LIGHTS, CAMERA, PROFIT!

Lily Goldwasser

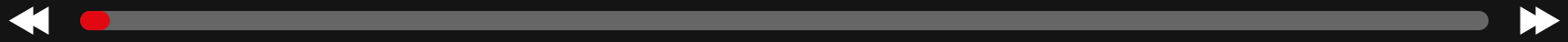
AGENDA

01 OVERVIEW

02 METHODS

03 RESULTS

04 CONCLUSIONS



OBJECTIVE

Build a Regression model that can predict Domestic Gross of movies.

CLIENT

Producer at Movie Studio

KEY FEATURES

Distributor, Domestic Opening, Budget, MPAA Rating, World Gross, Foreign Gross, Release Month, Release Year, Genre, Running Time

PROCESS OVERVIEW

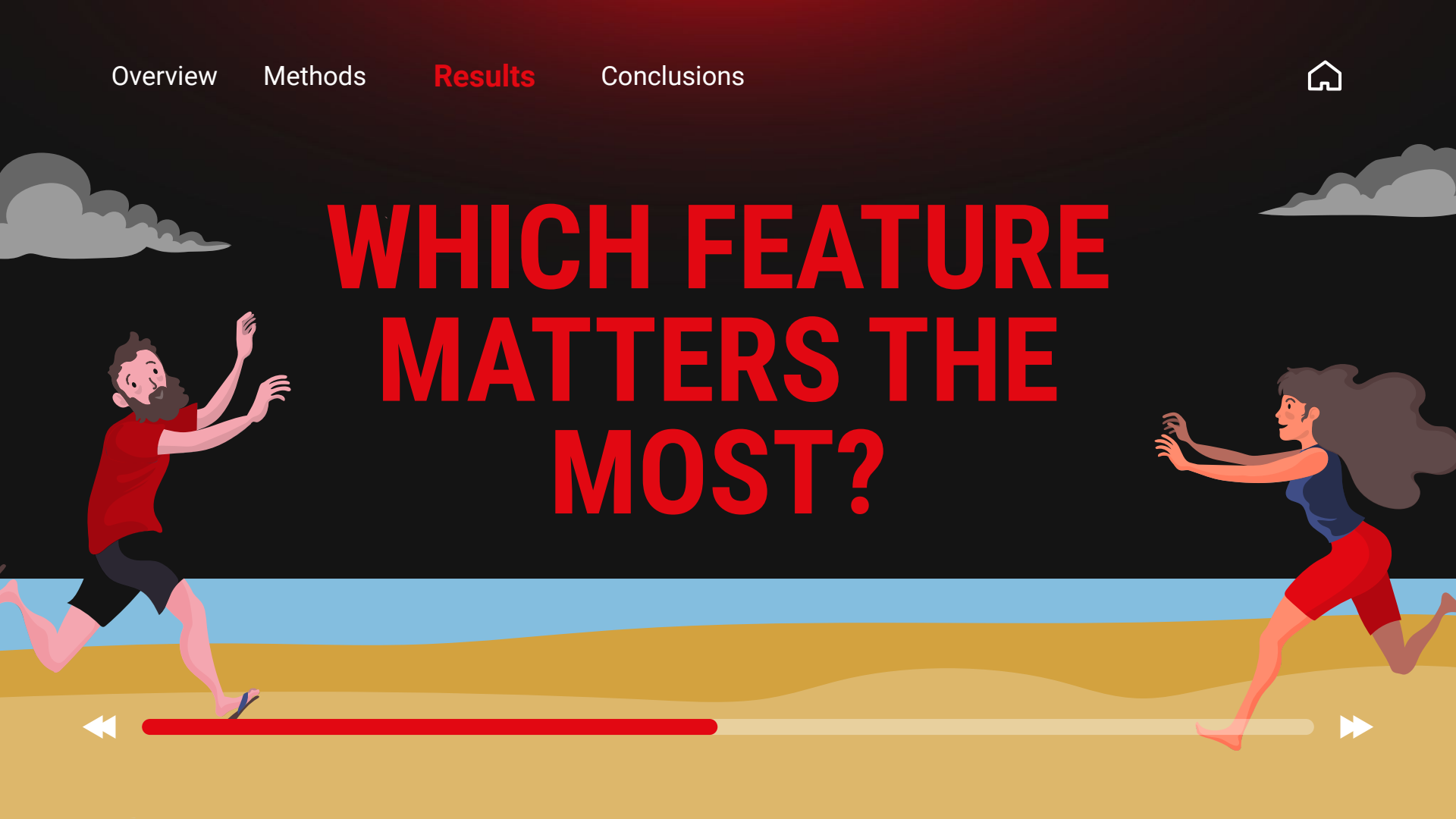
