

Module I: Microsoft Movies

Isiah Cruz

April 2020

Hello to the Microsoft Business Team, the Data Science Team is very excited to present to you today about our potential entry into the film industry! For those of you who I have not met yet, my name is Isiah and we are going to jump right into it!

TABLE OF CONTENTS

- | | |
|----|-------------------|
| 01 | PROBLEM STATEMENT |
| 02 | BUSINESS VALUE |
| 03 | METHODOLOGY |
| 04 | FINDINGS |

Here is an outline of the presentation, we will start with the problem statement (what are we trying to solve?), then we will look at the business value of the presentation (what does this mean for you?), and then we will close with the methodology & findings which put our analyses into context.

PROBLEM STATEMENT

1



MARKET

Our close competitors have started making **original content** thanks to proprietary film studios

2



MOVIE STUDIO

Out of our comfort zone so we need to do our **due diligence** to maintain high quality product

3



BOX OFFICE

Data analysis to explore **which films are currently doing best** in the box office

Just to recap the problem we are attempting to solve here. The reality is that some of our very close competitors like Amazon and Microsoft have entered the film industry with proprietary film studios. Therefore, it is important that we acknowledge our entry into an industry that is not necessarily in our wheelhouse. Furthermore, it is critical that we do our due diligence in entering this new industry that way we uphold the high quality product that clients & customers alike have come to expect from Microsoft. Most importantly, we are specifically looking at which films are currently doing best in the box office to better inform us on how to most effectively enter the industry with our first movie.

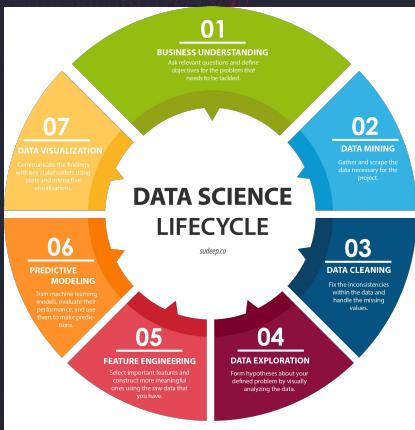
BUSINESS VALUE

- **LAUNCH**
- **RECENT SUCCESSES**
- **FILM CREATION**



Of course, the first impression is often the most important impression. So we really want to focus on maximizing our launch as much as possible. According to our analysis, there are 2 ways that we can do this: the popularity of our movies (indicated by the buzz that our movies generate, the amount of tickets we sell, and the conversations that result from our first movie) and the quality of our movies (including the ratings that our first movie receives from audiences and critics alike). In order to learn more about how to maximize our launch, we studied Recent Successes in the Box Office (so we considered movies only from 2015 on given the rapid innovation in the film industry) and we studied Film Creation (so the different factors that go into creating a movie that correlate with box office success).

METHODOLOGY



Business Understanding:

- Context of the business
- Available data sets (IMDB, Rotten Tomatoes, etc.)

Data Cleaning:

- Standardized data types
- Assumptions for empty data (median, ffill, etc)
- Removing unnecessary columns

Data Exploration:

- Searched for value in the data
- Combined and built data sets (joins, merges, etc)
- Sorted and arranged the data (top 5, top 100, etc)
- Customization (49+ votes only, 2015 or later only)

Data Visualization:

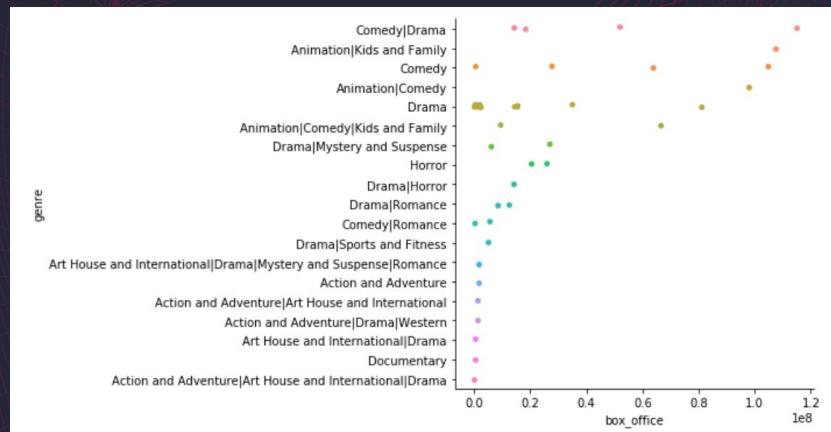
- Visualized findings in graphical form

Toolkits:

- Pandas (package for data manipulation and analysis)
- Numpy (package for scientific computing)
- Matplotlib (package for plotting)
- Seaborn (package for statistical graphics)

For our non-Data Scientists in the room, here a sneak peek of what we looked at our process. We specifically touched on some of the famous stages of the Data Science Lifecycle. Starting with Business Understanding, we contextualized ourselves with the state of our business at this current moment and we took stock of the available datasets out there (honing in on data from Rotten Tomatoes and IMDB). We then moves onto Data Cleaning, where we standardized our data and removes data that was unnecessary for the purposes of this analysis. Next, we explored the data, constructed new datasets by arranging, joining, and customizing the original data (i.e. we determined that any movie with 49+ votes would be considered for our analysis since a movie with 800K votes and another with 8 votes should not be judged the same, therefore we had to create a critical mass for the number of votes a movie had to receive). Finally, we visualized this data graphically and we used the above toolkits in case anyone wants to look a bit more into it.

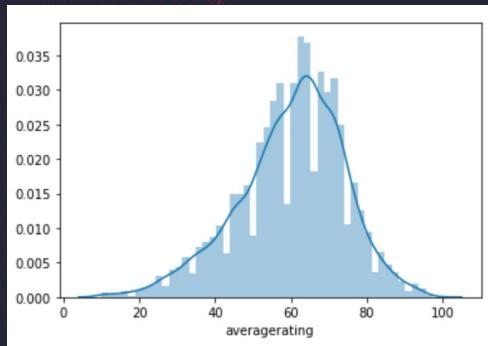
FINDINGS I



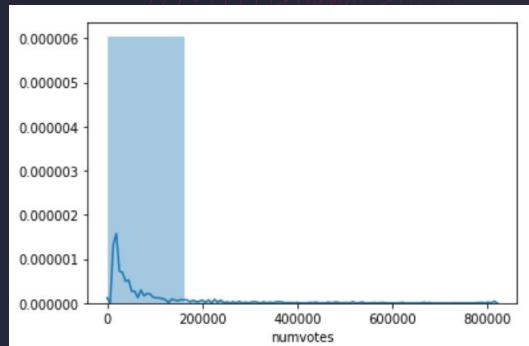
Genre types have the highest correlation to box office success (as evidenced by concentration distributions)

Our first finding was that Genre type is the greatest predictor of box office success. That is, of all the variables that go into film creation (cast, runtime, release month, etc), Genre has the highest concentration distribution when looking at the Top 100 movies. Of course, this prompted us to look further into specific genres and movie success.

FINDINGS 2



Average Rating = Quality

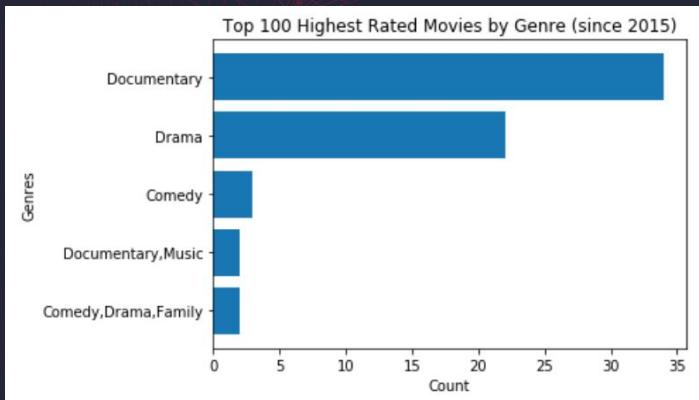


Number of Votes = Popularity

**Highly rated movies are few and far between
Very popular movies are few and far between**

Upon determining that genre is the greatest predictor of box office success, we wanted to better define box office success. We decide on Average audience rating as a good proxy for quality and Number of audience votes as a good proxy for popularity. Our conclusion: making a high quality and high popularity film is no walk in the park.

FINDINGS 3



Top 5:

1. Gini Helida Kathe (99 rating)
2. Once Upon a Time in Hollywood (97 rating)
3. Eghantham (97 rating)
4. Dosed (97 rating)
5. Ekvtime: Man of God (96 rating)

Documentary or Drama Movies

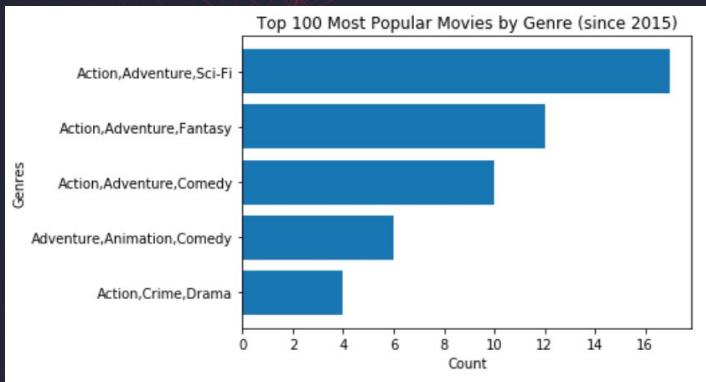
Documentary and Drama films account for over 56% of the Top 100 Highest Rated Movies since 2015

Takeaway

Create a Documentary or Drama film to maximize **audience ratings**

We then investigated the Top 100 Highest Rated Movies by Genre and found that 56%+ of these ranked movies were either Documentary or Drama films. Therefore, if our goal is to maximize audience ratings and the quality of our movie, then the suggestion is that we put out a Documentary film.

FINDINGS 4



Top 5:

1. Deadpool (820,847 votes)
2. Star Wars: Episode VII - The Force Awakens (784,780 votes)
3. Mad Max: Fury Road (780,910 votes)
4. The Martian (680,116 votes)
5. Avengers: Infinity War (670,926 votes)

Action + Adventure + X Movies

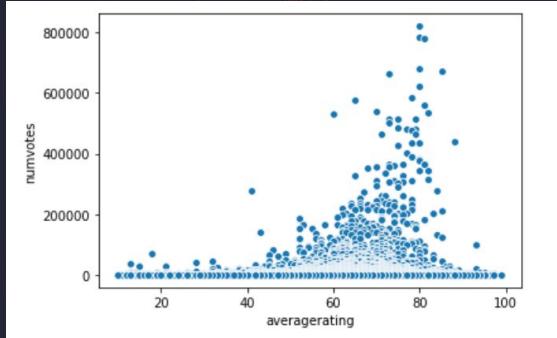
Action, Adventure and a mix of either Sci-fi, Fantasy, or Comedy account for over 40% of the Top 100 Most Popular Movies since 2015

Takeaway

Create an Action, Adventure, & Sci-Fi movie if optimizing for popularity and buzz

We also looked at the Top 100 Most Popular Movies since 2015 and found that Action/Adventure and some mix of Sci-fi, Fantasy, Or Comedy accounted for 40%+ of the most popular movies. Therefore, our recommendation is to create an Action, Adventure & Sci-Fi movies like Deadpool or Star Wars if we want to maximize popularity.

FINDINGS 5

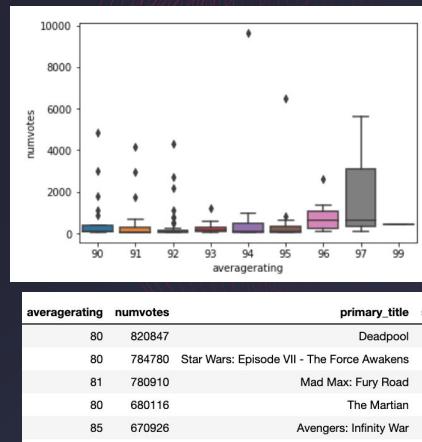


Positive Correlation between Average Rating & Number of Votes

The correlation coefficient between these two variables is **positive at 0.1012**, telling us that quality should not be sacrificed in chasing the box office hit.

Takeaway

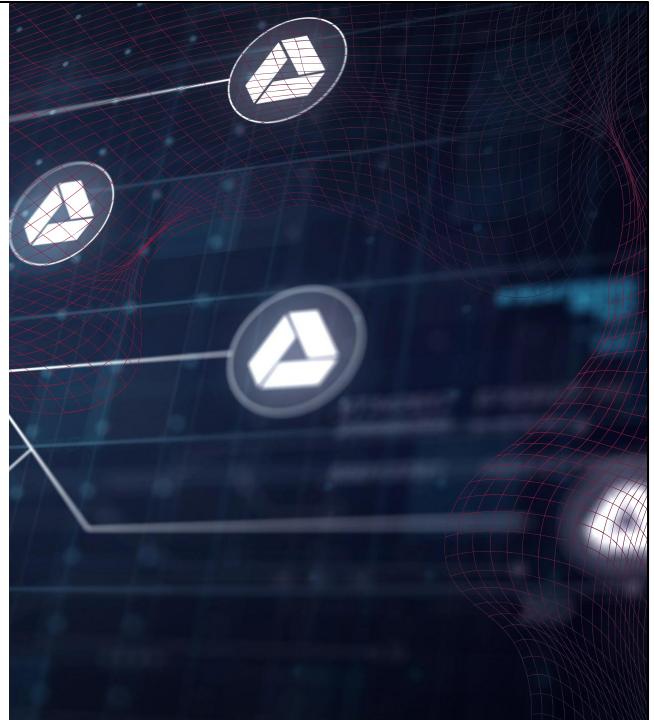
Aim to **create a highly rated movie** since this has shown to increase the probability of a box office hit, however, we also see that extremely high-rated movies tend to not be very popular (with **movies rated between 80-90** doing the best at the box office).



Our final finding is that there is a positive correlation between Average Rating & Number of Votes. We specifically measured the correlation coefficient and as you can see above, the number of votes (our proxy for popularity) increases as the average rating (our proxy for quality) increases. What this suggests is that quality should **not** be sacrificed in chasing the box office hit. Still, this does not necessarily mean that we need to create the greatest movie ever, it just means that we should aim for the rating sweet spot between 80-90 because these movies tend to be the most popular as we can see above.

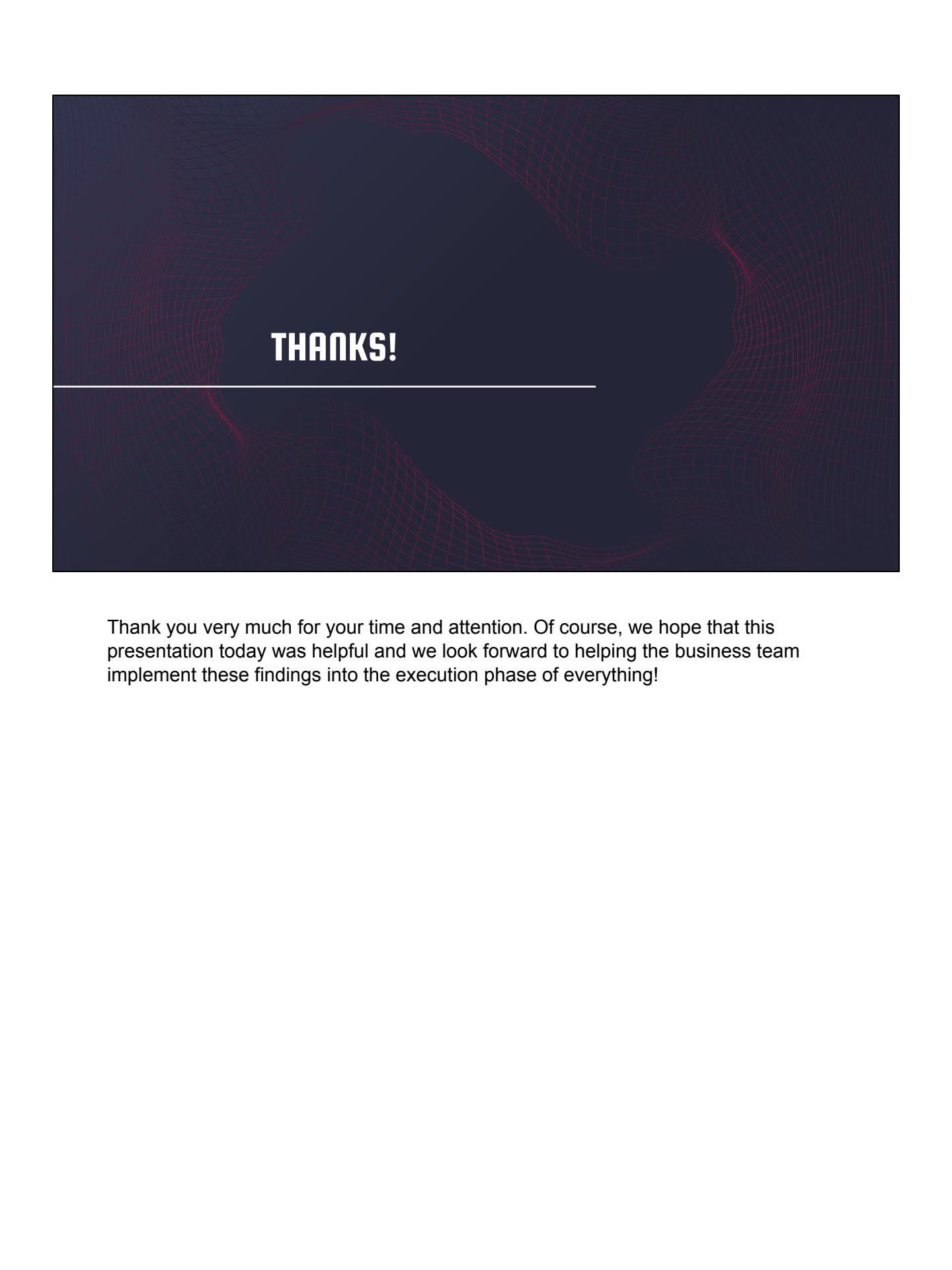
FUTURE WORK

1. Leveraging our legacy technology and proprietary data to create films
2. Exploring the viability and potential value in creating a streaming platform to house our films
3. Differentiating ourselves from competitors in what has become a saturated market



Looking at future work, we first suggest exploring ways in which we can leverage our technology to bolster our entry point into the film industry. For instance, perhaps we can apply some of our strengths in the software space to create top-notch Computer-Generated Imagery (CGI) for our first Sci-Fi movie. Next, we would advise looking into the viability of creating our own streaming platforms to distribute our films in-house, something that most players in the space are doing already. Finally, we suggest really thinking about how we are going to differentiate ourselves in what has become a saturated market (Netflix, Apple TV+, Amazon Prime Video, Hulu, Disney+, HBO MAX, etc).

Based on our data science, our final suggestion is to make a big splash in the movie industry with a high quality Action/Adventure/Sci-Fi film that blends popularity and quality in its release.

A dark blue rectangular background featuring a subtle, abstract red wireframe mesh pattern that forms a large, flowing, organic shape across the surface.

THANKS!

Thank you very much for your time and attention. Of course, we hope that this presentation today was helpful and we look forward to helping the business team implement these findings into the execution phase of everything!