**CA2 Data Analysis Report for Tools for Data Analytics**

**1. Data preparation:**

Data set is combined of 13 independent variables and 1 target variable. There are 2 c.. variables and 11 numeric variables

During data preparation, 13 variables have been divided into 23 variable. Variable Months as been divided into 10 independent variables. Variable Visitor type has been divided into 2 independent variables. By doing that we have all variables in numerical type and boolean type.

Boolean type are target variable; Transaction and independent variable; Weekend.

**2. Model Evaluation Strategy**

We took an approach of “minimising false positive”. We have decided for that approach for the following reason;

Let’s assume we have a web shop, as stated in an assignment.

Main question we asked our selves is what is worse for the company.

We came to an answer that it would be worse for the company if we predict that web shop would have for example 100 transactions and yet they achieved 50 transactions.

Why would that be worse? For the following reason: If company expect to have around 100 transaction they are counting on that profit, and yet they managed only 50 transactions. Thereby we consider we need to minimise false negative, to give as actuate prediction on transactions as possible.

**3. Model Building and Testing –**

Running test set with all features we came to this numbers: