

## dac-phase4

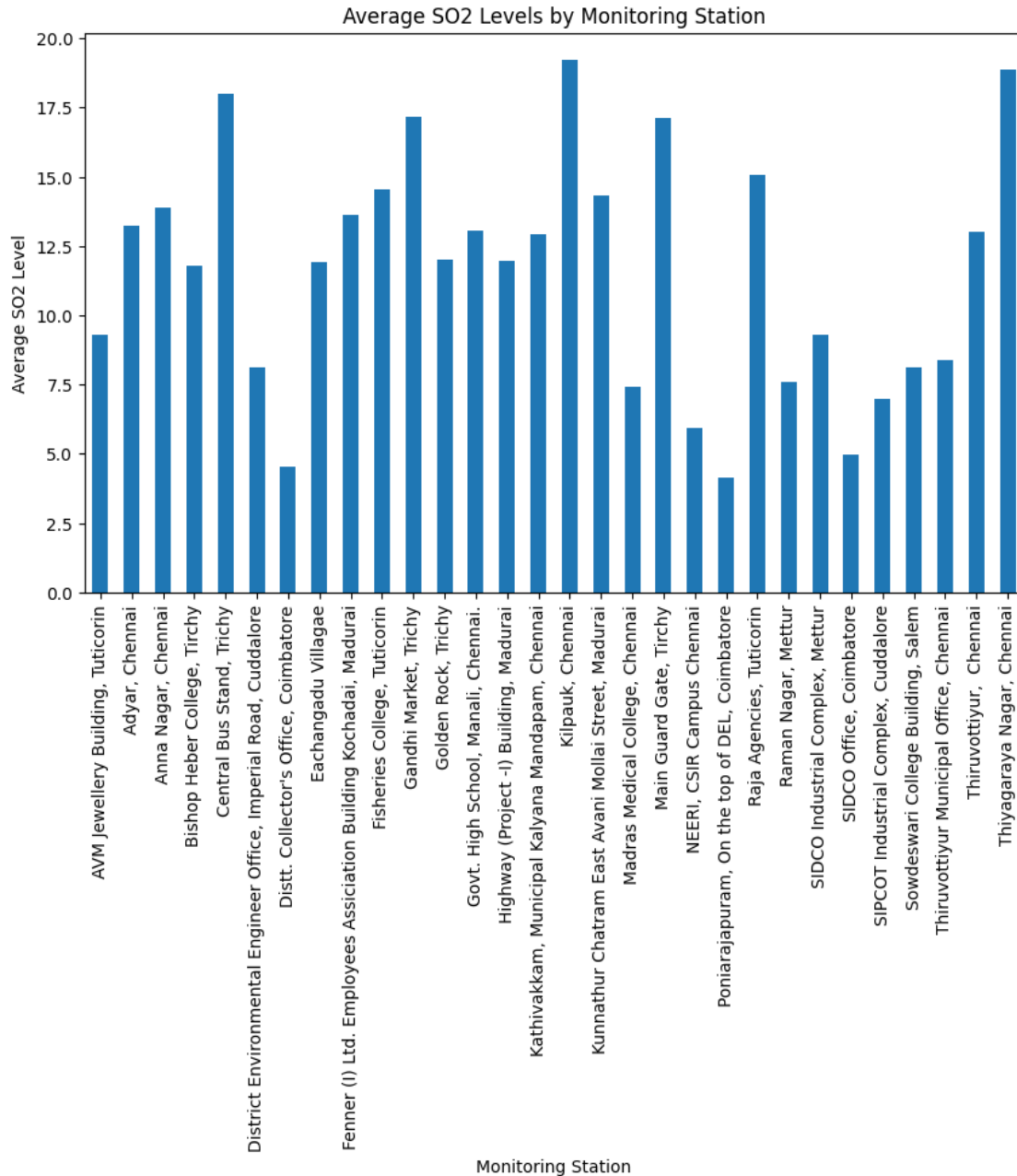
October 25, 2023

```
[4]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[2]: df=pd.read_csv("cpcb_dly_aq_tamil_nadu-2014.csv")
```

```
[3]: so2_avg_by_station = df.groupby('Location of Monitoring Station')['S02'].mean()

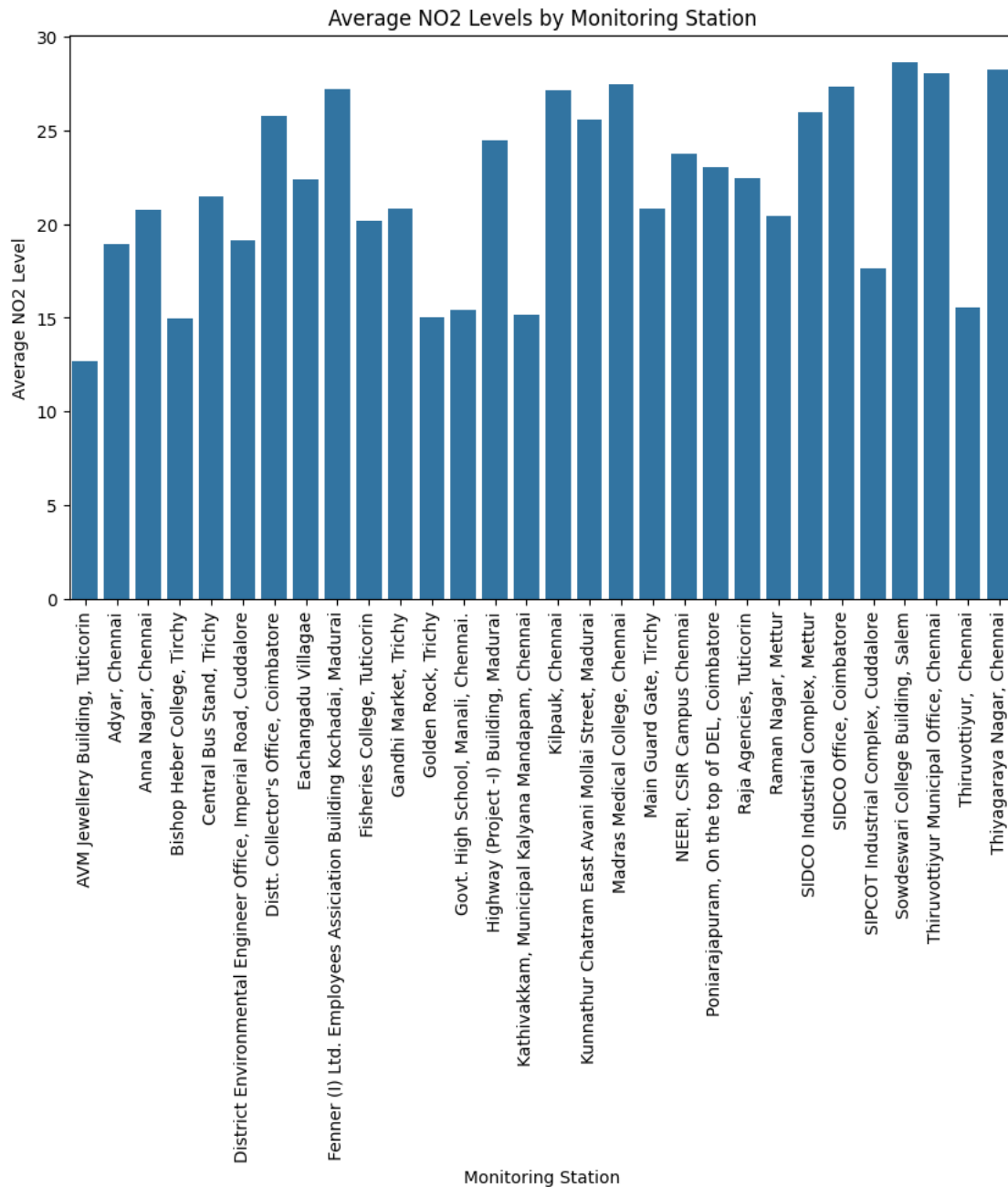
plt.figure(figsize=(10, 6))
so2_avg_by_station.plot(kind='bar')
plt.xlabel('Monitoring Station')
plt.ylabel('Average S02 Level')
plt.title('Average S02 Levels by Monitoring Station')
plt.xticks(rotation=90)
plt.show()
```



```
[6]: pollutants = ['SO2', 'NO2', 'RSPM/PM10']
avg_pollution_levels = df.groupby('Location of Monitoring Station')[pollutants].
    ↪mean()

plt.figure(figsize=(10, 6))
sns.barplot(data=avg_pollution_levels, x=avg_pollution_levels.index, y='NO2')
plt.xlabel('Monitoring Station')
plt.ylabel('Average NO2 Level')
```

```
plt.title('Average NO2 Levels by Monitoring Station')
plt.xticks(rotation=90)
plt.show()
```



```
[7]: plt.figure(figsize=(12, 6))
sns.barplot(data=avg_pollution_levels, x=avg_pollution_levels.index, y='S02',
            label='S02')
```

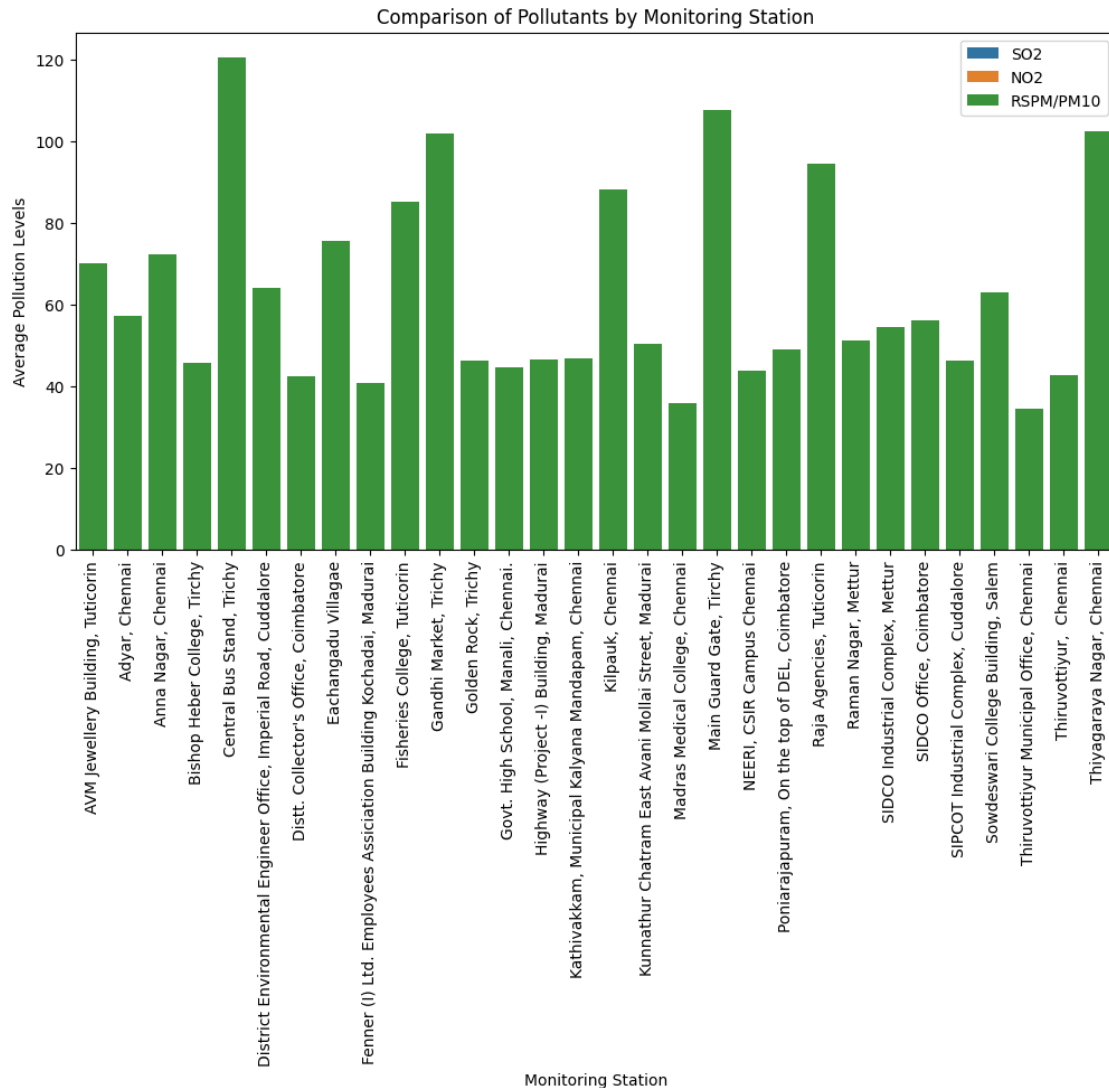
```

sns.barplot(data=avg_pollution_levels, x=avg_pollution_levels.index, y='NO2',
            label='NO2')
sns.barplot(data=avg_pollution_levels, x=avg_pollution_levels.index, y='RSPM/
            PM10', label='RSPM/PM10')
plt.xlabel('Monitoring Station')
plt.ylabel('Average Pollution Levels')
plt.title('Comparison of Pollutants by Monitoring Station')
plt.xticks(rotation=90)
plt.legend()
plt.show()

sns.pairplot(df[pollutants])
correlation_matrix = df[pollutants].corr()

df['Sampling Date'] = pd.to_datetime(df['Sampling Date'])
df.set_index('Sampling Date', inplace=True)
so2_time_series = df.groupby('Location of Monitoring Station')['SO2'].
    resample('M').mean()
so2_time_series.unstack().plot(figsize=(12, 6))

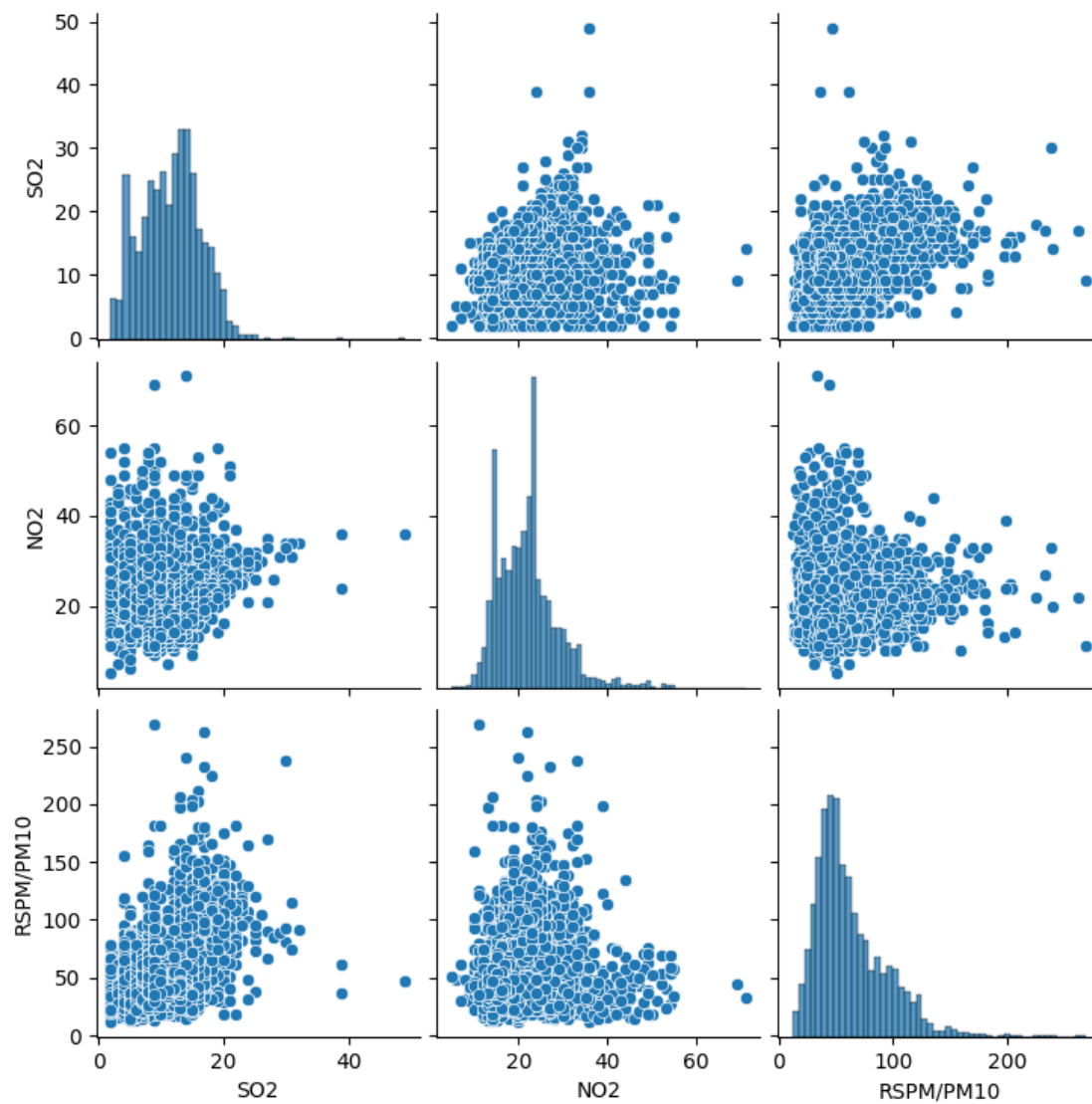
```



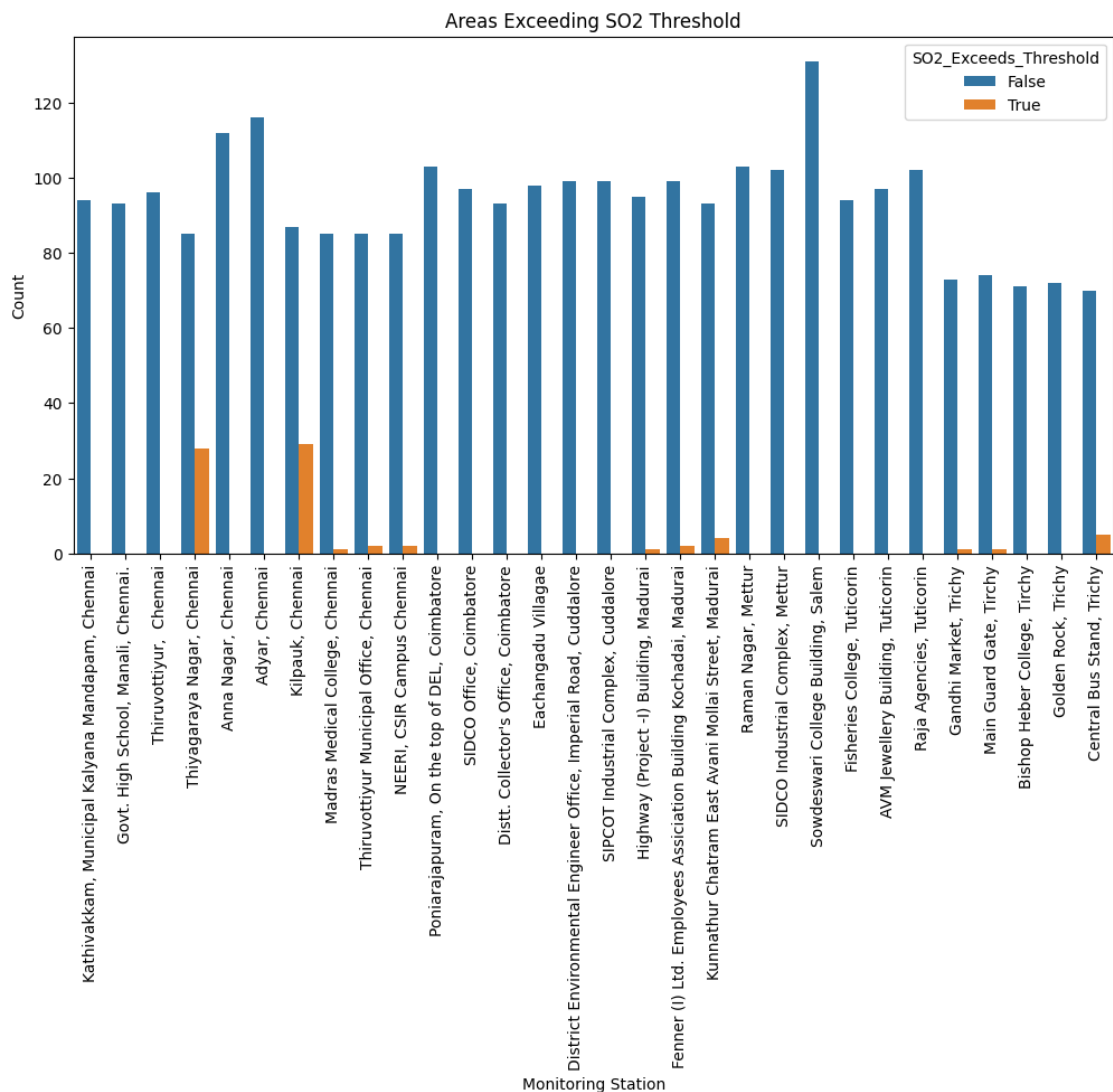
C:\Users\smile\AppData\Local\Temp\ipykernel\_5668\805127456.py:23: UserWarning: Could not infer format, so each element will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a format.

```
df['Sampling Date'] = pd.to_datetime(df['Sampling Date'])
```

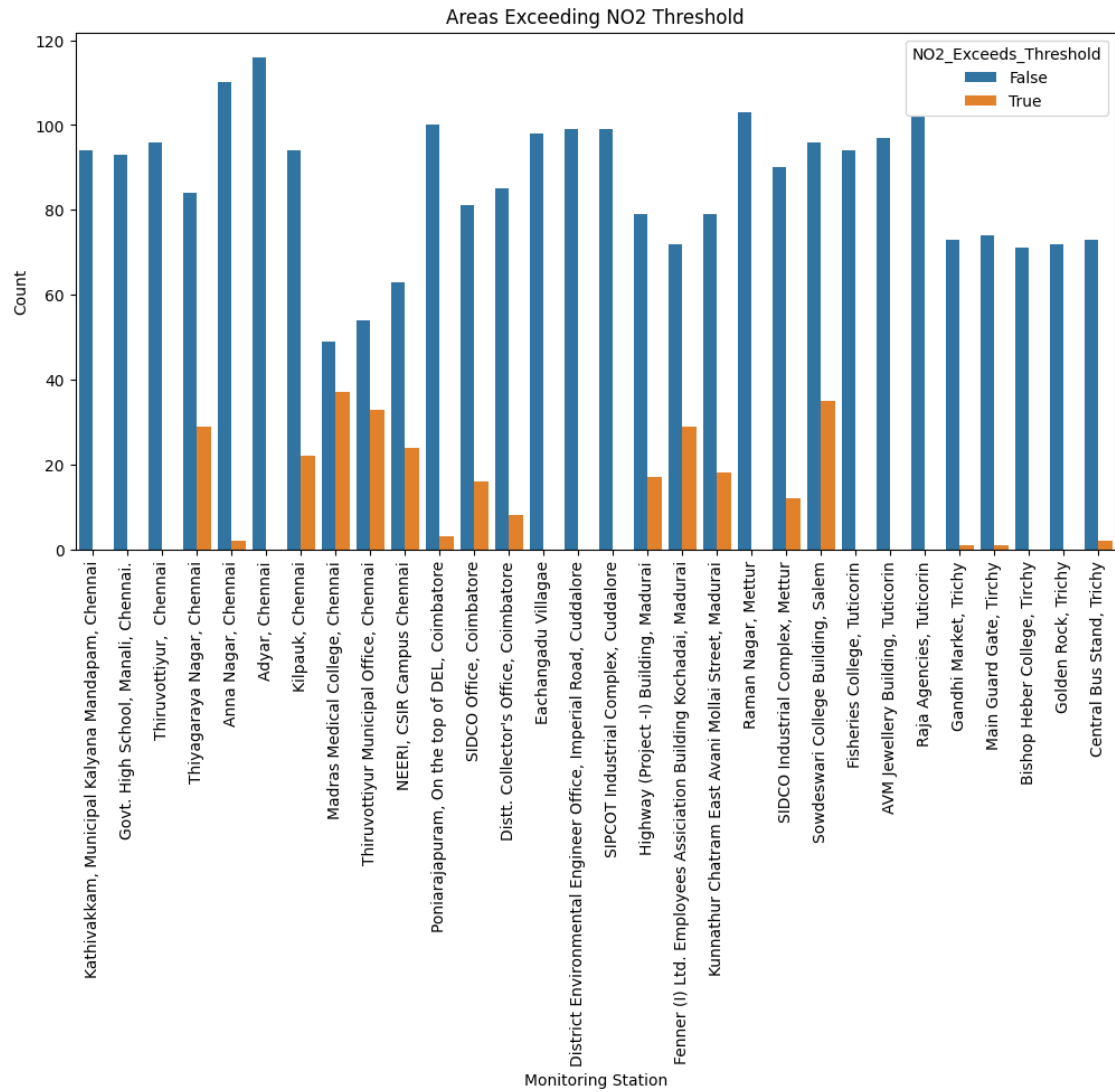
[7]: <Axes: xlabel='Location of Monitoring Station'>

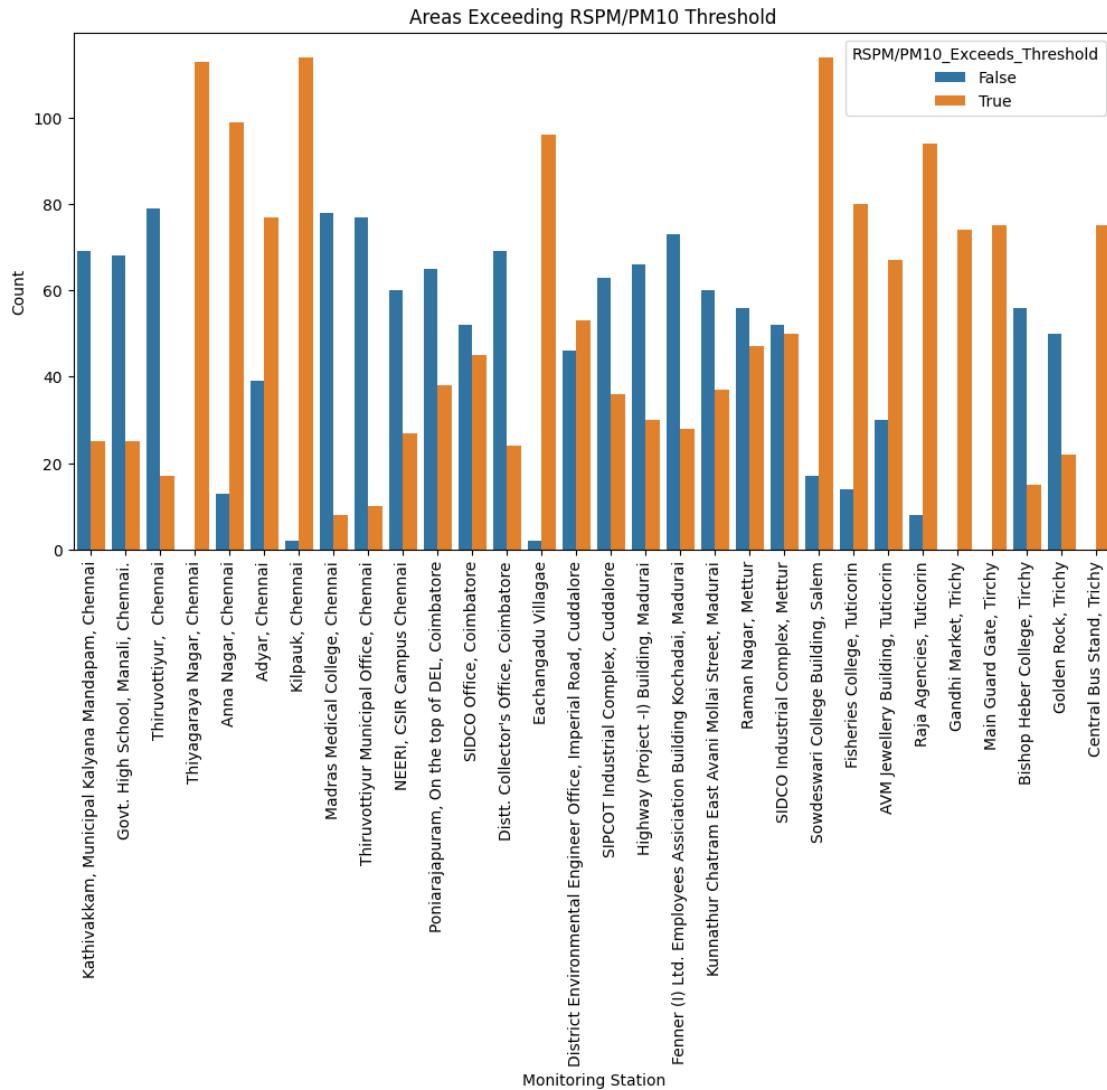












```
[9]: avg_pollution_levels.describe()
correlation_matrix
plt.figure(figsize=(12, 6))
sns.countplot(data=df, x='Location of Monitoring Station',
             hue='S02_Exceeds_Threshold')
plt.xlabel('Monitoring Station')
plt.ylabel('Count')
plt.title('Areas Exceeding S02 Threshold')
plt.xticks(rotation=90)
plt.savefig('S02_Exceedance.png')
```

