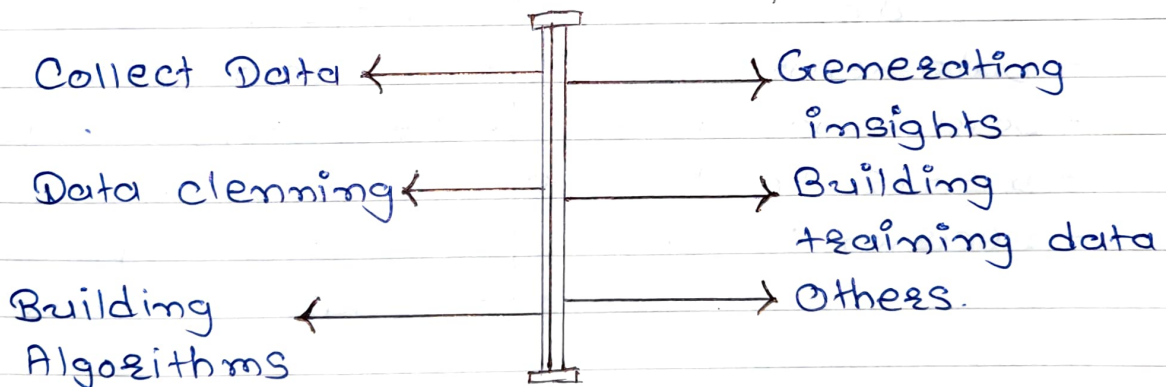


Lecture 1 - Fundamentals of Python

- Programming - Machine, Translator, Programming
- What is Python, Essential for Data Science?
- What is an IDE?
- Types of IDE in Python
- Jupyter Notebook
- What is Anaconda Navigator.
- Jupyter Notebook Features Explanation
- First Code in Python - Input, Output, Data
- Practice Questions
- Curious Data Minds.

* What a Data Scientist Do?



* Data Science

- | | |
|---------------|--------------------------|
| - Analysis | - Edtech |
| - Structure | - Fintech |
| - Algorithm | - D2C |
| - Process | - Healthtech |
| - Programming | - Aggitech |
| - Solving | - Media & Entertainment. |
| - Knowledge | |

Capture

- Data Acquisition
- Data Entry
- Signal Reception
- Data Extraction

Process

- Data Mining
- Clustering/Classification
- Data Modeling
- Data Summarization

Maintain

- Data Warehousing
- Data Cleaning
- Data Staging
- Data Processing
- Data Architecture

Communicate.

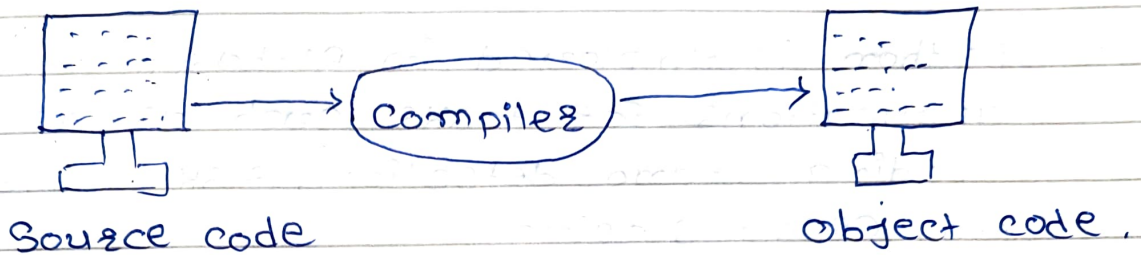
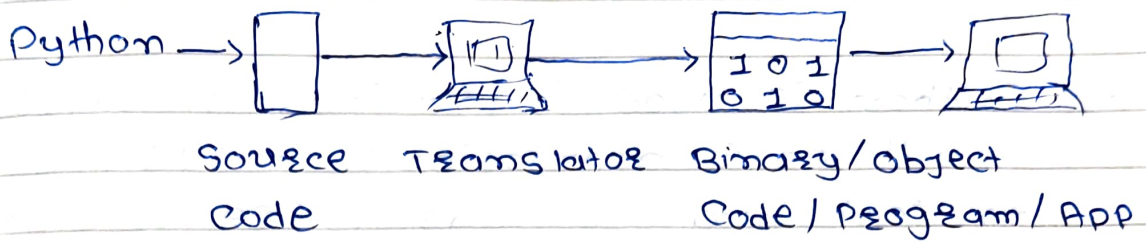
- Data Reporting
- Data Visualization
- Business Intelligence
- Decision Making

Analyze.

- Exploratory/Confirmatory
- Predictive Analysis
- Regression
- Text Mining
- Qualitative Analysis

* What is Programming?

=>



* Why are we studying Python for data science?

=> Python is commonly used for to develop chatbots - Chat GPT, Google Assistance, Siri, and chatbots used in computer service applications.

- Python is used to build recommendation system used by e-commerce platforms like Amazon and streaming service like Netflix.
- Python is employed in image recognition. Examples include facial recognition systems.
- TensorFlow is a free, open-source Python library developed by Google. Coca-Cola used TensorFlow to fuel its mobile proof of purchase model, which helped the company

save millions of dollars by eliminating production line printers.

- Python is used in various AI-based health care applications, including medical image analysis, disease diagnosis, drug discovery, and personalized medicine.
 - Python is employed in AI-based financial applications for tasks such as arithmetic trading, fraud detection, risk assessment, and credit scoring.
 - Python is utilized in speech recognition system, Examples include voice assistants like amazon Alexa.
 - PyTorch is an open-source library deployed by Facebook's (Meta) AI Research lab. Twitter uses PyTorch for natural language processing tasks. Microsoft employees PyTorch for research and development in artificial intelligence.
- * Why python is specifically used for data science ?
- ⇒ Python offers a wide range of libraries and tools specifically designed for data manipulation, analysis, and visualization. Libraries like pandas, Numpy, SciPy, and scikit-learn provide

Powerful functionalities for various data-related tasks.

- Its simplicity enables faster development, prototyping, and iteration, which is crucial in the iterative nature of data science projects.
- Python boasts a rich ecosystem of libraries and frameworks catering to different aspects of data science, including machine learning, deep learning, natural language processing and more.
- Python has a large and active community of developers, data scientists, and enthusiasts who contribute to its growth.
- Python has gained widespread adoption in the data science industry, with many companies and organizations using it as their primary programming language for data analysis and modeling.

* Brief History About Python

⇒ Python is High-level, general-purpose programming language that was created by Guido van Rossum and first released in 1991.

- On November 2020 Van Rossum holds the

title Distinguished Engineer at Microsoft.

- Van Rossum is now part of a team Microsoft working to speed up the language's performance. And recently a Microsoft blog post reported that Python 3.11 had brought speedups of 10-60% to some parts of the language.

* Anaconda Navigator.

=> Anaconda Navigator is a graphical user interface (GUI), which is a popular python distribution for data science and machine learning tasks.

- It provides an easy-to-use interface for managing packages, environments and applications related to python programming, especially in the realm of data science and scientific computing.
- Inside Anaconda Navigator, you have access to integrated development environments (IDEs) that are commonly used in Python development, particularly for data science and specific computing tasks.

* IDE - Integrated Development Environment
⇒ An Integrated Development Environment (IDE) is a software application that provides comprehensive facilities for software development.

- It typically include a code editor, a compiler or interpreter, build automation tools, debugging tools, and other features to streamline the process of writing, testing and debugging code.
- Data visualization: Many IDEs include integrated data visualization tools that allow data scientists to explore and visualize datasets directly within the IDE. This enables them to quickly generate plots, charts and graphs to better understand the data and identify patterns and trends.

9 Best Python IDEs And Code Editors

- 1) Pycharm
- 2) Spyder
- 3) Idle
- 4) VS Code
- 5) Sublime Text
- 6) Atom
- 7) Jupyter
- 8) Visual Studio
- 9) Vim

- * which of the following integrated development environments (IDEs) do you use on a regular basis?
- ⇒ - Jupyter (JupyterLab, Jupyter Notebooks, etc.) : 73%
- Visual Studio / VS code : 31%
 - R Studio : 30%
 - PyCharm : 29%
 - Notepad ++ : 22%
 - Spyder : 21%
 - Sublime Text : 18%
 - Vim / Emacs : 12%
 - MATLAB : 12%
 - Atom : 10%
 - Other : 5%
 - None : 2%

* Jupyter Notebook

- ⇒ - Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text.
- It's commonly used for data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.
 - One of the key features of Jupyter Notebook is its interactivity, which

allows you to ~~learn~~ run code in a step-by-step manner, making it excellent for exploratory data analysis and interactive computing.