

Name: Ankit Gaikwad

PRN: 202401040168

Roll No: CS4-57

Batch: C43

The screenshot shows the first 30 lines of the Python script 'Theory Activity.py'. The script is titled '# OpinRank Review Dataset Analysis using Numpy and Pandas'. It imports 'pandas as pd', 'numpy as np', and 'Counter' from 'collections'. It loads a dataset from 'ModelTrain.csv' into a DataFrame 'df'. It defines two helper functions: 'word_count(text)' which returns the length of a string split into words, and 'all_words(series)' which returns a list of all words in a series of reviews. The script then calculates several statistics: 1. Total number of reviews (df.shape[0]), 2. Count of POSITIVE and NEGATIVE reviews (df['Sentiment'].value_counts()), 3. Percentage of POSITIVE reviews (sentiment_counts['POSITIVE'] / total_reviews) * 100, and 4. Percentage of NEGATIVE reviews (sentiment_counts['NEGATIVE'] / total_reviews) * 100.

```
1 # OpinRank Review Dataset Analysis using Numpy and Pandas
2
3 import pandas as pd
4 import numpy as np
5 from collections import Counter
6
7 # Load the dataset
8 file_path = 'ModelTrain.csv'
9 df = pd.read_csv(file_path)
10
11 # Helper functions
12 def word_count(text):
13     return len(str(text).split())
14
15 def all_words(series):
16     words = []
17     for review in series:
18         words.extend(str(review).lower().split())
19     return words
20
21 # 1. Total number of reviews
22 total_reviews = df.shape[0]
23
24 # 2. Count of POSITIVE and NEGATIVE reviews
25 sentiment_counts = df['Sentiment'].value_counts()
26
27 # 3. Percentage of POSITIVE reviews
28 positive_percentage = (sentiment_counts['POSITIVE'] / total_reviews) * 100
29
30 # 4. Percentage of NEGATIVE reviews
```

The screenshot shows the continuation of the Python script 'Theory Activity.py' from line 30 to 58. It calculates the percentage of negative reviews, the average length of reviews in characters, the average length of positive and negative reviews, the review with maximum and minimum characters, the total number of words across all reviews, the average number of words per review, the most common word across all reviews, and the most common word in positive reviews.

```
30 # 4. Percentage of NEGATIVE reviews
31 negative_percentage = (sentiment_counts['NEGATIVE'] / total_reviews) * 100
32
33 # 5. Average length of reviews (characters)
34 avg_length_chars = df['Review'].apply(len).mean()
35
36 # 6. Average length of positive reviews
37 avg_length_positive = df[df['Sentiment'] == 'POSITIVE']['Review'].apply(len).mean()
38
39 # 7. Average length of negative reviews
40 avg_length_negative = df[df['Sentiment'] == 'NEGATIVE']['Review'].apply(len).mean()
41
42 # 8. Review with maximum characters
43 max_length_review = df.loc[df['Review'].apply(len).idxmax(), 'Review']
44
45 # 9. Review with minimum characters
46 min_length_review = df.loc[df['Review'].apply(len).idxmin(), 'Review']
47
48 # 10. Total number of words across all reviews
49 total_words = df['Review'].apply(word_count).sum()
50
51 # 11. Average number of words per review
52 avg_words_per_review = df['Review'].apply(word_count).mean()
53
54 # 12. Most common word across all reviews
55 all_words_list = all_words(df['Review'])
56 most_common_word = Counter(all_words_list).most_common(1)[0]
57
58 # 13. Most common word in positive reviews
59 positive_words_list = all_words(df[df['Sentiment'] == 'POSITIVE']['Review'])
```

```
File Edit Selection View Go Run ... eds
EXPLORER
OPEN EDITORS
Theory Activity.py
EDS
28.04.2025_22.00.02_REC.png
28.04.2025_22.00.40_REC.png
28.04.2025_22.00.59_REC.png
File Operations and Progra...
File Operations and Progra...
File Operations and Progra...
ModelTrain.csv
python-tutorial-thumbnailj...
Theory Activity.py
tryanderror.py
OUTLINE
TIMELINE
Theory Activity.py
79 reviews_more_than_100_words = df[df['Review'].apply(word_count) > 100].shape[0]
80
81 # 20. Average word length across all reviews
82 all_word_lengths = [len(word) for word in all_words_list]
83 average_word_length = np.mean(all_word_lengths)
84
85 # Displaying results
86 print("Total Reviews:", total_reviews)
87 print("Sentiment Counts:\n", sentiment_counts)
88 print(f"Positive %: {positive_percentage:.2f}")
89 print(f"Negative %: {negative_percentage:.2f}")
90 print(f"Average Review Length (chars): {avg_length_chars:.2f}")
91 print(f"Average Positive Review Length (chars): {avg_length_positive:.2f}")
92 print(f"Average Negative Review Length (chars): {avg_length_negative:.2f}")
93 print("Review with Maximum Characters (preview):", max_length_review[:300], "...")
94 print("Review with Minimum Characters:", min_length_review)
95 print("Total Words:", total_words)
96 print(f"Average Words per Review: {avg_words_per_review:.2f}")
97 print("Most Common Word (All Reviews):", most_common_word)
98 print("Most Common Word (Positive Reviews):", most_common_word_positive)
99 print("Most Common Word (Negative Reviews):", most_common_word_negative)
100 print("Reviews with 'clean':", reviews_with_clean)
101 print("Positive Reviews with 'excellent':", positive_with_excellent)
102 print("Negative Reviews with 'poor':", negative_with_poor)
103 print("Top 5 Frequent Words:", top_5_words)
104 print("Reviews > 100 Words:", reviews_more_than_100_words)
105 print(f"Average Word Length: {average_word_length:.2f}")
106
```

```
File Edit Selection View Go Run ... eds
EXPLORER
OPEN EDITORS
Theory Activity.py
EDS
28.04.2025_22.00.02_REC.png
28.04.2025_22.00.40_REC.png
28.04.2025_22.00.59_REC.png
28.04.2025_22.01.16_REC.png
File Operations and Progra...
File Operations and Progra...
File Operations and Progra...
ModelTrain.csv
python-tutorial-thumbnailj...
Theory Activity.py
tryanderror.py
OUTLINE
TIMELINE
PROBLEMS OUTPUT
[Running] python -u "d:\Documents\assignment\eds\Theory Activity.py"
Total Reviews: 123372
Sentiment Counts:
Sentiment
POSITIVE 62269
NEGATIVE 61103
Name: count, dtype: int64
Positive %: 50.47
Negative %: 49.53
Average Review Length (chars): 542.09
Average Positive Review Length (chars): 488.99
Average Negative Review Length (chars): 596.20
Review with Maximum Characters (preview): grander pricewe holiday virgin include flight transfer not give complete accurate
guide price would say look website amount pay safe say two weekend weeknight two locationthe grand hyatt excellent wish explore
whole new york situate next door grand central station not great shopping not much distanc ...
Review with Minimum Characters: k
Total Words: 10520526
Average Words per Review: 85.27
Most Common Word (All Reviews): ('not', 304477)
Most Common Word (Positive Reviews): ('not', 99465)
Most Common Word (Negative Reviews): ('not', 205012)
Reviews with 'clean': 44973
Positive Reviews with 'excellent': 12967
Negative Reviews with 'poor': 6753
Top 5 Frequent Words: [('not', 304477), ('stay', 167514), ('would', 103494), ('good', 94328), ('get', 92232)]
Reviews > 100 Words: 36477
Average Word Length: 5.37
[Done] exited with code=0 in 14.165 seconds
```