

**Modular Machine Learning Pipeline for Housing Price Prediction using GitHub CI/CD**

**Assignment-1**

**Machine Learning Operations**

**Submitted by**

Shreyas B Gaikwad

Roll No. G24AI1060

Post Graduate Diploma in Data Engineering

**Github Link**: <https://github.com/geek1060/HousingRegression>

Steps to follow:

1. Check for the availability of following environments; if not available install the required versions to ensure that further process can be carried out.

(a) WSL

(b) Python

(c) Conda

(d) Git

Microsoft Windows [Version 10.0.19045.5965]

(c) Microsoft Corporation. All rights reserved.

C:\Users\Shreyas>wsl --version

WSL version: 2.3.26.0

Kernel version: 5.15.167.4-1

WSLg version: 1.0.65

MSRDC version: 1.2.5620

Direct3D version: 1.611.1-81528511

DXCore version: 10.0.26100.1-240331-1435.ge-release

Windows version: 10.0.19045.5965

C:\Users\Shreyas>python --version

Python 3.13.5

C:\Users\Shreyas>conda --version

conda 25.5.1

C:\Users\Shreyas>git --version

git version 2.50.0.windows.1

2. Check the SSH key availability and connection to GitHub.

C:\Users\Shreyas>dir /a C:\Users\Shreyas\.ssh

Volume in drive C is OS\_SSD

Volume Serial Number is E6DB-D98C

Directory of C:\Users\Shreyas\.ssh

30-06-2025 21:33 <DIR> .

30-06-2025 21:33 <DIR> ..

30-06-2025 21:20 464 id\_ed25519

30-06-2025 21:20 103 id\_ed25519.pub

30-06-2025 21:33 831 known\_hosts

30-06-2025 21:32 93 known\_hosts.old

4 File(s) 1,491 bytes

2 Dir(s) 58,287,251,456 bytes free

C:\Users\Shreyas>ssh -T git@github.com

Enter passphrase for key 'C:\Users\Shreyas/.ssh/id\_ed25519':

Hi geek1060! You've successfully authenticated, but GitHub does not provide shell access.

C:\Users\Shreyas>

3. Create Conda environment.

4. Activate the created Conda Environment.

5. Change Directory to the project folder. (Additional step to ensure that compiling becomes easy and can be stored independently as required)

C:\Users\Shreyas>conda create -n HousingRegression python=3.13.5 -y

Channels:

- defaults

Platform: win-64

Collecting package metadata (repodata.json): done

Solving environment: done

## Package Plan ##

environment location: C:\Users\Shreyas\miniconda3\envs\HousingRegression

added / updated specs:

- python=3.13.5

The following NEW packages will be INSTALLED:

bzip2 pkgs/main/win-64::bzip2-1.0.8-h2bbff1b\_6

ca-certificates pkgs/main/win-64::ca-certificates-2025.2.25-haa95532\_0

expat pkgs/main/win-64::expat-2.7.1-h8ddb27b\_0

libffi pkgs/main/win-64::libffi-3.4.4-hd77b12b\_1

libmpdec pkgs/main/win-64::libmpdec-4.0.0-h827c3e9\_0

openssl pkgs/main/win-64::openssl-3.0.16-h3f729d1\_0

pip pkgs/main/noarch::pip-25.1-pyhc872135\_2

python pkgs/main/win-64::python-3.13.5-h286a616\_100\_cp313

python\_abi pkgs/main/win-64::python\_abi-3.13-0\_cp313

setuptools pkgs/main/win-64::setuptools-78.1.1-py313haa95532\_0

sqlite pkgs/main/win-64::sqlite-3.45.3-h2bbff1b\_0

tk pkgs/main/win-64::tk-8.6.14-h5e9d12e\_1

tzdata pkgs/main/noarch::tzdata-2025b-h04d1e81\_0

vc pkgs/main/win-64::vc-14.42-haa95532\_5

vs2015\_runtime pkgs/main/win-64::vs2015\_runtime-14.42.34433-hbfb602d\_5

wheel pkgs/main/win-64::wheel-0.45.1-py313haa95532\_0

xz pkgs/main/win-64::xz-5.6.4-h4754444\_1

zlib pkgs/main/win-64::zlib-1.2.13-h8cc25b3\_1

Downloading and Extracting Packages:

Preparing transaction: done

Verifying transaction: done

Executing transaction: done

#

# To activate this environment, use

#

# $ conda activate HousingRegression

#

# To deactivate an active environment, use

#

# $ conda deactivate

C:\Users\Shreyas>conda activate HousingRegression

(HousingRegression) C:\Users\Shreyas>cd /d D:\MLOPS\assign1

(HousingRegression) D:\MLOPS\assign1>

6. Create a new branch and push it to the repo.

(HousingRegression) D:\MLOPS\assign1>git push -u origin main

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

Enumerating objects: 3, done.

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

Writing objects: 100% (3/3), 249 bytes | 124.00 KiB/s, done.

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

To github.com:geek1060/HousingRegression.git

\* [new branch] main -> main

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

7. Initialise git, add README.md file to main branch and commit.

(HousingRegression) D:\MLOPS\assign1>echo "# HousingRegression" > README.md

(HousingRegression) D:\MLOPS\assign1>git init

Initialized empty Git repository in D:/MLOPS/assign1/.git/

(HousingRegression) D:\MLOPS\assign1>git add README.md

(HousingRegression) D:\MLOPS\assign1>git commit -m "Initial commit with README"

[master (root-commit) aadb6e4] Initial commit with README

1 file changed, 1 insertion(+)

create mode 100644 README.md

(HousingRegression) D:\MLOPS\assign1>git remote add origin git@github.com:geek1060/HousingRegression.git

(HousingRegression) D:\MLOPS\assign1>git branch -M main

(HousingRegression) D:\MLOPS\assign1>git push -u origin main

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

Enumerating objects: 3, done.

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

Writing objects: 100% (3/3), 249 bytes | 124.00 KiB/s, done.

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

To github.com:geek1060/HousingRegression.git

\* [new branch] main -> main

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

8. Print the required libraries to the file. Create a new branch reg\_branch and add the requirements.txt, utils.py and regression.py files to the branch.

(HousingRegression) D:\MLOPS\assign1>echo numpy >> requirements.txt

(HousingRegression) D:\MLOPS\assign1>echo pandas >> requirements.txt

(HousingRegression) D:\MLOPS\assign1>echo scikit-learn >> requirements.txt

(HousingRegression) D:\MLOPS\assign1>echo statsmodels >> requirements.txt

(HousingRegression) D:\MLOPS\assign1>echo matplotlib >> requirements.txt

(HousingRegression) D:\MLOPS\assign1>git checkout -b reg\_branch

Switched to a new branch 'reg\_branch'

(HousingRegression) D:\MLOPS\assign1>git add .

(HousingRegression) D:\MLOPS\assign1>git commit -m "Add regression code, utils, requirements"

[reg\_branch 8d1ffe5] Add regression code, utils, requirements

3 files changed, 66 insertions(+)

create mode 100644 regression.py

create mode 100644 requirements.txt

create mode 100644 utils.py

10. Create a new directory .github and add the workflows folder and the .yml file. Then push the yml file to the reg\_branch.

(HousingRegression) D:\MLOPS\assign1>mkdir .github

(HousingRegression) D:\MLOPS\assign1>mkdir .github\workflows

(HousingRegression) D:\MLOPS\assign1>notepad .github\workflows\ci.yml

(HousingRegression) D:\MLOPS\assign1>git add .github/workflows/ci.yml

(HousingRegression) D:\MLOPS\assign1>git commit -m "Add GitHub Actions CI workflow"

[reg\_branch 2e86d63] Add GitHub Actions CI workflow

1 file changed, 29 insertions(+)

create mode 100644 .github/workflows/ci.yml

(HousingRegression) D:\MLOPS\assign1>git push origin reg\_branch

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

Enumerating objects: 11, done.

Counting objects: 100% (11/11), done.

Delta compression using up to 4 threads

Compressing objects: 100% (8/8), done.

Writing objects: 100% (10/10), 2.03 KiB | 346.00 KiB/s, done.

Total 10 (delta 1), reused 0 (delta 0), pack-reused 0 (from 0)

remote: Resolving deltas: 100% (1/1), done.

remote:

remote: Create a pull request for 'reg\_branch' on GitHub by visiting:

remote: https://github.com/geek1060/HousingRegression/pull/new/reg\_branch

remote:

To github.com:geek1060/HousingRegression.git

\* [new branch] reg\_branch -> reg\_branch

11. Switch to the main branch and merge the reg\_branch with the main branch.

(HousingRegression) D:\MLOPS\assign1>git checkout main

Switched to branch 'main'

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

(HousingRegression) D:\MLOPS\assign1>git merge reg\_branch

Updating aadb6e4..2e86d63

Fast-forward

.github/workflows/ci.yml | 29 +++++++++++++++++++++++++++++

regression.py | 33 +++++++++++++++++++++++++++++++++

requirements.txt | 5 +++++

utils.py | 28 ++++++++++++++++++++++++++++

4 files changed, 95 insertions(+)

create mode 100644 .github/workflows/ci.yml

create mode 100644 regression.py

create mode 100644 requirements.txt

create mode 100644 utils.py

(HousingRegression) D:\MLOPS\assign1>git push

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

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

To github.com:geek1060/HousingRegression.git

aadb6e4..2e86d63 main -> main

12. Run the regression.py.

Code:

import numpy as np

import pandas as pd

from sklearn.linear\_model import LinearRegression

from sklearn.tree import DecisionTreeRegressor

from sklearn.ensemble import RandomForestRegressor

from utils import load\_data, prepare\_data, evaluate\_model

def run\_regression():

print("Loading data...")

df = load\_data()

X\_train, X\_test, y\_train, y\_test = prepare\_data(df)

models = {

"Linear Regression": LinearRegression(),

"Decision Tree": DecisionTreeRegressor(random\_state=42),

"Random Forest": RandomForestRegressor(n\_estimators=100, random\_state=42)

}

results = {}

for name, model in models.items():

print(f"Training {name}...")

model.fit(X\_train, y\_train)

metrics = evaluate\_model(model, X\_test, y\_test)

results[name] = metrics

print(f"{name} -> MSE: {metrics['MSE']:.2f}, R²: {metrics['R2']:.2f}")

result\_df = pd.DataFrame(results).T

result\_df.to\_csv("regression\_results.csv")

print("\nModel comparison saved to regression\_results.csv")

if \_\_name\_\_ == "\_\_main\_\_":

run\_regression()

Output:

(HousingRegression) D:\MLOPS\assign1>python regression.py

D:\MLOPS\assign1\utils.py:8: SyntaxWarning: invalid escape sequence '\s'

raw\_df = pd.read\_csv(data\_url, sep="\s+", skiprows=22, header=None)

Loading data...

Training Linear Regression...

Linear Regression -> MSE: 24.29, R²: 0.67

Training Decision Tree...

Decision Tree -> MSE: 10.42, R²: 0.86

Training Random Forest...

Random Forest -> MSE: 7.90, R²: 0.89

Model comparison saved to regression\_results.csv

(HousingRegression) D:\MLOPS\assign1>type regression\_results.csv

,MSE,R2

Linear Regression,24.2911194749737,0.6687594935356296

Decision Tree,10.416078431372549,0.8579634380978161

Random Forest,7.901513892156864,0.8922527442109116

13. Create a new branch hyper\_branch and add new regression.py file containing the hyper parameter tuning code. Commit the file and push the new branch to repo.

(HousingRegression) D:\MLOPS\assign1>git checkout -b hyper\_branch

Switched to a new branch 'hyper\_branch'

(HousingRegression) D:\MLOPS\assign1>git add regression.py

(HousingRegression) D:\MLOPS\assign1>git commit -m "Added hyperparameter tuning for 3 models in the regression.py file"

[hyper\_branch af441d3] Added hyperparameter tuning for 3 models in the regression.py file

1 file changed, 25 insertions(+), 12 deletions(-)

(HousingRegression) D:\MLOPS\assign1>git push origin hyper\_branch

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

Enumerating objects: 5, done.

Counting objects: 100% (5/5), done.

Delta compression using up to 4 threads

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

Writing objects: 100% (3/3), 999 bytes | 333.00 KiB/s, done.

Total 3 (delta 1), reused 0 (delta 0), pack-reused 0 (from 0)

remote: Resolving deltas: 100% (1/1), completed with 1 local object.

remote:

remote: Create a pull request for 'hyper\_branch' on GitHub by visiting:

remote: https://github.com/geek1060/HousingRegression/pull/new/hyper\_branch

remote:

To github.com:geek1060/HousingRegression.git

\* [new branch] hyper\_branch -> hyper\_branch

14. Run the new regression.py.

Code:

import numpy as np

import pandas as pd

from sklearn.linear\_model import LinearRegression

from sklearn.tree import DecisionTreeRegressor

from sklearn.ensemble import RandomForestRegressor

from utils import load\_data, prepare\_data, evaluate\_model

from sklearn.model\_selection import GridSearchCV

def tune\_hyperparameters(model, param\_grid, X\_train, y\_train):

grid\_search = GridSearchCV(model, param\_grid, cv=3, scoring='neg\_mean\_squared\_error')

grid\_search.fit(X\_train, y\_train)

return grid\_search.best\_estimator\_, grid\_search.best\_params\_

def run\_regression\_with\_tuning():

print("Loading data...")

df = load\_data()

X\_train, X\_test, y\_train, y\_test = prepare\_data(df)

models\_and\_params = {

"Linear Regression": (LinearRegression(), {}),

"Decision Tree": (

DecisionTreeRegressor(random\_state=42),

{'max\_depth': [None, 5, 10], 'min\_samples\_split': [2, 5]}

),

"Random Forest": (

RandomForestRegressor(random\_state=42),

{'n\_estimators': [50, 100], 'max\_depth': [None, 10], 'min\_samples\_split': [2, 5]}

)

}

results = {}

for name, (model, params) in models\_and\_params.items():

print(f"Tuning {name}...")

best\_model, best\_params = tune\_hyperparameters(model, params, X\_train, y\_train)

print(f"Best Params for {name}: {best\_params}")

metrics = evaluate\_model(best\_model, X\_test, y\_test)

results[name] = metrics

print(f"{name} -> MSE: {metrics['MSE']:.2f}, R²: {metrics['R2']:.2f}")

result\_df = pd.DataFrame(results).T

result\_df.to\_csv("hyperparameter\_results.csv")

print("\nHyperparameter tuning results saved to hyperparameter\_results.csv")

if \_\_name\_\_ == "\_\_main\_\_":

run\_regression\_with\_tuning()

(HousingRegression) D:\MLOPS\assign1>git add regression.py

Output:

(HousingRegression) D:\MLOPS\assign1>python regression.py

Loading data...

Tuning Linear Regression...

Best Params for Linear Regression: {}

Linear Regression -> MSE: 24.29, R²: 0.67

Tuning Decision Tree...

Best Params for Decision Tree: {'max\_depth': 5, 'min\_samples\_split': 2}

Decision Tree -> MSE: 8.55, R²: 0.88

Tuning Random Forest...

Best Params for Random Forest: {'max\_depth': 10, 'min\_samples\_split': 2, 'n\_estimators': 50}

Random Forest -> MSE: 7.42, R²: 0.90

Hyperparameter tuning results saved to hyperparameter\_results.csv

(HousingRegression) D:\MLOPS\assign1>type hyperparameter\_results.csv

,MSE,R2

Linear Regression,24.2911194749737,0.6687594935356296

Decision Tree,8.553906584646844,0.8833565347917995

Random Forest,7.4225563697558625,0.8987839430903407

15. Compare the results of both the models.

(HousingRegression) D:\MLOPS\assign1>type regression\_results.csv

,MSE,R2

Linear Regression,24.2911194749737,0.6687594935356296

Decision Tree,10.416078431372549,0.8579634380978161

Random Forest,7.901513892156864,0.8922527442109116

(HousingRegression) D:\MLOPS\assign1>type hyperparameter\_results.csv

,MSE,R2

Linear Regression,24.2911194749737,0.6687594935356296

Decision Tree,8.553906584646844,0.8833565347917995

Random Forest,7.4225563697558625,0.8987839430903407

15. The log generated is as follows

(HousingRegression) D:\MLOPS\assign1>git log --oneline

42ec32a (HEAD -> hyper\_branch) Add hyperparameter tuning results CSV

af441d3 (origin/hyper\_branch) Added hyperparameter tuning for 3 models in the regression.py file

57b65ac (origin/main, main) Add regression results

2e86d63 (origin/reg\_branch, reg\_branch) Add GitHub Actions CI workflow

8d1ffe5 Add regression code, utils, requirements

aadb6e4 Initial commit with README

**APPENDIX:**

**Output Snippet:**

Microsoft Windows [Version 10.0.19045.5965]

(c) Microsoft Corporation. All rights reserved.

C:\Users\Shreyas>wsl --version

WSL version: 2.3.26.0

Kernel version: 5.15.167.4-1

WSLg version: 1.0.65

MSRDC version: 1.2.5620

Direct3D version: 1.611.1-81528511

DXCore version: 10.0.26100.1-240331-1435.ge-release

Windows version: 10.0.19045.5965

C:\Users\Shreyas>python --version

Python 3.13.5

C:\Users\Shreyas>conda --version

conda 25.5.1

C:\Users\Shreyas>git --version

git version 2.50.0.windows.1

C:\Users\Shreyas>dir /a C:\Users\Shreyas\.ssh

Volume in drive C is OS\_SSD

Volume Serial Number is E6DB-D98C

Directory of C:\Users\Shreyas\.ssh

30-06-2025 21:33 <DIR> .

30-06-2025 21:33 <DIR> ..

30-06-2025 21:20 464 id\_ed25519

30-06-2025 21:20 103 id\_ed25519.pub

30-06-2025 21:33 831 known\_hosts

30-06-2025 21:32 93 known\_hosts.old

4 File(s) 1,491 bytes

2 Dir(s) 58,287,251,456 bytes free

C:\Users\Shreyas>ssh -T git@github.com

Enter passphrase for key 'C:\Users\Shreyas/.ssh/id\_ed25519':

Hi geek1060! You've successfully authenticated, but GitHub does not provide shell access.

C:\Users\Shreyas>conda create -n HousingRegression python=3.13.5 -y

Channels:

- defaults

Platform: win-64

Collecting package metadata (repodata.json): done

Solving environment: done

## Package Plan ##

environment location: C:\Users\Shreyas\miniconda3\envs\HousingRegression

added / updated specs:

- python=3.13.5

The following NEW packages will be INSTALLED:

bzip2 pkgs/main/win-64::bzip2-1.0.8-h2bbff1b\_6

ca-certificates pkgs/main/win-64::ca-certificates-2025.2.25-haa95532\_0

expat pkgs/main/win-64::expat-2.7.1-h8ddb27b\_0

libffi pkgs/main/win-64::libffi-3.4.4-hd77b12b\_1

libmpdec pkgs/main/win-64::libmpdec-4.0.0-h827c3e9\_0

openssl pkgs/main/win-64::openssl-3.0.16-h3f729d1\_0

pip pkgs/main/noarch::pip-25.1-pyhc872135\_2

python pkgs/main/win-64::python-3.13.5-h286a616\_100\_cp313

python\_abi pkgs/main/win-64::python\_abi-3.13-0\_cp313

setuptools pkgs/main/win-64::setuptools-78.1.1-py313haa95532\_0

sqlite pkgs/main/win-64::sqlite-3.45.3-h2bbff1b\_0

tk pkgs/main/win-64::tk-8.6.14-h5e9d12e\_1

tzdata pkgs/main/noarch::tzdata-2025b-h04d1e81\_0

vc pkgs/main/win-64::vc-14.42-haa95532\_5

vs2015\_runtime pkgs/main/win-64::vs2015\_runtime-14.42.34433-hbfb602d\_5

wheel pkgs/main/win-64::wheel-0.45.1-py313haa95532\_0

xz pkgs/main/win-64::xz-5.6.4-h4754444\_1

zlib pkgs/main/win-64::zlib-1.2.13-h8cc25b3\_1

Downloading and Extracting Packages:

Preparing transaction: done

Verifying transaction: done

Executing transaction: done

#

# To activate this environment, use

#

# $ conda activate HousingRegression

#

# To deactivate an active environment, use

#

# $ conda deactivate

C:\Users\Shreyas>conda activate HousingRegression

(HousingRegression) C:\Users\Shreyas>cd /d D:\MLOPS\assign1

(HousingRegression) D:\MLOPS\assign1>echo "# HousingRegression" > README.md

(HousingRegression) D:\MLOPS\assign1>git init

Initialized empty Git repository in D:/MLOPS/assign1/.git/

(HousingRegression) D:\MLOPS\assign1>git add README.md

(HousingRegression) D:\MLOPS\assign1>git commit -m "Initial commit with README"

[master (root-commit) aadb6e4] Initial commit with README

1 file changed, 1 insertion(+)

create mode 100644 README.md

(HousingRegression) D:\MLOPS\assign1>git remote add origin git@github.com:geek1060/HousingRegression.git

(HousingRegression) D:\MLOPS\assign1>git branch -M main

(HousingRegression) D:\MLOPS\assign1>git push -u origin main

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

Enumerating objects: 3, done.

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

Writing objects: 100% (3/3), 249 bytes | 124.00 KiB/s, done.

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

To github.com:geek1060/HousingRegression.git

\* [new branch] main -> main

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

(HousingRegression) D:\MLOPS\assign1>echo numpy >> requirements.txt

(HousingRegression) D:\MLOPS\assign1>echo pandas >> requirements.txt

(HousingRegression) D:\MLOPS\assign1>echo scikit-learn >> requirements.txt

(HousingRegression) D:\MLOPS\assign1>echo statsmodels >> requirements.txt

(HousingRegression) D:\MLOPS\assign1>echo matplotlib >> requirements.txt

(HousingRegression) D:\MLOPS\assign1>git checkout -b reg\_branch

Switched to a new branch 'reg\_branch'

(HousingRegression) D:\MLOPS\assign1>git add .

(HousingRegression) D:\MLOPS\assign1>git commit -m "Add regression code, utils, requirements"

[reg\_branch 8d1ffe5] Add regression code, utils, requirements

3 files changed, 66 insertions(+)

create mode 100644 regression.py

create mode 100644 requirements.txt

create mode 100644 utils.py

(HousingRegression) D:\MLOPS\assign1>mkdir .github

(HousingRegression) D:\MLOPS\assign1>mkdir .github\workflows

(HousingRegression) D:\MLOPS\assign1>notepad .github\workflows\ci.yml

(HousingRegression) D:\MLOPS\assign1>git add .github/workflows/ci.yml

(HousingRegression) D:\MLOPS\assign1>git commit -m "Add GitHub Actions CI workflow"

[reg\_branch 2e86d63] Add GitHub Actions CI workflow

1 file changed, 29 insertions(+)

create mode 100644 .github/workflows/ci.yml

(HousingRegression) D:\MLOPS\assign1>git push origin reg\_branch

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

Enumerating objects: 11, done.

Counting objects: 100% (11/11), done.

Delta compression using up to 4 threads

Compressing objects: 100% (8/8), done.

Writing objects: 100% (10/10), 2.03 KiB | 346.00 KiB/s, done.

Total 10 (delta 1), reused 0 (delta 0), pack-reused 0 (from 0)

remote: Resolving deltas: 100% (1/1), done.

remote:

remote: Create a pull request for 'reg\_branch' on GitHub by visiting:

remote: https://github.com/geek1060/HousingRegression/pull/new/reg\_branch

remote:

To github.com:geek1060/HousingRegression.git

\* [new branch] reg\_branch -> reg\_branch

(HousingRegression) D:\MLOPS\assign1>git checkout main

Switched to branch 'main'

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

(HousingRegression) D:\MLOPS\assign1>git merge reg\_branch

Updating aadb6e4..2e86d63

Fast-forward

.github/workflows/ci.yml | 29 +++++++++++++++++++++++++++++

regression.py | 33 +++++++++++++++++++++++++++++++++

requirements.txt | 5 +++++

utils.py | 28 ++++++++++++++++++++++++++++

4 files changed, 95 insertions(+)

create mode 100644 .github/workflows/ci.yml

create mode 100644 regression.py

create mode 100644 requirements.txt

create mode 100644 utils.py

(HousingRegression) D:\MLOPS\assign1>git push

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

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

To github.com:geek1060/HousingRegression.git

aadb6e4..2e86d63 main -> main

(HousingRegression) D:\MLOPS\assign1>pip install numpy pandas scikit-learn matplotlib statsmodels

Collecting numpy

Using cached numpy-2.3.1-cp313-cp313-win\_amd64.whl.metadata (60 kB)

Collecting pandas

Using cached pandas-2.3.0-cp313-cp313-win\_amd64.whl.metadata (19 kB)

Collecting scikit-learn

Using cached scikit\_learn-1.7.0-cp313-cp313-win\_amd64.whl.metadata (14 kB)

Collecting matplotlib

Using cached matplotlib-3.10.3-cp313-cp313-win\_amd64.whl.metadata (11 kB)

Collecting statsmodels

Downloading statsmodels-0.14.4-cp313-cp313-win\_amd64.whl.metadata (9.5 kB)

Collecting python-dateutil>=2.8.2 (from pandas)

Using cached python\_dateutil-2.9.0.post0-py2.py3-none-any.whl.metadata (8.4 kB)

Collecting pytz>=2020.1 (from pandas)

Using cached pytz-2025.2-py2.py3-none-any.whl.metadata (22 kB)

Collecting tzdata>=2022.7 (from pandas)

Using cached tzdata-2025.2-py2.py3-none-any.whl.metadata (1.4 kB)

Collecting scipy>=1.8.0 (from scikit-learn)

Using cached scipy-1.16.0-cp313-cp313-win\_amd64.whl.metadata (60 kB)

Collecting joblib>=1.2.0 (from scikit-learn)

Using cached joblib-1.5.1-py3-none-any.whl.metadata (5.6 kB)

Collecting threadpoolctl>=3.1.0 (from scikit-learn)

Using cached threadpoolctl-3.6.0-py3-none-any.whl.metadata (13 kB)

Collecting contourpy>=1.0.1 (from matplotlib)

Using cached contourpy-1.3.2-cp313-cp313-win\_amd64.whl.metadata (5.5 kB)

Collecting cycler>=0.10 (from matplotlib)

Using cached cycler-0.12.1-py3-none-any.whl.metadata (3.8 kB)

Collecting fonttools>=4.22.0 (from matplotlib)

Using cached fonttools-4.58.4-cp313-cp313-win\_amd64.whl.metadata (108 kB)

Collecting kiwisolver>=1.3.1 (from matplotlib)

Using cached kiwisolver-1.4.8-cp313-cp313-win\_amd64.whl.metadata (6.3 kB)

Collecting packaging>=20.0 (from matplotlib)

Downloading packaging-25.0-py3-none-any.whl.metadata (3.3 kB)

Collecting pillow>=8 (from matplotlib)

Downloading pillow-11.3.0-cp313-cp313-win\_amd64.whl.metadata (9.2 kB)

Collecting pyparsing>=2.3.1 (from matplotlib)

Using cached pyparsing-3.2.3-py3-none-any.whl.metadata (5.0 kB)

Collecting patsy>=0.5.6 (from statsmodels)

Downloading patsy-1.0.1-py2.py3-none-any.whl.metadata (3.3 kB)

Collecting six>=1.5 (from python-dateutil>=2.8.2->pandas)

Using cached six-1.17.0-py2.py3-none-any.whl.metadata (1.7 kB)

Using cached numpy-2.3.1-cp313-cp313-win\_amd64.whl (12.7 MB)

Using cached pandas-2.3.0-cp313-cp313-win\_amd64.whl (11.0 MB)

Using cached scikit\_learn-1.7.0-cp313-cp313-win\_amd64.whl (10.7 MB)

Using cached matplotlib-3.10.3-cp313-cp313-win\_amd64.whl (8.1 MB)

Downloading statsmodels-0.14.4-cp313-cp313-win\_amd64.whl (9.8 MB)

---------------------------------------- 9.8/9.8 MB 4.0 MB/s eta 0:00:00

Using cached contourpy-1.3.2-cp313-cp313-win\_amd64.whl (223 kB)

Using cached cycler-0.12.1-py3-none-any.whl (8.3 kB)

Using cached fonttools-4.58.4-cp313-cp313-win\_amd64.whl (2.2 MB)

Using cached joblib-1.5.1-py3-none-any.whl (307 kB)

Using cached kiwisolver-1.4.8-cp313-cp313-win\_amd64.whl (71 kB)

Downloading packaging-25.0-py3-none-any.whl (66 kB)

Downloading patsy-1.0.1-py2.py3-none-any.whl (232 kB)

Downloading pillow-11.3.0-cp313-cp313-win\_amd64.whl (7.0 MB)

---------------------------------------- 7.0/7.0 MB 5.7 MB/s eta 0:00:00

Using cached pyparsing-3.2.3-py3-none-any.whl (111 kB)

Using cached python\_dateutil-2.9.0.post0-py2.py3-none-any.whl (229 kB)

Using cached pytz-2025.2-py2.py3-none-any.whl (509 kB)

Using cached scipy-1.16.0-cp313-cp313-win\_amd64.whl (38.4 MB)

Using cached six-1.17.0-py2.py3-none-any.whl (11 kB)

Using cached threadpoolctl-3.6.0-py3-none-any.whl (18 kB)

Using cached tzdata-2025.2-py2.py3-none-any.whl (347 kB)

Installing collected packages: pytz, tzdata, threadpoolctl, six, pyparsing, pillow, packaging, numpy, kiwisolver, joblib, fonttools, cycler, scipy, python-dateutil, patsy, contourpy, scikit-learn, pandas, matplotlib, statsmodels

Successfully installed contourpy-1.3.2 cycler-0.12.1 fonttools-4.58.4 joblib-1.5.1 kiwisolver-1.4.8 matplotlib-3.10.3 numpy-2.3.1 packaging-25.0 pandas-2.3.0 patsy-1.0.1 pillow-11.3.0 pyparsing-3.2.3 python-dateutil-2.9.0.post0 pytz-2025.2 scikit-learn-1.7.0 scipy-1.16.0 six-1.17.0 statsmodels-0.14.4 threadpoolctl-3.6.0 tzdata-2025.2

(HousingRegression) D:\MLOPS\assign1>python regression.py

D:\MLOPS\assign1\utils.py:8: SyntaxWarning: invalid escape sequence '\s'

raw\_df = pd.read\_csv(data\_url, sep="\s+", skiprows=22, header=None)

Loading data...

