



CS 5830 Final Project:

Analyses of YouTube Data

Ann Marie Humble & Emma Lynn May

Dataset Creation

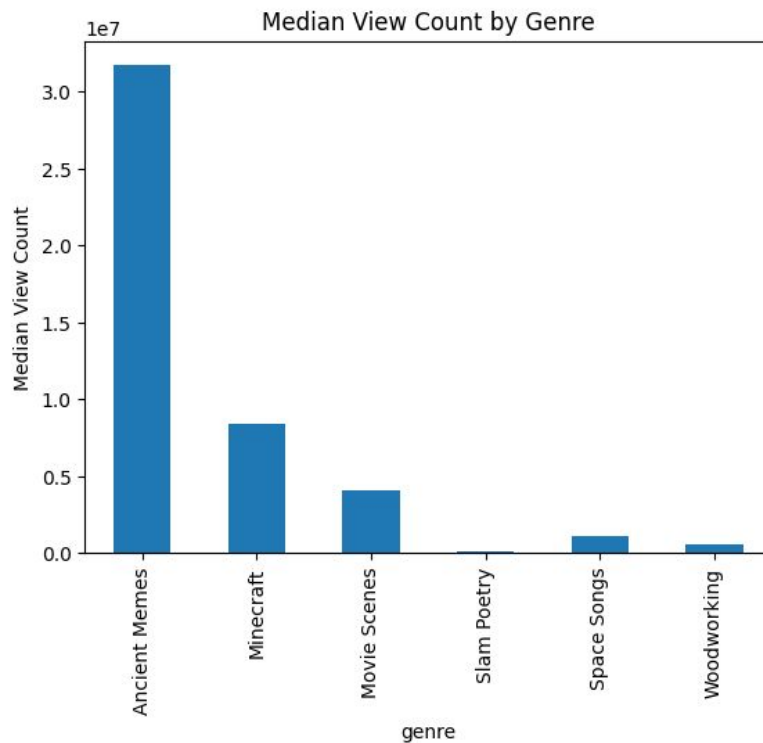
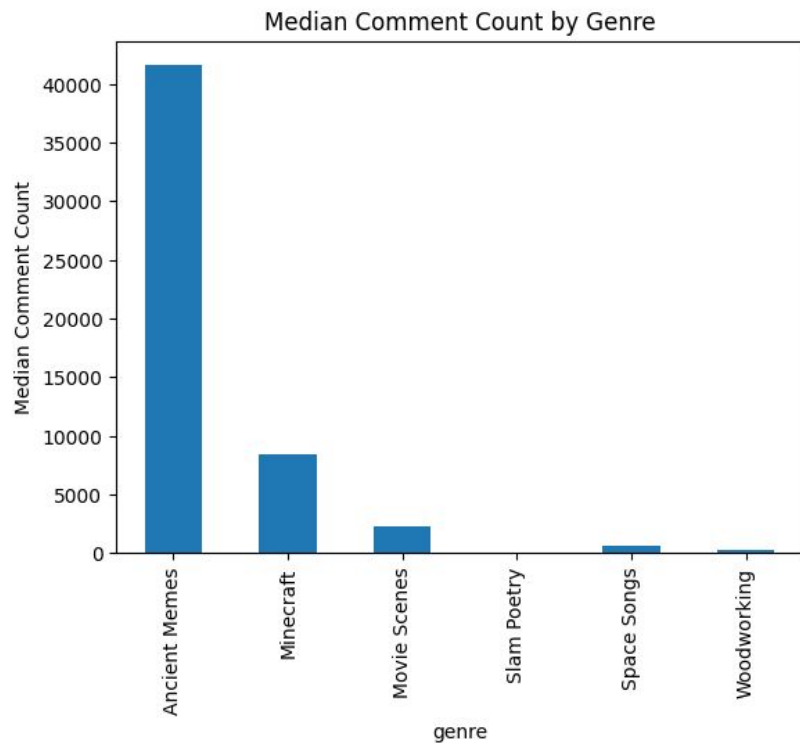
YouTube API

- We pulled info on videos from the following playlists:
 - Slam Poetry
 - Minecraft
 - Movie Scenes
 - Woodworking
 - Space Songs
 - Ancient Memes
 - Personal Finance
 - Travel
 - Health
- Endpoints used:
 - /videos
 - /playlistItems

```
[x] snippet
  [x] publishedAt*
  [x] title*
  [x] description*
  [x] tags
  [x] categoryId
  [x] defaultLanguage
[x] contentDetails
  [x] duration*
[x] statistics
  [x] viewCount*
  [x] likeCount*
  [x] dislikeCount*
    Not available unless you're the video owner
  [x] favoriteCount*
  [x] commentCount*
[x] paidProductPlacementDetails
  [x] hasPaidProductPlacement*
[x] topicDetails
  [x] topicIds
    Deprecated
  [x] relevantTopicIds
    Deprecated
  [x] topicCategories
```

Trends in Comment Sections

What videos do people tend to leave more comments on?



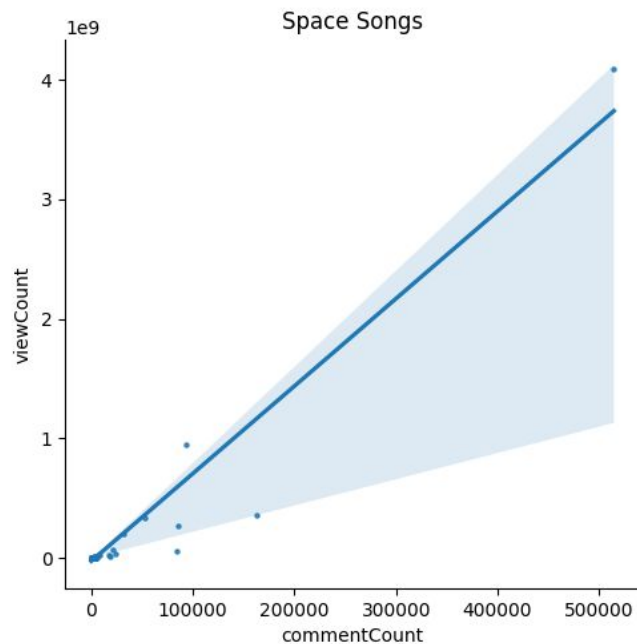
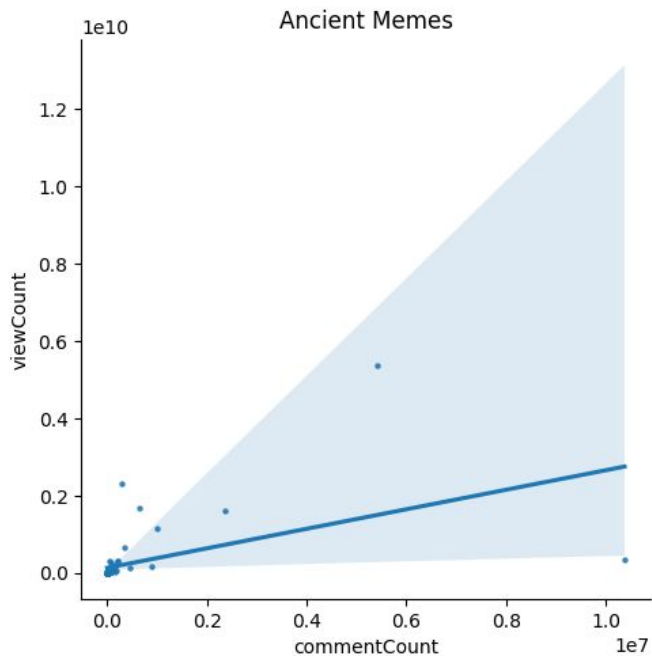
notice
how
similar
comment
count and
view
count are

Is there a correlation between comment count and view count?

