

# Movie Recommendations and Genome Tagging

One person out of the population labeled Plan 9 from Outer Space as a bad movie, but rated it as a four.

Why is that?

After careful consideration I started thinking about movie recommendations and how one determines when to recommend a bad rated movie.

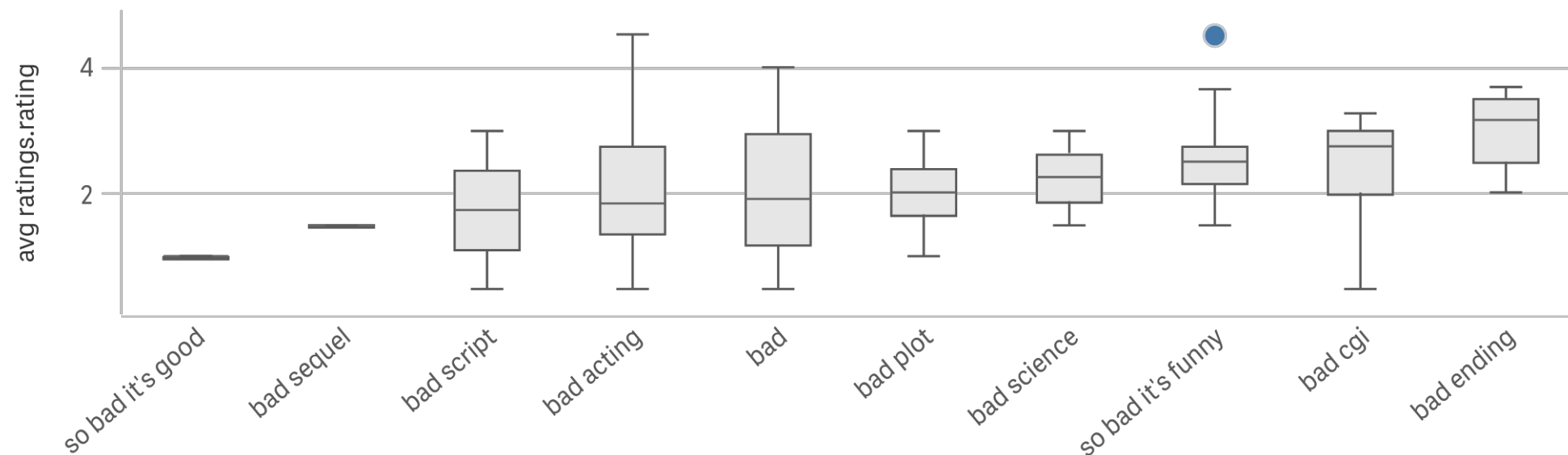
Do we recommend it based on user ratings? Or is there another solution we can use entirely?

## Plan 9 from Outer Space

Sci-Fi

bad  
4

## Tagged Bad Movie Ratings



I pulled out all the tags with a genome rating greater than 0.68% and were provided by users who submitted 3,000 or more reviews.

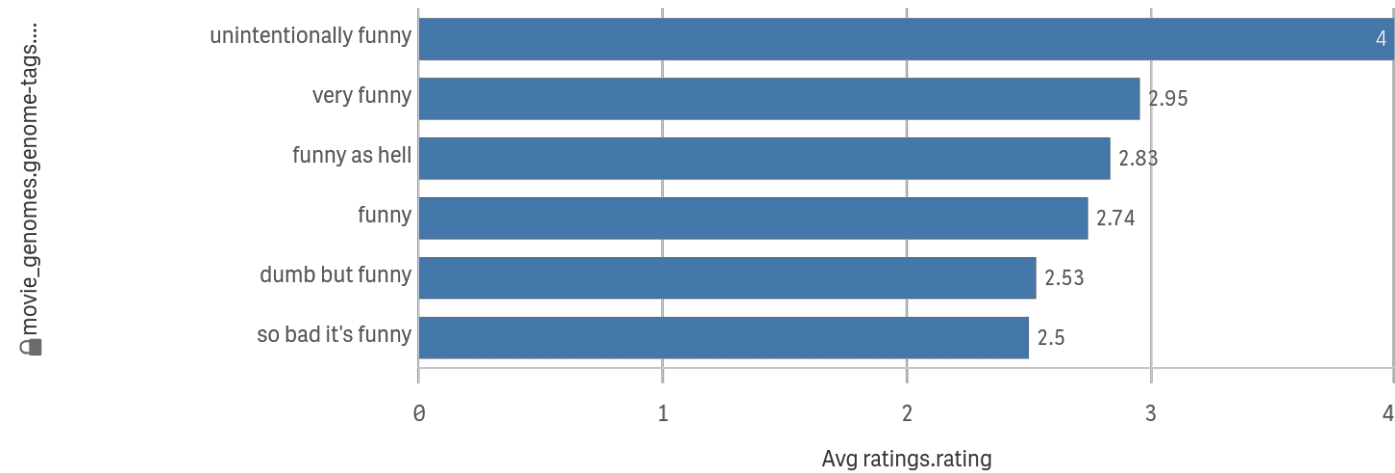
We're looking at the average rating for those movies that contained the word bad. The over all average rating for these users is:

Avg ratings.rating

2.94

Here we change the perspective, if I look at it from a different angle, does the average rating change for movies that were tagged as funny?

Avg(ratings.rating) by movie\_genomes.genome-tags.tag



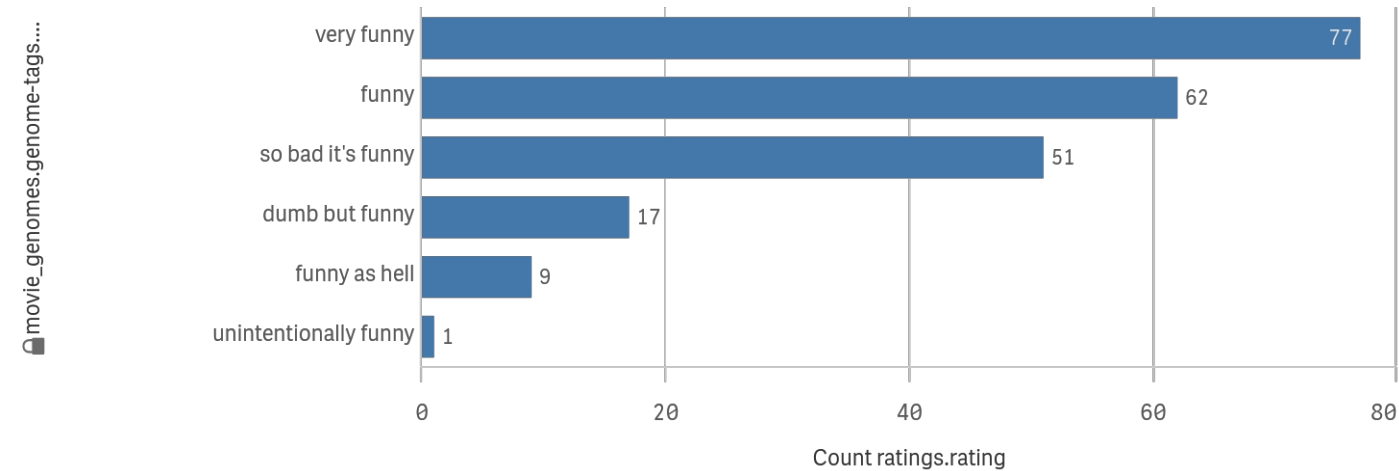
I decided to take a look at three categories.

So bad it's funny, dumb but funny, and unintentionally funny.

A drill down revealed that the most interesting films were tagged as so bad it's funny and unintentionally funny.

Dumb but funny were still just dumb.

Count(ratings.rating) by movie\_genomes.genome-tags.tag



# Must See Movies

movie_genomes.movies.title	Q	avg movie_genomes.relevance	Avg ratings.rating	
Totals		0.69234135	2.5288462	
Chairman of the Board		0.74	2.5	
Man's Best Friend		0.717	2.5	
King Kong vs. Godzilla (Kingukongu tai Gojira)		0.716	2	
Toxic Avenger, The		0.71225	1.5	
Make Them Die Slowly (Cannibal Ferox)		0.7095	3	
Wholly Moses		0.70775	2	
Billy Jack Goes to Washington		0.7075	1.5	
Beyond Borders		0.70725	3.5	
Two Moon Junction		0.7065	1	
Squirm		0.705	3	
Mole People, The		0.7035	2	
The Butterfly Effect 2		0.70225	2.5	
The Boy Next Door		0.70125	1	
Awakening, The		0.6985	0.5	
Doppelganger		0.698	4	
Garbage Pail Kids Movie, The		0.69625	0.5	
Born to Be Wild		0.69475	4.5	
My Baby's Daddy		0.6945	2	
Venom		0.694	4.5	
Space Truckers		0.69275	2	
Wake Up, Ron Burgundy		0.69175	3	
Resident, The		0.69075	1.5	
Ballistic: Ecks vs. Sever		0.6895	2	

Total  
52

# Project Recommendations:

movie_genomes.movies.title	Q	avg movie_genomes.relevance	Avg ratings.rating	
<b>Totals</b>		<b>0.68025</b>	<b>4.5</b>	
Rock-A-Doodle		0.68025	4.5	

I'm curious to see what the data looks like after doing [topic modeling](#) on the genome tags. After doing a little bit of analysis and learning more about the data - drill down a little bit more to determine whether or a prediction algorithm can be generated based on these genome tagging's and user ratings.

A bit of additional labeling will need to be done here. There are some cases where the movies were rated as a 0.5, but they're considered so bad it's funny. Do these users like these kinds of movies? Should we be showing them? Perhaps I need to investigate on the number of 0.5 ratings that their giving.

Mostly - I just need to get an LDA done, run a prediction model, and work from there.

Kalika Kay Curry - OCRUG Hackathon Team 5  
[Movie Lens Dataset](#) - [Full Version](#).