# The Potential and Misconceptions of Data Science

2024.12.6

PEOPLEDOT.

# Company Overview

**Company Name**

Peopledot (Formerly known as Datamix)

**Location**

Head Office : 2nd Ishizaka Building 2F, 2-44 Kanda Jinbocho, Chiyoda-ku, Tokyo

Tokyo Office : PMO Jinbocho 2F, 2-10-4 Kanda Jinbocho, Chiyoda-ku, Tokyo

**Established**

Feb 1, 2017

**Capital**

160,999,400 **JPY**

**Board of Directors**

| | |
|---|---|
| President & CEO | Yosuke Katada |
| Director & COO | Toshiki Kobayashi |
| Director | Yurika Ishii |
| External Directors | Daishi Matsukura |
| | Yohei Sakonju |
| | Ryo Nagai |

**Major Shareholders**

Innovation Engine
NTTファイナンス
insource
株式会社HCSホールディングス
Speee
GROUND
KIYO LEARNING
滋慶学園グループ
Kips
AI CROSS
FE Invest Co.,Ltd.

# Our Business

## AI & Data Science Division

- Consulting Service
- Product Development

## Business Insight Division

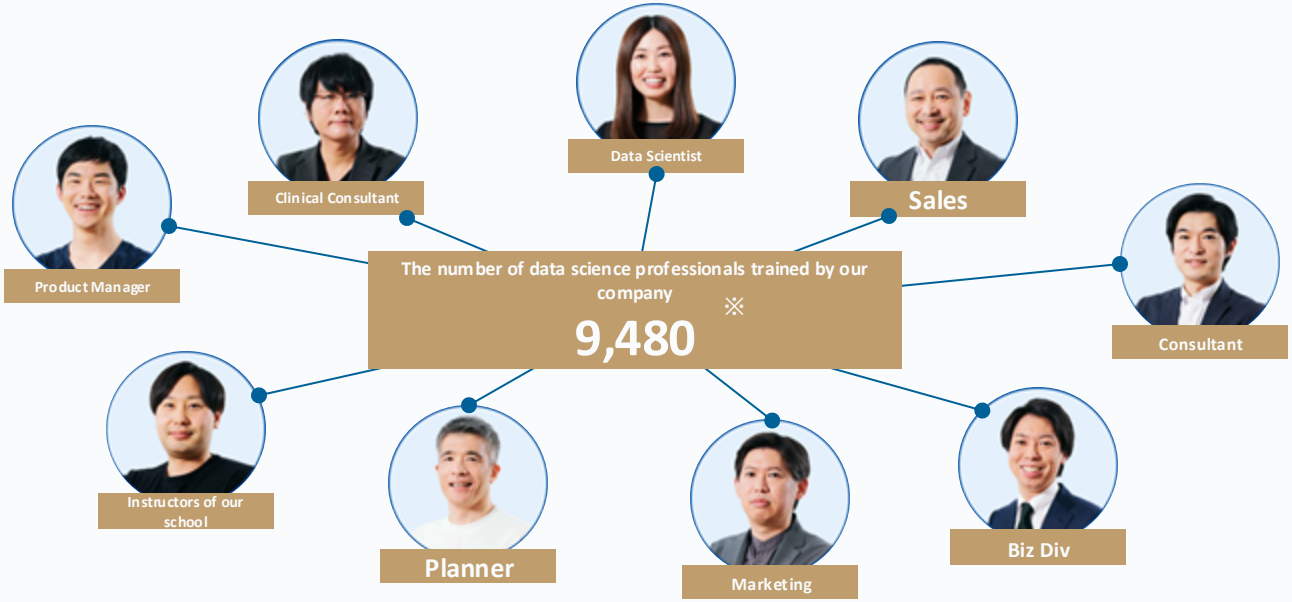
- Corporate Training
- School of Data Science (datamix)

## Laboratory

- A community of data science professionals who have graduated from our school

# Our Clients

# Our strength : a community of data professionals with diverse backgrounds.




※As of December 26, 2023, the number of participants in in-person and live classes (excluding e-learning, recorded viewing, and free event attendees)

# Self-Introduction

## Ryutaro Higuchi

(Group Chief of AI & DS Solution Division)

- Worked in sales, marketing, and business development at HR service company and a sports media company.
- After joining PeopleDot (formerly DataMix), completing our data scientist training program. Currently, in addition to sales for corporate clients, also engaged in training and consulting as an instructor.

**Speaking Engagements**

- Ad:tech 2016 Kansai
  “Developing Loyal Customers through the Intersection of Sports and Digital”
- Ad:tech 2017 Kyoto
  “The Business Potential of Native Advertising”

# The Potential and Misconceptions of Data Science

- What is Data Science?
- How is Data Utilized in Business?
- What are the Processes and Skills Required for Data Utilization?
- How to Acquire Data Science Skills?

# What are your thoughts on the following two cases?

## Case 1: Implementation of an Automated Clinical Trial Data Analysis System

**Background**

CRO Company A received a request from a pharmaceutical company to analyze data from the clinical trial phase.

Previously, patient data was manually aggregated and analyzed over an extended period to create clinical trial reports. Leveraging their expertise in data management and analysis, they developed an automated analysis system to efficiently process clinical trial data.

## Case 2: Implementation of a New Analysis Tool Capable of Handling Large Volumes of Data

**Background**

CRO Company B decided to implement a new analysis tool to enhance the content of reports provided to clients.

To collect more detailed patient data, they significantly expanded the survey items, gathering extensive information from participants. They then attempted to analyze the large volume of data using a popular, newly implemented analysis tool that enables rapid data processing.

# What are your thoughts on the following two cases?

## Case 1: Implementation of an Automated Clinical Trial Data Analysis System

| | |
|---|---|
| Result | The analysis of clinical trial data was completed in half the usual time, significantly shortening the trial period. This allowed the client pharmaceutical company to closely monitor trial progress, enabling faster decision-making and accelerating the product's time-to-market. As a result, Company A and the client pharmaceutical company decided to establish a long-term partnership. |

## Case 2: Implementation of a New Analysis Tool Capable of Handling Large Volumes of Data

| | |
|---|---|
| Result | The volume of collected data was far greater than anticipated, resulting in delays in both data collection and analysis. Additionally, much of the information gathered did not directly contribute to assessing the efficacy or safety of the trial drug, leading to challenges in organizing and interpreting the data. Consequently, despite significant investment of time and resources, they were unable to obtain useful insights regarding the drug’s effectiveness, causing delays in the overall schedule. |

# The purpose of data science

Many people view data science as a collective term for fields such as statistics, machine learning, AI, and programming. However, data science is not merely a collection of techniques or knowledge. It is a tool for identifying challenges through data and solving those challenges to achieve business results.

# The purpose of data science

Many people view data science as a collective term for fields such as statistics, machine learning, AI, and programming. However, data science is not merely a collection of techniques or knowledge. It is a tool for identifying challenges through data and solving those challenges to achieve business results.

**So, how can we utilize data to identify and solve business challenges, ultimately driving business outcomes?**

Q

# First of all, what do we mean by "business outcomes"?

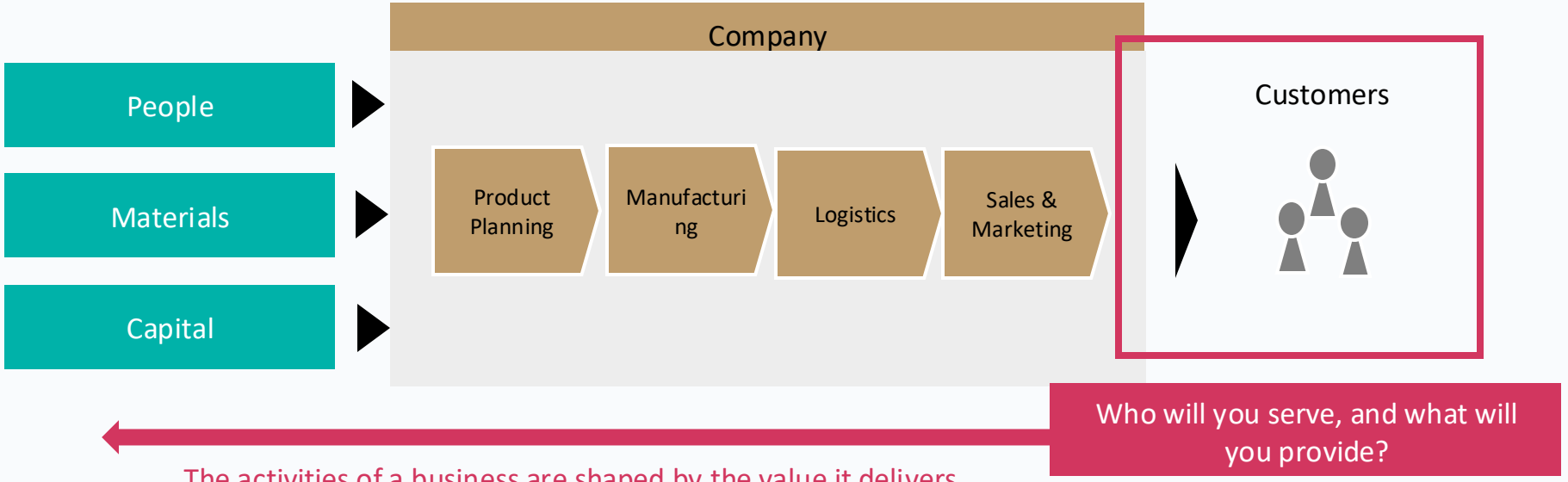
Q

If you were told to "start a business right now," what would you consider? And, what is the most important question to ask?

- Who should I hire?
- Who will be the customers?
- What should I sell?
- .
- .
- .
- Where should I source funding?
- What advertising methods will be effective?
- Where should the office be located?
- .
- .
- .

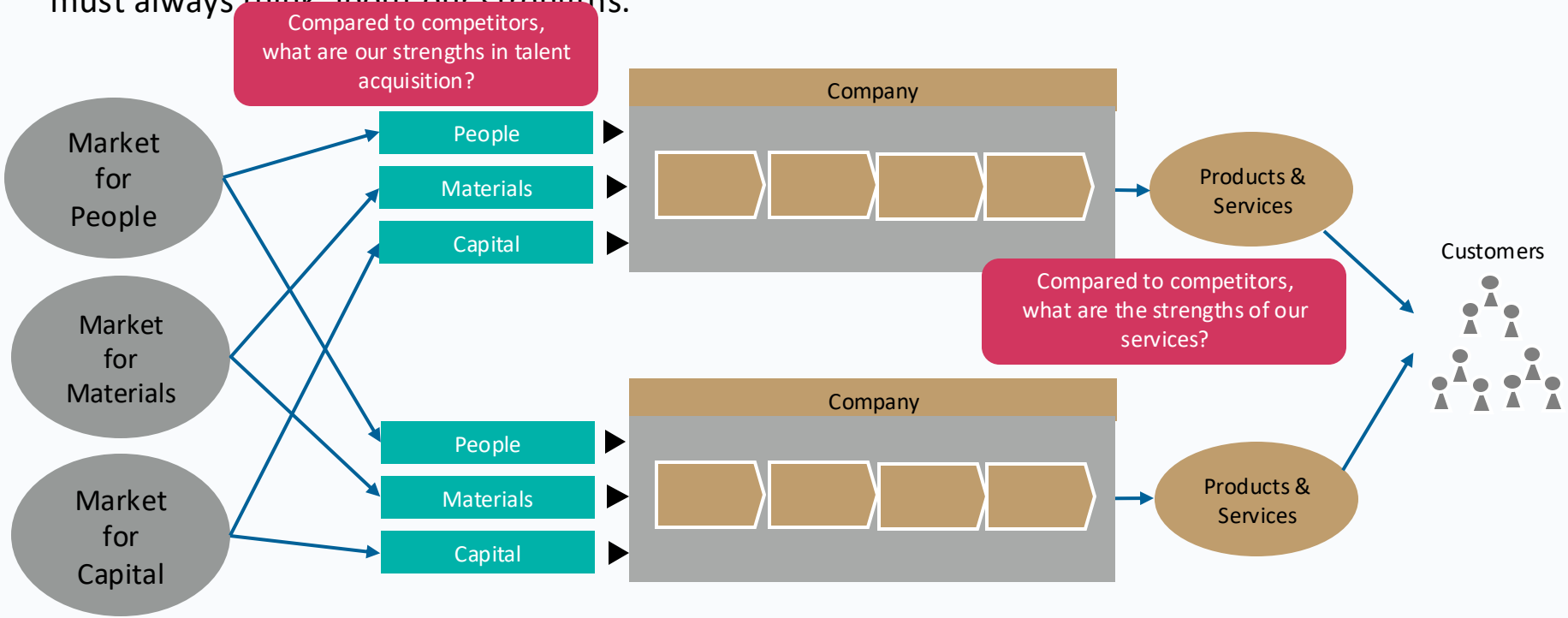
# The starting point of any business is to consider the value it provides to its customers.

The foundation of business lies in identifying “who” — the target customers — and “what” — the value — you will deliver.

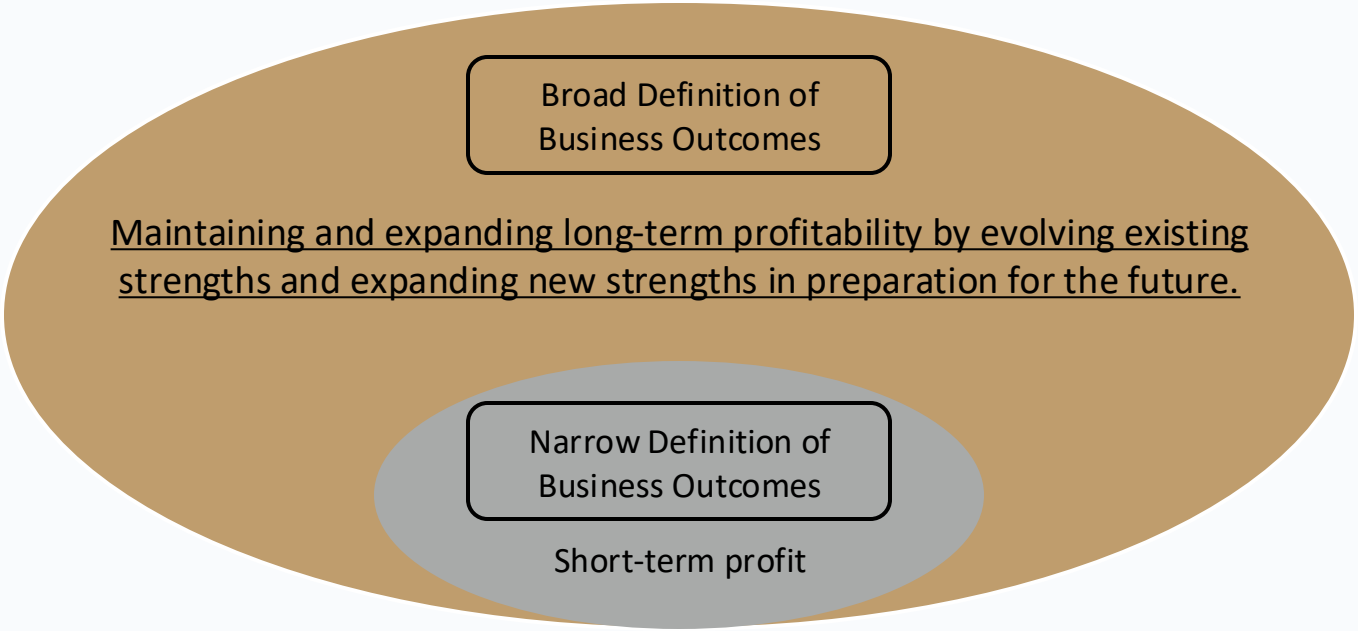


# To generate profit, it is essential to consider "what our strengths are."

Not only in acquiring customers but also in terms of securing people, materials, and capital, we must always think about our strengths.



# What exactly are business outcomes?



Q

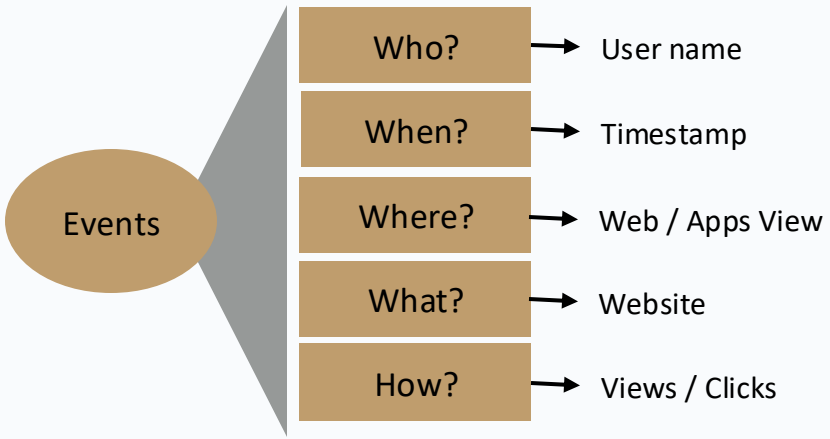
# What does it mean to utilize data?

# What is Data?

Data is a snapshot of events, consisting of both intentionally collected data and data gathered without specific intent.

## What is Data?

Data is a snapshot, an object representing an event of interest. Data alone has no inherent value.




## Examples

**Data gathered without specific intent**

- Mail
- Transaction history
- Invoices
- Sales reports
- Resumes

**Intentionally collected data**

- Web or app user behavior logs
- Customer demographic information
- Customer product reviews

## Data utilization can generally be divided into three categories.

| | |
|---|---|
| Decision Support | Use data to understand the current situation, formulate hypotheses, and validate outcomes. |
| Automation and Labor Reduction | Leverage machine learning and AI to automate parts of operations, reducing costs. |
| Personalization | Use machine learning and AI to provide tailored recommendations, aiming to increase customer lifetime value (LTV). |

# Data utilization begins with a question.

Questions

You

IT and mathematical skills are only valuable when guided by meaningful questions.

Events

Interpritation

Data Extraction & Analysis

Data Collection

# What is data utilization?

Data utilization is the process of supporting decision-making, automation, and personalization based on questions rooted in the strengths of one’s business.

It contributes not only to short-term profits but also to mid- to long-term profitability by deepening and expanding the business's core strengths.

# The Data Utilization Process

The process of data utilization consists of three steps: identifying the problem(Discovery), solving the problem(Solution), and implementing the analysis results(Implementation).

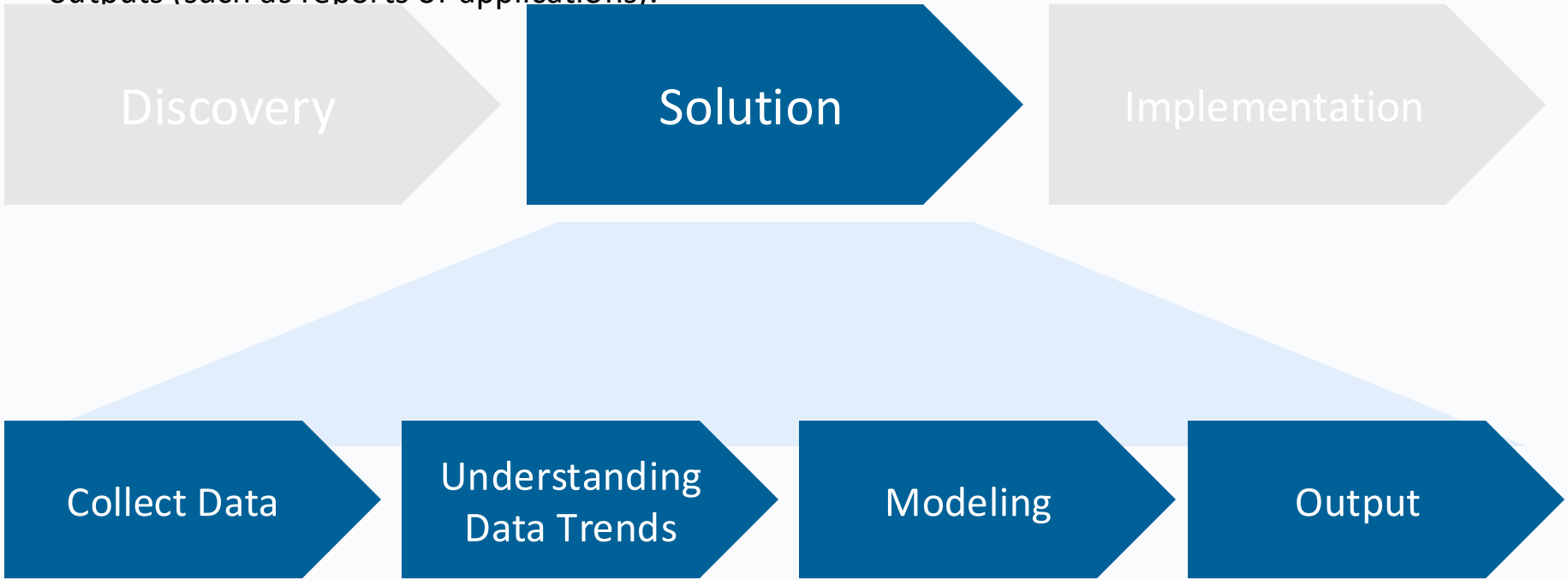
# The Data Utilization Process

## How can we identify problems?

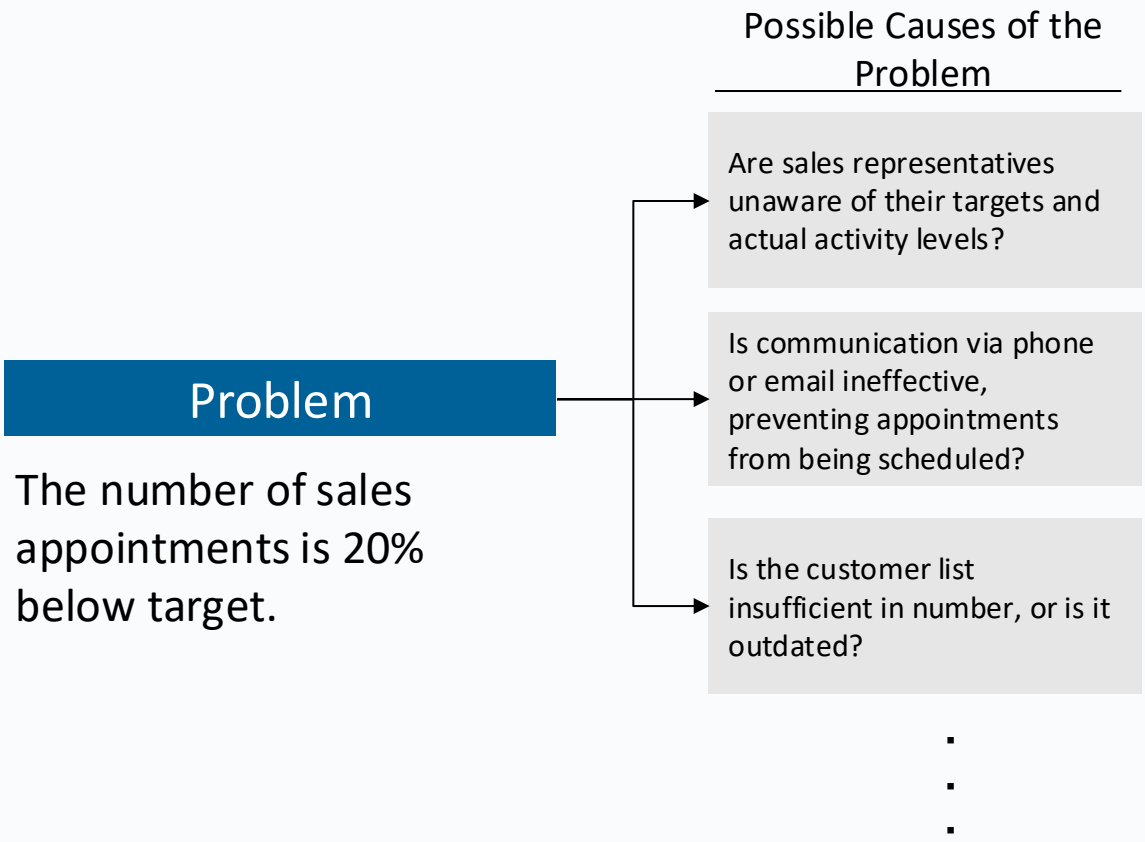
| | |
|---|---|
| In-Depth Understanding of the Business | ● Possesses detailed knowledge of customers and business processes<br>● Can quantitatively model the business |
| Current Situation Assessment through Data | ● Understands techniques for data visualization<br>● Able to monitor KPIs and formulate hypotheses on causes when deviations are identified |
| Problem Definition | ● Capable of structuring and understanding problems systematically<br>● Able to gather stakeholders and facilitate discussions |

# Steps to Solve Problems Using Data

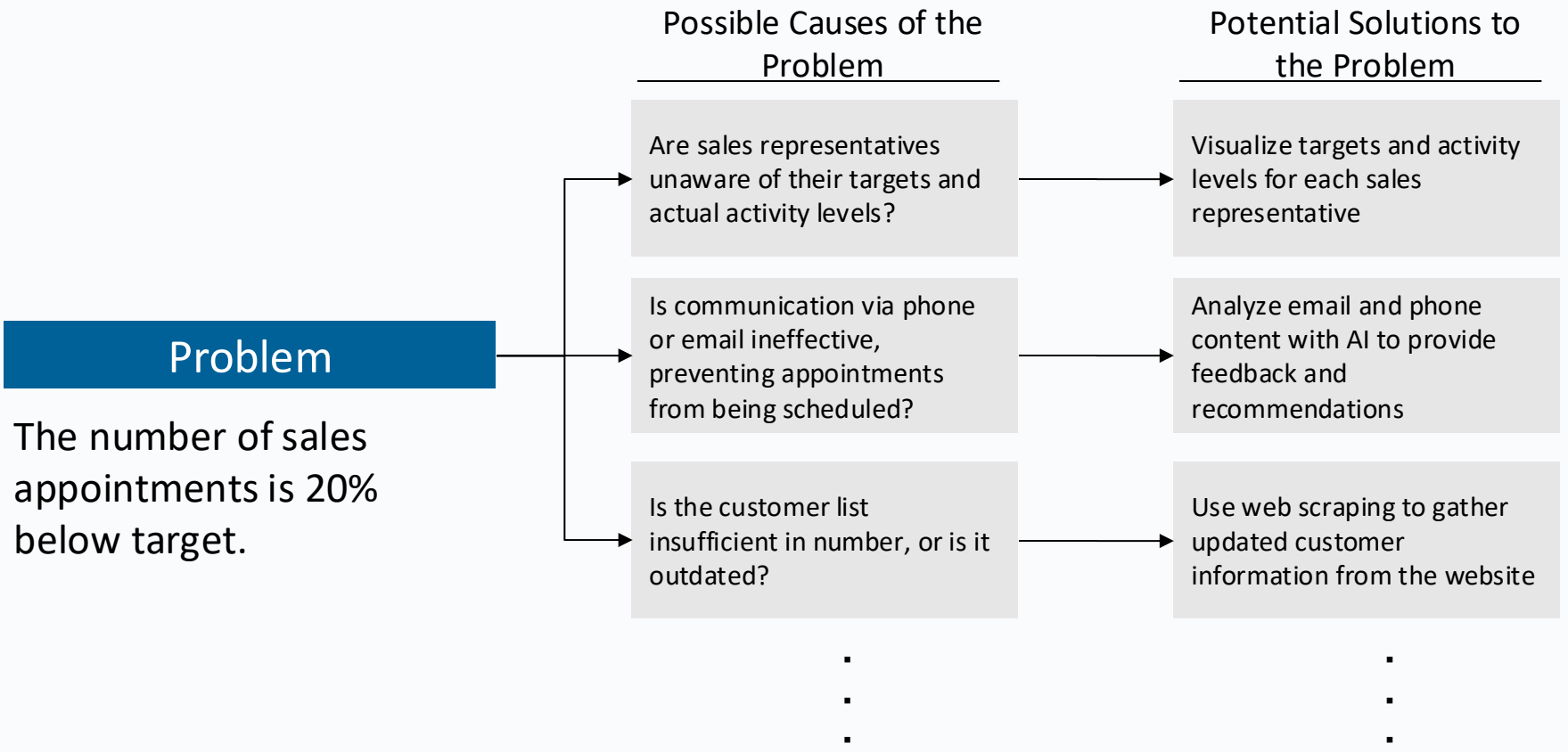
Collect data, understand trends through exploratory data analysis, perform modeling, and create outputs (such as reports or applications).



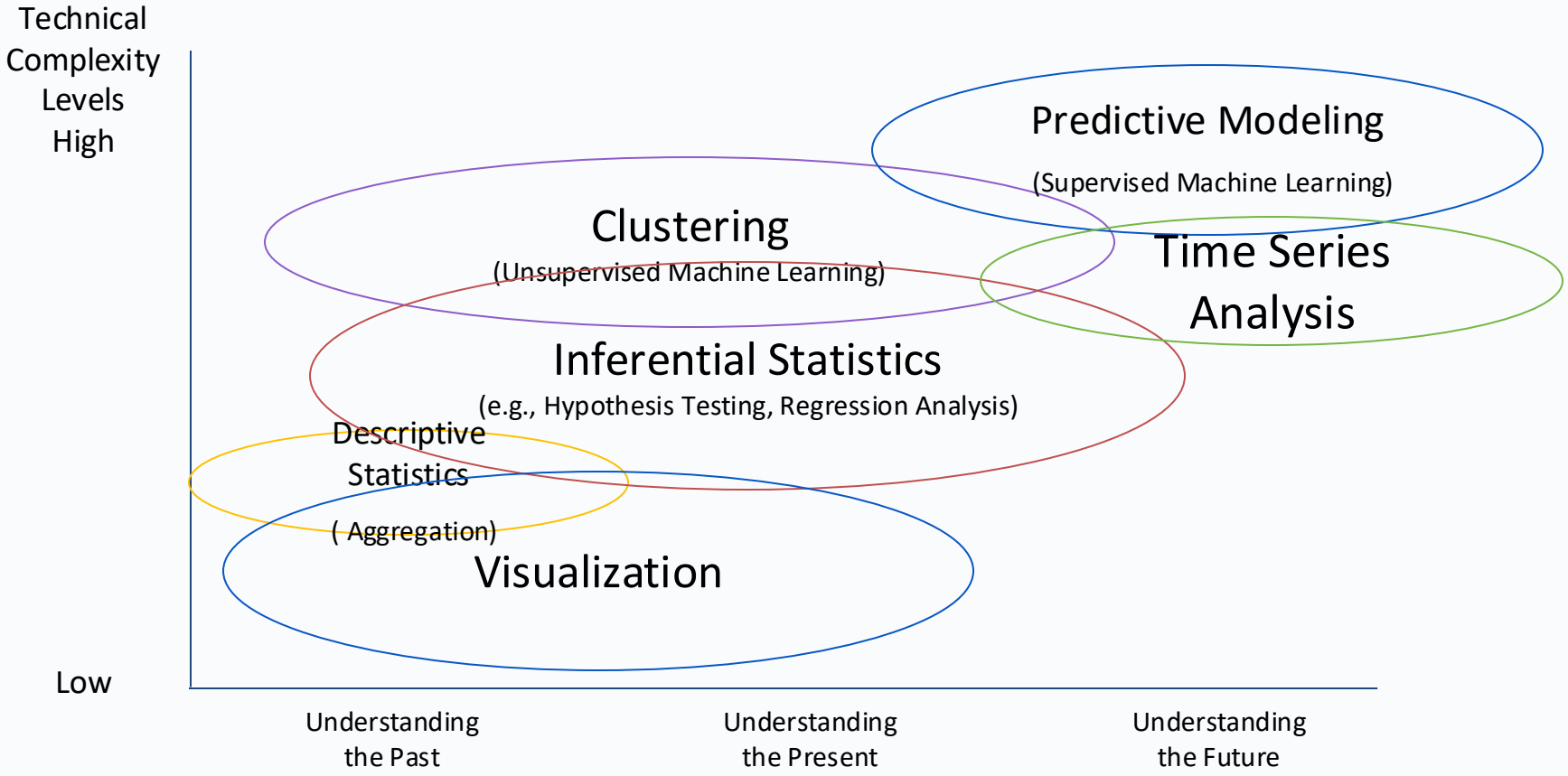
# To solve a problem, one must be able to identify causes and consider solutions.



# To solve a problem, one must be able to identify causes and consider solutions.



# Data Science Toolbox for Problem Solving



# Implementation

# How to Ensure Data is Used Effectively

| | |
|---|---|
| Understanding the Target User | ● Grasp the context in which they will use the data<br>● Understand their level of data literacy |
| Explanation & Presentation | ● Use their language (avoid jargon)<br>● Ensure clarity and logical flow |
| Follow-up | ● Simply saying “please use it” won’t lead to adoption; provide ongoing support until it is fully integrated into their workflow |

# Overview of Data Science Skills and Knowledge

| | Discovery | Solution | Implementation |
|---|---|---|---|
| Business | • Business Understanding<br>• KPI Understanding<br>• Problem Definition | • Client Interaction<br>• Report Creation | • Operational Insight<br>• Communication |
| Mathmatics | | • Statistics<br>• Machine Learning | • Statistics (Effectiveness Validation) |
| IT | • SQL<br>• BI Tools | • SQL<br>• BI Tools<br>• Cloud Services | • SQL<br>• BI Tools |

**Data science skills and knowledge are broad and must be applied contextually based on the situation.**

## Let me emphasize once again…

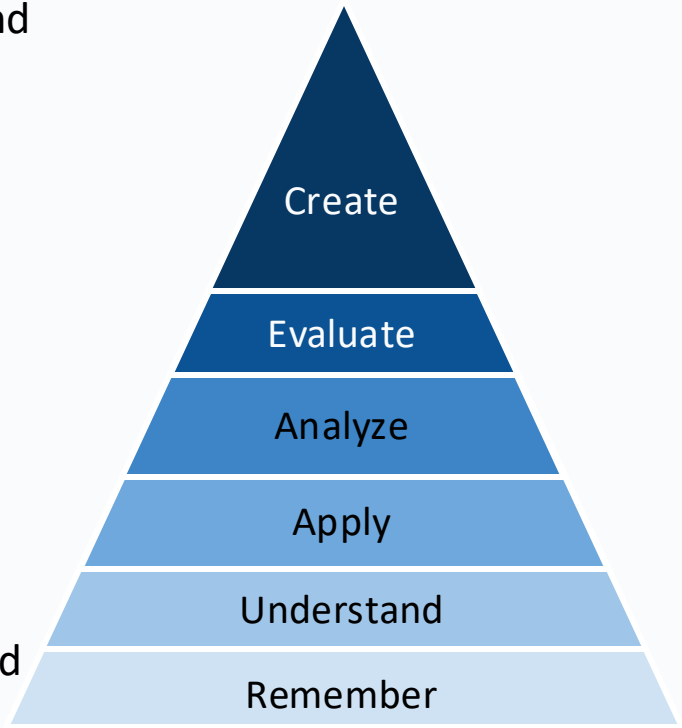
Data science projects are
“context-dependent.”

| | |
|---|---|
| Negotiations with Clients and Team Discussions | • Your approach may vary based on the flow of conversations with clients or team members. |
| Responding to Unexpected Issues | • The response required changes depending on the cause and impact of the issue. |
| Sports like Football or Baseball | • Actions to be taken vary depending on the tactics and game situation of the opposing team. |

# Bloom's Taxonomy

How to acquire skills?

Higher-Level Learning and Cognitive Skills




Lower-Level Learning and Cognitive Skills

## Experiential Learning

Involves reflecting on and internalizing one's own experiences to deepen understanding, often through self-reflection and feedback. It is also context-dependent, adapting to various situations.

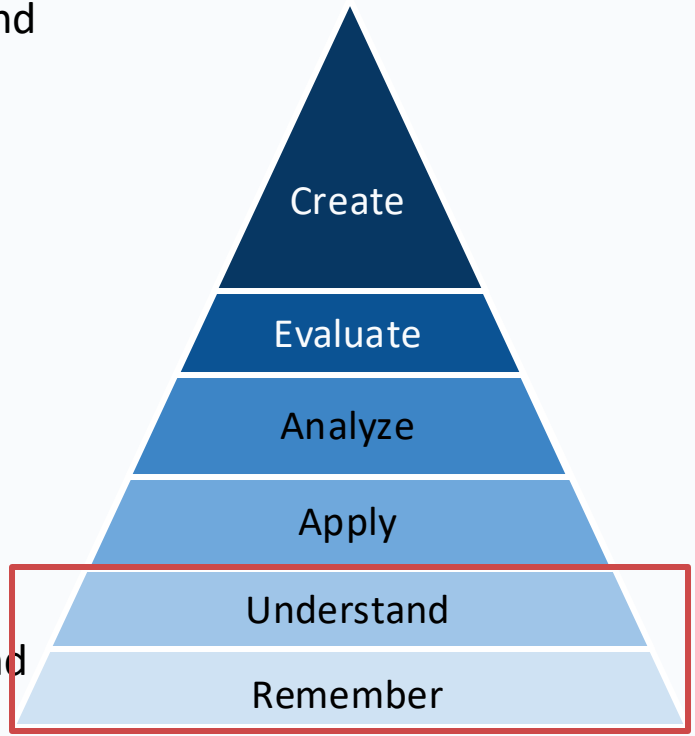
## Passive Learning

A format where information is received in a one-way manner, primarily focused on the intake of knowledge.

# Bloom's Taxonomy

How to acquire skills?

Higher-Level Learning and Cognitive Skills



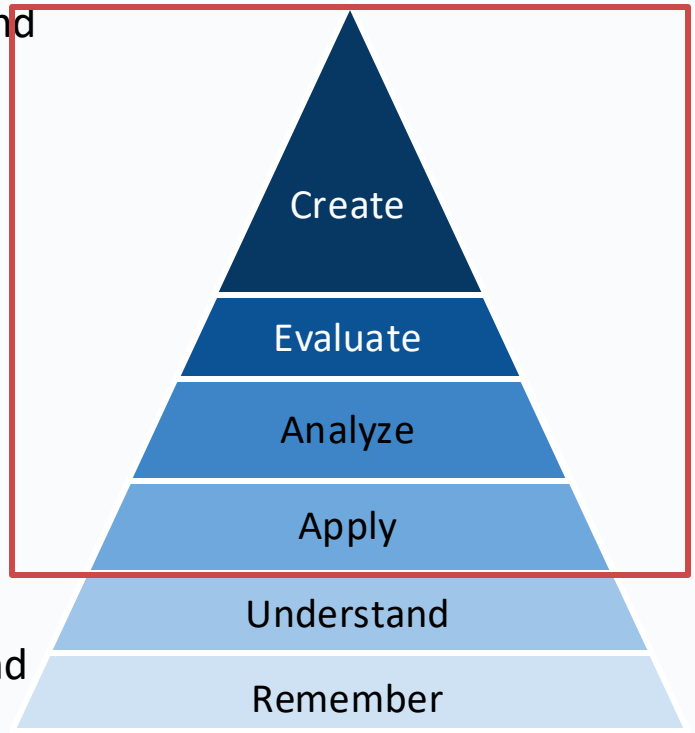

e-Learning and Books

Lower-Level Learning and Cognitive Skills

# Bloom's Taxonomy

How to acquire skills?

Higher-Level Learning and Cognitive Skills




Lower-Level Learning and Cognitive Skills

- Participation in hackathons and competitions
- Attending schools or graduate programs
- Taking on a second job or freelance work

# Summary

- Data science goes beyond techniques in statistics and machine learning; it is a tool for identifying and solving problems through data, driving both short- and long-term business outcomes. Achieving these outcomes requires meaningful questions derived from a deep understanding of customers and operations.
- The data utilization process—finding problems, solving them, and ensuring their use—requires a broad, adaptable skill set. Continuous learning and flexibility are essential to effectively connect data insights with business success.

Thank you for your attention.