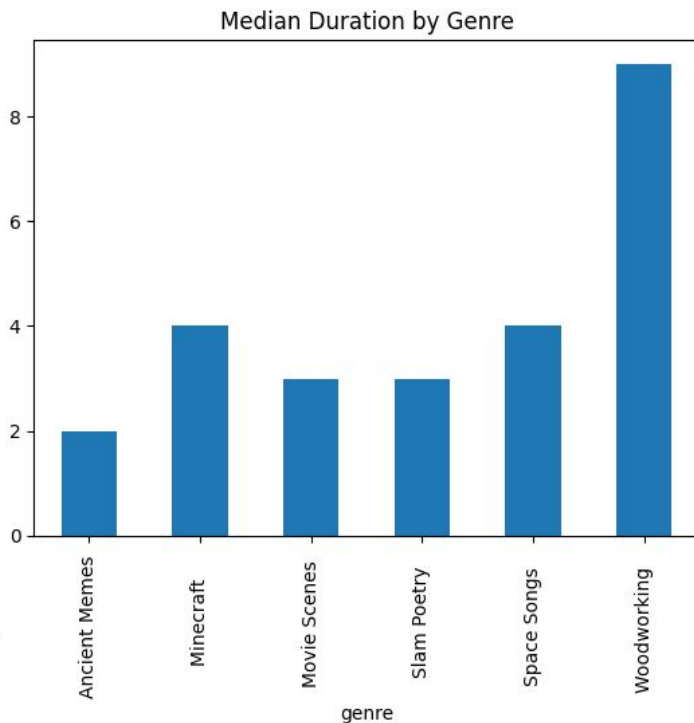
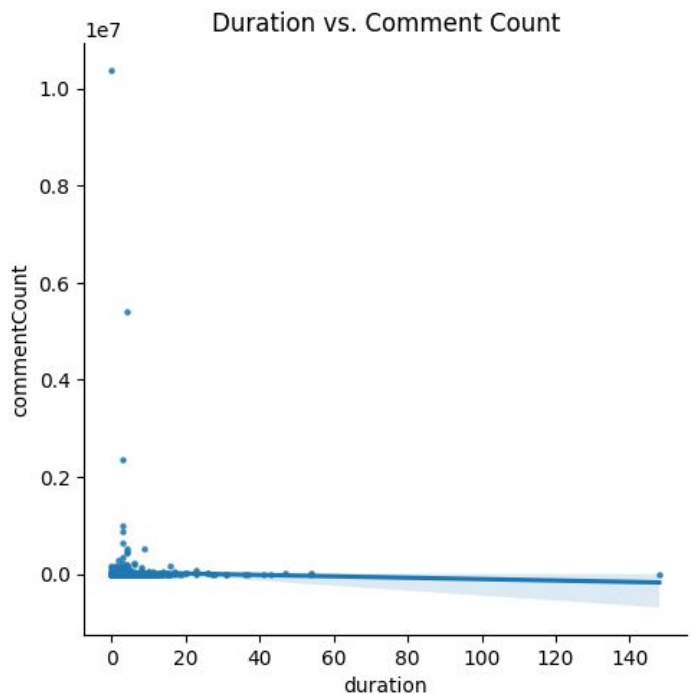
Space Songs	0.961765
Movie Scenes	0.919592
Slam Poetry	0.854547
Minecraft	0.719207
Woodworking	0.595980
Ancient Memes	0.477798

correlation coefficient

perhaps videos in the
woodworking and ancient
memes genre attract more
comments than other genres
do



Is there a correlation between comment count and duration?



Movie Scenes	0.269606
Woodworking	0.148625
Space Songs	0.047575
Slam Poetry	-0.020523
Minecraft	-0.045009
Ancient Memes	-0.056305

correlation coefficient

not a strong correlation –
the strongest is in Movie
Scenes

videos with highest comment count



Me at the zoo

341M views • 19 years ago



jawed ✓

00:00 Intro 00:05 The cool thing 00:17 End.

CC



Ancient Memes (comment count: 10376239)



Michael Jackson - Smooth Criminal (Official Video)

1B views • 14 years ago

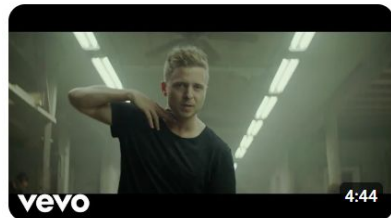


Michael Jackson 🌟

The short film for Michael Jackson's "Smooth Criminal" was the centerpiece of the feature film...



Movie Scenes (comment count: 527602)



OneRepublic - Counting Stars

4B views • 11 years ago



OneRepublic 🌟

#OneRepublic #CountingStars Music video by OneRepublic performing Counting Stars. (C) 2013...



Space Songs (comment count: 514504)



"Fallen Kingdom" - A Minecraft Parody of...

161M views • 12 years ago



CaptainSparklez ✓

Programs used: Maya After Effects Vegas Photoshop Pro Tools Lyrics: I used to rule the worl...



Minecraft (comment count: 464278)



I Will Not Let An Exam Result Decide My...

6.5M views • 11 years ago



Suli Breaks 🌟

Click(cc) for subtitles. I Will Not Let An Exam Result Decide My Fate - Suli Breaks,Vanity Fair,2009 "I Will...



Slam Poetry (comment count: 13144)



8ft Floating Shelf With Hidden Storage | DIY Build

3.5M views • 8 years ago



DIY Creators ✓

In this video I build an 8ft Floating Shelf With Hidd...



Woodworking (comment count: 4843)

YouTube Video Success

How to define the success of a video?

