

MODUL PRAKTIK DATA PREPARATION

DATA BINNING DAN GROUP BY

SIMPAN HASIL PRAKTIK KALIAN PADA GOOGLE COLAB DENGAN NAMA “pertemuan7_binning”

1. Buat data frame yang memuat nilai random integer dengan perintah

```
import pandas as pd
import numpy as np

np.random.seed(100)
age = pd.DataFrame({'Age':np.random.randint(0,100,100)})
```

2. Tampilkan data frame tersebut
3. Tampilkan 10 data teratas kemudian tampilkan juga data terbawah
4. Kategorikan data pada data frame diatas dengan perintah

```
a = pd.cut(age['Age'],[0,2,11,20,61,101])
a.head(10)
```

5. Berilah label untuk data “Age” yang tersedia

```
b=pd.cut(age['Age'],[0,2,11,20,61,101], labels = ['Bayi','Anak-anak','Remaja','Dewasa','Manula'])
b.head(10)
```

6. Buatlah grafik kategori umur dengan perintah

```
import matplotlib.pyplot as plt
age.plot(kind='hist', bins =[0,2,11,20,61,101], figsize =(8,5))
plt.grid()
```

7. Buat 1 kolom untuk menaruh data kategori

```
age['range'] = pd.cut(age['Age'], [0, 2, 11, 20, 61, 101])
age['group'] = pd.cut(age['Age'], [0, 2, 11, 20, 61, 101],
                      labels = ['bayi', 'anak-anak', 'Remaja', 'Dewasa', 'Lanjut Usia'])
```

8. Tampilkan data teratas hasil penambahan kolom
9. Hitunglah berapa jumlah data pada masing-masing kategori

SIMPAN HASIL PRAKTIK KALIAN PADA GOOGLE COLAB DENGAN NAMA “pertemuan7_groupby”

1. Baca file csv dengan nama “data.csv”

```
import pandas as pd
df = pd.read_csv('data.csv')
```

2. Tampilkan dataframe dari file csv di atas
3. Gunakan modul filterwarnings pada package warnings untuk tidak menampilkan warning untuk perintah di poin selanjutnya
4. Mencari rata-rata data berdasarkan alamat.

```
alamat = df.groupby('alamat').mean()
print(alamat)
```

5. Carilah rata-rata data berdasarkan jenis kelamin
6. Carilah total nilai berdasarkan alamat
7. Carilah total nilai berdasarkan jenis kelamin
8. Carilah jumlah data berdasarkan alamat
9. Carilah jumlah data berdasarkan jenis kelamin
10. Carilah nilai minimal dan maksimal dari keseluruhan data
11. Carilah nilai minimal dan maksimal berdasarkan alamat
12. Carilah nilai minimal dan maksimal berdasarkan jenis kelamin
13. Menggabungkan fungsi aggregation dengan group by

```
import numpy as np
jenis_kelamin = df.groupby('jenis kelamin').agg([np.mean, np.sum])
print(jenis_kelamin)
```