NAME: ISHART MAHAMADSHARIF MULANI

BRANCH:ENTC DIVISION: ET2 ROLL NO: ET2-58 SUBJECT: EDS

TOPIC: problem statements for a OpinRank Review dataset using Numpy and Pandas

1.Total number of review import pandas as pd df = pd.read\_csv('hotel.csv') total\_reviews = df.shape[0]

2.Find the number of missing values in each column. missing\_values = df.isnull().sum()

3.List unique authors. unique\_authors = df['author'].unique()

4.Find the review with the highest VADER rating. highest\_vader = df.loc[df['Vader\_rati'].idxmax()]

5.Find the review with the lowest VADER rating lowest\_vader = df.loc[df['Vader\_rati'].idxmin()]

6.First 5 record df.head()

7.Last 5 record df.tail()

8.first 10 record df.head(10)

9.No. of rows and columns df.shape()

10.Last 10 record df.tail(10)

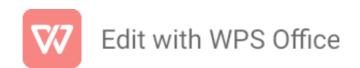
11.Find the top 3 reviews with the longest text after cleaning.

df['text\_length'] = df['cleaned\_te'].str.len()
longest\_reviews = df.sort\_values(by='text\_length', ascending=False).head(3)

12.Column name df.columns

13.Datatype df.dtypes

14.Find the median VADER rating.
import pandas as pd
import numpy as np
df = pd.read\_csv('hotel.csv')
median\_vader = np.median(df['Vader\_rati'].to\_numpy())



15.Find the number of reviews where VADER rating is above average.
vader\_scores = df['Vader\_rati'].to\_numpy()
average\_vader = np.mean(vader\_scores)
above\_avq\_count = np.sum(vader\_scores > average\_vader)

16.Find the 5 smallest VADER ratings. five\_smallest\_vader = np.sort(vader\_scores)[:5]

- 17.Find the average VADER rating for reviews with final\_sentiment = 1. vader\_positive = vader\_scores[df['final\_sentiment'].to\_numpy() == 1] average\_vader\_positive = np.mean(vader\_positive)
- 18.Normalize the VADER ratings between 0 and 1.

  vader\_min = np.min(vader\_scores)

  vader\_max = np.max(vader\_scores)

  vader\_normalized = (vader\_scores vader\_min) / (vader\_max vader\_min)
- 19.Find reviews whose VADER rating is in the top 10%. threshold = np.percentile(vader\_scores, 90) top\_10\_percent\_reviews = df[vader\_scores >= threshold]
- 20.Find the index of reviews where VADER score is negative. negative\_indices = np.where(vader\_scores < 0)[0]

