Pandas Cheat Sheet: Essential Commands for Data Analysis

1. Load Data

```
import pandas as pd

df = pd.read_csv("file.csv")          # Load CSV
df = pd.read_excel("file.xlsx")       # Load Excel
df = pd.read_json("file.json")        # Load JSON
```

2. Quick Look at Data

```
df.head()             # First 5 rows
df.tail(10)           # Last 10 rows
df.shape              # (rows, columns)
df.info()             # Summary (non-null, dtypes)
df.describe()         # Stats summary (numerical)
df.columns            # List of column names
df.dtypes             # Data types
```

3. Clean & Fix Data

```
df.dropna()                                          # Drop missing rows
df.fillna(0)                                         # Fill missing values
df.rename(columns={"old": "new"}, inplace=True)      # Rename columns
df["col"] = df["col"].astype("int")                  # Change column type
df.duplicated().sum()                                # Count duplicates
df.drop_duplicates(inplace=True)                     # Remove duplicates
```

4. Select Columns & Rows

```
df["col"]             # One column
df[["col1", "col2"]]  # Multiple columns
df.loc[0]             # Row by label
df.iloc[0]            # Row by index
df.loc[0, "col"]      # Specific cell
```

5. Filter Rows (Conditions)

```
df[df["col"] > 10]                    # Condition
df[(df["a"] > 10) & (df["b"] < 5)]    # Multiple conditions
df[df["col"].isin(["A", "B"])]        # Filter by values
df[~df["col"].isna()]                 # Remove NaNs
```

6. Aggregation & Stats

```
df["col"].mean()
df["col"].sum()
df["col"].value_counts()              # Frequency count
df.groupby("group_col").mean() # Group by and aggregate
df.pivot_table(index="A", columns="B", values="C", aggfunc="sum")
```

7. Modify Data

```
df["new"] = df["a"] + df["b"]                 # Create new column
df["col"] = df["col"].apply(func)             # Apply function
df["col"] = df["col"].str.lower()             # String ops
df["date"] = pd.to_datetime(df["date"]) # Convert to datetime
```

8. Work with Dates

```
df["date"].dt.year
df["date"].dt.month
df["date"].dt.day_name()
df.set_index("date", inplace=True)  # Make date the index
```

9. Plotting (Quick with Seaborn or Pandas)

```
df["col"].plot(kind="hist")                           # Histogram
df.plot(x="a", y="b", kind="scatter")                 # Scatter
df.groupby("col")["val"].mean().plot(kind="bar")  # Bar chart
```

10. Save Data

```
df.to_csv("cleaned.csv", index=False)
df.to_excel("cleaned.xlsx", index=False)
```