

## Python Pandas

Pandas is a popular library in Python used for data manipulation and analysis. It provides data structures like DataFrame and Series, which are particularly useful for handling structured data, such as tables and time series. The library is imported at the start of the script.

```
import pandas as pd

# Basic
mydataset = {
    'cars': ["BMW", "Volvo", "Ford"],
    'passings': [3, 7, 2]
}

myvar = pd.DataFrame(mydataset)
print(myvar)

print(pd.__version__)
```

	cars	passings
0	BMW	3
1	Volvo	7
2	Ford	2

2.2.3

```
# Pandas Series
# Create a pandas Series
my_series = pd.Series([1, 2, 3, 4])
print(my_series)

# Create a pandas Series with custom indices
my_series2 = pd.Series([10, 20, 30, 40], index=['a', 'b', 'c', 'd'])
print(my_series2)
```

0	1
1	2
2	3
3	4

dtype: int64

a	10
b	20
c	30
d	40

dtype: int64

```

# Pandas DataFrames
# Creating a DataFrame
data = {'Name': ['Alice', 'Bob', 'Charlie'], 'Age': [25, 30, 35]}
df = pd.DataFrame(data)

# Add a new column
df['City'] = ['New York', 'Los Angeles', 'Chicago']

# Select a column
print(df['Name'])

# Drop a column
df.drop('City', axis=1, inplace=True)

# Filter rows
print(df[df['Age'] > 25])

```

```

# Using csv
df = pd.read_csv('data.csv')
print(df.to_string())

# Using json
df = pd.read_json('data.json')
print(df.to_string())

```

```

0    Alice
1     Bob
2  Charlie
Name: Name, dtype: object
   Name Age
1   Bob  30
2  Charlie 35

```

Output prints the contents of the CSV and the JSON files.

CSV has 169 entries  
JSON has 168 entries

```

# Using csv
df = pd.read_csv('data.csv')
# Using json
df = pd.read_json('data.json')

# Pandas Analyzing DataFrames
print(df.describe()) # Summary statistics of the DataFrame
print(df.corr())     # Correlation between columns

```

	Duration ...	Calories
count	169.000000 ...	164.000000
mean	63.846154 ...	375.800000
std	42.299949 ...	266.377134
min	15.000000 ...	50.300000
25%	45.000000 ...	250.925000
50%	60.000000 ...	318.600000
75%	60.000000 ...	387.600000
max	300.000000 ...	1860.400000

[8 rows x 4 columns]

	Duration ...	Calories
Duration	1.000000 ...	0.922721
Pulse	-0.155408 ...	0.025120
Maxpulse	0.009403 ...	0.203814
Calories	0.922721 ...	1.000000

[4 rows x 4 columns]

Files

Search

data.csv

A

data.json

A

main.py

M

main.py

data.csv

data.json

+

data.csv

1 Duration,Pulse,Maxpulse,Calories

2 60,110,130,409.1

3 60,117,145,479.0

4 60,103,135,349.0

Files

Search

data.csv

A

data.json

A

main.py

M

main.py

data.csv

data.json

+

data.json

1 {

2 "Duration":{

3 "0":60,

