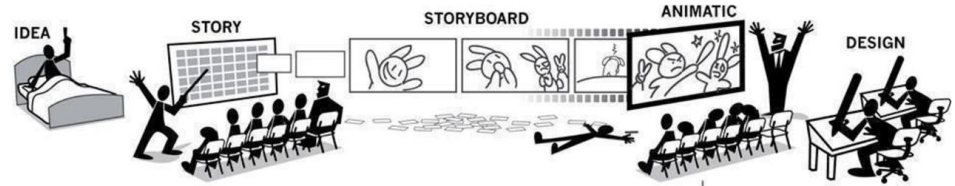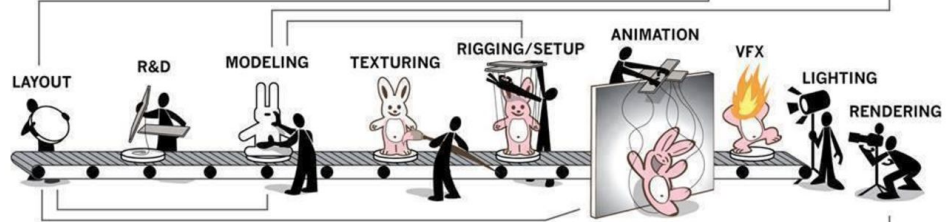# Predicting Revenue for Animation Films

Tiffany Li

Motivations

- The process for producing animation films is long, complex and expensive
- Joint efforts of up to 500 talents
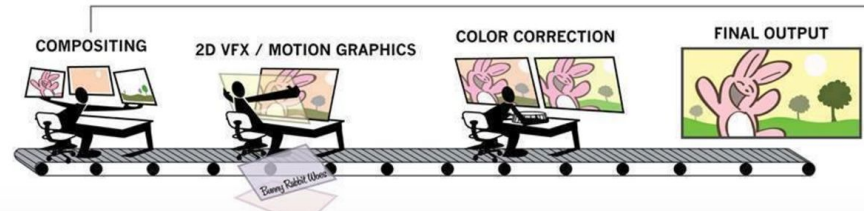- What are the keys to success?

PRE-PRODUCTION

IDEA · STORY · STORYBOARD · ANIMATIC · DESIGN

PRODUCTION

LAYOUT · R&D · MODELING · TEXTURING · RIGGING/SETUP · ANIMATION · VFX · LIGHTING · RENDERING

POST-PRODUCTION

COMPOSITING · 2D VFX / MOTION GRAPHICS · COLOR CORRECTION · FINAL OUTPUT

Web Scraping

Data collected from imdb.com

Data Processing

Using **beautiful soup**, scraped features for top 400 animation films, including movie title, budget, country, MPAA rating, opening weekend gross, production co, IMDB rating score, release date, runtime and awards.

Awards	Budget	Certificate	Color	Country	Filming Locations	Gross	Language	...	Movie Title	Official Sites	Opening Weekend
[' Nominated for 1 BAFTA Film Award.', ' A...	$200,000,000	NaN	Color	USA	Pixar Animation Studios - 1200 Park Avenue, Em...	$486,292,984	English	...	Finding Dory (2016) - IMDb	Official Facebook	$135,060,273
[' Nominated for 2 Oscars.', ' Another16 w...	$150,000,000	NaN	Color	USA	NaN	$436,471,036	English	...	Shrek 2 (2004) - IMDb	NaN	$108,037,878

Feature Engineering

- Dropped data outside of US or earlier than 1990
- Categorized movies into 'won awards', 'nominated' or 'not'
- Categorized production companies into large, medium and small based on total gross revenue
- Extracted year, month, day, day of the week and date ordinal
- Filled in missing value with median

223 data points as input

Box-Cox Transformation
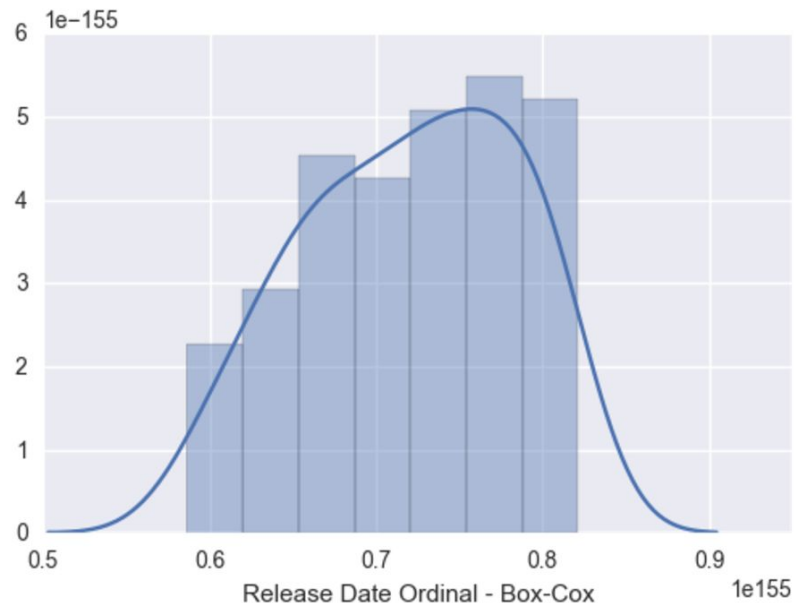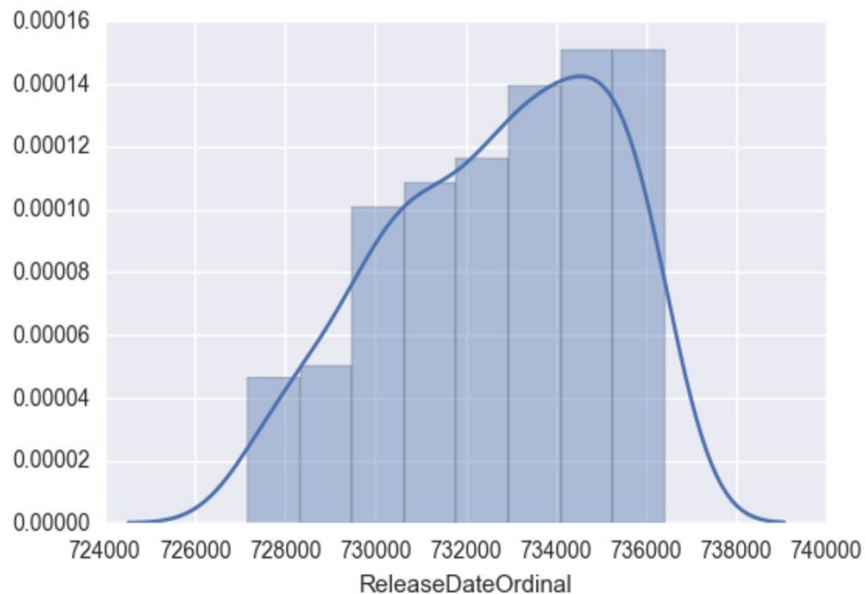
Finding the Best Model

- Used simply linear regression to compare effectiveness by adding/dropping variables
- Found the best predictor variables to be **opening weekend**, **rating score**, **release date ordinal**, **awards results** and **production company size.**

OLS Regression Results

Dep. Variable:	GrossM	R-squared:	0.831
Model:	OLS	Adj. R-squared:	0.826
Method:	Least Squares	F-statistic:	176.9
Date:	Thu, 02 Feb 2017	Prob (F-statistic):	1.81e-80
Time:	22:10:56	Log-Likelihood:	-1150.0
No. Observations:	223	AIC:	2314.
Df Residuals:	216	BIC:	2338.
Df Model:	6		
Covariance Type:	nonrobust		

Regularization

- Compared cross validated Lasso, Rigid and Elastic Net models

- **Cross validated Elastic Net model** produced the best mean squared error
  - Alpha = 0.17
  - R squared = 0.878

Predictive Model

Gross Revenues Equals to

$96 + 62 * \text{Opening Weekend} + 14 * \text{Rating Score} + 0.0 * \text{Release Date Ordinal} + 4 * \text{Nominated} + 42 * \text{Won} + 21 * \text{Large} - 23 * \text{Small}$

Example: Inside out

Model predicts: 340M
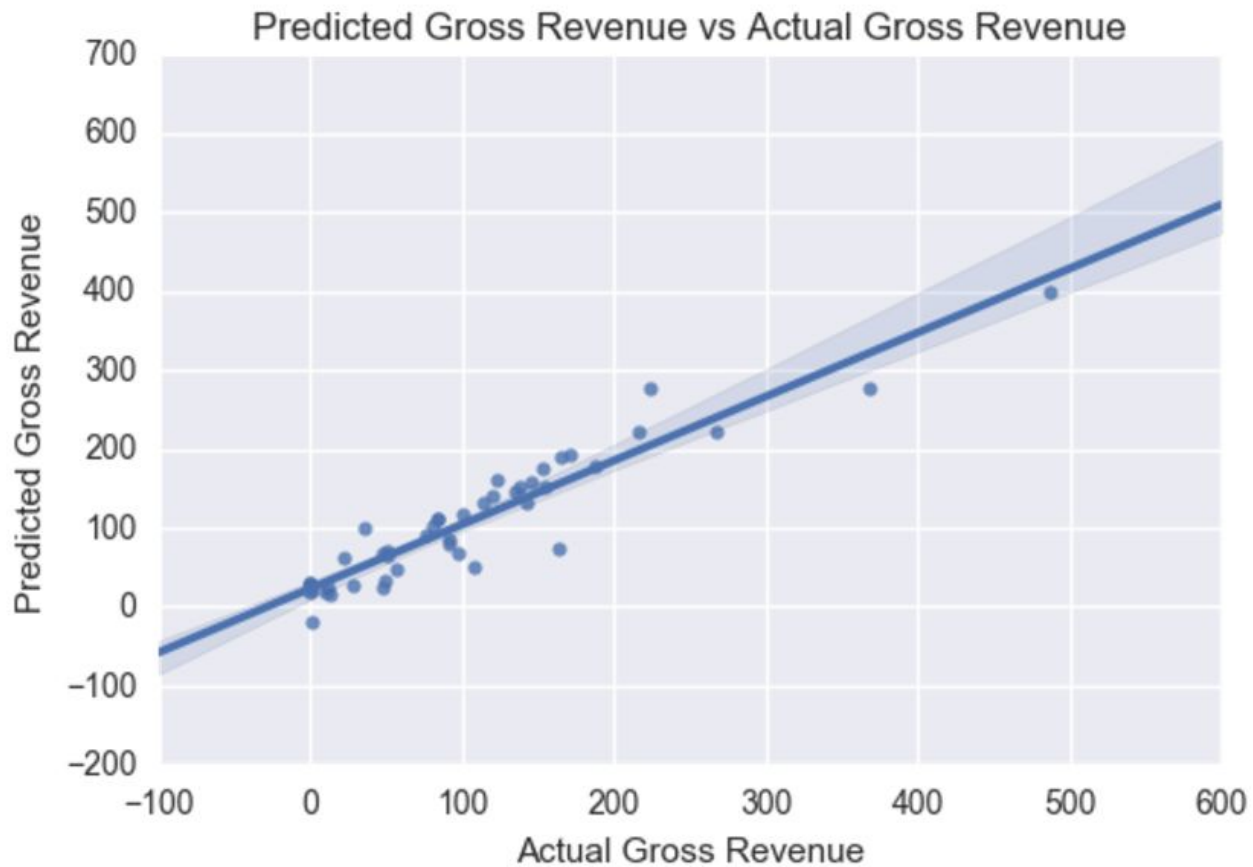
Predictive Model

Gross Revenues Equals to

$96 + 62 * \text{Opening Weekend} + 14 * \text{Rating Score} + 0.0 * \text{Release Date Ordinal} + 4 * \text{Nominated} + 42 * \text{Won} + 21 * \text{Large} - 23.4 * \text{Small}$

Example: Inside out

Model predicts: 340M

Actual: 356M

Predicted Gross Revenue vs Actual Gross Revenue

Implications

What's important?

- Opening Weekend
  - First impression matters! Put out more commercials
- IMDB Rating Score
  - Get that 'critically acclaimed' stamp
- Awards
  - Oscar, here I come...
- Production Company Size
  - Big studios still got the edge

Implications

What's not worth it:

- Budget
  - Maybe don't spend 6 months making a snowflake?
- Release Date Month / Quarter
  - Christmas doesn't guarantee the presents
- MPAA Rating
  - G or PG? Does not matter

In an effort to perfect Elsa's icy magic in Frozen, filmmakers called on Dr. Thomas Painter to learn about snowflakes from a molecular level.

Further Investigations

- More feature engineering
  - Break down the awards into different levels

- Incorporate data from other sources
  - Fill in missing values with actual values
  - Other rating metrics
  - Impact of 3D movies
  - Does the popularity of the voice over cast matter?

A minion thanks.