

# Behavioral Segmentation and Predictive Modeling of Purchasing Intent Among Takealot Online Shoppers

Leveraging Machine Learning Techniques for Enhanced Online Customer  
Purchasing Insights

**Name: M. Nsofu**

**Learner ID: 149050**

**Course: BAN6800**

**Date: 10<sup>th</sup> May 2025**

**Lecturer: Prof. Wanjiku**

**Assignment One**

**Nexford University**

## **1.0 Introduction:**

Takealot is an e-commerce company headquartered in Cape Town, South Africa, founded in 2011 by Kim Reid (Wikipedia, 2025). The company caters to South African online shoppers and has been the dominant player for the last 14 years. Amazon, the global online giant, recently entered the South African market (Fintechnews Africa, 2024). Amazon's strong brand reputation, advanced analytics, data-driven approach and technology threaten Takealot's market share and dominance. One of Takealot's weaknesses is the lack of diagnostic and predictive analytics capabilities that provide detailed insight into customer purchasing behaviour. Without actionable insights, Takealot cannot optimise marketing budgets, retain customers and protect market share. This project will use machine learning techniques to segment customer purchasing behavioural patterns and predict purchasing intent. This will culminate in improved customer engagement, an increase in conversion rates and customer retention through personalised recommendations.

## **2.0 Vision and Goal:**

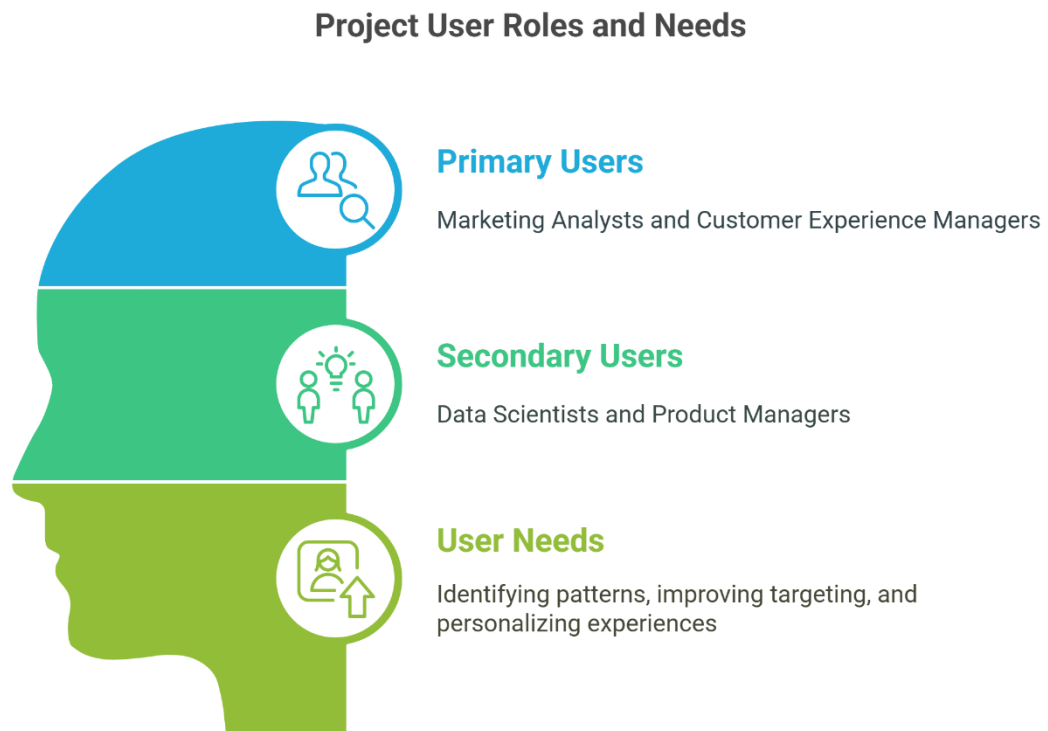
Despite Amazon's recent entry, Takealot aims to remain the market leader in the South African e-commerce industry. To remain the market leader, Takealot must better understand its customer behaviour. This project will use machine learning techniques to help Takealot find behavioural segments amongst its customers and predict purchasing intent. Modern analytics and machine learning techniques will help Takealot compete better against Amazon in South Africa.

### 3.0 Description of the users to be included in the project:

Figure 1 below captures the users and user needs for the project. The figure shows primary and secondary users coupled with the project's needs. The primary users' project needs include understanding online shopping behaviour, targeting relevant campaigns and predicting which online shoppers are high-conversion users.

**Figure 1**

Description of the users to be included in the project.

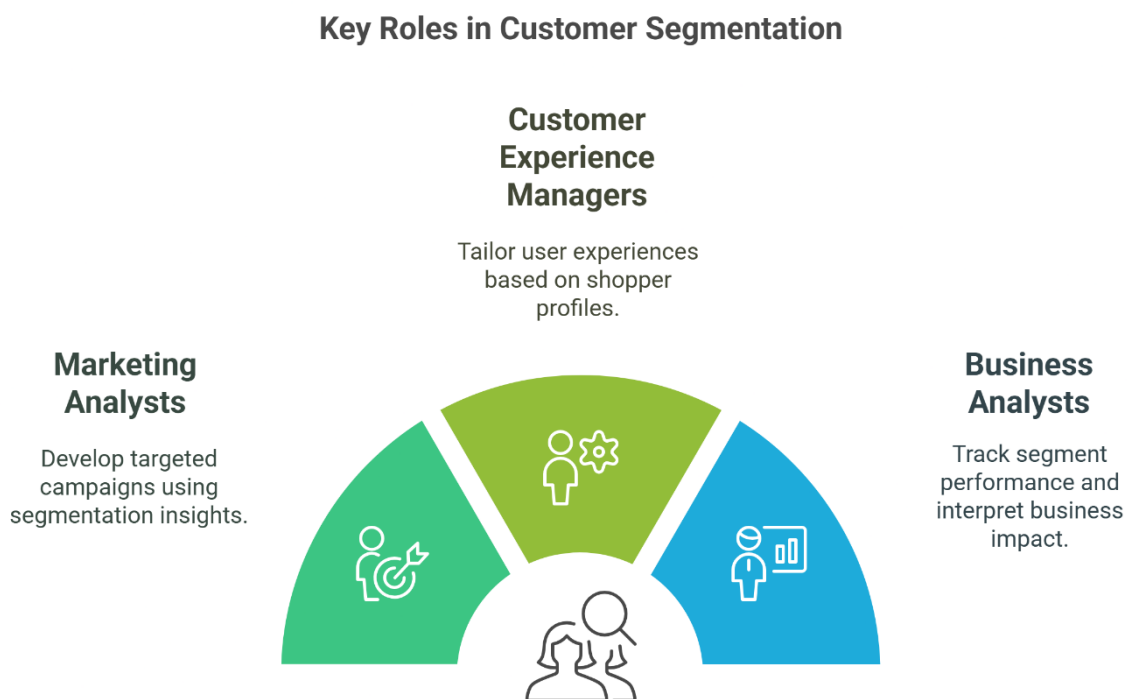


Note: Created by the Author (2025).

As illustrated in Figure 2, Takealot's primary users are those who will use customer segmentation insights. They include customer experience managers, business analysts, and marketing analysts.

**Figure 2**

Key roles in Customer Segmentation.

