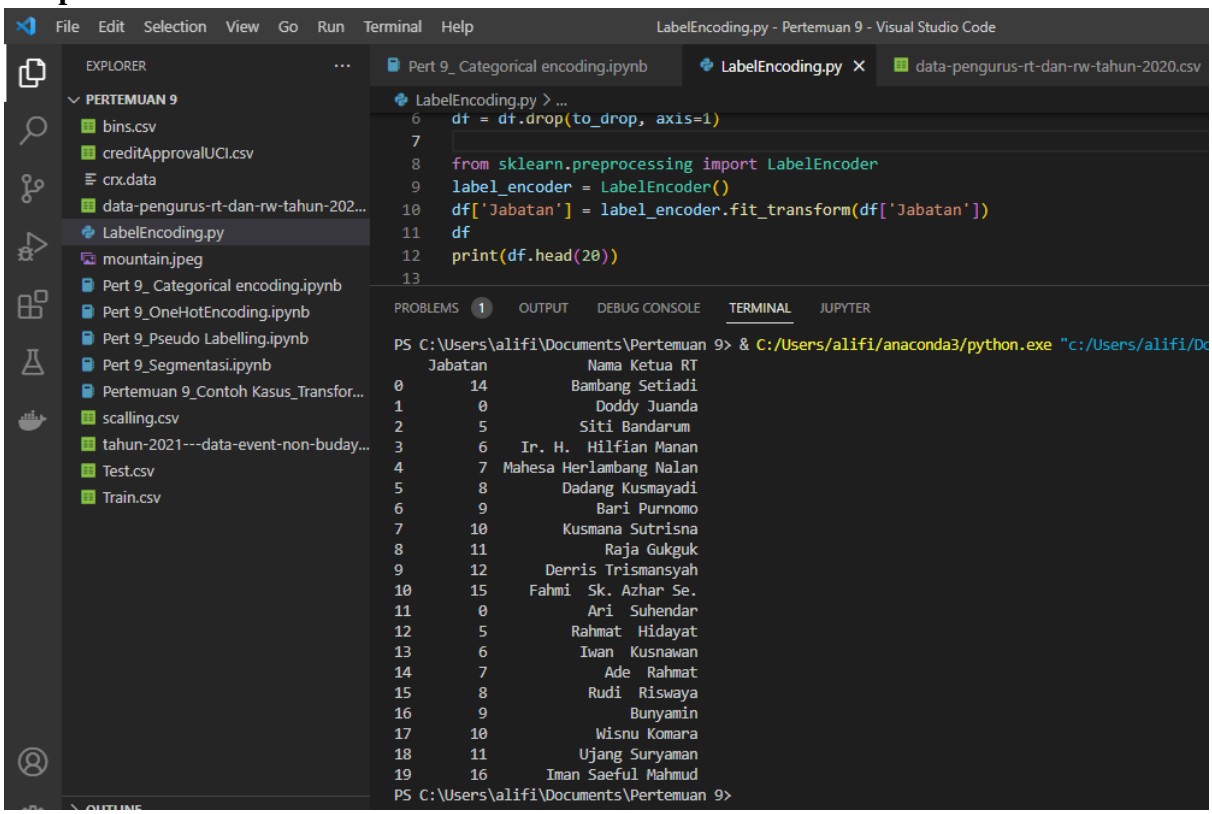
Tugas 5 Data Sains dan Analisis

1. Categorical Encoding (Label Encoding)
   Dari dataset, kolom jabatan diurutkan secara alfabet dan dipresentasikan dengan suatu nilai integer. Misalnya dalam outputnya kategori 0 = Ketua RT 01, dan seterusnya yang otomatis diurutkan alfabetikal.

   Source Code

```
import pandas as pd

dataset = "data-pengurus-rt-dan-rw-tahun-2020.csv"
df = pd.read_csv(dataset, encoding='cp1252')
to_drop = ['Kode Kecamatan','Kecamatan', 'Kode Kelurahan', 'Kelurahan']
df = df.drop(to_drop, axis=1)

from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
df['Jabatan'] = label_encoder.fit_transform(df['Jabatan'])
df

print(df.head(20))
```

   Output



2. Categorical Encoding (One-Hot Encoding)
   Dari dataset, kolom jabatan diurutkan secara alfabet dan diurutkan dalam bentuk bits atau setiap label dipetakan ke vector biner.

Souce Code

```
import pandas as pd

dataset = "data-pengurus-rt-dan-rw-tahun-2020.csv"
df = pd.read_csv(dataset, encoding='cp1252')
to_drop = ['Kode Kecamatan','Kecamatan', 'Kode Kelurahan', 'Kelurahan']
df = df.drop(to_drop, axis=1)

X = df['Jabatan'].values.reshape(-1,1)

from sklearn.preprocessing import OneHotEncoder
onehot_encoder = OneHotEncoder()
X = onehot_encoder.fit_transform(X).toarray()
print("One Hot Encoder")
print(X)

print(onehot_encoder.categories_)

df_onehot= pd.DataFrame(X, columns=[str(i) for i in range(X.shape[1])])
print(df_onehot.head(20))

df=pd.concat([df_onehot,df], axis=1)
print("menggabungkan dataframe")
print(df.head(20))

df = df.drop(['Jabatan'], axis=1)
print("Buang Atribut Country karena Direpresentasikan dengan
OneHotEncoder")
print(df.head(20))
```

Output

```
One Hot Encoder
[[0. 0. 0. ... 0. 0. 0.]
 [1. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]
[array(['Ketua RT 01', 'Ketua RT 010', 'Ketua RT 011', 'Ketua RT 012',
       'Ketua RT 013', 'Ketua RT 02', 'Ketua RT 03', 'Ketua RT 04',
       'Ketua RT 05', 'Ketua RT 06', 'Ketua RT 07', 'Ketua RT 08',
       'Ketua RT 09', 'Ketua RT 10', 'Ketua RW 01', 'Ketua RW 02',
       'Ketua RW 03', 'Ketua RW 04', 'Ketua RW 05', 'Ketua RW 06',
       'Ketua RW 07', 'Ketua RW 08', 'Ketua RW 09', 'Ketua RW 10',
       'Ketua RW 11', 'Ketua RW 12', 'Ketua RW 13', 'Ketua RW 14',
       'Ketua RW 15', 'Ketua RW 16', 'Ketua RW 17', 'Ketua RW 18',
       'Ketua RW 19', 'Ketua RW 20', 'Ketua RW 21', 'Ketua RW 6'],
      dtype=object)]
```

