DAY-3(FEB-14) U.Hasini

22A81A0658

SOC\_2

COURSE:ADVANCED PYTHON

**NOTES:TOPICS COVERED**

#To access rows based on index position.

#To get specific value of co\_ordinate in a data frame.

-> SYNTAX-->df.ioc[row,column]

#Retrieve data based on condition

#To get all values greather than 0

#To add specific column

#To handle missing data

#Create a dictionary

#To create dataframe

#To drop the values with Nan

#We use dropna() -->it deletes all rows with Nan

->dropna(how, thresh,axis,inplace)

#how="all" or "any" #if all -nan is there then only it will delete the rows

->It will be checking the rows with all Nan value to delete the row

->It will delete the rows which have even single Nan value

# Operating based on columns

#To check the sum of Nan values in each columns

#To check the sum of Nan values in each columns

#Thresh -->breaking point

-->min 2 value non nan values should be there in a row

#Based on column

# fillna()--->to avoid deleting wanting data

🡪#dummy values are added (0)

🡪#bydefault it takes 0.0 values

# To store average in the above set

#Group by

#To group based on column and perform aggregate functions

#Groupby based on aggregrate func

#To get the maximum salary person from each company

#Describe func:it gives mean,median,standard deviation..

#To access a csv-file

#Header-->is used to mention the column names ofthe dataframes

#To get first five rows

#To print(sep,end,file,finish)

#To get overall data

->use info()

#Matplotlib:

->It is data visualisation library ,which was inspired by matlab

->It is used for data visualisation in the form of various plots

->import matplotlib.pyplot as plt

#Bar plot,scatter plot ,hist plot,box plot

#seaborn:

->import seaborn as sns.

* Seaborn is a library for making statistical graphics in Python.
* It builds on top of [matplotlib](https://matplotlib.org/) and integrates closely with [pandas](https://pandas.pydata.org/) data structures.
* Seaborn helps you explore and understand your data.
* Its plotting functions operate on dataframes and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots.
* Its dataset-oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them.