In [214…
```python
import pandas as pd
from sklearn.preprocessing import LabelEncoder

# Load data
data = pd.read_csv('transformed_water-and-sanitation.csv')

# Inspecting the column names
print(data.columns)

# Selecting features and target variable
features = [
    'usage_of_improved_drinking_water_sources', 'usage_of_basic_drinking_water_services',
    'usage_of_limited_drinking_water_services', 'usage_of_unimproved_drinking_water_sources',
    'no_usage_of_drinking_water_facilities', 'usage_of_improved_sanitation_facilities',
    'usage_of_basic_sanitation_services'
]
target = 'usage_of_safely_managed_drinking_water_services'
print(data.head())
```

```
Index(['country', 'year', 'usage_of_improved_drinking_water_sources',
       'usage_of_basic_drinking_water_services',
       'usage_of_limited_drinking_water_services',
       'usage_of_unimproved_drinking_water_sources',
       'no_usage_of_drinking_water_facilities',
       'usage_of_safely_managed_drinking_water_services',
       'usage_of_improved_sanitation_facilities',
       'usage_of_basic_sanitation_services',
       ...
       'wat_pip_urban', 'wat_pip_number_rural', 'wat_pip_number',
       'wat_pip_number_urban', 'wat_sm_number_without_rural',
       'wat_sm_number_without', 'wat_sm_number_without_urban',
       'wat_sm_without_rural', 'wat_sm_without', 'wat_sm_without_urban'],
      dtype='object', length=206)
   country  year  usage_of_improved_drinking_water_sources  \
0  Algeria  2007                                      97.0
1  Algeria  2008                                      97.0
2  Algeria  2009                                      97.0
3  Algeria  2010                                      98.0
4  Algeria  2011                                      98.0

   usage_of_basic_drinking_water_services  \
0                                    18.0
1                                    18.0
2                                    17.0
3                                    17.0
4                                    17.0

   usage_of_limited_drinking_water_services  \
0                                       5.0
1                                       5.0
2                                       5.0
3                                       5.0
4                                       5.0

   usage_of_unimproved_drinking_water_sources  \
0                                         3.0
1                                         3.0
2                                         2.0
3                                         2.0
4                                         2.0

   no_usage_of_drinking_water_facilities  \
0                                    0.0
1                                    0.0
2                                    0.0
3                                    0.0
4                                    0.0

   usage_of_safely_managed_drinking_water_services  \
0                                             74.0
1                                             74.0
2                                             75.0
3                                             75.0
4                                             75.0

   usage_of_improved_sanitation_facilities  \
0                                     94.0
1                                     94.0
2                                     95.0
3                                     95.0
4                                     95.0

   usage_of_basic_sanitation_services  ...  wat_pip_urban  \
0                                25.0  ...      85.460900
1                                25.0  ...      84.741210
2                                25.0  ...      84.021520
3                                25.0  ...      83.301834
4                                25.0  ...      82.582146

   wat_pip_number_rural  wat_pip_number  wat_pip_number_urban  \
0             7648207.0      26627154.0            18978944.0
1             7566640.5      26929552.0            19362912.0