Pert 9_ Categorical encoding.ipynb | OneHot.py | data-pengurus-rt-dan-rw-tahun-2020.csv

OneHot.py

```
onehot_encoder = OneHotEncoder()
X = onehot_encoder.fit_transform(X).toarray()
print("One Hot Encoder")
print(X)
```

PROBLEMS 1 | OUTPUT | DEBUG CONSOLE | TERMINAL | JUPYTER

```
    0   1   2   3   4   5   6   7   8   9   10  11  12  13  14  ...  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35
0   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
1   1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
2   0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
3   0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
4   0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
5   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
6   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
7   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
8   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
9   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
10  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
11  1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
12  0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
13  0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
14  0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
15  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
16  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
17  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
18  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
19  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

[20 rows x 36 columns]
```

menggabungkan dataframe

```
    0   1   2   3   4   5   6   7   8   9   10  11  ...  26  27  28  29  30  31  32  33  34  35      Jabatan          Nama Ketua RT
0   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RW 01        Bambang Setiadi
1   1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 01           Doddy Juanda
2   0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 02           Siti Bandarum
3   0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 03    Ir. H.  Hilfian Manan
4   0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 04  Mahesa Herlambang Nalan
5   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 05       Dadang Kusmayadi
6   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 06          Bari Purnomo
7   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 07       Kusmana Sutrisna
8   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 08          Raja Gukguk
9   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 09     Derris Trismansyah
10  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RW 02   Fahmi  Sk. Azhar Se.
11  1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 01        Ari  Suhendar
12  0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 02     Rahmat  Hidayat
13  0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 03       Iwan  Kusnawan
14  0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 04        Ade  Rahmat
15  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 05       Rudi  Riswaya
16  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 06           Bunyamin
17  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 07       Wisnu Komara
18  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RT 08       Ujang Suryaman
19  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Ketua RW 03   Iman Saeful Mahmud

[20 rows x 38 columns]
```

Buang Atribut Country karena Direpresentasikan dengan OneHotEncoder

```
    0   1   2   3   4   5   6   7   8   9   10  11  12  ...  24  25  26  27  28  29  30  31  32  33  34  35          Nama Ketua RT
0   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0       Bambang Setiadi
1   1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0          Doddy Juanda
2   0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0          Siti Bandarum
3   0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0   Ir. H.  Hilfian Manan
4   0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 Mahesa Herlambang Nalan
5   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0      Dadang Kusmayadi
6   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0         Bari Purnomo
7   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0      Kusmana Sutrisna
8   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0         Raja Gukguk
9   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0    Derris Trismansyah
10  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Fahmi  Sk. Azhar Se.
11  1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0       Ari  Suhendar
12  0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0    Rahmat  Hidayat
13  0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0      Iwan  Kusnawan
14  0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0       Ade  Rahmat
15  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0      Rudi  Riswaya
16  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0          Bunyamin
17  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0      Wisnu Komara
18  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0      Ujang Suryaman
19  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ...  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  Iman Saeful Mahmud

[20 rows x 37 columns]
PS C:\Users\alifi\Documents\Pertemuan 9>
```

3. Segmentasi
   Dalam melakukan segmentasi gambar dapat dilakukan dengan berbagai pendekatan, manipulasi nilai pixel dengan menggunakan threshold, maupun dengan menerapkan 'filter' pengolahan citra. Pada hands-on ini Anda akan melakukan proses segmentasi.

   Source Code

```
#Import Library
from skimage.color import rgb2gray
import numpy as np
import matplotlib.pyplot as plt
from scipy import ndimage
from PIL import Image
from sklearn.cluster import KMeans
from skimage.filters import sobel
import skimage.segmentation
import skimage
import warnings
warnings.filterwarnings("ignore")
```

   Tampilkan gambar dengan format jpg dan lakukan resize gambar untuk memperkecil size gambar

```
image=Image.open('62f64cf93c182.jpg')
image=image.resize((320,225))
image=np.array(image)
plt.imshow(image)
```

   Output

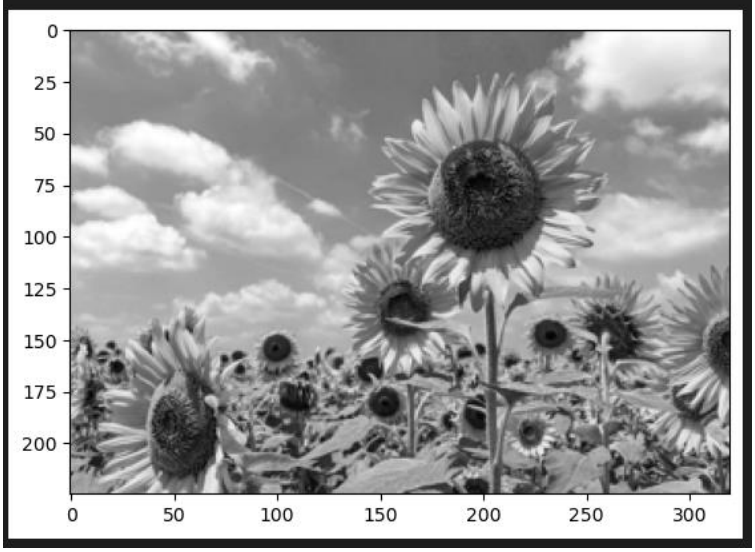
```
<matplotlib.image.AxesImage at 0x20123fa31c0>
```

Source Code

```
#Membuat gambar menjadi skala abu-abu.
gray = rgb2gray(image)
plt.imshow(gray, cmap='gray')
```

Output

```
<matplotlib.image.AxesImage at 0x2012400c4f0>
```

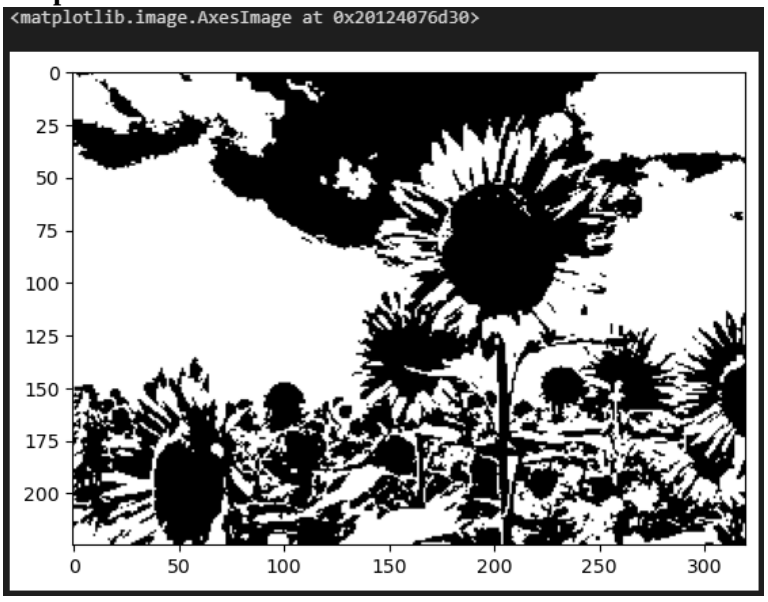


Source Code

```
#Segmentasi obyek menjadi 2 bagian yang berbeda berdasarkan nilai
threshold yang ditentukan
arr=gray.flatten()
for i in range(len(arr)):
    if arr[i]>=arr.mean() :
        arr[i]=1
    else:
        arr[i]=0
gray_segmented=arr.reshape(gray.shape[0],gray.shape[1])

plt.imshow(gray_segmented,cmap='gray')
```

Output

```
<matplotlib.image.AxesImage at 0x20124076d30>
```



Source Code

```
#Segmentasi obyek menjadi 5 bagian yang berbeda berdasarkan nilai
threshold yang ditentukan
arr=gray.flatten()
for i in range(len(arr)):
    if arr[i]>=arr.mean():
        arr[i]=4
    elif arr[i]>=0.75:
        arr[i]=3
    elif arr[i]>0.5 :
        arr[i]=2
    elif arr[i]>0.25:
        arr[i]=1
    else:
        arr[i]=0
gray_segmented_2=arr.reshape(gray.shape[0],gray.shape[1])

#There are 5 segments in the below image :)
plt.figure(figsize=(18,8))
plt.imshow(gray_segmented_2,cmap='pink')
plt.axis("off")
plt.show()
```

Output

Source Code

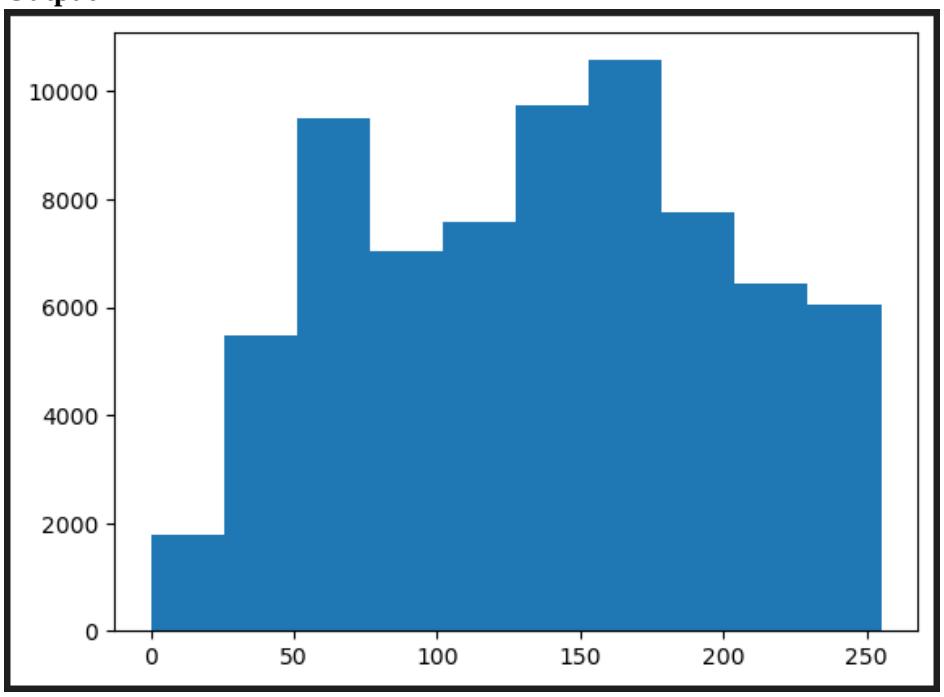
```
#Memanfaatkan informasi dari nilai histogram
imm=image[:,:,0]
elevation_map = sobel(imm)

fig, ax = plt.subplots(figsize=(18,8))
ax.imshow(elevation_map, cmap='gray', interpolation='nearest')
ax.axis('off')
ax.set_title('elevation_map')
plt.show()
```

Output

Source Code

```
plt.hist(imm.flatten())
plt.show()
```

Output

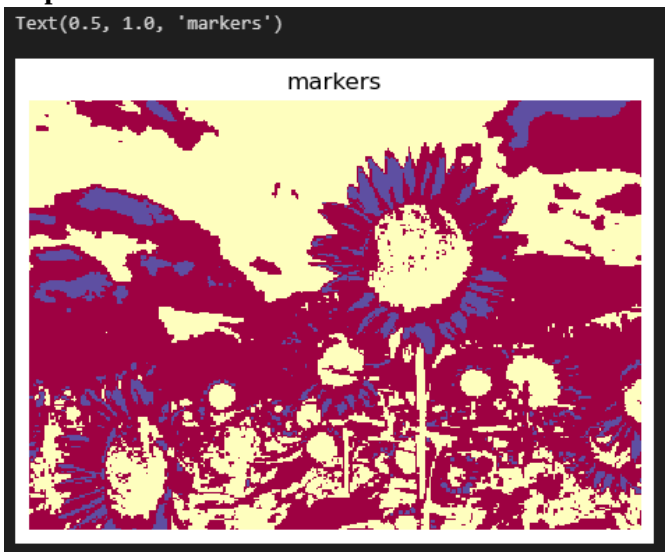


Source Code

```
#Melakukan pelabelan terhadap pixel berdasarkan nilai histogram
markers = np.zeros_like(imm)
markers[imm < 117] = 1
markers[imm > 232] = 2

fig, ax = plt.subplots(figsize=(8,4))
ax.imshow(markers, cmap='Spectral', interpolation='nearest')
ax.axis('off')
ax.set_title('markers')
```

Ouput

```
Text(0.5, 1.0, 'markers')
```

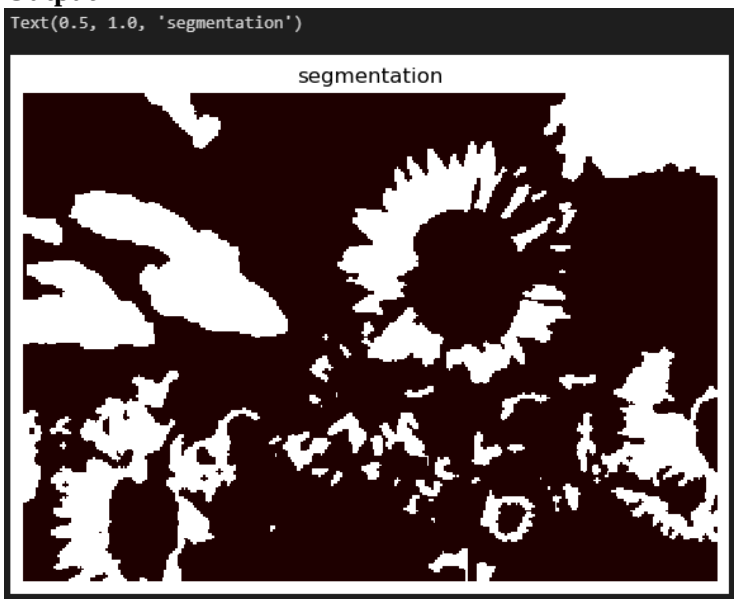


Source Code

```
segmentation = skimage.segmentation.watershed(elevation_map, markers)

fig, ax = plt.subplots(figsize=(10,5))
ax.imshow(segmentation, cmap='pink',interpolation='nearest')
ax.axis('off')
ax.set_title('segmentation')
```

Output

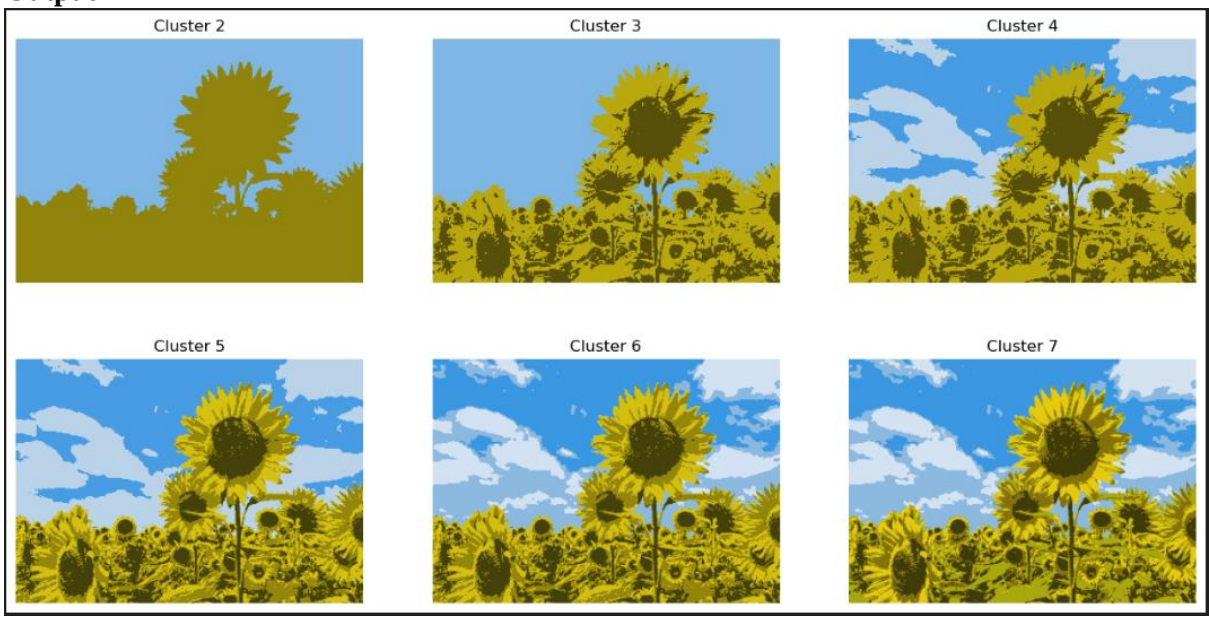
```
Text(0.5, 1.0, 'segmentation')
```



Source Code

```
#Segmentasi Dengan Memanfaatkan Nilai Cluster dari Pixel
im=image/255
pic=im.reshape(im.shape[0]*im.shape[1],im.shape[2])

fig, ax = plt.subplots(2, 3, figsize=(16, 8))
count=1
for i in range(2):
    for j in range(3):

        kmeans = KMeans(n_clusters=count+1, random_state=0).fit(pic)
        pic_print = kmeans.cluster_centers_[kmeans.labels_]
        clustered_pic=pic_print.reshape(im.shape[0],im.shape[1],im.shape[2
])
        count+=1
        ax[i][j].set_title('Cluster '+str(count))
        ax[i][j].imshow(clustered_pic)
        ax[i][j].set_axis_off()
plt.show()
```

Output



4. Transformasi Data (Imputasi Mean)
   Teknik ini mengganti nilai atau data yang hilang (NaN) dengan nilai mean (rata-rata). Dalam dataset terdapat data hilang (NaN) pada data ke 6 kolom LoftFrontage, diganti dengan nilai mean pada output dataset kedua.

Source Code

```
import pandas as pd
import numpy as np

dataset = "test123.csv"
df = pd.read_csv(dataset)
df = pd.DataFrame(df)
print(df.head(10))

df = df.fillna(df.mean())
print(df.head(10))
```

Output

```
PS C:\Users\alifi\Documents\Pertemuan 9> C:/Users/alifi/anaconda3/Scripts/activate
0  1461          20       RH         80.0    11622   Pave   NaN      Reg  ...    NaN  MnPrv         NaN       0      6   2010       WD        Normal
1  1462          20       RL         81.0    14267   Pave   NaN      IR1  ...    NaN    NaN        Gar2   12500      6   2010       WD        Normal
2  1463          60       RL         74.0    13830   Pave   NaN      IR1  ...    NaN  MnPrv         NaN       0      3   2010       WD        Normal
3  1464          60       RL         78.0     9978   Pave   NaN      IR1  ...    NaN    NaN         NaN       0      6   2010       WD        Normal
4  1465         120       RL         43.0     5005   Pave   NaN      IR1  ...    NaN    NaN         NaN       0      1   2010       WD        Normal
5  1466          60       RL         75.0    10000   Pave   NaN      IR1  ...    NaN    NaN         NaN       0      4   2010       WD        Normal
6  1467          20       RL          NaN     7980   Pave   NaN      IR1  ...    NaN  GdPrv        Shed     500      3   2010       WD        Normal
7  1468          60       RL         63.0     8402   Pave   NaN      IR1  ...    NaN    NaN         NaN       0      5   2010       WD        Normal
8  1469          20       RL         85.0    10176   Pave   NaN      Reg  ...    NaN    NaN         NaN       0      2   2010       WD        Normal
9  1470          20       RL         70.0     8400   Pave   NaN      Reg  ...    NaN  MnPrv         NaN       0      4   2010       WD        Normal

[10 rows x 80 columns]
c:\Users\alifi\Documents\Pertemuan 9\TransformasiData.py:9: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is
deprecated; in a future version this will raise TypeError.  Select only valid columns before calling the reduction.
  df = df.fillna(df.mean())
     Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  ... PoolQC  Fence MiscFeature MiscVal MoSold YrSold SaleType SaleCondition
0  1461          20       RH    80.000000    11622   Pave   NaN      Reg  ...    NaN  MnPrv         NaN       0      6   2010       WD        Normal
1  1462          20       RL    81.000000    14267   Pave   NaN      IR1  ...    NaN    NaN        Gar2   12500      6   2010       WD        Normal
2  1463          60       RL    74.000000    13830   Pave   NaN      IR1  ...    NaN  MnPrv         NaN       0      3   2010       WD        Normal
3  1464          60       RL    78.000000     9978   Pave   NaN      IR1  ...    NaN    NaN         NaN       0      6   2010       WD        Normal
4  1465         120       RL    43.000000     5005   Pave   NaN      IR1  ...    NaN    NaN         NaN       0      1   2010       WD        Normal
5  1466          60       RL    75.000000    10000   Pave   NaN      IR1  ...    NaN    NaN         NaN       0      4   2010       WD        Normal
6  1467          20       RL    68.580357     7980   Pave   NaN      IR1  ...    NaN  GdPrv        Shed     500      3   2010       WD        Normal
7  1468          60       RL    63.000000     8402   Pave   NaN      IR1  ...    NaN    NaN         NaN       0      5   2010       WD        Normal
8  1469          20       RL    85.000000    10176   Pave   NaN      Reg  ...    NaN    NaN         NaN       0      2   2010       WD        Normal
9  1470          20       RL    70.000000     8400   Pave   NaN      Reg  ...    NaN  MnPrv         NaN       0      4   2010       WD        Normal

[10 rows x 80 columns]
PS C:\Users\alifi\Documents\Pertemuan 9>
```

5. Transformasi Data (Imputasi Nilai Suka-suka (Arbitrary)
   Teknik menggantikan data yang hilang atau NaN dengan inputan tipe data numerik. Dalam dataset, dalam kolom LotFrontage terdapat data yang hilang dalam baris 7, 12, 14 dan digantikan dengan nilai 70 sesuai dengan nilai suka-suka yang diberikan.

Source Code

```
import pandas as pd
import numpy as np

dataset = "train123.csv"
df = pd.read_csv(dataset)
df = pd.DataFrame(df)
print(df.head(15))

df = df.fillna(70)
print(df.head(15))
```

Output

```
    Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  ...  Fence MiscFeature MiscVal MoSold YrSold SaleType SaleCondition SalePrice
0    1          60       RL         65.0     8450   Pave   NaN      Reg  ...    NaN         NaN       0      2   2008       WD        Normal    208500
1    2          20       RL         80.0     9600   Pave   NaN      Reg  ...    NaN         NaN       0      5   2007       WD        Normal    181500
2    3          60       RL         68.0    11250   Pave   NaN      IR1  ...    NaN         NaN       0      9   2008       WD        Normal    223500
3    4          70       RL         60.0     9550   Pave   NaN      IR1  ...    NaN         NaN       0      2   2006       WD        Abnorml   140000
6    7          20       RL         75.0    10084   Pave   NaN      Reg  ...    NaN         NaN       0      8   2007       WD        Normal    307000
7    8          60       RL          NaN    10382   Pave   NaN      IR1  ...    NaN        Shed     350     11   2009       WD        Normal    200000
8    9          50       RM         51.0     6120   Pave   NaN      Reg  ...    NaN         NaN       0      4   2008       WD        Abnorml   129900
9   10         190       RL         50.0     7420   Pave   NaN      Reg  ...    NaN         NaN       0      1   2008       WD        Normal    118000
10  11          20       RL         70.0    11200   Pave   NaN      Reg  ...    NaN         NaN       0      2   2008       WD        Normal    129500
11  12          60       RL         85.0    11924   Pave   NaN      IR1  ...    NaN         NaN       0      7   2006      New        Partial   345000
12  13          20       RL          NaN    12968   Pave   NaN      IR2  ...    NaN         NaN       0      9   2008       WD        Normal    144000
13  14          20       RL         91.0    10652   Pave   NaN      IR1  ...    NaN         NaN       0      8   2007      New        Partial   279500
14  15          20       RL          NaN    10920   Pave   NaN      IR1  ...   GdWo         NaN       0      5   2008       WD        Normal    157000

[15 rows x 81 columns]
    Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  ...  Fence MiscFeature MiscVal MoSold YrSold SaleType SaleCondition SalePrice
0    1          60       RL         65.0     8450   Pave    70      Reg  ...     70          70       0      2   2008       WD        Normal    208500
1    2          20       RL         80.0     9600   Pave    70      Reg  ...     70          70       0      5   2007       WD        Normal    181500
2    3          60       RL         68.0    11250   Pave    70      IR1  ...     70          70       0      9   2008       WD        Normal    223500
3    4          70       RL         60.0     9550   Pave    70      IR1  ...     70          70       0      2   2006       WD        Abnorml   140000
4    5          60       RL         84.0    14260   Pave    70      IR1  ...     70          70       0     12   2008       WD        Normal    250000
5    6          50       RL         85.0    14115   Pave    70      IR1  ...  MnPrv        Shed     700     10   2009       WD        Normal    143000
6    7          20       RL         75.0    10084   Pave    70      Reg  ...     70          70       0      8   2007       WD        Normal    307000
7    8          60       RL         70.0    10382   Pave    70      IR1  ...     70        Shed     350     11   2009       WD        Normal    200000
8    9          50       RM         51.0     6120   Pave    70      Reg  ...     70          70       0      4   2008       WD        Abnorml   129900
9   10         190       RL         50.0     7420   Pave    70      Reg  ...     70          70       0      1   2008       WD        Normal    118000
10  11          20       RL         70.0    11200   Pave    70      Reg  ...     70          70       0      2   2008       WD        Normal    129500
11  12          60       RL         85.0    11924   Pave    70      IR1  ...     70          70       0      7   2006      New        Partial   345000
12  13          20       RL         70.0    12968   Pave    70      IR2  ...     70          70       0      9   2008       WD        Normal    144000
13  14          20       RL         91.0    10652   Pave    70      IR1  ...     70          70       0      8   2007      New        Partial   279500
14  15          20       RL         70.0    10920   Pave    70      IR1  ...   GdWo          70       0      5   2008       WD        Normal    157000
```

6. Transformasi Data (Imputasi Frequent Category atau Modus)
   Teknik untuk menggantikan nilai atau data yang hilang atau dalam bentuk NaN dan digunakan bagi tipe data kategori. Dalam output terlihat data dalam kolom yang ada data NaN digantikan dengan data yang ada dalam satu kategori dan digantikan dengan nilai modus.

Source Code

```
import pandas as pd
import numpy as np
from sklearn.impute import SimpleImputer

dataset = "train123.csv"
df = pd.read_csv(dataset)
df = pd.DataFrame(df)
print(df.head(15))

imp = SimpleImputer(strategy='most_frequent')
df = pd.DataFrame(imp.fit_transform(df), columns=df.columns)

print(df.head(15))
```

