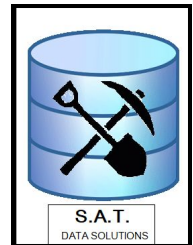




## Movie Data Analysis and Recommendations

By: Sachin Naik, Thomas Cornett and Amarnath Kuppannan



1

Hello everyone. We are S.A.T. Data Solutions hired by Microsoft to do data analysis on the movie industry and give recommendations.

# Contents

- Project overview
- The data given to us
- Where we found data to use
- Why we looked for this data
- What we looked for in the data
- Our findings
- Recommendations
- Future work
- Questions



2

- Project Overview
  - Our requirements:
    - Microsoft wants to enter the movie industry and we are here to give them insights on the best performing movies and more
- The data we have
  - The data that is currently given to us within the project
- Where we found data
  - Where we went to scrape for more data
  - We wanted the most recent data available to us
- Why we looked for this data
  - We brainstormed questions which would decide and have direct financial impacts to Microsoft
  - We wanted to gain insights to back up our recommendations
- What we found in the data
  - Popular movies over time
  - Budget analysis
  - Profit analysis
  - Runtime analysis
  - Gross Income
- Recommendations
  - What we believe that you should do
  - Based on the insights we derived from the data
- Future Work
- Questions

# Project Overview

- Microsoft is creating a new movie studio
- We were hired to do data analysis on the movie industry
- Find insights in the movie industry data and give recommendations

## Project Overview:

- Microsoft is creating a new movie studio
- We were hired to do data analysis on the movie industry
- Find insights in the movie industry data and give recommendations

## The Data Given To Us

- IMDb



- Rotten Tomatoes



- The Numbers



- The Movie Database



### The Data Given to Us:

We were given some data to work with mainly from four websites. We utilized some of the data given to us as well as acquiring fresh data to work with.

## Where We Found Data

- Our decision: Gather most recent movie data from IMDb
- Utilize budget data given to us



IMDb Charts

### Most Popular Movies

As determined by IMDb Users

Showing 100 Titles

Sort by: IMDb Rating

Rank & Title	IMDb Rating	Your Rating
 <b>The Shawshank Redemption</b> (1994) 34 (★ 8)	★ 9.3	☆
 <b>The Godfather</b> (1972) 54 (★ 1)	★ 9.2	☆
 <b>Hamilton</b> (2020) 1 (★ 1)	★ 9.0	☆
 <b>The Dark Knight</b> (2008) 67 (★ 5)	★ 9.0	☆

Where we found data:

We decided to stick with IMDb for multiple reasons. We also gathered data from it and merged that with the budget data given to us.

# Why We Looked For This Data

- We wanted the most recent information
- IMDb is trusted, organized and credible
- IMDb has many data points



So why did we stick with IMDb?

- We wanted the most recent information
  - IMDb is updated daily
- IMDb is trusted, organized and credible
  - IMDb has been around since 1990
  - It is over 29 years old
- IMDb has many data points
  - Over 40,000+ including TV shows!

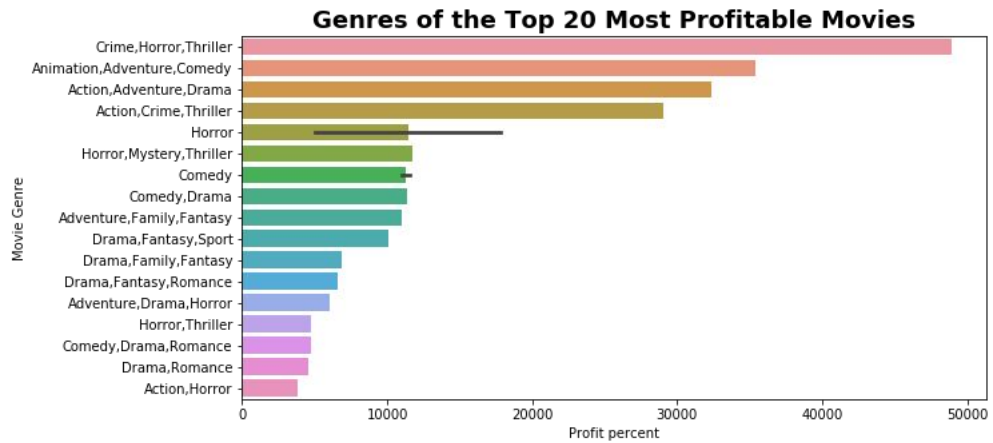
## What We Looked For In the Data

- What genre of movies will yield the best returns?
- What is the ideal runtime?
- What is the current trend of profit returns for the top genres?
- What are the popular genres?
- What is the expected budget and gross income?

7

- Questions
  - Initially, we came up with questions to ask before diving into the data to see what we find
  - We looked at what will have direct financial impacts on studio decisions for Microsoft

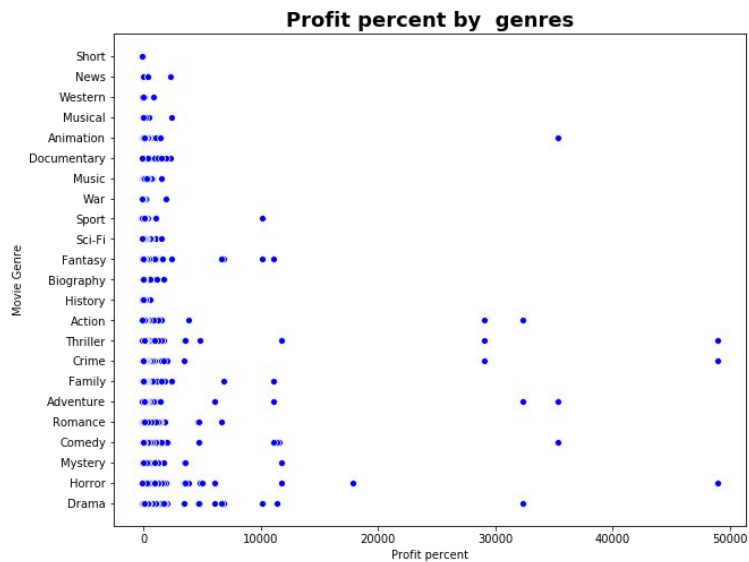
## Our Findings



So first we broke down into which genre combinations were the most profitable. We see here that “crime, horror, thriller” is the most profitable followed by “action, adventure, comedy.”

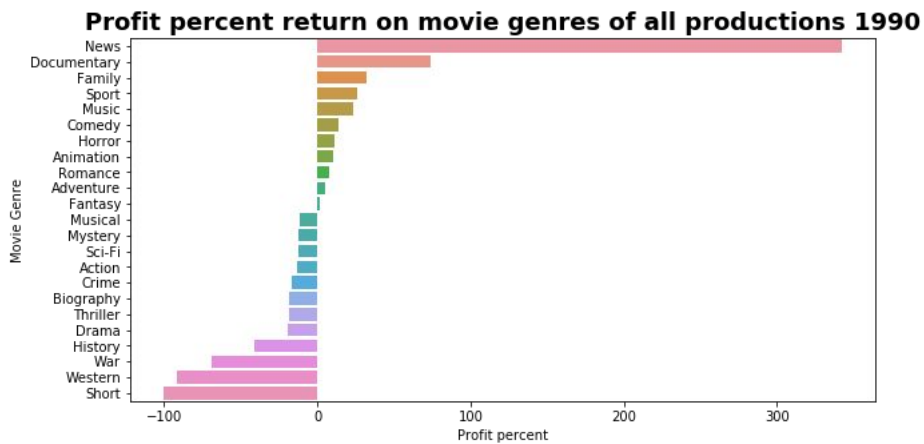


# Our Findings



We split up the genre clusters and looked at the profit percent return in every genre. We noticed that almost every genre has outliers. Because of these outliers, we decided to consider to using median values to estimate central tendency of the data.

# Our Findings

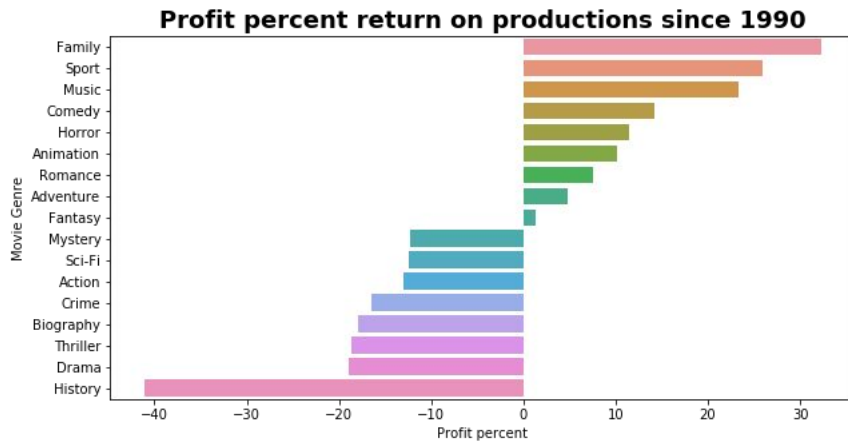


Here we looked at the profit percent returns on single genres.

We noticed here they might have been some skew in the data.

We decided here to narrow down our genres into genres that have at least 30 movies.

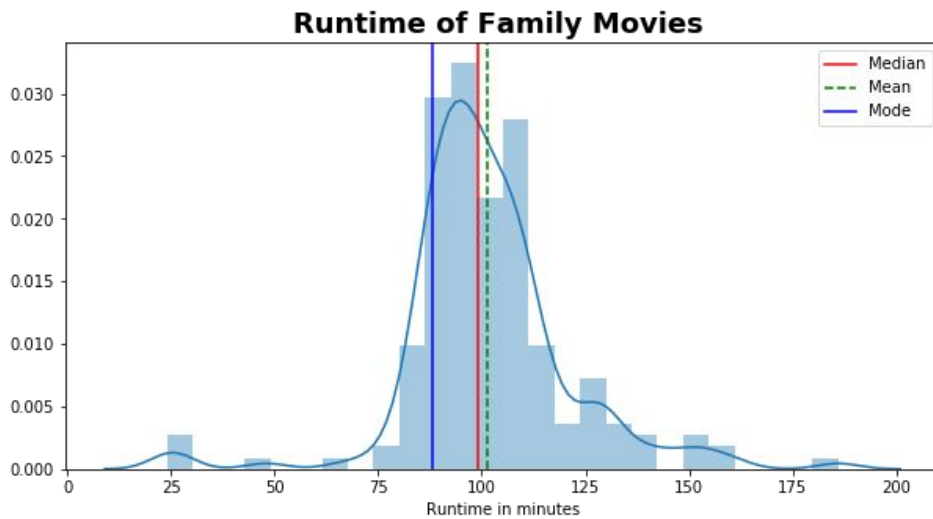
# Our Findings



11

This is after we did that cleanup where the genres should all have at least 30 movies. From this we see that family, sport, and music genres are the top 3 in positive profit percentage

## Our Findings



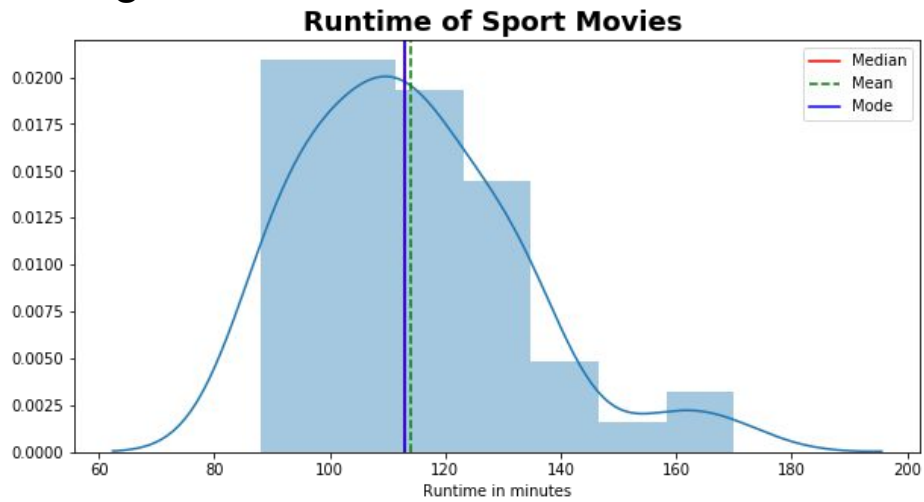
12

After looking at what our top 3 genres are, we are going to look for what the runtime should be for them.

Here, we are looking at the runtime analysis of family movies.

We see that the median here is 99 minutes.

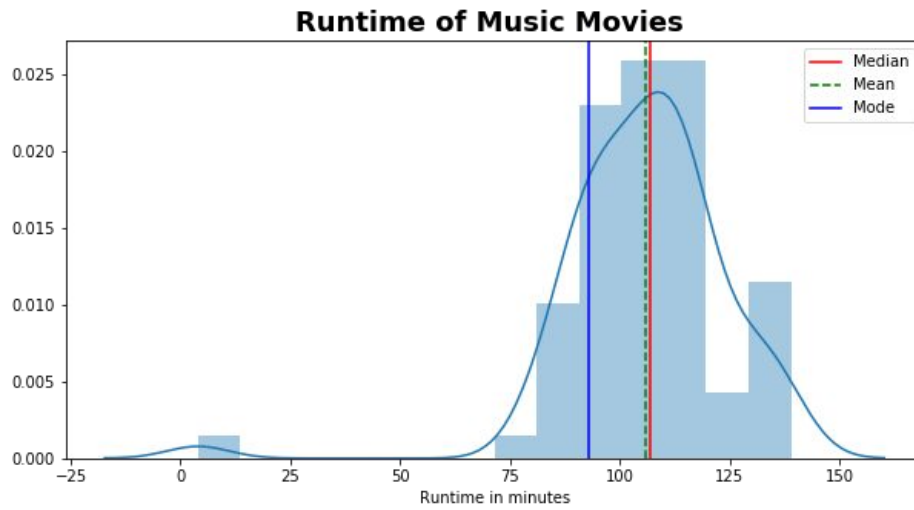
## Our Findings



13

Here is the runtime analysis for sports genre. We see that the median runtime here is 113 minutes.

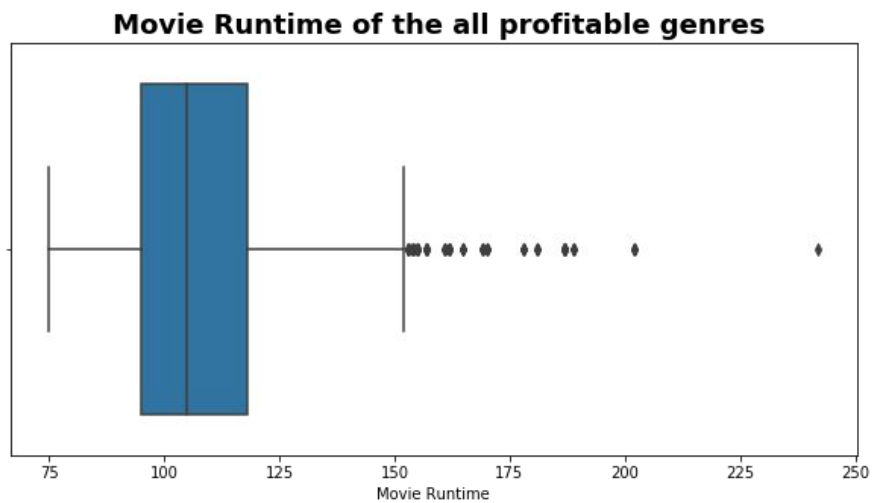
## Our Findings



14

Here is the runtime analysis of music movies. We see a median of 107 minutes for this genre.

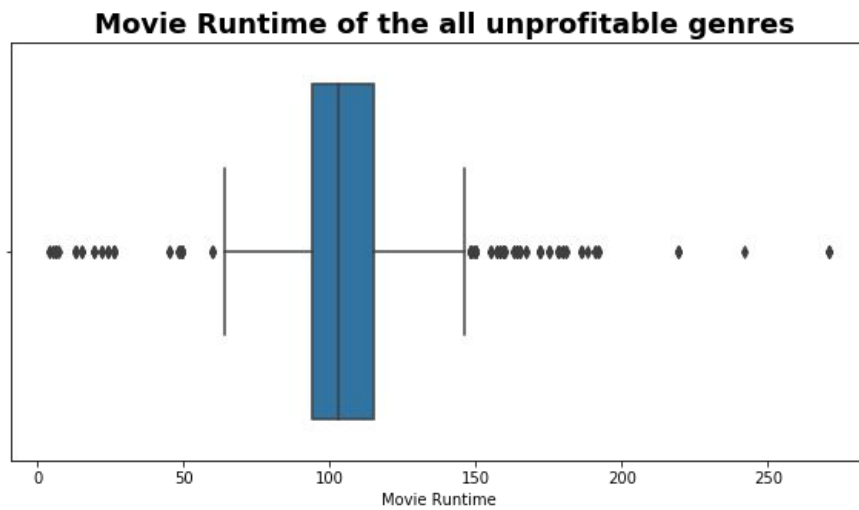
## Our Findings



15

Now we analyze the movie runtime of all the profitable genres to see what difference there is in the runtime. We see a median runtime here of 105 minutes with the outliers.

## Our Findings

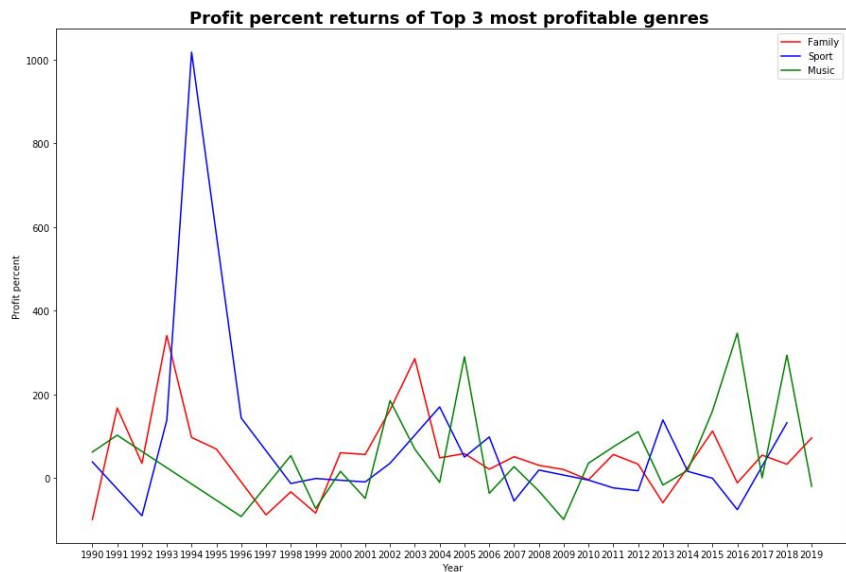


16

Additionally, we took a look at the runtime of all the unprofitable genres to see if runtime had an effect. We saw here a median runtime of 103 minutes which is not all too different.



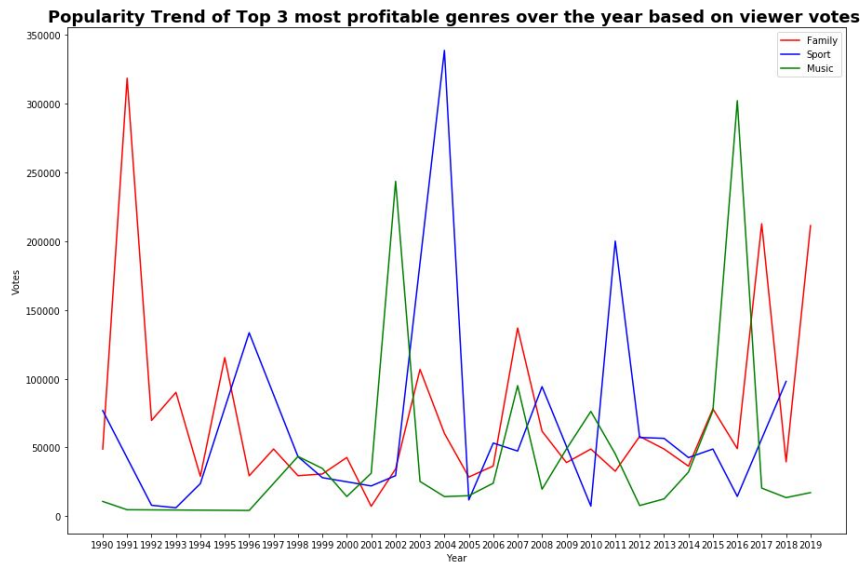
# Our Findings



17

Here we look at the release year of the genre of the movie vs. the year of release. We see here there is a spike in sports movie profits around the mid 1990s.

# Our Findings



18

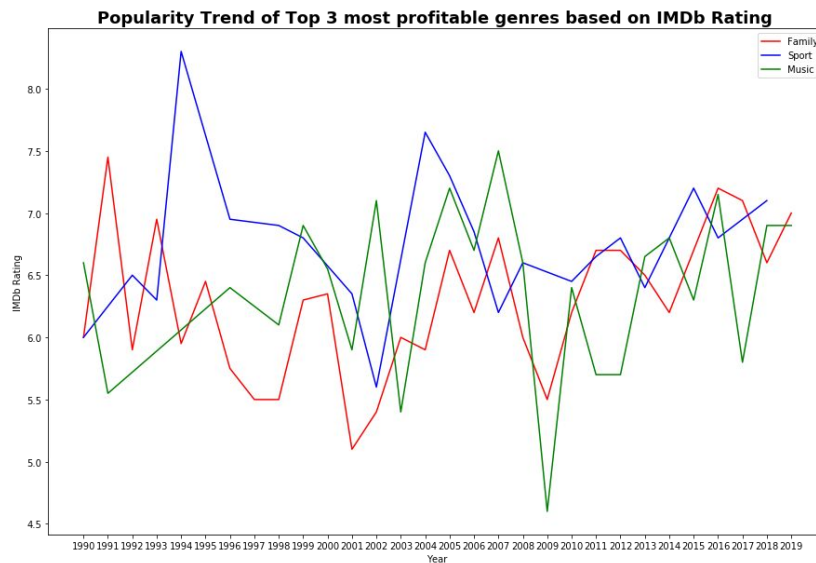
We also looked at the popularity of the genres over time. We noticed each has a spike in popularity during different time periods.

Family genre had a spike early 1990s.

Sports had highest spike early 2000s.

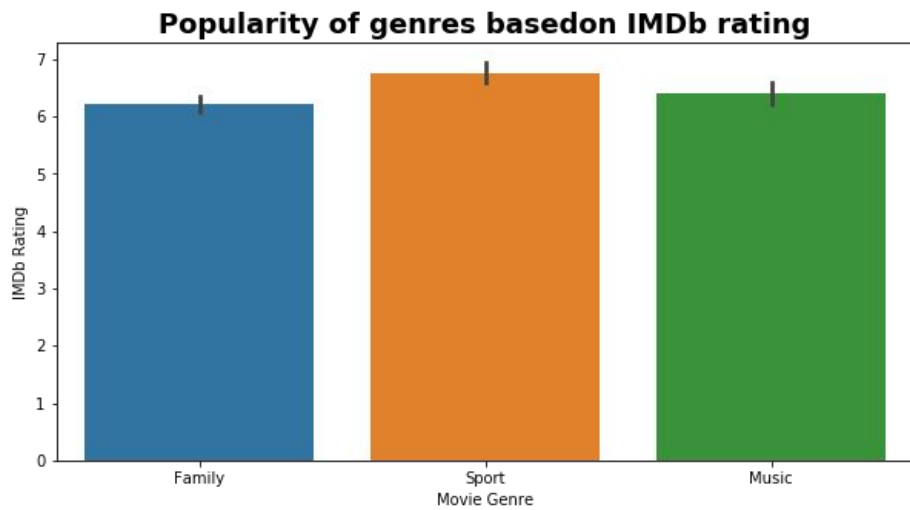
Music had lesser spikes with the highest only a few years ago.

# Our Findings



We also looked at IMDb rating over time for these three genres. We did not see much of a difference here. Each genre had fluctuations.

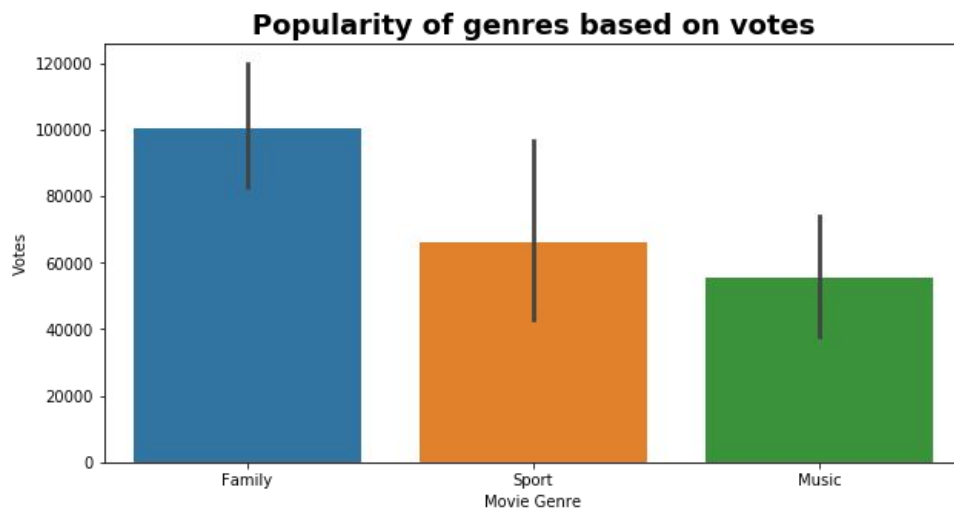
## Our Findings



20

We looked at the popularity of the genre on just IMDb ratings and we see that sports genre is ranked the highest.

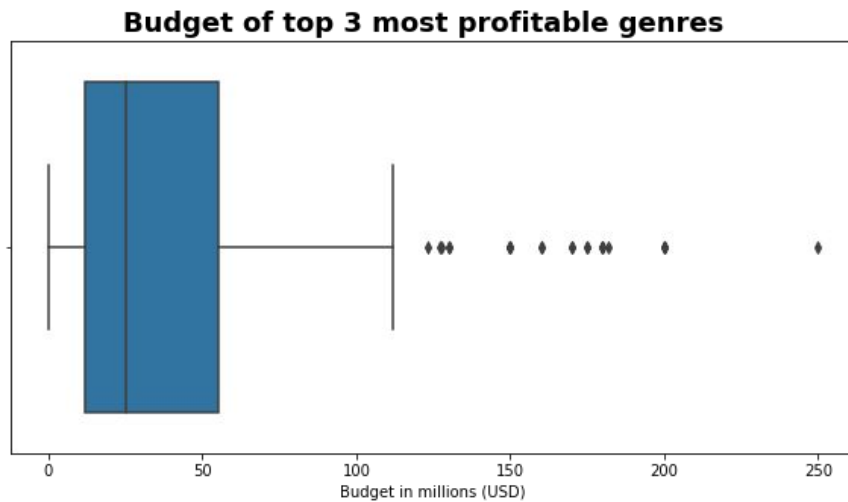
## Our Findings



21

We also looked at popularity of the genre based on votes. We see here that family is most popular based on individual votes.

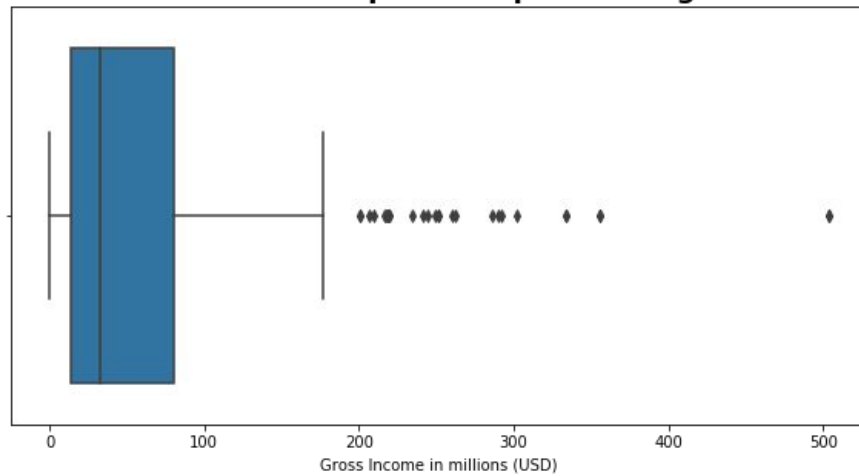
## Our Findings



Here we saw the budget of the top 3 most profitable genres. The median here is \$25 million.

## Our Findings

**Gross Income of top 3 most profitable genres**



This boxplot is to show the gross income of the top 3 most profitable genres. We see there are many outliers. The median gross income is around \$32 million.

## Recommendations

- Family, Music and Sports are the most profitable genres
- History, Drama, and Thriller are the least profitable genres
- Notably Sci-fi has been a popular overall genre but not very profitable

24

After cleaning up our data and only focusing on the genres with more than 30 movies, we saw that there were top 3 most profitable genres and bottom 3 least profitable genres.

We also saw that sci-fi is a popular genre but not very profitable.



# Recommendations

- Runtimes
  - Family - 99 minutes
  - Music - 107 minutes
  - Sports - 112 minutes
- However, not much of a difference in profit vs. runtime

25

We analyzed the runtimes for the top 3 profitable movies and recommend runtimes based on the median numbers we saw.

Although we have runtimes analyzed, we saw that there was not much of a difference between runtimes of profitable and unprofitable movie genres.

Every second of movie production costs money. We recommend focusing on the quality of the content rather than worrying about run time.

## Recommendations

- Focus on family and sports genres
- Family and sports genres both show recent upward trend
- Music shows downward trend

26

Based on the profit percent returns over release year of movie, we saw different trends emerge.

Family and sports look to be the best choices as they are a top 3 genre as well as trending upwards in the recent years.

## Recommendations

- Per IMDb ratings, sports genre is most popular
- Per individual votes, family genre is most popular

This also helps us show that family and sports are the best choices. They are both popular and profitable.

## Recommendations

- \$25 million budget is a good starting point
- Expect a gross income of \$32 million

Based on the analysis for budget and gross income these are our recommendations.

\$25 million should be a good budget to start with and there is a median gross income of \$32 million that can be expected.

# Recommendations

Overall:

- Start off with family or sports genre
  - 99 minute runtime for family movies
  - 112 minute runtime for sports movies

29

Overall, our final recommendation is to start off with family or sports genres with the following runtimes.

These genres are both trending and popular over time.

They are also profitable and should be good starting points for Microsoft's movie studio.

# Future Work

- Movie Directors
- Trends within movie franchises
- Trends within top production studios
- Movies based on other media
  - TV Shows
  - Video Games
  - Books
- Movie release timing
  - Month
  - Day of the week

30

- Movie Directors
  - We want to know who is the best talent we can hire
- Trends within movie franchises
  - There are certain movie franchises which do really well
- Trends within top production studios
  - For example, Pixar focuses solely on animation
- Movies based on other media - We think that there might be trends worth looking over here
  - TV Shows
  - Video Games
  - Books
- Movie release timing - We want to know if holidays are good days to release the movie
  - Month
  - Day of the week

Questions?



31

Questions?