Summer training report

by Bhanu Shri Nayak

MBA (Rural Management)

2019-21

IIHMR UNIVERSITY

(Python for data Science)

Python is a general-purpose programming language, meaning it can be used in the development of both web and desktop applications. It is also useful in the development of complex numeric and scientific applications. With this sort of versatility, it comes as no surprise that python is one of the fastest-growing programming languages in the world. Using this course, we will learn the essential concepts of python programming and gain in -depth, valuable knowledge in data analytics, machine learning, data visualization, web scraping, and natural language processing.

Based on recent surveys, python was and still is the most powerful, high -level, open -source programming language! It is a powerful, fast, and dynamic programming language the runs everywhere. And, interactive, object -oriented, and very easy to learn. It is great for web development, and scientific computing. It is also very useful for visualization and data analytics. This course covers python 3 and it has many features. It comprises of four modules. In this course we learnt the first module -The computer will run or execute in an expression or statement. The simplest program we can write is in a print statement. While running the print statement, python displays the value in the parentheses. The value in the parentheses is called the argument. If we use a Jupiter notebook in this course, we will see a small rectangle with the statement is called a cell. If we select this cell with our mouse, then we will click the run cell button. The statement will be executed. The result will be displayed down the cell. In second module we learnt about sets, dictionaries, lists and tuples these are called compound data types and are one of the key types of data structures in python. Tuples are an ordered sequence. Tuples are expressed as comma- separated elements within parentheses. These are values inside the parentheses. In python, there are different types strings, integer, float.

(Course description)

What lies beyond the Jupyter notebook? How can we elevate code from concept to production? What happens when scikit-learn is not enough? Will that last script die as a one-off or perform just as well for the next 10,000 inputs? The last decade has seen an amazing commoditization of cloud computing and scientific development tools that make it a truly glorious time to be a data scientist, yet the increasing ease-of-use can paradoxically hinder the development of more sophisticated tools if the scientist relies too heavily on magic and never opens the hood to explore how things really work. In this course, we explore the next level of fundamentals that make a difference for data science teams in real organizations using complex data. Key topics include formal collaboration techniques, testing, continuous integration and deployment, repeatable and intuitive workflows with directed graphs, recurring themes in practical algorithms, meta-programming and glue, performance optimization, and an emphasis on practical integration with tools in the broader data science ecosystem such as GitHub, Docker, Amazon Web Services, and Hadoop.

Objective of the Course

The objectives of this course are to get you started with Python as the programming language and give you a taste of how to start working with data in Python.

In this course we will learn about:

- What Python is and why is it useful
- The application of Python
- How to define **variables**
- **Sets** and **conditional statements** in Python
- The purpose of having **functions** in Python
- How to operate on files to **read and write data** in Python
- How to use **pandas**, a must have package for anyone attempting data analysis in Python.

Course Contents

Learning Methodology used in the course

Learning methodology is through online video and theory, it was difficult to understanding in Python for data since although in this course various case study are available.

Key Learning from the course

I learned from this course, hypothesis test, dealing with missing values in python and Identify the characteristics of datasets and compare the trivial data and big data for various applications.