



# BKR

BE KIND REWIND

## YouTube Sentiment Analysis



John Ian Castaneda

## Table of Contents

Description .....	2
Objectives.....	2
Steps.....	2
The Data .....	2
Tools.....	2
Data Cleaning and Pre-processing .....	3
Exploratory Data Analysis .....	3
1. Top performing videos by viewCount .....	4
2. Worst performing video by viewCount.....	4
3. Average view count of videos .....	5
4. Distribution of viewCount (visualization through Python) .....	5
5. Videos with the most comments .....	6
6. Videos with the least comments.....	6
7. Average number of comments .....	7
8. Video comments per 1000 views.....	7
9. Distribution of commentCount (visualization through Python) .....	8
10. Videos with the most likes .....	8
11. Days of the week that videos are posted.....	9
12. Day of week videos were published and their current view counts.....	9
13. Average duration of videos in minutes .....	10
14. Day of the week when comments are posted .....	10
15. Average number of words in comments for all videos.....	10
16. Top 10 videos with the highest average number of words in comments .....	11
17. Videos with the top 10 comments with the most number of words .....	11
18. Top 10 viewers with most comments on videos .....	12
19. Correlation of viewCount, likeCount, commentCount, duration, tagCount, and comment_view_ratio .....	13
20. Scatterplot of viewCount against commentCount, likeCount, tagCount, and durationSecs.....	13
21. Word Cloud .....	14
Sentiment Analysis.....	15
Overall Sentiment .....	15
Top 10 videos with the highest percentage of positive comments.....	16
Top 10 videos with the highest percentage of neutral comments.....	16
Top 10 videos with the highest percentage of negative comments.....	17
Top 10 users with most positive comments .....	17
Top 10 users with most neutral comments .....	17
Top 10 users with most negative comments .....	17
Word Cloud of Positive Video Comments.....	18
Word Cloud of Neutral Video Comments .....	19
Word Cloud of Negative Video Comments .....	20
Conclusion.....	21

# YouTube Sentiment Analysis - Be Kind Rewind

## Description

I've always been fascinated by women in films and the real women who play them. Every year I try to see who are nominated for the top acting prize for women and I make sure to watch the performances that garnered them that nomination. I would also often search for videos of past winners and watch their acceptance speeches and to also see which exceptional actresses they were able to beat.

So, I was very glad to discover this YouTube channel called **\*\*Be Kind Rewind\*\***. The channel is mainly about past best actress winners, women in film, and film and Hollywood history. The videos delve into the career of these accomplished women and the circumstances that lead to their win. The videos are well-researched and very insightful. I love history, I love film, and I love women in film so I truly enjoy every video from this channel. And so I wanted to see what other people thought about this channel and its videos.

In this project, I apply natural language processing (NLP) to determine the general sentiments on this channel. The results of this project will provide the creator behind this YouTube channel a good idea of how viewers perceive the content produced.

## Objectives

The goals of this project are:

- Use the YouTube API to collect the data for this project
- Analyze the performance of the videos of the channel
- Determine the polarity of sentiments in the video comments using NLP
- Get the most common words used in comments
- Determine the overall sentiment for the channel

## Steps

1. Download channel, video, and comments data from YouTube using the YouTube API
2. Clean and preprocess the data
3. Perform exploratory data analysis (EDA) on the data
4. Analyze the sentiment of all the comments on the videos and the channel as a whole
5. Conclusion

## The Data

The data was scraped using Google Youtube Data API version 3.0. There are two datasets: one for the details for each of the 67 videos on the channel and one for the comments posted for all of the videos on the channel. The first dataset was used to analyze the overall performance of the channel. The second dataset was used to get the sentiments from viewers.

## Tools

- **Python** - used to extract the data, clean and preprocess the data, get the sentiments from the comments, and analyze the data. The libraries used were:
  - NumPy
  - Pandas
  - Google API client
  - Matplotlib
  - Seaborn
  - WordCloud
  - Demoji
  - NLTK
  - TextBlob
- **SQL Server** – store the data and retrieve information for analysis
- **YouTube API** – source of the data

## Data Cleaning and Pre-processing

After importing the libraries and extracting the data of the channel from YouTube, we prepared the data for analysis by doing the following:

- Check the data types of each column and number of non-null values
- Convert numerical columns to numeric
- Remove the column 'favouriteCount' since this column has no data
- Convert 'duration' column to seconds
- Add a column that indicates the day of the week the videos and comments were published based on the column 'publishedAt'
- Add column 'tagCount'
- Remove emojis
- Insert a column to indicate the language the comment was written in and remove comments in languages other than English
- Remove stop words and special characters, lemmatize words
- Get the number of words in each comment

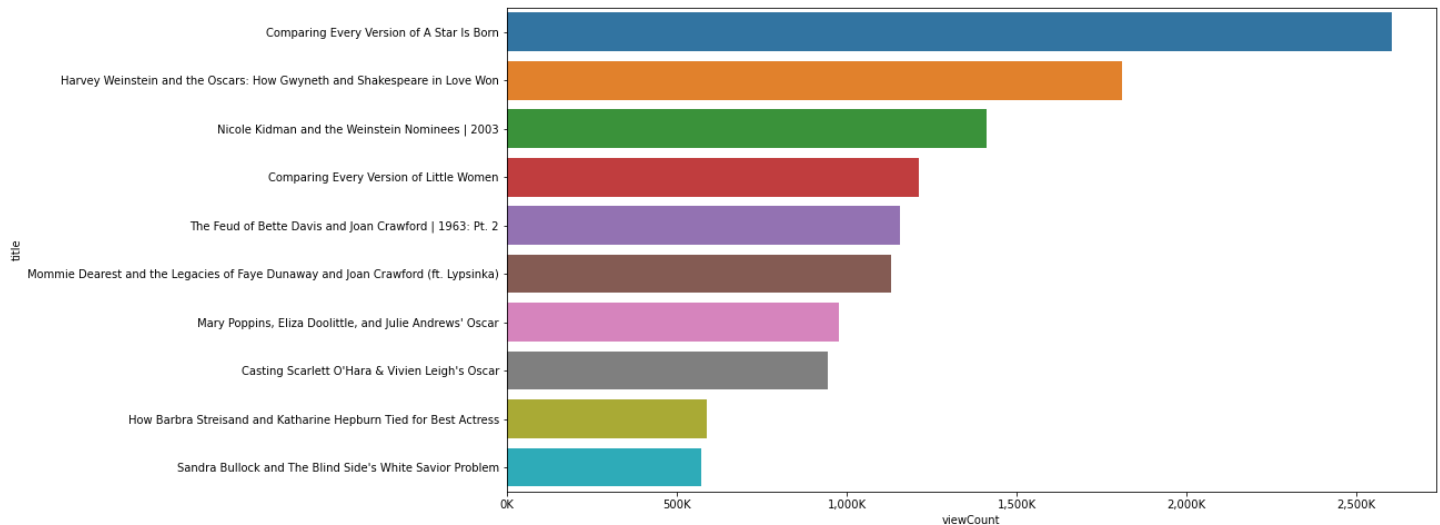
## Exploratory Data Analysis

The exploratory data analysis was performed using SQL to retrieve information for the datasets. Python was also used for the analysis to visualize the analysis. The EDA aimed to retrieve the following information:

1. [Top performing videos by viewCount](#)
2. [Worst performing video by viewCount](#)
3. [Average view count of videos](#)
4. [Distribution of viewCount \(visualization through Python\)](#)
5. [Videos with the most comments](#)
6. [Videos with the least comments](#)
7. [Average number of comments](#)
8. [Video comments per 1000 views](#)
9. [Distribution of commentCount \(visualization through Python\)](#)
10. [Videos with the most likes](#)
11. [Days of the week that videos are posted](#)
12. [Day of week videos were published and their current view counts](#)
13. [Average duration of videos in minutes](#)
14. [Day of the week when comments are posted](#)
15. [Average number of words in comments for all videos](#)
16. [Top 10 videos with the highest average number of words in comments](#)
17. [Videos with the top 10 comments with the most number of words](#)
18. [Top 10 viewers with most comments on videos](#)
19. [Correlation of viewCount, likeCount, commentCount, duration, tagCount, and comment\\_per\\_1k\\_view](#)
20. [Scatterplot of viewCount against commentCount, likeCount, tagCount, and durationSecs](#)
21. [Word Cloud](#)

## 1. Top performing videos by viewCount

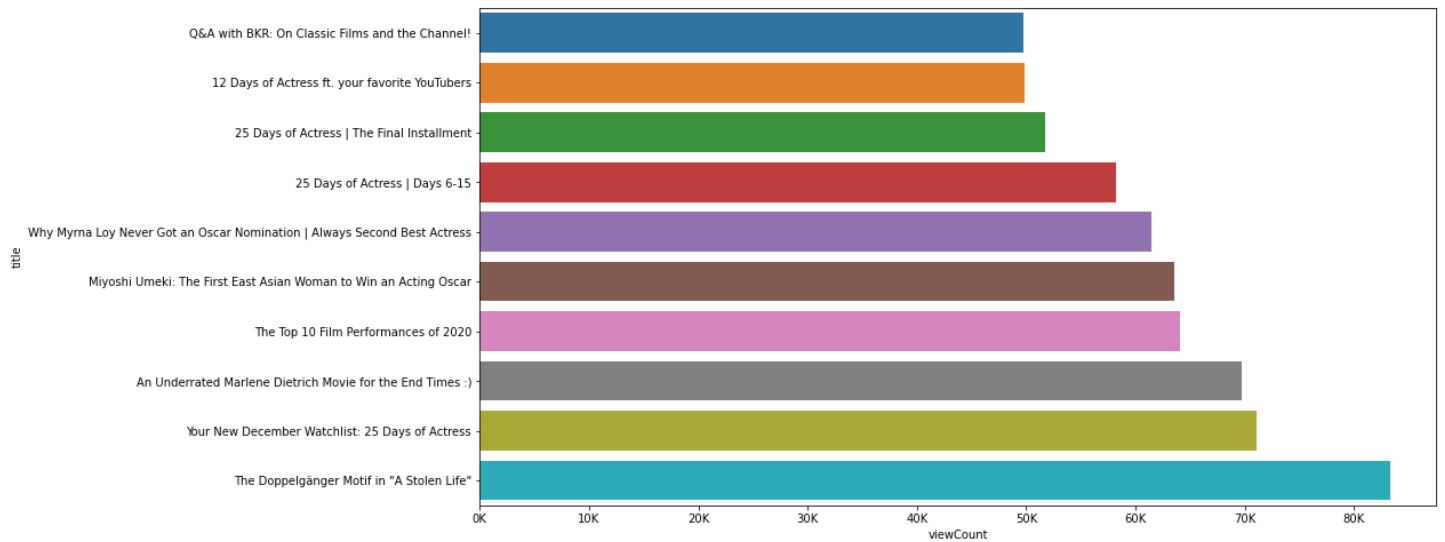
	title	viewcount
1	Comparing Every Version of A Star Is Born	2604287
2	Harvey Weinstein and the Oscars: How Gwyneth and ...	1810343
3	Nicole Kidman and the Weinstein Nominees   2003	1412620
4	Comparing Every Version of Little Women	1211742
5	The Feud of Bette Davis and Joan Crawford   1963: Pt. 2	1158232
6	Mommie Dearest and the Legacies of Faye Dunaway ...	1131213
7	Mary Poppins, Eliza Doolittle, and Julie Andrews' Oscar	976508
8	Casting Scarlett O'Hara & Vivien Leigh's Oscar	943599
9	How Barbra Streisand and Katharine Hepburn Tied for...	588030
10	Sandra Bullock and The Blind Side's White Savior Pro...	570741



The most viewed videos are those that discuss movies and personalities who are familiar to today's audiences. The channel goes over the story of past best actress winners all the way from the infancy of the Oscars. Average viewers may not be able to easily recognize and relate to the subject of some of the videos especially those from the Golden Age of Hollywood. The top 4 most viewed videos cover movies that were recently made (or remade) or actresses that are currently active who won their Oscar within the last 2 decades. The top 5 and 6 most watched video also has this characteristic in a way. They delve into the story of 2 actresses from the Golden Age of Hollywood who were the subjects of a recent popular TV series.

## 2. Worst performing video by viewCount

	title	viewCount
1	Q&A with BKR: On Classic Films and the Channel!	49722
2	12 Days of Actress ft. your favorite YouTubers	49836
3	25 Days of Actress   The Final Installment	51684
4	25 Days of Actress   Days 6-15	58194
5	Why Myrna Loy Never Got an Oscar Nomination   Al...	61398
6	Miyoshi Umeki: The First East Asian Woman to Win a...	63564
7	The Top 10 Film Performances of 2020	64040
8	An Underrated Marlene Dietrich Movie for the End Ti...	69661
9	Your New December Watchlist: 25 Days of Actress	71043
10	The Doppelgänger Motif in "A Stolen Life"	83333



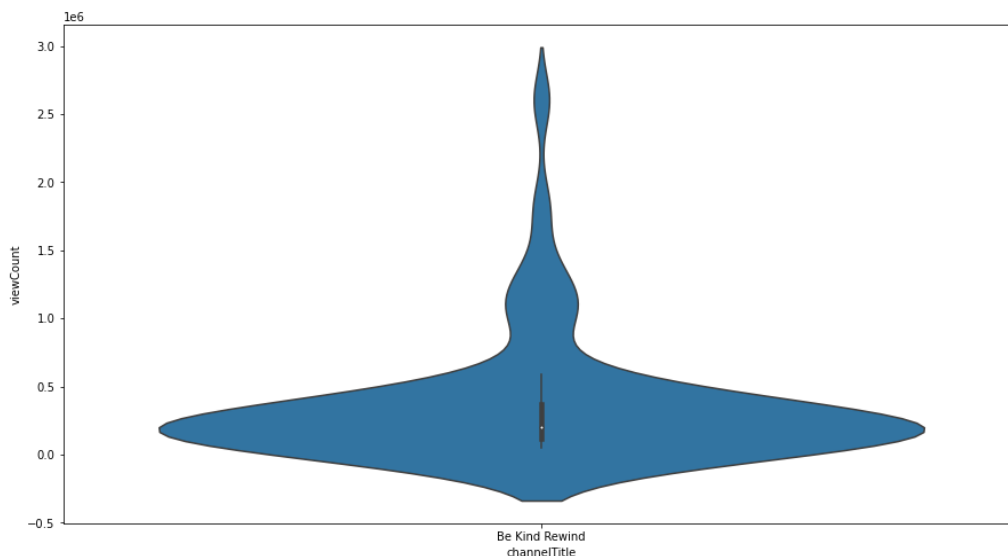
The top 4 least viewed videos are those that do not focus on a single specific actress or movie. The other least viewed videos are those the cover lesser know personalities to today's average movie-goer.

### 3. Average view count of videos

	average_views
1	352354

The average views of channel per video is 352,354.

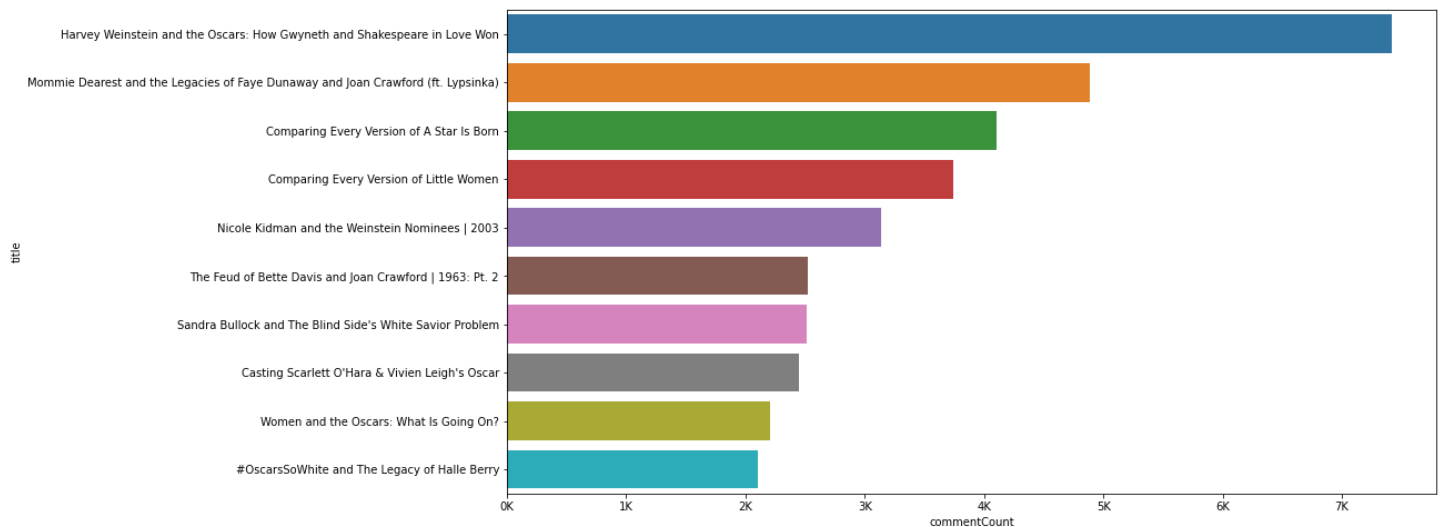
### 4. Distribution of viewCount (visualization through Python)



The plot shows the distributions of the view count. We can see that majority of the videos have view counts around the 300k range with the median at the 200k range. The mean is a little higher than the median because of some videos that have really high view counts.

## 5. Videos with the most comments

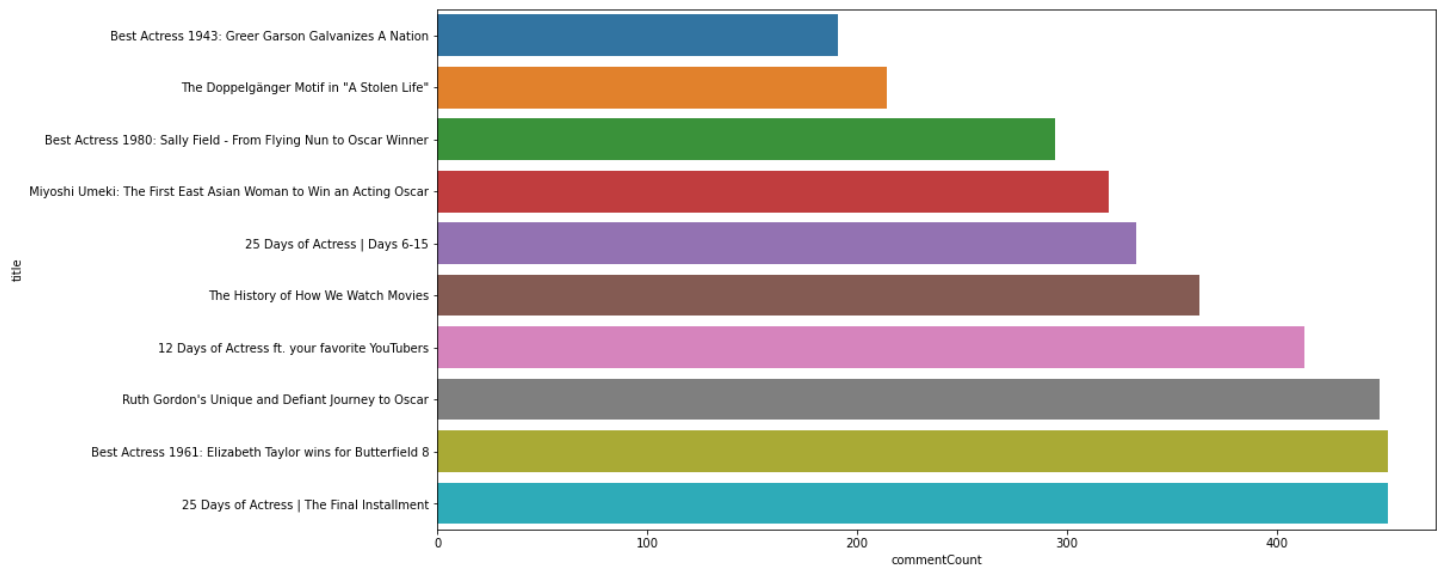
	title	commentCount
1	Harvey Weinstein and the Oscars: How Gwyneth and ...	7419
2	Mommie Dearest and the Legacies of Faye Dunaway ...	4889
3	Comparing Every Version of A Star Is Born	4104
4	Comparing Every Version of Little Women	3742
5	Nicole Kidman and the Weinstein Nominees   2003	3133
6	The Feud of Bette Davis and Joan Crawford   1963: Pt. 2	2520
7	Sandra Bullock and The Blind Side's White Savior Pro...	2513
8	Casting Scarlett O'Hara & Vivien Leigh's Oscar	2446
9	Women and the Oscars: What Is Going On?	2204
10	#OscarsSoWhite and The Legacy of Halle Berry	2103



Videos with the most comments are also those that cover actresses and movies that are more recent as they would be more familiar to today's viewers.

## 6. Videos with the least comments

	title	commentCount
1	Best Actress 1943: Greer Garson Galvanizes A Nat...	191
2	The Doppelgänger Motif in "A Stolen Life"	214
3	Best Actress 1980: Sally Field - From Flying Nun t...	294
4	Miyoshi Umeki: The First East Asian Woman to Wi...	320
5	25 Days of Actress   Days 6-15	333
6	The History of How We Watch Movies	363
7	12 Days of Actress ft. your favorite YouTubers	413
8	Ruth Gordon's Unique and Defiant Journey to Oscar	449
9	Best Actress 1961: Elizabeth Taylor wins for Butterf...	453
10	25 Days of Actress   The Final Installment	453



Videos with the least comments are those that feature actresses and films from the distant past and list-type videos that do not focus on a single actress.

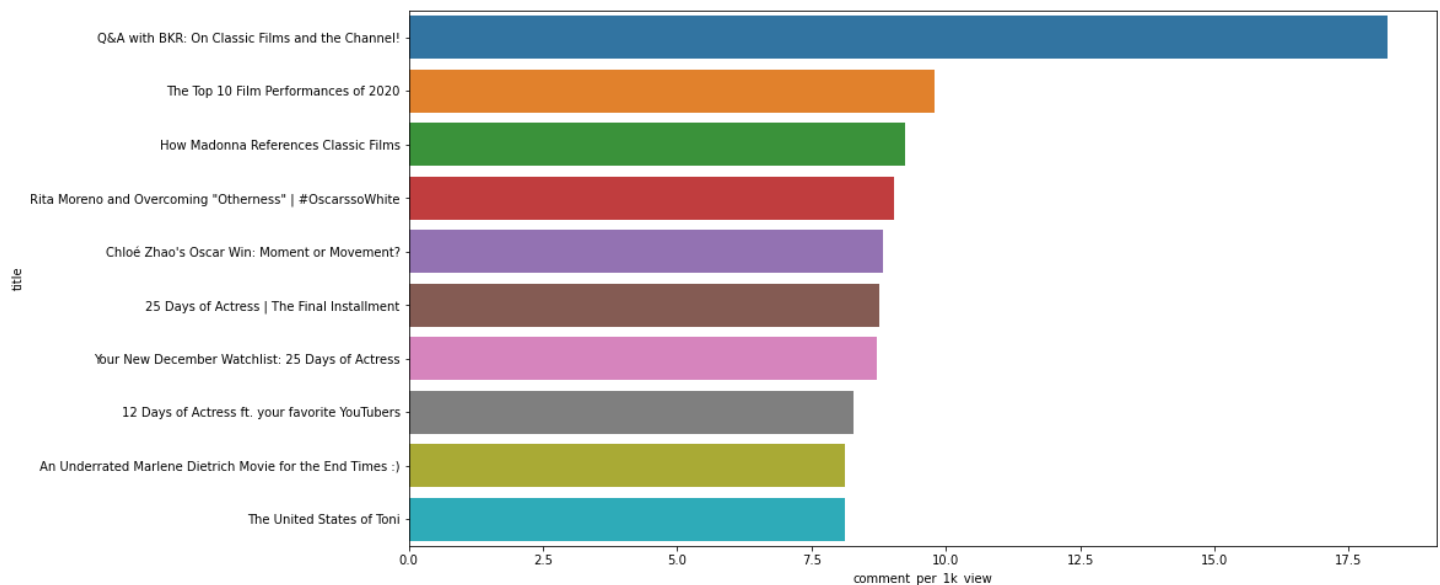
## 7. Average number of comments

	average_comments
1	1246

The average number of comments on the channel's videos is 1,246.

## 8. Video comments per 1000 views

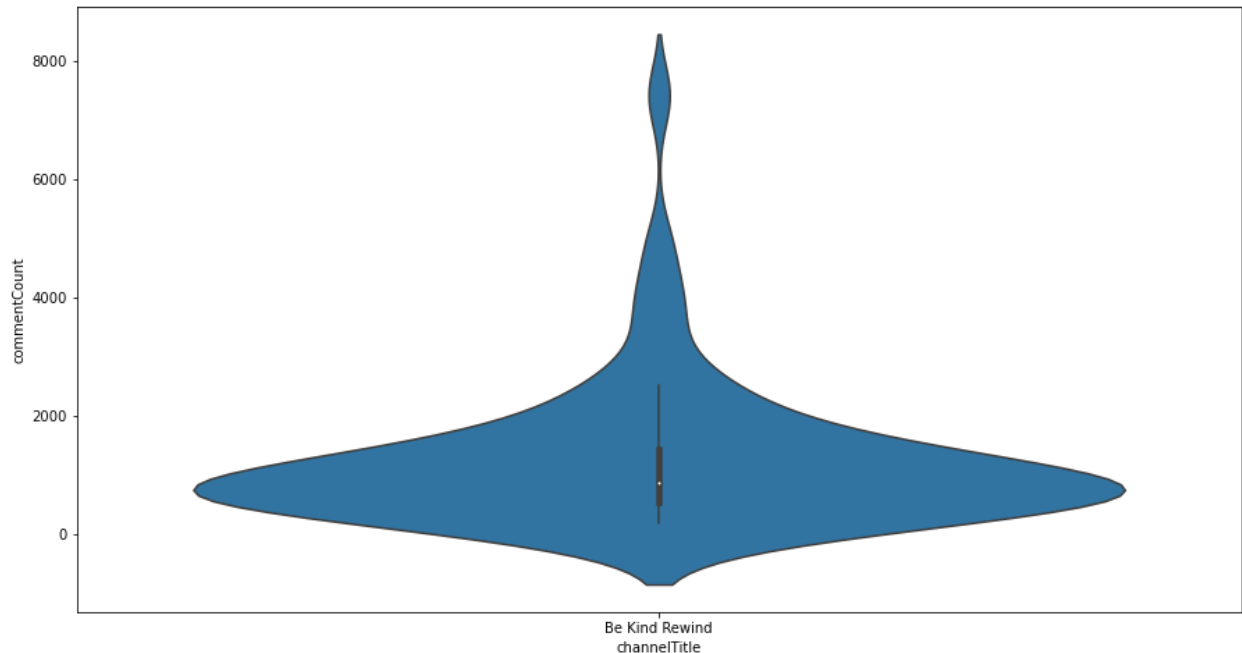
	title	comment_per_1k_view
1	Q&A with BKR: On Classic Films and the Channel!	18.22
2	The Top 10 Film Performances of 2020	9.79
3	How Madonna References Classic Films	9.24
4	Rita Moreno and Overcoming "Otherness"   #Os...	9.04
5	Chloé Zhao's Oscar Win: Moment or Movement?	8.84
6	25 Days of Actress   The Final Installment	8.76
7	Your New December Watchlist: 25 Days of Actress	8.71
8	12 Days of Actress ft. your favorite YouTubers	8.29
9	An Underrated Marlene Dietrich Movie for the En...	8.13
10	The United States of Toni	8.12





Video comments are correlated to views since the more views a video has, the more chance the video will have of getting comments. Thus, to be able to better compare the comment count of videos, we need to consider them in the context of their view counts. The comment count was divided by the total views of each video and then multiplied by 1000. Using this ratio, we can see that the most commented video per 1000 views is the channel's Q&A video. Perhaps the reason behind this is that this video is a way for the creator of the channel to engage with the channel's audience and this, in turn, resulted to more engagement in the comments section.

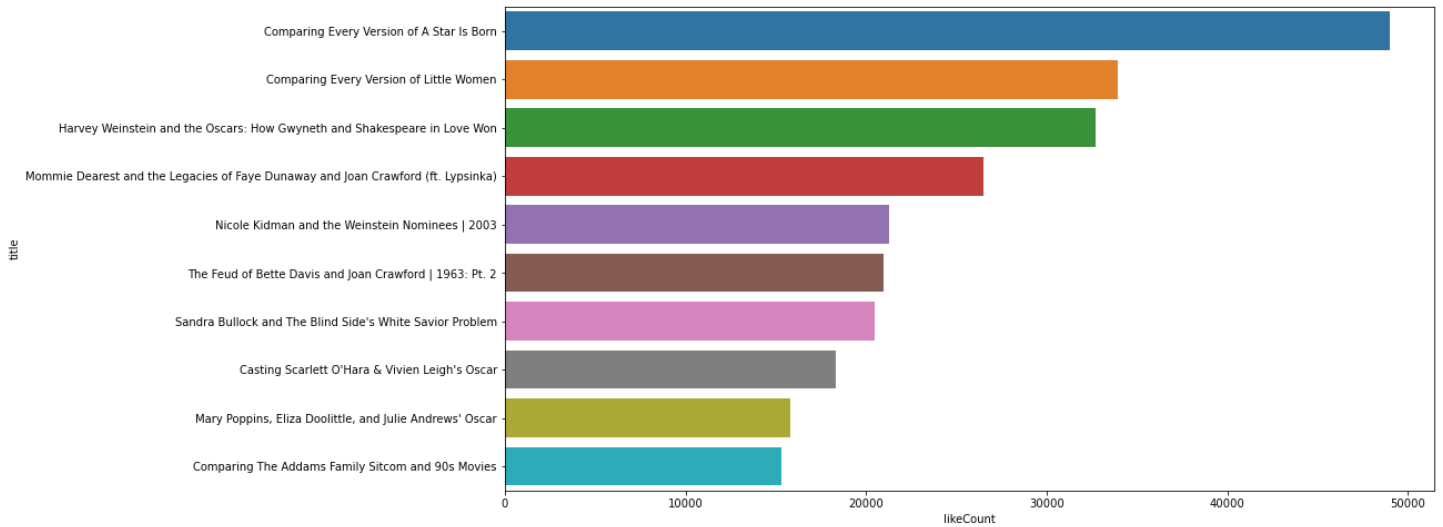
### 9. Distribution of commentCount (visualization through Python)



The plot show that the median of the comments is at around 800 with the mean at around 1200.

### 10. Videos with the most likes

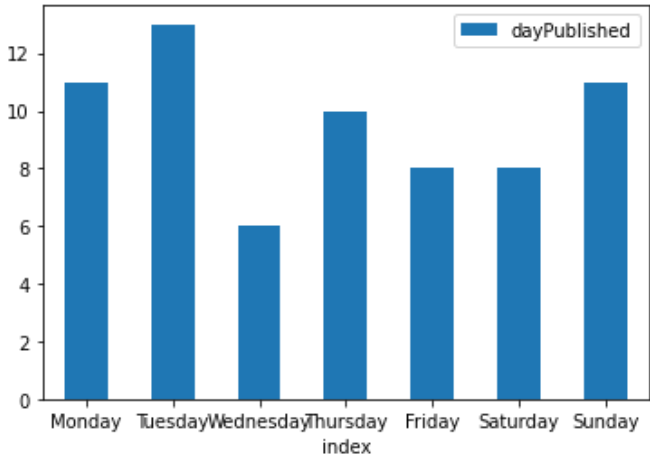
	title	likeCount
1	Comparing Every Version of A Star Is Born	48996
2	Comparing Every Version of Little Women	33953
3	Harvey Weinstein and the Oscars: How ...	32724
4	Mommie Dearest and the Legacies of Fa...	26508
5	Nicole Kidman and the Weinstein Nomin...	21277
6	The Feud of Bette Davis and Joan Crawf...	20953
7	Sandra Bullock and The Blind Side's Wh...	20445
8	Casting Scarlett O'Hara & Vivien Leigh's ...	18303
9	Mary Poppins, Eliza Doolittle, and Julie A...	15815
10	Comparing The Addams Family Sitcom ...	15290



Similar to view count and comment count, the videos the most likes are those that cover actresses and films in the more recent past of Hollywood. These actresses and films are more familiar to today’s viewers compared to actress and films in the distant past of Hollywood.

11. Days of the week that videos are posted

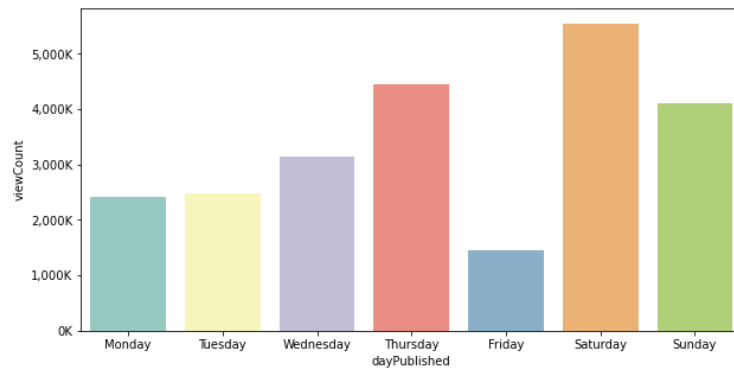
	dayPublished	videosPublished
1	Tuesday	13
2	Monday	11
3	Sunday	11
4	Thursday	10
5	Saturday	8
6	Friday	8
7	Wednesday	6



Most of the videos are posted on Sunday, Monday, and Tuesday.

12. Day of week videos were published and their current view counts

	dayPublished	total_views
1	Saturday	5545057
2	Thursday	4461961
3	Sunday	4102135
4	Wednesday	3149391
5	Tuesday	2479160
6	Monday	2425864
7	Friday	1444196



Videos that were posted on Saturday, Thursday, and Sunday have had the most views so far. However, this could be a function more of their subject rather than a function of the day of the week when they were posted.

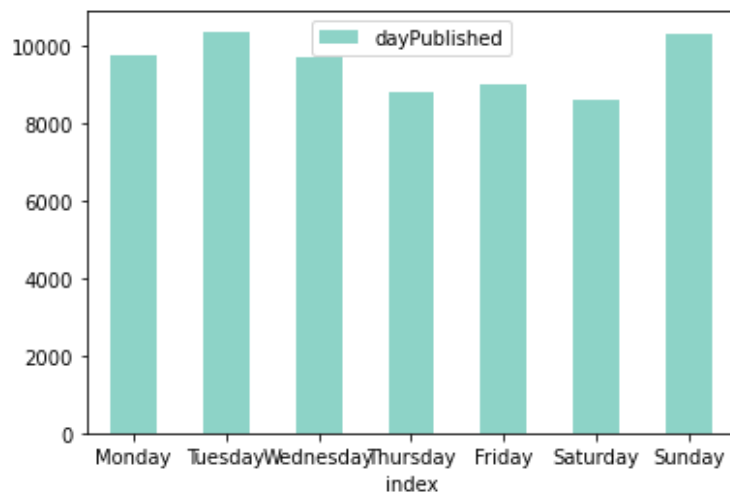
### 13. Average duration of videos in minutes

	average_video_duration_mins
1	23.13

The channel's videos have an average length of 23.13 minutes.

### 14. Day of the week when comments are posted

	dayPublished	comments_posted
1	Tuesday	10361
2	Sunday	10272
3	Monday	9759
4	Wednesday	9707
5	Friday	8976
6	Thursday	8819
7	Saturday	8603



The channel's videos receive roughly the same amount of comments for every day of the week.

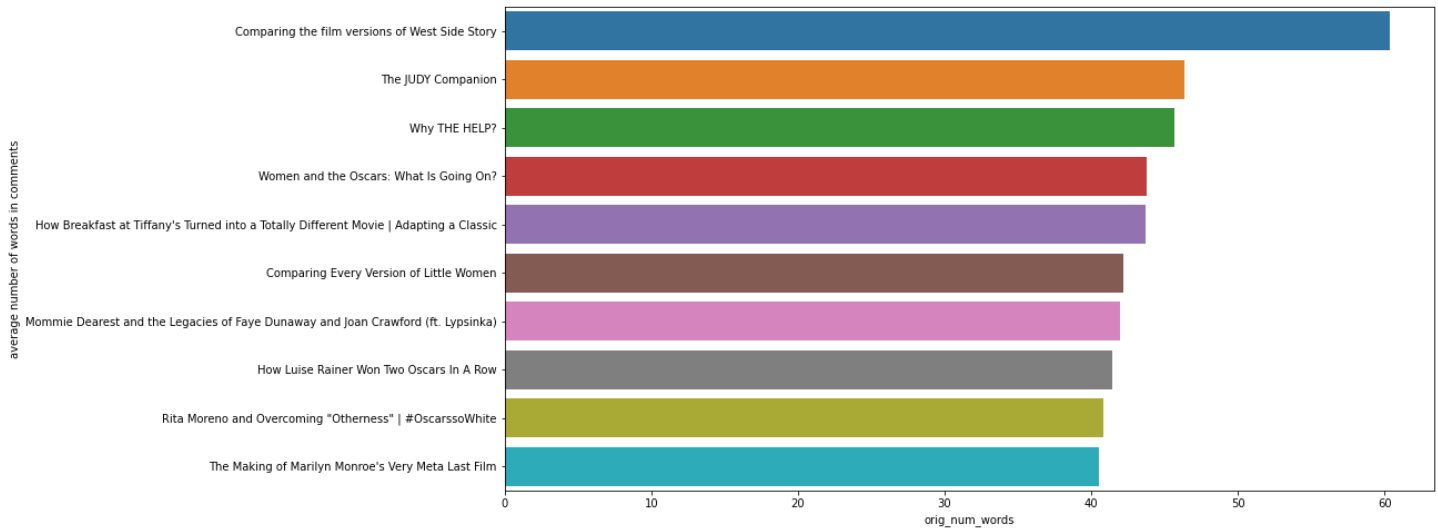
### 15. Average number of words in comments for all videos

	ave_num_words_all_vids
1	33

Comments have an average length of 33 words.

## 16. Top 10 videos with the highest average number of words in comments

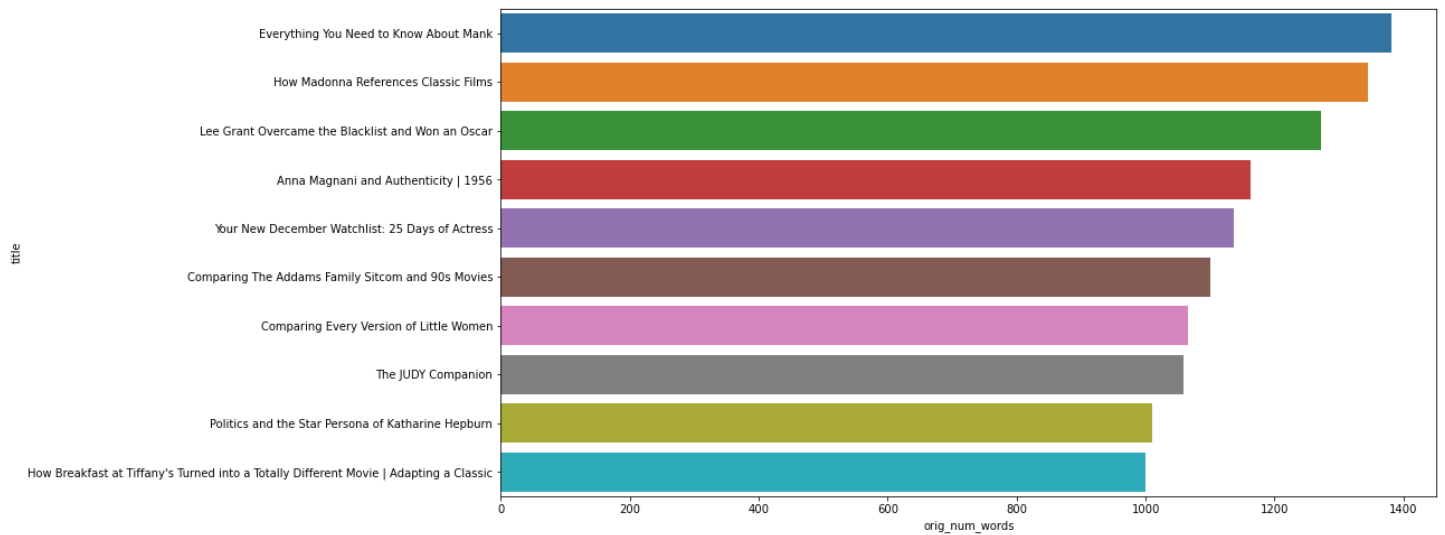
	title	ave_num_words
1	Comparing the film versions of West Side Story	60
2	The JUDY Companion	46
3	Why THE HELP?	45
4	Women and the Oscars: What Is Going On?	43
5	How Breakfast at Tiffany's Turned into a Total...	43
6	Comparing Every Version of Little Women	42
7	Mommie Dearest and the Legacies of Faye ...	41
8	How Luise Rainer Won Two Oscars In A Row	41
9	Rita Moreno and Overcoming "Otherness"   ...	40
10	The Making of Marilyn Monroe's Very Meta L...	40



The video “Comparing the film versions of West Side Story” has the highest average length of comments at 60 words.

## 17. Videos with the top 10 comments with the most number of words

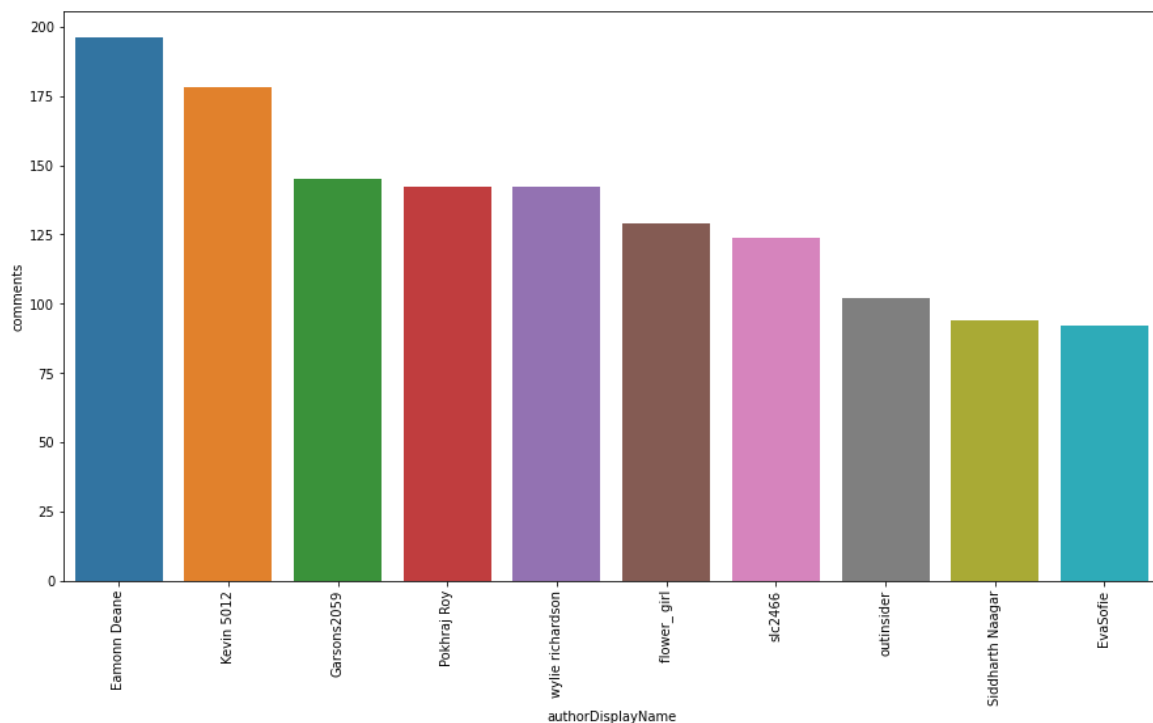
	title	orig_num_words
1	Everything You Need to Know About Mank	1382
2	How Madonna References Classic Films	1345
3	Lee Grant Overcame the Blacklist and W...	1272
4	Anna Magnani and Authenticity   1956	1162
5	Your New December Watchlist: 25 Days ...	1137
6	Comparing The Addams Family Sitcom a...	1101
7	Comparing Every Version of Little Women	1065
8	The JUDY Companion	1059
9	Politics and the Star Persona of Katharine...	1010
10	How Breakfast at Tiffany's Turned into a T...	999



This list shows the videos where the top 10 longest comments were posted. The longest comment with 1,382 comments was posted for the video “Everything You Need to Know About Mank”.

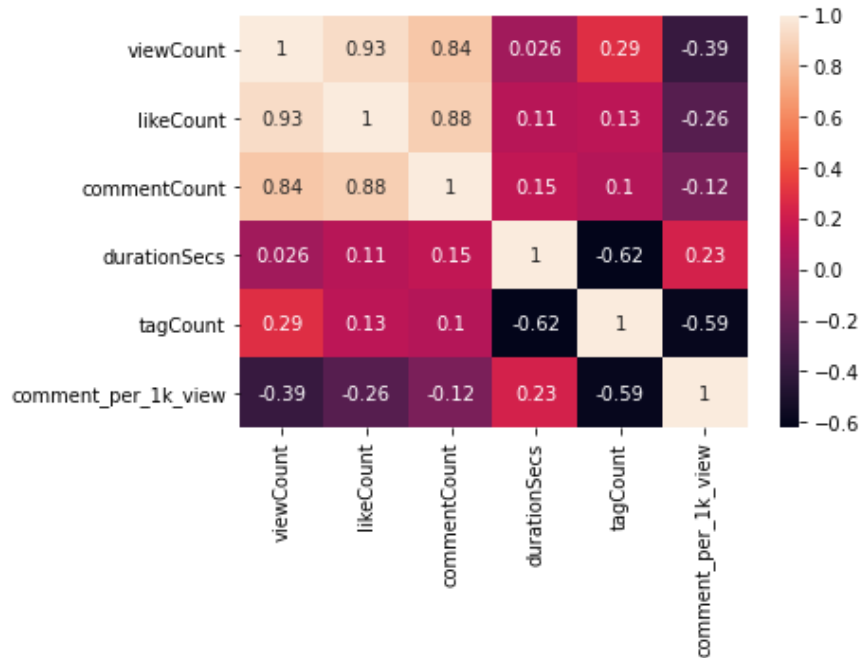
#### 18. Top 10 viewers with most comments on videos

	authorDisplayName	number_of_comments
1	Eamonn Deane	196
2	Kevin 5012	178
3	Garsons2059	145
4	wylie richardson	142
5	Pokhraj Roy	142
6	flower_girl	129
7	slc2466	124
8	outsider	102
9	Siddharth Naagar	94
10	EvaSofie	92



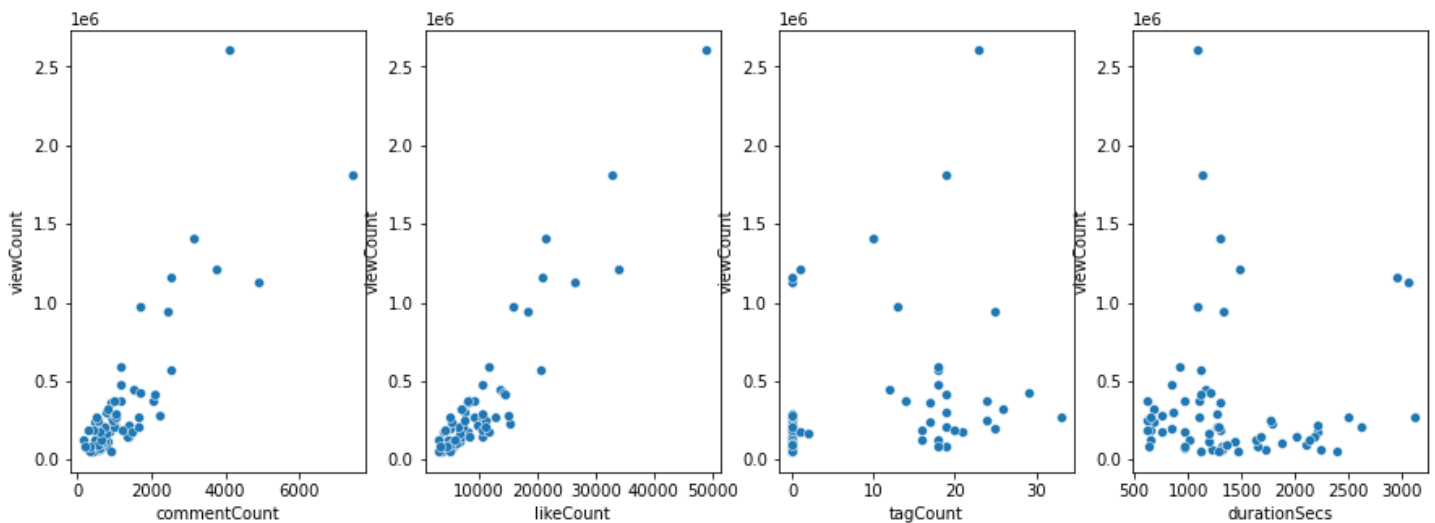
### 19. Correlation of viewCount, likeCount, commentCount, duration, tagCount, and comment\_view\_ratio

	viewCount	likeCount	commentCount	durationSecs	tagCount	comment_per_1k_view
viewCount	1.000000	0.932727	0.842766	0.025714	0.287985	-0.385730
likeCount	0.932727	1.000000	0.876643	0.108549	0.130585	-0.255519
commentCount	0.842766	0.876643	1.000000	0.146403	0.100263	-0.120073
durationSecs	0.025714	0.108549	0.146403	1.000000	-0.621275	0.228176
tagCount	0.287985	0.130585	0.100263	-0.621275	1.000000	-0.589845
comment_per_1k_view	-0.385730	-0.255519	-0.120073	0.228176	-0.589845	1.000000



There is a high correlation between viewCount, likeCount, and commentCount. The more a video is viewed, the chance of it getting likes and comments from the viewers.

### 20. Scatterplot of viewCount against commentCount, likeCount, tagCount, and durationSecs



As seen in the correlation matrix, there is a positive correlation between commentCount and viewCount as well as likeCount and viewCount. Videos with higher viewCounts have higher like counts and comment counts. There is no distinguishable relationship between durationSecs and viewCount as well as tagCount and viewCount.

Word Cloud of Video Comments



## Sentiment Analysis

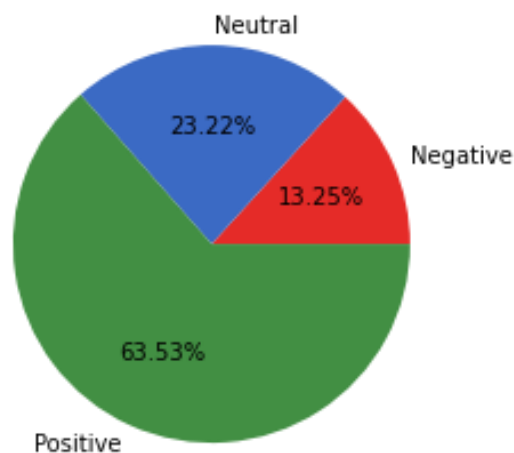
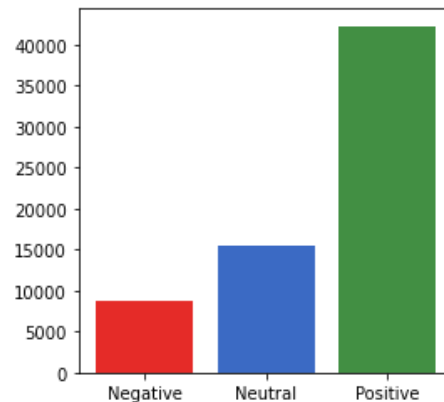
TextBlob was used to analyze each comment and determine their subjectivity and polarity. Polarity defines orientation of a statement whether it expresses a positive, neutral, or negative sentiment. It expressed as a score from -1 to +1 where -1 means a negative statement and +1 means a positive statement. Subjectivity measures the amount of personal opinion or factual information contained in a text. It's value is from 0 to 1 with 1 meaning it expresses personal opinion.

Based on the polarity score, each comment was tagged as 'Positive' if the score > 0, 'Neutral' if the score = 0, and 'Negative' if the score < 0.

### Overall Sentiment

	Sentiment	number_of_sentiment
1	Negative	8810
2	Neutral	15440
3	Positive	42247

avg_positive_percentage
64.88
avg_neutral_percentage
23.48
avg_negative_percentage
11.63

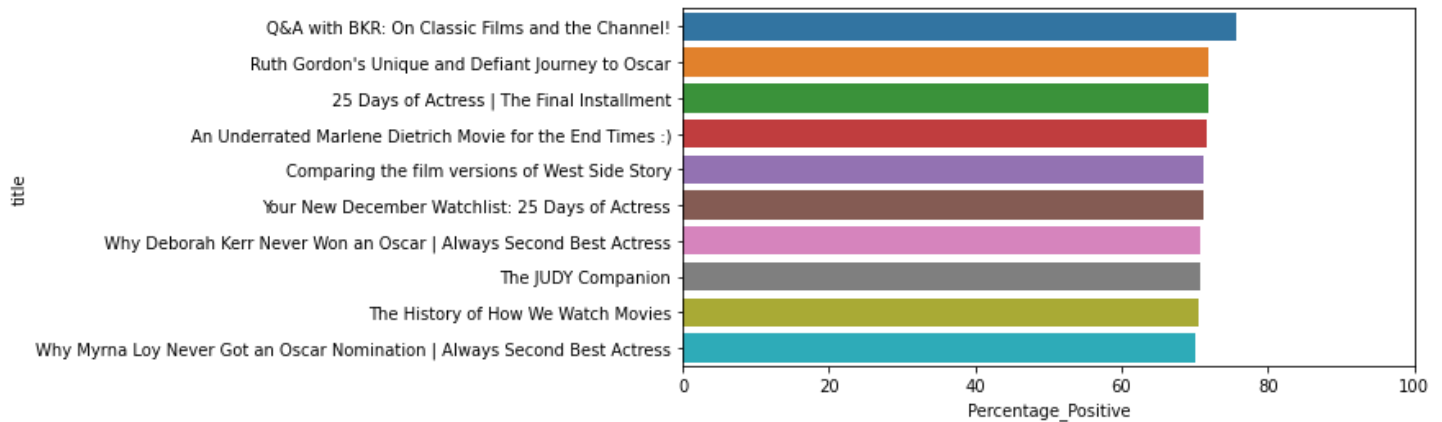


The channel has more positive comments than neutral or negative comments. With 63.53% of the comments being positive, viewers generally like the content available on the channel Be Kind Rewind. The videos have an average of 64.88% of their comments being positive.



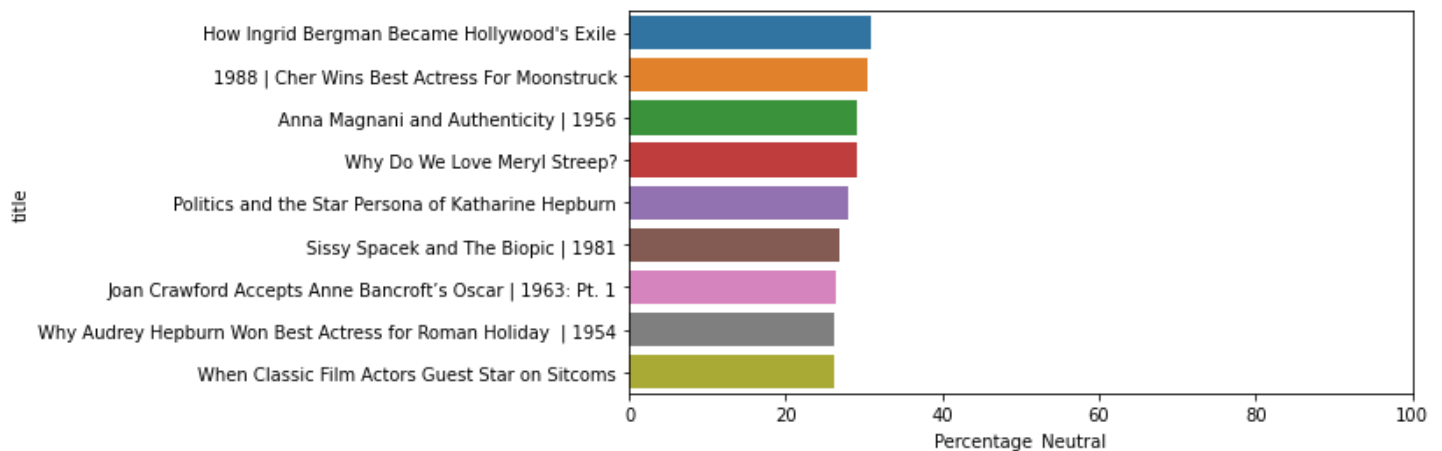
## Top 10 videos with the highest percentage of positive comments

	title	Positive_Percentage
1	Q&A with BKR: On Classic Films and the Channel!	75.71
2	Ruth Gordon's Unique and Defiant Journey to Oscar	71.95
3	25 Days of Actress   The Final Installment	71.82
4	An Underrated Marlene Dietrich Movie for the End...	71.69
5	Your New December Watchlist: 25 Days of Actress	71.09
6	Comparing the film versions of West Side Story	71.09
7	Why Deborah Kerr Never Won an Oscar   Always ...	70.84
8	The JUDY Companion	70.75
9	The History of How We Watch Movies	70.48
10	Why Myrna Loy Never Got an Oscar Nomination   ...	70.06



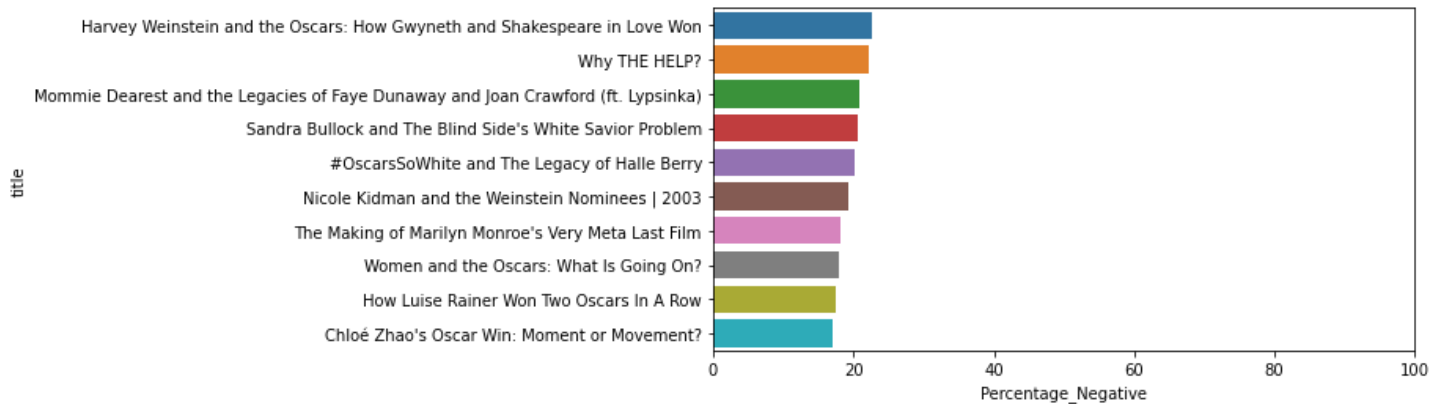
## Top 10 videos with the highest percentage of neutral comments

	title	Neutral_Percentage
1	How Ingrid Bergman Became Hollywood's Exile	30.81
2	1988   Cher Wins Best Actress For Moonstruck	30.37
3	Best Actress 1973: Liza Minnelli wins for Cabaret	30.34
4	Anna Magnani and Authenticity   1956	29.17
5	Why Do We Love Meryl Streep?	29.12
6	Politics and the Star Persona of Katharine Hep...	28.05
7	Sissy Spacek and The Biopic   1981	26.92
8	Joan Crawford Accepts Anne Bancroft's Oscar...	26.43
9	Why Audrey Hepburn Won Best Actress for R...	26.23
10	When Classic Film Actors Guest Star on Sitcoms	26.19



## Top 10 videos with the highest percentage of negative comments

	title	Negative_Percentage
1	Harvey Weinstein and the Oscars: How Gwyneth and...	22.74
2	Why THE HELP?	22.09
3	Mommie Dearest and the Legacies of Faye Dunaway ...	20.74
4	Sandra Bullock and The Blind Side's White Savior Pr...	20.53
5	#OscarsSoWhite and The Legacy of Halle Berry	20.07
6	Nicole Kidman and the Weinstein Nominees   2003	19.35
7	The Making of Marilyn Monroe's Very Meta Last Film	18.22
8	Women and the Oscars: What Is Going On?	17.98
9	How Luise Rainer Won Two Oscars In A Row	17.47
10	Chloé Zhao's Oscar Win: Moment or Movement?	16.94



## Top 10 users with most positive comments

	authorDisplayName	Positive
1	Kevin 5012	124
2	slc2466	102
3	Garsons2059	95
4	Pokhraj Roy	92
5	wylie richardson	86
6	outinsider	75
7	Eamonn Deane	67
8	Orpheus90	59
9	flower_girl	55
10	musicaltheaterge...	51

## Top 10 users with most neutral comments

	authorDisplayName	Neutral
1	Eamonn Deane	99
2	Kosta Jovanovic	51
3	flower_girl	45
4	amy clarke	45
5	RaymondHng	45
6	Siddharth Naagar	37
7	EvaSofie	37
8	Pokhraj Roy	36
9	Garsons2059	34
10	Abandoned Frag...	31

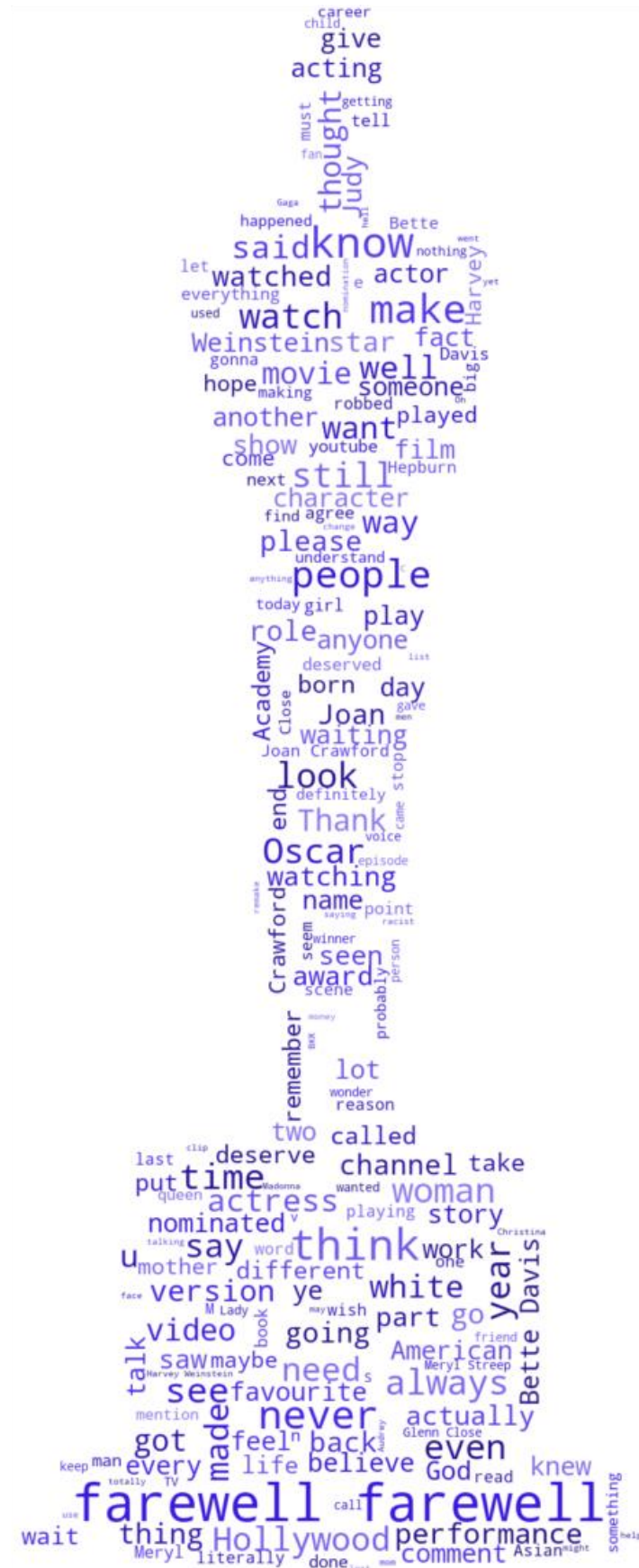
## Top 10 users with most negative comments

	authorDisplayName	Negative
1	wylie richardson	31
2	Eamonn Deane	30
3	flower_girl	29
4	Kevin 5012	24
5	Rickard Kaufman	21
6	Natalie Freeman	20
7	Jarred Knox-Ney...	16
8	Garsons2059	16
9	EvaSofie	15
10	Johanna Zamora	15

## Word Cloud of Positive Video Comments



## Word Cloud of Neutral Video Comments





## Word Cloud of Negative Video Comments



## Conclusion

Based on the sentiments of comments on the channel's videos, we can conclude that the channel is generally well received by viewers. Majority of the comments (63.53%) are positive. The videos have an average of 64.88% of their comments being positive. Some of the most used words in the positive comments are love, great, think, time, good, and woman.