2             7493047.0      27254994.0            19761948.0
3             7424406.5      27597924.0            20173518.0
4             7359657.0      27952218.0            20592560.0

   wat_sm_number_without_rural  wat_sm_number_without  \
0                    4905125.5              8930706.0
1                    4772996.0              8924893.0
2                    4645956.0              8930049.0
3                    4521925.5              8943716.0
4                    4400124.0              8964160.0

   wat_sm_number_without_urban  wat_sm_without_rural  wat_sm_without  \
0                    4025580.5             41.653310       26.279282
1                    4151897.0             40.724773       25.817177
2                    4284092.5             39.790870       25.372314
3                    4421790.0             38.851600       24.943192
4                    4564035.5             37.906970       24.530083

   wat_sm_without_urban
0             18.126915
1             18.170654
2             18.214600
3             18.258750
4             18.303108

[5 rows x 206 columns]
```

In [216…
```python
data.head(25)
```

Out[216…

	country	year	usage_of_improved_drinking_water_sources	usage_of_basic_drinking_water_services	usage_of_limited_drinking_water_services	usage_of_unimproved_drinking_water_sources	no_usage_of_dr
0	Algeria	2007	97.0	18.0	5.0	3.0	
1	Algeria	2008	97.0	18.0	5.0	3.0	
2	Algeria	2009	97.0	17.0	5.0	2.0	
3	Algeria	2010	98.0	17.0	5.0	2.0	
4	Algeria	2011	98.0	17.0	5.0	2.0	
5	Algeria	2012	98.0	17.0	5.0	2.0	
6	Algeria	2013	98.0	17.0	5.0	2.0	
7	Algeria	2014	98.0	17.0	5.0	1.0	
8	Algeria	2015	99.0	16.0	5.0	1.0	
9	Algeria	2016	99.0	17.0	5.0	1.0	
10	Algeria	2017	99.0	18.0	5.0	1.0	
11	Algeria	2018	99.0	19.0	5.0	1.0	
12	Algeria	2019	99.0	20.0	5.0	1.0	
13	Algeria	2020	99.0	22.0	5.0	0.0	
14	Algeria	2021	100.0	24.0	5.0	0.0	
15	Algeria	2022	100.0	24.0	5.0	0.0	
16	Bangladesh	2005	97.0	40.0	1.0	1.0	
17	Bangladesh	2006	97.0	41.0	1.0	1.0	
18	Bangladesh	2007	97.0	41.0	1.0	1.0	
19	Bangladesh	2008	97.0	41.0	1.0	1.0	
20	Bangladesh	2009	97.0	41.0	1.0	1.0	
21	Bangladesh	2010	98.0	42.0	1.0	1.0	
22	Bangladesh	2011	98.0	42.0	1.0	1.0	
23	Bangladesh	2012	98.0	42.0	1.0	1.0	
24	Bangladesh	2013	98.0	41.0	1.0	1.0	

25 rows × 206 columns

In [218…
```python
# Training model
from sklearn.ensemble import RandomForestRegressor

model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

Out[218…

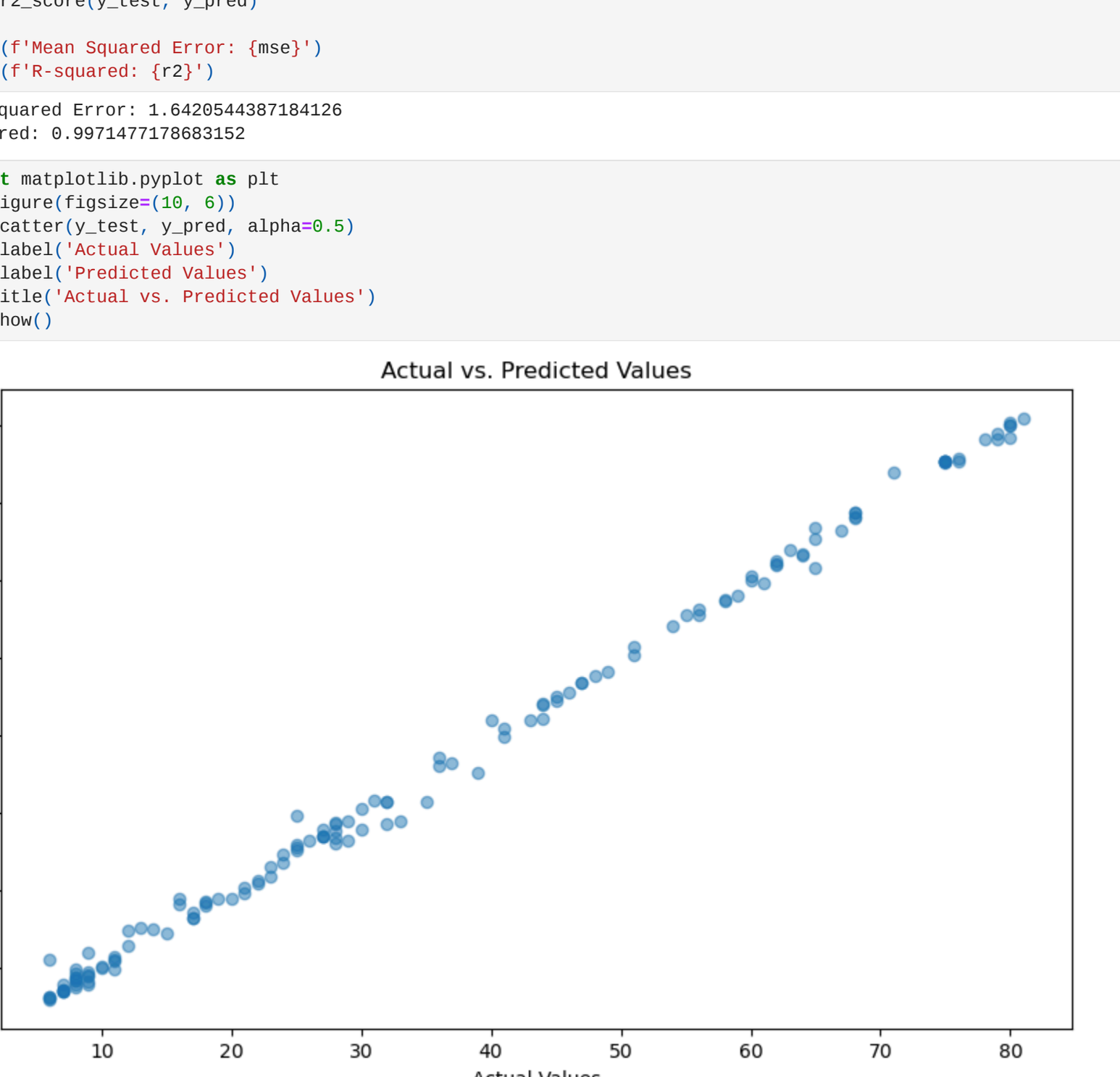
```
RandomForestRegressor
RandomForestRegressor(random_state=42)
```

In [220…
```python
y_pred = model.predict(X_test)
```

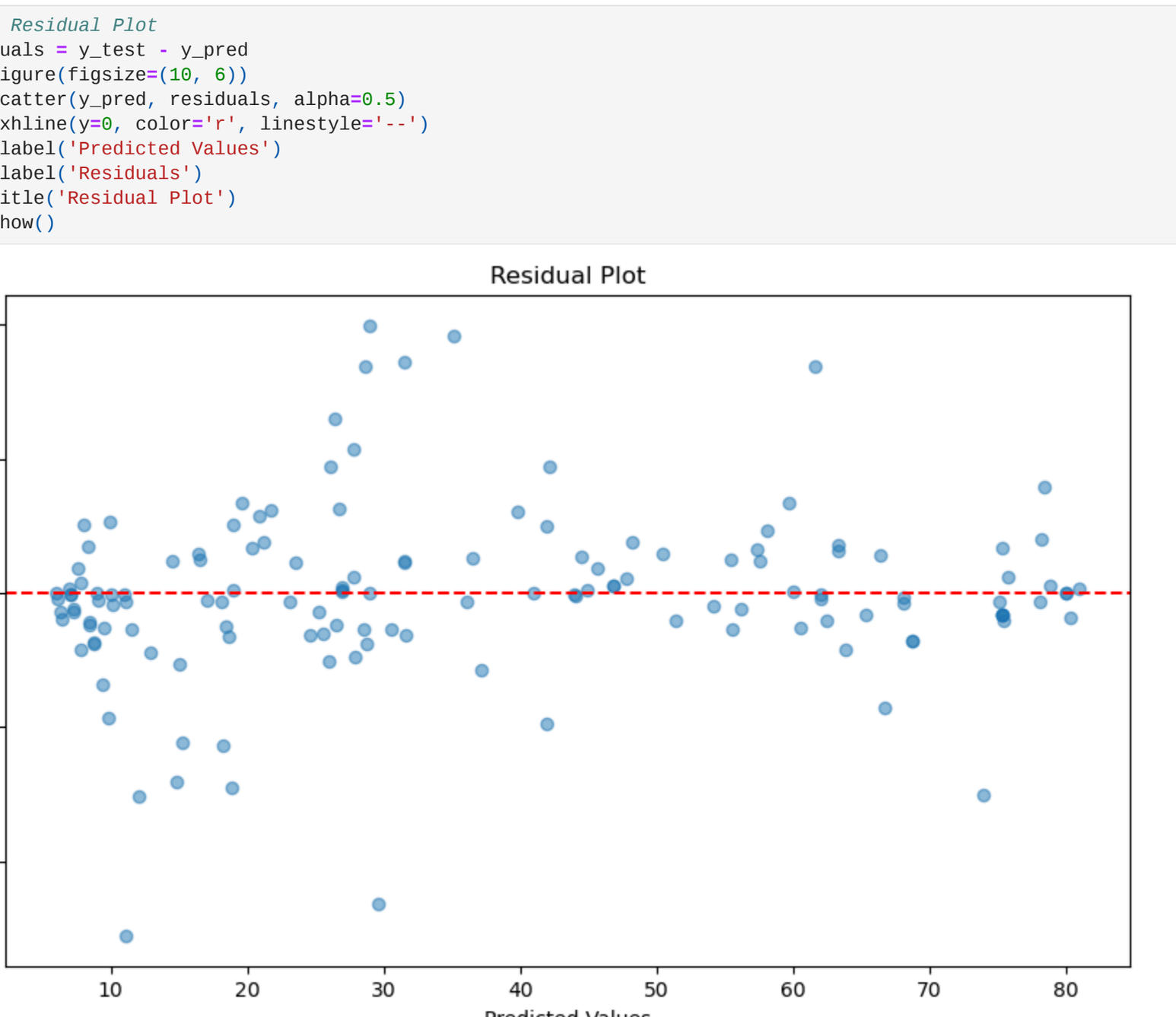
In [222…
```python
# Evaluating the model's performance
from sklearn.metrics import mean_squared_error, r2_score

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

print(f'Mean Squared Error: {mse}')
print(f'R-squared: {r2}')
```

```
Mean Squared Error: 1.6420544387184126
R-squared: 0.9971477178683152
```

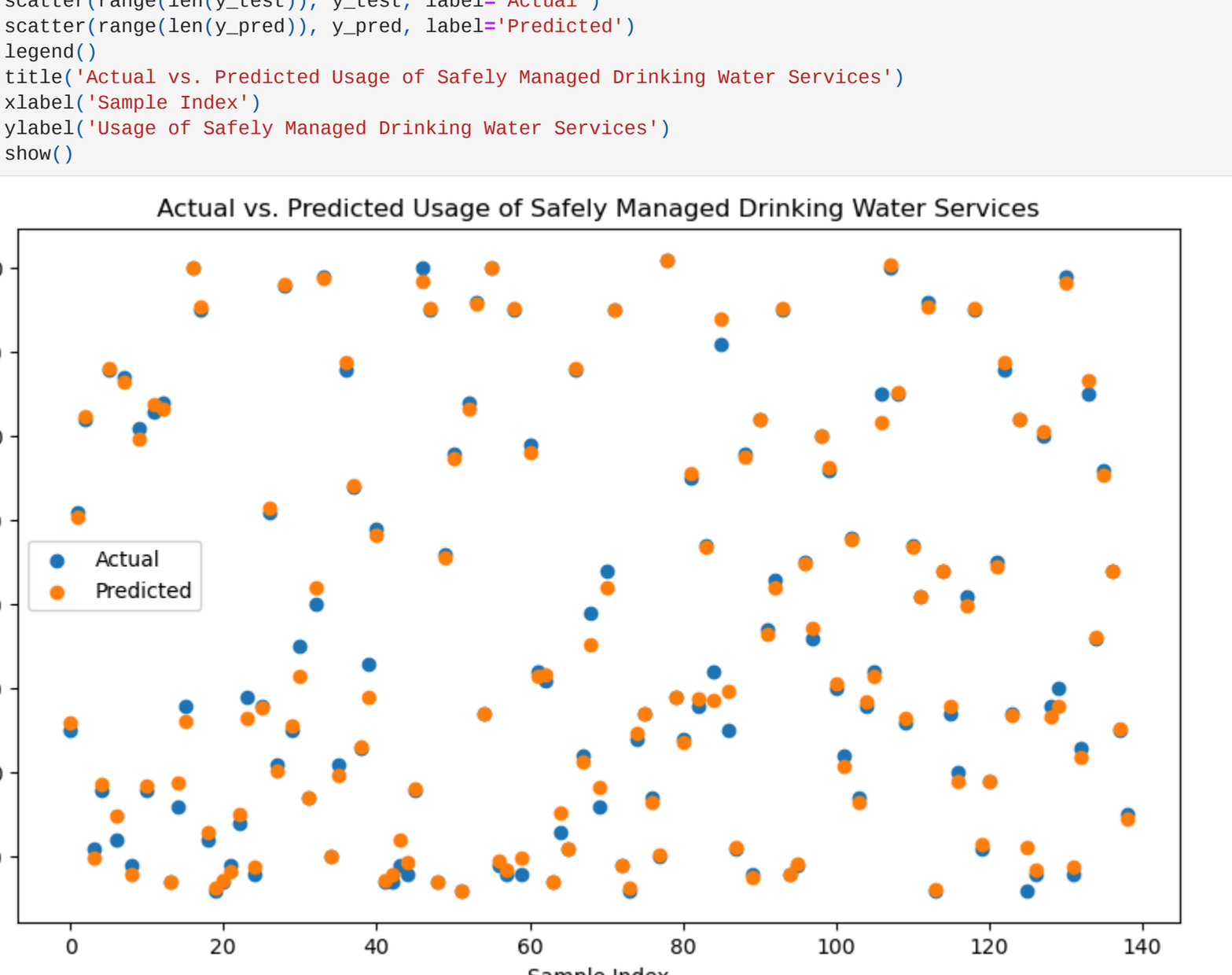
In [226…
```python
import matplotlib.pyplot as plt
plt.figure(figsize=(10, 6))
plt.scatter(y_test, y_pred, alpha=0.5)
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.title('Actual vs. Predicted Values')
plt.show()
```



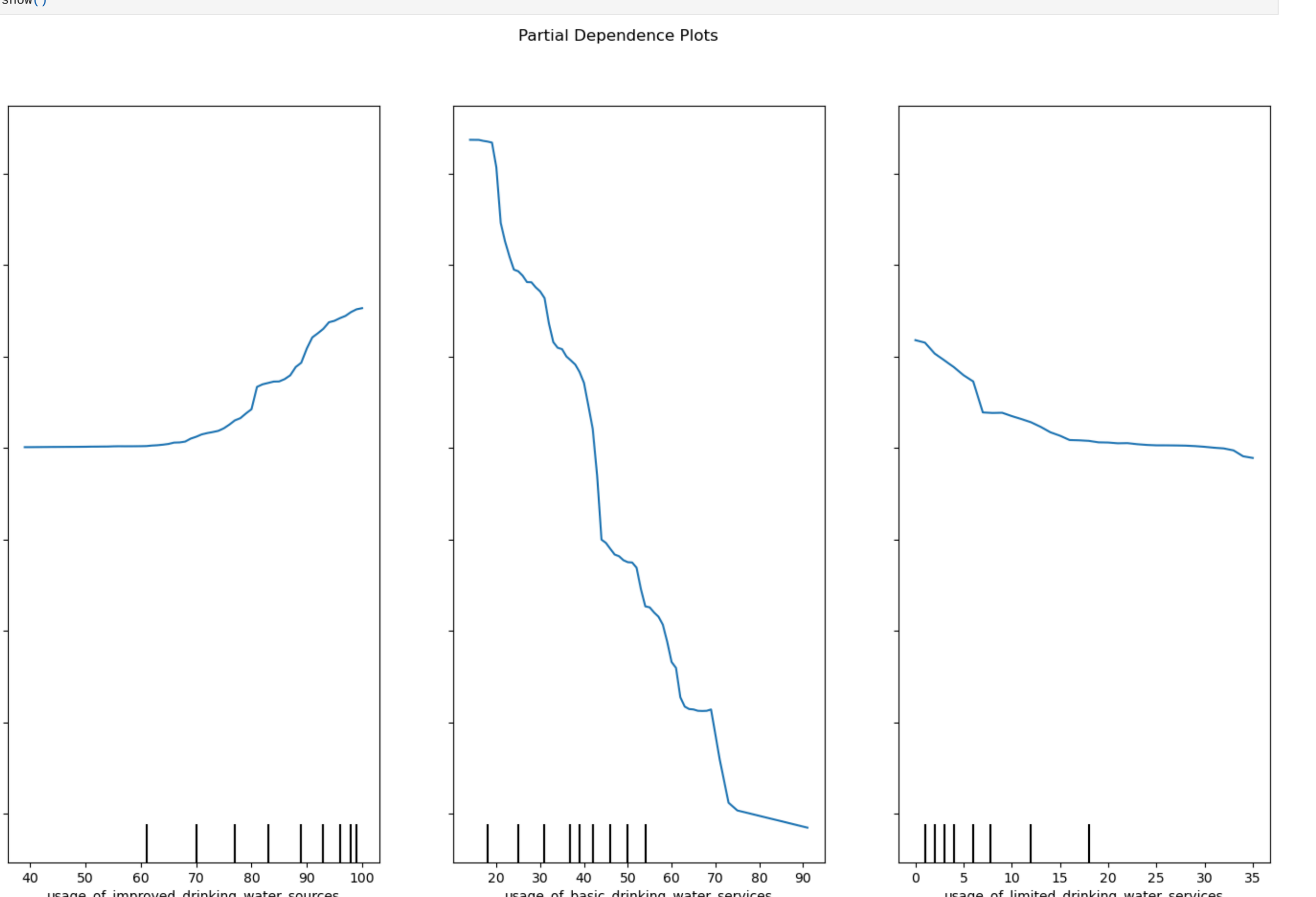
In [228…
```python
#Plot Residual Plot
residuals = y_test - y_pred
plt.figure(figsize=(10, 6))
plt.scatter(y_pred, residuals, alpha=0.5)
plt.axhline(y=0, color='r', linestyle='--')
plt.xlabel('Predicted Values')
plt.ylabel('Residuals')
plt.title('Residual Plot')
plt.show()
```



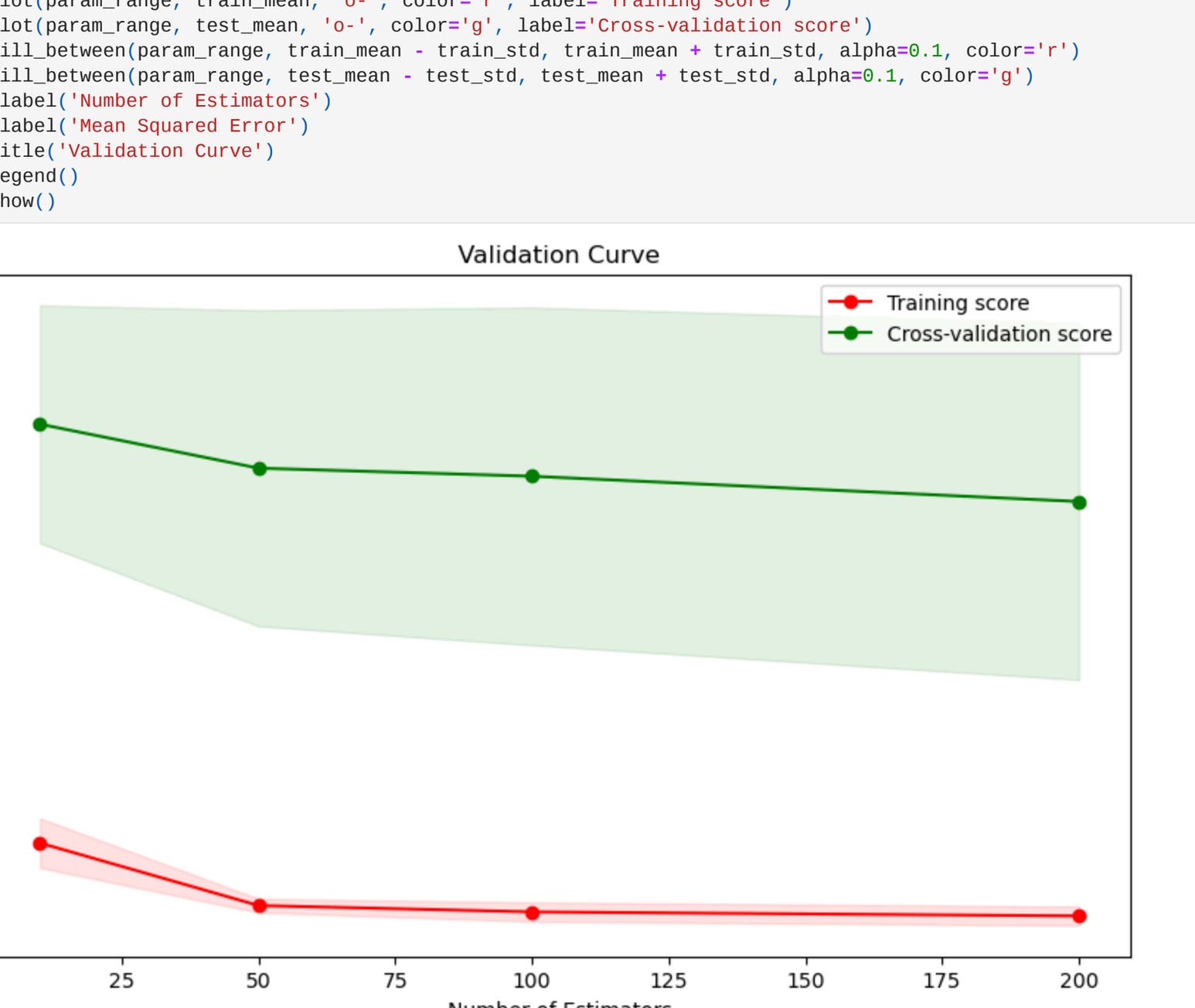
In [230…
```python
# Creating a dataframe for actual vs predicted values
predictions = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
# Plot actual vs. predicted values
plt.figure(figsize=(10, 6))
plt.scatter(range(len(y_test)), y_test, label='Actual')
plt.scatter(range(len(y_pred)), y_pred, label='Predicted')
