東京都データプラットフォーム

行政データ整備モデル事業

データ整備マニュアル

第 1 版

発行元：東京都

委託先：日本総合研究所、株式会社自動処理

2022 年 3 月

改訂履歴

版	日時	作成者/更新者	主な改訂内容
1.0	2022. 3. 31	株式会社自動処理/高木祐介	初版作成

目次

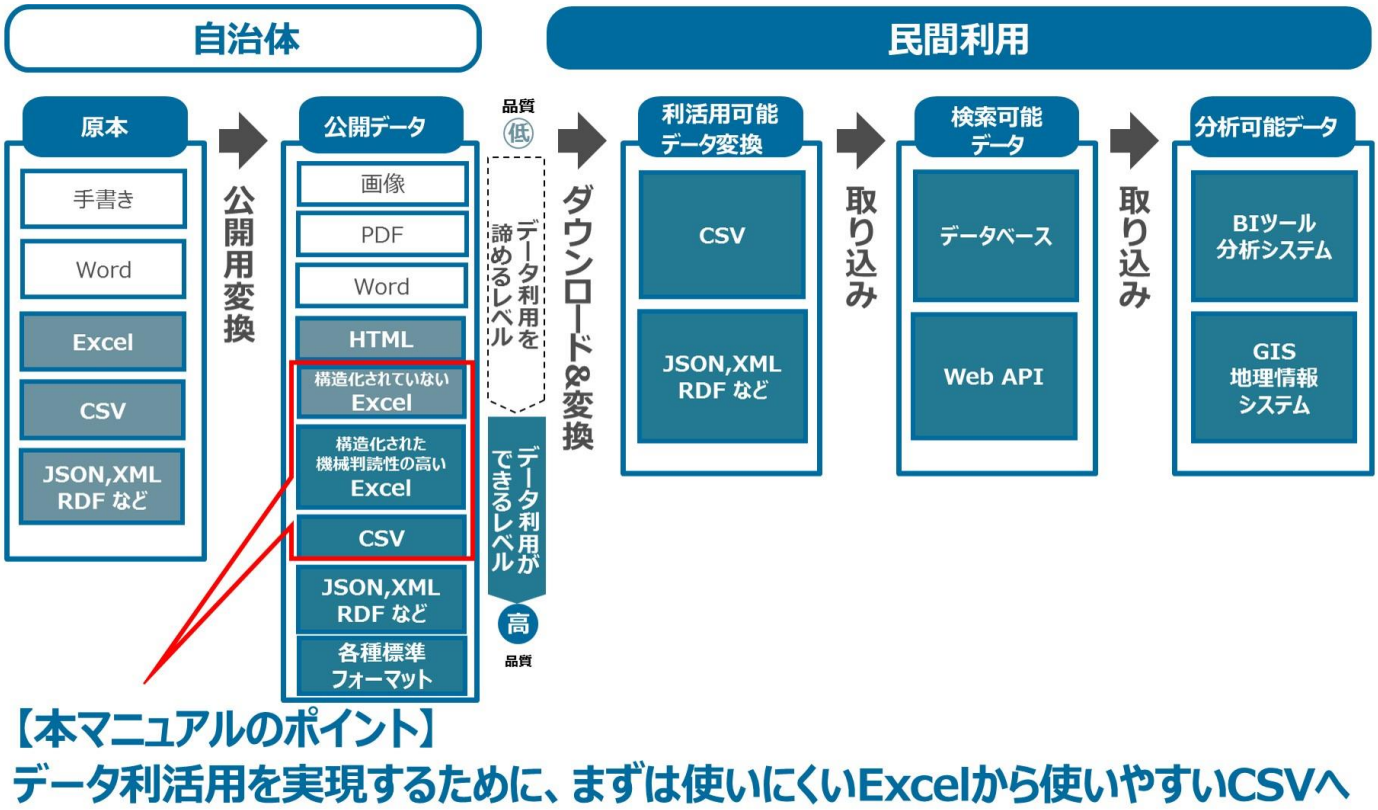
手順別目次

手順0 データ活用までのフローを知る

データ整備モデル事業について

民間利活用を目的にしたデータ化事業では
取り込まれたデータが、「業務システム用のマスタ」、「データ分析用の基礎データ」等として利用されます。企画の際の一過性な分析を行うケースを除き、民間利活用を考える場合、**簡単かつ安定的にデータベースに取り込む**ことができなければ、営業活動や、市場調査など実務に使う事が出来ません。その為、**データを公開する際には、データ利用者が利活用しやすい形（機械判読可能な形式）で提供する必要**があります。

データ化作業は、データベースを取り扱う**専門的な知識とスキルが必要**ですが、まずは**データ整備から活用までの一連の流れを把握することが重要**です。
例えば、手順４「データを適切に配置し、データの精度を高める」で、データベースのクレンジング作業は、データベースを精度良く取り扱うことのできる外注先（専門家）への発注や API の調達を前提にマニュアルを作成しています。
（マッピング、正規化する場合や、工程を自動化する場合も同様です。）



データ整備事業に使用するツール
本事業でのデータ化作業は、区市町村担当向けを想定している為、マニュアルや目視作業が必要な作業については、Excel で出来る範囲は極力 Excel にて対応を行うことを想定しています。
Excel 以外のツールについては**「作業に使用するアプリケーション/アドインを準備する」（P.5）**に記載しています。
またデータクレンジングなど作業に膨大な手間がかかるような処理については、API や外注など調達により作業を自動化する事を推奨します。

データ精度の扱い
Excel 等に格納されたデータは、安定的にデータベースに取り込むまでにさまざまな処理が必要になります。
正規化の過程では、人力でデータを加工したり、正規化ツールを使ったりする場面があります。東京都は 100%の精度で正規化することを必須とは考えていません。

作業に使用するアプリケーション/アドインを準備する

アプリケーション

本マニュアルに記載された手順では、以下のアプリケーションを使用します。一部、有償アプリケーションの入手が必要になります。
また、指定されたアプリケーションをお持ちでなくても、同等の機能を持つ互換アプリケーションを利用しても差し支えありません。

Microsoft Excel（互換アプリケーション含む）

バージョン指定：なし
CSV ファイルを取り扱える互換アプリケーションであれば、本書記載の操作は可能です。
Excel や互換アプリケーションで CSV ファイルを取り扱う場合は、電話番号等先頭に 0（ゼロ）を含むデータに気をつける必要があります。XLS 形式で保存すると、ゼロが消えてしまう場合があります。
データの誤変換を避けるためには、より本書の作業に適した CSV エディターでデータを変換いただくことをおすすめします。
例：CSV エディター
https://www.asukaze.net/soft/cassava/

CLOVA OCR Reader / AI-OCR 用

バージョン指定：なし
数字や文字をデータとして取り込めない場合は、OCR 機能を利用してデータを生成させることが必要になります。
本書では CLOVA OCR （Reader）の AI-OCR 機能を利用して、精度よくデータを生成する手順を案内しています。
詳細はこちら
https://clova.line.me/clova-ocr/

Adobe Acrobat DC（互換アプリケーション含む）/ AI-OCR 用

利用の目的：AI-OCR に取り込むファイルを生成
想定作業：保存、ページの回転、書き出し、の圧縮、アクションリストの利用
変換後のバージョン：Acrobat10.0
本書では Adobe Acrobat DC の書き出し機能を利用し、AI-OCR に文字が読み込まれやすいデータを生成する手順を案内しています。
詳細はこちら
https://www.adobe.com/jp/acrobat/features.html

アドイン

Power Query（Microsoft Excel の標準またはアドイン）

バージョン：**Excel2016 では標準機能**（追加インストール不要）。 Excel2010、Excel2013 ではアドインとしてインストール必要。
Power Query (以前のバージョンの Excel でデータを取得& 変換と呼ばれる) を使用すると、外部データをインポートまたは接続し、そのデータをニーズに合った方法で(列の削除、データ型の変更、テーブルの結合など) 整形できます。
本書では手順 2 で、Excel でデータを利用する際に使用します。
詳細はこちら
https://support.microsoft.com/ja-jp/office/excel-%E3%81%AE-power-query-%E3%81%AB%E3%81%A4%E3%81%84%E3%81%A6-7104fbee-9e62-4cb9-a02e-5bfb1a6c536a

手順1 データを変換する前に準備する（事前準備）

データ構造・形式統一前の準備として、データ形式別に変換処理を実施します。

1. **データ構造の検討** （参考 p. 7）
   データの利用目的等によって、必要とされるデータ構造は異なります。具体的に求められているデータ形式が定まっていない場合、データ構造がどうあるべきかを検討します。

2. **データ構造の調査** （参考 p. 8）
   検討したデータ構造に利用するデータ構造を調査します。
   ① **メタデータ調査** メタデータルールと利用イメージの検討等を参照します。メタデータとは、あるデータに関する属性情報などの付帯的なデータのことです。データのタイトルや説明、公開日などデータを検索する際などに使用します。
   ② **データ構造調査** 政府 CIO ポータルに配置された標準データセット等をはじめ、各種データ連携モデルやデータカタログサイト、公開 API 等を参照します。

3. **データフォーマットの検討** （参考 p. 14）
   推奨データセットや収集データを比較し、妥当なフォーマットを検討します。
   ① メタデータ検討
   ② 標準 CSV フォーマット検討

4. **出典元の収集** （参考 p. 18）

5. **機械判別可能なデータ作成のチェックポイント** （参考 p. 18）

6. **ケース一覧** （参考 p. 20）

1. データ構造の検討

データの利用目的等によって、必要とされるデータ構造は異なります。データ構造がどうあるべきかを検討します。

関連部署、団体から情報を収集

ここでは医療機関データをデータ化するための情報として、保健所や医師会からの診療所データを入手する例を示します。
入手元だけでなく、データの対象によっても、データの構造が異なっていることがあります。これからデータ化する Excel ファイルをどのデータ構造にするのかを検討します。

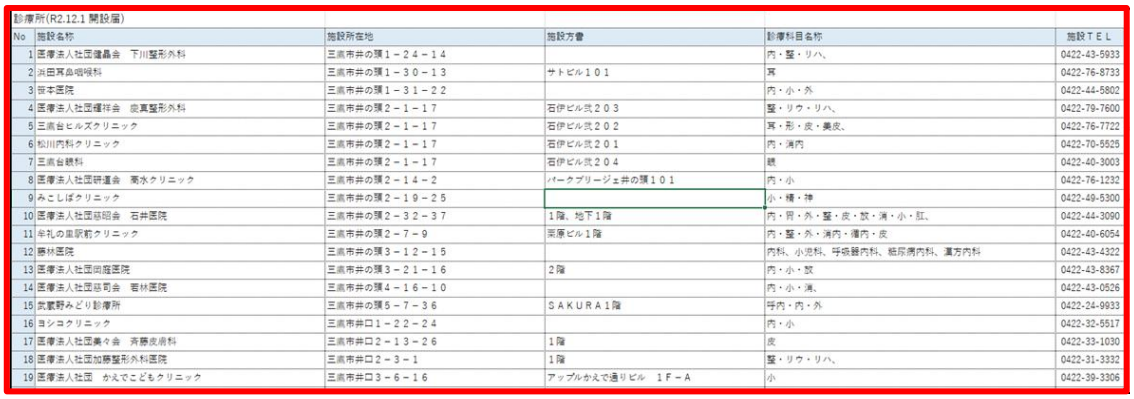
保健所からの病院データ

病院(R2.12.1 開設届)

No	施設名称	施設所在地	施設方書	診療科目名称	施設ＴＥＬ
1	医療法人社団宗理会　東京国際大畑病院	三鷹市下連雀４丁目８番４０号		内科、麻酔科、泌尿器科、放射線科、循環器内科、病理診断科、婦人科	0422-47-1000
2	医療法人財団紘友会　三鷹病院	三鷹市下連雀５丁目１番１２号		内科、外科、整形外科、眼科、リハビリテーション科、泌尿器科	0422-47-6101
3	篠原病院	三鷹市下連雀６丁目１３番１０号		内科、消化器科	0422-46-2251
4	医療法人財団慈生会　野村病院	三鷹市下連雀８丁目３番６号		外科、内科、整形外科、リハビリテーション科、形成外科、臨床検査科、神経内科、脳神経外科、糖尿病内科、循環器内科、呼吸器内科、放射線診断科、漢方内科、腹和ケア内科、内視鏡内科、脳細内科	0422-47-4848
5	公益財団法人　井之頭病院	三鷹市上連雀４丁目１４番１号			0422-44-5331
6	医療法人社団永寿会　三鷹中央病院	三鷹市上連雀５丁目２３番１０号		内科、消化器内科、消化器外科、循環器内科、外科、整形外科、脳神経外科、眼科、リハビリテーション科、放射線科、皮膚科、泌尿器科、呼吸器内科、呼吸器外科、神経内科、肛門外科、血液内科、乳腺外科、麻酔科、腫瘍内科、内分泌科	0422-44-6161
7	吉林大学医学部付属病院	三鷹市新川６丁目２０番２号		内科、小児科、皮膚科、精神科、外科、脳神経外科、整形外科、心臓血管外科、形成外科、小児外科、泌尿器科、眼科、耳鼻いんこう科、産科、婦人科、放射線科、麻酔科、歯科口腔外科、リハビリテーション科、美容外科、リウマチ科、呼吸器外科、神経内科、救急科、病理診断科、循環器科、呼吸器科、消化器科	0422-47-5511
8	医療法人社団裕友会　長谷川病院	三鷹市大沢２丁目２０番３６号		精神科、歯科、内科、心療内科	0422-31-8600

データ構造①

保健所からの診療所データ



診療所(R2.12.1 開設届)

No	施設名称	施設所在地	施設方書	診療科目名称	施設ＴＥＬ
1	医療法人社団健昌会　下川整形外科	三鷹市井の頭１－２４－１４		内・整・リハ、	0422-43-5933
2	浜田耳鼻咽喉科	三鷹市井の頭１－３０－１３	サトビル１０１	耳	0422-76-8733
3	笹本医院	三鷹市井の頭１－３１－２２		内・小・外	0422-44-5802
4	医療法人社団耀祥会　廣真整形外科	三鷹市井の頭２－１－１７	石伊ビル弐２０３	整・リウ・リハ、	0422-79-7600
5	三鷹台ヒルズクリニック	三鷹市井の頭２－１－１７	石伊ビル弐２０２	耳・形・皮・美皮、	0422-76-7722
6	松川内科クリニック	三鷹市井の頭２－１－１７	石伊ビル弐２０１	内・消内	0422-70-5525
7	三鷹台眼科	三鷹市井の頭２－１－１７	石伊ビル弐２０４	眼	0422-40-3003
8	医療法人社団研遵会　南水クリニック	三鷹市井の頭２－１４－２	パークプリージェ井の頭１０１	内・小	0422-76-1232
9	みこしぼクリニック	三鷹市井の頭２－１９－２５		小・精・神	0422-49-5300
10	医療法人社団延昭会　石井医院	三鷹市井の頭２－３２－３７	１階、地下１階	内・胃・外・整・皮・放・消・小・肛、	0422-44-3090
11	幸礼の里駅前クリニック	三鷹市井の頭２－７－９	岡原ビル１階	内・整・外・消内・循内・皮	0422-40-6054
12	藤林医院	三鷹市井の頭３－１２－１５		内科、小児科、呼吸器内科、糖尿病内科、漢方内科	0422-43-4322
13	医療法人社団岡庭医院	三鷹市井の頭３－２１－１６	２階	内・小・放	0422-43-8367
14	医療法人社団慈司会　若林医院	三鷹市井の頭４－１６－１０		内・小・消、	0422-43-0526
15	気観野みどり診療所	三鷹市井の頭５－７－３６	ＳＡＫＵＲＡ１階	呼内・内・外	0422-24-9933
16	ヨシコクリニック	三鷹市井口１－２２－２４		内・小	0422-32-5517
17	医療法人社団廣々会　斉藤皮膚科	三鷹市井口２－１３－２６	１階	皮	0422-33-1030
18	医療法人社団加藤整形外科医院	三鷹市井口２－３－１	１階	整・リウ・リハ、	0422-31-3332
19	医療法人社団　かえでこどもクリニック	三鷹市井口３－６－１６	アップルかえで通りビル　１Ｆ－Ａ	小	0422-39-3306

データ構造②

2. データ構造の調査

検討したデータ構造に利用するデータ構造を調査します。

- **メタデータ調査** メタデータルールと利用イメージの検討等を参照します。
  A：メタデータルールと利用イメージの検討等を参照　https://cio.go.jp/dp2021_07 (p. 8)
- **データ構造調査** 各種データ連係モデルやデータカタログサイト等を参照します。
  B：標準データセット等を参照　https://cio.go.jp/policy-opendata
  C：政府・自治体のデータカタログサイトを参照　https://datasetsearch.research.google.com/
  D：政府・自治体のホームページを参照　https://www.google.com

メタデータ調査（A：メタデータ調査ルールと利用イメージの検討等を参照）

政府 CIO ポータルサイトのメタデータルールと利用イメージの検討（ https://cio.go.jp/dp2021_07 ）にアクセスします。
BregDCAT(https://joinup.ec.europa.eu/collection/access-base-registries/solution/abr-bregdcat-ap)を基本としてメタデータを検討します。
サンプルの含まれる日本語の資料として、メタデータ比較検討用コンセプトペーパー（ https://cio.go.jp/sites/default/files/uploads/documents/dp2021_07_att.pdf ）がある為、参照しながら検討します。
※BregDCAT 方式の場合、データカタログサイトのデファクトスタンダードである CKAN のマッピングを行う事が出来る為、データカタログサイトに取り込む際には以下を参照に取り込む事としてください(https://github.com/ckan/ckanext-dcat#rdf-dcat-to-ckan-dataset-mapping)。

『2.4 メタデータの記載例』に基づきメタデータについて検討します。

記載例	
カタログ	
管理 ID	CA00001
タイトル	文字情報基盤
サブタイトル	
説明	行政で用いられる人名漢字、法人用漢字、かな等の約 6 万文字の漢字です。
キーワード	漢字、戸籍統一文字、住民基本台帳ネットワーク統一文字、登記統一文字、変体仮名
テーマ分類	文字
対象地域	全国
提供者	文字情報技術促進協議会
公開日	2011-10-26
最終更新日	2020-8-26
更新頻度	不定期
言語	ja
公開範囲	公開
公開条件	
ライセンス	IPA フォントライセンス v1.0
利用規約	クリエイティブ・コモンズ 表示 – 継承 2.1 日本 ライセンス条件
ホームページ	https://moji.or.jp/mojikiban/
カタログレコード	
トピックス	文字
説明	行政で用いられる人名漢字、法人用漢字、かな等の約 6 万文字の漢字です。
公開日	2011-10-26

記載例	
最終更新日	2020-8-26
言語	ja
データセット	
管理 ID	CA00001-DST001
タイトル	MJ 文字情報一覧表
サブタイトル	
バージョン	Ver.006.01
説明	各文字に関するコード、読み、字母、画数等をまとめた情報。
キーワード	フォント、ヨミガナ、画数
対象地域	全国
対象期間	
分類	全ての業務
提供者	文字情報技術促進協議会
作成者	情報処理推進機構
連絡先情報	組織名：一般社団法人 文字情報技術促進協議会 メールアドレス：info@moji.or.jp フォーム URL：https://moji.or.jp/about/contact/
タイプ	Strict Open XML
来歴情報	2020 年 10 月に、情報処理推進機構から文字情報技術促進協議会に信託譲渡
品質評価	正確性、網羅性
品質測定結果	公務で使うのに十分な品質
公開日	2011-10-26
最終更新日	2020-8-26
更新頻度	不定期
言語	ja
公開範囲	公開
公開条件	
準拠する標準	
関連ドキュメント	
ランディングページ	
データサービス	
管理 ID	DSV00001
タイトル	文字情報基盤検索サービス
説明	
キーワード	
対象データセット	
提供者	
タイプ	web
公開範囲	公開
公開条件	
ライセンス	
準拠する標準	
関連ドキュメント	
エンドポイント URL	
ランディングページ	
配信	
タイトル	mji.00601-xlsx.zip

記載例	
説明	
アクセスサービス	
バイトサイズ	7.3M
圧縮形式	zip
メディアタイプ	
公開日	
最終更新日	2019-05-01
期間	
ステータス	配信中
言語	ja
ライセンス	
利用規約	
準拠する標準	
関連ドキュメント	
アクセス URL	https://moji.or.jp/mojikiban/mjlist/
ダウンロード URL	https://moji.or.jp/wp-content/mojikiban/oscdl/mji.00601-xlsx.zip

メタデータの検討に当たっては、既に定義されているメタデータを参照します。
例えば DATA GO JP のサイト（ https://www.data.go.jp/ ）にアクセスし、メタ情報を調査する場合には以下のように調査を行います。

拠点のデータセットをクリックします。

必要なメタデータを入手します。

メタデータ①

次に Google Dataset Search（ https://www.data.go.jp/ ）を利用する方法もあります。
GoogleDataset Search では CKAN に登録されているデータが検索されるため、データカタログサイトに登録されているデータが全て検索されます。
例えば「医療機関」を調査する場合には、以下の通り検索を実施します。

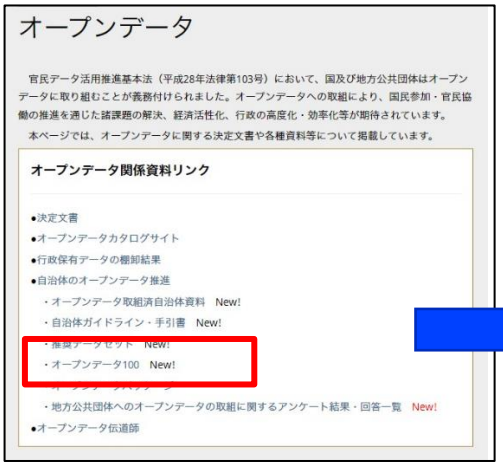
医療機関データをクリックします。　　必要なメタデータを入手します。

メタデータ②

データ構造調査（B：推奨データセットを参照）

政府 CIO ポータルサイトのオープンデータ（ https://cio.go.jp/policy-opendata ）にアクセスします。
推奨データセットに含まれるデータに関しては、推奨データセットを基本として、拡張を行います。
推奨データセットは、1 データ 1 ファイルを基本としている為、推奨データセットが含まれる場合には、手順にしたがってデータ定義を行います。
推奨データセットに含まれないデータ構造については、別途他のデータ構造を確認して検討を進めます。
データ項目までは推奨データセットに記載されていますが、文字列長、データ形式などデータの表現についての記載は少ない為、別途検討する必要があります。

オープンデータの中から、
「自治体のオープンデータ推進」推奨データセットを入手します。



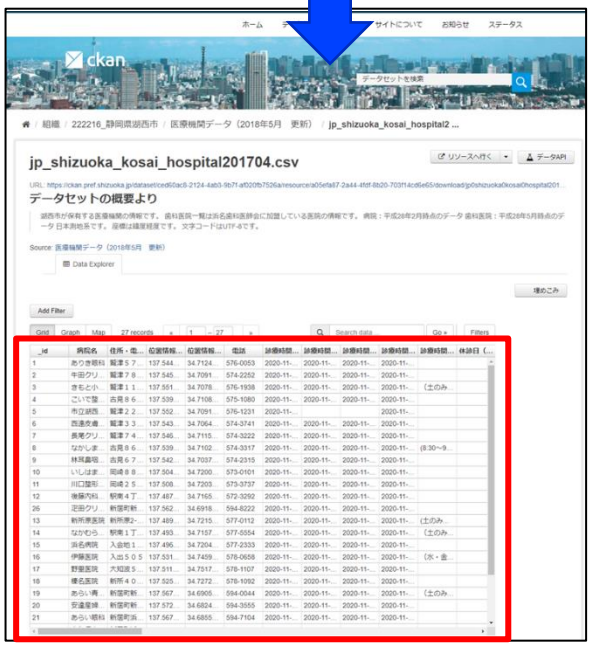
オープンデータの中から、「自治体のオープンデータ推進」推奨データセットを入手します。

標準データセット

データ構造調査（C：政府・自治体のデータカタログサイトを参照）

Google Dataset Search のサイト（ https://www.data.go.jp/ ）にアクセスし、「医療機関」を調査します。
GoogleDataset Search では CKAN に登録されているデータが検索されるため、データカタログサイトに登録されているデータが全て検索されます。

医療機関データをクリックします。　　必要なメタデータを入手します。



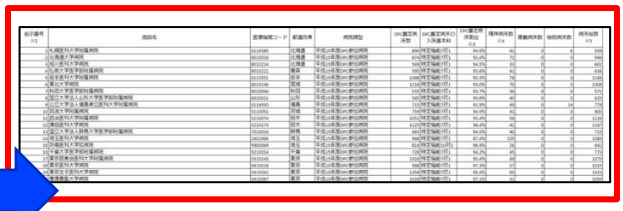
データ構造③

データ構造調査（D：政府・自治体のホームページを参照）

Google 検索（　https://www.google.com/　）にアクセスし、「**医療機関**一覧 filetype:xls site:go.jp」で検索します。
複数のデータ構造を閲覧し、1 行目が項目、2 行目以降がデータになっているデータを探します。

拠点の検索結果をクリックします。

必要な拠点データを入手します。

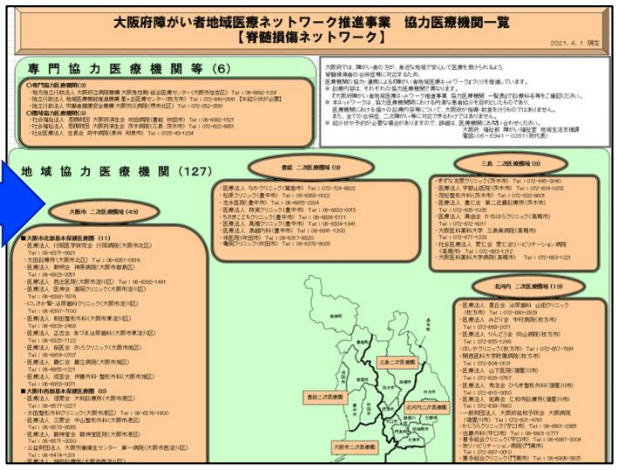


データ構造④

Google 検索（　https://www.google.com/　）にアクセスし、「**医療機関**一覧 filetype:xls site:lg.jp」で検索します。

拠点の検索結果をクリックします。

必要と想われる拠点データを入手しますが・・・



データにより、1 行目が項目、2 行目以降がデータになっていない場合があれば、採用しません。

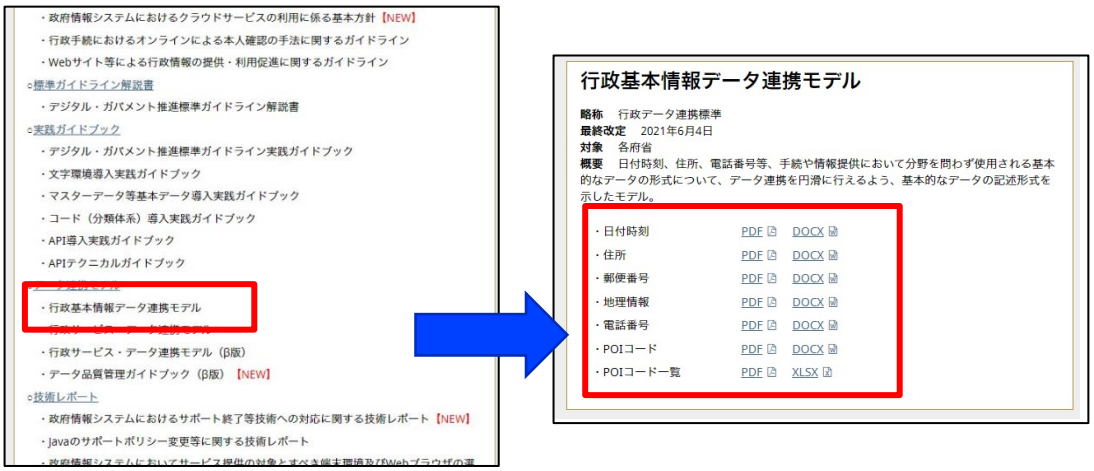
拠点の検索結果をクリックします。

必要な拠点データを入手します。

データ構造⑤

東京都のオープンデータカタログサイト（ https://portal.data.metro.tokyo.lg.jp/ ）にアクセスします。

標準ガイドライン群の中から、「行政基本情報データ連携モデル」推奨データセットを入手します。



データセットの中から、必要なデータセットを入手します。

3. CSV データ構造の決定

「データ構造の調査」の結果を元にまとめます。

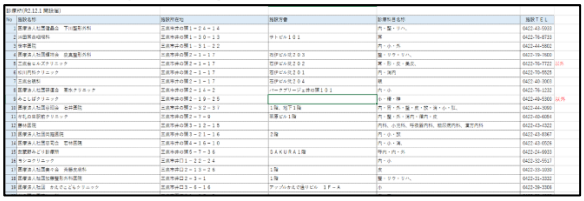
CSV データフォーマット検討

1. 推奨データセットと収集データを比較し、妥当なフォーマットを検討する

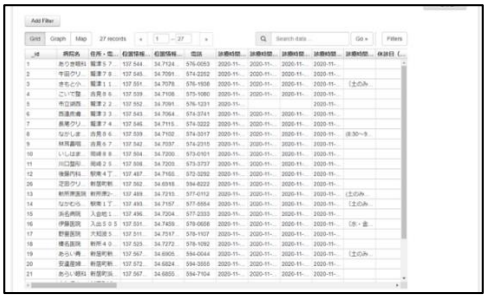
標準データセット（P. 11）

データ構造①（P. 7）

データ構造②（P. 7）

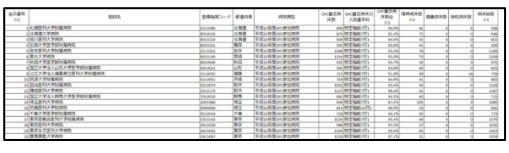


データ構造③（P. 11）

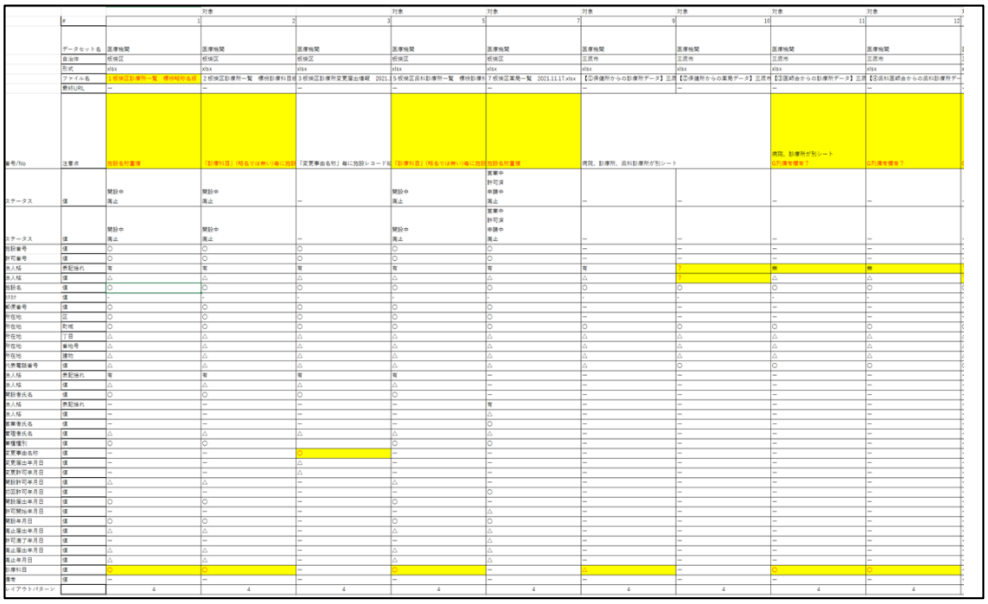


データ構造④（P. 11）

データ構造⑤（P. 11）



2. CSV フォーマットのデータ項目を整理する



今回はデータ処理を行う為のデータフォーマットを検討した為に、複数市区町村にてデータが格納されている共通項目をデータ項目とする方針としています。

3. CSV フォーマットのデータ項目を整理する

今回は推奨データセットと収集データをマッピングした結果を図の通りマッピングし、CSV フォーマットを検討しています。

推奨データセット	収集データマッピング	項目検討結果
都道府県コード又は市区町村コード	**市区町村コード**	都道府県コード又は市区町村コード
NO		NO
都道府県名		都道府県名
	都道府県コード	都道府県コード
市区町村名		市区町村名
	市区町村コード	市区町村コード
名称	**施設名**	名称
名称_カナ	**ｶﾀｶﾅ**	名称_カナ
医療機関の種類		医療機関の種類
	郵便番号	郵便番号
住所	**住所**	住所
方書		方書
緯度	**緯度**	緯度
経度	**経度**	経度
電話番号	**代表電話番号**	電話番号
内線番号		内線番号
FAX番号		FAX番号
	開設者法人格	開設者法人格
	開設者	開設者
法人番号	**法人の場合、法人番号**	法人番号
法人の名称	**法人の場合、法人名**	法人の名称
医療機関コード		医療機関コード
診療曜日		診療曜日
診療開始時間		診療開始時間
診療終了時間		診療終了時間
診療日時特記事項		診療日時特記事項
時間外における対応		時間外における対応
	業種種別	**業種種別**
診療科目	**診療科目**	診療科目
病床数		病床数
URL		URL
備考		備考

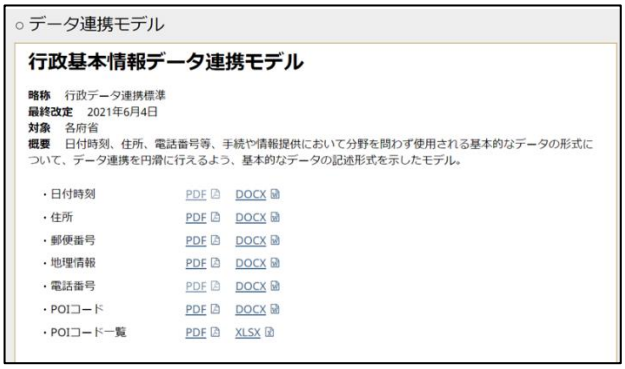
4. CSV データ形式の決定

ISO、JIS、WWC、RFC を参照しデータ形式を

- データ連携モデル　　https://cio.go.jp/guides
- ISO、JIS、WWC、RFC を検索　　https://www.google.com

1.　項目ごとにデータ形式を検討する（行政基本情報データ連携モデル）

行政基本情報データ連携モデルのデータフォーマットを確認し、一般的な項目については行政基本情報データ連携モデルを参照します。
https://cio.go.jp/guides
例えば、日付については ISO8601 及び JIS X 0301（日付及び時刻の表記） に準拠するようにします。



1．1　日付
日付のデータは以下の形式とする。半角を使用する。

YYYY-MM-DD
YYYY：西暦年 4 桁
MM　：月 2 桁（1 桁の場合には前に 0 をつける）
DD　：日 2 桁（1 桁の場合には前に 0 をつける）

YYYY-MM-DDTHH:MM:SS+hh:mm
hh　：UTC に対して進んでいる「時」
mm　：UTC に対して進んでいる「分」（通常は 00）

例）2017-09-01T09:30:00+09:00　（日本）
例）2017-09-01T00:30:00=+　（英国）

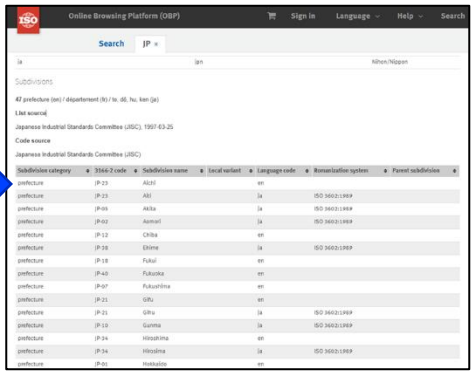
2.　項目ごとにデータ形式を検討する（ISO、JIS）

ISO、JIS にて定義されているコードがあるか確認します。
性別、都道府県、市区町村など既にコード化されているデータが存在します。その場合には出来る限り、既存のデータに追加してコードを付与して情報が公開される事が望ましいです。

例えば都道府県の標準コードを検索するには下記のように検索して調査を行います。

Google 検索（　https://www.google.com/　）にアクセスし、「“都道府県” JIS 標準　OR ISO　標準」で検索します。

調査する際には、出典元を確認し、最新のコードであることを確認する事が必要です。



3.　項目ごとにデータの表現を検討する（共通語彙基盤）

ISO、JIS にて定義されているコードがあるか確認します。
性別、都道府県、市区町村など既にコード化されているデータが存在します。その場合には出来る限り、既存のデータに追加してコードを付与して情報が公開される事が望ましいです。

例えば都道府県の標準コードを検索するには下記のように検索して調査を行います。

コア語彙（ https://imi.go.jp/core/ ）や共通語彙基盤コア語彙 2.4.2（ https://imi.go.jp/ns/core/Core242.html ）にアクセスし、表現を検索します。

4.　データ構造とデータ形式の検討結果をまとめる

これまでの結果をまとめ、データの表現についての検討結果を一元化します。

推奨データセット	収集データマッピング	項目検討結果	データの表現についての検討結果
都道府県コード又は市区町村コード	**市区町村コード**	都道府県コード又は市区町村コード	**JISX0402に準拠する**
NO		NO	数字
都道府県名		都道府県名	**総務省『全国地方公共団体コード』に準拠する**
	都道府県コード	都道府県コード	**JISX0401に準拠する**
市区町村名		市区町村名	**総務省『全国地方公共団体コード』に準拠する**
	市区町村コード	市区町村コード	**JISX0402に準拠する**
名称	**施設名**	名称	**英数字は半角、ひらがなカタカナ漢字は全角**
名称_カナ	**施設名_カタカナ**	名称_カナ	**名称からカタカナを作成する**
医療機関の種類		医療機関の種類	
	郵便番号	郵便番号	**行政基本情報データ連携モデルに準拠する**
住所	**住所**	住所	**行政基本情報データ連携モデルに準拠する**
方書		方書	
緯度	**緯度**	緯度	**行政基本情報データ連携モデルに準拠する**
経度	**経度**	経度	**行政基本情報データ連携モデルに準拠する**
電話番号	**代表電話番号**	電話番号	**行政基本情報データ連携モデルに準拠する**
内線番号		内線番号	
FAX番号		FAX番号	
	開設者法人格	開設者法人格	**法令に準拠する**
	開設者	開設者	**英数字は半角、ひらがなカタカナ漢字は全角**
法人番号	**法人の場合、法人番号**	法人番号	**国税庁法人番号公表サイトの公開情報に準拠する**
法人の名称	**法人の場合、法人名**	法人の名称	**国税庁法人番号公表サイトの公開情報に準拠する**
医療機関コード		医療機関コード	
診療曜日		診療曜日	
診療開始時間		診療開始時間	
診療終了時間		診療終了時間	
診療日時特記事項		診療日時特記事項	
時間外における対応		時間外における対応	
	業種種別	**業種種別**	確認中
診療科目	**診療科目**	診療科目	**診療科区分（厚生労働省 様式コード表）に準拠する**
病床数		病床数	
URL		URL	
備考		備考	

5. 機械判読可能なデータ作成のチェックポイント

データが機械判読可能かどうかで、データ化の精度が左右されます。精度の高いデータ化を目指すには、目視も含むチェックが必要です。下記を確認してください。

ファイル形式

□1　ファイル形式は Excel か CSV となっているか
□2　Word 形式のファイルが埋め込まれた場合は Excel ファイル等にコピーアンドペーストして保存したか

データ・項目

構造

□3　「1 セル 1 データ」となっているか ※1
□4　セルの結合は解除したか ※1
□5　オブジェクトを使用している場合、正しく変換されているか
□6　改行コードが入っていないか ※2※3
□7　ヘッダーは 1 行に収まっているか（可能な限りヘッダーは 1 行とする方が良い）※4※5※6
□8　同一列に内容の異なるデータが混在していないか ※6
□9　月毎、年毎等でブックを分ける場合は、ブック名、シート名のフォーマットとヘッダー列名等レイアウトは揃っているか

数値

□10　数値データは（文字列を含まない）数値属性、かつ半角数字となっているか ※2
□11　数式を使用している場合は、数値に修正しているか
□12　データの単位を記載しているか
□13　年月日記載の際は、半角数字で記載しているか(全角数字、漢数字、元年は使用しない) ※2
□14　英数字記載の際は半角英数字で記載しているか ※2

文字・体裁

□15　スペースや改行等で体裁を整えている場合、対策を行ったか ※2※3
□16　項目名等が省略されている場合、不足情報を補ったか
□17　機種依存文字を使用していないか
□18　e-Stat の時間軸コードの表記、西暦表記文は和暦に西暦の併記がされているか
□19　地域コードまたは地域名称が表記されているか
□20　数値データの同一列内に特殊記号（秘匿等）が含まれていないか
□21　取消線や装飾文字を使用していないか ※1
□22　データとして取り込むべき情報がコメントにて記載されていないか ※7

※1～※6 でエラーが発生する場合は下記のような整形処理が必要となります。

※1 フォーマットやレイアウトが異なる場合は個別に整形処理する
※2 値を置換する
※3 改行コードは半角スペースに変換する（改行コードが入っていると CSV 化する際、別行データとして扱われるため）
※4 列を結合/分離する
※5 行列を入れ替える
※6 同一列に内容の異なるデータが混在する場合、同一内容で列を分割するため、行列の入れ替えや列の分割等複雑な整形処理が必要となる
※7 Excel マクロで処理する

表の構成

□23　データが分断されていないか
□24　１シートに複数の表が掲載されていないか
□25　データが増える際に、列ではなく行が増える構成か（データ項目が縦に並ぶ「縦持ち」の表になっているか）

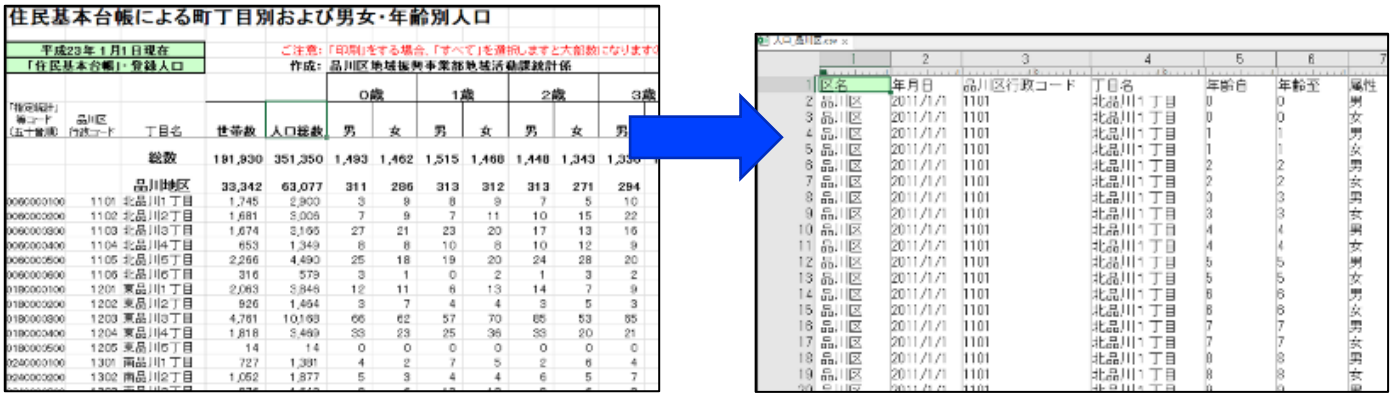
出典（□1、□3～□5、□10～□12、□15～□20、□23、□24）：総務省統計局. 統計表における機械判読可能なデータの表記方法の統一ルールの策定. 報道資料, 令和 2 年 12 月 18 日, 別紙, p.1-22. (online), https://www.soumu.go.jp/menu_news/s-news/01toukatsu01_02000186.html, (参照 2022-01-31) / 出典をもとに、チェックリスト形式に改変。

6. ケース一覧

データ変換作業において、本書ではデータ保持の状況ごとに 5 種類に分類し、各ケースにあわせてデータ化手順を掲載しています。
以下、5 種類のケースを示します。

ケース[A]：保存形式 XLS／縦横多段クロス集計

Excel 上でコピー、貼り付け等の操作や、結合セルを解除することで CSV データに変換できるデータのレイアウトの Excel データを、CSV データに変換します。



ケース[B]：保存形式 XLS／単純クロス集計

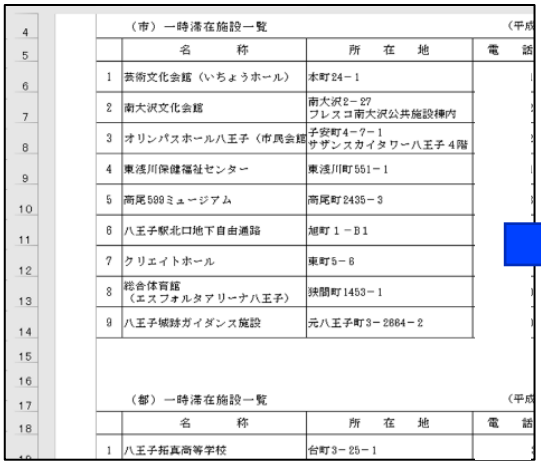
1 行目が項目行、2 行目以降がデータ行となり、1 レコードが 1 行で保持されたデータを、CSV データに変換します。

ケース[C]：保存形式 XLS／単純表形式または複数ファイル

1 行目が項目行、2 行目以降がデータ行となった、表形式のデータを、CSV データに変換します。
項目の粒度や目的が混在している場合は、変換したい CSV ファイルにあわせて表を分割する必要があります。

ケース[D]：保存形式 XLS／1 シートに単純表形式が種類別に縦に３つ並んでいる、等

1 シートに複数の単純表形式が混在しているレイアウトのデータを、CSV データに変換します。
CSV データにするにはファイルを分割する必要がある。例えば同一シート内に単純表形式が縦に 3 つあれば、1 シートに 1 表、かつ 3 ファイルに分割する必要があります。

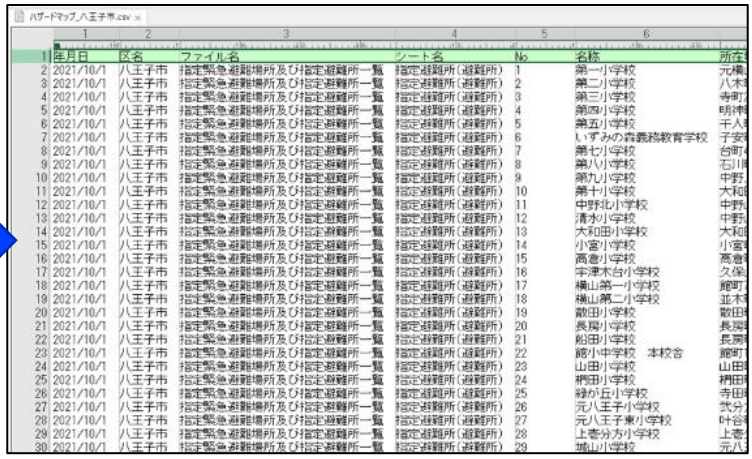


（市）一時滞在施設一覧　（平成

	名　称	所　在　地	電　話
1	芸術文化会館（いちょうホール）	本町24－1	
2	南大沢文化会館	南大沢2－27 プレスコ南大沢公共施設棟内	
3	オリンパスホール八王子（市民会館）	子安町4－7－1 サザンスカイタワー八王子４階	
4	東浅川保健福祉センター	東浅川町551－1	
5	高尾599ミュージアム	高尾町2435－3	
6	八王子駅北口地下自由通路	旭町１－Ｂ１	
7	クリエイトホール	東町5－6	
8	総合体育館 （エスフォルタアリーナ八王子）	狭間町1453－1	
9	八王子城跡ガイダンス施設	元八王子町3－2664－2	

（都）一時滞在施設一覧　（平成

	名　称	所　在　地	電　話
1	八王子拓真商等学校	台町3－25－1	



ハザードマップ_八王子市.csv

	年月日	区名	ファイル名	シート名	No	名称
2	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	1	第一小学校
3	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	2	第二小学校
4	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	3	第三小学校
5	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	4	第四小学校
6	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	5	第五小学校
7	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	6	いずみの森義務教育学校
8	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	7	第七小学校
9	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	8	第八小学校
10	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	9	第九小学校
11	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	10	第十小学校
12	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	11	中野北小学校
13	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	12	清水小学校
14	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	13	大和田小学校
15	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	14	小宮小学校
16	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	15	高倉小学校
17	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	16	宇津木台小学校
18	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	17	横山第一小学校
19	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	18	横山第二小学校
20	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	19	散田小学校
21	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	20	長房小学校
22	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	21	船田小学校
23	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	22	館小中学校　本校舎
24	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	23	山田小学校
25	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	24	柄田小学校
26	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	25	緑が丘小学校
27	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	26	元八王子小学校
28	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	27	元八王子東小学校
29	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	28	上壱分方小学校
30	2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	29	城山小学校

ケース[E]：保存形式 PDF

・AI-OCR を利用して読み取り、や PDF を CSV データに変換します。

区分		世帯数	人口			前月比増減	
			男	女	計	世帯数	人口
港区総数	日本人	119,108	99,438	113,124	212,562	20	113
	外国人	—	10,374	9,187	19,561	—	65
	合計	—	109,812	122,311	232,123	—	178
芝地区総合支所管内	日本人	20,888	16,589	17,998	34,587	22	27
	外国人	—	1,551	1,240	2,791	—	10
	合計	—	18,140	19,238	37,378	—	37
麻布地区総合支所管内	日本人	27,638	21,556	24,974	46,530	△ 7	19
	外国人	—	3,980	3,577	7,557	—	74
	合計	—	25,536	28,551	54,087	—	
赤坂地区総合支所管内	日本人	17,570	14,415	17,093	31,508	13	14
	外国人	—	1,742	1,577	3,319	—	17
	合計	—	16,157	18,670	34,827	—	31
高輪地区総合支所管内	日本人	29,000	23,751	29,574	53,325	△ 8	15
	外国人	—	1,487	1,370	2,857	—	△ 19

区分		世帯数	人口			前月比増減	
			男	女	計	世帯数	人口
港区総数	日本人	119,108	99,438	113,124	212,562	20	113
	外国人	—	10,374	9,187	19,561	—	65
	合計	—	109,812	122,311	232,123	—	178
芝地区総合支所管内	日本人	20,888	16,589	17,998	34,587	22	27
	外国人	—	1,551	1,240	2,791	—	10
	合計	—	18,140	19,238	37,378	—	37
麻布地区総合支所管内	日本人	27,638	21,556	24,974	46,530	△ 7	19
	外国人	—	3,980	3,577	7,557	—	74
	合計	—	25,536	28,551	54,087	—	93
赤坂地区総合支所管内	日本人	17,570	14,415	17,093	31,508	13	14
	外国人	—	1,742	1,577	3,319	—	17
	合計	—	16,157	18,670	34,827	—	31
高輪地区総合支所管内	日本人	29,000	23,751	29,574	53,325	△ 8	15
	外国人	—	1,487	1,370	2,857	—	△ 19

手順 2

データ構造を統一する

各種データの構造を統一します。

1. **PowerQurery によるデータ変換** （参考 p. 7）
   Excel を利用してデータ構造を統一します。

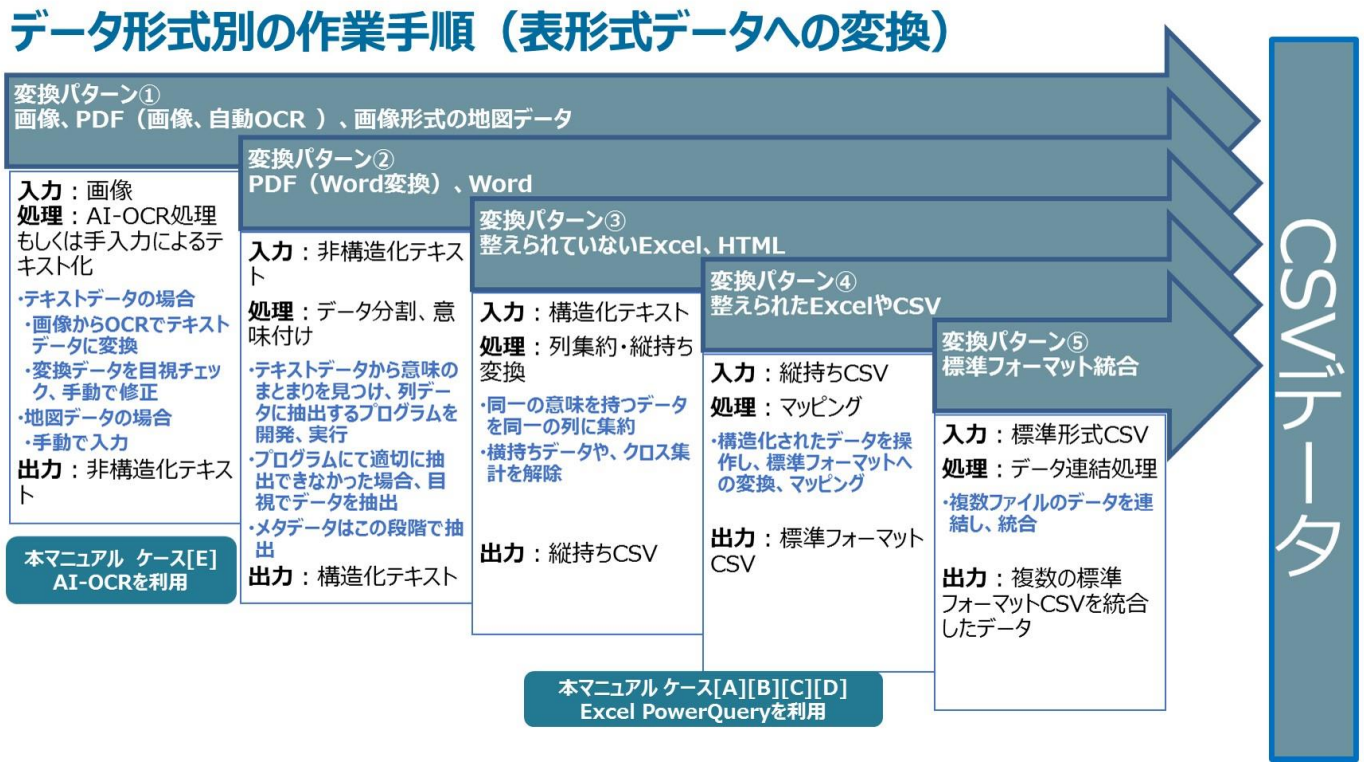
   ケース[A]　保存形式 XLS／縦横多段クロス集計 （参考 p. 7）
   ケース[B]　保存形式 XLS／単純クロス集計 （参考 p. 7）
   ケース[C]　保存形式 XLS／単純表形式または複数ファイル （参考 p. 7）
   ケース[D]　保存形式 XLS／1 シートに単純表形式が種類別に縦に３つ並んでいる （参考 p. 7）

2. **AI-OCR によるデータ変換** （参考 p. 8）
   AI-OCR を利用してデータ構造を統一します。

   ケース[E]　保存形式 PDF （参考 p. 7）

データ形式別の作業フロー（表形式データ変換）

最終的に公開データに変換する為には、データの種類に応じて複数のパターンがあり、この章では行政内でよく使われる Excel データと」PDF データを CSV 構造を統一する手順を説明しています。この手順を実施する事でその後のデータクレンジングなどのデータの精度向上手順を単純化する事が出来ます。
下記の図では「変換パターン①」の場合は変換パターン①から始まり、②、③、④、⑤までの作業がマッピング前までの CSV データ作成までに必要な手順となっており、「変換パターン③」の場合は、変換パターン③から始まり、④、⑤までの作業が必要な事を示しています。
（資料別目次（p. 3）では手順１、２として記載）

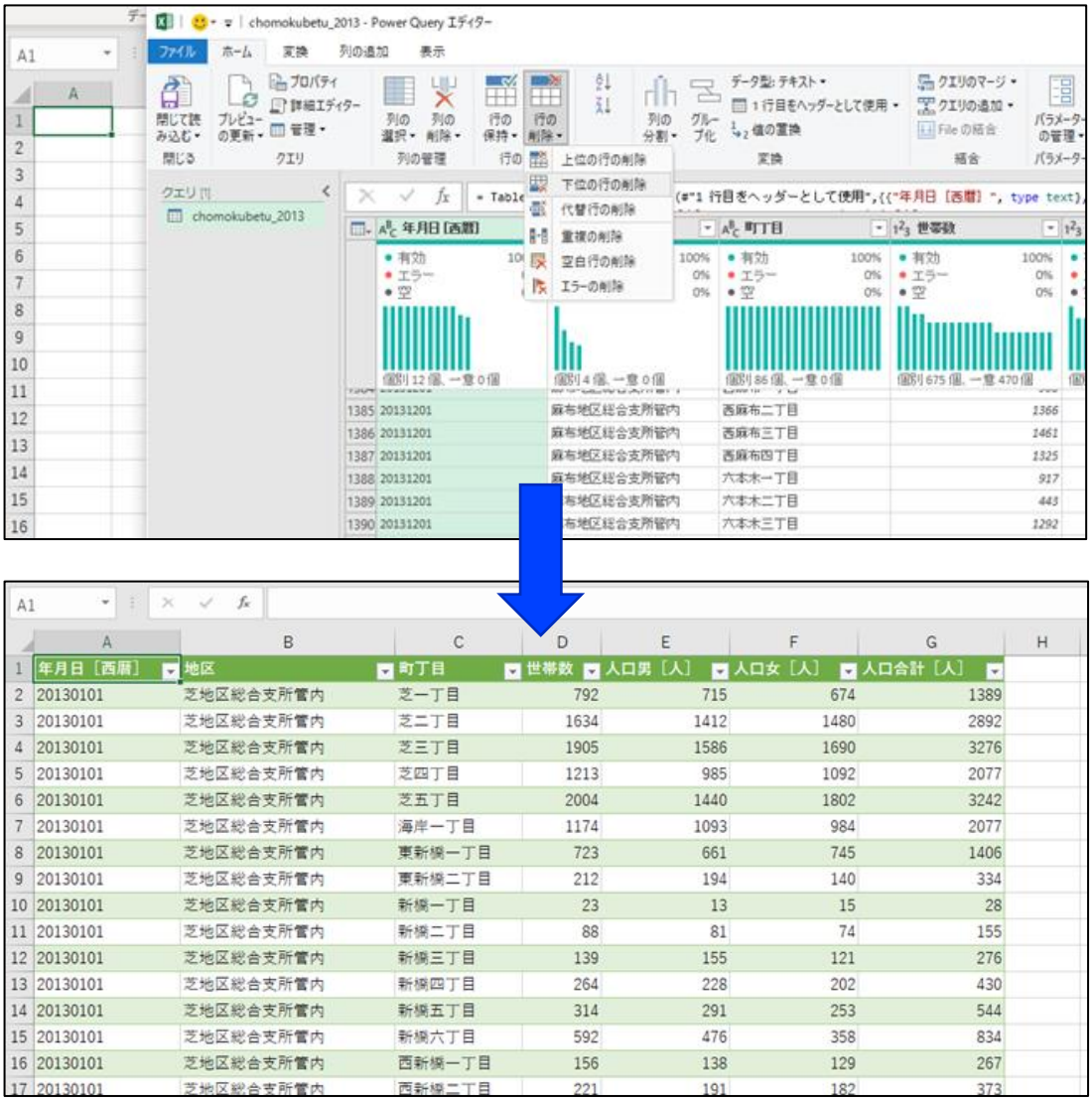


データ形式別の作業フロー（表形式にならないデータ）

他にも既に標準化されたデータやオープンデータにできないデータも存在しますが、本資料では取り扱いません

Power Query によるデータ変換

Power Query（パワークエリ）とは、マイクロソフト社の エクセルの機能の一つで、外部データとの連携や、連携してからのデータの加工（列の追加や抽出等）を自動化する機能です。
Excel の Power Query 機能を利用してデータを整形します。
その後、MySQL サーバーにデータをインポートし、検討した CSV データに変換します。
Power Query について詳細は、「Power Query（Microsoft Excel の標準またはアドイン）」(P. 5)をご覧ください。



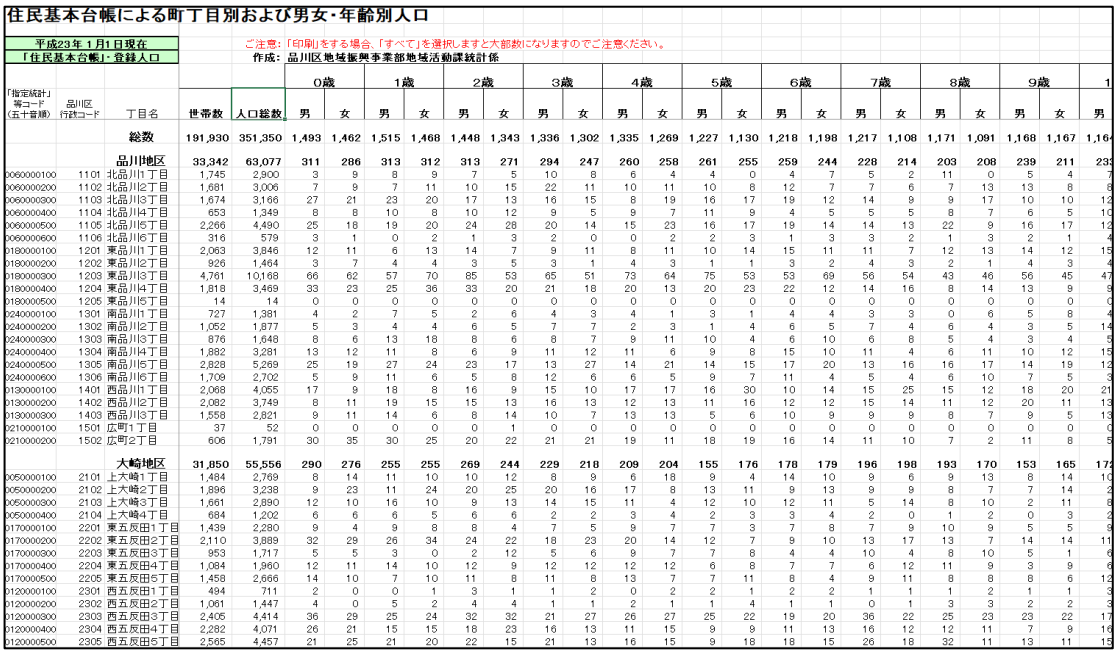
年月日［西暦］	地区	町丁目	世帯数	人口男［人］	人口女［人］	人口合計［人］
20130101	芝地区総合支所管内	芝一丁目	792	715	674	1389
20130101	芝地区総合支所管内	芝二丁目	1634	1412	1480	2892
20130101	芝地区総合支所管内	芝三丁目	1905	1586	1690	3276
20130101	芝地区総合支所管内	芝四丁目	1213	985	1092	2077
20130101	芝地区総合支所管内	芝五丁目	2004	1440	1802	3242
20130101	芝地区総合支所管内	海岸一丁目	1174	1093	984	2077
20130101	芝地区総合支所管内	東新橋一丁目	723	661	745	1406
20130101	芝地区総合支所管内	東新橋二丁目	212	194	140	334
20130101	芝地区総合支所管内	新橋一丁目	23	13	15	28
20130101	芝地区総合支所管内	新橋二丁目	88	81	74	155
20130101	芝地区総合支所管内	新橋三丁目	139	155	121	276
20130101	芝地区総合支所管内	新橋四丁目	264	228	202	430
20130101	芝地区総合支所管内	新橋五丁目	314	291	253	544
20130101	芝地区総合支所管内	新橋六丁目	592	476	358	834
20130101	芝地区総合支所管内	西新橋一丁目	156	138	129	267
20130101	芝地区総合支所管内	西新橋二丁目	221	191	182	373

取り扱うファイルのケースにあわせて、下記のケース別分類で Excel から CSV に変換します。

ケース[A] ／ 縦横多段クロス集計

適用データ例：
住民基本台帳のデータ

Excel 上でコピー、貼り付け等の操作や、結合セルを解除することで CSV データに変換できるデータのレイアウト。



住民基本台帳による町丁目別および男女・年齢別人口

平成23年１月１日現在
「住民基本台帳」・登録人口

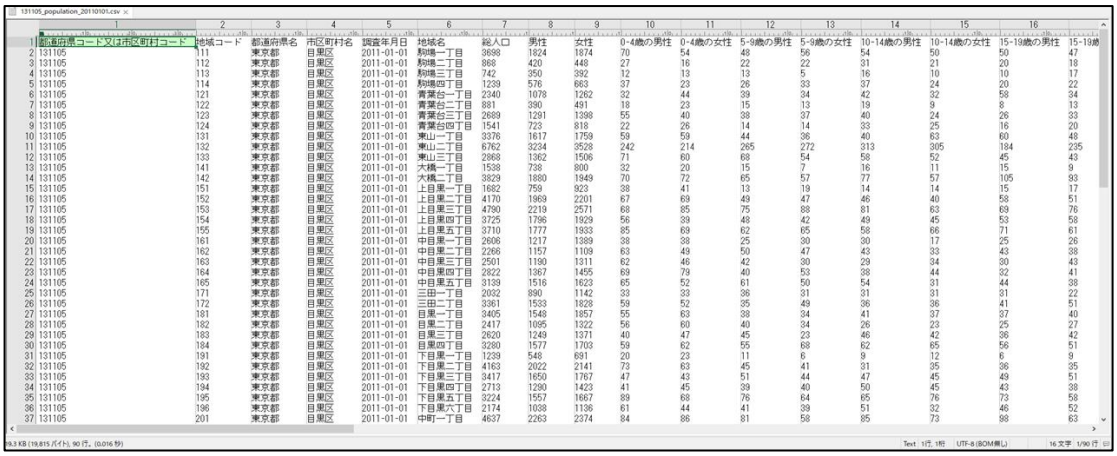
ご注意：「印刷」をする場合、「すべて」を選択しますと大部数になりますのでご注意ください。
作成：品川区地域振興事業部地域活動課統計係

「指定統計」等コード（五十音順）	品川区行政コード	丁目名	世帯数	人口総数	0歳		1歳		2歳		3歳		4歳		5歳		6歳		7歳		8歳		9歳		1
					男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男
		総数	191,930	351,350	1,493	1,462	1,515	1,468	1,448	1,343	1,336	1,302	1,335	1,269	1,227	1,130	1,218	1,198	1,217	1,108	1,171	1,091	1,168	1,167	1,164
		品川地区	33,342	63,077	311	286	313	312	313	271	294	247	260	258	261	255	259	244	228	214	203	208	239	211	233
0060000100	1101	北品川1丁目	1,745	2,900	3	9	8	9	7	5	10	8	6	4	4	0	4	7	5	2	11	0	5	4	7
0060000200	1102	北品川2丁目	1,681	3,006	7	9	7	11	10	15	22	11	10	11	10	8	12	7	7	6	7	13	13	8	8
0060000300	1103	北品川3丁目	1,674	3,166	27	21	23	20	17	13	16	15	8	19	16	17	19	12	14	9	9	17	10	10	12
0060000400	1104	北品川4丁目	653	1,349	8	8	10	8	10	12	9	5	9	7	11	9	4	5	5	5	8	7	6	5	10
0060000500	1105	北品川5丁目	2,266	4,490	25	18	19	20	24	28	20	14	15	23	16	17	19	14	14	13	22	9	16	17	12
0060000600	1106	北品川6丁目	316	579	3	1	0	2	1	3	2	0	0	2	2	3	1	3	3	2	1	3	2	1	4
0180000100	1201	東品川1丁目	2,063	3,846	12	11	6	13	14	7	9	11	8	11	10	14	15	11	11	7	12	13	14	12	15
0180000200	1202	東品川2丁目	926	1,464	3	7	4	4	3	5	3	1	4	3	1	1	3	2	4	3	2	1	4	3	4
0180000300	1203	東品川3丁目	4,761	10,168	66	62	57	70	85	53	65	51	73	64	75	53	53	69	56	54	43	46	56	45	47
0180000400	1204	東品川4丁目	1,818	3,469	33	23	25	36	33	20	21	18	20	13	20	23	22	12	14	16	8	14	13	9	9
0180000500	1205	東品川5丁目	14	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0240000100	1301	南品川1丁目	727	1,381	4	2	7	5	2	6	4	3	4	1	3	1	4	4	3	3	0	6	5	8	4
0240000200	1302	南品川2丁目	1,052	1,877	5	3	4	4	6	5	7	7	2	3	1	4	6	5	7	4	6	4	3	5	14
0240000300	1303	南品川3丁目	876	1,648	8	6	13	18	8	6	8	7	9	11	10	4	6	10	6	8	5	4	3	4	5
0240000400	1304	南品川4丁目	1,882	3,281	13	12	11	8	6	9	11	12	11	6	9	8	15	10	11	4	6	11	10	12	15
0240000500	1305	南品川5丁目	2,828	5,269	25	19	27	24	23	17	13	27	14	21	14	15	17	20	13	16	16	17	14	19	12
0240000600	1306	南品川6丁目	1,709	2,702	5	9	11	6	5	8	12	6	6	5	9	7	11	4	5	4	6	10	7	5	3
0130000100	1401	西品川1丁目	2,068	4,055	17	9	18	8	16	9	15	10	17	17	16	30	10	14	15	25	15	12	18	20	21
0130000200	1402	西品川2丁目	2,082	3,749	8	11	19	15	15	13	16	13	12	13	11	16	12	12	15	14	11	12	20	11	13
0130000300	1403	西品川3丁目	1,558	2,821	9	11	14	6	8	14	10	7	13	13	5	6	10	9	9	9	8	7	9	5	13
0210000100	1501	広町1丁目	37	52	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0210000200	1502	広町2丁目	606	1,791	30	35	30	25	20	22	21	21	19	11	18	19	16	14	11	10	7	2	11	8	5
		大崎地区	31,850	55,556	290	276	255	255	269	244	229	218	209	204	155	176	178	179	196	198	193	170	153	165	175
0050000100	2101	上大崎1丁目	1,484	2,769	8	14	11	10	10	12	8	9	6	18	9	4	14	10	9	6	9	13	8	14	10
0050000200	2102	上大崎2丁目	1,896	3,238	9	23	11	24	20	25	20	16	17	8	13	11	9	13	9	9	8	7	7	14	2
0050000300	2103	上大崎3丁目	1,661	2,890	12	10	16	10	9	13	14	15	11	4	12	10	12	11	5	14	8	10	2	11	8
0050000400	2104	上大崎4丁目	684	1,202	6	6	6	5	6	6	2	2	3	4	2	3	3	4	2	0	1	2	0	3	2
0170000100	2201	東五反田1丁目	1,439	2,280	9	4	9	8	8	4	7	5	9	7	7	3	7	8	7	9	10	9	5	5	9
0170000200	2202	東五反田2丁目	2,110	3,889	32	29	26	34	24	22	18	23	20	14	12	7	9	10	13	17	13	7	14	14	11
0170000300	2203	東五反田3丁目	953	1,717	5	5	3	0	2	12	5	6	9	7	7	8	4	4	10	4	8	10	5	1	6
0170000400	2204	東五反田4丁目	1,084	1,960	12	11	14	10	12	9	12	12	12	12	6	8	7	7	6	12	11	9	3	9	6
0170000500	2205	東五反田5丁目	1,458	2,666	14	10	7	10	11	8	11	8	13	7	7	11	8	4	9	11	8	8	8	6	12
0120000100	2301	西五反田1丁目	494	711	2	0	0	1	3	1	1	2	0	2	2	1	2	2	1	1	1	2	1	1	3
0120000200	2302	西五反田2丁目	1,061	1,447	4	0	5	2	4	4	1	1	2	1	1	4	1	1	0	1	3	3	2	2	3
0120000300	2303	西五反田3丁目	2,405	4,414	36	29	25	24	32	32	21	27	26	27	25	22	19	20	36	22	25	23	23	22	17
0120000400	2304	西五反田4丁目	2,282	4,071	26	21	15	15	18	23	16	13	11	15	9	9	11	13	16	12	12	11	7	9	16
0120000500	2305	西五反田5丁目	2,565	4,457	21	25	21	20	22	15	21	13	16	15	9	18	18	15	26	18	32	11	13	11	15

ケース[B] ／ 単純クロス集計

適用データ例：
地域・年齢別人口のデータ

1 行目が項目行、2 行目以降がデータ行となり、1 レコードが 1 行で保持されたデータ。



131105_population_20110101.csv

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	都道府県コード又は市区町村コード	地域コード	都道府県名	市区町村名	調査年月日	地域名	総人口	男性	女性	0-4歳の男性	0-4歳の女性	5-9歳の男性	5-9歳の女性	10-14歳の男性	10-14歳の女性	15-19歳の男性	15-19歳
2	131105	111	東京都	目黒区	2011-01-01	駒場一丁目	3698	1824	1874	70	54	48	56	54	50	50	47
3	131105	112	東京都	目黒区	2011-01-01	駒場二丁目	868	420	448	27	16	22	22	31	21	20	18
4	131105	113	東京都	目黒区	2011-01-01	駒場三丁目	742	350	392	12	13	13	5	16	10	10	17
5	131105	114	東京都	目黒区	2011-01-01	駒場四丁目	1239	576	663	37	23	26	33	37	24	20	22
6	131105	121	東京都	目黒区	2011-01-01	青葉台一丁目	2340	1078	1262	32	44	39	34	42	32	58	34
7	131105	122	東京都	目黒区	2011-01-01	青葉台二丁目	881	390	491	18	23	15	13	19	9	8	13
8	131105	123	東京都	目黒区	2011-01-01	青葉台三丁目	2689	1291	1398	55	40	38	37	40	24	26	33
9	131105	124	東京都	目黒区	2011-01-01	青葉台四丁目	1541	723	818	22	26	14	14	33	25	16	20
10	131105	131	東京都	目黒区	2011-01-01	東山一丁目	3376	1617	1759	59	59	44	36	40	63	60	48
11	131105	132	東京都	目黒区	2011-01-01	東山二丁目	6762	3234	3528	242	214	265	272	313	305	184	235
12	131105	133	東京都	目黒区	2011-01-01	東山三丁目	2868	1362	1506	71	60	69	54	58	52	45	43
13	131105	141	東京都	目黒区	2011-01-01	大橋一丁目	1538	738	800	32	20	15	7	16	11	15	9
14	131105	142	東京都	目黒区	2011-01-01	大橋二丁目	3829	1880	1949	70	72	65	57	77	57	105	93
15	131105	151	東京都	目黒区	2011-01-01	上目黒一丁目	1682	759	923	38	41	13	19	14	14	15	17
16	131105	152	東京都	目黒区	2011-01-01	上目黒二丁目	4170	1969	2201	67	69	49	47	46	40	58	51
17	131105	153	東京都	目黒区	2011-01-01	上目黒三丁目	4790	2219	2571	68	85	75	88	81	63	69	76
18	131105	154	東京都	目黒区	2011-01-01	上目黒四丁目	3725	1796	1929	56	39	48	42	49	45	53	58
19	131105	155	東京都	目黒区	2011-01-01	上目黒五丁目	3710	1777	1933	85	69	62	65	58	66	71	61
20	131105	161	東京都	目黒区	2011-01-01	中目黒一丁目	2606	1217	1389	38	38	25	30	30	17	25	28
21	131105	162	東京都	目黒区	2011-01-01	中目黒二丁目	2266	1157	1109	63	49	50	47	43	33	43	38
22	131105	163	東京都	目黒区	2011-01-01	中目黒三丁目	2501	1190	1311	62	46	42	30	29	34	30	43
23	131105	164	東京都	目黒区	2011-01-01	中目黒四丁目	2822	1367	1455	69	79	40	53	38	44	32	41
24	131105	165	東京都	目黒区	2011-01-01	中目黒五丁目	3139	1516	1623	65	52	61	50	54	31	44	38
25	131105	171	東京都	目黒区	2011-01-01	三田一丁目	2032	890	1142	33	33	36	31	31	31	31	22
26	131105	172	東京都	目黒区	2011-01-01	三田二丁目	3361	1533	1828	59	52	35	49	36	36	41	51
27	131105	181	東京都	目黒区	2011-01-01	目黒一丁目	3405	1548	1857	55	63	38	34	41	37	37	40
28	131105	182	東京都	目黒区	2011-01-01	目黒二丁目	2417	1095	1322	56	60	40	34	26	23	25	27
29	131105	183	東京都	目黒区	2011-01-01	目黒三丁目	2620	1249	1371	40	47	45	23	46	42	36	42
30	131105	184	東京都	目黒区	2011-01-01	目黒四丁目	3280	1577	1703	59	62	55	68	62	65	56	51
31	131105	191	東京都	目黒区	2011-01-01	下目黒一丁目	1239	548	691	20	23	11	6	9	12	6	9
32	131105	192	東京都	目黒区	2011-01-01	下目黒二丁目	4163	2022	2141	73	63	45	41	31	35	36	35
33	131105	193	東京都	目黒区	2011-01-01	下目黒三丁目	3417	1650	1767	47	43	51	44	47	45	49	51
34	131105	194	東京都	目黒区	2011-01-01	下目黒四丁目	2713	1290	1423	41	45	39	40	50	45	43	38
35	131105	195	東京都	目黒区	2011-01-01	下目黒五丁目	3224	1557	1667	89	68	76	64	65	76	73	58
36	131105	196	東京都	目黒区	2011-01-01	下目黒六丁目	2174	1038	1136	61	44	41	39	51	32	46	52
37	131105	201	東京都	目黒区	2011-01-01	中町一丁目	4637	2263	2374	84	86	81	58	85	73	98	63

ケース[C] ／ 単純表形式、または　複数ファイル

適用データ例：
医療機関データ

1 行目が項目行、2 行目以降がデータ行となった、表形式のデータ。
項目の粒度や目的が混在している場合は、変換したい CSV ファイルにあわせて表を分割する必要があります。

ステータ	施設番号	許可番号	施設名称	郵便番号	所在地	代表電話番号	開設者氏名	管理者氏名	業種種別	標榜診療科目名	開設許可年月日	開設届出年月日	開設年月日	廃止届出年月日	廃止年月日
開設中	00000299	H4板板保衛-00077	岩渕診療所	173-0001	板橋区本町17番1号	3961-2413	長尾 啓子	長尾 啓子	診療所	皮膚科	1992/4/30	1992/4/30	1992/4/19		
開設中	00000299	H4板板保衛-00077	岩渕診療所	173-0001	板橋区本町17番1号	3961-2413	長尾 啓子	長尾 啓子	診療所	内科	1992/4/30	1992/4/30	1992/4/19		
開設中	00000299	H4板板保衛-00077	岩渕診療所	173-0001	板橋区本町17番1号	3961-2413	長尾 啓子	長尾 啓子	診療所	小児科	1992/4/30	1992/4/30	1992/4/19		
開設中	00000299	H4板板保衛-00077	岩渕診療所	173-0001	板橋区本町17番1号	3961-2413	長尾 啓子	長尾 啓子	診療所	耳鼻いんこう科	1992/4/30	1992/4/30	1992/4/19		
開設中	00000726	H23板保生し-00066	鈴木整形外科クリニック	173-0001	板橋区本町17番5-101号ル・クール	5375-1049	鈴木　精	鈴木　精	診療所	整形外科	2011/7/19	2011/7/19	2011/7/19		
開設中	00000726	H23板保生し-00066	鈴木整形外科クリニック	173-0001	板橋区本町17番5-101号ル・クール	5375-1049	鈴木　精	鈴木　精	診療所	リハビリテーション科	2011/7/19	2011/7/19	2011/7/19		
開設中	00000726	H23板保生し-00066	鈴木整形外科クリニック	173-0001	板橋区本町17番5-101号ル・クール	5375-1049	鈴木　精	鈴木　精	診療所	皮膚科	2011/7/19	2011/7/19	2011/7/19		
廃止	00000300	H2板板保衛-00284	三瓶整形外科	173-0001	板橋区本町17番5号ル・クール1階	5375-1049	三瓶 晴雄	三瓶 晴雄	診療所	整形外科	1927/11/20	1927/11/20	1990/11/20	2011/6/23	2011/6/23
廃止	00000300	H2板板保衛-00284	三瓶整形外科	173-0001	板橋区本町17番5号ル・クール1階	5375-1049	三瓶 晴雄	三瓶 晴雄	診療所	リハビリテーション科	1927/11/20	1927/11/20	1990/11/20	2011/6/23	2011/6/23
開設中	00009364	H28板保生医し-00112	板橋本町皮膚・形成外科	173-0001	板橋区本町25番10号ファン・ビル2階	6912-4385	林　殿聡	林　殿聡	診療所	皮膚科		2016/10/3	2016/10/1		
開設中	00009364	H28板保生医し-00112	板橋本町皮膚・形成外科	173-0001	板橋区本町25番10号ファン・ビル2階	6912-4385	林　殿聡	林　殿聡	診療所	形成外科		2016/10/3	2016/10/1		
開設中	00000652	H18板保生し-00100	城西クリニック	173-0001	板橋区本町25番10号ファンビル4階	3579-3118	小沢 義行	小沢 義行	診療所	内科	2006/9/1	2006/9/1	2006/9/1		
開設中	00000652	H18板保生し-00100	城西クリニック	173-0001	板橋区本町25番10号ファンビル4階	3579-3118	小沢 義行	小沢 義行	診療所	外科	2006/9/1	2006/9/1	2006/9/1		
開設中	00000652	H18板保生し-00100	城西クリニック	173-0001	板橋区本町25番10号ファンビル4階	3579-3118	小沢 義行	小沢 義行	診療所	整形外科	2006/9/1	2006/9/1	2006/9/1		
廃止	00000301	M33板保旧し番-01087	清水医院	173-0001	板橋区本町25番6号	3961-2318	清水　宏	小高　久	診療所	内科	1947/12/2	1947/12/2	1947/12/2	2007/5/1	2007/4/30
廃止	00000301	M33板保旧し番-01087	清水医院	173-0001	板橋区本町25番6号	3961-2318	清水　宏	小高　久	診療所	小児科	1947/12/2	1947/12/2	1947/12/2	2007/5/1	2007/4/30
廃止	00000301	M33板保旧し番-01087	清水医院	173-0001	板橋区本町25番6号	3961-2318	清水　宏	小高　久	診療所	皮膚科	1947/12/2	1947/12/2	1947/12/2	2007/5/1	2007/4/30
廃止	00000301	M33板保旧し番-01087	清水医院	173-0001	板橋区本町25番6号	3961-2318	清水　宏	小高　久	診療所	放射線科	1947/12/2	1947/12/2	1947/12/2	2007/5/1	2007/4/30
廃止	00000016	H6板板保衛-00148	(医)育明会板橋本町クリニック	173-0001	板橋区本町25番7-101号	5248-0411	医療法人社団育明会	折口 信人	診療所	内科	1994/10/14	1994/11/1	1994/11/1	2004/7/8	2004/6/30
廃止	00000016	H6板板保衛-00148	(医)育明会板橋本町クリニック	173-0001	板橋区本町25番7-101号	5248-0411	医療法人社団育明会	折口 信人	診療所	小児科	1994/10/14	1994/11/1	1994/11/1	2004/7/8	2004/6/30
廃止	00000016	H6板板保衛-00148	(医)育明会板橋本町クリニック	173-0001	板橋区本町25番7-101号	5248-0411	医療法人社団育明会	折口 信人	診療所	外科	1994/10/14	1994/11/1	1994/11/1	2004/7/8	2004/6/30
廃止	00000016	H6板板保衛-00148	(医)育明会板橋本町クリニック	173-0001	板橋区本町25番7-101号	5248-0411	医療法人社団育明会	折口 信人	診療所	整形外科	1994/10/14	1994/11/1	1994/11/1	2004/7/8	2004/6/30
廃止	00000016	H6板板保衛-00148	(医)育明会板橋本町クリニック	173-0001	板橋区本町25番7-101号	5248-0411	医療法人社団育明会	折口 信人	診療所	皮膚科	1994/10/14	1994/11/1	1994/11/1	2004/7/8	2004/6/30
廃止	00000096	H11板保生し-00112	医療法人社団恵明会板橋本町外科医院	173-0001	板橋区本町25番7-102号	5248-3308	医療法人社団恵明会	増岡 宏昭	診療所	外科	1999/10/1	1999/10/4	1999/10/1	2001/4/11	2001/3/31
廃止	00000302	H10板保生し-00029	板橋本町外科医院	173-0001	板橋区本町25番7-102号	5248-3308	増岡　宏昭	増岡　宏昭	診療所	外科	1998/6/8	1998/6/8	1998/6/1	1999/10/4	1999/9/30
廃止	00000224	H23板保生し-00199	医療法人財団明理会大山クリニック	173-0001	板橋区本町36番14号	5943-9251	医療法人財団明理会	針生　恭一	診療所	泌尿器科	2012/3/26	2012/4/1	2012/4/1	2012/6/4	2012/5/31

ケース[D] ／ １シートに単純表形式が種類別に縦に３つ並んでいる、等

適用データ例：
ハザードマップ

1 シートに複数の単純表形式が混在しているレイアウト。
CSV データにするにはファイルを分割する必要がある。例えば同一シート内に単純表形式が縦に 3 つあれば、1 シートに 1 表、かつ 3 ファイルに分割する必要があります。

（市）一時滞在施設一覧　（平成31年1月末現在）

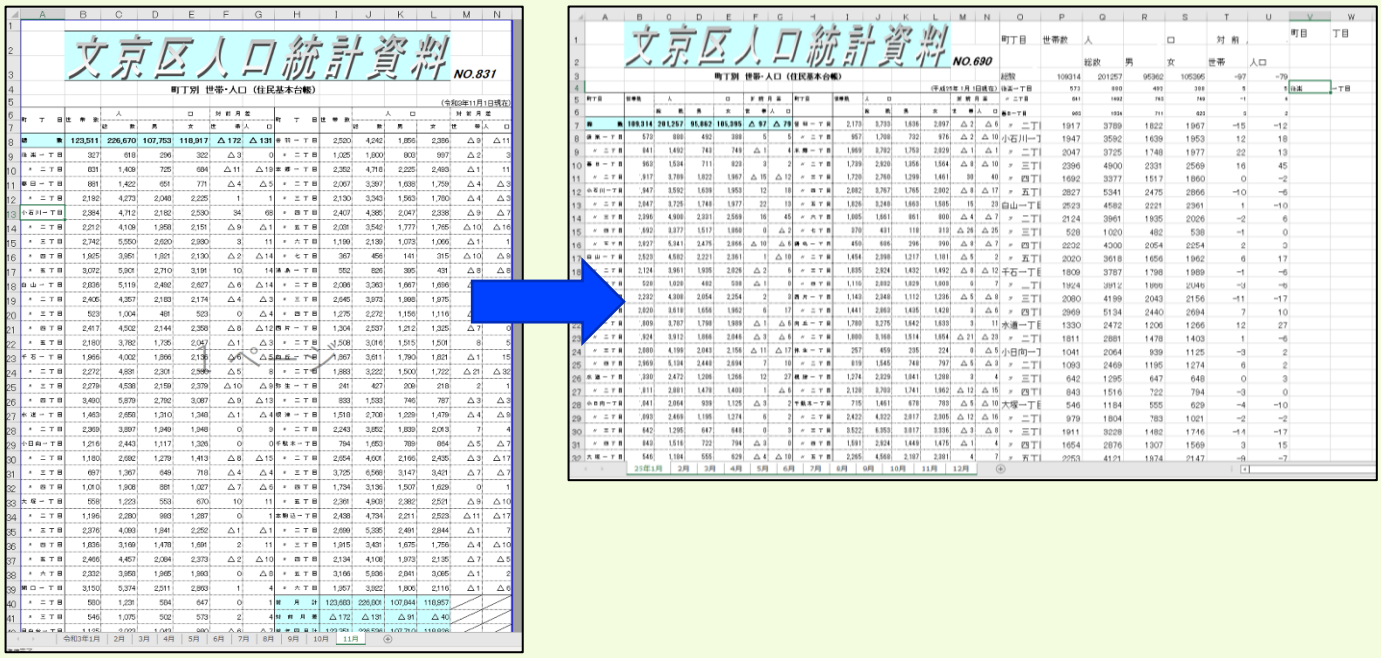
	名　　称	所　在　地	電　話	ＦＡＸ
1	芸術文化会館（いちょうホール）	本町24－1	非公開	
2	南大沢文化会館	南大沢2－27 フレスコ南大沢公共施設棟内		
3	オリンパスホール八王子（市民会館）	子安町4－7－1 サザンスカイタワー八王子４階		
4	東浅川保健福祉センター	東浅川町551－1		
5	高尾599ミュージアム	高尾町2435－3		
6	八王子駅北口地下自由通路	旭町１－Ｂ1		
7	クリエイトホール	東町5－6		
8	総合体育館 （エスフォルタアリーナ八王子）	狭間町1453－1		
9	八王子城跡ガイダンス施設	元八王子町3－2664－2		

（都）一時滞在施設一覧　（平成31年1月末現在）

	名　　称	所　在　地	電　話	ＦＡＸ
1	八王子拓真高等学校	台町3－25－1	非公開	
2	多摩職業能力開発センター八王子校	台町1－11－1		
3	松が谷高等学校	松が谷1772		
4	首都大学東京　南大沢キャンパス	南大沢1－1		

指定避難所（避難所）①　指定避難所（避難所）②　一時滞在施設

対象外のケース ／ シート毎に年月日が分かれている、1 枚のシート中にクロス集計表が横に並んでいる、等整えられていない Excel は、データをコピーと貼り付けすることにより、一括処理しやすい形式に変換します。



手順（ケース[A]）

下記の手順で Excel から CSV にデータを変換します。

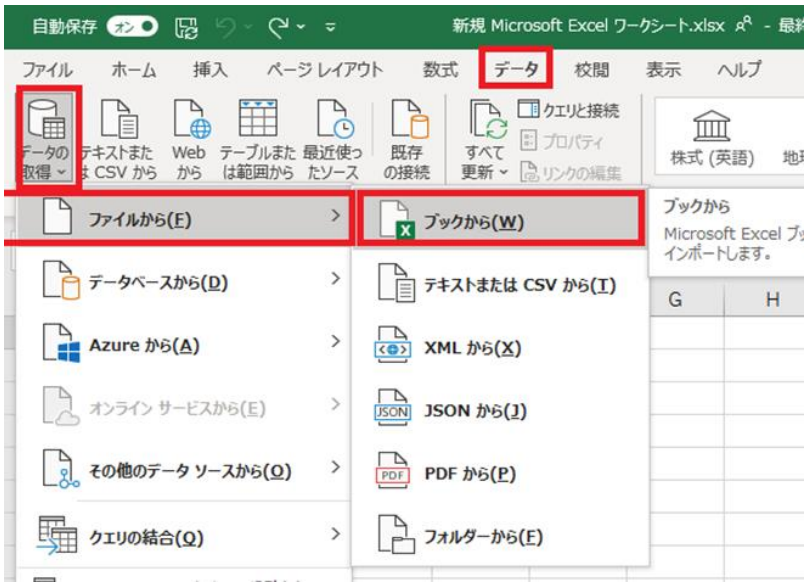
【お願い】： データ取得元（ソース）のファイルは閉じた状態で作業を行ってください。

① **Excel で整形、加工したいファイルをインポートします。** **（詳細： p.28）**
② **Excel の Power Query 機能で、データを整形します。** **（詳細： p.29）**
③ **Excel の Power Query 機能で（年次、月次等）定期発生するデータを整形する** **（詳細： p.43）**

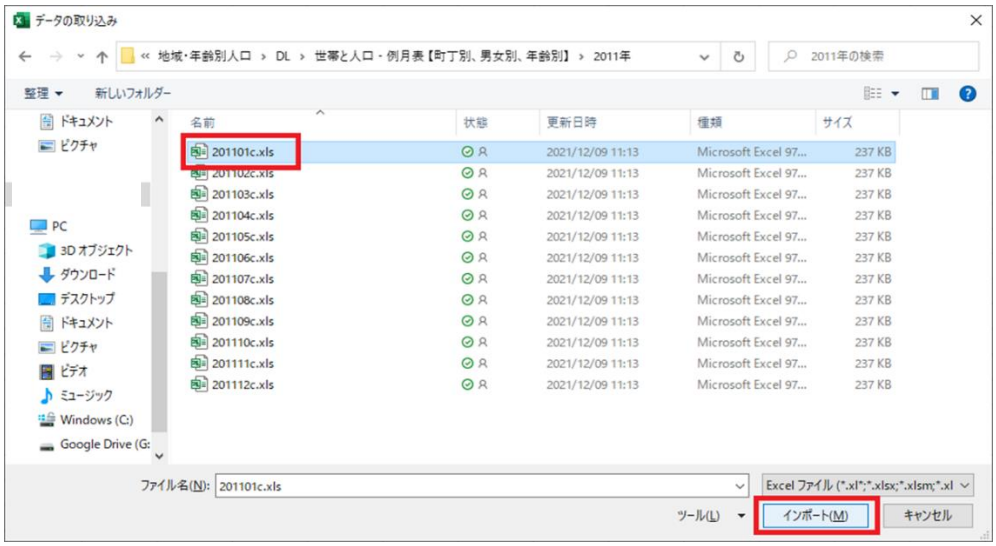
① Excel で整形、加工したいファイルをインポートする

1. データ取得元（ソース）のファイルを開いている場合は、閉じておく

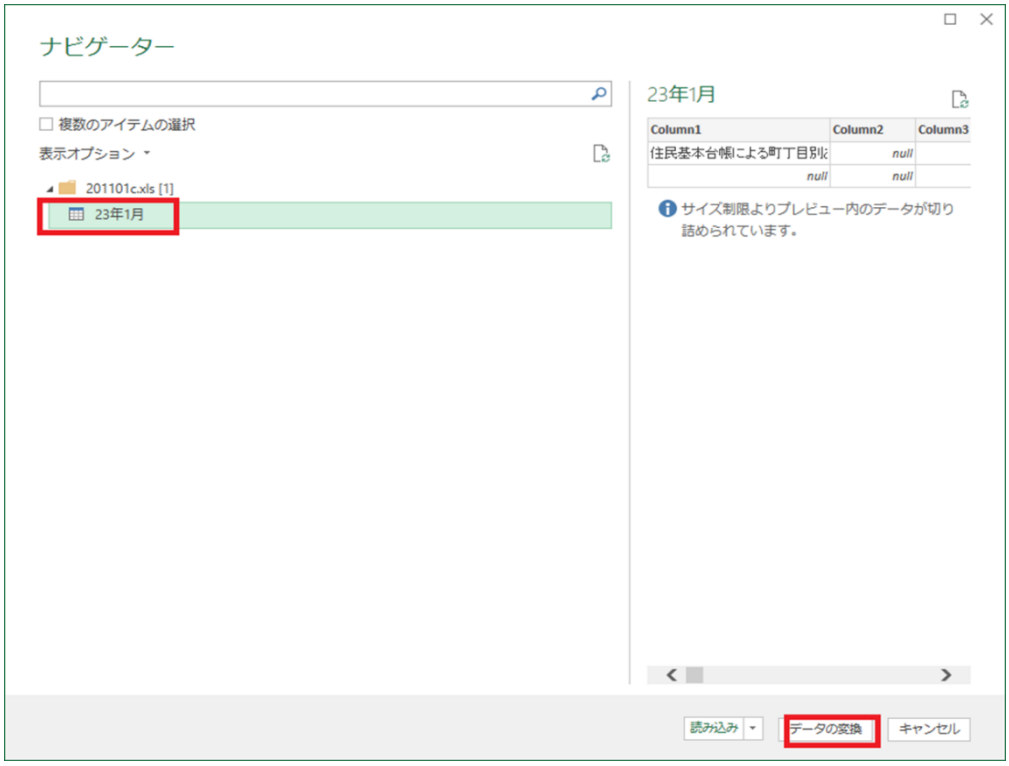
2. [データ]タブ→[データの取得]→「ファイルから」→「ブックから」をクリックする



3. Excel ファイルをクリックして選び、[インポート]をクリックする



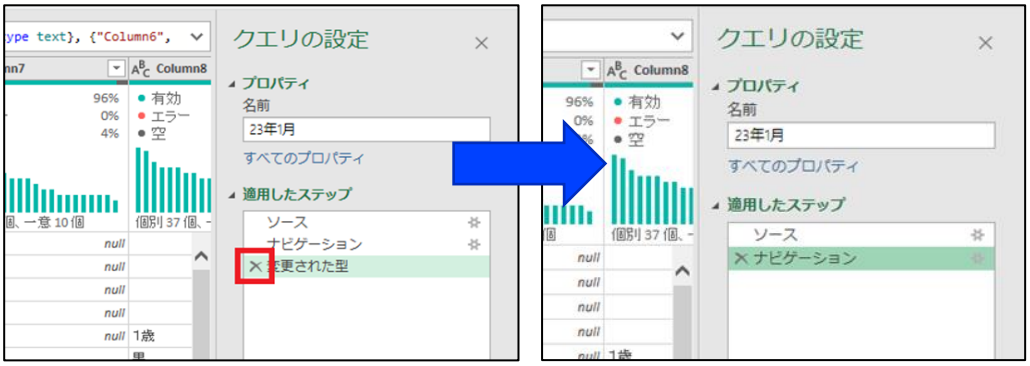
4.　変換するアイテムをクリックして選び、[データの変換]をクリックする



ここでの「アイテム」は、Excel のシートです。シート単位で複数表示されている場合は、必要なシートを選びます。

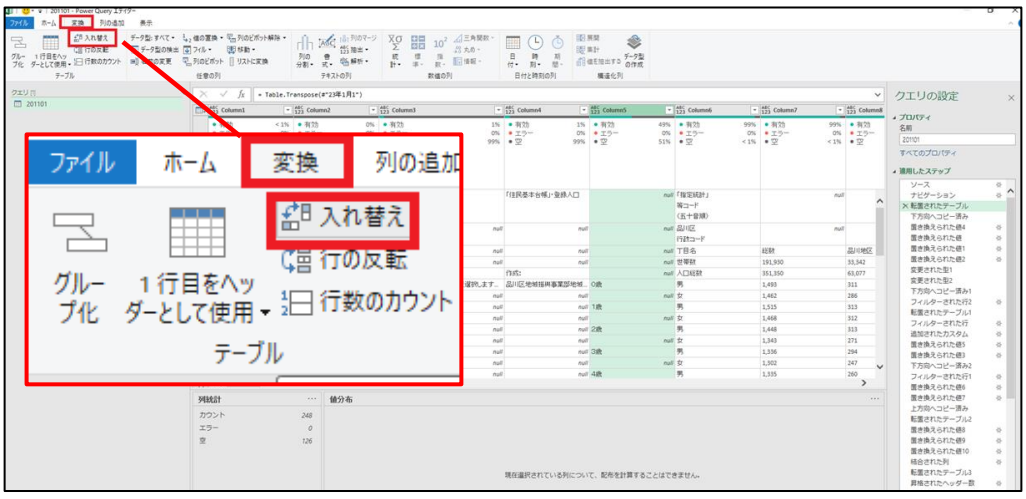
② Excel の Power Query 機能でデータを整形する

1.　「クエリの設定」で、「変更された型」ステップをクリックして削除する



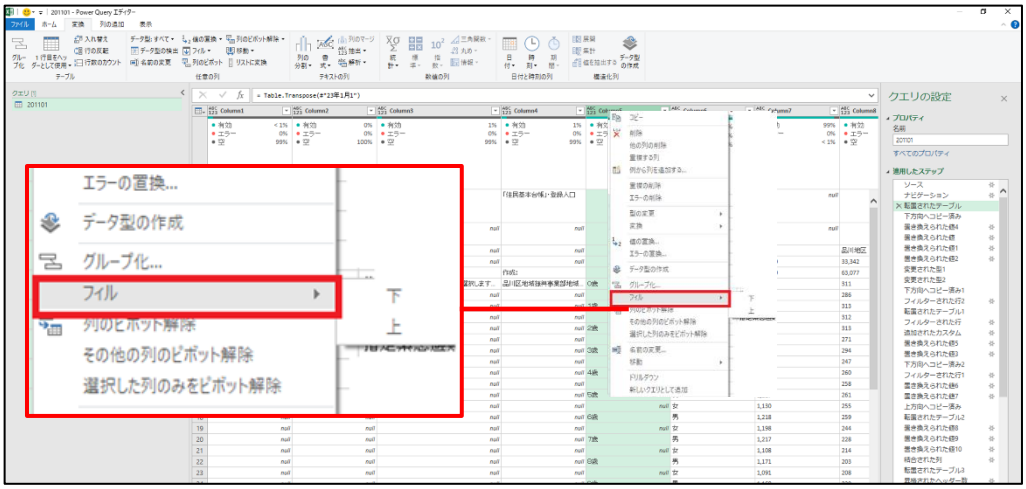
「変更された型」は操作を行うと自動的に追加されますが、手順は不要なので削除します。

2.　[変換]タブ→[入れ替え]をクリックする



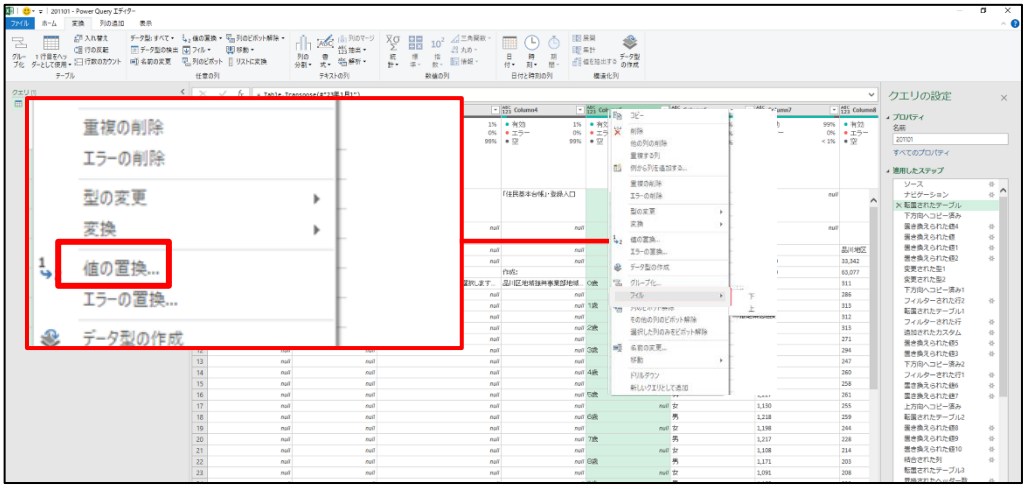
ここではヘッダー行を整形するため、行列を入れ替える処理を行います。

3. 年齢の列を右クリックし、[フィル]→[下]または[上]をクリックする

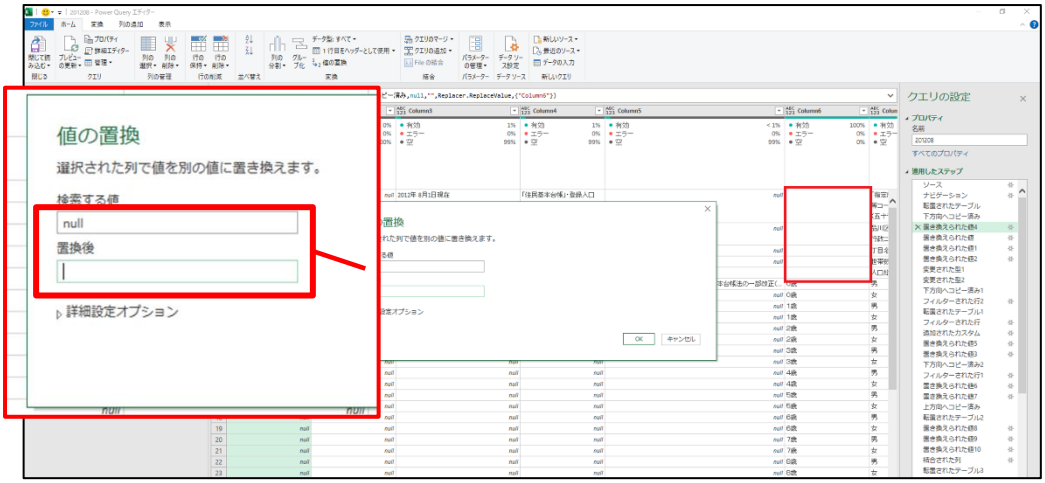


[下]または[上]は、コピーしたい方向を示します。
[下]は上から下の方向にデータをコピーし、[上]は下から上の方向にデータをコピーします。

4. 年齢の列が選択された状態で右クリックし、[値の置換]をクリックする



5. 置換する値に「null」、置換後に「“”」（ダブルクォーテーション）と入力し、[OK]をクリックする



ここでの作業は、後の手順で日付データをフィルするために行います。
「null」は何のデータも含まれない、または長さ 0 の空文字列という意味です。
画!像

6. 年月日の列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

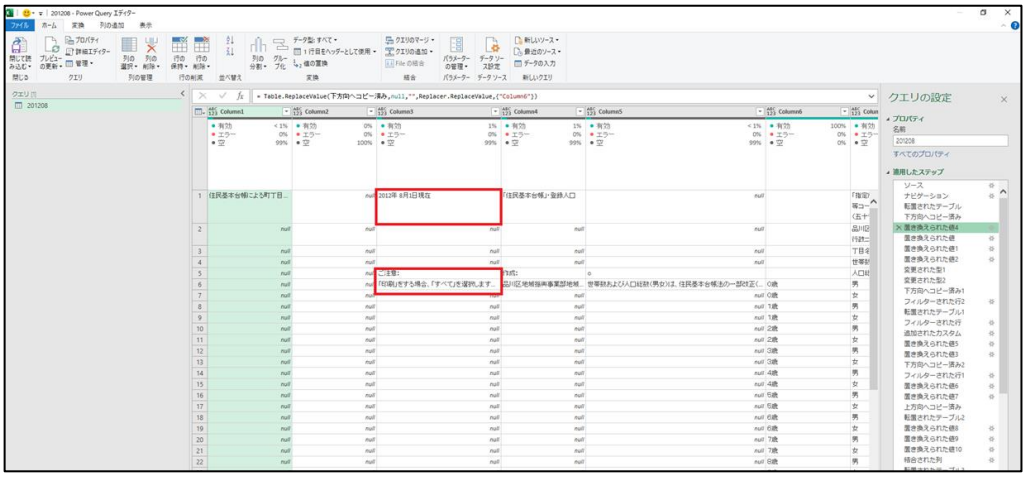
年月日列で不要データの値を置換します。

・対象列を選択した状態で右クリックし、「値の置換」を選ぶ。
　置換する値「日付データ以外の文言」 → 置換後「null」

年月日列を日付型でエラーとならない形式に整形します。

・対象列選択状態で　右クリック 「値の置換」
　置換する値：不要な()や現在、未等の文言 → 置換後「""」（ダブルクォーテーション）

ここでの不要データ置換作業は、後の手順で日付データをフィルするために行います。

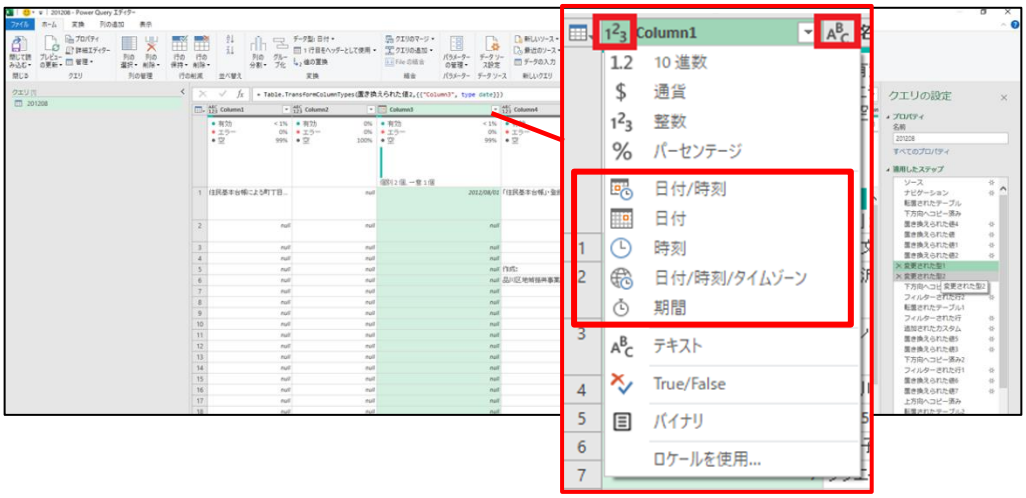


日付型の列での注意点です。

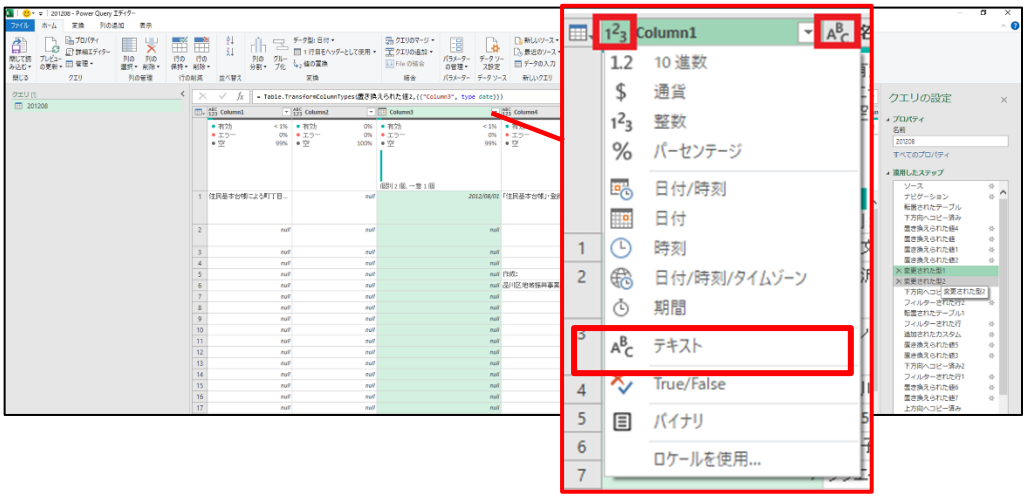
和暦表記も可能ですが、「元号」表記、「全角数字」、「漢数字」は取り扱えません。

× 令和元年１月１日　（元号）	○ 令和1年1月1日
× 令和１年１月１日　（全角数字）	○ 令和1年1月1日
× 令和一年一月一日　（漢数字）	○ 令和1年1月1日
× ２０１９０１０１	○ 20190101
× ２０１９/1/1	○ 2019/1/1

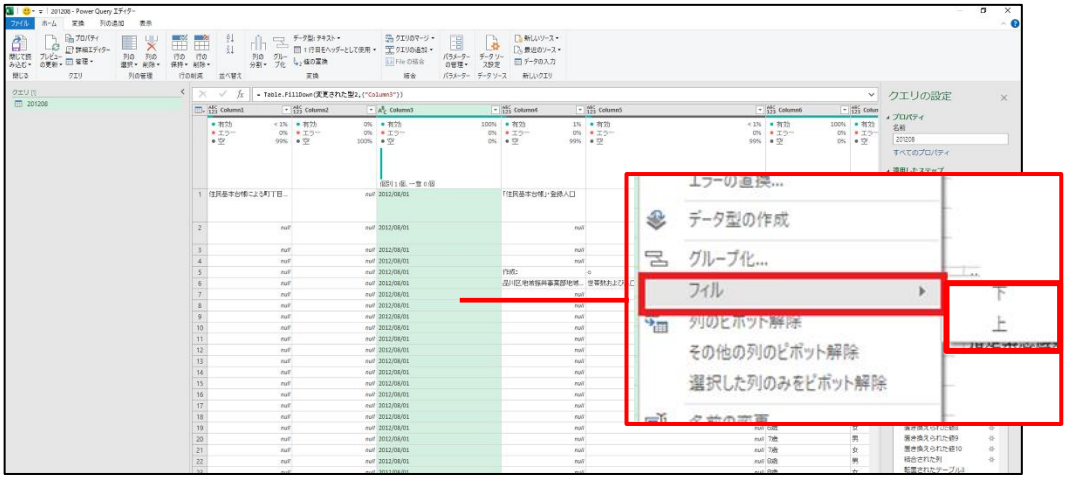
7. 年月日のヘッダー上部にある「ABC123」アイコンをクリックし、日付型を選択する



8. 年月日のヘッダー上部にある「ABC123」アイコンをクリックし、テキスト型を選択する



9. 年月日の列を右クリックし、[フィル]→[下]または[上]をクリックする



[下]または[上]は、コピーしたい方向を示します。
[下]は上から下の方向にデータをコピーし、[上]は下から上の方向にデータをコピーします。

ここから、不要な行を非表示にします。

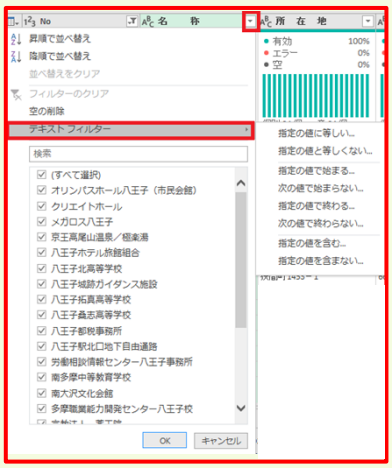
不要な行を非表示にする方法は 2 種類あります。

1. **対象列の右側「▼」をクリックし、数値フィルターまたはテキストフィルターを表示する**
2. **「規定の値と等しくない」を選び、[OK]をクリックする**

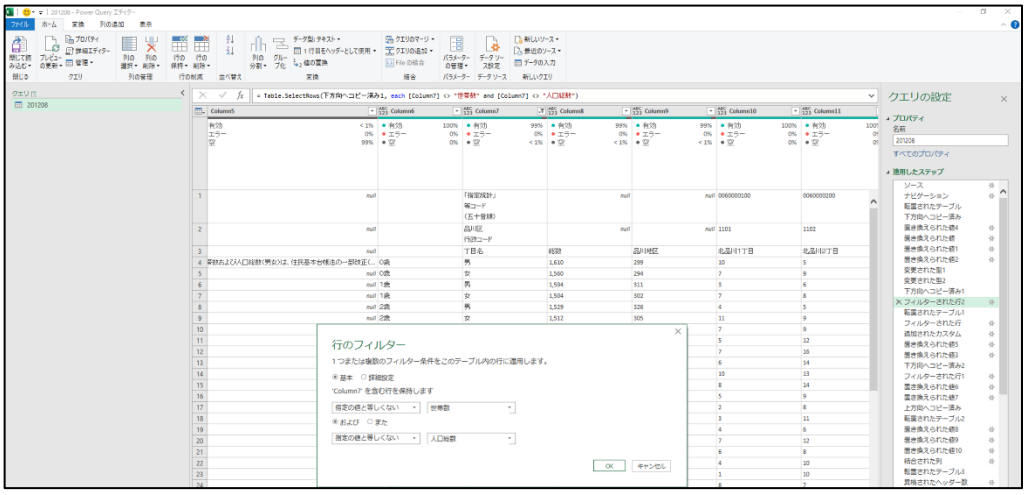
対象行が非表示になります。

数値フィルター

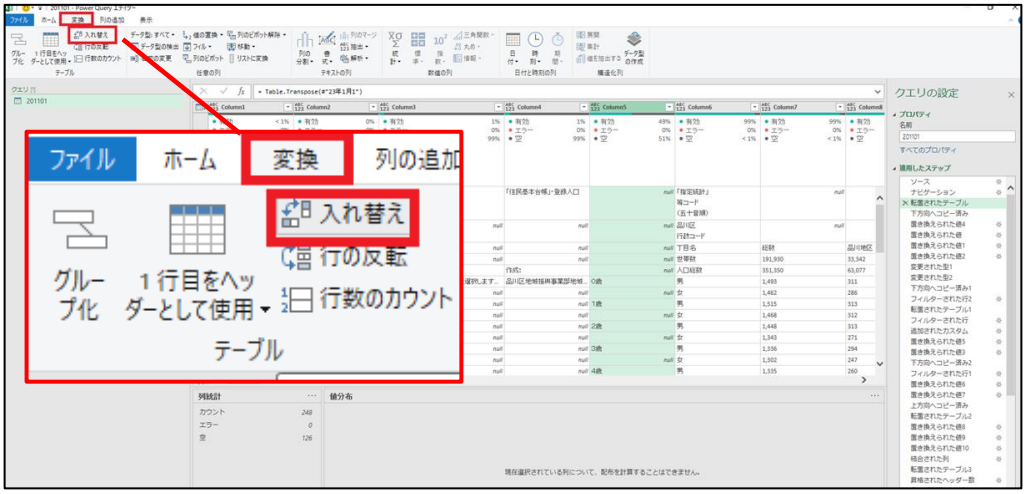
テキストフィルター



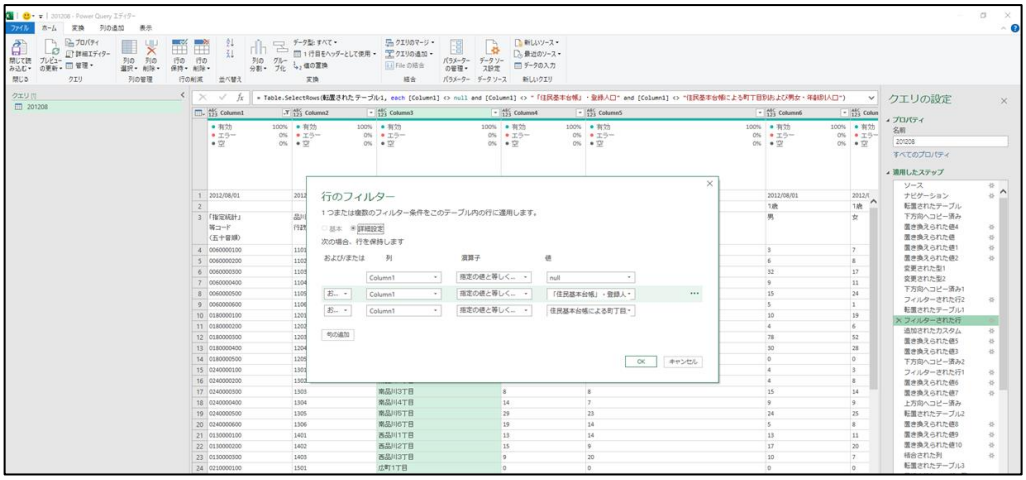
10. 年月日の列を右クリックし、除外するデータを非表示にする



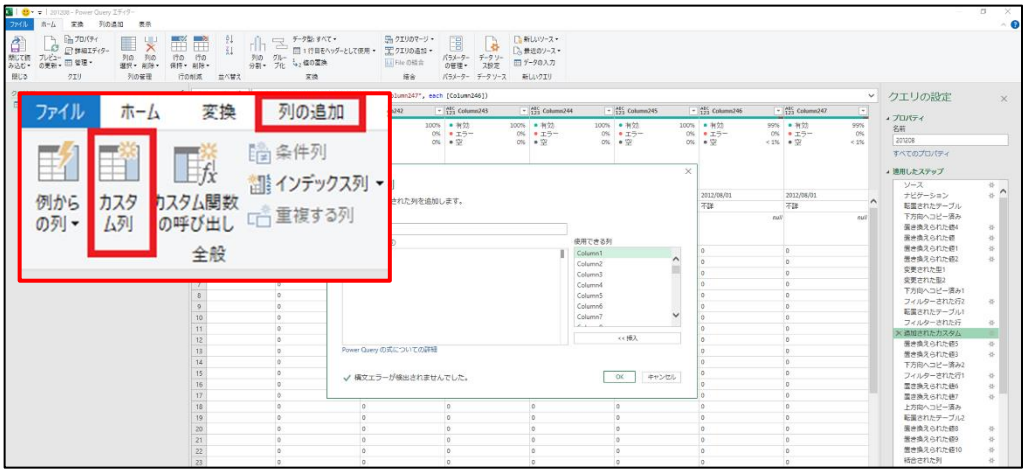
11. [変換]タブ→[入れ替え]をクリックする



12. 対象列の不要行を削除するため、対象列の右側「▼」をクリックし、除外するデータを非表示にする



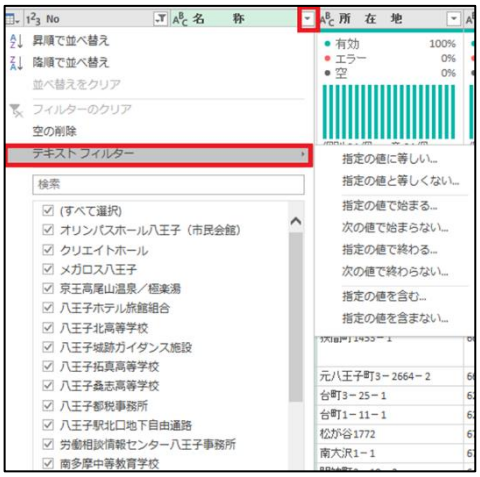
13. [列の追加]タブ→[カスタム列]をクリックする



今回は最終列をコピーし、年月日のカスタム列を作成します。

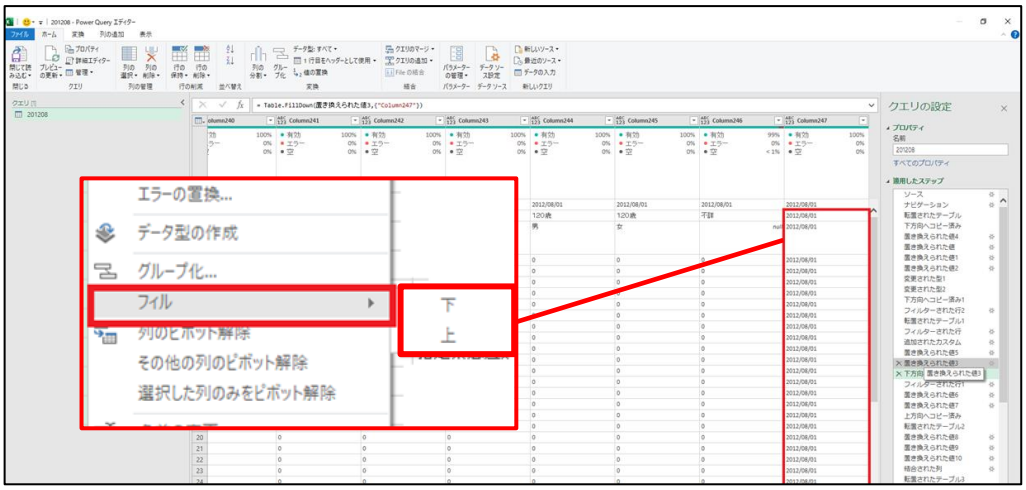
14. 追加したカスタム列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

置換する値「不詳」 → 置換後「null」
置換する値「0」 → 置換後「null」



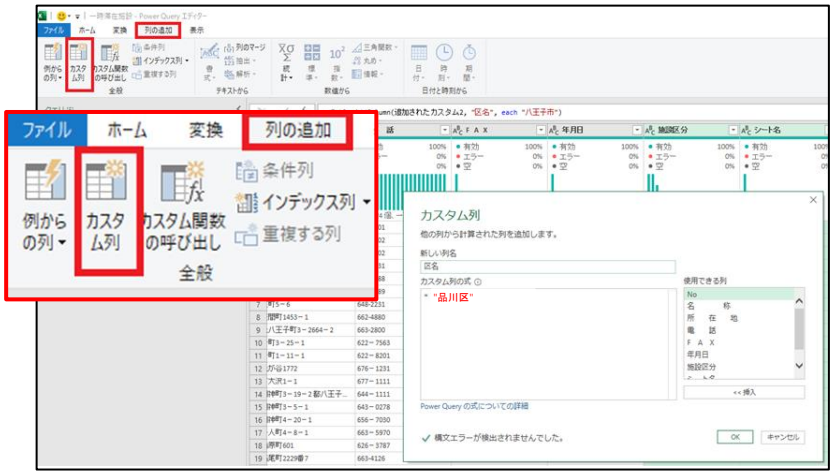
ここでの不要データ置換作業は、後の手順で日付データをフィルするために行います。

15. 年月日の列を右クリックし、[フィル]→[下]または[上]をクリックする



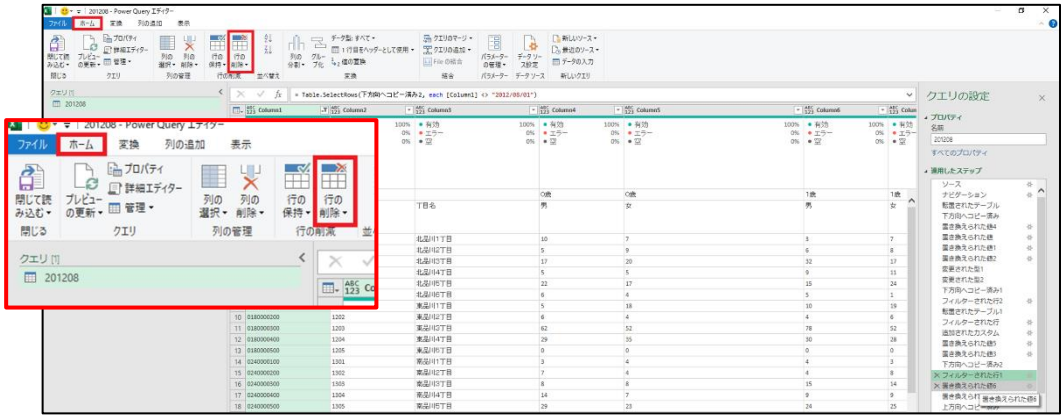
[下]または[上]は、コピーしたい方向を示します。
[下]は上から下の方向にデータをコピーし、[上]は下から上の方向にデータをコピーします。

16. [列の追加]タブ→[カスタム列]をクリックし、対象の区名をカスタム列の式に入力し、[OK]をクリックする



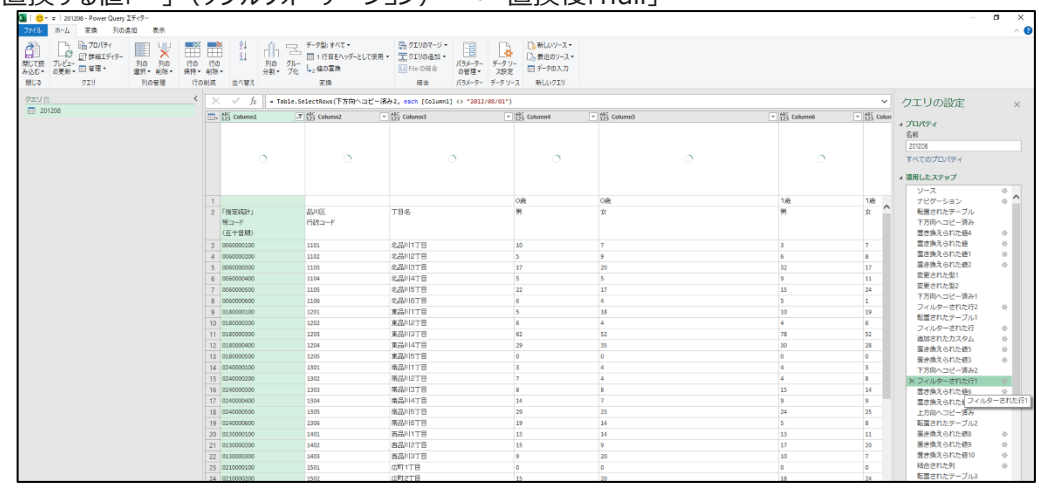
今回の例では”渋谷区”を入力します。

17. [ホーム]タブ→[行の削除]→「上位の行の削除」をクリックし、削除する上位の行数を入力する



18. 対象列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

置換する値「””」（ダブルクォーテーション）　→　置換後「null」

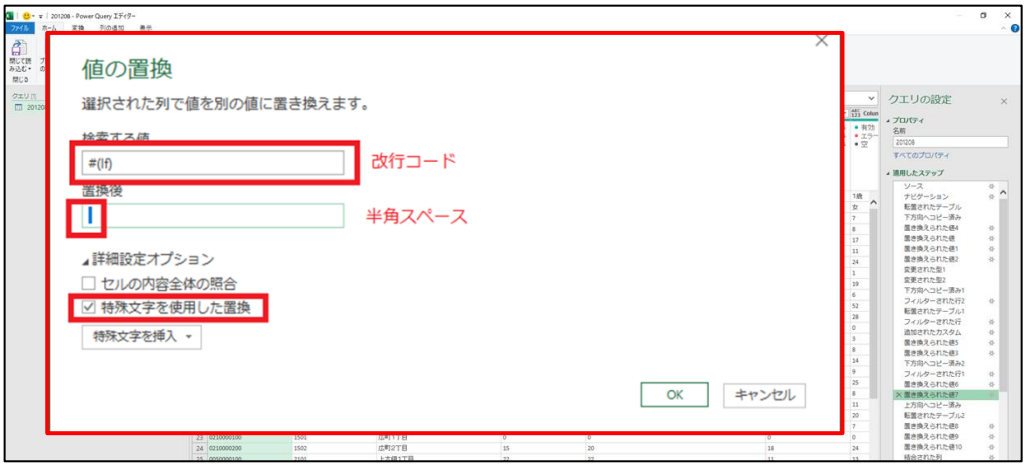


ここではヘッダー行ズレを整形します。

19. 対象列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

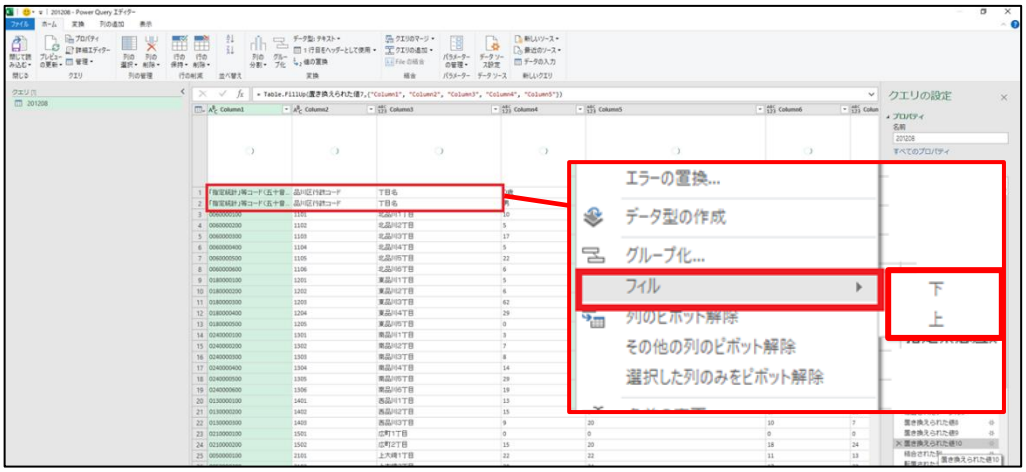
置換する値「#(lf)」（改行コード）　→　置換後「 」（半角スペース）
「詳細設定オプション」の「特殊文字を使用した置換」にチェックを付けておきます。

ここでは改行コードを削除しています。

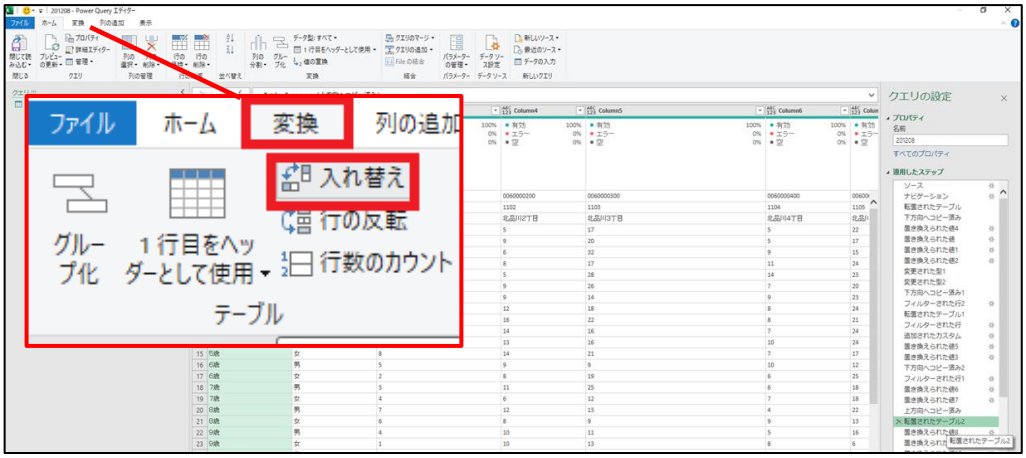


20. 対象列を右クリックし、[フィル]→[下]または[上]をクリックする

ここではヘッダー行データをコピーしています。



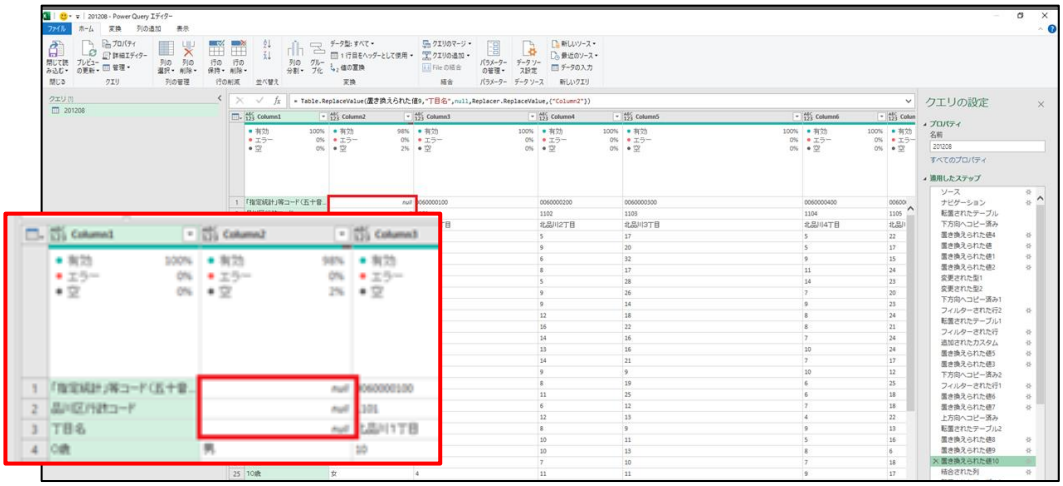
21. [変換]タブ→[入れ替え]をクリックする



22. 対象列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

置換する値「Column1 と Column2 で同じ文言」 → 置換後「Column2 の値を null」

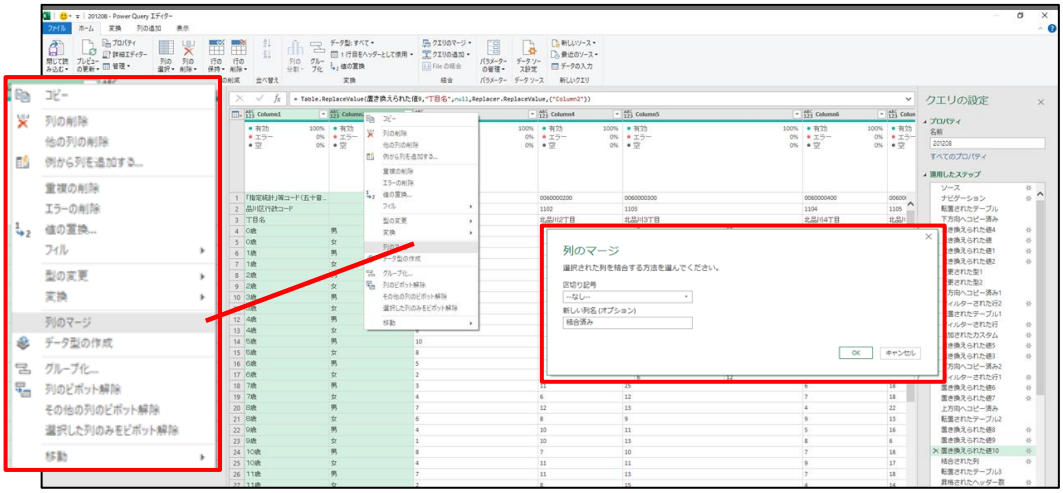
ここでは列をマージ（結合･統合）するため、不要な文言を削除しています。



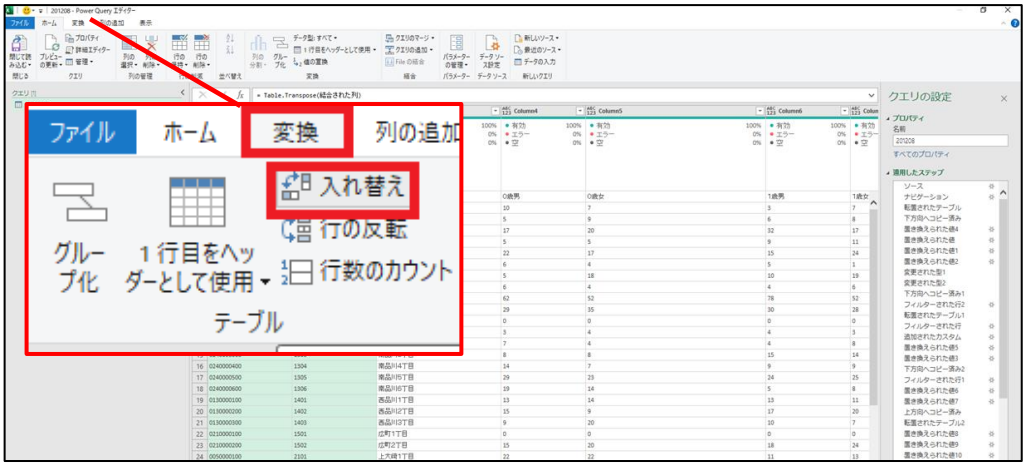
23. 対象列が選択された状態で右クリックし、[列のマージ]をクリックし、値を置換する

「段区切り記号」や「新しい列名」は内容に合わせて設定します。

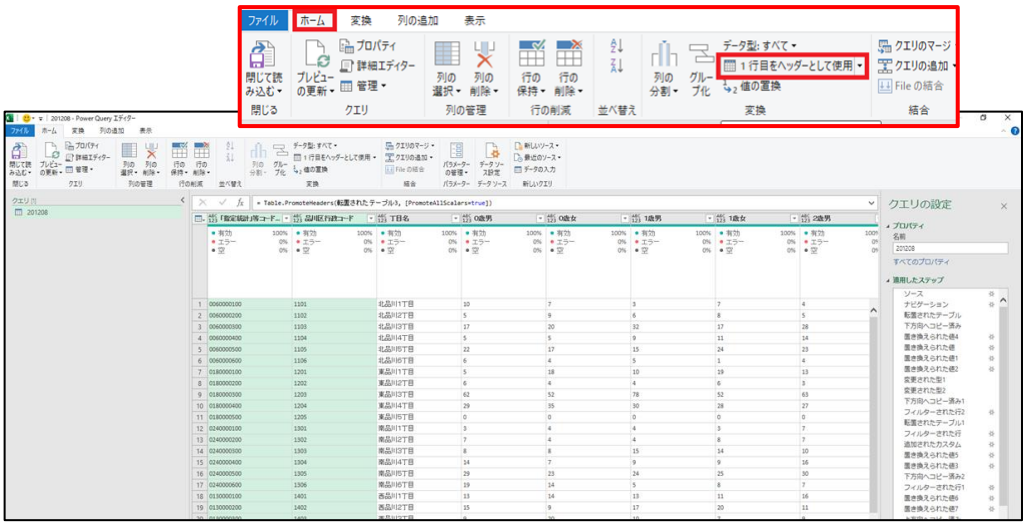
ここではヘッダー列を作成するため、列をマージしています。



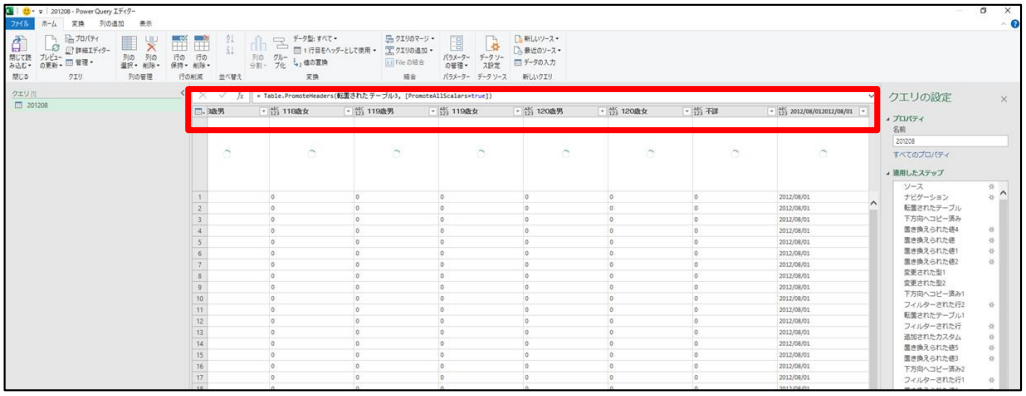
24. [変換]タブ→[入れ替え]をクリックする



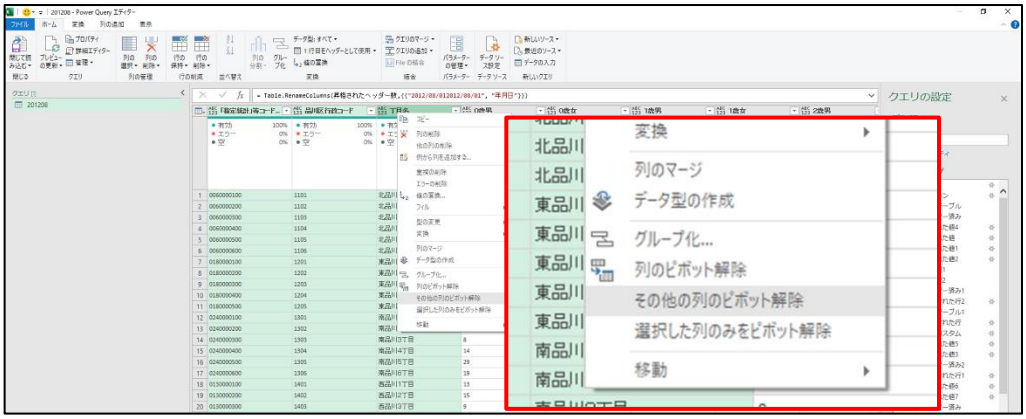
25. [ホーム]タブ→[1 行目をヘッダーとして使用]をクリックする



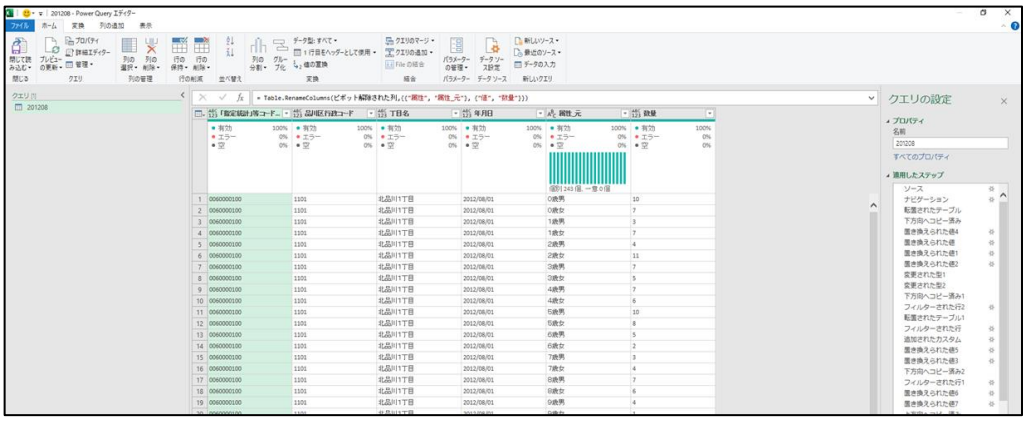
26. 追加した「カスタム列」や「列名」の名前を指定しなかった列名があれば、列をダブルクリックして任意の列名を入力する



27. 対象列が選択された状態で右クリックし、[そのほかの列のピボット解除]をクリックする



28. ピボット解除により作成されたヘッダー列名に、任意の列名を入力する

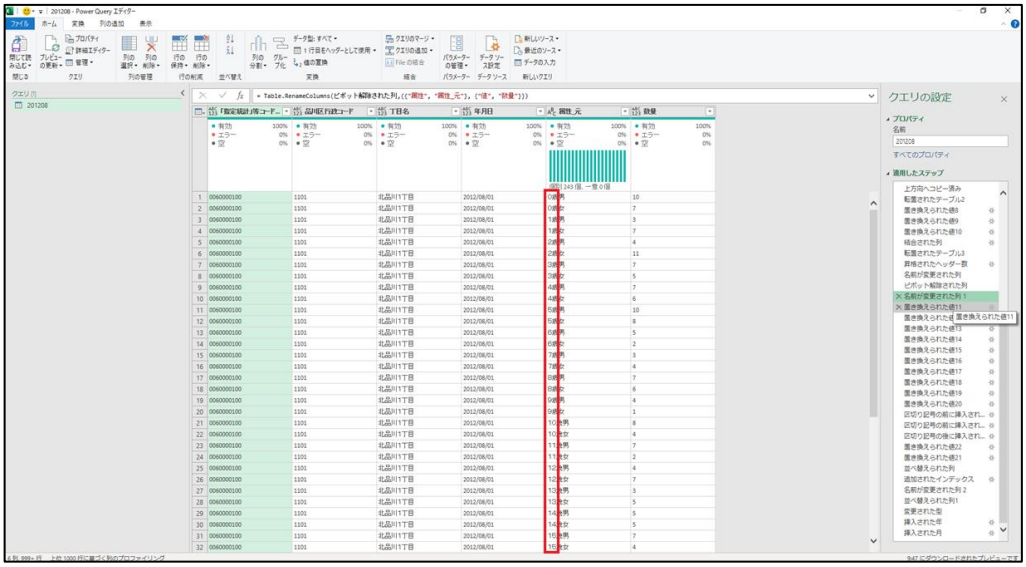


29. 対象列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

置換する値「全角数値０」 →　置換後「半角数値０」

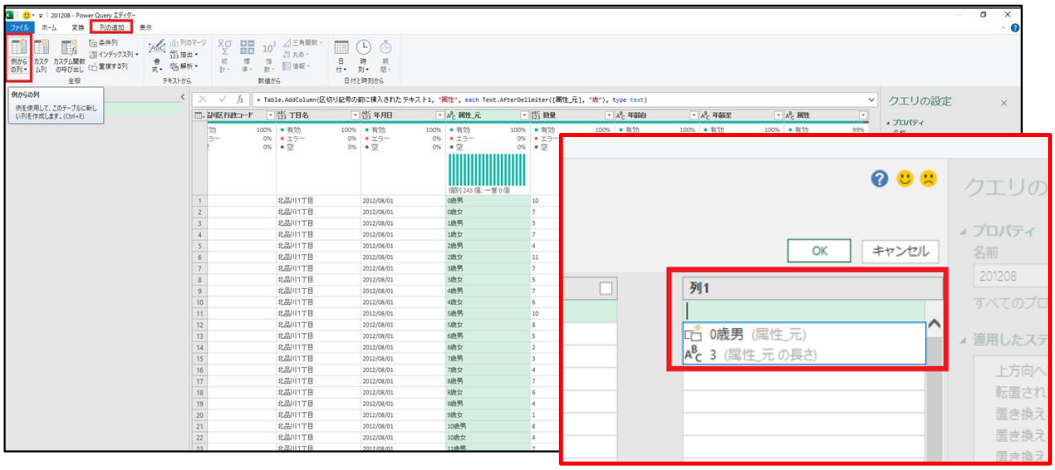
<この間の数字分も、手順を繰り返します。>

置換する値「全角数値９」 →　置換後「半角数値９」



30. 対象列が選択された状態で右クリックし、[列の追加]タブ→[例からの列]→[選択範囲から]をクリックし、任意の値に変更して例を作成する

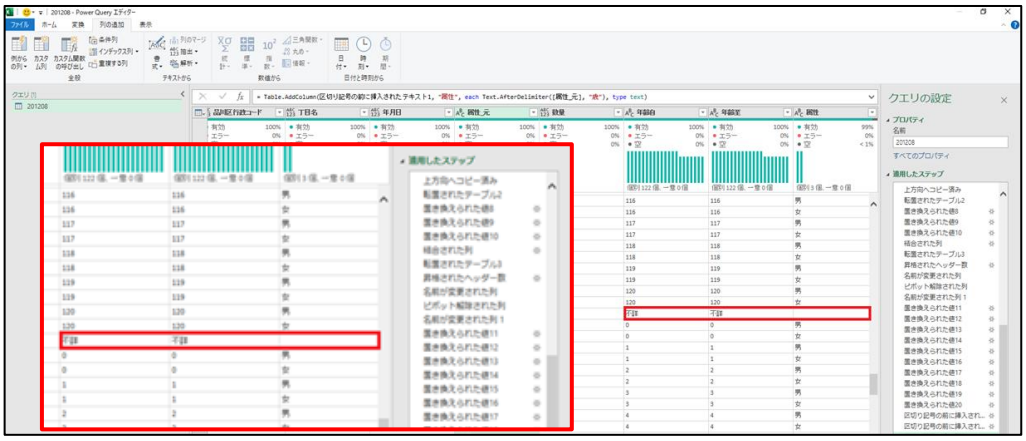
対象列：属性、年齢自、年齢列追加　属性_元列
それぞれの列から必要範囲を切り出して例を作成します。
全データで想定通りの値が入力されているか確認してください。



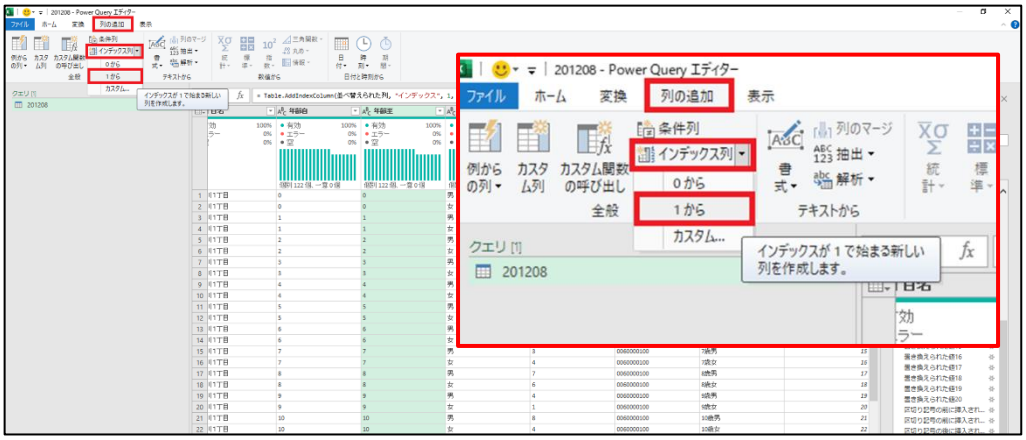
31. 対象列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

置換する値「不詳」 → 置換後「999」
置換する値「""」（ダブルクォーテーション） → 置換後「その他」

ここでは不詳データを整形しています。

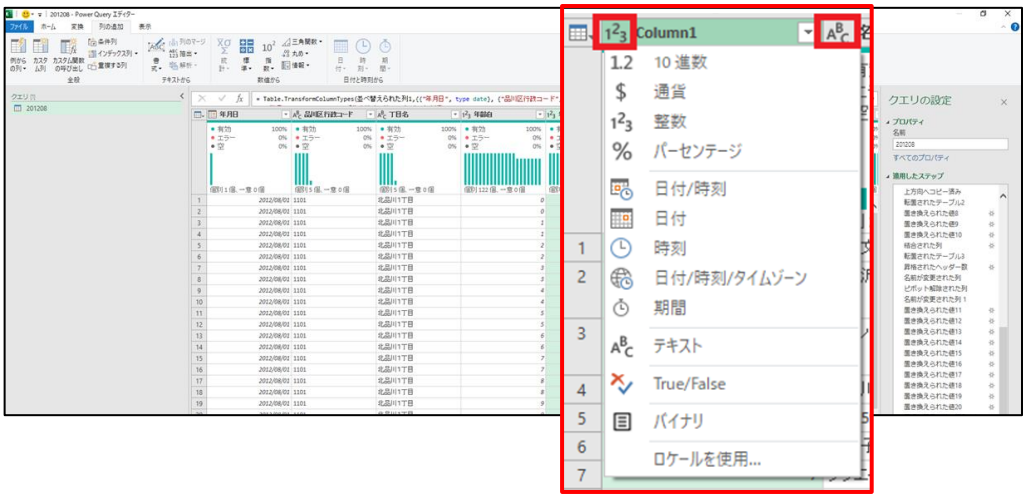


32. 追加する列が選択された状態で右クリックし、[列の追加]タブ→[インデックス列]→[1 から]をクリックする

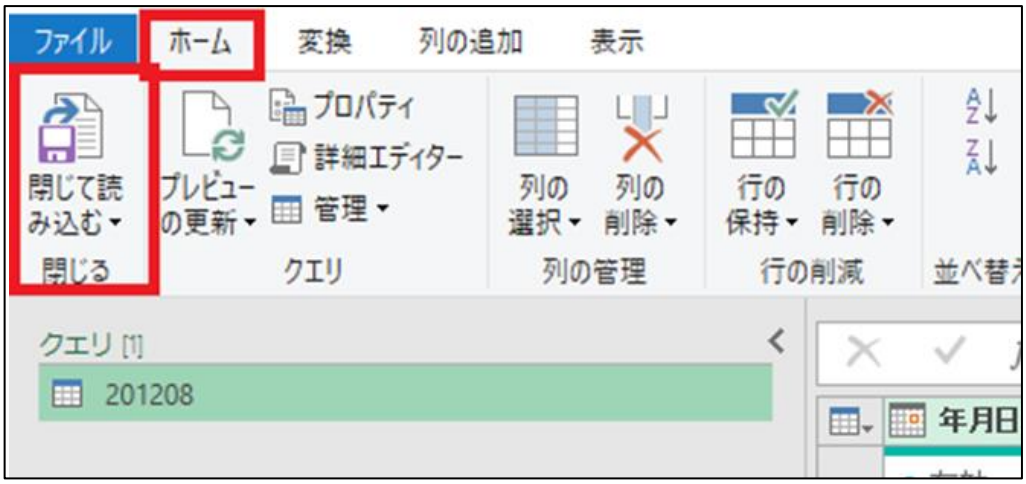


ここではインデックス列を追加しています。

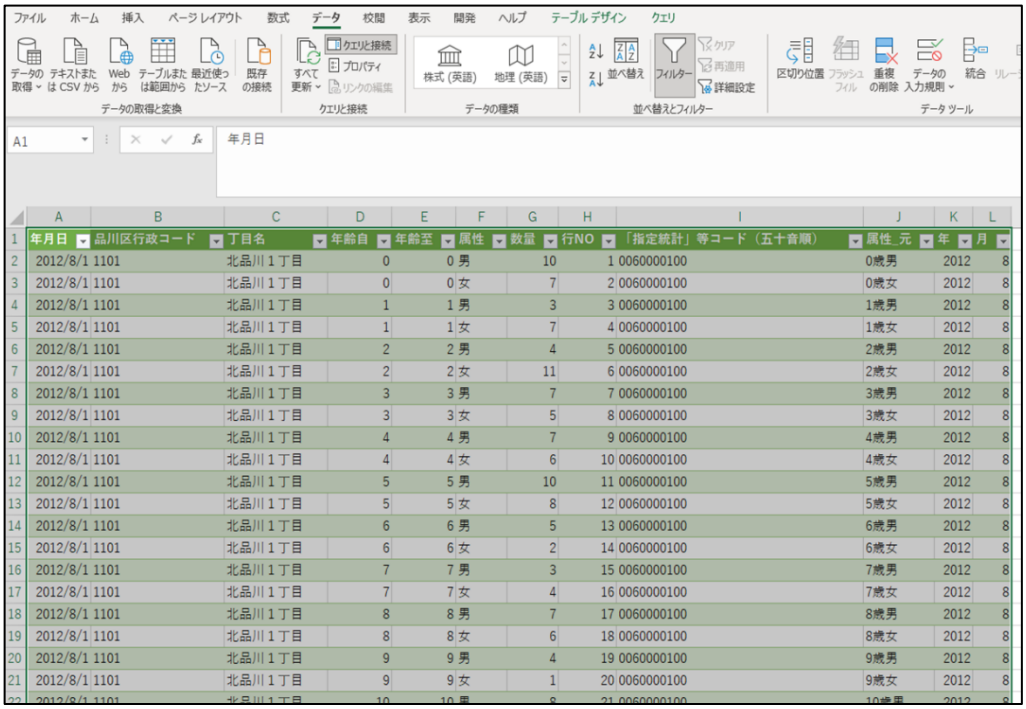
33. 各列のヘッダー上部にある「ABC123」アイコンをクリックし、任意の型を選択する



34. [ホーム]タブ→[閉じて読み込む]をクリックする



35. 読み込まれたデータを確認し、[ホーム]タブ→[閉じて読み込む]をクリックする



	A	B	C	D	E	F	G	H	I	J	K	L
1	年月日	品川区行政コード	丁目名	年齢自	年齢至	属性	数量	行NO	「指定統計」等コード（五十音順）	属性_元	年	月
2	2012/8/1	1101	北品川１丁目	0	0	男	10	1	0060000100	0歳男	2012	8
3	2012/8/1	1101	北品川１丁目	0	0	女	7	2	0060000100	0歳女	2012	8
4	2012/8/1	1101	北品川１丁目	1	1	男	3	3	0060000100	1歳男	2012	8
5	2012/8/1	1101	北品川１丁目	1	1	女	7	4	0060000100	1歳女	2012	8
6	2012/8/1	1101	北品川１丁目	2	2	男	4	5	0060000100	2歳男	2012	8
7	2012/8/1	1101	北品川１丁目	2	2	女	11	6	0060000100	2歳女	2012	8
8	2012/8/1	1101	北品川１丁目	3	3	男	7	7	0060000100	3歳男	2012	8
9	2012/8/1	1101	北品川１丁目	3	3	女	5	8	0060000100	3歳女	2012	8
10	2012/8/1	1101	北品川１丁目	4	4	男	7	9	0060000100	4歳男	2012	8
11	2012/8/1	1101	北品川１丁目	4	4	女	6	10	0060000100	4歳女	2012	8
12	2012/8/1	1101	北品川１丁目	5	5	男	10	11	0060000100	5歳男	2012	8
13	2012/8/1	1101	北品川１丁目	5	5	女	8	12	0060000100	5歳女	2012	8
14	2012/8/1	1101	北品川１丁目	6	6	男	5	13	0060000100	6歳男	2012	8
15	2012/8/1	1101	北品川１丁目	6	6	女	2	14	0060000100	6歳女	2012	8
16	2012/8/1	1101	北品川１丁目	7	7	男	3	15	0060000100	7歳男	2012	8
17	2012/8/1	1101	北品川１丁目	7	7	女	4	16	0060000100	7歳女	2012	8
18	2012/8/1	1101	北品川１丁目	8	8	男	7	17	0060000100	8歳男	2012	8
19	2012/8/1	1101	北品川１丁目	8	8	女	6	18	0060000100	8歳女	2012	8
20	2012/8/1	1101	北品川１丁目	9	9	男	4	19	0060000100	9歳男	2012	8
21	2012/8/1	1101	北品川１丁目	9	9	女	1	20	0060000100	9歳女	2012	8

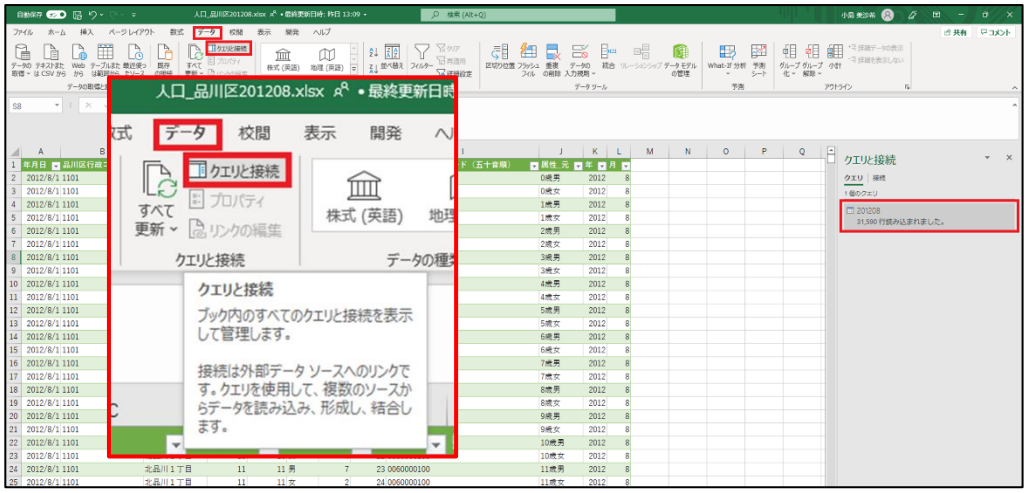
36. ファイルの種類欄で「テキスト（タブ区切り）」を指定して CSV を保存する

③ Excel の Power Query 機能で（年次、月次等）定期発生するデータを整形する

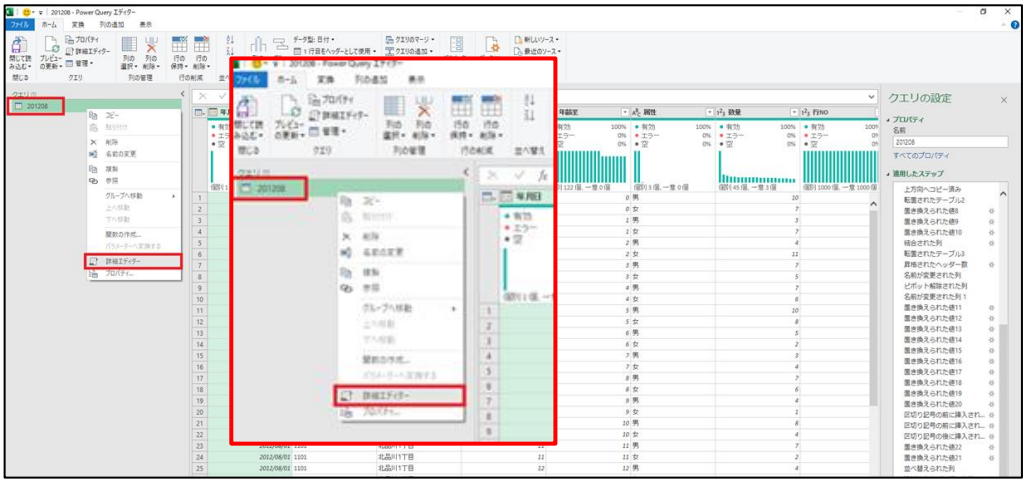
①②で処理した Power Query の詳細エディターを複製し、月毎等でエクセルブックが分かれている場合に発生する定期発生データも変換を行います。

【お願い】： データ取得元（ソース）のファイルは閉じた状態で作業を行ってください。

1. [データ]タブ→「クエリと接続」とクリックし、該当クエリをダブルクリックする



2. [ホーム]タブ→「該当クエリ」を右クリックし、「詳細エディター」をクリックする



3. 詳細エディターのコードを全選択し、コピーする

4. 新規エクセルブックで詳細エディターの編集を行う

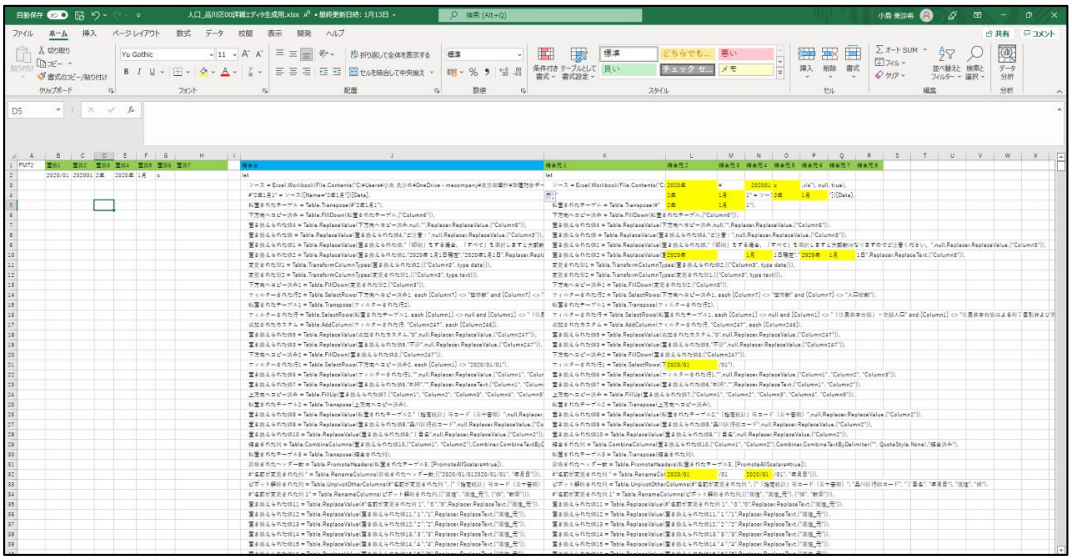
●B～H 列：フォルダ名、シート名、ブック名、年月部分等、置換用の文字列を入力します。
・B 列：YYYY/MM ※西暦(半角)
・C 列：YYYYMM　※西暦(半角)
・D 列：YY 年　　※和暦(シート名用)(半角)
・E 列：YYYY 年　※西暦(半角)
・F 列：MM 月　　※(半角)
・G 列：c ブック名 YYYYMM 以外文字列が c(小文字半角)以外の場合は個別対応必要
・H 列：※上記年月で全角文字使用等例外時に設定

ブック名、シート名、年月日に関わる部分で、記載内容と異なる状況がある場合（品川区の月ごとデータ等）、FAQ「Power Query」（P. 117）もご参考ください。

●J 列：生成後の文言出力用の数式を入力します。
・J 列：=K2&L2&M2&N2&O2&P2&Q2&R2

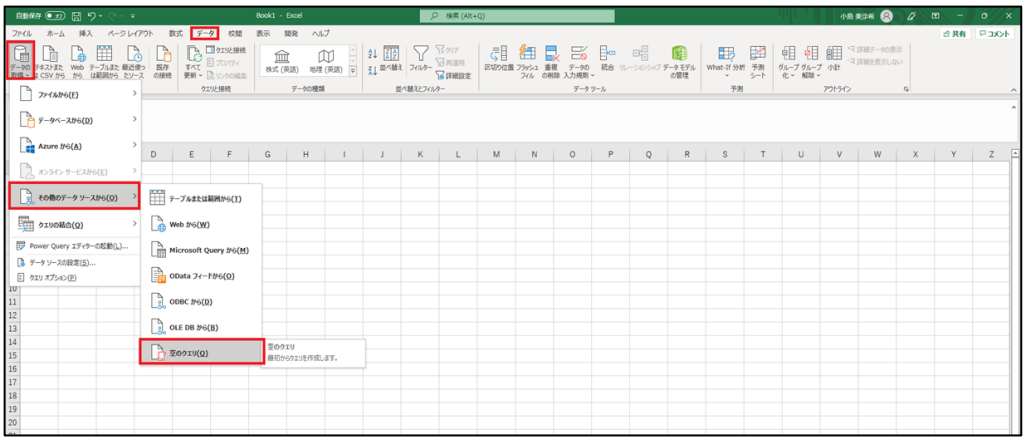
●K～R 列：コピーした詳細エディターを分割して入力
※B～H 列に設定した置換を行う必要のある文字列部分は列を分割します。
・置換文字列：=B2　※該当の置換文字列が設定されている B～H 列を設定すること。

上記生成用の範囲をコピーアンドペーストし、B～H 列の年月を該当年月に変更し、詳細エディターを生成します。

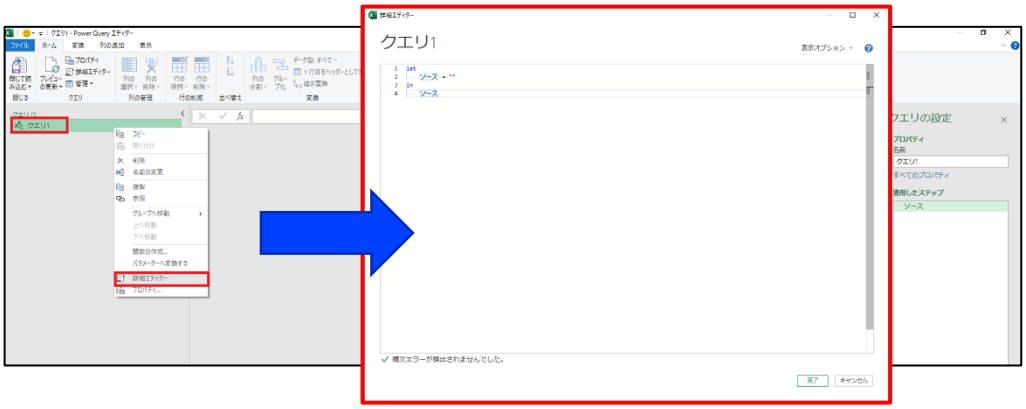


行については下記のように表記されます。
2 行目　ソース=…　：フォルダ名、ブック名記載箇所
3 行目　#"2 年 1 月 1" = ソース{[Name="2 年 1 月"]}[Data],：シート名記載箇所

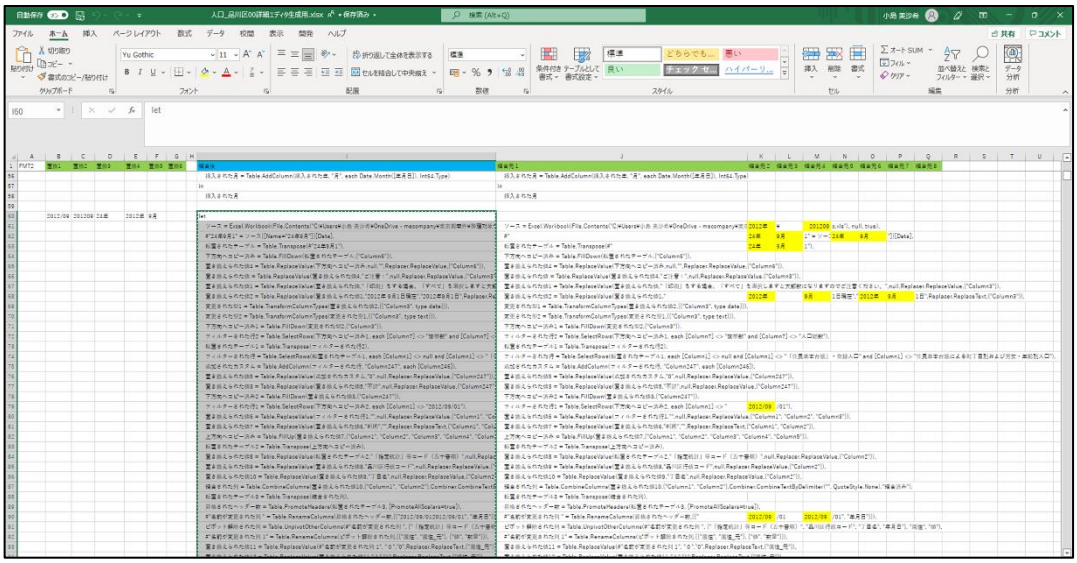
5. Excel で新規ブックを開き、[データ]タブ→[データの取得]→「その他のデータソースから」→「空のクエリ」をクリックする



6. [ホーム]タブ→「該当クエリ」を右クリックし、新規ブックの Power Query 詳細エディターを表示する

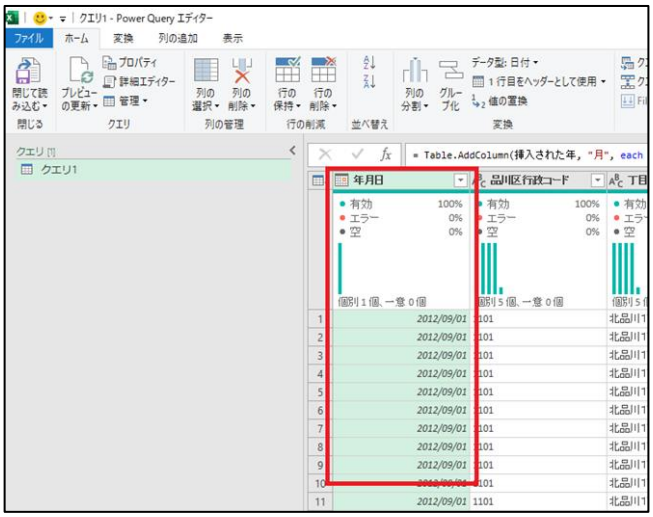


7. 詳細エディター生成用エクセルブックで、該当年月の J 列をコピーする

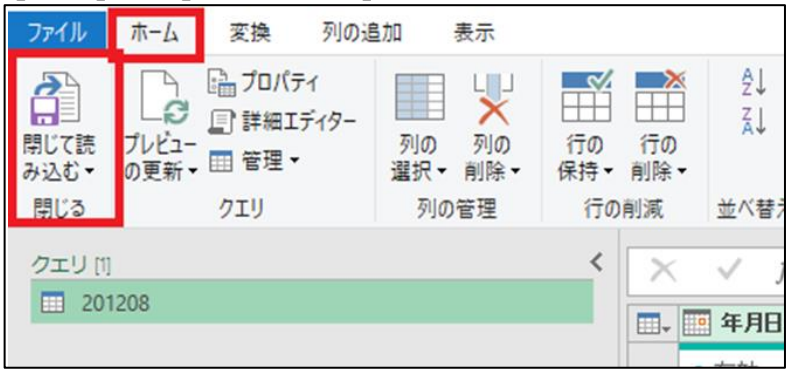


8. コピーした該当年月のＪ列を貼り付け、[完了]をクリックする

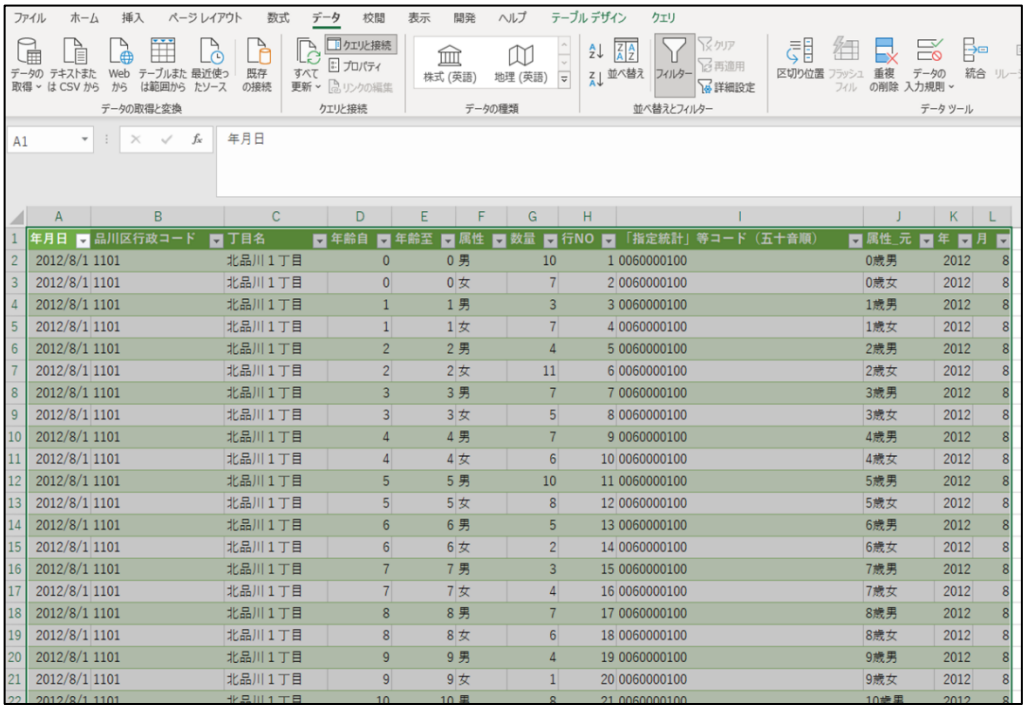
9. 正しく年月日が表示されていることを確認する



10. [ホーム]タブ→[閉じて読み込む]をクリックする



11. 読み込まれたデータを確認、データを全選択し、コピーする



年月日	品川区行政コード	丁目名	年齢自	年齢至	属性	数量	行NO	「指定統計」等コード（五十音順）	属性_元	年	月
2012/8/1	1101	北品川１丁目	0	0	男	10	1	0060000100	0歳男	2012	8
2012/8/1	1101	北品川１丁目	0	0	女	7	2	0060000100	0歳女	2012	8
2012/8/1	1101	北品川１丁目	1	1	男	3	3	0060000100	1歳男	2012	8
2012/8/1	1101	北品川１丁目	1	1	女	7	4	0060000100	1歳女	2012	8
2012/8/1	1101	北品川１丁目	2	2	男	4	5	0060000100	2歳男	2012	8
2012/8/1	1101	北品川１丁目	2	2	女	11	6	0060000100	2歳女	2012	8
2012/8/1	1101	北品川１丁目	3	3	男	7	7	0060000100	3歳男	2012	8
2012/8/1	1101	北品川１丁目	3	3	女	5	8	0060000100	3歳女	2012	8
2012/8/1	1101	北品川１丁目	4	4	男	7	9	0060000100	4歳男	2012	8
2012/8/1	1101	北品川１丁目	4	4	女	6	10	0060000100	4歳女	2012	8
2012/8/1	1101	北品川１丁目	5	5	男	10	11	0060000100	5歳男	2012	8
2012/8/1	1101	北品川１丁目	5	5	女	8	12	0060000100	5歳女	2012	8
2012/8/1	1101	北品川１丁目	6	6	男	5	13	0060000100	6歳男	2012	8
2012/8/1	1101	北品川１丁目	6	6	女	2	14	0060000100	6歳女	2012	8
2012/8/1	1101	北品川１丁目	7	7	男	3	15	0060000100	7歳男	2012	8
2012/8/1	1101	北品川１丁目	7	7	女	4	16	0060000100	7歳女	2012	8
2012/8/1	1101	北品川１丁目	8	8	男	7	17	0060000100	8歳男	2012	8
2012/8/1	1101	北品川１丁目	8	8	女	6	18	0060000100	8歳女	2012	8
2012/8/1	1101	北品川１丁目	9	9	男	4	19	0060000100	9歳男	2012	8
2012/8/1	1101	北品川１丁目	9	9	女	1	20	0060000100	9歳女	2012	8

12. ②で作成した CSV ファイルを開き、最終行の次行に貼り付ける
13. ファイルの種類欄で「テキスト（タブ区切り）」を指定して CSV を保存する
14. 全データの CSV 更新が完了したら、途中行に追加されたヘッダー行を検索し行を削除する

手順（ケース[B]）

下記の手順で Excel から CSV にデータを変換します。

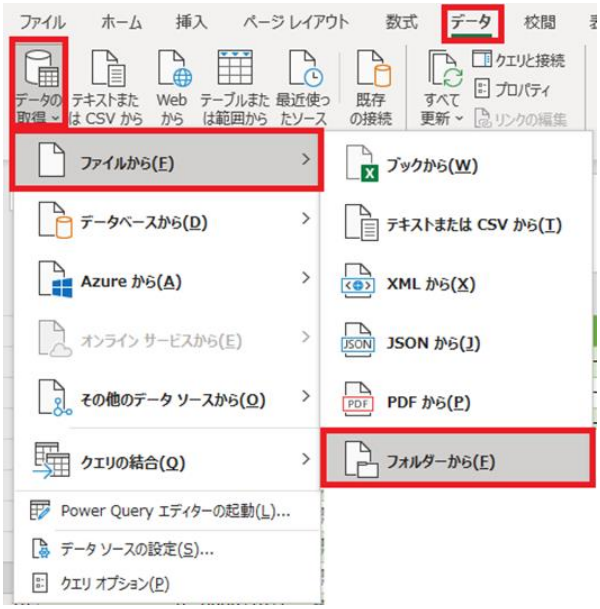
【お願い】： データ取得元（ソース）のファイルは閉じた状態で作業を行ってください。

① **Excel で整形、加工したいファイルをインポートします。** **（詳細： p. 48）**
② **Excel の Power Query 機能で、データを整形します。** **（詳細： p. 50）**

① Excel で整形、加工したいファイルをインポートする

1. **データ取得元（ソース）のファイルを開いている場合は、閉じておく**

2. **[データ]タブ→[データの取得]→「ファイルから」→「フォルダから」をクリックする**



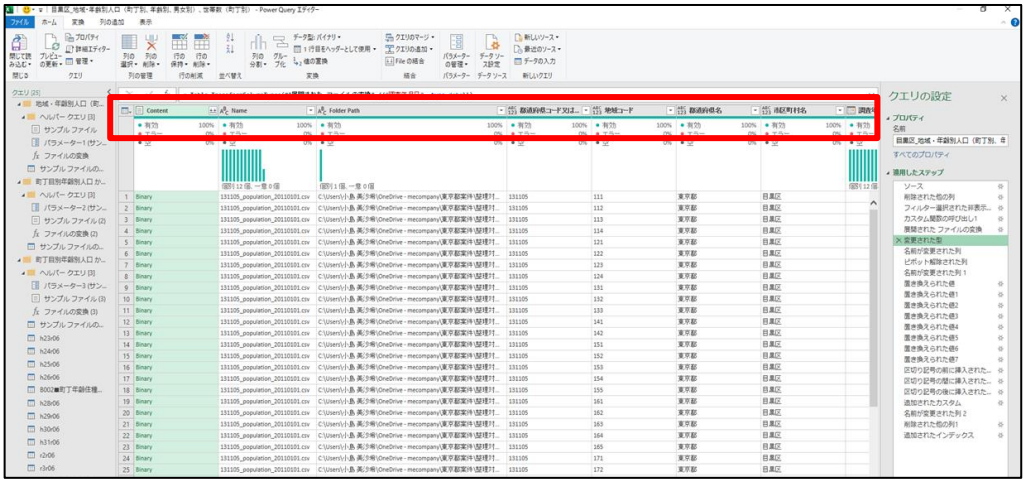
3. Excel ファイルをクリックして選び、[開く]をクリックする

4. [データの結合と変換]をクリックする

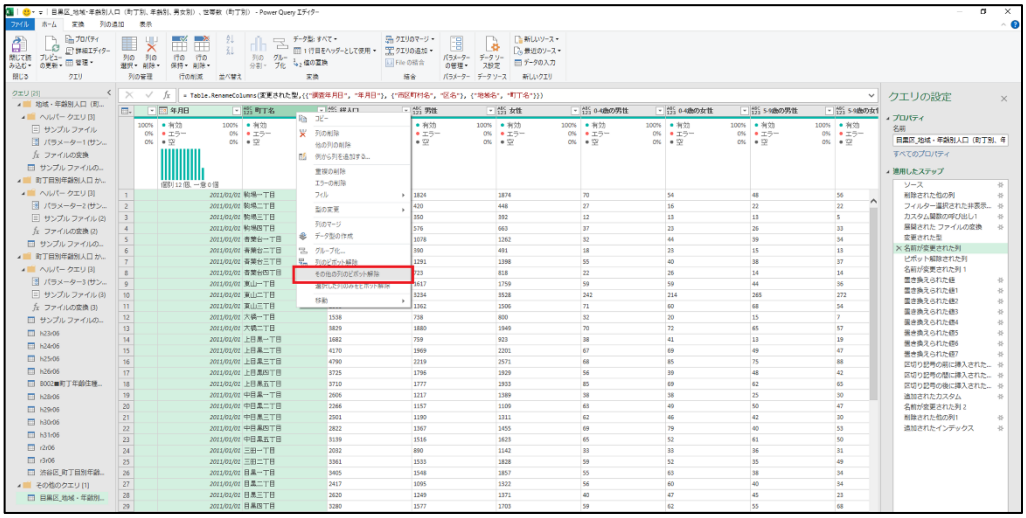
5. [OK]をクリックする

② Excel の Power Query 機能でデータを整形する

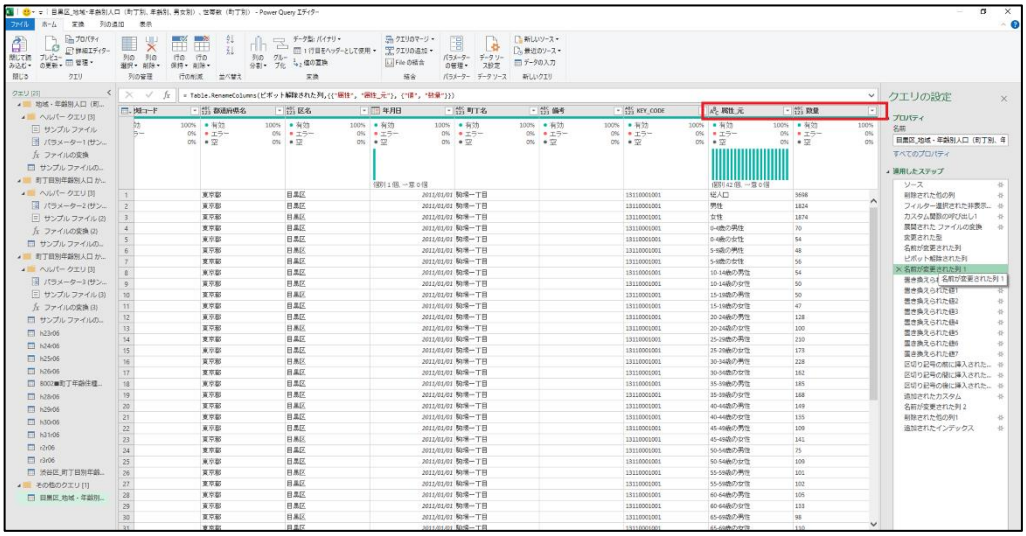
1. 「列名」をダブルクリックし、任意の列名を入力する



2. 対象列が選択された状態で右クリックし、[そのほかの列のピボット解除]をクリックする

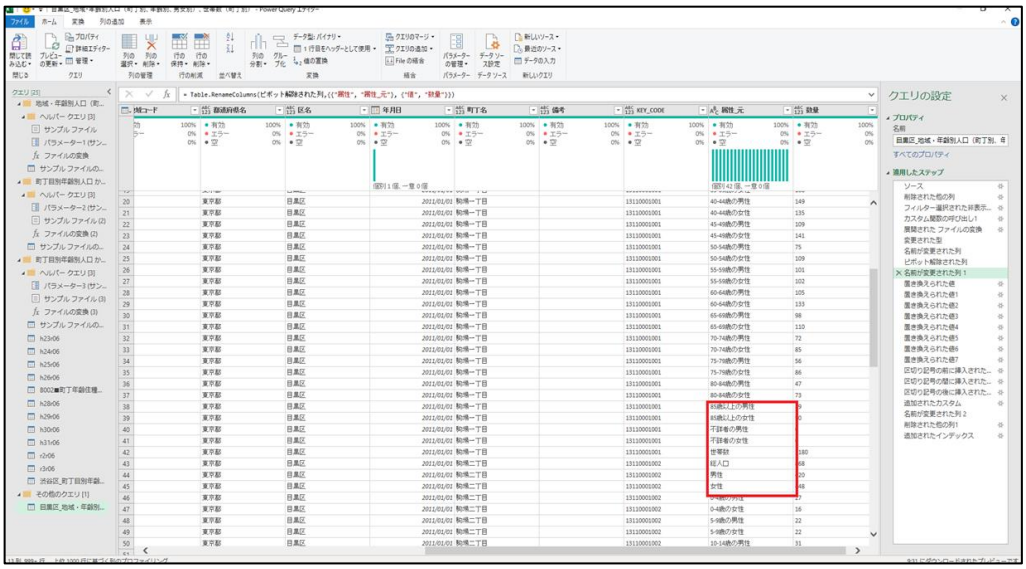


3. ピボット解除により作成されたヘッダー列名に、任意の列名を入力する



4. 対象列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

置換する値「000 歳-000 歳の〇〇〇形式となっていないデータ」
→ 置換後「000 歳-000 歳の〇〇〇」



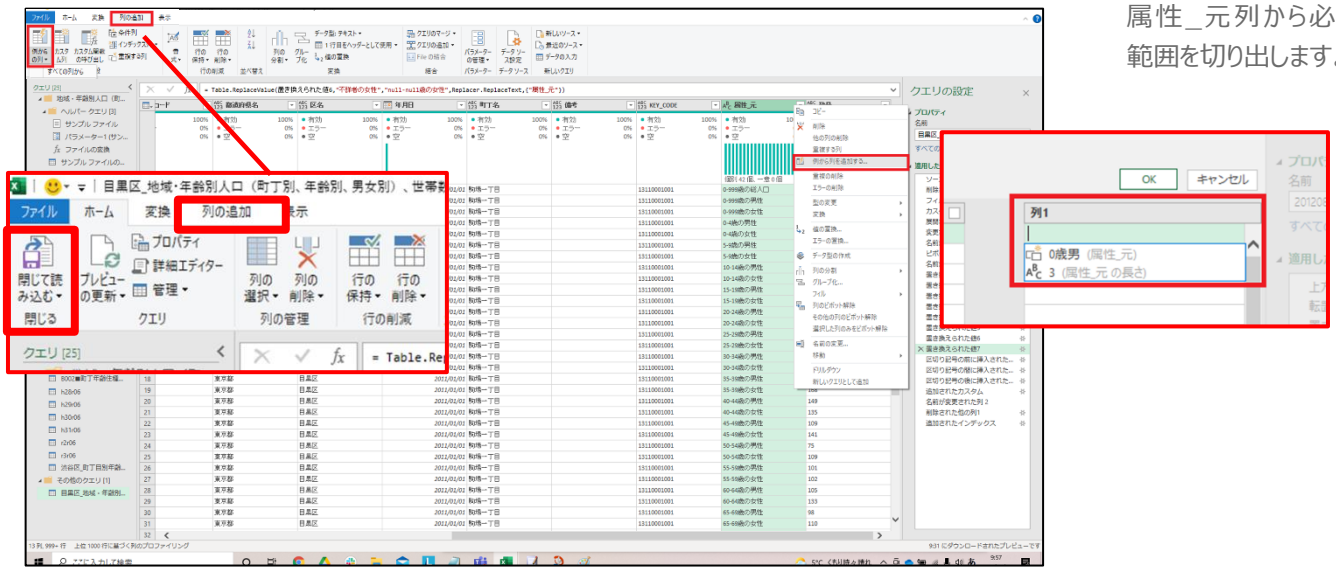
ここでは、属性_元列を年齢(開始)、年齢(終了)、属性に分割するため、入力フォーマットを揃えます。
属性_元列の値の置換を行います。

値の置換の際、"男性"、"女性"等、置換を行いたくない他のセル内容の一部と合致してしまう場合は、「セルの内容全体の照合」のチェックを ON にします。

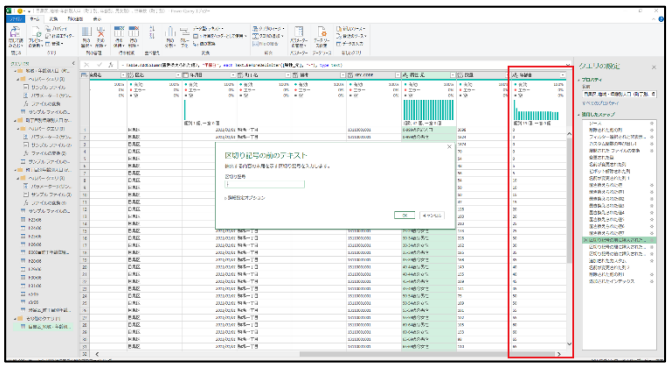
5. 対象列が選択された状態で右クリックし、[列の追加]→[例からの列]→[選択範囲から]をクリックし、例を作成する

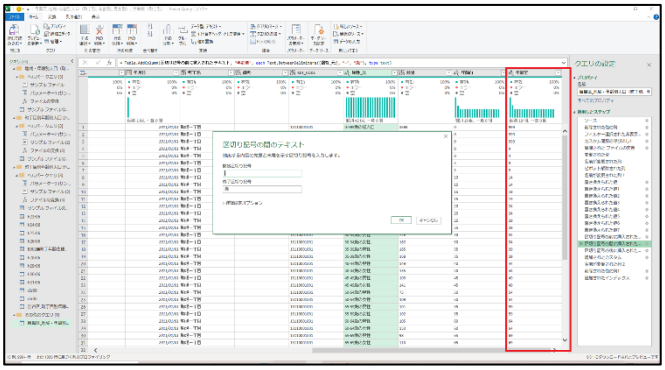
追加列のセルをダブルクリックし、選択セルのデータ内容をクリック、任意の値に変更して例を作成します。
作成後は、全データ想定通りの値が入力されている事を確認します。

ここでは、年齢自、年齢至、属性列追加属性_元列から必要範囲を切り出します。

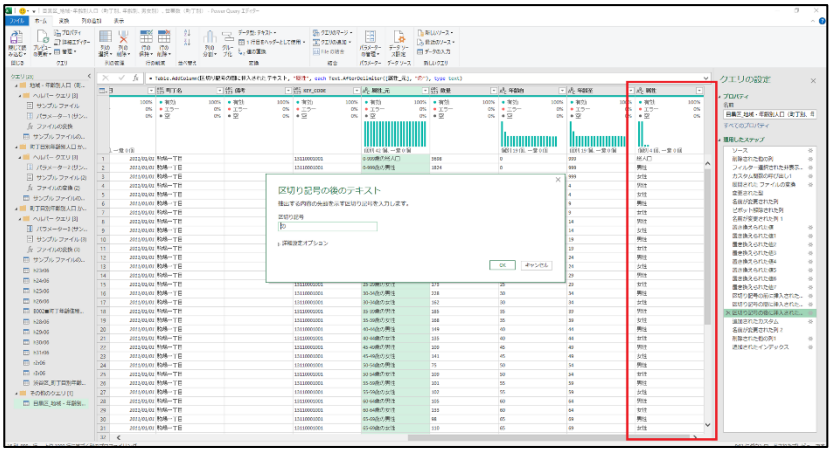


例からの列(年齢自列)追加後の画面（左）および例からの列(年齢至列)追加後（右）です。

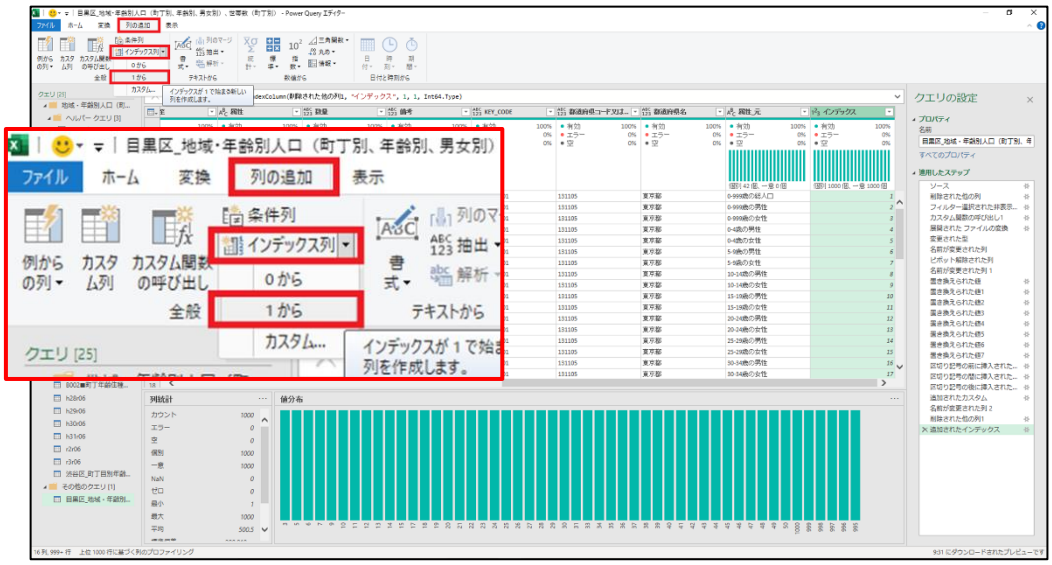




例からの列(年齢自列)追加後の画面です。

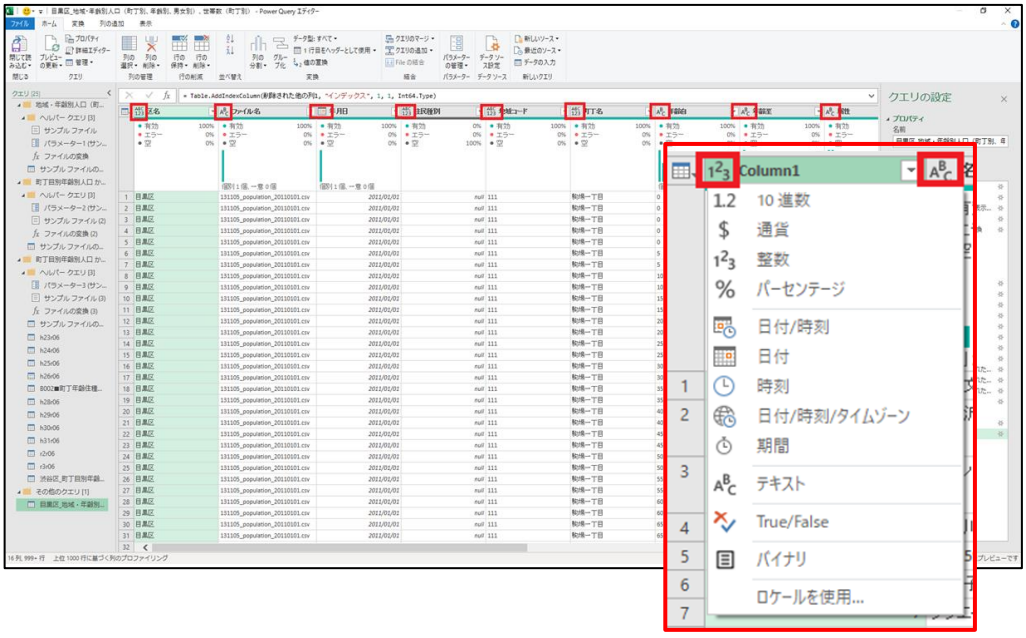


6. 追加する列が選択された状態で右クリックし、[列の追加]タブ→[インデックス列]→[1 から]をクリックする

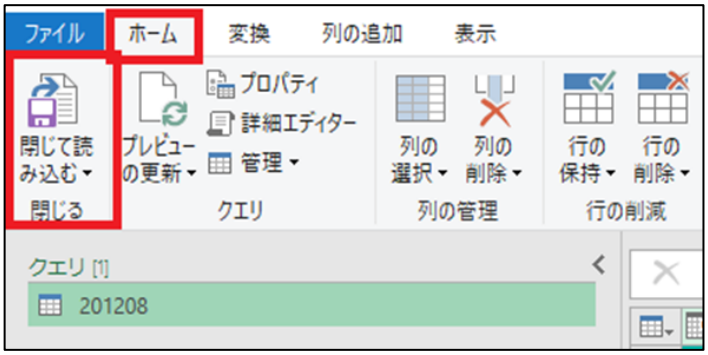


ここではインデックス列を追加しています。

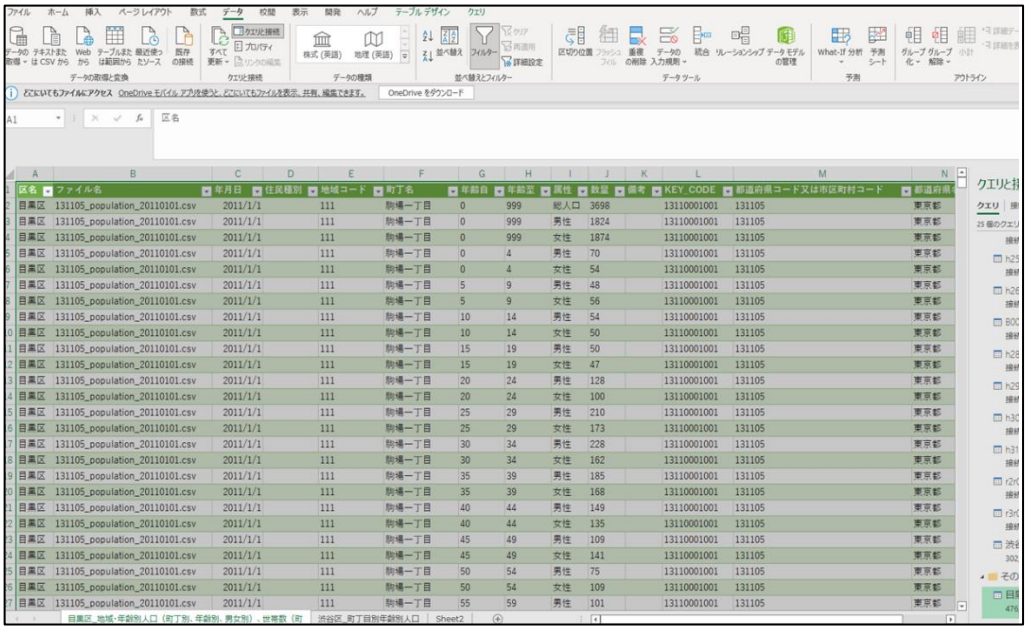
7. 各列のヘッダー上部にある「ABC123」アイコンをクリックし、任意の型を選択する



8. [ホーム]タブ→[閉じて読み込む]をクリックする



9. 読み込まれたデータを確認する



区名	ファイル名	年月日	住民種別	地域コード	町丁名	年齢自	年齢至	属性	数量	備考	KEY_CODE	都道府県コード又は市区町村コード	都道府県
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	0	999	総人口	3698		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	0	999	男性	1824		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	0	999	女性	1874		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	0	4	男性	70		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	0	4	女性	54		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	5	9	男性	48		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	5	9	女性	56		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	10	14	男性	54		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	10	14	女性	50		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	15	19	男性	50		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	15	19	女性	47		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	20	24	男性	128		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	20	24	女性	100		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	25	29	男性	210		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	25	29	女性	173		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	30	34	男性	228		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	30	34	女性	162		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	35	39	男性	185		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	35	39	女性	168		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	40	44	男性	149		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	40	44	女性	135		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	45	49	男性	109		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	45	49	女性	141		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	50	54	男性	75		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	50	54	女性	109		13110001001	131105	東京都
目黒区	131105_population_20110101.csv	2011/1/1		111	駒場一丁目	55	59	男性	101		13110001001	131105	東京都

10. ファイルの種類欄で「テキスト（タブ区切り）」を指定して CSV を保存する

手順（ケース[C]）

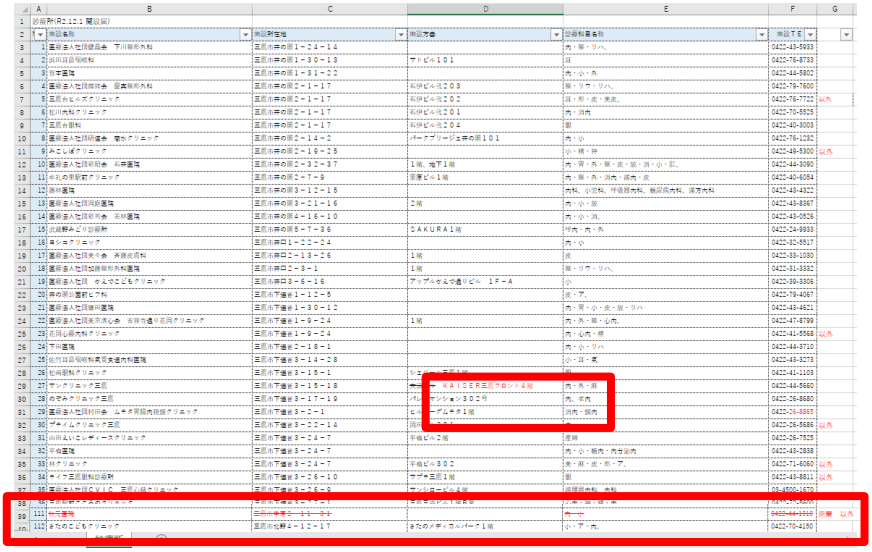
下記の手順で Excel から CSV にデータを変換します。

【お願い】： データ取得元（ソース）のファイルは閉じた状態で作業を行ってください。

① **Excel マクロにより、文字装飾等を削除する前処理を実施します。** **（詳細： p. 55）**
② **Excel で整形、加工したいファイルをインポートします。** **（詳細： p. 58）**
③ **Excel の Power Query 機能で、データを整形します。** **（詳細： p. 60）**
　クエリパターン 1 **（詳細： p. 92）**
　クエリパターン 2 **（詳細： p. 64）**
④ **Excel の Power Query 機能で、複数パターンが発生するクエリを結合します。** **（詳細： p. 71）**

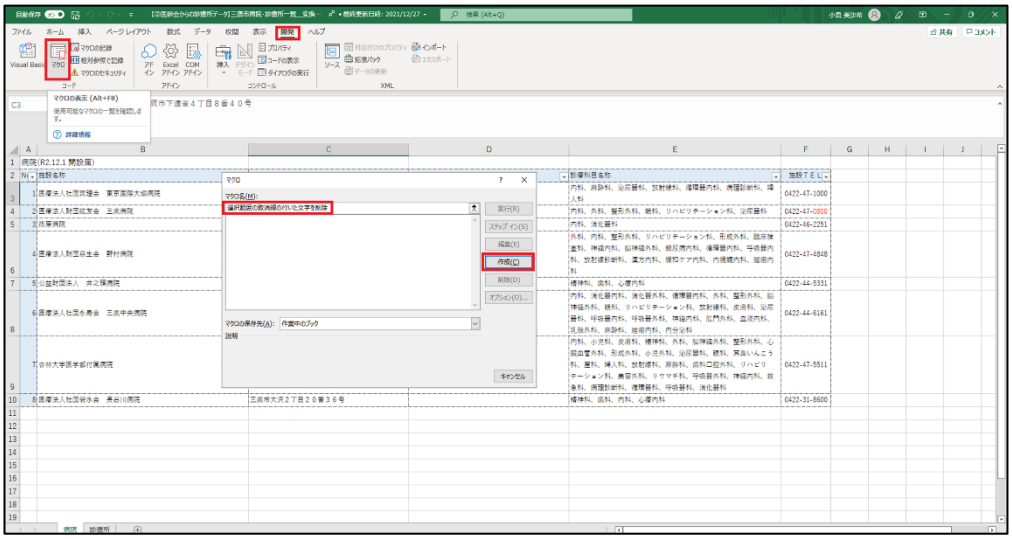
① Excel マクロにより、文字装飾等を削除する前処理の実施

1. 取消線や装飾文字のある箇所を確認する

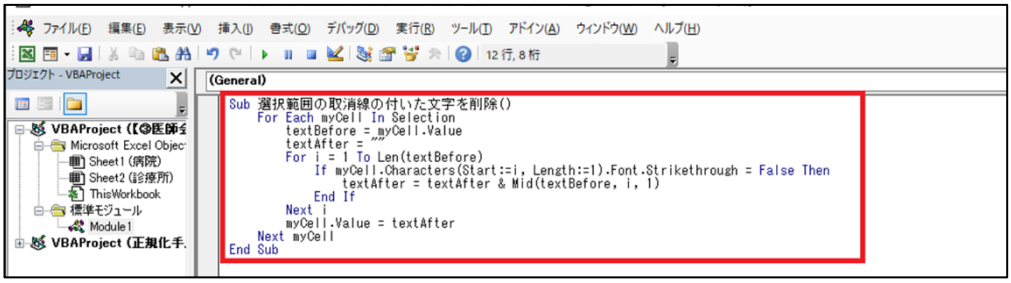


Excel に直接打ち込んだ、取り消し線等をマクロで削除します。

2.　[開発]→[マクロ]をクリックし、マクロ名を入力した後、[作成]をクリックする



3.　選択範囲を対象とした、取消線のついた文字の装飾を削除するマクロを入力する

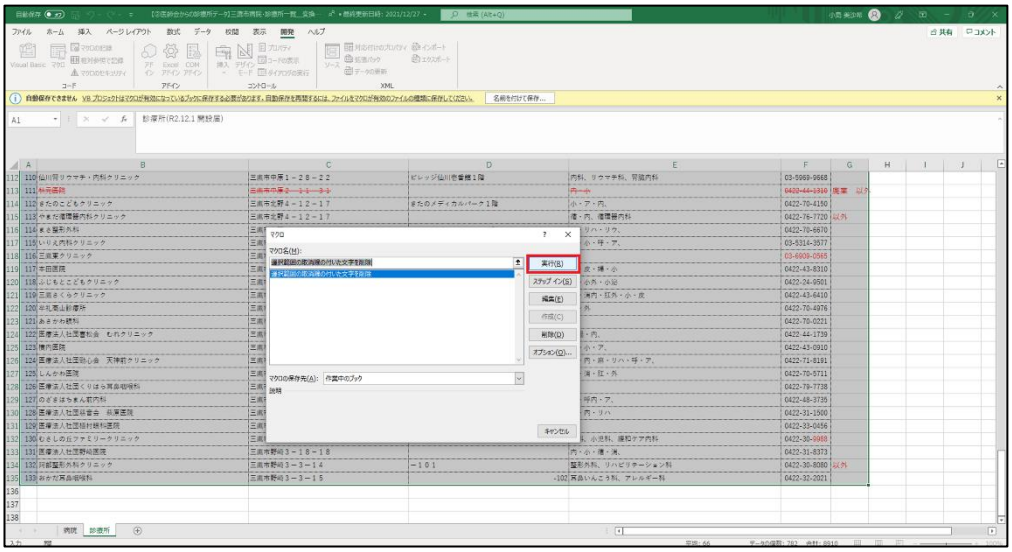


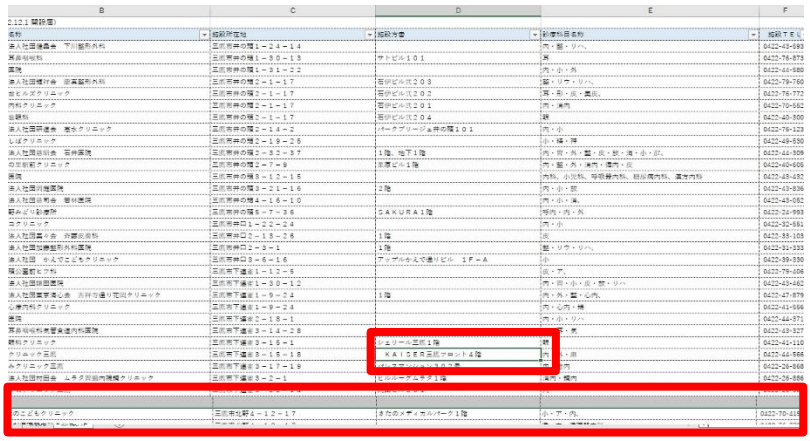
マクロの内容は下記です。

```
Sub 選択範囲の取消線の付いた文字を削除()
    For Each myCell In Selection
        textBefore = myCell.Value
        textAfter = ""
        For i = 1 To Len(textBefore)
            If myCell.Characters(Start:=i, Length:=1).Font.Strikethrough = False Then
                textAfter = textAfter & Mid(textBefore, i, 1)
            End If
        Next i
        myCell.Value = textAfter
    Next myCell
End Sub
```

4.　[ファイル]タブ→[終了して Microsoft Excel へ戻る]をクリックする

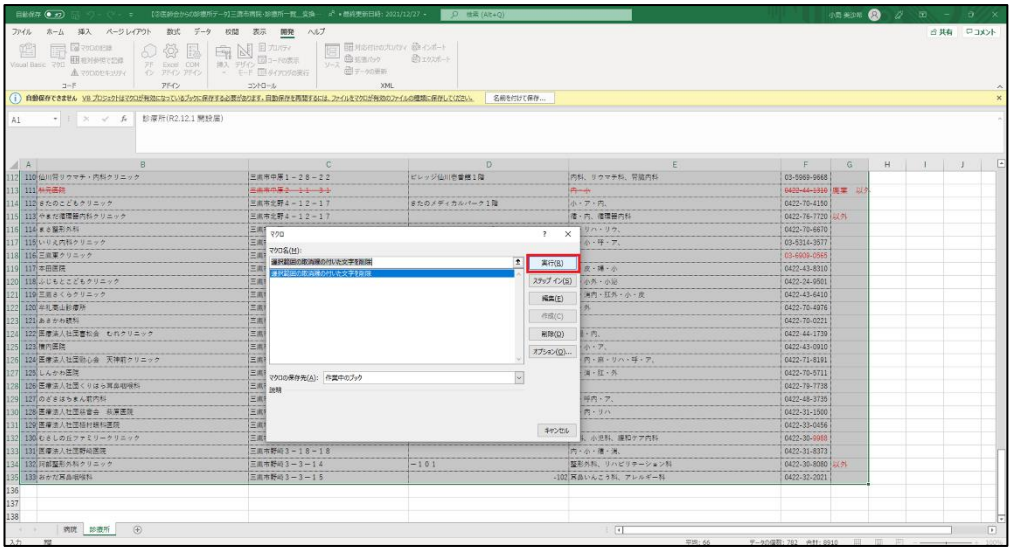
5. マクロを適用する範囲を選択し、「選択範囲の取り消し線の付いた文字を削除」マクロを選択し、[実行]をクリックする





マクロが実行され、選択範囲の文字装飾がなくなっていることを確認します。

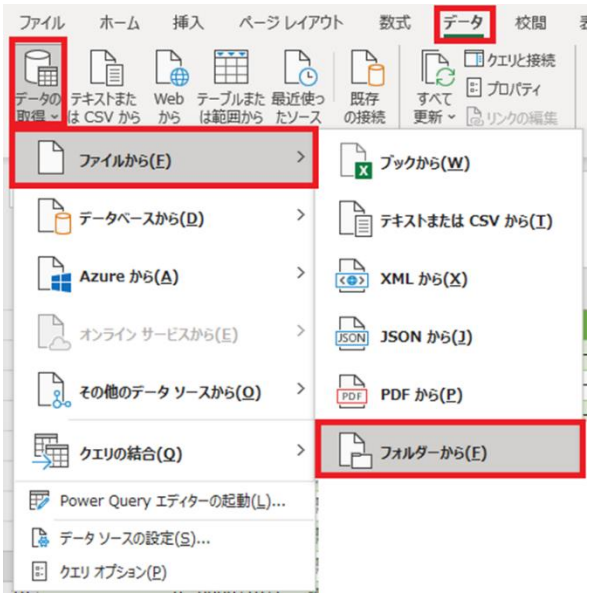
6. マクロを適用する範囲を選択し、「選択範囲の取り消し線の付いた文字を削除」マクロを選択し、[実行]をクリックする



② Excel で整形、加工したいファイルをインポートする

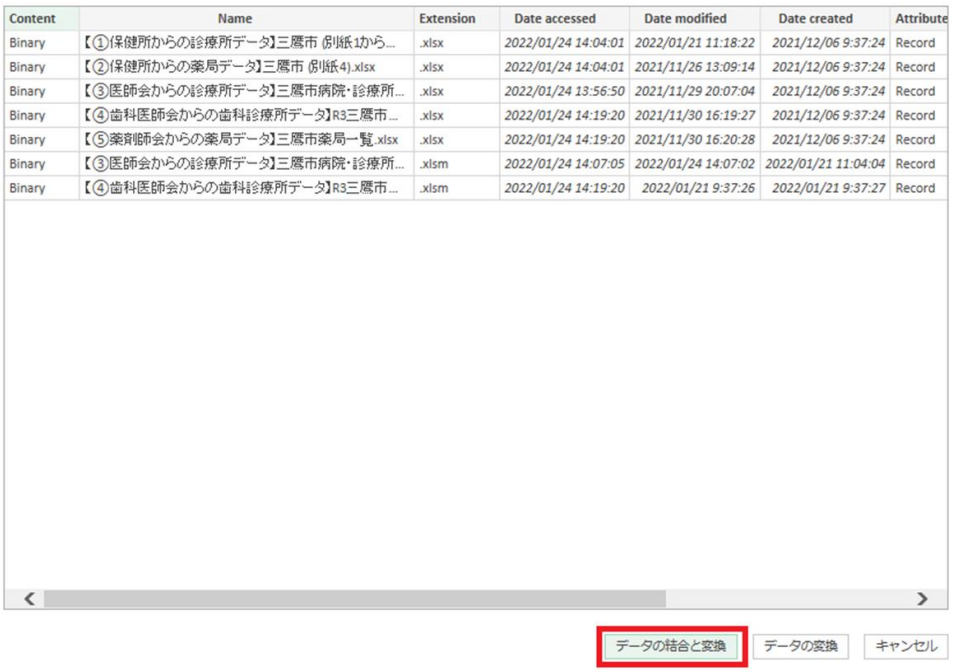
1. データ取得元（ソース）のファイルを開いている場合は、閉じておく

2. [データ]タブ→[データの取得]→「ファイルから」→「フォルダーから」をクリックする



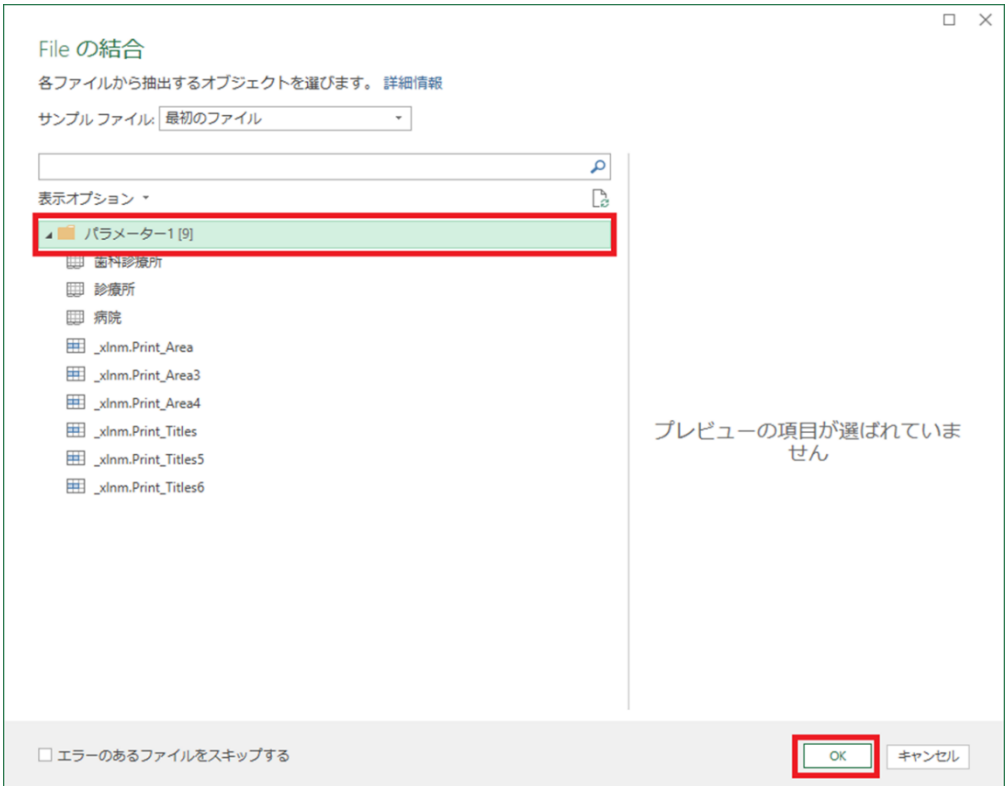
3. Excel ファイルをクリックして選び、[開く]をクリックする

4. [データの結合と変換]をクリックする



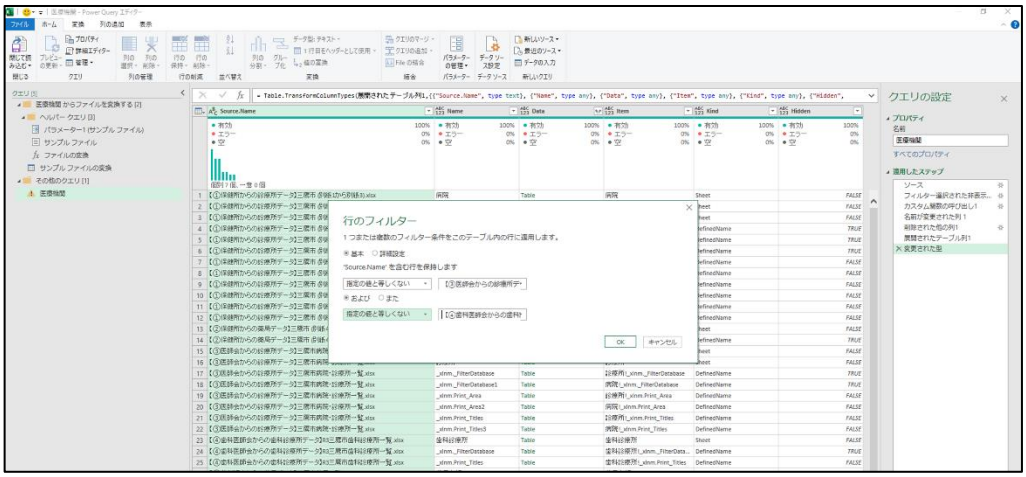
ここでの「アイテム」は、Excel のシートです。

5. フォルダをクリックして選び、[OK]をクリックする



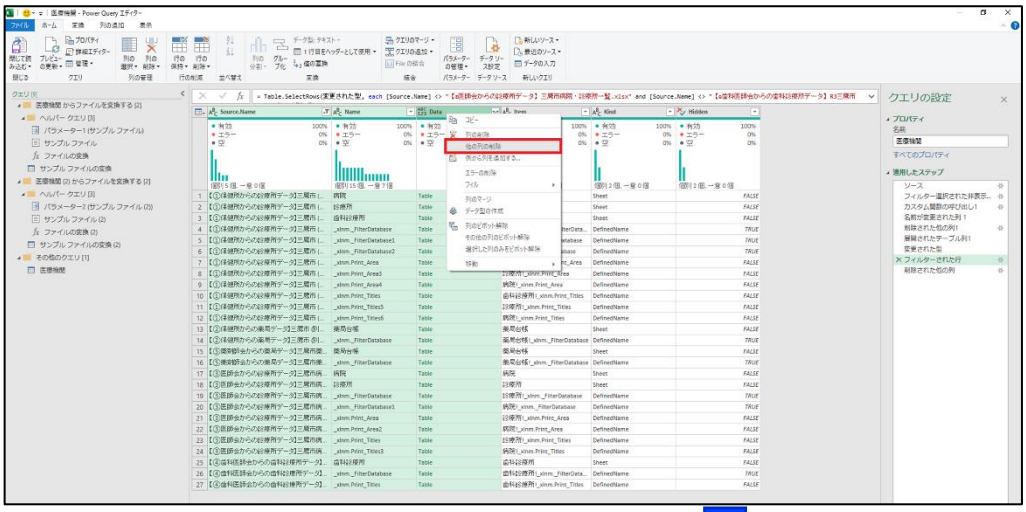
③ Excel の Power Query 機能でデータを整形する

1. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しくない」を選び、[OK]をクリックする

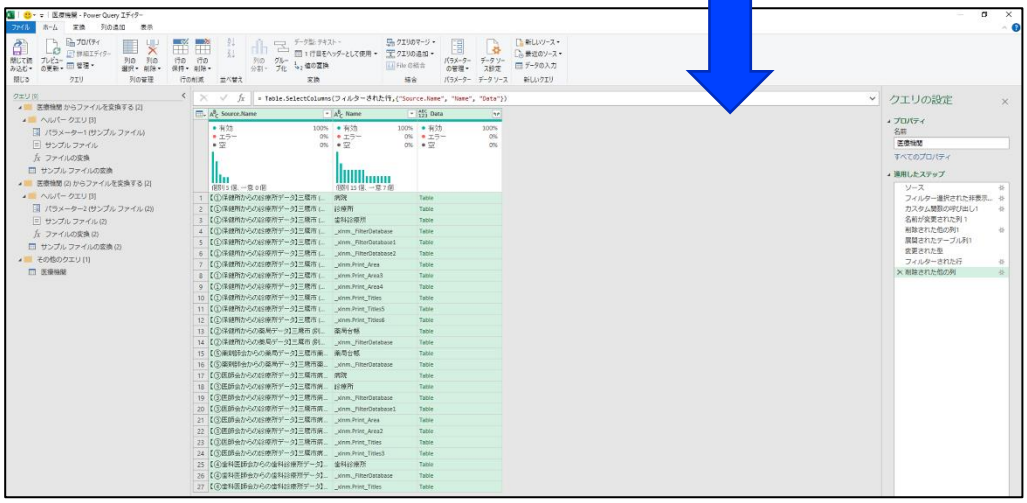


手順①でマクロ有(拡張子.xlsm)で保存したブックが存在する場合はマクロ無(拡張子.xlsx)を除外します。

2. 対象列を選択した後、右クリックし、「他の列の削除」をクリックする

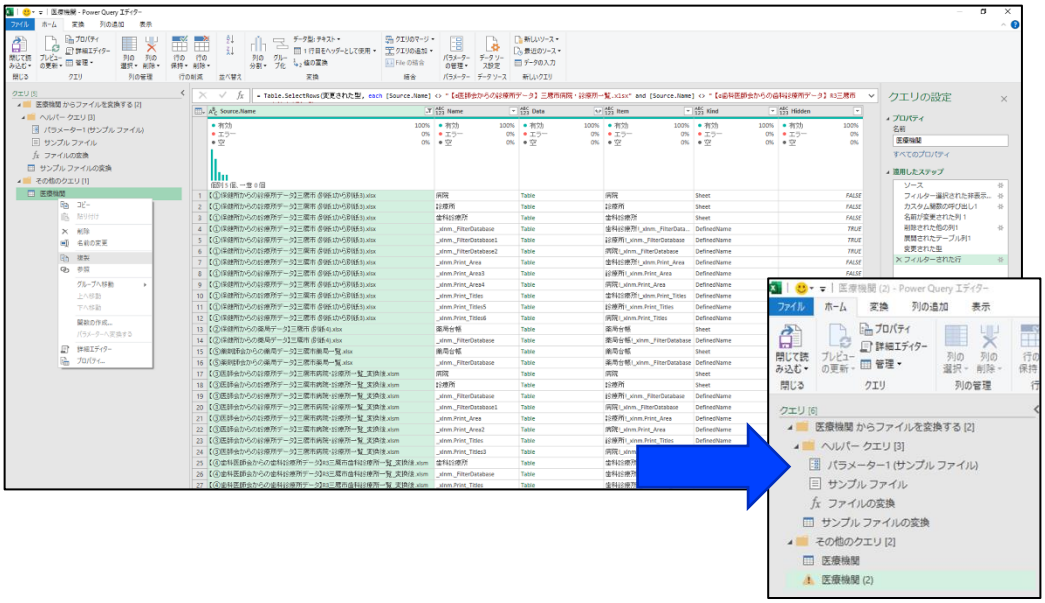


ここでは、対象列は、Source.Name　、Source、Data　例を指します。



対象列が残り、不要列を削除した状態です。

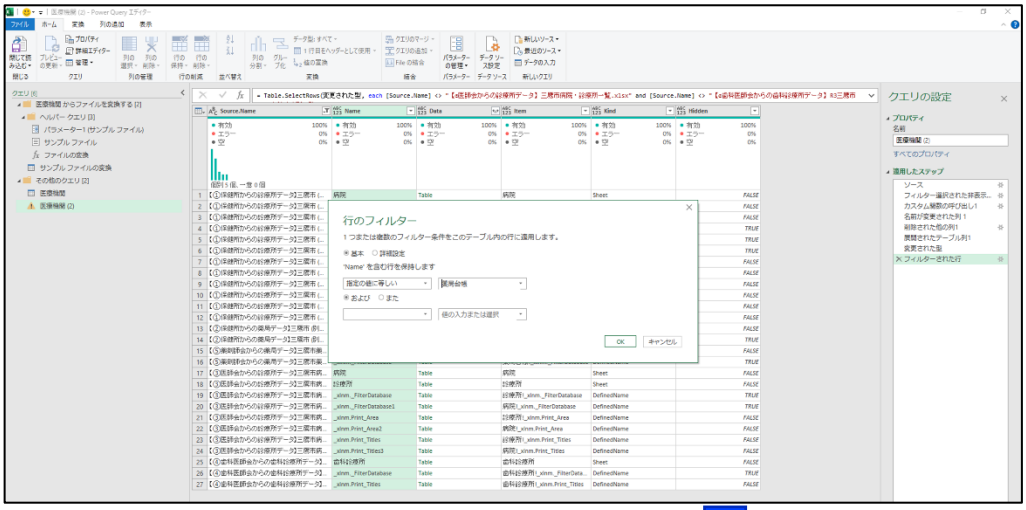
3. 該当するクエリを選択した後、右クリックし、「複製」をクリックする



ヘッダーパターン分のクエリを複製します。

<クエリパターン 1>

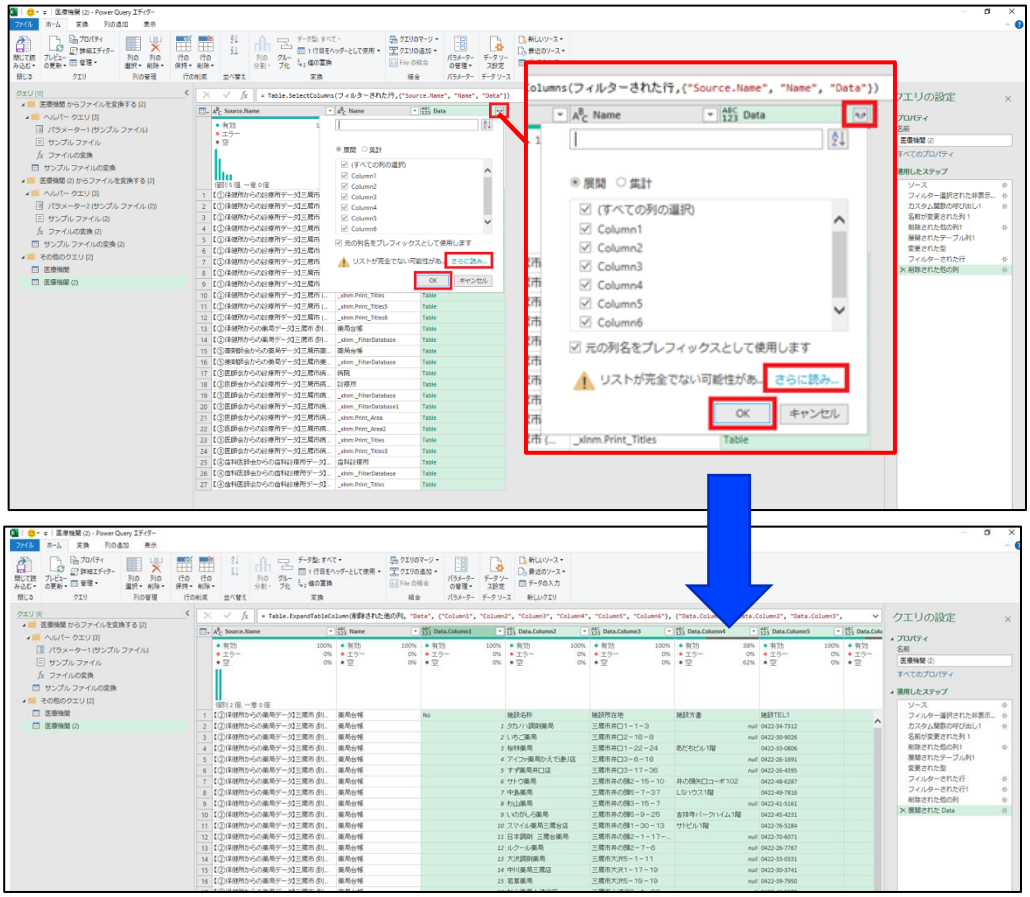
4. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しい」を選び、[OK]をクリックする



ここでは必要行を残すフィルターを設定しています。

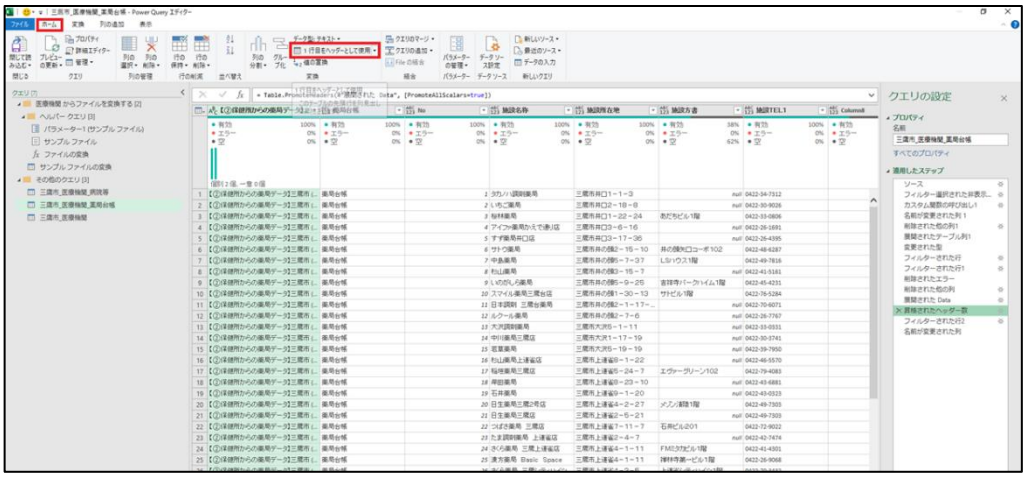
必要行が残った状態です。

5. 「Data」例　右上のアイコン→「さらに読み…」をクリックし、[OK]をクリックする



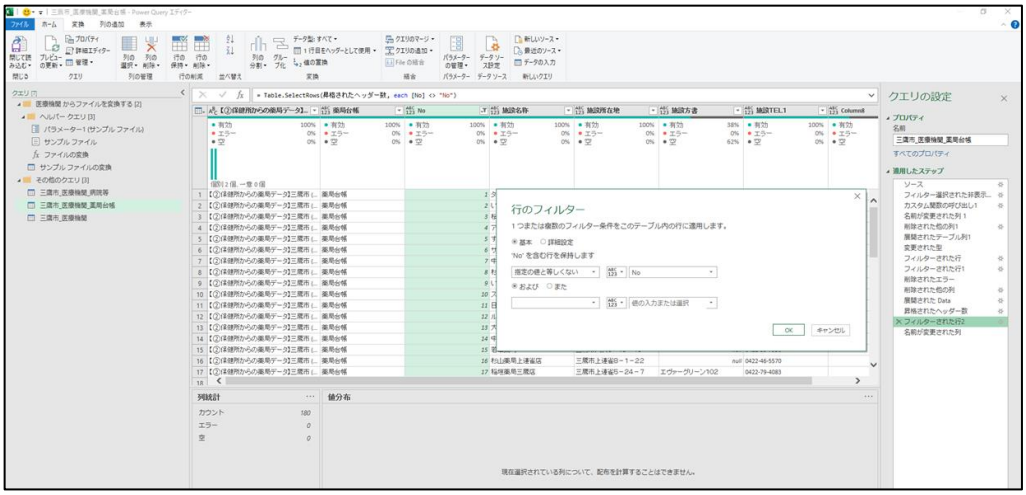
ここでは、Data 例を展開します。

6. [ホーム]タブ→「1 行目をヘッダーとして使用」をクリックする



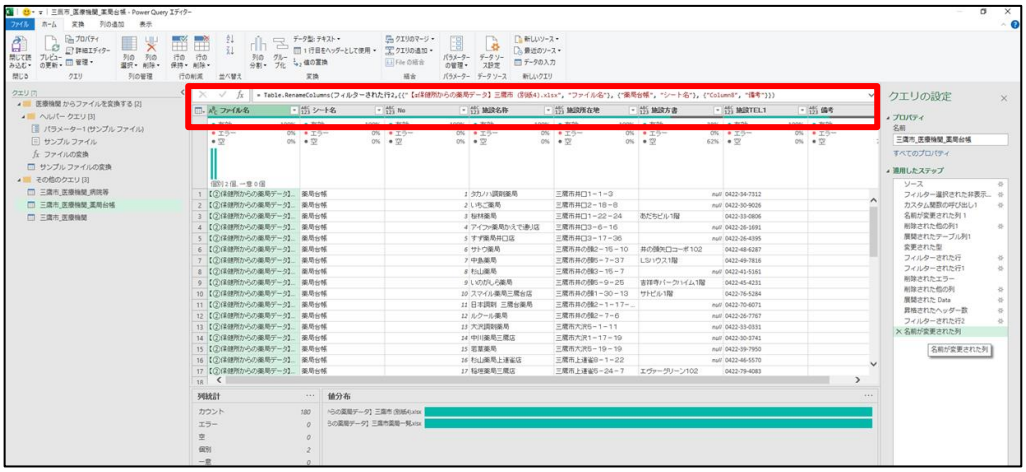
ここでは 1 行目をヘッダーとする整形を行っています。

7. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しくない」を選び、[OK]をクリックする

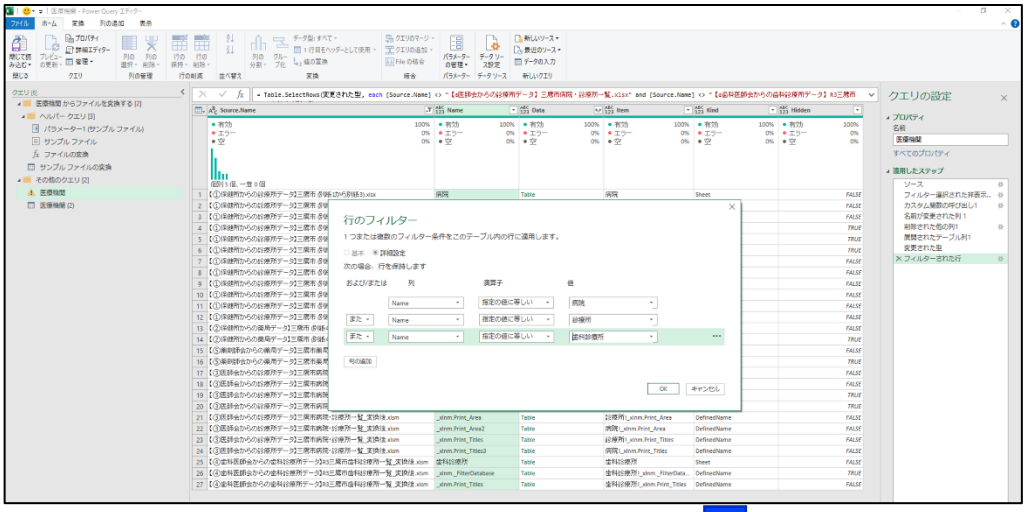


ここでは不要行を非表示する処理をしています。

8. 追加した「カスタム列」や「列名」の名前を指定しなかった列名があれば、列をダブルクリックして任意の列名を入力する



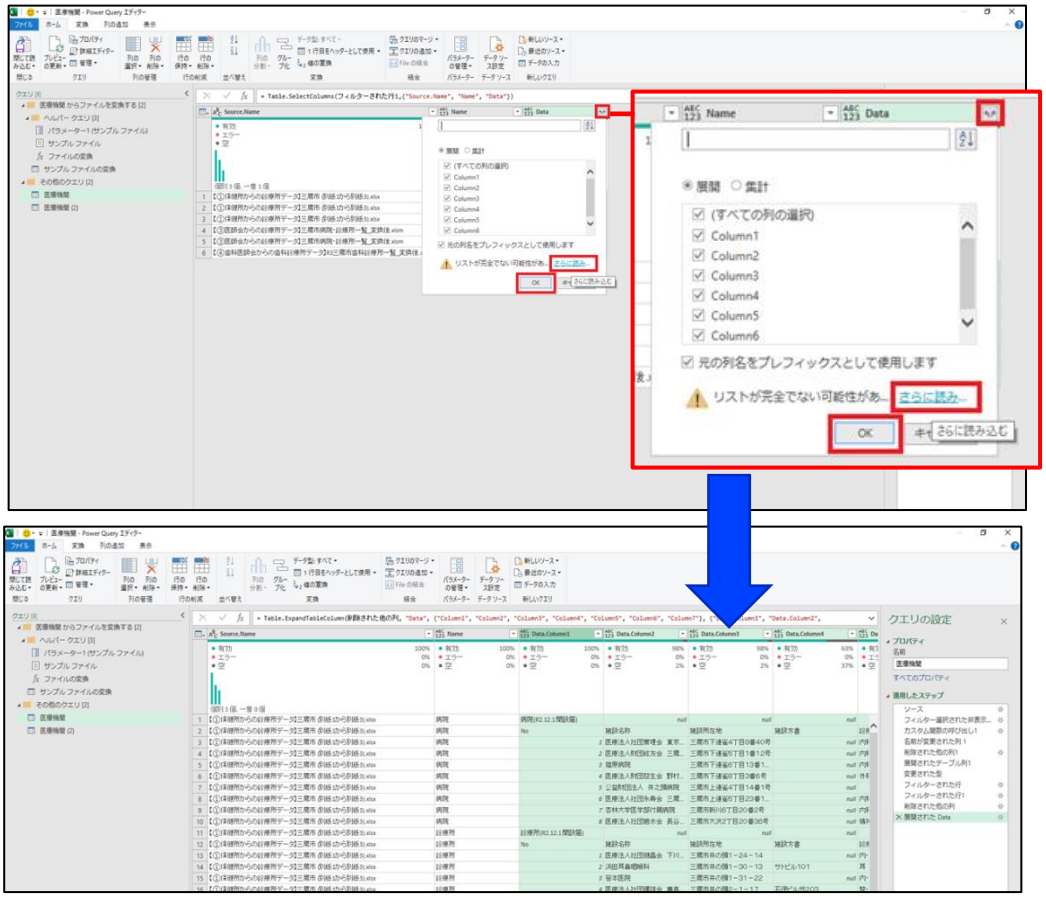
9. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しい」を選び、[OK] をクリックする



ここでは必要行を残すフィルターを設定しています。

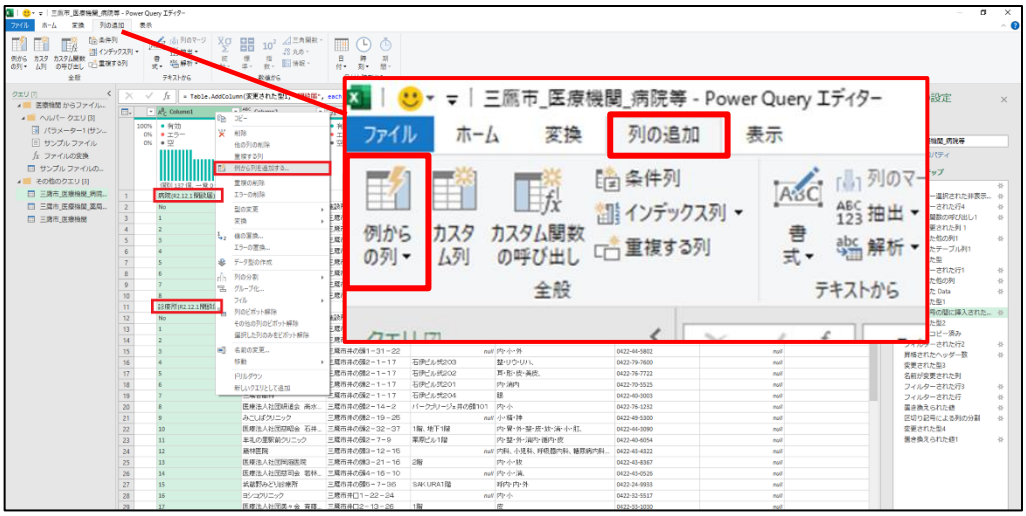
必要行が残った状態です。

10.「Data」例　右上のアイコン→「さらに読み…」をクリックし、[OK]をクリックする



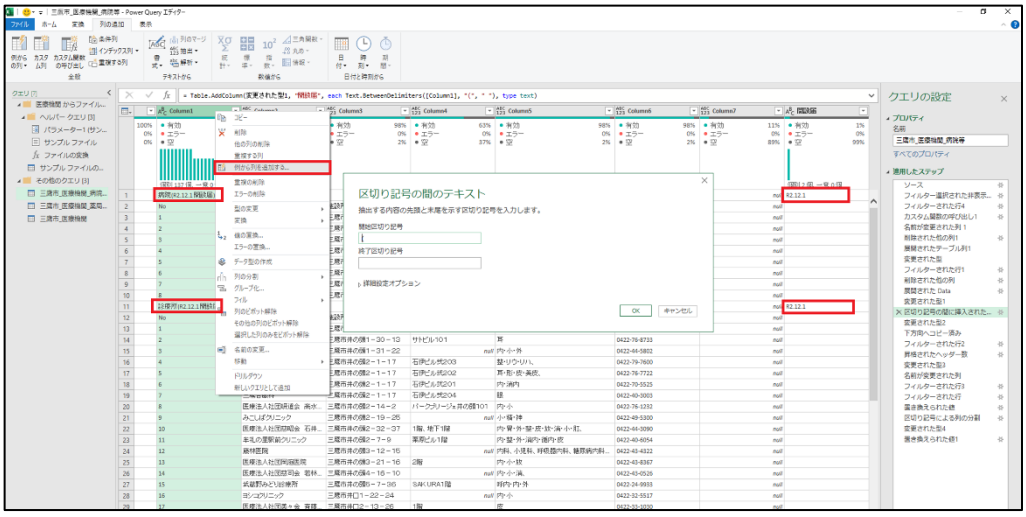
ここでは、Data 例を展開します。

11. 開設届の列が選択された状態で右クリックし、[列の追加]タブ→[例からの列]→[選択範囲から]をクリックする

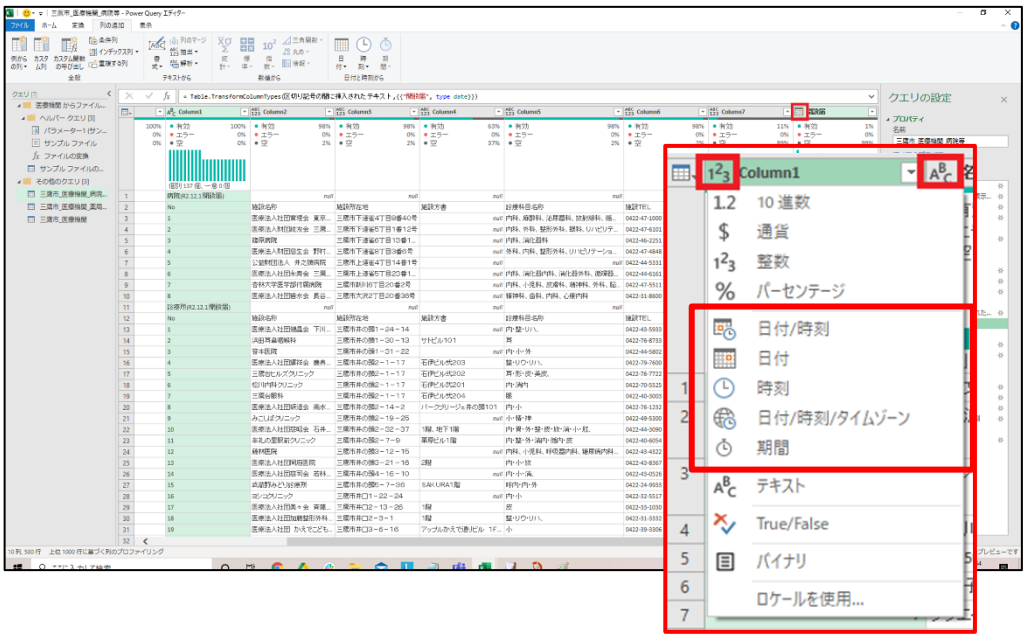


ここでは、No、開設届年月日が格納されている Column1 列から必要範囲を切り出し、開設届列を追加します。

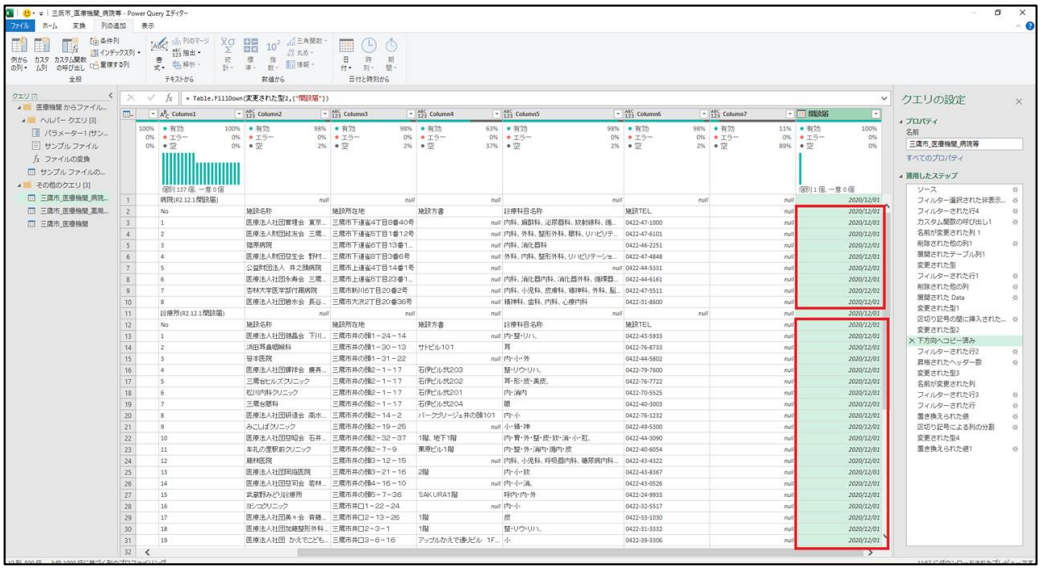
12. 追加列のセルをダブルクリック、選択セルのデータ内容をクリック、任意の値に変更して例を作成する



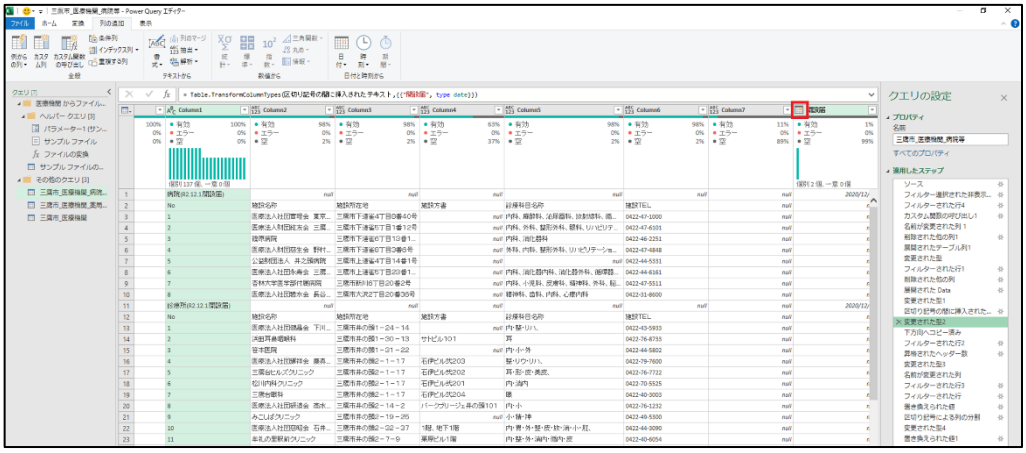
13. 開設届列のヘッダー上部にある「ABC123」アイコンをクリックし、日付の型を選択する



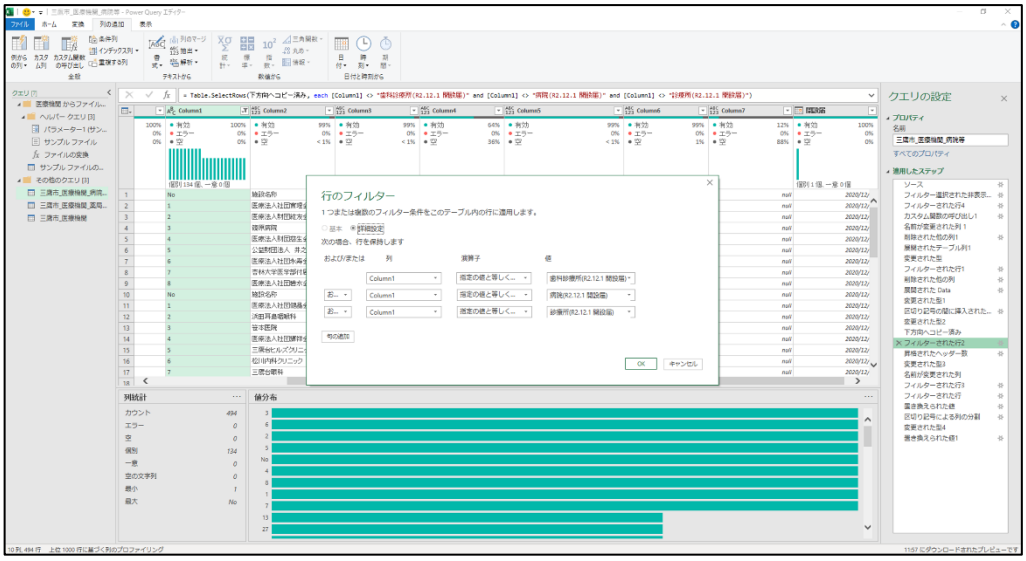
14. 開設届の列を右クリックし、[フィル]→[下]または[上]をクリックする



15. 開設届の列で、方を追加列のセルをダブルクリック、選択セルのデータ内容をクリック、任意の値に変更して例を作成する

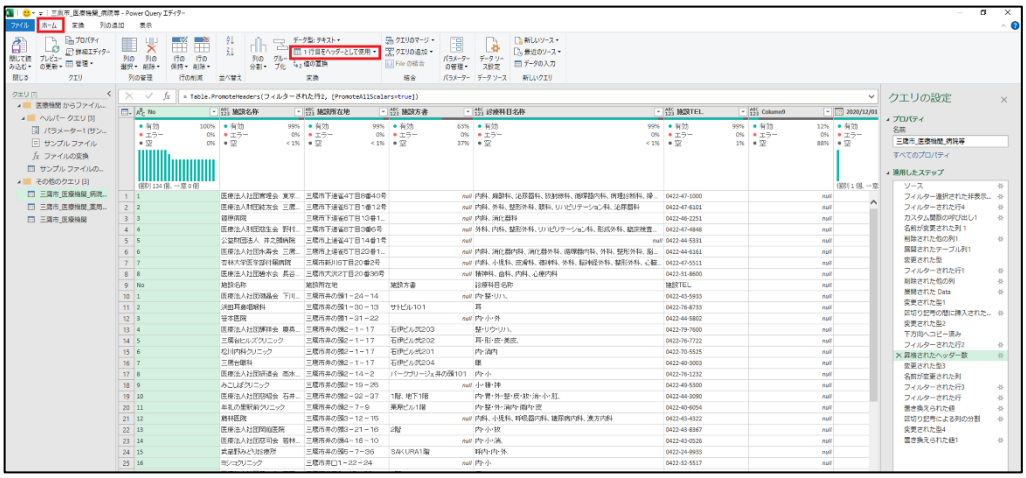


16. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しくない」を選び、[OK]をクリックする



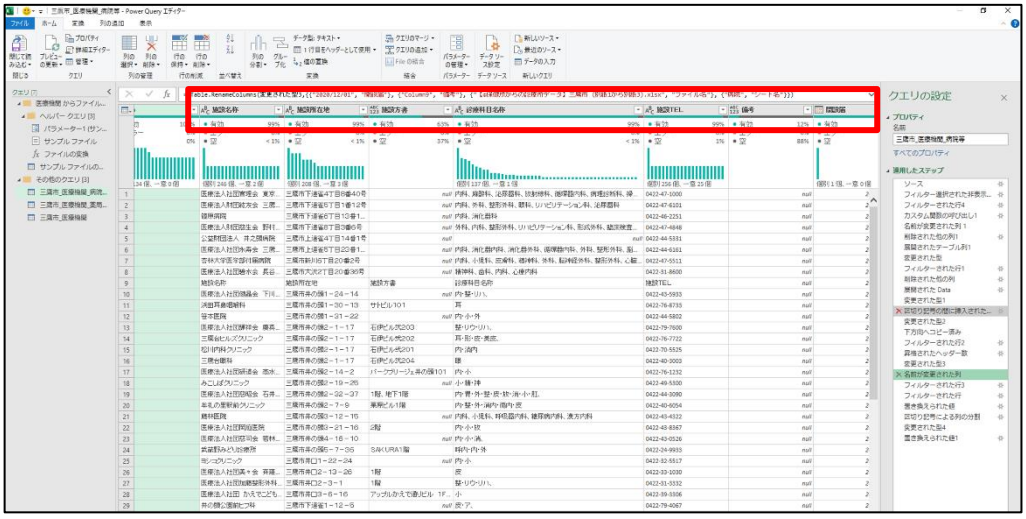
ここでは不要行を削除する処理をしています。
前の手順で別列へ分離した開設届年月日情報を含むデータ行は除外します。

17. [ホーム]タブ→「1 行目をヘッダーとして使用」をクリックする

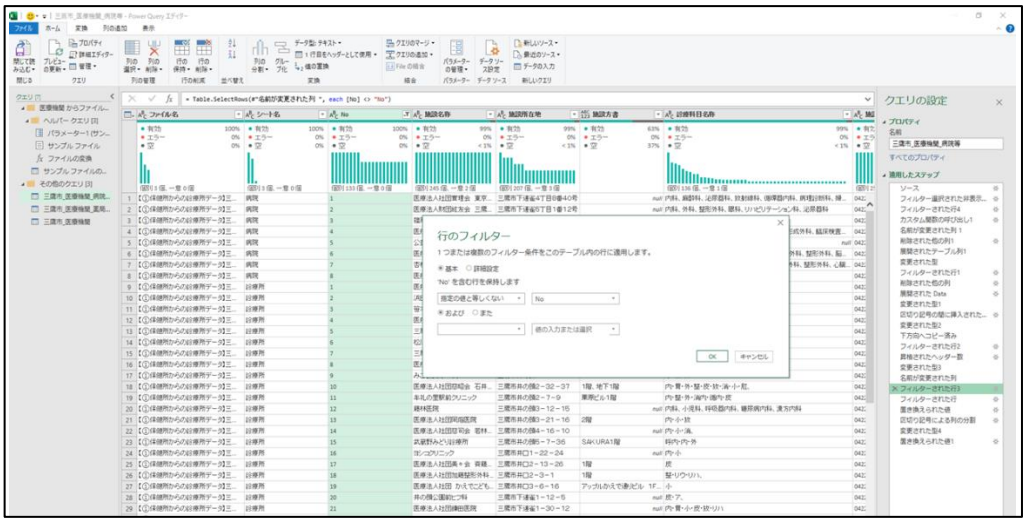


ここでは１行目をヘッダーとする整形を行っています。

18. 追加した「カスタム列」や「列名」の名前を指定しなかった列名があれば、任意の列名を入力する

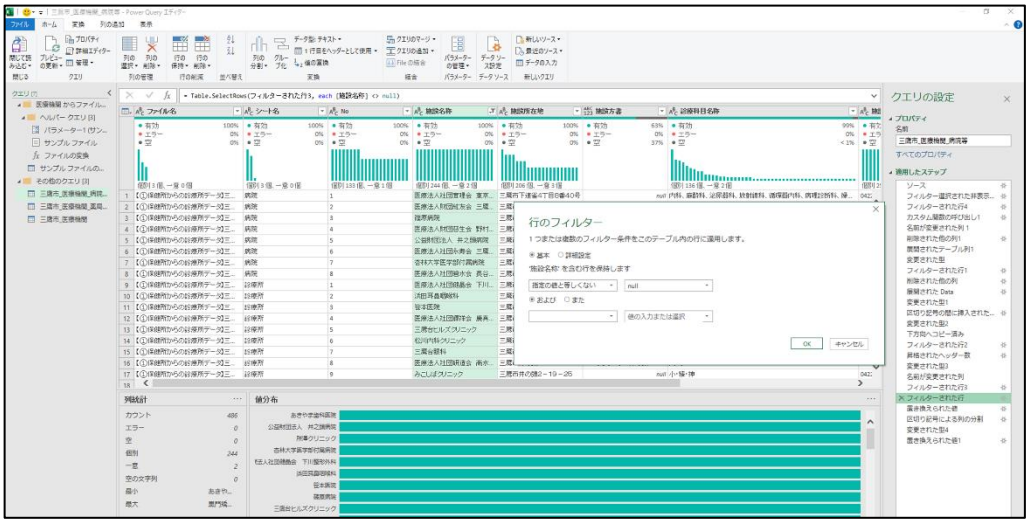


19. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しくない」を選び、[OK]をクリックする



ここでは不要行を削除する処理をしています。
ヘッダー行列名情報のデータ行は除外します。

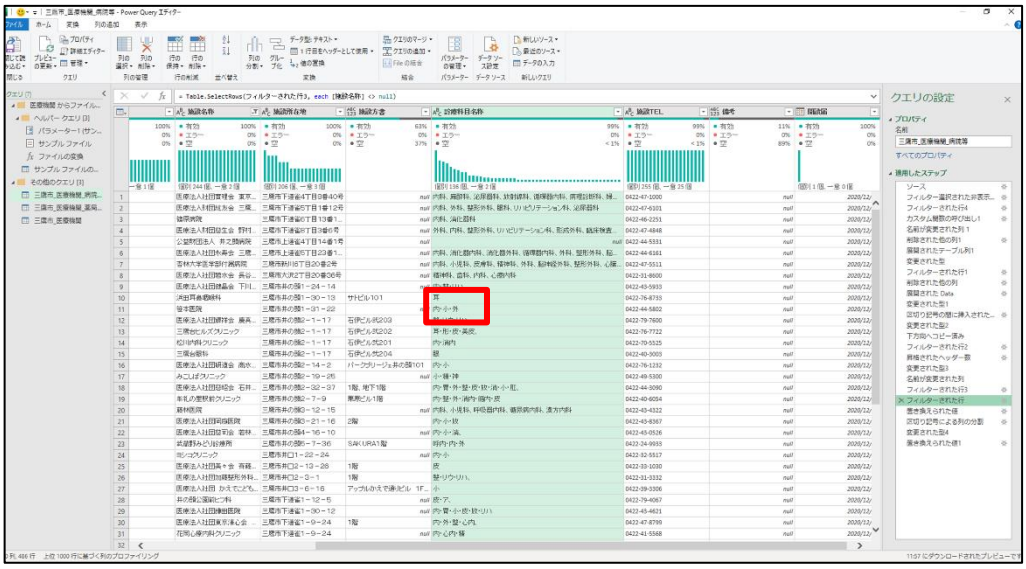
20. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しくない」を選び、[OK]をクリックする



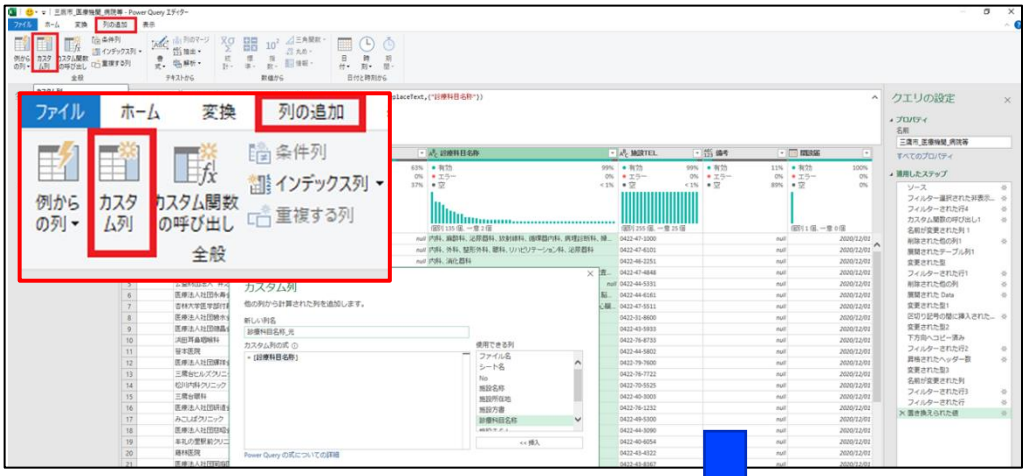
ここでは不要行を削除する処理をしています。
前の手順で、全て取消線で「null データ」となったデータ行は除外します。

21. 診療科目名称の列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

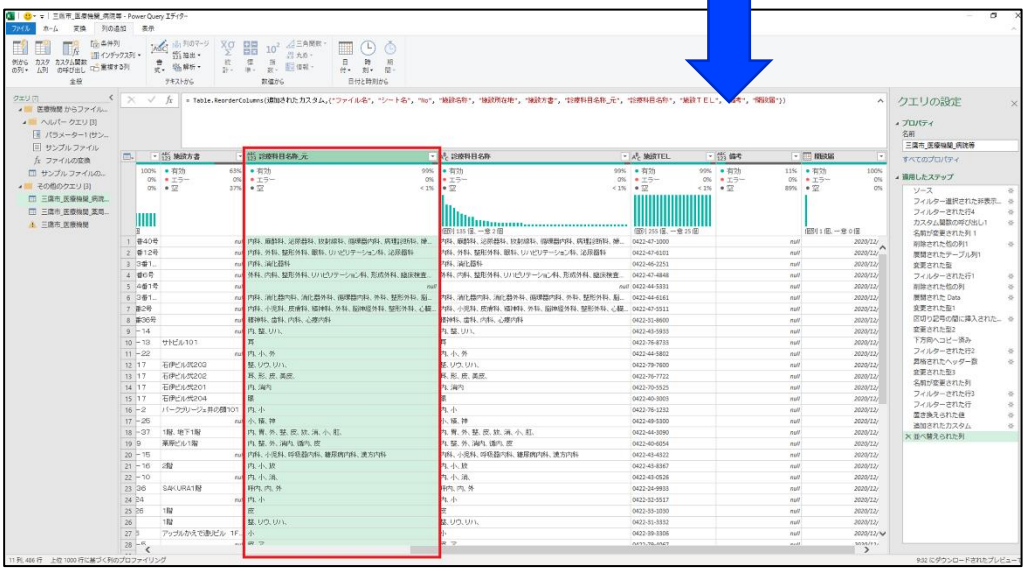
置換する値「“、”以外の区切文字」 → 置換後「“、”」



22. [列の追加]タブ→[カスタム列]をクリックする



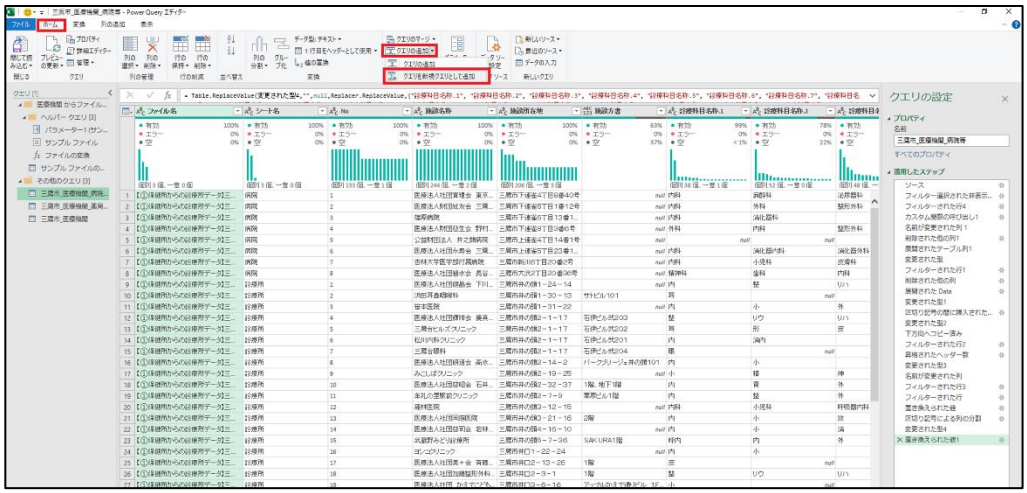
今回は元データ保持のため、診療科目名称をコピーした診療科目名称_元カスタム列を作成します。



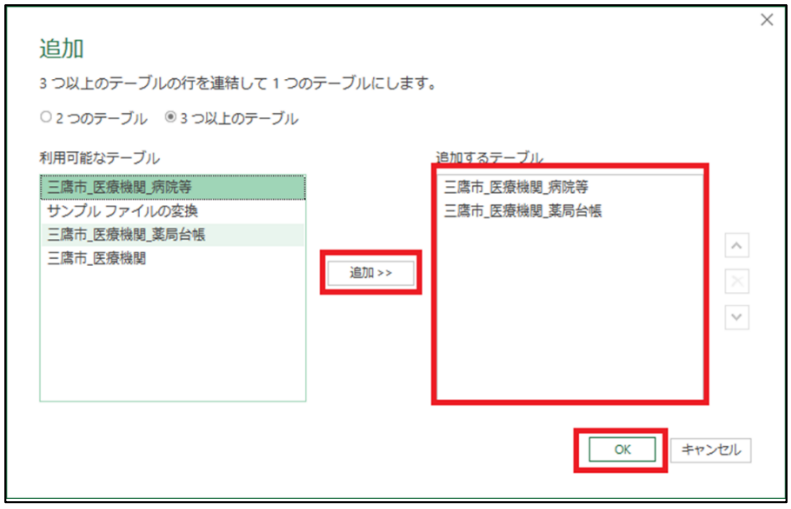
カスタム列を追加後の状態です。

④ Excel の Power Query 機能で、複数パターンが発生するクエリを結合する

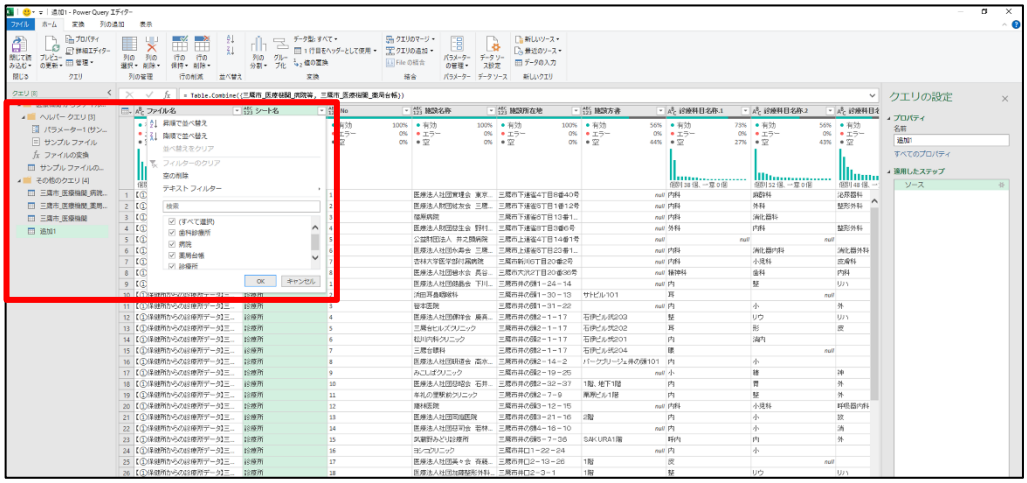
1. [ホーム]タブ→「クエリの追加」→「クエリを新規クエリとして追加」とクリックする



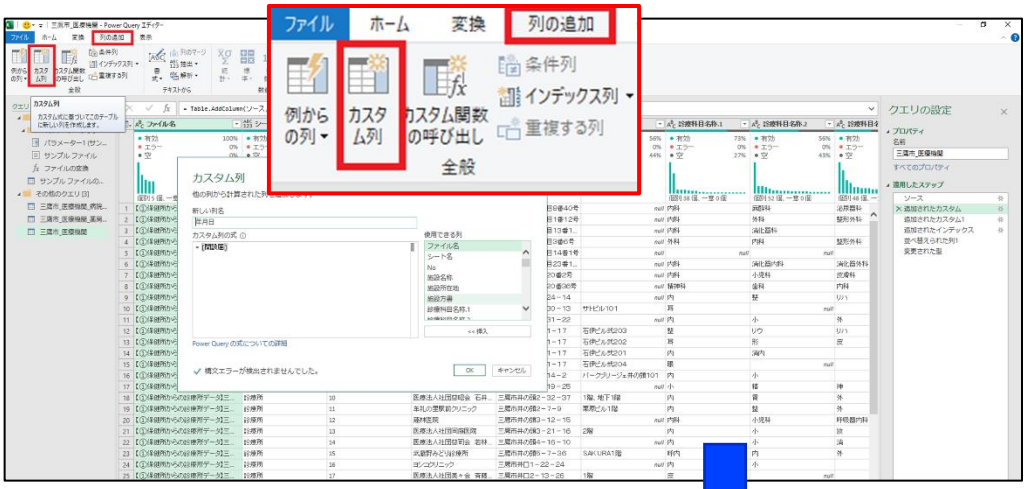
1. ①②③で作成したクエリを、利用可能なテーブルより選択し、[追加]→[OK]をクリックする



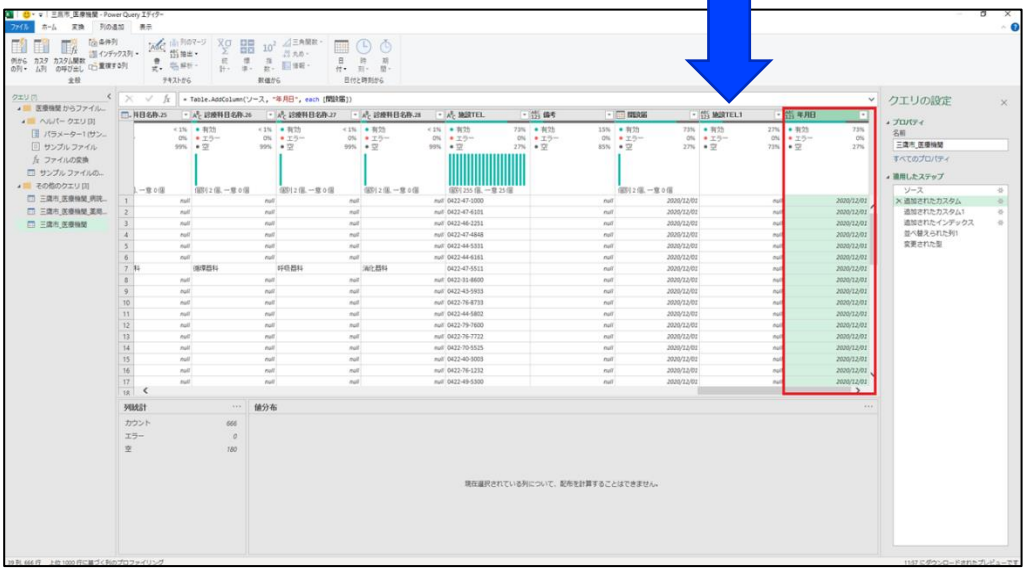
2. 選択した複数クエリのデータが行結合されていることを確認する



3. [列の追加]タブ→[カスタム列]をクリックする

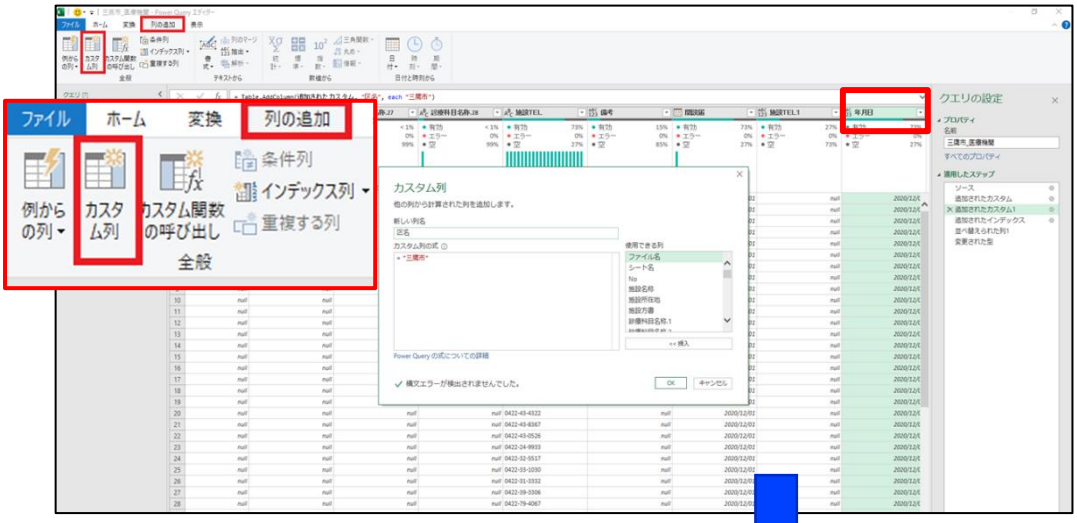


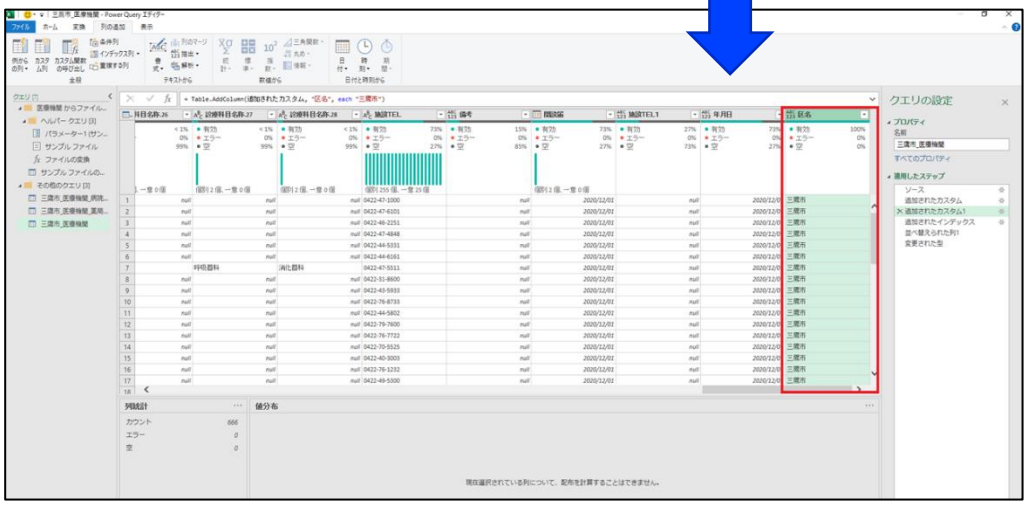
今回はカスタム列の式に開設届列を作成します。



年月日カスタム列を追加後の状態です。

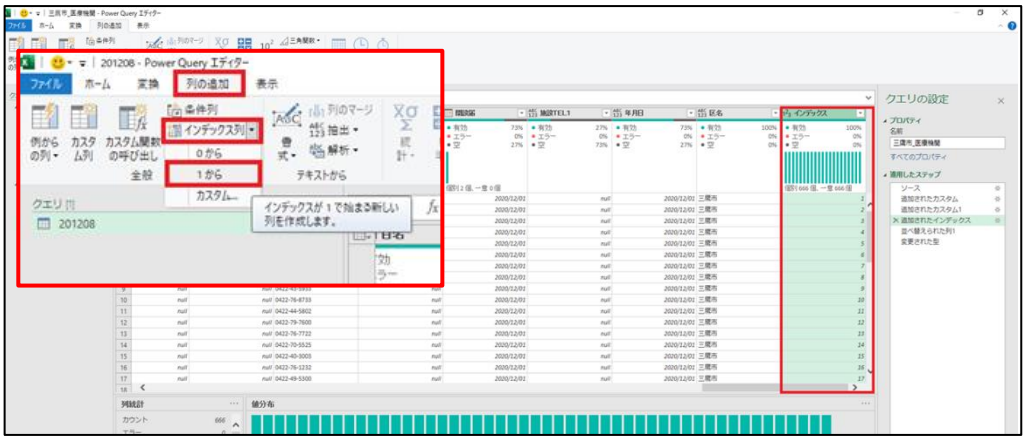
4. 追加した「カスタム列」に、対象の区名を入力する





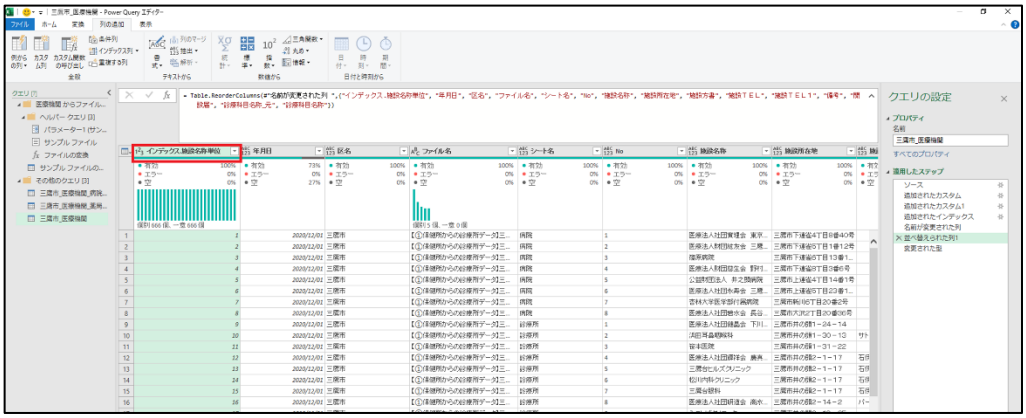
区名カスタム列を追加後の状態です。

5. 追加する列が選択された状態で右クリックし、[列の追加]タブ→[インデックス列]→[1 から]をクリックする



ここではインデックス列を追加しています。

6. 追加した列をダブルクリックし、任意の列名を入力する

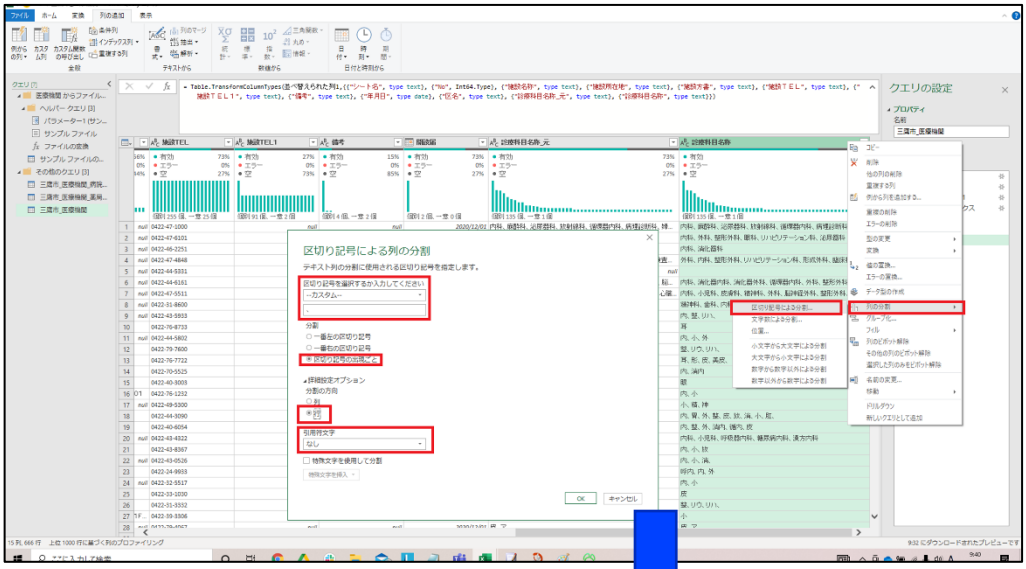


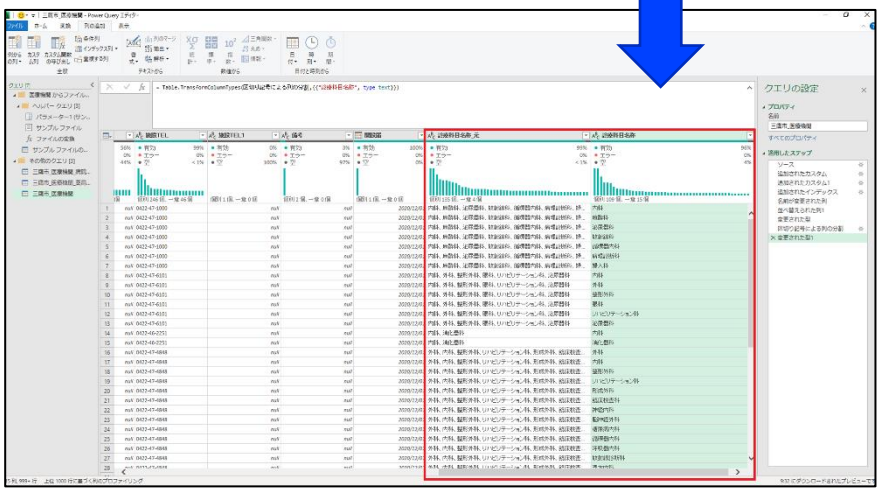
7. 対象列が選択された状態で右クリックし、「列の分割」→「区切記号による分割」をクリックする

「区切り記号による列の分割」では、下記の変更を行った後、[OK]をクリックしてください。

- 区切り記号：「カスタム」→「"、"」
- 分割：「区切り記号の出現ごと」
- 分割の方向：「行」
- 引用符文字：「なし」

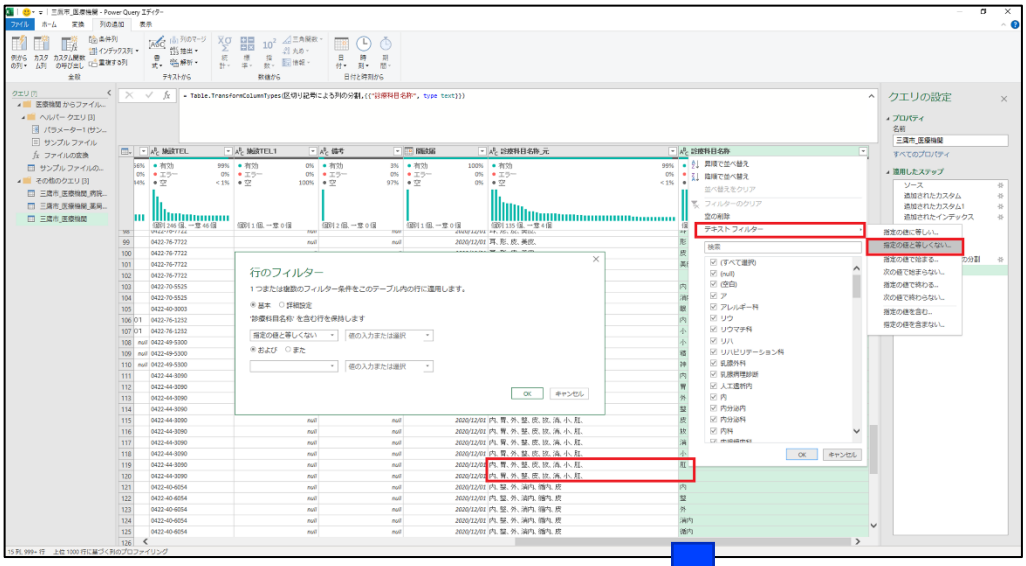
ここでは、同一列に複数列記されているデータで列を分割します。



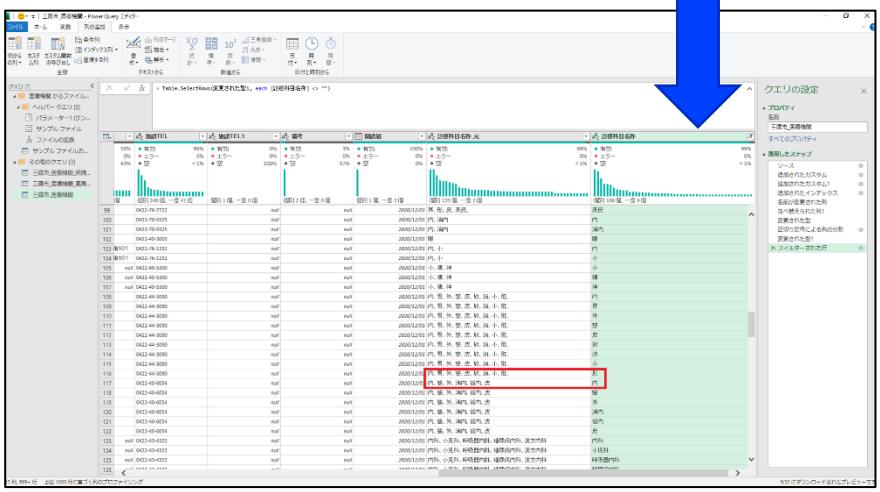


列が分割されます。

8. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しくない」を選び、[OK]をクリックする

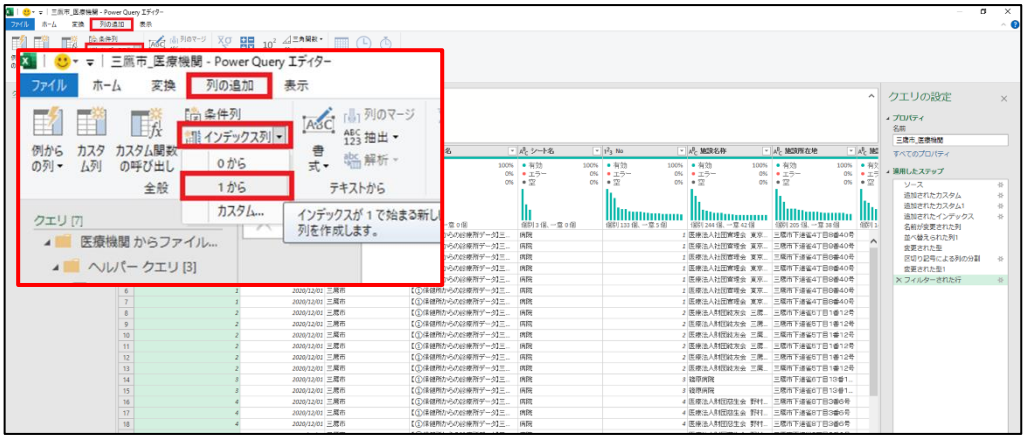


ここでは、列の分割前データの末尾に付与されていた区切り文字による空白データ(「皮」「ア」等も含む)を削除します。



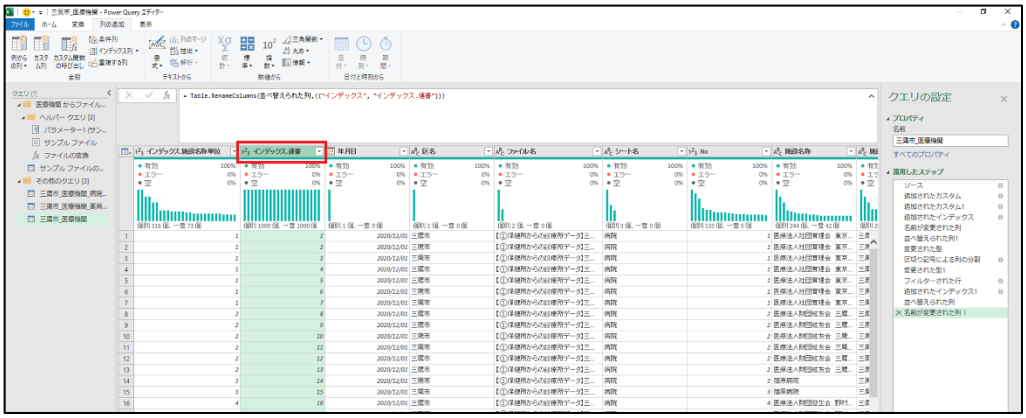
不要行（空白行）が削除されます。

9. 追加する列が選択された状態で右クリックし、[列の追加]タブ→[インデックス列]→[1 から]をクリックする

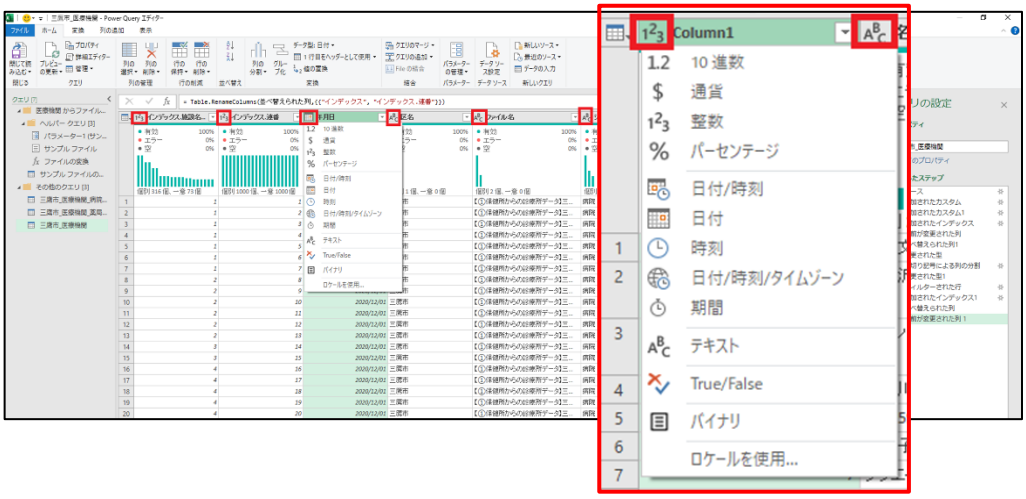


ここではインデックス列を追加しています。

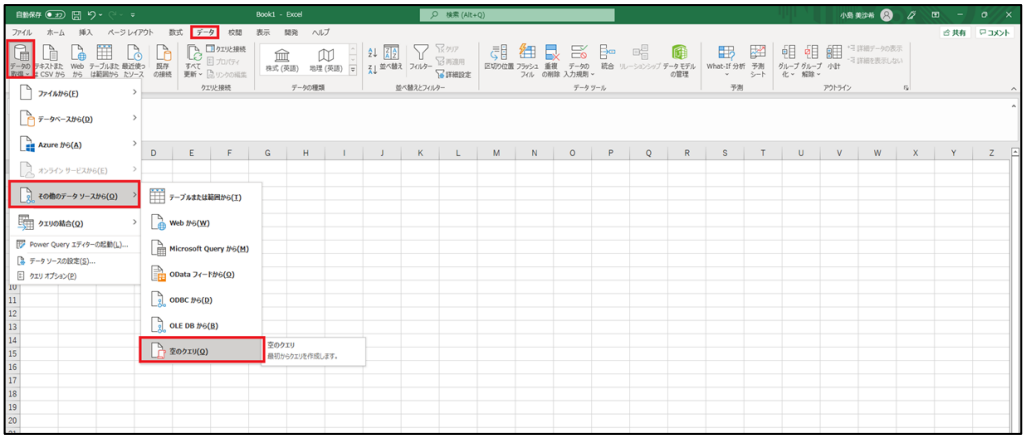
10. 追加した列をダブルクリックし、任意の列名を入力する



11. 各列のヘッダー上部にある「ABC123」アイコンをクリックし、任意の型を選択する

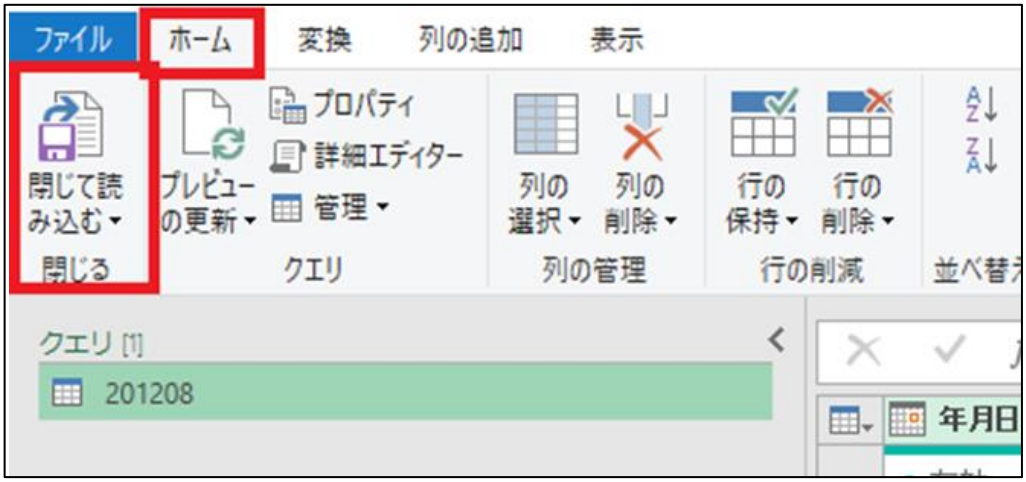


12. Excel で新規ブックを開き、[データ]タブ→[データの取得]→「その他のデータソースから」→「空のクエリ」をクリックする



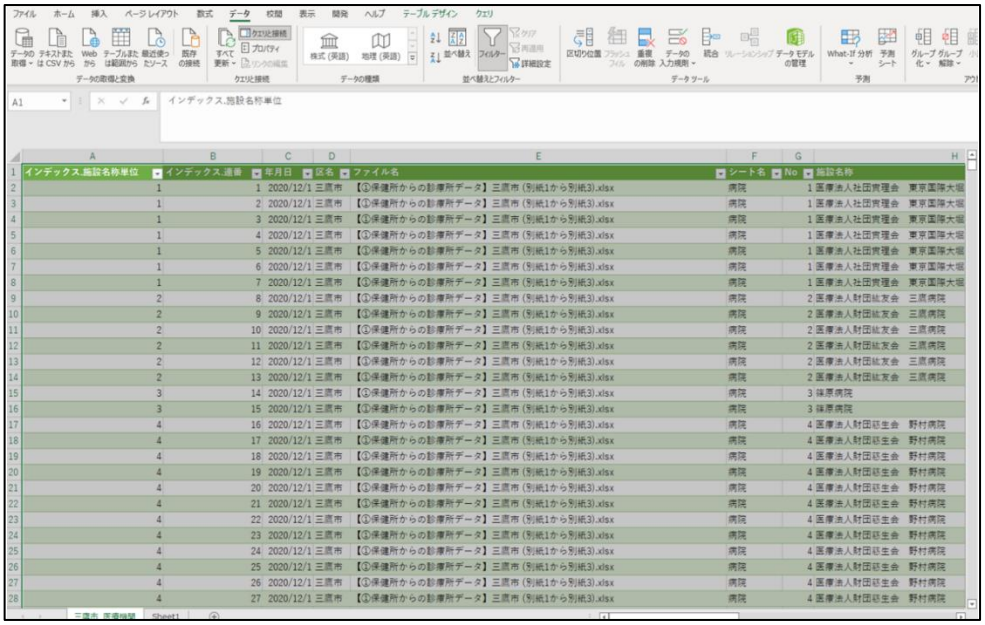
13. [ホーム]タブ→「該当クエリ」を右クリックし、新規ブックの Power Query 詳細エディターを表示する

14. [ホーム]タブ→[閉じて読み込む]をクリックする



【お願い】
データ取得元（ソース）のファイルは閉じた状態で作業を行ってください。

15. 読み込まれたデータを確認、データを全選択し、コピーする



16. CSV ファイルを開き、貼り付ける

17. ファイルの種類欄で「テキスト（タブ区切り）」を指定して CSV を保存する

手順（ケース[D]）

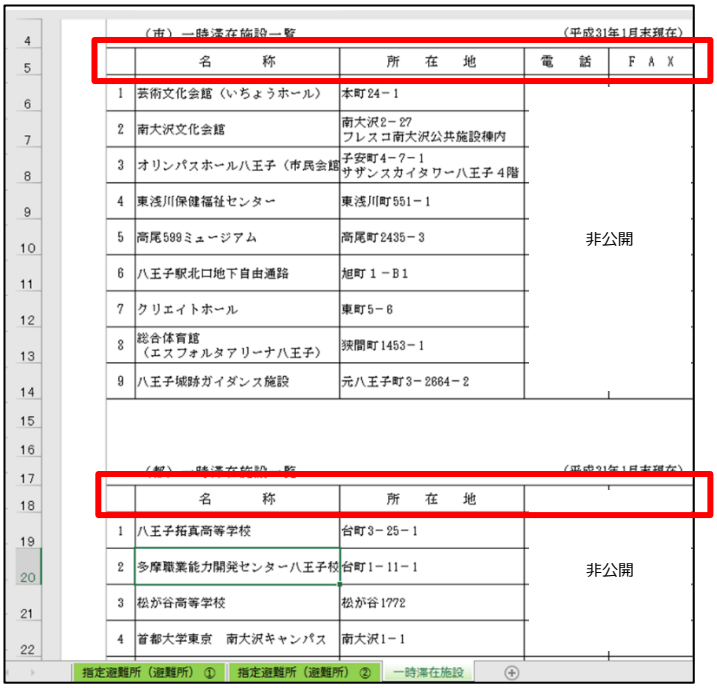
下記の手順で Excel から CSV にデータを変換します。

【お願い】： データ取得元（ソース）のファイルは閉じた状態で作業を行ってください。

① **Excel の表で、列名の状況を確認します。** **（詳細： p. 78）**
② **Excel で整形、加工したいファイルをインポートします。** **（詳細： p. 79）**
③ **Excel の Power Query 機能で、ブックごとにデータを整形します。** **（詳細： p. 80）**
④ **Excel の Power Query 機能で、Excel ブック用の処理を行います。**
※④はケース[C]の「Excel の Power Query 機能で、複数パターンが発生するクエリを結合する」 （詳細： p. 71）をご参照ください。

① Excel の表で、列名の状況を確認する

1. 複数ある表の、それぞれの列名が一致しているか目視で確認する



（市）一時滞在施設一覧　（平成31年1月末現在）

	名　称	所　在　地	電　話	F A X
1	芸術文化会館（いちょうホール）	本町24－1		
2	南大沢文化会館	南大沢2－27 フレスコ南大沢公共施設棟内		
3	オリンパスホール八王子（市民会館）	子安町4－7－1 サザンスカイタワー八王子 4階		
4	東浅川保健福祉センター	東浅川町551－1		
5	高尾599ミュージアム	高尾町2435－3	非公開	
6	八王子駅北口地下自由通路	旭町 1 －B 1		
7	クリエイトホール	東町5－6		
8	総合体育館 （エスフォルタアリーナ八王子）	狭間町1453－1		
9	八王子城跡ガイダンス施設	元八王子町3－2664－2		

（都）一時滞在施設一覧　（平成31年1月末現在）

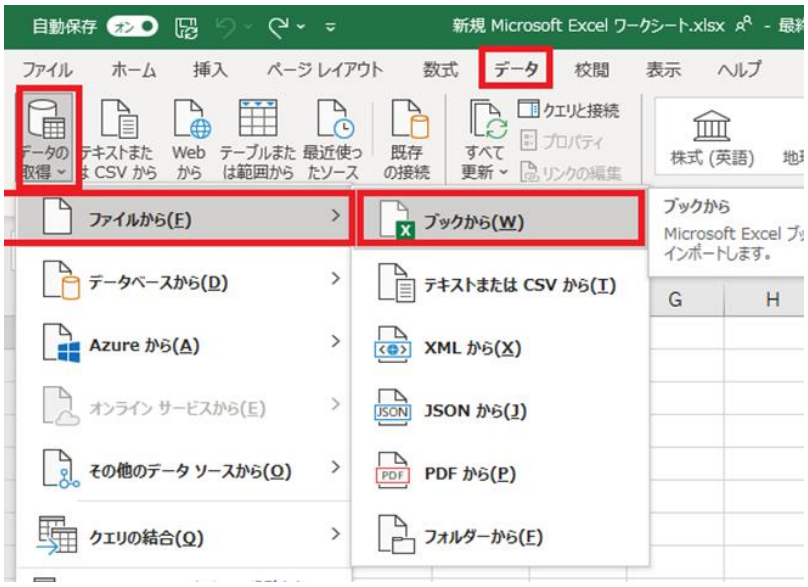
	名　称	所　在　地	
1	八王子拓真高等学校	台町3－25－1	
2	多摩職業能力開発センター八王子校	台町1－11－1	非公開
3	松が谷高等学校	松が谷1772	
4	首都大学東京　南大沢キャンパス	南大沢1－1	

指定避難所（避難所）①　指定避難所（避難所）②　一時滞在施設

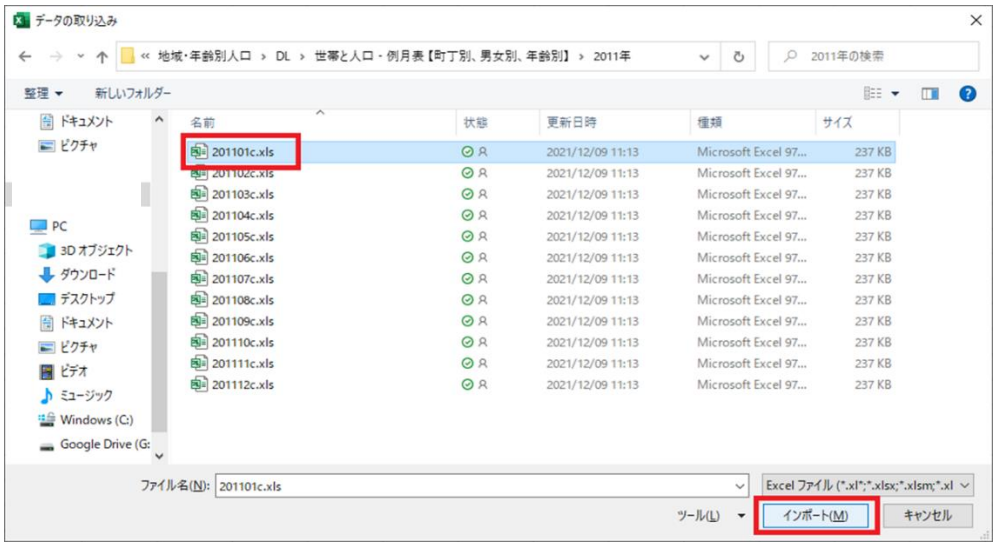
② Excel で整形、加工したいファイルをインポートする

1. データ取得元（ソース）のファイルを開いている場合は、閉じておく

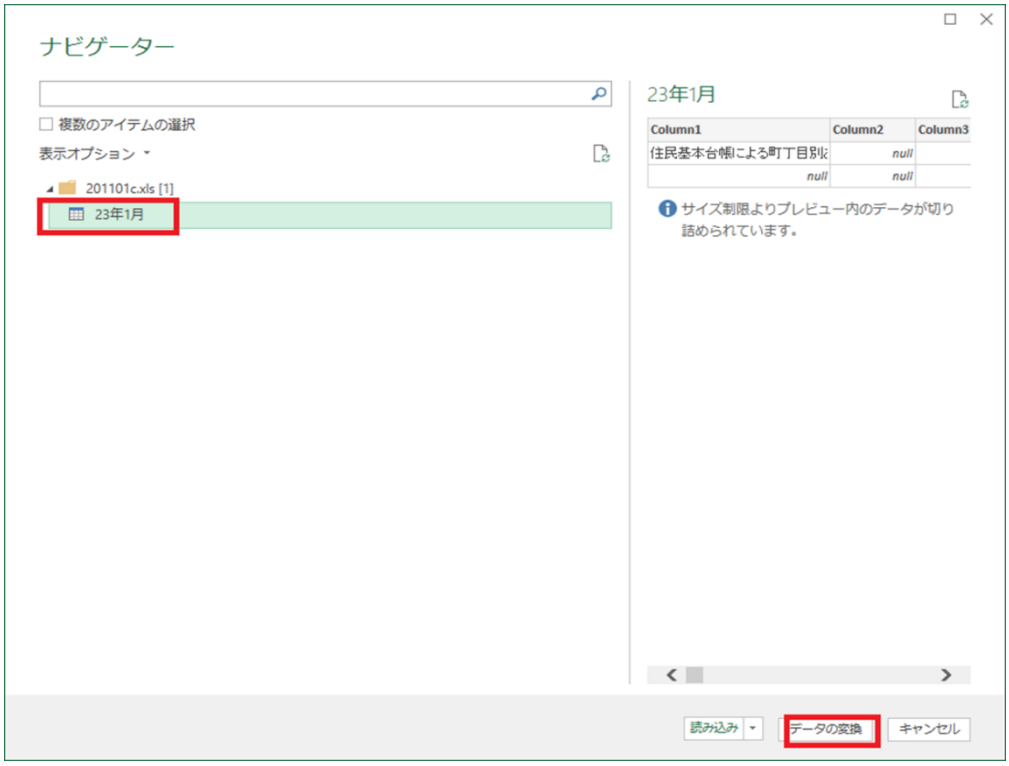
2. [データ]タブ→[データの取得]→「ファイルから」→「ブックから」をクリックする



3. Excel ファイルをクリックして選び、[インポート]をクリックする



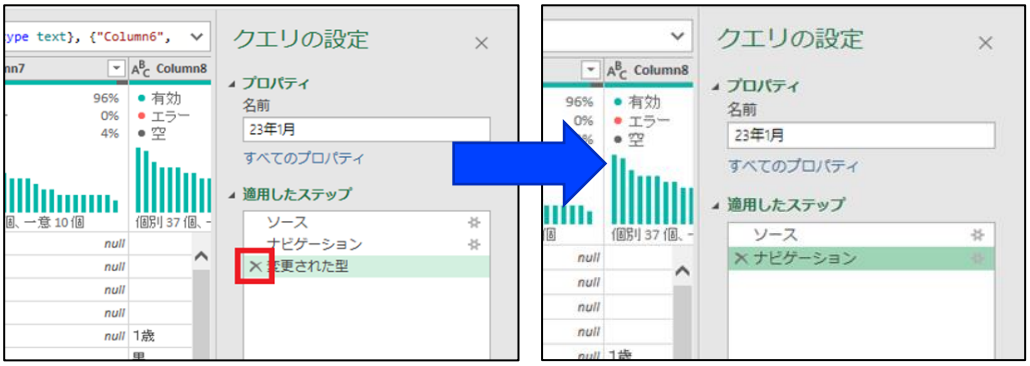
4.　変換するアイテムをクリックして選び、[データの変換]をクリックする



ここでの「アイテム」は、Excel のシートです。シート単位で複数表示されている場合は、必要なシートを選びます。

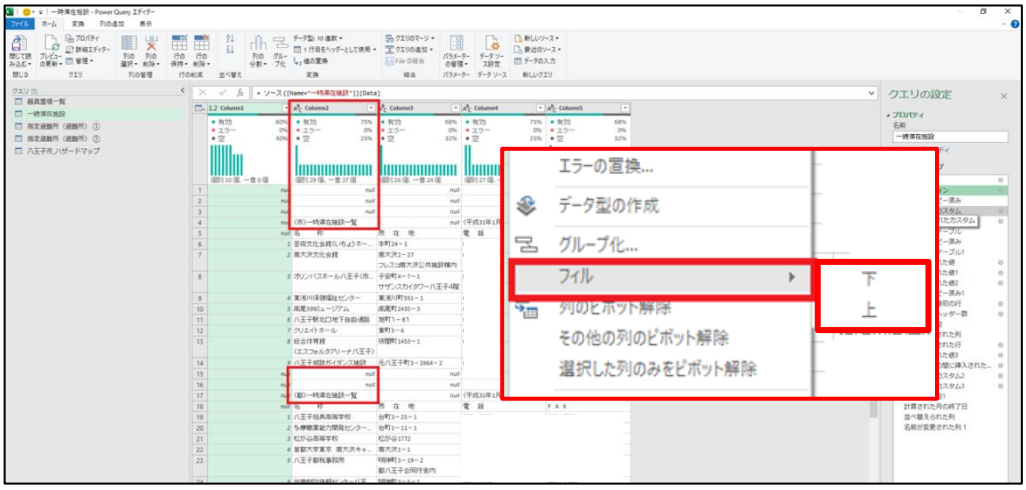
③　Excel の Power Query 機能でブックごとにデータを整形する

1.　「クエリの設定」で、「変更された型」ステップをクリックして削除する



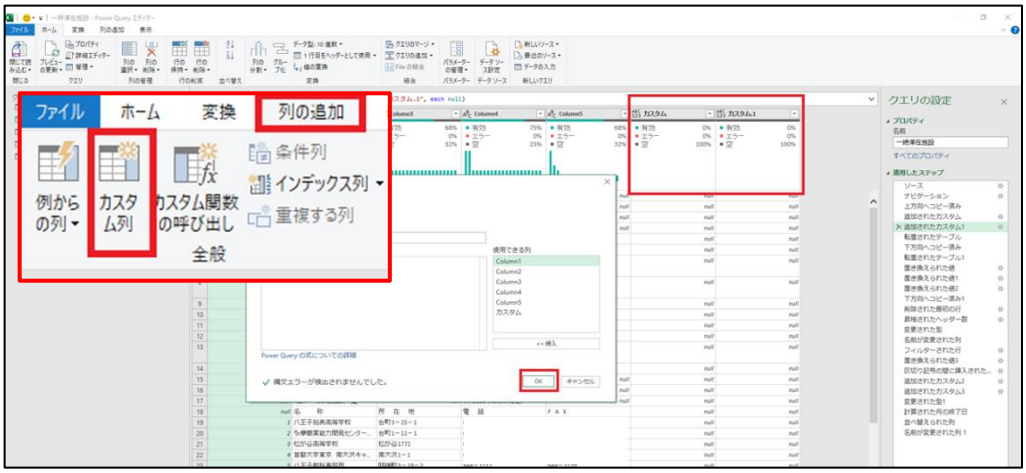
「変更された型」は操作を行うと自動的に追加されますが、手順は不要なので削除します。

2.　施設区分と名称の列を右クリックし、[フィル]→[下]または[上]をクリックする



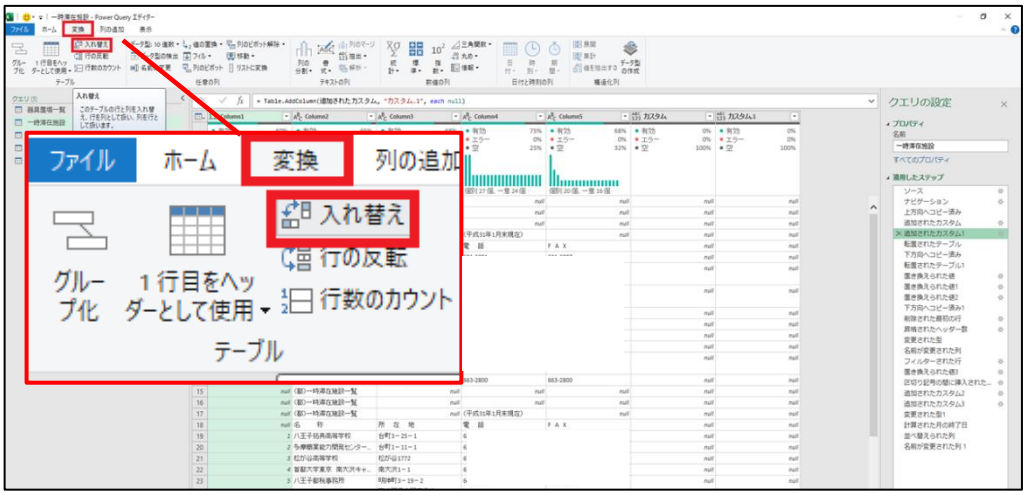
施設区分と名称が同列に混在するため分離します。

3. [列の追加]タブ→[カスタム列]をクリックする



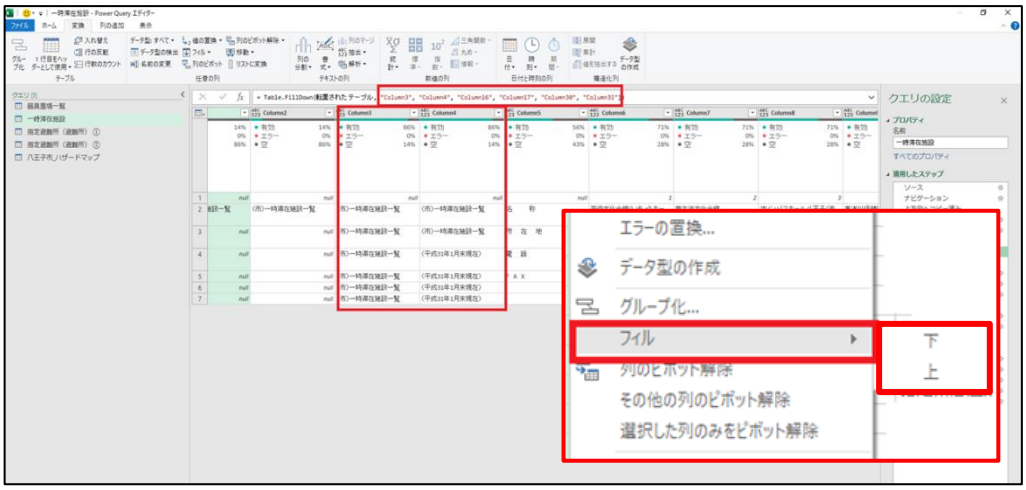
今回は最終列をコピーし、年月日のカスタム列を作成します。

4. [変換]タブ→[入れ替え]をクリックする

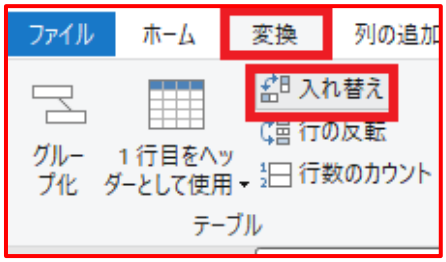


ここでは追加したカスタム列に値をコピーするため、行列を入れ替える処理を行います。

5. 施設区分と名称の列を右クリックし、[フィル]→[下]または[上]をクリックする



6. [変換]タブ→[入れ替え]をクリックする



7. 年月日の列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

施設区分列にフィルされている年月日データの値を置換します。

・対象列を選択した状態で右クリックし、「値の置換」を選ぶ。

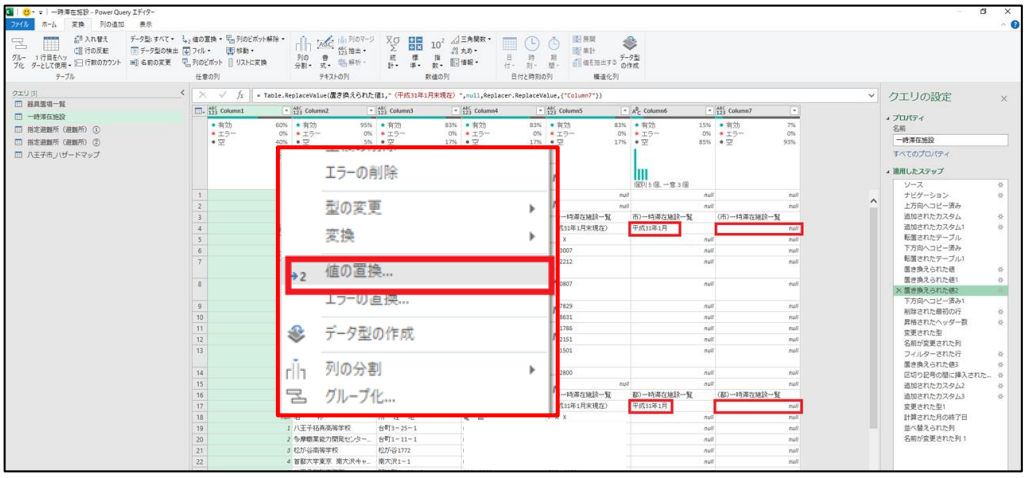
置換する値：年月日の文言　→　置換後「null」

年月日列を日付型でエラーとならない形式に整形します。

・対象列選択状態で　右クリック　「値の置換」

置換する値：不要な()や現在、未等の文言　→　置換後「""」（ダブルクォーテーション）

ここでの不要データ置換作業は、後の手順で日付データをフィルするために行います。

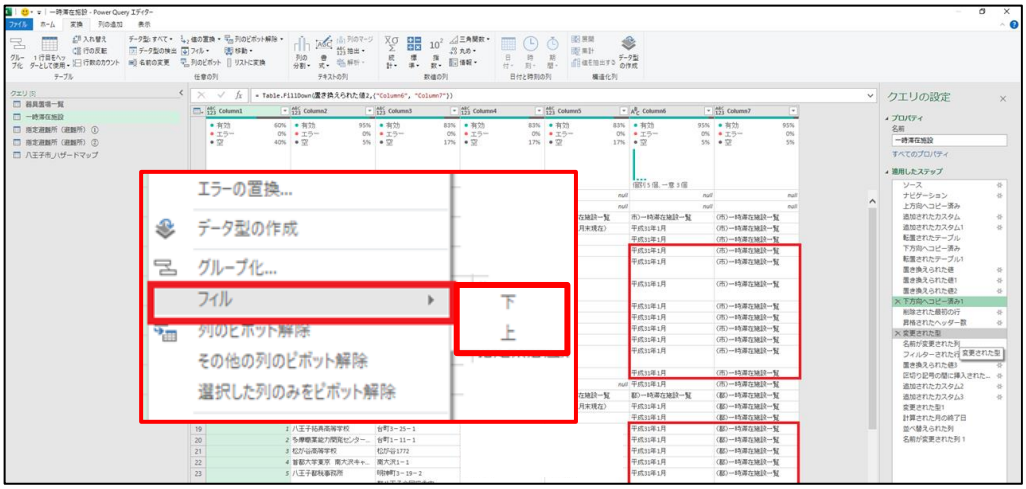


日付型の列での注意点です。

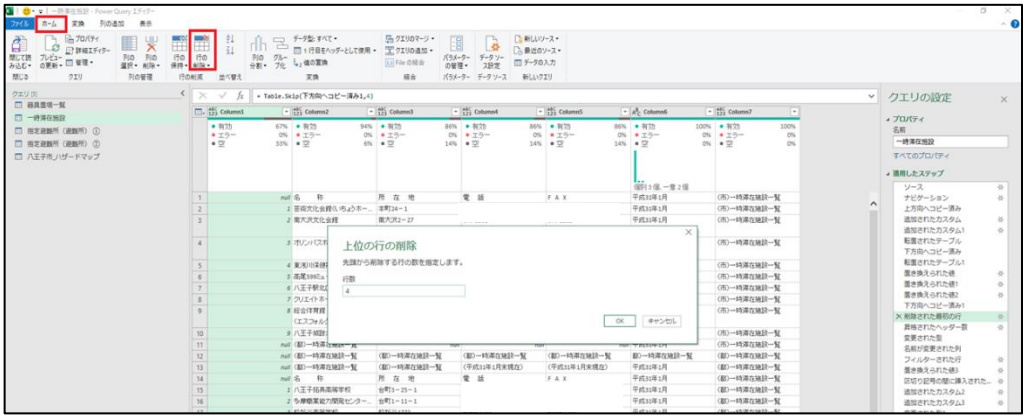
和暦表記も可能ですが、「元号」表記、「全角数字」、「漢数字」は取り扱えません。

× 令和元年１月１日	○ 令和１年１月１日
× 令和１年1月1日	○ 令和１年１月１日
× 令和一年一月一日	○ 令和１年１月１日
× ２０１９０１０１	○ 20190101
× ２０１９/1/1	○ 2019/1/1

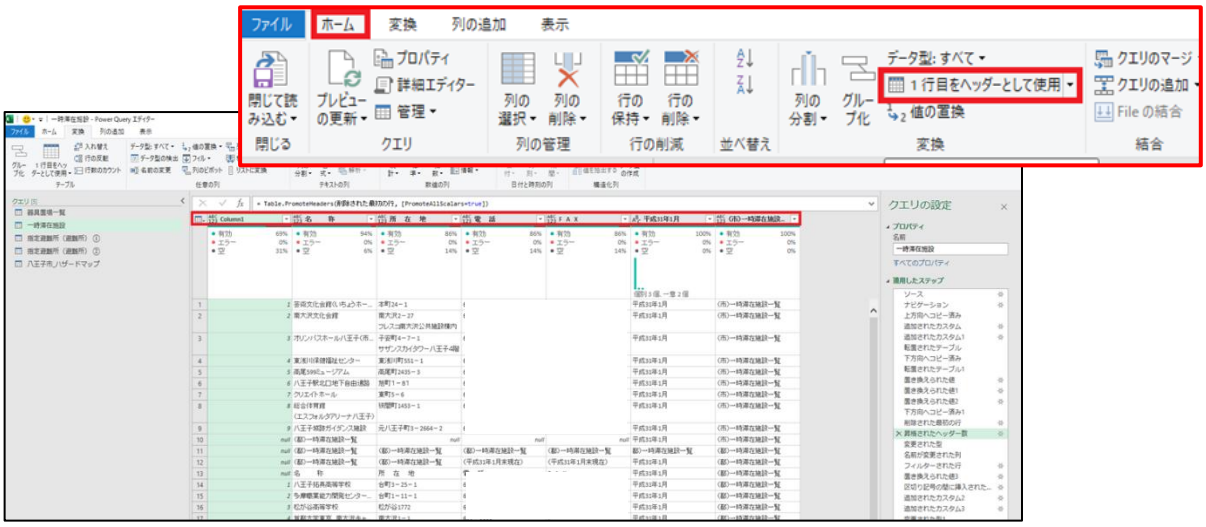
8. 施設区分と名称の列を右クリックし、[フィル]→[下]または[上]をクリックする



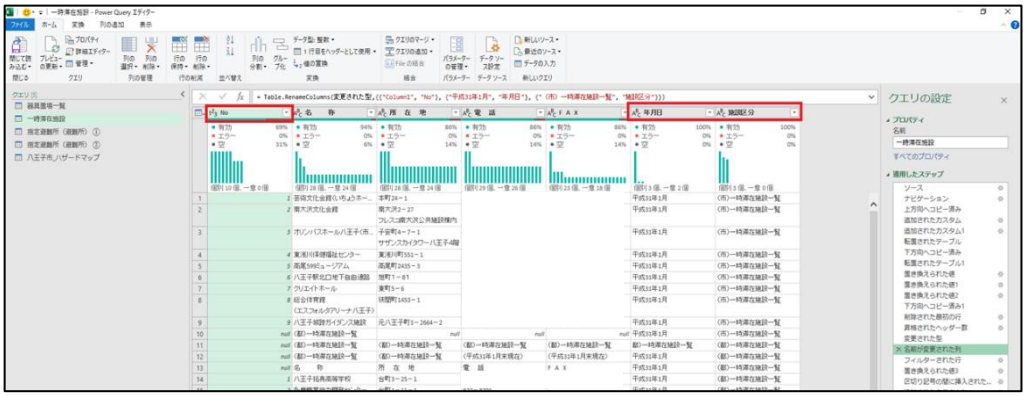
9. [ホーム]タブ→[行の削除]→「上位の行の削除」をクリックし、削除する上位の行数を入力する



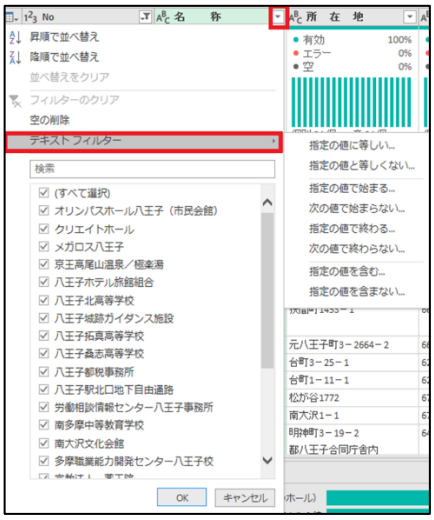
10. [ホーム]タブ→[1 行目をヘッダーとして使用]をクリックする



11. 追加した「カスタム列」や「列名」の名前を指定しなかった列名があれば、列をダブルクリックして任意の列名を入力する

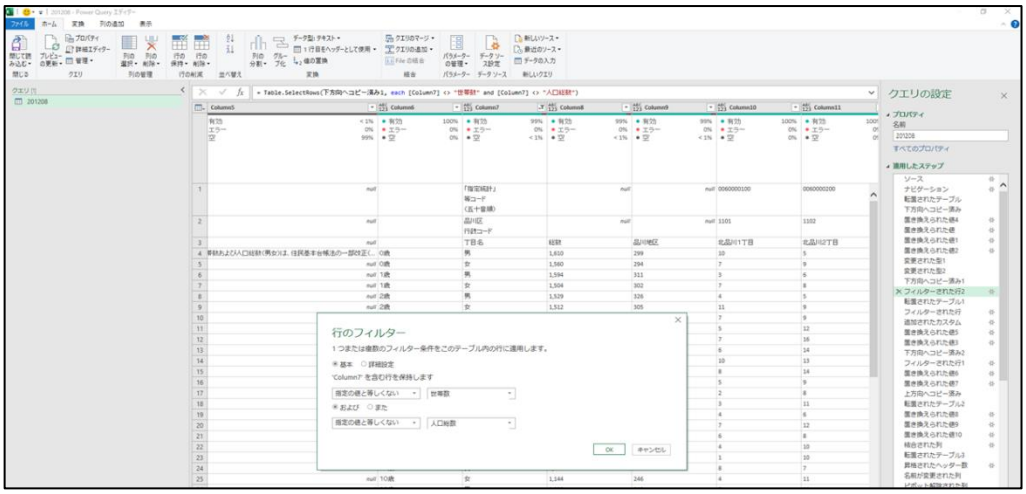


12. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しくない」を選び、[OK]をクリックする

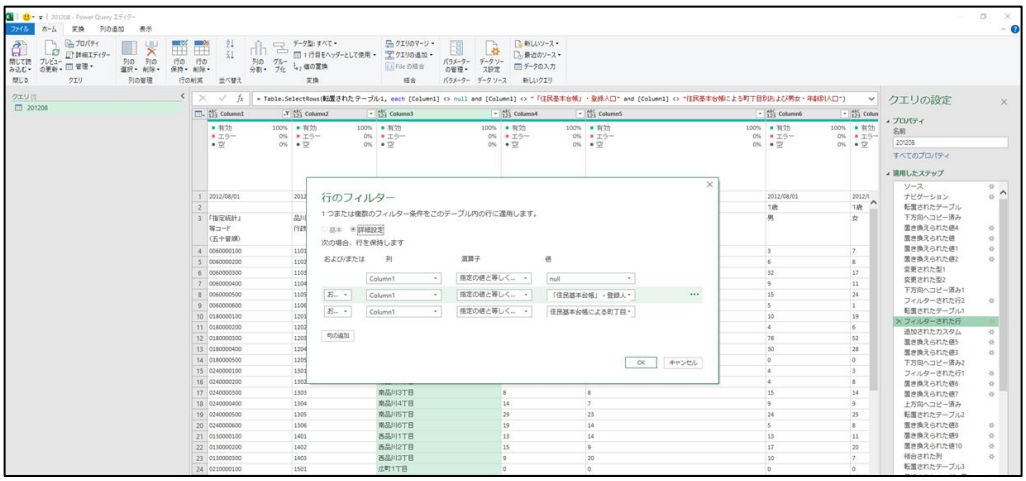


ここでは必要行を残すフィルターを設定しています。

13. 年月日の列を右クリックし、除外するデータを非表示にする



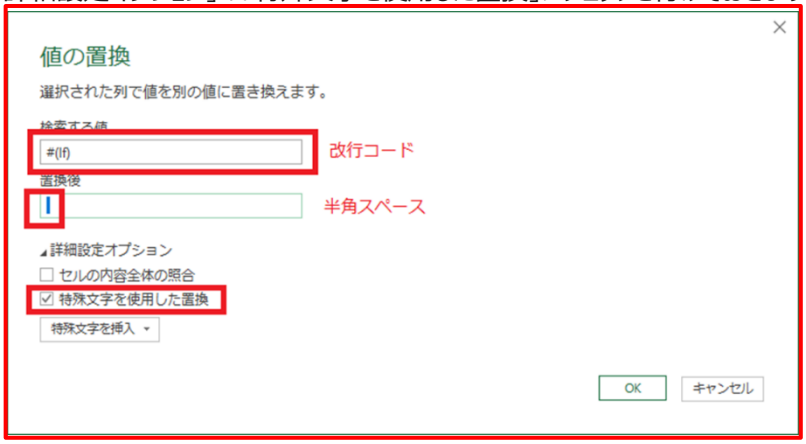
14. 対象列の不要行を削除するため、対象列の右側「▼」をクリックし、除外するデータを非表示にする



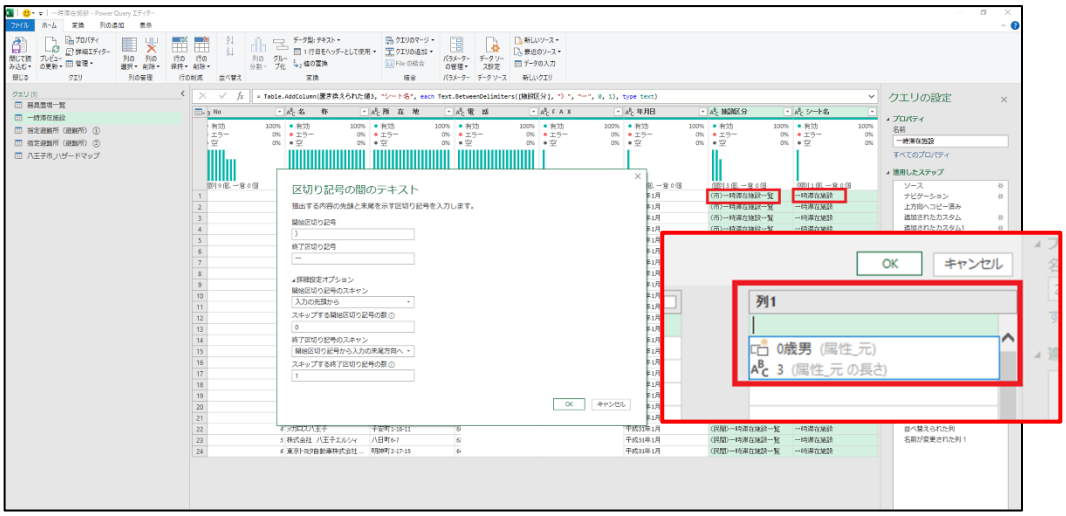
15. 対象列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

置換する値「#(lf)」（改行コード）　→　置換後「 」（半角スペース）
「詳細設定オプション」の「特殊文字を使用した置換」にチェックを付けておきます。

ここでは改行コードを削除しています。

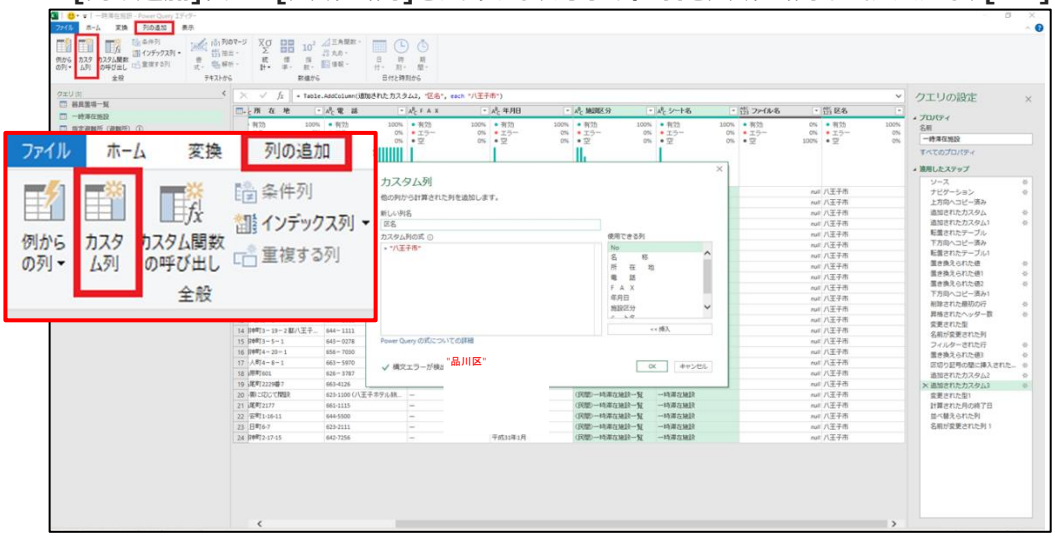


16. 開設届の列が選択された状態で右クリックし、[列の追加]タブ→[例からの列]→[選択範囲から]をクリックする

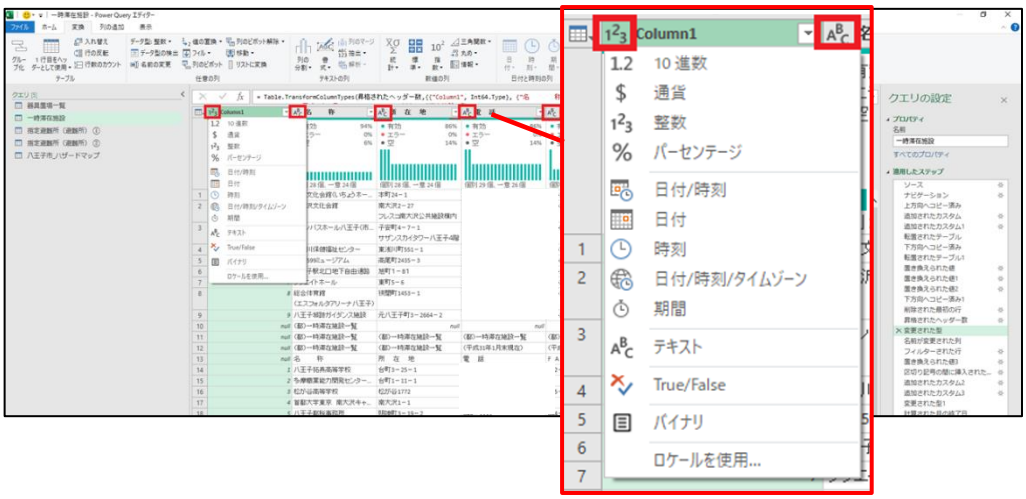


ここでは、施設区分列から必要範囲を切り出します。

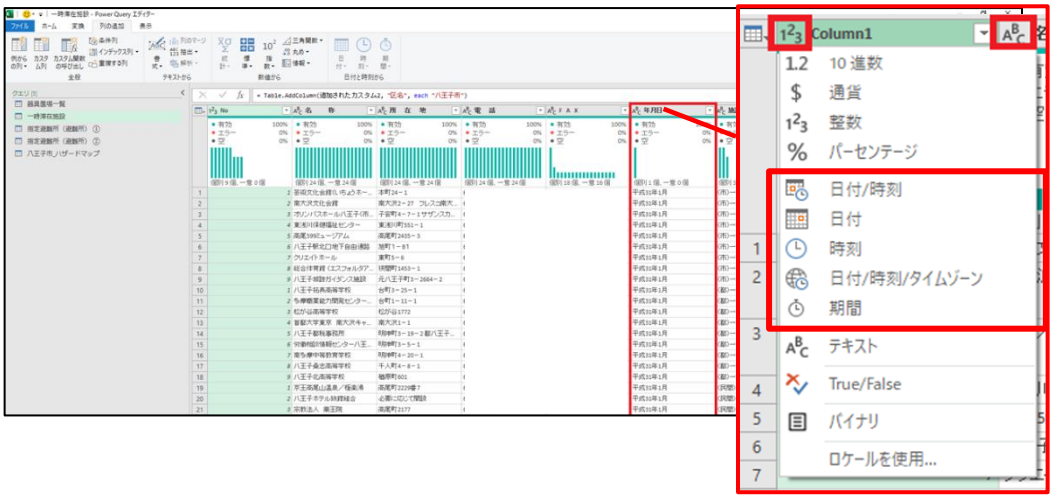
17. [列の追加]タブ→[カスタム列]をクリックし、対象の区名をカスタム列の式に入力し、[OK]をクリックする



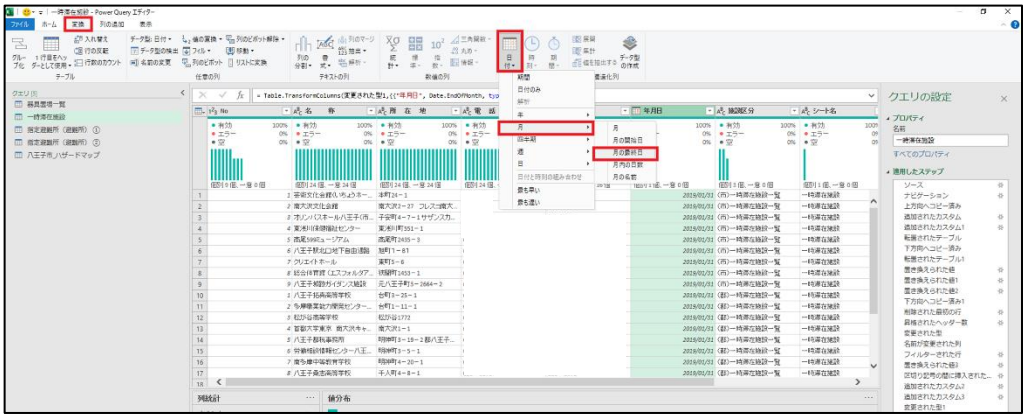
18. 列のヘッダー上部にある「ABC123」アイコンをクリックし、任意の型を選択する



19. 年月日のヘッダー上部にある「ABC123」アイコンをクリックし、日付型を選択する

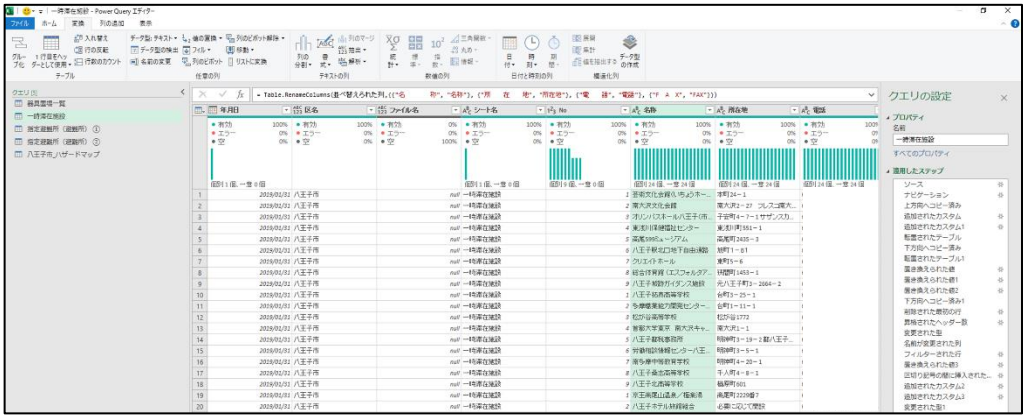


20. [変換]タブ→[日付] →[月] →[該当の日付を選択]をクリックし、年月日列に日付を付与する



例に使用したデータは日付が未となっていたため、「月の最終日」を選択しています。

21. ヘッダー列名に不要な空白が存在する場合は、ヘッダー列名をダブルクリックし、任意の列名を入力する



AI-OCR によるデータ変換

下記の手順で PDF から CSV データを生成する変換を進めます。
【準備】： Adobe Acrobat DC および CLOVA OCR READER を利用できる環境を確保します。（初回のみ）

① **Adobe Acrobat DC で、AI-OCR に読み込むアクションを登録します。** **（詳細： p. 88）**
② **アクションを実行し、データを作成します。** **（詳細： p. 92）**
③ **CLOVA OCR READER で、AI-OCR 処理するテンプレートを作成します。** **（詳細： p. 94）**
④ **CLOVA OCR READER で、AI-OCR 処理を行い、CSV ファイルを作成します。** **（詳細： p. 97）**

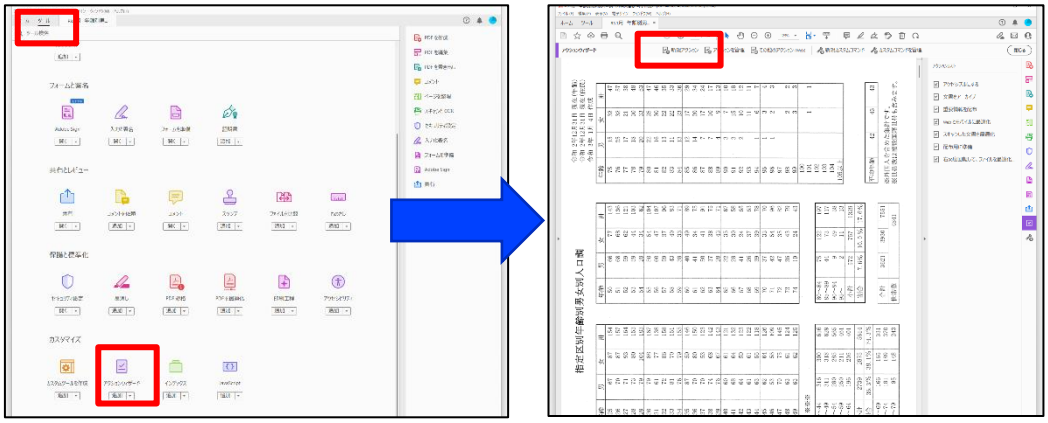
ご参考
本書記載以外の、CLOVA の詳しい操作については、下記をご参照ください。
https://clova.line.me/clova-ocr/

手順（ケース[E]）

① Adobe Acrobat DC で、AI-OCR に読み込むアクションを登録する

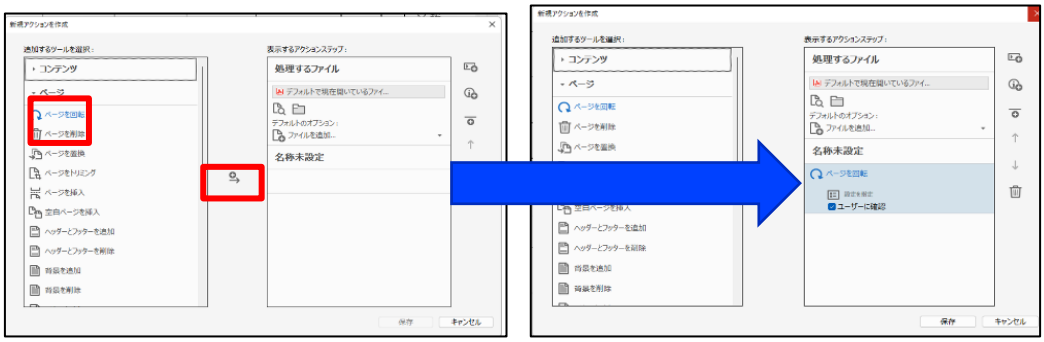
1. Adobe Acrobat DC を開く

2. ツールからアクションウイザードを開き、新規アクションをクリックする



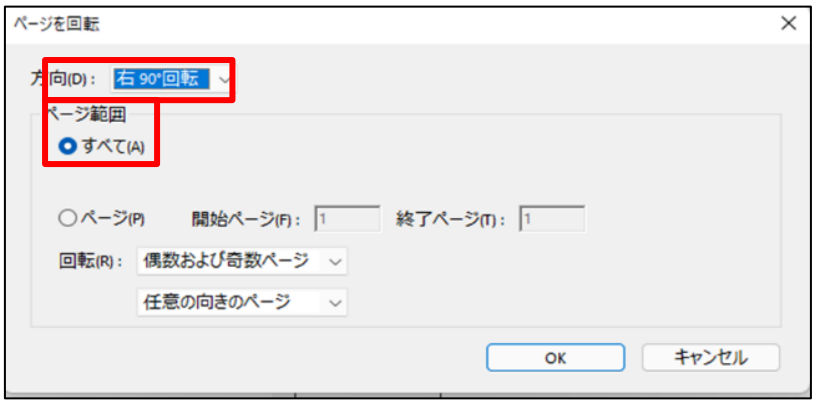
アクションは、一定の操作を記録し、繰り返し使うためのメニューです。

3. [ページ]→[ページを回転]をクリック、[→]をクリックして設定を右側パネルに移動させる



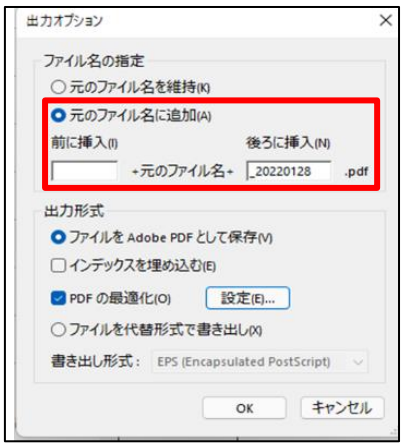
右側パネルに移動した操作が、上から順に実行されます。

4. 「すべて」と [右 90 度回転]を選択し、[OK]をクリックする



5. [保存と書き出し]→[保存]をクリック、[→]をクリックして設定を右側パネルに移動させる

6. 設定を変更し、[OK]をクリックする



ファイル名の指定：元のファイルに追加

- 後ろに挿入：_日付

7. PDF の最適化設定を変更し、[OK]をクリックする

互換性を確保：Acrobat10.0 及びそれ以降

- ダウンサンプルの解像度：300ppi
- カラーの圧縮：JPEG2000

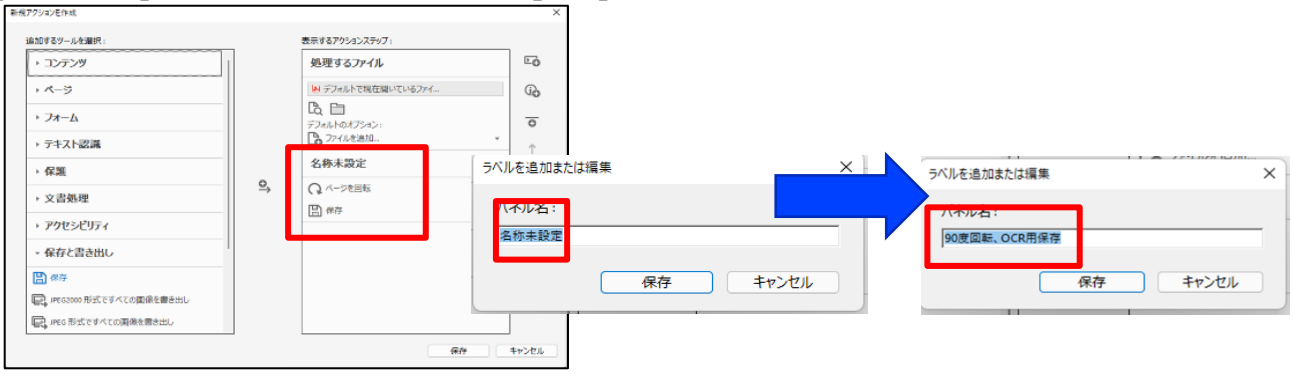
オブジェクトを破棄：すべてにチェックを入れる

ユーザーデータを破棄：すべてにチェックを入れる

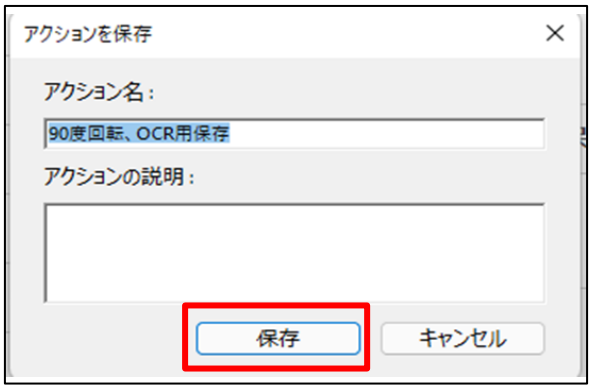
最適化：すべてにチェックを入れる

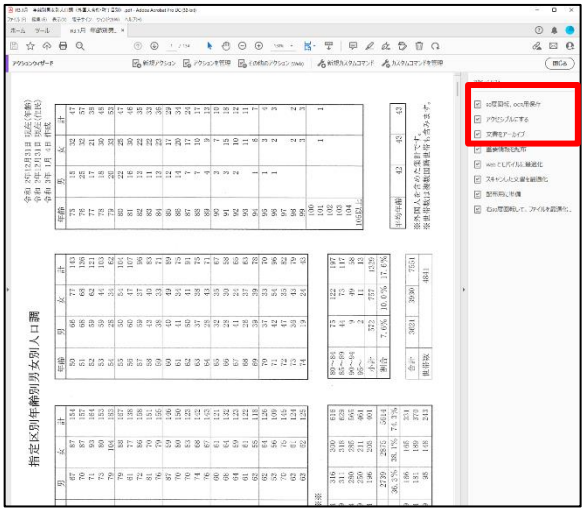
8. [保存]をクリック後、保存名を「OCR 用」と入力し、[OK]をクリックする

9. [名称未設定]をダブルクリック、処理名を入力し、[保存]をクリックする



10. アクション名を入力し、[保存]をクリックする

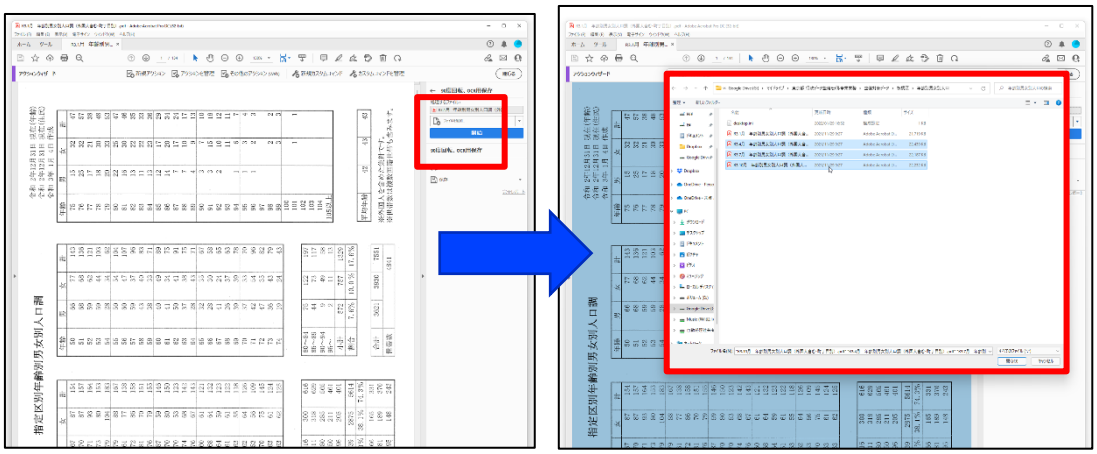




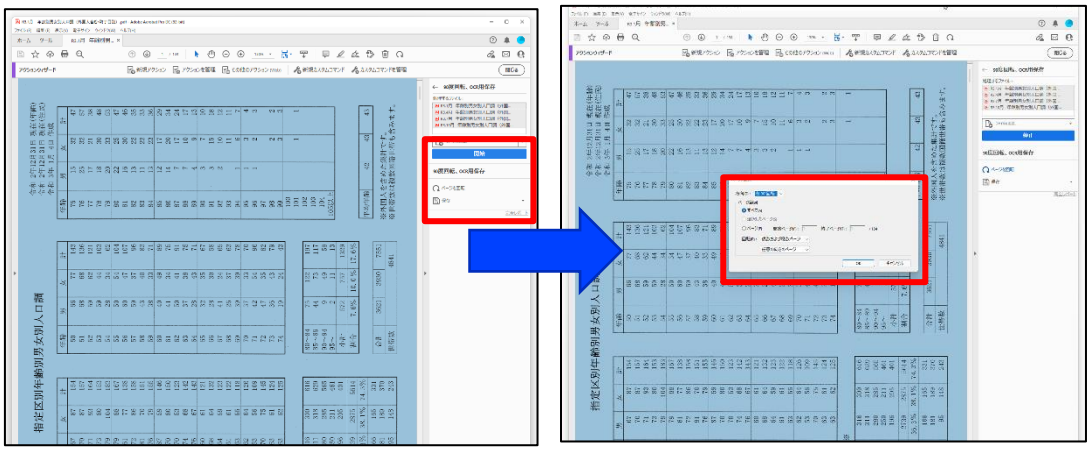
アクション名に追加されます。

② アクションを実行し、データを作成する

1. 作成したアクションリストをクリックし、処理するファイルを追加する

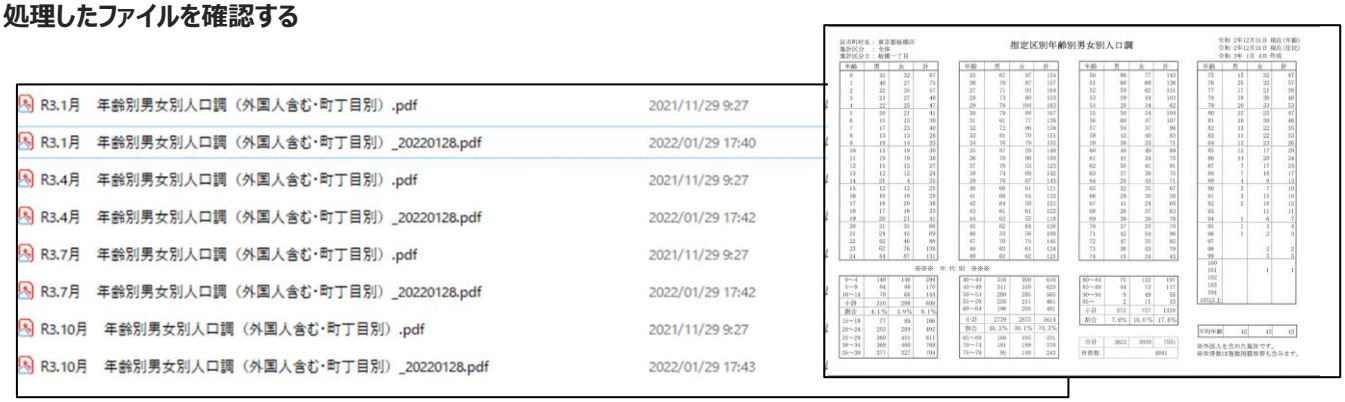


2. [開始]をクリックし、ページの回転で[OK]をクリックする



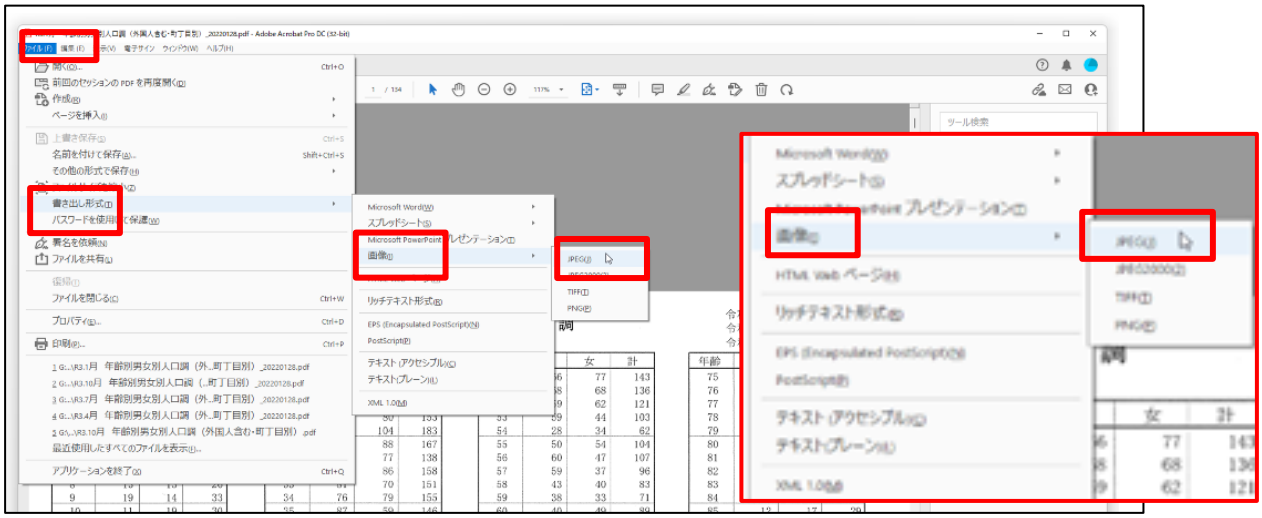
処理が実行されます。処理に時間がかかることがあります。

3. 処理したファイルを確認する

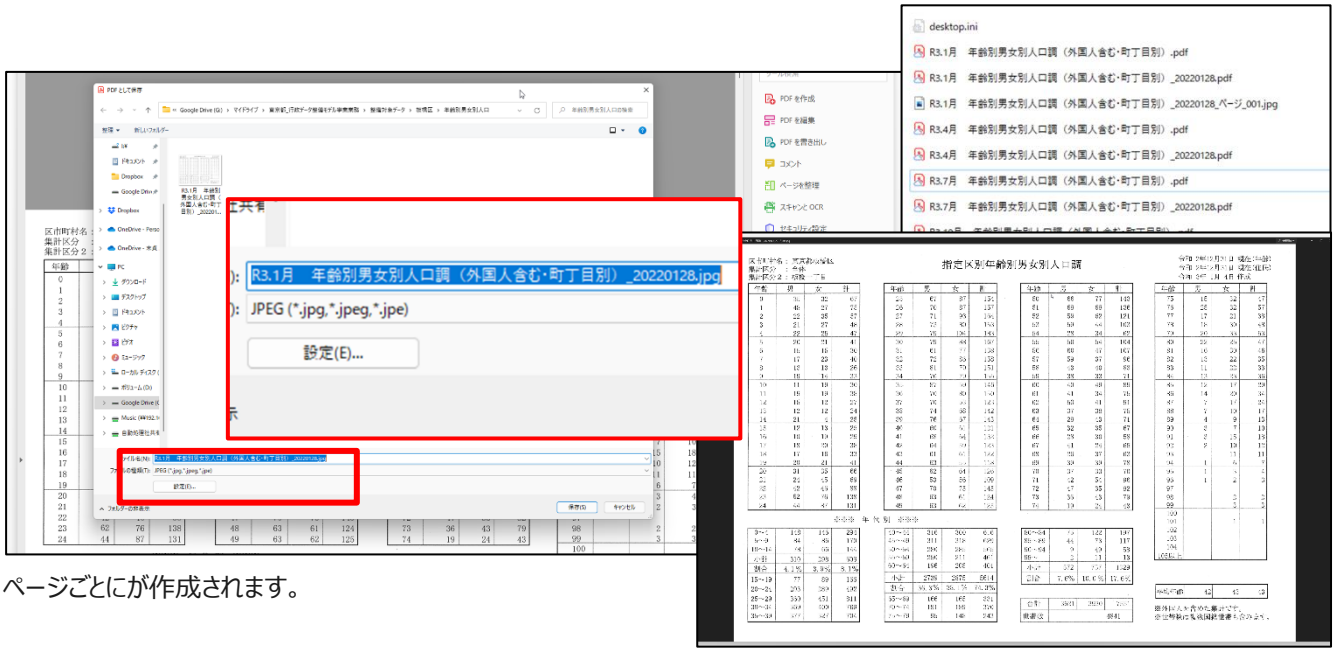


ファイル名末尾に指定した日付が入ったファイルが今回作成したデータです。

4. OCR 処理用の JPEG ファイルを 1 ページ出力する（[ファイル]→[書き出し形式]→[]→[JPEG]）



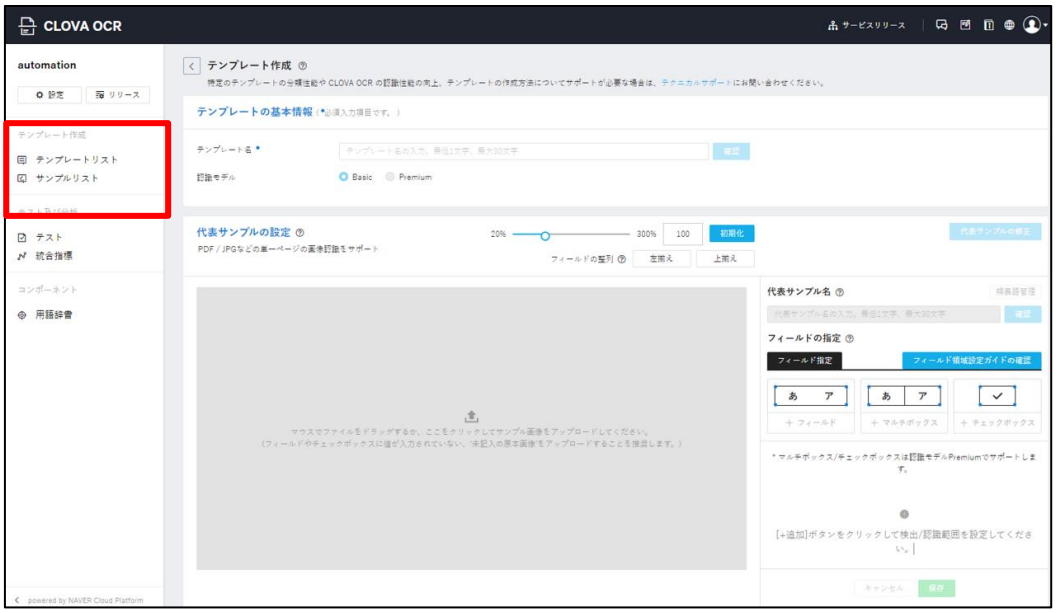
5. ファイル名を入力し、[保存]をクリックする



ページごとにが作成されます。

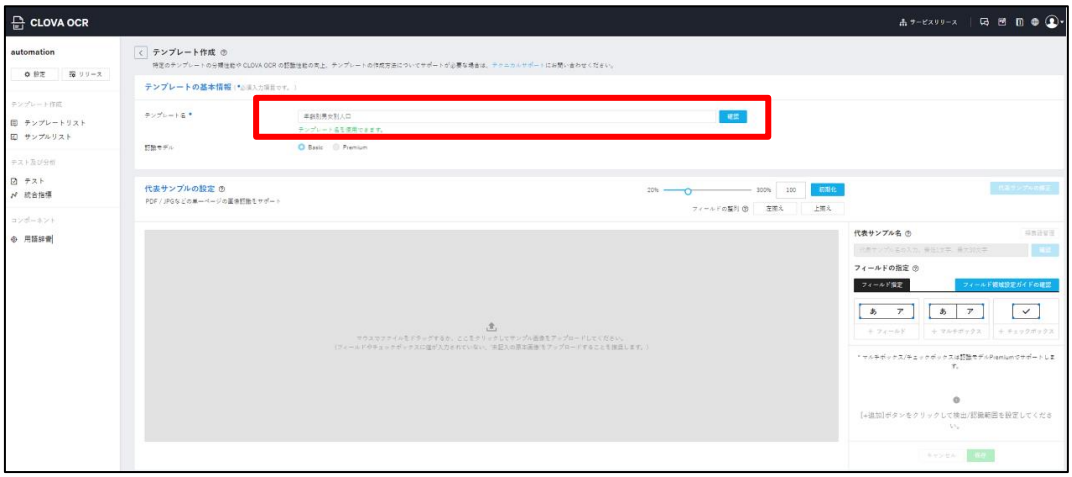
③ CLOVA OCR READER で、AI-OCR 処理するテンプレートを作成する

1. CLOVA OCR のテンプレートリスト画面にて、テンプレート作成ボタンをクリック



※CLOVA OCR の初期設定がすべて完了していることを前提に手順を記載しております。

2. テンプレート名を入力し、[確認]をクリック

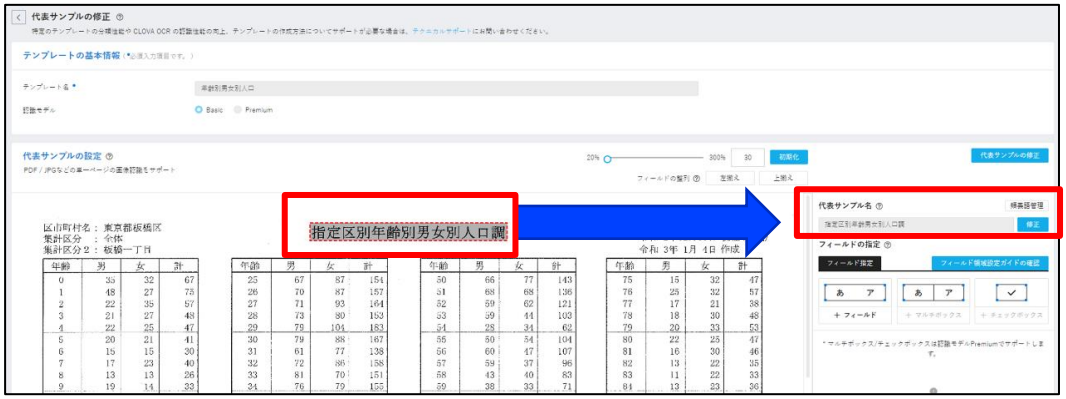


テンプレート名が確定されたことを確認してください。

3. ②で作成したファイル（JPG 形式）をドラッグ&ドロップで追加する

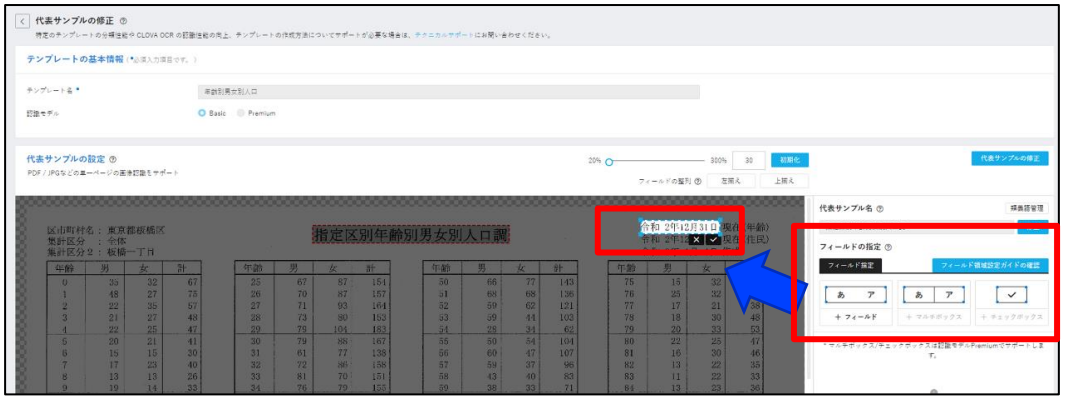
領域色が反転すると、ファイルが認識されています。そのまま保持状態を解除します。

4. PDF のタイトルにする領域を選択した後、「代表サンプル名」に『指定区別年齢男女別人口調』と入力、[確認]をクリックする



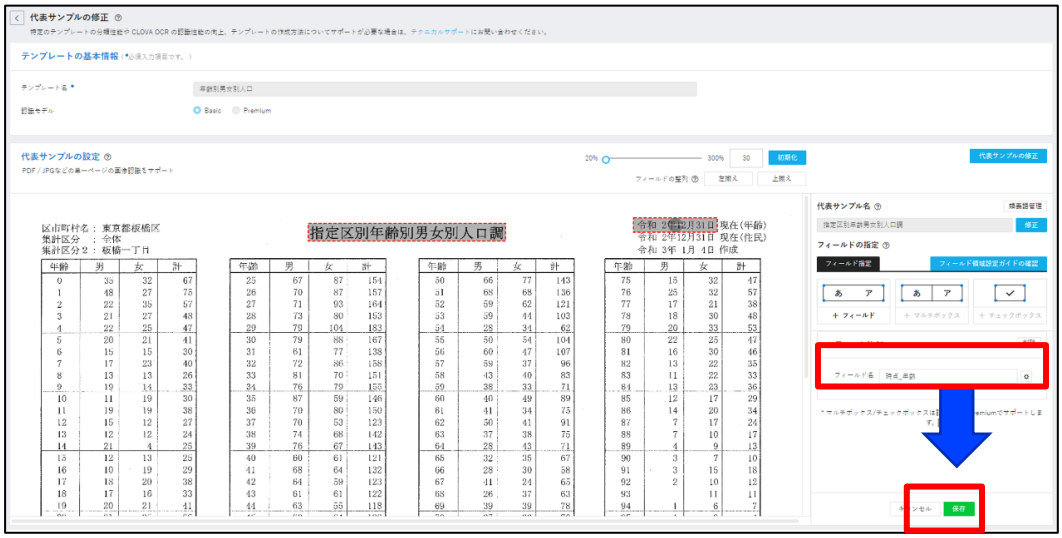
領域はマウスのドラッグ操作で選択できます。
選択部分は破線で囲まれ、色が反転します。

5. [+フィールド]をクリックし、読み取りたい領域を選択する



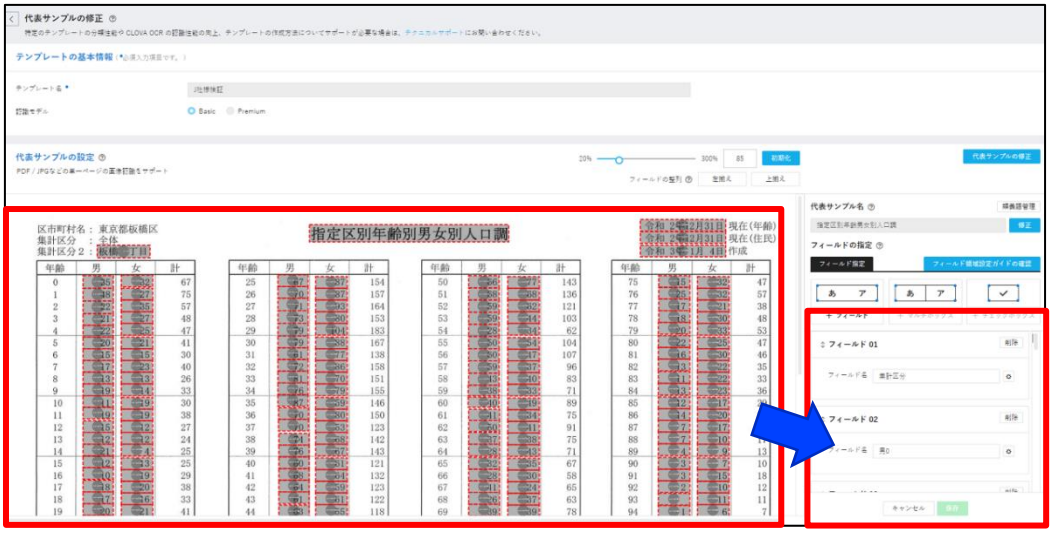
領域はマウスのドラッグ操作で選択できます。
選択部分は破線で囲まれ、色が反転します。

6. フィールド名を設定し、[保存]をクリックする



領域はマウスのドラッグ操作で選択できます。

7. 手順 5、手順 6 を繰り返し、読み取りたい領域すべてにフィールドを設定する



最後に[保存]をクリックするのを忘れないようにしてください。

④ CLOVA OCR READER で、AI-OCR 処理を行い、CSV ファイルを作成する

1. CLOVA OCR READER を開き、ファイルをアップロードする

※ CLOVA OCR READER の初期設定がすべて完了していることを前提に手順を記載しております。

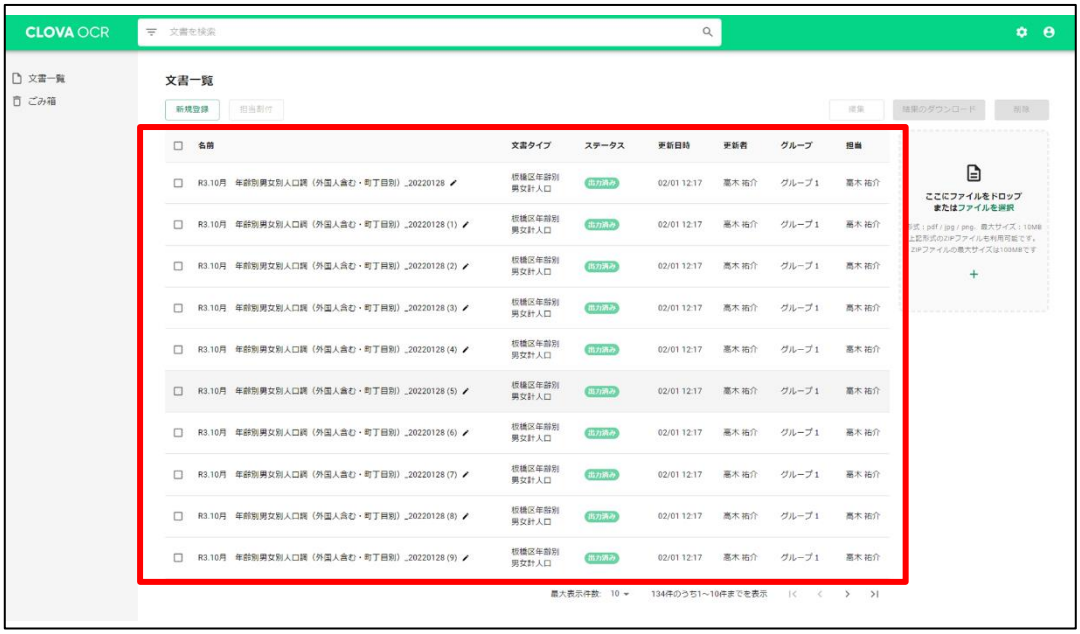
画面右側が、ファイルをアップロードする領域です。ここにドラッグ操作か、領域をクリックしてファイルをアップロードします。

ファイル選択画面では、[+]をクリックしてファイルの追加を繰り返します。

2. CLOVA OCR READER を開き、読み込む PDF ファイルをアップロードする

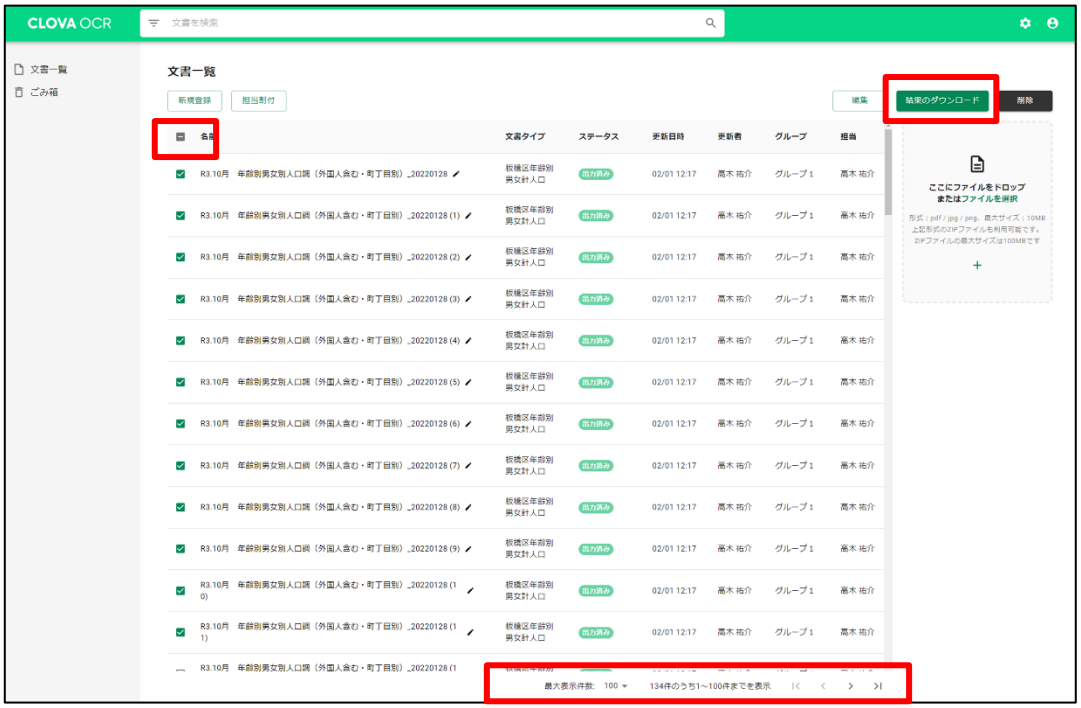
ファイルの追加が終わったら、[登録]をクリックします、

読み込みがはじまります。



読み込み後は、1 ファイルにつき、1 行の文書情報が表示されます。

3. CSV データを入手したい文書のチェックボックスにチェックを付け、[結果のダウンロード]をクリックする



数ページにわたる数の文書を一度に表示するには、画面下部の表示件数を変更します。

文書先頭のチェックボックスは、画面上部の「名前」左側のチェックボックスでまとめて選択できます。

	A	B	C	D	E	F	G	H	I	J	
1	title	集計区分	男0	男1	男2	男3	男4	男5	男6	男7	男8
2	指定区別年	板橋一丁目	37	33	29	19	24	17	22	11	
3	指定区別年	板橋二丁目	18	14	22	18	19	16	9	12	
4	指定区別年	板橋三丁目	16	26	23	19	22	18	22	31	
5	指定区別年	板橋四丁目	22	6	14	17	13	13	14	13	
6	指定区別年	加賀一丁目	33	48	59	42	58	67	58	61	
7	指定区別年	加賀二丁目	12	10	14	7	16	21	25	17	
8	指定区別年	大山東町	19	19	18	16	18	15	18	9	
9	指定区別年	大山金井町	25	12	16	14	10	20	8	14	
10	指定区別年	熊野町	7	17	10	9	15	22	13	12	
11	指定区別年	中丸町	19	28	30	26	16	18	31	19	
12	指定区別年	南町	16	9	7	10	10	22	5	14	

[結果のダウンロード]クリック後に、CSV ファイルがダウンロードされます。

手順 3

データを適切に配置し、データの精度を高める

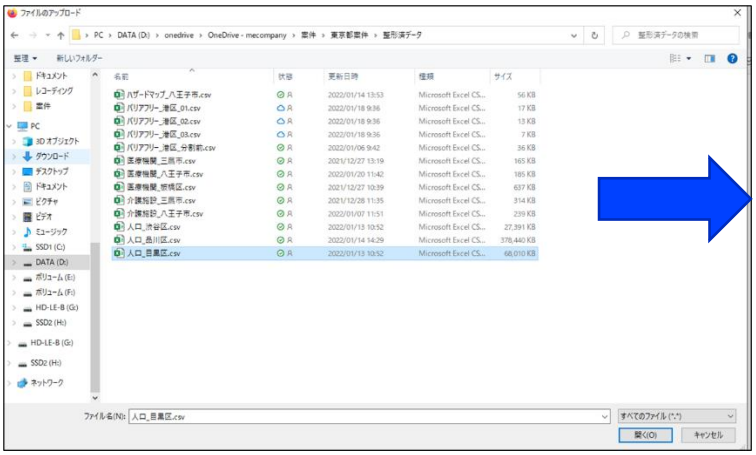
手順３はデータのマッピングとクレンジングの作業を行います。
この作業は、目視で実施すると膨大な作業が発生し、手入力で実施すると誤りが発生する原因になりますので、基本的には API や外注作業にて対応を行います。この章に記載している作業を参考に、調達を実施して適切なデータに変換を行ってください。
この章ではその際に自動化が難しいケースについても記載をしています。

① **これまでの手順で整備した CSV データをデータベースが読み込める場所に格納する** **（参考 p. 101）**
② **データをマッピングする** **（参考 p. 103）**
③ **マッピングデータを目視で確認する**
④ **ドメインデータを判別する**
⑤ **ドメインごとにクレンジングを行う** **（参考「クレンジング難度が高い例」p. 105）**
⑥ **データを正規化する** **（参考 p. 108）**
⑦ **ドメインデータをチェックする**

機能紹介／CSV 外形一括チェック

項目値の詳細チェック等が不要な場合等、簡易的に CSV ファイルの構造チェックが可能です。

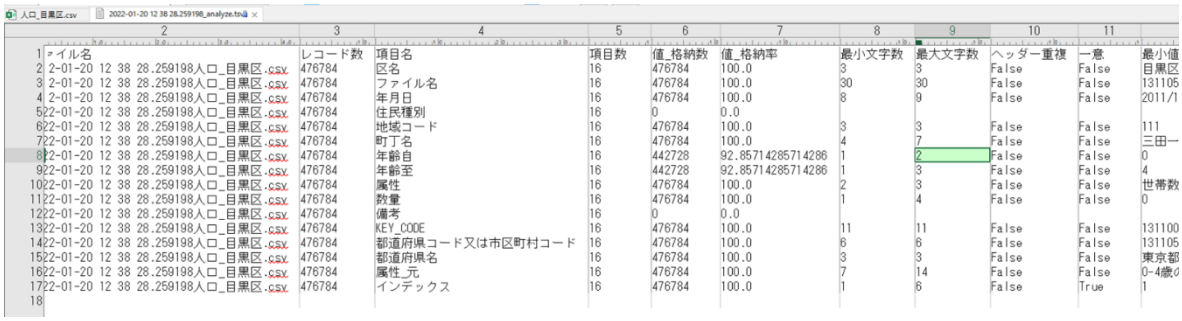
対象ファイルを選択



電話番号正規化関数
ミーカンパニー社独自

※複数ファイル一括チェックが可能です。
出力最終段にて全一括チェック等に利用します。

CSV 分析結果



	ファイル名	レコード数	項目名	項目数	値_格納数	値_格納率	最小文字数	最大文字数	ヘッダー重複	一意	最小値
2	2-01-20 12 38 28.259198人口_目黒区.csv	476784	区名	16	476784	100.0	3	3	False	False	目黒区
3	2-01-20 12 38 28.259198人口_目黒区.csv	476784	ファイル名	16	476784	100.0	30	30	False	False	131105
4	2-01-20 12 38 28.259198人口_目黒区.csv	476784	年月日	16	476784	100.0	8	9	False	False	2011/1
5	2-01-20 12 38 28.259198人口_目黒区.csv	476784	住民種別	16	0	0.0					
6	2-01-20 12 38 28.259198人口_目黒区.csv	476784	地域コード	16	476784	100.0	3	3	False	False	111
7	2-01-20 12 38 28.259198人口_目黒区.csv	476784	町丁名	16	476784	100.0	4	7	False	False	三田一
8	2-01-20 12 38 28.259198人口_目黒区.csv	476784	年齢自	16	442728	92.85714285714286	1	2	False	False	0
9	2-01-20 12 38 28.259198人口_目黒区.csv	476784	年齢至	16	442728	92.85714285714286	1	3	False	False	4
10	2-01-20 12 38 28.259198人口_目黒区.csv	476784	属性	16	476784	100.0	2	3	False	False	世帯数
11	2-01-20 12 38 28.259198人口_目黒区.csv	476784	数量	16	476784	100.0	1	4	False	False	0
12	2-01-20 12 38 28.259198人口_目黒区.csv	476784	備考	16	0	0.0					
13	2-01-20 12 38 28.259198人口_目黒区.csv	476784	KEY_CODE	16	476784	100.0	11	11	False	False	131100
14	2-01-20 12 38 28.259198人口_目黒区.csv	476784	都道府県コード又は市区町村コード	16	476784	100.0	6	6	False	False	131105
15	2-01-20 12 38 28.259198人口_目黒区.csv	476784	都道府県名	16	476784	100.0	3	3	False	False	東京都
16	2-01-20 12 38 28.259198人口_目黒区.csv	476784	属性_元	16	476784	100.0	7	14	False	False	0-4歳(
17	2-01-20 12 38 28.259198人口_目黒区.csv	476784	インデックス	16	476784	100.0	1	6	False	True	1

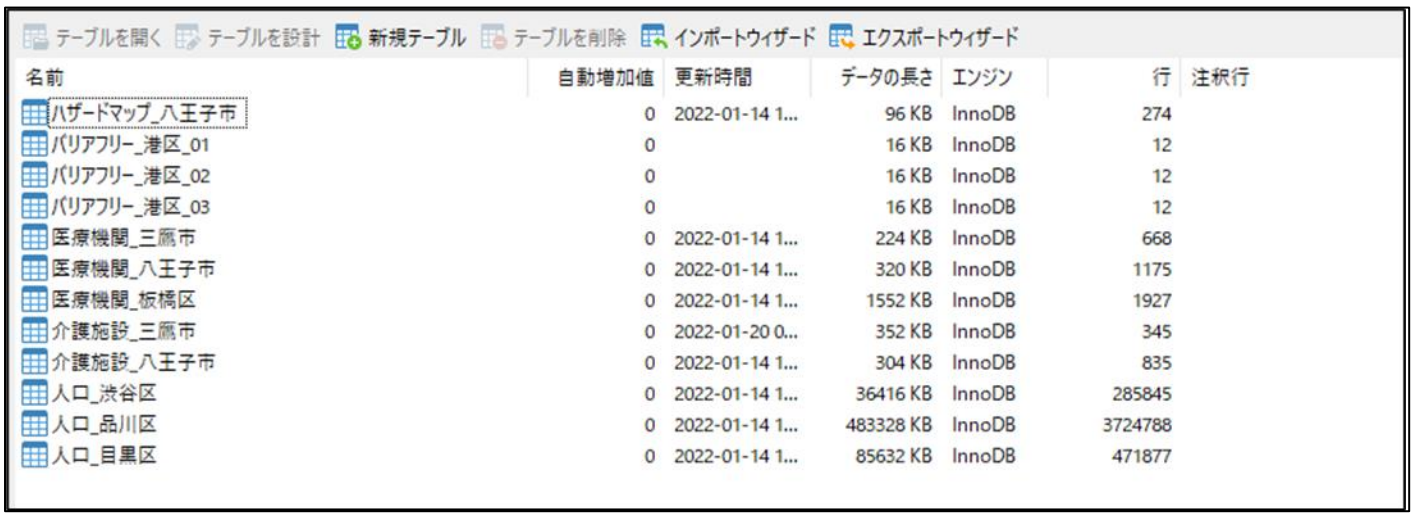
※格納率、一意、最大値、最小値、最大桁、最少桁等から、外れ値や作業ミスの簡易検出が可能です。

データの格納

整備対象データをデータベースに格納します。

参考ロジック

```
SELECT 項目 1、項目 2、項目 3、項目 4 FROM テーブル
```



名前	自動増加値	更新時間	データの長さ	エンジン	行	注釈行
ハザードマップ_八王子市	0	2022-01-14 1...	96 KB	InnoDB	274	
バリアフリー_港区_01	0		16 KB	InnoDB	12	
バリアフリー_港区_02	0		16 KB	InnoDB	12	
バリアフリー_港区_03	0		16 KB	InnoDB	12	
医療機関_三鷹市	0	2022-01-14 1...	224 KB	InnoDB	668	
医療機関_八王子市	0	2022-01-14 1...	320 KB	InnoDB	1175	
医療機関_板橋区	0	2022-01-14 1...	1552 KB	InnoDB	1927	
介護施設_三鷹市	0	2022-01-20 0...	352 KB	InnoDB	345	
介護施設_八王子市	0	2022-01-14 1...	304 KB	InnoDB	835	
人口_渋谷区	0	2022-01-14 1...	36416 KB	InnoDB	285845	
人口_品川区	0	2022-01-14 1...	483328 KB	InnoDB	3724788	
人口_目黒区	0	2022-01-14 1...	85632 KB	InnoDB	471877	

データの格納例（人口：目黒区）

区名	ファイル名	年月日	住民種別	地域コード	町丁名	年齢自	年齢至	属性	数量	備考	KEY_CODE	都道府県コード又は市区町村コー	都道府県名	属性_元	インデックス
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	0	999	総人口	3698	(Null)	13110001001	131105	東京都	0-999歳の総人口	1
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	0	999	男性	1824	(Null)	13110001001	131105	東京都	0-999歳の男性	2
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	0	999	女性	1874	(Null)	13110001001	131105	東京都	0-999歳の女性	3
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	0	4	男性	70	(Null)	13110001001	131105	東京都	0-4歳の男性	4
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	0	4	女性	54	(Null)	13110001001	131105	東京都	0-4歳の女性	5
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	5	9	男性	48	(Null)	13110001001	131105	東京都	5-9歳の男性	6
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	5	9	女性	56	(Null)	13110001001	131105	東京都	5-9歳の女性	7
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	10	14	男性	54	(Null)	13110001001	131105	東京都	10-14歳の男性	8
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	10	14	女性	50	(Null)	13110001001	131105	東京都	10-14歳の女性	9
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	15	19	男性	50	(Null)	13110001001	131105	東京都	15-19歳の男性	10
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	15	19	女性	47	(Null)	13110001001	131105	東京都	15-19歳の女性	11
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	20	24	男性	128	(Null)	13110001001	131105	東京都	20-24歳の男性	12
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	20	24	女性	100	(Null)	13110001001	131105	東京都	20-24歳の女性	13
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	25	29	男性	210	(Null)	13110001001	131105	東京都	25-29歳の男性	14
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	25	29	女性	173	(Null)	13110001001	131105	東京都	25-29歳の女性	15
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	30	34	男性	228	(Null)	13110001001	131105	東京都	30-34歳の男性	16
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	30	34	女性	162	(Null)	13110001001	131105	東京都	30-34歳の女性	17
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	35	39	男性	185	(Null)	13110001001	131105	東京都	35-39歳の男性	18
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	35	39	女性	168	(Null)	13110001001	131105	東京都	35-39歳の女性	19
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	40	44	男性	149	(Null)	13110001001	131105	東京都	40-44歳の男性	20
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	40	44	女性	135	(Null)	13110001001	131105	東京都	40-44歳の女性	21
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	45	49	男性	109	(Null)	13110001001	131105	東京都	45-49歳の男性	22
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	45	49	女性	141	(Null)	13110001001	131105	東京都	45-49歳の女性	23
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	50	54	男性	75	(Null)	13110001001	131105	東京都	50-54歳の男性	24
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	50	54	女性	109	(Null)	13110001001	131105	東京都	50-54歳の女性	25
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	55	59	男性	101	(Null)	13110001001	131105	東京都	55-59歳の男性	26
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	55	59	女性	102	(Null)	13110001001	131105	東京都	55-59歳の女性	27
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	60	64	男性	105	(Null)	13110001001	131105	東京都	60-64歳の男性	28
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	60	64	女性	133	(Null)	13110001001	131105	東京都	60-64歳の女性	29
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	65	69	男性	98	(Null)	13110001001	131105	東京都	65-69歳の男性	30
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	65	69	女性	110	(Null)	13110001001	131105	東京都	65-69歳の女性	31
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	70	74	男性	72	(Null)	13110001001	131105	東京都	70-74歳の男性	32
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	70	74	女性	85	(Null)	13110001001	131105	東京都	70-74歳の女性	33
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	75	79	男性	56	(Null)	13110001001	131105	東京都	75-79歳の男性	34
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	75	79	女性	86	(Null)	13110001001	131105	東京都	75-79歳の女性	35
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	80	84	男性	47	(Null)	13110001001	131105	東京都	80-84歳の男性	36
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	80	84	女性	73	(Null)	13110001001	131105	東京都	80-84歳の女性	37
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	85	999	男性	39	(Null)	13110001001	131105	東京都	85-999歳の男性	38
目黒区	131105_population_20110101.	2011/1/1	(Null)	111	駒場一丁目	85	999	女性	90	(Null)	13110001001	131105	東京都	85-999歳の女性	39

