

Pandas & Matplotlib Cheat Sheet (Syntax-Focused)

PANDAS – Dataset Manipulation

```
import pandas as pd
import numpy as np

# Load data
df = pd.read_csv("data.csv")
df = pd.read_excel("data.xlsx")

# Inspect
df.head()
df.tail()
df.shape
df.info()
df.describe()

# Selection
df["column"]
df[["col1", "col2"]]
df.loc[0]
df.iloc[1:5]

# Filtering
df[df["age"] > 30]
df[(df["age"] > 30) & (df["salary"] > 3000)]
df[df["city"].isin(["Paris", "Berlin"])]

# Sorting
df.sort_values("age")
df.sort_values("age", ascending=False)

# New columns
df["total"] = df["price"] * df["quantity"]
df["category"] = np.where(df["salary"] > 3000, "High", "Low")

# Missing values
df.isna().sum()
df.dropna()
df.fillna(0)
df.fillna(df.mean(numeric_only=True))

# Groupby
df.groupby("city")["salary"].mean()
df.groupby("city").agg({
    "salary": "mean",
    "age": ["min", "max"]
}).reset_index()

# Apply
df["salary_k"] = df["salary"].apply(lambda x: x / 1000)

# Merge & concat
pd.merge(df1, df2, on="id", how="inner")
pd.concat([df1, df2], axis=0)

# Rename & drop
df.rename(columns={"old": "new"})
df.drop(columns=["col1", "col2"])
```

MATPLOTLIB (pyplot) – Plotting

```
import matplotlib.pyplot as plt

# Line plot
plt.plot(df["x"], df["y"])
plt.xlabel("X")
plt.ylabel("Y")
plt.title("Line Plot")
plt.show()

# Scatter plot
plt.scatter(df["x"], df["y"])
plt.xlabel("X")
plt.ylabel("Y")
plt.show()

# Bar plot
plt.bar(df["category"], df["value"])
plt.show()
```

```
# Histogram
plt.hist(df["salary"], bins=10)
plt.xlabel("Salary")
plt.ylabel("Frequency")
plt.show()

# Boxplot
plt.boxplot(df["salary"])
plt.show()

# Multiple lines
plt.plot(df["x"], df["y1"], label="Y1")
plt.plot(df["x"], df["y2"], label="Y2")
plt.legend()
plt.show()

# Pandas plotting
df["salary"].plot(kind="hist", bins=10)
plt.show()

df.plot(kind="scatter", x="age", y="salary")
plt.show()

# Save figure
plt.savefig("plot.png", dpi=300, bbox_inches="tight")
```