

# 社会人のためのデータサイエンス演習

第1週：データサイエンスとは

第2回：データサイエンスが必要とされる背景

講師名：佐伯 諭

# ビジネスのネット化

## 金融サービス



インターネットバンキング



金融取引

## 情報サービス



音楽配信



ネットショッピング



ナビゲーションシステム

**ビッグデータ時代の到来**

# ビッグデータの種類と活用領域



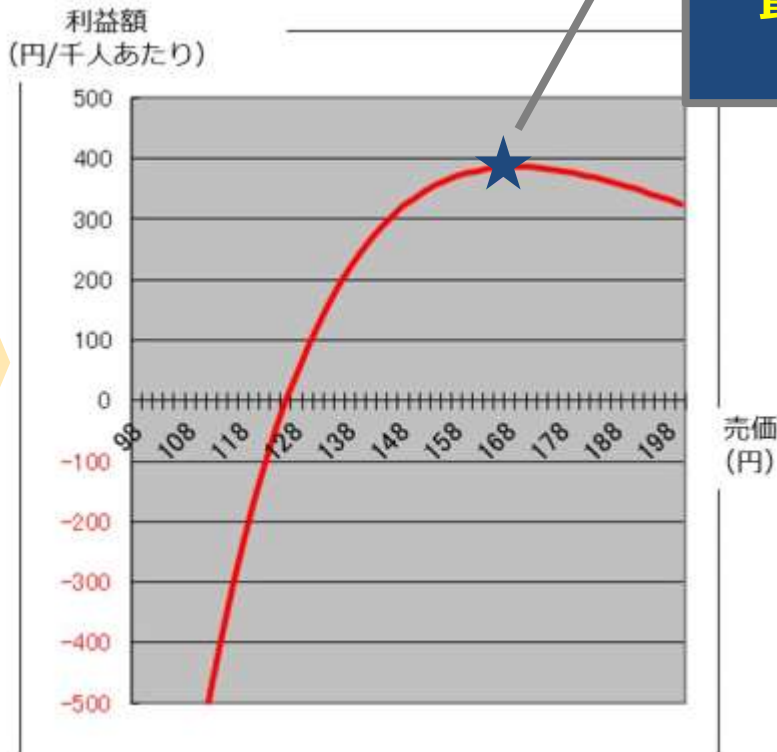
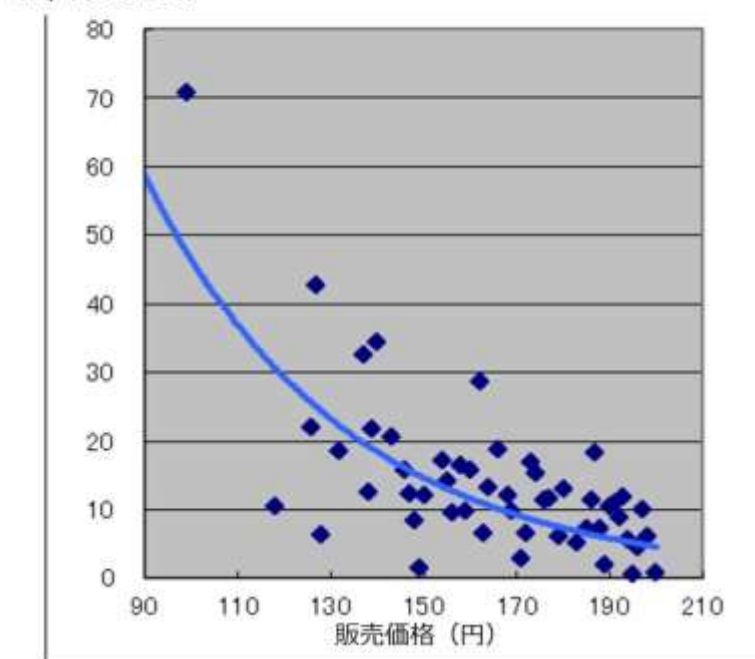


# ビッグデータ活用事例

## 全日食チェーン POSデータ分析で利益最大化

全日食チェーンは、月間 1 億 2 千万件の加盟店のPOSデータを分析し、売れ筋商品のエリア別選定や最適売価の算出を行い利益の最大化を行っている。最適売価は販売価格と販売数量の関係性を把握した上で統計処理を行い算出。算出された最適売価は加盟店各店へ展開される。

加盟店各店における原価128円の  
「食パン」の価格と数量



「食パン」の  
最適売価  
168円

# ビッグデータ×オープンデータ活用事例①

## スマートフォン向けゲームアプリ「Blossom Bristol」

Blossom Bristolは、英国の企業Mobile Pie社の提供する農作物育成ゲームです。このゲームはFacebookと連携しており、プレイヤーはFacebookでチェックインを行うと、ゲーム内にある実際の地名上に畑を作り、野菜を栽培することができるようになります。この各地点のリアルな環境データと併せて、「天候不順が続くと農作物がよく育たず販売価格が下がる」「カリフラワーは大気汚染度が高くてもよく育つ」といった農作物の特性が設定されており、プレイヤーには畑の位置と環境情報を踏まえた栽培戦略が求められます。このように、オープンデータという政府が公開する正確性の高い実データをゲームに活用することで、リアルとバーチャルを結ぶ新しいゲームビジネスが誕生しています。

出典：@IT 「Database Expert」 「5分でわかるオープンデータ」 2014年7月14日掲載  
[http://www.atmarkit.co.jp/ait/articles/1407/14/news006\\_3.html](http://www.atmarkit.co.jp/ait/articles/1407/14/news006_3.html)



Blossom Bristolのゲーム画面  
(出典：Google Play Blossom Bristolダウンロードページ)

# ビッグデータ×オープンデータ活用事例②

## ナビゲーションサービスで自治体提供の避難所情報検索

ナビタイムジャパンは各種サービスで検索できる避難所情報に、自治体が公開している避難所情報を加えて検索できるようにした。これまでナビタイムジャパンのサービスで検索できる避難場所は主に国土交通省が情報提供している避難場所に限られていた。

The screenshot displays the NAVITIME website interface for searching disaster shelters. On the left, a list of schools is provided as shelter locations, each with a link to a map:

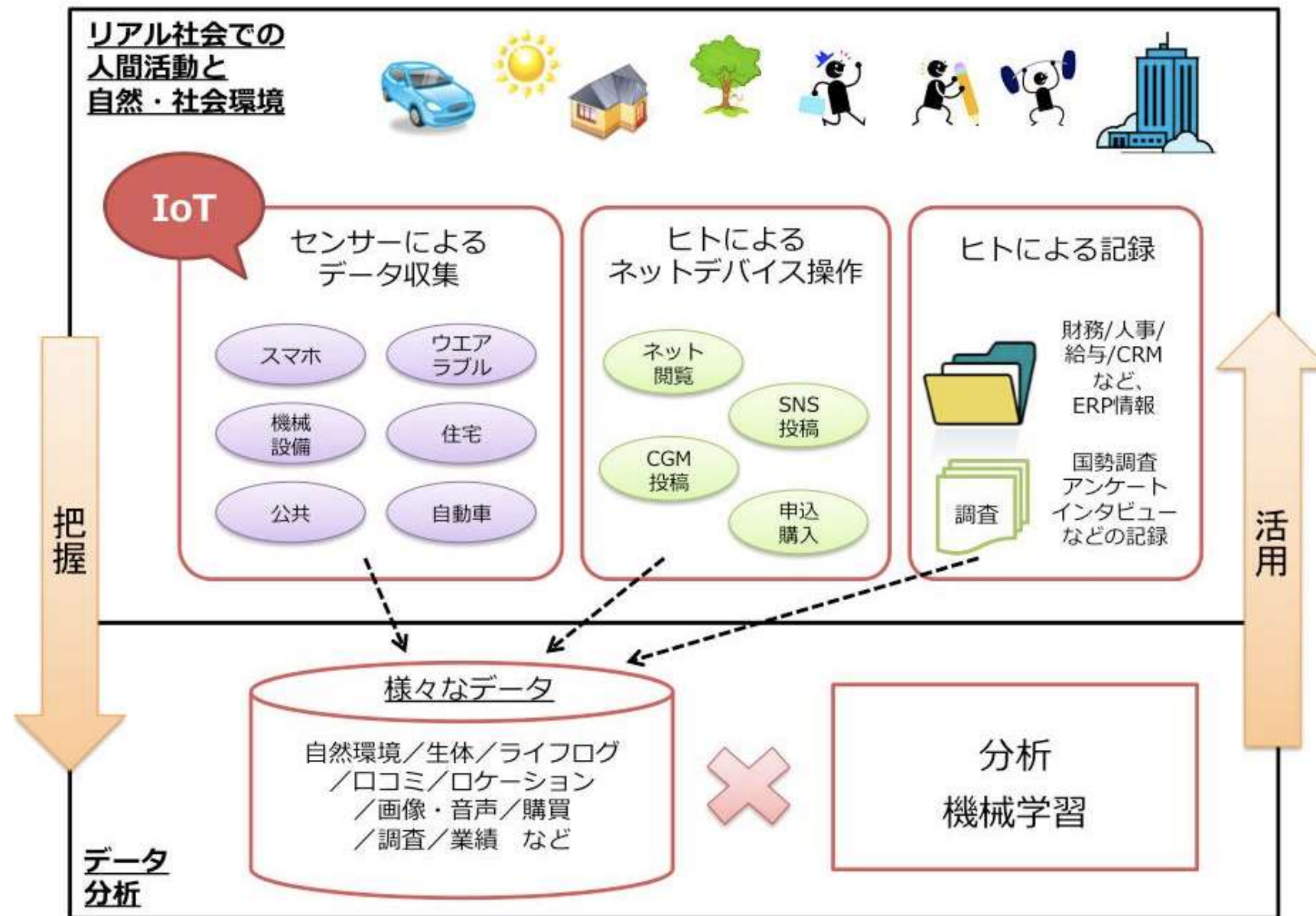
- 古市場小学校 (Fukumachi Elementary School) - 神奈川県川崎市幸区古市場1-1 [地図]
- 南原小学校 (Minahara Elementary School) - 神奈川県川崎市高津区上作延796 [地図]
- 子母口小学校 (Koboguchi Elementary School) - 神奈川県川崎市高津区子母口730 [地図]
- 幸町小学校 (Kōchō Elementary School) - 神奈川県川崎市幸区中幸町2-17 [地図]
- 千代ヶ丘小学校 (Chiyogakino Elementary School) - 神奈川県川崎市麻生区千代ヶ丘8-9-1 [地図]

The main content area shows the selected location, 古市場小学校, with its address (〒212-0052 神奈川県川崎市幸区古市場1-1) and a map. Below the map, the nearest bus stops and parking areas are listed:

最寄り駅	最寄りバス停	最寄り駐車場
1 鹿島田 約1.4km 徒歩で約19分 乗換案内   徒歩ルート	1 下平間 約785m 徒歩で約10分 バス乗換案内   バス系統/路線	1 リパーク川崎古市場 約643m 徒歩で約8分
2 平間 約1.7km 徒歩で約21分 乗換案内   徒歩ルート	2 武蔵館前 (神奈川県) 約674m 徒歩で約11分	2 ナビパーク川崎古市場第2 約152m 徒歩で約2分

# The Internet of Things (IoT)

(モノのインターネット)



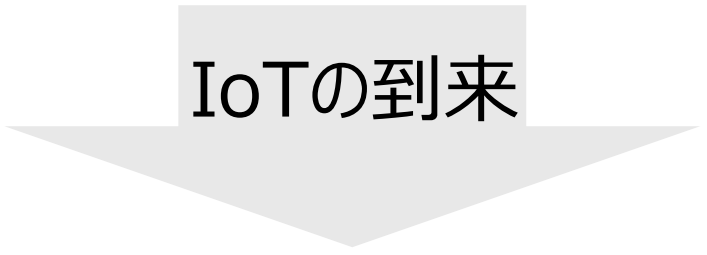
# データ収集・活用の考え方の遷移

---

従来のビッグデータ活用

社内に蓄積されたデータの利活用

IoTの到来



社内に蓄積されたデータの利活用

+

社内外のデータを生成・収集・蓄積

**積極的なデータの収集・活用に移行**

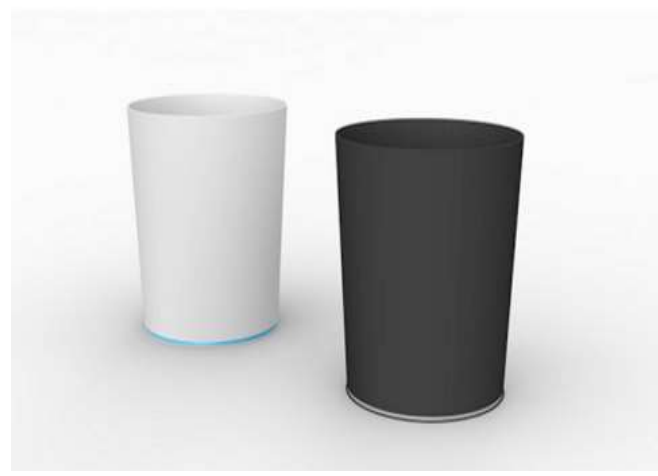


# IoT×オープンデータ活用事例

## KDDIと沖縄セルラー スマートフォンと連携して使用するインテリア雑貨販売



スマートフォンをもって傘立てに近づくと、傘の必要性に応じて傘立て底部に内蔵されたLEDが発光。同時にスマートフォンに通知が届き天気の詳細情報が確認できる。



スマートフォンをもってゴミ箱に近づくと、今日が何の種類のゴミの収集日かをLEDの色で教えてくれる。同時にスマートフォンにゴミの収集日に関する情報が通知される。



# ビッグデータの3つの特性

## Variety ～データの多様性～



様々なサービス・機器が増え  
データの種類が増加

## Volume ～量の膨大さ～



データ自体が大幅に増加

## Velocity ～スピード～



データが生成されるスピードも求められる  
意思決定のスピードも加速

# ビッグデータ・IoTに伴う変化

## ● 企業における2つの変化

### 顧客との対面接触機会の減少



顧客の行動心理がわかりづらくなる傾向

### 社内のデータ量の増加



社内に顧客に関するデータが増加

# データサイエンスの必要性

## ● データの活用が企業競争力になる時代へ



データ分析がうまくできれば



対面取引以上に  
顧客の行動心理を理解  
できる可能性が高い



データ分析ができないと



対面取引が減った分  
顧客の行動心理が  
理解しづらくなる可能性  
が高い



**データに基づき問題を解決する能力が求められている**



# 次回のテーマ

---

次回は

**「データサイエンスに求められるスキルや知識」**

お疲れ様でした！

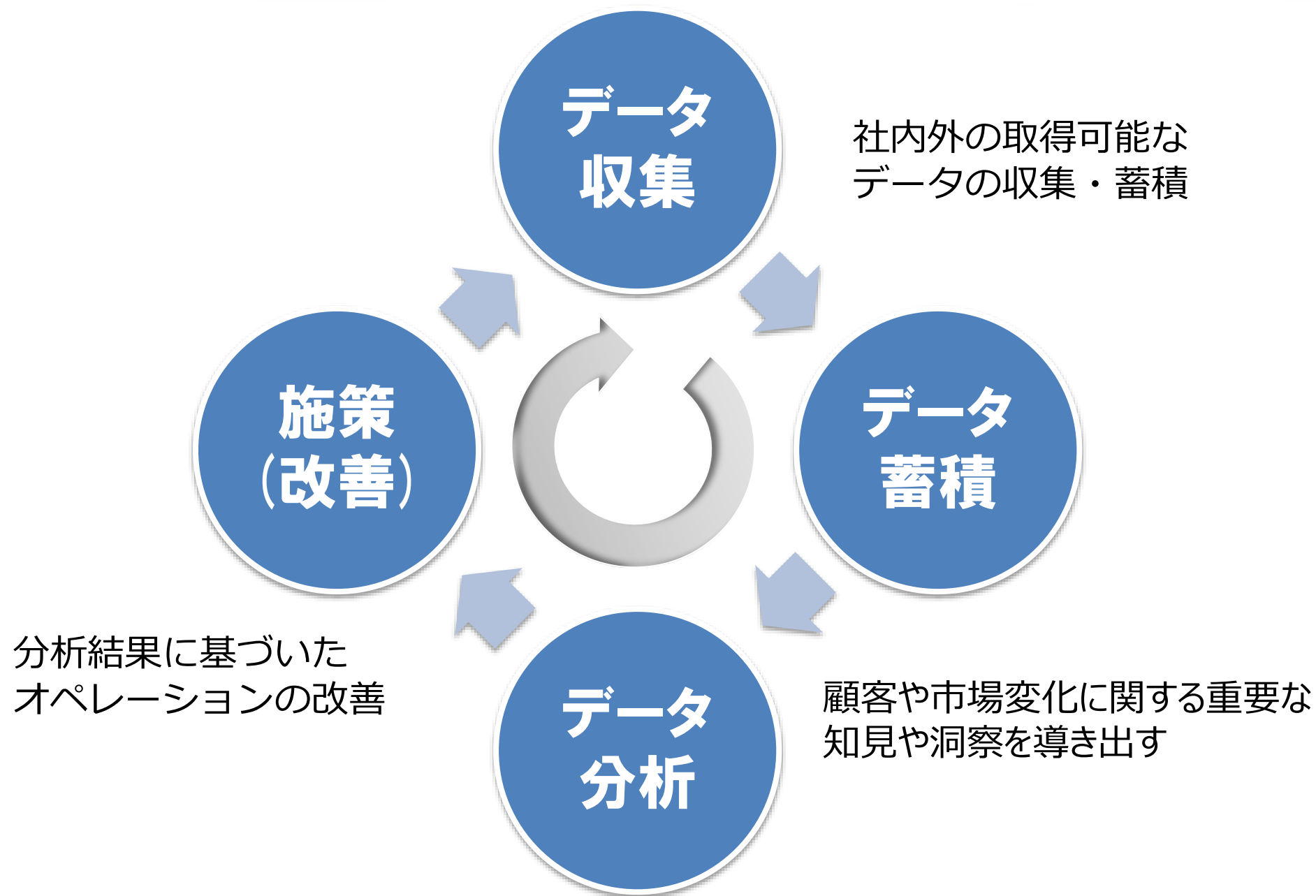
# 社会人のためのデータサイエンス演習

第1週：データサイエンスとは

第3回：データサイエンスに求められるスキルや知識

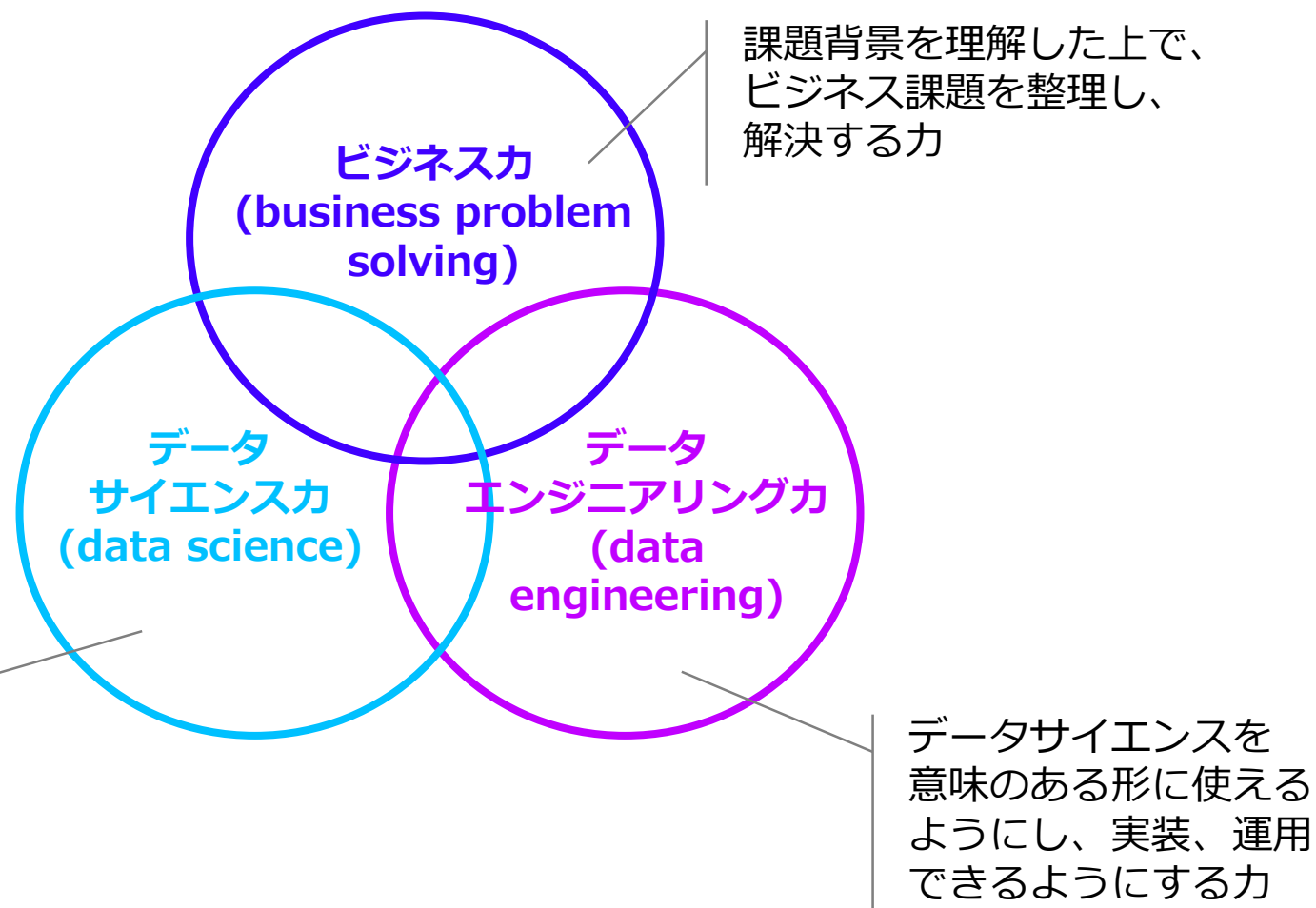
講師名：佐伯 諭

# 求められる能力



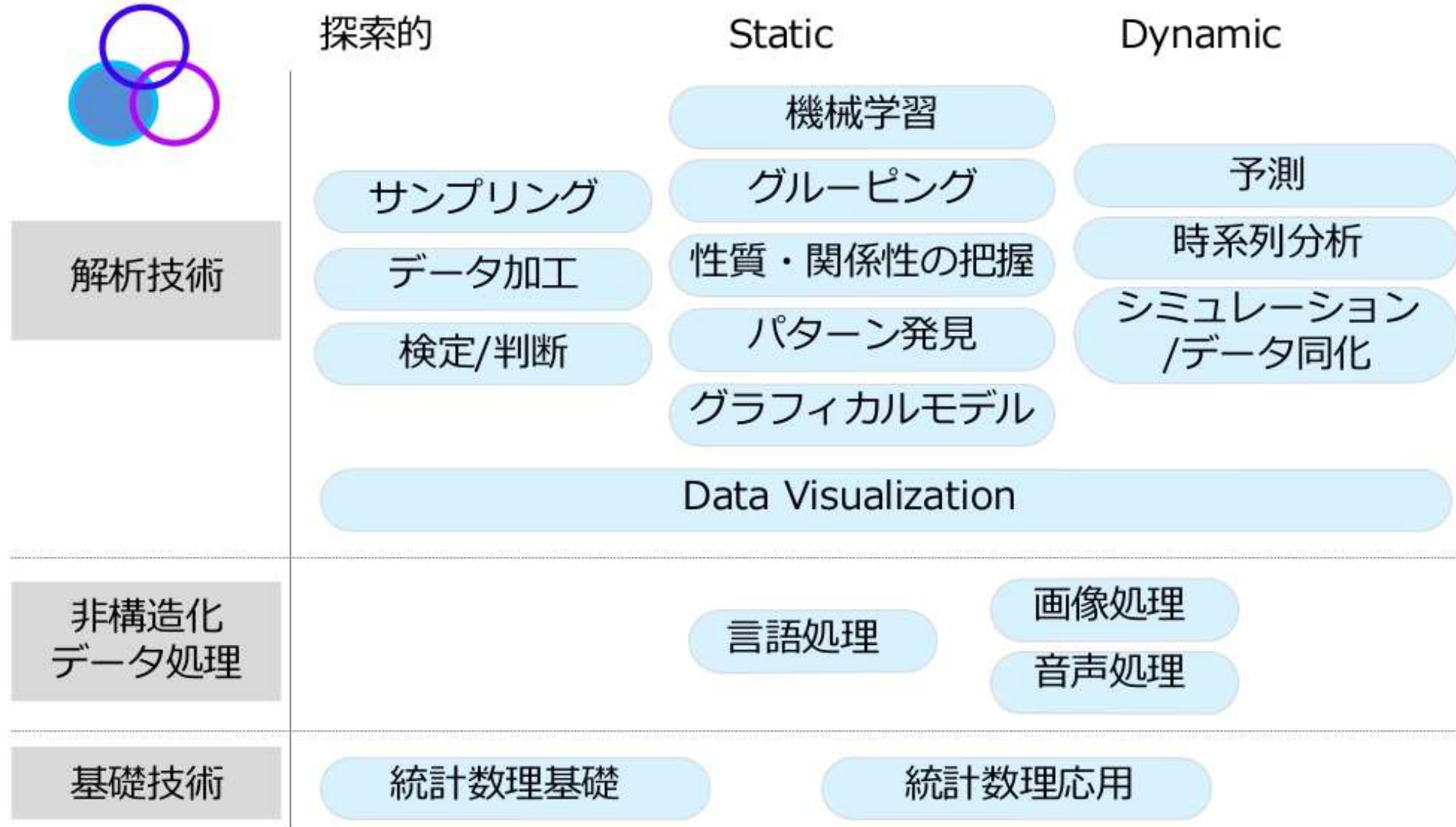
# データサイエンティストに求められるスキルセット

- データサイエンティスト協会では以下のように定義されている





# データサイエンスの領域



# 本講座のねらい

---

- データ分析が問題解決にどのように役立つかを理解
- 統計学に基づく、分析手法の理解と  
データ加工・分析スキルの向上
- データ分析をビジネス現場で活用できるようになる

# 講座内容

## 第1週

- データサイエンスとは

## 第2週

- 分析の概念と事例  
ビジネス課題解決のためのデータ分析基礎(事例と手法)①

## 第3週

- 分析の具体的手法  
ビジネス課題解決のためのデータ分析基礎(事例と手法)②

## 第4週

- ビジネスにおける予測と分析結果の報告  
ビジネス課題解決のためのデータ分析基礎(事例と手法)③

## 第5週

- ビジネスでデータサイエンスを実現するために

# 次回のテーマ

---

次回は

**「データサイエンスの将来」**

お疲れ様でした！



# 社会人のためのデータサイエンス演習

第1週：データサイエンスとは

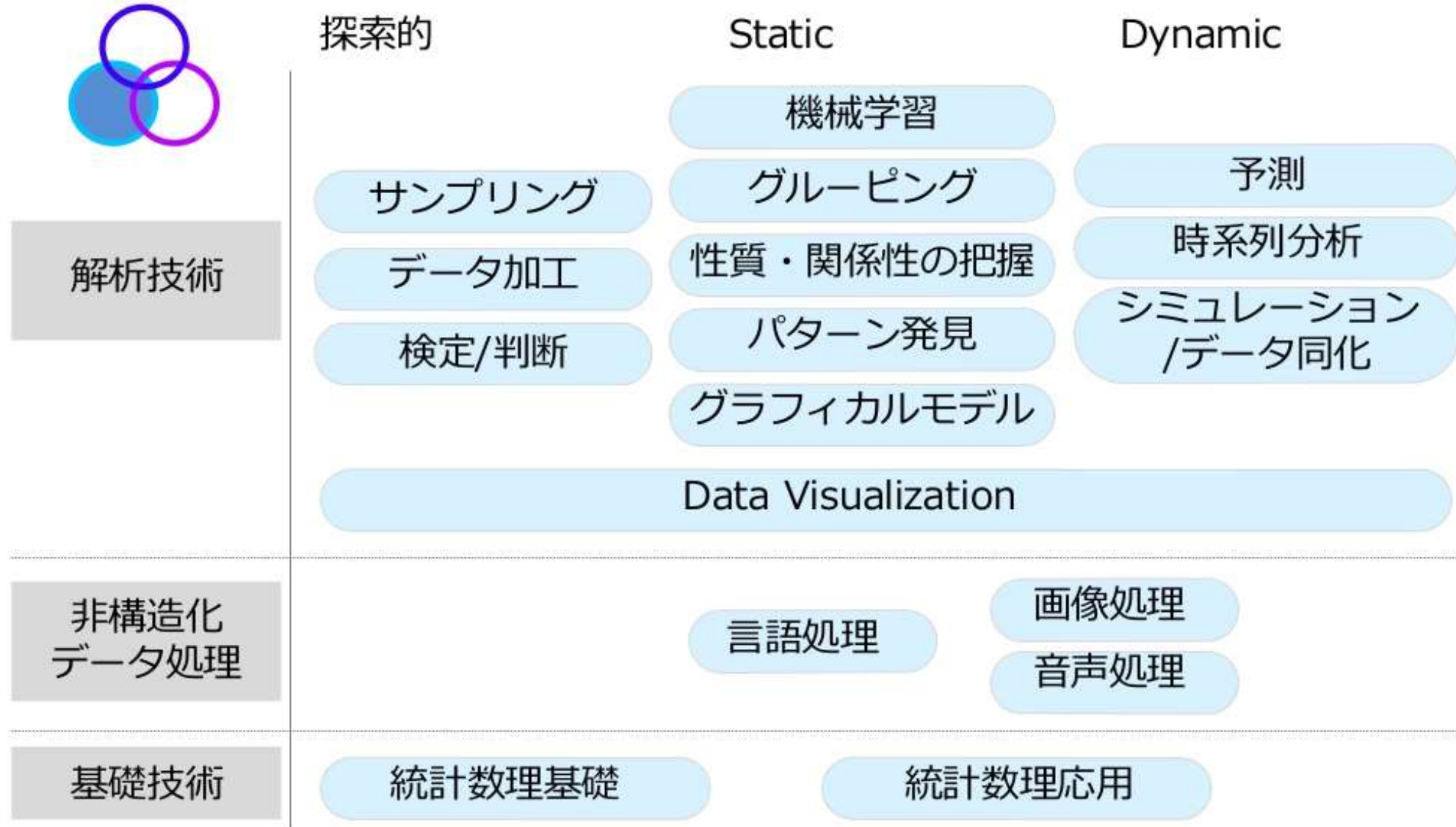
第4回：データサイエンスの将来

講師名：松尾 豊

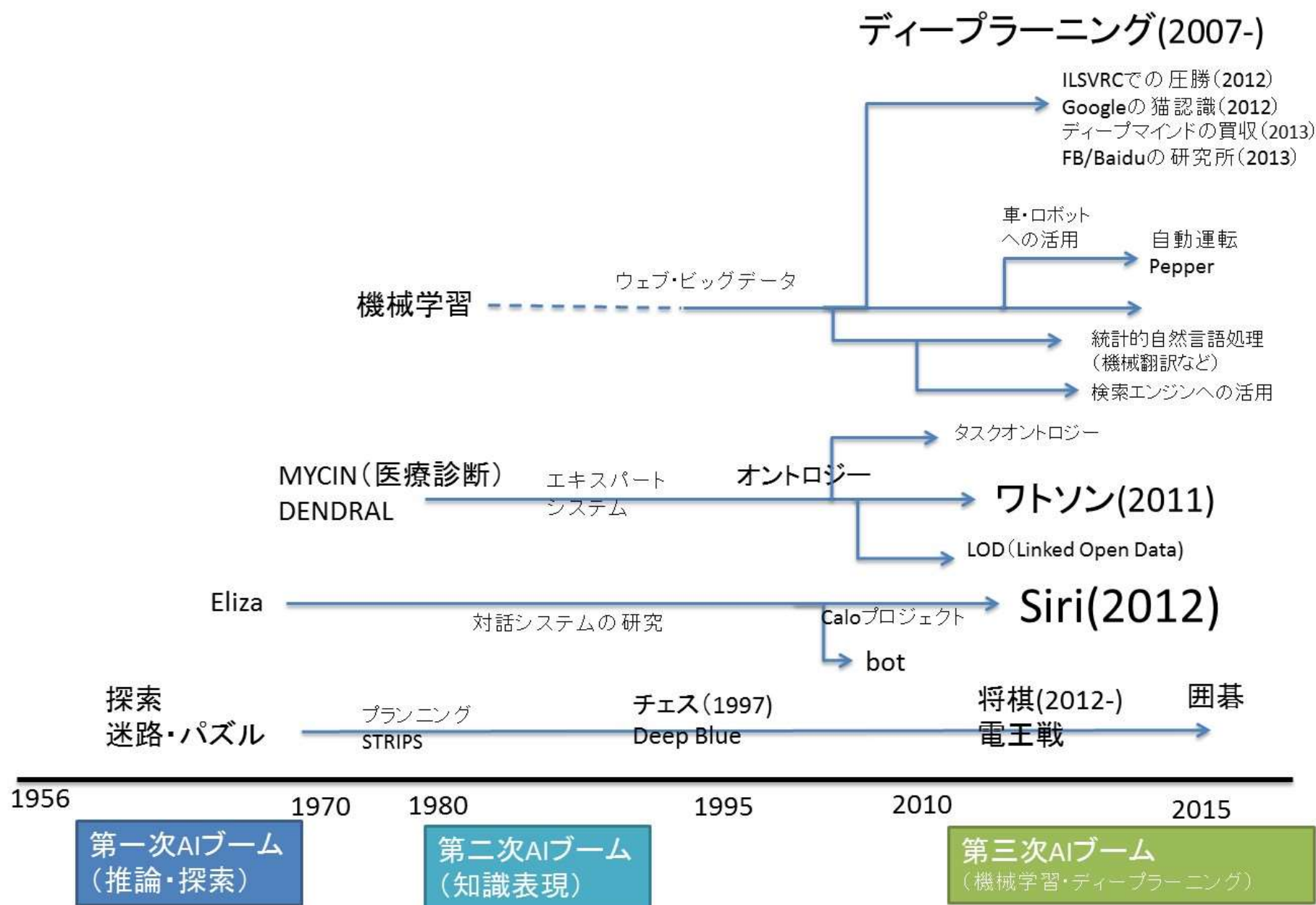
# 第1週の内容紹介

第1回	● “データサイエンス”力の高い人材育成について
第2回	● データサイエンスが必要とされる背景
第3回	● データサイエンスに求められるスキルや知識
第4回	● データサイエンスの将来
第5回	● PPDACサイクルに沿った問題解決の進め方
第6回	● 分析の設計手法

# データサイエンスの領域



# データサイエンス(人工知能)をめぐる動向





# データサイエンス活用事例(人工知能)

## Googleの猫

米Googleは現地時間2012年6月26日、同社研究チームの機械学習技術に関する研究成果を紹介した。脳をシミュレーションする大規模ネットワークを用いた新たな手法により、コンピュータが猫を認識する能力を自ら身につけることに成功したという。1週間にわたりYouTubeビデオを同ネットワークに見せたところ、ネットワークは猫の写真を識別することを学習した。

出典：ITPro「Google、大規模人工ニューロンネットワークを用いた研究成果を紹介」 2012年6月27日掲載  
<http://itpro.nikkeibp.co.jp/article/NEWS/20120627/405501/>



# 既存産業における今後の発展見込

画像認識の精度、運動の習熟向上により、今後多くの作業の効率化・自動化が可能となると考えられる

## ● 農業



収穫  
判定

- ✓ トラクター・コンバインの適用範囲の拡大
- ✓ 選別調整の効率化・自動化

収穫・耕うんの自動化

## ● 食品加工

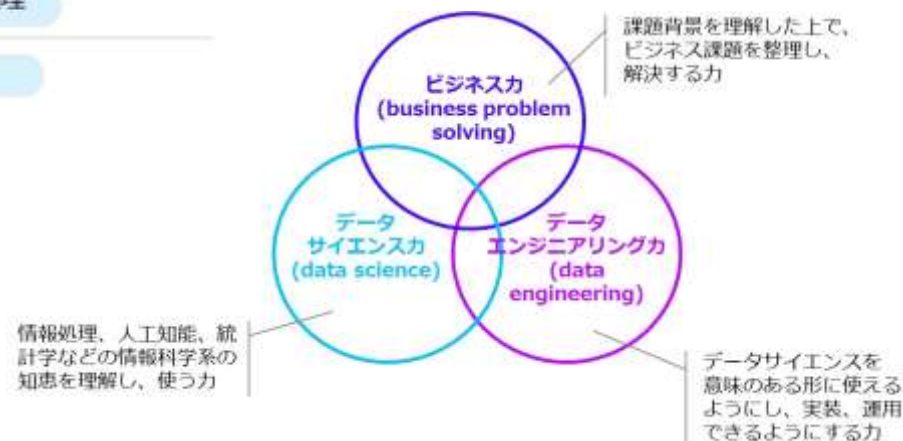


振り分け  
確認

- ✓ カット、皮むき、解体などの自動化

加工工場の自動化

# データに基づく問題解決能力が身につくと・・・



企業競争力のさらなる向上につながる

# 次回のテーマ

---

次回は

**「PPDACサイクルに沿った問題解決の進め方」**

お疲れ様でした！

# 社会人のためのデータサイエンス演習

第1週: データサイエンスとは

第5回: PPDACサイクルに沿った問題解決の進め方

講師名: 奥園 朋実



# 第1週の内容紹介

第1回	● “データサイエンス”力の高い人材育成について
第2回	● データサイエンスが必要とされる背景
第3回	● データサイエンスに求められるスキルや知識
第4回	● データサイエンスの将来
第5回	● PPDACサイクルに沿った問題解決の進め方
第6回	● 分析の設計手法

# 問題解決に至る一連のプロセス

- 実社会における問題には複数の要因が複雑に関係
- 問題解決に至るにはプロセスを繰り返す必要がある



**循環型の問題解決プロセスが存在**

# PPDACサイクルとは

---

**P**roblem (問題の発見)

**P**lan (調査の計画)

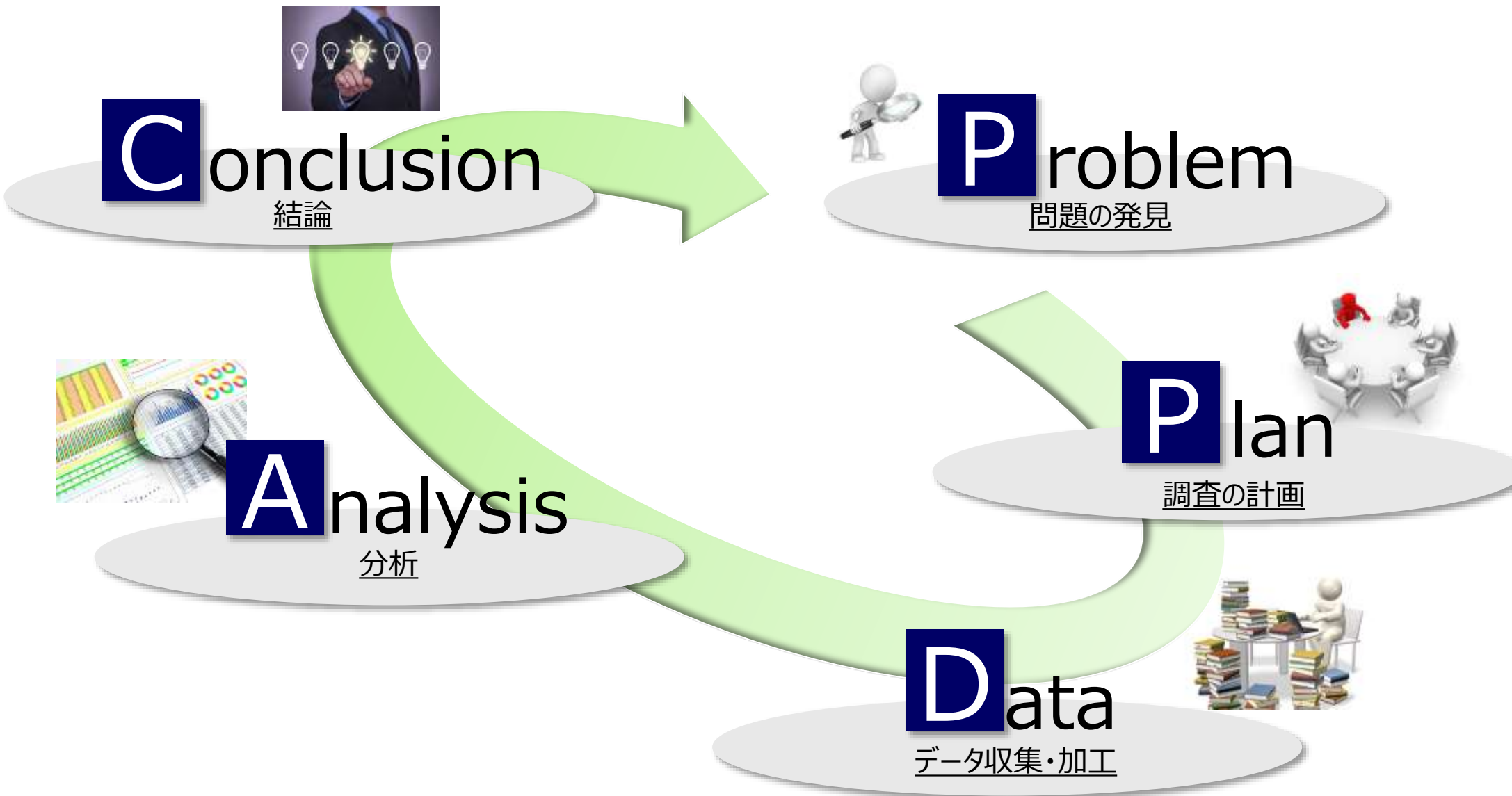
**D**ata (データの収集・加工)

**A**nalysis (分析)

**C**onclusion (結論)

**統計的な思考  
が求められる**

# PPDACサイクルとは



# Problem(問題の発見)

## 業績改善

- 売上
- 利益（売上総利益、営業利益など）

## 自社ブランドの認知度の向上

- アンケート時の認知数やブランド名検索からのサイト流入数

## 顧客満足度の向上

- 定量化された顧客満足度アンケート結果

**問題の明確化と定量化できる指標の決定**  
**KGI(Key Goal Indicator )の決定**



# Plan(調査の計画)



## アイスクリームの売上を改善したい

### 仮説①

気温が高く、日差しが強い日は  
アイスクリームの売上が高くなる

### 分析結果

気温25度以上で晴れの日  
25度未満で晴れの日と比べて売上が30%高い

売上増加のアクションにつなげにくい

# Plan(調査の計画)



## アイスクリームの売上を改善したい

### 仮説②

季節によって売れるアイスクリームの種類は異なる

### 分析結果

夏は氷・シャーベット系が売れ  
冬は濃厚なクリーム系が売れる

仮説を証明するデータがあれば  
売上増加のアクションにつながりやすい

# Plan(調査の計画)



## ● 分析内容のデザイン

- ✓ KGI達成のための因果の仮説の設定
- ✓ データ収集内容・分析手法などの設計

## ● 中間指標（KPI\*）の設定

\*日々の進捗を確認しやすくするために設定された中間指標

**分析結果の要因は数値からは読み取れないため  
分析設計前に仮説を立てておく**

# Data(データの収集・加工)



## ● KPIに対して必要十分で、かつ正確なデータ収集

- ✓ データの形式、単位の統一性、整合性等の確認
- ✓ データ更新状況や仕様変更といった時系列な変化の確認

## ● データの加工の正確性

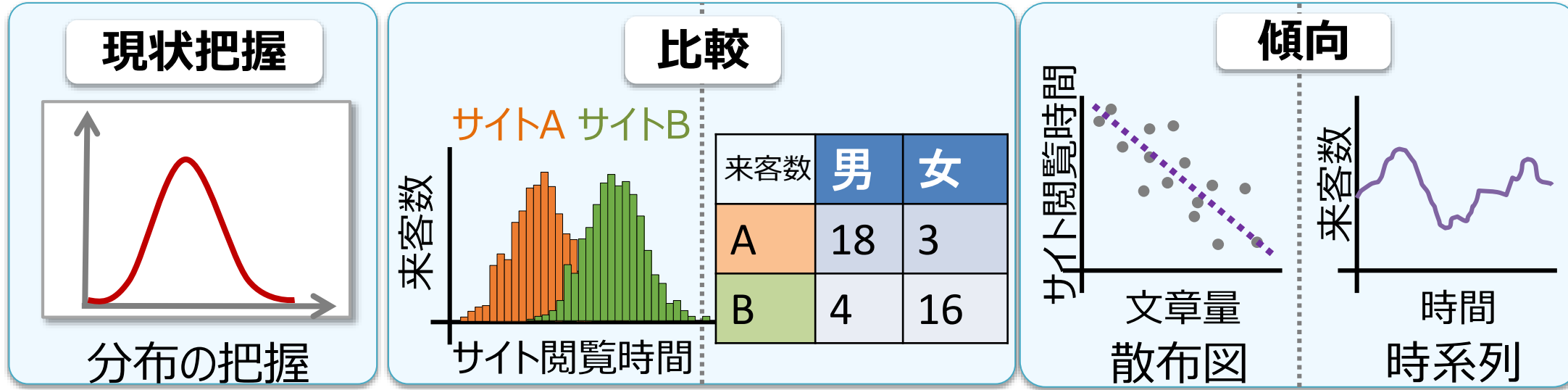
- ✓ 次に行う分析を考慮したデータ加工計画を立てる
- ✓ 手戻りがないよう作業の合間に第3者レビューを入れる

**分析目的に対して精度の高いデータを準備**

# Analysis (分析)



## ● データを表やグラフにして視覚的に見る



ヒストグラムの比較 クロス表

- ✓ 仮説設定時に考えた要因がKPIにどのような影響を与えているか
- ✓ KPIがKGIに対しどのような影響を与えているか

**分析結果から施策に活用するためのヒントを見出す**



# Conclusion (問題の解決)



- 分析結果をもとに、KGIの改善ポイントを判断
- 解釈はデータ収集方法や測定状況なども考慮

これまでのフェーズで出てきた知見をもとに次につながる  
アクションプランを考える

# 次回のテーマ

---

次回は

**「分析の設計手法」**

お疲れ様でした！

# 社会人のためのデータサイエンス演習

第1週：データサイエンスとは

第6回：分析の設計手法

講師名：奥園 朋実

# 第1週の内容紹介

第1回	● “データサイエンス”力の高い人材育成について
第2回	● データサイエンスが必要とされる背景
第3回	● データサイエンスに求められるスキルや知識
第4回	● データサイエンスの将来
第5回	● PPDACサイクルに沿った問題解決の進め方
第6回	● 分析の設計手法

# Problem(問題の発見)

---

あなたは、コンビニエンスストアのオーナーです。  
最近自社の業績が伸び悩んでいることから、  
売上拡大を目的としたマーケティング施策を  
行いたいと考えています



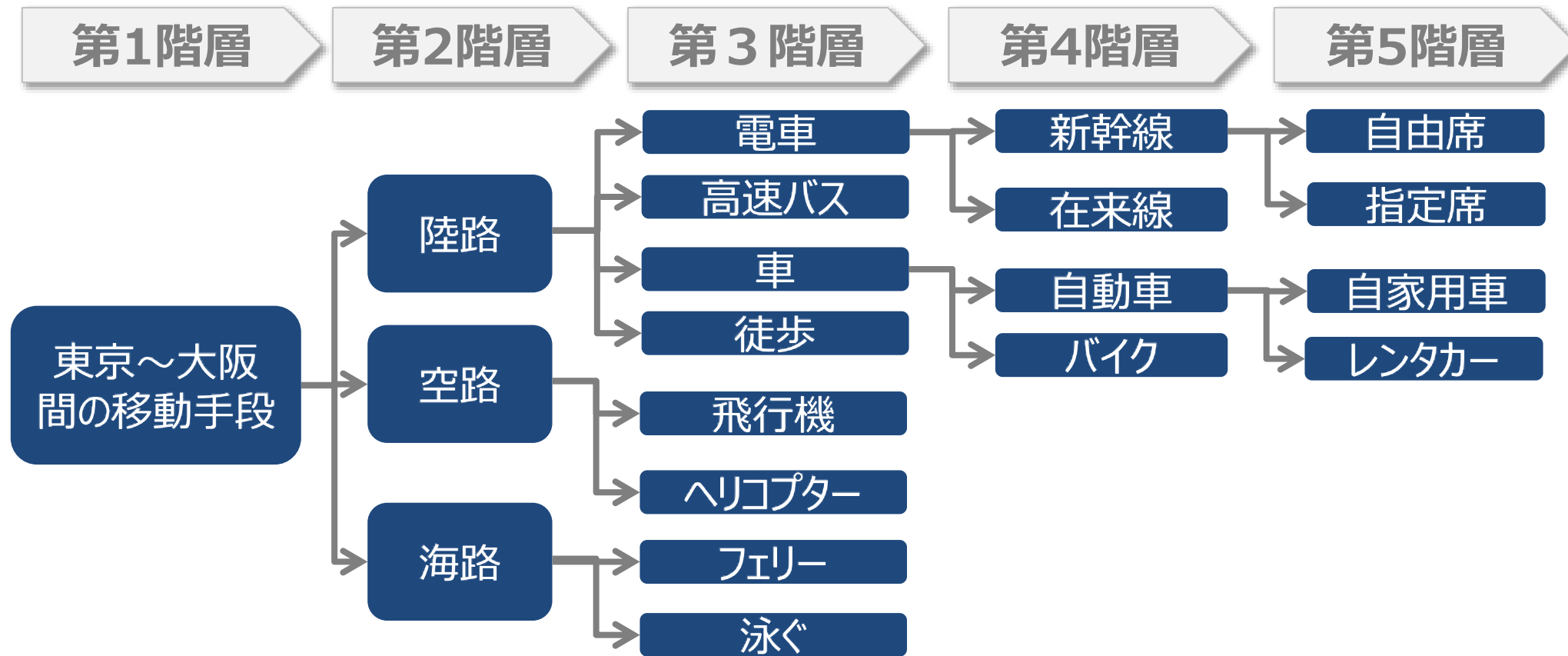
**経営者の悩み** : 業績の伸び悩み  
**問題解決のためのKGI** : 店舗の売上



# Plan(調査の計画) ロジックツリーとは



- 物事を論理的に分析・検討する際にその論理展開を  
樹形図に表現して考えていく思考技法

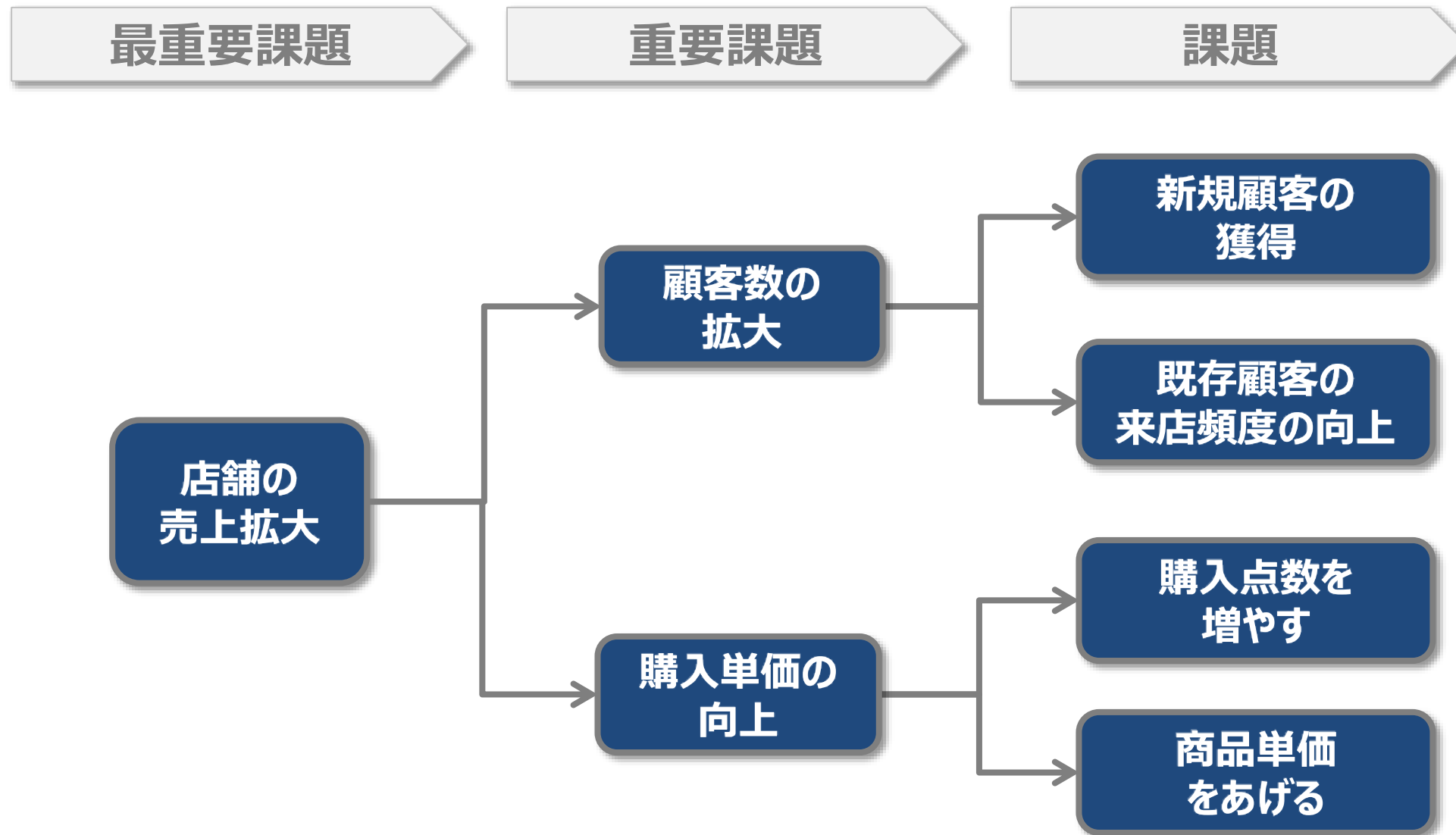


課題の全体把握が容易にでき、議論のズレを修正しやすい

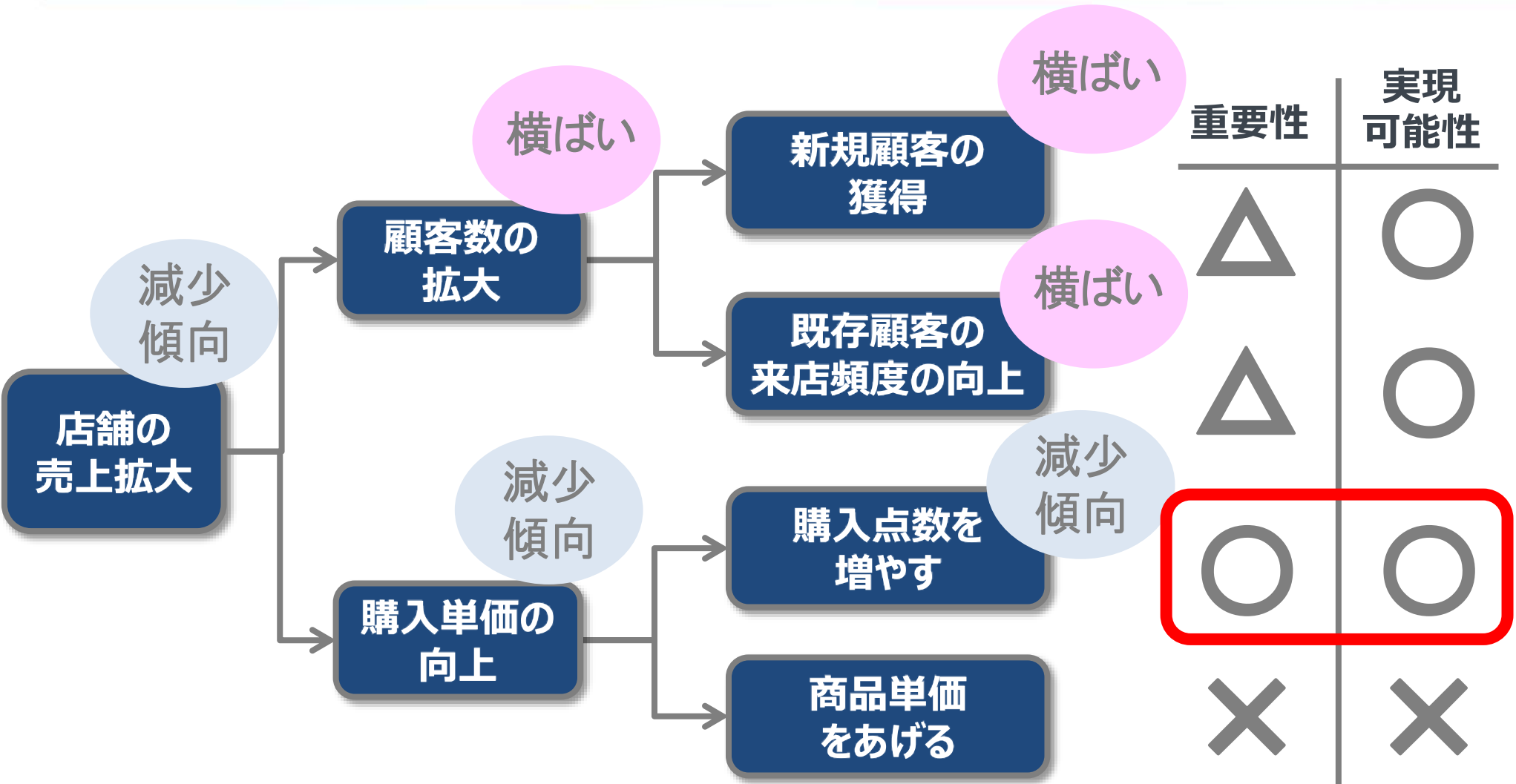
# Plan(調査の計画) ロジックツリーを活用した方針設計



## ● 表層に見えている問題から課題を発見する



# Plan(調査の計画) 現状把握と課題の抽出



各要素の現状確認と課題を抽出し、  
重要性和実現可能性から課題に優先順位をつける

# 次週のテーマ

---

次週は

**「分析の概念と事例」**

お疲れ様でした！