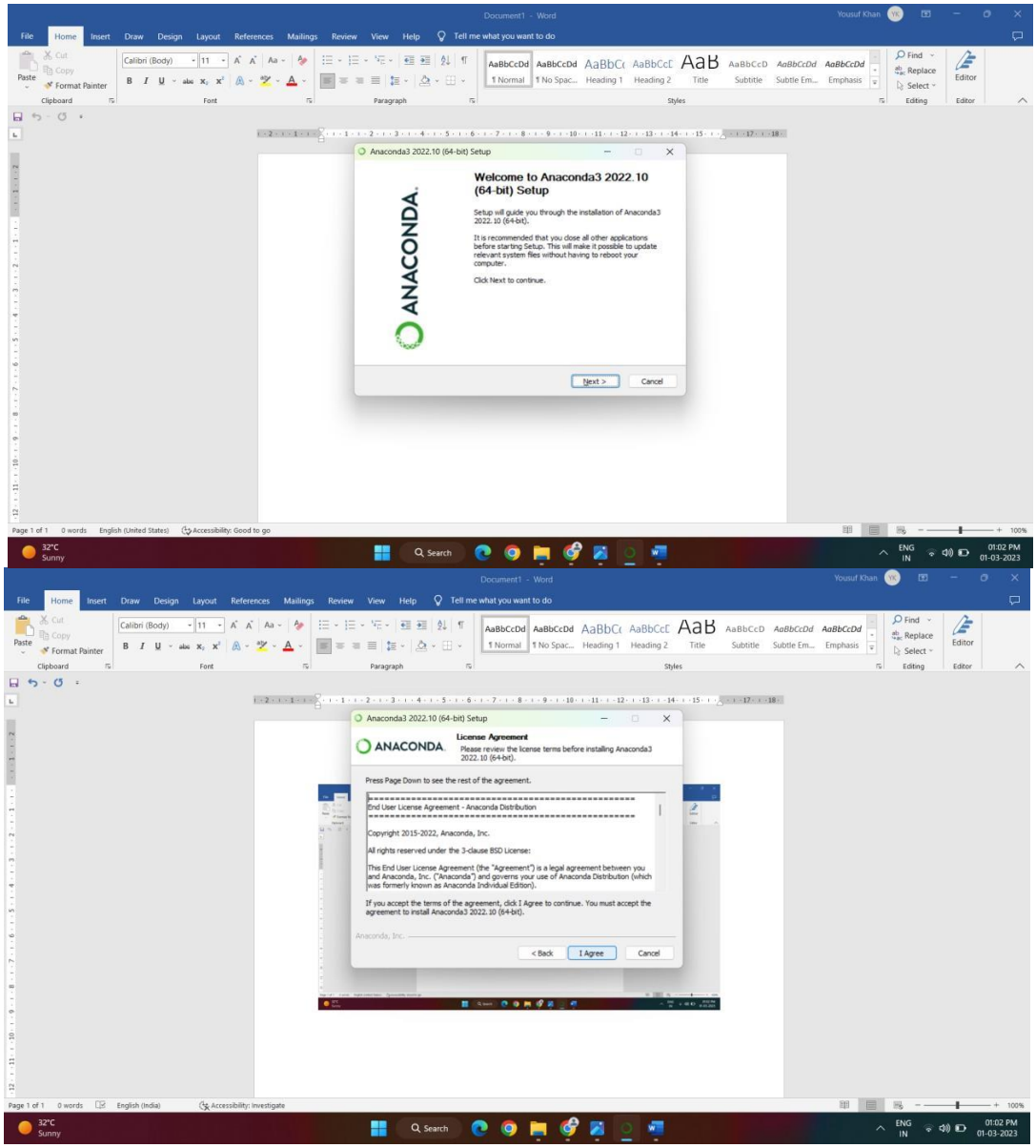
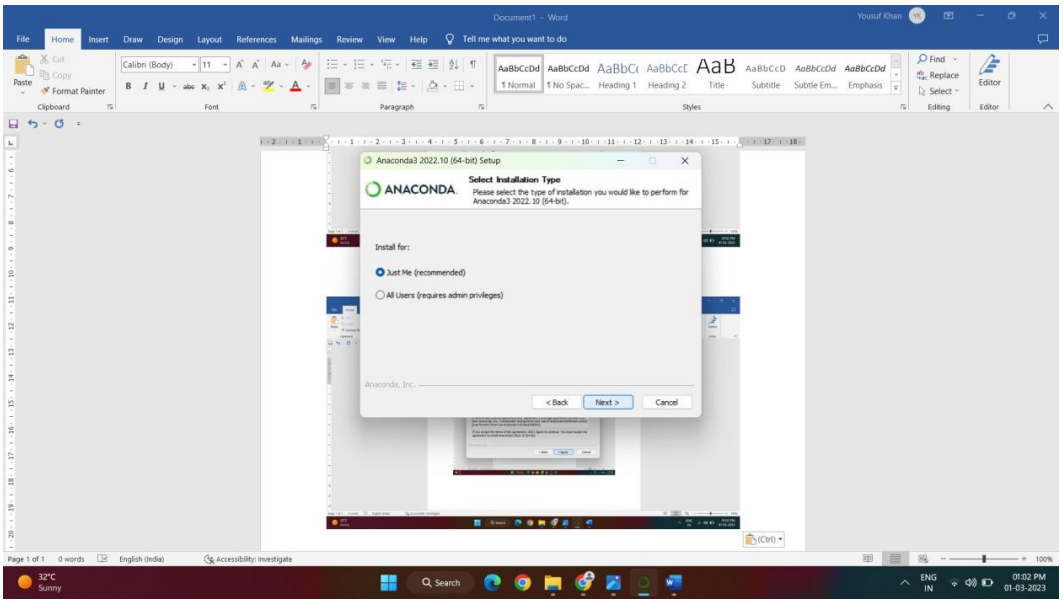
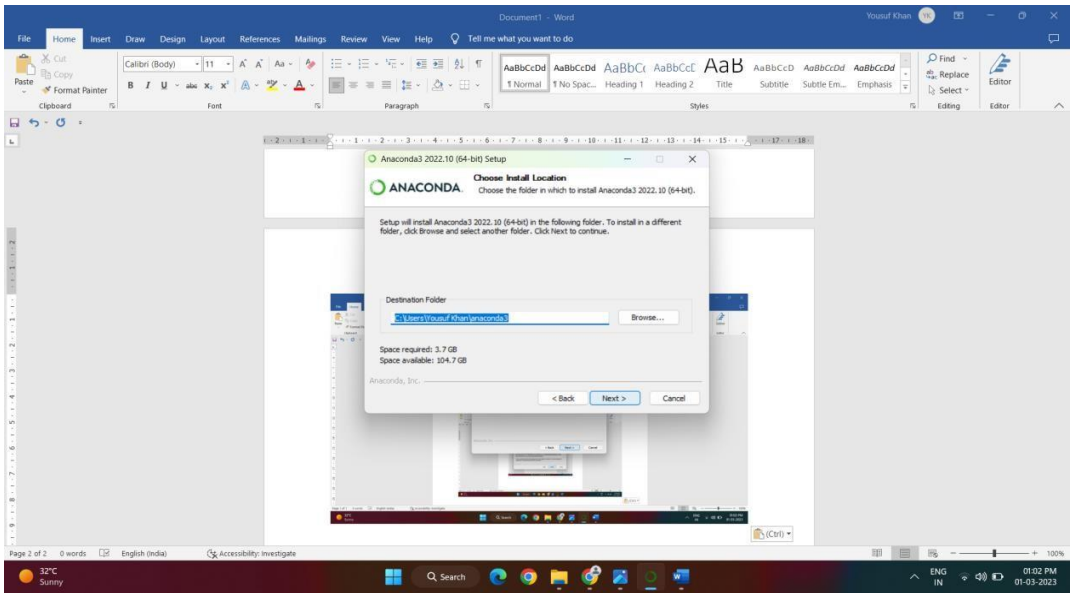
1. Basic Data Preprocessing

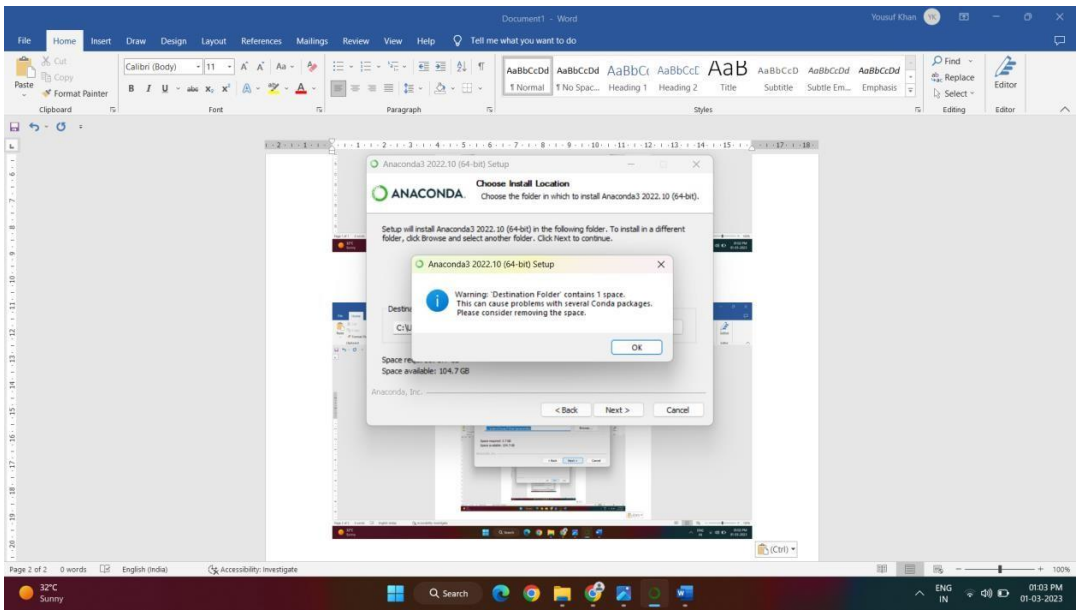
1a. Installation of python environment/Ananconda IDE for machine learning: installing python modules / packages like scikit-learn, Keras and Tensorflow etc.

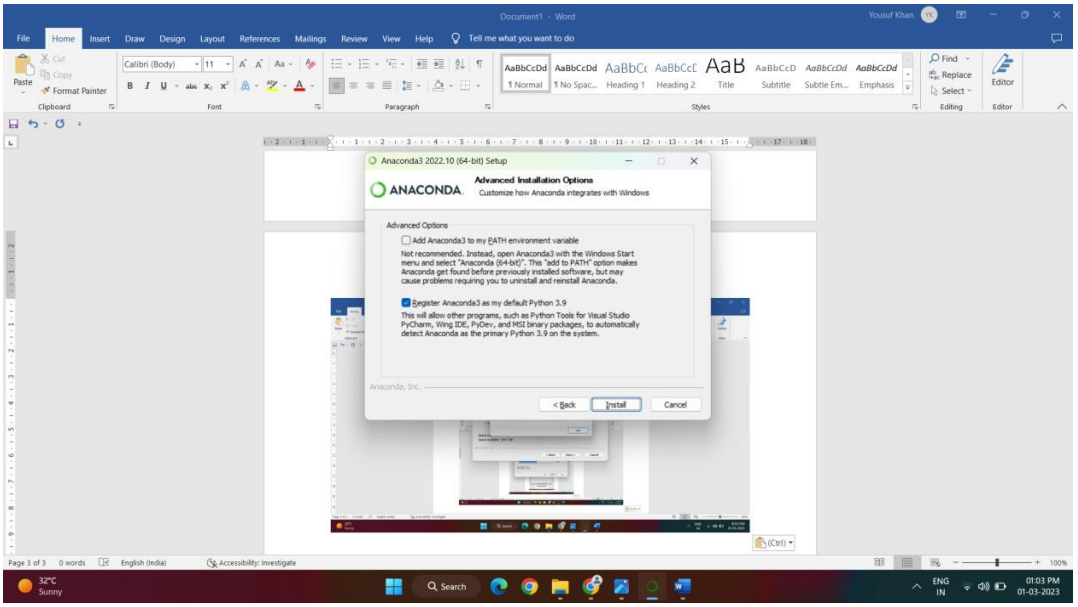
https://www.anaconda.com/products/distribution (link for anaconda installation) ON WINDOWS 10/11

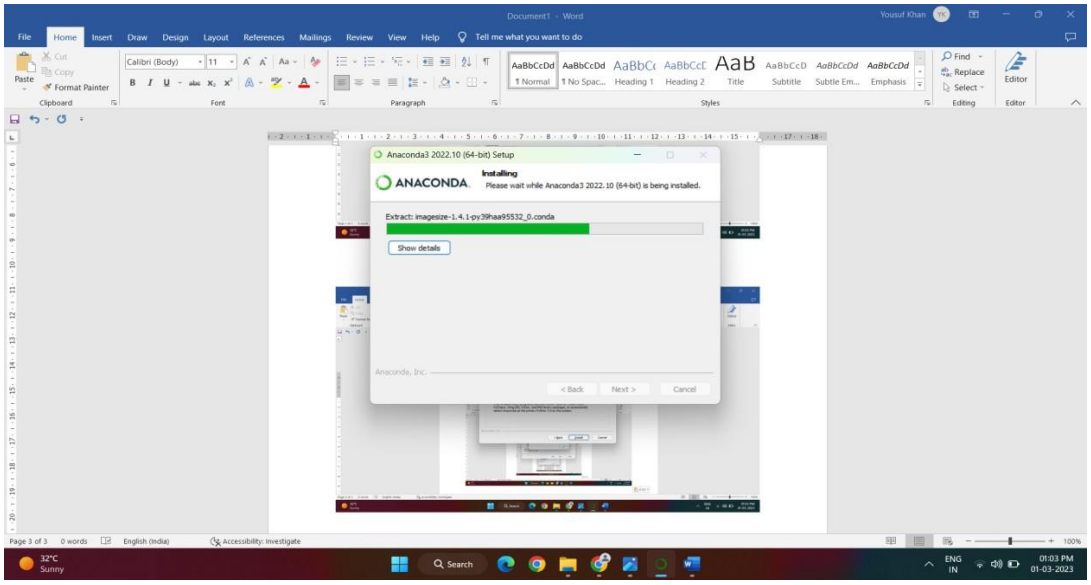


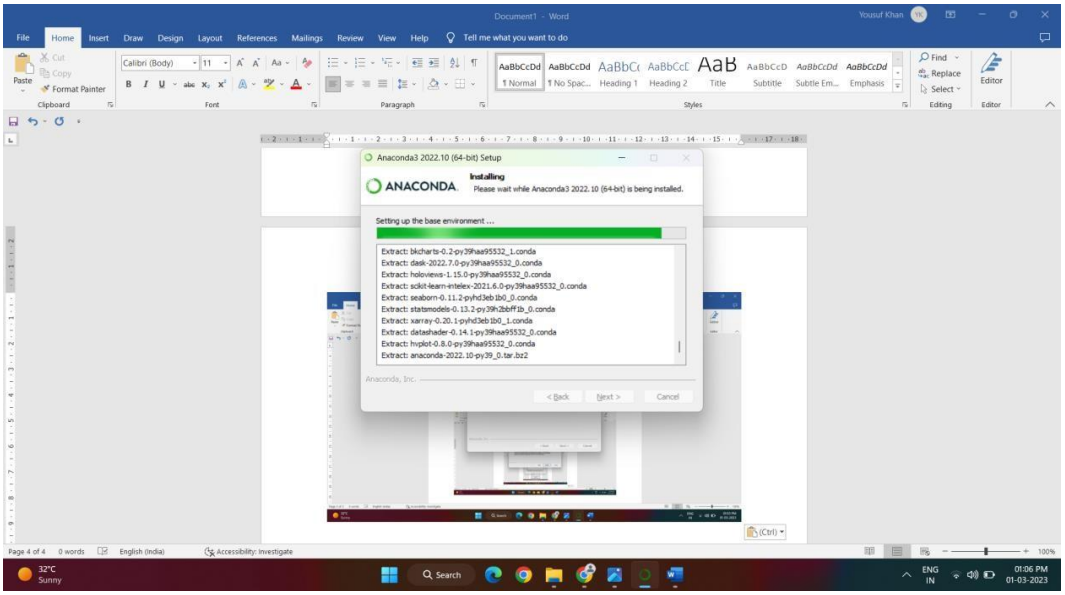

Anaconda3 2022.10 (64-bit) Setup
ANACONDA.
Select Installation Type
Please select the type of installation you would like to perform for Anaconda3 2022.10 (64-bit).
Install for:
Just Me (recommended)
All Users (requires admin privileges)
Anaconda, Inc.
< Back
Next >
Cancel

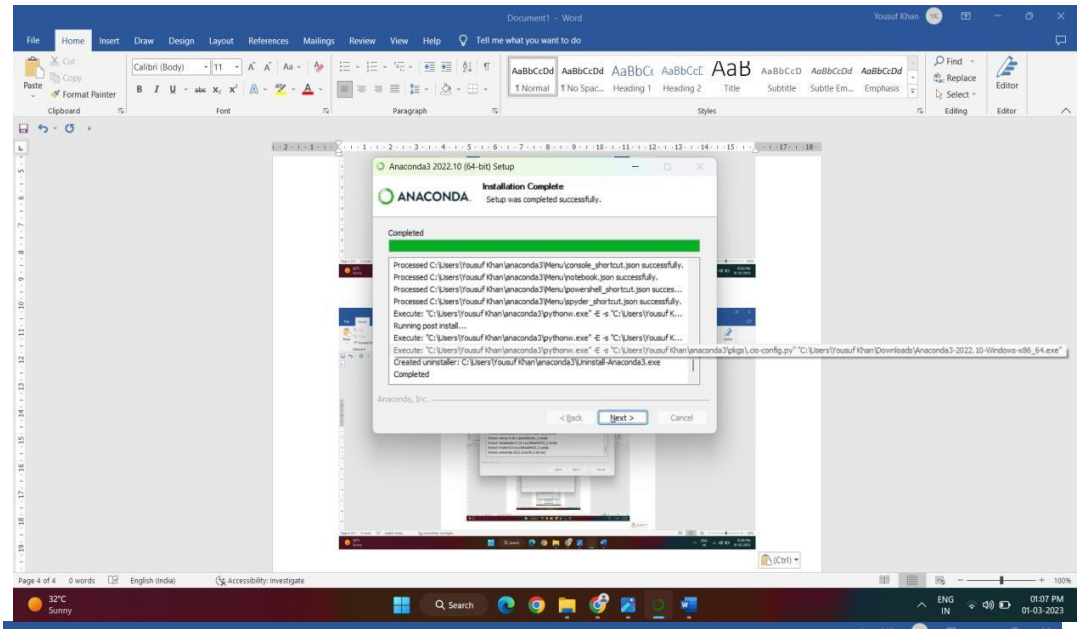

Anaconda3 2022.10 (64-bit) Setup
ANACONDA.
Choose Install Location
Choose the folder in which to install Anaconda3 2022.10 (64-bit).
Setup will install Anaconda3 2022.10 (64-bit) in the following folder. To install in a different folder, click Browse and select another folder. Click Next to continue.
Destination Folder
Browse...
Space required: 3.7 GB
Space available: 104.7 GB
Anaconda, Inc.
< Back
Next >
Cancel

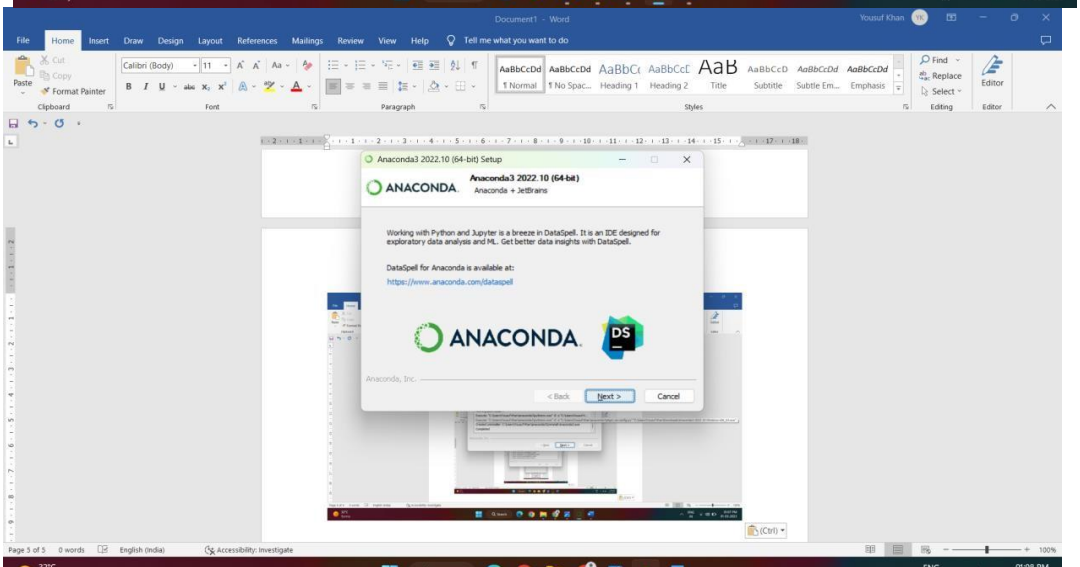

Anaconda3 2022.10 (64-bit) Setup
ANACONDA.
Choose Install Location
Choose the folder in which to install Anaconda3 2022.10 (64-bit).
Setup will install Anaconda3 2022.10 (64-bit) in the following folder. To install in a different folder, click Browse and select another folder. Click Next to continue.
Anaconda3 2022.10 (64-bit) Setup
Warning: 'Destination Folder' contains 1 space.
This can cause problems with several Conda packages.
Please consider removing the space.
OK
Space available: 104.7 GB
Anaconda, Inc.
< Back
Next >
Cancel

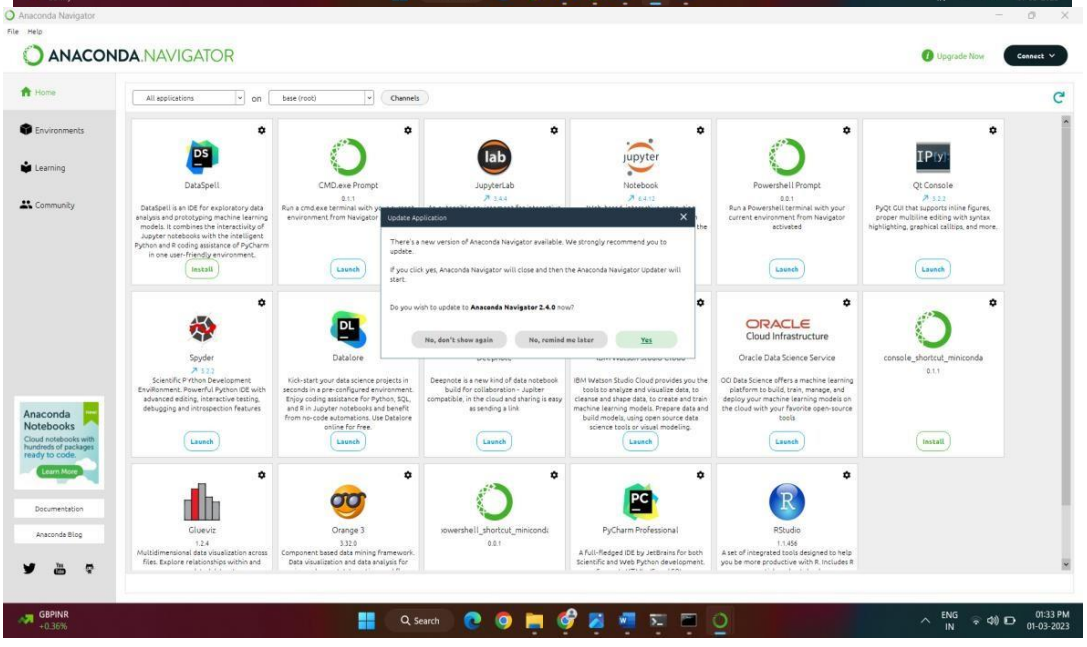

Anaconda3 2022.10 (64-bit) Setup
ANACONDA.
Advanced Installation Options
Customize how Anaconda integrates with Windows
Advanced Options
Add Anaconda3 to my PATH environment variable
Not recommended. Instead, open Anaconda3 with the Windows Start menu and select "Anaconda (64-bit)". This "add to PATH" option makes Anaconda get found before previously installed software, but may cause problems requiring you to uninstall and reinstall Anaconda.
Register Anaconda3 as my default Python 3.9
This will allow other programs, such as Python Tools for Visual Studio PyCharm, Wing IDE, PyDev, and MSI binary packages, to automatically detect Anaconda as the primary Python 3.9 on the system.
Anaconda, Inc.
< Back
Install
Cancel

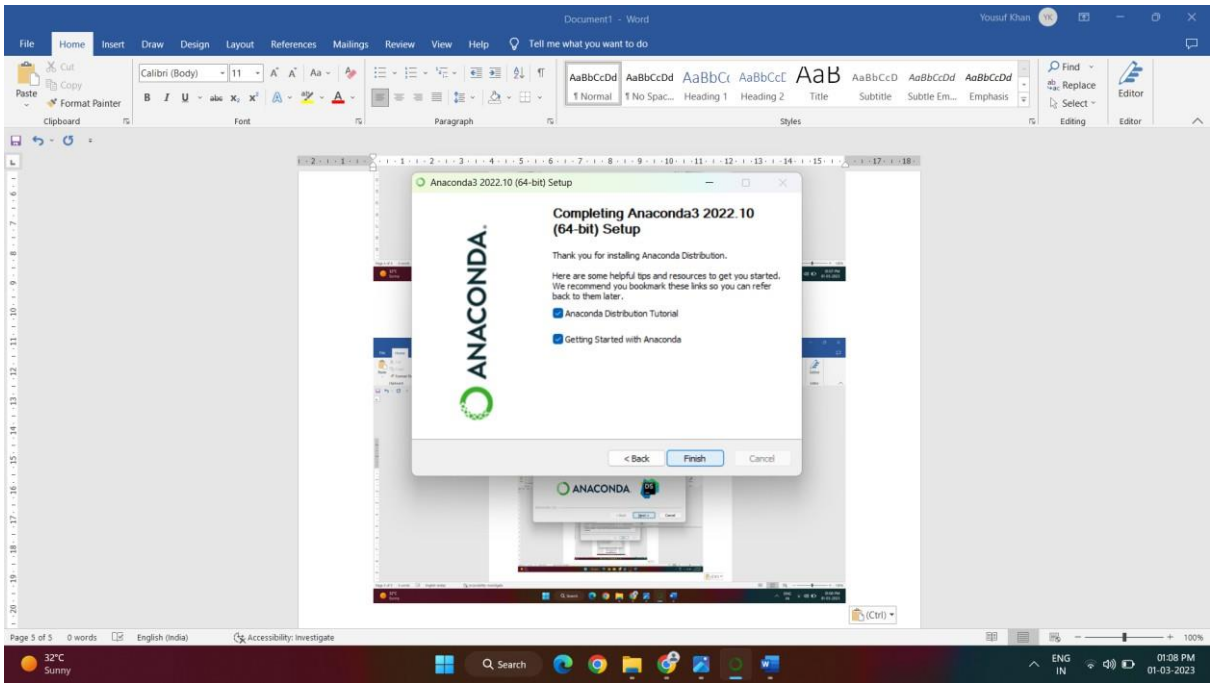

Anaconda3 2022.10 (64-bit) Setup
ANACONDA.
Installing
Please wait while Anaconda3 2022.10 (64-bit) is being installed.
Extract: imagesize-1.4.1-py39haa95532_0.conda
Show details
Anaconda, Inc.
< Back
Next >
Cancel


Anaconda3 2022.10 (64-bit) Setup
ANACONDA.
Installing
Please wait while Anaconda3 2022.10 (64-bit) is being installed.
Setting up the base environment ...
Extract: bkcharts-0.2-py39haa95532_1.conda
Extract: dask-2022.7.0-py39haa95532_0.conda
Extract: holoviews-1.15.0-py39haa95532_0.conda
Extract: scikit-learn-intelex-2021.6.0-py39haa95532_0.conda
Extract: seaborn-0.11.2-pyhd3eb1b0_0.conda
Extract: statsmodels-0.13.2-py39h2bbff1b_0.conda
Extract: xarray-0.20.1-pyhd3eb1b0_1.conda
Extract: datashader-0.14.1-py39haa95532_0.conda
Extract: hvplot-0.8.0-py39haa95532_0.conda
Extract: anaconda-2022.10-py39_0.tar.bz2
Anaconda, Inc.
< Back
Next >
Cancel


Anaconda3 2022.10 (64-bit) Setup
ANACONDA.
Installation Complete
Setup was completed successfully.
Completed
Next >
Cancel


Anaconda3 2022.10 (64-bit) Setup
ANACONDA.
Anaconda3 2022.10 (64-bit)
Anaconda + JetBrains
Working with Python and Jupyter is a breeze in DataSpell. It is an IDE designed for exploratory data analysis and ML. Get better data insights with DataSpell.
DataSpell for Anaconda is available at:
https://www.anaconda.com/dataspell
Next >
Cancel


ANACONDA.NAVIGATOR
Upgrade Now
Connect
Home
Environments
Learning
Community
DataSpell
CMD.exe Prompt
JupyterLab
Notebook
Powershell Prompt
Qt Console
Spyder
Datalore
Oracle Data Science Service
console_shortcut_miniconda
Gluoviz
Orange 3
PyCharm Professional
RStudio
Anaconda Notebooks
Update Application
There's a new version of Anaconda Navigator available. We strongly recommend you to update.
If you click yes, Anaconda Navigator will close and then the Anaconda Navigator Updater will start.
Do you wish to update to Anaconda Navigator 2.4.0 now?
No, don't show again
No, remind me later
Yes



Installing Package–Pandas: Pandas is a powerful data analysis library in Python that provides easy-to-use data structures and data analysis tools. It is widely used in data science and machine learning projects for data manipulation, cleaning, and analysis. Python is a widely used programming language for data science and analysis due to its versatility and powerful libraries. To install the Pandas library in Python, follow these steps:

Step 1: First, ensure that Python is already installed on your system. If not, you can easily download it from the official Python website (python.org).
Step 2: Next, open your terminal or command prompt to proceed with the installation.
Step 3: Type the command pip install pandas and hit Enter to initiate the installation process. This command will guide you on installing pandas in Python using pip.
Step 4: Finally, verify the successful installation of Pandas by running the command pip show pandas to confirm that pandas is successfully installed in your system.

Installing Package– Keras and Tensor flow: Keras and TensorFlow are open source Python libraries for working with neural networks, creating machine learning models and performing deep learning. Because Keras is a high level API for TensorFlow, they are installed together.

In general, there are two ways to install Keras and TensorFlow:

Install a Python distribution that includes hundreds of popular packages (including Keras and TensorFlow) such as ActivePython.
Use pip to install TensorFlow, which will also install Keras at the same time.

Pip Install TensorFlow
Instead of pip installing each package separately, the recommended approach is to install Keras as part of the TensorFlow installation. When you install TensorFlow 2.0+, Keras will be automatically

installed, as well. TensorFlow Requirements: TensorFlow and Keras require Python 3.6+ (Python 3.8 requires TensorFlow 2.2+) , and the latest version of pip.

Installing Packages-Scipy,numpy,Scikit: NumPy is a scientific computing library for Python. It provides powerful tools for manipulating and analyzing numerical data. It is used for array-based computations, linear algebra, Fourier transforms random number functions, and more.

Scikit-learn: Scikit-learn is a machine-learning library for Python.
SciPy is an open-source Python library which is used to solve scientific and mathematical problems. It is built on the NumPy extension and allows the user to manipulate and visualize data with a wide range of high-level commands. SciPy builds on NumPy and therefore if you import SciPy, there is no need to import NumPy.
Both NumPy and SciPy are Python libraries used for used mathematical and numerical analysis. NumPy contains array data and basic operations such as sorting, indexing, etc whereas, SciPy consists of all the numerical code.
Though NumPy provides a number of functions that can help resolve linear algebra, Fourier transforms, etc, SciPy is the library that actually contains fully-featured versions of these functions along with many others. However, if you are doing scientific analysis using Python, you will need to install both NumPy and SciPy since SciPy builds on NumPy.

At command prompt 1. pip install numpy 2. pip install scipy 3. pip install scikt-learn