

ご当地グルメマップを作ろう

Let's make! Local Food Map

PyCon APAC 2023 Day2

Hiroshi Sano / 佐野浩士



Today's Document

本日の資料は公開されています

- スライド: Speaker Deck

<https://speakerdeck.com/hrsono645/pyconapac2023-lets-make-localfood-map>

- GitHub リポジトリ: コードとスライド

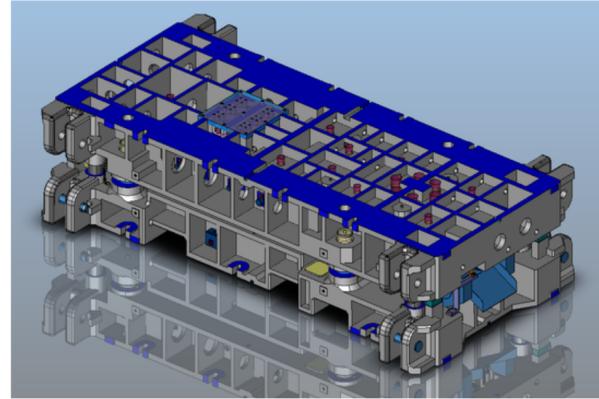
<https://github.com/hrsono645/pyconapac2023-local-food-map>

Self Infroduction

Hiroshi Sano (佐野浩士) [@hrs_sano645](#)

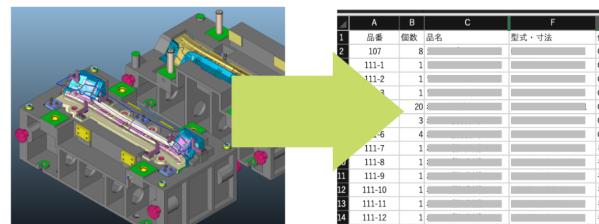
- 🗺️: Shizuoka, Eastern part 🏔️
- 🏢: 株式会社佐野設計事務所 CEO
- 💬🤝
 - 🐍: PyCon mini Shizuoka Stuff / Shizuoka.py / Unagi.py / Python 駿河
 - CivicTech, Startup Weekend Organizer
- Hobby: Camp 🌳, DIY 🔨, IoT💡





3D機械設計 / 3Dモデリング

自動車向けプレス金型の3D設計
金型設計関連の3Dモデリング
製品データの3Dモデリング



製造業DX支援



設計データをデジタル化
製造業DX取り組みサポート

Agenda

- トークのモチベーション
 - 今回のトークでできること
1. お店情報をデータにする: WEBスクレイピング
 2. データを整形をする: CSVファイルにする
 3. データを使ってみる: Googleマイマップで呼び出す
- まとめ

トークのモチベーション

- ご当地グルメを情報収集してマップを作りましょう！
 - ご当地グルメ=B級グルメのこと
 - とあるご当地グルメを例にしています
- PyCampを終えた人に：
データを集める、作る、利用するを元にプログラムの実装プロセスを学べます
 - 話すこと: トークお題をPythonで実装する過程
 - 話さないこと: Pythonの基礎や文法の解説

PyCamp:Python Boot Campとは

- 日本全国で開催されているPythonのチュートリアルイベント
 - 一般社団法人PyCon JP Associationが主催
- 半日でPythonの基礎を学び、簡単なプログラムを作る
 - 専用テキストを元に講師、TAがサポート



Python Boot Camp Text ドキュメント

Q. 検索

Python Boot Camp 運営マニュアル

Python Boot Camp テキスト

v: latest

Python Boot Camp Text

本テキストは日本各地での初心者向けPythonチュートリアルイベント「Python Boot Camp」で使用するため、Pythonの開発環境構築、基礎的な文法や実践応用についてまとめたものです。

- GitHub: <https://github.com/pyconjp/bootcamp-text/>
- Read the Docs: <https://readthedocs.org/projects/bootcamp-text/>

目次

- Python Boot Camp 運営マニュアル
 - 1. Python Boot Camp について
 - 2. Python Boot Camp 運営マニュアル
 - 3. Python Boot Camp エンタープライズマニュアル
 - 4. 貢献者一覧
- Python Boot Camp テキスト
 - 1. Pythonではじめる前に
 - 2. Pythonをはじめよう
 - 3. Pythonデータ型（基本編）
 - 4. Pythonデータ型（コレクション編）
 - 5. ファイル操作とモジュール
 - 6. サードパーティ製パッケージと venv
 - 7. Web API、スクレイピング
 - 8. 次のステップ

ライセンス

This work is licensed under a Creative Commons Attribution 4.0 International License.



PyCampを終えた方の次にチャレンジできるコンテンツを目指して作りました

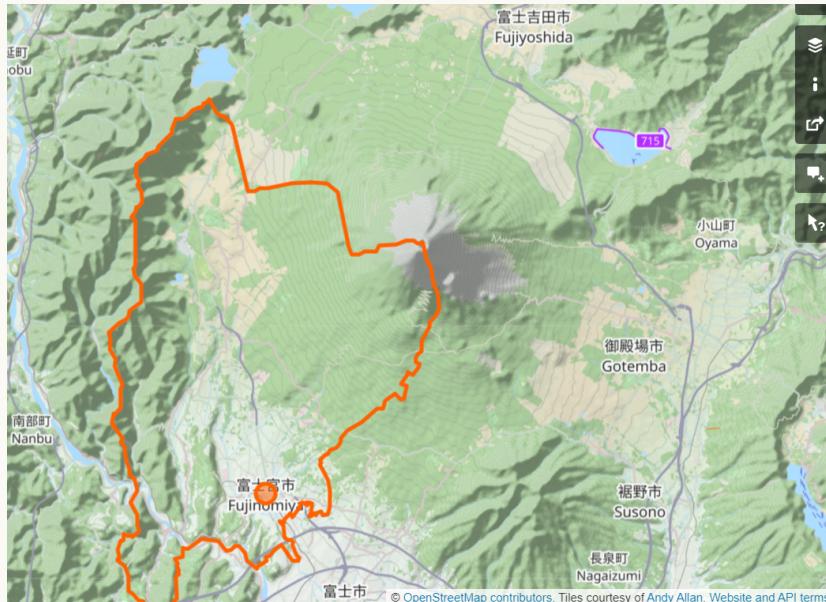
写真は静岡県3回目の様子

今回のお題: 本題のご当地グルメ



富士宮焼きそば

- 主に静岡県富士宮市周辺で食べられる焼きそば
- B-1グランプリ殿堂入り、NYで開催されたコナモングランプリで優勝



地元の人は、多分週一ぐらいで食べてる

ところで、みなさん、

もう食べたりましたね？ 😋

このあとはパーティ🎉ですし、お腹空きましたね 😭🍻

富士宮焼きそばマップを作りましょう🔍Chefcook emoji💪

今回のトークでできること

1. お店情報をデータにする: WEBスクレイピング
2. データを整形をする: CSVファイルにする
3. データを使ってみる: Googleマイマップで呼び出す

1. どこからデータを取り出すか



2. どんなデータを作るか



3. どこへデータを渡すか

今回のトークでできること

1. お店情報をデータにする: WEBスクレイピング
2. データを整形をする: CSVファイルにする
3. データを使ってみる: Googleマイマップ

抽象的に言い直すと...



1. どこから データを取り出すか
2. どんな データを作るか
3. どこへ データを渡すか

1. どこからデータを取り出すか



2. どんなデータを作るか



3. どこへデータを渡すか

1. どこからデータを取り出すか

- 🔎 WEBスクレイピングで収集する
- [付録] 📸 画像識別で収集する

利用するライブラリ

- requests: HTTPアクセス→情報取得（今回はHTML）
- BeautifulSoup4: HTML（マークアップ言語）解析と抽出

```
pip install requests  
pip install beautifulsoup4
```

*:スライドのコードは説明向けです。そのままだと動かないこともあります
資料のリポジトリから動作するスクリプトをDL可能です

ご当地グルメの情報はどこにあるか

- 地域情報を収集
- **その情報は機械可読ができるか？**
 - 大体が紙ベースが多い  パンフレットとか
 - (画像識別の手段は使える)

観光情報を探ってみる

- 市役所、観光協会のWEBサイトで紹介されていたり
- ご当地グルメの公式サイト（よく〇〇学会とも言われる）

今回は、富士宮焼きそば学会の公式サイトを例にしています
(左: お店一覧、右: お店の詳細)

The image shows two side-by-side browser windows displaying the Fujinomiya Yakisoba Society website. The left window shows a list of shops ('お店一覧') with details like name, address, phone number, and fax number. The right window shows a detailed view of a specific shop ('お好焼 あき'), including its address, area, map, business type, food type, and location.

Left Window (Shop List View):

- 要素 (Elements): Shows the HTML structure of the shop list page.
- ソース (Source): Shows the raw HTML code of the shop list page.
- Shop List Items (Excerpted):
 - [お好焼 あき](https://umya-yakisoba.com/shop/3776/)
 - [あさ家](https://umya-yakisoba.com/shop/3777/)
 - [あさ家](https://umya-yakisoba.com/shop/3774/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3616/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3772/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3779/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3618/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3778/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3780/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3782/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3627/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3783/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3784/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3633/)
 - [お好焼 あき](https://umya-yakisoba.com/shop/3769/)

Right Window (Shop Detail View):

- 要素 (Elements): Shows the HTML structure of the shop detail page for 'お好焼 あき'.
- ソース (Source): Shows the raw HTML code of the shop detail page for 'お好焼 あき'.
- Shop Detail Fields (Excerpted):
 - エリア: 富士宮市野中東町112-1
 - 地図: (Map)
 - 業種: 飲食店
 - 住所: 富士宮市野中東町112-1

```
> <header class="article-header entry-header">...</header>
<div class="article">
  <div class="p-shopList" style="flex">
    <a href="https://umya-yakisoba.com/shop/3776/">
      <figure class="p-shopthumb-ph-no">...</figure>
      <div class="p-shopthumb-text" style="flex"> == $0
        <div>お好焼 あき</div>
        <div>富士宮市野中東町112-1</div>
        <div>0544-27-0004</div>
        <div>定休日：火曜日</div>
      </div>
    </a>
    <a href="https://umya-yakisoba.com/shop/3777/">...</a>
    <a href="https://umya-yakisoba.com/shop/3774/">...</a>
    <a href="https://umya-yakisoba.com/shop/3616/">...</a>
    <a href="https://umya-yakisoba.com/shop/3772/">...</a>
    <a href="https://umya-yakisoba.com/shop/3779/">...</a>
    <a href="https://umya-yakisoba.com/shop/3618/">...</a>
    <a href="https://umya-yakisoba.com/shop/3778/">...</a>
    <a href="https://umya-yakisoba.com/shop/3780/">...</a>
    <a href="https://umya-yakisoba.com/shop/3782/">...</a>
    <a href="https://umya-yakisoba.com/shop/3627/">...</a>
    <a href="https://umya-yakisoba.com/shop/3783/">...</a>
    <a href="https://umya-yakisoba.com/shop/3784/">...</a>
    <a href="https://umya-yakisoba.com/shop/3633/">...</a>
    <a href="https://umya-yakisoba.com/shop/3769/">...</a>
```

お店一覧のaタグを収集する

```
import requests
from bs4 import BeautifulSoup

url = "https://umya-yakisoba.com/shop/"
shopinfo_list = []
res = requests.get(url)
soup = BeautifulSoup(res.text, 'html.parser')

# ここではdiv.p-shopList > a にURLがある
# その中にお店情報がまとまっているので、aタグから取り出す
shopinfo_tags = soup.find(
  'div', class_='p-shopList'
).find_all('a')
```

```
> <header class="article-header entry-header">...</header>
> <div class="article">
>   <div class="p-shopList" style="flex">
>     <a href="https://umya-yakisoba.com/shop/3776/">
>       <figure class="p-shopthumb-ph-no">...</figure>
>       <div class="p-shopthumb-text" style="flex"> == $0
>         <div>お好焼 あき</div>
>         <div>富士宮市野中東町112-1</div>
>         <div>0544-27-0004</div>
>         <div>定休日：火曜日</div>
>       </div>
>     </a>
>     <a href="https://umya-yakisoba.com/shop/3777/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3774/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3616/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3772/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3779/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3618/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3778/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3780/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3782/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3627/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3783/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3784/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3633/">...</a>
>     <a href="https://umya-yakisoba.com/shop/3769/">...</a>
```

aタグや、中にあるタグから必要な情報を取得する

```
for shopinfo_tag in shopinfo_tags:
    shopdata = {}
    # divタグの並びは上から店名、住所、電話番号、定休日
    # ここではurlと店名だけまとめたリストを作る
    shopdata['specurl'] = shopinfo_tag.get('href')
    shopdata['店名'] = shopinfo_tag.find_all("div")[1].text

    shopinfo_list.append(shopdata)
```

お店の詳細URL、店名が集まりました🎉

```
>>> shopinfo_list
[{'specurl': 'https://umya-yakisoba.com/shop/3776/', '店名': 'お好焼 あき'},
 {'specurl': 'https://umya-yakisoba.com/shop/3777/', '店名': 'あさ家'},
 {'specurl': 'https://umya-yakisoba.com/shop/3774/', '店名': 'あるばとろす'},
 {'specurl': 'https://umya-yakisoba.com/shop/3616/', '店名': 'いっぷく亭'},
 ...
]
```

*この例では店名に \u3000 (全角スペースの意味) が入ることがあり、半角スペースへ置き換え処理をした結果になります

```
> <header class="article-header entry-h
<div class="p-shopPh" > </div> flex
  <dl class="p-shopDetails" > flex
    <dt>エリア</dt>
    ▶ <dd>...</dd> == $0
    <dt>お店名ふりがな</dt>
    <dd> あき </dd>
    <dt>地図</dt>
    <dd> B6 </dd>
    <dt>業種</dt>
    <dd> 飲食店 </dd>
    <dt>住所</dt>
    <dd> 富士宮市野中東町112-1 </dd>
    <dt>TEL</dt>
    <dd> 0544-27-0004 </dd>
    <dt>営業時間</dt>
    ▶ <dd>...</dd>
    <dt>定休日</dt>
    <dd> 火曜日 </dd>
    <dt>受入人数</dt>
    <dd> 12 </dd>
    <dt>駐車場</dt>
    <dd> 4 </dd>
```

詳細URIへアクセスして、各お店の詳細情報を収集します

```
for shopinfo in shopinfo_list:
    # URLから店舗情報を取得
    res = requests.get(shopinfo['specurl'])
    soup = BeautifulSoup(res.text, 'html.parser')

    # dl.p-shopDetails > dt/dd構造
    # dtが項目名、ddが項目の値になっている。
    # zip関数で両者組み合わせて同時に取り出す
    shopspecs = {}
    for dt, dd in zip(
        soup.find('dl', class_='p-shopDetails').find_all('dt'),
        soup.find('dl', class_='p-shopDetails').find_all('dd')
    ):
        # 辞書形式にする
        # 値に 改行や空白文字があるので取り除く
        shopspecs[dt.text] = dd.text

    # 店舗情報をマップ情報に追加
    shopinfo.update(shopspecs)
    # INFO:ここにランダム待機時間があるとよさそう
```

最終的にできるデータ

```
>>> from pprint import pprint
>>> pprint(shopinfo_list)
[{'TEL': '0544-27-0004',
 'specurl': 'https://umya-yakisoba.com/shop/3776/',
 'お店名ふりがな': 'あき',
 'エリア': 'まちなか',
 '住所': '富士宮市野中東町112-1',
 '受入人数': '12',
 '営業時間': '10:00-21:00',
 '地図': 'B6',
 '定休日': '火曜日',
 '店名': 'お好焼 あき',
 '料金目安': '350~600円',
 '業種': '飲食店',
 '焼き方': 'お店',
 '調査員おすすめメニュー': 'キムチとチーズ入り',
 '調査員が見た特徴': 'キャベツとネギが多めに入っている',
 '駐車場': '4'},
 # 以下お店情報の辞書が続いて入る...
]
```

