In [115...

```
#Loading the Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.pipeline import Pipeline
from sklearn.metrics import r2_score, mean_absolute_error
```

In [133...

```
#Load the dataset
df=pd.read_csv('water-requirement-data.csv')
```

In [135...

```
df.head()
```

Out[135...

	CROP_TYPE	SOIL_TYPE	REGION	WEATHER_CONDITION	TEMP_MIN	TEMP_MAX	WATI
0	POTATO	DRY	DESERT	NORMAL	10	20	
1	POTATO	DRY	DESERT	SUNNY	10	20	
2	POTATO	DRY	DESERT	WINDY	10	20	
3	POTATO	DRY	DESERT	RAINY	10	20	
4	POTATO	DRY	DESERT	NORMAL	20	30	

In [121...

```
#Checking the null values
df.isnull().sum()
```

Out[121...

```
CROP _TYPE           0
SOIL _TYPE           0
REGION               0
WEATHER_CONDITION    0
TEMP_MIN             0
TEMP_MAX             0
WATER_REQUIREMENT    0
dtype: int64
```

In [123...

```
df.duplicated().sum()
```

Out[123...

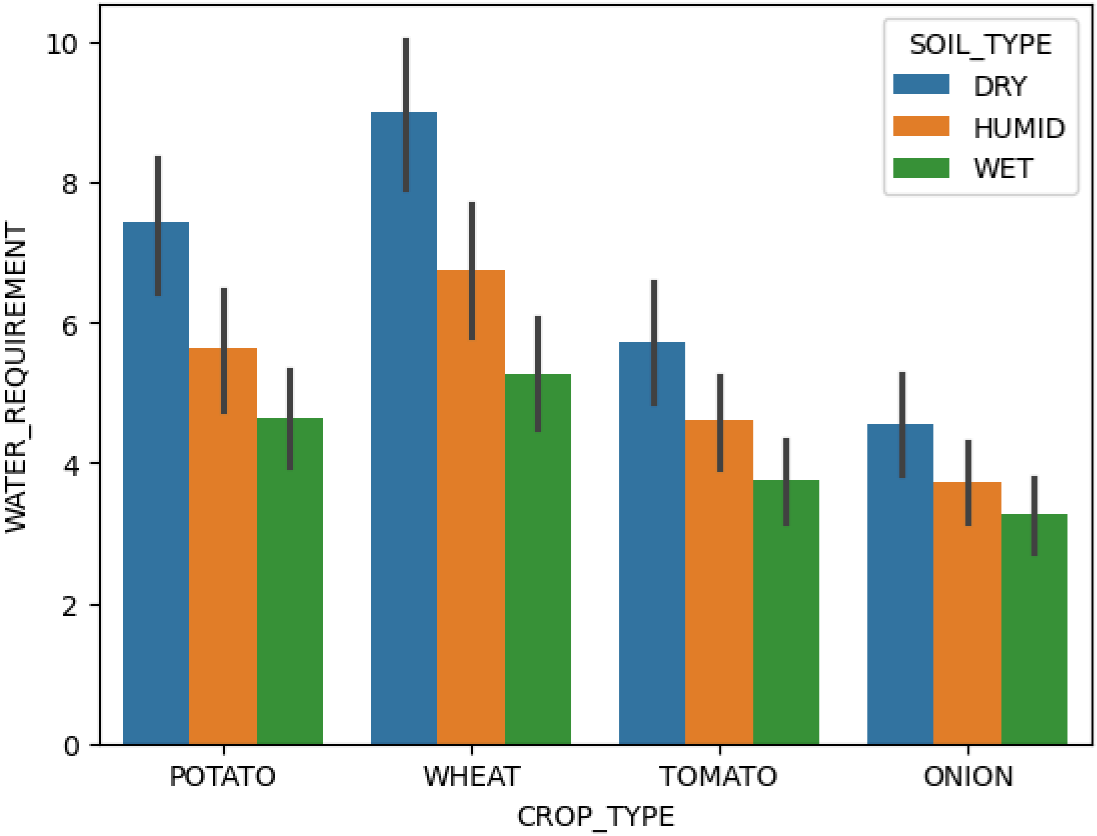
```
0
```

In [137...

```
#checking the graphical interpretations
sns.barplot(x='CROP_TYPE', y='WATER_REQUIREMENT', hue='SOIL_TYPE',data=df)
```

