**Numpy**

**MCQs**

1. B. Functions
2. D. Multi-dimensional
3. C. Evenly
4. D. Slicing
5. A. Ravel

**True or false**

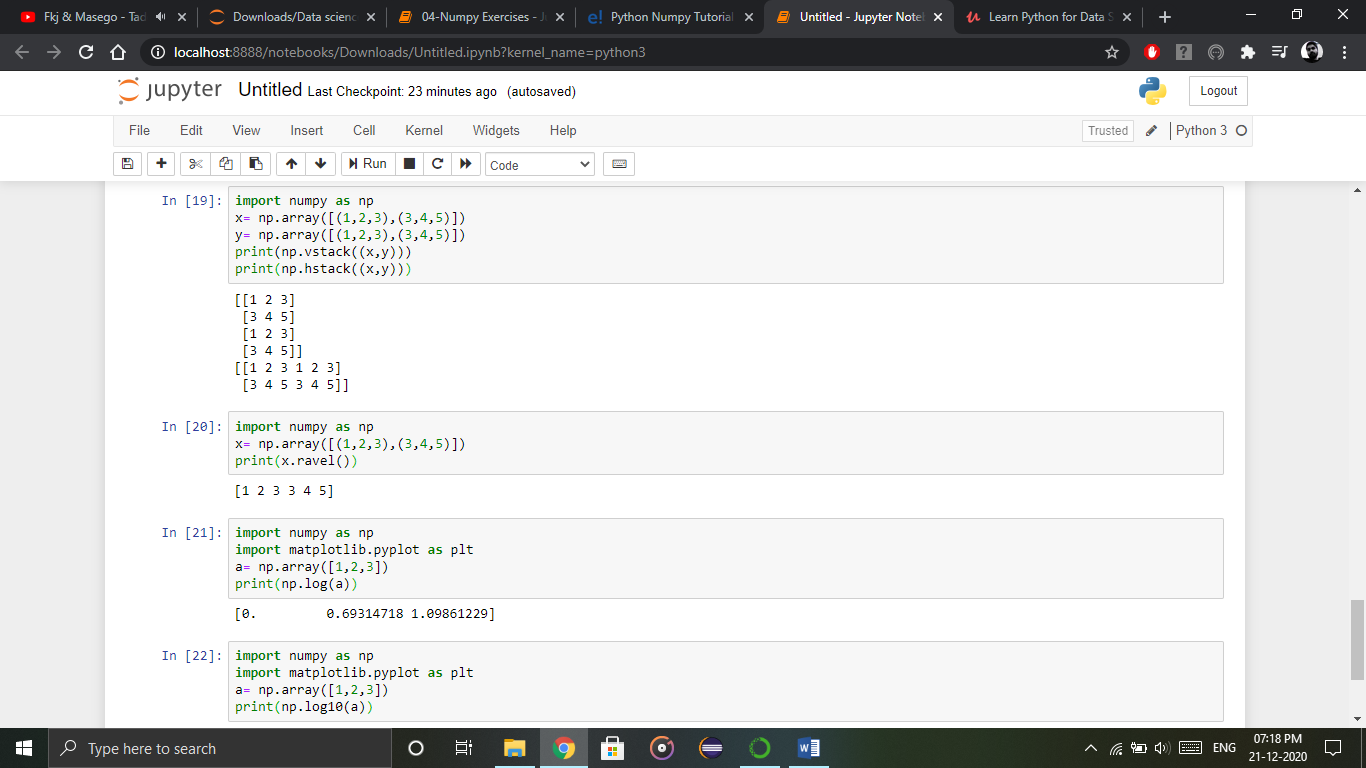
1. True
2. False
3. False
4. True
5. False

**Fill in the blanks**

1. Memory
2. Execution
3. List
4. Itemsize
5. Ravel

**Answer in short**

1. **NumPy** is a Python package that stands for ‘Numerical Python’. It is the core library for scientific computing, which contains a powerful n-dimensional array object.
2. Python NumPy arrays provide tools for integrating C, C++, etc. It is also useful in linear algebra, random number capability etc. NumPy array can also be used as an efficient multi-dimensional container for generic data. Now, let me tell you what exactly is a Python NumPy array.
3. To **install** Python NumPy, go to your command prompt and type “pip install numpy”. Once the installation is completed, go to your IDE (For example: Jupyter Notebook) and simply import it by typing: “import numpy as np”
4. The very first reason to choose python NumPy array is that it occupies **less memory** as compared to list. Then, it is pretty fast in terms of execution and at the same time, it is very convenient to work with NumPy. So these are the major advantages that Python NumPy array has over list. Don’t worry, I am going to prove the above points one by one practically in Jupyter Notebook
5. **Vertical\_&\_Horizontal\_Stacking**  
   Next, if you want to concatenate two arrays and not just add them, you can perform it using two ways – *vertical stacking* and *horizontal stacking*.



Output – [[1 2 3] [3 4 5] [1 2 3] [3 4 5]]  
[[1 2 3 1 2 3] [3 4 5 3 4 5]]

**Viva Questions**

1. What is Numpy?

Ans: NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. … A powerful N-dimensional array object. Sophisticated (broadcasting) functions.