1 Web scrape

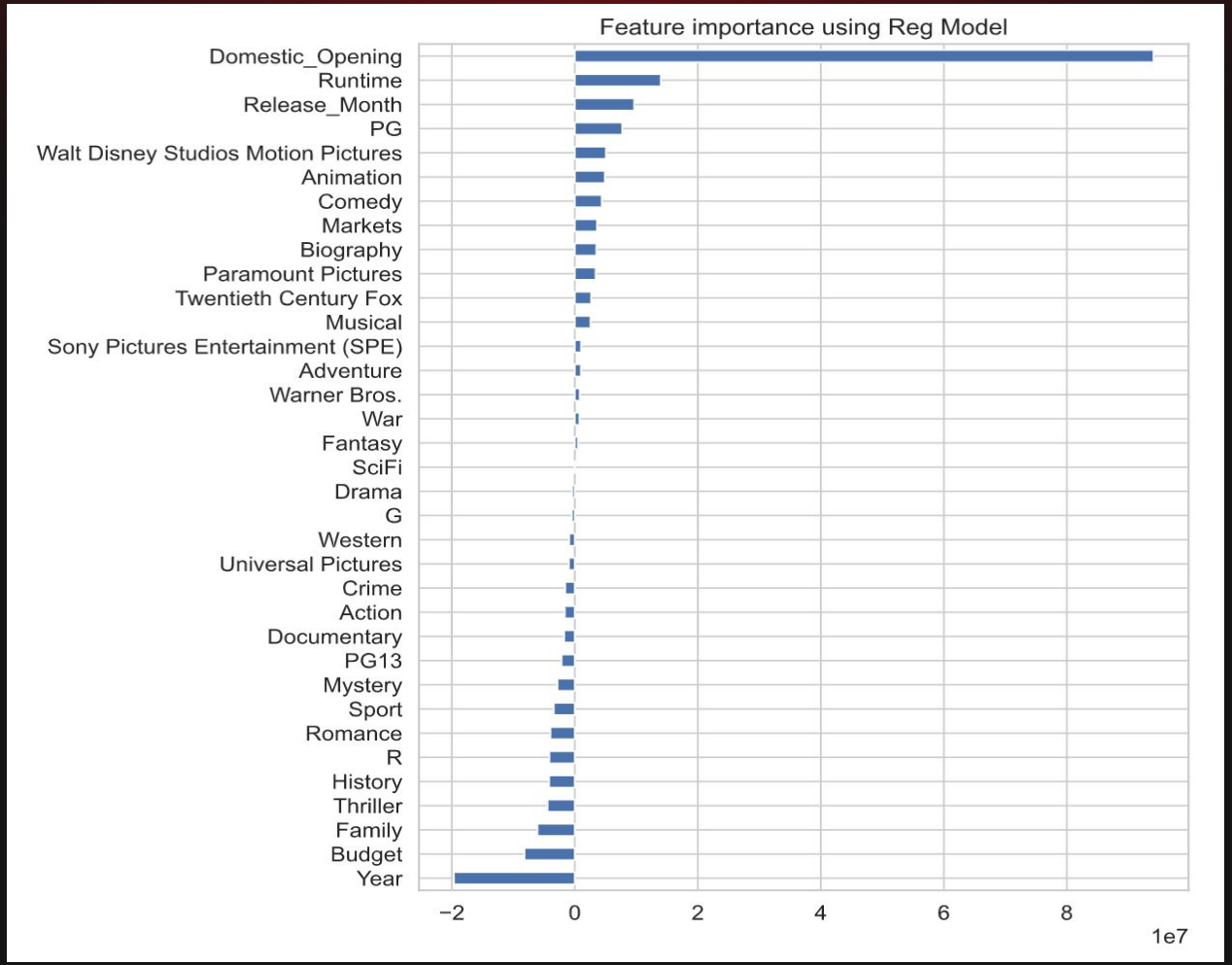
2 EDA to select Features

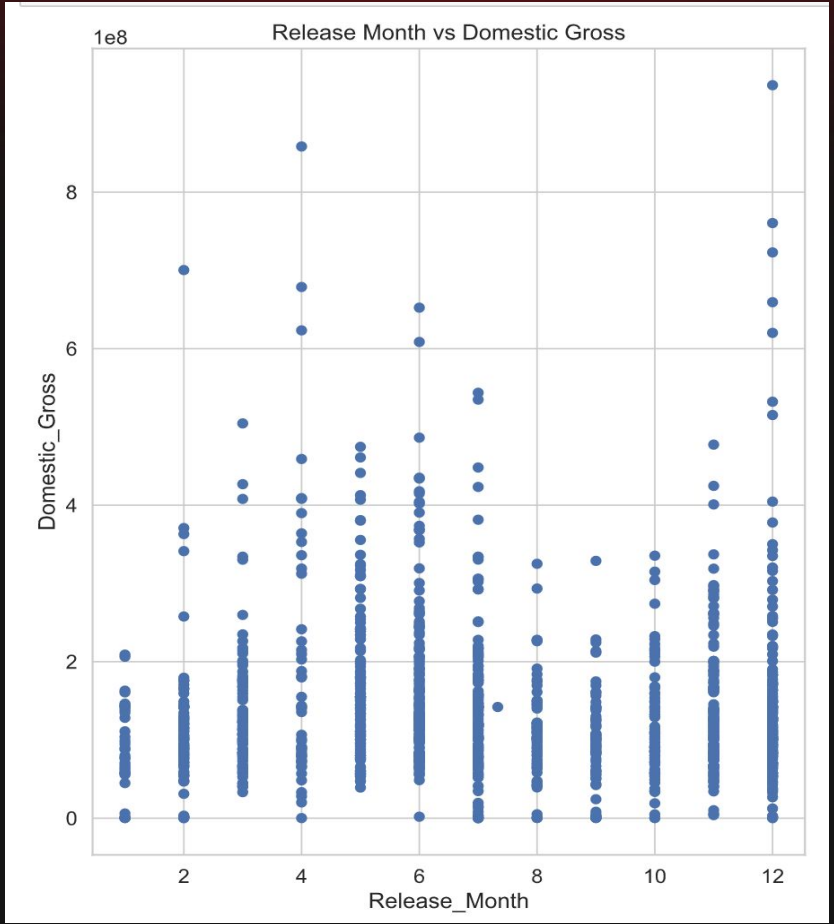
3 Regression Analysis

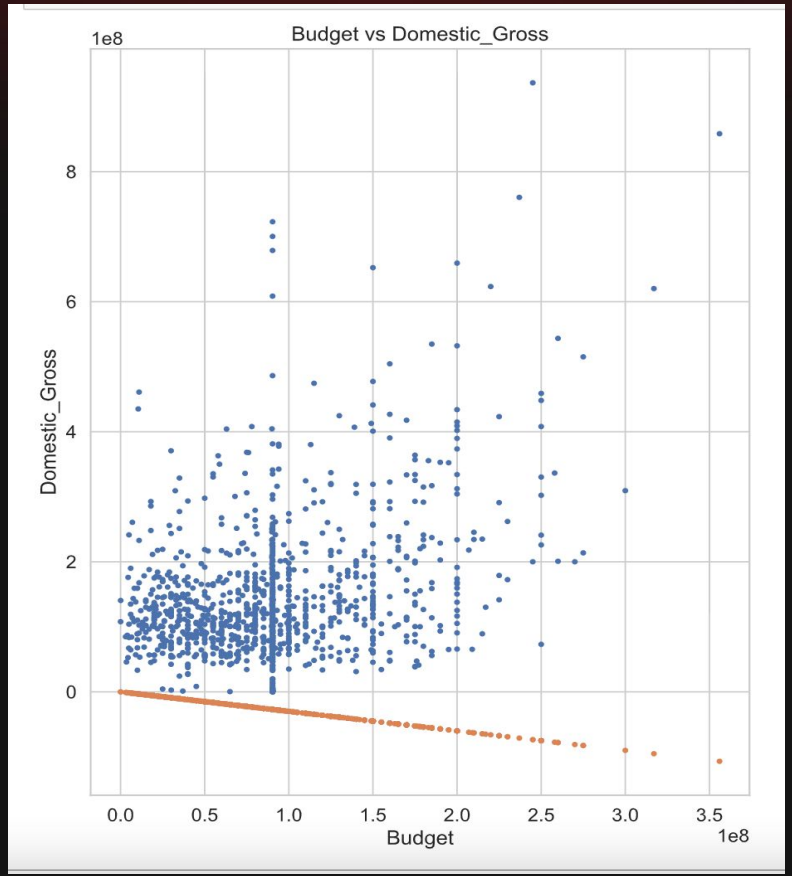
4 Prediction

WHICH FEATURE MATTERS THE MOST?

