The dataset is from [Kaggle](https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction/data), containing information about customers of an e-commerce company. There are 20 columns of data in total. The following summarizes the variable names and descriptions in the dataset.

CustomerID (Categorical - Nominal): Unique identifier assigned to each customer. This variable serves as an identifier and is not used for analytical purposes other than uniquely identifying customers.

Churn (Categorical - Binary): Indicates whether the customer has churned or not. It's a binary categorical variable with values TRUE (churned) or FALSE (not churned).

Tenure (Numerical - Interval): Represents the duration of the customer's association with the business. It is a quantitative and continuous variable, indicating the length of the customer's tenure.

PreferredLoginDevice (Categorical - Nominal): Represents the preferred device for customer login. This is a categorical variable with different device categories.

CityTier (Categorical - Ordinal): Indicates the tier of the city where the customer is located. It's an ordinal categorical variable with different city tier levels.

WarehouseToHome (Numerical - Interval): Represents the distance from the warehouse to the customer's home. It is a quantitative and continuous variable.

PreferredPaymentMode (Categorical - Nominal): Specifies the preferred mode of payment chosen by the customer. It's a categorical variable with different payment mode categories.

Gender (Categorical - Nominal): Represents the gender of the customer. It's a categorical variable with two possible values: Male or Female.

HourSpendOnApp (Numerical - Interval): Indicates the number of hours the customer spends on the mobile application. This variable is quantitative and continuous.

NumberOfDeviceRegistered (Numerical - Interval): Represents the number of devices registered by the customer. It is a quantitative and discrete variable.

PreferedOrderCat (Categorical - Nominal): Indicates the preferred category for ordering. This is a categorical variable with different order category options.

SatisfactionScore (Numerical - Interval): Reflects the satisfaction score given by the customer. This variable is quantitative and continuous.

MaritalStatus (Categorical - Nominal): Represents the marital status of the customer. It's a categorical variable with values like Single, Married, etc.

NumberOfAddress (Numerical - Interval): Indicates the number of addresses associated with the customer. It is a quantitative and discrete variable.

Complain (Categorical - Binary): Indicates whether the customer has lodged a complaint. It's a binary categorical variable with values TRUE or FALSE.

OrderAmountHikeFromlastYear (Numerical - Interval): Represents the percentage increase in order amount from the last year. This variable is quantitative and continuous.

CouponUsed (Numerical - Interval): Indicates the number of coupons used by the customer. It is a quantitative and discrete variable.

OrderCount (Numerical - Interval): Represents the count of orders placed by the customer. It is a quantitative and discrete variable.

DaySinceLastOrder (Numerical - Interval): Represents the number of days since the customer's last order. It is a quantitative and continuous variable.

CashbackAmount (Numerical - Interval): Reflects the cashback amount received by the customer. This variable is quantitative and continuous.

**Project Overview:**

In this comprehensive data analytics initiative, the seamless collaboration between SAS Enterprise Miner and Talend Data Preparation plays a pivotal role in addressing distinct facets of the data workflow. SAS Enterprise Miner assumes a leadership role in advanced analytics and predictive modeling, providing a comprehensive suite of tools for efficient data exploration, meticulous preparation, and robust model building. Leveraging its scalability and seamless integration capabilities, SAS Enterprise Miner proves to be a well-suited environment for data scientists, especially when dealing with large datasets.

Complementing this advanced analytics powerhouse, Talend Data Preparation excels in data cleansing, transformation, and integration tasks. Boasting a user-friendly interface, Talend empowers business users to actively engage in the data preparation process, fostering collaboration between technical and non-technical teams. The synergistic strengths of these tools converge to establish an end-to-end data pipeline. Talend's proficiency in data integration seamlessly complements SAS Enterprise Miner's advanced analytics capabilities, resulting in a streamlined and highly effective data analytics workflow.

**Objective:**

The primary objective of this project is to predict customer churn rate by analyzing various factors. This involves harnessing the combined strengths of SAS Enterprise Miner and Talend Data Preparation to create a cohesive and scalable solution. The predictive modeling capabilities of SAS Enterprise Miner, coupled with Talend's efficiency in data integration, ensure not only the accuracy of churn rate predictions but also the seamless scalability and effectiveness of the entire project.