

# 第 3 回 【データ活用 入門】

担当: 西川・豊原・勝又・好田

## 【目的】

- ビジネスにおけるデータ利用について知る
  - データ可視化・データ分析の重要性を知る
- ハンズオンを通じてデータ可視化 / データ分析の流れを体験する
  - SQLを用いたデータ可視化
  - SQLを用いたデータ分析

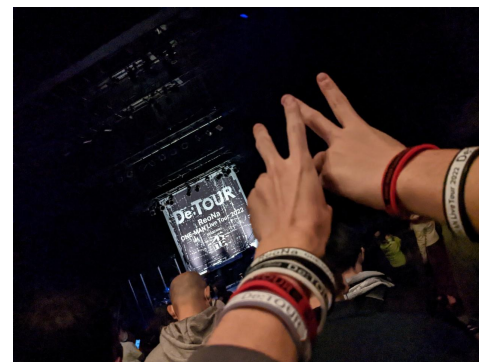
1. Introduction
2. データの可視化の重要性について
3. 可視化の手法について
4. KPIとは
5. 本日扱うデータについて
6. 演習
7. Q&A

## 担当講師紹介

- 所属: 理工学部 情報工学科 3年
  - 興味: AR/VR におけるインタラクション
- 担当回: 第3回
- 趣味: ライブ、野球観戦
- アルバイト:
  - データエンジニア、データサイエンティスト
  - サーバーサイドエンジニア



↑ slack のアイコン



# データの可視化(Data Visualization) ってなんで大事なの？

## Q. そもそもデータ活用とは？

A. データを蓄積することのみに留まらず、データに基づいた意思決定やシステム構築を行うこと

## Q. データ活用の目的は？

A. ビジネスにおいて競争力を維持・向上させるため

## Q. 例えば？

A. 一般的に行われているものは以下

- ・ 小売店が過去の売上を参考に仕入れ量を調節する
- ・ 地域の人流データから店舗の出店場所を決める

## Q. データ活用が最近になって活発化してきた理由は？

A. 情報通信技術 (ICT) の発展・ビジネスへの普及

- ・ 高速かつ大容量な通信
- ・ 大規模な計算を可能にする演算性能

慶應義塾大学 AI・高度プログラミングコンソーシアム

データマート

分析結果

用途ごとのデータの  
保管



<https://www.nicpartners.co.jp/report/92269/>

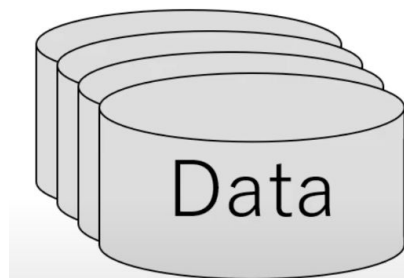


数値などのデータをグラフ・図・チャート・イラストといった形で  
「見える化」し、情報の理解促進や、**新しい法則・傾向の発見**を促すこと

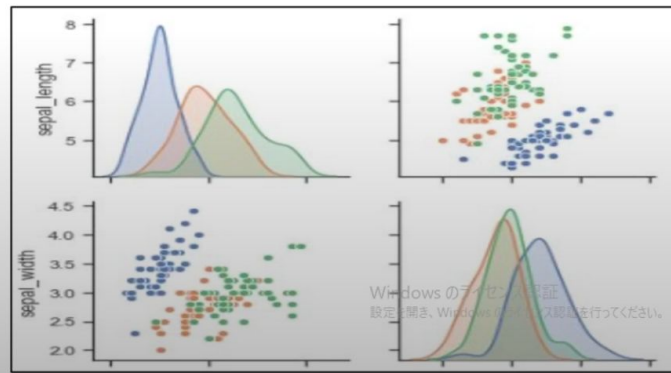
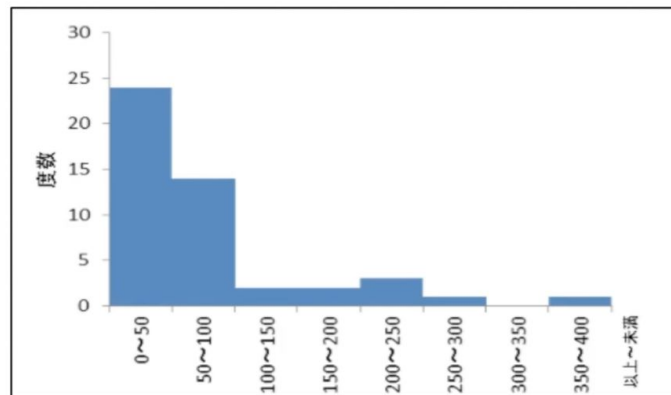
# このデータから何を読み取れますか、、？

登録日	ユーザーID	名前	性別	年齢	利用金額	誕生日
entrydate	userid	name	seibetsu	age	totalmoney	birthday
2021/01/06	KJ20452	ルカム	women	60	79000	1960/12/22
2020/07/27	OE93154	ケリウ	women	48	31000	1972/04/22
2020/04/12	WK51817	チネヌ	men	51	89000	1969/06/27
2020/10/01	DY41953	キヒラモクソ	women	45	72000	1975/10/29
2020/05/22	FK97065	メニテ	women	41	43000	1980/01/06
2020/05/04	DX90804	モシミタ	women	44	55000	1976/01/25
2020/10/31	ZP28414	カキアン	men	71	85000	1949/09/17
2020/03/14	XV40109	ケム	men	43	41000	1977/03/04
2021/01/09	ZO15902	ハマチオヒ	women	71	46000	1949/08/26
2020/02/29	CV64324	アケレ	men	73	42000	1947/11/21
2020/05/16	YV58267	ヒュシツヤ	men	55	23000	1965/02/09
2020/04/14	MO25293	ヘキスナスト	women	24	55000	1996/08/25
2020/02/09	FE87780	ヲイ	women	48	23000	1972/07/12
2020/02/27	FM28947	サアノセフ	men	79	76000	1941/07/22
2020/03/20	ID73056	サアオキ	men	31	62000	1989/04/23

# 人間は文字列や数字データに弱い、



可視化！



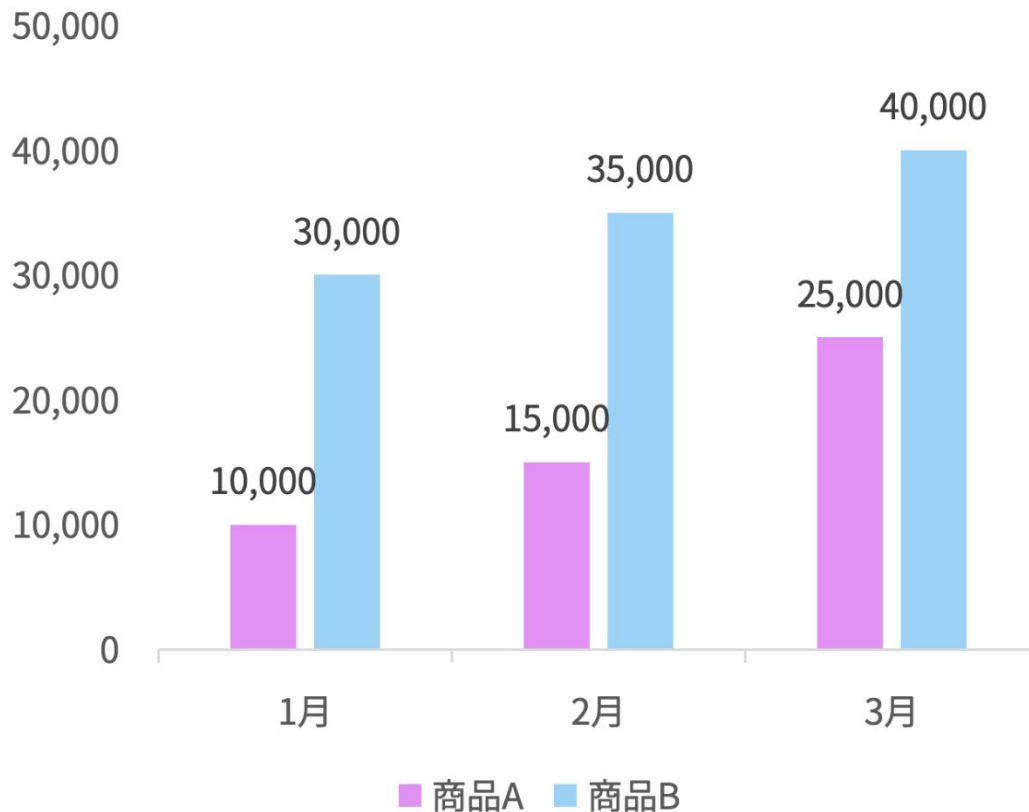
## 例えば...

ある会社における商品の売り上げを可視化したい！

月別の商品別売上高

	1月	2月	3月
商品A	10,000	15,000	25,000
商品B	30,000	35,000	40,000

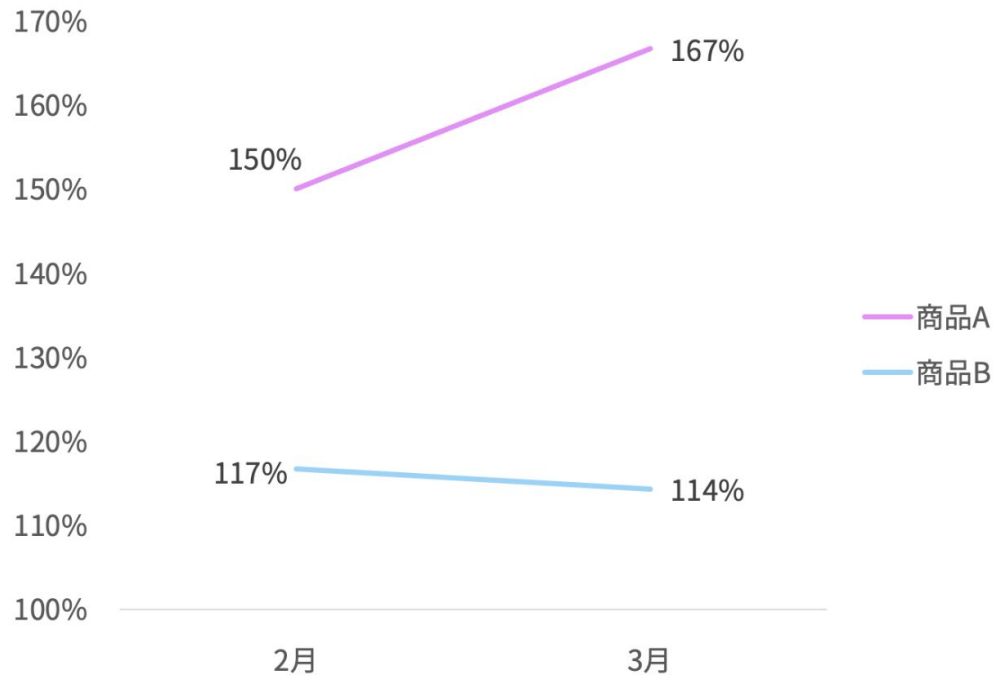
# より見やすくなる！



月別の商品別売上高と前月対比

		1月	2月	3月
商品A	売上高	10,000	15,000	25,000
	前月対比	-	150%	167%
商品B	売上高	30,000	35,000	40,000
	前月対比	-	117%	114%

# 可視化の方法を変えると伝わることが変わる！



1. 現状と課題の把握  
→課題解決や意思決定がしやすい
2. データ(課題)の共有による作業の効率化  
→前提条件を揃えやすい！
3. 蓄積データを活用した予測・分析で先手を打てる  
→今後どうなっていくかを統計的に予測

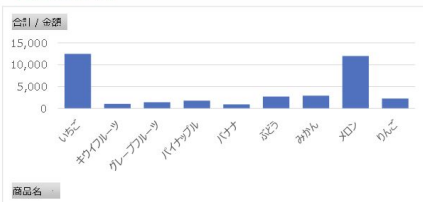


# データ可視化の手法について

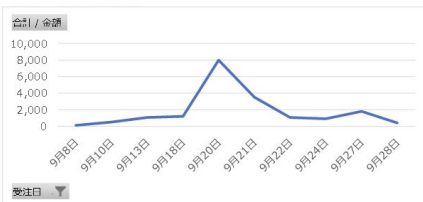


## Microsoft Excel

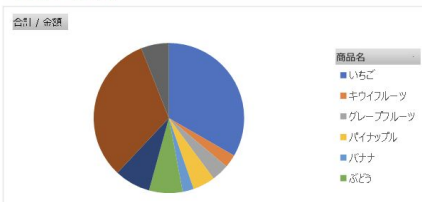
▼実数の大小



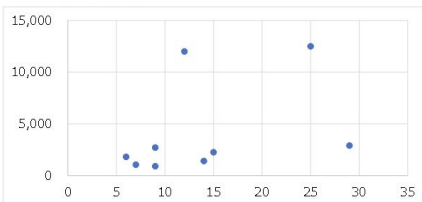
▼データのトレンド



▼比率の大小



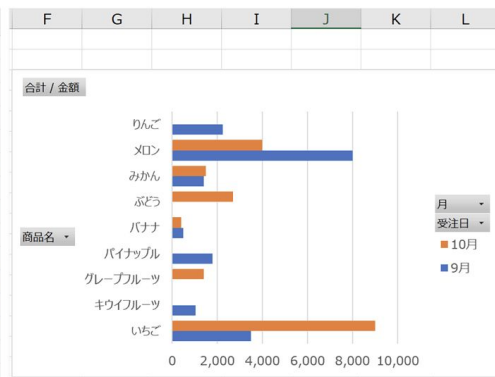
▼データのばらつき



▼表の場合

	A	B	C	D
1				
2				
3	合計 / 金額	列ラベル		
4		9月	10月	総計
5	行ラベル			
6	いちご	3,500	9,000	12,500
7	キウイフルーツ	1,050		1,050
8	グレープフルーツ		1,400	1,400
9	パイナップル	1,800		1,800
10	バナナ	500	400	900
11	ぶどう		2,700	2,700
12	みかん	1,400	1,500	2,900
13	メロン	8,000	4,000	12,000
14	りんご	2,250		2,250
15	総計	18,500	19,000	37,500

▼グラフの場合



<https://excel-master.net/> © Koushi Morita

<https://excel-master.net/> © Koushi Morita

- メリット

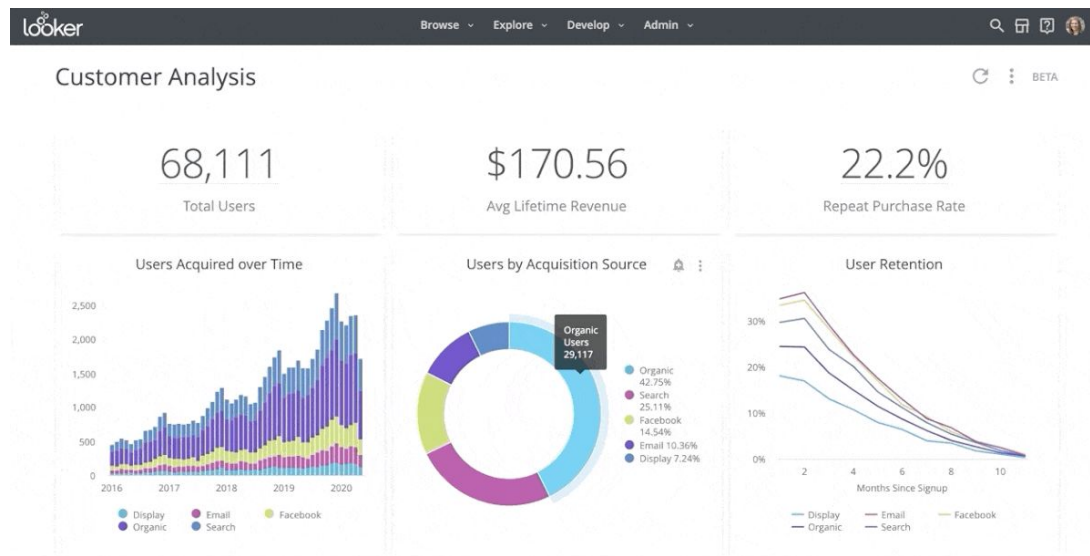
- 身近なソフトウェアである
- 簡単に使える
- 経験者は多い

- デメリット

- 機能への理解が必要
- 綺麗な可視化にはセンスが必要
- 連携が煩雑になりがち
- 共有が難しい
- 大容量データ/更新されるデータには向いていない

# ビジネスで使われているツール: BIツールとは、？

## BI(ビジネスインテリジェンス)



ビジネス上の意思決定に役立つ情報を  
収集、分析、可視化するために使用されるソフトウェア

【出典】

<https://www.looker.com/intl/ja/product/visualizations/>

## ● BIツールを使う理由

- 可視化に特化していて直感的に使えるものが多い
- 社内での共有が容易
- 大容量データも簡単に使用可能
- 定期更新等、DWHに接続して利用

ExcelとBIツールの違い		
業務内容	Excel	BIツール
01. 単一の表やグラフの作成	○	○
02. 表計算形式での分析	○	△
03. 膨大なデータを扱った分析	△	○
04. 複数のデータの同時抽出	△	○
05. 効率的なレポートニング	△	○

【出典】<https://yellowfin.co.jp/blog/21-blog2-summary-of-bi-tools>

# ビジネスで使われているツール: BIツールとは、？

- BIツールの例

コストと必要な機能に応じた選択が必要！



Google Data Portal



そもそも何を可視化するの、、？

A. KPI を可視化します！

Key Performance Indicator(重要業績評価指標)の略称で、  
設定した目標の達成度を表す「中間目標」



例えば、、企業なら  
目標:「利益を最大化する」

簡単な例なら、、

- 売上はどれぐらい？ → 売上(流通高)
- そのうちの利益の割合はどれぐらい？ → 利益率
- どのぐらいの人や会社(顧客)に購入された？ → 取引数, 購入UU
- 業界におけるシェアは？ → シェア率

## 利益を上げるには？

- 売り上げ
  - サイトに来た人数はどれぐらいか？
  - サイトに来て商品ページを開いた人はどれぐらいか？
  - 何人のお客さんが購入したか？
  - キャンセルされた割合は？
- 利益率
  - 原価率は？
  - コストはどこにかかっている？

- 売上高：全体の収益
  - 地域別売上：どの地域で売れているか？
  - カテゴリー別売上：どのカテゴリーが売れているか
  - 時期(季節、月、週、日)別売上：どの時期に売れているか
  - 新規とリピート顧客からの売上比率：新規と既存どちらを重視した施策をしたらいいか？
  - 取引数：どれぐらいの人が購入しているか
  - 取引単価：一回/一人あたりどれぐらい買っているか
  - キャンセル率：購入まで行ったけどキャンセルされた数は、、？
- 利益率：売上のうちどれぐらいが利益になるか
  - 広告コスト：広告コストはどの程度売り上げに対して占めているか
  - 商品カテゴリー別利益率：どのカテゴリーが一番利益が高いか
  - 地域別利益率：輸送コスト等も含めてどこが一番利益を出せているか？
  - 期間(季節、月、週、日)別利益率：いつ施策を打てばいいか？

- 大目的:

「コストパフォーマンスに優れた料理の提供を通じた各店舗の売上アップ」

→1人の1時間の生産性が圧倒的に高い状態を作る

KPI: 従業員1人が1時間で生み出す粗利益を示す「人時生産性」

- 開店前に1時間費やしていた作業が半分の30分に短縮
- 人時生産性の業界平均:3,000円程度→サイゼリヤ:4,000円
- コストパフォーマンスに優れた料理の提供 & 売り上げアップを実現



# 実際の企業の例: ハウステンボス株式会社

- 大目的:

「来場者数の増加」

→多くの『ここにしかないイベント』が開催できている状態をつくる

KPI: 『オンリーワン・ナンバーワン』のイベント展開数

- 話題性の高まり
- 入場者数はV字回復。140万人にまで減った来場者数も300万人以上に

→独自性を追求することをKPIに掲げたことで来場者数の増加を実現



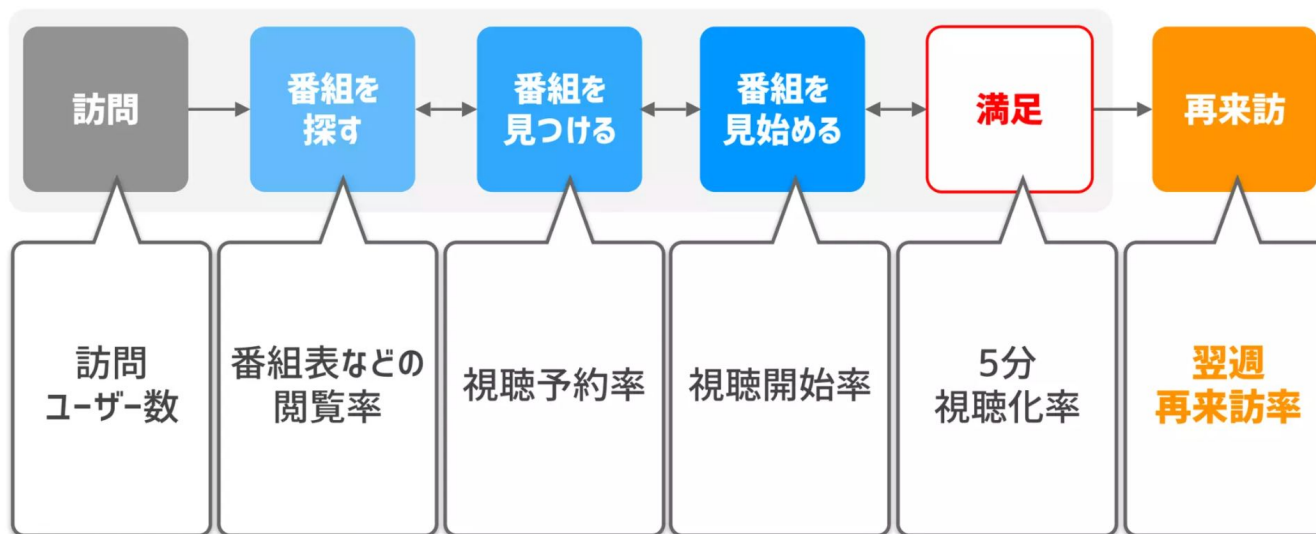
- 大目的:

「ユーザーの満足度向上」



導入・設計

## KPI設計



## 本日扱うデータについて

## 厚生労働省で公開しているコロナに関するデータをいじったもの

```
1 %sql
2 SELECT
3   *
4 FROM
5   main.covid19.daily_cases
6 LIMIT 10
```

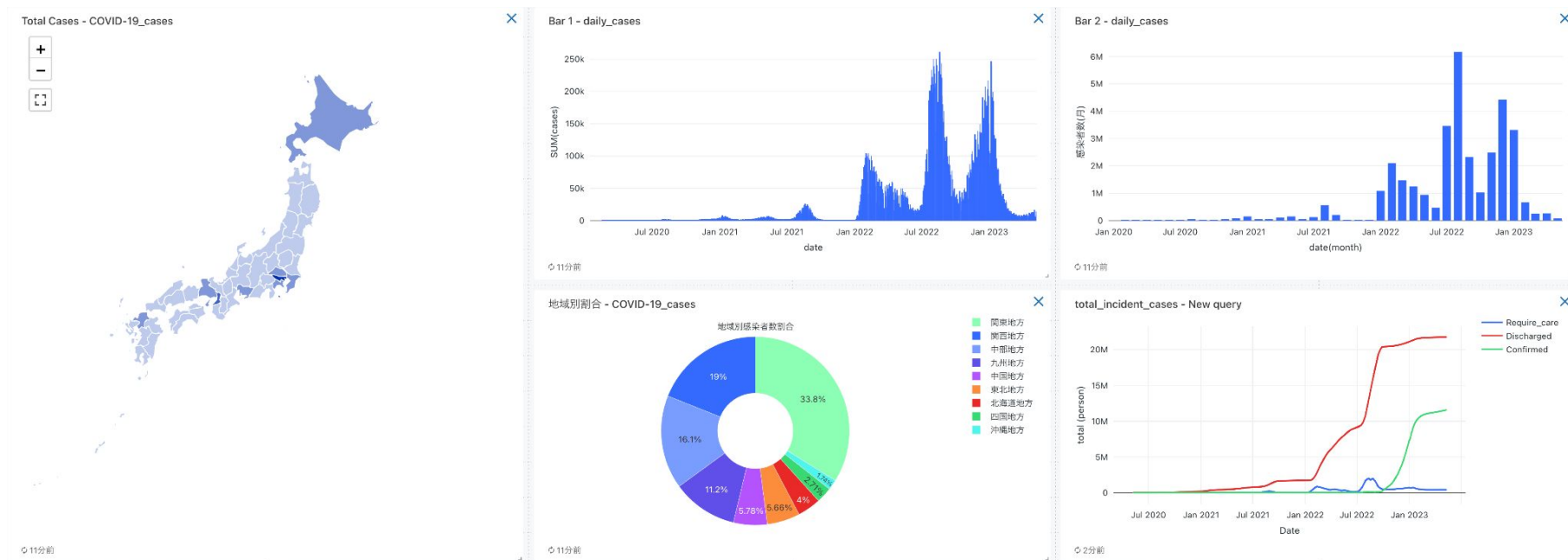
▶ (2) Sparkジョブ

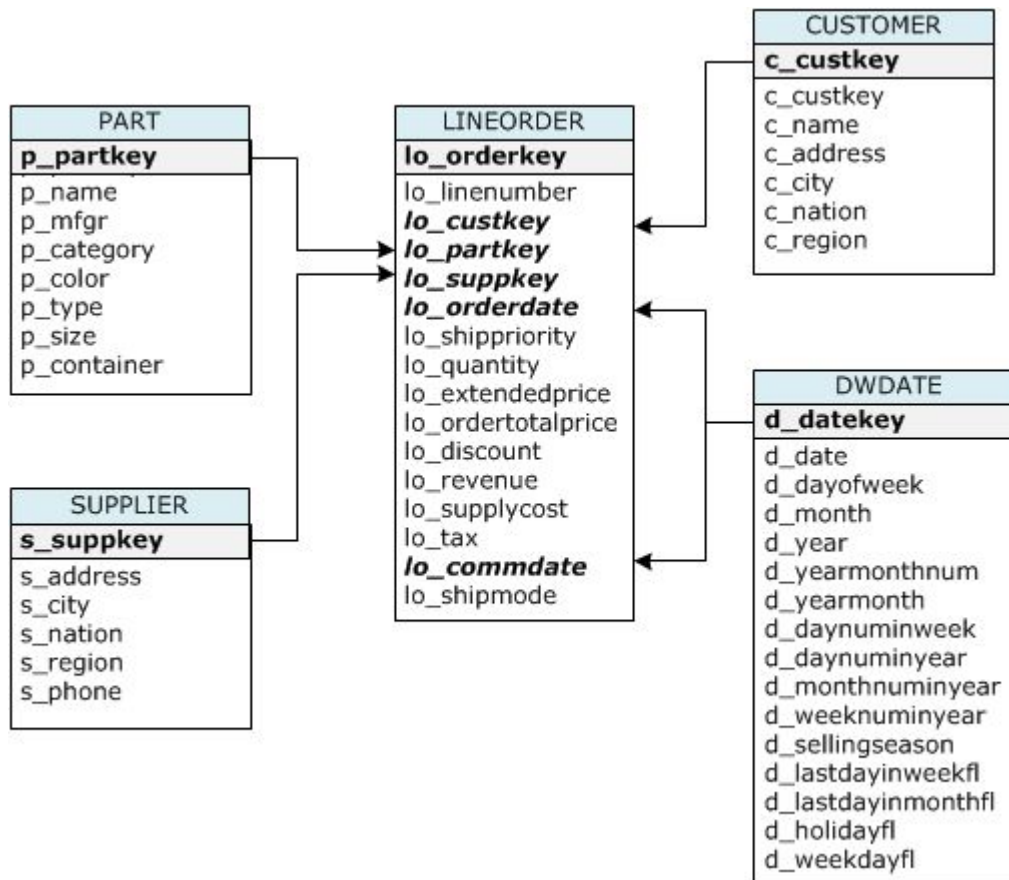
テーブル ▼ +

	date ▲	pref_no ▲	Prefecture ▲	Area ▲	cases ▲	Prefecture_code ▲	
1	2020-01-16	1	Hokkaido	北海道地方	0	JP-01	
2	2020-01-17	1	Hokkaido	北海道地方	0	JP-01	
3	2020-01-18	1	Hokkaido	北海道地方	0	JP-01	
4	2020-01-19	1	Hokkaido	北海道地方	0	JP-01	
5	2020-01-20	1	Hokkaido	北海道地方	0	JP-01	
6	2020-01-21	1	Hokkaido	北海道地方	0	JP-01	



## Dashboard 例





実際にやってみよう！

Databricksに移動します。