Resolute Films

Ealija Abbas

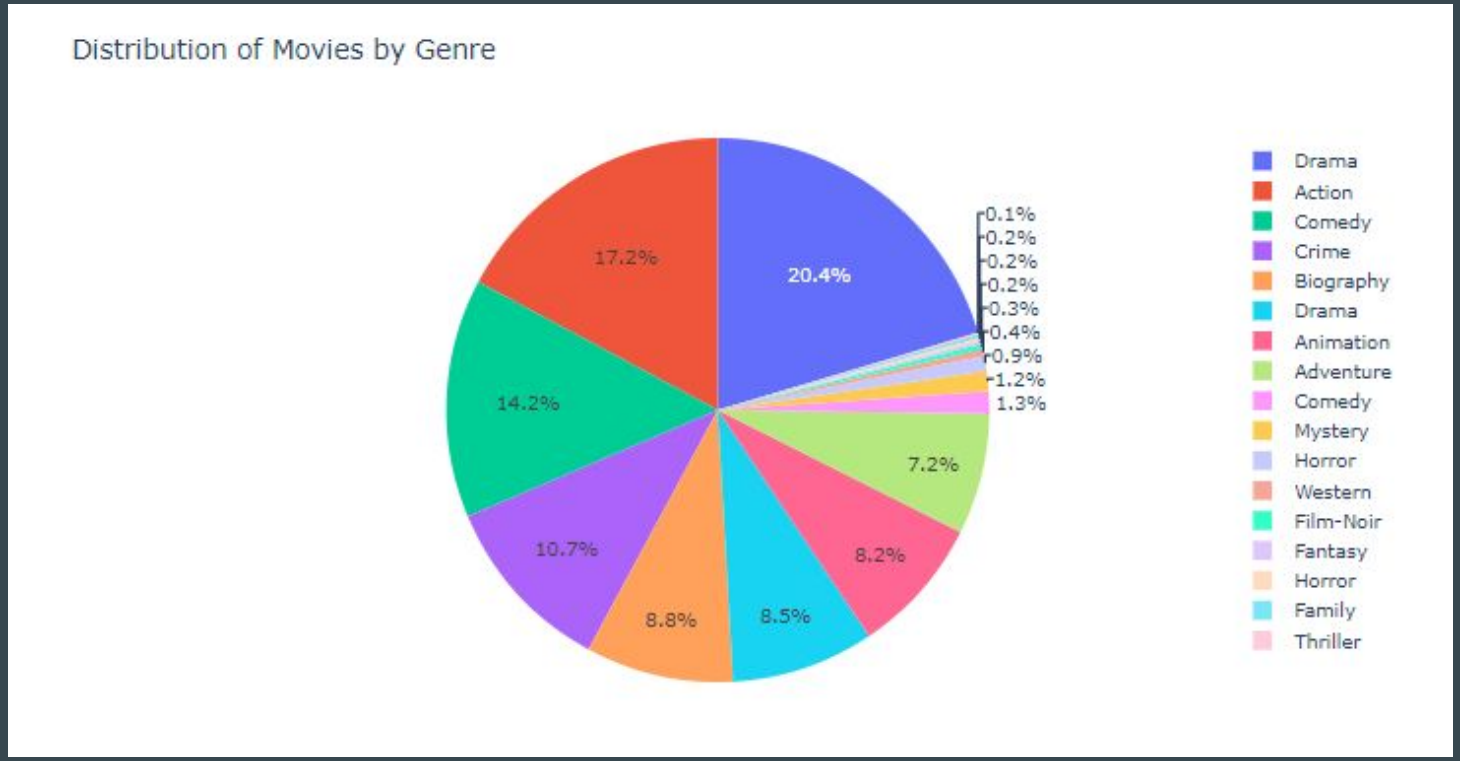
The problem statement.

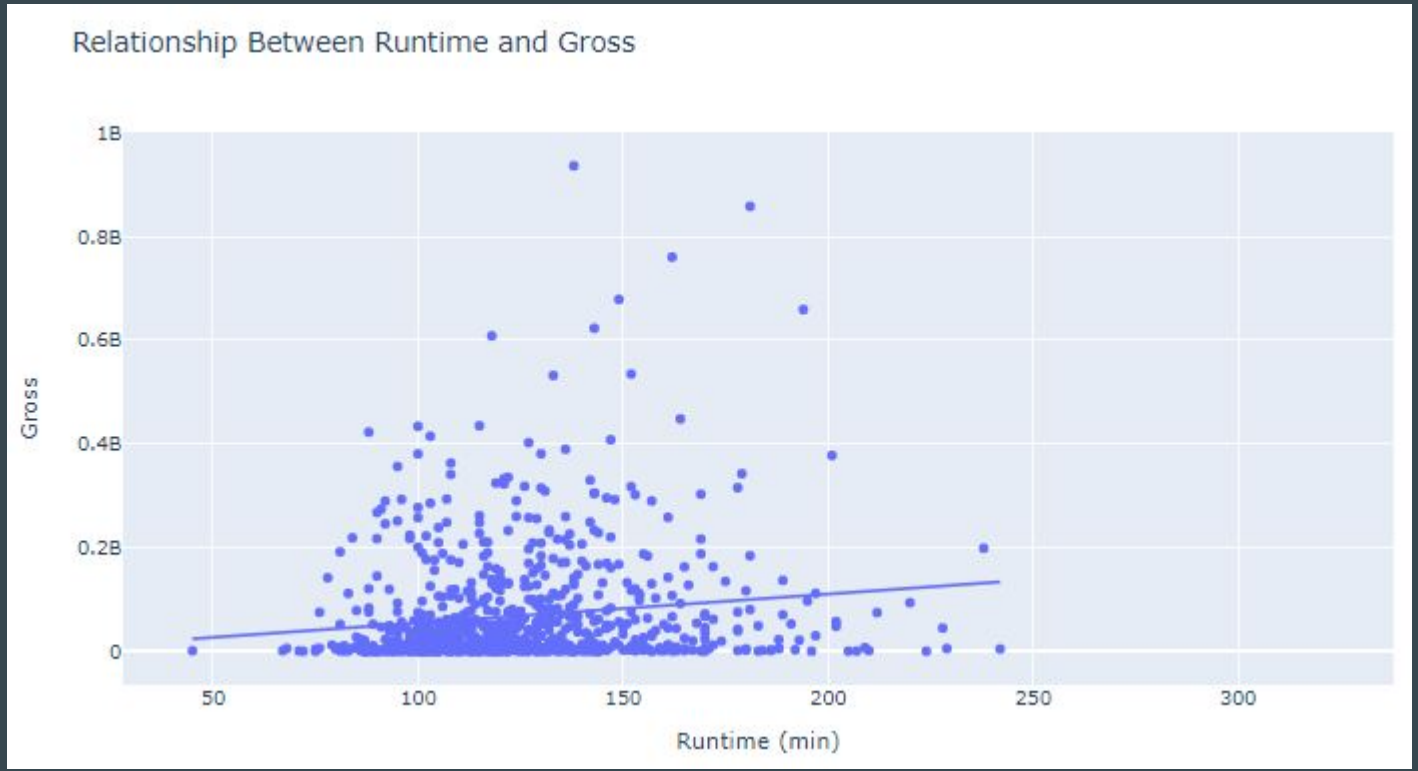
- Resolute Films ™ is a new Cinema in the Richmond Area that primarily shows top IMBD movies of the past. Before they start showing movies they asked me to help them find out which movies would be best.
- My use of python to find out which movies are best to show will help Resolute Cinema have success and not go bankrupt.

