

# MOVIE PROPHET

Nelson Dsouza , Maria George, Saurabh Seth



BCG  
Digital  
Ventures



## TABLE OF CONTENTS

Executive Summary
Boston Consulting Group – Digital Ventures
Hypothesis
Solution – Architecture Diagram
Feature Engineering
Exploratory Data Analysis
Statistical Modeling
Results – Top features
Movie Investor Assurance System
Conclusion
Testimonials
Acknowledgements

## ABSTRACT

Movie Prophet is a predictive engine that forecasts the box office revenue given a movie proposal. The project done in partnership with Boston Consulting Group - Digital Ventures (BCG-DV) will enable the stakeholders to have a competitive edge by investing in successful movie ventures. The project encompasses end-to-end spectrum of data science - web scraping, data wrangling, feature engineering (30+ novel features), building Machine Learning models and data visualization. The outcome of our research is an integrated platform that serves as a movie investor assurance system. The research identified key features that contribute to the success of a movie based on the range of the budget. While directors play an important role in low budget movies, it is the actors that contribute more to the box office revenue in high budget movies.

# BOSTON CONSULTING GROUP – DIGITAL VENTURES



BCG  
Digital  
Ventures

## MISSION

“Forward-thinking companies need to rewire themselves around digital innovation, creating venture-like environments and build new offerings around customer behaviors and needs. BCG Digital Ventures was founded on the belief that it’s possible to translate start-up based innovation into the corporate world by tapping into the trillions of capital that is parked on the balance sheets. By doing so, global organizations can create new and untapped profit pools enabling them to become the new disruptors, and not be disrupted.”

**Jeff Schumacher, CEO, BCG-DV**

Boston Consulting Group - Digital Ventures (BCG - DV), founded in 2014 is a new initiative by the Boston Consulting Group, a pioneer management consulting firm. BCG DV is a digital innovation, corporate venturing and incubation business unit. The specialties of the firm include digital innovation, product development, strategic design, data science and product management venture teams. The multidisciplinary team breaks the traditional consulting regime to form strategic venture team with the client partners.



# HYPOTHESIS: CAN MOVIE REVENUE BE PREDICTED BEFORE PRODUCTION?

The global movie industry brings in over \$38 billion a year in box office receipts. However, more than 10 movies released in 2016 have losses of \$60 million or more.

