**Reflections or Learning Outcomes**

**Learning outcomes**

1. Talend Data Preparation

* To perform data quality checks
* To utilise the features for remove duplicate column, change the data type and change data format
* To perform visualisation of the data

1. Talend Data Integration

* To integrate the data

1. SAS e-Miner

* To handle missing value
* To perform decision tree

**Changes faced and solutions**

**Talend Data Preparation**

1. Can not import the dataset directly to “preparation” in Talend Data Preparation

If import the dataset directly to “preparation” in Talend Data Preparation without import the dataset into the “dataset” feature, the quality bar unable to indicates the valid, empty and invalid fields.

To overcome this problem, I import the dataset into the “dataset” feature first. Select the dataset when add “preparation”. So that, the quality bar was functioned.

1. Maximum number of rows of 30,000

After add preparation, I found out that the data is limited to 30,000 rows while my dataset is more than that. I have manually reduced the dataset to 28,882 rows.

**SAS e-Miner**

1. ERROR: Maximum levels of 512 exceeded.

This error happened when I am handling missing value using “Impute”. My missing value is 51 but I still received the error messages.

To overcome this error, I amended the EM\_TRAIN\_MAXLEVELS to 20,000 instead of the value of 512.

A screenshot of a computer

Description automatically generated