

...

# Intro to Data Analytics

Karin Huangsuwan - 4 Plus Consulting ...



# Outline

---

- 1 บทนำ: ปลดล็อกพลังแห่ง Data**
  - 2 ทำความรู้จักกับ Data: ชนิด, การใช้งาน, และ ตัวอย่างจากชีวิตจริง**
  - 3 จาก Data สู่ Insights: ทำข้อมูลให้มีความหมาย**
  - 4 จัดระเบียบข้อมูลให้อยู่หมัด: รู้จัก Data Warehouse และ Data Modeling**
  - 5 เตรียมข้อมูลให้พร้อมใช้งาน: ETL, Data Cleansing และ Data Quality**
  - 6 Data Analytics: สร้างคุณค่าจากข้อมูลดิบ**
  - 7 พลังแห่ง Visualization: สื่อสารภาษา Data ในพริบตา**
-

# After this class, you will...

---

- **แยกประเภทข้อมูลพื้นฐานได้**  
 เช่น Structured, Unstructured, Semi-structured และรู้ว่าต้องจัดการยังไงกับแต่ละแบบ
  - เข้าใจเส้นทางของข้อมูลตั้งแต่ต้นทางถึงปลายทาง  
 ตั้งแต่การเก็บข้อมูล → การทำความสะอาด → การแปลง → การวิเคราะห์ → การนำเสนอ
  - เลือก **Visualization** ที่เหมาะสมกับประเภทข้อมูลและจุดประสงค์ของการสื่อสาร  
 เช่น เมื่อไหร่ควรใช้ Bar chart, Line chart, Pie chart หรือ Scatter plot
  - **รู้จักระบวนการ ETL (Extract, Transform, Load)**  
 เข้าใจว่าทำไมต้องเตรียมข้อมูลก่อนวิเคราะห์ และต้องทำอะไรบ้างในแต่ละขั้นตอน
  - **ตรวจสอบและปรับปรุงคุณภาพข้อมูลได้เบื้องต้น**  
 เช่น การจัดการ Missing values, Outliers, Duplicate records
    - อ่าน **Dashboard** และ **Report** ได้อย่างเข้าใจมากขึ้น
  - **เข้าใจการทำงานของ Data Warehouse และแนวคิด Star Schema**  
 รู้ว่าควรโฟกัสตรงไหน และสามารถอธิบายข้อมูลบนหน้าจอให้คนอื่นฟังได้
    - เข้าใจการทำงานของ **Data Warehouse** และแนวคิด **Star Schema**  
 พอยเข้าใจ Concept ของ Fact และ Dimension Table ในระดับเบื้องต้น
- **มีทักษะในการแปลงข้อมูลดิบให้เป็นข้อมูลที่พร้อมใช้งานวิเคราะห์**  
 ทั้งในเชิงโครงสร้างและคุณภาพ
  - **เข้าใจความหมายของคำว่า “Insight”**  
 รู้ว่าข้อมูลแบบไหนคือแค่ตัวเลขธรรมด้า และแบบไหนคือ Insight ที่นำไปใช้ตัดสินใจได้
    - เริ่มต้นวิเคราะห์ข้อมูลเชิงสำรวจ (EDA) ได้เองในระดับเบื้องต้น  
 รู้จักวิธีหาความสัมพันธ์ง่าย ๆ และแนวโน้มเบื้องต้นจากข้อมูล
    - **เข้าใจความสำคัญของ Data Governance และ Privacy**  
 รู้ว่าอะไรควรเปิดเผย อะไรควรรักษาไว้ทำงานกับข้อมูล
    - **เข้าใจ 4 ประเภทของ Data Analytics**  
 Descriptive / Diagnostic / Predictive / Prescriptive และรู้ว่าแต่ละแบบใช้เมื่อไหร่
      - **อธิบายได้ว่าทำไมองค์กรต้องลงทุนกับ Data Infrastructure**  
 เช่น การทำ Data Warehouse, Data Quality Improvement
        - มีพื้นฐานที่ดีสำหรับต่อยอดไปเรียน **Data Science** หรือ **BI Tools** ต่าง ๆ ในอนาคต
      - **เปลี่ยนมุมมองต่อข้อมูล จาก “ตัวเลขบน Excel” → “เครื่องมือช่วยตัดสินใจ”**  
 เริ่มมองเห็นว่าข้อมูลที่ดูธรรมด้า สามารถนำไปใช้สร้างคุณค่าในงานและชีวิตจริงได้ยังไง

เข้าใจและจัดการข้อมูลได้อย่างเป็นระบบ ตั้งแต่การแยกประเภท เตรียมความพร้อม วิเคราะห์ และสื่อสารผลลัพธ์ เพื่อสนับสนุนการตัดสินใจอย่างมีเหตุผล และสร้างคุณค่าทางธุรกิจ

---

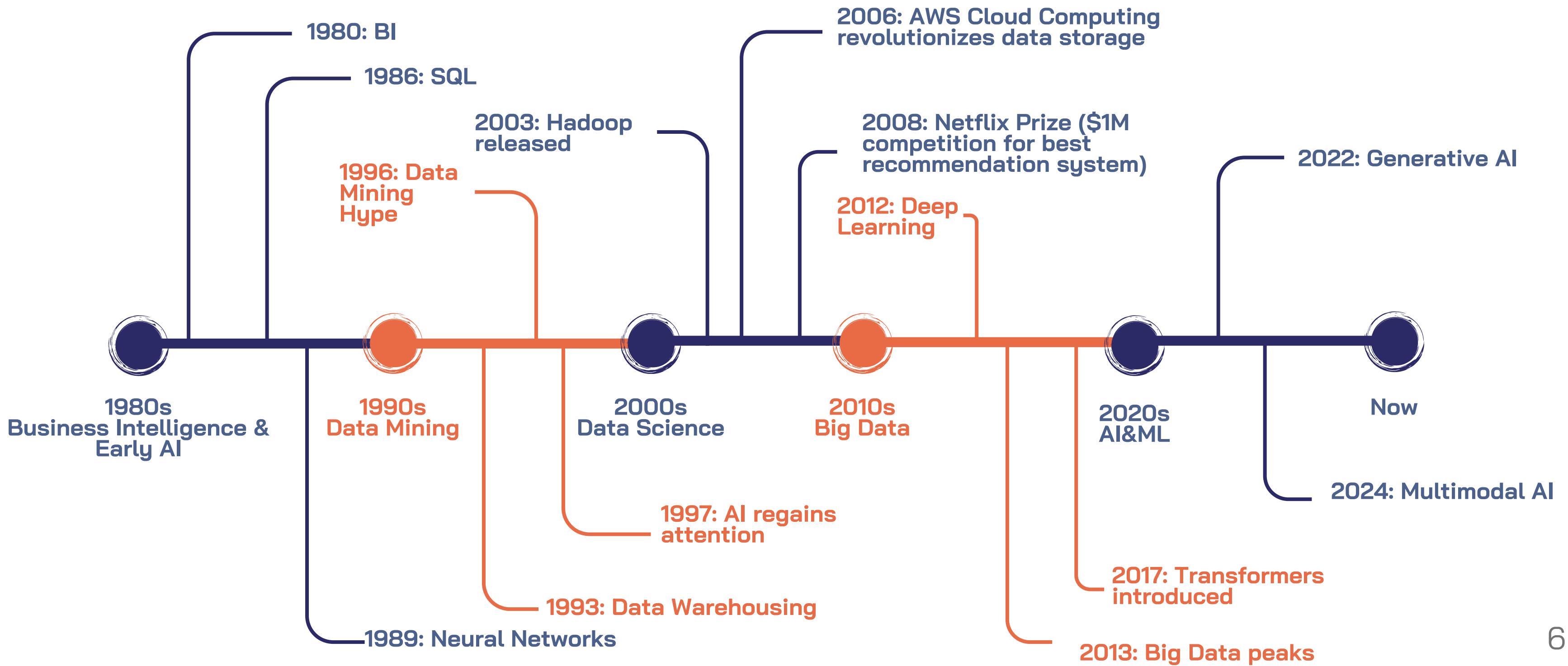
# 1 บทนำ: ปลดล็อกพลังแห่ง Data

---

# Data Analytics: คำพูดเหตุ ๆ หรืออาวุธลับเปลี่ยนโลก?



# Data Analytics: คำพูดเหตุ ฯ หรืออาวุธลับเปลี่ยนโลก?



# เบื้องหลังความซอตของ Data Analytics

## 1) วิวัฒนาการของหน่วยความจำ



5MB in 1956



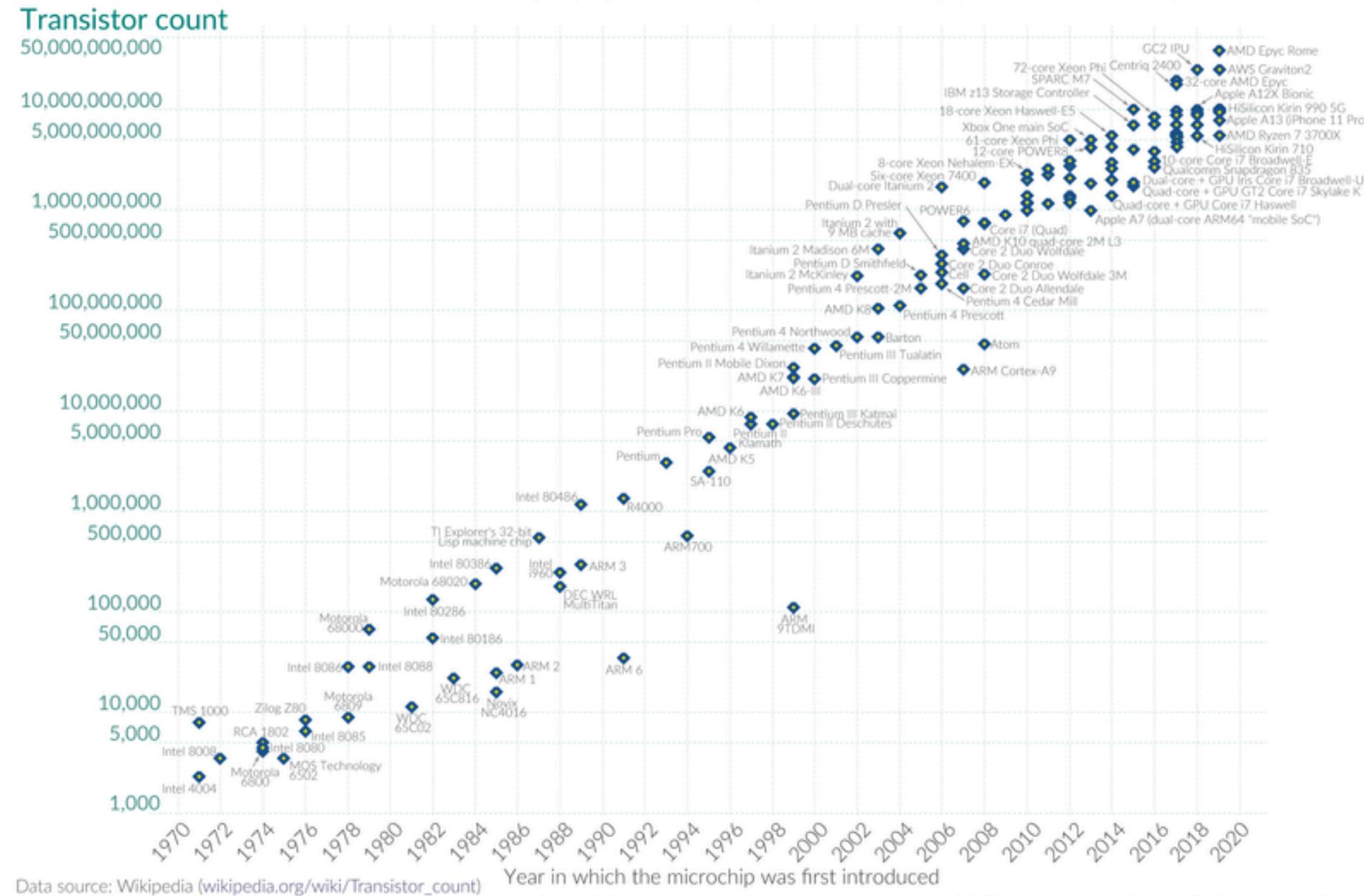
1TB in 2024  
(1,000,000 MB)

# เบื้องหลังความซottoของ Data Analytics

2) การเติบโตแบบก้าวกระโดดของพลัง computation

Moore's Law: The number of transistors on microchips doubles every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.



# เบื้องหลังความซottoของ Data Analytics

## 3) การมาเยือนของอินเตอร์เน็ตความเร็วสูงและ IoT



## 2 ทำความรู้จักกับ Data: ชนิด, การใช้งาน, และ ตัวอย่างจากชีวิตจริง

---

ข้อมูลที่มีโครงสร้างต่างกัน ก็จะมีวิธีจัดการไม่เหมือนกันด้วย

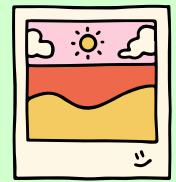
## โครงสร้าง (structure)



**Structured** จัดเรียงอยู่ในรูปแบบที่กำหนดไว้ล่วงหน้า เช่น ฐานข้อมูลแบบ relational



**Semi-structured** มีการจัดเรียงบางส่วน เช่น JSON, XML, NoSQL



**Unstructured** ไม่มีรูปแบบที่กำหนดไว้ล่วงหน้า เช่น วิดีโอ รูปภาพ เอกสารข้อความอิสระ

ถ้าแยกประเภทข้อมูลเหล่านี้ออก ก็จะทำงานง่ายขึ้นมาก

## ประเภทของข้อมูล

ข้อมูลบางประเภทต้องใช้ความระมัดระวังเป็นพิเศษในการทำงาน

## ความอ่อนไหว (sensitivity)



**Public** เปิดให้เข้าถึงทั่วไป เช่น ข้อมูลที่รัฐบาลเปิดเผยให้กับประชาชน



**Internal** จำกัดเฉพาะพนักงานในองค์กร เช่น รายงานภายใน



**Confidential** ต้องได้รับสิทธิ์หรือการอนุญาต เช่น ข้อมูลลูกค้า

ถ้าเข้าใจจะช่วยให้วิเคราะห์และแสดงผลได้ตรงจุด

## การใช้งาน (function)



**Transactional** ข้อมูลธุรกรรม ข้อมูลที่บันทึกเหตุการณ์เป็นรายครั้ง เช่น การซื้อขาย การโอนเงิน มีเวลาที่ชัดเจนและรายละเอียดของกิจกรรม



**Master** ข้อมูลที่ใช้อ้างอิงร่วมกับข้อมูลอื่น เช่น รายการสินค้า รายชื่อลูกค้า รหัสสาขา



**Metadata** ข้อมูลที่ใช้ธิบายลักษณะของข้อมูลอื่น ๆ เช่น ชื่อไฟล์ เวลาสร้าง/แก้ไข เจ้าของข้อมูล ประเภทของข้อมูล

## เวลา (time)



**Time-series** ข้อมูลที่จัดเก็บต่อเนื่องตามเวลา สามารถใช้ดูแนวโน้มที่เปลี่ยนแปลงได้ เช่น ยอดขายของสินค้ารายวัน



**Snapshot** ข้อมูล ณ เวลาหนึ่ง เช่น สถานะคลังสินค้าตอนลิ้นเดือนมิถุนายน ใช้เพื่อดูสถานการณ์ ณ เวลาหนึ่ง เช่น จำนวนสินค้าคงเหลือในวันที่ 30 มิ.ย.



**Static** ข้อมูลที่คงที่หรือเปลี่ยนแปลงน้อย ไม่ขึ้นอยู่กับเวลา เช่น รหัสประเทศ หรือหมวดหมู่สินค้า



## Structured + Transactional + Time series data

| Transaction ID | Customer ID | Product Name      | Quantity | Price  | Purchase Date | Payment Method   |
|----------------|-------------|-------------------|----------|--------|---------------|------------------|
| TX1001         | C001        | Wireless Mouse    | 1        | 450.0  | 2025-06-15    | Credit Card      |
| TX1002         | C002        | Bluetooth Speaker | 2        | 1200.0 | 2025-06-16    | Bank Transfer    |
| TX1003         | C001        | USB-C Charger     | 1        | 350.0  | 2025-06-17    | Credit Card      |
| TX1004         | C003        | Laptop Stand      | 1        | 800.0  | 2025-06-17    | Cash on Delivery |
| TX1005         | C004        | Webcam HD         | 1        | 900.0  | 2025-06-18    | E-Wallet         |



**Column:** แถวแนวตั้ง (Transaction ID, Customer ID, Product Name ...)

- อาจใช้คำเรียกอื่นเช่น **Feature, Attribute, Field, Variable, Dimension**

**Row:** แถวแนวนอน (TX1001, TX1002, TX1003, ...)

- ในบางทำร้าอาจใช้คำว่า **Record, Example, Instance, Observation** ขึ้นกับบริบท



## Structured + Master + Static data

| Product ID | Product Name      | Category    | Brand     | Unit Price | In Stock |
|------------|-------------------|-------------|-----------|------------|----------|
| 100101     | Wireless Mouse    | Accessories | LokiTech  | 450.0      | 150      |
| 100102     | Bluetooth Speaker | Audio       | UBL       | 1200.0     | 80       |
| 100103     | USB-C Charger     | Chargers    | SomeSoong | 350.0      | 200      |
| 100104     | Laptop Stand      | Office      | ReMin     | 800.0      | 75       |
| 100105     | Webcam HD         | Camera      | LokiTech  | 900.0      | 60       |



## Semi-Structured + Metadata + Static data

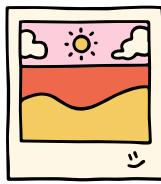
```
{
    Keys ↓           Values ↓
  "filename": "IMG_20250618.jpg",
  "resolution": "4032x3024",
  "filesize": "3.2MB",
  "timestamp": "2025-06-18T14:12:09Z",
  "location": {
      "latitude": 13.7563,
      "longitude": 100.5018
  },
  "camera": {
      "brand": "Apple",
      "model": "iPhone 14 Pro"
  }
}
```

นี่คือ metadata ของรูปภาพ

- ทำหน้าที่อธิบายรูปภาพนั้น ๆ
- อยู่ในรูปแบบ semi-structured (JSON)
- มี key-value
- บาง field ซ้อนกันหลายชั้น (ไม่จัดตารางแบบ SQL)

ข้อดี: ยืดหยุ่นกว่าในการจัดเก็บ

ข้อเสีย: ต้องไปเปลี่ยนเป็น structured หากต้องการนำมารวิเคราะห์



## Unstructured Data



รูปภาพและวิดีโอ

**Unstructured data** เพิ่มระดับความยากให้กับการวิเคราะห์ เพราะ **ข้อมูลไม่ได้อยู่ในรูปแบบที่คอมพิวเตอร์เข้าใจ** ต้องผ่าน process บางอย่าง (ปัจจุบันนิยมใช้ AI อย่าง Computer Vision และ GPT ในการจัดการข้อมูลประเภทนี้)

### *Lorem Ipsum*

"Neque porro quisquam est qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit..."  
"There is no one who loves pain itself, who seeks after it and wants to have it, simply because it is pain..."

#### What is Lorem Ipsum?

Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

#### Where does it come from?

Popular belief, Lorem Ipsum is not simply random text. It is a piece of classical Latin literature from 45 BC, over 2000 years old. Richard McClintock, a Latin teacher at Hampden-Sydney College in Virginia, looked up one of obscure Latin words, consectetur, from a Lorem Ipsum passage, and going through the cites of the word in classical literature, discovered the undoubtable source. Lorem Ipsum comes from sections 1.10.32 and 1.10.33 of "de Finibus Bonorum et Malorum" (The Extremes of Good and Evil) by Cicero, written in 45 BC.

#### Why do we use it?

It is a long established fact that a reader will be distracted by the readable content of a page when looking at its layout. That is why using Lorem Ipsum is that it has a more-or-less normal distribution of letters, as opposed to using 'Content here, content here', making it look like readable English. Many desktop publishing packages and web page editors now use Lorem Ipsum as their default model text, and a search for 'lorem ipsum' will reveal many web sites still in their infancy. Various versions have survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged.

#### Where can I get some?

There are many variations of passages of Lorem Ipsum available, but the majority have suffered alteration in some form, by injected humour, or randomised words which don't look even slightly believable. If you are going to use a passage of Lorem Ipsum, you need to be sure there isn't anything embarrassing hidden in the middle of text. All the Lorem Ipsum generators on the Internet tend to repeat predefined chunks as necessary, making them very much like pre-defined generator on the Internet. It uses a dictionary of over 200 Latin words, combined with a handful of model



เสียง

บทความ/เอกสาร

Public  
Data

Private  
Data

Confid  
ential  
data



## Activity: แบ่งประเภทข้อมูล

# ตัวอย่างจากชีวิตจริง: ธุรกิจต่าง ๆ นำข้อมูลไปใช้ยังไงบ้าง?



การตัดสินใจทาง  
ธุรกิจ



การปรับเนื้อหาให้  
ตรงกับผู้ใช้



สุขภาพและ  
การแพทย์



Data-Driven  
Use Cases



E-Commerce &  
Shopping

การเงินและการ  
ตรวจสอบการฉ้อโกง



แผนที่และ  
การนำทาง

# ตัวอย่างจากชีวิตจริง: ธุรกิจต่าง ๆ นำข้อมูลไปใช้ยังไงบ้าง?

## การตัดสินใจทางธุรกิจ



ข้อมูล search → ช่วยตัดสินใจในการโฆษณา  
ข้อมูลการชำระเงิน → จัดการสต็อกและปรับราคา

**ข้อมูล:** ข้อมูลการขาย ข้อมูลเกี่ยวกับลูกค้า แนวโน้มตลาดในปัจจุบัน ข้อมูลการดำเนินการต่างๆ

**การนำมาใช้:** ธุรกิจนำข้อมูลมาวิเคราะห์เพื่อวางแผนกลยุทธ์ให้เหมาะสม พยายารณ์ความต้องการ และปรับปรุงประสิทธิภาพการดำเนินงาน

**ประโยชน์:** ช่วยลดความเสี่ยง เพิ่มความพึงพอใจของลูกค้า และสร้างกำไรมากขึ้น

# ตัวอย่างจากชีวิตจริง: ธุรกิจต่าง ๆ นำข้อมูลไปใช้ยังไงบ้าง?

## การปรับเนื้อหาให้ ตรงกับผู้ใช้



**Spotify:** ประวัติการฟังเพลง/ค้นหา → เพลงที่แนะนำ

**N** พฤติกรรมการดูของแต่ละคน + เทคนิคปัจจุบัน → แนะนำหนัง/ซีรีส์

**ข้อมูล:** พฤติกรรมของผู้ใช้, ประวัติการรับชม, คำที่ค้นหา, ความสนใจของแต่ละคน

**การใช้งาน:** นำข้อมูล บวกกับความช่วยเหลือจาก AI เพื่อแนะนำเนื้อหาที่น่าสนใจ เหมาะสมกับผู้ใช้แต่ละราย

**ประโยชน์:** เพิ่ม engagement (การมีส่วนร่วม) ของผู้ใช้ ช่วยรักษาผู้ใช้ให้อยู่กับแพลตฟอร์มได้นานขึ้น

# ตัวอย่างจากชีวิตจริง: ธุรกิจต่าง ๆ นำข้อมูลไปใช้ยังไงบ้าง?

## E-Commerce & Shopping



วิเคราะห์ข้อมูลด้วย AI → แนะนำผลิตภัณฑ์, เพิ่มประสิทธิภาพให้ flash sales, ตรวจจับรีวิวปลอม

**amazon** ประวัติการซื้อ/ค้นหา → แนะนำผลิตภัณฑ์

**ข้อมูล:** ประวัติการสั่งซื้อ, พฤติกรรมการเรียกดูสินค้า, ผลกระทบสินค้า, รีวิวจากลูกค้า

**การใช้งาน:** แนะนำสินค้าที่เหมาะสม ปรับราคาสินค้าให้เหมาะสมกับผู้ใช้ และพยากรณ์ความต้องการล็อกอิน

**ประโยชน์:** เพิ่มยอดขาย ยกระดับประสบการณ์ผู้ใช้ และช่วยให้บริการล็อกอินได้อย่างมีประสิทธิภาพ

# ตัวอย่างจากชีวิตจริง: ธุรกิจต่าง ๆ นำข้อมูลไปใช้ยังไงบ้าง?

## แผนที่และการนำทาง



**Grab**

ข้อมูล GPS แบบเรียลไทม์ + การจราจร → แนะนำเส้นทางที่ดีที่สุด

ตำแหน่งของคนขับ + demand + การจราจร → จับคู่คนขับกับลูกค้า

**ข้อมูล:** ตำแหน่ง GPS รูปแบบการจราจร, เส้นทางที่ผ่านมา, รายงานจากผู้ใช้

**การใช้งาน:** เปิดนำทางวิเคราะห์ข้อมูลการจราจรทั้งแบบเรียลไทม์และย้อนหลัง เพื่อแนะนำเส้นทางที่เร็วที่สุด

**ประโยชน์:** ช่วยลดเวลาเดินทาง หลีกเลี่ยงรถติด และเพิ่มความปลอดภัยให้กับผู้ขับขี่

# ตัวอย่างจากชีวิตจริง: ธุรกิจต่าง ๆ นำข้อมูลไปใช้ยังไงบ้าง?

## การเงินและการตรวจสอบการฉ้อโกง



ลักษณะการใช้เงิน + ตำแหน่งที่จ่ายเงิน → ปกติ/ผิดปกติ



ข้อมูลธุกรรม → ความผิดปกติ



**ข้อมูล:** ข้อมูลธุกรรม, พฤติกรรมการใช้จ่าย, คะแนนเครดิต, ตำแหน่งที่ใช้จ่าย

**การใช้งาน:** ธนาคารและบริษัทการเงินใช้ Machine Learning เพื่อตรวจจับรูปแบบการใช้จ่ายที่ผิดปกติ และป้องกันการฉ้อโกง

**ประโยชน์:** ปกป้องผู้บริโภคจากการฉ้อโกงทางการเงิน เพิ่มความปลอดภัย และทำให้ธุรกิจมีความน่าเชื่อถือ

# ตัวอย่างจากชีวิตจริง: ธุรกิจต่าง ๆ นำข้อมูลไปใช้ยังไงบ้าง?

## สุขภาพและการแพทย์



อัตราการเต้นของหัวใจ + กิจกรรมที่ทำในแต่ละวัน → ช่วยเฝ้าระวังสุขภาพ



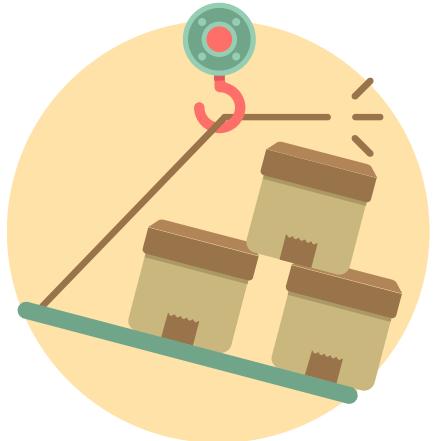
ระบบจดจำใบหน้าและตรวจสอบอาการเบื้องต้น → บริหารจัดการการฉุกเฉินได้

**ข้อมูล:** เวชระเบียนของผู้ป่วย, ข้อมูลสุขภาพจากอุปกรณ์สวมใส่, ผลตรวจทางห้องแล็บ, ข้อมูลพันธุกรรม

**การใช้งาน:** โรงพยาบาลและระบบ AI ใช้ข้อมูลสุขภาพในการวินิจฉัย วางแผนการรักษา และป้องกันโรค

**ประโยชน์:** ช่วยให้การดูแลผู้ป่วยมีประสิทธิภาพ ตรวจพบโรคได้เร็วขึ้น และสนับสนุนงานวิจัยทางการแพทย์

# ข้อควรระวังที่องค์กรต้องใส่ใจ หากเก็บข้อมูลไว้จำนวนมาก



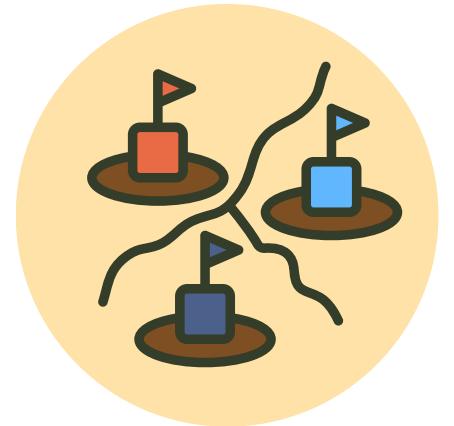
## ข้อมูลล้น (Data Overload)

หลายองค์กรเก็บข้อมูลจำนวนมากมหาศาล แต่ไม่สามารถนำมาใช้งานได้เท่าที่ควร บางครั้งอาจมีข้อมูลมากกว่าที่รับไหว



## คุณภาพข้อมูลไม่ดี (Data Quality Issues)

ข้อมูลผิดพลาด ขาดหาย หรือซ้ำซ้อน จากการกรอกข้อมูลที่ไม่เป็นระบบ



## ข้อมูลกระจัดกระจาย (Siloed Data)

ข้อมูลถูกแยกอยู่ตามฝ่ายต่าง ๆ เช่น HR, การตลาด, ฝ่ายขาย ทำให้มองไม่เห็นภาพรวม



## ความเสี่ยงด้านความเป็นส่วนตัว (Privacy & Security Risks)

ข้อมูลอาจรั่วไหล ถูกเข้าถึงโดยไม่ได้รับอนุญาต หรือใช้งานผิดวัตถุประสงค์ สร้างความเสียหายทั้งทางชื่อเสียงและกฎหมาย

# ข้อควรระวังที่องค์กรต้องใส่ใจ หากเก็บข้อมูลไว้จำนวนมาก



## ข้อมูลล้น (Data Overload)

ใช้เครื่องมือการองข้อมูล แดชบอร์ด และ AI วิเคราะห์แนวโน้มสำคัญ



## คุณภาพข้อมูลไม่ดี (Data Quality Issues)

ทำความสะอาดข้อมูล ตั้งกฎตรวจสอบ และใช้ระบบอัตโนมัติช่วยจัดการ



## ข้อมูลกระจัดกระจาย (Siloed Data)

รวมข้อมูลไว้ที่เดียวผ่านคลังข้อมูล (Data Warehouse) หรือระบบคลาวด์



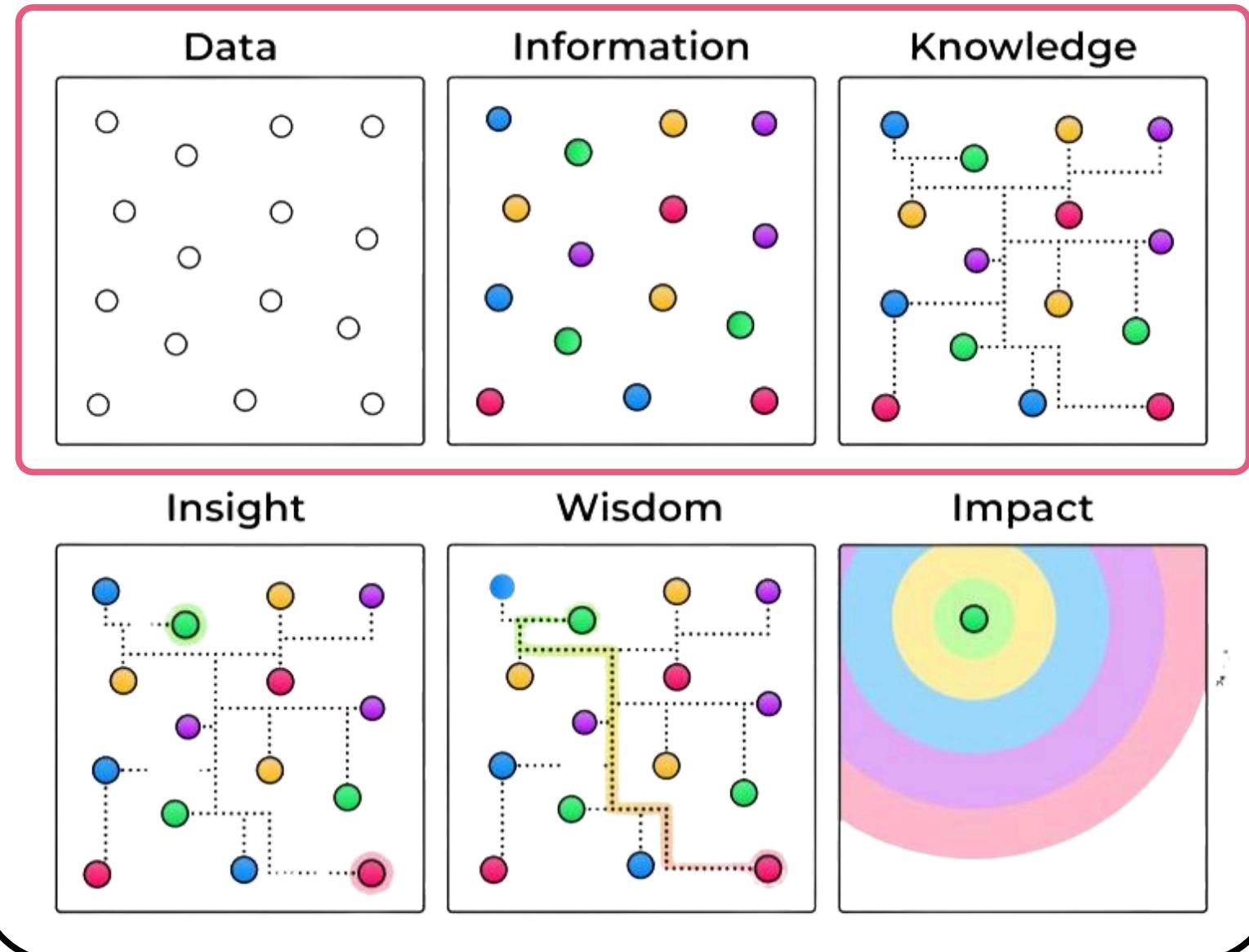
## ความเสี่ยงด้านความเป็นส่วนตัว (Privacy & Security Risks)

เข้ารหัสข้อมูล จำกัดสิทธิ์การเข้าถึง และปฏิบัติตามกฎหมาย

## 3 จาก Data สู่ Insights: ทำข้อมูลให้มีความหมาย

---

# The DIKIWI Model



**Data:** ข้อมูลที่ยังไม่ผ่านการประมวลผล (ข้อมูลดิบ) → ยังไม่รู้ว่าจะเอาไปทำอะไร

28, กรุงเทพ, 18:00

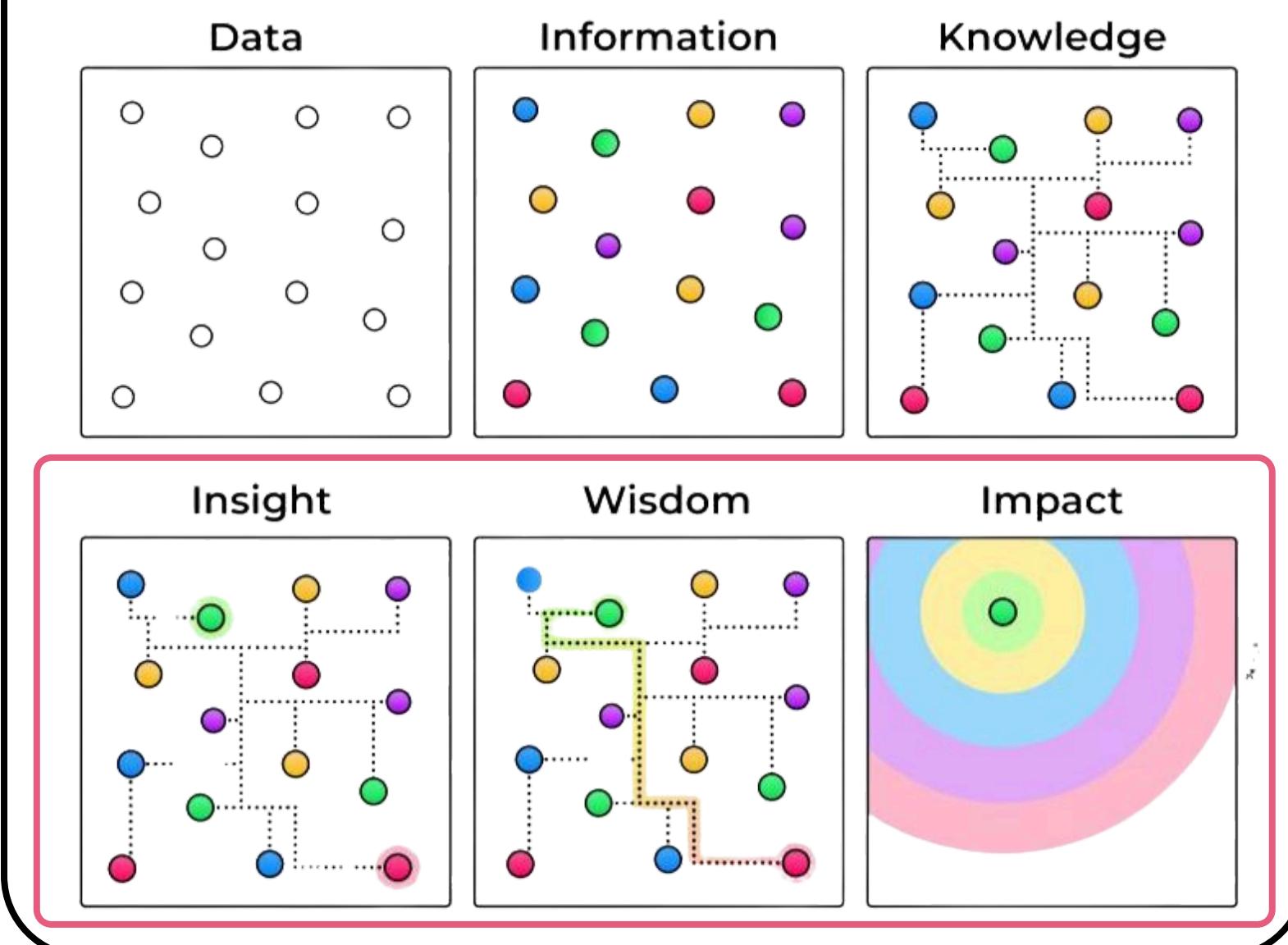
**Information:** เมื่อข้อมูลมีค่าหรือประโยชน์ ที่สามารถทำความเข้าใจได้

เวลา 18:00 อุณหภูมิ 28°C ที่กรุงเทพ

**Knowledge:** ความเข้าใจในความลับพันธ์และรูปแบบที่เชื่อมโยงกันระหว่างข้อมูลต่าง ๆ

ช่วงเย็นในกรุงเทพ มักอุณหภูมิประมาณ  
28°C

# The DIKIWI Model



**Insight:** ความเข้าใจในระดับลึกเกี่ยวกับ “เหตุผล”  
ที่อยู่เบื้องหลังความรู้

เพาะพะอาทิตย์ตก อุณหภูมิจึงลดลงในตอนเย็นทุกวัน

**Wisdom:** ใช้ Insight ในการตัดสินใจอย่างรอบคอบ มีเหตุผลรองรับ

ควรจัดกิจกรรมเต้นและโrobicตอนเย็น  
 เพราะไม่ร้อนเกินไป

**Impact:** ผลลัพธ์จริงจากการตัดสินใจนั้น

คนมาเข้าร่วมกิจกรรมเยอะ สุขภาพดีขึ้น และมีความพึงพอใจมากขึ้นในการใช้พื้นที่สาธารณะ

## **Activity: Get to know data, From data to Insights**

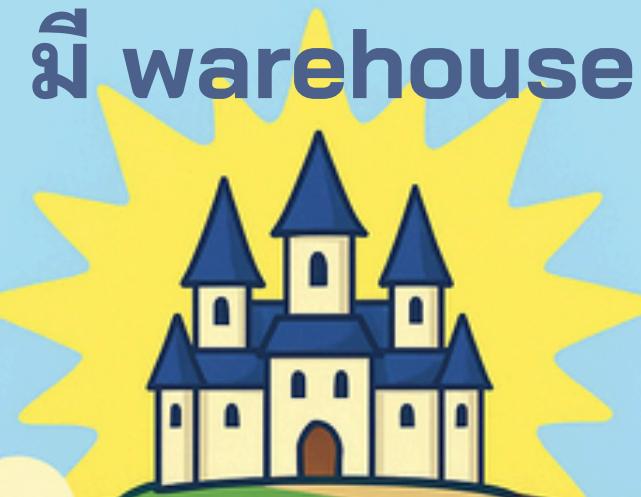
## 4 จัดระเบียบข้อมูลให้อยู่่หมวด: รู้จัก Data Warehouse และ Data Modeling

---

# Data Warehouse มีไว้ทำไม?



มี warehouse



ไม่มี warehouse



- เป็นระเบียบ
- เรียกดูง่าย
- ต้องดูแลรักษาให้เป็นระเบียบ



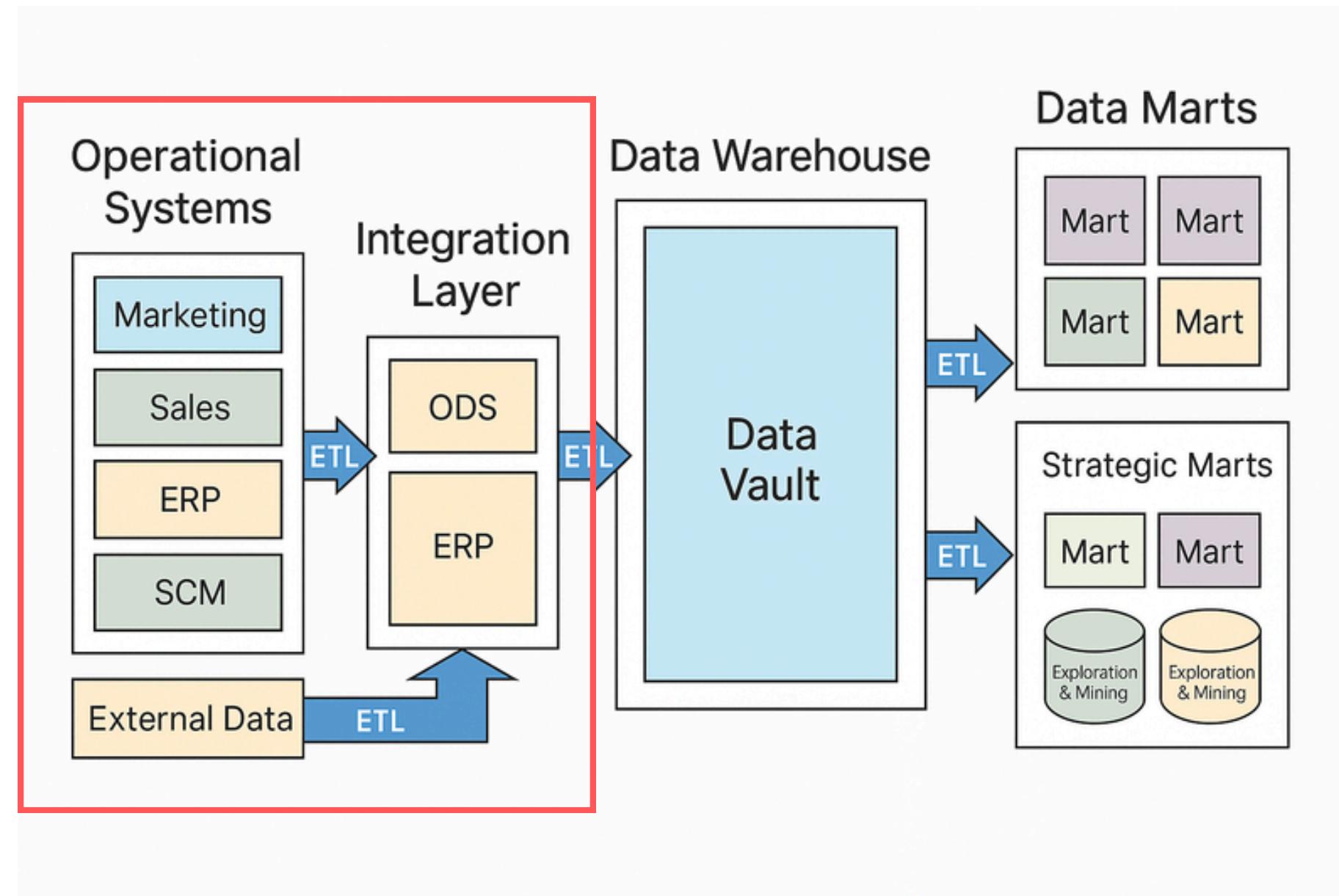
- ไม่ต้องทำอะไรกับข้อมูลมาก
- แต่เรียกข้อมูลใช้ยากและเสี่ยงต่อความผิดพลาด

## Data Warehouse มีไว้ทำไม?

- ปัจุบันในอดีต: ข้อมูลกระจายอยู่ตามแผนก → วิเคราะห์ซ้า → ไม่เห็นภาพรวม
- จุดเริ่มต้น: นักวิจัย IBM (Bill Inmon, 1990s) วางแผนคิด Data Warehouse
- แนวคิด: รวมข้อมูลจากหลายระบบเข้าเหลี่งเดียว เพื่อใช้วิเคราะห์ ไม่กระทบระบบจริง



# เข้าใจ Data Warehouse ใน 5 นาที



## Operational Systems

แหล่งข้อมูลต้นทาง เช่น ระบบการตลาด (Marketing), การขาย (Sales), ERP, SCM และ External Data

💡 ข้อมูลเหล่านี้เกิดขึ้นจากการดำเนินธุรกิจจริง

## Integration Layer

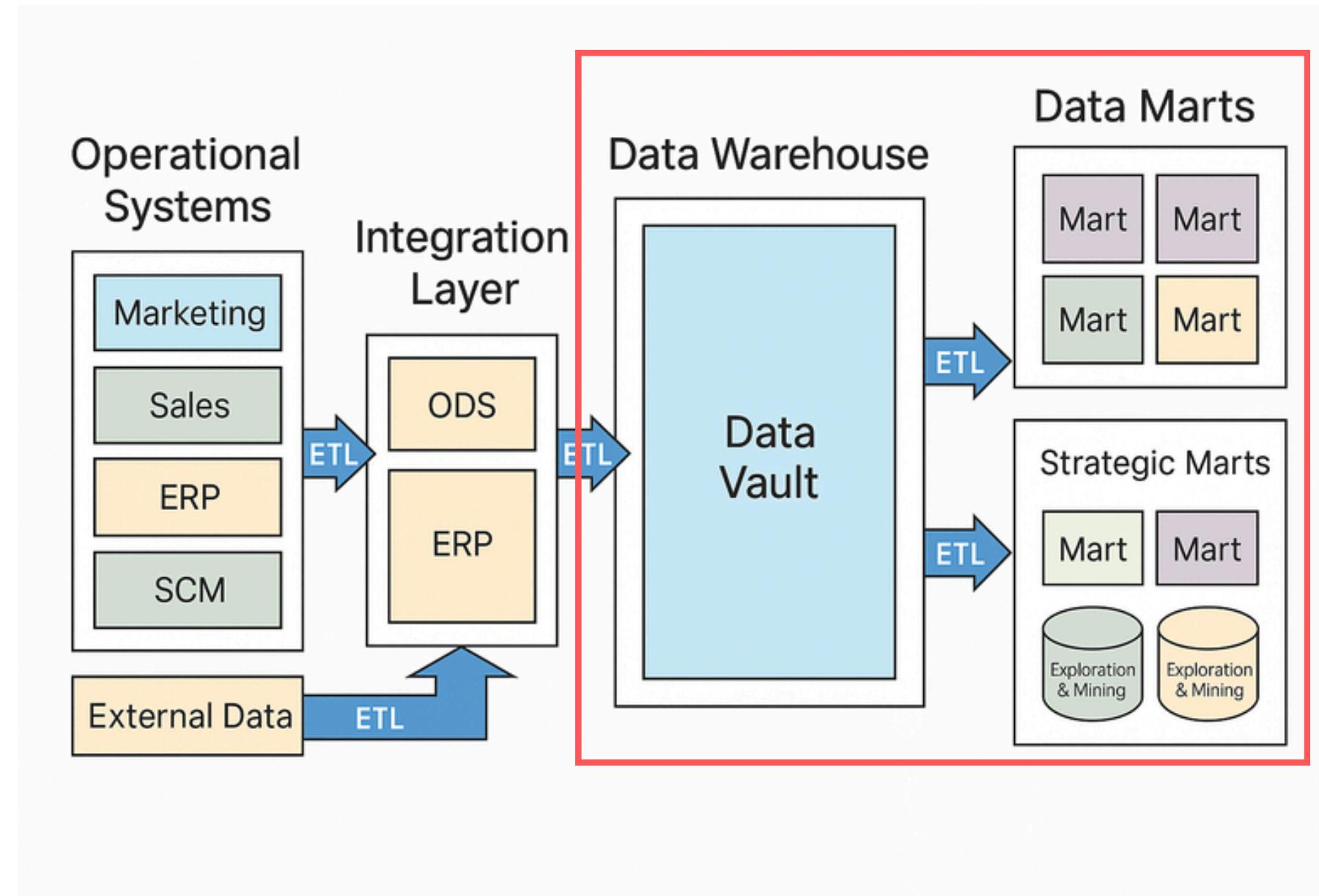
ทำหน้าที่รวมข้อมูลจากหลายระบบผ่านกระบวนการ ETL (Extract–Transform–Load)

📦 ส่งไปยัง Data Warehouse โดยไม่รบกวนระบบต้นทาง

ODS: สำหรับเก็บข้อมูลชั่วคราวหรือรายวัน

ERP Clone: สำเนาข้อมูลจากระบบ ERP เพื่อความปลอดภัย

# เข้าใจ Data Warehouse ใน 5 นาที



## Data Warehouse

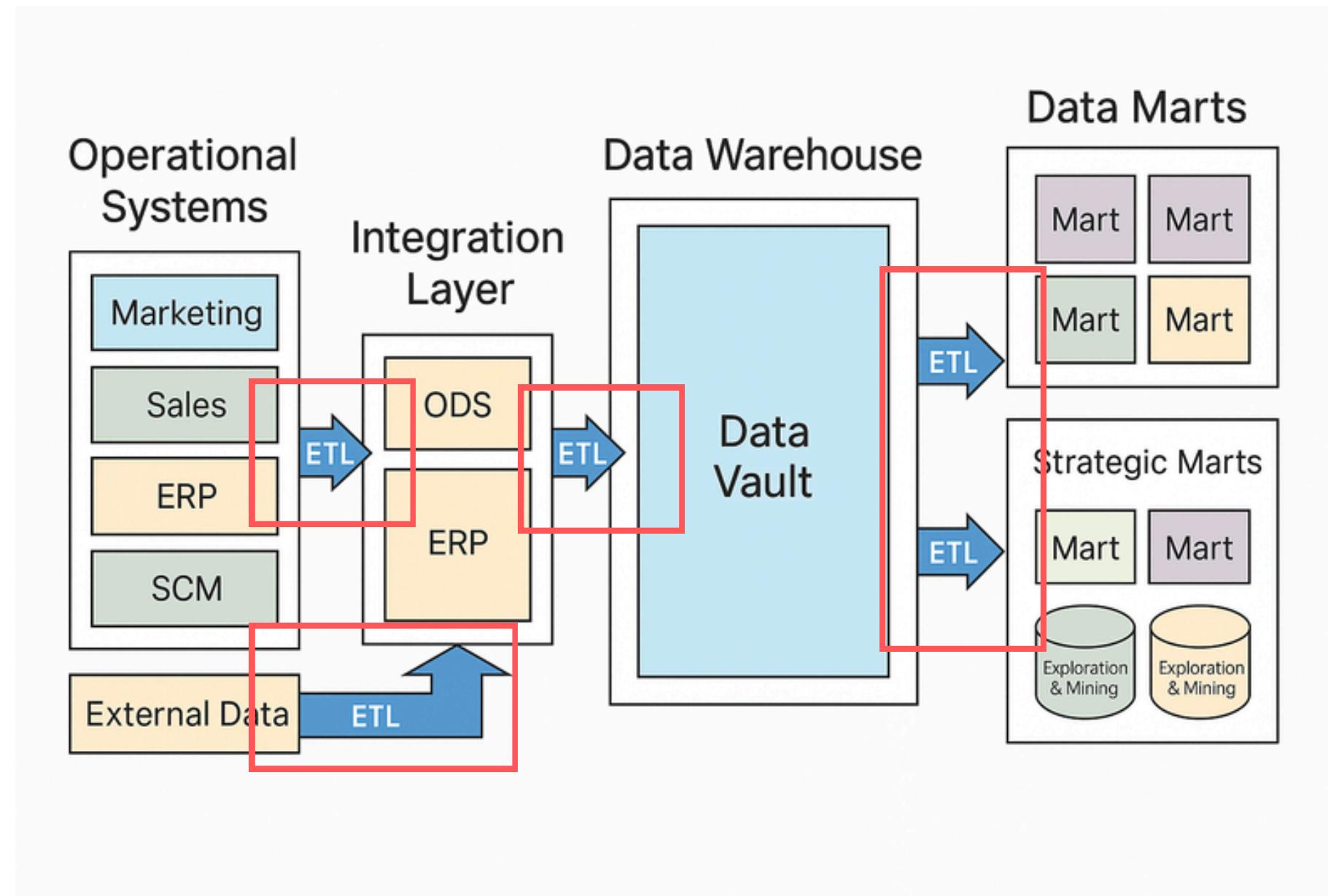
จุดศูนย์กลางข้อมูลขององค์กร  
 ใช้แบบ Data Vault ซึ่งยืดหยุ่น ปรับเปลี่ยนง่าย และติดตามได้ว่าข้อมูลมาจากไหน (traceable)

## Data Marts & Strategic Marts

แยกข้อมูลตามมุมมองผู้ใช้ เช่น Sales Mart, HR Mart

 Strategic Mart รองรับการวิเคราะห์ขั้นสูง เช่น Exploration / Data Mining

# เข้าใจ Data Warehouse ใน 5 นาที



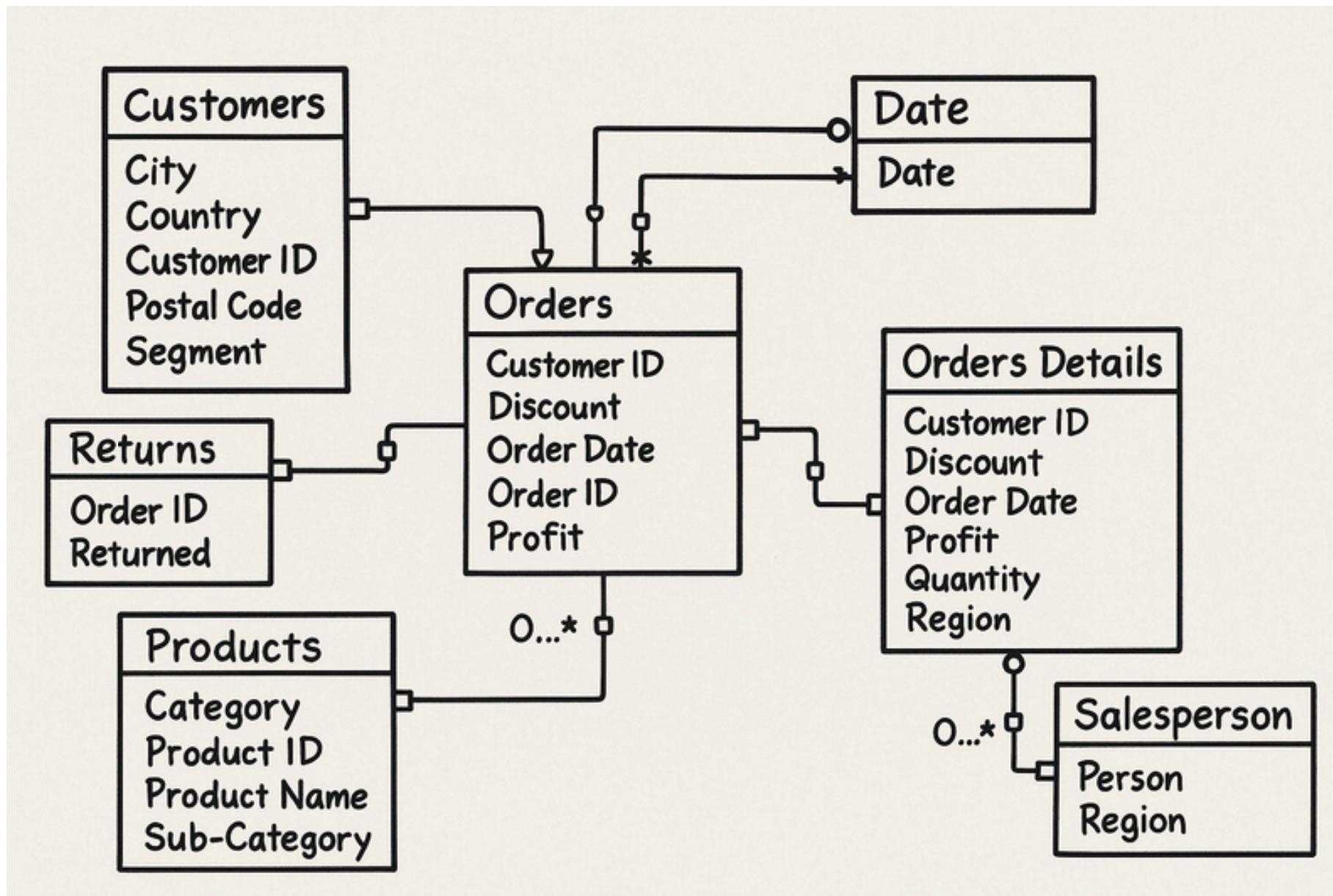
## บทบาทของ ETL (ลูกศรลีฟ้า)

- Extract: ดึงข้อมูลจากแหล่งต้นทาง
- Transform: แปลงให้อยู่ในรูปแบบเดียวกัน ตาม schema ที่กำหนด
- Load: นำเข้าคลังข้อมูลเพื่อวิเคราะห์ต่อ

**Schema** คือแบบแปลนของข้อมูล เช่น ตารางซึ่งมีคอลัมน์อะไรบ้าง ข้อมูลประเภทไหน

- ใช้กำหนดโครงสร้างให้ Transform ข้อมูลได้ตรง และ Load ได้ถูกที่
- จะได้ schema มาได้ยังไง? ต้องทำ **Data Modeling** ก่อน!

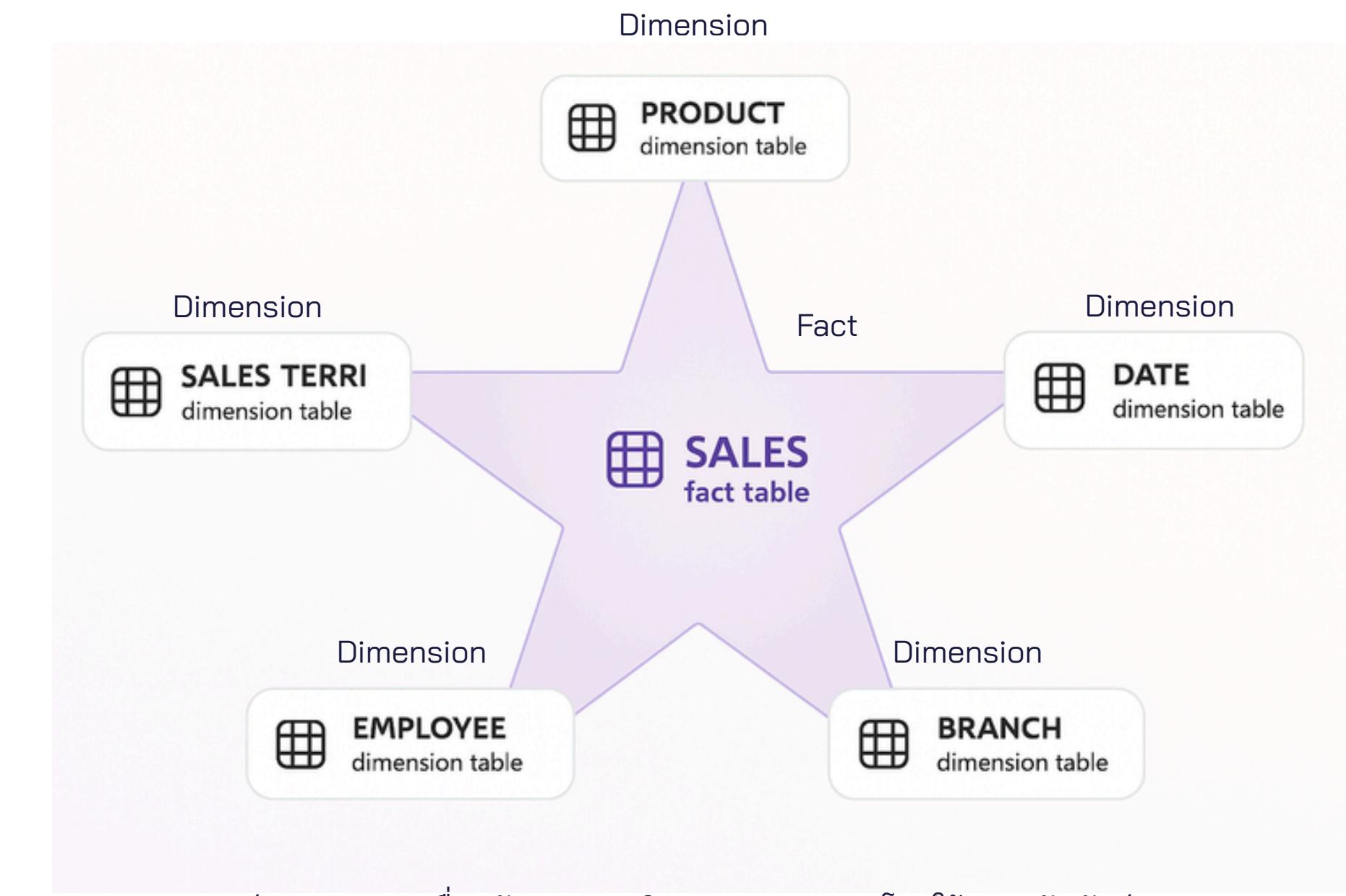
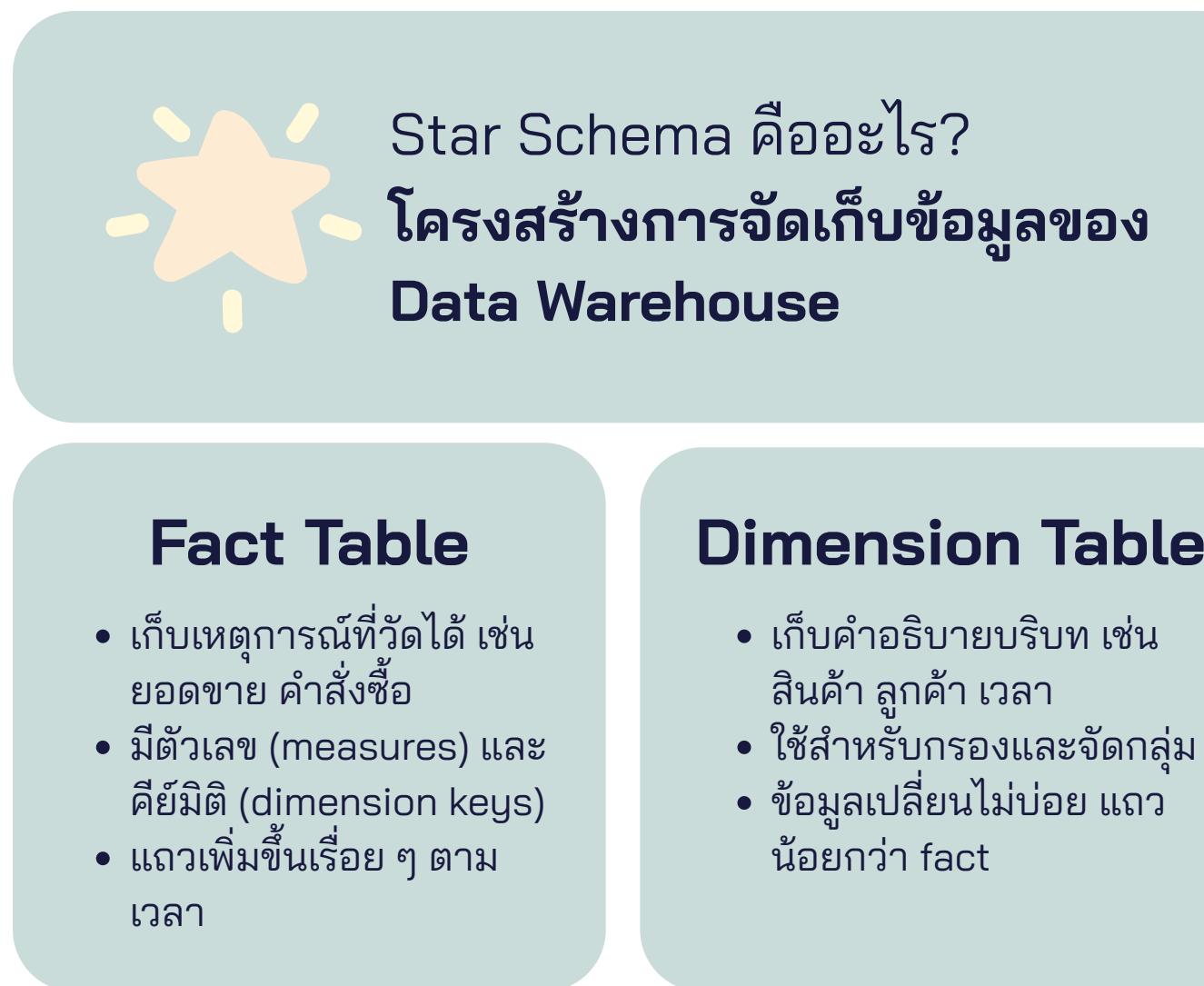
# Data Modeling



Data Modeling คือกระบวนการออกแบบโครงสร้างของข้อมูลเพื่อให้รู้ว่า:

- จะเก็บข้อมูลอะไรบ้าง (ชื่อ field, ประเภทข้อมูล)
- ข้อมูลแต่ละตารางเกี่ยวข้องกันยังไง (ความสัมพันธ์)
- ควรวางข้อมูลแบบไหนให้ใช้ง่ายและไม่ซ้ำซ้อน

# แนวคิด Star Schema และบทบาทของ Fact / Dimension Table



# ตารางแบบไม่แยก Fact/Dimension

ค่าที่ปรากฏช้า ๆ (ทำตารางให้ลูดโดยไม่จำเป็น)

เหตุการณ์ที่เก็บจริง ๆ

| OrderID | OrderDate  | ProductName | ProductCategory | CustomerName | Region     | Quantity | TotalPrice |
|---------|------------|-------------|-----------------|--------------|------------|----------|------------|
| 1001    | 2024-01-17 | Notebook    | Stationery      | Bob          | Chiang Mai | 4        | 164        |
| 1002    | 2024-01-28 | USB Drive   | Electronics     | Bob          | Chiang Mai | 2        | 40         |
| 1003    | 2024-01-16 | Stapler     | Office Supply   | Alice        | Bangkok    | 1        | 46         |
| 1004    | 2024-01-07 | Stapler     | Office Supply   | Bob          | Chiang Mai | 4        | 60         |
| 1005    | 2024-01-09 | Stapler     | Office Supply   | Alice        | Bangkok    | 2        | 50         |
| 1006    | 2024-01-09 | USB Drive   | Electronics     | Alice        | Bangkok    | 2        | 94         |
| 1007    | 2024-01-15 | USB Drive   | Electronics     | Alice        | Bangkok    | 3        | 66         |
| 1008    | 2024-01-15 | Stapler     | Office Supply   | Diana        | Bangkok    | 1        | 25         |
| 1009    | 2024-01-02 | Pen         | Stationery      | Charlie      | Phuket     | 1        | 14         |
| 1010    | 2024-01-01 | Notebook    | Stationery      | Charlie      | Phuket     | 5        | 100        |
| 1011    | 2024-01-02 | USB Drive   | Electronics     | Diana        | Bangkok    | 3        | 48         |

ทำเป็น dimension!

ทำเป็น fact

# Star Schema

ย้ายคอลัมน์ที่ซ้ำบ่อย ๆ ไปเป็น Dimension Table และเชื่อมกันด้วย “Key”



## Table: SalesFact

| OrderID | OrderDate  | ProductID | CustomerID | Quantity | TotalPrice |
|---------|------------|-----------|------------|----------|------------|
| 1001    | 2024-01-17 | 1         | 101        | 4        | 164        |
| 1014    | 2024-01-02 | 1         | 101        | 4        | 60         |
| 1017    | 2024-01-08 | 1         | 101        | 3        | 42         |
| 1002    | 2024-01-28 | 2         | 101        | 2        | 40         |
| 1004    | 2024-01-07 | 3         | 101        | 4        | 60         |
| 1012    | 2024-01-13 | 4         | 101        | 5        | 115        |
| 1010    | 2024-01-01 | 1         | 104        | 5        | 100        |
| 1009    | 2024-01-02 | 4         | 104        | 1        | 14         |
| 1015    | 2024-01-22 | 4         | 104        | 2        | 48         |
| 1019    | 2024-01-25 | 4         | 104        | 2        | 26         |
| 1006    | 2024-01-09 | 2         | 102        | 2        | 94         |
| 1007    | 2024-01-15 | 2         | 102        | 3        | 66         |
| 1018    | 2024-01-16 | 2         | 102        | 3        | 114        |

## Table: Product

| ProductID | ProductName | ProductCategory |
|-----------|-------------|-----------------|
| 1         | Notebook    | Stationery      |
| 2         | USB Drive   | Electronics     |
| 3         | Stapler     | Office Supply   |
| 4         | Pen         | Stationery      |

## Table: Customer

| CustomerID | CustomerName | Region     |
|------------|--------------|------------|
| 101        | Bob          | Chiang Mai |
| 102        | Alice        | Bangkok    |
| 103        | Diana        | Bangkok    |
| 104        | Charlie      | Phuket     |

# Fact Table: Primary Key, Foreign Key, Measures



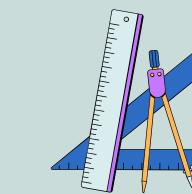
| OrderID |
|---------|
| 1001    |
| 1014    |
| 1017    |
| 1002    |
| 1004    |

**Primary Key:** คอลัมน์แสดงเลขเฉพาะของแต่ละเดา



| ProductID | CustomerID |
|-----------|------------|
| 1         | 101        |
| 1         | 101        |
| 1         | 101        |
| 2         | 101        |
| 3         | 101        |

**Foreign Key:** คอลัมน์ที่ใช้เชื่อมกับตาราง Dimension



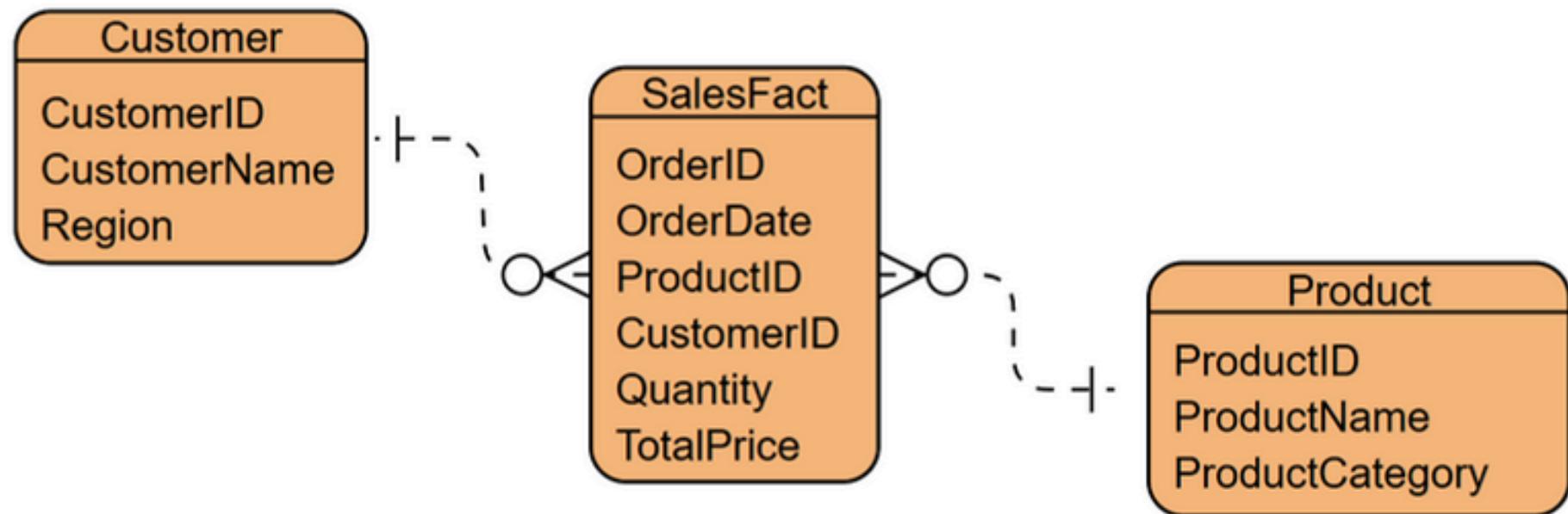
| Quantity | TotalPrice |
|----------|------------|
| 4        | 164        |
| 4        | 60         |
| 3        | 42         |
| 2        | 40         |
| 4        | 60         |

**Measures:** คอลัมน์แสดงค่าจากเหตุการณ์ที่บันทึกข้อมูลจริง

# Entity Relation Diagram (ERD)

ERD (Entity-Relationship Diagram) คืออะไร?

ERD คือแผนภาพที่ใช้แสดงโครงสร้างความสัมพันธ์ของข้อมูลในระบบฐานข้อมูลช่วยให้เราเห็น ตาราง (Entities) ที่สำคัญ และความเชื่อมโยงกันระหว่างตารางเหล่านั้น



💡 มีประโยชน์อย่างไร?

- เข้าใจโครงสร้างข้อมูลอย่างเป็นระบบ
- ใช้ในการออกแบบฐานข้อมูล หรือ วางแผน Data Warehouse
- สื่อสารกับทีมงาน (Dev, Analyst, BI) ได้ง่ายขึ้น

# ได้ Data Warehouse แล้วทำยังไงต่อ?



## Data governance & Security

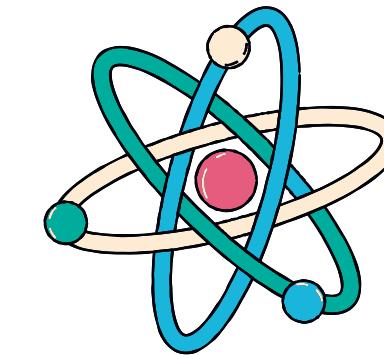
Data Steward / Data Governance Lead / IT Security

- ควบคุมสิทธิ์การเข้าถึงข้อมูล (Access Control)
- ตรวจสอบว่าใช้ข้อมูลสอดคล้องกับกฎหมาย (PDPA/GDPR)
- บันทึกการใช้ข้อมูล (audit log), ทำ data lineage

## Data Visualization & BI

BI Developer / Data Analyst / Business Analyst

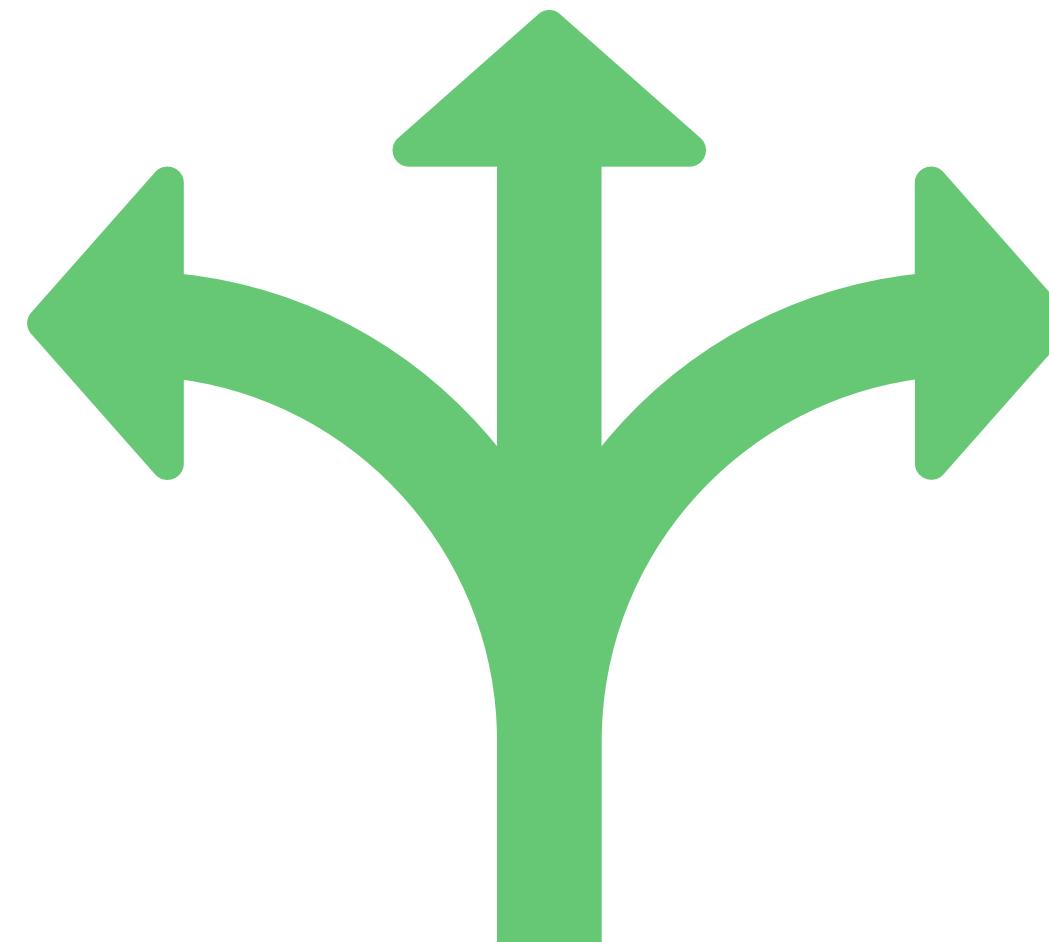
- เชื่อมต่อ DWH กับเครื่องมือ BI เช่น Power BI, Tableau
- สร้าง dashboard / report ที่ลือสารกับผู้บริหารหรือทีมงาน
- ตั้ง KPI, Visual Analysis



## Advanced Analytics & AI/ML

Data Scientist / ML Engineer / AI Team

- ใช้ข้อมูลจาก DWH ในการฝึก ML Model
- ค้นหา pattern ซ่อนอยู่ในข้อมูล (clustering, anomaly detection)
- สร้างระบบ recommendation / forecast / automation

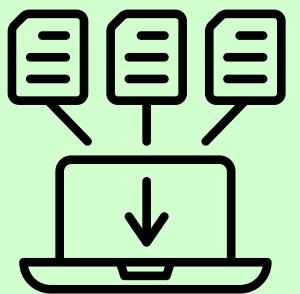


## Activity: Data Warehouse

## 5 เตรียมข้อมูลให้พร้อมใช้งาน: ETL, Data Cleansing และ Data Quality

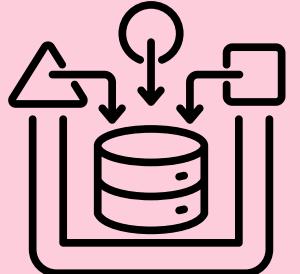
---

# ETL: เตรียมข้อมูลให้พร้อมใช้



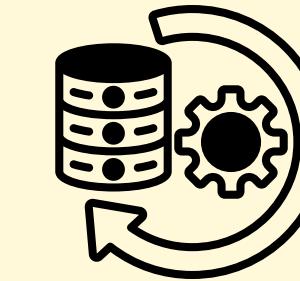
## Extract – ดึงข้อมูล

- ดึงข้อมูลจากแหล่งต่าง ๆ เช่น ฐานข้อมูล ERP, CRM, Excel, IoT ฯลฯ
- ข้อมูลดิบที่ได้อาจ ยังไม่ตรงกัน เช่น บางอัน เก็บวันที่เป็น 2024-06-23, บางอันเป็น 23/6/2024
- หรือบางแหล่งมีค่าว่าง, ข้อมูลซ้ำซ้อน, ใช้หน่วยไม่เหมือนกัน



## Load – โหลดข้อมูลเข้าสู่ปลายทาง

เมื่อข้อมูล "สะอาดและเป็นระเบียบแล้ว" เราจึงค่อยโหลดเข้า Database, Data Warehouse, Dashboard, หรือระบบ BI ต่าง ๆ ถ้าเราโหลดข้อมูลที่ยังไม่แปลง → ปลายทางจะอ่านไม่ออก / คำนวนผิด / ทำให้ระบบพัง



## Transform – แปลงข้อมูล

- ขั้นตอนนี้คือ "เตรียมและจัดระเบียบข้อมูลให้พร้อมใช้งาน"

### สิ่งที่มักต้องทำในขั้นนี้:

- ลบข้อมูลซ้ำ / ค่าว่าง
- รวม field หรือแยก field เช่น ชื่อ-นามสกุล → ชื่อ กับ นามสกุล
- แปลงหน่วย เช่น จาก inch เป็น cm
- แปลง format ให้เหมือนกัน เช่น วันที่, เบอร์โทรศัพท์
- สร้างคอลัมน์ใหม่ เช่น ยอดขายรวม = ราคาต่อหน่วย × จำนวน

ถ้าไม่ Transform ก่อน → ข้อมูล ม้วน/ซ้อน/ใช้ไม่ได้ -tonviradee-h

# จงหาความผิดปกติในข้อมูลนี้

| CustomerID | Name    | Age   | Email                 | Signup Date | Address       | Favorite Color |
|------------|---------|-------|-----------------------|-------------|---------------|----------------|
| 101        | Alice   | 25.0  | alice@example.com     | 2021-01-01  | 123 Main St   | Blue           |
| 102        | Bob     | 30.0  | bob@example.com       | 2021-02-15  | 456 Oak St    | Green          |
| 103        | Charlie | 22.0  | charlie@ex.com        | 2021-03-10  | 789 Pine St   | Red            |
| 103        | Charlie | 22.0  | charlie@ex.com        | 2021-03-10  | 789 Pine St   | Red            |
| 105        | Eva     | 29.0  | eva@example.com       | 2021-04-20  | 321 Maple St  | Blue           |
| 106        | Frank   |       | frank@example.co<br>m | 2021-05-30  | 654 Birch St  | Yellow         |
| 107        | Grace   | 31.0  | grace@example.co<br>m |             | 987 Cedar St  | Purple         |
| 108        | Henry   | 28.0  | henry@example.co<br>m | 2021-07-25  | 543 Elm St    | Green          |
| 109        | Ivy     | 999.0 | ivy@example.com       | 2021-08-05  | 111 Ash St    | Blue           |
| 110        | Jake    | 27.0  | jake@example.com      | 2021-09-10  | 222 Spruce St | Orange         |

# Data Cleansing: เพาะข้อมูลไม่ได้น่ารักสำหรับทุกคน

| CustomerID | Name    | Age   | Email             | Signup Date | Adress        | Favorite Color |
|------------|---------|-------|-------------------|-------------|---------------|----------------|
| 101        | Alice   | 25.0  | alice@example.com | 2021-01-01  | 123 Main St   | Blue           |
| 102        | Bob     | 30.0  | bob@example.com   | 2021-02-15  | 456 Oak St    | Green          |
| 103        | Charlie | 22.0  | charlie@ex.com    | 2021-03-10  | 789 Pine St   | Red            |
| 103        | Charlie | 22.0  | charlie@ex.com    | 2021-03-10  | 789 Pine St   | Red            |
| 105        | Eva     | 29.0  | eva@example.com   | 2021-04-20  | 321 Maple St  | Blue           |
| 106        | Frank   |       | ไม่มีข้อมูล       |             |               |                |
| 107        | Grace   | 31.0  | grace@example.com | ไม่มีข้อมูล | 987 Cedar St  | Purple         |
| 108        | Henry   | 28.0  | henry@example.com | 2021-07-25  | 543 Elm St    | Green          |
| 109        | Ivy     | 999.0 | ivy@example.com   | 2021-08-05  | 111 Ash St    | Blue           |
| 110        | Jake    | 27.0  | jake@example.com  | 2021-09-10  | 222 Spruce St | Orange         |

ไม่เกี่ยวข้อง (Irrelevant Columns)

→ ลบคอลัมน์ที่ไม่จำเป็นต่อการวิเคราะห์

พิมพ์ผิด (Typo / Structural Errors)

→ แก้ไขให้ตรงตามมาตรฐาน เช่น

"Adress" → "Address"

ซ้ำ (Duplicate)

→ ลบแถวที่ซ้ำกันออก (De-duplication)

ไม่มีข้อมูล (Missing Values)

→ ถ้าหายเสีย: ทิ้งทั้งแถว

→ ถ้าหายเล็กน้อย: แทนที่ด้วยค่ากลาง เช่น  
ค่าเฉลี่ย / ค่ากลาง / ค่าที่พบบ่อย

Outlier (ค่าผิดปกติ)

→ พิจารณาทิ้ง / แก้ไข / แยกออกไปวิเคราะห์ต่างหาก

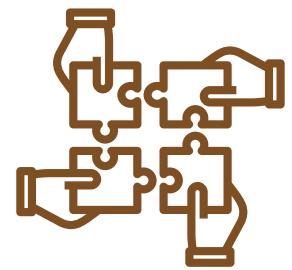
# 6 เสาหลักของ Data Quality

Data Quality คืออะไร?

**Data Quality** คือระดับความถูกต้อง ครบถ้วน ทันสมัย สม่ำเสมอ และเหมาะสมของข้อมูล  
ยิ่งข้อมูลคุณภาพดี → วิเคราะห์แม่นยำ → ตัดสินใจอย่างมั่นใจ

C L E A N S

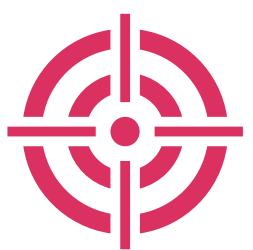
Consistency: ข้อมูล  
สอดคล้องกันทั้งระบบ



Level of Detail: ข้อมูล  
ละเอียดพอดี



Exactness: ข้อมูลถูก  
ต้องตามความจริง



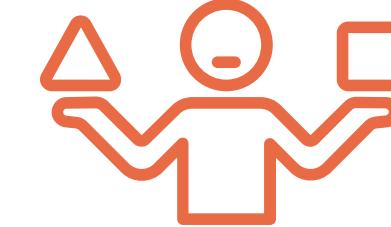
Availability: ข้อมูลครบ  
ถ้วนพร้อมใช้ ไม่มีช่องว่าง



Now: ข้อมูลเป็น  
ปัจจุบัน



Singularity: ข้อมูล  
ไม่ซ้ำซ้อน



## Activity: ETL, Data Cleansing, Data Quality

## 6 Data Analytics: สร้างคุณค่าจากข้อมูลดิบ

---

# กระบวนการวิเคราะห์ข้อมูล: จากข้อมูลดิบสู่ข้อมูลเชิงกลยุทธ์



Data  
Collection

เก็บข้อมูลดิบจากแหล่งต่าง ๆ เช่น ฐานข้อมูล, API, และแบบสอบถาม



Data  
Cleaning

ลบข้อผิดพลาด, จัดการข้อมูลที่หายไป และทำให้รูปแบบข้อมูลเป็นมาตรฐานเดียวกัน



Data  
Processing

แปลงข้อมูลที่สะอาดแล้วด้วยเทคนิคต่าง ๆ เช่น Feature Engineering และ Normalization



Data  
Analysis

สำรวจข้อมูล, ทดสอบสมมติฐาน และประยุกต์ใช้โมเดล เพื่อค้นหาข้อมูลเชิงลึก



Decision-  
Making

ใช้ข้อมูลเชิงลึกเพื่อกำหนดการดำเนินการ ผ่านรายงาน, แผนภาพ และแดชบอร์ด

Project start



ในหลาย ๆ โครงการ พบว่า Data cleaning กินเวลานานที่สุด เพราะข้อมูลที่ได้มามักไม่ได้พร้อมใช้งาน

# กระบวนการวิเคราะห์ข้อมูล: จากข้อมูลดิบสู่ข้อมูลเชิงกลยุทธ์



## Data Collection

**Task:** เก็บข้อมูลดิบจากหลายแหล่ง เช่น

- ฐานข้อมูล (databases)
- API
- การดึงข้อมูลจากเว็บ (web scraping)
- แบบสอบถาม (surveys)
- ชุดข้อมูลที่มีอยู่แล้ว (existing datasets)



## Data Cleaning

**Task:** จัดการข้อมูลดิบโดย

- เติมค่าที่หายไป (missing values)
- ลบข้อมูลซ้ำ (duplicates)
- ทำให้รูปแบบข้อมูลเป็นมาตรฐานเดียวกัน (standardizing formats)
- ทำให้ข้อมูลมีความสอดคล้อง (data consistency)



## Data Processing

**Task:** หลังจากทำความสะอาดแล้ว ต้องแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมต่อการวิเคราะห์

- การสร้างฟีเจอร์ (feature engineering)
- การแปลงข้อมูลประเภท category (encoding)
- การทำ normalization
- การรวมข้อมูล (aggregation)

### Challenges

คุณภาพของข้อมูล

ความเป็นส่วนตัวและ  
ความปลอดภัย

ใช้งบประมาณเยอะ

### Challenges

ข้อมูลไม่ครบถ้วน

ความผิดพลาดในข้อมูล

ปัญหาเรื่องรูปแบบข้อมูล

### Challenges

ปริมาณข้อมูลจำนวนมาก

การปรับแต่งอัลกอริธึม

ความซับซ้อนของการ  
สร้างฟีเจอร์

# กระบวนการวิเคราะห์ข้อมูล: จากข้อมูลดิบสู่ข้อมูลเชิงกลยุทธ์



## Data Analysis

**Task:** ขั้นตอนนี้ประกอบด้วย:

- การวิเคราะห์ข้อมูลเชิงสำรวจ (EDA)
- การทดสอบสมมติฐาน (Hypothesis Testing)
- การสร้างแบบจำลองทำนาย (Predictive Modeling)
- การแสดงผลข้อมูล (Visualization)
- การสร้างโมเดล Machine Learning



## Decision-Making

**Task:** นำข้อมูลเชิงลึกจากขั้นวิเคราะห์ไปใช้กำหนดกลยุทธ์ คำแนะนำ และการตัดสินใจต่าง ๆ

อาจรวมถึง:

- การจัดทำรายงาน (Reports)
- การสร้าง Visualization
- การทำ Dashboard สำหรับผู้มีส่วนเกี่ยวข้อง

### Challenges

เลือกวิธีการวิเคราะห์ให้เหมาะสม  
Storytelling  
Computation cost

### Challenges

การตีความผลวิเคราะห์ให้ถูกต้อง  
ทำให้สอดคล้องกับเป้าหมายทางธุรกิจ  
อคติทางความคิดหรืออัจฉริยา

# นักวิเคราะห์ 4 ผ่า

## Descriptive



- นักประวัติศาสตร์ (The Historian)
- สรุปและเล่าเรื่องเหตุการณ์ในอดีต
- ใช้รายงาน แดชบอร์ด และแนวโน้มต่าง ๆ เพื่ออธิบายว่า
- "เกิดอะไรขึ้น?"

## Diagnostic



- นักสืบ (The Investigator)
- ค้นหาสาเหตุว่าเหตุใดลิ่งนั้นจึงเกิดขึ้น
- วิเคราะห์รูปแบบและสาเหตุในข้อมูลเพื่อหาคำตอบว่า
- "ทำไมมันถึงเกิดขึ้น?"

## Predictive



- หมอดู (The Fortune Teller)
- ใช้ข้อมูลในอดีตเพื่อทำนายสิ่งที่จะเกิดในอนาคต
- พัฒนาโมเดลทางสถิติและ Machine Learning
- "แล้วต่อไปจะเกิดอะไรขึ้น?"

## Prescriptive



- นักวางแผนกลยุทธ์ (Strategist)
- แนะนำวิธีที่ดีที่สุดในการตัดสินใจ หรือดำเนินการ
- ใช้ AI เพื่อแนะนำทางเลือกเชิงกลยุทธ์และเพิ่มประสิทธิภาพธุรกิจ
- "ควรทำอย่างไรดี?"

# นักประวัติศาสตร์ (The Historians) – การวิเคราะห์เชิงพรรณนา (Descriptive Analysis)



เราตีแผ่ข้อมูลในอดีต เพื่อให้คุณ  
เข้าใจตนเอง



คำถามประจำตัว  
“เกิดอะไรขึ้น?”

- **Tools:** การแสดงผลข้อมูล (Data Visualization), กราฟ (Graphs), รายงาน (Reports), แดชบอร์ด (Dashboards)



ทำไมผ่านนี้จึงสำคัญ

- เป็นจุดเริ่มต้นของการทำความเข้าใจข้อมูล ก่อนนำไปทำอย่างอื่น
- สื่อสารและตัดสินใจง่ายขึ้น: แสดงข้อมูลด้วยกราฟ ตาราง หรือสถิติเบื้องต้น ทำให้เข้าใจง่ายและใช้ประกอบการตัดสินใจ
- เป็นพื้นฐานของการวิเคราะห์ขั้นสูง: ก่อนจะทำนายหรือปรับกลยุทธ์ ต้องเริ่มจากเข้าใจข้อมูลด้วยการวิเคราะห์เชิงพรรณนา

# นักสืบ (The Investigators) – การวิเคราะห์เชิงวินิจฉัย (Diagnostic Analysis)



เราค้นหาความจริงที่ซ่อนอยู่ เพื่อให้คุณ  
มั่นใจว่าจะไม่เกิดปัญหาเดิม ๆ

คำถามประจำตัว  
❓❓❓ “ทำไมสิ่งนี้ถึงเกิดขึ้น?”

**Tools:** BI Tools, Query & Exploration,  
Statistical Tests, Data Profiling Tools,  
Log/Trace Analysis



ทำไมผ่านนี้จึงสำคัญ

- ป้องกันปัญหาในอนาคต ด้วยการหาสาเหตุที่แท้จริง
- เปลี่ยนข้อมูลให้เป็น ทางแก้ปัญหาที่จับต้องได้ (Actionable Solutions) เช่น การปรับปรุงคุณภาพหรือประสิทธิภาพ
- ยกระดับการตัดสินใจ ด้วยการเปิดเผยตัวแปรสำคัญที่ทำให้  
สำเร็จหรือล้มเหลว

# หมอดู (The Fortune Tellers) – การวิเคราะห์เชิงพยากรณ์ (Predictive Analysis)

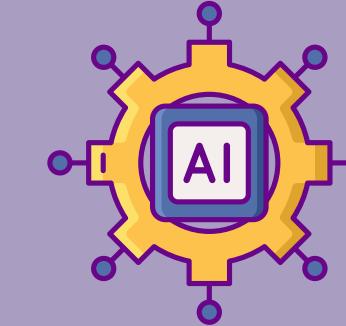


เราพยายามให้คุณเห็นอนาคต เพื่อให้เตรียมตัว  
รับมือได้อย่างมั่นใจ



คำถามประจำตัว  
“สิ่งใดจะเกิดขึ้นต่อไป?”

**Tools:** Machine learning models,  
statistical methods



## ทำไมเป่านี้จึงสำคัญ

- ลดความไม่แน่นอน ด้วยการให้ข้อมูลล่วงหน้าเกี่ยวกับแนวโน้มต่าง ๆ
- ช่วยให้องค์กร นำหน้าคู่แข่ง ด้วยการตัดสินใจเชิงรุกจากการวิเคราะห์แนวโน้ม
- ประหยัดทั้งเงินและทรัพยากร โดยการ ป้องกันความเสียหาย และเพิ่มประสิทธิภาพการดำเนินงาน

# นักวางแผนกลยุทธ์ (The Strategists) – การวิเคราะห์เชิงสั่งการ (Prescriptive Analysis)



เรารอออกแบบเส้นทางที่ดีที่สุด เพื่อให้คุณบรรลุ  
เป้าหมายอย่างแม่นยำ



คำถามประจำตัว  
“เราควรทำอะไรต่อ?”

**Tools:** อัลกอริธึมการ optimization  
 เช่น Linear Programming, แบบ  
 จำลองทางสถิติ



## ทำไมเฝ่านี้จึงสำคัญ

- เปลี่ยนจาก ข้อมูลเชิงลึก → การกระทำ **ไม่ใช่แค่คาดการณ์** แต่  
แนะนำทางเลือกที่เหมาะสมที่สุด
- ช่วยให้องค์กร เพิ่มประสิทธิภาพและผลกำไร ได้สูงสุด
- เป็น ขั้นสุดท้ายของการวิเคราะห์ขั้นสูง – เปลี่ยนข้อมูลเป็นการ  
ตัดสินใจที่ชาญฉลาด

## Activity: The 4 Types of Analytics

## 7 พลังแห่ง Visualization: สื่อสารภาษา Data ในพริบตา

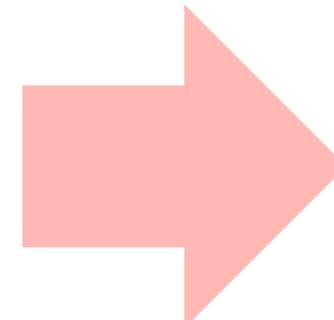
---

# Data Visualization คืออะไร

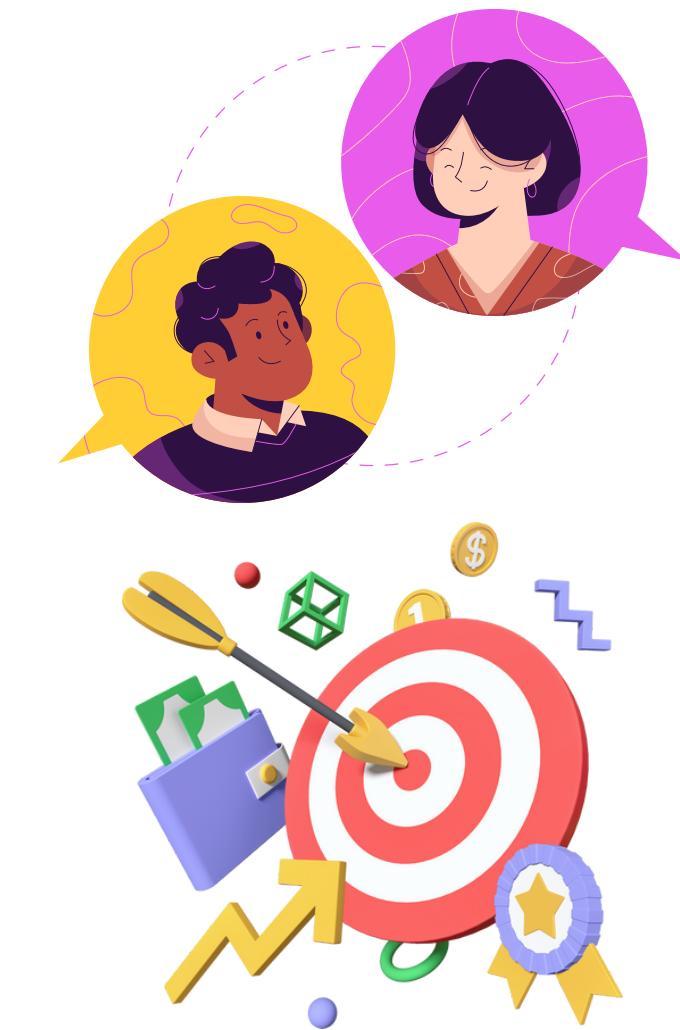
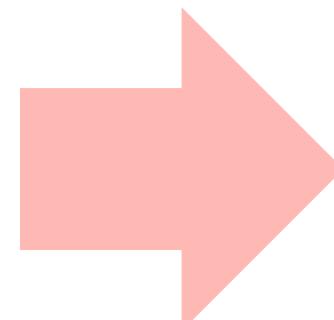
“กระบวนการแปลงข้อมูลดิบ ให้กลายเป็น รูปแบบภาพที่เข้าใจง่าย เพื่อให้สามารถสื่อสารและตัดสินใจได้อย่างมีประสิทธิภาพ”



**Raw Data**  
**(Hard to digest)**



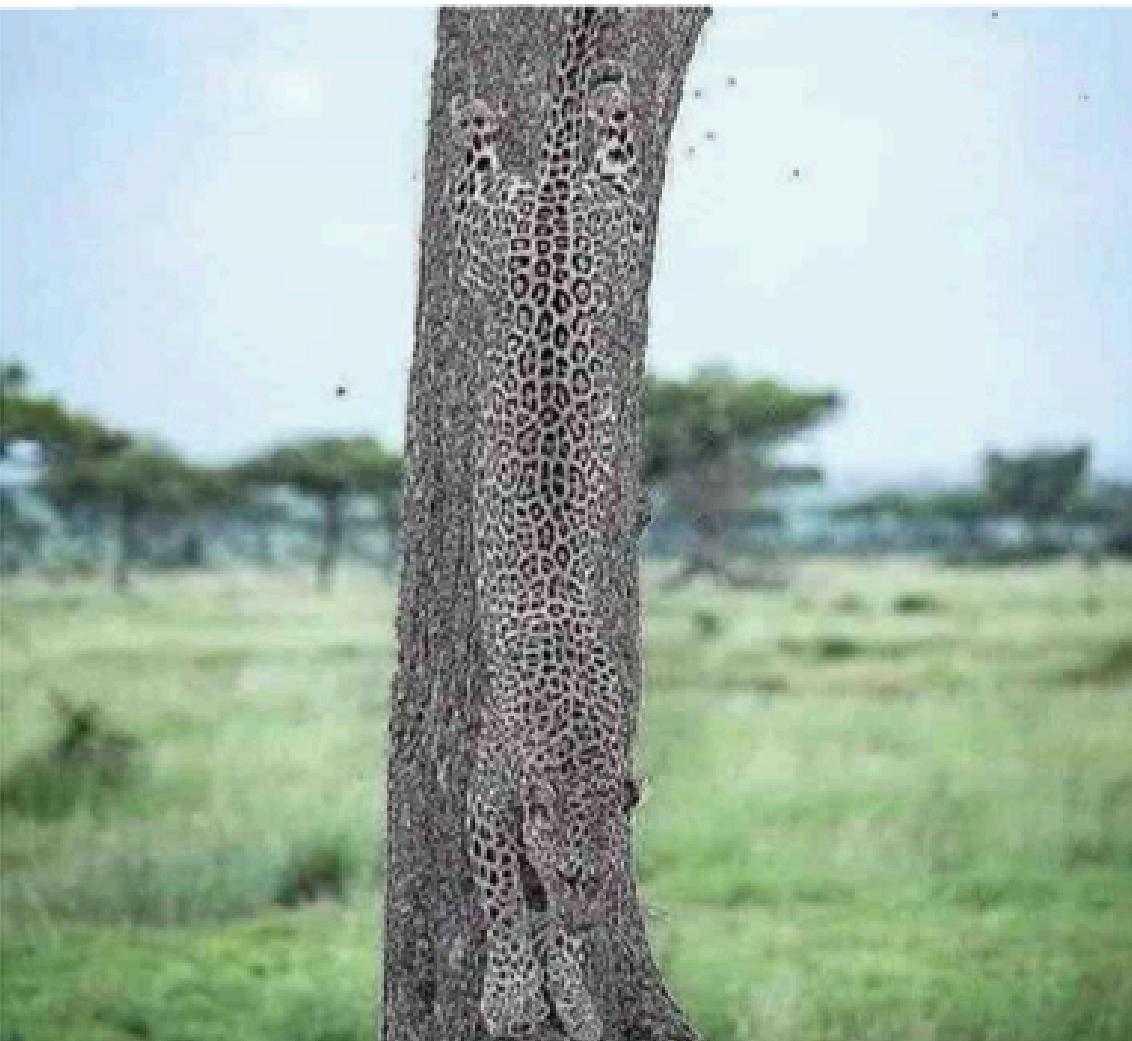
**Visual  
Representation**



**Better  
Communication**  
**Better Decision-  
Making**

# ทำไมต้อง Visualize

มนุษย์สามารถมองเห็นรูปแบบ (pattern) ได้รวดเร็ว ผ่านการมองเห็น  
 เพราะเรา vi วัฒนาการมาเพื่อลิ่งนี้



# เข้าใจประเภทของ Data ก่อนเลือก Chart



| ประเภท       | ความหมาย                                                      | ตัวอย่าง                   | ข้อสังเกต                                                      |
|--------------|---------------------------------------------------------------|----------------------------|----------------------------------------------------------------|
| Quantitative | ข้อมูลเชิงปริมาณ บวกลบ คุณหารได้                              | รายได้, คะแนน สอ卜          | เหมาะกับกราฟที่แสดงปริมาณ เช่น Bar, Line, Histogram            |
| Categorical  | ข้อมูลเชิงคุณภาพ การนำ มากบลบคุณหารทำไม่ได้ หรือไม่มีความหมาย | เพศ, ประเทศ, เบอร์โทรศัพท์ | เหมาะกับการเปรียบเทียบแต่ละ หมวดหมู่ เช่น Pie Chart, Bar Chart |

# เลือก Chart ให้เหมาะสมกับวัตถุประสงค์



| ประเภท     | ตัวอย่าง                  | วัตถุประสงค์                                              | กรณีที่ไม่ควรใช้                                       |
|------------|---------------------------|-----------------------------------------------------------|--------------------------------------------------------|
| Bar Chart  | ยอดขายแต่ละสาขา           | Convince ว่า “อันไหนสูง กว่า อันไหนต่ำกว่า”               | เมื่อมีหมวดหมู่เยอะมากจน แน่น (เช่น >15 categories)    |
| Line Chart | แนวโน้มยอดขายราย เดือน    | Convince ว่า “สิ่งนี้เพิ่ม ขึ้น / ลดลง / มีแนวโน้ม ยังไง” | ข้อมูลที่ไม่ใช่ลำดับเวลา เช่น เพศ, ประเทศ              |
| Pie Chart  | สัดส่วนยอดขายแต่ละ สินค้า | Convince ว่า “อันไหน เป็นส่วนใหญ่ของ ทั้งหมด”             | ค่าต่างกันเล็กน้อยจนดูไม่ ออก หรือมีหมวดหมู่มากเกิน ไป |

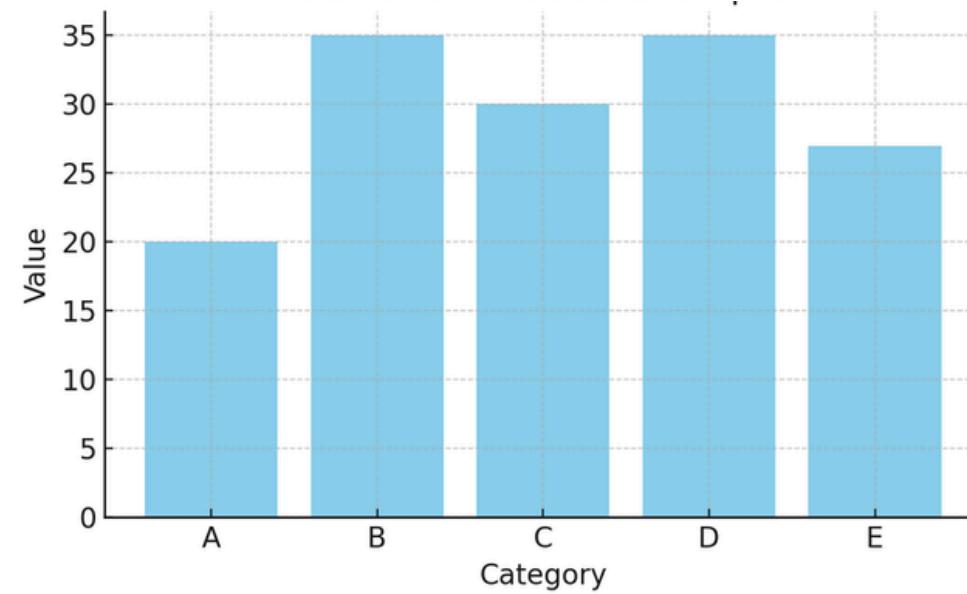
# เลือก Chart ให้เหมาะสมกับวัตถุประสงค์



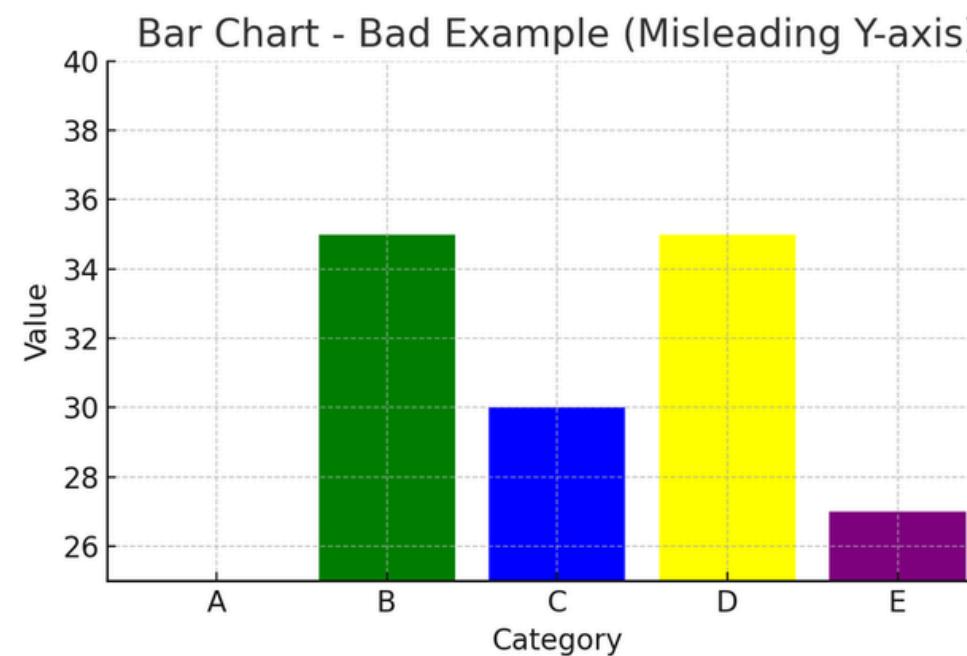
| ประเภท       | ตัวอย่าง                                    | วัตถุประสงค์                                               | กรณีที่ไม่ควรใช้                                 |
|--------------|---------------------------------------------|------------------------------------------------------------|--------------------------------------------------|
| Scatter Plot | ความสัมพันธ์ระหว่างอุณหภูมิกับยอดขายไอศครีม | Convince ว่า “สองสิ่งนี้มีความสัมพันธ์กัน (หรือไม่/แบบใด)” | กับข้อมูลที่ไม่เป็นตัวเลข หรือข้อมูลเชิงหมวดหมู่ |
| Histogram    | การกระจายคะแนนสอบ                           | Convince ว่า “ส่วนใหญ่ของข้อมูลอยู่ตรงไหน”                 | ถ้าเป็น categorical ห้ามใช้                      |

# Bar Chart (แผนภูมิแท่ง)

แกน Y (แกนตั้ง): ความสูงของแท่งแทนปริมาณ, แกน X (แกนนอน): แต่ละ Category



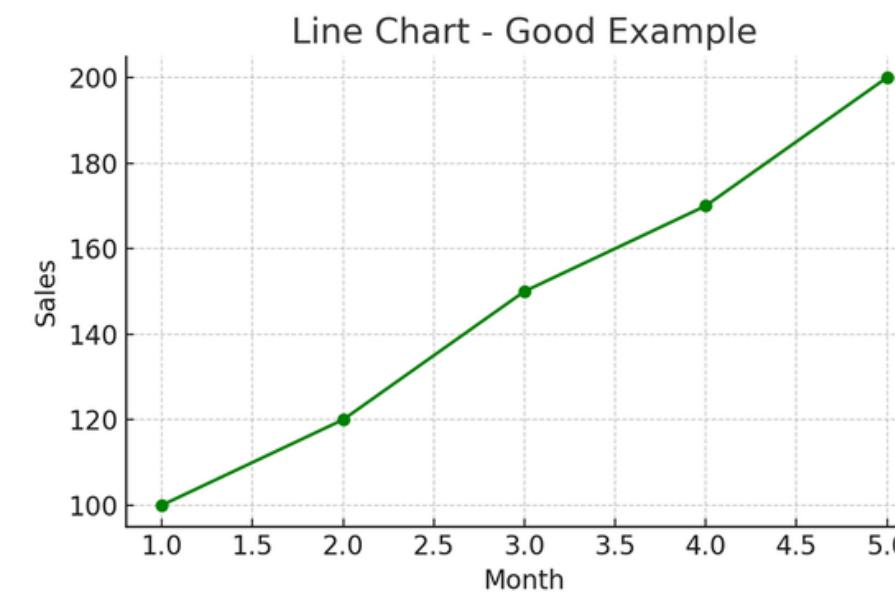
- แกน Y เริ่มที่ 0 → ไม่หลอกตา
- สีเรียบง่าย ใช้โทนเดียว (สีฟ้า) → โฟกัสที่ความสูงของแท่ง
- จัดเรียงหมวดหมู่ชัดเจน



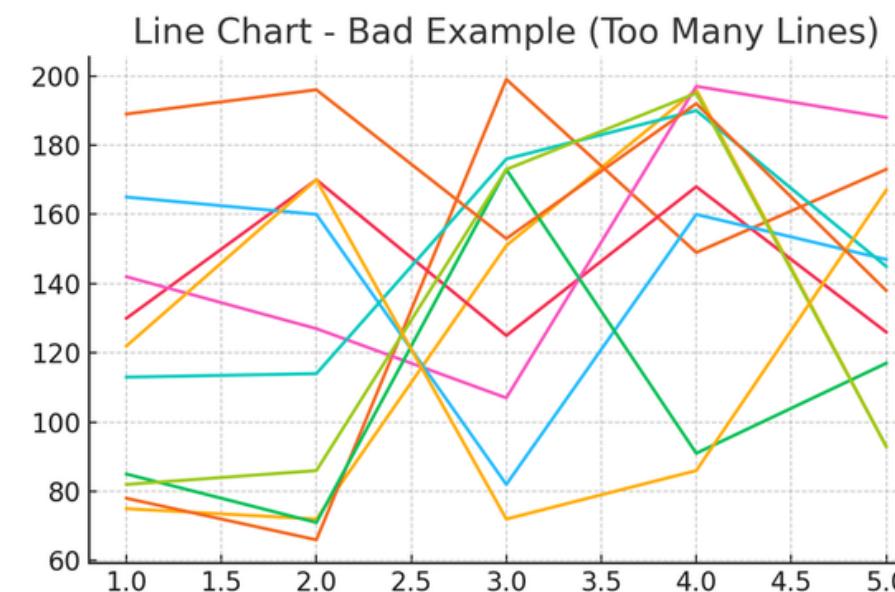
- แกน Y เริ่มที่ 25 → ทำให้ความต่างของแต่ละแท่งดูเว่อร์เกินจริง
- สีเยือกเกินไปแบบไม่มีความหมาย (Red, Green, Blue, ฯลฯ)
- ไม่มีการเรียงลำดับ ทำให้เปรียบเทียบยาก

# Line Chart (แผนภูมิเส้น)

แกน Y (แกนตั้ง): ปริมาณที่ต้องการแสดง, แกน X (แกนนอน): เวลา/รอบของการทำซ้ำ



- เส้นสีเด่นชัด
- มีจุด marker (วงกลม) ที่ช่วยเน้นจุดข้อมูลสำคัญ
- แกน X เป็นลำดับเวลา (Month)
- เห็นแนวโน้มชัดเจน (ยอดขายเพิ่มขึ้นเรื่อยๆ)

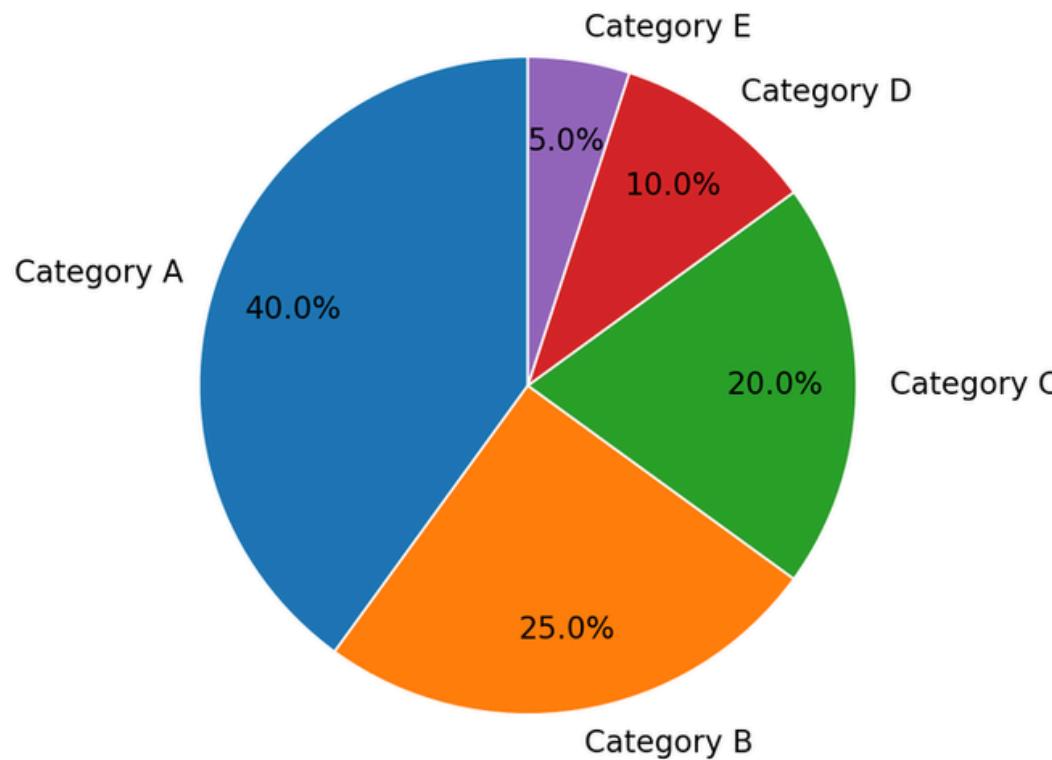


- วาดเส้น 10 เส้นซ้อนกันแบบ random color → เข้าใจยาก
- ไม่มี Legend / ไม่มี Label → ไม่รู้ว่าเส้นไหนคืออะไร
- Overplotting → อ่านแทรנד์ไม่ได้เลย

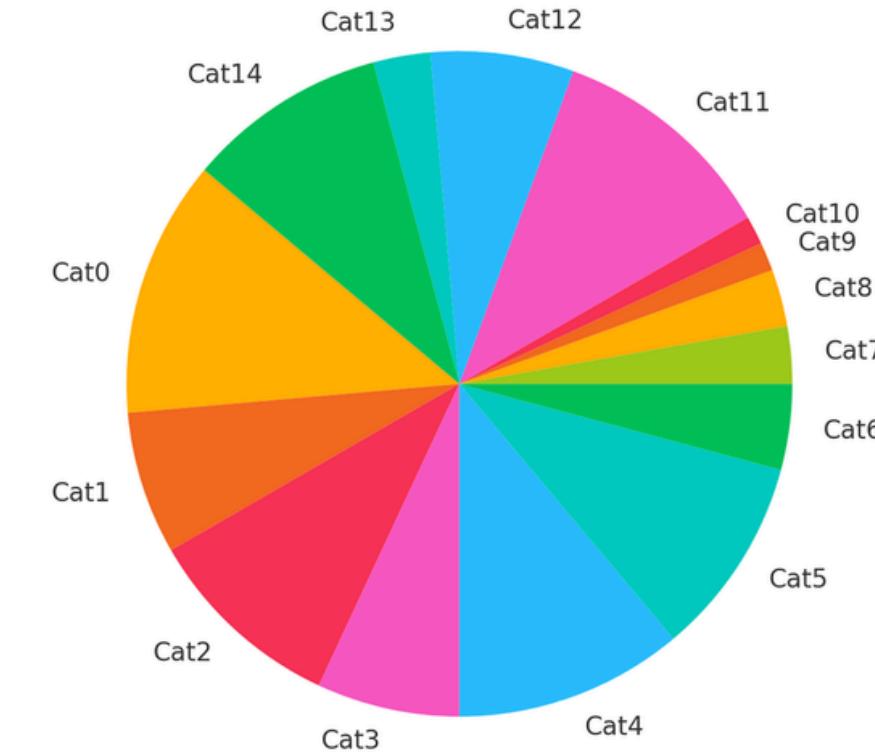
# Pie Chart (แผนภูมิวงกลม)

ลี แสดงแต่ละประเภท และขนาดของชีก (slice) แทนอัตราส่วน

Pie Chart - Good Example (Counterclockwise with Leader Lines)



Pie Chart - Bad Example (Too Many Slices)



- Slice ไม่เกิน 5-6 ชิ้น → อ่านง่าย
- สีต่างกันแบบ Contrast ชัดเจน
- มี % Label กำกับแต่ละ Slice
- เรียง slice ให้สุดเริ่มจากมุ่งซ้ายบน วนทวนเข็มนาฬิกา

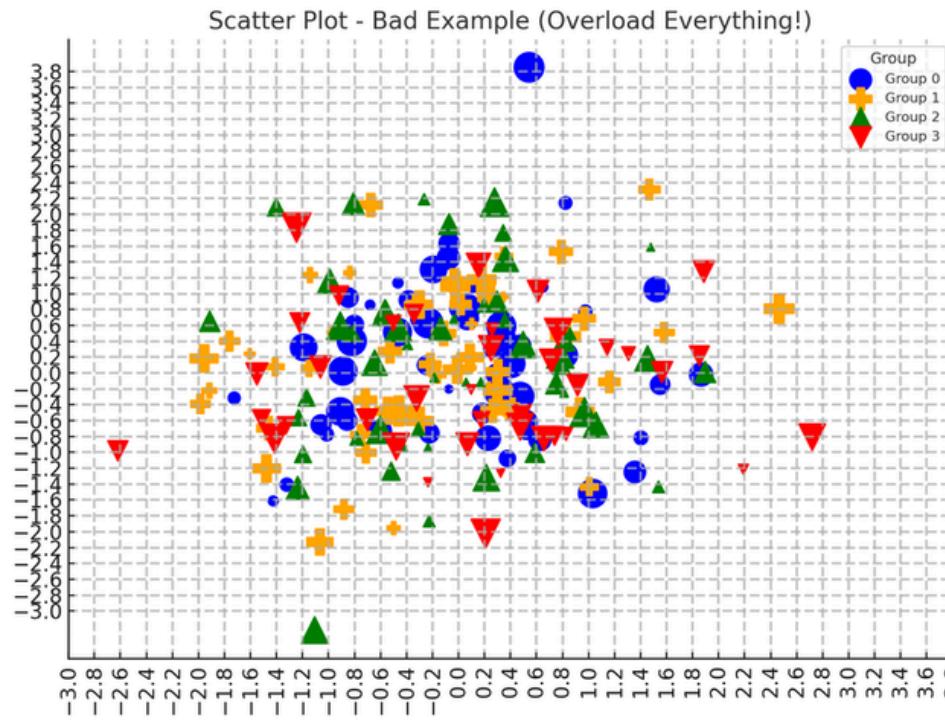
- จำนวน slice มากเกินไป
- ไม่มี % หรือ Label บอกว่าชิ้นไหนคืออะไร
- ขนาด slice ต่างกันน้อย ไม่แสดงให้เห็นเทรند์

# Scatter Plot (กราฟจุด)

แกน Y (แกนตั้ง): ตัวแปรตาม (ผล), แกน X (แกนนอน): ตัวแปรต้น (เหตุ)



- เห็นความสัมพันธ์ชัดเจน: จุดกระจายตัว มีแนวโน้มเชิงเส้นระหว่าง X กับ Y
- อ่านง่าย: ขนาดจุดพอดี สีไม่ล้ายตา Grid บางกำลังดี
- Label แกนครบ: มีชื่อแกน X และ Y ชัดเจน ช่วยให้เข้าใจความหมายของข้อมูล

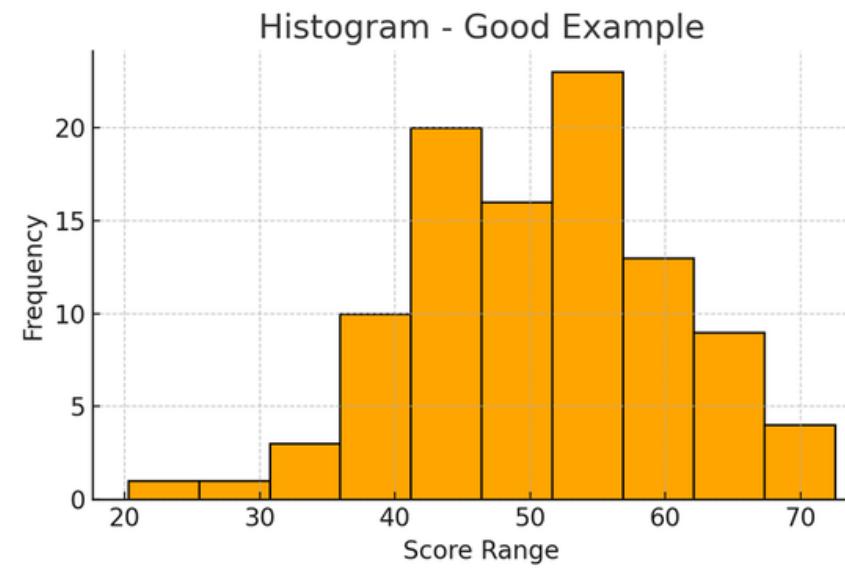


- สีและรูปทรงมากเกินไป: ทำให้อ่านยากและสับสน
- ขนาดจุดไม่สอดคล้อง: ไม่มีความหมายเชิงข้อมูลและربกวนสายตา
- Grid หนาแน่นเกินไป: ดึงความสนใจจากข้อมูลหลัก
- Tick บนแกนถี่เกิน: ทำให้แกนอ่านยากและรกสายตา
- Legend ไม่ชัดเจน: ขนาดและลักษณะไม่สัมพันธ์กับข้อมูลจริง

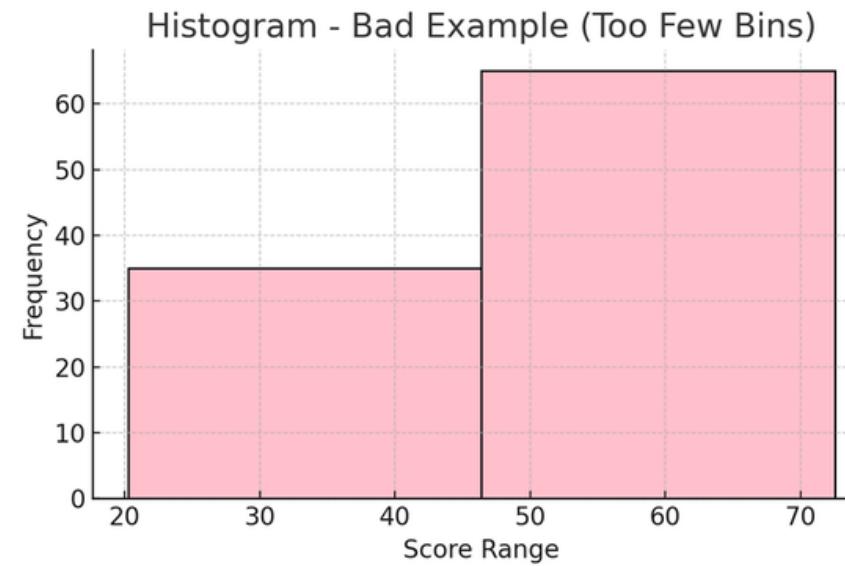
# Histogram (ฮิสโตแกรม)

แกน Y (แกนตั้ง): จำนวน (เห่านั้น), แกน X (แกนนอน): ตัวแปร (เชิงปริมาณเห่านั้น!)

- \* ตัวแปร หมายถึงลักษณะที่ต้องการอธิบาย เช่น ความสูง คะแนนสอบ อายุ
- \*\* ในทางสถิติ เราจะเรียกจำนวน ว่า ความถี่ (frequency)



- ใช้ Bin size ที่เหมาะสม (10 Bins) → เห็นการกระจายคะแนนชัดเจน
- สีสันเด่น + ขอบเลื่อนชัด (black edge)
- แกน X แสดงช่วงคะแนน, Y แสดงความถี่



- Bin แค่ 2 ช่อง → ข้อมูลรวมกันหมด อ่าน distribution ไม่ออกร

# Key Takeaway

---

1. เลือก Chart ให้เหมาะสมกับประเภทข้อมูลและวัตถุประสงค์
2. ออกแบบให้เรียบง่าย ชัดเจน และเน้นสาระสำคัญ
3. หลีกเลี่ยงความซับซ้อนเกินจำเป็น เช่น สีเยอรมัน รูปทรงหลากหลาย หรือ Grid ที่รบกวนสายตา
4. จัดการองค์ประกอบต่าง ๆ (Color, Size, Labels, Legend) ให้ช่วยเสริมการสื่อสาร ไม่ใช้สร้างความสับสน
5. เป้าหมายสำคัญ: ให้ผู้ชมเข้าใจข้อมูลและ Message ภายใต้รากฐานที่

*"Good visualization doesn't just show data. It tells a clear story."*

---

## Activity: Visualization Basics

# Q&A