Shifra Isaacs
Data Science Candidate

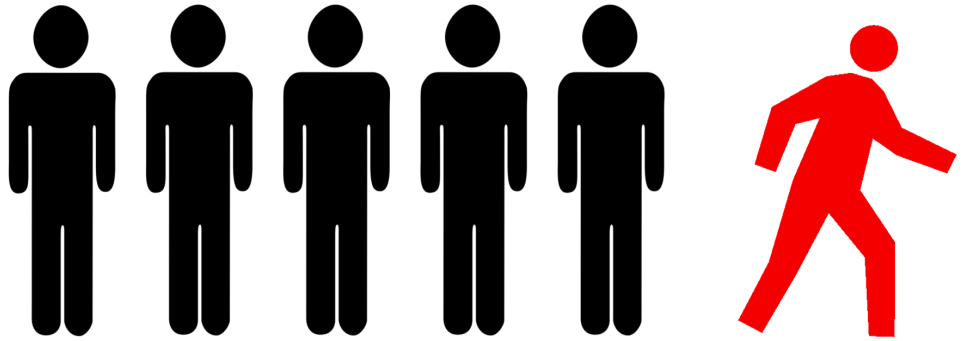
Case Study Presentation

November 14, 2022

Impact

Objective

- Used ML to identify schools where Marketing should spend resources to **prevent subscription churn**



Recommendation: Target Schools With…

- Less ability to pay for orders
- Poorly-funded book clubs and education groups
- Low enrollment and HHI
- High teacher turnover

SCHOLASTIC EDUCATION

Strategy

Exploratory Analysis

- Identify highest and lowest performing schools by subscription numbers
- Explore common features within these groups

Attrition Modeling

- Predict which schools' subscription rates decreased by **50%** or more from 2017 to 2019
- Identify buildings at risk of reducing subscriptions

SCHOLASTIC EDUCATION

Blockers

Data Questions

- Which arbitrary lines should we draw to define the schools we're looking for?
- How do we wrangle this data to address changes over time?

Data Challenges

- Big data
- Last day (11/14), realized I made a crucial mistake and needed to redefine and test my model
- Missing values in **Buildings** table
- Many schools ordered thousands of magazines with an overall revenue contribution of **$0** (Zero Orders)

SCHOLASTIC EDUCATION

Overcoming Blockers

Time-Efficient Solutions

- Considered **Zero Orders** to be special cases, i.e. donations or deals
- Left **Zero Orders** in YoY analysis but omitted them from the all-time analysis
- Filled missing values with column means, medians, and modes as needed
- Viewed **Buildings** columns as unchanging constants (i.e. enrollment, demographics, etc.)

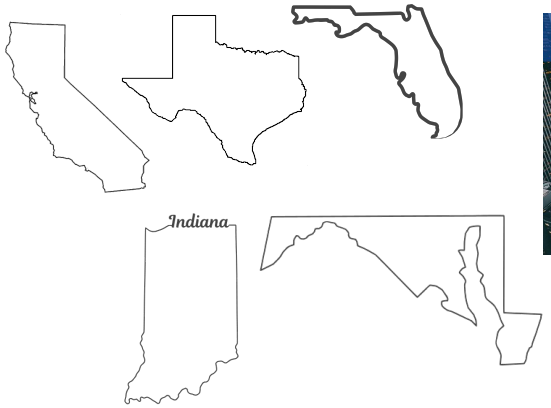
Snapshot of Aggregated Zero Orders

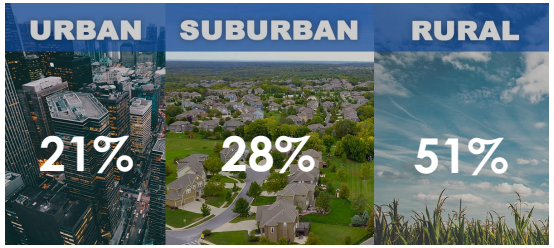
Building ID	Total Quantity	Total Revenue
600031079	100	$0
600031714	25	$0
600031934	1395	$0
600031940	1719	$0
600032227	885	$0

SCHOLASTIC EDUCATION

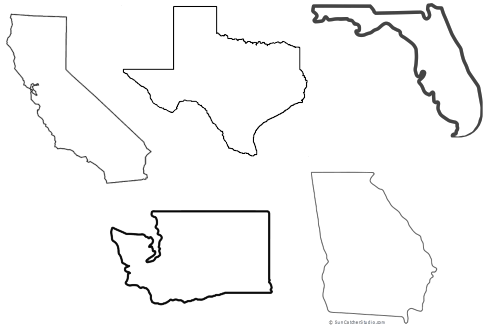
Qualitative + Quantitative Characterization

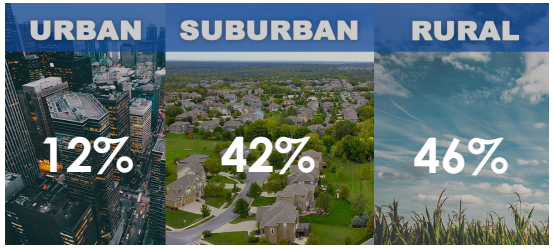
100 Highest Performers





100 Lowest Performers



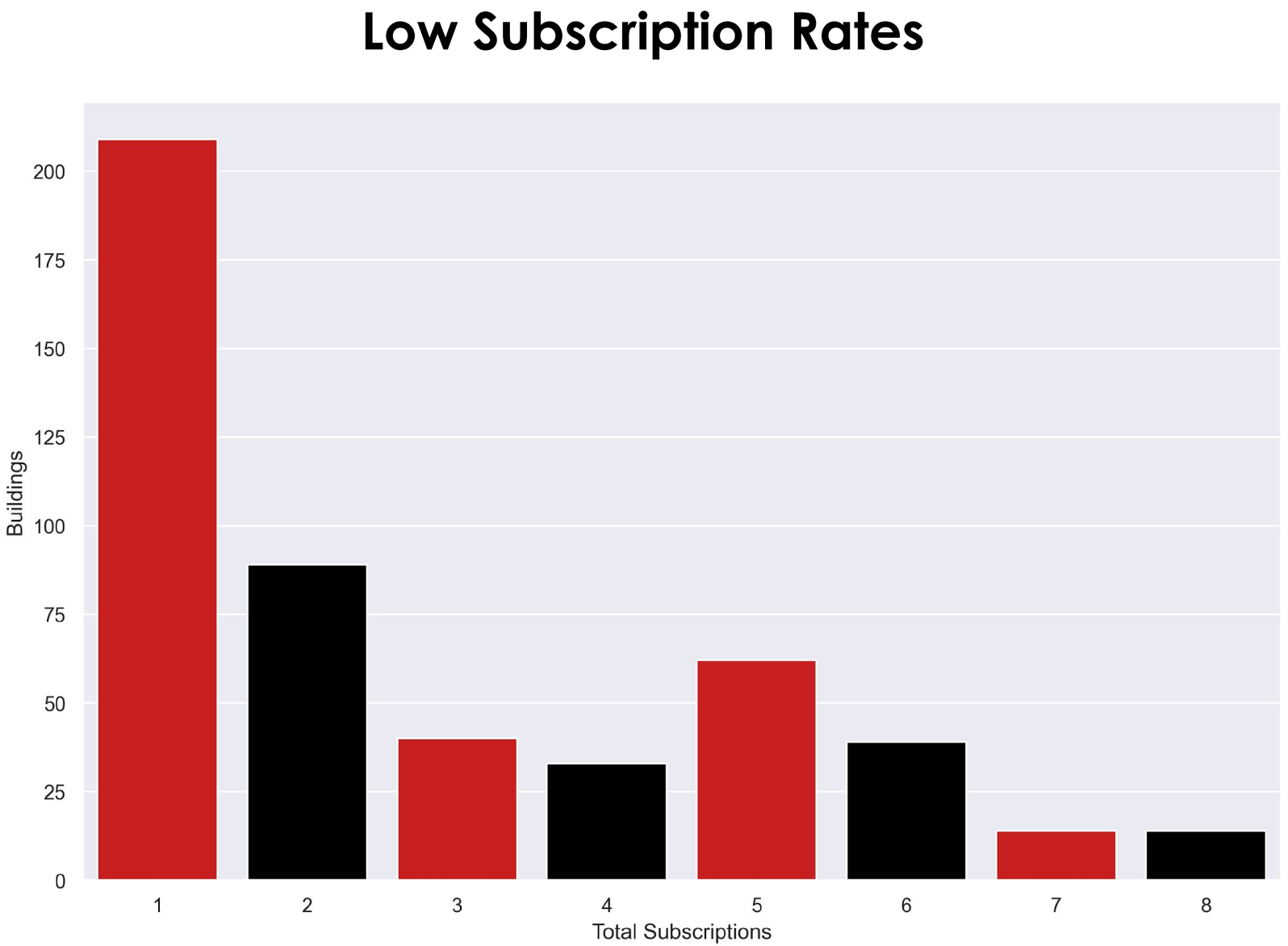


Group Averages	High Performers	Low Performers
Enrollment	481	429
HHI	$46K	$52K
T1 Eligible	66%	53%
6th-Grade Level	4.2%	4.7%
7th-Grade Level	2%	0%