plt.legend()
plt.title('Actual vs. Predicted Usage of Safely Managed Drinking Water Services')
plt.xlabel('Sample Index')
plt.ylabel('Usage of Safely Managed Drinking Water Services')
plt.show()
```



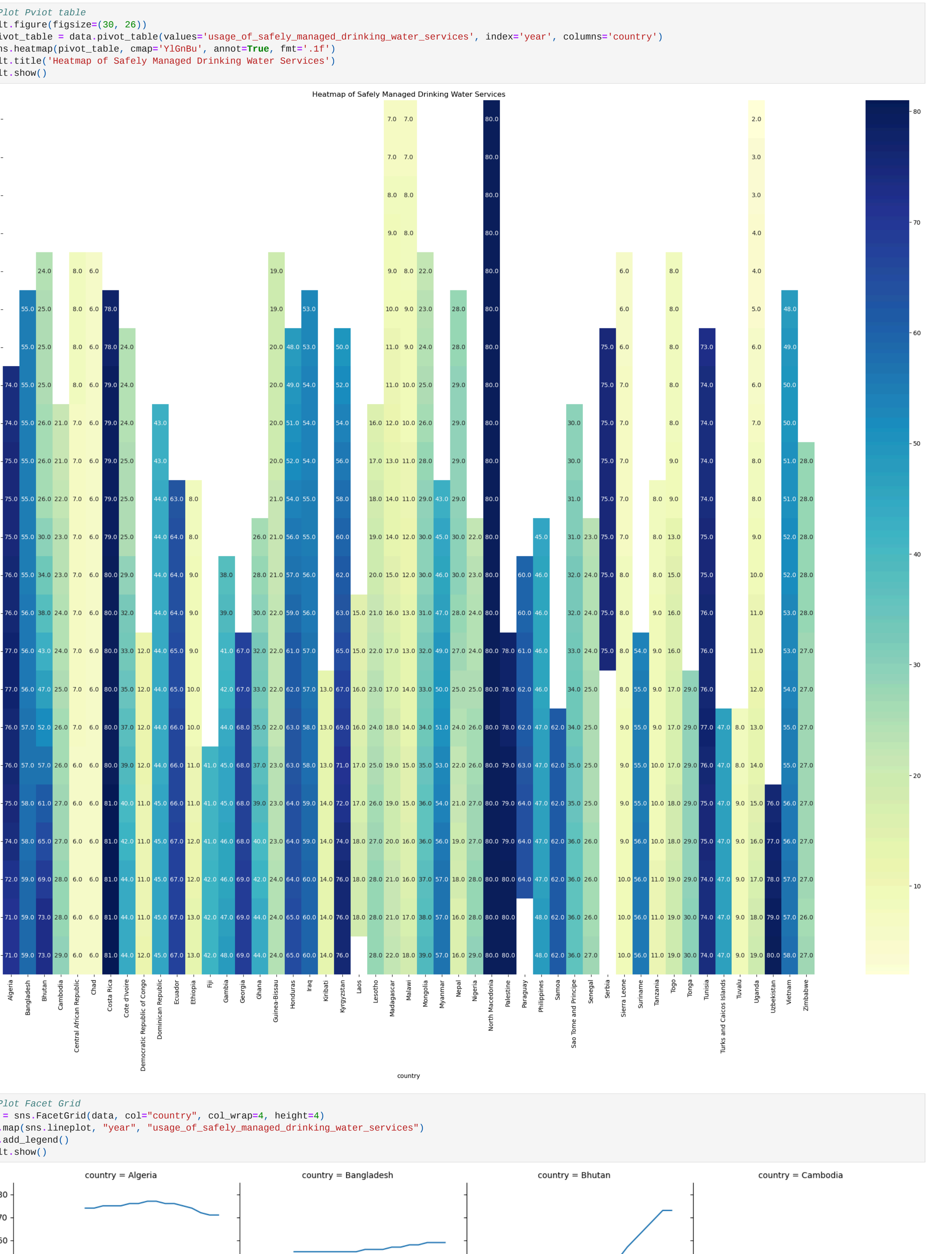
In [232…
```python
from sklearn.inspection import PartialDependenceDisplay
# Initializing the PartialDependenceDisplay
fig, ax = plt.subplots(figsize=(17, 10))
display = PartialDependenceDisplay.from_estimator(model, X_train, features=[0, 1, 2], ax=ax)
plt.suptitle('Partial Dependence Plots')
plt.subplots_adjust(top=0.9)
plt.show()
```



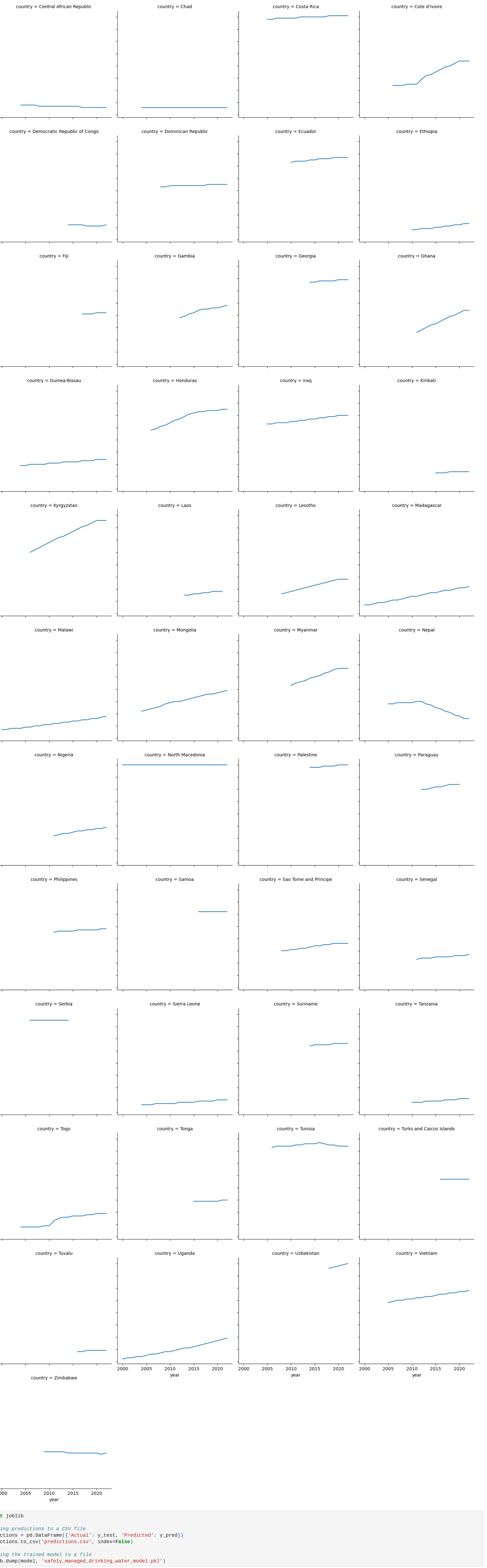
In [234…
```python
from sklearn.model_selection import validation_curve

# Computing validation curve
param_range = [10, 50, 100, 200]
train_scores, test_scores = validation_curve(model, X_train, y_train, param_name='n_estimators', param_range=param_range, cv=5, scoring='neg_mean_squared_error')

# Computing mean and standard deviation of training and testing scores
train_mean = -train_scores.mean(axis=1)
test_mean = -test_scores.mean(axis=1)
train_std = train_scores.std(axis=1)
test_std = test_scores.std(axis=1)

# Plots validation curve
plt.figure(figsize=(10, 6))
plt.plot(param_range, train_mean, 'o-', color='r', label='Training score')