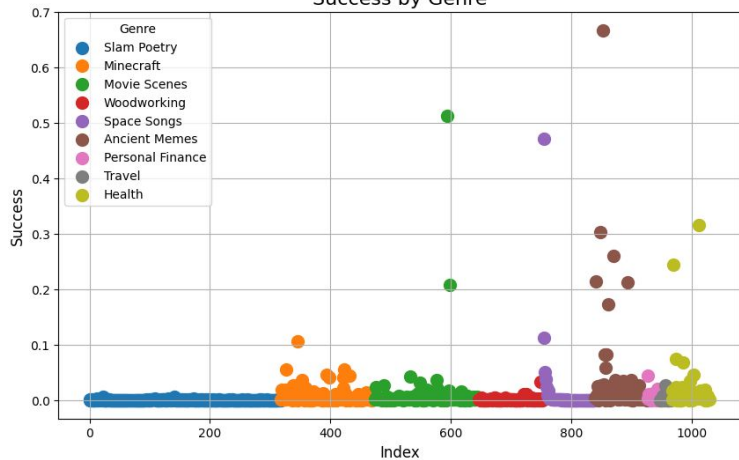
- We used these metrics to define a video's success:
 - Views
 - Likes
 - Estimated Ad Revenue
- Ad Revenue was estimated by multiplying a video's view count (in 1000's) by the genre's CPM (how much an advertiser will pay for every 1000 views)
- We looked at these metrics separately, as well as combined into a normalized overall success score

```
AD_REVENUE_CPM = {  
    'Slam Poetry': 0.36,  
    'Minecraft': 0.44,  
    'Movie Scenes': 1.24,  
    'Woodworking': 3.02,  
    'Space Songs': 0.01,  
    'Ancient Memes': 0.00,  
    'Health': 14.5,  
    'Travel': 17.5,  
    'Personal Finance': 18,
```

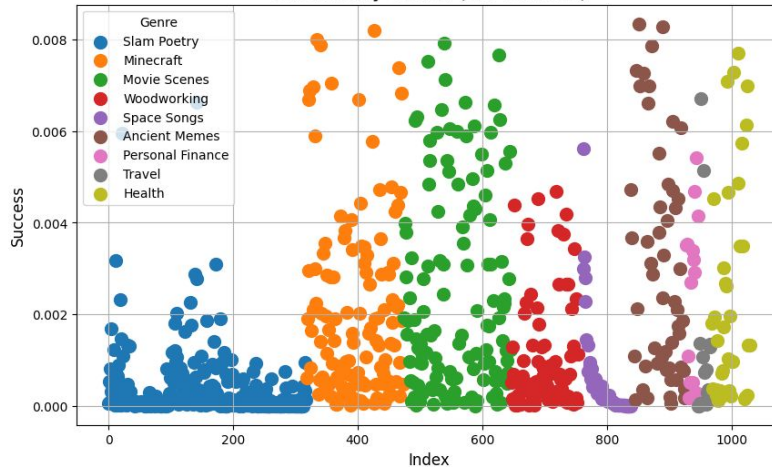
Sources: Yahoo Finance & UberSuggest
(2023)

Overall Success Score

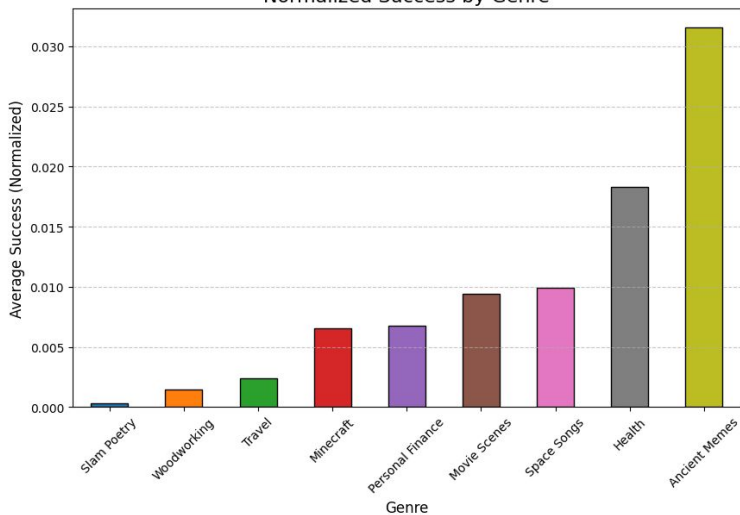
Success by Genre



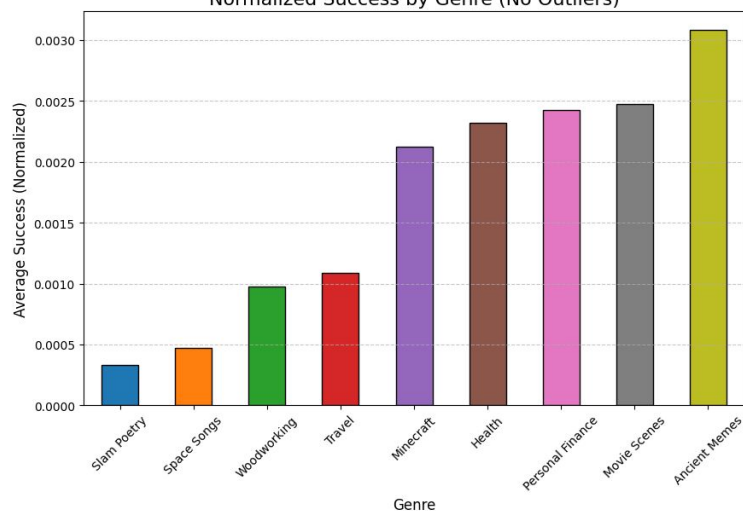
Success by Genre (No Outliers)



