```python
# Abdurrahman Bulut

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report

df = pd.read_csv("amazon_review.csv")

# Başlık Bilgisi
print(df.columns)
```

```
Index(['reviewerID', 'asin', 'reviewerName', 'helpful', 'reviewText',
       'overall', 'summary', 'unixReviewTime', 'reviewTime',
'day_diff',
       'helpful_yes', 'total_vote'],
      dtype='object')
```

```python
print(df.head())
```

```
       reviewerID        asin  reviewerName helpful  \
0  A3SBTW3WS4IQSN  B007WTAJTO           NaN  [0, 0]
1  A18K1ODH1I2MVB  B007WTAJTO          0mie  [0, 0]
2  A2FII3I2MBMUIA  B007WTAJTO           1K3  [0, 0]
3   A3H99DFEG68SR  B007WTAJTO           1m2  [0, 0]
4  A375ZM4U047O79  B007WTAJTO  2&amp;1/2Men  [0, 0]

                                          reviewText  overall  \
0                                         No issues.      4.0
1  Purchased this for my device, it worked as adv...      5.0
2  it works as expected. I should have sprung for...      4.0
3  This think has worked out great.Had a diff. br...      5.0
4  Bought it with Retail Packaging, arrived legit...      5.0

                                summary  unixReviewTime  reviewTime
\
0                            Four Stars      1406073600  2014-07-23

1                          MOAR SPACE!!!      1382659200  2013-10-25

2                nothing to really say....      1356220800  2012-12-23

3  Great buy at this price!!!  *** UPDATE      1384992000  2013-11-21

4                      best deal around      1373673600  2013-07-13
```

```
   day_diff  helpful_yes  total_vote
0       138            0           0
1       409            0           0
2       715            0           0
3       382            0           0
4       513            0           0
```

```
print(df.dtypes)
```

```
reviewerID                    object
asin                          object
reviewerName                  object
helpful                       object
reviewText                    object
overall                      float64
summary                       object
unixReviewTime                 int64
reviewTime            datetime64[ns]
day_diff                       int64
helpful_yes                    int64
total_vote                     int64
days_since_review              int64
time_weight                  float64
time_period                 category
dtype: object
```

```
print(df.isnull().sum())
```

```
reviewerID            0
asin                  0
reviewerName          1
helpful               0
reviewText            1
overall               0
summary               0
unixReviewTime        0
reviewTime            0
day_diff              0
helpful_yes           0
total_vote            0
days_since_review     0
time_weight           0
time_period          12
dtype: int64
```

```
print(df.describe())
```

```
           overall  unixReviewTime                      reviewTime  \
count  4915.000000    4.915000e+03                            4915
mean      4.587589    1.379465e+09  2013-09-26 15:11:27.772125952
min       1.000000    1.339200e+09             2012-01-09 00:00:00
```

```
25%       5.000000    1.365898e+09            2013-04-16 00:00:00
50%       5.000000    1.381277e+09            2013-10-03 00:00:00
75%       5.000000    1.392163e+09            2014-03-02 00:00:00
max       5.000000    1.406074e+09            2014-12-07 00:00:00
std       0.996845    1.581857e+07                            NaN

          day_diff  helpful_yes   total_vote  days_since_review
time_weight
count  4915.000000  4915.000000  4915.000000        4915.000000
4915.000000
mean    437.367040     1.311089     1.521465         436.367040
0.008708
min       1.000000     0.000000     0.000000           0.000000
0.000940
25%     281.000000     0.000000     0.000000         280.000000
0.001664
50%     431.000000     0.000000     0.000000         430.000000
0.002320
75%     601.000000     0.000000     0.000000         600.000000
0.003559
max    1064.000000  1952.000000  2020.000000        1063.000000
1.000000
std     209.439871    41.619161    44.123095         209.439871
0.058346
```

Task 1

Step 1

```
# Ürünün ortalama puanını hesaplama
average_rating = df['overall'].mean()
print(f"Average Rating: {average_rating}")

Average Rating: 4.587589013224822
```

Step 2

```
# İnceleme zamanını datetime formatına çevirme
df['reviewTime'] = pd.to_datetime(df['reviewTime'])

# En son incelemeden bu yana geçen gün sayısını hesaplma
df['days_since_review'] = (df['reviewTime'].max() -
df['reviewTime']).dt.days

# Zaman tabanlı ağırlık fonksiyonunu tanımlama
def time_weighted_rating(days):
    return 1 / (days + 1)  # Bölme hatasından kaçınmak için +1 ekliyoz

# Ağırlık fonksiyonu
```

```
df['time_weight'] =
df['days_since_review'].apply(time_weighted_rating)

# Ağırlıklı ortalama puanı hesaplama
weighted_average_rating = np.average(df['overall'],
weights=df['time_weight'])
print(f"Time-Weighted Average Rating: {weighted_average_rating}")
```

```
Time-Weighted Average Rating: 4.71371948084595
```

Step 3

```
# Zaman dilimlerini gruplama
df['time_period'] = pd.cut(df['days_since_review'],
                           bins=[0, 30, 90,
df['days_since_review'].max()],
                           labels=['Last 30 days', '31-90 days', '91+
days'])

# Her dönem için ortalama puanı hesaplama
periodic_avg_ratings = df.groupby('time_period')['overall'].mean()
print(periodic_avg_ratings)

# Karşılaştırma
print(f"Periodic Average Ratings:\n{periodic_avg_ratings}")
```

```
time_period
Last 30 days    4.753846
31-90 days      4.793103
91+ days        4.579627
Name: overall, dtype: float64
Periodic Average Ratings:
time_period
Last 30 days    4.753846
31-90 days      4.793103
91+ days        4.579627
Name: overall, dtype: float64
```

Task 2

Adım 1

```
# helpful_no değişkenini oluşturma
df['helpful_no'] = df['total_vote'] - df['helpful_yes']
```

Adım 2

```
# score_pos_neg_diff fonksiyonunu tanımlama
def score_pos_neg_diff(up, down):
```

```
    return up - down

# score_average_rating fonksiyonunu tanımlama
def score_average_rating(up, down):
    if up + down == 0:
        return 0
    return up / (up + down)

# wilson_lower_bound fonksiyonunu tanımlama
from scipy.stats import norm

def wilson_lower_bound(up, down, confidence=0.95):
    n = up + down
    if n == 0:
        return 0
    z = norm.ppf(1 - (1 - confidence) / 2)
    phat = 1.0 * up / n
    return (phat + z * z / (2 * n) - z * ((phat * (1 - phat) + z * z /
(4 * n)) / n)) / (1 + z * z / n)

# Skorları hesaplama ve veriye ekleme
df['score_pos_neg_diff'] = df.apply(lambda x:
score_pos_neg_diff(x['helpful_yes'], x['helpful_no']), axis=1)
df['score_average_rating'] = df.apply(lambda x:
score_average_rating(x['helpful_yes'], x['helpful_no']), axis=1)
df['wilson_lower_bound'] = df.apply(lambda x:
wilson_lower_bound(x['helpful_yes'], x['helpful_no']), axis=1)
```

Adım 3

```
# wilson_lower_bound'a göre ilk 20 yorumu belirleme ve sıralama
top_20_reviews = df.sort_values('wilson_lower_bound',
ascending=False).head(20)
print(top_20_reviews)
```

```
          reviewerID        asin                         reviewerName
\
2031  A12B7ZMXFI6IXY  B007WTAJTO                Hyoun Kim "Faluzure"

3449   AOEAD7DPLZE53  B007WTAJTO                   NLee the Engineer

4212   AVBMZZAFEKO58  B007WTAJTO                         SkincareCEO

4672  A2DKQQIZ793AV5  B007WTAJTO                             Twister

1835  A1J6VSUM80UAF8  B007WTAJTO                         goconfigure

317   A1ZQAQFYSXL5MQ  B007WTAJTO             Amazon Customer "Kelly"

3807   AFGRMORWY2QNX  B007WTAJTO                          R. Heisler
```

```
1609  A2TPXOZSU1DACQ  B007WTAJTO                                  Eskimo
1465   A6I8KXYK24RTB  B007WTAJTO                                D. Stein
3981  A1K91XXQ6ZEBQR  B007WTAJTO             R. Sutton, Jr. "RWSynergy"
4302  A2EL2GWJ9T0DWY  B007WTAJTO                              Stayeraug
4072  A22GOZTFA02O2F  B007WTAJTO                            sb21 "sb21"
4306   AOHXKM5URSKAB  B007WTAJTO                          Stellar Eller
315   A2J26NNQX6WKAU  B007WTAJTO              Amazon Customer "johncrea"
4596  A1WTQUOQ4WG9AI  B007WTAJTO            Tom Henriksen "Doggy Diner"
1142  A1PLHPPAJ5MUXG  B007WTAJTO  Daniel Pham(Danpham_X @ yahoo.  com)
121   A2Z4VVF1NTJWPB  B007WTAJTO                                  A. Lee
1753   ALPLKR59QMBUX  B007WTAJTO                               G. Becker
1072  A2O96COBMVY9C4  B007WTAJTO                          Crysis Complex
2583  A3MEPYZVTAV90W  B007WTAJTO                                J. Wong

           helpful                                          reviewText
\
2031  [1952, 2020]  [[ UPDATE - 6/19/2014 ]]So my lovely wife boug...
3449  [1428, 1505]  I have tested dozens of SDHC and micro-SDHC ca...
4212  [1568, 1694]  NOTE:  please read the last update (scroll to ...
4672      [45, 49]  Sandisk announcement of the first 128GB micro ...
1835      [60, 68]  Bought from BestBuy online the day it was anno...
317     [422, 495]  If your card gets hot enough to be painful, it...
3807      [22, 25]  I bought this card to replace a lost 16 gig in...
1609        [7, 7]  I have always been a sandisk guy.  This cards ...
1465        [7, 7]  I for one have not bought into Google's, or an...
3981    [112, 139]  The last few days I have been diligently shopp...
4302      [14, 16]  So I got this SD specifically for my GoPro Bla...
```

```
4072        [6, 6]  I used this for my Samsung Galaxy Tab 2 7.0 . ...
4306      [51, 65]  While I got this card as a "deal of the day" o...
315       [38, 48]  Bought this card to use with my Samsung Galaxy...
4596     [82, 109]  Hi:I ordered two card and they arrived the nex...
1142        [5, 5]  As soon as I saw that this card was announced ...
121         [5, 5]  Update: providing an update with regard to San...
1753        [5, 5]  Puchased this card right after I received my S...
1072        [5, 5]  What more can I say? The 64GB micro SD works f...
2583        [5, 5]  I bought this Class 10 SD card for my GoPro 3 ...


      overall                                                summary  \
2031      5.0  UPDATED - Great w/ Galaxy S4 & Galaxy Tab 4 10...
3449      5.0  Top of the class among all (budget-priced) mic...
4212      1.0  1 Star reviews - Micro SDXC card unmounts itse...
4672      5.0  Super high capacity!!!  Excellent price (on Am...
1835      5.0                                          I own it
317       1.0                                Warning, read this!
3807      3.0   Good buy for the money but wait, I had an issue!
1609      5.0                 Bet you wish you had one of these
1465      4.0                                          Finally.
3981      5.0  Resolving confusion between "Mobile Ultra" and...
4302      5.0                          Perfect with GoPro Black 3+
4072      5.0              Used for my Samsung Galaxy Tab 2 7.0
4306      5.0                                     Awesome Card!
315       5.0  Samsung Galaxy Tab2 works with this card if re...
4596      1.0     Designed incompatibility/Don't support SanDisk
1142      5.0                           Great large capacity card
121       5.0                    ready for use on the Galaxy S3
1753      5.0                   Use Nothing Other Than the Best
1072      5.0               Works wonders for the Galaxy Note 2!
2583      5.0                  Works Great with a GoPro 3 Black!

      unixReviewTime reviewTime  day_diff  helpful_yes  total_vote  \
2031      1367366400 2013-01-05       702         1952        2020
3449      1348617600 2012-09-26       803         1428        1505
4212      1375660800 2013-05-08       579         1568        1694
4672      1394150400 2014-07-03       158           45          49
1835      1393545600 2014-02-28       283           60          68
317       1346544000 2012-02-09      1033          422         495
3807      1361923200 2013-02-27       649           22          25
1609      1395792000 2014-03-26       257            7           7
```

```
1465      1397433600 2014-04-14       238            7            7
3981      1350864000 2012-10-22       777          112          139
4302      1395360000 2014-03-21       262           14           16
4072      1347321600 2012-11-09       759            6            6
4306      1339200000 2012-09-06       823           51           65
315       1344816000 2012-08-13       847           38           48
4596      1348272000 2012-09-22       807           82          109
1142      1396396800 2014-02-04       307            5            5
121       1346803200 2012-05-09       943            5            5
1753      1350864000 2012-10-22       777            5            5
1072      1349395200 2012-05-10       942            5            5
2583      1370649600 2013-08-06       489            5            5

      days_since_review  time_weight time_period  helpful_no  \
2031                701     0.001425    91+ days          68
3449                802     0.001245    91+ days          77
4212                578     0.001727    91+ days         126
4672                157     0.006329    91+ days           4
1835                282     0.003534    91+ days           8
317                1032     0.000968    91+ days          73
3807                648     0.001541    91+ days           3
1609                256     0.003891    91+ days           0
1465                237     0.004202    91+ days           0
3981                776     0.001287    91+ days          27
4302                261     0.003817    91+ days           2
4072                758     0.001318    91+ days           0
4306                822     0.001215    91+ days          14
315                 846     0.001181    91+ days          10
4596                806     0.001239    91+ days          27
1142                306     0.003257    91+ days           0
121                 942     0.001060    91+ days           0
1753                776     0.001287    91+ days           0
1072                941     0.001062    91+ days           0
2583                488     0.002045    91+ days           0

      score_pos_neg_diff  score_average_rating  wilson_lower_bound
2031                1884              0.966337            0.965420
3449                1351              0.948837            0.947631
4212                1442              0.925620            0.924577
4672                  41              0.918367            0.884445
1835                  52              0.882353            0.858691
317                  349              0.852525            0.849309
3807                  19              0.880000            0.819600
1609                   7              1.000000            0.798032
1465                   7              1.000000            0.798032
3981                  85              0.805755            0.795290
4302                  12              0.875000            0.785664
4072                   6              1.000000            0.772956
4306                  37              0.784615            0.763501
```

315	28	0.791667	0.763062
4596	55	0.752294	0.740315
1142	5	1.000000	0.740180
121	5	1.000000	0.740180
1753	5	1.000000	0.740180
1072	5	1.000000	0.740180
2583	5	1.000000	0.740180

En yararlı yorumların belirlenmesi, kullanıcıların en doğru bilgiye ulaşmasını sağlar. Wilson Lower Bound skoru, özellikle az sayıda oy alan yorumlarda daha güvenilir bir ölçüm sağlar.