Python Data Science Cheat Sheet

NumPy • Pandas • Matplotlib

NumPy

Import

```
import numpy as np
```

Array Creation

```
np.array([1, 2, 3])
np.zeros((2, 3))
np.ones((2, 3))
np.eye(3)
np.arange(0, 10, 2)
np.linspace(0, 1, 5)
```

Array Operations

```
a.shape
a.reshape((3, 2))
a.T
np.concatenate([a, b], axis=0)
np.vstack([a, b])
np.hstack([a, b])
```

Math & Stats

```
np.sum(a), np.mean(a), np.std(a)
np.max(a), np.min(a)
np.argmax(a), np.argmin(a)
np.dot(a, b)
np.sqrt(a), np.exp(a)
```

Indexing/Slicing

```
a[0, 1], a[:, 0], a[1:3]
a[a > 0]
```

Pandas

Import

```
import pandas as pd
```

Create Data

```
df = pd.DataFrame({'a': [1, 2], 'b': [3, 4]})
s = pd.Series([1, 2, 3])
```

Read / Write

```
pd.read_csv('file.csv')
df.to_csv('out.csv', index=False)
```

Inspect

```
df.head(), df.info(), df.describe()
df.columns, df.shape
```

Selection / Filtering

```
df['a'], df[['a', 'b']]
df.iloc[0], df.loc[0]
df[df['a'] > 1]
```

Manipulation

```
df['c'] = df['a'] + df['b']
df.drop('c', axis=1, inplace=True)
df.rename(columns={'a': 'A'}, inplace=True)
df.sort_values('a')
df.groupby('a').mean()
df.isnull().sum(), df.fillna(0)
```

Matplotlib

Import

```
import matplotlib.pyplot as plt
```

Basic Plotting

```
plt.plot(x, y)
plt.scatter(x, y)
plt.bar(x, height)
plt.hist(data)
```

Customization

```
plt.title("Title")
plt.xlabel("X Axis")
plt.ylabel("Y Axis")
plt.legend()
plt.grid(True)
```

Show / Save

```
plt.show()
plt.savefig('plot.png')
```