

In [ ]:

1

In [ ]:

1

## Nave Baye's Classifier

In [ ]:

1

Text classification is used in segregation of (movie reviews, hotel reviews, news data, customer support emails, complains, product review etc)

For many real life cases training a custom text model classification model proves to be much more accurate

In [164]:

```
1 import pandas as pd
2 import numpy as np
```

In [168]:

```
1 import seaborn as sns
2 import nltk
3 import re
4 from nltk.stem import PorterStemmer, WordNetLemmatizer
5 from nltk import pos_tag, wordnet
6 from nltk.tokenize import word_tokenize
7 from nltk.corpus import stopwords
8
9 from nltk.sentiment.vader import SentimentIntensityAnalyzer
10 from sklearn.feature_extraction.text import CountVectorizer
11 from sklearn.feature_extraction.text import TfidfTransformer
12 from sklearn.metrics import confusion_matrix
13 from sklearn.model_selection import train_test_split
14 from sklearn.naive_bayes import MultinomialNB
15 from wordcloud import WordCloud
```

In [215]:

```
1 import matplotlib.pyplot as plt
```

```
In [169]: 1 nltk.download('vader_lexicon')
          2
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data] /home/punit/nltk_data...
```

Out[169]: True

```
In [170]: 1 nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /home/punit/nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

Out[170]: True

```
In [171]: 1 nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to /home/punit/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

Out[171]: True

```
In [172]: 1 nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data] /home/punit/nltk_data...
[nltk_data] Package averaged_perceptron_tagger is already up-to-
[nltk_data] date!
```

Out[172]: True

```
In [173]: 1 nltk.download('wordnet')
```

```
[nltk_data] Downloading package wordnet to /home/punit/nltk_data...
[nltk_data] Package wordnet is already up-to-date!
```

Out[173]: True

```
In [ ]: 1
```

```
In [174]: 1 df = pd.read_csv('Womens Clothing E-Commerce Reviews.csv', index_col=0)
```

```
In [175]: 1 df.head()
```

```
Out[175]:
```

	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name
0	767	33	NaN	Absolutely wonderful - silky and sexy and comf...	4	1	0	Initmates	Intimate	Intimates
1	1080	34	NaN	Love this dress! it's sooo pretty. i happene...	5	1	4	General	Dresses	Dresses
2	1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	0	0	General	Dresses	Dresses
3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1	0	General Petite	Bottoms	Pants
4	847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	1	6	General	Tops	Blouses

```
In [177]: 1 df.columns
```

```
Out[177]: Index(['Clothing ID', 'Age', 'Title', 'Review Text', 'Rating',
                  'Recommended IND', 'Positive Feedback Count', 'Division Name',
                  'Department Name', 'Class Name'],
                  dtype='object')
```

```
In [ ]: 1
```

## EDA

```
In [179]: 1 df['Division Name'].unique()
```

```
Out[179]: array(['Initmates', 'General', 'General Petite', nan], dtype=object)
```

```
In [180]: 1 df['Department Name'].unique()
```

```
Out[180]: array(['Intimate', 'Dresses', 'Bottoms', 'Tops', 'Jackets', 'Trend', nan],  
              dtype=object)
```

```
In [ ]: 1
```

```
In [181]: 1 # to check for null values  
          2 df.isnull().sum()
```

```
Out[181]: Clothing ID          0  
          Age                0  
          Title             3810  
          Review Text       845  
          Rating             0  
          Recommended IND    0  
          Positive Feedback Count 0  
          Division Name      14  
          Department Name     14  
          Class Name         14  
          dtype: int64
```

```
In [182]: 1 len(df)
```

```
Out[182]: 23486
```

```
In [183]: 1 845/23486
```

```
Out[183]: 0.035978881035510515
```

## handling null value

```
In [184]: 1 df_review = df.copy()  
          2  
          3 df_review.dropna(axis=0,inplace=True)
```

```
In [186]: 1 df_review.isnull().sum()
```

```
Out[186]: Clothing ID          0
Age                          0
Title                        0
Review Text                  0
Rating                       0
Recommended IND              0
Positive Feedback Count      0
Division Name                0
Department Name              0
Class Name                   0
dtype: int64
```

```
In [ ]: 1
```

```
In [187]: 1 df_review.describe()
```

```
Out[187]:
```

	Clothing ID	Age	Rating	Recommended IND	Positive Feedback Count
<b>count</b>	19662.000000	19662.000000	19662.000000	19662.000000	19662.000000
<b>mean</b>	921.297274	43.260808	4.183145	0.818177	2.652477
<b>std</b>	200.227528	12.258122	1.112224	0.385708	5.834285
<b>min</b>	1.000000	18.000000	1.000000	0.000000	0.000000
<b>25%</b>	861.000000	34.000000	4.000000	1.000000	0.000000
<b>50%</b>	936.000000	41.000000	5.000000	1.000000	1.000000
<b>75%</b>	1078.000000	52.000000	5.000000	1.000000	3.000000
<b>max</b>	1205.000000	99.000000	5.000000	1.000000	122.000000

```
In [188]: 1 df_review.dtypes
```

```
Out[188]: Clothing ID          int64
Age              int64
Title            object
Review Text      object
Rating           int64
Recommended IND  int64
Positive Feedback Count  int64
Division Name    object
Department Name  object
Class Name       object
dtype: object
```

```
In [189]: 1 df_review['Clothing ID'].nunique()
```

```
Out[189]: 1095
```

```
In [190]: 1 df_review.head()
```

```
Out[190]:
```

	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name
2	1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	0	0	General	Dresses	Dresses
3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1	0	General Petite	Bottoms	Pants
4	847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	1	6	General	Tops	Blouses
5	1080	49	Not for the very petite	I love tracy reese dresses, but this one is no...	2	0	4	General	Dresses	Dresses
6	858	39	Cagrcoal shimmer fun	I aded this in my basket at hte last mintue to...	5	1	1	General Petite	Tops	Knits

```
In [193]: 1 df_review.shape
```

```
Out[193]: (19662, 10)
```

```
In [194]: 1 df.shape
```

```
Out[194]: (23486, 10)
```

- 1) The number of unique value in clothing id is significantly lesser than the number of rows.
- 2) This shows that clothing id are not unique. This is understandable given that clothing id is the product that is being reviewed
- 3) Hence this shows the same product was purchased and reviewed by multiple customers.
- 4) However, this also means that we can find out what are popular items (those appear more than 'n' times)
- 5) Lets set an arbitrary random number 'n' to be 100

```
In [198]: 1 famous_cloth_id = df_review['Clothing ID'].value_counts()
```

```
In [199]: 1 famous_cloth_id
```

```
Out[199]: 1078    871
          862    658
          1094   651
          1081   487
          829   452
          ...
          887     1
          48      1
          1127    1
          1143    1
          1199    1
          Name: Clothing ID, Length: 1095, dtype: int64
```

```
In [200]: 1 famous_cloth_id[famous_cloth_id>100]
```

```
Out[200]: 1078    871
          862    658
          1094   651
          1081   487
          829   452
          872   450
          1110   419
          868   370
          895   336
          867   291
          936   289
          1095   287
          850   280
          1077   251
          1059   247
          863   243
          1080   241
          1086   241
          860   237
          1083   214
          861   203
          873   197
          1033   190
          1092   187
          927   187
          828   181
          1056   180
          820   177
          1022   172
          836   172
          1072   166
          1008   163
          1104   147
          984   144
          831   138
          877   133
          1020   133
          833   132
          854   130
```



834	125
864	125
835	121
1082	119
1035	115
1099	115
940	113
1087	109
865	108
907	106
909	102
875	101

Name: Clothing ID, dtype: int64

```
In [201]: 1 len(famous_cloth_id[famous_cloth_id>100])
```

```
Out[201]: 51
```

```
In [ ]: 1
```

In [202]: 1 df\_review

Out[202]:

	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name
2	1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	0	0	General	Dresses	Dresses
3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1	0	General Petite	Bottoms	Pants
4	847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	1	6	General	Tops	Blouses
5	1080	49	Not for the very petite	I love tracy reese dresses, but this one is no...	2	0	4	General	Dresses	Dresses
6	858	39	Cagrcol shimmer fun	I aded this in my basket at hte last mintue to...	5	1	1	General Petite	Tops	Knits
...	...	...	...	...	...	...	...	...	...	...
23481	1104	34	Great dress for many occasions	I was very happy to snag this dress at such a ...	5	1	0	General Petite	Dresses	Dresses
23482	862	48	Wish it was made of cotton	It reminds me of maternity clothes. soft, stre...	3	1	0	General Petite	Tops	Knits
23483	1104	31	Cute, but see through	This fit well, but the top was very see throug...	3	0	1	General Petite	Dresses	Dresses
23484	1084	28	Very cute dress, perfect for summer parties an...	I bought this dress for a wedding i have this ...	3	1	2	General	Dresses	Dresses
23485	1104	52	Please make more like this one!	This dress in a lovely platinum is feminine an...	5	1	22	General Petite	Dresses	Dresses

19662 rows × 10 columns

In [ ]: 1

The age group of reviewers spans over wide range of age, we would now categorize them into separate bins (10-20] , (20-30] and so on ..

This enables subsequent analysis to be more meaningful because based on sentiments and ratings we can group according to age

```
In [203]: 1 bins = np.arange(0,100,10)
```

```
In [204]: 1 bins
```

```
Out[204]: array([ 0, 10, 20, 30, 40, 50, 60, 70, 80, 90])
```

```
In [205]: 1 df_review['Age group'] = pd.cut(df_review['Age'],bins)
```

```
In [206]: 1 df_review.head()
```

```
Out[206]:
```

	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name	Age group
2	1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	0	0	General	Dresses	Dresses	(50, 60]
3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1	0	General Petite	Bottoms	Pants	(40, 50]
4	847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	1	6	General	Tops	Blouses	(40, 50]
5	1080	49	Not for the very petite	I love tracy reese dresses, but this one is no...	2	0	4	General	Dresses	Dresses	(40, 50]
6	858	39	Cagrc coal shimmer fun	I aded this in my basket at hte last mintue to...	5	1	1	General Petite	Tops	Knits	(30, 40]

```
In [ ]: 1
```

## Plotting

```
In [208]: 1 ratings_count_df = df_review.groupby(['Rating',df_review['Age group']]).size().reset_index(name='n')
```

In [209]: 1 ratings\_count\_df

Out[209]:

	Rating	Age group	n
0	1	(0, 10]	0
1	1	(10, 20]	4
2	1	(20, 30]	97
3	1	(30, 40]	229
4	1	(40, 50]	170
5	1	(50, 60]	121
6	1	(60, 70]	60
7	1	(70, 80]	8
8	1	(80, 90]	1
9	2	(0, 10]	0
10	2	(10, 20]	3
11	2	(20, 30]	182
12	2	(30, 40]	488
13	2	(40, 50]	372
14	2	(50, 60]	206
15	2	(60, 70]	86
16	2	(70, 80]	18
17	2	(80, 90]	4
18	3	(0, 10]	0
19	3	(10, 20]	15
20	3	(20, 30]	375
21	3	(30, 40]	856
22	3	(40, 50]	645

	Rating	Age group	n
23	3	(50, 60]	365
24	3	(60, 70]	179
25	3	(70, 80]	24
26	3	(80, 90]	4
27	4	(0, 10]	0
28	4	(10, 20]	31
29	4	(20, 30]	569
30	4	(30, 40]	1460
31	4	(40, 50]	1138
32	4	(50, 60]	692
33	4	(60, 70]	349
34	4	(70, 80]	43
35	4	(80, 90]	6
36	5	(0, 10]	0
37	5	(10, 20]	70
38	5	(20, 30]	1460
39	5	(30, 40]	3493
40	5	(40, 50]	2693
41	5	(50, 60]	1886
42	5	(60, 70]	1048
43	5	(70, 80]	145
44	5	(80, 90]	59

In [211]: 1 rating\_pivot = pd.pivot\_table(ratings\_count\_df, index=['Age group'], values=['n'], columns=['Rating'], aggfu

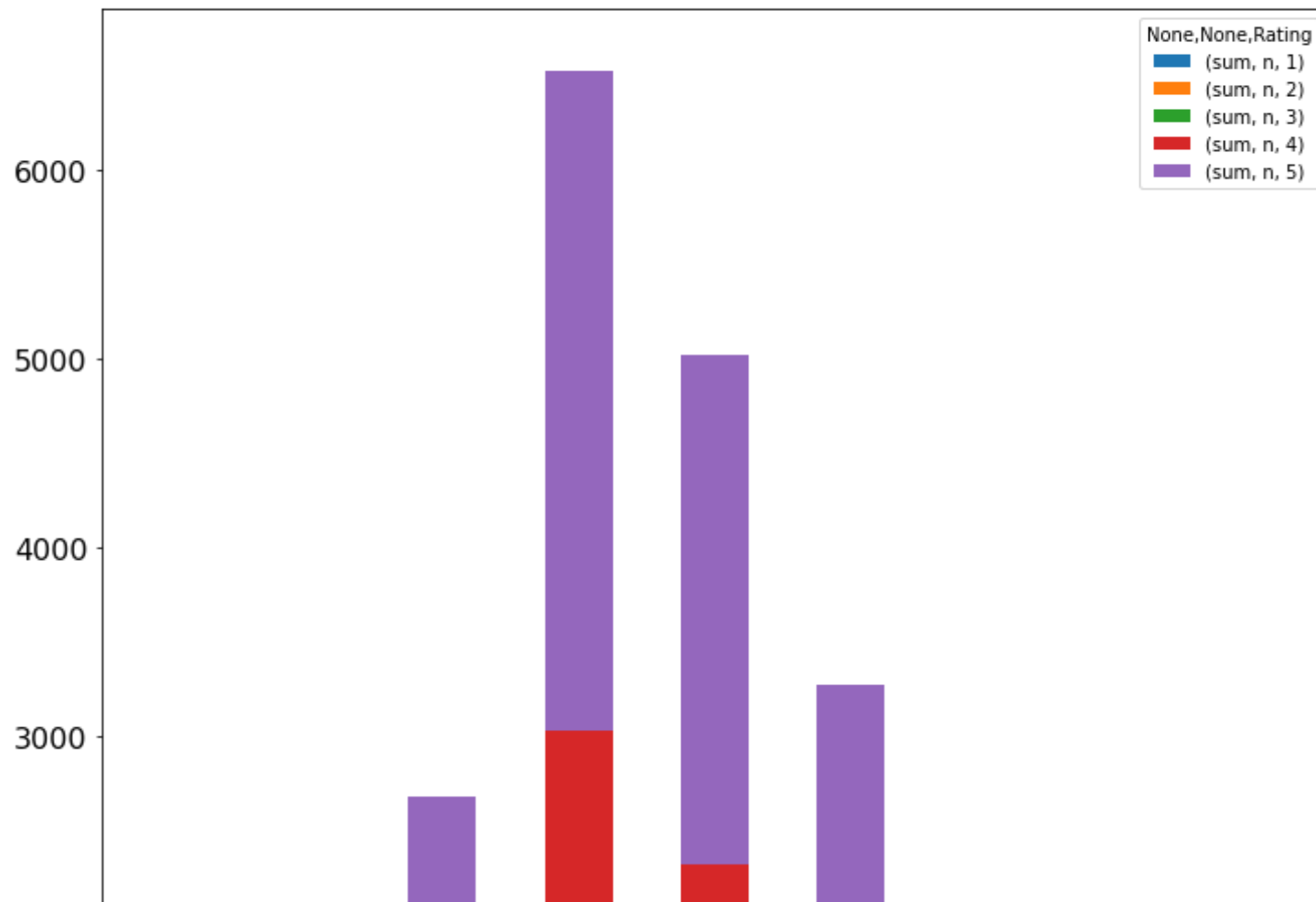
In [212]: 1 rating\_pivot

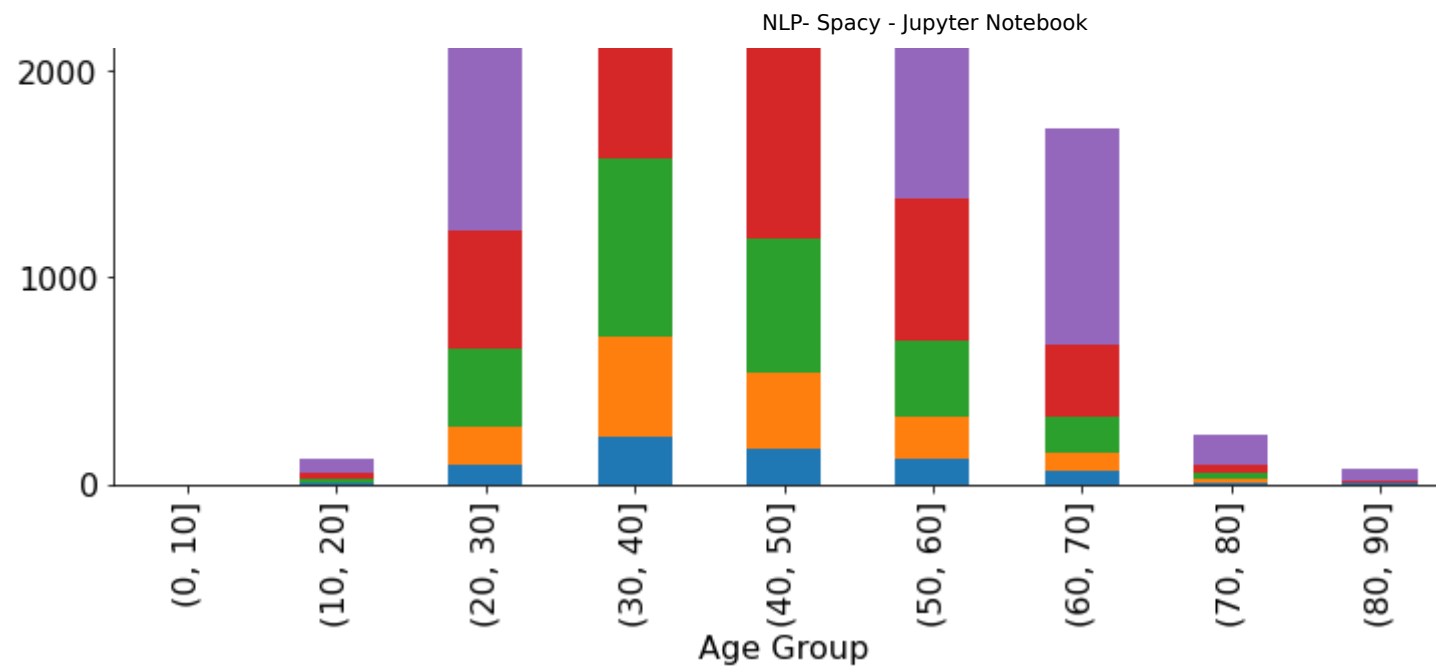
Out[212]:

Rating	sum				
	n				
Age group	1	2	3	4	5
(0, 10]	0	0	0	0	0
(10, 20]	4	3	15	31	70
(20, 30]	97	182	375	569	1460
(30, 40]	229	488	856	1460	3493
(40, 50]	170	372	645	1138	2693
(50, 60]	121	206	365	692	1886
(60, 70]	60	86	179	349	1048
(70, 80]	8	18	24	43	145
(80, 90]	1	4	4	6	59

```
In [220]: 1 fig = plt.figure(figsize=(12,13))
2 ax = fig.add_subplot(111)
3 rating_pivot.plot(kind='bar',stacked=True,fontsize=16,ax=ax)
4 ax.set_xlabel("Age Group",fontsize=16)
5 fig.suptitle("Stacked bar chart of ratings and cross age groups",fontsize=20)
6 plt.show()
```

Stacked bar chart of ratings and cross age groups





In [ ]:

1



In [221]: 1 df\_review

Out[221]:

	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name	Age group
2	1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	0	0	General	Dresses	Dresses	(50, 60]
3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1	0	General Petite	Bottoms	Pants	(40, 50]
4	847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	1	6	General	Tops	Blouses	(40, 50]
5	1080	49	Not for the very petite	I love tracy reese dresses, but this one is no...	2	0	4	General	Dresses	Dresses	(40, 50]
6	858	39	Cagrcoal shimmer fun	I aded this in my basket at hte last mintue to...	5	1	1	General Petite	Tops	Knits	(30, 40]
...	...	...	...	...	...	...	...	...	...	...	...
23481	1104	34	Great dress for many occasions	I was very happy to snag this dress at such a ...	5	1	0	General Petite	Dresses	Dresses	(30, 40]
23482	862	48	Wish it was made of cotton	It reminds me of maternity clothes. soft, stre...	3	1	0	General Petite	Tops	Knits	(40, 50]
23483	1104	31	Cute, but see through	This fit well, but the top was very see throug...	3	0	1	General Petite	Dresses	Dresses	(30, 40]
23484	1084	28	Very cute dress, perfect for summer parties an...	I bought this dress for a wedding i have this ...	3	1	2	General	Dresses	Dresses	(20, 30]
23485	1104	52	Please make more like this one!	This dress in a lovely platinum is feminine an...	5	1	22	General Petite	Dresses	Dresses	(50, 60]

19662 rows × 11 columns

```
In [222]: 1 import re
```

```
In [240]: 1 stopwords.words()
```

```
, 'ci'  
, 'c1'  
, 'cox'  
, 'cu'  
, 'cü'  
, 'çünki'  
, 'da'  
, 'daha'  
, 'də'  
, 'dedi'  
, 'dək'  
, 'dən'  
, 'dəqiqə'  
, 'deyil'  
, 'dir'  
, 'doqquz'  
, 'doqsan'  
, 'dörd'  
, 'düz'  
, 'ə'
```

## Text normalisation on Review Text

```

In [234]: 1 def clean_string(s):
2         r1 = re.sub('[^A-Za-z]+', ' ', str(s))
3         r1 = r1.strip().lower()
4
5         # removing stopwords
6
7         no_stp_wrd = []
8         if r1 not in stopwords.words():
9             no_stp_wrd.append(r1)
10
11        # performing stemming
12        stemmer = PorterStemmer()
13        final_list = []
14
15        for i in no_stp_wrd:
16            final_list.append(stemmer.stem(str(i)))
17
18        return final_list

```

```

In [225]: 1 reviews_df = df_review.copy()

```

```

In [226]: 1 reviews_df.head()

```

Out[226]:

	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name	Age group
2	1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	0	0	General	Dresses	Dresses	(50, 60]
3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1	0	General Petite	Bottoms	Pants	(40, 50]
4	847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	1	6	General	Tops	Blouses	(40, 50]
5	1080	49	Not for the very petite	I love tracy reese dresses, but this one is no...	2	0	4	General	Dresses	Dresses	(40, 50]
6	858	39	Cagrc coal shimmer fun	I aded this in my basket at hte last mintue to...	5	1	1	General Petite	Tops	Knits	(30, 40]

```
In [227]: 1 reviews_df = reviews_df[['Clothing ID', 'Review Text', 'Age group', 'Recommended IND']]
```

```
In [228]: 1 reviews_df
```

Out[228]:

	Clothing ID	Review Text	Age group	Recommended IND
2	1077	I had such high hopes for this dress and reall...	(50, 60]	0
3	1049	I love, love, love this jumpsuit. it's fun, fl...	(40, 50]	1
4	847	This shirt is very flattering to all due to th...	(40, 50]	1
5	1080	I love tracy reese dresses, but this one is no...	(40, 50]	0
6	858	I aded this in my basket at hte last mintue to...	(30, 40]	1
...	...	...	...	...
23481	1104	I was very happy to snag this dress at such a ...	(30, 40]	1
23482	862	It reminds me of maternity clothes. soft, stre...	(40, 50]	1
23483	1104	This fit well, but the top was very see throug...	(30, 40]	0
23484	1084	I bought this dress for a wedding i have this ...	(20, 30]	1
23485	1104	This dress in a lovely platinum is feminine an...	(50, 60]	1

19662 rows × 4 columns

```
In [235]: 1 reviews_df['Filtered Review'] = reviews_df['Review Text'].apply(clean_string)
```

In [236]: 1 reviews\_df

Out[236]:

	Clothing ID	Review Text	Age group	Recommended IND	Filtered Review
2	1077	I had such high hopes for this dress and reall...	(50, 60]	0	[i had such high hopes for this dress and real...
3	1049	I love, love, love this jumpsuit. it's fun, fl...	(40, 50]	1	[i love love love this jumpsuit it s fun flirt...
4	847	This shirt is very flattering to all due to th...	(40, 50]	1	[this shirt is very flattering to all due to t...
5	1080	I love tracy reese dresses, but this one is no...	(40, 50]	0	[i love tracy reese dresses but this one is no...
6	858	I aded this in my basket at hte last mintue to...	(30, 40]	1	[i aded this in my basket at hte last mintue t...
...	...	...	...	...	...
23481	1104	I was very happy to snag this dress at such a ...	(30, 40]	1	[i was very happy to snag this dress at such a...
23482	862	It reminds me of maternity clothes. soft, stre...	(40, 50]	1	[it reminds me of maternity clothes soft stret...
23483	1104	This fit well, but the top was very see throug...	(30, 40]	0	[this fit well but the top was very see throug...
23484	1084	I bought this dress for a wedding i have this ...	(20, 30]	1	[i bought this dress for a wedding i have this...
23485	1104	This dress in a lovely platinum is feminine an...	(50, 60]	1	[this dress in a lovely platinum is feminine a...

19662 rows × 5 columns

In [238]: 1 reviews\_df['Filtered Review']

Out[238]:

```

2      [i had such high hopes for this dress and real...
3      [i love love love this jumpsuit it s fun flirt...
4      [this shirt is very flattering to all due to t...
5      [i love tracy reese dresses but this one is no...
6      [i aded this in my basket at hte last mintue t...
      ...
23481  [i was very happy to snag this dress at such a...
23482  [it reminds me of maternity clothes soft stret...
23483  [this fit well but the top was very see throug...
23484  [i bought this dress for a wedding i have this...
23485  [this dress in a lovely platinum is feminine a...
Name: Filtered Review, Length: 19662, dtype: object

```

```
In [239]: 1 reviews_df['Review Text']
```

```
Out[239]: 2      I had such high hopes for this dress and reall...
3      I love, love, love this jumpsuit. it's fun, fl...
4      This shirt is very flattering to all due to th...
5      I love tracy reese dresses, but this one is no...
6      I aded this in my basket at hte last mintue to...
...
23481   I was very happy to snag this dress at such a ...
23482   It reminds me of maternity clothes. soft, stre...
23483   This fit well, but the top was very see throug...
23484   I bought this dress for a wedding i have this ...
23485   This dress in a lovely platinum is feminine an...
Name: Review Text, Length: 19662, dtype: object
```

```
In [241]: 1 reviews_df.head()
```

```
Out[241]:
```

	Clothing ID	Review Text	Age group	Recommended IND	Filtered Review
2	1077	I had such high hopes for this dress and reall...	(50, 60]	0	[i had such high hopes for this dress and real...
3	1049	I love, love, love this jumpsuit. it's fun, fl...	(40, 50]	1	[i love love love this jumpsuit it s fun flirt...
4	847	This shirt is very flattering to all due to th...	(40, 50]	1	[this shirt is very flattering to all due to t...
5	1080	I love tracy reese dresses, but this one is no...	(40, 50]	0	[i love tracy reese dresses but this one is no...
6	858	I aded this in my basket at hte last mintue to...	(30, 40]	1	[i aded this in my basket at hte last mintue t...

```
In [ ]: 1
```

## Sentiment Analysis on Filtered Review column

```
In [242]: 1 sia = SentimentIntensityAnalyzer()
```

```
In [255]: 1 num_records = reviews_df.shape[0]
```

In [256]: 1 num\_records

Out[256]: 19662

In [245]: 1 ones\_array = np.ones(num\_records)

In [246]: 1 ones\_array

Out[246]: array([1., 1., 1., ..., 1., 1., 1.])

In [247]: 1 reviews\_df['Compound Score'] = ones\_array

In [259]: 1 reviews\_df.head()

Out[259]:

	Clothing ID	Review Text	Age group	Recommended IND	Filtered Review	Compound Score
2	1077	I had such high hopes for this dress and reall...	(50, 60]	0	[i had such high hopes for this dress and real...	0.9398
3	1049	I love, love, love this jumpsuit. it's fun, fl...	(40, 50]	1	[i love love love this jumpsuit it s fun flirt...	0.7192
4	847	This shirt is very flattering to all due to th...	(40, 50]	1	[this shirt is very flattering to all due to t...	0.9162
5	1080	I love tracy reese dresses, but this one is no...	(40, 50]	0	[i love tracy reese dresses but this one is no...	0.9419
6	858	I aded this in my basket at hte last mintue to...	(30, 40]	1	[i aded this in my basket at hte last mintue t...	0.4576

```
In [264]: 1 reviews_df['Filtered Review']
```

```
Out[264]: 2      [i had such high hopes for this dress and real...
3      [i love love love this jumpsuit it s fun flirt...
4      [this shirt is very flattering to all due to t...
5      [i love tracy reese dresses but this one is no...
6      [i aded this in my basket at hte last mintue t...
...
23481  [i was very happy to snag this dress at such a...
23482  [it reminds me of maternity clothes soft stret...
23483  [this fit well but the top was very see throug...
23484  [i bought this dress for a wedding i have this...
23485  [this dress in a lovely platinum is feminine a...
Name: Filtered Review, Length: 19662, dtype: object
```

```
In [ ]: 1
```

```
In [257]: 1 len(reviews_df['Filtered Review'])
```

```
Out[257]: 19662
```

```
In [267]: 1 for i in range(len(reviews_df['Filtered Review'])):
2     element = reviews_df['Filtered Review'].iloc[i][0]
3     #     print(element)
4     #     break
5     score = sia.polarity_scores(element)
6     #     print(score)
7     compound = score['compound']
8     reviews_df['Compound Score'].iloc[i] = compound
```

/home/punit/anaconda3/lib/python3.8/site-packages/pandas/core/indexing.py:1637: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy))  
self.\_setitem\_single\_block(indexer, value, name)



In [268]: 1 reviews\_df.head()

Out[268]:

	Clothing ID	Review Text	Age group	Recommended IND	Filtered Review	Compound Score
2	1077	I had such high hopes for this dress and real...	(50, 60]	0	[i had such high hopes for this dress and real...	0.9398
3	1049	I love, love, love this jumpsuit. it's fun, fl...	(40, 50]	1	[i love love love this jumpsuit it s fun flirt...	0.7192
4	847	This shirt is very flattering to all due to th...	(40, 50]	1	[this shirt is very flattering to all due to t...	0.9162
5	1080	I love tracy reese dresses, but this one is no...	(40, 50]	0	[i love tracy reese dresses but this one is no...	0.9419
6	858	I aded this in my basket at hte last mintue to...	(30, 40]	1	[i aded this in my basket at hte last mintue t...	0.4576

In [ ]:

1

In [269]:

```

1 category = []
2 for i in reviews_df['Compound Score']:
3     if i > 0:
4         category.append("Positive")
5     elif i < 0:
6         category.append("Negative")
7     else:
8         category.append("Neutral")
9
10 reviews_df['Sentiment_Category'] = category

```

In [270]: 1 reviews\_df.head()

Out[270]:

	Clothing ID	Review Text	Age group	Recommended IND	Filtered Review	Compound Score	Sentiment_Category
2	1077	I had such high hopes for this dress and reall...	(50, 60]	0	[i had such high hopes for this dress and real...	0.9398	Positive
3	1049	I love, love, love this jumpsuit. it's fun, fl...	(40, 50]	1	[i love love love this jumpsuit it s fun flirt...	0.7192	Positive
4	847	This shirt is very flattering to all due to th...	(40, 50]	1	[this shirt is very flattering to all due to t...	0.9162	Positive
5	1080	I love tracy reese dresses, but this one is no...	(40, 50]	0	[i love tracy reese dresses but this one is no...	0.9419	Positive
6	858	I aded this in my basket at hte last mintue to...	(30, 40]	1	[i aded this in my basket at hte last mintue t...	0.4576	Positive

In [271]: 1 reviews\_df.tail()

Out[271]:

	Clothing ID	Review Text	Age group	Recommended IND	Filtered Review	Compound Score	Sentiment_Category
23481	1104	I was very happy to snag this dress at such a ...	(30, 40]	1	[i was very happy to snag this dress at such a...	0.9100	Positive
23482	862	It reminds me of maternity clothes. soft, stre...	(40, 50]	1	[it reminds me of maternity clothes soft stret...	0.6652	Positive
23483	1104	This fit well, but the top was very see through...	(30, 40]	0	[this fit well but the top was very see through...	0.9343	Positive
23484	1084	I bought this dress for a wedding i have this ...	(20, 30]	1	[i bought this dress for a wedding i have this...	0.8198	Positive
23485	1104	This dress in a lovely platinum is feminine an...	(50, 60]	1	[this dress in a lovely platinum is feminine a...	0.9286	Positive

In [272]: 1 reviews\_df.iloc[100:200]

Out[272]:

	Clothing ID	Review Text	Age group	Recommended IND	Filtered Review	Compound Score	Sentiment_Category
132	861	In my retailer this was hung over in the pj se...	(40, 50]	1	[in my retailer this was hung over in the pj s...	0.9675	Positive
133	966	I love this vest! there are so many ways to st...	(30, 40]	1	[i love this vest there are so many ways to st...	0.8720	Positive
134	845	This blouse is a perfect creation: perfect-wei...	(60, 70]	1	[this blouse is a perfect creation perfect wei...	0.9725	Positive
136	966	Love this vest! the color looks a little more ...	(30, 40]	1	[love this vest the color looks a little more ...	0.9791	Positive
137	836	I received the sky color in m online. am defi...	(50, 60]	1	[i received the sky color in m online am defin...	0.9623	Positive
...	...	...	...	...	...	...	...
250	1026	I love these pants. i have worn them a number ...	(40, 50]	1	[i love these pants i have worn them a number ...	0.6826	Positive
251	840	I love this tunic the natural color is just th...	(30, 40]	1	[i love this tunic the natural color is just t...	0.9401	Positive
252	1066	These cropped pants are very light weight and ...	(20, 30]	1	[these cropped pants are very light weight and...	0.7845	Positive
253	647	I have this dress on today in white and i am c...	(30, 40]	1	[i have this dress on today in white and i am ...	-0.0209	Negative
255	840	Cute, swing top that would be flattering on mo...	(50, 60]	1	[cute swing top that would be flattering on mo...	0.7748	Positive

100 rows × 7 columns

In [ ]:

1

