BAN6800: Business Analytics Capstone

Nexford University

Milestone Two Assignment

Title: *Behavioral Segmentation and Predictive Modeling of Purchasing Intent Among Takealot Online Shoppers*

Subtitle: *Business Analytics Model Results*

Name: Mubanga Nsofu
Learner ID: 149050
Date: 7th June 2025
Lecturer: Prof. Raphael Wanjiku

Introduction

Future-Proofing Takealot: A Machine Learning Response to Market Disruption

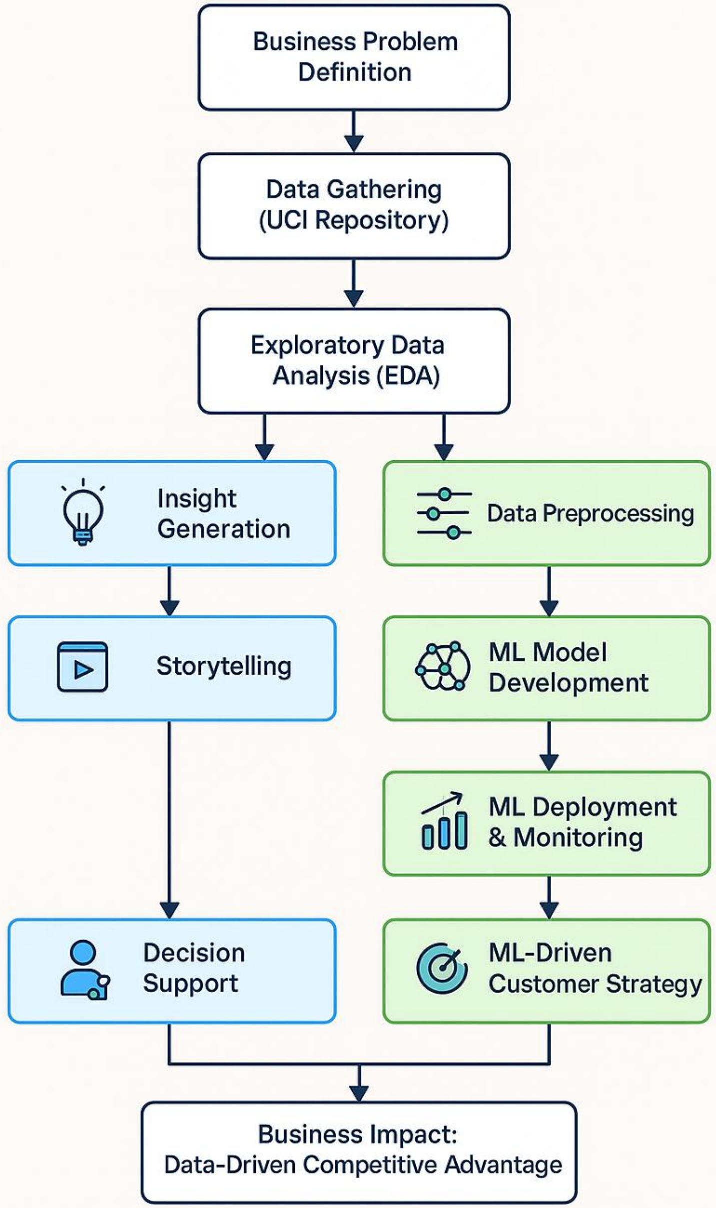
Situation
Takealot, a 14-year-old South African e-commerce leader, dominates the local market through scale and customer reach.

Complication
The recent market entry of global e-commerce glant Amazon threatens Takealot's market share and customer loyalty.

Question
How can Takealot protect and grow its competitive edge in this evolving digital retail landscape?

Answer
By leveraging applied machine learning–through customer behavioral clustering and purchase intent classification–to drive smarter, personalized engagement and strategy.

End-to-End Workflow for Takealot Customer Analytics

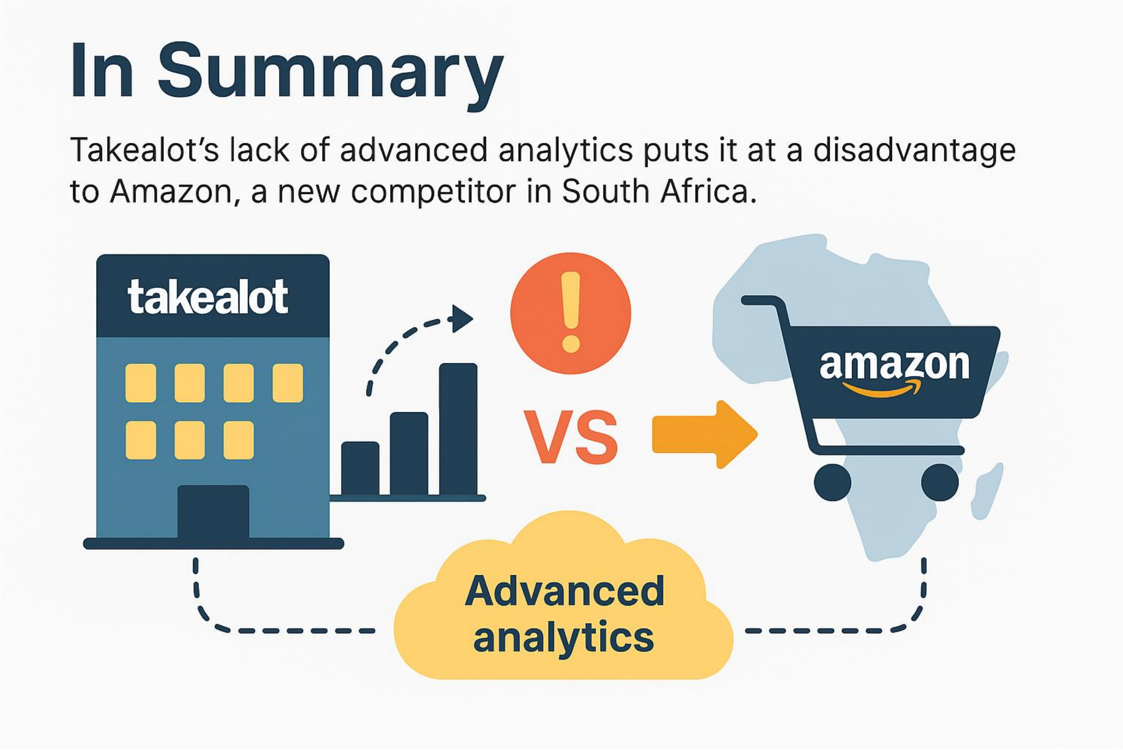


6-week project following the CRISP-DM process(Roy, 2018)

Takealot is a leading e-commerce player in South Africa (Wikipedia, 2025).

Nexford University

Business Problem Understanding: Takealot potential loss of market share and dominance are under threat from Amazon



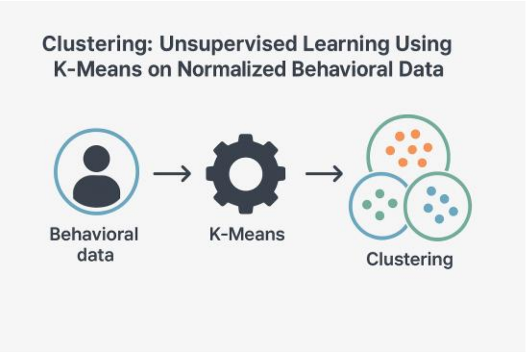
Nexford University

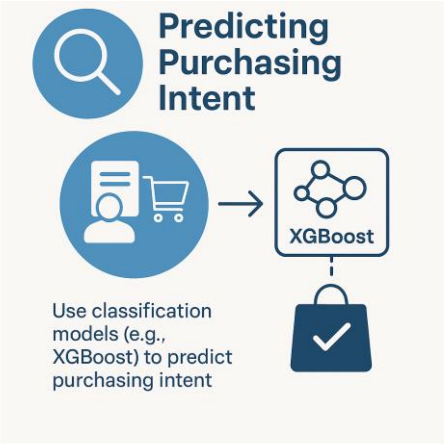
Solution Overview

PERFORM BEHAVIOIRAL CLUSTERING TO IDENTIFY CUSTOMER SEGMENTS

USE CLASSIFICATION MODELS TO PREDICT PURCHASING INTENT

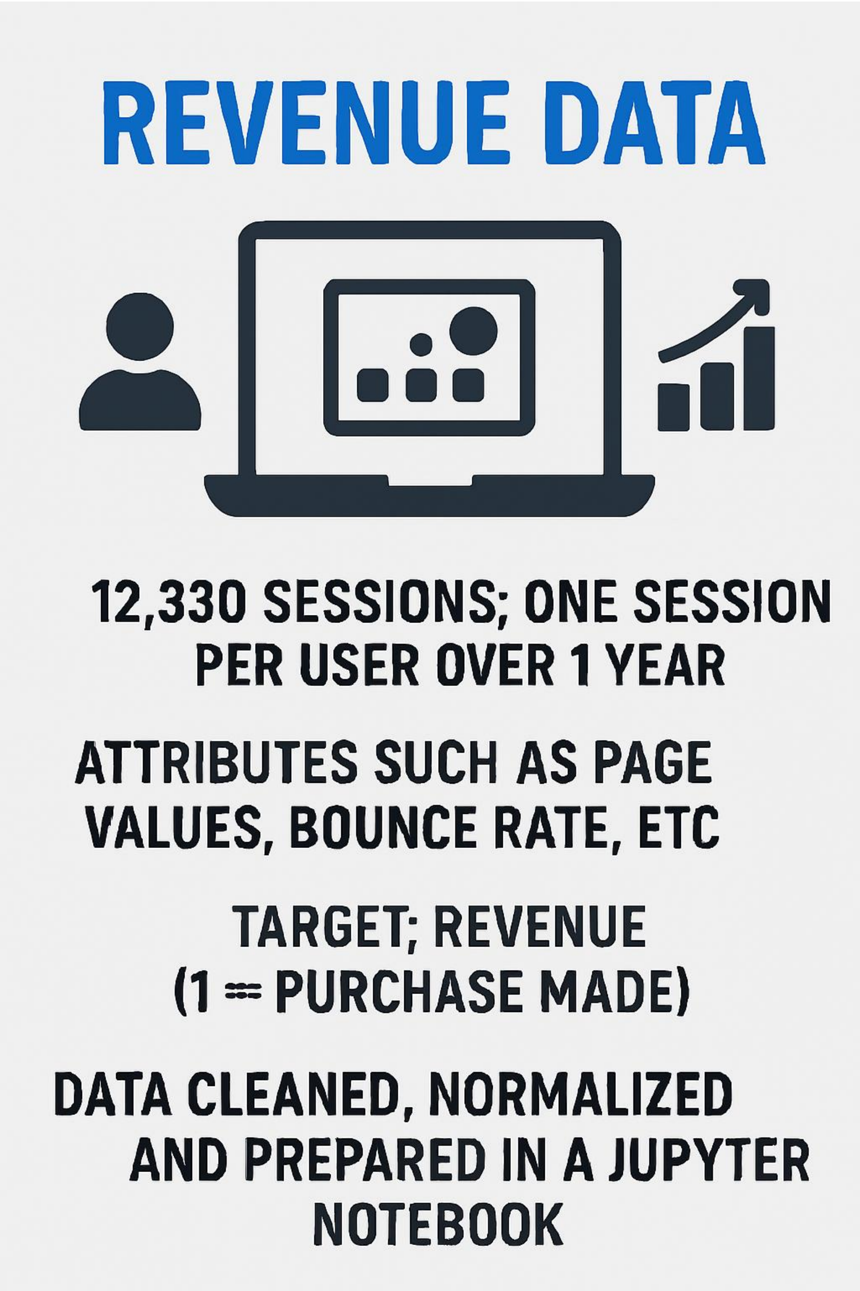
DERIVE ACTIONABLE INSIGHTS FOR USE BY ALL STAKEHOLDERS

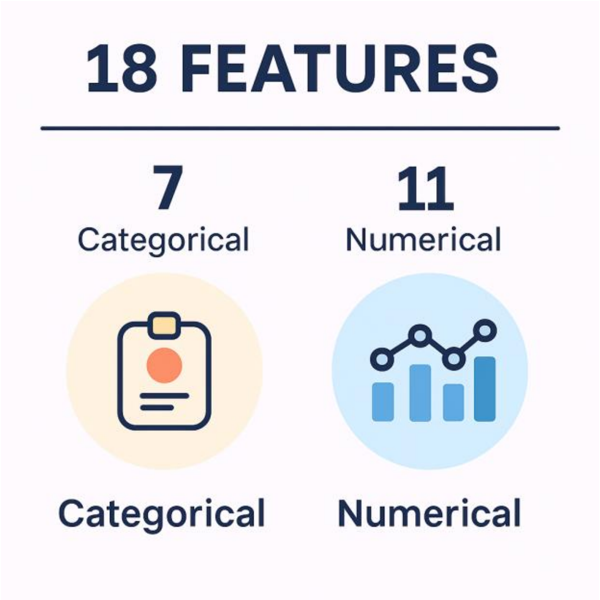




The solution combines techniques from two branches of machine learning to deliver actionable insights for Takealot to compete effectively

Dataset Summary: Online shoppers' intention dataset

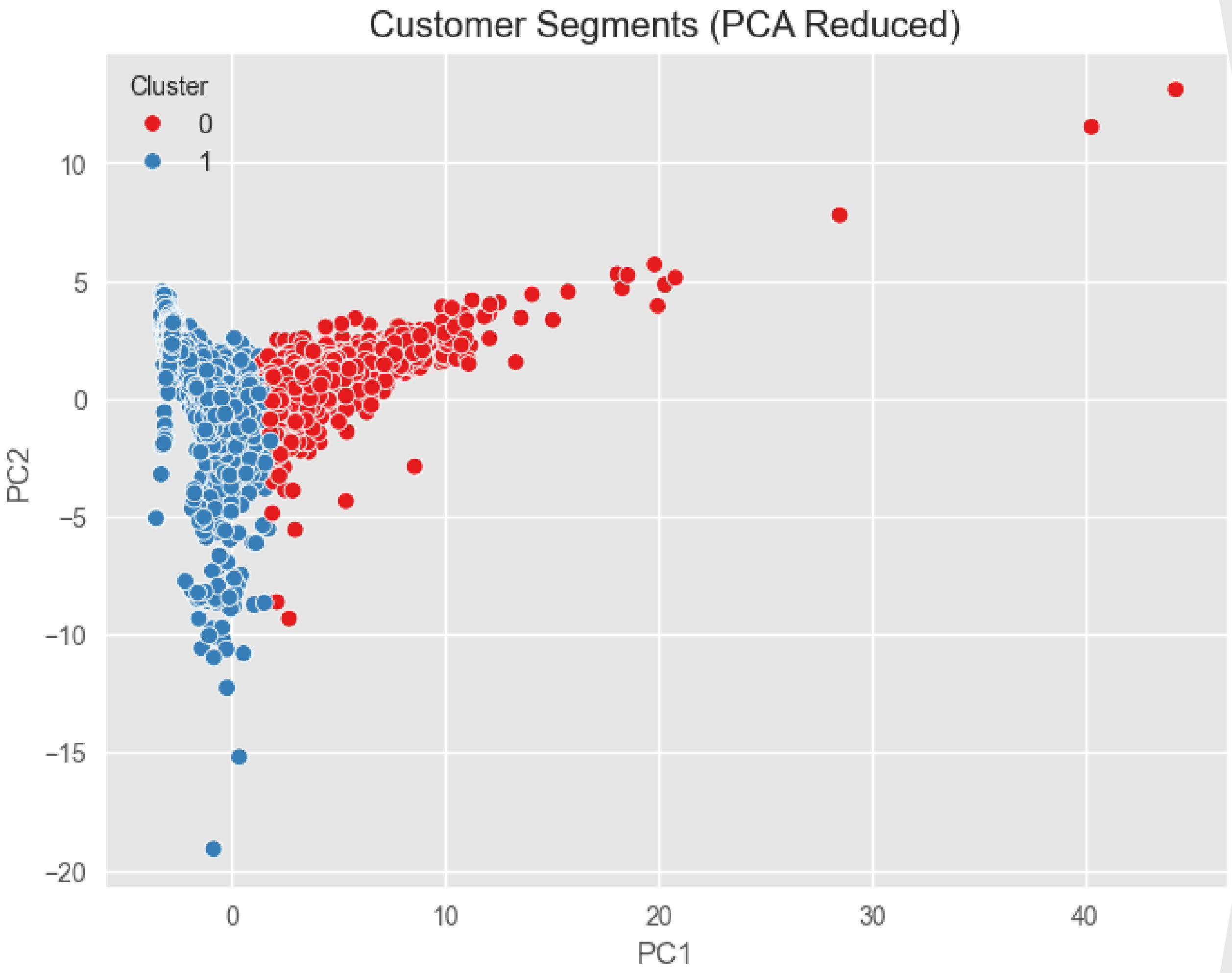




- The dataset was downloaded from the UCI repository (Sakar & Kastro, 2018).
- The dataset contained no missing values, but it was still cleaned, normalised and preprocessed (e.g. dimensionality reduction, class imbalance treatment etc.) to ensure it could be fed into machine learning algorithms.

Nexford University

Behavioral Clustering Results (1/2)



Clusters

Cluster 0

High-Intent Shoppers

- Likely to buy
- Strong engagement
- High product-related activity
- Low bounce and exit rates
- Browse deeply, higher conversion rates

Cluster 1

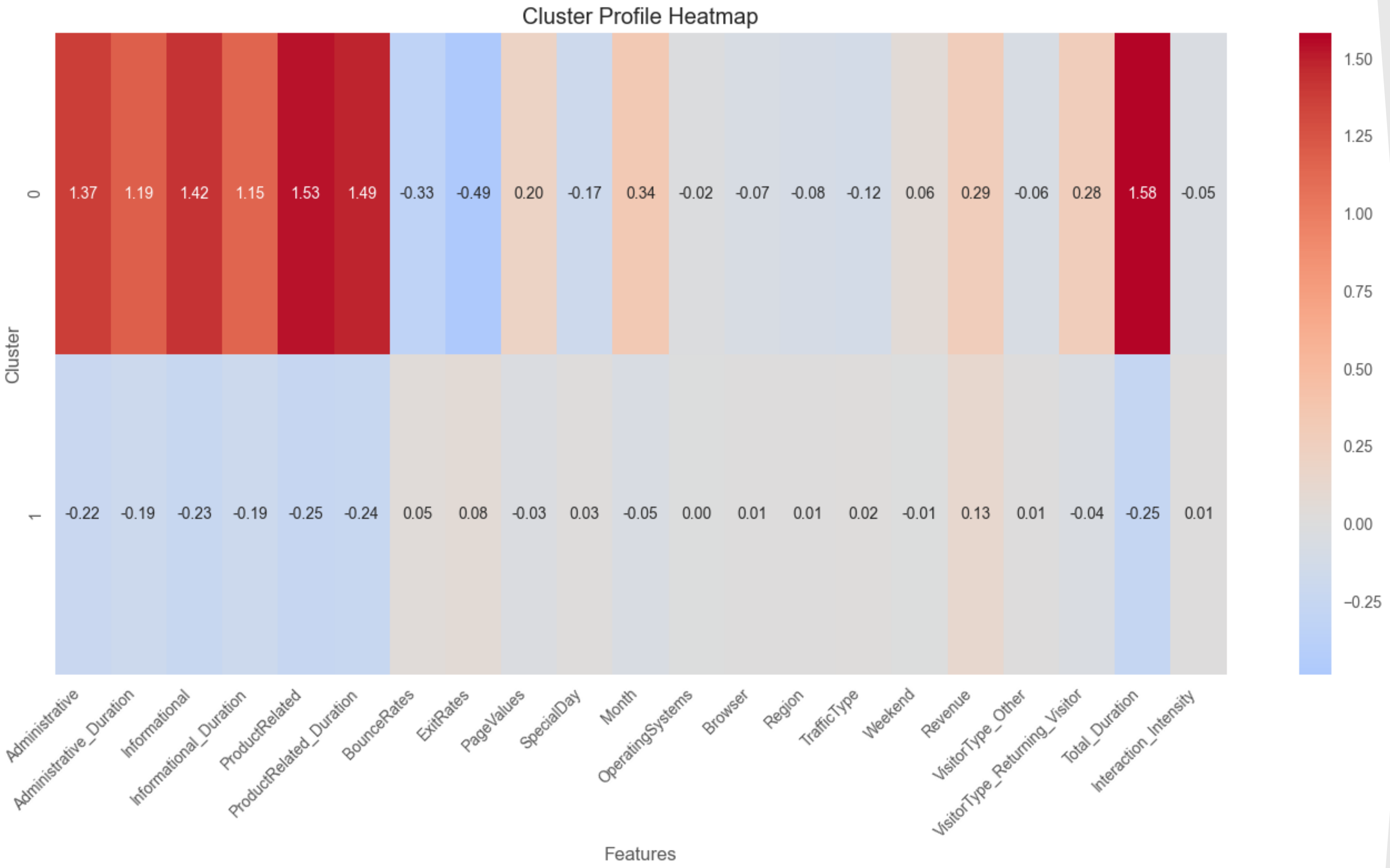
Casual Visitors

- Unlikely to buy
- Lower engagement
- Fewer page visits, shorter durations
- Slightly positive bounce and exit rates

Window shopping, less likely to convert

Actionable recommendations for both the CEO and Marketing teams are provided in slides 9 and 10, respectively

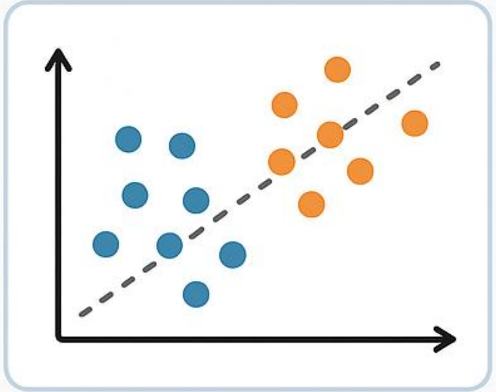
Behavioral Clustering Results (2/2)



Actionable recommendations for both the CEO and Marketing teams are provided in slides 9 and 10, respectively

Behavioral Clustering - Diagnostics

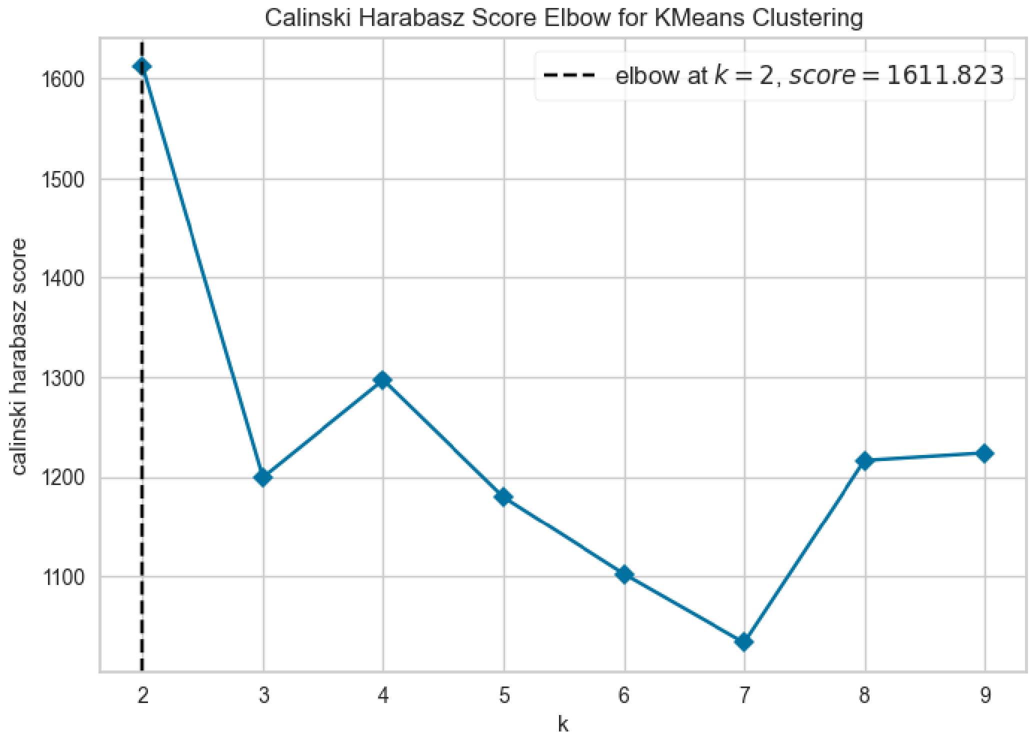
- Applied KMeans (k=2) with PCA for dimensionality reduction.



Davies-Bouldin Score: 1.5774 (moderate separation)

Homogeneity Score: 0.0219 (indicates weak label alignment – expected)

Clusters are used to understand behavioral patterns only (not for prediction)



Diagnostic Analysis of the Model

- The Calinksi-Harabasz score shows a cluster of two is optimal (high intent or casual visitor)
- The Davies-Bouldin score shows distinct clusters, making the model suitable for behavioral insights.
- Since the model is used for classification, low homogeneity is not necessarily a concern

CEO Strategic Insights derived from behavioral clustering

Strategic Recommendations

Focus on Cluster 0 for Revenue Growth

Insight: Cluster 0 shows significantly higher interaction with product-related pages, longer durations, and higher page values,

Action: Prioritize investment in user experience, recommendation engines, and fast checkout for these high-value users to boost conversions further.

Leverage Behavioral Segmentation in Strategy

Insight: The segmentation reveals two distinct customer archetypes –engaged buyers vs. casual browsers.

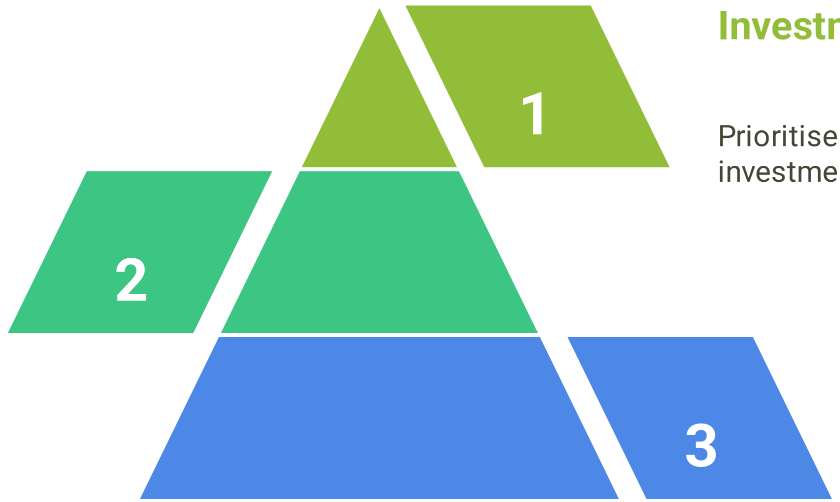
Action: Embed this segmentation into CRM, inventory, and pricing strategies. Consider exclusive campaigns or loyalty programs for cluster 0

Benchmark Against Amazon's UX

Insight: Cluster 0 exhibits low bounce and exit rates – meaning Takealot already captures attention effectively here.

Action: Benchmark these UX flows vs. Amazon to identify differentiators and opportunities to retain competitive edge.

Strategic Growth Pyramid



Investment Focus

Prioritise cluster 0 investments

Segmentation Integration

Embed segmentation into core strategies

Benchmarking

Benchmark Ux against Amazon

Tactial marketing insights derived from behavioral clustering

Marketing Actions from Segmentation

1. Retarget Cluster 1 with Personalized Campaigns

Insight: Cluster 1 has low engagement across the board (shorter time, fewer visits, slightly positive bounce/exit).

Action: Deploy personalized email offers or display ads tailored to browsing behavior to nudge them toward purchase.

2. Reinforce High-Intent Behavior (Cluster 0)

Insight: High product interaction & low bounce suggest readiness to buy.

Action: Use trigger-based marketing—e.g., abandoned cart emails, stock alerts, and limited-time offers.

3. A/B Test Landing Pages

Insight: Informational and administrative pages see higher traffic in Cluster 0.

Action: A/B test landing page content for different clusters to improve engagement and reduce drop-off in Cluster 1.

4. Use Interaction_Intensity to Score Leads

Classification Model – XGBoost Results (1/2)

XGBoost Model Performance Summary

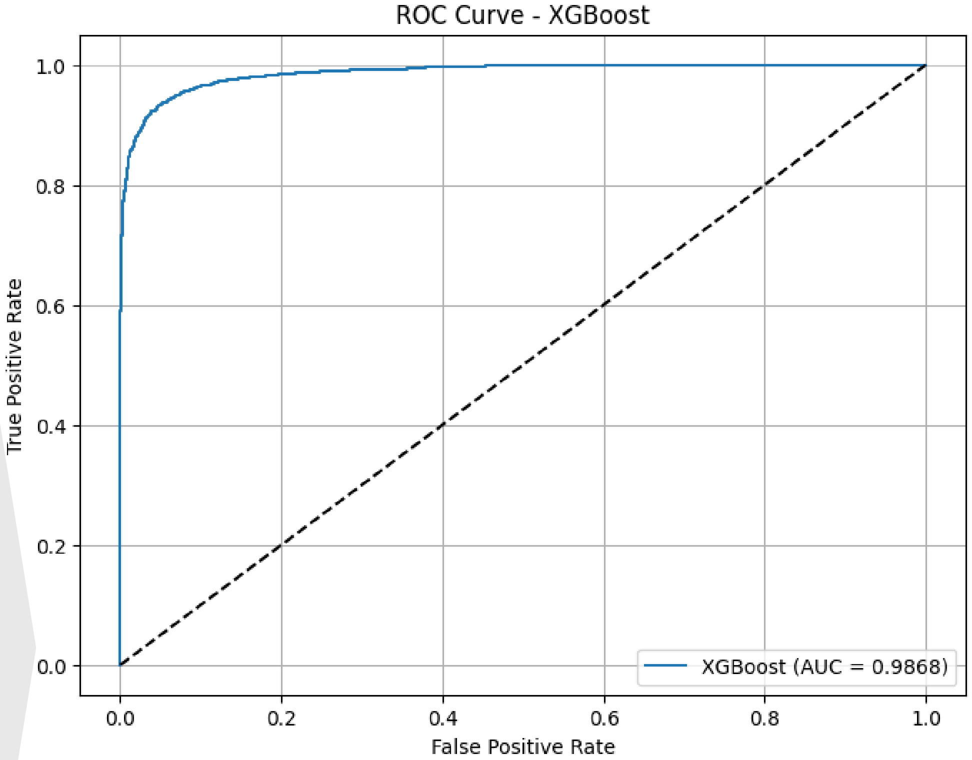
Predictive Modelling
Logistic Regression used as benchmark model

Excellent ROC AUC
0.9868 (XGBoost), SMOTE used for class imbalance

Feature Importance
SHAP identified the top 15 influential features

Avoid Overfitting
Early stopping, Optuna tuning, stratified CV

Model Deployment
Final model exported for reuse




```
XGBoost Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.94      0.94      3127
           1       0.94      0.94      0.94      3127

    accuracy                           0.94      6254
   macro avg       0.94      0.94      0.94      6254
weighted avg       0.94      0.94      0.94      6254

XGBoost ROC AUC: 0.9868269277281984
```

1. Our Model Predicts Customer Purchases with High Accuracy The curve indicates that our model can effectively distinguish between buyers and non-buyers.

2. An AUC of **98.7%** means it's **rarely wrong** — it almost always predicts correctly whether a customer will make a purchase or not.

3. This minimises both:

- **False positives** (saying someone will buy when they won't)
- **False negatives** (missing out on genuine buyers)

4. This gives us strong confidence to use the model for smarter marketing and customer targeting decisions.

XGBoost Summary Model Performance for IT

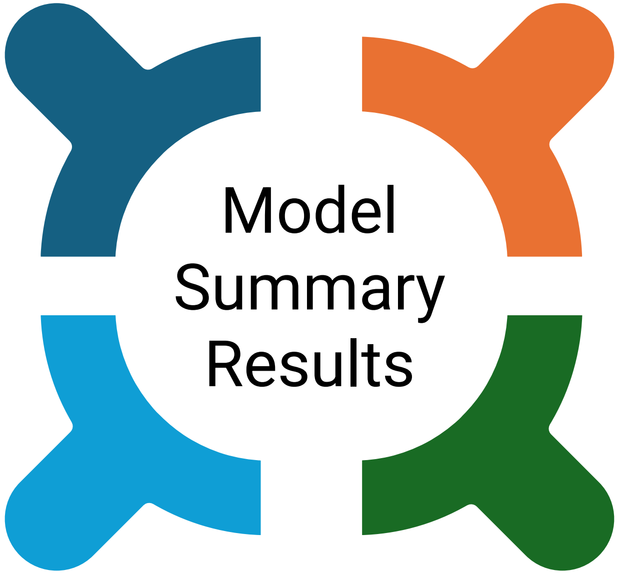
Recall is an important metric for the online purchasing behavior use case
The model predicts 94% of non-purchasing intent cases correctly- macro average
The model predicts 94% of purchasing intent cases correctly- weighted average

```
XGBoost Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.94      0.94      3127
           1       0.94      0.94      0.94      3127

    accuracy                           0.94      6254
   macro avg       0.94      0.94      0.94      6254
weighted avg       0.94      0.94      0.94      6254

XGBoost ROC AUC: 0.9868269277281984
```



1 Accuracy

The model correctly predicts 94% of the cases

How often the model is actually correct

2 AUC-ROC

The model's value is 98.6% which can predict purchasing intent

Out performs the logistic regression classifier baseline model

3 Weighted evaluation

- 94% precison
- **94% Recall**
- 94% F1- score

Depicts performance of each class using a weighted average

4 Macro evaluation

- 94% precison
- **94% Recall**
- 94% F1- score

Depicts performance of each class equally

Classification Model – XGBoost Results (2/2)

XGBoost Model Performance Summary

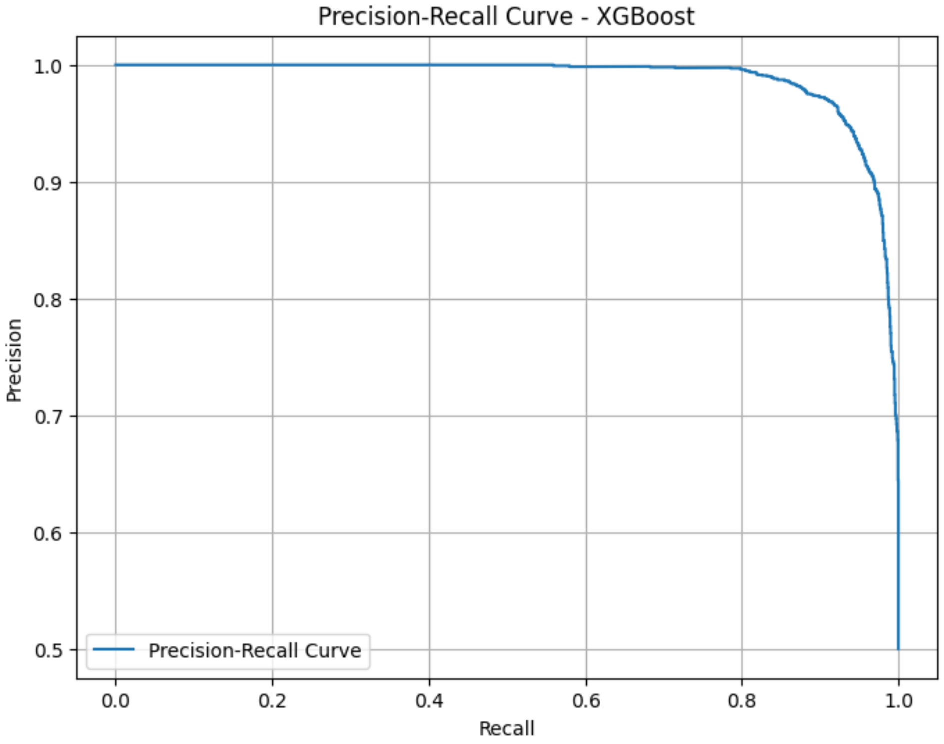
Predictive Modelling
Logistic Regression used as benchmark model

Excellent ROC AUC
0.9868 (XGBoost), SMOTE used for class imbalance

Feature Importance
SHAP identified the top 15 influential features

Avoid Overfitting
Early stopping, Optuna tuning, stratified CV

Model Deployment
Final model exported for reuse




```
XGBoost Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.94      0.94      3127
           1       0.94      0.94      0.94      3127

    accuracy                           0.94      6254
   macro avg       0.94      0.94      0.94      6254
weighted avg       0.94      0.94      0.94      6254

XGBoost ROC AUC: 0.9868269277281984
```

The chart illustrates how effectively our model balances accuracy and coverage in predicting who will make a purchase.

The curve stays near the top, meaning our predictive model is highly precise—we now rarely misidentify someone as a buyer when they're not.

The model maintains a strong recall, meaning we can successfully identify most true buyers. This enables us to identify actual buyers successfully: we can target real potential customers more confidently and waste less on people unlikely to convert.

Key Takeaways:
1. High precision = Marketing spends less on the wrong people.
2. High recall = We catch almost all real buyers.
3. Stable curve = Consistency in model performance—excellent for real-world decision-making

Nexford University

Logistic Regression Model Performance for IT

Recall is an important metric for the online purchasing behavior use case
The model predicts 79% of non-purchasing intent cases correctly- macro average
The model predicts 87% of purchasing intent cases correctly- weighted average

```
Best parameters for Logistic Regression: {'C': 9.978862418509248}

Logistic Regression Classification Report:
              precision    recall  f1-score   support

           0       0.79      0.88      0.83      3127
           1       0.87      0.76      0.81      3127

    accuracy                           0.82      6254
   macro avg       0.83      0.82      0.82      6254
weighted avg       0.83      0.82      0.82      6254

Logistic Regression ROC AUC: 0.9001514502416565
```

Accuracy

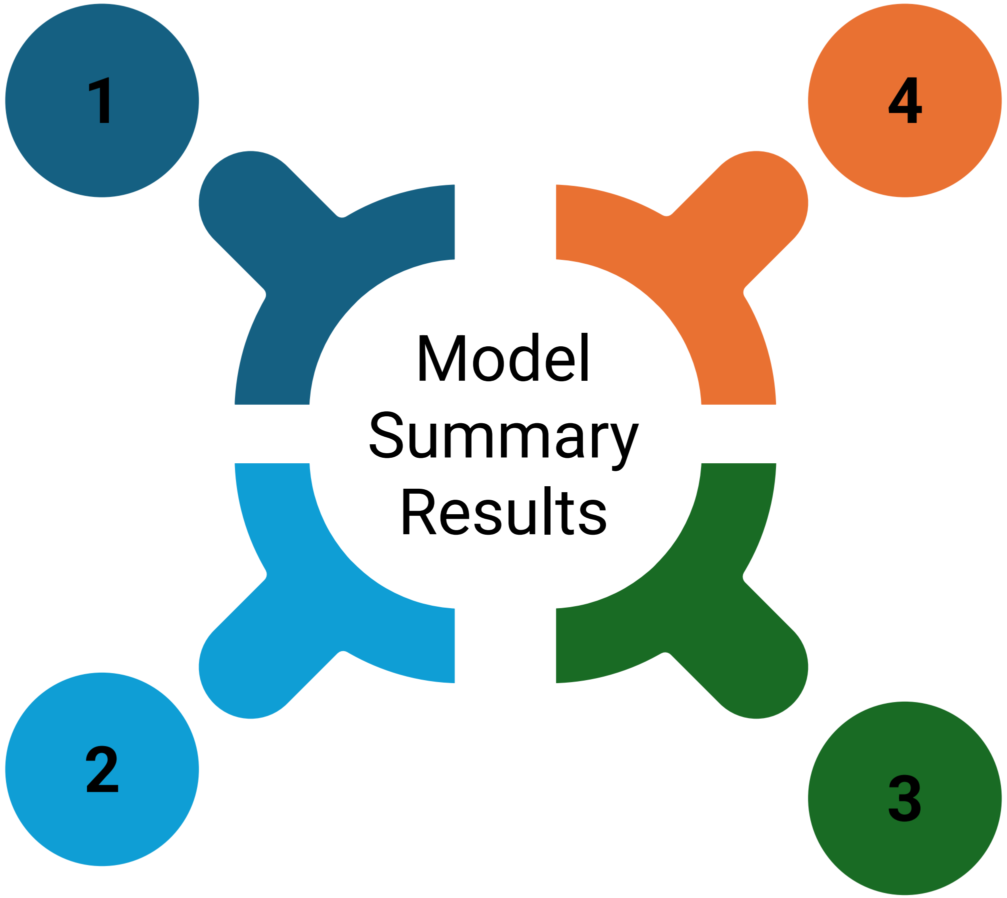
The model correctly predicts 79% of the non-purchasing cases

How often the model is actually correct

AUC-ROC

The model's value is 90% which can reasonably predict purchasing intent cases

Used as a baseline model for its simplicity



Macro evaluation

- 83% precison
- **82% Recall**
- 82% F1- score

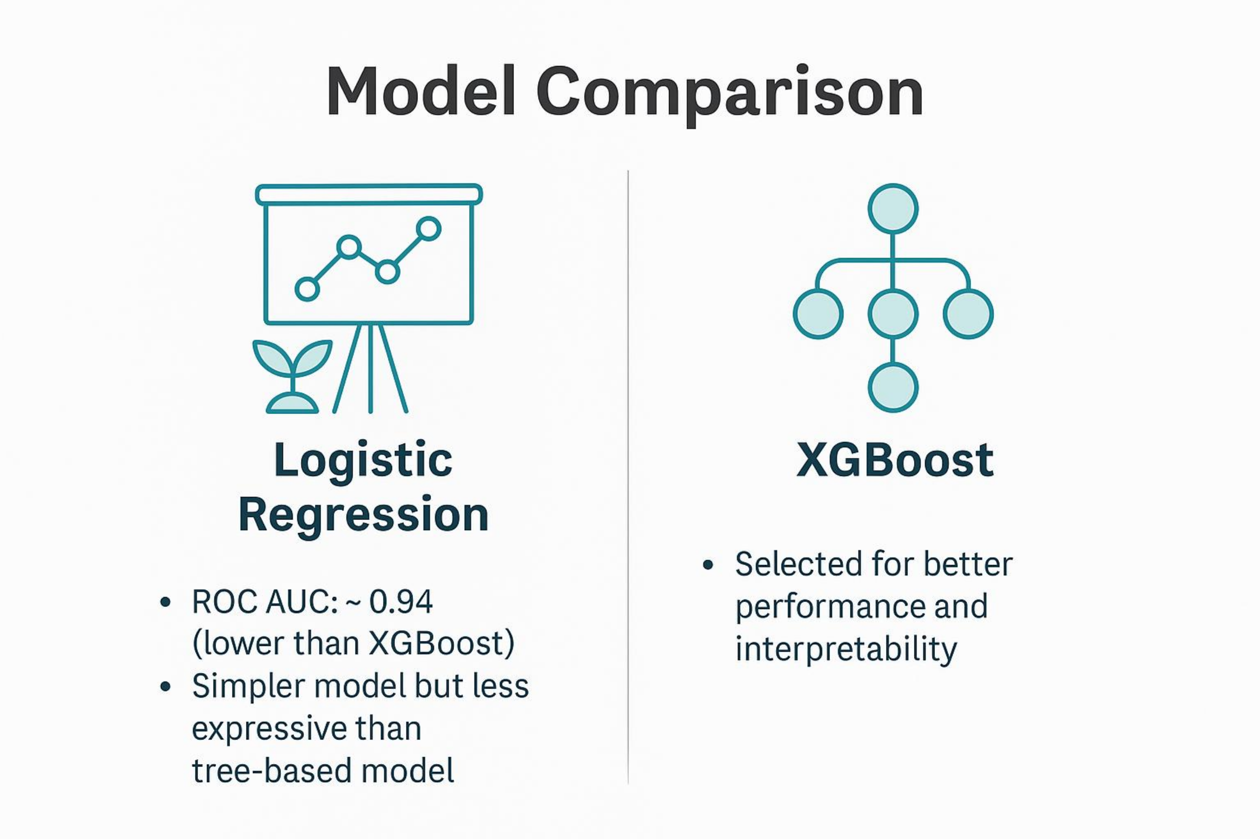
Depicts performance of each class equally

Weighted evaluation

- 83% precison
- **82% Recall**
- 82% F1- score

Depicts performance of each class using a weighted average

Classification Model – Model Comparison Logistic Regression




```
Best parameters for Logistic Regression: {'C': 9.978862418509248}

Logistic Regression Classification Report:
              precision    recall  f1-score   support

           0       0.79      0.88      0.83      3127
           1       0.87      0.76      0.81      3127

    accuracy                           0.82      6254
   macro avg       0.83      0.82      0.82      6254
weighted avg       0.83      0.82      0.82      6254

Logistic Regression ROC AUC: 0.9001514502416565
```

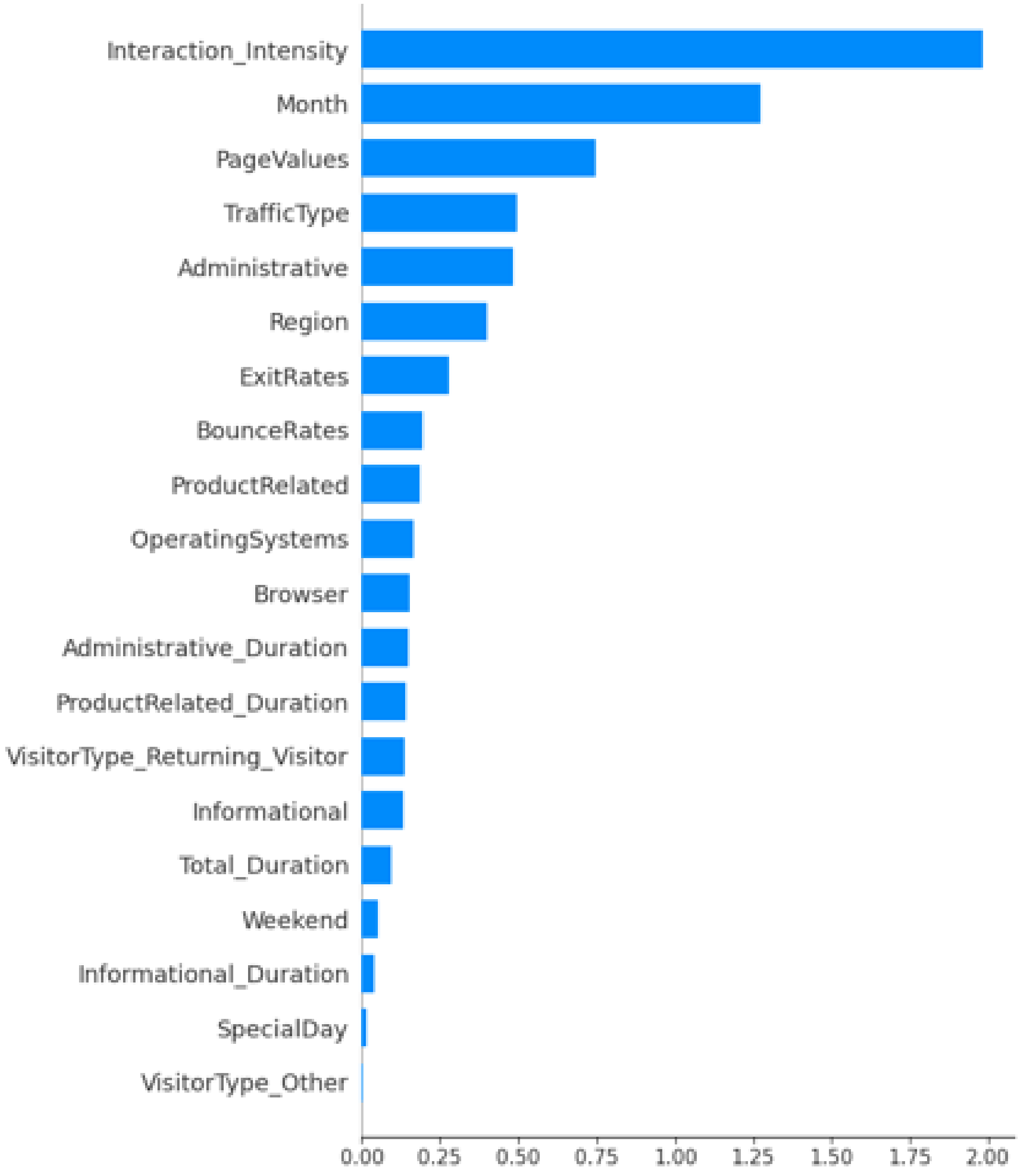
1. Logistic regression model is used as a Benchmark model; a production model should outperform this model ,our XGBoost model does, as the logistic regression model underperforms across all the metrics especially for recall

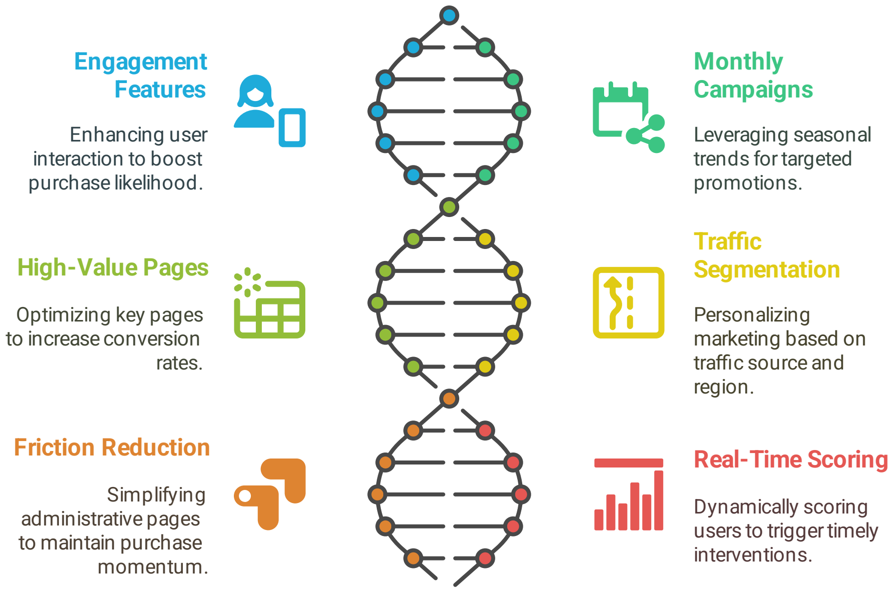
2. Logistic regression model chosen for benchmark because of its simplicity and considering the no free lunch theorem

3. XGBoost is the final predictive model deployed for purchasing intent prediction

Nexford University

Feature importance- What influences our customers to buy?





