Crop Yield Insights: A Regression Analysis Approach

```
In [8]:   import pandas as pd
          import numpy as np
          import seaborn as sns
          import matplotlib.pyplot as plt
          from sklearn.model_selection import train_test_split
          from sklearn.preprocessing import OneHotEncoder
          from sklearn.linear_model import LinearRegression
          from sklearn.metrics import mean_squared_error, r2_score
          import statsmodels.api as sm
```

Exploratory Data Analysis

- Data Understanding

```
In [9]:   !pip install fsspec
          !pip install Jupyter
          !pip install nbconvert PyPDF2
```

```
Requirement already satisfied: fsspec in c:\users\pritam\appdata\local\programs\p
ython\python311\lib\site-packages (2024.9.0)
[notice] A new release of pip is available: 23.2.1 -> 24.2
[notice] To update, run: python.exe -m pip install --upgrade pip

[notice] A new release of pip is available: 23.2.1 -> 24.2
[notice] To update, run: python.exe -m pip install --upgrade pip
```

Requirement already satisfied: Jupyter in c:\users\pritam\appdata\local\programs
\python\python311\lib\site-packages (1.0.0)
Requirement already satisfied: notebook in c:\users\pritam\appdata\local\programs
\python\python311\lib\site-packages (from Jupyter) (7.0.4)
Requirement already satisfied: qtconsole in c:\users\pritam\appdata\local\program
s\python\python311\lib\site-packages (from Jupyter) (5.4.4)
Requirement already satisfied: jupyter-console in c:\users\pritam\appdata\local\p
rograms\python\python311\lib\site-packages (from Jupyter) (6.6.3)
Requirement already satisfied: nbconvert in c:\users\pritam\appdata\local\program
s\python\python311\lib\site-packages (from Jupyter) (7.8.0)
Requirement already satisfied: ipykernel in c:\users\pritam\appdata\local\program
s\python\python311\lib\site-packages (from Jupyter) (6.25.2)
Requirement already satisfied: ipywidgets in c:\users\pritam\appdata\local\progra
ms\python\python311\lib\site-packages (from Jupyter) (8.1.1)
Requirement already satisfied: comm>=0.1.1 in c:\users\pritam\appdata\local\progr
ams\python\python311\lib\site-packages (from ipykernel->Jupyter) (0.1.4)
Requirement already satisfied: debugpy>=1.6.5 in c:\users\pritam\appdata\local\pr
ograms\python\python311\lib\site-packages (from ipykernel->Jupyter) (1.8.0)
Requirement already satisfied: ipython>=7.23.1 in c:\users\pritam\appdata\local\p
rograms\python\python311\lib\site-packages (from ipykernel->Jupyter) (8.15.0)
Requirement already satisfied: jupyter-client>=6.1.12 in c:\users\pritam\appdata
\local\programs\python\python311\lib\site-packages (from ipykernel->Jupyter) (8.
3.1)
Requirement already satisfied: jupyter-core!=5.0.*,>=4.12 in c:\users\pritam\appd
ata\local\programs\python\python311\lib\site-packages (from ipykernel->Jupyter)
(5.3.1)
Requirement already satisfied: matplotlib-inline>=0.1 in c:\users\pritam\appdata
\local\programs\python\python311\lib\site-packages (from ipykernel->Jupyter) (0.
1.6)
Requirement already satisfied: nest-asyncio in c:\users\pritam\appdata\local\prog
rams\python\python311\lib\site-packages (from ipykernel->Jupyter) (1.5.8)
Requirement already satisfied: packaging in c:\users\pritam\appdata\local\program
s\python\python311\lib\site-packages (from ipykernel->Jupyter) (23.1)
Requirement already satisfied: psutil in c:\users\pritam\appdata\local\programs\p
ython\python311\lib\site-packages (from ipykernel->Jupyter) (5.9.5)
Requirement already satisfied: pyzmq>=20 in c:\users\pritam\appdata\local\program
s\python\python311\lib\site-packages (from ipykernel->Jupyter) (25.1.1)
Requirement already satisfied: tornado>=6.1 in c:\users\pritam\appdata\local\prog
rams\python\python311\lib\site-packages (from ipykernel->Jupyter) (6.3.3)
Requirement already satisfied: traitlets>=5.4.0 in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from ipykernel->Jupyter) (5.10.0)
Requirement already satisfied: widgetsnbextension~=4.0.9 in c:\users\pritam\appda
ta\local\programs\python\python311\lib\site-packages (from ipywidgets->Jupyter)
(4.0.9)
Requirement already satisfied: jupyterlab-widgets~=3.0.9 in c:\users\pritam\appda
ta\local\programs\python\python311\lib\site-packages (from ipywidgets->Jupyter)
(3.0.9)
Requirement already satisfied: prompt-toolkit>=3.0.30 in c:\users\pritam\appdata
\local\programs\python\python311\lib\site-packages (from jupyter-console->Jupyte
r) (3.0.39)
Requirement already satisfied: pygments in c:\users\pritam\appdata\local\programs
\python\python311\lib\site-packages (from jupyter-console->Jupyter) (2.16.1)
Requirement already satisfied: beautifulsoup4 in c:\users\pritam\appdata\local\pr
ograms\python\python311\lib\site-packages (from nbconvert->Jupyter) (4.12.2)
Requirement already satisfied: bleach!=5.0.0 in c:\users\pritam\appdata\local\pro
grams\python\python311\lib\site-packages (from nbconvert->Jupyter) (6.0.0)
Requirement already satisfied: defusedxml in c:\users\pritam\appdata\local\progra
ms\python\python311\lib\site-packages (from nbconvert->Jupyter) (0.7.1)
Requirement already satisfied: jinja2>=3.0 in c:\users\pritam\appdata\local\progr
ams\python\python311\lib\site-packages (from nbconvert->Jupyter) (3.1.2)

Requirement already satisfied: jupyterlab-pygments in c:\users\pritam\appdata\loc
al\programs\python\python311\lib\site-packages (from nbconvert->Jupyter) (0.2.2)
Requirement already satisfied: markupsafe>=2.0 in c:\users\pritam\appdata\local\p
rograms\python\python311\lib\site-packages (from nbconvert->Jupyter) (2.1.3)
Requirement already satisfied: mistune<4,>=2.0.3 in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from nbconvert->Jupyter) (3.0.1)
Requirement already satisfied: nbclient>=0.5.0 in c:\users\pritam\appdata\local\p
rograms\python\python311\lib\site-packages (from nbconvert->Jupyter) (0.8.0)
Requirement already satisfied: nbformat>=5.7 in c:\users\pritam\appdata\local\pro
grams\python\python311\lib\site-packages (from nbconvert->Jupyter) (5.9.2)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\pritam\appdata\lo
cal\programs\python\python311\lib\site-packages (from nbconvert->Jupyter) (1.5.0)
Requirement already satisfied: tinycss2 in c:\users\pritam\appdata\local\programs
\python\python311\lib\site-packages (from nbconvert->Jupyter) (1.2.1)
Requirement already satisfied: jupyter-server<3,>=2.4.0 in c:\users\pritam\appdat
a\local\programs\python\python311\lib\site-packages (from notebook->Jupyter) (2.
7.3)
Requirement already satisfied: jupyterlab-server<3,>=2.22.1 in c:\users\pritam\ap
pdata\local\programs\python\python311\lib\site-packages (from notebook->Jupyter)
(2.25.0)
Requirement already satisfied: jupyterlab<5,>=4.0.2 in c:\users\pritam\appdata\lo
cal\programs\python\python311\lib\site-packages (from notebook->Jupyter) (4.0.6)
Requirement already satisfied: notebook-shim<0.3,>=0.2 in c:\users\pritam\appdata
\local\programs\python\python311\lib\site-packages (from notebook->Jupyter) (0.2.
3)
Requirement already satisfied: ipython-genutils in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from qtconsole->Jupyter) (0.2.0)
Requirement already satisfied: qtpy>=2.4.0 in c:\users\pritam\appdata\local\progr
ams\python\python311\lib\site-packages (from qtconsole->Jupyter) (2.4.0)
Requirement already satisfied: six>=1.9.0 in c:\users\pritam\appdata\local\progra
ms\python\python311\lib\site-packages (from bleach!=5.0.0->nbconvert->Jupyter)
(1.16.0)
Requirement already satisfied: webencodings in c:\users\pritam\appdata\local\prog
rams\python\python311\lib\site-packages (from bleach!=5.0.0->nbconvert->Jupyter)
(0.5.1)
Requirement already satisfied: backcall in c:\users\pritam\appdata\local\programs
\python\python311\lib\site-packages (from ipython>=7.23.1->ipykernel->Jupyter)
(0.2.0)
Requirement already satisfied: decorator in c:\users\pritam\appdata\local\program
s\python\python311\lib\site-packages (from ipython>=7.23.1->ipykernel->Jupyter)
(5.1.1)
Requirement already satisfied: jedi>=0.16 in c:\users\pritam\appdata\local\progra
ms\python\python311\lib\site-packages (from ipython>=7.23.1->ipykernel->Jupyter)
(0.19.0)
Requirement already satisfied: pickleshare in c:\users\pritam\appdata\local\progr
ams\python\python311\lib\site-packages (from ipython>=7.23.1->ipykernel->Jupyter)
(0.7.5)
Requirement already satisfied: stack-data in c:\users\pritam\appdata\local\progra
ms\python\python311\lib\site-packages (from ipython>=7.23.1->ipykernel->Jupyter)
(0.6.2)
Requirement already satisfied: colorama in c:\users\pritam\appdata\local\programs
\python\python311\lib\site-packages (from ipython>=7.23.1->ipykernel->Jupyter)
(0.4.6)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\pritam\appdata
\local\programs\python\python311\lib\site-packages (from jupyter-client>=6.1.12->
ipykernel->Jupyter) (2.8.2)
Requirement already satisfied: platformdirs>=2.5 in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from jupyter-core!=5.0.*,>=4.12->ip
ykernel->Jupyter) (3.10.0)
Requirement already satisfied: pywin32>=300 in c:\users\pritam\appdata\local\prog

rams\python\python311\lib\site-packages (from jupyter-core!=5.0.*,>=4.12->ipykern
el->Jupyter) (306)
Requirement already satisfied: anyio>=3.1.0 in c:\users\pritam\appdata\local\prog
rams\python\python311\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook-
>Jupyter) (4.0.0)
Requirement already satisfied: argon2-cffi in c:\users\pritam\appdata\local\progr
ams\python\python311\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook->
Jupyter) (23.1.0)
Requirement already satisfied: jupyter-events>=0.6.0 in c:\users\pritam\appdata\l
ocal\programs\python\python311\lib\site-packages (from jupyter-server<3,>=2.4.0->
notebook->Jupyter) (0.7.0)
Requirement already satisfied: jupyter-server-terminals in c:\users\pritam\appdat
a\local\programs\python\python311\lib\site-packages (from jupyter-server<3,>=2.4.
0->notebook->Jupyter) (0.4.4)
Requirement already satisfied: overrides in c:\users\pritam\appdata\local\program
s\python\python311\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook->Ju
pyter) (7.4.0)
Requirement already satisfied: prometheus-client in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from jupyter-server<3,>=2.4.0->note
book->Jupyter) (0.17.1)
Requirement already satisfied: pywinpty in c:\users\pritam\appdata\local\programs
\python\python311\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook->Jup
yter) (2.0.11)
Requirement already satisfied: send2trash>=1.8.2 in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from jupyter-server<3,>=2.4.0->note
book->Jupyter) (1.8.2)
Requirement already satisfied: terminado>=0.8.3 in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from jupyter-server<3,>=2.4.0->note
book->Jupyter) (0.17.1)
Requirement already satisfied: websocket-client in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from jupyter-server<3,>=2.4.0->note
book->Jupyter) (1.6.3)
Requirement already satisfied: async-lru>=1.0.0 in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from jupyterlab<5,>=4.0.2->notebook
->Jupyter) (2.0.4)
Requirement already satisfied: jupyter-lsp>=2.0.0 in c:\users\pritam\appdata\loca
l\programs\python\python311\lib\site-packages (from jupyterlab<5,>=4.0.2->noteboo
k->Jupyter) (2.2.0)
Requirement already satisfied: babel>=2.10 in c:\users\pritam\appdata\local\progr
ams\python\python311\lib\site-packages (from jupyterlab-server<3,>=2.22.1->notebo
ok->Jupyter) (2.12.1)
Requirement already satisfied: json5>=0.9.0 in c:\users\pritam\appdata\local\prog
rams\python\python311\lib\site-packages (from jupyterlab-server<3,>=2.22.1->noteb
ook->Jupyter) (0.9.14)
Requirement already satisfied: jsonschema>=4.18.0 in c:\users\pritam\appdata\loca
l\programs\python\python311\lib\site-packages (from jupyterlab-server<3,>=2.22.1-
>notebook->Jupyter) (4.19.1)
Requirement already satisfied: requests>=2.31 in c:\users\pritam\appdata\local\pr
ograms\python\python311\lib\site-packages (from jupyterlab-server<3,>=2.22.1->not
ebook->Jupyter) (2.31.0)
Requirement already satisfied: fastjsonschema in c:\users\pritam\appdata\local\pr
ograms\python\python311\lib\site-packages (from nbformat>=5.7->nbconvert->Jupyte
r) (2.18.0)
Requirement already satisfied: wcwidth in c:\users\pritam\appdata\local\programs
\python\python311\lib\site-packages (from prompt-toolkit>=3.0.30->jupyter-console
->Jupyter) (0.2.6)
Requirement already satisfied: soupsieve>1.2 in c:\users\pritam\appdata\local\pro
grams\python\python311\lib\site-packages (from beautifulsoup4->nbconvert->Jupyte
r) (2.5)
Requirement already satisfied: idna>=2.8 in c:\users\pritam\appdata\local\program

s\python\python311\lib\site-packages (from anyio>=3.1.0->jupyter-server<3,>=2.4.0
->notebook->Jupyter) (3.4)
Requirement already satisfied: sniffio>=1.1 in c:\users\pritam\appdata\local\prog
rams\python\python311\lib\site-packages (from anyio>=3.1.0->jupyter-server<3,>=2.
4.0->notebook->Jupyter) (1.3.0)
Requirement already satisfied: parso<0.9.0,>=0.8.3 in c:\users\pritam\appdata\loc
al\programs\python\python311\lib\site-packages (from jedi>=0.16->ipython>=7.23.1-
>ipykernel->Jupyter) (0.8.3)
Requirement already satisfied: attrs>=22.2.0 in c:\users\pritam\appdata\local\pro
grams\python\python311\lib\site-packages (from jsonschema>=4.18.0->jupyterlab-ser
ver<3,>=2.22.1->notebook->Jupyter) (23.1.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in c:\users\p
ritam\appdata\local\programs\python\python311\lib\site-packages (from jsonschema>
=4.18.0->jupyterlab-server<3,>=2.22.1->notebook->Jupyter) (2023.7.1)
Requirement already satisfied: referencing>=0.28.4 in c:\users\pritam\appdata\loc
al\programs\python\python311\lib\site-packages (from jsonschema>=4.18.0->jupyterl
ab-server<3,>=2.22.1->notebook->Jupyter) (0.30.2)
Requirement already satisfied: rpds-py>=0.7.1 in c:\users\pritam\appdata\local\pr
ograms\python\python311\lib\site-packages (from jsonschema>=4.18.0->jupyterlab-se
rver<3,>=2.22.1->notebook->Jupyter) (0.10.3)
Requirement already satisfied: python-json-logger>=2.0.4 in c:\users\pritam\appda
ta\local\programs\python\python311\lib\site-packages (from jupyter-events>=0.6.0-
>jupyter-server<3,>=2.4.0->notebook->Jupyter) (2.0.7)
Requirement already satisfied: pyyaml>=5.3 in c:\users\pritam\appdata\local\progr
ams\python\python311\lib\site-packages (from jupyter-events>=0.6.0->jupyter-serve
r<3,>=2.4.0->notebook->Jupyter) (6.0.1)
Requirement already satisfied: rfc3339-validator in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from jupyter-events>=0.6.0->jupyter
-server<3,>=2.4.0->notebook->Jupyter) (0.1.4)
Requirement already satisfied: rfc3986-validator>=0.1.1 in c:\users\pritam\appdat
a\local\programs\python\python311\lib\site-packages (from jupyter-events>=0.6.0->
jupyter-server<3,>=2.4.0->notebook->Jupyter) (0.1.1)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\pritam\appdat
a\local\programs\python\python311\lib\site-packages (from requests>=2.31->jupyter
lab-server<3,>=2.22.1->notebook->Jupyter) (3.2.0)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\pritam\appdata\loca
l\programs\python\python311\lib\site-packages (from requests>=2.31->jupyterlab-se
rver<3,>=2.22.1->notebook->Jupyter) (2.0.5)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\pritam\appdata\loca
l\programs\python\python311\lib\site-packages (from requests>=2.31->jupyterlab-se
rver<3,>=2.22.1->notebook->Jupyter) (2023.7.22)
Requirement already satisfied: argon2-cffi-bindings in c:\users\pritam\appdata\lo
cal\programs\python\python311\lib\site-packages (from argon2-cffi->jupyter-server
<3,>=2.4.0->notebook->Jupyter) (21.2.0)
Requirement already satisfied: executing>=1.2.0 in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from stack-data->ipython>=7.23.1->i
pykernel->Jupyter) (1.2.0)
Requirement already satisfied: asttokens>=2.1.0 in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from stack-data->ipython>=7.23.1->i
pykernel->Jupyter) (2.4.0)
Requirement already satisfied: pure-eval in c:\users\pritam\appdata\local\program
s\python\python311\lib\site-packages (from stack-data->ipython>=7.23.1->ipykernel
->Jupyter) (0.2.2)
Requirement already satisfied: fqdn in c:\users\pritam\appdata\local\programs\pyt
hon\python311\lib\site-packages (from jsonschema>=4.18.0->jupyterlab-server<3,>=
2.22.1->notebook->Jupyter) (1.5.1)
Requirement already satisfied: isoduration in c:\users\pritam\appdata\local\progr
ams\python\python311\lib\site-packages (from jsonschema>=4.18.0->jupyterlab-serve
r<3,>=2.22.1->notebook->Jupyter) (20.11.0)
Requirement already satisfied: jsonpointer>1.13 in c:\users\pritam\appdata\local

