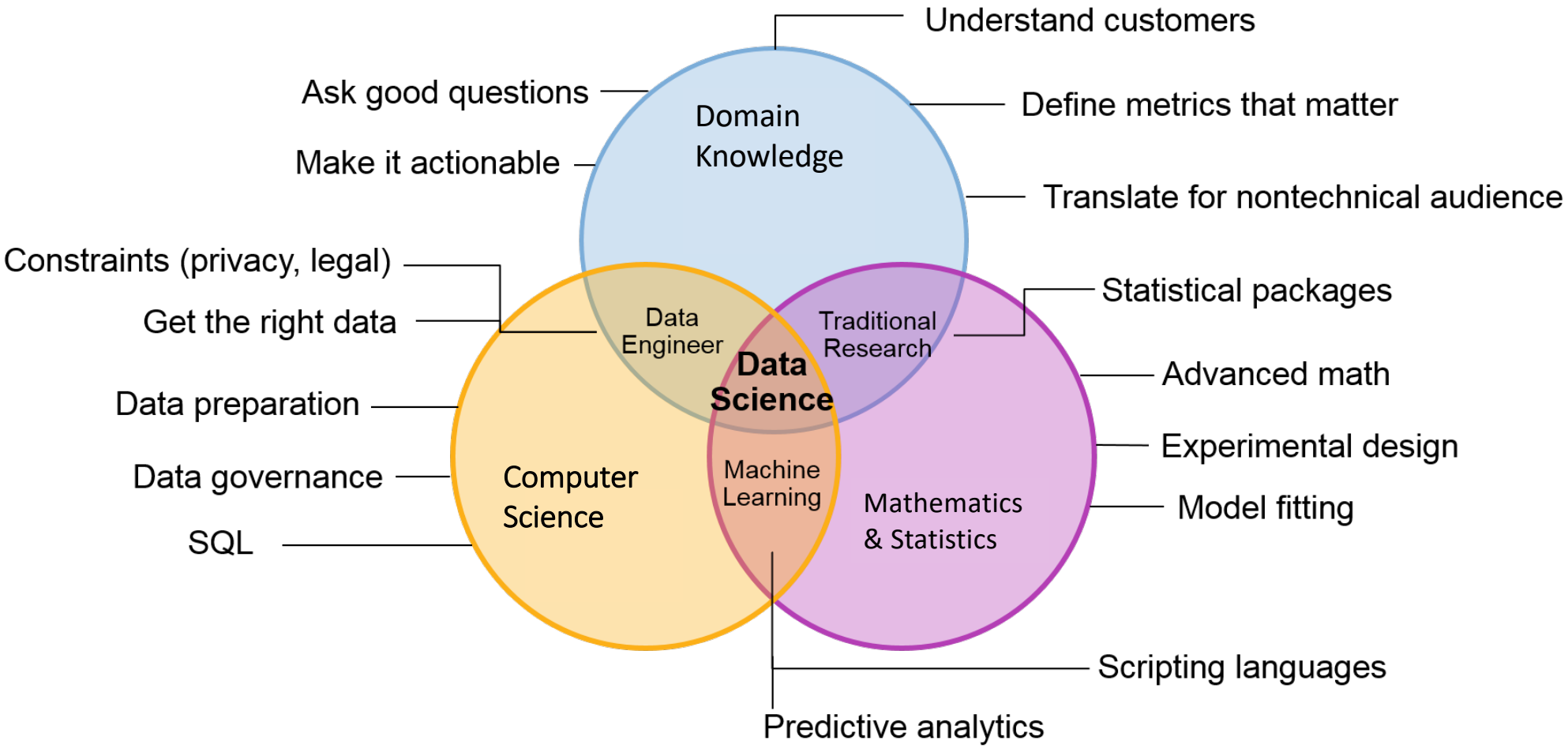
DS OCT 2022 Batch
Module 1 : Introduction to Data Science

Topics

- Introduction to Data Science
- Introduction to Programming
- Introduction to Python: Installation and Running
- Jupyter Notebook
- Running .py file from terminal
- Google Colab

Introduction to Data Science?


Understand customers
Ask good questions
Domain Knowledge
Define metrics that matter
Make it actionable
Translate for nontechnical audience
Constraints (privacy, legal)
Get the right data
Data Engineer
Traditional Research
Statistical packages
Data Science
Advanced math
Data preparation
Experimental design
Data governance
Computer Science
Machine Learning
Mathematics & Statistics
Model fitting
SQL
Scripting languages
Predictive analytics

Why Data Science?

- Every year, the world creates huge amounts of data that's only accelerating.
- Companies now have access to unprecedented amounts of data, which they need to process and understand.
- Data scientists are responsible for transforming this growing ocean of data into valuable and actionable information.
- The Bureau of Labor Statistics projects 31.4 percent employment growth for data scientists between 2020 and 2030. In that period, an estimated 19,800 jobs should open up.

Top Skills for Data Scientists

- Programming
- DBMS
- Mathematics and Statistics
- Data Visualisation
- Machine learning
- Deep Learning
- Communication Skills

Roles and Responsibilities

On a more technical level, the core job responsibilities of a data scientists include:

- Using database tools and programming languages to obtain, manipulate and analyse data.
- Creating machine learning algorithms across traditional and deep learning.
- Analysing historical data to identify trends and support optimal decision-making.
- Communicating with both technical and non-technical stakeholders.
- Keeping up-to-date with advancements in technology.

Introduction to Programming

What is Programming Language

- A Programming language is a set of rules that provides a way of telling a computer what operations to perform.
- A program is set of instructions that is executed by a computer to accomplish a particular task.
- It Provides a linguistic framework for describing computations.
- It is set of rules for communicating an algorithm
- It is a notational system for describing computation in a machine readable and human readable form.

What is Programming Language

- Any Common human language has words, vocabulary, symbols and Grammatical rules.
- Programming languages also have words, vocabulary, symbols and Grammatical rules.
- These grammatical rules are called as syntax.
- Each programming language has a different set of syntax rules.

What is Programming Language

Computer Programmer

Someone who can write computer programs or in other words, someone who can do computer programming is called a Computer Programmer.

Algorithm

An **algorithm** is a step-by-step procedure to resolve any problem. An algorithm is an effective method expressed as a finite set of well-defined instructions.

Example :

- Take list of 100 numbers (1 to 100).
- Create a list of squares of even numbers from above list

What is Python ?

Python

Python is a high-level, general-purpose and a very popular programming language.
Python programming language (latest Python 3) is being used in web development, Machine Learning applications, along with all cutting edge technology in Software Industry.

Advantages of Python

- Works on different platforms (Windows, Mac, Linux, etc).
- Simple syntax similar to the English language.
- Syntax that allows developers to write programs with fewer lines than some other programming languages.
- Runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.
- Supports functional and structured programming methods as well as OOP.
- Provides very high-level dynamic data types and supports dynamic type checking.
- Can be easily integrated with Other Languages.

Why Python is preferred in Data Science ?

Python in Data Science

As a new Data Scientist, you know that your path begins with programming languages you need to learn. Python is the most popular language for all Data Scientists.

But Why is this?

Reasons behind Python's popularity in Data Science

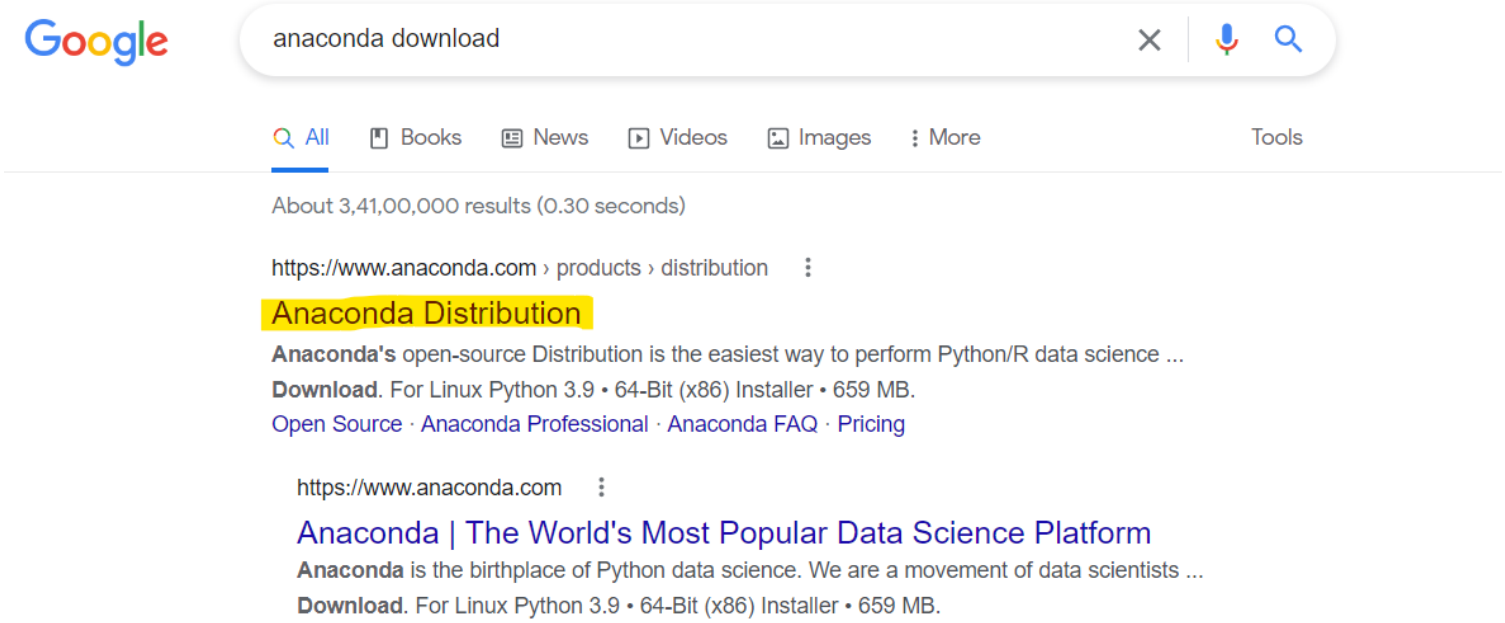
- Python is a free and open-source language
- Simple Syntax as compared to other languages
- Availability of large number of libraries and frameworks (NumPy, Pandas ,matplotlib …)
- Scalability
- Efficient Deep Learning
- Huge Community
- Jobs and Growth in market

Installation

Steps of installation

Step 1)

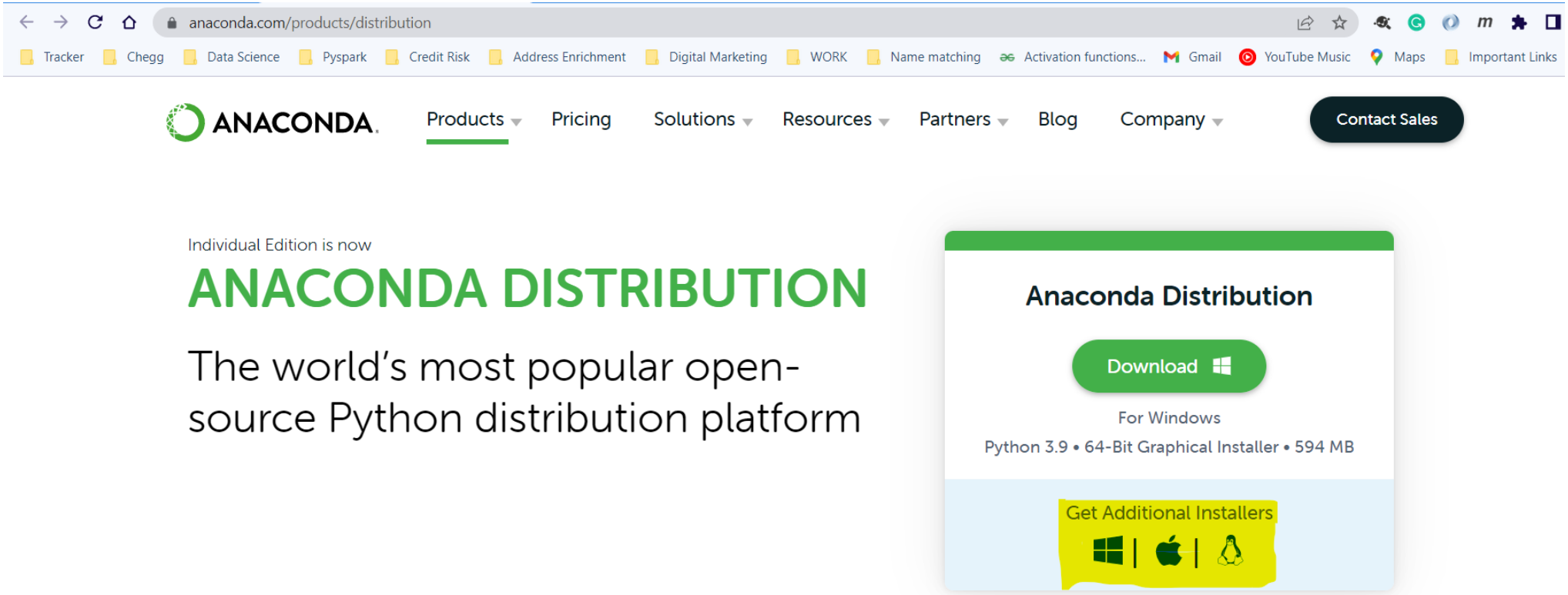
Go to
https://www.anaconda.com/products/distribution



Steps of installation

Step 2)

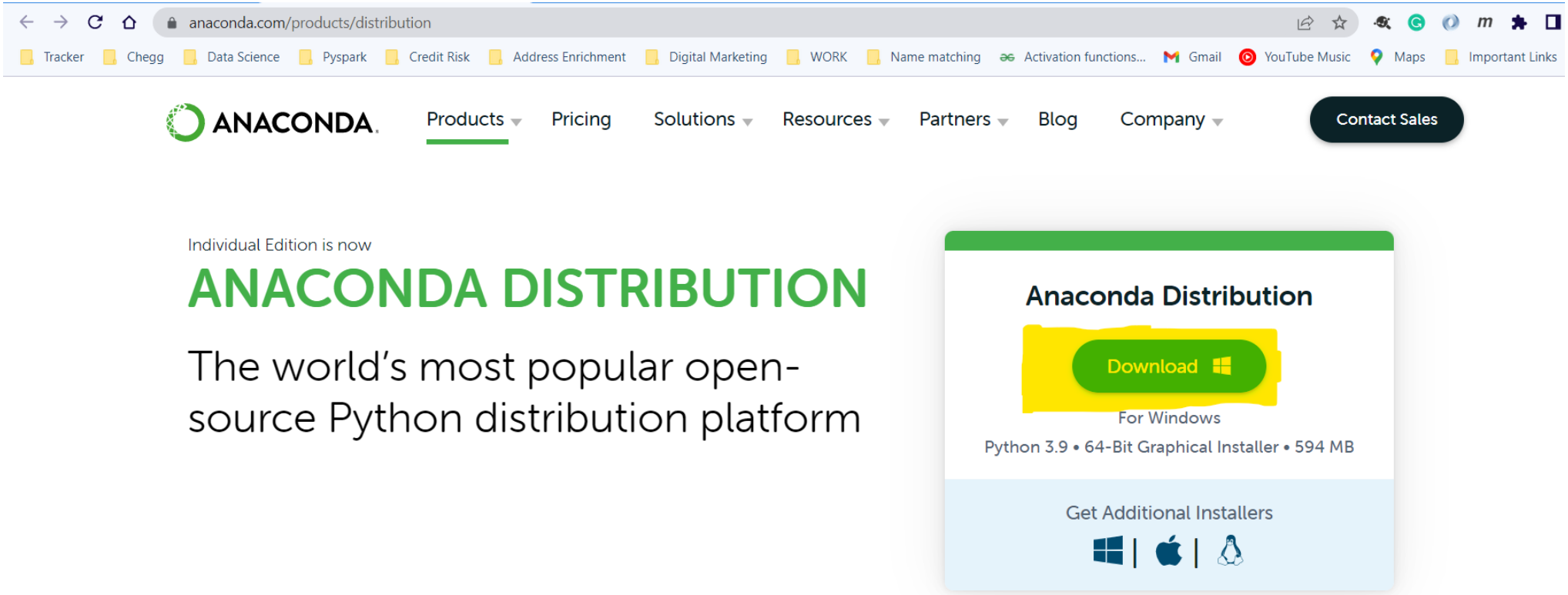
Select Your Operating System (Default download link is for windows operating system.)



Steps of installation

Step 3)

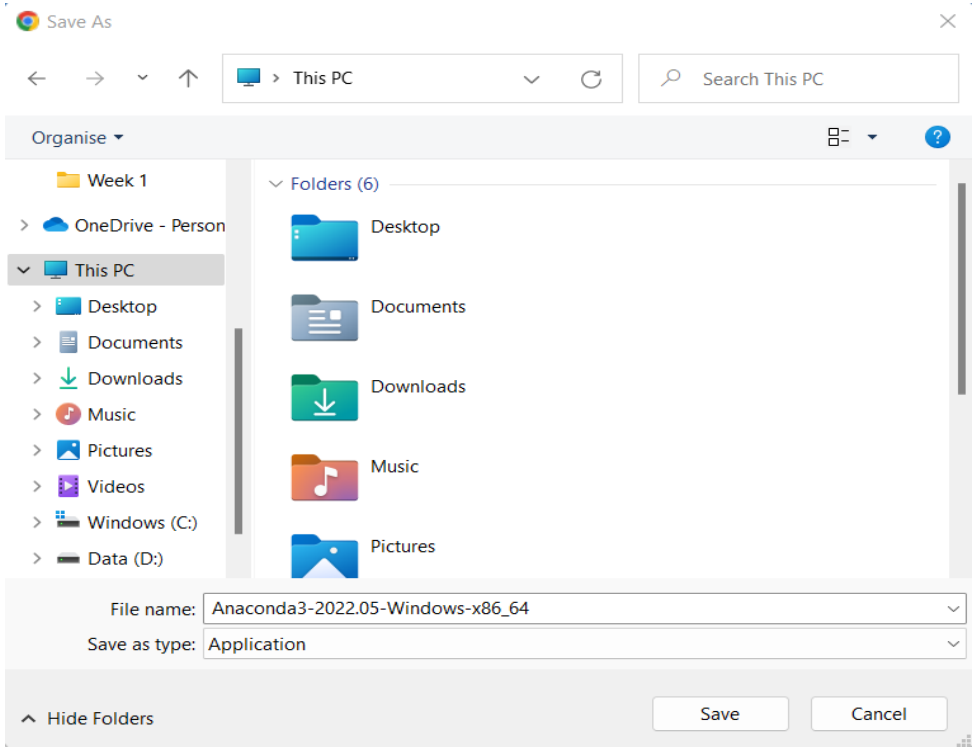
Once you have selected your operating system click on download button.

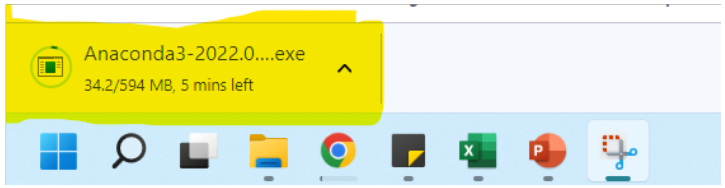


Steps of installation

Step 4)

Select the directory where you want to download your software and your download will start.

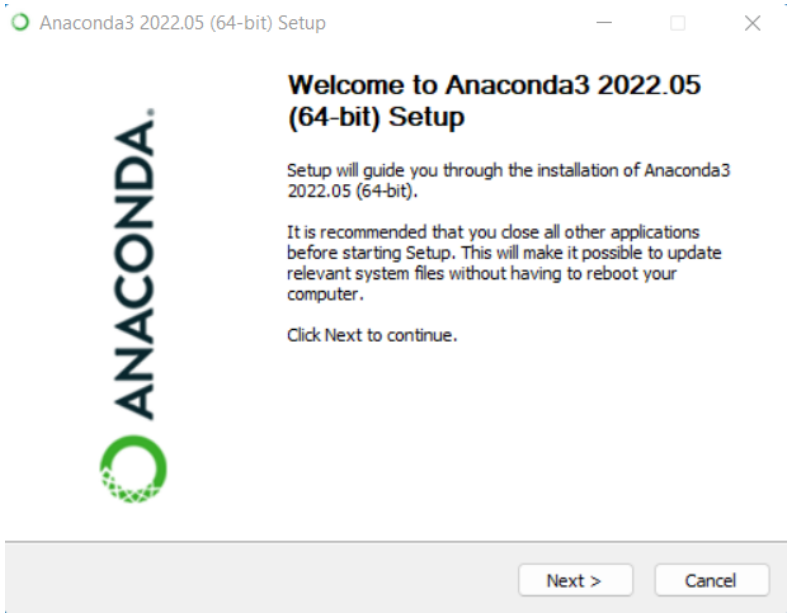




Steps of installation

Step 5)

Open your setup file which you have downloaded and select your default settings at each step and at last click on finish.



Introduction to Jupyter Notebook

Jupyter Notebooks

A Jupyter Notebook is an open source web application that allows data scientists to create and share documents that include live code, equations, and other multimedia resources.

Use of Jupyter Notebook:

Jupyter notebooks are used for all sorts of data science tasks such as exploratory data analysis (EDA), data cleaning and transformation, data visualization, statistical modeling, machine learning, and deep learning.

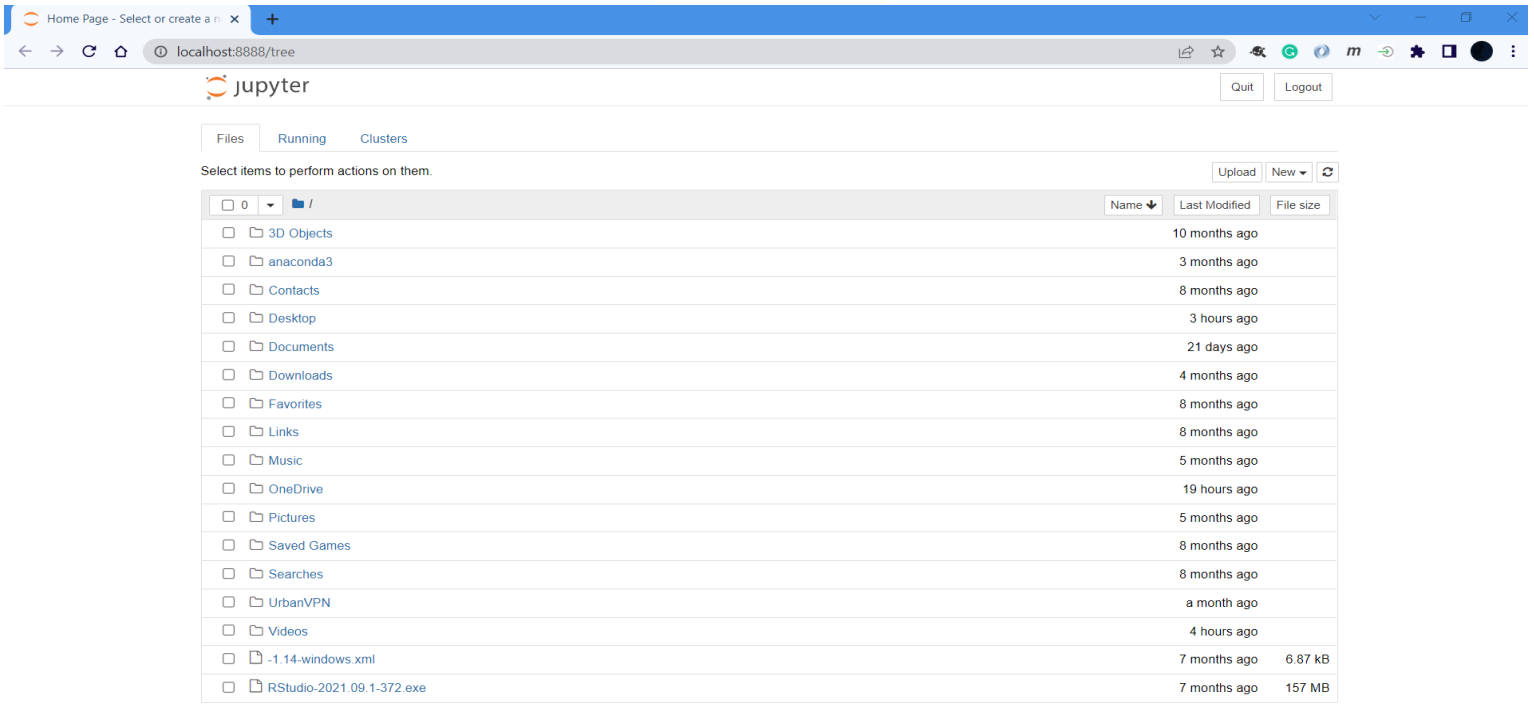
Advantage of Using Jupyter Notebook :

- Useful for "showing the work" through a combination of code, markdown, links, and images.
- Jupyter notebooks can also be converted to a number of standard output formats (HTML, PowerPoint, LaTeX, PDF, Restructured Text, Markdown, Python)

Working of Jupyter Notebook

Working of Jupyter Notebooks

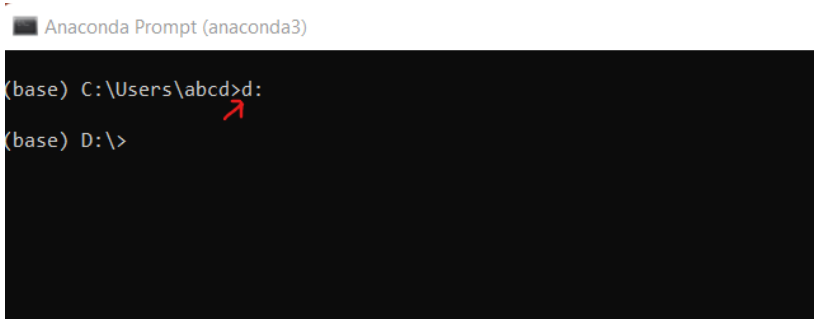
Step 1) Open Jupyter Notebook from apps
Step 2) In your default browser a web page will open like this and default directory will be C drive of your system.

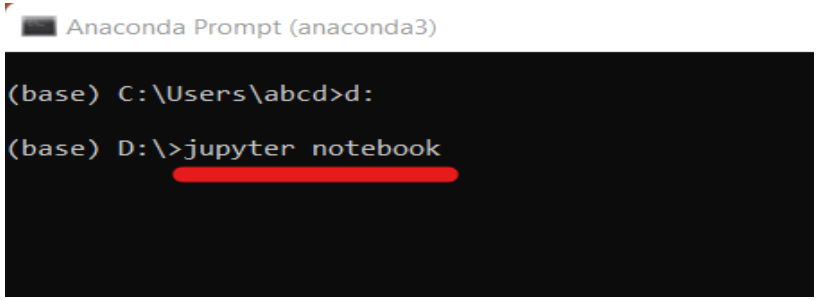


Working of Jupyter Notebooks

Step 3) If you want to run and save your code in other directory then follow these steps:

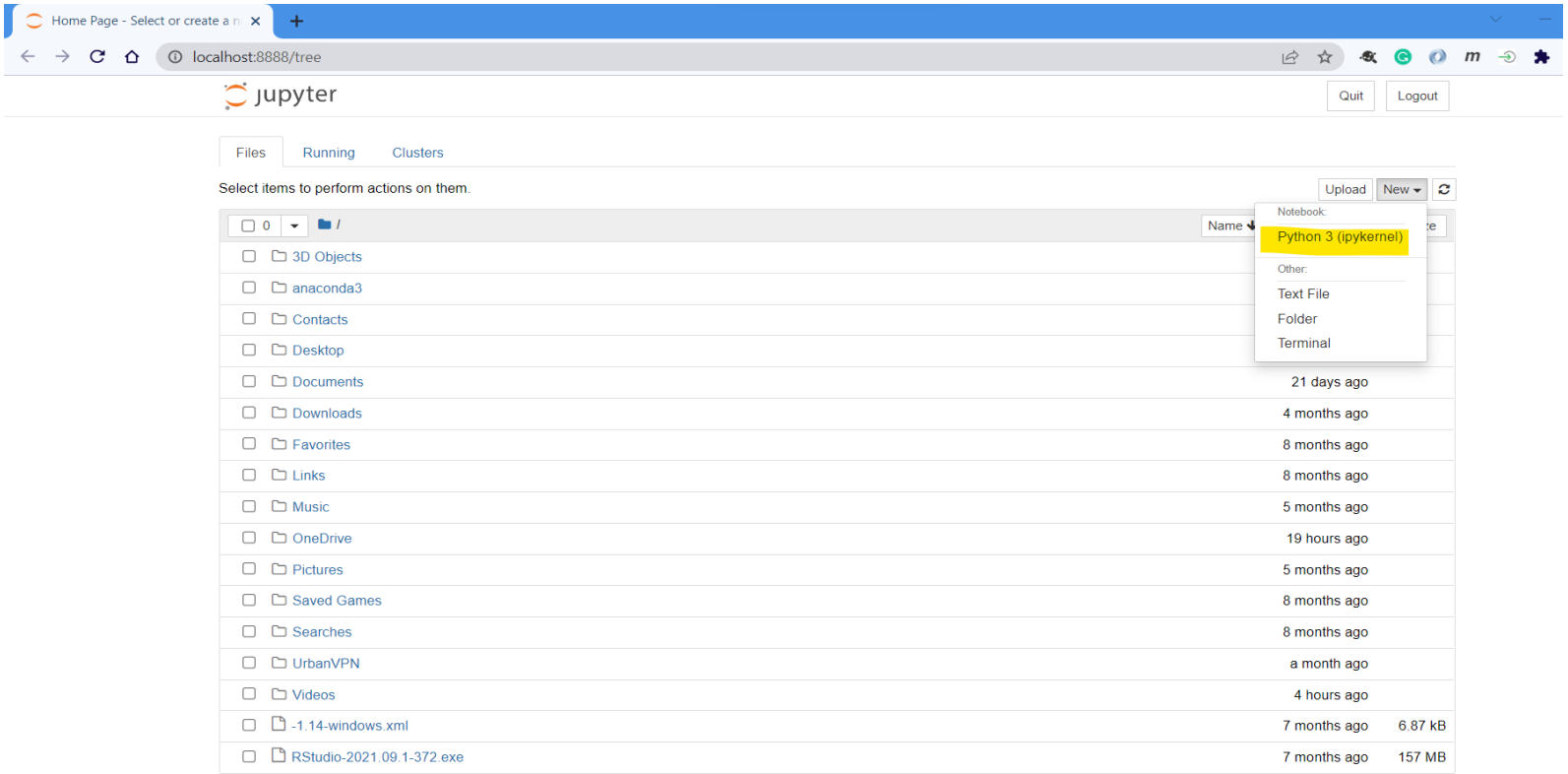
1) Open Anaconda Prompt (anaconda3) from start menu.
2) Type the directory name followed by colon (:) and then press Enter (Example – d:)
3) Type **jupyter notebook** and then Enter.





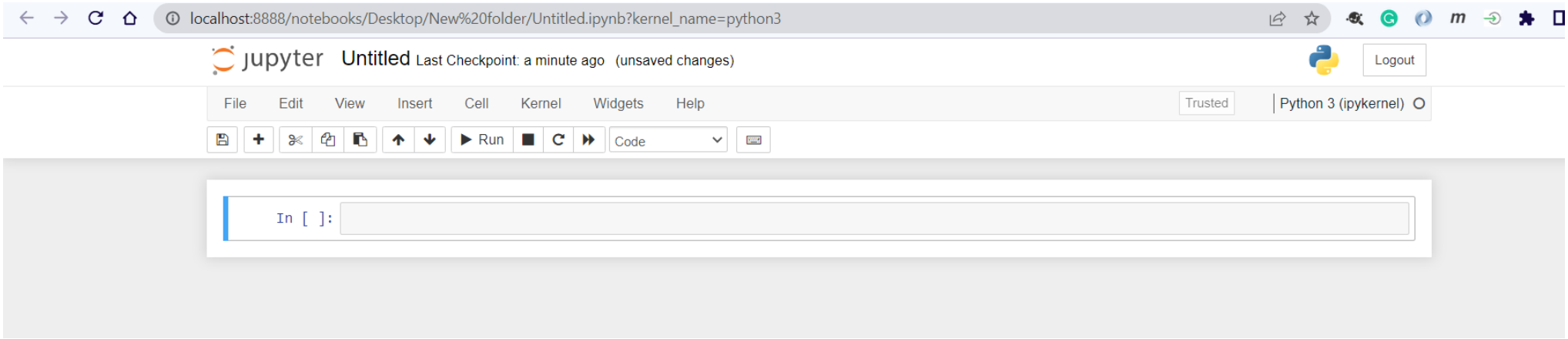
Working of Jupyter Notebooks

Step 4) Open your preferred folder and then click on New -> Python 3.



Working of Jupyter Notebooks

Step 5) A new notebook will be opened with name ‘Untitled.ipynb’ and you can run your code here.



Commands and Shortcut of Jupyter Notebooks

Action	Menu	Keyboard Shortcuts
Run selected cells	Cell -> Run Cells	Ctrl + Enter
Run the current cell and go down	Cell -> Run Cells and Select Below	Shift + Enter
Save and checkpoint	File -> Save and Checkpoint	Ctrl + S
Insert cell above	Insert -> Insert Cell Above	A
Insert cell below	Insert -> Insert Cell Below	B
Change the cell type to Code	Cell -> Cell Type -> Code	Y
Change the cell type to Markdown	Cell -> Cell Type -> Markdown	M
Code completion or indent		Tab

How to run .py file from Terminal

Spyder

Steps to Run .py scripts

1) Open Spyder from apps installed in your system.
2) Write and save your code in your preferred directory, it will be saved as .py extension.
3) Open Anaconda Command Prompt.
4) Go to the directory in which you have saved the .py file using below command
   cd path_to_directory
5) Run your script using below command:
   Python scriptname.py

Google Colab

Google Colab

- Colab is a free Jupyter notebook environment that runs entirely in the cloud.
- It does not require a setup.
- Notebooks that you create can be simultaneously edited.
- Supports many popular machine learning libraries which can be easily loaded in your notebook.

You can perform the following using Google Colab:

- Write and execute code in Python
- Create/Upload/Share notebooks
- Import/Save notebooks from/to Google Drive
- Import/Publish notebooks from GitHub
- Import external datasets e.g. from Kaggle
- Integrate PyTorch, TensorFlow, Keras, OpenCV
- Free Cloud service with free GPU/TPU.

Google Colab

Steps to use Google Colab

Step 1)
Log in to your Google account in your browser.

Step 2)
Open the following link : https://colab.research.google.com
Once you open this link a demo notebook will be opened.

Step 3)
Go to File -> New Notebook
A new notebook will open similar to Jupyter Notebook.

Step 4)
Write your code and Save it.
It will be saved in your Google Drive.

Note : If you want to find the location of your file then Go to File -> Locate in Drive


Thank you!