

# Documentation for Talend data Prep

- Before importing the dataset into SAS Enterprise Miner we needed be to Analysed and clean the dataset, for cleaning Talend Data Prep tool was used.

e\_commerce\_customer\_behaviour\_analysis\_n

10008/10008

Filters

Add a filter ...

	CustomerID	Age	Age	Gender	MembershipLevel	TotalPurchases	TotalSpent	Favor
	integer	integer	integer	gender	city	integer	decimal	
1	1	62	62	Female	Gold	158	3362.292989	Hor
2	2	65	65	Male	Gold	10	7630.458001	Hor
3	3	18	18	Other	Silver	113	8276.422097	Bei
4	4	21	21	Male	Bronze	103	9464.29354	Hor
5	5	21	21	Female	Gold	32	4103.204939	Boc
6	6	57	57	Male	Silver	27	5547.643896	Sp
7	7	27	27	Female	Platinum	197	7138.378812	El
8	8	37	37	Male	Bronze	72	7621.88857	Boc
9	9	39	39	Male	Bronze	10	9061.91898	Boc
10	10	68	68	Male	Bronze	143	8477.144206	Cl
11	11	54	54	Male	Gold	18	4230.132683	Sp
12	12	41	41	Female	Gold	78	8226.539905	Hor
13	13	24	24	Male	Bronze	49	9962.428441	Hor
14	14	42	42	Male	Silver	27	4651.319976	Boc
15	15	42	42	Male	Platinum	13	8082.666966	Cl
16	16	30	30	Female	Silver	192	2220.441548	Boc
17	17	19	19	Male	Silver	44	4313.368752	Boc

CustomerID

COLUMN ROW

Find a function ...

SUGGESTIONS

- Delete the rows with empty cell
- Fill empty cells with text...
- Compare numbers...
- Add, multiply, subtract or divide...

CHART VALUE PATTERN ADVANCED

Count: 10008 Min: 1

Distinct: 10001 Max: 10000

Duplicate: 7 Mean: 5000.5

Valid: 10000 Variance: 8334166.67

Empty: 8 Median: 5000.5

Invalid: 0 Lower quantile: 2500.25

Upper quantile: 7500.75

**Diagram 1**

- we can see in diagram 1 that there are 8 empty customer ID's , lets remove it

Filters

Add a filter ...

	CustomerID	Age	Age
	integer	integer	integer
1	1		
2	2		65
3	3		18

8 empty values (0%)

	CustomerID	Age	Age	Gender
	integer	integer	integer	gender
1				Female
2				Male
3				Other
4				Male
5	5	21	2	Female
6	6	57	57	Male

Select rows with empty values for CustomerID  
 Delete the rows with empty cell

8 empty values (0%)

- There is also a duplicate age column in the dataset lets remove that as well as.

Filters

Add a filter ...


	CustomerID	Age	Age	Gender
	integer			gender
1	1			Female
2	2			Male
3	3			Other
4	4			Male
5	5			Female
6	6			Male
7	7	27		

This column is a integer ▶  
 Duplicate column  
 Rename column  
 Create new column  
 Delete column

Delete column

- The current date format is in a different format, we can change it to look more understandable.

LastPurchaseDate≡ date	Re
10/18/2023	
6/4/2023	
10/6/2023	
12/9/2023	
1/7/2023	
9/17/2023	
2/12/2023	



LastPurchaseDate≡ date	Re
18/10/23	
04/06/23	
06/10/23	
09/12/23	
07/01/23	
17/09/23	
12/02/23	

- After then we can export the cleaned data. From the top right corner

