


§3.3.1 再構成ノードの概要

再構成ノードは、名義型フィールドとして定義されたカテゴリ値を元にして、複数の連続型フィールドを派生する場合に使用します。フラグ設定ノードでは、派生されたフィールドには**T**または**F**が保持されますが、再構成ノードでは連続型フィールドの値を保持させる場合に用います。

例えば、下記のように顧客IDごとに購入した商品が記録されている場合、商品フィールドは名義型フィールドになります。このフィールドに基づいて、カテゴリ値ごとに新規フィールドを構成し、それぞれの商品を購入した数量を表す連続型フィールドを構成する場合に使用できます。



ID	製品	数量	ID	パン	牛乳	チーズ	フルーツ
101	パン	3	101	3			
101	牛乳	2	101		2		
102	パン	2	102	2			
102	フルーツ	1	102				1
103	牛乳	5	103		5		

Figure3.3.1 再構成ノードによる再構成結果の例

再構成ノードでは、フラグ設定ノードと異なり連続型フィールドがデーセットに追加されることとなります。なお、該当する値がないセルは欠損値になります。

TIPS

フラグ設定ノードと異なり、再構成ノードでは集計キーによる集約を行うことはできません。再構成の結果について顧客IDなどによる集計を行う場合は、レコード設定パレットに含まれる**レコード集計**ノードを使用してください。

§3.3.2 再構成ノード

再構成ノードは、名義型フィールドとして定義されたカテゴリ値を元にして、複数の連続型フィールドを派生する場合に使用します。

ここでは、**支払方法**として、**現金**、**カード**、**ポイント**の3カテゴリを持つ名義型フィールドに基づいて、これらの**支払金額**を保持するフィールドに変換します。

操作手順

1. **フィールド設定**パレットを開きます。



Figure3.3.2 フィールド設定パレットの再構成ノード

操作手順

2. **再構成**ノードを選択ストリームキャンバスに挿入します。
3. **再構成**ノードを**データ型**ノードからリンクします。

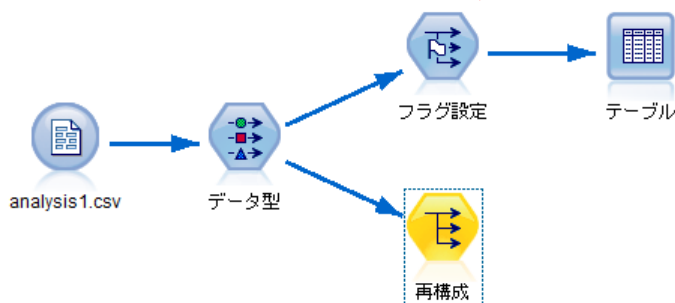


Figure3.3.3 データ型ノードからリンクした再構成ノード

操作手順

4. **再構成**ノードをダブルクリックして編集画面を開きます。

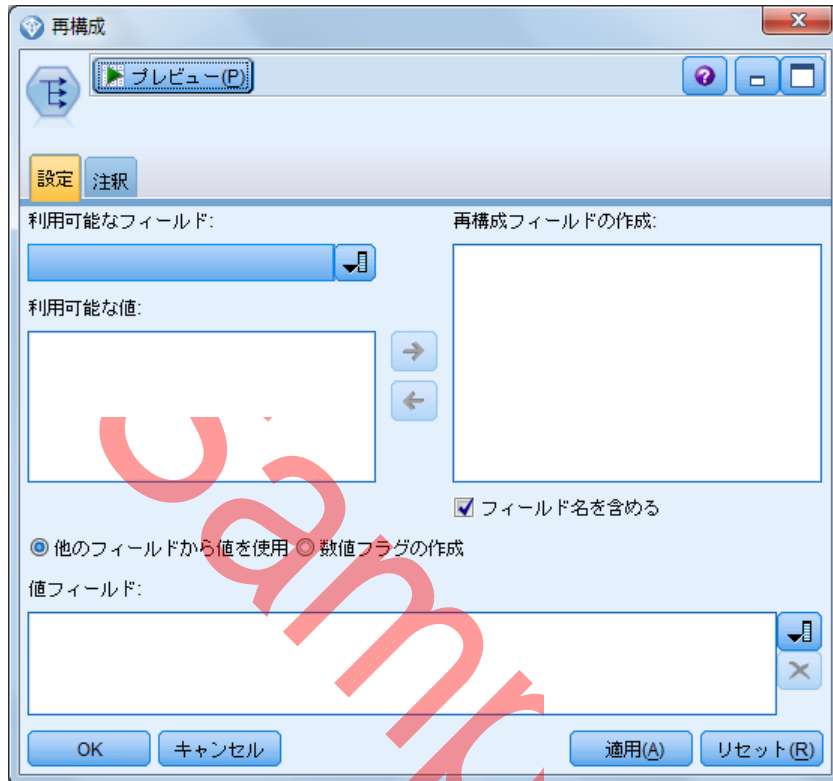


Figure3.3.4 フラグ設定ノードの編集画面

利用可能なフィールドとして、再構成の対象となる名義型フィールドを指定すると、利用可能なセット値にカテゴリ値が表示されます。再構成フィールドとして作成したいカテゴリ値を右側の作成リストに移動します。

操作手順

5. **利用可能なフィールド**のリストから**支払方法**を選択します。

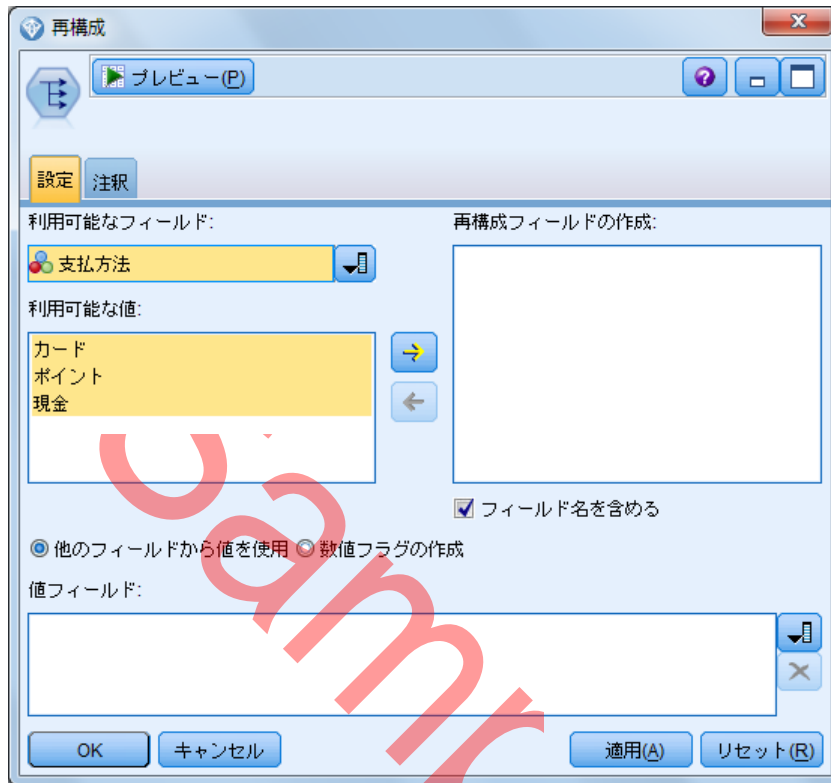


Figure3.3.5 フラグ設定ノードのセット型フィールドの指定

利用可能な値に、支払方法フィールドに含まれるカテゴリ値が表示されますので、再構成が必要なカテゴリ値を、再構成フィールドの作成リストに移動します。

操作手順

6. カード、ポイント、現金を再構成フィールドの作成リストに移動します。

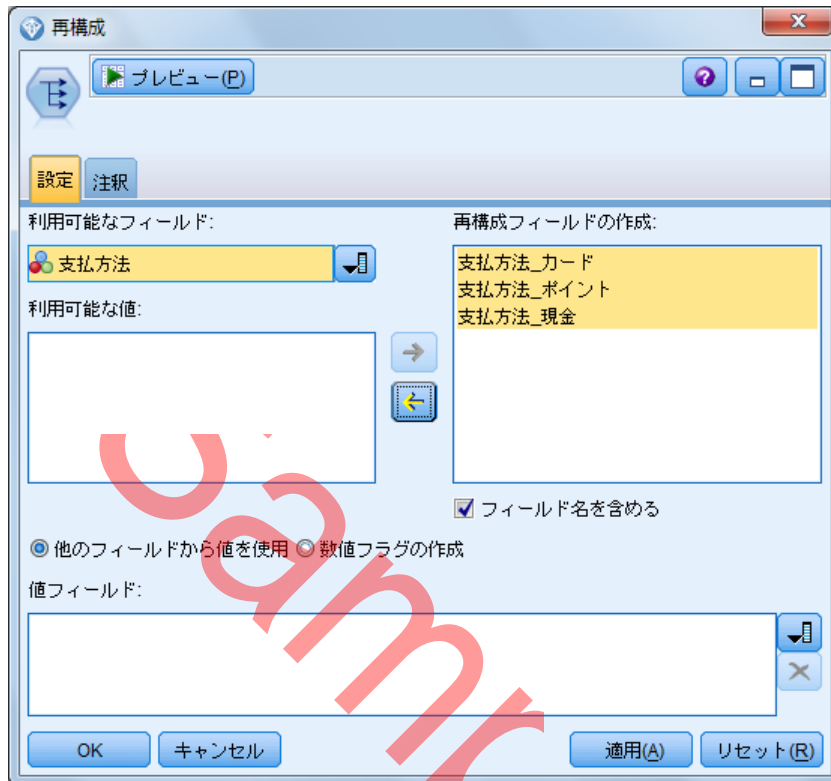


Figure3.3.6 値の挿入ダイアログボックス

カテゴリ値がそれぞれの新規フィールドとして作成されます。次に、再構成の結果として保持する連続型フィールドを**値フィールド**に指定します。この例では、**支払金額**を指定します。

