Dataset Premature Death by Non-Communicable Disease

import pandas as pd

data = pd.read\_csv('NCD Disease.csv')

df\_new = data.rename(columns = {'Unnamed: 0': 'Country', 'Unnamed: 1': 'Year', 'Unnamed: 2': 'Causes','Number of deaths attributed to non-communicable diseases, by type of disease and sex': 'Number of deaths attributed to non-communicable diseases, by type of disease and both sex', 'Number of deaths attributed to non-communicable diseases, by type of disease and sex.1': 'Male', 'Number of deaths attributed to non-communicable diseases, by type of disease and sex.2': 'Female'}, inplace = False)

print(df\_new)

df\_new.to\_csv(r'D:\BITI2513 INTRODUCTION TO DATA SCIENCE\NCD Diseases 3.csv')

import pandas as pd

df = pd.read\_csv('NCD Diseases 3.csv')

del df['Male']

del df['Female']

# Set index on a Dataframe

# Get indexes where name column has value john

df.set\_index("Country", inplace = True)

result = df.loc[["Malaysia", "Indonesia","Singapore","Philippines","Lao People's Democratic Republic","Cambodia","Myanmar","Viet Nam","Thailand","Brunei Darussalam"]]

result

result.to\_csv(r'D:\BITI2513 INTRODUCTION TO DATA SCIENCE\NCD Diseases 4.csv')

Coding (Partition the causes to integer values)

|  |
| --- |
| # Import package  import pandas as pd  # Load the CSV data into DataFrames  dataset = pd.read\_csv(r'C:\Users\ASUS\Documents\UTeM\SEM 4\DS\Project\NCD\_Diseases\_8.csv')  # Display data  df = pd.DataFrame(dataset)  df  # Categorize causes  df['Causes new'] = pd.factorize(df.Causes)[0]  # Display data  Df  # Drop unwanted column  df.drop('Causes', inplace = True, axis = 1)  # Display data after drop unwanted column  Df  # Save data to CSV file  df.to\_csv(r'C:\Users\ASUS\Documents\UTeM\SEM 4\DS\Project\Dataset\Dataset 1 [cleaned].csv', index = False) |