

# CORRELATION BETWEEN 2022 MOVIES AND 2022 HALLOWEEN COSTUMES



**Observation:** On Halloween people overwhelmingly wear costumes of popular movie characters

- Do the year's most popular movies have a significant correlation to the year's most popular Halloween costumes?
- How do the current year's movie-related costumes stack up against "traditional" Halloween costumes such as 'Vampire', 'Frankenstein', 'Zombie', 'Witch', 'Mummy', 'Pumpkin', 'Bat', 'Devil', and 'Ghost'?

## Methodology:

1. Collected data from the Google Frightgeist dataset that tracks and ranks Google search trends related to 2022 Halloween costumes
2. Obtain list of top 100 most popular movies from 2022 using Kaggle's "Latest Movie Dataset".
3. Using python and a NPL package to parse the movie titles and movie overview to get character names and match them to the costume list.
4. Using Python to filter out the "traditional" costumes from the costume's list
5. Compare the average costume ranking of the traditional vs. movie-related costumes.
6. Correlate movie-related costume's ranking to movie popularity

## Top 3 Movie Related Costumes

**1st Hocus Pocus - Hocus Pocus 2**  
(ranked 14 on Frightgeist)

**2nd Batman - The Batman**  
(ranked 19 on Frightgeist)

**3rd Top Gun - Top Gun: Maverick**  
(ranked 43 on Frightgeist)

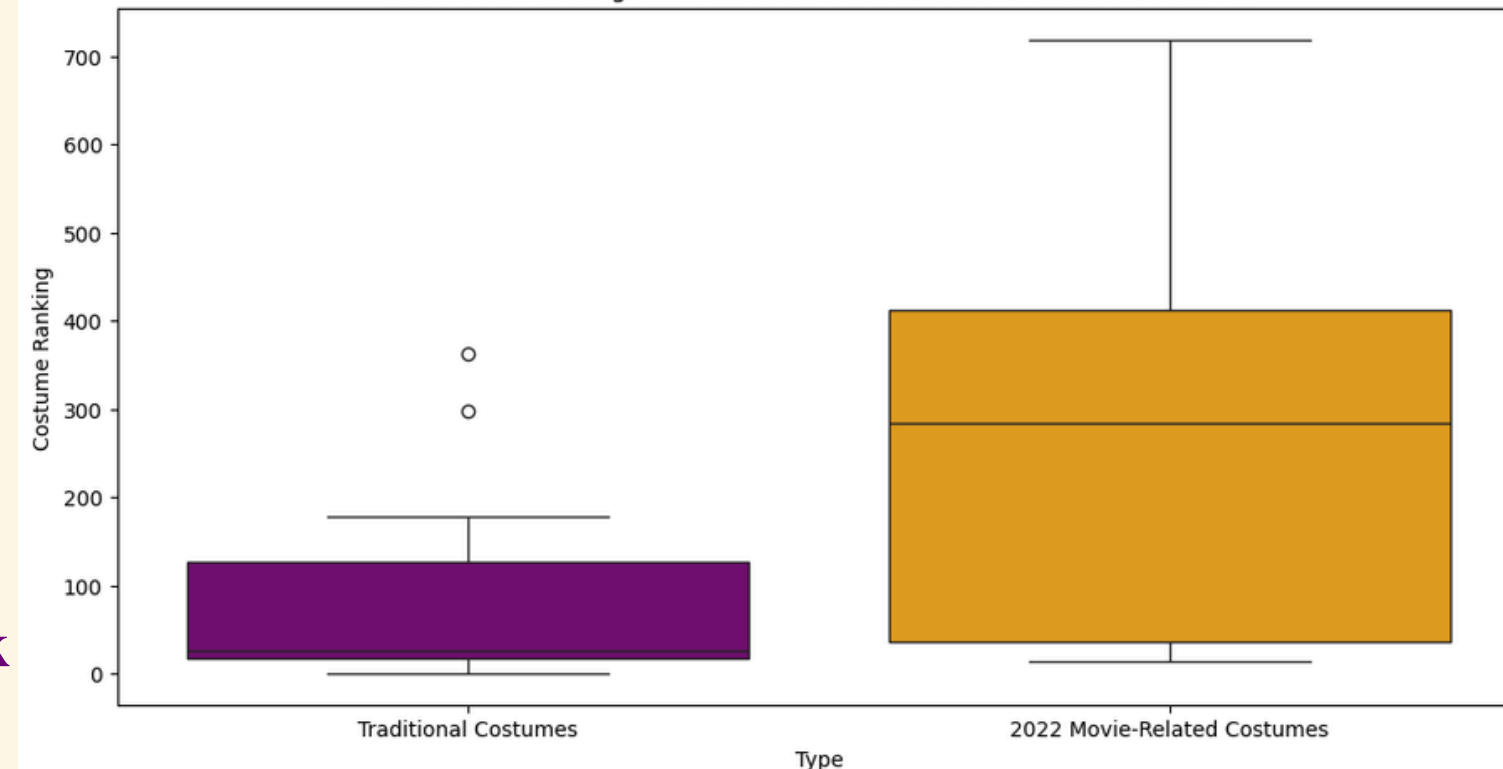
