HORROR MOVIE CRITIC RATINGS BASED ON EMOTIONS

TEAM JUST RAN A REGRESSION

AGENDA

1. Marketing Problem & Target Audience
2. Data Acquisition & Preliminary Analysis
3. Data Preparation
4. Modeling & Analysis

MEET OUR TEAM!


Hannah


Kristina

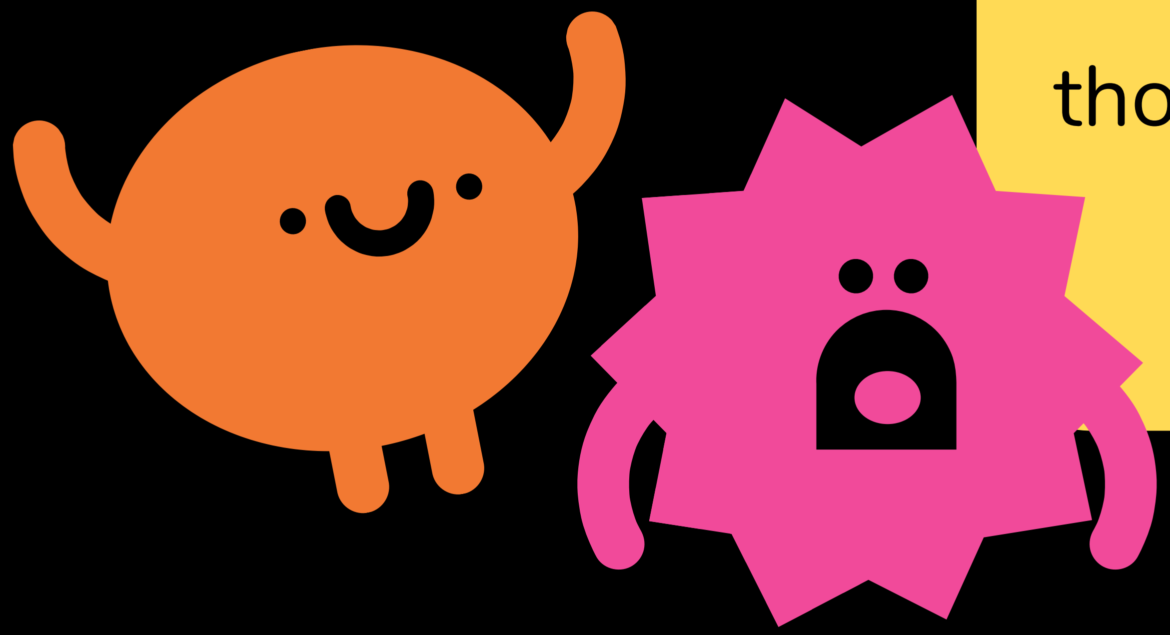

Mandy


Sally

MOTIVATION

Do movie critic reviews matter? Film criticism shapes the movie industry and how people view films. By highlighting the good and bad parts of a film, directors, writers, & actors can further better their skill sand learn from their mistakes.

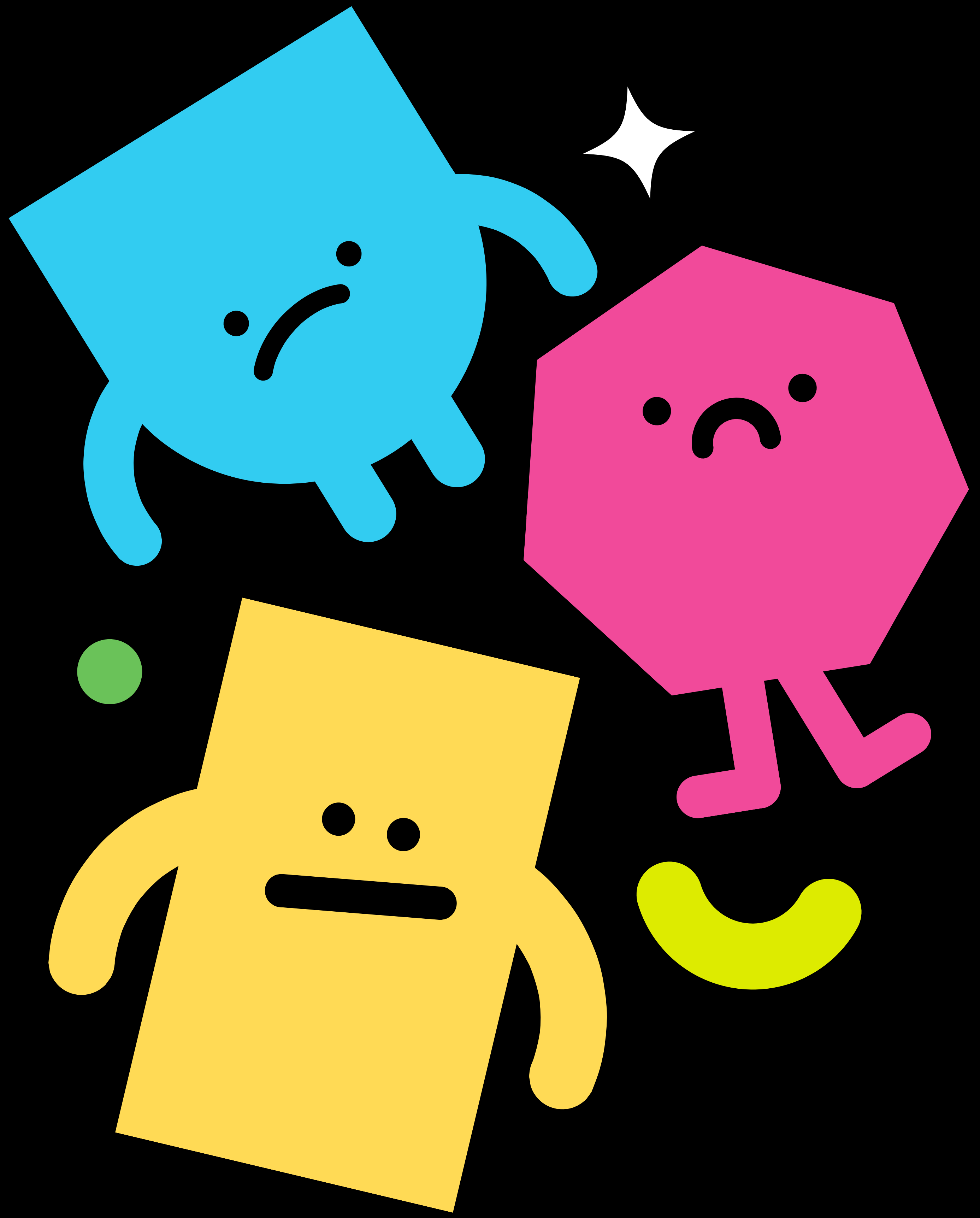
Additionally, movie reviews have a significant effect on a consumer's thought process and a predictor on how well the film performs financially.



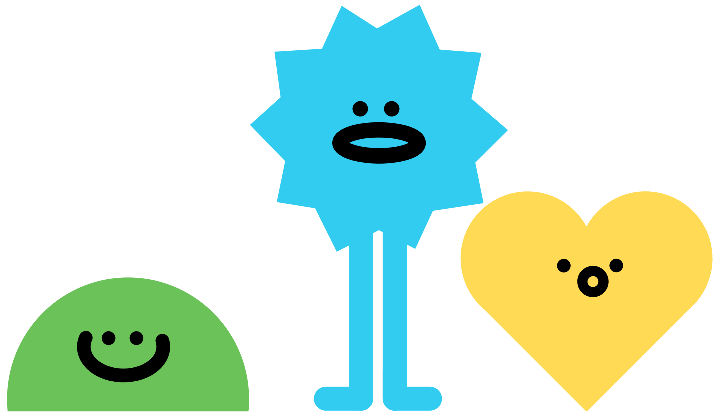
MARKETING QUESTION

Movie critic reviews are an essential marketing tool for the film industry. They unlock the door to a consumer's mind on whether a movie is worth the price of admission because they are viewed as film connoisseurs.

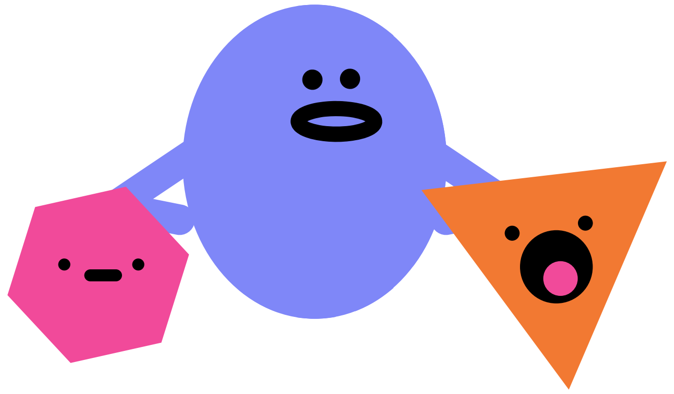
With that being said, our team had a question: **Can horror movies elicit certain feelings in movie critics to affect their film review and rate it higher on Rotten Tomatoes?**



TARGET AUDIENCE



- Film Production Companies
- Head of Production Teams - oversees the making of the film
- Involved in scripting, casting, corporate promotions, commercials, and marketing the films

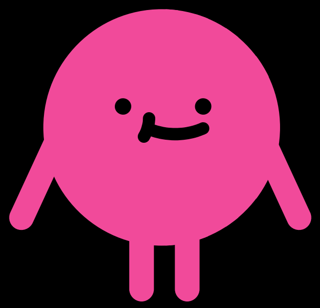


Blumhouse Productions & Platinum Dunes

Two notable production companies for producing horror & thriller films such as the Conjuring & The Purge series

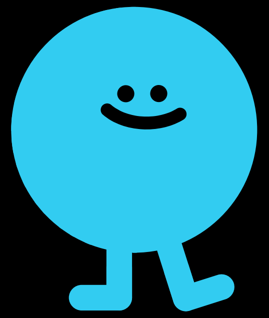
DATA ACQUISITION & PRELIMINARY ANALYSIS

2 Datasets - kaggle.com



Emotions Sensor Data Set

- Top 1100 English words classified statistically into 7 basic emotions: Disgust, Surprise, Neutral, Anger, Sad, Happy and Fear.
- The dataset contains 8 columns, and 1104 unique words.



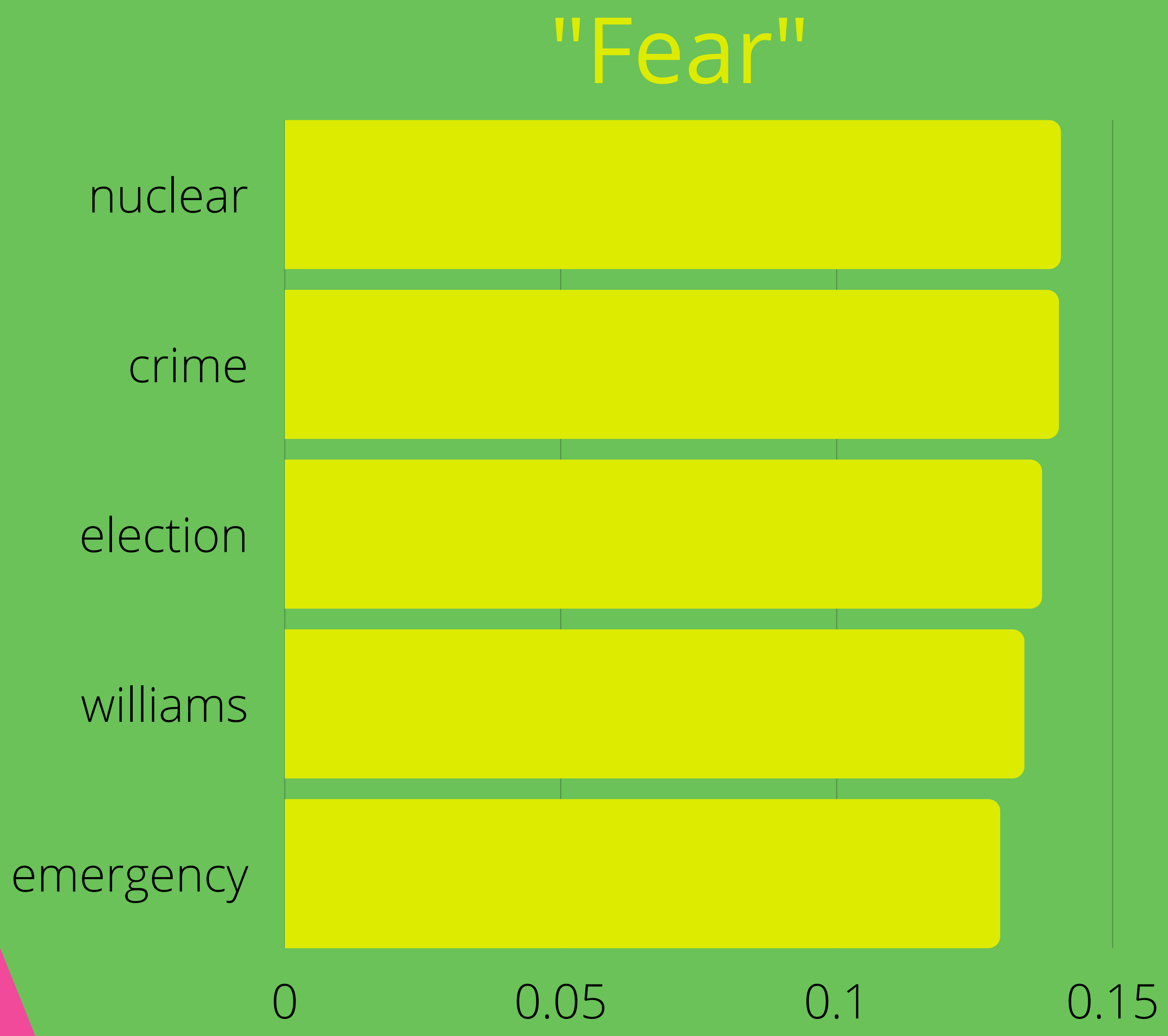
Rotten Tomatoes Movies and Critic Reviews Dataset

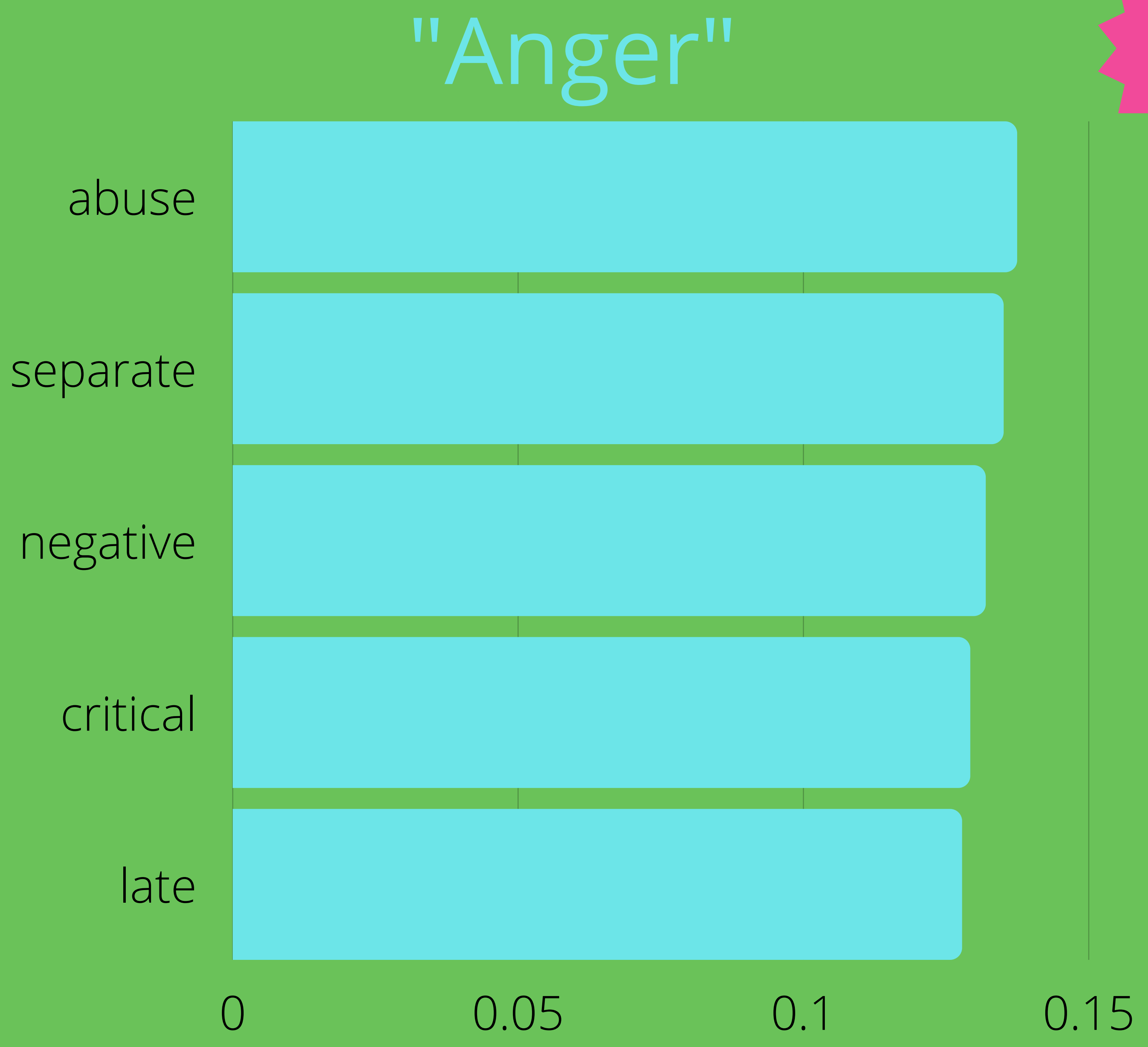
- **Movies dataset:** movie title, description, genres, duration, director, actors, users' ratings, and critics' ratings, etc
- 8 columns, 17712 unique values

word	disgust	surprise	neutral	anger	sad	happy	fear
ability	0.00446429	0.04783164	6.38E-04	0.02359694	0.01339286	0.01594388	0.04017858
able	1.73E-05	1.82E-04	4.09E-04	1.76E-04	2.19E-04	2.44E-04	1.86E-04
abuse	5.32E-04	1.77E-04	1.77E-04	0.13736264	0.00124069	0.00159518	0.00265863
academy	0.00714286	0.02142857	0.00714286	0.00714286	0.00714286	0.09285715	0.03571429
accept	0.00827068	0.00676692	7.52E-04	0.04887218	0.01879699	0.02481203	0.03834587
acceptance	0.00274725	0.00824176	0.00274725	0.01373627	0.02472528	0.09065934	0.01373627
accounting	0.01785714	0.01785714	0.01785714	0.01785714	0.05357143	0.08928572	0.01785714
accuracy	0.03571429	0.10714287	0.03571429	0.03571429	0.03571429	0.03571429	0.03571429
achieve	0.00138249	0.00414747	4.61E-04	0.00506912	0.00691244	0.12211982	0.00506912
acid	0.01785714	0.01785714	0.01785714	0.01785714	0.125	0.01785714	0.01785714
active	0.00238095	0.03571429	0.00238095	0.02619048	0.05	0.02619048	0.01190476
activities	3.15E-06	3.46E-05	0.0012618	4.72E-05	5.98E-05	2.20E-05	1.57E-05
adding	0.00324675	0.0487013	0.00324675	0.02272728	0.01623377	0.02922078	0.03571429
addresses	0.03571429	0.10714287	0.03571429	0.03571429	0.03571429	0.03571429	0.03571429
administrato	0.03571429	0.03571429	0.03571429	0.03571429	0.03571429	0.03571429	0.10714287
adobe	0.03571429	0.03571429	0.03571429	0.03571429	0.10714287	0.03571429	0.03571429
adult	0.0018797	0.05075188	0.0018797	0.02819549	0.02067669	0.01315789	0.03571429
adults	0.01428572	0.04285715	0.00476191	0.03333334	0.01428572	0.02380953	0.03333334
advanced	1.73E-05	3.45E-06	0.00138371	3.45E-06	1.73E-05	1.04E-05	[illegible]

Emotions Sensor Data Set

TOP 5 WORDS BY EMOTION





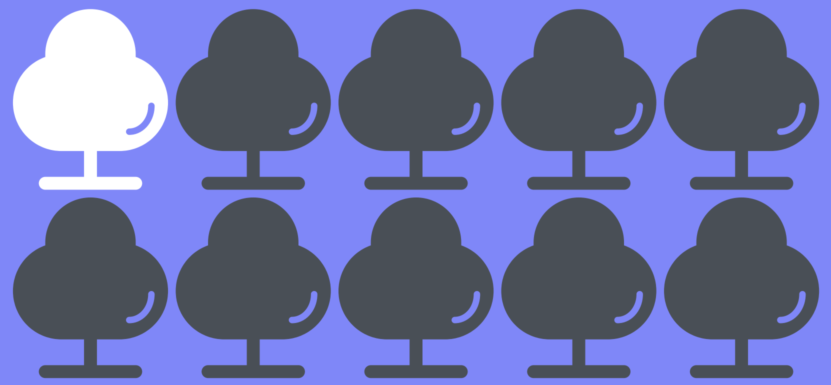
Movies Dataset

rotten_toma	movie_title	movie_info	critics_conse	content_rati	genres	directors	authors	actors	original_rele	streaming_r	runtime	production_	tomatomete	tomatomete	tomatomete	audience_st	audience_ra	audience_cc	tomatomete	tomatomete	tomatometer_rotten_critics_count
m/0814255	Percy Jacksc	Always troul	Though it m	PG	Action & Ad	Chris Columl	Craig Titley,	Logan Lerma	2/12/10	11/25/15	119	20th Century	Rotten	49	149	Spilled	53	254421	43	73	76
m/0878835	Please Give	Kate (Cather	Nicole Holof	R	Comedy	Nicole Holof	Nicole Holof	Catherine Ke	4/30/10	9/4/12	90	Sony Picture	Certified-Fre	87	142	Upright	64	11574	44	123	19
m/10	10	A successful	Blake Edwar	R	Comedy, Ro	Blake Edwar	Blake Edwar	Dudley Moc	10/5/79	7/24/14	122	Waner Bros.	Fresh	67	24	Spilled	53	14684	2	16	8
m/1000013	12 Angry M	Following th	Sidney Lume	NR	Classics, Dra	Sidney Lume	Reginald Ros	Martin Balsa	4/13/57	1/13/17	95	Criterion Coll	Certified-Fre	100	54	Upright	97	105386	6	54	0
m/1000079	20,000 Leag	In 1866, Pro	One of Disne	G	Action & Ad	Richard Fleis	Earl Felton	James Masc	1/1/54	6/10/16	127	Disney	Fresh	89	27	Upright	74	68918	5	24	3
m/10000_b	10,000 B.C.	Mammoth h	With attenti	PG-13	Action & Ad	Roland Emm	Harald Klose	Steven Strai	3/7/08	6/22/13	109	Warner Bros	Rotten	8	149	Spilled	37	411140	37	12	137
m/1000121	The 39 Step	While on va	Packed with	NR	Action & Ad	Alfred Hitchc	Alma Reville	Robert Dona	8/1/35	1/12/17	80	Gaumont Bri	Certified-Fre	96	51	Upright	86	23890	8	49	2
m/1000123	3:10 to Yum	Dan Evans (Van Heflin), a		NR	Classics, Dra	Delmer Dav	Halsted Wel	Glenn Ford, '	8/7/57	4/16/12	92	Columbia Pic	Fresh	96	28	Upright	79	9243	6	27	1
m/1000200	Charly (A He	Cultural differences, past		PG	Comedy, Dra	Adam Thom	Jack Weylar	Heather Bee	9/27/02	5/22/17	103	Excel Enterta	Rotten	20	10	Upright	87	4819	0	2	8
m/1000204	Abraham Lir	The 16th U.S. president ('		NR	Classics, Dra	D.W. Griffith	Gerrit J. Lloy	Walter Hust	11/8/30	12/3/13	97	United Artist	Fresh	82	11	Spilled	40	457	4	9	2
m/1000211	Dark Water	In this moody Japanese h		PG-13	Art House &	Hideo Nakat	Hideo Hakat	Hitomi Kurol	1/19/02	3/23/17	100	Toho Compa	Fresh	80	15	Upright	66	21475	3	12	3
m/1000224	The Accusec	Out drinking one night af		R	Drama, Mys	Jonathan Ka	Tom Topor	Jodie Foster,	10/14/88	10/4/16	110	Paramount I	Fresh	91	22	Upright	79	20821	1	20	2
m/1000251	The Lost City	Fico Fellove	Its heart is ir	R	Drama	Andy Garcia	G. Cabrera I	Andy Garcia	9/3/05	3/23/17	143	Magnolia Pi	Rotten	25	83	Upright	64	25944	37	21	62
m/1000251	The Breakin	A charter-boat captain wi		NR	Drama	Michael Cur	Ranald Mac	John Garfielc	10/6/50	8/25/16	97	Warner Hon	Fresh	100	10	Upright	86	335	1	10	0
m/1000253	Adam's Rib	A courtroom	Matched by	NR	Classics, Con	George Cuk	Garson Kani	Spencer Tra	11/18/49	5/1/08	101	MGM Home	Fresh	96	28	Upright	86	10563	6	27	1
m/1000263	The Bridge c	During the S	Despite an a	PG	Art House &	Mary McGu	Mary McGu	Gabriel Byrn	6/10/05	7/24/14	124	Fine Line Fe	Rotten	4	24	Spilled	35	1935	13	1	23
m/1000267	The Prowler	After being frightened by		PG	Drama, Mys	Joseph Lose	Robert Thoe	Van Heflin, E	12/3/51	12/5/16	92	VCI Entertai	Fresh	100	18	Upright	86	286	6	18	0
m/1000327	Criminal	Needing a n	If you saw N	R	Drama, Mys	Greggory Ja	Gregory Jac	John C. Reilly	9/24/04	4/27/16	87	Warner Bros	Fresh	69	124	Spilled	57	6711	35	85	39
m/1000334	The Adventu	In this claymation film, ce		G	Action & Ad	Will Vinton	Susan Shadl	James Whitr	3/1/85	2/4/17	86	Eureka Ente	Fresh	80	5	Upright	82	1046	1	4	1
m/1000343	Deep Blue	This nature	Full of visual	G	Documentar	Andy Byatt,	Alastair Fotl	Pierce Brosn	6/17/05	12/7/16	90	Miramax Fli	Fresh	67	52	Upright	80	4692	18	35	17
m/1000355	The Adventu	When King F	Errol Flynn t	PG	Action & Ad	Michael Cur	Norman Reil	Errol Flynn, (	5/14/38	1/23/14	102	Warner Bros	Certified-Fre	100	48	Upright	89	33946	8	48	0

- original_release_date
- streaming_release_date
- runtime
- production_company
- tomatometer_status
- **tomatometer_rating**
- tomatometer_count
- audience_status

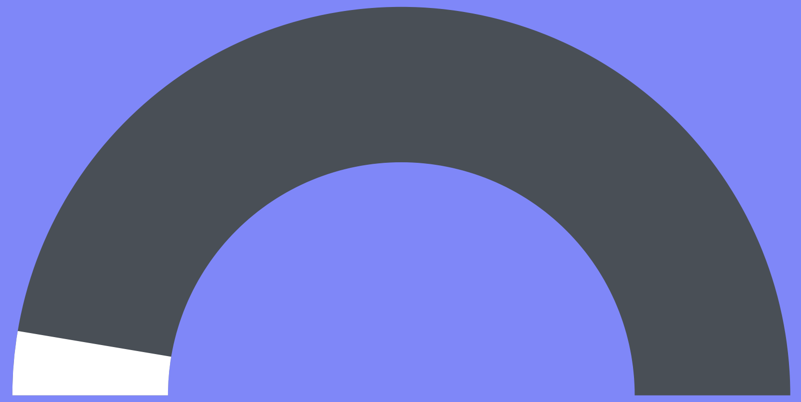
- rotten_tomatoes_link
- **movie_title**
- movie_info
- **critics_consensus**
- content_rating
- **genres**
- directors

- audience_rating
- audience_count
- tomatometer_top_critics_count
- tomatometer_fresh_critics_count
- tomatometer_rotten_critics_count



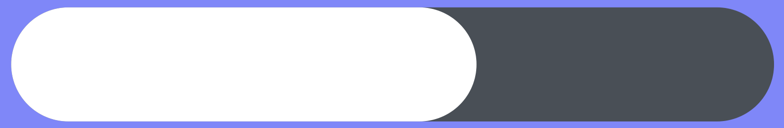
0.54 out of 10

Horror genre



954

Horror movies



60.88

Average rating for Horror genre

horror
enough
scare
thrill
genre
make
movie
gore
performance
may
offer
film
story
original
director
franchise
sequel
plot
fan
action
Dead
campy
strong
horror fan
without
genuine
surprising
keep
made
cliche
theme
well
blood
whose
game
standard
narrative
work
predecessor
ambitious
laughs
suffers
fresh
best
familiar
yet
far
find
prove
worth
dull
add
uneven
Stephen King
psychological
leave
title
script
bad
full
Though
visually
monster
come
wit
quite
truly
Eye
ultimately
visual
one
deliver
writer director
chill
audience
derivative
better
memorable
thanks
tropes
teen
series
fail
horror film
low budget
stylish
tension
direction
predictable
smart
will
acting
funny
feature
unsettling
chilling
sense
gory
little
remain
atmospheric
special effects
mark
twist
lacking
disturbing
creature feature
entertaining
back
frightening
cast
satisfying
give
manage
fun
lack
modern
Despite
dialogue
first
featuring
never
elements
rather
set
classic
premise
humor
zombie
time
violence
effectively
style
new
feel
vampire
installment
return
less
effect
sci fi
Halloween
adaptation
supernatural
good
although
part
often
muddled
turn
live
Dark
impressive
take
bring
use
Man
high
still
ingredient
Ghost
making
slasher
well acted
satisfy
many
cheap
way
star
surprisingly
another
atmosphere
bloody
compelling
silly
even
scary
terrifying
Night
entry
moments
create
creepy
remake
effective
idea
character
suspense
viewer
blend
comedy
much
House
nasty
effort
clever
horror movie
horror comedy

HYPOTHESIS

Movie reviews with words that are more likely to evoke the emotion "fear" will cause higher ratings for the movies

DATA PREPARATION

TASK:

Give each movie's critic consensus a "score" for each of the seven emotions based on the words used in the review

creating a list of reviews

```
[15] reviews = []
     for r in tomatoes["critics_consensus"]:
         reviews.append(r)
```

creating a list of popular words and the emotional scores

```
popular_words = []
for w in words["word"]:
  w = w.strip()
  popular_words.append(w)

disgust_scores = []
for d in words["disgust"]:
  disgust_scores.append(d)

surprise_scores = []
for s in words["surprise"]:
  surprise_scores.append(s)
```

Step 1: Turn everything into lists

Step 2: Create dictionaries for each of the emotions

```
disgust_dict = {}
index0 = 0
for w in popular_words:
    disgust_dict[w] = disgust_scores[index0]
    index0 += 1
```

Step 3: Create a loop for each review

```python
#a loop that splits review into words, then checks to see if each word in the review matches a popular word
for x in reviews:

  #creating empty lists for each emotion
  disgust_list = []
  surprise_list = []
  neutral_list = []
  anger_list = []
  sad_list = []
  happy_list = []
  fear_list = []

  #splitting review into invidial lists
  review_words = x.split(' ')
  for y in review_words:

      #if it does, adds the score for that word to their list
      if y in popular_words:
        disgust_list.append(disgust_dict[y])
        surprise_list.append(surprise_dict[y])
        neutral_list.append(neutral_dict[y])
        anger_list.append(anger_dict[y])
        sad_list.append(sad_dict[y])
        happy_list.append(happy_dict[y])
        fear_list.append(fear_dict[y])

  #averages the scores for each emotion
  if len(disgust_list) == 0:
    avg_disgust = 0
  else:
    avg_disgust = avg_fun(disgust_list)
```

```python
#a loop that splits review into words, then checks to see if each word in the review matches a popular word
for x in reviews:

  #creating empty lists for each emotion
  disgust_list = []
  surprise_list = []
  neutral_list = []
  anger_list = []
  sad_list = []
  happy_list = []
  fear_list = []

  #splitting review into invidial lists
  review_words = x.split(' ')
  for y in review_words:

      #if it does, adds the score for that word to their list
      if y in popular_words:
        disgust_list.append(disgust_dict[y])
        surprise_list.append(surprise_dict[y])
        neutral_list.append(neutral_dict[y])
        anger_list.append(anger_dict[y])
        sad_list.append(sad_dict[y])
        happy_list.append(happy_dict[y])
        fear_list.append(fear_dict[y])

  #averages the scores for each emotion
  if len(disgust_list) == 0:
    avg_disgust = 0
  else:
    avg_disgust = avg_fun(disgust_list)
```

Step 4: Split up review into individual words, then check to see if that word is one of the popular words

Step 5: If it is, adds the word's score for each emotion to empty lists

```
#a loop that splits review into words, then checks to see if each word in the review matches a popular word
for x in reviews:

  #creating empty lists for each emotion
  disgust_list = []
  surprise_list = []
  neutral_list = []
  anger_list = []
  sad_list = []
  happy_list = []
  fear_list = []

  #splitting review into invidial lists
  review_words = x.split(' ')
  for y in review_words:

    #if it does, adds the score for that word to their list
    if y in popular_words:
      disgust_list.append(disgust_dict[y])
      surprise_list.append(surprise_dict[y])
      neutral_list.append(neutral_dict[y])
      anger_list.append(anger_dict[y])
      sad_list.append(sad_dict[y])
      happy_list.append(happy_dict[y])
      fear_list.append(fear_dict[y])

  #averages the scores for each emotion
  if len(disgust_list) == 0:
    avg_disgust = 0
  else:
    avg_disgust = avg_fun(disgust_list)
```

Step 6: After running through each word in the review, calculates the average emotional score per emotion

```
#averages the scores for each emotion
if len(disgust_list) == 0:
  avg_disgust = 0
else:
  avg_disgust = avg_fun(disgust_list)
```

END RESULT:

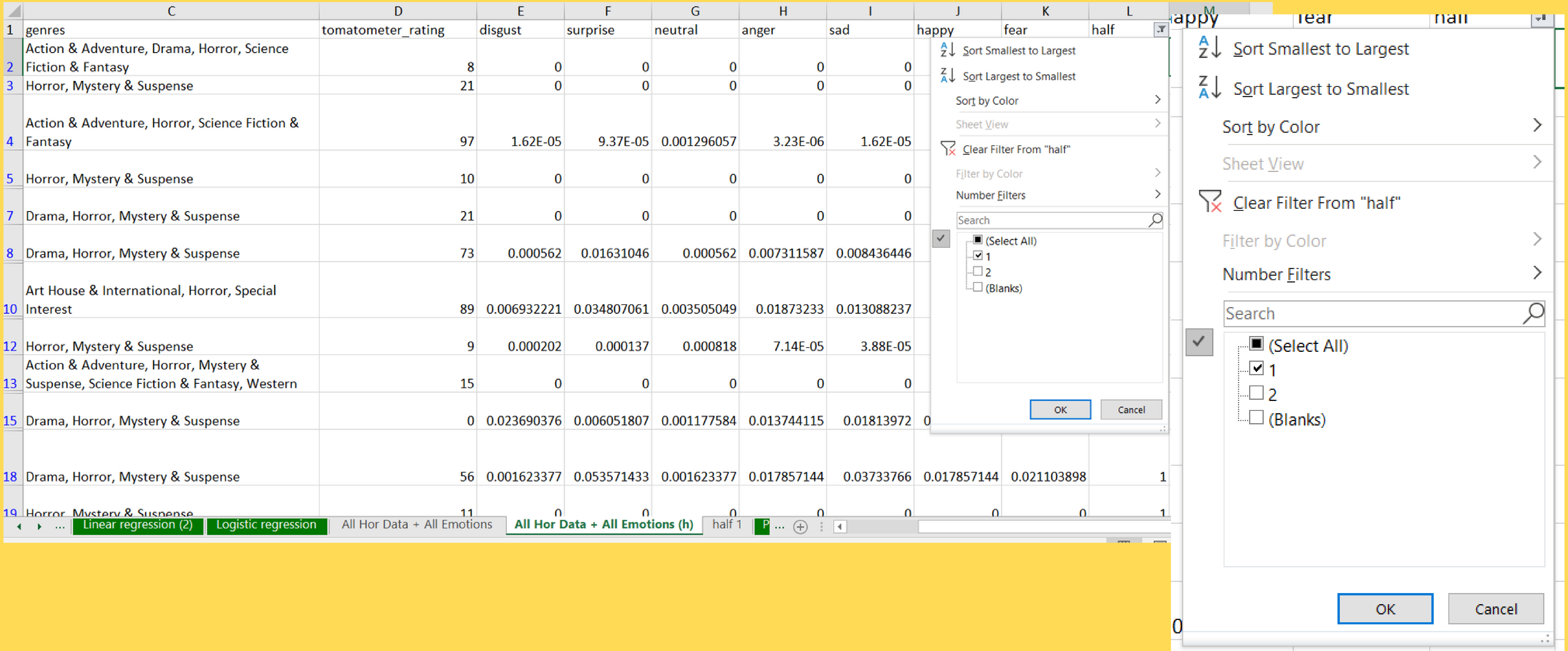
An output file with the average score per emotion based on the critic consensus for each horror movie

191 to 200 of 954 entries

disgust	surprise	neutral	anger	sad	happy	fear
0	0	0	0	0	0	0
0.007692308	0.010889112	9.990000000000002e-05	0.06883117	0.027072929	0.015084916	0.013686314
0.00027173333333333334	0.028978148666666665	0.0006995783333333333	0.0009151506666666667	0.0009047173333333334	0.016168389000000002	0.0009004673333333
0.002407705	0.012038523999999998	0.0008029999999999999	0.010433386999999999	0.01845907	0.098715894	0.004012841
0.000978	0.03033268	0.000978	0.014677105	0.034246575	0.036203522	0.03033268
0.008365508	0.022522522000000003	0.000644	0.027670529	0.01866152	0.045688547	0.0225225220000000
0.0156233285	0.0291064755	0.0027824525	0.0226859295	0.0355270205	0.0148207605	0.0362225810000000
0.0037440935	0.056706653999999995	0.0026534525000000003	0.0150490735	0.031661216	0.0278444215	0.0184660125000000
0.01084562025	0.0249609315	0.0015061299999999998	0.0175935025	0.020786613	0.024399588	0.05029576625
0.0024874556666666667	0.046304453999999995	0.0005545423333333334	0.02649584266666666	0.02441506266666664	0.0311452256666666665	[illegible]

MODELING & ANALYSIS

After calculating a score for each emotion, we split our data in half by randomly assigning each movie either a 1 or a 2.



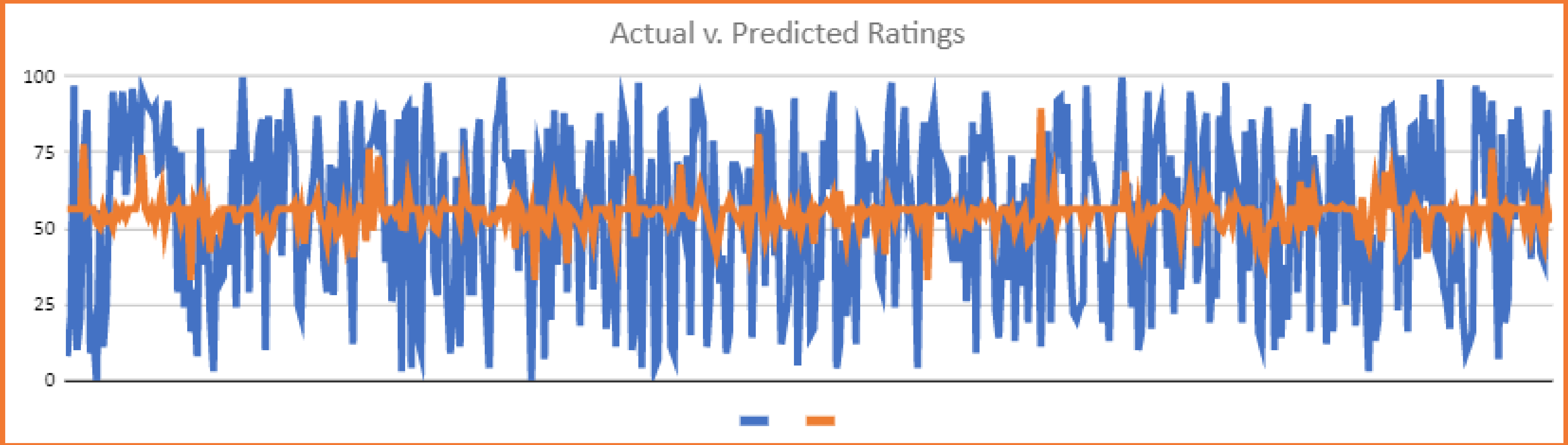
	C	D	E	F	G	H	I	J	K	L
1	genres	tomatometer_rating	disgust	surprise	neutral	anger	sad	happy	fear	half
2	Action & Adventure, Drama, Horror, Science Fiction & Fantasy	8	0	0	0	0	0			
3	Horror, Mystery & Suspense	21	0	0	0	0	0			
4	Action & Adventure, Horror, Science Fiction & Fantasy	97	1.62E-05	9.37E-05	0.001296057	3.23E-06	1.62E-05			
5	Horror, Mystery & Suspense	10	0	0	0	0	0			
7	Drama, Horror, Mystery & Suspense	21	0	0	0	0	0			
8	Drama, Horror, Mystery & Suspense	73	0.000562	0.01631046	0.000562	0.007311587	0.008436446			
10	Art House & International, Horror, Special Interest	89	0.006932221	0.034807061	0.003505049	0.01873233	0.013088237			
12	Horror, Mystery & Suspense	9	0.000202	0.000137	0.000818	7.14E-05	3.88E-05			
13	Action & Adventure, Horror, Mystery & Suspense, Science Fiction & Fantasy, Western	15	0	0	0	0	0			
15	Drama, Horror, Mystery & Suspense	0	0.023690376	0.006051807	0.001177584	0.013744115	0.01813972	0		
18	Drama, Horror, Mystery & Suspense	56	0.001623377	0.053571433	0.001623377	0.017857144	0.03733766	0.017857144	0.021103898	1
19	Horror, Mystery & Suspense	11	0	0	0	0	0	0	0	1

Response Variable = Rating

genres	tomatometer_rating	disgust	surprise	neutral	anger	sad	happy	fear
Action & Adventure, Drama, Horror, Science Fiction & Fantasy	8	0	0	0	0	0	0	0
Horror, Mystery & Suspense	21	0	0	0	0	0	0	0
Action & Adventure, Horror, Science Fiction & Fantasy	97	1.62E-05	9.37E-05	0.001296057	3.23E-06	1.62E-05	9.70E-06	9.70E-06
Horror, Mystery & Suspense	10	0	0	0	0	0	0	0
Art House & International, Drama, Horror, Mystery & Suspense	86	0	0	0	0	0	0	0
Drama, Horror, Mystery & Suspense	21	0	0	0	0	0	0	0
Drama, Horror, Mystery & Suspense	73	0.000562	0.01631046	0.000562	0.007311587	0.008436446	0.10179978	0.010686165
Drama, Horror, Mystery & Suspense	16	0	0	0	0	0	0	0
Art House & International, Horror, Special Interest	89	0.006932221	0.034807061	0.003505049	0.01873233	0.013088237	0.0274696	0.018829615
Horror, Mystery & Suspense, Science Fiction & Fantasy	44	1.62E-05	9.37E-05	0.001296057	3.23E-06	1.62E-05	9.70E-06	9.70E-06
Horror, Mystery & Suspense	9	0.000202	0.000137	0.000818	7.14E-05	3.88E-05	9.18E-05	7.96E-05
Action & Adventure, Horror, Mystery & Suspense, Science Fiction & Fantasy, Western	15	0	0	0	0	0	0	0

Explanatory Variable = Emotions

Next, we ran a log-linear regression on movies assigned to group 1.



The log-linear model fit our data better than log-log and linear models.

Out of the seven emotions we tested, anger, sad, and happy were the only statistically significant ones.

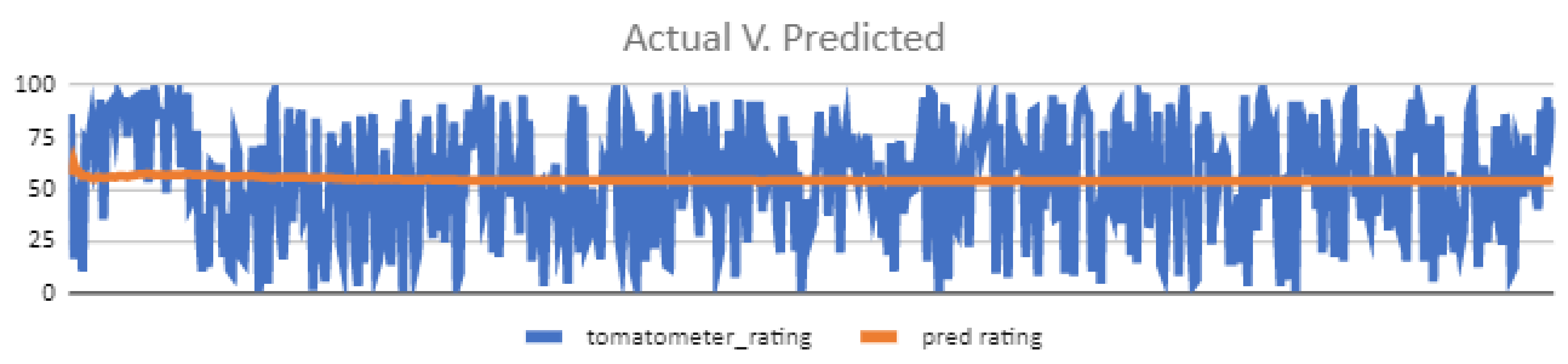
Source	Value	Standard error	Vald Chi-Squar	Pr > Chi²	Lower bound (	Upper bound (95%)
Intercept	4.034	0.009	200969.032	< 0.0001	4.016	4.052
anger	-4.916	0.495	98.784	< 0.0001	-5.886	-3.947
sad	-2.513	0.560	20.170	< 0.0001	-3.610	-1.416
happy	3.692	0.305	146.690	< 0.0001	3.094	4.289

Anger and sad both impacted rating negatively while happy impacted rating positively.

We applied our equation from the log-linear model to our holdout data (group 2).

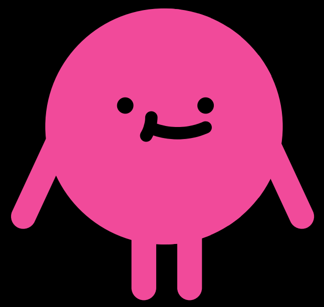
Equation of the model for the components (Variable tomatometer_rating):

Pred(tomatometer_rating) = exp(4.03405889085379-4.9162123172403*anger-2.51333995544336*sad+3.69184305708302*happy)



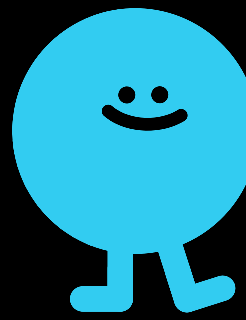
Ratings for group 2 movies were predicted based on how angry, sad and happy they were.

CONCLUSION



Reject our hypothesis

Movie reviews with words that evoke the emotion "fear" did not cause higher ratings for the movies.

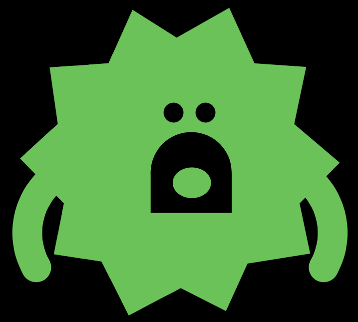


Predictive emotions

Although counterintuitive in a horror film, happy emotions contribute to higher ratings and anger & sad emotions lower the ratings.

Further Analysis

Our model contributes to a better review and predicts what will increase & decrease the rating. This model does not capture other variables that can contribute to a high review such as production value, acting quality, and editing techniques.



Limitations

We have to take into consideration movie reviews with high ratings tend to praise the film, thus using happier emotional verbiage.

THANK YOU!

Questions?

RESOURCE PAGE

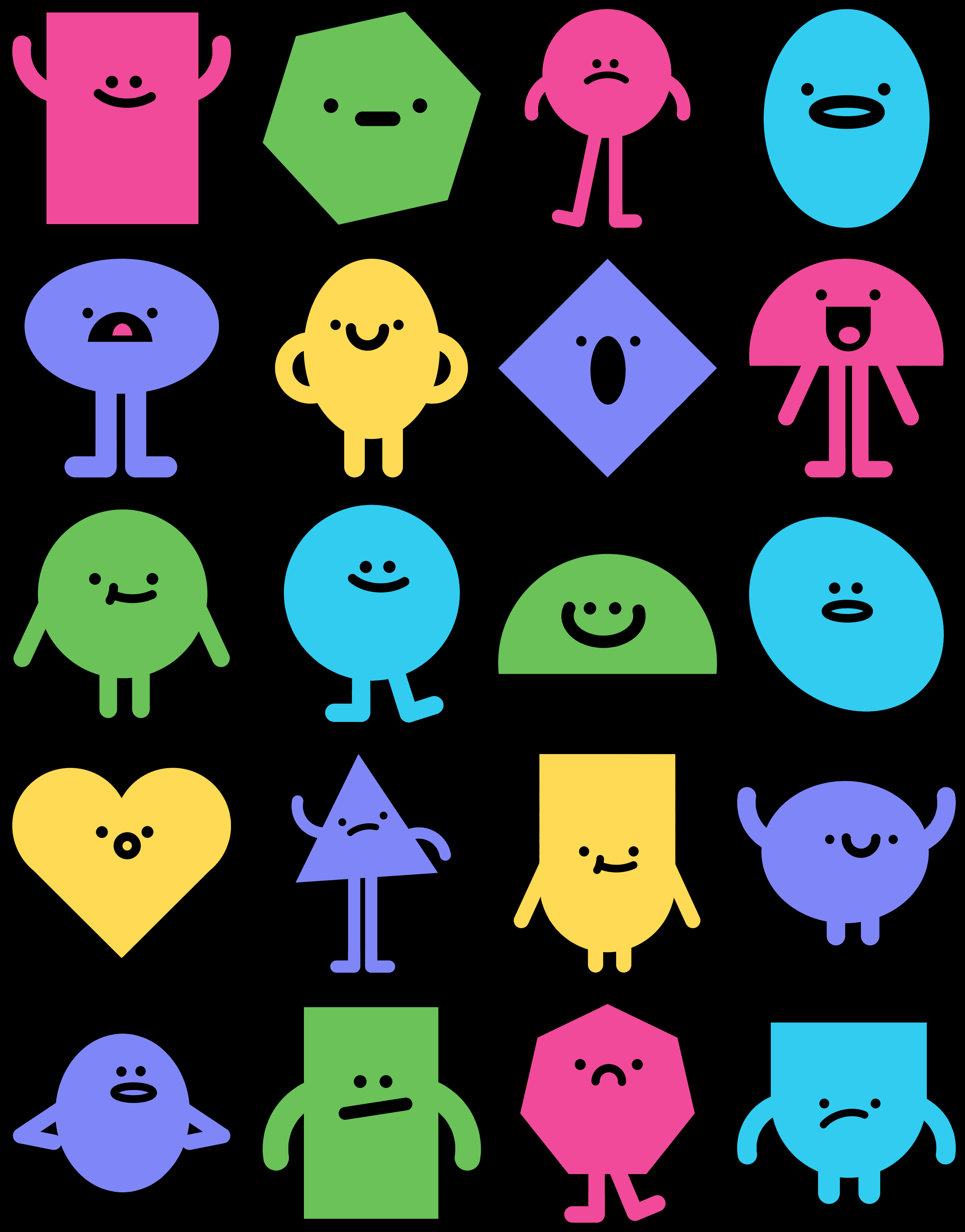
- https://www.kaggle.com/iwilldoit/emotions-sensor-data-set
- https://www.kaggle.com/stefanoleone992/rotten-tomatoes-movies-and-critic-reviews-dataset
- https://drive.google.com/file/u/0/d/1axF77q2HJXWFr0n0ZSLQ1nQaixw02OGi/edit

APPENDIX

- Emotional P-values & their significance

Model parameters for the components (Variable tomatometer_rating):

Source	Value	Standard error	Vald Chi-Squar	Pr > Chi^2	Lower bound (	Upper bound (95%)
Intercept	4.031	0.009	194246.314	< 0.0001	4.014	4.049
disgust	0.466	1.327	0.124	0.725	-2.134	3.067
surprise	0.455	0.471	0.931	0.335	-0.469	1.379
neutral	2.125	1.741	1.490	0.222	-1.287	5.538
anger	-5.054	0.512	97.417	< 0.0001	-6.057	-4.050
sad	-2.992	0.631	22.489	< 0.0001	-4.228	-1.755
happy	3.559	0.316	127.223	< 0.0001	2.941	4.178



- Code for wordcloud *text.txt includes all the reviews for horror movies

```
!pip install wordcloud
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS
import pandas as pd
```

```
Requirement already satisfied: wordcloud in /usr/local/lib/python3.7/dist-packages (1.5.0)
Requirement already satisfied: numpy>=1.6.1 in /usr/local/lib/python3.7/dist-packages (from wordcloud) (1.19.5)
Requirement already satisfied: pillow in /usr/local/lib/python3.7/dist-packages (from wordcloud) (7.1.2)
```

```
[ ] text = open('text.txt', mode='r', encoding = 'utf-8').read()
```

```
[ ] stopwords = STOPWORDS
```

```
[ ] wc = WordCloud(
        font_path = 'OpenSans-Regular.ttf',
        background_color = 'white',
        stopwords=stopwords,
        height=700,
        width=1300
    )
```

```
[ ] wc.generate(text)
```

```
<wordcloud.wordcloud.WordCloud at 0x7f1681778050>
```

```
[ ] wc.to_file('wordcloud3.png')
```