**NUMPY & PANDAS DOCUMENTATION**

**Created by:**

**Srinidhi Devan**

1. **NUMPY**:

* Python’s package to do math and is more advanced than +/-/\* etc.,
* Special functions: Cosine, Exponential, Square Root etc., are included.
* It is also useful in generating samples/arrays from many types of random variables.
* It also has powerful data types to define vectors, matrices and tensors.



1. **NumPy Basics**:



1. **Scalar Operations on NumPy:**



1. **NumPy Array Indexes:**





1. **Functions of NumPy Array:**



1. **Statistical Processing & Sketching Graph:**

Matplotlib is used for Visualization purposes in Python.





1. **Conditional Clauses & Boolean Operations:**

* We impose certain conditions to get the desired output.
* Disadvantage: The statements gets longer with the increase in the number of variables and arrays.
* To tackle this problem, we use ‘np.where()’ function.

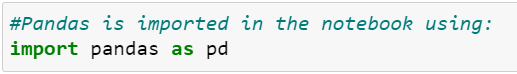


**Additional Note:**



1. **PANDAS**:

* It is a high-level data manipulation tool.
* It is built on NumPy library and extends the functionality of it.
* It has 2 main components: Series & DataFrame.



1. **Pandas Basics- SERIES:**



1. **Pandas Basics- DATAFRAME:**

* 2-D data structure with three principal components - Rows, Columns & Values.
* Similar to a table in excel sheets; Rows & Columns are labelled.
* Size of a dataframe can be changed anytime.



* Combining DataFrame in Pandas are done using: Concat, Join and/or Merge.

1. **Index in Pandas:**

* In both Series & DataFrame structure of Pandas, we use Index to refer to the row & column which is crucial in data analytics.



* ‘reindex() is done using:

1. ‘fill\_value’ method: Wherein values can be filled with a given number
2. ‘ffill’ method: Forward fill is an auto-fill method
3. **Dropping Entries in Series & DataFrame:**



1. **Handling Null/NaN Values in Pandas:**



1. **Selecting & Modifying Entries in Pandas:**

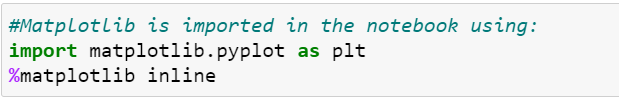


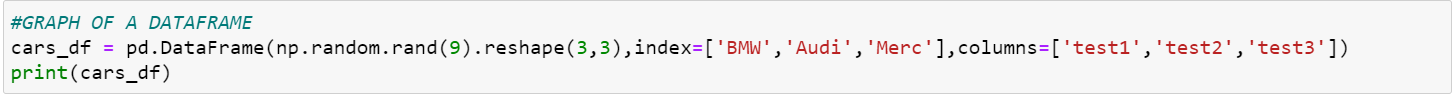
* DataFrame is accessed using ‘.loc()’ method.

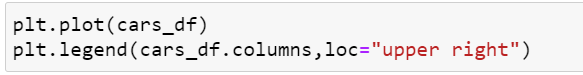
1. **Ranking & Sorting in Pandas Series:**
2. ‘sort\_index()’: This will facilitate sorting in ascending order
3. ‘sort\_values()’: To sort values in ascending order
4. ‘.rank()’: To rank the observations
5. **Graphs in Pandas DataFrame:**

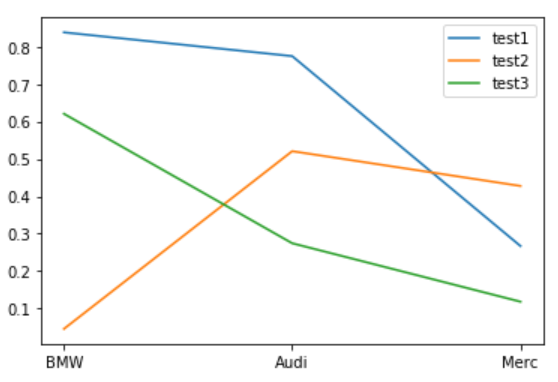
**Matplotlib:**

* Visualization tool in Python that contains various types of graphs for numerical as well as categorical variables.
* Pyplot is a sub element of Matplotlib.









1. **Statistics in Pandas DataFrame:**



**Additional Note:**

