Mhd Iqbal Pratama

105220042

- 1. Pertama kita harus mengecek apakah terdapat missing value, lalu setelah mengecek missing value yang ada pada kolom, kita menentukan tipe data dari kolom tersebut. Jika tipe data kolom tersebut kategori kita perlu melakukan encoder terlebih dahulu, setelah itu kita mengisi missing value dengan cara iterative imputer.
 - a. Import Library dan Membaca file

```
import pandas as pd
from sklearn.experimental import enable iterative imputer
from sklearn.impute import IterativeImputer
from sklearn.preprocessing import OneHotEncoder

110] 

0.1s

Python

df = pd.read_csv(r'C:\Users\LENGVO\Documents\PsdTask\volume-pengangkutan-sampah-di-kali-sungai-situ-waduk-bulan-janua
110] 

0.6s

Python
```

b. Melihat 5 data awal



c. Mengecek missing value

```
df.info()
✓ 0.1s
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 54343 entries, 0 to 54342
Data columns (total 8 columns):
     Column
                               Non-Null Count Dtype
                               54343 non-null int64
0
    bulan
    titik_lokasi
1
                               54343 non-null object
 2
    kecamatan
                               54343 non-null object
3
   wilayah
                               54343 non-null object
    panjang/luas
                               53475 non-null object
4
   satuan_panjang/luas
5
                               32388 non-null object
6
    tanggal
                               54343 non-null int64
    volume_sampah_perhari(m3)
                               54340 non-null object
7
dtypes: int64(2), object(6)
memory usage: 3.3+ MB
```

d. Mengedrop kolom yang tidak memiliki missing value

```
#drop beberapa fitur yang tidak ada missing value : titik_lokasi, kecamatan, dan bulan, wilayah dan tanggal kerna tid

df2 = df.drop(['titik_lokasi','kecamatan','bulan','wilayah', 'tanggal'], axis='columns')
    print(df2)
√ 0.41
       panjang/luas satuan_panjang/luas volume_sampah_perhari(m3)
                  464
54338
                   888
54339
                  788
54340
                  1000
54341
                 1666
54342
                   488
[54343 rows x 3 columns]
```

e. Mengecek missing value menggunakan fungsi

```
# cek missing value dengan fungsi
   def cek_null(df):
       col_na = df.isnull().sum().sort_values(ascending=False)
       percent = col_na/ len(df)*180
       missing_data = pd.concat((col_na, percent), axis=1, keys={'Total Missing Value', 'Percent'])
       print(missing_data[missing_data['Total Missing Value']>0])
   cek_null(df2)
                           Total Missing Value
                                                  Percent
                                         21955 40.400788
satuan_panjang/luas
panjang/luas
                                           868
                                                1.597262
volume_sampah_perhari(m3)
                                                 0.005520
```

f. Melakukan encoder

```
df2['satuan_panjang/luas']=df2['satuan_panjang/luas'].astype('category')
   df2['satuan_panjang/luas']=df2['satuan_panjang/luas'].cat.codes
   enc = OneHotEncoder()
   enc_data = pd.DataFrame(enc.fit_transform(df2[['satuan_panjang/luas']]).toarray())
   df3 = df2.join(enc_data)
   print(df3)
      panjang/luas
                    satuan_panjang/luas volume_sampah_perhari(m3)
0
               464
                                     10
                                                                   0.0
                                                                        0.0
               464
                                     10
                                                                   0.0
                                                                         0.0
2
               464
                                     10
               606
                                     10
                                                                   0.0
                                                                         0.0
4
               310
                                     10
                                                                   0.0
                                                                         8.8
54338
               800
                                     10
                                                                   0.0
                                                                        0.0
54339
               700
                                     10
                                                                   0.0
                                                                        0.0
54340
              1000
                                     10
                                                                    0.0
54341
                                     10
              1000
                                                                   0.0
                                                                        8.8
54342
               400
                                     10
                                                                2 0.0 0.0
                                            9 10 11
```

g. Mengisi missing value menggunakan iterative imputer

```
imputer = IterativeInputer(random_state = 42)
inputed = imputer.fit_transform(df2)
df_imputed = pd.DataFrame(imputed, columns=df2.columns)
```

- 2. Data tersebut memiliki outliers karena berdasarkan grafik scatter terdapat data yang tidak terdistribusi secara tidak normal.
 - a. Import library dan file

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

/ 2.5s

Off = pd.read_csv(r'C:\Users\LENOVO\Documents\PadTask\All_Diets.csv')

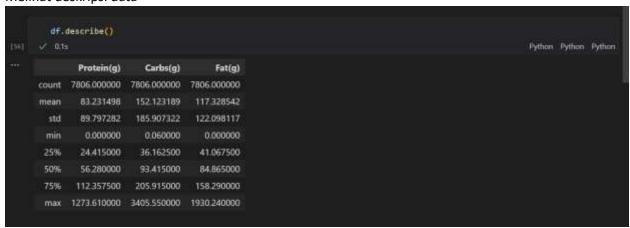
/ 0.1s

+ Code + Markdown
```

b. Melihat 5 data pertama



c. Melihat deskripsi data



d. Mengecek Null data

e. Scatting

