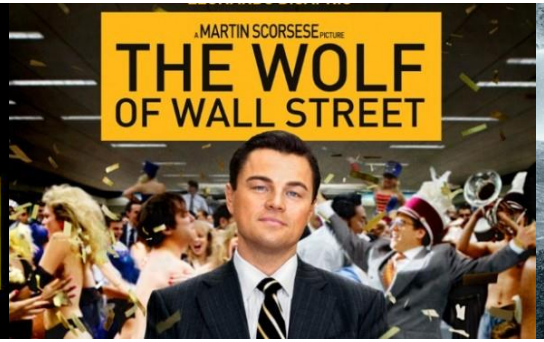
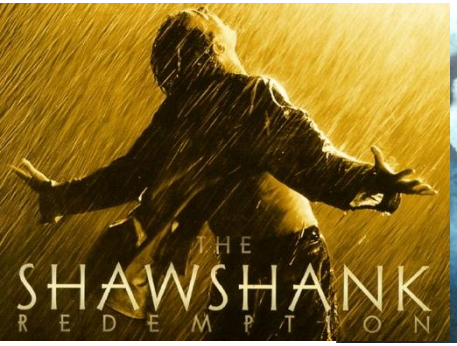


Movie Database

General Assembly Final Project

Kate Benjamin



Goals for Final Project

- I am **addressing** a data-related project in my field of work.
- My **data** is from Box Office Mojo **(http://www.boxofficemojo.com/)**
- My datasets **include** "*Average Gross Per Theatre from 1980-2014*" " *Average Ticket Price from 1980-2014,*" "*Movie Release by Season from 1980-2014,*" "*Studio Makeup by Year from 1980-2014,*" and my largest and most valuable one, "*Yearly Movie Gross from 1980-2014.*"
- My **dataset** "*Yearly Movie Gross from 1980-2014*" includes 13,113 movie title releases from 1980-2014. I ranked them by year and total gross. Included in the data is studio, total gross, opening week gross, opening day and gross per theatre. I created a new feature "2014 dollars." This is the amount the movie would have made if it were released in 2014.
- Using linear regression, I will **predict** the total gross of movies from 2000 to 2014 using opening week performance and the number of theatres it played in. I will then add in the studio name and year as new features to see if my prediction score becomes more accurate. I will continue to add new features to see if my score gets stronger.
- This summer, my **team** and I plan to update this dataset to include movies from 2015.

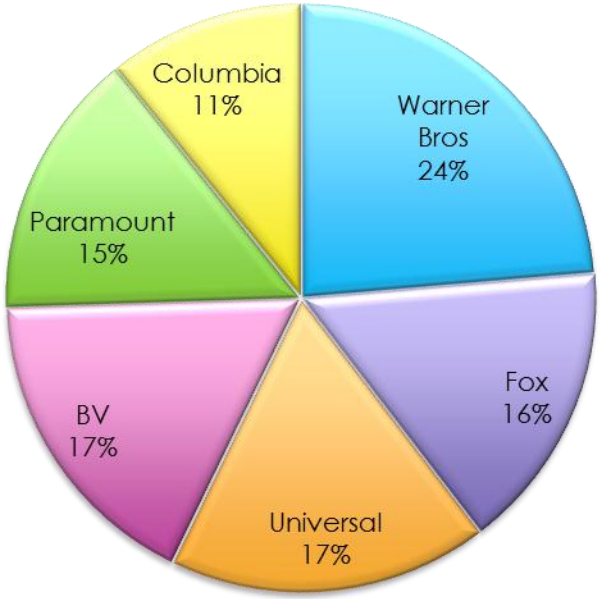
American Sniper was the highest grossing movie of 2014.

Rank	Studio	Movie Title	Total Gross	Opening	Theatres	Gross Per Theater
1	WB	American Sniper	$337,209,000	$633,456	3,885	$86,798
2	LGF	The Hunger Games: Mocking jay - Part 1	$336,962,000	$121,897,634	4,151	$81,176
3	BV	Guardians of the Galaxy	$333,176,600	$94,320,883	4,088	$81,501
4	BV	Captain America: The Winter Soldier	$259,766,572	$95,023,721	3,938	$65,964
5	WB	The LEGO Movie	$257,760,692	$69,050,279	3,890	$66,262
6	WB	The Hobbit: The Battle of the Five Armies	$251,678,000	$54,724,334	3,875	$64,949
7	Par.	Transformers: Age of Extinction	$245,439,076	$100,038,390	4,233	$57,982
8	BV	Maleficent	$241,410,378	$69,431,298	3,948	$61,148
9	Fox	X-Men: Days of Future Past	$233,921,534	$90,823,660	4,001	$58,466
10	BV	Big Hero 6	$218,056,000	$56,215,889	3,773	$57,794
11	Fox	Dawn of the Planet of the Apes	$208,545,589	$72,611,427	3,969	$52,544
12	Sony	The Amazing Spider-Man 2	$202,853,933	$91,608,337	4,324	$46,913
13	WB	Godzilla (2014)	$200,676,069	$93,188,384	3,952	$50,778
14	Sony	22 Jump Street	$191,719,337	$57,071,445	3,426	$55,960
15	Par.	Teenage Mutant Ninja Turtles (2014)	$191,204,754	$65,575,105	3,980	$48,041
16	Par.	Interstellar	$186,666,000	$47,510,360	3,561	$52,420
17	Fox	How to Train Your Dragon 2	$177,002,924	$49,451,322	4,268	$41,472
18	Fox	Gone Girl	$167,628,577	$37,513,109	3,284	$51,044
19	LG/S	Divergent	$150,947,895	$54,607,747	3,936	$38,351
20	Uni.	Neighbors	$150,157,400	$49,033,915	3,311	$45,351

Source: Box Office Mojo

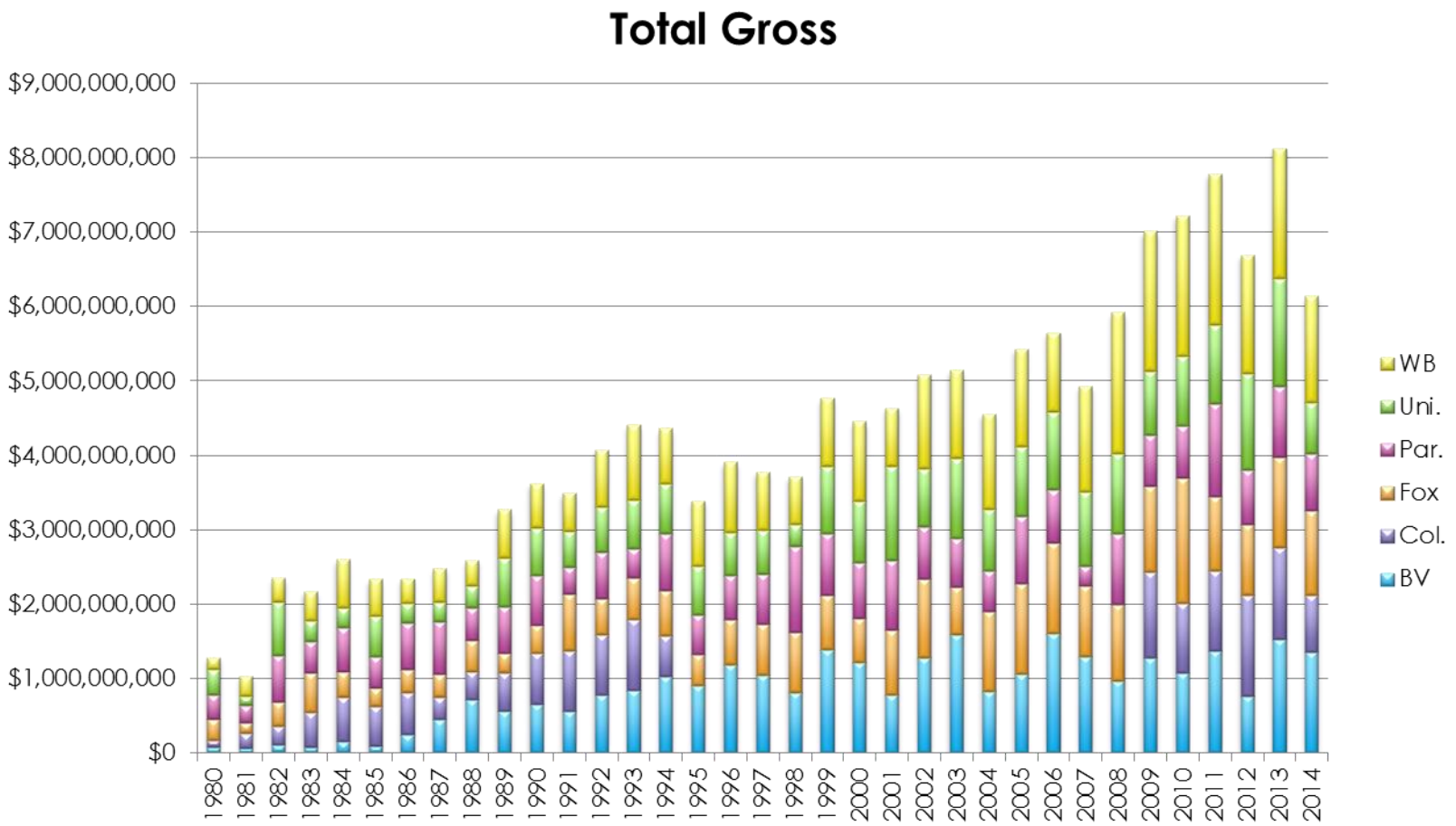
Since 1980, Warner Brothers has produced roughly ¼ of all new movies, in comparison to other top studios.

Top Studios by Presence Since 1980



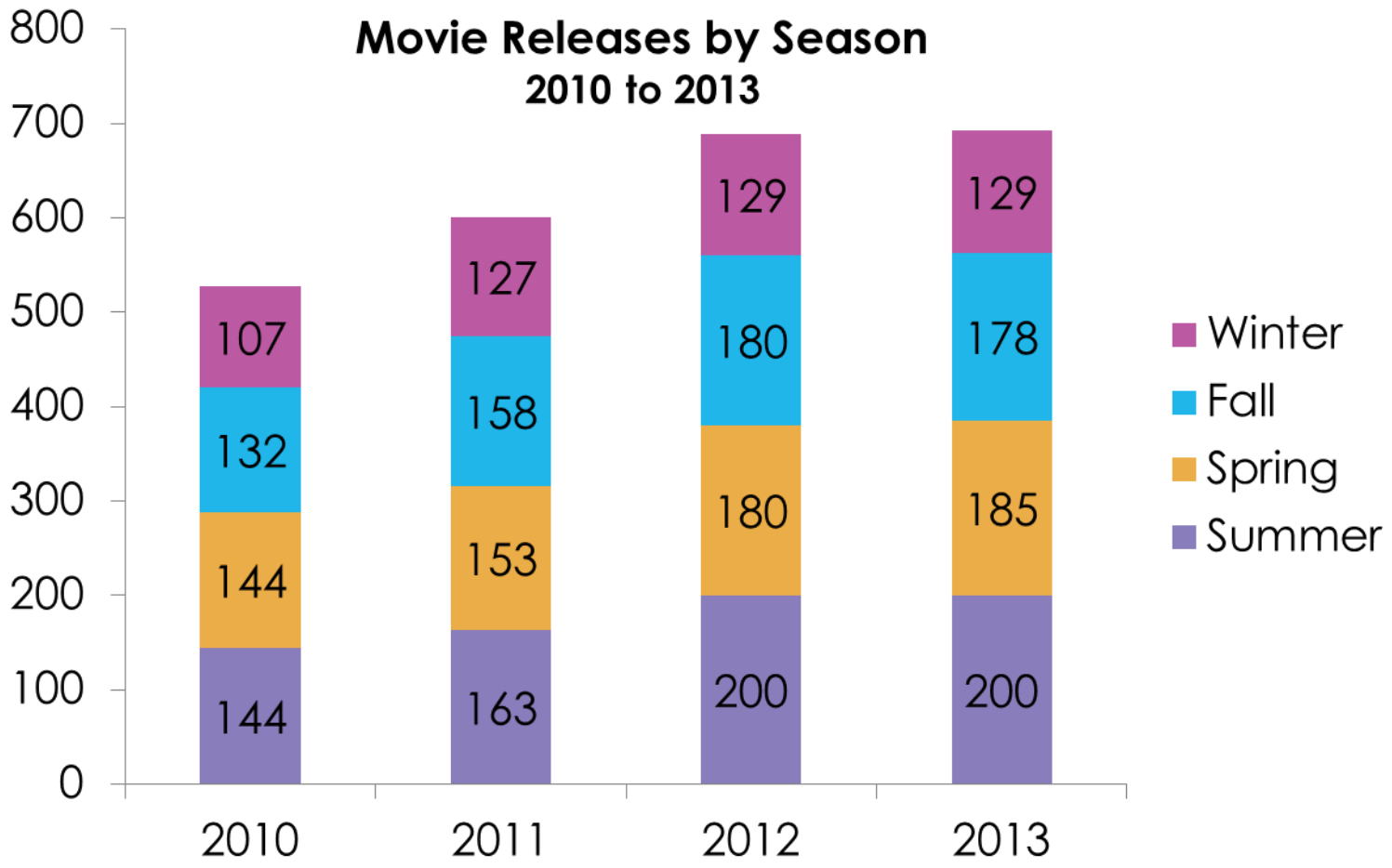
Source: Box Office Mojo

2013 was the highest grossing year for new movies, while 2012 and 2014 were unusually low.



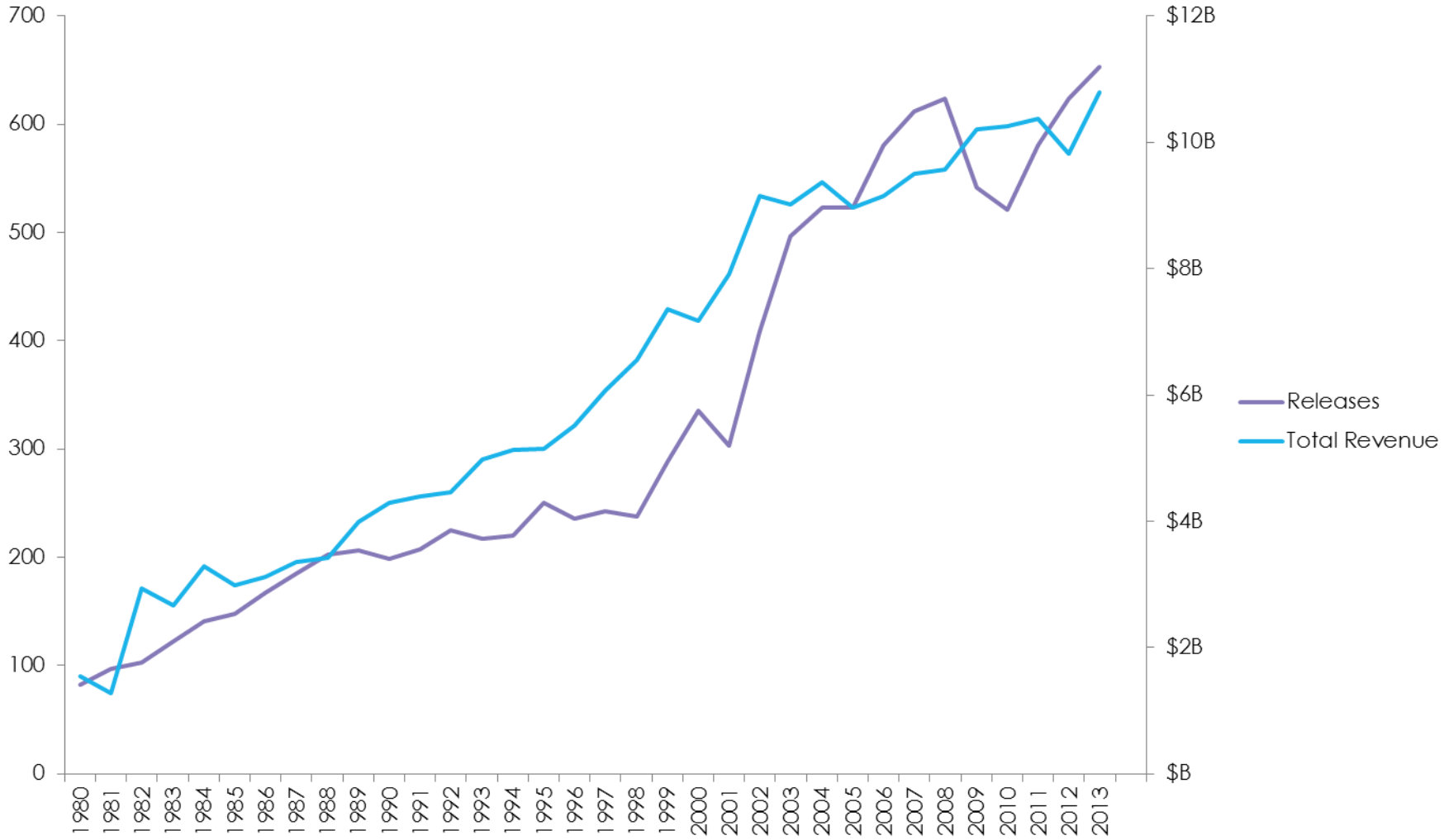
Source: Box Office Mojo
Data unavailable for Columbia Pictures 1995 to 2008.
BV (Buena Vista) stands for Walt Disney

Summer months tend to be the most popular time for new movie releases, while the winter sees a significant drop.



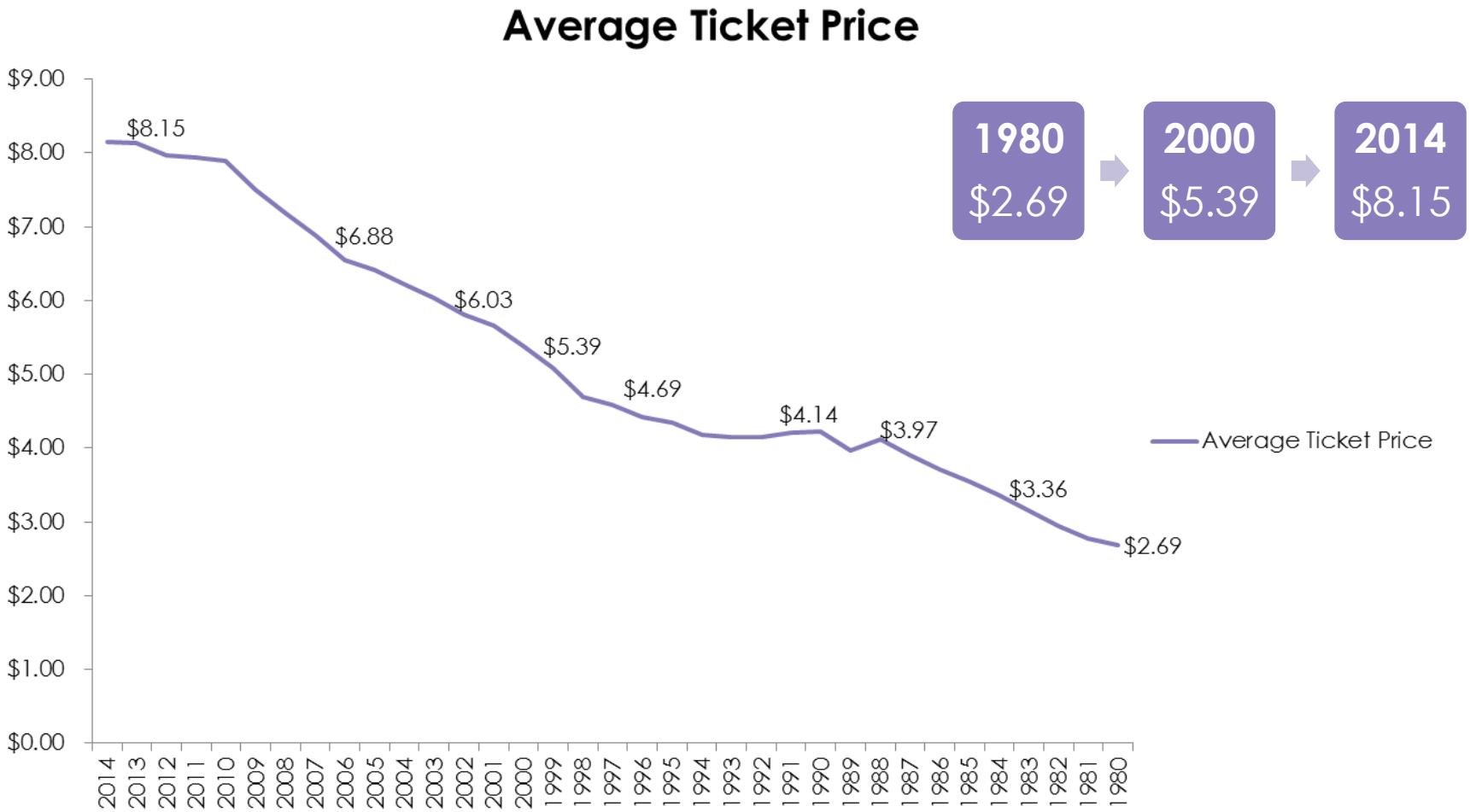
Source: Box Office Mojo

While the number of new movies released from 2008 to 2010 dropped significantly, total revenue was not impacted.



Source: Box Office Mojo
Data unavailable for 2014.

The average movie ticket price went up +203% from 1980 to 2014.



Source: Box Office Mojo

Using Linear Regression, my model was 92% accurate.

- Using **Linear Regression**, my model was 86% to 92% accurate, depending on my features.
- When I predicted total gross using just **opening week gross** and **length of weeks the movie was in theaters for** my model was 86% accurate.
- When I predicted **total gross** using **opening week gross, amount of theaters played in and year** my model was 86% accurate.
- Then I added in "weekly information" I had. When I predicted **total gross** using **opening gross, theaters, year, gross per week, amount of week movie was number 1, and the length of weeks the movie was in theaters** my model was 89% accurate.
- When I added **movie rank for the year, amount of theaters it opened in**, and **gross per theater** my model was 92% accurate.
- *Note: amount of theaters played in and amount of theaters opened in are different*
- I decided to not take into consideration studio name, quarter fiscal year and opening and closing dates. I'm sure my model would have been even more accurate with opening and closing dates (summer hits for example) but there was too much missing data for the feature for it to be reliable.
- The **opening weekly gross and the length of weeks the movie was in theaters for** were in my opinion the best indicator in predicting the total gross of a movie.

Conclusion

- In the future, I would be interested in using a **decision tree** or **random forest** to see if my results vary.
- Since this is an ongoing project at work, the dataset will be updated to include recent movies.
- In the future, I would like to predict total gross based on **studio and opening/closing day** as well – but right now there is too much missing data in those features, so I had to delete them from my dataset.
- The file was also to large for my laptop to include movies from 1980 – 2014, so I would also be interested if you used all of the data how the results would differ.


Thank you Mason, Lema and Data Science 30!
BELIEVE
GHOST
THE MUSICAL
TICKETMAIL.COM or 877-250-2929
COME IN FROM THE COLD!
OPEN 24 HOURS
FRIDAYS
FRIDAYS