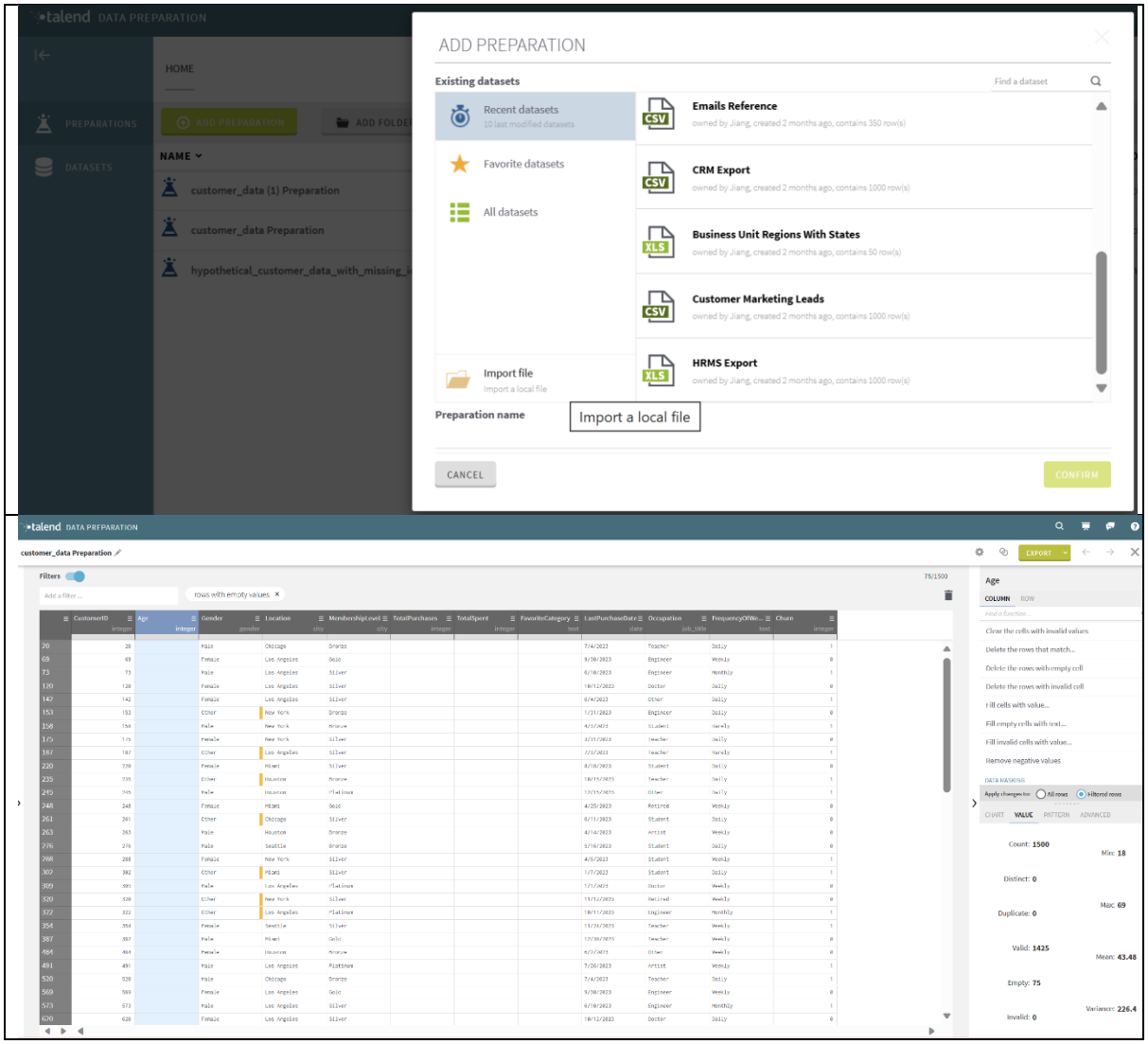
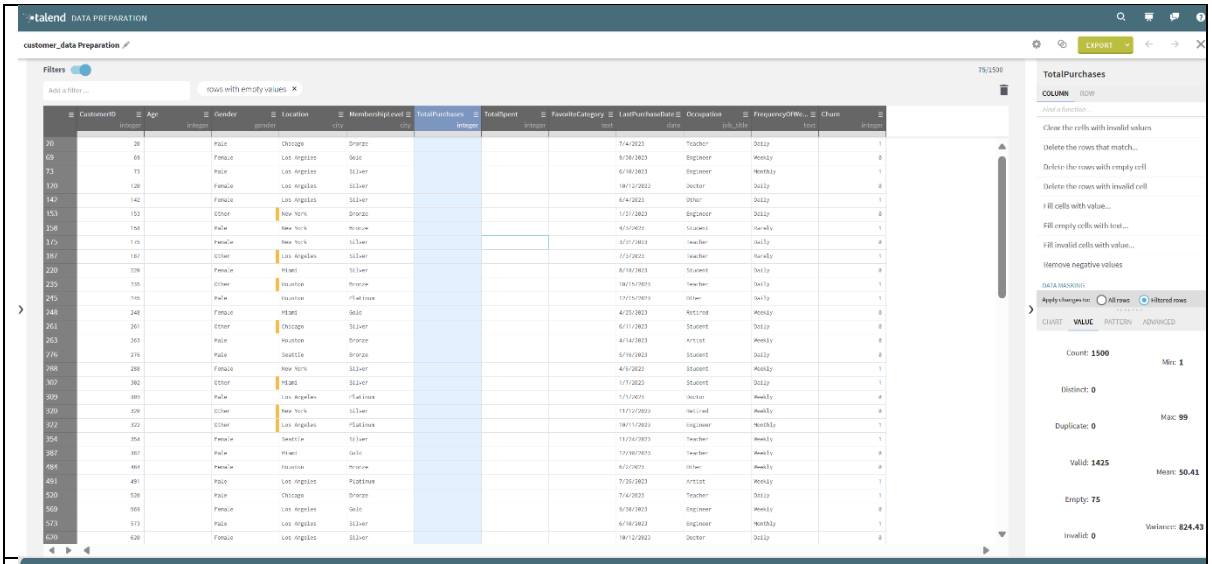
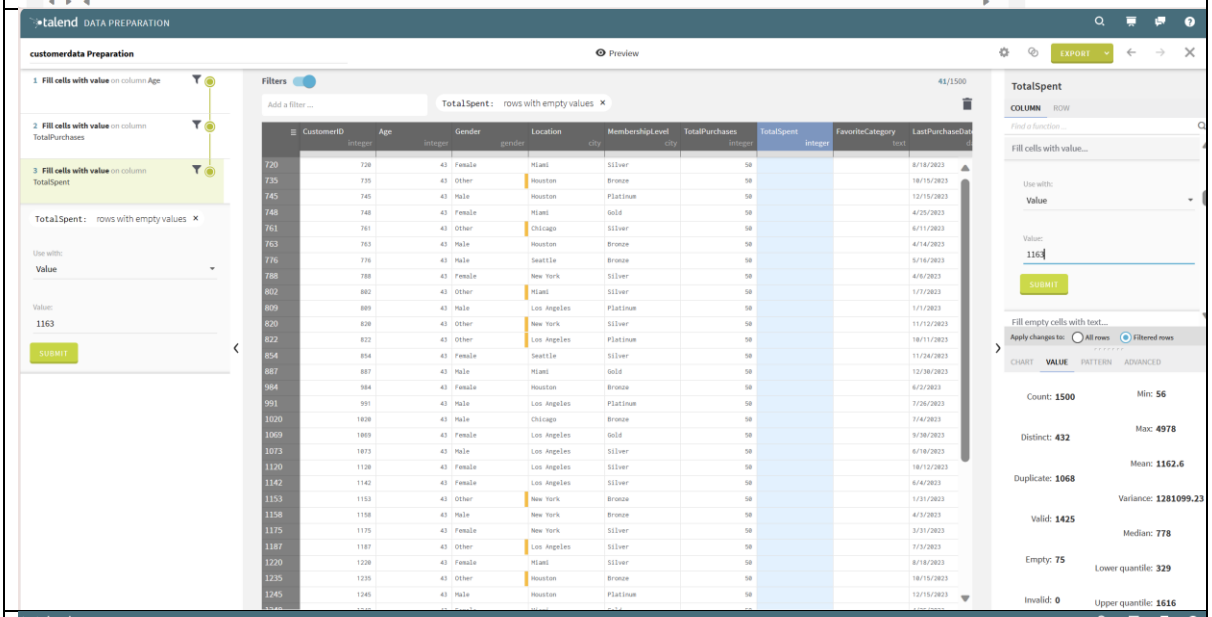
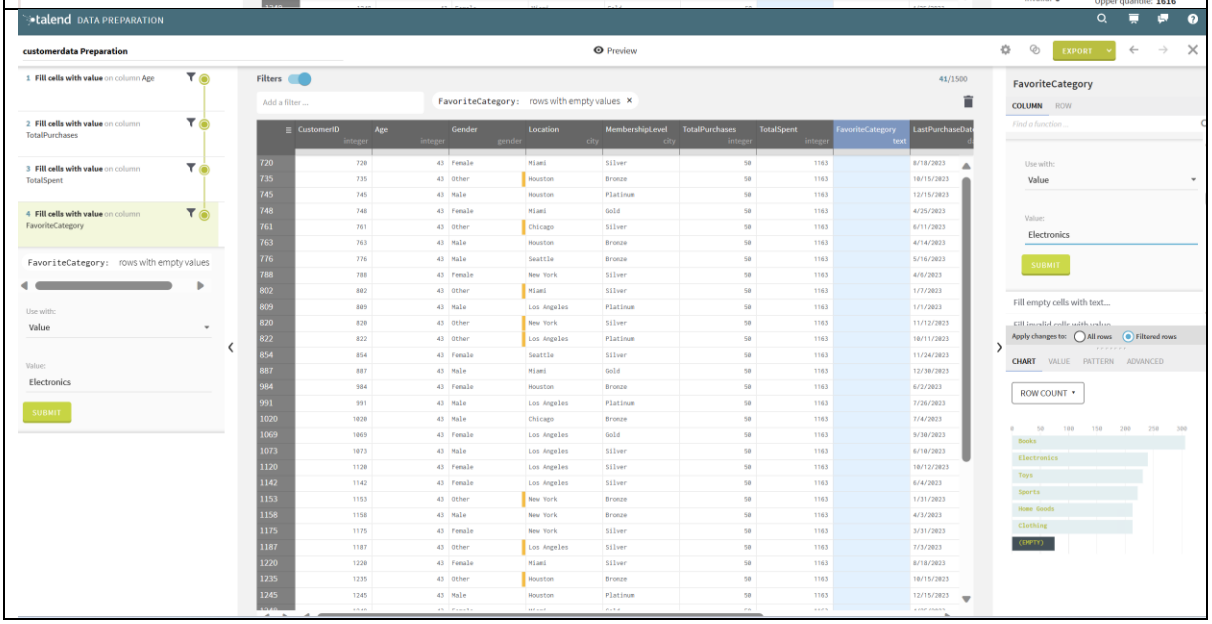
Talend Data Preparation

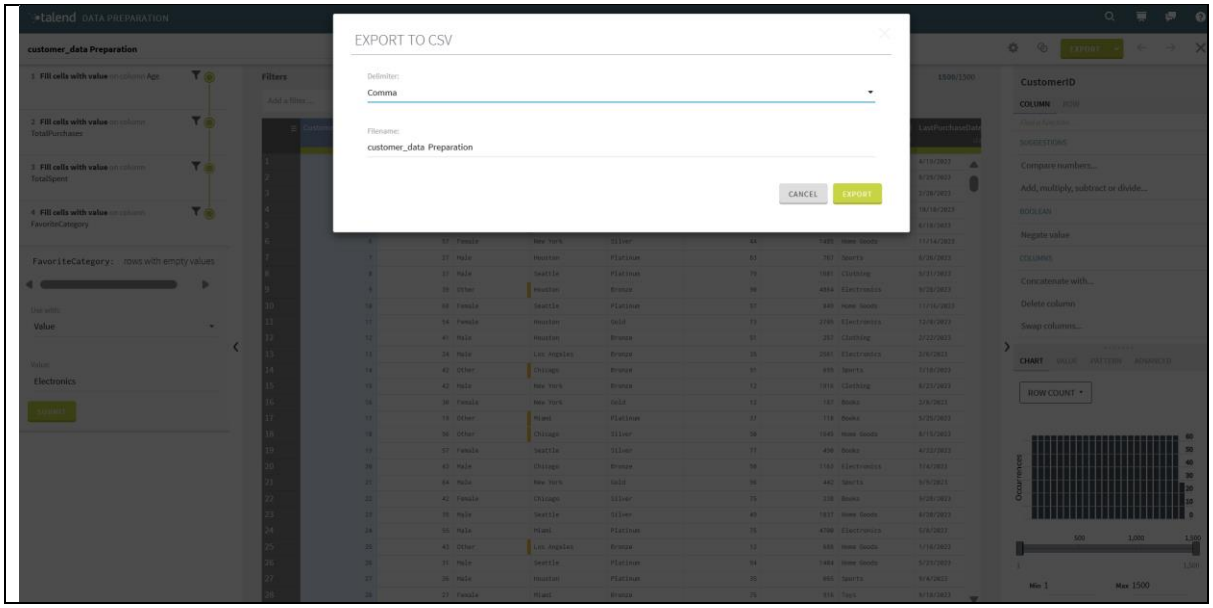
After merging the two dataset, the data preparation step is needed to do. In the Talend data preparation website page, we upload the dataset first then check the miss value first. It find there have two kind of missing class and interval. They are FavoriteCategory (75 missing values), Age(75 missing values), TotalPurchases (75 missing values), TotalSpent(75 missing values), then we impute the value, for FavoriteCategory the class the missing value it use