

Phase 3 Project

PREDICTING PURCHASE BEHAVIOURS IN ONLINE SHOPPERS

Margaret Sarah Wanjiru Njoroge



PROJECT OVERVIEW

I used machine learning to predict whether a visitor to an e-commerce website will make a purchase, taking into consideration the different types of Visitors and durations. I used a dataset from UC Irving machine learning repository: https://archive.ics.uci.edu/ml/machine-learning-databases/00468/online_shoppers_intention.csv

This predictive capability helps marketing and product teams improve user engagement and increase conversions.

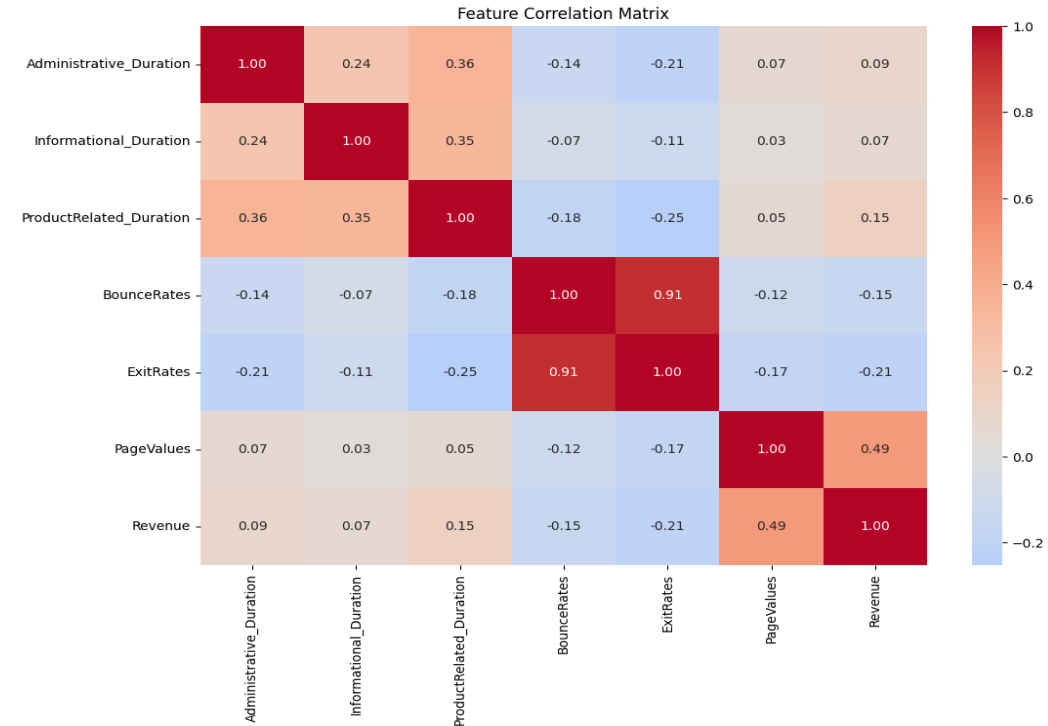
BUSINESS AND DATA UNDERSTANDING

1. Business Stakeholders: E-commerce marketing and product team.
2. Business Problem: How can we identify customer behaviour in sessions that are more likely to result in purchases?
3. Dataset: Online Shoppers Intention dataset with 12,330 sessions and 18 features.
4. Target Variable: Revenue (1 = purchase, 0 = no purchase).

KEY DATA INSIGHTS

Highlights

1. Most sessions do not lead to a purchase (Correct class imbalance).
2. Features like Page Values and Bounce Rates are highly predictive.
3. Returning visitors and sessions in November show higher conversion rates. Signifying the need for Marketing efforts focus on months with higher purchase activity likely in preparation to the holidays.



Talking Points on Purchasing behaviour:

- Page_Values has the strongest positive correlation with Revenue (0.49). Optimizing high-value pages (like product or checkout pages) to increase the chance of purchase.
- ExitRates and Bounce_Rates are strongly correlated (0.91) and negatively associated with Revenue. Reducing early exits by improving landing pages and navigation experience.
- Time spent on ProductRelated pages shows mild positive correlation with Revenue. Encourage more meaningful product exploration through smart recommendations or better filters.

MODELING EVALUATION AND PERFORMANCE

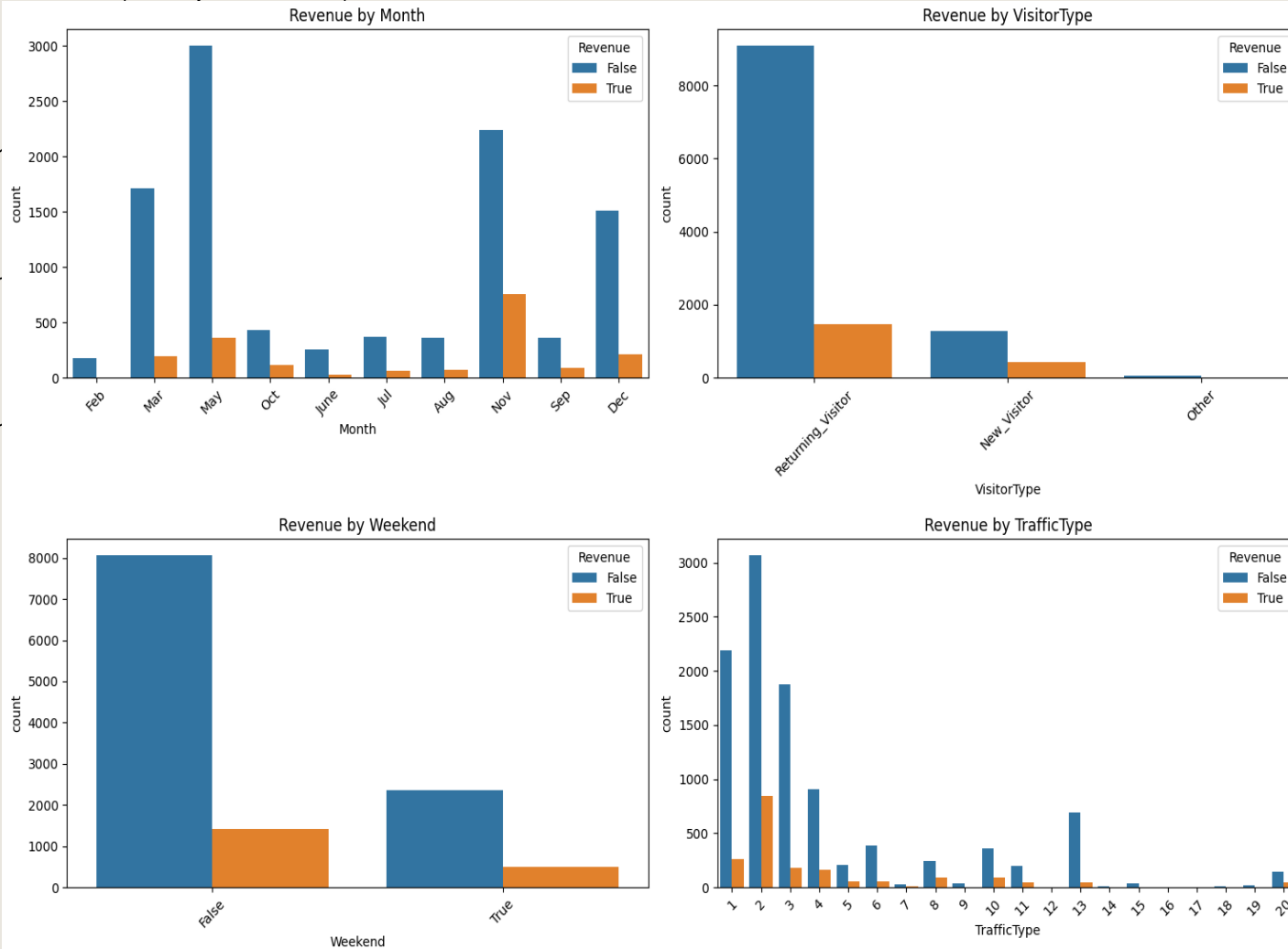
MODELLING

- Building multiple models to predict purchases.
- Baseline Model: Using Logistic Regression
- Improving the Models: Gradient Boosting and Neural Network
- Feature engineering helped improve performance (e.g. combining durations, interaction features).

EVALUATION

- Evaluated models using ROC AUC to balance sensitivity and specificity.
- **Best model:** Gradient Boosting, with high accuracy in identifying purchase sessions.
- Confusion matrix analysis revealed few false positives and negatives.

CATEGORICAL FEATURE IMPACT ON REVENUE



Most important factors:

1. High PageValue (users finding value)
 2. Low BounceRate (users engaging deeply)
 3. Total time spent on site
- Returning Visitors more likely to buy

Talking Points:

- November shows a purchase spike → likely due to holiday shopping (Assumption; Black Friday).
- Returning visitors contribute more to purchases than new/other.
- Purchases occur both on weekdays and weekends, but weekday sessions are more frequent.
- A few specific traffic types (e.g., 1, 2, 3) dominate revenue generation.



BUSINESS RECOMMENDATIONS

1. Enhance High-Value Pages to Drive Revenue

- Identify and optimize pages with high *PageValues* i.e., typically product pages, comparison tools, or checkout steps. These pages show the strongest link to purchases. Investing in their design, speed, and clarity can increase conversions.

2. Reduce Bounce and Exit Rates on Key Entry Pages

- Focus on improving user_experience landing pages and the homepage, ensure they load fast, show clear value, and guidance. High *BounceRates* and *ExitRates* are linked with fewer purchases. This means users are dropping off before engaging meaningfully.

3. Encourage Deeper Product Exploration

Use strategies like “related items,” filters, reviews, and better visuals to keep users on *ProductRelated* pages longer. More time on these pages slightly increases purchase likelihood — and it may become more impactful if paired with relevant content.

4. Segment Returning vs. New Visitors

Offer targeted messages with personalized content or incentives to *Returning Visitors*, who already have a higher intent to buy. The data shows returning visitors are much more likely to generate revenue than first-timers.

5. Re-evaluate Marketing Channels

Review *TrafficType* data to find out which channels (e.g., search, ads, referrals) drive high-converting users. Cut underperformers. Certain traffic sources bring more paying customers, leading to invest more where ROI

NEXT STEPS Q&A SESSIONS

1. Run a PageValue audit:

Using analytics tools we rank pages by PageValue and focus optimization on the top 10.

2. Improve bounce-prone pages:

A/B test layout and messaging for pages with high exits/bounces this will reduce the exits and bounce.

3. Build engagement journeys:

Develop pathways that guide users from informational pages toward product-related pages.

4. Target segments differently:

Using personalized email campaigns or promotions for returning visitors to convert them faster.

5. Marketing channel analysis:

Map TrafficType to AD spend and acquisition cost, reallocate budget to channels with higher conversion potential. ROI Analysis.

6. Measure changes:

For your better trackoing Set KPIs such as conversion rate, bounce rate reduction, and average session duration to track improvement.

A series of white, thin, overlapping geometric lines on a black background, forming a complex, abstract shape on the left side of the slide.

THANK YOU