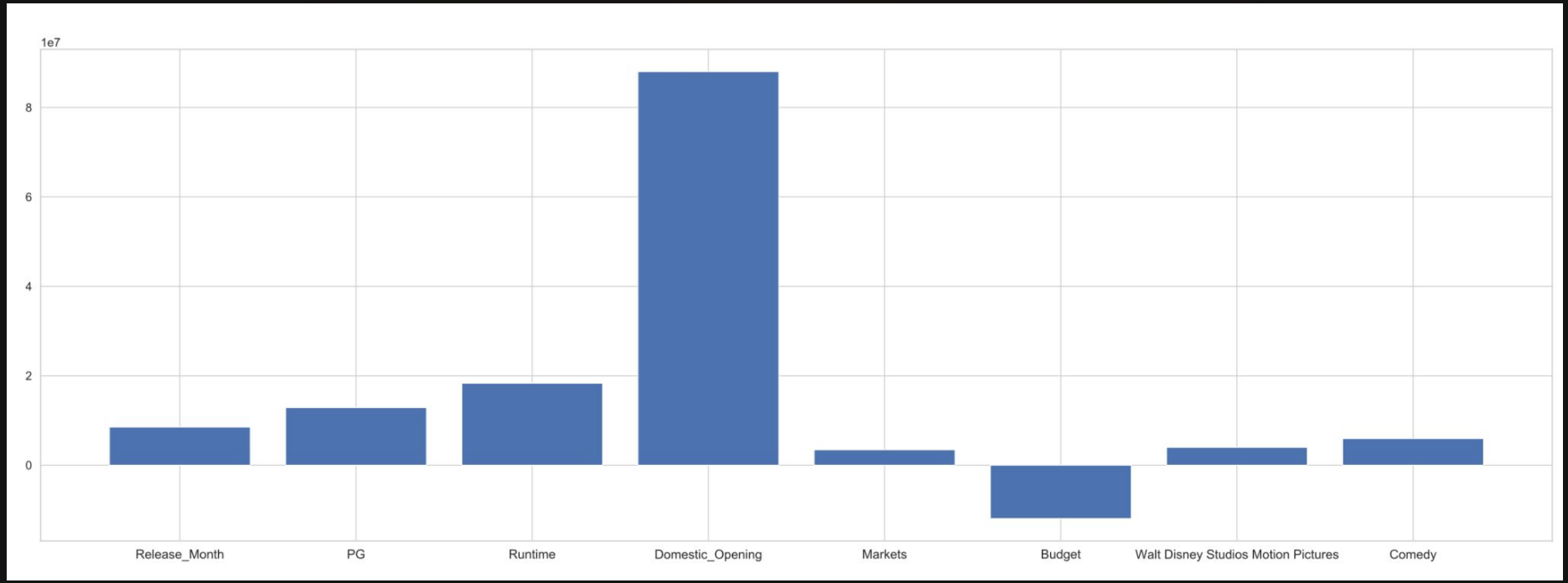



Feature importance using Reg Model
Domestic_Opening
Runtime
Release_Month
PG
Walt Disney Studios Motion Pictures
Animation
Comedy
Markets
Biography
Paramount Pictures
Twentieth Century Fox
Musical
Sony Pictures Entertainment (SPE)
Adventure
Warner Bros.
War
Fantasy
SciFi
Drama
G
Western
Universal Pictures
Crime
Action
Documentary
PG13
Mystery
Sport
Romance
R
History
Thriller
Family
Budget
Year
−2
0
2
4
6
8
1e7


Release Month vs Domestic Gross
1e8
Domestic_Gross
0
2
4
6
8
2
4
6
8
10
12
Release_Month


Budget vs Domestic_Gross
1e8
Domestic_Gross
8
6
4
2
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Budget
1e8

Feature Plot: Ridge Regression



Lasso performs best across LR features

Train/Test/Val	R^2
Linear Regression	0.711989934
Ridge Regression	0.71196667
Lasso Regularization	0.711989937

Cross Val/Test	R^2
Linear Regression	0.7048258991
Ridge Regression	0.70479173
Lasso Regularization	0.7048258995

CONCLUSIONS

- Use Domestic Opening Gross as an indicator for whether to ramp up or scale down on marketing costs
- Future work: TV series, Actors, Awards

THANKS!

Appendix

https://www.boxofficemojo.com/chart/top_lifetime_gross/?area=XWW

CREDITS: This presentation template was created by Slidesgo,
including icons by Flaticon and infographics & images by Freepik

Please keep this slide for attribution