1. Why NumPy is used in Python?

Ans: NumPy is a package in Python used for Scientific Computing. NumPy package is used to perform different operations. The ndarray (NumPy Array) is a multidimensional array used to store values of same datatype. These arrays are indexed just like Sequences, starts with zero.

1. What does NumPy mean in Python?

Ans: NumPy (pronounced /ˈnʌmpaɪ/ (NUM-py) or sometimes /ˈnʌmpi/ (NUM-pee)) is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

1. Where is NumPy used?

Ans: NumPy is an open source numerical Python library. NumPy contains a multi-dimentional array and matrix data structures. It can be utilised to perform a number of mathematical operations on arrays such as trigonometric, statistical and algebraic routines. NumPy is an extension of Numeric and Numarray.

1. How to import numpy in python?

Ans:

import numpy as np

1. How to create 1D Array ?

Ans:

num=[1,2,3]

num = np.array(num)

print(“1d array : “,num)

1. How to create 2D Array ?

Ans:

num2=[[1,2,3],[4,5,6]]

num2 = np.array(num2)

print(“\n2d array : “,num2)

**Pandas**

**MCQs**

* 1. D. Data
  2. A. Merging & Joining
  3. C. Concatenate
  4. A. Dataframe
  5. C. Pandas

**True or false**

1. False
2. True
3. True
4. True
5. False

**Fill in the blanks**

1. heterogeneously-typed
2. pandas
3. dataframes
4. joining
5. Concatenation

**Answer in short**

1. Pandas is used for data manipulation, analysis and cleaning. Python pandas is well suited for different kinds of data, such as:

* Tabular data with heterogeneously-typed columns
* Ordered and unordered time series data
* Arbitrary matrix data with row & column labels
* Unlabelled data
* Any other form of observational or statistical data sets

1. To install Python Pandas, go to your command line/ terminal and type “pip install pandas” or else, if you have anaconda installed in your system, just type in “conda install pandas”. Once the installation is completed, go to your IDE (Jupyter, Jupyter Notebook etc.) and simply import it by typing: “import pandas as pd”
2. import pandas as pd

df=pd.DataFrame({"Day":[1,2,3,4],"Visitors":[200,100,230,300],"Bounce\_Rate":[20,45,60,10]})

df.set\_index("Day", inplace= True)

print(df)

1. Python Pandas operations are:

* Slicing the Data Frame
* Merging & Joining
* Concatenation
* Change the index
* Change the Column Headers

1. df1=pd.DataFrame({"HPI":[80,90,70,60],"Int\_Rate":[2,1,2,3],"IND\_GDP":[50,45,45,67]}, index=[2001, 2002,2003,2004])

df2=pd.DataFrame({"HPI":[80,90,70,60],"Int\_Rate":[2,1,2,3],"IND\_GDP":[50,45,45,67]}, index=[2005, 2006,2007,2008])

concat= pd.concat([df1,df2])

print(concat)

**Viva Questions**

1. Define the Pandas/Python pandas?

Pandas is defined as an open-source library that provides high-performance data manipulation in Python. The name of Pandas is derived from the word Panel Data, which means an Econometrics from Multidimensional data. It can be used for data analysis in Python and developed by Wes McKinney in 2008. It can perform five significant steps that are required for processing and analysis of data irrespective of the origin of the data, i.e., load, manipulate, prepare, model, and analyze.

1. Mention the different types of Data Structures in Pandas?

Pandas provide two data structures, which are supported by the pandas library, Series, and DataFrames. Both of these data structures are built on top of the NumPy.

A Series is a one-dimensional data structure in pandas, whereas the DataFrame is the two-dimensional data structure in pandas.

1. Define Series in Pandas?

A Series is defined as a one-dimensional array that is capable of storing various data types. The row labels of series are called the index. By using a 'series' method, we can easily convert the list, tuple, and dictionary into series. A Series cannot contain multiple columns.

1. Define DataFrame in Pandas?

A DataFrame is a widely used data structure of pandas and works with a two-dimensional array with labeled axes (rows and columns) DataFrame is defined as a standard way to store data and has two different indexes, i.e., row index and column index. It consists of the following properties:

The columns can be heterogeneous types like int and bool.

It can be seen as a dictionary of Series structure where both the rows and columns are indexed. It is denoted as "columns" in the case of columns and "index" in case of rows.

1. What are the significant features of the pandas Library?

The key features of the panda's library are as follows:

* Memory Efficient
* Data Alignment
* Reshaping
* Merge and join
* Time Series

1. How will you create an empty DataFrame in Pandas?

A DataFrame is a widely used data structure of pandas and works with a two-dimensional array with labeled axes (rows and columns) It is defined as a standard way to store data and has two different indexes, i.e., row index and column index.

Create an empty DataFrame:

The below code shows how to create an empty DataFrame in Pandas:

# importing the pandas library

import pandas as pd

info = pd.DataFrame()

print (info)

Output:

Empty DataFrame

Columns: []

Index: []

1. How to Delete Indices, Rows or Columns From a Pandas Data Frame?

**Deleting an Index from Your DataFrame**

If you want to remove the index from the DataFrame, you should have to do the following:

Reset the index of DataFrame.

Executing del df.index.name to remove the index name.

Remove duplicate index values by resetting the index and drop the duplicate values from the index column.

Remove an index with a row.

**Deleting a Column from Your DataFrame**

You can use the drop() method for deleting a column from the DataFrame.

The axis argument that is passed to the drop() method is either 0 if it indicates the rows and 1 if it drops the columns.

You can pass the argument inplace and set it to True to delete the column without reassign the DataFrame.

You can also delete the duplicate values from the column by using the drop\_duplicates() method.

**Removing a Row from Your DataFrame**

By using df.drop\_duplicates(), we can remove duplicate rows from the DataFrame.

You can use the drop() method to specify the index of the rows that we want to remove from the DataFrame.

1. What is Time Series in Pandas?

The Time series data is defined as an essential source for information that provides a strategy that is used in various businesses. From a conventional finance industry to the education industry, it consists of a lot of details about the time.

Time series forecasting is the machine learning modeling that deals with the Time Series data for predicting future values through Time Series modeling.

**Matplotlib**

**MCQs**

1. B. 2-D
2. C. Library
3. A. Cartopy
4. B. Default
5. C. Mplot3D

**True or False**

1. False
2. True
3. False
4. True
5. False

**Fill in the blanks**

1. Proportional
2. APIs
3. Natgrid
4. Bin
5. Data

**Answer in short**

1. **matplotlib.pyplot** is a plotting library used for 2D graphics in python programming language. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits.
2. Matplotlib is a Python Library used for plotting, this python library provides and objected-oriented APIs for integrating plots into applications.
3. Matplotlib is not a part of the Standard Libraries which is installed by default when Python, there are several toolkits which are available that extend python matplotlib functionality. Some of them are separate downloads, others can be shipped with the matplotlib source code but have external dependencies.

* **Basemap**: It is a map plotting toolkit with various map projections, coastlines and political boundaries.
* **Cartopy**: It is a mapping library featuring object-oriented map projection definitions, and arbitrary point, line, polygon and image transformation capabilities.
* **Excel tools**: Matplotlib provides utilities for exchanging data with Microsoft Excel.
* **Mplot3d**: It is used for 3-D plots.
* **Natgrid**: It is an interface to the natgrid library for irregular gridding of the spaced data.

1. A bar graph uses bars to compare data among different categories. It is well suited when you want to measure the changes over a period of time. It can be represented horizontally or vertically. Also, the important thing to keep in mind is that longer the bar, greater is the value. Now, let us practically implement it using python matplotlib.
2. Histograms are used to show a distribution whereas a bar chart is used to compare different entities. Histograms are useful when you have arrays or a very long list.

**Viva Questions**

1. What Is Python Matplotlib?

matplotlib.pyplot is a plotting library used for 2D graphics in python programming language. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits.

1. What is Matplotlib used for?

Matplotlib is a Python Library used for plotting, this python library provides and objected-oriented APIs for integrating plots into applications.

1. Is Matplotlib Included in Python?

Matplotlib is not a part of the Standard Libraries which is installed by default when Python, there are several toolkits which are available that extend python matplotlib functionality. Some of them are separate downloads, others can be shipped with the matplotlib source code but have external dependencies.

* **Basemap**: It is a map plotting toolkit with various map projections, coastlines and political boundaries.
* **Cartopy**: It is a mapping library featuring object-oriented map projection definitions, and arbitrary point, line, polygon and image transformation capabilities.
* **Excel tools**: Matplotlib provides utilities for exchanging data with Microsoft Excel.
* **Mplot3d**: It is used for 3-D plots.
* **Natgrid**: It is an interface to the natgrid library for irregular gridding of the spaced data.

1. Explain Python Matplotlib Bar Graph.

A bar graph uses bars to compare data among different categories. It is well suited when you want to measure the changes over a period of time. It can be represented horizontally or vertically. Also, the important thing to keep in mind is that longer the bar, greater is the value. Now, let us practically implement it using python matplotlib.

1. What are histograms?

Histograms are used to show a distribution whereas a bar chart is used to compare different entities. Histograms are useful when you have arrays or a very long list. Let’s consider an example where I have to plot the age of population with respect to bin. Now, bin refers to the range of values that are divided into series of intervals. Bins are usually created of the same size. In the below code, I have created the bins in the interval of 10 which means the first bin contains elements from 0 to 9, then 10 to 19 and so on.

1. What is Python Scatter Plot?

Usually we need scatter plots in order to compare variables, for example, how much one variable is affected by another variable to build a relation out of it. The data is displayed as a collection of points, each having the value of one variable which determines the position on the horizontal axis and the value of other variable determines the position on the vertical axis.

1. What is area plot?

Area plots are pretty much similar to the line plot. They are also known as stack plots. These plots can be used to track changes over time for two or more related groups that make up one whole category.

1. What is a pie chart?

A pie chart refers to a circular graph which is broken down into segments i.e. slices of pie. It is basically used to show the percentage or proportional data where each slice of pie represents a category.