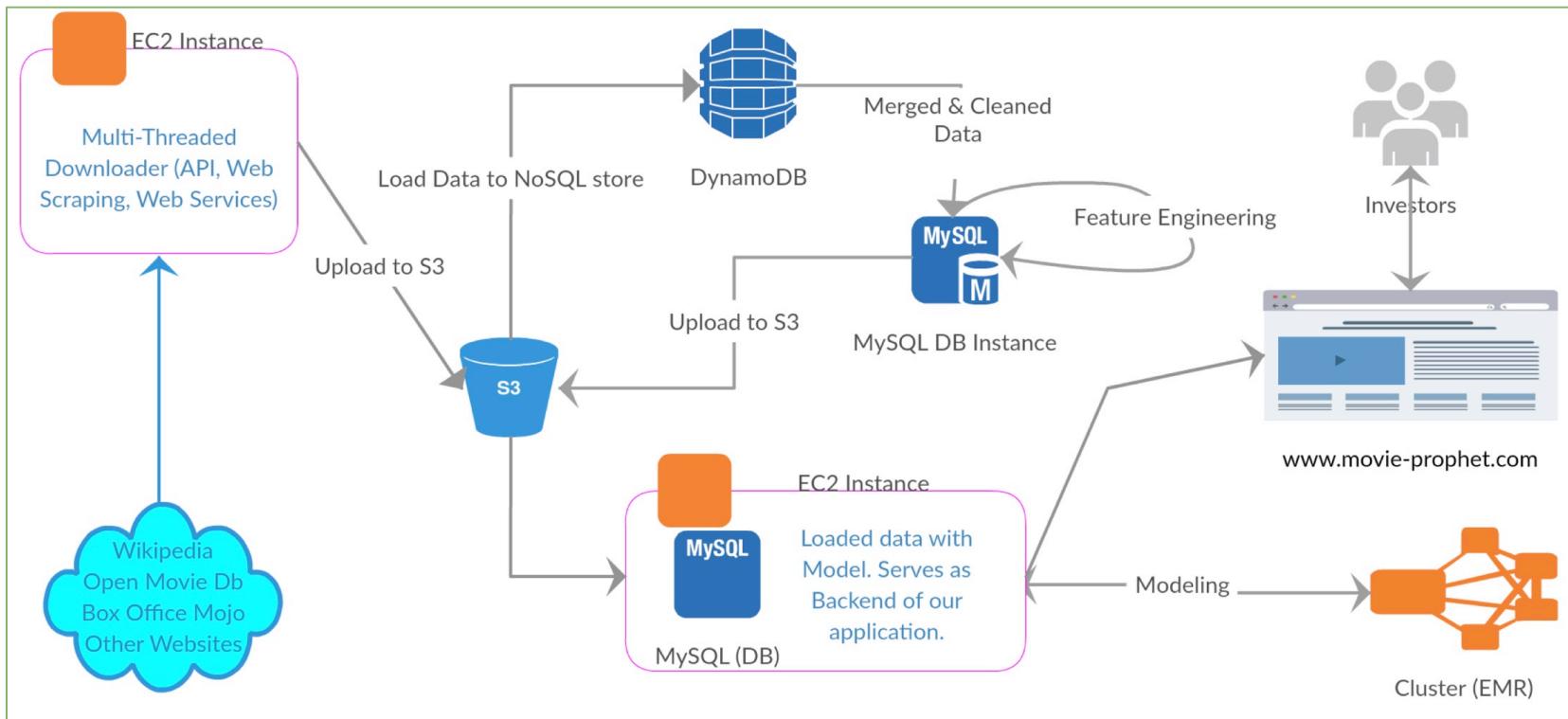


Movie studios pitch their story line to BCG-DV, who in turn advices people on investment opportunities. But how would BCG know which movies are worth investing?

*That's where we come in...*

- ❖ The initial hypothesis is that we can use the data about past movie releases to build a predictive engine that forecasts the movie proposal return-on-investment (ROI).
- ❖ Such an investor assurance system will enable BCG-DV to selectively stand behind successful movie ventures.

# THE SOLUTION



## Step 1: Fetching the data

- ❖ Fetched more than 45K movies from public web sources
- ❖ Cleaned and integrated the data using jaro-winkler algorithm



WIKIPEDIA  
The Free Encyclopedia

Box Office Mojo



## Step 2: Feature Engineering

created **30+ unique features**, including

- ❖ budget
- ❖ runtime
- ❖ release quarter
- ❖ release day of the year
- ❖ release week of the year
- ❖ award score
- ❖ holiday season
- ❖ mpaa rating
- ❖ genre
- ❖ composer score
- ❖ director score
- ❖ writer score
- ❖ distributor score
- ❖ actor score
- ❖ cinematographer score
- ❖ producer score

performed k-means clustering to identify 10 unique genres

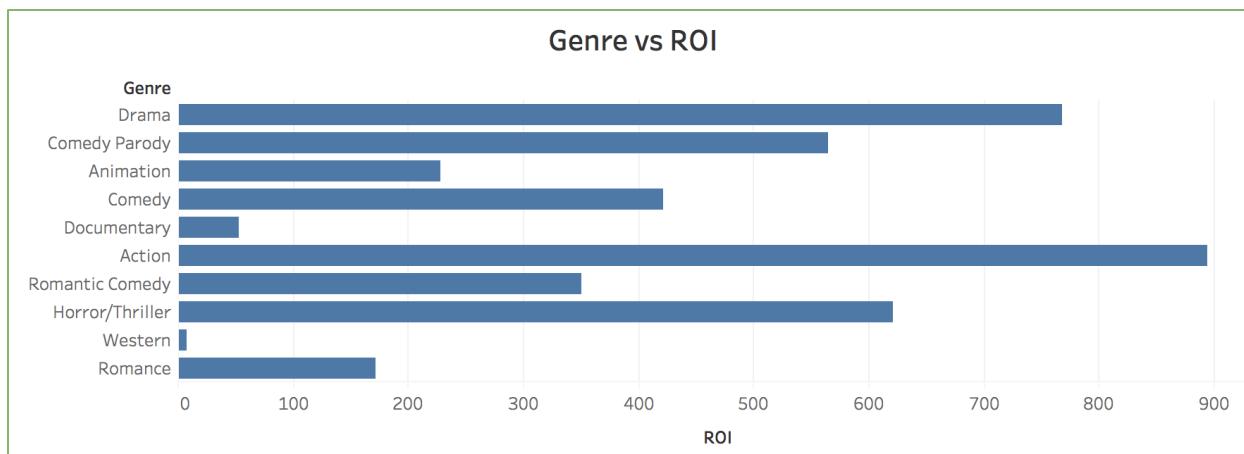
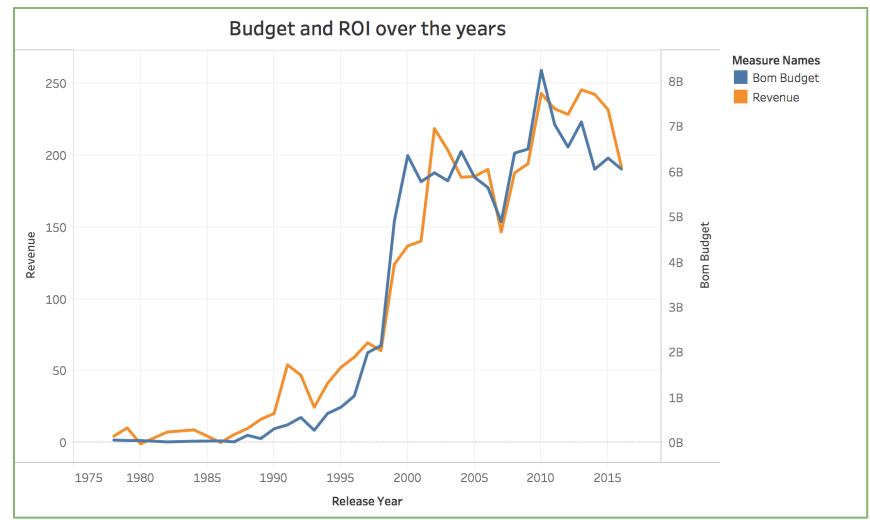
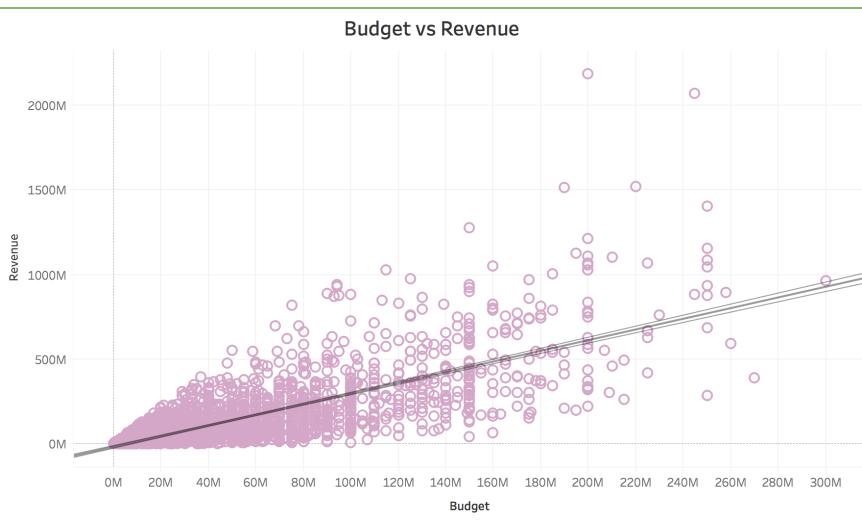
created a score for 4000+ unique actors based on the success rate of the past movies and awards won

sample output

<b>Christian Bale</b>	<b>3.5</b>
<b>Hugh Jackman</b>	<b>3.1</b>
<b>Adam Sandler</b>	<b>0.3</b>

## Step 3: Exploratory Data Analysis

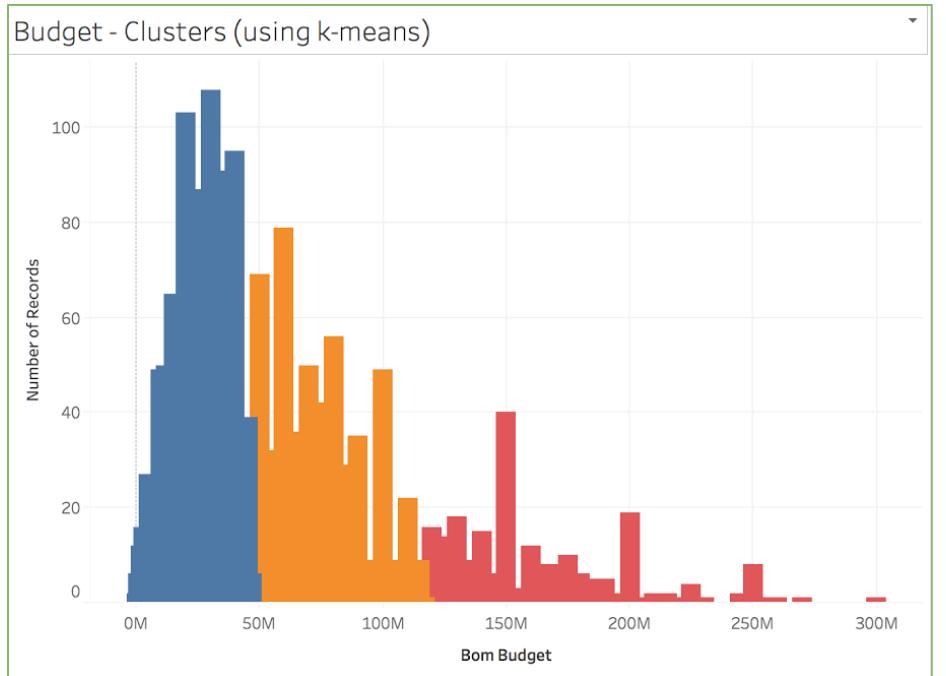
Performed Exploratory Data Analysis to analyze the relationship between response variable (ROI) and various features.



## Step 4: Building Machine Learning models

performed k-means clustering to categorize budget into **three** clusters

applied **gradient boosting algorithm** to each cluster to predict ROI



Low Budget

<47M \$

RMSE: 1.78

R-squared: 0.481

Medium Budget

48M - 117M \$

RMSE: 1.43

R-squared: 0.420

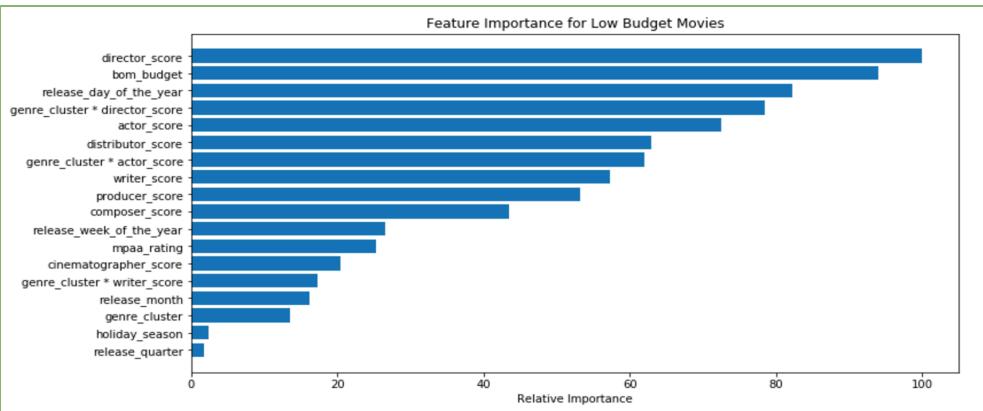
High Budget

>120 M \$

RMSE: 1.21

R-squared: 0.407

# RESULTS: Top predictors for ROI of a movie

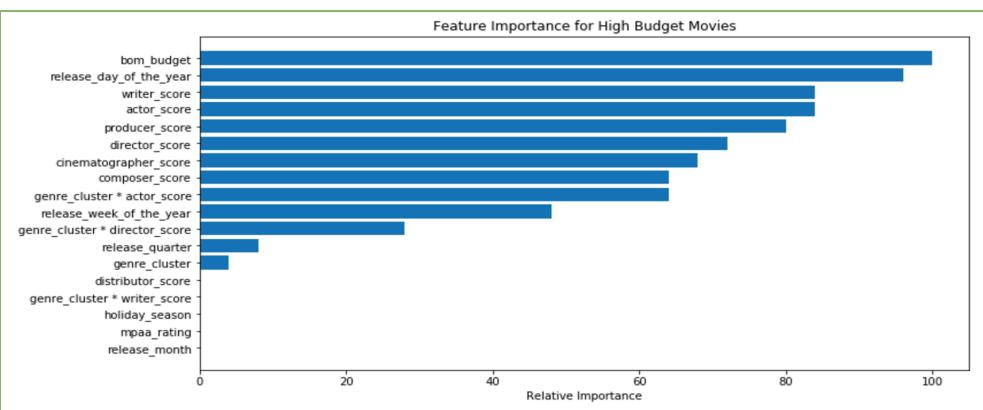
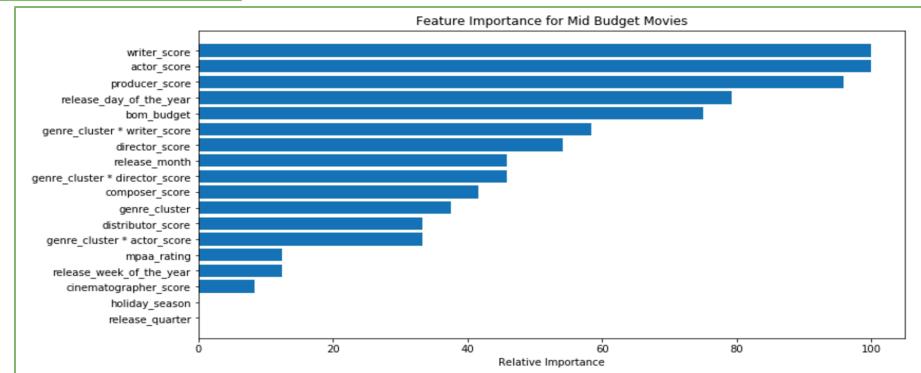


Low Budget  
Top 3 features

director  
budget  
release day of the year

Medium Budget  
Top 3 features

writer  
actor  
producer



High Budget  
Top 3 features

budget  
release day of the year  
writer

# MOVIE INVESTOR ASSURANCE SYSTEM

Integrated web interface that invokes the machine learning model based on user input parameters.

Movie Prophet Explore ▾

About Prediction Charts Tables

## Prediction Based on Movie Proposal

Movie Name\*  
Interstellar 2  
Movie Name will not impact prediction

Budget (\$M)\*  
300

Release Date\*  
06/03/2017

Genre\*  
Action

Actors\*  
Scarlett Johansson, Matthew McConaughey

Directors\*  
Christopher Nolan

**For the Movie:**  
Interstellar 2  
With budget:  
\$300 Million

**Revenue is:**  
 \$842 Million ↑  
ROI: 2.81%

**\*Details of your movie proposal:\***

**Release Date:** 2017-06-03  
**Genre:** Action  
**Actor/s:** Scarlett Johansson, Matthew McConaughey  
**Director/s:** Christopher Nolan  
**Producer/s:** James Cameron  
**Writer/s:** Sam Harper  
**Cinematographer/s:** Geoff Zanelli  
**Composer/s:** Tobias A. Schiessler  
**Distributor/s:** Warner Bros. (New Line)

Have a cool movie proposal? Find out how it fares at  
[www.movie-prophet.com](http://www.movie-prophet.com)

# CONCLUSION & NEXT STEPS

- ❖ Created an extensive dataset with information about more than 50K+ movies released between 1900 – 2017
- ❖ Engineered 30 + novel features including actor score and holiday season to improve prediction accuracy
- ❖ Built machine learning model that explains more than 50% of variance in the data based on range of the budget
- ❖ **Impact:** An integrated media decision platform for movies to be used by BCG-DV and movie studio executives

## Looking Forward:

- ❖ Incorporate sentiment analysis to understand the current popularity of actor/ director on social media
- ❖ Perform social network analysis to identify the best combination of actor-actor and actor-director to improve ROI
- ❖ Include more predictors in the data including weather, topic identification on plot to improve prediction accuracy
- ❖ Provide recommendation on actors/ directors based on genre of the movie

## TECHNOLOGY STACK



# TESTIMONIALS

"I had the pleasure of working with Maria, Nelson and Saurabh during their Capstone Project at BCG DV. They showed great dedication, problem-solving and technical capabilities during all phases. Their project involved four main disciplines: Data engineering, data science, full-stack application engineering, and design. The project showcased their multiple talents and execution. The team effectively divided tasks and integrated components. Each member exceeded my expectations and will bring creativity and calculated execution to any companies they join."

**Farshad Kheiri, Sr Data Scientist, BCG-DV**

"The UW team Nelson, Saurabh and Maria did a fabulous job in executing their Capstone Project. They first started in understanding and mapping what needs to be done and quickly came up with estimates. Once they had estimates they went on full throttle on executing the project. They kept to their timeline fairly well. The team did encounter several technical hurdles which they overcame in a steadfast manner. They were always learning and coming up with innovative solutions to the problems they encountered. Most of all it was impressive to see great team work with each individual working in unison and assisting each other when needed. I am very pleased with the final product and their dedication in executing it in conjuncture with BCG Digital Venture"

**Mario Gerard, TPM, BCG-DV**

# ACKNOWLEDGEMENTS

We would like to express our gratitude to Mario Gerard and Farshad Kheiri from BCG –Digital Ventures for all the guidance and mentorship.

We would also like to thank the iSchool faculty Sean T McGann and Teaching Assistant, Stephen Keating for all the continued support and encouragement.

Finally, to all our friends for their valued feedback and constructive criticism that has contributed immensely to the successful completion of the project.

Nelson, Maria, Saurabh