END TO END MACHINE LEARNING PROJECT

WHAT WE COVER---

1. SET UP PROJECT WITH GITHUB
2. DATA INGESTION
3. DATA TRANSFORMATION
4. MODEL TRAINER
5. MODEL EVALUATION
6. MODEL DEPLOYMENT
7. CI/CD PIPELINES-GITHUB ACTIONS
8. DEPLOYEMENT – AWS

PREREQUISTE- PYTHON PROGRAMMING LANGUAGE, MODULAR CODING, DEEP LEARNING

Operating System – Windows

Download the Anaconda- https://repo.anaconda.com/archive/Anaconda3-2020.02-Windows-x86\_64

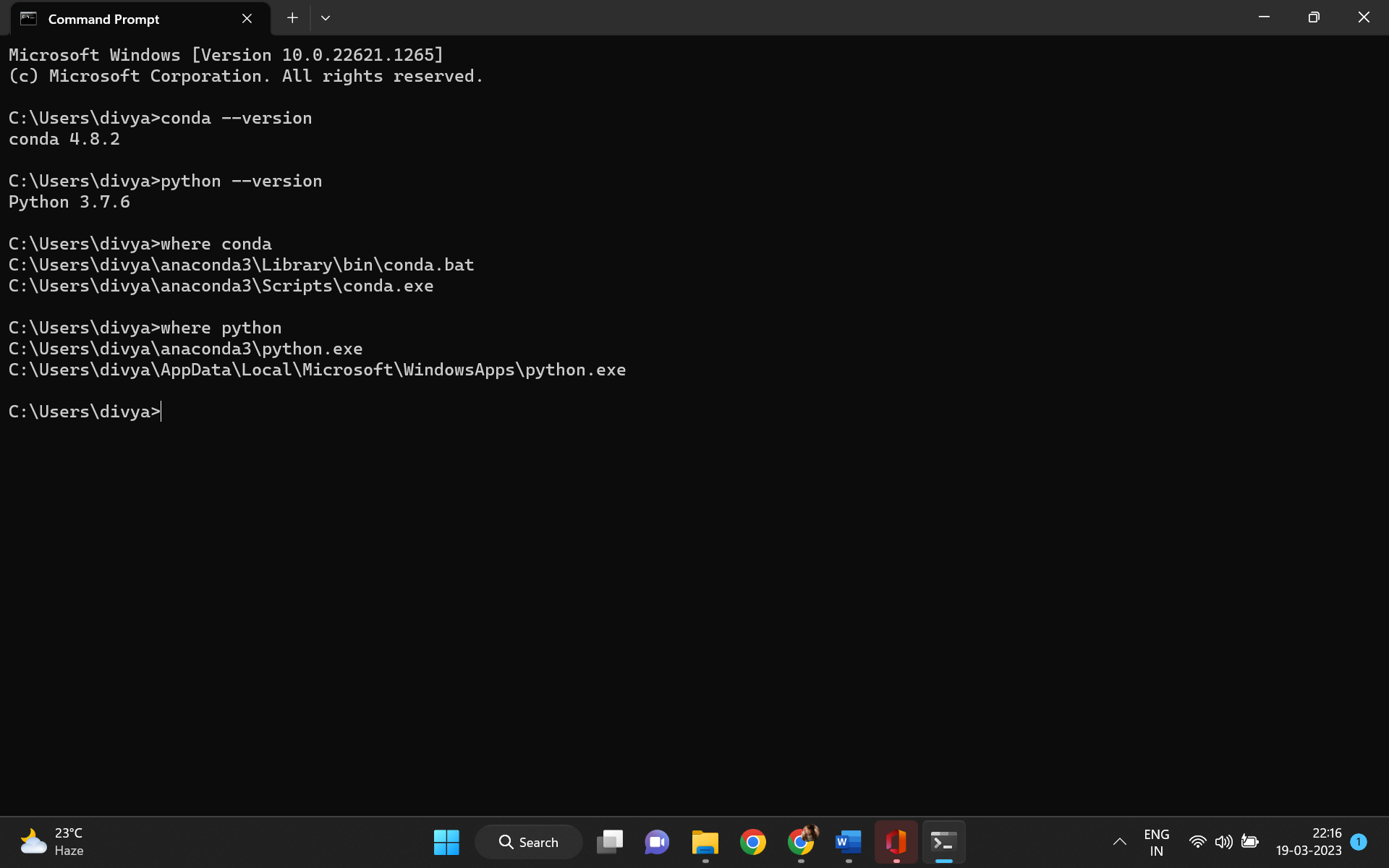
Installation of Anaconda – https://www.datacamp.com/tutorial/installing-anaconda-windows

Text Editor – Visual studio

Install Visual Studio - https://code.visualstudio.com/docs/setup/windows

Python version – 3.7.6

Conda version- 4.8.2



Tutorial –1 SET UP PROJECT WITH GITHUB

AGENDA—

1. SET UP THE GITHUB {REPOSITORY}

* NEW ENVIRONMENT
* Setup.py
* Requirements.txt

1. Src folder and Build the package

Prerequisite –

1. Basic knowledge about GitHub
2. Basic Knowledge about Python
3. Difference – git vs github

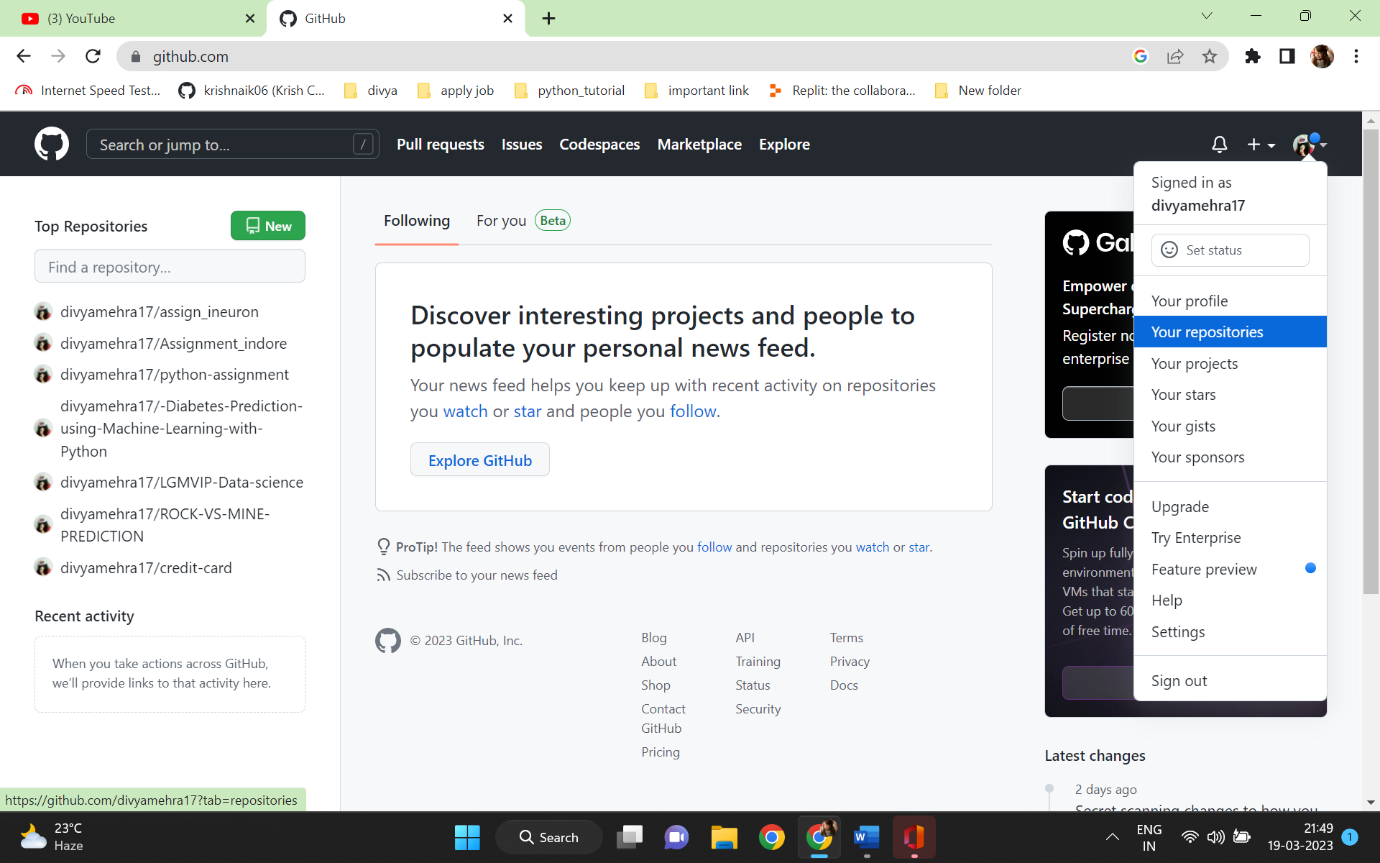
Git is used to storing the source code for a project and track the complete history of all changes to that code, while GitHub is a cloud-based platform built around the Git tool. The major difference is that git is software that a developer can locally install on it machine to manage source code while GitHub is an online service to which developers who use Git can connect and upload or download resources. Github is **hosting service** for **git repositories**.

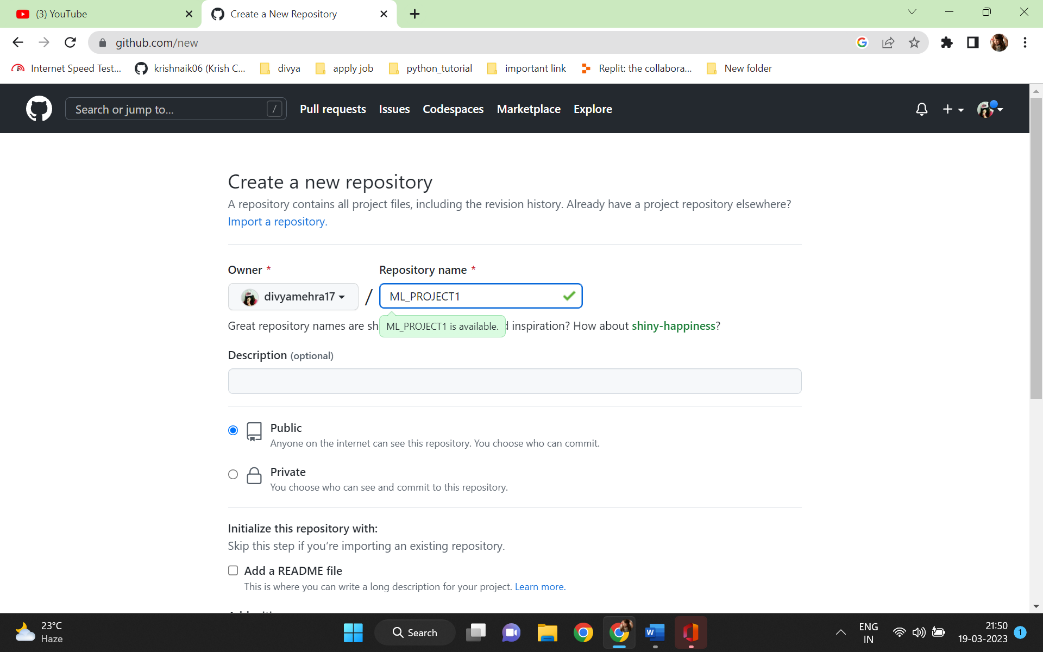
1. Difference git clone vs git init

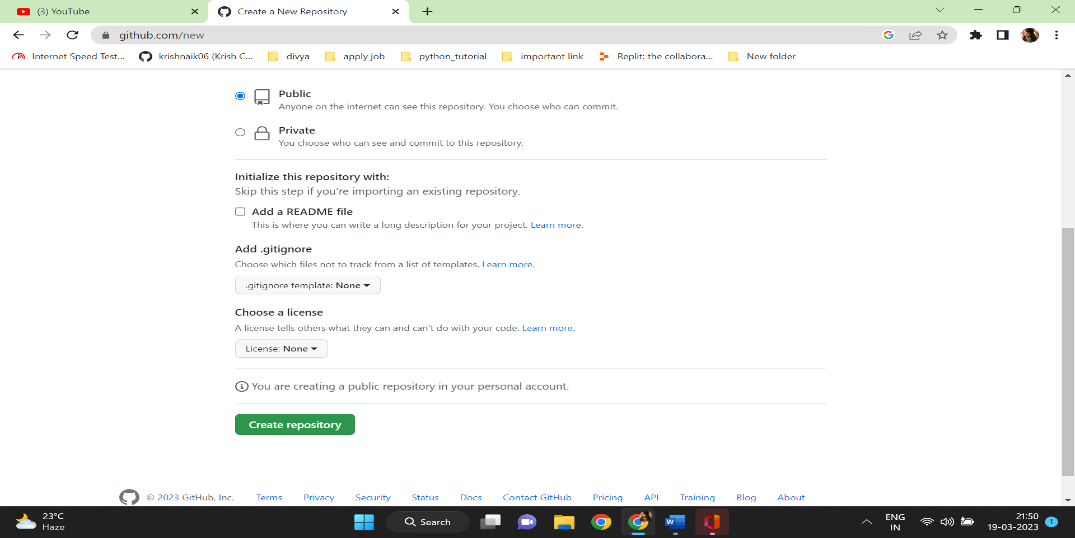
git clone is dependent on git init. git clone is used to create a copy of existing repository but git clone itself first calls git init to create new repository.

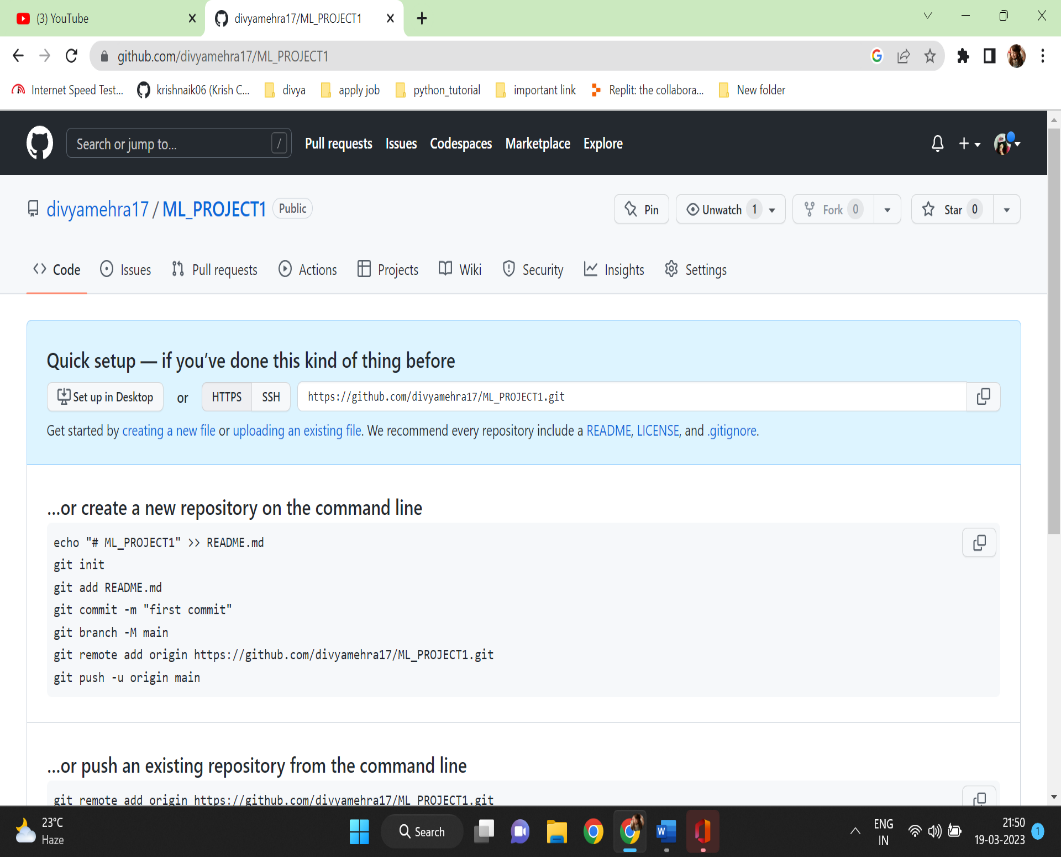
Step:

1. Create a github account and Create a Repository



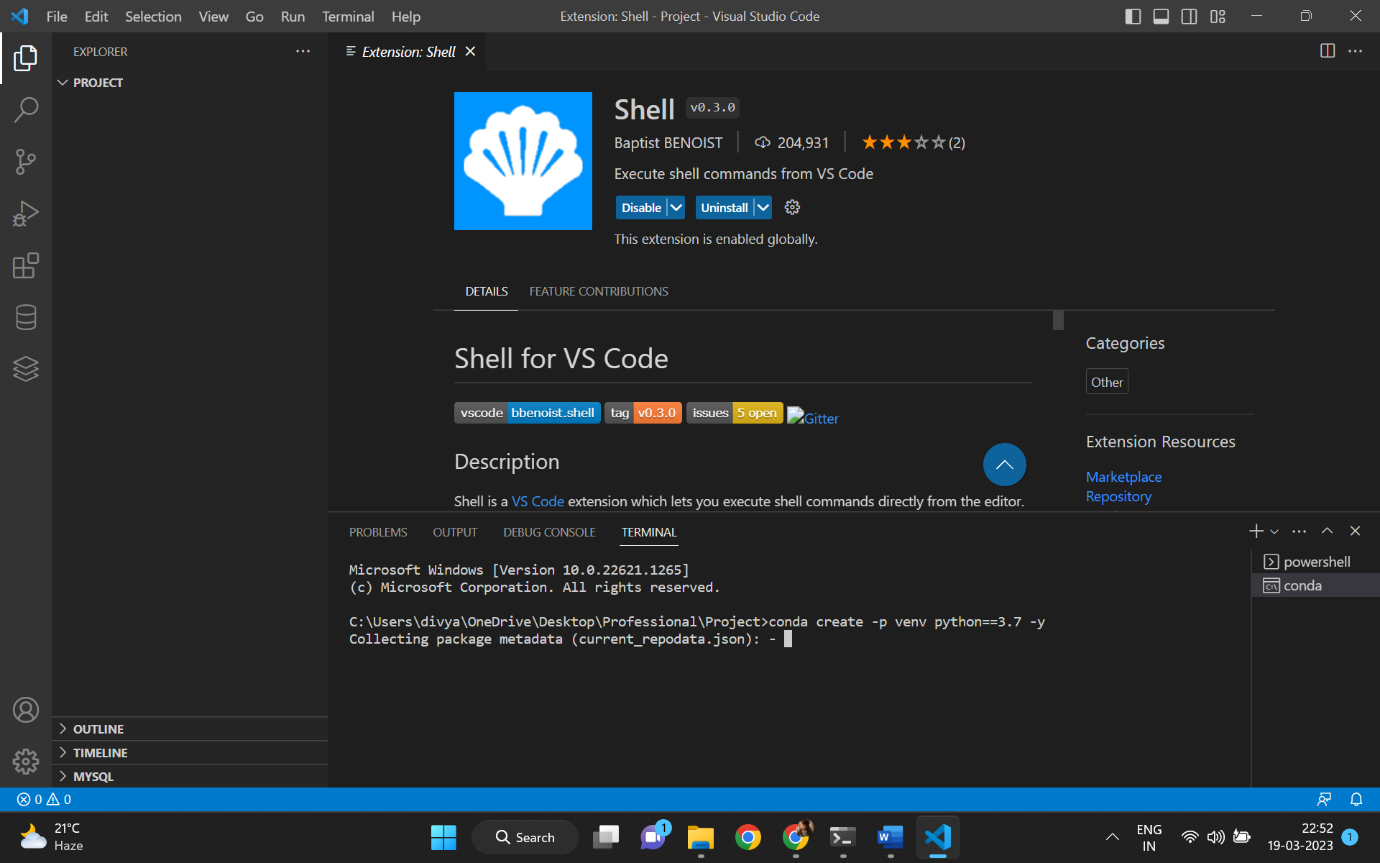






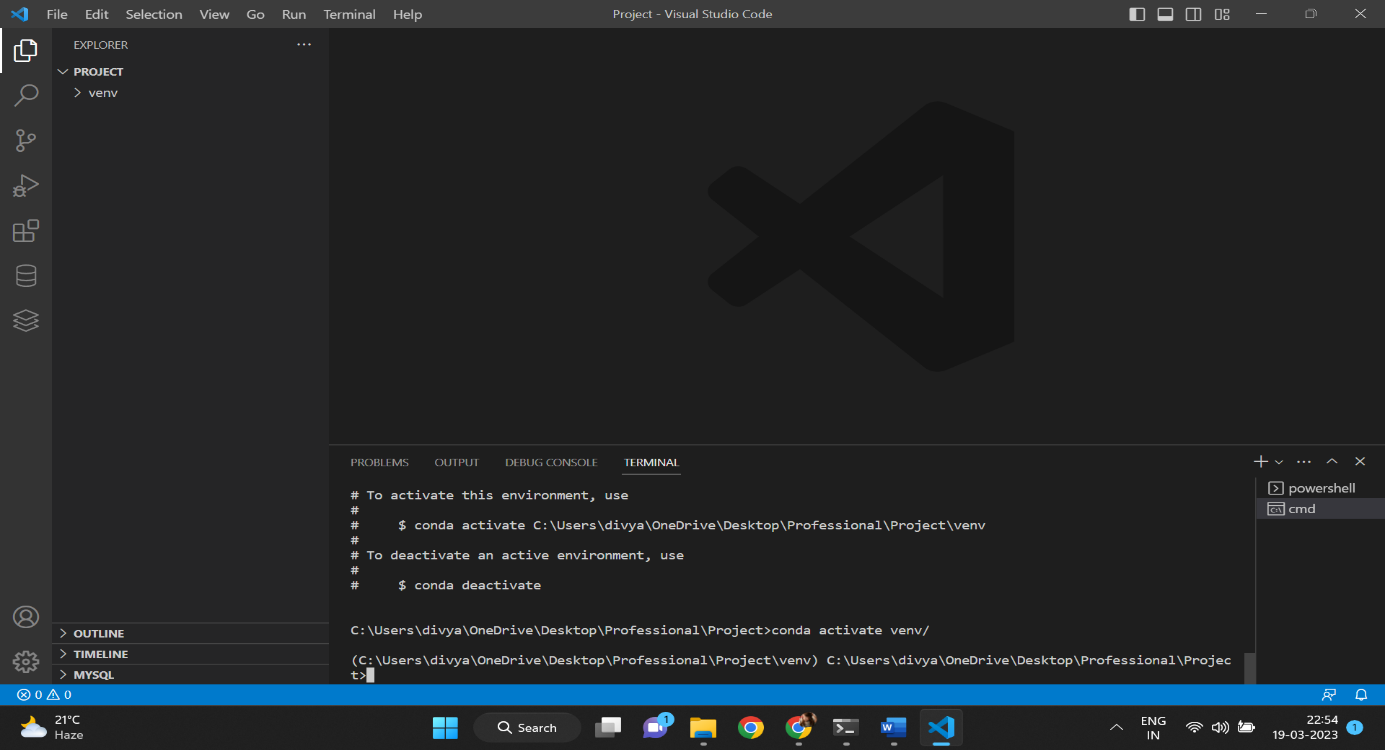
1. Creating a virtual environment in Visual Studio

To create a virtual environment syntax - C:\Users\divya\OneDrive\Desktop\Professional\Project>conda create -p venv python==3.7 -y



To activate the virtual environment syntax-

C:\Users\divya\OneDrive\Desktop\Professional\Project>conda activate venv/



1. Cloning with your repository

(C:\Users\divya\OneDrive\Desktop\Professional\Project\venv) C:\Users\divya\OneDrive\Desktop\Professional\Project>git init

Reinitialized existing Git repository in C:/Users/divya/OneDrive/Desktop/Professional/Project/.git/

(C:\Users\divya\OneDrive\Desktop\Professional\Project\venv) C:\Users\divya\OneDrive\Desktop\Professional\Project>git add README.md

(C:\Users\divya\OneDrive\Desktop\Professional\Project\venv) C:\Users\divya\OneDrive\Desktop\Professional\Project>git commit -m "first Commit"

On branch main

Your branch is up to date with 'origin/main'.

Untracked files:

(use "git add <file>..." to include in what will be committed)

venv/

nothing added to commit but untracked files present (use "git add" to track)

(C:\Users\divya\OneDrive\Desktop\Professional\Project\venv) C:\Users\divya\OneDrive\Desktop\Professional\Project>git branch -M main

(C:\Users\divya\OneDrive\Desktop\Professional\Project\venv) C:\Users\divya\OneDrive\Desktop\Professional\Project>git remote add origin https://github.com/divyamehra17/ML\_PROJECT1.git

error: remote origin already exists.

(C:\Users\divya\OneDrive\Desktop\Professional\Project\venv) C:\Users\divya\OneDrive\Desktop\Professional\Project>git push -u origin main

Enumerating objects: 3, done.

Counting objects: 100% (3/3), done.

Writing objects: 100% (3/3), 211 bytes | 211.00 KiB/s, done.

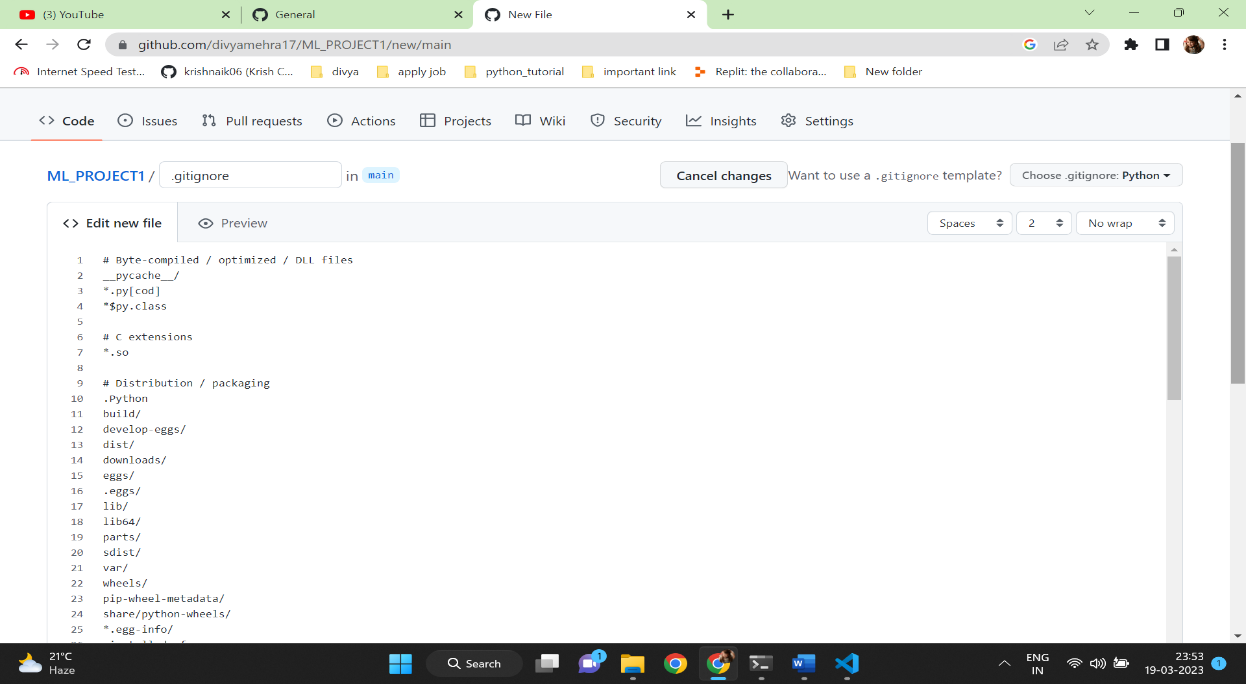
Total 3 (delta 0), reused 0 (delta 0), pack-reused 0

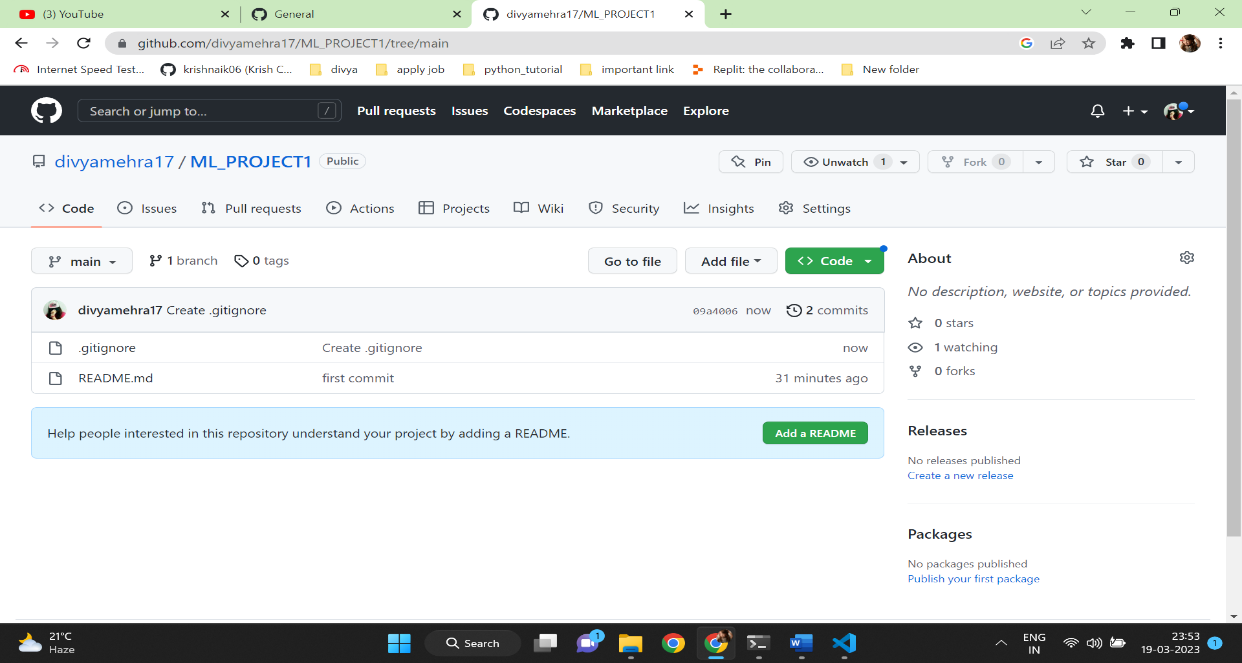
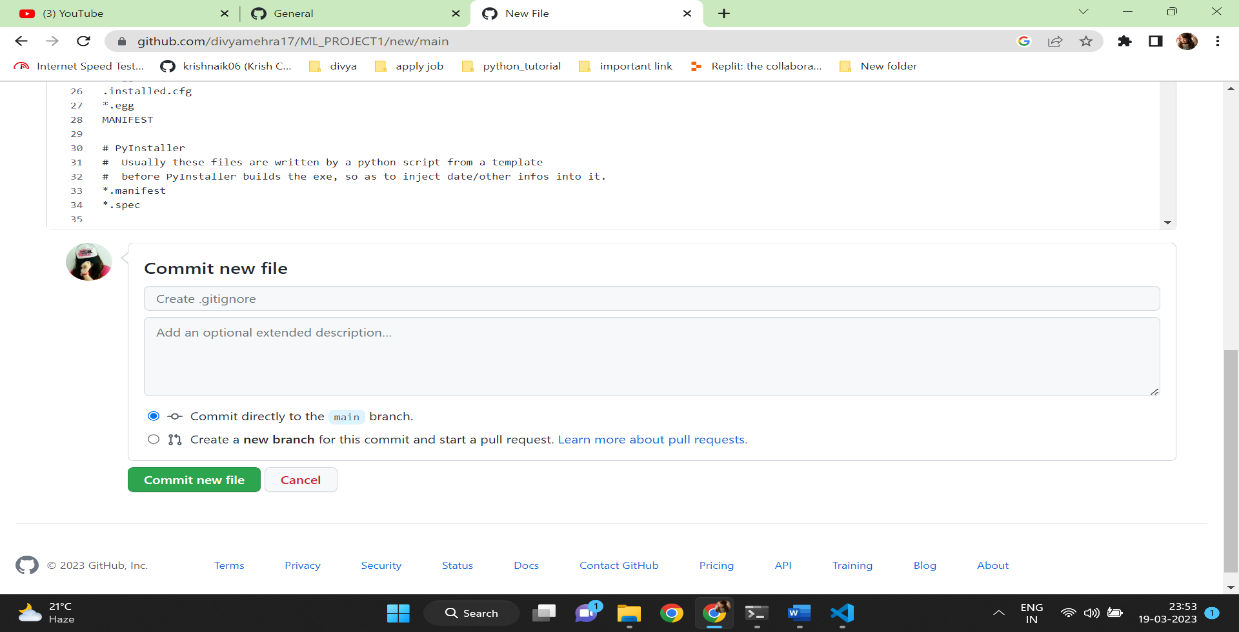
To https://github.com/divyamehra17/ML\_PROJECT1.git

\* [new branch] main -> main

branch 'main' set up to track 'origin/main'.

1. Create a .gitignore

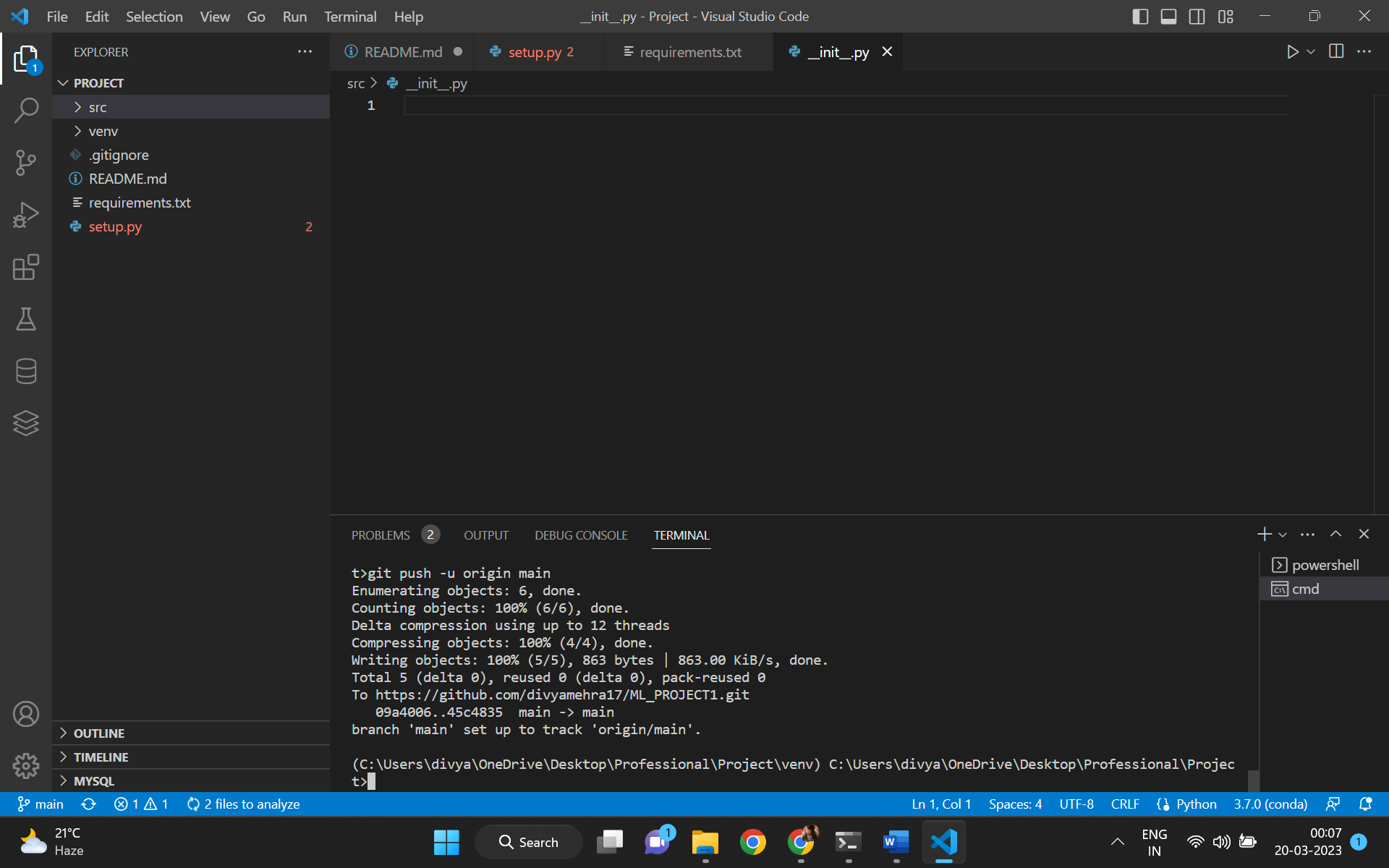




Syntax in visual studio – git pull (for updation)

1. Creating a setup.py and requirements.txt file in visual studio

* Create a setup.py and requirements.txt file and then src folder inside this folder create \_\_init\_\_.py
* Run in the terminal : pip install -r requirements.txt
* Run the git command-



(C:\Users\divya\OneDrive\Desktop\Professional\Project\venv) C:\Users\divya\OneDrive\Desktop\Professional\Project>git add .

(C:\Users\divya\OneDrive\Desktop\Professional\Project\venv) C:\Users\divya\OneDrive\Desktop\Professional\Project>git commit -m "setup and requirement"

[main 45c4835] setup and requirement

3 files changed, 36 insertions(+)

create mode 100644 requirements.txt

create mode 100644 setup.py

create mode 100644 src/\_\_init\_\_.py

(C:\Users\divya\OneDrive\Desktop\Professional\Project\venv) C:\Users\divya\OneDrive\Desktop\Professional\Project>git branch -M main

(C:\Users\divya\OneDrive\Desktop\Professional\Project\venv) C:\Users\divya\OneDrive\Desktop\Professional\Project>git remote add origin https://github.com/divyamehra17/PROJECT\_ML1.git

error: remote origin already exists.

(C:\Users\divya\OneDrive\Desktop\Professional\Project\venv) C:\Users\divya\OneDrive\Desktop\Professional\Project>git push -u origin main

Enumerating objects: 6, done.

Counting objects: 100% (6/6), done.

Delta compression using up to 12 threads

Compressing objects: 100% (4/4), done.

Writing objects: 100% (5/5), 863 bytes | 863.00 KiB/s, done.

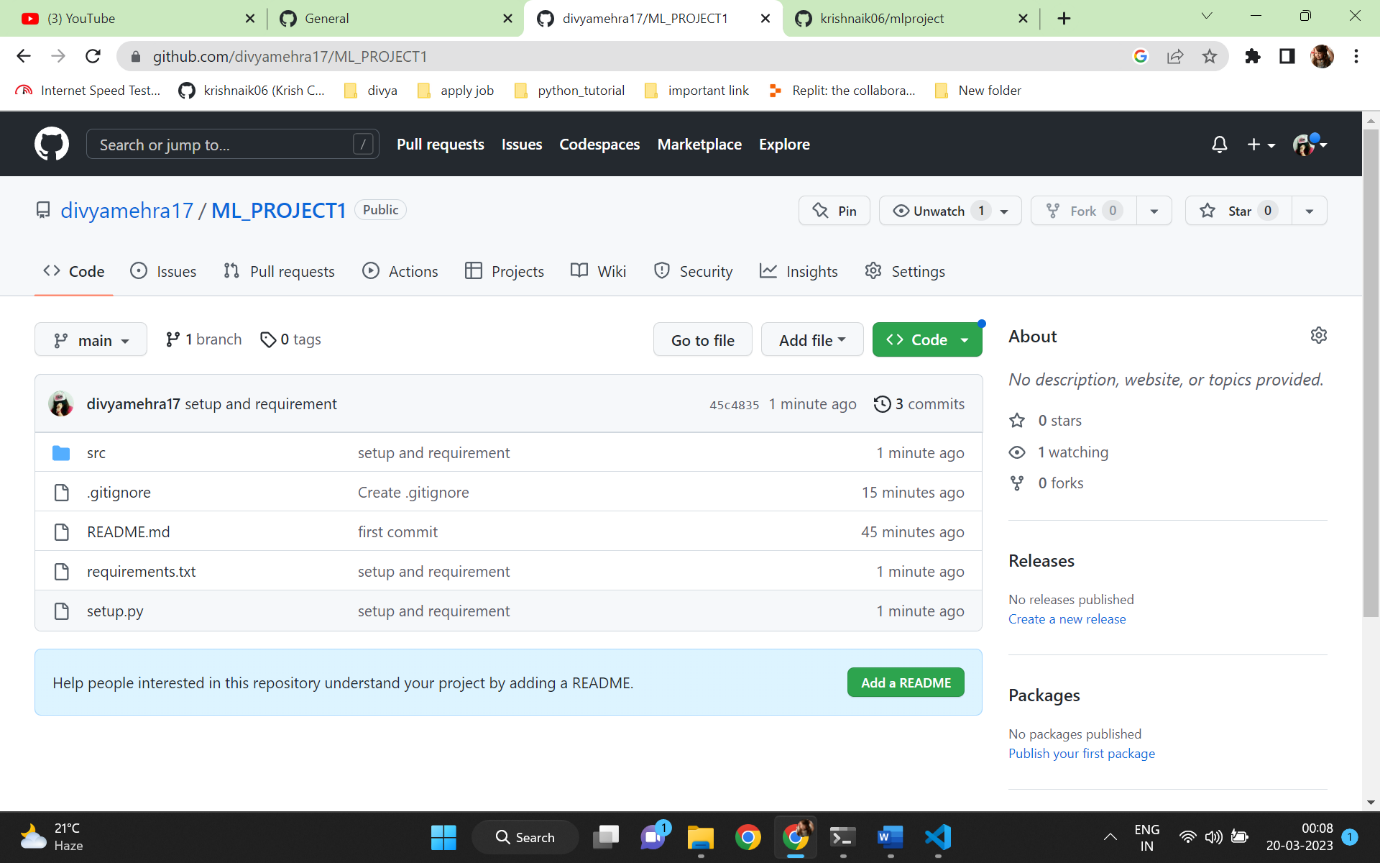
Total 5 (delta 0), reused 0 (delta 0), pack-reused 0

To https://github.com/divyamehra17/ML\_PROJECT1.git

09a4006..45c4835 main -> main

branch 'main' set up to track 'origin/main'.

Final output in github---



Tutorial -2

Create a src folder inside src folder –

1. Components – inside this folder data\_ingestion,data\_transformation,model\_trainer.
2. Pipeline – train\_pipeline,predict\_pipeline
3. Logger.py and exception.py

