

PowerQuery

written by *M. Hosoda*

OutLine

Power Queryは、外部データの取得、加工、整形を自動化するツール。

様々な形態で散在するデータを「抽出 (Extract)」「変換・加工 (Transform)」「格納 (Load)」し、分析可能な形に一元化するのでETLツールと呼ばれる。

Excelに搭載されているが、**PowerBI**のデータ処理部分もPowerQuery。

SPOのファイルリストをExcelに出力してダウンロードできるが、実際にダウンロードされるのはPowerQueryのクエリデータで、ローカルではそれをPowerQueryで展開する。

PowerPivotという強力なデータ分析ツールがある。通常のピボットテーブルとの大きな違いは、ピボットテーブルはシートのデータを対象とするが、PowerPivotは**データモデル**というオブジェクト(の塊)を対象にする。

シートのデータは当然データ数に制限があるが、データモデルはデータベースのテーブルやそれらをリレーションしたクエリの集合のようなもので、理論上はデータ数制限はない。(メモリ内に展開されるので、メモリの制限はある)

シートのデータは関連するデータをつなげるには、Vlookup(Xlookup)などの関数で処理するしかないが、データモデルではリレーションで解決されている。

PowerQueryはまさにそのデータモデルを作れるツールなので、PowerPivotの前段階として使える。(PowerBIと同じ立場)

いろいろなデータソースをつなぎ合わせて、問い合わせることができる。

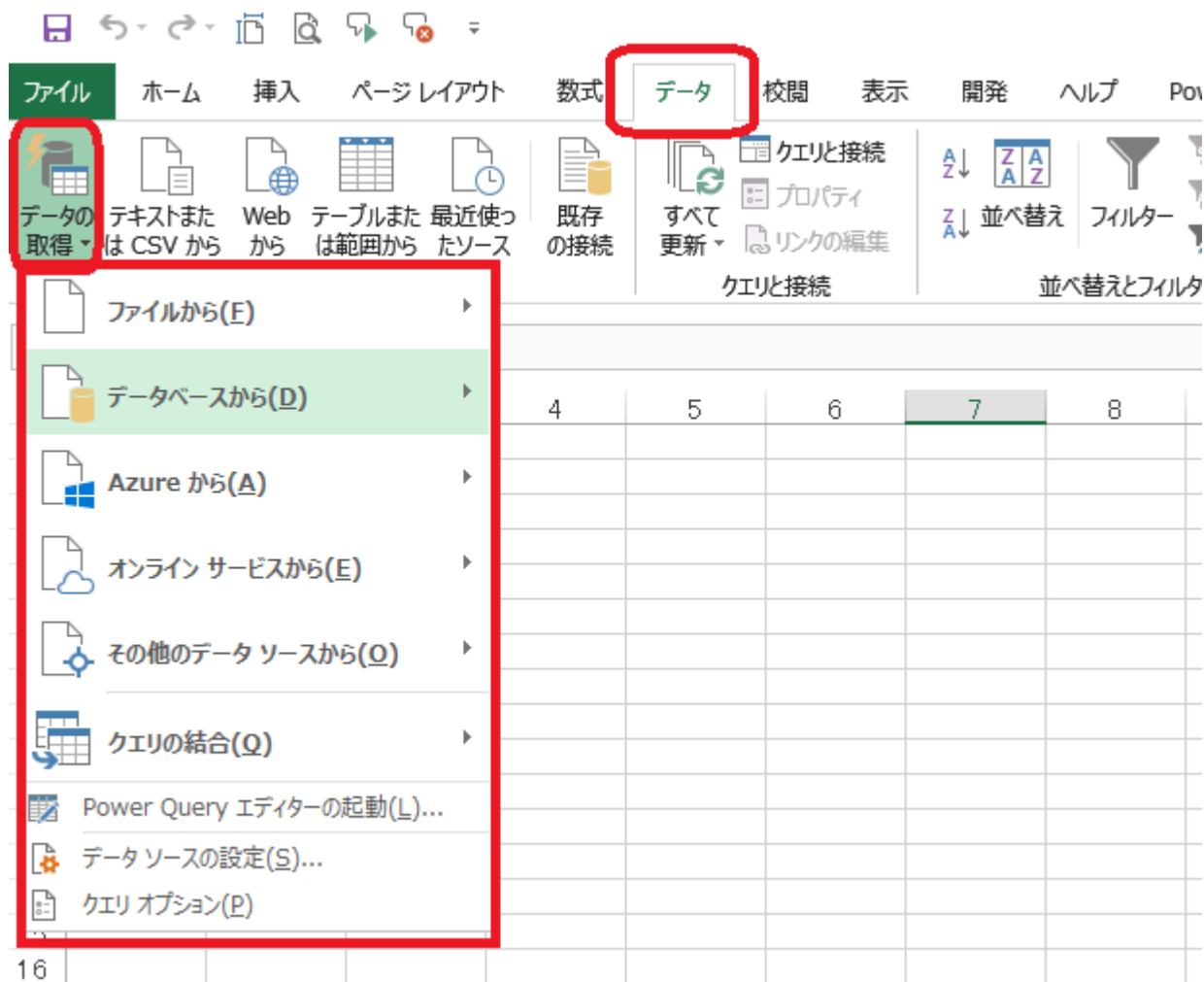
また、アクセスにあるデータをPowerQueryから利用するというのも便利かもしれない。

いろいろなデータソースに対応しているので、e-Statのデータ(RESTful API)への対応も強力。

ということで、PowerQueryに慣れておくことは有用だと考える。

利用方法

Excelから



基本操作

(ネットからのパクリ ☹️) <https://youseful.jp/microsoft/excel/powerquery-1/>

データ取得

ファイルを指定するのが普通だが、フォルダに同じ様式のExcel Bookが入っている場合フォルダごと読むことができる。

[データ] タブ → [データの取得] → [ファイルから] → [フォルダから] で、そのフォルダ内のエクセルを全部開いてくれる。

データ編集

列の削除

不要な列を選択して右クリックで削除(Deleteでも可)

列を加工して追加

[列の追加] タブで 条件処理や演算を施して新しい列を作成できる。

粗利=売上高-原価 の例。

「売上高」を選択して、[Ctrl] を押しながら「原価」を選択（複数選択）

[列の追加] タブ→ [標準] → [減算]

タイトルを粗利に変更

Sort

列の [▼] をクリックして [昇順で並べ替え] を押すと小さい値順に並び変わり、[降順で並べ替え] を押すと大きい値順に並び変わる。

出力

指定したエクセルのシートに出力

[ホーム] タブ→ [閉じて次に読み込む]

[テーブル] を選択

[既存のワークシート] を選択して先頭アドレスを入力してOK

記録

自動化のため、操作が記録されている。

上の方で取り消すと、それ以降の操作が全部ロールバックされる。

データ更新

対象のフォルダに新しいデータが追加されたら、[クエリタブの「更新」](#) で自動的に追加される。

ユースケース

PDFからの取り込み

PDFファイルを直接読み込んでPDFに含まれるテーブルデータを取得してテーブル化が可能。[Webから] を用いてURLを指定すれば、Webサイト上の表形式のデータ（PDFファイル含む）を読み込める。（相手によると思う。ビックカメラは接続できなかった）

データのプレビュー

Power Queryエディタ内の[表示] タブ → [データのプレビュー]内のチェックボックスをチェック。（あまり使わないと思うけど）

データ型の変換・統一

取り込んだ列のデータ型が“abc”や“123”のアイコンで表示
[変換] → [データ型] から適切な型を選択

文字コード自動変換

ExcelはShiftJISでないと文字化けする。(最新バージョンは大丈夫) PowerQueryは文字コードがShiftJISでなくても大丈夫

ピボット解除

非正規化された繰り返し構造を正規化するみたいな。
3文字ファイルを縦構造に変換できるということ。

データの重複を削除

データに重複レコードがある、あるいは欠損値が含まれている場合に重複を削除したり、欠損を一括置換したりすることで、分析に適したクリーンなデータを得られる。
操作の記録が残り元に戻せるので、やりやすい。

データモデルの作成

Excelのワークシートに読み込むことのできない104万行を超えるデータであっても、「接続のみ」、「データモデル」にチェックを入れた状態で読み込みを行える。
データモデルはPowerPivotで利用できる。

e-Stat対応

例えば <https://www.e-stat.go.jp/stat-search/file-download?statInfId=000032189720&fileKind=4> をWebからとってみる。ファイルをダウンロードせず、データを取得できる。
この過程はクエリとして登録できるので、URLをパラメータとして取り換えるだけで、毎月の処理ができる。
これはVBAで管理できる。

以上