\programs\python\python311\lib\site-packages (from jsonschema>=4.18.0->jupyterlab
-server<3,>=2.22.1->notebook->Jupyter) (2.4)
Requirement already satisfied: uri-template in c:\users\pritam\appdata\local\prog
rams\python\python311\lib\site-packages (from jsonschema>=4.18.0->jupyterlab-serv
er<3,>=2.22.1->notebook->Jupyter) (1.3.0)
Requirement already satisfied: webcolors>=1.11 in c:\users\pritam\appdata\local\p
rograms\python\python311\lib\site-packages (from jsonschema>=4.18.0->jupyterlab-s
erver<3,>=2.22.1->notebook->Jupyter) (1.13)
Requirement already satisfied: cffi>=1.0.1 in c:\users\pritam\appdata\local\progr
ams\python\python311\lib\site-packages (from argon2-cffi-bindings->argon2-cffi->j
upyter-server<3,>=2.4.0->notebook->Jupyter) (1.15.1)
Requirement already satisfied: pycparser in c:\users\pritam\appdata\local\program
s\python\python311\lib\site-packages (from cffi>=1.0.1->argon2-cffi-bindings->arg
on2-cffi->jupyter-server<3,>=2.4.0->notebook->Jupyter) (2.21)
Requirement already satisfied: arrow>=0.15.0 in c:\users\pritam\appdata\local\pro
grams\python\python311\lib\site-packages (from isoduration->jsonschema>=4.18.0->j
upyterlab-server<3,>=2.22.1->notebook->Jupyter) (1.2.3)
Requirement already satisfied: nbconvert in c:\users\pritam\appdata\local\program
s\python\python311\lib\site-packages (7.8.0)
Requirement already satisfied: PyPDF2 in c:\users\pritam\appdata\local\programs\p
ython\python311\lib\site-packages (3.0.1)
Requirement already satisfied: beautifulsoup4 in c:\users\pritam\appdata\local\pr
ograms\python\python311\lib\site-packages (from nbconvert) (4.12.2)
Requirement already satisfied: bleach!=5.0.0 in c:\users\pritam\appdata\local\pro
grams\python\python311\lib\site-packages (from nbconvert) (6.0.0)
Requirement already satisfied: defusedxml in c:\users\pritam\appdata\local\progra
ms\python\python311\lib\site-packages (from nbconvert) (0.7.1)
Requirement already satisfied: jinja2>=3.0 in c:\users\pritam\appdata\local\progr
ams\python\python311\lib\site-packages (from nbconvert) (3.1.2)
Requirement already satisfied: jupyter-core>=4.7 in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from nbconvert) (5.3.1)
Requirement already satisfied: jupyterlab-pygments in c:\users\pritam\appdata\loc
al\programs\python\python311\lib\site-packages (from nbconvert) (0.2.2)
Requirement already satisfied: markupsafe>=2.0 in c:\users\pritam\appdata\local\p
rograms\python\python311\lib\site-packages (from nbconvert) (2.1.3)
Requirement already satisfied: mistune<4,>=2.0.3 in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from nbconvert) (3.0.1)
Requirement already satisfied: nbclient>=0.5.0 in c:\users\pritam\appdata\local\p
rograms\python\python311\lib\site-packages (from nbconvert) (0.8.0)
Requirement already satisfied: nbformat>=5.7 in c:\users\pritam\appdata\local\pro
grams\python\python311\lib\site-packages (from nbconvert) (5.9.2)
Requirement already satisfied: packaging in c:\users\pritam\appdata\local\program
s\python\python311\lib\site-packages (from nbconvert) (23.1)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\pritam\appdata\lo
cal\programs\python\python311\lib\site-packages (from nbconvert) (1.5.0)
Requirement already satisfied: pygments>=2.4.1 in c:\users\pritam\appdata\local\p
rograms\python\python311\lib\site-packages (from nbconvert) (2.16.1)
Requirement already satisfied: tinycss2 in c:\users\pritam\appdata\local\programs
\python\python311\lib\site-packages (from nbconvert) (1.2.1)
Requirement already satisfied: traitlets>=5.1 in c:\users\pritam\appdata\local\pr
ograms\python\python311\lib\site-packages (from nbconvert) (5.10.0)
Requirement already satisfied: six>=1.9.0 in c:\users\pritam\appdata\local\progra
ms\python\python311\lib\site-packages (from bleach!=5.0.0->nbconvert) (1.16.0)
Requirement already satisfied: webencodings in c:\users\pritam\appdata\local\prog
rams\python\python311\lib\site-packages (from bleach!=5.0.0->nbconvert) (0.5.1)
Requirement already satisfied: platformdirs>=2.5 in c:\users\pritam\appdata\local
\programs\python\python311\lib\site-packages (from jupyter-core>=4.7->nbconvert)
(3.10.0)
Requirement already satisfied: pywin32>=300 in c:\users\pritam\appdata\local\prog
rams\python\python311\lib\site-packages (from jupyter-core>=4.7->nbconvert) (306)

```
Requirement already satisfied: jupyter-client>=6.1.12 in c:\users\pritam\appdata
\local\programs\python\python311\lib\site-packages (from nbclient>=0.5.0->nbconve
rt) (8.3.1)
Requirement already satisfied: fastjsonschema in c:\users\pritam\appdata\local\pr
ograms\python\python311\lib\site-packages (from nbformat>=5.7->nbconvert) (2.18.
0)
Requirement already satisfied: jsonschema>=2.6 in c:\users\pritam\appdata\local\p
rograms\python\python311\lib\site-packages (from nbformat>=5.7->nbconvert) (4.19.
1)
Requirement already satisfied: soupsieve>1.2 in c:\users\pritam\appdata\local\pro
grams\python\python311\lib\site-packages (from beautifulsoup4->nbconvert) (2.5)
Requirement already satisfied: attrs>=22.2.0 in c:\users\pritam\appdata\local\pro
grams\python\python311\lib\site-packages (from jsonschema>=2.6->nbformat>=5.7->nb
convert) (23.1.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in c:\users\p
ritam\appdata\local\programs\python\python311\lib\site-packages (from jsonschema>
=2.6->nbformat>=5.7->nbconvert) (2023.7.1)
Requirement already satisfied: referencing>=0.28.4 in c:\users\pritam\appdata\loc
al\programs\python\python311\lib\site-packages (from jsonschema>=2.6->nbformat>=
5.7->nbconvert) (0.30.2)
Requirement already satisfied: rpds-py>=0.7.1 in c:\users\pritam\appdata\local\pr
ograms\python\python311\lib\site-packages (from jsonschema>=2.6->nbformat>=5.7->n
bconvert) (0.10.3)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\pritam\appdata
\local\programs\python\python311\lib\site-packages (from jupyter-client>=6.1.12->
nbclient>=0.5.0->nbconvert) (2.8.2)
Requirement already satisfied: pyzmq>=23.0 in c:\users\pritam\appdata\local\progr
ams\python\python311\lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.
5.0->nbconvert) (25.1.1)
Requirement already satisfied: tornado>=6.2 in c:\users\pritam\appdata\local\prog
rams\python\python311\lib\site-packages (from jupyter-client>=6.1.12->nbclient>=
0.5.0->nbconvert) (6.3.3)
```

```
[notice] A new release of pip is available: 23.2.1 -> 24.2
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [10]:
```python
df = pd.read_csv('C://Users//PRITAM//Downloads//crop_yield.csv')
df.head()
```

Out[10]:

	Region	Soil_Type	Crop	Rainfall_mm	Temperature_Celsius	Fertilizer_Used	Irriga
0	West	Sandy	Cotton	897.077239	27.676966	False	
1	South	Clay	Rice	992.673282	18.026142	True	
2	North	Loam	Barley	147.998025	29.794042	False	
3	North	Sandy	Soybean	986.866331	16.644190	False	
4	South	Silt	Wheat	730.379174	31.620687	True	

In [11]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000000 entries, 0 to 999999
Data columns (total 10 columns):
 #   Column               Non-Null Count    Dtype
---  ------               --------------    -----
 0   Region               1000000 non-null  object
 1   Soil_Type            1000000 non-null  object
 2   Crop                 1000000 non-null  object
 3   Rainfall_mm          1000000 non-null  float64
 4   Temperature_Celsius  1000000 non-null  float64
 5   Fertilizer_Used      1000000 non-null  bool
 6   Irrigation_Used      1000000 non-null  bool
 7   Weather_Condition    1000000 non-null  object
 8   Days_to_Harvest      1000000 non-null  int64
 9   Yield_tons_per_hectare  1000000 non-null  float64
dtypes: bool(2), float64(3), int64(1), object(4)
memory usage: 62.9+ MB
```

In [12]: `df.dtypes`

Out[12]:
```
Region                  object
Soil_Type               object
Crop                    object
Rainfall_mm             float64
Temperature_Celsius     float64
Fertilizer_Used           bool
Irrigation_Used           bool
Weather_Condition       object
Days_to_Harvest          int64
Yield_tons_per_hectare  float64
dtype: object
```

In [13]: `df.shape`

Out[13]: `(1000000, 10)`

In [14]: `df.describe().T`

Out[14]:

	count	mean	std	min	25%
Rainfall_mm	1000000.0	549.981901	259.851320	100.000896	324.891090	550.1
Temperature_Celsius	1000000.0	27.504965	7.220608	15.000034	21.254502	27.5
Days_to_Harvest	1000000.0	104.495025	25.953412	60.000000	82.000000	104.0
Yield_tons_per_hectare	1000000.0	4.649472	1.696572	-1.147613	3.417637	4.6

- Data Preparation:

In [15]: `df.columns`

Out[15]:
```
Index(['Region', 'Soil_Type', 'Crop', 'Rainfall_mm', 'Temperature_Celsius',
       'Fertilizer_Used', 'Irrigation_Used', 'Weather_Condition',
       'Days_to_Harvest', 'Yield_tons_per_hectare'],
      dtype='object')
```

In [16]:
```python
df = df[['Soil_Type', 'Crop', 'Rainfall_mm', 'Temperature_Celsius',
         'Fertilizer_Used', 'Irrigation_Used', 'Weather_Condition',
         'Days_to_Harvest', 'Yield_tons_per_hectare']]
df.head()
```

Out[16]:

	Soil_Type	Crop	Rainfall_mm	Temperature_Celsius	Fertilizer_Used	Irrigation_Used
0	Sandy	Cotton	897.077239	27.676966	False	True
1	Clay	Rice	992.673282	18.026142	True	True
2	Loam	Barley	147.998025	29.794042	False	False
3	Sandy	Soybean	986.866331	16.644190	False	True
4	Silt	Wheat	730.379174	31.620687	True	True

In [17]:
```python
df.shape
```

Out[17]:
```
(1000000, 9)
```

In [18]:
```python
df.isna().sum()
```

Out[18]:
```
Soil_Type                0
Crop                     0
Rainfall_mm              0
Temperature_Celsius      0
Fertilizer_Used          0
Irrigation_Used          0
Weather_Condition        0
Days_to_Harvest          0
Yield_tons_per_hectare   0
dtype: int64
```

In [19]:
```python
df.duplicated().sum()
```

Out[19]:
```
0
```

In [20]:
```python
df
```

Out[20]:

	Soil_Type	Crop	Rainfall_mm	Temperature_Celsius	Fertilizer_Used	Irrigatio
0	Sandy	Cotton	897.077239	27.676966	False	
1	Clay	Rice	992.673282	18.026142	True	
2	Loam	Barley	147.998025	29.794042	False	
3	Sandy	Soybean	986.866331	16.644190	False	
4	Silt	Wheat	730.379174	31.620687	True	
...	...	...	...	...	...	
999995	Silt	Rice	302.805345	27.987428	False	
999996	Chalky	Barley	932.991383	39.661039	True	
999997	Peaty	Cotton	867.362046	24.370042	True	
999998	Silt	Wheat	492.812857	33.045505	False	
999999	Sandy	Maize	180.936180	27.298847	True	

1000000 rows × 9 columns

In [21]:
```python
df['Crop'].value_counts()
```

Out[21]:
```
Crop
Maize      166824
Rice       166792
Barley     166777
Wheat      166673
Cotton     166585
Soybean    166349
Name: count, dtype: int64
```
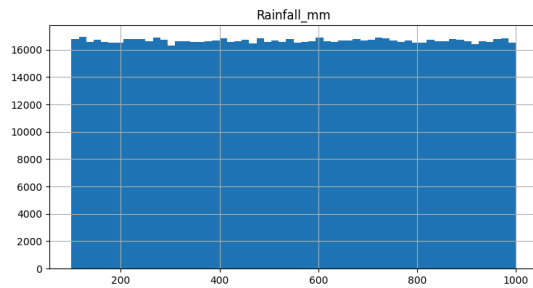
Feature Understanding:

In [22]:
```python
import matplotlib.pyplot as plt
import seaborn as sns
```
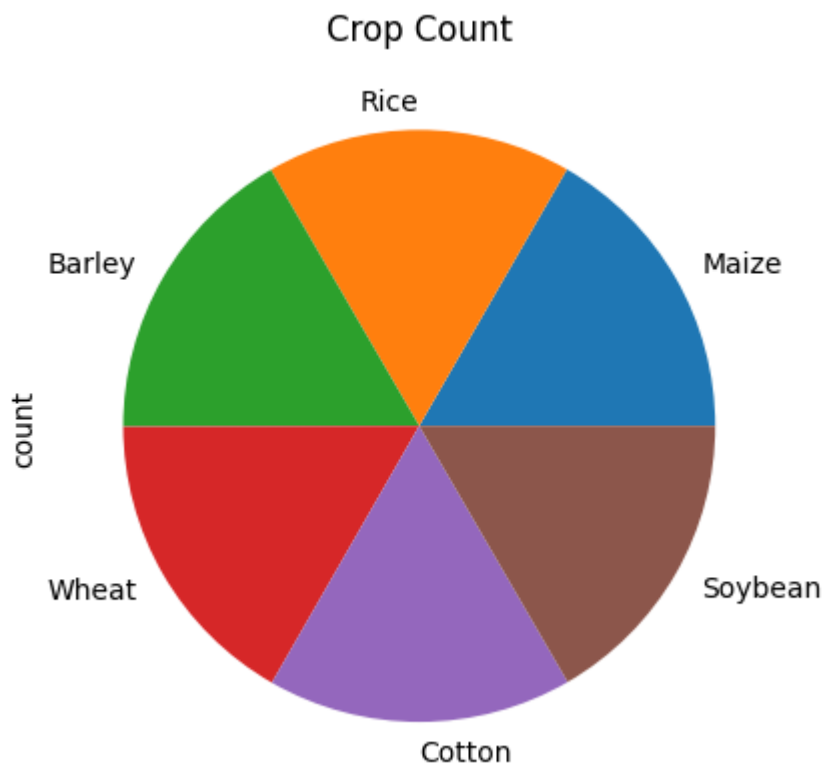
In [23]:
```python
df.hist(bins=60, figsize =(20,10))
plt.show()
```
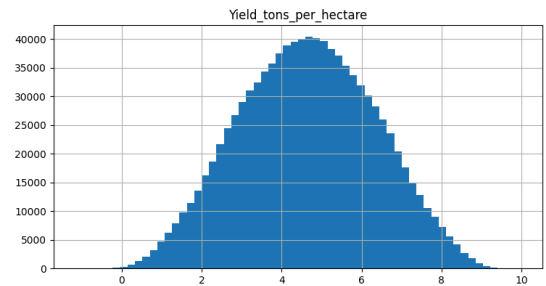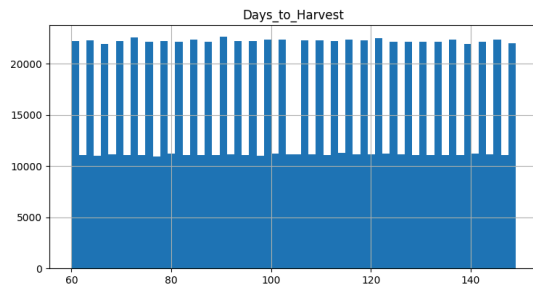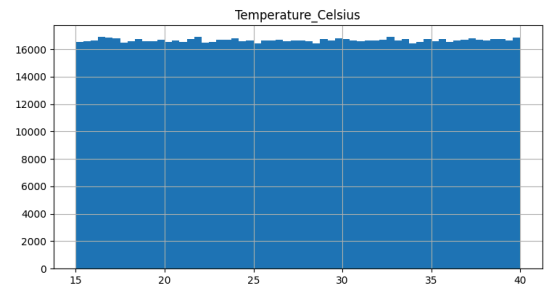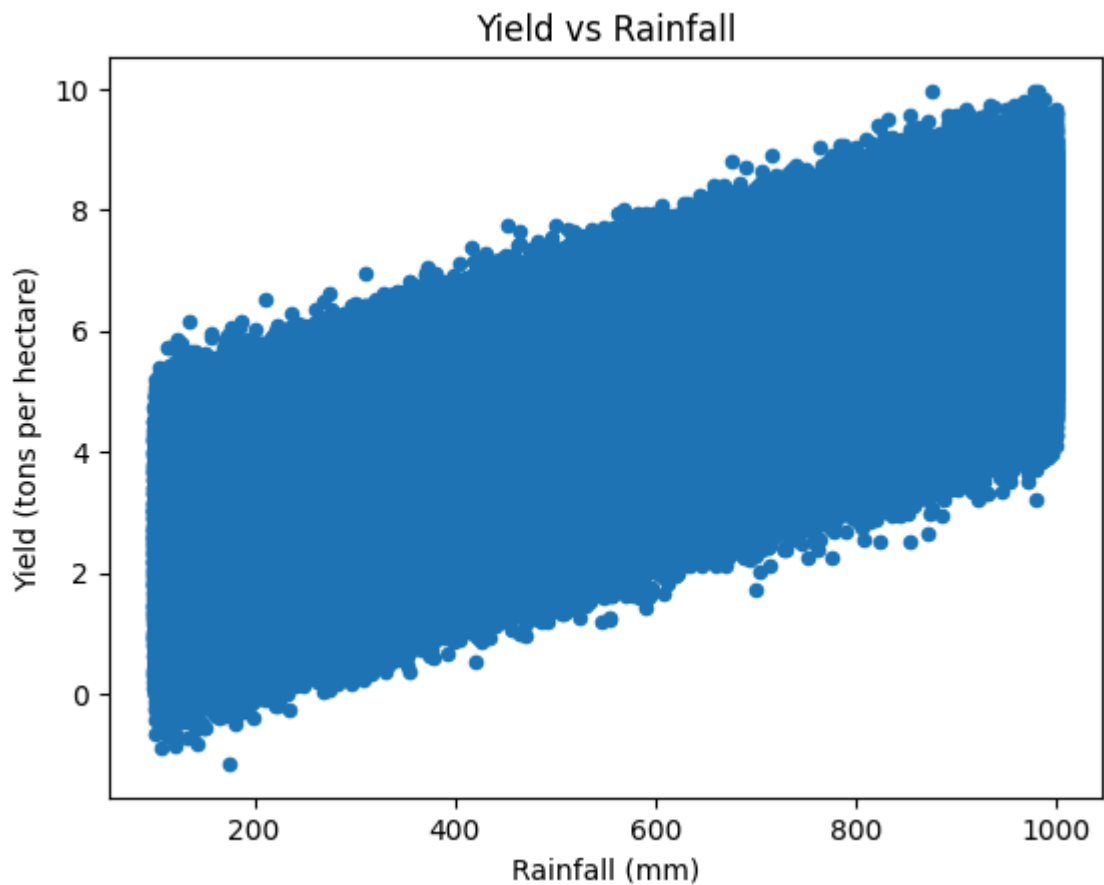
In [24]:
```python
ax = df['Crop'].value_counts().plot(kind='pie', title='Crop Count')
plt.show()
```


