



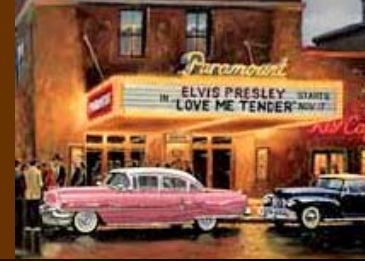
BUDT 733 Data Analysis for Decision Maker

Andreana Able | Chu-Tzu Ko | Kumiko Kotera | Thomas Lotze | Sakol Chompoonich

# Understanding the Factors Creating a Box Office Hit

May 10, 2007

# Agenda



➤ Project Overview

➤ Data Exploration

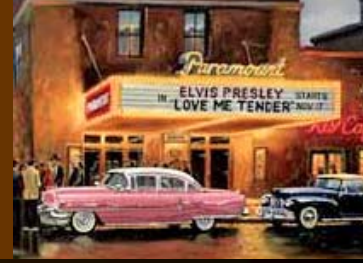
➤ Data Analysis

➤ Our Findings

➤ Moving Forward

# Project Overview

## Study Objectives



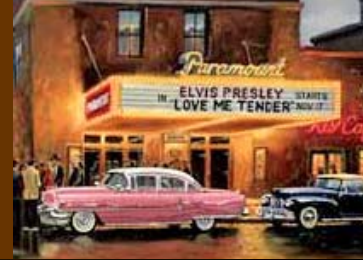
- To profile top 60 box office movies from 2004 to 2006 that “breakeven”
  - Total gross box office revenue exceed budget
  
- To answer some of the underlying questions
  - High budget or low budget movie?
  - Which movie genre to produce?
  - When is the best release date?
  - Does the distribution matter?
  - Does quality matter?

*(and more...)*



# Project Overview

## Business Challenges



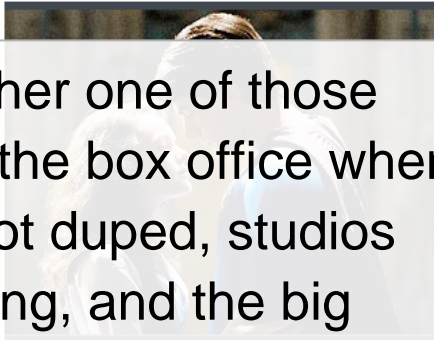
### Selected industry publications...

#### BOX OFFICE REPORT

### Disappointing 'Returns'

Joshua Rich's report: The three-day and five-day totals for Bryan Singer's "Superman" don't live up to expectations

By Joshua Rich



...It was another one of those weekends at the box office when forecasters got duped, studios started spinning, and the big numbers weren't nearly as impressive as they appear...

-- *Entertainment Weekly*

### Paramount Is Right

By Steven Mallas  
August 25, 2006

0

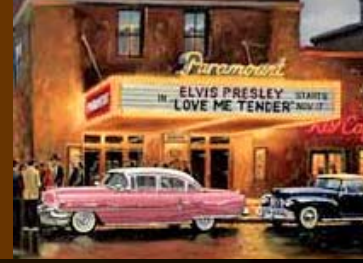
Recommendations

Recommend This

By now, everyone has heard about the soured relationship between Tom "Couch Jumpin'" Cruise and **Viacom's** (NYSE: **VIA**) Paramount movie studio. What does it all mean? ...Harrison Ford can have a flop just as easily as a newcomer could... You can spend a ton of money or a minimal amount, but there's not always a direct relationship between big budget and viewership...

-- *The Motley Fool*

# Agenda



➤ Project Overview

➤ Data Exploration

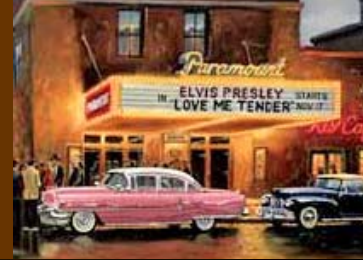
➤ Data Analysis

➤ Our Findings

➤ Moving Forward

# Data Exploration

## Data Source



### Original Data

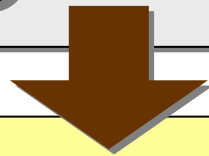
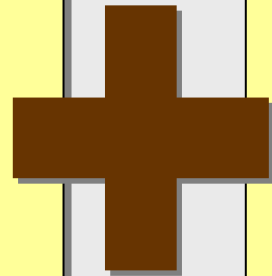
([www.the-movie-times.com](http://www.the-movie-times.com))

- Movie
- Opening Gross
- Total Gross
- # of Theaters
- Weeks in Top 60
- Release Date

### Additional Data

([www.imdb.com](http://www.imdb.com))

- Rating (10 scale)
- Number of Votes
- Primary Genre (*binned*)
- Est. Budget
- Is Sequel
- MPAA Rating (*binned*)

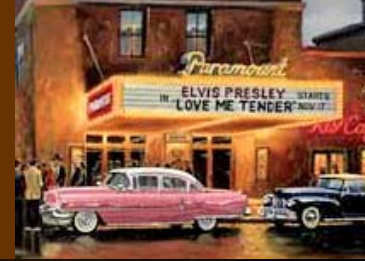


### Derived Data

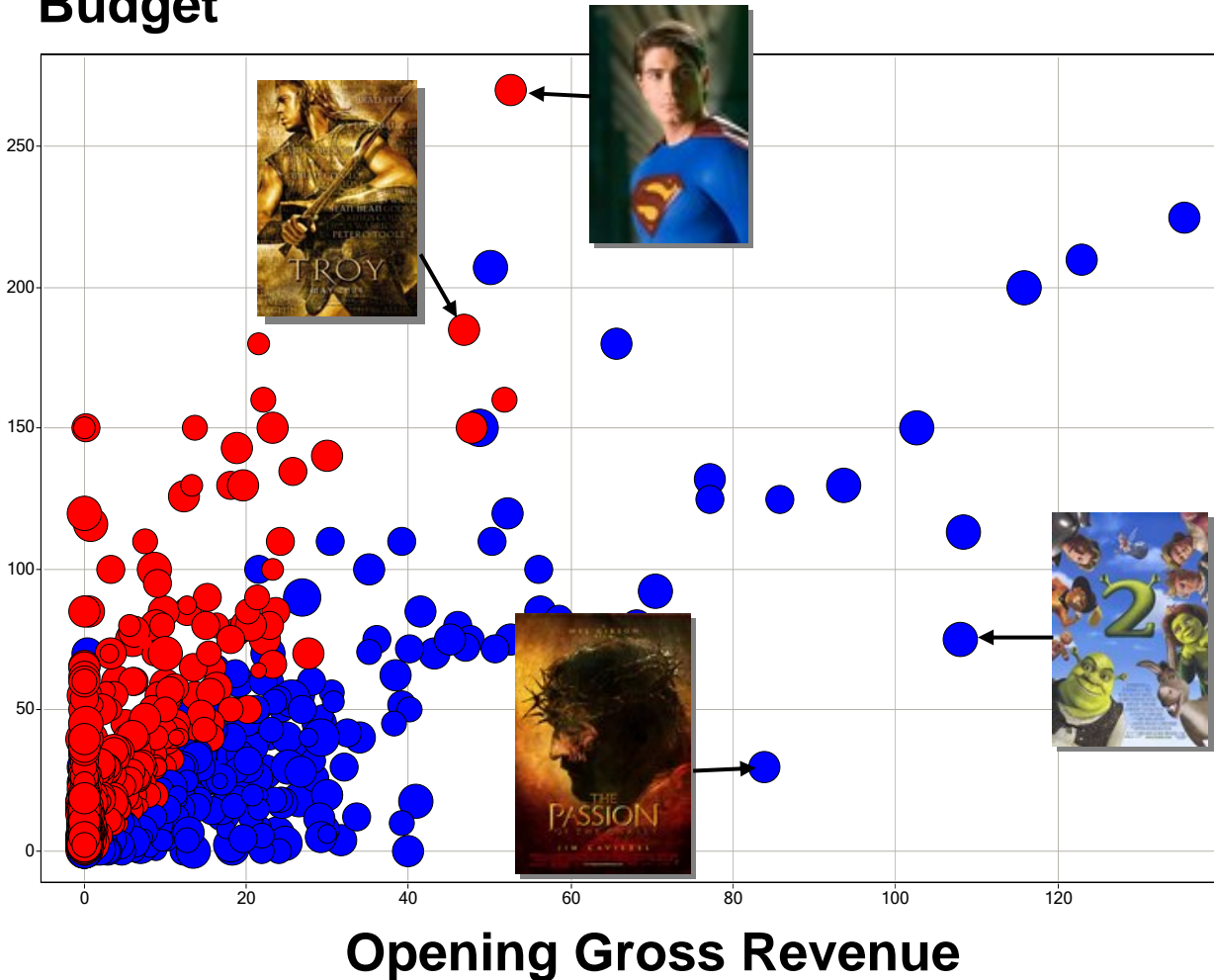
- Summer Release
- Christmas Release
- Breakeven

# Data Exploration

## Budget, Opening Gross, and Review Rating



### Budget



- High budget movies are less likely to breakeven
- High opening gross movies are more likely to breakeven

