

# Predicting NPS

---

EXPLORATORY ANALYSIS

WILL SANDERS

# Situation

---

- The year, 2006, with only a week until the next blockbuster that produces a great return needs to be selected
- The executive producer recently read an article about data drive decisions
- To avoid reading all of the 10,000+ scripts, he decides to hire a data scientist

# Proposal

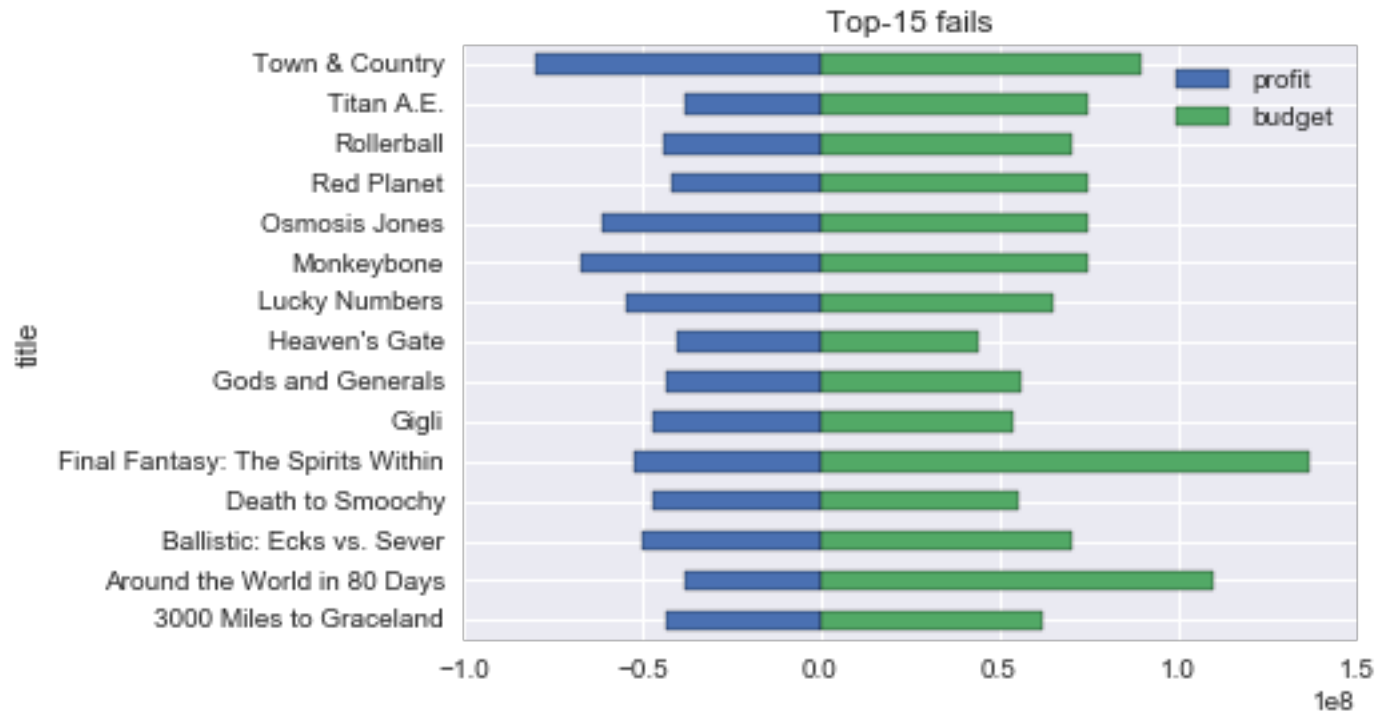
---

Can we predict the next blockbuster and maximize profit knowing the variables like genre, distributor, and mpaa rating?

Looking at data from 1980 to 2005, we were able to come up with a model for selecting the best movie to make the biggest return on our investment

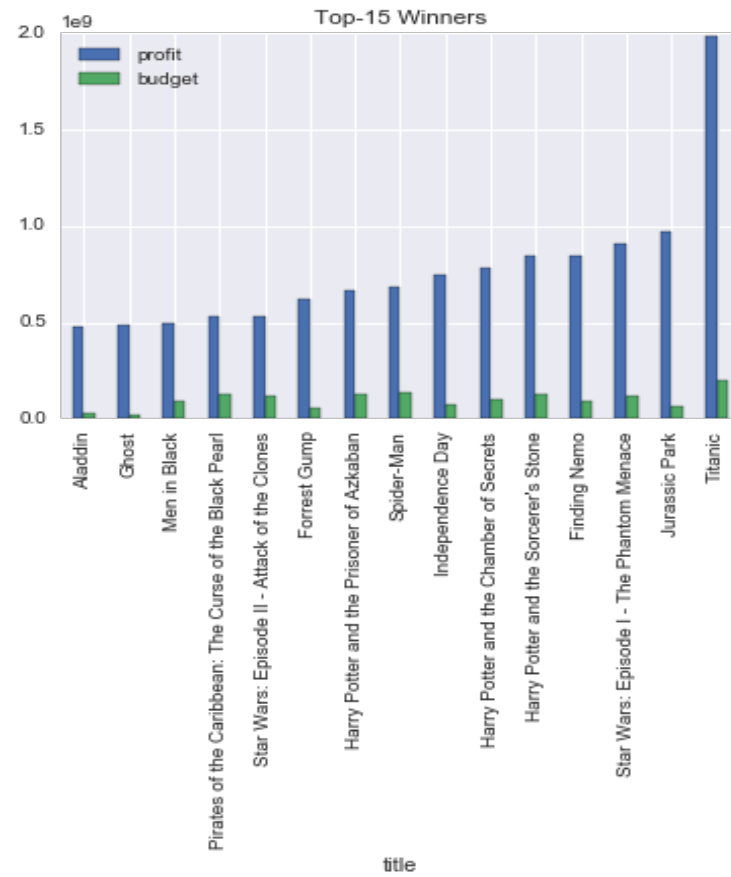
# What movies lost the most money?

---



# Winners

---



# Regression Model

---

Evaluated OLS and had 776 observations after cleaning the data

***Profit = 1 + mpaa[PG] + mpaa[PG-13] + mpaa[R] + votes count + budget\_bin + year + length + Romance + Action + distributor\_bin + opening\_gross\_bin***

	R^2	Adj. R^2
OLS	.567	.560

# Additional Ideas

---

Get release date data to better understand how that changes when to release movies

Compare other rating systems besides IMDB

# Data

---

Scraping Box Office Mojo from 1980 – 2005

Using IMDB Database from 1980 – 2005 that includes ratings

Joined both of the data sets together



# Movie NPS

---

