

# Analysis of Amazon Musical Instruments Reviews

## I. Introduction

For my mini project 2, I will perform analysis on the json file download from Kaggle (<https://www.kaggle.com/eswarchandt/amazon-music-reviews>). And dealing with the following questions with python:

1. Compare the most common words used in each overall level of reviews in each year.
2. Group the reviews by year and month, summarizing the month trend and annual trend of each overall level reviews.
3. Separate the reviews into two groups: positive reviews and negative reviews, performing a sentiment analysis on them.

## II. Method

### i. Preparing Data

There are 10255 lines reviews in this dataset, with calling `readlines()` and `json.loads()` function, I import the json file into my notebook, and there are 9 keys in this json file.

```
# find out the keys in this json file
set(test)

{'asin',
 'helpful',
 'overall',
 'reviewText',
 'reviewTime',
 'reviewerID',
 'reviewerName',
 'summary',
 'unixReviewTime'}
```

Figure 1. all the keys in this json file

The next step I will store this json data into a dataframe, and below is what the dataframe looks like.

	asin	helpful	overall	reviewText	reviewTime	reviewerID	reviewerName	summary	unixReviewTime
0	1384719342	[0, 0]	5.0	Not much to write about here, but it does exac...	02 28, 2014	A2IBPI20UZIR0U	cassandra tu "Yeah, well, that's just like, u...	good	1393545600
1	1384719342	[13, 14]	5.0	The product does exactly as it should and is q...	03 16, 2013	A14VAT5EAX3D9S	Jake	Jake	1363392000
2	1384719342	[1, 1]	5.0	The primary job of this device is to block the...	08 28, 2013	A195EZSQDW3E21	Rick Bennette "Rick Bennette"	It Does The Job Well	1377648000
3	1384719342	[0, 0]	5.0	Nice windscreen protects my MXL mic and preven...	02 14, 2014	A2C00NNG1ZQQG2	RustyBill "Sunday Rocker"	GOOD WINDSCREEN FOR THE MONEY	1392336000
4	1384719342	[0, 0]	5.0	This pop filter is great. It looks and perform...	02 21, 2014	A94QU4C90B1AX	SEAN MASLANKA	No more pops when I record my vocals.	1392940800

Figure 2. overview of the dataframe in this step

And in the following step, I combine the 'summary' column and 'reviewText' column. Then I find the reviewTime column need to be converted into date format, and there are some unhelpful columns like 'helpful', 'asin', and 'unixReviewTime' need to be dropped. And after the cleaning the final version of dataframe is shown below.

	overall	reviewTime	text
0	5.0	2014-02-28	Not much to write about here, but it does exac...
1	5.0	2013-03-16	The product does exactly as it should and is q...
2	5.0	2013-08-28	The primary job of this device is to block the...
3	5.0	2014-02-14	Nice windscreen protects my MXL mic and preven...
4	5.0	2014-02-21	This pop filter is great. It looks and perform...

**Figure 3.** overview of the final version dataframe

## ii. Modeling

### Question 1:

To answer this question, I only the year of the date in 'reviewTime' column, so I use '.dt.year' to extract the year of each date, and I find the data covers reviews from 2004 to 2014.

```
df_q1 = df_new[['overall', 'reviewTime', 'text']]
df_q1['reviewTime'] = df_q1['reviewTime'].dt.year
set(df_q1['reviewTime'])

{2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014}
```

**Figure 4.** year extracting solution

To summarize the words, I need to do vectorization first, then display the 25 (maximum) most common used words of each overall level of reviews. And for the vectorization part, I set the n-gram equal to (1, 3), which means I will extract phrases that contains 1 words, 2 words, and 3 words, in this way I can cover most of the sentence components.

### Question 2:

As for this question, I only need to use columns 'overall', 'reviewTime'. Then use pivot\_table() function to slice the dataframe with month and year, drawing a line plot for each dataframe.

	overall	year
0	5.0	2014
1	5.0	2013
2	5.0	2013
3	5.0	2014
4	5.0	2014

**Figure 5.** dataframe of year

	overall	month
0	5.0	2
1	5.0	3
2	5.0	8
3	5.0	2
4	5.0	2

**Figure 6.** dataframe of month

### Question 3:

In this question, I used columns 'overall', 'text' for my analysis. The first step I take is to define the positive and negative label for each review. And the rule I set is the review will be considered as positive when their overall are higher than 3, otherwise they will be label as negative. Then I add a new column in the dataframe named 'POS' to store the label, which positive equals to 1, negative equals to 0.

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

**Figure 7.** dataframe of month

In the next step, I will do vectorization and visualization with this dataframe, to compare the negative reviews and positive reviews.

## III. Results & Conclusion

### Question 1:

The results of this question are shown below.

```
there are none or only few reviews in year: 2004
=====
there are none or only few reviews in year: 2005
=====
there are none or only few reviews in year: 2006
=====
there are none or only few reviews in year: 2007
=====
there are none or only few reviews in year: 2008
=====
there are none or only few reviews in year: 2009
=====
here are the top 4 words been used in year: 2010
[('one', 3), ('may', 2), ('good', 1), ('buy', 0)]
=====
here are the top 25 words been used in year: 2011
[('would', 31), ('work', 30), ('weight', 29), ('way', 28), ('use', 27), ('top', 26), ('stand', 25), ('really', 24), ('purchase', 23), ('one', 22), ('never', 21), ('much', 20), ('long', 19), ('like', 18), ('hold', 17), ('guitars', 16), ('guitar', 15), ('great', 14), ('good', 13), ('going', 12), ('go', 11), ('get', 10), ('even', 9), ('enough', 8), ('could', 7)]
=====
here are the top 25 words been used in year: 2012
[('zero', 282), ('yet', 281), ('year', 280), ('wrong', 279), ('would', 278), ('worst', 277), ('works', 276), ('worked', 275), ('work', 274), ('without', 273), ('whole', 272), ('went', 271), ('well', 270), ('weeks', 269), ('way', 268), ('watt', 267), ('waste', 266), ('want', 265), ('volume', 264), ('using', 263), ('useless', 262), ('used', 261), ('use', 260), ('usb', 259), ('us', 258)]
=====
here are the top 25 words been used in year: 2013
[('years', 272), ('would', 271), ('worth', 270), ('worst', 269), ('works', 268), ('working', 267), ('work', 266), ('without', 265), ('went', 264), ('well', 263), ('weeks', 262), ('week', 261), ('wear', 260), ('way', 259), ('want', 258), ('using', 257), ('useless', 256), ('used', 255), ('use', 254), ('update', 253), ('unless', 252), ('unit', 251), ('uncomfortable', 250), ('two', 249), ('tuned', 248)]
=====
here are the top 25 words been used in year: 2014
[('years', 223), ('would', 222), ('worth', 221), ('works', 220), ('working', 219), ('work', 218), ('without', 217), ('whether', 216), ('went', 215), ('well', 214), ('way', 213), ('wanted like', 212), ('wanted', 211), ('want', 210), ('using', 209), ('useless', 208), ('used', 207), ('use', 206), ('update', 205), ('unless', 204), ('unit', 203), ('two', 202), ('trying', 201), ('try', 200), ('tried', 199)]
```

**Figure 7.1.** most common words in 1 star reviews

Junfan Chen  
IST652  
MiniProject2

```
there are none or only few reviews in year: 2004
=====
there are none or only few reviews in year: 2005
=====
there are none or only few reviews in year: 2006
=====
there are none or only few reviews in year: 2007
=====
there are none or only few reviews in year: 2008
=====
there are none or only few reviews in year: 2009
=====
here are the top 11 words been used in year: 2010
[('use', 10), ('sound', 9), ('purchased', 8), ('product', 7), ('price', 6), ('one', 5), ('need', 4), ('mic', 3), ('get', 2),
('first', 1), ('buy', 0)]
=====
here are the top 25 words been used in year: 2011
[('would better', 150), ('would', 149), ('worth', 148), ('works', 147), ('work', 146), ('without', 145), ('wise', 144), ('well', 143),
('way', 142), ('want', 141), ('using', 140), ('used', 139), ('use', 138), ('two', 137), ('turn', 136), ('try', 135),
('tried', 134), ('tone', 133), ('tiny', 132), ('time', 131), ('thought', 130), ('think', 129), ('things', 128), ('thing', 127),
('thin', 126)]
=====
here are the top 25 words been used in year: 2012
[('years', 147), ('would', 146), ('worth', 145), ('works', 144), ('worked', 143), ('work well', 142), ('work', 141), ('without', 140),
('well', 139), ('way', 138), ('want', 137), ('used', 136), ('use', 135), ('unit', 134), ('two', 133), ('turn', 132),
('tuner', 131), ('try', 130), ('tone', 129), ('time', 128), ('though', 127), ('think', 126), ('thing', 125), ('terrible', 124),
('take', 123)]
=====
here are the top 25 words been used in year: 2013
[('yeti', 457), ('yet', 456), ('years', 455), ('yamaha', 454), ('xlr', 453), ('wrong', 452), ('wound', 451), ('would recommend', 450),
('would probably', 449), ('would make', 448), ('would give', 447), ('would', 446), ('worth money', 445), ('worth', 444), ('works', 443),
('working', 442), ('worked', 441), ('work', 440), ('without', 439), ('within', 438), ('wish', 437), ('wire', 436), ('whatever', 435),
('well made', 434), ('well', 433)]
=====
here are the top 25 words been used in year: 2014
[('yet', 311), ('years', 310), ('year', 309), ('would', 308), ('worth', 307), ('works fine', 306), ('works', 305), ('working', 304),
('work well', 303), ('work', 302), ('without', 301), ('well', 300), ('weeks', 299), ('week', 298), ('ways', 297), ('way', 296),
('waves capo', 295), ('waves', 294), ('want', 293), ('volume', 292), ('version', 291), ('using', 290), ('user', 289), ('used', 288),
('use', 287)]
```

**Figure 7.2.** most common words in 2 stars reviews

Junfan Chen  
IST652  
MiniProject2

```
there are none or only few reviews in year: 2004
=====
there are none or only few reviews in year: 2005
=====
there are none or only few reviews in year: 2006
=====
there are none or only few reviews in year: 2007
=====
here are the top 8 words been used in year: 2008
[('well', 7), ('sound', 6), ('really', 5), ('pedal', 4), ('one', 3), ('need', 2), ('like', 1), ('good', 0)]
=====
here are the top 23 words been used in year: 2009
[('would', 22), ('use', 21), ('turn', 20), ('though', 19), ('thing', 18), ('stick', 17), ('sound', 16), ('small', 15), ('place', 14), ('one', 13), ('old', 12), ('nice', 11), ('need', 10), ('make', 9), ('great', 8), ('good', 7), ('get', 6), ('far', 5), ('different', 4), ('cool', 3), ('better', 2), ('back', 1), ('also', 0)]
=====
here are the top 25 words been used in year: 2010
[('years', 62), ('wrong', 69), ('would', 68), ('works', 67), ('well', 66), ('way', 65), ('using', 64), ('used', 63), ('use', 62), ('two', 61), ('try', 60), ('tried', 59), ('time', 58), ('though', 57), ('sure', 56), ('strings', 55), ('string', 54), ('sounds', 53), ('sound', 52), ('seems', 51), ('said', 50), ('reviews', 49), ('recommend', 48), ('really', 47), ('product', 46)]
=====
here are the top 25 words been used in year: 2011
[('yet', 599), ('years', 598), ('year', 597), ('xlr', 596), ('would recommend', 595), ('would', 594), ('worth', 593), ('works well', 592), ('works pretty', 591), ('works fine', 590), ('works', 589), ('worked', 588), ('work well', 587), ('work', 586), ('wood', 585), ('without', 584), ('within', 583), ('wish', 582), ('whatever', 581), ('went', 580), ('well', 579), ('weeks', 578), ('way', 577), ('waves', 576), ('watt', 575)]
=====
here are the top 25 words been used in year: 2012
[('zero', 738), ('yet', 737), ('yes', 736), ('years ago', 735), ('years', 734), ('year', 733), ('xlr', 732), ('wrong', 731), ('would recommend', 730), ('would buy', 729), ('would', 728), ('worth money', 727), ('worth', 726), ('worst', 725), ('worse', 724), ('worried', 723), ('works well', 722), ('works great', 721), ('works', 720), ('working', 719), ('worked well', 718), ('worked', 717), ('work well', 716), ('work', 715), ('wonder', 714)]
=====
here are the top 25 words been used in year: 2013
[('yet', 999), ('yes', 998), ('years ago', 997), ('years', 996), ('year', 995), ('wrong', 994), ('writing', 993), ('wound', 992), ('would work', 991), ('would use', 990), ('would say', 989), ('would recommend', 988), ('would never', 987), ('would like', 986), ('would great', 985), ('would give', 984), ('would buy', 983), ('would', 982), ('worth money', 981), ('worth', 980), ('worse', 979), ('worry', 978), ('world', 977), ('works well', 976), ('works okay', 975)]
=====
here are the top 25 words been used in year: 2014
[('yet', 690), ('yes', 689), ('years', 688), ('year', 687), ('wrong', 686), ('write', 685), ('would work', 684), ('would recommend', 683), ('would', 682), ('worth', 681), ('worn', 680), ('works well', 679), ('works really', 678), ('works', 677), ('working', 676), ('worked', 675), ('work well', 674), ('work fine', 673), ('work', 672), ('wood', 671), ('without', 670), ('wish', 669), ('winder', 668), ('whole', 667), ('went', 666)]
```

Figure 7.3. most common words in 3 stars reviews

Junfan Chen  
IST652  
MiniProject2

```
there are none or only few reviews in year: 2004
=====
there are none or only few reviews in year: 2005
=====
there are none or only few reviews in year: 2006
=====
here are the top 10 words been used in year: 2007
[('would', 9), ('use', 8), ('one', 7), ('like', 6), ('great', 5), ('good', 4), ('go', 3), ('get', 2), ('even', 1), ('controls',
0)]
=====
here are the top 25 words been used in year: 2008
[('would', 126), ('works well', 125), ('works', 124), ('without', 123), ('well', 122), ('wanted', 121), ('want', 120), ('vocal
s', 119), ('using', 118), ('uses', 117), ('useful', 116), ('used', 115), ('use', 114), ('type', 113), ('two', 112), ('try', 11
1), ('top', 110), ('time', 109), ('things', 108), ('thing', 107), ('tell', 106), ('switch', 105), ('studio', 104), ('string', 1
03), ('still', 102)]
=====
here are the top 25 words been used in year: 2009
[('yet', 205), ('would', 204), ('works', 203), ('work', 202), ('without', 201), ('wide', 200), ('well made', 199), ('well', 19
8), ('way', 197), ('want', 196), ('using', 195), ('used', 194), ('use sound', 193), ('use', 192), ('usb', 191), ('update', 19
0), ('two', 189), ('try', 188), ('tried', 187), ('tone', 186), ('time', 185), ('thus', 184), ('though', 183), ('think', 182),
('thick', 181)]
=====
here are the top 25 words been used in year: 2010
[('yet', 487), ('years ago', 486), ('years', 485), ('year', 484), ('wrong', 483), ('would recommend', 482), ('would', 481), ('w
orth', 480), ('works great', 479), ('works', 478), ('working', 477), ('work', 476), ('without', 475), ('within', 474), ('wish',
473), ('whatever', 472), ('went', 471), ('well made', 470), ('well', 469), ('weight', 468), ('wear', 467), ('way', 466), ('want
ed', 465), ('want', 464), ('wah', 463)]
=====
here are the top 25 words been used in year: 2011
[('yet', 999), ('yes', 998), ('years ago', 997), ('years', 996), ('year', 995), ('yamaha', 994), ('wrong', 993), ('wow', 992),
('wound', 991), ('would work', 990), ('would suggest', 989), ('would say', 988), ('would recommend', 987), ('would', 986), ('wo
rth', 985), ('worse', 984), ('worry', 983), ('world', 982), ('works well', 981), ('works great', 980), ('works advertised', 97
9), ('works', 978), ('working', 977), ('worked', 976), ('work well', 975)]
=====
here are the top 25 words been used in year: 2012
[('youtube', 999), ('yet', 998), ('yes', 997), ('years ago', 996), ('years', 995), ('year', 994), ('xlr', 993), ('wrong', 992),
('would recommend', 991), ('would nice', 990), ('would', 989), ('worth', 988), ('works well', 987), ('works great', 986), ('wor
ks fine', 985), ('works', 984), ('working', 983), ('worked fine', 982), ('worked', 981), ('work well', 980), ('work', 979), ('w
ood', 978), ('without', 977), ('within', 976), ('wish', 975)]
=====
here are the top 25 words been used in year: 2013
[('yet', 999), ('yes', 998), ('years ago', 997), ('years', 996), ('year', 995), ('yamaha', 994), ('xlr', 993), ('wrong', 992),
('writing', 991), ('would recommend', 990), ('would buy', 989), ('would', 988), ('worth', 987), ('worry', 986), ('works well',
985), ('works great', 984), ('works fine', 983), ('works', 982), ('working', 981), ('worked', 980), ('work well', 979), ('work
fine', 978), ('work', 977), ('wood', 976), ('wont', 975)]
=====
here are the top 25 words been used in year: 2014
[('zipper', 999), ('youtube', 998), ('yet', 997), ('yes', 996), ('years', 995), ('year', 994), ('xlr', 993), ('wrong', 992),
('would recommend', 991), ('would nice', 990), ('would expect', 989), ('would buy', 988), ('would', 987), ('worth', 986), ('wor
ry', 985), ('works well', 984), ('works like', 983), ('works great', 982), ('works good', 981), ('works fine', 980), ('works ad
vertised', 979), ('works', 978), ('working', 977), ('worked', 976), ('work well', 975)]
```

Figure 7.4. most common words in 4 stars reviews



```

here are the top 16 words been used in year: 2004
[('using', 15), ('use', 14), ('tension', 13), ('sound', 12), ('set', 11), ('right', 10), ('quickly', 9), ('position', 8), ('playing', 7), ('perfect', 6), ('model', 5), ('lever', 4), ('get', 3), ('either', 2), ('drum', 1), ('bass', 0)]
=====
there are none or only few reviews in year: 2005
=====
here are the top 20 words been used in year: 2006
[('without', 19), ('well', 18), ('way', 17), ('used', 16), ('use', 15), ('tried', 14), ('still', 13), ('sound', 12), ('set', 11), ('one', 10), ('need', 9), ('low', 8), ('like', 7), ('instrument', 6), ('however', 5), ('guitar', 4), ('good', 3), ('get', 2), ('built', 1), ('available', 0)]
=====
here are the top 25 words been used in year: 2007
[('zoom', 106), ('years ago', 105), ('years', 104), ('would', 103), ('works great', 102), ('works', 101), ('without', 100), ('well', 99), ('want', 98), ('using', 97), ('used', 96), ('use', 95), ('unlike', 94), ('two', 93), ('tune', 92), ('time', 91), ('three', 90), ('thought', 89), ('though', 88), ('thing', 87), ('taylor', 86), ('strings', 85), ('string', 84), ('standard', 83), ('stand', 82)]
=====
here are the top 25 words been used in year: 2008
[('years ago', 170), ('years', 169), ('would', 168), ('worth', 167), ('works', 166), ('work', 165), ('without', 164), ('wish', 163), ('went', 162), ('well constructed', 161), ('well', 160), ('way', 159), ('want', 158), ('version', 157), ('using', 156), ('used', 155), ('use', 154), ('tuner', 153), ('tune', 152), ('try', 151), ('tried', 150), ('took', 149), ('tone', 148), ('time', 147), ('time', 146)]
=====
here are the top 25 words been used in year: 2009
[('yet', 397), ('years', 396), ('xlr', 395), ('would', 394), ('worth', 393), ('worry', 392), ('world', 391), ('works perfectly', 390), ('works great', 389), ('works', 388), ('working', 387), ('worked', 386), ('work well', 385), ('work', 384), ('wood', 383), ('without', 382), ('whatever', 381), ('well made', 380), ('well', 379), ('way', 378), ('wanted', 377), ('want', 376), ('volume', 375), ('vocals', 374), ('value', 373)]
=====
here are the top 25 words been used in year: 2010
[('yet', 999), ('yes', 998), ('years ago', 997), ('years', 996), ('year', 995), ('yamaha', 994), ('xlr', 993), ('wrong', 992), ('write', 991), ('wow', 990), ('would recommend', 989), ('would like', 988), ('would also', 987), ('would', 986), ('worth price', 985), ('worth money', 984), ('worth', 983), ('worry', 982), ('world', 981), ('works well', 980), ('works perfectly', 979), ('works great', 978), ('works fine', 977), ('works exactly', 976), ('works advertised', 975)]
=====
here are the top 25 words been used in year: 2011
[('youtube', 999), ('yet', 998), ('yes', 997), ('years', 996), ('year', 995), ('xlr', 994), ('wrong', 993), ('wow', 992), ('wound', 991), ('would recommend', 990), ('would', 989), ('worth', 988), ('worry', 987), ('works well', 986), ('works great', 985), ('works fine', 984), ('works', 983), ('worked', 982), ('work well', 981), ('work', 980), ('wood', 979), ('wonderful', 978), ('without', 977), ('wish', 976), ('wire', 975)]
=====
here are the top 25 words been used in year: 2012
[('youtube', 999), ('yet', 998), ('yes', 997), ('years ago', 996), ('years', 995), ('year', 994), ('xlr', 993), ('wrong', 992), ('wow', 991), ('would recommend', 990), ('would', 989), ('worth', 988), ('worry', 987), ('works well', 986), ('works great', 985), ('works fine', 984), ('works', 983), ('working', 982), ('worked', 981), ('work well', 980), ('work great', 979), ('work', 978), ('word', 977), ('wood', 976), ('wonderful', 975)]
=====
here are the top 25 words been used in year: 2013
[('zero', 999), ('youtube', 998), ('yet', 997), ('yes', 996), ('years ago', 995), ('years', 994), ('year', 993), ('yamaha', 992), ('xlr', 991), ('wrong', 990), ('wow', 989), ('would recommend', 988), ('would', 987), ('worth', 986), ('worry', 985), ('works well', 984), ('works like', 983), ('works great', 982), ('works fine', 981), ('works', 980), ('working', 979), ('worked', 978), ('work well', 977), ('work great', 976), ('work', 975)]
=====
here are the top 25 words been used in year: 2014
[('youtube', 999), ('yet', 998), ('yes', 997), ('years', 996), ('year', 995), ('yamaha', 994), ('xlr', 993), ('wrong', 992), ('wow', 991), ('wound', 990), ('would recommend', 989), ('would', 988), ('worth', 987), ('worry', 986), ('world', 985), ('works well', 984), ('works perfectly', 983), ('works great', 982), ('works fine', 981), ('works', 980), ('working', 979), ('worked', 978), ('work well', 977), ('work great', 976), ('work', 975)]

```

**Figure 7.5.** most common words in 5 stars reviews

As the results shown, these items are not very popular until 2010, since there is none or very few reviews before 2010. And people prefer giving a high overall (4, 5) review than giving a low one, which can tell us that these products may have a good quality, also the words used in 4 stars and 5 stars reviews like ‘worth money’, ‘worth price’, ‘works perfectly’, ‘wonderful’ etc. can proof that. But in the common words used in overall 2 and 1, I do not find the feature words that can tell the negative feeling about product quality, maybe these customers met some problems other than the quality of Instruments.

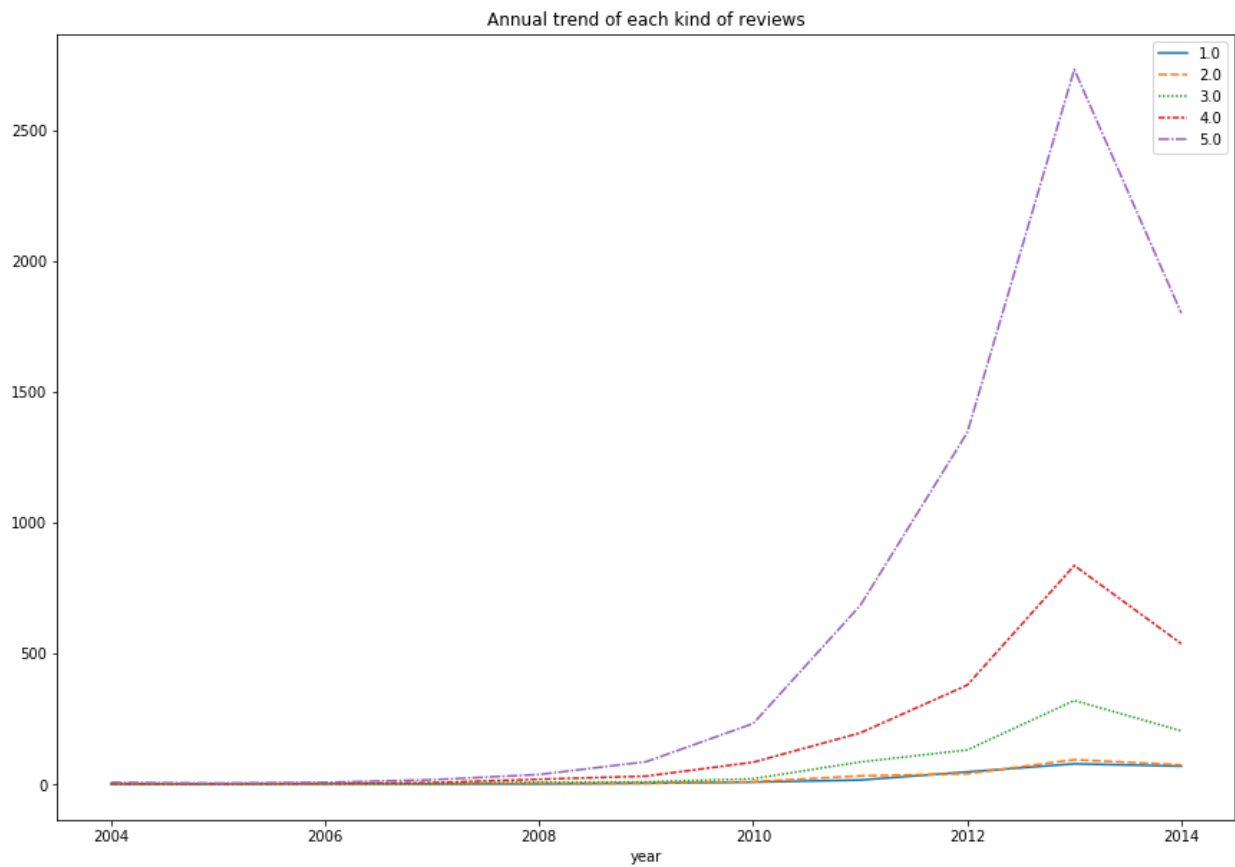
## Question 2:



Here are the results of this question.

overall	1.0	2.0	3.0	4.0	5.0
year					
2004	0.0	0.0	0.0	1.0	6.0
2005	0.0	0.0	0.0	1.0	3.0
2006	1.0	0.0	1.0	2.0	6.0
2007	0.0	1.0	0.0	5.0	16.0
2008	0.0	2.0	7.0	18.0	36.0
2009	3.0	2.0	8.0	30.0	85.0
2010	7.0	9.0	20.0	83.0	231.0
2011	15.0	31.0	84.0	195.0	682.0
2012	46.0	39.0	130.0	378.0	1343.0
2013	77.0	93.0	319.0	835.0	2731.0
2014	68.0	73.0	203.0	536.0	1799.0

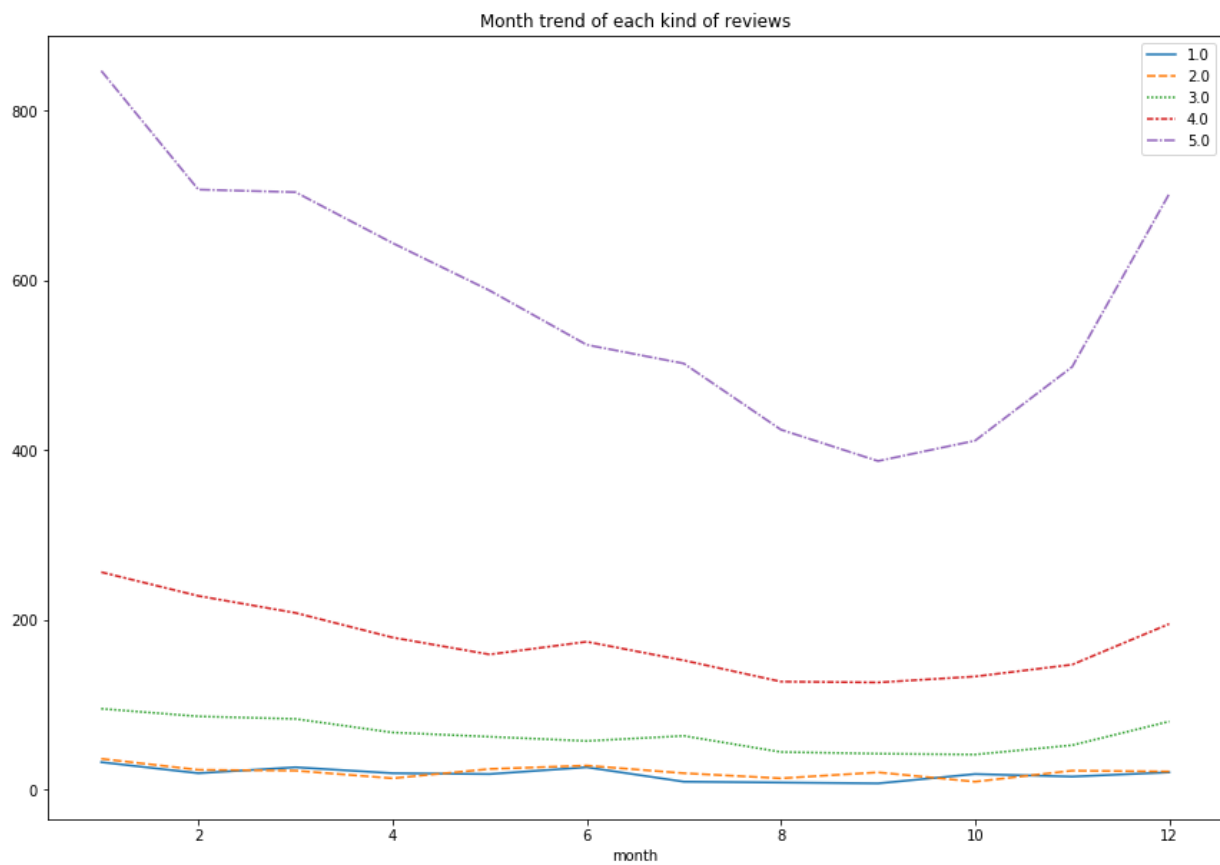
**Figure 8.1.** dataframe of the summary of reviews in each year



**Figure 8.2.** annual trend of each kind of reviews

overall	1.0	2.0	3.0	4.0	5.0
month					
1	32	36	95	256	847
2	19	23	86	228	707
3	26	22	83	208	704
4	19	13	67	179	644
5	18	24	62	159	588
6	26	28	57	174	524
7	9	19	63	152	502
8	8	13	44	127	424
9	7	20	42	126	387
10	18	9	41	133	411
11	15	22	52	147	498
12	20	21	80	195	702

**Figure 9.1.** dataframe of the summary of reviews in each month

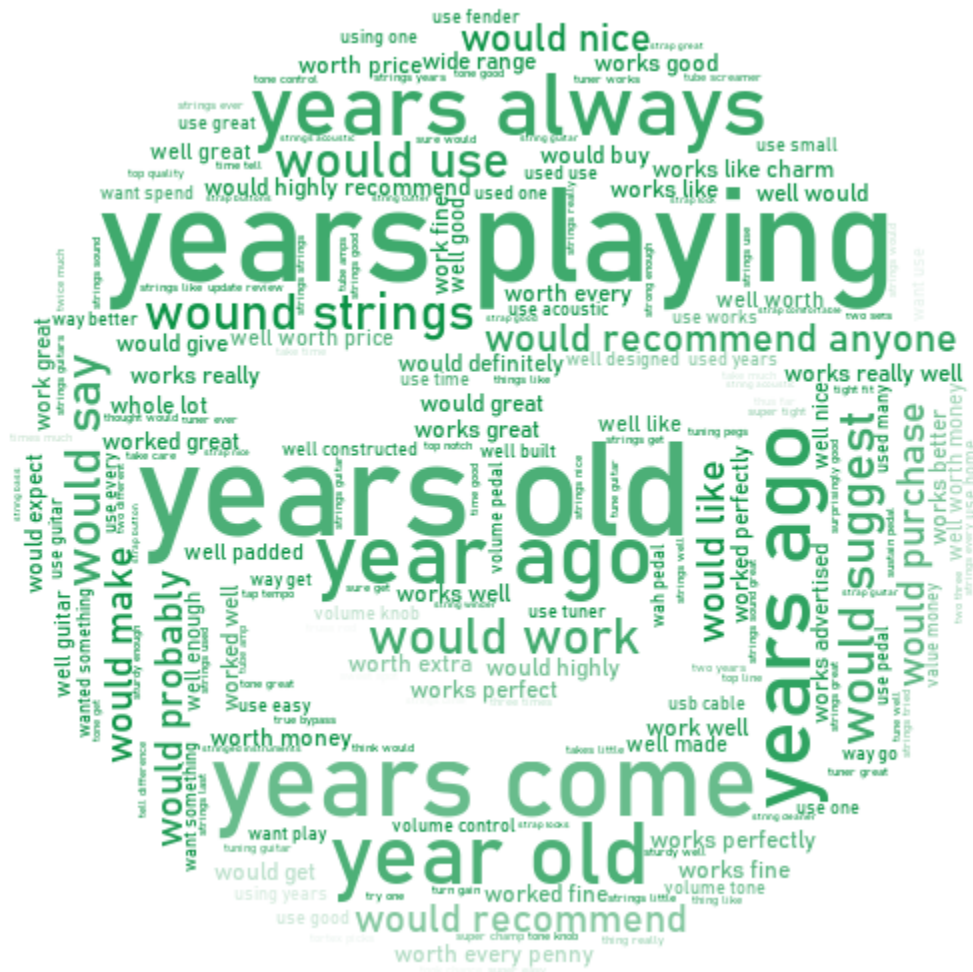


**Figure 9.2.** month trend of each kind of reviews

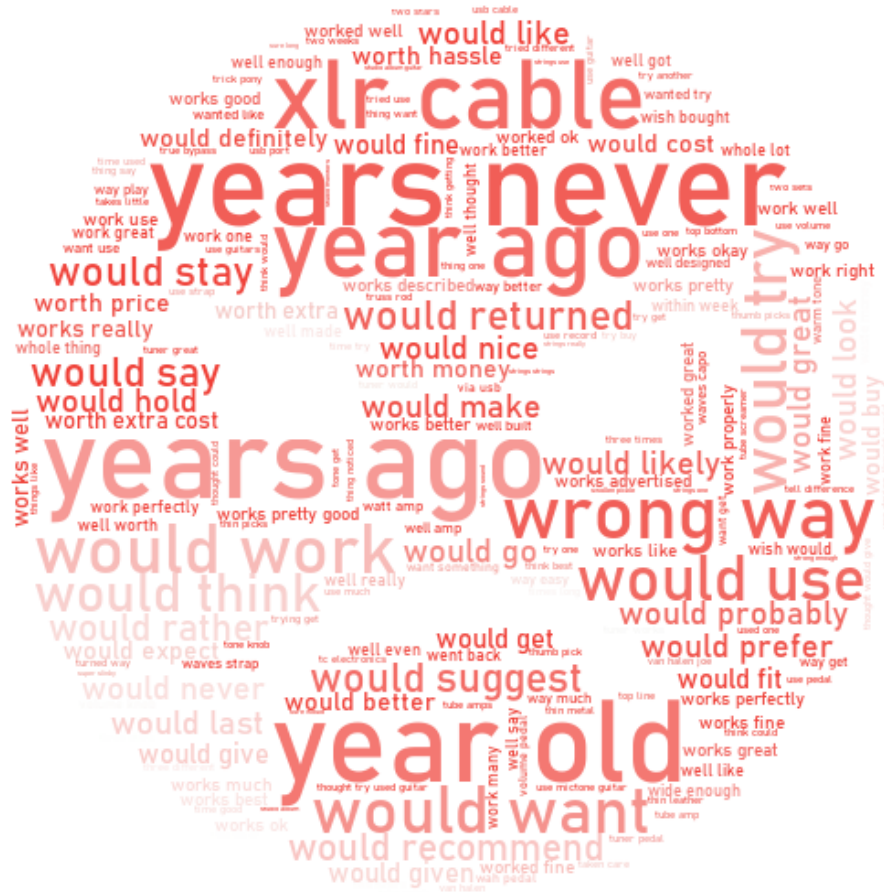
As the plots and the dataframes shown, the sales volume reaches the peak in 2013, the reason may be the increase in popularity and production capacity, and people are more likely to purchase for these instruments in December and January, maybe there are some deals during this period of time.

### Question 3:

In order to filter the unhelpful words, I do the bigram and trigram vectorization in this step and here are the word cloud of positive features and negative features.



**Figure 10.1.** positive features word cloud



**Figure 10.2.** negative features word cloud

From the word cloud we can find there are some common words, which are used in both negative reviews and positive reviews, such as ‘year ago’, ‘year old’, ‘would suggest’ etc., but still cannot find the features that can express the negative emotions that people understand, maybe I need to do a further processing on text data, make the computer think like a human.