DATA SCIENCE USING PYTHON

Presented by:
Ms. Shraddha S. More
Assistant Professor
Department of Information Technology
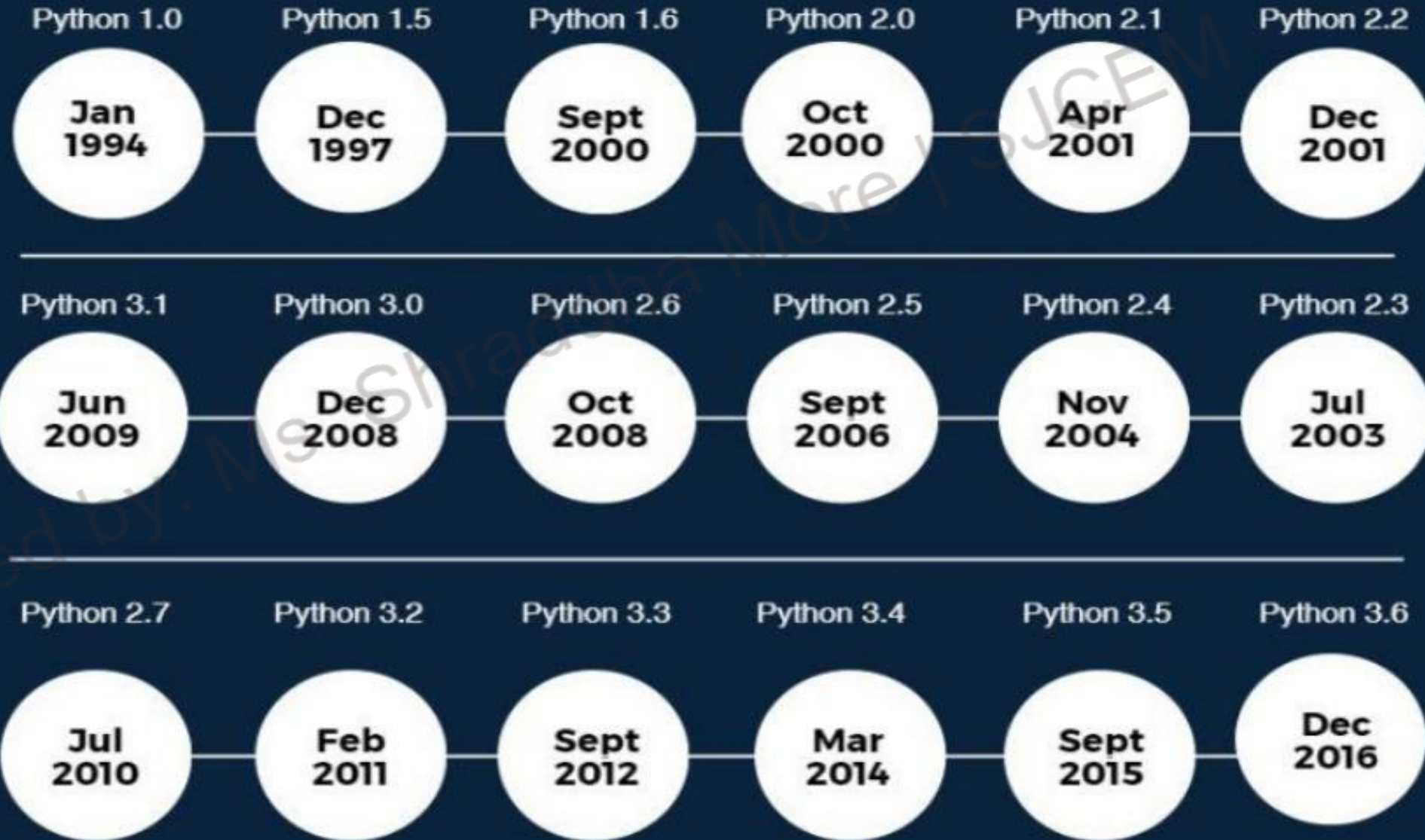**St. John College of Engineering and Management, Palghar**

HISTORY OF PYTHON

- Python was conceptualized by **Guido Van Rossum** in the late **1980s**.

- Rossum published the first version of Python code (0.9.0) in February **1991** at the CWI (Centrum Wiskunde & Informatica) in the Netherlands , Amsterdam.

- Python is derived from **ABC** programming language, which is a general-purpose programming language that had been developed at the CWI.

- Rossum chose the name "**Python**", since he was a big fan of Monty Python's Flying Circus.

- Python is now maintained by a core development team at the institute, although Rossum still holds a vital role in directing its progress.

Python Versions

Latest Version Python 3.10.2

Python 1.0	Python 1.5	Python 1.6	Python 2.0	Python 2.1	Python 2.2
Jan 1994	Dec 1997	Sept 2000	Oct 2000	Apr 2001	Dec 2001

Python 3.1	Python 3.0	Python 2.6	Python 2.5	Python 2.4	Python 2.3
Jun 2009	Dec 2008	Oct 2008	Sept 2006	Nov 2004	Jul 2003

Python 2.7	Python 3.2	Python 3.3	Python 3.4	Python 3.5	Python 3.6
Jul 2010	Feb 2011	Sept 2012	Mar 2014	Sept 2015	Dec 2016

What is Python?

- Python is a powerful, open-source high-level & popular programing language.

- Used for web development, scientific & mathematical application development etc.

- Provides excellent library support and has a large developer community.

- Provides easy integration with web services & GUI-based desktop applications.

- Used by most of the companies such as YouTube, Instagram, Pinterest etc.

- Extensively used in Data Science and in for developing Machine Learning projects

Features of Python

Easy	01	Extensible	07
Expressive	02	Embeddable	08
Free and Open Source	03	Interpreted	09
High-Level	04	Large Standard Library	10
Portable	05	GUI Programming	11
Object Oriented	06	Dynamically Typed	12

Why Python?

01 Simple Syntax		**06** Use in Web Development	
02 Abundance of Libraries and Frameworks		**07** Useful in ML and AI	
03 Portable Feature		**08** Highly Secure	
04 Encourages Automation		**09** Offers Appreciable Jobs	
05 Useful in Data Science		**10** Offers a Satisfactory Salary	

Benefits of Python

Applications of Python
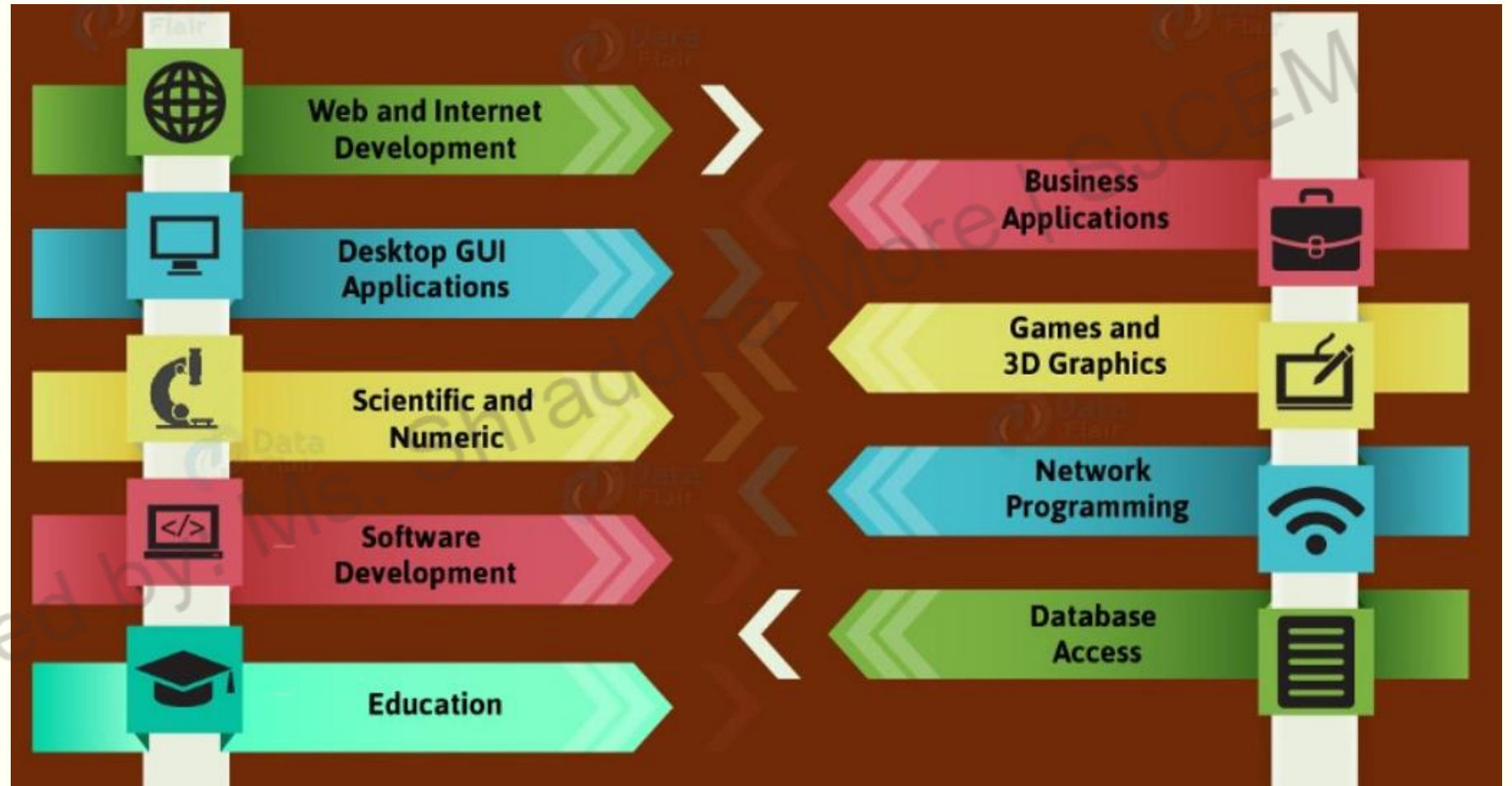
Top Companies Using Python

