MOVIE REVENUE PREDICTION

Chetana Vyas

Modelling Framework

Box Office Mojo Website **(Domestic Gross)** : https://www.boxofficemojo.com/

Original data: 6360 data points + 11 features

Final data: 2199 data points + 46 features

REGRESSION ALGORITHMS

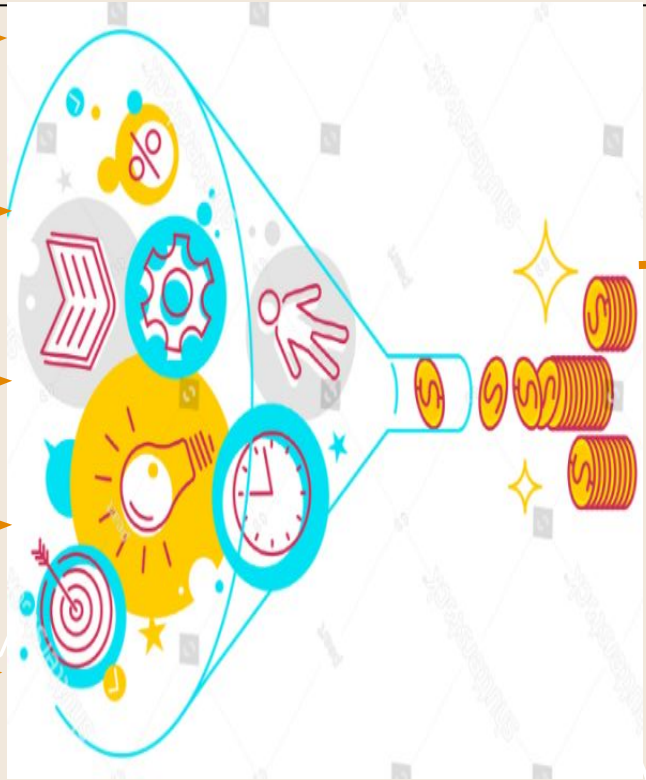
Budget $$$

MPAA Rating
[G, PG, PG-13, R,..]

Running Time

Release Month
(Seasonality changes)

Genres
(Animation, Comedy, Romance,...)



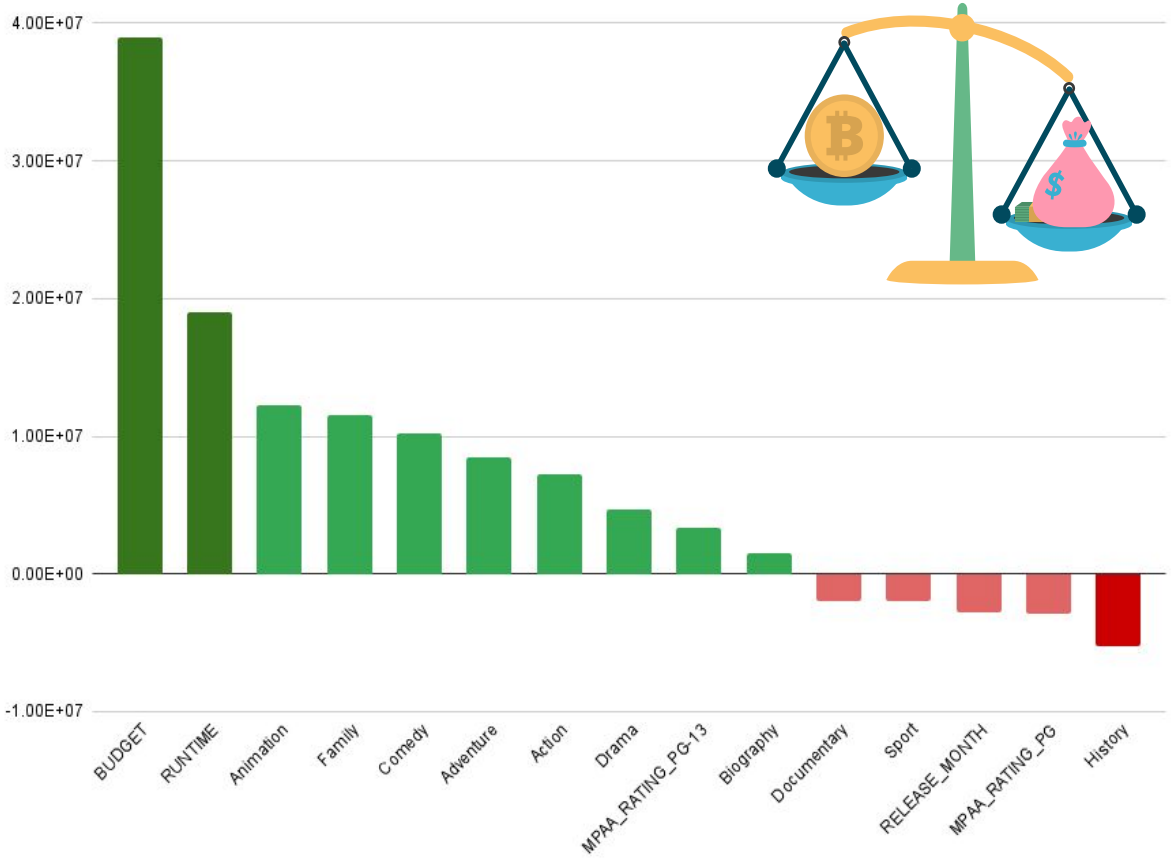
PREDICTED REVENUE $$$

	Final Model
R^2	0.47
RMSE	$ 62 MIL

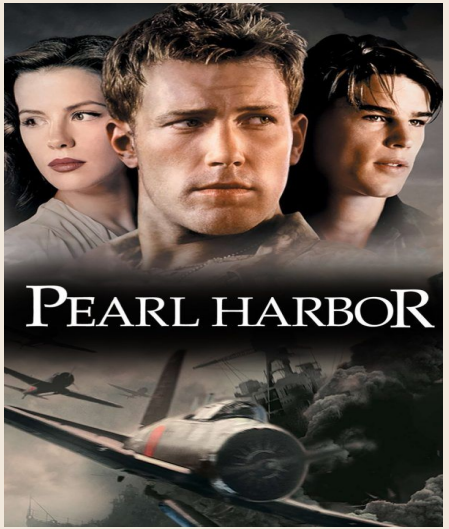
Rank, Domestic Opening, IMDB Rating ?

Recommendations

- High Budget Movie $$$
- Genres (Animation, Comedy, Action)
- MPAA Rating PG-13
- Avoid Winter releases. Consider Summer or Holiday Blockbuster release

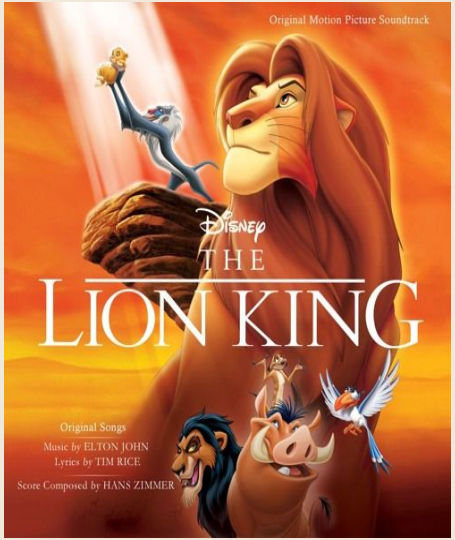


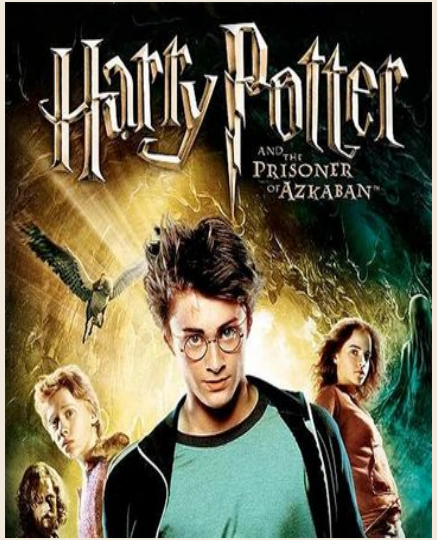
MODEL PREDICTICTIONS

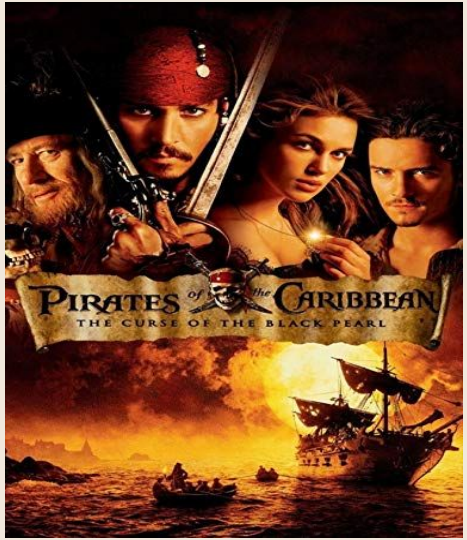


ACTUAL	$ 161 M	$ 209 M	$ 144 M	$ 199 M
PREDICTED	$ 153 M	$ 196 M	$ 146 M	$ 177 M
DIFF	$ 8 (5%)	$ 13 M (6%)	$ 13 M (8%)	$ 22 M (11%)

WHAT HAPPENED HERE?



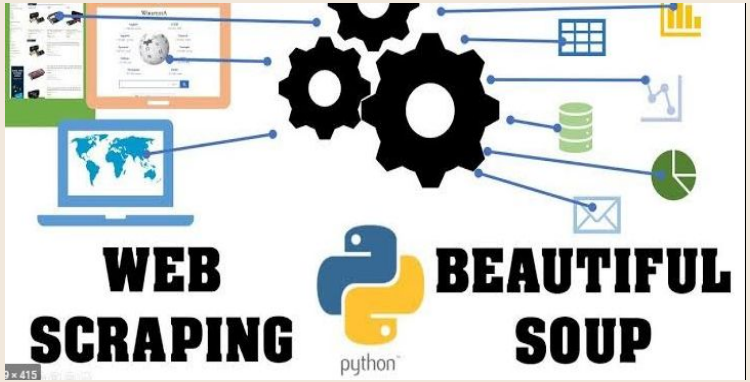




ACTUAL	$ 544 M	$ 243 M	$ 305 M	$ 652 M
PREDICTED	$ 313 M	$ 110 M	$ 189 M	$ 185 M
DIFF	$ 231 M (42%)	$ 133 M (54%)	$ 116 M (38%)	$ 468 M (72%)

Can we build the perfect model (say $R^2 = 1$) if we had all possible pre release movie features?

APPENDIX - TOOLS



NumPy

matplotlib

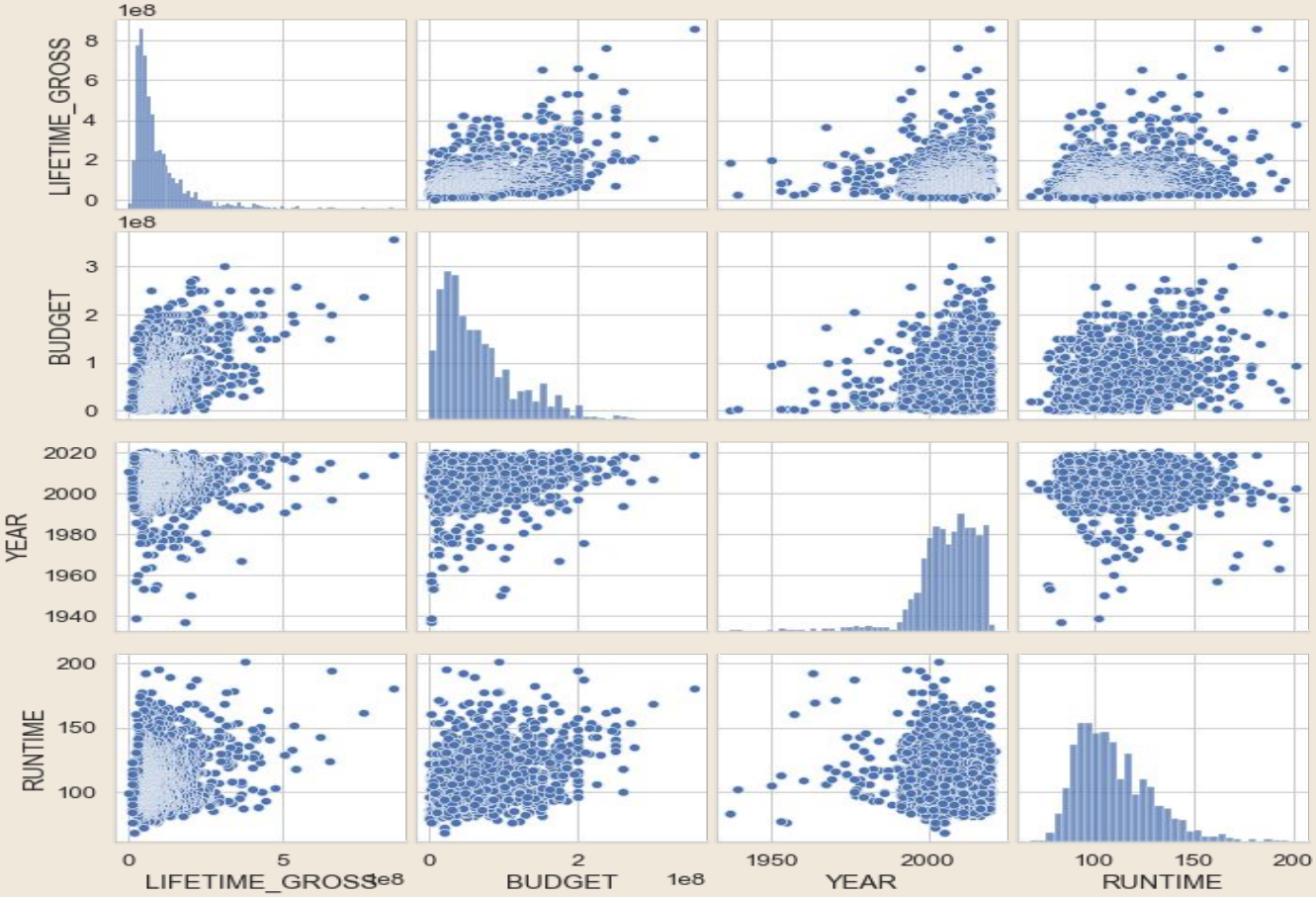
Pandas

scikit learn

Seaborn

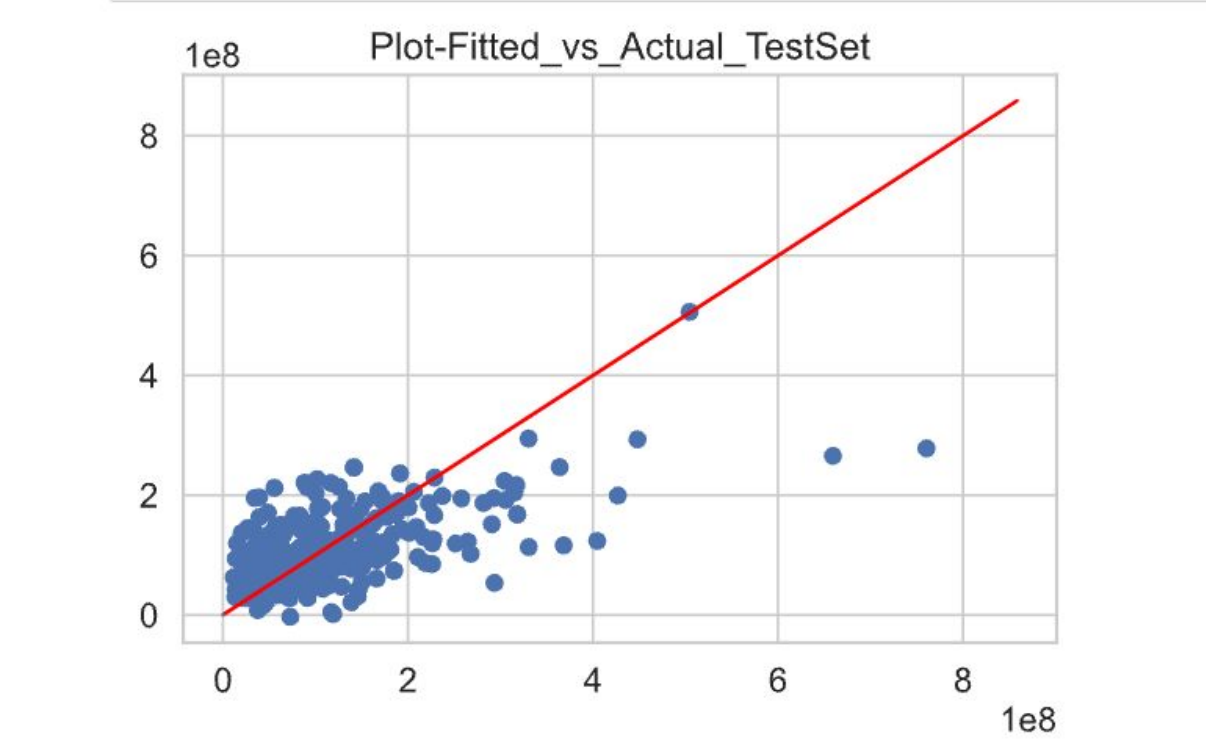
statsmodels

APPENDIX


LIFETIME_GROSS
BUDGET
YEAR
RUNTIME
1e8
1e8
1e8
0
5
0
2
1950
2000
100
150
200
LIFETIME_GROSS
BUDGET
YEAR
RUNTIME

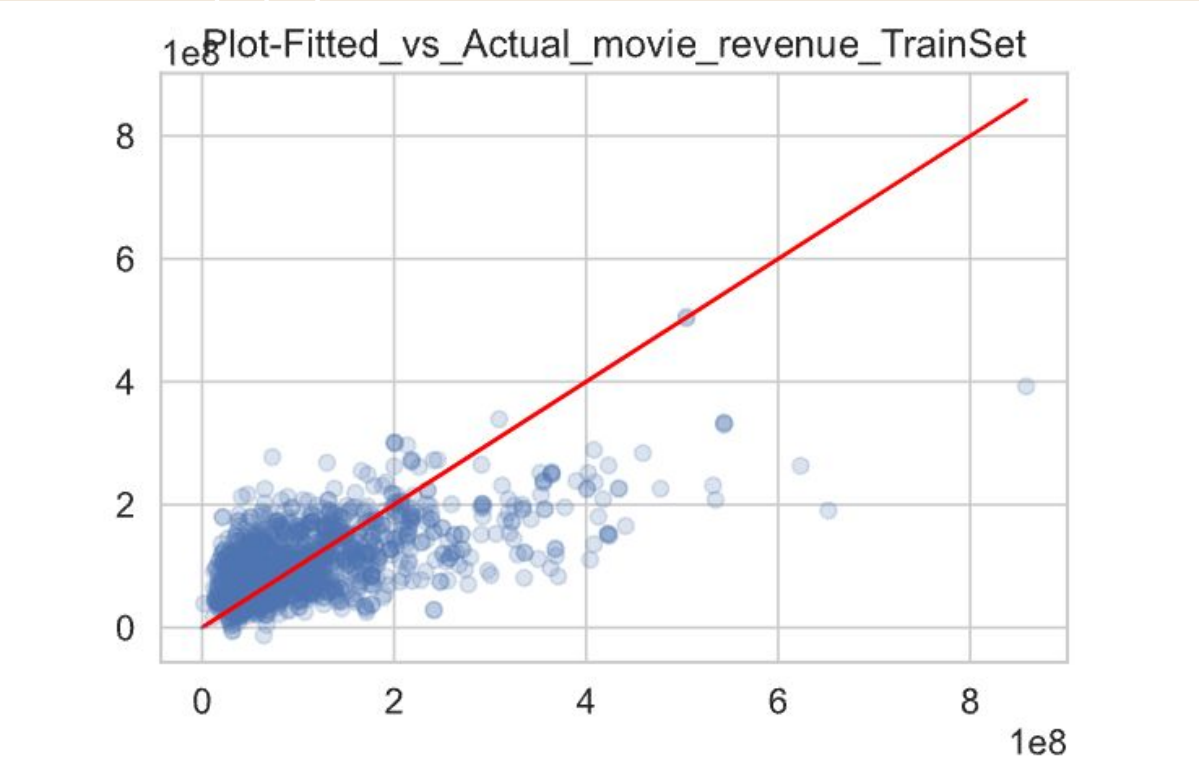
APPENDIX

Actual Revenue v/s Predicted Revenue on Testing DataSet



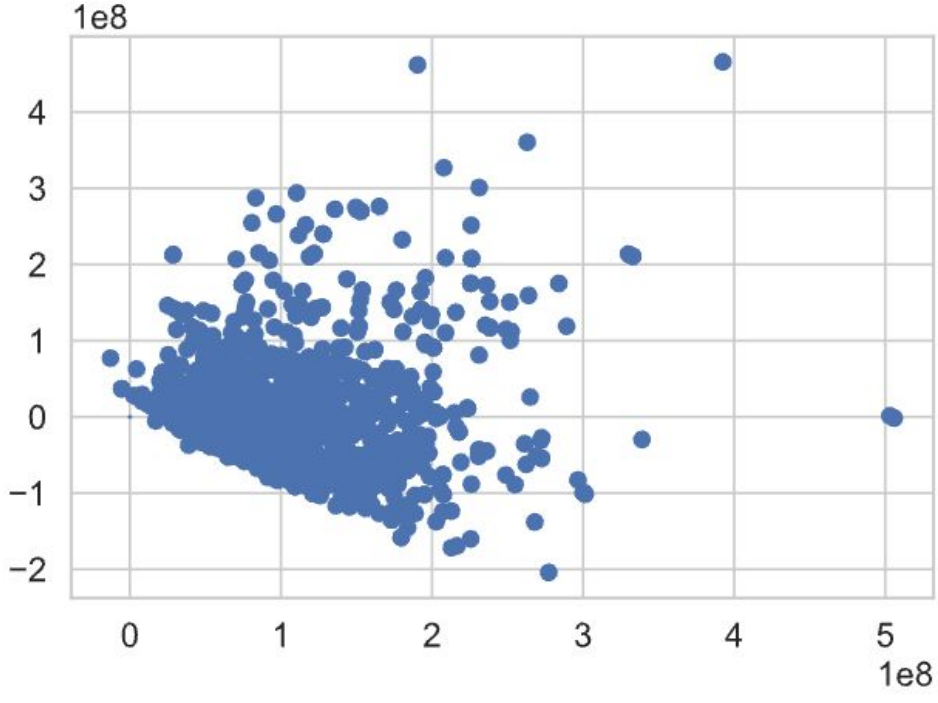
APPENDIX

Actual Revenue v/s Predicted Revenue on Training DataSet



APPENDIX

Residual Analysis



THANK YOU!