By understanding our customers' decision drivers—like interaction intensity and page value—Takealot can create a more innovative strategy to boost conversions, enhance personalisation, and maintain our competitive edge against Amazon's aggressive entry into the market.

Nexford University

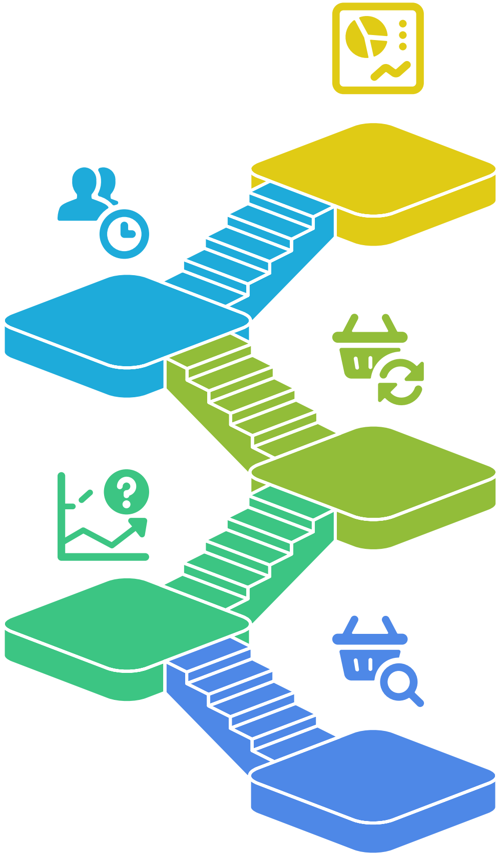
Business Value for Takealot

Key Benefits

- Enables customer-centric marketing and promotions
- Reduces revenue loss by identifying high-intent users
- Improves targeting efficiency via behavioral insights
- Supports Takealot's digital transformation goals

Supports Takealot's digital transformation goals

Business Value



1 Identify Patterns
Analyze customer data to understand purchasing behaviors.

2 Improve Conversion
Use predictive analytics to increase customer conversion rates.

3 Reduce Churn
Enhance shopping experience to retain customers.

4 Monitor Behavior
Implement real-time monitoring of customer interactions.

5 Track Performance
Evaluate segment performance to assess business impact.

Next steps and call action

FINAL RECOMMENDATIONS & STRATEGIC CALL TO ACTION

1 Leverage Behavioral Segmentation for Strategy Alignment

- Deploy clustering model to segment customers based on behavieral patterns.
- Prioritize Cluster 0 (high-intent users) for investment in experience, fargeting, and retention.
- Embed segmentation insights into CRM, Inventory planning, and dynamic pricing strategies.

2 Maximize Conversion with Predictive Intelligence

- Deploy the XGBoost model for real-time purchasing intent prediction.
- Use top predictive features (e.g, interaction intensity. Product Duration, Page Values) to guide personalized campaigns.
- Enable real-time marketing triggers such as cart reminders, special offers or chat prompts

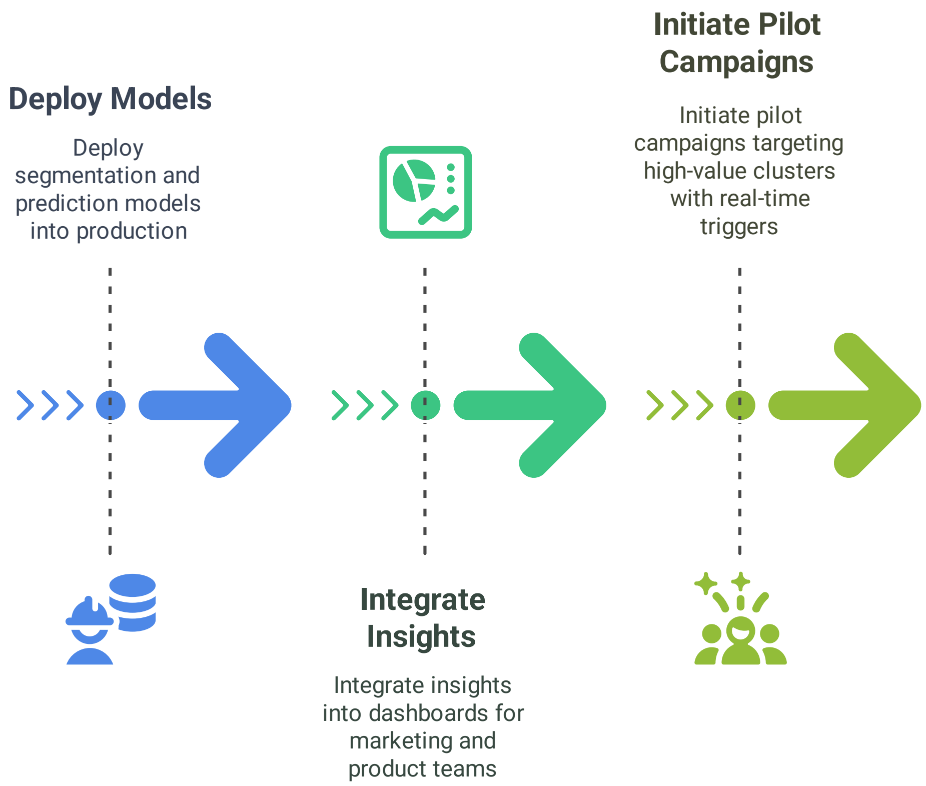
3 Enhance Marketing Effectiveness

- Retarget Cluster 1 (low-intent users) with personalized offers and hudges.
- Launch A/B tests for landing pages to failor engagement per segment.
- Use traffic source and region to personalize content and promotional timing

4 Optimize User Experience to Match Market Leaders

- Benchmark Takealot's UX against Amazon to identify gaps and opportunities

Immediate next steps



References

- Wikipedia. (2025). *Takealot.com*. Wikipedia. https://en.wikipedia.org/wiki/Takealot.com
- Sakar, C. & Kastro, Y. (2018). Online Shoppers Purchasing Intention Dataset [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5F88Q.
- Roy, A. (2018). Chapter 1 - Introduction to CRISP DM Framework for Data Science and Machine Learning | LinkedIn. https://www.linkedin.com/pulse/chapter-1-introduction-crisp-dm-framework-data-science-anshul-roy/
- Google. (2024). Classification on imbalanced data. TensorFlow Core. https://www.tensorflow.org/tutorials/structured_data/imbalanced_data

Thank you!

Nexford University