Disney Movie Success

Presented By:
Andres Gonzalez
Christina Lopez
Katya Mora

Table of Contents

The Data Set

Two data files:

- Disney*movies*total*gross.csv*
  - Contains data on the total gross of Walt Disney Studios movies from 1937 to 2016
- Disneyrevenue_1991-2016
  - Contains data on the revenue of the Walt Disney Company from 1991 to 2016

Subset of Data to Analyze:

- Merged data set of movies and revenue
- Years 1992 to 2016

Source : https://www.kaggle.com/datasets/thedevastator/disney-character-success-a-comprehensive-analysi?resource=download

Variables For Each Individual Movie

Response:

- Gross: Gross inflation-adjusted revenue per movie (Millions of US $)

Variables:

- Year: Year of movie release date
- Genre: The genre of the movie
- Rating: The MPAA rating for the movie
- Movie Title: The title of the movie

Variables for Yearly Summary

Responses:

- Revenue: Total revenue of the Disney company per year (Millions of US $)

Variables:

- Year: year of total revenue
- Movie_count: count of movies released every year
- Musical: count of movies released per year with the genre “musical”
- Comedy: count of movies released per year with the genre “comedy”
- Action: count of movies released per year with the genre “action”
- Adventure: count of movies released per year with the genre “adventure”
- Drama: count of movies released per year with the genre “drama”

Analysis Questions

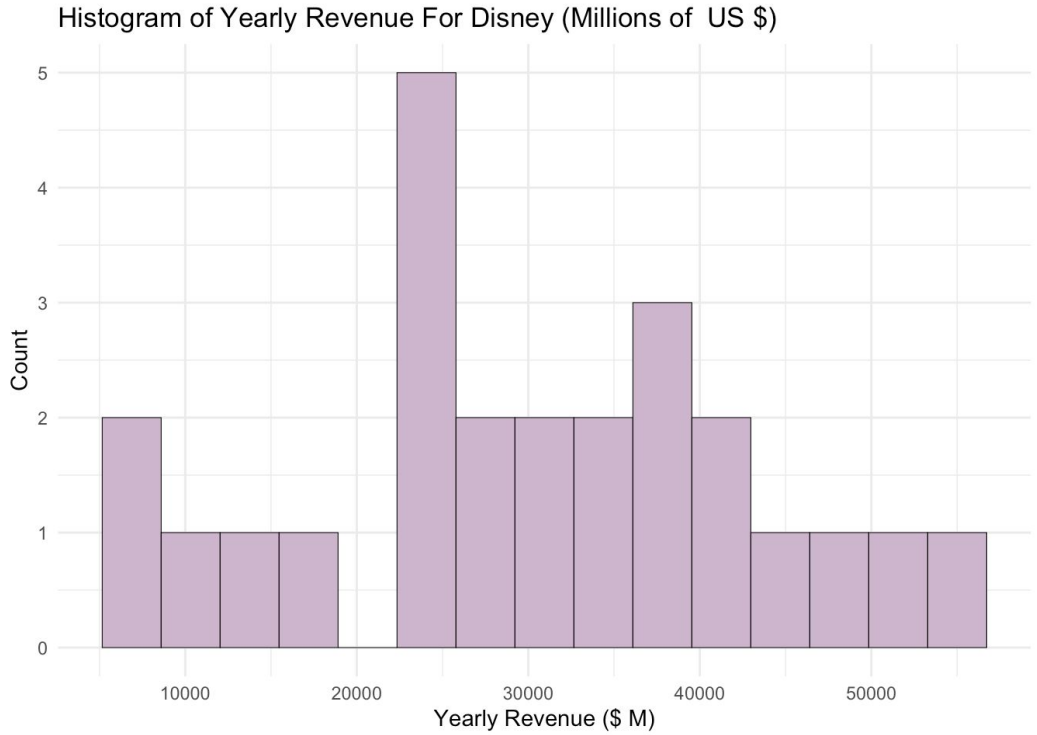
1. **What are the highest and lowest grossing movies?**
   → Bar chart of gross revenue per movie

2. **What is the most common genre produced by Disney?**
   → Bar chart of count of movie genre

3. **Which variables best predict the actual revenue per year?**
   → AIC variable selection and multiple linear regression

4. **What is Disney's expected total revenue in a year where they release 10 movies and 2 of them are comedies?**
   → 95% prediction interval using the multiple linear regression model

Exploratory Data Analysis

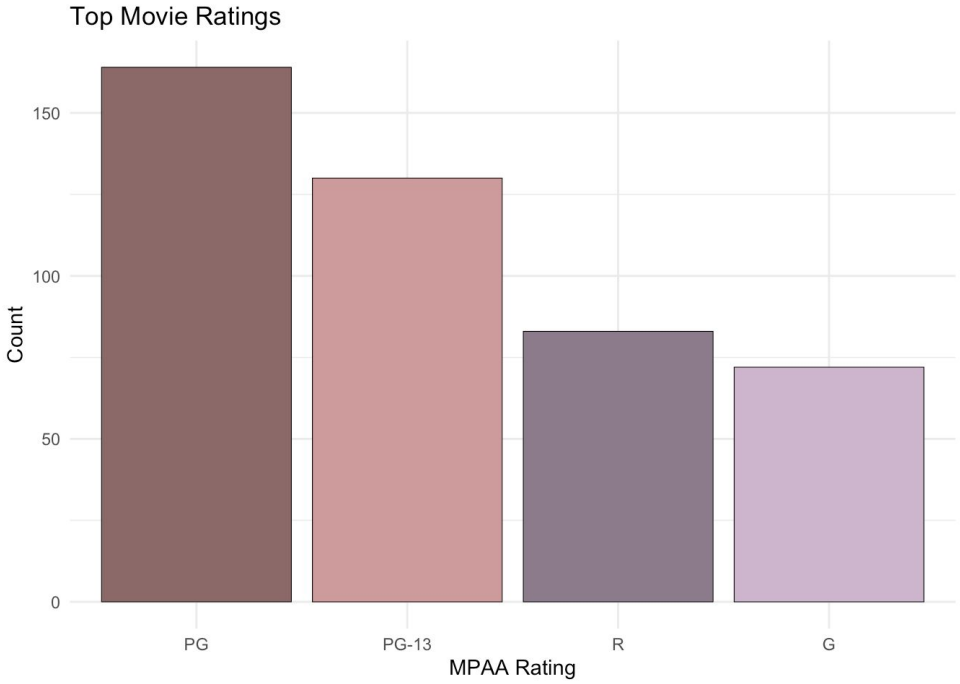


- As we can see at the 20,000 mark: most years made more than 20 billion dollars yearly

Exploratory Data Analysis

- The Rating with the most produced movies is PG.

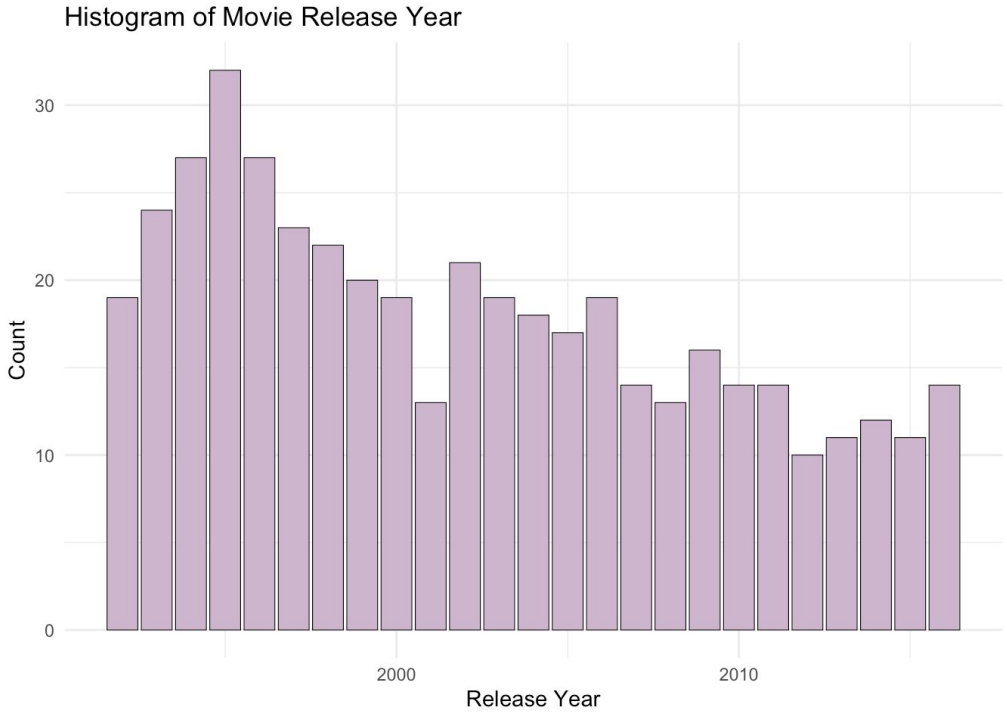
- PG movie examples:
  - Lilo & Stitch
  - The Lizzie McGuire Movie
  - The Haunted Mansion



Exploratory Data Analysis

- The year with the most produced movies is 1995.

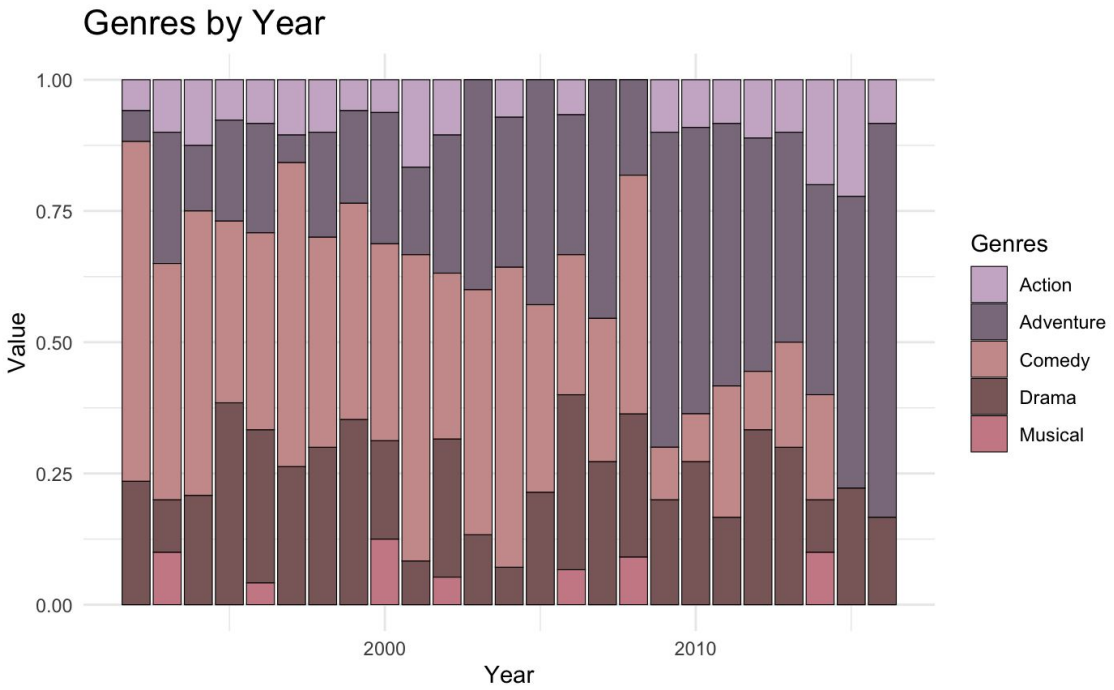
- Movies this year included
  - A Goofy Movie
  - Toy Story
  - Pocahontas



Exploratory Data Analysis

- We can see the comedy genre decrease substantially throughout the years

- Adventure movies have increased from 1992 to 2016.



Question 1: What is the highest grossing movie?

- Star Wars Ep VII was the highest grossing movie bringing in 936.7 million dollars.

Source: Disney Character Success from Kaggle

Summary of the $ Gross Revenue Per Movie from 1992 to 2016

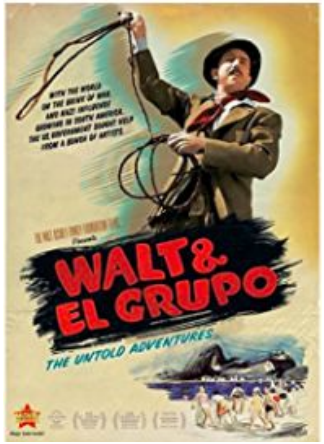
Million US $

Movies	Genre	Rating	Year	$ Gross
Star Wars Ep. VII: The Force Awakens	Adventure	PG-13	2015	936.7
The Lion King	Adventure	G	1994	761.6
The Avengers	Action	PG-13	2012	660.1
Pirates of the Caribbean: Dead Man'...	Adventure	PG-13	2006	544.8
Rogue One: A Star Wars Story	Adventure	PG-13	2016	529.5
Finding Nemo	Adventure	G	2003	518.1
Finding Dory	Adventure	PG	2016	486.3
The Sixth Sense	Thriller/Suspense	PG-13	1999	485.4

This file contains data on the Revenue and Gross of the Walt Disney Company from 1992 to 2016

What is the lowest grossing movie?

- Walt and El Grupo was the lowest grossing movie only making $23,064.



Source: Disney Character Success from Kaggle

Summary of the Gross Revenue Per Movie from 1992 to 2016

Million US $

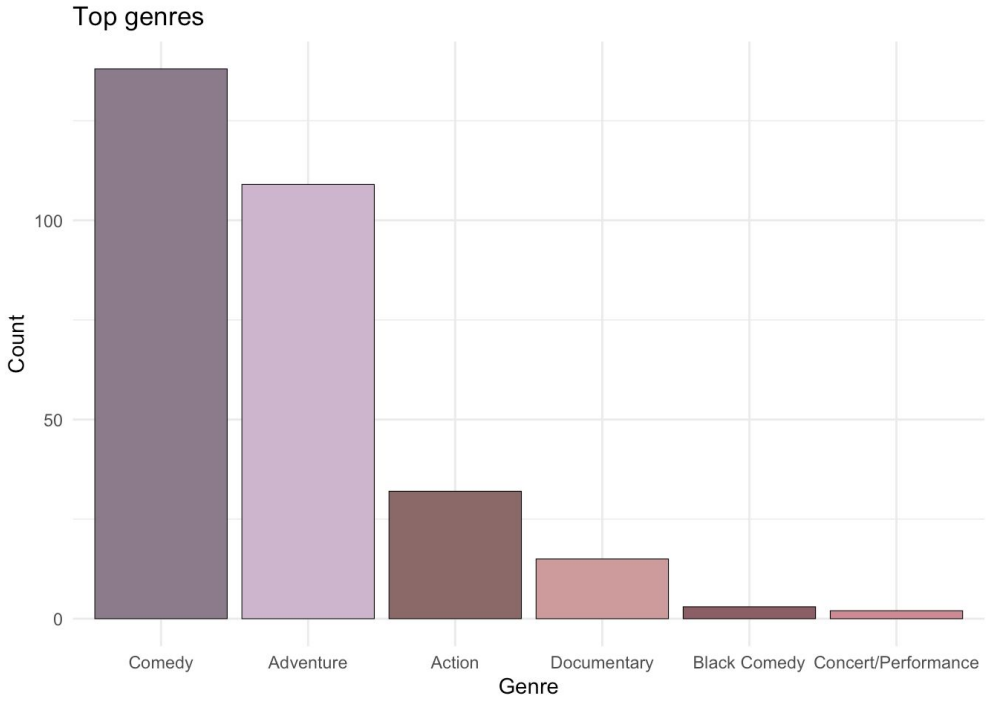
Movies	Genre	Rating	Year	$ Gross
Walt and El Grupo	Documentary	PG	2009	0.0
Zokkomon	Adventure	PG	2011	0.0
An Alan Smithee Film: Burn Hollywood ...	Comedy	R	1998	0.1
Waking Sleeping Beauty	Documentary	PG	2010	0.1
Gedo Senki (Tales from Earthsea)	Adventure	PG-13	2010	0.1
Breakfast of Champions	Comedy	R	1999	0.3
Goal! 2: Living the Dream...	Drama	PG-13	2008	0.3
Morning Light	Documentary	PG	2008	0.3

This file contains data on the Revenue and Gross of the Walt Disney Company from 1992 to 2016

Question 2: What is the most common genre produced by Disney?

- The most common genre produced by Disney overall is comedy.

- Movies include
  - Recess: School's Out
  - Air Bud
  - Aladdin



Question 3: Which variables best predict total revenue of the Disney Company?

```
Step:  AIC=431.19
total_revenue ~ comedy + movie_count

               Df Sum of Sq        RSS    AIC
<none>                      608473521 431.19
+ musical       1  23449554  585023967 432.21
+ action_count  1  19633920  588839601 432.37
+ adventure     1   5720004  602753516 432.95
+ drama         1    137329  608336192 433.18
- movie_count   1 105911512  714385033 433.20
- comedy        1 785647804 1394121325 449.92
```

- We used Stepwise Regression AIC for variable selection.
- Just a quick glance we get a 2 predictor model with total_revenue as the response and comedy and movie_count as our predictors.

Summary of full model without interaction effect

- R- squared is .856
- 85.6% of variation in annual revenue is explained by the number of movies released and the number of comedies released per year
- Predictors are significant
- Both coefficients are negative, which suggests that Disney's revenue decreases for each additional movie produced.

```
mod.full <- lm(total_revenue ~ comedy + movie_count, data = yearly2.0)
summary(mod.full)
```

```
##
## Call:
## lm(formula = total_revenue ~ comedy + movie_count, data = yearly2.0)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9454.0 -3619.1   706.5  3128.4  8969.2
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  55173.2     4001.6  13.788 2.64e-12 ***
## comedy       -2511.2      471.2  -5.330 2.38e-05 ***
## movie_count   -607.9      310.6  -1.957   0.0632 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5259 on 22 degrees of freedom
## Multiple R-squared:  0.856,  Adjusted R-squared:  0.8429
## F-statistic:  65.4 on 2 and 22 DF,  p-value: 5.509e-10
```

Summary of full model with interaction effect

- Interaction term is not significant
- We will not include in our model

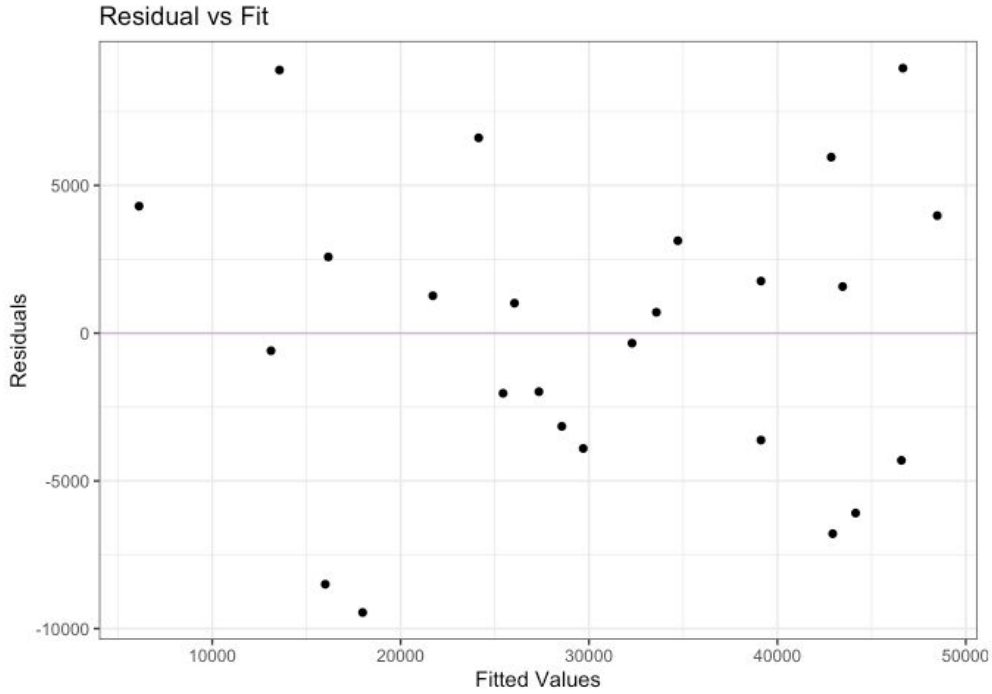
```
add1(mod.full, ~.+movie_count*comedy, test = 'F')
```

```
## Single term additions
##
## Model:
## total_revenue ~ comedy + movie_count
##                      Df Sum of Sq       RSS    AIC F value Pr(>F)
## <none>                             608473521 431.19
## comedy:movie_count  1  45954919 562518602 431.23  1.7156 0.2044
```

LINE Conditions: (Linearity and Equal Variance)

- Plot is “Well-Behaved”
  - Non-linearity is not a problem.
  - The errors have equal variances.

- Transformation will not be required on x or y.

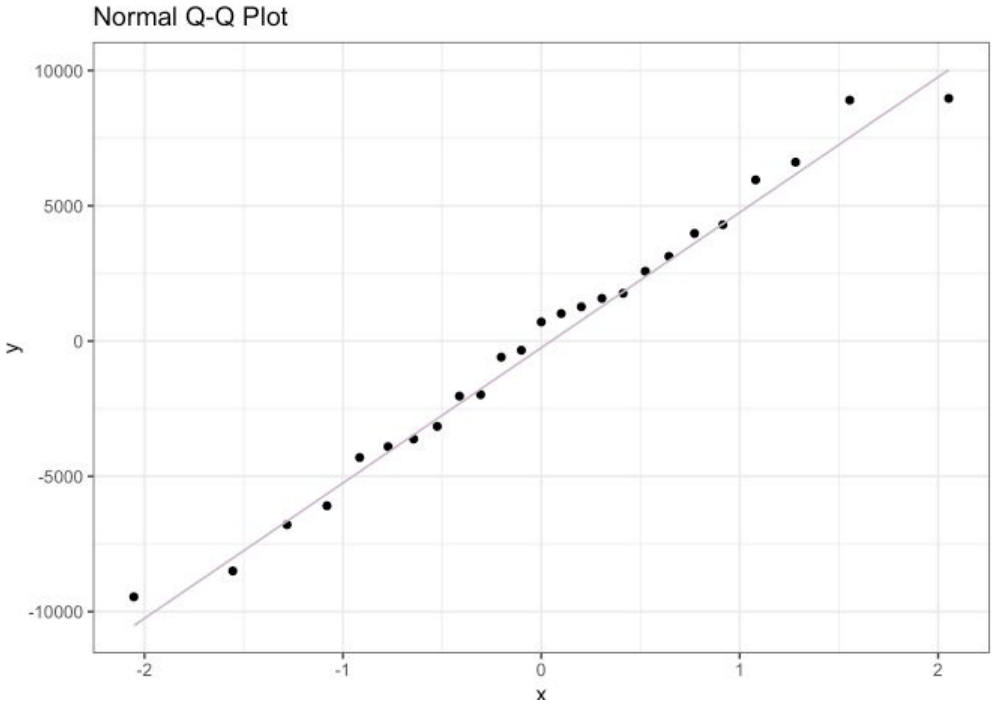


LINE Conditions: (Normality)

- It is normally distributed.
- Also, in the shapiro test we fail to reject the null hypothesis since 0.91 > 0.05 = alpha, so we don't do any transformation on y.

```
Shapiro-Wilk normality test

data:  resid(mod.cp)
W = 0.98134, p-value = 0.9102
```



Question 4: What is Disney's expected total revenue in a year where they releases 10 movies and 2 of them are comedies?

We are 95% confident that the Disney's total gross revenue in a year where they releases 10 movies and 2 of them are comedies will be between 32500.31 and 55643.5 in millions US $.

```
new = data.frame(Comedy = 2, movie_count = 10)
prediction = predict(mod.full,new,interval = "prediction", level = 0.95)
prediction
```

	fit	lwr	upr
1	44071.91	32500.31	55643.5

Conclusion

- Comedy is the most common movie genre created by Disney between the years of 1992 and 2016.

- We can successfully predict Disney's annual total revenue by counting the number of total movies released each year and counting the number of comedy movies released each year

- We are 95% confident that Disney will make between $32,500 M and $55,643 M

Future Work

- Someone can pull in the data set of main characters per movie and define each main character by certain characteristics (funny / quirky / hero / anti-hero / etc) and predict the success of a movie based on characteristics of the main character

- Include data beyond 2016 to see the most recent changes in movies Disney releases.

Limitations

- Translating gross revenue from a text variable to a numeric variable

- We can't get a very high R^2 value to predict gross revenue from movies. Adding more variables from other data sets might help predict individual movie revenue, but we decided to focus on revenue of Disney as a whole.


Thank you