkap dan data yang dapat dipercaya untuk membantu dalam pemrosesan data science agar sesuai dengan tujuan yang dikehendaki.



Mengapa Kaggle?

- Merupakan komunitas data science terbesar dan belum ada yang dapat menyaingi.
- Menyediakan project dan kompetisi data science secara umum dimana peminat data dapat bergabung untuk belajar dan berlatih.
- Terdapat ruang diskusi bagi sesama peminat data dari bermacam-macam latar belakang dan industri.

# INTRODUCTION TO KAGGLE

## Kaggle Services

### 1. Competitions

Fitur utama dari Kaggle dimana suatu perusahaan memberikan permasalahan yang harus dikerjakan oleh peserta menggunakan ML sesuai kebutuhan untuk membuat algoritma atau model terbaik dengan tujuan meningkatkan kinerja perusahaan.

- **Datasets**

Kaggle menyediakan banyak datasets yang bisa dengan mudah diunduh dimana bersumber dari berbagai latar belakang dan industri serta dapat dimanfaatkan untuk mengunggah dataset. Format dari datasets yang paling banyak dipakai adalah CSV, selain itu ada juga format lain seperti JSON.

- **Kernels**

Merupakan fitur yang sama fungsinya dengan Google Colab, yang dapat membuat kodingan sendiri melihat dan berbagi kodingan dengan pengguna lain sehingga antar pengguna dapat berkolaborasi. Bahasa pemrograman yang dapat digunakan yaitu Python, R, R Markdown, dll.

- **Discussions**

Berupa forum atau ruang untuk berdiskusi antar pengguna untuk menyelesaikan masalah apabila menghadapi suatu kesulitan sehingga bisa membangun networking dan bekerja bersama.

- **Courses**

Kaggle menyediakan pelatihan gratis dalam bahasa inggris untuk banyak topik seputar Data Science.

# INTRODUCTION TO KAGGLE

- Kaggle Progression System**

Setiap pengguna dapat menggunakan seluruh fitur Kaggle mulai dari level pemula (Novice) sampai Grandmaster.





# CRITICAL DAN ANALYTICAL THINKING

- **DESIGN THINKING**

Merupakan sebuah pemikiran yang digunakan untuk memecahkan masalah dengan orientasi pada pengalaman yang dimiliki oleh user

Design thinking akan berusaha memahami kebutuhan user bersangkutan agar dapat mengidentifikasi solusi terbaik terhadap permasalahan yang ada.

- **Design Thinking Framework**

The principles : Fokus pada kebutuhan user, memperhatikan perbedaan pandangan pada sebuah tim.

The loops : Observe, Reflect and make. Pertama lihat dan bicaralah kepada user lalu sampaikan ide yang dimiliki, setelah itu hubungkan Kembali ide tersebut dengan orang-orang yang paham dengan project tersebut.

The Keys: Menjelaskan sesuatu yang dapat dilakukan oleh seorang user .

# CRITICAL DAN ANALYTICAL THINKING

## DESIGN THINKING IN ACTION

- Seluruh inisiatif dari *design thinking* dimulai dengan sebuah permasalahan bisnis, seperti menemukan segmentasi pasar terbaru
- Mengadopsi *design thinking* dan memperoleh *outcome* yang besar yang kita baru dengar/tahu tidak terjadi dalam waktu yang sebentar
- *Tools* dan pengetahuan yang dibutuhkan untuk menghadapi tantangan dan memanfaatkan peluang akan dirasakan ketika telah mengerjakan Latihan yang ada
- Mengatasi permasalahan bisnis dengan melakukan *framing* pada masalah tersebut berdasarkan pengalaman orang lain.

## BUSINESS PROBLEMS TO HUMAN-CENTERED PROBLEMS

- Masalah bisnis yang coba diselesaikan seringkali berfokus pada manusia
- Untuk mulai berfokus pada hal tersebut maka perlu diperhatikan permasalahan user yang mendasari permasalahan bisnis.

# CRITICAL DAN ANALYTICAL THINKING

## YOU AND USERS

- Bagaimanapun juga kita bukanlah users
- Walaupun kita memiliki pengalaman yang sama, namun dengan latar belakang berbeda apa yang didapat bisa jadi akan berbeda
- Untuk mengetahui apa yang user benar-benar alami, kita harus melandasi asumsi kita dengan terus belajar, mengamati dan berlatih

## OBSERVATION BUILDS EMPATHY

- Kita disini tidak sekedar mendesain suatu hal, melainkan untuk memahami apa yang dibutuhkan dan diinginkan seseorang.
- Membangun empati kepada *user* dan mengungkapkan kebutuhan mereka akan membuat kita lebih dekat untuk memecahkan masalah bisnis dibanding hanya berfokus pada bisnisnya saja

## RELENTLESS REINVENTION

- Prinsip dari suatu pemikiran adalah bahwasannya kita harus selalu berfikir untuk kedepannya dan melakukan percobaan dan pembelajaran untuk mendapatkan sebuah solusi yang lebih baik.

# CRITICAL DAN ANALYTICAL THINKING

## *Talking About Failure*

- Kebanyakan orang takut pada kegagalan sehingga tidak mau mengambil sebuah resiko. Tapi, untuk mendapatkan pemikiran yang **hebat dan luar biasa**, terkadang kita harus mengambil sebuah resiko namun resiko tersebut yang dapat **dikelola** (meskipun dapat mengakibatkan kegagalan).

## HOW THINGS STICK

- Untuk menjalankan sebuah kolaborasi yang sehat diantara sebuah kelompok yang beragam perludanya "*storytelling*"
- Ide yang baik dapat ditunjukkan melalui suatu gambar yang tepat
- Playbacks merupakan kesempatan untuk mengungkapkan suatu cerita yang mudah diingat dan berfokus pada manusia untuk membagikan ide- ide, prototipe, strategi, dan hal lainnya.

# CRITICAL DAN ANALYTICAL THINKING

## CASES

1. Apakah ada dokumen yang dapat dipelajari?
2. Meminta pengguna untuk membuat suatu deck
3. Membuat list pertanyaan untuk membuat hipotesa singkat mengenai permasalahan yang dihadapi user
4. Pada saat meeting ajukan banyak pertanyaan yang relevan untuk kebutuhan pengguna

Melaporkan hasil dengan membuat sebuah rangkuman pada suatu laporan sederhana

## DATA SCIENCE AGILE SYSTEM

- Melakukan pengaplikasian *data science* menggunakan metode *agile*
- Bagaimana *data scientist* dapat melakukan *tracking* terhadap proses yang sedang dikerjakan sehingga tidak kehilangan fokus pada tujuan yang sudah ditetapkan
- Dapat mempercepat pekerjaan tanpa meningkatkan beban pekerjaan karena dibuat berdasarkan prioritas pekerjaan