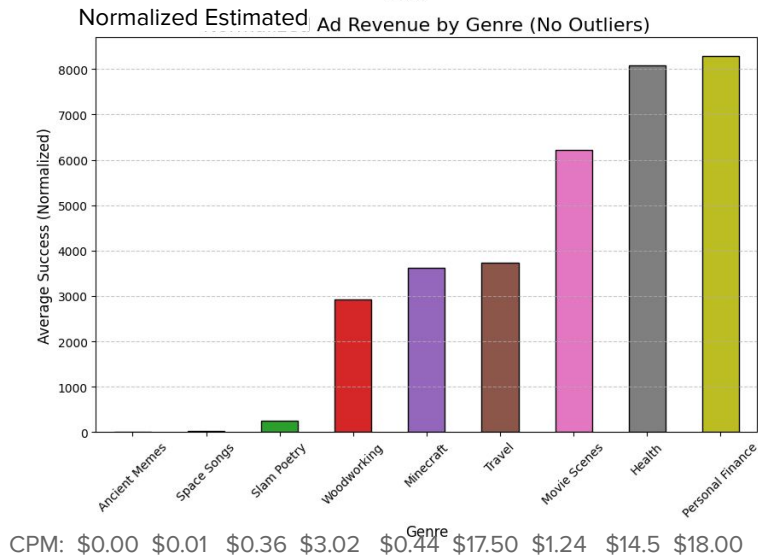
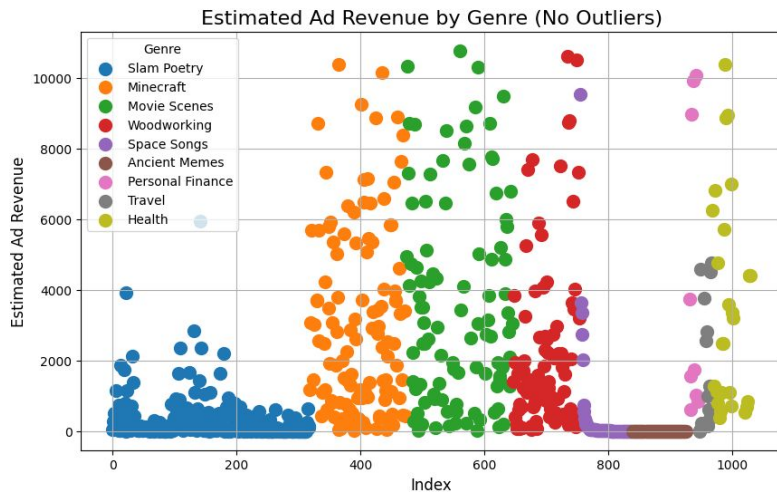
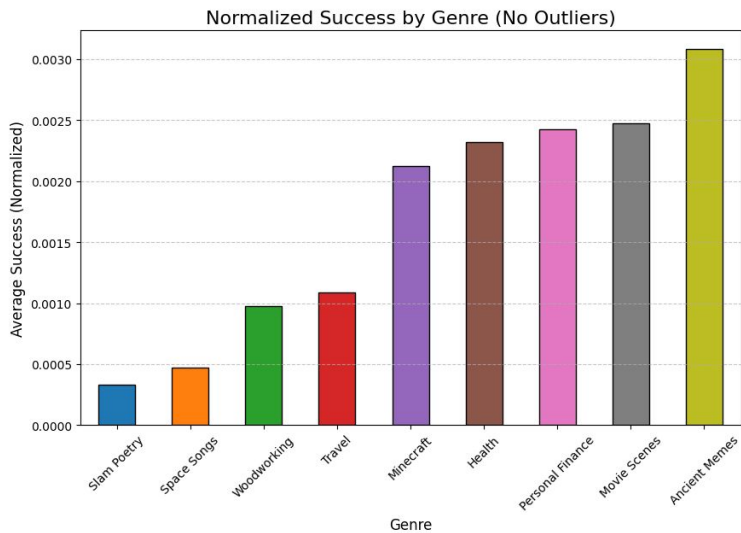
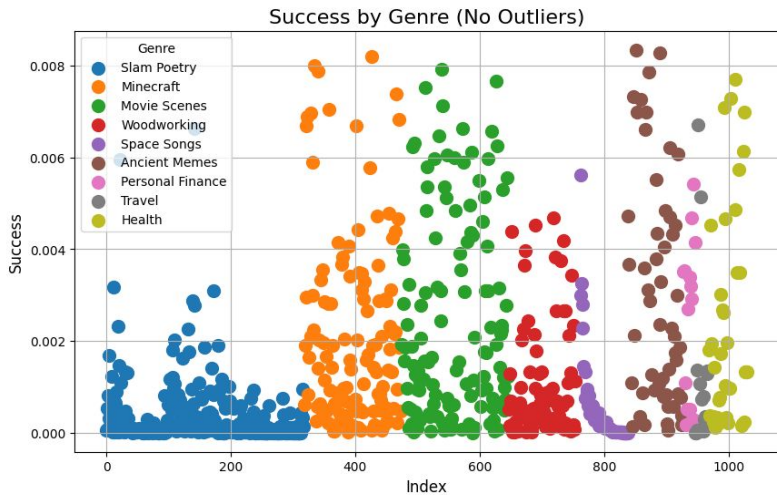
Normalized Success by Genre



