"I'm Just Browsing"
Predicting the Value of Prospective Customers

Estella Dentinger & Brittany Nguyen

Faculty Advisor: Professor Georgia Perakis

PhD Advisor: Leann Thayaparan

Macy's Team: Iris Singhania, Yixin Cai, Donghao Pei

macys inc

macy's bloomingdale's bluemercury

What is a Prospective Customer?

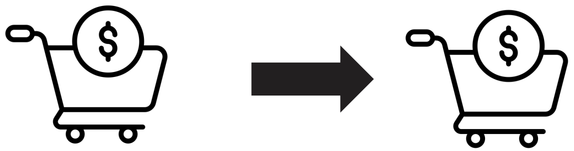
Online activity | Did not make a purchase | Prospective Customer

- **Why are they important?** Understanding the potential future value of customers who have engaged with Macy's, but have not made any purchases is critical to **new customer acquisition**
- **Our Project:** Understand who the valuable prospective customers are, and how to activate their first purchase and retain them

Problem Overview

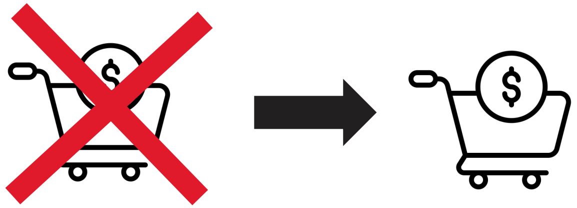
Current State

Macy's has models to predict the future value of **active customers** by using their historical purchase data



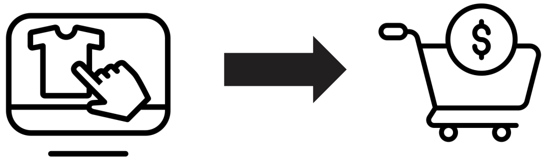
Limitation

Prospective customers, by definition, do not have purchase history



Our Approach

Use prospective customer **online activity data** to predict their value



Our Approach: Constructing the Dataset

Feb 2020 – Jan 2022

Prospective customers had online activity but no purchases

Feb 2022 – Jan 2023

Predict prospective customer **value** in 2022

Inactive Customers
Have purchase history prior to Feb 2020

New Customers
Have never made a purchase prior to 2022

Features

Click behaviors
Search, browse, add to cart, page view, abandon cart, and others

User Profile
Loyalty status, length of loyalty, new/inactive

Data Limitations

Imbalanced Dataset
Only 8% of prospective customers made a purchase in 2022 - spend is skewed

Skewed Distribution for Online Activity
Majority of values indicate little activity

Missing Values
Removed demographic and income features

Our Approach: Predictive Modeling

	1	2	3
3 Key Questions	Which prospective customers will make a purchase?	How much will prospective customers spend?	Who are the high value prospective customers?
3 Models	**Binary Classification** model to predict whether a customer will purchase in next fiscal year	**Regression** model to predict the dollar amount that a customer will spend in next fiscal year	**Multi-Classification** model to predict zero/low/high spend in next fiscal year

Model Validation through Backtesting

Backtesting: training on recent customer trends and testing on historical data

1 Train model on more recent data

Jul 2020 – Jun 2022 → Jul 2022 – Jun 2023

Prospective customer online activity → Predict prospective customer CLV

2 Test model on older data

Jul 2019 – Jun 2021 → Jul 2021 – Jun 2022

Prospective customer online activity → Predict prospective customer CLV

Our Models Improve upon Existing Methodologies

84% **Accuracy** of our Best Model

+7% Over Baseline

73% **Recall** of our Best Model

+19% Over Baseline

79% **AUC** of our Best Model

+12% Over Baseline

Out-of-Sample Backtesting Results

Model	Accuracy	Recall	AUC
GBM Binary	0.84	**0.73**	0.79
GBM Multiclass	0.79	**0.79**	0.75
Baseline (Active Customer Churn)	0.77	**0.54**	0.67

macys inc

MIT MANAGEMENT BUSINESS ANALYTICS

OPERATIONS RESEARCH CENTER

Our Best Model: Binary CatBoost GBM

Baseline: Existing Customer Churn Models

Top Drivers of Prospective Customer Value

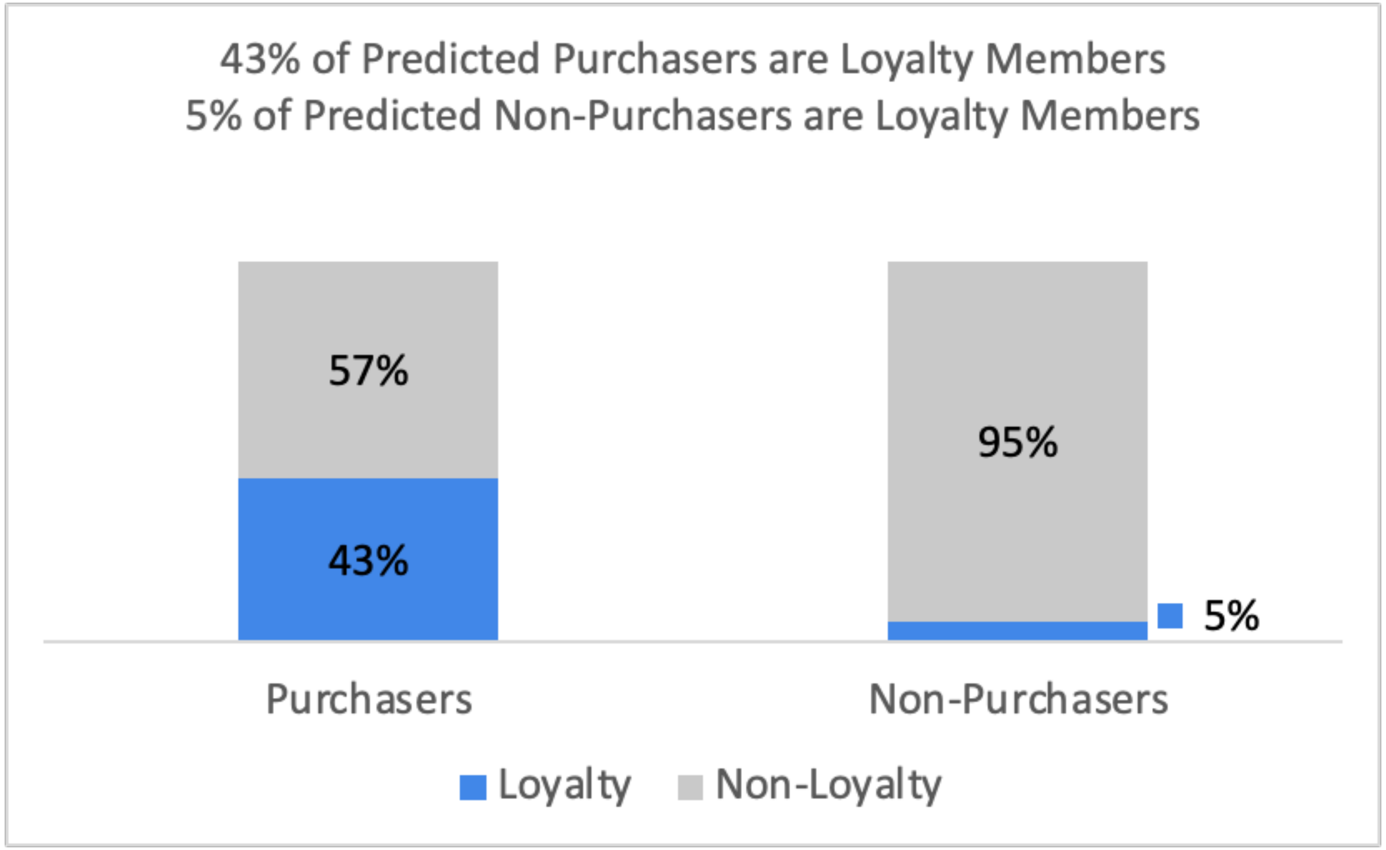
1	**New vs. Inactive Customer**	Whether a customer is new or inactive
2	**Account Creation**	Whether or not a customer signed up for a Macy's account as a loyalty member or non-loyalty member
3	**Email Opt-In**	Whether or not a customer opted-in to email marketing
4	**Count of SMS Sent**	Number of SMS messages delivered to customer
5	**Search**	Number of days with a search in the past 720 days

*Top Drivers of Binary GBM Model

Post-Modeling Analysis: Binary Model

Feature	Predicted Purchasers	Predicted Non-Purchasers
% Inactive Customers	91%	0.3%
% Acct Creation Loyal	6%	1%
% Email Opt-In	6%	0.5%
Count of SMS Sent	8	0.2
Avg Number of Days w/ Searches Past 720 Days	4	3.8

Post-Modeling Analysis: Binary Model



macys inc

MIT MANAGEMENT BUSINESS ANALYTICS

OPERATIONS RESEARCH CENTER

Business Impact

- Targeted email campaigns to valuable prospective customers
- Guide customer personalization, engagement, and retention efforts and act as a data resource for teams across Macy's

Next Steps

- Integration into active customer CLV workflow
- Predict CLV for future time frame 2023-2024
- Deployment of prospective customer CLV models

Acknowledgements

Thank you to:

- Iris Singhania, Yixin Cai, and Donghao Pei
- Professor Georgia Perakis and Leann Thayaparan
- MIT MBAn Program Team

macys inc

MIT MANAGEMENT
BUSINESS ANALYTICS

OPERATIONS
RESEARCH
CENTER

APPENDIX

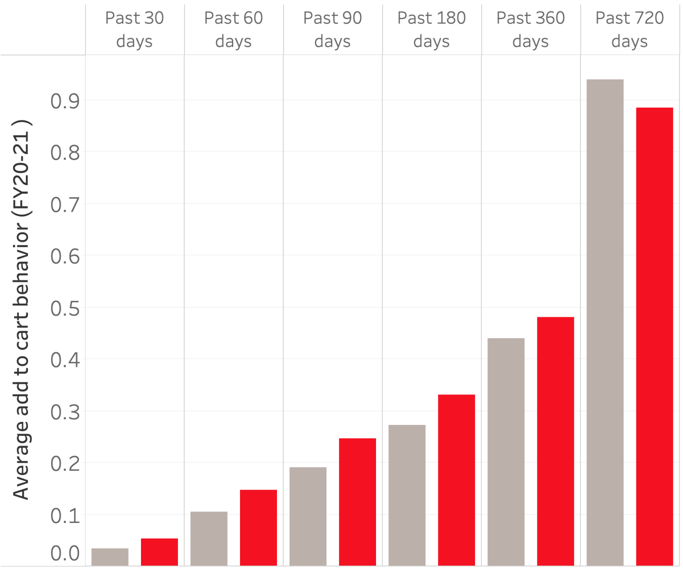
Predictive Features Selected

- Online activity metrics (search, browse product, page view, add to cart, abandon) *
- Loyalty tier & age of loyalty
- Email opt-in & SMS opt-in flag
- App download flag
- Prospective customer flag (1=never made a purchase)
- Count & Duration (seconds) of visits
- Device medium for visits (mobile phone, mobile app, tablet, desktop)
- Source sites (Google, Facebook, Bing, etc.)
- SMS data (sent, clicked, ordered, click rate, days since sms sent, days since sms clicked, days since sms ordered) *

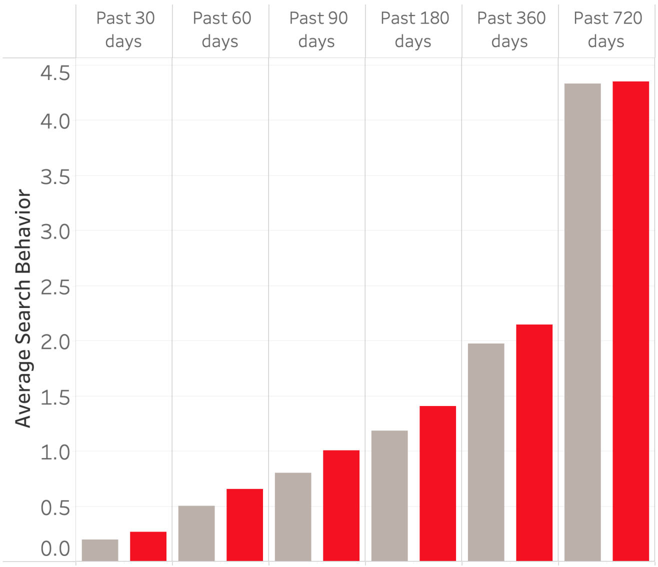
- ***note: time frame: across 30, 60, 90, 180, 360, 720 days, 2 years**

Online Activity Average Counts are Similar Across Purchasers and Non-Purchasers

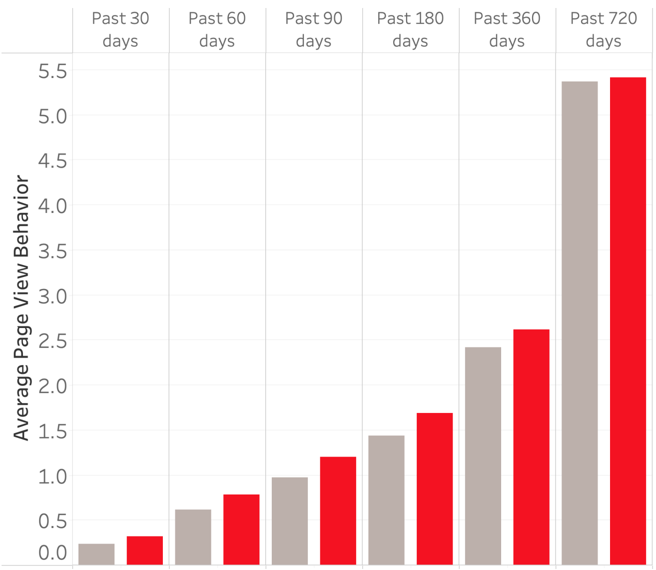
Add to Cart



Search



Page View



FY22 Purchase Flag

No Purchase

Purchase

*Outliers filtered out for all

macys inc

MIT MANAGEMENT BUSINESS ANALYTICS

OPERATIONS RESEARCH CENTER

App Download, Email Marketability, Session Length

Feature	Description
App Download Flag	• Binary • Whether user downloaded Macy's app
Email Marketability Flag	• Binary • Whether user was email marketable at time of downloading Macy's app
Session Length	• Total duration spent in seconds over last 7 days

SMS Features

Feature	Description
# SMS Sent	• # of SMS messages sent to user
# SMS Clicked (Total)	• # of SMS messages clicked by user
# SMS Clicked (Unique)	• # of unique SMS messages clicked by user
Click Rate	• # SMS messages clicked / # sent
Days Since SMS Sent	• # of days since SMS was sent
Days Since SMS Clicked	• # of days since SMS was clicked

*Measured for all features over 30, 60, 90, 360, 720 days

Post-Modeling Analysis: Multi-Class Model

Feature	Predicted Zero Tier (spend = $0)	Predicted Low Tier* (spend <= $119)	Predicted High Tier (spend > $119)
% Acct Creation Loyal	0%	10%	11%
% New Customers	99.9%	15%	19%
Avg Number of Days w/ Searches Past 360 Days	0.3	0.4	10
Avg Number of Days w/ Abandons Past 720 Days	1	0.5	3.5
Avg Number of Days w/ Page Views in Past 360 Days	0.4	0.5	12
% Email Opt-In	0%	5%	18%
Count of SMS Sent	0.4	6.6	6.8

*$119 cutoff determined by median spend value

Post-Modeling Analysis: Multi-Class Model

Feature	Predicted Zero Tier (spend = $0)	Predicted Low Tier* (spend <= $119)	Predicted High Tier (spend > $119)
% Acct Creation Loyal	0%	10%	11%
% Inactive Customers	0.1%	85%	81%
Avg Number of Days w/ Searches Past 360 Days	0.3	0.4	10
% Email Opt-In	0%	5%	18%
Count of SMS Sent	0.4	6.6	6.8

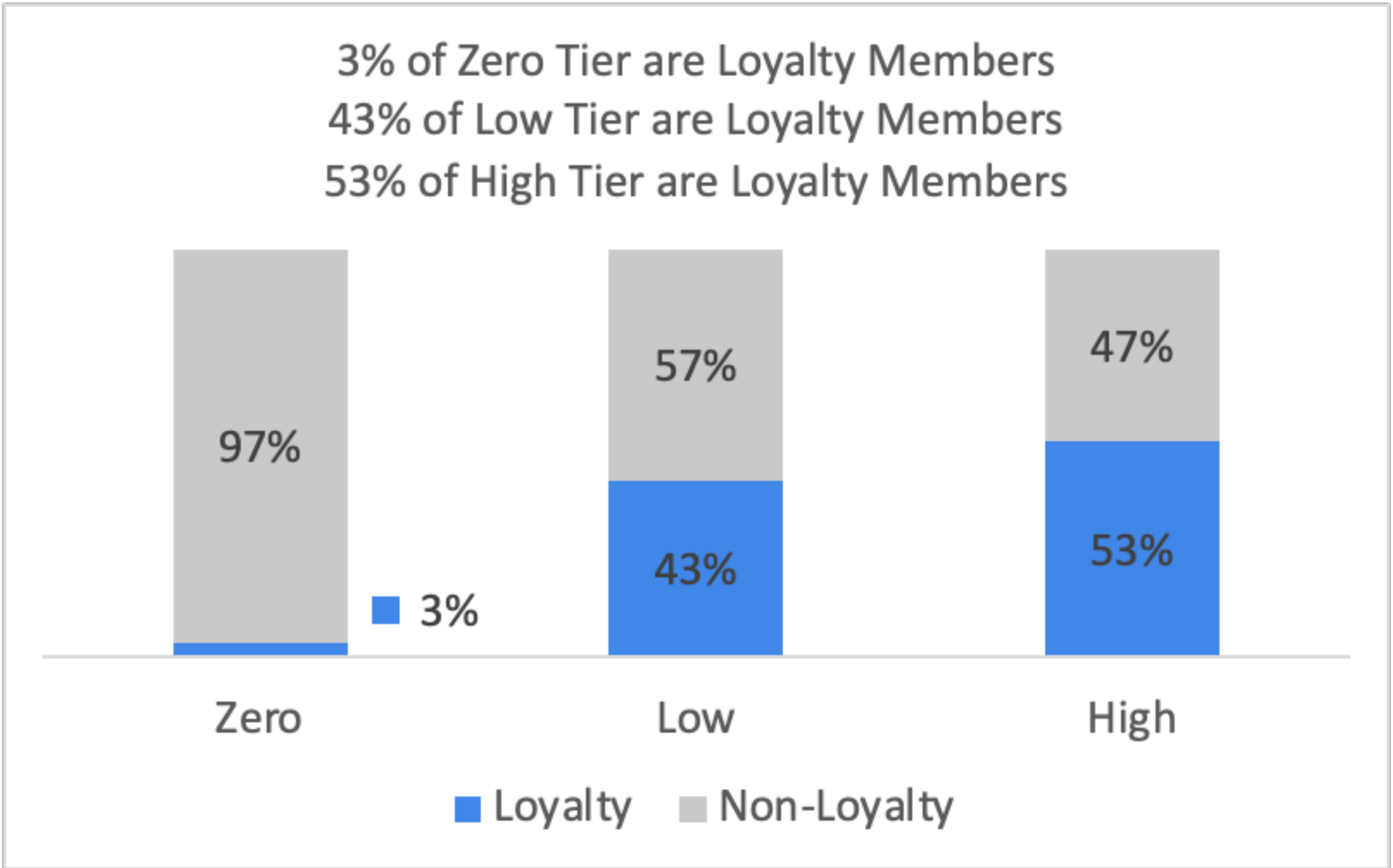
macys inc

MIT MANAGEMENT BUSINESS ANALYTICS

OPERATIONS RESEARCH CENTER

*$119 cutoff determined by median spend value

Post-Modeling Analysis: Multi-Class Model



macys inc

MIT MANAGEMENT BUSINESS ANALYTICS

OPERATIONS RESEARCH CENTER