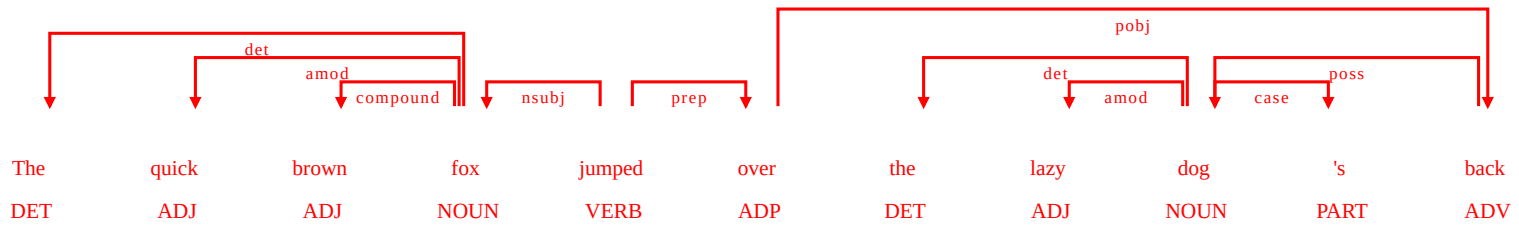
In [72]:
```
for token in sentence:
    print(f'{token.text:{10}} {token.pos_:{7}} {token.dep_:{7}} {spacy.explain(token.dep_)}')
```

```
The        DET     det     determiner
quick      ADJ     amod    adjectival modifier
brown      ADJ     compound compound
fox        NOUN    nsubj   nominal subject
jumped     VERB    ROOT    None
over       ADP     prep    prepositional modifier
the        DET     det     determiner
lazy       ADJ     amod    adjectival modifier
dog        NOUN    poss    possession modifier
's         PART    case    case marking
back       ADV     pobj    object of preposition
```

In [73]:
```
options = {'distance':95,'compact':True,'color':'red','bg':'#09a3e0','font':'Times'}
displacy.render(sentence,style='dep',jupyter=True,options=options)
```



In [ ]:
```

```

In [ ]:
```

```

Nave Baye's Classifier

In [ ]:
```

```

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 [74]:
```
import pandas as pd
import numpy as np
```

In [75]:
```
import seaborn as sns
import nltk
import re
from nltk.stem import PorterStemmer,WordNetLemmatizer
from nltk import pos_tag,wordnet
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords

from nltk.sentiment.vader import SentimentIntensityAnalyzer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from wordcloud import WordCloud
```

In [76]:
```
import matplotlib.pyplot as plt
```

In [77]:
```
nltk.download('vader_lexicon')

```

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

Out[77]:
```
True
```

In [78]:
```
nltk.download('punkt')
```

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

Out[78]:
```
True
```

In [79]:
```
nltk.download('stopwords')
```

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

Out[79]:
```
True
```

In [80]:
```
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[80]:
```
True
```

In [81]:
```
1 nltk.download('wordnet')
```

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

Out[81]: True

In [ ]:
```
1
```

In [82]:
```
1 df = pd.read_csv('Womens Clothing E-Commerce Reviews.csv',index_col=0)
```

In [83]:
```
1 df.head()
```

Out[83]:

	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 [84]:
```
1 df.columns
```

Out[84]:
```
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 [85]:
```
1 df['Division Name'].unique()
```

Out[85]: array(['Initmates', 'General', 'General Petite', nan], dtype=object)

In [86]:
```
1 df['Department Name'].unique()
```

Out[86]:
```
array(['Intimate', 'Dresses', 'Bottoms', 'Tops', 'Jackets', 'Trend', nan],
      dtype=object)
```

In [ ]:
```
1
```

In [87]:
```
# to check for null values
df.isnull().sum()
```

Out[87]:
```
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 [88]:
```
len(df)
```

Out[88]:
```
23486
```

In [89]:
```
845/23486
```

Out[89]:
```
0.035978881035510515
```

handling null value

In [90]:
```
df_review = df.copy()

df_review.dropna(axis=0,inplace=True)
```

In [91]:
```
df_review.isnull().sum()
```

Out[91]:
```
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 [ ]:
```

```

In [92]:
```
1 df_review.describe()
```

Out[92]:

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

In [93]:
```
1 df_review.dtypes
```

Out[93]:
```
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 [94]:
```
1 df_review['Clothing ID'].nunique()
```

Out[94]:
```
1095
```

In [95]:
```
1 df_review.head()
```

Out[95]:

	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 [96]:
```
1 df_review.shape
```

Out[96]:
```
(19662, 10)
```

In [97]:
```
1 df.shape
```

Out[97]:
```
(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 reviewd

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 arbitary random number 'n' to be 100

In [98]:
```
1 famous_cloth_id = df_review['Clothing ID'].value_counts()
```

In [99]:
```
1 famous_cloth_id
```

Out[99]:
```
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 [100]:
```
1 famous_cloth_id[famous_cloth_id>100]
```

Out[100]:
```
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 [101]:
```
len(famous_cloth_id[famous_cloth_id>100])
```

Out[101]: 51

In [ ]:
```

```

In [102]:
```
df_review
```

Out[102]:

	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
...	...	...	...	...	...	...	...	...	...	...
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 [ ]:
```

```

The age group of reviewers spans over wide range of age, we would now categorize them into seperate 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 accodrding to age

In [103]:
```
bins = np.arange(0,100,10)
```

In [104]:
```
bins
```

Out[104]:
```
array([ 0, 10, 20, 30, 40, 50, 60, 70, 80, 90])
```

In [105]:
```
df_review['Age group'] = pd.cut(df_review['Age'],bins)
```

In [106]:
```
df_review.head()
```

Out[106]:

	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]

In [ ]:
```

```

Plotting

In [107]:
```
ratings_count_df = df_review.groupby(['Rating',df_review['Age group']]).size().reset_index(name='n')
```

In [108]:
```
1 ratings_count_df
```

Out[108]:

	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
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

	Rating	Age group	n
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 [109]:
```
1 rating_pivot = pd.pivot_table(ratings_count_df,index=['Age group'],values=['n'],columns=['Rating'],aggfunc=[np.sum])
```

