YouTube comments on Oscar-nominated movie trailers

Data Mining and Business Intelligence

Group 1: Ekaterina Burkhanova, Neehar Namjoshi, Pooja Shah

Table of Contents

Introduction

Database pertains to Youtube comments on Oscar nominated movie trailers for 4 select few movies:

- **32070** rows x **8** columns
- Comments on movie trailers with date timestamp
- Movies selected are as below:
  - **Dunkirk** - History/War
  - **The Shape of Water** - Romance/Fantasy
  - **Phantom Thread** - Romance/Drama
  - **Get Out** - Psychological Thriller/Horror

Data pre-processing:

timestamp	comment_text	likes	has_replies	number_of_replies	Movies	Time	Date
1.52E+12	OSCAR	0	FALSE	0	Dunkirk	11:40:00	Monday, March 5, 2018
1.52E+12	Christopher Nolan + Hans Zimmer = perfection	0	FALSE	0	Dunkirk	11:40:00	Monday, March 5, 2018
1.52E+12	This deserves an Oscar and not shape of you.	0	FALSE	0	Dunkirk	11:40:00	Monday, March 5, 2018
1.52E+12	This movie was like watching a loud plane fly over you for 3 hours...there was r	0	FALSE	0	Dunkirk	11:40:00	Monday, March 5, 2018
1.52E+12	Get fucked Nolan, no oscars for you.	0	FALSE	0	Dunkirk	11:40:00	Monday, March 5, 2018
1.52E+12	HARRYYYYYYÃ°Å¸ËœÂÃÂ°Å¸ËœÂÃÂ°Å¸ËœÂ like if you want to see it for harry	0	FALSE	0	Dunkirk	11:40:00	Monday, March 5, 2018

In the process of cleaning the dataset:
- Combine all 4 selected movies in one dataset
- And keep all 4 selected movies separated
- Change timestamp to Time and Date
- Added one more column to name the movies

SAS Enterprise Miner Workstation process:
- Importing the file:
  - All 4 selected movies in one dataset
  - Separated 4 selected movie name in 4 dataset
- In Text Parsing, we added the table: *type of emotions* in Filter tab for separated movie dataset

Name	Role
comment_text	Text
Date	Time ID
has_replies	Rejected
likes	Rejected
Movies	Label
number_of_replies	Rejected
Time	Rejected
timestamp	Rejected

	TERMS
1	TERMS
2	accused
3	acquitted
4	adorable
5	adored
6	affected
7	afflicted
8	aghast
9	agog
10	agonized

Exploration of Data : 4 Movies together

| Topic	Category	Term Cutoff	Document Cutoff	Number of Terms	# Docs
Racism	User	0.011	0.094	154	1820
junk	User	0.01	0.147	11	0
War + History	User	0.017	0.07	245	2542
Raw excitement emotions	User	0.014	0.061	52	461
Harry Styles Fanbase	User	0.014	0.093	33	1154
Christopher Nolan Movie Reviews	User	0.015	0.084	44	1233
People's Choice of Oscar Accolades	User	0.015	0.079	52	1342
Marine + Related Movies to SOW	User	0.014	0.072	48	542
Mixed Emotions	User	0.015	0.067	62	753
Disgust + Fear	User	0.015	0.069	95	1441
Movie Soundtrack	User	0.014	0.061	28	449
Dunkirk Actors Fanbase	User	0.015	0.066	58	520

- For all 4 movies together, mixed set of reactions
- Racism is a widely talked about topic
- Nolan movies carry a huge fan base

Diving deeper into data

Exploration of Data - Dunkirk

Key topics

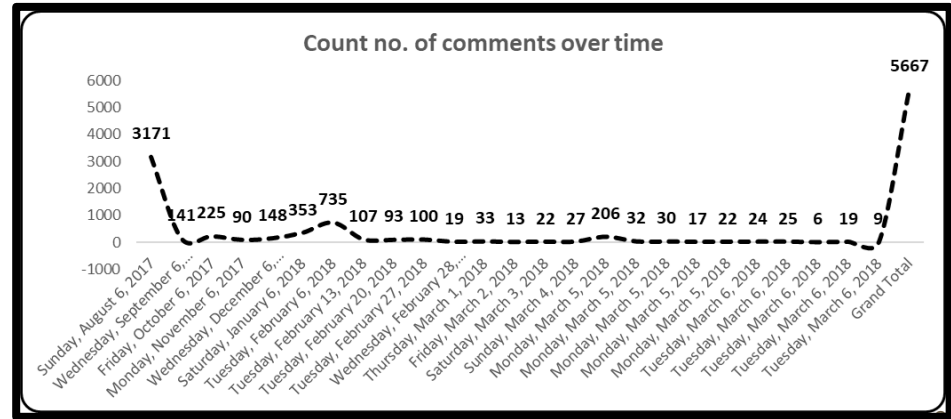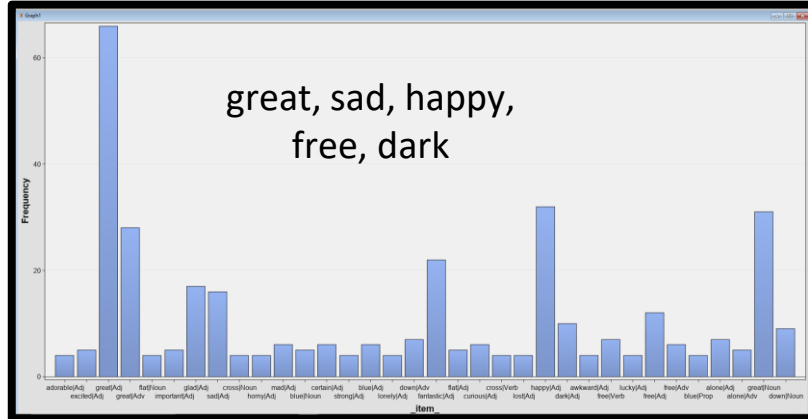| Topic	Category
hardy,+sad,+happy,+dark,+great	Multiple
+great,+happy,+dark,important,significant	Multiple
+great,+great,hardy,great,important	Multiple
hardy,hardy,+great,+dark,+great	Multiple
+proud,+great,+great,important,+lucky	Multiple
great,+great,hardy,important,+proud	Multiple
+sad,+proud,+dark,important,+great	Multiple
excited,great,+proud,+happy,hardy	Multiple
+happy,+dark,+sad,+proud,glad	Multiple
hardy,mad,+great,important,+great	Multiple

great, hardy, proud, fantastic, sad, important, glad

Count no. of comments over time
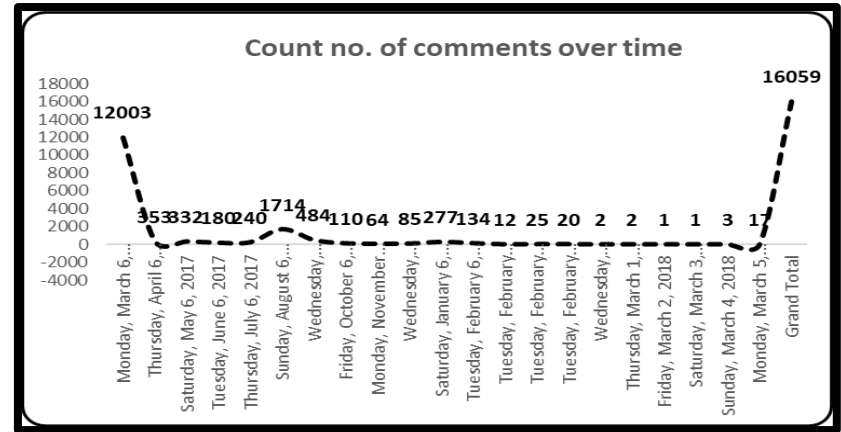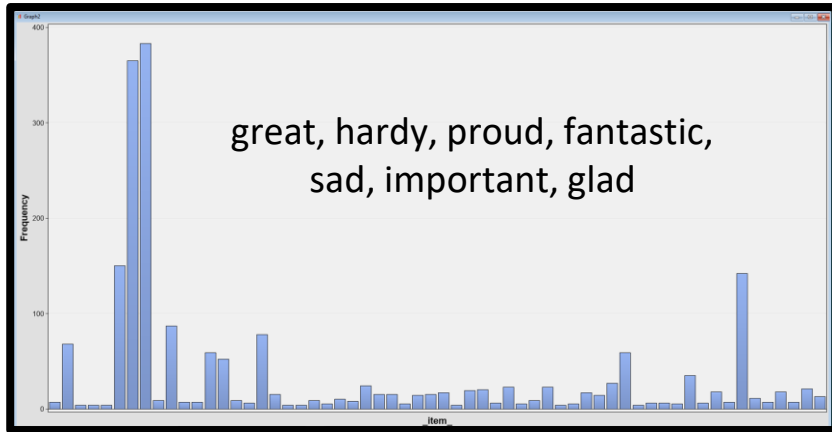
Sentiment Analysis - Dunkirk
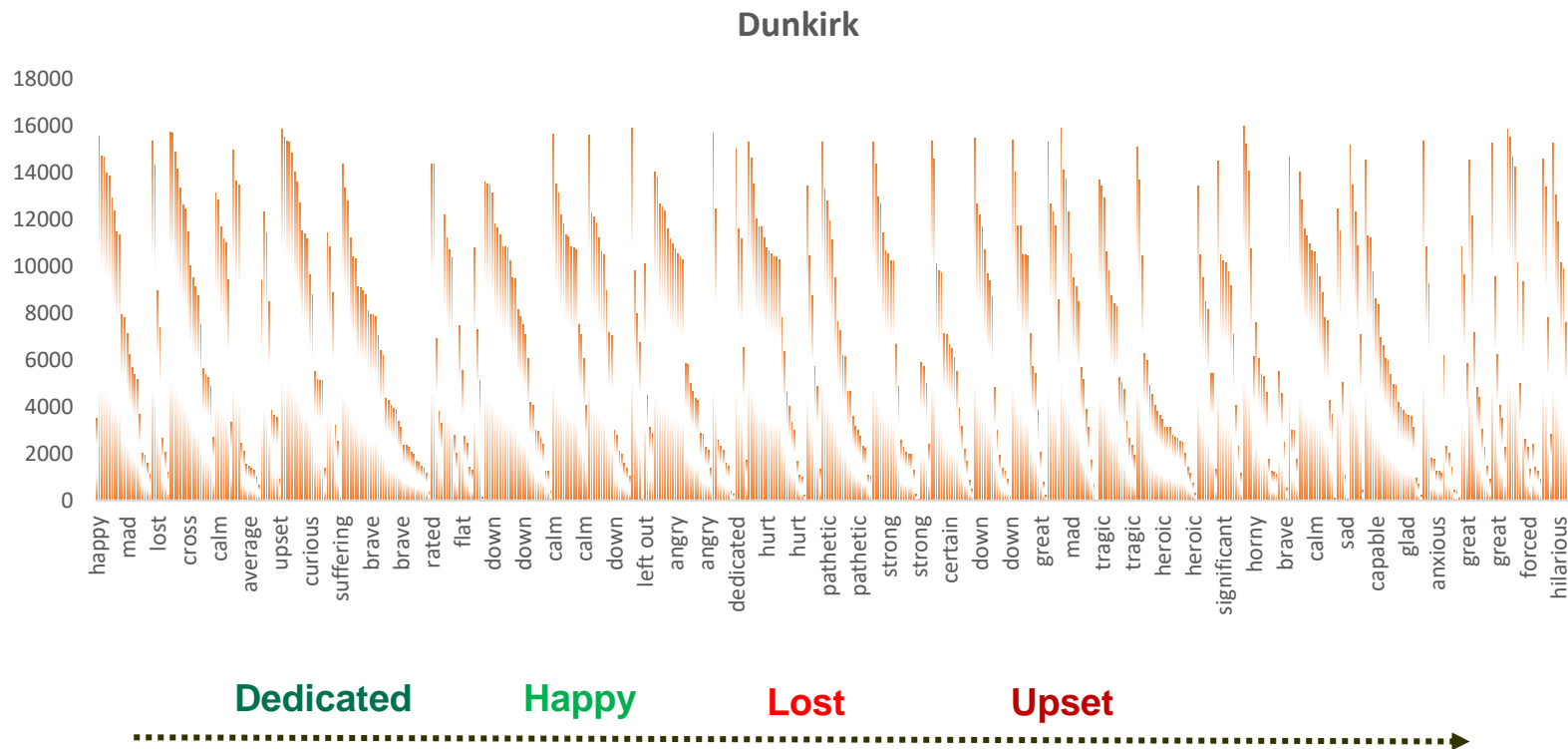
Dunkirk

Exploration of Data - Dunkirk

Exploration of Data - The Shape of Water

Key topics

Topic	Category
+great,cross,happy,+free,important	Multiple
+great,mad,great,+great,happy	Multiple
happy,sad,lucky,adorable,excited	Multiple
great,+great,happy,awkward,adorable	Multiple
fantastic,horny,+great,great,adorable	Multiple
glad,happy,great,adorable,+great	Multiple

great, sad, happy, free, dark

Count no. of comments over time

Sentiment Analysis - The Shape of Water
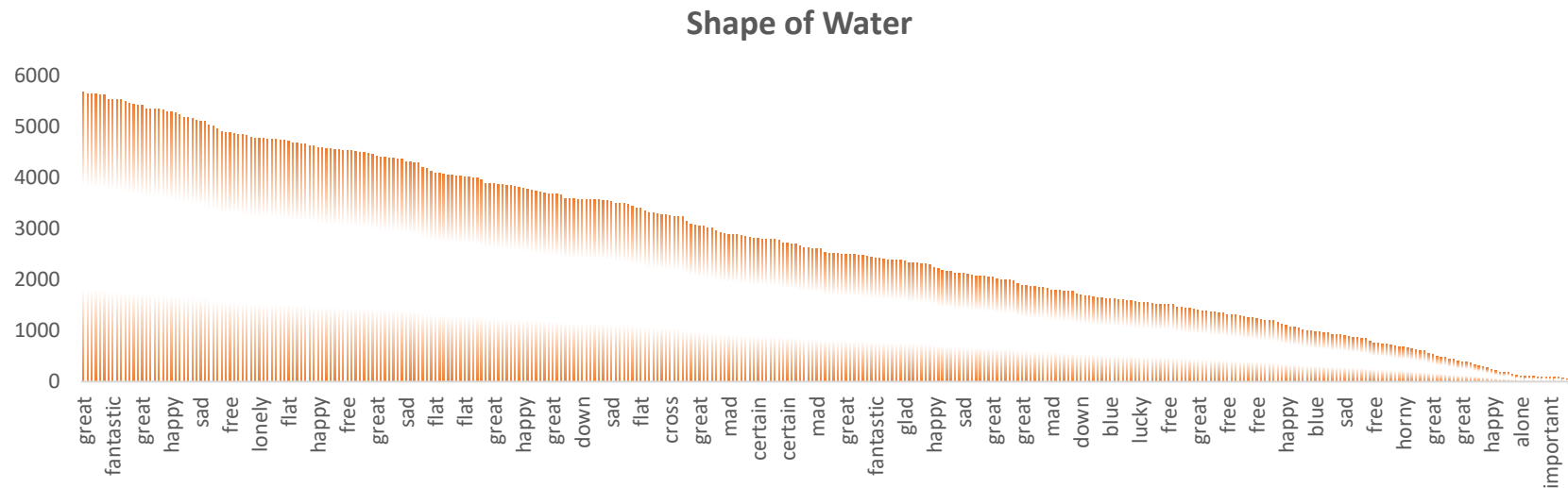
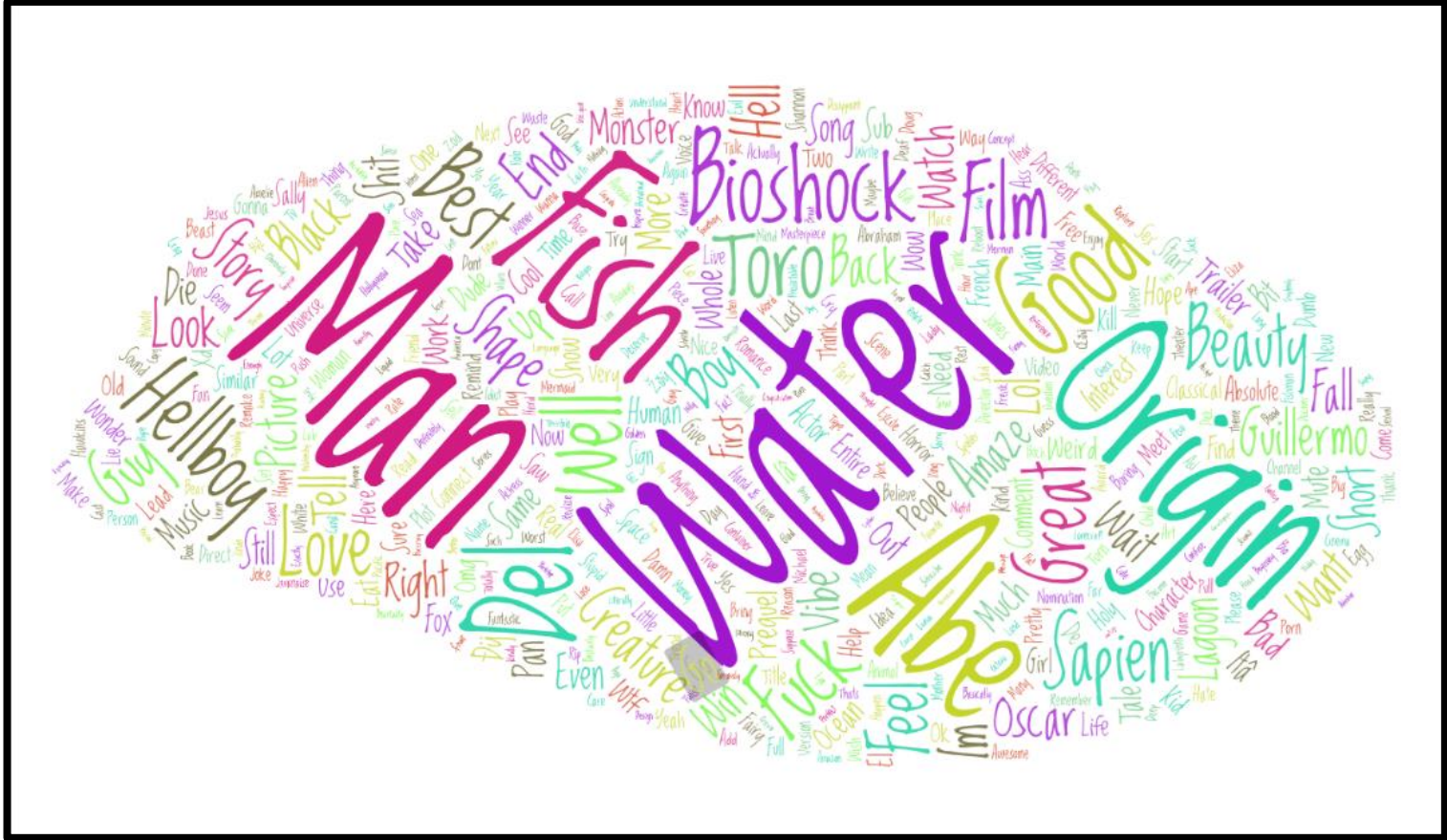Shape of Water

Great   **Fantastic**   **Happy**

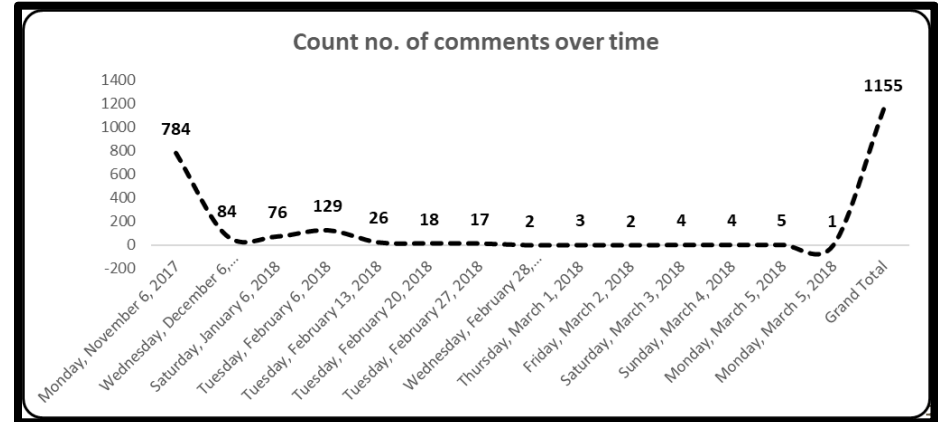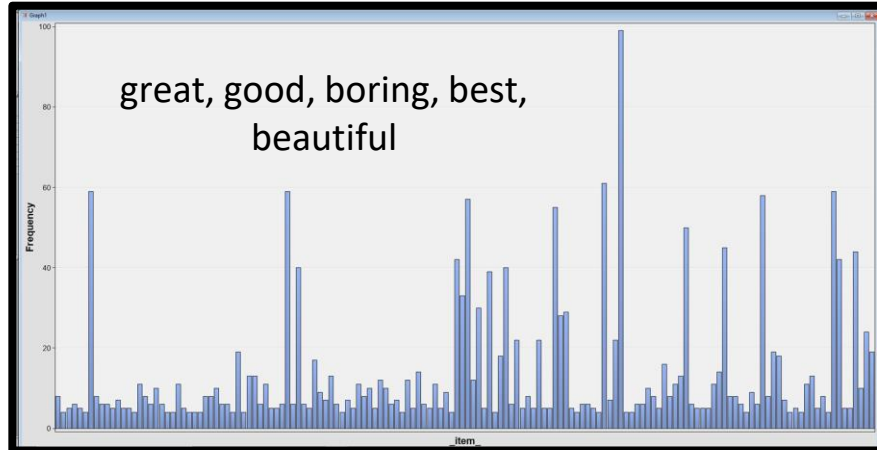# Exploration of Data - The Shape of Water

Exploration of Data - Phantom Thread

Key topics

Topic	Category
not,+do,+wait,+film,+have	Multiple
daniel,+day,lewis,daniel day,+retire	Multiple
anderson,paul,thomas,+be,+make	Multiple
+oscar,+go,+get,4th,bait	Multiple
ðÿ,â,thread,+movie,phantom	Multiple
+be,+movie,+think,blood,i	Multiple
+look,boring,+good,+movie,+be	Multiple
+movie,+see,i,+make,+want	Multiple
pta,ddl,+film,+see,+year	Multiple
+good,+actor,day-lewis,+film,+year	Multiple

great, good, boring, best, beautiful

Count no. of comments over time

784   84   76   129   26   18   17   2   3   2   4   4   5   1   1155

# Sentiment Analysis - Phantom Thread
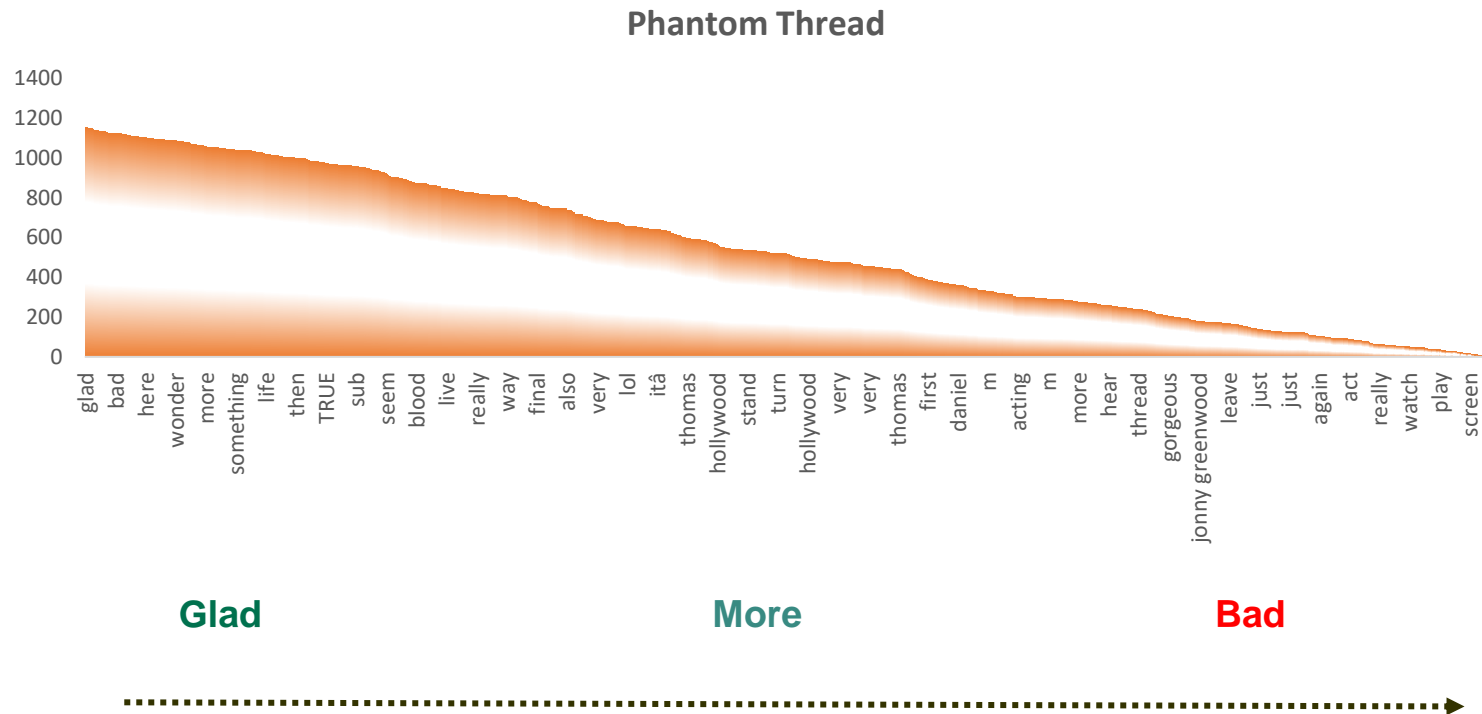


Phantom Thread

Exploration of Data - Phantom Thread

Exploration of Data - Get Out

Key topics

Topic	Category
+great,glad,important,alone,uncomfortable	Multiple
glad,great,mad,nervous,fantastic	Multiple
hilarious,glad,uncomfortable,curious,afraid	Multiple
great,+great,common,+sad,angry	Multiple
mad,hilarious,angry,+down,+upset	Multiple
nervous,uncomfortable,angry,afraid,+upset	Multiple
+happy,+sad,nervous,+hurt,excited	Multiple
great,alone,free,+great,+sad	Multiple
+sad,guilt,+down,+dark,+upset	Multiple
down,calm,+down,calm,+dark	Multiple

great, hilarious, glad, sad, mad

Count no. of comments over time
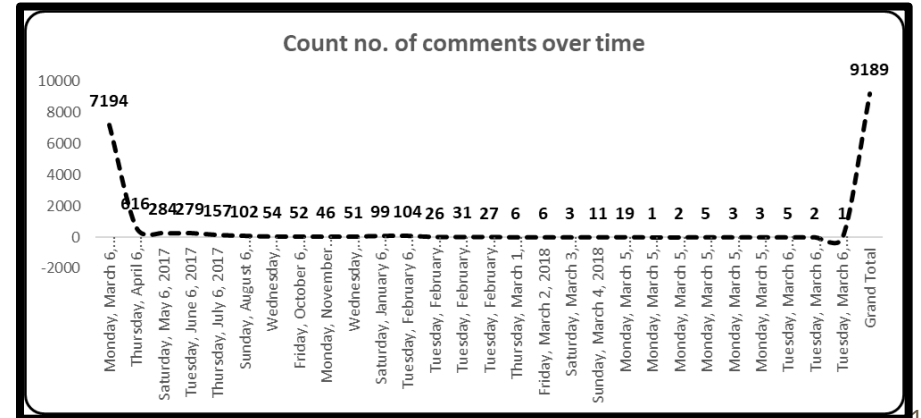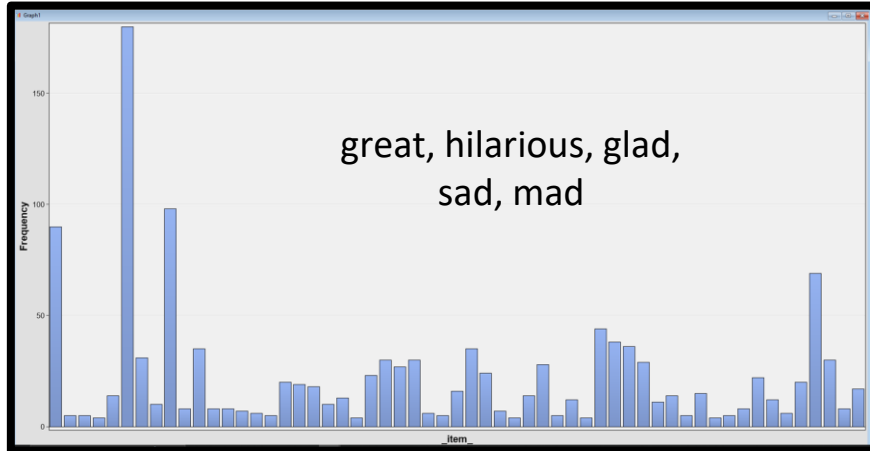
Sentiment Analysis – Get Out

Get Out

Loving   Uncomfortable   Dark
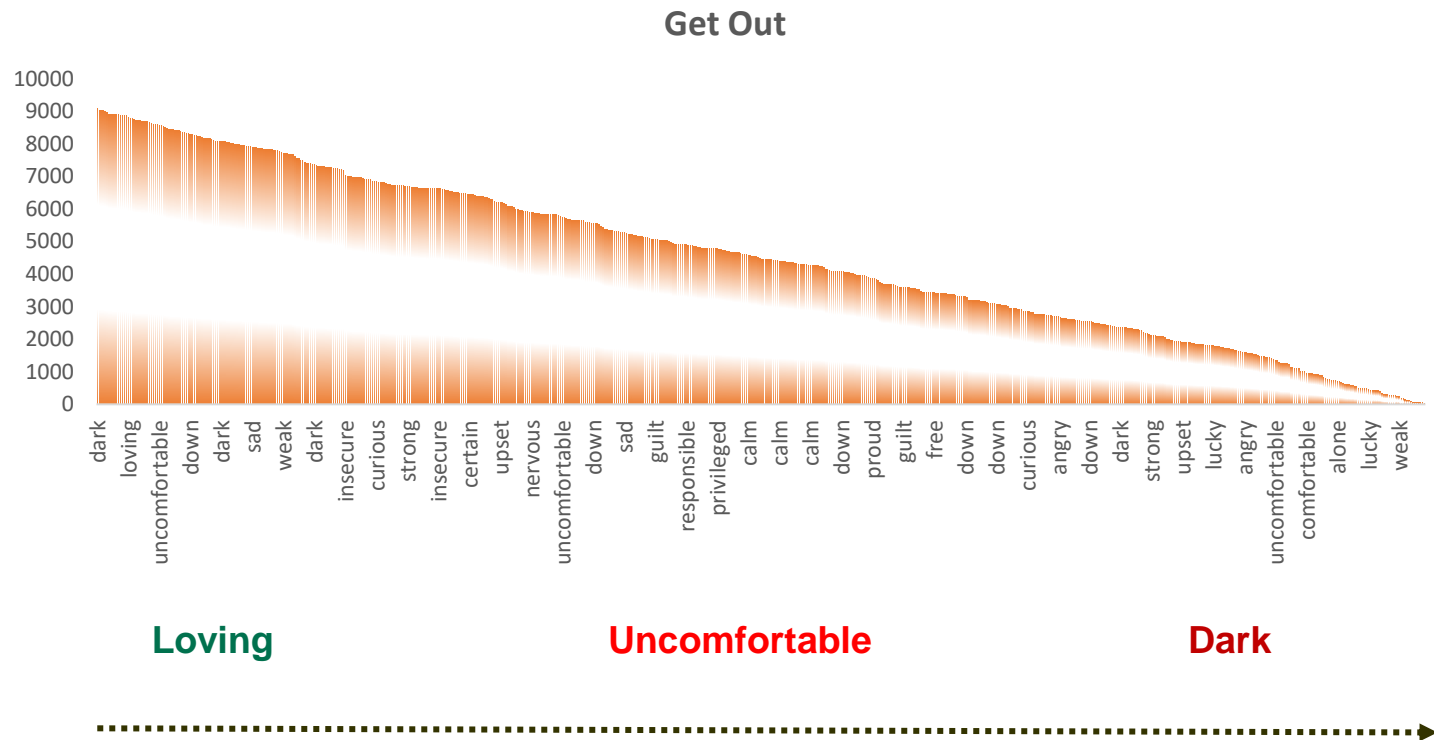
Exploration of Data - Get Out

Key trends and insights

Key trends and insights

People gravitate towards war and action movies

People tend to reflect negatively on topics which are sensitive and attracts a lot of flak amongst viewers if not portrayed well

Actors and Actresses can overpower the fame a movie can get and influence viewership driven by crazy fan-base to a very large extend

Business Recommendations

01 **War** and **history** are the topics of interest to the reviewers. Youtube and other media platforms can look at creating content along those lines.

02 A huge chunk of revenue generation depends on the actor / actress casted. **Highly acclaimed actors** earn higher viewership and hence increase revenue generation.

03 The degree of viewership tends to increase as there are **events** like nomination announcements, trailer launches, etc. Doing events around those time frames can **help drive revenue generation** due to higher CTR and viewerships.

21