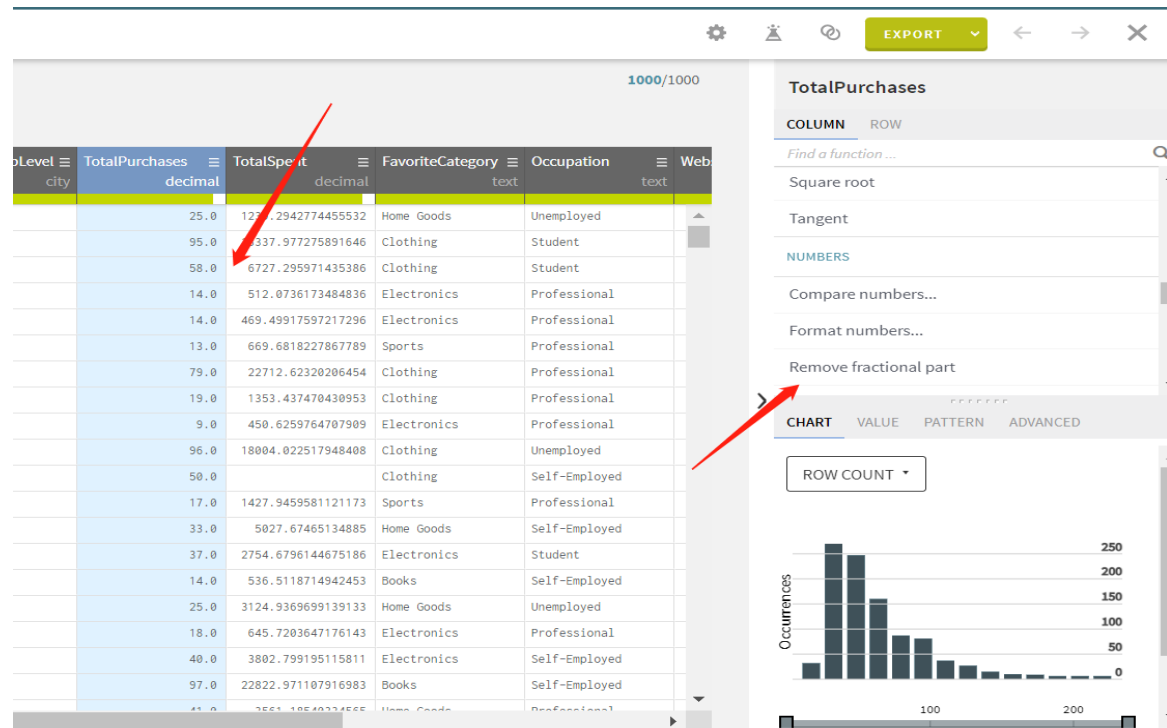


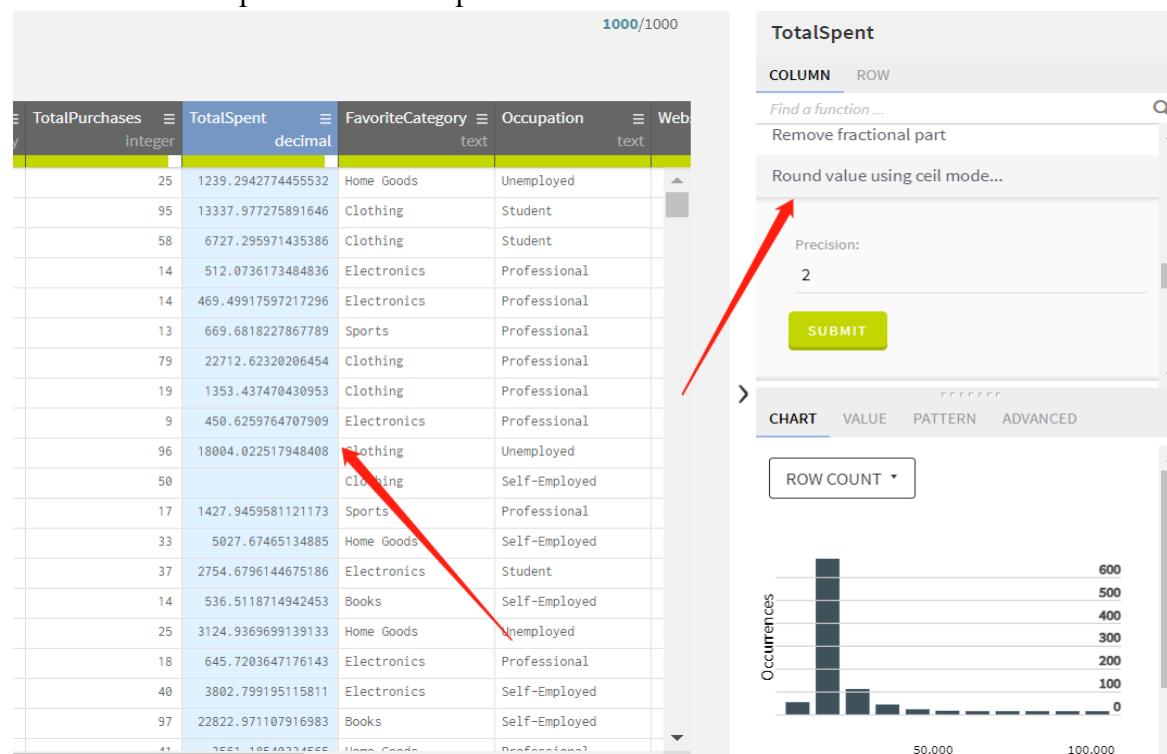
## Data Preparation (DP)

After merging the distinct datasets, we need to conduct initial exploration and processing of the data.

- 2.1 It is not reasonable for TotalPurchases to have decimal points, so I removed the decimal points from TotalPurchases.

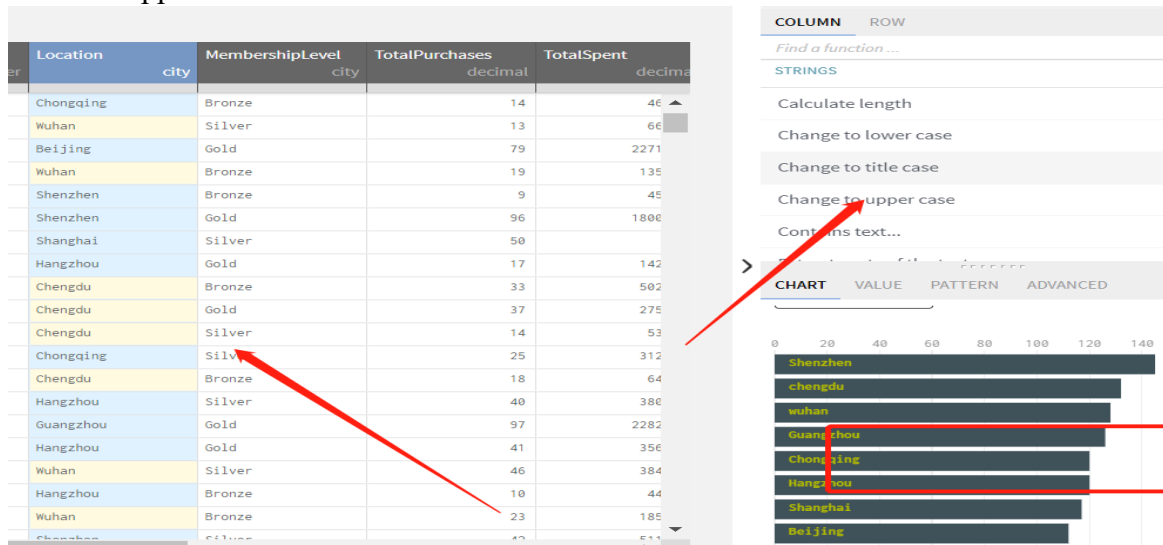


- 2.2 Many decimal places in TotalSpent are not reasonable, so I retained two decimal places for TotalSpent.

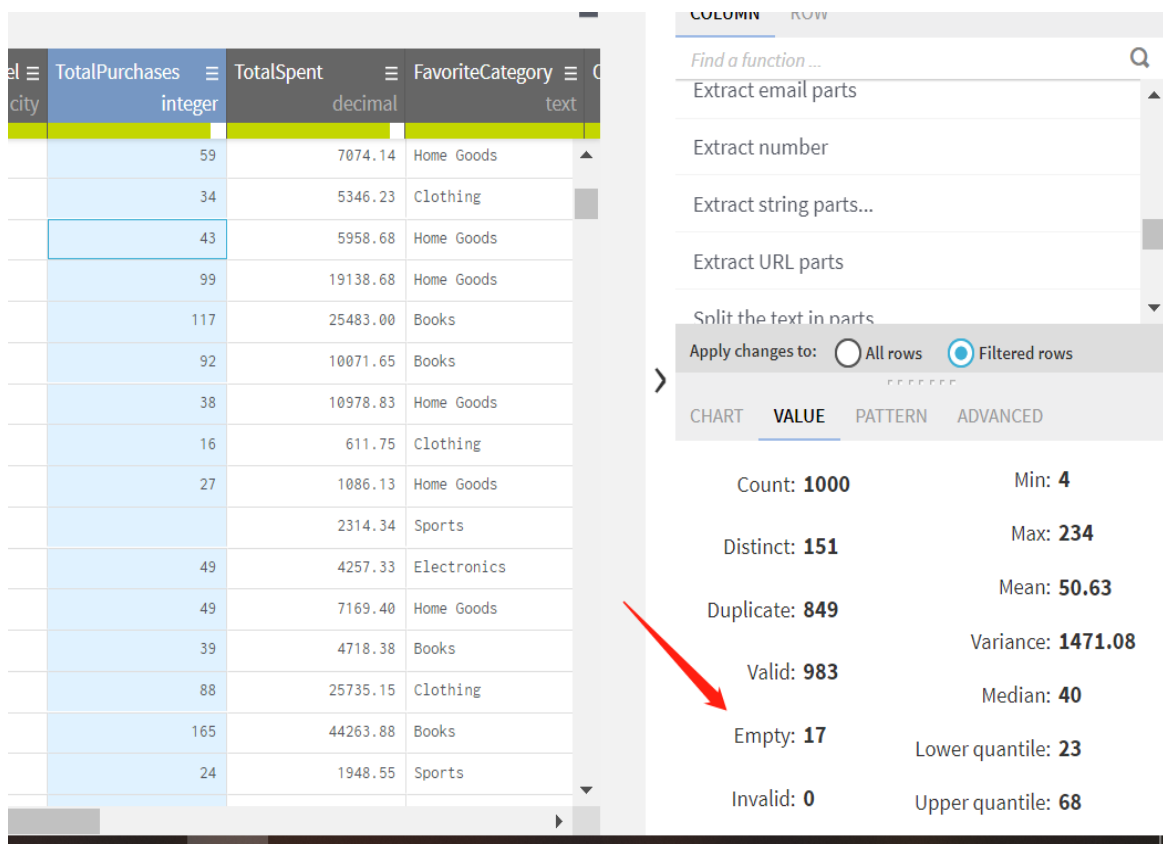


- 2.3 The capitalization of city names in the location data is inconsistent, with

some having uppercase initial letters and others having lowercase initial letters. This is not reasonable, so I have standardized all city names to have uppercase initial letters.



- 2.4 I observed that both TotalPurchases and TotalSpent have missing values, with 17 and 39 missing values respectively. I won't handle them at this point; instead, I will examine the overall structure of the data inside SAS EM before proceeding with any imputation or treatment.



TotalSpent	FavoriteCategory
decimal	text
7074.14	Home Goods
5346.23	Clothing
5958.68	Home Goods
19138.68	Home Goods
25483.00	Books
10071.65	Books
10978.83	Home Goods
611.75	Clothing
1086.13	Home Goods
2314.34	Sports
4257.33	Electronics
7169.40	Home Goods
4718.38	Books
25735.15	Clothing
44263.88	Books
1948.55	Sports

Find a function ...

SPLIT

Extract email parts

Extract number

Extract string parts...

Extract URL parts

Apply changes to: ☐ All rows ☒ Filtered rows

CHART VALUE PATTERN ADVANCED

Count: **1000** Min: **89.28**

Distinct: **962** Max: **109341.17**

Duplicate: **38** Mean: **9625.96**

Valid: **961** Variance: **210668700**

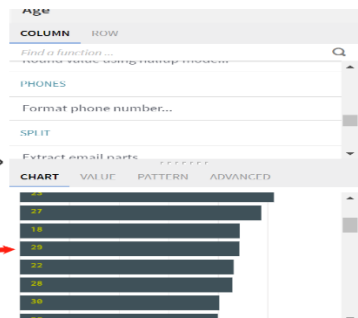
Empty: **39** Median: **4038.71**

Invalid: **0** Lower quantile: **1455.95**

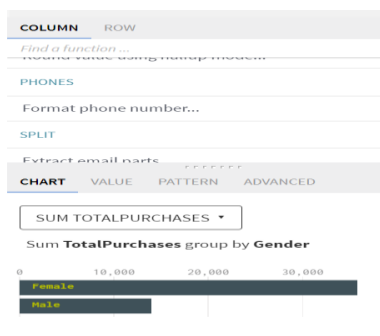
Upper quantile: **10333.2**

- 2.5 Observing this dataset, it appears that the majority of individuals are young, and the spending capacity of females is significantly higher than that of males. This aligns well with real-world scenarios.

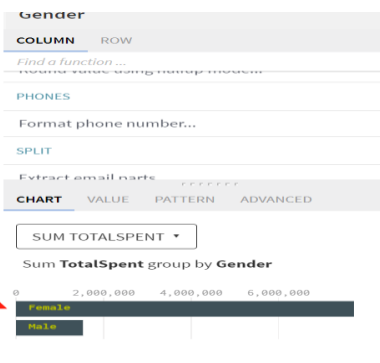
Age	Gender	Location	MembershipLevel	TotalPurchases	TotalSpent
integer	gender	city	city	integer	
86667	49	Female	Shenzhen	Bronze	25
86533	21	Male	Guangzhou	Gold	95
85972	24	Female	Hangzhou	Bronze	58
86227	64	Male	Chongqing	Bronze	14
86346	23	Male	Chongqing	Bronze	14
86645	32	Male	Wuhan	Silver	13
85781	65	Female	Beijing	Gold	79
85985	38	Female	Wuhan	Bronze	19
85980	45	Male	Shenzhen	Bronze	9
86388	26	Female	Shenzhen	Gold	96
86178	42	Female	Shanghai	Silver	50
86232	36	Male	Hangzhou	Gold	17
86468	28	Female	Chengdu	Bronze	33
85913	20	Male	Chengdu	Gold	37



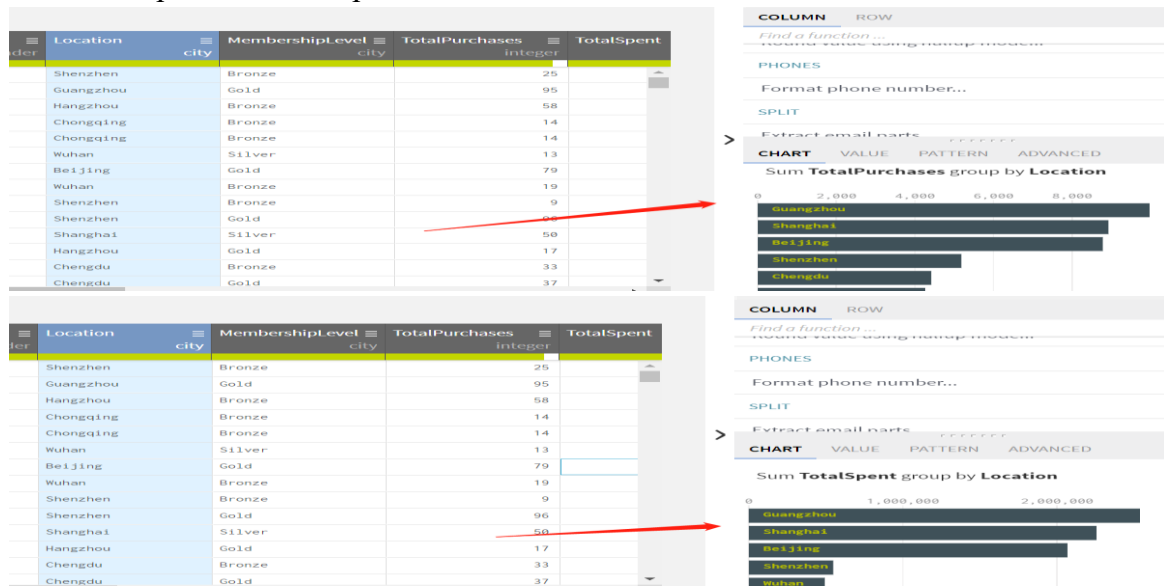
Gender	Location	MembershipLevel	TotalPurchases	TotalSpent
gender	city	city	integer	
49	Female	Shenzhen	Bronze	25
21	Male	Guangzhou	Gold	95
24	Female	Hangzhou	Bronze	58
64	Male	Chongqing	Bronze	14
23	Male	Chongqing	Bronze	14
32	Male	Wuhan	Silver	13
65	Female	Beijing	Gold	79
38	Female	Wuhan	Bronze	19
45	Male	Shenzhen	Bronze	9
25	Female	Shenzhen	Gold	96
42	Female	Shanghai	Silver	50
36	Male	Hangzhou	Gold	17
28	Female	Chengdu	Bronze	33
20	Male	Chengdu	Gold	37



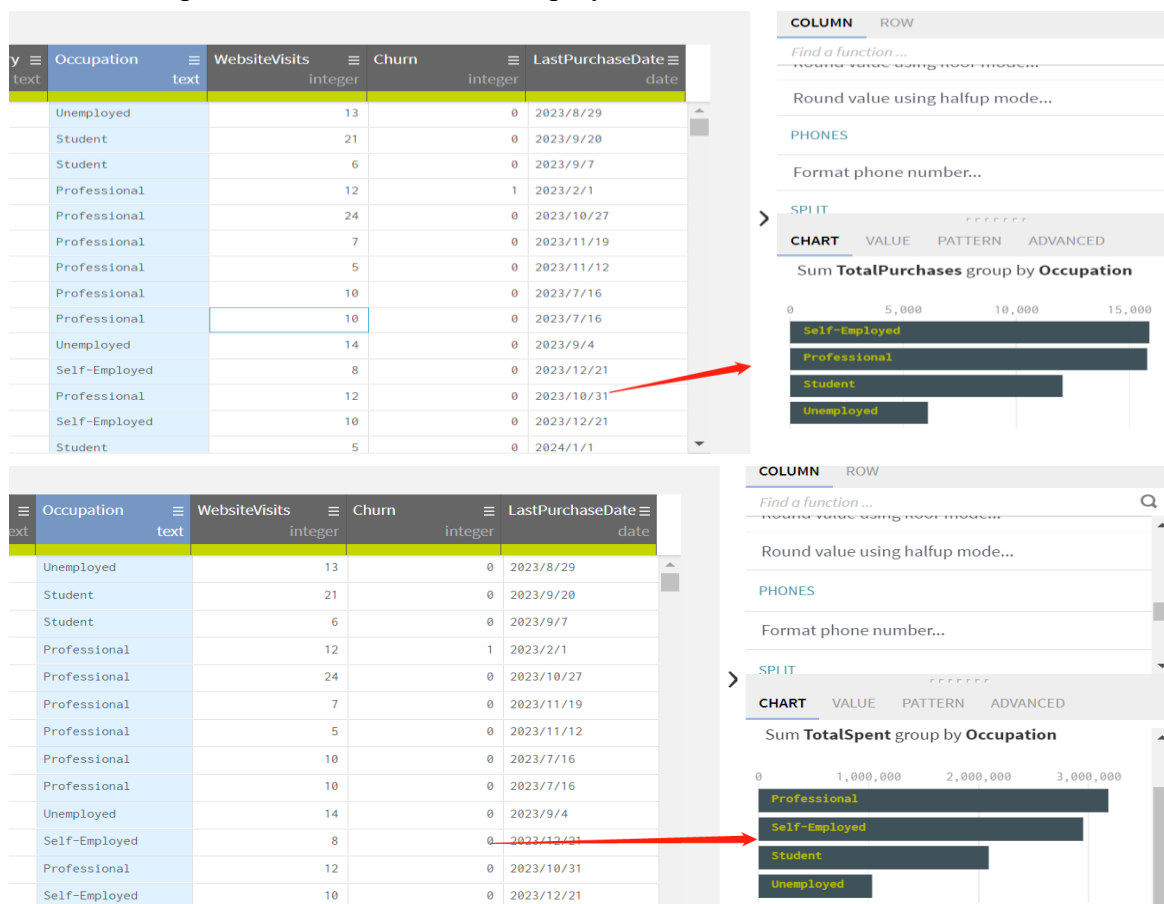
Gender	Location	MembershipLevel	TotalPurchases	TotalSpent
gender	city	city	integer	
49	Female	Shenzhen	Bronze	25
21	Male	Guangzhou	Gold	95
24	Female	Hangzhou	Bronze	58
64	Male	Chongqing	Bronze	14
23	Male	Chongqing	Bronze	14
32	Male	Wuhan	Silver	13
65	Female	Beijing	Gold	79
38	Female	Wuhan	Bronze	19
45	Male	Shenzhen	Bronze	9
25	Female	Shenzhen	Gold	96
42	Female	Shanghai	Silver	50
36	Male	Hangzhou	Gold	17
28	Female	Chengdu	Bronze	33
20	Male	Chengdu	Gold	37



- 2.6 It can be observed that major cities like Beijing, Guangzhou, and Shanghai exhibit significantly higher purchase frequencies and total expenditures compared to other cities.



- 2.7 It is noticeable that individuals with employment or student status tend to have significantly higher purchase frequencies and total expenditures compared to those who are unemployed.



- 2.8 After completing the conversion of the data format and cleaning the data, as well as the initial exploration of data correlations, we exported the data and continued to SAS EM to process the analysis

EXPORT TO CSV

Delimiter:  
Semicolon

Filename:  
Aggregate customer information PREPARATION

CANCEL EXPORT