plt.plot(param_range, test_mean, 'o-', color='g', label='Cross-validation score')
plt.fill_between(param_range, train_mean - train_std, train_mean + train_std, alpha=0.1, color='r')
plt.fill_between(param_range, test_mean - test_std, test_mean + test_std, alpha=0.1, color='g')
plt.xlabel('Number of Estimators')
plt.ylabel('Mean Squared Error')
plt.title('Validation Curve')
plt.legend()
plt.show()
```



In [235…
```python
#Plot Pviot table
plt.figure(figsize=(30, 26))
pivot_table = data.pivot_table(values='usage_of_safely_managed_drinking_water_services', index='year', columns='country')
sns.heatmap(pivot_table, cmap='YlGnBu', annot=True, fmt='.1f')
plt.title('Heatmap of Safely Managed Drinking Water Services')
plt.show()
```



In [236…
```python
#Plot Facet Grid
g = sns.FacetGrid(data, col="country", col_wrap=4, height=4)
g.map(sns.lineplot, "year", "usage_of_safely_managed_drinking_water_services")
g.add_legend()
plt.show()
```



In [253…
```python
import joblib

# Saving predictions to a CSV file
predictions = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
predictions.to_csv('predictions.csv', index=False)

# Saving the trained model to a file
joblib.dump(model, 'safely_managed_drinking_water_model.pkl')
```

Out[253…
```
['safely_managed_drinking_water_model.pkl']
```

In [ ]: