Leveraging Modeling & Analytics to Derive Insights for Loyalty Campaigns

Yifan Gu, Sai Nithin Godi, Nikitha Balaji, Muskan Aggarwal, Harshal Amin, Dhairya Dedhia, Matthew A. Lanham
Purdue University, Krannert School of Management
gu277@purdue.edu; sgodi@purdue.edu; balaji76@purdue.edu; aggar158@purdue.edu; amin82@purdue.edu; ddedhia@purdue.edu; lanhamm@purdue.edu

BACKGROUND

Loyalty programs drive customer retention and revenue growth by enhancing engagement and influencing spending patterns. **Loyalty offers**, a key feature of these programs, provide incentives for specific actions in exchange for rewards. This study focuses on optimizing **a well-known grocery retailer's loyalty program** by analyzing **overall** and **business segment level Loyalty offer effectiveness**, and **digital engagement metrics effectiveness**. The goal is using **data analytics** and **predictive modeling** to generate insights that optimize offer strategies for greater customer engagement and sales.

BUSINESS PROBLEM FRAMING

Business Problem

This project aims to:
- **Assess Overall Loyalty Offer Effectiveness** – How do different offer types influence customer behavior?
- **Analyze Business Segment Level Loyalty Offer Impact** – Does loyalty offers received in the campaign period and specific business segments lift sales?
- **Analyze Digital Engagement Metrics Impact** – Does digital engagement metrics contribute meaningfully to predicting sales in the campaign period?

Importance & Business Benefits

Business Importance
- Loyalty programs drive retention & revenue growth

Business Benefits
- Improved digital engagement & store visits
- Increased customer spending
- Higher loyalty & retention
- Smarter strategy through behavior forecasting

Stakeholders, Context & Constraints

This project serves **Marketing, Finance, Customer Experience, UI/UX, and Data Science teams** by providing insights to refine loyalty strategies. It leverages **sales, engagement, and behavioral data** to optimize Earn offers. Additionally, the study considers **customer segmentation** to balance personalization with scalability.
However, a key challenge is **data imbalance**—the pre-campaign period spans one year, while during-campaign period data is limited to one month, potentially affecting model accuracy and generalizability.

ANALYTICS PROBLEM FRAMING

EDA reveals distinct sales differences between different relevant factors, prompting a **causal-inference modeling** approach.

Assumptions
- Pre-period sales data reflects baseline customer purchasing behavior
- Treatment effects are linear and additive
- Behavioral trends are stable for prediction

Model Building Strategy

- **Multiple Linear Regression**
  - control for past behavior, minimizing confounding effects
  - quantify effect sizes and statistical significance, helping identify the most impactful drivers of customer response

- **Dependent Variable**: DuringSales (sales during campaign period)
- **Independent Variables:**
  - **PreSales** (1-year sales data before campaign period)
  - Model 1: **Offer Types** (PointsPerTrip, PointsMultiplier, NormalTargeting)
  - Model 2: **Digital Activity** (WebVisitDays, ClaimDays, EarnPageViewDays)
  - Model 3: **DuringOfferFlag** (indicate if offer is given in campaign period)

Success Metrics ★★★

- **R^2** (explanatory power)
- **Business Interpretability** (actionability): **Coefficient, p-value**
  - Coefficients present predictor's effect on DuringSales, controlling for PreSales
  - effects with $p < 0.05$ are considered statistically significant

DATA FOUNDATIONS

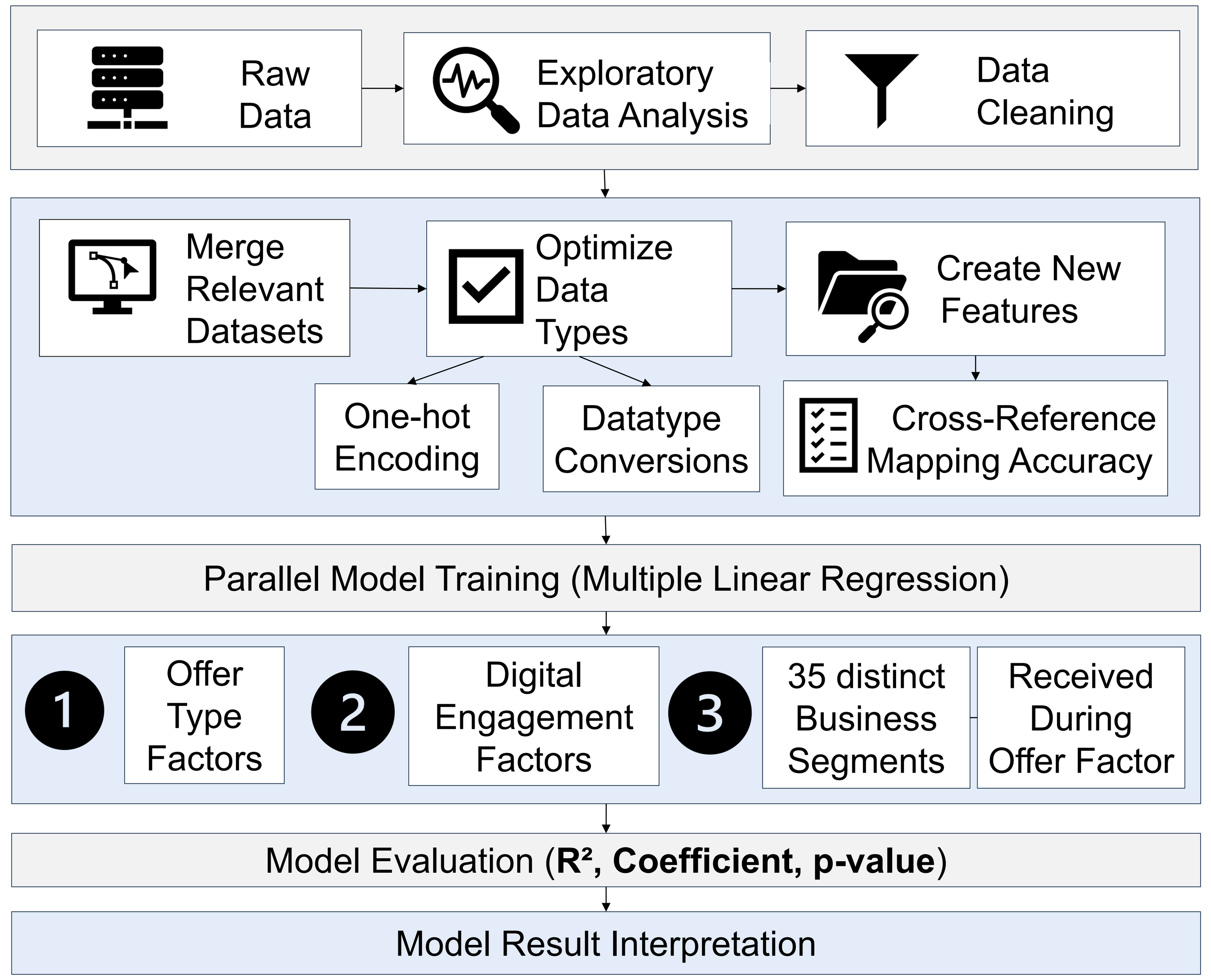
Data Sources & Processing

- **Merged datasets** include customer segmentation, sales (DuringSales, PreSales), digital engagement metrics (WebVisitDays, ClaimDays, EarnPageViewDays), and offer information from multiple data sources
- **Inner joins on customer ID** ensure data integrity
- **4,259 missing sales entries** are filled with zeros for completeness
- Applied **one-hot encoding** and created new features for modeling

Alignment with Business & Analytics Goals
- **Merged data reflects business goal**: links relevant factors to customer purchasing behavior for targeted marketing and informed decisions
- **Data aligns with analytics goal:** data merging and preprocessing ensured clean, complete inputs with relevant dependent and independent variables for modeling

MOTHODOLOGY

✓ Used Python for data preprocessing and modeling



MODEL BUILDING & KEY INSIGHTS

Descriptive Analytics: Digital Engagement Metrics

- EarnPageViewDays: median = 0 → most customers rarely use this
- ClaimDays: mean = 1.4, SD = 1.5 → less frequent, but impactful
- WebVisitDays: mean = 12, SD = 7.9 → active channel

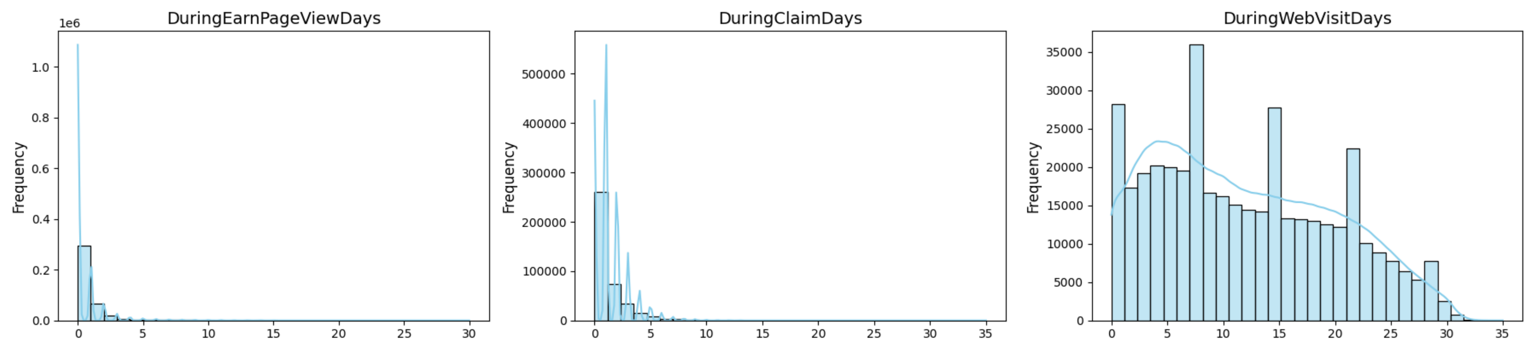


Fig 1. Distributions for Digital Engagement Metrics

Model Results & Interpretation

Model 1 (Offer Type Factors)
- **PointsMultiplier** (coef = -2.45, p = 0.01) and **NormalTargeting** (coef = 1.57, p = 0.13) are both less effective than **PointsPerTrip**; only **PointsMultiplier** shows a statistically significant impact.

Model 2 (Digital Engagement Factors)

- Digital engagement metrics significantly predict DuringSales, with **DuringClaimDays** (coef = 61.14, p < 0.001) showing the strongest effect.

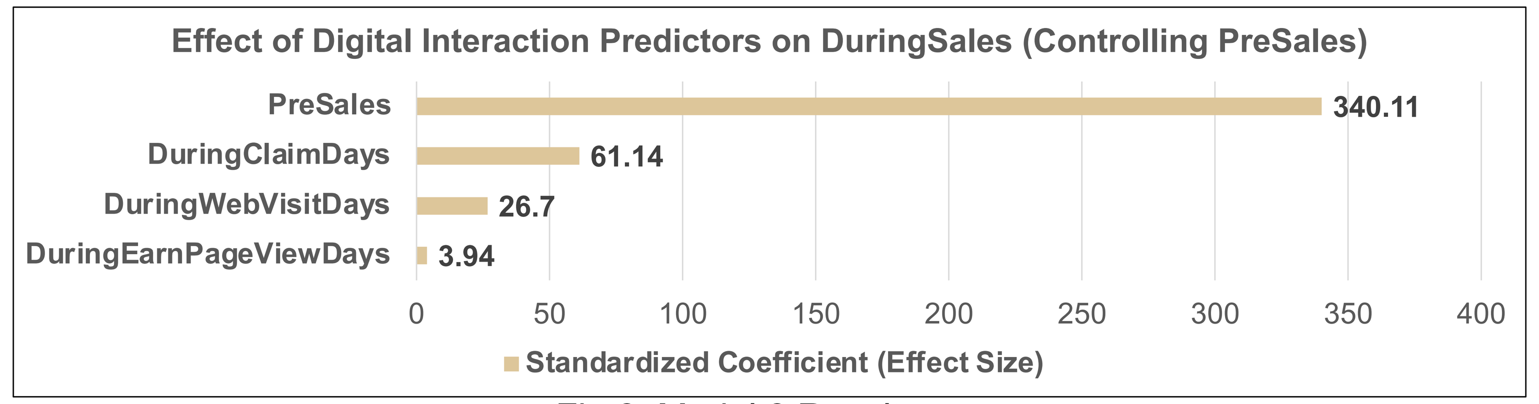


Fig 2. Model 2 Result

Model 3 (Business Segments Level - During Offer Factor)
- Among 35 business segments, **PETS** and **BABY** were the only ones where during-campaign offers led to a significant increase in customer sales, with **PETS** showing the largest effect (coef = 15.42, p < 0.001).

- **Model Robustness:** Reusable with updated inputs and consistent preprocessing steps; reliable under linear regression assumptions.
- **Limitations:** Model prioritizes interpretability over prediction; may need other machine learning models if input patterns shift or accuracy is key

BUSINESS VALIDATION & IMPACT SUMMARY

Business Validation

The models aligned strongly with business goals by identifying effective offer types, high-impact segments, and key digital engagement drivers. Data-driven decisions could be made for optimizing current campaign.

Model Impact
- **Offer Effectiveness:** promote **PointsPerTrip** offers (Model 1)
- **Customer Engagement:** direct traffic to **Claim** landing page (Model 2)
- **Campaign Optimization:** design and introduce more offers in **PETS** and **BABY** business segments (Model 3)

Stakeholder Feedback

Stakeholders described results as "impressive", especially findings on digital engagement support efforts to drive targeted traffic.

Future Scope
- Explore advanced machine learning models for better prediction power
- Align model findings with long-term marketing strategies

ACKNOWLEDGEMENTS

We would like to thank Professor Matthew Lanham and our industry partner for their guidance and support on this project.