操作手順

7. 値フィールドのリストから**支払金額**を指定します。

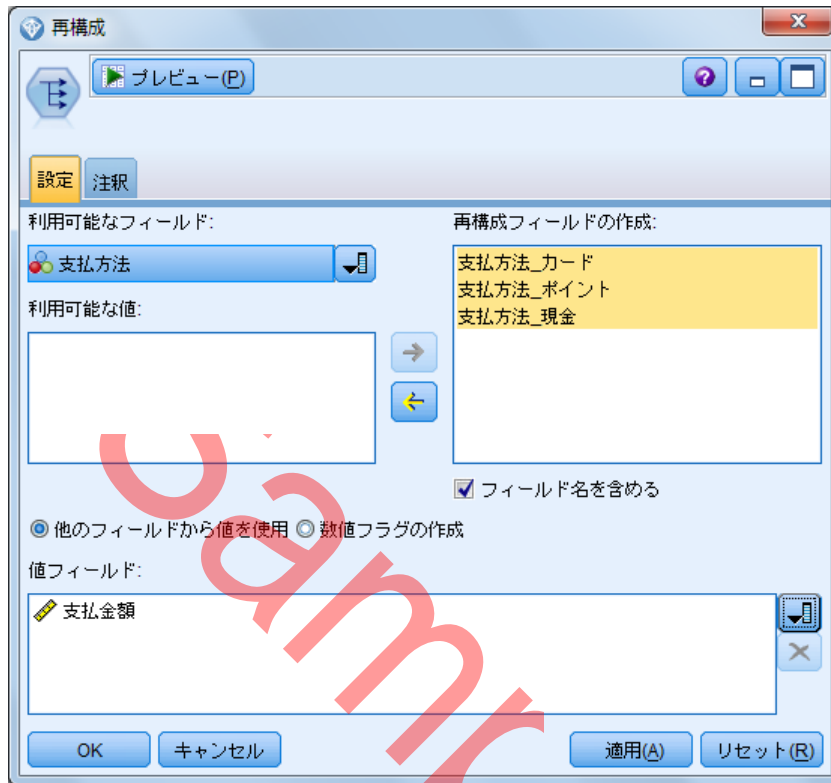


Figure3.3.7 値フィールドの指定

以上の設定で、再構成されたフィールドには支払金額が保持されます。

操作手順

8. **OK**ボタンをクリックします。

操作手順

9. 出力パレットから**テーブル**ノードを挿入し、**再構成**ノードにリンクします。

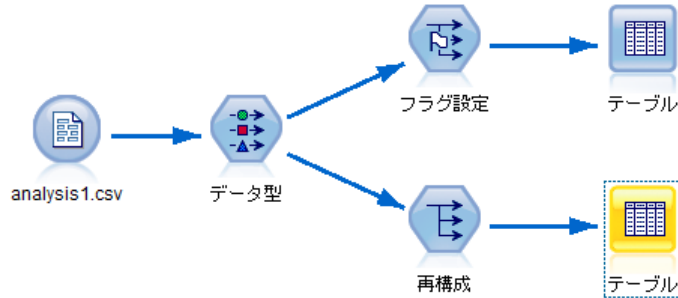


Figure3.3.8 テーブルノードの挿入

操作手順

10. **テーブル**ノードを実行します。

	契約	支払方法	支払金額	支払方法_カード_支払金額	支払方法_ポイント_支払金額	支払方法_現金_支払金額
1	0	ポイント	3179	\$null\$	3179	\$null\$
2	0	現金	9888	\$null\$	\$null\$	9888
3	1	現金	9173	\$null\$	\$null\$	9173
4	0	ポイント	4279	\$null\$	4279	\$null\$
5	1	カード	5720	5720	\$null\$	\$null\$
6	0	カード	6164	6164	\$null\$	\$null\$
7	0	ポイント	4365	\$null\$	4365	\$null\$
8	0	現金	14901	\$null\$	\$null\$	14901
9	1	現金	9440	\$null\$	\$null\$	9440
10	0	現金	9049	\$null\$	\$null\$	9049
11	0	ポイント	4319	\$null\$	4319	\$null\$
12	0	現金	6518	\$null\$	\$null\$	6518
13	0	現金	14281	\$null\$	\$null\$	14281
14	1	カード	10639	10639	\$null\$	\$null\$
15	1	カード	12762	12762	\$null\$	\$null\$
16	1	現金	4644	\$null\$	\$null\$	4644
17	0	現金	5427	\$null\$	\$null\$	5427
18	0	カード	20065	20065	\$null\$	\$null\$
19	0	ポイント	6470	\$null\$	6470	\$null\$
20	0	現金	11751	\$null\$	\$null\$	11751

Figure3.3.9 再構成ノードによって作成されたフィールド(フィールド順序変更済)

該当する値がない場合は、欠損値として\$null\$が含まれます。

操作手順

11. **OK**ボタンをクリックして、**テーブル**ノードを閉じます。

POINT

再構成ノードを使用することで、名義型フィールドのカテゴリ値に基づいて、その他の連続型フィールドの値を保持する新しいフィールドを作成することができます。

Udacity

§3.4.1 置換ノードの概要

置換ノードは、フィールド値の置換やストレージの変更のために使用することができます。CLEM式による条件や、空白値やヌル値を特定の値に変換することもできます。置換ノードは、データ型ノードと組み合わせて、欠損値を置き換えるためによく用いられます。

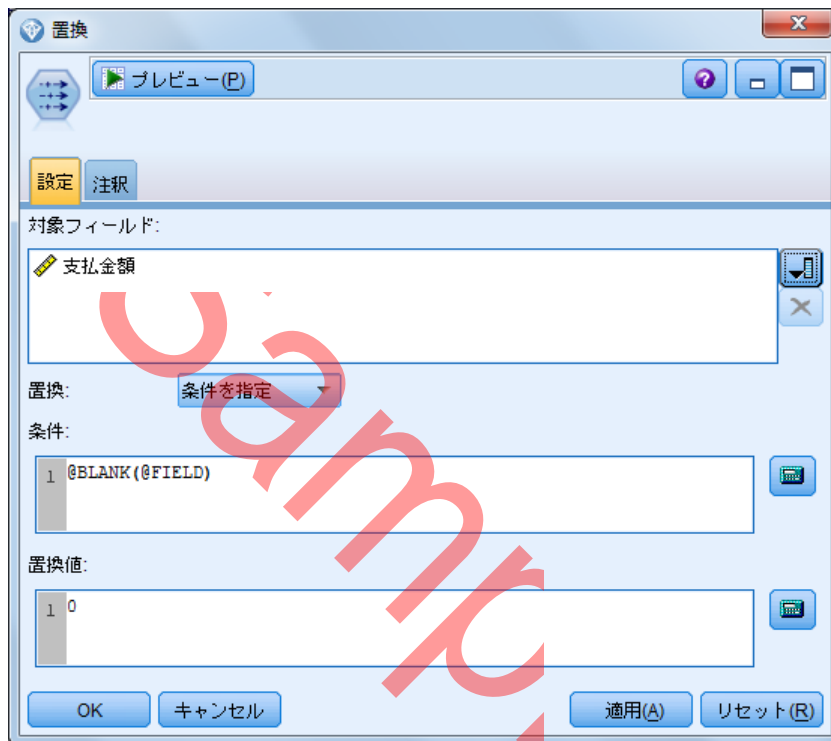


Figure3.4.1 置換ノードによる欠損値を0に置換する例

置換ノードの置換条件を使用すれば、単一または複数のフィールドのストレージタイプを簡単に変換することができます。例えば、関数`to_integer`を使用すれば、文字型のストレージを整数に変換することができます。

TIPS

例えば、`@GLOBAL_MEAN`のような式を指定することで、空白をフィールドの平均値で置き換えることができます。

§3.4.2 置換ノード

ここでは、**再構成**ノードによって作成された、**支払方法_現金_支払金額**、**支払方法_ポイント_支払金額**、**支払方法_カード_支払金額**の3つのフィールドに含まれている欠損値をすべて**0**に置き換えます。

操作手順

1. **フィールド設定**パレットを開きます。



Figure3.4.2 フィールド設定パレットの置換ノード

操作手順

2. **置換**ノードを選択ストリームキャンバスに挿入します。
3. **置換**ノードを**再構成**ノードと**テーブル**ノードの間にバインドします。

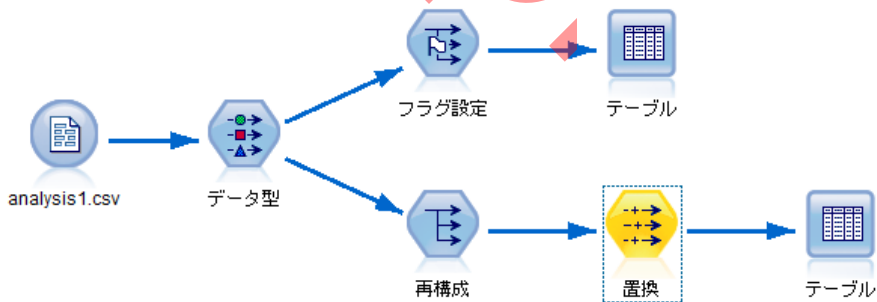


Figure3.4.3 再構成ノードとテーブルノードの間にバインドした置換ノード

操作手順

4. **置換**ノードをダブルクリックして編集画面を開きます。



Figure3.4.4 置換ノードの編集画面

対象フィールドとして、置換が必要なフィールドを指定します。フィールドは一度に複数
を指定することができます。

TIPS

対象フィールドに複数フィールドを指定する場合は、それらのストレージがすべて同じで
ある必要があります。

操作手順

5. 対象フィールドのリストから**支払方法_カード_支払金額**、**支払方法_ポイント_支払金額**、**支払方法_現金_支払金額**を選択します。

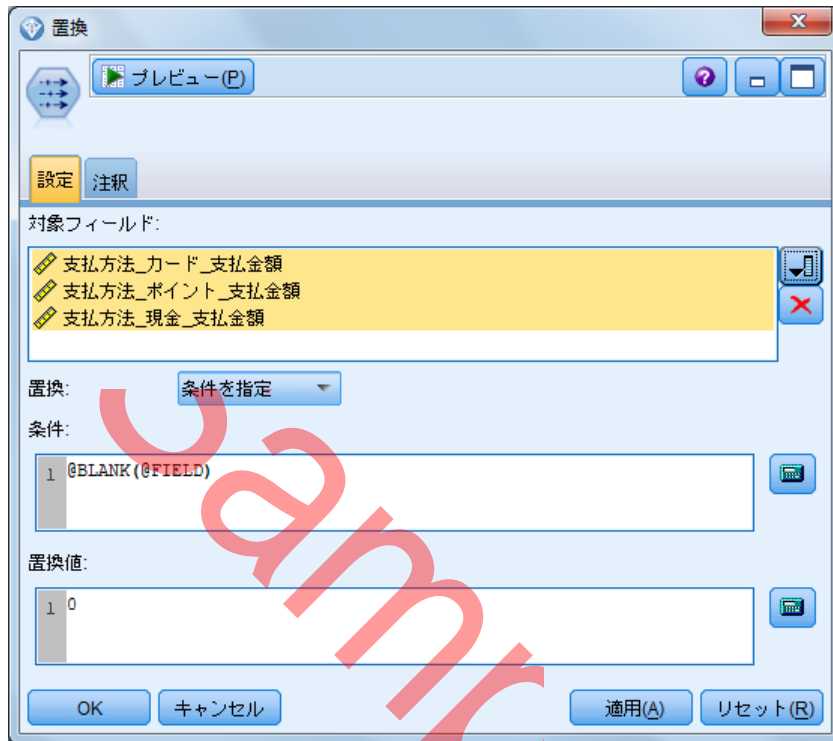


Figure3.4.5 フラグ設定ノードのセット型フィールドの指定

置換の設定として、条件を指定にすると、条件式に合致するレコードが置換対象となります。そのほか、**常時**、**空白値**、**ヌル値**、**空白値とヌル値**を指定することができます。

TIPS

ヌル値は数値型フィールドの欠損値を意味します。また、**空白値**は有効な値を欠損値として指定した値(ユーザー指定の欠損値)を意味します。空白値は、データ型ノードで各フィールドに設定することができます。

