

การจัดการข้อมูล

เปลี่ยนข้อมูลดิบให้เป็นขุมทรัพย์ทางธุรกิจ

ในยุคที่ “Data is a new Oil”
การรู้วิธีการกรองและสกัดมูลค่าออกมา
คือความได้เปรียบที่แท้จริง



เรากำลังอยู่ในมหาสมุทรของข้อมูล

ในยุคปัจจุบัน ปริมาณข้อมูลเพิ่มขึ้นมหาศาลทุกวัน ทำให้ข้อมูลกลายเป็นสิ่งที่มีมูลค่าสูงสุด ข้อมูลขนาดมหึมาี้เรียกว่า "Big Data" ซึ่งมีลักษณะสำคัญ 4 ประการ (The 4Vs) ที่ทำให้การจัดการมีความท้าทาย:

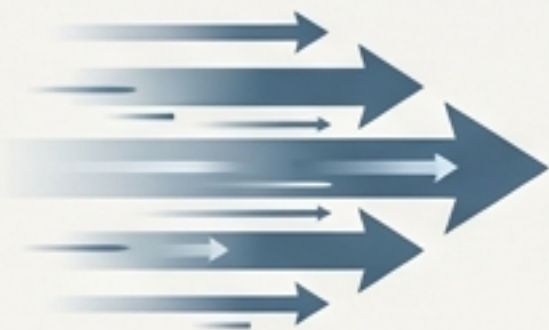


Volume (ปริมาณ): ข้อมูลมีขนาดใหญ่และปริมาณมหาศาล ทั้งแบบ offline และ online



Variety (ความหลากหลาย): รูปแบบข้อมูลหลากหลาย ทั้งที่มีโครงสร้างชัดเจนและไม่มีโครงสร้าง (unstructured)

Big Data



Velocity (ความเร็ว): ข้อมูลเปลี่ยนแปลงและส่งผ่านอย่างรวดเร็วต่อเนื่องแบบ real-time streaming



Veracity (ความไม่แน่นอน): ข้อมูลอาจมีความไม่ชัดเจน ไม่น่าเชื่อถือ หรือยังไม่ผ่านการตรวจสอบ (untrusted, uncleaned)

ผู้เชี่ยวชาญสองบทบาทในโลกแห่งข้อมูล

ศาสตร์ที่ใช้จัดการข้อมูลเรียกว่า "วิทยาการข้อมูล" (Data Science) ซึ่งประกอบด้วยผู้เชี่ยวชาญหลักสองตำแหน่ง:



นักวิศวกรรมข้อมูล (Data Engineer)

- **หน้าที่:** เปรียบเสมือนสถาปนิกและผู้สร้าง "โรงกลั่น"
- **ความรับผิดชอบหลัก:** เตรียมข้อมูลให้พร้อมสำหรับการนำไปวิเคราะห์ จัดการโครงสร้างพื้นฐานและกระบวนการนำส่งข้อมูล

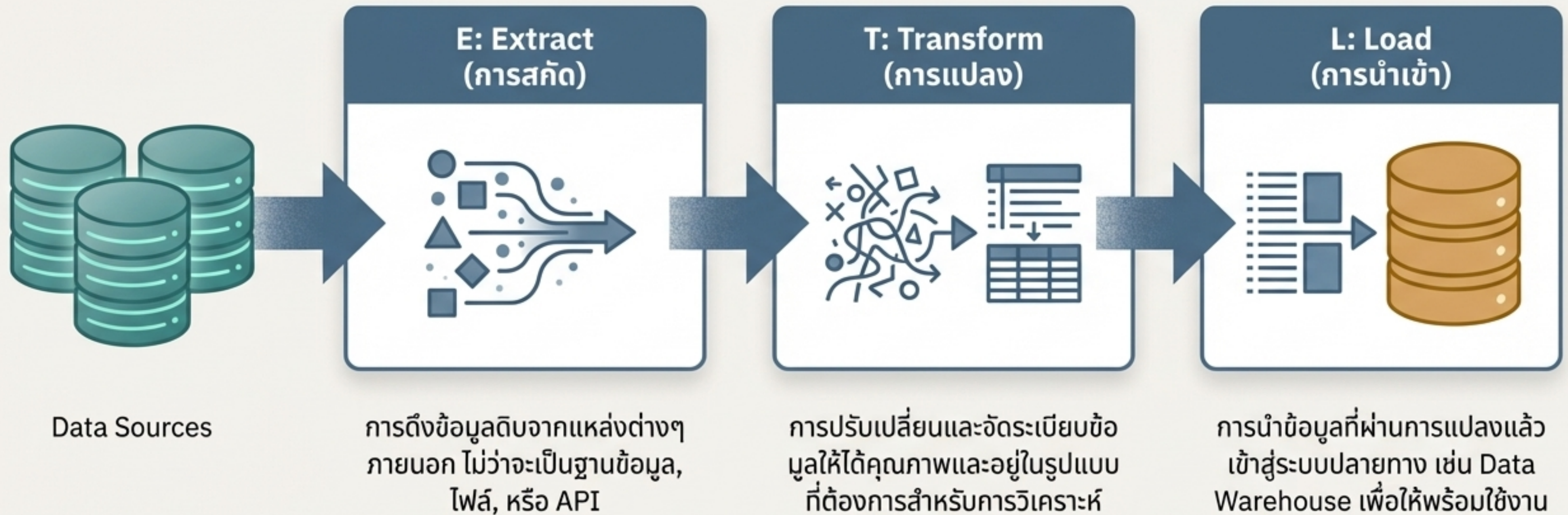


นักวิทยาศาสตร์ข้อมูล (Data Scientist)

- **หน้าที่:** เปรียบเสมือนนักวิเคราะห์และนักประดิษฐ์
- **ความรับผิดชอบหลัก:** นำข้อมูลที่ผ่านการเตรียมแล้วมาวิเคราะห์หารูปแบบ (patterns) สร้างโมเดลพยากรณ์ และสกัดข้อมูลเชิงลึก (insights) ที่มีคุณค่าออกมา

ETL: หัวใจของกระบวนการกลั่นกรองข้อมูล

ETL Model คือโมเดลที่อธิบายการทำงานของ Data Engineer ได้ดีที่สุด ประกอบด้วย 3 ขั้นตอนหลัก:



Data Warehouse: คลังเก็บข้อมูล

Data Warehouse คือ 'คลังข้อมูล' ขนาดใหญ่ที่ออกแบบมาเพื่อจัดเก็บข้อมูลสารสนเทศ (Information) ที่ผ่านกระบวนการแล้ว รวบรวมข้อมูลจากหลายแหล่งทั่วทั้งองค์กร ตั้งแต่อดีตจนถึงปัจจุบัน เพื่อใช้ในการวิเคราะห์เชิงลึก

ความแตกต่างจาก Database ทั่วไป

Database



เก็บข้อมูลการดำเนินงานปัจจุบัน แยกเป็นหมวดหมู่
(เช่น ข้อมูลสินค้า, ข้อมูลพนักงาน)

Data Warehouse



เก็บข้อมูลในอดีตและปัจจุบันทั้งหมดเพื่อการ
วิเคราะห์ภาพรวมและแนวโน้มทางธุรกิจ

ส่วนประกอบย่อย

Data Mart: คลังข้อมูลขนาดเล็กที่เน้นเฉพาะเรื่อง เหมือน 'ร้านค้าปลีก' ที่หยิบใช้ข้อมูลได้สะดวก



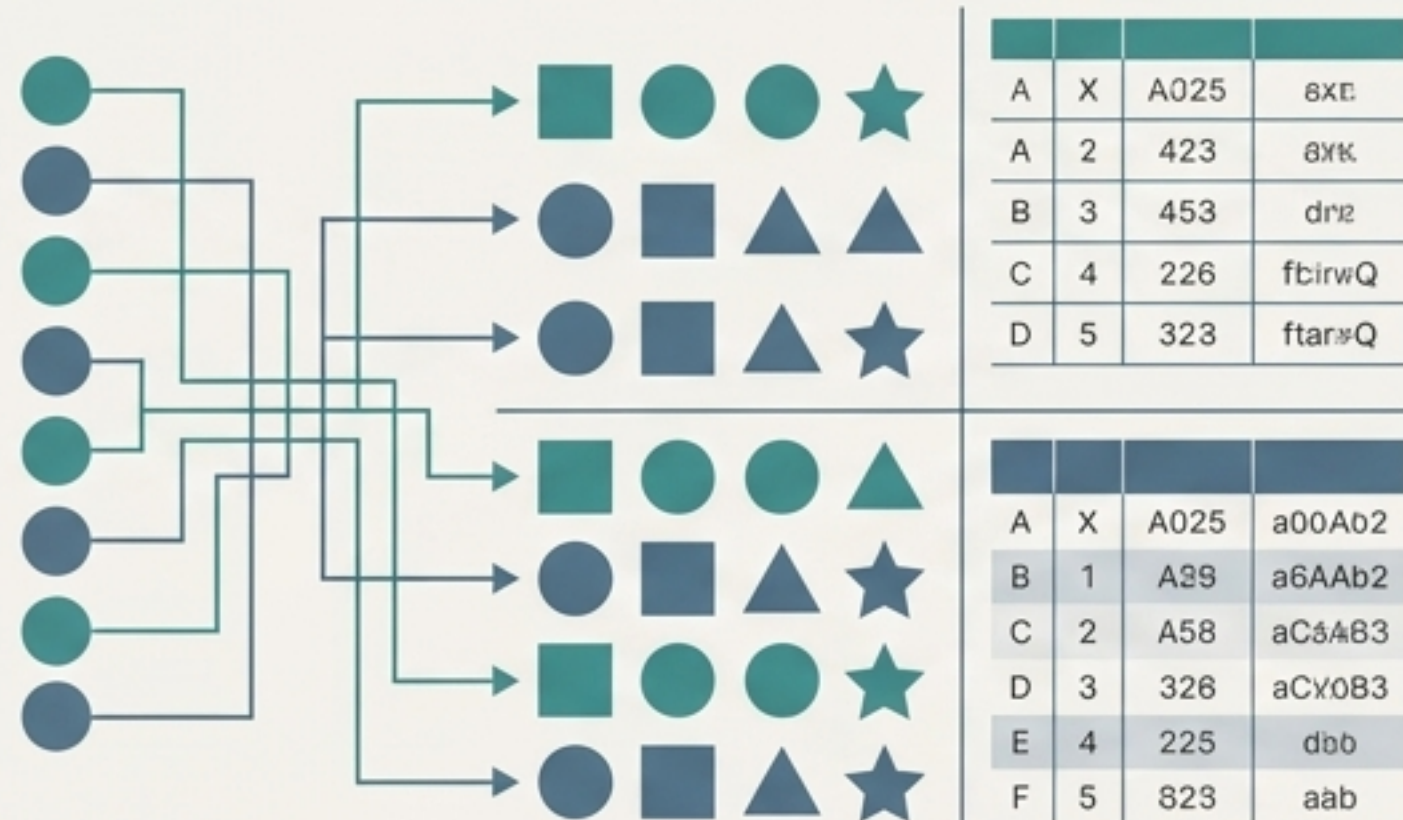
คุณภาพคือสิ่งสำคัญ: การรวบรวมและทำความสะอาดข้อมูล

เพื่อให้ได้ข้อมูลที่มีคุณภาพสูงสุดใน Data Warehouse ต้องผ่านสองกระบวนการสำคัญ:

ข้อมูลดิบ (Raw Data)



ข้อมูลสะอาด (Clean Data)



Data Gathering (การรวบรวมข้อมูล)

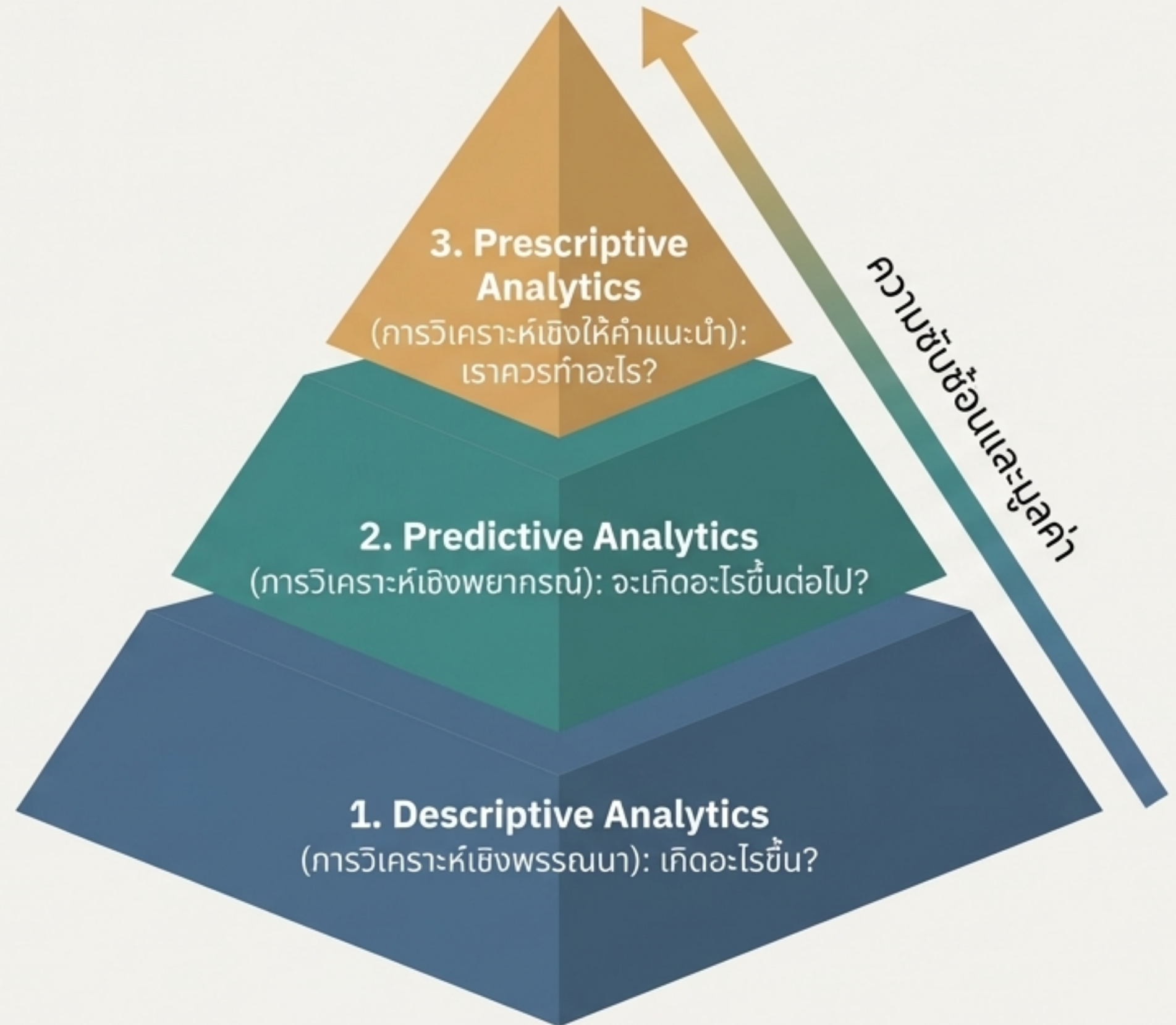
ทำหน้าที่เหมือน "นักสืบ" ที่ต้องรวบรวมข้อมูลทุกอย่างที่คาดว่าจะมีประโยชน์จากทุกแหล่งที่เป็นไปได้ แม้กระทั่งใบเสร็จหรือเอกสารกระดาษ

Data Cleansing (การทำความสะอาดข้อมูล)

กระบวนการ "คัดแยกขยะ" เพื่อลบข้อมูลที่ไม่ถูกต้อง ไม่สมบูรณ์ หรือไม่มีประโยชน์ทิ้งไป เพื่อให้เหลือแต่ข้อมูลที่สะอาดและน่าเชื่อถือสำหรับการวิเคราะห์

จากข้อมูลสู่ปัญญา: ระดับของการวิเคราะห์

การวิเคราะห์ Big Data คือการดึงข้อมูลสำคัญออกจากชุดข้อมูลขนาดใหญ่เพื่อหา pattern ที่มองไม่เห็นด้วยตาเปล่า นำไปสู่การตัดสินใจทางธุรกิจที่ดีขึ้น โดยแบ่งระดับการวิเคราะห์ได้ 3 รูปแบบหลัก:



ระดับที่ 1: Descriptive Analytics

เกิดอะไรขึ้นในอดีต? (What Happened in the Past?)

เป็นการวิเคราะห์พื้นฐานที่สุดเพื่อทำความเข้าใจเหตุการณ์ที่เกิดขึ้นแล้ว โดยตอบคำถามเช่น:

- **เกิดอะไรขึ้น?** (ผ่าน Standard Report)
- **จำนวนเท่าไร บ่อยแค่ไหน ที่ไหน?** (ผ่าน Ad hoc Report)
- **อะไรคือปัญหาที่แท้จริง?** (ผ่าน Query/Drill down)
- **ต้องมีการดำเนินการอะไร?** (ผ่าน Alerts)

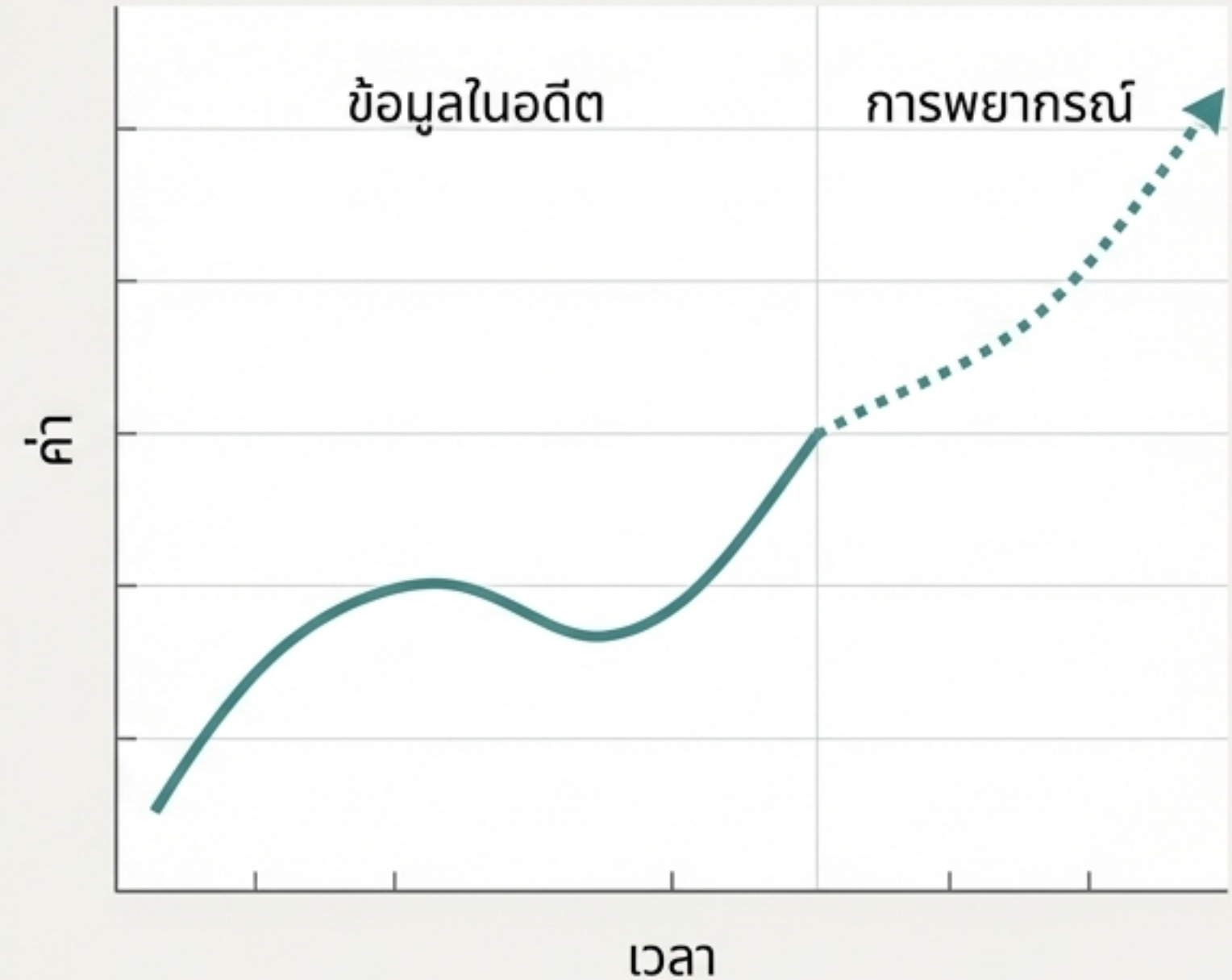


ระดับที่ 2: Predictive Analytics

จะเกิดอะไรขึ้นต่อไป? (What Will Happen Next?)

เป็นการวิเคราะห์ที่ซับซ้อนขึ้นเพื่อประเมินสิ่งที่จะเกิดขึ้นในอนาคต โดยอาศัยแนวโน้มจากข้อมูลในอดีต เพื่อตอบคำถามเช่น:

- **ทำไมสิ่งนี้จึงเกิดขึ้น?** (ผ่าน Statistical Analysis)
- **จะเกิดอะไรขึ้นถ้าเราลองทำแบบนี้?** (ผ่าน Randomized Testing)
- **จะเกิดอะไรขึ้นต่อไป?** (ผ่าน Predictive Modeling)
- **(ผ่าน ข้อมูลการกั้นเดี่ยอะไร?)** (ผ่าน Optimization)
- **สถานการณ์ที่ดีที่สุดคืออะไร?** (ผ่าน Optimization)



ระดับที่ 3: Prescriptive Analytics

เราควรตัดสินใจอย่างไร? (What Should We Do?)

เป็นการวิเคราะห์ขั้นสูงสุด ที่สังเคราะห์ข้อมูลทั้งหมดเพื่อวิเคราะห์แนวโน้ม และ "เสนอทางเลือก" ในการตัดสินใจที่เหมาะสมที่สุดโดยอัตโนมัติบนพื้นฐานของการคาดการณ์

ตัวอย่าง: ระบบ GPS ที่ไม่เพียงแค่บอกว่ารถติด (Descriptive) หรือคาดการณ์ว่าจะถึงที่หมายใน 30 นาที (Predictive) แต่ยังแนะนำเส้นทางที่ดีที่สุด ณ เวลานั้น (Prescriptive)



พลังของข้อมูล: 5 ประโยชน์ที่เปลี่ยนโลกธุรกิจ

ข้อมูลที่ผ่านการวิเคราะห์อย่างเป็นระบบสามารถสร้างประโยชน์ได้มหาศาล ดังนี้:



1. การอธิบายสิ่งที่เกิดขึ้น (Explaining What Happened):
ทำความเข้าใจสถานการณ์ปัจจุบันและเหตุการณ์ในอดีต



2. เปลี่ยนมุมมองต่อโลก (Changing Perspectives):
ค้นพบความจริงหรือความสัมพันธ์ที่ไม่เคยเห็นมาก่อน



3. ช่วยในการวางแผน (Aiding in Planning): คาดการณ์อนาคตเพื่อวางแผนกลยุทธ์ได้อย่างแม่นยำ



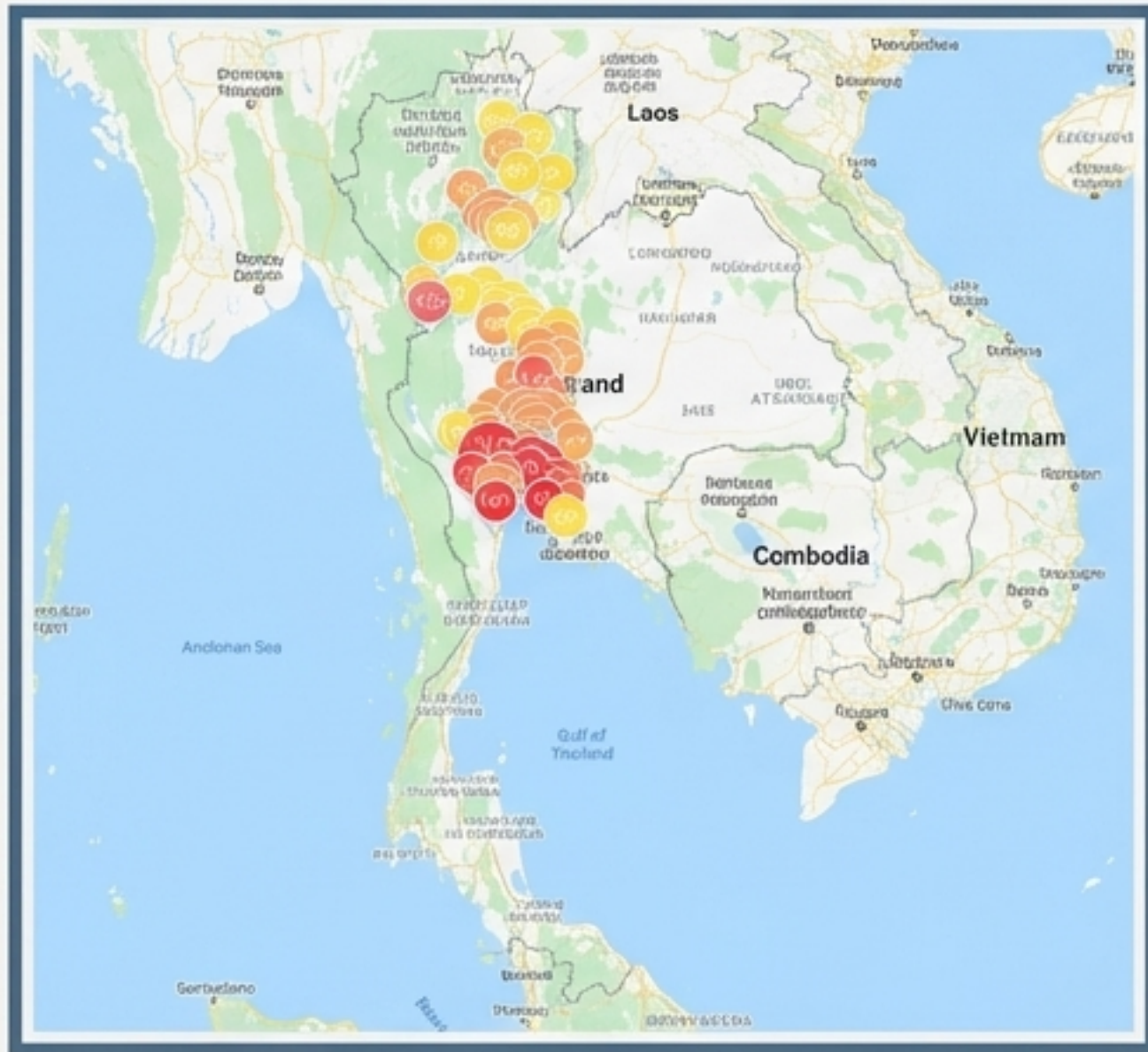
4. ช่วยในการตัดสินใจ (Supporting Decisions):
ให้ข้อมูลที่เป็นข้อเท็จจริงเพื่อประกอบการตัดสินใจที่สำคัญ



5. เพิ่มมูลค่าให้สินค้าและบริการ (Adding Value to Products & Services): สร้างสรรค์ผลิตภัณฑ์และบริการที่ตอบโจทย์ลูกค้าได้ดีขึ้น

ข้อมูลในโลกแห่งความจริง

อธิบายสิ่งที่เกิดขึ้น



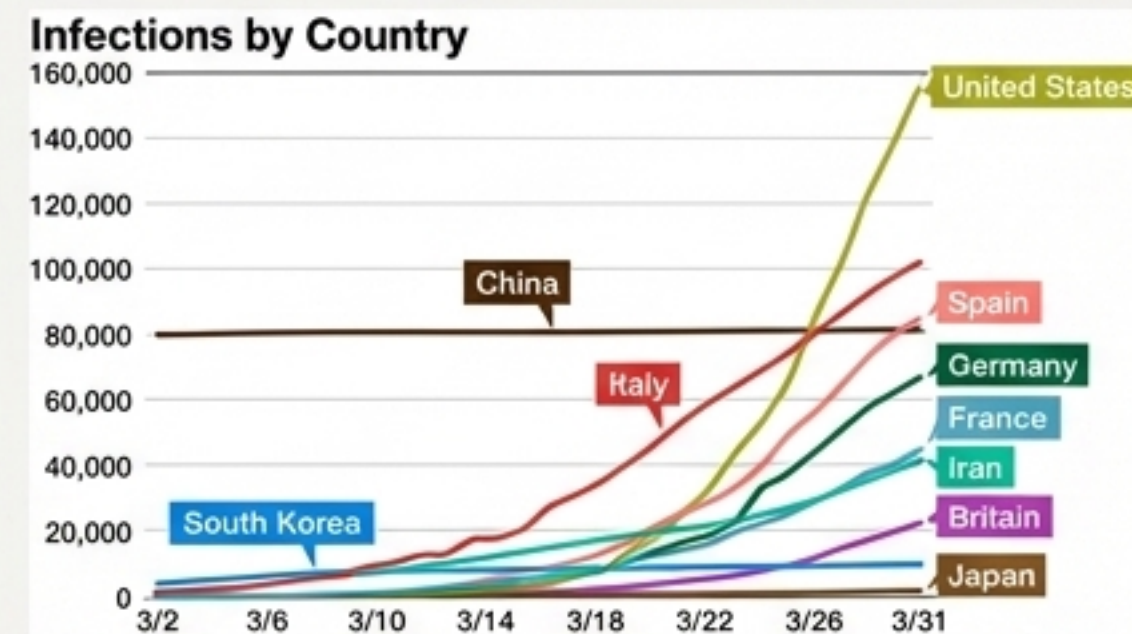
แสดงข้อมูลคุณภาพอากาศตามเวลาจริง
ทำให้เห็นว่าพื้นที่ใดมีความเสี่ยง

เปลี่ยนมุมมอง



แสดงความสัมพันธ์ระหว่างรายได้และอายุขัย
เผยให้เห็นพัฒนาการของโลกในมุมที่ไม่เคยเห็น

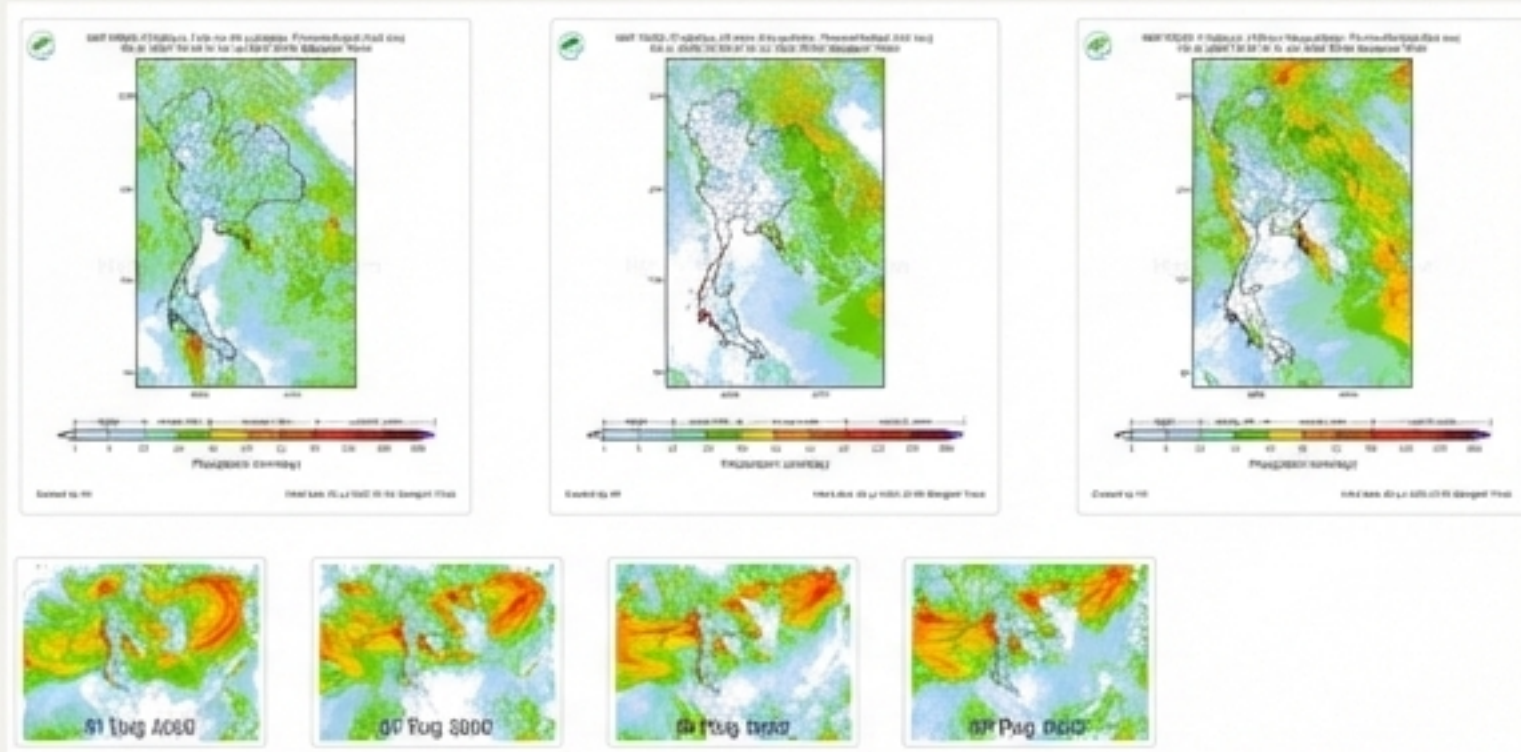
ช่วยในการตัดสินใจ



เปรียบเทียบอัตราการระบาดเพื่อใช้ประกอบการตัดสินใจด้านนโยบายสาธารณสุข

จากการวางแผนสู่การสร้างมูลค่า

ช่วยในการวางแผน



พยากรณ์ปริมาณฝนล่วงหน้า 7 วัน
ช่วยในการวางแผนจัดการน้ำและการเกษตร

เพิ่มมูลค่าให้สินค้าและบริการ



ใช้ข้อมูลภูมิศาสตร์และผลผลิตเพื่อสร้างแพลตฟอร์มที่
เชื่อมโยงเกษตรกรกับตลาด เพิ่มมูลค่าให้สินค้าเกษตร

กระบวนการวิทยาการข้อมูล: จากคำถามสู่คำตอบ



กรณีศึกษา: การวางแผนรับนักศึกษาด้วยข้อมูล



ทุกคำถามทางธุรกิจ สามารถเริ่มต้นหาคำตอบได้ด้วยข้อมูล