Patryk Ostrowski, mod\_4\_zad\_3

import pandas as pd  
  
# Configuration: display all  
pd.set\_option('display.max\_rows', None)  
pd.set\_option('display.max\_columns', None)  
  
df = pd.read\_csv('27\_\_dane.csv', sep=';')  
  
print('Data set:')  
print(df.sample(20).to\_string())  
print()  
  
set\_lenght = len(df)  
print(f'Data set contains {set\_lenght} rows.')  
print()  
  
# Pozbądź się wszystkich rzędów, które zawierają duplikaty  
duplicates\_sum = df.duplicated().sum()  
print(f'Number of duplicates (as exist): {duplicates\_sum}.')  
print()  
  
print('Duplicates sorted (as exist):')  
print(df[df.duplicated(keep=False)].sort\_values(by='Name').to\_string())  
print()  
  
df = df.drop\_duplicates()  
print('Duplicates are being dropped...')  
duplicates\_sum = df.duplicated().sum()  
print(f'Number of duplicates (after drop): {duplicates\_sum}.')  
print()  
  
print('Duplicates sorted (after drop):')  
print(df[df.duplicated(keep=False)].sort\_values(by='Name').to\_string())  
print()  
  
  
  
# Rozbij kolumnę 'Name’ na dwie kolumny 'Imie’ i 'Nazwisko’  
df['Imię'] = df['Name'].str.partition()[0]  
df['Nazwisko'] = df['Name'].str.partition()[2]  
print("Column 'Name' after split:")  
print(df.sample(10).to\_string())  
print()  
  
  
  
# Usuń kolumnę 'Name’  
df = df.drop('Name', axis=1)  
print(df.sample(10).to\_string())  
print()

# Napraw rzędy, które zawierają brakujące dane  
print('Show rows with missing values and the sum of them:')  
print(df[['Age', 'Address', 'Height', 'Weight', 'Imię', 'Nazwisko']].isnull().sum())  
df = df.fillna(  
 {  
 'Age' : df['Age'].median(),  
 'Address' : 'unknown',  
 'Height' : df['Height'].mean(),  
 'Weight' : df['Weight'].mean(),  
 'Imię' : 'unknown',  
 'Nazwisko' : 'unknown'  
 }  
)  
print()  
  
print('Filling up missing data in progress...')  
print()  
print('Show rows with missing values and the sum of them:')  
print(df[['Age', 'Address', 'Height', 'Weight', 'Imię', 'Nazwisko']].isnull().sum())  
print()  
  
print('Show again random data:')  
print(df.sample(20).to\_string())  
print()  
  
  
  
# Zmień nazwy kolumn:  
#  
# ’Age’ -> 'Wiek’  
# ’Height’ -> 'Wzrost’  
# ’Weight’ -> 'Waga’  
# ’Address’ -> 'Adres’  
df = df.rename(columns={'Age' : 'Wiek', 'Height' : 'Wzrost', 'Weight' : 'Waga', 'Address' : 'Adres'})  
print('Column names have been changed:')  
print()  
print(df.sample(1).to\_string())  
print()  
  
  
  
# Dodaj kolumnę 'BMI’ obliczoną jako `Waga / (Wzrost / 100) ^ 2`  
df['BMI'] = df['Waga'] / (df['Wzrost'] / 100) \*\* 2  
print('BMI column added:')  
print(df['BMI'].sample(5).to\_string())  
print()  
print('Entire data frame:')  
print(df.sample(5).to\_string())  
print()

# Posortuj DataFrame po kolumnie 'BMI’ malejąco  
print('All the data frame sorted by BMI descending:')  
print()  
df = df.sort\_values(by='BMI', ascending=False)  
print(df.to\_string())  
  
  
  
# Zapisz zmodyfikowany DataFrame do pliku CSV  
df.to\_csv('Patryk Ostrowski - mod\_4\_zad\_3.csv', index=False)