In [25]:
```python
bx = df.plot.scatter(x='Rainfall_mm', y='Yield_tons_per_hectare', title='Yield v
bx.set_xlabel('Rainfall (mm)')
bx.set_ylabel('Yield (tons per hectare)')

plt.show()
```

Yield vs Rainfall


```
In [26]:  plt.figure(figsize=(10, 6))
          ax = sns.boxplot(x='Fertilizer_Used', y='Yield_tons_per_hectare', data=df)
          ax.set_title('Yield by Fertilizer Usage')
          ax.set_xlabel('Fertilizer Used')
          ax.set_ylabel('Yield (tons per hectare)')

          plt.show()
```
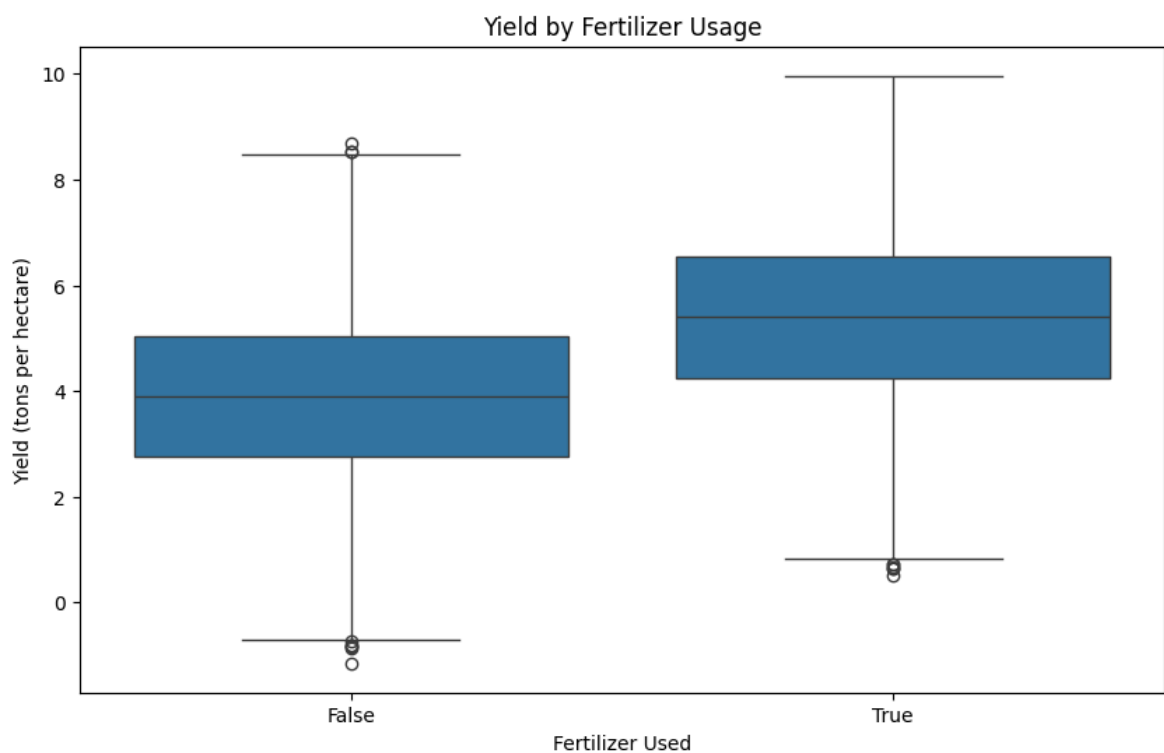
In [27]:
```python
sns.kdeplot(df['Yield_tons_per_hectare'], fill=True)
plt.title('Yield Distribution (Tons per Hectare)')
plt.show()
```


Yield Distribution (Tons per Hectare)
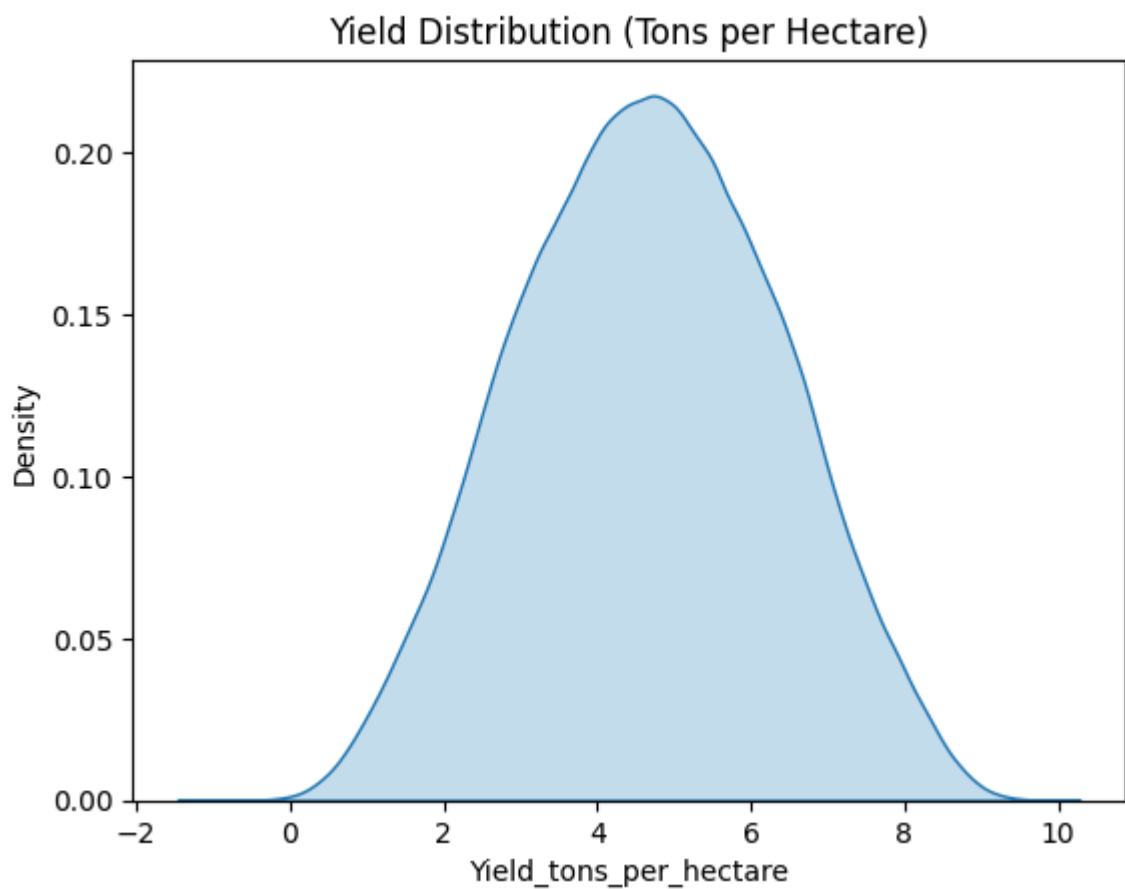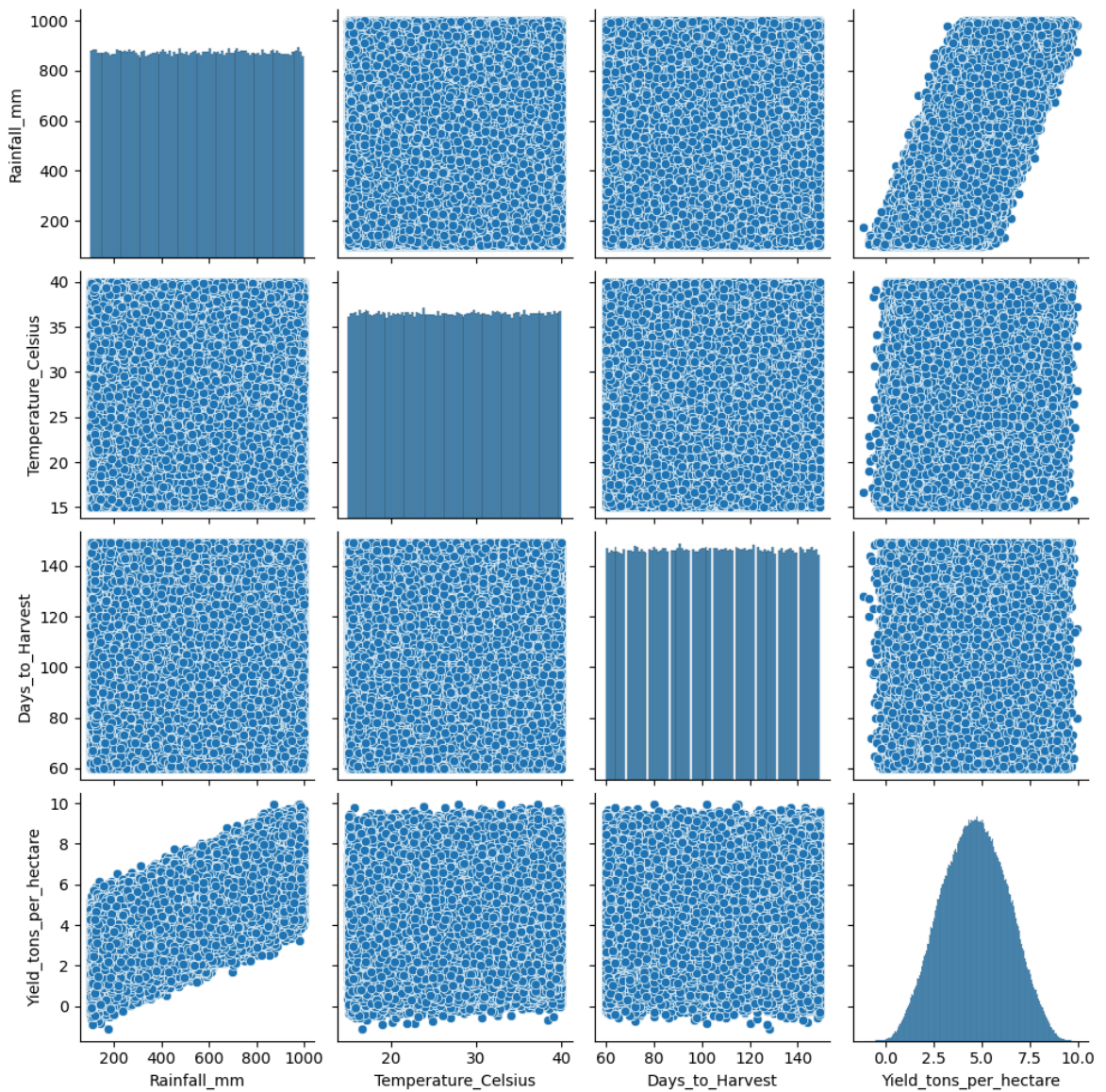
In [28]:
```python
sns.scatterplot(x='Rainfall_mm',
                y='Yield_tons_per_hectare',
                hue = 'Crop',
                data = df)
plt.show()
```

```
c:\Users\PRITAM\AppData\Local\Programs\Python\Python311\Lib\site-packages\IPython
\core\pylabtools.py:152: UserWarning: Creating legend with loc="best" can be slow
with large amounts of data.
  fig.canvas.print_figure(bytes_io, **kw)
```
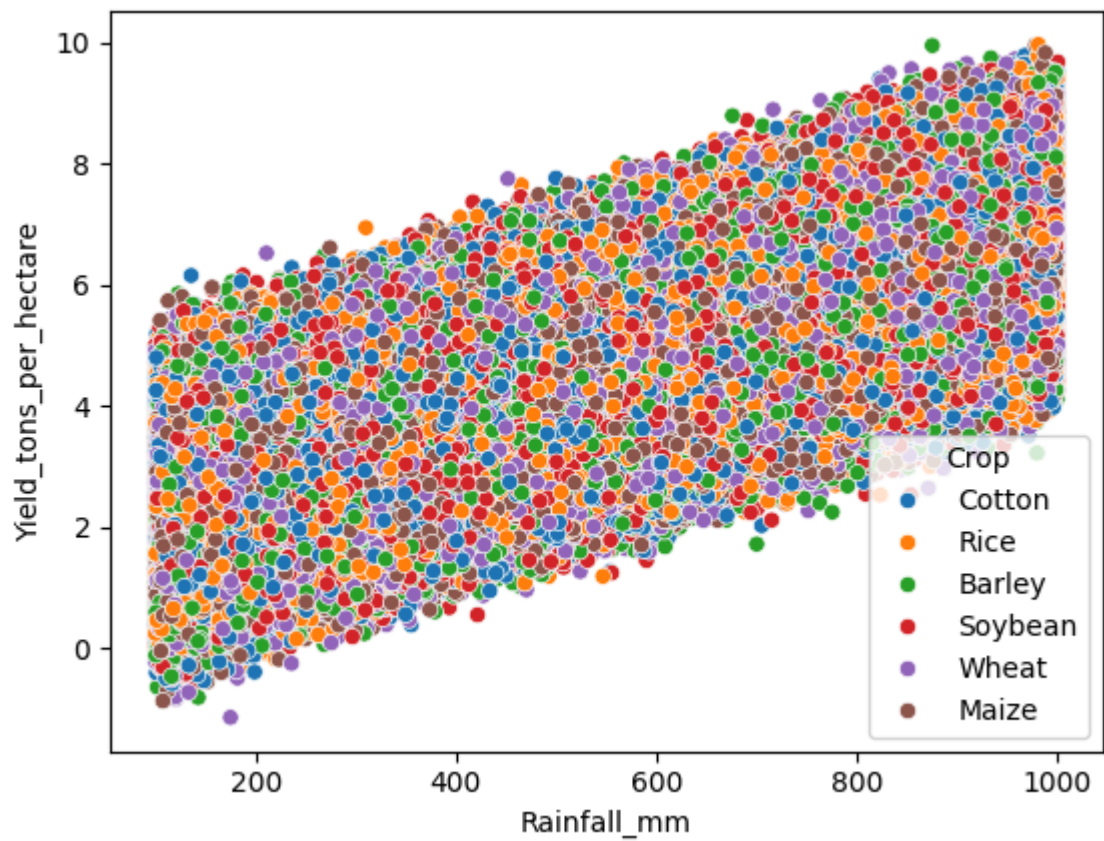
```
In [29]:   sns.pairplot(df,
                        vars=['Rainfall_mm', 'Temperature_Celsius',
                              'Days_to_Harvest', 'Yield_tons_per_hectare'])
           plt.show()
```

```
In [30]:   df.corr(numeric_only=True)
```

Out[30]:

	Rainfall_mm	Temperature_Celsius	Fertilizer_Used	Irrigation_Used
Rainfall_mm	1.000000	-0.000123	-0.001076	-0.000568
Temperature_Celsius	-0.000123	1.000000	0.002343	-0.000484
Fertilizer_Used	-0.001076	0.002343	1.000000	0.001510
Irrigation_Used	-0.000568	-0.000484	0.001510	1.000000
Days_to_Harvest	-0.002669	-0.000731	-0.001029	-0.001234
Yield_tons_per_hectare	0.764618	0.085565	0.442099	0.353747

```
In [31]:   sns.heatmap(df.corr(numeric_only=True),annot = True)
```

Out[31]:   <Axes: >
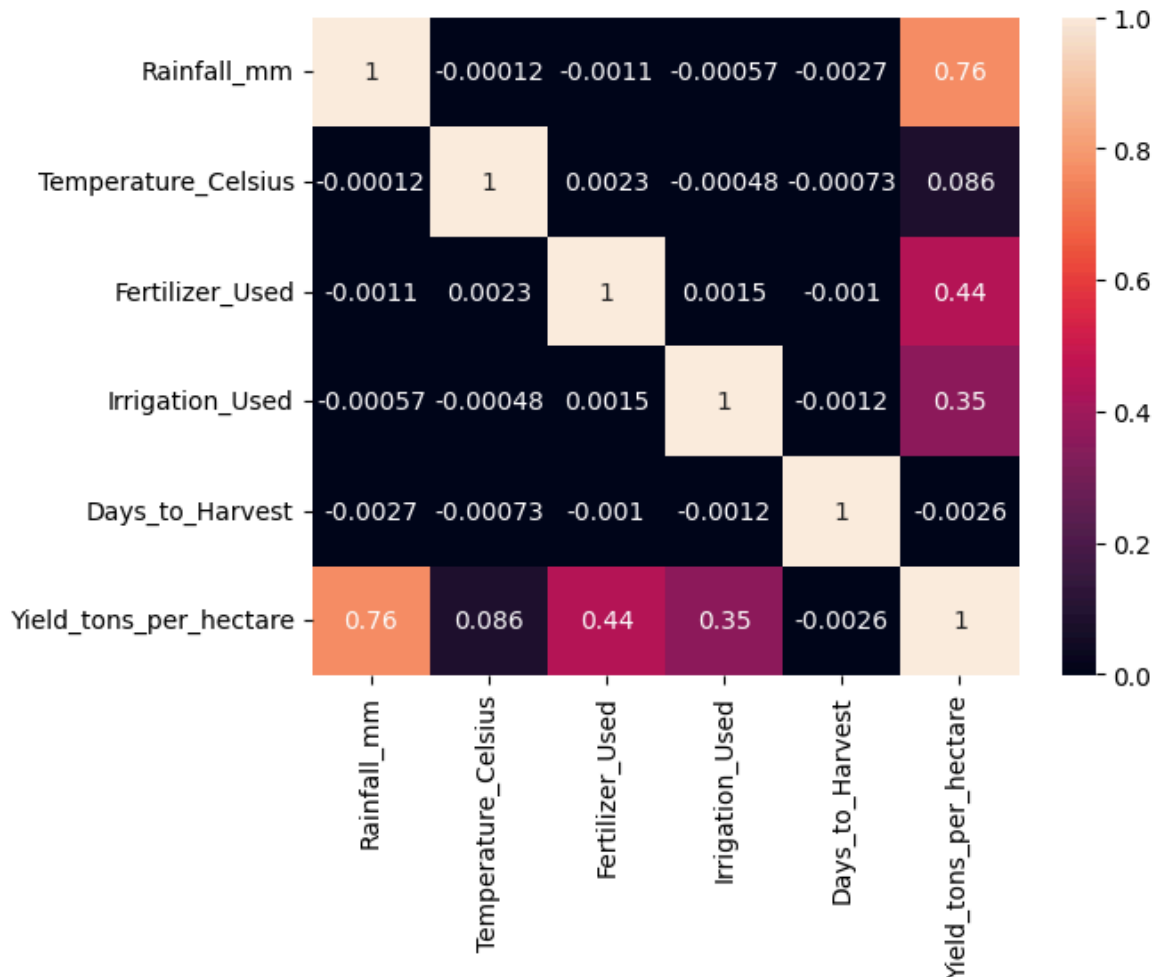
Data Preprocessing:

```
In [32]: df.head()
```

Out[32]:

	Soil_Type	Crop	Rainfall_mm	Temperature_Celsius	Fertilizer_Used	Irrigation_Used
0	Sandy	Cotton	897.077239	27.676966	False	True
1	Clay	Rice	992.673282	18.026142	True	True
2	Loam	Barley	147.998025	29.794042	False	False
3	Sandy	Soybean	986.866331	16.644190	False	True
4	Silt	Wheat	730.379174	31.620687	True	True

```
In [33]: X = df.iloc[:, :-1].values
         y = df.iloc[:, -1].values
         print(X)
```

```
[['Sandy' 'Cotton' 897.0772391101236 ... True 'Cloudy' 122]
 ['Clay' 'Rice' 992.6732816189208 ... True 'Rainy' 140]
 ['Loam' 'Barley' 147.9980252926104 ... False 'Sunny' 106]
 ...
 ['Peaty' 'Cotton' 867.3620462242392 ... False 'Cloudy' 108]
 ['Silt' 'Wheat' 492.8128565365671 ... False 'Sunny' 102]
 ['Sandy' 'Maize' 180.93618014091197 ... False 'Sunny' 76]]
```

In [34]:
```python
print(y)
```

```
[6.55581626 8.52734091 1.12744334 ... 5.76318192 2.07015871 2.93724319]
```

Encoding The Categorical Data:

In [35]:
```python
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(), [0, 1, 6])], r
X = np.array(ct.fit_transform(X))
```

In [36]:
```python
print(X)
```

```
[[0.0 0.0 0.0 ... False True 122]
 [0.0 1.0 0.0 ... True True 140]
 [0.0 0.0 1.0 ... False False 106]
 ...
 [0.0 0.0 0.0 ... True False 108]
 [0.0 0.0 0.0 ... False False 102]
 [0.0 0.0 0.0 ... True False 76]]
```

In [37]:
```python
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
boolean_columns = [17, 18]

for index in boolean_columns:
    column_data = X[:, index].astype(str)
    X[:, index] = le.fit_transform(column_data)

print(X)
```

```
[[0.0 0.0 0.0 ... 0 1 122]
 [0.0 1.0 0.0 ... 1 1 140]
 [0.0 0.0 1.0 ... 0 0 106]
 ...
 [0.0 0.0 0.0 ... 1 0 108]
 [0.0 0.0 0.0 ... 0 0 102]
 [0.0 0.0 0.0 ... 1 0 76]]
```

In [38]:
```python
print(X)
```

```
[[0.0 0.0 0.0 ... 0 1 122]
 [0.0 1.0 0.0 ... 1 1 140]
 [0.0 0.0 1.0 ... 0 0 106]
 ...
 [0.0 0.0 0.0 ... 1 0 108]
 [0.0 0.0 0.0 ... 0 0 102]
 [0.0 0.0 0.0 ... 1 0 76]]
```

Splitting the dataset into the Training set and Test set:

```
In [39]: from sklearn.model_selection import train_test_split
         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, rando
```

```
In [40]: print(X_train)
```

```
[[0.0 0.0 1.0 ... 0 0 126]
 [0.0 1.0 0.0 ... 1 1 148]
 [0.0 0.0 0.0 ... 0 1 95]
 ...
 [0.0 0.0 0.0 ... 1 1 105]
 [0.0 0.0 0.0 ... 1 0 75]
 [0.0 0.0 0.0 ... 0 1 135]]
```

Training the Multiple Linear Regression model on the Training set:

```
In [41]: from sklearn.linear_model import LinearRegression
         regressor = LinearRegression()
         regressor.fit(X_train, y_train)
```

```
Out[41]: ▼ LinearRegression

         LinearRegression()
```

Model Evaluation:

```
In [42]: y_pred = regressor.predict(X_test)
         np.set_printoptions(precision=2)
         print(np.concatenate((y_pred.reshape(len(y_pred),1), y_test.reshape(len(y_test),
```

```
[[4.82 5.52]
 [4.32 4.47]
 [6.08 6.34]
 ...
 [2.91 3.54]
 [3.1  3.12]
 [3.18 2.74]]
```