Output

```
    Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  ...  Fence MiscFeature MiscVal MoSold YrSold SaleType SaleCondition SalePric
0    1          60       RL         65.0     8450   Pave   NaN      Reg  ...    NaN         NaN       0      2   2008       WD        Normal    20850
1    2          20       RL         80.0     9600   Pave   NaN      Reg  ...    NaN         NaN       0      5   2007       WD        Normal    18150
2    3          60       RL         68.0    11250   Pave   NaN      IR1  ...    NaN         NaN       0      9   2008       WD        Normal    22350
3    4          70       RL         60.0     9550   Pave   NaN      IR1  ...    NaN         NaN       0      2   2006       WD       Abnorml    14000
4    5          60       RL         84.0    14260   Pave   NaN      IR1  ...    NaN         NaN       0     12   2008       WD        Normal    25000
5    6          50       RL         85.0    14115   Pave   NaN      IR1  ...  MnPrv        Shed     700     10   2009       WD        Normal    14300
6    7          20       RL         75.0    10084   Pave   NaN      Reg  ...    NaN         NaN       0      8   2007       WD        Normal    30700
7    8          60       RL          NaN    10382   Pave   NaN      IR1  ...    NaN        Shed     350     11   2009       WD        Normal    20000
8    9          50       RM         51.0     6120   Pave   NaN      Reg  ...    NaN         NaN       0      4   2008       WD       Abnorml    12990
9   10         190       RL         50.0     7420   Pave   NaN      Reg  ...    NaN         NaN       0      1   2008       WD        Normal    11800
10  11          20       RL         70.0    11200   Pave   NaN      Reg  ...    NaN         NaN       0      2   2008       WD        Normal    12950
11  12          60       RL         85.0    11924   Pave   NaN      IR1  ...    NaN         NaN       0      7   2006      New       Partial    34500
12  13          20       RL          NaN    12968   Pave   NaN      IR2  ...    NaN         NaN       0      9   2008       WD        Normal    14400
13  14          20       RL         91.0    10652   Pave   NaN      IR1  ...    NaN         NaN       0      8   2007      New       Partial    27950
14  15          20       RL          NaN    10920   Pave   NaN      IR1  ...   GdWo         NaN       0      5   2008       WD        Normal    15700

[15 rows x 81 columns]
    Id MSSubClass MSZoning LotFrontage LotArea Street Alley LotShape  ...  Fence MiscFeature MiscVal MoSold YrSold SaleType SaleCondition SalePrice
0    1         60       RL        65.0    8450   Pave  Grvl      Reg  ...  MnPrv        Shed       0      2   2008       WD        Normal    208500
1    2         20       RL        80.0    9600   Pave  Grvl      Reg  ...  MnPrv        Shed       0      5   2007       WD        Normal    181500
2    3         60       RL        68.0   11250   Pave  Grvl      IR1  ...  MnPrv        Shed       0      9   2008       WD        Normal    223500
3    4         70       RL        60.0    9550   Pave  Grvl      IR1  ...  MnPrv        Shed       0      2   2006       WD       Abnorml    140000
4    5         60       RL        84.0   14260   Pave  Grvl      IR1  ...  MnPrv        Shed       0     12   2008       WD        Normal    250000
5    6         50       RL        85.0   14115   Pave  Grvl      IR1  ...  MnPrv        Shed     700     10   2009       WD        Normal    143000
6    7         20       RL        75.0   10084   Pave  Grvl      Reg  ...  MnPrv        Shed       0      8   2007       WD        Normal    307000
7    8         60       RL        60.0   10382   Pave  Grvl      IR1  ...  MnPrv        Shed     350     11   2009       WD        Normal    200000
8    9         50       RM        51.0    6120   Pave  Grvl      Reg  ...  MnPrv        Shed       0      4   2008       WD       Abnorml    129900
9   10        190       RL        50.0    7420   Pave  Grvl      Reg  ...  MnPrv        Shed       0      1   2008       WD        Normal    118000
10  11         20       RL        70.0   11200   Pave  Grvl      Reg  ...  MnPrv        Shed       0      2   2008       WD        Normal    129500
11  12         60       RL        85.0   11924   Pave  Grvl      IR1  ...  MnPrv        Shed       0      7   2006      New       Partial    345000
12  13         20       RL        60.0   12968   Pave  Grvl      IR2  ...  MnPrv        Shed       0      9   2008       WD        Normal    144000
```

7. Transformasi Data (Imputasi Nilai Nol/Konstanta)
   Digunakan untuk menggantikan nilai yang hilang dengan imputasi nilai nol atau konstanta. Dalam output yang data terdapat NaN diganti dengan nilai 0.

Source Code

```
import pandas as pd
import numpy as np
from sklearn.impute import SimpleImputer

dataset = "train123.csv"
df = pd.read_csv(dataset)
df = pd.DataFrame(df)
print(df.head(12))

df = df.fillna(0)
print(df.head(12))
```

Output

```
    Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  ...  Fence MiscFeature MiscVal MoSold YrSold SaleType SaleCondition SalePrice
0    1          60       RL         65.0     8450   Pave   NaN      Reg  ...    NaN         NaN       0      2   2008       WD        Normal    208500
1    2          20       RL         80.0     9600   Pave   NaN      Reg  ...    NaN         NaN       0      5   2007       WD        Normal    181500
2    3          60       RL         68.0    11250   Pave   NaN      IR1  ...    NaN         NaN       0      9   2008       WD        Normal    223500
3    4          70       RL         60.0     9550   Pave   NaN      IR1  ...    NaN         NaN       0      2   2006       WD       Abnorml    140000
4    5          60       RL         84.0    14260   Pave   NaN      IR1  ...    NaN         NaN       0     12   2008       WD        Normal    250000
5    6          50       RL         85.0    14115   Pave   NaN      IR1  ...  MnPrv        Shed     700     10   2009       WD        Normal    143000
6    7          20       RL         75.0    10084   Pave   NaN      Reg  ...    NaN         NaN       0      8   2007       WD        Normal    307000
7    8          60       RL          NaN    10382   Pave   NaN      IR1  ...    NaN        Shed     350     11   2009       WD        Normal    200000
8    9          50       RM         51.0     6120   Pave   NaN      Reg  ...    NaN         NaN       0      4   2008       WD       Abnorml    129900
9   10         190       RL         50.0     7420   Pave   NaN      Reg  ...    NaN         NaN       0      1   2008       WD        Normal    118000
10  11          20       RL         70.0    11200   Pave   NaN      Reg  ...    NaN         NaN       0      2   2008       WD        Normal    129500
11  12          60       RL         85.0    11924   Pave   NaN      IR1  ...    NaN         NaN       0      7   2006      New       Partial    345000

[12 rows x 81 columns]
    Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  ...  Fence MiscFeature MiscVal MoSold YrSold SaleType SaleCondition SalePrice
0    1          60       RL         65.0     8450   Pave     0      Reg  ...      0           0       0      2   2008       WD        Normal    208500
1    2          20       RL         80.0     9600   Pave     0      Reg  ...      0           0       0      5   2007       WD        Normal    181500
2    3          60       RL         68.0    11250   Pave     0      IR1  ...      0           0       0      9   2008       WD        Normal    223500
3    4          70       RL         60.0     9550   Pave     0      IR1  ...      0           0       0      2   2006       WD       Abnorml    140000
4    5          60       RL         84.0    14260   Pave     0      IR1  ...      0           0       0     12   2008       WD        Normal    250000
5    6          50       RL         85.0    14115   Pave     0      IR1  ...  MnPrv        Shed     700     10   2009       WD        Normal    143000
6    7          20       RL         75.0    10084   Pave     0      Reg  ...      0           0       0      8   2007       WD        Normal    307000
7    8          60       RL          0.0    10382   Pave     0      IR1  ...      0        Shed     350     11   2009       WD        Normal    200000
8    9          50       RM         51.0     6120   Pave     0      Reg  ...      0           0       0      4   2008       WD       Abnorml    129900
9   10         190       RL         50.0     7420   Pave     0      Reg  ...      0           0       0      1   2008       WD        Normal    118000
10  11          20       RL         70.0    11200   Pave     0      Reg  ...      0           0       0      2   2008       WD        Normal    129500
11  12          60       RL         85.0    11924   Pave     0      IR1  ...      0           0       0      7   2006      New       Partial    345000

[12 rows x 81 columns]
PS C:\Users\alifi\Documents\Pertemuan 9>
```

