

# PR Modul 4

## Data Analytics

Tim Materi PTI-B 2018/2019

2018-10-15

### Petunjuk

1. Kerjakan modul ini sesuai dengan materi yang diujikan. Gunakan materi yang diujikan (data analytics). Tidak perlu menggunakan materi yang belum diujikan.
2. Buatlah program yang menjawab soal. Untuk tiap soal, tuliskan kembali soalnya dalam komentar, diikuti kode yang menjawab soal tersebut, diikuti jawaban untuk data yang diberikan dalam komentar. Contoh:

Matlab

```
M = readtable('data.csv');

% Banyaknya data
fprintf('%d\n', height(M))
% 100000

% lanjutkan nomor selanjutnya
```

Python

```
df = pd.read_csv("data.csv")

# Banyaknya data
print(len(df))
# 100000

# lanjutkan nomor selanjutnya
```

3. Perhatikan penamaan file terutama untuk ekstensi file (\*.m, \*.py). File tanpa ekstensi beresiko tidak dapat dibuka oleh asisten sehingga kode program tidak dapat dikoreksi (nilai 0)
4. Pastikan program lulus compile dan dapat dijalankan.
5. Untuk setiap file source code program berikan identitas, minimum:

Matlab

```
% NIM>Nama :
% Tanggal :
% Deskripsi :
```

Python

```
# NIM>Nama :
# Tanggal :
# Deskripsi :
```

6. Seluruh file kode program di-compress dengan nama **H04-NIM.zip** sebelum dikumpulkan. Tidak perlu mengumpulkan data yang diberikan.
7. Penulisan kode sebaiknya menggunakan indentasi yang baik dan menambahkan komentar (kegunaan sebuah variabel, percabangan, pengulangan, fungsi dan prosedur) sehingga mempermudah proses pencarian kesalahan pada program (debugging)
8. Kecurangan berupa copy-paste kode program dari peserta atau sumber lain akan memperoleh sanksi tegas.
9. Jika ada perbedaan antara instruksi di sini dan instruksi asisten, ikuti instruksi asisten.
10. Selamat Mengerjakan!

## Problem 1

Simpan dengan nama file: **H04-NIM-01.\***.

Anda diberi data nilai mahasiswa. Tuliskan:

1. Banyaknya data (poin 0)
2. Nilai matematika, fisika, dan kimia dari mahasiswa bernama Tuan Yon. (poin 10).
3. Banyaknya mahasiswa dengan nilai matematika di atas 80 (poin 10).
4. Banyaknya mahasiswa yang mendapat nilai kurang dari 40 di matkul apapun (poin 10).
5. Banyaknya mahasiswa yang mendapat nilai terendah di fisika (poin 10).
6. Data 10 besar peserta peraih nilai tertinggi di fisika. Jika ada yang nilainya sama, ambil mahasiswa dengan nama lebih kecil (poin 10).
7. Tabel frekuensi masing-masing fakultas (poin 10).
8. Rata-rata dari nilai matematika semua mahasiswa (poin 10).
9. Standar deviasi dari nilai fisika semua mahasiswa (poin 10).
10. Rata-rata dari nilai kimia STEI (poin 10).
11. Dengan nilai manakah (matematika / kimia) nilai fisika berkorelasi? Berapa koefisien korelasinya? (poin 10).

Nomor 1 sudah dijawab untuk Anda, sesuai kode berikut (perhatikan kalau jumlah data mungkin saja berbeda dengan milik Anda):

Matlab

```
M = readtable('data.csv');  
  
% Banyaknya data  
fprintf('%d\n', height(M))  
% 100000  
  
% lanjutkan nomor selanjutnya
```

Python

```
df = pd.read_csv("data.csv")  
  
# Banyaknya data  
print(len(df))  
# 100000  
  
# lanjutkan nomor selanjutnya
```

## Problem 2

Simpan dengan nama file: **H04-NIM-02.\***.

Anda diberi data penerbangan. Tentukan:

1. Banyaknya data (poin 0)
2. Data penerbangan yang memiliki harga 1.999.999 (poin 10).
3. Banyaknya penerbangan dengan harga di atas 3.500.000 (poin 10).
4. Banyaknya penerbangan di tahun 2014 dengan asal atau tujuan bukan kota PDG (poin 10).
5. Data penerbangan maskapai macan dengan harga minimum (poin 10).
6. Data 10 penerbangan rajawali dengan penumpang terbanyak. Jika ada penerbangan yang sama banyak, urutkan berdasar harga dari yang termurah (poin 10).
7. Tabel frekuensi penerbangan country link masing-masing tahun (poin 10).
8. Rata-rata harga seluruh penerbangan (poin 10).
9. Standar deviasi jumlah penumpang maskapai mataram (poin 10).
10. Rata-rata dari harga maskapai country link (poin 10).
11. Koefisien korelasi dari jumlah penumpang dengan harga (poin 10).

Nomor 1 sudah dijawab untuk Anda, sesuai kode berikut (perhatikan kalau jumlah data mungkin saja berbeda dengan milik Anda):

Matlab

```
M = readtable('data.csv');  
  
% Banyaknya data  
fprintf('%d\n', height(M))  
% 100000  
  
% lanjutkan nomor selanjutnya
```

Python

```
df = pd.read_csv("data.csv")  
  
# Banyaknya data  
print(len(df))  
# 100000  
  
# lanjutkan nomor selanjutnya
```

## Problem 3

Simpan dengan nama file: **H04-NIM-03.\***.

Anda diberi data karyawan. Tentukan:

1. Banyaknya data (poin 0)
2. Data karyawan bernama Tuan Yon (poin 10).
3. Banyaknya karyawan dengan usia kurang dari atau sama dengan 30 (poin 10).
4. Banyaknya karyawan personalia dengan gaji di luar rentang 4 - 5 juta (poin 10).
5. Data karyawan produksi dengan gaji maksimum (poin 10).
6. Data 10 karyawan teknologi dengan gaji terbanyak. Jika ada gaji yang sama banyak, urutkan dari usia yang termuda (poin 10).
7. Tabel frekuensi banyaknya karyawan tiap golongan (poin 10).
8. Tabel frekuensi banyaknya karyawan tiap departemen yang masuk sebelum tahun 2015 (poin 10).
9. Rata-rata gaji seluruh karyawan (poin 10).
10. Standar deviasi usia karyawan operasional (poin 10).
11. Dengan apakah gaji berkorelasi? Usia, tahun masuk, atau golongan? Tuliskan koefisien korelasinya (poin 10).

Nomor 1 sudah dijawab untuk Anda, sesuai kode berikut (perhatikan kalau jumlah data mungkin saja berbeda dengan milik Anda):

Matlab

```
M = readtable('data.csv');  
  
% Banyaknya data  
fprintf('%d\n', height(M))  
% 100000  
  
% lanjutkan nomor selanjutnya
```

Python

```
df = pd.read_csv("data.csv")  
  
# Banyaknya data  
print(len(df))  
# 100000  
  
# lanjutkan nomor selanjutnya
```