Out[137...

```
<Axes: xlabel='CROP_TYPE', ylabel='WATER_REQUIREMENT'>
```

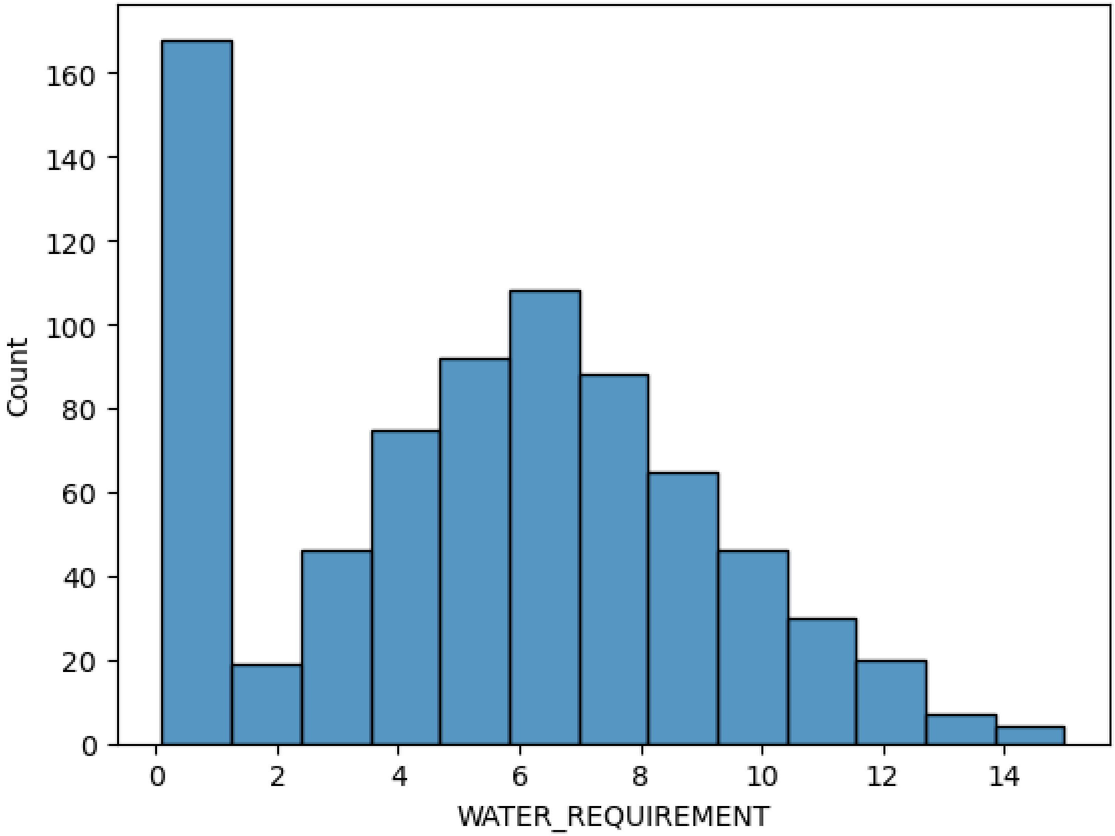


In [139…

```
#Making the Histogarm to check the weither the data is normal or not
sns.histplot(x='WATER_REQUIREMENT', data=df,)
```

Out[139…

```
<Axes: xlabel='WATER_REQUIREMENT', ylabel='Count'>
```



In [141…

```
# Seperating input and target columns
x=df.iloc[:,:6]
y=df.iloc[:,6:]
```

In [143…

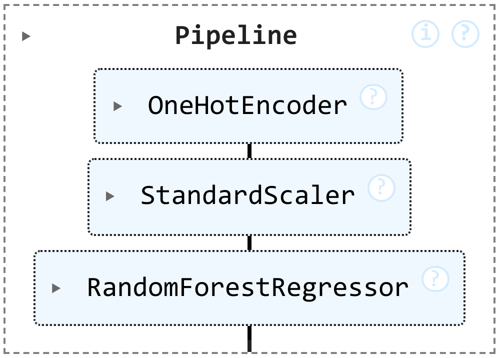
```
#Applying the train test split
X_train,X_test,y_train, y_test=train_test_split(x,y,test_size=0.2, random_state=42)
```

In [153…

```
#Applying the pipline function
pipeline=Pipeline([
    ('OneHot',OneHotEncoder(sparse_output=False, handle_unknown='ignore')),
    ('scalar',StandardScaler()),
    ('model',RandomForestRegressor())
])
pipeline.fit(X_train,y_train)
```

```
C:\Users\ALI\anaconda3\Lib\site-packages\sklearn\base.py:1473: DataConversionWarnin
g: A column-vector y was passed when a 1d array was expected. Please change the shap
e of y to (n_samples,), for example using ravel().
  return fit_method(estimator, *args, **kwargs)
```

Out[153…



In [155…

```
#Checking the accuracy of the Model
print(r2_score(y_pred,y_test))
print(mean_absolute_error(y_pred,y_test))
```

```
0.8981898859009387
0.6584203927025354
```

In [157…

```
#saving the model in pickle Library
import pickle
pickle.dump(pipeline,open('Crop_Water_Requirement.pkl','wb'))
```

In [ ]: