

# テーブルから一部の行を取り出す

<b>ここで使う！</b>	特定の商品カテゴリのデータを取り出す、昨年上期のデータを取り出す、属性が不明（空欄）のデータを取り除く、…のように、データテーブルから一部分を抜き出してから集計や分析を行う場合に有効です。また、集計結果や分析結果の一部分を取り出す場合などにも用いられます。
<b>どうする？</b>	テーブルから条件に適合する行を取り出すには、「 <b>データハンドリング</b> 」アイコンの「 <b>行選択</b> 」を使用します。

次の3つの例を挙げて、それぞれ操作手順をご紹介します。

- 任意の商品カテゴリの行を抽出する
- 任意の価格帯の行を抽出する
- 平日の購買データの行を抽出する

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。それぞれの例に入る前に次のように準備を行ってください。

## 準備

The screenshot shows the 'Object Browser' window in Visual Mining Studio. The left pane displays a tree view of data sources and operations. The right pane shows a workspace with two icons: '個人ID付きPOSデータ' (Personal ID Attached POS Data) and 'データハンドリング' (Data Handling). Three callout boxes provide instructions:

- ① 「個人 ID 付き POS データ.vdt」と「データハンドリング」アイコンを、それぞれプロジェクトボードヘッダラッグ&ドロップします。
- ② データアイコンからデータハンドリングアイコンへ、マウスの中ボタンで矢印を繋ぎます。
- ③ データハンドリング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。

## ■ 任意の商品カテゴリの行を抽出する

### 手順

① 対象データとして「個人 ID 付 POS データ」を選択します。

② 「行選択」タブを選択します。

③ この処理で得られるテーブルの名前を記入します。

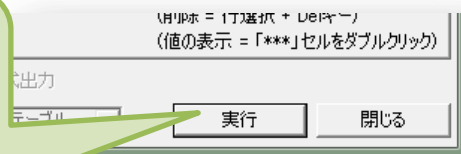
④ 商品カテゴリを表す列を選択します。

⑤ 展開ボタン[+]をクリックし、商品カテゴリ一覧を表示させます。

⑥ 取り出したい商品カテゴリを一覧から選択します。

⑦ 右下の「実行」ボタンを押すと該当商品カテゴリの行が抽出され、新たなテーブルとして追加されます。

⑧ 表示されている「データ&グラフビュー」で内容を確認し、良ければ「閉じる」ボタンを押し、終了します。



### ワンポイント アドバイス

#### ◆ 条件に合致「しない」行の抽出

上の例では、分類名が「IT」または「ビジネス」である商品の行だけを、「IT\_ビジネス」というテーブル名（手順③）で抽出しました。このとき、「条件に合致した結果名」の欄に「IT\_ビジネス」というテーブル名を記述しましたが、となりの「合致しない結果名」に例えば「その他」と記述することで、条件に合致する行と合致しない行を同時に抽出することができます。

#### ◆ 各商品カテゴリの件数

上の例では、手順⑤で商品カテゴリの一覧を表示させました。この一覧の「件数」や「パーセンテージ」の「\*\*\*」をダブルクリックすることで、カテゴリごとの件数と件数の構成比（パーセンテージ）を表示させることができます。

## ■ 任意の価格帯の行を抽出する

### 手順

① 対象データとして「個人 ID 付き POS データ」を選択します。

② 「行選択」タブを選択します。

③ この処理で得られるテーブルの名前を記入します。

④ 価格を表す列を選択します。

⑤ 不等式「 $\geq$ 」を選択します。

⑥ 価格の下限値（円）を入力します。

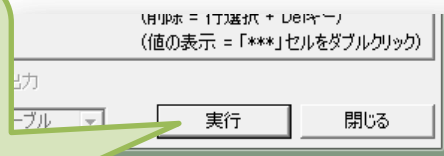
⑦ 続けて価格の上限を条件に加えるために、「条件追加」ボタンを押します。

条件結合	列名	列属性	条件演算	対象	最小
	価格	整数	$\geq$	500	**

⑧ ④～⑥と同様に、価格の上限値を設定します。

条件結合	列名	列属性	条件演算	対象	最小
	価格	整数	$\geq$	500	**
かつ	価格	整数	$\leq$	1000	**

- ⑨ 右下の「実行」ボタンを押すと設定した条件に該当する価格帯の商品（行）が抽出され、新たなテーブルとして追加されます。
- ⑩ 表示されている「データ&グラフィックビュー」で内容を確認し、良ければ「閉じる」ボタンを押し、終了します。



### ワンポイント アドバイス

#### ◆ 平均値以上の価格の商品を抽出する

上の例では、手順⑥で価格の下限値を「500（円）」と入力しました。この右側にある「平均」欄の「\*\*\*」をダブルクリックすることで、全体の平均値を表示させることができます。さらにその表示された平均値をクリックすると、「500」と記述した欄（「対象」欄）がこの平均値で置き換わります。この機能を利用することで、簡単に平均値などの統計量が算出でき、その値で条件付けすることができます。

## ■ 平日の購買データの行を抽出する

次のような手順を踏みます。

1. 「平日」を特定するために、休日フラグを各行に付加します。
2. 付加した休日フラグを頼りにして、平日の行だけを抽出します。

### 手順1 休日フラグを付加

① 対象データとして「個人 ID 付き POS データ」を選択します。

② 「日付時刻処理」タブを選択します。

③ この処理で得られる結果の名前を記入します。

④ 「列名」にチェックを入れ、日付の列「年月日」を選択します。

⑤ 日付の形式を選択します。

元データ出力

⑥ 「元データ出力」にチェックを入れることで、この処理の結果に、元のデータを含ませることができます。

⑦ 「分解出力」にチェックを入れ、「休日フラグ」にチェックを入れます。

分解出力

出力項目

<input type="checkbox"/> 年	<input checked="" type="checkbox"/> 休日フラグ
<input type="checkbox"/> 月	<input type="checkbox"/> 休日名
<input type="checkbox"/> 日	<input type="checkbox"/> 曜日
<input type="checkbox"/> 時	<input type="checkbox"/> 月末
<input type="checkbox"/> 分	<input type="checkbox"/> 次月初
<input type="checkbox"/> 秒	
<input type="checkbox"/> 午前・午後	
<input type="checkbox"/> 年通算週	
<input type="checkbox"/> 月通算週	
<input type="checkbox"/> 年通算日	

曜日形式  
月,火,水,木,金,土,日,Error

期間計算

終期

列名 ID  
(日付形式=自動検出)

日時 2014年06月13日 16:29:15

計算の向き 始期から終期

出力項目

<input type="checkbox"/> 年数	<input type="checkbox"/> 満年数
<input type="checkbox"/> 月数	<input type="checkbox"/> 満月数
<input type="checkbox"/> 週数	<input type="checkbox"/> 満週数
<input type="checkbox"/> 日数	
<input type="checkbox"/> 時数	
<input type="checkbox"/> 分数	
<input type="checkbox"/> 秒数	

⑧ 右下の「実行」ボタンを押すと、「休日フラグ」列が付加され、新たなテーブルとして追加されます。

⑨ 表示されている「データ&グラフィックビュー」で内容を確認し、次の手順へ進みます。

実行

閉じる

## 手順2 平日の行だけ抽出

① 手順1の結果を選択します。

② 「行選択」タブを選択します。

③ この処理で得られる結果の名前を記入します。

④ 「休日フラグ」列を選択します。

⑤ 展開ボタン[+]をクリックし、「休日フラグ」の一覧を表示させます。

⑥ 「平日」にチェックを入れます。

⑦ 右下の「実行」ボタンを押すと設定した条件に該当する購買データ（の行）が抽出され、新たなテーブルとして追加されます。

⑧ 表示されている「データ&グラフィックビュー」で内容を確認し、良ければ「閉じる」ボタンを押し、終了します。

条件に合致した結果名	条件結合	列名	列属性	条件演算	対象	最小
平日の購買データ		休日フラグ	文字列	次のいずれか	平日	

選択	内容	件数	パーセンテージ
<input type="checkbox"/>	休日	***	***
<input checked="" type="checkbox"/>	平日	***	***

## ワンポイント アドバイス

### ◆ 一覧表示されるのは「文字列」だけ

上の例では、展開ボタン[+]をクリックすることで「休日フラグ」の一覧を表示させました。この一覧表示機能は対象の列（「列名」で選択された列）が文字列である場合にのみ有効です（「休日フラグ」は、「休日」「平日」という2種類の文字列で表現されている）。「列名」で選択された列が数値（整数または実数）の場合には、「条件演算」の欄に等式または不等式を指定して条件設定を行います。

### ◆ 任意の期間のデータを抽出する

上の例では、手順1-⑦で「休日フラグ」にチェックを入れることで、元データに「休日フラグ」の列を加え、「平日」の判定ができるようにしました。これと同様にして、「月」にチェックを入れることで、月を表す列を加えることができ、例えば上期（4月～9月）のデータを抽出する、ということが可能になります。