Gandhinagar Institute of Technology
Computer Engineering & Information Technology Department
A Report on
"Expert Lecture on Python for Data Science"
(7th September, 2021)

Objective:

The objective of the session was to provide the programming knowledge using 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. Moreover it is being continuously upgraded in form of new addition to its plethora of libraries aimed at different programming requirements. This involves optimizing the production methods, the storage and distribution mechanisms as well as studying the customers consumption patterns. Linking the data from all these sources and deriving insight seems a daunting task. This is made easier by using the tools of data science.

About Expert Lecture:

This conference provided a valuable platform for the company's industry experts to exchange and discuss the latest views on the technology of the Python for Data Science. 40 participants have been attended from Sem-5 Students from Computer Engineering and Information Technology Department. Mr. Raj Makhijani is a senior Python Developer in Brainy Beam Technology, Ahmedabad. He has wide experience in Web App Development with Python, Django, Machine Learning and Data Science technology.

Structure of Lecture:

Time: 11:25 am to 1:00 pm

Mr. Raj Makhijani discussed about Python distribution is available for a wide variety of platforms. You need to download only the binary code applicable for your platform and install Python. If the binary code for your platform is not available, you need a C compiler to compile the source code manually. Compiling the source code offers more flexibility in terms of choice of features that you require in your installation. Many of your coding projects may require you to pull a bunch of information from an HTML or XML page. This task can be really tedious and boring, that is until you learn how to scrape the web with an HTML Parser! That's where Beautiful Soup comes in. This Python package allows you to parse HTML and XML pages with ease and pull all sorts of data off the web.

Missing data is always a problem in real life scenarios. Areas like machine learning and data mining face severe issues in the accuracy of their model predictions because of poor quality of data caused by missing values. In these areas, missing value treatment is a major point of focus to make their models more accurate and valid. Pandas is an open-source Python Library used for high-performance data manipulation and data analysis using its powerful data structures. Python
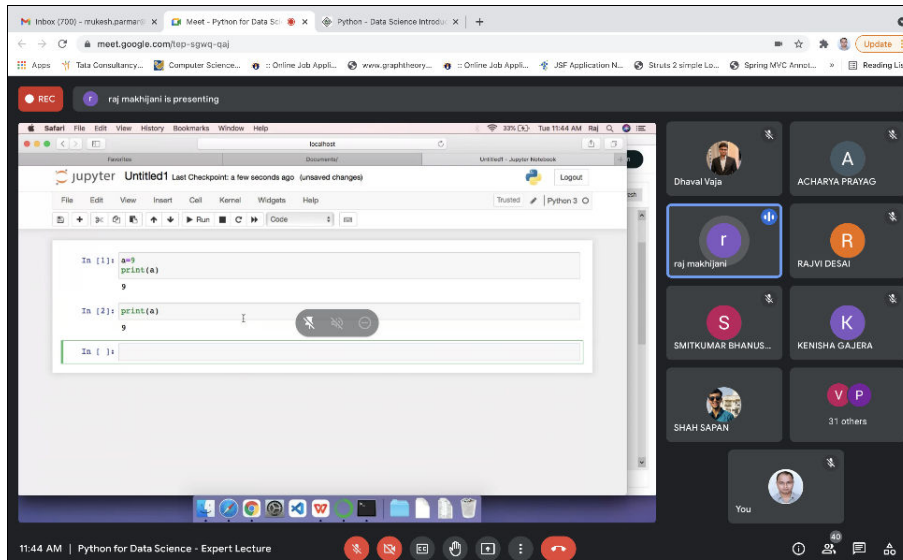
with pandas is in use in a variety of academic and commercial domains, including Finance, Economics, Statistics, Advertising, Web Analytics, and more. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data — load, organize, manipulate, model, and analyse the data.

NumPy is a Python package which stands for 'Numerical Python'. It is a library consisting of multidimensional array objects and a collection of routines for processing of array. NumPy is often used along with packages like SciPy (Scientific Python) and matplotlib (plotting library). This combination is widely used as a replacement for MATLAB, a popular platform for technical computing. However, Python alternative to MATLAB is now seen as a more modern and complete programming language. It is open source, which is an added advantage of NumPy.

