

# Introduction to Data Models

CPE 232: Data Models

*Dr. Sansiri Tarnpradab*

*Department of Computer Engineering, KMUTT*

# Welcome!

- Syllabus
- TAs' Office Hours
- **Remarks:**
  - Be Proactive
  - Teaching → Guiding
- Let's get to know you!

# Outline

## Introduction to Data Models

- Data
- Data Modeling
- Data Visualization

# Data

# Where data come from



# Significance

## The Economist



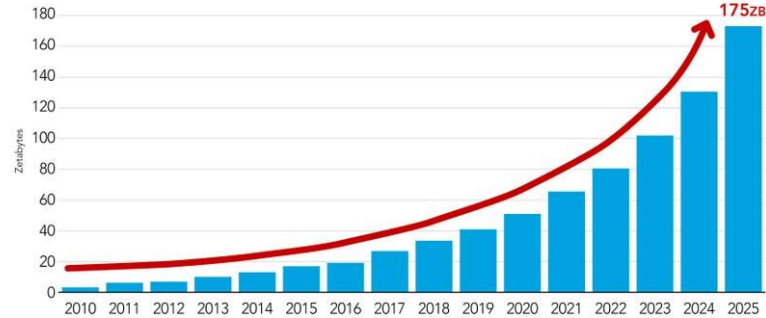
Ref: <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>



# Quantitatively speaking..

## HOW THE AMOUNT OF DIGITAL DATA IS INCREASING

Annual size of the global data sphere 2010–25



Source: IDC Global DataSphere, November 2018

[accaglobal.com/machine-learning](http://accaglobal.com/machine-learning)

It is estimated that around

**90%**

of all the digital data in the world has been created since 2016



Think Ahead

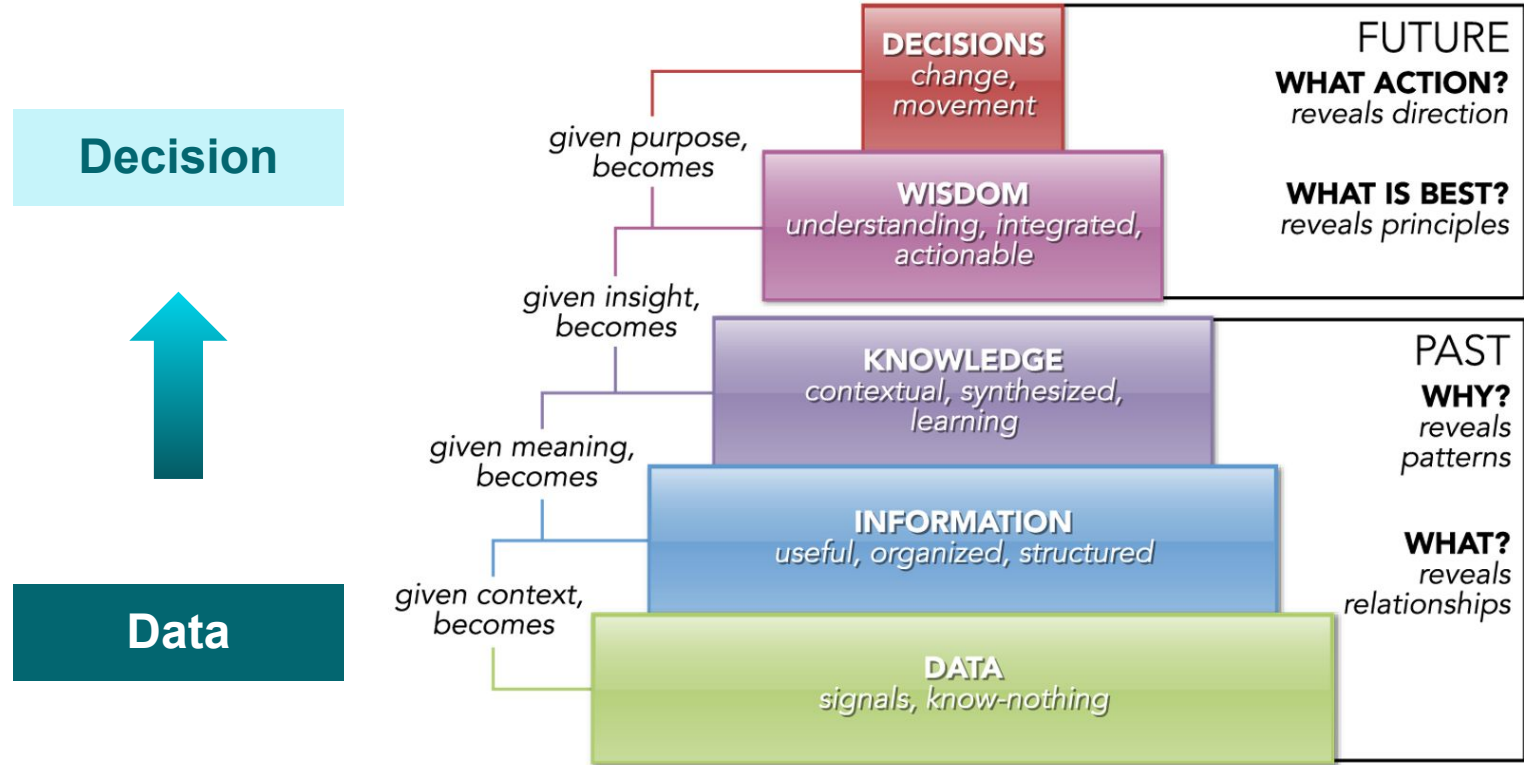
Ref (left): <https://www.domo.com/data-never-sleeps>

Ref (right): <https://www.accaglobal.com/gb/en/professional-insights/technology/machine-learning.html>

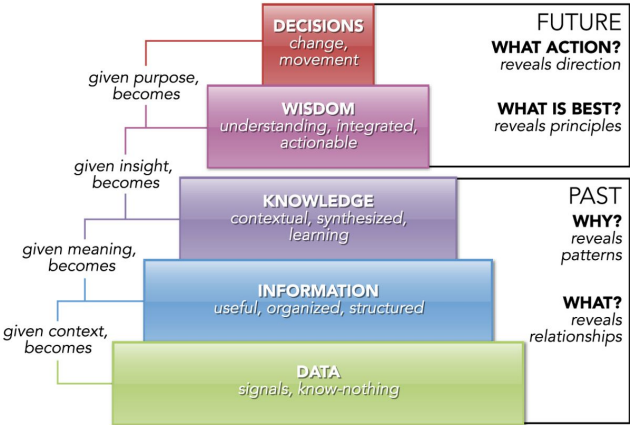




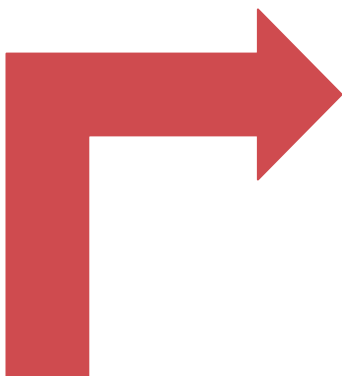
# DIKW Pyramid



# DIKW Example #1



W		
K		
I		
D	Raw	A list of numbers: 55, 62, 75, 89



# DECISIONS

The organization officially announced a work-from-home policy.

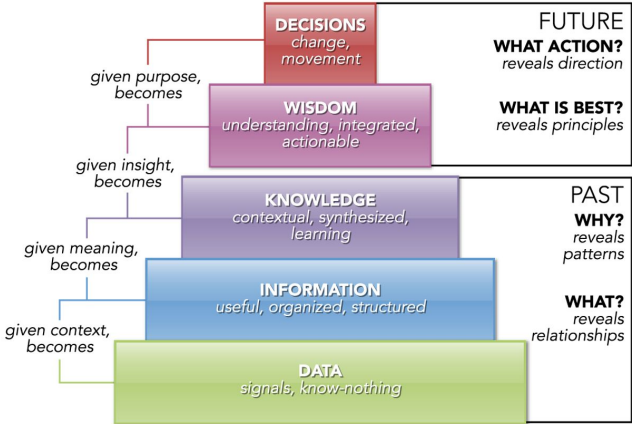


ประกาศมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี  
เรื่อง แนวปฏิบัติเพื่อรองรับมาตรการจัดการด้านการเรียนการสอนการวิจัยและการปฏิบัติงาน  
ในสถานการณ์ปริมาณฝุ่น PM 2.5

ตามที่กระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม (อว.) ได้ประกาศนโยบายขอความร่วมมือจากทุกหน่วยงานในกระทรวงที่อยู่ในพื้นที่ที่มีฝุ่น PM2.5 เกินค่ามาตรฐานให้พิจารณาจัดการเรียนการสอนแบบออนไลน์ และหน่วยงานอื่นด้านวิทยาศาสตร์พิจารณาให้ข้าราชการ พนักงานและเจ้าหน้าที่ มีการทำงานแบบ Work From Home (WFH) ตามความเหมาะสมตั้งแต่วันที่ 27-31 มกราคม 2568 หรือจนกว่าสถานการณ์จะดีขึ้นนั้น มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรีตั้งอยู่ในพื้นที่ที่ได้รับผลกระทบจากสถานการณ์ปริมาณฝุ่น PM2.5 ที่มีปริมาณสูงในระดับที่ส่งผลกระทบต่อสุขภาพของนักศึกษาและบุคลากร เพื่อให้การปฏิบัติงาน การจัดกิจกรรม การจัดการเรียนการสอน และการวิจัยเป็นไปด้วยความเรียบร้อย มหาวิทยาลัยจึงขอให้ทุกหน่วยงานดำเนินการเพื่อลดผลกระทบจากสถานการณ์ปริมาณฝุ่น PM 2.5 ตั้งแต่วันที่ 27-31 มกราคม 2568 ดังนี้

1. การปฏิบัติงาน การจัดกิจกรรมหรือการจัดการจัดการเรียนการสอนที่มีความเสี่ยงสูงต่อสุขภาพของนักศึกษาและบุคลากร ให้รองอธิการบดี คณบดี ผู้อำนวยการสำนัก/สถาบัน หรือหัวหน้าหน่วยงานพิจารณาอนุญาตให้บุคลากรในสังกัดปรับเปลี่ยนการปฏิบัติงานเป็นแบบ Work from Home (WFH) หรือปรับเปลี่ยนกิจกรรมหรือการเรียนการสอนให้เหมาะสมเป็นรายกรณี
2. หลีกเลี่ยงการจัดกิจกรรมหรือการเรียนการสอนกลางแจ้ง หรือให้ปฏิบัติงานกลางแจ้งเท่าที่จำเป็น
3. แนะนำให้นักศึกษาและบุคลากรสวมหน้ากากอนามัยที่ป้องกันฝุ่น PM2.5
4. กรณีนักศึกษาหรือบุคลากรในสังกัดมีอาการผิดปกติจากฝุ่น PM 2.5 ให้ไปพบแพทย์และตรวจสุขภาพโดยเร็ว
5. ผู้ที่มีหน้าที่เฉพาะ เช่น การรักษาความปลอดภัย การดำเนินการตามข้อมูลพื้นที่อันตรายที่มีการกำหนดระยะเวลา การดูแลโครงสร้างพื้นฐานของมหาวิทยาลัย และโครงสร้างพื้นฐานสำคัญทางสารสนเทศ เป็นต้น รวมถึงผู้ที่

# DIKW Example #2

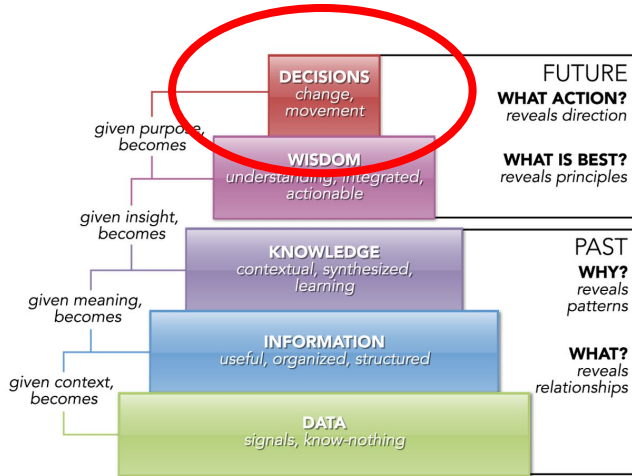
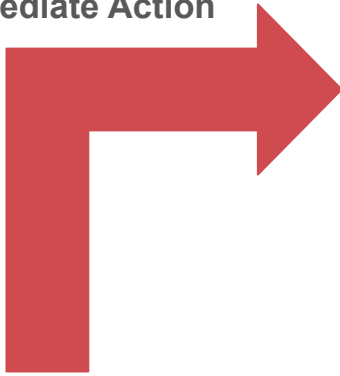


## DECISIONS

Issue a public advisory warning communities downstream not to use the river water for drinking or bathing until further notice

W		
K		
I		
D	Raw	A list of numbers: 6.2, 6.8, 7.1, 5.9

## Immediate Action



## DECISIONS

Issue a public advisory warning communities downstream not to use the river water for drinking or bathing until further notice

### Investigation:

Conduct a thorough investigation to identify the sources of pollution, such as nearby factories or agricultural runoff.

### Policy Implementation:

Enforce stricter environmental regulations for industrial discharge and establish regular water quality monitoring systems.

### Community Engagement:

Educate local residents about water safety practices and encourage the use of alternative water sources where available.

### Long-Term Solution:

Collaborate with industries to adopt sustainable practices and introduce water treatment systems to minimize future pollution.

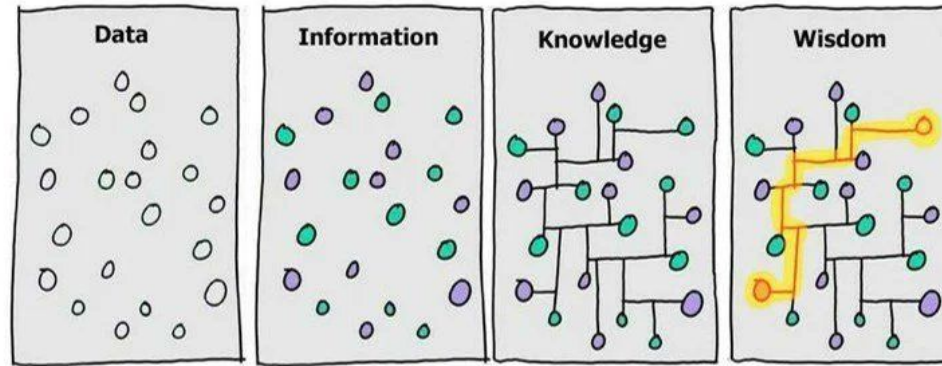
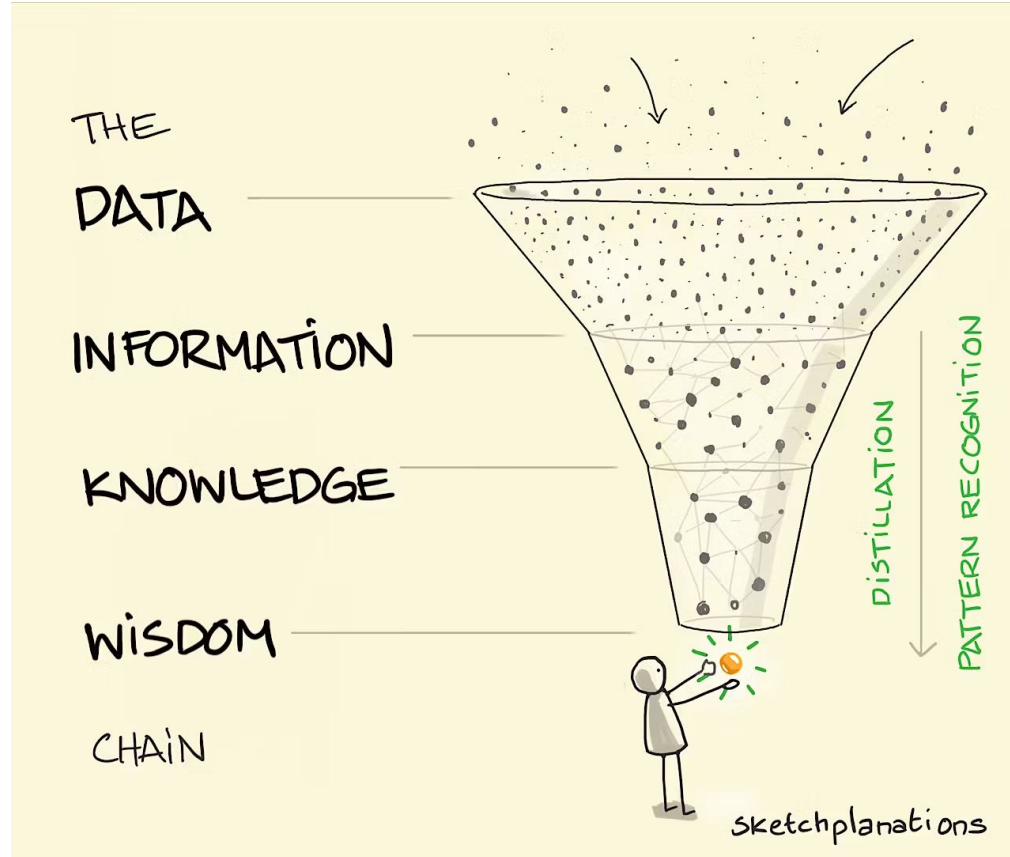
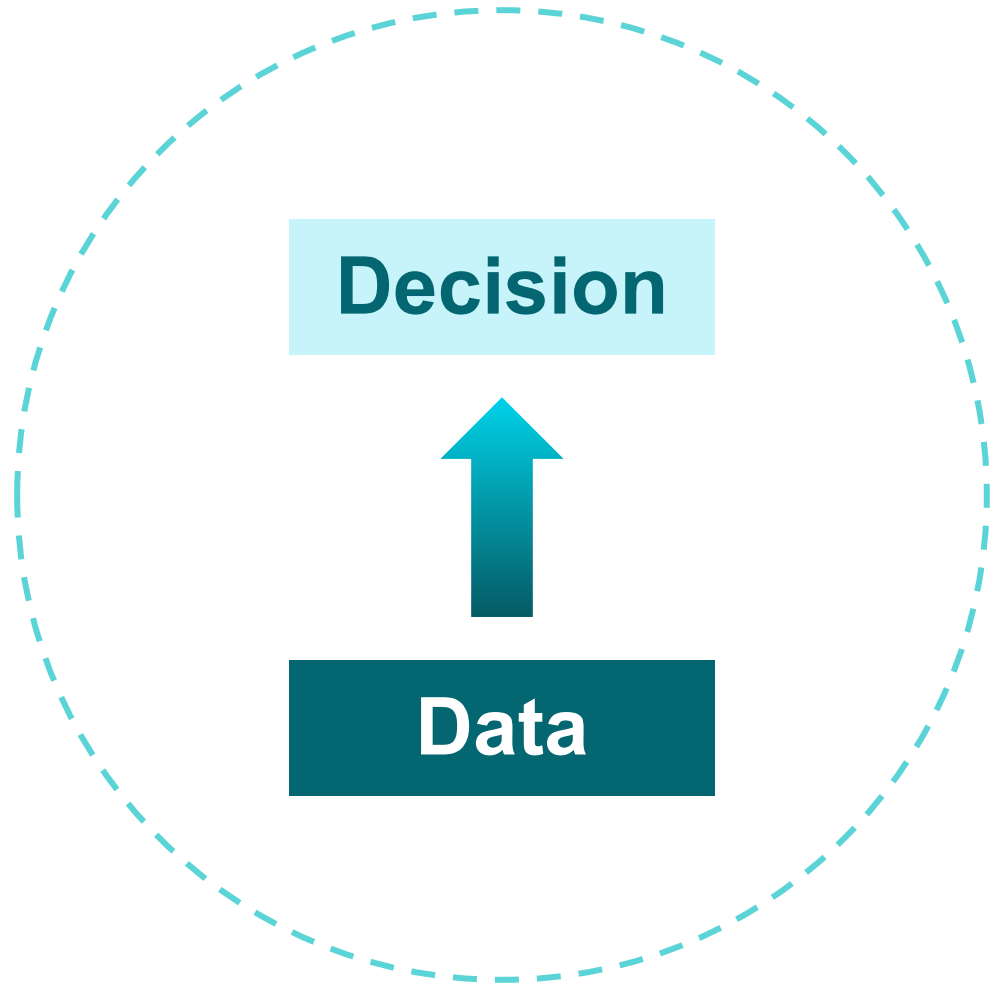


Image Adapted from: [https://x.com/prathviraj\\_com/status/1462482013332643843/photo/1](https://x.com/prathviraj_com/status/1462482013332643843/photo/1)



Ref: <https://sketchplanations.com/dikw>

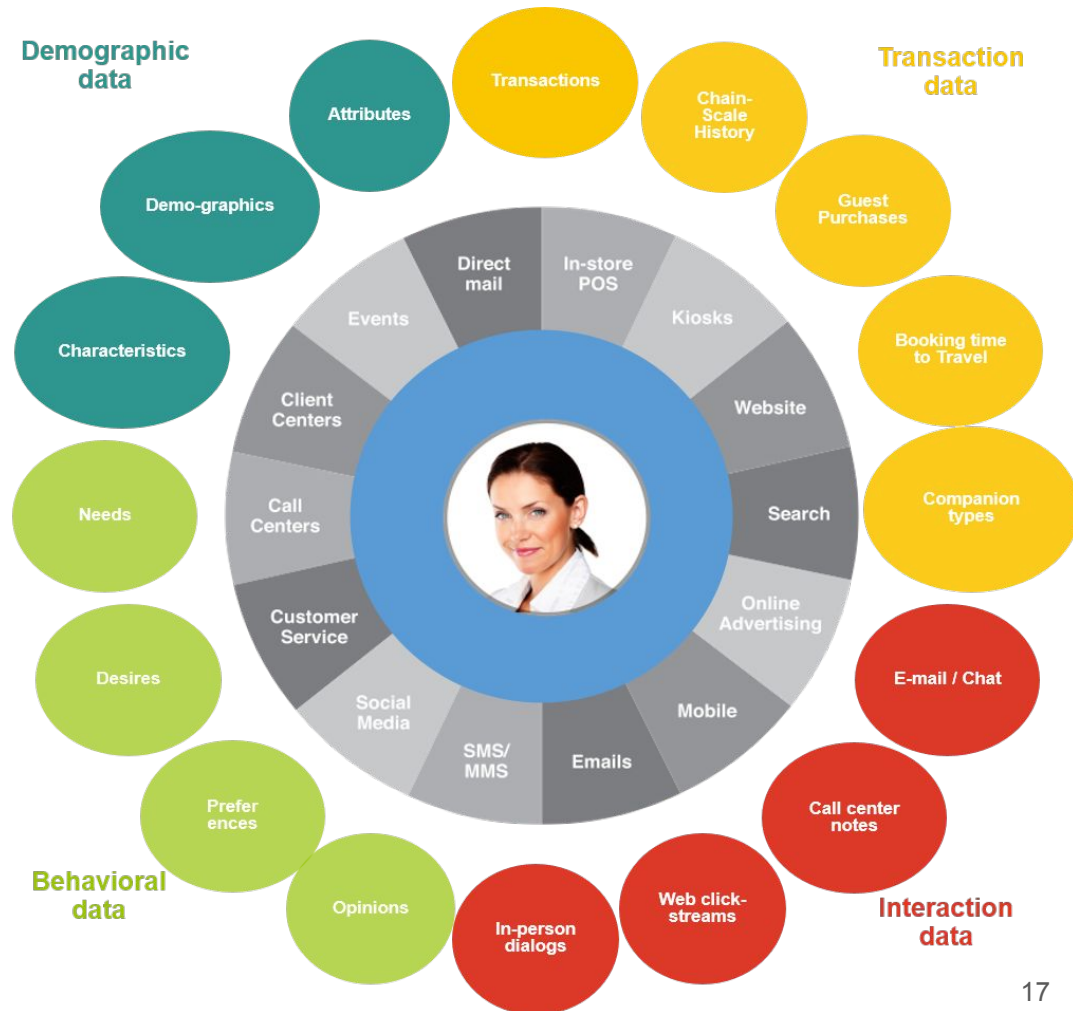
## Data-driven Decision Making





# Types of Data

- Demographic
- Transactional
- Interaction
- Behavioral



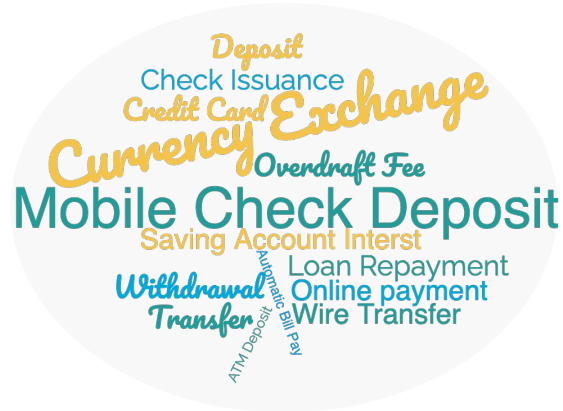
# Demographic Data

- Demography is the study of human populations
- Information about groups of people according to certain attributes:
  - age, gender, and place of residence
  - **socioeconomic attributes**: occupation, family status, income
- Collected to gain a deeper insight into the target group



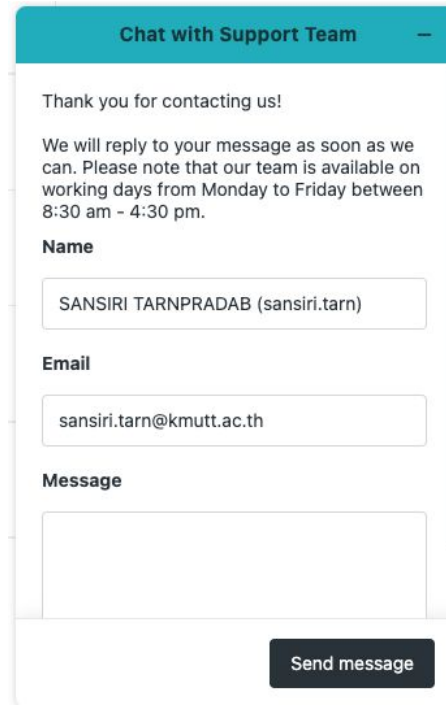
# Transactional Data

- Information that is captured from transactions
- All those records, all those details..
- Help in tracking and managing operations



# Interaction Data

- The record that describes an **interaction** between individuals/entities
- Examples:
  - Web click streams
  - In-person dialogs
  - Call center notes
  - Email / Chat log



The image shows a mobile app interface for a chat support team. At the top is a teal header with the text 'Chat with Support Team' and a minus icon. Below the header, the chat content area contains a message from the support team: 'Thank you for contacting us! We will reply to your message as soon as we can. Please note that our team is available on working days from Monday to Friday between 8:30 am - 4:30 pm.' Below this message are three input fields: 'Name' with the value 'SANSIRI TARNPRADAB (sansiri.tarn)', 'Email' with the value 'sansiri.tarn@kmutt.ac.th', and 'Message' which is currently empty. At the bottom right of the chat area is a dark blue button labeled 'Send message'.

# Behavioral Data

- Needs
- Desires
- Preferences
- Opinions



*Ref: <https://newlifectr.org/news-and-events/support-new-life-center-though-our-amazon-wish-list/>*

- Significance:
  - Strengthen organizations
  - Help to better understand your customers at the individual level
  - Transform marketing campaigns and personalized recommendations

# Types of Data

- Demographic
- Transactional
- Interaction
- Behavioral

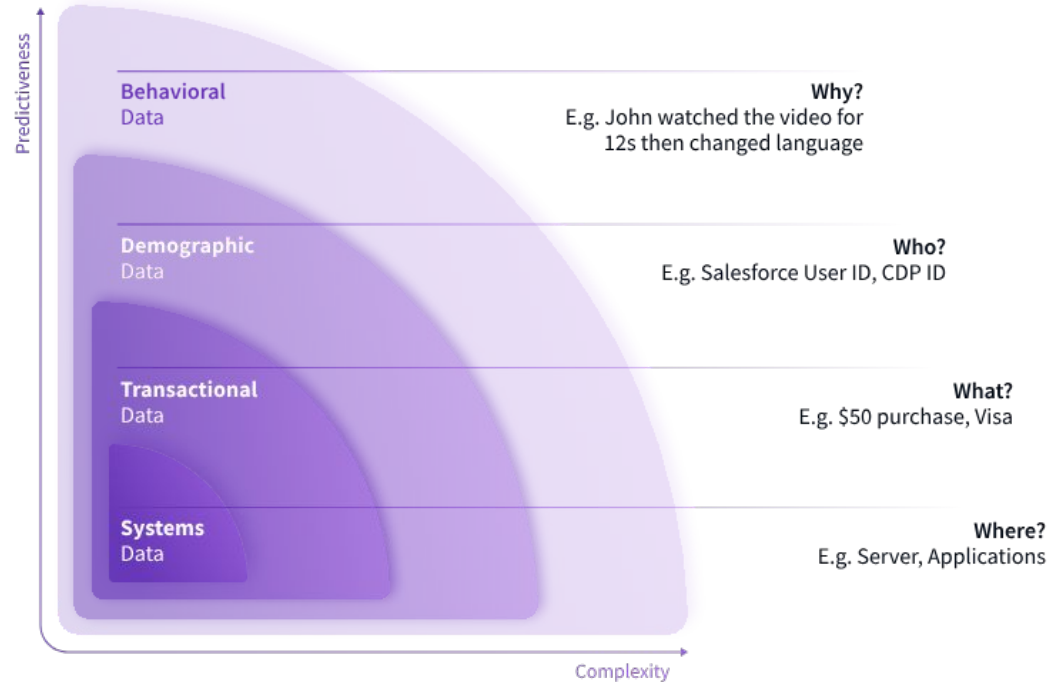
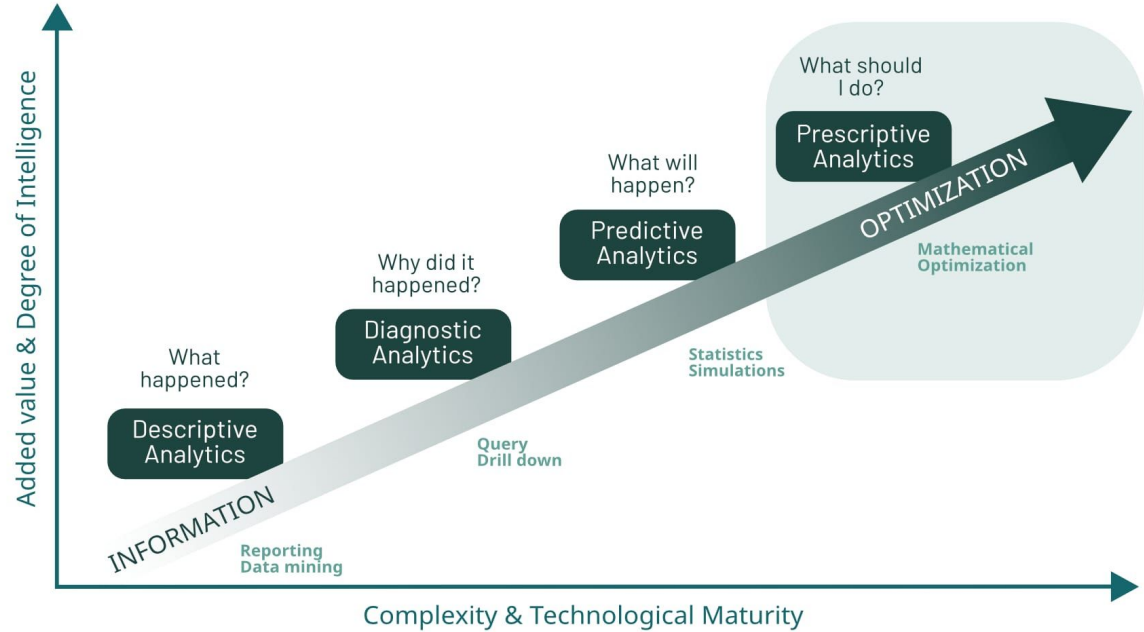


Figure 1. How behavioral data compares to other data types

Ref: <https://snowplow.io/what-is-behavioral-data/>

# Data Analytics

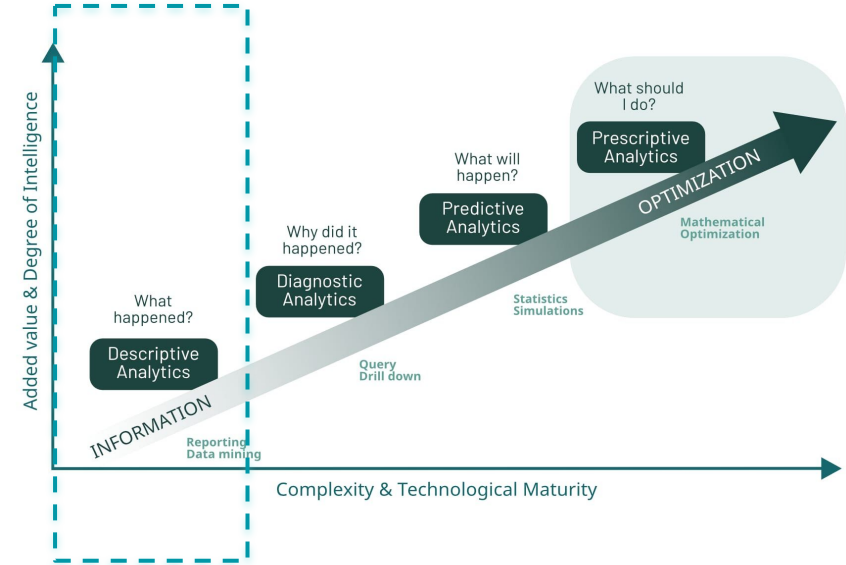
Make the most out of  
**DATA**



Ref: [https://atoptima.com/blog/prescriptive\\_analytics\\_improve\\_decision\\_making/](https://atoptima.com/blog/prescriptive_analytics_improve_decision_making/)

# Descriptive Analytics

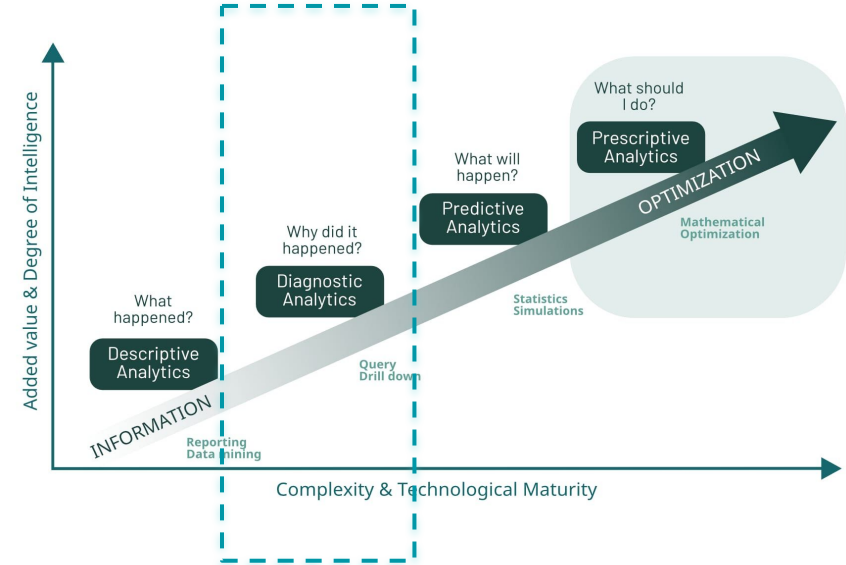
- What happened?
  - Analyzing historical data to understand what has happened.
- Reactive
  - Response to events/issues after they happen.
- Resources/Tools:
  - Report
  - Dashboard
  - Scorecards
- Outcome:
  - Well-defined problems and opportunities





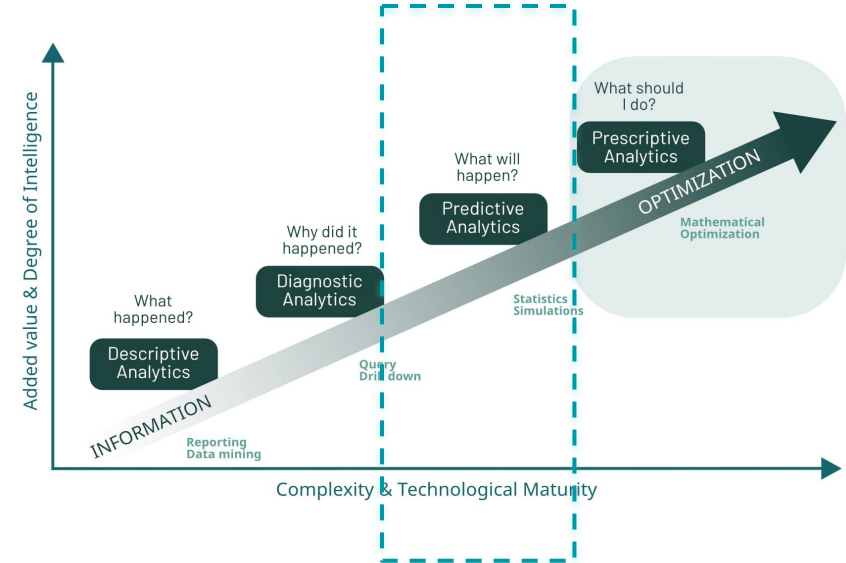
# Diagnostic Analytics

- Why did it happen?
  - Analyzing historical data and identifying patterns, relationships, and root causes.
  - Uncovering underlying reasons behind trends or anomalies
- Reactive
  - Response to events/issues after they happen.
- Resources/Tools:
  - Cause and effect analysis
  - Correlation
- Outcome:
  - Cause and effect of changes in activities



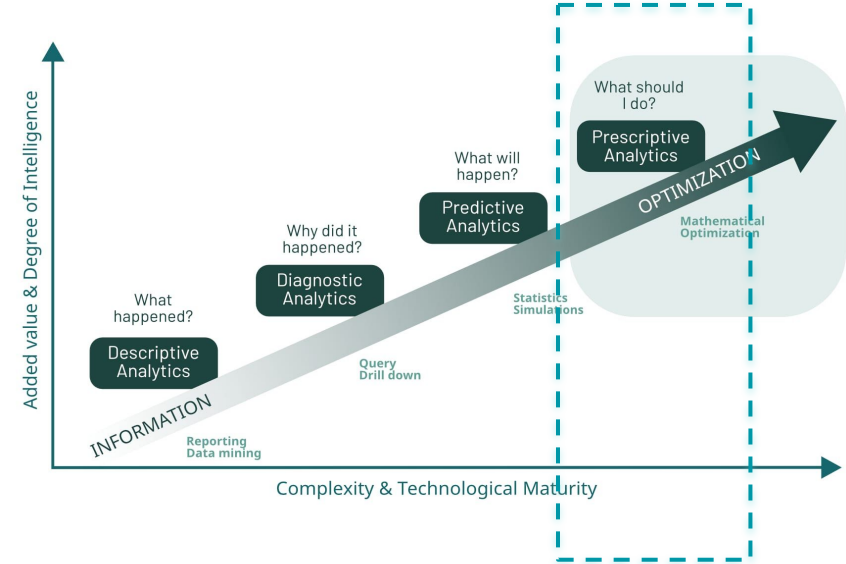
# Predictive Analytics

- What will happen?
  - Using historical data to identify patterns and trends to make forecasts and anticipate outcomes.
- Proactive
  - Taking measures before issues arise to prevent or minimize them.
- Resources/Tools:
  - Forecasting
  - Simulations
- Outcome:
  - Accurate projections of the future state



# Prescriptive Analytics

- What should I do?
  - Building on predictive analytics
  - Providing recommendations or solutions for decision-making
- Proactive
  - Taking measures before issues arise to prevent or minimize them.
- Resources/Tools:
  - Optimization
  - Decision Modeling
- Outcome:
  - Best possible decision





	Descriptive	Diagnostic	Predictive	Prescriptive
<b>Question</b>	What happened	Why did it happen?	What will happen?	What should I do?
<b>Response Strategies</b>	Reactive	Reactive	Proactive	Proactive
<b>Enablers</b>	Report, Dashboard, Scorecards	Cause and effect analysis, Correlation	Forecasting, Simulations	Optimization, Decision Modeling
<b>Outcomes</b>	Well-defined problems and opportunities	Cause and effect of changes in activities	Accurate projections of the future state	Best possible decision

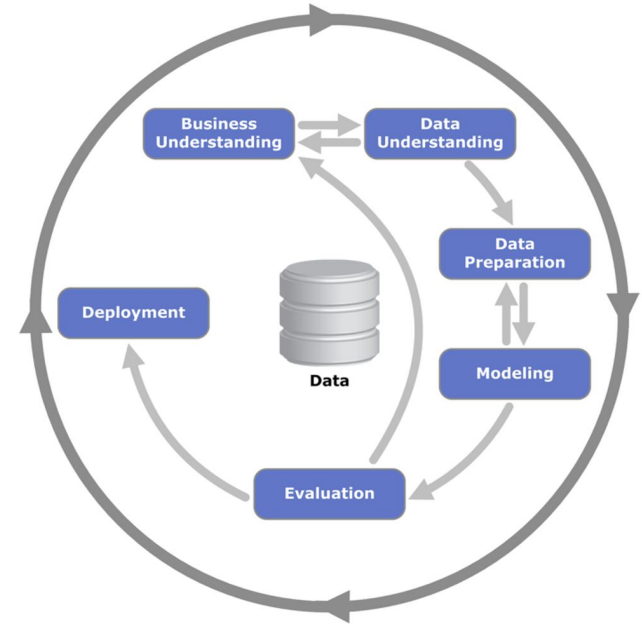


# Data Modeling

# What is Data Modeling?

A process of:

- Creating an abstract representation of a structure in a dataset or a system.
- Analyzing and defining all the different data types, as well as the relationships among them.
- Involves creating a visual representation to communicate connections between data points and structures.

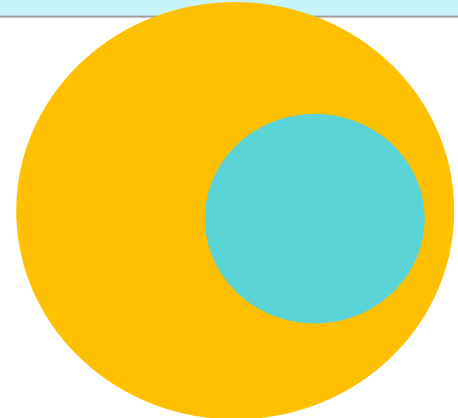


# Data-driven modeling vs Machine Learning

Data-driven modeling	Machine learning
<p>The process of using data to derive the functional form of a model or the parameters of an algorithm.</p> <ul style="list-style-type: none"><li>• Statistical methods</li><li>• Rule-based methods</li><li>• Data analysis methods</li></ul>	<ul style="list-style-type: none"><li>• The process of fitting parameters to data to minimize a cost function when the model is applied to the data.</li><li>• Parameters are updated through learning</li></ul>

## Examples:

- Descriptive statistics for sales (mean, median, mode, visualization)
- Customer segmentation (via demographic analysis)
- Credit scoring assessment (rule-based criteria)



