* INTRODUCTION

PYTHON LANGUAGE

[Python](https://www.geeksforgeeks.org/python-programming-language/) is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code.

**LANGUAGE FEATURES**

* **Interpreted**
  + There are no separate compilation and execution steps like C and C++.
  + Directly *run* the program from the source code.
  + Internally, Python converts the source code into an intermediate form called bytecodes which is then translated into native language of specific computer to run it.
  + No need to worry about linking and loading with libraries, etc.
* **Platform Independent**
  + Python programs can be developed and executed on multiple operating system platforms.
  + Python can be used on Linux, Windows, Macintosh, Solaris and many more.
* **Free and Open Source**
* Redistributable
* **High-level Language**
  + In Python, no need to take care about low-level details such as managing the memory used by the program.
* **Simple**
  + Closer to English language
  + Easy to Learn
  + More emphasis on the solution to the problem rather than the syntax
* **Robust**
  + Exceptional handling features
  + Memory management techniques in built

**Softwares making use of Python**

Python has been successfully embedded in a number of software products as a scripting language.

1. [GNU Debugger](https://en.wikipedia.org/wiki/GNU_Debugger) uses Python as a **pretty printer** to show complex structures such as C++ containers.
2. Python has also been used in artificial intelligence
3. Python is often used for **natural language processing** tasks.

**Advantages of Python:**

* Presence of third-party modules
* Extensive support libraries(NumPy for numerical calculations, Pandas for data analytics etc)
* Open source and community development
* Easy to learn
* User-friendly data structures
* High-level language
* Dynamically typed language
* Object-oriented language
* Portable and Interactive
* Portable across Operating systems

**Applications :**

* GUI based desktop applications(Games, Scientific Applications)
* Web frameworks and applications
* Enterprise and Business applications
* Operating Systems
* Language Development
* Prototyping

**Organizations using Python :**

* Google(Components of Google spider and Search Engine)
* Yahoo(Maps)
* YouTube
* Mozilla
* Dropbox
* Microsoft
* Cisco
* Spotify
* Quora

**WHY PYTHON IS POPULAR?**

There are many factors that contribute to the success of Python. One of these is its popularity in web development. Many top companies such as Google, Facebook, Mozilla, Quora, etc. use Python Web Framework. Another well-known factor for the rapid growth of Python is its use in Artificial Intelligence.

**1. Python is Easy To Use**

Python is simple with an easily readable syntax and that makes it well loved by both seasoned developers and experimental students. In addition to this, Python is also supremely efficient. It allows developers to complete more work using fewer lines of code.

**2. Python has a Supportive Community**

Python has been around since 1990 and that is ample time to create a supportive community. Because of this support, Python learners can easily improve their knowledge, which only leads to increasing popularity. And that’s not all! There are many resources available online to promote Python, ranging from official documentation to YouTube tutorials that are a big help for learners.

**3. Python has multiple Libraries and Frameworks**

Python is already quite popular and consequently, it has hundreds of different libraries and frameworks that can be used by developers. These libraries and frameworks are really useful in saving time which in turn makes Python even more popular.

**4. Python has Corporate Support**

Many top companies such as Google, Facebook, Mozilla, Amazon, Quora, etc use Python for their products. In fact, Google has practically adopted Python for many of its platforms and applications. It also provides various guides and tutorials for working with Python in [Google’s Python Class](https://developers.google.com/edu/python/).

**5. Python is used in Big Data and Machine Learning**

Big Data and Machine Learning are the hottest trends in modern times. And Python is used for much of the research and development in these fields. Python is also crucial in related fields such as Artificial Intelligence. So it stands to reason that this would be a huge factor in the rapid growth of Python.

**6. Python is used in Web Development**

Python is very popular in Web development. It is easy enough to learn and yet capable of powering some of the most popular sites in the world like Spotify, Instagram, Pinterest, Mozilla, Yelp, etc. (It’s a win-win situation!!!).

**7. Python is used in Academics**

Python is such a fundamental part of the programming world currently that it is even taught in schools and colleges as a core language requirement. This is because Python is in trend with its myriad uses in data science, machine learning, deep learning, artificial intelligence, etc. And since so many of the current students are learning Python, it is obvious that its importance will increase even more in the future.

**Python Numpy**

**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.  
Besides its obvious scientific uses, Numpy can also be used as an efficient multi-dimensional container of generic data.

**ARRAY IN NUMPY**

Array in Numpy is a table of elements (usually numbers), all of the same type, indexed by a tuple of positive integers. In Numpy, number of dimensions of the array is called rank of the array.A tuple of integers giving the size of the array along each dimension is known as shape of the array. An array class in Numpy is called as **ndarray**. Elements in Numpy arrays are accessed by using square brackets.

**Python Pandas**

**Pandas**  is two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns. Pandas DataFrame consists of three principal components, the **data, rows,** and**columns.**

#### **Creating a Pandas DataFrame**

A Pandas DataFrame will be created by loading the datasets from existing storage, storage can be SQL Database, CSV file, and Excel file. Pandas DataFrame can be created from the lists, dictionary, and from a list of dictionary etc.

## **Key Features of Pandas**

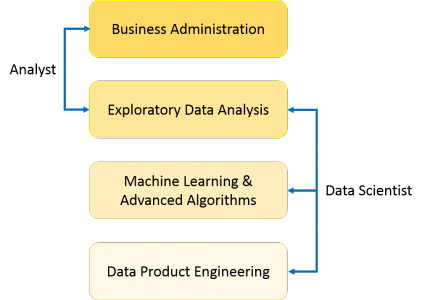
* Fast and efficient DataFrame object with default and customized indexing.
* Tools for loading data into in-memory data objects from different file formats.
* Data alignment and integrated handling of missing data.
* Reshaping and pivoting of date sets.
* Label-based slicing, indexing and subsetting of large data sets.
* Columns from a data structure can be deleted or inserted.
* Group by data for aggregation and transformations.
* High performance merging and joining of data.
* Time Series functionality.

## **WHAT IS DATA SCIENCE??**

Data Science is increasingly common*,*but what does it exactly mean? What skills do you need to become Data Scientist? What is the difference between BI and Data Science? How are decisions and predictions made in Data Science? These are some of the questions that will be answered further.

First, let’s see what is Data Science. Data Science is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the raw data. How is this different from what statisticians have been doing for years?

The answer lies in the difference between explaining and predicting.



 Data Science is primarily used to make decisions and predictions making use of predictive causal analytics, prescriptive analytics (predictive plus decision science) and machine learning.

* **Predictive causal analytics :**If you want a model which can predict the possibilities of a particular event in the future, you need to apply predictive causal analytics. Say, if you are providing money on credit, then the probability of customers making future credit payments on time is a matter of concern for you. Here, you can build a model which can perform predictive analytics on the payment history of the customer to predict if the future payments will be on time or not.
* **Prescriptive analytics:** If you want a model which has the intelligence of taking its own decisions and the ability to modify it with dynamic parameters, you certainly need prescriptive analytics for it. This relatively new field is all about providing advice. In other terms, it not only predicts but suggests a range of prescribed actions and associated outcomes.
* **Machine learning for making predictions** :If you have transactional data of a finance company and need to build a model to determine the future trend, then machine learning algorithms are the best bet. This falls under the paradigm of supervised learning. It is called supervised because you already have the data based on which you can train your machines. For example, a fraud detection model can be trained using a historical record of fraudulent purchases.
* **Machine learning for pattern discovery** : If you don’t have the parameters based on which you can make predictions, then you need to find out the hidden patterns within the dataset to be able to make meaningful predictions.
* **Advantages of Data Analytics**
* Following are the **advantages of data Analytics**:

➨It detects and correct the errors from data sets with the help of data cleansing. This helps in improving quality of data and consecutively benefits both customers and institutions such as banks, insurance and finance companies.   
➨It removes duplicate informations from data sets and hence saves large amount of memory space. This decreases cost to the company.   
➨It helps in displaying relevant advertisements on the online shopping websites based on historic data and purchase behaviour of the users. Machine learning algorithms are applied for the same. This helps in increasing revenue and productivity of the companies.   
➨It reduces banking risks by identifying probable fraudulent customers based on historic data analysis. This helps institutes in deciding whether to issue loan or credit cards to the applicants or not.   
➨It is used by security agencies for surveillane and monitoring purpose based on informations collected by huge number of sensors. This helps in preventing any wrongdoings and/or calamities.

### Disadvantages Data Analytics

Following are the **disadvantages of data Analytics**:  
➨This may breach privacy of the customers as their information such as purchases, online transactions, subscriptions are visible to their parent companies. The companies may exchange these useful customer databases for their mutual benefits.   
➨The cost of data analytics tools vary based on applications and features supported. Moreover some of the data analytics tools are complex to use and require training. This increases cost to the company willing to adopt data analytics tools or softwares.   
➨The information obtained using data analytics can also be misused against group of people of certain country or community or caste.   
➨It is very difficult to select the right data analytics tools. This is due to the fact that it requires knowledge of the tools and their accuracy in analysing the relevant data as per applications. This increases time and cost to the company.

* Data Science Application

1. Banking.

2. Finance.

3. Manufacturing.

4. Transport.

5.Healthcare.

a. Medical image analysis.

b. Genomic Data Analysis.

c. Drug Discovery.

6. E-Commerce.