Python Libraries for Data Science
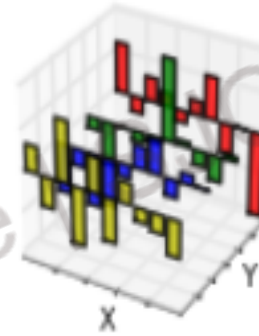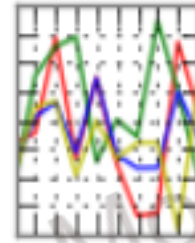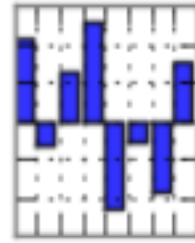
From Data Exploration to visualization to analysis – Pandas is the almighty library you must master!
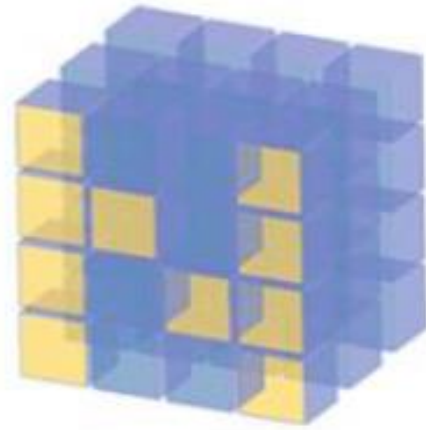
Pandas is an open-source package. It helps you to perform data analysis and data manipulation in Python language. Additionally, it provides us with fast and flexible data structures that make it easy to work with Relational and structured data.

SciPy (Scientific Python) is the go-to library when it comes to scientific computing used heavily in the fields of mathematics, science, and engineering. It is equivalent to using Matlab which is a paid tool.

NumPy is one of the most essential Python Libraries for scientific computing and it is used heavily for the applications of Machine Learning and Deep Learning. NumPy stands for NUMerical PYthon. Machine learning algorithms are computationally complex and require multidimensional array operations. NumPy provides support for large multidimensional array objects and various tools to work with them.

Matplotlib is the most popular library for exploration and data visualization in the Python ecosystem. Every other library is built upon this library.

Matplotlib offers endless charts and customizations from histograms to scatterplots, matplotlib lays down an array of colors, themes, palettes, and other options to customize and personalize our plots. matplotlib is useful whether you're performing data exploration for a machine learning project or building a report for stakeholders, it is surely the handiest library!

scikits
learn
machine learning in Python

Sklearn is the Swiss Army Knife of data science libraries. It is an indispensable tool in your data science armory that will carve a path through seemingly unassailable hurdles. In simple words, it is used for making machine learning models.

Scikit-learn is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering, and dimensionality reduction.

Seaborn

Seaborn is a free and open-source data visualization library based on Matplotlib. Many data scientists prefer seaborn over matplotlib due to its high-level interface for drawing attractive and informative statistical graphics.

Scrapy is a Python framework for large scale web scraping. It gives you all the tools you need to efficiently *extract* data from websites, *process* them as you want, and store them in your preferred *structure* and format.

HANDS-ON

Python Career Opportunities

THANK YOU !!!

shraddham@sjcem.edu.in   linkedin.com/in/shraddha-more-1052b4104   shraddhamore22

Q&A

You can reach out to me on

shraddham@sjcem.edu.in

linkedin.com/in/shraddha-more-1052b4104

shraddhamore22