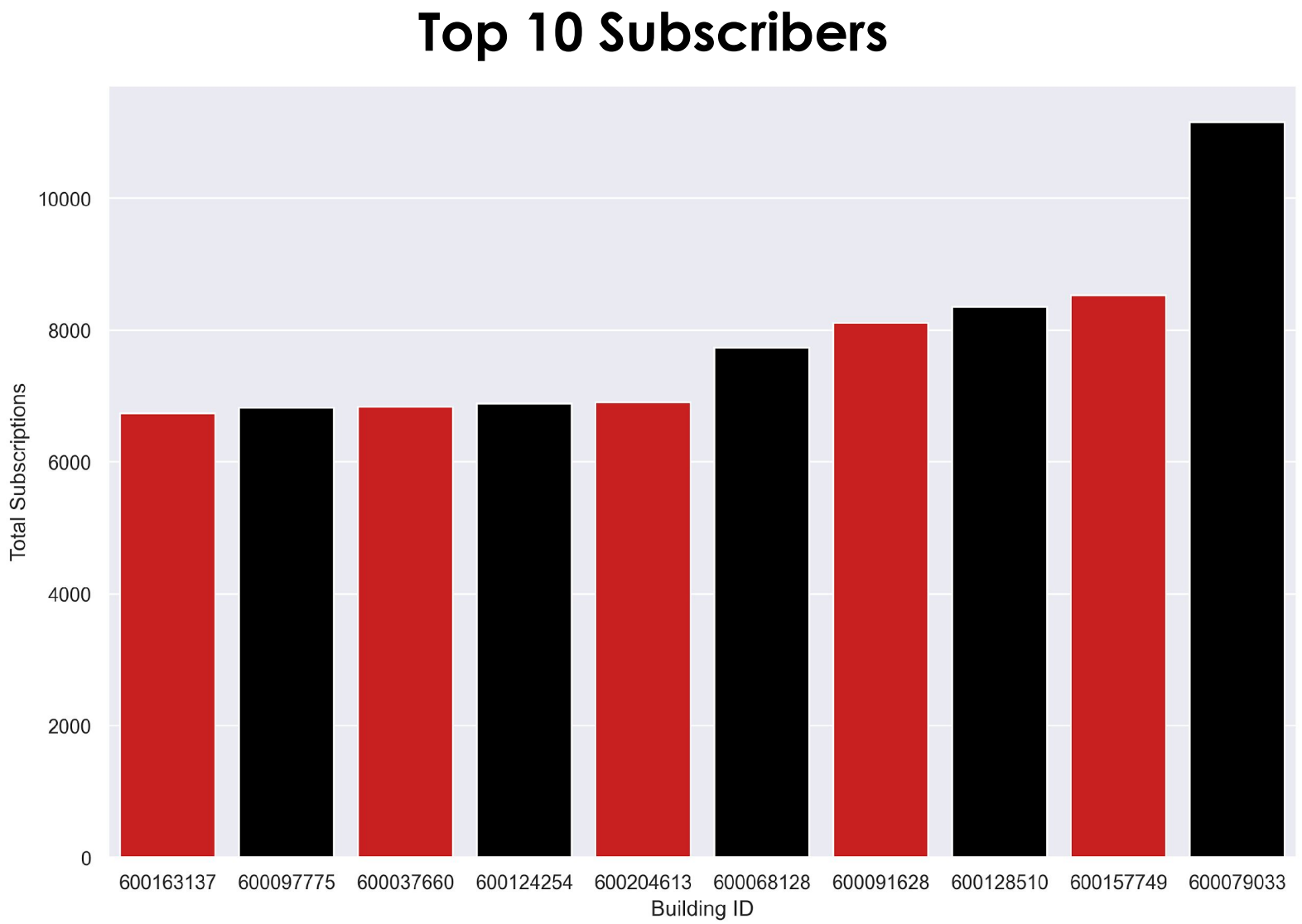
**All table values are means

SCHOLASTIC EDUCATION

Low Building Performance



High Building Performance



Feature Engineering

New/Transformed Features

- *Deltas (Δ)*: Teacher Count, Order Amount, Book Club Revenue, Education Group Revenue
- *Flags*: 7-8th Grade Reading Levels, Washington, Georgia, Indiana, Maryland, Rural, Suburban, Urban
- *Target*: **Paid Quantity Δ** (2 classes)

Procedure

- Data Scaling: Made features easy to interpret later (Min Max Scaler)
- Automatic backward selection:

SCHOLASTIC EDUCATION

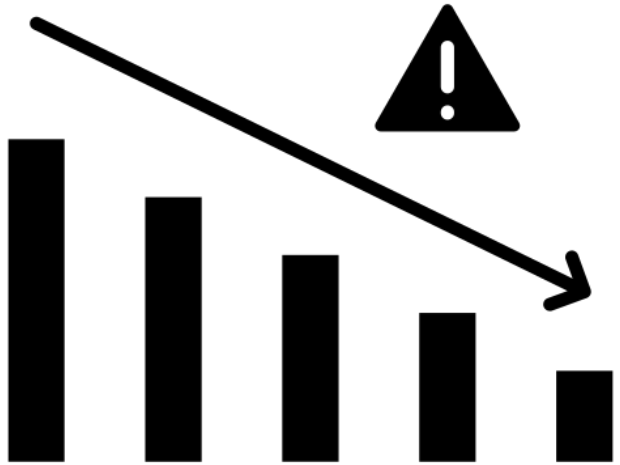
Modeling Strategy

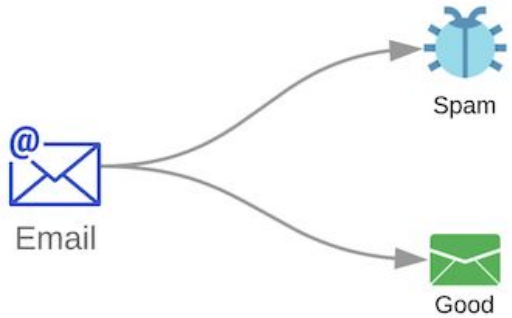
Goal

- Predict **Drastic Churn (DC):** buildings whose subscription rates fell by **50%** or more between 2017 and 2019
- Incorporate helpful features from all tables provided

