

Here's a concise **Python Cheat Sheet for Data Analysis** to help you get started or refresh your knowledge:

1. Importing Libraries

```
import pandas as pd # For data manipulation
import numpy as np  # For numerical operations
import matplotlib.pyplot as plt # For visualization
import seaborn as sns # For advanced visualizations
```

2. Loading and Inspecting Data

```
# Load data
df = pd.read_csv('file.csv') # Read CSV file
df.head()                    # Display first 5 rows
df.info()                    # Summary of the dataset
df.describe()                # Statistical summary
df.shape                     # Rows and columns count
```

3. Data Cleaning

```
# Handling missing values
df.dropna(inplace=True)      # Remove rows with missing values
df.fillna(value, inplace=True) # Fill missing values

# Renaming columns
df.rename(columns={'old_name': 'new_name'}, inplace=True)

# Removing duplicates
df.drop_duplicates(inplace=True)
```

4. Data Manipulation

```
# Selecting columns
df['column_name']          # Select single column
df[['col1', 'col2']]       # Select multiple columns

# Filtering rows
df[df['column'] > 10]       # Filter rows based on condition

# Adding new columns
df['new_col'] = df['col1'] + df['col2']

# Grouping and aggregating
df.groupby('column').mean() # Group by and calculate mean
```

5. Visualization

```
# Line plot
plt.plot(df['column'])
plt.show()

# Histogram
df['column'].hist()

# Scatter plot
sns.scatterplot(x='col1', y='col2', data=df)

# Heatmap
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
```

6. Statistical Analysis

```
# Summary statistics
df['column'].mean()        # Mean
df['column'].median()      # Median
df['column'].std()         # Standard deviation
```

```
# Correlation
```

```
df.corr()
```

```
# Correlation matrix
```

7. Exporting Data

```
df.to_csv('output.csv', index=False) # Save DataFrame to CSV
```

This cheat sheet covers the essentials for data analysis in Python. Let me know if you'd like more details on any specific topic! 😊