



Successful Movie Prediction

DIEGO DUQUE



Agenda

1

INTRO

2

APPROACH

3

RESULTS

4

FUTURE WORK

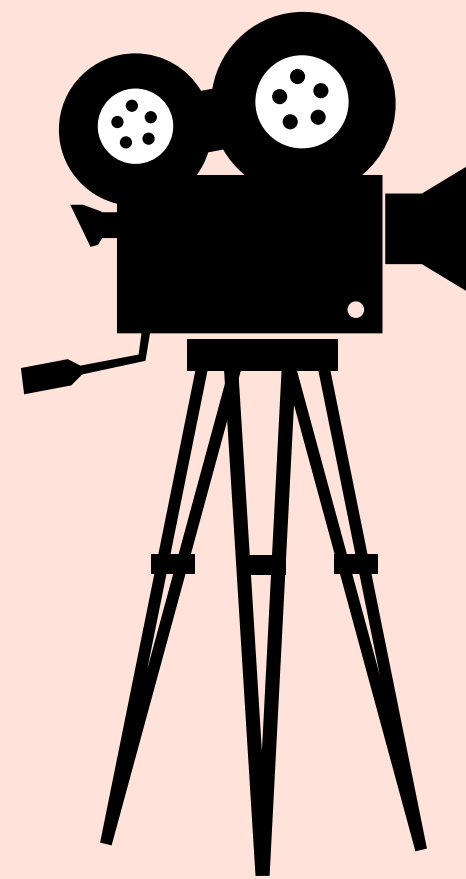
5

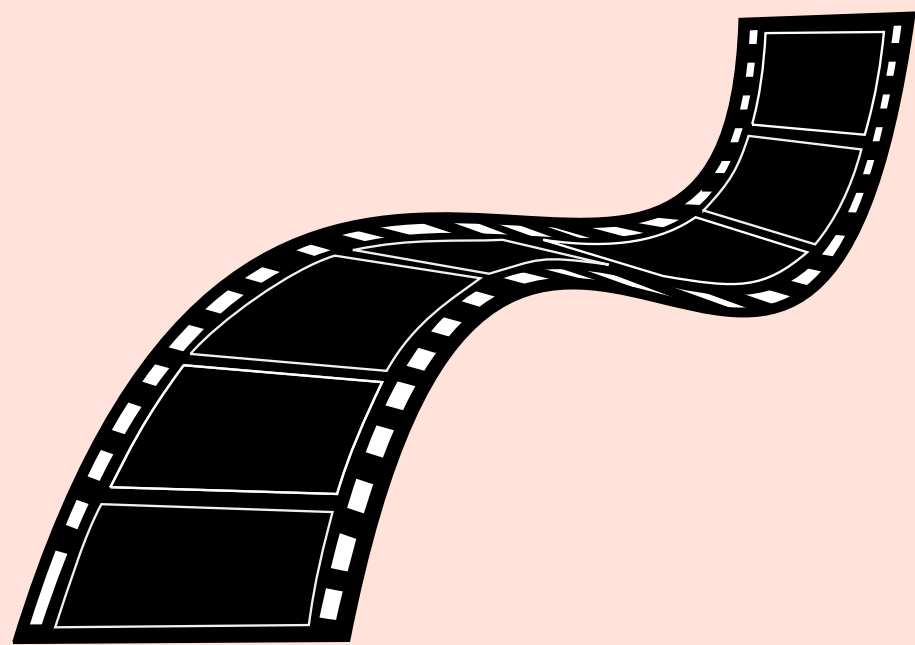
APPENDIX



1- INTRO

THE CLIENT





2- APPROACH

OUTCOME



PREDICT GLOBAL GROSS
FOR A SUCCESSFUL MOVIE

Scrape



BoxOfficeMojo.com



2010-2019



2000 movies

DATA

Numerical

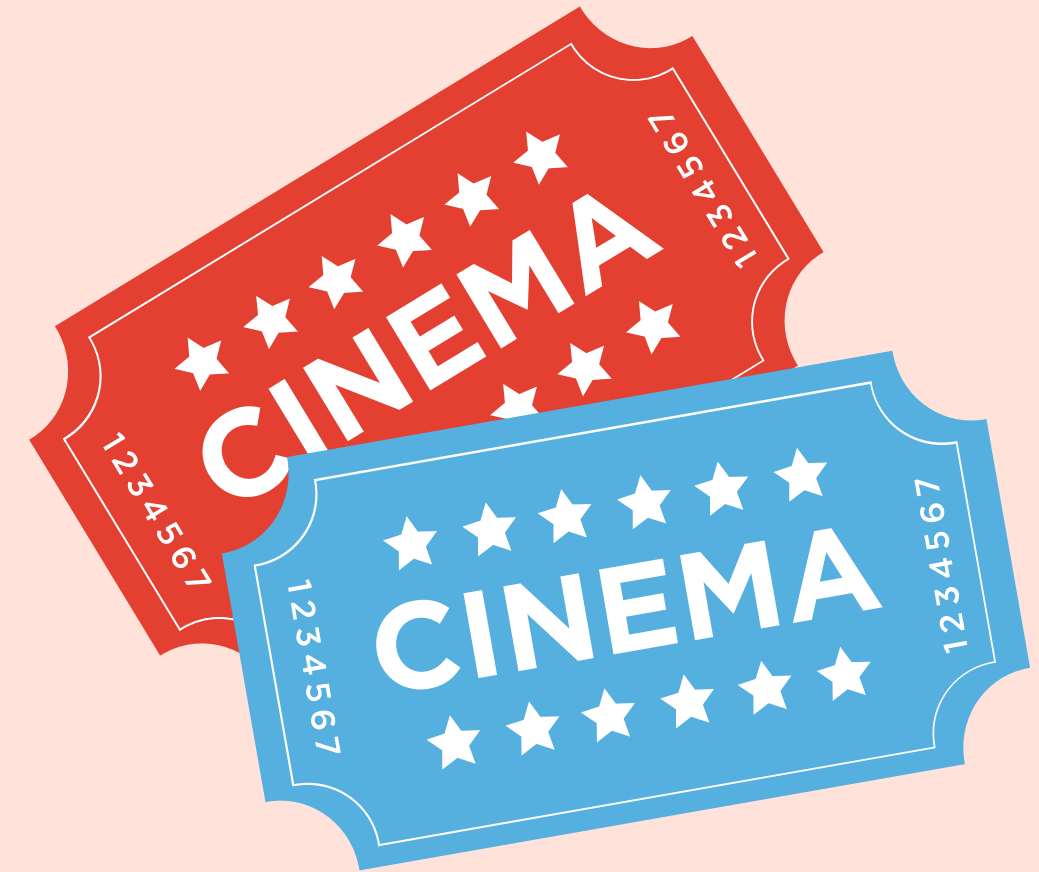
- WORLDWIDE GROSS
- BUDGET
- WIDEST RELEASE
- RUNTIME

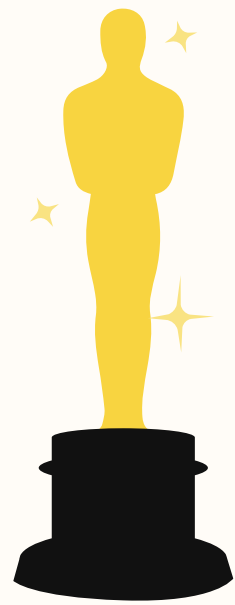
Categorical

- TITLE
- MPAA RATING
- RELEASE DATE
- GENRES
- SEASON

3- RESULTS

MODELING



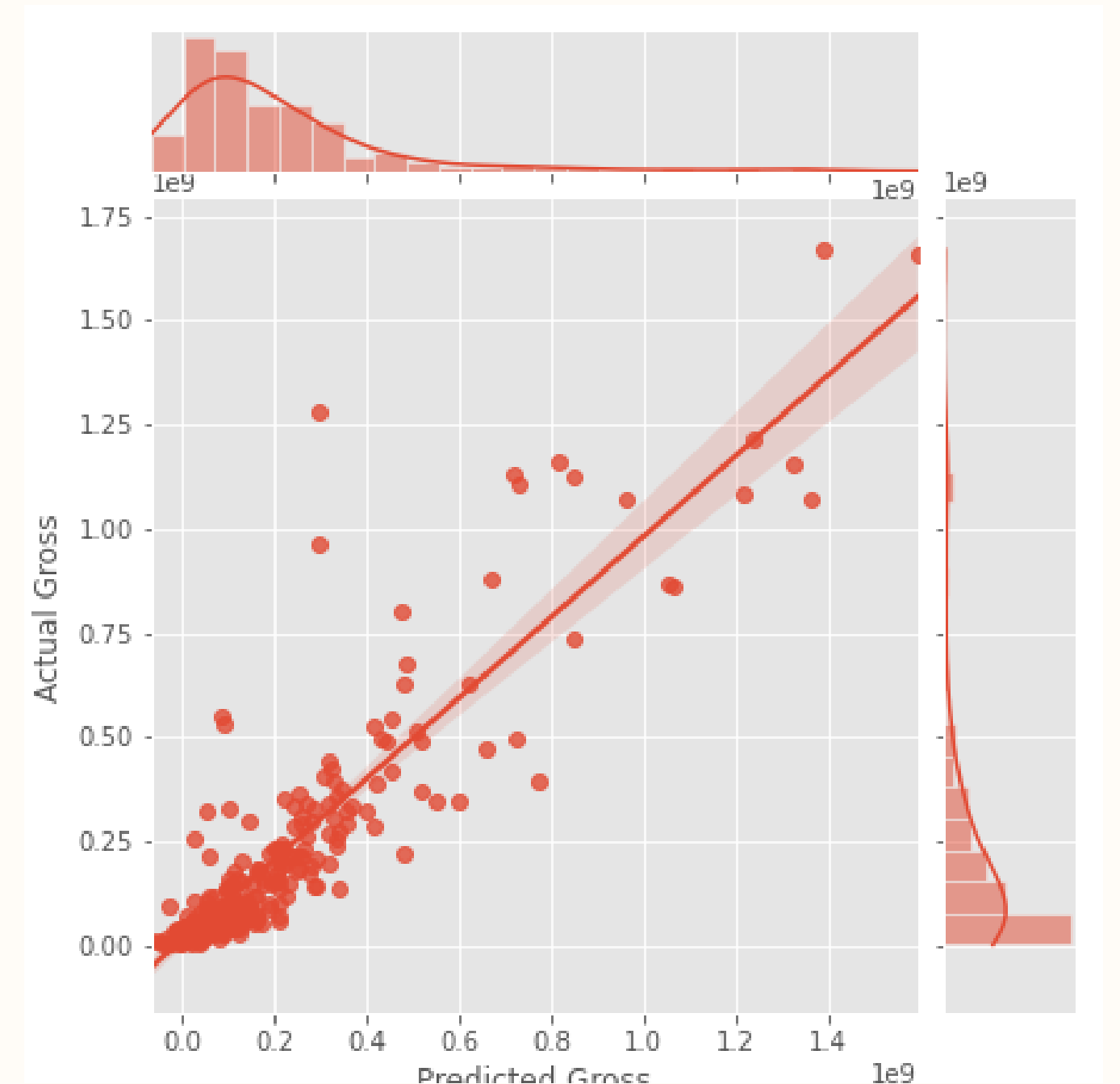


Ridge CV

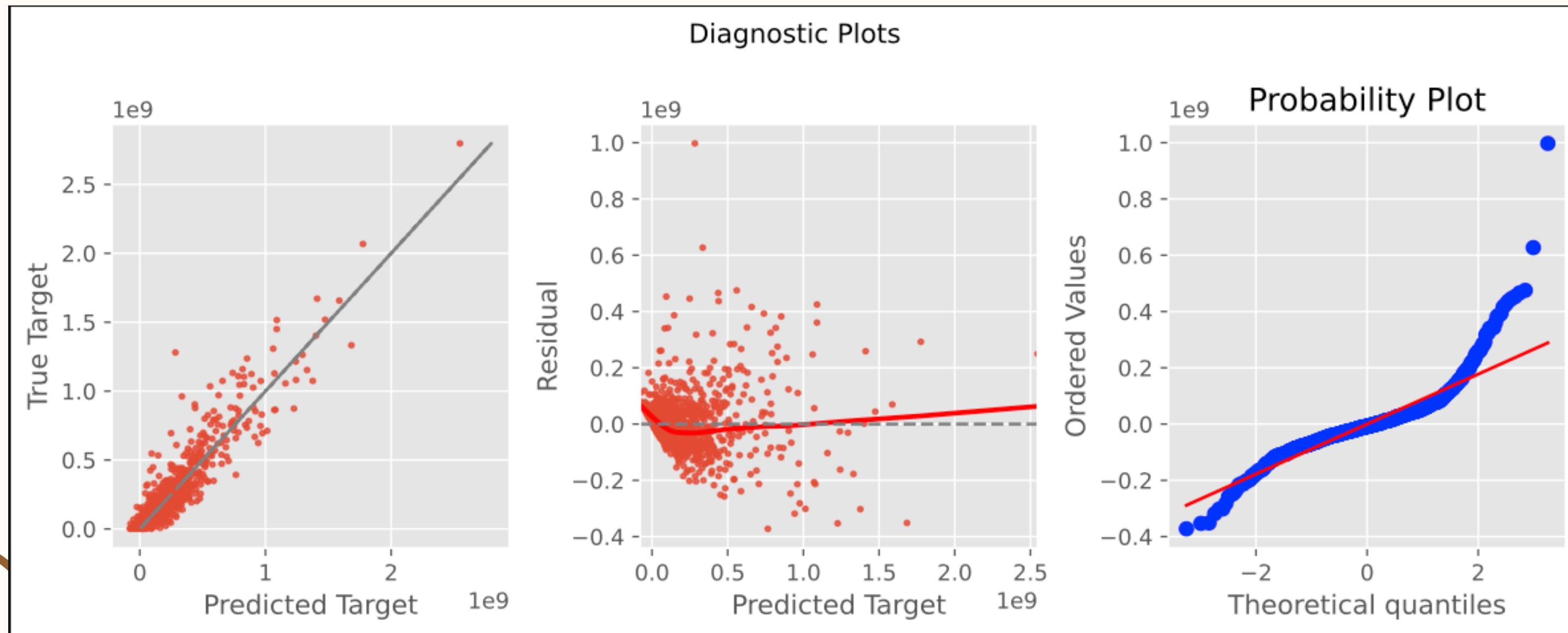
$R^2 = 0.896$

Lasso CV

$R^2 = 0.894$







Ridge CV



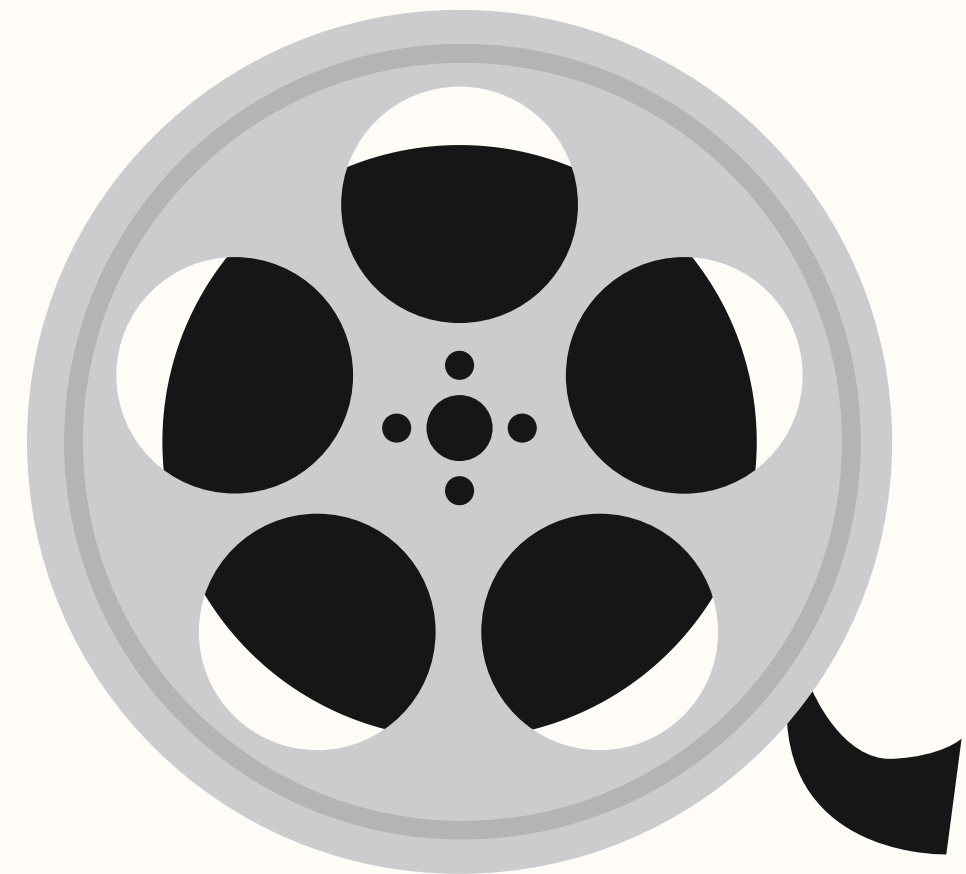


Presenting Key Attributes

-  **BUGET**
High
-  **RATING**
PG-13
-  **DISTRIBUTOR**
Walt Disney Studios Motion Pictures
-  **SEASON**
Summer

4 - FUTURE WORK

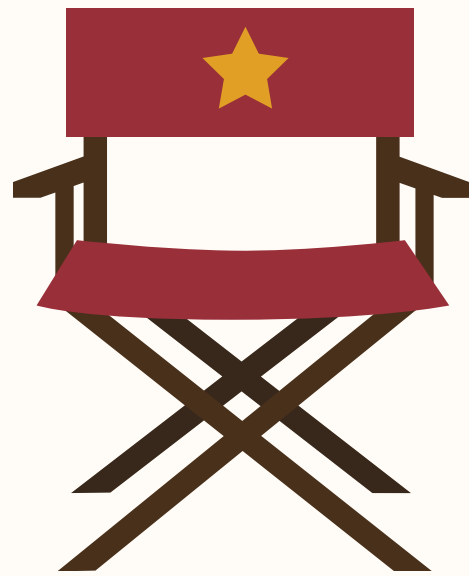
- 1 DIRECTOR
AND CAST
- 2 TIME SERIES
- 3 COMBINE WITH
OTHER WEBSITES





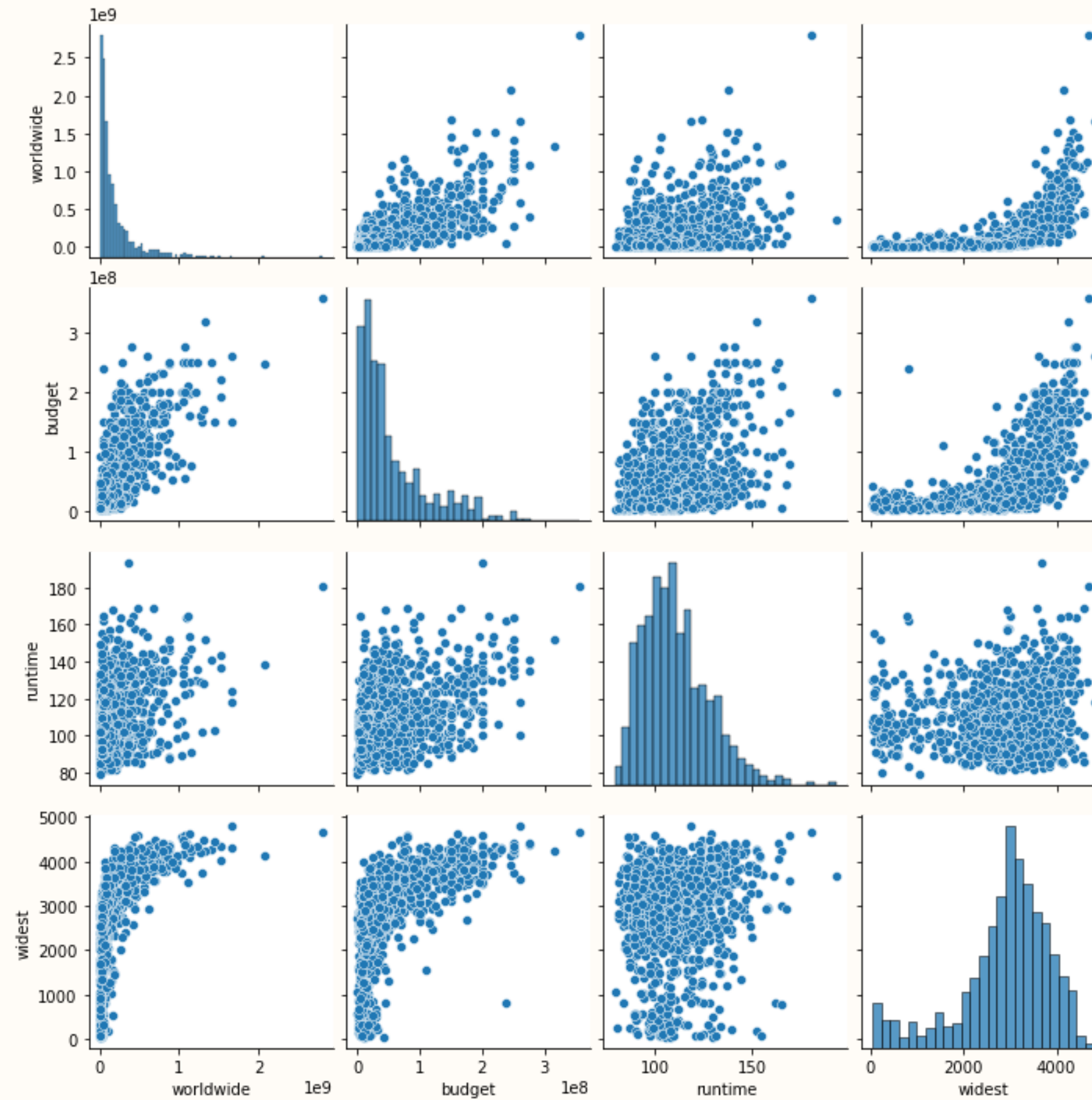
THANK YOU

DIEGO DUQUE



5- APPENDIX

Additional charts



PAIRPLOT (Numerical values)

OLS

(Numerical values)

OLS Regression Results						
Dep. Variable:	worldwide		R-squared:	0.587		
Model:	OLS		Adj. R-squared:	0.586		
Method:	Least Squares		F-statistic:	544.1		
Date:	Fri, 09 Jul 2021		Prob (F-statistic):	7.74e-220		
Time:	03:25:42		Log-Likelihood:	-23454.		
No. Observations:	1151		AIC:	4.692e+04		
Df Residuals:	1147		BIC:	4.694e+04		
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-2.516e+08	3.93e+07	-6.396	0.000	-3.29e+08	-1.74e+08
budget	2.7531	0.121	22.727	0.000	2.515	2.991
runtime	1.11e+06	3.27e+05	3.396	0.001	4.69e+05	1.75e+06
widest	5.859e+04	6506.779	9.005	0.000	4.58e+04	7.14e+04
Omnibus:	664.497	Durbin-Watson:	1.176			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	10855.225			
Skew:	2.321	Prob(JB):	0.00			
Kurtosis:	17.311	Cond. No.	6.27e+08			

Ridge - The relationship between alpha and coef