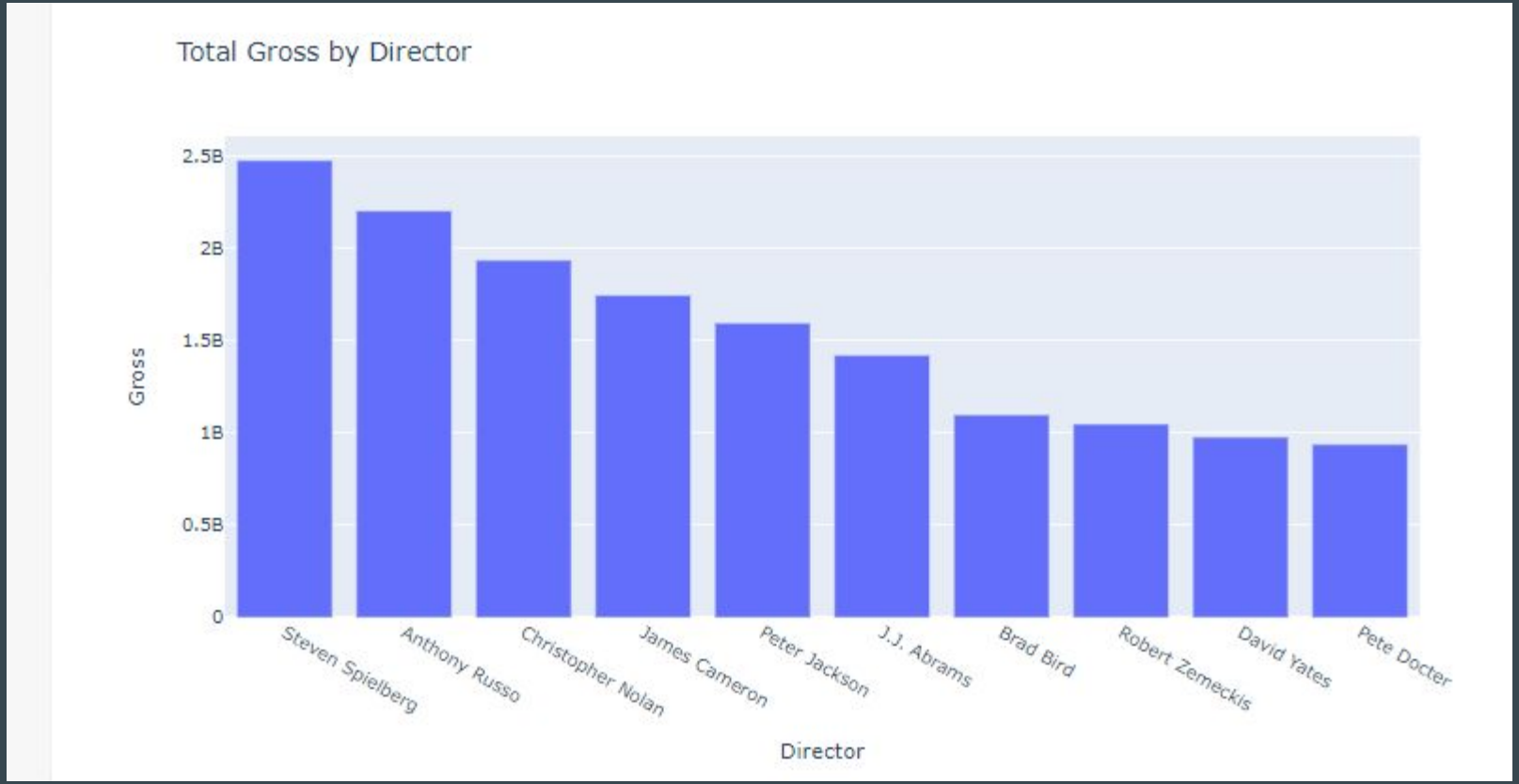
Methodology

- Data collected from Kaggle
- Data was cleaned. merged, and aggregated
- Used pandas to read in data
- Visualizations were created with plotly, mathplotlib, and seaborn
- Used a linear regression model to predict the movie gross based on ratings, production budget, and other variables.
- Had some trouble with missing values in gross and Meta_score columns. To get around this we substituted mean values for missing data.