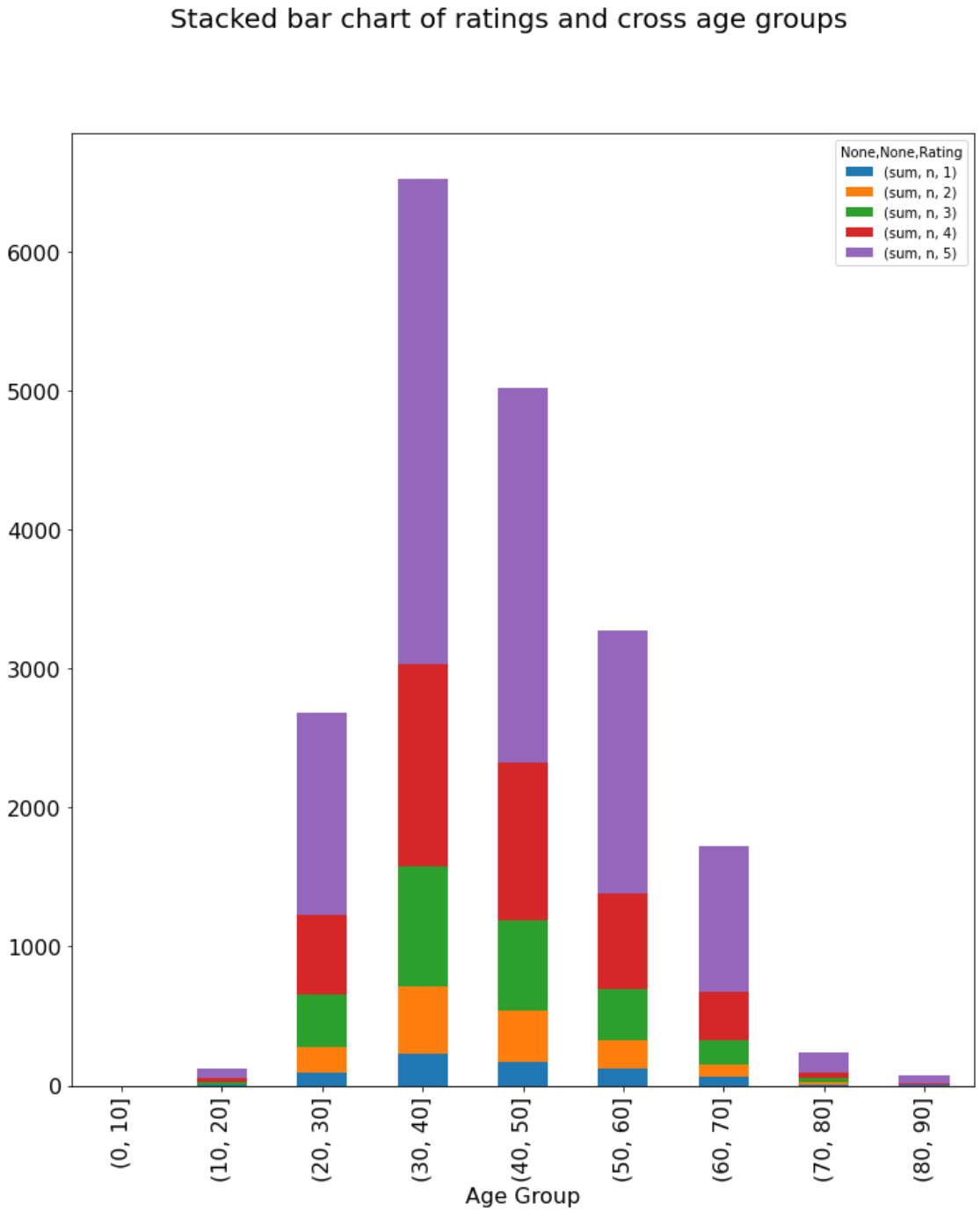
In [110]:
```
1 rating_pivot
```

Out[110]:

	sum				
	n				
Rating	**1**	**2**	**3**	**4**	**5**
Age group					
(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 [111]:

```
fig = plt.figure(figsize=(12,13))
ax = fig.add_subplot(111)
rating_pivot.plot(kind='bar',stacked=True,fontsize=16,ax=ax)
ax.set_xlabel("Age Group",fontsize=16)
fig.suptitle("Stacked bar chart of ratings and cross age groups",fontsize=20)
plt.show()
```


Stacked bar chart of ratings and cross age groups
None,None,Rating
(sum, n, 1)
(sum, n, 2)
(sum, n, 3)
(sum, n, 4)
(sum, n, 5)
6000
5000
4000
3000
2000
1000
0
(0, 10]
(10, 20]
(20, 30]
(30, 40]
(40, 50]
(50, 60]
(60, 70]
(70, 80]
(80, 90]
Age Group

In [ ]:
```
1
```

In [112]:
```
1 df_review
```

Out[112]:

	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 [113]:
```
1 import re
```

In [114]:
```
stopwords.words()
```

Out[114]:
```
,'إذ']
,'إذا'
,'إذما'
,'إذن'
,'أف'
,'أقل'
,'أكثر'
,'ألا'
,'إلا'
,'التي'
,'الذي'
,'الذين'
,'اللاتي'
,'اللائي'
,'اللتان'
,'اللتيا'
,'اللتين'
,'اللذان'
,'اللذين'
```

Text normalisation on Review Text

In [115]:
```
def clean_string(s):
    r1 = re.sub('[^A-Za-z]+',' ',str(s))
    r1 = r1.strip().lower()

    # removing stopwords

    no_stp_wrd = []
    if r1 not in stopwords.words():
        no_stp_wrd.append(r1)

    # performing stemming
    stemmer =  PorterStemmer()
    final_list = []

    for i in no_stp_wrd:
        final_list.append(stemmer.stem(str(i)))

    return final_list
```

In [116]:
```
reviews_df = df_review.copy()
```

In [117]:
```
1 reviews_df.head()
```

Out[117]:

	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]

In [118]:
```
1 reviews_df = reviews_df[['Clothing ID','Review Text','Age group','Recommended IND']]
```

In [119]:
```
1 reviews_df
```

Out[119]:

	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 [120]:
```
1 reviews_df['Filtered Review'] = reviews_df['Review Text'].apply(clean_string)
```

In [121]:
```
1 reviews_df
```

Out[121]:

	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 [122]:
```
1 reviews_df['Filtered Review']
```

Out[122]:
```
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 [123]:
```
1 reviews_df['Review Text']
```

Out[123]:
```
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 [124]:
```
1 reviews_df.head()
```

Out[124]:

	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 [125]:
```
1 sia = SentimentIntensityAnalyzer()
```

In [126]:
```
1 num_records = reviews_df.shape[0]
```

In [127]:
```
1 num_records
```

Out[127]: 19662

In [128]:
```
1 ones_array = np.ones(num_records)
```

In [129]:
```
1 ones_array
```

Out[129]: array([1., 1., 1., ..., 1., 1., 1.])

In [130]:
```
1 reviews_df['Compound Score'] = ones_array
```

In [131]:
```
1 reviews_df.head()
```

Out[131]:

	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...	1.0
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...	1.0
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...	1.0
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...	1.0
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...	1.0

In [132]:
```
reviews_df['Filtered Review']
```

Out[132]:
```
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 [ ]:
```

```

In [133]:
```
len(reviews_df['Filtered Review'])
```

Out[133]:
```
19662
```

In [134]:
```
for i in range(len(reviews_df['Filtered Review'])):
    element = reviews_df['Filtered Review'].iloc[i][0]
#     print(element)
#     break
    score = sia.polarity_scores(element)
#     print(score)
    compound = score['compound']
    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)
  self._setitem_single_block(indexer, value, name)
```

In [135]:
```
reviews_df.head()
```

Out[135]:

	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 [ ]:
```

```

In [136]:
```
category = []
for i in reviews_df['Compound Score']:
    if i > 0:
        category.append("Positive")
    elif i < 0:
        category.append("Negative")
    else:
        category.append("Neutral")

reviews_df['Sentiment_Category'] = category
```

In [137]:
```
reviews_df.head()
```

Out[137]:

	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 [138]:
```
reviews_df.tail()
```

Out[138]:

	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 throug...	(30, 40]	0	[this fit well but the top was very see throug...	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 [139]:
```
1 reviews_df.iloc[100:200]
```

Out[139]:

	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
```

In [140]:
```
1 # calcualting the total of positive, negative and neutral sentiments
```

In [141]:
```
1 sentiment_df = reviews_df.copy()
```

In [142]:
```
1 sentiment_df.columns
```

Out[142]:
```
Index(['Clothing ID', 'Review Text', 'Age group', 'Recommended IND',
       'Filtered Review', 'Compound Score', 'Sentiment_Category'],
      dtype='object')
```

In [143]:
```
1 sentiment_df = sentiment_df[['Clothing ID','Sentiment_Category']]
```

In [144]:
```
1 sentiment_df
```

Out[144]:

	Clothing ID	Sentiment_Category
2	1077	Positive
3	1049	Positive
4	847	Positive
5	1080	Positive
6	858	Positive
...	...	...
23481	1104	Positive
23482	862	Positive
23483	1104	Positive
23484	1084	Positive
23485	1104	Positive

19662 rows × 2 columns

In [145]:
```
1 sentiment_df_agg = sentiment_df.groupby('Sentiment_Category').size().reset_index(name='No of Sentiments')
```

In [146]:
```
1 sentiment_df_agg
```

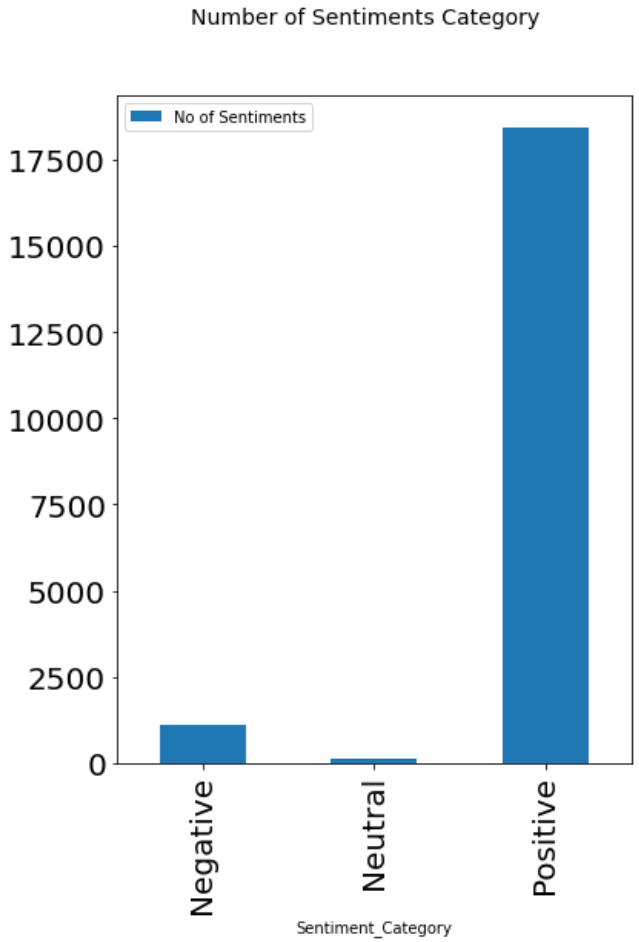
Out[146]:

	Sentiment_Category	No of Sentiments
0	Negative	1094
1	Neutral	121
2	Positive	18447

In [ ]:
```
1
```

In [152]:
```
fig,ax = plt.subplots()
sentiment_df_agg.plot(kind='bar',x='Sentiment_Category',y='No of Sentiments',figsize=(6,8),fontsize=20,ax=ax)
fig.suptitle('Number of Sentiments Category',fontsize=14)
```

Out[152]: Text(0.5, 0.98, 'Number of Sentiments Category')



In [ ]:
```
1
```

In [153]:
```
1 reviews_df.head()
```

Out[153]:

	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

There are huge amoount of positive reviews which we have figured out in the above process, We can find the list of all such words which shows the positive sentiments of the user for their purchase.

Use of word-cloud

In [159]:
```
1 positive_review = reviews_df.loc[reviews_df['Sentiment_Category'] == 'Positive', 'Filtered Review'].tolist()
```

In [172]:
```
1 positive_review
```

Out[172]:
```
[['i had such high hopes for this dress and really wanted it to work for me i initially ordered the petite small my usual size but i found this
to be outrageously small so small in fact that i could not zip it up i reordered it in petite medium which was just ok overall the top half was
comfortable and fit nicely but the bottom half had a very tight under layer and several somewhat cheap net over layers imo a major design flaw
was the net over layer sewn directly into the zipper it c'],
 ['i love love love this jumpsuit it s fun flirty and fabulous every time i wear it i get nothing but great compli'],
 ['this shirt is very flattering to all due to the adjustable front tie it is the perfect length to wear with leggings and it is sleeveless so
it pairs well with any cardigan love this shirt'],
 ['i love tracy reese dresses but this one is not for the very petite i am just under feet tall and usually wear a p in this brand this dress w
as very pretty out of the package but its a lot of dress the skirt is long and very full so it overwhelmed my small frame not a stranger to alt
erations shortening and narrowing the skirt would take away from the embellishment of the garment i love the color and the idea of the style bu
t it just did not work on me i returned this dress'],
 ['i aded this in my basket at hte last mintue to see what it would look like in person store pick up i went with teh darkler color only becaus
e i am so pale hte color is really gorgeous and turns out it mathced everythiing i was trying on with it prefectly it is a little baggy on me a
nd hte xs is hte msallet size bummer no petite i decided to jkeep it though because as i said it matvehd everything my ejans pants and the skir
ts i waas trying on of which i kept all oop'],
 ['i love this dress i usually get an xs but it runs a little snug in bust so i ordered up a size very flattering and feminine with the usual r
etailer flair for styl'],
 ['dress runs small esp where the zipper area runs i ordered the sp which typically fits me and it was very tight the material on the top looks
and feels very cheap that even just pulling on it will cause it to rip the fabric pretty disappointed as it was going to be my christmas dress
this year needless to say it will be going back']
```

In [165]:
```
for i in positive_review:
    print(i)
    print('\n')
```

```
material at my feet the gaps in the front are much wider than they look felt like the dress just fell flat both were returned im usually a larg
e and the med fit better d in jean']


['i have been waiting for this sweater coat to ship for weeks and i was so excited for it to arrive this coat is not true to size and made me l
ook short and squat the sleeves are very wide although long as a light weight fall coat the sleeves don t need to be as wide because you wouldn
t be layerng too much underneath the buttons need to be moved at least three inches in for a nicer fit i thought about redoing the buttons myse
lf but the sleeves looked even more out of proportion with a tigh']


['the colors weren t what i expected either the dark blue is much more vibrant and i just couldn t find anything to really go with it fabric is
thick and good quality has nice weight and movement to it the skirt just wasn t for me in the end']


['i have several of goodhyouman shirts and i get so many compliments on them especially the one that says forehead kisses are underrated don t
hesitate buy this shirt you won t be sorri']


['this sweater is so comfy and classic it balances a quirky hand knit look with a beautiful color and practical fit it is a bit cropped and box
```

In [166]:
```
pos_rev_desc = ''

for i in positive_review:
    for j in i:
        pos_rev_desc = pos_rev_desc + '' + j

print(pos_rev_desc)
```

```
IOPub data rate exceeded.
The notebook server will temporarily stop sending output
to the client in order to avoid crashing it.
To change this limit, set the config variable
`--NotebookApp.iopub_data_rate_limit`.

Current values:
NotebookApp.iopub_data_rate_limit=1000000.0 (bytes/sec)
NotebookApp.rate_limit_window=3.0 (secs)
```

In [175]:
```
fig = plt.figure(figsize=(13,15))

descr_word_cloud = WordCloud(width=900,height=400,background_color="white",
            max_words=450,relative_scaling=1.0).generate(pos_rev_desc)
plt.imshow(descr_word_cloud)
plt.axis("off")
plt.title('WordCloud of Positive Reviews',fontsize=15)
plt.show()
```

In [167]:
```
len(pos_rev_desc)
```

Out[167]: 5663913

In [ ]:
```

```

In [176]:
```
neg_review = reviews_df.loc[reviews_df['Sentiment_Category'] == 'Negative', 'Filtered Review'].tolist()
```

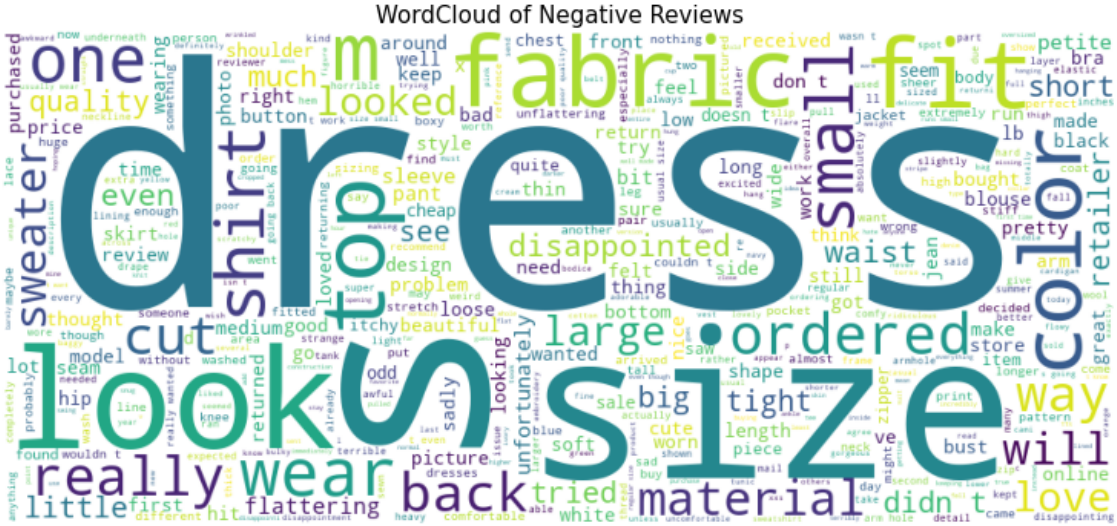
In [177]:
```
neg_review
```

```
were multiple snags along the seams additionally at lbs i was swimming in the size i would have tried sizing down if the quality was bett'],
 ['why is it so difficult to find a shirt that doesn t expose half of my bra if i lift my arm this is cute beyond the arm hole issue but overpr
iced going back'],
 ['this dress is most definitely not a maxi barely goes below the knees also you do not get the cute detail shown on bodice there are no folds
which were an added plus on picture waist is higher not at natural waist very pretty silk though and still cute just not as pictur'],
 ['too big and it s not secure enough to wear for swim'],
 ['i was so excited to receive these jeans typically wide leg jeans are a no brainer for me super flattering and in chambray i thought they wou
ld be stylish and comfortable but i found these to be verging on mom jean something about them seems not so cool maybe the color of the wash th
e weight of the fabric not sure but they don t lend themselves to current looks sadly i will be sending them back'],
 ['i found these to fit odd i am also use to wearing the skinny jeans so trying a wide leg was a bit weird anyway the placement of the pockets
and the way it hit my waist was awkward and not flattering i am usually a and the fit but just did not look good'],
 ['the dress arrived with a few snags and the fabric already pilling the fabric will only get worse with wear return cute concept poor fabric q
u'],
 ['luv these jeans they felt so comfy the first time i put them on it s like butter i m a but went down a size and wear a belt they said they r
high rise but i can t see how but no complaints as i can breathe in them'],
 ['i originally got these in a petite and thought they fit great right out of the bag i m usually a and within an hour of wearing them to retai
ler actually the knees had gotten so baggy that they looked absolutely ridiculous i ve never had jeans do this the salesperson at retailer agre
ed that they seemed very odd i exchanged for a p at her suggestion and i just hate them i d say the smaller size seems every so slightly small
but nothing major but the waist actually folds over a'],
 ['i ll start with saying that it s a beautiful coat i get tons of compliments and love the color it runs a little big but it s meant to have t
```

In [178]:
```
neg_rev_desc = ''

for i in neg_review:
    for j in i:
        neg_rev_desc = neg_rev_desc + '' + j

print(neg_rev_desc)
```

```
size or smali think in general it is just hard to buy a bathing suit online but i really underestimated the v neck of this suit i read the revi
ews and noted it was not recommended for girls with larger size breast as i m basically an a cup i figured i would be fine the top portion basi
cally has no coverage or hold even when tied very tight the top still opened making me feel exposed and like my entire breast would fall out of
the suit needless to say my fianc was not a fan of the suit either evencan be a bit itchy sometimes i guess depending on your skin haven t had
a problem with the sweater up to nowi had tried this on in the beige loved it but could not find it in my size so i got it in the pink did not
look at good as the beigthere s are just what i was looking for these are perfect in texture and thickness not too thick but not too thin super
soft fleece on the inside i both both the black and green i will most likely but the grey soon enoughi am floored by the amount of positive rev
iews on this dress when i received it it looked nothing like it does on the model the bottom looked like dirty sand and was completely wrinkled
if you have anything above a c cup the top looks completely unflattering i looked so top heavy in this dress definitely not worth the price so
disappointi was so excited to order this sweater but was so disappointed when it came the knit is somewhat stiff and the sleeves are not as bel
l shaped as pictured on the model overall i would of kept it if it were much less money for the price it is just an ordinary sweater with no st
ylthis sweater was a big let down i am so i ordered it in a petite it was so short that if i lifted my arms you d see my bra there is no way i
could wear it without something under it even with high waisted bottoms and the bell sleeves look nothing like the pic the entire sleeve is jus
t wide and it continues with little increase into the cuff it is just short and boxy it is going back for suri was excited to see this top in p
erson but once i got a close up look at it and also tried it on i was disappointed first the color is much more persimmon orangey in person and
that s a hard color to pull off for a lot of people it just pulls any pink or red in your skin and highlights it second the top was poorly fini
shed i had to pick up two to try on because the lace was wonky across the front of the first one i grabbed so much so that the lace was puckere
d because it had beenthis sweater looks soft and flowy on the model but in the store it is rather stiff lace in the back and very plain in the
front i did not like the ribbing on the front bottom of the sweater because it pulled in places and didn t lie flat way too expensive for what
you are getthis shirt is not a good look for a gal with hips the fit at the top is tight i ordered a small but there is so much material at the
```

In [179]:
```
fig = plt.figure(figsize=(13,15))

descr_word_cloud = WordCloud(width=900,height=400,background_color="white",
            max_words=450,relative_scaling=1.0).generate(neg_rev_desc)
plt.imshow(descr_word_cloud)
plt.axis("off")
plt.title('WordCloud of Negative Reviews',fontsize=15)
plt.show()
```



In [ ]:
```

```

Applying Naive Bayes Classification Algorithm

In [181]:
```
reviews_df.head()
```

Out[181]:

	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 [180]:
```
X = reviews_df['Sentiment_Category']

y = reviews_df['Recommended IND']
```

In [183]:
```
1 y
```

Out[183]:
```
2        0
3        1
4        1
5        0
6        1
        ..
23481    1
23482    1
23483    0
23484    1
23485    1
Name: Recommended IND, Length: 19662, dtype: int64
```

In [184]:
```
1 from sklearn.model_selection import train_test_split
```

In [185]:
```
1 X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.25)
```

In [217]:
```
1 X_test
```

Out[217]:
```
17091    Positive
12664    Positive
6948     Positive
9289     Positive
15882    Positive
           ...
21178    Positive
20223    Positive
14465    Positive
16794    Positive
19440    Positive
Name: Sentiment_Category, Length: 4916, dtype: object
```

In [213]:
```
1 print(train_df_tfidf)
```

```
  (0, 2)        1.0
  (1, 2)        1.0
  (2, 2)        1.0
  (3, 2)        1.0
  (4, 2)        1.0
  (5, 2)        1.0
  (6, 2)        1.0
  (7, 0)        1.0
  (8, 0)        1.0
  (9, 2)        1.0
  (10, 2)       1.0
  (11, 2)       1.0
  (12, 2)       1.0
  (13, 2)       1.0
  (14, 2)       1.0
  (15, 2)       1.0
  (16, 2)       1.0
  (17, 2)       1.0
  (18, 2)       1.0
  (19, 2)       1.0
  (20, 2)       1.0
  (21, 2)       1.0
  (22, 2)       1.0
  (23, 2)       1.0
  (24, 2)       1.0
  :     :
  (14721, 2)    1.0
  (14722, 2)    1.0
  (14723, 2)    1.0
  (14724, 2)    1.0
  (14725, 2)    1.0
  (14726, 2)    1.0
  (14727, 2)    1.0
  (14728, 2)    1.0
  (14729, 2)    1.0
  (14730, 2)    1.0
  (14731, 2)    1.0
  (14732, 2)    1.0
  (14733, 2)    1.0
  (14734, 2)    1.0
  (14735, 2)    1.0
  (14736, 2)    1.0
  (14737, 2)    1.0
  (14738, 2)    1.0
  (14739, 2)    1.0
  (14740, 0)    1.0
  (14741, 2)    1.0
  (14742, 2)    1.0
  (14743, 2)    1.0
  (14744, 2)    1.0
  (14745, 2)    1.0
```

In [190]:
```
count_vect = CountVectorizer()
tfidf_tfm = TfidfTransformer()

train_df_counts = count_vect.fit_transform(X_train)

train_df_tfidf = tfidf_tfm.fit_transform(train_df_counts)
```

In [215]:
```
clf = MultinomialNB().fit(train_df_tfidf,y_train)
```

In [219]:
```
1 print(count_test_vect)
```

```
(0, 2)        1
(1, 2)        1
(2, 2)        1
(3, 2)        1
(4, 2)        1
(5, 2)        1
(6, 2)        1
(7, 2)        1
(8, 2)        1
(9, 2)        1
(10, 2)       1
(11, 2)       1
(12, 2)       1
(13, 2)       1
(14, 2)       1
(15, 2)       1
(16, 2)       1
(17, 2)       1
(18, 2)       1
(19, 2)       1
(20, 2)       1
(21, 2)       1
(22, 2)       1
(23, 2)       1
(24, 2)       1
:     :
(4891, 2)     1
(4892, 2)     1
(4893, 2)     1
(4894, 2)     1
(4895, 2)     1
(4896, 2)     1
(4897, 2)     1
(4898, 2)     1
(4899, 0)     1
(4900, 2)     1
(4901, 2)     1
(4902, 2)     1
(4903, 2)     1
(4904, 2)     1
(4905, 2)     1
(4906, 0)     1
(4907, 2)     1
(4908, 2)     1
(4909, 2)     1
(4910, 2)     1
(4911, 2)     1
(4912, 2)     1
(4913, 2)     1
(4914, 2)     1
(4915, 2)     1
```

In [197]:
```
count_test_vect = count_vect.transform(X_test)
y_predict = clf.predict(count_test_vect)
```

In [198]:
```
y_predict
```

Out[198]:
```
array([1, 1, 1, ..., 1, 1, 1])
```

In [199]:
```
from sklearn.metrics import confusion_matrix
```

In [200]:
```
cm = confusion_matrix(y_test,y_predict)
```

In [201]:
```
print(cm)
```

```
[[ 174  736]
 [ 104 3902]]
```

In [206]:
```
print("True Negative",cm[0][0])
print("True Positive",cm[1][1])
print("False Negative",cm[0][1])
print("Total",sum(sum(cm)))
print('Accuracy',(cm[0][0]+cm[1][1])/sum(sum(cm)))
```

```
True Negative 174
True Positive 3902
False Negative 736
Total 4916
Accuracy 0.8291293734743694
```

In [ ]:
```

```