

# Franchise Movies

## Box Office Performance

6<sup>th</sup> Oct 2017

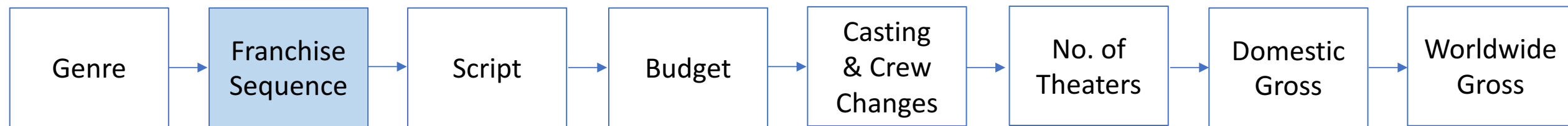
Sanghamitra Dutta



# Objective

Predicting World Wide Gross of a Franchise Movie on the basis of its **sequence** in the Franchise

## Movie Project Pipeline



# Data

Data acquired through Web scrapping the following sites:



## Description of Data Acquired:

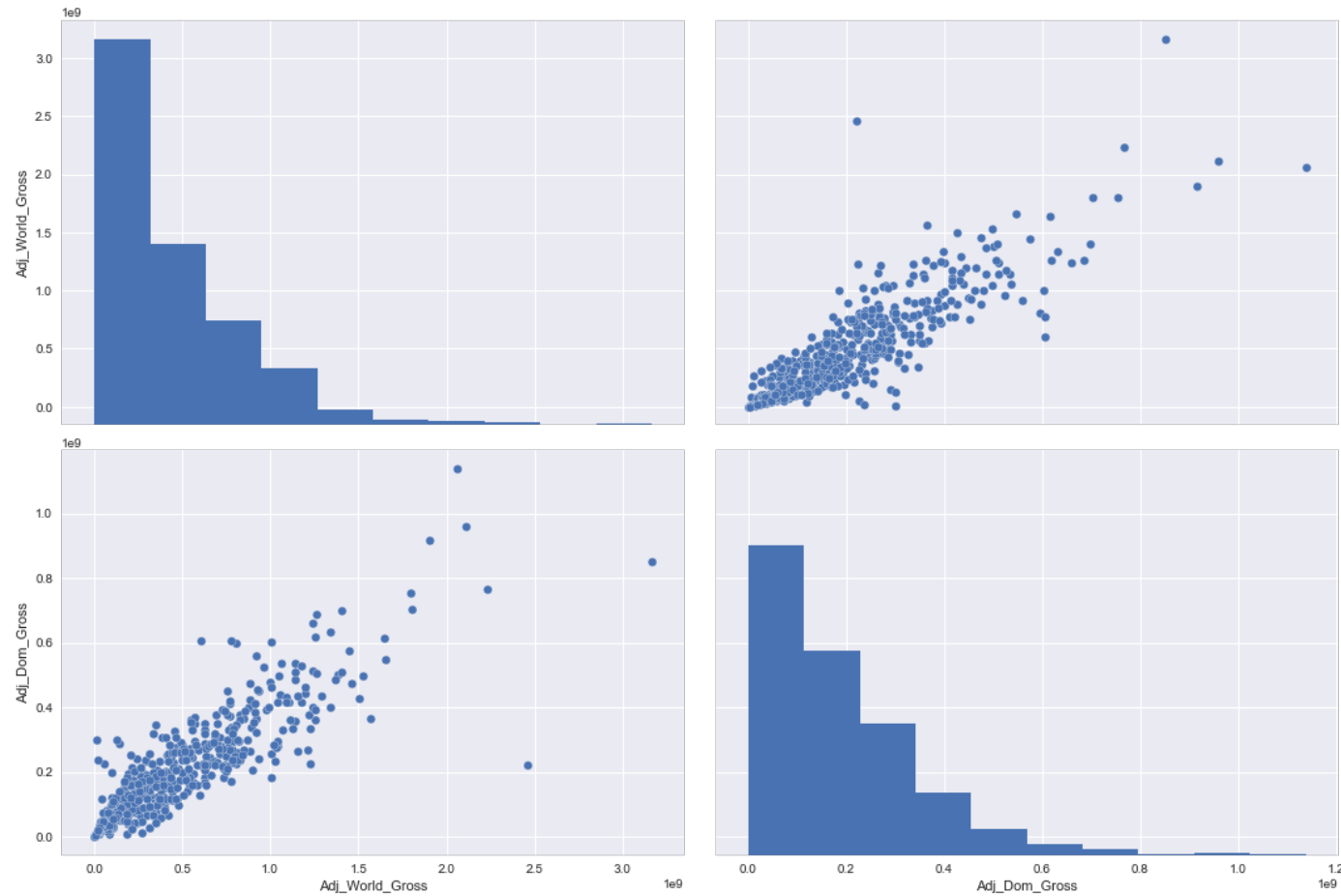
- 205 Franchises
- 593 Movies

# Data Exploration

# Features Analyzed

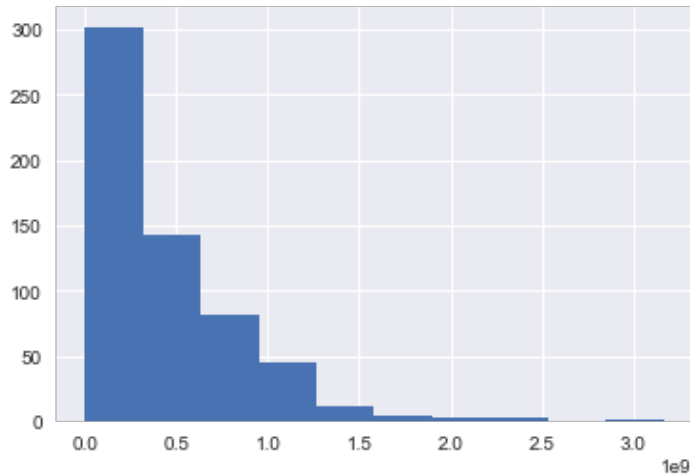
- Adjusted Domestic Gross
- Adjusted Budget
- Widest Release
- Run Time
- Franchise Sequence Number
- Genre

Dropped Domestic Gross as it is highly correlated to Worldwide Gross

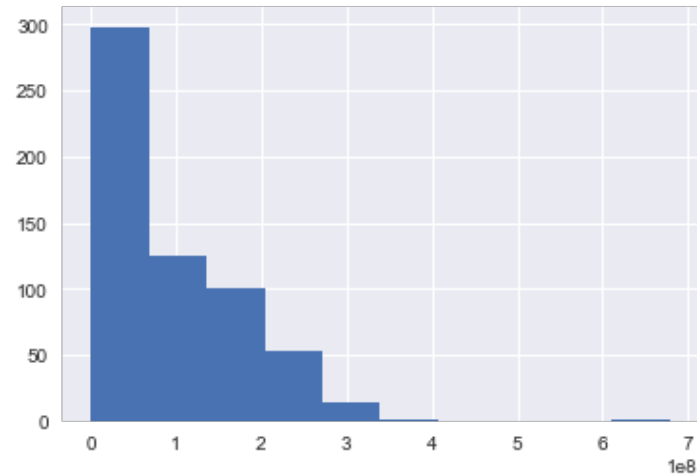


Correlation: 0.879480

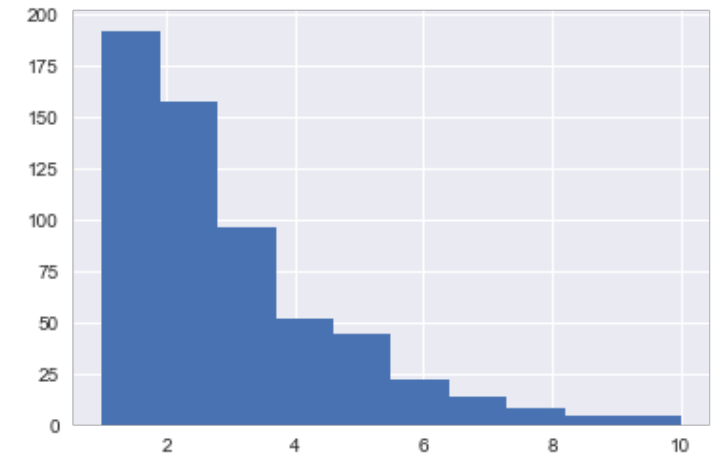
# Log Transformed Right Skewed Data



Adjusted Worldwide Gross



Adjusted Budget



Franchise Number

## Low Correlation of Dependent Variable with Franchise Sequence

Adj_World_Gross	
Adj_Budget	<u>0.667993</u>
Adj_Dom_Gross	0.879480
Adj_World_Gross	1.000000
Widest_Release	<u>0.417215</u>
Month	0.067967
Run_Time	<u>0.525915</u>
FranchiseNum	0.104562



# Linear Regression Modelling

# Model Feature Selection

Dep. Variable:	log_Adj_World_Gross	R-squared:	0.456
Model:	OLS	Adj. R-squared:	0.454
Method:	Least Squares	F-statistic:	247.6
Date:	Thu, 05 Oct 2017	Prob (F-statistic):	8.58e-79
Time:	21:19:03	Log-Likelihood:	-798.89
No. Observations:	593	AIC:	1604.
Df Residuals:	590	BIC:	1617.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	8.1755	0.527	15.499	0.000	7.139	9.211
log_Adj_Budget	0.5663	0.033	16.909	0.000	0.501	0.632
Run_Time	0.0098	0.002	4.575	0.000	0.006	0.014

Omnibus:	246.929	Durbin-Watson:	1.647
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3721.008
Skew:	-1.423	Prob(JB):	0.00
Kurtosis:	14.937	Cond. No.	1.56e+03

log (Adjusted Budget)  
Run Time

## Adjusted R<sup>2</sup> values reported for log(Adjusted Worldwide Gross)

Adjusted R<sup>2</sup> = 0.467

+  
log(Franchise No .)

Adjusted R<sup>2</sup> 0.570

+  
log(Franchise No \*Widest Release)  
log(Franchise No .)

Adjusted R<sup>2</sup> = 0.463

+  
log(Franchise No \*Widest Release)

# Used LassoCV to Cross Validate and Regularize Model

- $\log(\text{Franchise No.})$  is the most important Feature
- Followed by interaction between  $\log(\text{Franchise No. and Widest Release})$
- While the 1<sup>st</sup> feature impacts the Box Office Gross negatively the 2<sup>nd</sup> feature has a positive impact

**Best Alpha selected by Lasso CV**

<b>Feature Name</b>	<b>Alpha = 0.0</b>
$\log(\text{Franchise No.})$	-1.100
$\log(\text{Franchise No. X Widest Release})$	0.769
$\log(\text{Adj. Budget})$	0.379
Run Time	0.013

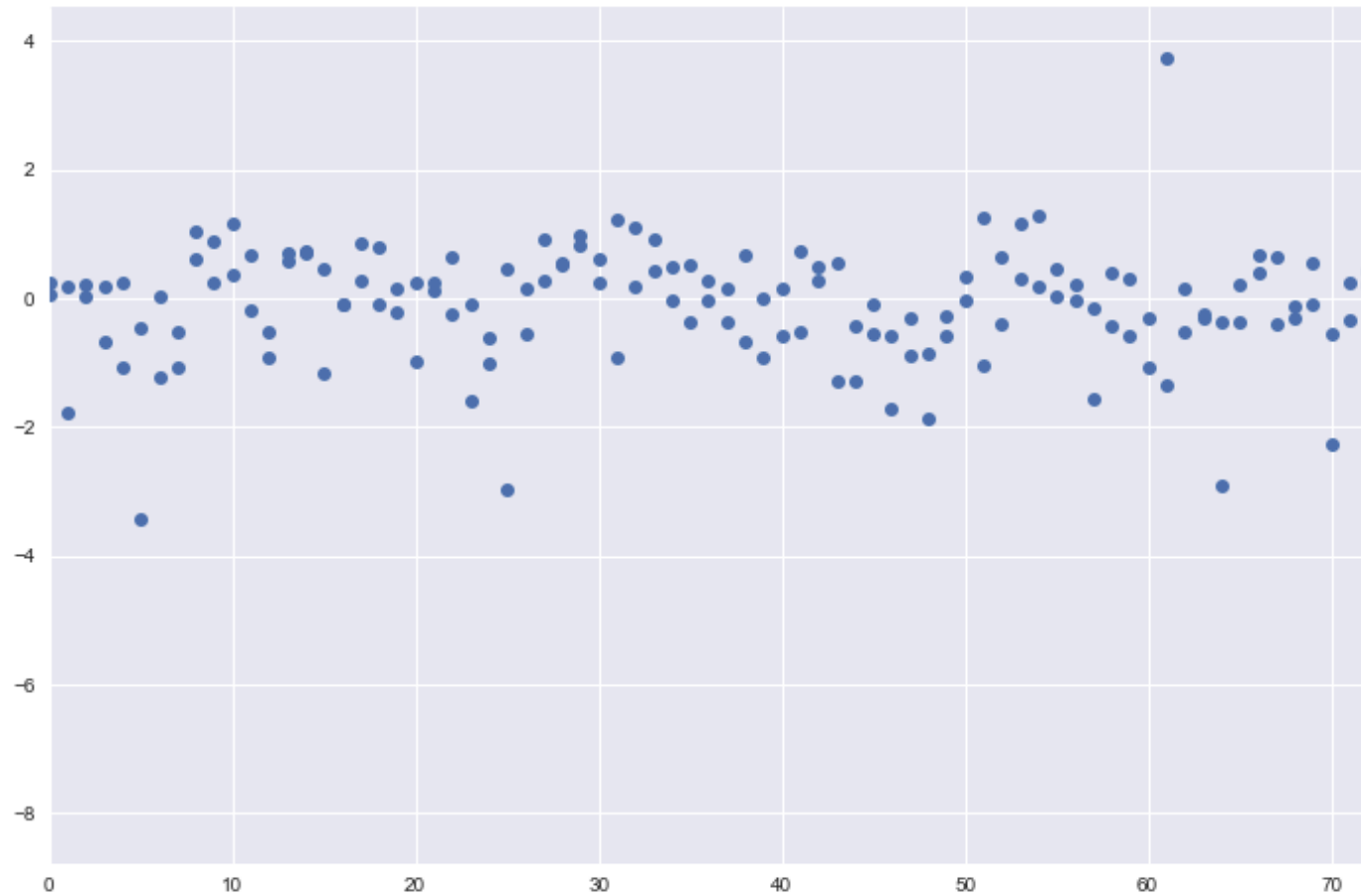
**For  $\log(\text{Adjusted Worldwide Gross})$ :**

**LassoCV  $R^2 = 0.573$**

**For Adjusted Worldwide Gross:**

**$R^2 = 0.427$**

Residual plot is normally distributed



Prediction

vs

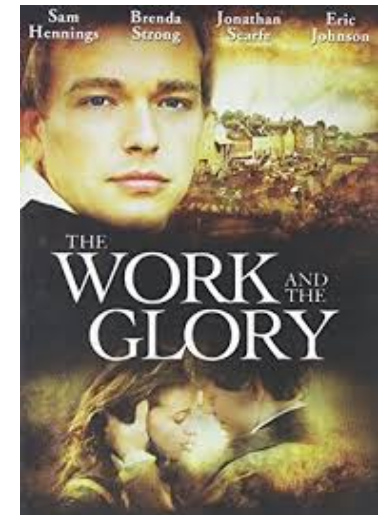
Actual



# Movies Over Predicted by 10% or more

MovieTitle	Franchise	genre	Release_Date	FranchiseNum	Adj_World_Gross	Adj_Budget	Widest_Release	Predict	PredictDiff
The Giant Mechanical Man	Mechanic	Romantic Comedy	2012-04-27	1	8.300000e+03	1.010005e+06	1.0	17.218447	-8.194436
Atlas Shrugged Part III: Who Is John Galt?	AtlasShruggedFranchise	Drama	2014-09-12	3	9.356777e+05	5.493065e+06	242.0	17.120782	-3.371756
The Work and the Glory	TheWorkandtheGlory	Period Drama	2004-11-24	1	4.691000e+06	1.050962e+07	112.0	18.344704	-2.983548
Atlas Shrugged: Part II	AtlasShruggedFranchise	Drama	2012-10-12	2	3.684200e+06	1.104359e+07	1012.0	18.044561	-2.924997
The Work and the Glory II: American Zion	TheWorkandtheGlory	Period Drama	2005-10-21	2	2.805000e+06	9.003561e+06	206.0	17.595170	-2.748256
Mallrats	ViewAskewniverse	Comedy	1995-10-20	2	4.308792e+06	1.246635e+07	852.0	17.927787	-2.651619
Hot Tub Time Machine 2	HotTubTimeMachine	Comedy	2015-02-20	2	1.363343e+07	1.532756e+07	2901.0	18.698256	-2.270220

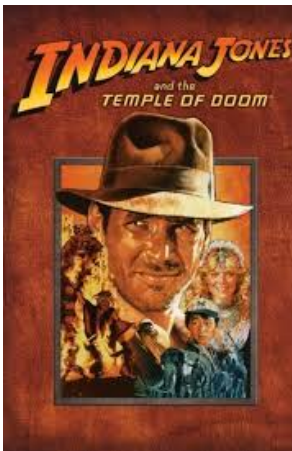
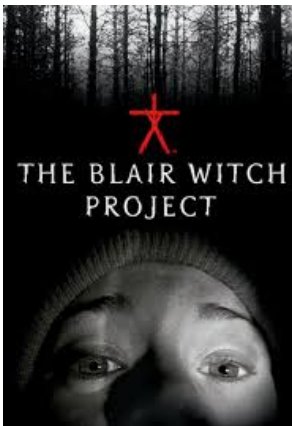
- Very Low Budget Movies
  - In bottom 10 percentile of budgets
- Low Widest Release



# Movies Under Predicted by 10% or more

MovieTitle	Franchise	genre	Release_Date	FranchiseNum	Adj_World_Gross	Adj_Budget	Widest_Release	Run_Tin	Run_Time	Predict	PredictDiff
Jaws	Jaws	Horror Thriller	1975-06-20	3	2.061341e+09	3.065517e+07	675.0	125	125.0	18.109727	3.336896
The Blair Witch Project	BlairWitch	Horror	1999-07-16	1	4.345250e+08	1.050000e+05	2538.0	86	86.0	17.017169	2.872595
Paranormal Activity	ParanormalActivity	Horror	2009-09-25	1	2.268332e+08	1.752212e+04	2712.0	99	99.0	16.458692	2.781034
Indiana Jones and the Temple of Doom	IndianaJones	Period Adventure	1984-05-23	4	8.812748e+08	7.408332e+07	1687.0	118	118.0	18.664788	1.932092
The Exorcist	TheExorcist	Horror	1973-12-26	2	1.902624e+09	5.704813e+07	3181.0	122	122.0	19.522171	1.844329
Jaws 2	Jaws	Horror Thriller	1978-06-16	2	7.936429e+08	7.598291e+07	640.0	116	116.0	18.700848	1.791296

- Horror flicks have a better performance than predictions
- Cult flicks also were under predicted
- All of them have much lower budgets
  - Some extremely low budget independent movies – Blair Witch Project & Paranormal Activity



# Investigating Genre

Do genres significantly impact Franchise Movie Box Office Gross?



None of the movie genres were statistically significant

	coef	std err	t	P> t	[0.025	0.975]
<b>Intercept</b>	6.3980	0.881	7.265	0.000	4.668	8.128
<b>Run_Time</b>	0.0208	0.002	9.435	0.000	0.016	0.025
<b>log_Adj_Budget</b>	0.3377	0.035	9.517	0.000	0.268	0.407
<b>log_FranNum_X_WidRel</b>	0.7598	0.060	12.720	0.000	0.642	0.877
<b>log_FranchiseNum</b>	-1.1013	0.085	-12.941	0.000	-1.268	-0.934
<b>Horror</b>	-0.9108	0.771	-1.182	0.238	-2.425	0.603
<b>Action</b>	-1.1188	0.774	-1.446	0.149	-2.639	0.401
<b>Comedy</b>	-1.1839	0.771	-1.535	0.125	-2.698	0.331
<b>Animation</b>	-0.5516	0.782	-0.705	0.481	-2.088	0.985
<b>Drama</b>	-1.3760	0.778	-1.769	0.077	-2.903	0.151
<b>Sci_Fi</b>	-0.9595	0.780	-1.230	0.219	-2.492	0.573

# Conclusions

- $\log(\text{Franchise Sequence})$  &  $\log$  of interaction between Franchise Sequence & Widest Release are good predictors of Worldwide Gross
- While the 1<sup>st</sup> feature has a negative impact on Worldwide Gross, the 2<sup>nd</sup> feature impacts it positively
- Genres don't have a significant impact on predicting Worldwide Gross

# Next Steps

- Acquire data on World Wide Gross for all the movies in a Franchise
- Re-investigate the prediction basis Worldwide Gross of the previous movie in the Franchise

Thank You!

# Appendix

	<b>budget</b>	<b>Domestic Gross</b>	<b>Adj_Dom_Gross</b>	<b>Widest Release</b>	<b>Month</b>
<b>count</b>	5.930000e+02	5.930000e+02	5.930000e+02	593.000000	593.000000
<b>mean</b>	6.986330e+07	1.302218e+08	1.873410e+08	2943.782462	6.883642
<b>std</b>	6.538902e+07	1.142830e+08	1.597862e+08	1000.568626	3.070394
<b>min</b>	7.000000e+03	7.396000e+03	8.300000e+03	1.000000	1.000000
<b>25%</b>	1.900000e+07	4.807130e+07	6.544570e+07	2396.000000	5.000000
<b>50%</b>	4.200000e+07	9.104708e+07	1.493454e+08	3181.000000	7.000000
<b>75%</b>	1.100000e+08	1.810306e+08	2.609284e+08	3692.000000	10.000000
<b>max</b>	2.700000e+08	9.366622e+08	1.138621e+09	4468.000000	12.000000
	<b>Run_Time</b>	<b>FranchiseNum</b>	<b>WORLD_GROSS</b>	<b>Adj_Budget</b>	<b>Adj_World_Gross</b>
<b>count</b>	593.000000	593.000000	5.930000e+02	5.930000e+02	5.930000e+02
<b>mean</b>	110.202361	2.684654	3.162379e+08	9.332014e+07	4.368697e+08
<b>std</b>	20.395901	1.863605	3.221080e+08	8.073125e+07	4.167060e+08
<b>min</b>	63.000000	1.000000	7.396000e+03	1.503121e+04	8.300000e+03
<b>25%</b>	94.000000	1.000000	8.527871e+07	3.030891e+07	1.179016e+08
<b>50%</b>	106.000000	2.000000	2.028124e+08	6.709433e+07	3.105092e+08
<b>75%</b>	123.000000	3.000000	4.350404e+08	1.428342e+08	6.352604e+08
<b>max</b>	200.000000	10.000000	2.783919e+09	6.786261e+08	3.166146e+09