**Research paper on Data Science & Analysis using Python**

**Introduction:**

Data is the new Oil. This statement shows how every modern IT system is driven by capturing, storing and analyzing data for various needs. Be it about making decision for business, forecasting weather, studying protein structures in biology or designing a marketing campaign. All of these scenarios involve a multidisciplinary approach of using mathematical models, statistics, graphs, databases and of course the business or scientific logic behind the data analysis. So we need a programming language which can cater to all these diverse needs of data science. Python shines bright as one such language as it has numerous libraries and built in features which makes it easy to tackle the needs of Data science.

Below we will see some example scenarios where Data science is used.

### Recommendation systems

As online shopping becomes more prevalent, the e-commerce platforms are able to capture users shopping preferences as well as the performance of various products in the market. This leads to creation of recommendation systems which create models predicting the shoppers needs and show the products the shopper is most likely to buy.

### Financial Risk management

The financial risk involving loans and credits are better analysed by using the customers past spend habits, past defaults, other financial commitments and many socio-economic indicators. These data is gathered from various sources in different formats. Organising them together and getting insight into customers profile needs the help of Data science.

### Improvement in Health Care services

The health care industry deals with a variety of data which can be classified into technical data, financial data, patient information, drug information and legal rules. All this data need to be analyzed in a coordinated manner to produce insights that will save cost both for the health care provider and care receiver while remaining legally compliant.

### Computer Vision

The advancement in recognizing an image by a computer involves processing large sets of image data from multiple objects of same category. For example, Face recognition. These data sets are modelled, and algorithms are created to apply the model to newer images to get a satisfactory result. Processing of these huge data sets and creation of models need various tools used in Data science.

### Efficient Management of Energy

As the demand for energy consumption soars, the energy producing companies need to manage the various phases of the energy production and distribution more efficiently. This involves optimizing the production methods, the storage and distribution mechanisms as well as studying the customers consumption patterns.

## Python in Data Science -

The programming requirements of data science demands a very versatile yet flexible language which is simple to write the code but can handle highly complex mathematical processing. Python is most suited for such requirements as it has already established itself both as a language for general computing as well as scientific computing. More over it is being continuously upgraded in form of new addition to its plethora of libraries aimed at different programming requirements. Below we will discuss such features of python which makes it the preferred language for data science.

* A simple and easy to learn language which achieves result in fewer lines of code than other similar languages like R. Its simplicity also makes it robust to handle complex scenarios with minimal code and much less confusion on the general flow of the program.
* It is cross platform, so the same code works in multiple environments without needing any change. That makes it perfect to be used in a multi-environment setup easily.
* It executes faster than other similar languages used for data analysis like R and MATLAB.
* Its excellent memory management capability, especially garbage collection makes it versatile in gracefully managing very large volume of data transformation, slicing, dicing and visualization.
* Most importantly Python has got a very large collection of libraries which serve as special purpose analysis tools. For example – the NumPy package deals with scientific computing and its array needs much less memory than the conventional python list for managing numeric data. And the number of such packages is continuously growing.
* Python has packages which can directly use the code from other languages like Java or C. This helps in optimizing the code performance by using existing code of other languages, whenever it gives a better result.

**Modules in Python-**

A module is a file containing Python definitions and statements. A module can define functions, classes and variables. A module can also include runnable code. Grouping related code into a module makes the code easier to understand and use.

1. PANDAS (for Data Analysis)
2. NUMPY (for numerical analysis and formation)
3. MATPLOTLIB (for data visualization )
4. SCIPY (for scientific computing)
5. SEABORN (for data visualization)
6. TENSORFLOW (used in deep learning)
7. SCIKIT-LEARN (used in machine learning)

**Why Data Analytics?**

Data Analytics is needed in Business to Consumer applications (B2C). Organisations collect data that they have gathered from customers, businesses, economy and practical experience. Data is then processed after gathering and is categorised as per the requirement and analysis is done to study purchase patterns and etc.

**Why Your Company Needs Data Analytics?**

The idea is to make sense of the data you have, to analyse it and share better business prospects in the near future and how you’re going to do it, is with the concepts of analytics. Data Science involves extraction of trends, patterns and useful information from a set of existing data which will be of no use if not analyzed. It is a kind of business intelligence that is now used for gaining profits and making better use of resources. This can also help in improving managerial operations and leverage organizations to next level.

If not analyzed this data is going to get wasted whereas if analyzed properly this data can help us in finding information that is powerful to bring in a change in the patterns of how business is already working or going. Just imagine a source of unleashed information exists and you haven’t dived in yet to get the grip of it. Your business can take a competitive advantage of it and do wonders with the data. This is going to extract insights that will allow an advantage to a business or an organization in an economy.

Data and information are increasing rapidly, the growth rate of the information is so high that the information available to us in the near future is going to unpredictable. data is generated through hundreds of users, businesses and industries on a whole. Try to amalgamate if this data, not the big data but the data you have gathered from your business if wasted what you’ll be losing on.

Modelling and visualizing is one of the major aspects of analytics and so to get an up gear from this, you really need to understand the intricacies of it as a whole. Earlier data needed a number of skilled analysts to process data whereas we now have tools that are used in running high-speed data analytics on massive amounts of data, and this gives an opportunity to the entrepreneurs to incorporate data analytics when making decisions.

Different decisions can be made as far as your target audience is concerned, your audience can change on the basis of the analysis you have done with the help of data analytics. Social media is another example that has increased the growth of the data and your organization can make changes based on that too. As the communication between you and consumer if analyzed can also help in making snap decisions.