Predicting Customer Purchases

Turning Website Visits into Revenue

Exploring how understanding user behavior can unlock significant revenue growth and enhance customer experience.

Oge Ohia

Click here to interactively view this presentation

The Big Question

Can we predict, in real-time, which website visit sessions are most likely to end in a customer buying something?

Why This Matters

Boost Sales

Identify high-potential customers and proactively guide them toward conversion.

Smarter Marketing

Target promotions and messages precisely, reducing wasted effort and improving ROI.

Enhanced Experience

Personalize the website journey, making it more intuitive and effective for motivated buyers.

The Data Behind Our Predictions

We meticulously analyzed data from 12,330 past visitor sessions on our online store. Each session provided a rich tapestry of behavioral and contextual details, forming the foundation of our predictive model.

Key Data Points Captured:

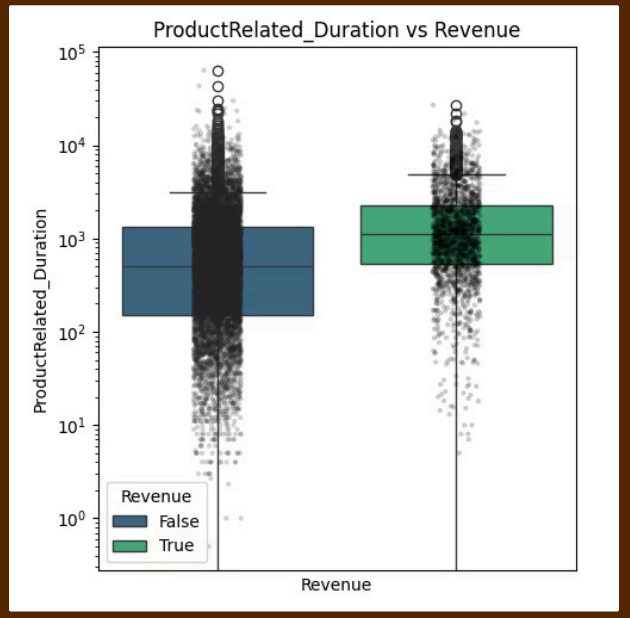
- **Actions Taken:** Pages viewed, time spent on product, information, and administrative sections.
- **Navigation Flow:** Bounce rates, exit patterns, and sequence of page interactions.
- **Session Context:** Month, day of the week, traffic source, geographical region, and device used.
- **Visitor Type:** Whether the visitor was new, returning, or categorized differently. The ultimate goal: Did they make a purchase?

Data Exploration Insights

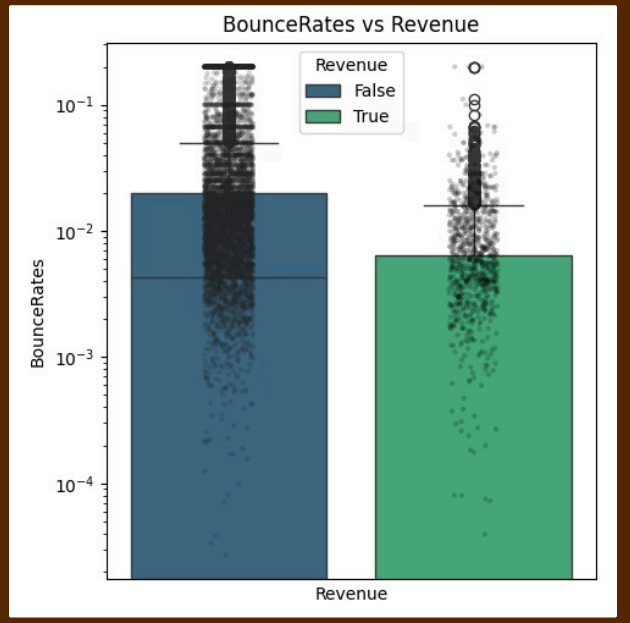
Engagement is Key to Conversion

Our initial data exploration clearly highlighted the critical role of user engagement in predicting purchase intent. Sessions with higher levels of interaction consistently correlate with increased revenue.

Time & Pages: Sessions with extended durations and more pages viewed, particularly product-related pages, are strong positive indicators.



Bounce & Exit Rates: Conversely, high bounce and exit rates are clear negative signals, indicating low engagement and conversion probability.

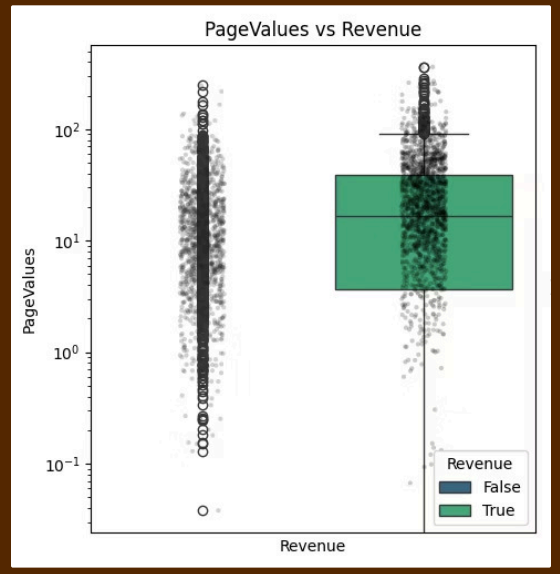


Digging Deeper into Signals

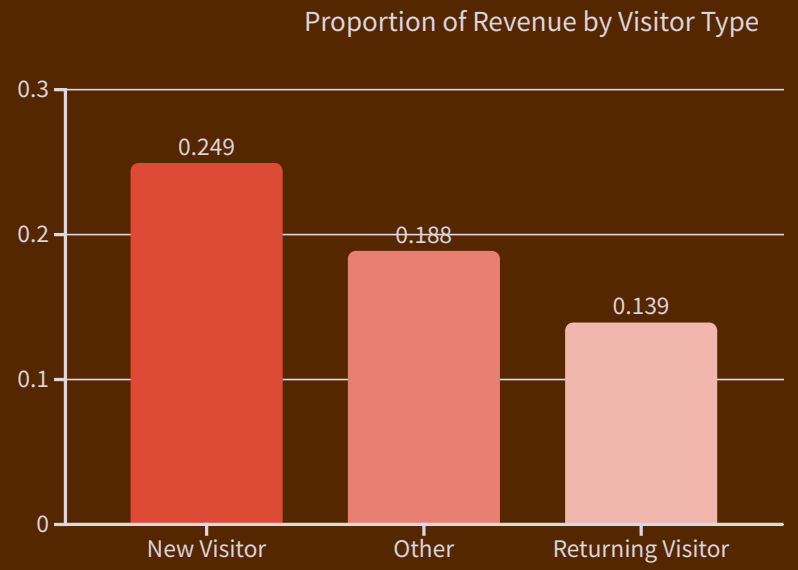
Page Value and Visitor Dynamics

Beyond general engagement, specific page interactions and the nature of the visitor reveal crucial insights into purchase intent.

Page Value: Sessions that include visits to high-value pages, such as checkout or product detail pages, are significantly more likely to culminate in a purchase. These pages represent key milestones in the conversion funnel.



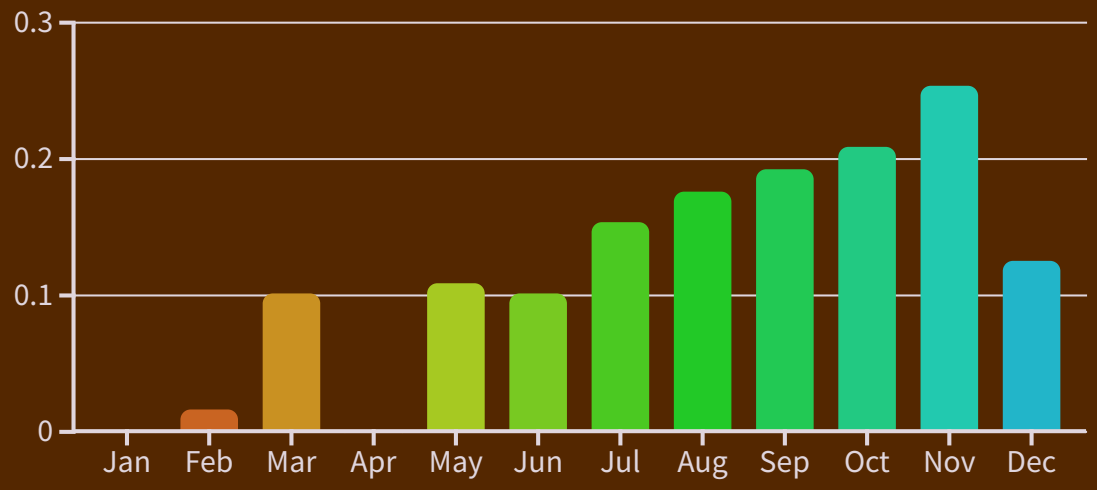
Visitor Type: While returning visitors constitute the majority of sessions, new visitors and those categorized as 'Other' exhibit a disproportionately higher purchase rate. This suggests a unique opportunity for targeted engagement with these segments.



The Power of Seasonality

Timing is a critical factor in customer purchasing behavior. Our analysis reveals distinct seasonal patterns that significantly influence conversion rates, offering clear opportunities for strategic planning.

Proportion of Revenue by Month



Peak Purchase Periods

Purchases are significantly more likely to occur in **October and November**. This aligns with typical holiday shopping trends and major sales events, underscoring the importance of intensified marketing and inventory management during these months.

Understanding these seasonal peaks allows us to optimize resource allocation, launch targeted campaigns, and prepare our infrastructure to handle increased demand, maximizing revenue potential.

Understanding User Behavior with Visitor Groups

We used clustering to identify **three distinct visitor types** based on their behavior patterns:

Group 0

Moderately Engaged Browsers

- Moderate activity levels
- Many returning visitors
- 25% purchase likelihood

Group 1

Highly Engaged Shoppers

- Very active users
- Low exit rates
- Mostly returning visitors
- **33% purchase likelihood**
- Our most valuable group!

Group 2

Low Engagement Visitors

- Little site activity
- High exit rates
- More new visitors
- Only 13% purchase likelihood

Takeaway: The visitor group is a powerful predictor of purchase intent that we can identify early in the session.

From Data to Decisions

Leveraging these insights, we developed a machine learning model designed to predict purchase intent with high accuracy. This model processes real-time session data to assign a probability score to each visitor.

Real-Time Data Feed

Visitor actions and session attributes are fed into the system instantly.

Predictive Algorithm

The model analyzes patterns and assigns a "purchase probability" score.

Actionable Insights

Scores trigger immediate, personalized actions on the website.

This proactive approach allows us to engage with visitors at their most receptive moments, significantly improving the chances of conversion.

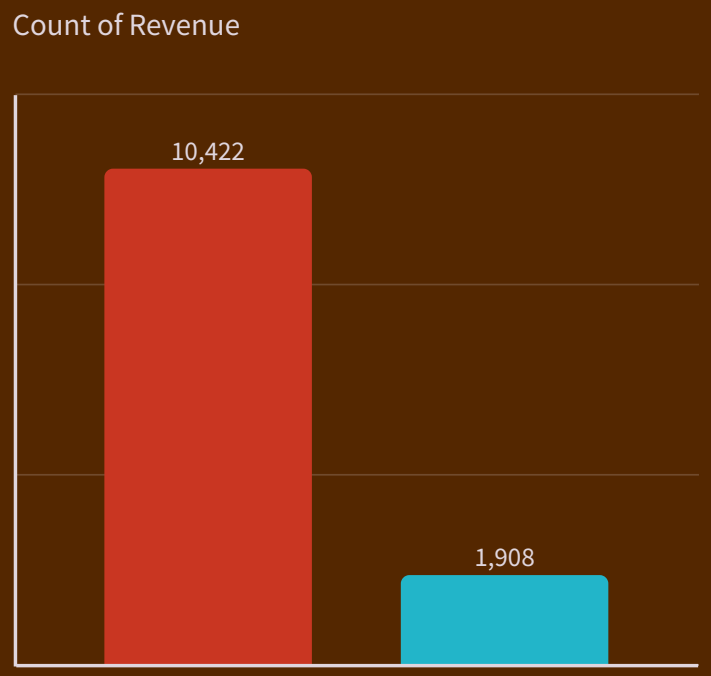
Building the Prediction Model

Addressing Data Imbalance

One of our biggest challenges was the natural imbalance in the data: most visits didn't result in a purchase. We implemented specialized techniques to help the model learn better from the smaller group of "purchase" sessions.

Model Development Process

1. **Feature Engineering**
   Created derived metrics from raw behavioral data to enhance prediction power
2. **Model Selection**
   Tested two algorithms - Random Forest and LightGBM
3. **Hyperparameter Tuning**
   Fine-tuned model settings to optimize for purchase prediction accuracy
4. **Validation**
   Rigorously tested on previously unseen data to ensure real-world performance



The significant imbalance between purchase (15%) and non-purchase (85%) sessions required specialized sampling techniques to ensure model effectiveness.

Model Performance: How Well Can We Predict?

Key Performance Metrics

Accuracy

Overall correctness of predictions across all sessions

Precision

When we predict a purchase, how often are we right? (Minimizes false alarms)

Recall

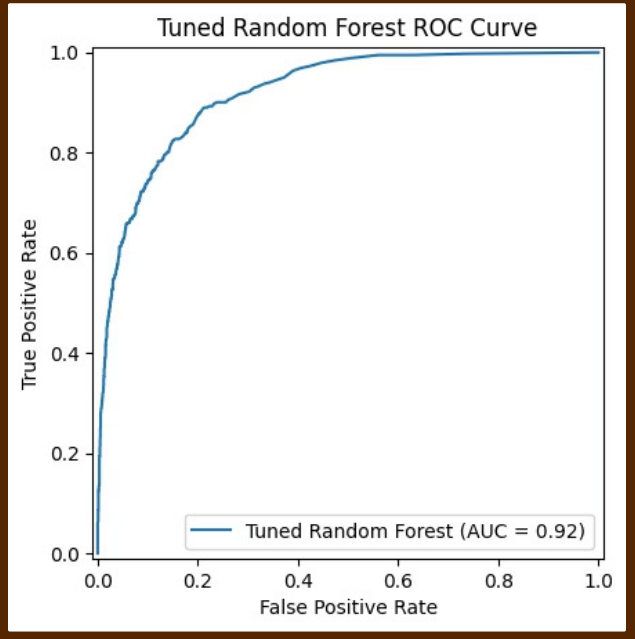
Out of all actual purchases, how many did we correctly identify? (Finds potential buyers)

ROC-AUC

Overall ability to distinguish between buyers and non-buyers (Higher is better)

Results Overview

Our models achieved excellent performance with **ROC-AUC scores around 0.92** (a perfect model would have 1.0).



Handling the data imbalance significantly increased our ability to find more potential buyers (improved Recall), with only a slight increase in false positives (reduced Precision).

Our Chosen Model:

Tuned Random Forest

92%

ROC-AUC Score

Overall ability to distinguish between buyers and non-buyers

75%

Precision

When we predict a purchase, we're right 3 out of 4 times - excellent for targeted actions!

55%

Recall

We successfully identify more than half of all purchase-intent sessions

After evaluating both models, the **Tuned Random Forest** provides the best balance of precision and recall, making it ideal for real-time intervention without overwhelming customers with irrelevant prompts.

Actionable Insights & Next Steps

Identify High-Potential Sessions

The model can predict in real-time which active user sessions have the highest purchase probability.

Enable Targeted Interventions

- Show personalized offers or recommendations based on browsing history
- Offer real-time chat support to users showing high intent but encountering difficulty
- Dynamically tailor website layout or messaging to maximize conversion probability

Implementation Roadmap

1. Week 1-2
   ### Deploy the Model
   Integrate prediction API with website backend for real-time scoring
2. Week 3-4
   ### A/B Testing
   Test intervention strategies with different user segments to identify optimal approaches
3. Week 5-8
   ### Monitor & Learn
   Track performance metrics, gather feedback, and refine approach
4. Ongoing
   ### Continuous Improvement
   Retrain model with new data monthly to improve prediction accuracy

From Reactive to Proactive

Business Impact

- **Increase Conversion Rates**

  By targeting the right users with the right interventions at the right time

- **Optimize Marketing Spend**

  Focus resources on high-potential customers likely to convert

- **Improve User Experience**

  Provide personalized support and recommendations based on real-time intent

- **Drive Revenue Growth**

  Convert browsers to buyers through data-driven, targeted engagement

Questions?

Thank You!