Normalized Success by Genre (No Outliers)



Overall Success Score



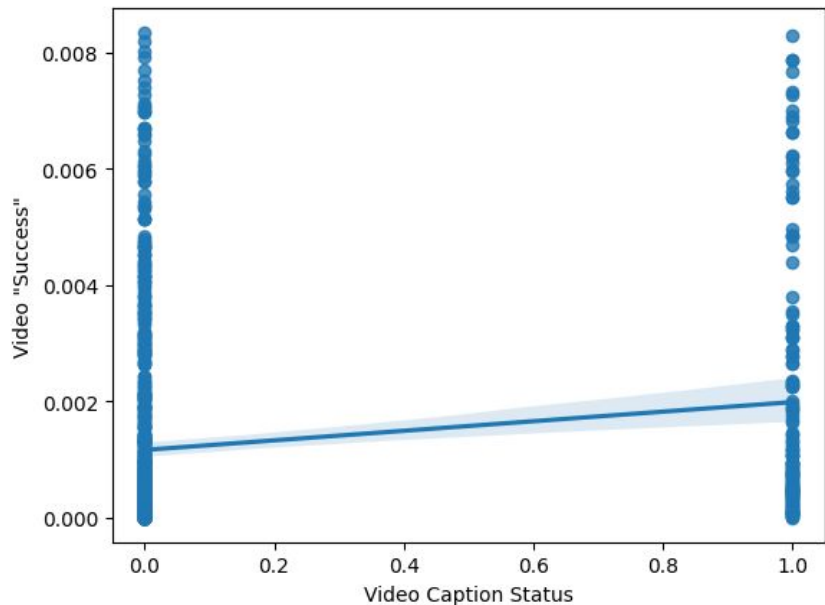
Estimated Ad Revenue

What video attributes are most correlated with success?

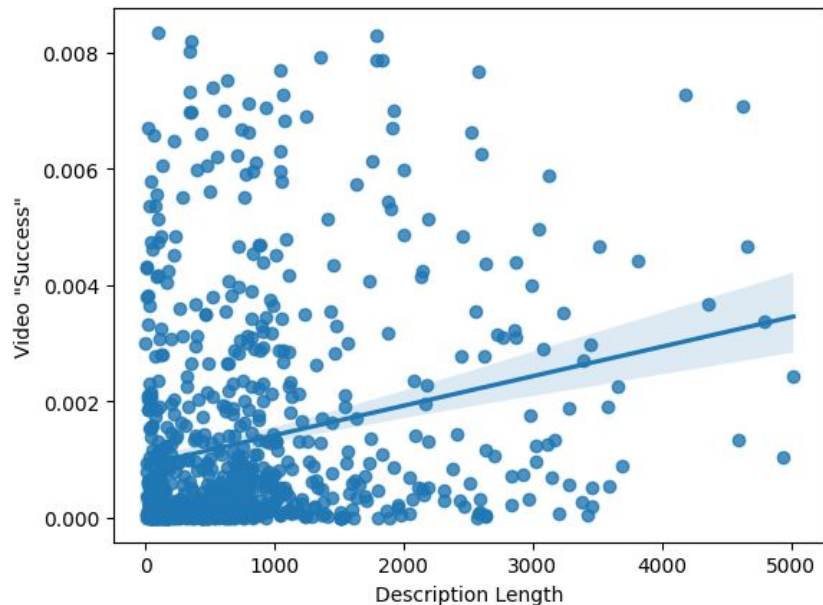
Significance Level = .05

Video Attributes	Pearson Correlation Coefficient	p-value
Length	0.037	0.284
Caption Status	0.168	0.0
Description Length	0.237	0.0
Title Length	0.008	0.813
Paid Product Placement	0.01	0.773

What video attributes are most correlated with success?

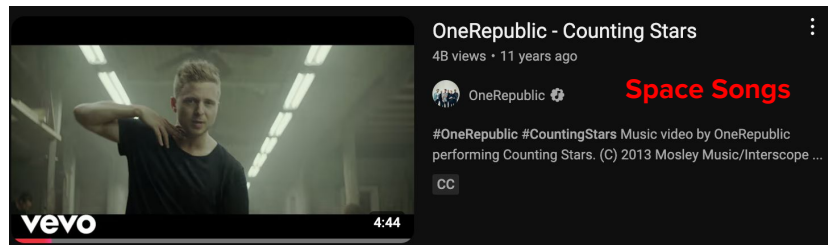
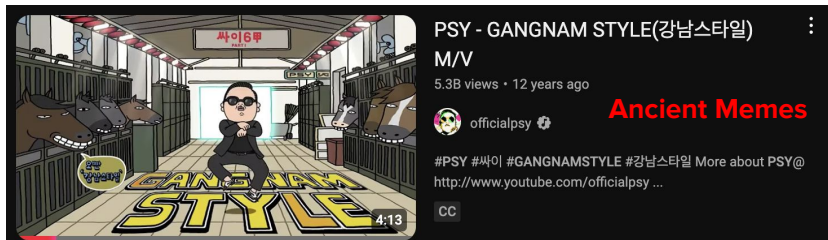


Caption Correlation: 0.168



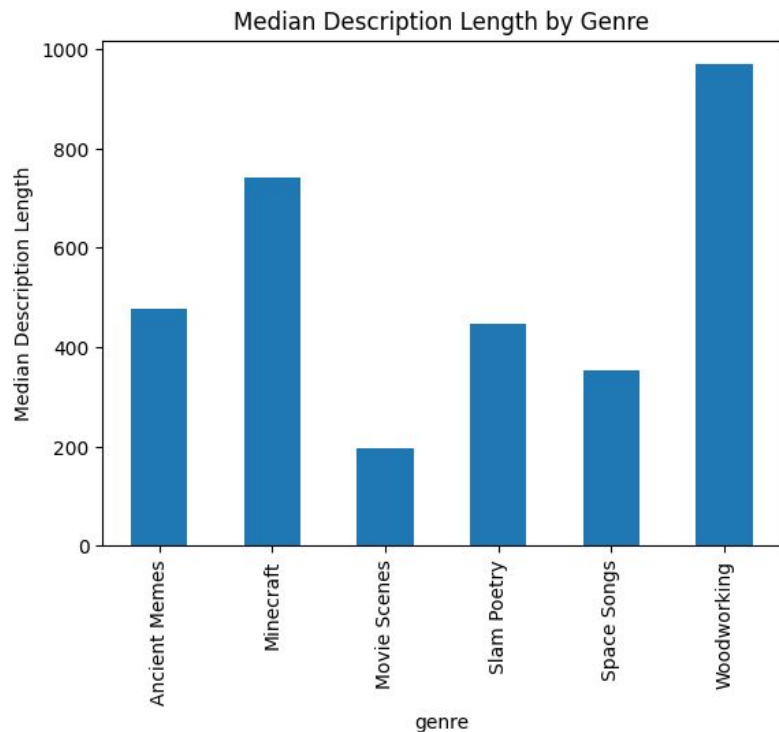
Caption Correlation: 0.237

Notable Outliers (Most “Successful” Videos)



Analyzing Descriptions

median and max description length



Max description length in each genre:

Ancient Memes	22077
Minecraft	5034
Movie Scenes	8604
Slam Poetry	3424
Space Songs	8899
Woodworking	5018

woodworking has
highest median
description length –
perhaps because of
all the technical info in
their video
descriptions

Max description for all:

"Weird Al Yankovic - Trapped In The Drive-Thru"
(Ancient Memes)

