1. Dataset Overview

This report examines the question: **What drives online purchase conversion?** It uses the Online Shoppers Purchasing Intention Dataset from the UCI Machine Learning Repository (Sakar and Kastro, 2018), containing 12,330 e-commerce browsing sessions recorded over one year. It includes **10 numerical variables and 8 categorical variables**, with **no missing values**. Each row represents a single session and includes a binary purchase outcome (**Revenue**), allowing purchase behaviour to be analysed at the session level.

The dataset also contains behavioural and contextual variables, including browsing duration of three types of pages (**xx_Duration**), month, and visitor type.

ExitRates captures the proportion of page views in a session that ended with an exit from the site, making it a useful indicator of drop-off behaviour.

Due to data privacy restrictions, even though **Region** is provided only as a coded variable and cannot be matched to specific geographic areas, it identifies the visitor's geographic market, allowing the analysis to consider whether purchase patterns vary across locations.

We also derive a **ConversionRate** measure to represent the purchase conversion level of each group; the formula is:

$$\text{Conversion Rate} = \frac{\text{Number of sessions with Revenue=True}}{\text{Total number of sessions in the group}}$$

2. Research Structure

To answer the main question, the analysis is organised into three parts:

1. **Conversion Context**, which identifies where conversion is more likely to occur
2. **Browsing Behaviour**, which examines duration patterns and drop-off signals in purchasing and non-purchasing sessions
3. **Segment Differences**, which compares these patterns across visitor groups.

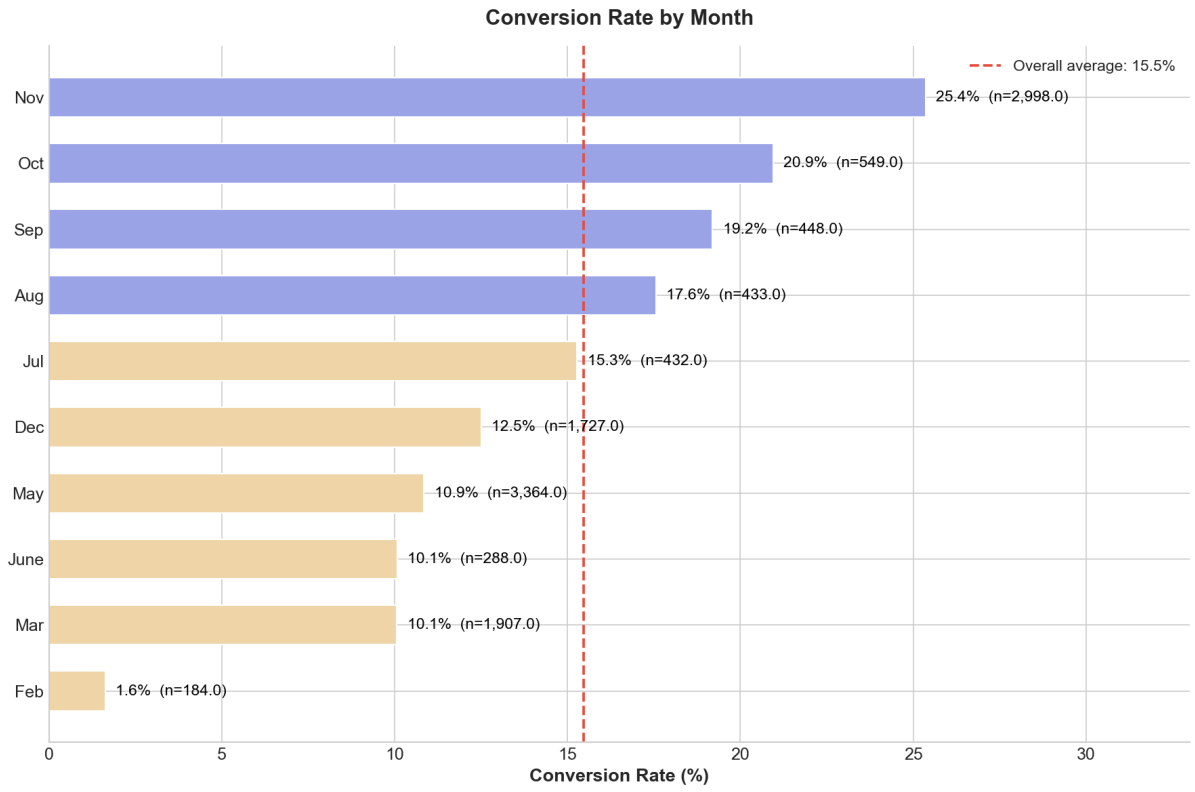
This structure links the dataset clearly to the research question and supports practical recommendations for the marketing team.

3. Analysis

Module 1. Conversion Context

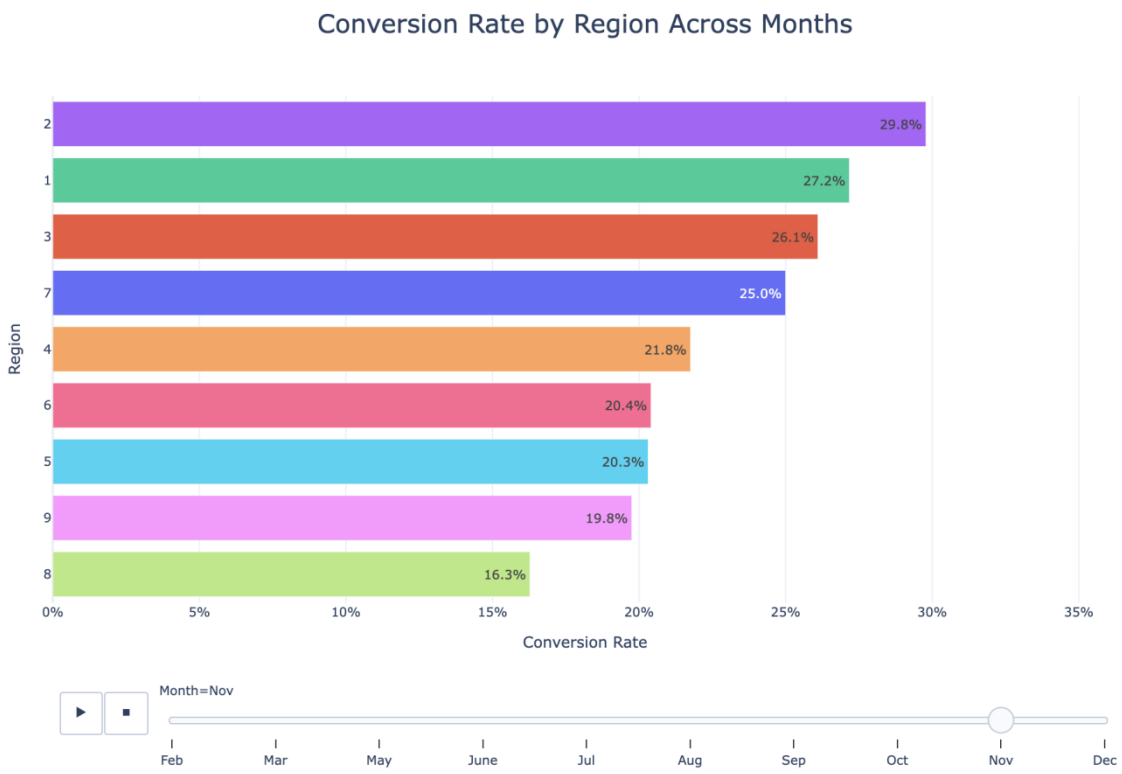
RQ_1: In what session contexts is conversion more likely to occur?

Figure 1. Conversion Rate by Month



The horizontal bar chart shows substantial monthly differences in conversion rate, with November, October and September performing well above the overall average, while February and several earlier months remain well below it. This suggests that month is an important session context associated with purchase conversion.

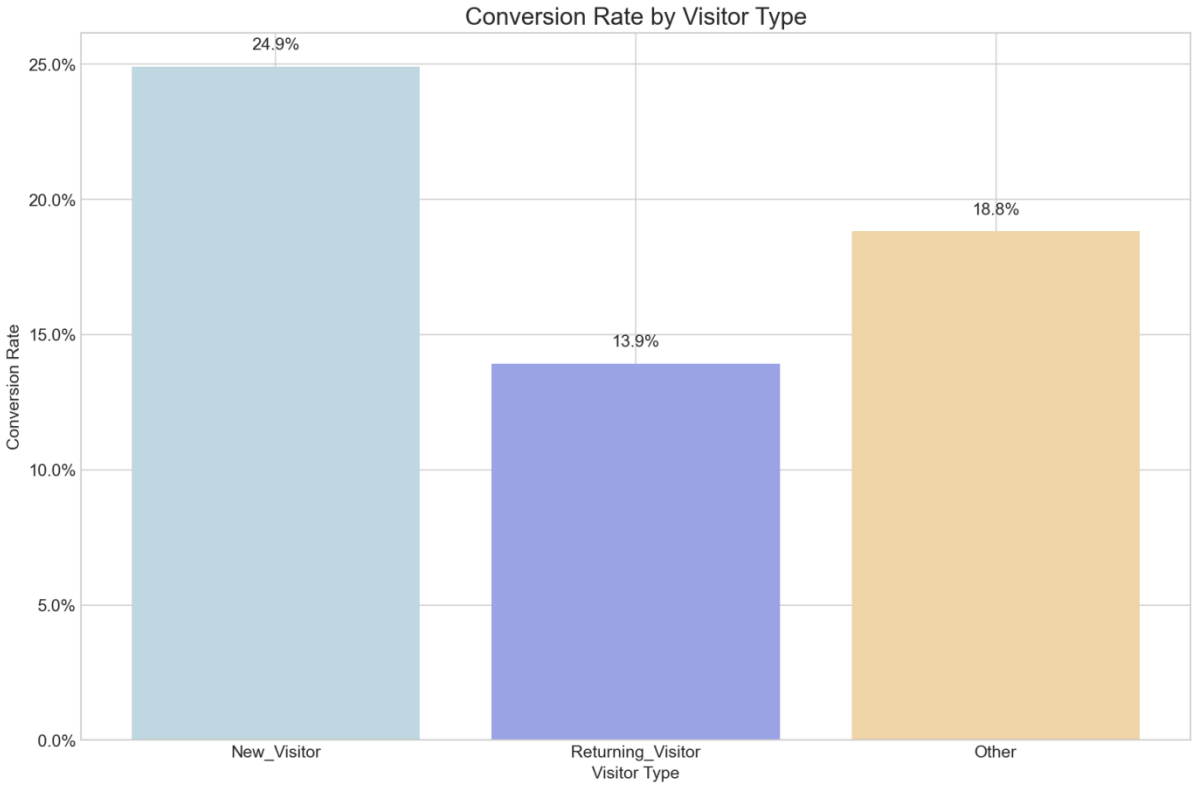
Figure 2. Conversion Rate by Region Across Months



Link：https://poetic-cendol-94e7b0.netlify.app/

The animated ranked bar chart highlights clear ranking shifts across months, showing that regions do not maintain the same relative conversion performance over time. This suggests that regional differences in conversion are time-sensitive rather than fixed.

Figure 3. Conversion Rate by Visitor Type



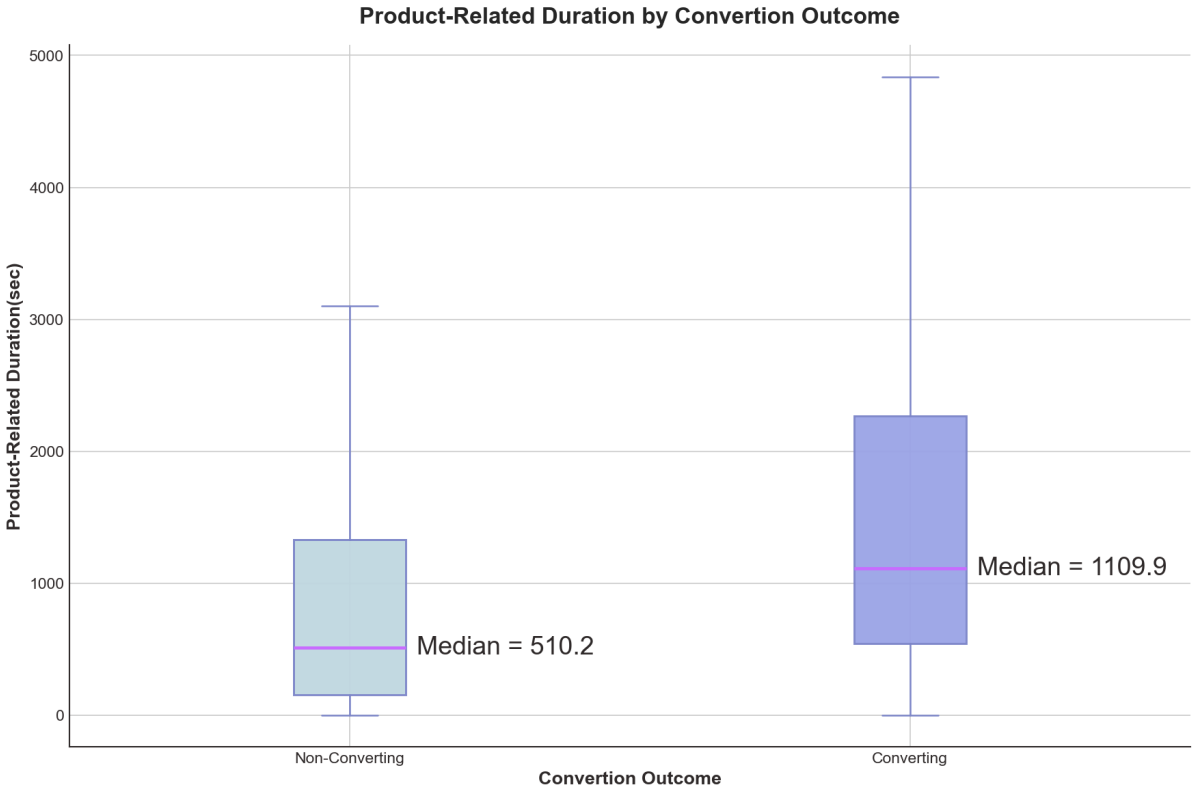
Conversion also varies by visitor type. New visitors show the highest conversion rate (24.9%), while returning visitors convert at a substantially lower rate (13.9%), suggesting that visitor type may reflect differences in purchase intent rather than familiarity alone. The "Other" category is excluded from later analysis because its sample size is too small for robust interpretation.

Overall, these findings show that conversion is not randomly distributed across sessions, but already differs systematically by entry context, including month, region, and visitor type.

Module 2. Browsing Behaviours

$RQ_{2.1}$: How do converting and non-converting sessions differ in time-based engagement patterns?

Figure 4. Product-Related Duration by Conversion Outcome

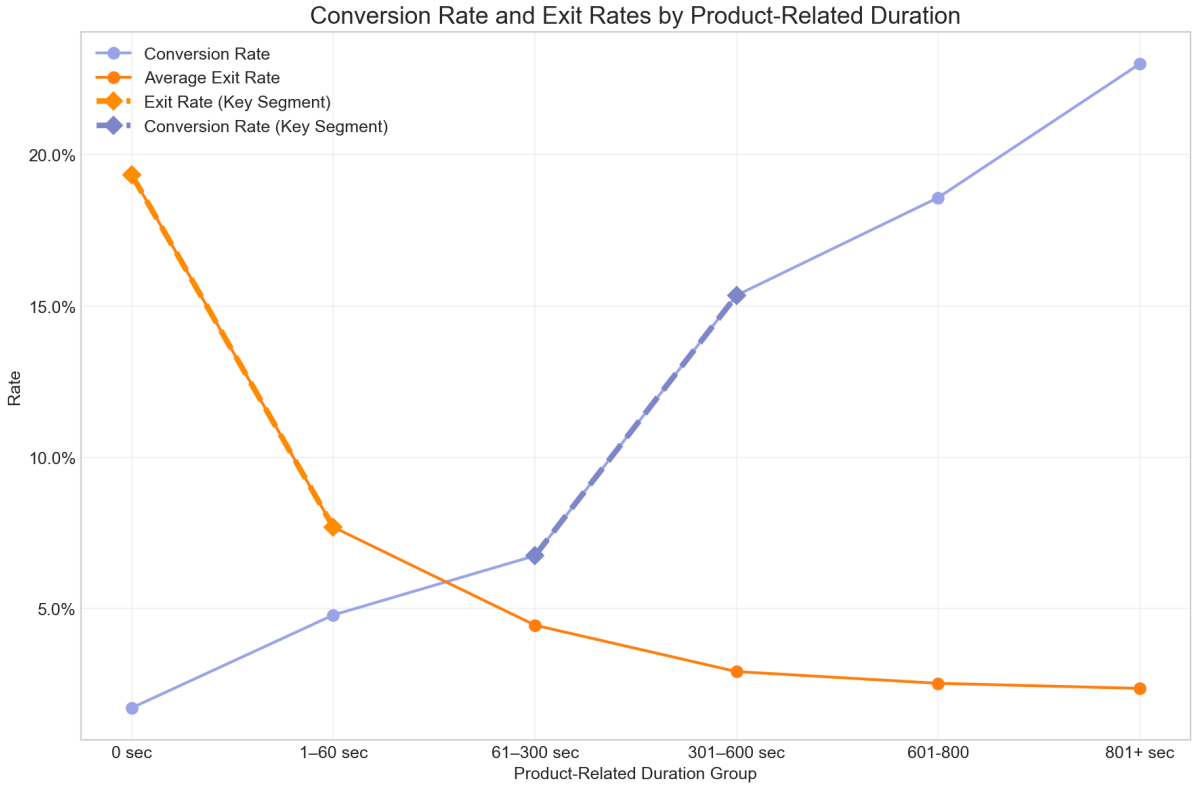


Among the three page-duration variables, **ProductRelated_Duration** shows the strongest contrast between converting and non-converting sessions; therefore, this section focuses only on this variable. Also, outliers were removed during data preprocessing to ensure distributional clarity.

The boxplot shows that converting sessions have a substantially higher median Product-Related Duration than non-converting sessions, as well as a higher overall distribution. This suggests that converting sessions exhibit deeper product browsing and stronger engagement than non-converting sessions.

$RQ_{2.2}$: In what browsing contexts is drop-off most pronounced?

Figure 5. Conversion Rate and Exit Rates by Product-Related Duration

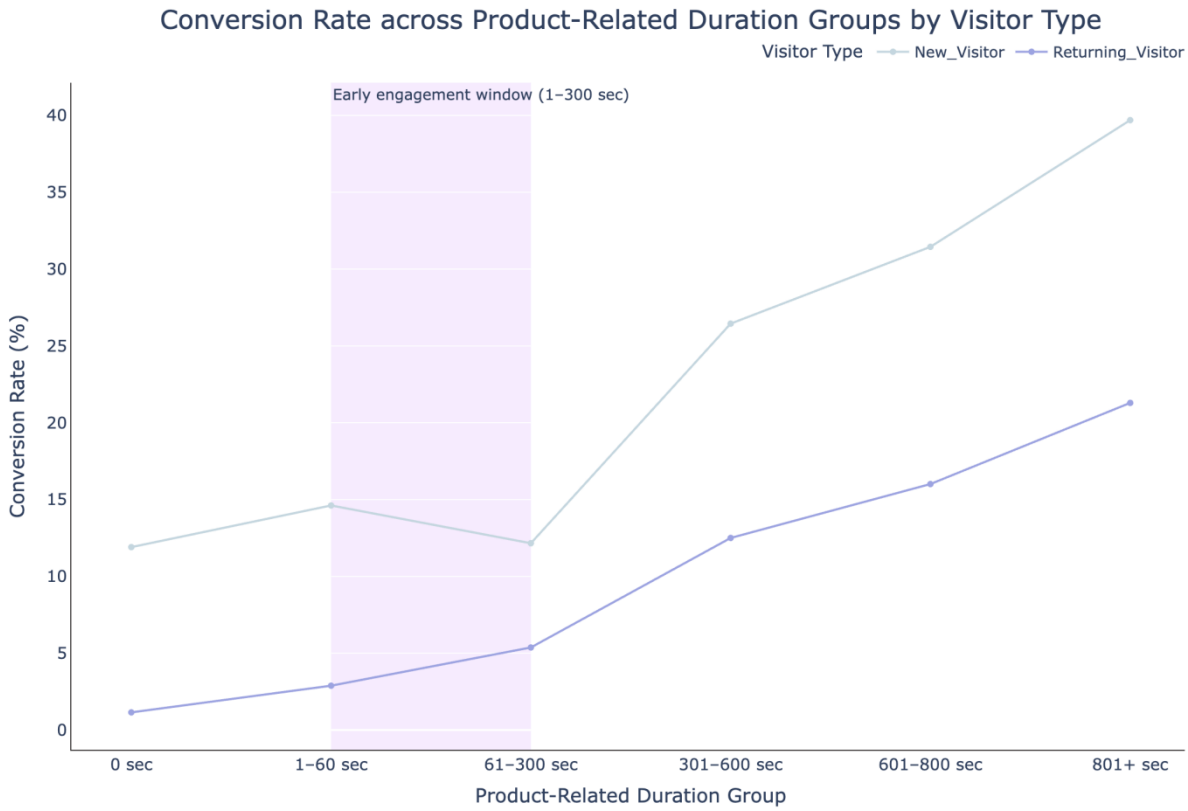


The line chart shows the changing trends of **ConversionRate** and **ExitRates** under different intervals of **ProductRelated_Duration**. For Exit Rate, it declines sharply within the 0-60 sec range, indicating that drop-off risk is concentrated in the early stage of product-related browsing; beyond 300 sec, the decline becomes much slower. Meanwhile, Conversion Rate rises most rapidly between 61-600 sec, suggesting that this is the key period when sessions begin to shift towards purchase.

Module 3. Segment Differences

RQ_3: Do the same browsing patterns carry different implications across visitor segments?

Figure 6. Conversion Rate across Product-Related Duration Groups by Visitor Type



https://jovial-raindrop-13a5b2.netlify.app/

The line chart shows that Conversion Rate generally rise with longer Product-Related Duration for both new and returning visitors. However, the pattern differs across segments. Specifically, Conversion Rate for New Visitors declines slightly, while Conversion Rate for Returning Visitors continues to increase between 1-300 sec. Therefore, even in the same browsing patterns, different segments of visitors have different conversion outcomes.

4. Recommendations

1) Prioritise budget allocation toward high-conversion months and high-intent users

Marketing resources should be concentrated in seasonally stronger conversion periods, particularly the late-year months, and directed toward new visitors with clear product interest, as these sessions are more likely to convert.

2) Optimise product pages to support purchase decisions

Because longer product-related browsing is associated with higher conversion, product pages should be improved to make that time more effective. Key actions include better product images, clearer user reviews, more scannable selling points, and easier access to stock, delivery, and pricing information. Purchase buttons should also be more visually prominent to reduce hesitation and help turn product interest into conversion.

3) Use session-duration-based interventions to reduce early drop-off and capture the decision window

For sessions lasting less than 60 seconds, light-touch prompts such as personalised pop-ups or limited-time discounts can be used to intercept users who are at high risk of leaving early. For sessions between 61 and 600 seconds, stronger conversion signals, including more prominent purchase buttons, stock alerts, and customer reviews, can help capture users during the key decision window.