WEEK 5 DAY – 1

19/09/2023

Nabiha Khan

Python

- args (*)
- Do not know how many arguments to add
- Inside a function -> behavior
- kwargs (**)
- Keyword arguments -> key -> value
- Like dictionary
- Number not defined
- Keyword & Positional Parameter
- Keyword-only argument
- Position-only argument

Keyword-only	Position-only
Parameter names are used to pass the arguments during function call	We need to maintain the order of the passing arguments. The arguments should be defined before passing parameters
Order of the parameters can be changed while passing the information	Order must not be changed
function_name(parametername="value")	function_name(value1,value2)

OOPs

- Abstraction (.exe)
- Abstract it on –computer
- Encapsulation
- In a capsule
- Inheritance
- parent-to-child
- Polymorphism
- Many forms

Class

- Service to serve the requirements
- Uses the appropriate functions, objects, and variables to serve the requirements

List

- Store multiple items in a single variable
- Built-in data type (list, tuple, set, dictionary)
- Square brackets
- Allows duplicates since it's indexed
- List length len()
- List items – data types
- Multiple data types

Python – Access List Items

- Access Items – referring to the index number
- Negative indexing – start from the end
- Range of indexes – [n:m] a new list with new start & end
- Sorting list
- zip() function – used for tuples
- Enumerate function

Python – Access List Items

- Lambda

- Anonymous function (=> :)

DAY – 2

20/09/2023

Nabiha Khan

NUMPY

- Library
- Primary use for numerical computations
- Data science & Data analytics
- Extension of list
- N-dimensional arrays (32)
- type(a)
- a.ndim
- a.shape
- a.size
- a.dtype
- np.clip
- astype – type conversion
- np.sin, np.cos, np.tan
- np.round(np.sin())
- np.argmax() – index of max
- np.power

NUMPY

- np.arange
- np.abs – only positive values return (mod)
- np.exp – exponential
- np.pi
- np.linspace
- np.unique(a,return_counts=True)
- np.ones(5) – array of 5 1's
- np.random.randint(0,10,2) – 2 random values from 0-10
- np.random.seed – initialize and fix the random
- Transpose of a matrix = .T

MatPlotLib

- Line Graph

- Color

- Pie Chart

- Bar Graph

- Horizontal Bar graph

- Directed Line Graphs

Pandas

- Line Graph
- Color
- Pie Chart
- Bar Graph
- Horizontal Bar graph
- Directed Line Graphs

DAY – 3

Nabiha Khan

PySpark

- Handles both batch and stream data
- Supports in-memory calculations
- Transformations – RAM -> Speed

- Basic Architecture
- Driver Program – 1
- Worker Node – N
- Cluster Manager – Connector

- Deployment Modes
- Client Mode – driver is outside the cluster
- Cluster Mode – driver is inside the cluster

PySpark

- Spark's Toolset

- RDD (Resilient Distributed Dataset) -> immutable collection of data->fault tolerant ->

- DataFrame --> looks like a table but it's not a table

- Spark Context --> entry point of spark functionality -> communicate with cluster & create RDD -> only 1 spark context per JVM

- Spark Session -> overcomes the problem of 1 sc per JVM -> create rdd, dataframe, & dataset

PySpark

- DataFrame functions
- Select