EXPERIMENT 4

# Import necessary libraries

import pandas as pd

# Function to load data from a file

def load\_data(file\_path):

# Read the dataset

data = pd.read\_csv('/content/household-living-costs-price-indexes.csv')

return data

# Function to export data to a file

def export\_data(data, output\_path):

# Save the dataset to a new file

data.to\_csv(output\_path, index=False)

# Function to display dataset details

def dataset\_details(data):

print("Number of rows:", data.shape[0])

print("Number of columns:", data.shape[1])

print("\nFirst five rows of the dataset:\n", data.head())

print("\nDataset size:", data.size)

print("\nNumber of missing values in each column:\n", data.isnull().sum())

# For numerical columns only

numerical\_data = data.select\_dtypes(include=['number'])

print("\nSum of numerical columns:\n", numerical\_data.sum())

print("\nAverage of numerical columns:\n", numerical\_data.mean())

print("\nMinimum values of numerical columns:\n", numerical\_data.min())

print("\nMaximum values of numerical columns:\n", numerical\_data.max())

# Main execution

if \_name\_ == "\_main\_":

# Load dataset from a CSV file

file\_path = 'your\_input\_file.csv' # Update this path to your dataset file

data = load\_data(file\_path)

# Show dataset details

dataset\_details(data)

# Export the dataset to a new CSV file

output\_path = 'your\_output\_file.csv' # Update this path where you want to export

export\_data(data, output\_path)

Number of rows: 70

Number of columns: 15

First five rows of the dataset:

hlpi\_name year hlpi tot\_hhs own own\_wm own\_prop \

0 All households 2008 allhh 1560859 1087580 574406 69.7

1 Beneficiary 2008 benef 185965 71256 39405 38.3

2 Income quintile 1 (low) 2008 disq1 312376 191470 48424 61.3

3 Income quintile 2 2008 disq2 312333 196203 84171 62.8

4 Income quintile 3 2008 disq3 312240 217657 141318 69.7

own\_wm\_prop prop\_hhs age size income expenditure eqv\_income eqv\_exp

0 36.8 100.0 35.9 2.7 46704 42394 26869 25132

1 21.2 11.9 29.9 2.6 23404 25270 14258 15824

2 15.5 20.0 40.0 2.3 16747 21145 13402 14408

3 26.9 20.0 34.7 2.8 31308 29855 18917 18266

4 45.3 20.0 31.5 3.0 49106 46561 26870 24672

Dataset size: 1050

Number of missing values in each column:

hlpi\_name 0

year 0

hlpi 0

tot\_hhs 0

own 0

own\_wm 0

own\_prop 0

own\_wm\_prop 0

prop\_hhs 0

age 0

size 0

income 0

expenditure 0

eqv\_income 0

eqv\_exp 0

dtype: int64

Sum of numerical columns:

year 140980.0

tot\_hhs 28876349.0

own 18827977.0

own\_wm 9074762.0

own\_prop 4445.4

own\_wm\_prop 2124.4

prop\_hhs 1749.0

age 2679.9

size 183.8

income 3968635.0

expenditure 3448699.0

eqv\_income 2341605.0

eqv\_exp 2065733.0

dtype: float64

Average of numerical columns:

year 2014.000000

tot\_hhs 412519.271429

own 268971.100000

own\_wm 129639.457143

own\_prop 63.505714

own\_wm\_prop 30.348571

prop\_hhs 24.985714

age 38.284286

size 2.625714

income 56694.785714

expenditure 49267.128571

eqv\_income 33451.500000

eqv\_exp 29510.471429

dtype: float64

Minimum values of numerical columns:

year 2008.0

tot\_hhs 132215.0

own 30080.0

own\_wm 14220.0

own\_prop 22.8

own\_wm\_prop 5.1

prop\_hhs 7.8

age 27.3

size 1.6

income 16747.0

expenditure 16413.0

eqv\_income 13402.0

eqv\_exp 11015.0

dtype: float64

Maximum values of numerical columns:

year 2020.0

tot\_hhs 1756740.0

own 1125000.0

own\_wm 574662.0

own\_prop 88.1

own\_wm\_prop 58.2

prop\_hhs 100.0

age 70.3

size 3.2

income 146672.0

expenditure 123424.0

eqv\_income 79607.0

eqv\_exp 71985.0

dtype: float64