Training Linear Regression...

Linear Regression -> MSE: 24.29, R²: 0.67

Training Decision Tree...

Decision Tree -> MSE: 10.42, R²: 0.86

Training Random Forest...

Random Forest -> MSE: 7.90, R²: 0.89

Model comparison saved to regression\_results.csv

(HousingRegression) D:\MLOPS\assign1>type regression\_results.csv

,MSE,R2

Linear Regression,24.2911194749737,0.6687594935356296

Decision Tree,10.416078431372549,0.8579634380978161

Random Forest,7.901513892156864,0.8922527442109116

(HousingRegression) D:\MLOPS\assign1>git add regression\_results.csv

(HousingRegression) D:\MLOPS\assign1>git commit -m "Add regression results"

[main 57b65ac] Add regression results

1 file changed, 4 insertions(+)

create mode 100644 regression\_results.csv

(HousingRegression) D:\MLOPS\assign1>git push

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

Enumerating objects: 4, done.

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

Delta compression using up to 4 threads

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

Writing objects: 100% (3/3), 409 bytes | 204.00 KiB/s, done.

Total 3 (delta 1), reused 0 (delta 0), pack-reused 0 (from 0)

remote: Resolving deltas: 100% (1/1), completed with 1 local object.

To github.com:geek1060/HousingRegression.git

2e86d63..57b65ac main -> main

(HousingRegression) D:\MLOPS\assign1>git checkout -b hyper\_branch

Switched to a new branch 'hyper\_branch'

(HousingRegression) D:\MLOPS\assign1>git add regression.py

(HousingRegression) D:\MLOPS\assign1>git commit -m "Added hyperparameter tuning for 3 models in the regression.py file"

[hyper\_branch af441d3] Added hyperparameter tuning for 3 models in the regression.py file

1 file changed, 25 insertions(+), 12 deletions(-)

(HousingRegression) D:\MLOPS\assign1>git push origin hyper\_branch

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

Enumerating objects: 5, done.

Counting objects: 100% (5/5), done.

Delta compression using up to 4 threads

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

Writing objects: 100% (3/3), 999 bytes | 333.00 KiB/s, done.

Total 3 (delta 1), reused 0 (delta 0), pack-reused 0 (from 0)

remote: Resolving deltas: 100% (1/1), completed with 1 local object.

remote:

remote: Create a pull request for 'hyper\_branch' on GitHub by visiting:

remote: https://github.com/geek1060/HousingRegression/pull/new/hyper\_branch

remote:

To github.com:geek1060/HousingRegression.git

\* [new branch] hyper\_branch -> hyper\_branch

(HousingRegression) D:\MLOPS\assign1>python regression.py

Loading data...

Tuning Linear Regression...

Best Params for Linear Regression: {}

Linear Regression -> MSE: 24.29, R²: 0.67

Tuning Decision Tree...

Best Params for Decision Tree: {'max\_depth': 5, 'min\_samples\_split': 2}

Decision Tree -> MSE: 8.55, R²: 0.88

Tuning Random Forest...

Best Params for Random Forest: {'max\_depth': 10, 'min\_samples\_split': 2, 'n\_estimators': 50}

Random Forest -> MSE: 7.42, R²: 0.90

Hyperparameter tuning results saved to hyperparameter\_results.csv

(HousingRegression) D:\MLOPS\assign1>type hyperparameter\_results.csv

,MSE,R2

Linear Regression,24.2911194749737,0.6687594935356296

Decision Tree,8.553906584646844,0.8833565347917995

Random Forest,7.4225563697558625,0.8987839430903407

(HousingRegression) D:\MLOPS\assign1>git push origin hyper\_branch

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

Everything up-to-date

(HousingRegression) D:\MLOPS\assign1>git add hyperparameter\_results.csv

(HousingRegression) D:\MLOPS\assign1>git push origin hyper\_branch

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

Everything up-to-date

(HousingRegression) D:\MLOPS\assign1>type regression\_results.csv

,MSE,R2

Linear Regression,24.2911194749737,0.6687594935356296

Decision Tree,10.416078431372549,0.8579634380978161

Random Forest,7.901513892156864,0.8922527442109116

(HousingRegression) D:\MLOPS\assign1>type hyperparameter\_results.csv

,MSE,R2

Linear Regression,24.2911194749737,0.6687594935356296

Decision Tree,8.553906584646844,0.8833565347917995

Random Forest,7.4225563697558625,0.8987839430903407

(HousingRegression) D:\MLOPS\assign1>git push

fatal: The current branch hyper\_branch has no upstream branch.

To push the current branch and set the remote as upstream, use

git push --set-upstream origin hyper\_branch

To have this happen automatically for branches without a tracking

upstream, see 'push.autoSetupRemote' in 'git help config'.

(HousingRegression) D:\MLOPS\assign1>git branch

\* hyper\_branch

main

reg\_branch

(HousingRegression) D:\MLOPS\assign1>git add hyperparameter\_results.csv

(HousingRegression) D:\MLOPS\assign1>git commit -m "Add hyperparameter tuning results CSV"

[hyper\_branch 42ec32a] Add hyperparameter tuning results CSV

1 file changed, 4 insertions(+)

create mode 100644 hyperparameter\_results.csv

(HousingRegression) D:\MLOPS\assign1>git log --oneline

42ec32a (HEAD -> hyper\_branch) Add hyperparameter tuning results CSV

af441d3 (origin/hyper\_branch) Added hyperparameter tuning for 3 models in the regression.py file

57b65ac (origin/main, main) Add regression results

2e86d63 (origin/reg\_branch, reg\_branch) Add GitHub Actions CI workflow

8d1ffe5 Add regression code, utils, requirements

aadb6e4 Initial commit with README

(HousingRegression) D:\MLOPS\assign1>git checkout main

Switched to branch 'main'

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

(HousingRegression) D:\MLOPS\assign1>git merge hyper\_branch

Updating 57b65ac..42ec32a

Fast-forward

hyperparameter\_results.csv | 4 ++++

regression.py | 37 +++++++++++++++++++++++++------------

2 files changed, 29 insertions(+), 12 deletions(-)

create mode 100644 hyperparameter\_results.csv

(HousingRegression) D:\MLOPS\assign1>git push

Enter passphrase for key '/c/Users/Shreyas/.ssh/id\_ed25519':

Enumerating objects: 4, done.

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

Delta compression using up to 4 threads

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

Writing objects: 100% (3/3), 432 bytes | 216.00 KiB/s, done.

Total 3 (delta 1), reused 0 (delta 0), pack-reused 0 (from 0)

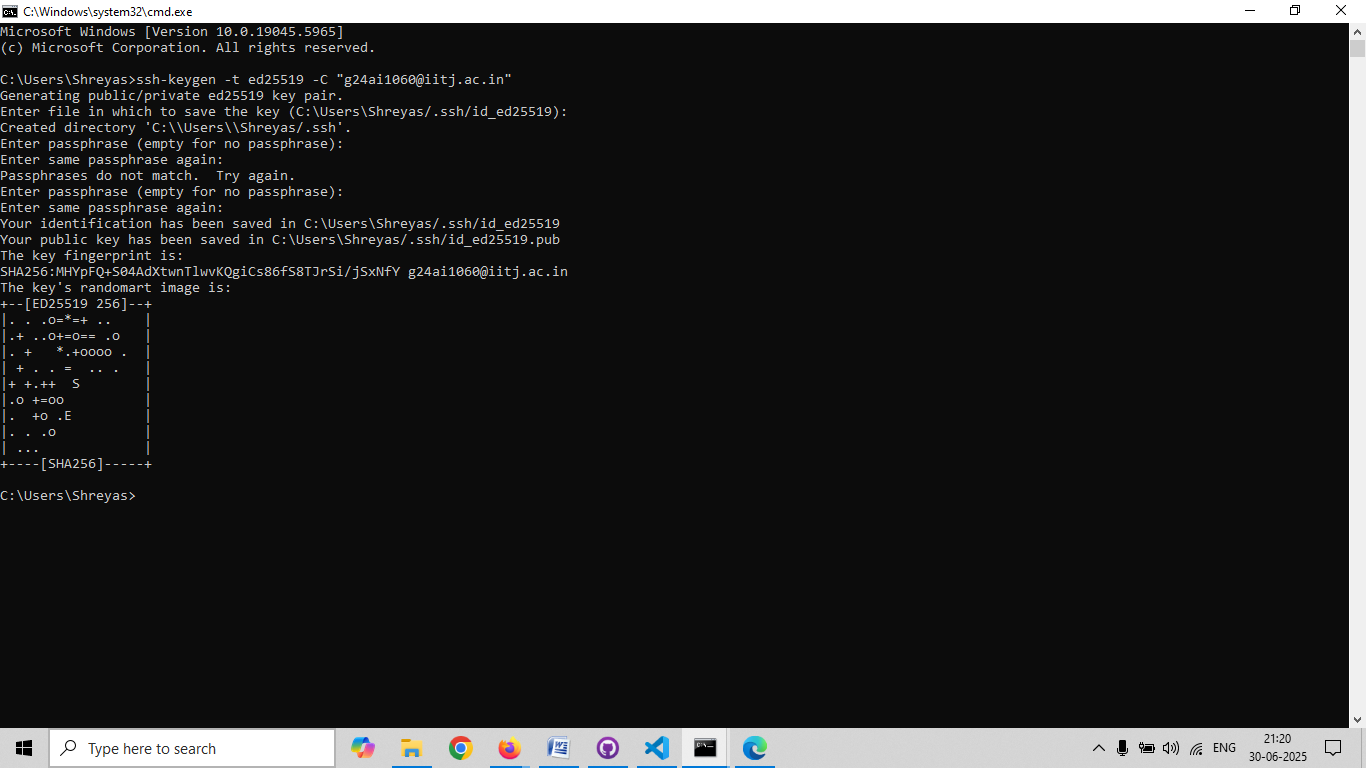
remote: Resolving deltas: 100% (1/1), completed with 1 local object.

To github.com:geek1060/HousingRegression.git

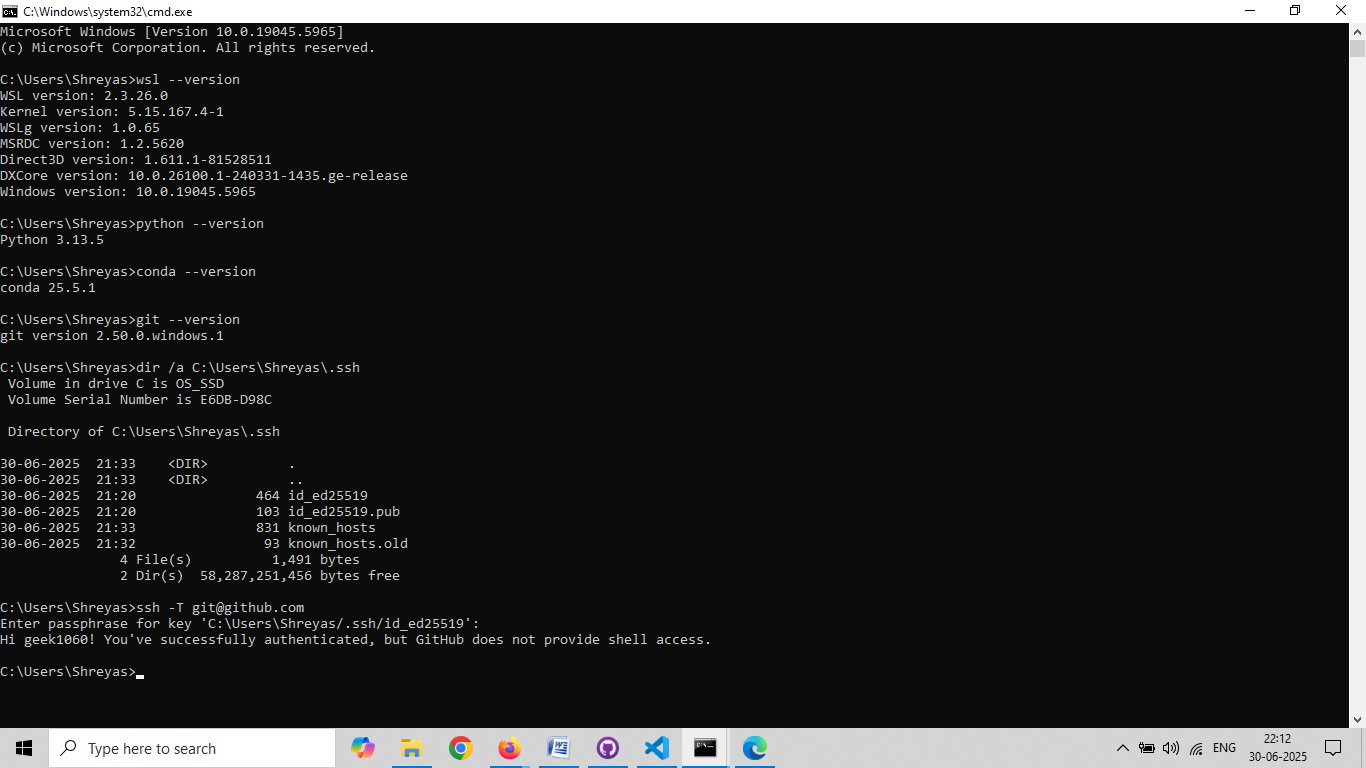
57b65ac..42ec32a main -> main

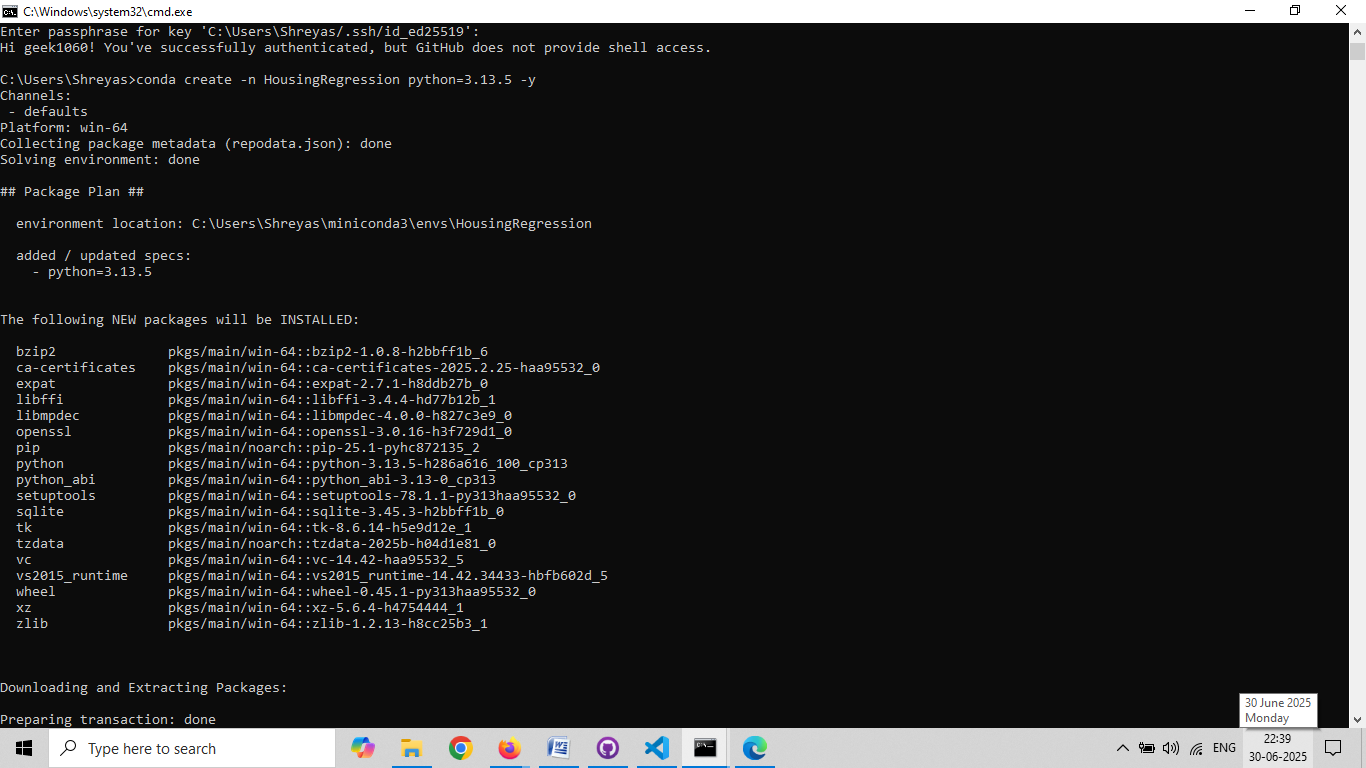
(HousingRegression) D:\MLOPS\assign1>

**Generating and Saving SSH Key for access to GitHub**

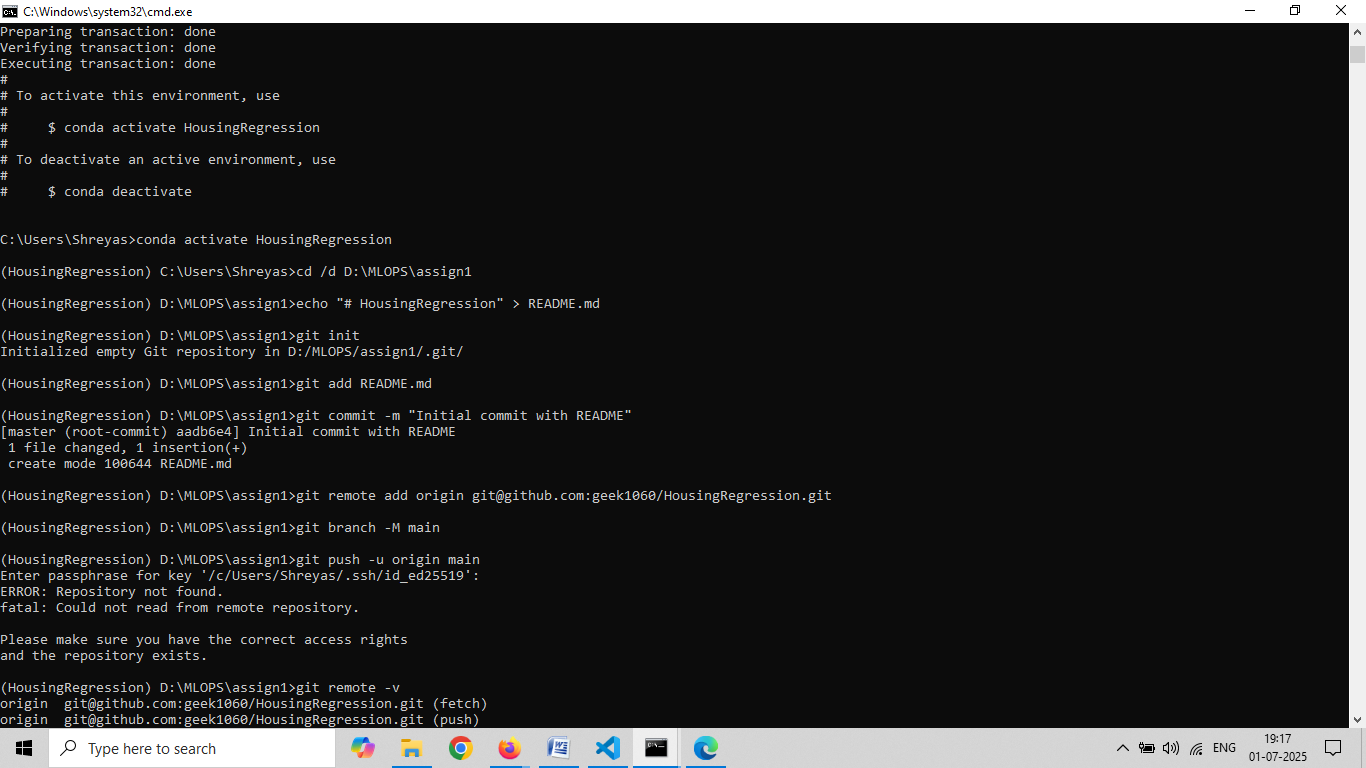


**Screenshot of the Entire Procedure performed on the Cmd Prompt Window.**

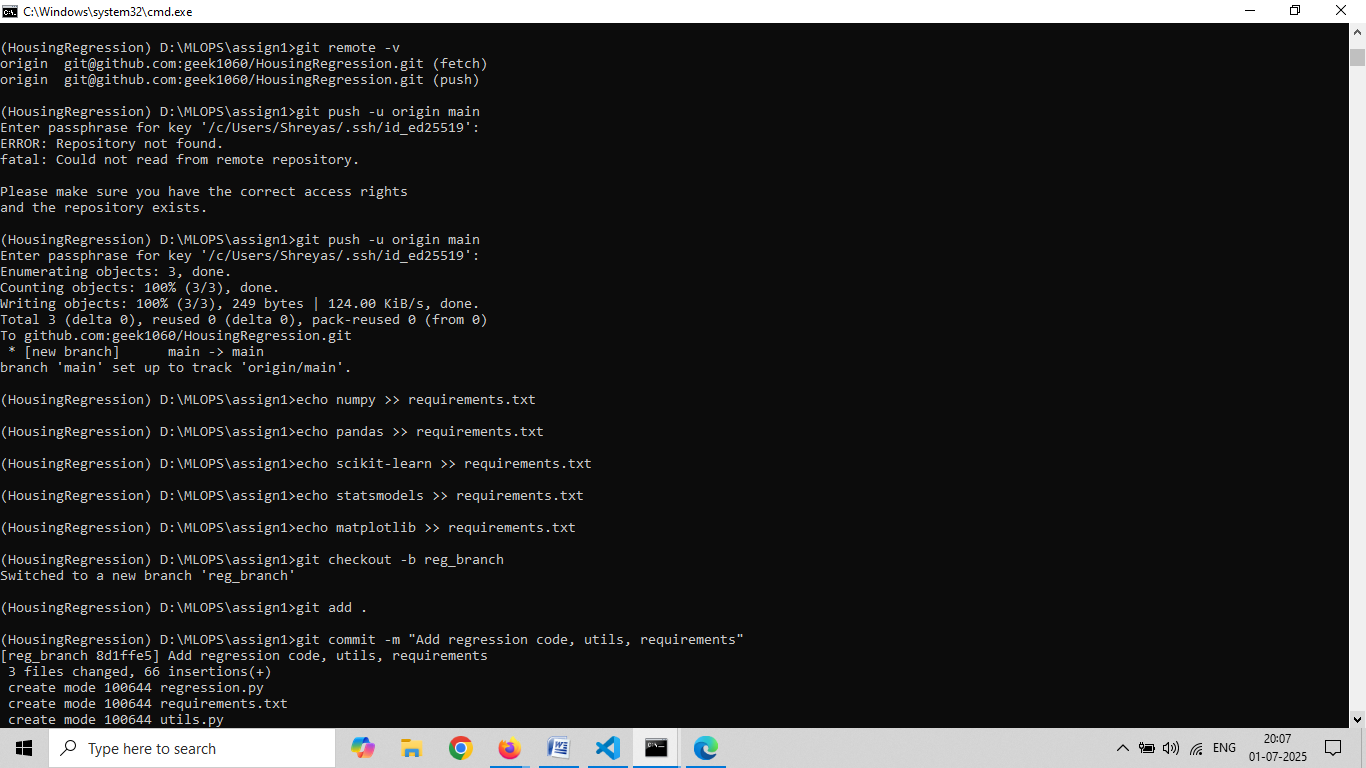




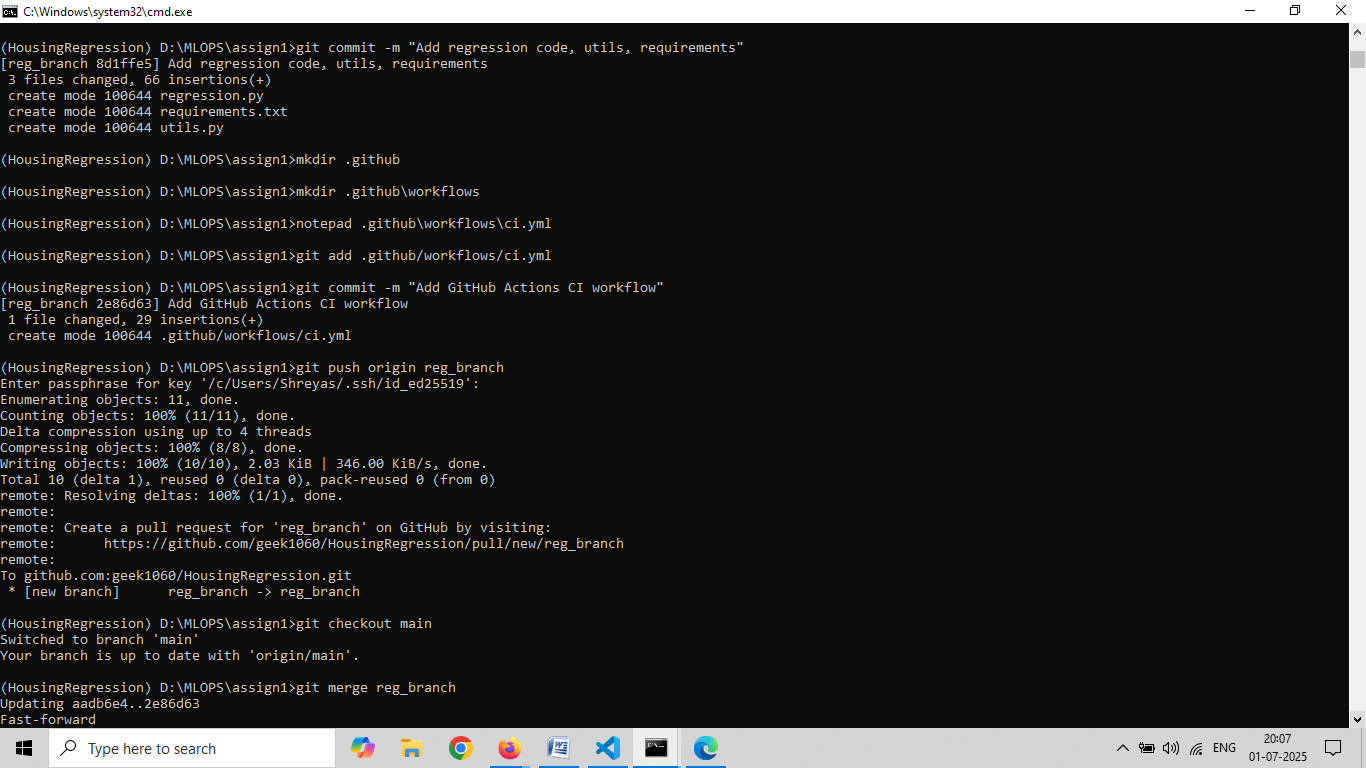
**Packages Installed and selected conda environment HousingRegression.**

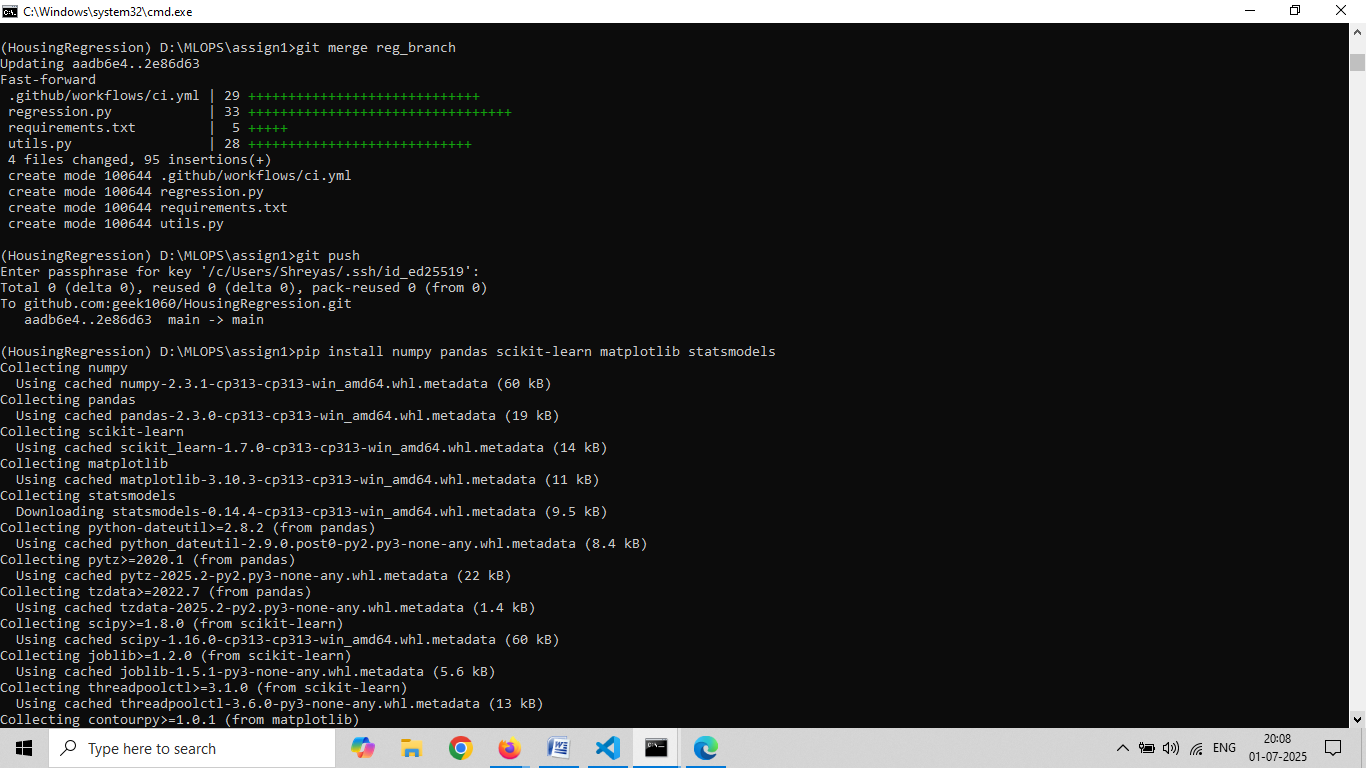
****

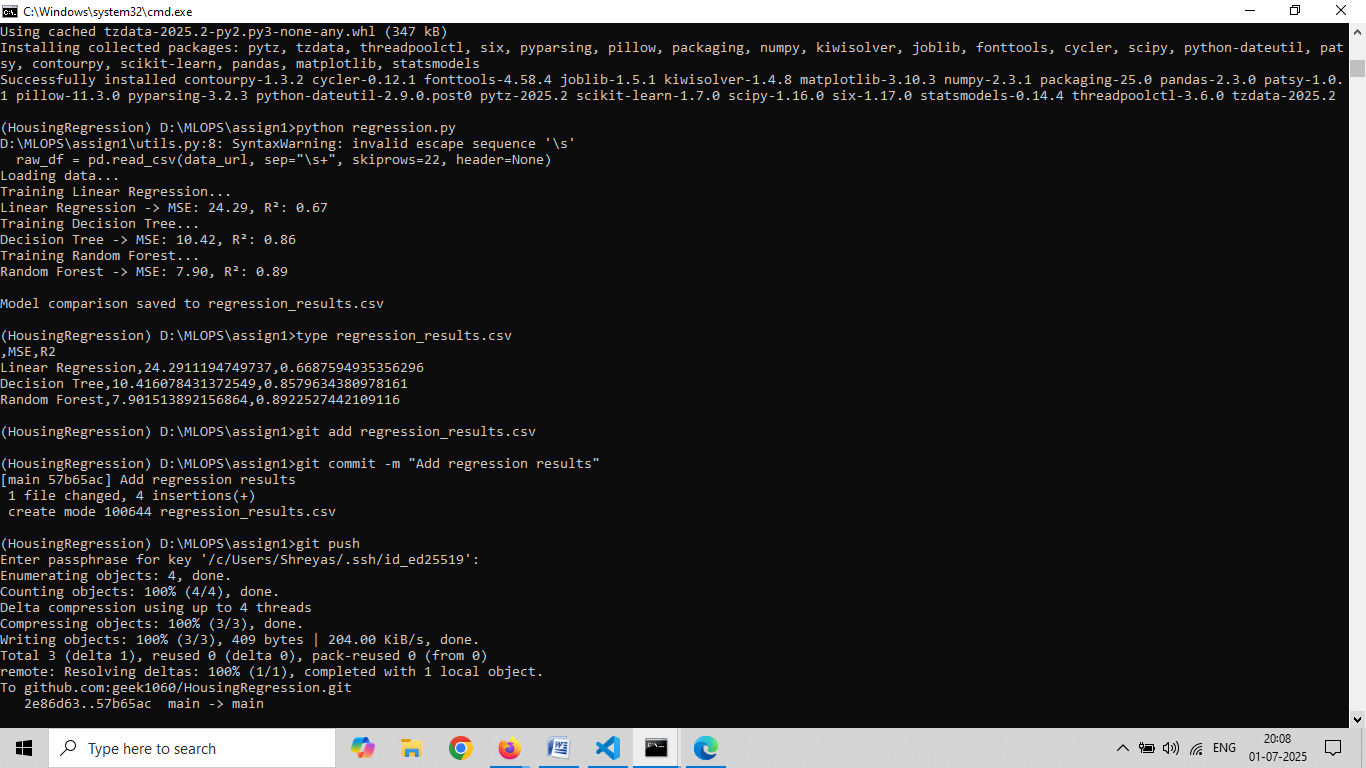
**Created main branch, added the requirements.txt and pushed it to the reg\_branch, further added all files to the reg\_branch.**

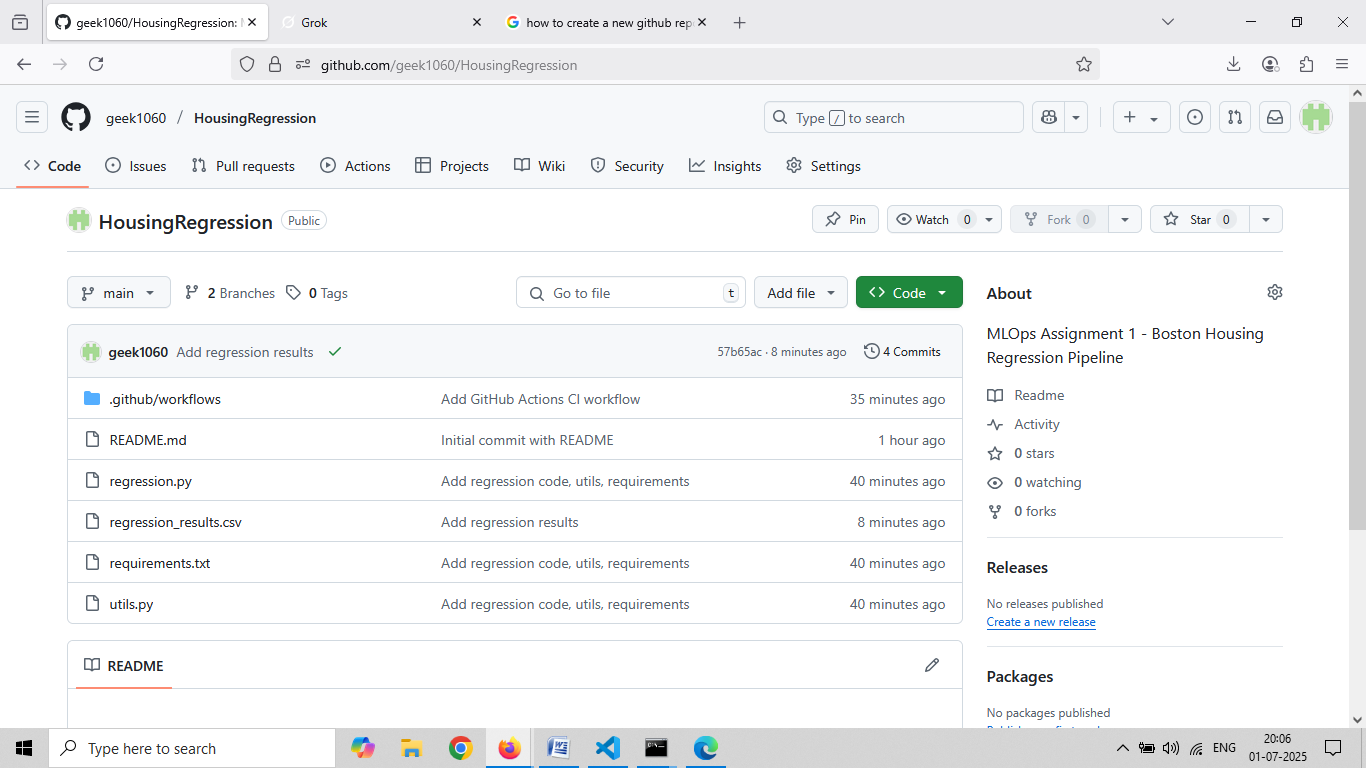
****

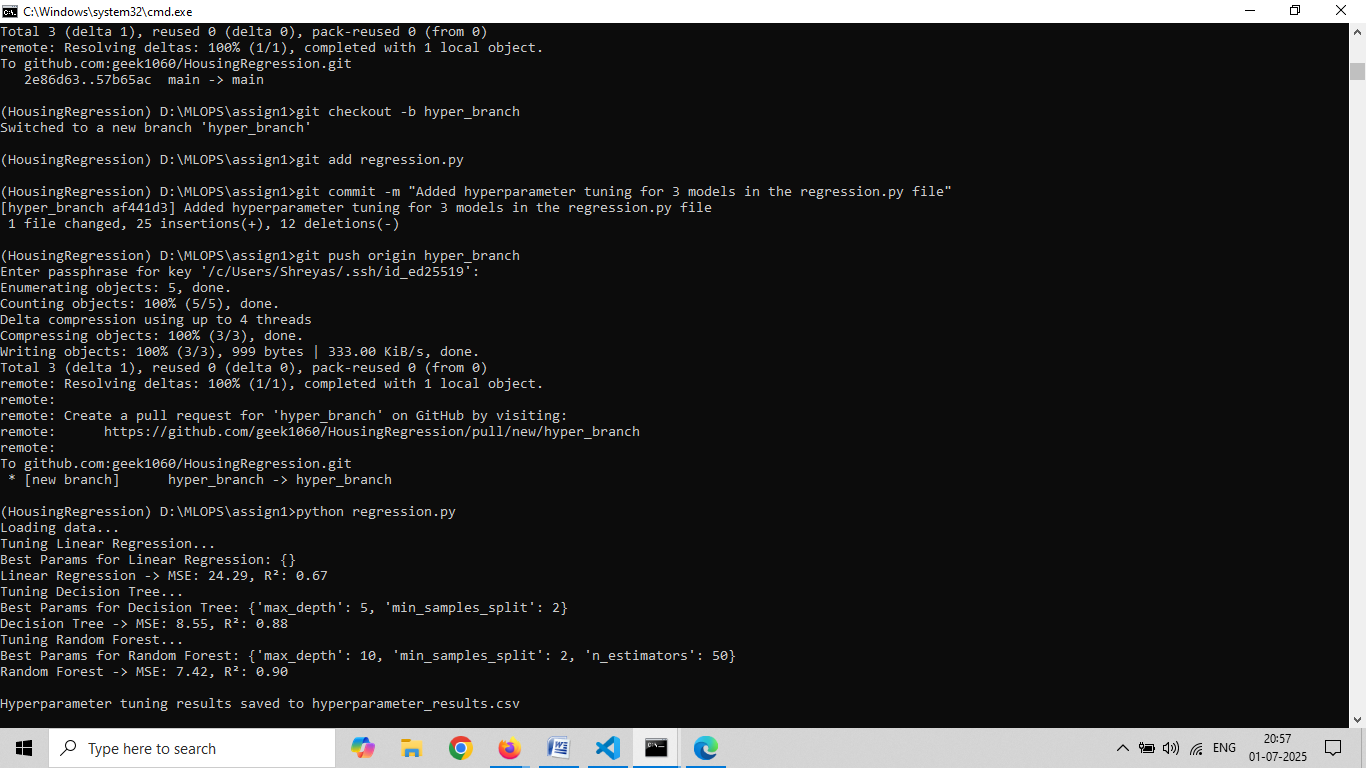
**Added workflow yml file to the reg\_branch and merged it with main branch.**

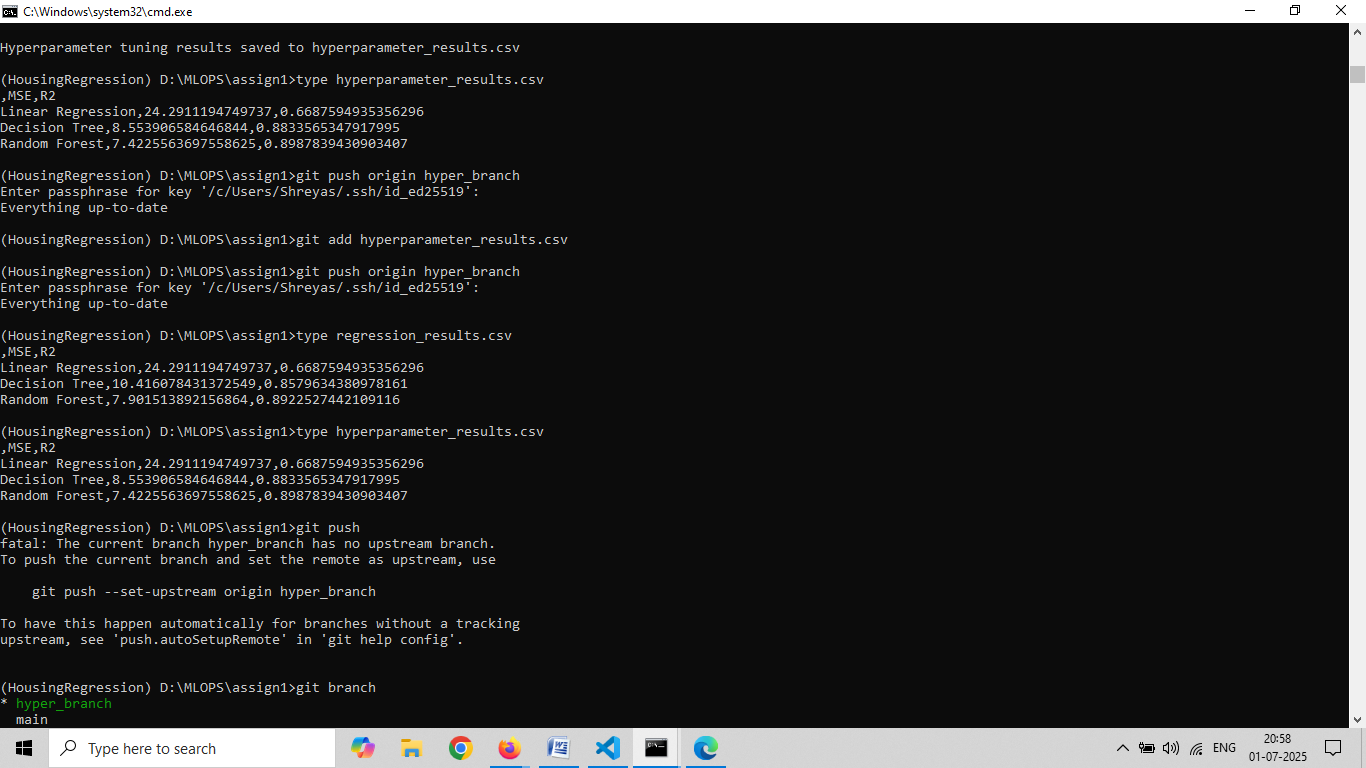
****

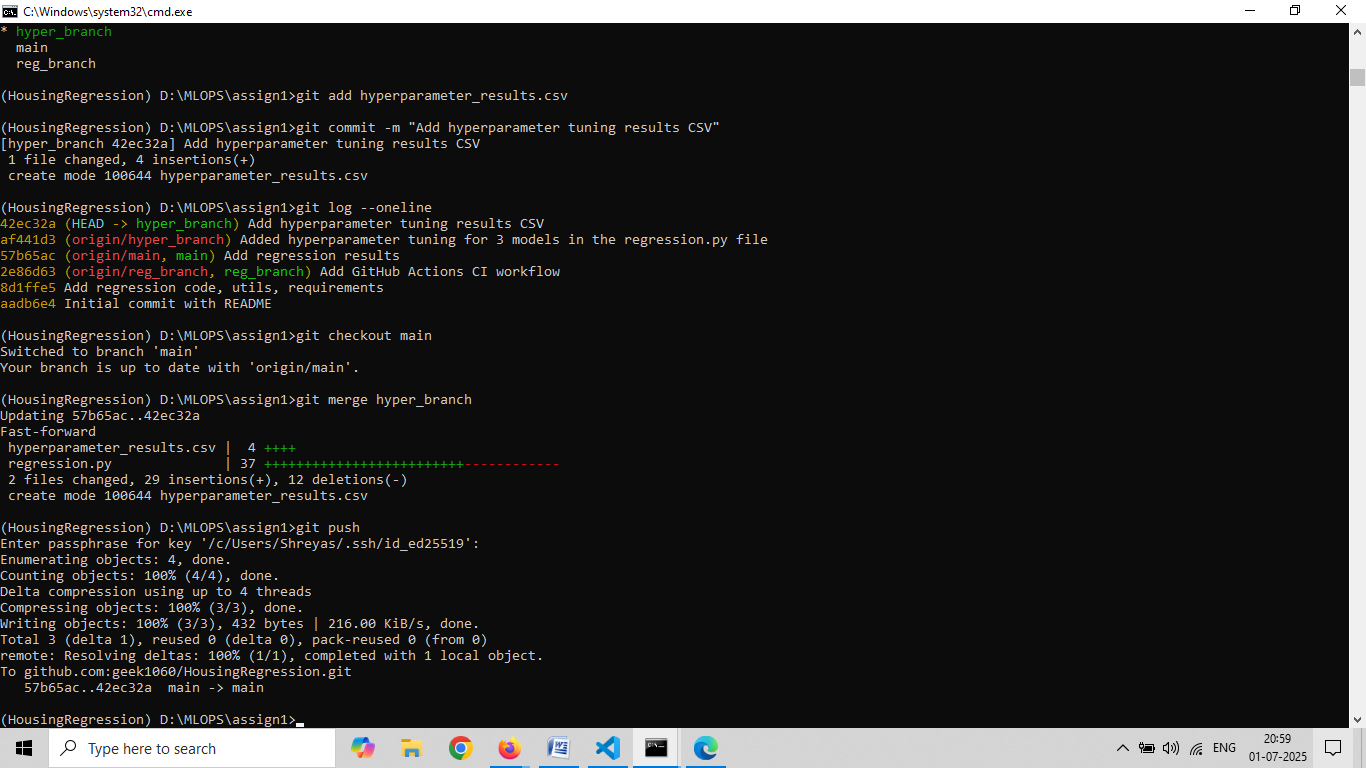
****

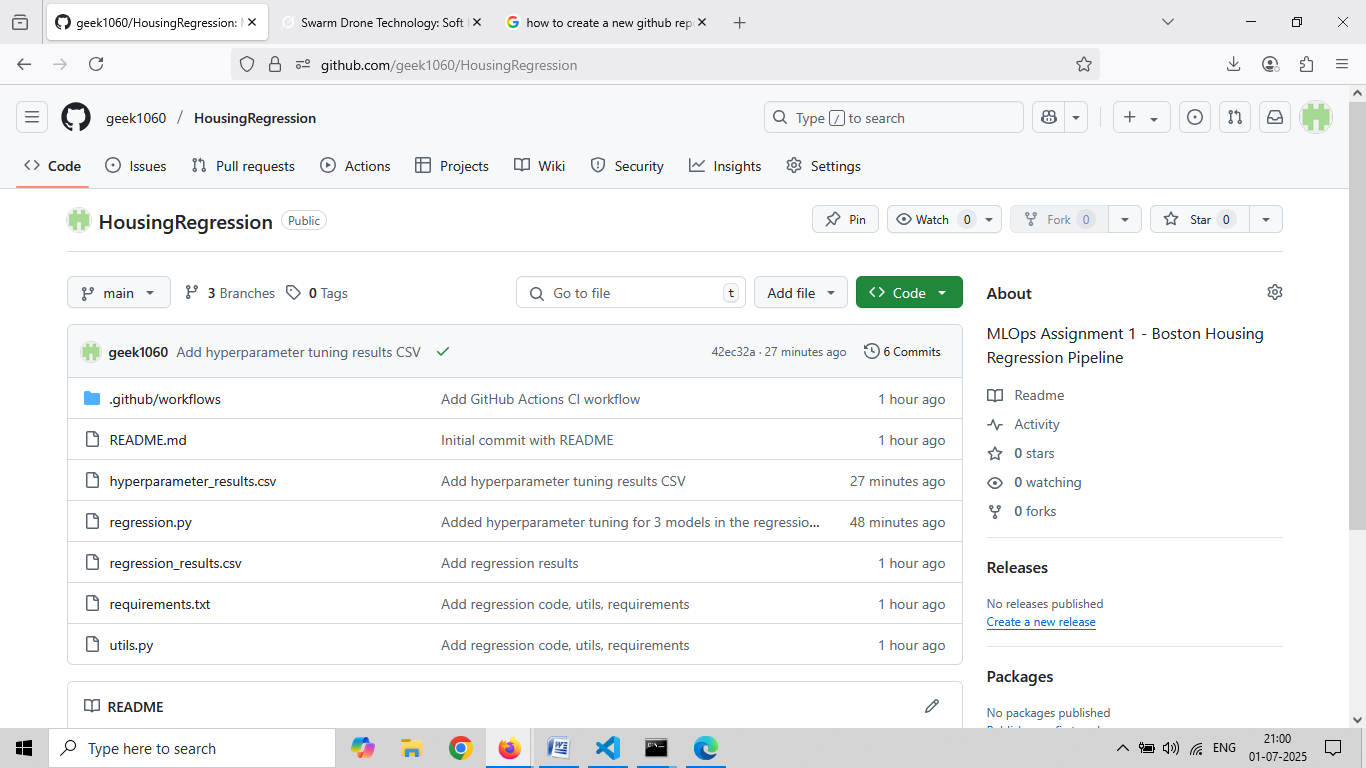
****

****









**The build logs are as follows:**

