# Introduction

Variable Information:

• “Transaction” is the target variable. FALSE where no transaction occurred during a session, and TRUE where a transaction occurred.

• "Administrative", "Administrative\_Duration", "Informational", "Informational\_Duration", "ProductRelated" and "ProductRelated\_Duration" represent the number of different types of pages visited in a session and total time spent in each of these page categories.

• "Bounce Rate", "Exit Rate" and "Page Value" represent the metrics measured by "Google Analytics". The value of "Bounce Rate" refers to the percentage of visitors who entered the site from a specific page (landing page for a session) and then left ("bounced") without visiting any other page during that session. The value of "Exit Rate" is calculated as for all pageviews to a specific page (exit page for a session), the percentage that were the last in that session. The "Page Value" represents the average value for a web page that a user visited before landing on the goal page or completing an e-commerce transaction (or both) in the given session. Goal page is a page that e-commerce company wants visitors to reach during a session. This page can be a transaction page. Shopping cart pages often have high page values.

• "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 ecommerce 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. • “Month” indicates the month when session took place.

• “VisitorType” indicates the type of visitor for a session – Returning\_Visitor or New\_Visitor

• “Weekend” is FALSE where session did not occur on weekend, and TRUE where session occurred on weekend.

# Part 1: Random Forest

## Data Preparation

We started our data preparation by analyzing each variable from the provided dataset E-shop.csv. **Transaction** is our target variable and what we want to predict.

## Model Evaluation Strategy

## Model Building and Testing

## Identifying Best Model

## Generating Recommendations

# Part 2: PCA & K-Means Clustering

## PCA Implementation to visualize dataset

## Elbow Plot Creation

## K-Means Implementation