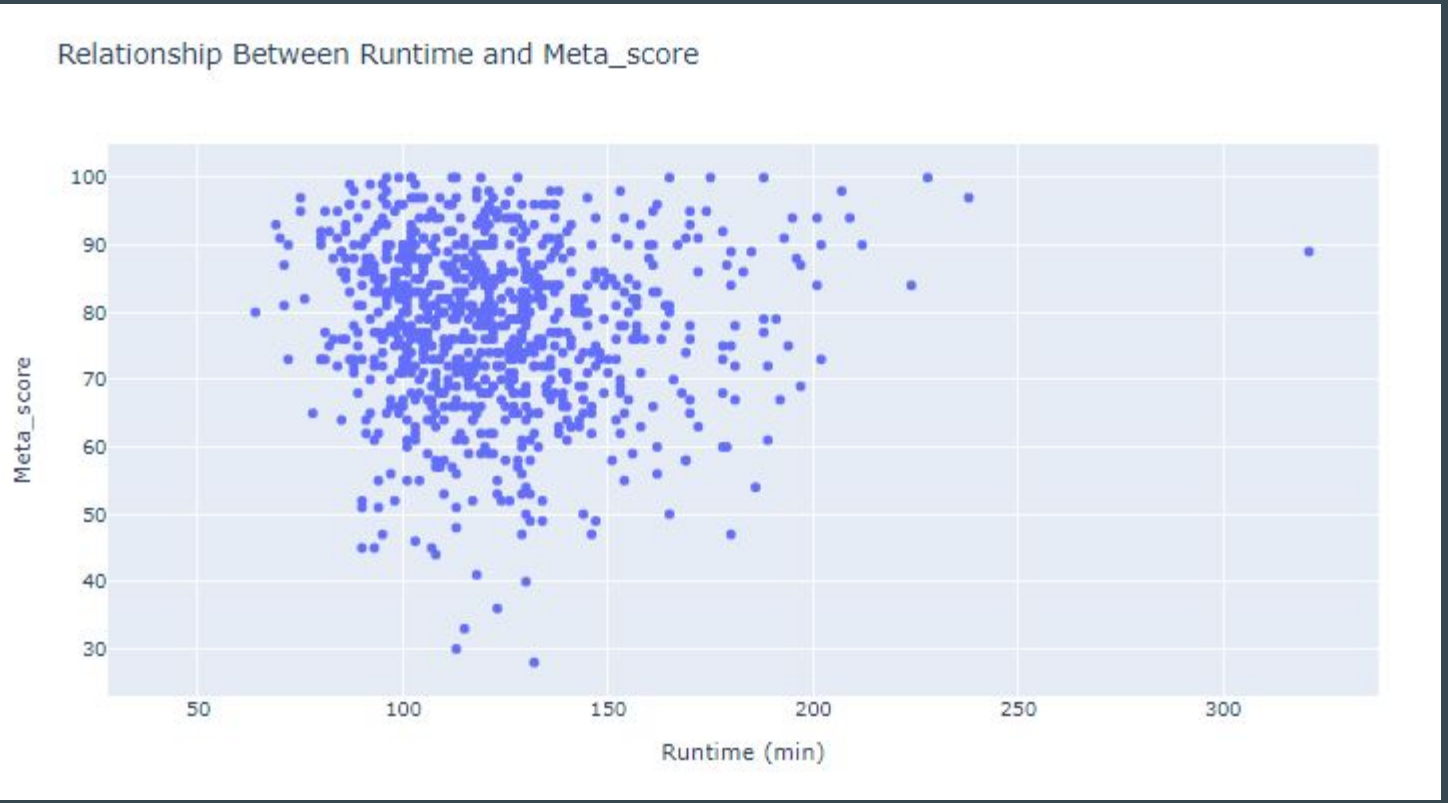
Visualizations




Relationship Between Runtime and Gross
1B
0.8B
0.6B
0.4B
0.2B
0
Gross
50
100
150
200
250
300
Runtime (min)


Total Gross by Director
Gross
2.5B
2B
1.5B
1B
0.5B
0
Steven Spielberg
Anthony Russo
Christopher Nolan
James Cameron
Peter Jackson
J.J. Abrams
Brad Bird
Robert Zemeckis
David Yates
Pete Docter
Director


Trend of Gross over the Years
Gross
3B
2.5B
2B
1.5B
1B
0.5B
0
1920
1922
1925
1927
1930
1932
1934
1936
1938
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
Released_Year


Relationship Between Runtime and Meta_score
Meta_score
100
90
80
70
60
50
40
30
50
100
150
200
250
300
Runtime (min)


Distribution of IMDB Ratings by Certificate
IMDB_Rating
9.2
9
8.8
8.6
8.4
8.2
8
7.8
7.6
A
UA
U
PG-13
R
PG
G
Passed
TV-14
16
TV-MA
Unrated
GP
Approved
TV-PG
U/A
Certificate

	Genre	Gross	Runtime (min)
0	Action	2.001680e+10	129.046512
1	Adventure	5.273754e+09	134.111111
2	Animation	8.573824e+09	99.585366
3	Biography	4.750170e+09	136.022727
4	Comedy	6.200388e+08	95.000000
5	Comedy	3.544773e+09	113.697183
6	Crime	3.179785e+09	126.392523
7	Drama	2.020898e+09	127.211765
8	Drama	7.029586e+09	123.705882
9	Family	4.391106e+08	107.500000
10	Fantasy	0.000000e+00	85.000000
11	Film-Noir	2.557251e+06	104.000000
12	Horror	2.355221e+08	111.000000
13	Horror	5.003356e+08	100.111111
14	Mystery	2.739558e+08	119.083333
15	Thriller	1.755074e+07	108.000000
16	Western	5.822151e+07	148.250000

	Director	IMDB_Rating
0	Aamir Khan	8.40
1	Aaron Sorkin	7.80
2	Abdellatif Kechiche	7.70
3	Abhishek Chaubey	7.80
4	Abhishek Kapoor	7.70
..	...	...
543	Zack Snyder	7.60
544	Zaza Urushadze	8.20
545	Zoya Akhtar	8.05
546	Çagan Irmak	8.30
547	Ömer Faruk Sorak	8.00

[548 rows x 2 columns]

	Series_Title	IMDB_Rating_x	Gross_x	IMDB_Rating_y	Gross_y
0	The Shawshank Redemption	9.3	28341469.0	NaN	NaN
1	The Godfather	9.2	134966411.0	NaN	NaN
2	The Dark Knight	9.0	534858444.0	9.0	534858444.0
3	The Godfather: Part II	9.0	57300000.0	NaN	NaN
4	12 Angry Men	9.0	4360000.0	NaN	NaN

Issues to overcome

5]:
```
#Added mean values to missing data points in GRoss
mean_gross = df["Gross"].mean()
df["Gross"].fillna(mean_gross, inplace=True)
```

6]:
```
#Added mean values to missing data points in meta_score

df["Meta_score"].fillna(df["Meta_score"].mean(), inplace=True)
```

Findings

- We found that the genre of a movie is the most significant predictor of its success. We also found that movies with higher budgets tend to perform better at the box office.

- Our findings support the theory that the genre of a movie is a significant predictor of success. However, our analysis also shows that other factors, such as budget and rating, can also contribute to a movie's success.

- Our machine learning model achieved an R score of .62 which suggests that our model is decent at predicting the success of a movie based on its features.

Recommendations

Movie producers should focus on creating movies in popular genres, such as action and comedy. They should also allocate larger budgets to movies with high potential for success.

Our analysis provides valuable insights into the factors that contribute to a movie's success. This information can be used by movie producers to make more informed decisions about which movies to produce and how much to invest in them.

Future Work

In the future, we plan to expand our analysis to include more variables, such as cast and crew data. We also plan to analyze the success of movies in different regions of the world to identify regional trends

By analyzing more variables, we can create more accurate models that predict the success of a movie. We can also improve the accuracy of our models by using more advanced machine learning algorithms