



Project #1 - Group 10



Rob Savage
Nicole Cook
Hermela Mekonnen
Franklin Truong



Analysis of Movie Box Office Data

Questions of Importance -

- Do popular genres change over time?
- Does the audience's taste for genre change over time?
- Do seasons have an effect on the metrics of a movie?
- Do seasons impact runtime and audience ratings?
- Has movie revenue (adjusted for inflation) fluctuated through the decades?
- Does duration have any impact on the financial success of a movie?

Analysis of Movie Box Office Data

Why those particular questions? -

- Wide-range of interest from audience
- Innate curiosity of the psychology of moviegoers
- As screens have become more of a dominant presence in the lives of humans, attention span may see a significant decrease
- Interest in seeing if the lack of available activities (seasons) impact the film industry

Analysis of Movie Box Office Data

Data Resources -

- ❑ Kaggle - IMDb Movies Extensive Data Set:
<https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset>
- ❑ Types of data included:
 - ❑ Title, Year of Release, Genre, Duration, Country of Origin, Language, Director, Writer(s), Actor(s), Production Company, Average User Rating, and # of Average User Ratings.

Analysis of Movie Box Office Data

Data Resources -

- ❏ data.world - BoxOfficeMojo All-time Domestic Data Set:
<https://data.world/eliasdabbas/boxofficemojo-alltime-domestic-data>
- ❏ Types of data included:
 - ❏ Title, Studio, Lifetime Gross, and Year.



Data Cleaning



Data Cleaning

Goals -

- Merge data sets on Title and Year
- Bring over Lifetime Gross Revenue into the primary IMDb data frame
- Adjust Lifetime Gross for inflation
- Explode data on Genre
- Re-format headers

Original IMDb data

	imdb_title_id	title	original_title	year	date_published	genre	duration	country	language	director	...	actors	description	a
0	tt0000574	The Story of the Kelly Gang	The Story of the Kelly Gang	1906	1906-12-26	Biography, Crime, Drama	70	Australia	NaN	Charles Tait	...	Elizabeth Tait, John Tait, Norman Campbell, Be...	True story of notorious Australian outlaw Ned ...	
1	tt0001892	Den sorte drøm	Den sorte drøm	1911	1911-08-19	Drama	53	Germany, Denmark	NaN	Urban Gad	...	Asta Nielsen, Valdemar Psilander, Gunnar Helse...	Two men of high rank are both wooing the beaut...	
2	tt0002101	Cleopatra	Cleopatra	1912	1912-11-13	Drama, History	100	USA	English	Charles L. Gaskill	...	Helen Gardner, Pearl Sindelar, Miss Fielding, ...	The fabled queen of Egypt's affair with Roman ...	
3	tt0002130	L'Inferno	L'Inferno	1911	1911-03-06	Adventure, Drama, Fantasy	68	Italy	Italian	Francesco Bertolini, Adolfo Padovan	...	Salvatore Papa, Arturo Pirovano, Giuseppe de L...	Loosely adapted from Dante's Divine Comedy and...	
4	tt0002199	From the Manger to the Cross; or, Jesus of Naz...	From the Manger to the Cross; or, Jesus of Naz...	1912	1913	Biography, Drama	60	USA	English	Sidney Olcott	...	R. Henderson Bland, Percy Dyer, Gene Gauntier,...	An account of the life of Jesus Christ, based ...	

- Messy headers
- No Lifetime Gross
- Multiple Genres

Original Box Office data

rank		title	studio	lifetime_gross	year
0	1	Star Wars: The Force Awakens	BV	936662225	2015
1	2	Avengers: Endgame	BV	857190335	2019
2	3	Avatar	Fox	760507625	2009
3	4	Black Panther	BV	700059566	2018
4	5	Avengers: Infinity War	BV	678815482	2018
5	6	Titanic	Par.	659363944	1997
6	7	Jurassic World	Uni.	652270625	2015
7	8	Marvel's The Avengers	BV	623357910	2012
8	9	Star Wars: The Last Jedi	BV	620181382	2017
9	10	Incredibles 2	BV	608581744	2018

- Messy headers
- Un-adjusted Gross
- Limited information



(Jupyter Notebook)



Analysis



Analysis

Questions to Answer -

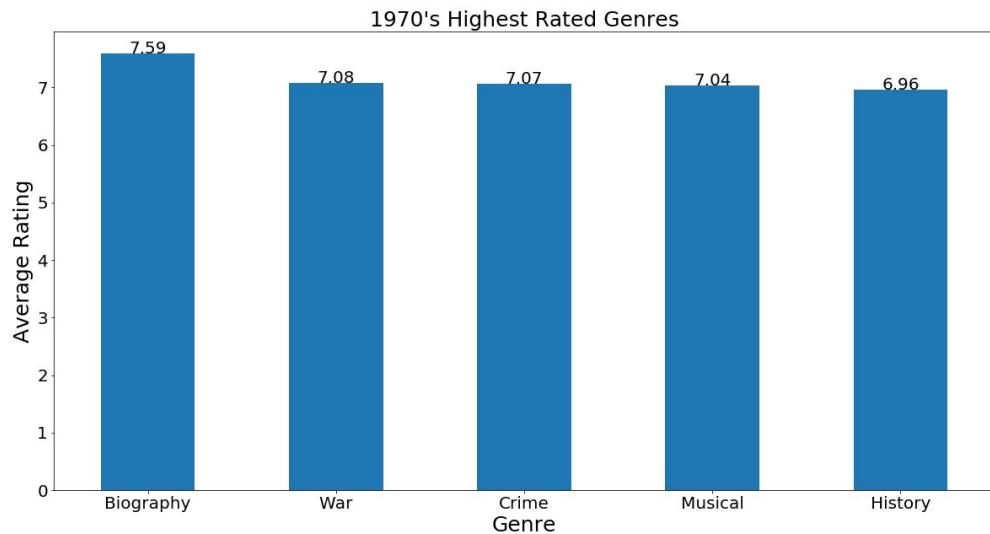
- Do popular genres change over time?
- Does the audience's taste for genre change over time?
- Do seasons have an effect on the metrics of a movie?
- Do seasons impact runtime and audience ratings?
- Has movie revenue (adjusted for inflation) fluctuated through the decades?
- Does duration have any impact on the financial success of a movie?



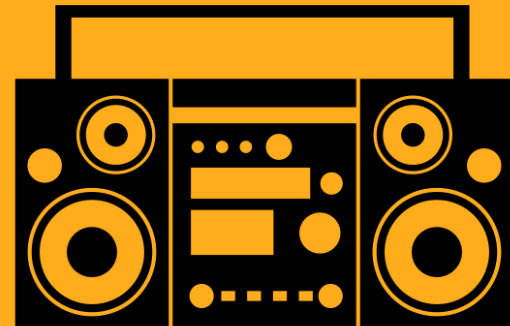
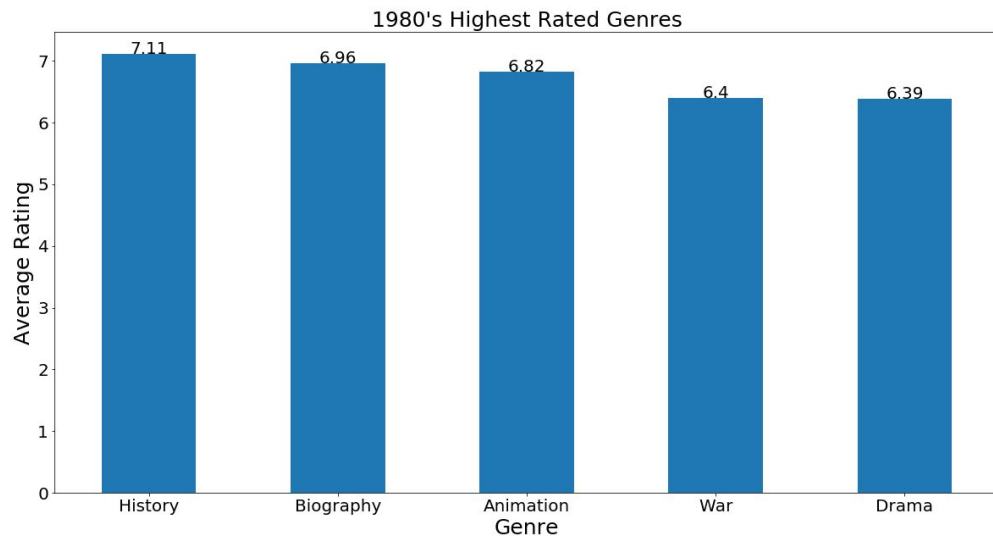
Highest Rated Genre by Decade



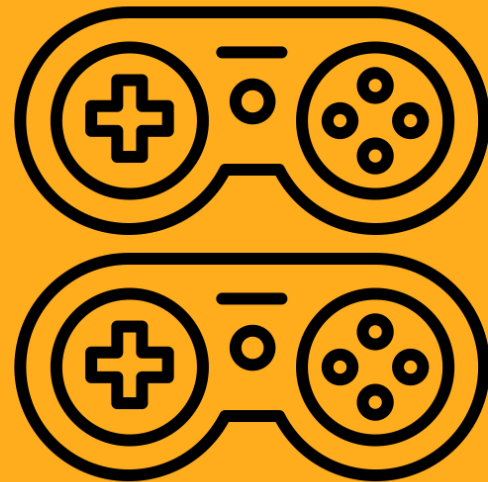
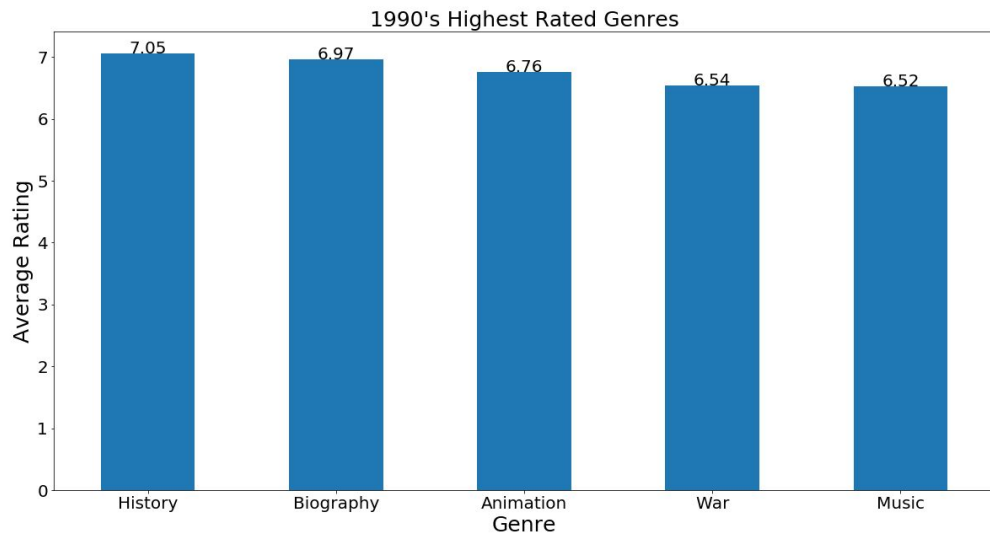
1970's



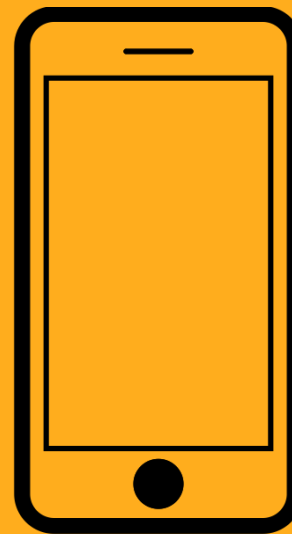
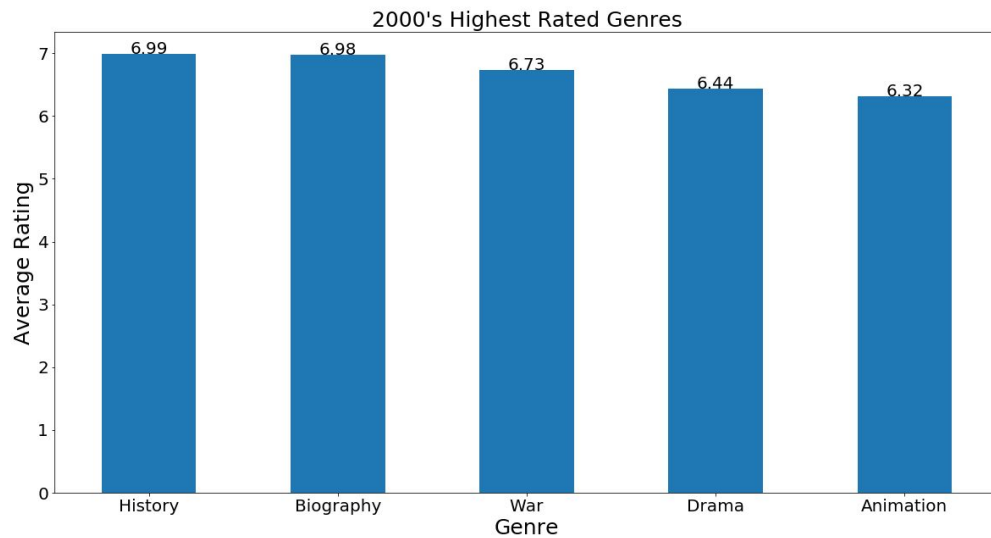
1980's



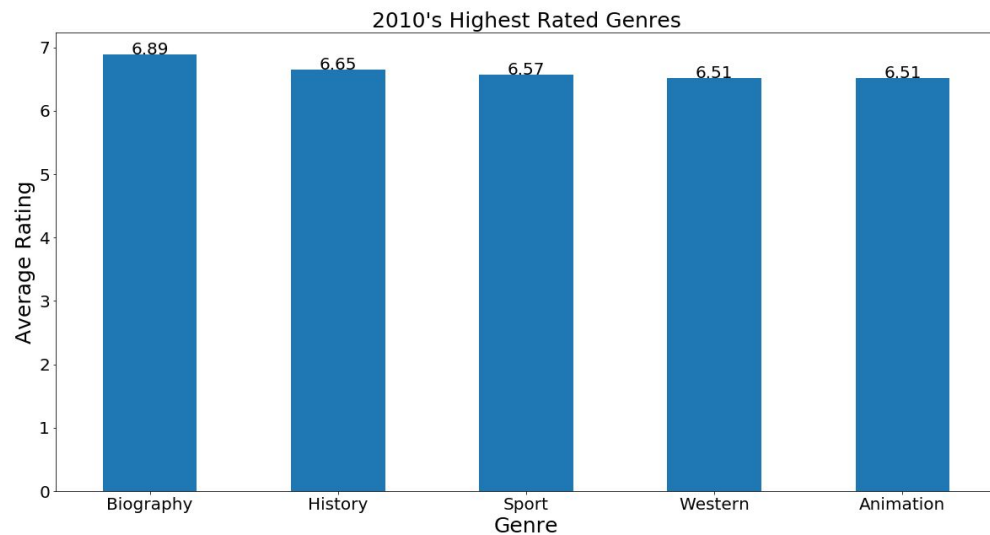
1990's



2000's



2010's

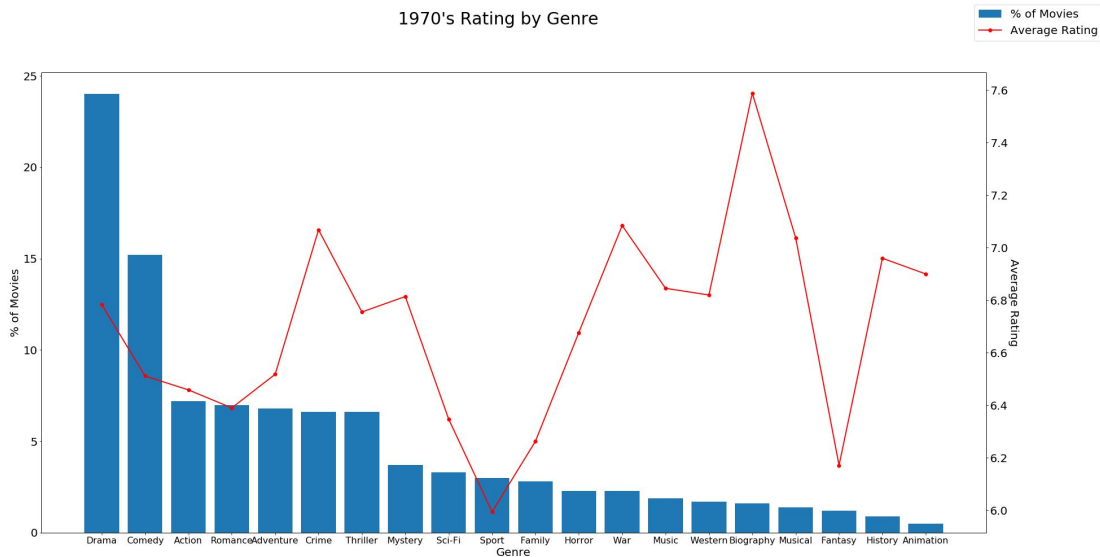




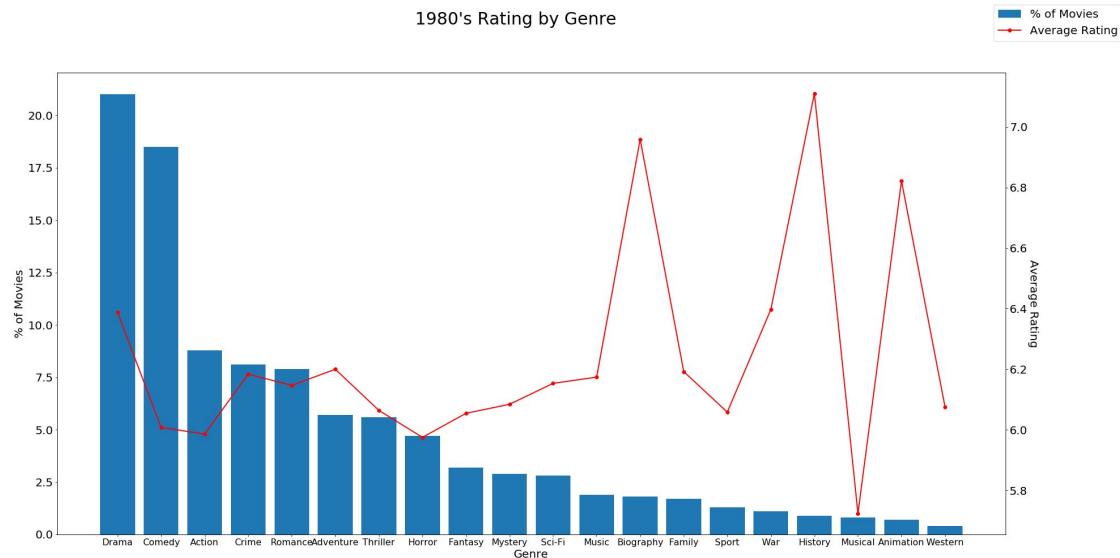
Ratings vs. Popularity (Expanded)



1970's

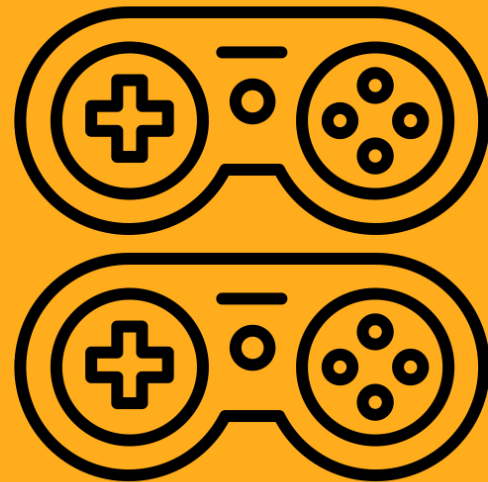
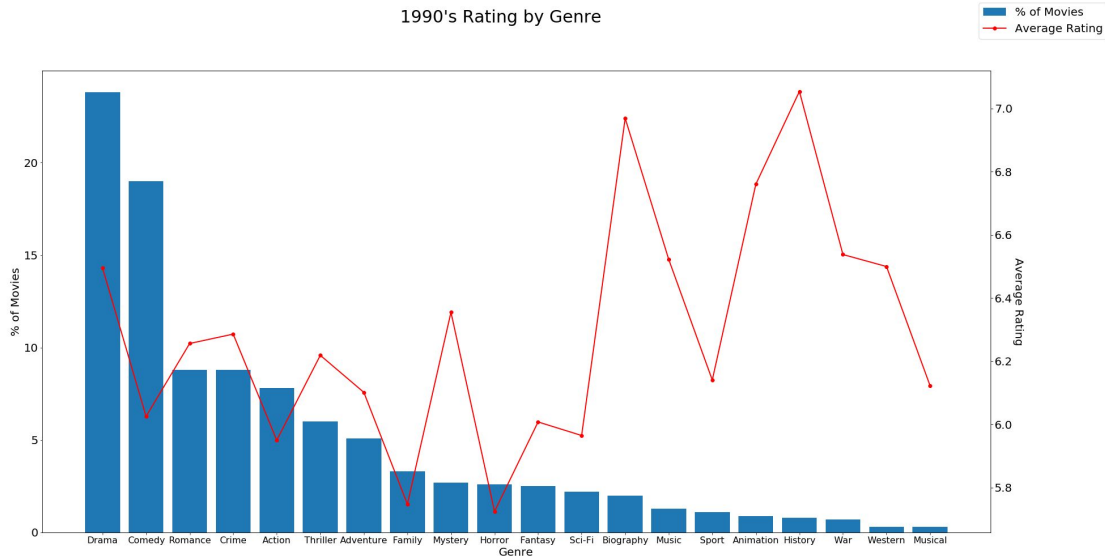


1980's

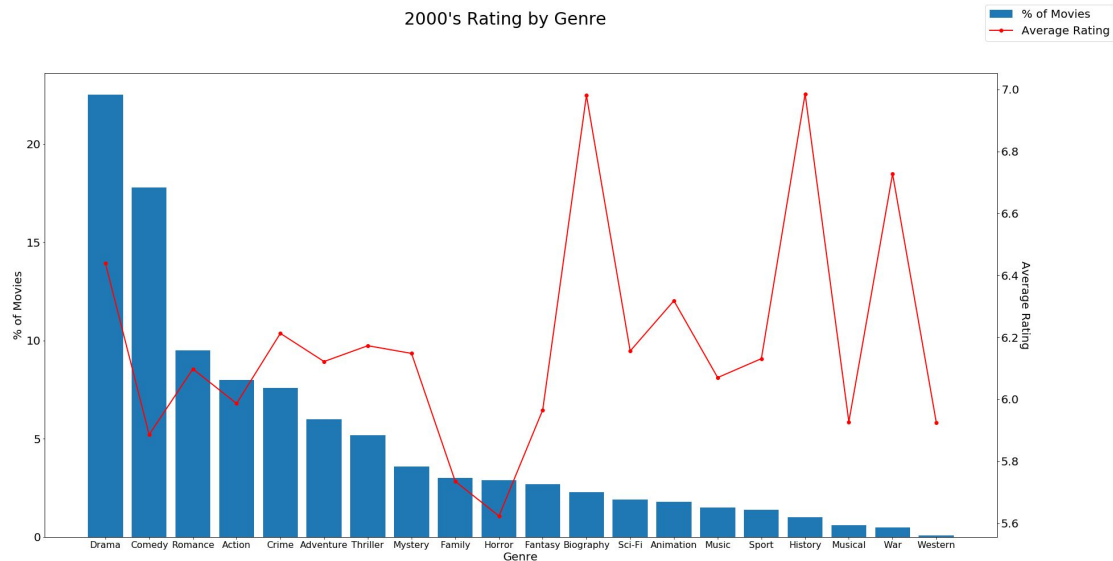


1990's

1990's Rating by Genre

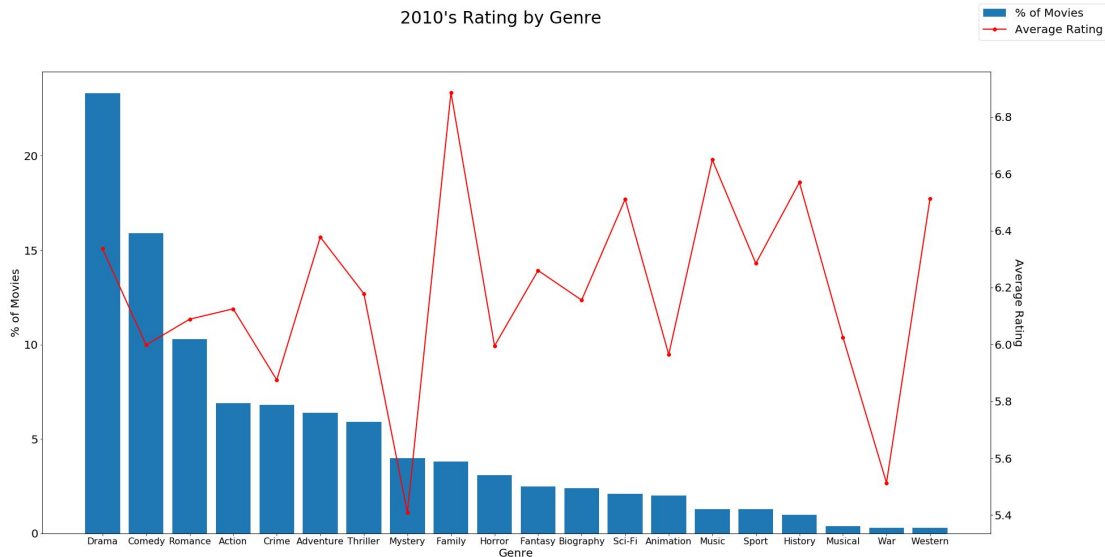


2000's



2010's

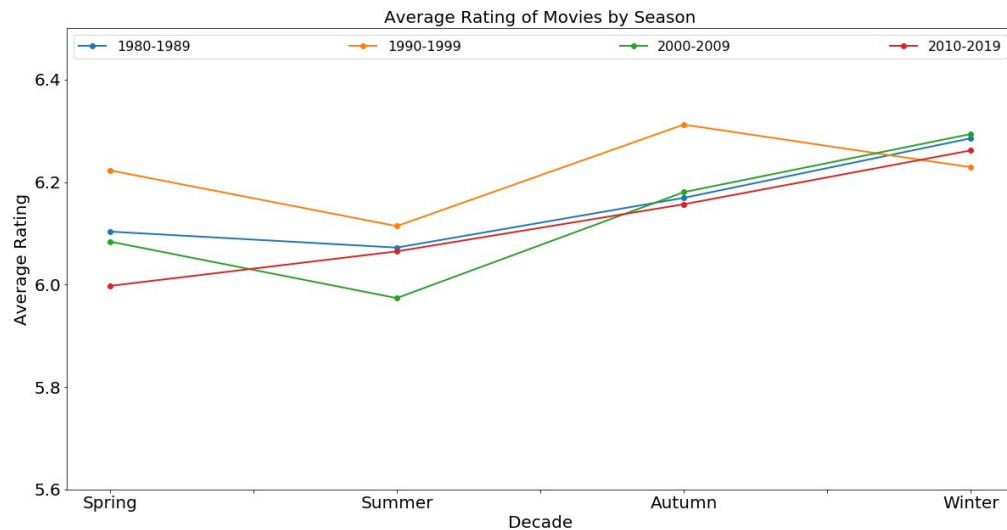
2010's Rating by Genre



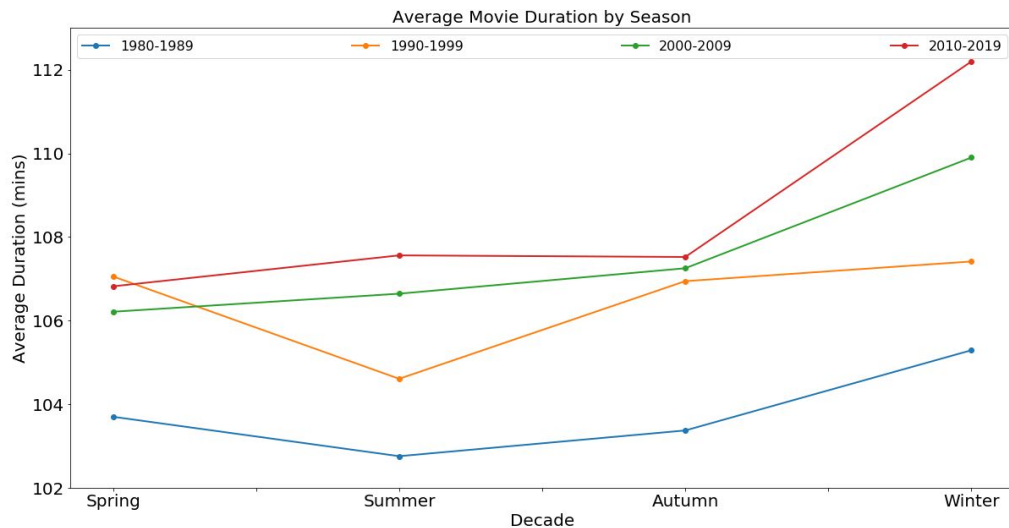


Impact of Seasons on Revenue, Rating, & Runtime

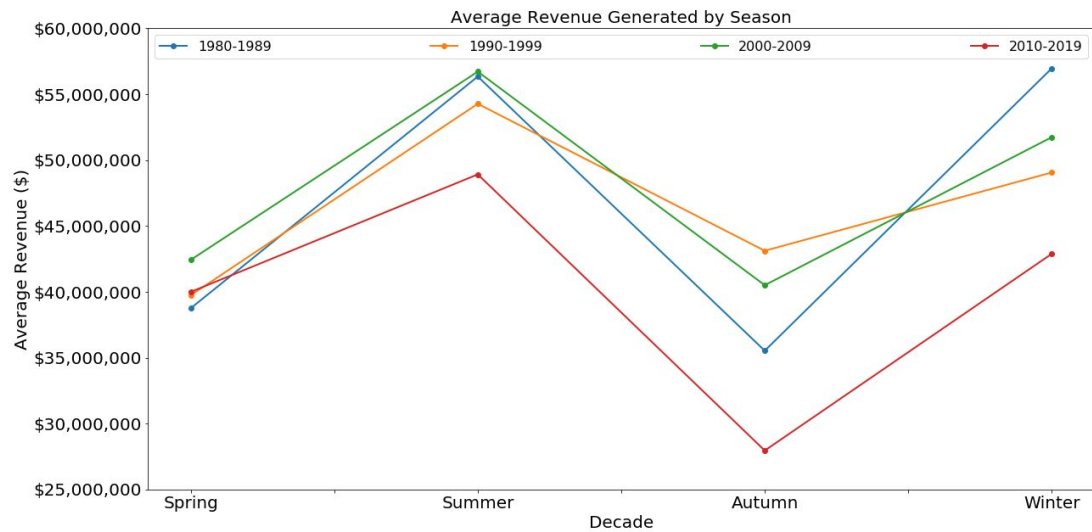
Seasons vs. Rating



Seasons vs. Duration



Seasons vs. Revenue





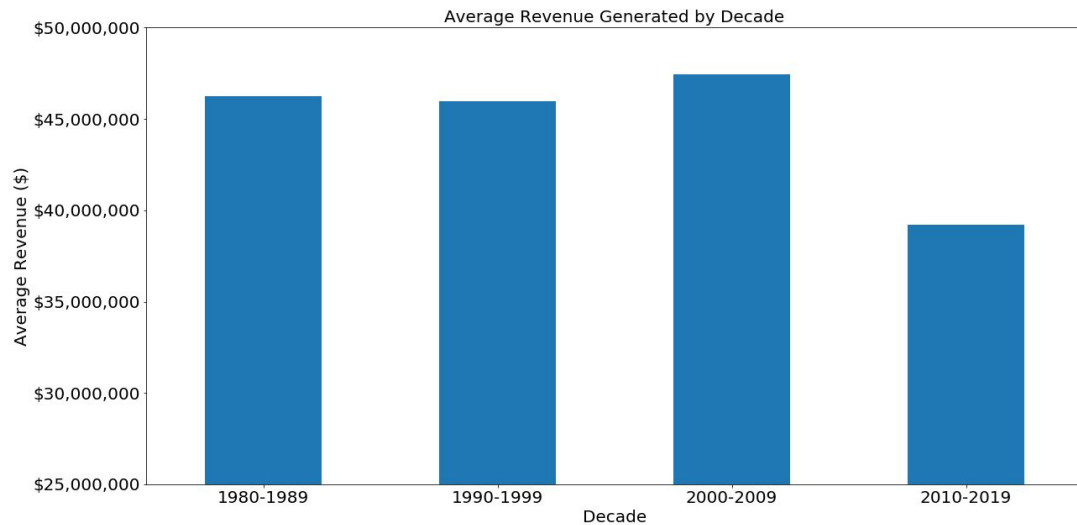
(Jupyter Notebook)



Revenue through the Decades



Avg. Revenue by Decade

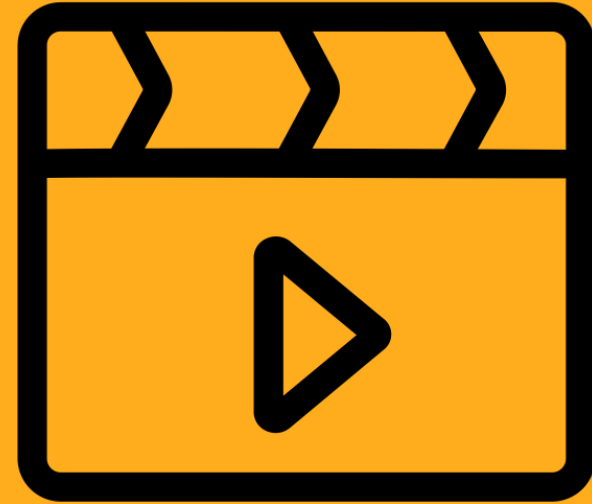
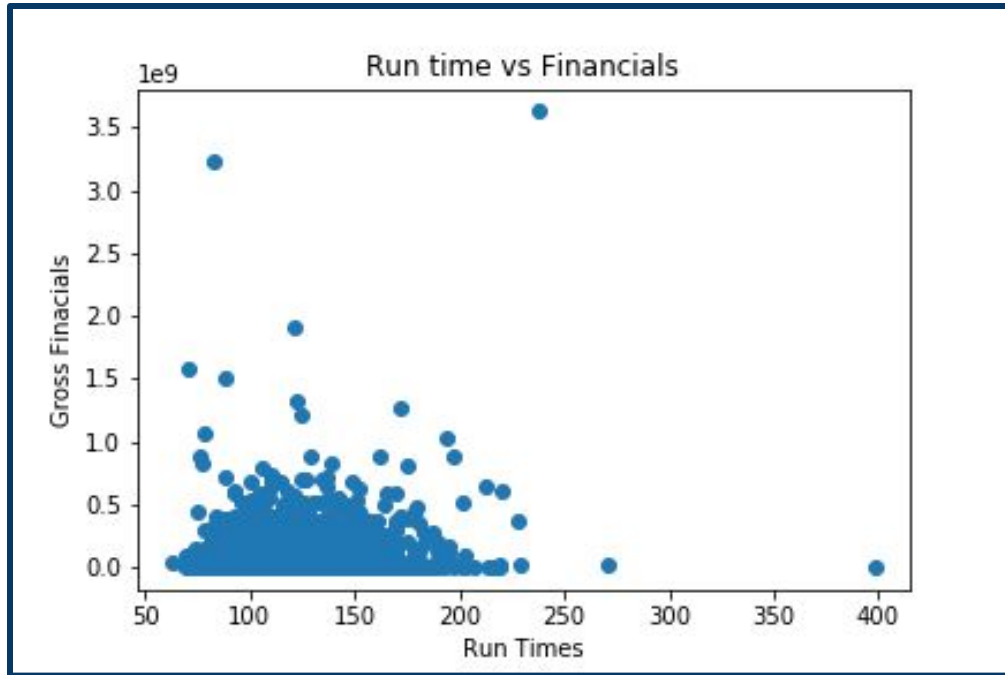


** Adjusted for inflation



Revenue vs. Duration

Revenue vs. Duration





Conclusion & Findings

Things to Ponder

Genres over Time

Not much variation in film genres over the last 50 years. Dramas and Comedies dominate that time frame with 40% of films over that period.

\$\$\$ vs. Critical Acclaim

Summer blockbusters proved to be hits at the box office, but not in the hearts of critics.

Biographies?

Biographies churned out the highest ratings, but weren't players at the box office.

Sign of things to come?

The 2010's saw a major hit in box office revenue. With at-home viewing becoming the new norm, how will this affect films moving forward?

Golden Era

Movies have come a long way, but according to audience ratings the 70's serve as the pinnacle of filmmaking.

Duration vs. Revenue

No significant correlations were found between movie durations and the financial success of a film.



Thank you!