## **Online Shoppers Purchasing Intention Dataset**

**The raw data contains 12,330 rows and 17 columns.**

**The “Revenue” column is our target.**

1. **Continuous Variables:**Administrative Duration (time spent on administrative pages)

Informational Duration (time spent on informational pages)

Product-Related Duration (time spent on product-related pages)

Bounce Rates (percentage of single-page visits)

Exit Rates (percentage of exits from each page)

Page Values (value of each page view)

Special Day Indicator (indicates closeness to a special day)

1. **Non-Continuous Variables:**Month

Visitor Type (New Visitor, Returning Visitor, or Other)

Weekend Indicator (whether the visit occurred on a weekend)

Operating System

Browser

Region

Traffic Type

## **`**

The dataset consists of 10 numerical and 8 categorical attributes. Whereas, out of 12,330 sessions in the dataset, 84.5% (10,422) were negative class samples that did not end with shopping, and the rest (1908) were positive class samples ending with shopping. The Online Shoppers Purchasing Intention Dataset contains various features that capture user behavior on an e-commerce website. These features can help in predicting whether a visitor will make a purchase.

**Here are the key features:**

**Revenue (Target Variable)** – This is a binary variable indicating if the session ended with a purchase. Values are True (purchase made) or False (no purchase). This is the target variable for classification tasks.

**Behavioural Features:**

**Administrative** – The number of pages related to administrative functions visited during the session.

**Administrative Duration** – The time spent on administrative pages during the session.

**Informational** – The number of pages related to informational content visited during the session.

**Informational Duration** – The time spent on informational pages during the session.

**Product-Related** – The number of pages related to product listings or details visited during the session.

**Product-Related Duration** – The time spent on product-related pages during the session.

**Bounce Rates** – The percentage of single-page sessions (sessions where the user left the website from the landing page).

**Exit Rates** – The percentage of exits from each page.

**Page Values** – The average value attributed to a page based on transactions (i.e., the page's contribution to revenue).

**Special Day** – A continuous variable that indicates how close the session is to a special day, such as a holiday. Higher values indicate proximity to a special day.

**Categorical Features**

**Month** – The month in which the session occurred (e.g., Jan, Feb, Mar).

**Operating System** – The operating system used by the visitor.

**Browser** – The type of browser used during the session.

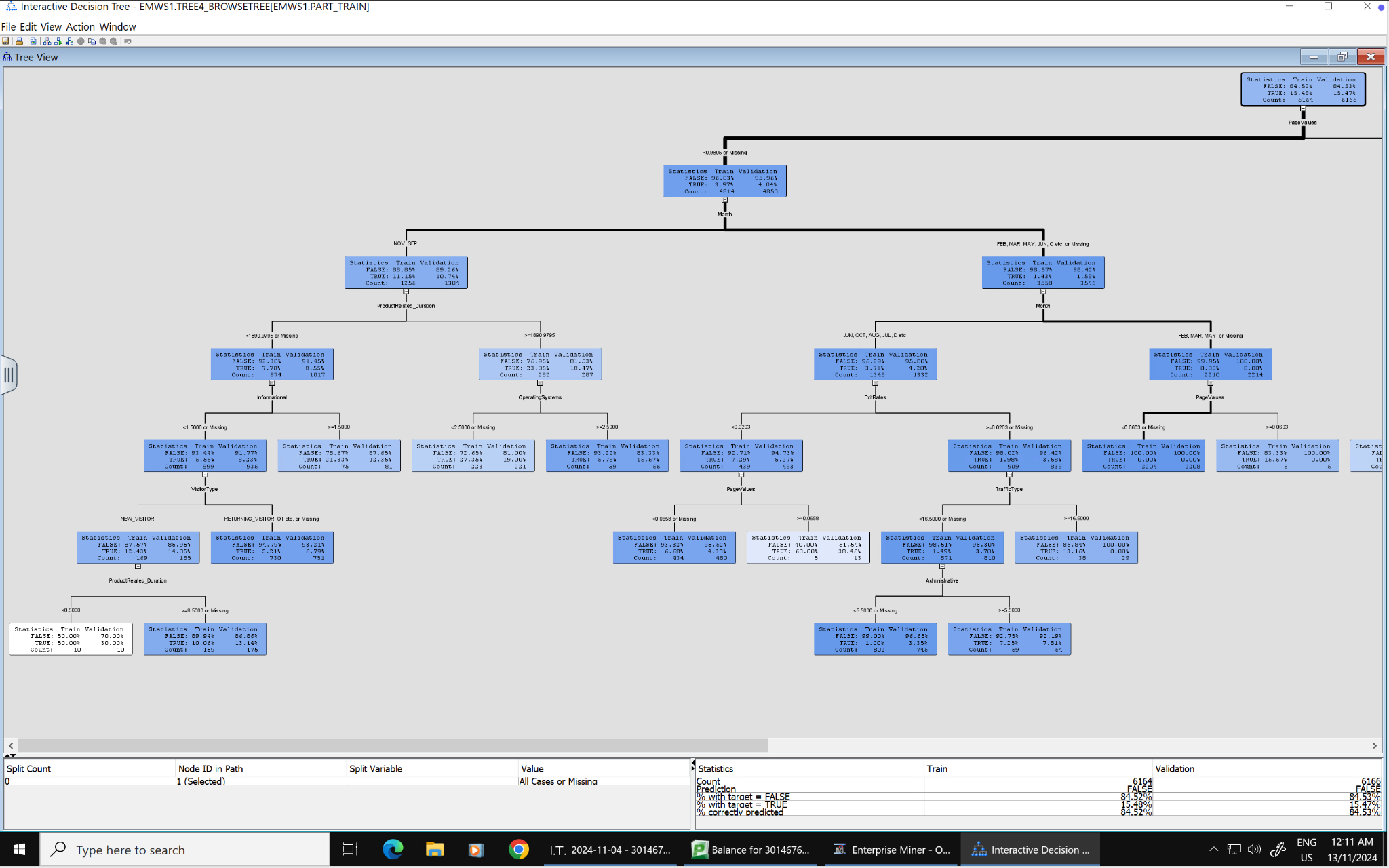
**Region** – The geographical region of the visitor.

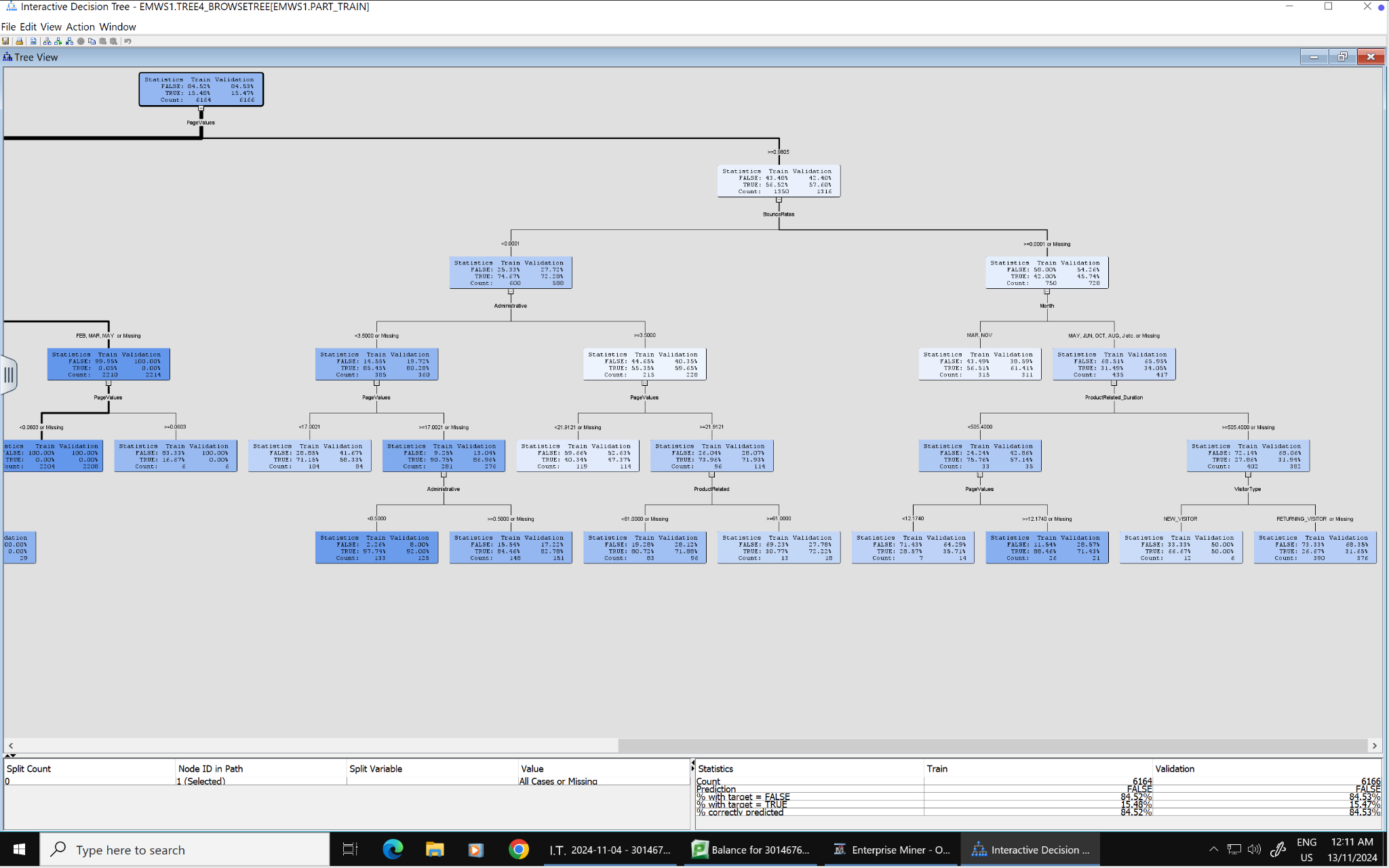
**Traffic Type** – The source of traffic, coded as different categories.

**Visitor Type** – The type of visitor, either Returning Visitor, New Visitor, or Other.

**Weekend** – A binary variable indicating if the session occurred on a weekend (True or False).

**Maximal Tree:**

****

****