

Assignment - 4

PANDAS

What is Pandas?

Pandas is a python library for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data.

The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis". and was created by wes mckinney in 2008.

Why Use Pandas?

Pandas allows us to analyze big data and make conclusions based on statistical theories.

Pandas can clean messy data sets, and make them readable and relevant. Relevant data is very important in data science.

Functions

1. Read Data

import pandas as pd

df = pd.read_csv("heart-failure.csv")

2. Head & Tail

df.head() & df.tail()

3. shape size and info

df.shape() - total no of source & then columns

df.size() - returns the number of source times
number of columns in the data
frame

df.info() - gives such as resource from
Range Index Data columns and
then data type of each column

4. Data Structures

- Series is one dimensional array like object
that can hold data of any type.

- Data frame: A two-dimensional table-like
structure with rows and columns, similar to

a spreadsheet or SQL table

8. Reading & Writing data:

Pandas can read data from various sources, such as CSV files, Excel spreadsheets, SQL databases & more. It can also write data to their format.

6. Data manipulation:

Pandas provides a wide range of operations for data manipulation, including selecting, filtering, merging, reshaping and group data.

7. Data Analysis & Exploration

- You can perform various statistical & descriptive analysis using pandas, including mean, median, standard deviation & more.
- Data Exploration tasks, such as identifying missing data or unique values, are easily accomplished with pandas.

8. Data cleaning and processing:

- You can clean & process data by handling missing values, converting data types, and applying function to columns.

9. Data visualization:

- While Pandas itself is not a data visualization library like Matplotlib and Seaborn to create plots and charts.

→ Some essential concepts & tasks you can perform with Pandas:

1. Indexing & Selection:

- Pandas provides various methods for selecting and indexing data within Data Frames. You can use `loc []` for label-based indexing and `iloc []` for integer-based indexing. You can also use Boolean indexing for conditional selection.