

# Online Shoppers Intention Analysis

Machine Learning for Purchase Prediction

By Ujjwal Bhardwaj | September 2025



# Analytical Methodology: A Data-Driven Approach

Our project followed a structured, five-phase methodology to uncover and predict online shopper intentions.



## Data Exploration

Initial analysis of behavioral features and target conversion rates.



## Feature Analysis

Identifying key correlations and customer behavioral patterns.



## Preprocessing

Transforming raw data for optimal model performance.



## Model Training

Comparative evaluation and optimization of predictive models.



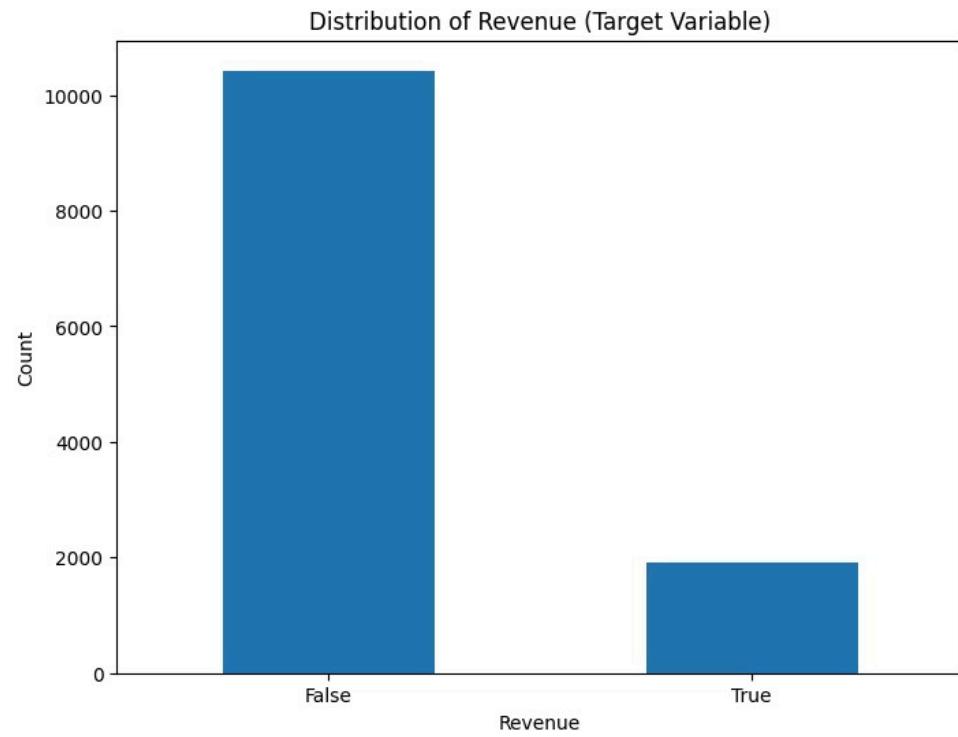
## Deployment

Translating insights into actionable business recommendations.

# Phase 1 & 2: Understanding the Data Landscape

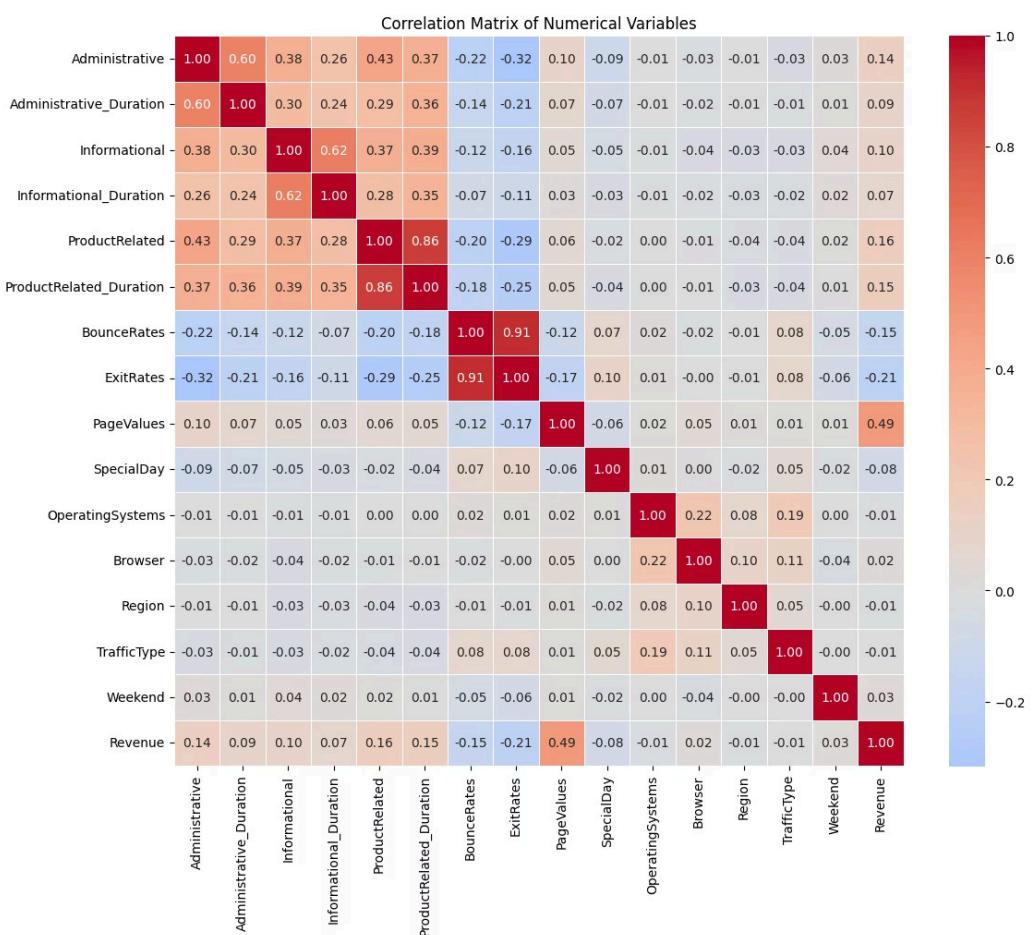
## Data Exploration

Our dataset comprised **12,330 unique shopping sessions**, each enriched with **18 distinct behavioral features**. A critical discovery was the target analysis: only **15.5% of these sessions resulted in a purchase conversion**, highlighting the predictive challenge.



## Feature Analysis

Through rigorous **correlation analysis**, we identified significant relationships between various features and purchase intent. This allowed us to segment customers based on their **distinct online behavioral patterns**, providing a foundational understanding for targeted strategies.



# Phase 3 & 4: Data Preparation and Model Training

Rigorous preparation and strategic model selection were key to building a robust predictive system.

## Preprocessing



We applied **Boolean encoding** to categorical variables and performed **feature scaling** to normalize data. A **stratified train-test split (80/20)** ensured balanced representation for robust model evaluation.

## Model Training



Our core objective involved comparing **Random Forest** against **Logistic Regression**. Special emphasis was placed on **performance optimization for imbalanced data** to accurately predict the minority class (purchases).

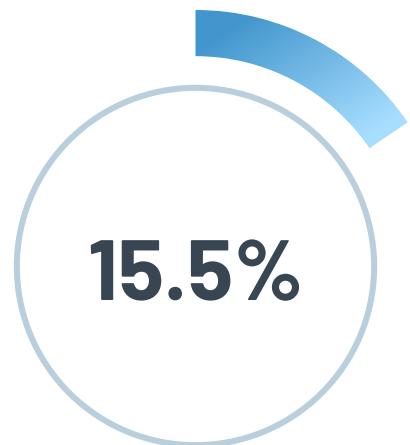
Data  
Preprocessing

Feature  
Engineering

Model  
Training

Evaluation

# Critical Business Insights

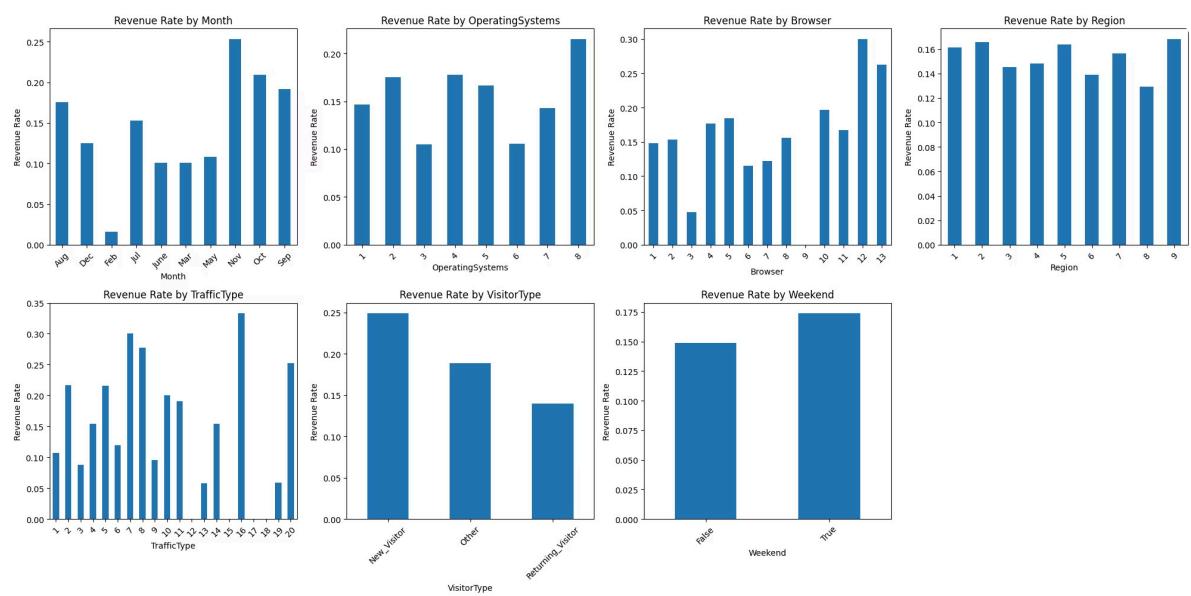


## Purchase Conversion

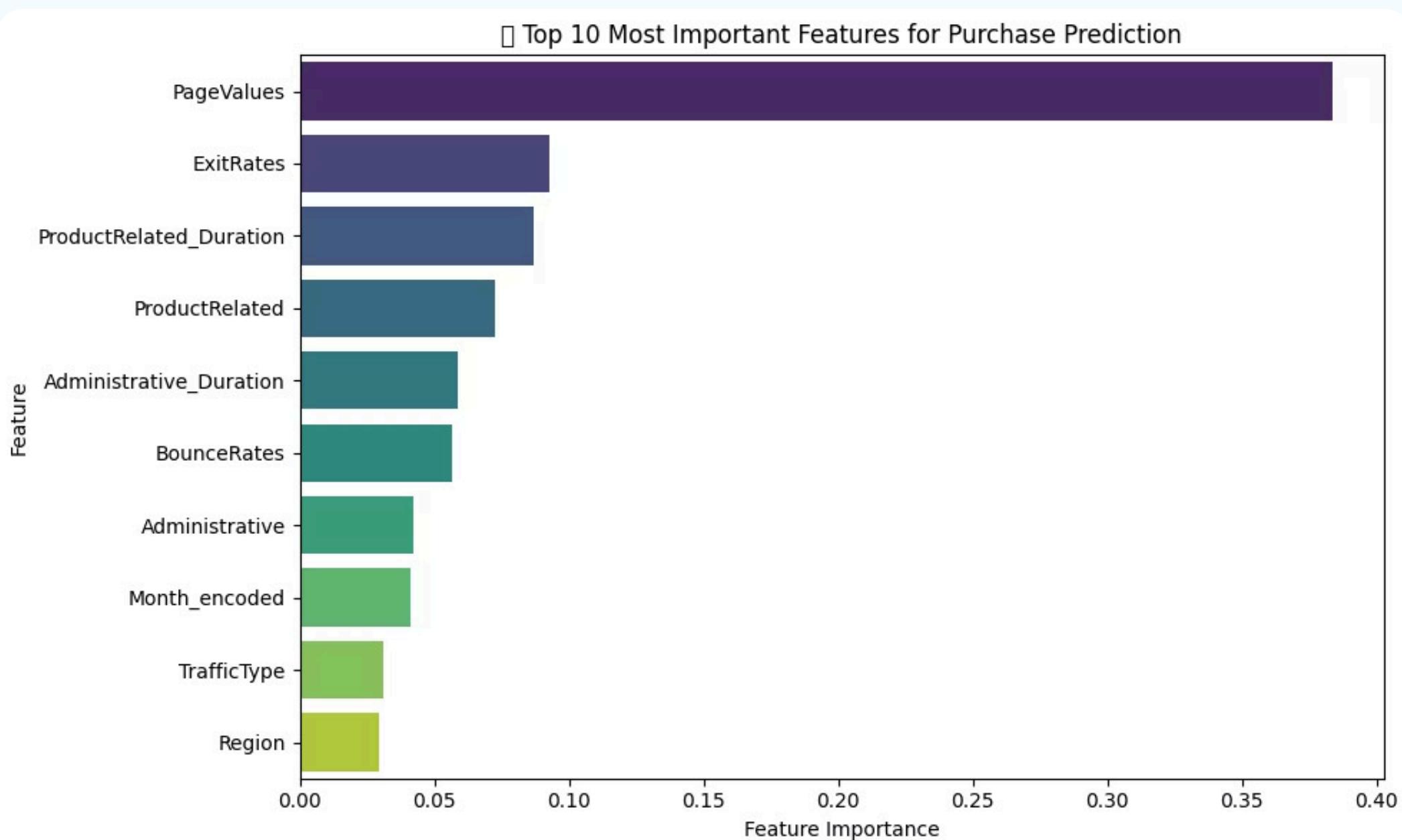
Only a small fraction of visitors convert, highlighting areas for optimization.



- **Weekend Shopping Premium:** Weekend sessions convert **17% higher than weekdays** (17.5% vs 15%), indicating that weekend-specific promotions, enhanced customer support, and mobile optimization should be prioritized for maximum impact.
- **Customer Acquisition Focus:** **New Visitors convert at 25%** compared to only 19% for returning visitors, suggesting that first-impression optimization and new customer targeting campaigns will yield higher ROI than retention-focused efforts.
- **Seasonal trends affect purchase likelihood:** November shows the highest conversion rate at **25%** - nearly double the baseline rate - indicating strong holiday shopping behavior that should drive concentrated marketing investment and inventory planning for Q4.
- **Technical Performance Gaps:** Significant browser and operating system variations (ranging from 5% to 30% conversion rates) reveal technical optimization opportunities that could improve overall site performance across different user environments.



# Top Predictive Features: Driving Intent



This chart illustrates the relative importance of each feature in predicting online purchase intentions. **PageValues** stands out as the most influential factor, emphasizing the critical role of user engagement with specific product pages.

# Model Performance: Random Forest Triumphs

A direct comparison between our chosen models reveals superior predictive power of the Random Forest classifier.

| Model               | Accuracy | Precision | Recall |
|---------------------|----------|-----------|--------|
| Random Forest       | 90.15%   | 73.0%     | 58.0%  |
| Logistic Regression | 88.32%   | 76.0%     | 36.0%  |



## Winner: Random Forest

Achieved best overall performance, especially on imbalanced datasets.



## Feature Interaction

Excels in handling complex feature interactions for better predictions.

# Driving Business Value & ROI

Our predictive model is poised to deliver significant improvements across key business metrics.

**15-25%**

## Marketing Efficiency

Targeted campaigns leading to reduced wasted ad spend.

**10-15%**

## Conversion Rates

Directly boosting the percentage of visitors who make a purchase.

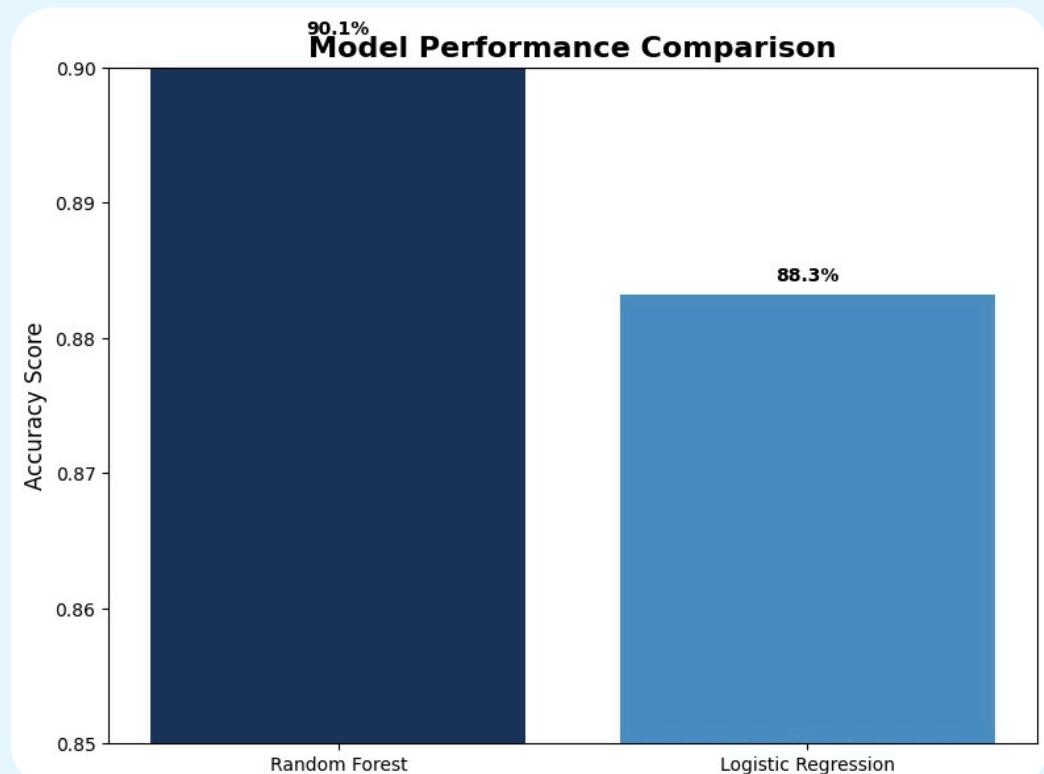
**20-30%**

## Wasted Ad Spend

Optimizing ad placements by focusing on high-potential shoppers.

### Deployment Status:

- Achieved **90.1% accuracy** on purchase prediction.
- Production-ready models saved for immediate integration.
- Real-time prediction pipeline completed for dynamic insights.



# Strategic Implementation & Future Outlook

The finalized model is ready for seamless integration, offering immediate and long-term benefits.

1

Model serialized for easy loading and execution in any environment.

2

Prediction pipeline enables real-time scoring of new user sessions.

3

Integration with existing marketing automation platforms.

This robust infrastructure ensures that actionable insights are consistently delivered to optimize online marketing and sales strategies.

# Questions & Discussion

Thank you for your attention. We welcome your questions and feedback.

→ **Project Repository:**

[github.com/Ujjwal012003/ExaltAI\\_Capstone\\_Project](https://github.com/Ujjwal012003/ExaltAI_Capstone_Project)  
Ujjwal

→ **Trained Models:**

Production-ready Random Forest classifier for immediate use.

→ **Complete Analysis:**

Detailed Jupyter Notebook with full methodology and code.

→ **Documentation:**

Comprehensive guidelines for implementation and maintenance.

**Contact:**

ujjwaal05@gmail.com