FULL EMOTION DATA ANALYSIS & CLASSIFICATION PIPELINE

```
First 5 rows of the dataset:
   Unnamed: 0.1  Unnamed: 0  \
0             0           0
1             1           1
2             2           2
3             3           3
4             4           4


                                         Text      Sentiment  \
0  Enjoying a beautiful day at the park!       ...  Positive
1  Traffic was terrible this morning.          ...  Negative
2  Just finished an amazing workout! 💪         ...   Positive
3  Excited about the upcoming weekend getaway!  ...  Positive
4  Trying out a new recipe for dinner tonight.  ...  Neutral


            Timestamp          User      Platform  \
0  15-01-2023 12:30    User123        Twitter
1  15-01-2023 08:45    CommuterX      Twitter
2  15-01-2023 15:45    FitnessFan     Instagram
3  15-01-2023 18:20    AdventureX     Facebook
4  15-01-2023 19:55    ChefCook       Instagram


                        Hashtags   Retweets   Likes        Country
0   #Nature #Park                       15        30     USA
1   #Traffic #Morning                    5        10     Canada
2   #Fitness #Workout                   20        40     USA
3   #Travel #Adventure                   8        15     UK
4   #Cooking #Food                      12        25     Australia
```
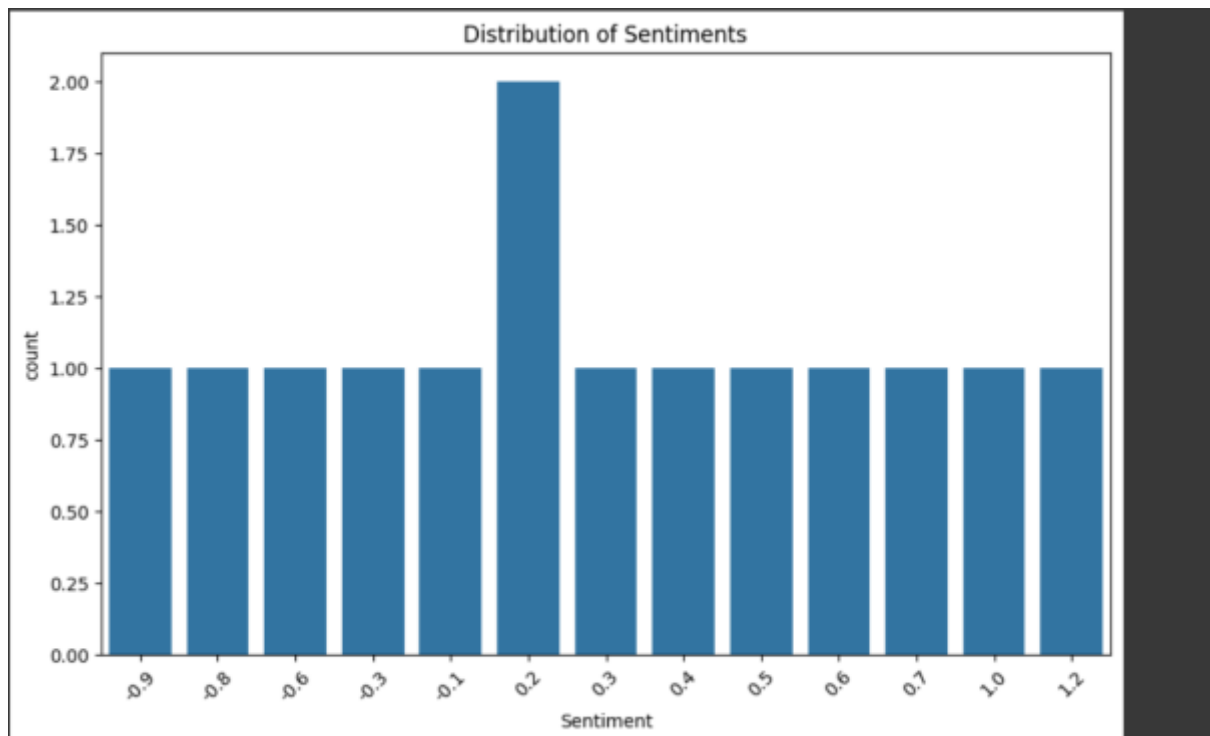
```
   Year  Month  Day  Hour
0  2023      1   15    12
1  2023      1   15     8
2  2023      1   15    15
3  2023      1   15    18
4  2023      1   15    19

Column names: ['Unnamed: 0.1', 'Unnamed: 0', 'Text', 'Sentiment', 'Timestamp', 'User', 'Platform', 'Hashtags', 'Retweets', 'Likes', 'Country', '
```

#DISTRIBUTION OF SENTIMENT

Distribution of Sentiments

```
Class distribution:
Sentiment
 0.2    2
-0.8    1
-0.3    1
 0.4    1
 0.7    1
 1.0    1
-0.9    1
 0.3    1
-0.1    1
 0.6    1
-0.6    1
 1.2    1
 0.5    1
Name: count, dtype: int64
```
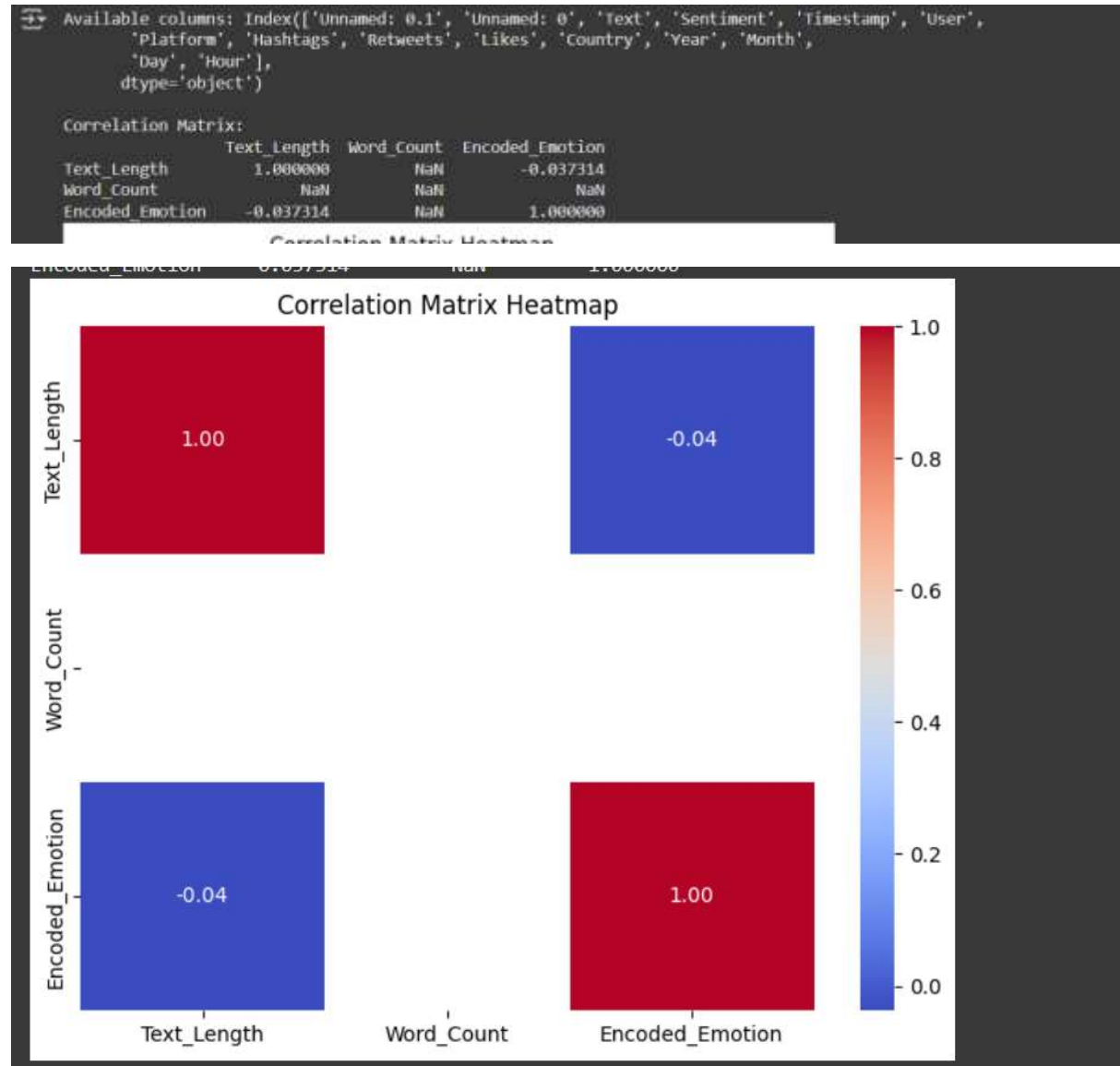
#Encode labels


```
Encoded sentiments:
{np.float64(-0.9): np.int64(0), np.float64(-0.8): np.int64(1), np.float64(-0.6): np.int64(2), np.float64(-0.3): np.int64(3), np.float64(-0.1): np
```

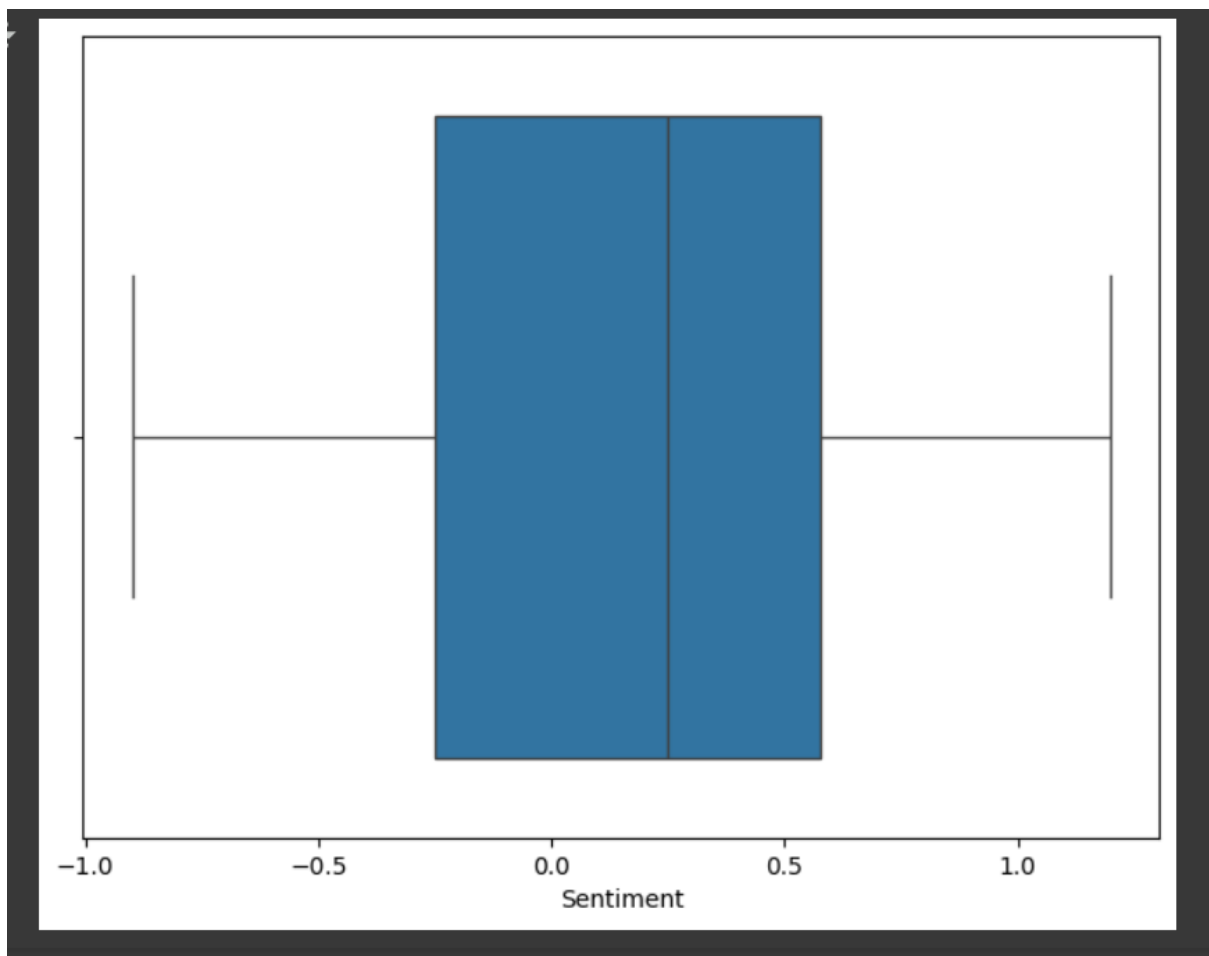#TRAINING THE MODE


```
Training the model...
```

#HEATMAP

```
Available columns: Index(['Unnamed: 0.1', 'Unnamed: 0', 'Text', 'Sentiment', 'Timestamp', 'User',
       'Platform', 'Hashtags', 'Retweets', 'Likes', 'Country', 'Year', 'Month',
       'Day', 'Hour'],
      dtype='object')

Correlation Matrix:
                 Text_Length  Word_Count  Encoded_Emotion
Text_Length         1.000000         NaN        -0.037314
Word_Count               NaN         NaN              NaN
Encoded_Emotion    -0.037314         NaN         1.000000
```


Correlation Matrix Heatmap

#BOXPLOT

#Z-SCORE

Z-scores: [np.float64(-0.26940795384816235), np.float64(0.14967100502231233), np.float64(0.35921060405354976), np.float64(-0.898026510133745),