Nama : Daru Arya Anugra

Nim : A11.2022.14737

Kelompok : Data Science 05

TUGAS 1 - CAPSTONE PROJECT

1. Dari Google Collab Import ke Drive, Lalu mengkoneksikan ke Drive

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call dr
```

2. Membaca data di dalam My Drive ObesityDataSet.csv

```
'/content/drive/MyDrive/ObesityDataSet.csv'
```

```
'/content/drive/MyDrive/ObesityDataSet.csv'
```

```
import pandas as pd

file_path = '/content/drive/MyDrive/ObesityDataSet.csv'

df = pd.read_csv('/content/drive/MyDrive/ObesityDataSet.csv')
```

3. Menampilkan baris, kolom, deskripsi statistik, beberapa baris utama dalam csv.

```
print("Shape (baris, kolom):", df.shape)
print("\nInfo dataset:")
print(df.info())
print("\nDeskripsi statistik:")
print(df.describe(include='all'))
print("\nBeberapa baris pertama:")
print(df.head())
```

```
Shape (baris, kolom): (2111, 17)

Info dataset:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
Data columns (total 17 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Age                             2097 non-null   object
 1   Gender                          2102 non-null   object
 2   Height                          2099 non-null   object
 3   Weight                          2100 non-null   object
 4   CALC                            2106 non-null   object
 5   FAVC                            2100 non-null   object
 6   FCVC                            2103 non-null   object
 7   NCP                             2099 non-null   object
 8   SCC                             2101 non-null   object
 9   SMOKE                           2106 non-null   object
 10  CH2O                            2105 non-null   object
 11  family_history_with_overweight  2098 non-null   object
 12  FAF                             2103 non-null   object
 13  TUE                             2102 non-null   object
 14  CAEC                            2100 non-null   object
 15  MTRANS                          2105 non-null   object
 16  NObeyesdad                      2111 non-null   object
dtypes: object(17)
memory usage: 280.5+ KB
None

Deskripsi statistik:
          Age Gender Height Weight       CALC  FAVC  FCVC   NCP   SCC SMOKE  \
count    2097   2102   2099   2100       2106  2100  2103  2099  2101  2106
unique   1394      3   1562   1518          5     3   808   637     3     3
top        18   Male    1.7     80  Sometimes   yes     3     3    no    no
```

FileNotFoundError common.py F •••

Please explain the error:

```
FileNotFoundError: [Errno 2] No such file or
```

Hypothesis

The error `FileNotFoundError: [Errno 2] No such file or directory: '/content/drive/My Drive/ObesityDataSet.csv'` indicates that the Python interpreter cannot find the file 'ObesityDataSet.csv' at the specified path '/content/drive/My Drive/'. This is likely due to one of the following reasons:

1. **The file does not exist at that location:** Make sure the file 'ObesityDataSet.csv' is actually present in your Google Drive under the 'My Drive' folder. Double-check the file name for any typos as well.
2. **Google Drive is not mounted properly:** The code uses `drive.mount('/content/drive')` to mount your Google Drive to the Colab environment. If this step was not executed successfully, the path `/content/drive/My Drive/` will not be accessible.
3. **Incorrect file path:** The path `/content/drive/My Drive/ObesityDataSet.csv` might be incorrect. Google Drive might have subfolders or a different structure than assumed.

Suggested Changes

1. **Verify the file's existence and location in Google Drive:**
   - Go to your Google Drive in your web browser.
   - Navigate to the 'My Drive' folder.
   - Ensure the file 'ObesityDataSet.csv' is present there.
   - If not present, upload the file to the correct location.
2. **Confirm Google Drive Mounting:**
   - Run the cell containing `from google.colab import drive; drive.mount('/content/drive')` in your Colab notebook.
   - Follow the authentication prompts to grant Colab access to your Drive.
   - A message should appear confirming that the drive was mounted successfully. If not mounted, re-run the code and follow the authorization instructions provided by Google Colab.
3. **Double-check the file path:**
   - If you made any changes to the file's location within Google Drive, update the `file_path` variable in the code:

```
import pandas as pd

file_path = '/content/drive/My Drive/Obes
df = pd.read_csv(file_path)
```

Use code with caution

Rate this answer

```
freq     124   1056     58     58       1386  1844   647  1183  1997  2054

        CH2O family_history_with_overweight   FAF   TUE       CAEC  \
count   2105                            2098  2103  2102       2100
unique  1263                               3  1186  1130          5
top        2                             yes     0     0  Sometimes
freq     441                            1705   404   552       1747

                       MTRANS      NObeyesdad
count                    2105            2111
unique                      6               7
top     Public_Transportation  Obesity_Type_I
freq                     1572             351

Beberapa baris pertama:
  Age  Gender Height Weight        CALC FAVC FCVC NCP  SCC SMOKE CH2O  \
0  21  Female   1.62     64          no   no    2   3   no    no    2
1  21  Female   1.52     56   Sometimes   no    3   3  yes   yes    3
2  23    Male    1.8     77  Frequently   no    2   3   no    no    2
3  27    Male    1.8     87  Frequently   no    3   3   no    no    2
4  22    Male   1.78   89.8   Sometimes   no    2   1   no    no    2

  family_history_with_overweight FAF TUE       CAEC                MTRANS  \
```

4. Mengecheck Jumlah Missing Values Per Kolom dari csv.

```
print("\nJumlah missing values per kolom:")
print(df.isnull().sum())
```

```
Jumlah missing values per kolom:
Age                               14
Gender                             9
Height                            12
Weight                            11
CALC                               5
FAVC                              11
FCVC                               8
NCP                               12
SCC                               10
SMOKE                              5
CH2O                               6
family_history_with_overweight    13
FAF                                8
TUE                                9
CAEC                              11
MTRANS                             6
NObeyesdad                         0
dtype: int64
```

5. Menghitung Jumlah Nilai Unik Per Kolom, Dari CSV.

```
print("\nJumlah nilai unik per kolom:")
print(df.nunique())
```

```
Jumlah nilai unik per kolom:
Age                               1394
Gender                               3
Height                            1562
Weight                            1518
CALC                                 5
FAVC                                 3
FCVC                               808
NCP                                637
SCC                                  3
SMOKE                                3
CH2O                              1263
family_history_with_overweight       3
FAF                               1186
TUE                               1130
CAEC                                 5
MTRANS                               6
NObeyesdad                           7
dtype: int64
```

6. Check Data Duplikat Yang Kembar, di dalam csv

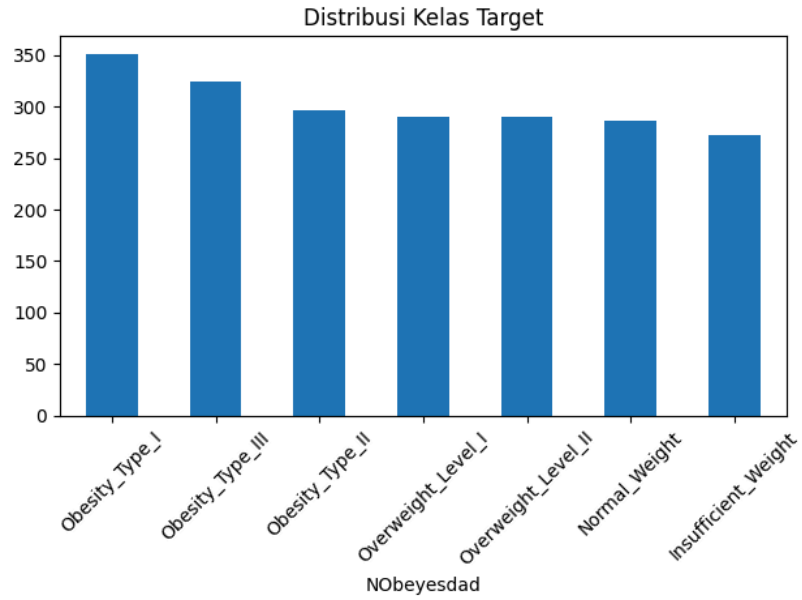
```
print("\nJumlah data duplikat:", df.duplicated().sum())
```

```
Jumlah data duplikat: 18
```

7. Mengecheck Distribusi Kelas Target, Pada Kolom

```
if 'NObeyesdad' in df.columns:
    print("\nDistribusi kelas target (NObeyesdad):")
    print(df['NObeyesdad'].value_counts())
    df['NObeyesdad'].value_counts().plot(kind='bar')
    plt.title('Distribusi Kelas Target')
    plt.xticks(rotation=45)
    plt.tight_layout()
    plt.show()
```

