

第二回 OOL Tech Connect

生成AIを活用したオープンデータ活用

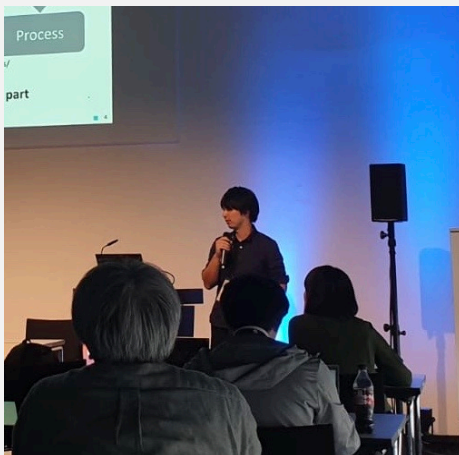
TIS株式会社

デジタル社会サービス企画ユニット

デジタル社会サービス企画部

坂本 諒太

1. 自己紹介
2. OOLでの取り組み
3. オープンデータの背景
4. オープンデータの課題を生成AIを使って解決、改善
5. 生成AIによる「オープンデータ作成」
6. 生成AIによる「オープンデータ活用」
7. 全体のまとめ



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、テニス

SNS

Twitter
@SkeinIn

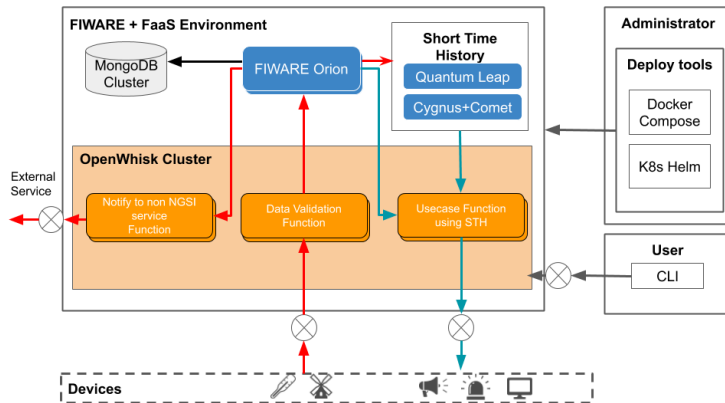
Github
@ryo-ma

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

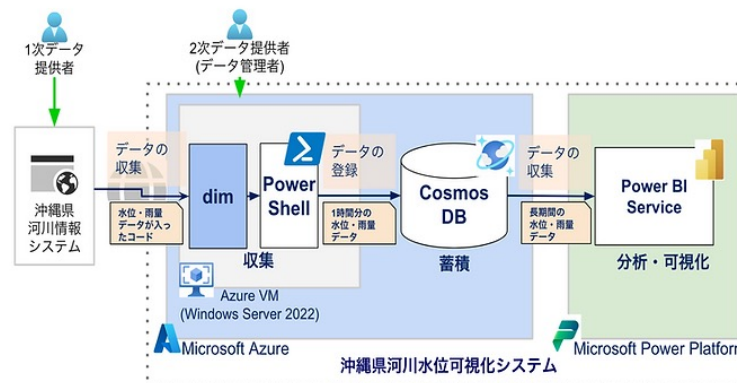
2019~ IoTプラットフォームPJ

- FIWARE 検証
- FIWARE Meteoroid開発



2021~ DataOps PJ

- DataOpsプロセス検討
- DataOps基盤検証

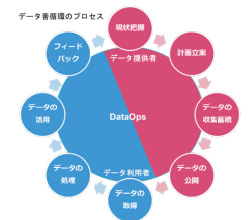
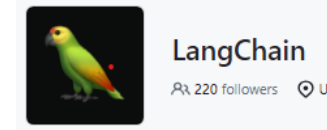


現在

2023~ DataOps PJ

- AI × DataOps検証
- ユースケース検証

 **OpenAI**



オープンデータの背景

- **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
- **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

オープンデータの課題感は？

オープンデータの提供者側も利用者側も、オープンデータを活用したいと切に願っている。
しかしお互いの思いが直接伝わらないため、「とりあえずできるところから始めてみる」の先へ進むことができていない。

データ提供者（自治体や地場の企業等）



データ利用者（住民やシビックハッカー、企業等）



オープンデータの課題を生成AIを使って解決、改善

オープンデータの作成と活用における課題では生成AIを使うことで解決、改善につながると考えた

データ提供者の課題

オープンデータの作成

- データを綺麗に作れる
- データの間違いを知る
- 間違ったデータを修正する

生成AI活用



ChatGPT 等

データ利用者の課題

オープンデータの活用

- データを使えるように加工
- 地図にデータを出す
- グラフにデータを出す

1. オープンデータの知識がなくても作れる

2. 開発知識がなくてもデータが活用できるようになる

生成AIによる「オープンデータ作成」

世の中のオープンデータは使いにくい形式(PDF、XLSX)だったり
必要な項目が不足していたりする



正しいオープンデータの作り方を知る必要がある



オープンデータのレビューAIを作る



OpenAI社が公開したPythonの実行とアップロード・ダウンロードが可能なモデル
現在はChatGPT Plusでベータ版として提供が開始されている

◆ 特徴

- 外部環境から切り離されたサンドボックス上でPythonコードを実行可能
- 実行したコードはチャットを消さない限り、コードの再利用が可能
- チャットへのファイルアップロードが可能
- 実行結果のダウンロードが可能

◆ ユースケース

- 数学的な問題の解決 (定量的 / 定性的)
- データの分析と可視化
- ファイルのフォーマット変換

◆ 安全性について

- ネットワークの制限をかけることで現実世界へ予期せぬ作用をもたらさないように配慮
- それぞれのセッションのリソースを一定に制限

2. オープンデータのルールに基づいて回答

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

Advanced Data Analysis
(旧Code Interpreter)

ファイル 文字コード変換
文字コードをshift-jis -> utf-8

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。

新たな項目の提案が多かったため、既存の項目に対するフィードバックを追加で要求

※実際はChatGPTといくつかのやり取りが発生



ChatGPT

Advanced Data Analysis
(旧Code Interpreter)

命令：既に記載されているデータに関して、データ項目定義書に則ったフィードバックはありますか？

回答

- 施設名称（名称）：既存のデータは定義書に準拠しています。特に改善点は見られません。
- 所在地: 定義書では所在地情報は「都道府県」、「市区町村」、「町字」、「番地以下」のように細かく分けられていますが、提供されたデータではこれらが一つの列にまとめられています。これらを個別の列に分けることで、データの活用性が高まります。
- 緯度・経度: これらのデータは定義書に準拠しています。ただし、測地系（例：世界測地系（WGS）や日本測地系）が明示されていない場合、これらの情報を明記することをお勧めします。
- 電話番号: 定義書には電話番号の項目は見られません。しかし、この情報は施設への連絡方法として重要なため、現状維持でも問題ないと思われます。
- URL: 定義書にはURLの項目は見られませんが、施設やAEDの詳細情報を提供する上で有用な情報であるため、現状維持でも問題ないと思われます。ただし、これらがリンク切れでないこと、または最新の情報を反映していることを定期的を確認することをお勧めします。

- **良い点**

- データ項目定義書を一度学習させることで、インプットに合わせたシートを参照してフィードバックをしてくれる
- インプットファイルの文字コード変換などの処理を自動でやってくれる

- **課題点**

- 一定時間経過するとデータ項目定義書の再インプットが必要になる
- Code Interpreterを利用するモデルでは、インターネットアクセスが無効化されており、リンク切れの確認などは自分自身で行う必要がある
- プロンプトを工夫しなければならない
 - フィードバックが欲しい項目を詳しく指定する必要がある

