In [13]:
```python
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
```

In [15]:
```python
df = pd.read_csv("C:/Users/gundr/Downloads/Civil_Engineering_Regression_Dataset.csv")
```

In [17]:
```python
X = df.drop(columns=["Project_ID", "Construction_Cost"])  # Exclude ID and target variable
Y = df["Construction_Cost"]
```

In [19]:
```python
X = sm.add_constant(X)
```

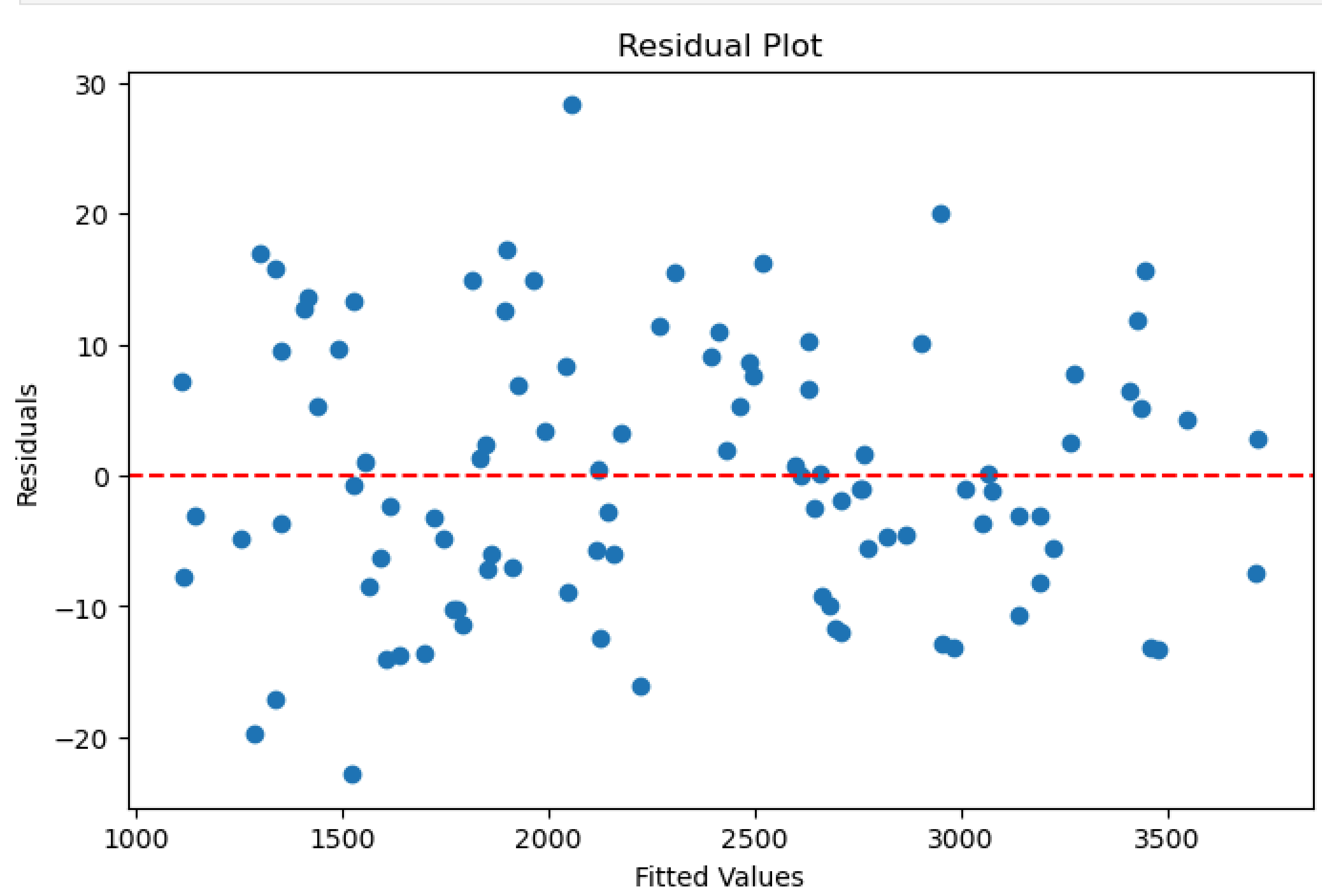
In [21]:
```python
model = sm.OLS(Y, X).fit()
```

In [23]:
```python
print(model.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:      Construction_Cost   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 7.593e+04
Date:                Wed, 12 Feb 2025   Prob (F-statistic):          2.94e-169
Time:                        22:36:00   Log-Likelihood:                -372.01
No. Observations:                 100   AIC:                             758.0
Df Residuals:                      93   BIC:                             776.3
Df Model:                           6                                         
Covariance Type:            nonrobust                                         
==========================================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                    -17.6374      6.972     -2.530      0.013     -31.482      -3.792
Building_Height           49.8823      0.080    622.215      0.000      49.723      50.042
Material_Quality_Index    10.7094      0.525     20.408      0.000       9.667      11.751
Labor_Cost                 0.5197      0.015     33.734      0.000       0.489       0.550
Concrete_Strength         20.3016      0.115    176.312      0.000      20.073      20.530
Foundation_Depth          29.9854      0.434     69.106      0.000      29.124      30.847
Weather_Index              0.4170      0.553      0.754      0.453      -0.681       1.515
==============================================================================
Omnibus:                        1.749   Durbin-Watson:                   1.764
Prob(Omnibus):                  0.417   Jarque-Bera (JB):                1.667
Skew:                           0.220   Prob(JB):                        0.435
Kurtosis:                       2.546   Cond. No.                     1.38e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.38e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [25]:
```python
plt.figure(figsize=(8,5))
plt.scatter(model.fittedvalues, model.resid)
plt.axhline(y=0, color='r', linestyle='--')
plt.xlabel("Fitted Values")
plt.ylabel("Residuals")
plt.title("Residual Plot")
plt.show()
```



In [27]:
```python
print("Model Interpretation & Conclusion")
print("\nKey Takeaways:")
print("1. Significant predictors identified: Building Height, Material Quality Index, Labor Cost, Concrete Strength, and Foundation Depth.")
print("2. High R-squared value suggests strong predictive power, but potential overfitting.")
print("3. Weather Index is not statistically significant.")
```

```
Model Interpretation & Conclusion

Key Takeaways:
1. Significant predictors identified: Building Height, Material Quality Index, Labor Cost, Concrete Strength, and Foundation Depth.
2. High R-squared value suggests strong predictive power, but potential overfitting.
3. Weather Index is not statistically significant.
```

In [29]:
```python
print("\nHow Construction Companies Can Use Regression Analysis:")
print("- Helps estimate project costs early in planning.")
print("- Allows for cost optimization by focusing on key variables.")
print("- Enables better budgeting and financial planning.")
```

```
How Construction Companies Can Use Regression Analysis:
- Helps estimate project costs early in planning.
- Allows for cost optimization by focusing on key variables.
- Enables better budgeting and financial planning.
```

In [31]:
```python
print("\nLimitations:")
print("- Possible overfitting due to high R-squared value.")
print("- Multicollinearity concerns among independent variables.")
print("- Missing factors like site location, equipment costs, and project duration.")
```

```
Limitations:
- Possible overfitting due to high R-squared value.
- Multicollinearity concerns among independent variables.
- Missing factors like site location, equipment costs, and project duration.
```

In [33]:
```python
print("\nPotential Model Improvements:")
print("- Include additional variables such as equipment cost and site location.")
print("- Apply feature selection techniques to avoid overfitting.")
print("- Use regularization methods like Ridge or Lasso regression.")
```

```
Potential Model Improvements:
- Include additional variables such as equipment cost and site location.
- Apply feature selection techniques to avoid overfitting.
- Use regularization methods like Ridge or Lasso regression.
```

In [35]:
```python
print("\nContribution of Regression Analysis to Civil Engineering:")
print("- Provides data-driven insights for cost estimation.")
print("- Enhances resource allocation and budgeting.")
print("- Improves decision-making for sustainable project management.")
```

```
Contribution of Regression Analysis to Civil Engineering:
- Provides data-driven insights for cost estimation.
- Enhances resource allocation and budgeting.
- Improves decision-making for sustainable project management.
```

In [ ]:
```python

```

In [38]:
```python
print("\nConclusion:")
print("Regression analysis plays a critical role in optimizing construction costs, enabling firms to plan efficiently and reduce unexpected expenses. Data science techniques help drive cost-effective planning, making construction projects
```

Conclusion:
Regression analysis plays a critical role in optimizing construction costs, enabling firms to plan efficiently and reduce unexpected expenses. Data science techniques help drive cost-effective planning, making construction projects more predictable and financially sound.

In [ ]:
```

```