## Top 3 Traditional Costumes

**1st Witch**  
(ranked 1 on Frightgeist)

**2nd Vampire**  
(ranked 13 on Frightgeist)

**3rd Pumpkin**  
(ranked 15 on Frightgeist)

Box Plot of Rankings for Traditional vs. 2022 Movie-Related Costumes



Traditional costumes on average ranked higher in the **top 99**, while 2022 Movie-related costumes on average ranked in the **top 275** and have considerably more variability. Though there are a lot of movie characters on the costume list, only 8 costumes match with the the Top 100 movies of 2022. This indicates characters from movies are popular, but just not 2022s and there are other factors to consider (cultural impact, marketing, sequels etc.) It is of note that the top movie-related costume is a movie about witches and the top traditional costume is also witch.

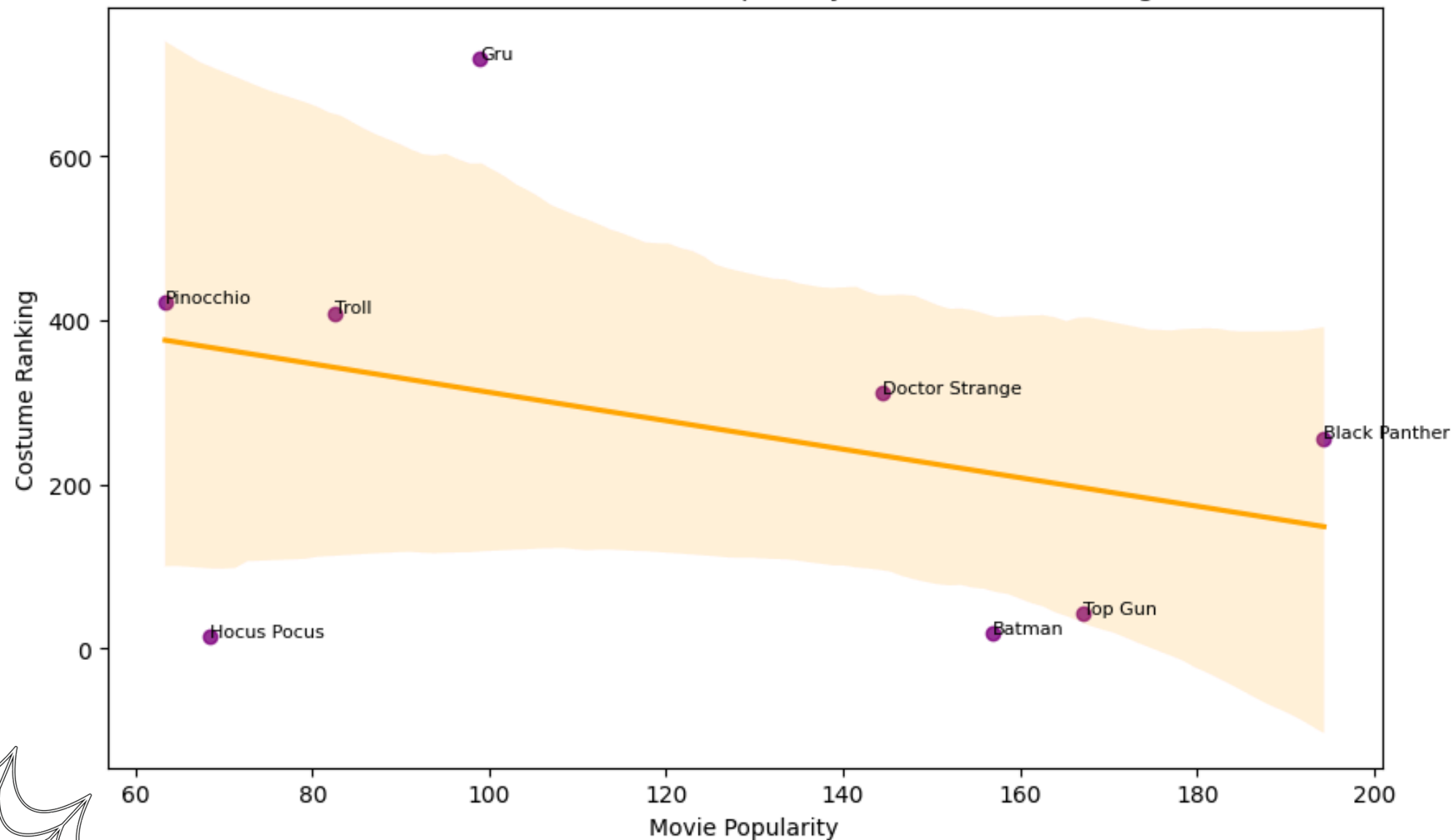
There is a **moderate negative correlation (-0.35)** of the popularity of 2022 movies to the ranking the matched costume received on Googles' Frightgeist list. This points to a relationship where the more popular the movie is the higher ranking the related costume is going to have. **(no surprise)**

However, there is a considerable amount of **variance in the data and a poor fit** to the linear model. This is most likely because of the quality of the data, and how little data points there are (8 movie-related costume matches).

