**Preprocessing using Talend Data Preparation**

For the next preprocessing, I utilized Talend Data Preparation platform. I uploaded the integrated data from Talend Data Integration, which was produced in CSV delimiter format generated by the tFileOutputDelimited component. The figure below shows the uploaded CSV file in the talend data Preparation.

A screenshot of a computer

Description automatically generated

**Data Cleaning**

The data cleaning process involves individually clicking on each column. During this, I noticed that the 'TotalSpend' column contains 14 missing values.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

For the "TotalSpend" column, I identified 14 missing values. These were replaced with the mean value of 5028.48 to ensure that the column is free from any empty entries.

**Data Standardization**

The data standardization process involves individually clicking on each column. During this, I noticed that the 'Gender' column contains varying letter cases, and the 'Date' column also lacks standardization.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

The figure above shows the gender contains varying letter cases.

A screenshot of a cell test

Description automatically generated A screenshot of a graph

Description automatically generated

The figure above shows after standardized the gender.

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

After identifying the date columns, a function within Talend was selected to standardize the dates. The ISO 8601 format was specified within this function to transform all date entries to this uniform standard. Following the transformation, the data was reviewed to ensure that the dates were correctly standardized, with any anomalies or errors requiring further attention being identified and possibly corrected.

A white background with blue lines

Description automatically generated

The final step involves exporting the file from Talend Data Preparation in CSV format, ensuring that the delimiter is set to a comma.

**APPENDIX**

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a cell test

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A white background with blue lines

Description automatically generated