

ROCKBUSTER

stealth

Data Analysis

Q3 '05 Rentals

Aaron Manzano





Rental Data: Q3 '05

(15/06/2005-02/09/2005)

Table of Contents

Introduction	1-4
Geographic Analysis	5-8
Data Metrics	9-13
Modeling	14-20
Key Optimizations	21-25
Recommendations	26-27
Conclusion	28



RockBuster *stealth*

Customers

Global Movie Rentals in • 109 Countries

• 14,413 Transactions

• \$4.25 Average transaction amount:

• 599 Customers

• \$102 Avg customer spend

• (~24 rentals)



RockBuster *stealth*

Films

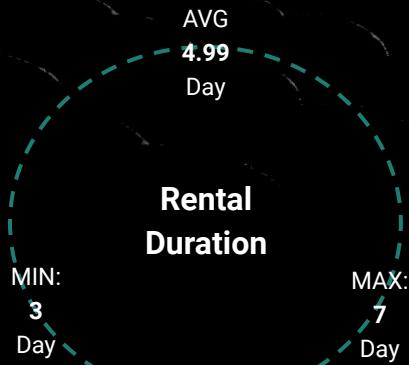
Film Catalog with • 958 Titles

• 17 Categories

• 1-8 Copies per title

• 4,581 Total movies

• \$10-\$30 Replacement cost
• (avg \$20)





Overview

Define Key Metrics

Gross Profit Analyzes

Return on Investment.

- Total Revenue
 - For 3 Months
- Inventory Cost
 - 1-8 film copies
- Identify Important Variables

Optimize Catalog

Catalog Growth Strategy:

- Identify weak titles
 - Low ROI Films
- Remove weak titles
 - Sell physical copies
 - Don't renew digital licenses

Geographic Analysis

Identify Top Markets:

- Top Countries
 - By Customer Base
- Top Customers
 - By Total Spend
- Top Categories
 - In Top Markets

Key Performance Metric

Gross Profit

Total Revenue

Gross Revenue from all payments with that Film. Overall, about 30% of revenue is from late fees.

Inventory Cost

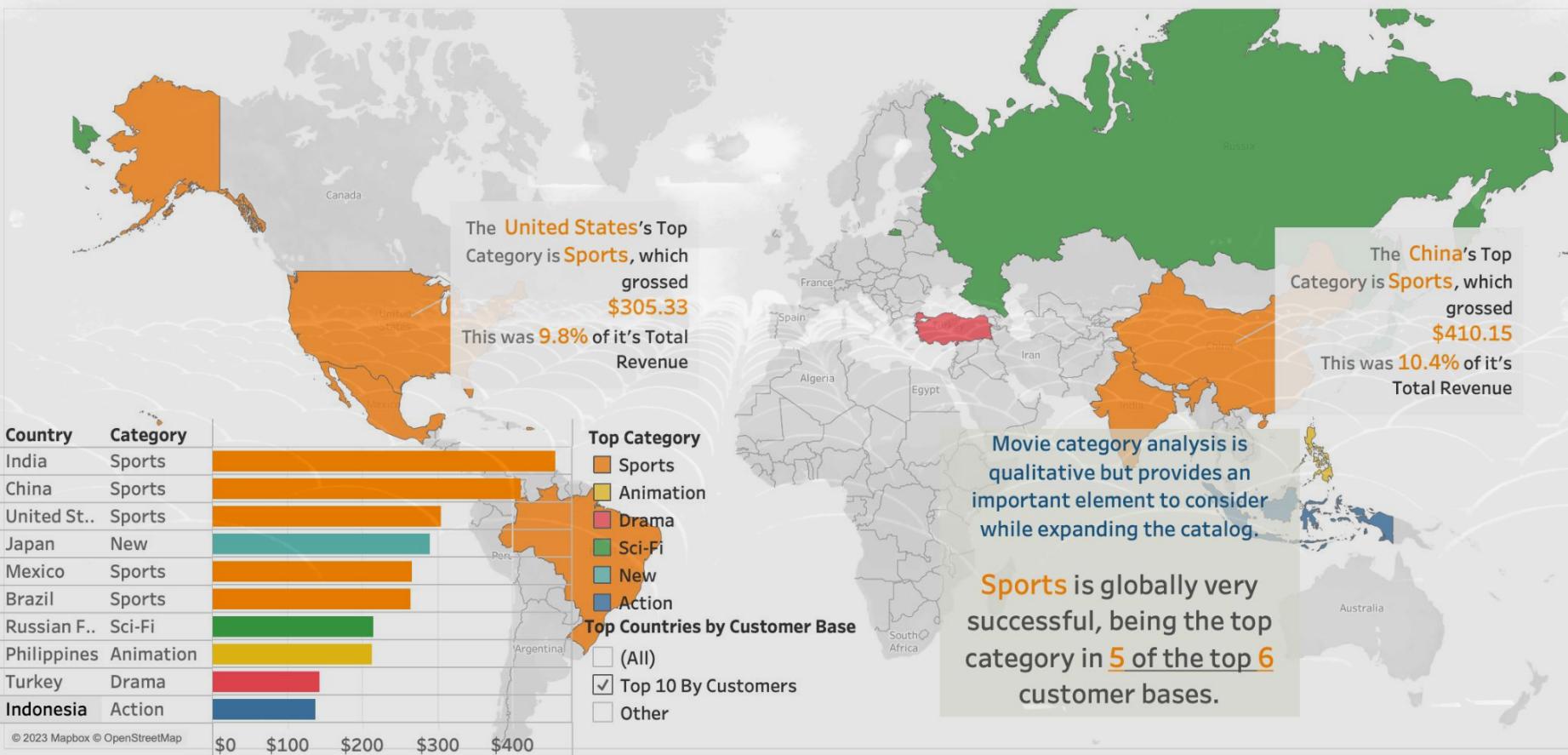
Total Cost from multiplying the replacement cost by the number of copies in inventory.



Geographic Analysis

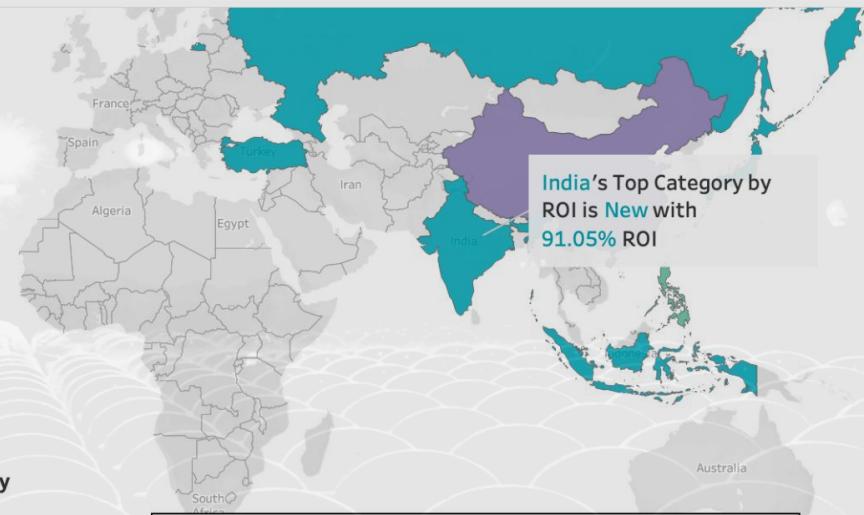
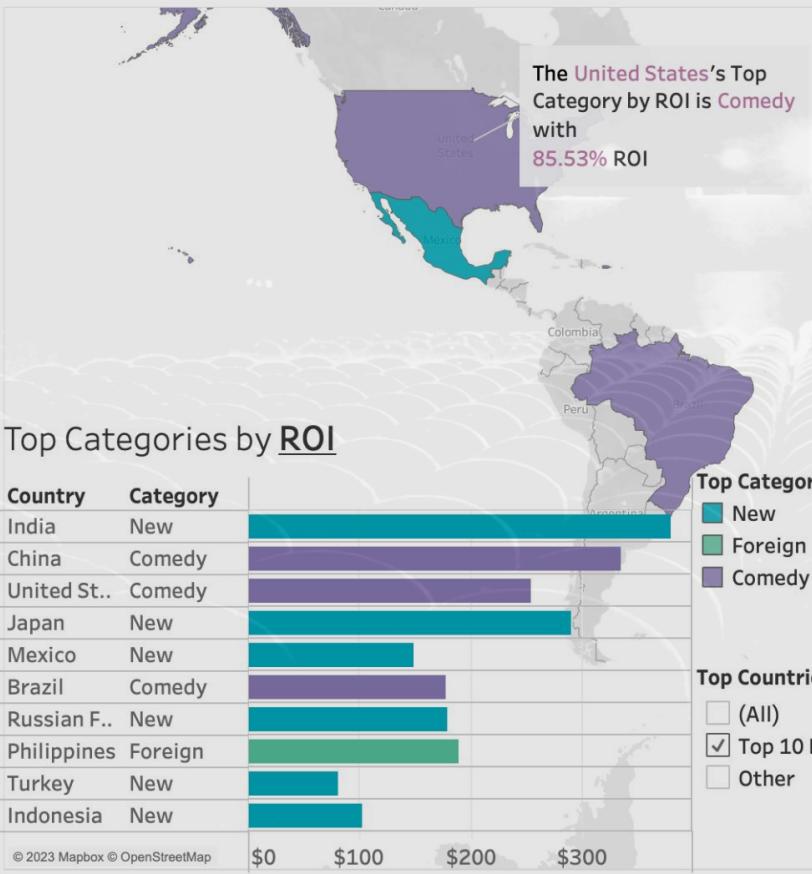
5

Top Categories in Top Markets



Geographic Analysis

Top Categories in Top Markets: By Return on Investment



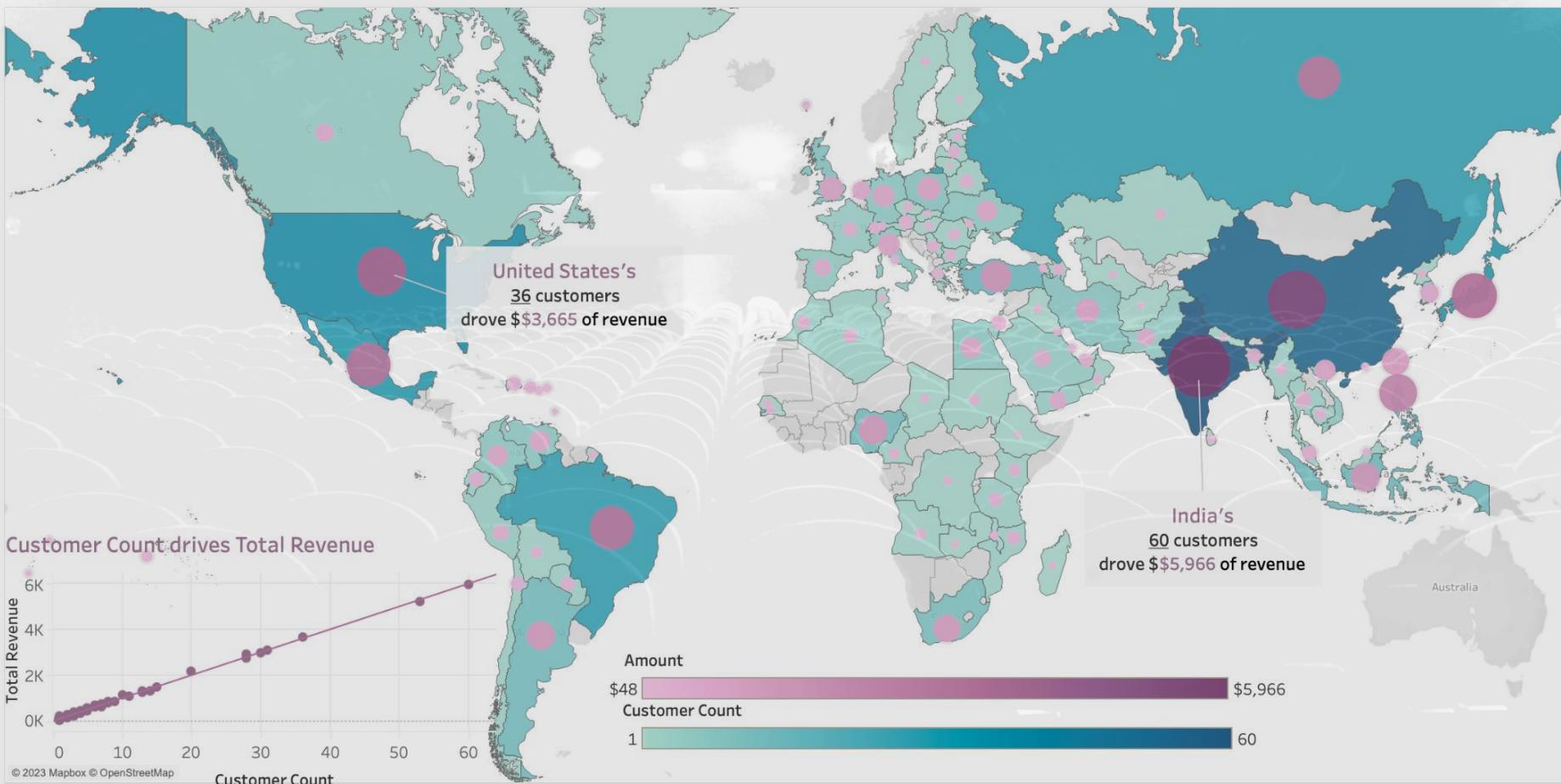
New is the top category by ROI in 6 of the top 10 countries by customer base.



Geographic Analysis

7

Map: Customer Base and Total Revenue





Geographic Analysis

8

Map: Top Countries and Customers

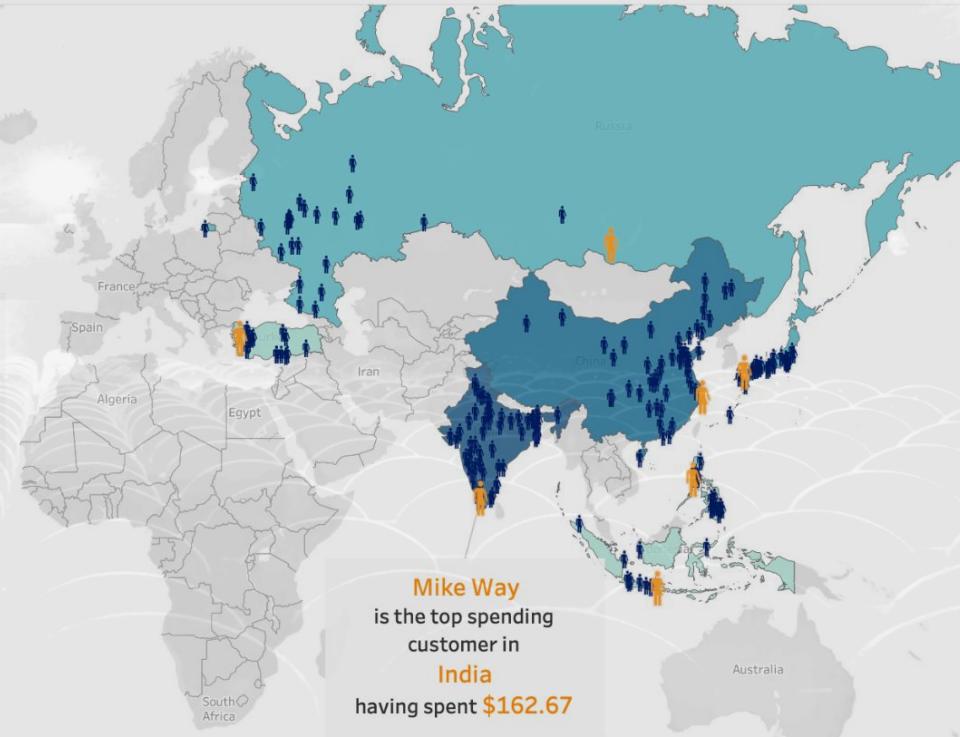


Top Customers

Customers

Top Countries by Customer Count ↓

Top Countries	Top Customer	Customer Count	Top Spend
India	Mike Way	60	\$162.67
China	Guy Brownlee	53	\$151.69
United States	Karl Seal	36	\$208.58
Japan	Angela Hernandez	31	\$138.65
Mexico	Alex Gresham	30	\$143.68
Brazil	Marion Snyder	28	\$194.61
Russian Federation	Warren Sherrod	28	\$152.69
Philippines	Marcia Dean	20	\$166.61
Turkey	Brandon Huey	15	\$145.64
Indonesia	Steve Mackenzie	14	\$152.68



Top Countries by Customer Base

(All)

Top 10 By Customers

Other

Customer Count

14

60

50 unknown

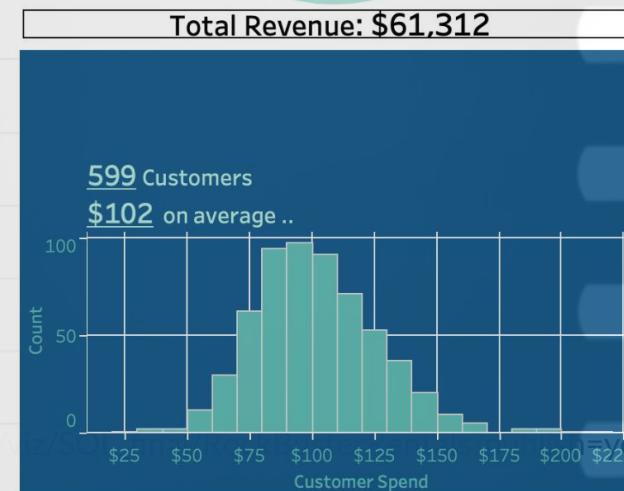
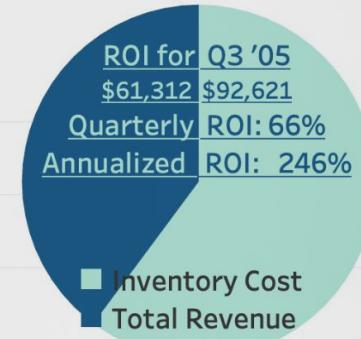
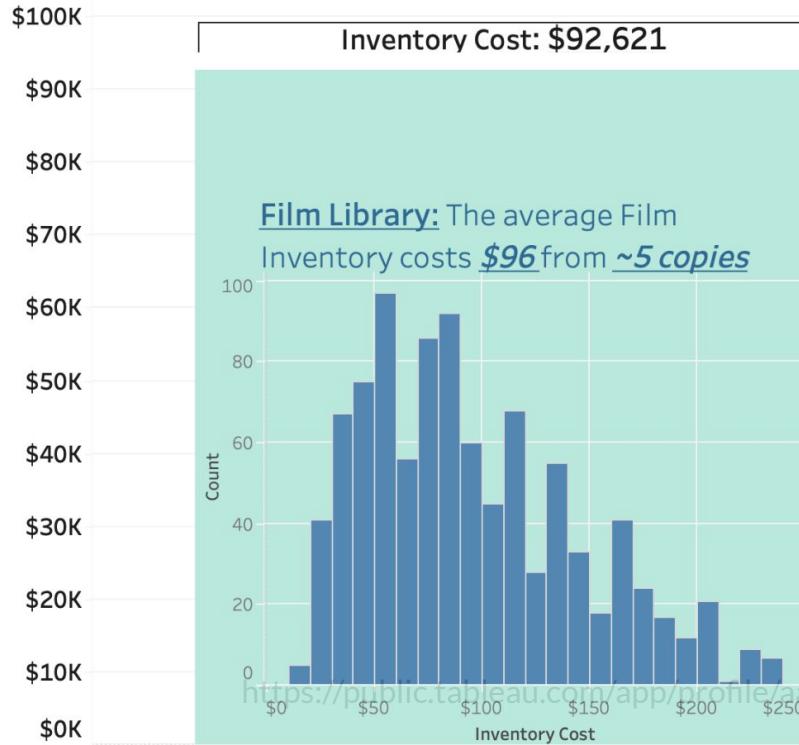


Data Metrics

9

Cost and Revenue:

Breaking down the Key Metric: Return on Investment

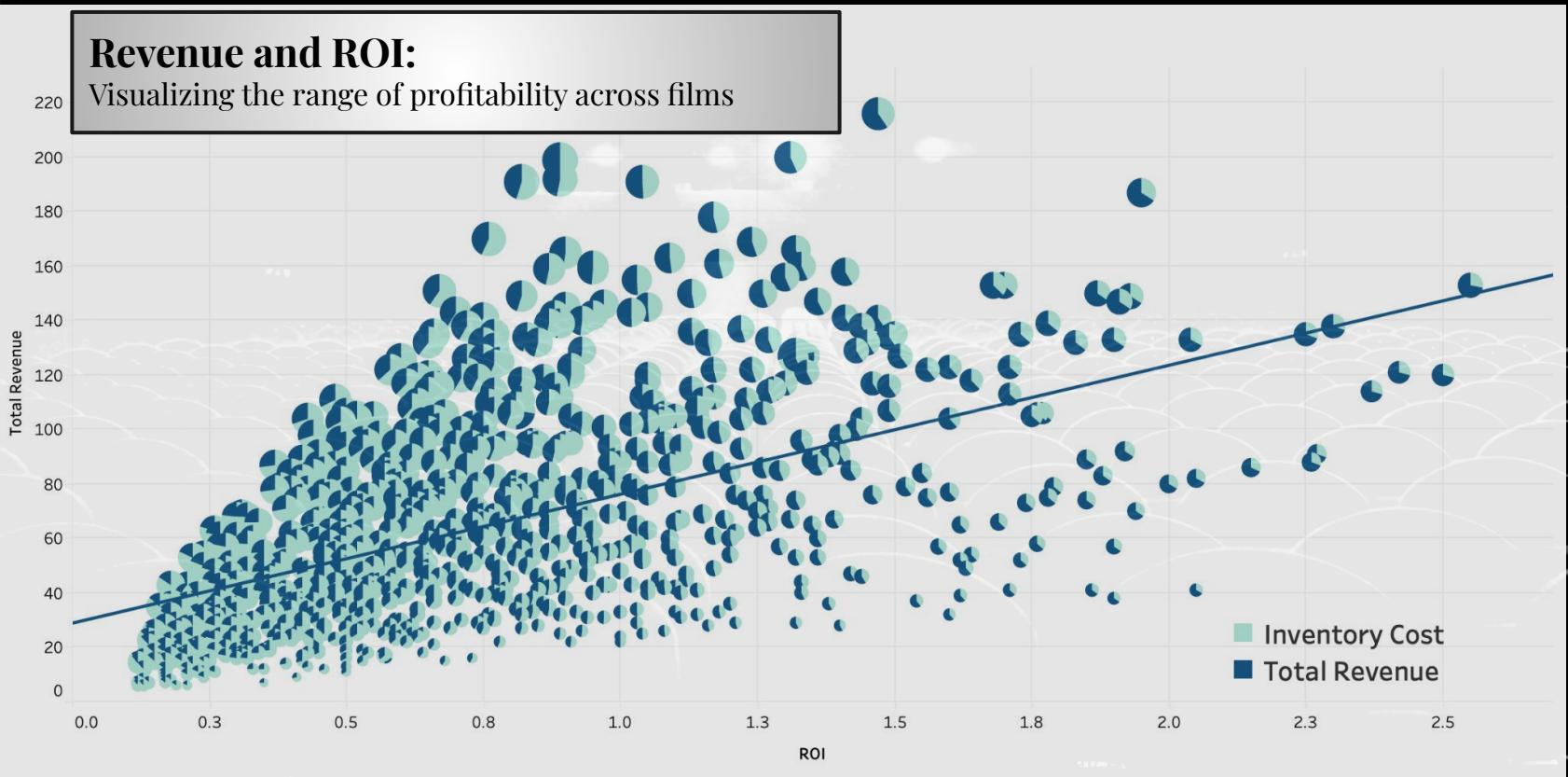


<https://public.tableau.com/app/profile/aaron.franzana>



Revenue and ROI:

Visualizing the range of profitability across films





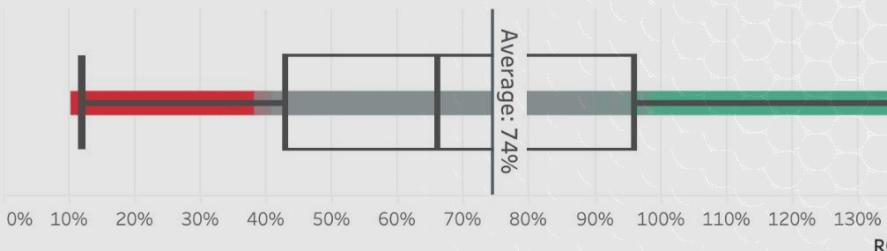
Determining the Least Profitable Titles:

The *Shapely Values* indicate that [Quantitative > Qualitative](#).

Instead of choosing between categories or ratings, we select the [bottom quartile](#) of all categories by Gross Profit.

This measure of [Inventory Cost / Total Revenue](#) provides

Gross Profit by Category Quartiles



Low Profit Titles (244)↑

Cruelty Unforgiven: 12% Japanese Run: 13%
Freedom Cleopatra: 12% Lawrence Love: 13%
Hollywood Anonymous: 12% Silence Kane: 13%
Sassy Packer: 12% Texas Watch: 13%
Clockwork Paradise: 13% Bride Intrigue: 14%

Profitable Titles (467)↑

Hills Neighbors: 33% Chitty Lock: 39%
Dalmations Sweden: 35% Midnight Westward: 39%
Glass Dying: 36% Army Flintstones: 40%
Extraordinary Conquerer: 38% Doctor Grail: 40%
Reunion Witches: 38% Carrie Bunch: 41%

Top Quartile + Outliers (247)↓

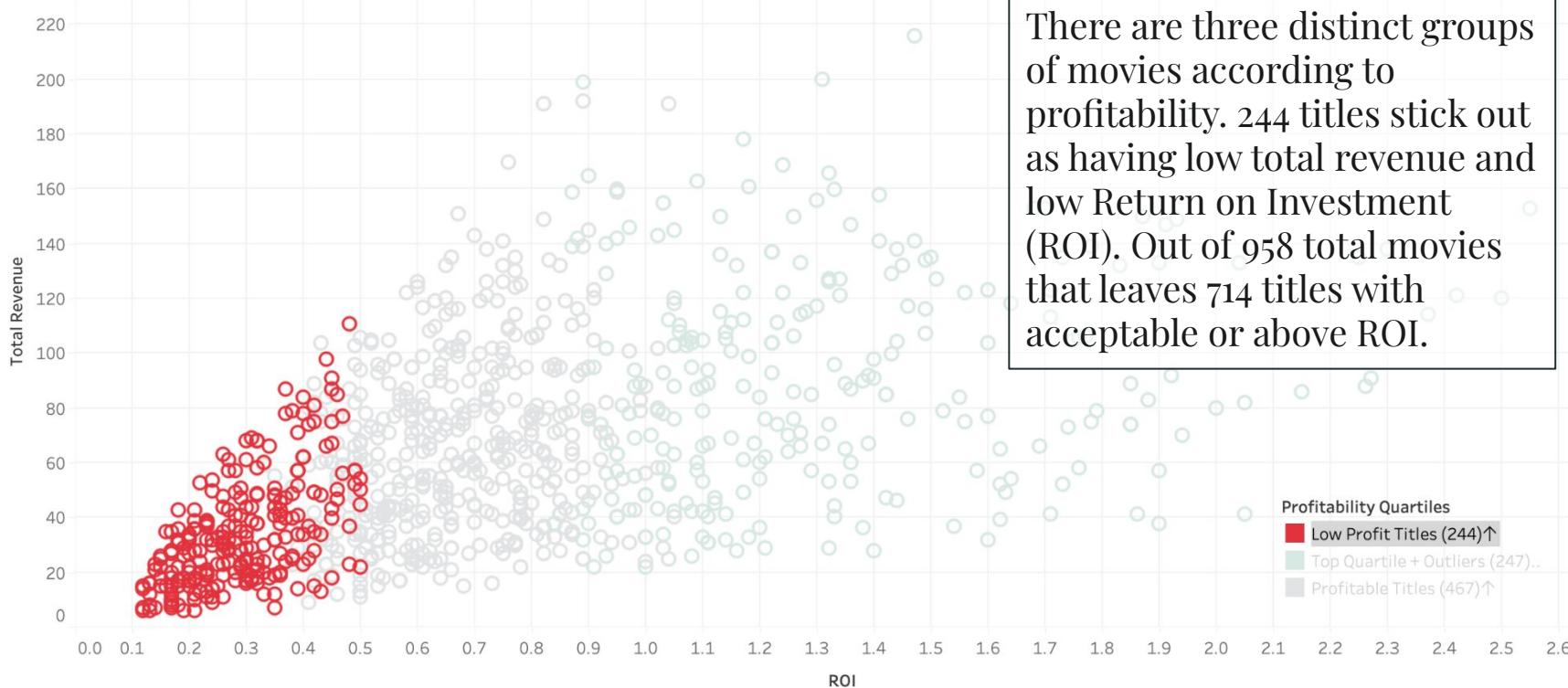
Maiden Home: 255% Dude Blindness: 227%
Trap Guys: 250% Bright Encounters: 226%
Kissing Dolls: 242% Flintstones Happiness: 225%
Whale Bikini: 237% Daisy Menagerie: 215%
Fellowship Autumn: 230% Paths Control: 205%





Profitability Quartiles:

Identifying low profit titles

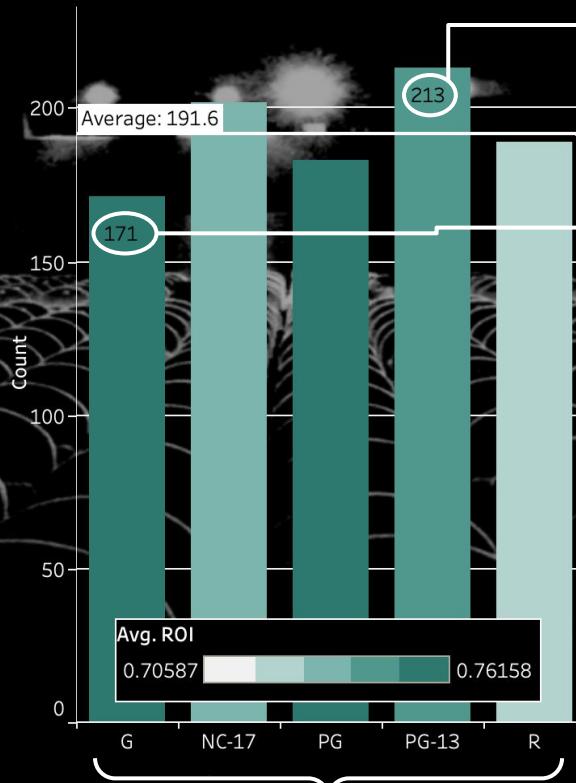




5 Ratings

Rating Count:

Higher film count
doesn't always lead to higher ROI



Max Count: 213
Avg Count: 191.6
Min count: 171

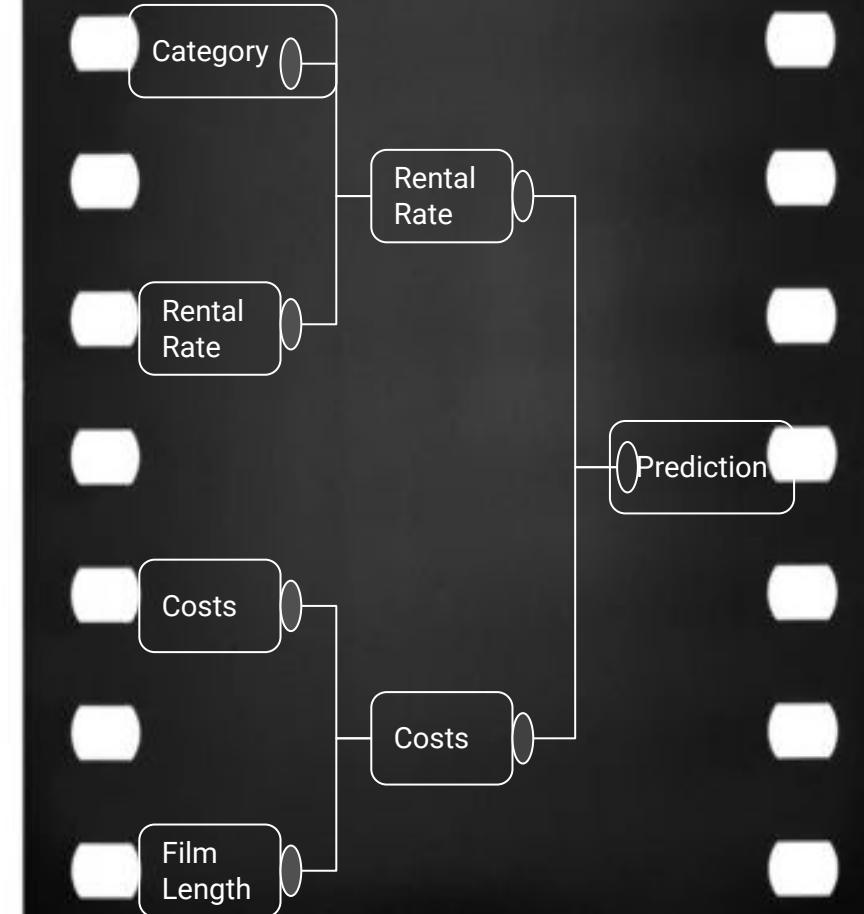
The darker the Hue, the
higher the Rating's ROI.

Higher number of films in
that Rating don't seem
associated with higher ROI,
as the Min Count has very
high ROI.

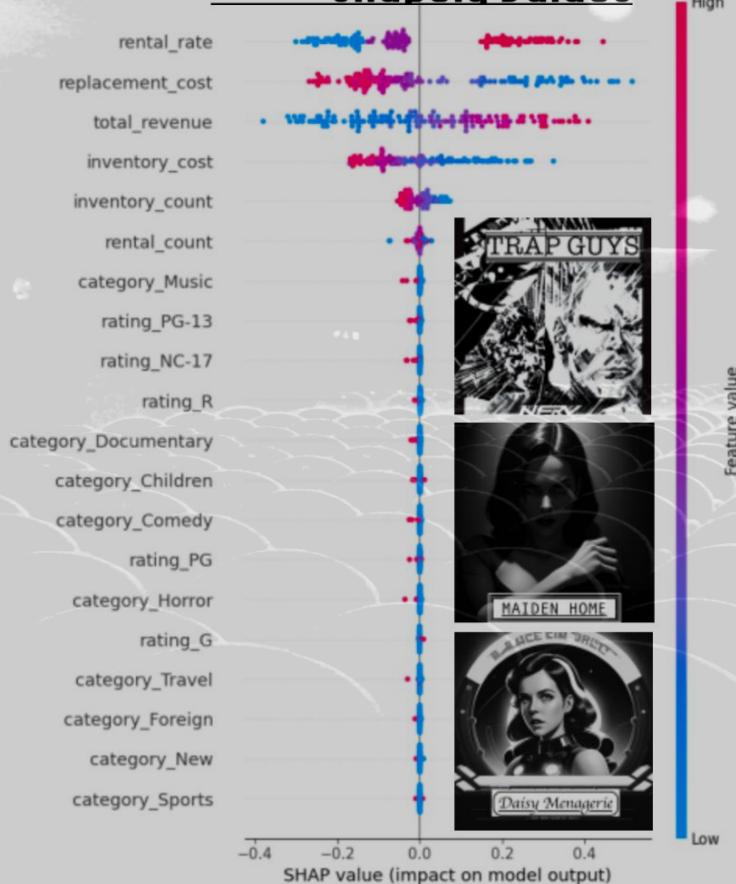


How to determine which features are important in *driving ROI?*

- Machine Learning is an asset
- Model interpretability is important to understand how the model is weighing different features
 - Features: Rental rate, replacement cost, categories, etc
- Train the model on predicting ROI
 - Use interpretability metrics to define a path towards higher ROI



Shapely Values



XGBoost is a preferred choice among experts for tackling intricate data challenges and generating robust predictive models.

- Decision Tree ML model trained to predict film's Profit Ratio ($\frac{\text{total revenue}}{\text{inventory cost}}$)
- Low MSE (0.0093) of the model indicates reliability

Shapely values offer a pragmatic framework for fairly attributing feature importance in these models, quantifying the practical significance of each variable's role in the model.

- SHAP values closer to zero indicate features with less predictive value in the model. Ratings/categories aren't predictive.
- Feature values are shown in blue (low values) to red (high values)

- 1) Rental_rate, Replacement_cost, Total_revenue are furthest from 0 and are therefore contribute most to the model.
- 2) Quantitative data is much more predictive than qualitative data (like categories and ratings).
- 2) High rental rates and low replacement costs are the key drivers of profitability.



PROFIT

Interpreting a model
to determine feature
importance.

Shapely Values *far from 0* are more important

Blue values are low variables, red is high.

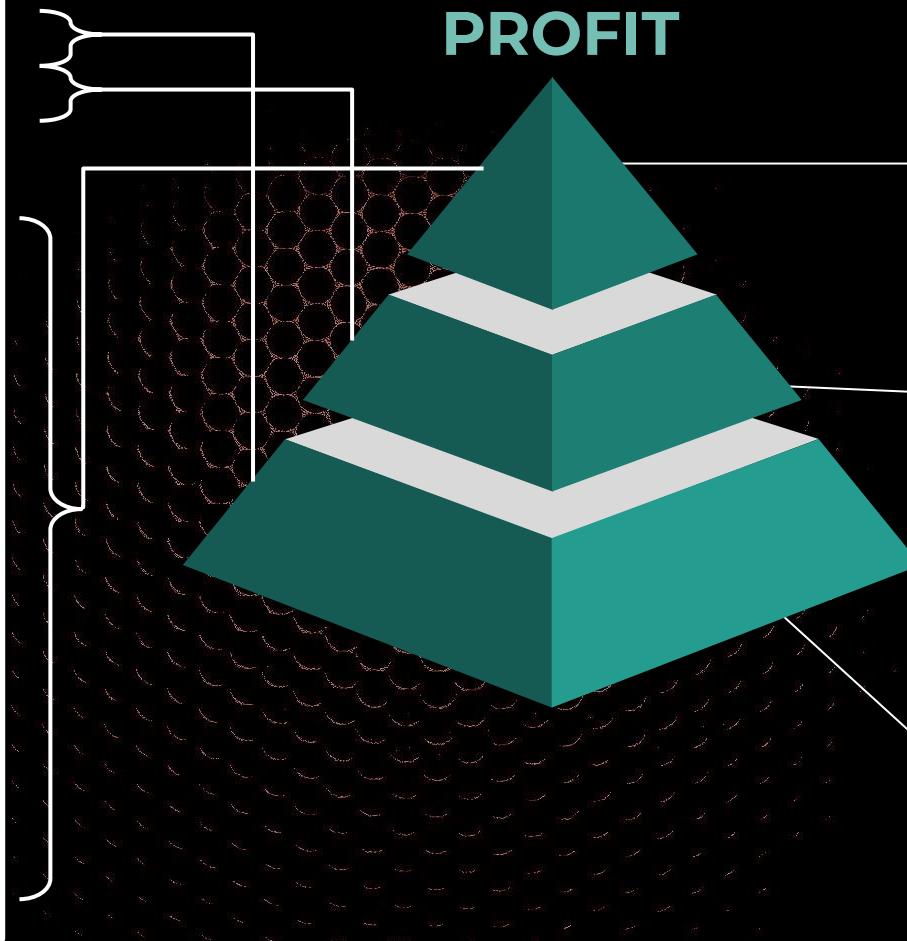
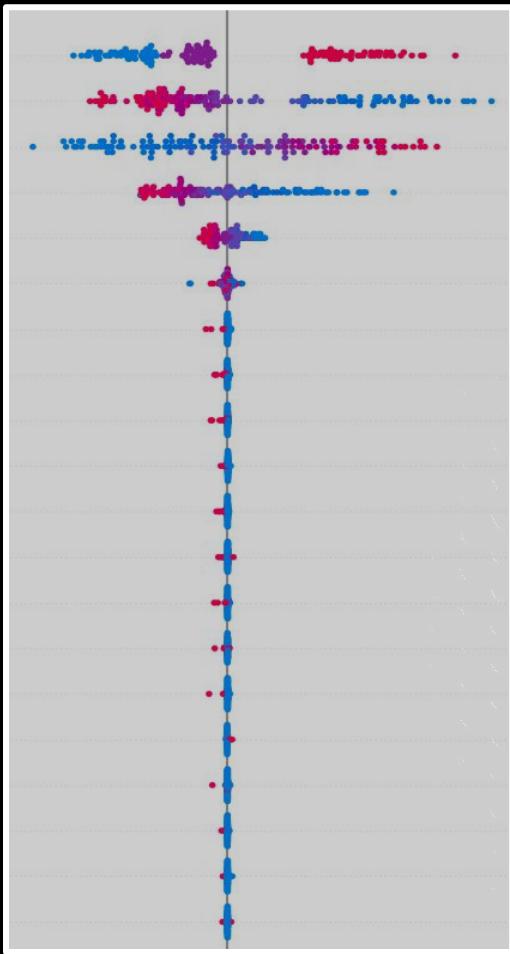
Low rates and high costs predict Low Profit

High rates and low costs
predict High Profit

rental_rate

replacement_cost

0



Categories and ratings

1 Qualitative data is closer to zero and less useful

Replacement Costs

Maintaining low inventory costs predicts profitability

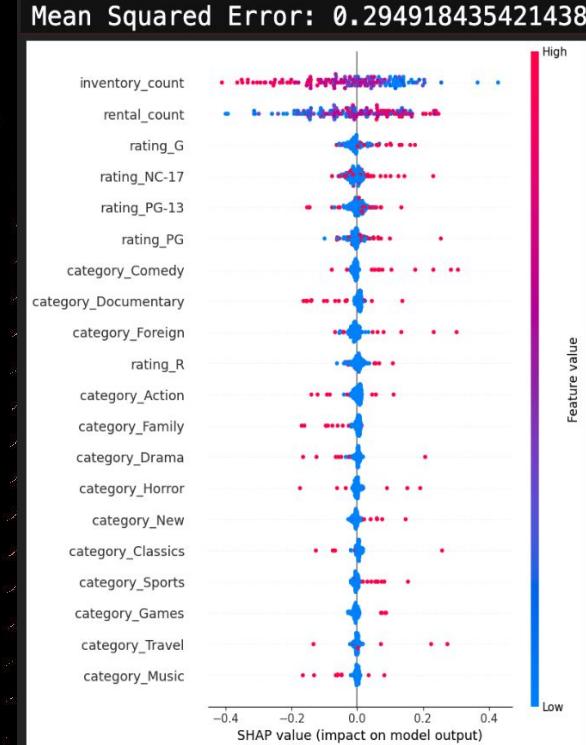
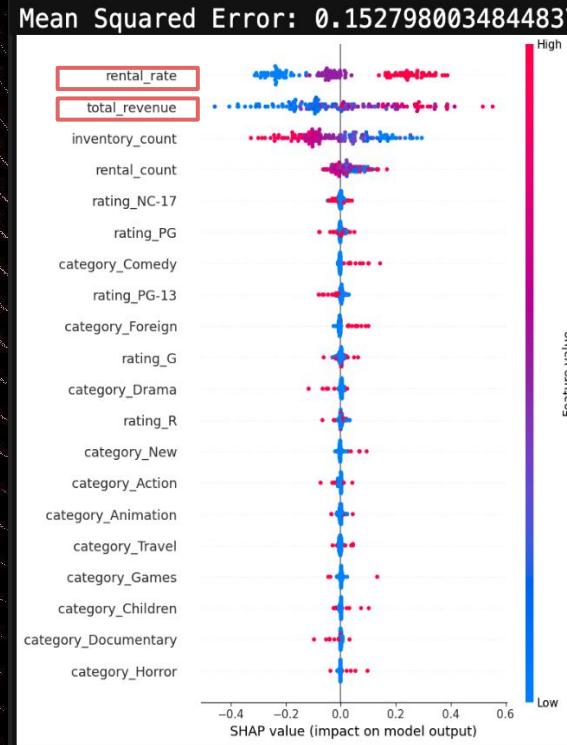
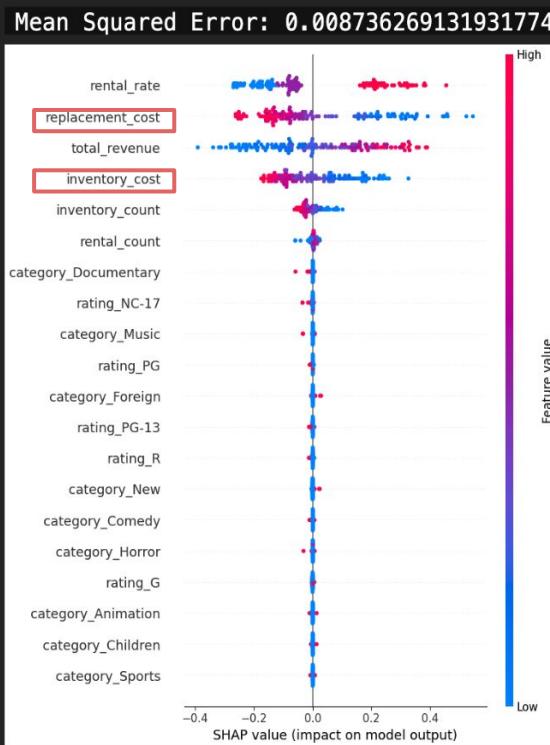
Rental Rates

High Fees are the foundation of profitability



Feature Leakage

Feature leakage is when information from the target variable is directly or indirectly being used by the model to predict said target variable. This hinders generalization to new data and predictive capabilities of the model. Notice how the MSE degrades as leaking data highlighted in red are removed.

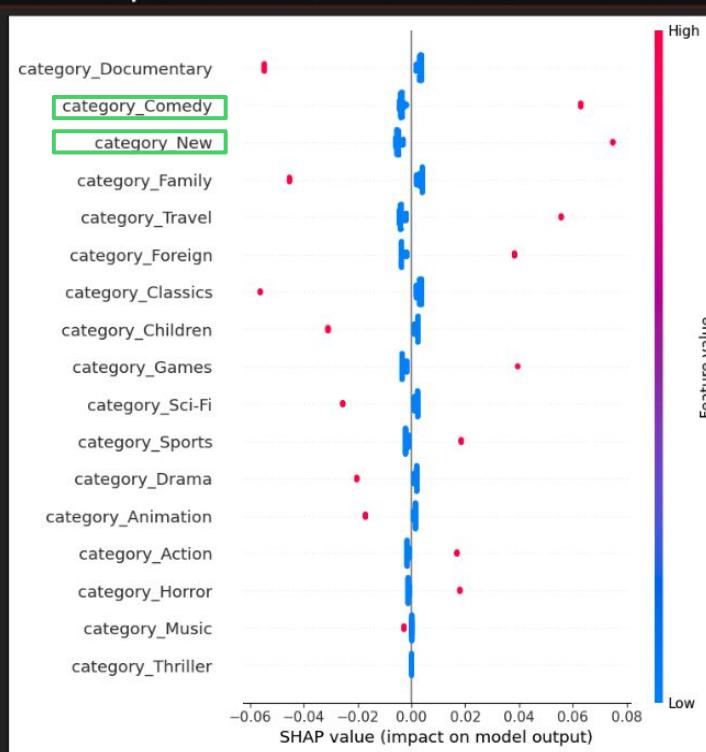




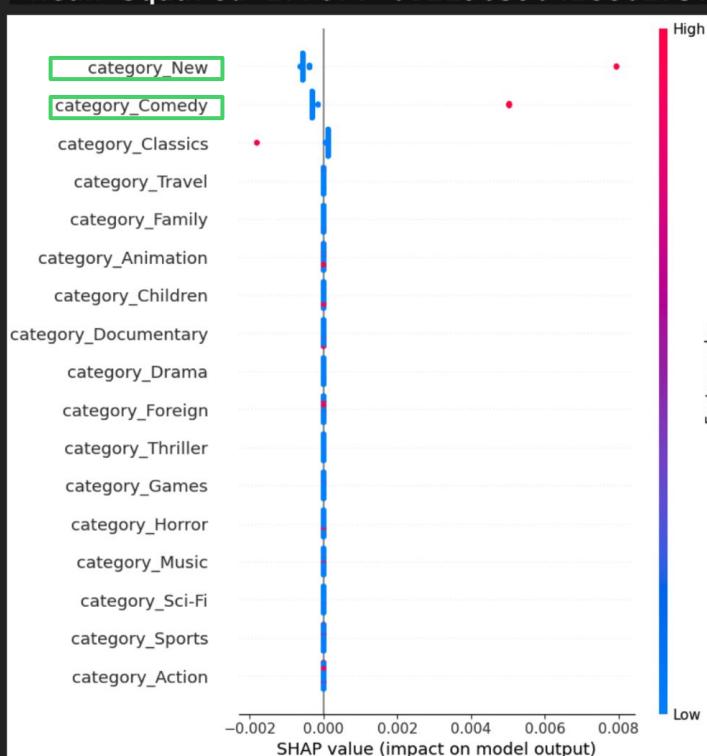
Categories | GridSearch

Here we focus only on categories and train for gross profit, then again after performing grid search for optimal hyperparameter settings. The MSE is high on both, and so predictive accuracy is questionable, but the feature importance can help separate noise from signal. Comedies and New stand out as profitable categories.

Mean Squared Error: 0.21681947275400443



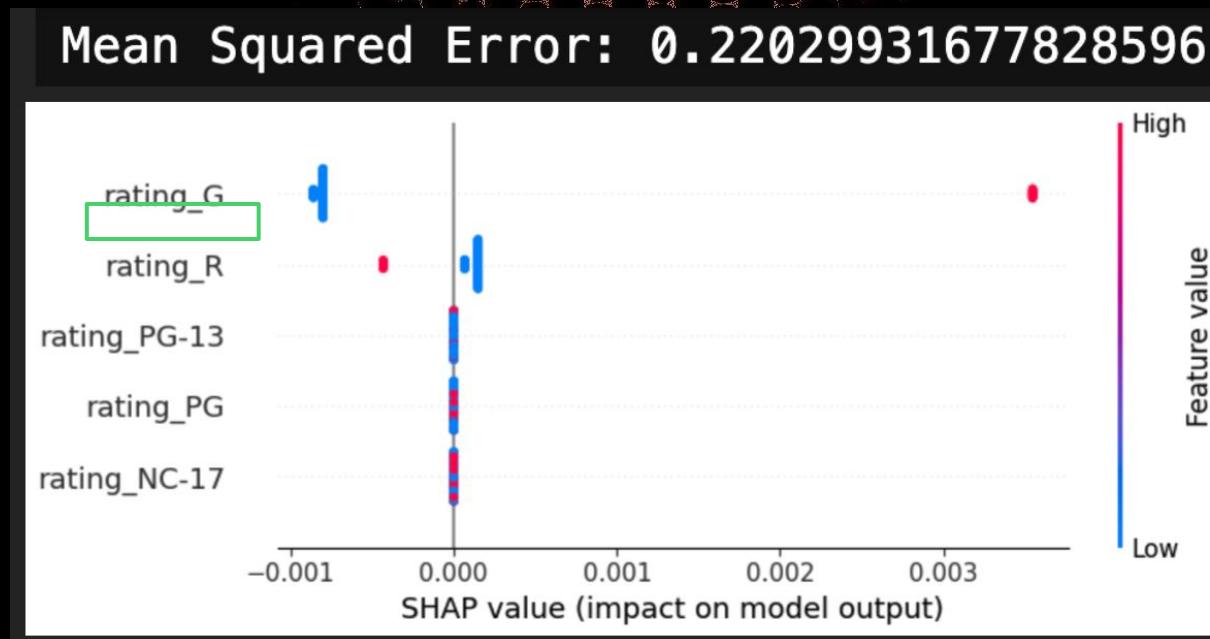
Mean Squared Error: 0.2196390428802757





Ratings | GridSearch

Here we focus only on rating and train for gross profit with grid search for optimal hyperparameters. The MSE is not very low again, but G ratings stand out as higher impact on positive ROI than other ratings.





Rental Rates:

High importance for predicting ROI

\$0.99

Low rate

\$2.99

Medium rate

\$4.99

High rate

Low rates are easily associated with lower ROI, as it takes more rentals to reach the same revenue as higher rates.

What do the rental rates look like in the context of ROI?

Do higher rates really result in more return on investment?

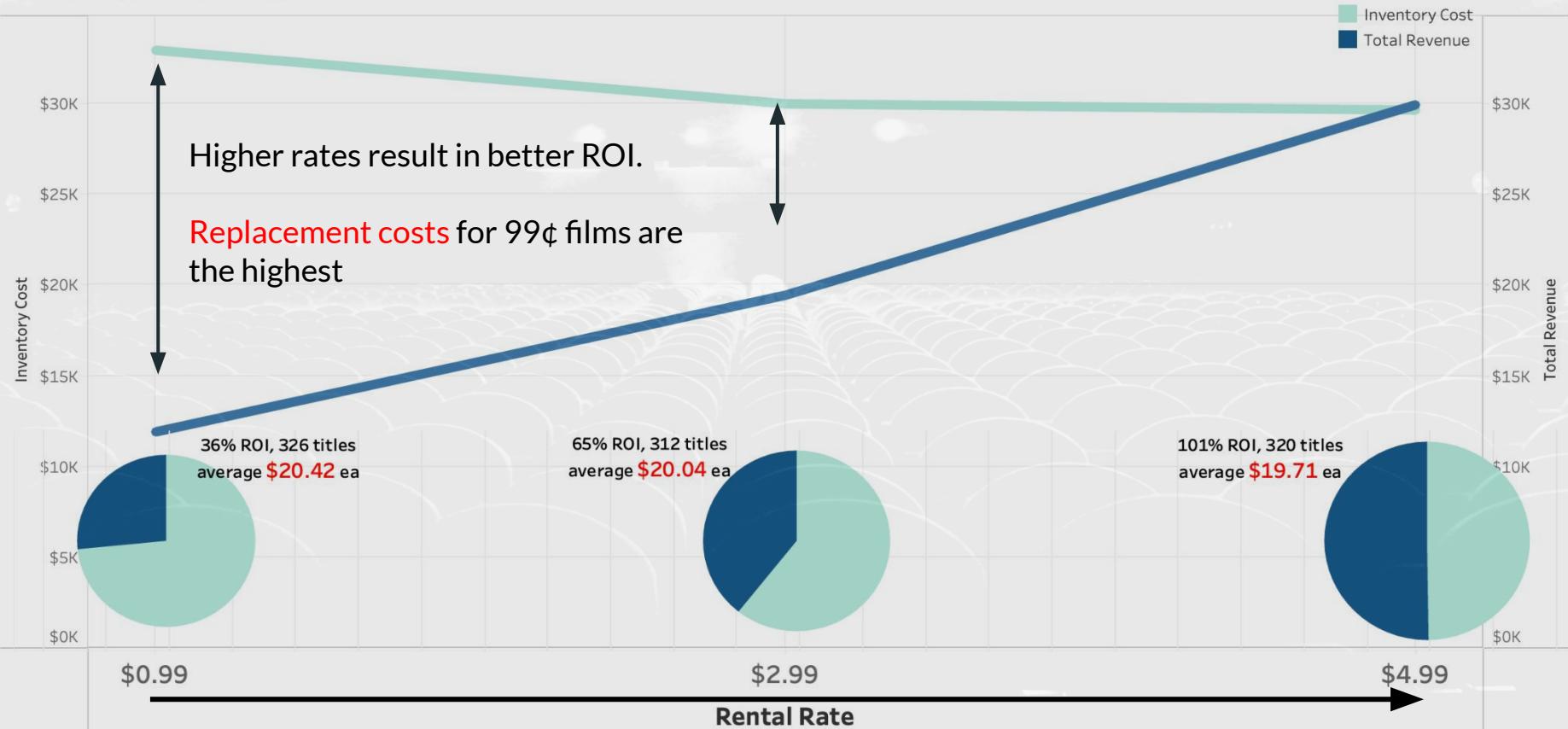


Key Optimizations

22

Rental Rates and ROI:

Films for 99¢ drive much less profit

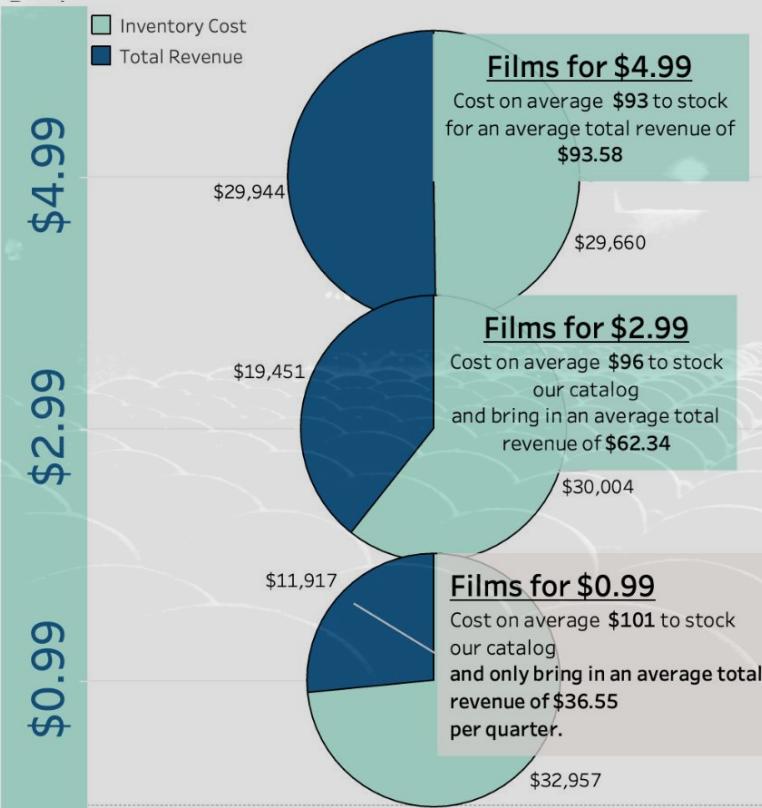




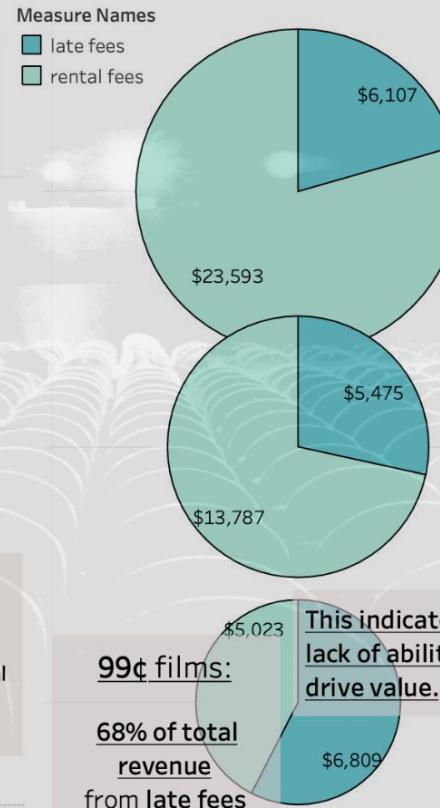
Key Optimizations

23

Rental Rates: Profit



Late Fees

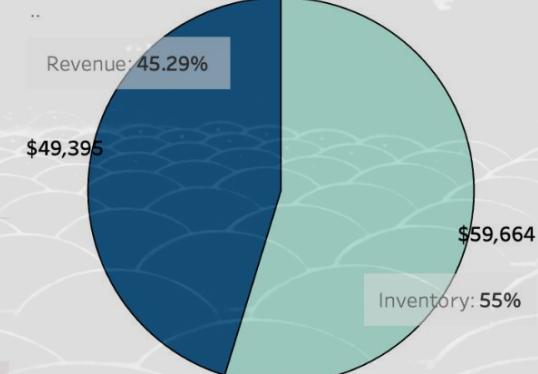


Removing \$0.99 Rentals

There are 326 titles rentable for \$0.99, removing them changes the catalog's profitability:

Total Revenue- $\$61,312 - \$49,395 = \$11,917$
less revenue

Inventory Cost- $\$92,621 - \$59,664 = \$32,957$
less inventory cost



However, cutting 33% of the catalog may be too high a threshold and can be improved with similar results.

Rate Groups

- (All)
- \$0.99
- \$2.99 & \$4.99

Profitability Quartiles

There are plenty of ways to prune the catalog, the strategy is to cut the bottom ~20% in terms of ROI each quarter by category.

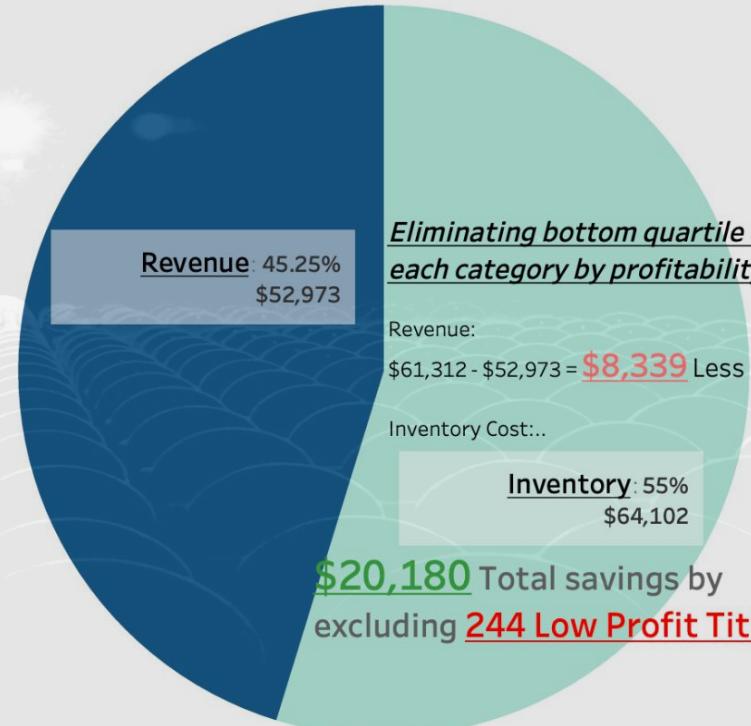
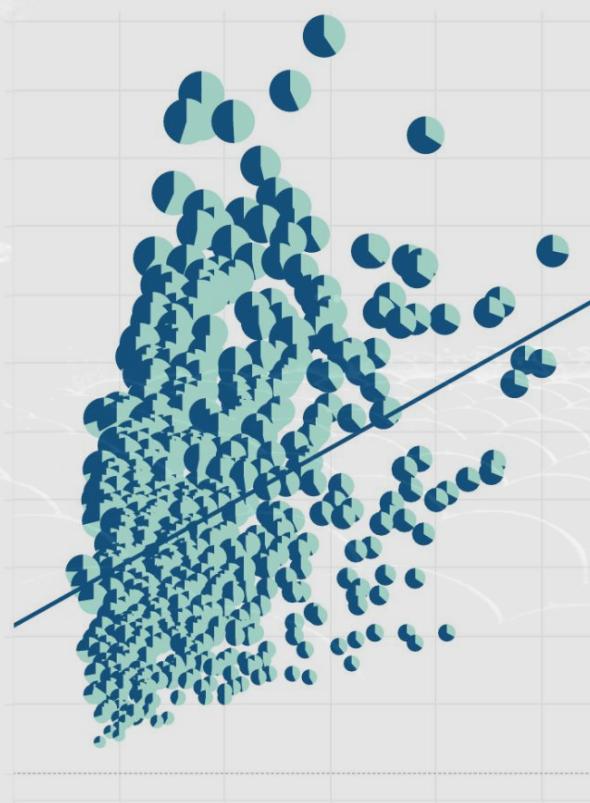
This will result in less revenue but much less cost, and will continuously update the catalog to be popular and profitable.

What happens to total ROI when we remove the Low Profit Quartile by Category?



Key Optimizations

25



- (All)
- Low Profit Titles (244) ↑
- Top Quartile + Outliers (247) ↓



40% ROI

Define Key Metrics

Goal: Improve 40% ROI

Optimize

Solve for Key Metric
(ROI)

Remove Low ROI
Titles

45% ROI

Save \$20,180

Collect data for next
optimization cycle



Rates

Avoid 99¢ titles

Maximize \$4.99 rates

Categories

Largest 10 Markets

Sports is the Highest
Grossing, but New
has the Highest ROI

Loyalty

Top customers in our
largest markets

Identify and Reward



Spend *less* to get more

Find the *weak links*

Strengthen and Optimize
the film catalog

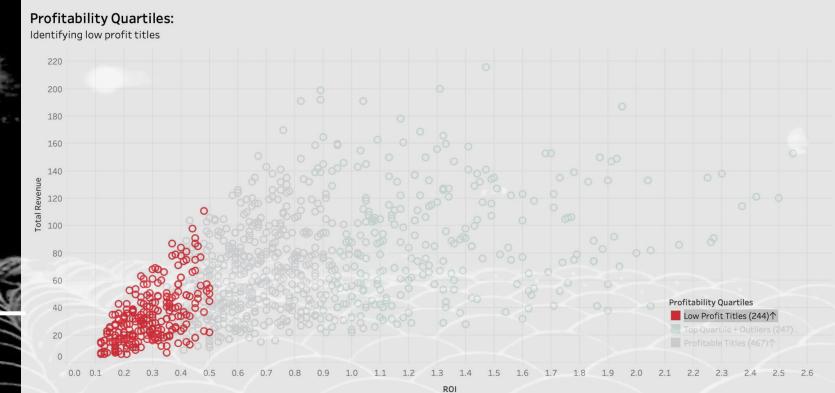


Tableau Link:

https://public.tableau.com/app/profile/aaron.manzano/viz/SQL_final/RockBusterRentals?publish=yes