Assignment3_AI

June 25, 2023

0.0.1 Introduction

[1]:
```
import pandas as pd

#Get the dataframe
path="Downloads/Musical_instruments_reviews.csv"
df = pd.read_csv(path)
df.head()
```

[1]:
```
       reviewerID        asin  \
0  A2IBPI20UZIR0U  1384719342
1  A14VAT5EAX3D9S  1384719342
2  A195EZSQDW3E21  1384719342
3  A2C00NNG1ZQQG2  1384719342
4   A94QU4C90B1AX  1384719342

                                       reviewerName   helpful  \
0  cassandra tu "Yeah, well, that's just like, u…    [0, 0]
1                                              Jake  [13, 14]
2                     Rick Bennette "Rick Bennette"    [1, 1]
3                         RustyBill "Sunday Rocker"    [0, 0]
4                                      SEAN MASLANKA    [0, 0]

                                          reviewText  overall  \
0  Not much to write about here, but it does exac…      5.0
1  The product does exactly as it should and is q…      5.0
2  The primary job of this device is to block the…      5.0
3  Nice windscreen protects my MXL mic and preven…      5.0
4  This pop filter is great. It looks and perform…      5.0

                                  summary  unixReviewTime   reviewTime
0                                    good      1393545600  02 28, 2014
1                                    Jake      1363392000  03 16, 2013
2                    It Does The Job Well      1377648000  08 28, 2013
3           GOOD WINDSCREEN FOR THE MONEY      1392336000  02 14, 2014
4  No more pops when I record my vocals.      1392940800  02 21, 2014
```

```
[2]: #Get summary column
     summary=df['summary']
     summary
```

```
[2]: 0                                                     good
     1                                                     Jake
     2                                     It Does The Job Well
     3                            GOOD WINDSCREEN FOR THE MONEY
     4                    No more pops when I record my vocals.
                                    …
     10256                                           Five Stars
     10257    Long life, and for some players, a good econom…
     10258                                     Good for coated.
     10259                                          Taylor Made
     10260    These strings are really quite good, but I wou…
     Name: summary, Length: 10261, dtype: object
```

0.0.2 Tokenizer

```
[3]: import nltk
     from nltk.tokenize import word_tokenize
     import spacy
```

```
C:\Users\annet\anaconda3\lib\site-packages\scipy\__init__.py:146: UserWarning: A
NumPy version >=1.16.5 and <1.23.0 is required for this version of SciPy
(detected version 1.24.3
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

```
[4]: # Tokenize Summary column
     nltk.download('punkt')
     df['text_token'] = df.apply(lambda row: word_tokenize(row['summary']), axis=1)
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\annet\AppData\Roaming\nltk_data…
[nltk_data]   Package punkt is already up-to-date!
```

```
[5]: df['text_token']
```

```
[5]: 0                                                   [good]
     1                                                   [Jake]
     2                               [It, Does, The, Job, Well]
     3                      [GOOD, WINDSCREEN, FOR, THE, MONEY]
     4          [No, more, pops, when, I, record, my, vocals, .]
                                    …
     10256                                        [Five, Stars]
     10257    [Long, life, ,, and, for, some, players, ,, a,…
     10258                                [Good, for, coated, .]
     10259                                       [Taylor, Made]
```

```
10260    [These, strings, are, really, quite, good, ,, …
Name: text_token, Length: 10261, dtype: object
```

0.0.3 Stemmer

```
[6]: from nltk.stem import SnowballStemmer
     stemmer = SnowballStemmer('english')
     df['stemmed_summary'] = summary.apply(lambda x: ' '.join([stemmer.stem(word)␣
       ↪for word in x.split()]))
```

```
[7]: df['stemmed_summary']
```

```
[7]: 0                                                     good
     1                                                     jake
     2                                      it doe the job well
     3                            good windscreen for the money
     4                     no more pop when i record my vocals.
                                    …
     10256                                            five star
     10257    long life, and for some players, a good econom…
     10258                                     good for coated.
     10259                                          taylor made
     10260    these string are realli quit good, but i would…
     Name: stemmed_summary, Length: 10261, dtype: object
```

```
[8]: from nltk.stem import SnowballStemmer
     stemmer = SnowballStemmer('english')
     df['stemmed_summary2'] = df['text_token'].apply(lambda tokens: [stemmer.
       ↪stem(word) for word in tokens])
     df['stemmed_summary2']
```

```
[8]: 0                                                   [good]
     1                                                   [jake]
     2                                [it, doe, the, job, well]
     3                      [good, windscreen, for, the, money]
     4           [no, more, pop, when, i, record, my, vocal, .]
                                    …
     10256                                         [five, star]
     10257    [long, life, ,, and, for, some, player, ,, a, …
     10258                                 [good, for, coat, .]
     10259                                       [taylor, made]
     10260    [these, string, are, realli, quit, good, ,, bu…
     Name: stemmed_summary2, Length: 10261, dtype: object
```

0.0.4 Lemmatization

[9]:
```
from nltk.stem import WordNetLemmatizer
nlp = spacy.load('en_core_web_sm')
df['lemmatized_summary'] = df['summary'].apply(lambda x: [token.lemma_ for␣
 ↪token in nlp(str(x))])
```

[10]:
```
df['lemmatized_summary']
```

[10]:
```
0                                                   [good]
1                                                   [Jake]
2                                 [it, do, the, Job, well]
3                      [good, WINDSCREEN, for, the, money]
4           [no, more, pop, when, I, record, my, vocal, .]
                               …
10256                                         [five, star]
10257    [long, life, ,, and, for, some, player, ,, a, …
10258                                 [good, for, coat, .]
10259                                       [Taylor, make]
10260    [these, string, be, really, quite, good, ,, bu…
Name: lemmatized_summary, Length: 10261, dtype: object
```

[12]:
```
df
```

[12]:
```
           reviewerID        asin  \
0      A2IBPI20UZIR0U  1384719342
1      A14VAT5EAX3D9S  1384719342
2      A195EZSQDW3E21  1384719342
3      A2C00NNG1ZQQG2  1384719342
4       A94QU4C90B1AX  1384719342
…                 …          …
10256  A14B2YH83ZXMPP  B00JBIVXGC
10257   A1RPTVW5VEOSI  B00JBIVXGC
10258   AWCJ12KBO5VII  B00JBIVXGC
10259  A2Z7S8B5U4PAKJ  B00JBIVXGC
10260  A2WA8TDCTGUADI  B00JBIVXGC


                                             reviewerName   helpful  \
0      cassandra tu "Yeah, well, that's just like, u…    [0, 0]
1                                                    Jake  [13, 14]
2                                 Rick Bennette "Rick Bennette"    [1, 1]
3                                 RustyBill "Sunday Rocker"    [0, 0]
4                                            SEAN MASLANKA    [0, 0]
…                                                     …       …
10256                                     Lonnie M. Adams    [0, 0]
10257                                   Michael J. Edelman    [0, 0]
10258                                     Michael L. Knapp    [0, 0]
10259                               Rick Langdon "Scriptor"    [0, 0]
```

```
10260                                   TheTerrorBeyond    [0, 0]

                                              reviewText  overall  \
0      Not much to write about here, but it does exac…      5.0
1      The product does exactly as it should and is q…      5.0
2      The primary job of this device is to block the…      5.0
3      Nice windscreen protects my MXL mic and preven…      5.0
4      This pop filter is great. It looks and perform…      5.0
…                                                    …      …
10256            Great, just as expected.  Thank to all.      5.0
10257  I've been thinking about trying the Nanoweb st…      5.0
10258  I have tried coated strings in the past ( incl…      4.0
10259  Well, MADE by Elixir and DEVELOPED with Taylor…      4.0
10260  These strings are really quite good, but I wou…      4.0

                                                 summary  unixReviewTime  \
0                                                   good      1393545600
1                                                   Jake      1363392000
2                                   It Does The Job Well      1377648000
3                          GOOD WINDSCREEN FOR THE MONEY      1392336000
4                  No more pops when I record my vocals.      1392940800
…                                                    …             …
10256                                         Five Stars      1405814400
10257  Long life, and for some players, a good econom…      1404259200
10258                                   Good for coated.      1405987200
10259                                        Taylor Made      1404172800
10260  These strings are really quite good, but I wou…      1405468800

        reviewTime                                         text_token  \
0      02 28, 2014                                             [good]
1      03 16, 2013                                             [Jake]
2      08 28, 2013                         [It, Does, The, Job, Well]
3      02 14, 2014                [GOOD, WINDSCREEN, FOR, THE, MONEY]
4      02 21, 2014   [No, more, pops, when, I, record, my, vocals, .]
…            …                                                 …
10256  07 20, 2014                                      [Five, Stars]
10257   07 2, 2014  [Long, life, ,, and, for, some, players, ,, a,…
10258  07 22, 2014                              [Good, for, coated, .]
10259   07 1, 2014                                     [Taylor, Made]
10260  07 16, 2014  [These, strings, are, really, quite, good, ,, …

                                         stemmed_summary  \
0                                                   good
1                                                   jake
2                                    it doe the job well
3                          good windscreen for the money
4                   no more pop when i record my vocals.
```

```
…                                                  …
10256                                          five star
10257  long life, and for some players, a good econom…
10258                                   good for coated.
10259                                        taylor made
10260  these string are realli quit good, but i would…

                                        stemmed_summary2  \
0                                                 [good]
1                                                 [jake]
2                              [it, doe, the, job, well]
3                    [good, windscreen, for, the, money]
4         [no, more, pop, when, i, record, my, vocal, .]
…                                                  …
10256                                       [five, star]
10257  [long, life, ,, and, for, some, player, ,, a, …
10258                               [good, for, coat, .]
10259                                     [taylor, made]
10260  [these, string, are, realli, quit, good, ,, bu…

                                      lemmatized_summary
0                                                 [good]
1                                                 [Jake]
2                               [it, do, the, Job, well]
3                    [good, WINDSCREEN, for, the, money]
4         [no, more, pop, when, I, record, my, vocal, .]
…                                                  …
10256                                       [five, star]
10257  [long, life, ,, and, for, some, player, ,, a, …
10258                               [good, for, coat, .]
10259                                     [Taylor, make]
10260  [these, string, be, really, quite, good, ,, bu…

[10261 rows x 13 columns]
```

```
[ ]:
```

```
[ ]:
```