```
Distribusi kelas target (NObeyesdad):
NObeyesdad
Obesity_Type_I         351
Obesity_Type_III       324
Obesity_Type_II        297
Overweight_Level_I     290
Overweight_Level_II    290
Normal_Weight          287
Insufficient_Weight    272
Name: count, dtype: int64
```

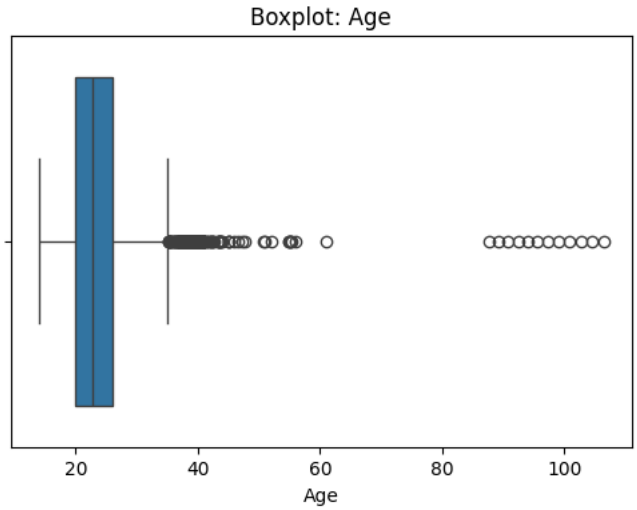


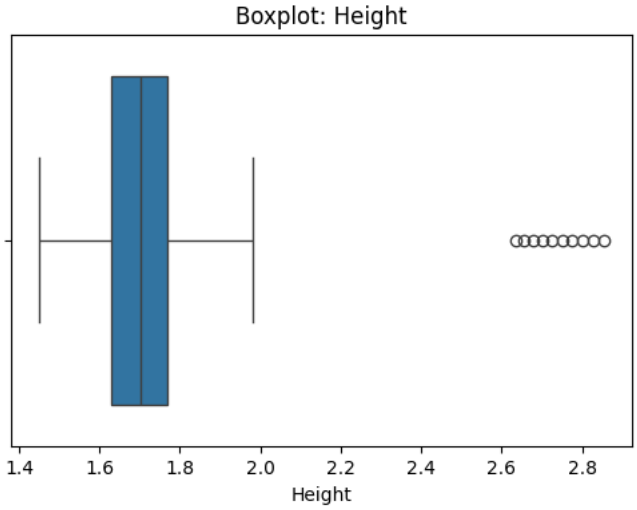
8. Mengkonversikan Objek Ke Float Dalam Bentuk Numerik.

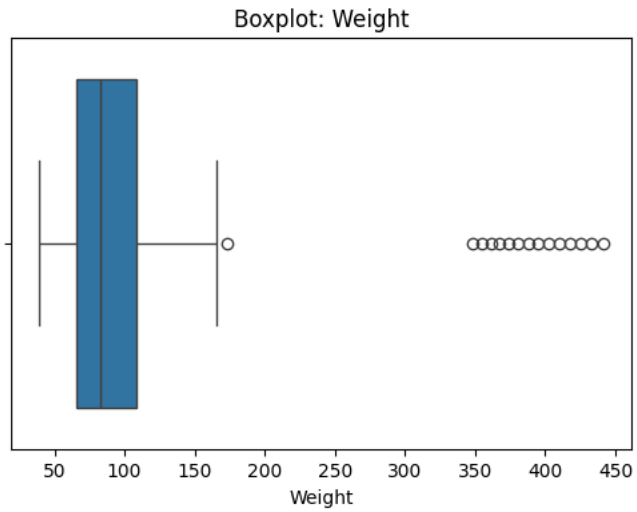
```
kolom_numerik = ['Age', 'Height', 'Weight', 'FCVC', 'NCP', 'CH2O', 'FAF', 'TUE']
for col in kolom_numerik:
    df[col] = pd.to_numeric(df[col], errors='coerce')  # invalid parsing jadi NaN
```

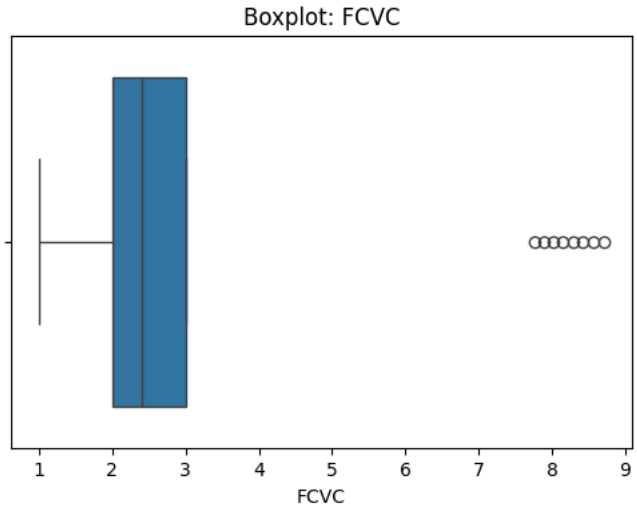
9. Mengecheck Heat Map Yang Berkorelasi dengan kolom Numerik.

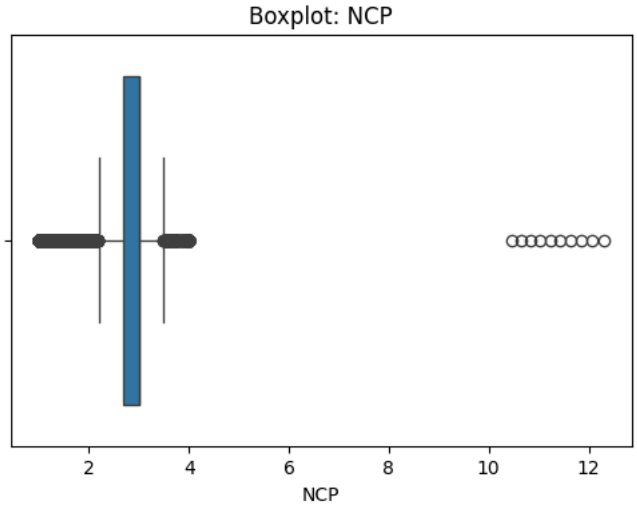
```
for col in kolom_numerik:
    plt.figure(figsize=(6, 4))
    sns.boxplot(x=df[col])
    plt.title(f'Boxplot: {col}')
    plt.show()
```

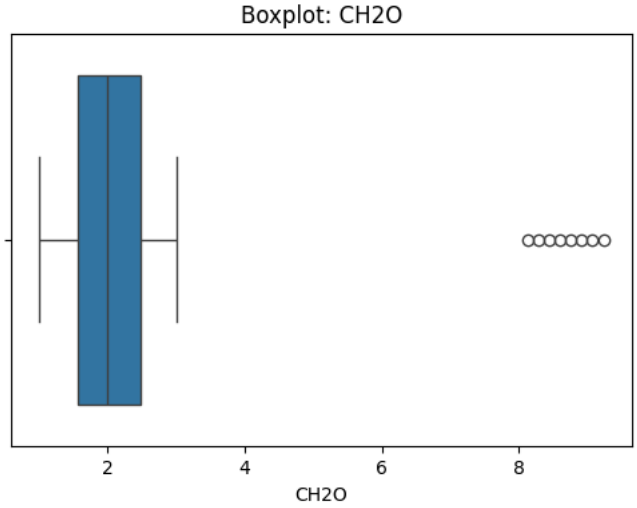

Boxplot: Age
20
40
60
80
100
Age

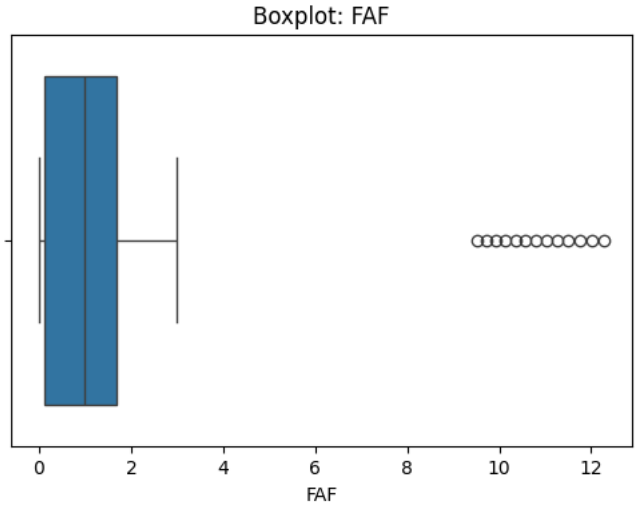

Boxplot: Height
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
Height

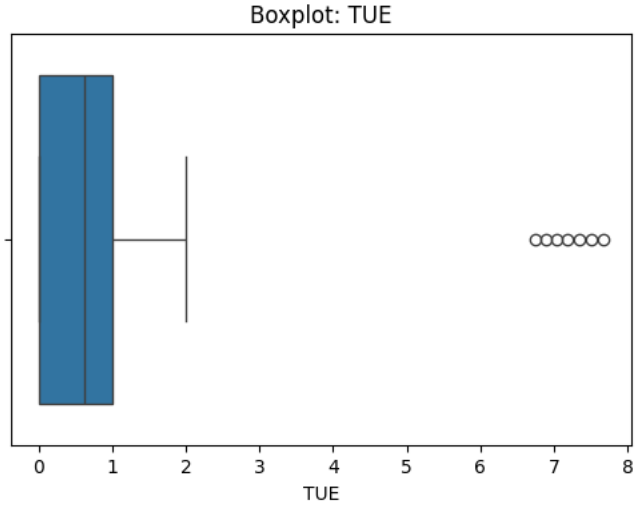

Boxplot: Weight
50
100
150
200
250
300
350
400
450
Weight


Boxplot: FCVC
1
2
3
4
5
6
7
8
9
FCVC


Boxplot: NCP
2
4
6
8
10
12
NCP


Boxplot: CH2O
2
4
6
8
CH2O


Boxplot: FAF
0
2
4
6
8
10
12
FAF


Boxplot: TUE
0
1
2
3
4
5
6
7
8
TUE