In [145…
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.impute import SimpleImputer
```

In [147…
```
# Load the dataset
file_path_ce = "C:/Users/gundr/Downloads/Civil_Engineering_Regression_Dataset.csv"
df_ce = pd.read_csv(file_path_ce)
```

In [149…
```
df_ce.info()
print(df_ce.head())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 8 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Project_ID            100 non-null    int64
 1   Building_Height       100 non-null    float64
 2   Material_Quality_Index  100 non-null    int64
 3   Labor_Cost            100 non-null    float64
 4   Concrete_Strength     100 non-null    float64
 5   Foundation_Depth      100 non-null    float64
 6   Weather_Index         100 non-null    int64
 7   Construction_Cost     100 non-null    float64
dtypes: float64(5), int64(3)
memory usage: 6.4 KB
   Project_ID  Building_Height  Material_Quality_Index  Labor_Cost  \
0           1        21.854305                       9   70.213332
1           2        47.782144                       9  142.413614
2           3        37.939727                       3  110.539985
3           4        31.939632                       6  250.784939
4           5        12.020839                       7  167.575159

   Concrete_Strength  Foundation_Depth  Weather_Index  Construction_Cost
0          45.326394          8.804790              4        2400.287931
1          47.900505          6.727632              6        3705.461312
2          22.112484          8.208544              8        2653.631004
3          26.267562          7.094515              4        2534.099466
4          40.134306          6.160303              6        1741.179333
```

In [151…
```
independent_vars = ['Building_Height', 'Material_Quality_Index', 'Labor_Cost',
                    'Concrete_Strength', 'Foundation_Depth', 'Weather_Index']
dependent_var = 'Construction_Cost'

missing_values = df_ce.isnull().sum()
print("Missing Values:\n", missing_values)
```

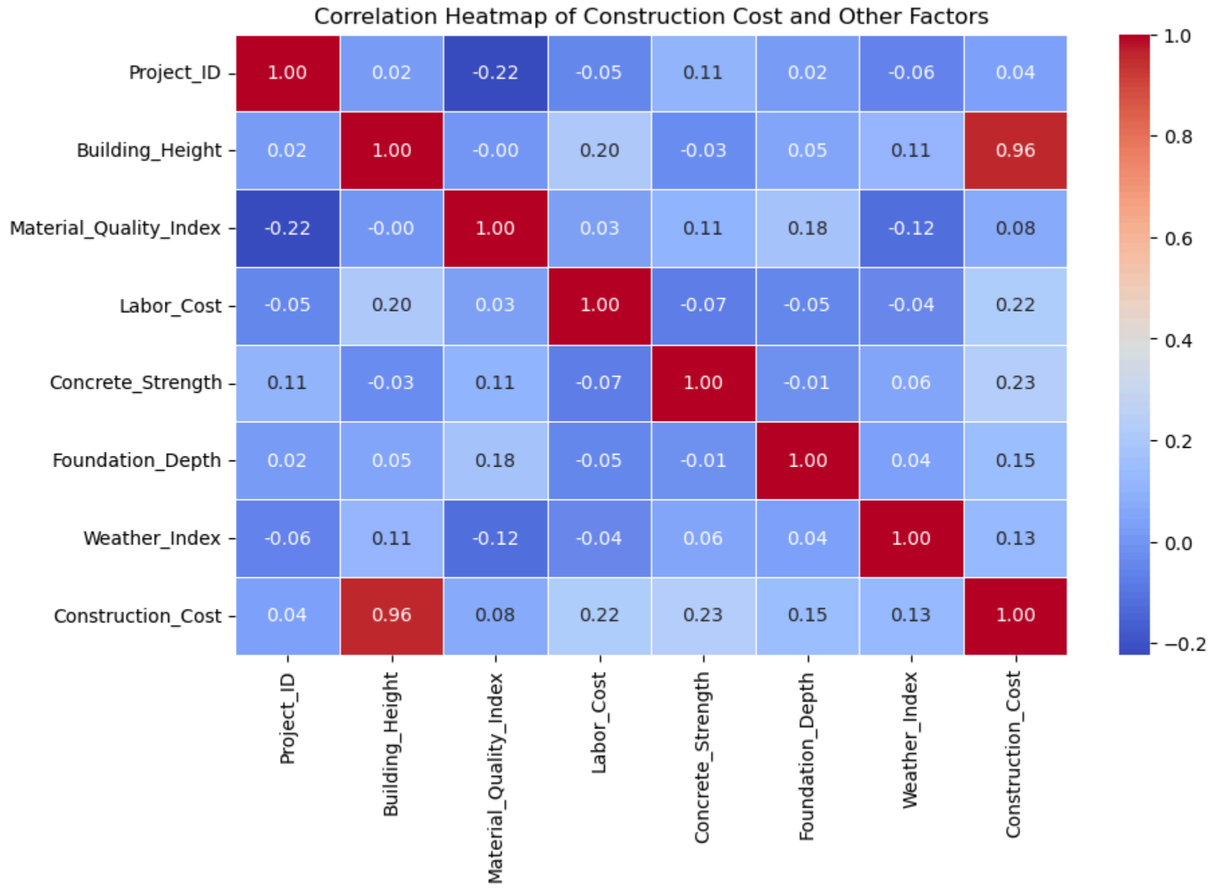
```
Missing Values:
 Project_ID              0
Building_Height          0
Material_Quality_Index   0
Labor_Cost               0
Concrete_Strength        0
Foundation_Depth         0
Weather_Index            0
Construction_Cost        0
dtype: int64
```

In [153…
```
summary_stats = df_ce.describe()
print("Summary Statistics:\n", summary_stats)
```

```
Summary Statistics:
        Project_ID  Building_Height  Material_Quality_Index  Labor_Cost  \
count  100.000000       100.000000              100.000000  100.000000
mean    50.500000        26.158133                5.940000  188.582366
std     29.011492        13.387023                2.048996   69.448489
min      1.000000         5.248495                3.000000   54.518841
25%     25.750000        13.694034                4.000000  130.371094
50%     50.500000        25.886410                6.000000  182.399916
75%     75.250000        37.859140                8.000000  251.182520
max    100.000000        49.409912                9.000000  299.063425

       Concrete_Strength  Foundation_Depth  Weather_Index  Construction_Cost
count         100.000000        100.000000     100.000000         100.000000
mean           33.624225          5.784751       5.900000        2307.354667
std             9.154865          2.450679       1.920122         702.491423
min            20.155546          1.350668       3.000000        1108.639036
25%            25.519003          3.898884       4.000000        1735.221415
50%            32.721430          5.957537       6.000000        2244.061942
75%            41.055857          7.639785       7.000000        2825.075251
max            49.455227          9.963381       9.000000        3723.127092
```

In [155…
```
plt.figure(figsize=(10, 6))
sns.heatmap(df_ce.corr(), annot=True, cmap="coolwarm", fmt=".2f", linewidths=0.5)
plt.title("Correlation Heatmap of Construction Cost and Other Factors")
plt.show()
```



In [164…
```
cleaned_file_path = "C:/Users/gundr/Downloads/Civil_Engineering_Regression_Dataset.csv"
df_ce.to_csv(cleaned_file_path, index=False)
```

In [166…
```
print("Cleaned dataset saved to:", cleaned_file_path)
```

```
Cleaned dataset saved to: C:/Users/gundr/Downloads/Civil_Engineering_Regression_Dataset.csv
```

In [ ]: