Project 1: Draft Milestone 3 – Disney Gross Prediction before Covid

Thip Rattanavilay

Master of Science data science

Bellevue University

Abstract

There are multiple factors and national trends that could affect the output and consumption of the Walt Disney Company. As we enter the outbreak of Covid-19 last year, the world has seen an enormous impact on Gross Domestic Product and unemployment. The entertainment industry and the Walt Disney Company felt the same effects. The Walt Disney Company furloughed 77,000 employees as they were forced to shut down all major parks, cruises, and other entertainment facilities because of the coronavirus outbreak worldwide. This project analyzes the use of features extracted from network representations of the Disney Movie Gross profit found in Kaggle etc. and before Covid-19. I am hoping to show that through the use of these features, it is possible to build more powerful prediction models compared to common baseline methods. Movies make a high profile, billion-dollar industry and prediction of movie revenue can be very lucrative. Predicted revenues can be used for planning both the production and distribution stages. Hence, the tough job of predicting a movie's gross revenue can be simplified with the help of modern computing power and the historical data available as movie databases.

Introduction

The Walt Disney Company is a global entertainment and media company headquartered in Burbank, California. It was founded on October 16, 1923, as the Disney Brothers Studio, an animation company. In the past 89 years, the company has flourished, employing 156,000 people globally. Its business can be broken into five segments: Media Networks, Parks & Resorts, Studio Entertainment, Consumer Products, and the Disney Interactive Media Group. The Walt Disney Company first began as the Disney Brothers Studio; this business segment was the platform on which it all began. Today, Disney Studios continues to bring high quality films, music, and stage plays worldwide. Over the year Disney has acquired or begun ventures such as Marvel Studios, Touchstone Pictures, Pixar Animation Studios, Disney Music Group, Disneynature and Walt Disney Studios Motion Pictures.

Business Problem

Box office revenue prediction is an important problem in the film industry that governs financial decisions made by producers and investors. Generally, these predictions are made using data science algorithms and statistical techniques as described in this analysis. While these approaches are common practice, they often only provide a coarse estimate of revenue prediction before a film has been released. This project aims to develop a computational model for predicting Disney's box office revenues based on public data for movies extracted from popular movie databases (post covid). Whereas the fact that approximately 25% of the gross revenue gets accumulated in the first weekend of screening; indulged them to use the first weekend collection and the number of screenings to build another more accurate prediction

model. Furthermore, to find out the impact of film-critics and award nominations, they built another model using rating given by a well-known film critic and academy award nominations. There results showed that the use of opening weekend business predicts the gross revenue most accurately among all the other models.

Dataset

The data contains 579 Disney movies with list of predictor variables includes genre, MPAA ratings, country of origin, star power, production budget, indicator variable for sequels of earlier movies, indicator variables for release during certain holiday periods of the year, number of screenings in the first weekend, rating of the movie by well-known film critic and the academy award nominations.

The variables that I will be using for this project are:

- movie_title
- release_data
- genre
- mpaa_rating
- total_gross
- inflation_adjusted gross

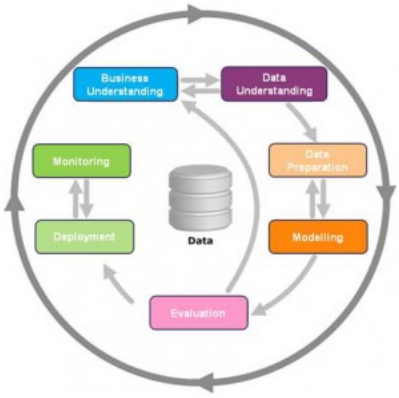
Therefore, most of our data preprocessing involved rectifying those inconsistencies.

This includes:

- Filtering out voice actors/roles that were not in a Disney animated feature film
- Filtering out live action movies that are not relevant to our project
- Adding missing movie box office revenues
- Adding missing voice actors & roles of Disney animated feature films
- Formatting the dates in the movies dataset for consistency
- Filling in IMDB ratings for the movies
- Joining the Academy Award dataset with the movies dataset to obtain the wins for the movies of interest
- Joining the Academy Award dataset with the voice actor's dataset to obtain the wins for the voice actors of interest

Methods

The goal is to use Multiple Linear Regression to have an accurate predictive model using the CRISP-DM method, the approach was broken down into six simple steps. Each step consists of the tasks that need to be performed before moving on to the next phase.



Business understanding: Determine resources availability, project requirements, assess risks and contingencies, and conduct a cost-benefit analysis.

Data understanding: Examine the data and document its surface properties like data format, number of records, or field identities. Dig deeper into the data. Query it, visualize it, and identify relationships among the data.

Data preparation: Determine which data sets will be used and document reasons for inclusion/exclusion. A common practice during this task is to correct, impute, or remove erroneous values.

Modeling: Determine which algorithms to try (e.g. regression, neural net). Pending your modeling approach, you might need to split the data into training, test, and validation sets. Generally, multiple models are competing against each other, and the data scientist needs to interpret the model results based on domain knowledge, the pre-defined success criteria, and the test design.

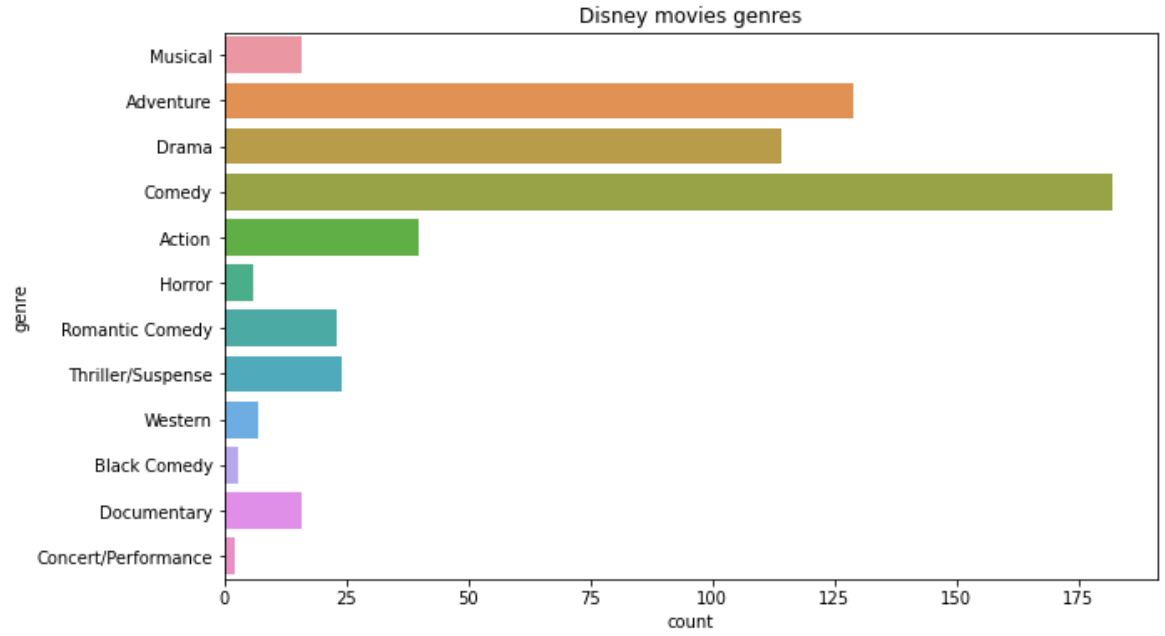
Evaluation: Do the models meet the business success criteria? Which one(s) should we approve for the business? Based on the previous three tasks, determine whether to proceed to deployment, iterate further, or initiate new projects.

Deployment: Report final results. Develop and document a plan for deploying the model. Conduct a project retrospective about what went well, what could have been better, and how to improve in the future.

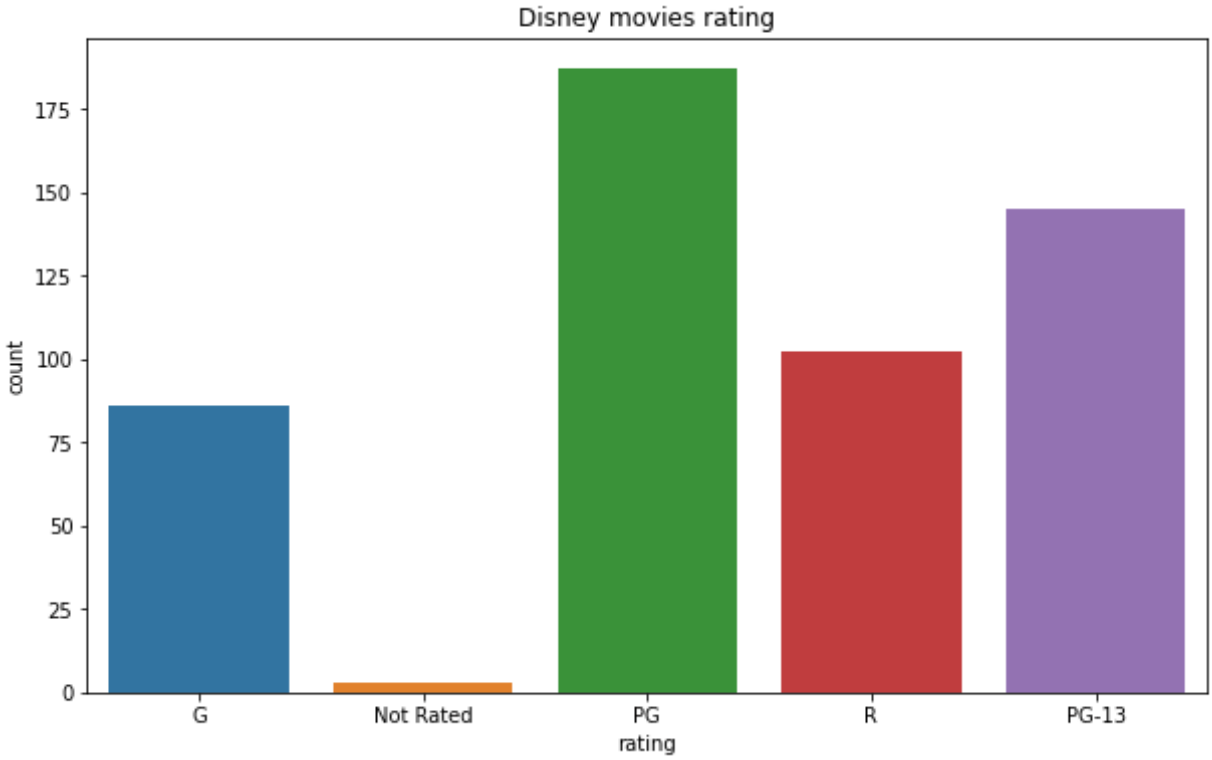
Illustrations

Disney genre trend:

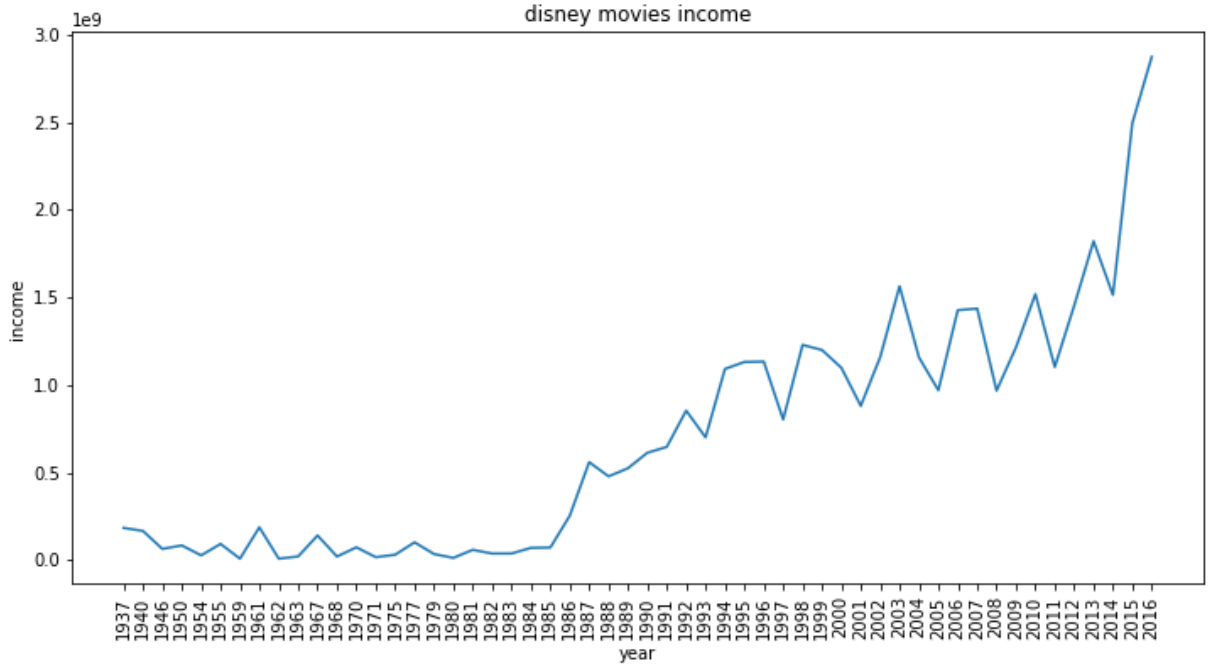
	genre	release_year	total_gross	inflation_adjusted_gross
0	Action	1981	0.00	0.00
1	Action	1982	26918576.00	77184895.00
2	Action	1988	17577696.00	36053517.00
3	Action	1990	59249588.50	118358772.00
4	Action	1991	28924936.50	57918572.50
5	Action	1992	29028000.00	58965304.00
6	Action	1993	21943553.50	44682157.00
7	Action	1994	19180582.00	39545796.00
8	Action	1995	63037553.50	122162426.50
9	Action	1996	135281096.00	257755262.50



Disney rating:



Disney Movie Rev:



Appendix

The data contains 579 Disney movies with six features: movie title, release date, genre, MPAA rating, total gross, and inflation-adjusted gross.

The variables that are used:

- movie_title
- release_data
- genre
- mpaa_rating
- total_gross
- inflation_adjusted gross

EDA

	movie_title	release_date	genre	mpaa_rating	total_gross	inflation_adjusted_gross
0	Snow White and the Seven Dwarfs	1937-12-21	Musical	G	184925485	5228953251
1	Pinocchio	1940-02-09	Adventure	G	84300000	2188229052
2	Fantasia	1940-11-13	Musical	G	83320000	2187090808
3	Song of the South	1946-11-12	Adventure	G	65000000	1078510579
4	Cinderella	1950-02-15	Drama	G	85000000	920608730
...	...	...	...	...	...	...
574	The Light Between Oceans	2016-09-02	Drama	PG-13	12545979	12545979
575	Queen of Katwe	2016-09-23	Drama	PG	8874389	8874389
576	Doctor Strange	2016-11-04	Adventure	PG-13	232532923	232532923
577	Moana	2016-11-23	Adventure	PG	246082029	246082029
578	Rogue One: A Star Wars Story	2016-12-16	Adventure	PG-13	529483936	529483936

Feature

	movie_title	release_date	genre	mpaa_rating	total_gross	inflation_adjusted_gross	decade
0	Snow White and the Seven Dwarfs	1937-12-21	Musical	G	184925485	5228953251	<1950
1	Pinocchio	1940-02-09	Adventure	G	84300000	2188229052	<1950
2	Fantasia	1940-11-13	Musical	G	83320000	2187090808	<1950
3	Song of the South	1946-11-12	Adventure	G	65000000	1078510579	<1950
4	Cinderella	1950-02-15	Drama	G	85000000	920608730	<1950
...	...	...	...	...	...	...	...
574	The Light Between Oceans	2016-09-02	Drama	PG-13	12545979	12545979	2010-2020
575	Queen of Katwe	2016-09-23	Drama	PG	8874389	8874389	2010-2020
576	Doctor Strange	2016-11-04	Adventure	PG-13	232532923	232532923	2010-2020
577	Moana	2016-11-23	Adventure	PG	246082029	246082029	2010-2020
578	Rogue One: A Star Wars Story	2016-12-16	Adventure	PG-13	529483936	529483936	2010-2020

10 Question

1. What does the business need?
2. What modeling techniques should we apply?
3. How do stakeholders access the results?
4. What is Disney's internal weakness?
5. How many genres are there and are they mixed?
6. Which production group made the most revenue for Disney?
7. Which genre of movie is earning the highest Total Gross Income?
8. Which movie earns the highest income?
9. What is the trend of movies over the years (1937-2016)?
10. Which model best meets the business objectives?

References

The Walt Disney Company. (n.d.). Retrieved from https://www.thewaltdisneycompany.com/

Forte, D. (2017, September 28). Walt Disney launches new online, store shopping experiences. Retrieved March 17, 2019, from https://multichannelmerchant.com/ecommerce/walt-disneylaunches-new-online-store-shoppingexperiences/

Bonomolo, C. (2018, August 29). Disney's streaming service gets its name as new details emerge. Retrieved October 8, 2018, from https://comicbook.com/movies/2018/08/26/disneystreaming-service-new-details-disneyplay/

Alexander, J. (2018, February 06). Disney's stand-alone streaming service won't compete with Netflix in scale. Retrieved October 8, 2018, from https://www.polygon.com/2018/2/6/16981884/disneystreaming-netflix-marvel-star-warspixar-lucasfilm-svod

Alvarez, E. (2018, April 12). ESPN offers a first look at Disney's big plans for streaming. Retrieved October 8, 2018, from https://www.engadget.com/2018/04/12/espn-plusdisney-streamingservice-bamtech/

Anderson, M. (2018, February 6). Will Disney's streaming service roar - or squeak? Retrieved October 8, 2018, from https://www.usnews.com/news/business/articles/2018-02-06/will-disneysstreaming-service-roar-or-squeak

Bartiromo, M. (2019, March 13). Walt Disney World increases ticket prices for most popular days. Retrieved March 14, 2019, from https://www.foxnews.com/travel/waltdisney-world-increasesticket-prices-for-most-popular-days

Thompson, D. (2018, May 17). Disneyflix is coming. And Netflix should be scared. Retrieved October 8, 2018, from https://www.theatlantic.com/magazine/archive/2018/05/disneyflix-netflix/556895/

Mitovich, M. W. (2018, February 09). Disney streaming service FAQ: Will Marvel heroes leave Netflix? Multiple Star Wars series? And more answers. Retrieved October 8, 2018, from https://tvline.com/2018/02/08/star-wars-tv-series-disney-streaming-service-2019/

Krämer, P. (2000). Entering the Magic Kingdom: The Walt Disney company, the Lion King and the limitations of criticism. Film Studies, (2), 44-50,2.