8. Transformasi Data (Imputasi Regresi: Deterministik)

Pertemuan 9_Contoh Kasus_Transformasi data.ipynb ● TranformasiData.ipynb ●

TranformasiData.ipynb > (mno.matrix(df, figsize = (20, 8)))

+ Code + Markdown | Run All Clear All Outputs Restart | Variables Outline ··· base (Python 3.9.13)

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import linear_model
import missingno as mno
```

[15] ✓ 0.2s Python

```
dataset = "train123.csv"
df = pd.read_csv(dataset)
print(df.head(12))
print(df.info())
print(df.describe())
```

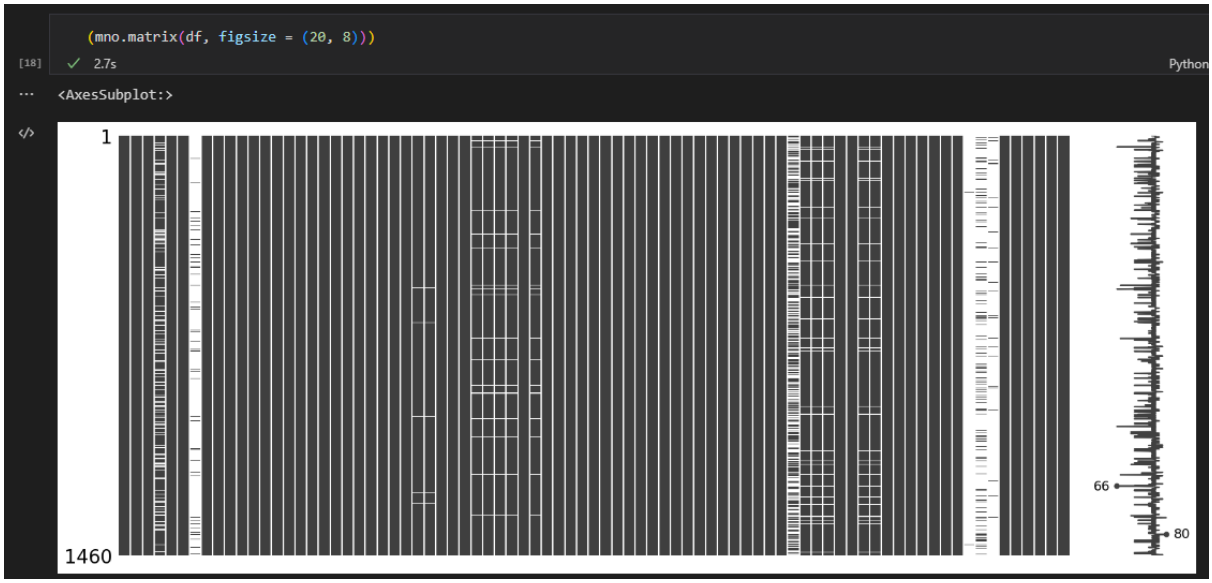
[16] ✓ 1.1s Python

```
... Output exceeds the size limit. Open the full output data in a text editor
    Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  \
0    1          60       RL         65.0     8450   Pave   NaN      Reg
1    2          20       RL         80.0     9600   Pave   NaN      Reg
2    3          60       RL         68.0    11250   Pave   NaN      IR1
3    4          70       RL         60.0     9550   Pave   NaN      IR1
4    5          60       RL         84.0    14260   Pave   NaN      IR1
5    6          50       RL         85.0    14115   Pave   NaN      IR1
6    7          20       RL         75.0    10084   Pave   NaN      Reg
7    8          60       RL          NaN    10382   Pave   NaN      IR1
8    9          50       RM         51.0     6120   Pave   NaN      Reg
9   10         190       RL         50.0     7420   Pave   NaN      Reg
10  11          20       RL         70.0    11200   Pave   NaN      Reg
11  12          60       RL         85.0    11924   Pave   NaN      IR1

   LandContour Utilities  ... PoolArea PoolQC  Fence MiscFeature MiscVal  \
0          Lvl    AllPub  ...        0    NaN    NaN         NaN       0
1          Lvl    AllPub  ...        0    NaN    NaN         NaN       0
2          Lvl    AllPub  ...        0    NaN    NaN         NaN       0
3          Lvl    AllPub  ...        0    NaN    NaN         NaN       0
4          Lvl    AllPub  ...        0    NaN    NaN         NaN       0
5          Lvl    AllPub  ...        0    NaN  MnPrv        Shed     700
6          Lvl    AllPub  ...        0    NaN    NaN         NaN       0
7          Lvl    AllPub  ...        0    NaN    NaN        Shed     350
8          Lvl    AllPub  ...        0    NaN    NaN         NaN       0
9          Lvl    AllPub  ...        0    NaN    NaN         NaN       0
...
75%       0.000000      0.000000     8.000000  2009.000000  214000.000000
max     738.000000  15500.000000    12.000000  2010.000000  755000.000000

[8 rows x 38 columns]
```

```
df.loc[df["MSSubClass"] == 0.0, "MSSubClass"] = np.nan
df.loc[df["LotArea"] == 0.0, "LotArea"] = np.nan
df.loc[df["LotFrontage"] == 0.0, "LotFrontage"] = np.nan
df.isnull().sum()[1:4]
```

[17] ✓ 0.1s Python

```
... MSSubClass       0
MSZoning         0
LotFrontage    259
dtype: int64
```


(mno.matrix(df, figsize = (20, 8)))
[18]
2.7s
Python
<AxesSubplot:>
1
1460
66
80