東京都データプラットフォーム
データ整備事業
データ整備マニュアル

第２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
2023．3．31

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
2.0	2023. 3. 31	株式会社自動処理/伊藤輝和	以下のページ追加 データの格納・判別 文字列正規化 データ構造の統一化 - R4 版データのページを追加 データの値チェック・クレンジング データの正規化 - 名称、住所、電話番号 1 項目に複数情報が設定されている場合 - 営業日、定休日、料金情報 複雑なデータ補完

ⓒ東京都

目次

手順別目次 4
データ活用までのフローを知る 5
データ整備事業について 5
作業に使用するアプリケーション/アドインを準備する 6
アプリケーション 6
アドイン 6
データを変換する前に準備する（事前準備） 7
1. データ構造の検討 8
2. データ構造の調査 9
3. CSV データ構造の決定 17
4. CSV データ形式の決定 19
5. 機械判読可能なデータ作成のチェックポイント 21
6. ケース一覧 23
データ構造を統一する 25
データ形式別の作業フロー（2 次元形式データ変換） 25
データ形式別の作業フロー（2 次元形式にならないデータ） 26
POWER QUERY によるデータ変換 27
AI-OCR によるデータ変換 91
データを適切に配置し、データの精度を高める 103
データの格納 103
データの判別 105
文字列正規化 107
データのマッピング 110
データの値チェック 114
データの目視によるクレンジング 116
データの正規化 121
1 項目に複数情報が設定されている場合 134
複雑なデータ補完 140
データの保存（エクスポート） 144
データベース接続の設定 144
データのエクスポート 145
FAQ 147

ⓒ東京都

手順別目次

手順	プロセス	概要	備考	参照先
⓪	**データの整備概要の理解**	データ活用までのフローを知る データ整備事業について 作業に使用するアプリケーション/アドインを準備する		P5
①	**データ構造・形式の検討**	データを変換する前の事前準備 1.データ構造の検討 2.データ構造の調査 3.CSV データ構造の決定 4.CSV データ形式の決定 5.機械判別可能なデータ作成のチェックポイント 6.ケース一覧		P7
②	**Excel や PDF からデータ変換**	データ構造を統一する Power Query によるデータ変換 ✓ ケース[A]　保存形式 XLS／縦横多段クロス集計 ✓ ケース[B]　保存形式 XLS／単純クロス集計 ✓ ケース[C]　保存形式 XLS／単純表形式または複数ファイル ✓ 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 ✓ ケース[E]　保存形式 PDF AI-OCR によるデータ変換	データ内容によりケース[A]～ケース[E]に分けられるので、最適な手法を選択	P25
③	**利活用しやすいデータへ変換**	データを適切に配置し、データの精度を高める データの格納 データのマッピング データのクレンジング データの正規化 ※データの正規化とは、一定のルールに基づいて、データを整える作業になります。	API、もしくは外部調達によって実施想定	P103
④	**データベースの作成**	データの保存（エクスポート）		P146

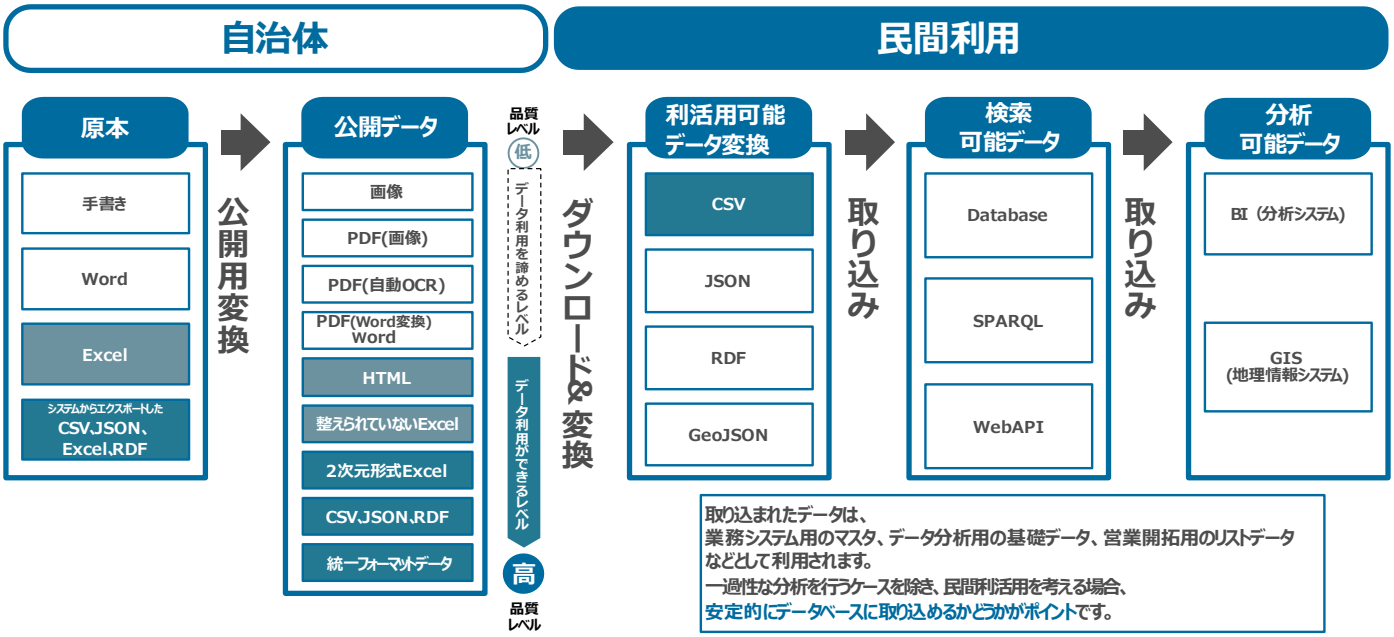
ⓒ東京都

手順0 データ活用までのフローを知る

データ整備事業について

民間利活用を目的にしたデータ整備では、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」、「営業開拓用のリストデータ」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。**

データ整備作業は、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。
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も作成しています。
（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、市区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.6）**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、安定的な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。

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

ⓒ東京都

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/

AI-OCR ツール

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。
AI 技術を用いて柔軟な文字認識、帳票データの読取などが可能な AI-OCR 製品として、AI inside「DX Suite」、Cogent Labs「tegaki」、インフォディオ「スマート OCR」、NTT データ NJK「FormOCR」、PFU「DynaEye」等、様々な製品がリリースされていますが、本書では LINE「CLOVA OCR（Reader）」の AI-OCR 機能を利用して、データを生成する手順を案内しています。
詳細はこちらをご参照ください。
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。詳細はこちらをご参照ください。
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です。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。詳細はこちらをご参照ください。
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

ⓒ東京都

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. 8）**
   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. 9）**
   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。
   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. **データ構造の検討・決定** **（参考 p.17）**
   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. **CSV データ形式の決定** **（参考 p.19）**

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. 21）**

6. **ケース一覧** **（参考 p. 23）**

ⓒ東京都

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

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。
入手元により、データ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。

保健所からの診療所データ

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　東京国際大垣病院	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	原原病院	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、緩和ケア内科、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、脳細内科、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	医療法人社団緑光会　長谷川病院	三鷹南大沢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

医師会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	医療法人社団健晶会　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　鹿嘉整形外科	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	医療法人社団研道会　南水クリニック	三鷹市井の頭２－１４－２	パークブリージェ井の頭１０１	内・小	0422-76-1232
9	みこしばクリニック	三鷹市井の頭２－１９－２５		小・精・神	0422-49-5300
10	医療法人社団基昭会　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	半礼の里駅前クリニック	三鷹市井の頭２－７－９	関原ビル１階	内・整・外・消内・循内・皮	0422-40-6054
12	勝林医院	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	医療法人社団基司会　菊林医院	三鷹市井の頭４－１６－１０		内・小・消、	0422-43-0526
15	武蔵野みどり診療所	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	医療法人社団美々会　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

ⓒ東京都

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
- **データ構造調査**　各種データ連携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）を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/

ⓒ東京都

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
	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	

ⓒ東京都

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
	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	2019/5/1
	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

ⓒ東京都

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。
Google Dataset 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。

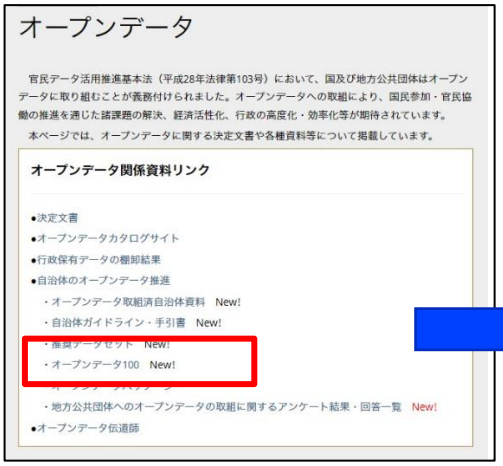
メタデータ②

ⓒ東京都

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

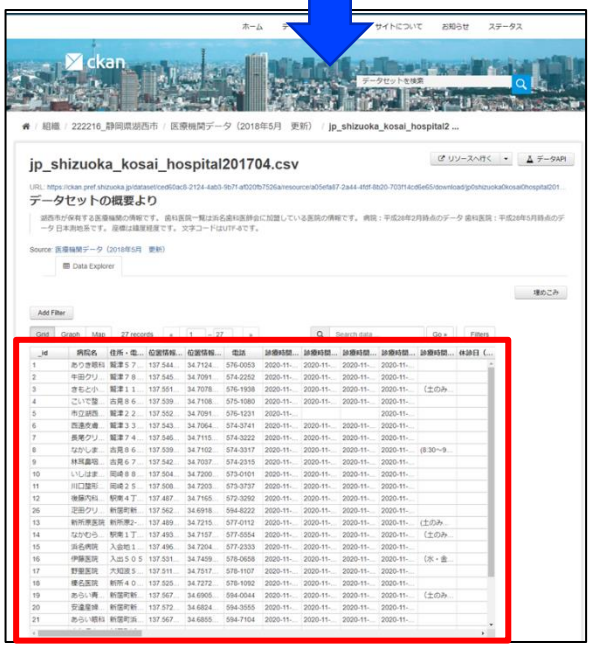
ⓒ東京都

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。
Google Dataset 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

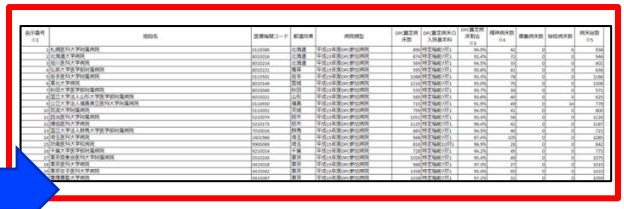
ⓒ東京都

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。

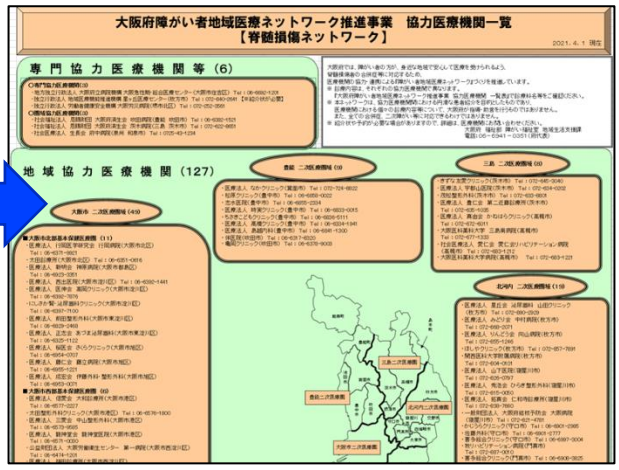


データ構造④

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。

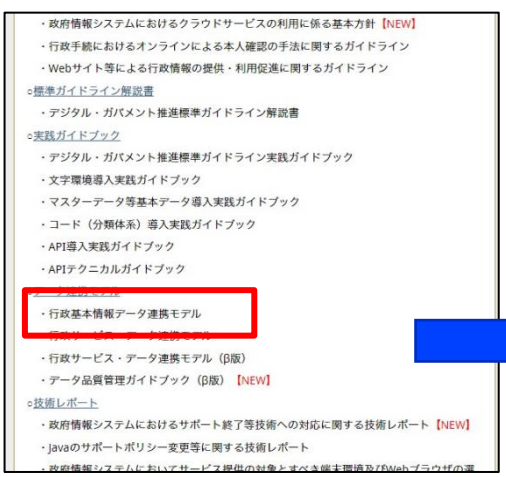
必要な拠点データを入手します。

データ構造⑤

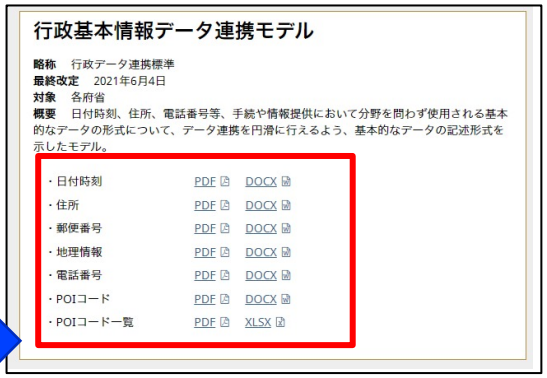
ⓒ東京都

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。



ⓒ東京都

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。

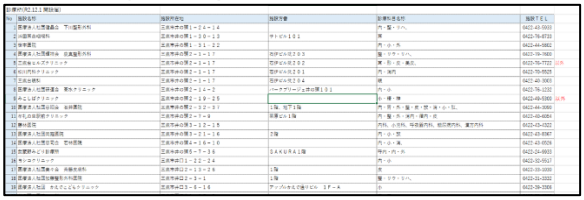
CSV データフォーマット検討

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

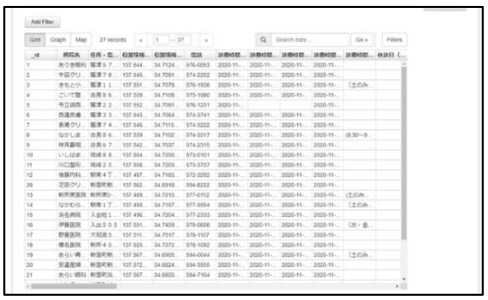
標準データセット（P. 13）

データ構造①（P. 8）

データ構造②（P. 8）

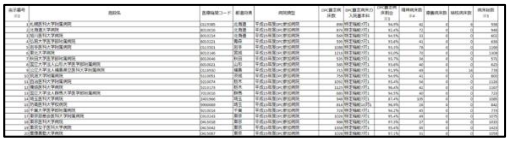


データ構造③（P. 14）

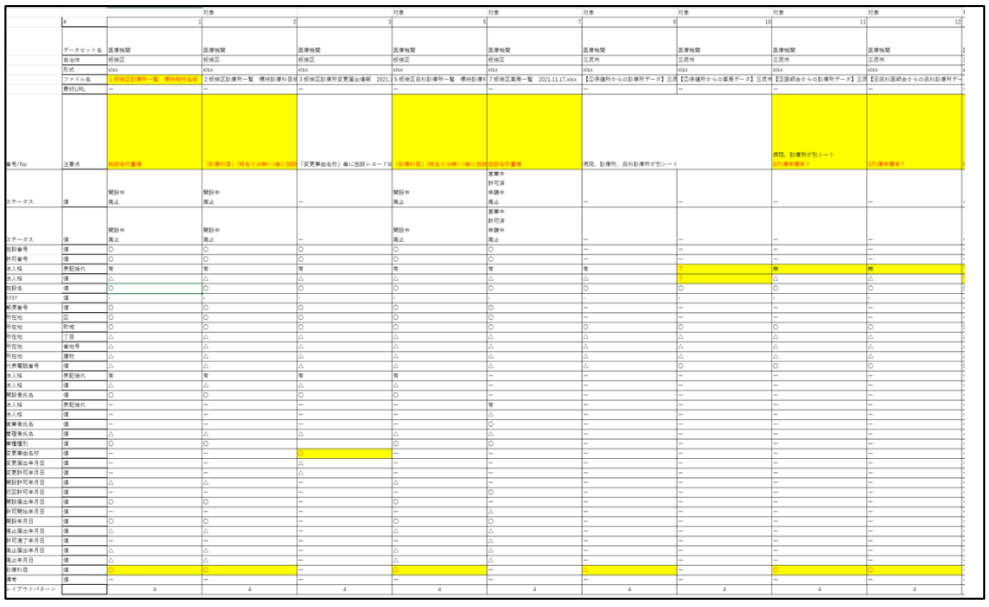


データ構造④（P. 14）

データ構造⑤（P. 14）



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。

ⓒ東京都

3.　CSV フォーマットのデータ項目を整理する

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

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。
※Excel2016 以降の場合、変更履歴を保存する機能がありますので、データ構造の変更を管理したい場合には、ご利用下さい。

ⓒ東京都

4. CSV データ形式の決定

ISO、JIS、WWC、RFC を参照しデータ形式を検討します。

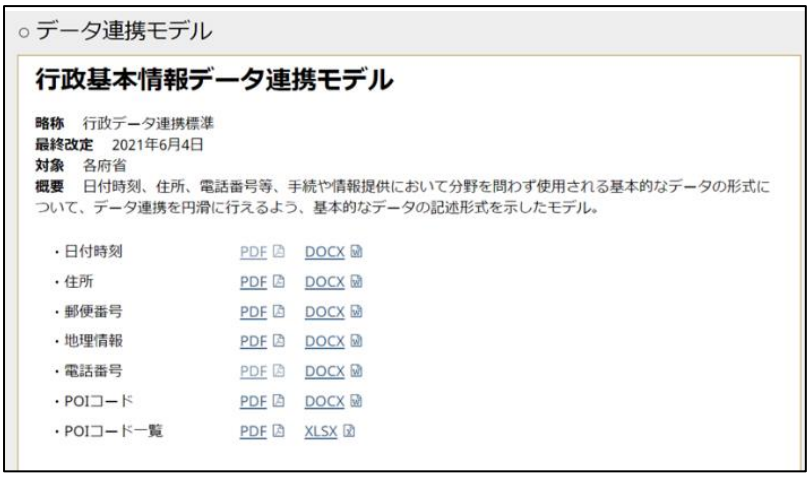
- データ連携モデル　　https://cio.go.jp/guides
- ISO、JIS、WWC、RFC を検索　　https://www.google.com

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）

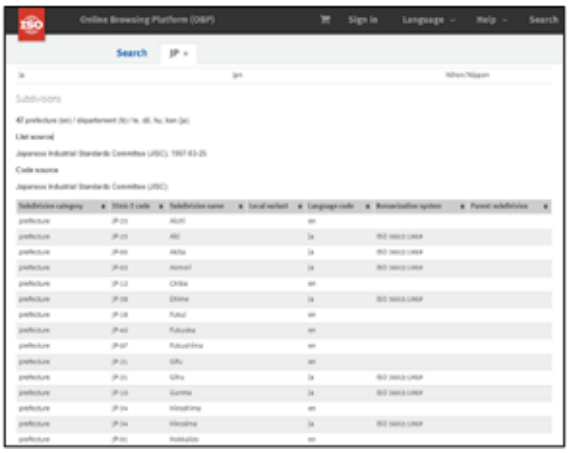
ISO、JIS にて定義されているコードがあるか確認します。

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。



ⓒ東京都

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	

ⓒ東京都

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、オブジェクトが残っていないか
□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　データの単位をデータの項目名、もしくは、データ項目の説明覧等に記載しているか ※7
□13　数字記載の際は半角数字で記載しているか ※2

日付
□14　年月日記載の際は、日付書式かつ、半角数字で記載しているか
(全角数字、漢数字、元年、末日、上旬などは使用しない) ※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、データ整備方針に基づいて、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 ※8

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

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)
をもとに、チェックリスト形式に改変。

ⓒ東京都

※1～※7 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 数値の後ろに直接単位の記載やエクセルのセルのユーザー定義で単位を追加する方法は、機械で判読する際に数値として扱えなくなるため実施してはいけない
※8 Excel マクロで処理する

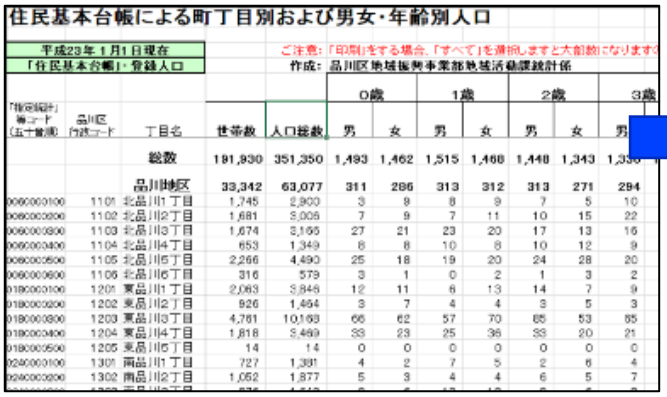
ⓒ東京都

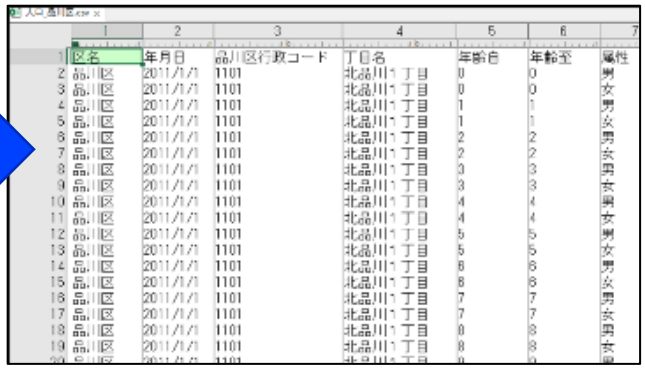
6. ケース一覧

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

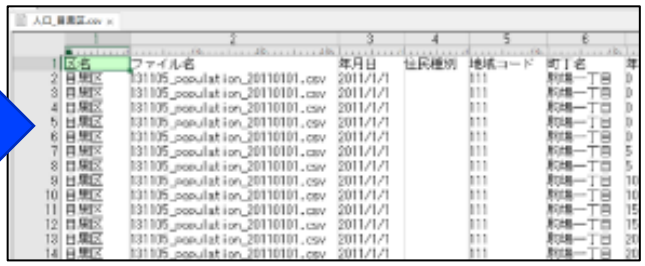
Excel 上でコピー、貼り付け等の操作や、結合セルを解除することで CSV データに変換できるデータのレイアウトの Excel データを、CSV データに変換します。





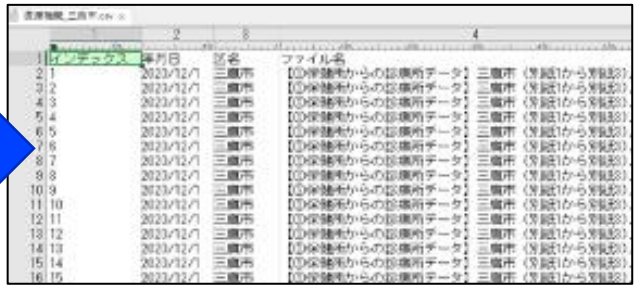
ケース[B]：保存形式 XLS／単純クロス集計

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

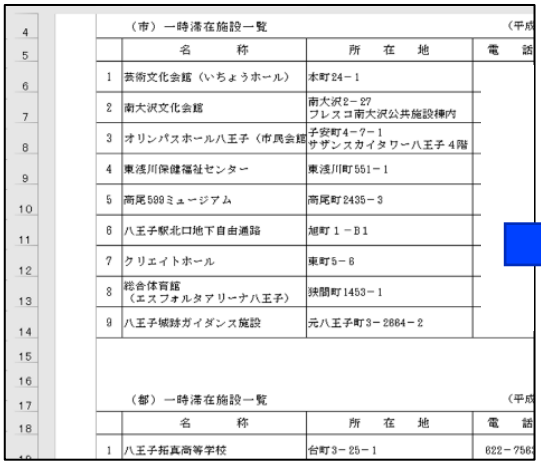
1 行目が項目行、2 行目以降がデータ行となった、表形式のデータを、CSV データに変換します。
項目の粒度や目的が混在している場合は、変換したい CSV ファイルにあわせて表を分割する必要があります。



ⓒ東京都

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。

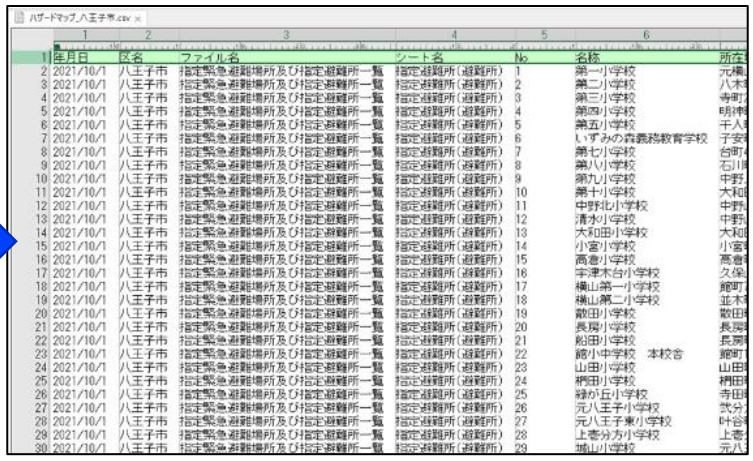


（市）一時滞在施設一覧　（平成

	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5－8	
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3－2884－2	

（都）一時滞在施設一覧　（平成

	名　称	所　在　地	電　話
1	八王子拓真高等学校	台町3－25－1	622－756

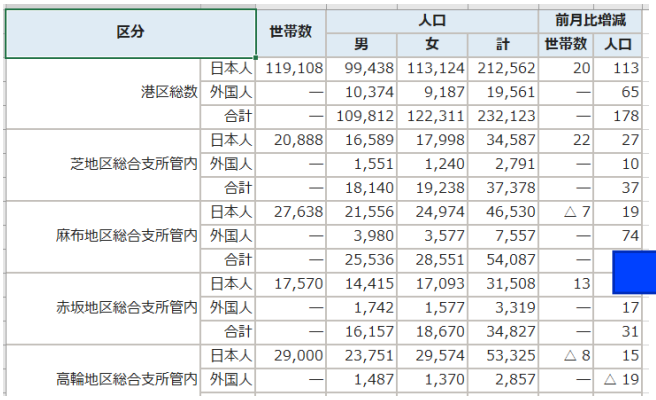


ハザードマップ_八王子市.csv

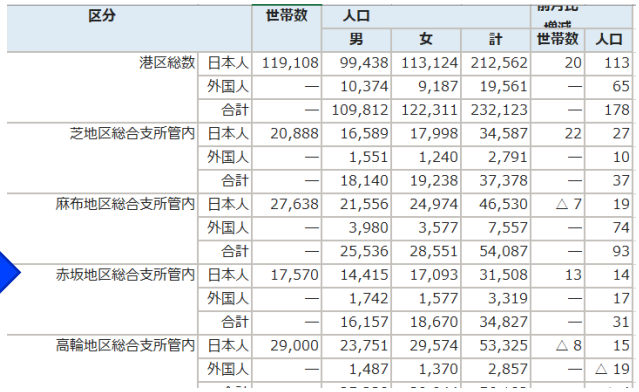
作成日	区名	ファイル名	シート名	No	名称	所在
2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	1	第一小学校	元横
2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	2	第二小学校	八木
2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	3	第三小学校	寺町
2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	4	第四小学校	明神
2021/10/1	八王子市	指定緊急避難場所及び指定避難所一覧	指定避難所（避難所）	5	第五小学校	千人
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	子安
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	台町
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	石川
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	中野
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	大和
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	中野
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	中野
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	大和
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	小宮
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	高倉
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	久保
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	館町
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	並木
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	散田
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	長房
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	長房
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	館町
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	山田
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	柄田
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	寺田
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	犬分
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	叶谷
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	上壱
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	元八

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	
	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

ⓒ東京都

手順 2

データ構造を統一する

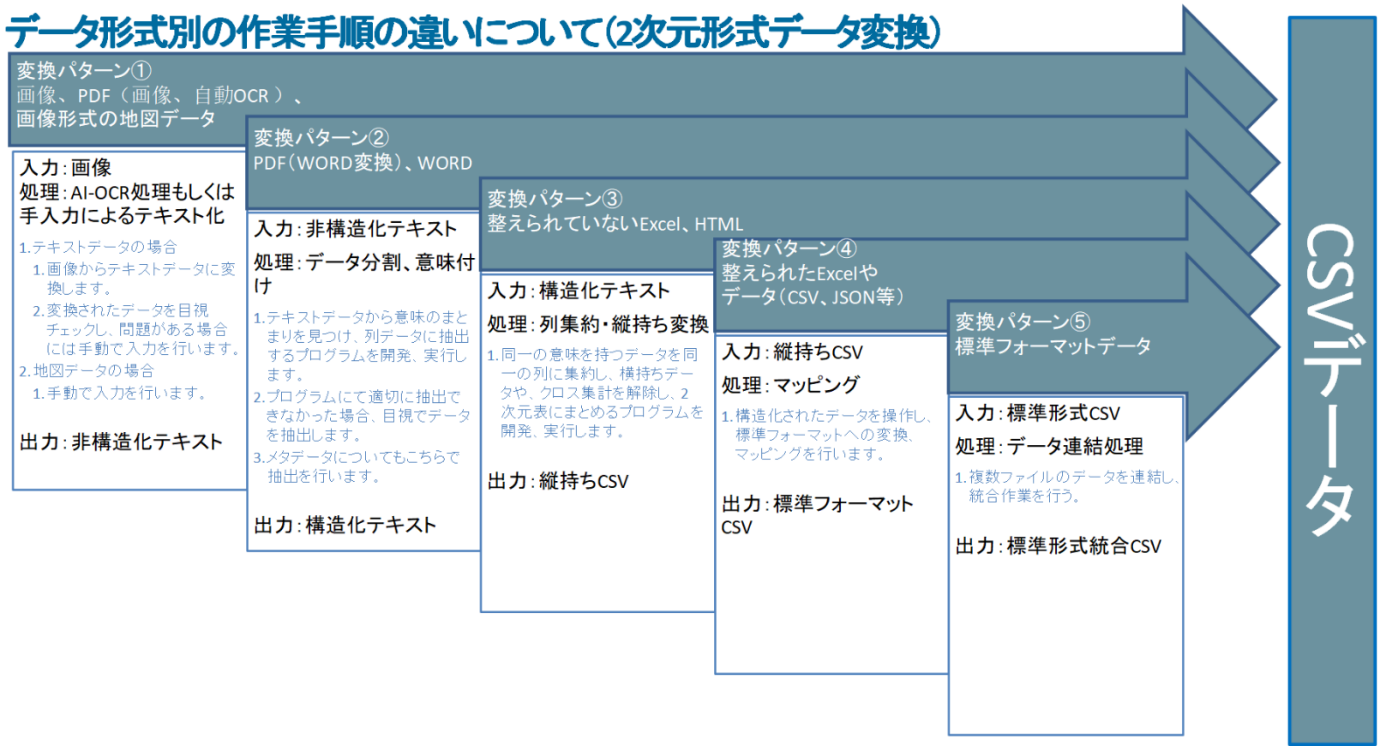
各種データの構造を統一します。

1. **PowerQurery によるデータ変換** **（参考 p.27）**
   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.30）**
   ケース[B]　保存形式 XLS／単純クロス集計 **（参考 p.51）**
   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.58）**
   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.81）**

2. **AI-OCR によるデータ変換** **（参考 p.91）**
   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.91）**

データ形式別の作業フロー（2 次元形式データ変換）

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 構造に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）



ⓒ東京都

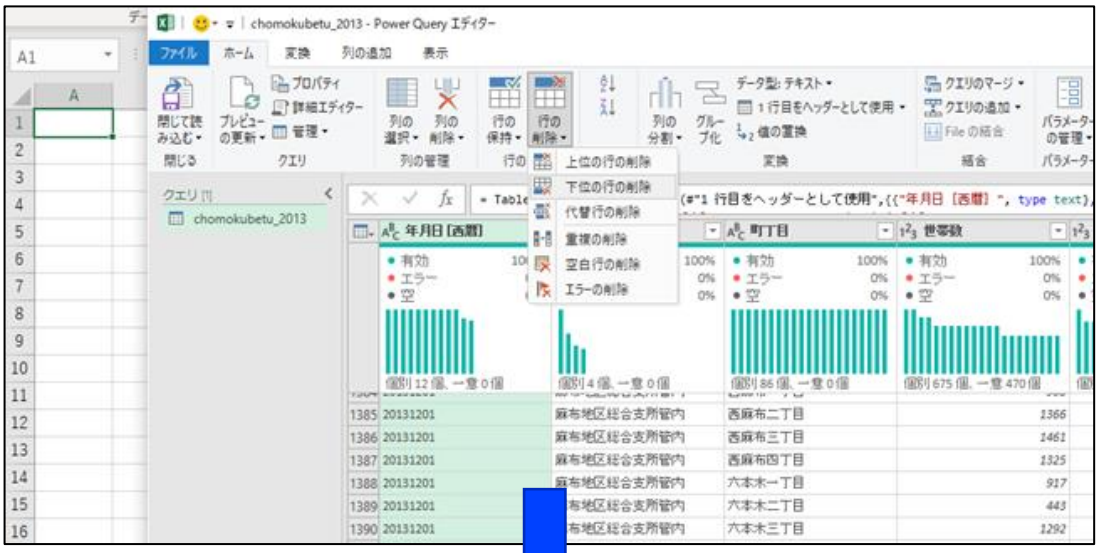
データ形式別の作業フロー（2 次元形式にならないデータ）

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。

ⓒ東京都

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. 6)をご覧ください。



	A	B	C	D	E	F	G
1	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］
2	20130101	芝地区総合支所管内	芝一丁目	792	715	674	1389
3	20130101	芝地区総合支所管内	芝二丁目	1634	1412	1480	2892
4	20130101	芝地区総合支所管内	芝三丁目	1905	1586	1690	3276
5	20130101	芝地区総合支所管内	芝四丁目	1213	985	1092	2077
6	20130101	芝地区総合支所管内	芝五丁目	2004	1440	1802	3242
7	20130101	芝地区総合支所管内	海岸一丁目	1174	1093	984	2077
8	20130101	芝地区総合支所管内	東新橋一丁目	723	661	745	1406
9	20130101	芝地区総合支所管内	東新橋二丁目	212	194	140	334
10	20130101	芝地区総合支所管内	新橋一丁目	23	13	15	28
11	20130101	芝地区総合支所管内	新橋二丁目	88	81	74	155
12	20130101	芝地区総合支所管内	新橋三丁目	139	155	121	276
13	20130101	芝地区総合支所管内	新橋四丁目	264	228	202	430
14	20130101	芝地区総合支所管内	新橋五丁目	314	291	253	544
15	20130101	芝地区総合支所管内	新橋六丁目	592	476	358	834
16	20130101	芝地区総合支所管内	西新橋一丁目	156	138	129	267
17	20130101	芝地区総合支所管内	西新橋二丁目	221	191	182	373

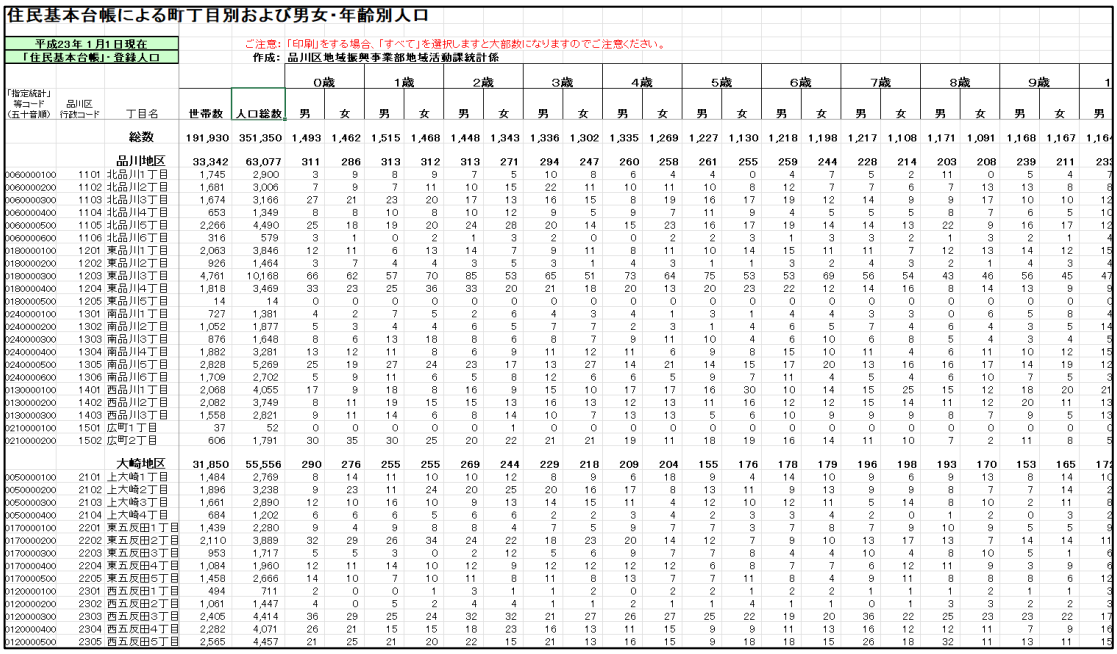
ⓒ東京都

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

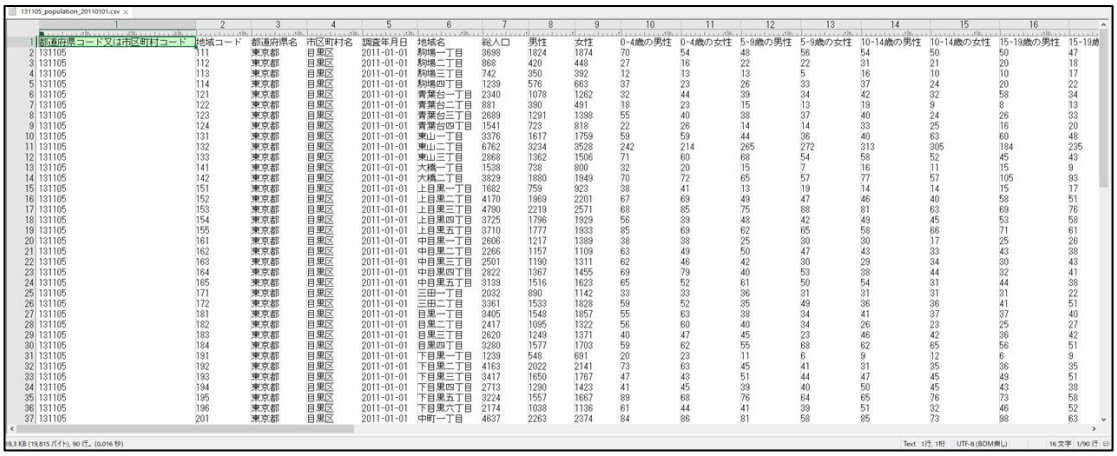
ご注意：「印刷」をする場合、「すべて」を選択しますと大部数になりますのでご注意ください。
作成：品川区地域振興事業部地域活動課統計係

「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	173
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
0170000300	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	68	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	26
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	72	63	45	41	31	35	38	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

ⓒ東京都

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

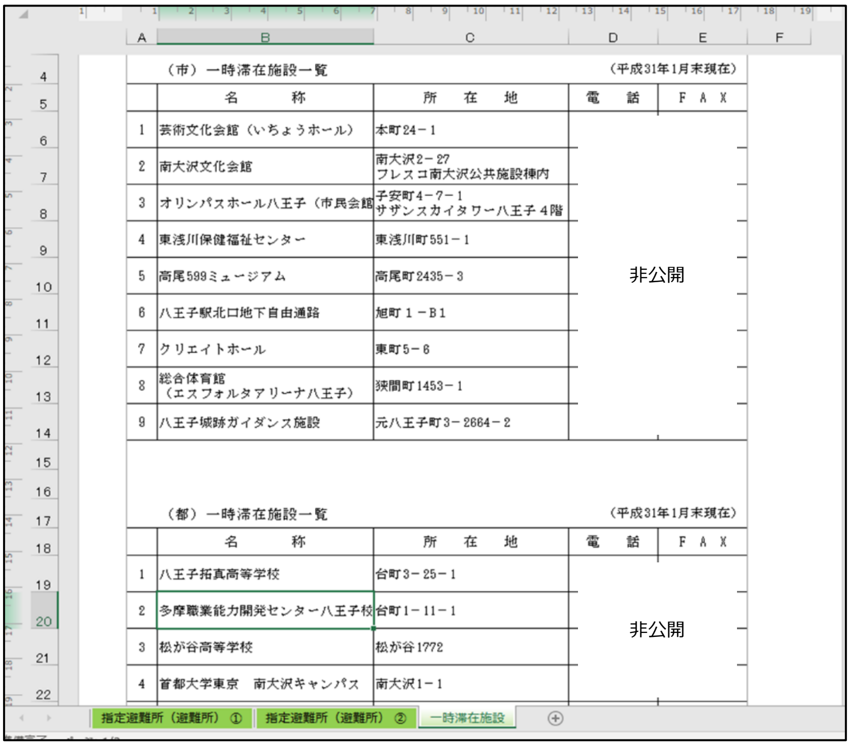
1 行目が項目行、2 行目以降がデータ行となった、表形式のデータ。
項目の粒度や目的が混在している場合は、変換したい CSV ファイルにあわせて表を分割する必要があります。

ステータ	施設番	許可番号	施設名称	郵便番号	所在地	代表電話番	開設者氏	管理者氏	業種種別	標榜診療科目名	開設許可年月	開設届出年月	開設年月	廃止届出年月	廃止年月
開設中	0000299	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	0000299	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	0000299	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	0000299	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	[illegible]	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	[illegible]	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	F A X
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		

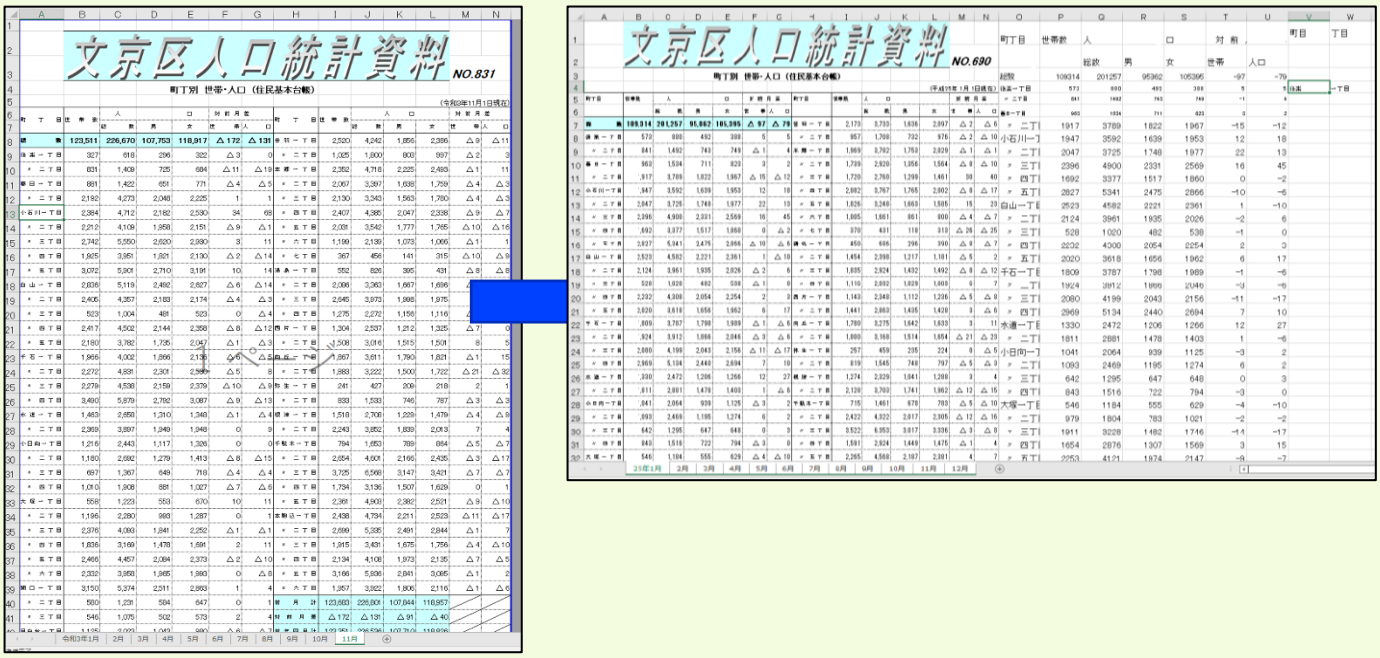
（都）一時滞在施設一覧　（平成31年1月末現在）

	名　称	所　在　地	電　話	F A X
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

ⓒ東京都

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。

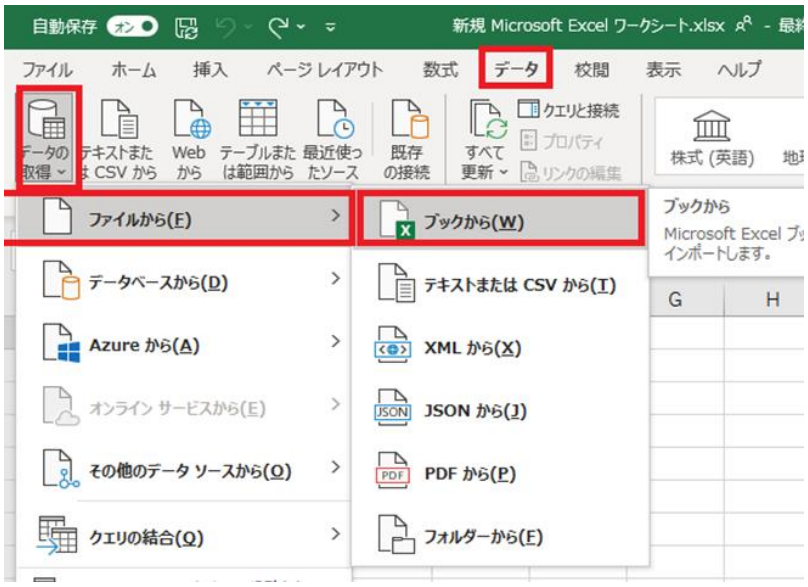
① **Excel で整形、加工したいファイルをインポートします。** **（詳細： p.31）**
② **Excel の Power Query 機能で、データを整形します。** **（詳細： p.32）**
③ **Excel の Power Query 機能で（年次、月次等）定期発生するデータを整形する** **（詳細： p.46）**

ⓒ東京都

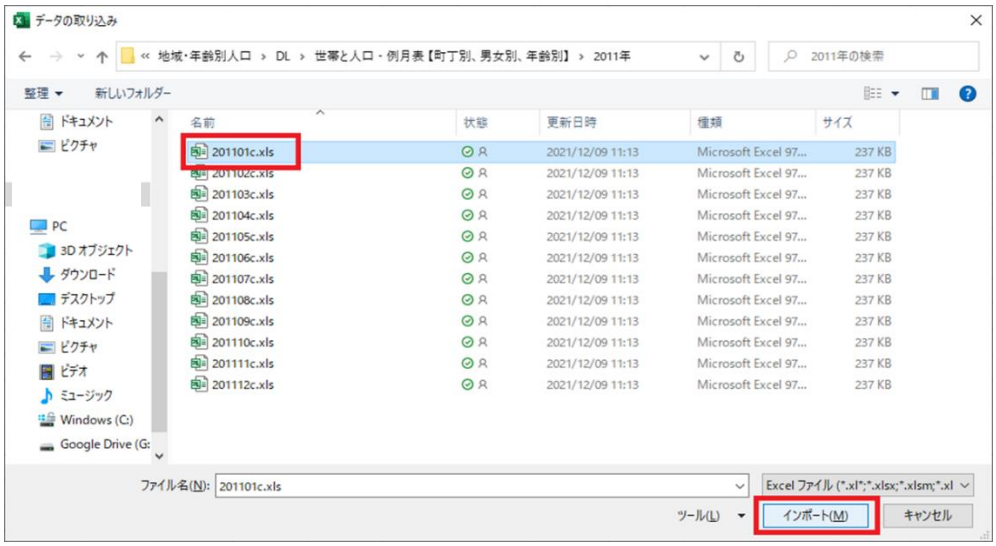
① Excel で整形、加工したいファイルをインポートする

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

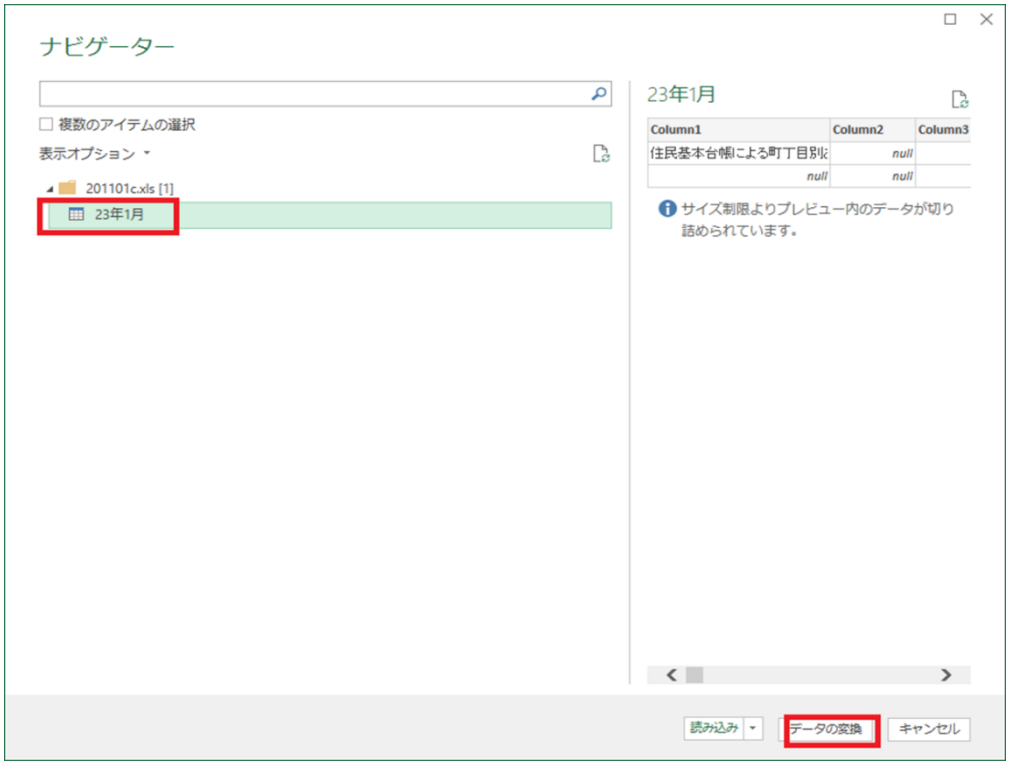


3. Excel ファイルをクリックして選び、[インポート]をクリックする



ⓒ東京都

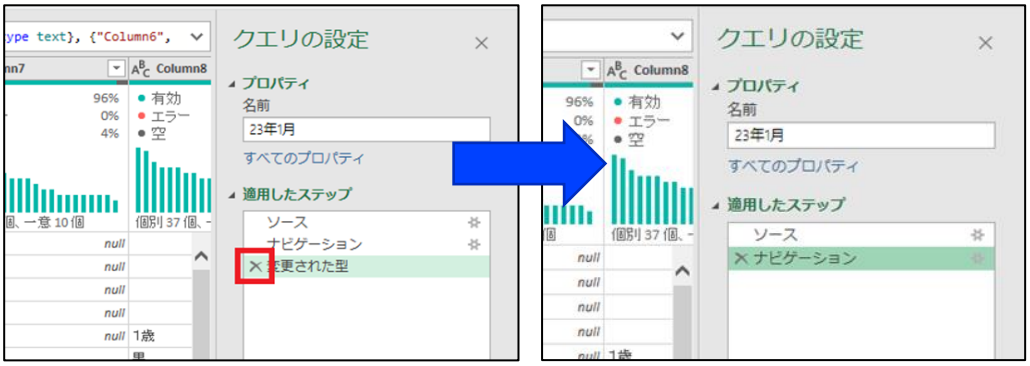
4.　変換するアイテムをクリックして選び、[データの変換]をクリックする



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。

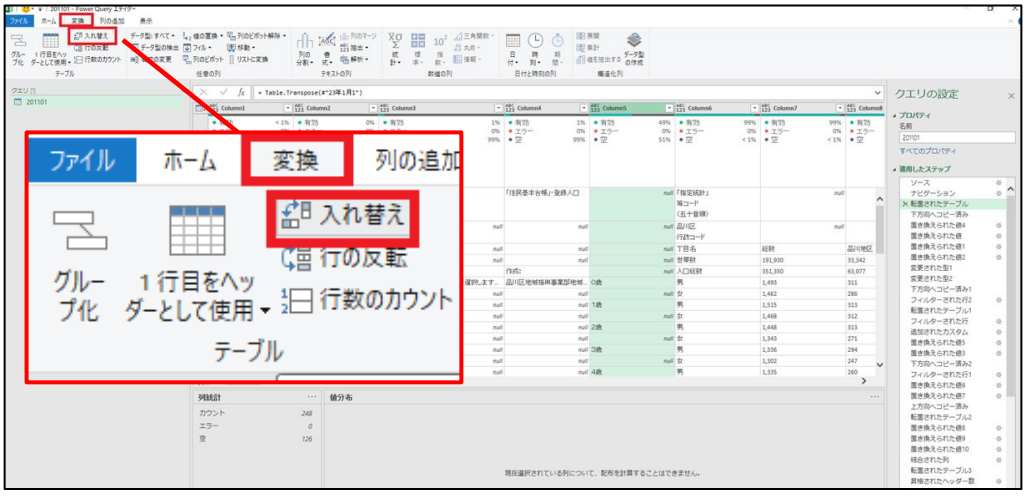
② Excel の Power Query 機能でデータを整形する

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。

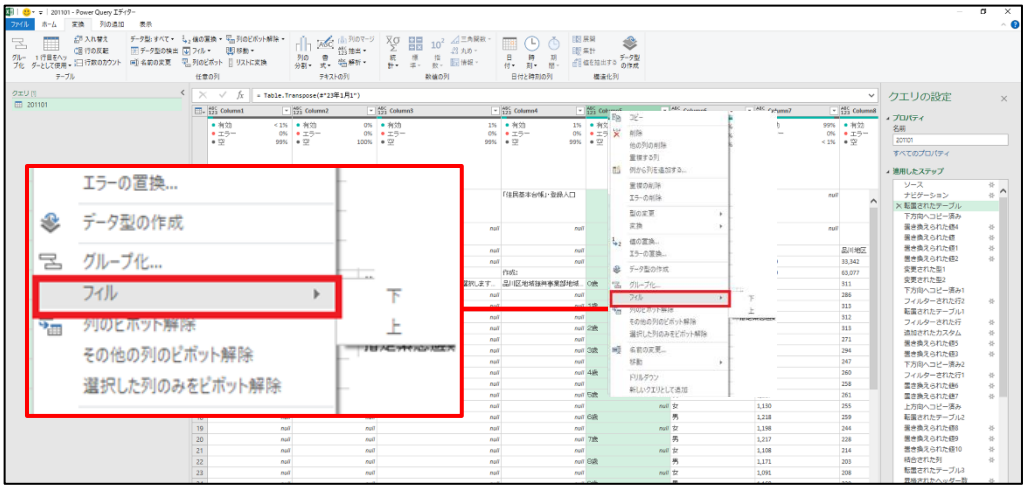
2.　[変換]タブ→[入れ替え]をクリックする



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。

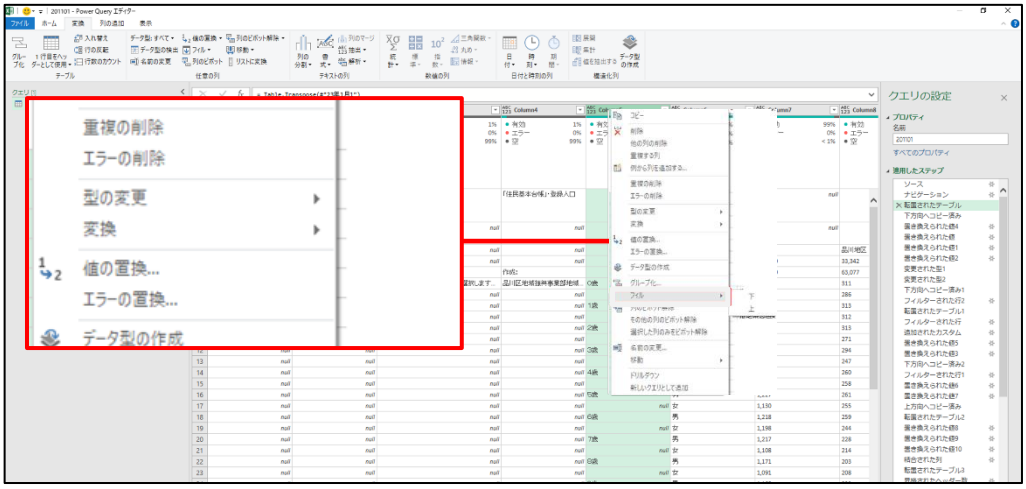
ⓒ東京都

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

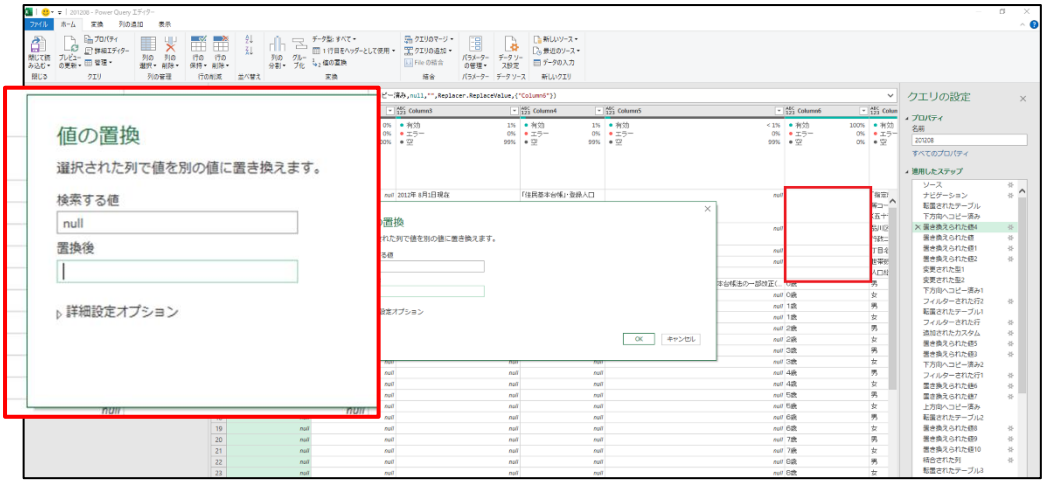


[下]または[上]は、コピーしたい方向を示します。
[下]は上から下の方向にデータをコピーし、[上]は下から上の方向にデータをコピーします。

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。

ⓒ東京都

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

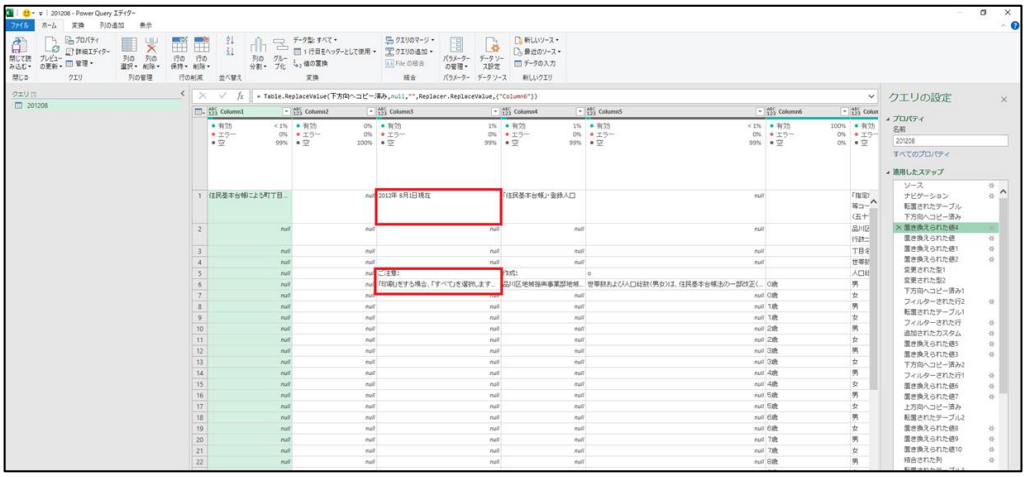
ここでの不要データ置換作業は、後の手順で日付データをフィルするために行います。

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。

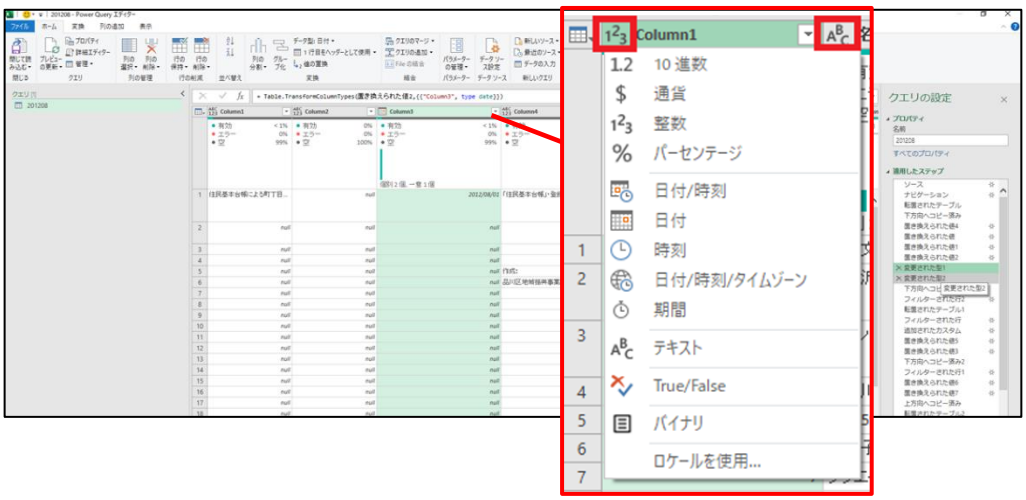
・対象列選択状態で　右クリック 「値の置換」
　置換する値：不要な()や現在、未等の文言 → 置換後「""」（ダブルクォーテーション）



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。

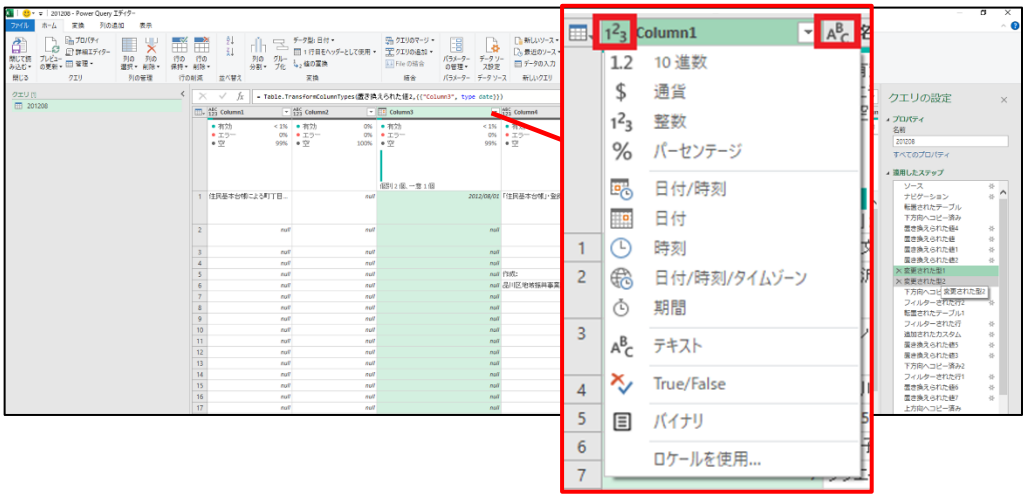
× 令和元年１月１日　（元号）	○ 令和1年1月1日
× 令和１年１月１日　（全角数字）	○ 令和1年1月1日
× 令和一年一月一日　（漢数字）	○ 令和1年1月1日
× ２０１９０１０１	○ 20190101
× ２０１９/1/1	○ 2019/1/1

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

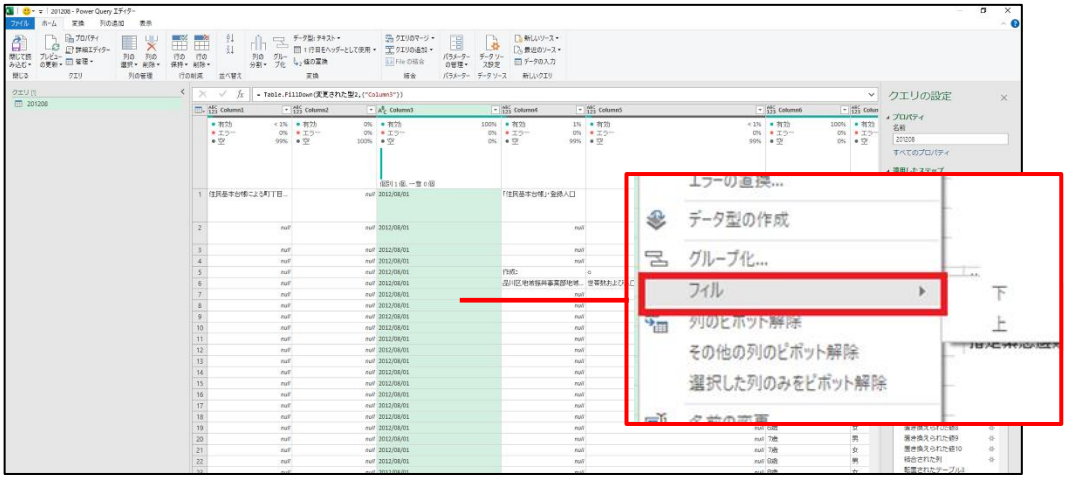


ⓒ東京都

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。

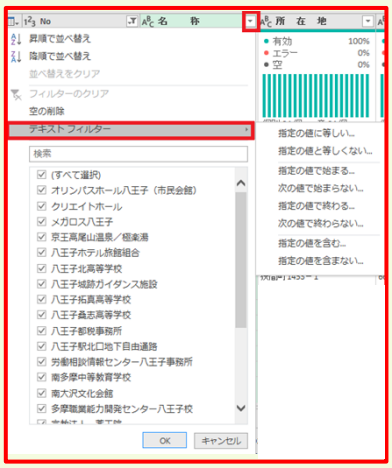
不要な行を非表示にする方法は 2 種類あります。

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。

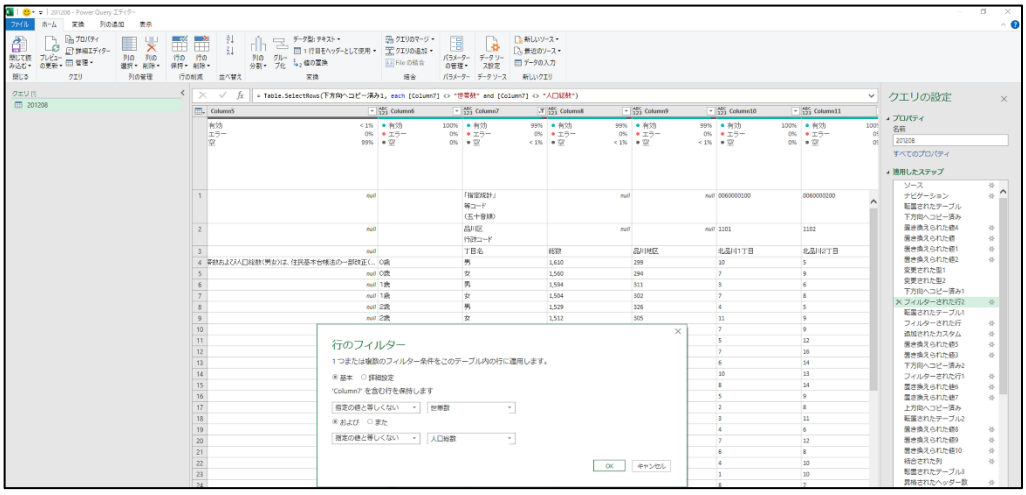
数値フィルター

テキストフィルター

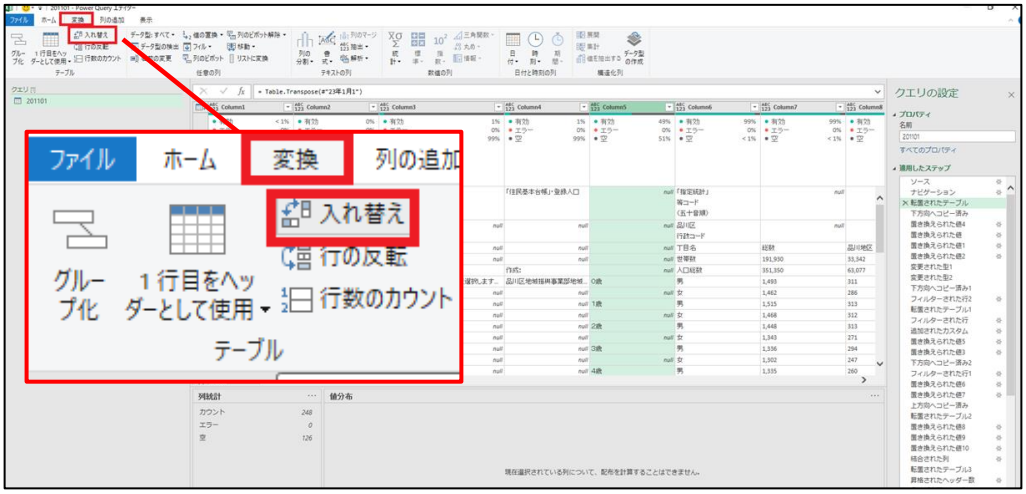


ⓒ東京都

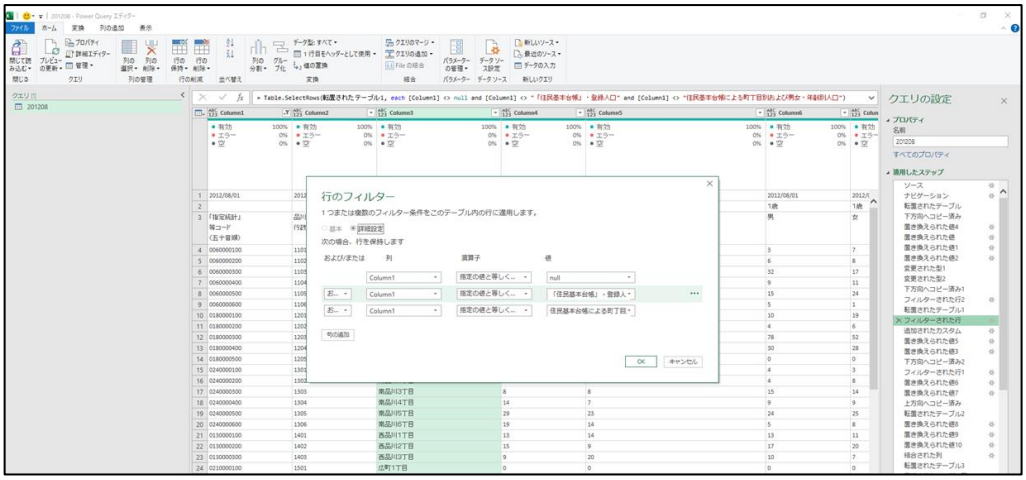
10. 年月日の列を右クリックし、除外するデータを非表示にする



11. [変換]タブ→[入れ替え]をクリックする

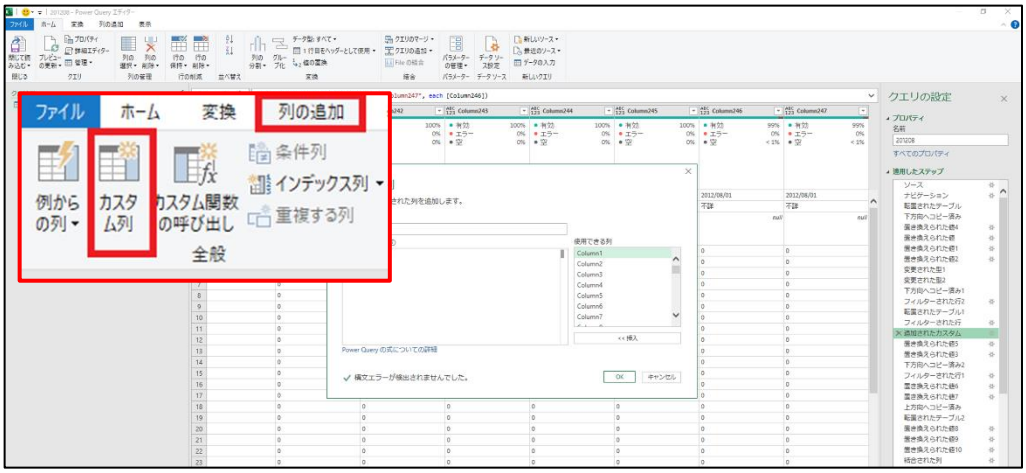


12. 対象列の不要行を削除するため、対象列の右側「▼」をクリックし、除外するデータを非表示にする



ⓒ東京都

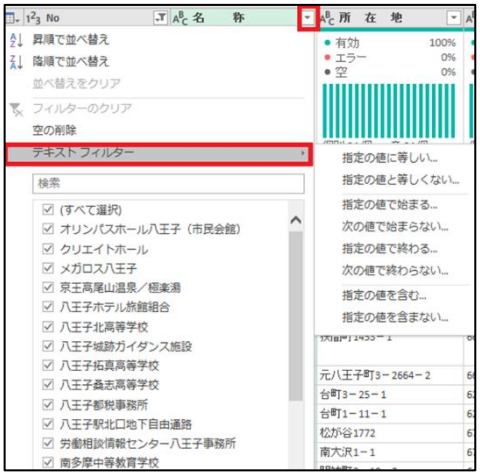
13. [列の追加]タブ→[カスタム列]をクリックする



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。

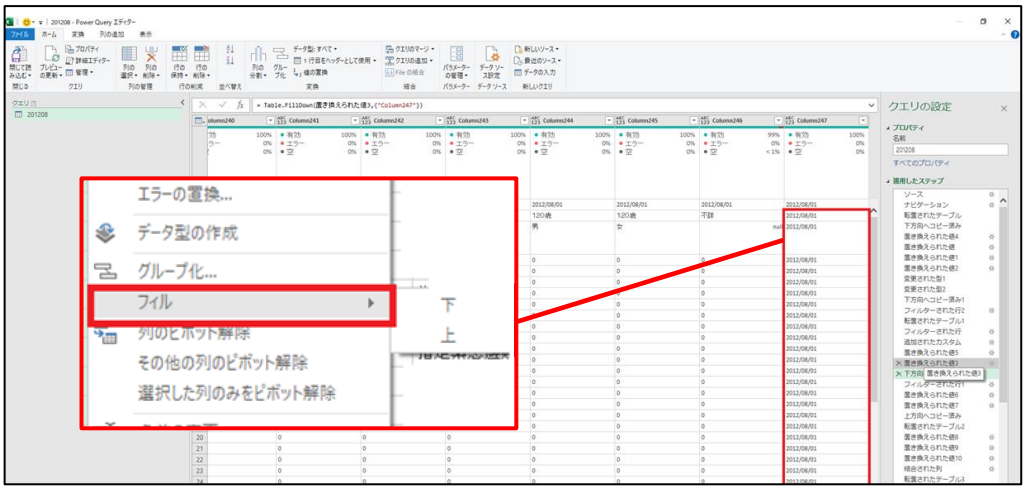
14. 追加したカスタム列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

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

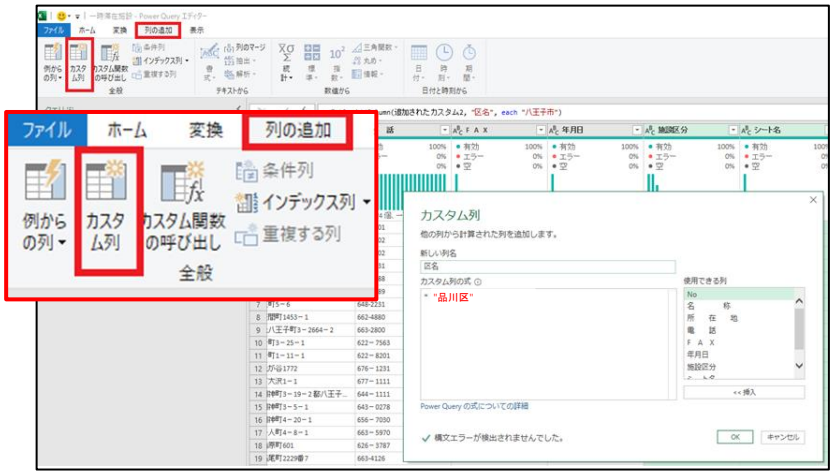


[下]または[上]は、コピーしたい方向を示します。
[下]は上から下の方向にデータをコピーし、[上]は下から上の方向にデータをコピーします。

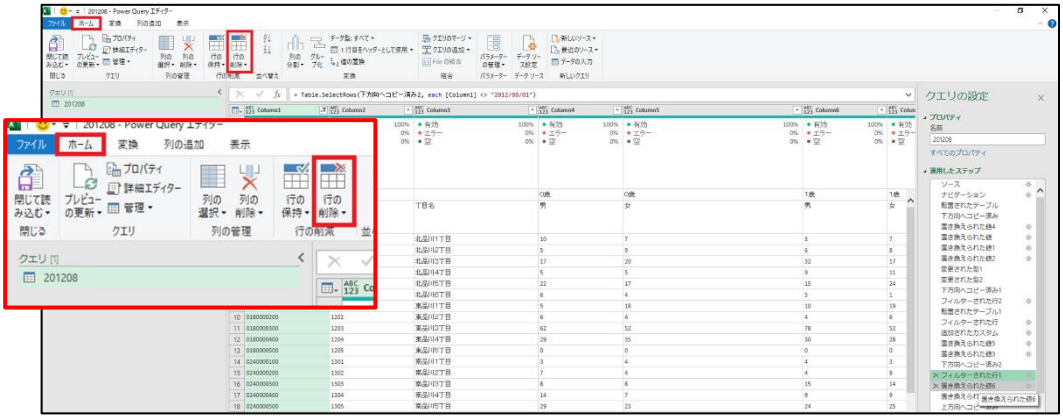
ⓒ東京都

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。



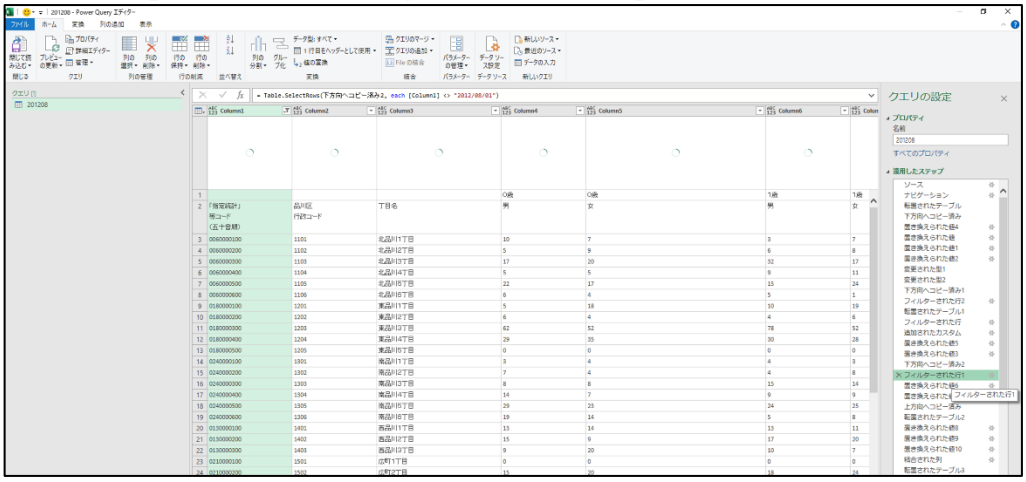
17. [ホーム]タブ→[行の削除]→「上位の行の削除」をクリックし、削除する上位の行数を入力する



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」

ヘッダー行ズレを整形します。

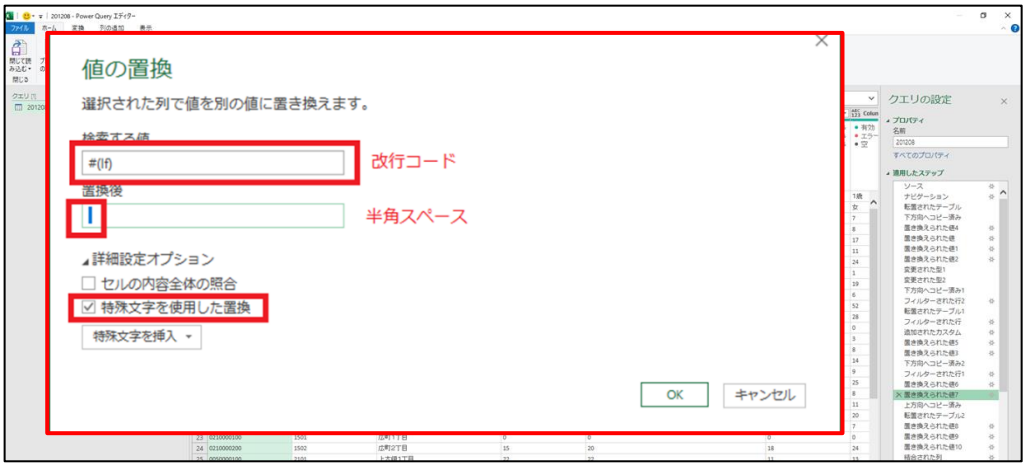


ⓒ東京都

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

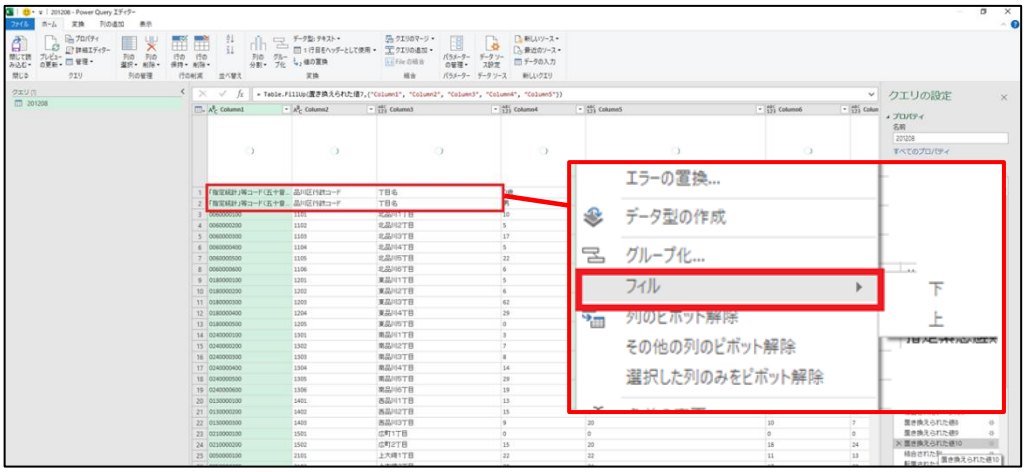
置換する値「#(lf)」（改行コード）　→　置換後「 」（半角スペース）
「詳細設定オプション」の「特殊文字を使用した置換」にチェックを付けておきます。

改行コードを削除しています。

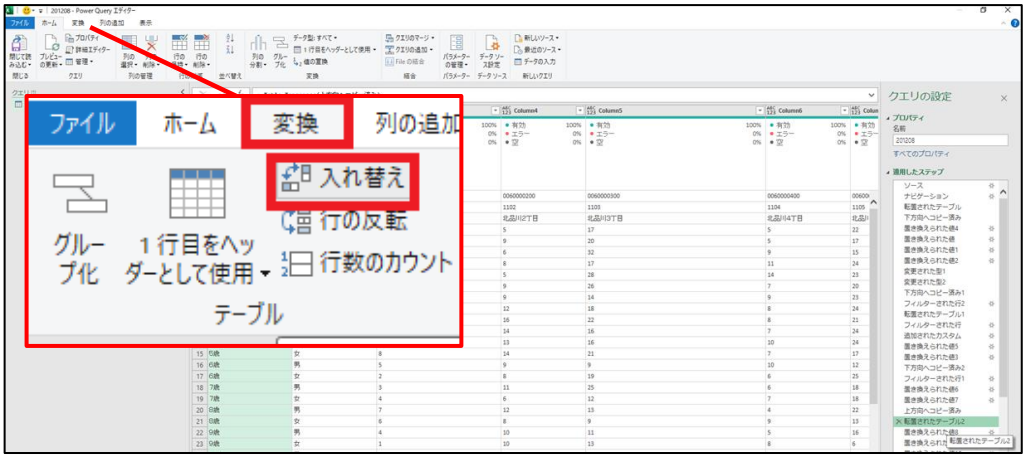


20. 対象列を右クリックし、[フィル]→[下]または[上]をクリックする

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

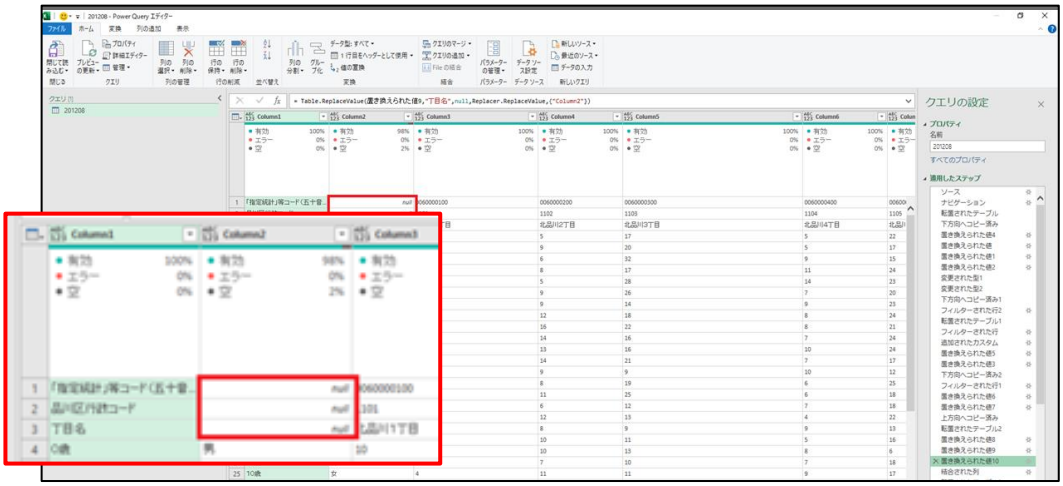


ⓒ東京都

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」

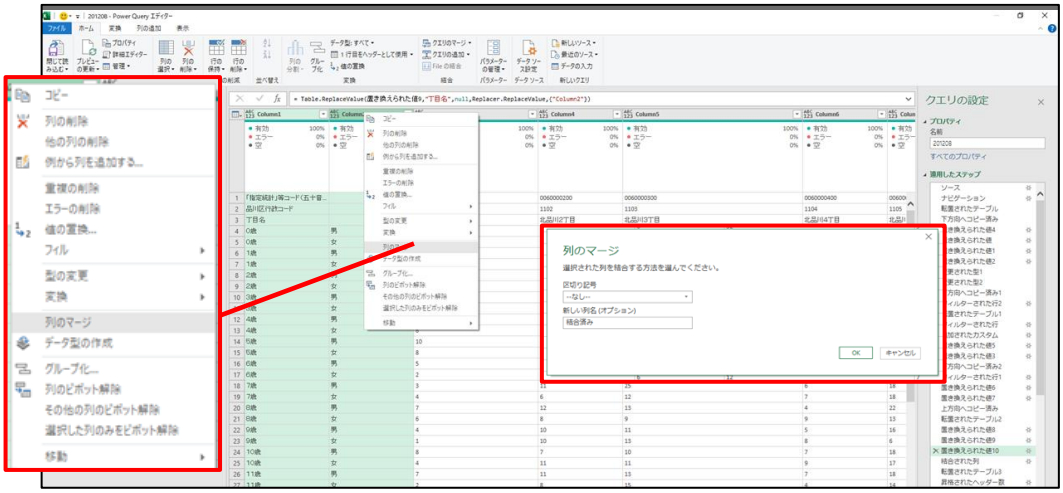
列をマージ（結合･統合）するため、不要な文言を削除しています。



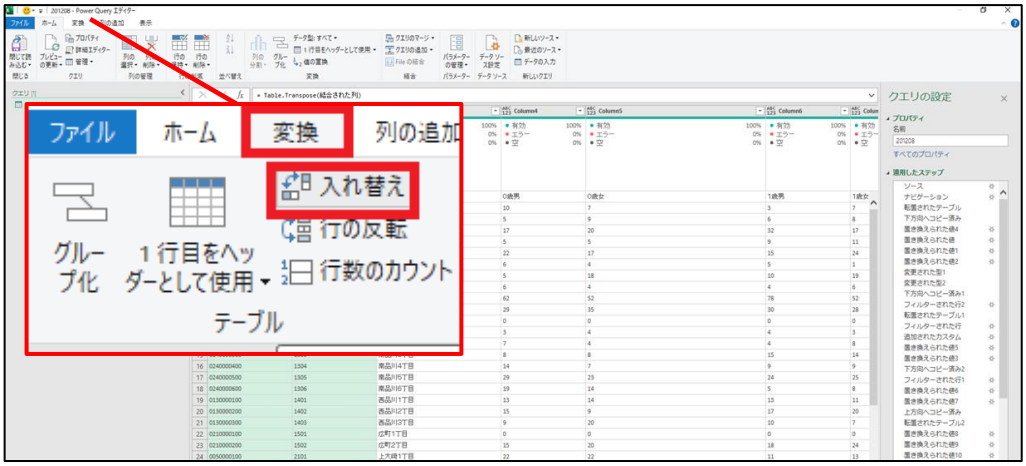
23. 対象列が選択された状態で右クリックし、[列のマージ]をクリックし、値を置換する

「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。

ヘッダー列を作成するため、列をマージしています。

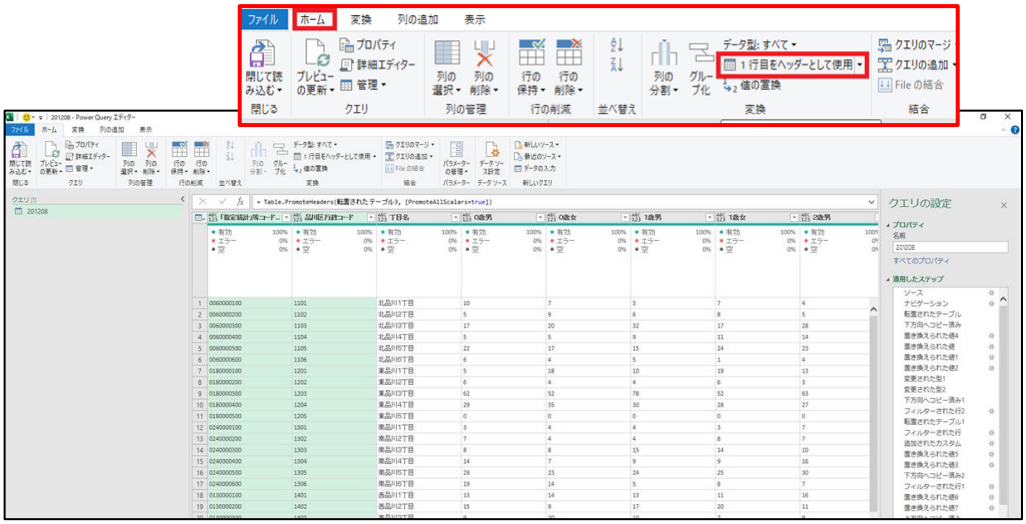


24. [変換]タブ→[入れ替え]をクリックする

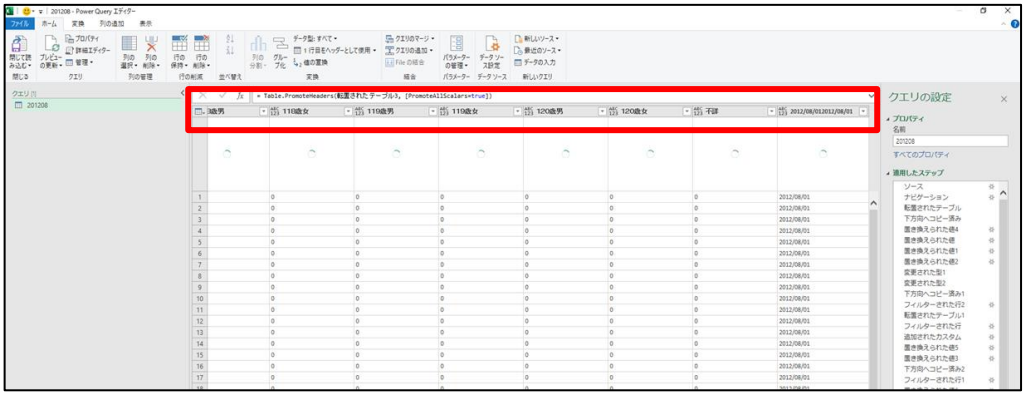


ⓒ東京都

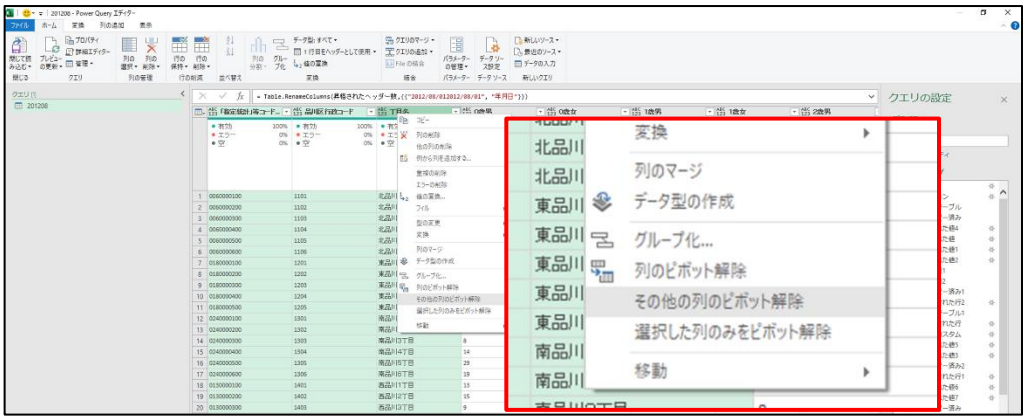
25. [ホーム]タブ→[1 行目をヘッダーとして使用]をクリックする



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

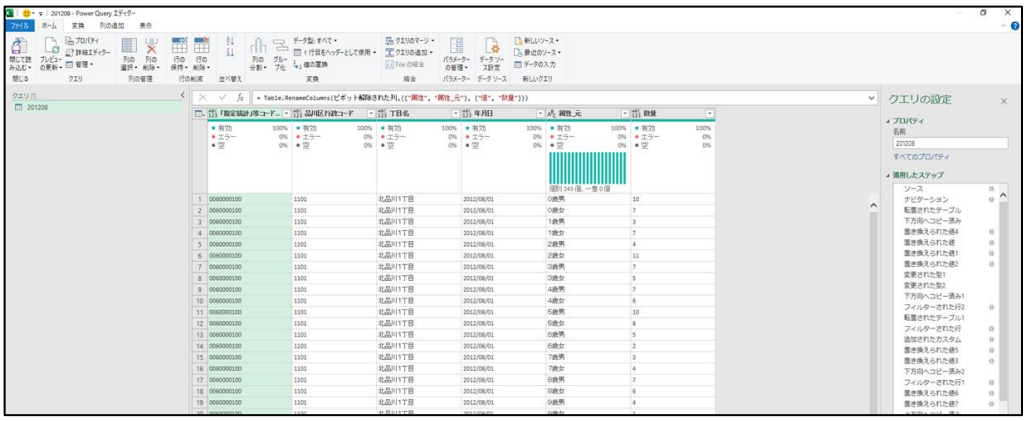


27. 対象列が選択された状態で右クリックし、[そのほかの列のピボット解除]をクリックする



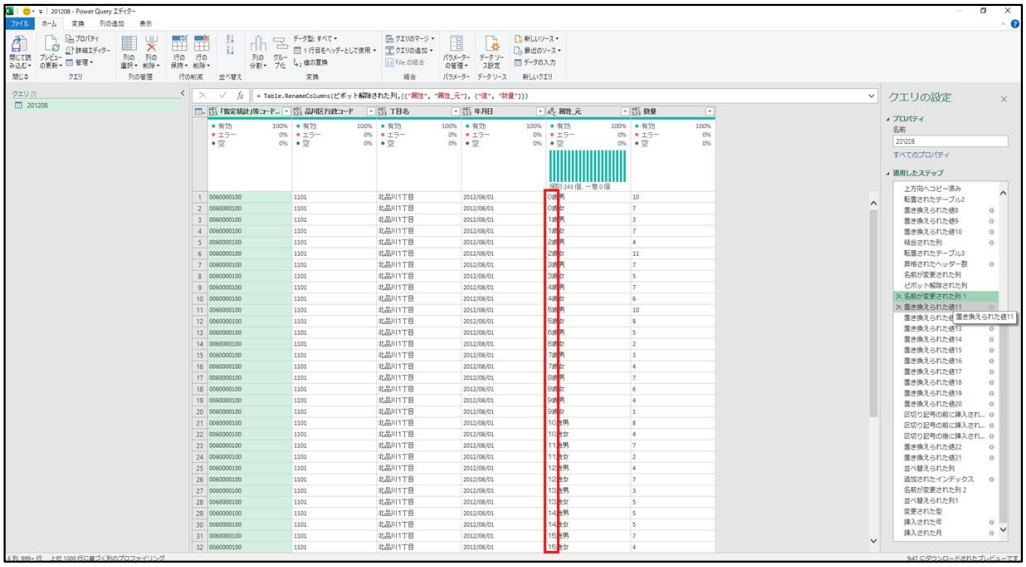
ⓒ東京都

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

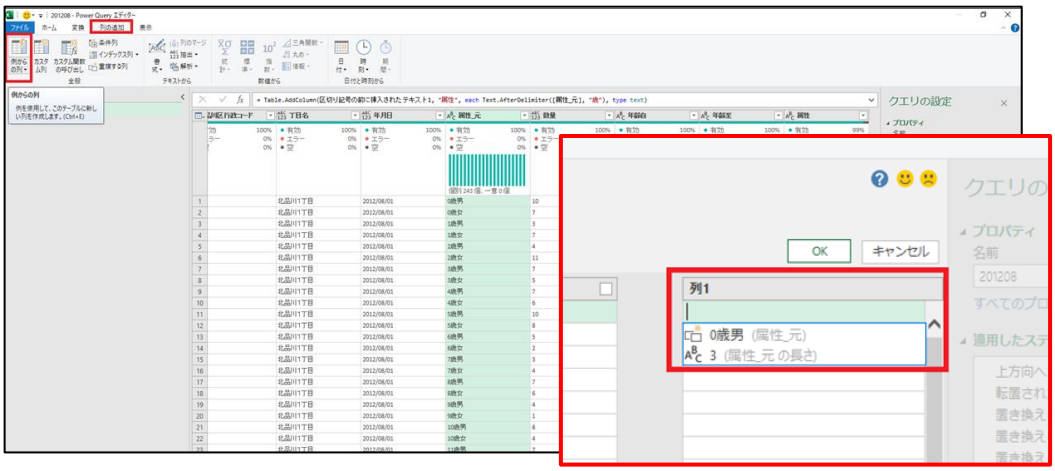
置換する値「全角数値０」 → 置換後「半角数値０」
<この間の数字分も、手順を繰り返します。>
置換する値「全角数値９」 → 置換後「半角数値９」



ⓒ東京都

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

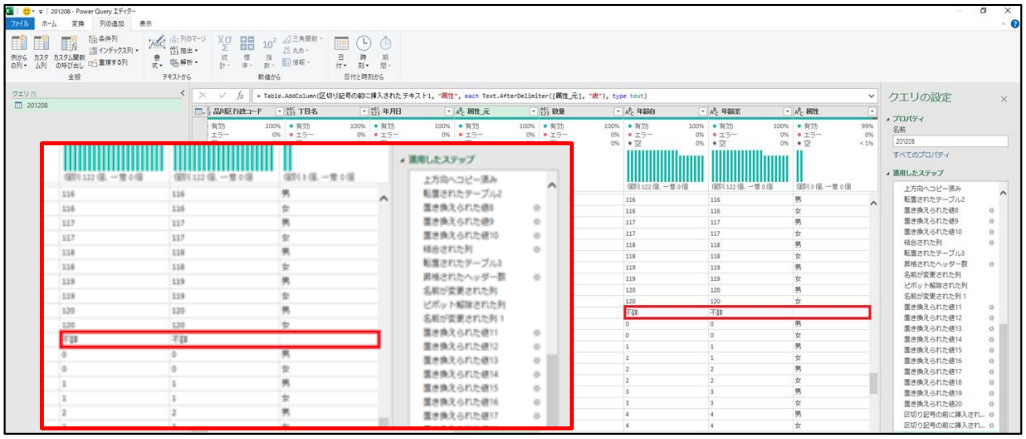
対象列：属性、年齢自、年齢列追加　属性_元列
それぞれの列から必要範囲を切り出して例を作成します。
全データで想定通りの値が入力されているか確認してください。



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

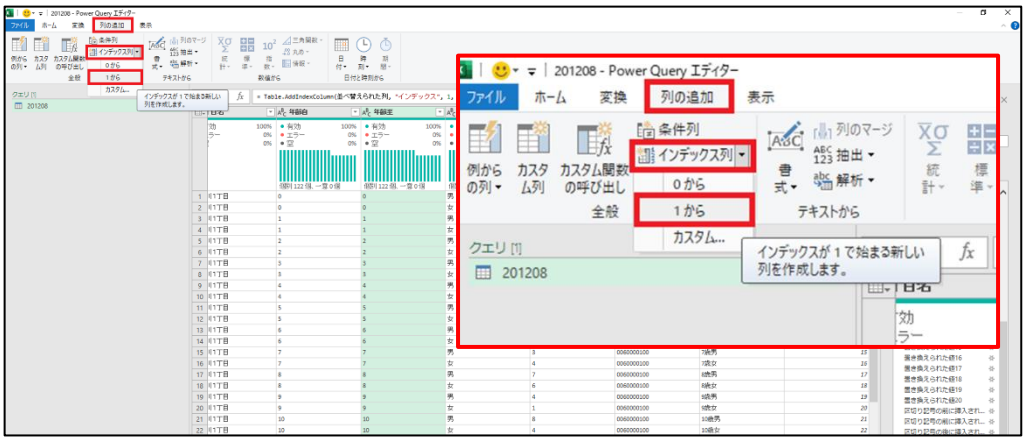
置換する値「不詳」　→　置換後「999」
置換する値「""」（ダブルクォーテーション）　→　置換後「その他」

不詳データを整形しています。



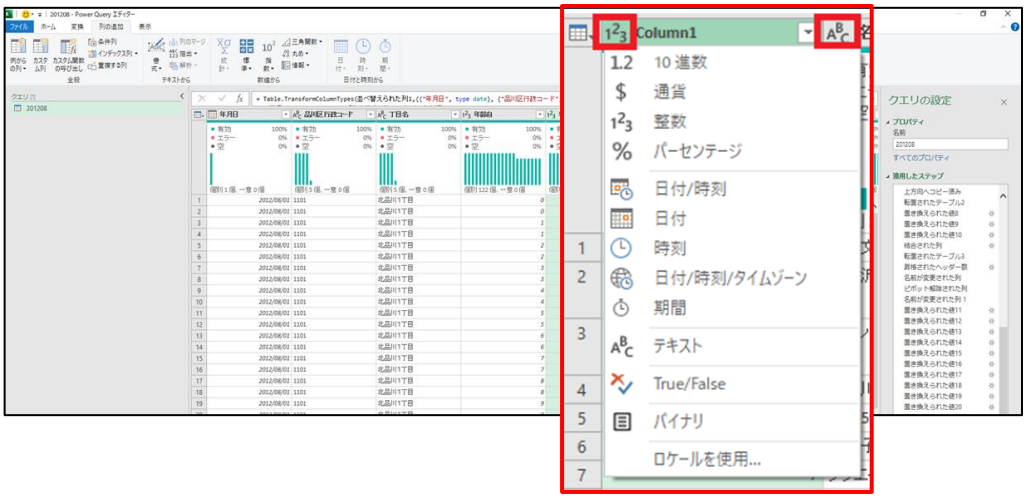
ⓒ東京都

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

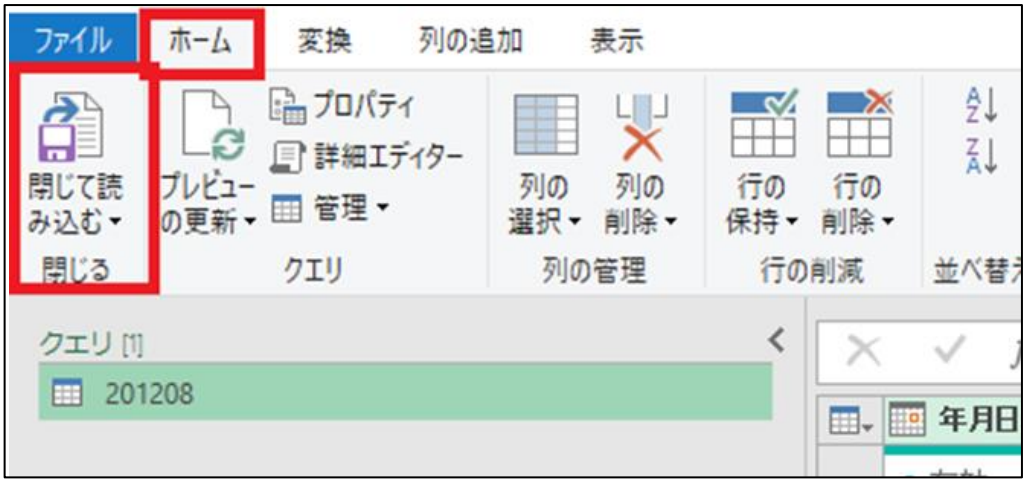


インデックス列を追加しています。

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

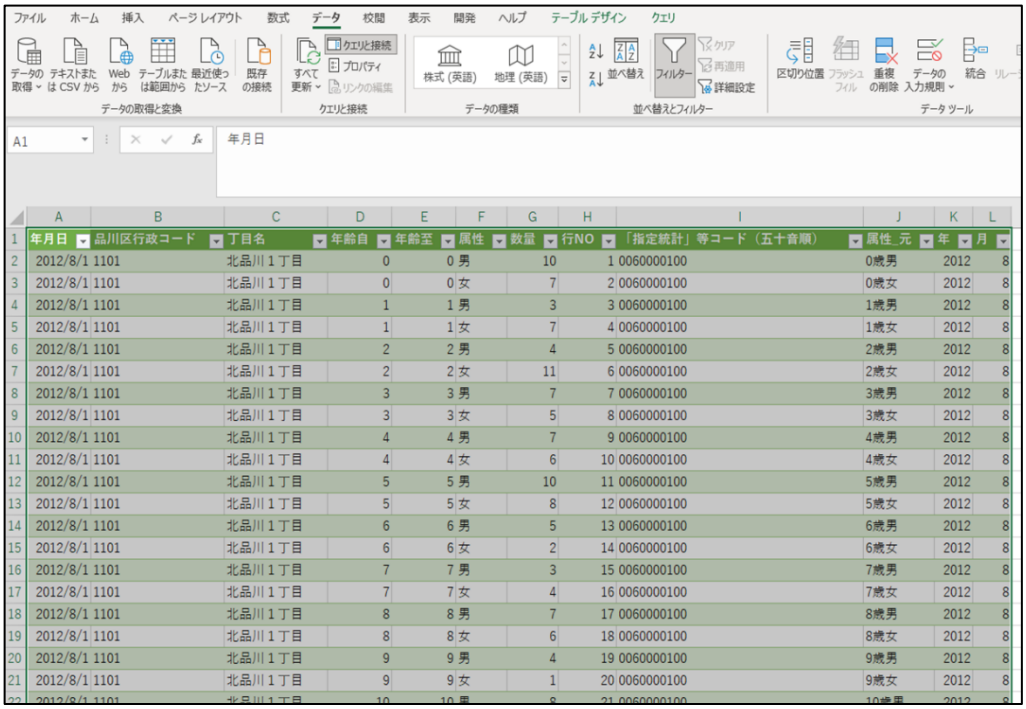


34. [ホーム]タブ→[閉じて読み込む]をクリックする



ⓒ東京都

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

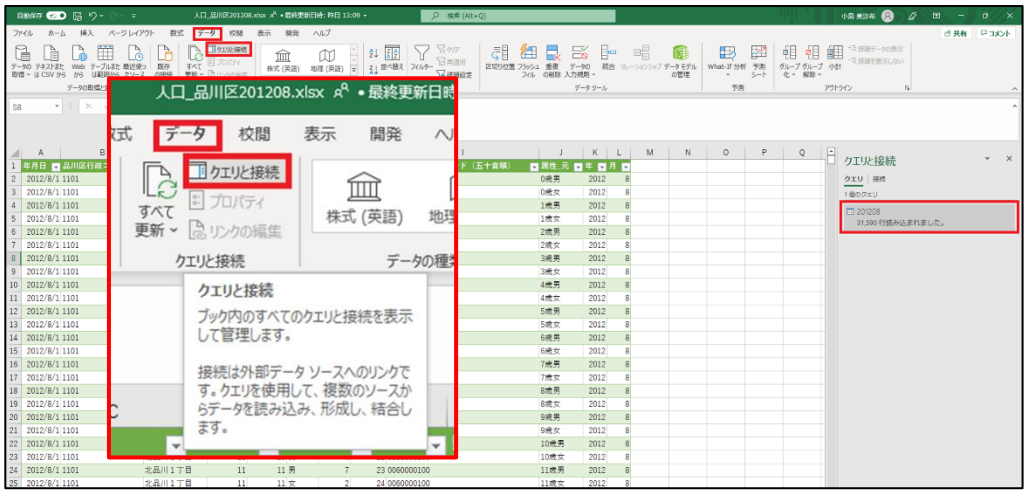
ⓒ東京都

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

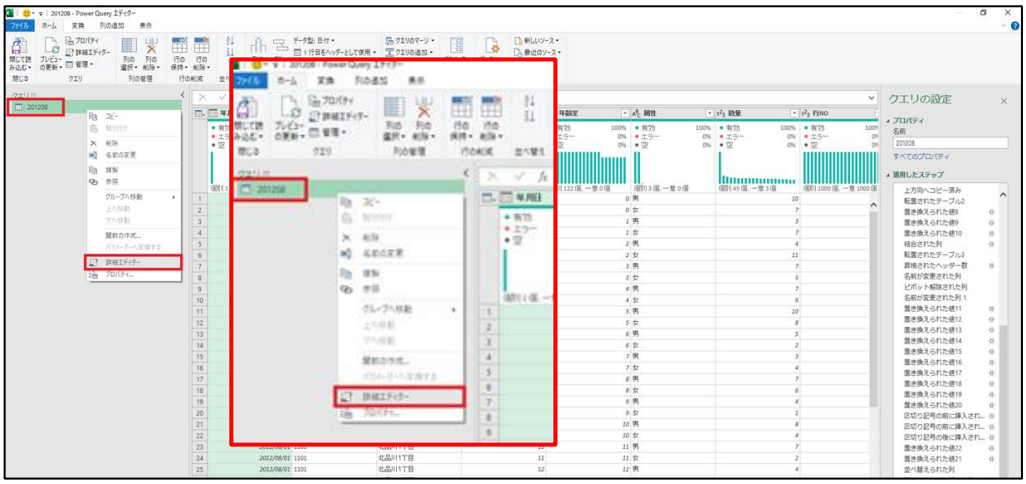
①②で処理した Power Query の詳細エディターを複製し、月毎等でエクセルブックが分かれている場合に発生する定期発生データも変換を行います。

【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

ⓒ東京都

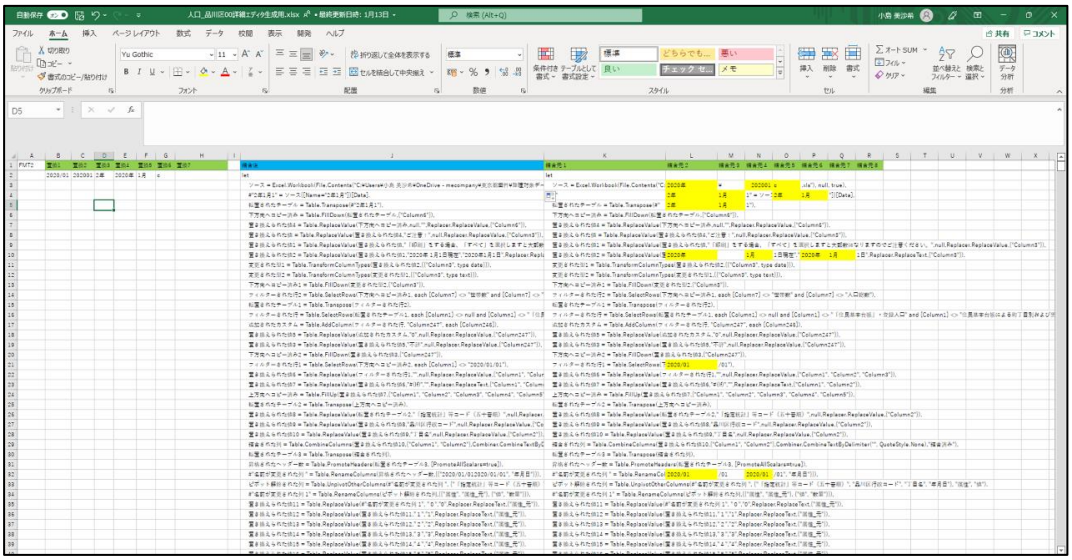
4. 新規エクセルブックで詳細エディターの編集を行う

●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

●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。

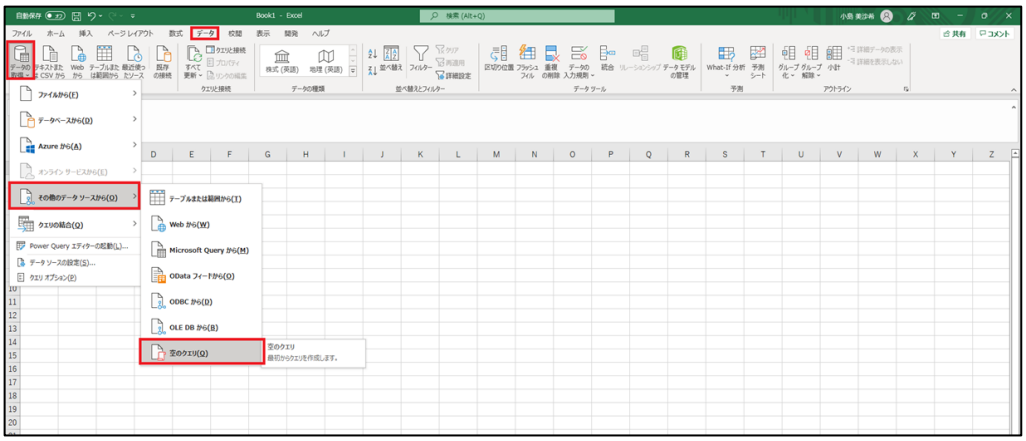


行については下記のように表記されます。
2 行目　ソース=…　：フォルダ名、ブック名記載箇所
3 行目　#"2 年 1 月 1" = ソース{[Name="2 年 1 月"]}[Data], ：シート名記載箇所

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. 147）もご参考ください。

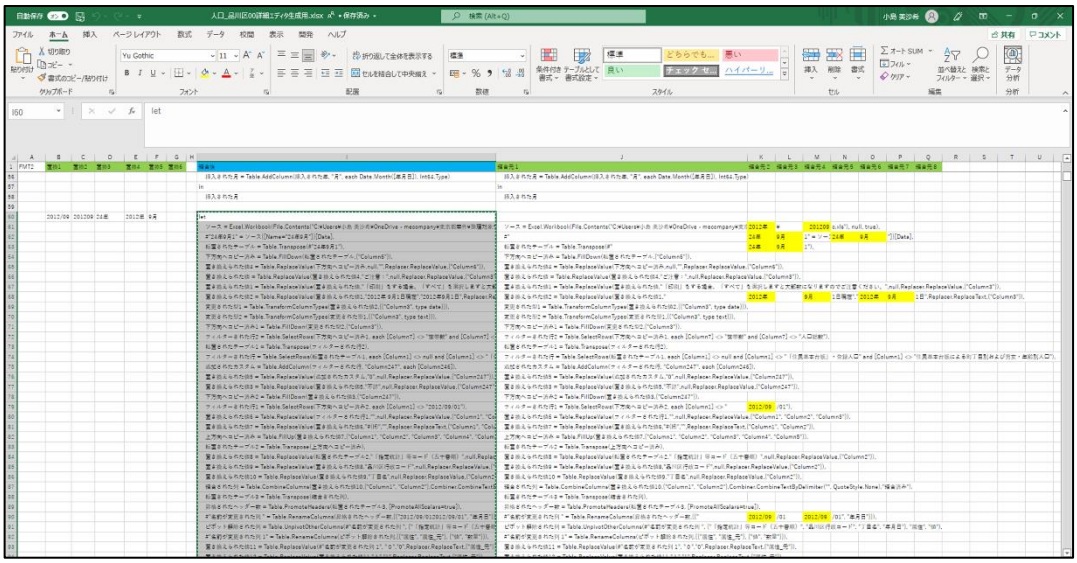
ⓒ東京都

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

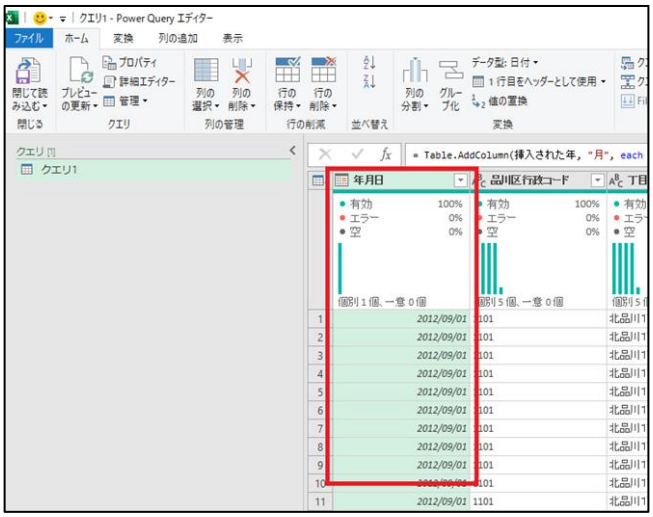
7. 詳細エディター生成用エクセルブックで、該当年月の J 列をコピーする



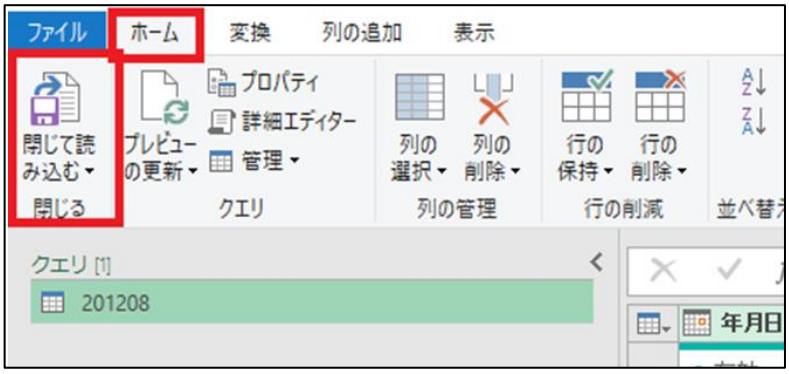
ⓒ東京都

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

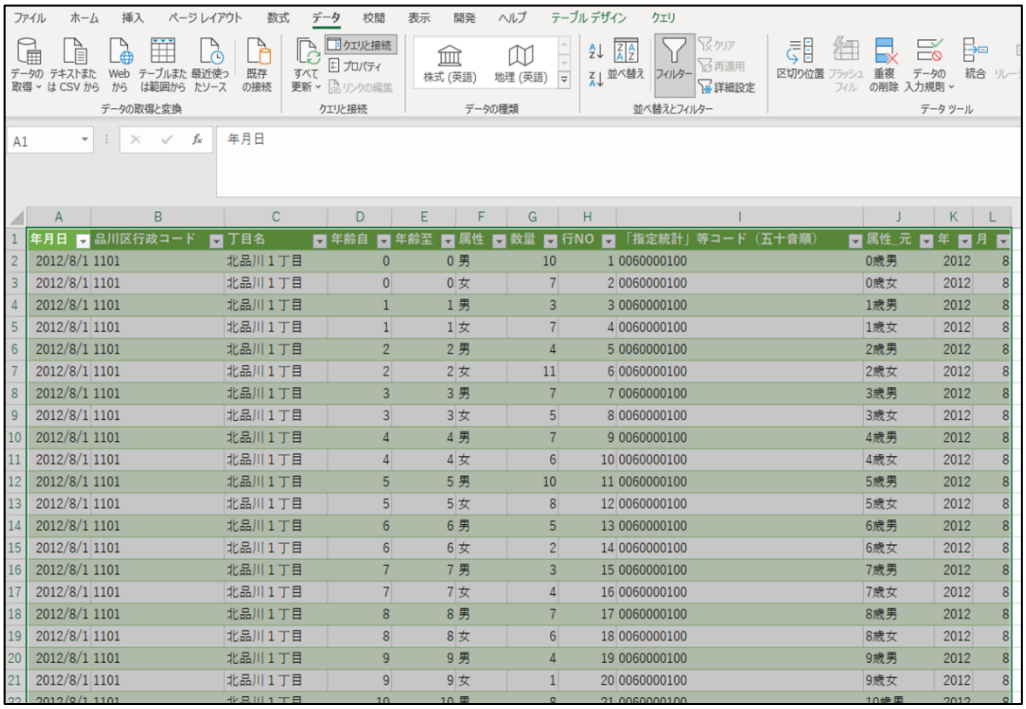


10. [ホーム]タブ→[閉じて読み込む]をクリックする



ⓒ東京都

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

ⓒ東京都

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。

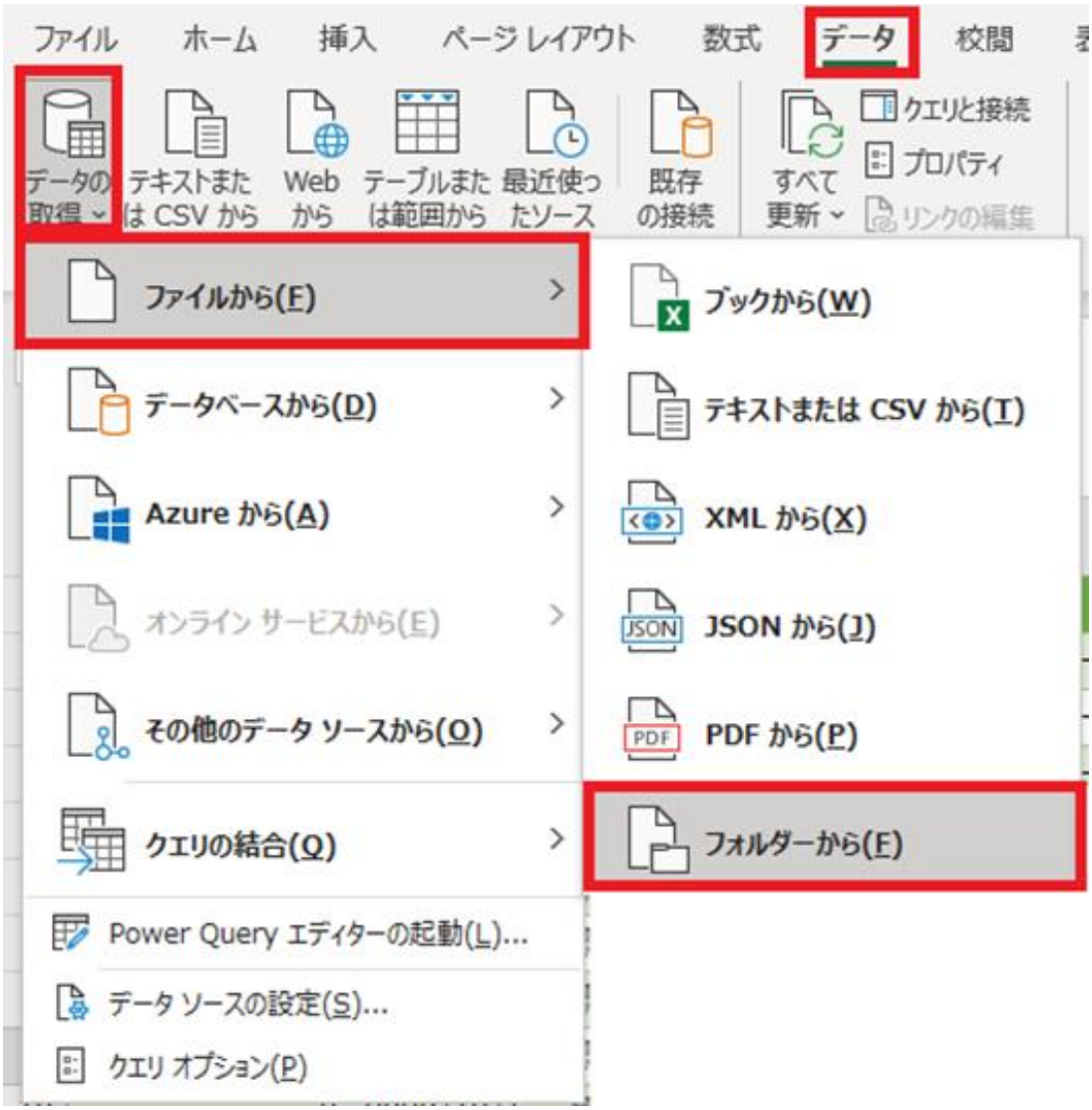
【お願い】： データ取得元（ソース）のファイルは閉じた状態で作業を行ってください。

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. 51）
② **Excel の Power Query 機能で、データを整形します。** （詳細： p. 53）

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**

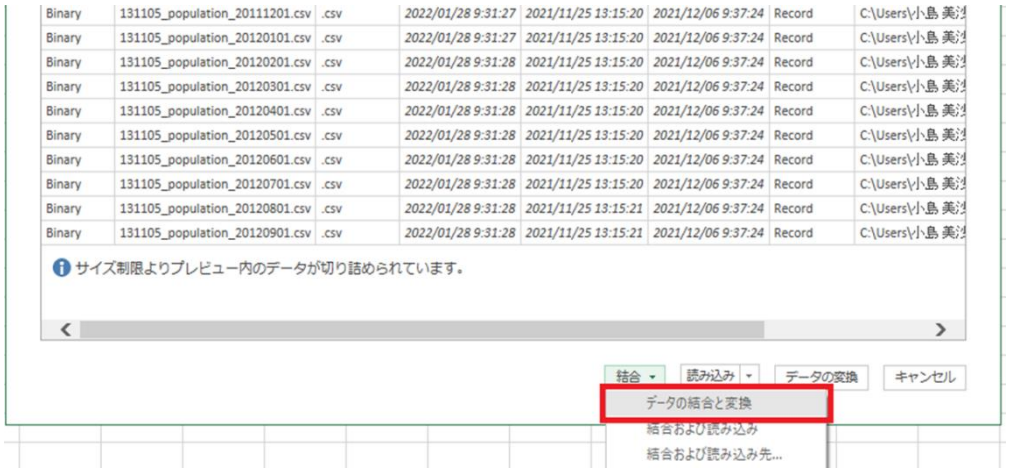
2. **[データ]タブ→[データの取得]→「ファイルから」→「フォルダから」をクリックする**



ⓒ東京都

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

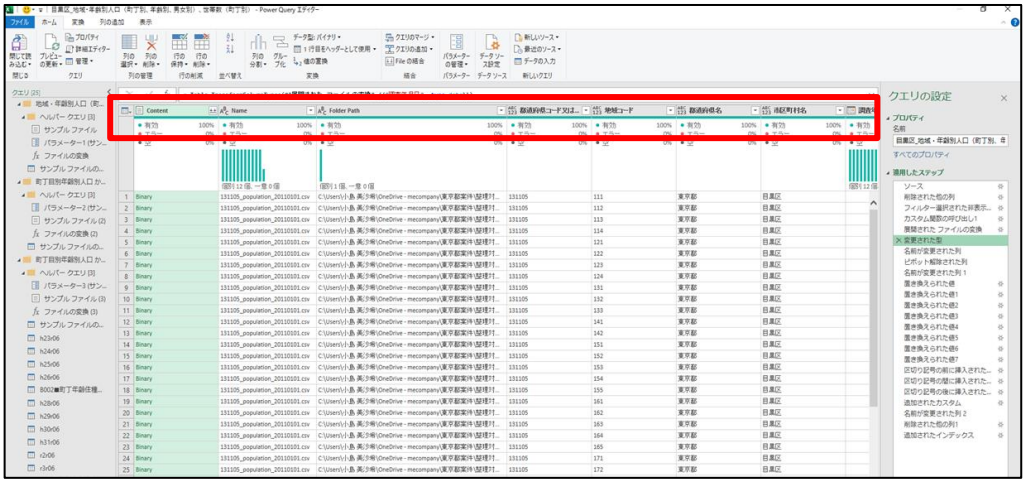


5. [OK]をクリックする

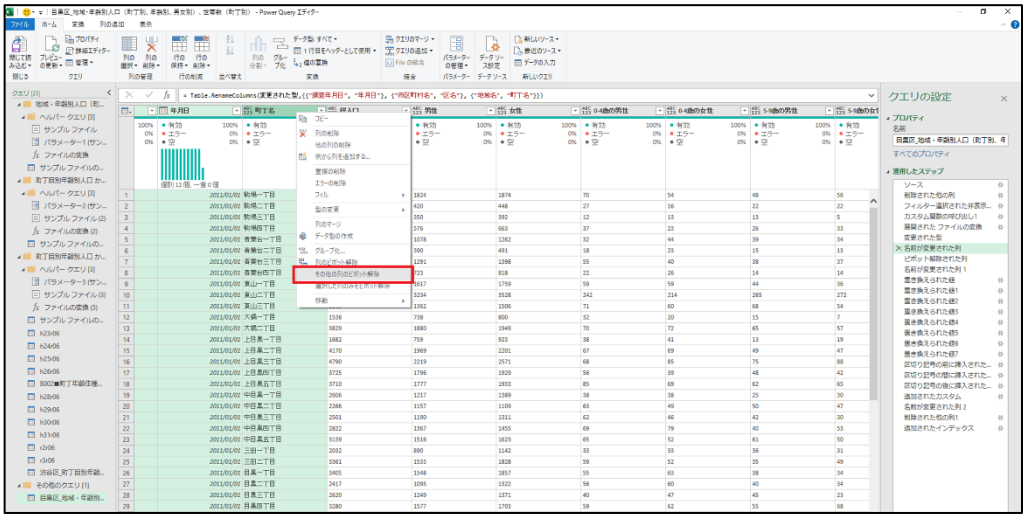
ⓒ東京都

② Excel の Power Query 機能でデータを整形する

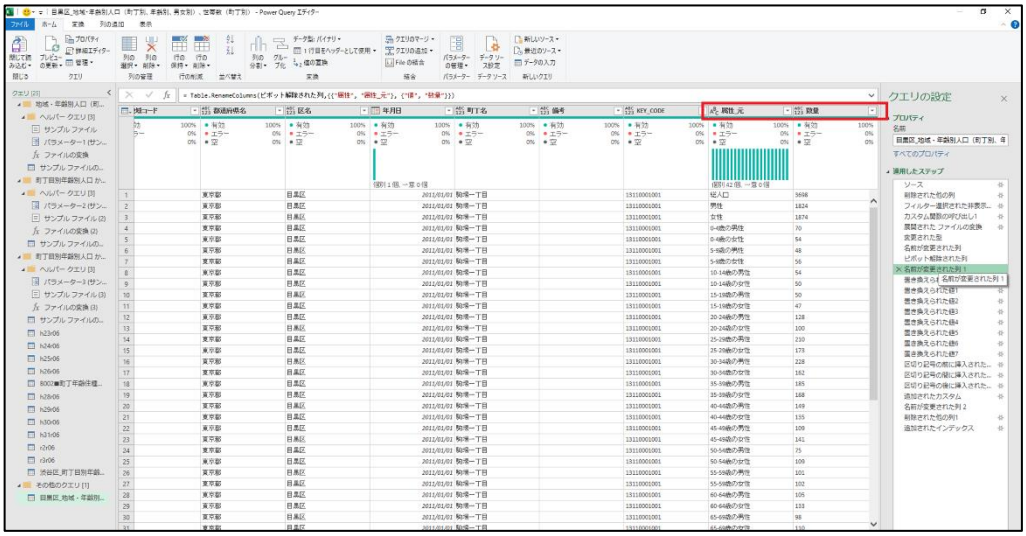
1. 「列名」をダブルクリックし、任意の列名を入力する



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



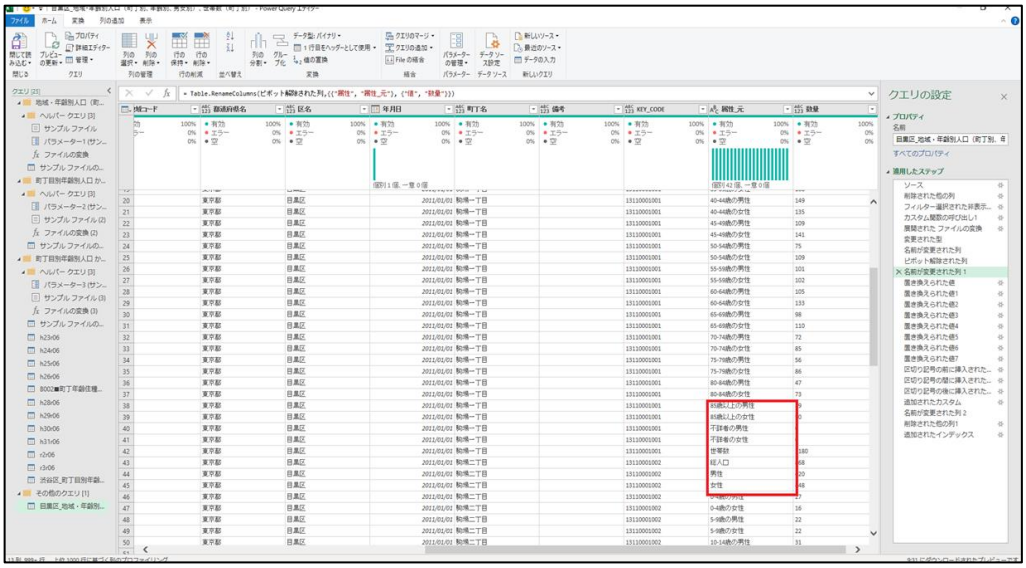
3. ピボット解除により作成されたヘッダー列名に、任意の列名を入力する



ⓒ東京都

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」



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。

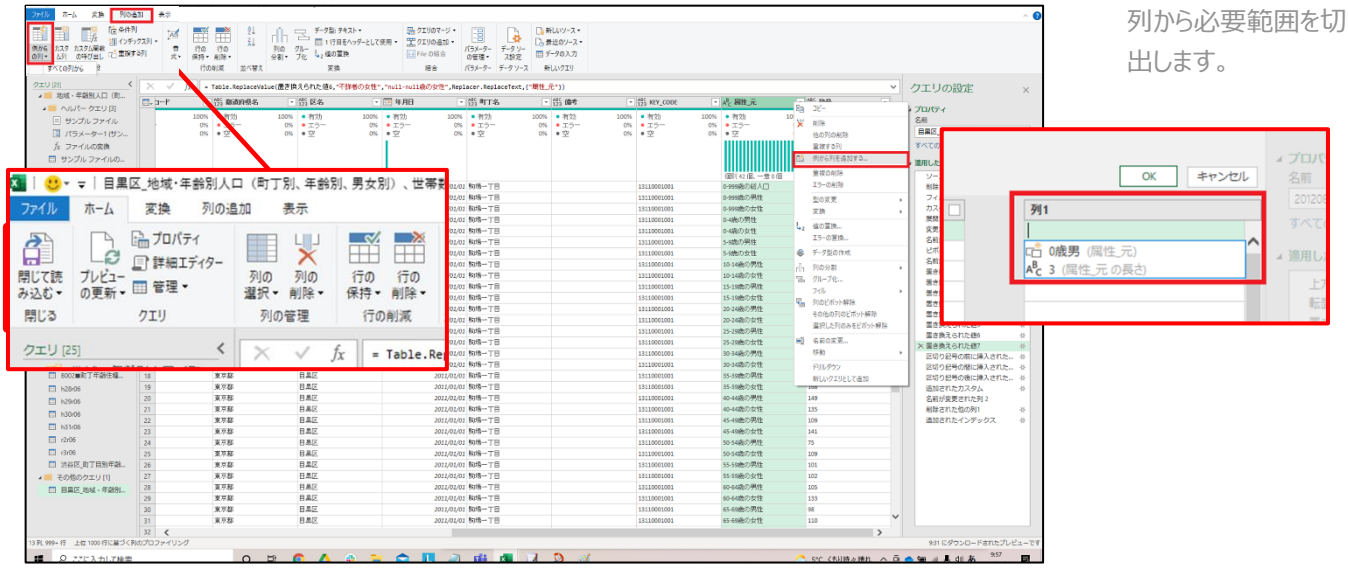
値の置換の際、"男性"、"女性"等、置換を行いたくない他のセル内容の一部と合致してしまう場合は、「セルの内容全体の照合」のチェックを ON にします。

ⓒ東京都

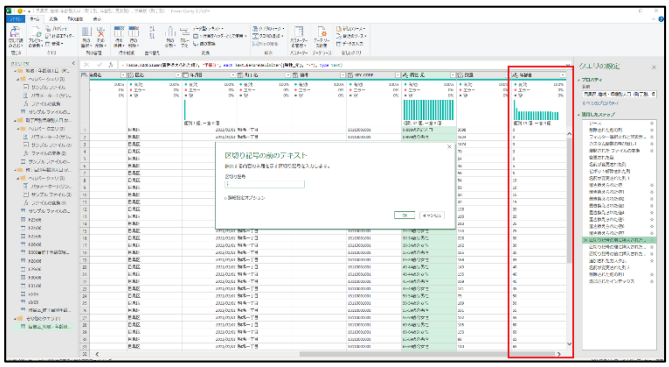
5.　対象列が選択された状態で右クリックし、[列の追加]→[例からの列]→[選択範囲から]をクリックし、例を作成する

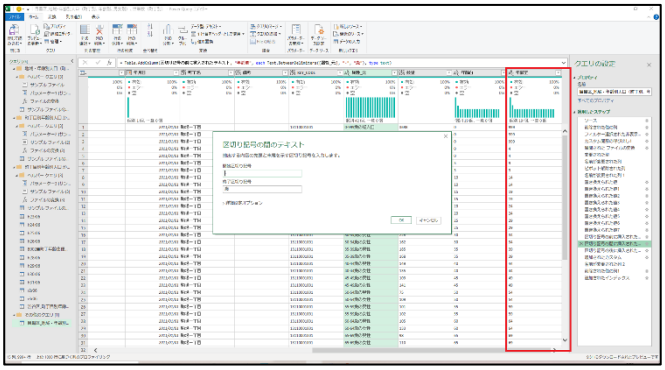
追加列のセルをダブルクリックし、選択セルのデータ内容をクリック、任意の値に変更して例を作成します。
作成後は、全データ想定通りの値が入力されている事を確認します。

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。

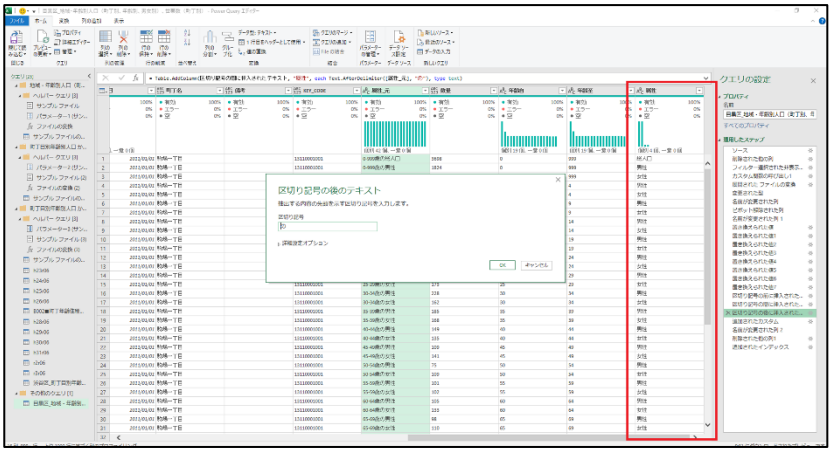


例からの列(年齢自列)追加後の画面（左）および例からの列(年齢至列)追加後（右）です。





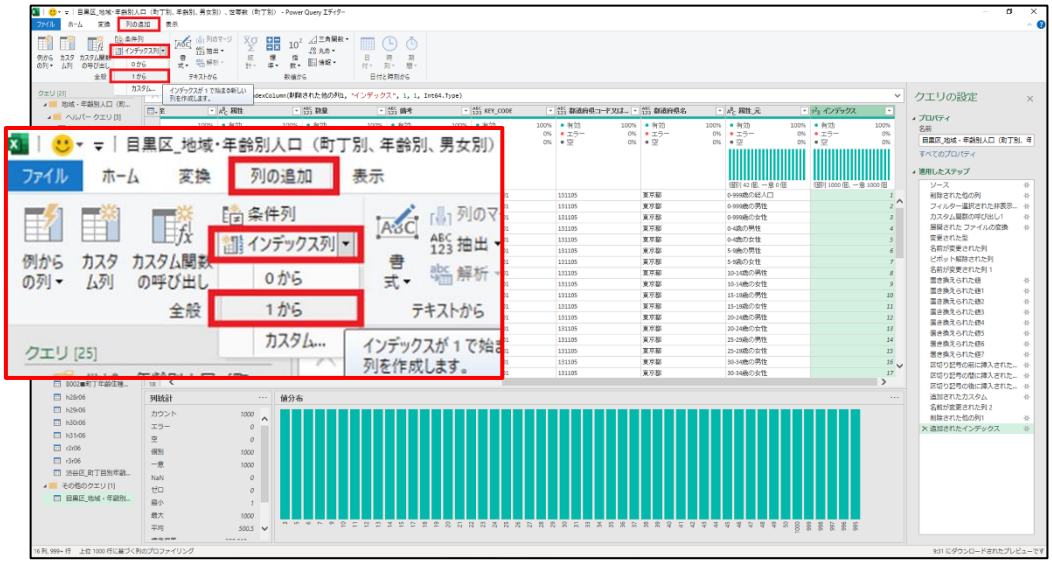
例からの列(年齢自列)追加後の画面です。



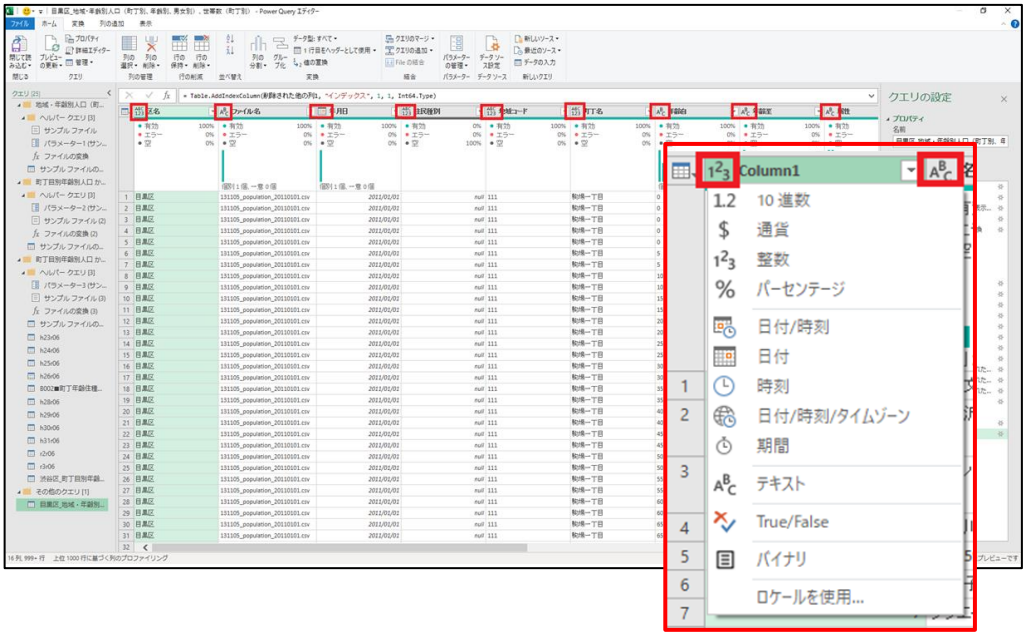
ⓒ東京都

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

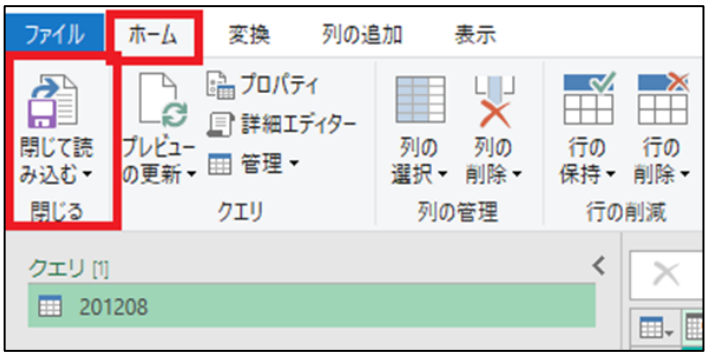
インデックス列を追加しています。



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

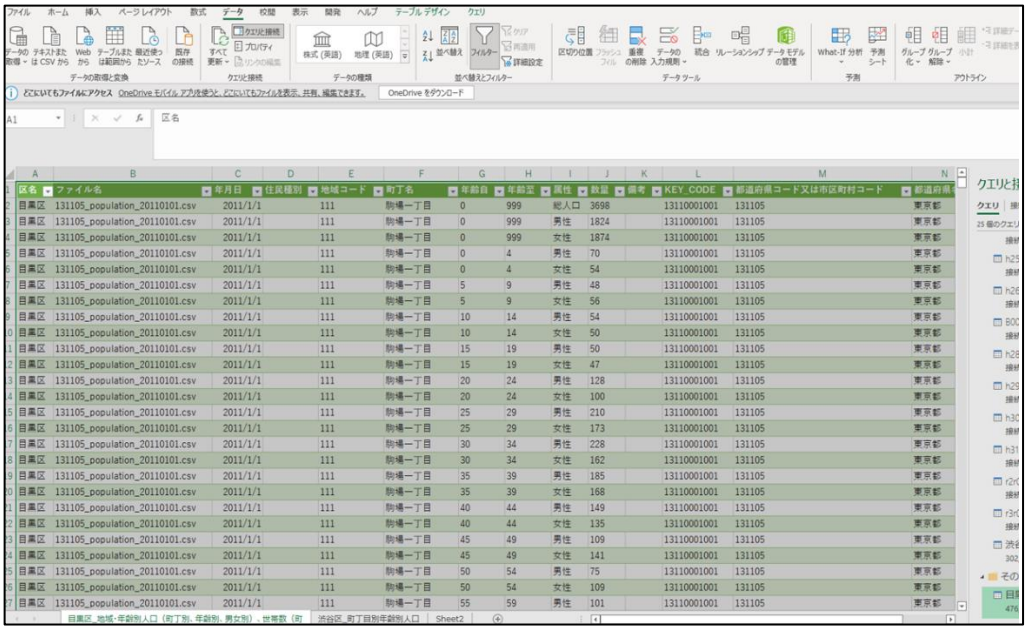


8. [ホーム]タブ→[閉じて読み込む]をクリックする



ⓒ東京都

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

ⓒ東京都

手順（ケース[C]）

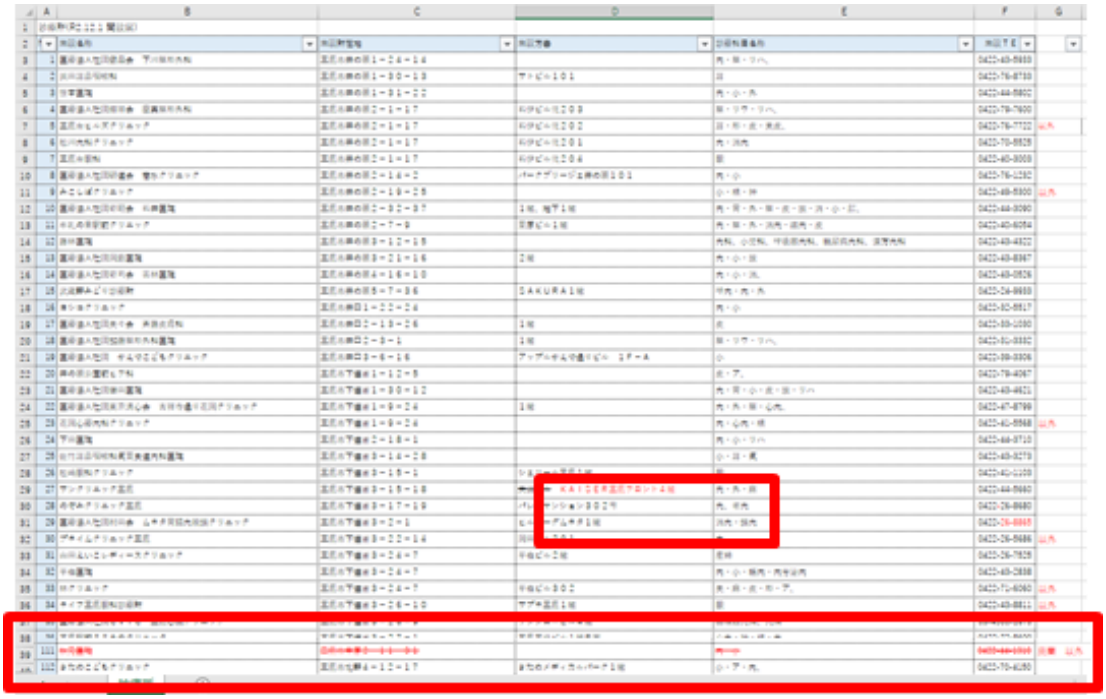
下記の手順で Excel から CSV にデータを変換します。

【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. 58）
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. 61）
③ Excel の Power Query 機能で、データを整形します。 （詳細： p. 63）
　クエリパターン 1 （詳細： p. 64）
　クエリパターン 2 （詳細： p. 67）
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. 74）

① Excel マクロにより、文字装飾等を削除する前処理の実施

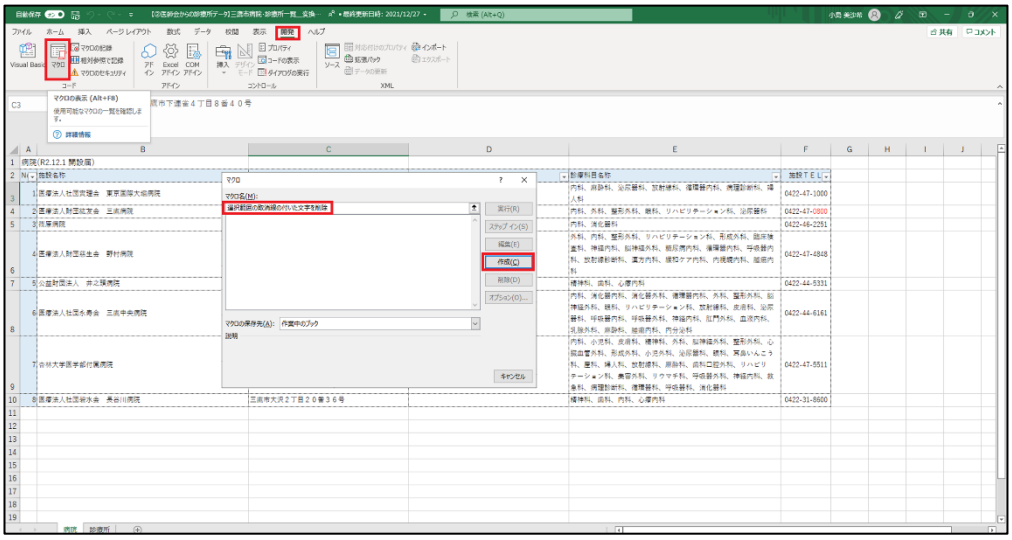
1. 取消線や装飾文字のある箇所を確認する



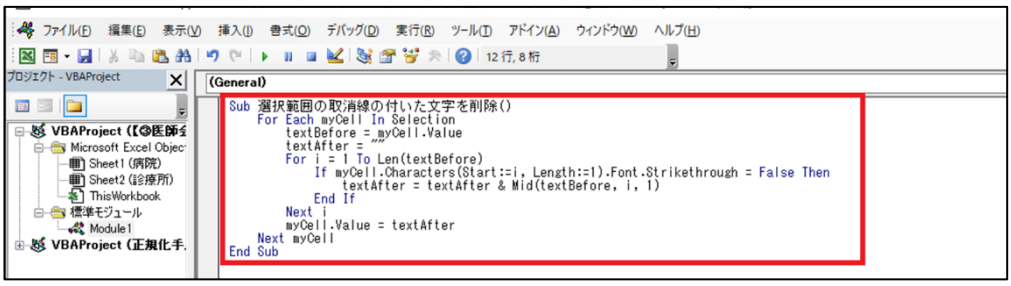
Excel に直接打ち込んだ、取り消し線等をマクロで削除します。

ⓒ東京都

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。

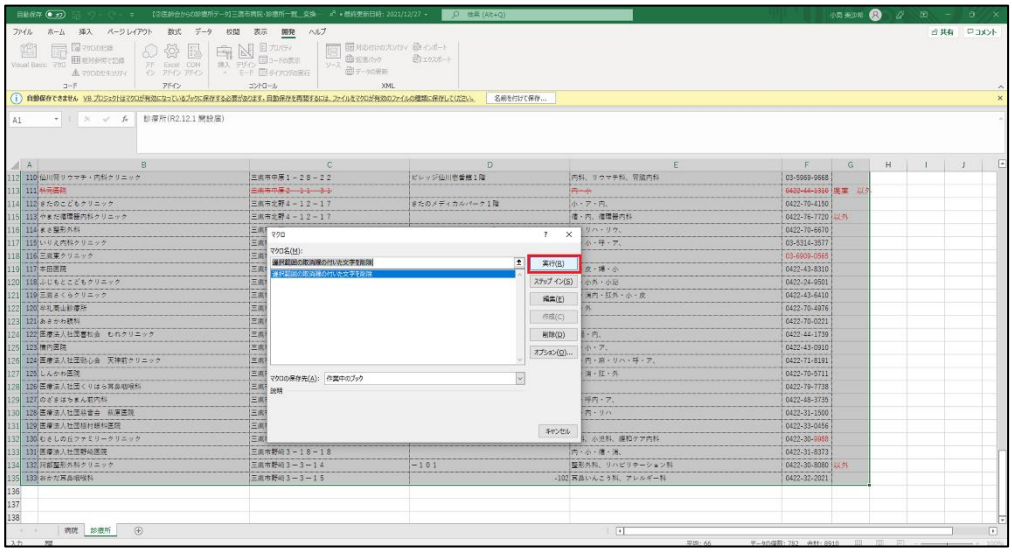
```
Sub 選択範囲の取消線の付いた文字を削除()
    For Each myCell In Selection
        textBefore = myCell.Value
        textAfter = ""
        For i = 1 To Len(textBefore)
            If myCell.Characters(Start:=i, Length:=1).Font.Strikethrough = False Then
                textAfter = textAfter & Mid(textBefore, i, 1)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

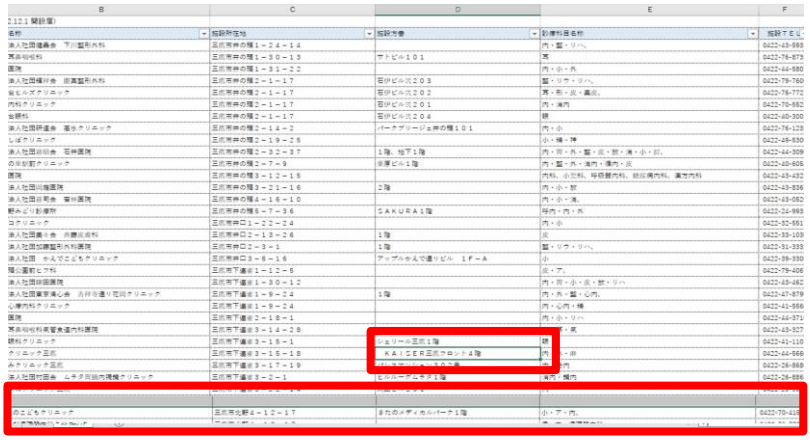
※上記 Excel マクロは Office365 にて動作確認

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

ⓒ東京都

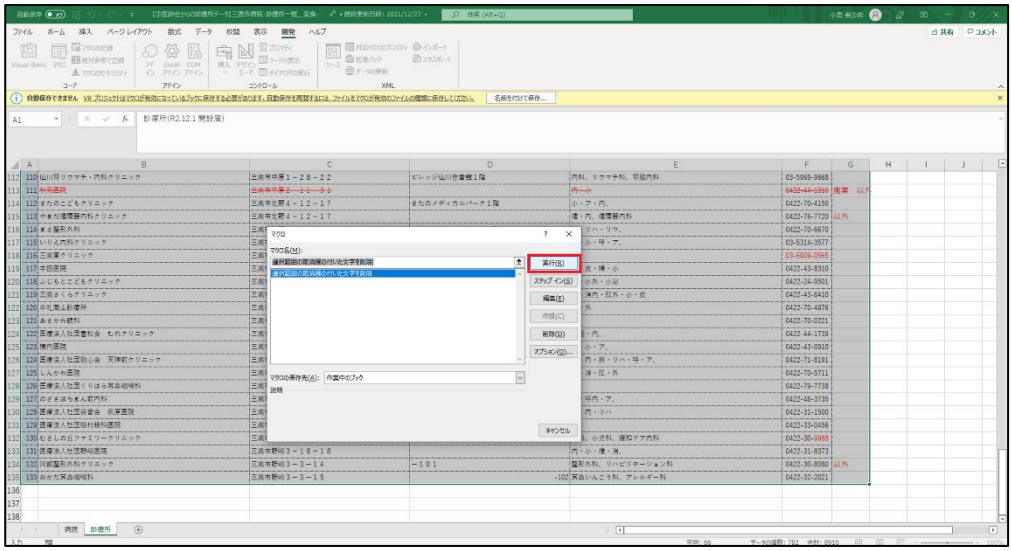
5. マクロを適用する範囲を選択し、「選択範囲の取り消し線の付いた文字を削除」マクロを選択し、[実行]をクリックする





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

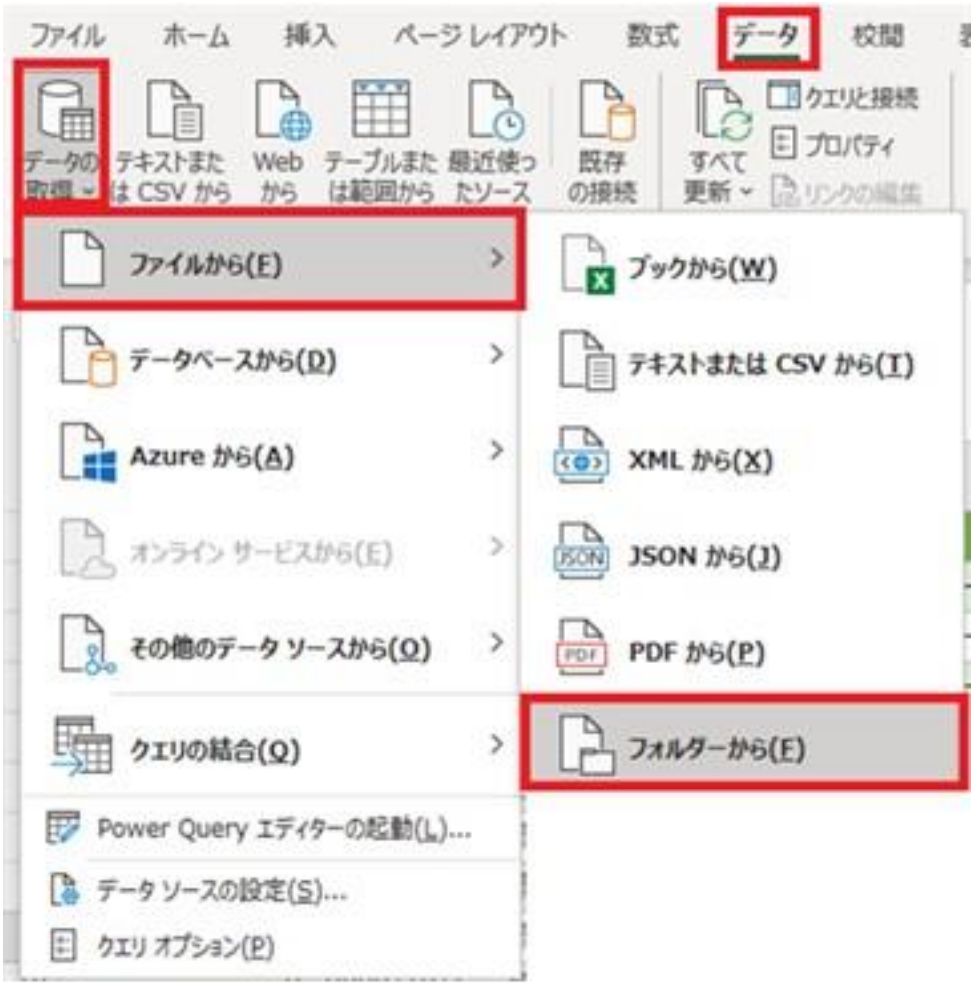


ⓒ東京都

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

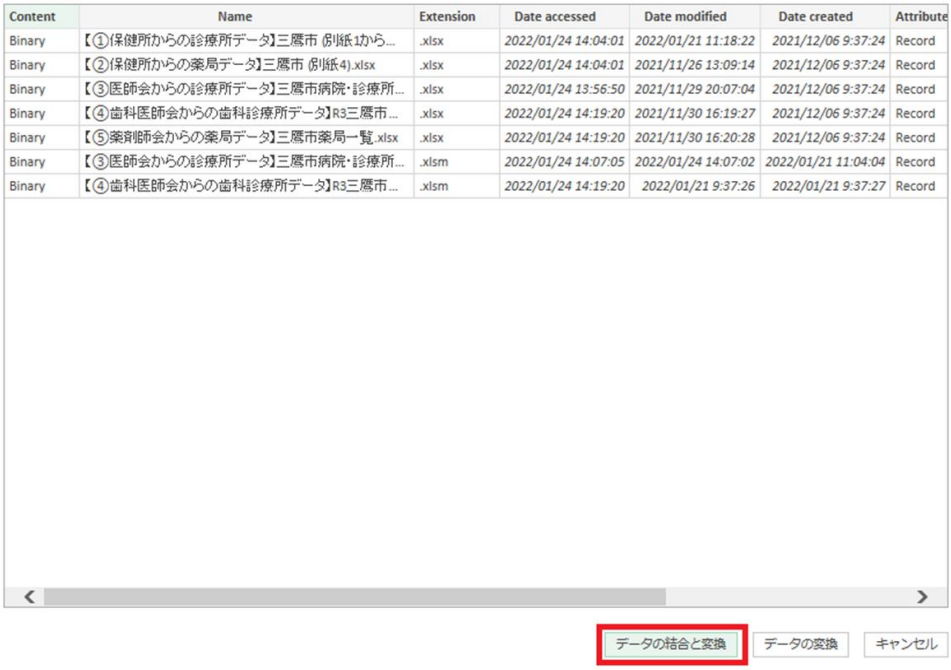


3. Excel ファイルをクリックして選び、[開く]をクリックする

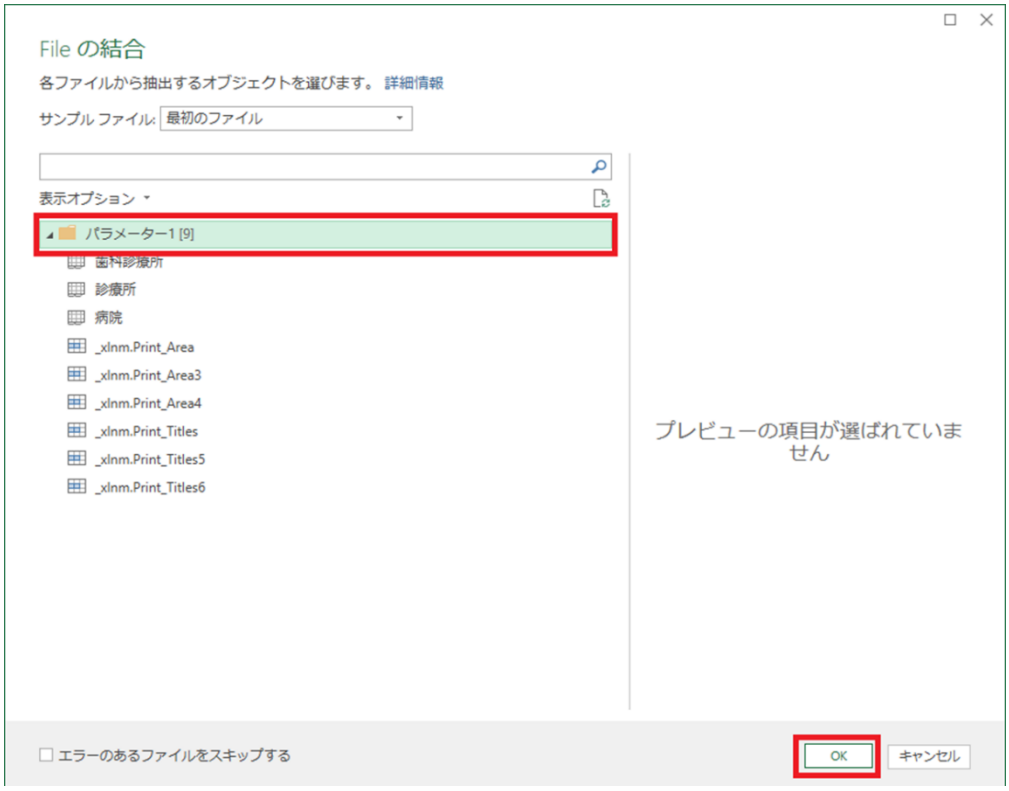
ⓒ東京都

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。



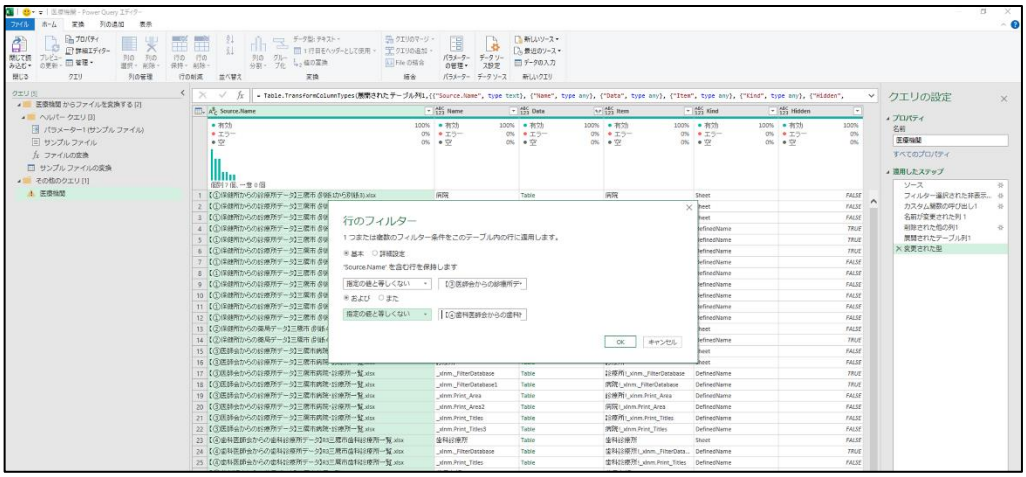
5. フォルダをクリックして選び、[OK]をクリックする



ⓒ東京都

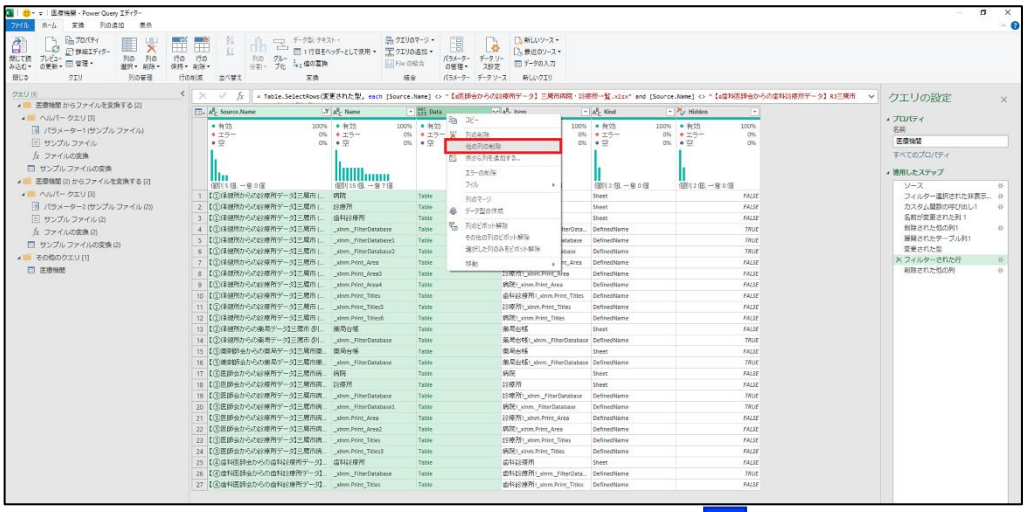
③ Excel の Power Query 機能でデータを整形する

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

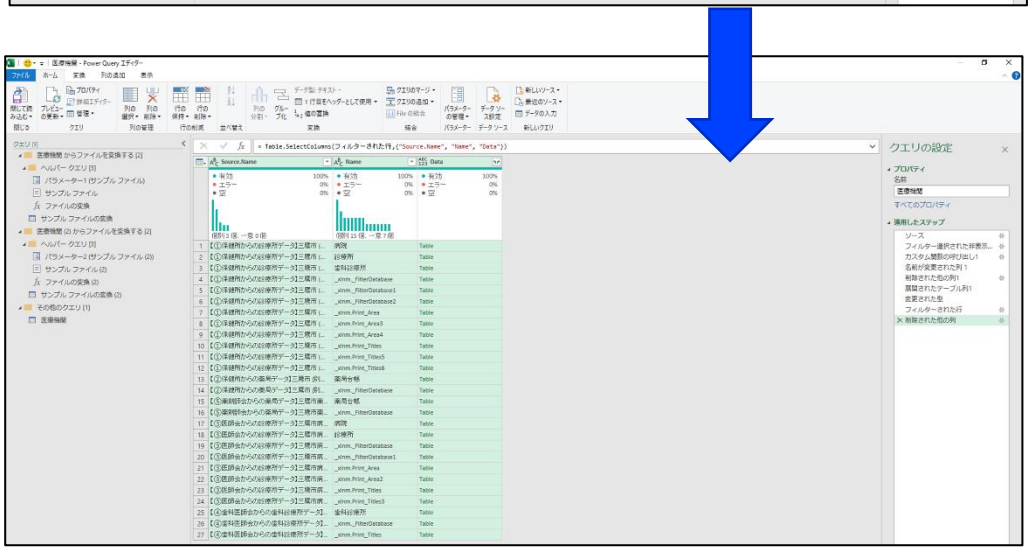


手順①でマクロ有(拡張子.xlsm)で保存したブックが存在する場合はマクロ無(拡張子.xlsx)を除外します。

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



対象列は、Source.Name、Source、Data 例を指します。

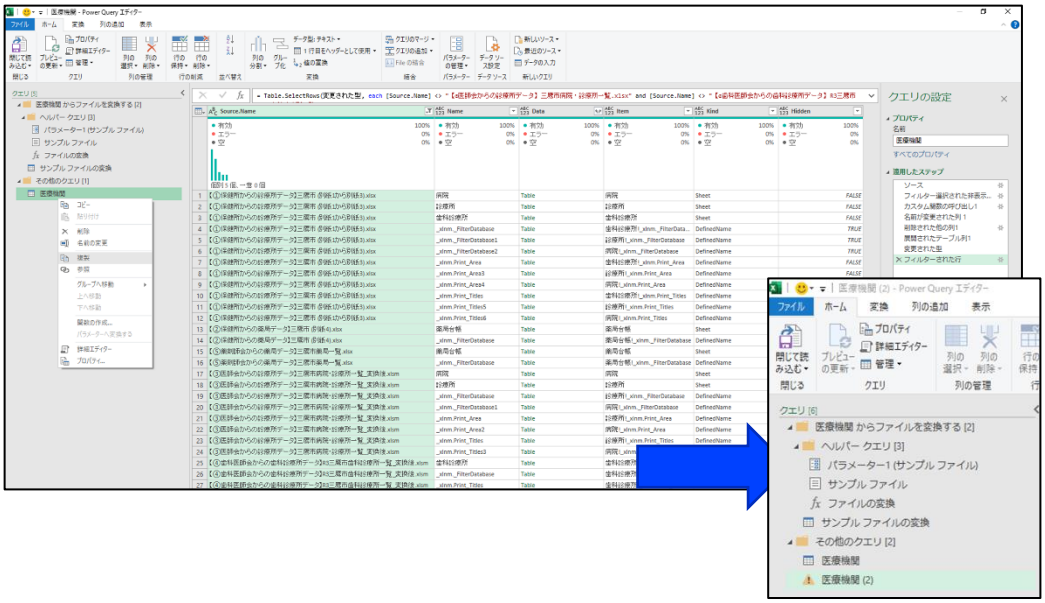


対象列が残り、不要列を削除した状態です。

ⓒ東京都

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

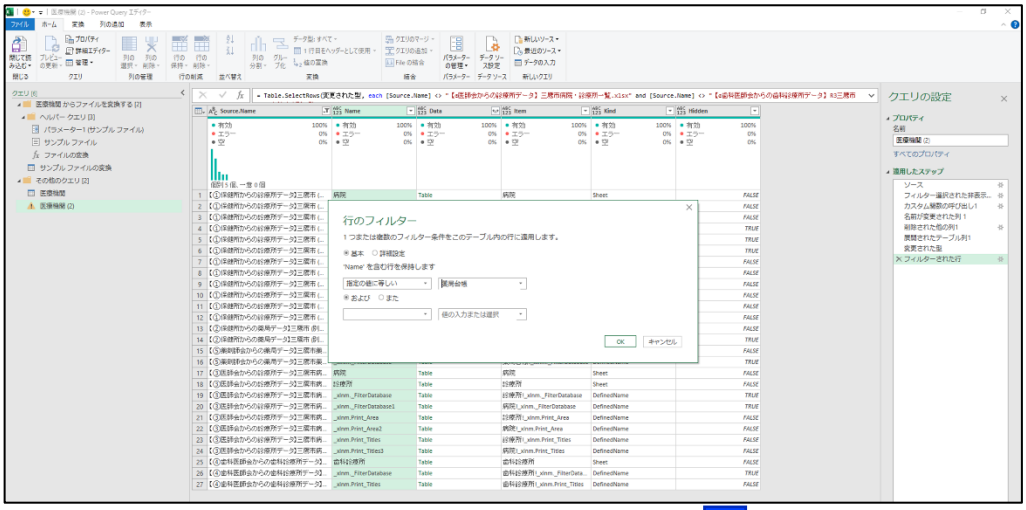
ヘッダーパターン分のクエリを複製します。



<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

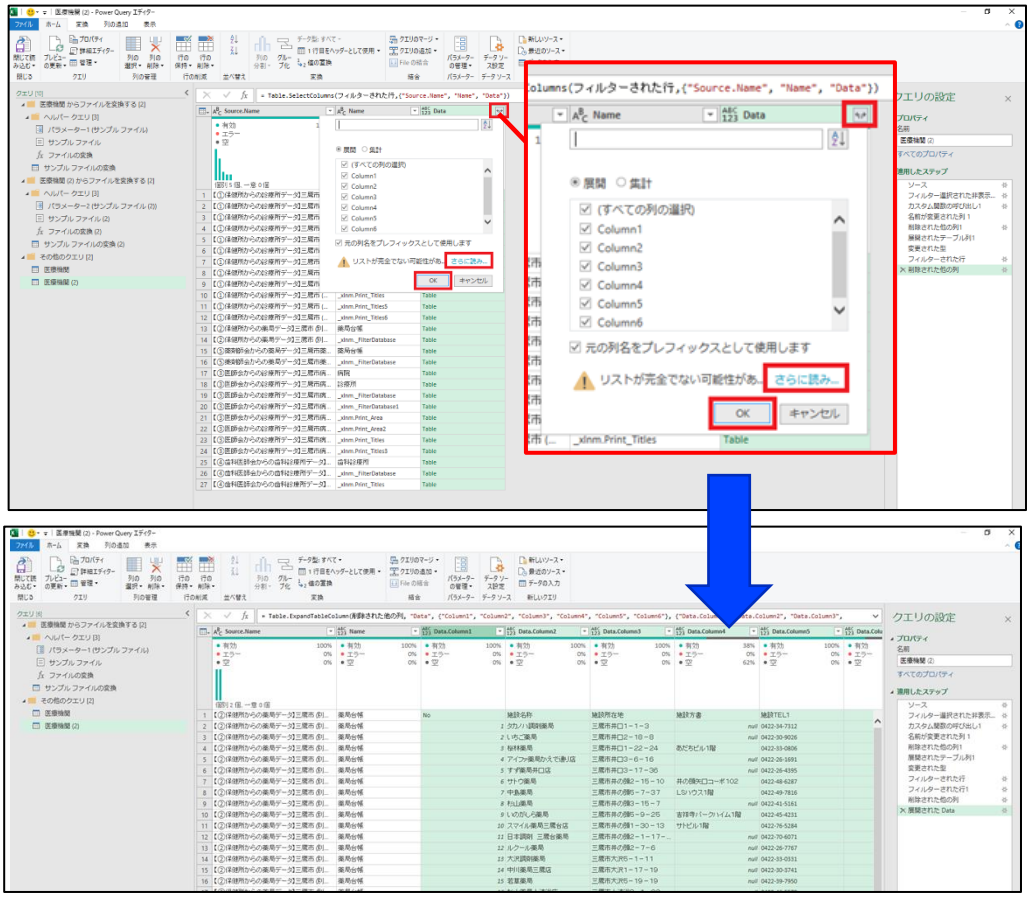
必要行を残すフィルターを設定しています。



必要行が残った状態です。

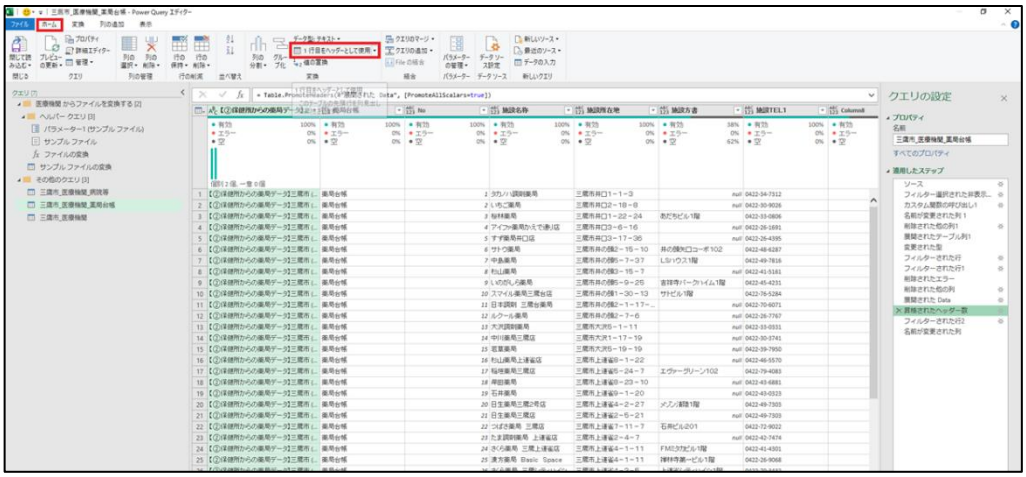
ⓒ東京都

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



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

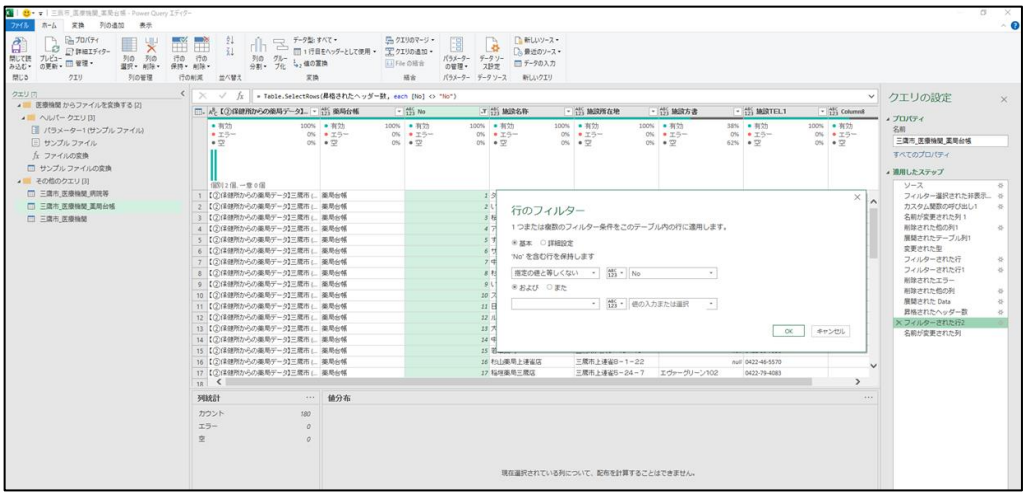


1 行目をヘッダーとする整形を行っています。

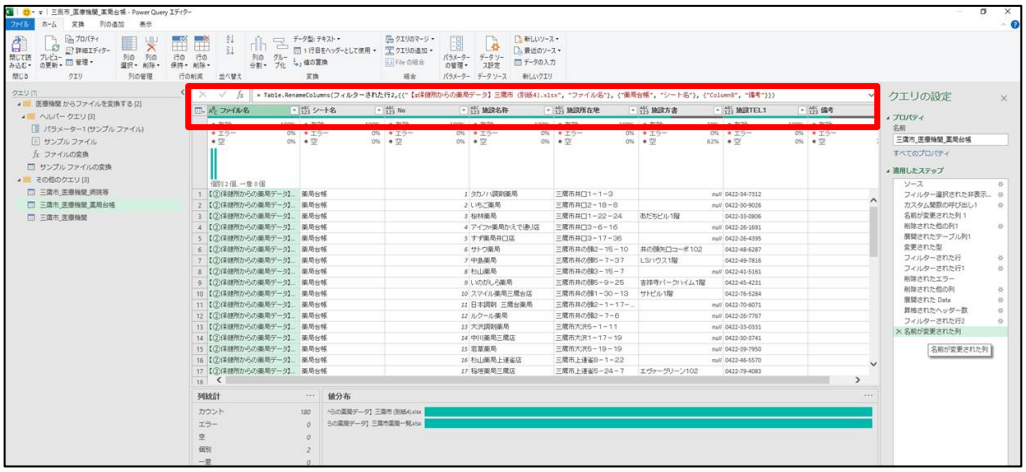
ⓒ東京都

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

不要行を非表示する処理をしています。

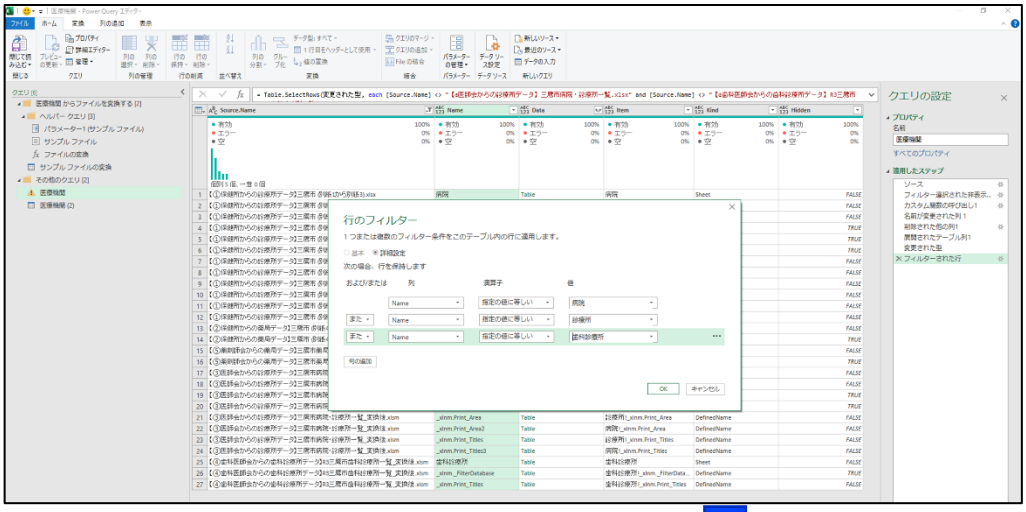


8.　追加した「カスタム列」や「列名」の名前を指定しなかった列名があれば、列をダブルクリックして任意の列名を入力する



ⓒ東京都

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

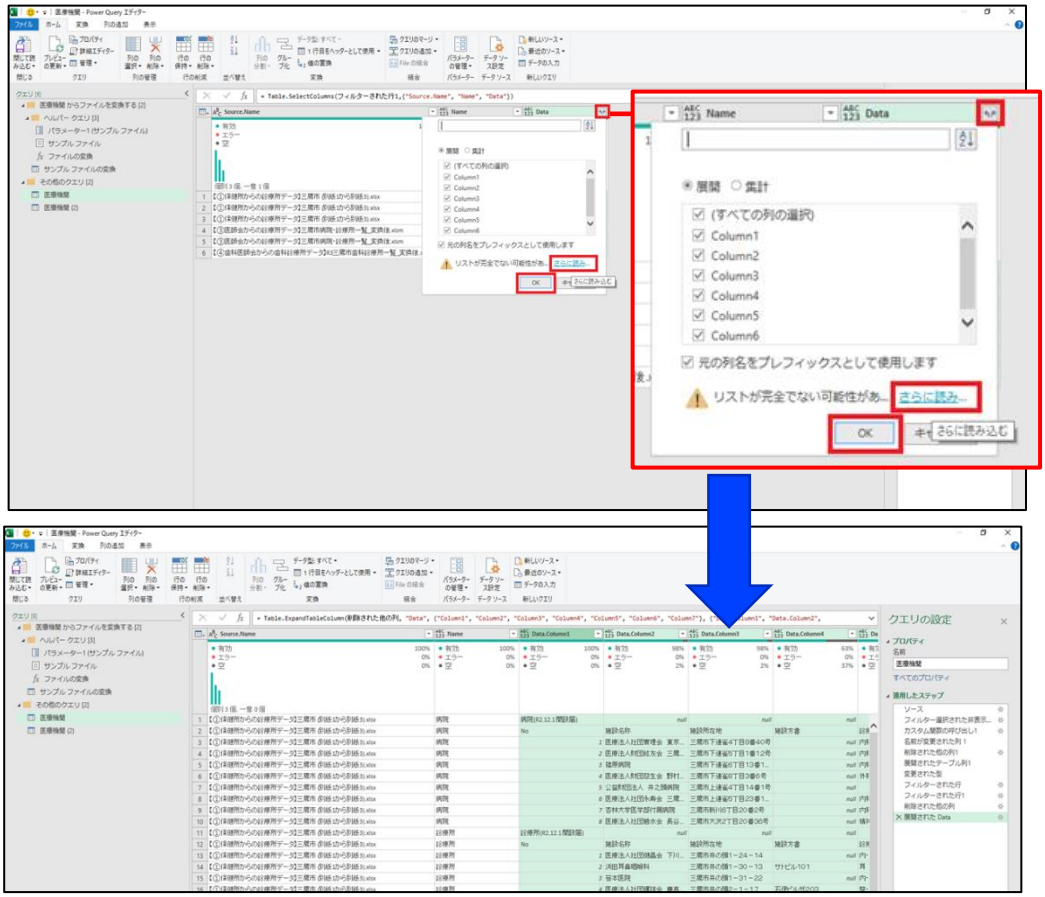


必要行を残すフィルターを設定しています。

必要行が残った状態です。

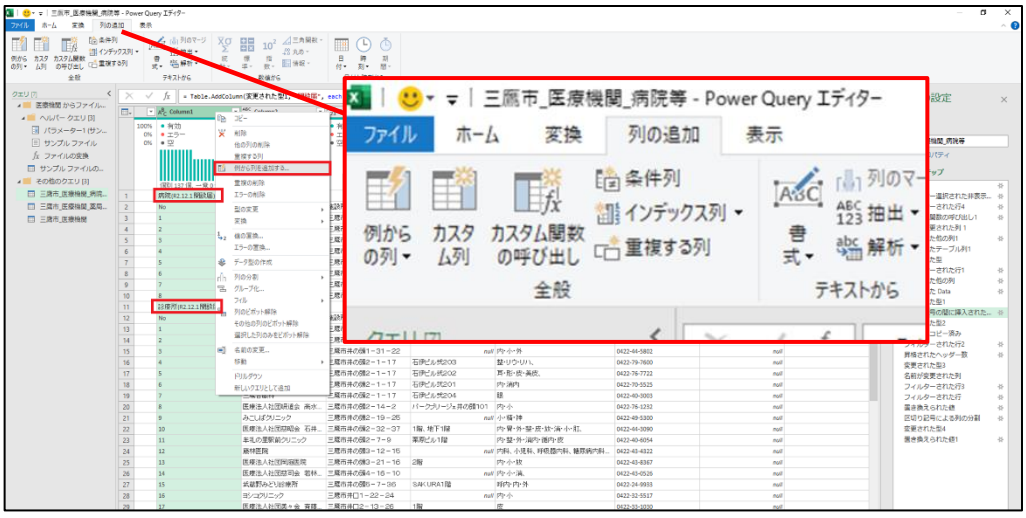
ⓒ東京都

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



Data 例を展開します。

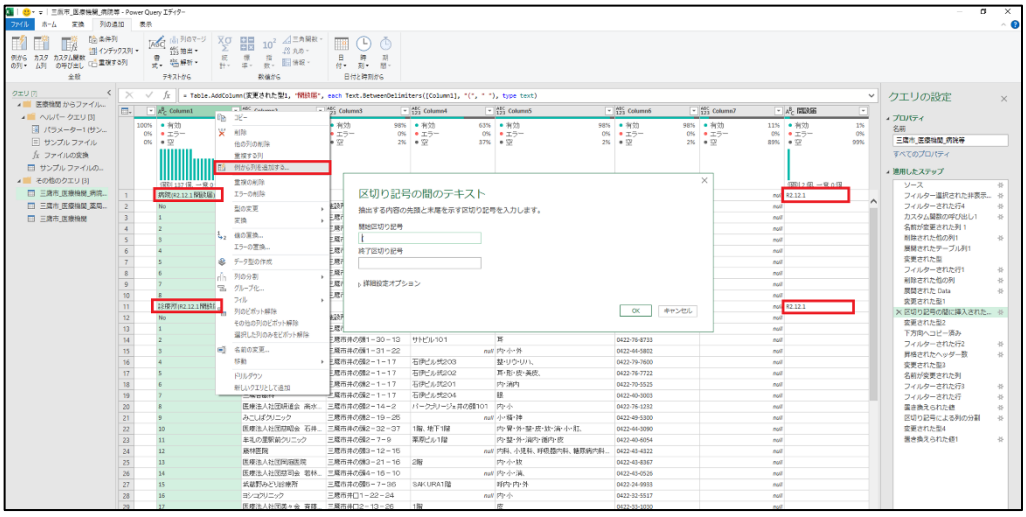
11. 開設届の列が選択された状態で右クリックし、[列の追加]タブ→[例からの列]→[選択範囲から]をクリックする



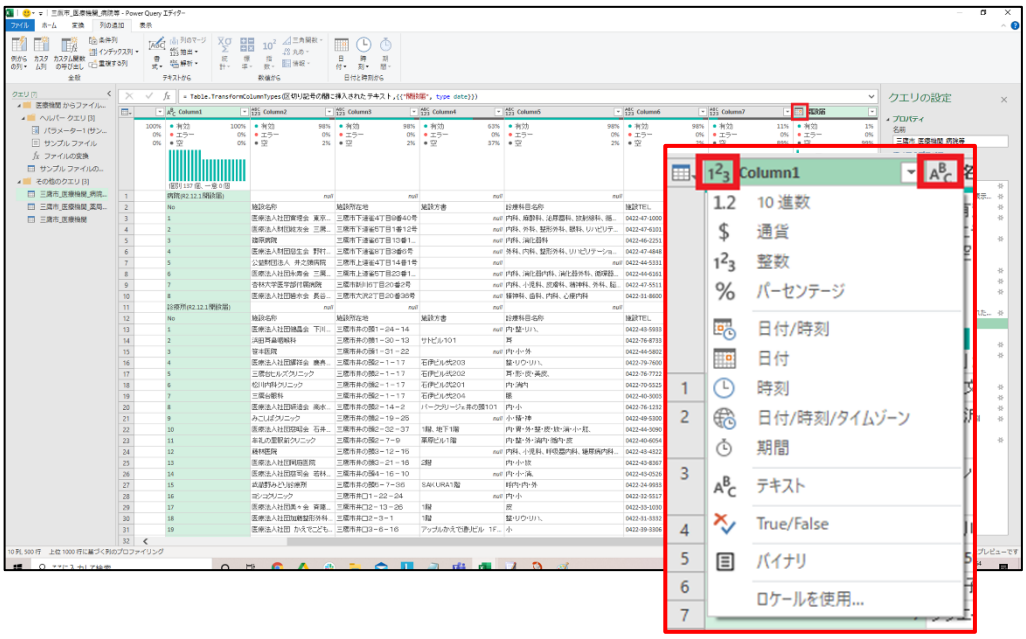
No、開設届年月日が格納されている Column1 列から必要範囲を切り出し、開設届列を追加します。

ⓒ東京都

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

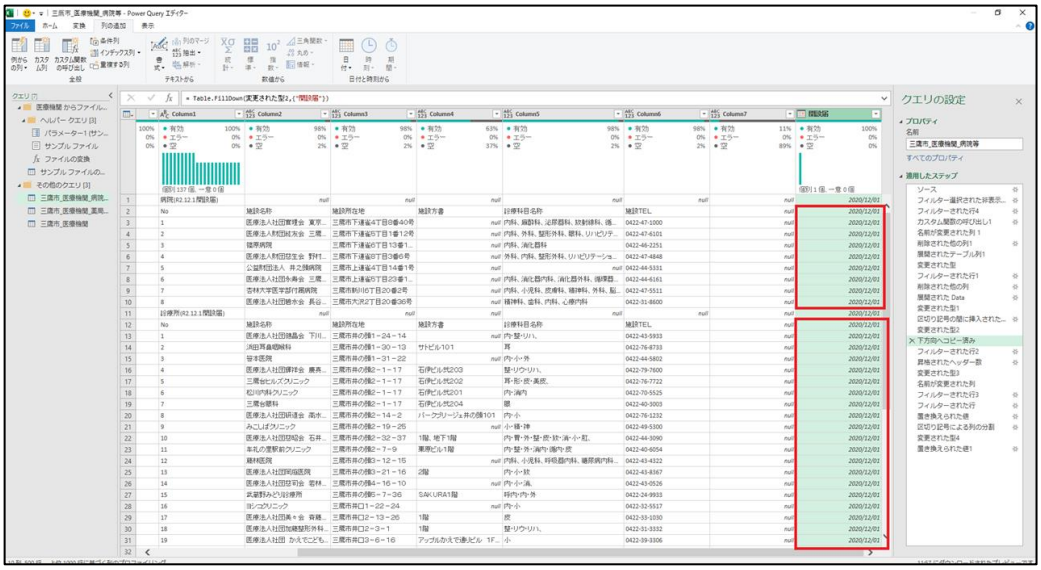


13. 開設届列のヘッダー上部にある「ABC123」アイコンをクリックし、日付の型を選択する

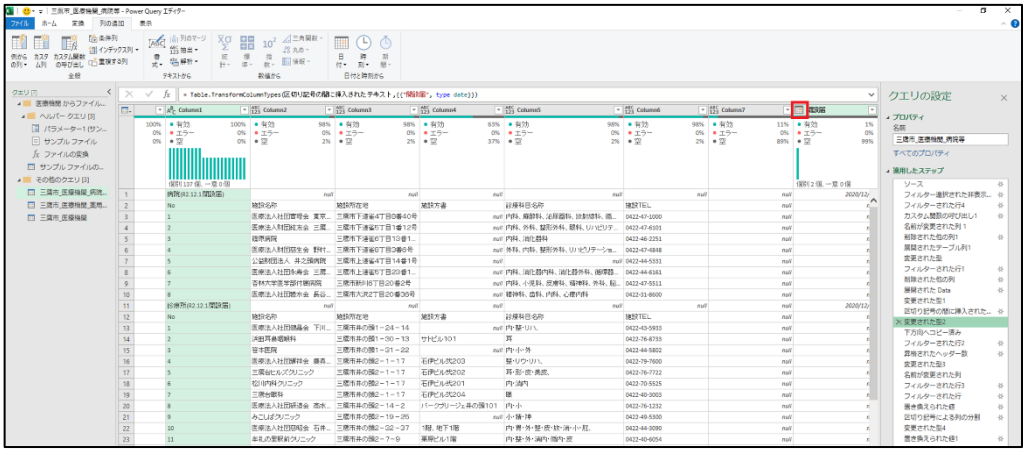


ⓒ東京都

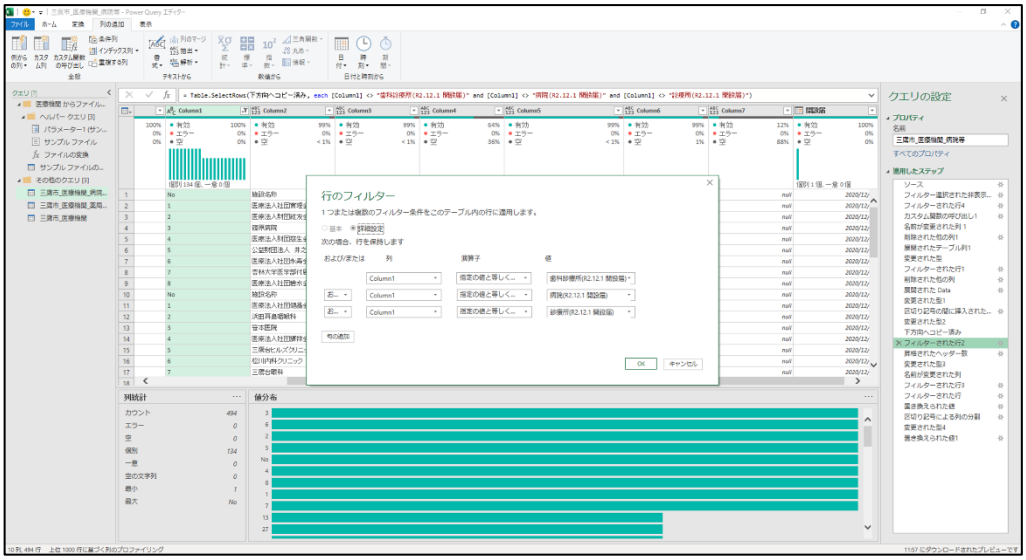
14. 開設届の列を右クリックし、[フィル]→[下]または[上]をクリックする



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



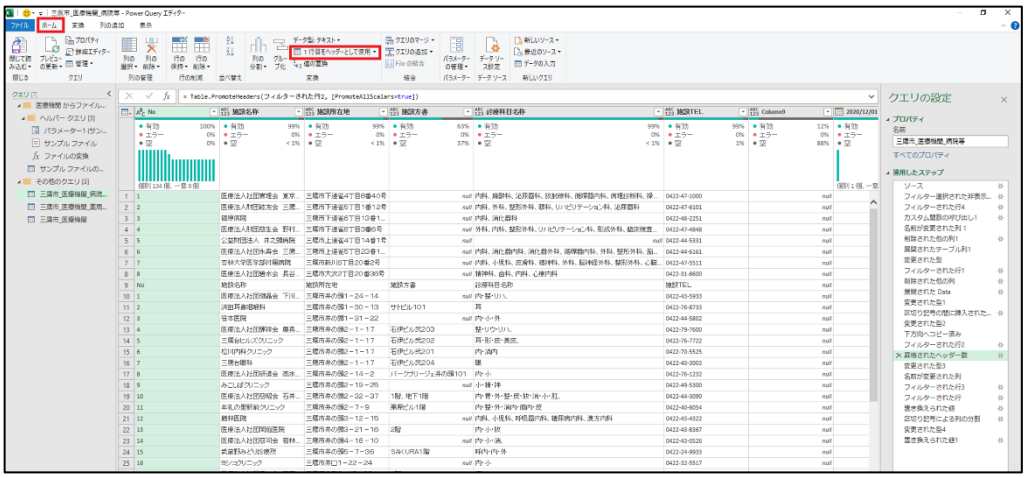
16. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しくない」を選び、[OK]をクリックする



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。

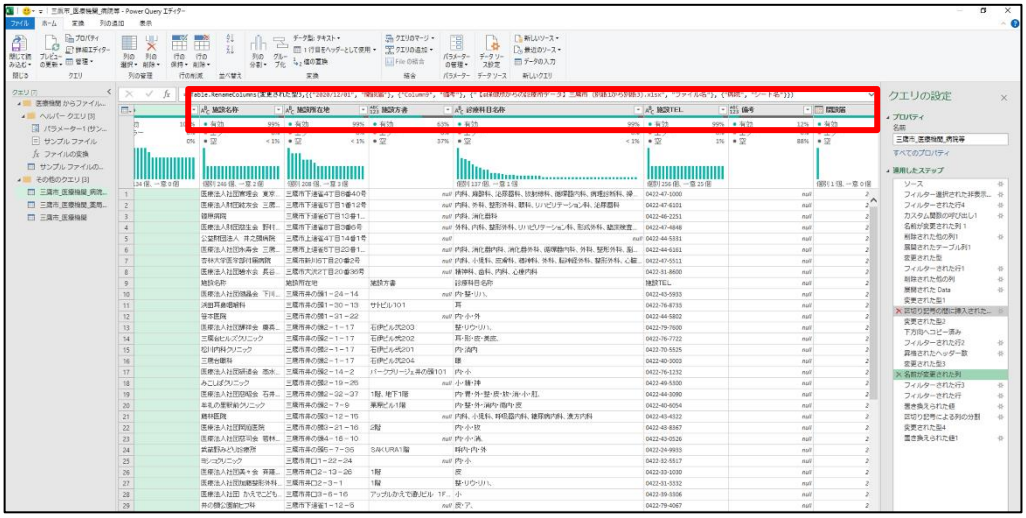
ⓒ東京都

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

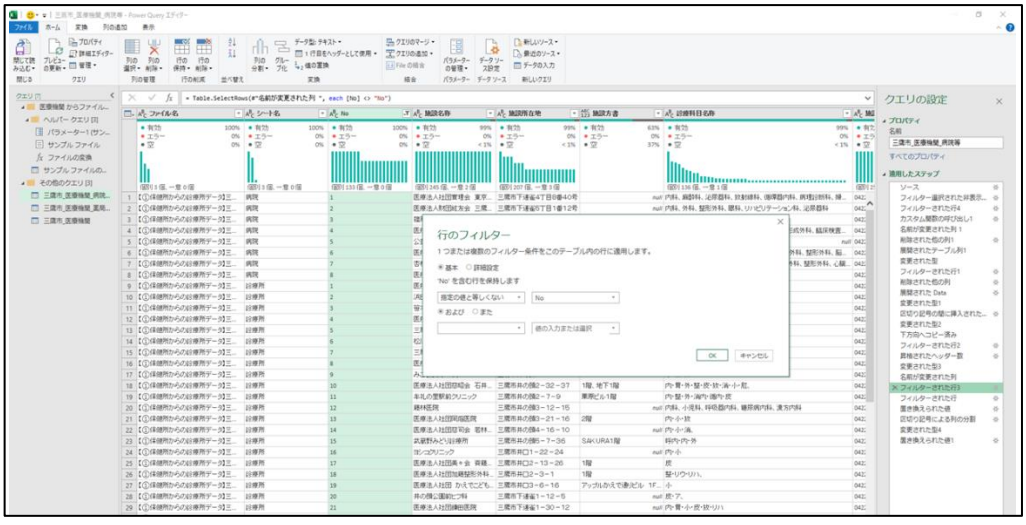


1 行目をヘッダーとする整形を行っています。

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



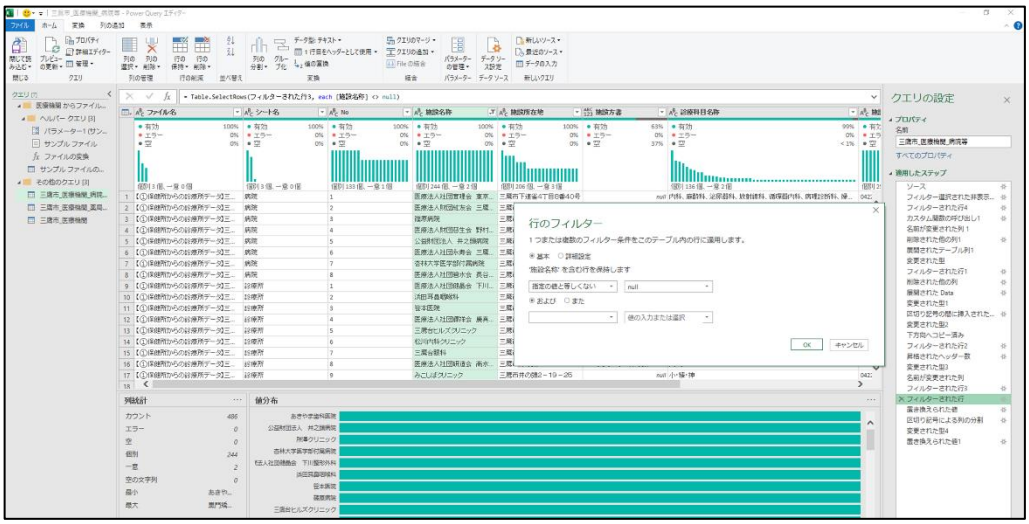
19. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しくない」を選び、[OK]をクリックする



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。

ⓒ東京都

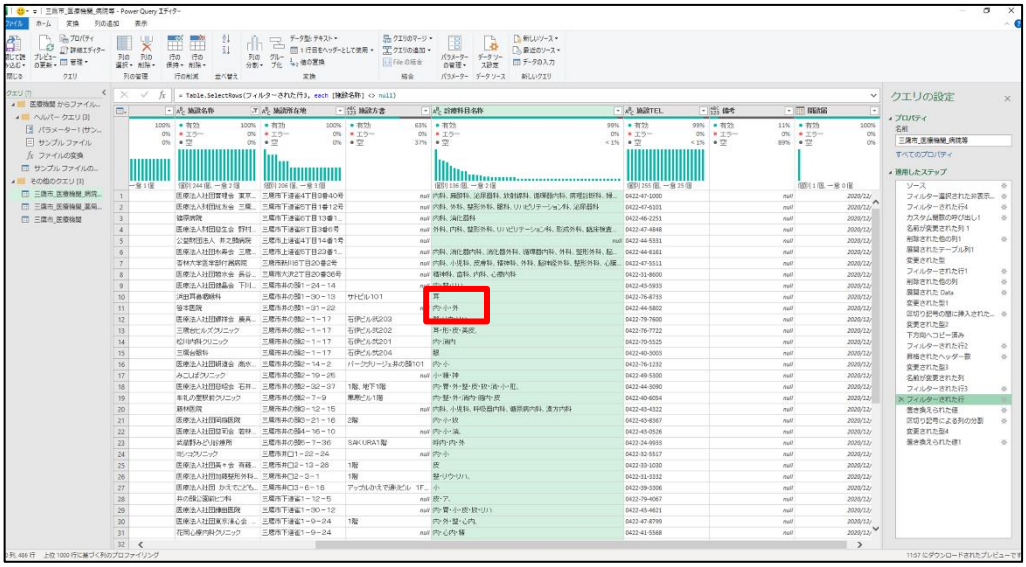
20. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しくない」を選び、[OK]をクリックする



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。

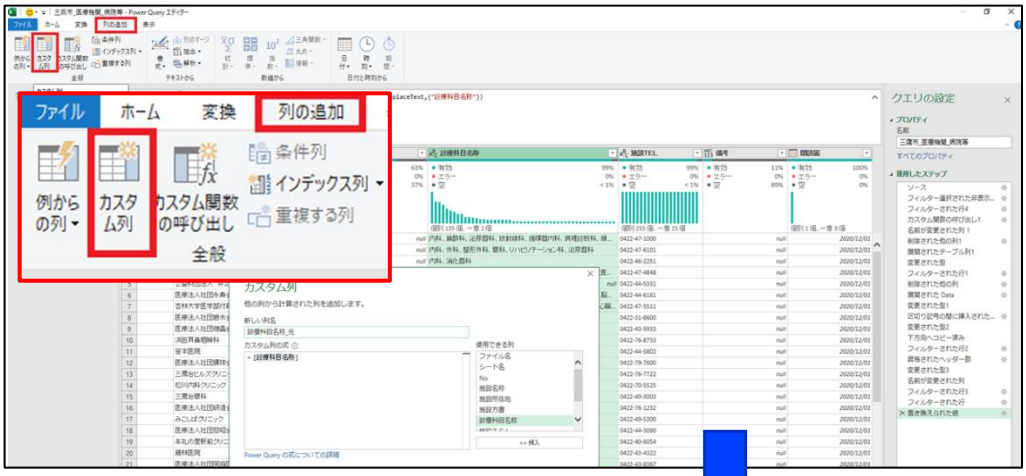
21. 診療科目名称の列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

置換する値「"、"以外の区切文字」 → 置換後「"、"」

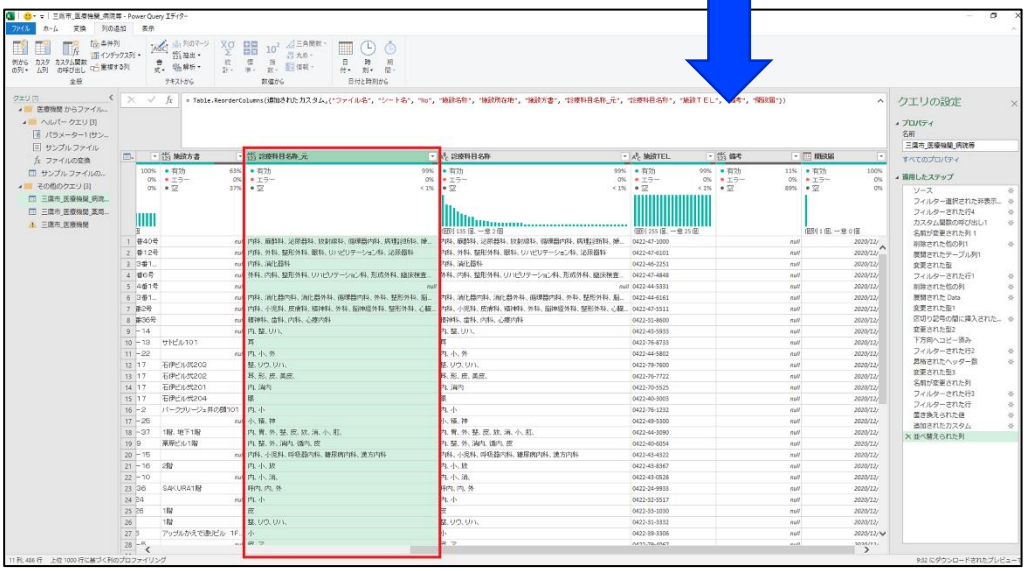


ⓒ東京都

22. [列の追加]タブ→[カスタム列]をクリックする



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。

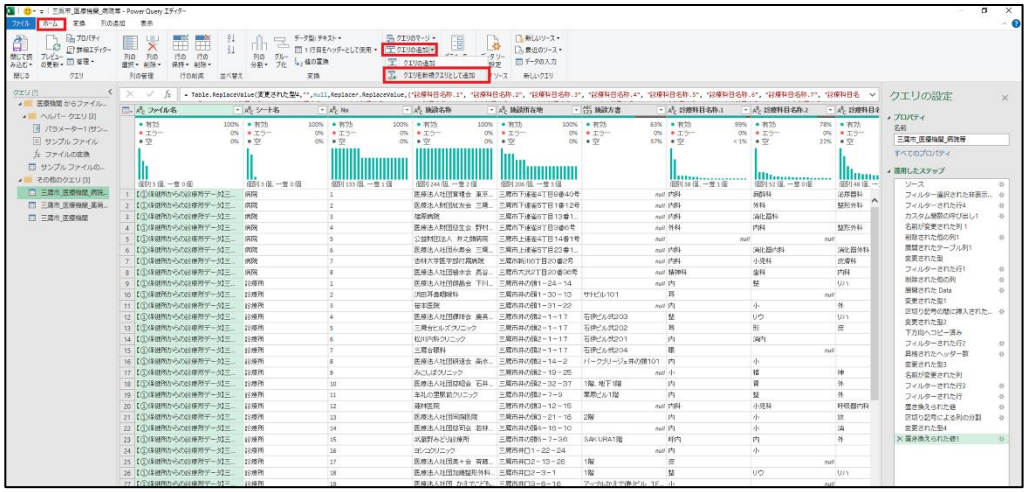


カスタム列を追加後の状態です。

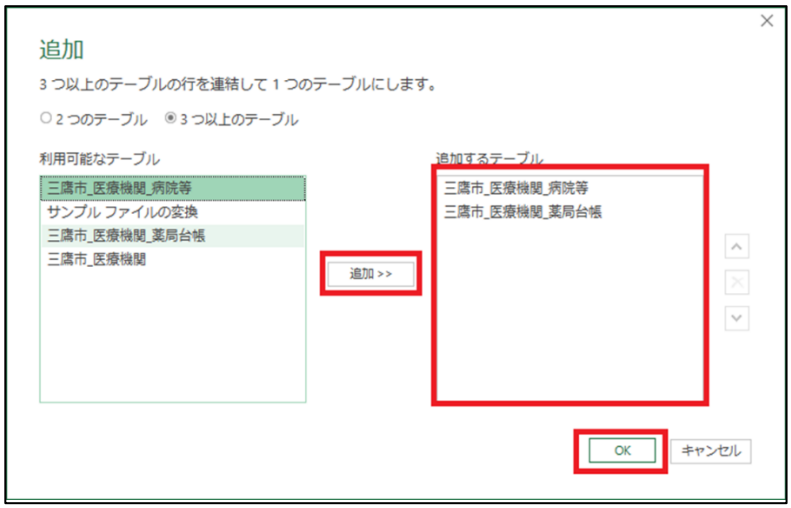
ⓒ東京都

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

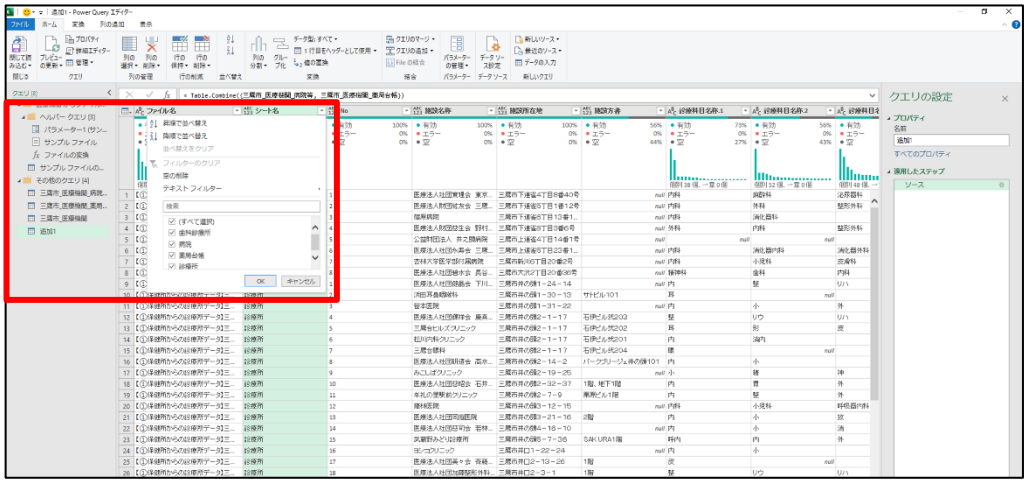
1. [ホーム]タブ→「クエリの追加」→「クエリを新規クエリとして追加」とクリックする



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

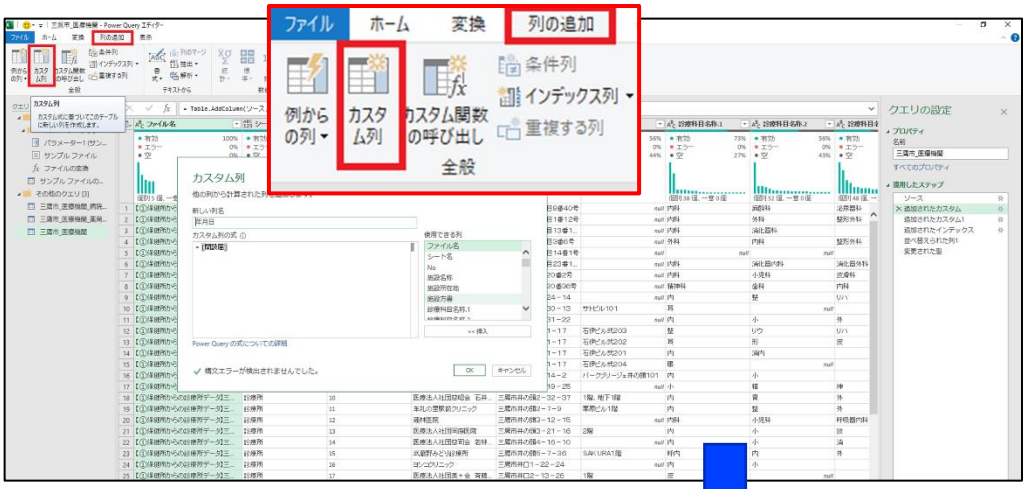


2. 選択した複数クエリのデータが行結合されていることを確認する

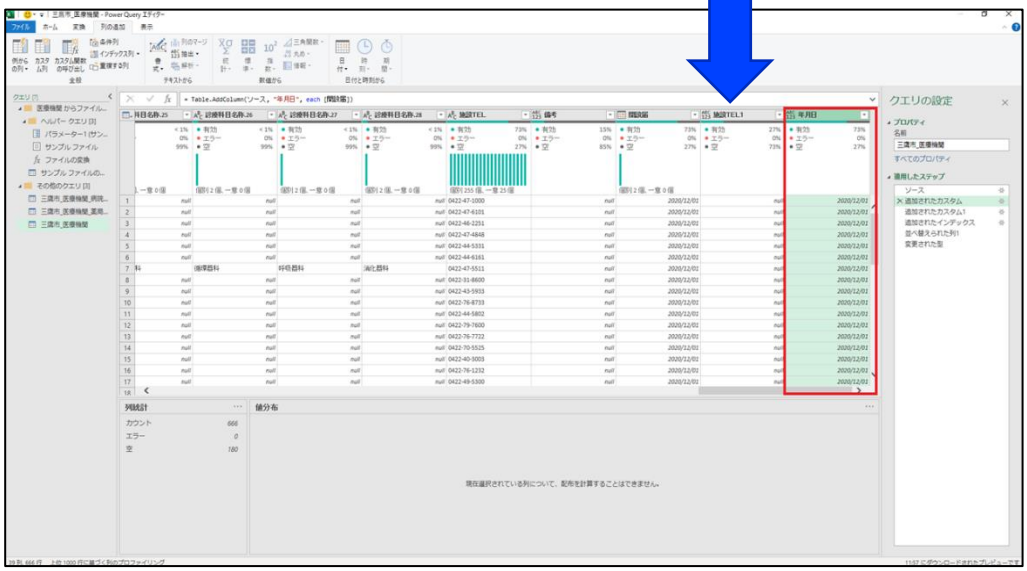


ⓒ東京都

3. [列の追加]タブ→[カスタム列]をクリックする



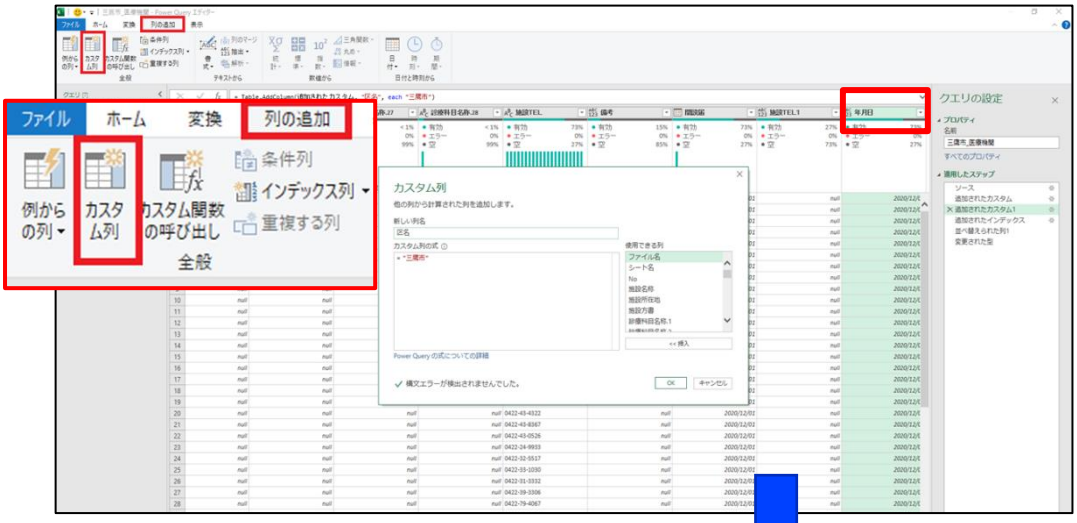
カスタム列の式に開設届列を作成します。

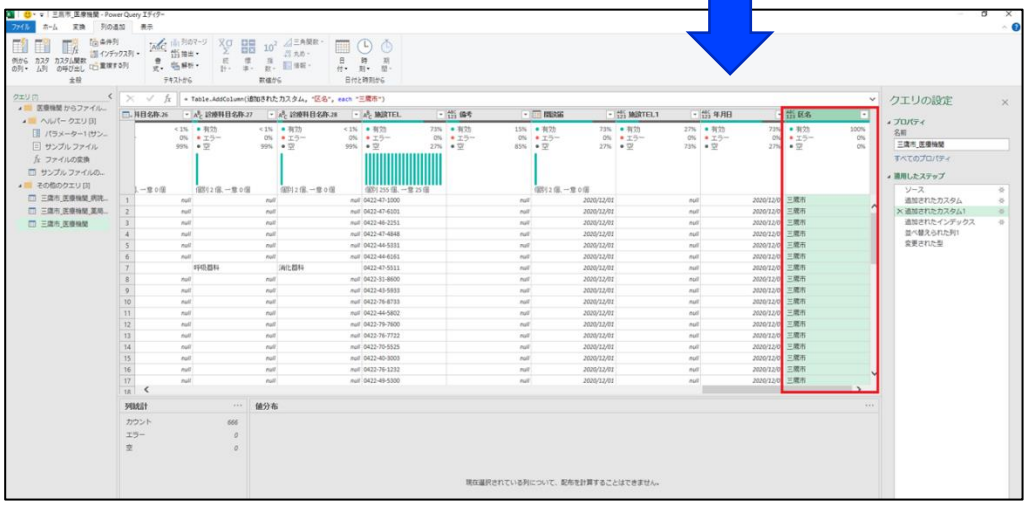


年月日カスタム列を追加後の状態です。

ⓒ東京都

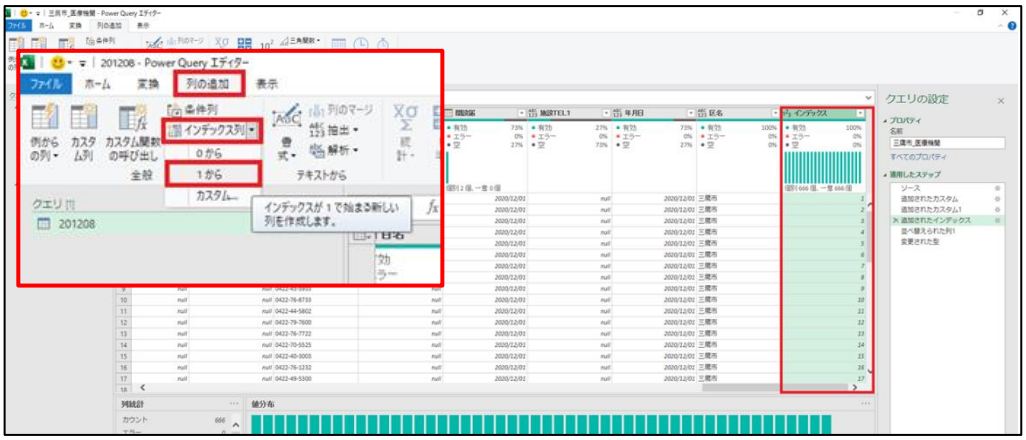
4. 追加した「カスタム列」に、対象の区名を入力する





区名カスタム列を追加後の状態です。

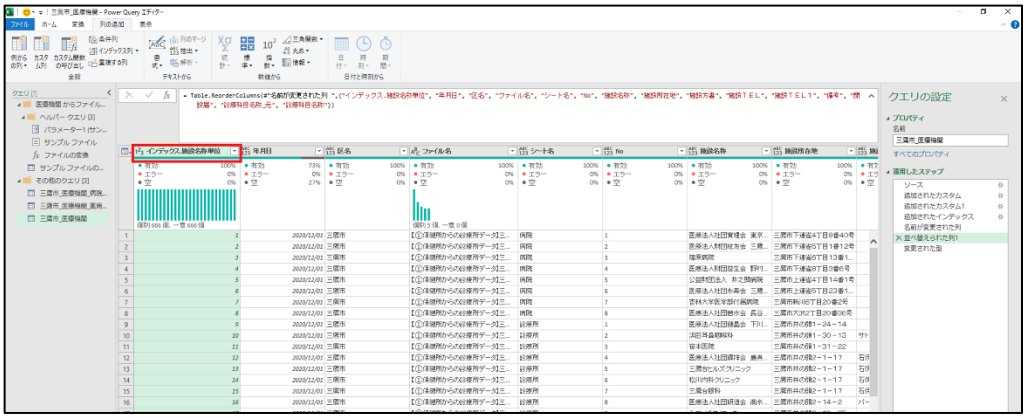
5. 追加する列が選択された状態で右クリックし、[列の追加]タブ→[インデックス列]→[1 から]をクリックする



インデックス列を追加しています。

ⓒ東京都

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

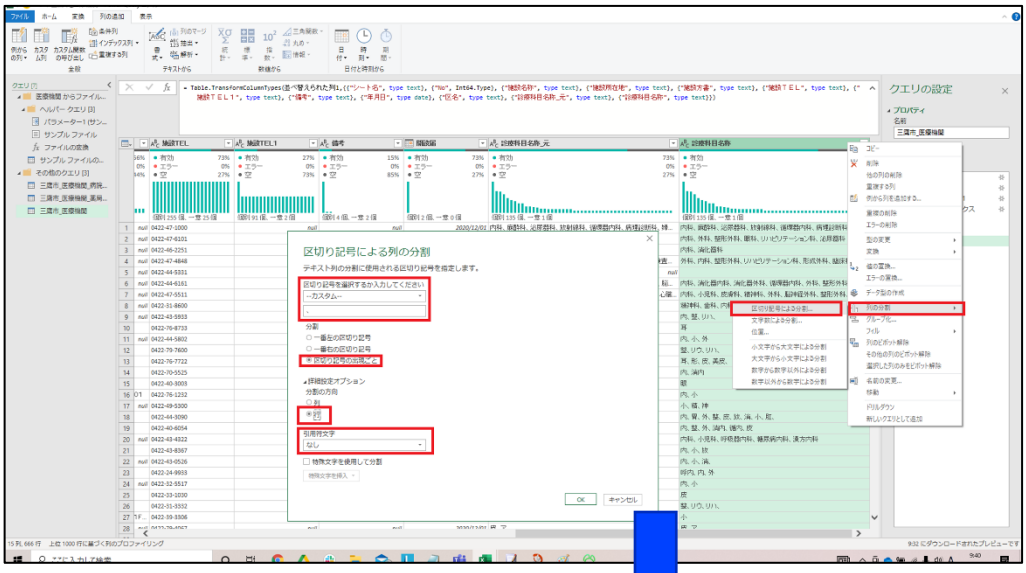


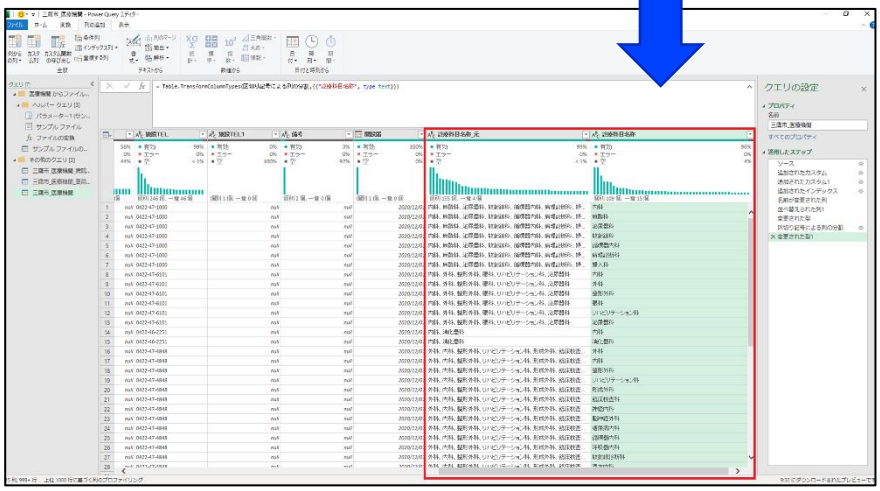
7. 対象列が選択された状態で右クリックし、「列の分割」→「区切記号による分割」をクリックする

「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」

同一列に複数列記されているデータで列を分割します。

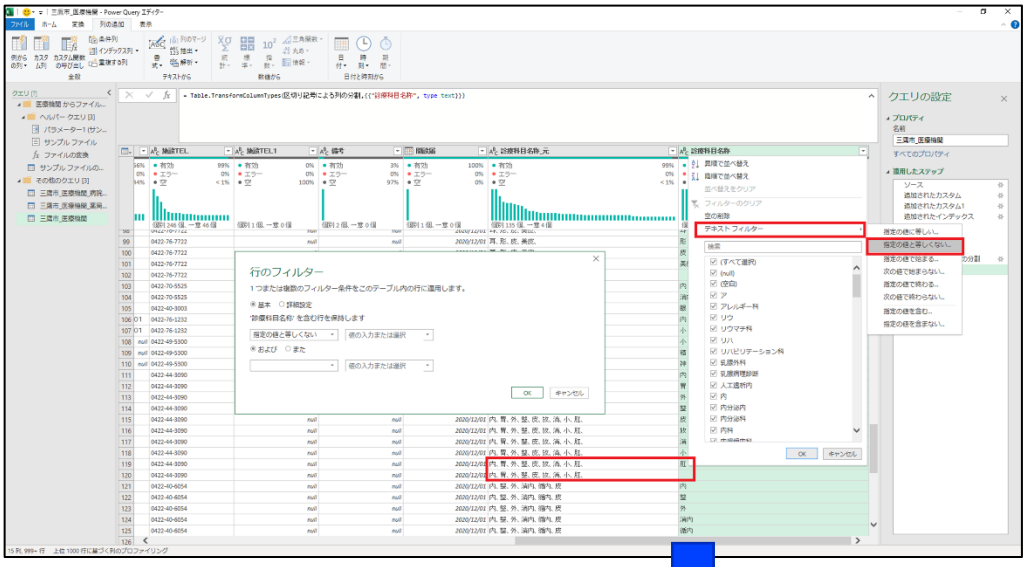




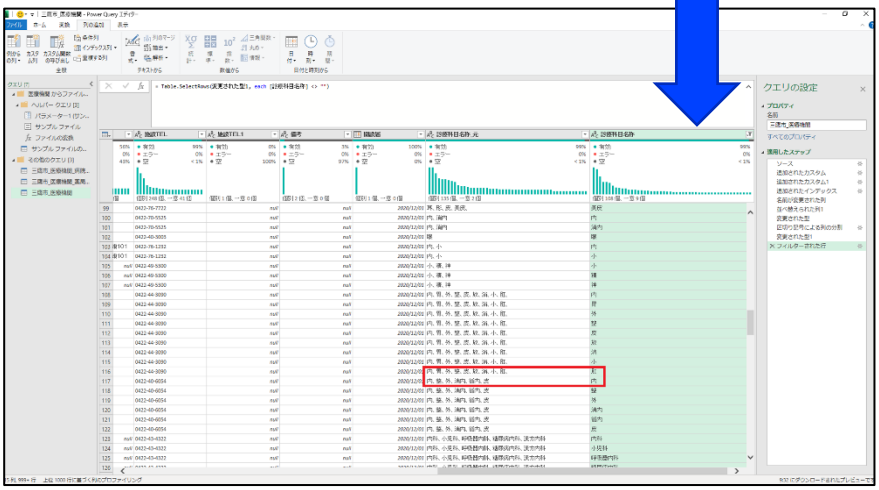
列が分割されます。

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

ⓒ東京都

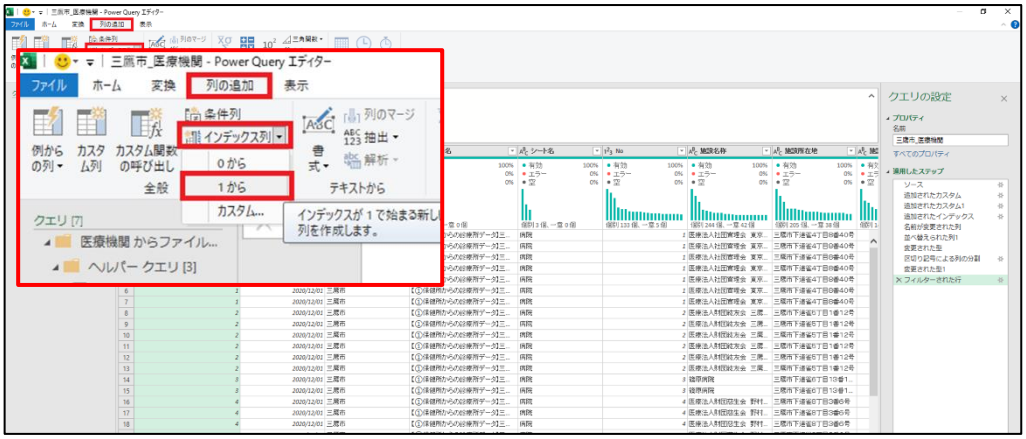


列の分割前データの末尾に付与されていた区切り文字による空白データ(「皮」「ア」等も含む)を削除します。



不要行（空白行）が削除されます。

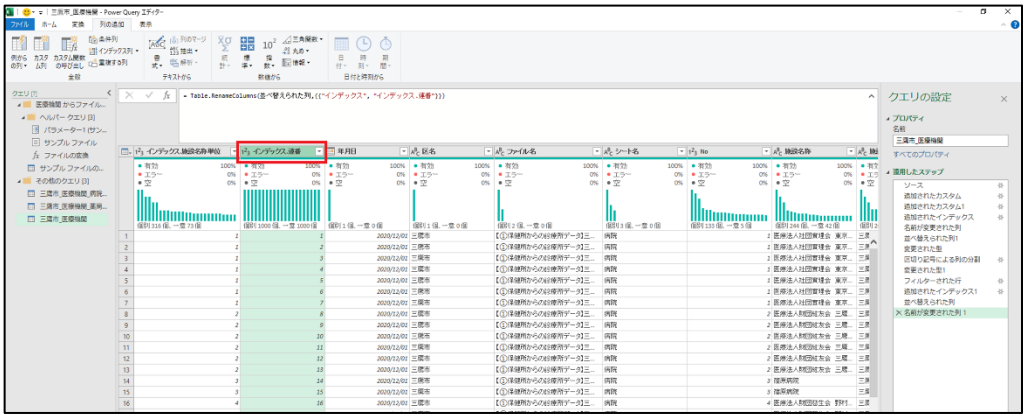
9. 追加する列が選択された状態で右クリックし、[列の追加]タブ→[インデックス列]→[1 から]をクリックする



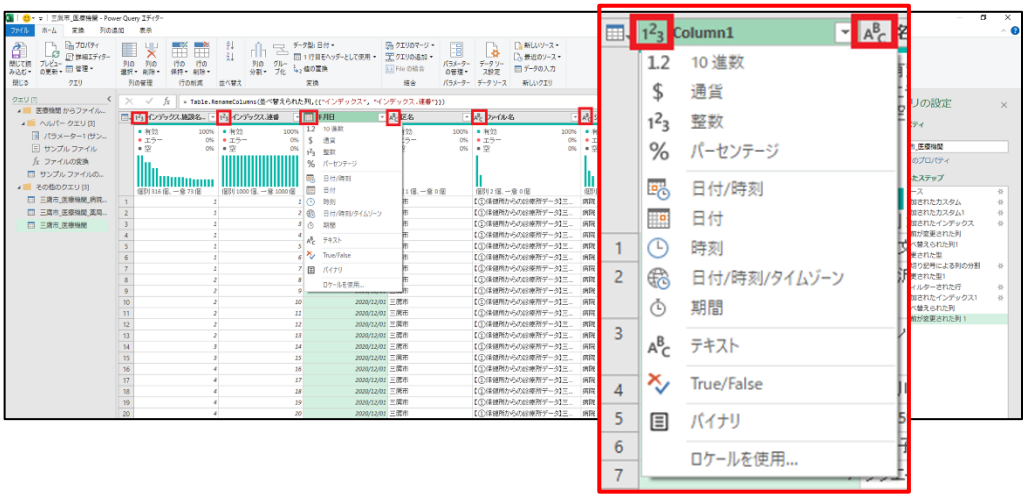
インデックス列を追加しています。

ⓒ東京都

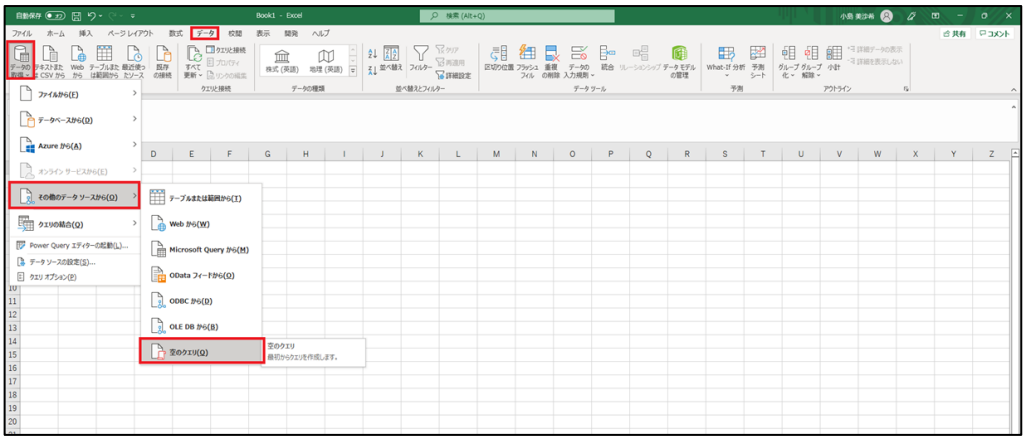
10. 追加した列をダブルクリックし、任意の列名を入力する



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

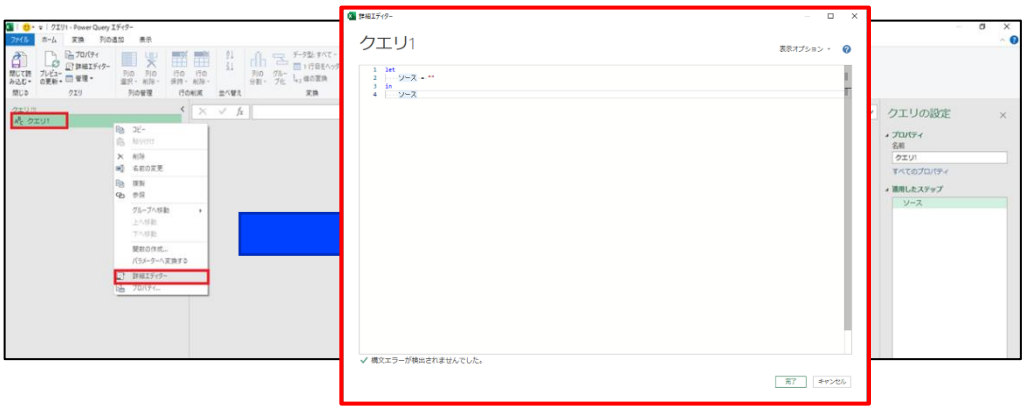


12. Excel で新規ブックを開き、[データ]タブ→[データの取得]→「その他のデータソースから」→「空のクエリ」をクリックする

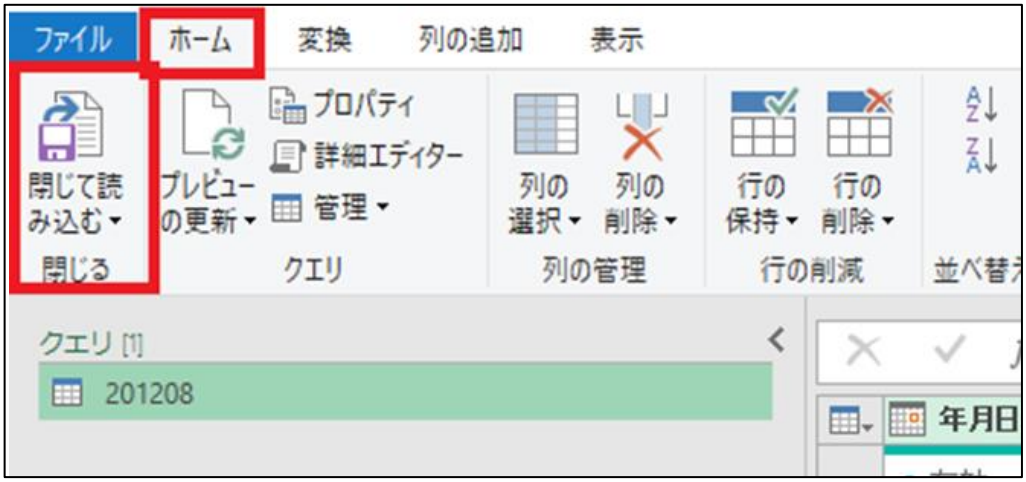


ⓒ東京都

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

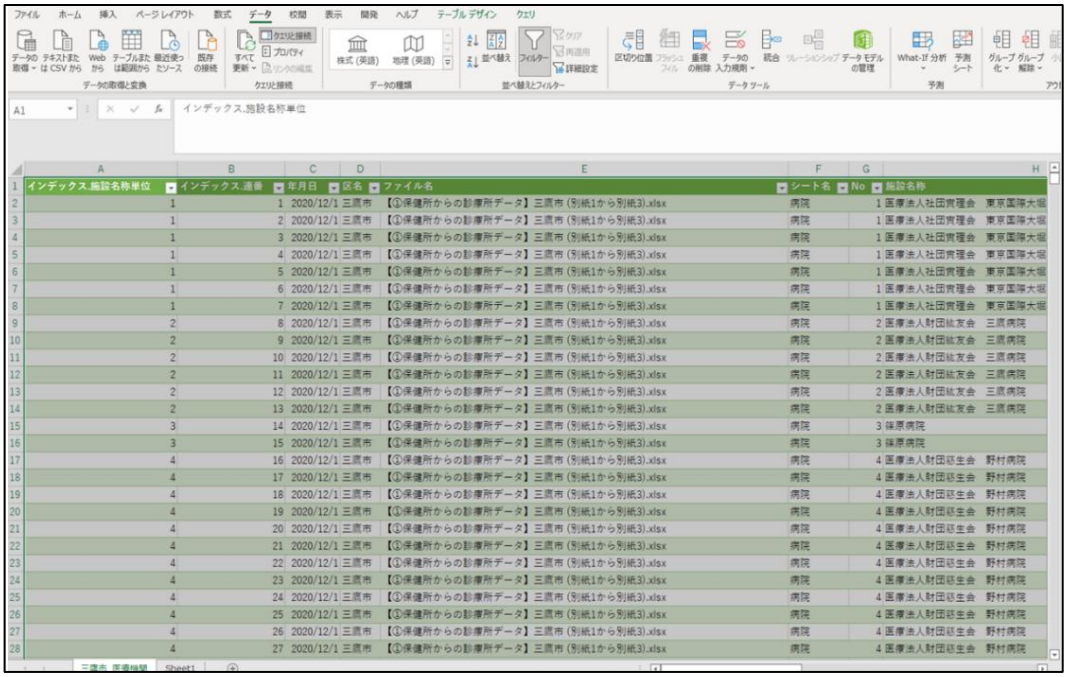


14. [ホーム]タブ→[閉じて読み込む]をクリックする



【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

ⓒ東京都

手順（ケース[D]）

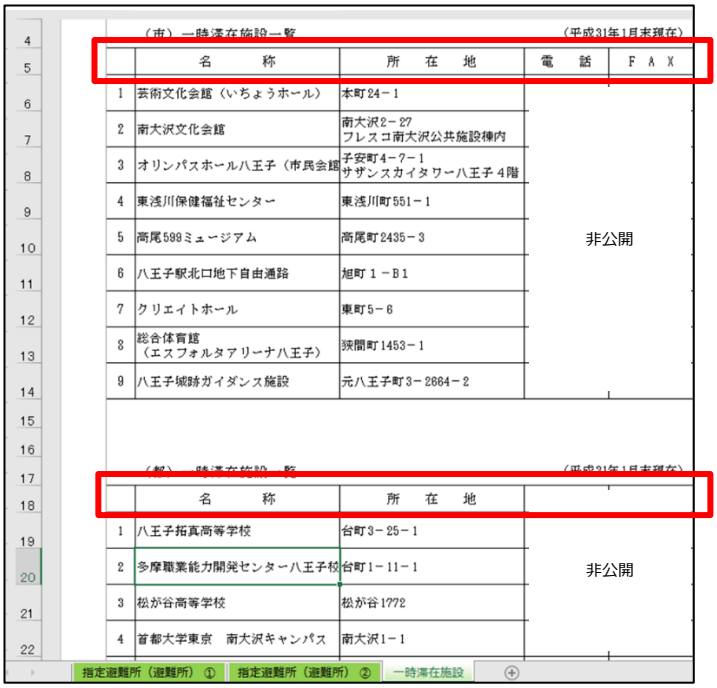
下記の手順で Excel から CSV にデータを変換します。

【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. 81）
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. 82）
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. 83）
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. 74）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	FAX
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	

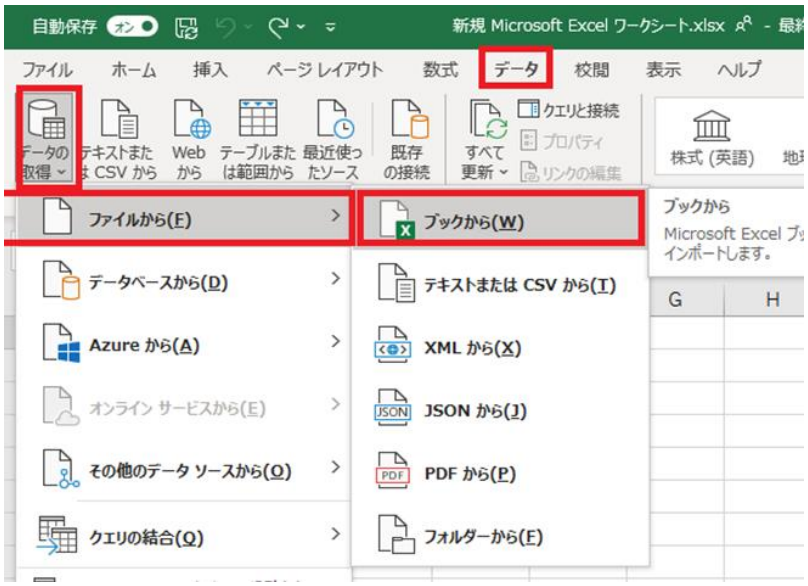
指定避難所（避難所）①　指定避難所（避難所）②　一時滞在施設

ⓒ東京都

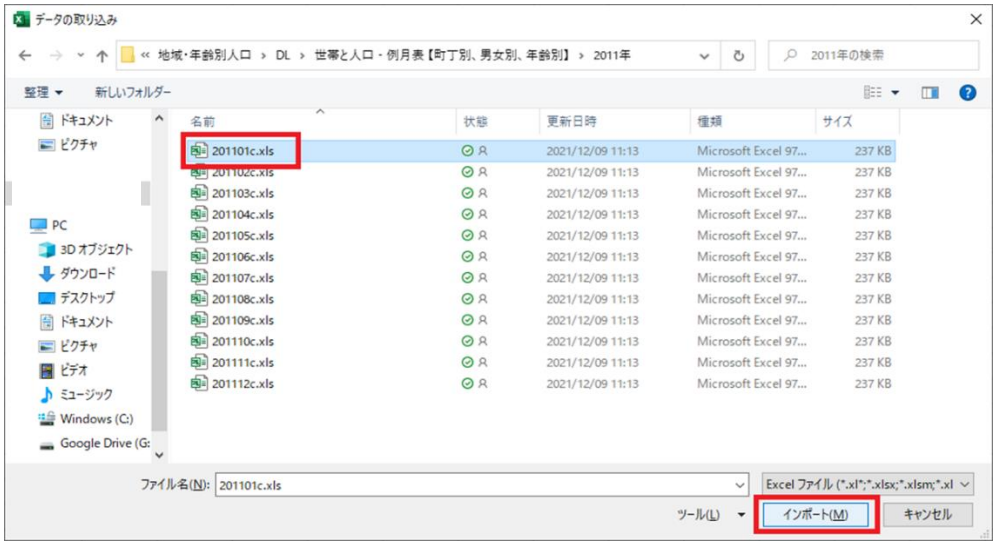
②　Excel で整形、加工したいファイルをインポートする

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

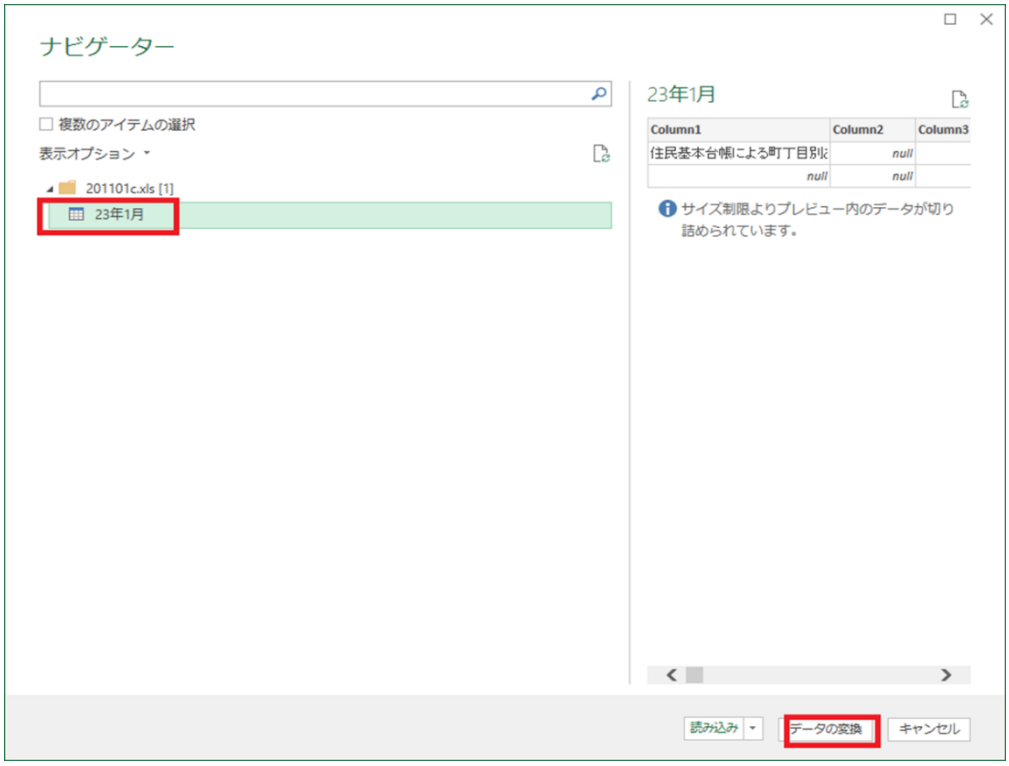


3. Excel ファイルをクリックして選び、[インポート]をクリックする



ⓒ東京都

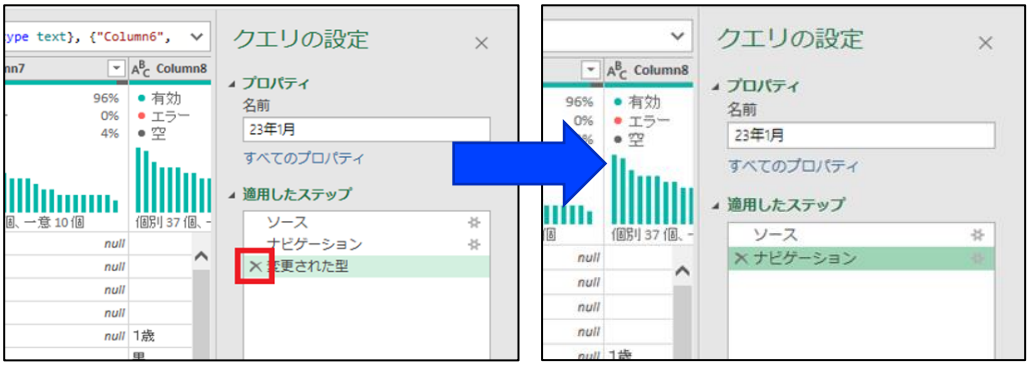
4.　変換するアイテムをクリックして選び、[データの変換]をクリックする



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。

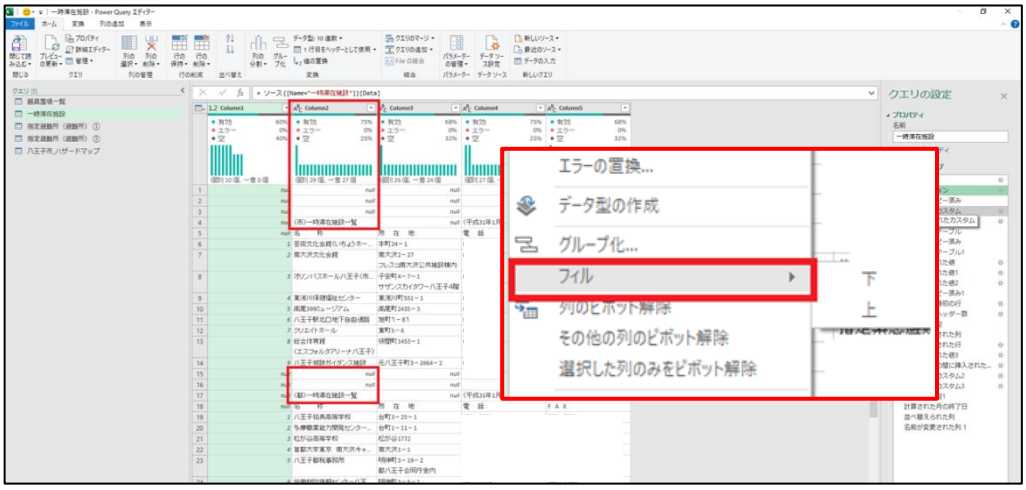
③　Excel の Power Query 機能でブックごとにデータを整形する

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。

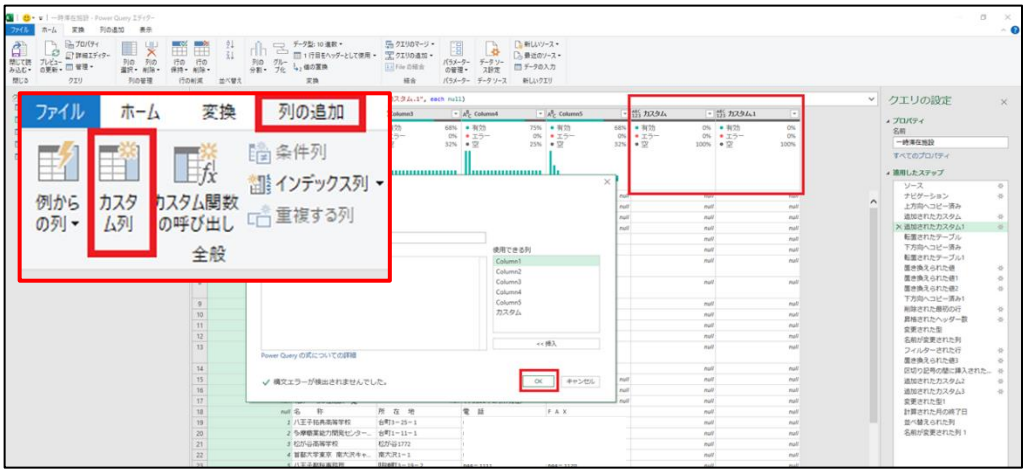
2.　施設区分と名称の列を右クリックし、[フィル]→[下]または[上]をクリックする



施設区分と名称が同列に混在するため分離します。

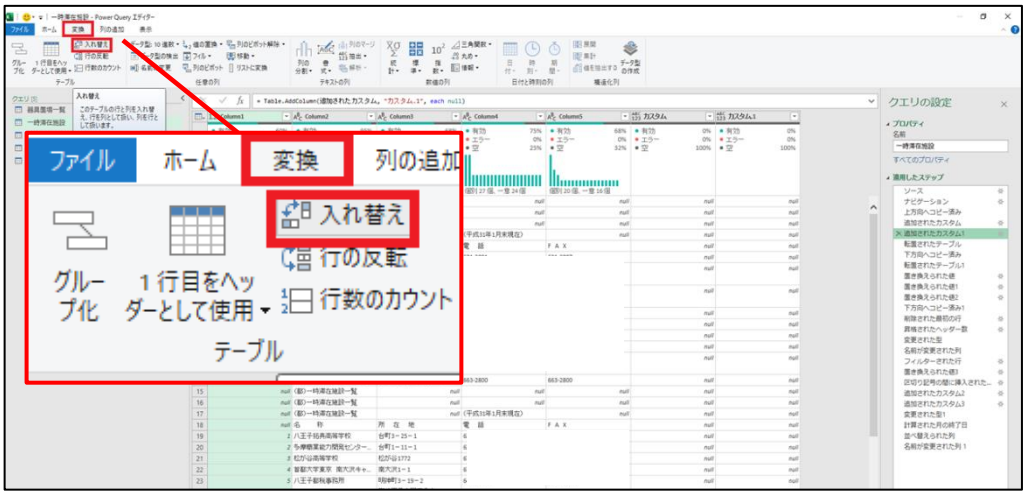
ⓒ東京都

3. [列の追加]タブ→[カスタム列]をクリックする



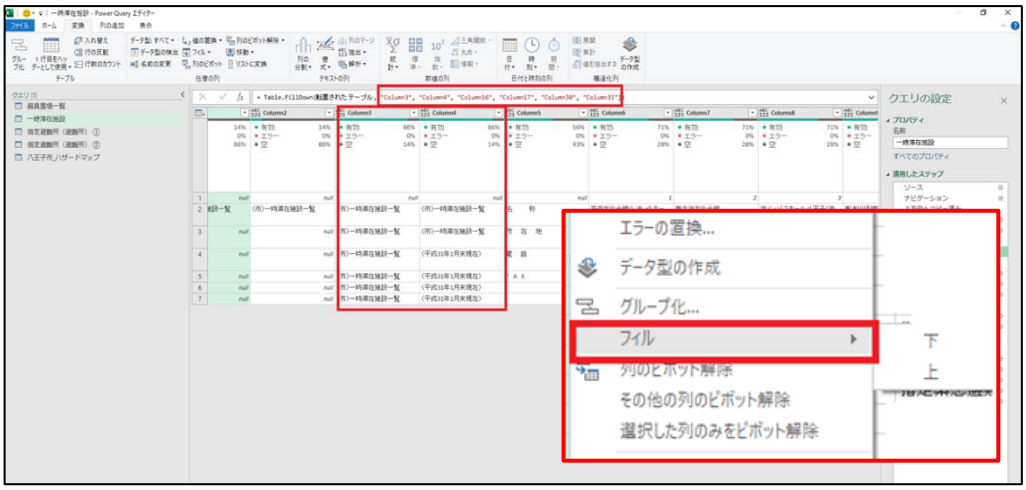
今回は最終列をコピーし、年月日のカスタム列を作成します。

4. [変換]タブ→[入れ替え]をクリックする



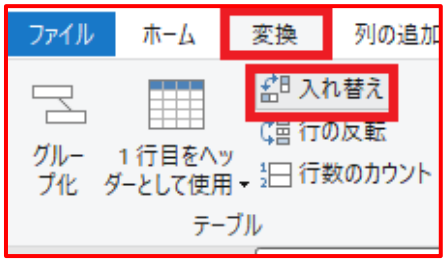
追加したカスタム列に値をコピーするため、行列を入れ替える処理を行います。

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



ⓒ東京都

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。

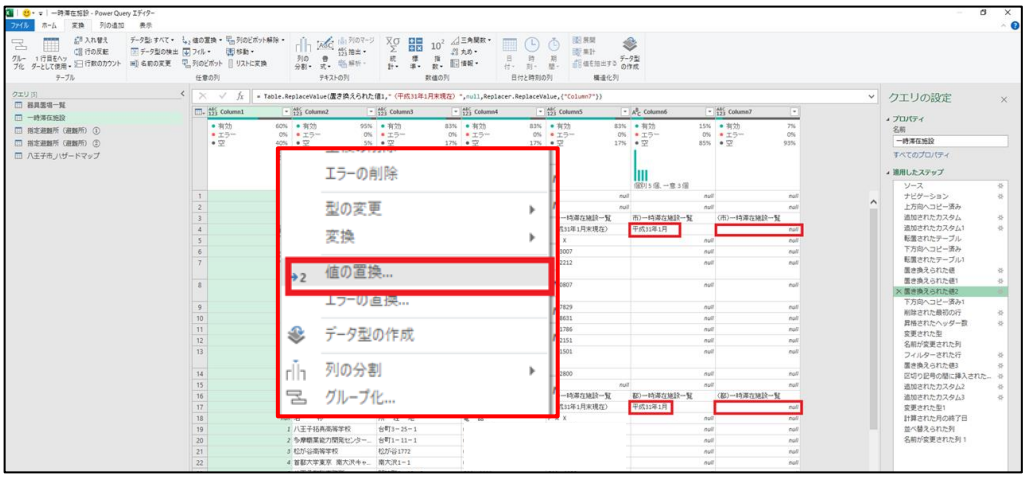
置換する値：年月日の文言　→　置換後「null」

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。



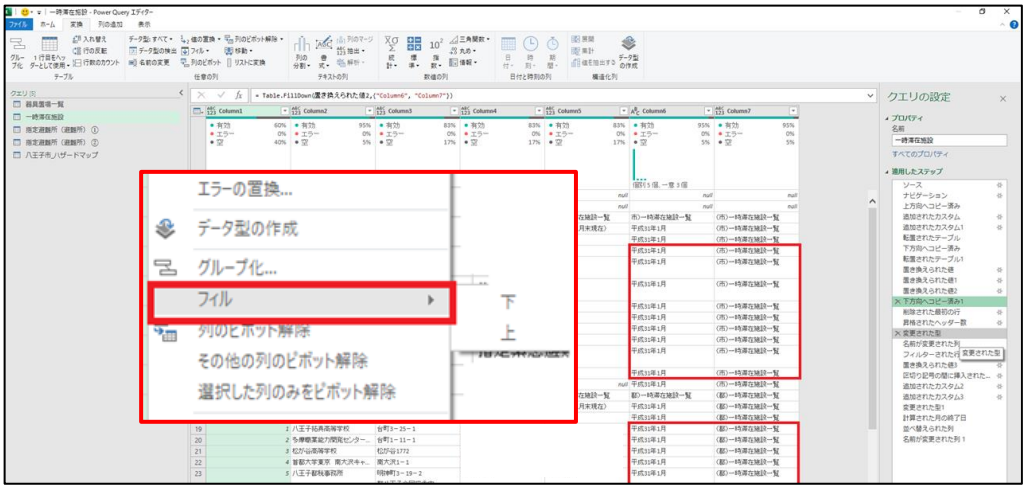
日付型の列での注意点です。

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。

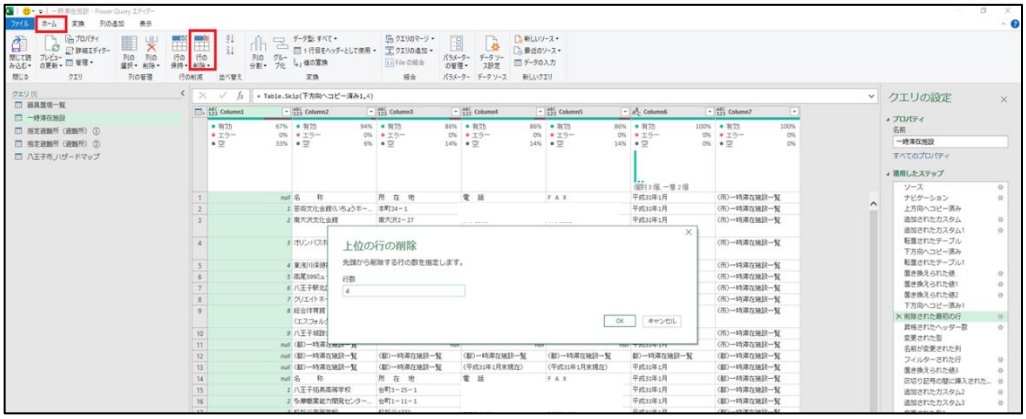
× 令和元年１月１日	○ 令和１年１月１日
× 令和１年1月1日	○ 令和１年１月１日
× 令和一年一月一日	○ 令和１年１月１日
× ２０１９０１０１	○ 20190101
× ２０１９/1/1	○ 2019/1/1

ⓒ東京都

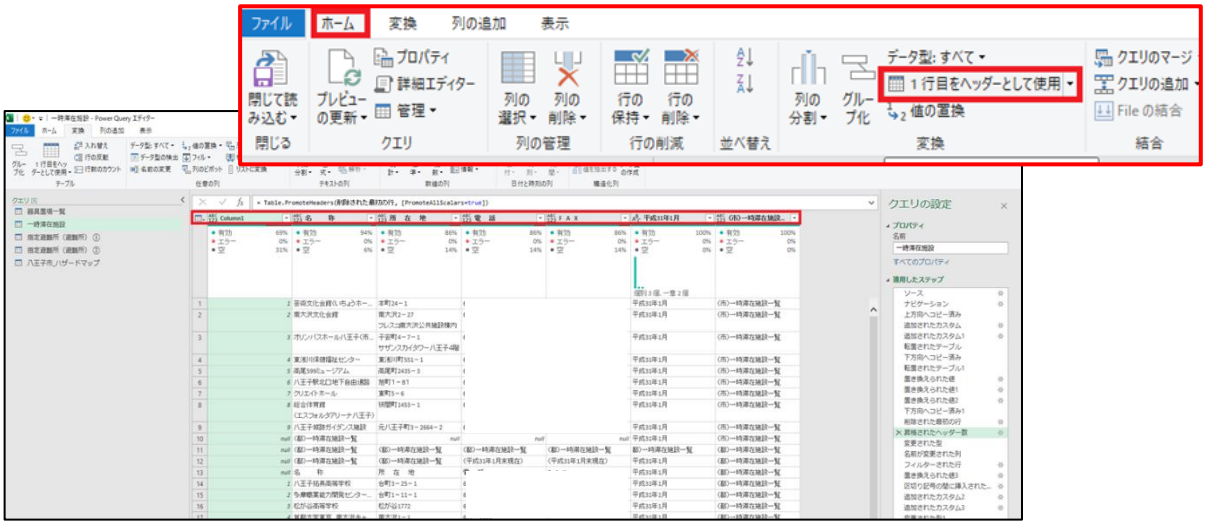
8. 施設区分と名称の列を右クリックし、[フィル]→[下]または[上]をクリックする



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

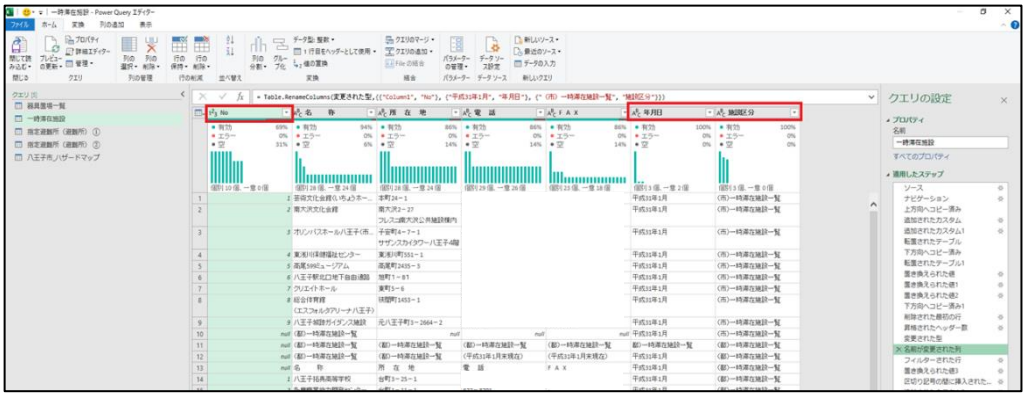


10. [ホーム]タブ→[1 行目をヘッダーとして使用]をクリックする



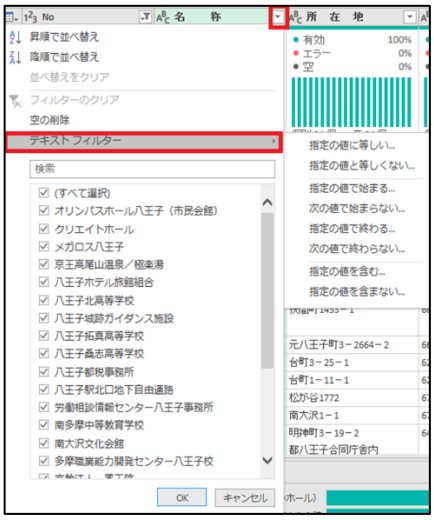
ⓒ東京都

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

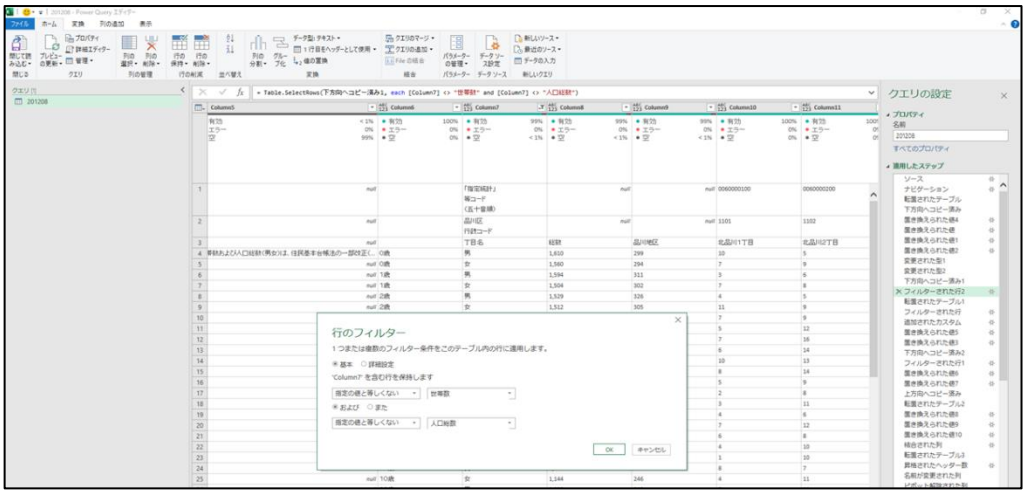


12. 対象列の右側「▼」をクリックして、数値フィルターまたはテキストフィルターを表示した後、「規定の値と等しくない」を選び、[OK]をクリックする

必要行を残すフィルターを設定しています。

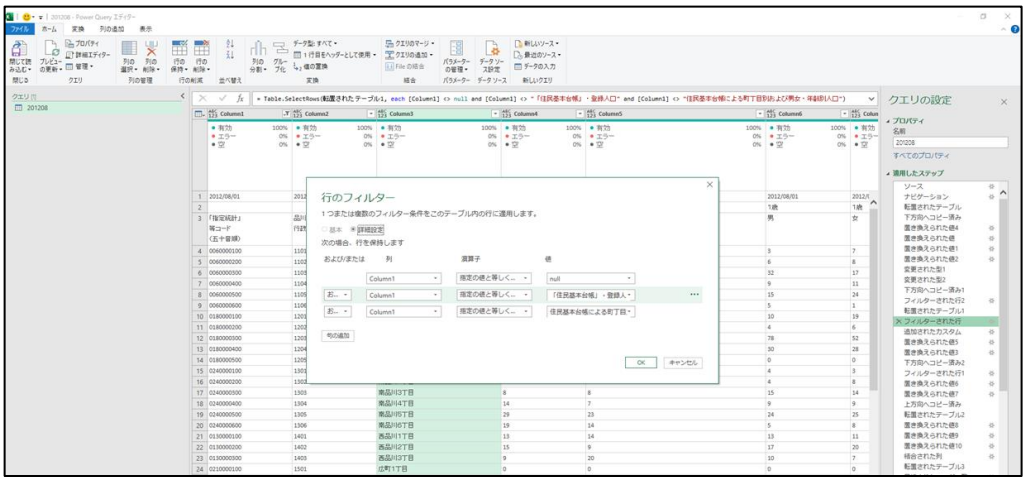


13. 年月日の列を右クリックし、除外するデータを非表示にする



ⓒ東京都

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

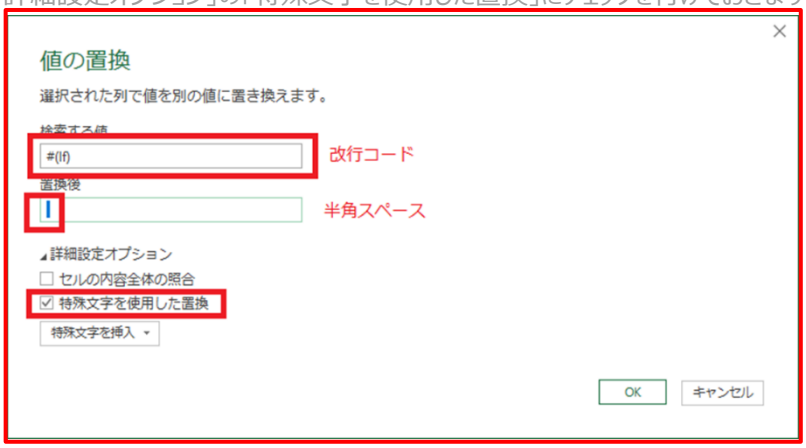


15. 対象列が選択された状態で右クリックし、[値の置換]をクリックし、値を置換する

置換する値「#(lf)」（改行コード）　→　置換後「 」（半角スペース）

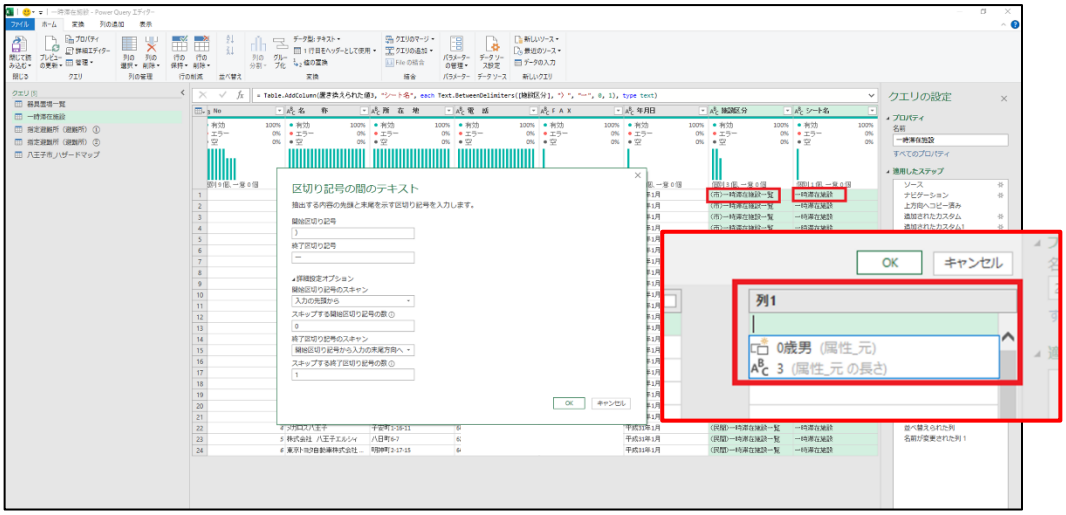
「詳細設定オプション」の「特殊文字を使用した置換」にチェックを付けておきます。

改行コードを削除しています。



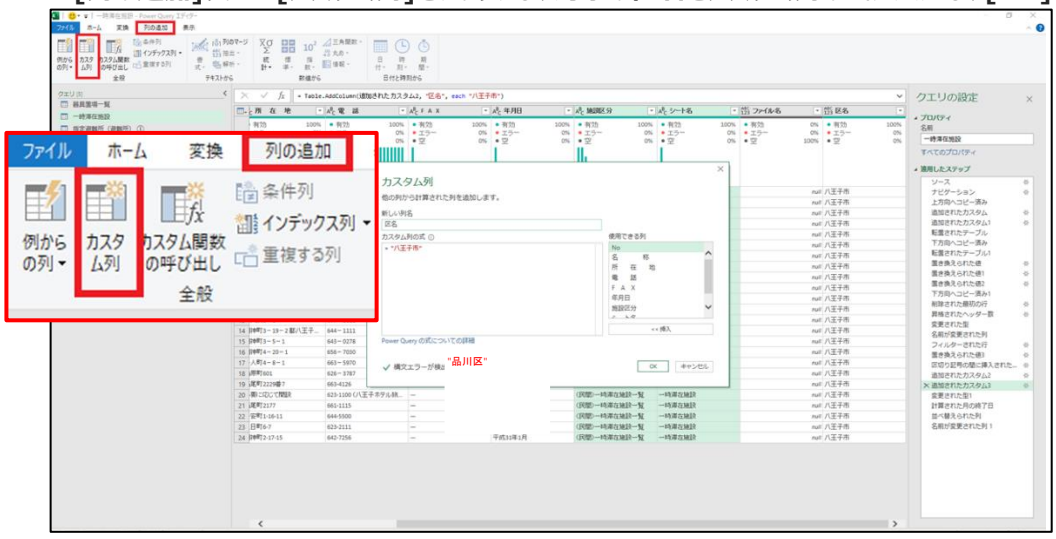
16. 開設届の列が選択された状態で右クリックし、[列の追加]タブ→[例からの列]→[選択範囲から]をクリックする

施設区分列から必要範囲を切り出します。

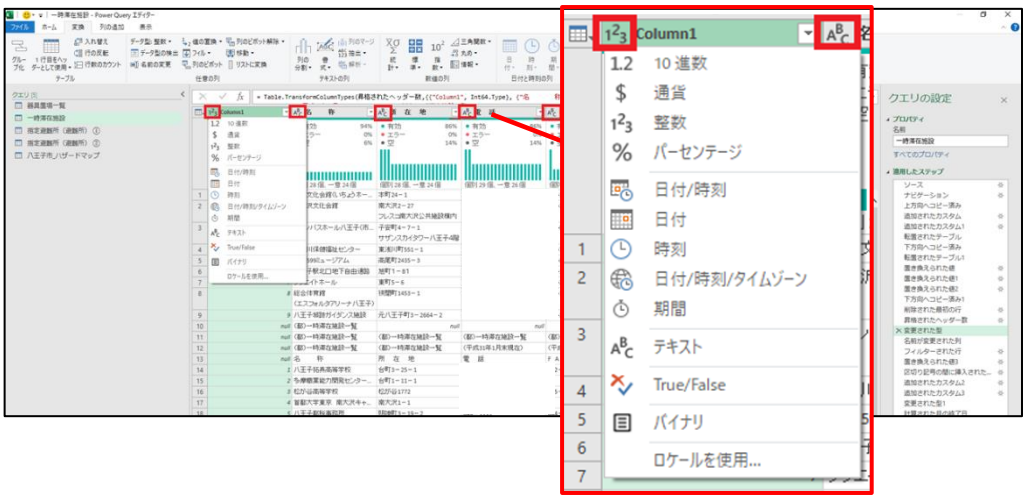


ⓒ東京都

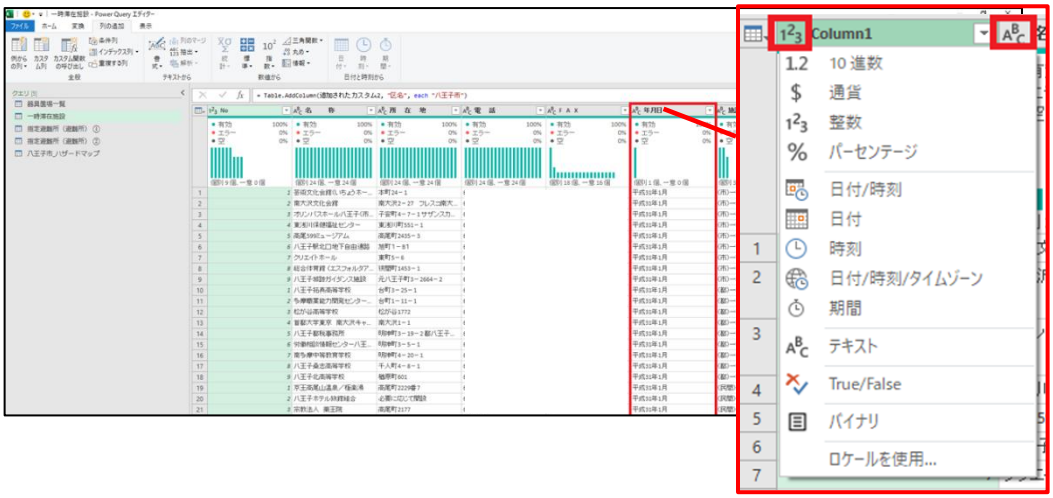
17. [列の追加]タブ→[カスタム列]をクリックし、対象の区名をカスタム列の式に入力し、[OK]をクリックする



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

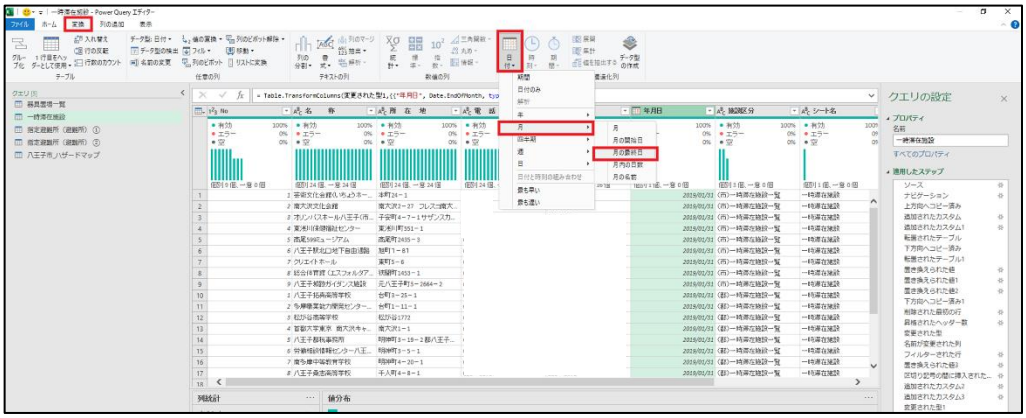


19. 年月日のヘッダー上部にある「ABC123」アイコンをクリックし、日付型を選択する



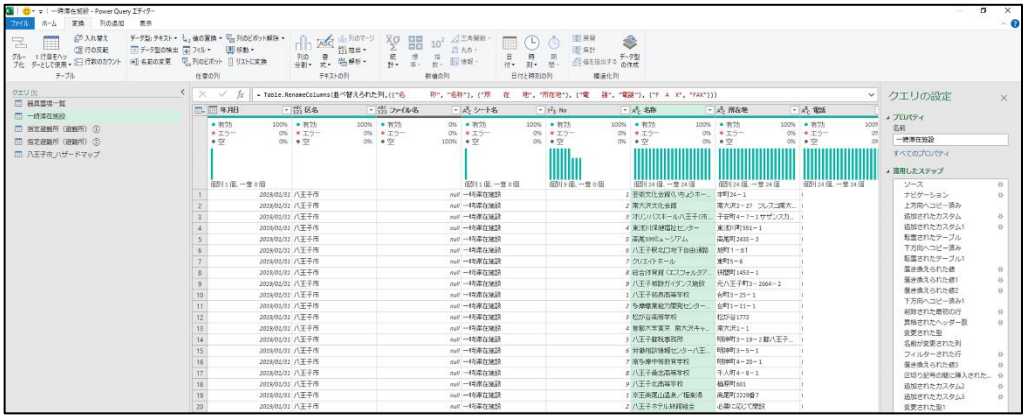
ⓒ東京都

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



ⓒ東京都

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. 91）**
② **アクションを実行し、データを作成します。** **（詳細： p.95）**
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.97）**
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.100）**
⑤ **処理結果を目視にて確認しクレンジング・正規化方針を決定します。** **（詳細： p.102）**

ご参考

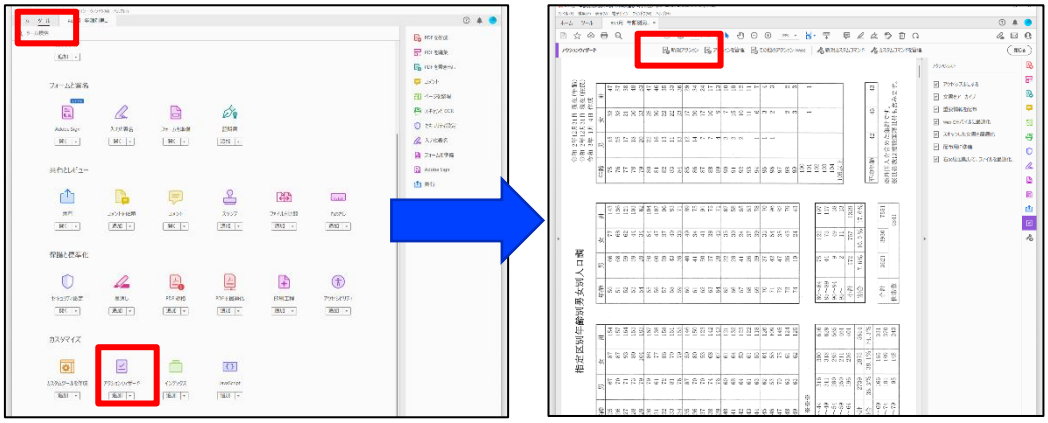
本書記載以外の、CLOVA の詳しい操作については、下記をご参照ください。
https://clova.line.me/clova-ocr/

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

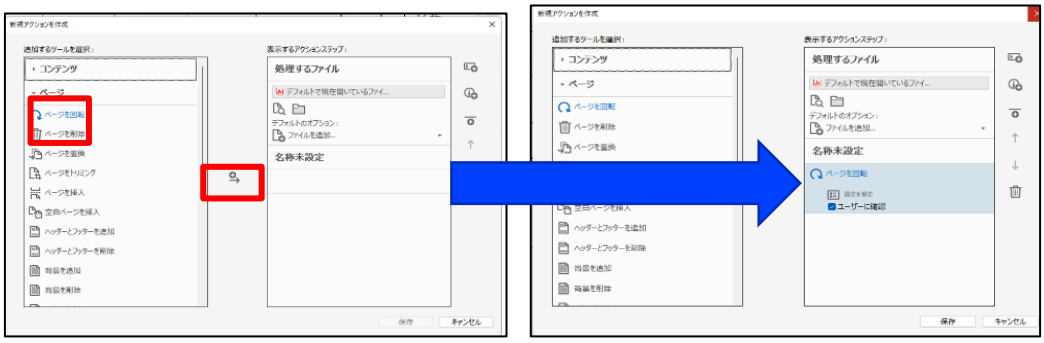
1. Adobe Acrobat DC を開く

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。

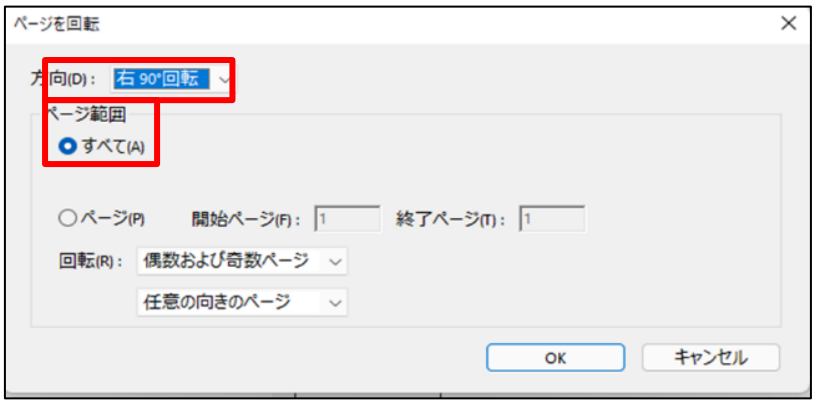
3. [ページ]→[ページを回転]をクリック、[→]をクリックして設定を右側パネルに移動させる



右側パネルに移動した操作が、上から順に実行されます。

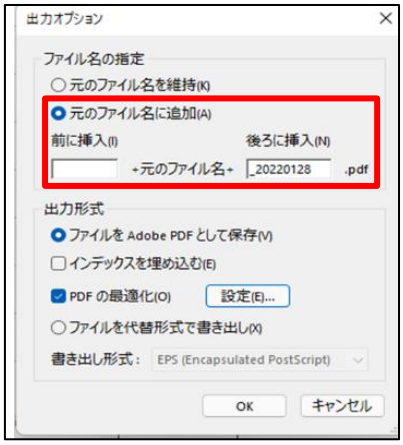
ⓒ東京都

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

ⓒ東京都

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

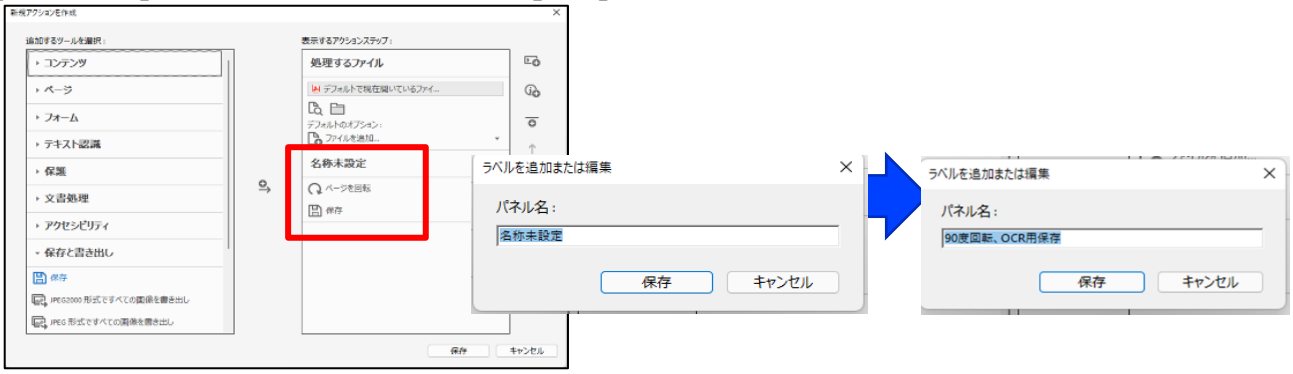
ユーザーデータを破棄：すべてにチェックを入れる

最適化：すべてにチェックを入れる

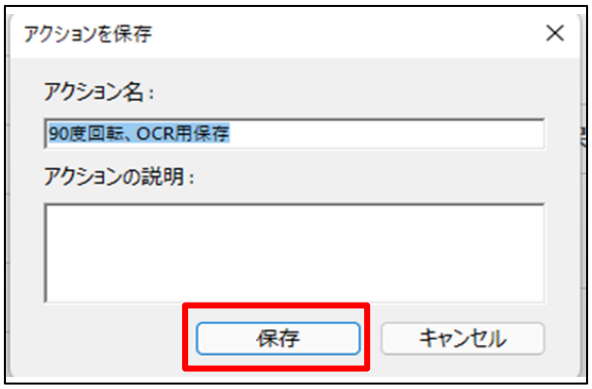
ⓒ東京都

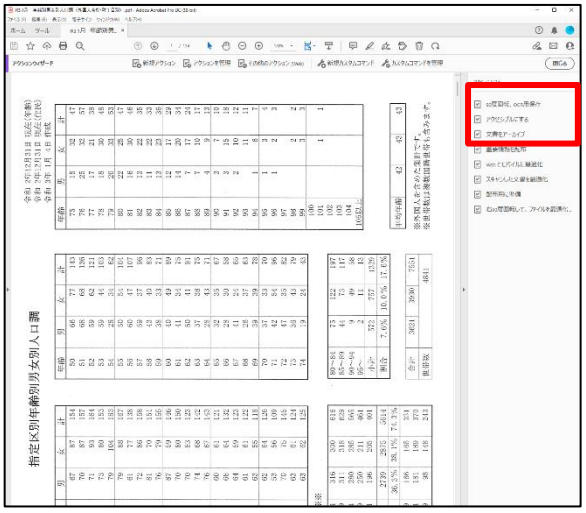
8. [保存]をクリック後、保存名を「OCR 用」と入力し、[OK]をクリックする

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



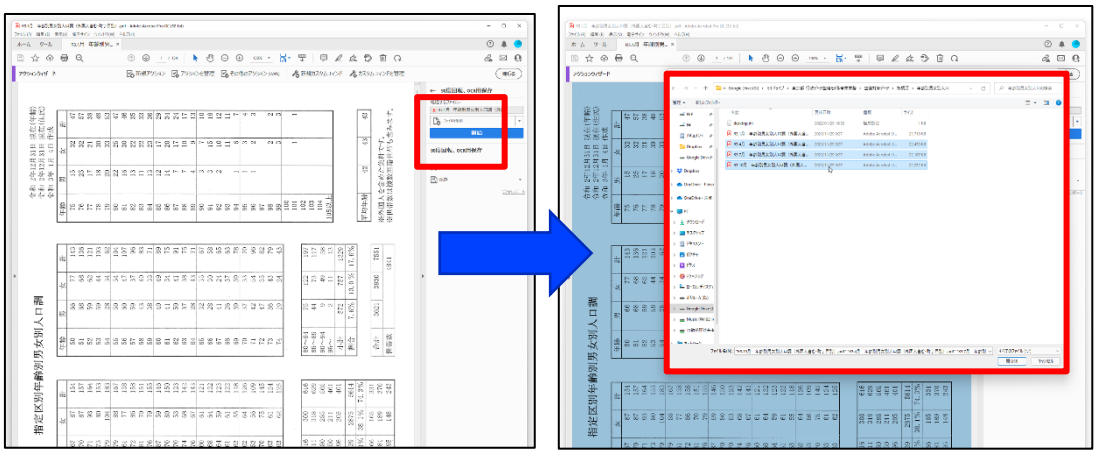


アクション名に追加されます。

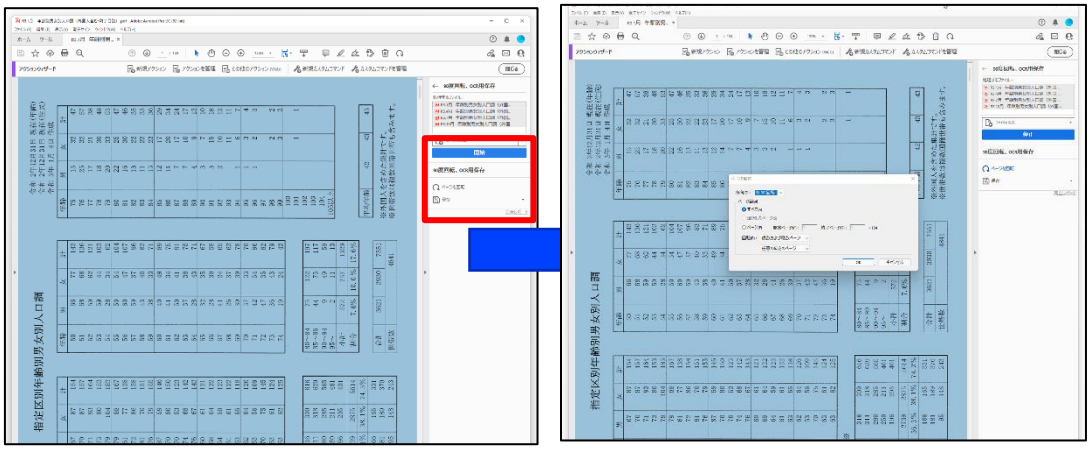
ⓒ東京都

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

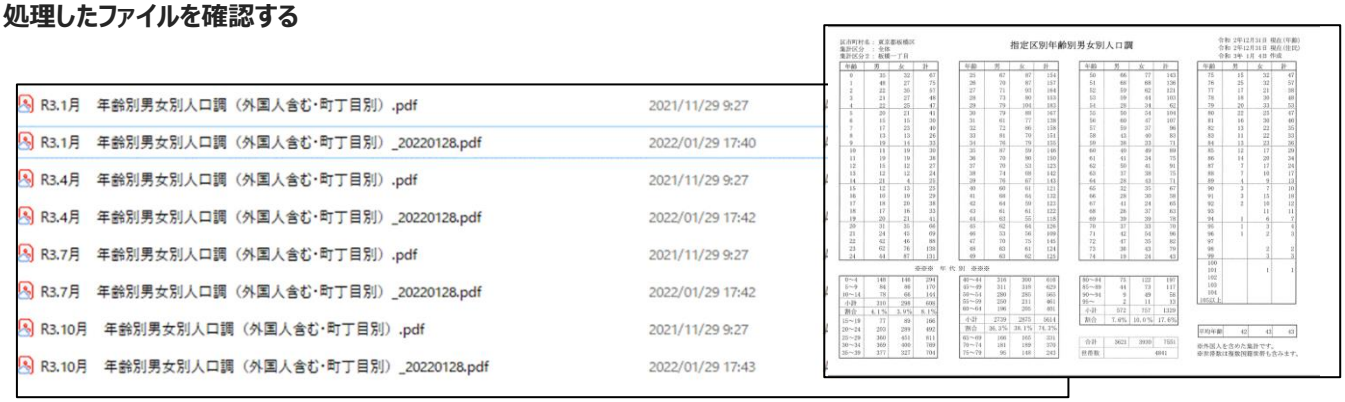


2. [開始]をクリックし、ページの回転で[OK]をクリックする



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

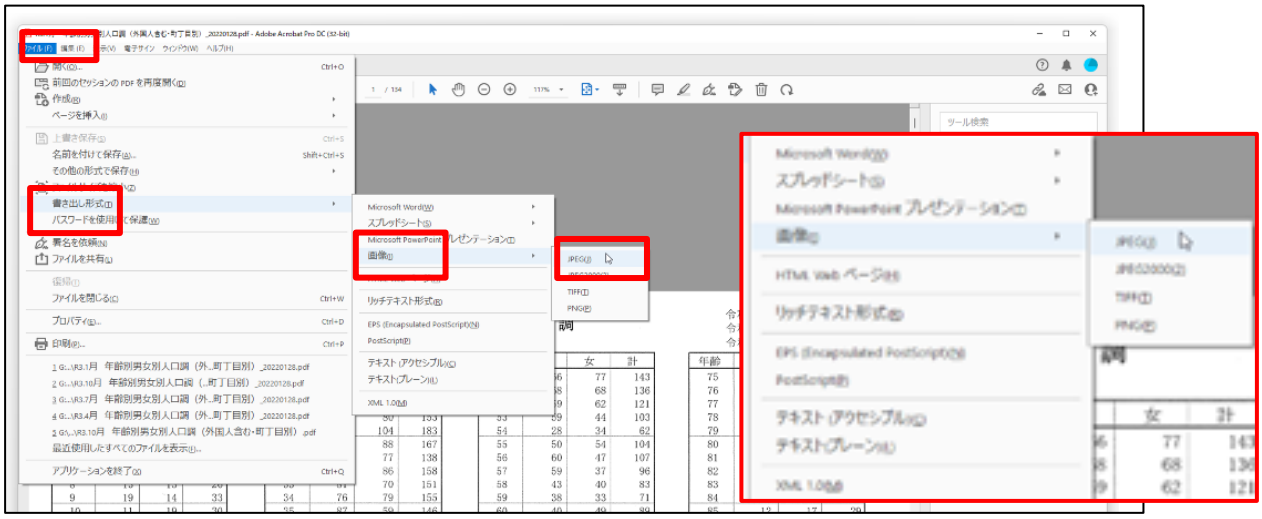


R3.1月　年齢別男女別人口調（外国人含む･町丁目別）.pdf	2021/11/29 9:27
R3.1月　年齢別男女別人口調（外国人含む･町丁目別）_20220128.pdf	2022/01/29 17:40
R3.4月　年齢別男女別人口調（外国人含む･町丁目別）.pdf	2021/11/29 9:27
R3.4月　年齢別男女別人口調（外国人含む･町丁目別）_20220128.pdf	2022/01/29 17:42
R3.7月　年齢別男女別人口調（外国人含む･町丁目別）.pdf	2021/11/29 9:27
R3.7月　年齢別男女別人口調（外国人含む･町丁目別）_20220128.pdf	2022/01/29 17:42
R3.10月　年齢別男女別人口調（外国人含む･町丁目別）.pdf	2021/11/29 9:27
R3.10月　年齢別男女別人口調（外国人含む･町丁目別）_20220128.pdf	2022/01/29 17:43

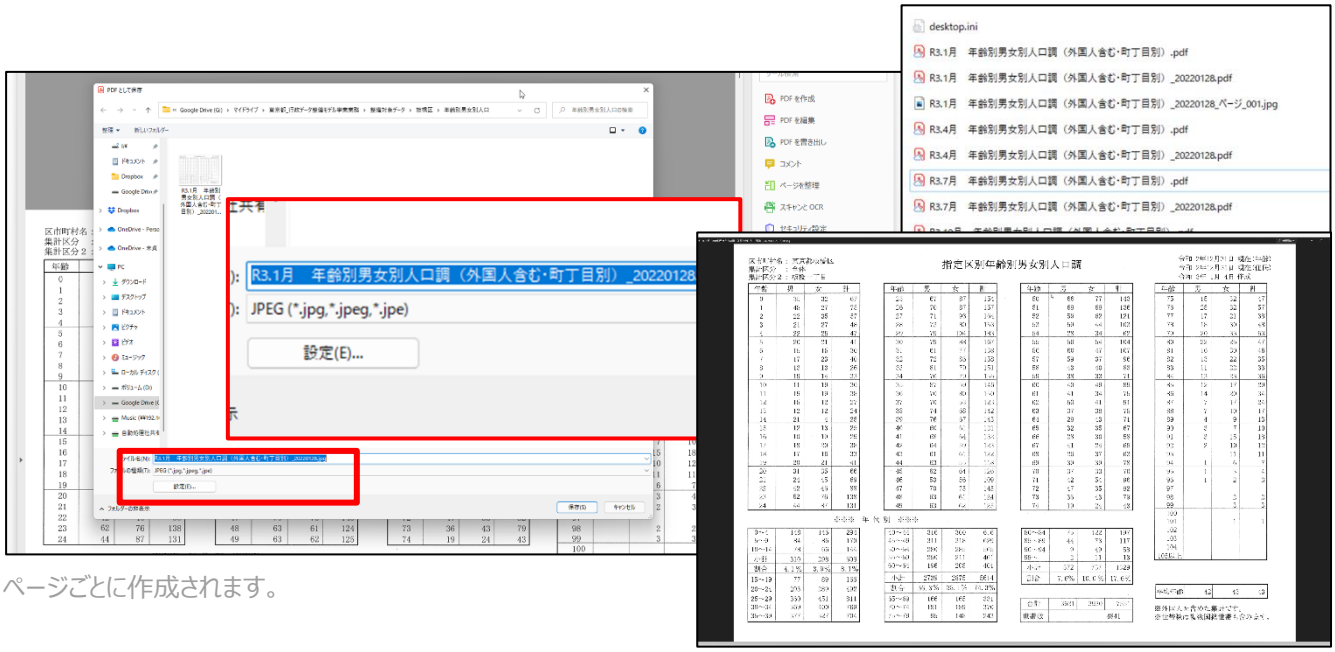
ファイル名末尾に指定した日付が入ったファイルが今回作成したデータです。

ⓒ東京都

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

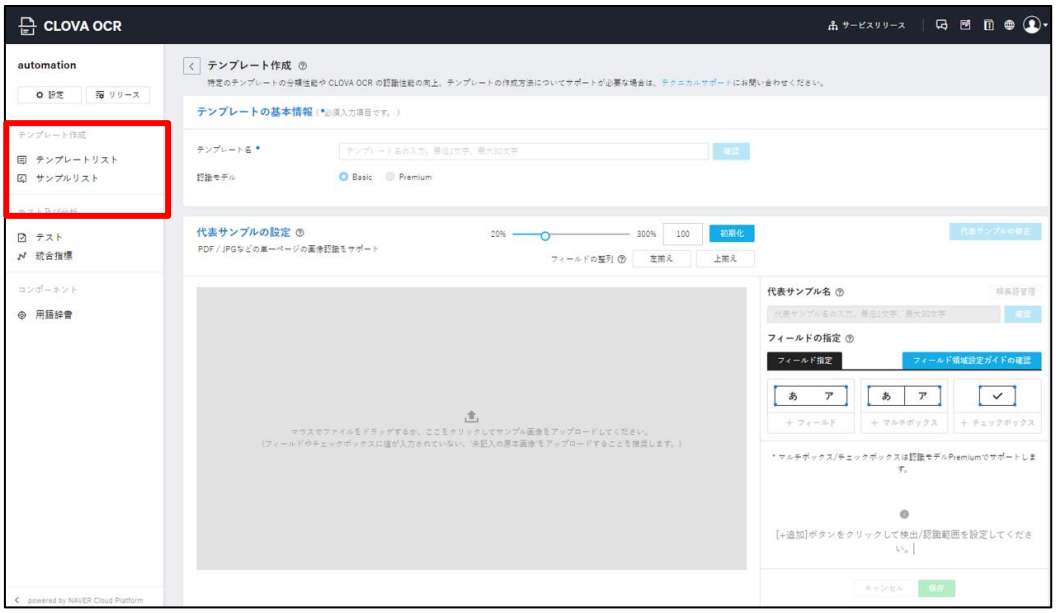


ページごとに作成されます。

ⓒ東京都

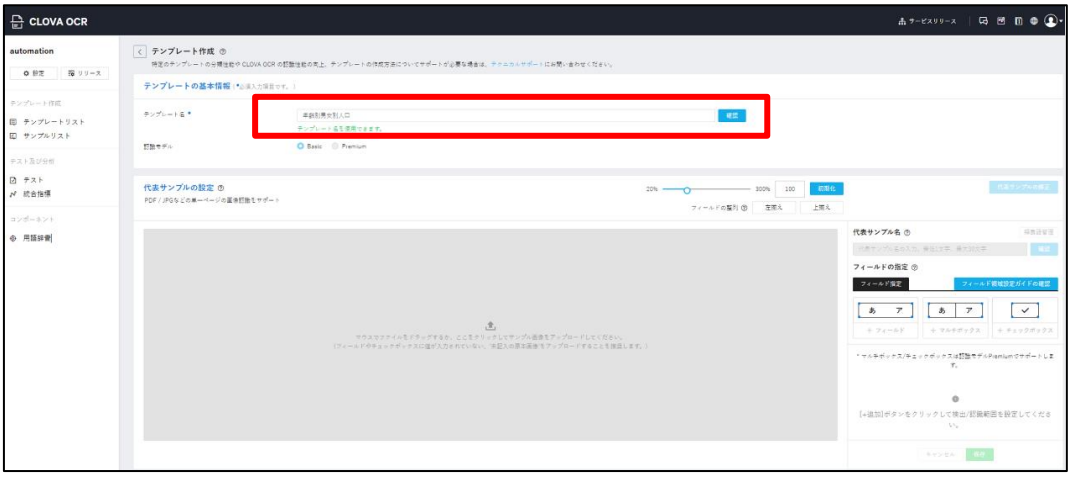
③ CLOVA OCR READER で、AI-OCR 処理するテンプレートを作成する

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



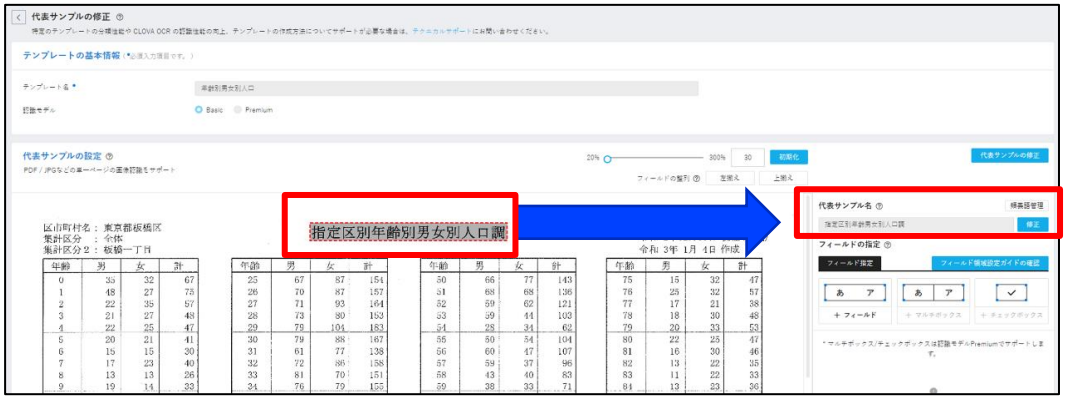
テンプレート名が確定されたことを確認してください。

ⓒ東京都

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

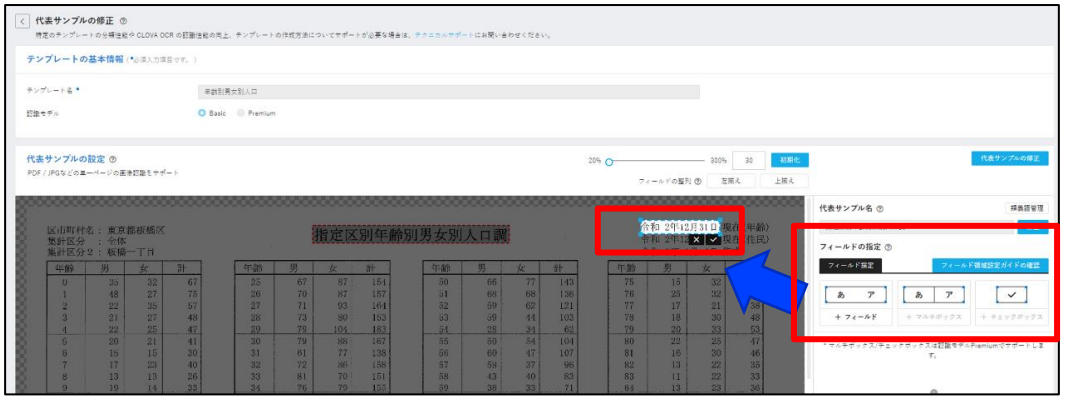
領域色が反転すると、ファイルが認識されています。そのまま保持状態を解除します。

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。

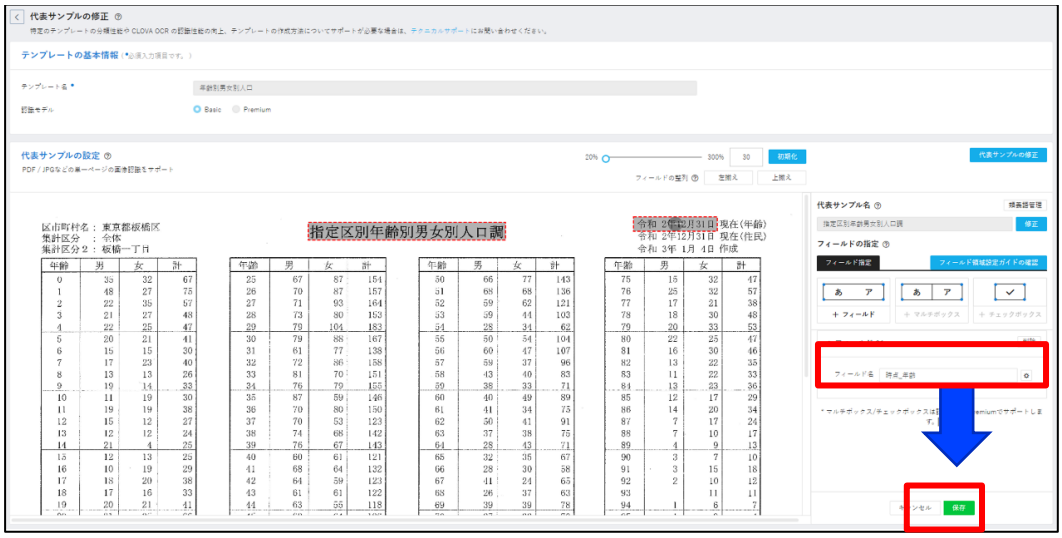
5. [+フィールド]をクリックし、読み取りたい領域を選択する



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。

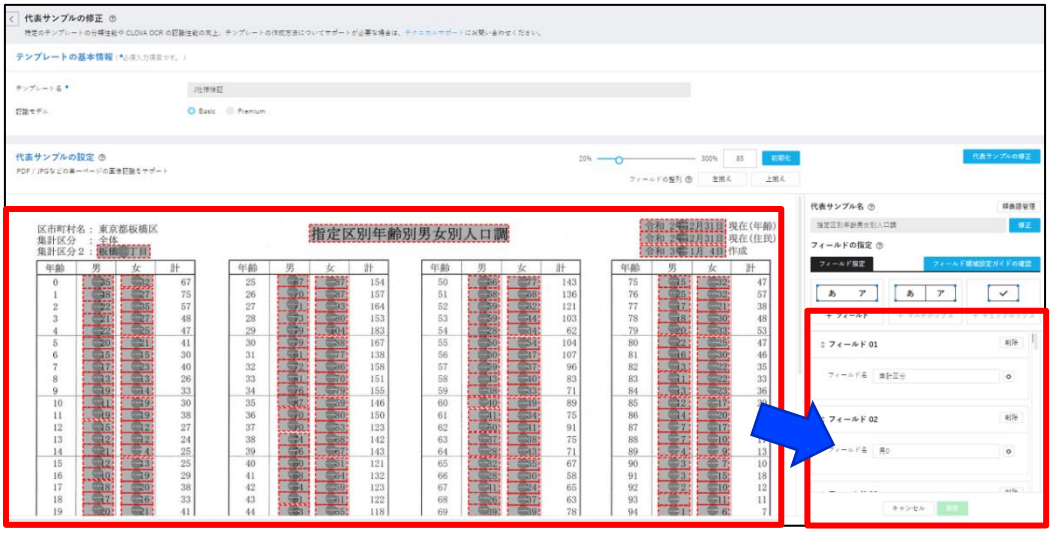
ⓒ東京都

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。

ⓒ東京都

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。

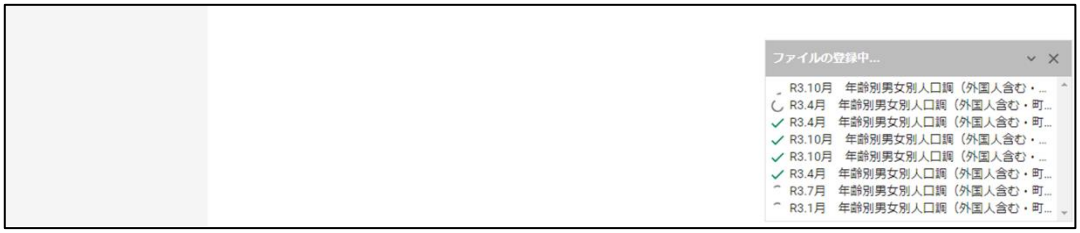
画面右側が、ファイルをアップロードする領域です。ここにドラッグ操作か、領域をクリックしてファイルをアップロードします。

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。

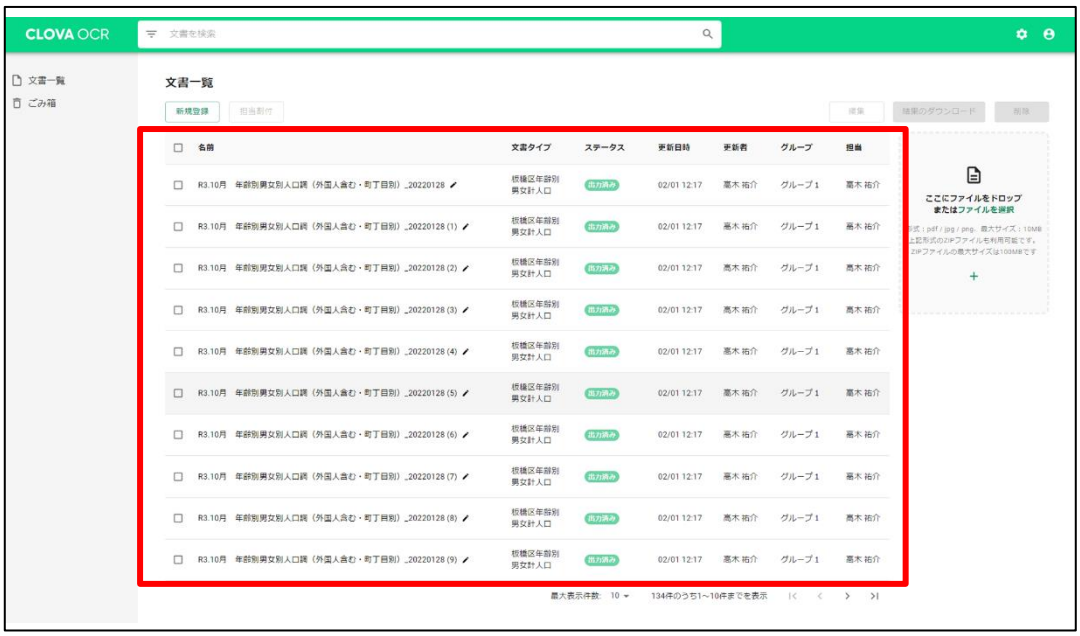
ⓒ東京都

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、



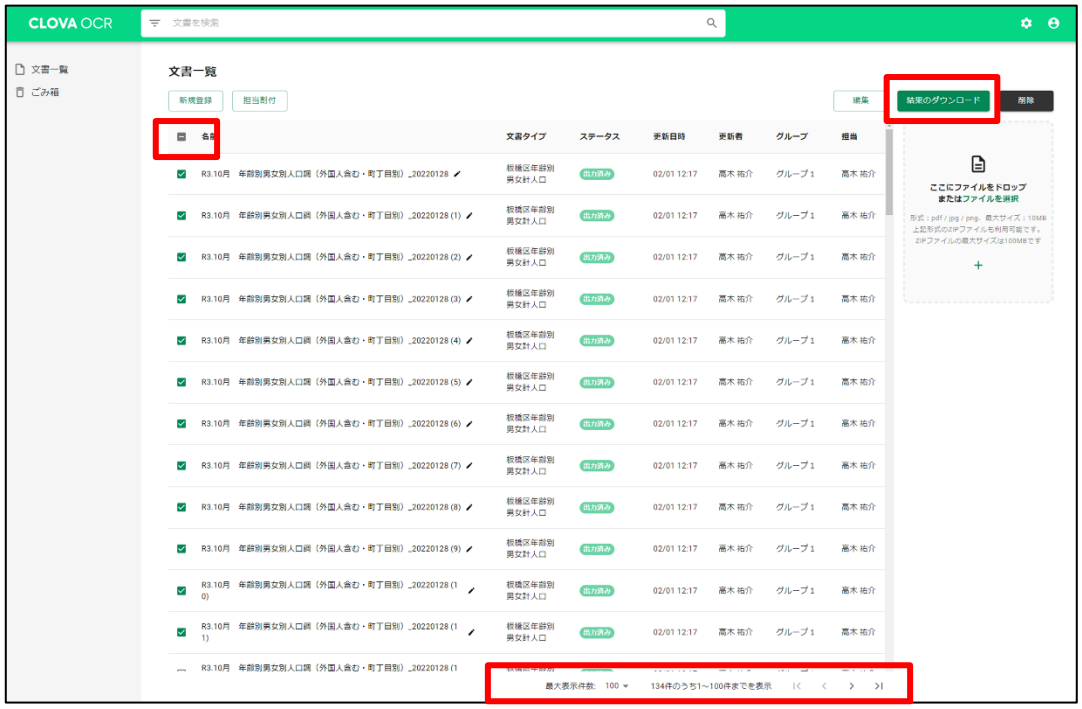
読み込みがはじまります。



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。

ⓒ東京都

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。

⑤ 処理結果を目視にて確認しクレンジング・正規化方針を決定する

AI-OCR は、全ての文字を認識できるわけではなく、癖の強い手書き文字や印刷時に潰れた文字等は誤認識や、帳票のフォーマットの崩れ（罫線位置等の帳票毎の揺れ）等によって、意図しない変換結果となるケースもあります。
そのため、目視でデータの荒れや特性を確認し、データの値の格納場所のズレや、誤認識等のパターンを把握、データのクレンジングや正規化といった後続処理について方針を決定する必要があります。結果、手作業でのクレンジングが必要となるケースもあります。
データの確認や後続処理については、後述の「データの値チェック」「文字列正規化」「データの目視によるクレンジング」をご参照ください。

OCR 起因のデータ荒れの例：

有料老人ホーム「シタシ郷」　：ハイフンや記号の誤認識（長音の「ー」を漢字の「一」（いち）と認識）
サンプル神社　：類似の漢字や潰れた文字のご認識（「社」を「杜」と認識）
その他、打消し線も含めた文字読取や、枠外にはみ出たデータの認識漏れ等

ⓒ東京都

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うことが望ましいです。この章に記載している作業や各パートの最後に記載されている委託事業者へ依頼時の仕様イメージを参考に、外部委託を行い、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。

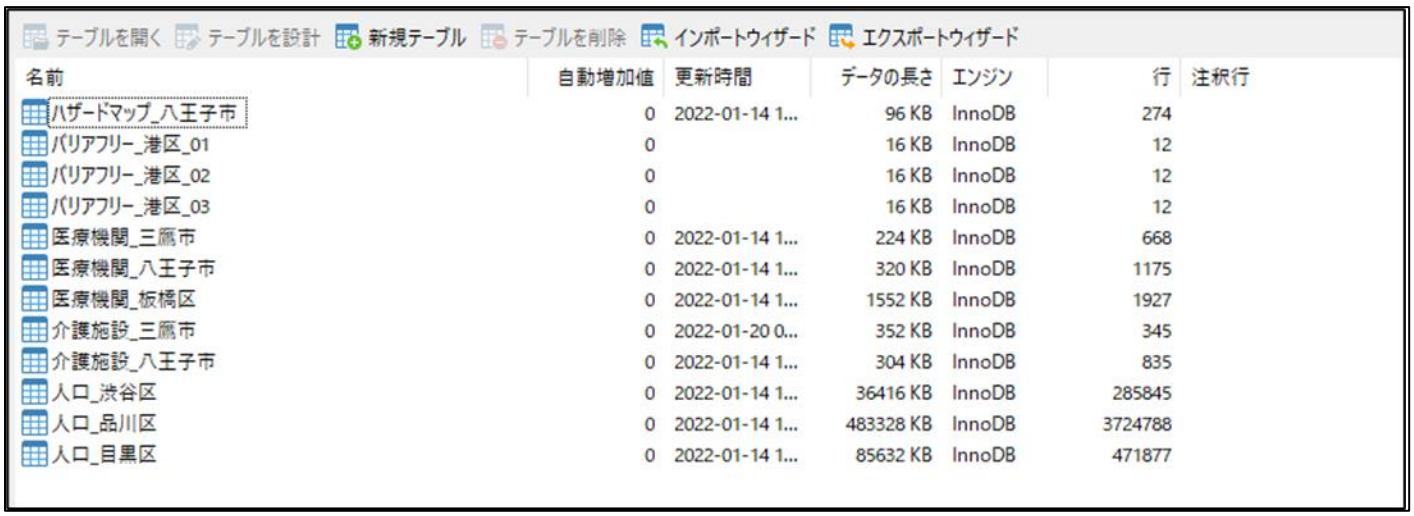
① これまでの手順で整備した CSV データをデータベースが読み込める場所に格納する （参考 p. 103）
② データをマッピングする （参考 p.110）
③ マッピングデータを目視で確認する
④ ドメインデータを判別する
⑤ ドメインごとにクレンジングを行う （参考「クレンジング難度が高い例」p. 116）
⑥ データを正規化する （参考 p.121）
⑦ ドメインデータをチェックする

データの格納

件数にもよりますが、本作業は委託業者での目視による手作業を想定しています。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

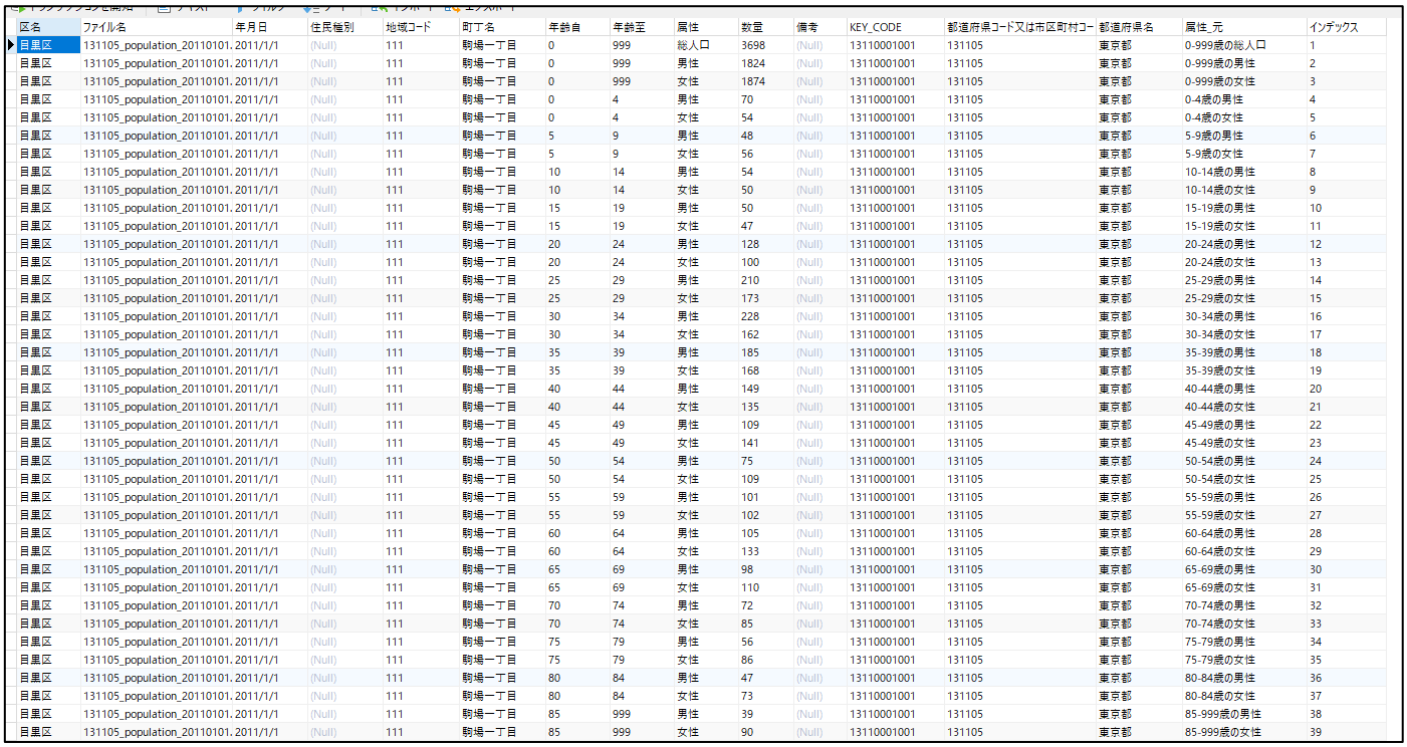
```
SELECT 項目 1、項目 2、項目 3、項目 4 FROM テーブル
```



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	

ⓒ東京都

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）

【データの格納における委託事業者へ依頼時の仕様イメージ】

項目	内容	備考
作業の目的・概要	データ整備作業や出力を効率的に行うために、CSV・エクセルのデータをデータベースに格納	
アウトプット	整備対象データが格納されたデータベース （元データの項目・データは基本的にすべて格納されている）	
作業内容の想定例	・ 元データにセルの結合が含まれる場合は解除し、値を補完 ・ 項目名が空欄の場合は適切な名称を追加 ・ 格納後に全データが格納されたかのチェック　等	

ⓒ東京都

データの判別

件数にもよりますが、本作業は委託業者でのプログラムによる自動処理もしくは、目視による手作業を想定しています。

各データ項目に対してデータプロファイルを実施し、データの型や格納箇所間違いを判別する。必要に応じでデータを修正します。

データプロファイル関数を使用し、データ型及び格納列の間違いを検出し後続作業へ適用していきます。

参考ロジック ： 営業所所在地に実施する SAMPLE

```
select
    SF_GET_DOMAIN(営業所所在地)
from
    渋谷区_届出施設一覧_131130_food_business_todokede_all
```

実行結果

DOMAIN	RESULT
事業所名称	99%
法人名称	0%
電話番号	0%
住所	1%
郵便番号	0%
事業所コード	0%
名称_かな	0%
名称_英語	0%

格納されているデータをプロファイリングした結果、データ型は事業所名称であり、
一部住所データらしいデータが格納されていることが判別できます。
間違えて格納されたと思われる住所データは、適切な項目へ再設定します。

ⓒ東京都

【データの判別における委託事業者へ依頼時の仕様イメージ】

項目	内容	備考
作業の目的・概要	文字列の正規化等の実施前に、各データ項目別に、想定しているデータの型に合わせたデータチェックを行い、データの型や格納箇所間違いを発見し、修正する	
アウトプット	基本的な機械的なチェックによりデータの型や格納間違いが修正されたデータ	・ 定量的にアウトプットの品質を定義しにくいため、作業内容を明確に指示することが望ましいです。
作業内容の想定例	・ データの型にエラーがないかのチェック（数字が入るべきデータ項目に数字以外のデータが入っていないかなど） ・ 文字列の長さのチェック（住所等で文字数が 5 文字以下などは明らかにエラーデータ） ・ その他各データの型に合わせたエラーチェック作業 ・ エラーデータに対するデータの補完・修正	・ 委託事業者別に独自ノウハウが存在している部分であり、委託事業者に相談した上で、作業内容等を具体化することが望ましいです。

ⓒ東京都

文字列正規化

件数にもよりますが、本作業は委託業者でのプログラムによる自動処理もしくは、目視による手作業を想定しています。

各データ項目に対して文字列正規化処理を実施し、文字単位の基本的な表記ゆれを解消します。改行などの制御文字についても削除を実施します。縮退マップを利用した文字列の縮退や記号の異体字についても考慮し作業を実施します。人間が目視すれば同じ情報だと判断できる場合でも、文字情報が別になると、機械では判別できません。その為、出来る限る統一した文字にして情報公開を行う必要があります。
※文字列正規化方針は、各ドメイン（名称、住所等）毎に自動処理社基準にて設定しています。

代表的な文字列正規化

アルファベット(半角文字に統一)

スポーツオアシス　ＡＢＣＤＥＦ	→	スポーツオアシス　ABCDEF
ｘｘ-ｘｘｘ代官山	→	xx-xxx 代官山

カナ(全角文字に統一)

株式会社ｼﾞﾄﾞｳｼｮﾘ	→	株式会社ジドウショリ
株式会社 ﾄｳｷｮｳﾄ	→	株式会社トウキョウト

数字(半角に統一)

株式会社サンプル　９Ｆ	→	株式会社サンプル　9F
日本著作権協会　４Ｆ　カップ	→	日本著作権協会　4F　カップ

無駄な余白の表記統一(全角空有白一つに統一)

abcde　タカギ薬局原宿店	→	abcde　タカギ薬局原宿店
学校法人 IT 学園　　新プラザ棟	→	学校法人 IT 学園　新プラザ棟

法人格表示揺れ統一

（株）自動処理	→	株式会社自動処理
かぶ）自動処理	→	株式会社自動処理
㈲手動処理	→	有限会社　手動処理

MJ 縮退マップによる異体字の統一

髙橋太郎	→	高橋太郎

ハイフンや記号の異体字の統一

-	→	―
-	→	―

※上記全て同じ文字に見えますが、別の文字です。

参考 ： その他文字列正規化(今回処理には使用していないが、参考掲載)

読み仮名付与、ローマ字変換、常用漢字への縮退等

ⓒ東京都

【文字列正規化における委託事業者へ依頼時の仕様イメージ】

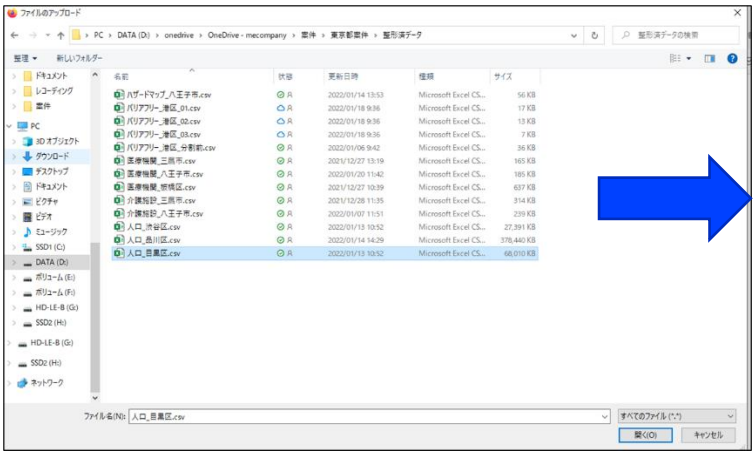
項目	内容	備考
作業の目的・概要	各データ項目に対して、文字単位の基本的な表記揺れを判別し、想定している文字列の型に合わせて修正する	
アウトプット	基本的な機械的なチェックにより、文字単位の表記揺れが解消された状態	・定量的にアウトプットの品質を定義しにくいため、作業内容を明確に指示することが望ましいです。
作業内容の想定例	・アルファベット、カナ、数字、漢字、記号等の、文字列種別毎の表記揺れのチェック（半角/全角、空白の長さ、常用漢字/常用外漢字など） ・MJ 縮退マップを利用して、異体字を変換 ・記号の異体字を統一 ・省略表現による表記揺れのチェック（法人格の表記など） ・その他の表記揺れのチェック（読み仮名、ローマ字など） ・表記揺れの修正方針の決定、統一化作業	・委託事業者別に独自ノウハウが存在している部分であり、委託事業者に相談した上で、作業内容や統一化方針等を具体化することが望ましいです。

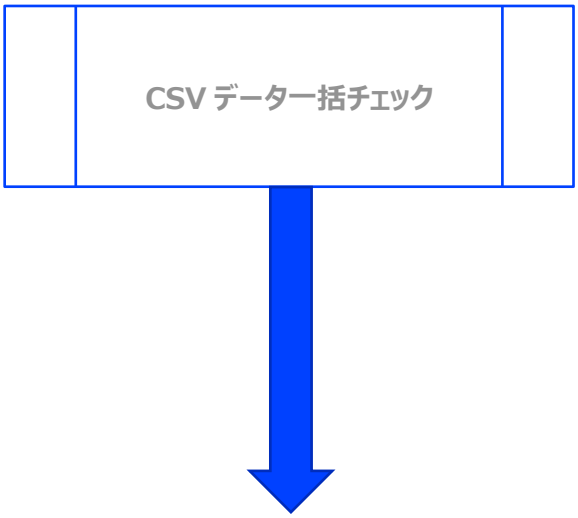
ⓒ東京都

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。

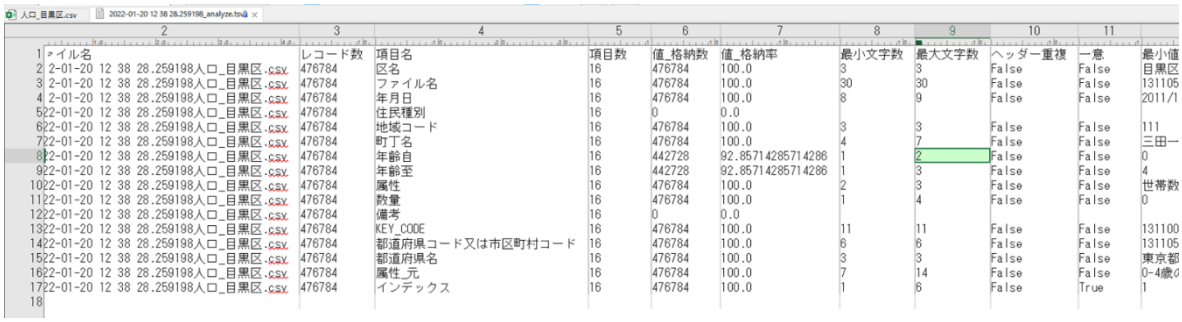
対象ファイルを選択





※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	22-01-20 12 38 28.259198人口_目黒区.csv	476784	住民種別	16	0	0.0					
6	22-01-20 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	22-01-20 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	22-01-20 12 38 28.259198人口_目黒区.csv	476784	年齢自	16	442728	92.85714285714286	1	2	False	False	0
9	22-01-20 12 38 28.259198人口_目黒区.csv	476784	年齢至	16	442728	92.85714285714286	1	3	False	False	4
10	22-01-20 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	22-01-20 12 38 28.259198人口_目黒区.csv	476784	数量	16	476784	100.0	1	4	False	False	0
12	22-01-20 12 38 28.259198人口_目黒区.csv	476784	備考	16	0	0.0					
13	22-01-20 12 38 28.259198人口_目黒区.csv	476784	KEY_CODE	16	476784	100.0	11	11	False	False	131100
14	22-01-20 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	22-01-20 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	22-01-20 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	22-01-20 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
18											

※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。

ⓒ東京都

データのマッピング

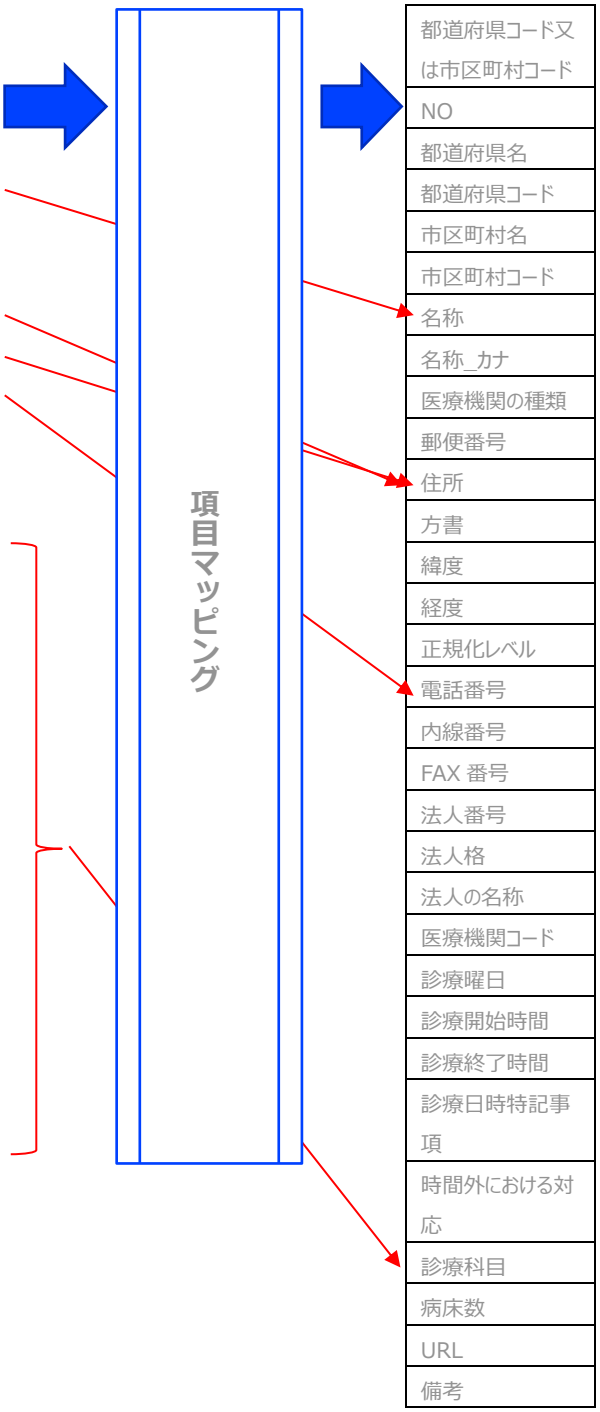
本作業は委託業者でのプログラムによる自動処理を想定しています。

データのフォーマット統一化

各様式・フォーマットを、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



ⓒ東京都

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`
```


ⓒ東京都

例 ： 食品営業届出(渋谷区)

ID、住所、名称、POI コード、緯度経度、法人情報などについては、データ正規化、データ補完にて実施します。

インデックス	1	2
都道府県コード又は市区町村コード	131130	131130
NO	2100400521	2100400526
都道府県名	東京都	東京都
市区町村名	渋谷区	渋谷区
営業所名称	ＳＡＭＰＬＥ　バークレー東京店	データオフィス冷凍冷蔵庫１２３４５６
営業所名称_カナ		
営業の種類	乳類販売業	その他の食料・飲料販売業
業態		
営業所所在地	東京都渋谷区渋谷 1-2-3	東京都渋谷区神南 4-5-6
営業所方書	青ビル１１１．１１２号	皆　1F
緯度		
経度		
営業所電話番号	03-6666-1111	050-7777-6666
法人名	有限会社イーアンドエム	グリコチャネルクリエイト株式会社
法人番号		
許可番号	4 渋保生食届第 999 号	4 渋保生食届第 888 号
初回許可年月日		
許可年月日		
許可開始日		
許可満了日		
廃業年月日		
申請区分		
許可条件		
備考		
届出日	2022/11/30	2022/11/30
個人営業者氏名		
代表者肩書	取締役	
代表者名	山田　真弓	鈴木太郎
営業者住所	神奈川県横浜市鶴見区下末吉３丁目４番５号	大阪府大阪市西淀川区歌島５丁目６－７
営業者方書		
営業者電話番号	045-555-3393	050-7777-6666
廃業番号		

ⓒ東京都

参考ロジック

```
select
　　営業所名称,
　　都道府県名,
　　市区町村名,
　　営業所所在地,
　　営業所電話番号
　　# ・・・・・・・・・
from
　　渋谷区_届出施設一覧_131130_food_business_todokede_all
```

主体別のデータ正規化・マッピング作業に関する所感

民間データは、データ値の表記揺れは少なく、正規化やクレンジング、データチェックの作業時間が少ない一方で、データフォーマットの独自色が強く、データマッピングの作業時間がかかる傾向にあります。

対して自治体データは、データ値の表記ゆれが多く、正規化やクレンジング、データチェックの作業時間が多い一方で、データフォーマットについては共通フォーマットに近しい形のデータが多く、データマッピングの作業時間は少ない傾向にあります。

主体	フォーマット	値の表記揺れ
民間データ	独自色が強い	少ない
自治体データ	標準的なフォーマットが多い	多い

【データのマッピングにおける委託事業者へ依頼時の仕様イメージ】

項目	内容	備考
作業の目的・概要	各データのフォーマットから、想定する共通フォーマットへ統一化する	
アウトプット	想定する共通フォーマットのデータ項目に、データ値が正しく格納された状態	
作業内容の想定例	・統一化する共通フォーマットの検討・決定（事前準備） ・統一化する共通フォーマットのデータ項目と、各データの様式・フォーマットのデータ項目の対応関係の整理 ・統一化する共通フォーマットのデータ項目に対応するデータ値の格納（マッピング） ・データ項目別のデータ正規化、不足するデータ項目に対するデータ補完の並行実施	・統一化するフォーマットの検討方法は、手順 1 の「データ構造の検討」「データ構造の調査」をご参照ください。 ・データ正規化とデータ補完については、後述の「データの正規化」「複雑なデータ補完」をご参照ください。

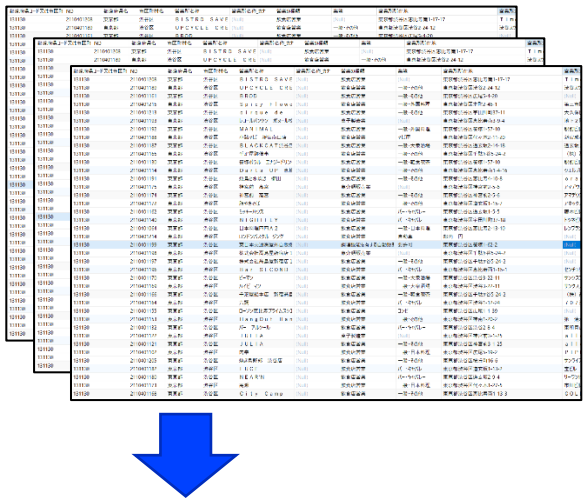
ⓒ東京都

データの値チェック

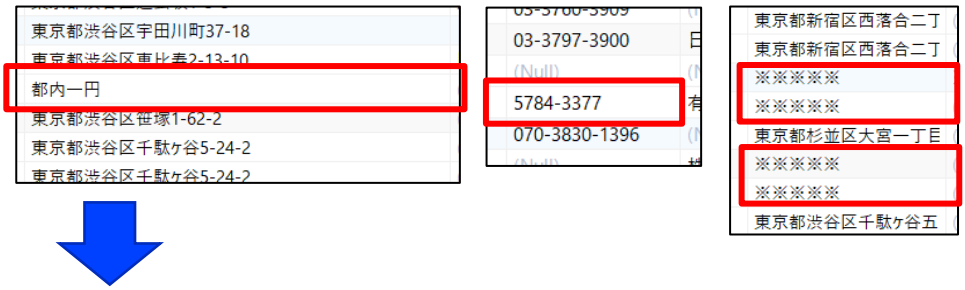
件数によりますが、本作業は委託業者での目視による手作業を想定しています。

目視によるデータ荒れ傾向の把握

各データに対して、目視で荒れデータを確認し、後続処理について方針を決定します。



目視にて、データ荒れや特性をチェックします。



データ荒れや特性に応じて、データのクレンジングや正規化方針を決定します。

データ荒れや特性の例

有料老人ホーム「付きの郷」	：介護サービス種別 ＋ 施設名が括弧付き
水曜は午後診あり 14:00～17:00	：時間表示が自由記載
※※※※※	：住所値として成り立たない値
（３階）、＃７０１、-602、（ＴＯＳグループ内）	：方書として荒れ地と思われる値
3518-2111、58306130	：市外局番表記なし。ハイフン無しなど表記の不統一
Mushizuka(Insect Grave)	：施設名_英語にカッコ表記
336917	：桁不足（本来は 10 桁）のコード値
（有）SAMPLE 新聞店、自動屋　上野店　（キーコーヒー）	：法人格表記ゆれ、カッコ表記
株式会社　　自動処理	：スペースが複数
山田太郎、鈴木　一郎（佐藤　太郎）	：姓名間のスペースなし。カッコ表記
．（ドット）、手動（てどう）	：施設名表記が自由記載
仕出し・定期、仕出し、一般・スナック、一般、店舗・一般	：業務形態について表記ゆれ有り

ⓒ東京都

【データの値チェックおける委託事業者へ依頼時の仕様イメージ】

項目	内容	備考
作業の目的・概要	データのクレンジングや正規化の実施前に、各データの荒れ傾向や特性を把握し、クレンジングや正規化の方針を決定する	
アウトプット	各データの荒れ傾向や特性と、それらに対応したデータのクレンジングや正規化方針の一覧	
作業内容の想定例	・ 目視によるデータ項目別のデータ値の制御状況の把握 （自由記載か否か、電話番号、氏名、施設名などの表記が統一されているか）	・ 対象データの質や量に応じ、作業量が変化するため、サンプルデータや課題感を事前に提示した上で、作業内容・範囲を決める形が望ましいです。

ⓒ東京都

データの目視によるクレンジング

本作業は委託業者でのプログラムによる自動処理もしくは、目視による手作業を想定しています。

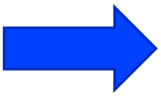
クレンジング難度が高い例（データ構造の課題）

「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

ⓒ東京都

経年でデータが変わる

平成31年3月1日現在

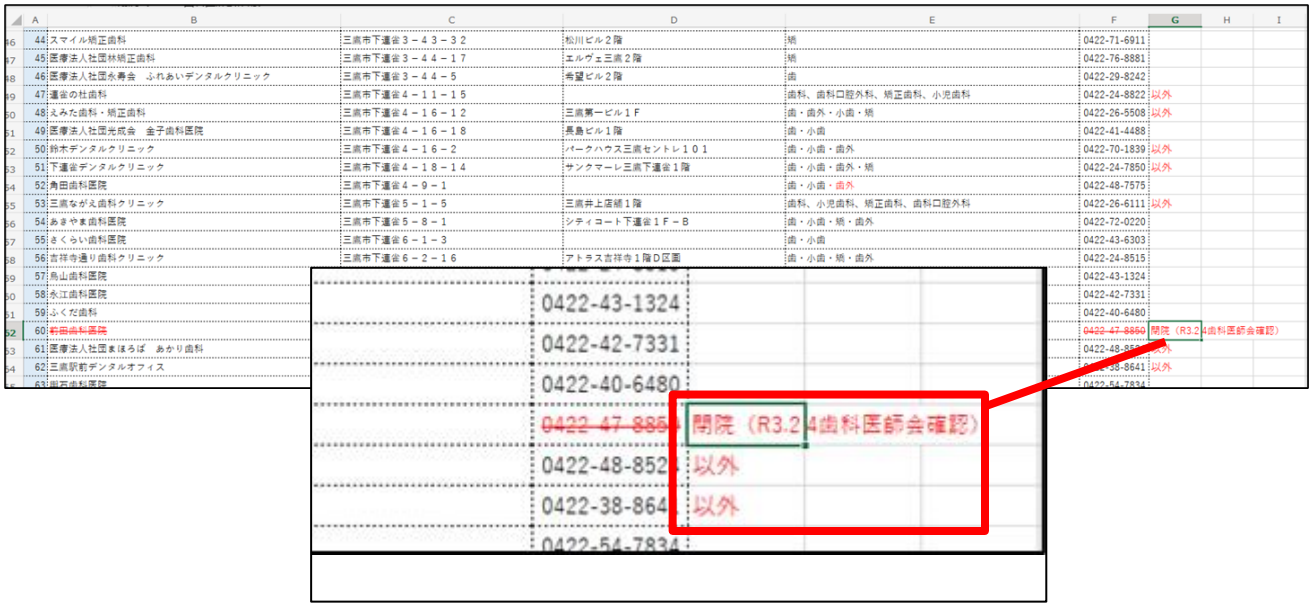
年齢	総数	男	女
総数	258,093	121,704	136,389
0歳～4歳	14,651	7,544	7,107
0歳	2,876	1,501	1,375
1歳	2,958	1,509	1,449
2歳	2,987	1,505	1,482
3歳	3,025	1,601	1,424
4歳	2,805	1,428	1,377
5歳～9歳	11,629	5,877	5,752
5歳	2,513	1,256	1,257
6歳	2,540	1,261	1,279
7歳	2,289	1,169	1,120
8歳	2,192	1,114	1,078
90歳～94歳	2,113	512	1,601
90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



ⓒ東京都

【データ構造に関するクレンジングにおける委託事業者へ依頼時の仕様イメージ】

項目	内容	備考
作業の目的・概要	目視で確認を行い、機械判読性の高いデータ構造に変換する	
アウトプット	機械判読性の高いデータ構造に変換されたデータ	・ 対象となるデータの構造により、作業量や実現可能なアウトプットの品質が異なるため、サンプルデータを提示した上で、作業内容・アウトプットを決める形が望ましいです。
作業内容の想定例	・ 1 セル 1 データへの変換 ・ データ項目の表記ゆれの統一 ・ 経年でデータ粒度が異なる場合の粒度統一 ・ 枠外にデータが追加されている場合のデータの取込	

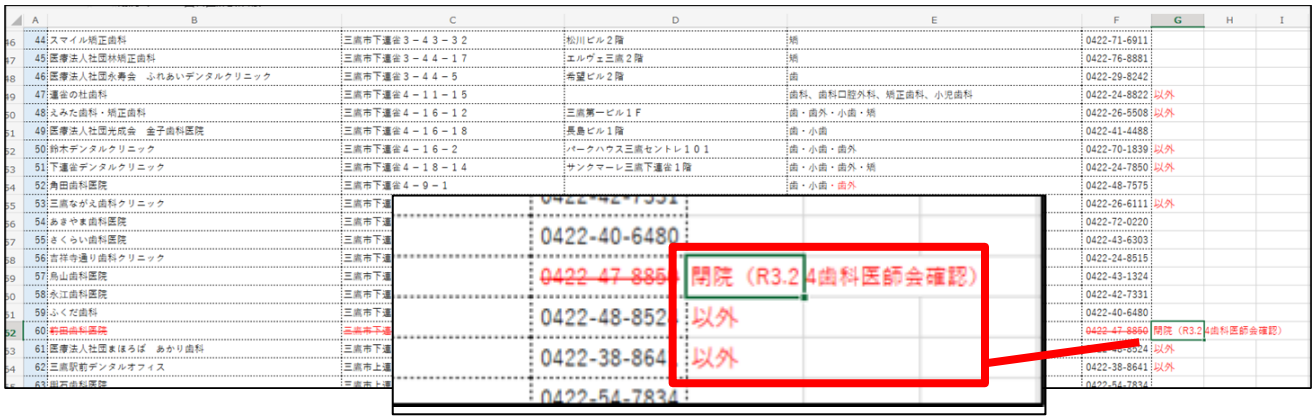
ⓒ東京都

クレンジング難度が高い例（データの課題）

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。

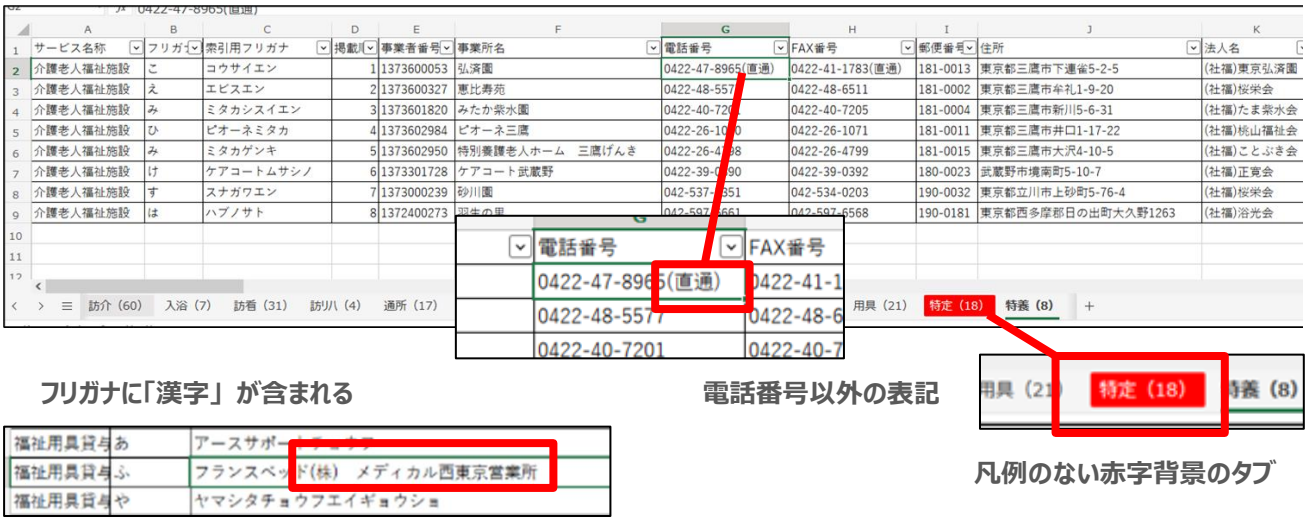
枠外に閉院データが表記されている



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

ⓒ東京都

【データに関するクレンジングにおける委託事業者へ依頼時の仕様イメージ】

項目	内容	備考
作業の目的・概要	目視で確認を行い、データの内容を機械判読性の高いデータに変換する	
アウトプット	機械判読性の高いデータに変換されたデータ ※アウトプットデータ検証のために「クレンジング作業別の修正箇所一覧」もあると望ましい	・ 対象となるデータの形式や値により、作業量や実現可能なアウトプットの品質が異なるため、サンプルデータを提示した上で、作業内容・アウトプットを決める形が望ましいです。
作業内容の想定例	・ 枠外にデータが追加されている場合のデータの取込 ・ 打消し線で修正されているデータへの対応（打消し線部分のデータの削除） ・ 異表記なデータ形式の修正	

ⓒ東京都

データの正規化

件数にもよりますが、本作業は委託業者でのプログラムによる自動処理もしくは、目視による手作業を想定しています。
データの正規化とは、一定のルールに基づいて、データを整える作業になります。

事業所名

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。
事業所名、名称のみが格納されていると思われがちですが、グループ法人名、法人名、事業所名、店舗名、サービス名など複数の情報が 1 つの項目に含まれている場合があります。特定の事業所を探している時に、事業所以外の情報が含まれていた場合、情報を確認して指示した事業所かどうかを人間が判別しないと、適切な情報かどうか判断がつきません。データを検索する際に適切に情報を格納しておくことで、事業所を簡単に検索する事が出来るようになります。

代表的な文字列正規化

法人、事業所、サービスの混在表記を、「グループ、法人、事業所、サービス名、店舗名」に分離

有料老人ホーム「自動処理苑」	→	- / - /自動処理苑 / 有料老人ホーム / -
ミニマート四谷駅前店	→	- / - /ミニマート / - / 四谷駅前店
医療法人社団 サンプル会 xx クリニック	→	- / 医療法人社団 サンプル会 / xx クリニック / -/ -

法人格表記ゆれを統一

（株）自動処理	→	株式会社自動処理
㈲手動処理	→	有限会社 手動処理

文字列（アルファベット、数字、漢字（異体字等））の表記ゆれを統一

株式会社 手動処理 １号店 Ｏｎｅ	→	株式会社 手動処理 １号店 One

参考ロジック ： 営業所名称に実施する SAMPLE

```
select
    SF_FMT_NAME(営業所名称)
from
    渋谷区_届出施設一覧_131130_food_business_todokede_all
```


ⓒ東京都

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
鳴呼、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）

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」

ⓒ東京都

※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」

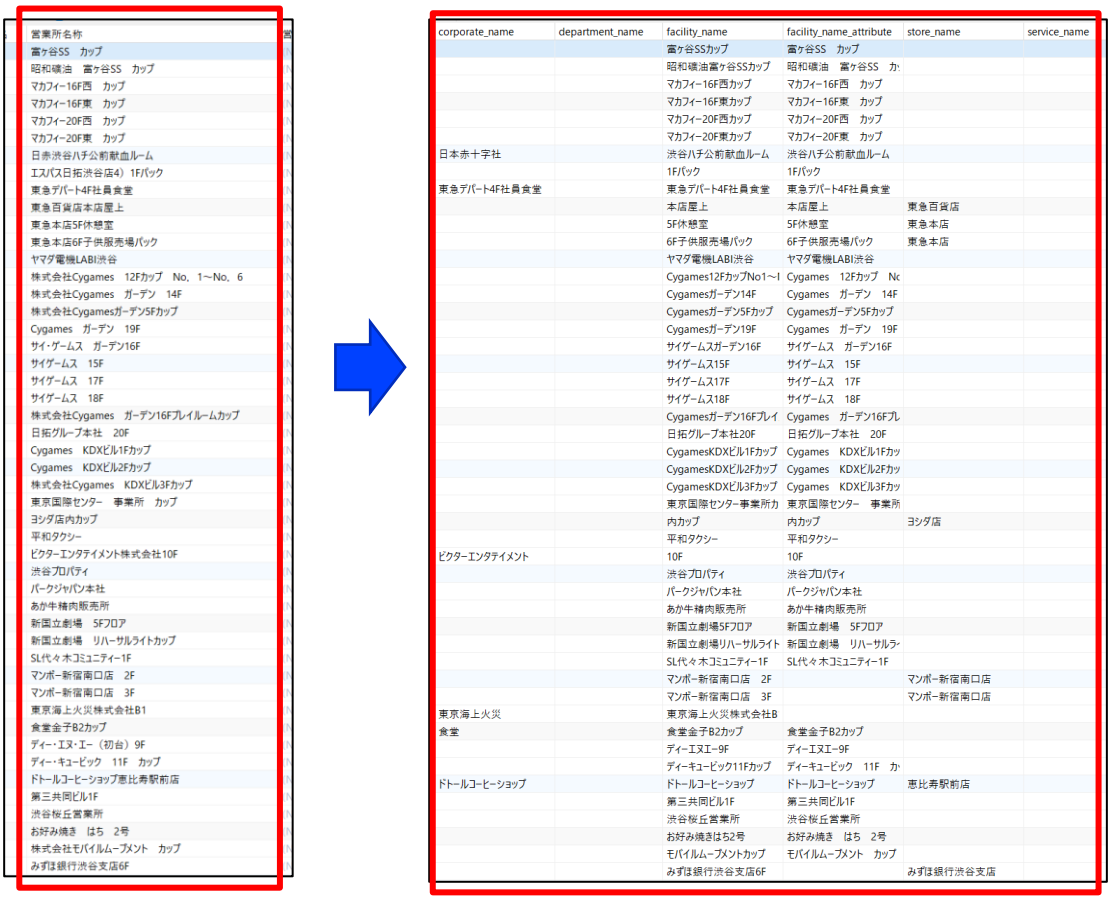
「表記揺れ統一後の正規化データ」では、法人格表記統一、スペース統一、記号全半角統一等、正規化内容の詳細な制御が可能です。

例 ： 法人格は削除等

ⓒ東京都

※前ページからの続き

json 形式で得た名称正規化の処理結果を、項目マッピングすることで正規化、分割、補完された利用可能データとなります。



営業所名称
富ヶ谷SS　カップ
昭和礦油　富ヶ谷SS　カップ
マカフィー16F西　カップ
マカフィー16F東　カップ
マカフィー20F西　カップ
マカフィー20F東　カップ
日赤渋谷ハチ公前献血ルーム
エスパス日拓渋谷店4）1Fパック
東急デパート4F社員食堂
東急百貨店本店屋上
東急本店5F休憩室
東急本店6F子供服売場パック
ヤマダ電機LABI渋谷
株式会社Cygames　12Fカップ　No. 1～No. 6
株式会社Cygames　ガーデン　14F
株式会社CygamesガーデンSFカップ
Cygames　ガーデン　19F
サイ・ゲームス　ガーデン16F
サイゲームス　15F
サイゲームス　17F
サイゲームス　18F
株式会社Cygames　ガーデン16Fプレイルームカップ
日拓グループ本社　20F
Cygames　KDXビル1Fカップ
Cygames　KDXビル2Fカップ
株式会社Cygames　KDXビル3Fカップ
東京国際センター　事業所　カップ
ヨシダ店内カップ
平和タクシー
ピクターエンタテイメント株式会社10F
渋谷プロパティ
パークジャパン本社
あか牛精肉販売所
新国立劇場　5Fフロア
新国立劇場　リハーサルライトカップ
SL代々木コミュニティー1F
マンボー新宿南口店　2F
マンボー新宿南口店　3F
東京海上火災株式会社B1
食堂金子B2カップ
ディー・エヌ・エー（初台）9F
ディー・キュービック　11F　カップ
ドトールコーヒーショップ恵比寿駅前店
第三共同ビル1F
渋谷桜丘営業所
お好み焼き　はち　2号
株式会社モバイルムーブメント　カップ
みずほ銀行渋谷支店6F

corporate_name	department_name	facility_name	facility_name_attribute	store_name	service_name
		富ヶ谷SSカップ	富ヶ谷SS　カップ		
		昭和礦油富ヶ谷SSカップ	昭和礦油　富ヶ谷SS　カ		
		マカフィー16F西カップ	マカフィー16F西　カップ		
		マカフィー16F東カップ	マカフィー16F東　カップ		
		マカフィー20F西カップ	マカフィー20F西　カップ		
		マカフィー20F東カップ	マカフィー20F東　カップ		
日本赤十字社		渋谷ハチ公前献血ルーム	渋谷ハチ公前献血ルーム		
		1Fパック	1Fパック		
東急デパート4F社員食堂		東急デパート4F社員食堂	東急デパート4F社員食堂		
		本店屋上	本店屋上	東急百貨店	
		5F休憩室	5F休憩室	東急本店	
		6F子供服売場パック	6F子供服売場パック	東急本店	
		ヤマダ電機LABI渋谷	ヤマダ電機LABI渋谷		
		Cygames12FカップNo1～I	Cygames　12Fカップ　No		
		Cygamesガーデン14F	Cygames　ガーデン　14F		
		CygamesガーデンSFカップ	CygamesガーデンSFカップ		
		Cygamesガーデン19F	Cygames　ガーデン　19F		
		サイゲームスガーデン16F	サイゲームス　ガーデン16F		
		サイゲームス15F	サイゲームス　15F		
		サイゲームス17F	サイゲームス　17F		
		サイゲームス18F	サイゲームス　18F		
		Cygamesガーデン16Fプレイ	Cygames　ガーデン16Fプレ		
		日拓グループ本社20F	日拓グループ本社　20F		
		CygamesKDXビル1Fカップ	Cygames　KDXビル1Fカッ		
		CygamesKDXビル2Fカップ	Cygames　KDXビル2Fカッ		
		CygamesKDXビル3Fカップ	Cygames　KDXビル3Fカッ		
		東京国際センター事業所カ	東京国際センター　事業所		
		内カップ	内カップ	ヨシダ店	
		平和タクシー	平和タクシー		
ピクターエンタテイメント		10F	10F		
		渋谷プロパティ	渋谷プロパティ		
		パークジャパン本社	パークジャパン本社		
		あか牛精肉販売所	あか牛精肉販売所		
		新国立劇場5Fフロア	新国立劇場　5Fフロア		
		新国立劇場リハーサルライト	新国立劇場　リハーサルラ		
		SL代々木コミュニティー1F	SL代々木コミュニティー1F		
		マンボー新宿南口店　2F		マンボー新宿南口店	
		マンボー新宿南口店　3F		マンボー新宿南口店	
東京海上火災		東京海上火災株式会社B			
食堂		食堂金子B2カップ	食堂金子B2カップ		
		ディーエヌエー9F	ディーエヌエー9F		
		ディーキュービック11Fカップ	ディーキュービック　11F　カ		
ドトールコーヒーショップ		ドトールコーヒーショップ	ドトールコーヒーショップ	恵比寿駅前店	
		第三共同ビル1F	第三共同ビル1F		
		渋谷桜丘営業所	渋谷桜丘営業所		
		お好み焼きはち2号	お好み焼き　はち　2号		
		モバイルムーブメントカップ	モバイルムーブメント　カップ		
		みずほ銀行渋谷支店6F		みずほ銀行渋谷支店	

ⓒ東京都

住所

住所は基本４情報に含まれる一般的なデータですが、実は正式な定義が難しいデータです。例えば住所に含まれる長目の後ろの字といわれる地名は、旧住所情報が含まれることがあり、字がなくても成立する住所が数多くあります。その為、住所として字を入れる場合と入れない場合でもどちらでも住所として成立します。また表記についても、漢数字と数字の混在や、郵便番号を含む事で、郵便番号までの住所を省略しているケースや、建物名を省略する郵便バーコードによる表記、漢字の異体字を含むか含まないかなど、様々な表記パターンがあります。その為、住所が同一である事をデータ上で判別する事は簡単なようで、機械には非常に難しいです。
その為、自動的に正規化処理を施して、画一的な住所情報を記載するか、将来的には、デジタル庁で進めているアドレスベースレジストリ（ベータ版）を元に住所を統一して表記するか、町字 ID と一緒に記載する事が望ましいですが、**現時点では、標準的な定義がない為、デジタル庁の標準化を待つのが望ましいです。現時点では民間企業各社が独自データを元に正規化精度を競っている状態ですので、公平な基準で精度を測り、作業依頼を行う事が大事です。**
正規化処理処理を行う事で、別のファイルに保存されている住所が同じ場所を表すという事を示す事が機械的に判別しやすくすることが出来ます。

代表的な文字列正規化

都道府県、市区町村が全欠損、一部欠損

新宿区四谷 4-33-2 マーキュリースクエア 2F	→	**東京都**新宿区四谷 4-33-2 マーキュリースクエア 2F
東京都四谷 4-33-2 マーキュリースクエア 2F	→	東京都**新宿区**四谷 4-33-2 マーキュリースクエア 2F

同一住所異表記

東京都四谷４丁目-33-2 マーキュリースクエア 2F	→	東京都新宿区四谷 4-33-2 マーキュリースクエア 2F
東京都四谷四丁目-33-2 マーキュリースクエア 2F	→	東京都新宿区四谷 4-33-2 マーキュリースクエア 2F
東京都四谷四丁目-33-2-2F	→	東京都新宿区四谷 4-33-2 マーキュリースクエア 2F
160-0004 33-2-2F	→	東京都新宿区四谷 4-33-2 マーキュリースクエア 2F

ヶ、ケ、が、ガ、の、ノ、ツ、っ、ッの表記混在

東京都千代田区霞**が**関 2-1-3	→	東京都千代田区霞**ヶ**関 2-1-3

市区町村合併前の表記、建物番号や階数の混在表記

東京都**保谷市**南町 1-2-3	→	東京都**西東京市**南町 1-2-3

異体字混在

東京都葛飾区立石 5-13-1	→	東京都葛飾区立石 5-13-1

町字 ID

0002002	→	東京都千代田区霞が関二丁目

郵送確認（不達の為、データを削除した例）

東京都葛飾区立石 5-13-1	→	

ⓒ東京都

システム処理例 ： 東京都渋谷区渋谷 1-2-3

JSON

```
{
  results : {
    id : 11832822
    area : "渋谷"
    city : "渋谷区"
    ward : ""
    chome : "2丁目"
    geoid : 20035182
    county : ""
    remain : "青山台ビル101．102号"
    azacode : "13113012002"
    govcode : "13113"
    is_only : true
    jisx401 : "13"
    zipcode : "1500002"
    latitude : 35.660742
    street_1 : ""
    street_2 : ""
    longitude : 139.70738
    subarea_1 : "9"
    subarea_2 : "10"
    subarea_3 : ""
    subarea_4 : ""
    subarea_5 : ""
    subarea_6 : ""
    subarea_7 : ""
    norm_level : 7
    prefecture : "東京都"
    building_info : {
      building_name : "青山台ビル"
      building_room : "101"
      building_floor : ""
      building_number : ""
      building_remain : "102号"
      remain_banchi_fussy : ""
      building_room_suffix : ""
      building_floor_prefix : ""
      building_floor_suffix : ""
      building_number_suffix : ""
    }
    before_building : "青山台ビル101．102号"
    delivery_address : "東京都渋谷区渋谷2-9-10青山台ビル101．102号"
    matching_address : "東京都渋谷区渋谷2丁目-9-10"
    scuel_address_id : "OMA8Y8YSFNMV19M"
    normalize_address : "東京都渋谷区渋谷2丁目-9-10"
    remain_after_area : "9-10青山台ビル101．102号"
    remain_after_ward : "渋谷2-9-10青山台ビル101．102号"
    is_discontinue_area : 0
    is_discontinue_city : 0
    largeoffice_zipcode : ""
    remain_after_subarea : "青山台ビル101．102号"
    machiaza_resolve_address : "東京都渋谷区渋谷2丁目9-10青山台ビル101．102号"
    normalize_address_lossy1 : "東京都渋谷区渋谷2丁目-9"
    normalize_address_lossy2 : "東京都渋谷区渋谷2丁目"
    normalize_postal_address : "東京都渋谷区渋谷2丁目-9-10-101"
    delivery_normalize_address : "東京都渋谷区渋谷2-9-10青山台ビル101．102"
    scuel_normalize_string_code : "SSNC-552E-PH0C-5GH9"
    machiaza_resolve_normalize_address : "東京都渋谷区渋谷2丁目9-10青山台ビル101．102"
  }
```

```
  metadata : {
    log : {
      log : {
        is_text : false
        is_long_text : false
        is_occur_error : false
        is_found_compcode : false
        not_found_pref_city : false
        is_pref_city_and_number : false
      }
      errors : {
        code : 0
        target : ""
        more_info : ""
        user_message : ""
        internal_message : ""
      }
      log_level : ""
    }
    type : ""
    title : "scuel_normalizer#convert_domain_data"
    status : "Success"
    version : "635576f"
    parameter : {
      options : [
        0 : ""
      ]
      log_level : ""
      input_data : "東京都渋谷区渋谷2-9-10青山台ビル１０１．１０２号"
      output_type : "230"
    }
    result_set : {
      count : 1
      status : {
        has_aza : false
        has_oaza : false
        is_area_fussy : false
        is_found_area : true
        is_found_city : true
        is_found_ward : false
        is_old_address : false
        no_suffix_area : false
        no_suffix_city : false
        no_suffix_ward : false
        complement_area : false
        complement_city : false
        complement_ward : false
        is_found_county : false
        is_subarea_fussy : false
        complement_county : false
        complement_street : false
        no_suffix_subarea1 : false
        complement_subarea1 : false
        convert_new_address : false
        is_found_prefecture : true
        no_suffix_prefecute : false
        complement_prefecute : false
        is_found_subarea1_by_area : false
        street_has_maybe_many_area : false
      }
      normalize_address_id : 0
      scuel_normalize_address_id : "SNAD-552E-PH0C-5GHA"
    }
  }
}
```


参考ロジック ： 営業所所在地に実施する SAMPLE

```
select
    SF_FMT_ADDRESS(営業所所在地)
from
    渋谷区_届出施設一覧_131130_food_business_todokede_all
```


ⓒ東京都

例（住所：八王子）

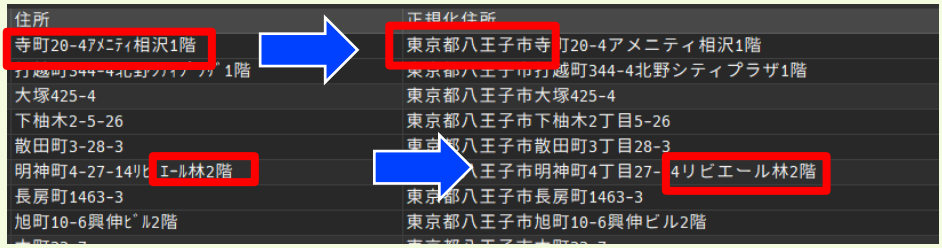
住所
寺町20-4ｱﾒﾆﾃｨ相沢1階
打越町344-4北野ｼﾃｨﾌﾟﾗｻﾞ1階
大塚425-4
下柚木2-5-26
散田町3-28-3
明神町4-27-14ﾘﾋﾞｴｰﾙ林2階
長房町1463-3
旭町10-6興伸ﾋﾞﾙ2階
本町33-7
大塚622-9ﾆｭｰﾊｲﾑ井上201号
高尾町1602
元横山町2-1-20
狭間町1450-1めじろ台ｺｰﾎﾟﾗｽ
みなみ野3-9-7
八日町4-9幸ﾋﾞﾙ3階
長沼町200-3旗野ｺｰﾎﾟﾗｽ1階
片倉町443-4
子安町4-6-1Phiﾋﾞﾙ3階
子安町4-6-1Phiﾋﾞﾙ2階
明神町3-26-10土屋ﾋﾞﾙ1階
東浅川町879-6
子安町1-2-8守屋ﾋﾞﾙ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ﾌﾚｽｺ南大沢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現時点ではない為、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



ⓒ東京都

※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	{"results": {"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北野ｼﾃｨﾌﾟﾗｻﾞ1階	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"id":	{"results": {"normalize_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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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	{"results": {"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。

ⓒ東京都

※前ページからの続き

json 形式で得た正規化（分割）処理結果を、項目マッピングすることで正規化、分割、補完された利用可能データとなります。

営業所所在地
東京都渋谷区宇田川町7-1
東京都渋谷区宇田川町12-18
東京都渋谷区宇田川町16-4
東京都渋谷区宇田川町21-1
東京都渋谷区恵比寿1丁目23-13
東京都渋谷区恵比寿南2丁目4-1
東京都渋谷区恵比寿南3丁目5-7
東京都渋谷区大山町45-14
東京都渋谷区渋谷1丁目11-12
東京都渋谷区渋谷3丁目28-7
東京都渋谷区神泉町18-11
東京都渋谷区神宮前4丁目26-5
東京都渋谷区神宮前5丁目53-67
東京都渋谷区神宮前5丁目53-67
東京都渋谷区神宮前6丁目20-10
東京都渋谷区神南1丁目22-6
東京都渋谷区千駄ヶ谷5丁目21-3
東京都渋谷区富ヶ谷1丁目10-8
東京都渋谷区富ヶ谷1丁目15-1
東京都渋谷区富ヶ谷2丁目20-3
東京都渋谷区道玄坂1丁目12-1

govcode	prefecture	county	city	ward	remain	remain_after_ward	remain_after_area	remain_after_subarea	street_1	street_2	area	chome	subarea_1	subarea_2	subarea_3	subarea_4	subarea_5	subarea_6	subarea_7	jisx401	azacode	zipcode	norm_level
13113	東京都	(Null)	渋谷区	(Null)	アムウェイ本社ビル13階	宇田川町7-1アムウェイ本社	7-1アムウェイ本社ビル13階	アムウェイ本社ビル13階	(Null)	(Null)	宇田川町	(Null)	7	1	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113003000	1500042	7
13113	東京都	(Null)	渋谷区	(Null)	東急ハンズ渋谷店7階	宇田川町12-18東急ハンズ	12-18東急ハンズ渋谷店7階	東急ハンズ渋谷店7階	(Null)	(Null)	宇田川町	(Null)	12	18	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113003000	1500042	7
13113	東京都	(Null)	渋谷区	(Null)	-1階	宇田川町16-4-1階	16-4-1階	-1階	(Null)	(Null)	宇田川町	(Null)	16	4	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113003000	1500042	7
13113	東京都	(Null)	渋谷区	(Null)	(Null)	宇田川町21-1	21-1	(Null)	(Null)	(Null)	宇田川町	(Null)	21	1	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113003000	1500042	7
13113	東京都	(Null)	渋谷区	(Null)	アルカイビル2階	恵比寿1-23-13アルカイビル	23-13アルカイビル2階	アルカイビル2階	(Null)	(Null)	恵比寿	1丁目	23	13	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113004001	1500013	7
13113	東京都	(Null)	渋谷区	(Null)	-3階	恵比寿南2-4-1-3階	4-1-3階	-3階	(Null)	(Null)	恵比寿南	2丁目	4	1	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113006002	1500022	7
13113	東京都	(Null)	渋谷区	(Null)	(Null)	恵比寿南3-5-7	5-7	(Null)	(Null)	(Null)	恵比寿南	3丁目	5	7	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113006003	1500022	7
13113	東京都	(Null)	渋谷区	(Null)	(Null)	大山町45-14	45-14	(Null)	(Null)	(Null)	大山町	(Null)	45	14	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113007000	1510065	7
13113	東京都	(Null)	渋谷区	(Null)	-4階	渋谷1-11-12-4階	11-12-4階	-4階	(Null)	(Null)	渋谷	1丁目	11	12	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113012001	1500002	7
13113	東京都	(Null)	渋谷区	(Null)	青ビル6階	渋谷3-28-7青ビル6階	28-7青ビル6階	青ビル6階	(Null)	(Null)	渋谷	3丁目	28	7	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113012003	1500002	7
13113	東京都	(Null)	渋谷区	(Null)	(Null)	神泉町18-11	18-11	(Null)	(Null)	(Null)	神泉町	(Null)	18	11	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113014000	1500045	7
13113	東京都	(Null)	渋谷区	(Null)	-3階	神宮前4-26-5-3階	26-5-3階	-3階	(Null)	(Null)	神宮前	4丁目	26	5	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113015004	1500001	7
13113	東京都	(Null)	渋谷区	(Null)	コスモス青山1階	神宮前5-53-67コスモス青山	53-67コスモス青山1階	コスモス青山1階	(Null)	(Null)	神宮前	5丁目	53	67	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113015005	1500001	7
13113	東京都	(Null)	渋谷区	(Null)	コスモス青山西棟4階休憩1	神宮前5-53-67コスモス青山	53-67コスモス青山西棟4階	コスモス青山西棟4階休憩1	(Null)	(Null)	神宮前	5丁目	53	67	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113015005	1500001	7
13113	東京都	(Null)	渋谷区	(Null)	MIYASHITAPARKSouth3F	神宮前6-20-10MIYASHIT	20-10MIYASHITAPARKSc	MIYASHITAPARKSouth3F	(Null)	(Null)	神宮前	6丁目	20	10	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113015006	1500001	7
13113	東京都	(Null)	渋谷区	(Null)	渋谷マルイ1階イベントスペ	神南1-22-6渋谷マルイ1階	22-6渋谷マルイ1階イベント	渋谷マルイ1階イベントスペ	(Null)	(Null)	神南	1丁目	22	6	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113016001	1500041	7
13113	東京都	(Null)	渋谷区	(Null)	-5階	千駄ヶ谷5-21-3-5階	21-3-5階	-5階	(Null)	(Null)	千駄ヶ谷	5丁目	21	3	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113017005	1510051	7
13113	東京都	(Null)	渋谷区	(Null)	(Null)	富ヶ谷1-10-8	10-8	(Null)	(Null)	(Null)	富ヶ谷	1丁目	10	8	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113019001	1510063	5
13113	東京都	(Null)	渋谷区	(Null)	(Null)	富ヶ谷1-15-1	15-1	(Null)	(Null)	(Null)	富ヶ谷	1丁目	15	1	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113019001	1510063	7
13113	東京都	(Null)	渋谷区	(Null)	(Null)	富ヶ谷2-20-3	20-3	(Null)	(Null)	(Null)	富ヶ谷	2丁目	20	3	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113019002	1510063	7
13113	東京都	(Null)	渋谷区	(Null)	-1渋谷マークシティ16階	道玄坂1-12-1渋谷マークシ	12-1渋谷マークシティ16階	-1渋谷マークシティ16階	(Null)	(Null)	道玄坂	1丁目	12	1	(Null)	(Null)	(Null)	(Null)	(Null)	13	13113020001	1500043	7

※norm_level から、住所値の正規化精度の取得も可能である（高いほど正確）。

ⓒ東京都

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。
デジタル庁が政府相互運用性フレームワーク（GIF）のコアデータパーツ 電話番号[1]では、電話番号は RFC3966 に準拠するように求められています。自治体が公開するデータは市外局番が概ね決まっている事が多く、市外局番を省略する事も多いですが、複数自治体のデータを一括で集める際には、市外局番を補完しなければ電話番号が欠けられない場合があります。今回の事業の中では住所情報を元に、市外局番の補完を行い、表記を統一し、データを扱いやすくしています。

代表的な文字列正規化

電話番号以外の情報記載（内線番号等）。有効な電話番号情報のみ抽出

03-2222-3333(内線:999)	→	(03)2222-3333

電話番号区切り一の不正。区切り文字の欠落。表記の統一。RFC3966 形式に変換

1111-1111	→	(03)1111-1111
444455-55	→	(03)4444-4444
03-1234-5678	→	(03)1234-5678

市外局番の欠落。住所情報より市外局番、市内局番の補完

1111-1111	→	(03)1111-1111
03--5555	→	(03)4444-4444

架電確認（不通の為、データを削除した例）

1111-1111	→	

名称と電話番号から電話番号を補完

住所	東京都渋谷区神宮前 3-21-21	
電話番号	6464-2626	→ (03)6464-2626

市外局番補完、表記フォーマット統一されたデータを項目マッピングすることで正規化、分割、補完された利用可能データとなります。

[1]コアデータパーツ 電話番号 https://github.com/JDA-DM/GIF/blob/main/440_%E3%82%B3%E3%82%A2%E3%83%87%E3%83%BC%E3%82%BF%E3%83%91%E3%83%BC%E3%83%84/md/445_core_dataparts_phonenumber.md

ⓒ東京都

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

（次ページ）

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

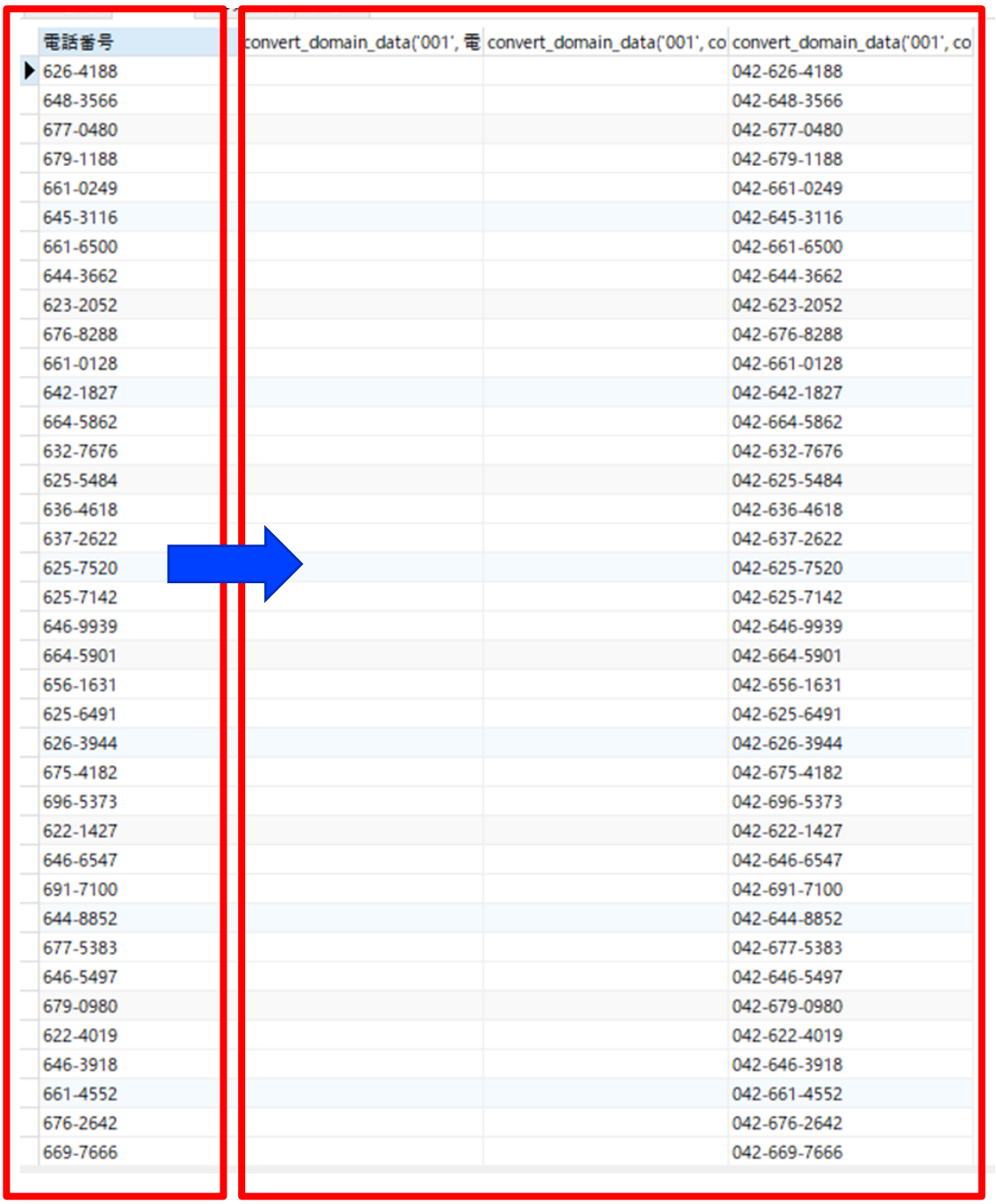
一部地域で 03 以外の番号を付与しています。

- 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

ⓒ東京都

※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

ⓒ東京都

【データの正規化における委託事業者へ依頼時の仕様イメージ】

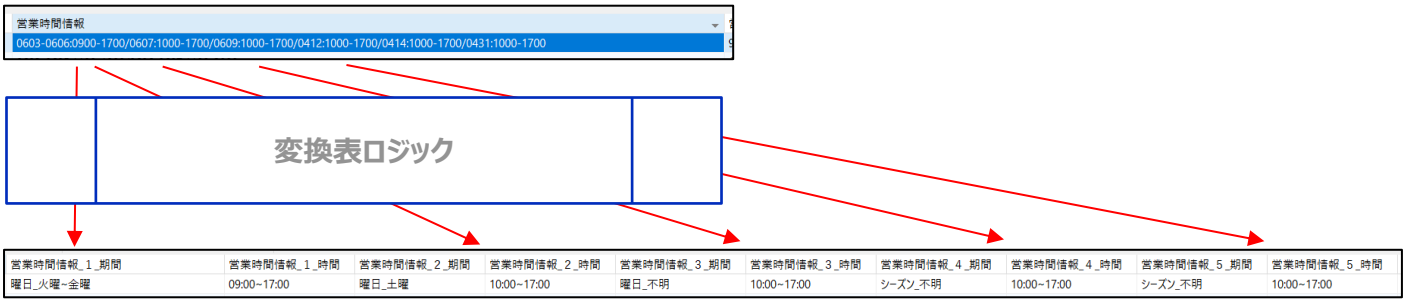
項目	内容	備考
作業の目的・概要	データを最終化（オープンデータとしての公開等）を行う前に住所・事業所名・電話番号に文字欠け・表記ゆれなどの修正を行う	
アウトプット	事業所名・住所・電話番号が正規化された（半角・全角の統一や誤字・脱字の修正等が施された）データベース 【事業所の正規化の状態の定義例】 ◆法人、事業所、サービスの混在表記を、「グループ、法人、事業所、サービス名、店舗名」に分離されている ◆法人格は省略表記しない ◆法人格と会社名の間に空白は含めない ◆アルファベット、数字は半角統一 ◆異体字・旧字は常用漢字に統一 【住所の正規化の状態の定義例】 ◆都道府県、市区町村、住所 1（丁目レベル）、住所 2（号レベル ＋ 建物名他）の 4 つの構成にする（都道府県や県名や郡名が省略されていない） ◆住所の表記（地名に含まれる「ヶ、ケ、が、ガ」「の、ノ」「ツ、っ、ッ」等）は郵便局の住所データに合わせる ◆異体字・旧字は常用漢字に統一 ◆旧住所は新住所に統一（市区町村合併処理） ◆住所 1、住所 2 の表記の統一 ◆郵送確認を行い正しい住所か郵送確認 【電話番号】 ◆RFC3966 形式　（XX）XXXX-XXXX ◆文字列はすべて半角 ◆有効な電話番号情報のみ抽出 ◆有効な電話番号か架電確認	・データ項目別に正規化された状態の定義を明確化することが重要です。 ・データの具体的な処理については、データの状態により正規化出来るかどうかや、整備目標が変わります。また要件についても明確に記載する事は難しいですが、調達する際に『電話番号の値が存在するデータに対して、99.9%以上の整備が出来る事を目標とする』など目標値を設定する事で、調達要件とすることが出来ます。 ・妥当な数値目標を設定するには、整備難度に応じた網羅的なケースを用意し、調達前にあらかじめ提示した上で、目標値を達成できるかどうか提案してもらうなどの方法をとることが出来ます。 ・求められるデータ品質によっては、住所は郵送確認、電話は架電確認などを含めると良いです。
作業内容の想定例	・住所・事業所名・電話番号に文字欠け・表記ゆれ等の確認 ・関数（住所正規化関数・名称正規化関数・電話番号正規化関数）の利用や、手動での一括変換により、文字欠け・表記ゆれ等を修正 ・修正後のデータに項目マッピングを行い、正規化・分割・補完されたデータ作成	・作業内容は、元のデータの状態、正規化された状態の定義により異なるため、元のデータのサンプルデータと正規化された状態の定義を委託事業者に定義した上で、作業内容を明確化することが望ましいです。

ⓒ東京都

1 項目に複数情報が設定されている場合

営業時間情報

値展開ロジック イメージ



・スラッシュ区切りで、1 項目に複数値が設定されている。

営業時間情報 1/営業時間情報 2 …営業時間情報 5

※営業時間は 0 個～ 5 個と格納数は様々（実データより）。

・変換表を参照しないと、意味のわかるデータとならない。

値フォーマット ：「大区分_小区分(-大区分_小区分:時間)」のデータフォーマットとなっている。

変換表

大区分	小区分
00 ： 不明	00 ： 不明
01 ： 月	01～12 ： 1月～12月、 21 ： 毎月
02 ： 日	01～31 ： 1日～31日
03 ： 旬	01 ： 初(上)旬、 02 ： 中旬、 03 ： 下旬
04 ： ｼｰｽﾞﾝ	01 ： ｵｰﾙｼｰｽﾞﾝ、 02 ： ｵﾌｼｰｽﾞﾝ、 03 ： ｵﾝｼｰｽﾞﾝ、 04 ： 通年、 05 ： 通常 11 ： 春、 12 ： 夏、 13 ： 秋、 14 ： 冬 21:年末、 22:年始、 23:年末年始、 24:正月、 25:お盆、 26:ｺﾞｰﾙﾃﾞﾝｳｨｰｸ 31:春休み、 32:夏休み、 33:冬休み
05 ： 週	01～07 ： 毎週日曜～毎週土曜 11～17 ： 第1日曜～第1土曜 21～27 ： 第2日曜～第2土曜 31～37 ： 第3日曜～第3土曜 41～47 ： 第4日曜～第4土曜 51～57 ： 第5日曜～第5土曜 61～67 ： 隔週日曜～隔週土曜 71～77 ： 最終日曜～最終土曜
06 ： 曜日	01～07 ： 日曜～土曜、 08 ： 月～金、 09 ： 日･祝祭日、 10 ： 祝祭日
07 ： 自然条件	01 ： 雨天、 02 ： 濃霧、 03 ： 積雪、 04 ： 降雪、 05 ： 日の出、 06 ： 日没
08 ： 限定期間	01 ： 期間条件、 02 ： 開業期間中(開催･開園･開校含む)、 03 ： 不定
09 ： 休日	01 ： 年中無休

変換例 ：0602-0606:0800-1700

変換表：06 は曜日データ、02 は月曜、06 は金曜

データ変換後：

0602-0606:0800-1700

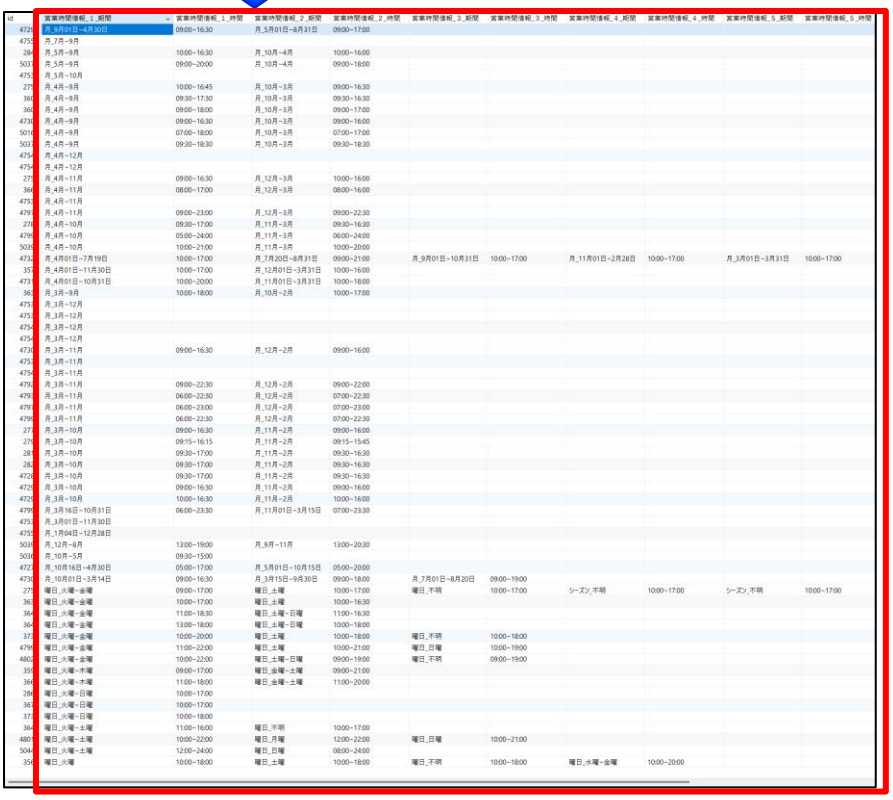
⇒曜日_月曜～金曜 0800-1700

値 Sample

月_3 月~10 月	09:30~17:00
曜日_土曜~日曜	11:00~16:30
週_毎週月曜	
月_4 月 01 日~7 月 19 日	10:00~17:00

ⓒ東京都

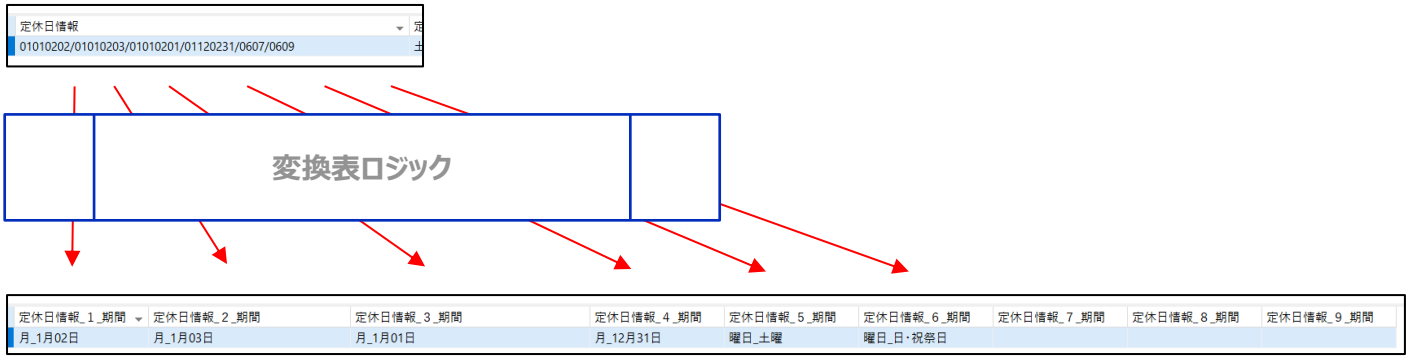
システム変換処理実施前後のデータイメージ



ⓒ東京都

定休日情報

値展開ロジック イメージ



・スラッシュ区切りで、1 項目に複数値が設定さている。

定休日情報 1/定休日情報 2/…定休日情報 9

※定休日は 0 個～9 個と格納数は様々（実データより）。

・変換表を参照しないと、意味のわかるデータとならない。

値フォーマット ：「大区分_小区分(-大区分_小区分)」のデータフォーマットとなっている。

変換表

大区分	小区分
00 ： 不明	00 ： 不明
01 ： 月	01～12 ： 1月～12月、 21 ： 毎月
02 ： 日	01～31 ： 1日～31日
03 ： 旬	01 ： 初(上)旬、 02 ： 中旬、 03 ： 下旬
04 ： ｼｰｽﾞﾝ	01 ： ｵｰﾙｼｰｽﾞﾝ、 02 ： ｵﾌｼｰｽﾞﾝ、 03 ： ｵﾝｼｰｽﾞﾝ、 04 ： 通年、 05 ： 通常 11 ： 春、 12 ： 夏、 13 ： 秋、 14 ： 冬 21:年末、 22:年始、 23:年末年始、 24:正月、 25:お盆、 26:ｺﾞｰﾙﾃﾞﾝｳｲｰｸ 31:春休み、 32:夏休み、 33:冬休み
05 ： 週	01～07 ： 毎週日曜～毎週土曜 11～17 ： 第1日曜～第1土曜 21～27 ： 第2日曜～第2土曜 31～37 ： 第3日曜～第3土曜 41～47 ： 第4日曜～第4土曜 51～57 ： 第5日曜～第5土曜 61～67 ： 隔週日曜～隔週土曜 71～77 ： 最終日曜～最終土曜
06 ： 曜日	01～07 ： 日曜～土曜、 08 ： 月～金、 09 ： 日・祝祭日、 10 ： 祝祭日
07 ： 自然条件	01 ： 雨天、 02 ： 濃霧、 03 ： 積雪、 04 ： 降雪、 05 ： 日の出、 06 ： 日没
08 ： 限定期間	01 ： 期間条件、 02 ： 開業期間中(開催・開園・開校含む)、 03 ： 不定
09 ： 休日	01 ： 年中無休

変換例 ：0404

変換表：04 はシーズンデータ、04 は通年

データ変換後：

0404

⇒シーズン_通年

値 Sample

シーズン_年末年始

月_12 月 29 日~1 月 03 日

休日_年中無休

曜日_日・祝祭日

ⓒ東京都

システム変換処理実施前後のデータイメージ

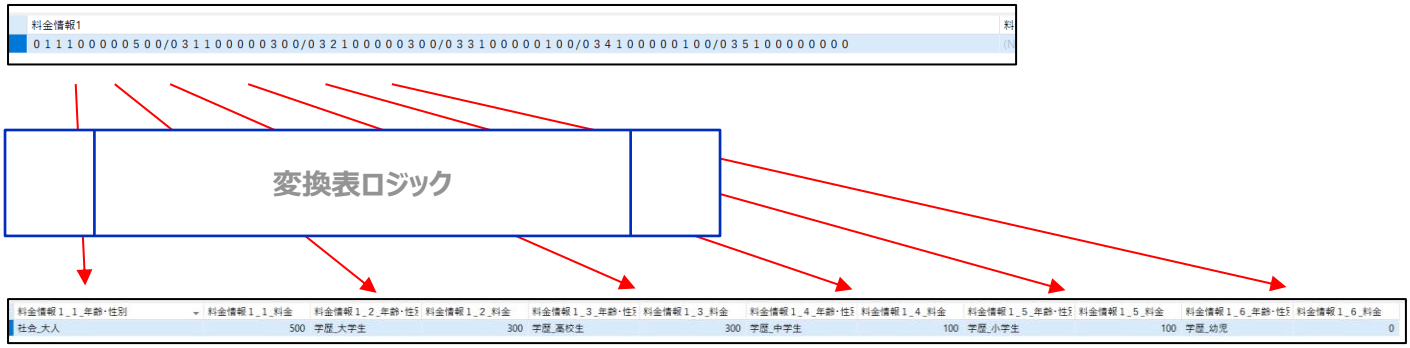
定休日情報
01120230-01010204/01040215/0609
01120229-01010203/0609
01120229-01010203/0607/0609
01120229-01010203
01120229-01010203
01120229-01010203
01120225-01010203/0506/0502
01110215/0423/0609
01110215/0423/0609
01110215/0423/0609
01110215/01120229-01010203
01080211-01080213/01120222-01010204
01060521/0423
01060511/0423
01050201/01110204
01050201/01100215/0423/0609
01040215/0423/0609
0102:0542/0102:0543/01120231/01010201
01020212/0527/0423/0609
01010544/01010554/01020504
01010202/01010203/01010201/01120231/0607/0609
01010202
01010201/01120231/0503
01010201/01120231
01010201/01020531
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202/01010203
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202
01010201/01010202

定休日情報_1_期間	定休日情報_2_期間	定休日情報_3_期間	定休日情報_4_期間	定休日情報_5_期間	定休日情報_6_期間	定休日情報_7_期間	定休日情報_8_期間	定休日情報_9_期間
曜日_祝祭日	週_第5日曜							
週_毎週月曜	週_第4火曜	シーズン_年末年始						
週_第2火曜	週_第4火曜							
週_毎週月曜	週_第4火曜							
週_第2火曜	週_第4火曜							
週_第2火曜	週_第4火曜							
週_第2水曜	週_第4水曜							
週_第2木曜	週_第4木曜							
週_第2木曜	週_第4木曜	曜日_日・祝祭日						
週_第2月曜	週_第4月曜	シーズン_年末年始						
週_第2月曜	週_第4月曜	月_12月29日~1月03日						
週_第2月曜	週_第4月曜	月_12月29日	月_12月30日	月_12月31日	月_1月01日	月_1月02日	月_1月03日	
週_第2月曜	週_第4月曜							
週_第2日曜	週_第4日曜							
週_第2土曜	週_第4土曜	曜日_日・祝祭日	シーズン_年末年始					
週_第2土曜	週_第4土曜	曜日_日・祝祭日	月_12月29日~1月03日	月_6月				
曜日_日・祝祭日	週_第4土曜	月_11月05日	シーズン_年末年始					
週_第2土曜	週_第4土曜	曜日_日・祝祭日	月_11月01日	月_12月29日~1月03日				
週_第2土曜	週_第4土曜	曜日_日・祝祭日	月_4月13日	月_4月14日	月_12月29日~1月03日			
週_第2土曜	週_第4土曜	曜日_日・祝祭日						
週_第2土曜	週_第4土曜	曜日_日・祝祭日	月_12月29日~1月03日					
週_第2土曜	週_第4土曜	曜日_日・祝祭日	月_12月30日~1月03日					
週_第2土曜	週_第4土曜	曜日_日・祝祭日						
週_第2土曜	週_第4土曜	曜日_日・祝祭日						
週_第2土曜	週_第4土曜	曜日_日・祝祭日						
週_第2土曜	週_第4土曜	曜日_日・祝祭日						
週_第2土曜	週_第4土曜	週_第5土曜	曜日_日曜					
週_第2土曜	週_第4土曜	曜日_日・祝祭日						
週_第2土曜	週_第4土曜	月_4月	シーズン_年末年始	曜日_日・祝祭日				
週_毎週月曜	週_第3木曜	月_12月28日~1月04日						
週_第1月曜	週_第3月曜							
週_第1月曜	週_第3月曜							
週_第1月曜	週_第3月曜	月_12月29日~1月03日						
週_毎週月曜	週_第3日曜							
週_毎週月曜	週_第3日曜	曜日_祝祭日	シーズン_年末年始					
週_第1日曜	週_第3日曜							
週_第1土曜	週_第3土曜	月_1月10日	月_4月23日	シーズン_年末年始	曜日_日・祝祭日			
週_第1土曜	週_第3土曜	週_第5土曜	曜日_日・祝祭日	月_12月29日~1月03日				
週_第1土曜	週_第3土曜	週_第5土曜	曜日_日・祝祭日	月_12月30日~1月03日				
曜日_日・祝祭日	週_第3土曜	月_12月25日	シーズン_年末年始					
週_第1土曜	週_第3土曜	曜日_日・祝祭日	月_12月30日~1月03日					
週_第1土曜	週_第3土曜	週_第5土曜	曜日_日・祝祭日	シーズン_年末年始				
週_第1土曜	週_第3土曜	週_第5土曜	曜日_日・祝祭日					
曜日_水曜	週_第3土曜	曜日_日・祝祭日						
週_第1土曜	週_第3土曜	週_第5土曜	曜日_日・祝祭日					
週_第1土曜	週_第3土曜	月_11月01日	シーズン_年末年始	曜日_日・祝祭日				
週_第1土曜	週_第3土曜	週_第5土曜	曜日_日・祝祭日	月_12月29日~1月03日				
曜日_月曜	週_第2火曜	週_第3火曜						
週_毎週月曜	週_第2木曜							
月_2月12日	週_第2土曜	シーズン_年末年始	曜日_日・祝祭日					
曜日_日・祝祭日	週_第2土曜	週_第3土曜	週_第5土曜	シーズン_年末年始				
週_毎週月曜	週_毎週金曜							
週_毎週月曜	週_毎週金曜	シーズン_年末年始						
曜日_月曜~火曜	週_毎週金曜							
月_12月25日~1月03日	週_毎週金曜	週_毎週月曜						
週_毎週月曜	週_毎週火曜							

ⓒ東京都

料金情報

値展開ロジック イメージ



・スラッシュ区切りで、1 項目に複数値が設定さている。

料金情報 1/料金情報 2 /…料金情報 6

※料金情報は 0 個 6 個と格納数は様々（実データより）。

・変換表を参照しないと、意味のわかるデータとならない。

値フォーマット ：「分類_年齢/性別_金額」のデータフォーマットとなっている。

変換表

分類コード	年齢／性別コード
01 ： 社会	11 ： 大人、 12 ： 大人男性、 13 ： 大人女性、 14 ： 大人会員 21 ： 中人、 22 ： 中人男性、 23 ： 中人女性、 24 ： 中人会員 31 ： 小人、 32 ： 小人男性、 33 ： 小人女性、 34 ： 小人会員 61 ： 社会人、 62 ： 社会人男性、 63 ： 社会人女性、 64 ： 社会人会員 71 ： 学生、 72 ： 学生男性、 73 ： 学生女性、 74 ： 学生会員
02 ： 性別	21 ： 男性 31 ： 女性
03 ： 学歴	11 ： 大学生、 12 ： 大学生男子、 13 ： 大学生女子、 14 ： 大学生会員 21 ： 高校生、 22 ： 高校生男子、 23 ： 高校生女子、 24 ： 高校生会員 26 ： 高校生以上、 27 ： 高校生以下 31 ： 中学生、 32 ： 中学生男子、 33 ： 中学生女子、 34 ： 中学生会員 36 ： 中学生以上、 37 ： 中学生以下 41 ： 小学生、 42 ： 小学生男子、 43 ： 小学生女子、 44 ： 小学生会員 46 ： 小学生以上、 47 ： 小学生以下 51 ： 幼児、 52 ： 幼児男子、 53 ： 幼児女子、 54 ： 幼児会員 56 ： 幼児以上、 57 ： 幼児以下
04 ： 年齢	16 ： 3歳以上、 17 ： 3歳以下、 18 ： 3歳未満 26 ： 4歳以上、 27 ： 4歳以下、 28 ： 4歳未満 36 ： 5歳以上、 37 ： 5歳以下、 38 ： 5歳未満 46 ： 15歳以上、 47 ： 15歳以下、 48 ： 15歳未満 56 ： 18歳以上、 57 ： 18歳以下、 58 ： 18歳未満 66 ： 20歳以上、 67 ： 20歳以下、 68 ： 20歳未満 76 ： 60歳以上、 77 ： 60歳以下、 78 ： 60歳未満 86 ： 65歳以上、 87 ： 65歳以下、 88 ： 65歳未満 96 ： 6歳以上、 97 ： 6歳以下、 98 ： 6歳未満
05 ： 特約	31 ： 一般、 32 ： 一般男性、 33 ： 一般女性 41 ： 会員、 42 ： 会員男子、 43 ： 会員女性
06 ： 組み合わせ	11:大人(平日) 12:大人(日・祝祭日) 21:小人(平日) 22:小人(日・祝祭日)

変換例 ：011100001000

変換表：01 は社会、11 は大人、00001000 は 1000 円

データ変換後：

011100001000

⇒社会_大人　1000 円

値 Sample

特約_一般	1000
学歴_高校生	500
社会_大人	200
学歴_中学生以下	200

ⓒ東京都

システム変換処理実施前後のデータイメージ

料金情報1
053100002800
053100002400
053100002300
053100002100
053100002000
053100001500/031100001000/032100001000/033700000500
053100001500
053100001300/031100000800/032100000500/033100000500
053100001030/031100000730/033700000420
053100001000/034600000600
053100001000/034100000600
053100001000/032700000800
053100001000/032100000500/045800000500
053100001000/031100000800/032100000500/033100000500
053100001000/031100000700/033700000500
053100001000/031100000700/033700000400
053100001000/031100000700/032100000700
053100001000/031100000700
053100001000/031100000500/032100000500/033100000200/034100000200
053100001000/031100000500/032100000500
053100001000/017100000800
053100001000
053100000880/048600000780/031100000680/032100000580/033100000580/034100000280
053100000800/034700000500
053100000800/033100000500/034100000500
053100000800/032100000500/034100000300
053100000800/031100000800/032100000500/033100000500
053100000800/031100000600/034100000400
053100000800/031100000500/033700000200
053100000800/031100000500/032100000500/033100000300/034100000300
053100000800/031100000500/032100000500
053100000800/017100000700/013100000400
053100000800/017100000700/013100000400
053100000800/017100000500
053100000800
053100000700/033100000250/048600000350
053100000700/031100000350/032100000350/033100000350/034100000350
053100000600/033100000200/048600000300
053100000600/033100000200/048600000300
053100000600/032100000400/033100000400/042600000250
053100000600/031100000600
053100000600/031100000500/033700000400
053100000600/031100000500/032100000500/033100000200/034100000200
053100000600/031100000480/032100000300/033100000300
053100000600/031100000400/033700000300
053100000600/031100000400
053100000600/017100000300
053100000600
053100000600
053100000525/017100000420/033100000210/034100000200
053100000525
053100000500/048600000400/048600000250/032700000250
053100000500/048600000250/017100000400/032100000250
053100000500/034100000250/033100000250/032100000250
053100000500/033100000300/034100000300/035100000000
053100000500/033100000250/034100000250
053100000500/032100000400/034100000200
053100000500/032100000300/033100000300/034100000300
053100000500/032100000250/033100000250/034100000250

料金情報１_１_年齢・性別	料金情報１_１_料金	料金情報１_２_年齢・性別	料金情報１_２_料金	料金情報１_３_年齢・性別	料金情報１_３_料金	料金情報１_４_年齢・性別	料金情報１_４_料金	料金情報１_５_年齢・性別	料金情報１_５_料金	料金情報１_６_年齢・性別	料金情報１_６_料金
社会_小人	500	社会_大人	700		(Null)		(Null)		(Null)		(Null)
社会_大人	200	学歴_中学生以下	100		(Null)		(Null)		(Null)		(Null)
社会_大人	3800	社会_小人	1900		(Null)		(Null)		(Null)		(Null)
社会_大人	2250		(Null)		(Null)		(Null)		(Null)		(Null)
社会_大人	200	学歴_中学生	100	学歴_小学生	100		(Null)		(Null)		(Null)
社会_大人	300		(Null)		(Null)		(Null)		(Null)		(Null)
社会_大人	2550	社会_小人	1500		(Null)		(Null)		(Null)		(Null)
社会_大人	400	社会_小人	200		(Null)		(Null)		(Null)		(Null)
社会_大人	300	学歴_中学生	200	学歴_小学生	200		(Null)		(Null)		(Null)
社会_大人	500	社会_学生	300		(Null)		(Null)		(Null)		(Null)
社会_大人	550	学歴_小学生以下	200		(Null)		(Null)		(Null)		(Null)
社会_大人	700	学歴_高校生	300	学歴_小学生	200		(Null)		(Null)		(Null)
社会_大人	1000	社会_小人	500		(Null)		(Null)		(Null)		(Null)
社会_大人	1500	社会_小人	1000		(Null)		(Null)		(Null)		(Null)
社会_大人	520	社会_学生	300		(Null)		(Null)		(Null)		(Null)
社会_大人	1000	学歴_大学生	800	学歴_高校生	800		(Null)		(Null)		(Null)
社会_大人	300	学歴_中学生	100	学歴_小学生	100		(Null)		(Null)		(Null)
社会_大人	300	学歴_大学生	200	学歴_高校生	200	学歴_中学生	200	学歴_小学生	200		(Null)
社会_大人	100	学歴_大学生	100	学歴_高校生	50	学歴_中学生	50	学歴_小学生	50		(Null)
社会_大人	1000	学歴_大学生	500	学歴_高校生	500		(Null)		(Null)		(Null)
社会_大人	300	学歴_高校生	50		(Null)		(Null)		(Null)		(Null)
社会_大人	300		(Null)		(Null)		(Null)		(Null)		(Null)
社会_大人	400	学歴_中学生以下	200		(Null)		(Null)		(Null)		(Null)
社会_大人	200		(Null)		(Null)		(Null)		(Null)		(Null)
社会_大人	600	社会_小人	300		(Null)		(Null)		(Null)		(Null)
社会_大人	800		(Null)		(Null)		(Null)		(Null)		(Null)
社会_大人	1260	年齢_3歳以上	735		(Null)		(Null)		(Null)		(Null)
社会_大人	3000	社会_小人	700		(Null)		(Null)		(Null)		(Null)
社会_大人	3150		(Null)		(Null)		(Null)		(Null)		(Null)
社会_大人	1800	社会_小人	1000		(Null)		(Null)		(Null)		(Null)
社会_大人	840	社会_小人	420		(Null)		(Null)		(Null)		(Null)
社会_大人	800	学歴_小学生	400		(Null)		(Null)		(Null)		(Null)
社会_大人	2000	社会_小人	1500		(Null)		(Null)		(Null)		(Null)
社会_大人	200	社会_小人	100		(Null)		(Null)		(Null)		(Null)
社会_大人	500	学歴_高校生	200	学歴_中学生	200	学歴_小学生	200		(Null)		(Null)
社会_大人	200	社会_小人	50		(Null)		(Null)		(Null)		(Null)
社会_大人	500	学歴_大学生	300	学歴_高校生	300	学歴_中学生	300	学歴_小学生以下	0		(Null)
社会_大人	500	学歴_大学生	300	学歴_高校生	300	学歴_中学生	100	学歴_小学生	100	学歴_幼児	0
社会_大人	1000	学歴_大学生	1000	学歴_高校生	700	学歴_中学生	700	学歴_小学生	400	学歴_幼児	100
社会_大人	900	学歴_大学生	700	学歴_高校生	700	学歴_中学生	400	学歴_小学生	400		(Null)
社会_大人	700	学歴_中学生以下	250		(Null)		(Null)		(Null)		(Null)
社会_大人	300	学歴_高校生	150	学歴_中学生	150	学歴_小学生	150		(Null)		(Null)
社会_大人	100		(Null)		(Null)		(Null)		(Null)		(Null)
社会_大人	500	学歴_中学生	200	年齢_65歳以上	250		(Null)		(Null)		(Null)
社会_大人	250	学歴_中学生	100		(Null)		(Null)		(Null)		(Null)
社会_大人	500	年齢_65歳以上	250	学歴_高校生以上	250		(Null)		(Null)		(Null)
社会_大人	2000	学歴_高校生	1500	学歴_中学生	1000	学歴_小学生	1000	年齢_3歳以上	600		(Null)
社会_大人	385		(Null)		(Null)		(Null)		(Null)		(Null)
社会_大人	385		(Null)		(Null)		(Null)		(Null)		(Null)
社会_大人	1000	学歴_小学生以下	500		(Null)		(Null)		(Null)		(Null)
社会_大人	385		(Null)		(Null)		(Null)		(Null)		(Null)

ⓒ東京都

複雑なデータ補完

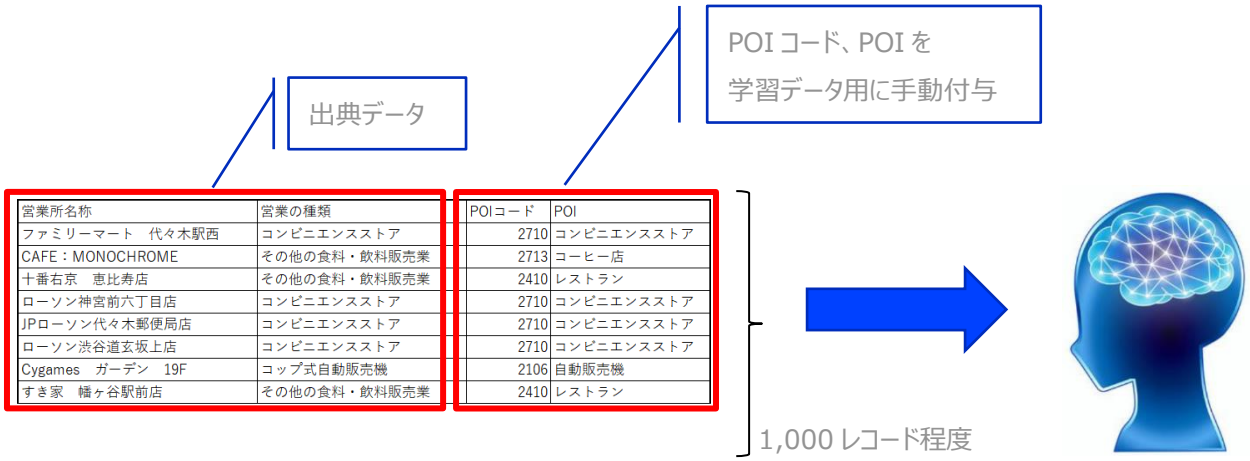
POI コード、緯度経度、法人番号付与は委託業者によるシステム付与を想定しています。

POI コード

名称、分類を元に AI を用いて POI コードの推測補完を行います。

事前学習(1,000 件程度で充分な精度確保が可能)

施設名称、分類に対して、正解と思われる POI コード、POI を付与し学習データとします。



営業所名称	営業の種類	POIコード	POI
ファミリーマート　代々木駅西	コンビニエンスストア	2710	コンビニエンスストア
CAFE：MONOCHROME	その他の食料・飲料販売業	2713	コーヒー店
十番右京　恵比寿店	その他の食料・飲料販売業	2410	レストラン
ローソン神宮前六丁目店	コンビニエンスストア	2710	コンビニエンスストア
JPローソン代々木郵便局店	コンビニエンスストア	2710	コンビニエンスストア
ローソン渋谷道玄坂上店	コンビニエンスストア	2710	コンビニエンスストア
Cygames　ガーデン　19F	コップ式自動販売機	2106	自動販売機
すき家　幡ヶ谷駅前店	その他の食料・飲料販売業	2410	レストラン

学習済モデルを用いて、POI コード機械付与

例　：

処理対象データ：株式会社高島屋新宿店 11 開催会場　大北海道展

大分類：飲食

```
JSON
{
  処理対象データ : "株式会社高島屋新宿店11開催会場　大北海道展"
  大分類 : "飲食"
  POI : "レストラン"
  POIコード : 2401
}
```

Json　値の処理結果を元に、POI、POI コードの補完を実施します。

学習済モデルに、処理対象データ全件を入力すると、POI、POI コードが補完されます。

処理対象データ	大分類	POIコード	POI
株式会社高島屋新宿店11開催会場　大北海道展	飲食	2401	レストラン

【手作業で POI コードを調べる場合の方法について】

① 政府相互運用性フレームワーク(GIF) 490 その他 491-3-1_POI コード表(xlsx) （ https://github.com/JDA-DM/GIF/raw/main/490_その他/491-3-1_POI コード表.xlsx)から POI コード表をダウンロード

② 大分類、POI などを絞り込み、POI コードの自動付与を実施する。

ⓒ東京都

緯度経度

住所情報を元に緯度経度情報の推測補完を行います。
株式会社自動処理社では住所の表記ゆれの訂正と共に、枝番レベルまで解析が出来た住所については、緯度経度情報を取得する事が出来ます。緯度経度情報については人間にはわかりませんが、GIS（地理情報システム）と呼ばれる地図システムにて利用する事が出来ます。緯度経度情報がある事で、地図を明確に指定することが出来ます。住所の項目にて住所情報の解析が難しい旨記載しましたが、緯度経度を利用する事によって、場所を明確に示す事が出来ます。

緯度経度のライセンスについて
緯度経度情報については、緯度経度情報は、地理的な位置を特定するために広く利用されています。これらの情報は、地図やナビゲーションアプリ、位置情報サービスなど、多くの産業で重要な役割を果たしています。オープンデータ政策が普及するにつれ、政府や自治体の中で、緯度経度情報を付与してオープンデータ化する事が増えていますが、ライセンスを確認しないまま、オープンデータとして公開する事で、ライセンス違反になっているケースがあります。
今回の作業では株式会社自動処理社がライセンスを保有している緯度経度を利用していますが、自治体の皆様がライセンスを違反しないように緯度経度を付与する際には、ライセンスを購入するか、オープンストリートマップを利用して、緯度経度を取得するか、下記の地理院マップシートを利用するかのどれかになると思います。特に Google Map を利用して緯度経度を取得するライセンス違反がよく見られる為、気を付けてください。
今回の作業の中では株式会社自動処理がライセンス許諾を受けた緯度経度情報をご提供している為問題が発生しておりません。

表記揺れ統一後の正規化データ

営業所所在地	緯度	経度
東京都台東区秋葉原1番5	35.7020182292	139.7752284071
東京都台東区秋葉原3番7	35.7022222222	139.7748787435
東京都台東区秋葉原3番1	35.7020610894	139.7750165473
東京都台東区秋葉原3番1	35.7020610894	139.7750165473
東京都台東区浅草1丁目1	35.7109556749	139.7974820964
東京都台東区浅草1丁目1	35.7109556749	139.7974820964
東京都台東区浅草1丁目1	35.7109556749	139.7974820964
東京都台東区浅草1丁目1	35.7109754774	139.7973019748
東京都台東区浅草1丁目1	35.7109754774	139.7973019748
東京都台東区浅草1丁目1	35.7110061306	139.7971964518
東京都台東区浅草1丁目1	35.7111564128	139.7971199544
東京都台東区浅草1丁目1	35.7112689887	139.7971397569
東京都台東区浅草1丁目1	35.7113864475	139.7971411133
東京都台東区浅草1丁目1	35.7113864475	139.7971411133
東京都台東区浅草1丁目1	35.7115386285	139.7971568468
東京都台東区浅草1丁目1	35.7116238064	139.7973258464
東京都台東区浅草1丁目1	35.7116238064	139.7973258464
東京都台東区浅草1丁目1	35.7116080729	139.7974948459

【手作業で緯度経度を調べる場合の方法について】

① 地理院マップシートダウンロードページ(https://renkei2.gsi.go.jp/renkei/130326mapsh_gijutu/index.html)にアクセス
② 「地理院マップシート（Ver.2.0.0.4）」ダウンロードファイルをダウンロードする
③ ご利用上の注意に留意しながら、Excel に住所を入力し、緯度経度の調査を行う

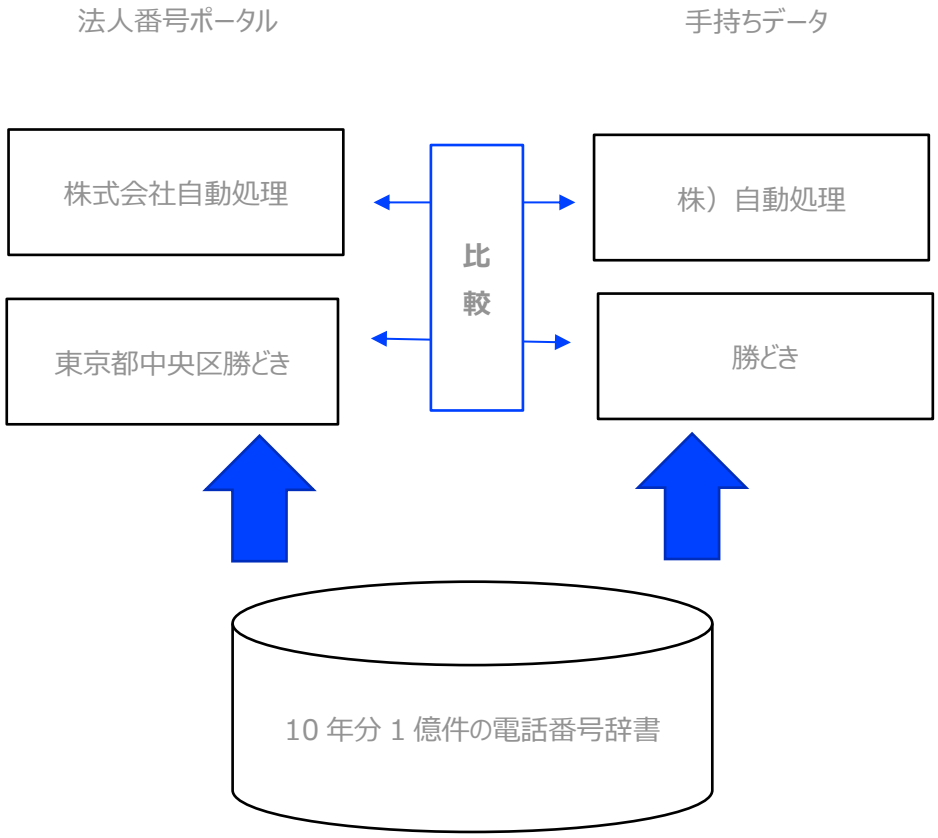
※地理院マップシートについては、国土地理院の緯度経度の更新頻度が高くない事や住所情報のカバー範囲が広くない事から、利用ケースによっては緊急性の高い情報（例えば AED 設置箇所一覧や医療機関一覧、消防水利施設一覧など）の場合は、データの更新頻度が高く、網羅性の高い地図データを保有する事業者に緯度経度付与を依頼する事を推奨とする。

ⓒ東京都

法人番号付与

事業所の名称、住所、電話番号情報を元に法人番号の補完を行います。
法人利用の場合、手持ちの情報に法人情報が付与されていると、顧客管理や取引先管理などの利用に非常に便利になりますが、法人番号に保存されている情報は法人名称と住所しかない為、この 2 つの情報を元に名寄せを行う事に法人番号を検索する事になります。法人番号を調べる為には、名称と住所をクレンジングし、検索しやすくした上で、検索処理を行います。株式会社自動処理では、それに加えて、過去 10 年分のハローページ情報から情報を強化し、より法人番号を誤りなく正確に付与できるように工夫をしています。
法人番号を付与する際には、様々な手法がありますが、誤った番号を付与してしまう可能性がある事を念頭に注意しながら、作業を行ってください。法人番号の付与率と、誤り率を両方とも確認しながら作業を行うようにしてください。
法人番号付与は以下の手順で行います。

① データの正規化
　(ア) 事業所の名称のクレンジング　法人名やサービス名の混ざる事業所名から法人名を抜き出し、文字列を変換します。
　(イ) 住所のクレンジング　住所情報を統一的なデータとしてクレンジングします。
　(ウ) 電話番号のクレンジング　電話番号情報を統一的なデータとしてクレンジングします。
② 国税庁法人番号公表サイト
　(ア) 法人の名称のクレンジング　法人名の文字列を変換します。
　(イ) 住所のクレンジング　住所情報を統一的なデータとしてクレンジングします。
③ データの補完
　(ア) 法人番号ポータルのデータと、法人番号付与対象のデータに対して、ハローページ情報を元にしてデータの強化します
④ 法人番号検索
　(ア) 法人番号検索を行います。



ⓒ東京都

【手作業で法人番号を調べる場合の方法について】

① 法人番号公表サイト(https://www.houjin-bangou.nta.go.jp/)にアクセス

② 名称所在地などから調べるフォームを利用して検索を行う

※基本３情報ダウンロードから全件ダウンロードする事は出来るが、かなりのデータ量になる為、単純に検索するよりは、ウェブフォームから検索する事をお勧めします。複数企業が見つかる場合や、検索しても見つからない場合は住所を絞り込むか、部分一致検索にて検索すると絞り込める場合があります。

【複雑なデータ補完における委託事業者へ依頼時の仕様イメージ】

項目	内容	備考
作業の目的・概要	POI コード・緯度経度・法人番号を新たに付与し、データ項目を増やす	
アウトプット	POI コード・緯度経度・法人番号が付与されたデータベース	・ データの具体的な処理については、データの状態により正規化出来るかどうかや、整備目標が変わります。 ・ 調達する際に『名称の値が存在するデータに対して、90%以上の整備が出来る事を目標とする』など目標値を設定する事で、調達要件とすることが出来ます。 ・ 妥当な数値目標を設定するには、整備難度に応じた網羅的なケースを用意し、調達前にあらかじめ提示した上で、目標値を達成できるか提案してもらう等の方法を取ることが出来ます。
作業内容の想定例	・ 付与対象データの準備 ➢ POI コード：分類項目及びコードの取得 ➢ 緯度経度：ライセンス取得・購入 ➢ 法人番号：国税庁法人番号サイトより取得 ・ 付与対象データと元データの対応関係整理 ➢ POI コード：AI による推測補完 （事前学習作業：施設名称、分類に対応する POI、POI コードが格納された教師データの用意、AI モデルの学習） ・ データ付与の実施	・ 緯度経度情報について、ライセンス上問題がないかを事前に確認し、必要に応じてライセンス取得・購入を行います。

ⓒ東京都

手順 4

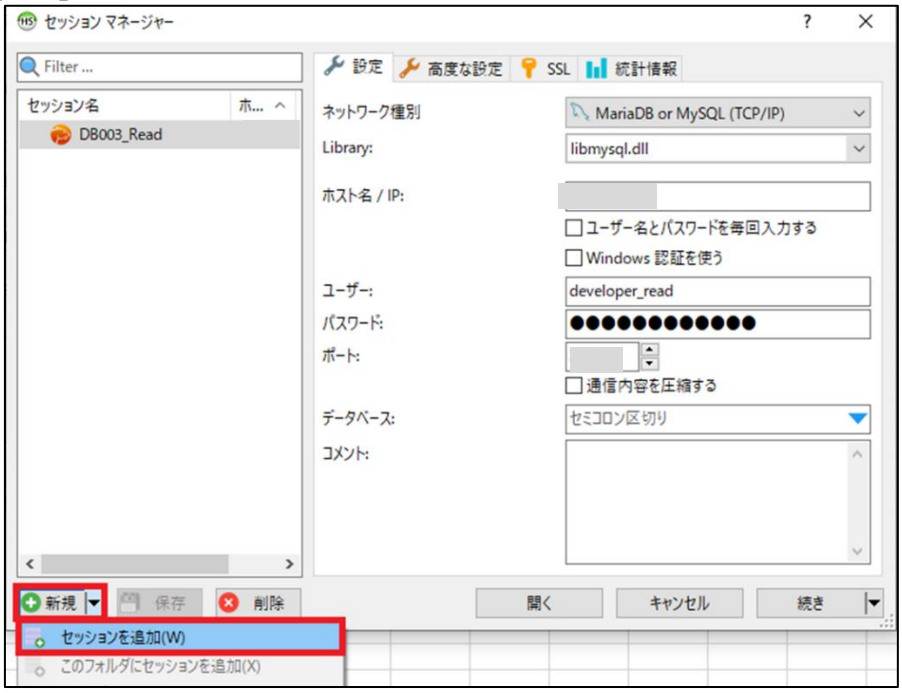
データの保存（エクスポート）

データベース接続の設定

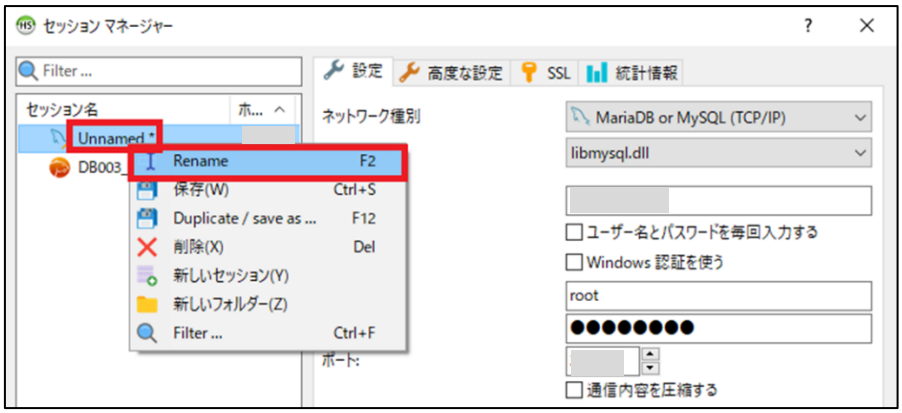
HeidiSQL を用いて、データベースに接続する設定を行います。
任意のデータベース接続アプリケーションを使用される場合は、説明文をお使いのアプリケーションに適宜読み替えてご参照ください。

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**

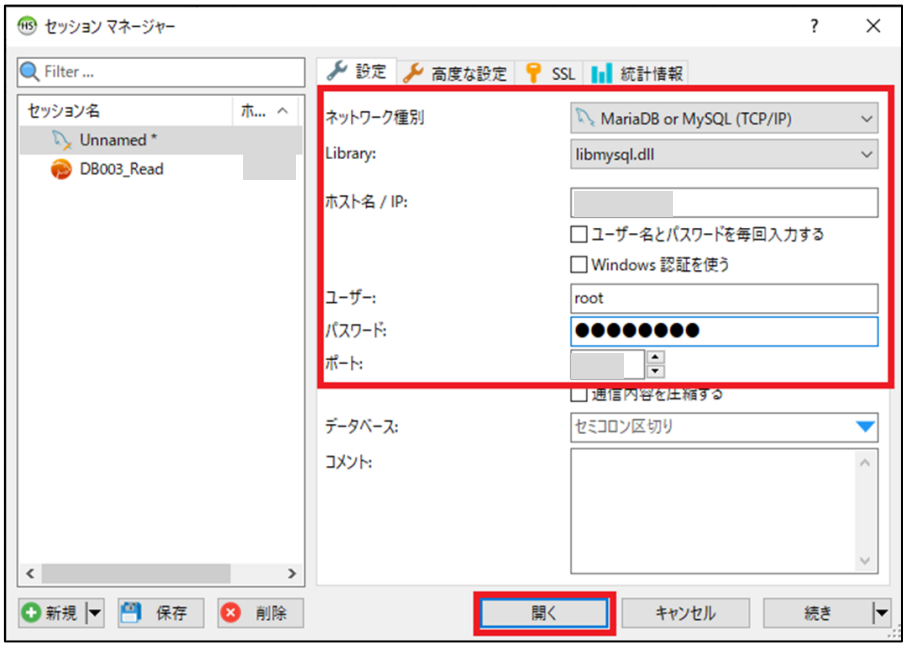


3. **セッション名を右クリック、「Rename」をクリックした後、任意の名前に変更する**



ⓒ東京都

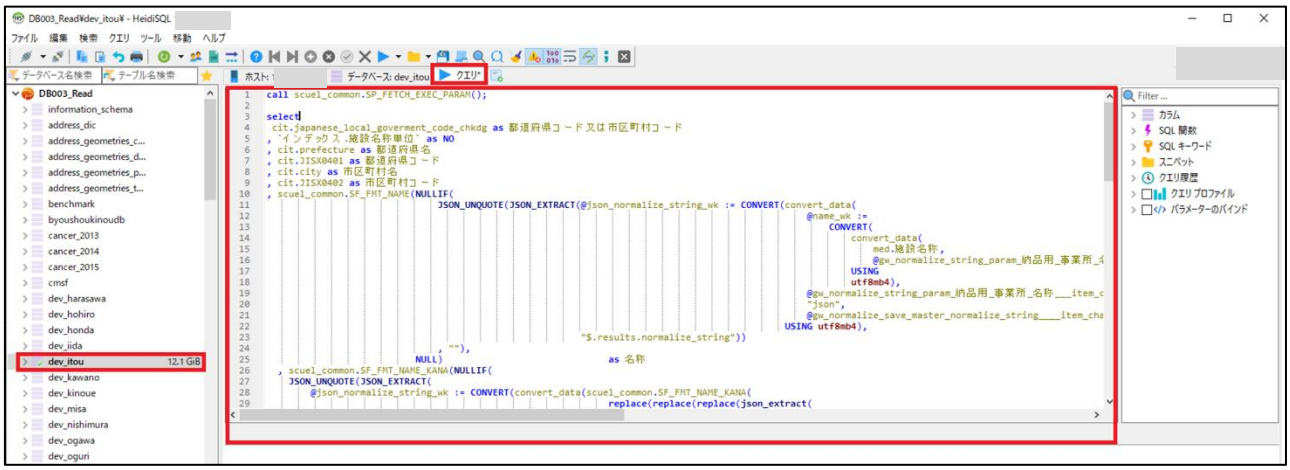
4. 「設定」タブをクリック、設定を変更した後、[開く]をクリックする



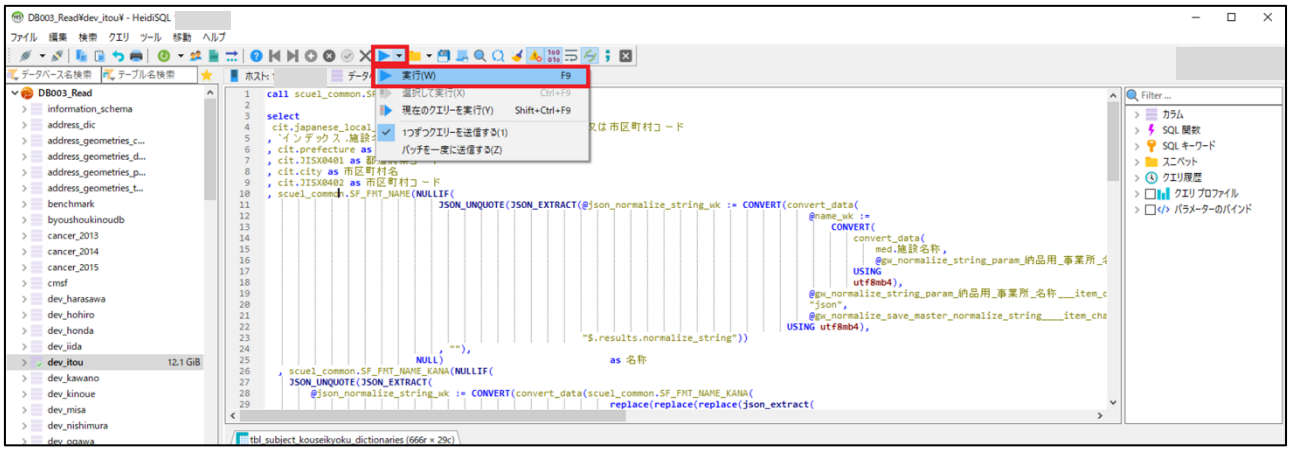
ネットワーク種別：MariaDB or MySQL(TCP/IP)を選択
Library: libmysql.dll を選択
ホスト名/IP:
接続先 DB のホスト名/IP を入力
ユーザー：
接続先 DB のユーザーを入力
パスワード：
接続先 DB のパスワードを入力
ポート：
接続先 DB のポートを入力

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

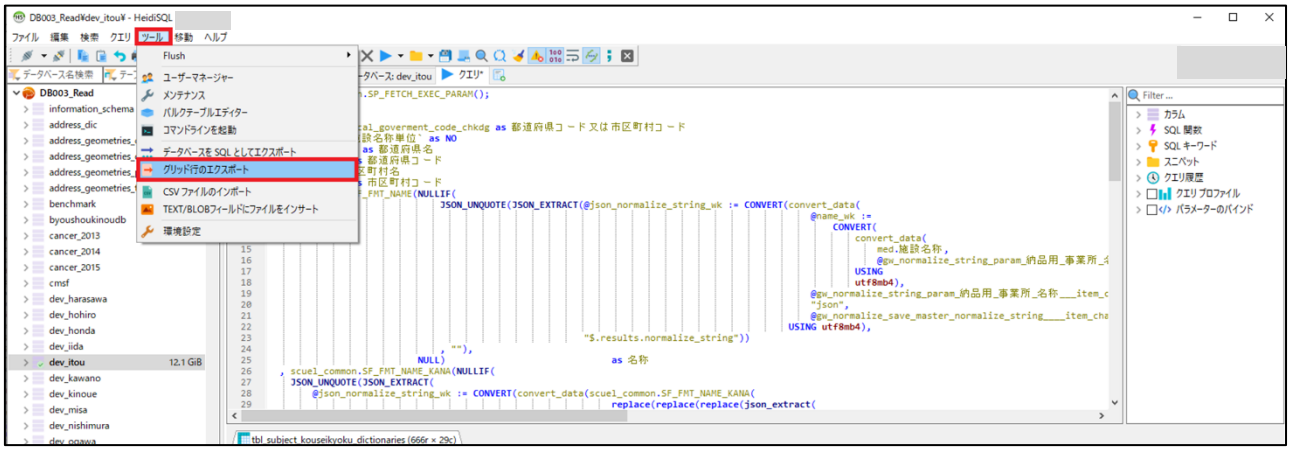


2. 「▶」→「実行」をクリックして、クエリを実行する

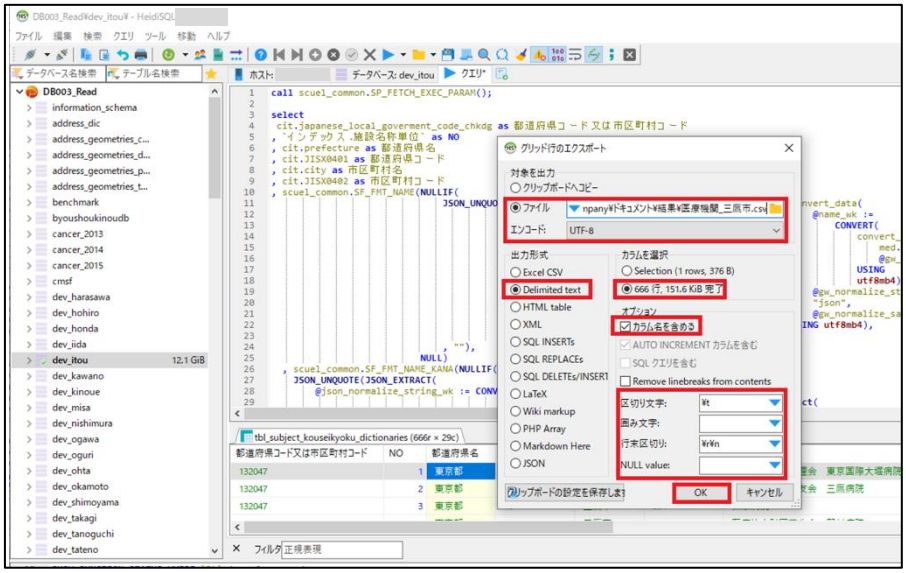


ⓒ東京都

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

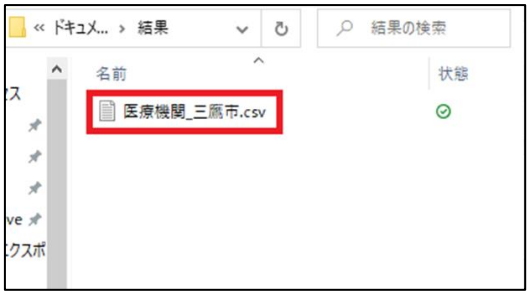


4. 「対象を出力」で任意の出力先を指定した後、設定を変更し、[OK]をクリックする



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）

ⓒ東京都

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）

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」

ⓒ東京都

マッピング

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

東京都データプラットフォーム

データ整備事業

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

2023 年 3 月　V2.0　第 2 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 2023 年 3 月末現在の情報です。
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。

ⓒ東京都