

DAL

1. What is Python? What are the benefits of using Python?

AND – (1) Python is a popular programming language.

(2) Python supports modules and packages, which encourages program modularity and code reuse.

(3) The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms,

and can be freely distributed.

It is used for:

- web development (server-side),
- software development,
- mathematics,
- system scripting.

2 Explain libraries.

A Python library is a **collection of related modules**. It contains bundles of code that can be used repeatedly in different programs. It makes Python Programming simpler and convenient for the programmer. As we don't need to write the same code again and again for different programs.

[“There are a lot of python libraries which could be used to build

visualization like matplotlib, vispy, bokeh, seaborn, pygal, folium, plotly,

cufflinks, and networkx. Of the many, matplotlib and seaborn seems to be

very widely used for basic to intermediate level of visualizations.”]

3 Define:

- (1) data analytics

Data Analysis. Data Analysis is **the process of systematically applying statistical and/or logical techniques to describe and evaluate data.**

- (Steps of data analysis - Step One: Ask The Right Questions. So you're ready to get started. ...
- Step Two: Data Collection. This brings us to the next step: data collection. ...
- Step Three: Data Cleaning. ...
- Step Four: Analyzing The Data. ...
- Step Five: Interpreting The Results.)

2. Python Packages for data science

Procedure:

The Python community has adopted a number of naming conventions for

commonly used modules:

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
import pandas as pd
```

```
import seaborn as sns
```

```
import statsmodels as sm
```

```
(pandas, numpy,scipy,matplotlib,seaborn,scikit  
learn,statsmodels,ploty,)
```

```
[1]. pandas
```

- (1) Pandas is an open-source Python package that provides high-performance, easy-to-use data structures and data analysis
 - (2) Pandas stand for Python Data Analysis Library.
 - (3) Pandas is a perfect tool for data wrangling or munging. It is designed for quick and easy data manipulation, reading, aggregation, and visualization.
 - (4) Applications of pandas -
 - 1. Indexing, manipulating, renaming, sorting, merging data frame
 - 2. Update, Add, Delete columns from a data frame
 - 3. Impute missing files, handle missing data or NaNs
 - 4. Plot data with histogram or box plot
- import pandas as pd**

[2]. NumPy

- (1) One of the most fundamental packages in Python, NumPy is a general-purpose array-processing package.
- (2) It provides high-performance multidimensional array objects
- (3) NumPy is used to process arrays that store values of the same datatype.
- (4) Applications of NumPy
 - 1. Basic array operations: add, multiply, slice, flatten, reshape, index arrays
 - 2. Advanced array operations: stack arrays, split into sections, broadcast arrays

3. Work with DateTime or Linear Algebra

4. Basic Slicing and Advanced Indexing in NumPy Python

import numpy as np

[3] SciPy

(1) The SciPy library is one of the core packages that make up the SciPy stack.

(2) SciPy builds on the NumPy array object and is part of the stack which includes tools like Matplotlib, Pandas, and SymPy with additional tools,

(3) SciPy library contains modules for efficient mathematical routines

as linear algebra, interpolation, optimization, integration,

(4) SciPy uses arrays as its basic data structure. It has various modules

to perform common scientific programming

“Integration, calculus, ordinary differential equations, and signal processing.”

[5] Matplotlib

(1) Matplotlib is the plotting library for Python that provides an object-oriented API for embedding plots into applications.

(2) close resemblance to MATLAB embedded in Python programming.

(3) Applications

1. Line plots

2. Scatter plots

3. Area plots

4. Bar charts and Histograms

5. Pie charts

6. Stem plots

7. Contour plots

8. Quiver plots

9. Spectrograms

Matplotlib also facilitates labels, grids, legends, and some more formatting entities with Matplotlib. Basically, everything that can be drawn!

3. Data Wrangling

- (1) Data wrangling is the process of cleaning and unifying messy and complex
- (2) data sets for easy access and analysis
- (3) This process
- (4) typically includes manually converting and mapping data from one raw form into another format to allow for more convenient consumption and organization of the data.
- (5) Applications

1. Merging multiple data sources into a single dataset for analysis
2. Identifying gaps in data (for example, empty cells in a spreadsheet) and
 - either filling or deleting them
3. Deleting data that's either unnecessary or irrelevant to the project you're working on
4. Identifying extreme outliers in data and either explaining the discrepancies or removing them so that analysis can take place

4. Data Visualization

- (1) Data visualization is an easier way of presenting the data, however complex it is, to analyze trends and relationships
- (2) advantages of Data Visualization:-
 - Easier representation of compels data
 - Highlights good and bad performing areas
 - Explores relationship between data points
 - Identifies data patterns even for larger data points

While building visualization, it is always a good practice to keep some below mentioned points in mind

 - Ensure appropriate usage of shapes, colors, and size while building visualization
 - Plots/graphs using a co-ordinate system are more pronounced
 - Knowledge of suitable plot with respect to the data types brings more clarity to the information
 - Usage of labels, titles, legends and pointers passes seamless information the wider audience

5. Descriptive Statistics

(1) **The study of numerical and graphical ways to describe and display your data** is called descriptive statistics.

6. Steps for Importing and Exporting Data in Python

1- Excel File:

2- Capture the file path

3- Apply the Python code

4- Run the Python code to import the Excel file

7. Steps for Analyzing data in python

8. Steps for Data Visualization in Python

1. Step 1: Importing Matplotlib in your Jupiter Notebook. ...
2. Step 2: Create Simple Line Function. ...
3. Step 3: Visualize Your Function. ...
4. Step 4: Working with Title, Label, and Legend. ...
5. Step 4: Saving Your Plot.

9. Steps for making Group in python

10. Step 1 - Import the library.
11. Step 2 - Setting up the Data.
12. Step 3 - Grouping Rows.

10.Numeric Types

Ans: integers, floating-point numbers, and complex numbers.

11. What is `__init__`?

- (1) **The `__init__` method is the Python equivalent of the C++ constructor in an object-oriented approach.**
- (2) **The `__init__` function is called every time an object is created from a class.**
- (3) **The `__init__` method lets the class initialize the object's attributes and serves no other purpose. It is only used within classes.**

13. What is slicing in Python?

- (1) **Slicing in Python is a function that allows us to access parts of sequences like strings, tuples, lists, arrays, etc.**
- (2) **We can specify where to start the sequence and where to end it.**

14. Define pandas dataframe.

Ans- `dataframe = pd.DataFrame(data, index, columns, dtype)`

15. How to add new column to pandas dataframe?

- (1) **Method #1: By declaring a new list as a column.**
- (2) **Method #2: By using `DataFrame.insert()`. It gives the freedom to add a column at any position we like and not just at the end. It**

also provides different options for inserting the column values.

- (3) **Method #3: Using `Dataframe.assign()` method**
This method will create a new dataframe with a new column added to the old dataframe.

- (4) **Method #4: By using a dictionary**
We can use a Python dictionary to add a new column in pandas DataFrame. Use an existing column as the key values and their respective values will be the values for a new column.

16. How will you delete indices, rows and columns from a dataframe?

Example #1: Dropping Rows by index label

In his code, A list of index labels is passed and the rows corresponding to those labels are dropped using `.drop()` method.

