Pandas Cheat Sheet for Data Analysis #1

Import pandas library

import pandas as pd

Load a CSV file into a DataFrame

df1 = pd.read_csv("Example.csv")

Meaning: reads data from a CSV file and stores it in a variable called df1

Get total number of elements (rows × columns)

print(df1.size)

Meaning: .size returns the total count of all cells in the DataFrame

View the shape of the DataFrame (rows, columns)

print(df1.shape)

Meaning: .shape returns a tuple: (number of rows, number of columns); (df1.shape[0]) means rows and (df1.shape[1]) means columns

Display data types of each column

print(df1.dtypes)

Meaning: .dtypes shows the data type (e.g., object, int64, float64) of each column

Show the first 5 rows of the DataFrame

print(df1.head(5))

Meaning: .head(n) displays the first n rows (default is 5)

Show the last 5 rows of the DataFrame

print(df1.tail(5))

Meaning: .tail(n) displays the last n rows (default is 5)

Get concise summary of the DataFrame

print(df1.info())

Meaning:.info() shows column names, non-null counts, data types, and memory usage

Count occurrences of unique values in a column

domain_counts = df1['example_column].value_counts(dropna=True)

Meaning:.value_counts() returns a Series with frequency counts of unique values ; dropna=True excludes missing (NaN) values from the count

To count combinations of multiple columns, use: df1[['column1', 'column12]].value_counts()

Always check for missing data before analysis

print(df1.isnull().sum())

Meaning: you can use df1.isnull().sum() to see how many missing values per column