Dataset Description

The data provides historical customer transactions for a retailer, for purchases made on their e-

commerce platform between Dec 2010 to Dec 2011. Customers are uniquely identified by their

Customer ID (Col G).

Problem Statement

Assume that the company has been experiencing high customer churn and a reduction in repeat

customers.

Churn Definition (Target Variable definition)

For this problem, the churn should be defined as follows:

The training data should include all customers who have made a transaction between 1 st Dec 2010 to

31 st Aug 2011. Customers from this set who have NOT made any subsequent purchase in the period

Sep 2011 to Dec 2011 should be labeled as “Churn”. All other customers with purchases made

during this follow-up period should be labelled as “Not Churn”