

## Problem Statement

You are a data scientist at *Megakkart* the largest ecommerce platform in the country. Owing to the festive season *Megakkart* is planning a '**Super Festive Sale**' where it wants to achieve a GMV sales of **INR 500 crores**. You are being provided with the user data for the past one year on various parameters mentioned below. You need to identify whether the user on the web portal will make a purchase or not. The company plans to work on various hooks for the customers who are at a risk of not making a purchase. Your analysis and machine learning model will help the product team to plan these hooks.

## Data Description

"**Administrative**", "**Administrative Duration**", "**Informational**", "**Informational Duration**", "**Product Related**" and "**Product Related Duration**" represent the number of different types of pages visited by the visitor in that session and total time spent in each of these page categories. The values of these features are derived from the URL information of the pages visited by the user and updated in real time when a user takes an action, e.g. moving from one page to another.

The "**Bounce Rate**", "**Exit Rate**" and "**Page Value**" features represent the metrics measured by "Google Analytics" for each page in the e-commerce site.

The value of "**Bounce Rate**" feature for a web page refers to the percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session.

The value of "**Exit Rate**" feature for a specific web page is calculated as for all pageviews to the page, the percentage that were the last in the session.

The "**Page Value**" feature represents the average value for a web page that a user visited before completing an e-commerce transaction.

The "**Special Day**" feature indicates the closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with transaction. The value of this attribute is determined by considering the dynamics of e-commerce such as the duration between the order date and delivery date. For example, for Valentine's day, this value takes a nonzero value between February 2 and February 12, zero before and after this date unless it is close to another special day, and its maximum value of 1 on February 8.

The dataset also **includes operating system, browser, region, traffic type, visitor type** as returning or new visitor, a Boolean value indicating whether the date of the **visit is weekend**, and **month of the year**.