Cheat Sheet : Data Preprocessing Tasks in Pandas

Task	Syntax	Description
Load CSV data	`pd.read_csv('filename.csv')`	Read data from a CSV file into a Pandas DataFrame
Handling Missing Values	`df.dropna()`	Drop rows with missing values
	`df.fillna(value)`	Fill missing values with a specified value
Removing Duplicates	`df.drop_duplicates()`	Remove duplicate rows
Renaming Columns	`df.rename(columns={'old_name': 'new_name'})`	Rename one or more columns
Selecting Columns	`df['column_name']` or `df.column_name`	Select a single column
	`df[['col1', 'col2']]`	Select multiple columns
Filtering Rows	`df[df['column'] > value]`	Filter rows based on a condition
Applying Functions to Columns	`df['column'].apply(function_name)`	Apply a function to transform values in a column
Creating New Columns	`df['new_column'] = expression`	Create a new column with values derived from existing o
Grouping and Aggregating	`df.groupby('column').agg({'col1': 'sum', 'col2': 'mean'})`	Group rows by a column and apply aggregate functions
Sorting Rows	`df.sort_values('column', ascending=True/False)`	Sort rows based on a column
Displaying First n Rows	`df.head(n)`	Show the first n rows of the DataFrame
Displaying Last n Rows	`df.tail(n)`	Show the last n rows of the DataFrame
Checking for Null Values	`df.isnull()`	Check for null values in the DataFrame
Selecting Rows by Index	`df.iloc[index]`	Select rows based on integer index
	`df.iloc[start:end]`	Select rows in a specified range
Selecting Rows by Label	`df.loc[label]`	Select rows based on label/index name
	`df.loc[start:end]`	Select rows in a specified label/index range
Summary Statistics	`df.describe()`	Generates descriptive statistics for numerical columns

Cheat Sheet : Plot Libraries

Library	Main Purpose	Key Features	Programming Language	Level of Customization	Dashboard Cap
Matplotlib	General-purpose plotting	Comprehensive plot types and variety of customization options	Python	High	Requires additic
Pandas	Fundamentally used for data manipulation but also has plotting functionality	Easy to plot directly on Panda data structures	Python	Medium	Can be combine creating dashbc
Seaborn	Statistical data visualization	Stylish, specialized statistical plot types	Python	Medium	Can be combine plots on dashbc
Plotly	Interactive data visualization	interactive web-based visualizations	Python, R, JavaScript	High	Dash framework interactive dash
Folium	Geospatial data visualization	Interactive, customizable maps	Python	Medium	For incorporatir integrated with
PyWaffle	Plotting Waffle charts	Waffle charts	Python	Low	Can be combine waffle chart on