LAPORAN LEMBAR KERJA PERTEMUAN 4

Langkah 1

```python
import pandas as pd

# Membuat data
data = {
    "IPK": [3.8, 2.5, 3.4, 2.1, 3.9],
    "Jumlah_Absensi": [3, 8, 4, 12, 2],
    "Waktu_Belajar_Jam": [10, 5, 7, 2, 12],
    "Lulus": [1, 0, 1, 0, 1]
}

# Ubah menjadi DataFrame
df = pd.DataFrame(data)

# Simpan ke file CSV
df.to_csv("Kelulusan_mahasiswa.csv", index=False)

print("File CSV berhasil dibuat di folder kerja saat ini!")
print(df)
```

```
File CSV berhasil dibuat di folder kerja saat ini!
   IPK  Jumlah_Absensi  Waktu_Belajar_Jam  Lulus
0  3.8               3                 10      1
1  2.5               8                  5      0
2  3.4               4                  7      1
3  2.1              12                  2      0
4  3.9               2                 12      1
```

langkah 2

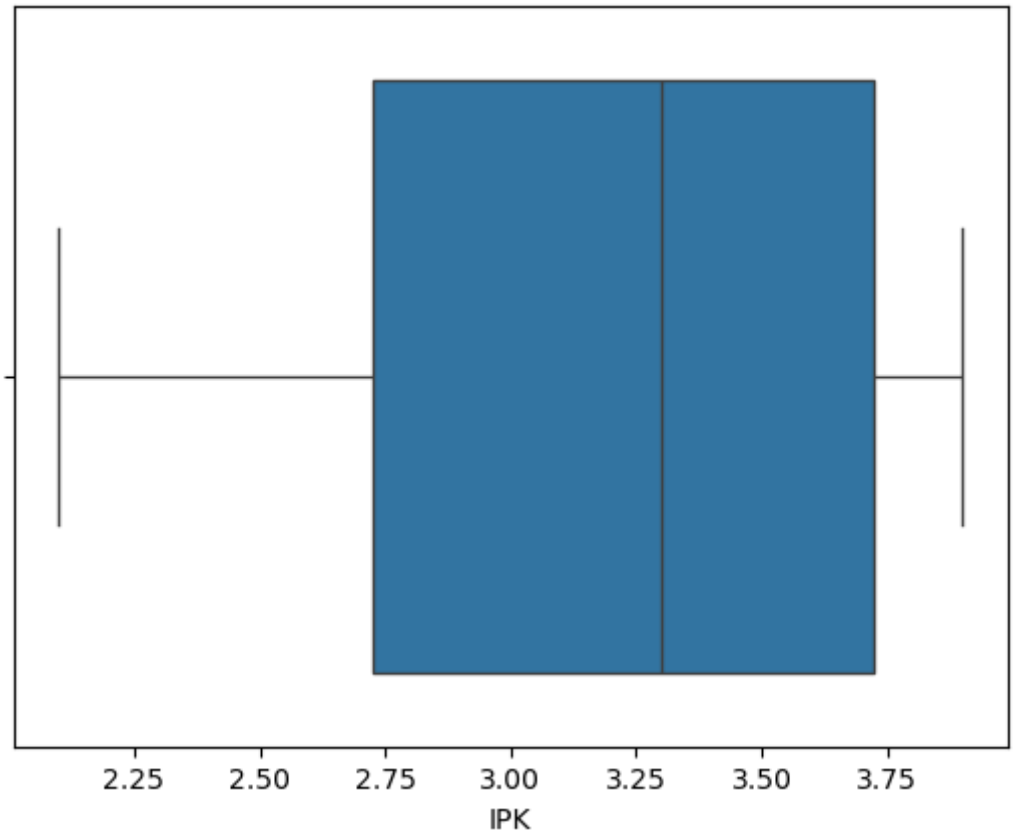
```python
import pandas as pd

# Pastikan file X_train.csv sudah ada di folder ini
df = pd.read_csv(r"C:\Users\Dedeh\OneDrive\Desktop\Text Document\X_train.csv")

print("=== 5 Data Pertama ===")
print(df.head())

print("\n=== Jumlah Data Kosong per Kolom ===")
print(df.isnull().sum())
```

```
=== 5 Data Pertama ===
   IPK  Jumlah_Absensi  Waktu_Belajar_Jam  Lulus
0  3.8               3                 10      1
1  2.5               8                  5      0
2  3.4               4                  7      1
3  2.1              12                  2      0
4  3.9               2                 12      1

=== Jumlah Data Kosong per Kolom ===
IPK                 0
Jumlah_Absensi      0
Waktu_Belajar_Jam   0
Lulus               0
dtype: int64
```

Langkah 3

```
print(df.isnull().sum())
df = df.drop_duplicates()

import seaborn as sns
sns.boxplot(x=df['IPK'])
```

```
IPK                 0
Jumlah_Absensi      0
Waktu_Belajar_Jam   0
Lulus               0
Rasio_Absensi       0
IPK_x_Study         0
dtype: int64
<Axes: xlabel='IPK'>
```



Langkah 4

```
import seaborn as sns
import matplotlib.pyplot as plt
print(df.describe())
sns.histplot(df['IPK'], bins=10, kde=True)
sns.scatterplot(x='IPK', y='Waktu_Belajar_Jam', data=df, hue='Lulus')
sns.heatmap(df.corr(), annot=True, cmap="coolwarm")
```

```
             IPK  Jumlah_Absensi  Waktu_Belajar_Jam     Lulus  Rasio_Absensi  \
count   8.000000        8.000000           8.000000  8.000000       8.000000
mean    3.175000        5.500000           7.375000  0.625000       0.392857
std     0.654108        3.338092           3.113909  0.517549       0.238435
min     2.100000        2.000000           2.000000  0.000000       0.142857
25%     2.725000        3.000000           5.750000  0.000000       0.214286
50%     3.300000        4.500000           7.500000  1.000000       0.321429
75%     3.725000        7.250000           9.250000  1.000000       0.517857
max     3.900000       12.000000          12.000000  1.000000       0.857143

       IPK_x_Study
count     8.000000
mean     25.125000
std      13.994565
min       4.200000
25%      15.725000
50%      24.700000
75%      34.475000
max      46.800000
<Axes: >
```