the most frequentable one, in this case we choose books. For the other interval missing in this case we choose mean value to impute. After impute all the missing value export the file.




talend DATA PREPARATION
customer_data Preparation
EXPORT
Filters
rows with empty values
TotalPurchases
COLUMN
ROW
Clear the cells with invalid values
Delete the rows that match...
Delete the rows with empty cell
Delete the rows with invalid cell
Fill cells with value...
Fill empty cells with text...
Fill invalid cells with value...
Remove negative values
DATA MASKING
CHART
VALUE
PATTERN
ADVANCED
Count: 1500
Min: 1
Distinct: 0
Max: 99
Duplicate: 0
Valid: 1425
Mean: 50.41
Empty: 75
Variance: 824.43
Invalid: 0


talend DATA PREPARATION
customerdata Preparation
Preview
EXPORT
1 Fill cells with value on column Age
2 Fill cells with value on column TotalPurchases
3 Fill cells with value on column TotalSpent
TotalSpent: rows with empty values
Use with:
Value
Value:
1163
SUBMIT
Filters
41/1500
TotalSpent
COLUMN
ROW
Fill cells with value...
Fill empty cells with text...
Apply changes to:
All rows
Filtered rows
CHART
VALUE
PATTERN
ADVANCED
Count: 1500
Min: 56
Distinct: 432
Max: 4978
Mean: 1162.6
Duplicate: 1068
Variance: 1281099.23
Valid: 1425
Median: 778
Empty: 75
Lower quantile: 329
Invalid: 0
Upper quantile: 1616


talend DATA PREPARATION
customerdata Preparation
Preview
EXPORT
1 Fill cells with value on column Age
2 Fill cells with value on column TotalPurchases
3 Fill cells with value on column TotalSpent
4 Fill cells with value on column FavoriteCategory
FavoriteCategory: rows with empty values
Use with:
Value
Value:
Electronics
SUBMIT
Filters
41/1500
FavoriteCategory
COLUMN
ROW
Fill cells with value...
Fill empty cells with text...
Apply changes to:
All rows
Filtered rows
CHART
VALUE
PATTERN
ADVANCED
ROW COUNT
Books
Electronics
Toys
Sports
Home Goods
Clothing
(EMPTY)


talend DATA PREPARATION
customer_data Preparation
EXPORT TO CSV
Delimiter:
Comma
Filename:
customer_data Preparation
CANCEL
EXPORT
CustomerID
ROW COUNT
Occurrences
Min 1
Max 1500