

Tool 2: Use KNIME to transform and process the newdataset.

1. Check the null value situation of the newdataset.

Null value: none

The workflow diagram shows a sequence of nodes: CSV Reader (newdataset) → String Manipulation (r[s]) → Rule Engine (Coupon_Status (0/1)) → Rule Engine (Gender (0/1)). Below the workflow, a statistics table is displayed for 12 rows and 14 columns.

Name	Type	# Missing val...	# Unique valu...	Minimum	Maximum	25% Quantile	50% Quantile ...	75% Quantile	Mean
CustomerID	Number (inte...	0	1468	12,346	18,283	13,869	15,311	16,996.75	15,346.71
Gender	String	0	2	⓪	⓪	⓪	⓪	⓪	⓪
Location	String	0	5	⓪	⓪	⓪	⓪	⓪	⓪
Tenure_Months	Number (inte...	0	49	2	50	15	27	37	26.128
Transaction_ID	Number (inte...	0	25061	16,679	48,497	25,384	32,625.5	39,126.75	32,409.826
Transaction_D...	String	0	365	⓪	⓪	⓪	⓪	⓪	⓪
Product_SKU	String	0	1145	⓪	⓪	⓪	⓪	⓪	⓪
Product_Cate...	String	0	20	⓪	⓪	⓪	⓪	⓪	⓪
Quantity	Number (inte...	0	151	1	900	1	1	2	4.498
Avg_Price	Number (dou...	0	546	0.39	355.74	5.7	16.99	102.13	52.238
Delivery_Char...	Number (dou...	0	267	0	521.36	⓪	⓪	⓪	10.518
Coupon_Status	String	0	3	⓪	⓪	⓪	⓪	⓪	⓪

2. Convert the Coupon_Status from a ternary classification: Used, Not Used, Clicked, into a binary classification of Used and Not Used, where Clicked is considered as Not Used, in order to make the data more balanced.

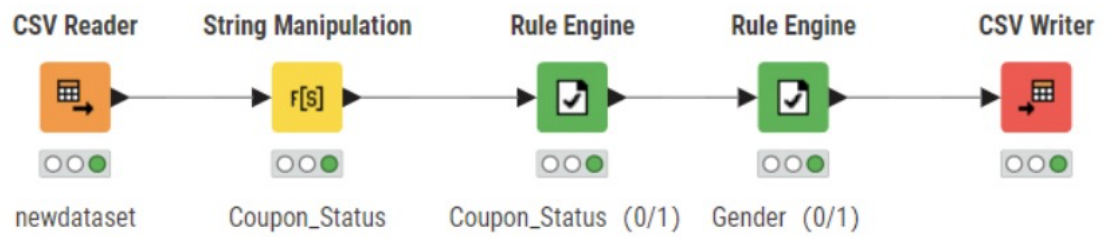
```
Expression
1 replace($Coupon_Status$, "Clicked", "Not Used")
2
```

3. Conduct some conversions from string to numerical values to facilitate better modeling, such as binary variables for gender and Coupon.

```
5 $Gender$ = "F" => "1"
6 $Gender$ = "M" => "0"
7 TRUE => $Gender$

5 $Coupon_Status$ = "Used" => "1"
6 $Coupon_Status$ = "Not Used" => "0"
7 TRUE => $Coupon_Status$
```

4. The overall workflow, exported as datasetfinal.csv.



5. Completed the imbalance processing of the data in KINME