生成AIによる「オープンデータ活用」



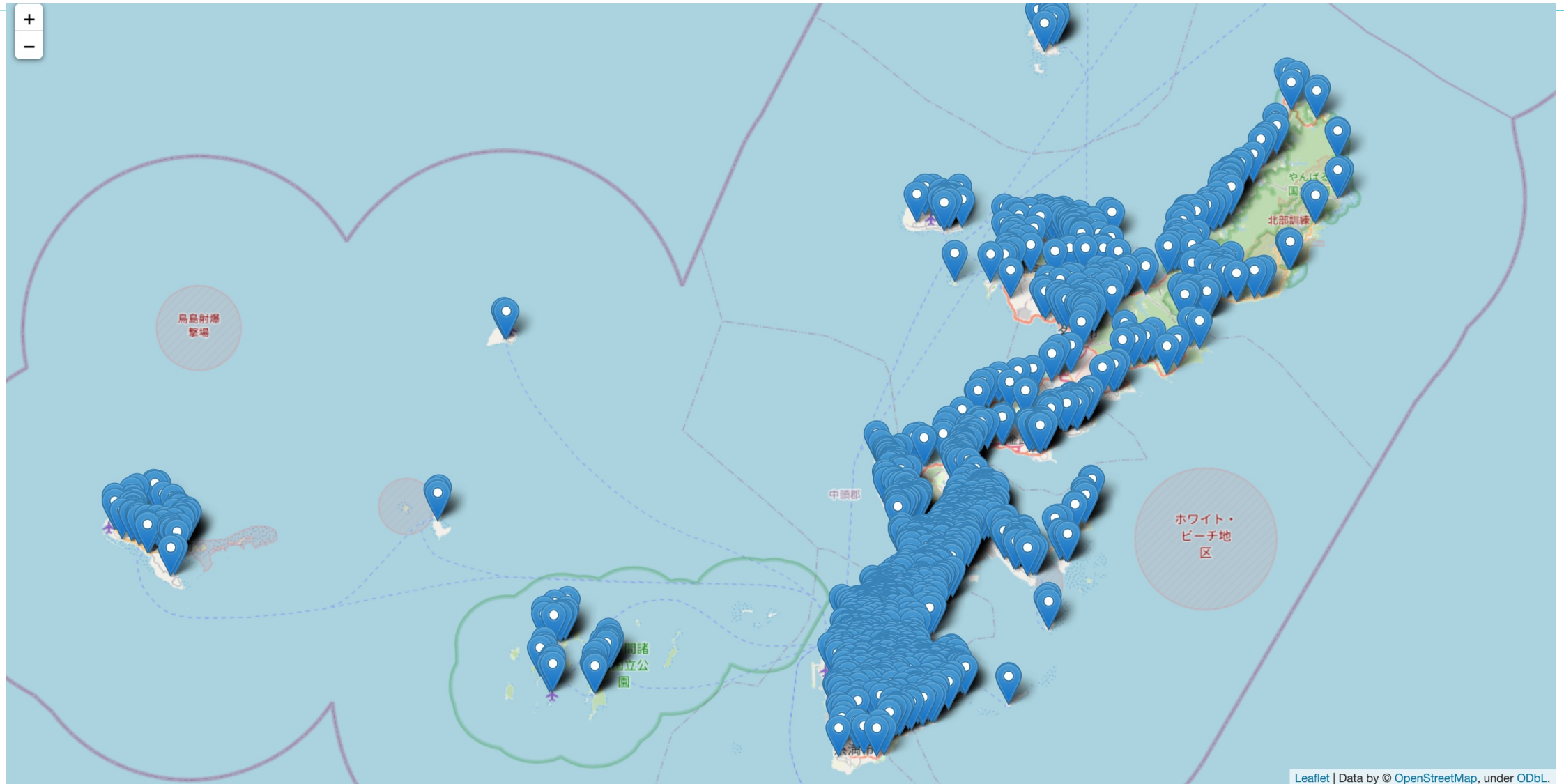
ChatGPT

Advanced Data Analysis
(旧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

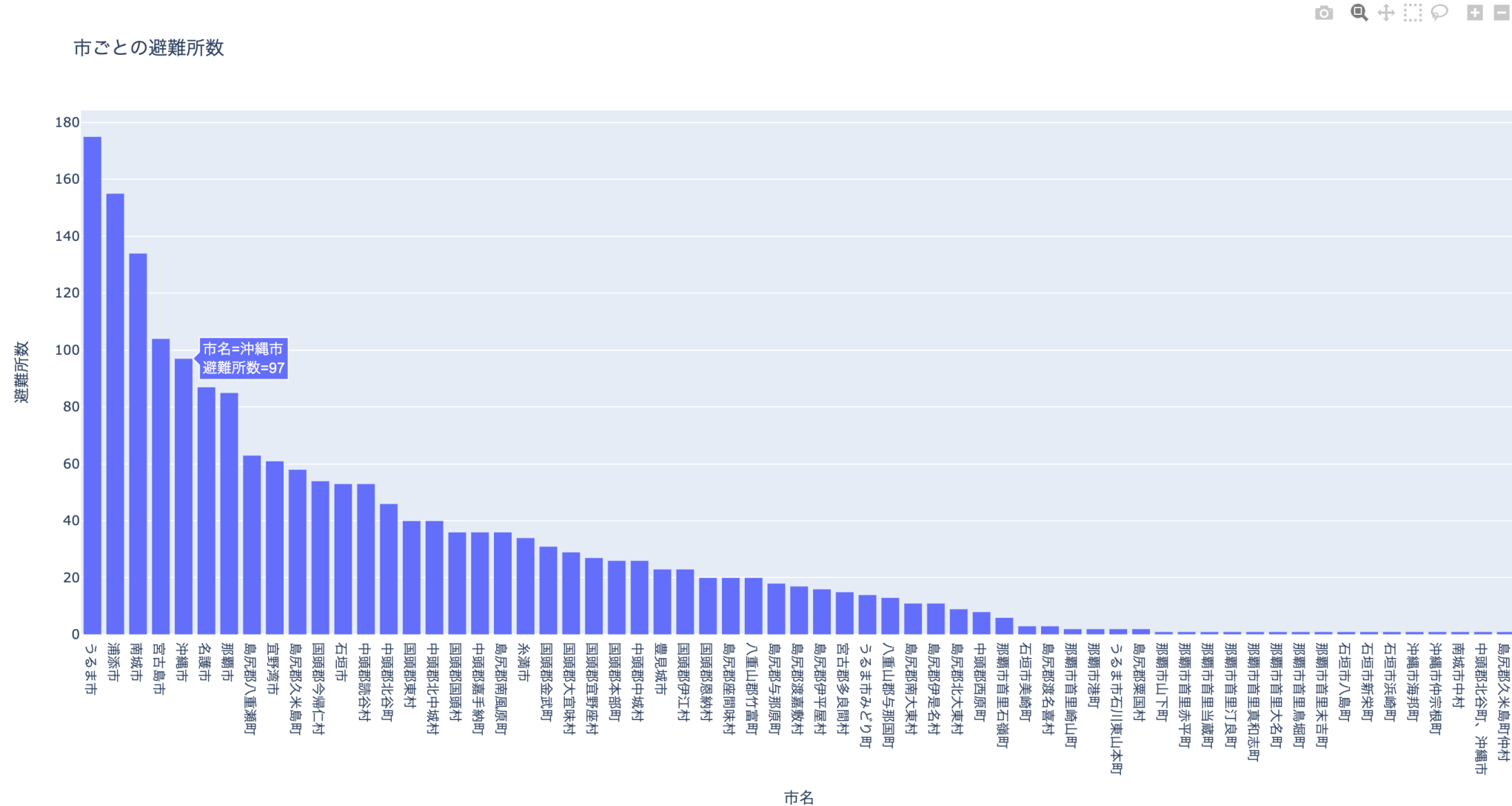
```
1 {  
2   "type": "FeatureCollection",  
3   "name": "KS-META-P20-12_47",  
4   "crs": { "type": "name", "properties": { "name": "urn:ogc:def:crs:EPSG::4612" } },  
5   "features": [  
6     { "type": "Feature", "properties": { "行政区域": "47201", "名称": "J A会館", "住所":  
       "沖縄県那覇市", "施設の種類": "緊急避難場所", "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>



OpenStreetMap上に避難所のピンが立ったHTMLが生成された
※実際はChatGPTといくつかのやり取りが発生

先ほどのデータを市ごとに避難所数を集計してグラフ化を命令



Plotlyを使ったインタラクティブなグラフのHTMLページが生成された
※実際はChatGPTといくつかのやり取りが発生

- **良い点**

- 可視化、集計などコーディングの手間が無くなった
- ファイルを読み込んでいい感じに内容を解釈、必要な方法を随時提案してくれる

- **課題点**

- 一発でファイルが生成されたわけではなく、ChatGPTもトライアンドエラーで進んでいく(課題点というほどでもない)
- うまくいかない場合もあるので、回避策を命令するためにITの知識が必要な場合があった
- 誤りもあるので人間のチェックは必要

- **オープンデータの課題感に対してChatGPTを活用した**
 - データ作成の課題感に対して生成AIの活用
 - オープンデータのレビューAIを作成
 - データ利用の課題感に対して生成AIの活用
 - オープンデータの可視化をChatGPTで実現
- **今後**
 - 他のオープンデータの課題感に対してもアプローチ
 - 例：生成AIを使ってより使いやすい形式のオープンデータへ整形など

ご清聴ありがとうございました

ITで、社会の願い叶えよう。

