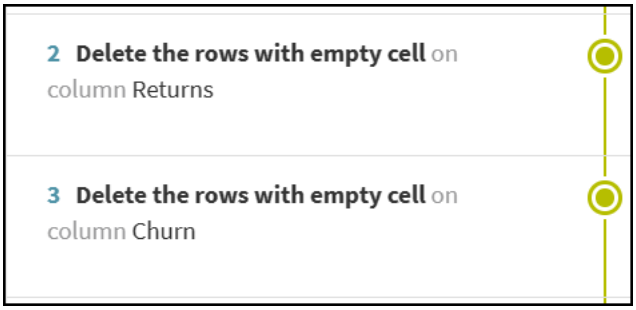
Talend Data Preparation

1. The dataset "ecommerce_customer_data_large (dirty)" was successfully loaded into Talend Data Preparation for cleaning and preparation.

2. Data Cleaning:

Addressing Missing Values:

- Rows with missing values in the "Returns" and "Churn" columns were deleted due to their low count (less than 10), minimizing impact on overall analysis.

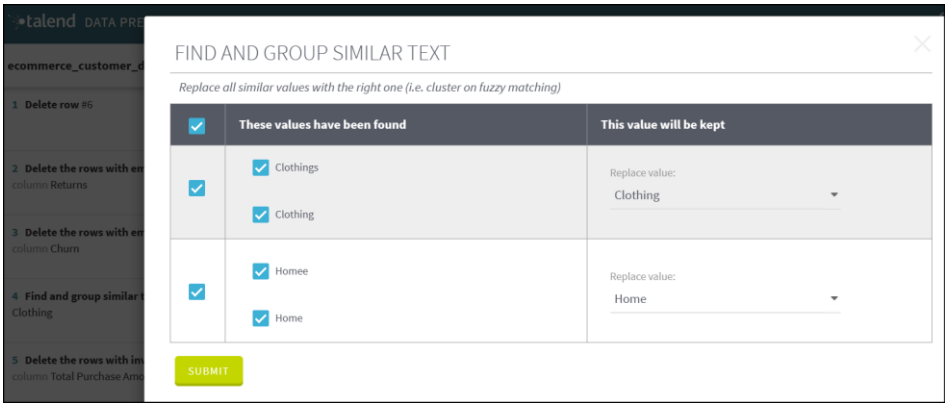


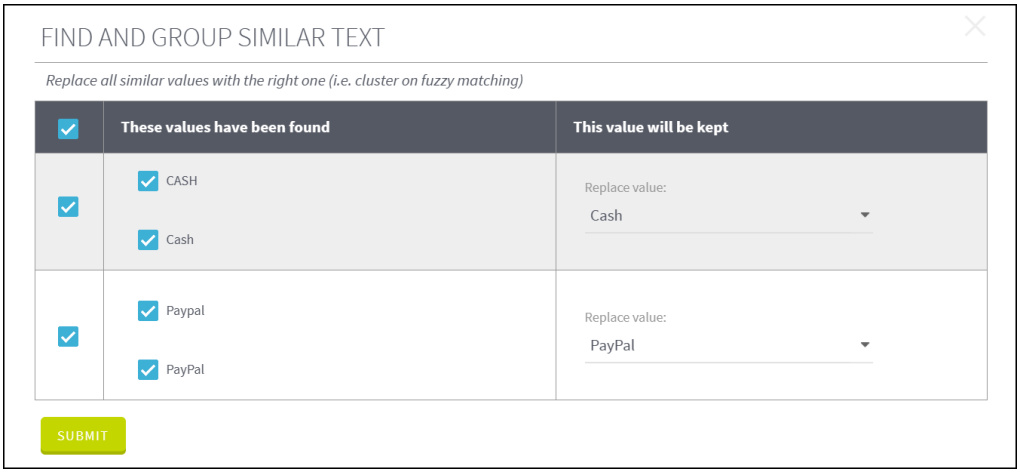
3. Correcting Data Inconsistencies:

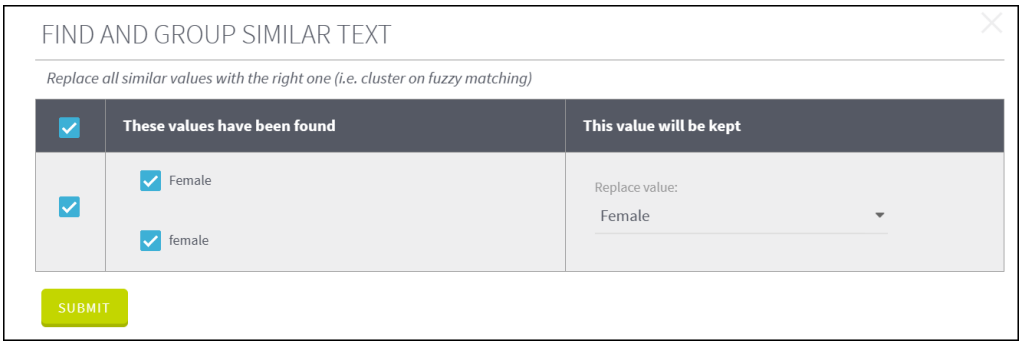
- The "Find and group similar text" function was used to:

  Address typo errors in the "Product_Category" column.

  Standardize data formats in the "Gender" and "Payment Method" columns.

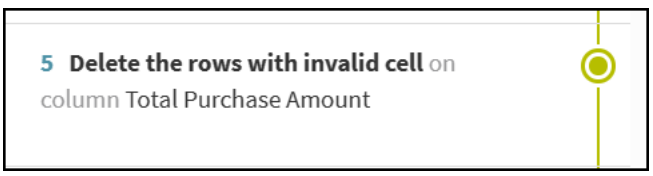




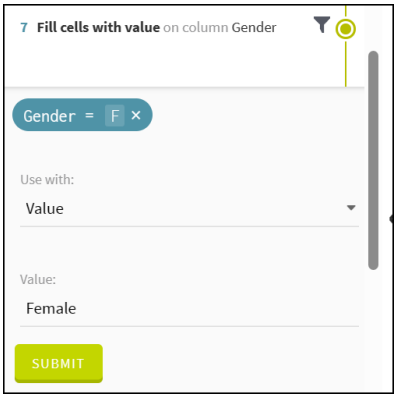


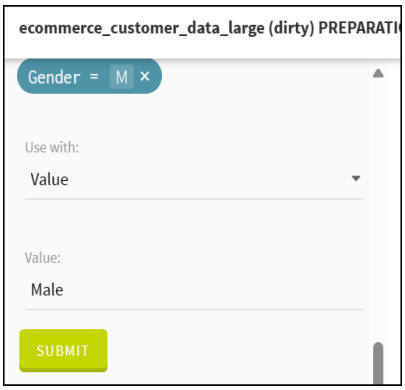
4. Removing Invalid Data:

- Rows with invalid records in the "Total Purchase Amount" column were deleted due to their limited quantity (less than 5).



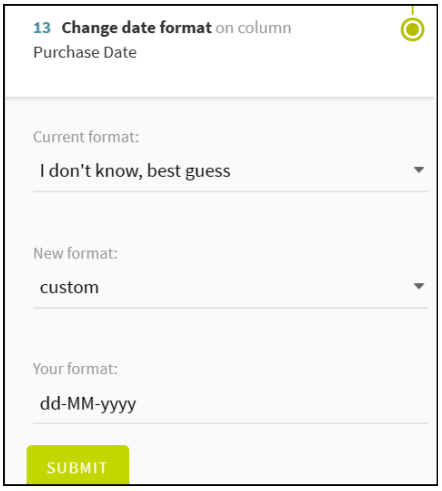
5. The "Fill cells with value" function rectified inconsistencies in the "Gender" column.





6. Standardizing Date Formats:

- The "Change date format" function ensured consistency in date representations throughout the dataset.



7. Overview of the data preprocessing. With the completion of data preprocessing, the dataset is now prepared for further analysis and modeling.