```
# importing pandas module
import pandas as pd

# making data frame from csv file
data = pd.read_csv("nba.csv", index_col ="Name" )

# dropping passed values
data.drop(["Avery Bradley", "John Holland", "R.J.
Hunter",
           "R.J. Hunter"], inplace
= True)
```

```
# display  
data
```

17. What do you understand by NumPy?

- (1) **NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Numerical Python.**

- (2) **The NumPy library is a popular Python library used for scientific computing applications, and is an acronym for "Numerical Python". NumPy's operations are divided into three main categories: Fourier Transform and Shape Manipulation, Mathematical and Logical Operations, and Linear Algebra and Random Number Generation**

18. Define PIP.

PIP is a package manager for Python packages, or modules if you like. Note: If you have Python version 3.4 or later, PIP is included by default.

AND IF WE NEEDS UP TO DATE OUR LIBRARIES

WE USE ONE COMMAND - pip install pandas-profiling // Jupiter library installation command

19. Sampling Theorem

20. Explain Probability with types

- (1) Probability is the branch of mathematics concerning the occurrence of a random event
- (2) four main types of probability exist: **classical, empirical, subjective and axiomatic.**
- (3) Probability is synonymous with possibility, so you could say it's the possibility that a **particular event will happen.**

21. Explain linear and multiple regression

Multiple regression is like [linear regression](#), but with more than one independent value, meaning that we try to predict a value based on two or more variables.

1.In Python we have modules that will do the work for us. Start by importing the Pandas module.

2. Then make a list of the independent values and call this variable `x`.

Put the dependent values in a variable called `y`.

```
3. from sklearn import linear_model
```

22. What is regression