データの格納例（医療機関：八王子）

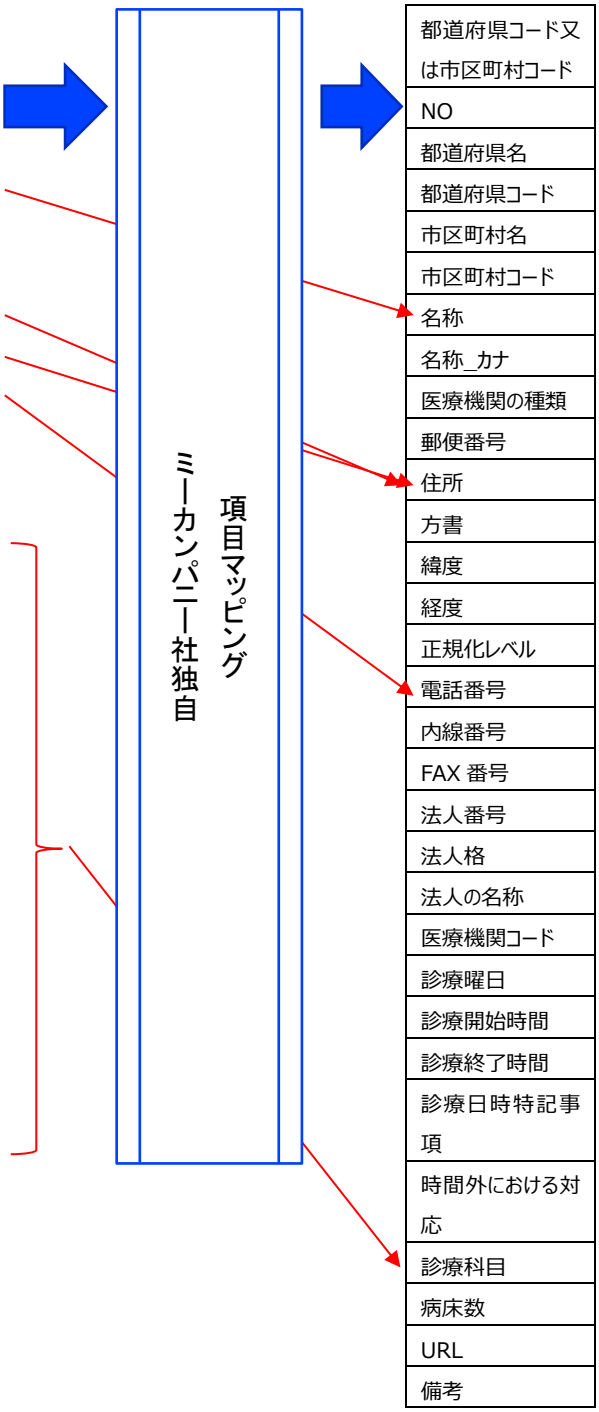
データのマッピング

データのフォーマット統一化

出典各様式を、共通フォーマットへ統一する作業です。

例 ： 医療機関(八王子_医療機関)

インデックス	1	2
年月日	44166	44166
区名	三鷹市	三鷹市
ファイル名	○○○.xlsx	□□□.xlsx
シート名	病院	病院
No	1	2
施設名称	A 病院	B 病院
施設所在地	三鷹市△△	三鷹市▽▽
施設方書		
施設ＴＥＬ	0422-47-????	0422-47-????
施設ＴＥＬ１		
備考		
開設届	44166	44166
診療科目名称.1	内科	内科
診療科目名称.2	麻酔科	外科
診療科目名称.3	泌尿器科	整形外科
診療科目名称.4	放射線科	眼科
診療科目名称.5	循環器内科	リハビリテーション科
診療科目名称.6	病理診断科	泌尿器科
診療科目名称.7	婦人科	
診療科目名称.8		
診療科目名称.9		
・		
・		
診療科目名称.27		
診療科目名称.28	1	2



参考ロジック

```
select
   市区町村コード_デジット as 都道府県コード又は市区町村コード
 , `インデックス` as NO
 , '    東京都' as 都道府県名
 , 13 as 都道府県コード
 , `区名` as 市区町村名
 , 市区町村コード as 市区町村コード
 , `施設名称` as 名称
 , null as 名称_カナ
 , null as 医療機関の種類
 , 郵便番号 as 郵便番号
 , `施設所在地` as 住所
 , null as 方書
 , null as 緯度
 , null as 経度
 , null as 正規化レベル
 , case when `施設ＴＥＬ` IS not null then `施設ＴＥＬ` else `施設ＴＥＬ１` end as 電話番号
 , null as 内線番号
 , null as FAX 番号
 , null as 法人番号
 , null as 法人格
 , null as 法人の名称
 , null as 医療機関コード
 , null as 診療曜日
 , null as 診療開始時間
 , null as 診療終了時間
 , null as 診療日時特記事項
 , null as 時間外における対応
 , concat_ws('|', `診療科目名称.1`, `診療科目名称.2`, `診療科目名称.3`, `診療科目名称.4`, `診療科目名称.5`, `診療科目名称.6`, `診療科目名称.7`, `診療科目名称.8`, `診療科目名称.9`, `診療科目名称.10`,
                    `診療科目名称.11`, `診療科目名称.12`, `診療科目名称.13`, `診療科目名称.14`, `診療科目名称.15`, `診療科目名称.16`, `診療科目名称.17`, `診療科目名称.18`, `診療科目名称.19`, `診療科目名称.20`,
                    `診療科目名称.21`, `診療科目名称.22`, `診療科目名称.23`, `診療科目名称.24`, `診療科目名称.25`, `診療科目名称.26`, `診療科目名称.27`, `診療科目名称.28`
      ) as 診療科目
 , null as 病床数
 , null as URL
 , `備考` as 備考
   from `医療機関_三鷹市`
```

データのクレンジング

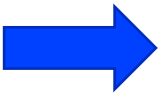
クレンジング難度が高い例（データ構造の課題）

「1 セル 1 データ」となっていない。

営業日・営業時間の表記ゆれ

営業時間
窓口受付時間
受付営業時間
サービス提供時間
通いサービス提供時間
訪問可能時間帯
24 時間対応
休業日
窓口定休日
受付休業日
通いサービスの定休日

マッピング難度が高い



推奨データセット

介護事務所営業時間項目

利用可能曜日
利用可能曜日特記事項

経年でデータが変わる

平成31年3月1日現在

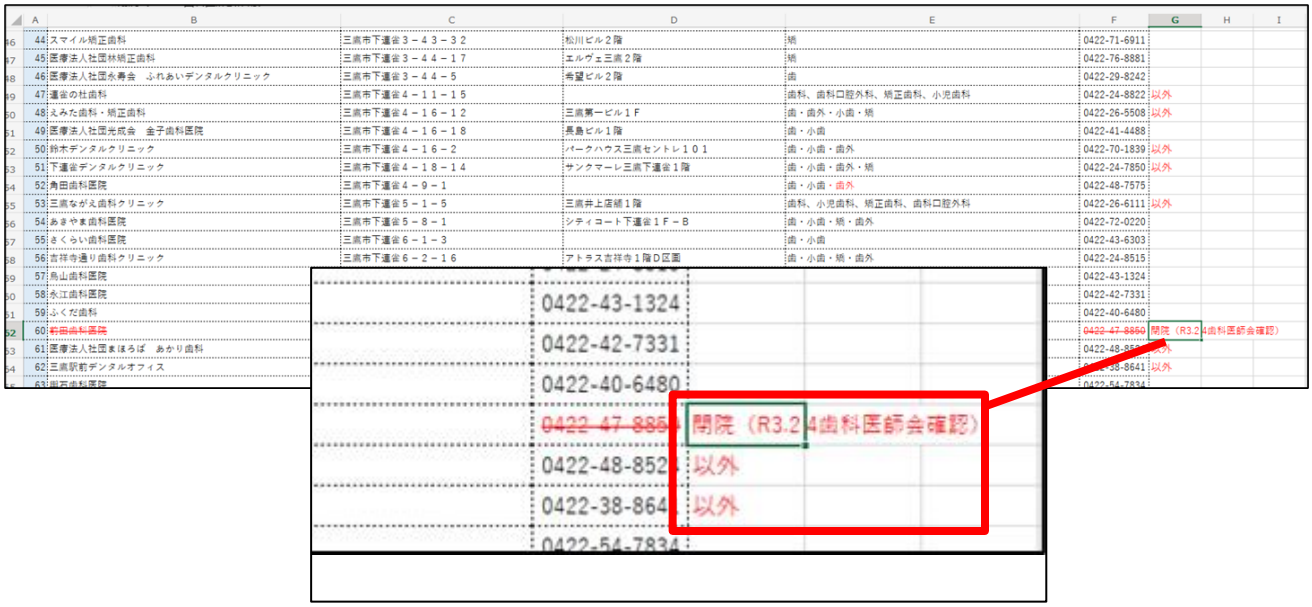
年齢	総数	男	女
総数	258,093	121,704	136,389
0歳～4歳	14,651	7,544	7,107
0歳	2,876	1,501	1,375
1歳	2,958	1,509	1,449
2歳	2,987	1,505	1,482
3歳	3,025	1,601	1,424
4歳	2,805	1,428	1,377
5歳～9歳	11,629	5,877	5,752
5歳	2,513	1,256	1,257
6歳	2,540	1,261	1,279
7歳	2,289	1,169	1,120
8歳	2,192	1,114	1,078
90歳～94歳	2,113	512	1,601
90歳	607	149	458
91歳	473	114	359
92歳	394	103	291
93歳	340	66	274
94歳	299	80	219
95歳～99歳	690	117	573
100歳以上	116	17	99

平成31年4月1日現在

年齢	総数	男	女
総数	258,696	121,989	136,707
0歳～4歳	14,516	7,473	7,043
0歳	2,802	1,457	1,345
1歳	2,954	1,519	1,435
2歳	2,976	1,498	1,478
3歳	2,968	1,567	1,401
4歳	2,816	1,432	1,384
5歳～9歳	11,622	5,866	5,756
5歳	2,533	1,263	1,270
6歳	2,514	1,240	1,274
7歳	2,268	1,154	1,114
8歳	2,216	1,132	1,084
95歳～99歳	690	118	572
95歳	242	44	198
96歳	162	21	141
97歳	131	29	102
98歳	97	12	85
99歳	58	12	46
100歳以上	127	19	108
100歳	48	9	39
101歳	31	4	27
102歳	14	4	10
103歳以上	34	2	32

2019 年以降は 100 歳以上も 1 歳刻みで表示されている。

枠外にデータが表記されている

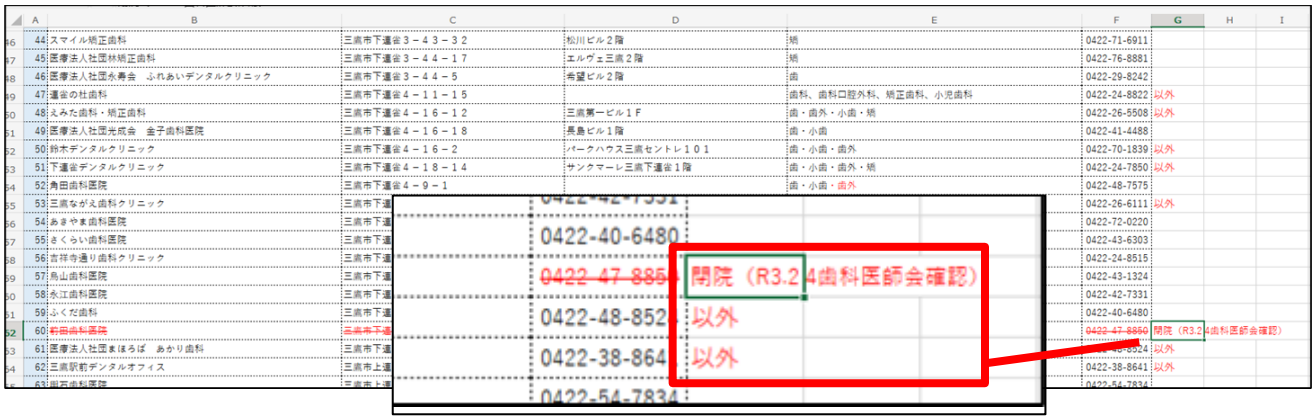


クレンジング難度が高い例（データの課題）

ここでは単純なデータクレンジング作業では対応できず、目視を伴う作業が発生するケースを記載しています。

以下のようなケースでは、100%自動化は難しく、目視作業が必要になります。作業を依頼する場合には、こういった作業が発生する事を念頭に依頼を行ってください。

枠外に閉院データが表記されている

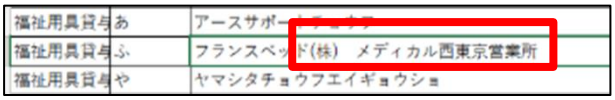


打ち消し線でデータが修正されている

打ち消し線　　赤文字　　「以外」表記

異表記がデータに含まれる

フリガナに「漢字」が含まれる　　電話番号以外の表記



凡例のない赤字背景のタブ

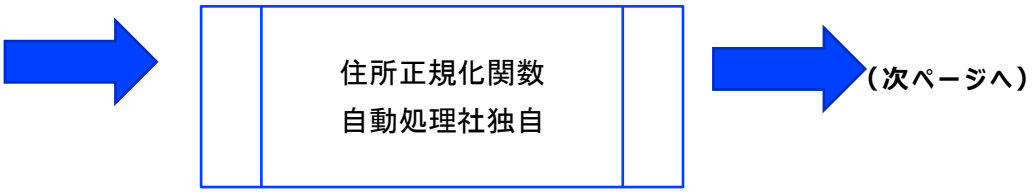
データの正規化

住所

データ最終化前の、文字欠け、文字化け、表記揺れ対応、対応方針の正確性確認、不正値検知等のために必要な作業です。

例（住所：八王子）

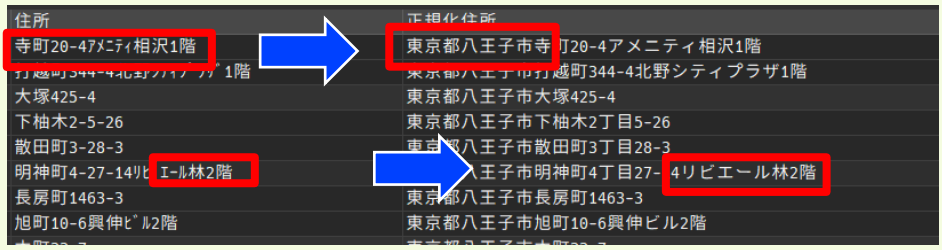
住所
寺町20-4ｱﾒﾆﾃｨ相沢1階
打越町344-4北野ｼﾃｨﾌﾟﾗｻﾞ1階
大塚425-4
下柚木2-5-26
散田町3-28-3
明神町4-27-14ﾘﾋﾞｴｰﾙ林2階
長房町1463-3
旭町10-6興伸ﾋﾞﾙ2階
本町33-7
大塚622-9ﾆｭｰﾊｲﾑ井上201号
高尾町1602
元横山町2-1-20
狭間町1450-1めじろ台ｺｰﾎﾟﾗｽ
みなみ野3-9-7
八日町4-9幸ﾋﾞﾙ3階
長沼町200-3旗野ｺｰﾎﾟﾗｽ1階
片倉町443-4
子安町4-6-1Phitﾋﾞﾙ3階
子安町4-6-1Phitﾋﾞﾙ2階
明神町3-26-10土屋ﾋﾞﾙ1階
東浅川町879-6
子安町1-2-8守屋ﾋﾞﾙ3階
元本郷町3-16-7
中野上町5-5-3
松木14-3ｻﾞ・ﾃﾗｽ松木103号
高月町365-1　2階
子安町4-21-7
横山町8-19牛久保ﾋﾞﾙ3階
久保山町2-43-4
大和田町6-5-17
南大沢2-27ﾚｽｺ南大沢4階
横山町11-4井藤ﾋﾞﾙ4階FC
別所1-17-1ｸﾚｾﾝﾄﾋﾙｽﾞ2階-B
日吉町13-31
北野町546-6小山ﾋﾞﾙ2階
長房町375
鹿島6
東浅川町1028
本町3-4
めじろ台2-49-19
子安町1-4-5
大和田町4-11-19
片倉町492-1綱木ﾋﾞﾙ1階
明神町4-12-2
川口町1088-4
北野台4-27-2
高倉町48-2高橋ﾋﾞﾙ2階
北野町555-17
散田町5-1-6
横川町536-3
打越町1515



住所正規化関数の役割

SQL 関数にて郵便番号を追加します。
住所を都道府県、市区町村を補完し、市区町村合併を反映、緯度経度を付与します。
※住所については日本に標準となるものがない為、100%精度の適切な住所が存在しません。その為、住所の正規化は非常に難度の高い作業になります。また日本全国の建物や番地に緯度経度がついていない為、緯度経度を 100%つける事はほぼ不可能です。緯度経度を公開する際には、正規化精度を必ず一緒につけて公開するようにしてください。
※Google Place API など緯度経度の情報を安く取得できる方法はありますが、ほとんどの場合は、取得したデータをそのまま公開してはいけない利用規約になっており、公開すると利用規約違反になってしまう事が多く、莫大な損害賠償を請求される可能性があります。緯度経度をオープンデータとして公開する際には、許諾を適切に取ったデータのみ公開するようにしてください。

例：住所正規化



※前ページからの続き

元情報住所　　表記揺れ統一後の正規化データ　　元住所

元住所	正規化住所結果JSON	@json_normalize_string_wk :=	正規化住所	@json_normalize_string_wk :=
東京都寺町20-4ｱﾒﾆﾃｨ相沢1階	esults": {"id": 12428729, "	{"results": {"normalize_string"	東京都八王子市寺町20-4アメニティ相沢1階	{"results": {"normalize_string": "
東京都打越町344-4北野ｼﾃｨﾌﾞﾗｻﾞ1階	esults": {"id": 12375697, "	{"results": {"normalize_string"	東京都八王子市打越町344-4北野シティプラザ1階	{"results": {"normalize_string": "
東京都大塚425-4	esults": {"id": 12380392, "	{"results": {"normalize_string"	東京都八王子市大塚425-4	{"results": {"normalize_string": "
東京都下柚木2-5-26	esults": {"id": 12416142, "	{"results": {"normalize_string"	東京都八王子市2-5-26	{"results": {"normalize_string": "
東京都散田町3-28-3	esults": {"id": 12411251, "	{"results": {"normalize_string"	東京都八王子市3-28-3	{"results": {"normalize_string": "
東京都明神町4-27-14ﾘﾋﾞｴｰﾙ林2階	esults": {"id": 12462204, "	{"results": {"normalize_string"	東京都八王子市4-27-14リビエール林2階	{"results": {"normalize_string": "
東京都長房町1463-3	esults": {"id": 12437918, "	{"results": {"normalize_string"	東京都八王子市長房町1463-3	{"results": {"normalize_string": "
東京都旭町10-6興伸ﾋﾞﾙ2階	esults": {"id": 12367358, "	{"results": {"normalize_string"	東京都八王子市旭町10-6興伸ビル2階	{"results": {"normalize_string": "
東京都本町33-7	esults": {"id": 0, "area": "",	{"results": {"normalize_string"	東京都本町33-7	{}
東京都大塚622-9ﾆｭｰﾊｲﾑ井上201	esults": {"id": 12379784, "	{"results": {"normalize_string"	東京都八王子市大塚622-9ニューハイム井上201	{"results": {"normalize_string": "
東京都高尾町1602	esults": {"id": 12419282, "	{"results": {"normalize_string"	東京都八王子市高尾町1602	{"results": {"normalize_string": "
東京都元横山町2-1-20	esults": {"id": 12470417, "	{"results": {"normalize_string"	東京都八王子市町2-1-20	{"results": {"normalize_string": "
東京都狭間町1450-1めじろ台ｺｰﾎﾟﾗｽ	esults": {"id": 12446314, "	{"results": {"normalize_string"	東京都八王子市狭間町1450-1めじろ台コーポラス	{"results": {"normalize_string": "
東京都みなみ野3-9-7	esults": {"id": 12489190, "	{"results": {"normalize_string"	東京都八王子市野3-9-7	{"results": {"normalize_string": "
東京都八日町4-9幸ﾋﾞﾙ3階	esults": {"id": 12473681, "	{"results": {"normalize_string"	東京都八王子市八日町4-9幸ビル3階	{"results": {"normalize_string": "
東京都長沼町200-3旗野ｺｰﾎﾟﾗｽ1階	esults": {"id": 12434268, "	{"results": {"normalize_string"	東京都八王子市長沼町200-3旗野コーポラス1階	{"results": {"normalize_string": "
東京都片倉町443-4	esults": {"id": 12388043, "	{"results": {"normalize_string"	東京都八王子市片倉町443-4	{"results": {"normalize_string": "
東京都子安町4-6-1Phiﾋﾞﾙ3階	esults": {"id": 12408873, "	{"results": {"normalize_string"	東京都八王子市4-6-1Phiビル3階	{"results": {"normalize_string": "
東京都子安町4-6-1Phiﾋﾞﾙ2階	esults": {"id": 12408873, "	{"results": {"normalize_string"	東京都八王子市4-6-1Phiビル2階	{"results": {"normalize_string": "
東京都明神町3-26-10土屋ﾋﾞﾙ1階	esults": {"id": 12462013, "	{"results": {"normalize_string"	東京都八王子市3-26-10土屋ビル1階	{"results": {"normalize_string": "
東京都東浅川町879-6	esults": {"id":	ormalize_string"	東京都八王子市東浅川町879-6	{"results": {"normalize_string": "
東京都子安町1-2-8守屋ﾋﾞﾙ3階	esults": {"id":	{"results": {"normalize_string"	東京都八王子市1-2-8守屋ビル3階	{"results": {"normalize_string": "
東京都元本郷町3-16-7	esults": {"id": 12469867, "	{"results": {"normalize_string"	東京都八王子市町3-16-7	{"results": {"normalize_string": "
東京都中野上町5-5-3	esults": {"id": 0, "area": "",	{"results": {"normalize_string"	東京都中野区上町5-5-3	{}
東京都松木14-3ｻﾞ・ﾃﾗｽ松木103	esults": {"id": 12456134, "	{"results": {"normalize_string"	東京都八王子市松木14-3ザ・テラス松木103	{"results": {"normalize_string": "
東京都高月町365-1 2階	esults": {"id": 12420849, "	{"results": {"normalize_string"	東京都八王子市高月町365-1-2階	{"results": {"normalize_string": "
東京都子安町4-21-7	esults": {"id": 12409000, "	{"results": {"normalize_string"	東京都八王子市4-21-7	{"results": {"normalize_string": "
東京都横山町8-19牛久保ﾋﾞﾙ3階	esults": {"id": 12476628, "	{"results": {"normalize_string"	東京都八王子市横山町8-19牛久保ビル3階	{"results": {"normalize_string": "
東京都久保山町2-43-4	esults": {"id": 12483042, "	{"results": {"normalize_string"	東京都八王子市町2-43-4	{"results": {"normalize_string": "
東京都大和田町6-5-17	esults": {"id": 12384179, "	{"results": {"normalize_string"	東京都八王子市町6-5-17	{"results": {"normalize_string": "
東京都南大沢2-27ﾌﾚｽｺ南大沢4階	esults": {"id": 12458559, "	{"results": {"normalize_string"	東京都八王子市2-27フレスコ南大沢4階	{"results": {"normalize_string": "
東京都横山町11-4井藤ﾋﾞﾙ4階FC	esults": {"id": 12476589, "	{"results": {"normalize_string"	東京都八王子市横山町11-4井藤ビル4階FC	{"results": {"normalize_string": "
東京都別所1-17-1ｸﾚｾﾝﾄﾋﾙｽﾞ2階-B	esults": {"id": 12452677, "	{"results": {"normalize_string"	東京都八王子市−17-1クレセントヒルズ2階−B	{"results": {"normalize_string": "
東京都日吉町13-31	esults": {"id": 12450910, "	{"results": {"normalize_string"	東京都八王子市日吉町13-31	{"results": {"normalize_string": "
東京都北野町546-6小山ﾋﾞﾙ2階	esults": {"id": 12399888, "	{"results": {"normalize_string"	東京都八王子市北野町546-6小山ビル2階	{"results": {"normalize_string": "
東京都長房町375	esults": {"id": 12436119, "	{"results": {"normalize_string"	東京都八王子市長房町375	{"results": {"normalize_string": "
東京都鹿島6	esults": {"id": 12385230, "	{"results": {"normalize_string"	東京都八王子市鹿島6	{"results": {"normalize_string": "
東京都東浅川町1028	esults": {"id": 12448064, "	{"results": {"normalize_string"	東京都八王子市東浅川町1028	{"results": {"normalize_string": "
東京都本町3-4	esults": {"id": 12455033, "	{"results": {"normalize_string"	東京都八王子市本町3-4	{"results": {"normalize_string": "
東京都めじろ台2-49-19	esults": {"id": 12463627, "	{"results": {"normalize_string"	東京都八王子市台2-49-19	{"results": {"normalize_string": "
東京都子安町1-4-5	esults": {"id": 12407438, "	{"results": {"normalize_string"	東京都八王子市1-4-5	{"results": {"normalize_string": "
東京都大和田町4-11-19	esults": {"id": 12383556, "	{"results": {"normalize_string"	東京都八王子市町4-11-19	{"results": {"normalize_string": "
東京都片倉町492-1綱木ﾋﾞﾙ1階	esults": {"id": 12386681, "	{"results": {"normalize_string"	東京都八王子市片倉町492-1綱木ビル1階	{"results": {"normalize_string": "
東京都明神町4-12-2	esults": {"id": 12462095, "	{"results": {"normalize_string"	東京都八王子市4-12-2	{"results": {"normalize_string": "
東京都川口町1088-4	esults": {"id": 12394063, "	{"results": {"normalize_string"	東京都八王子市川口町1088-4	{"results": {"normalize_string": "
東京都北野台4-27-2	esults": {"id": 12485392, "	{"results": {"normalize_string"	東京都八王子市4-27-2	{"results": {"normalize_string": "
東京都高倉町48-2高橋ﾋﾞﾙ2階	esults": {"id": 12420541, "	{"results": {"normalize_string"	東京都八王子市高倉町48-2高橋ビル2階	{"results": {"normalize_string": "
東京都北野町555-17	esults": {"id": 12400354, "	{"results": {"normalize_string"	東京都八王子市北野町555-17	{"results": {"normalize_string": "
東京都散田町5-1-6	esults": {"id": 12412005, "	{"results": {"normalize_string"	東京都八王子市5-1-6	{"results": {"normalize_string": "
東京都横川町536-3	esults": {"id": 12474858, "	{"results": {"normalize_string"	東京都八王子市横川町536-3	{"results": {"normalize_string": "

「表記揺れ統一後の正規化データ」では、先頭の「都道府県」は、前処理にて付与されます。
元住所に東京都が記載されており、「東京都東京都板橋区」となっても問題は発生しません。
※余計な空白は削除（全半角統一等）されます。

参考

「元住所」では、郵便番号マスタ比較にて、有効であるか判別ができます。

事業所名（文字列）

データ最終化前の、表記揺れ対応、対応方針の正確性確認、不正値検知等のために必要な作業です。

例（事業所名：三鷹市　介護施設）

事業所名
SOMPOケア　三鷹下連雀　居宅介護支援
下連雀ケアセンター
NPO・ACTたま居宅介護支援事業所
テルウェル東日本　多摩東ケアプランセンタ
東京海上日動みずたま介護ST三鷹ケアプランセンター
居宅介護支援事業所太郎
やさしい手三鷹居宅介護支援事業所
弘済ケアセンター
居宅介護支援事業所　菜の花
ケアプランみんなの家
(医財)慈生会　野村訪問看護ステーション
三鷹中央リハケア居宅介護支援事業所
あじさい居宅介護支援事業所
高齢者在宅サービスセンター恵比寿苑
ケアプランはちどり
はなかいどう指定居宅介護支援事業所
介護サービス西野
みのり屋
やさしい手三鷹北野居宅介護支援事業所
ケアプランとんぼ
ケアプランみずほ
トータス長寿ケアセンター
エバーグリーン三鷹新川ヴィレッジ
嗚呼、ケアプラン介護隊
居宅介護支援事業所おしどり
ケアサポートきんぎょ
三鷹市社会福祉協議会　居宅介護支援事業所
ラ　ヴィータ三鷹
ソフィアケアプラン三鷹
ケアプラン　なのはな
(株)三鷹ナース・ヘルパーセンター
ツクイ三鷹
オレンジハート 指定居宅介護支援事業所
クローベルケア
泉のほとり
けやき苑指定居宅介護支援事業所
長谷川病院居宅介護支援事業所
青い空
ケアプランたんぽぽ
ケアプランきづな
ラ　ヴィータ吉祥寺
スカイリフレ介護サービス
日介センター吉祥寺
居宅介護支援事業所　吉祥寺南
ムウケア武蔵野
稲垣薬局介護サービス
アクセスポイント吉祥寺ケアプラン
ヘルパーステーション金太郎
SOMPOケア　武蔵野　居宅介護支援
東電さわやかケア武蔵野・居宅介護支援

（次ページへ）

文字列正規化関数の役割

SQL 関数にて、名称を一括で英数字を半角、ひらがなカタカナを全角に変換、漢字を縮退マップに基づいて縮退します。

例：文字列正規化

事業所名	正規化済み事業所名
相沢歯科医院	相沢歯科医院
青木歯科医院	青木歯科医院
あきいけ歯科医院	あきいけ歯科医院
あさかわ歯科医院	あさかわ歯科医院
天野歯科医院	天野歯科医院
彩デンタルｸﾘﾆｯｸ	ンタルクリニック
新井歯科医院	新井歯科医院

出所：三鷹市・三鷹市介護保険事業者連絡協議会「介護サービス事業者ガイドブック」

※前ページからの続き

元事業所名	表記揺れ統一後の正規化データ
事業所名	results
SOMPOケア　三鷹下連雀　居宅介護支援	SOMPOケア　三鷹下連雀　居宅介護支援
下連雀ケアセンター	下連雀ケアセンター
NPO・ACTたま居宅介護支援事業所	特定非営利活動法人・ACTたま居宅介護支援事業所
テルウェル東日本　多摩東ケアプランセンタ	テルウェル東日本　多摩東ケアプランセンタ
東京海上日動みずたま介護ST三鷹ケアプランセンター	東京海上日動みずたま介護ST三鷹ケアプランセンター
居宅介護支援事業所太郎	居宅介護支援事業所太郎
やさしい手三鷹居宅介護支援事業所	やさしい手三鷹居宅介護支援事業所
弘済ケアセンター	弘済ケアセンター
居宅介護支援事業所　菜の花	居宅介護支援事業所　菜の花
ケアプランみんなの家	ケアプランみんなの家
(医財)慈生会　野村訪問看護ステーション	医療法人財団慈生会　野村訪問看護ステーション
三鷹中央リハケア居宅介護支援事業所	三鷹中央リハケア居宅介護支援事業所
あじさい居宅介護支援事業所	あじさい居宅介護支援事業所
高齢者在宅サービスセンター恵比寿苑	高齢者在宅サービスセンター恵比寿苑
ケアプランはちどり	ケアプランはちどり
はなかいどう指定居宅介護支援事業所	はなかいどう指定居宅介護支援事業所
介護サービス西野	介護サービス西野
みのり屋	みのり屋
やさしい手三鷹北野居宅介護支援事業所	やさしい手三鷹北野居宅介護支援事業所
ケアプランとんぼ	ケアプランとんぼ
ケアプランみずほ	ケアプランみずほ
トータス長寿ケアセンター	トータス長寿ケアセンター
エバーグリーン三鷹新川ヴィレッジ	エバーグリーン三鷹新川ヴィレッジ
鳴呼、ケアプラン介護隊	鳴呼、ケアプラン介護隊
居宅介護支援事業所おしどり	居宅介護支援事業所おしどり
ケアサポートきんぎょ	ケアサポートきんぎょ
三鷹市社会福祉協議会　居宅介護支援事業所	三鷹市社会福祉協議会　居宅介護支援事業所
ラ　ヴィータ三鷹	ラ　ヴィータ三鷹
ソフィアケアプラン三鷹	ソフィアケアプラン三鷹
ケアプラン　なのはな	ケアプラン　なのはな
(株)三鷹ナース・ヘルパーセンター	株式会社三鷹ナース・ヘルパーセンター
ツクイ三鷹	ツクイ三鷹
オレンジハート 指定居宅介護支援事業所	オレンジハート　指定居宅介護支援事業所
クローベルケア	クローベルケア
泉のほとり	泉のほとり
けやき苑指定居宅介護支援事業所	けやき苑指定居宅介護支援事業所
長谷川病院居宅介護支援事業所	長谷川病院居宅介護支援事業所
青い空	青い空
ケアプランたんぽぽ	ケアプランたんぽぽ
ケアプランきづな	ケアプランきづな
ラ　ヴィータ吉祥寺	ラ　ヴィータ吉祥寺
スカイリフレ介護サービス	スカイリフレ介護サービス

出所：三鷹市・三鷹市介護保険事業者連絡協議会「介護サービス事業者ガイドブック」

「表記揺れ統一後の正規化データ」では、法人格表記統一、スペース統一、記号全半角統一等、正規化内容の詳細な制御が可能です。

例 ： 法人格は削除等

電話番号

データ最終化前の、表記揺れ対応、対応方針の正確性確認、不正値検知等のために必要な作業です。

例（電話番号：八王子　医療機関）

電話番号
626-4188
648-3566
677-0480
679-1188
661-0249
645-3116
661-6500
644-3662
623-2052
676-8288
661-0128
642-1827
664-5862
632-7676
625-5484
636-4618
637-2622
625-7520
625-7142
646-9939
664-5901
656-1631
625-6491
626-3944
675-4182
696-5373
622-1427
646-6547
691-7100
644-8852
677-5383
646-5497
679-0980
622-4019
646-3918
661-4552
676-2642
669-7666
622-0142
663-2421
645-4018
646-1010
637-1324
642-5652
654-7418

電話番号正規化関数
自動処理社独自

（次ページへ）

電話番号正規化関数の役割

SQL 関数にて、電話番号は住所を元に市外局番を補完し、行政基本情報データ連携モデル準拠に変換します。

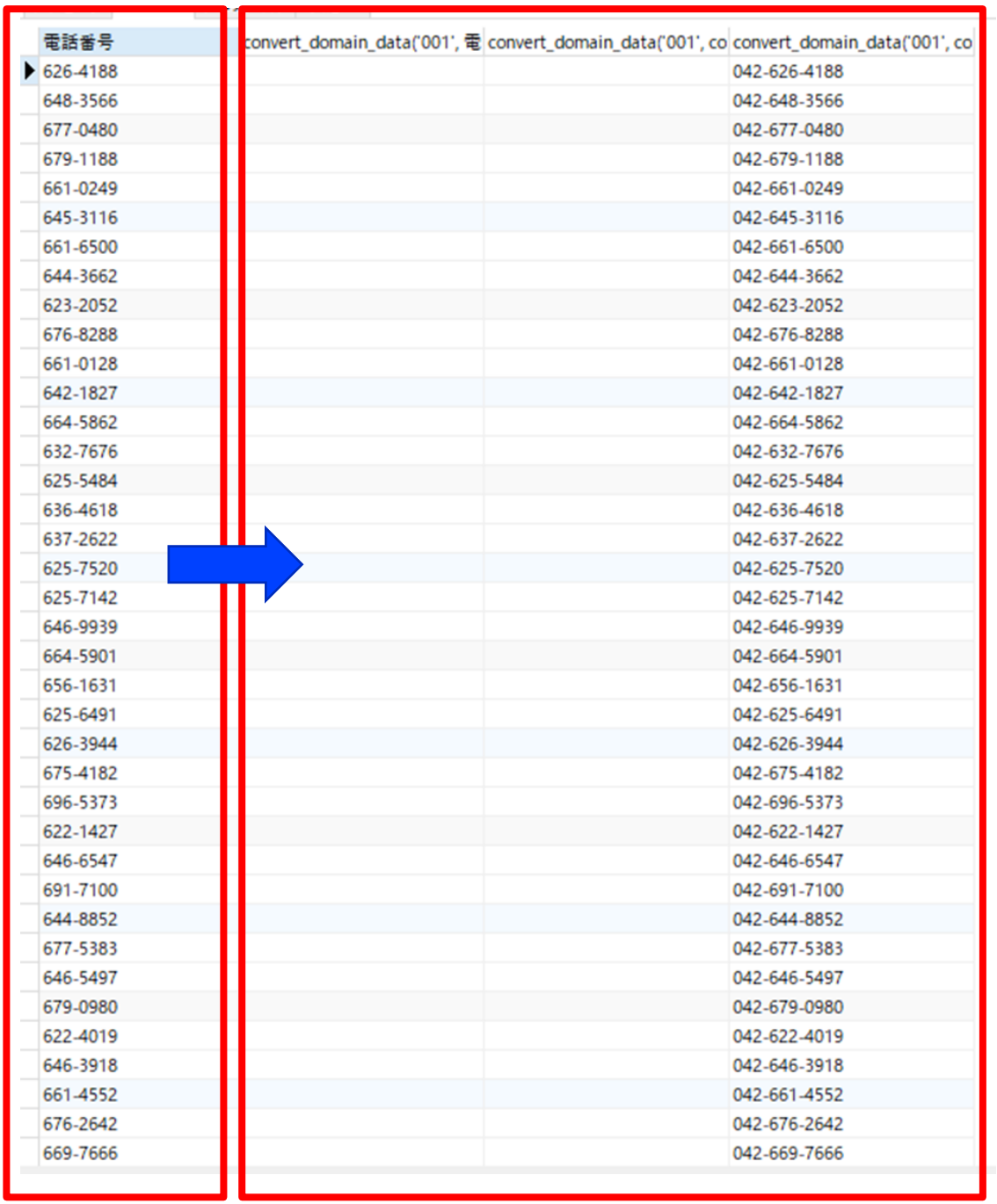
例：電話番号正規化

一部地域で 03 以外の番号を付与しています。

- 三鷹市中原一丁は 03、それ以外は 0422
- 八王子市は 042
  ※2006/3/5 までは 0426

※前ページからの続き

電話番号　　**表記揺れ統一後の正規化データ**



電話番号	convert_domain_data('001', 電	convert_domain_data('001', co	convert_domain_data('001', co
626-4188			042-626-4188
648-3566			042-648-3566
677-0480			042-677-0480
679-1188			042-679-1188
661-0249			042-661-0249
645-3116			042-645-3116
661-6500			042-661-6500
644-3662			042-644-3662
623-2052			042-623-2052
676-8288			042-676-8288
661-0128			042-661-0128
642-1827			042-642-1827
664-5862			042-664-5862
632-7676			042-632-7676
625-5484			042-625-5484
636-4618			042-636-4618
637-2622			042-637-2622
625-7520			042-625-7520
625-7142			042-625-7142
646-9939			042-646-9939
664-5901			042-664-5901
656-1631			042-656-1631
625-6491			042-625-6491
626-3944			042-626-3944
675-4182			042-675-4182
696-5373			042-696-5373
622-1427			042-622-1427
646-6547			042-646-6547
691-7100			042-691-7100
644-8852			042-644-8852
677-5383			042-677-5383
646-5497			042-646-5497
679-0980			042-679-0980
622-4019			042-622-4019
646-3918			042-646-3918
661-4552			042-661-4552
676-2642			042-676-2642
669-7666			042-669-7666

「表記揺れ統一後の正規化データ」で、値が空欄な箇所は、元データが電話番号として無効な場合です。（桁不足等）
赤枠内 1 列目 ： 元データを電話番号正規化した場合
赤枠内 2 列目 ： 元データ先頭に 03 を付記して、電話番号正規化した場合
赤枠内 3 列目 ： 元データ先頭に 042 を付記して、電話番号正規化した場合

手順 4

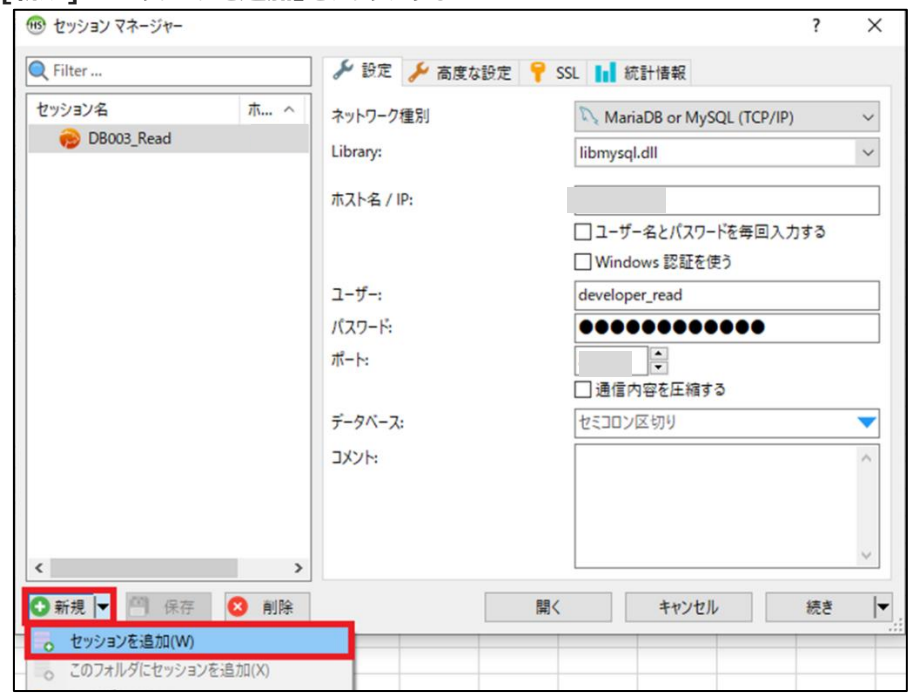
データの保存（エクスポート）

データベース接続の設定

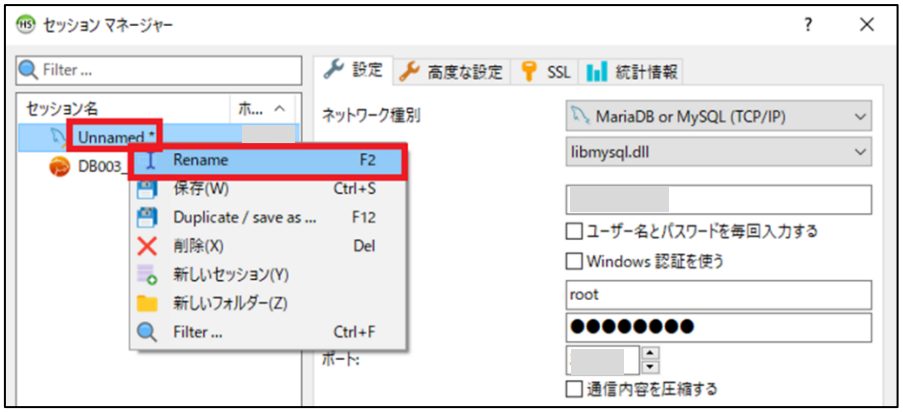
ここでは、HeidiSQL を用いて、データベースに接続する設定を行います。

任意のデータベース接続アプリケーションを使用される場合は、説明文をお使いのアプリケーションに適宜読み替えてご参照ください。

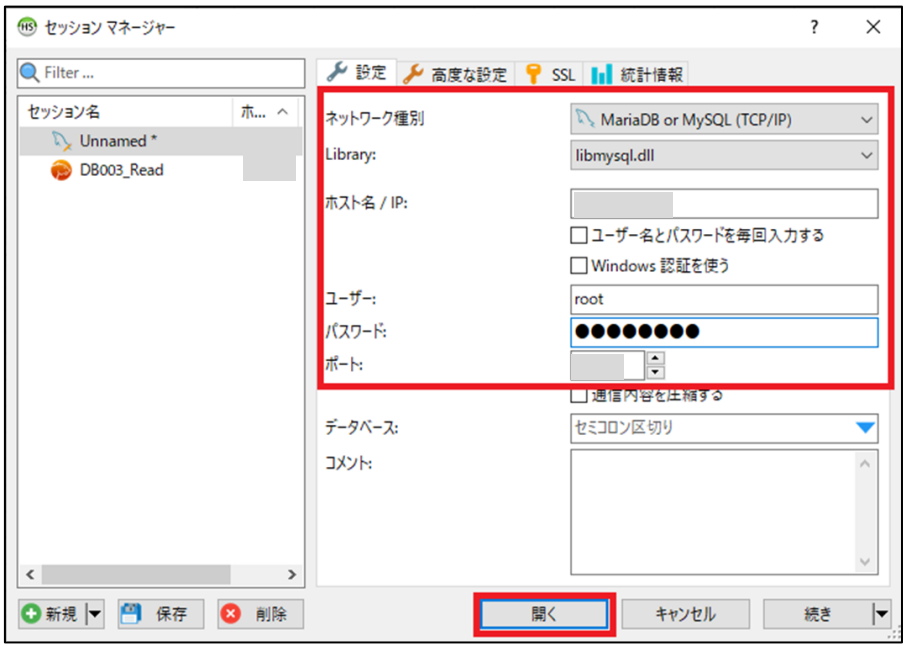
1. **HeidiSQL 等、データベースに接続できるアプリケーションを開く**
2. **[新規]→「セッションを追加」をクリックする**



3. **セッション名を右クリック、「Rename」をクリックした後、任意の名前に変更する**



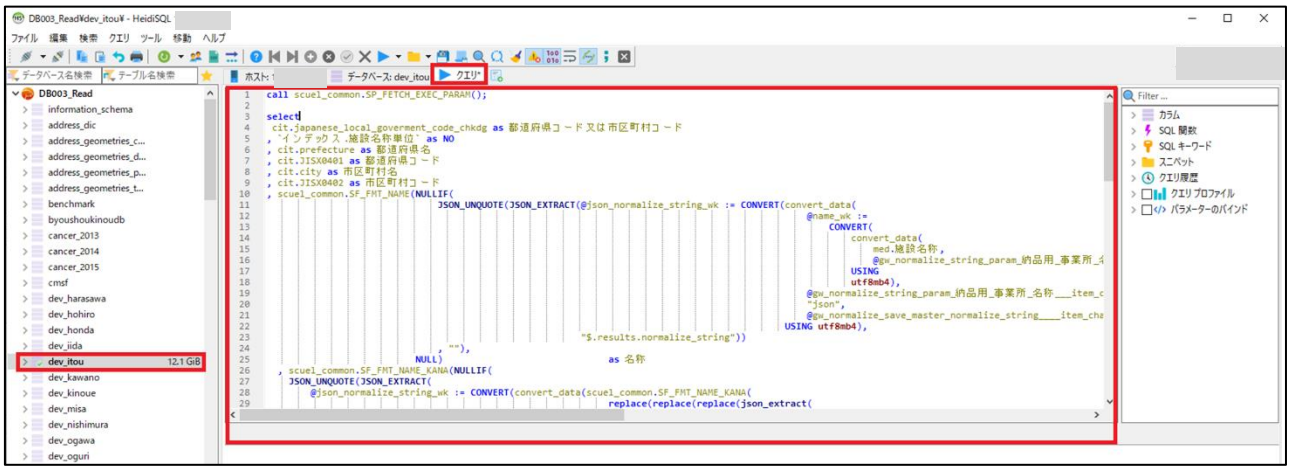
4. 「設定」タブをクリック、設定を変更した後、[開く]をクリックする



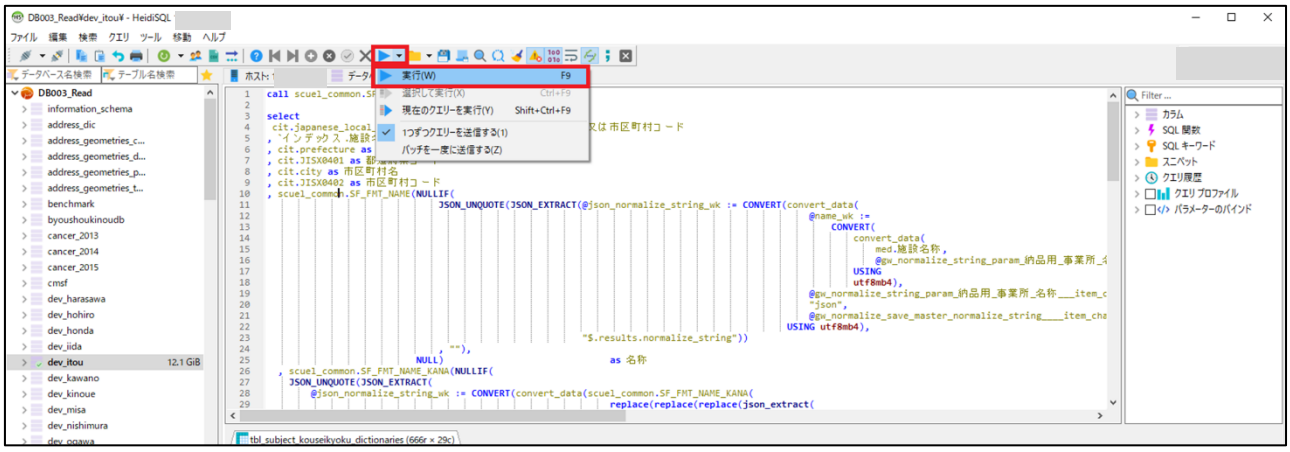
ネットワーク種別： MariaDB or MySQL(TCP/IP)を選択
Library: libmysql.dll を選択
ホスト名/IP:
接続先 DB のホスト名/IP を入力
ユーザー：
接続先 DB のユーザーを入力
パスワード：
接続先 DB のパスワードを入力
ポート：
接続先 DB のポートを入力

データのエクスポート

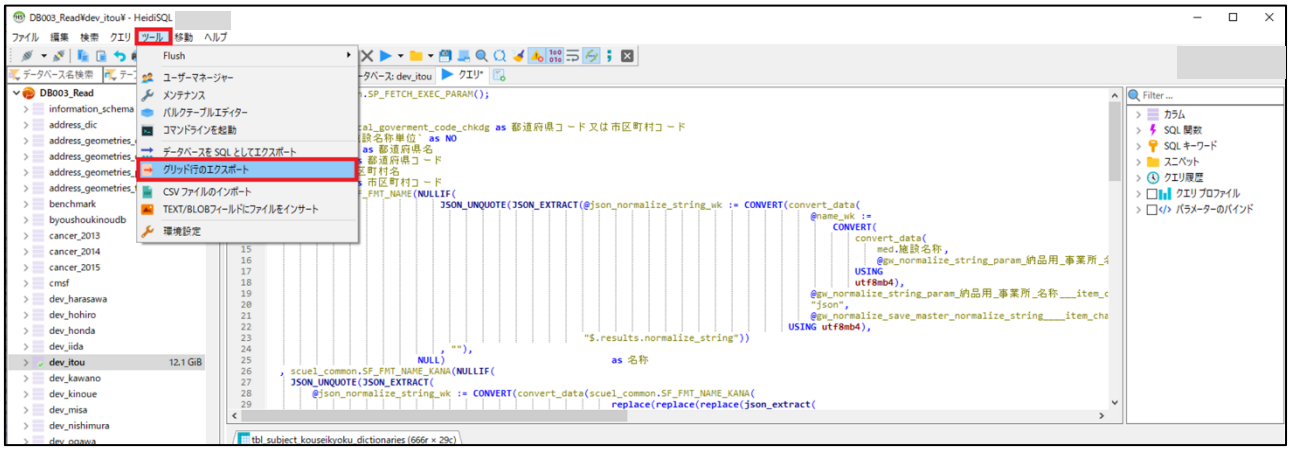
1. エクスポート対象のデータベースをクリックして選択、「▶クエリ」タブをクリックし、整形・正規化を行う SQL クエリを入力する



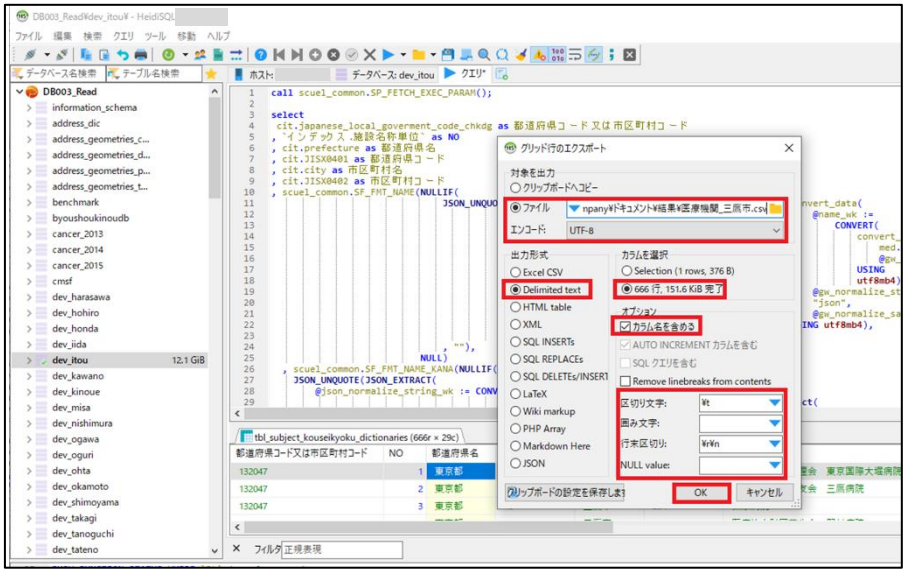
2. 「▶」→「実行」をクリックして、クエリを実行する



3. 「ツール」タブ→「グリッド行のエクスポート」をクリックする



4. 「対象を出力」で任意の出力先を指定した後、設定を変更し、[OK]をクリックする



エンコード：「UTF-8」を選択
出力形式：「Delimitied text」を選択
カラムを選択： 下側(複数行)を選択
オプション：「カラム名を含める」にチェック
区切文字：「¥t」を選択
囲み文字：（なし）
行末区切り：「¥r¥n」（Windows 形式の改行）を選択
NULL value：（なし）

5. 指定したパスにファイルがエクスポートされるので、ファイルの内容を確認する

エクスポートされた内容（例）

FAQ

アプリケーション

項目名	詳細
CSV ファイルの保存形式について教えてください。	CSV は、「テキスト（タブ区切り）」で保存しています。 データにカンマが含まれる可能性を考慮し、タブ区切りを指定しています。

データ

項目名	詳細
重複データが存在する。	重複データが目視で見つかった場合は、1 行分を手作業で削除してください。（存在が確認できている資料：板橋区データ）
名寄せや ID 付与がわからない。	東京都では来年度以降にもデータ整備モデル事業の実施を予定しています。名寄せや ID 付与については来年度のマニュアルに実装される予定です。

Power Query

項目名	詳細
ブック名、シート名、年月日に関わる部分で、記載内容と異なる状況がある場合（品川区の月ごとデータ等）	**個別対応が必要となる場合、下記の情報をご参考ください。** **●ブック名 YYYYMM 以外文字列が c ではない：**2012/05、2012/12、2019/05、2021/08、2021/11、2021/12 **●シート名の和暦年が全角：**2016/01、2021/12 **●シート名の月が全角：**2014/07、2020/10、2021/09 **●シート名の和暦年月が全角：**2020/05、2020/06、2020/08、2020/09、2020/11、2020/12、2021/01、2021/02、2021/03,2021/06、2021/08 **●シート名和暦年月以外：**2018/08(201808c)、2018/09(201809c)、2019/05(2019 年 5 月)、2019/06(C,町丁別年齢別《保存用》)、2019/07(2019 年 7 月)、2019/08(2019 年 8 月)、2019/09(2019 年 9 月)、2019/10(2019 年 10 月)、2019/11(2019 年 11 月)、2019/12(2019 年 12 月)、2020/02(2020 年 2 月) **●シート名に元号が追加：**2015/01～2015/04 **●令和元年のため、手順 3 年月日列の整形変更有：**2019/05～2019/12 **●2019/05、2019/06　：**値の置換「令和元年 5 月 1 日現在」　→「2019 年 5 月 1 日」 **●2019/07、2019/08、2019/10、2019/12：**値の置換「令和元年 7 月 1 日現在」　→「2019 年 7 月 1 日」 **●2019/09　：**値の置換「令和元年 9 月 1 日 現在」→「2019 年 9 月 1 日」 **●2019/11　：**値の置換「令和元年 11 月 1 日 現在」→「2019 年 11 月 1 日」 **●元号()で記載のため、手順 3 年月日列の整形変更有：**2021/11 **●2021/11　：**値の置換「2021 年（令和 3 年）11 月 1 日現在」→「2021 年 11 月 1 日」

項目名	詳細
マッピング操作でエラーが出る	データ群が多い場合（例：八王子のハザードマップでは約 270 件）は、マッピング時にデータ群を分割する必要があります。 データを分割の後、再度マッピングを実施してください。
市外局番付与が難しい	**八王子市** **市外局番は、「0426」であるが、** **市外局番末尾の「6」は元データに付与されている。** **混在しており、個別判断は困難な状態。** **三鷹市** **23 区外であるが、「03」「0422」**

東京都データプラットフォーム

行政データ整備モデル事業

データ整備マニュアル

マニュアルの改善に向けて

本マニュアルの改善に向けて、気づいた点などございましたら、データ利活用の推進のページにある「ご意見・ご感想」フォームからお願いいたします。

https://www.digitalservice.metro.tokyo.lg.jp/society5.0/index.html

2022 年　3 月　第 1 版
発行元・著作　東京都
委託先　日本総合研究所、株式会社自動処理

本書掲載の内容は 2022 年 3 月現在の情報です。
内容には万全を期しておりますが、システム改変等により実際の内容と異なる可能性があることをご了承ください。
操作説明に使用している画面は代表的なデータを使用するため、実際の利用画面と異なることがあります。

本書に記載されている会社名、システム名、製品名は一般に各社の登録商標または商標です。

なお、本文および図表中では、「™」、「®」は明記しておりません。