

Untitled0.ipynb - Colab

colab.research.google.com/drive/1Eyzq_iEz1KgQmUSNPao0geBjC3rd5yH6H#scrollTo=w7k1M38laloe

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

```
from google.colab import files
import pandas as pd

# Upload file
uploaded = files.upload()

# Load into pandas
df = pd.read_csv("cleaned_global_water_consumption.csv")

print(df.head())
```

Choose File cleaned_gl...umpton.csv

cleaned_global_water_consumption.csv(text/csv) - 48212 bytes, last modified: 8/30/2025 - 100% done
Saving cleaned_global_water_consumption.csv to cleaned_global_water_consumption.csv

	Country	Year	Total Water Consumption (Billion Cubic Meters)
0	Argentina	2000	481.490000
1	Argentina	2001	455.063000
2	Argentina	2002	482.749231
3	Argentina	2003	452.660000
4	Argentina	2004	634.566000

	Per Capita Water Use (Liters per Day)	Agricultural Water Use (%)
0	235.431429	48.550000
1	299.551000	48.465000
2	340.124615	50.375385
3	326.756667	49.086667

Variables Terminal

3:31 PM Python 3

Untitled0.ipynb - Colab

colab.research.google.com/drive/1Eyzq_iEz1KgQmUSNPao0geBjC3rd5yH6H#scrollTo=w7k1M38laloe

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

```
Industrial Water Use (%) Household Water Use (%) \
0 20.844286 30.100000
1 26.943000 22.550000
2 29.042308 23.349231
3 30.476000 24.440000
4 36.670000 23.924000

Rainfall Impact (Annual Precipitation in mm) \
0 1288.698571
1 1371.729000
2 1590.305385
3 1816.012667
4 815.998000

Groundwater Depletion Rate (%) Water Scarcity Level
0 3.255714 Moderate
1 3.120000 Moderate
2 2.733846 Moderate
3 2.708000 Moderate
4 1.902000 Moderate
```

```
[2] print("Dataset Shape:", df.shape)
print("\nData Types:\n", df.dtypes)
print("\nFirst 5 Rows:\n", df.head())
```

Dataset Shape: (500, 10)

Variables Terminal

3:31 PM Python 3

Untitled0.ipynb - Colab

colab.research.google.com/drive/1Eyzq_iEz1KgQmUSNPaoGeBjC3rd5yH6H#scrollTo=w7k1M3Blaloe

File Edit View Insert Runtime Tools Help

Q Commands + Code + Text ▶ Run all

```
print("Dataset Shape:", df.shape)
print("\nData Types:\n", df.dtypes)
print("\nFirst 5 Rows:\n", df.head())
```

Dataset Shape: (500, 10)

Data Types:

Country	object
Year	int64
Total Water Consumption (Billion Cubic Meters)	float64
Per Capita Water Use (Liters per Day)	float64
Agricultural Water Use (%)	float64
Industrial Water Use (%)	float64
Household Water Use (%)	float64
Rainfall Impact (Annual Precipitation in mm)	float64
Groundwater Depletion Rate (%)	float64
Water Scarcity Level	object
dtype:	object

First 5 Rows:

	Country	Year	Total Water Consumption (Billion Cubic Meters) \
0	Argentina	2000	481.490000
1	Argentina	2001	455.063000
2	Argentina	2002	482.749231
3	Argentina	2003	452.660000
4	Argentina	2004	634.566000

Per Capita Water Use (Liters per Day) Agricultural Water Use (%) \

0	235.431429	48.550000
---	------------	-----------

Variables Terminal

3:31 PM Python 3

15:46 30-08-2025

Untitled0.ipynb - Colab

colab.research.google.com/drive/1Eyzq_iEz1KgQmUSNPaoGeBjC3rd5yH6H#scrollTo=w7k1M3Blaloe

File Edit View Insert Runtime Tools Help

Q Commands + Code + Text ▶ Run all

```
df["Industrial Water Use (%)"]
df["Household Water Use (%)"]
df["Rainfall Impact (Annual Precipitation in mm)"]
df["Groundwater Depletion Rate (%)"]
df["Water Scarcity Level"]
df.dtypes
```

First 5 Rows:

	Country	Year	Total Water Consumption (Billion Cubic Meters) \
0	Argentina	2000	481.490000
1	Argentina	2001	455.063000
2	Argentina	2002	482.749231
3	Argentina	2003	452.660000
4	Argentina	2004	634.566000

Per Capita Water Use (Liters per Day) Agricultural Water Use (%) \

0	235.431429	48.550000
1	299.551000	48.465000
2	340.124615	50.375385
3	326.756667	49.086667
4	230.346000	38.670000

Industrial Water Use (%) Household Water Use (%) \

0	20.844286	30.100000
1	26.943000	22.550000
2	29.042308	23.349231
3	30.476000	24.440000
4	36.670000	23.924000

Rainfall Impact (Annual Precipitation in mm) \

Variables Terminal

3:31 PM Python 3

15:46 30-08-2025

Untitled0.ipynb - Colab

colab.research.google.com/drive/1Eyzq_iEz1KgQmUSNPaoGeBjC3rd5yH6H#scrollTo=w7k1M38laloe

Untitled0.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

RAM Disk

```
1 299.551000 48.465000
2 340.124615 50.375385
3 326.756667 49.086667
4 230.346000 38.670000

Industrial Water Use (%) Household Water Use (%) \
0 20.844286 30.100000
1 26.943000 22.550000
2 29.042308 23.349231
3 30.476000 24.440000
4 36.670000 23.924000

Rainfall Impact (Annual Precipitation in mm) \
0 1288.698571
1 1371.729000
2 1590.305385
3 1816.012667
4 815.998000

Groundwater Depletion Rate (%) Water Scarcity Level
0 3.255714 Moderate
1 3.120000 Moderate
2 2.733846 Moderate
3 2.708000 Moderate
4 1.902000 Moderate
```