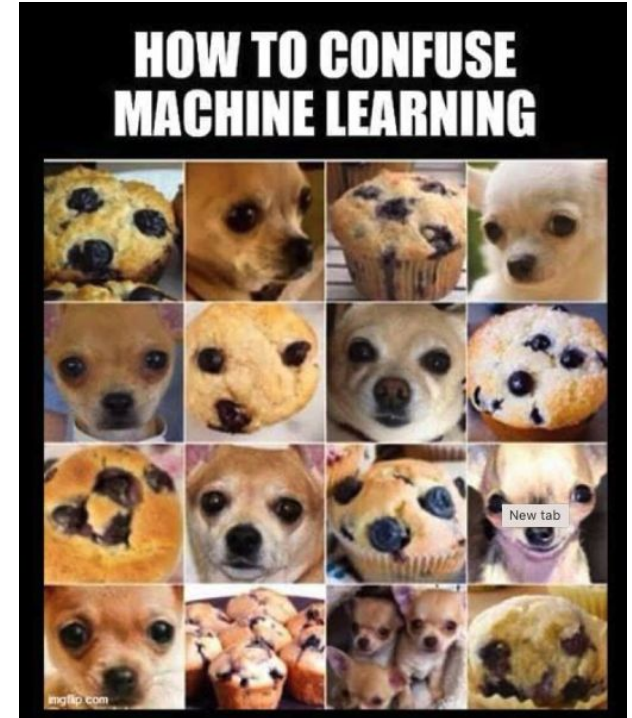


# A Glimpse of Machine Learning

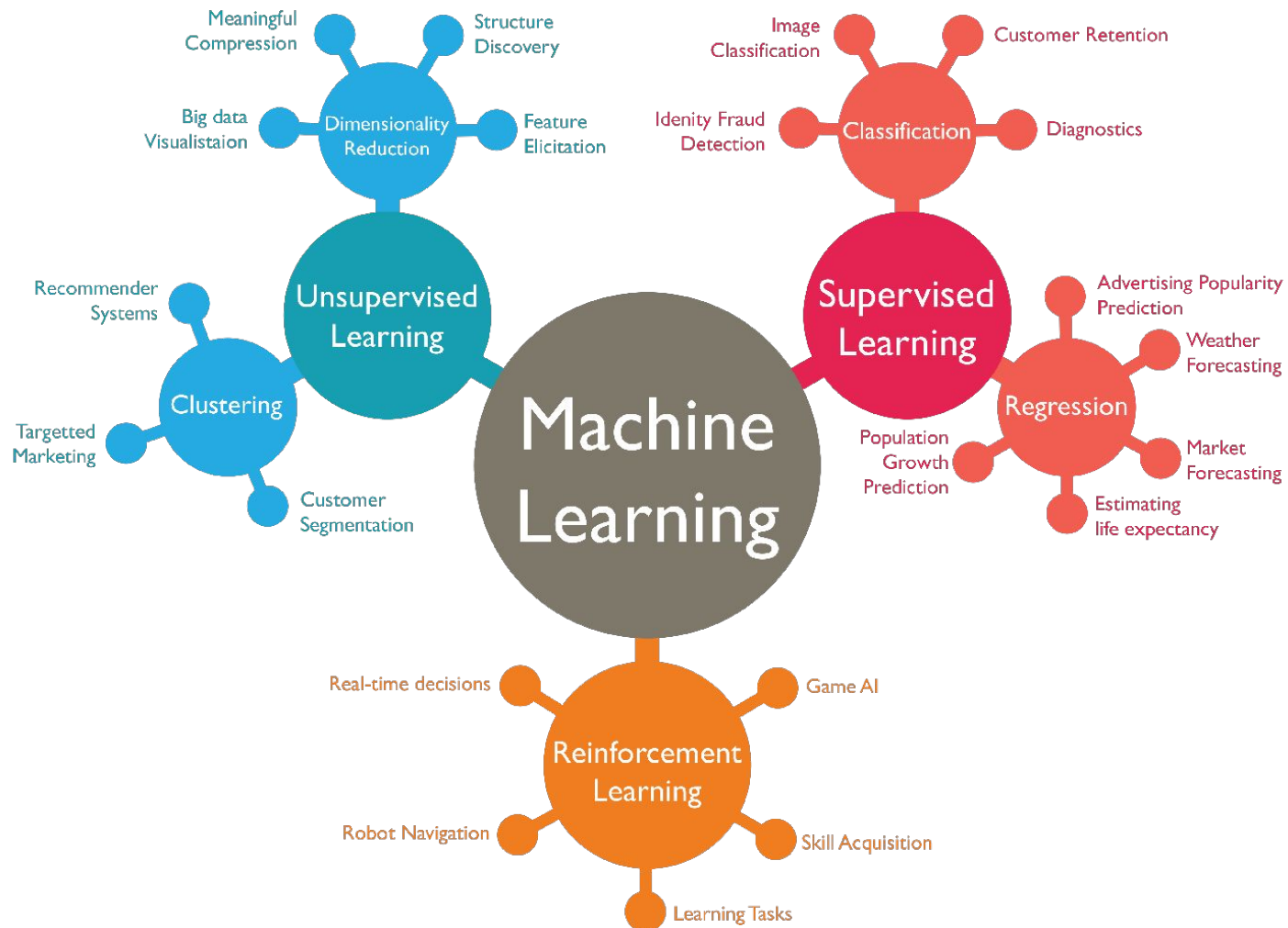
*“The science of getting computers to learn from data  
without having to be explicitly programmed by humans.”*

# ML Examples

- Speech & Image Recognition
- Traffic alerts (Google Map)
- Chatbot (Online Customer Support, ChatGPT)
- Google Translation
- Prediction
- Extraction
- Auto-Friend Tagging Suggestion
- Self-driving Cars
- Ads Recommendation
- Email Filtering



Ref: [https://www.youtube.com/watch?v=dWmP7\\_l1rFo](https://www.youtube.com/watch?v=dWmP7_l1rFo)



# Data Visualization

# What is Data Visualization?

## What:

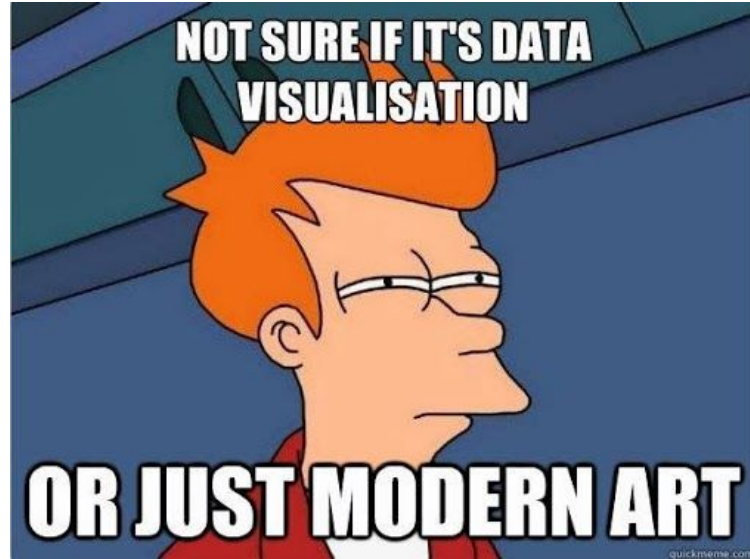
- Technique to present data in a pictorial/graphical format

## Significance & Benefits:

- Gain insights into an information space by mapping data onto graphical primitives
- Provide qualitative overview of large data sets
- Search for patterns, trends, structure, irregularities, relationships among data
- Help find interesting regions and suitable parameters for quantitative analysis
- Simplification

# Examples

- Scatterplot
- Heatmap
- Bar chart
- Line chart
- Pie chart
- Histogram
- Clustering
- Dashboard
- ... and more



Ref: [https://hackmd.io/@cs1951a/HJ\\_Q25wio](https://hackmd.io/@cs1951a/HJ_Q25wio)

# In Summary

## Introduction to Data Models

- Data
  - Significance
  - Decision Making
  - Types of Data
- Data Modeling
- Data Visualization



Ref:

<https://medium.sqldbm.com/database-modeling-relational-vs-transformational-be7b131ec73b>

# Q & A