*この例では各項目に \n (改行コード) が入ることがあり、空白へ置き換え処理をした結果になります

上記コードの注意点

※:  WEBスクレイピングは注意が必要です

- 短時間で多数アクセスはしないように注意
- 規約やポリシーを守りましょう

※: サイト上に見えない文字があることがあります → 文字列置換をしましょう

※: この例ではサイトのページネーションに対応していません
ページネーションについては資料のコードで対応しています

WEBスクレイピングでランダム時間待機できる関数の例

```
import random
from time import sleep

def random_sleep(a: int,b: int) -> None:
    """
    aからbまでのランダムな秒数を待つ
    """
    time.sleep(random.randint(a,b))

# 例: 2~5秒の間でランダムに待つ
# 複数回アクセスする時、ループの最後に差し込むと良い
random_sleep(2, 5)
```

1. どこからデータを取り出すか



2. どんなデータを作るか



3. どこへデータを渡すか

2. どんなデータを作るか

3. で利用するためのデータ（ファイル）を作成します

- 情報を整理して表形式ファイルで書き出す
- [付録] 地理情報を集める

情報を整理

どのフォーマットで書き出すか？

よくある地理データ向けファイルフォーマット

- CSV（カンマ区切り表形式、汎用性高）
- GeoJSON（WEB APIで広く流通しているJSON形式の地理情報向け）
- KML（XML形式）

Python標準のCSVライブラリを使って書き出します

`csv.DictWriter` を使うと辞書形式のデータをCSVに書き出せます

```
with open('mapdata.csv', 'w', newline='') as csvfile:  
    # ※:お店の詳細情報の各項目:辞書のキーが部分的に異なるため、  
    # 全ての項目名:辞書のキーを集めて重複を取り除いたリストを作成しています  
    csv_fieldnames = list(set()).union(*shopinfo_list)  
  
    writer = csv.DictWriter(csvfile, fieldnames=csv_fieldnames)  
    writer.writeheader()  
    for shopinfo in shopinfo_list:  
        writer.writerow(shopinfo)
```

* 変数:`csv_fieldname` では項目名のばらつきがあるため、

出力できたCSVファイル

	mapdata.csv
1	住所,営業時間,使用麺,ホームページ,駐車場,焼き方,調査員おすすめメニュー,TEL,エリア,お店名ふりがな,受入人数,specurl,地図,店名,業種,定休日,調査員が見た特徴,料金目安
2	富士宮市野中東町112-1,10:00~21:00,,,4,お店,キムチとチーズ入り,0544-27-0004,まちなか,あき,12, https://umya-yakisoba.com/shop/3776/ ,B6,お好焼 あき,飲食店,火曜日,キャベツとネギが多めに入っている,350~600円
3	富士宮市元城町13-20,7:00~19:00,,,なし,お店,塩焼そば,0544-22-3029,まちなか,あさや,10, https://umya-yakisoba.com/shop/3777/ ,C4,あさ家,飲食店,なし,具たくさん太盛、朝定食あり,400円~
4	富士宮市淀師523-4,11:00~21:00(休憩14:00~17:00),,,10,,,0544-27-6005,郊外・北部,あるばとろす,30, https://umya-yakisoba.com/shop/3774/ ,あるばとろす,飲食店,月曜日(第4週日・月連休),,500円~
5	富士宮市宮町2-9,平日11:00~16:00土日祝11:00~17:00,曾我麺,,なし,お店,焼きそば、ピリ辛焼きそば,0544-26-3849,まちなか,いっぷく亭,20, https://umya-yakisoba.com/shop/3616/ ,B4,いっぷく亭,鉄板焼・軽食,月曜・第3日曜,麺はもちもちでコシのある歯ごたえが好評で、野菜は細く切り麺とソースと絡んで食べやすく,550~750円
6	富士宮市淀師468-2,10:30~17:00※材料がなくなり次第終了,, https://tabelog.com/shizuoka/A2204/A220401/22002448/ ,25,お店,五目焼きそば,0544-27-6494,まちなか,いとう,50(3部屋あり), https://umya-yakisoba.com/shop/3772/ ,A1,好み食堂伊東,飲食店,日曜・火曜,大きな鉄板アツアツをどうぞ,550~850円

2	店名	住所	営業時間	使用麺
3	お好焼 あき	富士宮市野中東町112-1	10:00~21:00	
4	あさ家	富士宮市元城町13-20	7:00~19:00	
5	あるばとろす	富士宮市淀師523-4	11:00~21:00(休憩14:00~17:00)	
6	いっぷく亭	富士宮市宮町2-9	平日11:00~16:00土日祝11:00~17:00	曾我麺
7	好み食堂伊東	富士宮市淀師468-2	10:30~17:00※材料がなくなり次第終了	h
8	焼きそば好みうめづ	富士宮市外神1273-5	11:00~19:30	
9	うるおいてい	富士宮市淀師415-2	火~金11:30~14:00 16:30~18:30土日祝 11:00~18:00	h
10	大阪屋	富士宮市元城町12-10	10:00~20:00	h
11	海友	富士宮市西町28-9	18:00~22:30	
12	キッチントロ	富士宮市中央町7-18	11:00~17:00	
13				
14				
15				
16				
17				
18				

