**PYTHON FOR DATA SCIENCE AND DATA ANALYSIS  
NAME: BHANU CHANDER KURETI  
COLLEGE: ILLINOIS INSTITUTE OF TECHNOLOGY**

**COURSE: ITMD - 513 - PYTHON**

**PYTHON FOR DATA SCIENCE AND DATA ANALYSIS**

Data Science has been established as an important emergent scientific field and paradigm driving research evolution in such disciplines as statistics, computing science and intelligence science, and practical transformation in such domains as science, engineering, the public sector, business, social sci­ence, and lifestyle. The field encompasses the larger ar­eas of artificial intelligence, data analytics, machine learning, pattern recognition, natural language understanding, and big data manipulation. It also tackles related new sci­entific chal­lenges, ranging from data capture, creation, storage, retrieval, sharing, analysis, optimization, and vis­ualization, to integrative analysis across heterogeneous and interdependent complex resources for better decision-making, collaboration, and, ultimately, value creation.

In a Short Note, Data science is a multidisciplinary blend of data inference, algorithm development, and technology in order to solve analytically complex problems.

The aspect of data science is all about uncovering findings from data. Diving in at a granular level to mine and understand complex behaviors, trends, and inferences. It's about surfacing hidden insight that can help enable companies to make smarter business decisions. For example: Netflix data mines movie viewing patterns to understand what drives user interest and uses that to make decisions on which Netflix original series to produce.

A "data product" is a technical asset that: (1) utilizes data as input, and (2) processes that data to return algorithmically generated results. The classic example of a data product is a recommendation engine, which ingests user data, and makes personalized recommendations based on that data. Example: Amazon Recommendation engine .

**Python good choice of Data Science:**

Python is a very powerful tool and has a user-friendly interface, which is comfortable to learn. It is a type of interpreted language because as it gets the input it provides the output simultaneously. It also has a base of all stored data.

This program also supports Internet Communications Engines (ICE) and other integration technology as well. It also consists of many add-on packages which can tackle specific tasks. Its utility as a scripting language is limited but finds a broader use in non-scripting contexts. Python also has a very clear and readable syntax, making it ideal for users. Writing programs becomes very easy with the use of this language.

It has vast use in imagining application because of which Python is the best choice for data scientists. Gnofract 4D is one of the flexible fractal programming software which allows the user to generate beautiful images. Using mathematical principles, the computer automatically generates images. The user doesn’t need to apply maths and can use keyboard and mouse inputs to generate the images.

Gogh is another one painting program or image editor which is based on PyGTK which supports pressure-sensitive tablets. It is also a photo collection manager and viewer which allows you to search depending on the content. VPython is a programming language with 3D graphics which is called “visual”. It also helps to display the object on the computer screen. It focuses more on the computational aspect. It is very user-friendly and imported by the Python module from Python programs. It can also be scripted from the Python interpreter.

**Python implementation in Data Analysis:**

Data analysts are responsible for interpreting data and analyzing the results utilizing statistical techniques and providing ongoing reports. They develop and implement data analyses, data collection systems, and other strategies that optimize statistical efficiency and quality. They are also responsible for acquiring data from primary or secondary data sources and maintaining databases. Besides, they identify, analyze, and interpret trends or patterns in complex data sets. Data analysts review computer reports, printouts, and performance indicators to locate and correct code problems. By doing this, they can filter and clean data.

Data analysts conduct full lifecycle analyses to include requirements, activities, and design, as well as developing analysis and reporting capabilities. They also monitor performance and quality control plans to identify improvements. Finally, they use the results of the above responsibilities and duties to better work with management to prioritize business and information needs. One needs only to briefly glance over this list of data-heavy tasks to see that having a tool that can handle mass quantities of data easily and quickly is an absolute must. Considering the proliferation of Big Data (and it’s still on the increase), it is important to be able to handle massive amounts of information, clean it up, and process it for use. Python fits the bill since its simplicity and ease of performing repetitive tasks means less time needs to be devoted to trying to figure out how the tool works.

References

**John Terra**2019 Python for Data Science and Data Analysis. Available at <https://www.simplilearn.com/why-python-is-essential-for-data-analysis-article>.

Understanding how python is used for data science. Available at <https://www.datasciencegraduateprograms.com/python/>.

Why is python a language of choice for data science. Available at <https://www.quora.com/Why-is-Python-a-language-of-choice-for-data-scientists> .