19/10/23 : Pandas

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

→ Import pandas as pd

Once you import it you can take your data analysis to whole new level.

→ df = pd.read_csv("data file name.csv")

Used to read the csv file imported.

→ df.head.

Find the headings from the csv file

→ df.tail

Find the last 5 heading

→ df['Name'].dtype

df.dtypes

Used to find the data types in it

→ df.columns

Used to find the number of columns and heading

→ df.shape

→ df['name'].mean()

find the mean value.

→ df.group by('name')[''] - mean()

* → df.describe().

It will give you count, mean standard deviation and also 5 number summary

→ df.nunique()

It will give all the unique values a variable counting

↔ df.anaemia.value_counts().

Panda dtype	Pythontype	Usage
Object	Str. Oxmixed	Txt a mixed numeric & non-numeric values
int 64	int	Integer numbers
Float 64	float	Floating point numbers
bool	bool	True / False values

1. Data Structure:
Series: A one-dimensional array like object that hold data of any type
2. Reading & writing Data:
Pandas can read data from various sources such as csv files, SQL database & more. It can also write data to these format
3. Data Manupulation:
Pandas provids a vaid range of operation for data manupulation, including selecting, Filtering, Merging, reshaping and grouping data.

4. Data Analysis & Exploration:
You can program various statistical & description analysis using pandas, including mean, median, standard deviation & mode.
Data Exploration tasks such as identifying missing data or unique values are usually accomplished with pandas

5. Data Cleaning & pre processing:
You can clean & preprocess data by handling missing values, converting data types and applying functions to colourns

6. Data Visulization:
While pandas itself is not a data visualization ~~library~~ library like matplotlib & seaborn to create plots and charts

Some essential concepts & tasks you can perfor with pandas

1. Identifying & Selection:
Pandas provides various methods for selecting & indexing data within Data frames. You can use '.loc[]' for label-based indexing & '.iloc[]' for integer-based indexing. Also Boolen is used for indexing conditional selection.

2. Grouping & Aggregation:
You can group data within a Data frame based

on one or more columns & the apply aggregation function like sum, mean, count, etc to these groups.

3. Sorting & Ranking:
   Data within a data frame can be sorted based on one or more columns. You can assign ranks to the data based on certain criteria

4. Merging & Joining:
   Pandas allows how to merge multiple DataFrames together using methods like 'merge()' or concat(), similar to SQL