

Triggering the Buy: What Drives Spending, and How Can Pricing Win?

a data-driven approach

BACKGROUND

Industry: Online Education (EdTech)

Client: ExtraLearn, an emerging EdTech startup.

Opportunity: The post-COVID-19 era has accelerated adoption of online learning, drawing a surge of new customers and creating a booming market with increasing digital engagement.

Business Problem: With rapid industry growth comes intense competition. ExtraLearn must effectively identify, segment, and prioritize high-quality leads to optimize conversion rates.

Goal: Improve lead conversion by targeting high-potential customers, optimizing marketing and sales resources. Customer segmentation insights can also guide personalized pricing, such as tiered plans or discounts based on engagement and interest.



OBJECTIVES

- To develop a machine learning model to predict leads with a higher likelihood of converting to paid customers, using demographic, platform interactions, and marketing engagement data
- To conduct cluster analysis to segment customers into meaningful groups based on their interaction patterns with the platform

ABOUT THE DATA

Demographics

Spending Platform Interaction

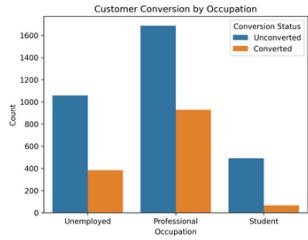
Marketing

Conversion Outcome



EXPLORATORY DATA ANALYSIS

1. Conversion Rate by Occupation



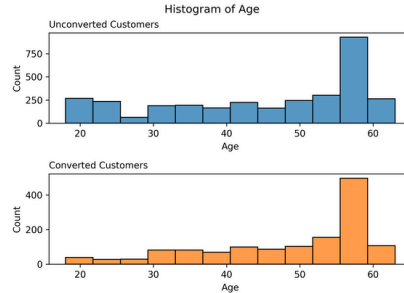
Observations:

- Professionals convert most; students least - shows intent vs. price sensitivity
- Unemployed moderate - cautious spending

Strategies:

- Students:** Introduce entry-level pricing, learning bundles, or gamified incentives to boost perceived value
- Professionals:** Upsell premium, career-focused plans; emphasize ROI
- Unemployed:** Offer trial access or flexible payment (lower entry barriers)

2. Age Distribution by Conversion Status



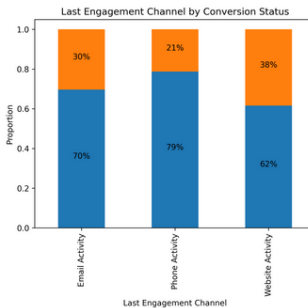
Observations:

- Higher conversion among old age groups
- Low conversion in young segments - less engaged to current offerings

Strategies:

- Consider premium packages or value-driven bundles tailored for older customers who may be more willing to spend for quality, trust, or simplicity
- Reevaluate how current products are perceived by the <30 age group

3. Last Engagement Channel by Conversion Outcome



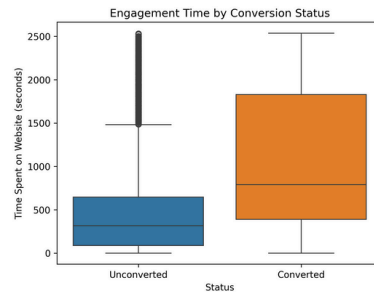
Observations:

- Website and email activities drive most conversions - clear information, easy access
- Phone has the lowest conversion - likely seen as intrusive or outdated

Strategies:

- Website & Email:** Optimize pricing visibility, clear call-to-actions, and mobile UX
- Phone:** Limit to high-intent leads or support, avoid cold calls
- Channel Strategy:** Shift focus to website or email, and the use of chatbots to boost engagement

4. Engagement Time by Conversion Outcome



Observations:

- Converters spend 3-4x more time on website
- Spending platform engagement correlates with conversion decision
- Some non-converters show high engagement - likely outliers

Strategies:

- Use time-based prompts (e.g., nudges, limited-time offers) to convert active users
- Improve UX for long sessions (save, resume), progress bars, and content personalization
- Use urgency messaging to highlight limited-time deals to prompt action

CUSTOMER SEGMENTATION ANALYSIS WITH K-MEANS CLUSTERING

Cluster 1: Engaged Non-Converters

- High visit frequency with moderate time spent
- Low conversion rate (12%)
- Likely browsing regularly without intent to act

Cluster 2: Quick Converters

- Low time spent and few visits
- High conversion rate (38%)
- Decisive users with clear goals

Cluster 3: Deep Researchers Who Convert

- Very high time spent with moderate visits
- Highest conversion rate (40%)
- Highly engaged and thorough decision-makers

Cluster 4: Casual Browsers

- High page views per visit with average time spent
- Low conversion rate (10%)
- Curious users who explore but rarely take action

Cluster	Average Time Spent on Website	Average Website Visits	Average Page Views per Visit	Percentage of Converted Customers
1	482.49	8.52	2.66	12%
2	282.47	2.36	2.31	38%
3	1861.36	3.3	2.75	40%
4	557.75	3.27	6.6	10%



PREDICTIVE MODEL DEVELOPMENT AND SELECTION

Logistic Regression

- Used for binary classification that predicts the probability of an outcome based on input features
- Hyperparameters:** Penalty, Solver, Regularization Strength, Maximum Iterations

Decision Tree

- A tree-like model that splits data into subsets based on feature values to make predictions
- Hyperparameters:** Maximum Depth, Split Criterion, Maximum Features

Gradient Boosting

- A sequential method where each tree corrects the previous one's errors using gradient descent
- Hyperparameters:** Maximum Depth, Learning Rate, Number of Estimators

XGBoost

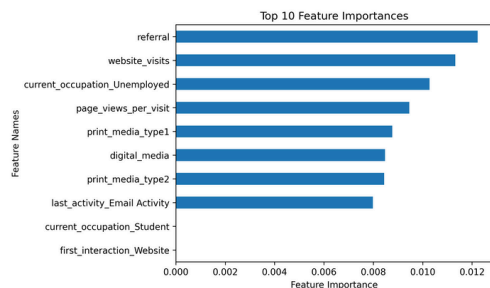
- A more efficient gradient boosting algorithm that uses second-order gradients for faster and more accurate predictions
- Hyperparameters:** Maximum Depth, Learning Rate, Subsample, Number of Estimators

Random Forest

- An ensemble method that builds multiple decision trees and combines their outputs to reduce overfitting
- Hyperparameters:** Maximum Depth, Number of Estimators, Minimum Samples Split and Leaves, Maximum Features

PREDICTIVE MODEL EVALUATION

Champion Model: XGBoost



Key Aspect	Finding	Strategy
Referrals	Leads coming through referrals are significantly more likely to convert	Boost referral programs with incentives (e.g., discounts, free course trials)
Spending Platform (Website)	High website visits and page views per visit correlate with conversion	Target highly engaged visitors with time-sensitive offers or retargeting ads and consider implementing behavioral triggers (e.g., pop-ups with discounts after 3+ visits)
Occupation	The unemployed have much greater interest in the product offerings	Craft messaging that emphasizes career advancement, certifications, and job-readiness
Marketing Approach	Offline and online media both influence conversions	Continue omnichannel marketing
Engagement Channel	Email engagement prior to conversion is meaningful	Incorporate personalized pricing offers into email campaigns

Deployment Plan:

- Model deployed for real-time lead scoring
- Integrate into CRM pipeline for automated lead prioritization
- Launch targeted marketing based on predictive scores
- Implement feedback loop for continuous model improvement

Accuracy: 88.30% of the predictions were correct overall

Recall: 74.82% of actual converters were identified

Precision: 84.02% accurate when it predicted a customer would convert



PRICING STRATEGIES

Tiered Pricing

- Use occupation and engagement data to guide a tiered pricing model
 - Students/Unemployed:** Offer discounts, freemium access, or pay-as-you-learn plans
 - Engaged Users:** Those with high visits/page views may be nudged into upselling bundles or premium subscriptions

Behavior-Based Discounts

- Trigger discount offers based on observed behaviors like repeated visits or deep content engagement

Referral-Based Rewards

- Give price reductions or credits to users who refer others - both referrer and referee benefit