**Ad Hoc Analysis of E-Commerce User Behavior and Session Data**

*To analyze user behavior and session data, enforce business rules, address data inconsistencies, and provide actionable insights to improve engagement and conversion rates.*

**Prepared By**

[Dat Quang Nguyen] [Toronto - ON]  
[Business Analytics student – Humber College]  
[curtis.nguyen2305@gmail.com]

**I. Introduction**

This project aims to evaluate user behavior and session data within an e-commerce environment to identify actionable insights for enhancing engagement and increasing conversion rates. By enforcing business rules and addressing key data inconsistencies, the analysis seeks to uncover patterns in user behavior, highlight areas for targeted interventions, and provide recommendations for improving overall business performance.

**II. Executive Summary**

This document outlines a comprehensive analysis of user behavior and session data, guided by predefined business rules and focused on addressing misalignments with business scenarios. Key findings and actions include:

1. **Business Rules and Data Issues:**
   * Established rules for traffic sources, device usage, action sequences, timestamps, financial consistency, and session-user relationships to ensure data integrity.
   * Identified critical issues, such as invalid campaign attributions, skipped action sequences, and inconsistent user or product IDs, with corresponding treatment plans to rectify these anomalies.
2. **Behavioral Insights:**
   * Analysis revealed two main categories of non-purchasing users:
     + **Highly Engaged Users**: Strong purchase intent but did not complete transactions.
     + **Interested but Hesitant Users**: Moderate purchase intent, requiring interventions to overcome hesitation.
   * Individual **user\_id** emerged as the **most influential feature** in predicting purchase behavior, underscoring the importance of personalized strategies.
3. **Prioritization of Interventions:**
   * Users were ranked based on engagement level and purchase intent to allocate resources effectively.
   * Behavioral patterns among the top 70% of prioritized users were categorized, revealing significant opportunities to target:
     + Highly Engaged Sessions for immediate conversion.
     + Interested but Hesitant Sessions to address hesitation.
     + Broader strategies for Low-Value Sessions and undefined behaviors.
4. **Actionable Recommendations:**
   * Highly Engaged Users: Implement personalized marketing campaigns and streamlined checkout processes.
   * Interested but Hesitant Users: Offer product comparisons, tailored recommendations, and promotional incentives.
   * Low-Value Sessions: Deploy broader engagement strategies, such as retargeting or educational content to drive awareness.

By addressing these findings and implementing tailored interventions, this project provides a clear roadmap for enhancing user engagement and maximizing conversion rates, ultimately driving business growth.

**III. Business Rules and Problems Document**

**A. Business Rules**

1. **Traffic Source Rules:**
   * source = "Direct" must not have a non-null reference\_link.
   * source = "Email" or "Referral" must have a valid reference\_link.
   * Ad spend (ads\_spent) should not occur when source = "Direct".
2. **Device Rules:**
   * Valid devices include "Desktop," "Mobile," and "Tablet." Invalid devices include "Smart TV" or "Gaming Console."
   * Actions like "Add to Cart" or "Purchase" should not originate from invalid devices.
   * action\_detail = "Physical Store Visit" must not occur on "Desktop" or "Mobile."
3. **Reference\_link rules:**
4. Reference\_link must only apply to ads traffic and it’s the crucial element determining a traffic is from ads or not**Action Sequence Rules:**
   * Intermediate actions, such as "Add to Cart," should not be skipped before "Purchase."
5. **Time and Session Rules:**
   * Timestamps (time) must be sequential within a session.
   * Future dates (e.g., "2050-01-01") or invalid times are not allowed.
   * Sessions with actions logged at improbable times (e.g., midnight purchases for business hours) should be flagged.
   * A session (session\_id) should not involve multiple devices unless cross-device tracking is enabled.
6. **Financial Consistency Rules:**
   * Discounts (discount\_amount) must only apply to sessions with relevant actions (e.g., "Purchase") and in ads traffic (reference\_link non null)
   * Ads spent must only apply to ads traffic (reference\_link nun null)
7. **User and Session Rules:**
   * Each user\_id must be unique and numeric.
   * A session (session\_id) must belong to only one user\_id.

**B. Problems Found Misaligned with Business Scenario**

1. **Traffic Source Contradictions:** 
   * source = "Direct" with a non-null reference\_link.
   * source = "Email" or "Referral" without associated campaigns or links in reference\_link.
2. **Invalid Campaign Attribution:** 
   * Ad spend (ads\_spent) assigned to sessions with no incoming traffic source (e.g., source = "Direct").

* Treatment plan: set to 0

1. **Action Detail Inconsistencies:** 
   * Missing intermediate actions (e.g., "Add to Cart" skipped before a "Purchase").

* Treatment plan: for later analysis

1. **Financial and Discount Problems:**
   * discount\_amount applied to sessions without purchases or relevant actions.

* Treatment plan: set to 0

1. **Inconsistency of Product ID in Product Intent, Non Product Intent sessions**

* Session linked to NaN product\_id when there are ‘Product View’, ‘Add to Cart’, ‘Purchase’ actions (product intent)
* Treatment plan: impute NaN product to the product\_id that an user engage the most (Mode)
* Session linked to non NaN product\_id when there are ‘Page View’ action (non product intent)
* Treatment plan: set to Null

1. **Format Anomalies:**
   * Sessions linked to invalid user\_id, product\_id formats (e.g., generic IDs like "user\_8126" or non-existent users).

* Treatment plan: Trim the ‘user\_’ part for user\_id and ‘Unknown’ for invalid product\_id

**IV. Analysis**

1. **Rise in Sessions and Conversion Rate (01 Dec to 05 Dec)**

A graph with a line going up

Description automatically generatedA graph with a line going up

Description automatically generated

Between **01 Dec and 05 Dec**, both the **number of sessions** and the **conversion rate** increased, suggesting overall improved user engagement and purchasing activity during this period. However, a closer look at the data from **02 Dec to 03 Dec** reveals a disproportionate growth rate between the two metrics:

* **Number of sessions increased significantly**, but **conversion rate growth lagged behind**.
* This discrepancy indicates a rise in sessions that did **not result in purchases**.

**Action Point**: Investigate the root cause of the mismatch between increased sessions and stagnant conversion rates.

**2.** **Behavioral Analysis of Sessions Not Leading to Purchases**

A screenshot of a graph

Description automatically generated

**Insights**:

A deeper examination of sessions not resulting in purchases reveals two prominent behavioral patterns:

**Highly Engaged Users**:

* + - These users spend a significant amount of time exploring products and exhibit strong purchase intent through frequent "add to cart" actions.

**Interested but Hesitant Users**:

* + - These users also spend substantial time browsing products but show only moderate purchase intent, with fewer "add to cart" actions compared to Highly Engaged Users.

**Next Steps**: To understand these behaviors better, we need to identify the **features** that have the most significant impact on distinguishing between purchasing and non-purchasing sessions.

**3. Identifying Features importance in Influence to Purchase Behavior**

A graph of a bar graph

Description automatically generated

To deepen the understanding of non-purchase behaviors, we analyzed various features to determine which has the **greatest impact** on **distinguishing** between **purchasing** and **non-purchasing** behavior. This was done by calculating the **coefficient of variation** (CV) of the **purchase-to-non-purchase ratio** for each feature.

**Findings**:

* Among all the features analyzed, **user\_id** emerged as the most important factor influencing purchasing behavior. This indicates that individual users demonstrate distinct patterns that strongly correlate with whether or not they make a purchase.
* By focusing on **user\_id**, we can identify unique behavioral tendencies and develop personalized strategies for intervention.

**Implications**:

* Individual users exhibit distinct purchasing behaviors, meaning targeted interventions can be tailored to their specific engagement patterns and intent.
* This finding directs efforts towards understanding and segmenting users based on their unique behaviors, rather than relying on broader session or demographic features.

**Next Steps**:

* Analyze user-specific engagement patterns and purchase history to categorize behaviors more effectively.
* Prioritize high-impact users for targeted actions based on their engagement and purchase intent metrics.

**4. Prioritizing Users for Interventions**

A graph with a red line

Description automatically generated

To effectively allocate resources and prioritize users for interventions:

* Rank users based on:

**Engagement Level**: Time spent on the website.

**Purchase Intent**: Number of "add to cart" actions.

* Users with both higher engagement and intent scores should be prioritized, as they are more likely to convert with targeted efforts.

**Action Point**: Point out the behaviors from high prioritized users

**5. Behavioral Patterns Among Top Priority Users**

A graph with blue squares

Description automatically generatedAnalyzing the top **70% of users by priority score**, the distribution of behaviors reveals the following insights:

* **Highly Engaged Sessions**: **1,204 users**.  
  These users exhibit significant engagement, spending substantial time browsing and adding multiple items to their cart, indicating strong purchase intent.
* **Interested but Hesitant Users**: **312 users**.  
  These users also spend considerable time browsing but demonstrate moderate purchase intent with fewer "add to cart" actions.
* **Low-Value Sessions**: **270 users**.  
  These sessions are characterized by minimal engagement and low purchase intent, contributing less to overall conversion opportunities.
* **Other Sessions**: **499 users**.  
  This group includes users with behaviors that do not fall into the predefined categories, potentially indicating inconsistent or undefined patterns of engagement.

**Implications**:

* **Highly Engaged Sessions** should be the primary focus for targeted interventions, as these users are closest to making a purchase.
* **Interested but Hesitant Users** could benefit from strategies aimed at reducing hesitation, such as offering product comparisons, personalized recommendations, or limited-time promotions.
* **Low-Value Sessions** may require broader engagement strategies, such as retargeting campaigns or incentives to explore more products.

**Action Points**:

* Develop personalized marketing strategies for **Highly Engaged Sessions** to convert them into buyers.
* Identify and address potential barriers for **Interested but Hesitant Users**, such as unclear pricing or insufficient product information.
* Investigate patterns in **Other Sessions** to determine if new behavior categories need to be defined for further analysis.

**V. Summary:**

This project evaluated user behavior and session data in an e-commerce environment to enhance engagement and conversions. It enforced business rules to ensure data integrity, identified **user\_id** as the most impactful feature influencing purchasing behavior, and prioritized high-intent users for targeted interventions. These insights enable actionable strategies to improve user experience and drive business performance.