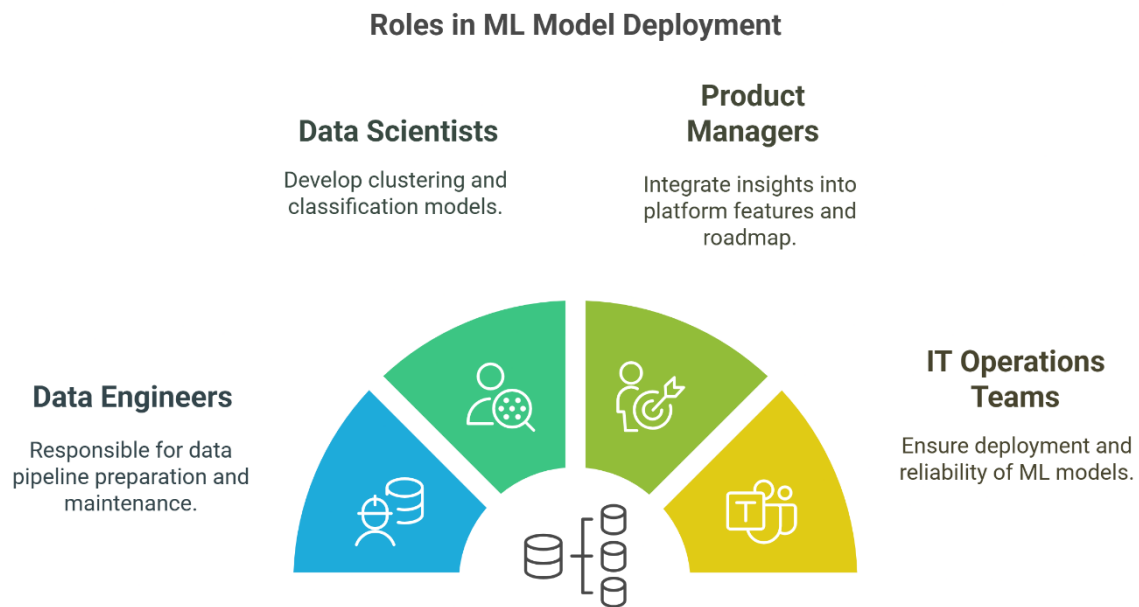


Note: Created by the Author (2025).

The secondary users are responsible for developing, deploying, and maintaining the solution for Takealot's primary users. They include Data Scientists, Data Engineers, Product Managers, and IT Operations teams. Figure 3 depicts these role players.

**Figure 3**

Key roles in ML Model Deployment.



Note: Created by the Author (2025).

#### **4.0 Project Stakeholders:**

- **Internal Stakeholders:**

- ***C-Suite (CEO, CMO and CIO):*** Require actionable insights to guide tactical and strategic decisions and ensure alignment with Takealot's goals.
- ***Project Sponsors (CMO):*** They will approve resources and promote the project across the company.
- ***Marketing & Business Analysts:*** End users of the solution.
- ***Data Engineers, Data Scientists and IT Operations:*** Develop, deploy and maintain the solution.

- **External Stakeholders:**

- **Online Shoppers:** Beneficiaries of personalised experiences from Takealot.
- **Regulatory Officers:** Ensure compliance with regulations like GDPR or similar, depending on the jurisdiction.

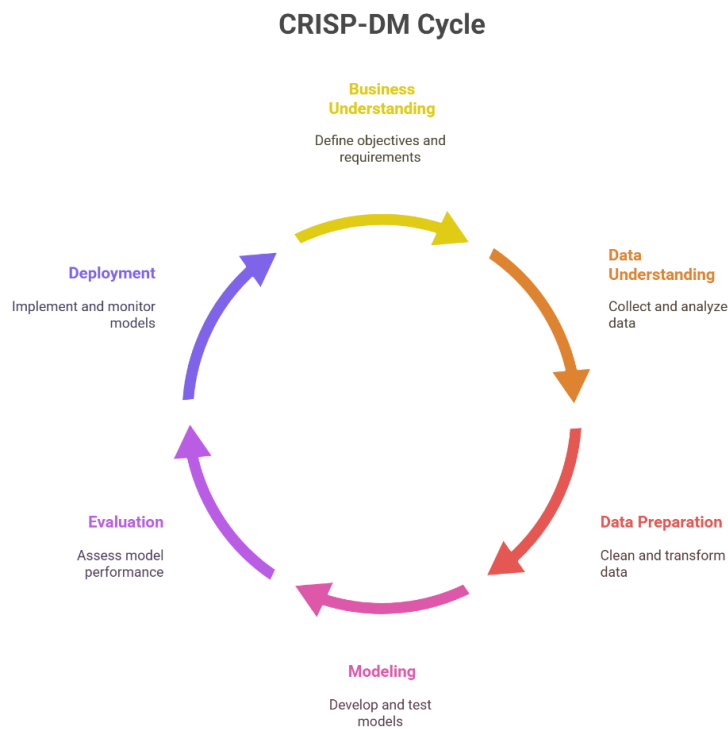
## **5.0 Overview of the Product to be Developed:**

The digital artefact will consist of a machine learning pipeline that performs two key actions: Clustering and Classification. Clustering will provide insights into customer behaviour segmentation based on online behaviour, whilst classification will be used to predict the likelihood of purchase. These ML engines will be integrated into an internal dashboard so primary users can understand insights. Python and R technology stacks will be used to build the solution, which is expected to be modular and scalable for the future. The project will be developed using the CRISP-DM process (Roy, 2018), shown in Figure 4.

In summary, the solution will incorporate descriptive analytics, diagnostic analytics and elements of predictive analytics to help Takealot minimise churn risks, optimise marketing and advertising budgets and improve overall return on investment by predicting which customers are likely to purchase.

**Figure 4**

## CRISP-DM Process



Note: Created by the Author (2025).

### 6.0 Features of the Product:

- Dimensionality Reduction (e.g. PCA, t-SNE, etc.).
- Clustering Algorithms (e.g. K-Means, DBSCAN, etc.).
- Ensemble of Classification Algorithms (e.g. XGBoost, Logistic Regression, etc.).
- Data Preprocessing and Feature Engineering.
- Exploratory Data Analysis.

- Export Functionality for Marketing Campaign.
- Real-time or batch scoring capabilities for new customer data.
- Nice to have: a forecasting module.

## 7.0 Product Requirements:

- **Dataset:** Online shoppers' purchasing data.
- **Tools:** Python, Pandas, Scikit-learn or PyTorch, R, Tidyverse, Tidymodels, MLR3, Shiny or Streamlit.
- **Infrastructure:** Cloud-based deployment using technologies such as Streamlit or Shiny.

## 8.0 Expected Risks and Constraints:

Anticipated risks in developing the product include an imbalanced dataset, which could affect the models' performance. Overfitting and potential issues with feature interpretability could also occur. Data quality issues must be dealt with during the product's development. Furthermore, ethical concerns such as privacy and algorithm bias could arise, considering that user data is being used in the ML models. Finally, there could be inertia in adopting the solution, especially among non-technical stakeholders.

## 9.0 Documentation Requirements:

The code base will be well documented following best practices such as PEP8 (Rossum, Warsaw, & Coghlan, 2013) and the Tidyverse style guide (The tidyverse team, n.d.). A data dictionary will be provided to ensure no ambiguity. Throughout the project, version control shall be maintained using git. Finally, model evaluation reports will also be documented.



## 6. References

Wikipedia. (2025). *Takealot.com*. Wikipedia. <https://en.wikipedia.org/wiki/Takealot.com>

Fintechnews Africa. (2024). *Amazon's Entry into South Africa Set to Shake up E-Commerce Landscape*. Fintech News Africa. <https://fintechnews.africa/44008/fintech-south-africa/amazons-entry-into-south-africa-set-to-shake-up-e-commerce-landscape/>

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/>

Rossum, G. van, Warsaw, B., & Coghlan, A. (2013). *PEP 8 – Style Guide for Python Code*. Peps.Python.Org. <https://peps.python.org/pep-0008/>

The tidyverse team. (n.d.). *Tidyverse style guide*. Retrieved May 10, 2025, from <https://style.tidyverse.org/>