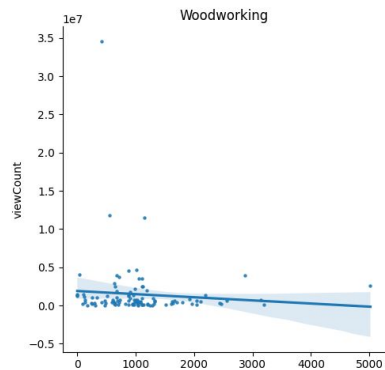
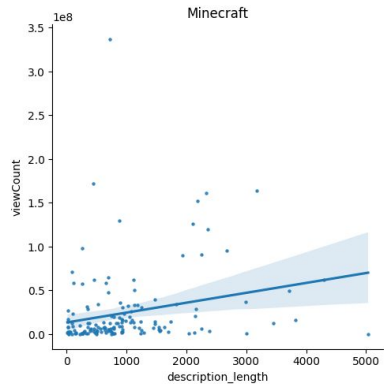
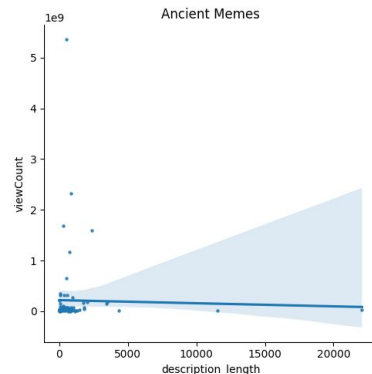
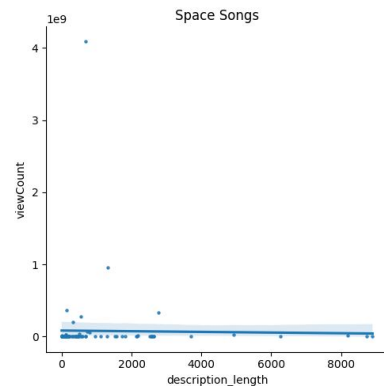
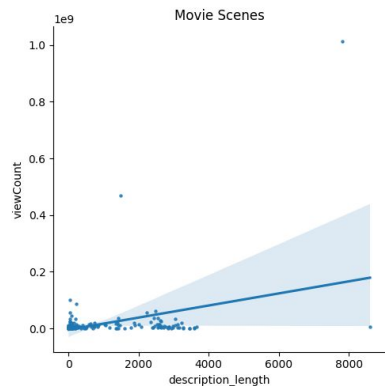
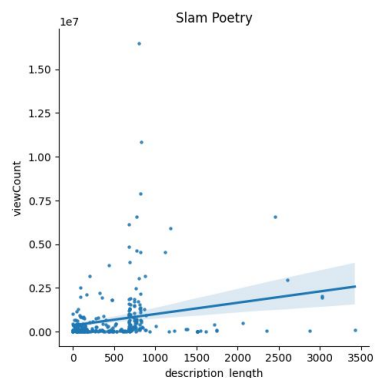
(no obvious trend in song description length over
time)

Predicting genre using description length

Overall F1-Score of 0.30 (not very good)

	precision	recall	f1-score	support
Ancient Memes	0.04	0.05	0.04	20
Minecraft	0.31	0.33	0.32	46
Movie Scenes	0.25	0.33	0.29	42
Slam Poetry	0.52	0.37	0.43	83
Space Songs	0.12	0.12	0.12	17
Woodworking	0.28	0.29	0.29	24

Is there a correlation between description length and view count?



Movie Scenes	0.350554
Minecraft	0.236534
Slam Poetry	0.217899
Space Songs	-0.018624
Ancient Memes	-0.024420
Woodworking	-0.084239

it's interesting that Woodworking doesn't have a higher correlation, since Woodworking descriptions are probably more vital to the Woodworking audience

Predicting genre using description content

Using TF-IDF Vectorization (TfidfVectorizer) gave us an overall F1-Score of 0.66

Using a Count Vectorizer (CountVectorizer) have us an overall F1-Score of 0.81

	precision	recall	f1-score	support
Ancient Memes	0.75	0.30	0.43	20
Minecraft	0.76	0.89	0.82	46
Movie Scenes	0.58	0.88	0.70	42
Slam Poetry	0.97	0.89	0.93	83
Space Songs	1.00	0.65	0.79	17
Woodworking	0.95	0.75	0.84	24

highest F1-Score
in Slam Poetry,
then
Woodworking,
then Minecraft

(these are very
distinctive and
probably more
consistent
genres)

lowest score (by far) is Ancient Memes – not surprising because this genre is very inconsistent :)

Which words are the best predictors of each genre?

Ancient Memes

funny, dancing, haha, spartaaaa

Minecraft

minecraft, victim, server, dennis

Movie Scenes

movie, scene, film, disney, speech

Slam Poetry

poetry, slam, poem, spoken, performs

Space Songs

album, space, moon, music, futuristic

Woodworking

make, table, drill, press, woodworking,
saw

Questions?