[3] print("\nMissing Values:\n", df.isnull().sum())

Missing Values:

Variables Terminal

3:31 PM Python 3

Untitled0.ipynb - Colab

colab.research.google.com/drive/1Eyzq_iEz1KgQmUSNPaoGeBjC3rd5yH6H#scrollTo=w7k1M38laloe

Untitled0.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

RAM Disk

```
4 1.902000 moderate
```

[3] print("\nMissing Values:\n", df.isnull().sum())

Missing Values:

Country	Year	Total Water Consumption (Billion Cubic Meters)	Per Capita Water Use (Liters per Day)	Agricultural Water Use (%)	Industrial Water Use (%)	Household Water Use (%)	Rainfall Impact (Annual Precipitation in mm)	Groundwater Depletion Rate (%)	Water Scarcity Level
0	0	0	0	0	0	0	0	0	0

dtype: int64

[4] print("\nDuplicate Rows:", df.duplicated().sum())

Duplicate Rows: 0

[5] print("\nSummary Statistics:\n", df.describe())

Summary Statistics:

Year	Total Water Consumption (Billion Cubic Meters)
0	0

Variables Terminal

3:31 PM Python 3

Untitled0.ipynb - Colab

colab.research.google.com/drive/1Eyzq_iEz1KgQmUSNPa0geBjC3rd5yH6H#scrollTo=w7k1M3Blaloe

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

RAM Disk

```
[5] print("\nSummary Statistics:\n", df.describe())
```

Summary Statistics:

	Year	Total Water Consumption (Billion Cubic Meters) \
count	500.000000	500.000000
mean	2012.000000	501.224430
std	7.218324	96.078937
min	2000.000000	129.636667
25%	2006.000000	441.447385
50%	2012.000000	502.197154
75%	2018.000000	563.849594
max	2024.000000	798.418000

	Per Capita Water Use (Liters per Day)	Agricultural Water Use (%) \
count	500.000000	500.000000
mean	276.004782	50.180829
std	42.669593	5.566886
min	111.708333	28.905000
25%	250.225406	46.513611
50%	276.430556	50.318482
75%	300.221750	54.061964
max	484.350000	66.520000

	Industrial Water Use (%)	Household Water Use (%) \
count	500.000000	500.000000
mean	27.792837	24.832515
std	4.361660	2.956135
min	13.276667	13.668333
25%	25.018000	23.015355
50%	27.648539	25.071944
75%	30.698333	26.720167
max	43.583333	34.202000

Variables Terminal

3:31 PM Python 3

15:47 30-08-2025

Untitled0.ipynb - Colab

colab.research.google.com/drive/1Eyzq_iEz1KgQmUSNPa0geBjC3rd5yH6H#scrollTo=w7k1M3Blaloe

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

RAM Disk

```
[6] numeric_cols = df.select_dtypes(include=['float64','int64']).columns
```

	Year	Total Water Consumption (Billion Cubic Meters) \
count	500.000000	500.000000
mean	2012.000000	501.224430
std	7.218324	96.078937
min	2000.000000	129.636667
25%	2006.000000	441.447385
50%	2012.000000	502.197154
75%	2018.000000	563.849594
max	2024.000000	798.418000

	Per Capita Water Use (Liters per Day)	Agricultural Water Use (%) \
count	500.000000	500.000000
mean	276.004782	50.180829
std	42.669593	5.566886
min	111.708333	28.905000
25%	250.225406	46.513611
50%	276.430556	50.318482
75%	300.221750	54.061964
max	484.350000	66.520000

	Industrial Water Use (%)	Household Water Use (%) \
count	500.000000	500.000000
mean	27.792837	24.832515
std	4.361660	2.956135
min	13.276667	13.668333
25%	25.018000	23.015355
50%	27.648539	25.071944
75%	30.698333	26.720167
max	43.583333	34.202000

	Rainfall Impact (Annual Precipitation in mm) \
count	500.000000
mean	1544.824300
std	292.786579
min	700.230000
25%	1353.734583
50%	1537.537724
75%	1746.402425
max	2533.678000

	Groundwater Depletion Rate (%)
count	500.000000
mean	2.573037
std	0.480630
min	1.300000
25%	2.235000
50%	2.555778
75%	2.887202
max	4.322000

Variables Terminal

3:31 PM Python 3

15:47 30-08-2025

Untitled0.ipynb - Colab

colab.research.google.com/drive/1Eyzq_iEz1KgQmUSNPa0geBjC3rd5yH6H#scrollTo=w7k1M38laloe

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

max 4.322000

```
numeric_cols = df.select_dtypes(include=['float64', 'int64']).columns
for col in numeric_cols:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = ((df[col] < lower_bound) | (df[col] > upper_bound)).sum()
    if outliers > 0:
        print(f"{col}: {outliers} outliers")
```

Total Water Consumption (Billion Cubic Meters): 10 outliers
Per Capita Water Use (Liters per Day): 8 outliers
Agricultural Water Use (%): 4 outliers
Industrial Water Use (%): 6 outliers
Household Water Use (%): 12 outliers
Rainfall Impact (Annual Precipitation in mm): 4 outliers
Groundwater Depletion Rate (%): 6 outliers

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(10,6))
sns.heatmap(df.select_dtypes(include=['float64', 'int64']).corr(), annot=True, cmap="Blues")
plt.title("Correlation Heatmap")
```

Variables Terminal

3:31 PM Python 3

15:47 30-08-2025





