

# 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.

The screenshot displays the Talend Data Preparation interface. The top section shows the 'ADD PREPARATION' dialog, which includes a sidebar for 'Existing datasets' (Recent, Favorite, All) and a list of datasets like 'Emails Reference', 'CRM Export', 'Business Unit Regions With States', 'Customer Marketing Leads', and 'HRMS Export'. The 'Import file' button is highlighted. Below the dialog, the 'customer\_data Preparation' dataset is shown in a table view. The table has columns for CustomerID, Age, Gender, Location, Membership, TotalPurchases, TotalSpent, FavoriteCategory, LastPurchaseDate, Occupation, FrequencyOfVisits, and Churn. The 'Age' column is highlighted, and the 'rows with empty values' filter is applied. The right sidebar shows the 'Age' column settings, including a 'COUNT' of 1500 and a 'Max' of 18.

CustomerID	Age	Gender	Location	Membership	TotalPurchases	TotalSpent	FavoriteCategory	LastPurchaseDate	Occupation	FrequencyOfVisits	Churn
20	20	Male	Chicago	Bronze				7/4/2023	Teacher	Daily	1
69	69	Female	Los Angeles	Gold				9/30/2023	Engineer	Weekly	0
73	73	Male	Los Angeles	Silver				6/18/2023	Engineer	Monthly	1
130	130	Female	Los Angeles	Silver				10/12/2023	Doctor	Daily	0
147	142	Female	Los Angeles	Silver				6/4/2023	Engineer	Daily	1
153	153	Other	New York	Bronze				1/31/2023	Engineer	Daily	0
158	158	Male	New York	Bronze				4/5/2023	Student	Weekly	1
175	175	Female	New York	Silver				12/1/2023	Teacher	Daily	0
187	187	Other	Los Angeles	Silver				7/1/2023	Teacher	Weekly	1
220	220	Female	Miami	Silver				8/18/2023	Student	Daily	0
225	225	Other	Houston	Bronze				10/15/2023	Teacher	Daily	1
255	255	Male	Houston	Platinum				12/15/2023	Other	Daily	1
268	268	Female	Miami	Gold				4/20/2023	Artist	Monthly	0
261	261	Other	Chicago	Silver				6/11/2023	Student	Daily	0
263	263	Male	Houston	Bronze				4/14/2023	Artist	Weekly	0
276	276	Male	Seattle	Bronze				9/16/2023	Student	Daily	0
288	288	Female	New York	Silver				4/6/2023	Student	Weekly	1
357	357	Other	Miami	Silver				1/7/2023	Student	Daily	1
395	395	Male	Los Angeles	Platinum				1/5/2023	Doctor	Weekly	0
580	580	Other	New York	Silver				11/12/2023	Teacher	Weekly	0
597	597	Other	Los Angeles	Platinum				10/17/2023	Engineer	Monthly	1
594	594	Female	Seattle	Silver				11/24/2023	Teacher	Weekly	1
587	587	Male	Miami	Gold				12/30/2023	Teacher	Weekly	0
664	664	Female	Houston	Bronze				8/2/2023	Other	Weekly	0
491	491	Male	Los Angeles	Platinum				7/26/2023	Artist	Weekly	1
530	530	Male	Chicago	Bronze				7/4/2023	Teacher	Daily	1
569	569	Female	Los Angeles	Gold				9/30/2023	Engineer	Weekly	0
573	573	Male	Los Angeles	Silver				6/18/2023	Engineer	Monthly	1
620	620	Female	Los Angeles	Silver				10/12/2023	Doctor	Daily	0

**talend DATA PREPARATION**

customer\_data Preparation

Filters: Add a filter... rows with empty values

CustomerID	Age	Gender	Location	MembershipLevel	TotalPurchases	TotalSpent	FavoriteCategory	LastPurchaseDate	Occupation	FrequencyOfVisits	Churn
20	28	Male	Chicago	Bronze			7/4/2023	Teacher	Daily		1
69	69	Female	Los Angeles	Gold			9/26/2023	Engineer	Weekly		0
73	73	Male	Los Angeles	Silver			6/18/2023	Engineer	Weekly		1
120	120	Female	Los Angeles	Silver			18/7/2023	Doctor	Daily		0
147	142	Female	Los Angeles	Silver			6/4/2023	Other	Daily		1
153	153	Other	New York	Bronze			1/31/2023	Engineer	Daily		0
156	156	Male	New York	Bronze			4/2/2023	Student	Weekly		1
175	175	Female	New York	Silver			3/27/2023	Teacher	Daily		0
187	187	Other	Los Angeles	Silver			7/7/2023	Teacher	Weekly		1
220	220	Female	Miami	Silver			8/18/2023	Student	Daily		0
235	235	Other	Houston	Bronze			18/5/2023	Teacher	Weekly		1
245	245	Male	Houston	Platinum			12/1/2023	Other	Weekly		1
248	248	Female	Miami	Gold			4/25/2023	Retired	Weekly		0
261	261	Other	Chicago	Silver			6/17/2023	Student	Daily		0
263	263	Male	Houston	Bronze			4/14/2023	Artist	Weekly		0
276	276	Male	Seattle	Bronze			5/18/2023	Student	Daily		1
288	288	Female	New York	Silver			4/4/2023	Student	Weekly		1
307	307	Other	Miami	Silver			1/7/2023	Student	Daily		1
309	309	Male	Los Angeles	Platinum			1/15/2023	Doctor	Weekly		0
320	320	Other	New York	Silver			11/7/2023	Retired	Weekly		0
322	322	Other	Los Angeles	Platinum			18/7/2023	Engineer	Weekly		1
354	354	Female	Seattle	Silver			7/27/2023	Teacher	Weekly		1
367	367	Male	Miami	Gold			12/26/2023	Teacher	Weekly		0
484	484	Female	Houston	Bronze			6/27/2023	Other	Weekly		0
481	481	Male	Los Angeles	Platinum			7/26/2023	Artist	Weekly		1
520	520	Male	Chicago	Bronze			7/4/2023	Teacher	Daily		1
569	569	Female	Los Angeles	Gold			5/28/2023	Engineer	Weekly		0
573	573	Male	Los Angeles	Silver			6/18/2023	Engineer	Weekly		1
626	626	Female	Los Angeles	Silver			18/7/2023	Doctor	Daily		0

TotalPurchases

Count: 1500

Distinct: 0

Duplicate: 0

Valid: 1425

Empty: 75

Invalid: 0

Min: 1

Max: 99

Mean: 50.41

Variance: 824.43

**talend DATA PREPARATION**

customerdata Preparation

Filters: Add a filter... TotalSpent: rows with empty values

CustomerID	Age	Gender	Location	MembershipLevel	TotalPurchases	TotalSpent	FavoriteCategory	LastPurchaseDate
720	720	43	Female	Miami	Silver	50		8/18/2023
735	735	43	Other	Houston	Bronze	50		18/15/2023
745	745	43	Male	Houston	Platinum	50		12/15/2023
748	748	43	Female	Miami	Gold	50		4/25/2023
761	761	43	Other	Chicago	Silver	50		6/11/2023
763	763	43	Male	Houston	Bronze	50		4/14/2023
776	776	43	Male	Seattle	Bronze	50		5/16/2023
788	788	43	Female	New York	Silver	50		4/8/2023
802	802	43	Other	Miami	Silver	50		1/7/2023
809	809	43	Male	Los Angeles	Platinum	50		1/7/2023
820	820	43	Other	New York	Silver	50		11/12/2023
822	822	43	Other	Los Angeles	Platinum	50		18/11/2023
854	854	43	Female	Seattle	Silver	50		11/24/2023
887	887	43	Male	Miami	Gold	50		12/26/2023
984	984	43	Female	Houston	Bronze	50		6/2/2023
991	991	43	Male	Los Angeles	Platinum	50		4/2/2023
1020	1020	43	Male	Chicago	Bronze	50		7/26/2023
1069	1069	43	Female	Los Angeles	Gold	50		5/38/2023
1073	1073	43	Male	Los Angeles	Silver	50		6/18/2023
1120	1120	43	Female	Los Angeles	Silver	50		18/12/2023
1142	1142	43	Female	Los Angeles	Silver	50		6/4/2023
1153	1153	43	Other	New York	Bronze	50		1/31/2023
1158	1158	43	Male	New York	Bronze	50		4/3/2023
1175	1175	43	Female	New York	Silver	50		3/31/2023
1187	1187	43	Other	Los Angeles	Silver	50		7/3/2023
1220	1220	43	Female	Miami	Silver	50		8/18/2023
1235	1235	43	Other	Houston	Bronze	50		18/15/2023
1245	1245	43	Male	Houston	Platinum	50		12/15/2023

TotalSpent

Count: 1500

Distinct: 432

Duplicate: 1068

Valid: 1425

Empty: 75

Invalid: 0

Min: 56

Max: 4978

Mean: 1162.6

Variance: 1281099.23

Median: 778

Lower quantile: 329

Upper quantile: 1616

**talend DATA PREPARATION**

customerdata Preparation

Filters: Add a filter... FavoriteCategory: rows with empty values

CustomerID	Age	Gender	Location	MembershipLevel	TotalPurchases	TotalSpent	FavoriteCategory	LastPurchaseDate
720	720	43	Female	Miami	Silver	50	1163	
735	735	43	Other	Houston	Bronze	50	1163	
745	745	43	Male	Houston	Platinum	50	1163	
748	748	43	Female	Miami	Gold	50	1163	
761	761	43	Other	Chicago	Silver	50	1163	
763	763	43	Male	Houston	Bronze	50	1163	
776	776	43	Male	Seattle	Bronze	50	1163	
788	788	43	Female	New York	Silver	50	1163	
802	802	43	Other	Miami	Silver	50	1163	
809	809	43	Male	Los Angeles	Platinum	50	1163	
820	820	43	Other	New York	Silver	50	1163	
822	822	43	Other	Los Angeles	Platinum	50	1163	
854	854	43	Female	Seattle	Silver	50	1163	
887	887	43	Male	Miami	Gold	50	1163	
984	984	43	Female	Houston	Bronze	50	1163	
991	991	43	Male	Los Angeles	Platinum	50	1163	
1020	1020	43	Male	Chicago	Bronze	50	1163	
1069	1069	43	Female	Los Angeles	Gold	50	1163	
1073	1073	43	Male	Los Angeles	Silver	50	1163	
1120	1120	43	Female	Los Angeles	Silver	50	1163	
1142	1142	43	Female	Los Angeles	Silver	50	1163	
1153	1153	43	Other	New York	Bronze	50	1163	
1158	1158	43	Male	New York	Bronze	50	1163	
1175	1175	43	Female	New York	Silver	50	1163	
1187	1187	43	Other	Los Angeles	Silver	50	1163	
1220	1220	43	Female	Miami	Silver	50	1163	
1235	1235	43	Other	Houston	Bronze	50	1163	
1245	1245	43	Male	Houston	Platinum	50	1163	

FavoriteCategory

Count: 1500

Distinct: 432

Duplicate: 1068

Valid: 1425

Empty: 75

Invalid: 0

Min: 56

Max: 4978

Mean: 1162.6

Variance: 1281099.23

Median: 778

Lower quantile: 329

Upper quantile: 1616

talend DATA PREPARATION

customer\_data Preparation

1. Fill cells with value on columns: customer Age

2. Fill cells with value on columns: totalPurchases

3. Fill cells with value on columns: totalSpent

4. Fill cells with value on columns: FavoriteCategory

FavoritesCategory: rows with empty values

Use with: Value

Output: Electronics

EXPORT

Filters

Add a filter...

customer\_data Preparation

EXPORT TO CSV

Delimiter: Comma

Filename: customer\_data Preparation

CANCEL EXPORT

CustomerID

COLUMN: ROW

SUBSET ROWS

Compare numbers...

Add, multiply, subtract or divide...

ROUNDS

Negate value

Catascenate with...

Delete column

Swap columns...

CHART

ROW COUNT

Chart visualization showing a grid of data points.

Max 1 Max 1500

1	20	Female	New York	Silver	15	100	New Goods	5/15/2023
2	21	Male	Houston	Platinum	61	767	Sports	6/26/2023
3	21	Male	Seattle	Platinum	76	1061	Clothing	6/21/2023
4	30	Other	Houston	Bronze	36	4854	Electronics	6/28/2023
5	30	Female	Seattle	Platinum	97	849	New Goods	1/16/2023
6	34	Female	Houston	Gold	71	2185	Electronics	12/18/2023
7	41	Male	Houston	Bronze	91	281	Clothing	2/22/2023
8	36	Male	Los Angeles	Bronze	38	2081	Electronics	5/18/2023
9	42	Other	Chicago	Bronze	37	895	Sports	5/18/2023
10	42	Male	New York	Bronze	11	1816	Clothing	6/21/2023
11	36	Female	New York	Gold	18	187	Books	5/18/2023
12	18	Other	Miami	Platinum	21	718	Books	5/25/2023
13	36	Other	Chicago	Silver	58	1645	New Goods	6/15/2023
14	37	Female	Seattle	Silver	71	458	Books	4/22/2023
15	42	Male	Chicago	Bronze	98	1163	Electronics	1/14/2023
16	36	Male	New York	Gold	36	442	Sports	6/10/2023
17	42	Female	Chicago	Silver	75	218	Books	6/28/2023
18	36	Male	Seattle	Silver	40	1817	New Goods	4/28/2023
19	42	Male	Miami	Platinum	75	4758	Electronics	5/18/2023
20	42	Other	Los Angeles	Bronze	11	668	New Goods	1/18/2023
21	31	Male	Seattle	Platinum	74	1484	New Goods	6/21/2023
22	36	Male	Houston	Platinum	35	861	Sports	6/10/2023
23	27	Female	Miami	Bronze	76	916	Tools	6/18/2023