**What is Data Science?**

**Data Science** or **Data Analytics** is a process of Analyzing large set of data points to get answers on questions related to that data set.

**Pandas** is a python module that makes **Data science** easy and effective.

**Process** of **cleaning messy data** is called **data munging** or **data wrangling**.

**Dataframe?**

**Dataframe** is a main object in pandas. It is used to represent data with rows and columns (tabular or excel spread sheet like data).

**Topics,**

* Creating Dataframe
* Dealing with rows and columns
* Operations: **min**, **max**, **std**, **describe**
* Conditional Selection
* Set\_index

**Creating Dataframe?**

**Using File**import **pandas** as **pd**

**df**=**pd**. **read\_csv(**"wh.csv"**)**

print(df)

**Using Dictionary**

**weather\_data**= {

'**day**’: ['2025/03/16','2025/03/17','2025/03/18','2025/03/19'],

'**temperature**’: [32,40,23,25],

'**windspeed**’: [6,7,8,5],

'**event**’: ['Rain','sunny','snow','cloudy']

}

df=**pd. DataFrame**(**weather\_data**)

print(df)

**shape property**

S=**df. Shape** [It returns tuple with the values of **rows**, **columns**]

print(S)

(4,4)

Rows, columns=**df. Shape**

print(Rows)

4

Print(columns)

4

**head()**

It returns initial starting few rows in a table.

**df**=**pd. read\_csv**('wh.csv')

s=**df.head**()

print(s)

**To print two rows in a table**

**df**=**pd. read\_csv**('wh.csv')

s=**df. head** (2)

print(s)

**tail()**

It returns last few rows in a table.

s=df.tail()

print(s)

To return last two rows in a table.

s=df.tail(2)

print(s)

**Indexing**

df is a data structure which supports the indexing numbers to access its data.

df=pd.read\_csv('wh.csv')

print(df[1:4])

**columns**

It returns the no of columns present in a data frame.

df=pd.read\_csv('wh.csv')

print(type(df.columns))

**To Access particular column data**

df=pd.read\_csv('wh.csv')

print(**df['Date']**)

(or)

print(**df.Date**)

**max**

**df**['Temperature']. **max ( )**

**min**

**df**['Temperature']. **min ( )**

**mean**

**df**['Temperature']. **mean ( )**

**std**

**df**['Temperature']. **std ( )**

**Describe**

**df**. **describe()**

**index**

df.index

**set\_index()**

df.set\_index('Date')

**reset\_index()**

df.reset\_index(inplace=True)

**Methods & properties:** -

* read\_csv()
* DataFrame()
* Shape
* head()
* tail()
* columns
* max()
* min()
* mean()
* std()
* describe()
* index
* set\_index()
* loc[]
* reset\_index()

**Different ways of creating dataframe,**

**Topics,**

* Using CSV
* Using Excel
* From python dictionary
* From list of tuples
* From list of dictionaries

**Topics,**

* Read CSV
* Write CSV
* Read Excel
* Write Excel