Machine Learning Techniques

- 2 data classes: DC and Non-DC
- Binary classification modeling





Example: Binary Classification

SCHOLASTIC EDUCATION

Findings

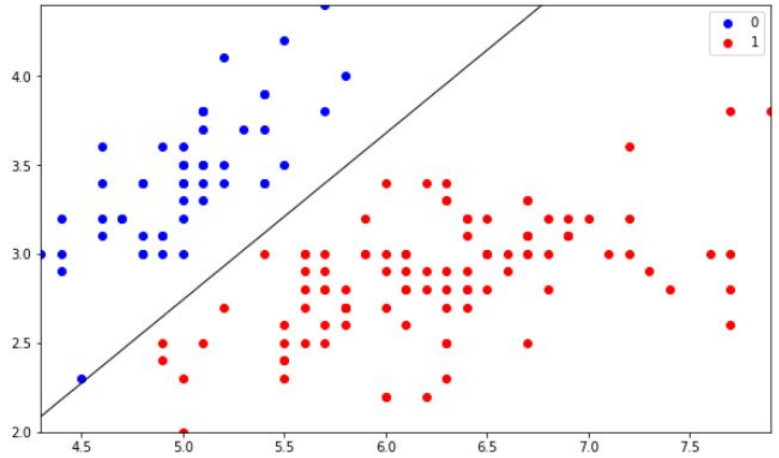
Model Results

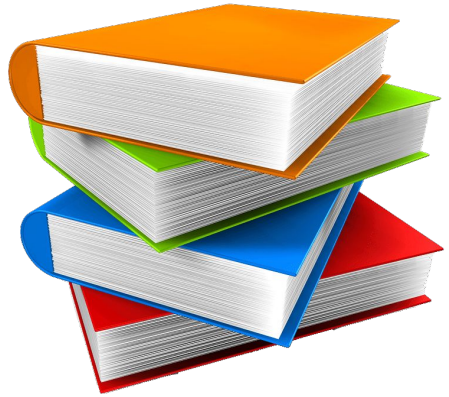
- Logistic regression predicts **DC** buildings with **87%** accuracy
- **Gaussian Naïve Bayes** yields best overall metrics (App. A)

Selected Model Features

- Book Club Revenue Δ
- Revenue Ed Group Δ
- Teacher Count Δ
- % Students at 3rd-Grade Level
- % Students at 4th-Grade Level

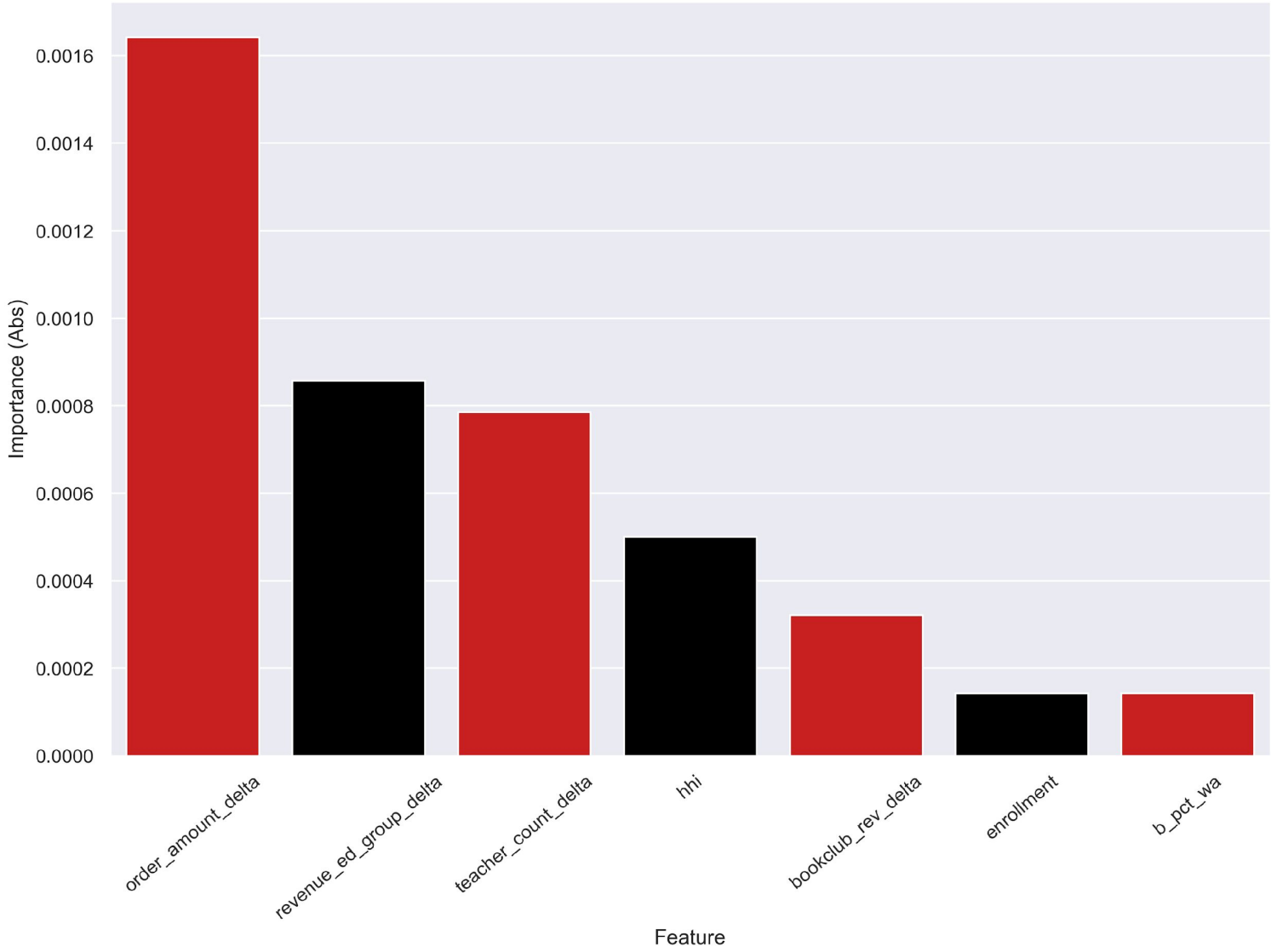
2D Visual Representation of Logistic Regression





SCHOLASTIC EDUCATION

Feature Importance (Gaussian NB)

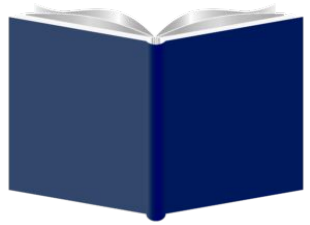


SCHOLASTIC EDUCATION

Recommendation

Target Schools

- **Book/Education/Order Revenue Δ □** Schools with poorly-funded book clubs and education groups
- **Teacher Count Δ □** Schools with high teacher turnover
- Schools with low **Enrollment**
- Schools with low **Household Income**



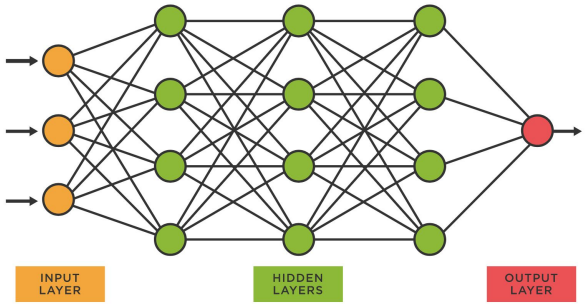
Social Action

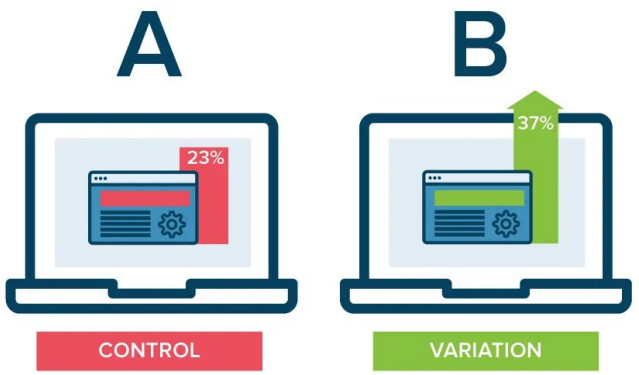
- Fundraisers and incentivized reading contests for target schools to

Next Steps

Further Analysis

- Tune models and try other powerful binary classifiers like XGBoost and Neural Networks
- A/B Test data-driven marketing efforts
- Evaluate data quality; Explore other methods to fill null values





References

- GitHub Repository
- Used knowledge and code from previous portfolio project

[redacted]

SCHOLASTIC EDUCATION

Impact Revisited

Conclusion

- Machine learning model identifies schools that will see **drastic subscription churn**

Business Value

- **Marketing:** Focus on **retention** of schools at risk of DC and **prevention** of subscription churn
- **Sales:** Use similar school spending patterns to inform pitches
- **Overall:** Increase operational efficiency by directing resources where they count most

SCHOLASTIC EDUCATION

Thank you!

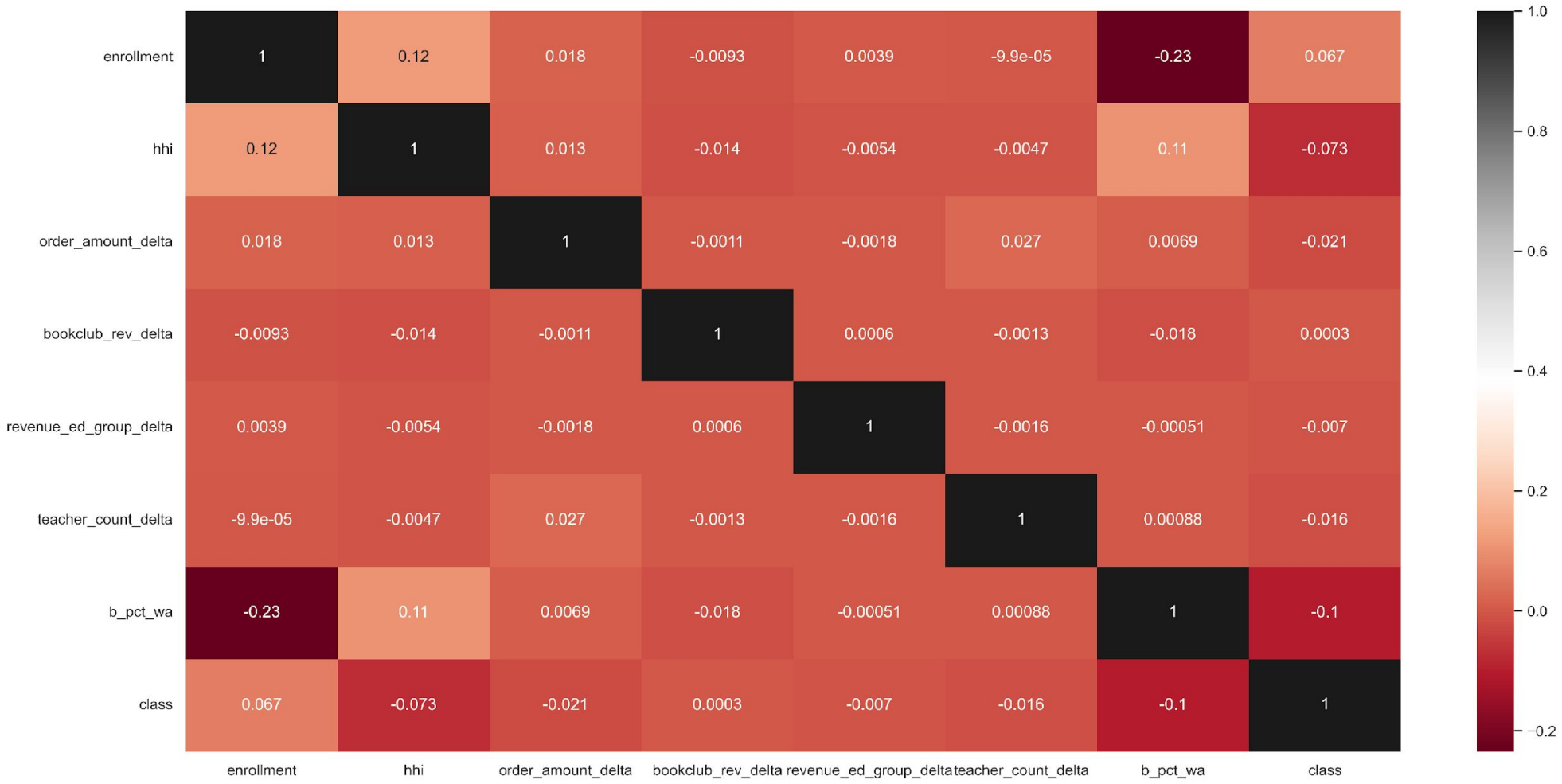
Sincerely,
Shifra Isaacs

Appendix A: Binary Classification Model Results

Model	Accuracy	Precision	Recall	F1 Score
Logistic Reg	87%	0%	0%	0%
SVC	87%	0%	0%	0%
Linear SVC	87%	0%	0%	0%
Gaussian NB	16%	98%	14%	23%
Random Forest	86%	2%	1%	36%
KNN	84%	2%	11%	4%

SCHOLASTIC EDUCATION

Appendix B: Correlation Matrix of Final Features



SCHOLASTIC EDUCATION