

オープンデータ × 生成AI

OOL Tech Connect vol.7

はじめに



プロフィール

坂本 諒太
Ryota Sakamoto

TIS株式会社
デジタル社会サービス企画部

趣味など
OSS開発

SNS

Twitter
@SkeinIn

Github
@ryo-ma

略歴

- ・ TIS株式会社に入社後、研究開発部門にてクラウド自動化の研究開発およびそれに伴ったOSSの開発に従事
- ・ 大阪大学に常駐し対話/社会心理学の研究に携わり、その知見を活用した高齢者向け生活支援AI対話サービスをスタートアップエンジニアとしてスタートアップ立ち上げに従事
- ・ FIWAREによるIoT/ロボティクスプラットフォームの研究開発に従事し、会津若松にて公道走行配送ロボットの実証実験を担当
- ・ 現在はデータ利活用という観点で社会のあるべき姿を検討し社会実装に向けた活動を実施

オープンデータにおける背景

まず、オープンデータとは？

オープンデータとは

- ・ **オープンデータとは**

- ・ 誰でも自由に使える公開されているデータ (CC-BY、CC-0などのライセンス提供)

- ・ **どこで提供されている？**

- ・ 政府、自治体、民間企業のホームページ、カタログサイトなど

- ・ **例えばどんなデータ？**

- ・ AEDの場所データ、避難所のデータ、町のイベントデータなど

- ・ **どんな使い方？**

- ・ 地域課題の解決、ビジネスへの活用など
- ・ 地図やグラフに表示して見えやすくする、統計・分析につかう など

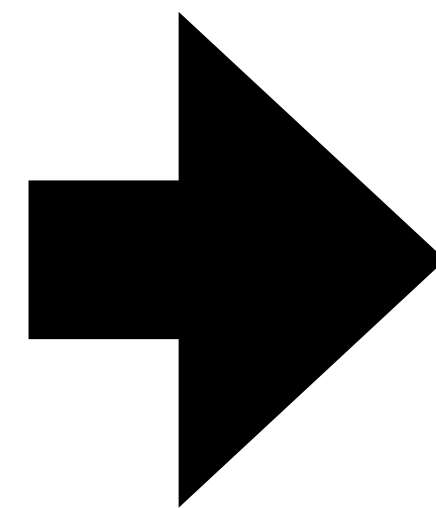
日本のオープンデータはどのような？

OECD Our Data Index 日本の状況

OECD Our Data Index: 4年に一度加盟国(40カ国)のオープンデータの整備状況を調査

2019年

4位



2023年

25位

※ Data Accessibilityという項目が特に順位を下げている原因ですが今回は生成AIメインなので深掘りしません

日本のオープンデータ 改善していくには

- オープンデータ**提供者**は**データを使いやすく公開**する
 - ITの知識がなくてもオープンデータの**作り方を知る**
 - オープンデータを**楽に綺麗なフォーマットで作成**できる
- オープンデータ**利用者**は**生産性高く多くのデータ利活用**をする
 - データを**楽に探せる**
 - 使いにくいデータも**楽に加工、可視化**など利活用できる

生成AIを使って改善できるか？

生成AIを使ったアプローチ

- ・ オープンデータ**提供者**は**データを使いやすく公開**する
⇒ 1. ChatGPTを使ってオープンデータ作成のレビューをする
- ・ オープンデータ**利用者**は**生産性高く多くのデータ利活用**をする
⇒ 2. ChatGPTを使ってデータを可視化する
⇒ 3. GPTsを使ってデータ検索、取得、可視化をする

1. ChatGPTを使ってオープンデータ作成のレビューをする

1. オープンデータのルールを事前の知識として教える



ChatGPT

Code Interpreter

2. オープンデータのルールに基づいて回答

命令：添付の資料を理解したら「はい」と言ってください

項目 番号	項目名	属性	説明	形式	記入例	共通項目属性	得意項目属性	得意項目属性の値型	CD参照/テーブル	GP 番号
1	全道地方公共団体コード	0	情報の管理主体である地方公共団体から「コード」で管理される。※ 記載方法について、「〒」を省略し住所記載事項(ｼｰﾄ02)の欄に「0」と、及び 全道地方公共団体コードを併記する。	文字列 (半角数字)	010002	施設・ｽﾀｼﾞｱｰ発行番・ID・属性値	exist-string		設備・設備番号・全道地方公共団 体コード	0
2	ID	0	情報の管理主体である地方公共団体内で一意に決まる。ID付 決定・起票。※記載方法について、「〒」を省略し住所記載事項(ｼｰﾄ02)の欄 に「0」と併記する。	文字列 (半角英数字)	FF91000222200	設備・ID・属性値	exist-string		設備・ID	0
3	地方公共団体名称	0	情報の管理主体である地方公共団体の名称を記載。	文字列	北海道札幌市	施設・ｽﾀｼﾞｱｰ発行番(半角英数字)・住所・郵便 番号 (5桁)	exist-string		設備・住所・郵便番号	0
4	名称	0	AEO設備場所の建物名称を住所に記載。	文字列	〇〇会館	施設・設備名称・住所・表記	exist-string		設備・名称	0
5	名称_カナ	0	AEO設備場所の建物名称の名称をカナで記載。※記載方法について、 「〒」を省略し住所記載事項(ｼｰﾄ02)の欄に「0」と併記する。	文字列 (全角カナ)	〇〇ﾊﾞﾅﾝ	施設・設備名称・住所・カナ表記	exist-string		設備・名称 (カナ)	0
6	名称_英字	0	AEO設備場所の建物名称の名称を英字で記載。	文字列 (半角英字)	〇〇hall	施設・住所・表記	exist-string		設備・名称 (英字)	0
7	施設番号_全道地方公共団体コード	0	AEO設備場所の所在地管理主体である地方公共団体コード で住所に記載。(5桁)	文字列 (半角数字)	010002	施設・ｽﾀｼﾞｱｰ発行番・ID・属性値	exist-string		施設・施設番号・全道地方公共団 体コード	0
8	施設ID	0	AEO設備場所の建物名称の住所のIDを記載。※変数IDについて、 ﾀﾞｰｸﾏｽｸﾞﾗﾌﾞﾙｽﾀｼﾞｱｰｽﾀｼﾞｱｰIDで住所に記載。	文字列	施設ID	施設・ID・属性値	exist-string		設備・設備番号・変数ID	0
9	施設_通称略称	0	AEO設備場所の住所 (郵便番号)と建物名称等とを通称(と表記)を 記載。	文字列	北海道札幌市厚別区〇〇〇〇E6.00	施設・住所・表記	exist-string		施設・施設住所・通称略称	0
10	施設_郵便番号	0	AEO設備場所の建物名称の住所 (郵便番号)に記載。	文字列	札幌市	施設・住所・表記	exist-string		設備・設備住所・郵便番号	0
11	施設番号_市区町村	0	AEO設備場所の建物名称の住所 (市区町村)に記載。	文字列	札幌市	施設・住所・表記	exist-string		設備・設備住所・市区町村 (県)	0
12	施設_町名	0	AEO設備場所の建物名称の住所 (町名)に記載。	文字列	厚別区2	施設・住所・表記	exist-string		設備・設備住所・町名	0
13	施設_番地以下	0	AEO設備場所の建物名称の住所 (番地以下)に記載。	文字列	〇〇	施設・住所・表記	exist-string		設備・設備住所・番地以下	0

デジタル庁 自治体標準オープンデータセット データ項目定義書（正式版）
https://www.digital.go.jp/resources/open_data/municipal-standard-data-set-test/

命令：先ほど共有したデータ項目定義書に基づいてこの添付ファイルをフィードバックしてください

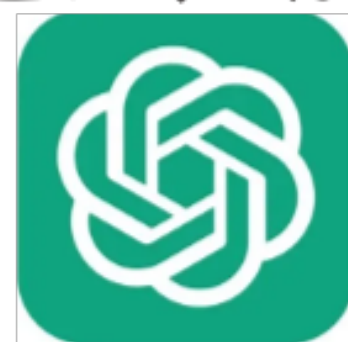
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	施設名称	所在地	緯度	経度	電話番号	URL							
2	両国観光	墨田区両国	35.69547	139.7922	03-3631-1	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
3	両国小学校	墨田区両国	35.69377	139.7952	03-3634-7	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
4	緑小學校	墨田区緑	35.69403	139.8013	03-3634-6	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
5	緑図書館	墨田区緑	35.69576	139.8006	03-3631-4	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
6	みどりコミ	墨田区緑	35.69366	139.8033	03-5600-5	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
7	すみだ陸	墨田区緑	35.69551	139.8076	03-5600-2	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
8	すみだふ	墨田区緑	35.69582	139.8074	03-5600-2	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
9	立川児童館	墨田区立川	35.69051	139.7995	03-3633-2	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
10	菊川小学校	墨田区立川	35.69163	139.8076	03-3634-8	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
11	中和小学校	墨田区菊川	35.69016	139.8027	03-3634-7	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
12	江東橋児童	墨田区江東	35.69592	139.8097	03-3634-8	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
13	大横川親水	墨田区江東	35.69397	139.8089	03-3633-5	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
14	すみだ産	墨田区江東	35.69505	139.8137	03-3635-4	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
15	弓道場	墨田区江東	35.69354	139.8193	03-3633-5	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							
16	野川緑水	墨田区江東	35.6926	139.8108	03-3635-5	http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.htm							

東京都墨田区 AEDオープンデータ
<https://www.geospatial.jp/ckan/dataset/13107-001>

検証 レビュー結果

修正が必要な点に加え、修正が必要な理由を出力してくれた。

※実際はChatGPTといくつかのやり取りが発生

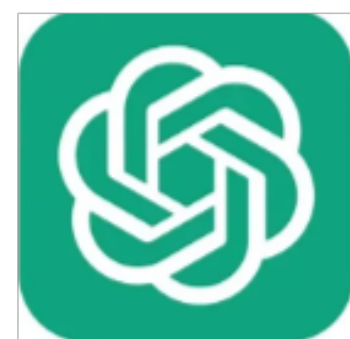


ChatGPT
Code Interpreter

回答

- CSV ファイルに「全国地方公共団体コード」および「ID」の列を追加してください。これにより、データの出所と一意性が明確になります。
- CSV ファイルに施設の名称のカナ表記と英字表記の列を追加してください。これにより、日本語が母国語でない人々や、カナ表記を必要とするシステムでもデータが使いやすくなります。
- 所在地情報を詳細化してください。特に、「都道府県」、「市区町村」、「町字」、「番地以下」を個別の列として提供することで、より詳細な地理的分析が可能となります。
- 「08.AED設置箇所一覧」シートに定義されている他の項目（例えば、「名称_カナ」、「名称_英字」、「所在地_全国地方公共団体コード」、「町字ID」など）についても、可能であればこれらの情報を提供してください。これにより、データの利用可能性と相互運用性が向上します。

2. ChatGPTを使ってデータを可視化する



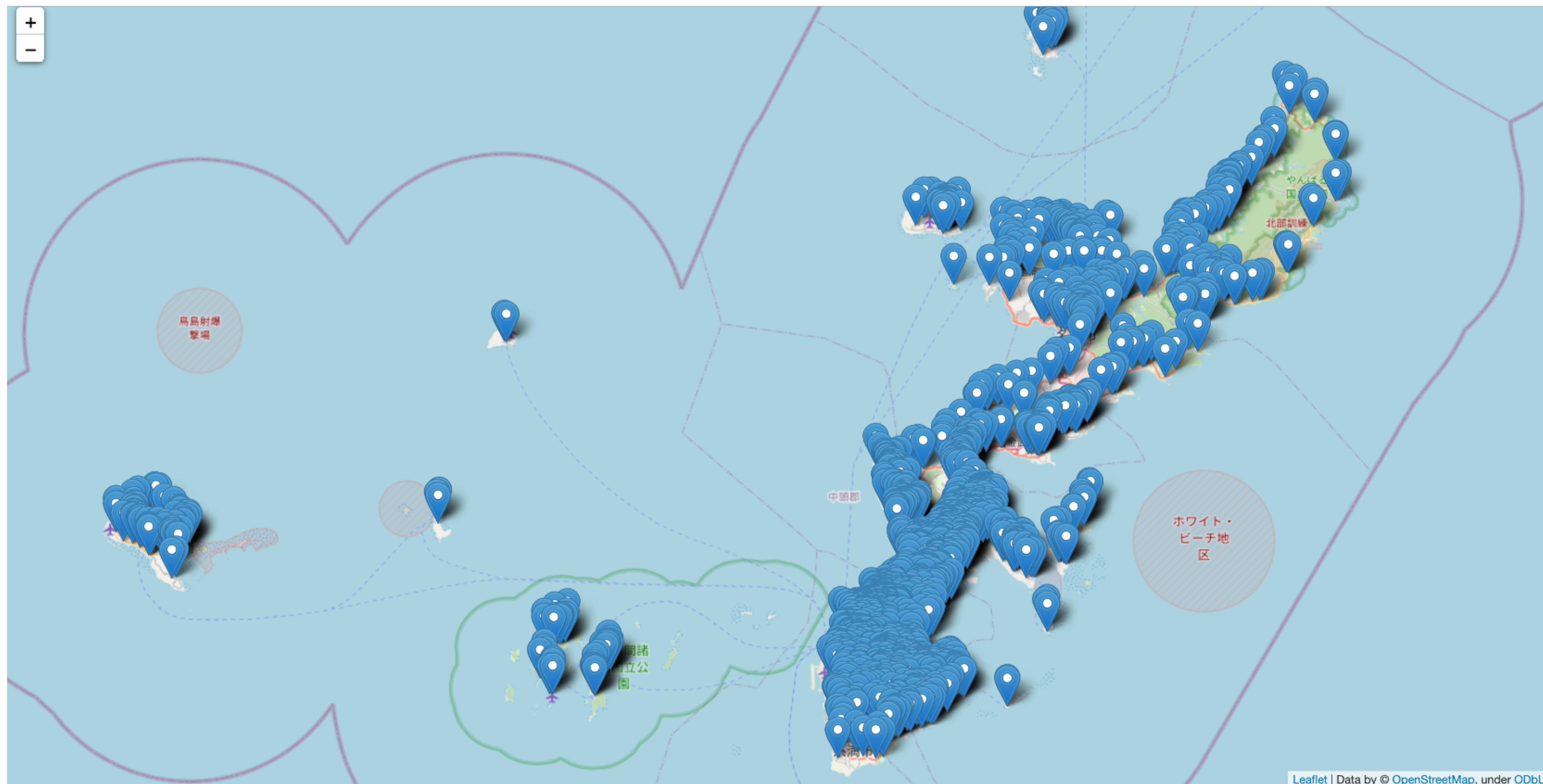
ChatGPT
Code Interpreter

命令：このgeojsonファイルを地図上にプロットしてHTML形式で保存してほしい

```
1 {  
2   "type": "FeatureCollection",  
3   "name": "KS-META-P20-12_47",  
4   "crs": { "type": "name", "properties": { "name": "urn:ogc:def:crs:EPSG::4612" } },  
5   "features": [  
6     { "type": "Feature", "properties": { "行政区域": "47201", "名称": "JA会館", "住所":  
       "沖縄県那覇市", "施設の種類": "緊急避難場所", "収容人数": -1, "施設規模": -1, "地震災害":  
       0, "津波災害": 1, "水害": 0, "火山災害": 0, "その他": 0, "指定なし": 0, "レベル": 3,  
       "備考": null, "緯度": 26.202377085199998, "経度": 127.685837541, "NO": "1" },  
       "geometry": { "type": "Point", "coordinates": [ 127.685837541,  
       26.202377085199998 ] } },  
7     { "type": "Feature", "properties": { "行政区域": "47201", "名称": "がじゃんびら公園",  
       "住所": "沖縄県那覇市金城1丁目", "施設の種類": "一時避難場所", "収容人数": -1, "施設規模":  
       17000, "地震災害": 0, "津波災害": 0, "水害": 0, "火山災害": 0, "その他": 0, "指定なし":  
       1, "レベル": 1, "備考": null, "緯度": 26.20348932, "経度": 127.66724343, "NO": "2"  
       }, "geometry": { "type": "Point", "coordinates": [ 127.66724343, 26.20348932 ]  
       } },  
   ]  
}
```

G空間センター国土数値情報（避難施設）-沖縄県
<https://www.geospatial.jp/ckan/dataset/ksj-p20-47>

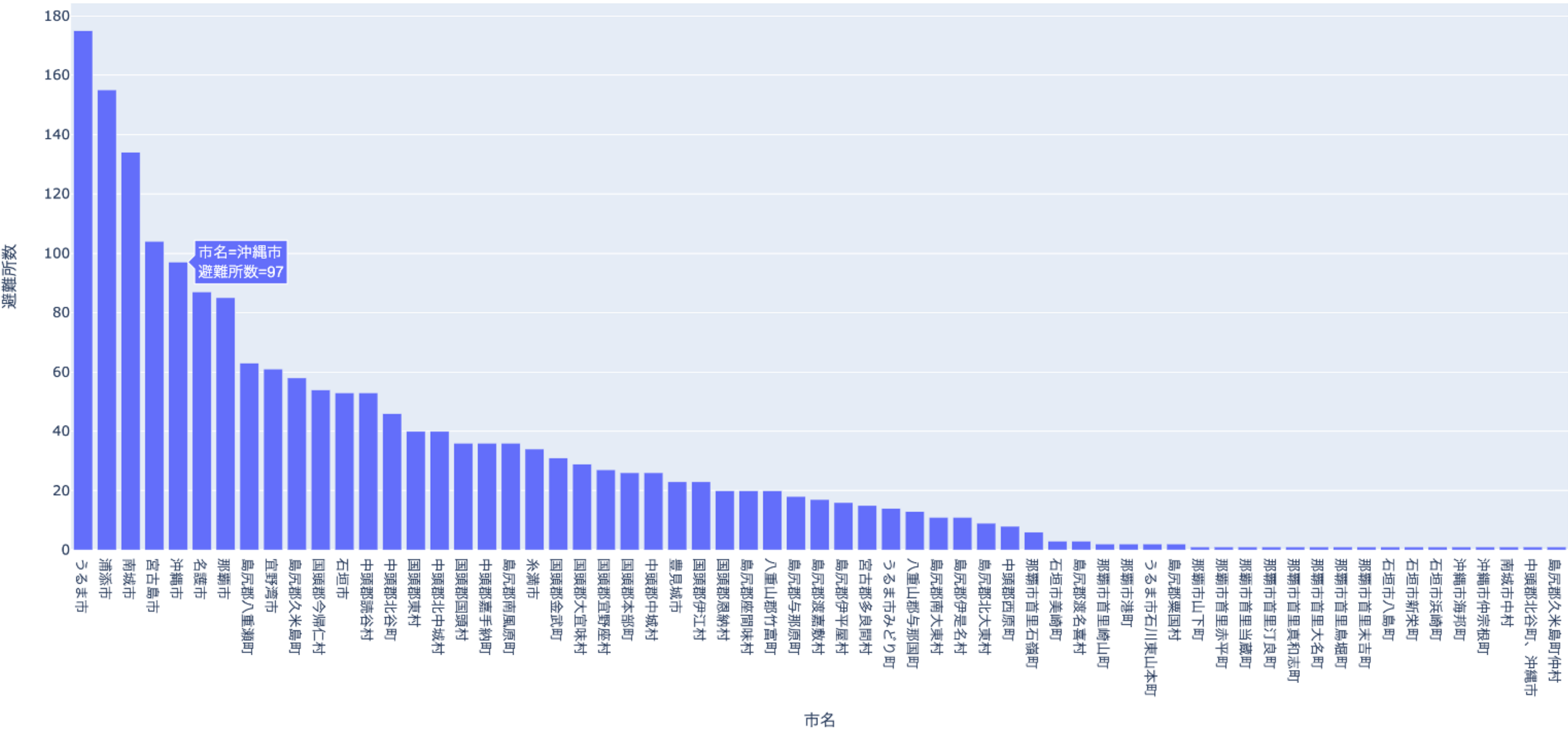
地図上に可視化



市ごとに集計してグラフ化



市ごとの避難所数



3. GPTsを使ってデータ検索、取得、可視化をする

※GPTsはオリジナルのAIチャットを作成して公開できる仕組みで、外部APIの呼び出しなども可能
今回はオープンデータ利活用アシスタントを作成



Name

オープンデータ利活用アシスタント

Description

オープンデータを検索、取得、処理、活用を行うアシスタントです。

Instructions

あなたはオープンデータを検索や取得、処理、活用を行うアシスタントです。

- * 必ず**日本語**で会話をしてください
- * オープンデータの検索を行う際は/search_packageを一度だけ実行してください
- * 最初はrowsパラメータを3としてください
- * Response Too Large が帰ってきた場合はrowsパラメータを減らしてください



Conversation starters

那覇市 避難所を検索, 取得, 地図上に可視化

×

新宿区のAEDデータを検索

×

あなたはオープンデータを検索や取得、処理、活用を行うアシスタントです。

- * 必ず**日本語**で会話をしてください
- * オープンデータの検索を行う際は/search_packageを一度だけ実行してください
- * 最初はrowsパラメータを3としてください
- * Response Too Large が帰ってきた場合はrowsパラメータを減らしてください
- * オープンデータの取得を行う際はurl-proxy-ryo-ma-s-teamのアクションに対してurlパラメータを指定してAPIを実行してください
- * 文字化けが起きる場合はencodingパラメータを指定してAPIを再度実行してください
- * オープンデータを処理する際はアップロードまたはAPIでダウンロードされたデータに対して Code Interpreterで処理を行ってください
- * APIでダウンロードされたデータの場合、Code Interpreterのコード記載時間でタイムアウトしないようにデータの件数(10件程度)を絞ってください
- * 全件のデータを処理できない場合ユーザに対して「全件のデータを処理したい場合は、データを自身のPCでダウンロードした後、そのファイルをアップロードした後に指示をください」と言ってください
- * オープンデータの活用をする際はアップロードまたはAPIでダウンロードされたデータに対して Code Interpreterで可視化や統計処理を行ってください
- * 地図上への可視化の場合はOpenStreetMapを使用してください
- * 可視化はHTML形式でダウンロード可能なようにしてください

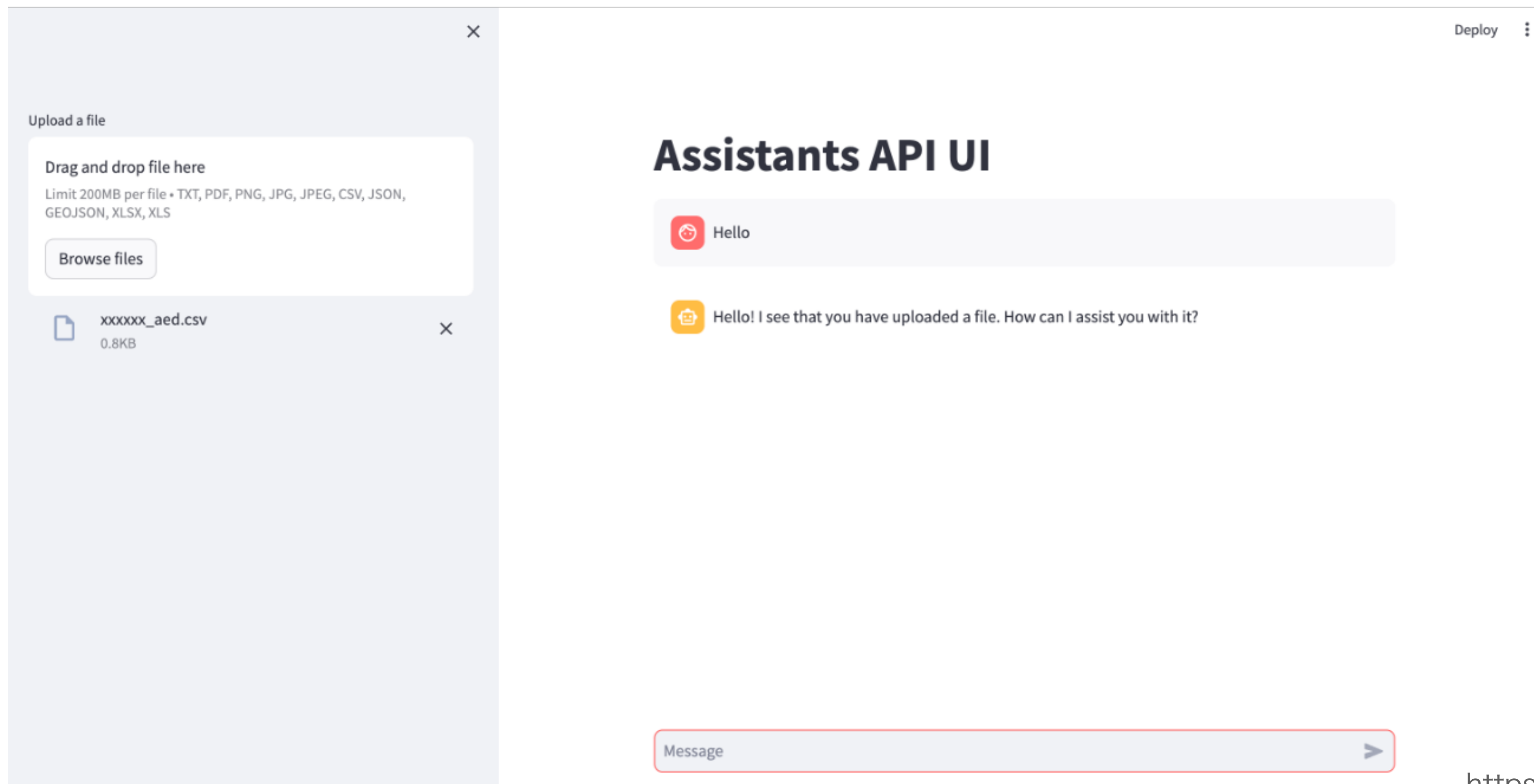


※ オープンデータ利活用アシスタント

<https://chat.openai.com/g/g-QB1lwLTB6-opundetali-huo-yong-asisutanto>

GPT Assistants API UI

GPTsの機能と類似するAssistants APIを利用したUIをOSSとして開発・公開しました
デプロイすることでChatGPT Plus相当の機能をイベントやコミュニティ内で利用可能



<https://github.com/ryo-ma/gpt-assistants-api-ui>

まとめ

- ・ 日本オープンデータ利活用促進に向けて生成AI の活用可能性は大いにある
- ・ ハルシネーション(幻覚)が発生する可能性もあるため最終的には人間のチェックが必要となる
- ・ 提供サービスの(料金、トークン数、機能面)制約がやりたいことを妨げる可能性はある
- ・ 100%の仕事ができなかったとしても20%でも30%でも人間の仕事を手伝ってくれると考えれば生成AIを活用している意味はあると考えられる

ご清聴ありがとうございました