

Numpy and Pandas Cheat Sheet

Common Imports

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Vectorized Operations

```
xs + ys ..... Element-wise addition
xs + z ..... Adding a scalar
xs & ys ..... Bitwise (boolean) and
xs | ys ..... bitwise (boolean) or
xs ..... Bitwise (boolean) not
xs < ys ..... Less than
```

Subtraction (-), multiplication (*), division (/), exponentiation (**), and other comparison operators (<=, >, >=, ==, !=) work similarly.

matplotlib plotting

```
plt.hist(xs) ..... Histogram
plt.scatter(xs, ys) ..... Scatterplot
plt.plot(xs, ys) ..... Line plot
```

Array / Series functions

```
max() ..... Maximum
min() ..... Minumum
mean() ..... Mean (average)
median() ..... Median
sum() ..... Sum (total)
```

Accessing Data in a Series

```
s.iloc[i] ..... Get element by position
s.loc[x] ..... Get element by index
s.values ..... Get NumPy array
```

Plotting for Series

```
s.hist() ..... Histogram
s.plot() ..... Line plot
```

Apply Functions

```
s.apply(value -> value) ..... returns a Series
df.applymap(value -> value) ..... returns a DataFrame
df.apply(series -> value) ..... returns a Series
df.apply(series -> series) ..... returns a DataFrame
```

Accessing Data in a DataFrame

```
df['col'] ..... Get column by name
df.iloc[i] ..... Get row by position
df.loc[x] ..... Get row by index
df.iloc[i, j] ..... Get element by position
df.loc[x, y] ..... Get element by index
df.values ..... Get 2D NumPy array
```

DataFrame Summarization

```
df.describe() ..... Stats about each column
df.head(n) ..... First n rows
df.tail(n) ..... Last n rows
df.columns ..... List of column names
```

Axis Argument

```
df.mean(axis=0) ..... mean of each column
df.mean(axis=1) ..... mean of each row
df.mean(axis='index') ..... mean of each column
df.mean(axis='columns') ..... mean of each row
```

Plotting for DataFrames

```
df.plot() ..... Line plot with one line per column
```