Viral Video Data Analysis

Project Deliverable 1

Team Members: Abhishek Ashwin Bhale (20144060)
Shivani Shashikant Achrekar (20212861)
Harshita Menon Achat (20212860)
Divya Peddapeta (20228697)

Oklahoma State University

MSIS 5600 Data Science Programming 1

Dr. Bryan Hammer

December 10, 2019

Table of Contents

Executive Summary

YouTube is one of the largest video sharing platforms where users and the general public can watch, like, share, comment and upload their videos. YouTube platform consists of two types of users: Video creators (people who have channels and upload videos to them) and Video viewers (people who watch videos, interact with videos and subscribe to channels). Many users use it for entertainment purposes, some to watch tutorials or for keeping up with their favorite artists' latest music videos and so much more. (source: https://www.lifewire.com/youtube-101-3481847)

As of February 2017, there were more than 400 hours of content uploaded to YouTube each minute, and one billion hours of content being watched on YouTube every day. As of August 2018, the website is ranked as the second-most popular site in the world, according to Alexa Internet.

Videos uploaded on YouTube get circulated through the viral process of Internet sharing. These videos may be informative, comic or sometimes deeply emotional. Our objective through this project is to perform analysis on these viral videos.

Statement of Scope

The purpose of this project is to analyze the viral YouTube videos based on the comments, likes, dislikes and the view counts given by the users. This analysis will be done on the viral videos that show up when searched for the Viral Videos on YouTube. Text mining and Sentiment analysis will be performed on the comments of the top viral videos.

Project Objectives

- To perform Visualization on the numeric data to explain what the data tells us.
- To perform Descriptive Statistics on views, likes and dislikes on viral videos. This will help us to analyze the opinions of the users on these videos.
- To perform Text Mining and Sentiment Analysis on the comments of the videos to determine the users' emotions like happiness, anger, sarcasm, etc.
- To build a classification model using a Target Variable and predictor variable.
- To perform Named-Entity-Recognition analysis and generate word clouds.

  The analysis which we performed on the videos published in the month of October is just a sample where we can generalize the analysis to any number of videos.

Variables

We are going to scrape the name of the video, video uploader, number of views, likes, dislikes, video id, length of the video in seconds, duration and the published date. These are the variables on which we will perform Descriptive Statistics. Name of the video, video uploader and video id are the qualitative text variables.

We will also scrape the comments from the viral videos to perform Text mining and Sentiment Analysis on them. For the ease of analysis, we have selected 5 videos from the search results.

Project Schedule

We have used two GANTT charts for our project schedule. The first chart was for our previous schedule and the second chart is for the revised schedule. Every week we meet on Wednesdays and plan to meet on every other following Wednesday and plan to work for 3 to 4 hours. All the tasks are done together as a group. The Project might take around 5 weeks more to get completed. In every meeting new plan is created and we work on these plans for the following week. As we had exams in the month of September there was a slight change in the schedule which lead to a delay in our schedule. The previous chart depicts that there was a Video presentation to be completed but as a revised schedule, the video presentation got cancelled. And the word cloud function was being added and that took up a while. Though there was a change in the pattern and schedule of the project, we were resilient to the changes and have gathered all the data so that the project could be completed on time.

Research Task	August				September				October				November				December	
	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2
1. Project Proposal																		
2. Executive Summary																		
3. Statement of Scope																		
4. Data Access																		
5. Data Consolidation																		
6. Data Cleaning																		
7. Data Transfromation																		
8. Data Reduction																		
9. Data Dictionary																		
Submission of Deliverable 1																		
1. Adjust Deliverable 1 Requirements																		
2. Descriptive Statistics, Visualizations, and Analysis																		
3. Text mining and Sentiment Analysis																		
4. Formatting, Style, Grammar and Spelling																		
Submission of Deliverable 2																		
1. Compile Final Report																		
2. Presentation																		
3. Record Presentation																		
Final Deliverable: Presentation																		

Research Task	August				September				October				November				December	
	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2
1. Project Proposal																		
2. Executive Summary																		
3. Statement of Scope																		
4. Data Access																		
5. Data Consolidation																		
6. Data Cleaning																		
7. Data Transfromation																		
8. Data Reduction																		
9. Data Dictionary																		
Submission of Deliverable 1																		
1. Adjust Deliverable 1 Requirements																		
2. Descriptive Statistics, Visualizations, and Analysis																		
3. Text mining and Sentiment Analysis																		
4. Formatting, Style, Grammar and Spelling																		
Submission of Deliverable 2																		

Data Preparation

Data Access

Here, we searched for the 'Viral Videos' on YouTube as our search query and analyze the top 100 videos that show up after the search. For accessing the data, we are using a code which will scrape the links to the top videos that were obtained with the help of Selenium and CSS selectors. Out of all the links that are obtained using Selenium, we will only be using the top 100 links for the videos. We used a library in Python named pafy which gives out the likes, dislikes, view count, name of the video and the publisher and many more. All of the data comes from YouTube itself.

Source: (https://www.youtube.com/, https://pythonhosted.org/Pafy/)

Data Consolidation

We have done the consolidation process using Python code which we have attached in the Appendix at the end of this document.

CSV file of YouTube Data.

	A	B	C	D	E	F	G	H	I	J
1	X	VideoId	Title	Uploader	PublishedDateTime	Viewcount	Likes	Dislikes	Length_in_	Duration
2	0	rqnAPo-r8	Top 100 Be	Newsflare	11/30/2019 17:15	310390	2677	179	1941	0:32:21
3	1	db9EJdbKK	3, 2, 1.... F	America's	12/4/2019 14:00	58175	903	82	626	0:10:26
4	2	nTWnU-Ac	Top Viral V	Newsflare	8/29/2019 18:15	2673083	20718	1977	1188	0:19:48
5	3	mq7yszUJI	Top 50 Be:	Newsflare	11/9/2019 17:30	876521	5389	516	822	0:13:42
6	4	hFLZgMHd	PART 1 | V	Raffy Tulfo	12/2/2019 11:22	4882606	87206	3547	1231	0:20:31
7	5	NC__B8j_	Top 60 Vir	Newsflare	10/26/2019 16:00	1664966	12552	1079	1042	0:17:22
8	6	vGVatXSG	50 Best Vir	Newsflare	9/29/2019 18:15	1041456	6969	547	953	0:15:53
9	7	v9pOERMr	Top 40 Vir	Newsflare	7/1/2019 14:00	900744	6846	698	1104	0:18:24
10	8	mwENYk6	Hilarious C	Newsflare	5/7/2019 10:08	12233245	119978	5392	557	0:09:17
11	9	uQl-r3pqn	Top 100 Vi	This Is Hap	12/30/2018 16:48	41228504	223496	22796	1844	0:30:44
12	10	xQGniTrrf1	PART 3 | V	Raffy Tulfo	12/4/2019 13:22	1628757	32617	1438	1824	0:30:24
13	11	l4a7MBfz-	50 Best Vir	Newsflare	8/18/2019 17:00	5375783	52638	4993	898	0:14:58
14	12	FC31lxet5E	11 Most U	#Mind Wa	8/18/2019 18:01	6207175	54418	2828	719	0:11:59
15	13	VFn3HxKm	She CAN'T	America's	11/30/2019 14:00	256073	1954	117	625	0:10:25
16	14	1asKDNnjZ	Dogs hilari	Newsflare	11/22/2019 16:30	586690	16855	372	81	0:01:21
17	15	cWiPROxS	Top 100 Vi	This Is Hap	12/23/2017 16:00	65694818	437043	34372	1757	0:29:17
18	16	Z9jnQdmlI	PART 2 | V	Raffy Tulfo	12/3/2019 15:10	3498769	48797	2094	882	0:14:42
19	17	WxRw73Q	Top 50 Vir	FailSnare	10/9/2019 15:40	150015	601	64	602	0:10:02
20	18	2_NzEpRV	Top 40 Vir	Newsflare	12/22/2018 19:52	13001037	80260	6248	487	0:08:07
21	19	k8-S7VkQp	The 100 BE	America's	1/9/2019 14:00	12815758	72277	6749	718	0:11:58
22	20	68ORjuHT:	Top 50 Fur	America's	9/27/2018 15:15	10058776	22109	3048	495	0:08:15
23	21	iCBPf4alAF	Top Viral V	Newsflare	4/28/2019 18:45	1751992	11307	1103	804	0:13:24
24	22	mFLoerQIF	Top 5 Hot	Shams Stu	1/7/2019 16:30	58107	162	86	360	0:06:00
25	23	Qk4acvdyy	Top 10 Vir	WatchMoj	12/12/2014 15:30	13774051	91807	6143	850	0:14:10
26	24	yDTk6u2o'	TOP VIRAL	Mr Kuchbh	2/15/2019 17:21	40712	70	27	100	0:01:40
27	25	ys_uWgg8	Dressed Fo	America's	11/29/2019 14:00	566850	2948	311	635	0:10:35
28	26	oSdUTQM	Organizati	RED BD	10/1/2017 17:36	358165	404	138	430	0:07:10
29	27	b-1l-Am55	Trump call	CityNews	12/4/2019 22:50	360	41	4	188	0:03:08

We have created 5 different CSV files for the comments on different videos to make the analyis simple.

CSV file for comments on the 1st video.

X	Comments1
0	Anyone else feel majorly uncomfortable looking into the underwater depths of that crystal clear lake ??
1	That little girl with the Star Wars characters was so sweet
2	Anyone else waiting on the shark in the crystal clear lake lol
3	Please show us the door not closing a few times more so we really get it.
4	Well 0:30 made me realise that I'm terrified of deep and clear water
5	That little Star Wars girl is awesome.
6	What did we learn? Blue kayaks are faster than red ones.
7	Some great vids, but thumbs down for unnecessary repeats
8	18:11 give that chicken momma an umbrella please<3
9	If only more people could parallel park like that kid, wow!
10	3:09-3:44
11	That monk's finger punch had me like 'F**K!!!' That dude can crush faces with two freakin fingers.
12	0:30
13	Stop showing clips twice and we can get through this a bit quicker
14	4:15 when your tools broke so now you gotta mine with your bare hands.
15	Poor duck he gives shoes back and a week later they will eat him
16	2:04 Litterally suffering from anxiety...he jump so close towards the pool wall ðŸ˜¶
17	Take the door off the hinges then back on, better to not be able to open all the way than not close at all
18	4:10 that monkey was like, oh crap! I didn't do it, their fault!
19	For some reason at 8:00 the red boat spinning makes me laugh so hard XD
20	0:22 why is a shark the only thing I can think of?
21	"oh daddy cakes where are my goggles ?"
22	If anyone has any objections to this marriage.. GOD: Bzaaap
23	There were some really cool clips but the helicopter and camera being synched up at #34 was amazing.
24	13:14 That cat is already missing part of it's tail- does it have to put up with a kitten attacking what's left? Poor tubby cat
25	15:56 Persian Cat Massages Owner Persian Cat Rips Owner's Skin
26	0:24 is the most scared video a ever seen in my life, I fell like a monsters was going jump from the deep of that water.
27	0:30 was getting anxious waiting for a damn shark or something to swim up to him fastðŸ˜°

CSV file for comments on the 2nd video.

X	Comments2
0	2019 and people still filming with the phone the wrong the wrong way
1	The sleep walking girl on piano was scarier than the haunted house
2	1:20 that Halloween decoration is so simple and yet so brilliant. Well done sir ðŸ‘ŒðŸ½
3	Um sleepwalking piano player needs an exorcism lol that would freak me the eff out!!
4	8:40 Best friends or trying to fatten the mouse up??
5	Don't let your kids eat whilst falling asleep. They can breathe food into their lungs which can cause choking and aspiration pneumonia
6	Thumbs up forthe Halloween house. :)
7	Whose head is spinning while watching the guy with parachute..
8	The House was too cute (those eyesðŸ‘€ðŸ‘€)ðŸ‘ðŸ‘ðŸ‘
9	The power of the sea frightens me to no end.
10	8:53 Plot twist Cat feeds mouse so that it can taste cat food also when it eats the mouse later
11	9:30 That's doggo for You took er jerb.
12	Giraffe: Look! It has a long neck just like me! A bird-giraffe!
13	I say we put the sleep walking girl and the robot in a piano competition against one another
14	I'll stick to my Schick Hydro 5 for shaving my hair, thank you. Not sure about that wood plane.
15	3:35 Yeah. It's way more important to tap on your smartphone than put that damn thing aside and reach for your baby... ðŸ™„
16	That guy with the parachute is nuts. He'll end up dead one day
17	6:35
18	That guy running that pony on cement should be shot.
19	@5:19 Nope.
20	2:51 Please help the bulldog into the car. He wants to drive it so bad.ðŸ˜,
21	Oh my god. You saved that babe.ðŸ˜-ðŸ˜- god bless you.
22	2:52: me trying to get in bed
23	:13 Dang! That was totally worth...
24	Always the practical one, I am watching the guy flying through the air over the mountains and wonder how he is going to get home?
25	2:51 my mom trying to get into my truck
26	Not gonna lie that spooky house looks kinda cool.
27	Being a thief now adays is easy especially in Halloween.

CSV file for comments on the 3rd video.

X	Comments3
0	Check out our September top virals of the month!
1	That little girl is without a doubt the cutest thing I've ever seen in my life.
2	I call BS on the "fastest readers ever" clip lol.
3	How many of you cried when you heard that what she says "My heart fell in love with you"
4	The little girl with her adopted parents is the cutest thing I've ever seen.
5	The little adopted girl is the cutest video of the decade. What a precious little soul.
6	That little girl is wise and eloquent beyond her years! Why are my eyes leaking?!
7	The adopted asian girl made my dayðŸ˜ªðŸ‘¤
8	The adopted little girl was outstanding and so sweet.
9	That little girl @ 10:45 in amazing. I have never in my life heard someone so young talk so...adult like. And when she whispered about her heart like it was some huge secret she was c
10	On the train where the guy is singing Bon Jovi the others passengers should have totally sung the chorus
11	That little adopted girl is freakinâ€™ adorable and those fingernails are freakinâ€™ hideous.
12	My husband gagged like that with every poopy diaper. But he still changed them - and never complained. Because thatâ€™s what good dads do. ðŸ¥°
13	Those long fingernails are disgusting!
14	That cat loves him so much. So cute
15	I refuse to believe the kid flipping the book was reading anything. The book wasnâ€™t even open entirely. Canâ€™t see all the words on the page no matter how fast you read.
16	I can't believe they actually tried to push the car sideways to get it into the spot haha.
17	When after he falls in the muck and then they said they aren't paying him. That's not right
18	Lol I used to mimick my son crying and got the same effect. Heâ€™d just stop and look at me in amazement too.
19	The longest nails in the world made me want to throw up.
20	That beautiful little adopted girl speaking such awesome love. ðŸ¥°11:45 is just the most sweetest precious thing!
21	the first one is sad. the poor calf had been separated from his mother.
22	This is first time I told you.
23	Omg when the little girl said her heart fell in live with her mom I cried it wad so sweet and cute.
24	That Asian girl who would profess love towards her adoptive parents â€¦ I would dedicate my eternal love and soul and life fortune (were she to out-live me) to someone like that.
25	Lol.. that one guy changing a diaper with gloves on ðŸ¤£ðŸ¤£ðŸ¤£
26	my heart fell in love with somebody today, yet i'm crying my eyes out... has to be due to that luck. right?
27	How is that girl adopted that girl is so cuteðŸ˜-ðŸ˜-

CSV file for comments on the 4th video.

X	Comments4
0	#1 wow that bird sounds like a baby. i would make u show me the box before i walked away
1	That bird has it's human trained well.
2	11:12 -- Little girl: "I'm not a princess, I'm a queen!"
3	6:53 the guy playing soccer with his dog ðŸ•¶ awesome ðŸ‘ðŸ½
4	Why on earth is there a toddler on a boat without a life vest?
5	The glass walkway with the little girl made my heart stop
6	Boy...the guy with the scooter...5 accidents in 1 minute...
7	That art teacher tho! Spectacular! Wish heâ€™d been my art teacher.
8	I mean,is pretending to bang your kids head and making them cry a craze?It's just weird.
9	The teacher that draws on the board ðŸ¥°ðŸ˜¥ðŸ˜¥
10	"I'm not a princess, I'm a queen".
11	1:45 I have arachnophobia now.
12	Lmfao omg he sounds like a baby !!
13	"I'm not a princess. I'm a Queen" YAAASSSS GIRL!!! Love her so much and super adorable!!!
14	This is a spectacular viral video that crying bird tho lol love this channel
15	Wtf fell in a mega hole after all the accidents. Hahahaha
16	2:16, What!? How did they do that?
17	teacher does a drawing stretching over three blackboards length,so i suppose the natural thing to do these days is to hold the camera vertical so you can only see a narrow strip at a time.
18	9:57 minute epic ðŸ˜,ðŸ˜, ðŸ˜, hence
19	That parrot so smart! Learned the one sound that would make just about any human, and apparently also some other animals, investigate to see what the hell is going on
20	Parrot crying like a baby is creepy
21	All your clip amazing the quality fabulous wow wow thank you ðŸ‘â¤ðŸ‘ðŒðŸ‘šðŸ‘šðŸ‘šðŸ‘š
22	The dog in the stairwell was priceless
23	12:28 nice jobðŸ‘ðŸ¼ now the left side.. gotta even it out.
24	Hhhh a hahaha retina scanner... Such a hilarious father in law.. Love him so hilariously funny.
25	Baby telling dad her way that she doesnâ€™t like his singing ðŸ˜,.
26	9:52 ðŸ˜,ðŸ˜,ðŸ˜,ðŸ‘
27	Oh yeah thatâ€™s my art teacher

CSV file for comments on the 5th video.

X	Comments5
0	Which one is best
1	Who is a like
2	For that hanger jacket thing WHO EVEN SNEEZES OF COUGHS LIKE THAT
3	Which one is the best
4	This is how many people love 123 GO.
5	Hi most people will not read this but
6	Who else watches these, but never tries them? I do!
7	1:11 She is so funny throwing the dust to those who think I am mean i meant when laughing has her entertaining
8	I never knew you guys cursed! Btw you said â€œDamn Vickey!â€
9	5:08 they said damn
10	3:03 could see the hole.like if yes.
11	The subway one was apperent motion
12	1:13 Alan and Alex stokes do that
13	VickyðŸ˜ðŸ˜˜ on bloopers your actions are like Mr.BeanðŸ˜
14	Love â¤ï¸ you
15	I love this vid THUMBS UP!!
16	When you put the powder I sneezeðŸ˜ª
17	Leave a like if you love 123 go.
18	Vickie is my favorite.
19	"That can be our little secret!"
20	6:48 te dio sida? :'v
21	Thatâ€™s sneaky
22	Roses are red
23	2020 I think u should reveal your voice
24	Well Vicky is already in 5 minute crafts but I am going to have to say 123 go
25	6:29 look in the pot its water she said its noodle soup
26	5:13 she DID eat "his" face off
27	3:50 Lyp synch sister we know

Data Cleaning

There are few videos that have missing data for Likes and Dislikes columns. So, as to handle this missing data, we did Listwise deletion to perform descriptive analytics and visualization. The comments consisted of some unwanted data like the timings when the comments were posted and emoticons also appeared in the comments' section. We used regular expressions to remove these timings and emoticons. In this way, we cleaned the comments and made them useful for text mining and sentiment analysis. There is no more specific data that needs to be cleaned or any variable that needs to be dropped.

Data Transformation

For the variable Viewcount, the data that is obtained is in lakhs, to plot this data we are normalizing it by dividing it by 1000. We are doing this so that the visuals of the plot can be easily analyzed.

We are transforming the comments that were obtained for the viral videos by converting them to strings. This will help us in performing Text Mining and Sentiment Analysis.

Data Reduction

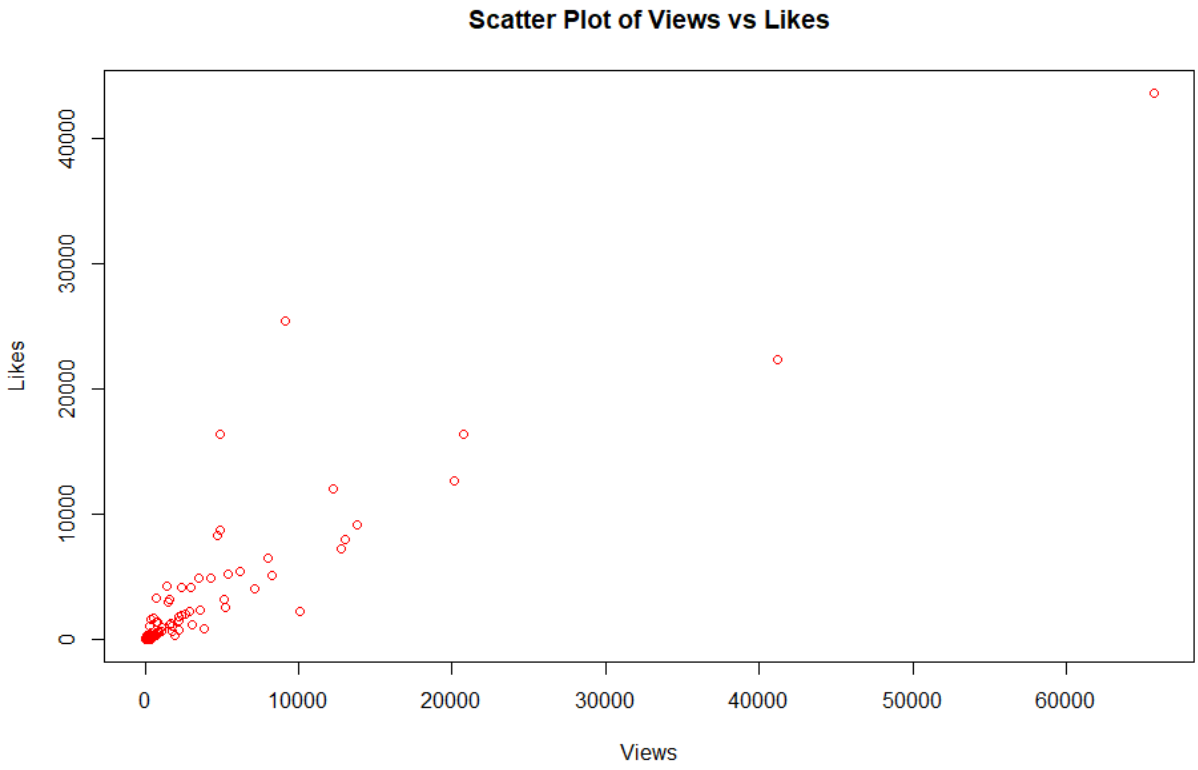
As of now, we are pulling data from only 100 YouTube videos. And the number of variables we have are 9. So, this is the exact data that we will need for our analysis. If needed, we can increase the number of YouTube videos that we will be pulling the data from. For now, we won't be performing any Data Reduction on the data that we have scraped.

Data Dictionary

Attribute Name	Description	Data Type	Source
VideoId	It is the alphanumeric unique id of the individual videos	String	https://www.youtube.com/
Title	Name of the video	Char(30)	https://www.youtube.com/
Uploader	Name of the channel that published the video	Char(30)	https://www.youtube.com/
PublishedDateTime	The time and date the video was uploaded	Date time format as String	https://www.youtube.com/
Viewcount	The number of users who viewed the video	Integer	https://www.youtube.com/
Likes	Users who liked the video	Integer	https://www.youtube.com/
Dislikes	Users who disliked the video	Integer	https://www.youtube.com/
Length_in_sec	The length of the video in seconds	Integer	https://www.youtube.com/
Duration	The duration of video in HH:MM:SS format	String	https://www.youtube.com/
Comments	Feedback given on videos	Char (30)	https://www.youtube.com/

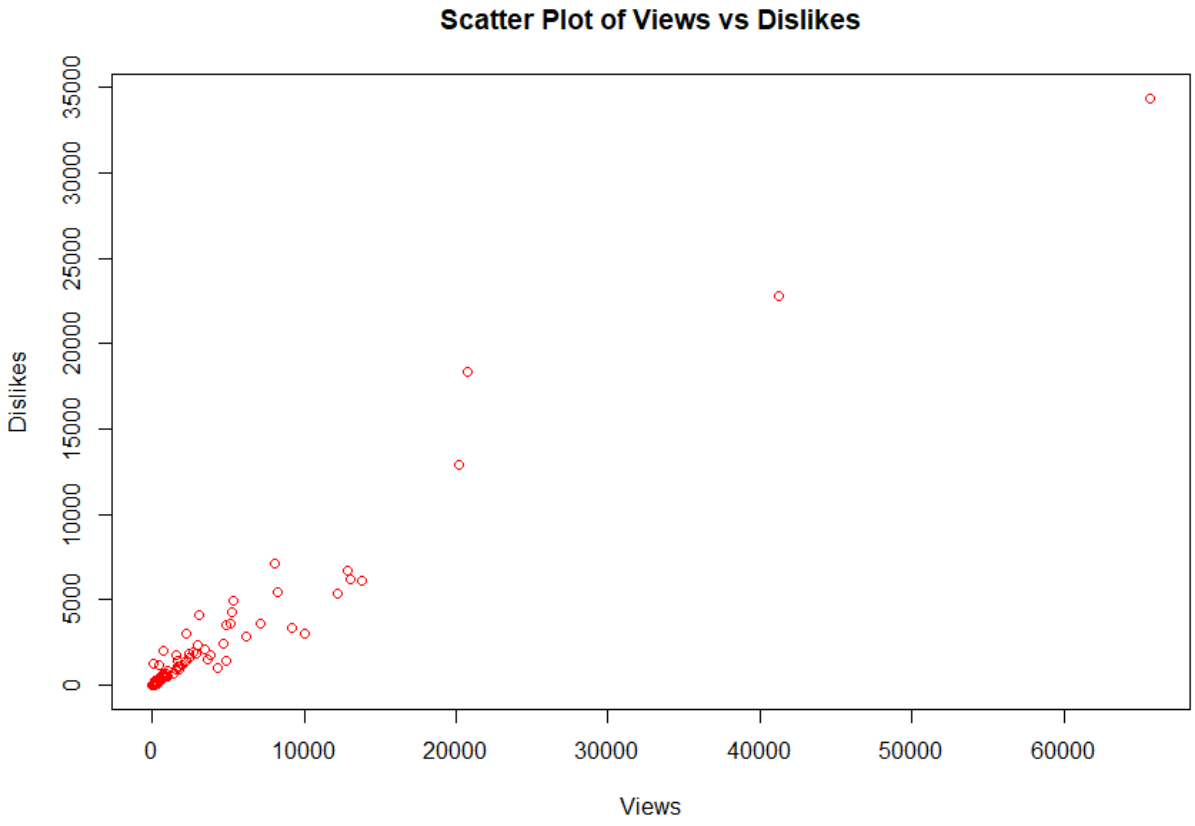
Visualizations

Data visualizations help us depict and understand data using graphs and various kinds of plots. The pictorial representation of data helps us to interpret data more effectively. A scatterplot shows the relationship between two continuous variables. In our project, the proposed target variables are number of Likes, Dislikes and Length_in_sec. The predictor variable is the Viewcount.