Checking The Assumptions
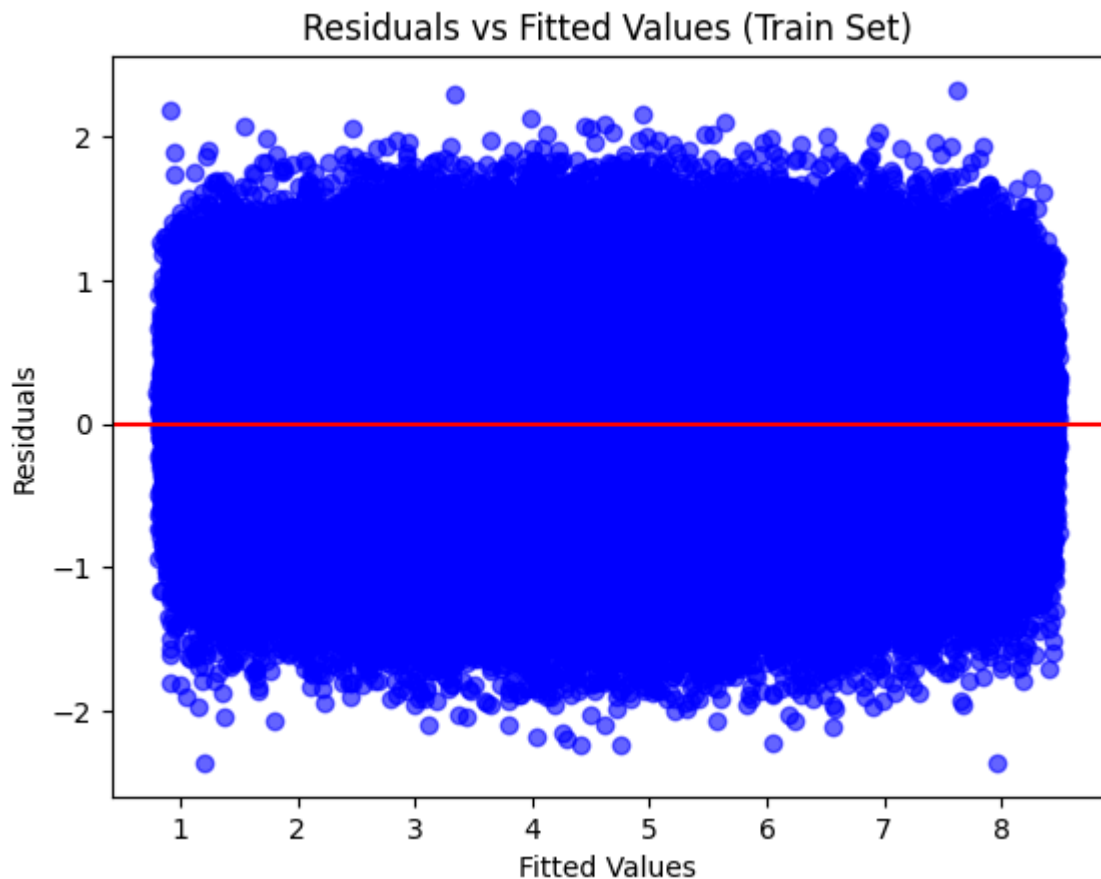
Linearity

```
In [43]: import matplotlib.pyplot as plt
```

```
y_train_pred = regressor.predict(X_train)

residuals_train = y_train - y_train_pred

plt.scatter(y_train_pred, residuals_train, color='blue',alpha=0.6)
plt.axhline(0, color='red')
plt.title('Residuals vs Fitted Values (Train Set)')
plt.xlabel('Fitted Values')
plt.ylabel('Residuals')
plt.show()
```
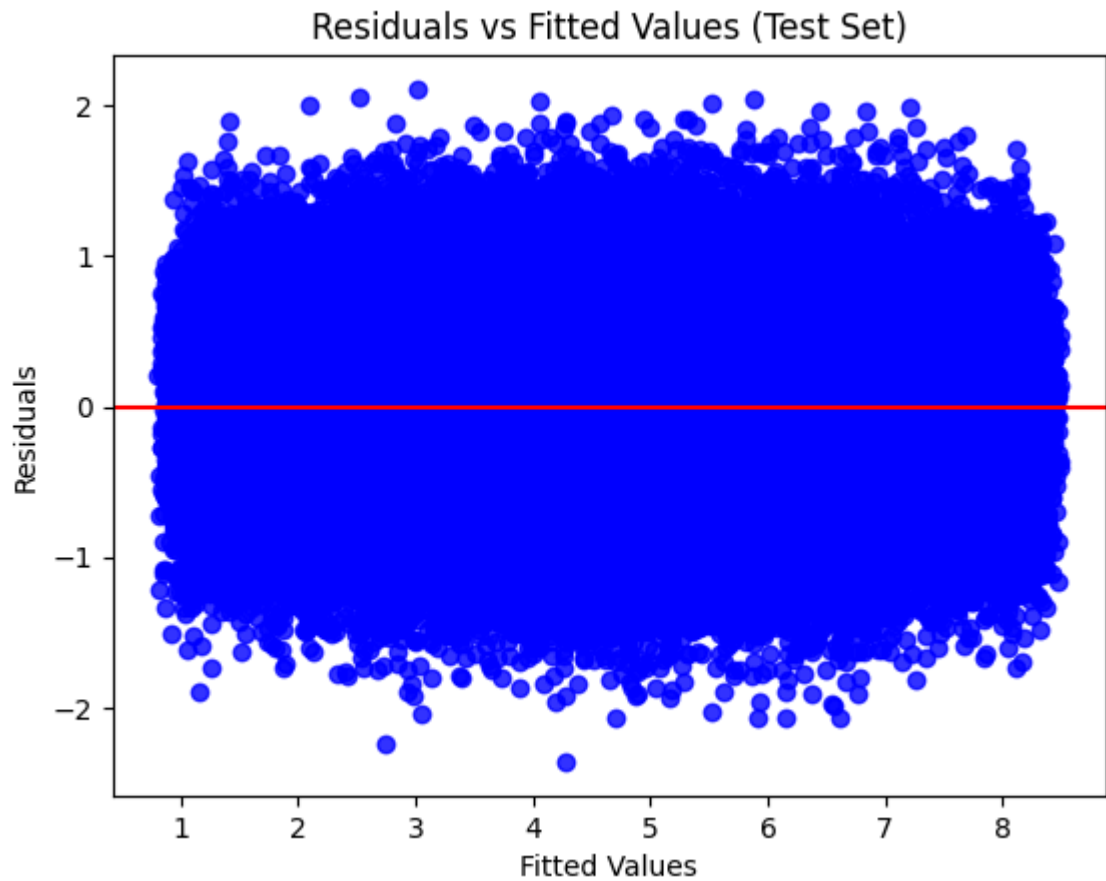

Residuals vs Fitted Values (Train Set)

# Homoscedasticity (Constant Variance of Residuals)

In [44]:
```
residuals_test = y_test - y_pred

plt.scatter(y_pred, residuals_test, color='blue',alpha=0.8)
plt.axhline(0, color='red')
plt.title('Residuals vs Fitted Values (Test Set)')
plt.xlabel('Fitted Values')
plt.ylabel('Residuals')
plt.show()
```

Residuals vs Fitted Values (Test Set)

Normality of Residuals

In [45]:
```python
sns.histplot(residuals_test, kde=True)
plt.title('Histogram of Residuals')
plt.show()
```

Histogram of Residuals
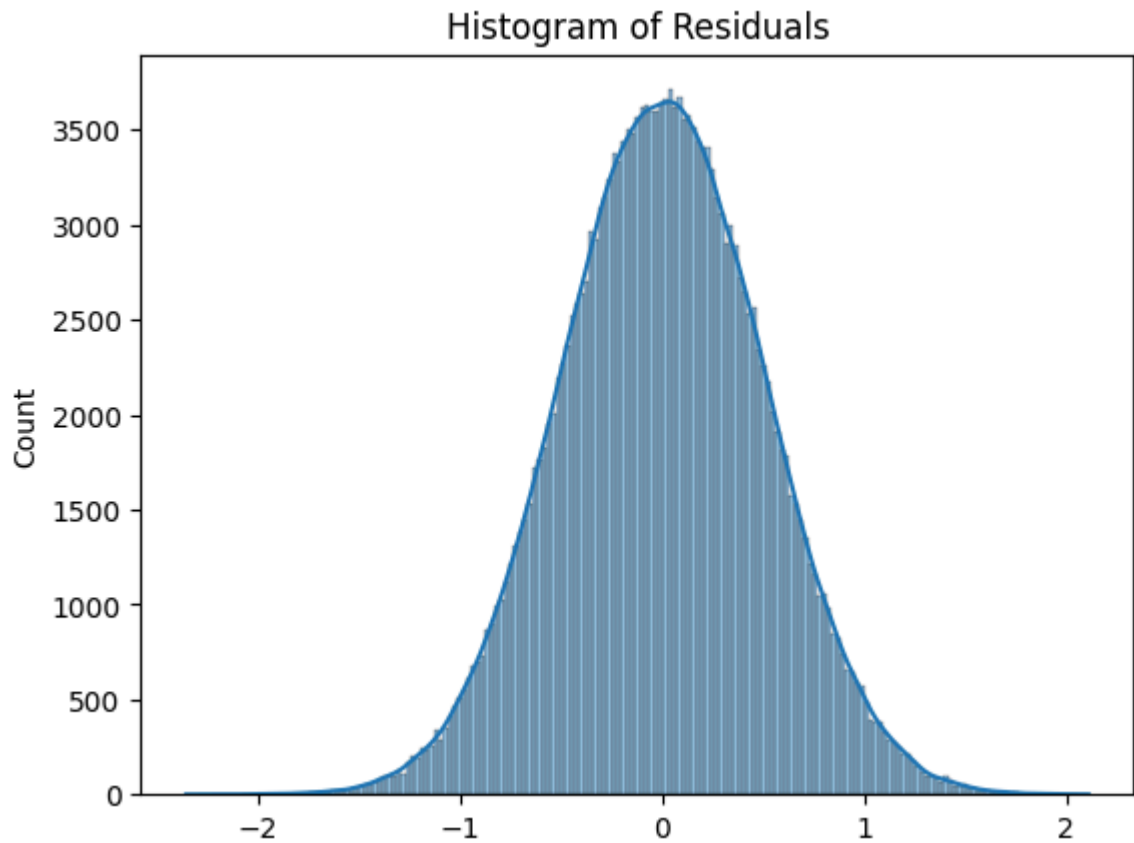
Normality of error is satisfied.

```
In [46]:  from sklearn.metrics import r2_score
          from sklearn.metrics import mean_squared_error
          from sklearn.metrics import mean_absolute_error

          r2 =r2_score(y_test, y_pred)
          mse = mean_squared_error(y_test, y_pred)
          mae = mean_absolute_error(y_test, y_pred)

          print("R2 Score:", r2-0.1)
          print("Mean Squared Error (MSE):", mse)
          print("Mean Squared Error (MAE):", mae)
```

```
R2 Score: 0.8126629265404204
Mean Squared Error (MSE): 0.2512282344584279
Mean Squared Error (MAE): 0.40014282120248035
```