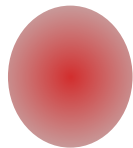
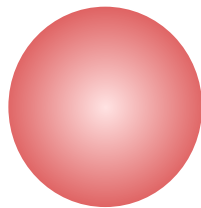
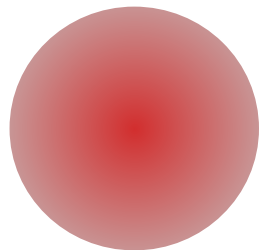
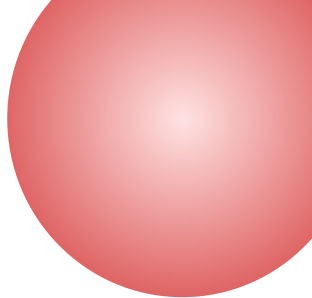




# データ分析を

# 生成AIとPythonで！

合同会社 長目  
Founder & CEO  
小川 英幸



# 自己紹介

合同会社長目（ちょうもく）



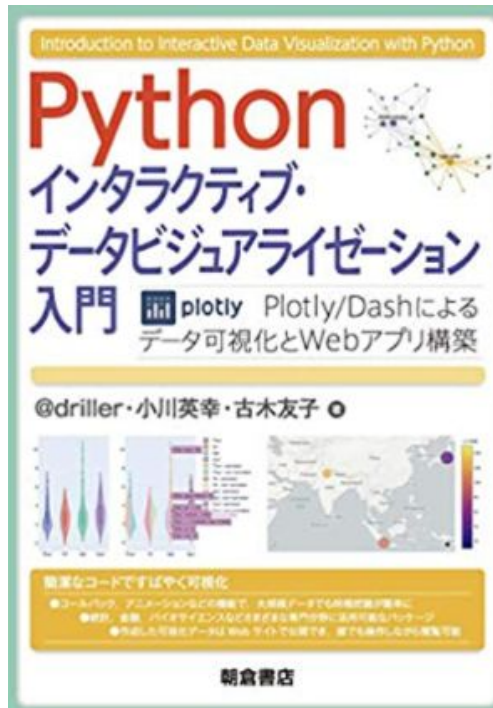
- ▷ 設立 : 2016年6月1日
- ▷ 社員 : 2名
- ▷ 事業内容 : データ分析受託・アプリ作成
- ▷ 採択: 第4期浜松市実証実験サポート事業 オー  
プンデータの利活用 (2022/10-2023/9)
- ▷ ウェブサイト: <https://www.chomoku.info>

代表社員 小川 英幸



- ▷ 証券会社でトレーダー、アナリストとして18年間働いたのち、長目を設立
- ▷ プログラミング書籍の執筆が2冊（日経BP、朝倉書店より）
- ▷ 趣味: 筋トレ、銭湯、歳なので何事も新しいものを好むようにしている。

会社の経営をしていて、九州でのビジネスチャンスにも興味があります。京都からは熊本がすごいということだけ聞こえてきていて、それにすごい興味があるのですが、それ以外もいろいろ教えていただけると嬉しいです！（地理を全然分かってなくて鹿児島から熊本へは30分くらいで行けるとあって申し込んだ）



小川英幸 OGAWA HIDEYUKI

# Python

## データ分析

### ハンズオンセミナー

日経BP

前処理 可視化 詳細分析

こうすれば  
ビジネスに  
役に立つ!

リアルな統計データで  
分析の始めから完了までを実体験

# Pythonデータ分析ハンズオンセミナー

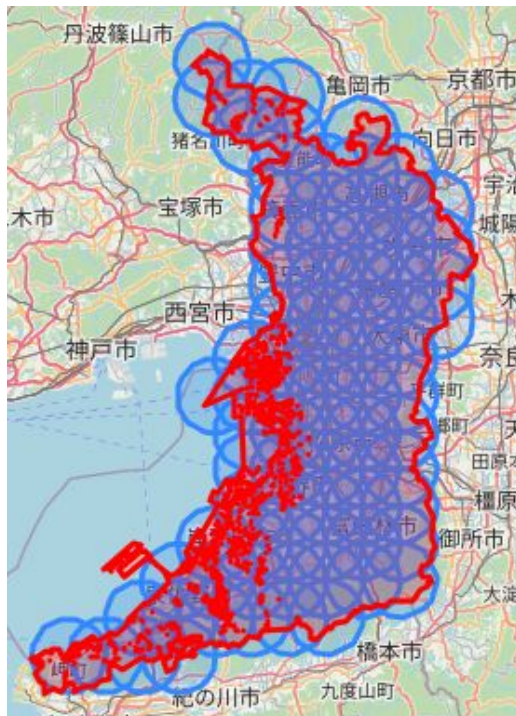
飛耳長目

chomoku.info



書籍のコードはすべてgithubに

<https://github.com/mazarimono/python-data-seminar>



実際の国勢調査のデータを使って解説。一個のミスが命取りになる、筆者ヒヤヒヤのチャレンジング書籍。商圈を作ってその人口のデータを自前で作れる。

データ分析についての解説も力を入れた。

# 本日のトークのモチベーション

- 人がネックになっていたデータ分析ビジネス
- そのネックがなくなり、スケールしやすくなるというアイデア
- 実際どんな感じになってきているかをアイデアなどを紹介。皆さんと話し合って色々と思いを先に進めたい

- 生成AIを活用すべき理由
- データ分析について
- 国勢調査のデータを使って色々妄想
- まとめ

# なぜ生成AIを活用すべきか？

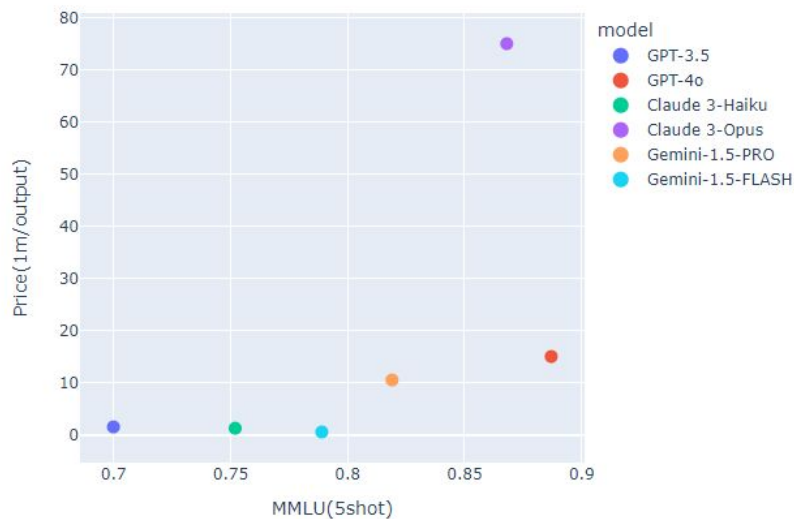
- デジタル、データ活用の人材不足
- より生成AIが賢く、APIが安くなっている現状

実際にコードを動かしながらご紹介！  
(超緊張する・・・)

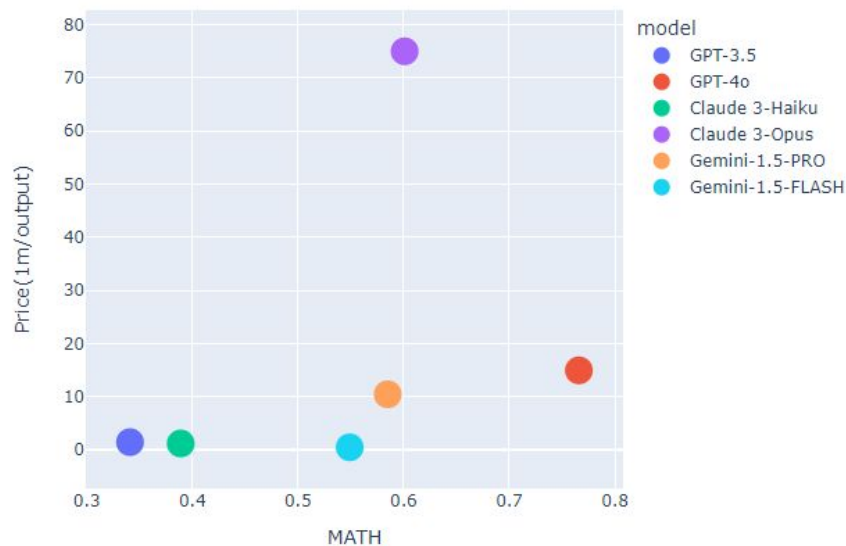


# なぜ生成AIを活用すべきか？

MMLU(5shot)とPrice(1m/output)の関係



MATHとPrice(1m/output)の関係



CODE:

[https://colab.research.google.com/drive/1R047AOGv4mBwBUwhr\\_guRImplqtr5sdU?usp=sharing](https://colab.research.google.com/drive/1R047AOGv4mBwBUwhr_guRImplqtr5sdU?usp=sharing) データは各社ウェブページから長目が作成

# プログラミング業界は活用が速い

- Github Copilot
- Colab AI
- などなど

今回のトークのモチベーション:

データ分析の生成AI活用の可能性を、国勢調査のデータ（鹿児島）を使いながら、将来の方向を検討したい。どうせなら**楽しく働ける方向で発展させたい**。

私がアプリケーション開発、利用していて思ったところを話す、色々と**意見交換させていただけると幸いです**。

- 生成AIを活用すべき理由
- データ分析とは
- 国勢調査のデータを使って色々妄想
- まとめ

課題解決の場面で、自社に適した意思決定を行う確率を上げる

- アパホテル土地選び
  - 夜間人口よりも昼間人口が多いところでホテルを作る
  - 出張需要・観光需要などが取り込める（推測）
- 楽待 【アパの戦略】 アパホテル流、土地選びの秘訣とは？
- <https://www.youtube.com/watch?v=3Ypve9dp-J4>

確実に将来が見える水晶玉ではない



## データ分析の限界

未来は  
常に不確実

アップサイドリスクと  
ダウンサイドリスクを  
観察し続ける

## 企業全体で向上の必要

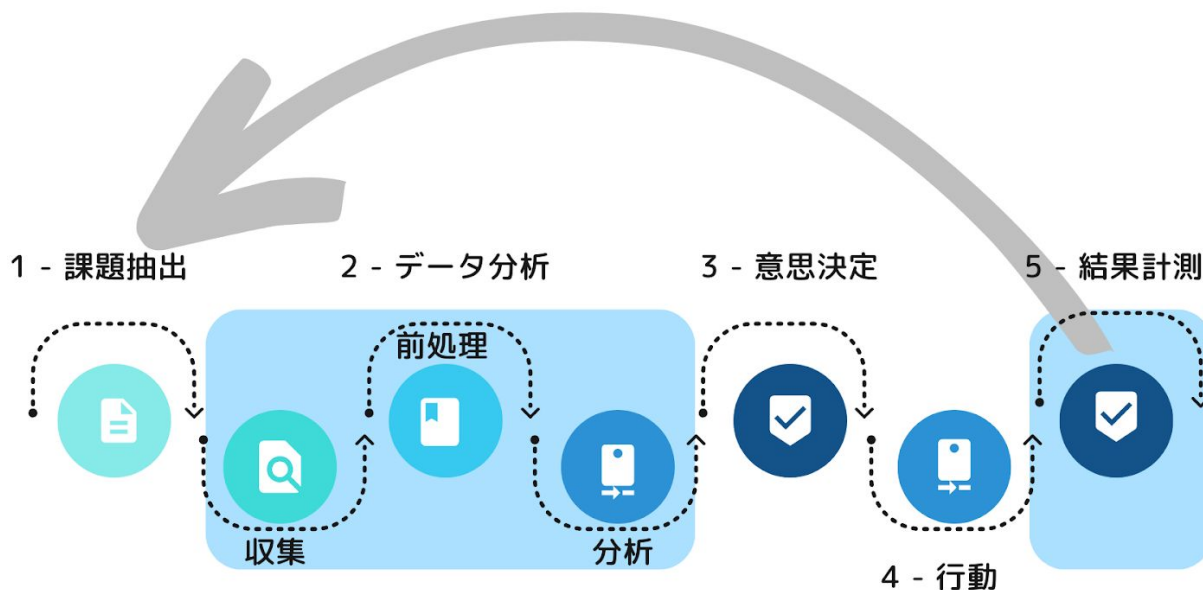
データ収集力

データ分析力

データ分析脳

# データ分析のステップ

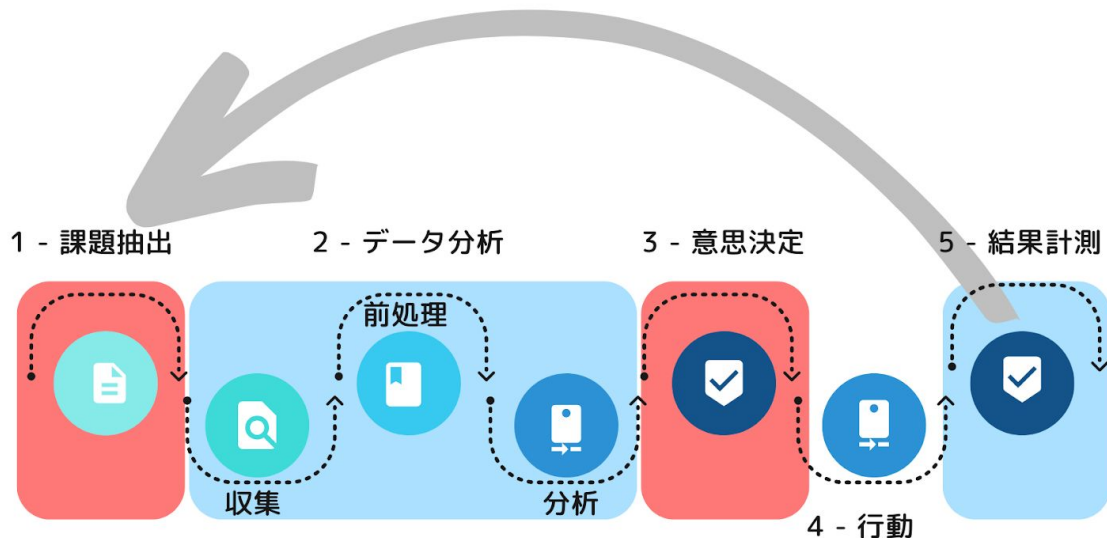
## 今までのデータ分析の流れ



Pythonデータ分析ハンズオンセミナー 小川英幸p17より

今まで：主な企業では、データサイエンティストが役立つ  
場面は青い部分だった。  
全体的なイテレーション+部分的なイテレーション

## これからのデータ分析の流れ



Pythonデータ分析ハンズオンセミナー 小川英幸p17更新

赤い部分もデータサイエンティストが役立つ  
色のついた工程、LLMと人間の協業で効率アップ？



これから: 赤の部分をLLMを活用することにより効率的になりそう。元の青い部分もLLMと人間の協業により、精度なども上がる可能性。

- 例:
  - 非構造化データの活用(課題抽出、意思決定)
  - 思考・アイデアを深める (課題抽出)
  - 意思決定: よりロジカルに決定できる可能性を秘める
- データ分析のステップもかなり変わる
  - LLMがアイデアを基にコードを書いてくれる
  - アイデアをLLMと人間が作る
  - 一方で、LLMに難しい作業もある

- 人間の強みは行動？
  - ここから得た知識を全体にフィードバック？
  - 逆に意思決定に関するデータは揃いやすくなる可能性もある
- しっかりとした意思決定が本当に役立つのかは、よくわからない部分もあるだろう
  - 人間的に堅物よりも、ちょっと融通が利く方が何事もうまくいったりする
  - 情けは人の為ならず / ゲーム理論的な思考を組み込むのは難しそう
  - 意思決定の種類によるかも: 定量的にどっちというのは得意そう (AI)
- データを整理して上手く使えるようにすることが重要ってのは変わらなさそう
  - LLM活用しかり

- 生成AIを活用すべき理由
- データ分析について
- 国勢調査のデータを使ってデータ分析ステップの進め方を観察
- まとめ

- 5年に一回日本に住んでいるすべての人と世帯を対象する、国のもっとも重要な統計調査。
- 年齢別の人口、家族構成、働いている人や日本に住んでいる外国人人数などの数を提供
- e-Stat: 国勢調査の説明より
  - <https://www.e-stat.go.jp/gis/statmap-search?type=1>
  - 統計地理情報システム: 位置情報と統計情報がある
- 今回は1kmメッシュのデータを使って1995年と2020年の人口増減を調査する

- 主に分析ステップだけの事例ですが
- Pythonデータ分析ハンズオンセミナーのリポジトリ参考
  - <https://github.com/mazarimono/python-data-seminar>
  - 順番にコードを作りながら、進めていく
  - 全部のコードを人間が色々調べながら作っていく

- 本もそうでしたが、登壇でも、実際のデータを使って何についてデータ分析をするかというのは、頭の痛いところ
- **課題の明確化、深堀、アイデアの拡張に生成AIが役立つ**
- Q: データ分析には課題が必要ですが、どのようなものが良いでしょうか？鹿児島県の特徴を踏まえてアイデアを出してください。
- A:
  - 人口減少と高齢化対策の立案
  - 離島振興策の効果検証
  - 農林水産業の担い手確保策の立案
  - 観光振興のためのターゲット分析
  - 地域コミュニティの維持・活性化策の提案

- 本もそうでしたのですが、登壇でも、実際のデータを使って何についてデータ分析をするかというのは、頭の痛いところ
- **課題の明確化、深堀、アイデアの拡張に生成AIが役立つ**
- Q: データ分析には課題が必要ですが、どのようなものが良いでしょうか？鹿児島県の特徴を踏まえてアイデアを出してください。
- A:
  - **人口減少と高齢化対策の立案**
  - 離島振興策の効果検証
  - 農林水産業の担い手確保策の立案
  - 観光振興のためのターゲット分析
  - 地域コミュニティの維持・活性化策の提案

- 本もそうでしたのですが、登壇でも、実際のデータを使って何についてデータ分析をするかというのは、頭の痛いところ
- **課題の明確化、深堀、アイデアの拡張に生成AIが役立つ**
- Q: データ分析には課題が必要ですが、どのようなものが良いでしょうか？鹿児島県の特徴を踏まえてアイデアを出してください。
- A:
  - **人口減少と高齢化がどの地域で起きているか？を明確に**
  - 離島振興策の効果検証
  - 農林水産業の担い手確保策の立案
  - 観光振興のためのターゲット分析
  - 地域コミュニティの維持・活性化策の提案



- Colab AI を使う
  - Google Colaboratory についているコード入力補助ツール
  - Colab AI: 自然言語での対話でコードを作ってくれる
  - Github Copilot みたいな感じ
- 指示を出すとColab AIが働いてくれる
  - 対話的にコードだけを書いてくれる
  - 指示が難しいことも多い
  - 躓くところは、プログラミングと同じようなところだったりする。
    - header行の指定
    - csvの読み込みのencodeの指定などなど

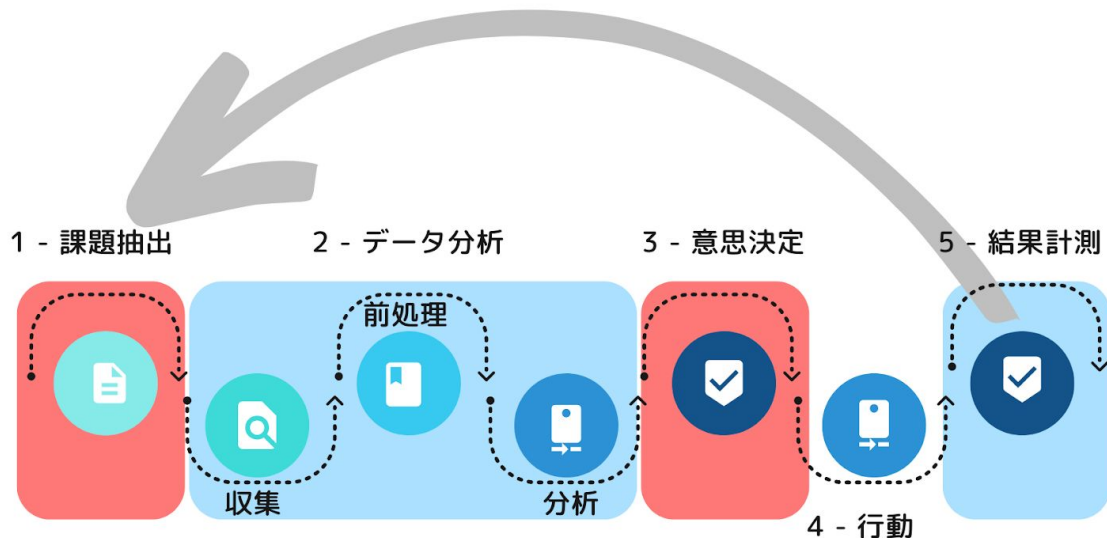
notebook: <https://colab.research.google.com/drive/166RV1kGh4dgO7-Vfmie554aMQG9T8qNO?usp=sharing>

- Agentを作る
  - 先の話にあったpandasaiもよい。が、自作したい欲
  - お二人のコードを参考に作ってみた（対話できる+）
    - 生ビールさん:  
[https://twitter.com/wmoto\\_ai/status/1769296451920937240](https://twitter.com/wmoto_ai/status/1769296451920937240)
    - からあげさん:  
<https://zenn.dev/karaage0703/articles/f6a1df0b2eabf4>
- 色々お手本はある
  - pandasai: <https://github.com/Sinaptik-AI/pandas-ai>
  - open interpreter: <https://github.com/OpenInterpreter/open-interpreter>
  - crewAI: <https://github.com/joaomdmoura/crewAI>
  - Microsoft Autogen: <https://www.microsoft.com/en-us/research/project/autogen/>

- 最初のLLMと価格、性能などのデータをLLM同士(上司とデータ分析官)で話し合わせたノート
  - なんか結論を急ごうとする
  - もっと興味深くするためには複雑なデータがいいのかもしれない
  - プロンプトの工夫も必要？
  - GPT-3.5-turbo:  
<https://colab.research.google.com/drive/19SEBU4EUkoU1oW2xAF3TxOdRDSBwarM7?usp=sharing>
  - GPT-4o:  
<https://colab.research.google.com/drive/1UbH2PDNNNfG6jdeIVszeE56nopBpiJQx?usp=sharing>

# データ分析: 将来像 (Multi) Agentとの作業

これからのデータ分析の流れ



Pythonデータ分析ハンズオンセミナー 小川英幸p17更新

どのようにAgentを配備するかは興味深いところ

- 生成AIを活用すべき理由
- データ分析について
- 国勢調査のデータを使って色々妄想
- まとめ

- 生成AIにより、作業効率はより上がりそう
- 何でもよいから、使い始めてみよう
- 仕事で使うところに迷うなら、まずはデータ分析に使ってみては？（ChatGPTと一緒にがおすすめかも）
- 人間が不得意な部分を、生成AIとの協業で乗り越える
- Multi Agentをマネージする未来はすぐそこにあるのでは？

# まとめ（ビジネス観点）

- データ分析は人が必要だった
- それが少人数でこなせるため、サービスがスケールしやすくなる
- データドリブンなアプリでスケールするビッグチャンス
- データ分析系の人にはものすごいチャンスが広がっているように思う（ので私も頑張ろうと思っています。）
  - モックは素早く作れる（streamlit / Gradio / Dash）
  - AIへの理解が深い
- ビジネスのヒント！
  - Andrew Ng: Opportunities AI - 2023
  - <https://www.youtube.com/watch?v=5p248yoa3oE>

# 生成AIを活用して

# 楽しい未来を創造しよう



ありがとうございました



X: @OgawaHideyuki