

Microsoft Movie Recommendations

Iman Kumarasinghe

Summary

- Microsoft aims to create film content in their new movie studio
- Three separate dataframes have been analysed
- Gathered from databases such as IMDb, Box Office Mojo, TMDb and The Numbers.
- Three findings and three recommendations from this analysis

Summary

- Adventure and action were the most frequent genres in the movies that made the most revenue from the around the world
- The average runtime for the movies with highest popularity score was around 132 minutes
- There is a positive correlation between budget and revenue - average was \$192 USD million

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

- ❖ Microsoft looking to know which films have done well
- ❖ Asking questions such as
 - What genres were the most popular?
 - How long do successful movies run for
 - How much do movie studios budget

Data

- ❖ Datasets from four different databases
 - IMDb (Internet Movie Database)
 - Box Office Mojo
 - TMDB (The Movie Database)
 - The Numbers

Dataframe 1

:

	title	studio	domestic_gross	foreign_gross	year	tconst	primary_title	original_title	start_year	runtime_minutes	genres
0	Toy Story 3	BV	415000000.0	652000000	2010	tt0435761	Toy Story 3	Toy Story 3	2010	103.0	Adventure,Animation,Comedy
1	Inception	WB	292600000.0	535700000	2010	tt1375666	Inception	Inception	2010	148.0	Action,Adventure,Sci-Fi
2	Shrek Forever After	P/DW	238700000.0	513900000	2010	tt0892791	Shrek Forever After	Shrek Forever After	2010	93.0	Adventure,Animation,Comedy
3	The Twilight Saga: Eclipse	Sum.	300500000.0	398000000	2010	tt1325004	The Twilight Saga: Eclipse	The Twilight Saga: Eclipse	2010	124.0	Adventure,Drama,Fantasy
4	Iron Man 2	Par.	312400000.0	311500000	2010	tt1228705	Iron Man 2	Iron Man 2	2010	124.0	Action,Adventure,Sci-Fi
...
3361	Souvenir	Strand	11400.0	NaN	2018	tt2389092	Souvenir	Souvenir	2014	86.0	Comedy,Romance
3362	Souvenir	Strand	11400.0	NaN	2018	tt3478898	Souvenir	Souvenir	2014	86.0	Documentary
3363	Beauty and the Dogs	Osci.	8900.0	NaN	2018	tt6776572	Beauty and the Dogs	Aala Kaf Ifrit	2017	100.0	Crime,Drama,Thriller
3364	The Quake	Magn.	6200.0	NaN	2018	tt6523720	The Quake	Skjelvet	2018	106.0	Action,Drama,Thriller
3365	An Actor Prepares	Grav.	1700.0	NaN	2018	tt5718046	An Actor Prepares	An Actor Prepares	2018	97.0	Comedy

Dataframe 2

	title	studio	domestic_gross	foreign_gross	year	tconst	primary_title	original_title_x	start_year	runtime_minutes	...	domestic_and_foreign
3454	Avengers: Infinity War	BV	678800000.0	1.369500e+09	2018	tt4154756	Avengers: Infinity War	Avengers: Infinity War	2018	149.0	...	2.048300e+09
1577	John Wick	LG/S	43000000.0	4.570000e+07	2014	tt2911666	John Wick	John Wick	2014	101.0	...	8.870000e+07
1494	The Hobbit: The Battle of the Five Armies	WB (NL)	255100000.0	7.009000e+08	2014	tt2310332	The Hobbit: The Battle of the Five Armies	The Hobbit: The Battle of the Five Armies	2014	144.0	...	9.560000e+08
1495	Guardians of the Galaxy	BV	333200000.0	4.402000e+08	2014	tt2015381	Guardians of the Galaxy	Guardians of the Galaxy	2014	121.0	...	7.734000e+08
3074	Blade Runner 2049	WB	92100000.0	1.672000e+08	2017	tt1856101	Blade Runner 2049	Blade Runner 2049	2017	164.0	...	2.593000e+08
3073	Blade Runner 2049	WB	92100000.0	1.672000e+08	2017	tt1856101	Blade Runner 2049	Blade Runner 2049	2017	164.0	...	2.593000e+08
3463	Fantastic Beasts: The Crimes of Grindelwald	WB	159600000.0	4.941000e+08	2018	tt4123430	Fantastic Beasts: The Crimes of Grindelwald	Fantastic Beasts: The Crimes of Grindelwald	2018	134.0	...	6.537000e+08
3468	Ralph Breaks the Internet	BV	201100000.0	3.281000e+08	2018	tt5848272	Ralph Breaks the Internet	Ralph Breaks the Internet	2018	112.0	...	5.292000e+08
3008	Spider-Man: Homecoming	Sony	334200000.0	5.460000e+08	2017	tt2250912	Spider-Man: Homecoming	Spider-Man: Homecoming	2017	133.0	...	8.802000e+08
3464	Ant-Man and the Wasp	BV	216600000.0	4.060000e+08	2018	tt5095030	Ant-Man and the Wasp	Ant-Man and the Wasp	2018	118.0	...	6.226000e+08
1880	Avengers: Age of Ultron	BV	459000000.0	9.464000e+08	2015	tt2395427	Avengers: Age of Ultron	Avengers: Age of Ultron	2015	141.0	...	1.405400e+09

Dataframe 3

	id	release_date	movie	production_budget	domestic_gross	worldwide_gross	domestic_worldwide	worldwide_minus_budget
0	1	Dec 18, 2009	Avatar	425000000.0	760507625.0	2.776345e+09	3.536853e+09	3.111853e+09
5	6	Dec 18, 2015	Star Wars Ep. VII: The Force Awakens	306000000.0	936662225.0	2.053311e+09	2.989973e+09	2.683973e+09
42	43	Dec 19, 1997	Titanic	200000000.0	659363944.0	2.208208e+09	2.867572e+09	2.667572e+09
6	7	Apr 27, 2018	Avengers: Infinity War	300000000.0	678815482.0	2.048134e+09	2.726950e+09	2.426950e+09
33	34	Jun 12, 2015	Jurassic World	215000000.0	652270625.0	1.648855e+09	2.301125e+09	2.086125e+09
26	27	May 4, 2012	The Avengers	225000000.0	623279547.0	1.517936e+09	2.141215e+09	1.916215e+09
41	42	Feb 16, 2018	Black Panther	200000000.0	700059566.0	1.348258e+09	2.048318e+09	1.848318e+09
66	67	Apr 3, 2015	Furious 7	190000000.0	353007020.0	1.518723e+09	1.871730e+09	1.681730e+09
43	44	Jun 15, 2018	Incredibles 2	200000000.0	608581744.0	1.242521e+09	1.851102e+09	1.651102e+09
4	5	Dec 15, 2017	Star Wars Ep. VIII: The Last Jedi	317000000.0	620181382.0	1.316722e+09	1.936903e+09	1.619903e+09
134	35	Mar 17, 2017	Beauty and the Beast	160000000.0	504014165.0	1.259200e+09	1.763214e+09	1.603214e+09
260	61	Jul 15, 2011	Harry Potter and the Deathly Hallows: Part II	125000000.0	381193157.0	1.341693e+09	1.722886e+09	1.597886e+09
112	13	Jun 22, 2018	Jurassic World: Fallen Kingdom	170000000.0	417719760.0	1.305773e+09	1.723493e+09	1.553493e+09
3	4	May 1, 2015	Avengers: Age of Ultron	330600000.0	459005868.0	1.403014e+09	1.862020e+09	1.531420e+09
155	56	Nov 22, 2013	Frozen	150000000.0	400738009.0	1.272470e+09	1.673208e+09	1.523208e+09
425	26	Dec 17, 2003	The Lord of the Rings: The Return of the King	94000000.0	377845905.0	1.141403e+09	1.519249e+09	1.425249e+09

Methods

- ❖ Cleaning columns in domestic and foreign gross due to commas and NaNs
 - Removed the commas and changed NaNs to 0
 - Made sure the values had the correct number of digits
 - Cross checked the changed values with website data
- ❖ Combined domestic and foreign gross for worldwide revenue
- ❖ Cleaned commas from genres and generated a list within the top 15
- ❖ Cleaned commas and dollar signs from budget and revenue
- ❖ Calculated profit

Methods

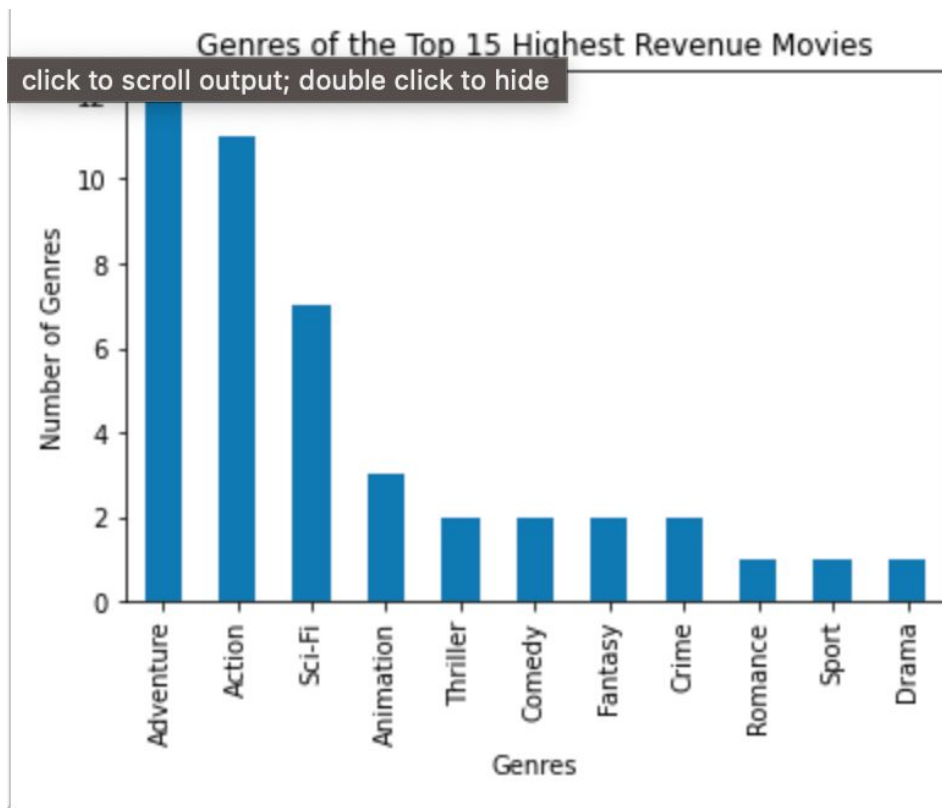
- ❖ Sorted the data by
 - which made the most revenue
 - the most popular
 - which made the most profit

- ❖ Gathered the top amounts e.g. top 15, top 30

Results from Dataframe One

- Top 15 movies included Avengers, Jurassic World, Furious 7, Black Panther, Frozen, Incredibles, The Fate of the Furious, Iron Man 3, Minions and Capitan America.

Results from Dataframe One

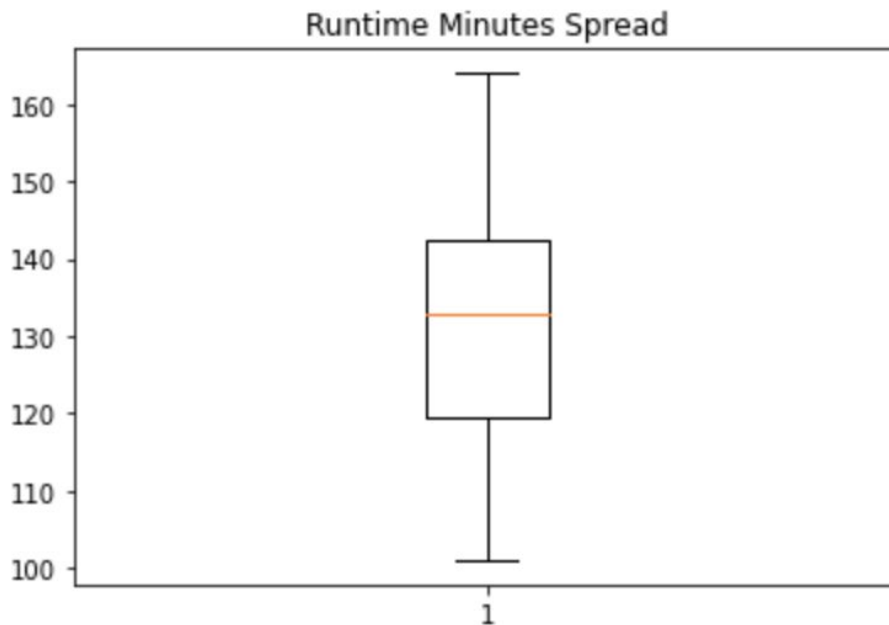


Results from Dataframe Two

- Highest popularity score movies included Avengers, John Wick, Spiderman, The Hobbit, Blade Runner, Guardians of the Galaxy, Logan, Ralph Breaks the Internet

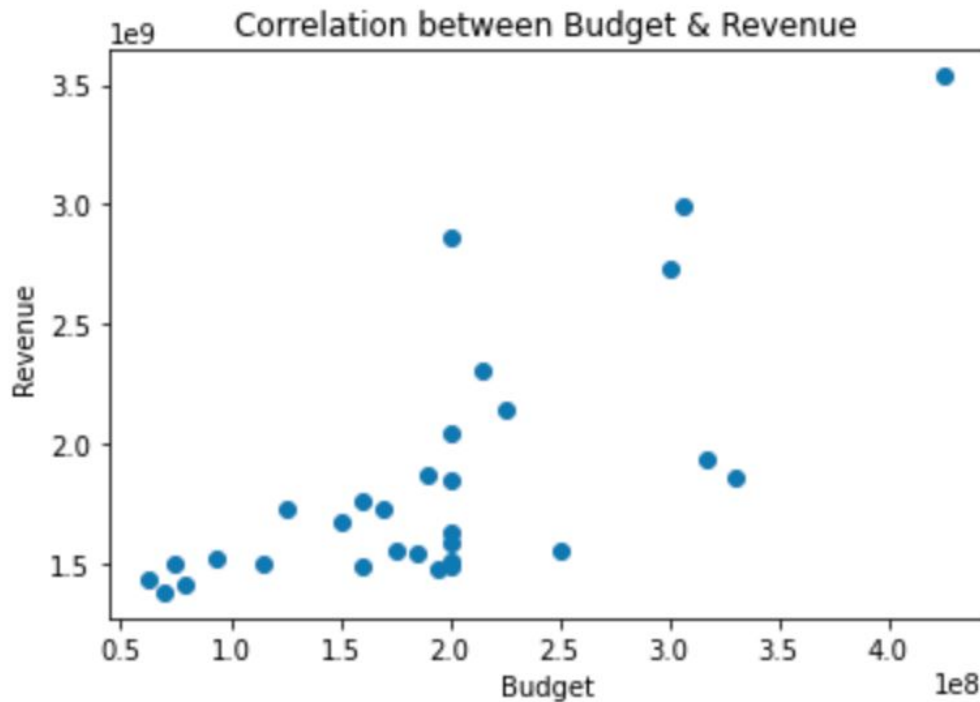
Results from Dataframe Two

- Runtimes ranged from 101 minutes to 164 minutes
- Mean of 132 minutes



Results from Dataframe Three

- Correlation coefficient of 0.73
- Average budget of \$192 million USD



Recommendations

- Recommendation One: To create a movie that includes the genres of Adventure or Action
- Recommendation Two: Create a movie around 130 minutes
- Recommendation Three: Invest around \$192 mil for a high return

Limitations

- Different movies in different datasets
- Data didnt include that from streaming sites
- Movies looked at weren't as recent
-

Future Analysis

- Actors & Actresses
- Popularity of movies streamed online

Thank You!