1. どこからデータを取り出すか



2. どんなデータを作るか



3. どこへデータを渡すか

3. どこへデータを渡すか

旅行中に使うためのツールとして

- 巨人の肩に乗る:

 Googleマイマップで使おう

- [付録]  ポータブルに扱う: 印刷をする
- [付録]  専用のWEBアプリを作ろう

Googleマイマップとは

- オリジナルマップを作成できる
- スマホ版Googleマップでも表示可能

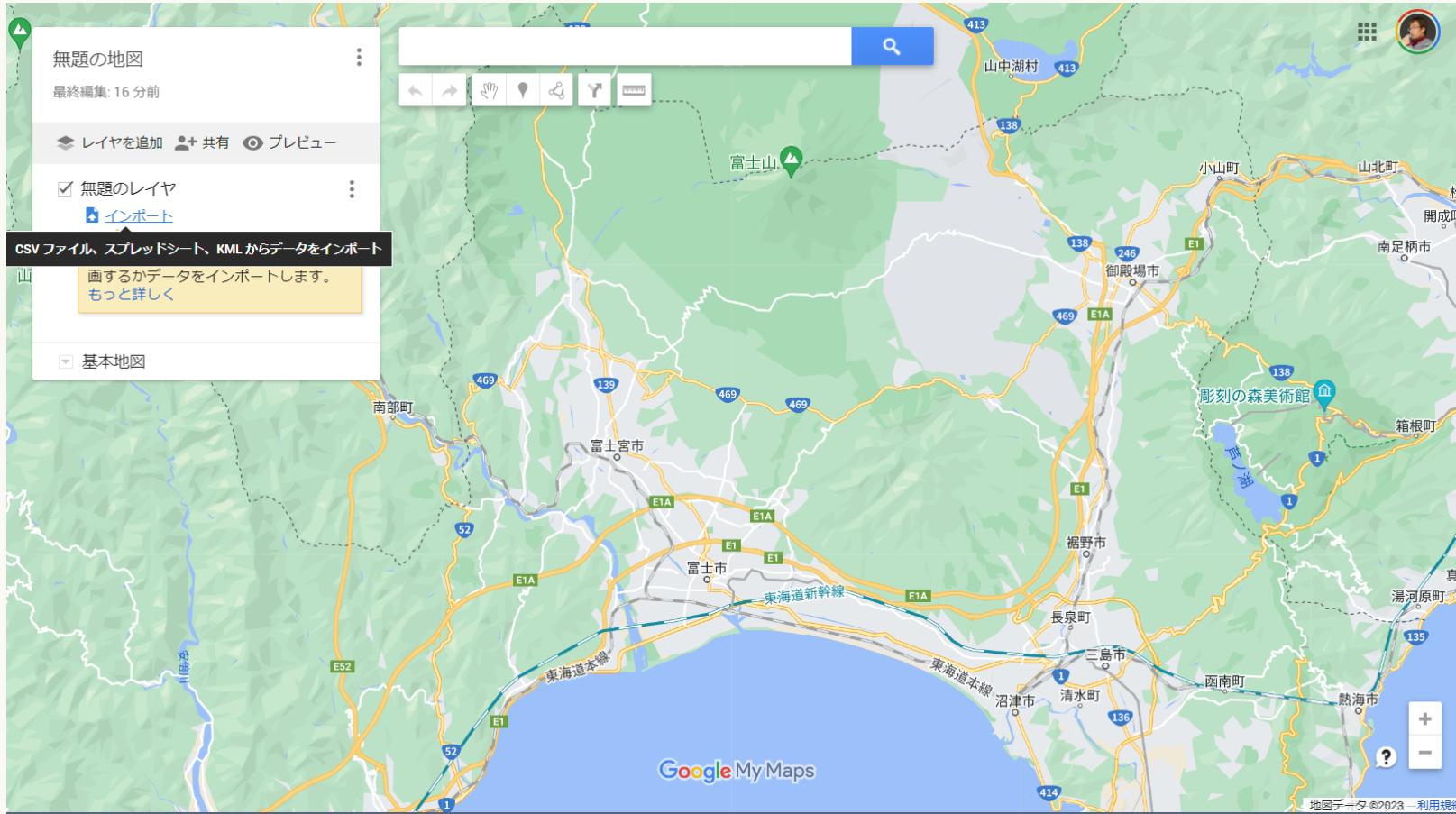
扱い方

<https://www.google.com/maps/d/> ヘアクセスして利用します

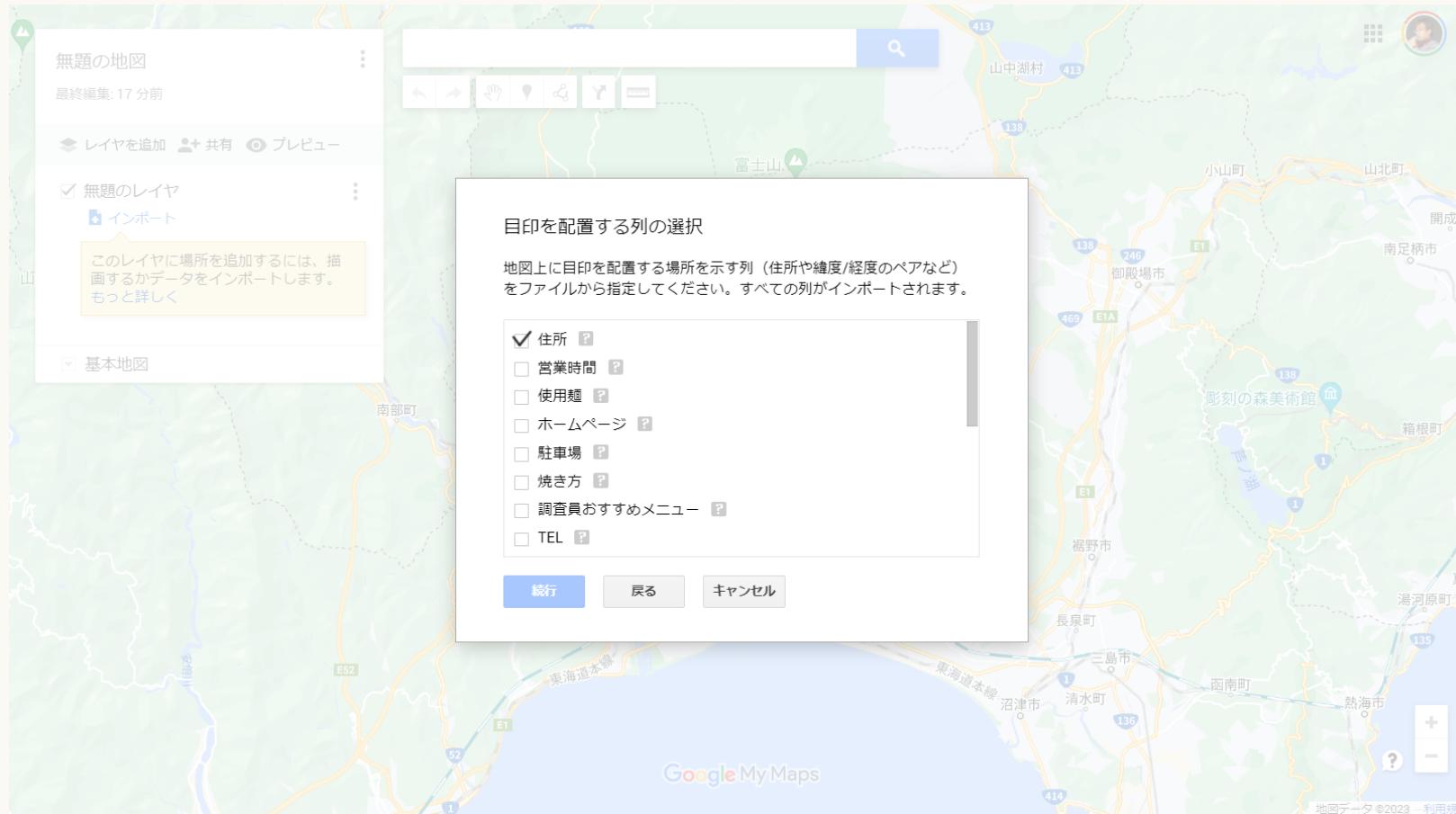
「新しい地図を作成」ボタンを押す



新規レイヤーへCSVファイルをインポートします

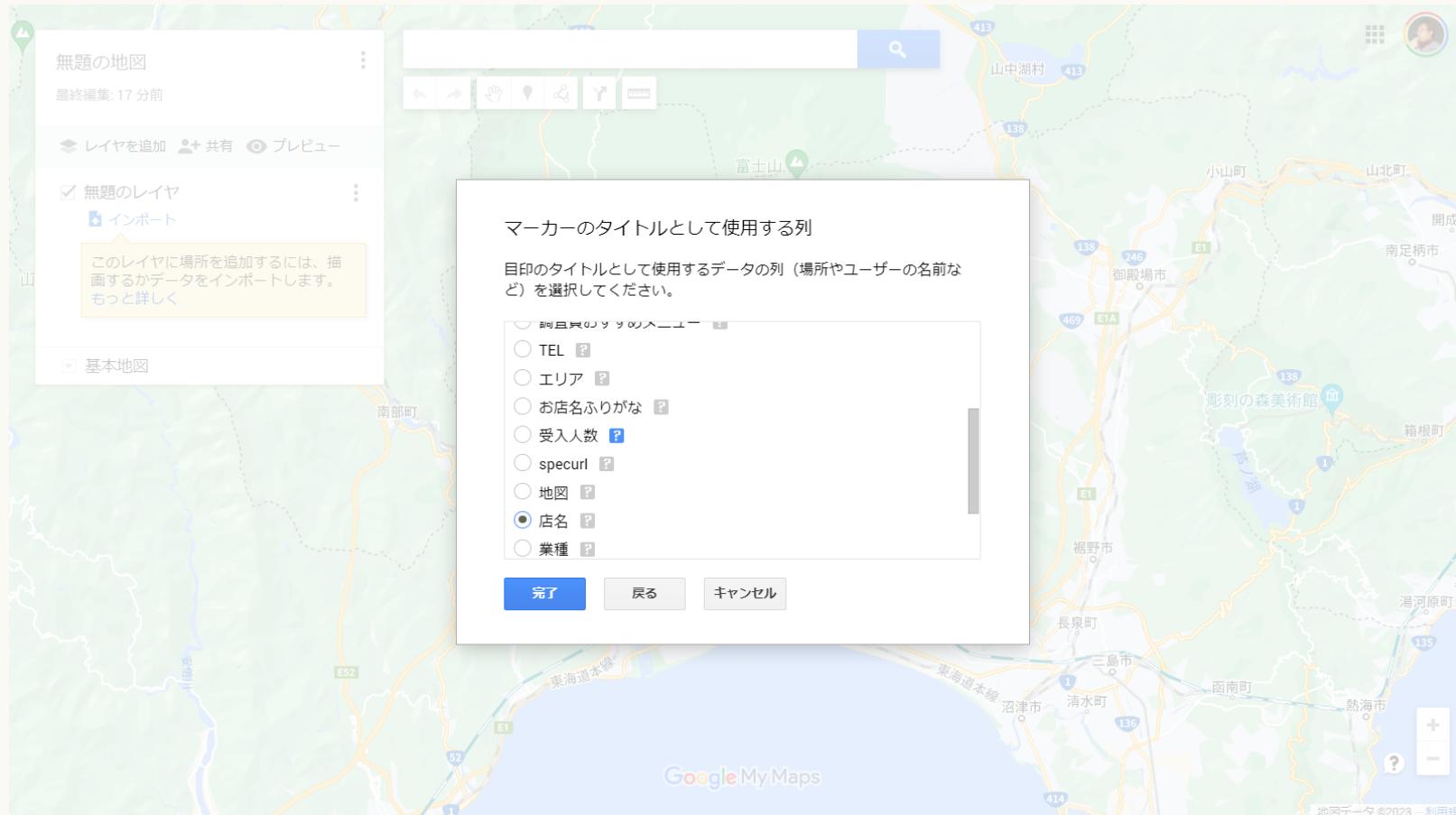


マップへのポイントは住所を使います

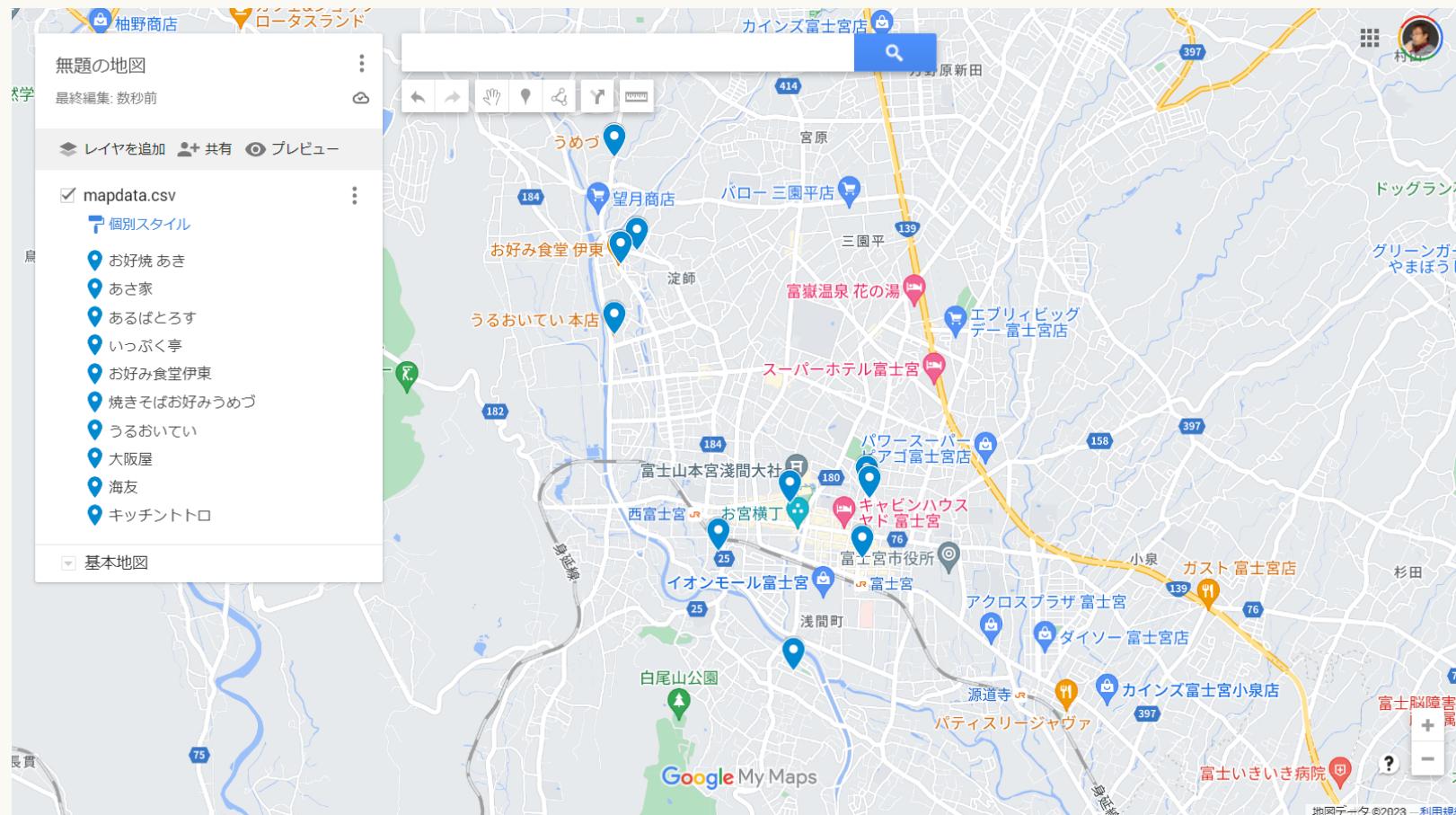


ポイントした部分への簡易説明（マーカー）を入れる

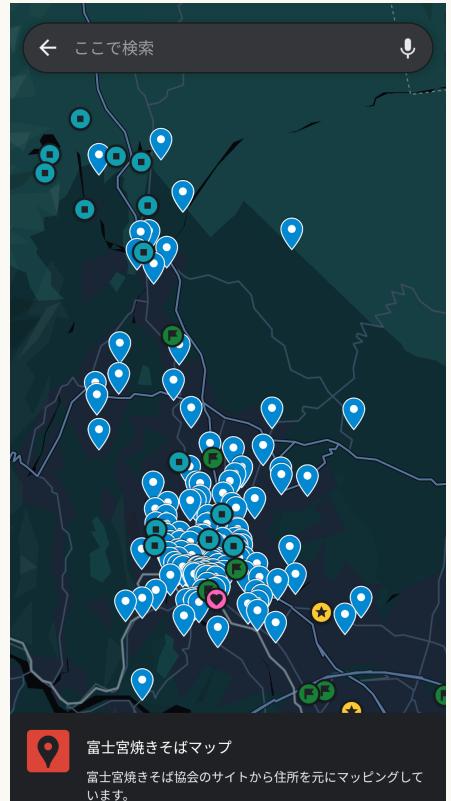
今回は店名にしました



マッピングされました!



スマホで見ることもできます！



そのほか選択肢

今回はGoogleマイマップを利用しましたが、推奨しているわけではなくてデータを作るいろんなサービスと連携できることをお伝えしました。

オリジナルのマップを作るサービスの一例です

- OpenStreetMap uMap: <https://umap.openstreetmap.fr/ja/>
- proxi: <https://www.proxi.co/>
- [日本向け]国土地理院: <https://maps.gsi.go.jp/>
- etc...

※: それぞれ特徴や無料有料があるので、使いやすいものを探すと良いと思います

トークのまとめ

1. どこからデータを取り出すか



2. どんなデータを作るか



3. どこへデータを渡すか

今回のトークでできること

1. お店情報をデータにする: WEBスクレイピング
2. データを整形をする: CSVファイルにする
3. データを使ってみる: Googleマイマップ

抽象的に言い直すと...



1. どこから データを取り出すか
2. どんな データを作るか
3. どこへ データを渡すか

プログラミングでもっとも行われる行為

- どこから データを取り出すか
- どんな データを作るか
- どこへ データを渡すか

プログラミングはデータの流れを意識しよう

- どこから データを取り出すか
 - データベースから取り出す → データベース接続、API
 - 現実の統計データを使う → データセット、オープンデータ
 - センサーデータで現実環境を扱う → ハードウェア操作
- どんな データを作るか
 - データを加工して新しいデータを作る → 前処理、後処理
 - データを整形する → ファイル形式、データ構造
- どこへ データを渡すか
 - データの可視化をして分析 → グラフ、ダッシュボード
 - データを使って業務工程改善 → オートメーション、システム連携
 - データを使ってAIで予測 → 機械学習、ディープラーニング

最後に伝えたいこと

PyCampやPythonの基礎を学んだ方の一歩先として。オススメです！

自分が使いたい、利用したい、ものやことがあればチャレンジしましょう！

できたら、とても楽しいし、感動します 😊

Happy Hacking!!

and, Have a nice trip!!

付録

今回扱わなかった他の方法については、またどこかで解説できたらと思います

- WEBスクレイピングしづらいサイトもカバーする
 - selenium + headless chrome
- 画像識別でお店情報を収集する
 - OCR, Googleなど
- 緯度経度を収集
 - 世界: Google?
 - 日本: 東大CSIS
- 印刷物を作る -> テンプレートエンジンで印刷しやすいHTMLを生成
- fletで自分専用のマップアプリを作る

WEBスクレイピング: Seleniumの例

- 動的な（javascriptなど利用した）サイトはrequestsで対応が難しいことがあります
- 本物のブラウザ + ブラウザ自動操作ツール(selenium)を使った例も載せておきます

-> url

画像識別

結構難しい分野（ちょっと自信ない）

- Google Cloud Vision
 - -> url

位置情報

- 世界: Google Maps Platform ->
<https://developers.google.com/maps/documentation/geocoding/overview?hl=ja>
 - -> url
- 日本: 東大CSIS -> jageocoder
<https://github.com/t-sagara/jageocoder>
 - 無料で利用できる（リクエスト数は少なめがよし
 - -> url

印刷用マップを作ってみる

印刷用のHTMLファイルを作って印刷してみる

- Mapboxで概要と詳細の地図を用意
- お店の情報をテーブルで埋め込む
- 印刷用にCSSで調整する

-> url

モバイル向けのWEBアプリ

実験的な扱いです。うまく動かなければ...

- fletを使って作ってみました
- staticなページでWASMとして埋め込んで作ってみる
 - <https://flet.dev/docs/guides/python/publishing-static-website/>
- ※サイト自体の公開はしません

-> url