Phase One Project

Author: Alexander Wood

Overview

This project is aimed to provide Microsoft with a few short reccomendations for their burgeoninging new movie studio. While they look to the future with a fresh young face, we took the time to analyze current movie trends to discover diabolically useful suggestions for their new direction. The Microsoft movie executives have nothing to fear with us on the case.

Business Problem

With all their money lying around but no experience whatsoever with making movies, Microsoft stands at a metaphorical crossroads with a metaphorical racecar. This project will provide the blinking lights that direct them to fiery success, rather than inevitable doom. Following this almost divinely inspired guide could (hypothetically) double, triple, or even halve their investment! The shareholders have never been more ambivalent.

Data Understanding

This study uses ratings from IMDB and gross income from from Box Office Mojo to explore several burning questions in the movie industry. IMDB is one of the biggest movie databases in the world, with records of nearly everything to do with movies. Importantly for this study it also has a robust rating system where both professional critics and regular citizens alike can vote for their favourite (or least favourite!) movie. How democratic! Box Office Mojo is an American website that algorithmically tracks box office revenue. Soon they'll be tracking your runaway profits too! All the revenue data in this study comes from their databanks.

Data Preparation

When we were beginning preparations for this study we decided to compare most of our variables to the IMDB rating. While this isn't a direct comparison to gross revenue, the raw girth of the IMDB data dwarfed the BOM revenue data. Of course, we wouldn't take such a drastic step without first confirming that there was some tangible relation between rating and gross revenue, so we put our systems to work and out spawned a chart several hours later and with only some frustration!
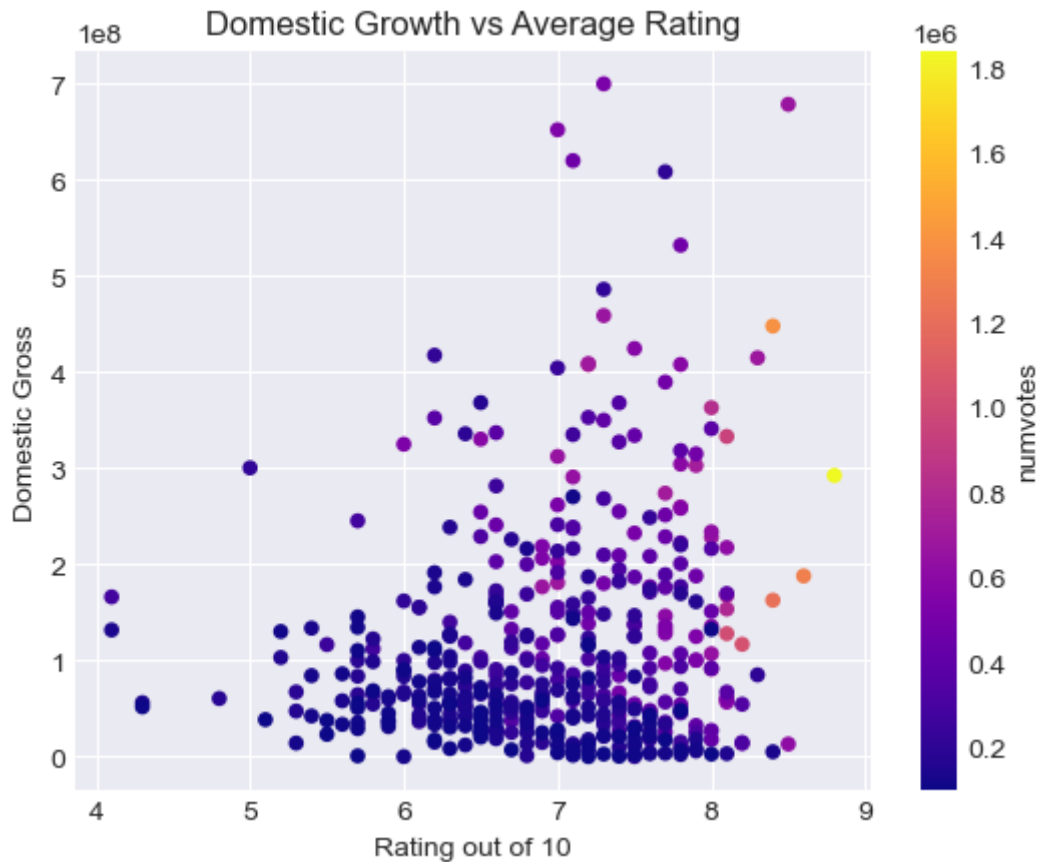
As can be seen in our elegant, perhaps even sublime, graph as rating numbers go up, so too do profits. Not only that, but the highest rated movies also have the most user engagement in the form of more ratings. It's like free advertisement that you paid for!

Data Modeling

We set out to immediately determine the three most vital aspects of what makes the perfect movie. What could these be? Passion? Drive? The constant need for external validation? Of course not! These three aspects are the genre, the director, and the length of the movie. These were chosen in an extremely non-arbritray way that we assure you we have fantastic reasons for, and you can't legally prove otherwise. Some claim that there is no such thing as a movie that everyone will enjoy, but we will prove them wrong by appealing to the lowest possible denominator! Using math!

Average Rating of Genres
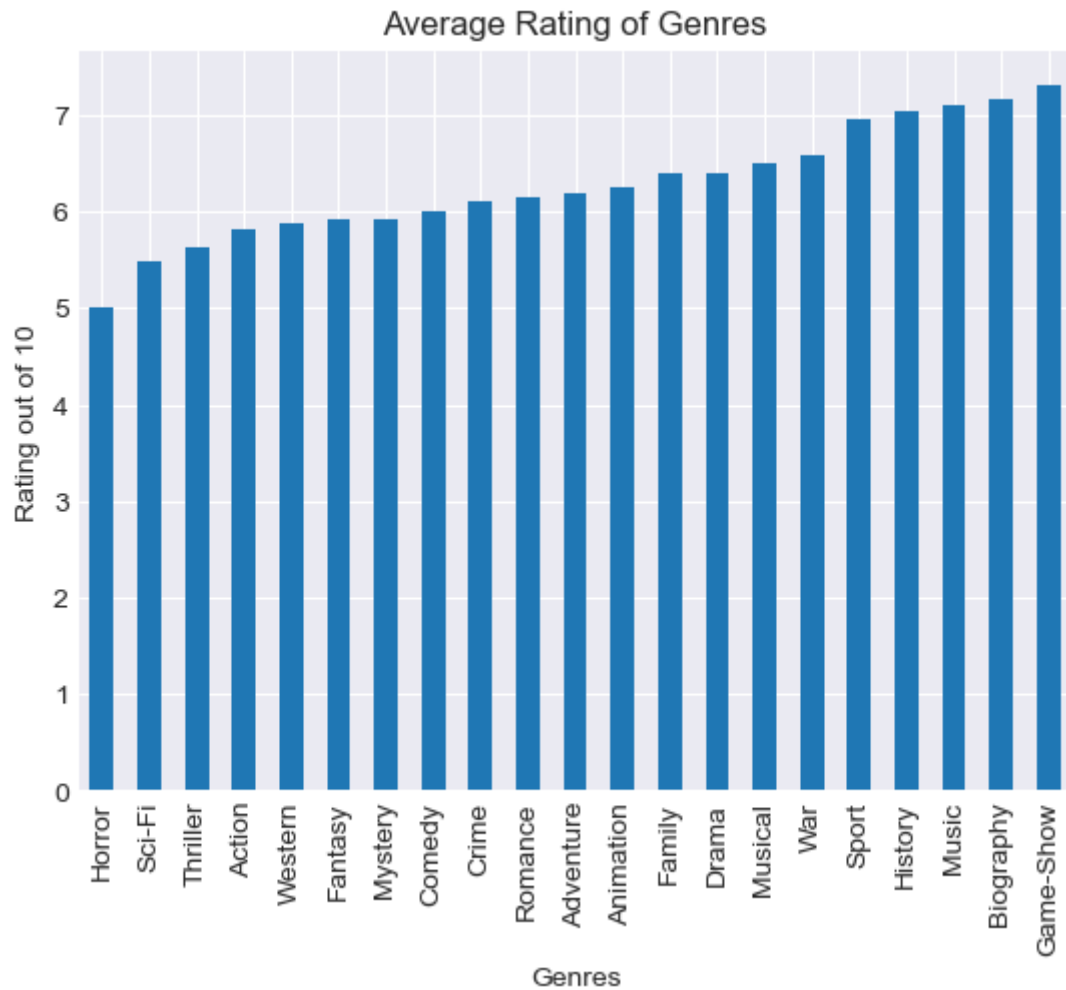
By carefully scrying into the hearts of the worldwide populace with our magic computers we were successfully able to determine that the most popular type of movie among the people is the ever festive game show! Coming in a close second is the long-time favourite of the youth, biographies. It may alarm you that both of these categories aren't generally considered movies at all. It's true! We probably should have removed them. We can. But we didn't! But rather than thinking of this as a blunder or some sort of flaw, consider instead the rampant possibilies for innovation. Armed with this data you can make the most popular, and profitable, game-show-biography-music-video the world has ever seen! But who could possibly direct such a masterpiece of questionable-fiction. Well, I am glad you asked.

Director	Average Rating	Number of Votes
Quentin Tarantino	8.675000	4.138442e+05
Bob Persichetti	8.500000	2.108690e+05
Rodney Rothman	8.500000	2.108690e+05
Christopher Nolan	8.425000	1.248687e+06
Adrian Molina	8.400000	2.771940e+05
Lee Unkrich	8.350000	4.797060e+05
Anthony Russo	8.225000	5.904550e+05
Ronnie Del Carmen	8.200000	5.361810e+05
Malik Bendjelloul	8.200000	5.865400e+04
Mel Gibson	8.100000	3.669040e+05
Damián Szifron	8.100000	1.511230e+05
Spike Jonze	8.050000	2.338690e+05
Rajkumar Hirani	8.050000	8.924250e+04
Jared Bush	8.000000	3.834460e+05
Banksy	8.000000	5.872100e+04
Stephen Chbosky	8.000000	2.671515e+05
Tim Miller	8.000000	8.208470e+05
Damien Chazelle	7.966667	3.909983e+05
Wes Anderson	7.933333	3.445330e+05
Byron Howard	7.900000	3.749060e+05

Look at those numbers! Look at those ratings! The clear candidate to lead your bright new venture is Quentin Tarantino! We have here displayed the 20 most highly rated directors in all of IMDB, disregarding the miserably decrepit peasants who couldn't even muster a measly 50,000 ratings across all their movies. Now you may be wondering how Mr. Tarantino's distinctive style would gel with our bold concept of a game-show based movie and frankly, I don't know. I don't even know who half these people are! But math has never lied to me, so I can only presume that under Quentin's comsopolitan leadership this terrific venture will make record profits. But of course there is one more question that must be answered, how long must we inflict our non-captive (for legal reasons) audience to our brand new Experience?
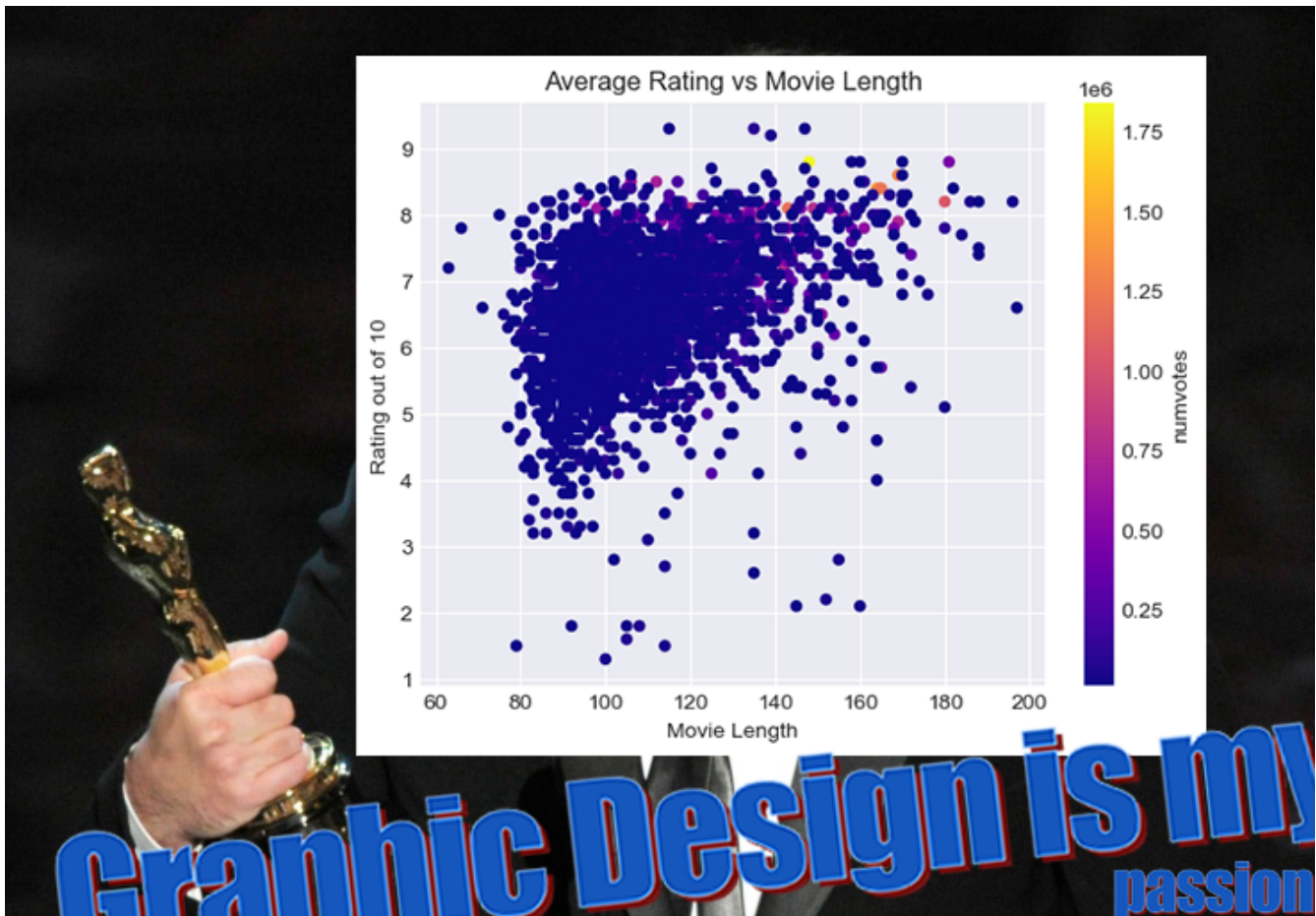
This graph shows a very interesting trend. The longer the movie, the better the rating! That means for maximum gaurantee for success you should logically make your movie as long as humanly possible. Of course, if you were to interpret that selfsame data in a more logical way, trying to put in an optimal amount of work for reward, you may note that there appears to be an optimal area in around the 120 minute mark. Bollocks to that, I say! Go big or go home. If you aren't on the bus then the bus is on you! Let's make some money!

Evaluation

Now we arrive, sadly, to the part of this study where we must look inwards and examine our study for any failures. Where did we go wrong? Can art not be constructed by the machinations of our constantly evolving fancy calculators? Will machines ever be able to understand the human heart? Of course not! We set out to determine three actionable insights for Microsofts sudden and mysterious venture into the movie industry and by jove we delivered! Your three-step recipe for ultimate success is as follows: Hire Quentin Tarantino. Have him direct a game-show themed movie. Make the experience as long as possible to maximize the enjoyment of your loving audience. That's it! The perfect movie, as determined by flawless math.

Conclusions

Once again math has proven that it can do what no mortal can. Perfect in it's angularity, sublime in its multiplicity. Some would say that this study has no flaws, and it's true, I am quite smart. But I am also humble in my intelligence, and so I present to you some future directions for your fortuitous future. In future studies it may be worth considering the genres the directors you are looking to hire specialise in. More study would be needed to create a chart showing the leaders of each genre. So too it might be

worth further study into how genre relates to the length of a movie. Is it possible that a short action film may be more popular than that same movie arbitrarily streched to meet some arbitrary length enforced by a man blindly putting numbers into a computer? We may never know. I certainly won't ever know. We've reached the end! Farewell!

For more details on how we came to our ~~disastrous~~ conclusions please check our work in the attached [document (files/project.ipynb)](files/project.ipynb).