

クロス集計表を作成する

ここで使う！

POS 販売データのような形式のデータからクロス集計表を作成します。

どうする？

表形式のデータをリスト化するには、「再配置」アイコンの「横展開」を使用します。

Visual Mining Studio サンプルデータの「個人 ID 付き POS データ」の、「ID」「分類名」ごとに「価格」を合計します。その上で、行方向に「ID」、列方向に「分類名」とした集計金額データを作成します。購買がなかった分類名の列データには、ゼロが入力されます。

【変換前のデータ】

	ID.Key	分類名.Key	価格
1	10001	IT	400.000
2	10001	ノンフィクション	2495.000
3	10001	ビジネス	1690.000
4	10001	ミステリ	1880.000
5	10002	IT	2480.000
6	10002	タレント	490.000
7	10002	ビジネス	2140.000

【変換後のデータ】

	ID.Key	IT	ノンフィクション	ビジネス	ミステリ
1	10001	400.000	2495.000	1690.000	1880.000
2	10002	2480.000	0.000	2140.000	5290.000
3	10003	10935.000	3570.000	4410.000	48115.000
4	10004	2250.000	0.000	0.000	0.000

データは、Visual Mining Studio のサンプルデータである「個人 ID 付き POS データ.vdt」を用います。データ処理の操作に入る前に次のように準備を行ってください。

ワンポイント アドバイス

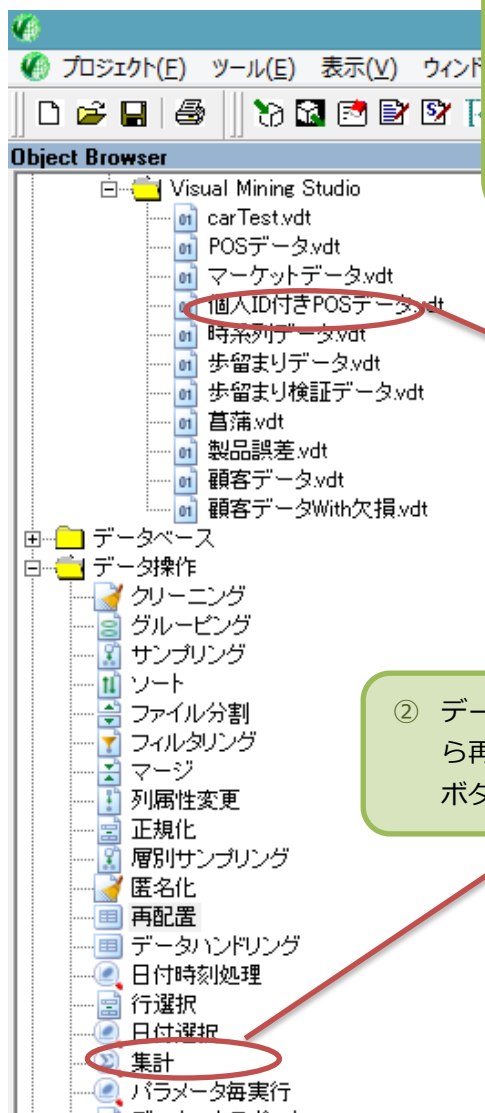
◆ データ量

POS データ、WEB アクセスログなど、【変換前のデータ】のようになっているデータはたくさんあります。一般的に、考えられる組み合わせのパターンは非常に多いものの、実際に出現するパターンは少ない場合には、このようなデータ形式を用いてデータ量を抑えることができます。

POS データ： ユーザー × 商品
WEB アクセスログ： 閲覧者 × WEB ページ

しかし、【変換後】のようなデータに変換すると、実際に出現しなかったパターンにも 0 という値を埋めてしまいますので、結果のデータサイズが大きくなります。データの内容によっては、処理時間と変換後のデータサイズが非常に大きくなってしまう場合があるのでご注意ください。

準備



① 画面左側のオブジェクトブラウザ（Object Browser）から、「データ/Sample/Visual Mining Studio」の「個人ID付きPOSデータ.vdt」と「データ操作」の「集計」「再配置」アイコンを、それぞれプロジェクトボードへドラッグ&ドロップします。

② データアイコンから集計、集計から再配置アイコンへ、マウスの中ボタンで矢印を繋ぎます。

③ 集計アイコンをダブルクリックして、設定画面を立ち上げます。

■ リスト形式のデータを集計する

手順

① 「集計方法」で「サマリ」を選びます。

② 「サマリ項目」で「合計」を選択します。

③ 「集計キー列名」で「ID」、「分類名」を選択します。

※ここで選択した列の項目がクロス集計表の表頭、表側になります。

④ 「集計対象列名」で「価格」を選択します。

【集計】を右ボタンクリック、【データビュー / 開く】を選択すると、次の内容になっています。

	ID.Key	分類名.Key	Summary.Op.Name	価格
1	10001	IT	合計	400.000
2	10001	ノンフィクション	合計	2495.000
3	10001	ビジネス	合計	1690.000
4	10001	ミステリ	合計	1880.000
5	10002	IT	合計	2480.000
6	10002	タレント	合計	490.000
7	10002	ビジネス	合計	2140.000
8	10002	ミステリ	合計	5290.000

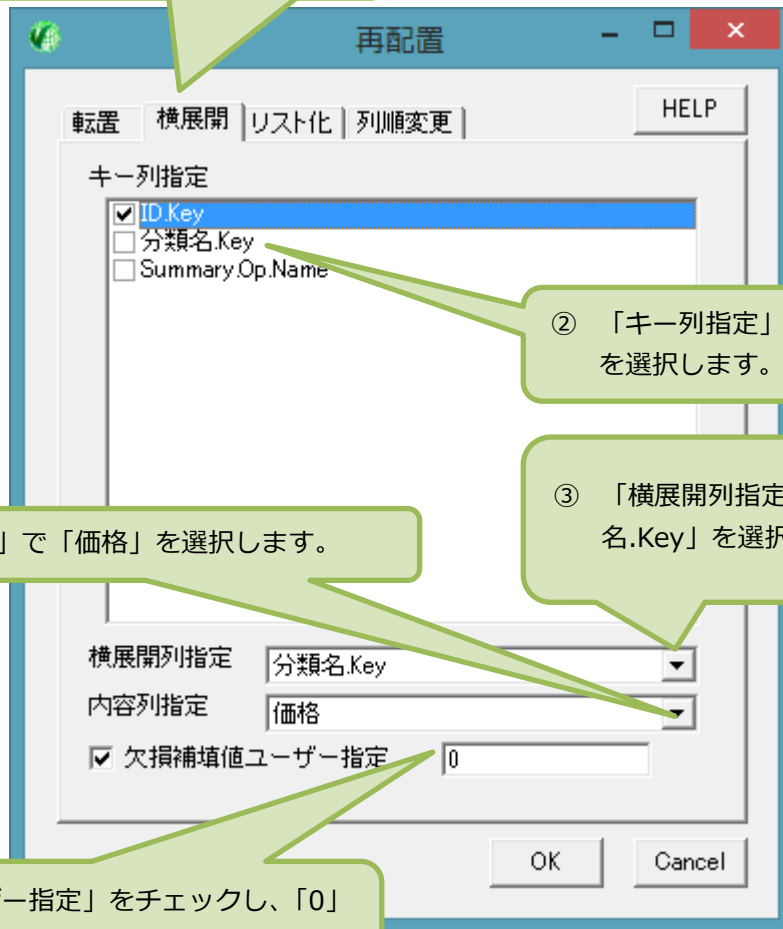
テーブル: 5,918 行 4 列

続いて、クロス表形式のデータにします。

■ 集計結果をクロス表形式データにします

手順

① 「横展開」タブを選択します。



② 「キー列指定」で「ID.Key」を選択します。

③ 「横展開列指定」で「分類名.Key」を選択します。

⑤ 「内容列指定」で「価格」を選択します。

④ 「欠損補填値ユーザー指定」をチェックし、「0」を入力します。

ワンポイント アドバイス

◆ 「キー列」とは？

リストデータで同一人物、あるいは同一観測であることを示すキーとなる列です。通常はIDや連続した番号が使われます。

◆ 横展開列指定

表形式の横方向に展開される列です。新しいデータの列名になります。

◆ 内容列指定

新しいデータのセルの値になる列です。

◆ 欠損補填値ユーザー指定

新しいデータのセルに埋める値がない場合、ここで指定した値を使います。

◆ 集計内容を変更する

この例では合計価格のクロス集計表を作成しましたが、集計アイコンの「サマリ項目」を「平均」に設定すれば、平均価格のクロス集計表を作成することもできます。