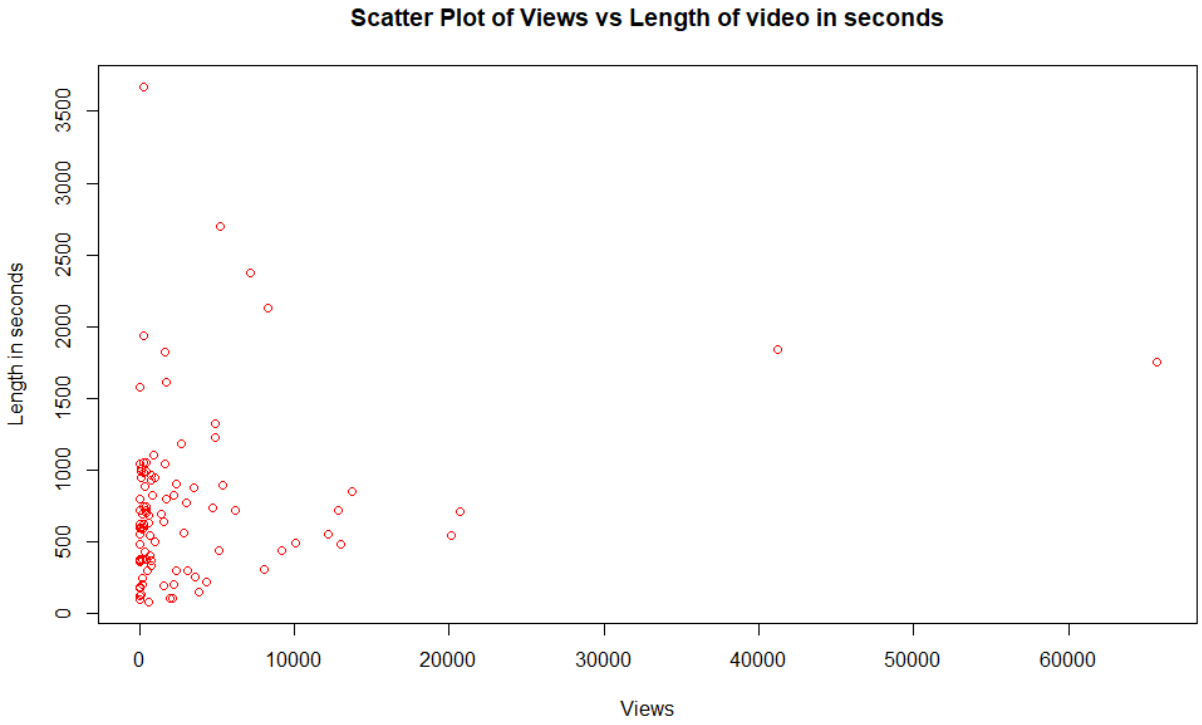


The graph above depicts the relationship between Views and Likes. The scatterplot depicts a strong positive linear relationship between Views of videos and Likes of videos. But there appears to be a couple of outliers in the data. Most of the videos' views range from 10 views to 6,000,000 views. The views are in millions, so we have divided them by 1000. The scatterplot also depicts potential outliers. The Likes on the video are in the range of 10,000 so for better visualization we have divided it by 10. A few videos have more than 6,000,000 views and by here we eventually conclude

that most of the times as the views increase the Likes also increase i.e., if more people tend to watch the video, they would tend to have an opinion about the video.



The graph above depicts the relationship between Views and Dislikes. The scatterplot depicts a strong positive linear relationship between Views of videos and Dislikes of videos. But there appears to be a couple of outliers in the data. Most of the views of the videos range from 10 views to a 6 million views. The view values are divided by 1000. The scatterplot also depicts potential outliers. A few videos have more than 20,000 views and by here we eventually conclude that most of the times as the views increase there is a chance of people not liking the video and hence the dislikes increase. i.e., if more people tend to watch the video, they would tend to have an opinion about the video.



From the graph above it can be interpreted that, the duration of the video does not affect the views. Irrespective of the duration in minutes the number of views does not change. The plot is a vertical line depicting that the views are constant. There are also a few potential outliers in the data where the views for an 80-minute video is more than 60,000 according to the plot.

In our modelling technique, we are selecting Viewcount as our predictor variable and Likes, Dislikes and Length in seconds as our dependent variable.

Descriptive Statistics and Analysis

```
> summary(youtube_data)
       X                VideoId
 Min.   : 0.00   _k112zTWVsM: 1
 1st Qu.:24.25   0GUT3Egdd5k: 1
 Median :48.50   1asKDNnjZlc: 1
 Mean   :48.88   2_NzEpRVGxQ: 1
 3rd Qu.:73.75   23f94jncwmg: 1
 Max.   :99.00   2fSV5LrxISM: 1
                 (Other)    :92
                                                                                                  Title
 viral video of the week  | Top 10 viral                                                             : 3
 Trump calls Trudeau â€˜two-facedâ€™ after NATO video goes viral                                      : 2
 #Today Vigo viral videos#trending Vigo comedy video songs | latest tiktok funny video jokes 2019: 1
 (use headphones) musically adult viral video ðŸ'žðŸ'ž                                               : 1
 10 FAKE Viral Videos That FOOLED Everyone                                                        : 1
 11 Most Unusual Kids in the World                                                                : 1
 (Other)                                                                                          :89
                          Uploader            PublishedDateTime   Viewcount            Likes
 Newsflare                    :17   2011-09-08 16:56:47: 1      Min.   :      16   Min.   :     1
 Funny Vines                  : 8   2014-12-12 15:30:52: 1      1st Qu.:  165715   1st Qu.:  1509
 This Is Happening            : 7   2015-01-03 18:00:06: 1      Median :  717613   Median :  6002
 America's Funniest Home Videos: 5   2015-11-30 18:00:00: 1      Mean   : 3552095   Mean   : 29989
 Raffy Tulfo in Action        : 3   2015-12-26 18:00:00: 1      3rd Qu.: 3058179   3rd Qu.: 31096
 TOP 10 VIRAL                 : 3   2016-11-26 01:00:01: 1      Max.   :65694818   Max.   :437043
 (Other)                      :55   (Other)            :92
    Dislikes       Length_in_sec        Duration
 Min.   :    0.0   Min.   :  81.0   00:01:50: 2
 1st Qu.:  136.5   1st Qu.: 373.5   00:03:08: 2
 Median :  562.5   Median : 639.0   00:06:19: 2
 Mean   : 2148.2   Mean   : 750.7   00:08:07: 2
 3rd Qu.: 1949.0   3rd Qu.: 947.2   00:11:33: 2
 Max.   :34372.0   Max.   :3673.0   00:15:53: 2
                                    (Other) :86
```

```
> sd(youtube_data$Viewcount)
[1] 8494648
> sd(youtube_data$Likes)
[1] 61825.47
> sd(youtube_data$Dislikes)
[1] 4746.949
> sd(youtube_data$Length_in_sec)
[1] 590.1963
```

The minimum view counts on a video is 16 while the maximum is 65694818. On an average, the videos have 3552095 views. There is a video with a single like and one has 437043 with maximum likes. The mean value for number of likes is 29989 and for number of dislikes is 2148. The standard deviation for Dislikes is 4746.949. The shortest YouTube video is of 81 seconds where as the longest You tube video is of 3673 seconds.

Text Mining and Sentiment Analysis

Following are the outputs of the top ten words generated after performing text mining on the comments of the selected 5 videos.

```
# A tibble: 327 x 2
   word        n
   <chr>   <int>
 1 cat         7
 2 girl        7
 3 water       7
 4 time        6
 5 finger      5
 6 lol         5
 7 pizza       5
 8 babi        4
 9 tornado     4
10 xd          4
# ... with 317 more rows
```

```
# A tibble: 405 x 2
   word       n
   <chr>  <int>
 1 cat       13
 2 guy        9
 3 love       9
 4 hous       8
 5 watch      8
 6 dollar     7
 7 mous       7
 8 piano      7
 9 video      7
10 eat        6
# ... with 395 more rows
```

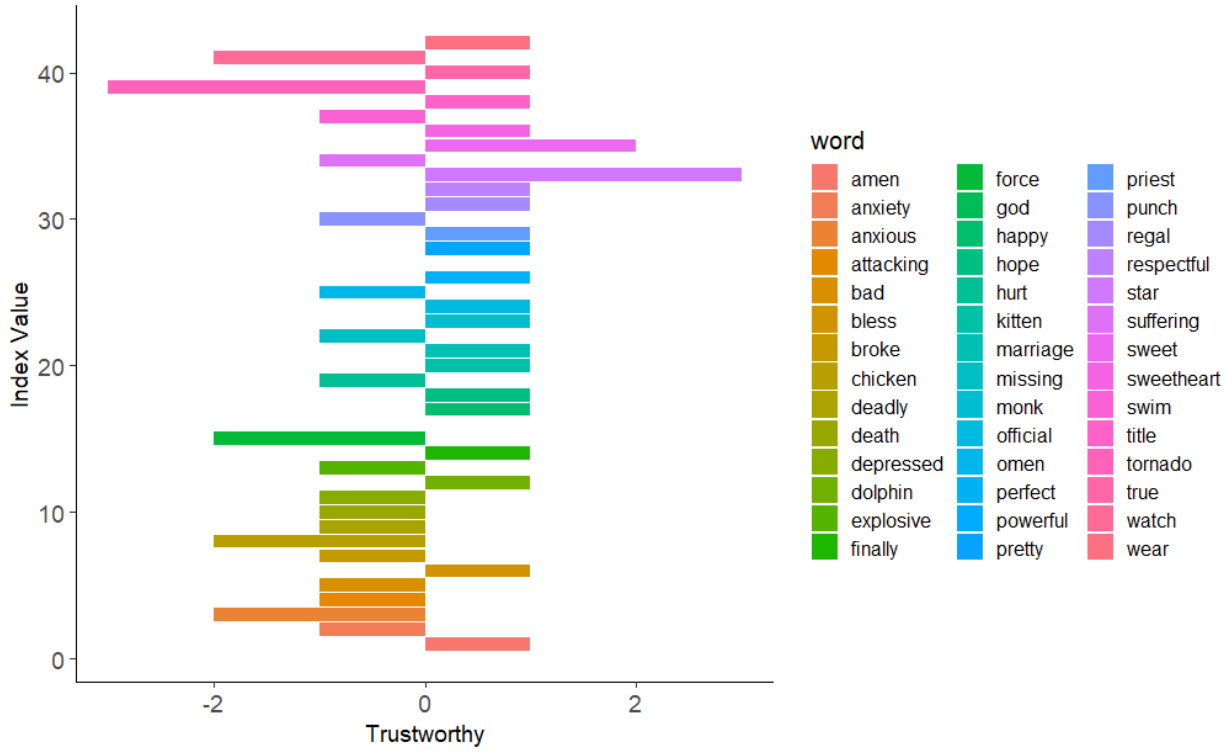
```
# A tibble: 264 x 2
   word      n
   <chr> <int>
 1 girl     25
 2 love     15
 3 heart    14
 4 adopt    10
 5 fell      8
 6 cri       7
 7 kid       7
 8 book      6
 9 talk      6
10 ador      5
# ... with 254 more rows
```

```
# A tibble: 278 x 2
   word        n
   <chr>   <int>
 1 babi       17
 2 dog        12
 3 guy        11
 4 cri        10
 5 love        8
 6 scooter     7
 7 bird        6
 8 life        6
 9 boat        5
10 girl        5
# ... with 268 more rows
```

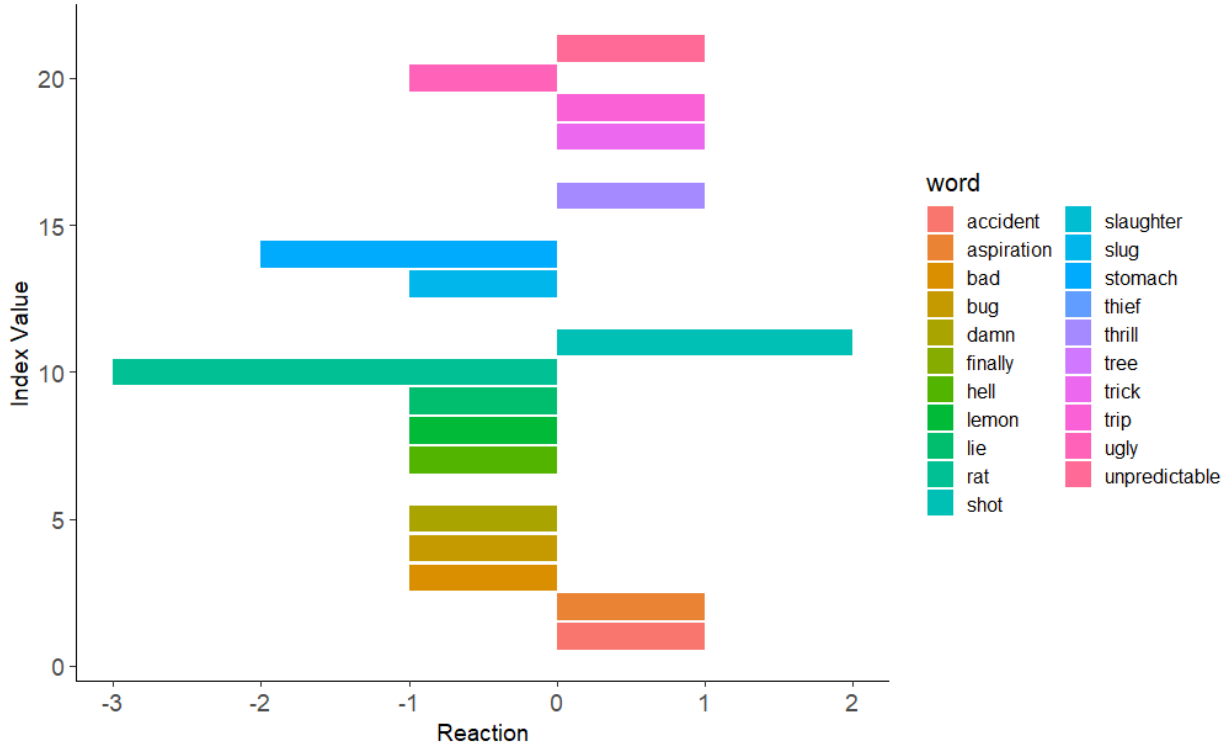
```
# A tibble: 251 x 2
   word        n
   <chr>   <int>
 1 love       30
 2 vicki      12
 3 edit       10
 4 peopl       9
 5 guy         8
 6 comment     7
 7 craft       5
 8 damn        5
 9 favorit     5
10 minut       5
# ... with 241 more rows
```

Following are the outputs for the Sentiment Analysis that was performed on the comments of the YouTube data.

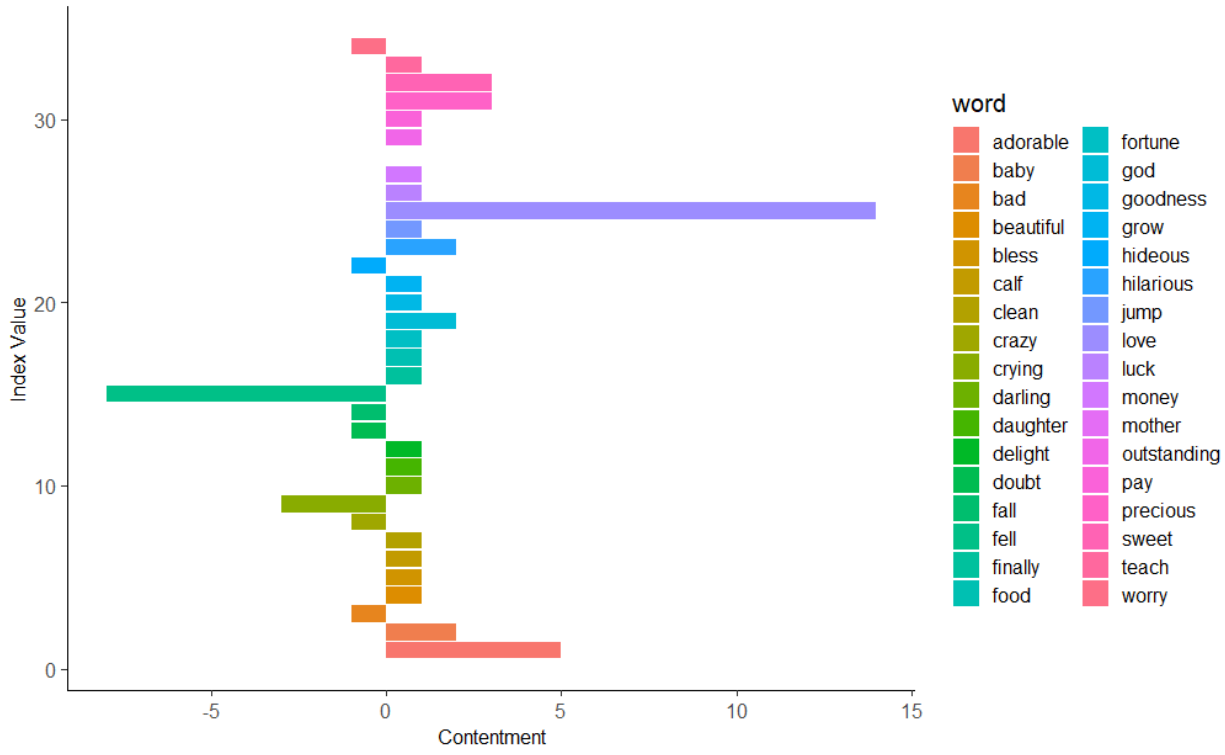
For the comments on the first video, we have taken trust and fear as the two sentiments for assessing the feeling of Trustworthiness.



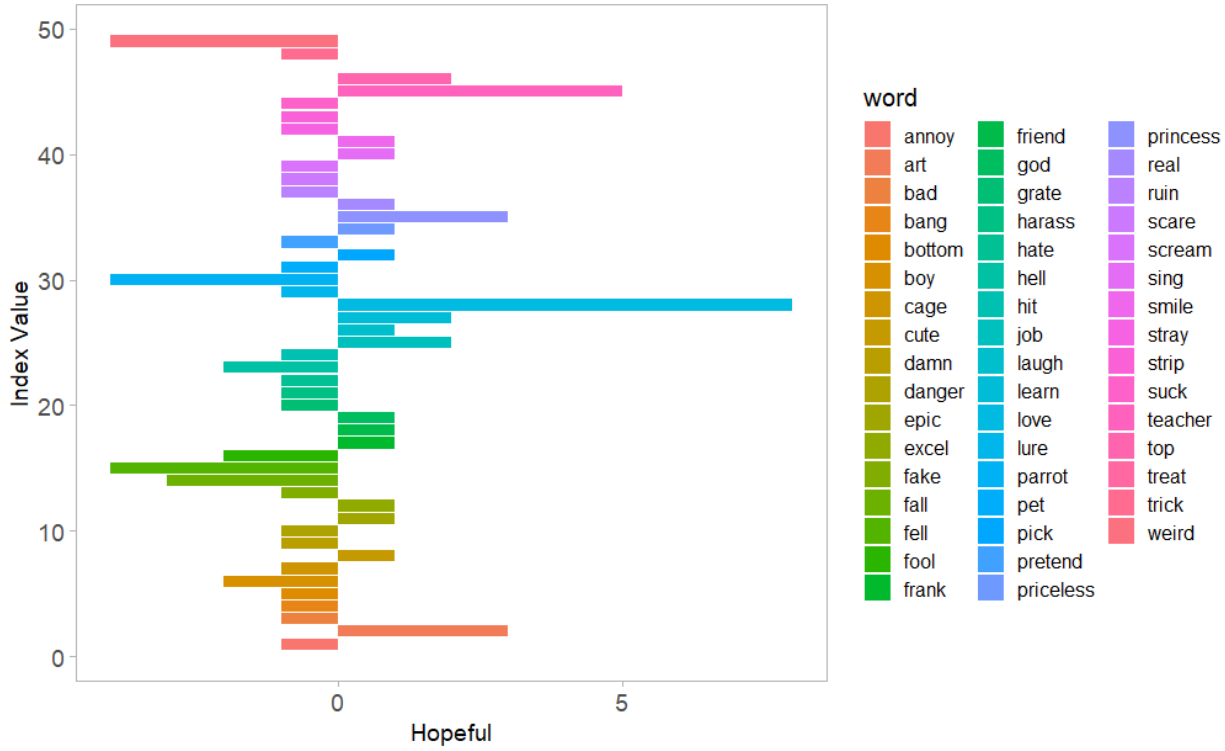
Sentiment analysis on the second video is performed using the sentiments surprise and disgust giving us the kind of reaction that the users have.



For the third video, joy and sadness are the two sentiments which we have used to assess the feeling of Contentment.



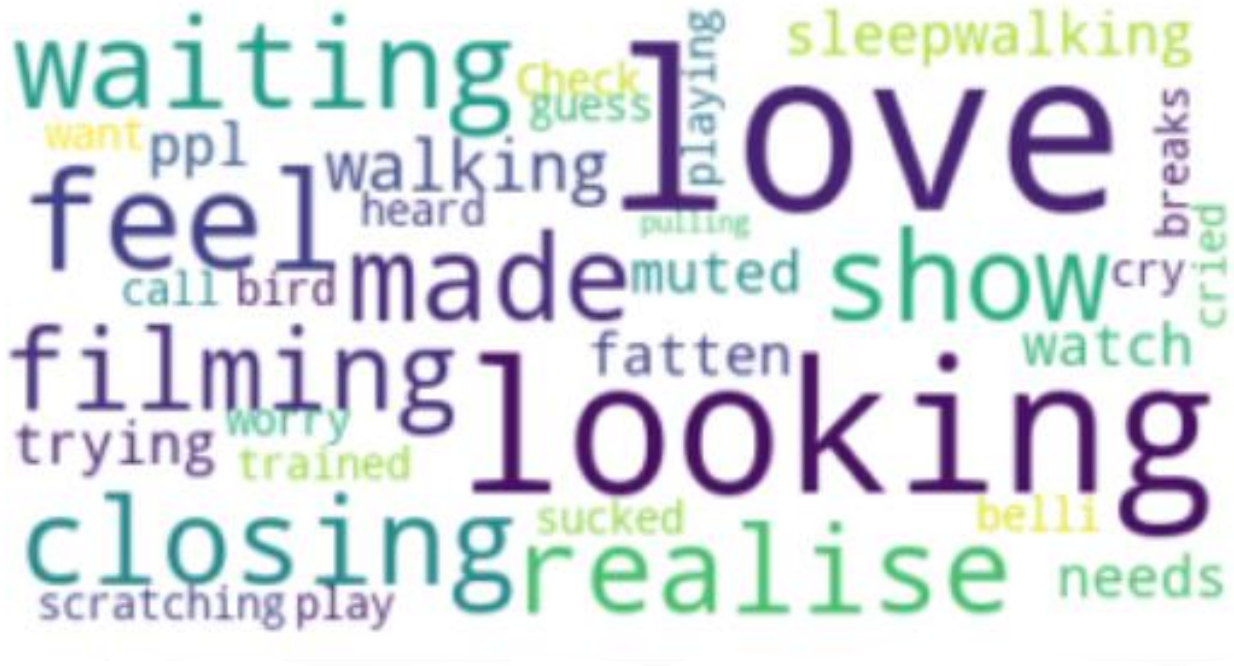
Similarly, for the comments on the fourth video, we have done sentiment analysis using positive and negative as the two sentiments to analyze the feeling whether the users are hopeful or not.



Classification Model

```
> summary(logimod1)

Call:
glm(formula = hit_video ~ Viewcount + Length_in_sec, family = "binomial",
    data = df)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.86862  -0.46061  -0.37853   0.07613   2.10508

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.587e+00  6.816e-01  -3.795 0.000148 ***
Viewcount      1.739e-06  3.900e-07   4.460  8.2e-06 ***
Length_in_sec -4.838e-05  6.581e-04  -0.074 0.941401
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 134.384  on 97  degrees of freedom
Residual deviance:  59.496  on 95  degrees of freedom
AIC: 65.496

Number of Fisher Scoring iterations: 8
```

Word Cloud for Verbs



In this word cloud, love, looking, feel, realise, closing, waiting are the entities that stand out to us.

The entity ‘love’ has occurred in the top 10 words of the 4 videos. So, it’s the word that catches the eye.

Word Cloud for Adjectives

In this word cloud, little, many, lol, uncomfortable, deadly and terrified are the entities that stand out to me.

The entity ‘lol’ has appeared in the top 10 word of the first video and we can see that it stands out in the word cloud.

Appendix

Data Consolidation

```
## Putting all the data into the
dataframe#########################################################################

linku = []
id = []
name = []
count = []
like = []
dislike = []
lent = []
dur = []
aut = []
pub = []



for i in range(long):
    url = links[i]
    v = pafy.new(url, gdata = True)
    video_id = v.videoid
    id.append(video_id)
    v_title = v.title
    name.append(v_title)
    views = v.viewcount
    count.append(views)
    likes = v.likes
    like.append(likes)
    dislikes = v.dislikes
    dislike.append(dislikes)
    v_length = v.length
    lent.append(v_length)
    v_duration = v.duration
    dur.append(v_duration)
    v_author = v.author
    aut.append(v_author)
    v_published = v.published
    pub.append(v_published)


## Creating the
DataFrame#########################################################################################

list = {'VideoId':id, 'Title':name, 'Uploader':aut, 'PublishedDateTime':pub,
'Viewcount':count, 'Likes':like, 'Dislikes': dislike, 'Length_in_sec': lent, 'Duration':
dur}
youtube_data = pd.DataFrame(list)
youtube_data.to_csv(r'C:\Users\abhis\OneDrive\Desktop\Assignments\Project\youtubedata.csv

')
```

```
# Consolidation code for comments of the first video


come = []
for y in comments0:
    come.append(y.text)
comment_data1 = pd.DataFrame(come)
comment_data1.to_csv(r'C:\Users\abhis\OneDrive\Desktop\Assignments\Project\youtubecomment
data1.csv')
```

Similarly, we have consolidated the remaining 4 videos.