

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



# 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

# Today's Document

本日の資料は公開されています

- GitHubリポジトリ: コードとスライド  
<https://github.com/hrsano645/pyconapac2023-local-food-map>
- スライド: Speaker Deck  
<URL載せ替えます>

後日トライしたい方は参考にしてみてください

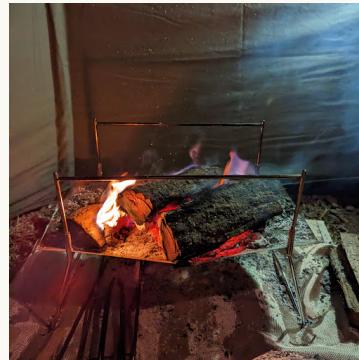
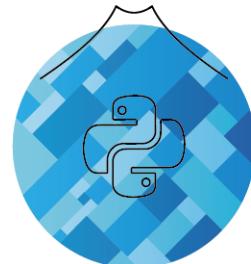
# Self Infroduction

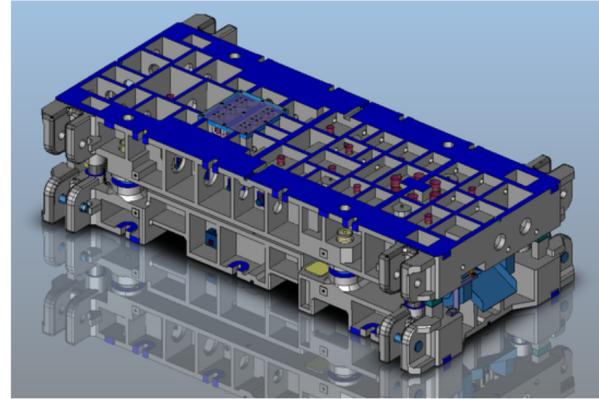
Hiroshi Sano(佐野浩士) [@hrs\\_sano645](#)

- 🗺️: 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 💡



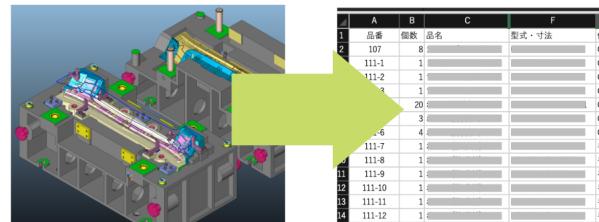
**PyCon  
mini  
SHIZUOKA**





## 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

# 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！** 食べに行きましょう！
  - ご当地グルメ=B級グルメのこと
- PyCampを終えた人に: **Pythonでデータを集めて作り利用するプロセス** を学べます
  - 初学者向けの内容です
  - 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
- 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



The screenshot shows the homepage of the "Python Boot Camp Text" documentation. At the top left is a logo featuring a stylized figure made of code symbols like '#', '=', '+', and '-' on a blue background, with the text "Python Boot Camp" below it. The main title "Python Boot Camp Text" is centered at the top. Below the title is a brief description: "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」で使用するために、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。" There are two links: "GitHub: <https://github.com/pyconjp/bootcamp-text/>" and "Read the Docs: <https://readthedocs.org/projects/bootcamp-text/>". A sidebar on the left contains navigation links: "Python Boot Camp Text ドキュメント", "Q. 検索", "Python Boot Camp 運営マニュアル", and "Python Boot Camp テキスト". The main content area is titled "目次" and lists several sections with their sub-sections. At the bottom, there's a "ライセンス" section with a Creative Commons Attribution 4.0 International License logo and the text "This work is licensed under a <a href='https://creativecommons.org/licenses/by/4.0/">Creative Commons Attribution 4.0 International License. v: latest".

このトークはその続きからトライできるコンテンツを目指して作りました

# PyCampの様子

静岡県3回目の様子

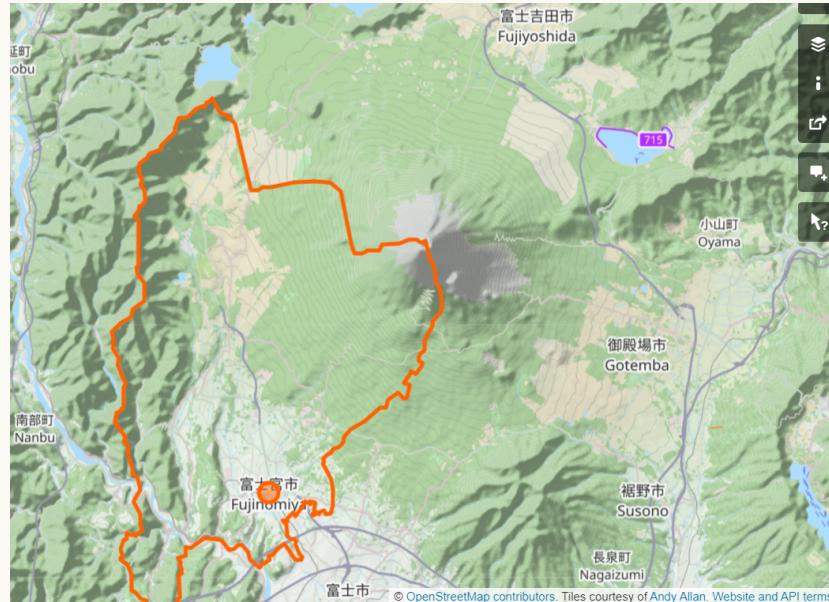


今回のお題: 本題のご当地グルメ

# 富士宮焼きそば



- 主に静岡の富士宮市周辺で食べられる焼きそば
- 麺は富士宮周辺でしか手に入らない。まさにローカルフード
- 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した



地元の人は、多分月に数回は食べてる

ところで、みなさん

もう食べたくなりましたね？ 😭

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

# 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



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] 📸 画像識別で加工を試みる

機械可読性はどちらも微妙

WEBスクレイピングはまだやりやすい

## ご当地グルメの情報はどこにあるか

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

観光情報を探ってみると…

- 市役所で紹介されたり
- 観光協会で紹介されたり
- 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。

<https://umya-yakisoba.com/shop/>

## 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。

この構造からBeautifulSoup4を使って必要な情報を取り出します。

要素 コンソール ソース »

```
<main>
  <!-- /#breadcrumb -->
  <header class="article-header entry-header">...</header>
  <div class="article">
    <div class="p-shopList">...</div>
      <a href="https://umya-yakisoba.com/shop/3776/">
        <figure class="p-shopThumb-ph-no">...</figure>
        <div class="p-shopThumb-text">== $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>
      <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>
    </div>
  </div>

```

要素 コンソール ソース »

```
ema.org/Blog"
  <div id="breadcrumb" class="breadcrumb breadcrumb-category sbp-main-top">...</div>
  <!-- /#breadcrumb -->
  <article>
    <header class="article-header entry-header">...</header>
    <div class="p-shopPh">...</div>
    <dl class="p-shopDetails">...</dl>
      <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>
    </dl>
  </article>

```

```
> <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" style="flex">...</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>
```

構造の中にあるタグから必要な情報を取得する

```
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")
```

```
... , "p-shopthumb-text"> = $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タグの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'] = replace_text(
        shopinfo_tag.find_all("div")[1].text
    )

    shopinfo_list.append(shopdata)
```

※ replace\_text関数:

店名に出てくる空白文字を置き換える独自関数  
(後述します)

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'},
 ...
]
```

## 収集した詳細URLのリストを使って、お店情報収集

The screenshot shows a browser window with the URL [www.fujiyaki-soba.com/fujiyaki/akihikoshiro](http://www.fujiyaki-soba.com/fujiyaki/akihikoshiro). The page title is "お好焼 あき". On the left, there's a sidebar with categories like "エリア", "まちなか", "お店名ふりがな", "あき", "地図", "B6", "業種", "飲食店", and "住所". The main content area displays a shop detail card with fields for address, phone number, business hours, etc. On the right, the browser's element inspector is open, showing the DOM structure. A red box highlights the `<dl class="p-shopDetails">` element, which contains `<dt>` and `<dd>` pairs. The `<dt>` tags include "エリア", "お店名ふりがな", "地図", "業種", "飲食店", and "住所". The `<dd>` tags provide the corresponding values: "あき", "あき", "地図", "B6", "飲食店", "住所", "富士宮市野中東町112-1", "TEL", "0544-27-0004", "営業時間", "火曜日", "受入人数", "12", "駐車場", and "4".

`dt` と `dd` タグは記述リスト、説明リストと呼ばれるHTMLタグ

```
> <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>
```

```
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構造でdtが項目、
    # ddが値になっている。これを辞書形式にする
    shopspecs = {}
    for dt, dd in zip(
        soup.find('dl', class_='p-shopDetails').find_all('dt'),
        soup.find('dl', class_='p-shopDetails').find_all('dd')
    ):
        # 値に 改行や空白文字があるので取り除く
        shopspecs[dt.text] = replace_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
 # ...]
```

## 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やめましょう  
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スクレイピングの注意点 ⚠

※スクレイピング対象のサイトへ短時間に多数のアクセスはしないように注意しましょう

- よくあるトラブル
  - ループ構造が深くてアクセス回数が膨れる
  - エラーの時にリトライし続ける
- 少し時間を置きながらアクセス
  - ランダム時間置いてみる
  - 回数リミットをつけて待つ
- サイトポリシーがあればそれに従う
  - 利用規約を見る
  - (クローラー向けの) robot.txtを見る

## 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))

# 2~5秒の間でランダムに待つ
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

マップのもとになるデータを作成します

-  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構造、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）

CSVライブラリを使って書き出せます

`csv.DictWriter` を使うと辞書のキーを使って列見出しを用意できる

```
with open('mapdata.csv', 'w', newline='') as csvfile:  
    # お店の詳細情報の各項目:辞書のキー が部分的にあったりなかったりしたので  
    # 全ての辞書のキーから全ての項目をカバーしたリストを生成する  
    csv_fieldnames = list(set().union(*shopinfo_list))  
  
    writer = csv.DictWriter(csvfile, fieldnames=csv_fieldnames)  
    writer.writeheader()  
    for shopinfo in shopinfo_list:  
        writer.writerow(shopinfo)
```

お店ごとに詳細情報の項目数が違う

```
>>> shopinfo_list[0].keys()
dict_keys(['TEL', 'specurl', 'お店名ふりがな', 'エリア', '住所', '受入人数', '営業時間',
'地図', '定休日', '店名', '料金目安', '業種', '焼き方',
'調査員おすすめメニュー', '調査員が見た特徴', '駐車場'])
>>> shopinfo_list[1].keys()
dict_keys(['TEL', 'specurl', 'お店名ふりがな', 'エリア', '住所', '受入人数', '営業時間',
'地図', '定休日', '店名', '料金目安', '業種', '焼き方'])
```

それぞれの項目をすべて合わせたリストを作る

```
>>> print(list(set().union(*shopinfo_list)))
['エリア', '駐車場', '焼き方', '受入人数', 'specurl', 'TEL', '店名', '定休日', '業種',
'調査員が見た特徴', '住所', '営業時間', 'お店名ふりがな', '地図', '料金目安', '調査員おすすめメニュー']
```

# 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, <a href="https://umya-yakisoba.com/shop/3776/">https://umya-yakisoba.com/shop/3776/</a> ,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, <a href="https://umya-yakisoba.com/shop/3777/">https://umya-yakisoba.com/shop/3777/</a> ,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, <a href="https://umya-yakisoba.com/shop/3774/">https://umya-yakisoba.com/shop/3774/</a> ,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, <a href="https://umya-yakisoba.com/shop/3616/">https://umya-yakisoba.com/shop/3616/</a> ,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,, <a href="https://tabelog.com/shizuoka/A2204/A220401/22002448/">https://tabelog.com/shizuoka/A2204/A220401/22002448/</a> ,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), <a href="https://umya-yakisoba.com/shop/3772/">https://umya-yakisoba.com/shop/3772/</a> ,A1,好み食堂伊東,飲食店,日曜・火曜,大きな鉄板でアツアツをどうぞ,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

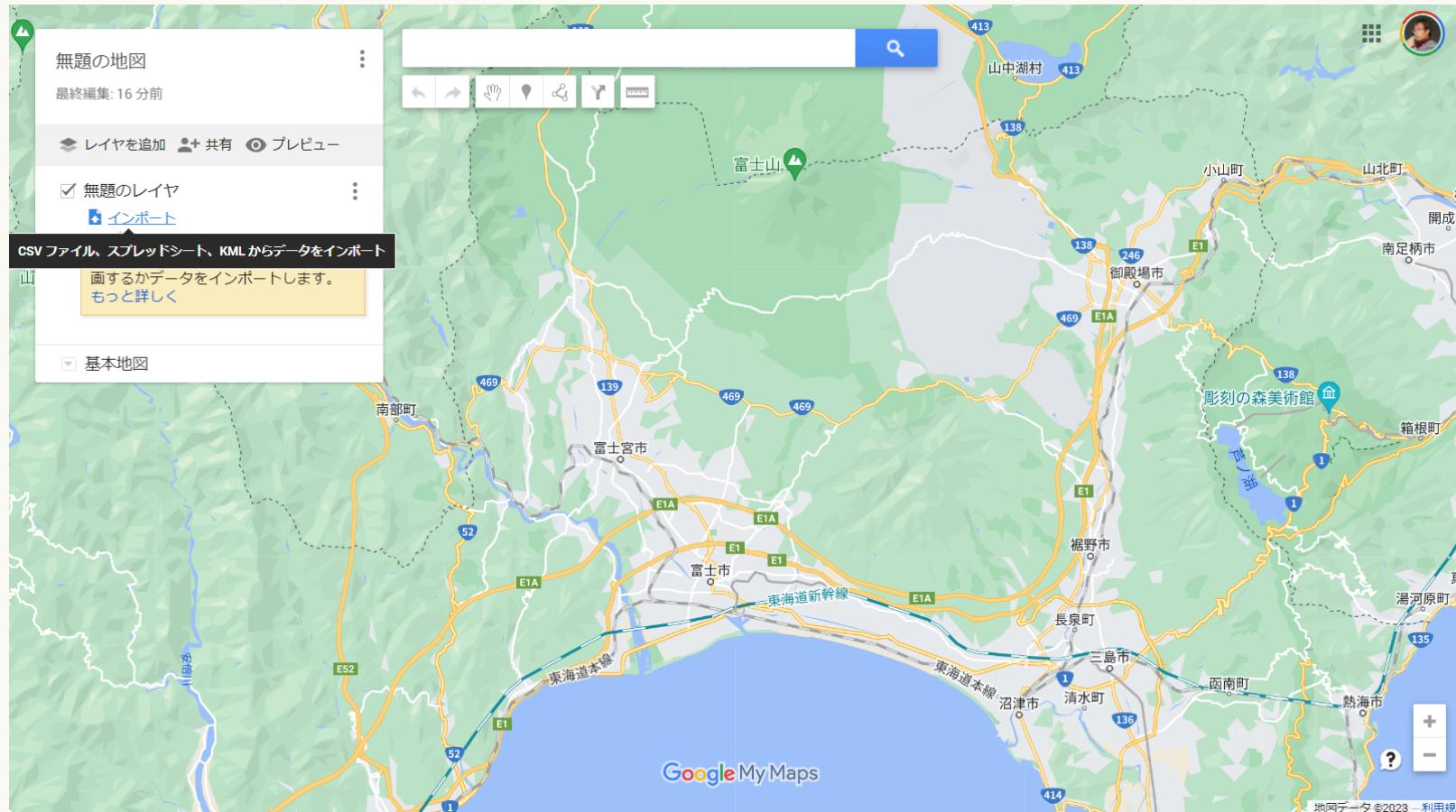
# 扱い方

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

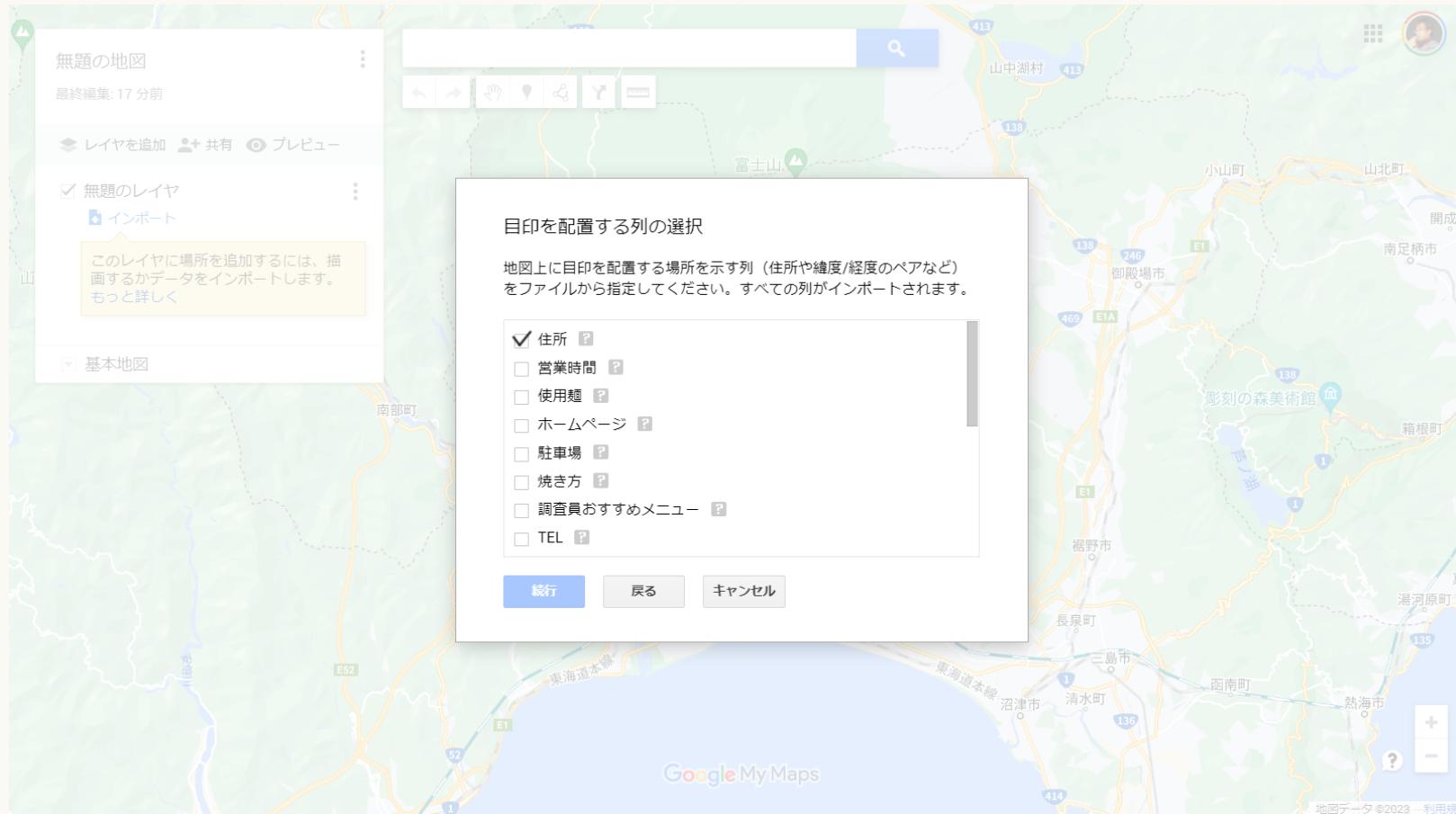
「新しい地図を作成」ボタンを押す



新規レイヤーへCSVファイルをインポートします

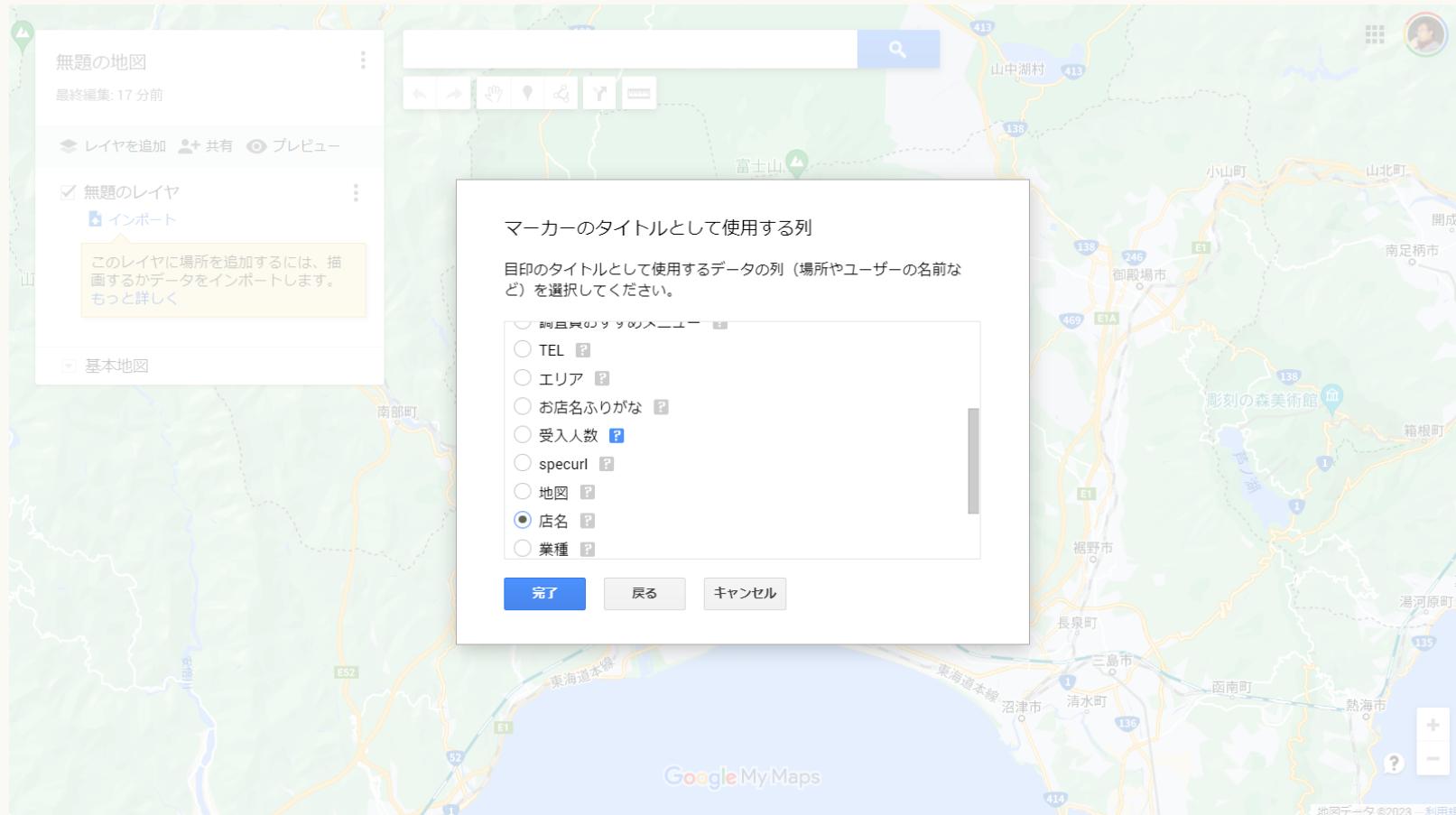


## マップへのポイントは住所を使います

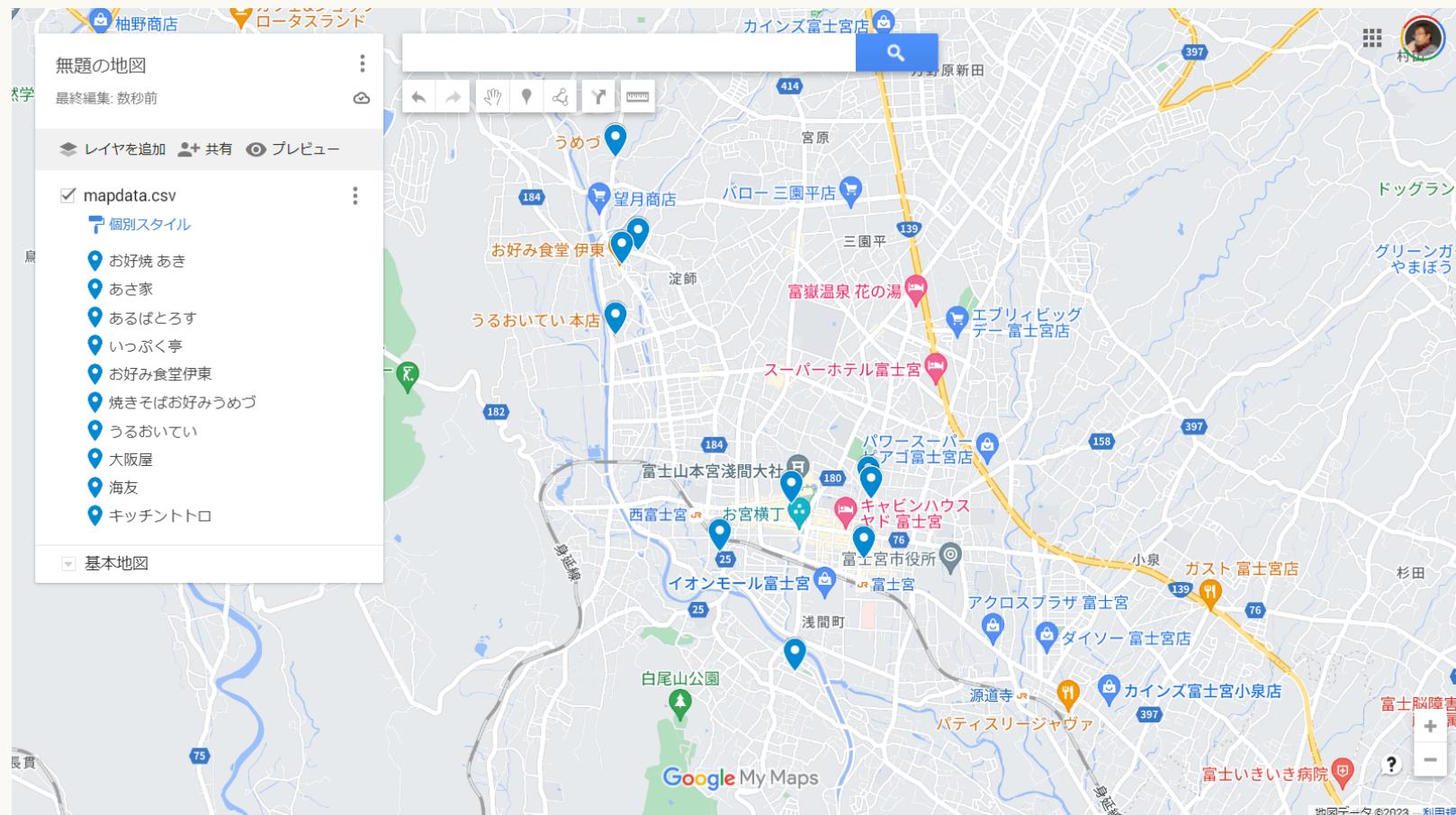


ポイントした部分への簡易説明（マーカー）を入れる。

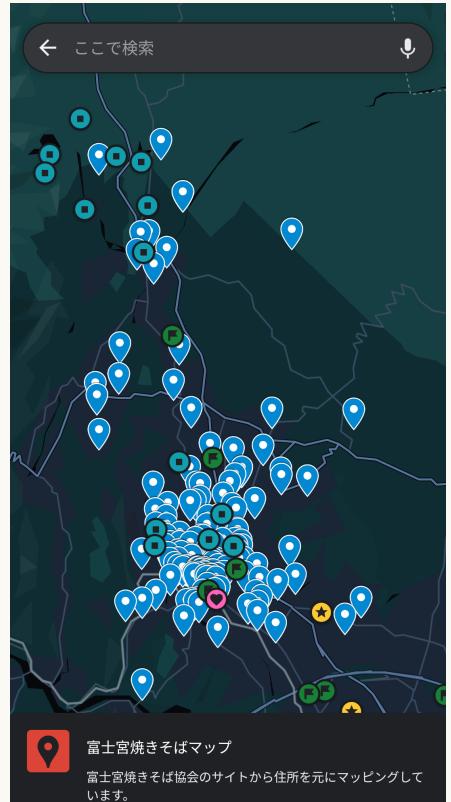
今回は店名にしました。



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は他にも多数あります。一例を載せます
- ※: 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
- 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...

# 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



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

- **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
  - 現実の統計データを使う
  - センサーデータで現実環境を扱う
- **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
  - データを整形する
- **どこへ** データを渡すか
  - データの可視化をして分析
  - データを使って業務工程改善
  - データを使ってAIで予測

- 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！

自分が使いたい、利用したい、ものやことでトライしましょう

それができれば、とても楽しいし感動します。オススメです 😊

**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