Patryk Ostrowski, mod_4_zad_3

```python
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()
```

```python
# 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)
```