この例では、数値型フィールドの欠損値を**0**に変換するため、**置換**に**ヌル値**を指定し、**置換値**はデフォルトの**0**のまま進めます。

操作手順

6. **置換**の**ヌル値**を指定します。

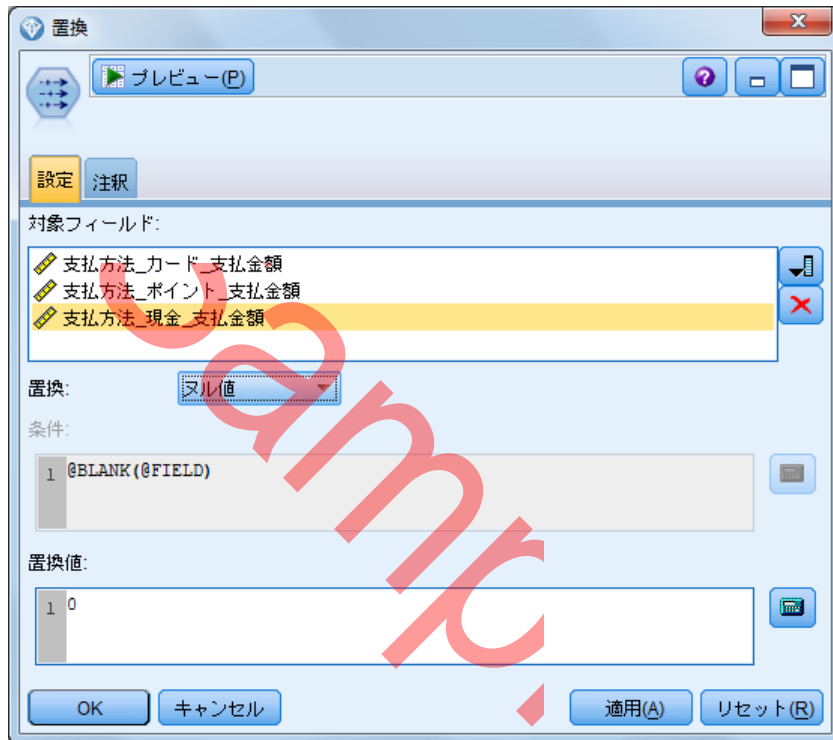


Figure3.4.6 値フィールドの指定

操作手順

7. **OK**ボタンをクリックします。

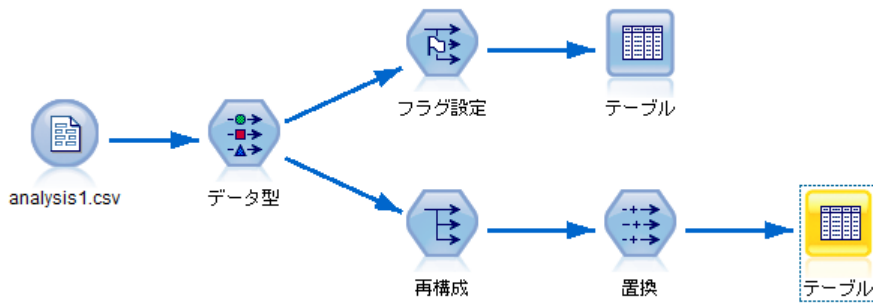


Figure3.4.7 テーブルノードの実行

操作手順

8. テーブルノードを実行します。

テーブル (20 フィールド、 8,632 レコード) #8

ファイル(F) 編集(E) 生成(G)

テーブル 注釈

	支払方法	支払金額	支払方法_カード_支払金額	支払方法_ポイント_支払金額	支払方法_現金_支払金額
1	0 ポイント	3179	0	3179	0
2	0 現金	9888	0	0	9888
3	1 現金	9173	0	0	9173
4	0 ポイント	4279	0	4279	0
5	1 カード	5720	5720	0	0
6	0 カード	6164	6164	0	0
7	0 ポイント	4365	0	4365	0
8	0 現金	14901	0	0	14901
9	1 現金	9440	0	0	9440
10	0 現金	9049	0	0	9049
11	0 ポイント	4319	0	4319	0
12	0 現金	6518	0	0	6518
13	0 現金	14281	0	0	14281
14	1 カード	10639	10639	0	0
15	1 カード	12762	12762	0	0
16	1 現金	4644	0	0	4644
17	0 現金	5427	0	0	5427
18	0 カード	20065	20065	0	0
19	0 ポイント	6470	0	6470	0
20	0 現金	11751	0	0	11751

OK

Figure3.3.9 置換によって変換されたフィールド(フィールド順序変更済)

欠損値\$null\$が0に置き換わっていることが確認できます。

操作手順

9. **OK**ボタンをクリックして、**テーブル**ノードを閉じます。

POINT

置換ノードを使用することで、指定したフィールドの値を、特定の値に置換することができます。例えば、欠損値を**0**に変換することができます。

TIPS

置換値にストレージの**変換関数**を指定することで、フィールドのストレージを変換することができます。変換関数には、**to_integer(ITEM)**、**to_string(ITEM)**、**to_date(ITEM)**などがあります。