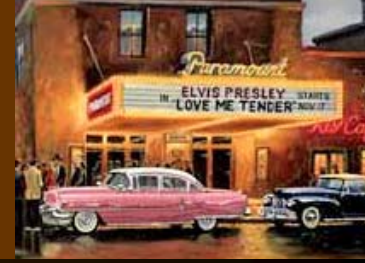
● Not Breakeven

● Breakeven

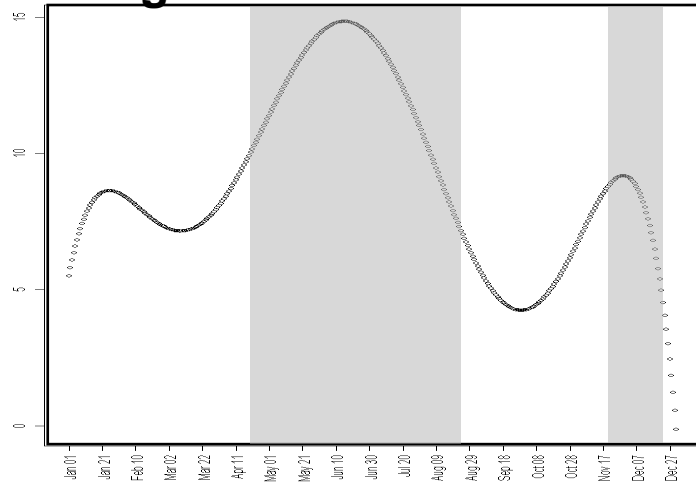
Size: Rating

# Data Exploration

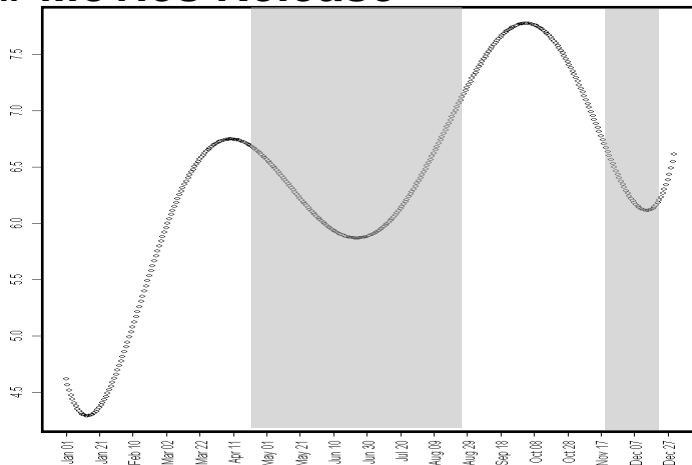
## Movie Data at a Glance



### Average Total Gross Revenue



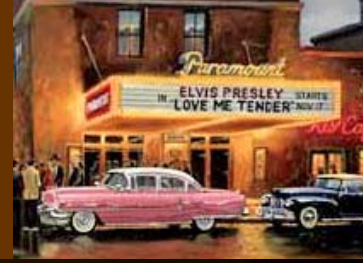
### # Movies Release



- Represent strong seasonality pattern
- High Gross Revenue during summer and winter
- Fewer movies are released during summer and winter (but stay longer)



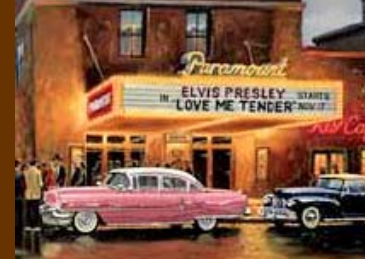
# Agenda



- Project Overview
- Data Exploration
- Data Analysis
- Our Findings
- Moving Forward

# Data Analysis

## Final Model – Breakeven Analysis



### The Regression Model

Input variables	Coefficient	Std. Error	p-value	Odds
Constant term	-5.30040979	0.70259899	0	*
Budget	-0.1557765	0.01487826	0	0.85575044
OpeningGross	0.39828286	0.04490587	0	1.4892652
IMDBRating	0.52095139	0.09765893	0.0000001	1.68362868
MaxTheaters	0.00137735	0.00016614	0	1.0013783
Summer Release	0.48275226	0.23492539	0.03988699	1.62052834
Christmas Release	1.22813332	0.3376188	0.00027516	3.41484904

Residual df	1001
Residual Dev.	593.456604
% Success in training data	33.03571429
# Iterations used	10
Multiple R-squared	0.53600442

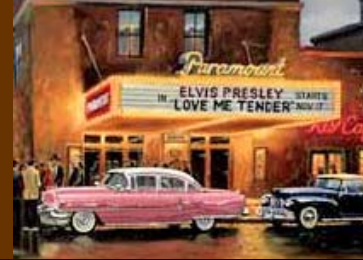
### Training Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable)	<b>0.5</b>
---	------------

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	247	86
0	26	649

Error Report			
Class	# Cases	# Errors	% Error
1	333	86	25.83
0	675	26	3.85
<b>Overall</b>	<b>1008</b>	<b>112</b>	<b>11.11</b>

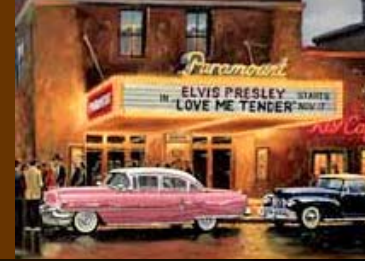
# Agenda



- Project Overview
- Data Exploration
- Data Analysis
- Our Findings
- Moving Forward

# Our Findings

## Factors that Help Create a Box Office Hit



### ➤ High budget means high risk

- Higher production budget movies are less likely to breakeven but may offer higher reward

### ➤ Quality still counts

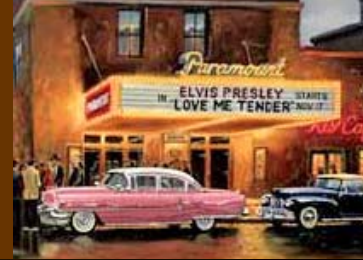
- Quality of the movie (as measured by IMDB rating) improves the chance to breakeven

### ➤ Action movies are as risky as documentary movies (...and others)

- With other predictors, movie genre is not a significant classifier

# Our Findings

## Factors that Help Create a Box Office Hit (cont.)



### ➤ Opening weekend success is important

- Opening gross is the key predictor of the success

### ➤ Distribution *DOES* matter

- In order to breakeven, a movie must make it into a large number of theaters

### ➤ Summer and Christmas

- Releasing movies during Summer and Christmas increase chance to breakeven



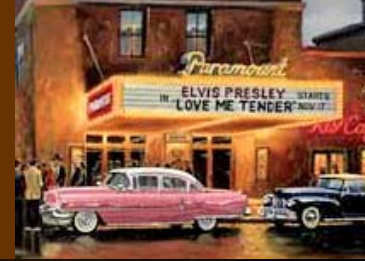
# Agenda



- Project Overview
- Data Exploration
- Data Analysis
- Our Findings
- Moving Forward

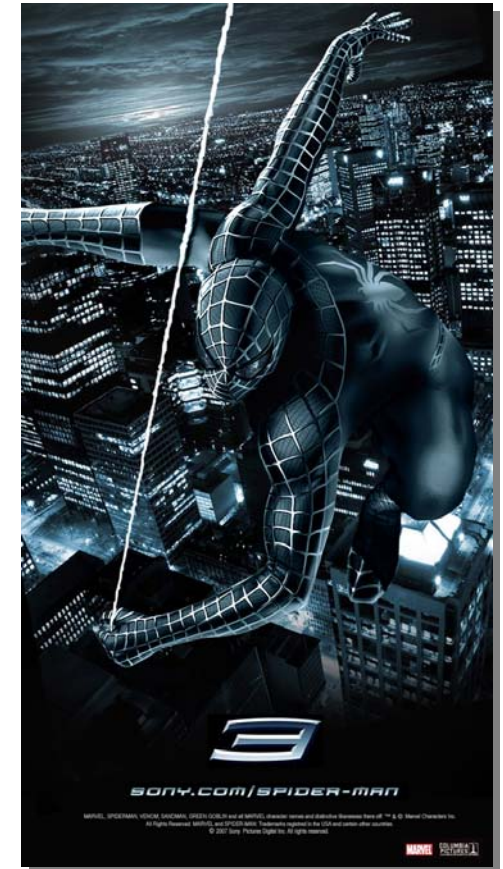
# Moving Forward

## What We Can Improve



### ➤ Obtain additional variables

- Promotion and Advertising budget
- Lead actors and actress salary
- Movies from comics or novel
- Payment to copyright owner



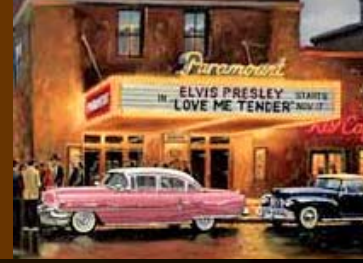
# Q&A





# Our Findings

Compare to Dr. Simonoff's Model



- **Opening gross** revenue is the most **powerful predictor** for both breakeven and total gross revenue
- Without other predictors, **documentary** movie performs the worse, the 3 best genres are
  - **Animation**
  - **Crime & Horror**
  - **Action & Adventure**
- **Increase budget** will **increase total gross** revenue but make the movie **more risky**