

# Homework

## Data Processing



## Estimasi Waktu Pengerjaan

 **1 - 2 jam**

## Jumlah Soal

 **7 Soal**

## Total Point

 **100 poin**

# Teknis Pengerjaan

1. Pekerjaan dilakukan secara **individu**, dengan **menggunakan template yang disediakan**, download [di sini](#)
2. Soal-soal berupa **pertanyaan bisnis** dan dibagi menjadi beberapa bagian berdasarkan tingkat kesulitannya.
  - ***Introduction to Pandas DataFrame and Transformation***: Total 45 Points (2 Soal Beginner + 1 Soal Intermediate)
  - ***DataFrame Aggregation***: Total 25 Points (2 Soal Intermediate)
  - ***DataFrame Combination***: Total 30 Points (1 Soal Beginner + 1 Soal Advance)
3. Upload hasil pengerjaanmu melalui LMS dengan format nama file sebagai berikut **Nama Lengkap\_Batch\_XX** dalam format .html (cara save dalam format .html [disini](#))

# **Introduction to Pandas DataFrame and Transformation**

(Total 45 Points)

1. Buatlah dataframe secara manual seperti gambar dibawah ini.

	name	age	phone_number	ielts_score
0	fiqry	23	+62813123414	6.5
1	iqbal	21	+6287842464	NaN
2	monica	22	+62813125554	7.5
3	rama	24	+6287834464	6.5
4	johan	26	+62813113414	8.0

**(Beginner: 5 poin)**



**5-10 menit**



2. Lakukan import `application_processed.csv` data kedalam notebook (**5pts**) dan carilah beberapa informasi pada dataset `application_processed` (**5pts**) dan jawablah beberapa pertanyaan sebagai berikut:

- a. Column apa saja yang memiliki nilai **NaN** / **None**?
- b. Berapa `rata-rata` pada column `amount_income_total`?
- c. Berapa `median` pada column `amount_income_total`?

### Hint:

- Menggunakan operasi dasar DataFrame pada topik Introduction Pandas DataFrame and Transformation

**(Beginner: 10 poin)**



**5-10 menit**

3. Dengan menggunakan dataset **application\_processed.csv** buatlah dataframe baru dengan ketentuan sebagai berikut:

1. Ambil semua users yang memiliki `email` . (ditandai dengan nilai pada column `flag_email` = 1) dan tidak memiliki data `duplicates` . (5pts)
2. Buatlah column baru bernama `has_car_and_property` dimana nilainya merupakan range dari 0-1 . 1 apabila `flag_own_car` bernilai 1 dan `flag_own_property` bernilai 1 . 0 apabila terdapat atau semua nilai pada column `flag_own_car` dan `flag_own_property` terdapat angka 0. (5pts)
3. Ambil column yang diperlukan yaitu `id` , `gender` , `has_car_and_property` , `income_type` , `education_type` , `family_status` , `housing_type` , `occupation_type` , `amount_income_total` . (5pts)
4. Lakukan filter dengan ketentuan sebagai berikut: (5pts)
  - Memiliki `car` dan `property`
  - Tidak ada nilai null pada column `occupation_type`
  - Memiliki value pada column `occupation_type` dengan pola `staff` .
5. Sudah dilakukan sorting berdasarkan `income_type` dari A-Z dan `amount_income_total` dari besar ke kecil . (5pts)
6. Simpan kedalam bentuk `.csv` dengan nama `homework_3_result.csv` . (5pts)



# DataFrame Aggregation

(Total 25 Points)



4. Dengan menggunakan data `titanic.csv`, buatlah DataFrame untuk melihat informasi `max` dan `min` pada column `Age` dan juga informasi `mean` dan `median` pada column `Fare` pada masing-masing `PassengerClass` atau `Pclass`.

**Expected result:**

	Age		Fare	
	max	min	mean	median
Pclass				
1	80.0	0.92	84.154687	60.2875
2	70.0	0.67	20.662183	14.2500
3	74.0	0.42	13.675550	8.0500

**(Intermediate: 10 poin)**



**10-20 menit**

5. Buatlah DataFrame baru menggunakan fungsi pivot table untuk melihat **Embarked** apa saja yang **average\_female\_ticket\_price** lebih besar daripada **average\_male\_ticket\_price**. Informasi **average\_female\_ticket\_price** dan **average\_male\_ticket\_price** didapatkan dari rata-rata pada column **Fare**.

### Hint:

1. menggunakan pivot table
2. dapat dilihat pada topik **DataFrame Aggregation** bagian **Pivot Table | Reshape Columns and Rows**

### Expected result:

	Embarked	avg_female_ticket_price	avg_male_ticket_price
0	C	75.169805	48.262109
2	S	38.740929	21.711996

(Intermediate: 15 poin)



15-30 menit

# DataFrame Combination

(Total 30 Poin)

6. Buatlah DataFrame seperti pada soal **no.1**. Kemudian tambahkan data tersebut dengan data users yang baru seperti gambar dibawah berikut.

**New Users Data:**

	name	age	phone_number	ielts_score
0	ali	37	None	5.5
1	adit	32	+62152155	6.0

**Expected Result:**

	name	age	phone_number	ielts_score
0	fiqry	23	+62813123414	6.5
1	iqbal	21	+6287842464	NaN
2	monica	22	+62813125554	7.5
3	rama	24	+6287834464	6.5
4	johan	26	+62813113414	8.0
5	ali	37	None	5.5
6	adit	32	+62152155	6.0

**(Beginner: 5 poin)**



**20-30 menit**

7. Buatlah 2 dataframe mengikuti petunjuk berikut:

**df\_1**: Filter dataset **01 telecom\_revenue.csv** dengan multiple kondisi

- MonthlyRevenue > 10
- Occupation terdiri dari Professional, Student, and Crafts (5 Poin)

**df\_2**: Filter dataset **02 telecom\_usage.csv** dengan multiple kondisi

- nilai UnansweredCalls > .BlockedCalls (5 Poin)
- Hitunglah total CustomerID dan rata-rata DroppedCalls untuk masing-masing occupation. (10 Poin)
- pada occupation apa, rata-rata DroppedCalls paling besar? (5 Poin)

**(Advanced: 25 poin)**



**20-30 menit**

# **Selamat Mengerjakan!**