

Pythonで見る私の行動変容 –COVID-19前後の比較–

4th July 2020, Python Charity Talks in Japan
Yuta Kanzawa @yutakanzawa

Data Science Senior Analyst at Janssen Pharmaceutical K.K., Tokyo
A Family Company of Johnson & Johnson



I am...

- 神沢雄大 **Yuta Kanzawa** (twitter: [@yutakanzawa](https://twitter.com/@yutakanzawa))
- Data scientist at **Janssen Japan**, Tokyo
 - A pharmaceutical company of **J&J**
- Opera & wine lover
 - Wagner
 - Bourgogne
- 7 languages
 - Human: Japanese, English, German
 - Computer: R, Python, SAS, SQL



アジェンダ

- 今日話すこと
 - 行動ログデータ
 - データと分析
- 今日話さないこと
 - COVID-19
 - R

行動ログデータ

When, where

Swarm (4sq): 行動ログ記録サービス (アプリ) *1

- 「チェックイン」をユーザーが**自分で**記録
 - スマートフォン
 - 要アカウント
 - GPS
- 公式**API***2からデータを取得可能
 - 本来はアプリ開発者向け
 - Pythonライブラリー : foursquare*3
 - 非公式



*1 <https://www.swarmapp.com/>

*2 <https://developer.foursquare.com/docs/api-reference/users/checkins/>

*3 <https://pypi.org/project/foursquare/>

データ

- APIの返り値 →
 - 入れ子のJSON
- pandasのデータフレームに変換
 - 分析に使う項目のみ
 - 全期間：22,000行
 - 分析用：4,300行
 - 2017-20年の1-6月

```
{'id': '56f3714c4be3f1902800a11d',  
  'createdAt': 1234567890,  
  'type': 'checkin',  
  'timeZoneOffset': 540,  
  'venue': {'id': '5164e907e4b0e029c7c9cf23',  
            'name': 'BeProud Inc. (株式会社ビープライド)',  
            'location': {'address': '千駄ヶ谷5-32-7',  
                          'crossStreet': 'NOF南新宿ビル4F',  
                          'lat': 35.68653340974678,  
                          'lng': 139.70382938037164,  
                          'labeledLatLngs': [{'label': 'display',  
                                                'lat': 35.68653340974678,  
                                                'lng': 139.70382938037164}],  
                          'postalCode': '151-0051',  
                          'cc': 'JP',  
                          'city': 'Tokyo',  
                          'state': 'Tokyo',  
                          'country': 'Japan',  
                          'formattedAddress': ['千駄ヶ谷5-32-7 (NOF南新宿ビル4F)',  
                                               '渋谷区, 東京都',  
                                               '151-0051']},  
            'categories': [{'id': '4bf58dd8d48988d125941735',  
                             'name': 'Tech Startup',  
                             'pluralName': 'Tech Startups',  
                             'shortName': 'Tech Startup',  
                             'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/shops/technology_',  
                                       'suffix': '.png'},  
                             'primary': True}],  
            'likes': {'count': 0,  
                      'groups': []},  
            'like': False,  
            'isMayor': False,  
            'photos': {'count': 0,  
                       'items': []},  
            'posts': {'count': 0,  
                     'textCount': 0},  
            'comments': {'count': 0},  
            'source': {'name': 'Swarm for iOS',  
                      'url': 'https://www.swarmapp.com'}}},
```

使ったライブラリー（主なもの）

- データ取得、加工
 - requests
 - foursquare
 - json
 - pandas
- データ分析
 - pandas
 - 集計、可視化
 - geopy
 - 緯度、経度からの距離の計算
 - matplotlib
 - 可視化

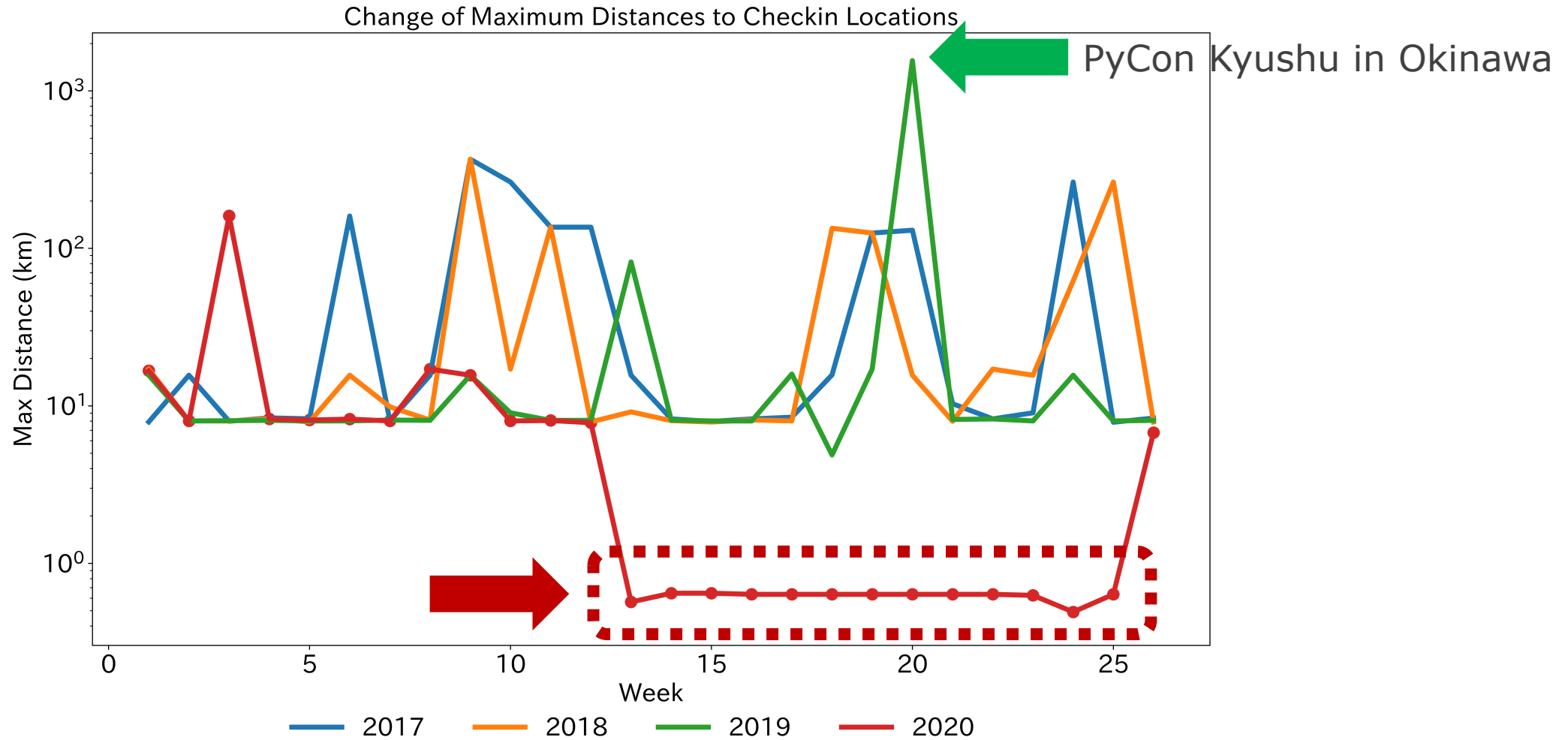
*1 <https://www.swarmapp.com/>

*2 <https://pypi.org/project/foursquare/>

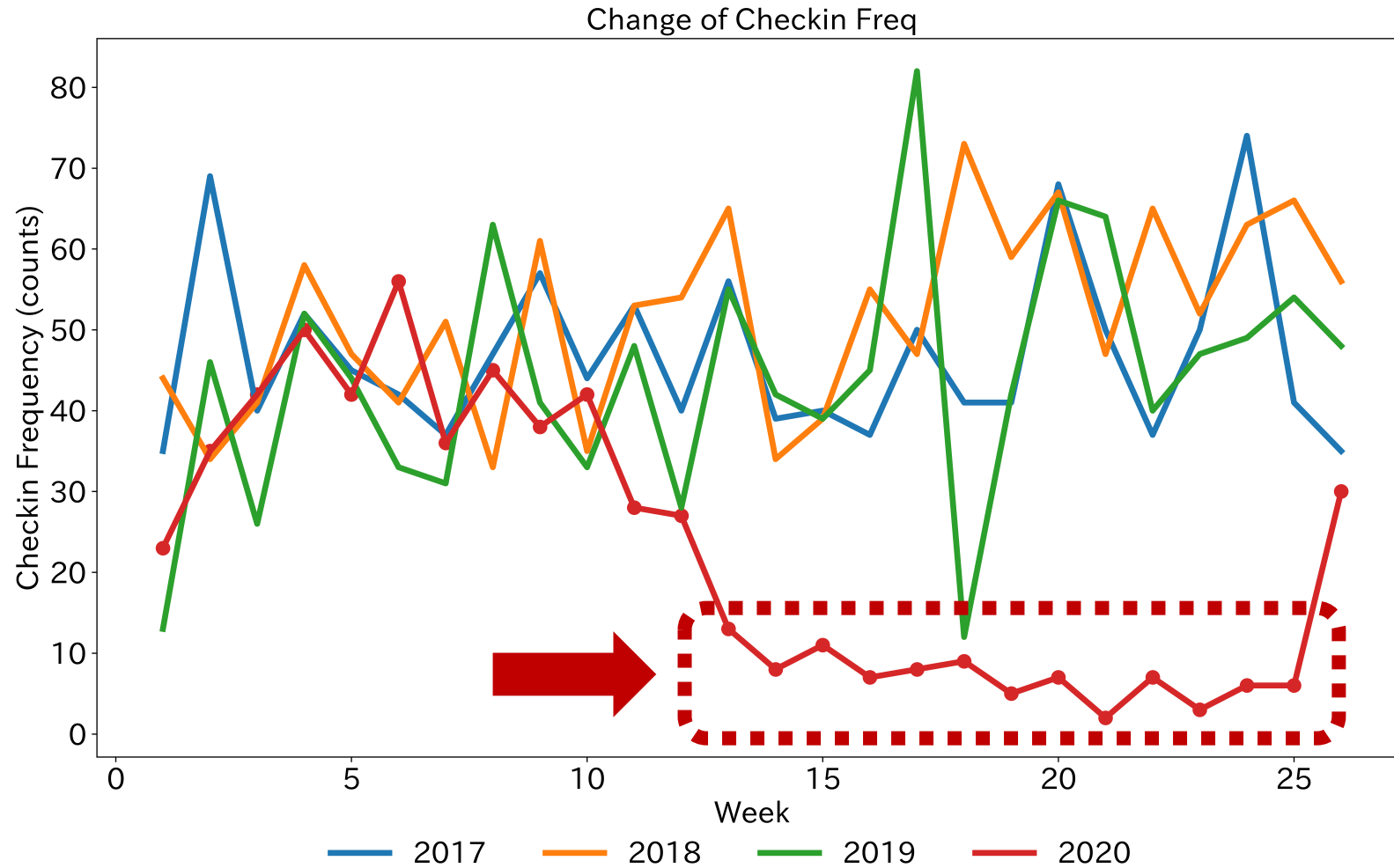
分析（結果）

Most fun part

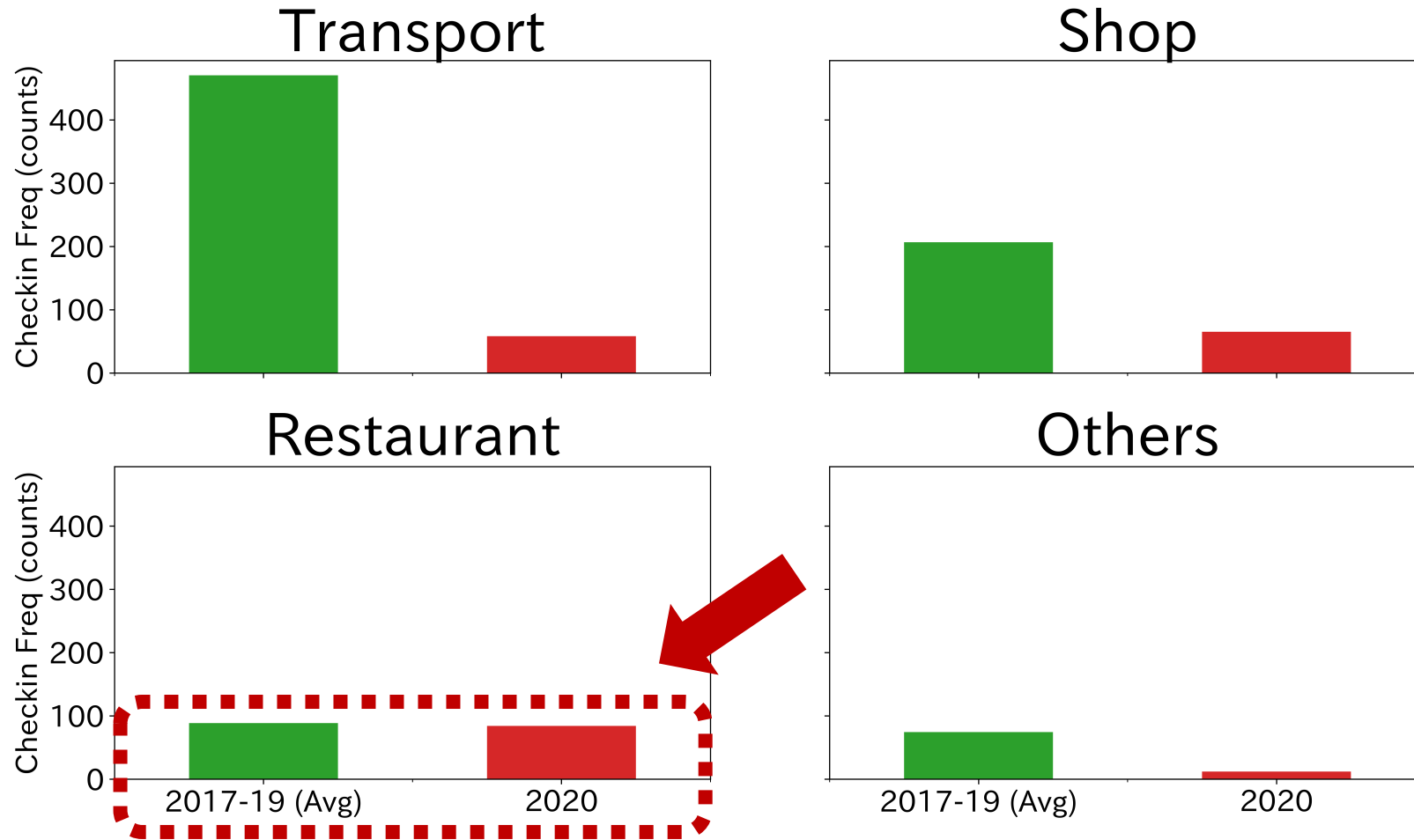
3月下旬から6月中旬まで：自宅から1km圏内を維持



チェックイン回数 ~ 1日1回以下



テイクアウトを活用して近所の飲食店を応援



* 3月下旬から6月下旬（各年の第10-26週）について算出。

ちょっとマジメな話

- こんな分析ができるのは、データがあるから！
 - データがないと始まらない。

- でも、データがあるだけでは始まらない！
 - 目的（今回は後付け...）
 - 品質
 - 環境（プラットフォーム）

まとめ

Long story short

Long story short

- 3月以降引きこもっていた自宅で人の命を守っていたことが示せた。
 - オンライン勉強会
 - テイクアウトや出前は大体どれも◎
 - 近所の飲食店を開拓
- データがあって初めて分析できる。
 - 一方、目的、品質、環境が伴わない → 無価値なデータ

Enjoy!