Glimpse of Expert Lecture on Python for Data Science

Installation of Python

Installation of Anaconda

Explanation of jupyter
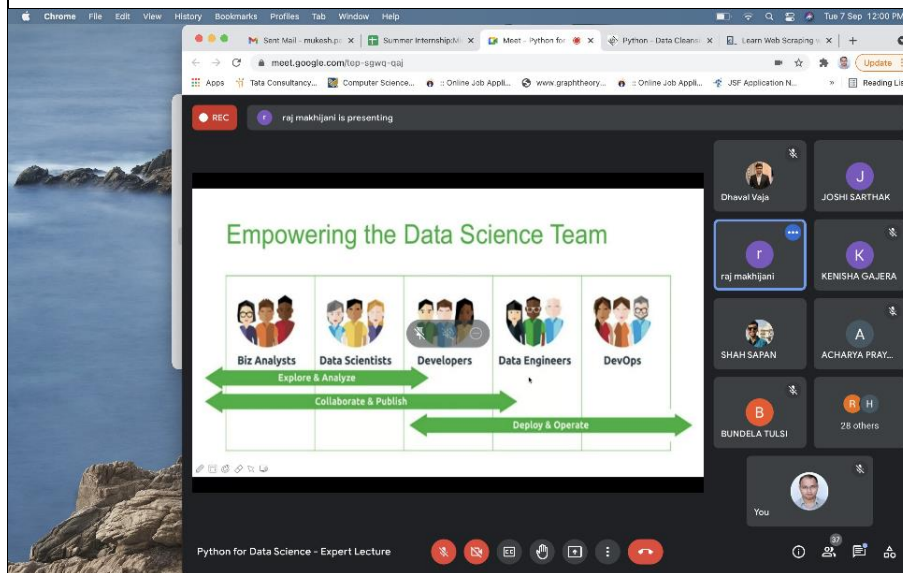
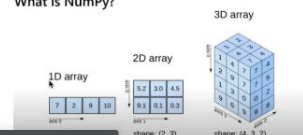Data Science Team
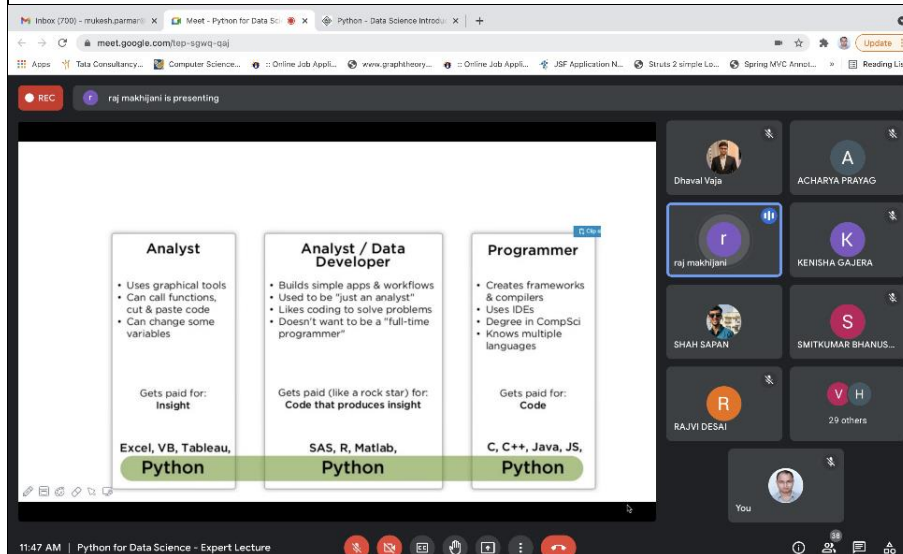
Explanation Analyst, Data Developer, Programmer

Numpy Explanation