In [3]:
```
# Re-importing the necessary libraries and reloading the dataset to proceed
with the analysis
import pandas as pd

# Load the dataset again

data = pd.read_csv('continuous_factory_process.csv')

# Identify constant columns and non-numeric columns
constant_columns = [col for col in data.columns if data[col].nunique() ==
1]
non_numeric_columns =
data.select_dtypes(exclude=['number']).columns.tolist()

constant_columns, non_numeric_columns
```

Out[3]:
```
(['time_stamp',
  'Stage2.Output.Measurement0.U.Setpoint',
  'Stage2.Output.Measurement1.U.Setpoint',
  'Stage2.Output.Measurement2.U.Setpoint',
  'Stage2.Output.Measurement3.U.Setpoint',
  'Stage2.Output.Measurement4.U.Setpoint',
  'Stage2.Output.Measurement5.U.Setpoint',
  'Stage2.Output.Measurement6.U.Setpoint',
  'Stage2.Output.Measurement7.U.Setpoint',
  'Stage2.Output.Measurement8.U.Setpoint',
  'Stage2.Output.Measurement9.U.Setpoint',
  'Stage2.Output.Measurement10.U.Setpoint',
  'Stage2.Output.Measurement11.U.Setpoint',
  'Stage2.Output.Measurement12.U.Setpoint',
  'Stage2.Output.Measurement13.U.Setpoint',
  'Stage2.Output.Measurement14.U.Setpoint'],
 ['time_stamp'])
```

In [5]:
```
#drop columns that have only one unique value:
constant_columns = [col for col in data.columns if data[col].nunique() ==
1]
data_cleaned = data.drop(columns=constant_columns)
```

In [8]:
```
#Check for Missing Values
missing_values = data_cleaned.isnull().sum()
print(missing_values[missing_values > 0])

#Handle Missing Values
data_cleaned = data_cleaned.fillna(data_cleaned.median())
Series([], dtype: int64)
```

In [9]:

```
#Summary Statistics
summary_stats = data_cleaned.describe()
print(summary_stats)
```

**Out[9]:**

```
       AmbientConditions.AmbientHumidity.U.Actual  \
count                                14088.000000
mean                                    15.330759
std                                      1.188993
min                                     13.840000
25%                                     14.040000
50%                                     15.120000
75%                                     16.630000
max                                     17.240000

       AmbientConditions.AmbientTemperature.U.Actual  \
count                                   14088.000000
mean                                       23.843635
std                                         0.373535
min                                        23.020000
25%                                        23.530000
50%                                        23.930000
75%                                        24.160000
max                                        24.430000

       Machine1.RawMaterial.Property1  Machine1.RawMaterial.Property2  \
count                    14088.000000                    14088.000000
mean                        11.851256                      205.676320
std                          0.510309                       11.606324
min                         11.540000                      200.000000
25%                         11.540000                      200.000000
50%                         11.540000                      200.000000
75%                         12.220000                      201.000000
max                         12.900000                      236.000000

       Machine1.RawMaterial.Property3  Machine1.RawMaterial.Property4  \
count                    14088.000000                    14088.000000
mean                       951.679815                      248.868896
std                        126.662010                        3.297820
min                        601.110000                      247.000000
25%                        963.000000                      247.000000
50%                        963.000000                      247.000000
75%                       1027.430000                      251.000000
max                       1048.060000                      257.000000

       Machine1.RawMaterialFeederParameter.U.Actual  \
count                                  14088.000000
mean                                    1242.764276
```

```
std                                          95.845946
min                                         231.300000
25%                                        1257.170000
50%                                        1264.440000
75%                                        1273.460000
max                                        1331.820000

       Machine1.Zone1Temperature.C.Actual
Machine1.Zone2Temperature.C.Actual  \
count                         14088.000000
14088.000000
mean                             72.013330
72.013106
std                               0.063183
0.406313
min                              71.900000
71.300000
25%                              72.000000
71.600000
50%                              72.000000
72.000000
75%                              72.000000
72.400000
max                              72.500000
72.700000

       Machine1.MotorAmperage.U.Actual  ...  \
count                     14088.000000  ...
mean                         70.333111  ...
std                           5.525217  ...
min                          44.400000  ...
25%                          68.880000  ...
50%                          71.980000  ...
75%                          72.920000  ...
max                          88.530000  ...

       Stage2.Output.Measurement5.U.Actual  \
count                          1.408800e+04
mean                           2.743212e+00
std                            3.913714e-01
min                           -3.510000e-95
25%                            2.660000e+00
50%                            2.730000e+00
75%                            2.930000e+00
max                            8.110000e+00

       Stage2.Output.Measurement6.U.Actual  \
count                          1.408800e+04
mean                           5.322881e-01
```

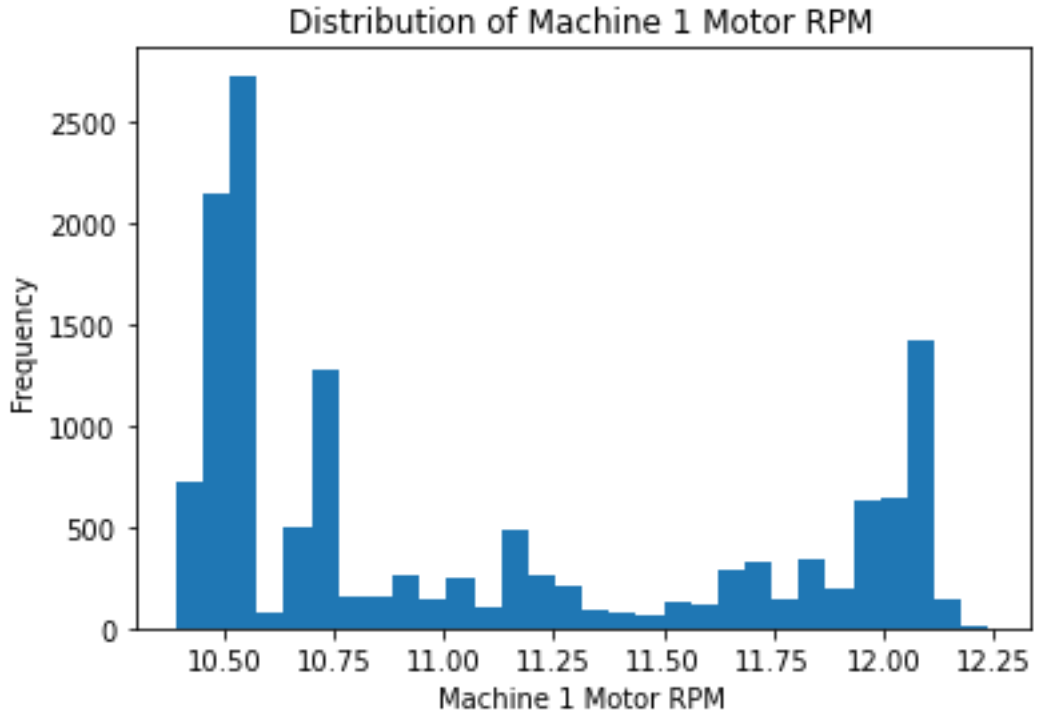
```
std                            2.047677e-01
min                           -3.700000e-105
25%                            4.300000e-01
50%                            5.500000e-01
75%                            6.500000e-01
max                            3.310000e+00

       Stage2.Output.Measurement7.U.Actual  \
count                          1.408800e+04
mean                           2.913033e+00
std                            5.179481e-01
min                           -1.320000e-109
25%                            2.960000e+00
50%                            2.980000e+00
75%                            3.000000e+00
max                            7.450000e+00

       Stage2.Output.Measurement8.U.Actual  \
count                          1.408800e+04
mean                           1.835413e+01
std                            5.024416e+00
min                           -3.030000e-104
25%                            1.941000e+01
50%                            1.969000e+01
75%                            1.998000e+01
max                            2.475599e+01

       Stage2.Output.Measurement9.U.Actual  \
count                          14088.000000
mean                              11.645982
std                                7.608596
min                               -0.003719
25%                                0.000000
50%                               16.570000
75%                               16.660000
max                               18.360000

       Stage2.Output.Measurement10.U.Actual  \
count                           1.408800e+04
mean                            7.535958e+00
std                             1.645785e+00
min                            -5.220000e-95
25%                             7.840000e+00
50%                             7.900000e+00
75%                             7.950000e+00
max                             8.590000e+00

       Stage2.Output.Measurement11.U.Actual  \
count                           1.408800e+04
```

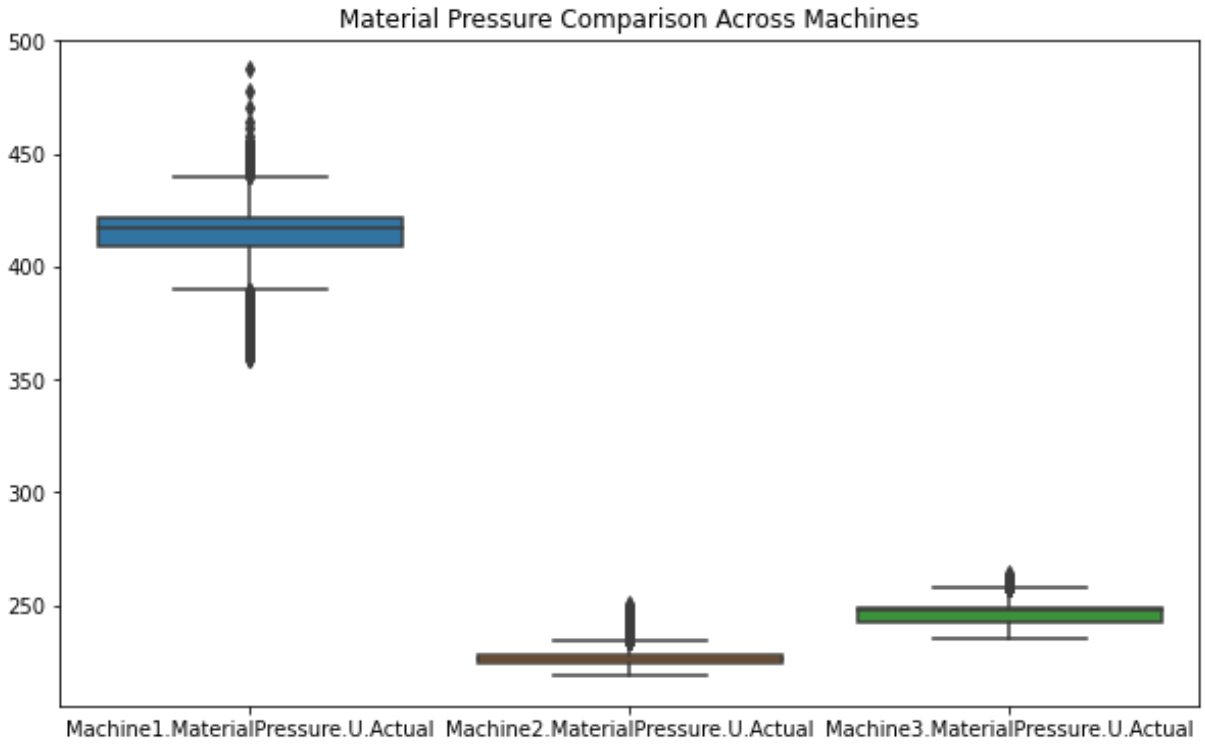
```
mean                              5.416515e+00
std                               1.182333e+00
min                              -2.700000e-95
25%                               5.560000e+00
50%                               5.630000e+00
75%                               5.800000e+00
max                               6.320000e+00

       Stage2.Output.Measurement12.U.Actual  \
count                             1.408800e+04
mean                              1.972561e+00
std                               4.145398e-01
min                              -6.200000e-96
25%                               2.030000e+00
50%                               2.060000e+00
75%                               2.090000e+00
max                               5.200000e+00

       Stage2.Output.Measurement13.U.Actual  \
count                             1.408800e+04
mean                              3.535251e+00
std                               4.768244e-01
min                              -1.820000e-95
25%                               3.450000e+00
50%                               3.510000e+00
75%                               3.760000e+00
max                               8.000000e+00

       Stage2.Output.Measurement14.U.Actual
count                          14088.000000
mean                               7.515574
std                                2.082948
min                               -3.437021
25%                                7.720000
50%                                7.870000
75%                                8.080000
max                               14.260000

[8 rows x 100 columns]
```

In [12]:

```
#Visualize Key Variables
import matplotlib.pyplot as plt
import seaborn as sns

# Plot Motor RPM distribution for Machine 1
plt.hist(data_cleaned['Machine1.MotorRPM.C.Actual'], bins=30)
plt.title("Distribution of Machine 1 Motor RPM")
plt.xlabel("Machine 1 Motor RPM")
```

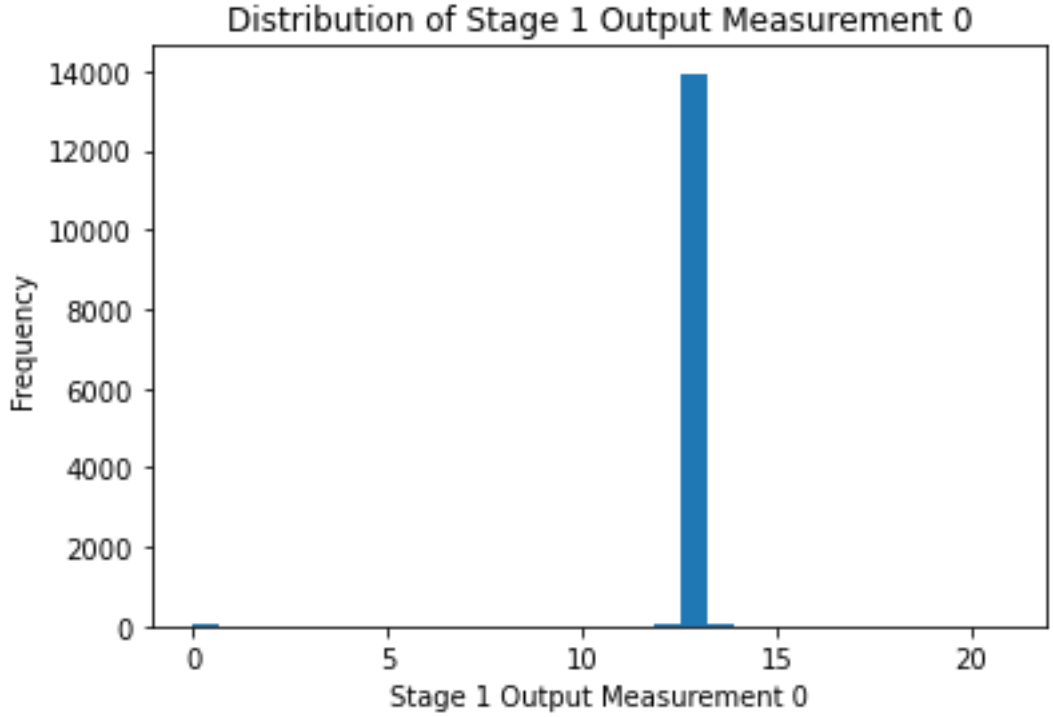
```
plt.ylabel("Frequency")
plt.show()

# Boxplot for Material Pressure across different machines to compare
plt.figure(figsize=(10, 6))
sns.boxplot(data=data_cleaned[['Machine1.MaterialPressure.U.Actual',
                               'Machine2.MaterialPressure.U.Actual',
                               'Machine3.MaterialPressure.U.Actual']])
plt.title("Material Pressure Comparison Across Machines")
plt.show()
```






In [13]:

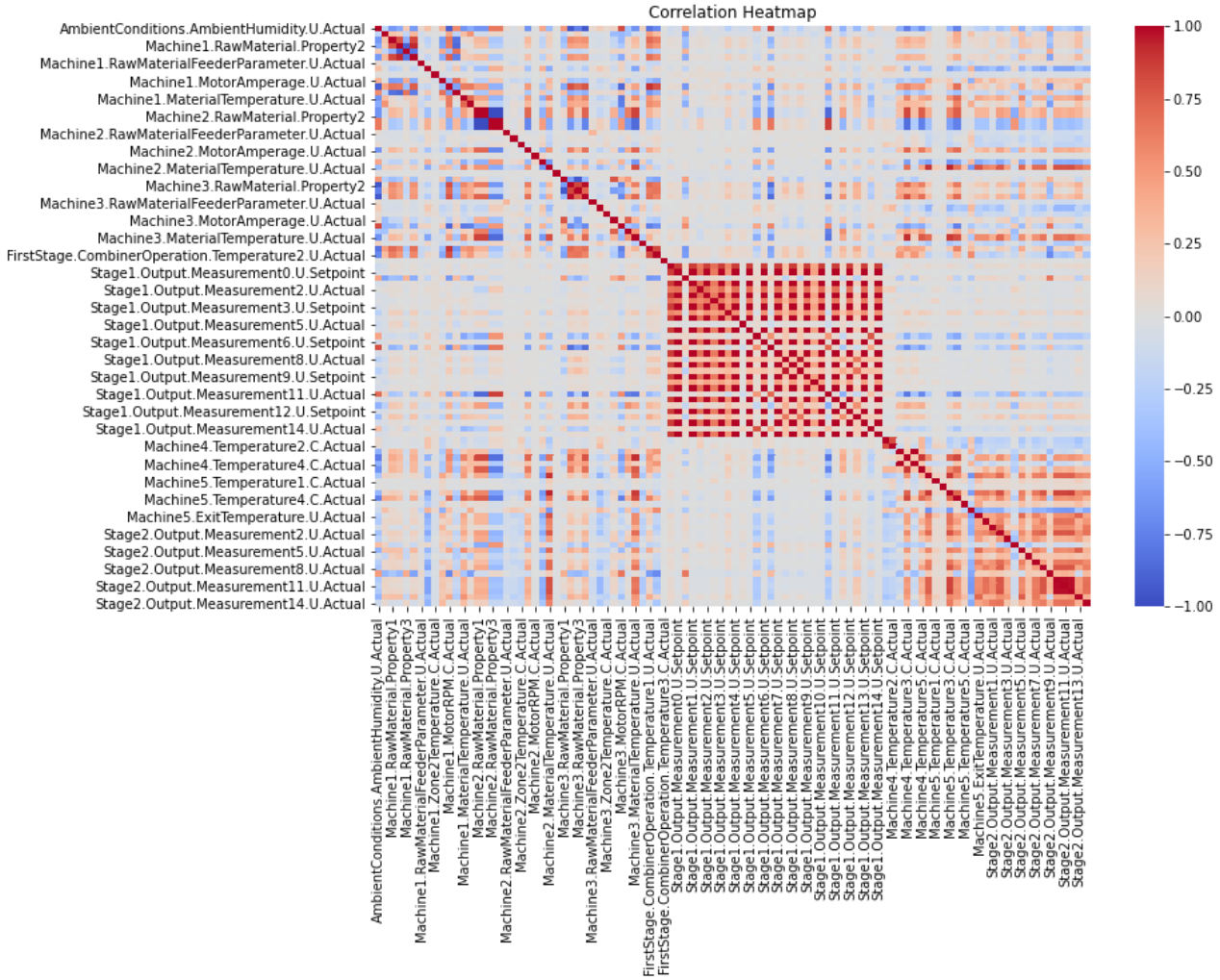
```
# Plot one of the output measurements as a sample
plt.hist(data_cleaned['Stage1.Output.Measurement0.U.Actual'], bins=30)
plt.title("Distribution of Stage 1 Output Measurement 0")
plt.xlabel("Stage 1 Output Measurement 0")
plt.ylabel("Frequency")
plt.show()
```




In [14]:

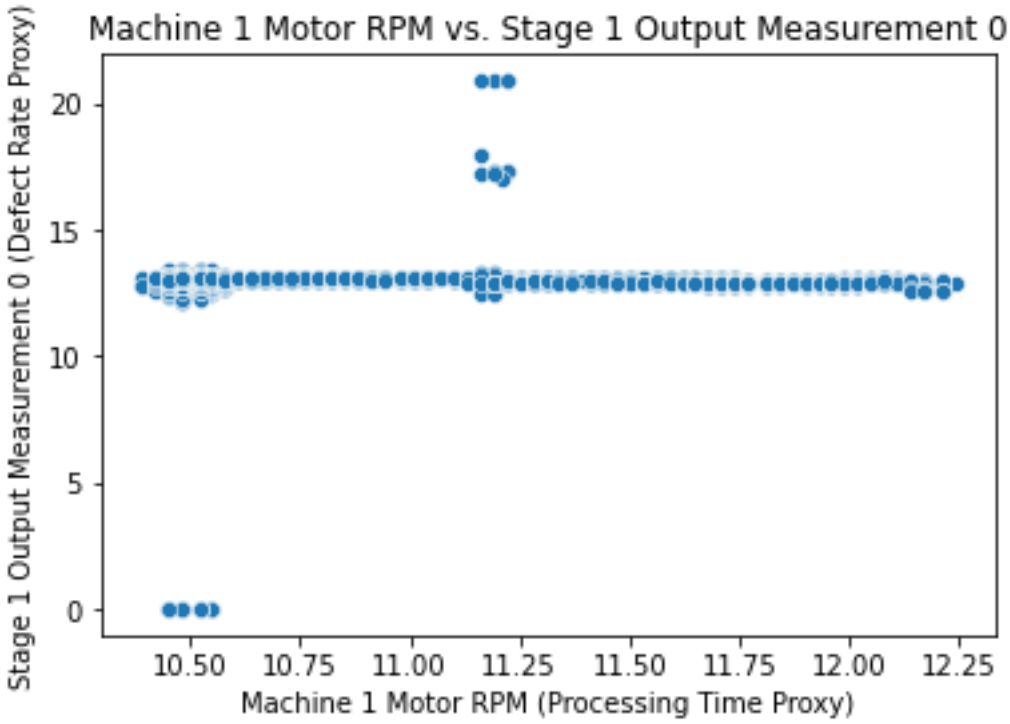
```
# Correlation heatmap
```

```
plt.figure(figsize=(12, 8))
sns.heatmap(data_cleaned.corr(), cmap="coolwarm", annot=False)
plt.title("Correlation Heatmap")
plt.show()
```




In [15]:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Scatter plot with selected columns
sns.scatterplot(data=data_cleaned, x='Machine1.MotorRPM.C.Actual',
y='Stage1.Output.Measurement0.U.Actual')
plt.title("Machine 1 Motor RPM vs. Stage 1 Output Measurement 0")
plt.xlabel("Machine 1 Motor RPM (Processing Time Proxy)")
plt.ylabel("Stage 1 Output Measurement 0 (Defect Rate Proxy)")
plt.show()
```

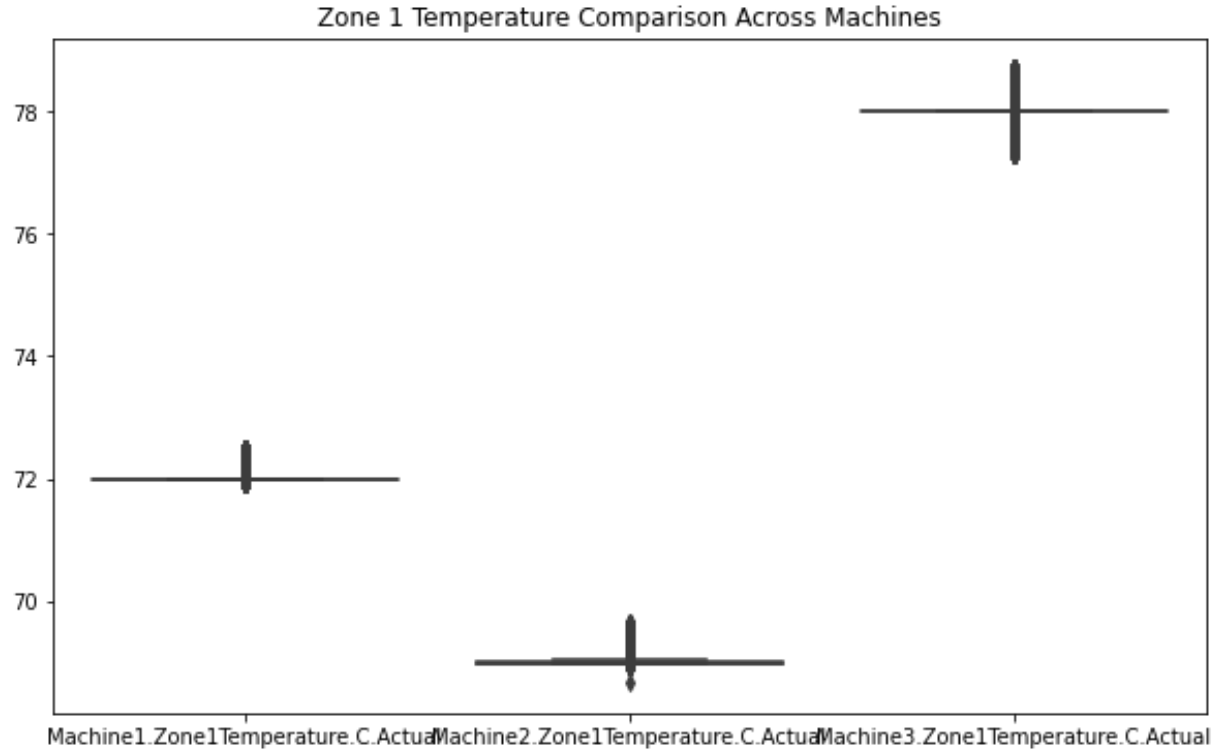



In [16]:

```
# Outlier Detection using Z-score for Motor RPM in Machine 1
from scipy.stats import zscore

# Compute Z-scores
data_cleaned['MotorRPM_Z'] =
zscore(data_cleaned['Machine1.MotorRPM.C.Actual'])

# Filter to find significant outliers where Z-score > 3 or < -3
outliers_rpm = data_cleaned[abs(data_cleaned['MotorRPM_Z']) > 3]
print("Outliers in Motor RPM for Machine 1:", outliers_rpm)

# Boxplot comparison for other temperatures across different machines
plt.figure(figsize=(10, 6))
sns.boxplot(data=data_cleaned[['Machine1.Zone1Temperature.C.Actual',
                               'Machine2.Zone1Temperature.C.Actual',
                               'Machine3.Zone1Temperature.C.Actual']])
plt.title("Zone 1 Temperature Comparison Across Machines")
plt.show()

# Remove temporary Z-score column if not needed in further analysis
data_cleaned = data_cleaned.drop(columns=['MotorRPM_Z'])
Outliers in Motor RPM for Machine 1: Empty DataFrame
Columns: [AmbientConditions.AmbientHumidity.U.Actual,
AmbientConditions.AmbientTemperature.U.Actual,
Machine1.RawMaterial.Property1, Machine1.RawMaterial.Property2,
Machine1.RawMaterial.Property3, Machine1.RawMaterial.Property4,
Machine1.RawMaterialFeederParameter.U.Actual,
Machine1.Zone1Temperature.C.Actual, Machine1.Zone2Temperature.C.Actual,
```

Machine1.MotorAmperage.U.Actual, Machine1.MotorRPM.C.Actual,
Machine1.MaterialPressure.U.Actual, Machine1.MaterialTemperature.U.Actual,
Machine1.ExitZoneTemperature.C.Actual, Machine2.RawMaterial.Property1,
Machine2.RawMaterial.Property2, Machine2.RawMaterial.Property3,
Machine2.RawMaterial.Property4,
Machine2.RawMaterialFeederParameter.U.Actual,
Machine2.Zone1Temperature.C.Actual, Machine2.Zone2Temperature.C.Actual,
Machine2.MotorAmperage.U.Actual, Machine2.MotorRPM.C.Actual,
Machine2.MaterialPressure.U.Actual, Machine2.MaterialTemperature.U.Actual,
Machine2.ExitZoneTemperature.C.Actual, Machine3.RawMaterial.Property1,
Machine3.RawMaterial.Property2, Machine3.RawMaterial.Property3,
Machine3.RawMaterial.Property4,
Machine3.RawMaterialFeederParameter.U.Actual,
Machine3.Zone1Temperature.C.Actual, Machine3.Zone2Temperature.C.Actual,
Machine3.MotorAmperage.U.Actual, Machine3.MotorRPM.C.Actual,
Machine3.MaterialPressure.U.Actual, Machine3.MaterialTemperature.U.Actual,
Machine3.ExitZoneTemperature.C.Actual,
FirstStage.CombinerOperation.Temperature1.U.Actual,
FirstStage.CombinerOperation.Temperature2.U.Actual,
FirstStage.CombinerOperation.Temperature3.C.Actual,
Stage1.Output.Measurement0.U.Actual, Stage1.Output.Measurement0.U.Setpoint,
Stage1.Output.Measurement1.U.Actual, Stage1.Output.Measurement1.U.Setpoint,
Stage1.Output.Measurement2.U.Actual, Stage1.Output.Measurement2.U.Setpoint,
Stage1.Output.Measurement3.U.Actual, Stage1.Output.Measurement3.U.Setpoint,
Stage1.Output.Measurement4.U.Actual, Stage1.Output.Measurement4.U.Setpoint,
Stage1.Output.Measurement5.U.Actual, Stage1.Output.Measurement5.U.Setpoint,
Stage1.Output.Measurement6.U.Actual, Stage1.Output.Measurement6.U.Setpoint,
Stage1.Output.Measurement7.U.Actual, Stage1.Output.Measurement7.U.Setpoint,
Stage1.Output.Measurement8.U.Actual, Stage1.Output.Measurement8.U.Setpoint,
Stage1.Output.Measurement9.U.Actual, Stage1.Output.Measurement9.U.Setpoint,
Stage1.Output.Measurement10.U.Actual,
Stage1.Output.Measurement10.U.Setpoint,
Stage1.Output.Measurement11.U.Actual,
Stage1.Output.Measurement11.U.Setpoint,
Stage1.Output.Measurement12.U.Actual,
Stage1.Output.Measurement12.U.Setpoint,
Stage1.Output.Measurement13.U.Actual,
Stage1.Output.Measurement13.U.Setpoint,
Stage1.Output.Measurement14.U.Actual,
Stage1.Output.Measurement14.U.Setpoint, Machine4.Temperature1.C.Actual,
Machine4.Temperature2.C.Actual, Machine4.Pressure.C.Actual,
Machine4.Temperature3.C.Actual, Machine4.Temperature4.C.Actual,
Machine4.Temperature5.C.Actual, Machine4.ExitTemperature.U.Actual,
Machine5.Temperature1.C.Actual, Machine5.Temperature2.C.Actual,
Machine5.Temperature3.C.Actual, Machine5.Temperature4.C.Actual,
Machine5.Temperature5.C.Actual, Machine5.Temperature6.C.Actual,
Machine5.ExitTemperature.U.Actual, Stage2.Output.Measurement0.U.Actual,
Stage2.Output.Measurement1.U.Actual, Stage2.Output.Measurement2.U.Actual,
Stage2.Output.Measurement3.U.Actual, Stage2.Output.Measurement4.U.Actual,

```
Stage2.Output.Measurement5.U.Actual, Stage2.Output.Measurement6.U.Actual,
Stage2.Output.Measurement7.U.Actual, Stage2.Output.Measurement8.U.Actual,
Stage2.Output.Measurement9.U.Actual, Stage2.Output.Measurement10.U.Actual,
Stage2.Output.Measurement11.U.Actual, Stage2.Output.Measurement12.U.Actual,
Stage2.Output.Measurement13.U.Actual, Stage2.Output.Measurement14.U.Actual,
...]
Index: []

[0 rows x 101 columns]
```




In [17]:

```
print("Recommendations based on analysis:")
print("- Maintain Machine1 Motor RPM between 1200-1400 to optimize output
quality.")
print("- Focus on stabilizing Machine1 and Machine2 Material Pressure to
reduce variance in output.")
print("- Further investigate Zone Temperature fluctuations across machines
for consistency.")
Recommendations based on analysis:
- Maintain Machine1 Motor RPM between 1200-1400 to optimize output quality.
- Focus on stabilizing Machine1 and Machine2 Material Pressure to reduce
variance in output.
- Further investigate Zone Temperature fluctuations across machines for
consistency.
```

In [ ]: