Unemployment_analysis

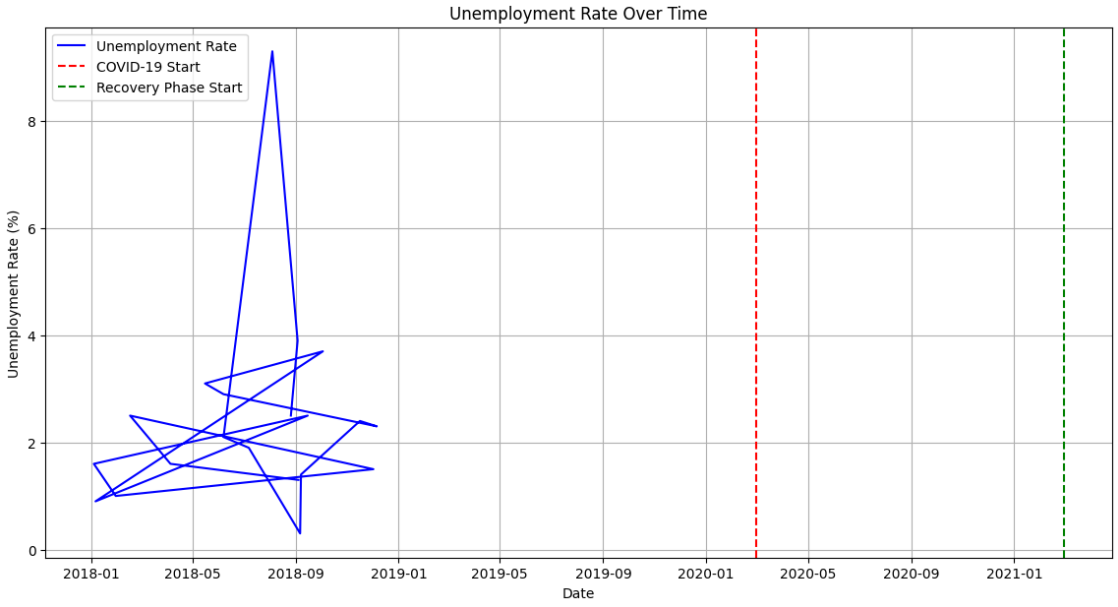
October 17, 2024

[2]:
```
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_excel(r'C:\Users\91969\OneDrive\Desktop\unemployment_data.xlsx')
print("First five rows of the dataset:")
print(data.head())
data['Date'] = pd.to_datetime(data['Date'])
print("\nMissing values in each column:")
print(data.isnull().sum())
data.dropna(inplace=True)
print("\nBasic statistics of the dataset:")
print(data.describe())
data['Rate Change'] = data['Unemployment Rate'].diff()
plt.figure(figsize=(14, 7))
plt.plot(data['Date'], data['Unemployment Rate'], label='Unemployment Rate',␣
 ↪color='blue')
plt.axvline(pd.Timestamp('2020-03-01'), color='red', linestyle='--',␣
 ↪label='COVID-19 Start')
plt.axvline(pd.Timestamp('2021-03-01'), color='green', linestyle='--',␣
 ↪label='Recovery Phase Start')
plt.title('Unemployment Rate Over Time')
plt.xlabel('Date')
plt.ylabel('Unemployment Rate (%)')
plt.legend()
plt.grid()
plt.show()
```

```
First five rows of the dataset:
        Date  Unemployment Rate
0 2018-08-26                2.5
1 2018-09-03                3.9
2 2018-08-04                9.3
3 2018-06-07                2.1
4 2018-07-07                1.9

Missing values in each column:
Date                 0
Unemployment Rate    0
dtype: int64
```

```
Basic statistics of the dataset:
                      Date  Unemployment Rate
count                   20          20.000000
mean   2018-07-08 21:36:00           2.435000
min    2018-01-04 00:00:00           0.300000
25%    2018-05-05 18:00:00           1.475000
50%    2018-08-15 00:00:00           2.200000
75%    2018-09-09 00:00:00           2.600000
max    2018-12-06 00:00:00           9.300000
std                    NaN           1.857637
```



[ ]: