CLICK TO SEE SPOILER

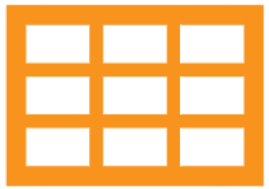
Detecting Movie Spoilers

BRIAN TRACY

Background

- Information that reveals important plot details or surprises
- Reviews influence consumer attitude and decisions on seeing a movie
- Users will avoid resources that present them with spoilers

Business Problem



Current model relies on site moderation and user reporting

Automated flagging systems can free resources for other problems

Goal: increase user confidence they will not be presented with spoilers

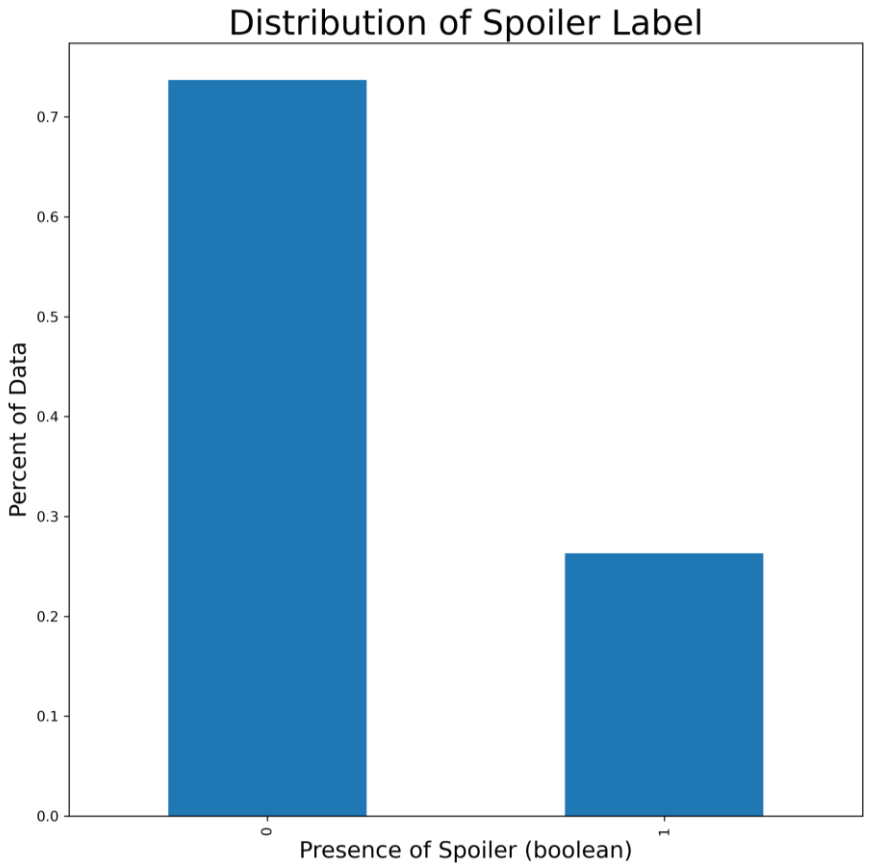
Data Understanding

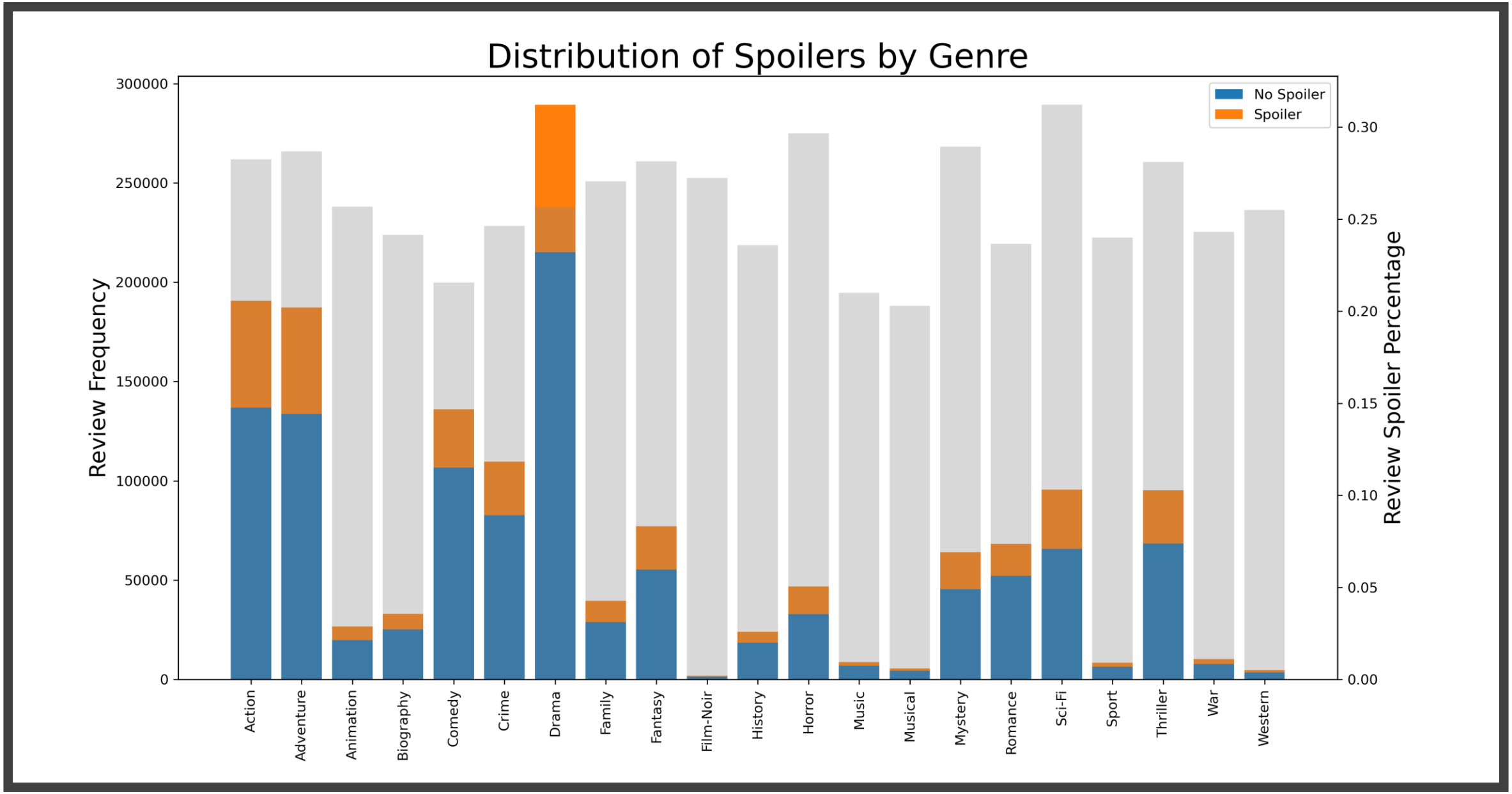
Data sourced from Kaggle

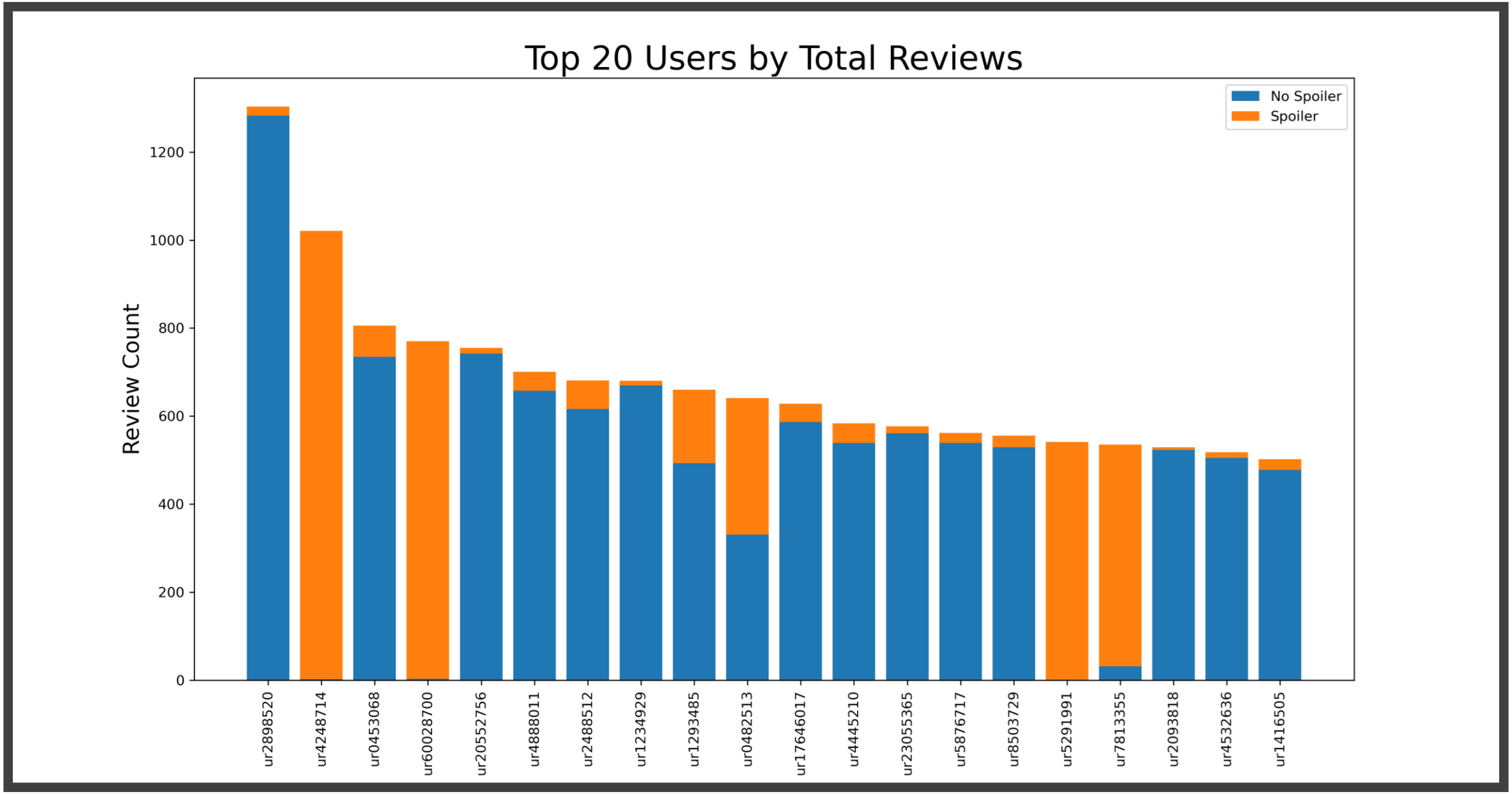
573,913 reviews between 1998 and 2018

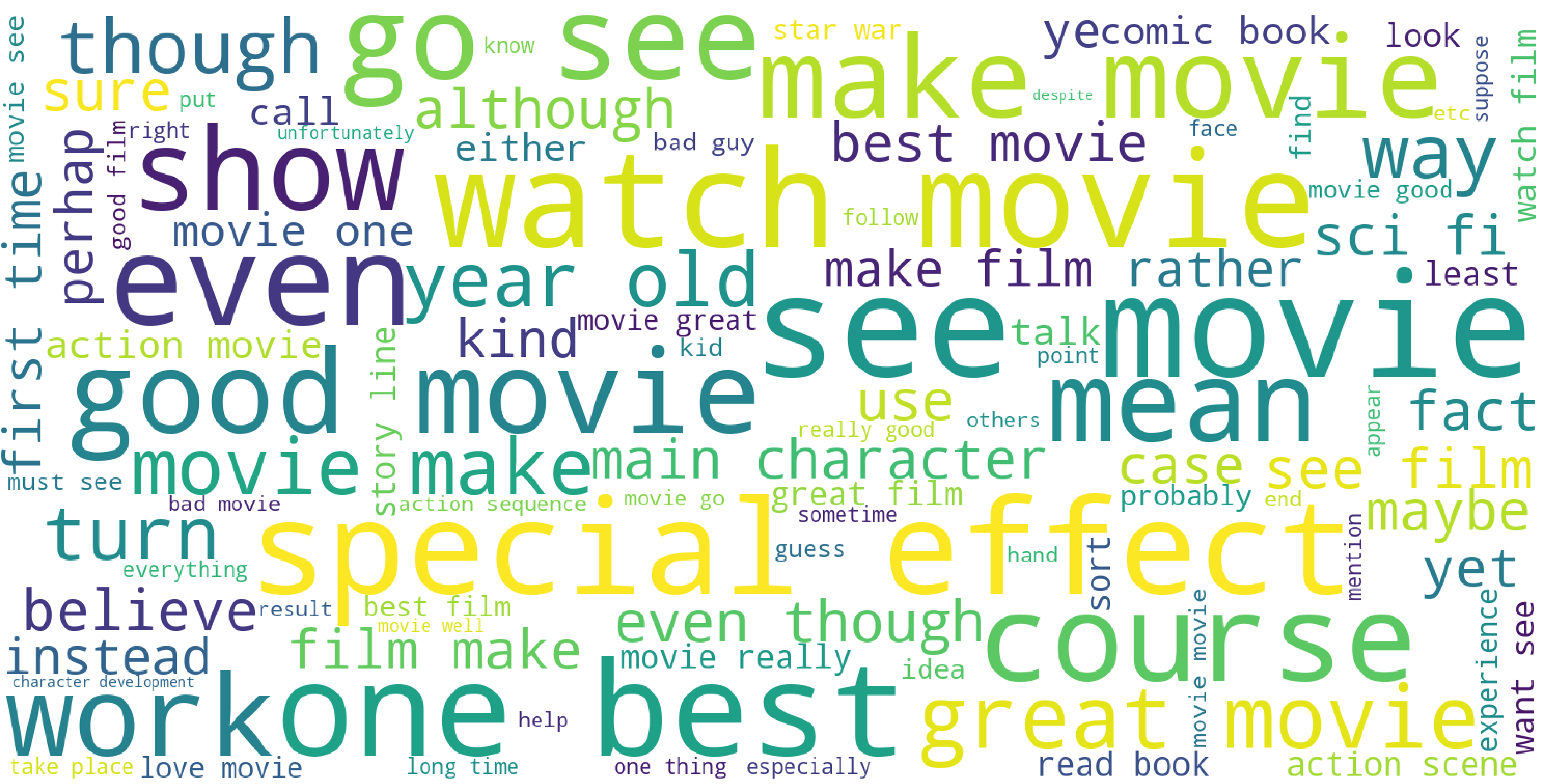
1572 unique films being reviewed

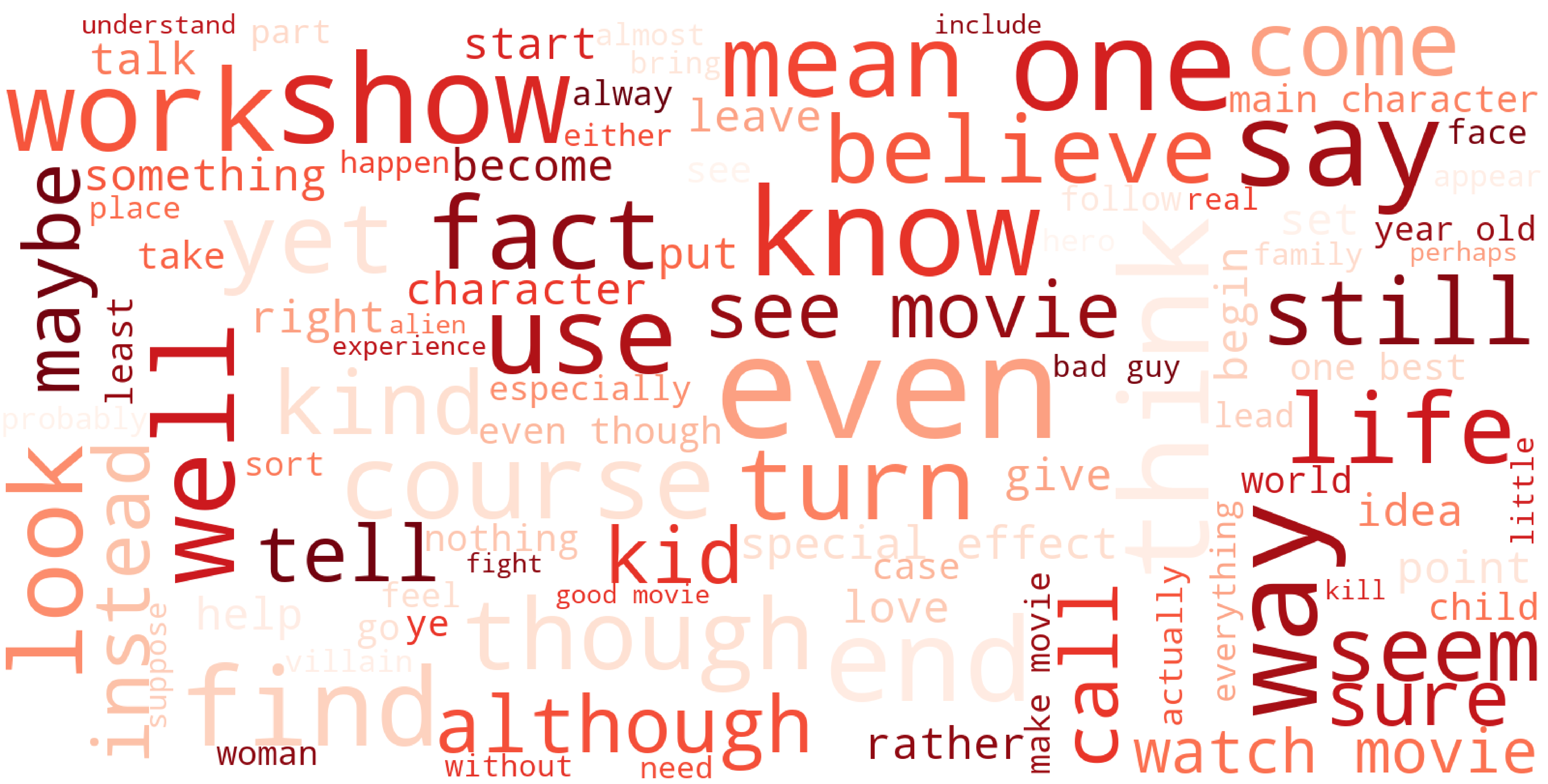
Additional film metadata and spoiler tagging




Distribution of Spoilers by Genre
Review Frequency
Review Spoiler Percentage
No Spoiler
Spoiler
300000
250000
200000
150000
100000
50000
0
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Action
Adventure
Animation
Biography
Comedy
Crime
Drama
Family
Fantasy
Film-Noir
History
Horror
Music
Musical
Mystery
Romance
Sci-Fi
Sport
Thriller
War
Western


Top 20 Users by Total Reviews
No Spoiler
Spoiler
Review Count
0
200
400
600
800
1000
1200
ur2898520
ur4248714
ur0453068
ur60028700
ur20552756
ur4888011
ur2488512
ur1234929
ur1293485
ur0482513
ur17646017
ur4445210
ur23055365
ur5876717
ur8503729
ur5291991
ur7813355
ur2093818
ur4532636
ur1416505

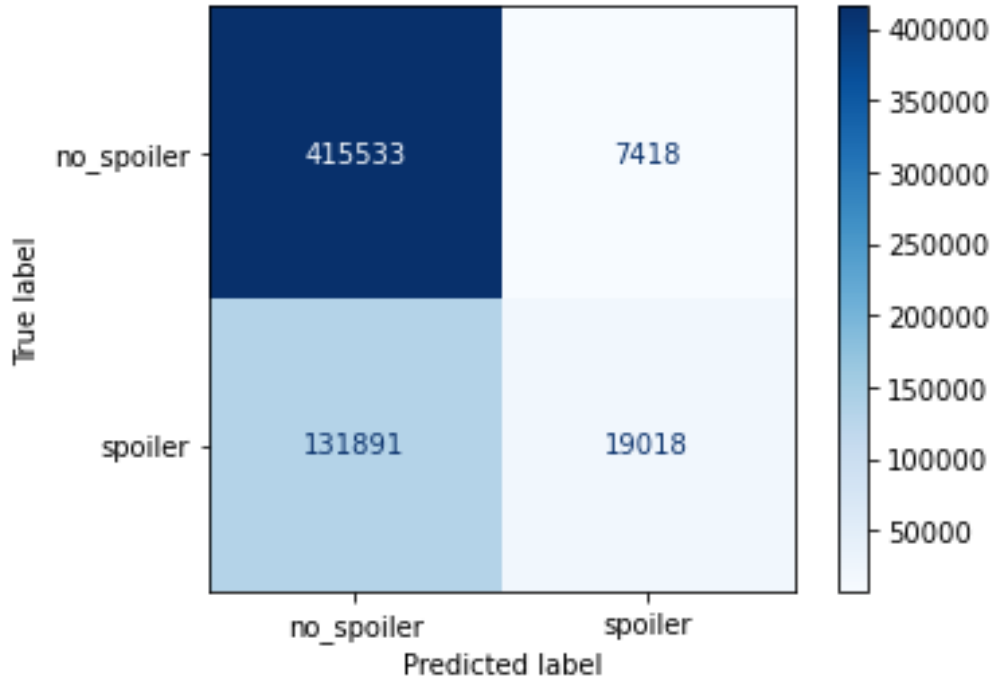

though
go
see
star war
know
make movie
ye
comic book
look
sure
put
call
although
despite
etc
suppose
watch film
movie see
right
unfortunately
show
either
bad guy
best movie
face
find
way
movie time
perhap
good film
watch movie
movie good
follow
sci fi
movie one
even
year old
make film
rather
least
first
action movie
kind
movie great
kid
see movie
talk
point
good movie
story line
use
mean
fact
others
really good
appear
movie
make
main character
case
see film
must see
bad movie
action sequence
movie go
great film
probably
end
sometime
maybe
turn
special effect
guess
hand
sort
mention
yet
everything
believe
result
best film
movie well
even though
course
movie movie
experience
want see
instead
film make
movie really
idea
character development
work
one
best
great movie
help
take place
love movie
long time
one thing
especially
read book
action scene


understand
part
start
almost
include
talk
bring
mean
one
come
workshow
alway
leave
main character
either
believe
say
face
something
happen
become
see
appear
maybe
place
fact
know
follow
real
yet
set
year old
take
put
hero
family
perhaps
character
think
begin
right
alien
see movie
still
least
experience
use
well
kind
even
bad guy
one best
especially
probably
lead
life
even though
sort
course
turn
give
little
look
instead
world
idea
tell
nothing
kid
special effect
fight
case
point
way
kill
feel
good movie
love
child
help
go
ye
though
end
seem
suppose
villain
make movie
call
actually
everything
find
although
sure
rather
woman
without
need
watch movie

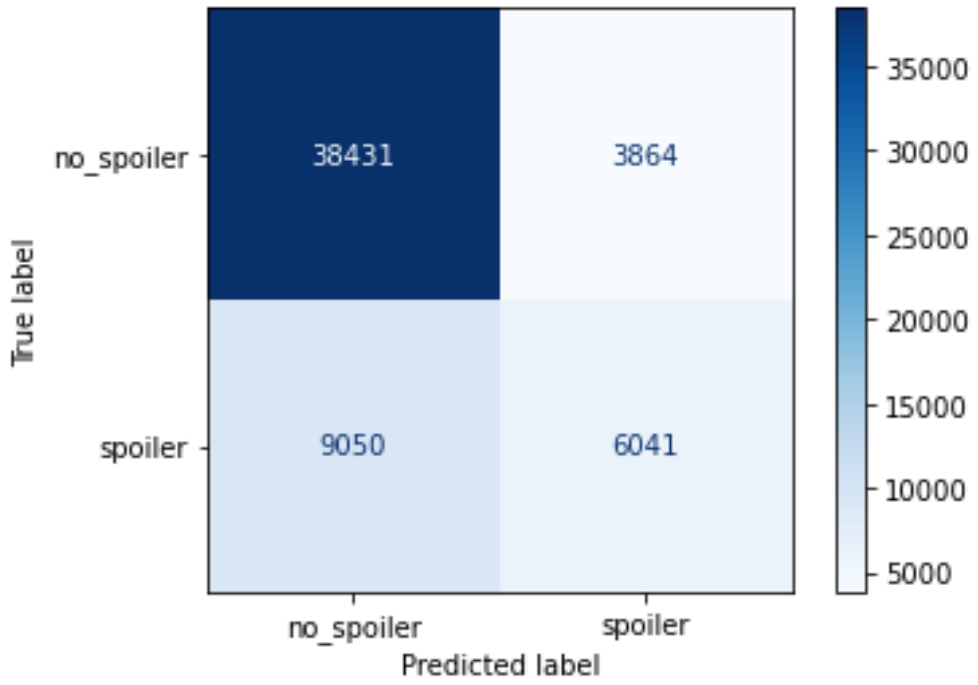
Baseline modeling

- Premise: users will self-tag spoilers
  - Does not account for negation
- Also modeled non-language features
  - Results in all reviews predicted as no spoiler

	precision	recall	f1-score	support
no_spoiler	0.76	0.98	0.86	422951
spoiler	0.72	0.13	0.21	150909
accuracy			0.76	573860
macro avg	0.74	0.55	0.54	573860
weighted avg	0.75	0.76	0.69	573860



	precision	recall	f1-score	support
no_spoiler	0.81	0.91	0.86	42295
spoiler	0.61	0.40	0.48	15091
accuracy			0.77	57386
macro avg	0.71	0.65	0.67	57386
weighted avg	0.76	0.77	0.76	57386



Results

- Iterative testing with different encoding tools and parameters
- Evaluating using f1-score and overall accuracy
- Utilized several different encoding methods for bag of words modeling

Next Steps

- Scrape movie-centric subreddits
- Explore cosine similarity
- Use transfer learning

Questions?

brtracy1984@gmail.com

https://github.com/brtracy