It would be interesting to expand on this relationship with more data points. For example, looking at how the relationship is effected over the **span of 5 years** versus just looking at 2022 or also including TV shows. It would also be interesting to see how popularity of movie characters and their subsequent costumes persist over time. (ie. There was no Spider-man movie in 2022. Instead 2022 was flanked by Spider-man movies in 2021 and 2023, and yet in 2022 "Spider-man" costumes ranked #2 on the Frightgeist list.)



Correlation between Movie Popularity and Costume Ranking



Costume	Ranking	Popularity
Hocus Pocus	14	68.501
Batman	19	156.961
Top Gun	43	167.162
Black Panther	256	194.346
Doctor Strange	311	144.519
Troll	409	82.599
Pinnochio	422	63.372



# METADATA

RangeIndex: 865 entries, 0 to 864

Data columns (total 2 columns):

#	Column	Non-Null Count	Dtype
0	Rankings for Halloween costumes, Based on US searches in September 2022	865 non-null	object
1	Unnamed: 1	865 non-null	object

dtypes: object(2)  
memory usage: 13.6+ KB

Rankings for Halloween costumes,\nBased on US searches in September 2022			Unnamed: 1
0	Costume	Ranking	
1	Witch	1	
2	Spider-Man	2	
3	Dinosaur	3	
4	Stranger Things	4	

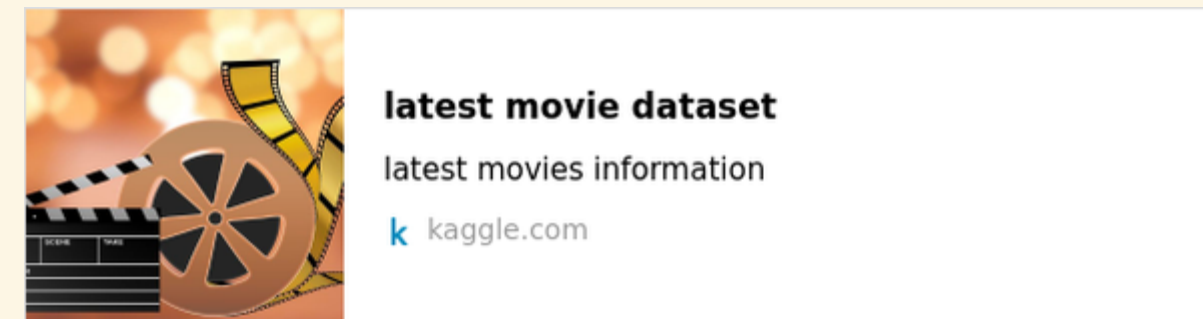
RangeIndex: 10000 entries, 0 to 9999

Data columns (total 9 columns):

#	Column	Non-Null Count	Dtype
0	Unnamed: 0	10000 non-null	int64
1	id	10000 non-null	int64
2	genre_ids	10000 non-null	object
3	title	10000 non-null	object
4	overview	9998 non-null	object
5	popularity	10000 non-null	float64
6	release_date	10000 non-null	object
7	vote_average	10000 non-null	float64
8	vote_count	10000 non-null	int64

dtypes: float64(2), int64(3), object(4)  
memory usage: 703.2+ KB

Unnamed: 0	id	genre_ids	title	overview	popularity	release_date	vote_average	vote_count	
0	0	238	[18, 80]	The Godfather	Spanning the years 1945 to 1955, a chronicle o...	119.438	1972-03-14	8.7	18448
1	1	278	[18, 80]	The Shawshank Redemption	Framed in the 1940s for the double murder of h...	90.415	1994-09-23	8.7	24376
2	2	240	[18, 80]	The Godfather Part II	In the continuing saga of the Corleone crime f...	70.637	1974-12-20	8.6	11144
3	3	424	[18, 36, 10752]	Schindler's List	The true story of how businessman Oskar Schind...	48.096	1993-12-15	8.6	14421
4	4	19404	[35, 18, 10749]	Dilwale Dulhania Le Jayenge	Raj is a rich, carefree, happy-go-lucky second...	26.588	1995-10-20	8.6	4225



Original Data and Data Analysis here: git clone <https://github.com/MayaTalpai/cs131/tree/main/a5>