import pandas as pd

df = pd.read\_csv (r"data.csv")

pd.set\_option("display.max\_columns", None)

# drop unnecessary columns

df = df.drop(["1985", "1986","1987","1988", "1989"], axis=1)

# determine the number of missing values

print("List of missing values in each column")

print("--------------------------------")

print(df.isnull().sum())

# since each column is the next year, it makes the most sense

# to fill all missing values with the number/percentage

# that comes before it in the previous column (previous year)

print()

print("Table with Missing Values")

print("------------------------")

print(df.loc[8:11, ["country", "2001", "2002", "2003", "2004"]])

df = df.fillna(method='ffill', axis=1)

print()

print("Table with Filled in Values")

print("------------------------")

print(df.loc[8:11, ["country", "2001", "2002", "2003", "2004"]])

# determine the number of duplicate rows

print()

print("Number of Duplicate Rows")

print("------------------------")

print(df.duplicated().sum())

# find the duplicated row(s)!

print()

print("Duplicate Row")

print("----------------")

print(df.loc[df.duplicated(), "country"])

# drop all duplicate rows

df.drop\_duplicates(inplace=True)

print()

print()

print("Number of Duplicate Rows Now")

print("------------------------")

print(df.duplicated().sum())