Pandas

Cheat Sheet

A Beginner's Guide

Wadeed Madni
linkedin.com/in/wadeedmadni

What is Pandas?

Pandas is a powerful and flexible open-source data analysis and manipulation library for Python.

Importance and Use-Cases

Pandas is widely used in data science and analytics for its ability to handle large datasets and perform various operations on them. It is commonly used in tasks such as data cleaning, data transformation, and data exploration. Some real-life applications of Pandas include data analysis for financial forecasting, customer segmentation for marketing campaigns, and data preprocessing for machine learning models.

Wadeed Madni
linkedin.com/in/wadeedmadni

Reading
&
Writing Data

- **pd.read_csv('file.csv') :** Read a CSV file into a DataFrame
- **df.to_csv('file.csv') :** Write a DataFrame to a CSV file
- **pd.read_excel('file.xlsx') :** Read an Excel file into a DataFrame
- **df.to_excel('file.xlsx') :** Write a DataFrame to an Excel file

Wadeed Madni
linkedin.com/in/wadeedmadni

Data Inspection

- **df.head() :** Display the first 5 rows of a DataFrame
- **df.tail() :** Display the last 5 rows of a DataFrame
- **df.info() :** Display information about a DataFrame, including data types and memory usage
- **df.describe() : Display summary statistics of numerical columns in a DataFrame**

Wadeed Madni
linkedin.com/in/wadeedmadni

Data Selection

- **df[col] :** Select a single column by name as a Series
- **df[[col1, col2]] :** Select multiple columns by name as a DataFrame
- **df.loc[row, col] :** Select a single value by row and column label
- **df.iloc[row, col] :** Select a single value by row and column index

Wadeed Madni
linkedin.com/in/wadeedmadni

Data Manipulation

- **df['new_col'] = value :** Add a new column to a DataFrame
- **df.drop(col, axis=1, inplace=True) :** Remove a column from a DataFrame
- **df.drop(row, axis=0, inplace=True) :** Remove a row from a DataFrame
- **df.sort_values(by=col, ascending=True) :** Sort a DataFrame by a column

Wadeed Madni
linkedin.com/in/wadeedmadni

Grouping & Aggregation

- **df.groupby(col).sum() :** Group a DataFrame by a column and compute the sum of each group
- **df.groupby(col).median() :** Group a DataFrame by a column and compute the median of each group
- **df.groupby(col).max() :** Group a DataFrame by a column and compute the maximum of each group
- **df.groupby(col).first() :** Group a DataFrame by a column and return the first row of each group
- **df.groupby(col).size() :** Group a DataFrame by a column and return the size of each group
- **df.groupby(col).agg(func) :** Group a DataFrame by a column and apply a specific aggregation function to each group

Wadeed Madni
linkedin.com/in/wadeedmadni