

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:社会で起きている変化～データサイエンスの必要性～

講師名:石川 恵理香

講義内容

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

- 分析的具体的手法
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

第1週の内容紹介

第1回

- データサイエンスの発展

第2回

- 社会で起きている変化 ~データサイエンスの必要性~

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未来と発展

第5回

- データサイエンスのサイクルと課題解決の進め方

第6回

- 分析の手法の選択

第2回の目的

社会において、データサイエンスがどのように役立てられているかを知る

データサイエンスを取り巻くさまざまなキーワードとその意味を理解する

ビッグデータの定義とその特徴を理解する

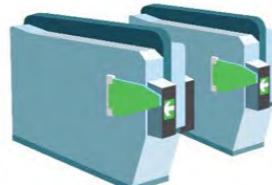
社会における変化

大量のデータの取得が可能になり、さらにデータの処理性能は向上。それらのデータと複数の技術を組み合わせたサービスが誕生している。

私達の日常で取得されるデータの一例



オンラインショッピング
SNS
動画配信サイト など



交通ICカードなどの
移動履歴データ



車載カーナビの
移動履歴データ

データサイエンスを活用し、社会課題の解決へ



快適な買い物の実現
生活者ニーズの把握

最適ルートの提示



混雑・渋滞緩和
など



社会における変化

データサイエンスを取り巻くキーワード

ビッグデータ



ICT情報通信技術の進展によって生成・収集・蓄積等が可能となった多種多量のデータ。

AI



人工知能。大量の知識データを活用し、日常の身近な商品やサービスに組み込まれている。

ロボット



AIやデータサイエンスの技術を用いて、人間の労働の代替として動作するロボットが開発されている。

IoT



Internet of Thingsの略称。さまざまな物体に通信機能を搭載し、センサーから情報を取得する。

DX

DX=Digital Transformation
ICTの浸透により、人々の生活をあらゆる面でより良い方向に変化させる。

ウメオ大学（スウェーデン）エリック・ストルターマン教授提唱

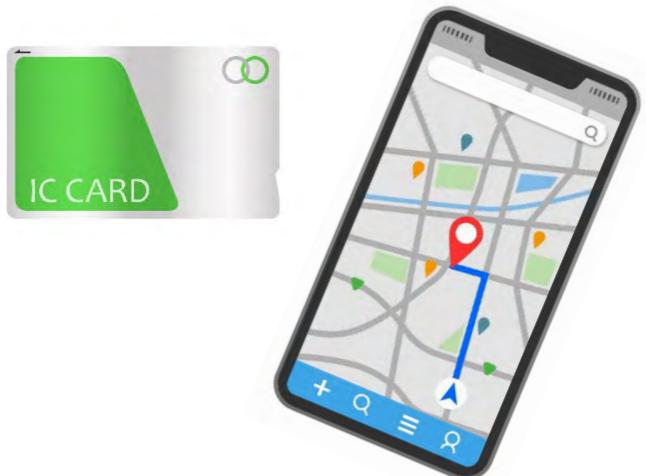
ビッグデータとは

ビッグデータは、社会において生成・取得される様々な形式の大容量データ群。データから有用な知見を見出し、ビジネスや社会の課題解決に活用されることが期待される。

身近なビッグデータの一例

センサーデータ

GPS、交通系ICカード、RFID等で検知される位置、乗車履歴、温度、加速度等



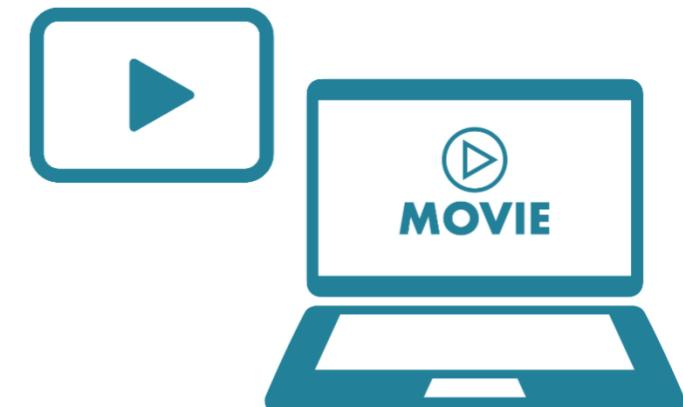
ソーシャルメディアデータ

ソーシャルメディア（SNS）において
参加者が投稿するコメント等



マルチメディアデータ

ウェブ上の動画・音楽配信サービスなどで提供されるデータ



ビッグデータとは

ビッグデータの特徴は以下の「5つのV」で説明される。

Volume (容量の大きさ)

IoTの発達によって現代では膨大な量のデータを取得することが可能になりました。収集したデータは今後継続して集めていくことを想定し、活用しやすい状態にしておく必要があります。

Variety (多様性、種類)

数字だけではなく、テキスト、音声、画像、位置情報、センサー、SNSなどあらゆる形態のデータを様々な場所から得ることができます。データが示す内容の種類も豊富で多岐にわたります。

Velocity (スピード、頻度)

世の中の移り変わりの早い現代では、時代に合ったデータ活用をするためにはリアルタイムに近い新しいデータを活用することが重要です。新しく更新頻度が早いデータは、時代に合ったサービスの提供を可能にします。

Veracity (正確さ)

正しいデータ活用のためには正確なデータが必要不可欠です。正確なデータを集めるためには、手元のデータが事実に沿ったものであるか見極めが必要です。

Value (価値)

データの活用は、社会に価値をもたらしたり課題を解決することにより価値を創造します。

事例：地図と政府統計データを活用したGISサービスの提供

株式会社ゼンリンマーケティングソリューションズの活用事例

概要

自社で保有する地図データを用いたコンサルティング事業において、
政府統計データを活用。企業のマーケティング領域の支援を行っている。

顧客の課題・活用背景

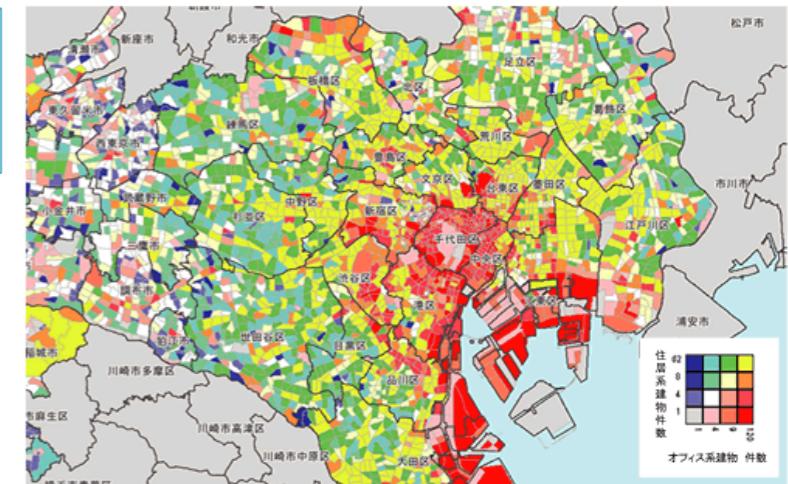
店舗展開をする企業などにおいて、出店エリアの見極め、既存店舗の統廃合などにおいてデータを活用した意思決定のニーズが高まっていた。

活用データ

国勢調査、経済センサス、商業統計調査、家計調査、住宅・土地統計調査 など

効果

- 多店舗展開をする際に、出店すべき商圈エリアの見極めが行えるようになった
- 店舗への物流配送における最適ルート、配達人員の見極めが行えるようになった



行政区分地図対応版

行政界ごとに集計・統計化。ゼンリン行政区分地図データと組み合わせてご利用いただけます。

メッシュ対応版

メッシュごとに集計・統計化。お手持ちの汎用的なデータと組み合わせてご利用いただけます。

※画像は500mメッシュの例です。



株式会社ゼンリンマーケティングソリューションズ提供
「建物統計データ2021」より

事例：過去の気象・被害データに基づいた災害規模予測（仙台市）

概要

気象情報と市の過去の被害記録を収集し、未来の災害規模を予測する。それらの結果を職員が用いて、大雨時に災害対策を迅速に行うべく、共通の危機意識を持てる仕組みを構築。

活用データ

気象情報、市の過去の被害記録データ

プロジェクトの流れ

Problem

大雨災害時に、予想雨量などの気象情報のみでは、被害予測が困難となり、対応に迷いが生じていた。

Data

気象情報と市の過去の被害記録を収集

Analysis

重回帰分析とニューラルネットワークの手法により、気象データと被害記録を統計分析。起こり得る災害規模を予測した。

Conclusion

大雨による災害発生が予想される数日前から、災害規模予測の結果に基づいた迅速な災害対策立案が可能になった



BO sai O peration S tandard T ool

仙台市 災害規模予測ツール



■ 予測メニュー

予測する方法を選択してください。

予測雨量と警報から予測する

【重回帰分析による予測】

気圧配置と予測雨量から予測する

【NNによる予測】

■ 設定メニュー

設定する内容を選択してください。

新規事例を追加する

【重回帰分析による予測】

予測計算パラメータを再調整する

【重回帰分析による予測】

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 謙

第1週の内容紹介

第1回

- データサイエンスの発展

第2回

- 社会で起きている変化 ~データサイエンスの必要性~

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未来と発展

第5回

- データサイエンスのサイクルと課題解決の進め方

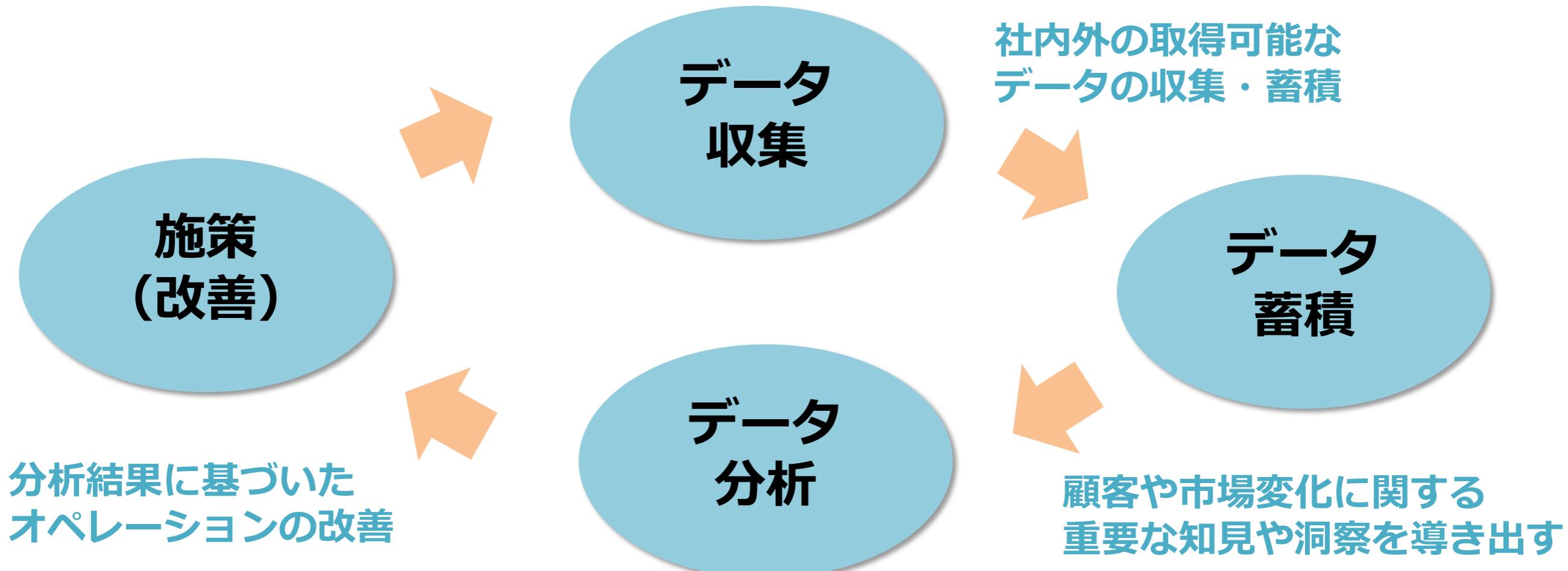
第6回

- 分析の手法の選択

データサイエンスとは

データサイエンスとは、「アルゴリズムや統計などといった情報科学系の理論を活用してデータを分析し、有益な知見を見出すことを追求する新しいアプローチ」。

独立行政法人情報処理推進機構（IPA）・データサイエンティスト協会（著）
「データサイエンティストのためのスキルチェックリスト/タスクリスト概説」より



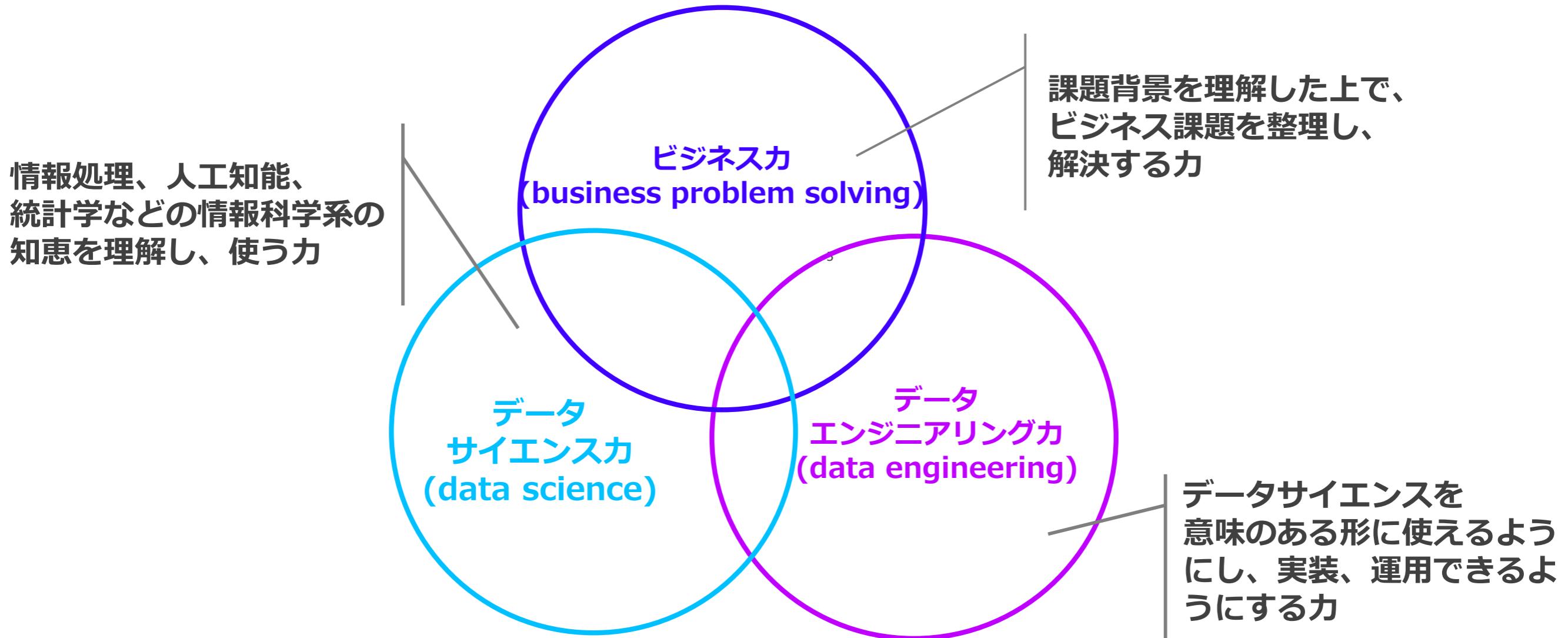
データサイエンティストの定義

データサイエンス力とデータエンジニアリング力をベースに、データから価値を創出し、ビジネス課題に答えを出すプロフェッショナル

* ここでの「ビジネス」とは社会に役に立つ意味のある活動全般を指す

データサイエンティストに求められるスキル

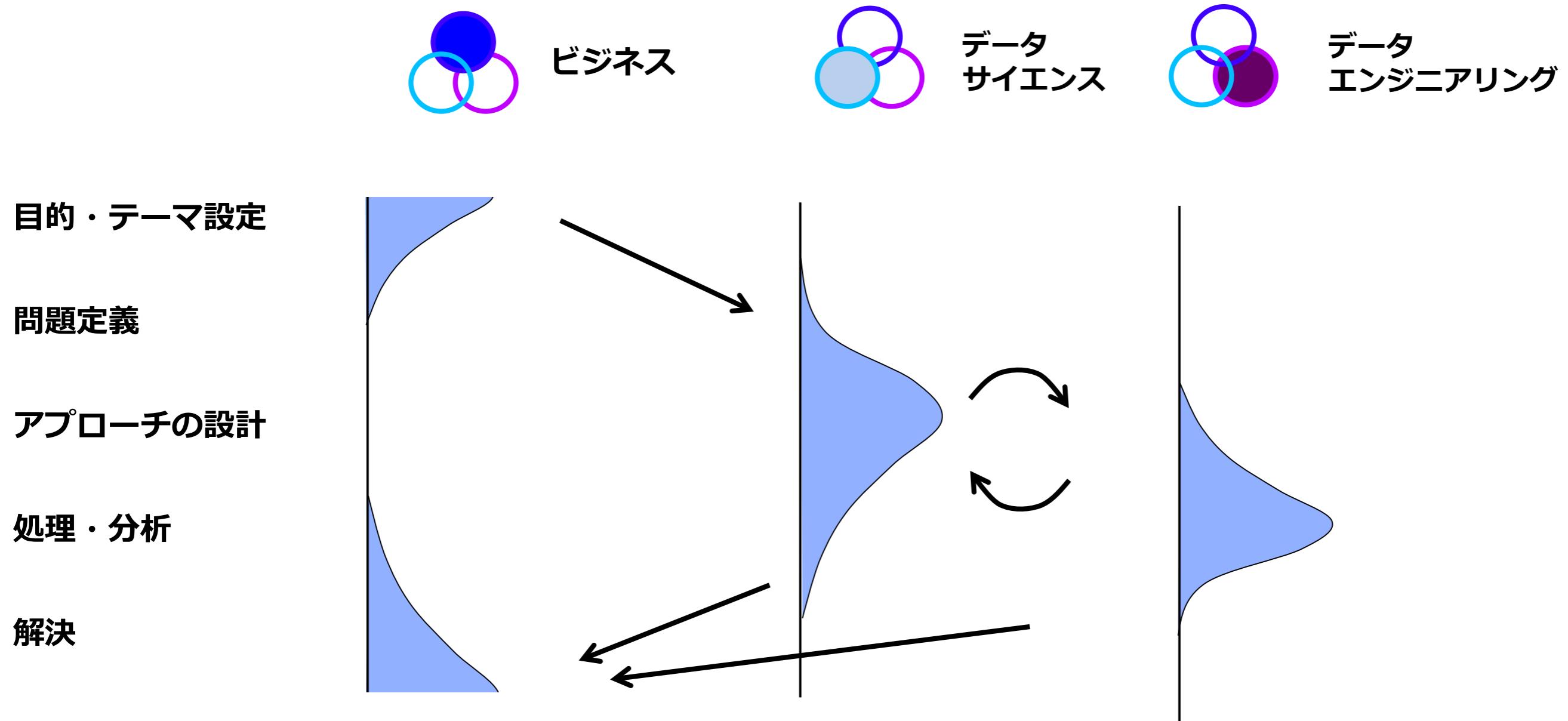
3つのスキルセット



参照：データサイエンティスト協会プレスリリース (2014.12.10)

<http://www.datascientist.or.jp/files/news/2014-12-10.pdf>

課題解決のフェーズによって変わるスキル



データサイエンティストの役割

ビジネス現場において、データサイエンティストは実際に以下のような役割が求められる

データサイエンティストの役割



ビジネス

- ✓ 課題の発見、定義
- ✓ データの入手
- ✓ データにひそむ動向やパターンの発見、意味合い抽出

データサイエンス

- ✓ 最新動向の把握
- ✓ AI開発のためのパッケージ選定
- ✓ データクレンジング
- ✓ データ加工処理
- ✓ データの可視化
- ✓ 分析結果の評価

データエンジニアリング

- ✓ データの保管、管理
- ✓ データ変換
- ✓ 分析プログラムの実行
- ✓ ITのリスク回避

※データサイエンティスト協会スキルチェックリスト★1の項目より抜粋

本講座の狙い

データ分析における基本プロセスを理解する

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

ビジネス現場においてデータ分析を行えるようにする

次回のテーマ

次回は

「データサイエンスの未来と発展」

お疲れ様でした！

社会人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未来と発展

講師名:菅 由紀子

第1週の内容紹介

第1回

- データサイエンスの発展

第2回

- 社会で起きている変化 ~データサイエンスの必要性~

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未来と発展

第5回

- データサイエンスのサイクルと課題解決の進め方

第6回

- 分析の手法の選択

産業におけるデータサイエンス活用への期待

多くの産業において、データサイエンスの有効性に期待が集まる。
単なるスマート化ではない、新たな産業への変革へ。

各産業におけるデータサイエンス活用例

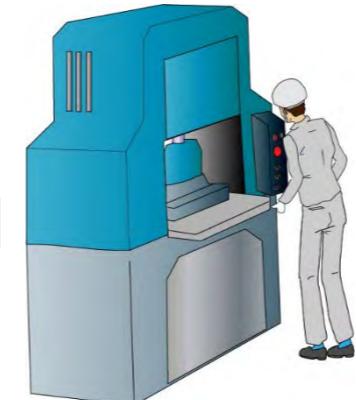
道路・交通

- ✓ 自動運転
- ✓ 交通量の予測



製造

- ✓ 生産量の予測
- ✓ 人員最適化



医療

- ✓ 早期の疾病予測
- ✓ 画像診断



小売

- ✓ 需要予測
- ✓ 顧客分析
- ✓ 販売計画

アパレル

- ✓ 新素材の開発
- ✓ 類似画像解析

農業

- ✓ 農作物の生産管理
(スマート農業)

社会問題の解決

- ✓ CO2削減
- ✓ 災害対策

データサイエンスとAI

- データサイエンスにおける課題解決において、AI技術の活用は欠かせない

AI（人工知能）とは

人間の思考プロセスと同じような形で動作するプログラム、あるいは人間が知的と感じる情報処理・技術といった広い概念で理解されている。



AIが行う主な5つのタスク

分類

予測

認識

パターン発見

生成

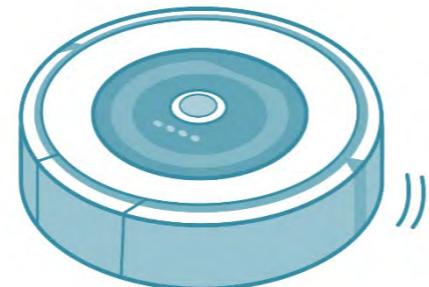
人々の生活に身近なAI

スマートスピーカー



AIによるアシスタント機能を搭載しており、対話型の音声操作に対応しているスピーカー

ロボット掃除機



本体にセンサーを内蔵し、対象の空間の広さ・形状などを把握したうえで自律的に掃除を行う掃除機。

接客ロボット



対人向けの接客を行うことを目的に製造されたロボット。対話による応答が可能なものもある。

自動運転



車など、乗り物や移動体の操縦を人の手を介さず、機械が自立的に行うことができる技術。

使用されている技術の一例

音声認識
自然言語処理 など

位置認識
物体検知 など

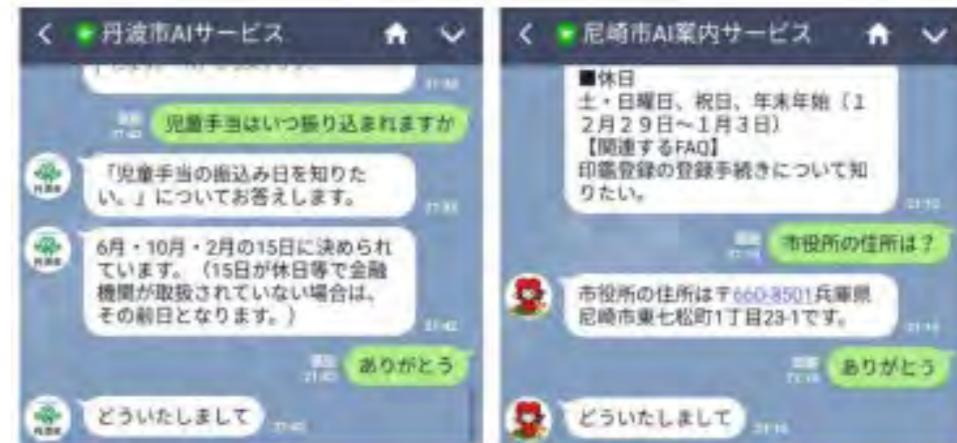
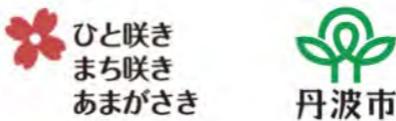
物体検知
画像認識
音声合成
パターン発見 など

危険予測
物体認識
ルート最適化
位置情報特定 など

事例：AIチャットボットによる問い合わせ対応の自動化（尼崎市・丹波市）

概要

LINE公式アカウント上で行政サービスに関する対話システムを運用する実証実験を実施。問合せ傾向の分析から、社会課題の解決に資する人工知能（AI）の開発、並びに利用者の潜在的ニーズを把握し、住民サービスの向上や地域活性化施策の新たな創出に役立てる。



チャットボット導入による 市民/市 それぞれにとっての主なメリット

市民のメリット

24時間対応により休日・夜間の市民サービスが向上

災害時に問合せが急増してもスムーズに解決

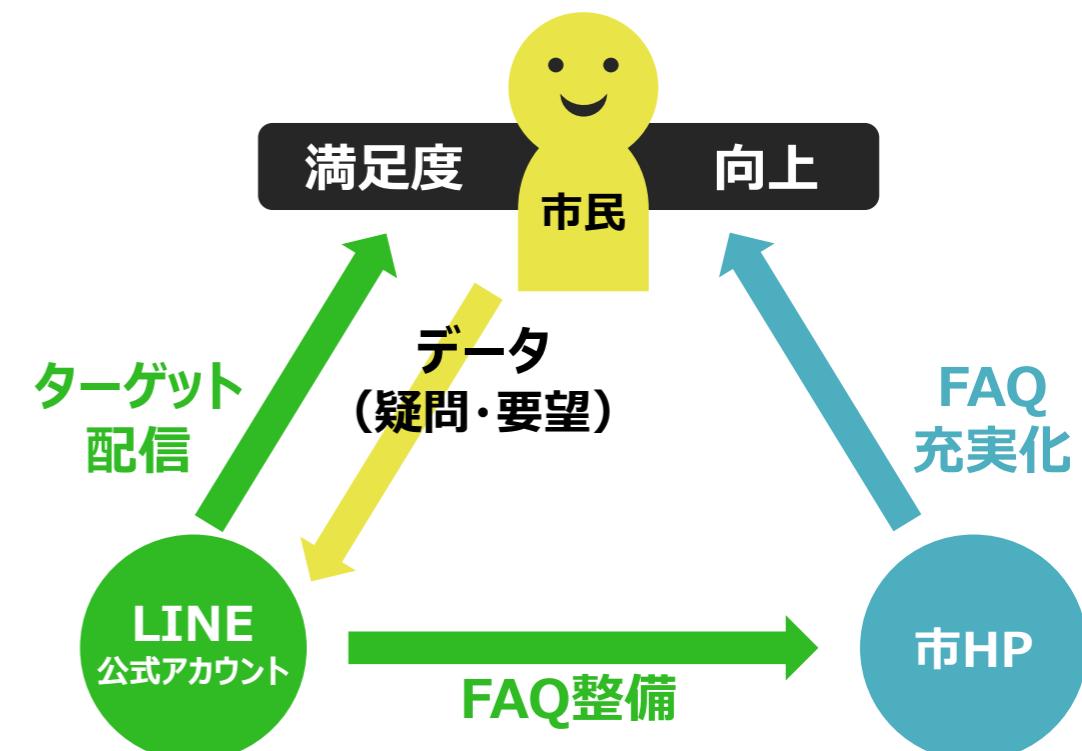
AI案内サービスだからできる相談ごとの増加
(いじめなどのセンシティブな悩み)

既存のHPで不足しているFAQデータの可視化

市民の要望を把握

(近隣地域や自分の特性にあった情報配信を希望する声が多く見られた)

市のメリット



事例：AIを活用し、倉庫運営における人員配置の最適化を実現

概要

日本電気株式会社／東邦ホールディングス株式会社

NECが持つ最先端AI技術と、東邦HDの医薬品管理におけるノウハウを組み合わせ、本来、管理責任者が経験に基づいて行っていた倉庫運営における人員配置等の業務をAIで再現。

顧客の課題・活用背景

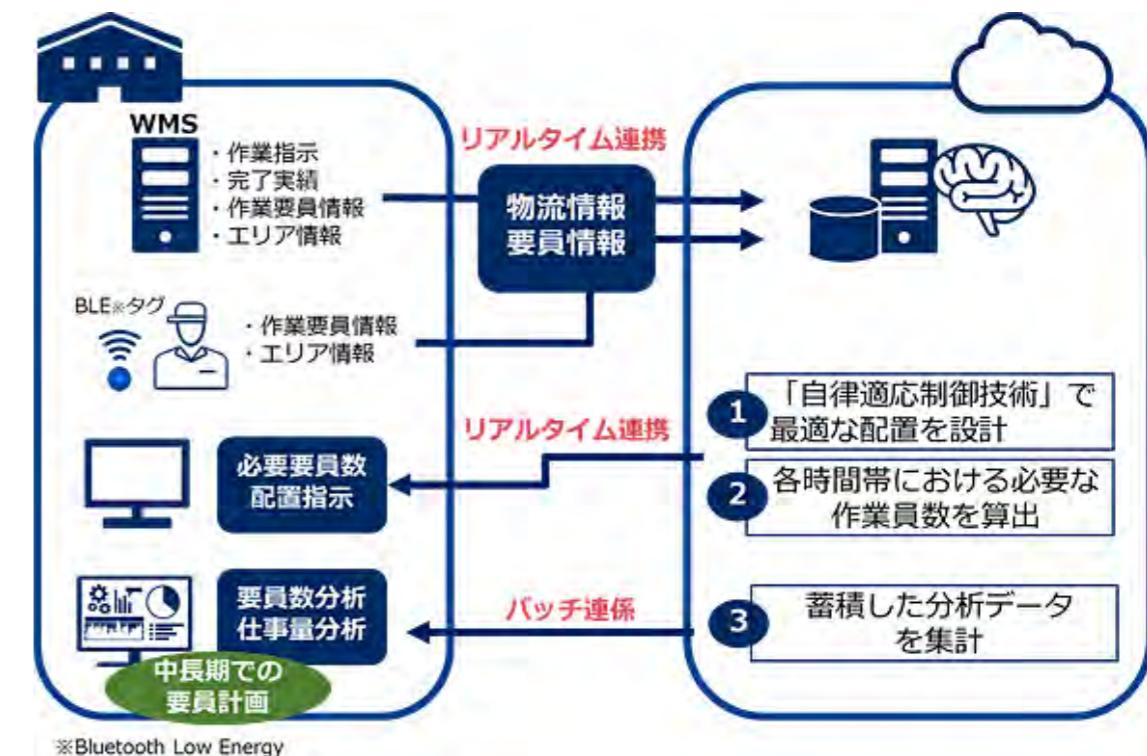
現場の最適な人員配置は管理責任者の経験によって実現していたが、労働力不足や人材の高齢化によって困難に。効率的な業務プロセスの設計や人員配置の適正化が求められていた。

活用データ

作業指示、完了実績/作業要員情報/エリア情報/物流情報など

導入効果

- ✓ 人員配置の最適化で、倉庫内作業は最大20%生産性が向上
- ✓ 作業時間の短縮とコスト抑制に貢献
- ✓ 作業員の現場における状況理解を促進



出典：日本電気株式会社提供
「NEC、東邦HDと倉庫運営における人員配置の最適化をAIで実現」より

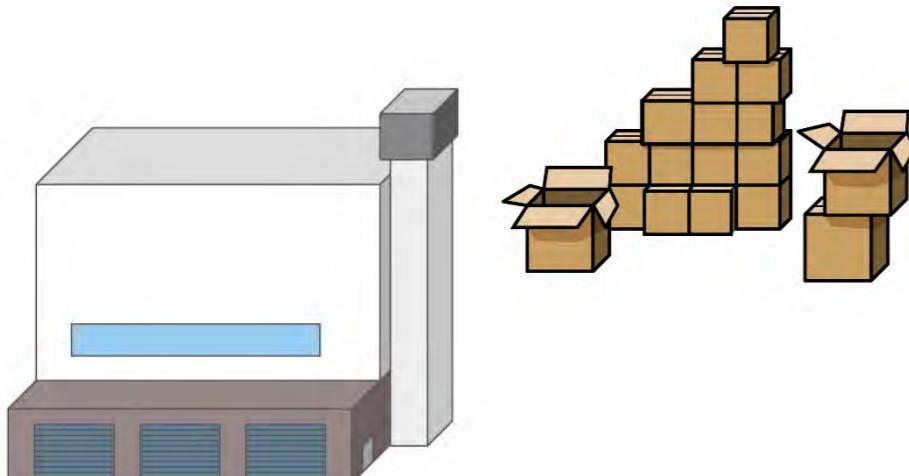
これらの実績による有効性が期待され、NECが保有するICT・DXのアセットは、東邦HD以外にも日本通運をはじめとする様々な企業で導入されている

各業界におけるAI技術を用いた最適化

製造業

原材料の在庫管理をAIで実施。
仕入れコストや在庫リスク削減へ

原材料を安価に入手するほか、売れ残りや機会損失を減少させるため、経済動向・市場の需要等の情報から適正な仕入れ価格の管理と在庫量のコントロールが可能。



医療

患者のモニタリング結果から異常を検知し、重症化を防ぐ

レントゲン画像や超音波など従来のモニタリング結果から異常を検知し、疾病の重症化や発病を未然に防ぐ活用が可能。



流通

世の中のニーズに合わせて商品の販売価格を自動調整

市場動向、実績値のほか天気などさまざまなデータを用いて販売数の予測を実施。常に適正な在庫管理が行えるほか、市場動向に応じた価格の変動により、需要・売上を安定させることができる
(ダイナミックプライシング)



次回のテーマ

次回は

「データサイエンスのサイクルと課題解決の進め方」

お疲れ様でした！



総務省統計局

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:データサイエンスのサイクルと課題解決の進め方

講師名:小野 陽子

第1週の内容紹介

第1回

- データサイエンスの発展

第2回

- 社会で起きている変化 ~データサイエンスの必要性~

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未来と発展

第5回

- **データサイエンスのサイクルと課題解決の進め方**

第6回

- 分析の手法の選択

データサイエンスを用いた課題解決のサイクル

ビジネスにおける問題は、さまざまな要因が複合的に関係している
問題解決にいたるには、仮説検証のためのプロセスを繰り返す必要がある

例) 顧客起因による要因

要因

要因

要因



DATA

例) 環境起因の要因

要因

要因

例) 自社起因の要因

要因

要因

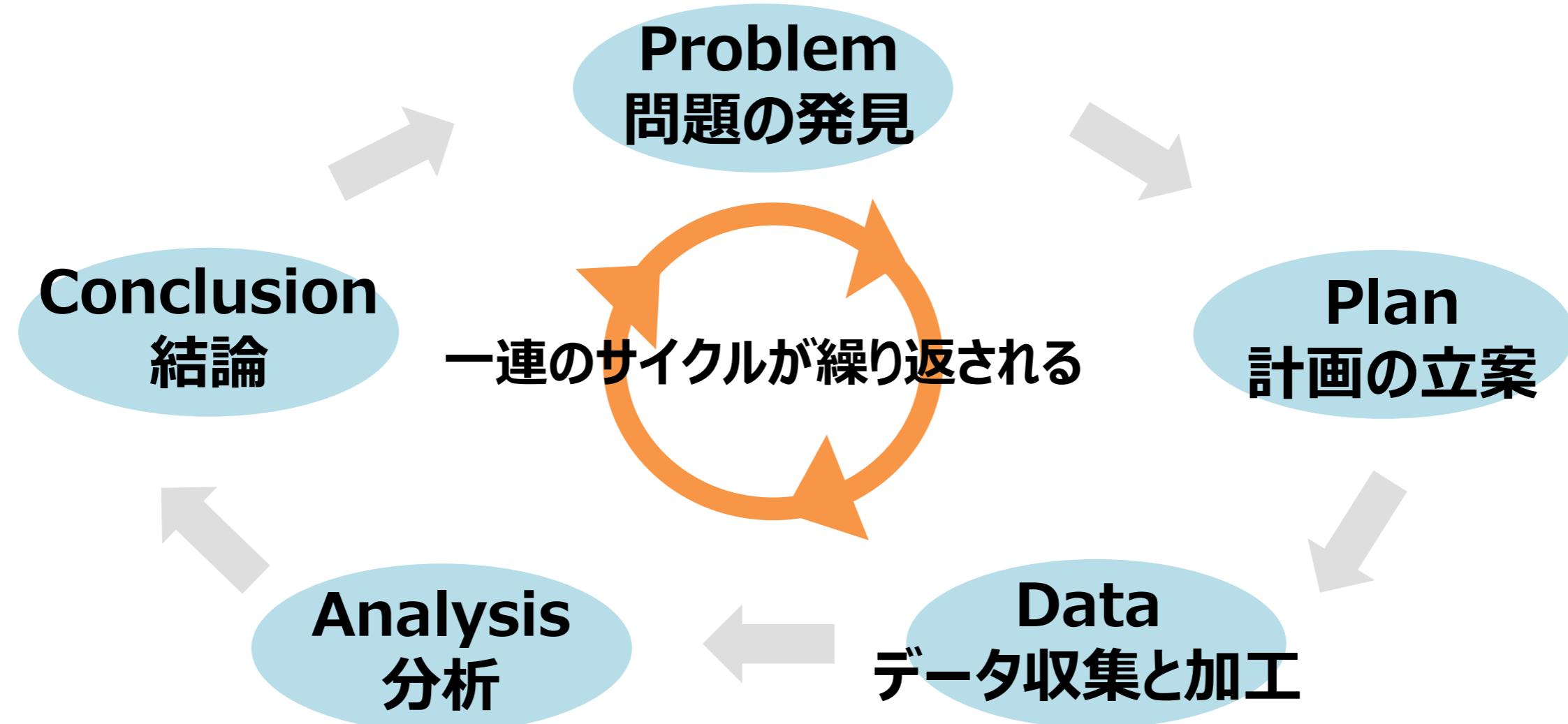
循環型の課題解決プロセスへ

データサイエンスを用いた課題解決のサイクル

PPDACサイクル

Problem (問題)
Analysis (分析)

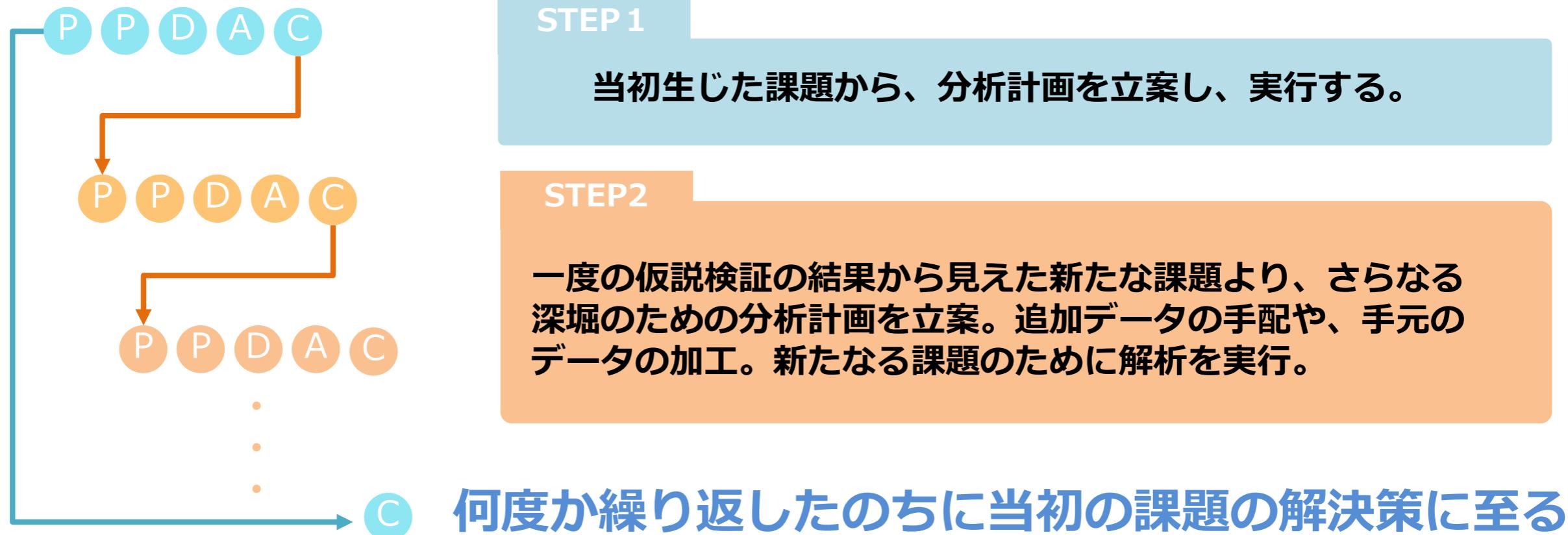
Plan (計画)
Conclusion (結論)



データサイエンスを用いた課題解決のサイクル

- 分析結果からさらなる課題発見につなげ、同様のサイクルを繰り返す。

ビジネス現場における課題解決



Problem (問題の発見)

P P D A C

解くべきビジネスの問題点と、それらを解決に導くための課題を明確化する。分析を行う上でのコンセプトや目標を定義する。

問題の認識

「何が問題か？」を発見・認識する
事業環境における共通の問題点を認識する

- 売上が伸び悩んでいる
- 新商品が売れない

具体的な 課題の明確化

要因を深堀し、解決すべき課題を明確化
問題を構成する要素を明確にし、解決すべき
課題を見つけ出す。

- 新規顧客を獲得する
- 商品の認知度を上げる

目標の明確化

「どのように測るか？」達成時の評価指標を決定
課題解決における目標の設定と、評価方法を具体化。
判定時期と定量的な基準を設定する。

達成すべき目標 : KGI(Key Goal Indicator)
中間指標 : KPI(Key Performance Indicator)



Plan (調査の計画)

P P D A C

課題解決に向けたデータ分析プロセスの設計と
必要データの収集、モデル構築方法を検討する。

仮説の検討

「何があれば課題が解決するか？」
評価指標に影響を与える要素を推定する。

- 20代女性の認知を上げる
 - ポイントカード利用者を増やす
- etc

データ分析手法 の検討

どのようなデータで解き明かすか？
仮説を検証するうえで必要データの種類
とデータ分析手法やモデル要件を検討。

- 年代別認知度の分析
- 購入価格帯の分布を把握

評価・検証 方法の検討

どのような分析手法と評価方法をすべきか？
モデルの構築手法、評価手法、具体的な検証
方法を検討する



Data (データの収集と加工)

P P D A C

分析目的に即したデータを収集し、データの状況把握とデータ分析に必要な事前処理を行う。

データ収集

プロジェクトに必要なデータをどのように集めるか？
分析目的に合わせて利用・収集可能なデータとその収集方法を決定する。収集したデータで分析目的が達成可能か評価する。

- 年代別アンケート
- 顧客属性データ
- 購買データ

データ量

欠損・重複

データ型

表記ゆれ

etc

データ加工

解析に向けたデータの事前準備「前処理」
除外または補正すべきデータのクレンジングの実行。必要十分なデータをそろえるための加工と整形を行う。

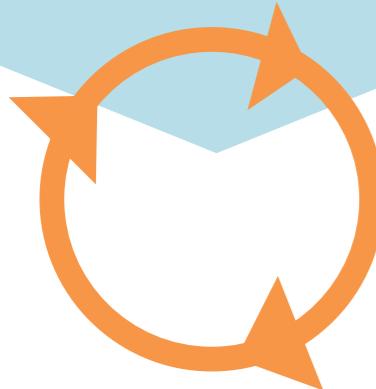


Analysis (分析)

P P D A C

データの性質や関係性を把握し、適切な解析処理を行う。解析や可視化の結果によって、仮説検証を繰り返す。

解析



可視化

基本統計量などを把握し、データの性質や関係性を把握

収集したデータを様々な切り口で集計し、データが本来持つべき意味や活動目的の視点から違和感がないかどうかを見極める。

データ集計・抽出

性質・関係性の把握

予測

グルーピング

パターン発見

最適化

etc

データをグラフなどで表現し、示唆や新たな仮説を見出す

表現したい内容を明確にし、目的に合った表現手法で可視化する。得られた示唆により、新たな仮説とデータ取得の検討を再度行う。

二軸チャート化

多次元可視化

リアルタイム可視化

etc

Conclusion (結論)

P P D A C

立案したモデルの評価から、結論を導く。
課題を解決する施策につなげる。

モデル評価

最終的に採用するモデルを選定。
あらかじめ定めた検証手法により、モデルパフォーマンスの検証を実施。必要に応じて説明変数やパラメータのチューニングを行い、最終モデルを選定する。

分析評価

新たな知識の発見から、業務の組み込み準備
分析結果に基づき、法則性などから新たな知識を見出す。分析プロジェクトの継続要否や、施策の業務への組み込みを検討する。

次回のテーマ

次回は

「分析の手法の選択」

お疲れ様でした！



総務省統計局

社会人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:分析の手法の選択

講師名:小野 陽子

第1週の内容紹介

第1回

- データサイエンスの発展

第2回

- 社会で起きている変化 ~データサイエンスの必要性~

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未来と発展

第5回

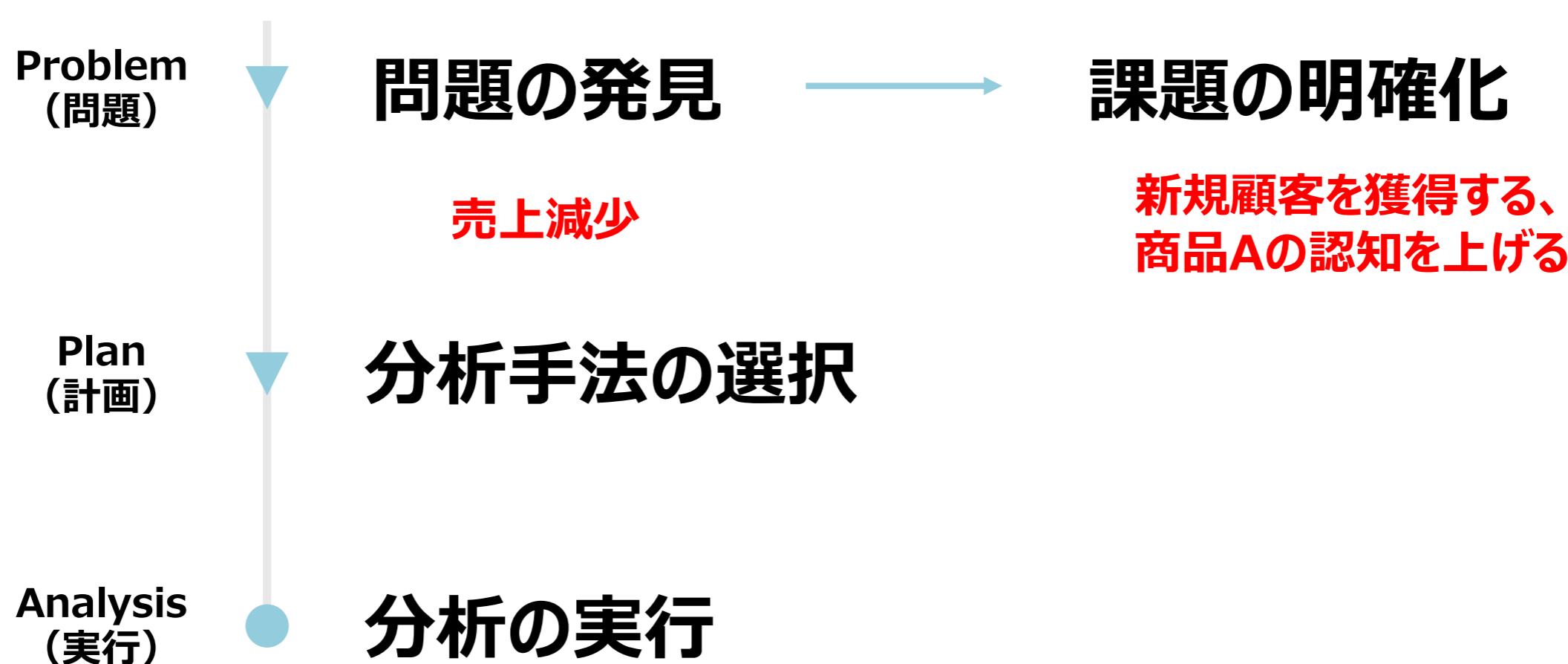
- データサイエンスのサイクルと課題解決の進め方

第6回

- 分析の手法の選択

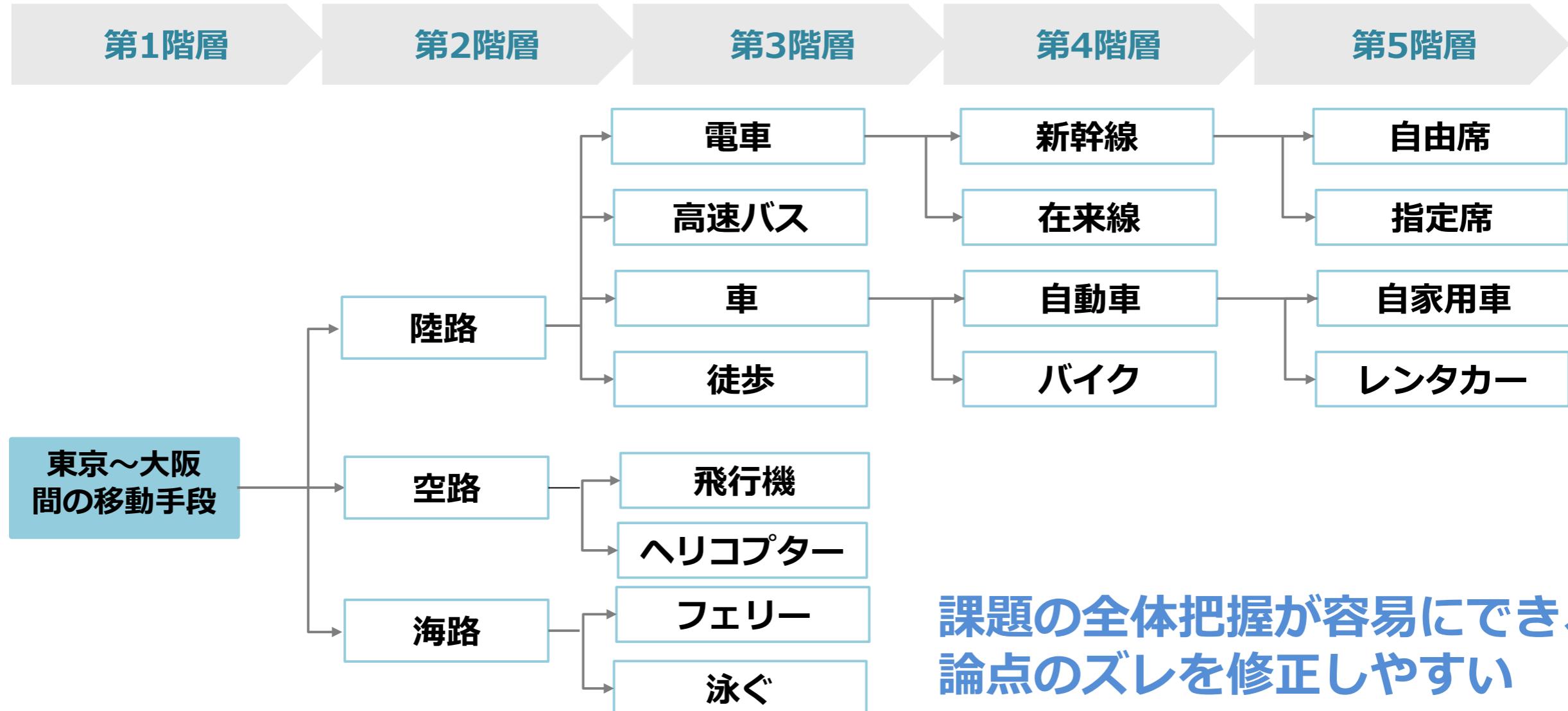
データ分析のプロセス

解き明かすべき問題から課題を明確にし、課題に応じた分析手法の選択を行う。



ロジックツリーについて

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して考える思考技法



Problem (問題の発見)

例題

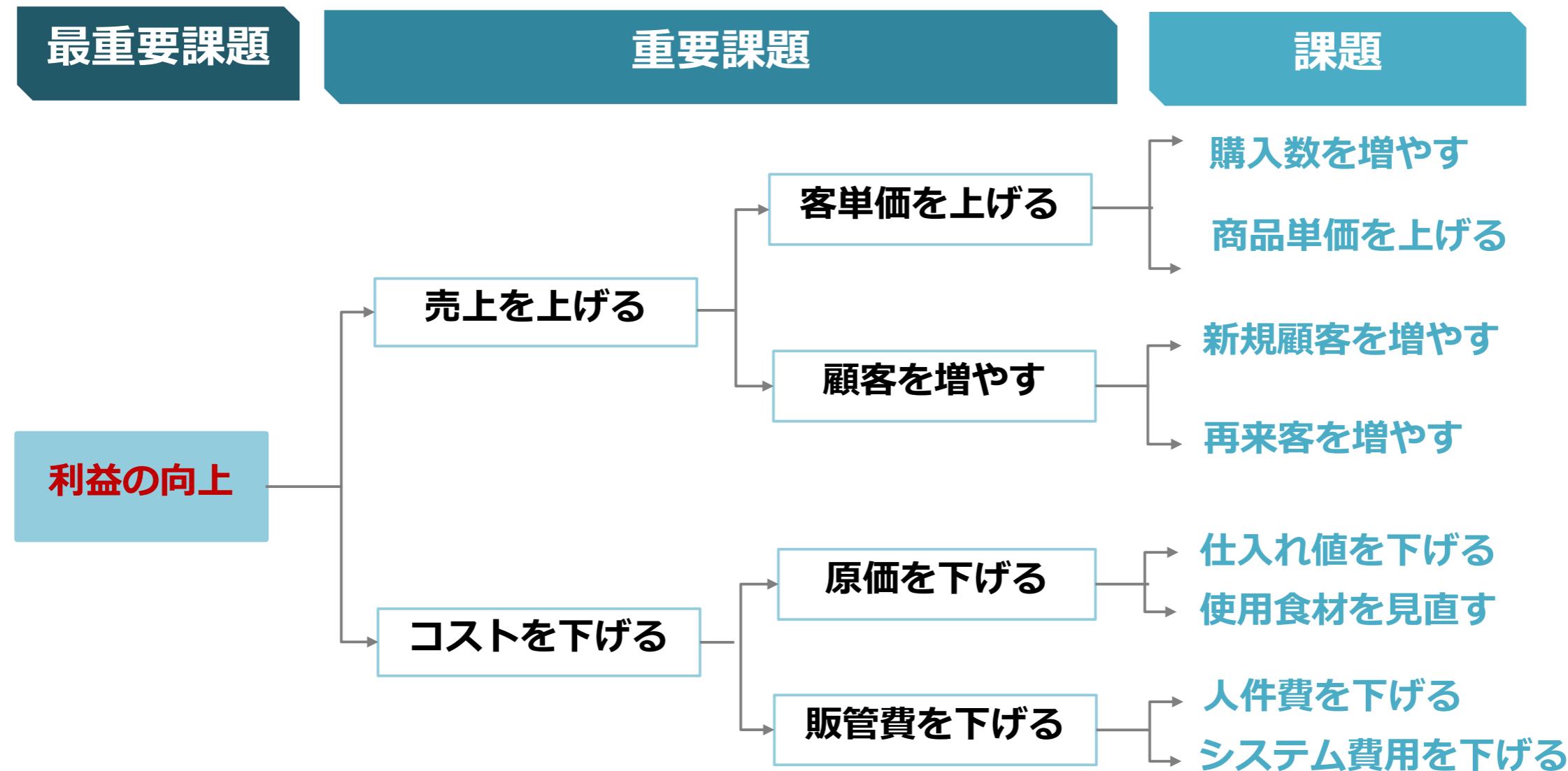
あなたはある飲食店のオーナーです。
仕入れている食材の高騰などが原因で
お店の利益が低下しています。
今後の継続的な店舗運営のために
お店の利益を増やしたいと考えています。



問題解決のためのKGI：店舗の利益

ロジックツリーを活用した方針設計

最重要課題：利益を向上させる



課題の優先順位決め

各要素の現状確認をし、課題に優先順位をつける



重要性	実現可能性
△	○
△	×
○	○
○	○

目的別分析手法の選定

目的

性質・関係性の把握

推定・検定

予測・識別

パターン発見

分析手法の例

基本統計量、クロス集計、相関分析 など

t検定、カイ二乗検定 など

回帰分析、時系列予測、ニューラルネットワーク など

クラスター分析、k近傍法、アソシエーション分析 など

次週のテーマ

次週は

「分析の概念と事例」

お疲れ様でした！