

KNOWLEDGE MANAGEMENT

หลักสูตรทักษะการจัดการข้อมูล และการวิเคราะห์ข้อมูล โดยใช้ Big Data Big Data Analytics Skill

ภายใต้แผนงานพัฒนาความสามารถทางเทคโนโลยีของบุคลากร
ภาคอุตสาหกรรม
โครงการสร้างกำลังคนและทักษะแห่งอนาคตในภูมิภาคเพื่อตอบโจทย์
การพัฒนานวัตกรรมของประเทศ ประจำปีงบประมาณ 2563



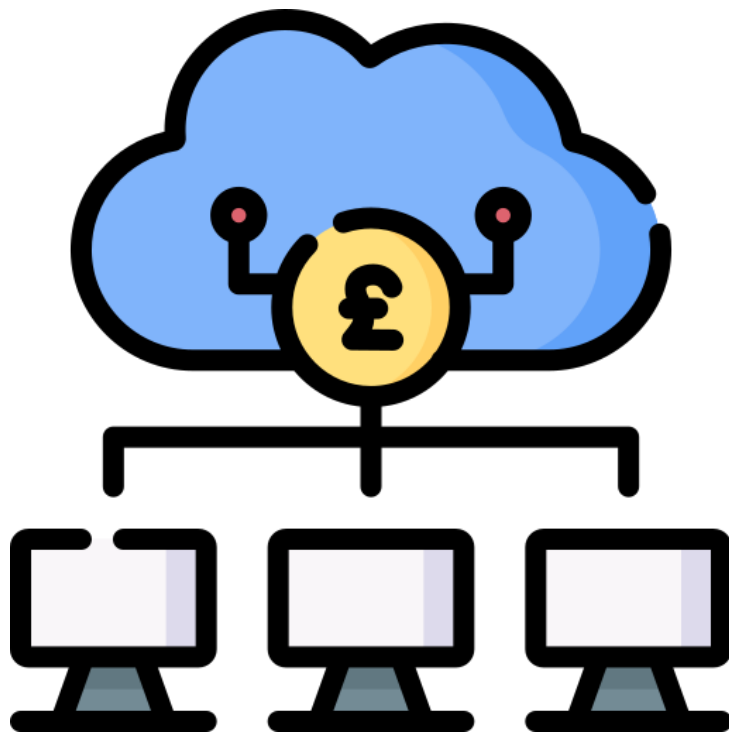
สารบัญ

	หน้า
บทที่ 1 : คำอธิบายโครงการ และหลักสูตร	1
1.1 แนะนำโครงการ	2
1.2 คำอธิบายหลักสูตร	8
บทที่ 2 : แบบทดสอบและประเมินผลก่อนเรียน	13
2.1 แบบทดสอบก่อนพัฒนากิจกรรม (Pre-Test)	14
2.2 แบบประเมินทักษะก่อนการพัฒนากิจกรรม (Pre-Embedded Skill)	21
บทที่ 3 : การพัฒนากิจกรรม	23
3.1 บทที่ 1 : Understanding Big Data	24
3.2 บทที่ 2 : Business Intelligence	106
3.3 บทที่ 3 : Big Data Planning	127
3.4 บทที่ 4 : Big Data Analytic	224
3.5 บทที่ 5 : Data Preparation	262
3.6 บทที่ 6 : Data Visualization	286
บทที่ 4 : แบบทดสอบและประเมินผลหลังเรียน	370
4.1 แบบทดสอบหลังพัฒนากิจกรรม (Post-Test)	371
4.2 แบบประเมินทักษะหลังการพัฒนากิจกรรม (Post-Embedded Skill)	378
4.3 เฉลยแบบทดสอบก่อน-หลังพัฒนากิจกรรม	380
บทที่ 5 : แผนงาน (Action Plan)	387
5.1 แบบฟอร์มแผนงาน (Action Plan)	388

บทที่ 1 : คำอธิบายโครงการ และหลักสูตร



1.1 แนะนำโครงการ





BRAIN POWER SKILL UP

ภายใต้
แผนงานการพัฒนาความสามารถทางเทคโนโลยีของบุคลากร
ภาคอุตสาหกรรม
โครงการสร้างกำลังคนและทักษะแห่งอนาคตในภูมิภาคเพื่อตอบโจทย์
การพัฒนานวัตกรรมของประเทศ

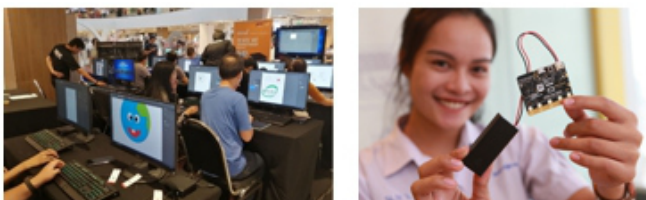


สร้างทักษะกำลังคนขั้นสูง
เพื่อรับมือความเปลี่ยนแปลง
วางแผนขับเคลื่อนธุรกิจสู่อนาคต

1 ภาพรวมแผนงาน “การพัฒนาความสามารถทางเทคโนโลยีของบุคลากรภาคอุตสาหกรรม (Brain Power Skill Up)”

รายละเอียด

หลักสูตรพัฒนาทักษะสำหรับภาคอุตสาหกรรม (upskill for future technology) จำนวน 20 หลักสูตร เพื่อรองรับการเปลี่ยนแปลง (transform) เทคโนโลยีของบริษัท เช่น AI, Data science, Big Data เป็นต้น



กลุ่มเป้าหมาย

บุคลากรในภาคอุตสาหกรรม

วัตถุประสงค์

1. เพื่อสร้างกำลังคนและทักษะแห่งอนาคตในภูมิภาค ให้ตอบโจทย์การพัฒนาอุตสาหกรรมของประเทศ
2. เพื่อสร้างระบบนิเวศเทคโนโลยีและนวัตกรรมในสถาบันการศึกษา

อุตสาหกรรมเป้าหมาย

อุตสาหกรรมเดิมที่มีศักยภาพในการต่อยอด



ยานยนต์สมัยใหม่



อิเล็กทรอนิกส์อัจฉริยะ



ท่องเที่ยวกลุ่มรายได้ดีและการท่องเที่ยวเชิงสุขภาพ



การเกษตรและเทคโนโลยีชีวภาพ



การแปรรูปอาหาร

อุตสาหกรรมอนาคต



หุ่นยนต์เพื่ออุตสาหกรรม



การบินและโลจิสติกส์



เชื้อเพลิงชีวภาพและเคมีชีวภาพ



ดิจิทัล



การแพทย์ครบวงจร

อุตสาหกรรมที่มีศักยภาพในภาคเหนือ



การแปรรูปอาหาร



เกษตร



กาแฟ



Herb & Cosmetics

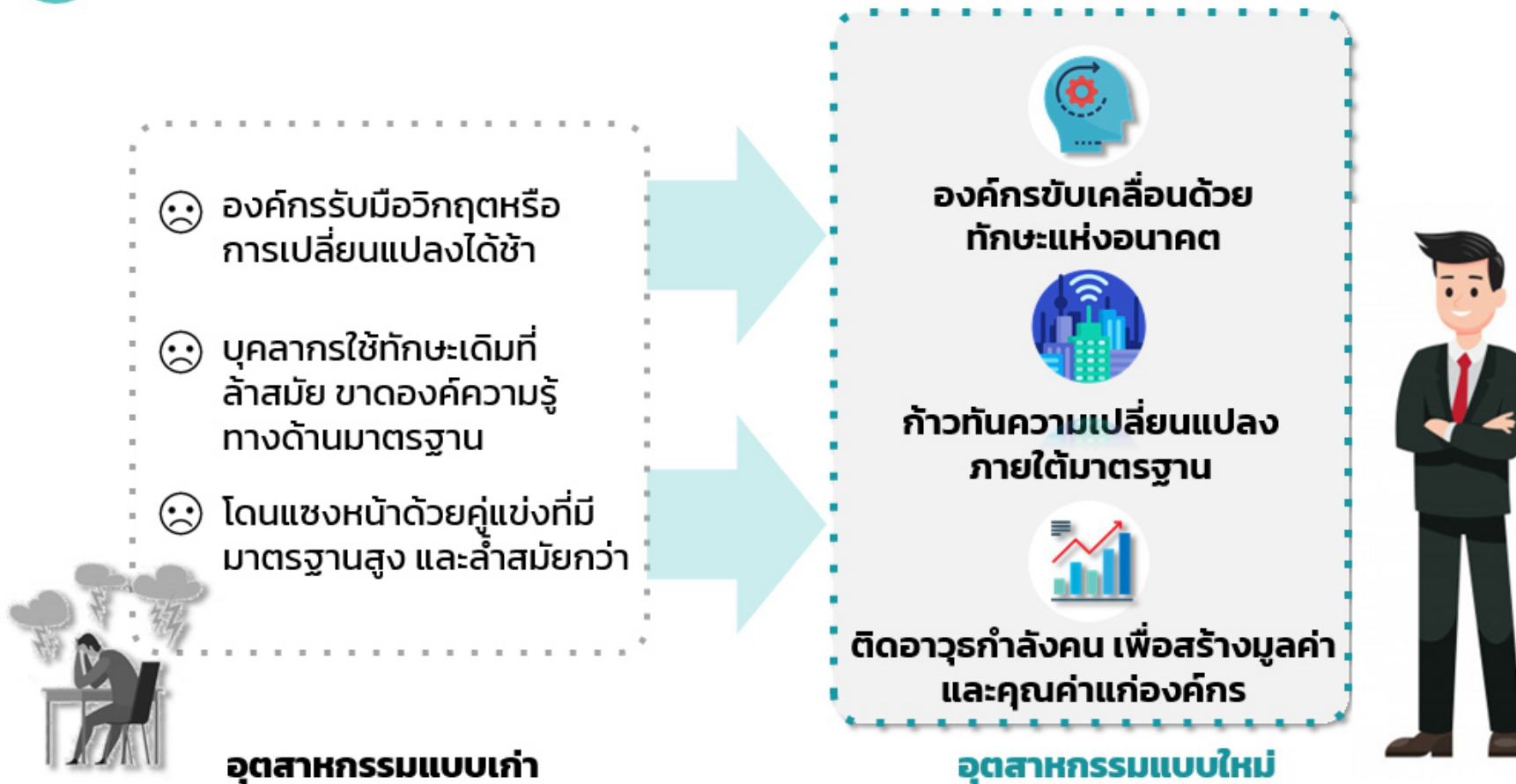


Fashion & Jewelry



Gift & Lifestyle

2 หลักการและความสำคัญ



3 กลไกการสร้างทักษะ (10 ขั้นตอน)



- 1 แนะนำโครงการ และ กิจกรรมสร้างเครือข่าย (Networking Workshop)
- 2 แบบทดสอบก่อนพัฒนาทักษะ (Pre-Test)
แบบประเมินทักษะก่อนเรียน (Pre-Embedded Skill Evaluation)
- 3 เรียนภาคทฤษฎี (Lecture)
- 4 การอบรมเชิงปฏิบัติการ (Case-Studies & Workshop)
- 5 เรียนภาคปฏิบัติ (Hands-On)
- 6 การเขียนแผนงาน **ทุกวันหลังเลิกเรียน** (Assignment – Action Plan)
- 7 การให้คำปรึกษาแผนงานโดยผู้เชี่ยวชาญ (Feedback) **ทุกวันหลังเลิกเรียน**
- 8 บันทึกความก้าวหน้าการพัฒนาทักษะ: **Learning Curve Record** **ทุกวันหลังเลิกเรียน**
- 9 แบบทดสอบหลังพัฒนาทักษะ (Post-Test)
แบบประเมินทักษะหลังเรียน (Post-Embedded Skill Evaluation)
- 10 ผู้เรียนนำเสนอแผนงาน Action Plan

4 ผลลัพธ์ทักษะ (Learning Output)



ผู้เรียนได้ทดสอบ
องค์ความรู้
ผ่าน **Pre-Test**
และ **Post-Test**



ผู้เรียนบันทึกและวางแผนการ
พัฒนาทักษะของตนเอง ผ่าน
Learning Curve Record

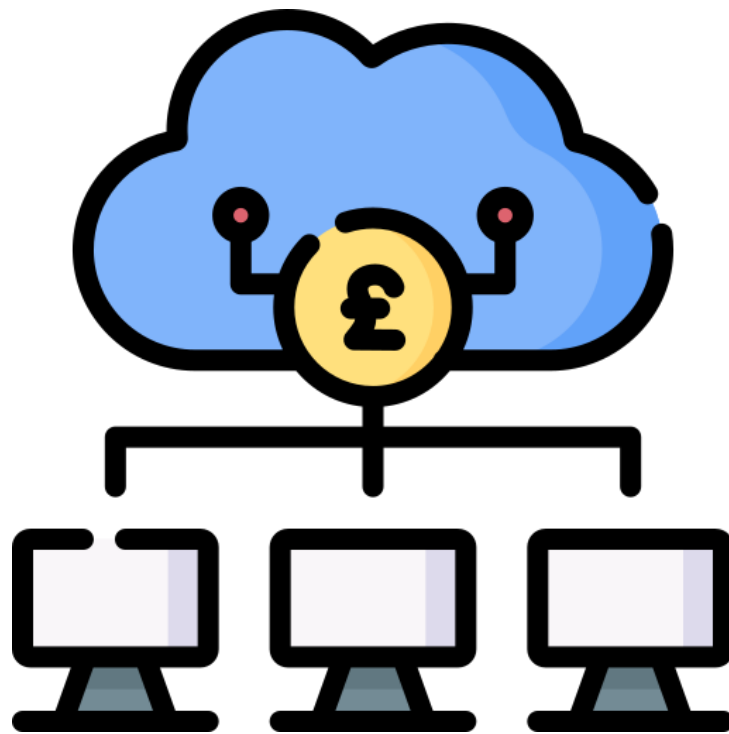


ผู้เรียนได้แผน
Action Plan
รายคน



ผู้เรียนได้รับการวัดผลสำเร็จ
การพัฒนาทักษะ ก่อนและหลัง
**(Pre-Post Embedded Skill
Evaluation)**
โดยวัดผลลัพธ์ 5 ทักษะ
เพื่อนำมาทำ Radar Chart

1.2 คำอธิบายหลักสูตร



ประเภท 1

การพัฒนาทักษะความสามารถทางเทคโนโลยีของบุคลากรขั้นพื้นฐาน

**หลักสูตรที่ 6 | ทักษะการจัดการข้อมูลและการวิเคราะห์ข้อมูล โดยใช้ Big Data
Big Data Analytics Skill**

คำอธิบาย : เสริมสร้างทักษะและสร้างความรู้ความเข้าใจในการจัดทำระบบ Big data ที่สามารถนำไปใช้ได้จริง โดยเรียนรู้การนำข้อมูลขนาดใหญ่ในปริมาณมหาศาลมาเข้าสู่ระบบวิเคราะห์โดยใช้เครื่องมือที่เหมาะสม นำไปสู่การใช้งานในภาคธุรกิจได้จริง




- วัตถุประสงค์ :**
1. เพื่อสร้างทักษะในการบริหารจัดการข้อมูลขนาดใหญ่ (Big data) สำหรับผู้ปฏิบัติงาน
 2. เพื่อเสริมสร้างทักษะและประสบการณ์ในการวิเคราะห์ข้อมูลขนาดใหญ่ด้วยเทคนิคต่าง ๆ
 3. เข้าใจหลักการของ Big data การวางกลยุทธ์ด้าน Big Data
 4. เรียนรู้การทำ Big Data Life Cycle และการบริหารโครงการ Big Data



- ผลลัพธ์ทักษะ :**
1. ทักษะการกำหนดปัญหาในธุรกิจด้วย Big Data
 2. ทักษะการรวบรวมข้อมูลเพื่อประกอบการใช้เทคนิค Big Data
 3. ทักษะการวิเคราะห์ข้อมูลโดยใช้ Big Data เพื่อสร้างประสิทธิภาพให้กับธุรกิจ
 4. ทักษะการเลือกเครื่องมือและการนำเสนอให้สอดคล้องกับธุรกิจ
 5. ทักษะการประยุกต์ใช้ Big Data

อุตสาหกรรมเป้าหมาย : ทุกกลุ่มอุตสาหกรรม

กลุ่มเป้าหมาย : ผู้บริหาร หัวหน้างาน พนักงาน

วิทยากร :

ลำดับ	ชื่อ-นามสกุล	หน่วยงาน	ความเชี่ยวชาญ	รูปภาพ
1	พศ. ดร. สิริกร สันติโรจนกุล	อาจารย์ประจำ หลักสูตรการ จัดการสมัยใหม่ และเทคโนโลยี สารสนเทศ วิทยาลัยศิลปะ สื่อ และเทคโนโลยี มหาวิทยาลัย เชียงใหม่	-เหมืองข้อมูล (Data mining) - การวิเคราะห์ระบบ (System Analysis) - พาณิชย์ อิเล็กทรอนิกส์ (E-Commerce)	
2	ดร.เอกสิทธิ์ เทียมแก้ว	อาจารย์ประจำ ภาควิชาวิทยาการ คอมพิวเตอร์และ เทคโนโลยี สารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัย นเรศวร จังหวัด พิษณุโลก	- Big Data and Data Analytics - Big Data Adoption - Big Data Technologies	
3	รศ.ดร.จรรย์ทร ไชยจารุณิช	หัวหน้าศูนย์วิจัย วิทยาการข้อมูล ภาควิชาวิทยาการ คอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัย เชียงใหม่	- Algorithms - Informatics - Data Mining - Bioinformatics - Machine Learning	

ลำดับ	ชื่อ-นามสกุล	หน่วยงาน	ความเชี่ยวชาญ	รูปภาพ
4	อ.ภูษิต ก้อนสุรินทร์	อาจารย์ประจำ สาขาวิชาวิศวกรรม ซอฟต์แวร์ คณะ วิทยาศาสตร์ มหาวิทยาลัยพายัพ	- การประยุกต์ใช้ งาน Data Science, Data Visualization, Artificial Intelligence	
5	คุณกัศสปะ ศักดิ์ศรี	Technical Specialist of Aware Corporation Limited.	- พัฒนา Software - Geographic Information Systems - Data Analysis	

เนื้อหาที่เรียน :

ลำดับที่	เนื้อหาที่เรียน
1	<p>Understanding Big Data</p> <ul style="list-style-type: none"> • Big Data Concepts and Terminology • Characteristics of Big Data • Types of Data • Big Data Ecosystem • Big Data for Organization Data Disruption • Storing Big Data • Extract Transform Load (ETL) <p>Business Intelligence</p> <ul style="list-style-type: none"> • Traditional BI and Big Data BI
2	Big Data and Data Analytics

ลำดับที่	เนื้อหาที่เรียน
	Big Data Adoption Big Data Technologies Big Data Analytics Lifecycle Business Case Evaluation Data Identification Data Acquisition & Filtering Data Extraction Data Validation & Cleansing Data Aggregation & Representation
3	Big Data Analysis Techniques
4	Data Preparation (Workshop) <ul style="list-style-type: none"> • Setting ZOHO • Import Data from different source • Data Cleansing Technic • Data Correction Technic
5	Data Visualization (Workshop) <ul style="list-style-type: none"> • Create Chart

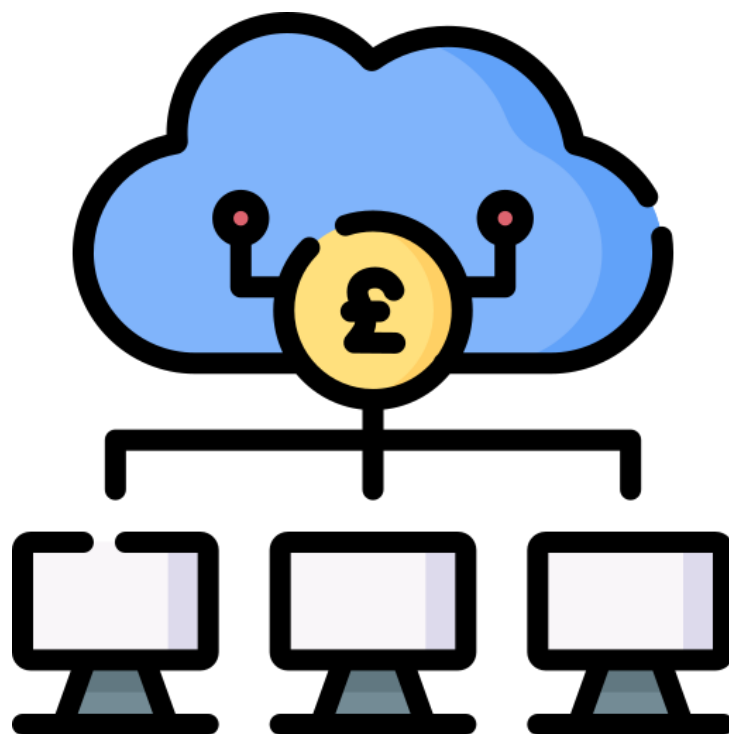
วิธีการเรียน :

1. แนะนำโครงการ Brain Power Skill Up
2. แบบทดสอบก่อนพัฒนาดัชนี (Pre-Test)/แบบประเมินทักษะก่อนเรียน (Pre-Embedded Skill Evaluation)
3. เรียนภาคทฤษฎี (Lecture)
4. การเขียนแผนงาน ทุกวันหลังเลิกเรียน (Assignment – Action Plan)
5. บันทึกความก้าวหน้าการพัฒนาดัชนี Learning Curve Record ทุกวันหลังเลิกเรียน
6. แบบทดสอบหลังพัฒนาดัชนี (Post-Test)/แบบประเมินทักษะหลังเรียน (Post-Embedded Skill Evaluation)

บทที่ 2 : แบบทดสอบและ ประเมินผลก่อนเรียน



2.1 แบบทดสอบก่อน พัฒนาทักษะ (Pre-Test)



ส่วนที่ 1 ลงทะเบียน

1. กรุณากรอกชื่อ-นามสกุล.....
2. สถานประกอบการ.....
3. Email.....
4. เบอร์โทร.....

ส่วนที่ 2 แบบทดสอบก่อนพัฒนาทักษะ: (Pre-Test)

- คำชี้แจง 1. แบบทดสอบฉบับนี้เป็นแบบปรนัยชนิดเลือกตอบ 4 ตัวเลือก จำนวน 16 ข้อ 16 คะแนน
2. จงเลือกคำตอบที่ถูกต้องที่สุดเพียงข้อเดียว

1. ข้อมูลในข้อใดต่อไปนี้เป็นข้อมูลเชิงคุณภาพ (Qualitative Data)

- เพศ
- สัญชาติ
- ยอดขาย
- ความพึงพอใจ

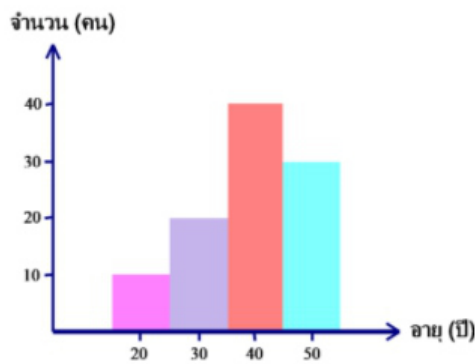
2. ข้อมูลในข้อใดต่อไปนี้เป็นข้อมูลเชิงปริมาณ (Quantitative Data)

- อุณหภูมิ
- ระดับการศึกษา
- ราคาสินค้า
- น้ำหนัก

3. ข้อใดไม่ใช่องค์ประกอบสำคัญของข้อมูลขนาดใหญ่ (Big Data)

- Volume (ข้อมูลที่มีปริมาณมาก)
- Variety (ข้อมูลที่มีหลากหลายรูปแบบ)
- Velocity (ข้อมูลที่มีการเปลี่ยนแปลงอย่างรวดเร็ว)
- Visitor (ข้อมูลจากผู้เข้าเยี่ยมชม)

4. จากกราฟแท่งต่อไปนี้ แสดงการกระจายของอายุประชาชนในหมู่บ้านแห่งหนึ่ง ประชาชนในหมู่บ้านแห่งนี้มีทั้งหมดกี่คน



- 100 คน
- 80 คน
- 70 คน
- 60 คน

5. จากการสำรวจอายุของเด็กจำนวน 100 คนพบว่า เด็กอายุ 4 ขวบ มีจำนวน 20 คน เด็กอายุ 5 ขวบ มีจำนวน 40 คน เด็กอายุ 6 ขวบ มีจำนวน 40 คน จงหาค่าเฉลี่ยของอายุของเด็กกลุ่มนี้

- 5.2 ขวบ
- 5.3 ขวบ
- 5.4 ขวบ
- 5.5 ขวบ

6. ชุดข้อมูลหรือกลุ่มข้อมูลใดที่เกี่ยวข้องกัน

- Datasets
- Data analytics
- Data Visualization
- Data Cleansing

7. ข้อใดเป็นการวิเคราะห์เชิงวินิจฉัยภาษาเหตุ โดยเปรียบเทียบข้อมูลในอดีตเทียบกับข้อมูลอื่นเพื่อหาคำตอบ

- Descriptive Analytics
- Predictive Analytics
- Diagnostic Analytics
- Prescriptive Analytics

8. ข้อใดคือคุณสมบัติของ Big data

- ปริมาณข้อมูลขนาดใหญ่
- ความหลากหลายของกลุ่มข้อมูล
- คุณภาพหรือความน่าเชื่อถือของข้อมูล
- ถูกทุกข้อ

9. แหล่งข้อมูลของBig data ที่ข้อมูลสร้างมาจากโซเชียลมีเดีย Blog posts อีเมลล์ ภาพ ข้อความในออนไลน์ เป็นแหล่งข้อมูลจากข้อใด

- Machine Generated
- Software Generated
- Human-generated data
- Social Network Generated

10. ข้อมูลชนิดวิดีโอคลิป ที่เผยแพร่ หรือแชร์ทาง YouTube.com นั้นเป็นประเภทข้อมูลในข้อใด

- ข้อมูลที่ไม่มีโครงสร้าง
- ข้อมูลกึ่งมีโครงสร้าง
- ข้อมูลเชิงโครงสร้าง
- ข้อมูลโซเชี่ยลเน็ตเวิร์ค

11. กระบวนการ Big data ในข้อใดจัดเรียงตามลำดับได้ถูกต้อง

- การจัดเก็บรวบรวมข้อมูล, การประมวลผล, สร้างรายงาน
- การจัดเก็บรวบรวมข้อมูล, การประมวลผล, การเผยแพร่, สร้างรายงาน
- การประมวลผล, การจัดเก็บรวบรวมข้อมูล, การเผยแพร่
- การประมวลผล, การจัดเก็บรวบรวมข้อมูล, สร้างรายงาน, การเผยแพร่

12. ข้อใดต่อไปนี้ ไม่ใช่การทำ Data Preparation

- Data Visualization
- Data Cleansing
- Data Collection
- Find & Replace Data

13. ข้อใดไม่ใช่สาเหตุของการทำ Data Cleansing

- เพราะแหล่งข้อมูลมาจากหลายที่
- เพราะข้อมูลนำเข้ามามีจำนวนมาก
- เพราะข้อมูลที่นำเข้ามามีไม่ standard
- เพราะผู้นำเข้าข้อมูลมาหลากหลาย

14. สำนักข่าวต้องการรวบรวมความคิดเห็นสาธารณะในเหตุการณ์หนึ่ง โดยที่มีการพิจารณาถึงกลยุทธ์ที่แตกต่างกัน 2 กลยุทธ์

ในการรวบรวมข้อมูลออนไลน์ โดย 2 กลยุทธ์ดังกล่าว มีดังนี้

กลยุทธ์ที่ 1:

- ใช้ฐานข้อมูลในการเก็บข้อมูลแบบสอบถามทั้งหมด
- เก็บข้อมูลเป็นข้อความ และ ตัวเลข
- ติดตามข้อมูลเพิ่มเติมเกี่ยวกับผู้รับการสำรวจที่จะไม่ปรากฏต่อสาธารณะ

กลยุทธ์ที่ 2:

- ใช้ spreadsheet เดียวเพื่อจัดเก็บคำตอบแบบสำรวจทั้งหมด
- เก็บข้อมูลทั้งหมดเป็นตัวเลข
- ไม่มีการติดตามข้อมูลใด ๆ นอกเหนือจากแบบสอบถาม

ข้อความใดต่อไปนี้เป็นารเปรียบเทียบที่ถูกต้องที่สุดของกลยุทธ์เหล่านี้

- กลยุทธ์ที่หนึ่งจะช่วยให้จัดเรียงและกรองข้อมูลได้ง่ายขึ้นในขณะที่กลยุทธ์ที่สองจะช่วยให้สร้างกราฟข้อมูลได้ง่ายขึ้น
- กลยุทธ์ที่หนึ่งจะทำให้เกิดปัญหาเนื่องจากประเภทข้อมูลที่หลากหลายในขณะที่กลยุทธ์ที่สองจะทำให้ง่ายต่อการค้นหาข้อมูลเฉพาะ
- กลยุทธ์ที่หนึ่งจะช่วยให้หน่วยงานสามารถสรุปเกี่ยวกับความคิดเห็นของประชาชนได้มากขึ้นเนื่องจากติดตามข้อมูล metadata เพิ่มเติมในขณะที่กลยุทธ์ที่สองจะทำให้ยากที่จะมองแนวโน้มและเข้าถึงข้อมูลบางส่วน
- กลยุทธ์ที่หนึ่งลดเวลาสำหรับการเตรียมและจัดการข้อมูลให้น้อยลงในขณะที่กลยุทธ์ที่สองต้องการการคำนวณเพิ่มเติมจำนวนมากเพื่อใช้ข้อมูล

15. เจ้าของเว็บไซต์ต้องการวางแผนการแยกแหล่งที่มาของการเข้าชมเว็บไซต์ โดยมี 4 หมวดหมู่ ได้แก่ การเข้าชมโดยตรง (Direct Traffic), เครื่องมือค้นหา (Search Engines), ไซต์อ้างอิง (Referring Sites) และ อื่น ๆ (Other) แผนภูมิ (Chart) ประเภทใด

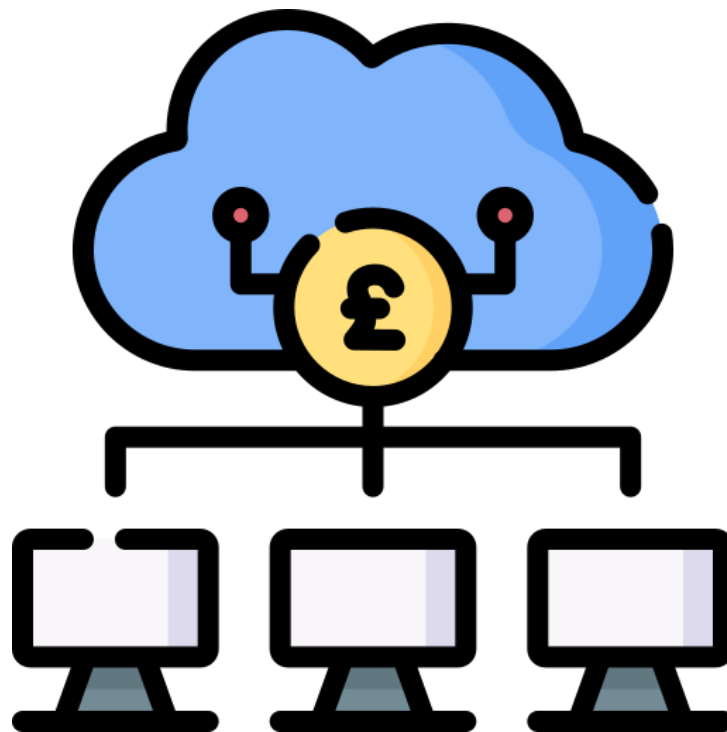
ที่จะเป็นทางเลือกที่ดีที่สุด

- Line Chart
- Dial Chart
- Pie Chart
- Bubble Chart

16. นักวิเคราะห์ข้อมูล สามารถทำนาย (Forecasting) ผลได้จากการอ่าน แผนภูมิ (Chart) ยกเว้น ข้อมูลจาก Chart ประเภทใด

- Line Chart
- Bar Chart
- Pie Chart
- Area Chart

2.2 แบบประเมินทักษะก่อนการพัฒนา ทักษะ (Pre-Embedded Skill)



ส่วนที่ 1 สำหรับ ผู้เรียน

ข้อมูลทั่วไป
 ชื่อ-นามสกุล

ชื่อสถานประกอบการ

ส่วนที่ 2 สำหรับ เจ้าของกิจการ หรือ หัวหน้างาน

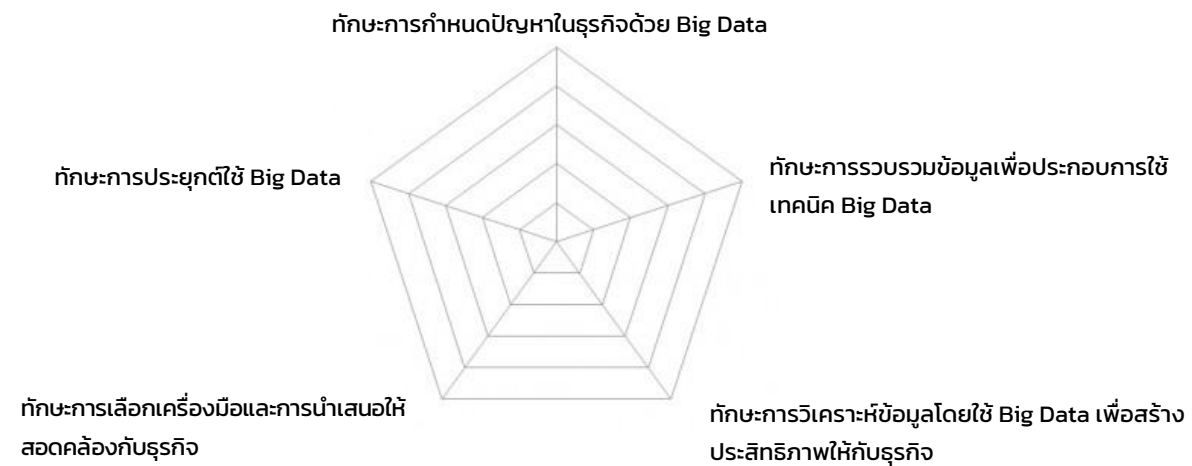
2.1 การประเมินผู้เรียน
ความหมายระดับคะแนน

- 0 = Beginner ไม่มีความรู้ ไม่มีทักษะ
- 1 = Learner มีความเข้าใจในทฤษฎีเบื้องต้น
- 2 = Practitioner มีความเข้าใจในทฤษฎีอย่างเต็มที่ มีความรู้ด้านปฏิบัติเล็กน้อย สามารถตอบคำถามหรือแก้ไขปัญหาคือไม่ซับซ้อนได้
- 3 = Experienced มีความเข้าใจในทฤษฎีและปฏิบัติอย่างเต็มที่ สามารถประยุกต์ใช้ความรู้เพื่อแก้ไขปัญหาคือซับซ้อนปานกลางได้
- 4 = Embedded เกิดทักษะติดตัว สามารถเชื่อมโยงความรู้ในการแก้ไขปัญหาที่ซับซ้อนมากได้ และสามารถกำหนดแผนเพื่อปรับปรุงและพัฒนาประสิทธิภาพการทำงานในองค์กรได้และนำไปสู่การต่อยอดเพื่อลงมือทำจริง
- 5 = Broaden เกิดทักษะอย่างทอ่งแท้ในระดับผู้เชี่ยวชาญ และสามารถถ่ายทอดทักษะให้แก่ผู้อื่นได้

กรุณา (✓) ในช่องระดับคะแนน

ผลลัพธ์ทักษะ	ระดับคะแนน					
	0	1	2	3	4	5
1. ทักษะการกำหนดปัญหาในธุรกิจด้วย Big Data						
2. ทักษะการรวบรวมข้อมูลเพื่อประกอบการใช้เทคนิค Big Data						
3. ทักษะการวิเคราะห์ข้อมูลโดยใช้ Big Data เพื่อสร้างประสิทธิภาพให้กับธุรกิจ						
4. ทักษะการเลือกเครื่องมือและการนำเสนอให้สอดคล้องกับธุรกิจ						
5. ทักษะการประยุกต์ใช้ Big Data						

การวิเคราะห์ผลการพัฒนาทักษะด้วยกราฟเรดาร์ (Radar Chart)



บทที่ 3 : การพัฒนาทักษะ



3.1 บทที่ 1: Understanding Big Data



Understanding Big Data

8 - 9 กันยายน 2563

ผศ.ดร.สิริกร สันติโรจนกุล

อาจารย์ภูษิต ก้อนสุรินทร์

อาจารย์ ดร.เอกสิทธิ์ เทียมแก้ว

Digital Marketing Solution Providers Center

เพจ @digitalSPC

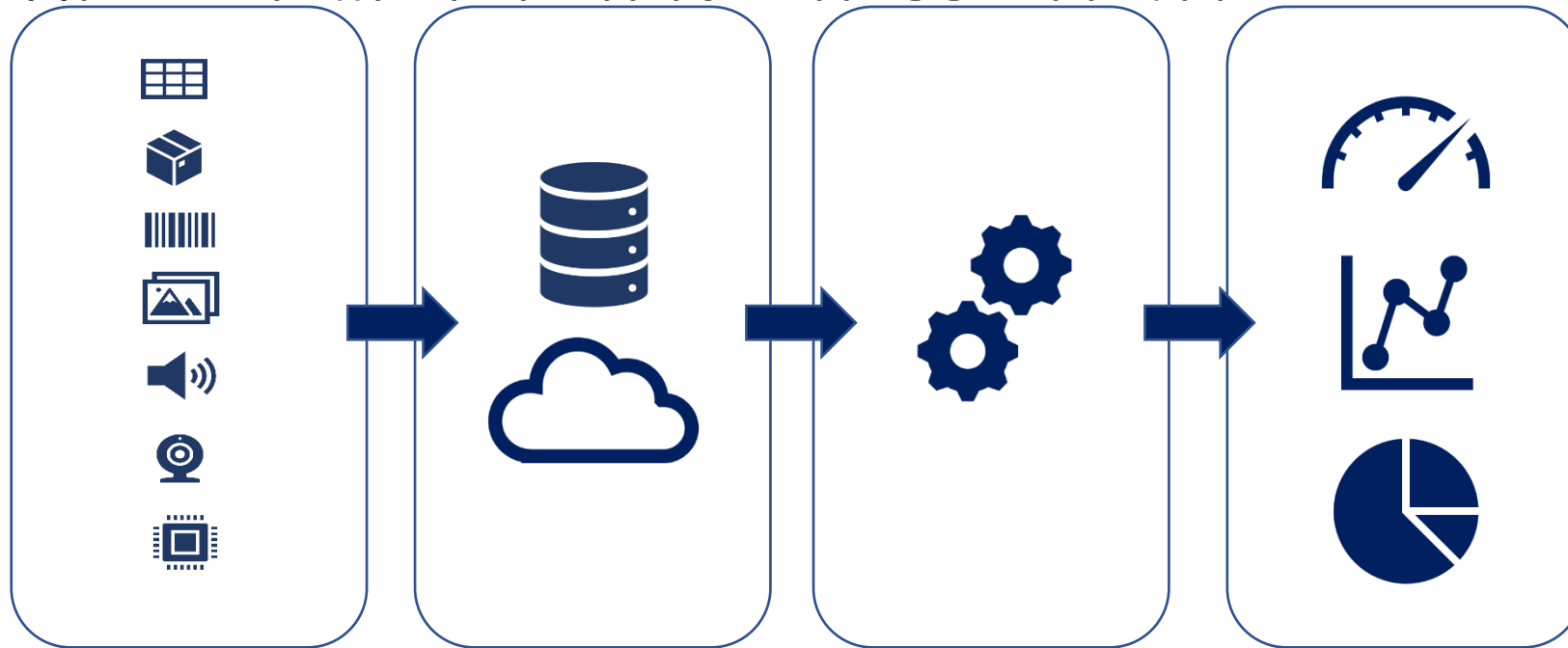


Understanding Big Data

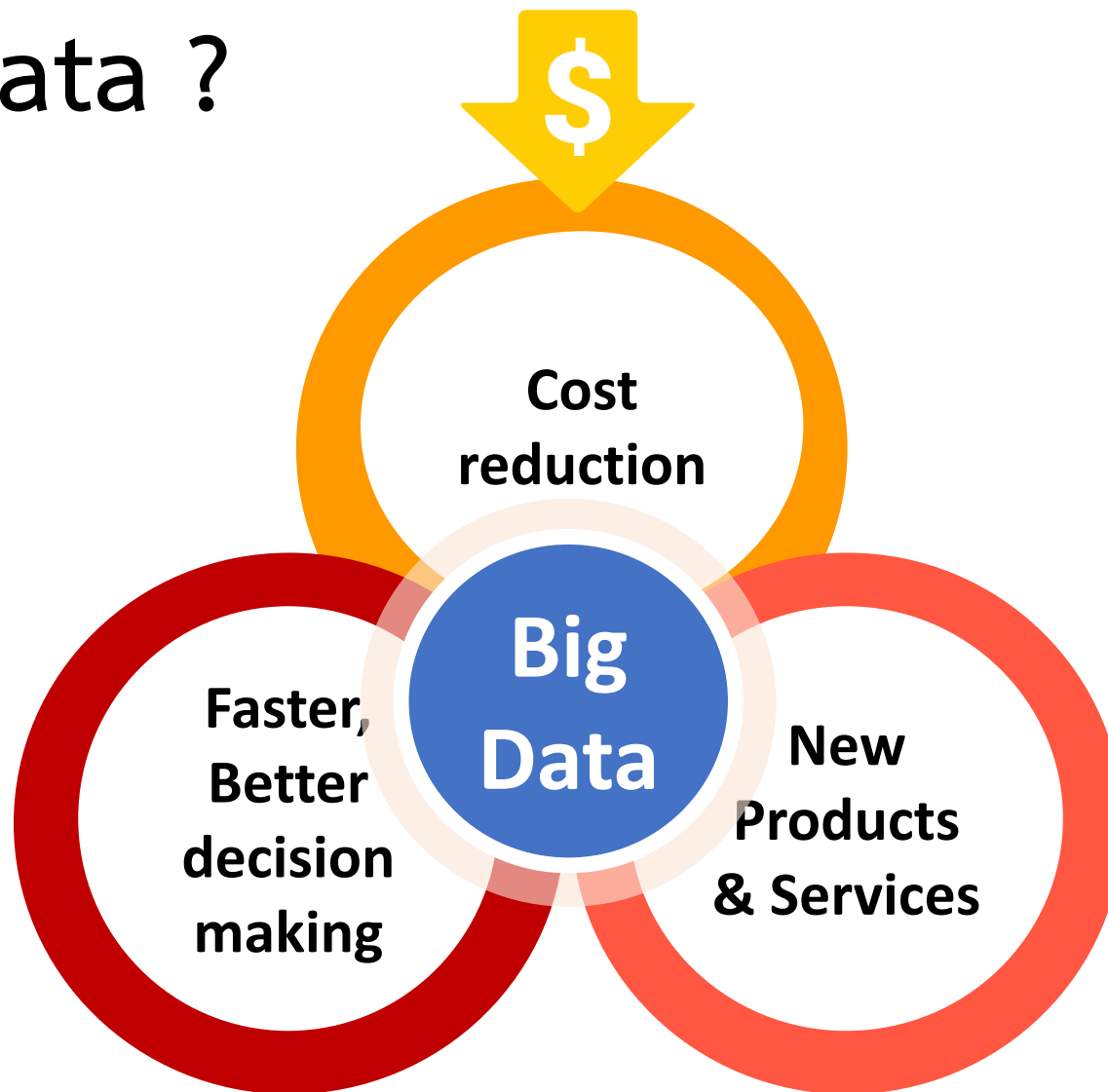
- Big Data Concepts and Terminology
- Characteristics of Big Data
- Types of Data
- Big Data Ecosystem
- Big Data for Organization Data Disruption
- Storing Big Data
- Extract Transform Load (ETL)

Big Data ?

- Big Data = การใช้ประโยชน์จากกระบวนการรวบรวม การจัดเก็บ การประมวลผล และการวิเคราะห์ข้อมูล ที่มีความหลากหลาย มีปริมาณมหาศาล และสามารถใช้งานได้อย่างรวดเร็ว

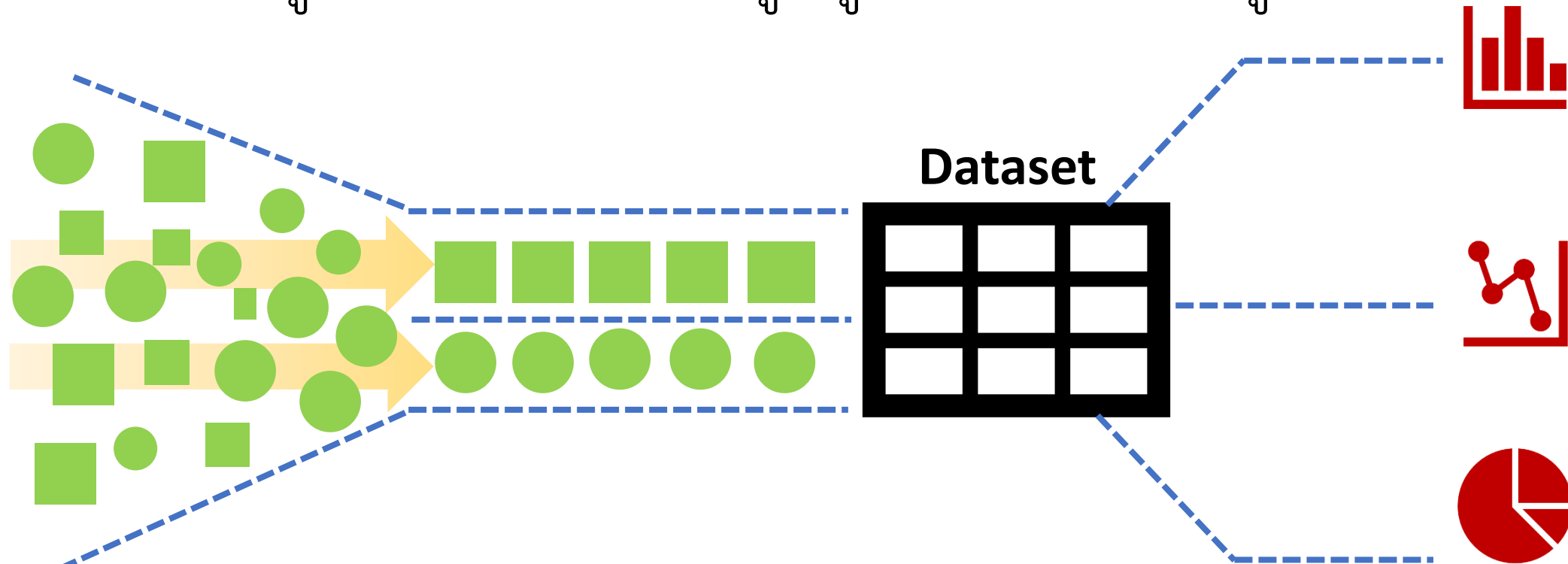


Why Big Data ?



Big Data Terminology

- **Datasets** : ชุดข้อมูลหรือกลุ่มข้อมูลที่เกี่ยวข้งกัน ซึ่งอาจจะอยู่ในหลากหลายรูปแบบ เช่น ไฟล์ข้อมูล รูปภาพ ตารางข้อมูล เป็นต้น



Sources of Open Datasets

- UCI Machine Learning Repository!

<http://archive.ics.uci.edu/ml/index.php>

- World Bank Open Data

<https://data.worldbank.org/>

- Kaggle Open Datasets and Machine Learning Projects

<https://www.kaggle.com/datasets>

- UNICEF Dataset

<https://data.unicef.org/>

- Google Public Data Explorer

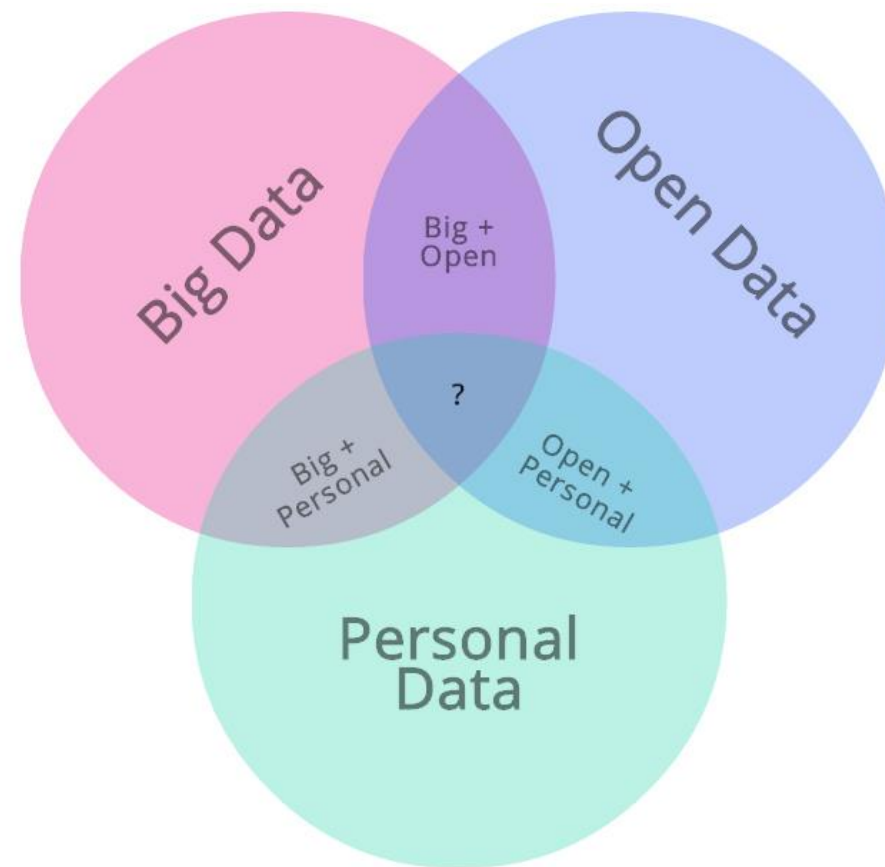
<https://www.google.com/publicdata/directory>

- Google Dataset Search

<https://datasetsearch.research.google.com/>

- Open Data Thailand

<https://data.go.th/>



<https://data.go.th/>

หน้าหลัก | ชุดข้อมูล | อองค์กร | **กลุ่มชุดข้อมูล** | กิจกรรมและข่าวสาร | เอกสารเผยแพร่ | ร้องขอชุดข้อมูล

ขนาดตัวอักษร | ก | ก | ก

กลุ่มชุดข้อมูล

ค้นหาหมวดหมู่...

พบ 15 หมวดหมู่ | เรียงโดย : เรียงชื่อตามลำดับตัวอักษร (ก-อ)

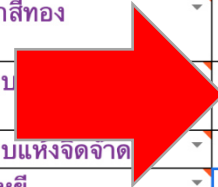
ศาสนา ศิลปะ และ วัฒนธรรม (140)	กฎหมาย ศาล และ อาชญากรรม (79)	การศึกษา (26)	เศรษฐกิจ การเงิน และ อุตสาหกรรม (284)	สาธารณสุข (85)	เกษตรกรรม (75)
สถิติทางการ (469)	วิทยาศาสตร์ เทคโนโลยี ดิจิทัล และนวัตกรรม (63)	เมืองและภูมิภาค (172)	สังคมและสวัสดิการ (260)	คมนาคมและโลจิสติกส์ (242)	การท่องเที่ยวและกีฬา (49)

Big Data Terminology

Data Cleansing : คือกระบวนการในการ



C	D	E	F	G	H	I	J	K	L	M
เลิกเพื่อป้อนหรือเลือกซื้อสินค้า	ชื่อนี้อย่าลบนะคะ	Lot #1				Lot #2				
ชื่อสินค้า	แบรนด์	รับ	ขาย	ทดสอบ	วางจำหน่าย	รับ	ขาย	ทดสอบ	วางจำหน่าย	รับ
างกุ้งขอสมะขาม	อำพัน ออร์แกนิกฟาร์ม	4			0				0	
อยจ้อกรเชียงปู่	เบ๊ ภัทตาการ	6			0				0	
โด้ถั่งเจ้าสีทอง	CMU CORDYCEPS	1			0				0	
ลเบอร์รี่	#N/A	8			0				0	
ลเบอร์รี่แห้งจัดจาด	#N/A	2			0				0	
กหม่อนหยี	#N/A	1			0				0	
unsnack	#N/A	1			0	1			0	
ปประสทอดกรอบ	#N/A									
ะขามแช่อิ่มแบบกระปุก	#N/A	3			0				0	
นมเบียร์ทุเรียน พูเล็ก (size S)	นางสาวทุเรียน	8			0				0	
ะขามแช่อิ่มมอบแห้งของใหญ่	#N/A	2			0				0	
ะขามแช่อิ่มมอบแห้งของใหญ่ลายไทย	#N/A	2			0				0	



Data Cleansing

ข้อมูลเป็นแบบนี้จะทำอย่างไรให้ข้อมูลถูกต้อง
อยู่ในรูปแบบเดียวกัน
นำไปใช้วิเคราะห์ได้ และเชื่อถือได้ ?

	A	B	D	E	F	G	H	I
1	วันที่	รายการสินค้า	จำนวน	ราคาขาย	รวมรายได้	ราคาส่ง	ต้นทุน	กำไร
2	3/4/2562	นมดื่มเกษตรมช.	20	20.00	400.00	22.00	440.00	-40.00
3	3/4/2562	ข้าวแช่คว (ซอง)	3	15.00	45.00	15.00	45.00	0.00
4	3/4/2562	ถั่วมิกส์นัท (กระป๋อง)	3	15.00	45.00	15.00	45.00	0.00
5	3/4/2562	ถั่วลายเสือคว (กระป๋อง)	1	63.00	63.00	63.00	63.00	0.00
6	4/4/2562	ถั่วมิกส์นัท (ซอง)	1	15.00	15.00	15.00	15.00	0.00
7	4/4/2562	ข้าวแช่คว (กระป๋อง)	1	59.00	59.00	59.00	59.00	0.00
8	4/4/2562	ข้าวแช่คว (ซอง)	3	15.00	45.00	15.00	45.00	0.00
9	4/4/2562	ถั่วมิกส์นัท (กระป๋อง)	1	75.00	75.00	75.00	75.00	0.00
10	4/4/2562	นมดื่มเกษตรมช.	20	20.00	400.00	18.00	360.00	40.00
11	4/4/2562	ถั่วลายเสือคว (กระป๋อง)	1	63.00	63.00	59.00	59.00	4.00
12	4/4/2562	ถั่วเหลืองคว (กระป๋อง)	1	57.00	57.00	57.00	57.00	0.00
13	5/4/2562	ถั่วมิกส์นัท (ซอง)	2	15.00	30.00	15.00	30.00	0.00
14	5/4/2562	ข้าวแช่คว (ซอง)	1	15.00	15.00	15.00	15.00	0.00
15	5/4/2562	ข้าวแช่คว (ซอง)	1	15.00	15.00	15.00	15.00	0.00

แบรนด์	จังหวัด
นมดื่มเกษตรมช.	เชียงใหม่
ถั่วคุณลุง	แม่ฮ่องสอน
ถั่วคุณลุง	MHS
ถั่วคุณลุง	แม่ฮ่องสอน
ถั่วคุณลุง	แม่ฮ่องสอน
ถั่วคุณลุง	แม่ฮ่องสอน
ถั่วคุณลุง	แม่ฮ่องสอน
ถั่วคุณลุง	มฮส.
นมดื่มเกษตรมช.	ชม.

Garbage in Garbage out
หากไม่ทำ Cleansing

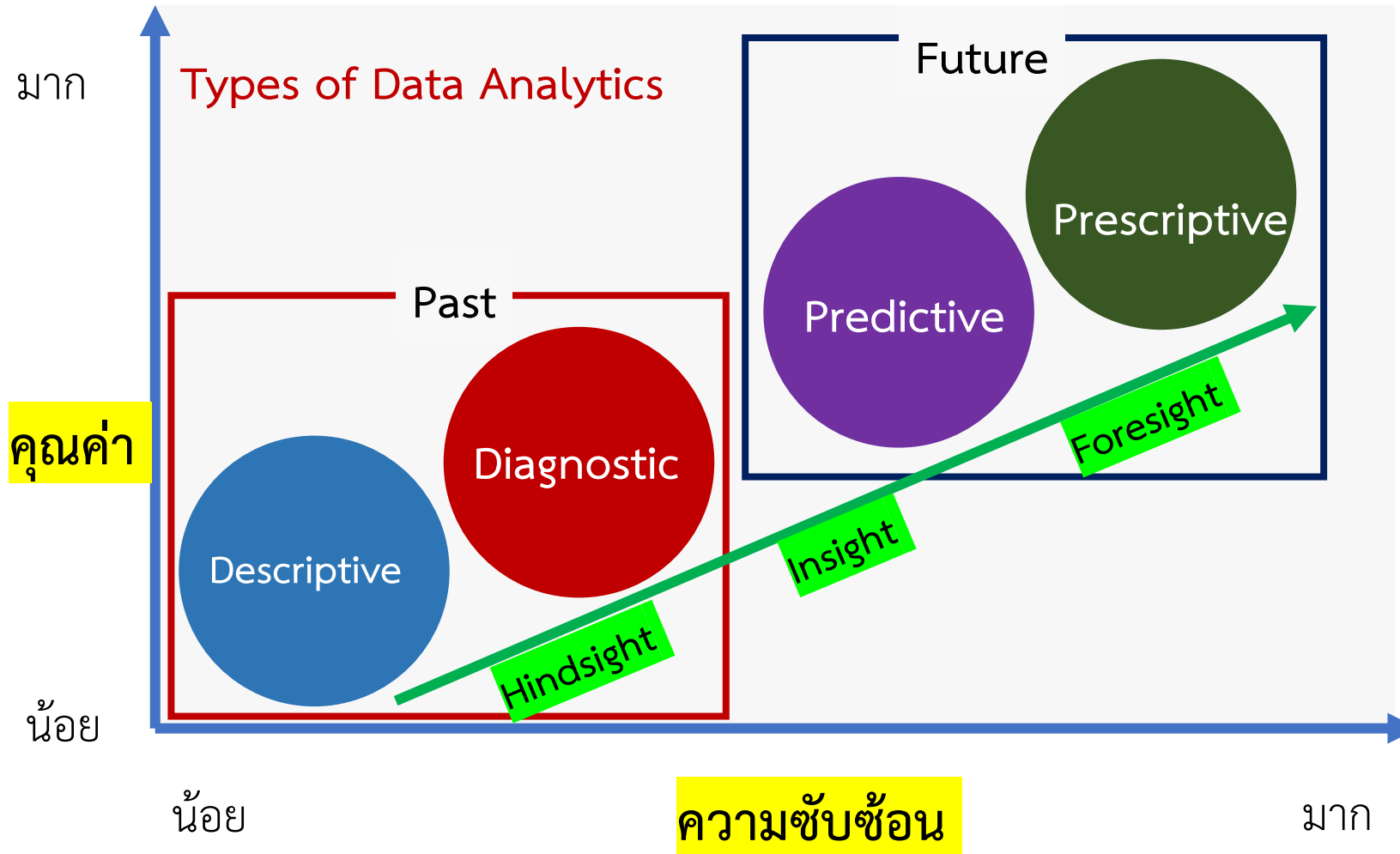
Big Data Terminology

- **Data Analysis** หรือการวิเคราะห์ข้อมูล เป็นกระบวนการตรวจสอบข้อมูลเพื่อค้นหาข้อเท็จจริงความสัมพันธ์รูปแบบข้อมูลเชิงลึกและ/หรือแนวโน้ม
 - เป้าหมายโดยรวมของคือเพื่อใช้ในการสนับสนุนการตัดสินใจได้ดีขึ้น
- ตัวอย่าง เช่น
- “การวิเคราะห์ข้อมูลการขายไอศกรีมเพื่อดูว่าจำนวนไอศกรีมที่ขายมีความสัมพันธ์กับอุณหภูมิประจำวันอย่างไร”
 - ผลของการวิเคราะห์ดังกล่าวจะใช้สนับสนุนการตัดสินใจเกี่ยวกับปริมาณไอศกรีมที่ร้านค้าควรสต็อกไว้โดยสัมพันธ์กับข้อมูลพยากรณ์อากาศ

Big Data Terminology

- **Data Analytics** : การวิเคราะห์ข้อมูลมีความหมายที่กว้างขึ้นซึ่งครอบคลุมถึงการวิเคราะห์ข้อมูล
- การวิเคราะห์ข้อมูลเป็นระเบียบวิธีที่รวมถึงการจัดการวงจรชีวิตของข้อมูลที่สมบูรณ์ซึ่งรวมถึงการรวบรวมทำความสะอาดจัดระเบียบจัดเก็บวิเคราะห์และควบคุมข้อมูล

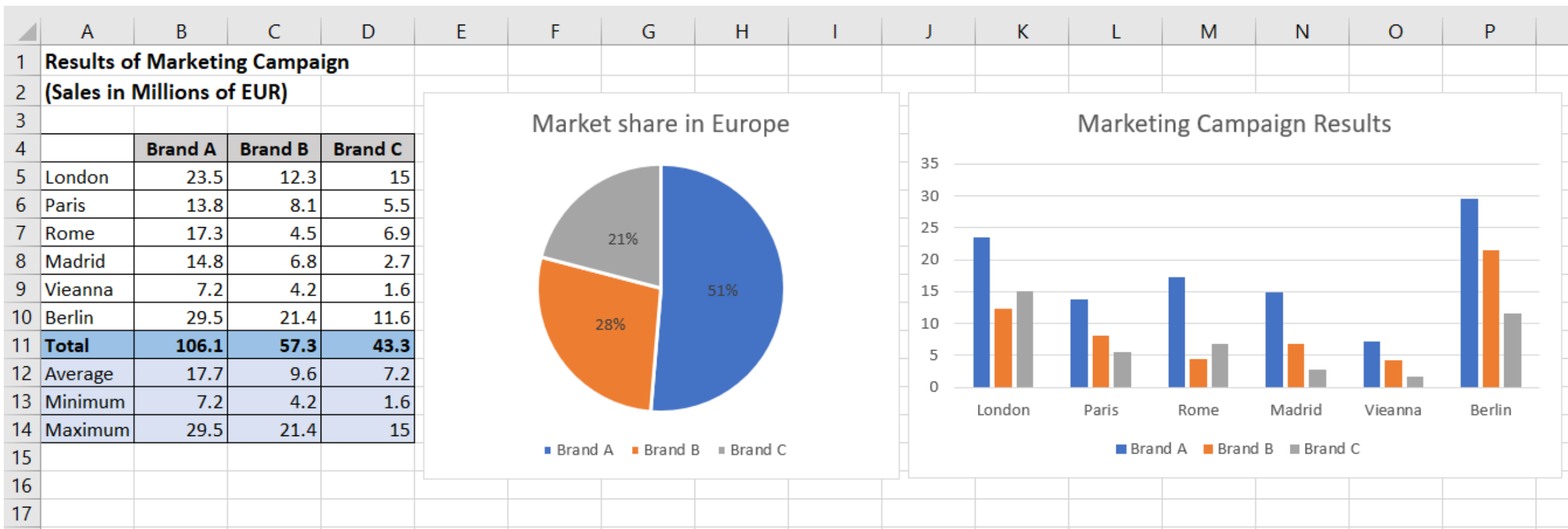
Big Data Terminology



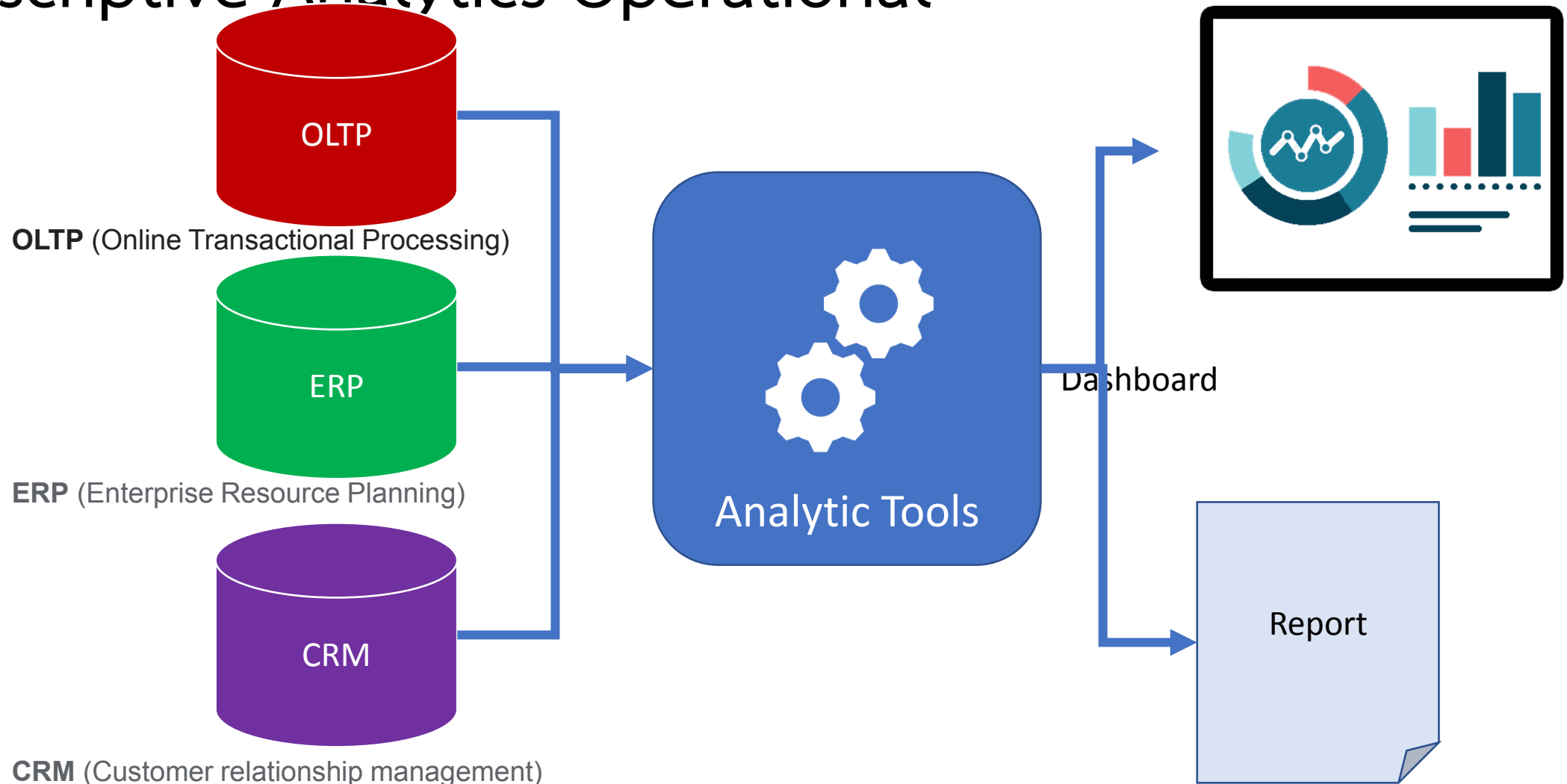
Big Data Terminology

- **Descriptive Analytics : “What happened ?”**
- การวิเคราะห์เชิงพรรณนาเป็นรูปแบบการใช้ข้อมูลแบบพื้นฐาน โดยเน้นการอธิบายการเกิดเหตุการณ์ต่าง ๆ ในอดีต
- เช่น รายงานธุรกิจ รายงานด้านการทำโฆษณา หรือรายงานผลดำเนินงานที่ผ่านมา เป็นข้อมูลพื้นฐานที่แสดงผลในการทำกิจกรรมต่าง ๆ

Descriptive Analytics case study



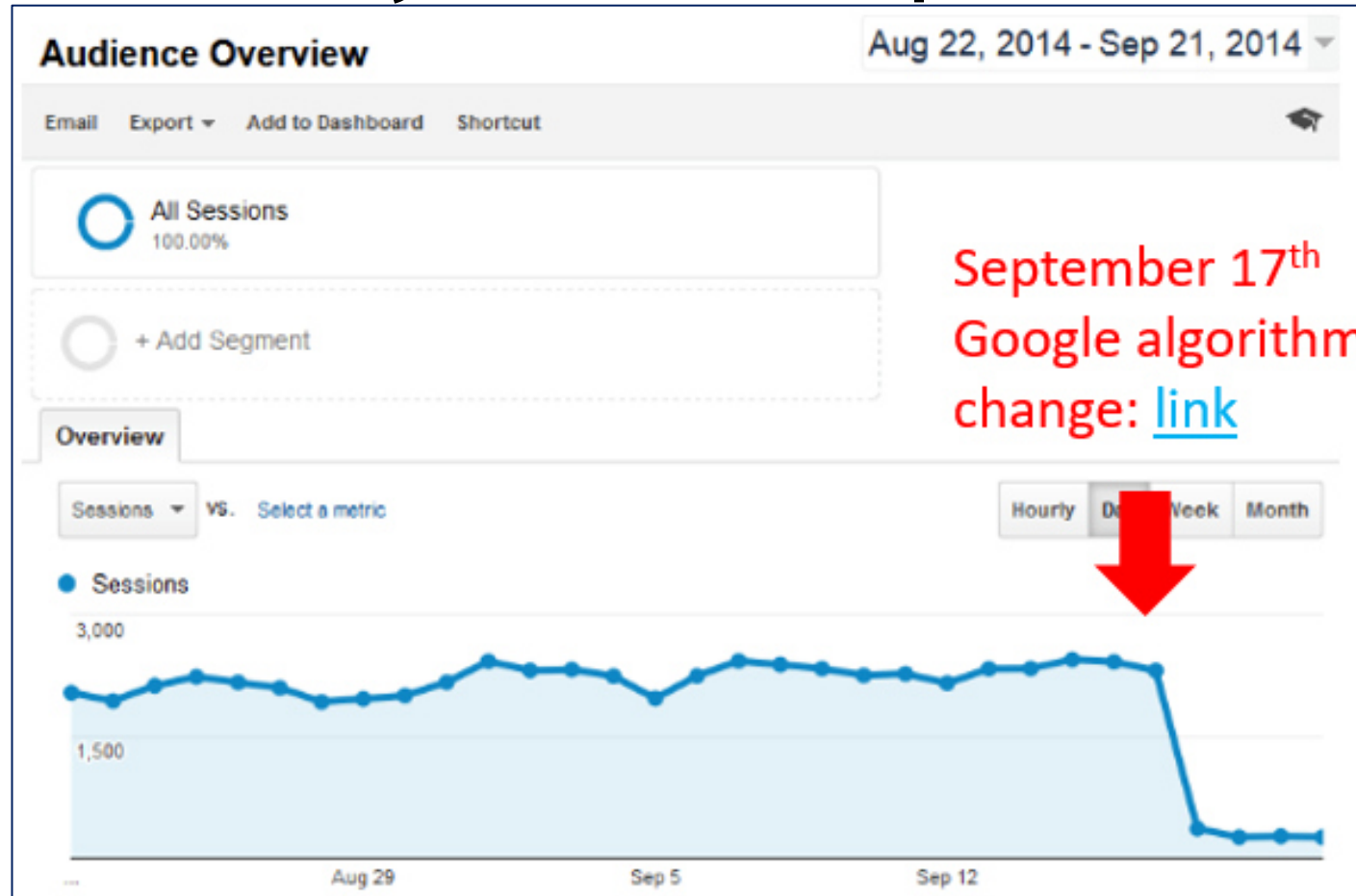
Descriptive Analytics Operational



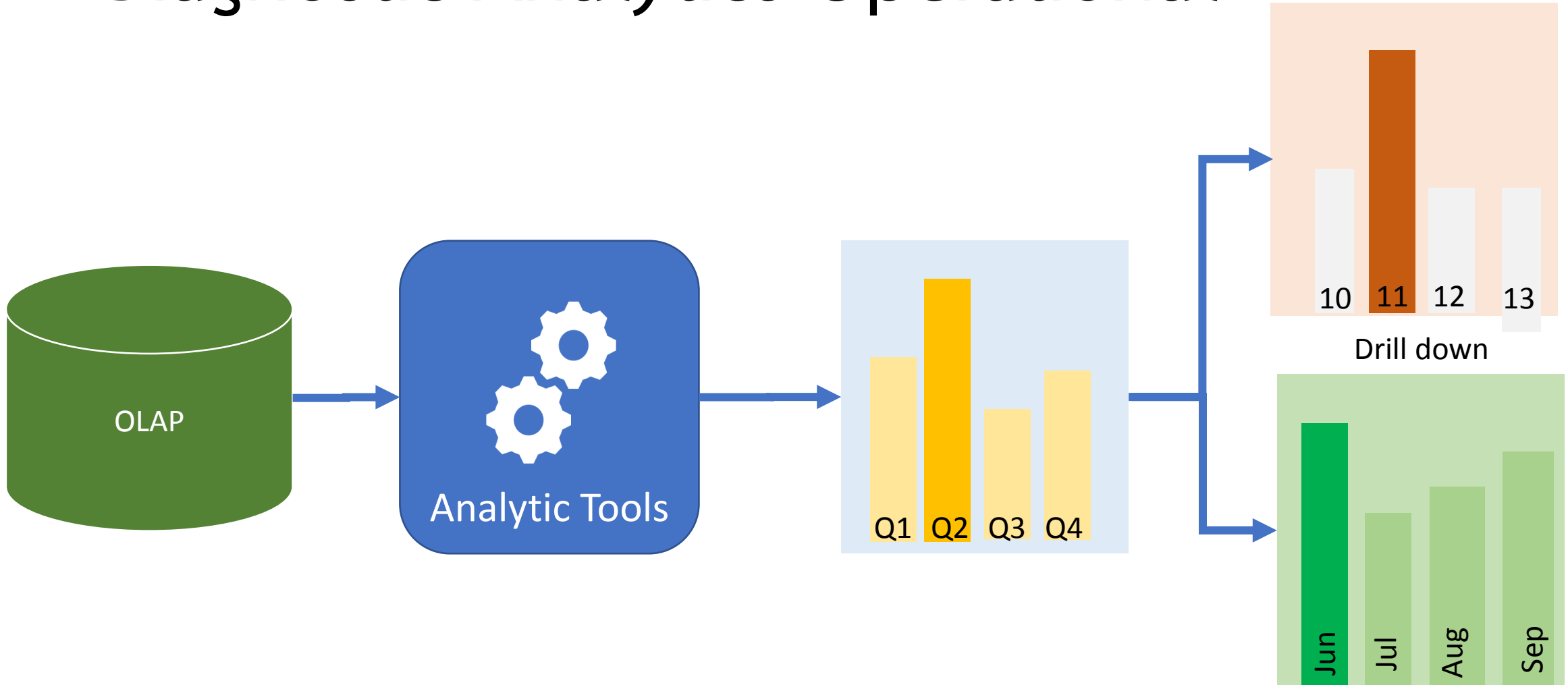
Big Data Terminology

- **Diagnostic Analytics : “Why it happened ?”**
- เป็นการวิเคราะห์เชิงวินิจฉัยหาสาเหตุ โดยเปรียบเทียบข้อมูลในอดีตเทียบกับข้อมูลอื่นเพื่อตอบคำถามว่า ทำไมบางสิ่งเกิดขึ้น
- การวิเคราะห์เชิงวินิจฉัยทำให้มีความเป็นไปได้ที่จะเจาะลึกค้นหา การอ้างอิงและระบุรูปแบบ จากข้อมูลเชิงลึกเกี่ยวกับปัญหาเฉพาะ แต่จะต้องมีข้อมูลที่มีความละเอียดมากเพียงพอ

Diagnostic Analytics Example



Diagnostic Analytics Operational

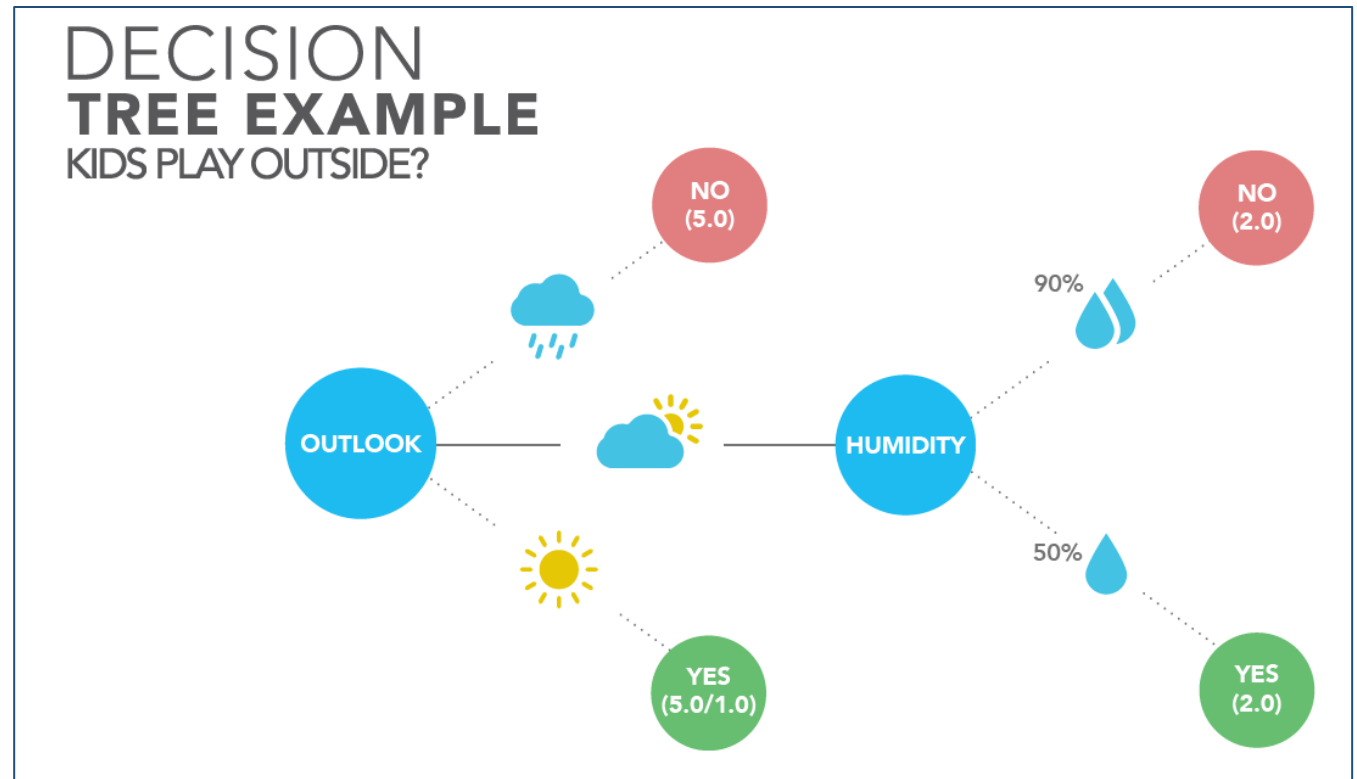


Big Data Terminology

- **Predictive Analytics : “What might happen ?”**
- เป็นการวิเคราะห์เชิงพยากรณ์ รูปแบบการใช้ข้อมูลที่มีความซับซ้อน โดยเป็นการ “พยากรณ์” หรือ “ทำนาย” สิ่งที่กำลังเกิดขึ้น โดยใช้ข้อมูลในอดีต ร่วมกับโมเดลทางคณิตศาสตร์ต่าง ๆ ร่วมกับการทำ Data Mining
- นอกจากนี้ยังทำให้เราสามารถวิเคราะห์หาโอกาสและความเสี่ยงต่าง ๆ ที่จะเกิดขึ้นในอนาคตได้ด้วย เช่น
 - การรู้แนวโน้มทางการตลาด
 - การพยากรณ์ยอดการขายหรือการทำโฆษณาว่าจะมีคนร่วมเท่าไร หรือ
 - มีค่า ROI (Return of Investment) เท่าไร

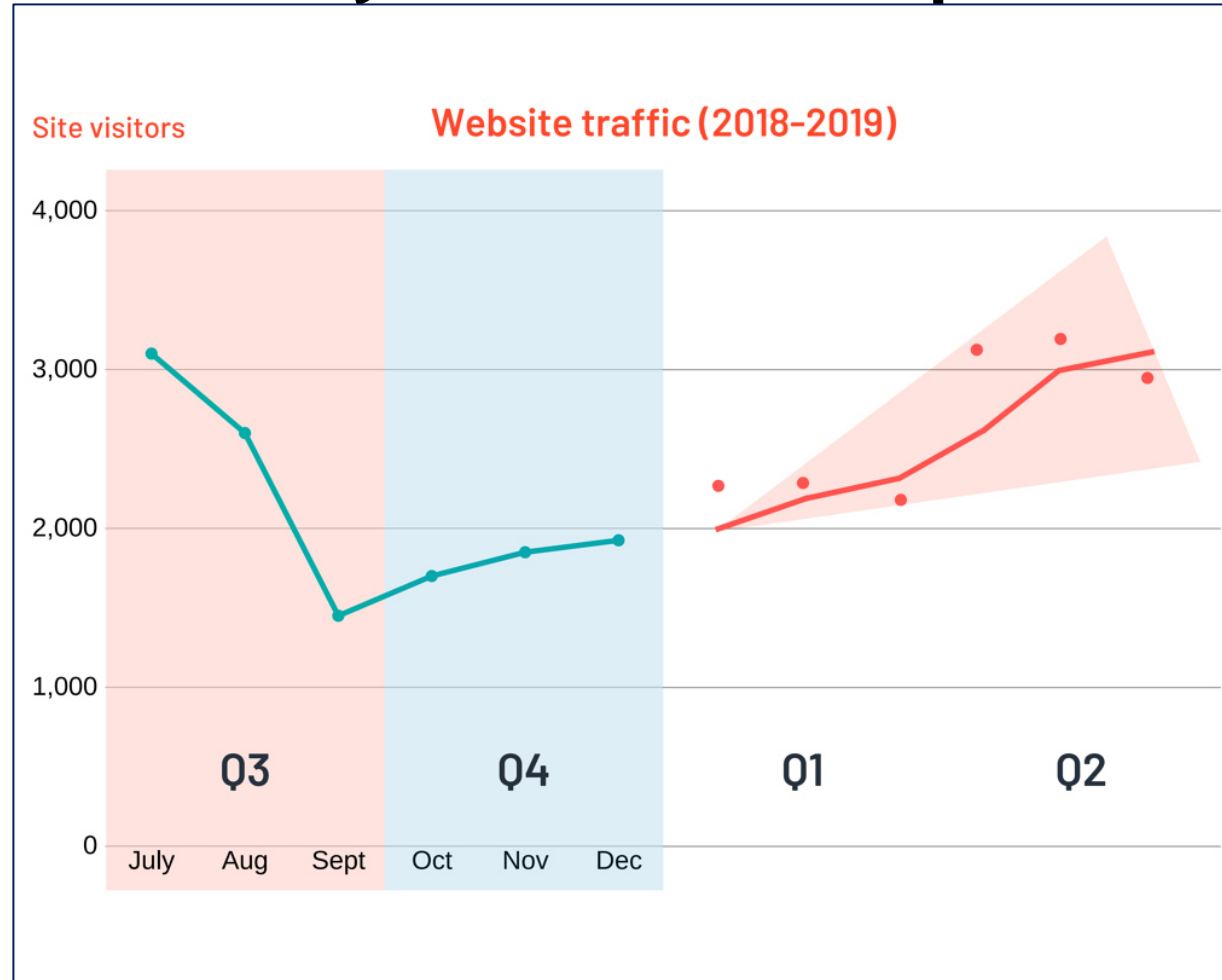
Predictive Analytics Example

Day #	Forecast	Temperature	Humidity	Play outside
Day 1	sunny	hot	high	yes
Day 2	sunny	hot	high	yes
Day 3	cloudy	hot	high	no
Day 4	rainy	mild	high	no
Day 5	rainy	cool	normal	no
Day 6	rainy	cool	normal	no
Day 7	cloudy	cool	normal	yes
Day 8	sunny	mild	high	yes
Day 9	sunny	cool	normal	no
Day 10	rainy	mild	normal	no
Day 11	sunny	mild	normal	yes
Day 12	cloudy	mild	high	no
Day 13	cloudy	hot	normal	yes
Day 14	rainy	mild	high	no

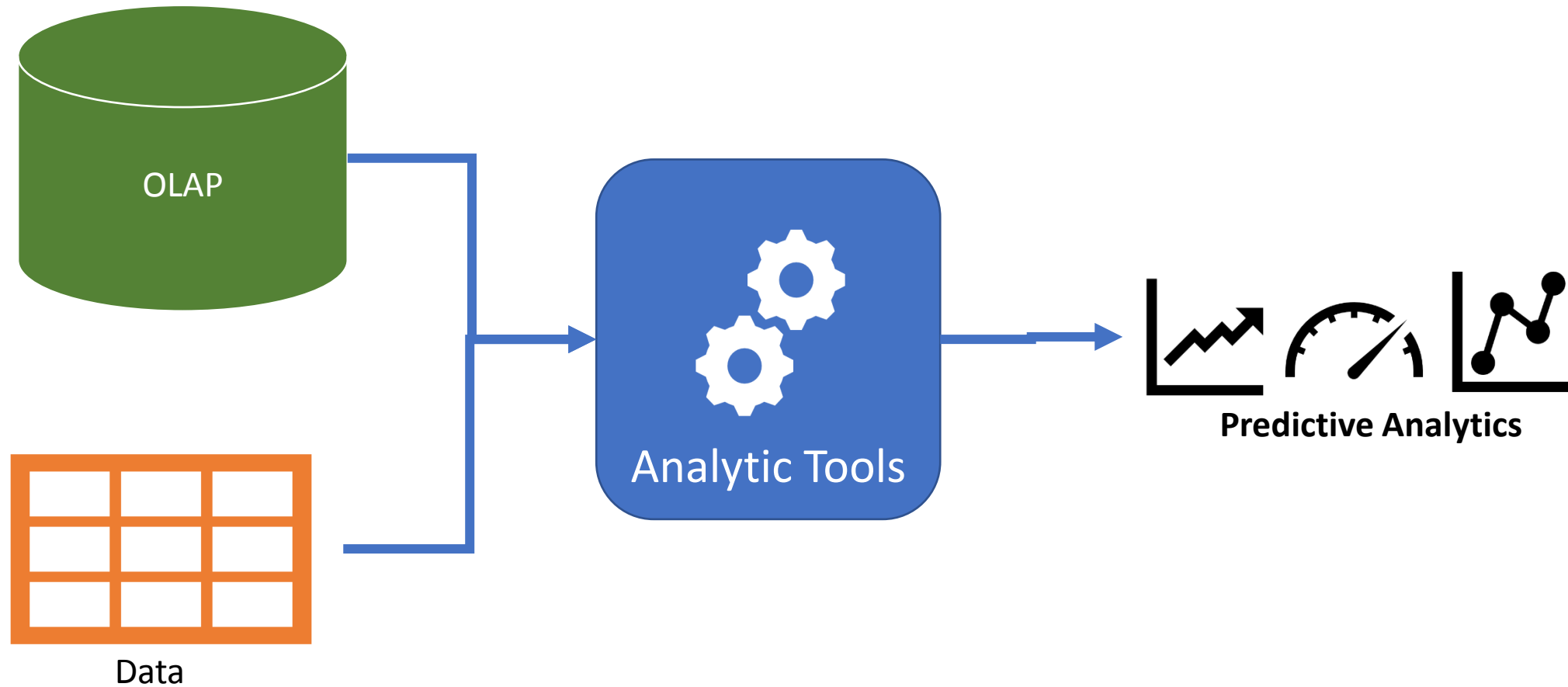


<https://www.analyticsinhr.com/blog/predictive-analytics-human-resources/>

Predictive Analytics Example



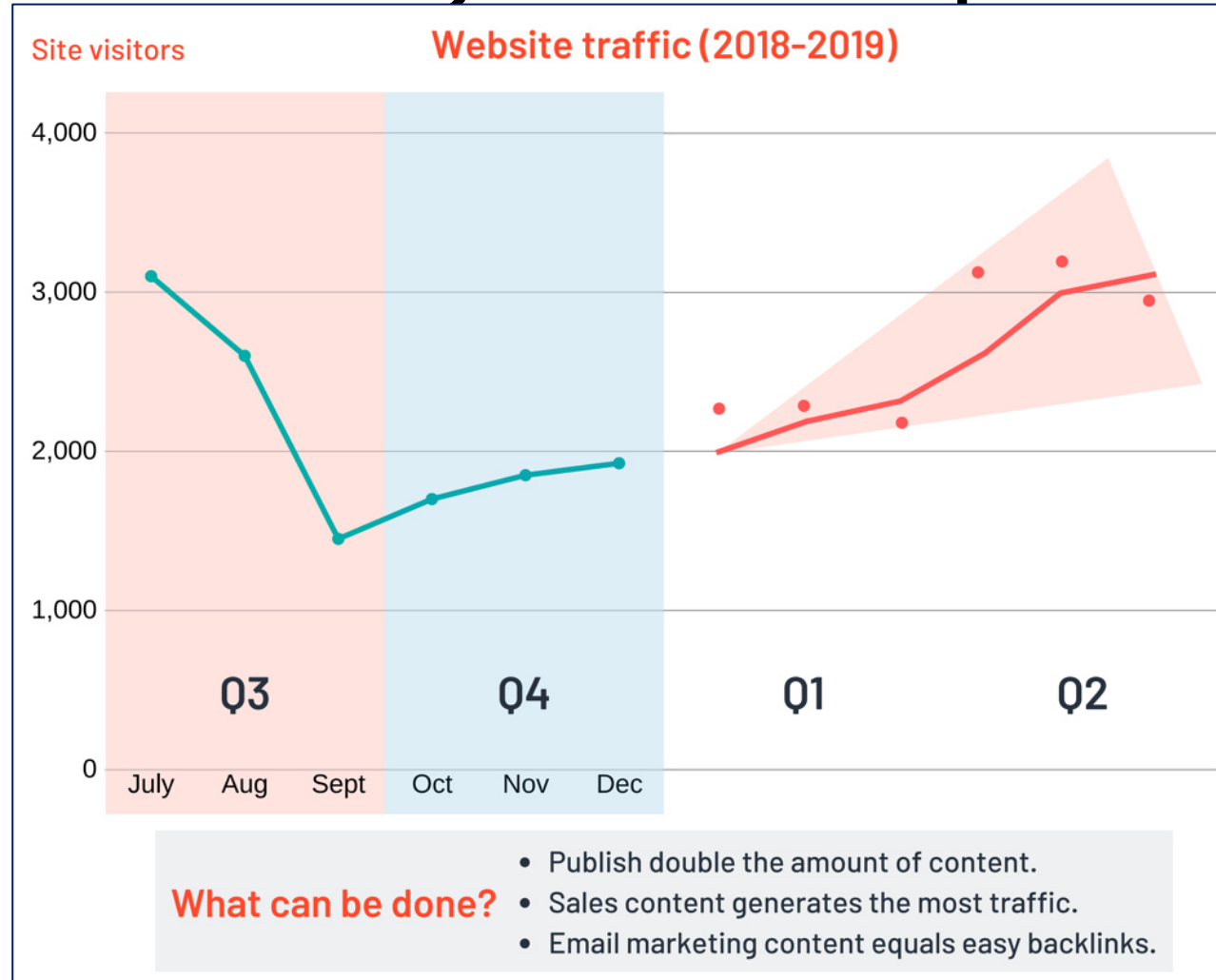
Predictive Analytics Operational



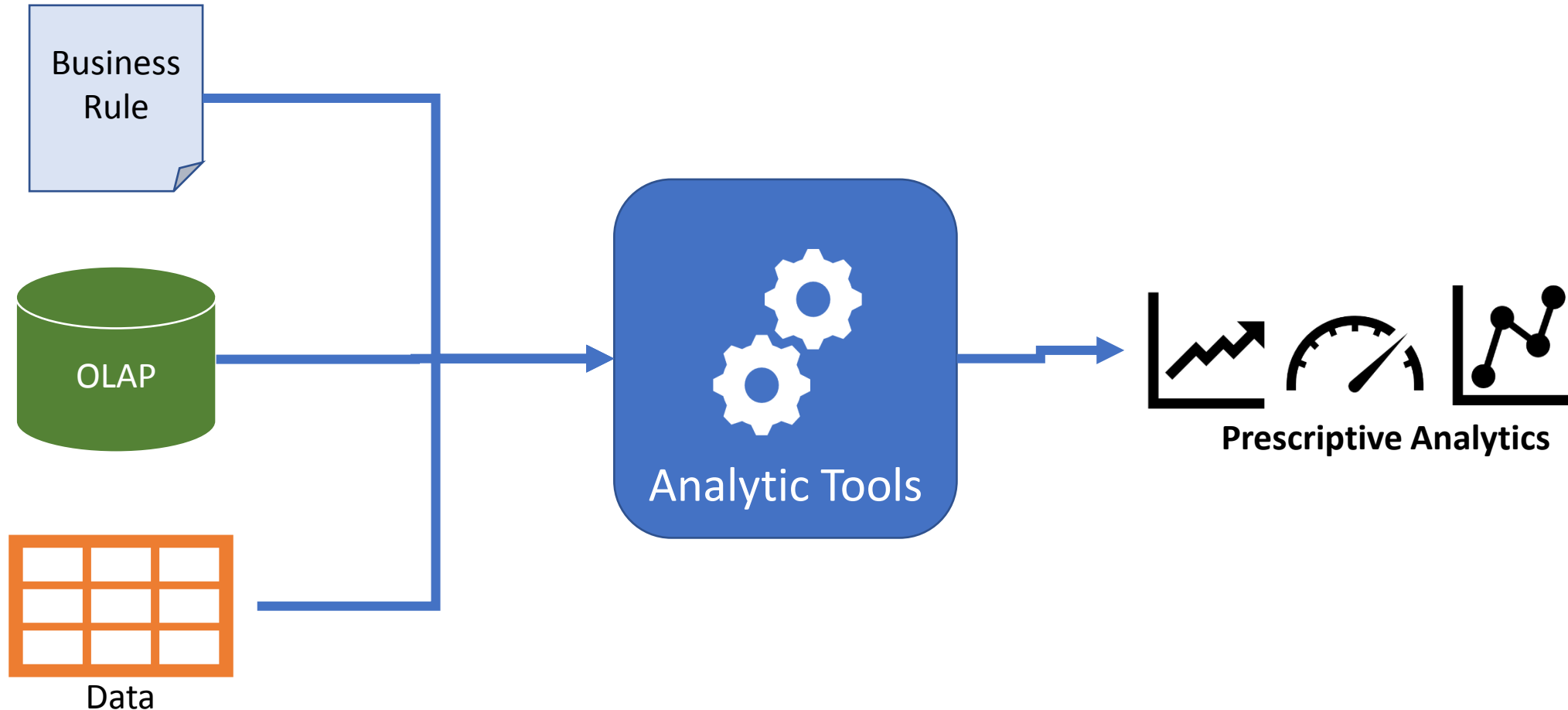
Big Data Terminology

- **Prescriptive Analytics : “What should we do?”**
- เป็นการวิเคราะห์เชิงพยากรณ์และทางเลือก เป็นรูปแบบที่มีความซับซ้อนและยากที่สุด
- เพราะไม่เพียงพยากรณ์หรือทำนายว่าอะไรจะเกิดขึ้น แต่ยังให้คำแนะนำในทางเลือกต่าง ๆ และผลแต่ละทางเลือกกว่าจะมี Pros & Consอย่างไร
- โมเดลสามารถปรับเปลี่ยนได้ตามข้อมูลที่เพิ่มเติมเข้ามา ในรูปแบบนี้ก็ยังเป็นการใช้ข้อมูลที่มากที่สุด และเกี่ยวข้องกับเรื่อง Big Data มากที่สุด

Prescriptive Analytics Example



Prescriptive Analytics Operational



Descriptive Analytics

What happened?

Univariate Analysis

Diagnostic Analytics

Why it happened?

Multivariate and Multidimensional

Predictive Analytics

What might happen?

Data mining and Machine learning

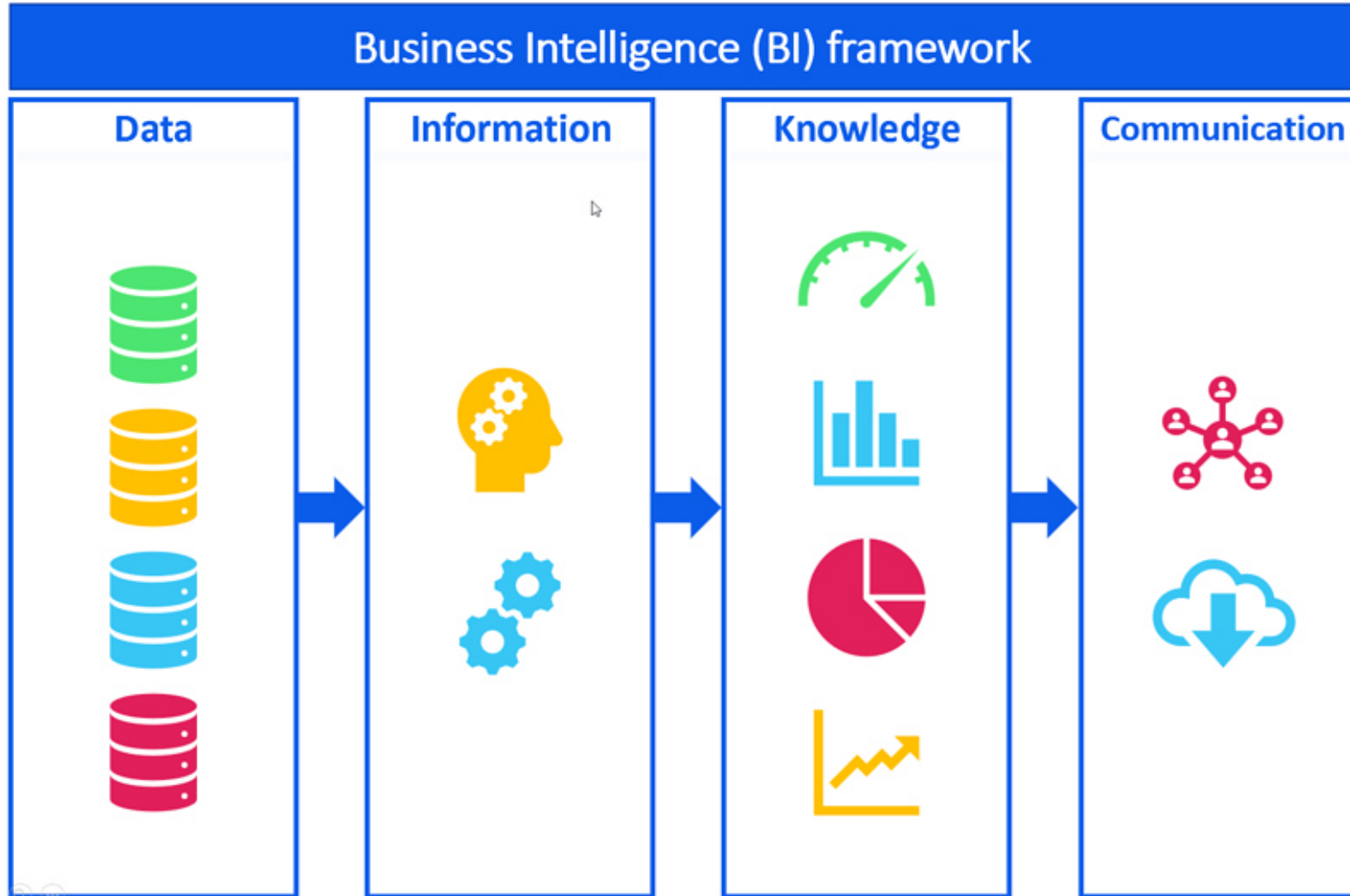
Prescriptive Analytics

What should we do?

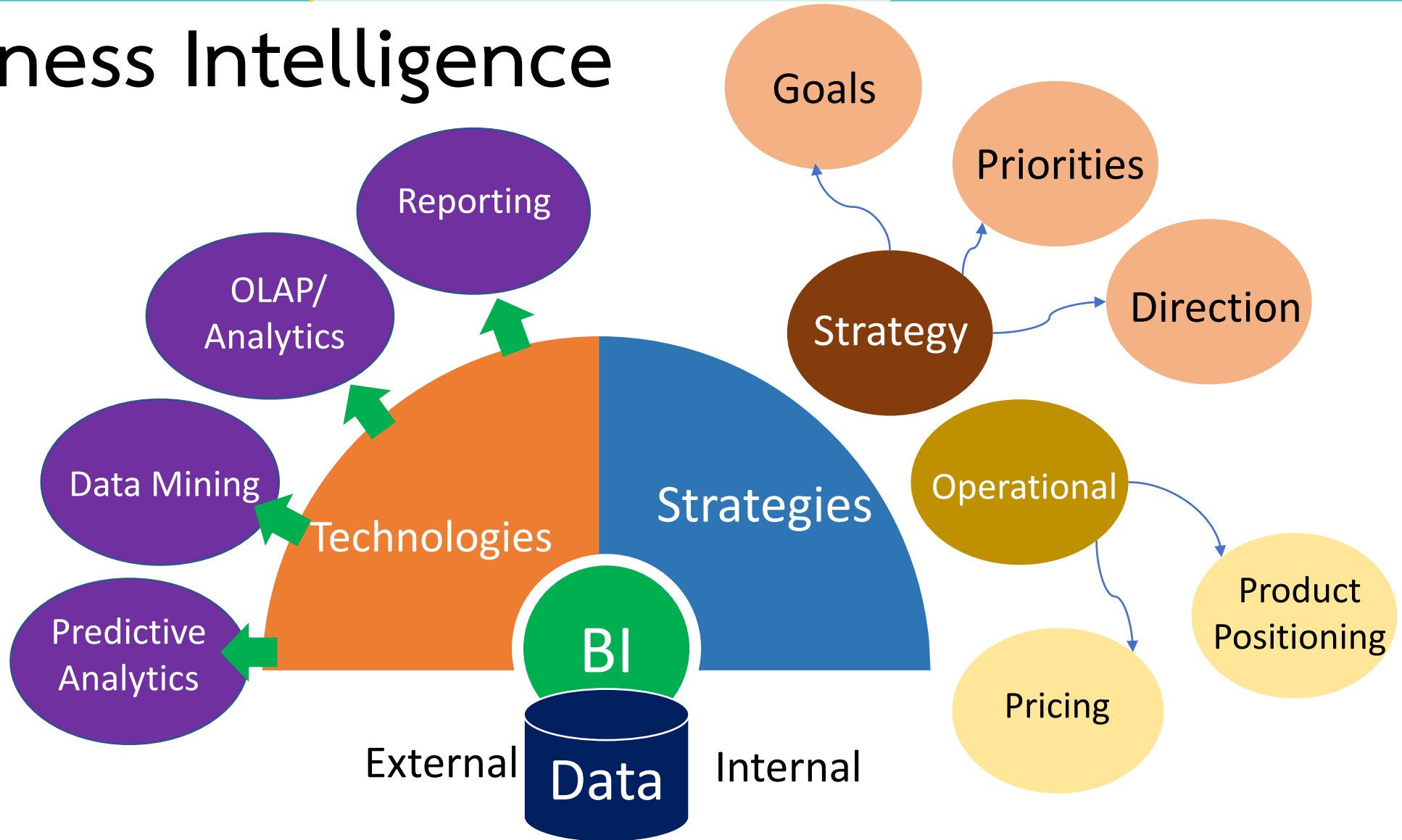
Optimization

Big Data Terminology

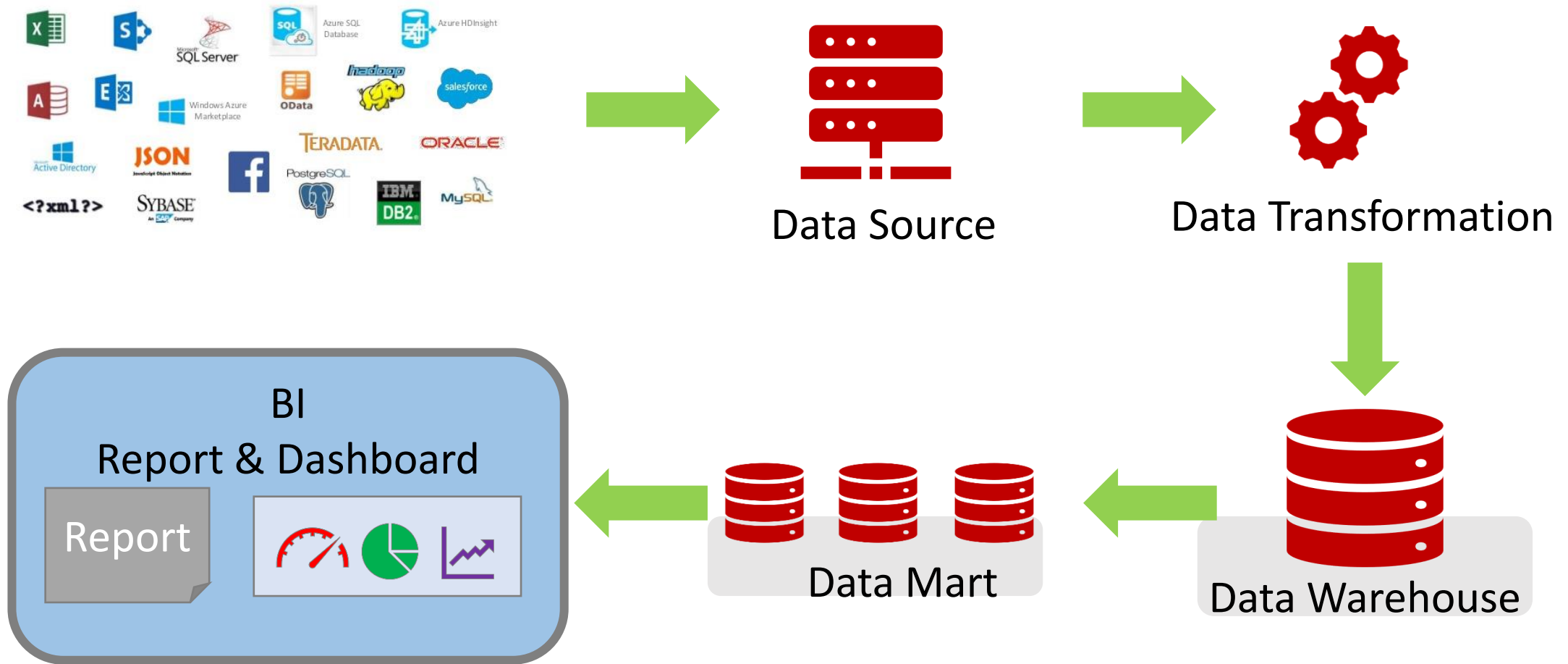
- Business Intelligence (BI)
- เป็นกระบวนการที่องค์กรจะใช้ขับเคลื่อนการดำเนินงานด้วยเทคโนโลยี โดยเน้นในด้านการวิเคราะห์ข้อมูลและนำเสนอข้อมูล
- สามารถนำไปช่วยให้ผู้บริหาร ผู้จัดการและผู้ใช้ใน ขององค์กรสามารถ ตัดสินใจทางธุรกิจได้อย่างมีข้อมูล



Business Intelligence



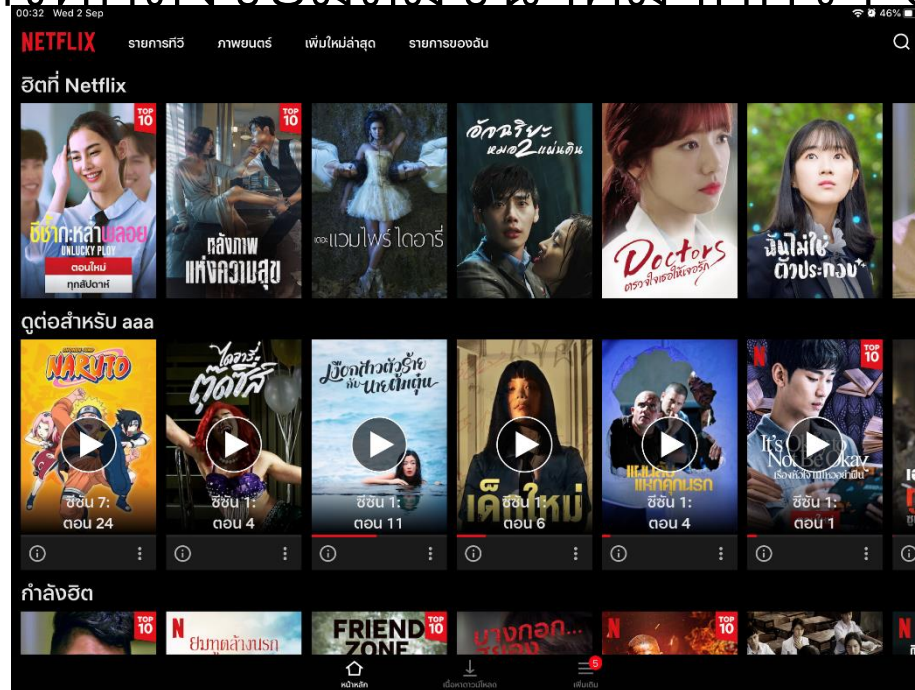
องค์ประกอบของ Business Intelligence (BI)



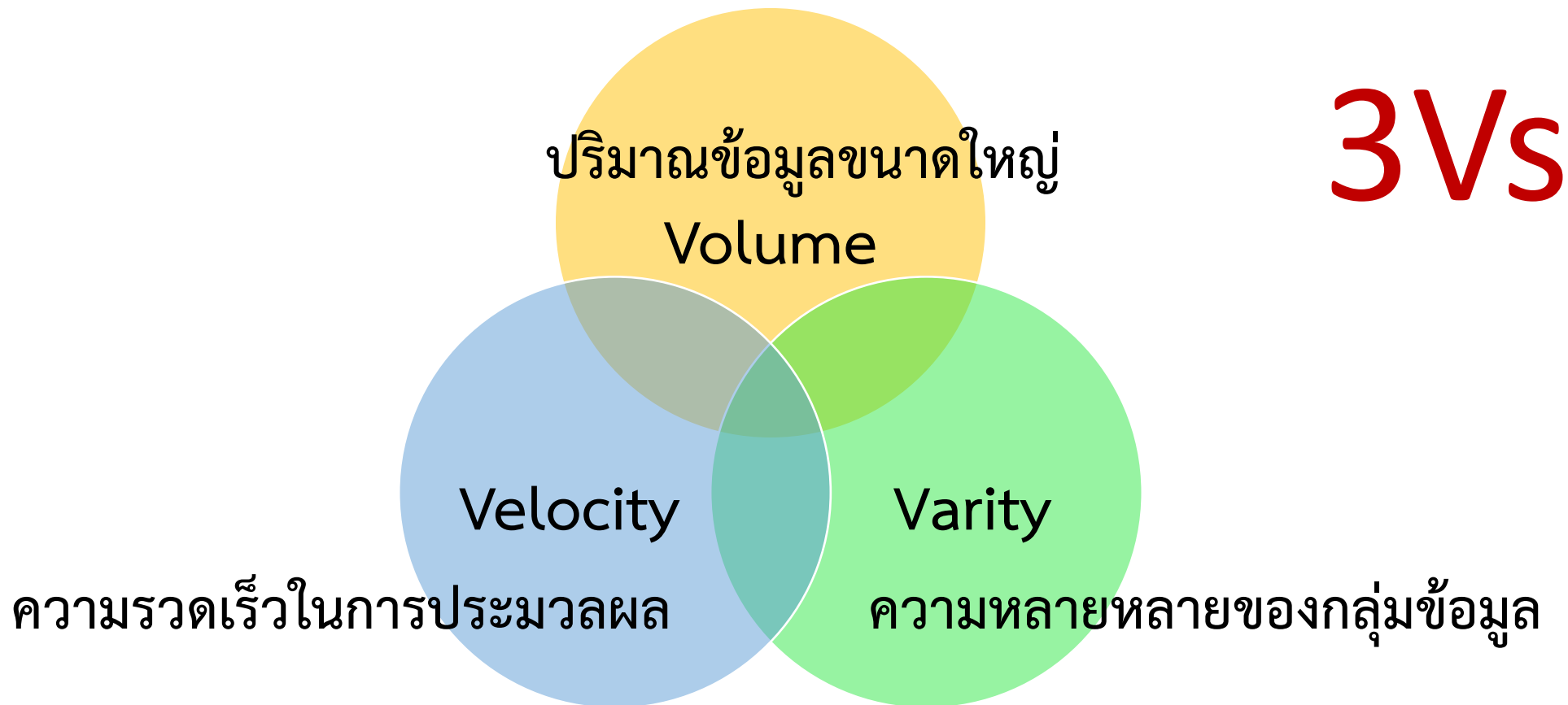
ตย. คุณลักษณะ Big data :Volume

- Netflix มีสมาชิกมากกว่า 86 ล้านคนทั่วโลกสตรีมเนื้อหามากกว่า 125 ล้านชั่วโมงต่อวัน ส่งผลให้คลังข้อมูลมีขนาดมากกว่า 60 Petabyte(PB)

NETFLIX

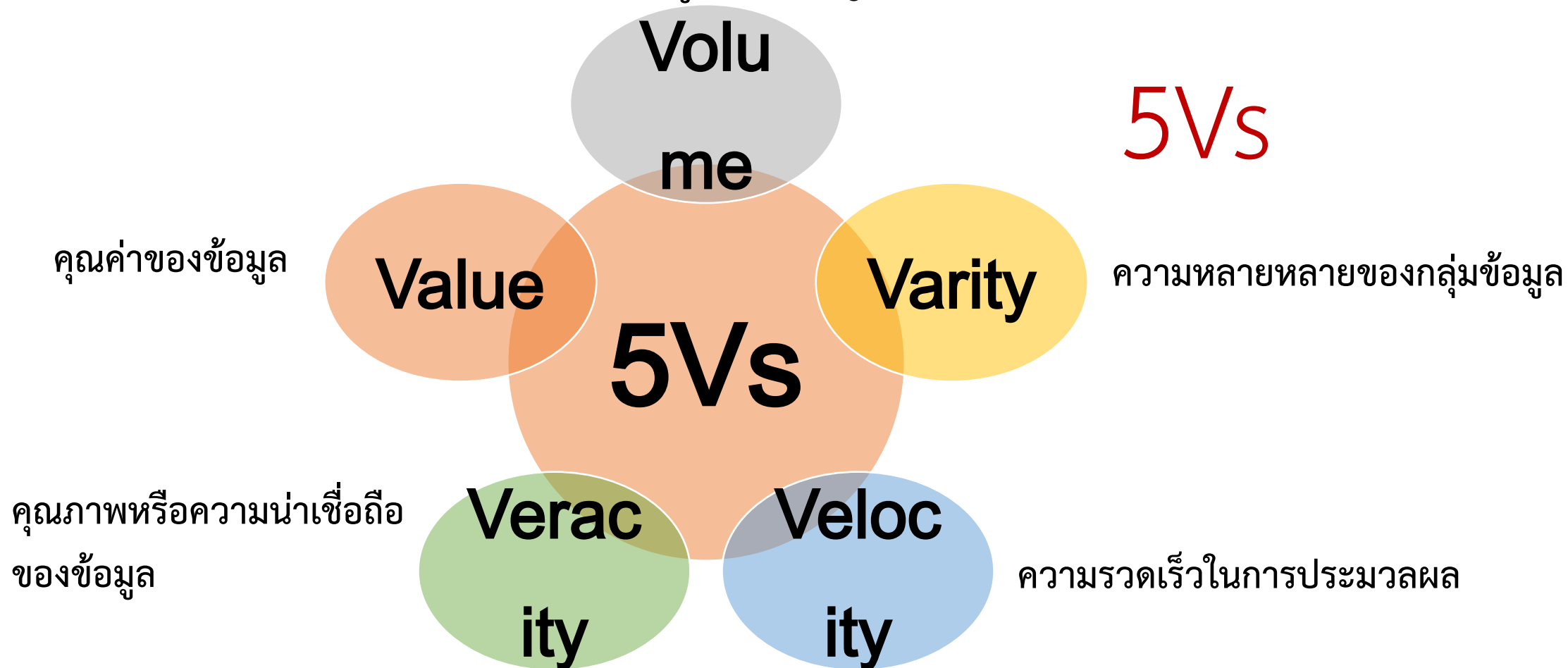


Characteristics of Big Data



Characteristics of Big Data

ปริมาณข้อมูลขนาดใหญ่



ตย. คุณลักษณะ Big data : Velocity

- Google ได้รับการค้นหาจากผู้ใช้ในการค้นหาข้อมูลมากกว่า 63,000 ครั้งต่อวินาทีในแต่ละวัน



ตย. คุณลักษณะ Big data : Variety

- ข้อมูลใน Facebook ประกอบด้วยข้อมูลทั้ง Unstructured และ Structured Data ในโพสต์ต่าง ได้แก่ ข้อความ ตัวเลข รูปภาพ เสียง วิดีโอ ลิงค์ เว็บไซต์ อีเมล เป็นต้น



facebook

Big data : Variety

Bundled flight

Cheapest flight

Depart on Thu, 15 Oct
CNX 18:10 → BKK 19:30
VietJet Air
1h 20m | Direct

Return on Sun, 18 Oct
BKK 19:20 → CNX 20:40
VietJet Air
1h 20m | Direct

We have selected the cheapest flight to give you the best bundle deal, feel free to change timings, airlines, stops and duration for flight.

Hurry! 53% of properties on our site are fully booked!
Rooms in Bangkok are in high demand on your selected dates. Reserve yours now before prices go up.

SORT: BEST MATCH | BUNDLE SAVINGS | LOWEST PRICE FIRST | DISTANCE | TOP REVIEWED

These properties were selected just for you.

Agoda Preferred Property

Best seller | Frequently re-booked by guests | Agoda PREFERRED

Jasmine City Hotel
★★★★★ | Sukhumvit, Bangkok - View on map
Exceptional location | Convenient city access
Breakfast
Recommended by 93% of guests
AGODASPONSORED | Coupon Code AGODASPONSORE...
BUNDLE SAVINGS
฿ 8,637
฿ 15,014
฿ 6,377
Flight + Hotel
Per person + 3 nights
Select Room

Agoda Preferred Property

Agoda EasyCancel | Agoda PREFERRED

Al Meroz Hotel Bangkok- The Leading Halal Hotel
★★★★★ | Ratchada, Bangkok - View on map
Excellent 8.6 (1,105 reviews)
9.1 cleanliness

ข้อมูลเที่ยวบิน
เชียงใหม่ – กรุงเทพฯ

ข้อมูลห้องพักในกรุงเทพฯ

ข้อมูลภาพนิ่ง ข้อความ ข้อมูลสถิติ
ข้อมูลรูปแบบแผนที่

Price per person: ฿ 0 - ฿ 6,473,568

Star rating: ★★★★★ (444)

Location rating: 9+ Exceptional location (666)

Search results for: What are you looking for?

Agoda Preferred Property

Jasmine City Hotel
★★★★★ | Excellent (7148 reviews)
฿ 6,377

Agoda Preferred Property

Al Meroz Hotel Bangkok- The Leading Halal Hotel
★★★★★ | Excellent (1105 reviews)
฿ 4,909

Agoda Preferred Property

Vic3 Bangkok

ตย. คุณลักษณะ Big data : Veracity

- คุณภาพหรือความน่าเชื่อถือของข้อมูล
- ข้อมูลที่ไม่มีอคติ ไม่มีBugs Noise ข้อมูลมีความปลอดภัย แหล่งที่มา ข้อมูล ความไม่แน่นอน Out of Date หรือเกิดจาก Human error

ตย. คุณลักษณะ Big data : Value

Usefulness of data for an enterprise

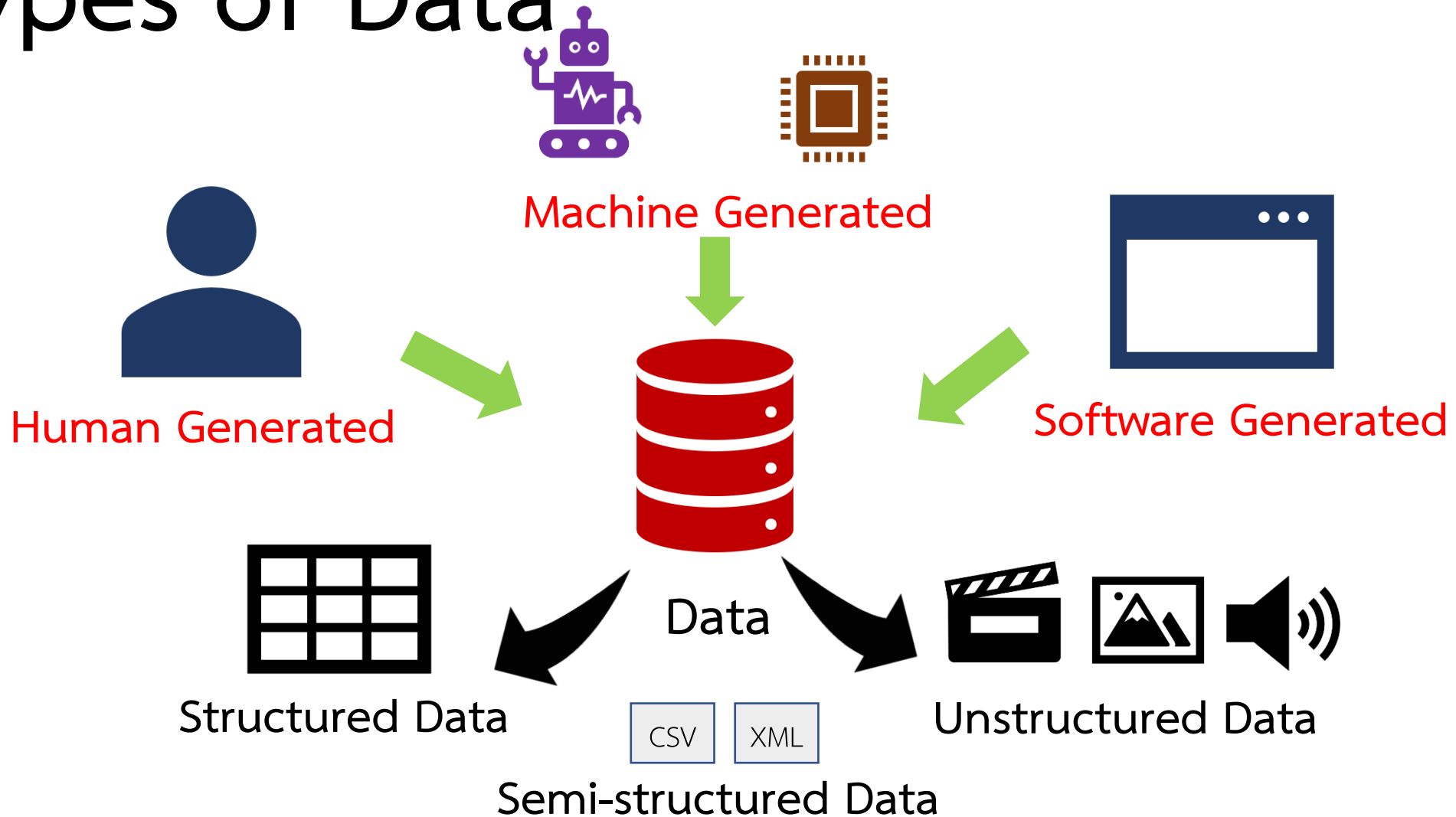
Valuable Data


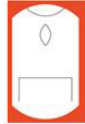
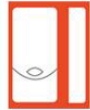


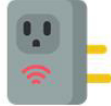











































- Customer Lifetime Value
- Average Order Value
- Cancellation Rate

Data With No Value

- Data with missing or corrupt values
- Data missing key structured elements such as customer reference or date

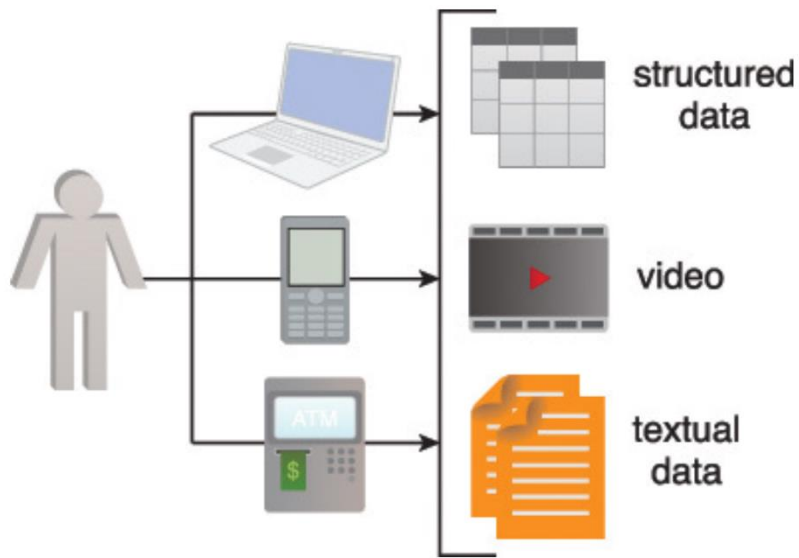
Types of Data



 Connected building	 Motion sensor	 Door sensor	 Button	 Bed	 Plug	 Light	 Medical	 Camera	 Central
 Fibaro flood sensor  Intervox smoke detector  Sensing labs humidity sensor  Sensing labs temperature sensor  FHE Sensor Pilot	 Tunstall  Beewi  Eldes  Intervox  Fibaro  FHE	 Tunstall  Beewi  Eldes  Intervox  Fibaro  FHE	 Tunstall  Intervox  Vitalbase Fall detector  Vivago watch	 Emfit  Telehealth	 MyPlug - Orange  Smartplug - Beewi  Otio  FHE Smart Plug	 Intervox	 Bewell glucose meter  Bewell blood pressure monitor  Bewell oximeter  Bewell thermometer  Bewell electro-cardiogram  Bathroom scale  Connected watch	 Vivotek  Kinect	 Tunstall  Intervox

Sensor data:

Human-generated data



- แหล่งสร้างข้อมูล เช่น โซเชียลมีเดีย Blog posts อีเมลล์ ภาพข้อความที่แชร์กัน

ตัวอย่าง

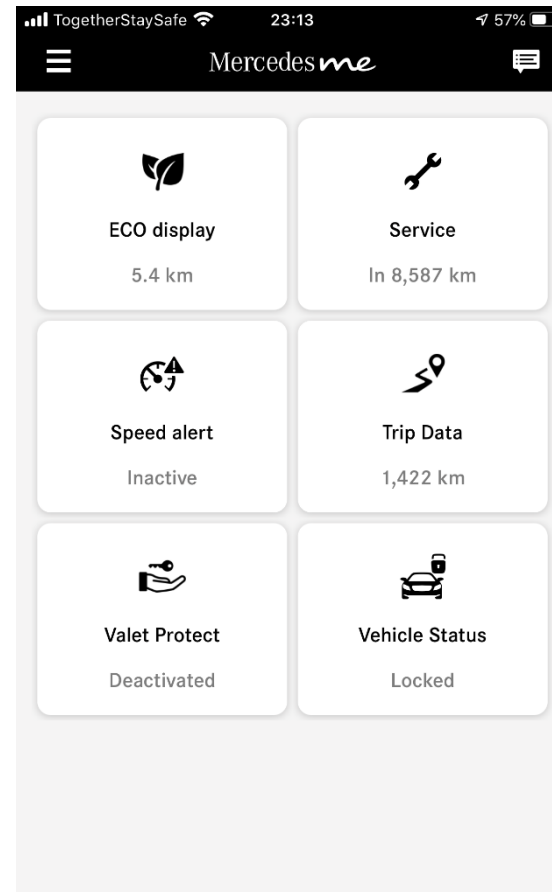
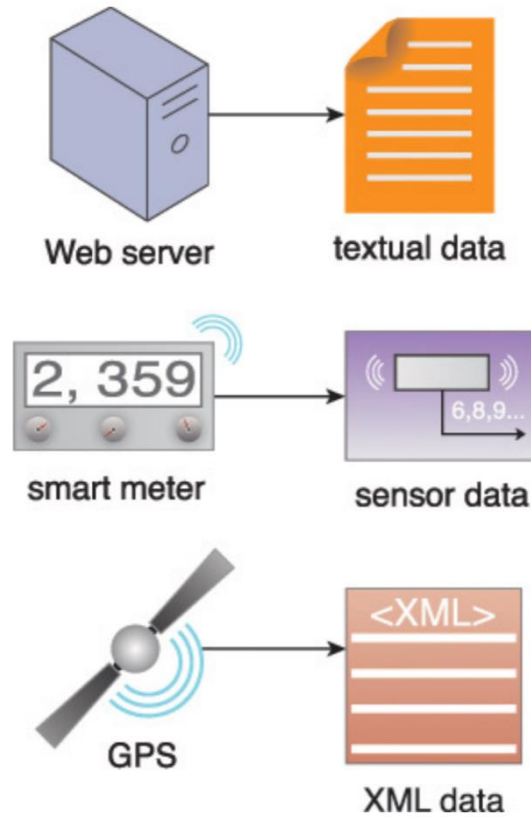
- ผู้ประกอบการ 1 ราย Live จำหน่ายสินค้า มียอดวิว 210 K มีคนมาเมนต์ 709 ข้อความ ยอดแชร์จากคลิปนี้ 276 ครั้ง
- คุณเคยนับหรือไม่ว่าสร้างข้อมูลในช่องทางต่าง ๆ เท่าไหร่ อย่างไร



ภาพจาก: Erl, T., Khattak, W. and Buhler, P. (2016). Big Data Fundamentals: Concepts, Divers and Techniques

Machine-generated data

แหล่งข้อมูลจาก web logs ข้อมูลจากเซ็นเซอร์ ข้อมูลจากเครื่องใช้ ตู้เย็น Smart meter และอื่น ๆ




ภาพจาก: Erl, T., Khattak, W. and Buhler, P. (2016). Big Data Fundamentals: Concepts, Divers and Techniques

Structured Data Example

📶 สินค้า + สินค้าใหม่

🔍 นมดื่ม

ผลการค้นหา 2 รายการ แสดง ทั้งหมด [เปิดใช้งาน](#)




นมดื่มเกษตรมช. ขนาด 1 ลิตร (Pre-Order 2 วัน)

🛒 ของดื่มมช. cmu #CMUCarelicious #อาหารปลอดภัย #ข้าวพิภพไม่ปลอดสารพิษ #สินค้าปลอดภัย

#ของดื่มมช. #นมCMU

ราคา ฿65.00



นมดื่มเกษตร มช.

🛒 ของดื่มมช. cmu #CMUCarelicious #อาหารปลอดภัย #ข้าวพิภพไม่ปลอดสารพิษ #สินค้าปลอดภัย

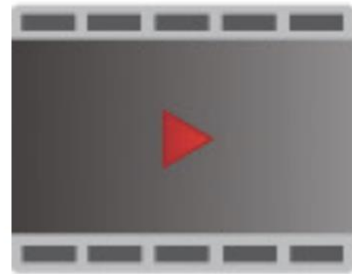
#ของดื่มมช. #นมCMU

ราคา ฿20.00

- E-Commerce ประกอบด้วย
- Product Catalog Systems
 - Member System
 - Shopping Cart

ภาพจาก: cmuselectedproducts.com

Unstructured Data Example



video



image
files



audio

ภาพจาก: Erl, T., Khattak, W. and Buhler, P. (2016). Big Data Fundamentals: Concepts, Divers and Techniques

Semi-structured Data Example



ตัวอย่าง

- E-mails
- XML and other markup languages
- Binary executables
- TCP/IP packets
- Zipped files
- Integration of data from different sources
- Web pages

ภาพจาก: Erl, T., Khattak, W. and Buhler, P. (2016). Big Data Fundamentals: Concepts, Divers and Techniques

Advantages of Semi-structured Data

- The data is not constrained by a fixed schema
- Flexible i.e Schema can be easily changed.
- Data is portable
- It is possible to view structured data as semi-structured data
- Its supports users who can not express their need in SQL
- It can deal easily with the heterogeneity of sources.

Difference between Structured, Semi-structured and Unstructured data

ข้อมูลมีโครงสร้าง	ข้อมูลกึ่งโครงสร้าง	ข้อมูลไม่มีโครงสร้าง
ข้อมูลจัดเก็บในฐานข้อมูล แบบ Database SQL รูปแบบตารางข้อมูล (แถวและคอลัมน์)	ไม่ขึ้นหรือผูกติดกับฐานข้อมูลใด ฐานข้อมูลหนึ่ง	ข้อมูลที่ไม่มีโครงสร้างเป็นข้อมูลที่ไม่ได้จัดระเบียบตามลักษณะที่กำหนดไว้ล่วงหน้า
ออกแบบรูปแบบข้อมูลที่จัดเก็บแต่ละ คอลัมน์	โครงสร้าง และคุณสมบัติเหมาะกับการนำไปวิเคราะห์ต่างแพลตฟอร์ม	ไม่มีโมเดลข้อมูลที่กำหนดไว้ล่วงหน้า
ระบบฐานข้อมูลเชิงสัมพันธ์	ข้อมูลที่ไม่ได้อยู่ในฐานข้อมูลเชิงสัมพันธ์ แต่มีคุณสมบัติขององค์กรบางประการที่ช่วยให้วิเคราะห์ได้ง่ายขึ้น	ไม่เหมาะสำหรับฐานข้อมูลเชิงสัมพันธ์ กระแสหลัก

ในองค์กรของคุณมีข้อมูลแบบนี้ไหม

จับผิดข้อมูลจากสองภาพนี้

GUEST DETAILS



Flight 1: Chiang Mai → Penang	
Chiang Mai to Kuala Lumpur	
Ms Sirikorn Santirojanakul	• Checked baggage 20kg
Mr Phusit Kornsurin	• No Add-ons
Kuala Lumpur to Penang	
Ms Sirikorn Santirojanakul	• Checked baggage 20kg
Mr Phusit Kornsurin	• No Add-ons
Flight 2: Penang → Chiang Mai	
Penang to Bangkok - Don Mueang	
Ms Sirikorn Santirojanakul	• Checked baggage 20kg
Mr Phusit Kornsurin	• No Add-ons
Bangkok - Don Mueang to Chiang Mai	
Ms Sirikorn Santirojanakul	• Checked baggage 20kg



AIRASIA BERHAD (284669-W)
RedQ, Jalan Pekeliling 5,
Lapangan Terbang Antarabangsa Kuala Lumpur (KLIA2)
64000 KLIA, Selangor Darul Ehsan
Malaysia

(GST Reg No: 001062207488)

Tax Invoice Number: AQTM6K_AK1804030002952
Printed Date: 4 Apr, 2018

TAX INVOICE

MR Sirikorn Santirojanakul
, Thailand
Mobile Phone: 66635395159

Date of Invoice: 3/4/2018 (02:57:50 PM)

No.	Description	Total Excl. GST	GST @ 6%
	FARE	THB	THB
1	2x Guests (CNX-PEN) **	4250.68	0.00
2	2x Guests (PEN-CNX) **	5190.46	0.00
3	4x Advance Passenger Processing Service **	140.00	0.00
4	4x International Arrival and Departure Fees (Thailand) **	60.00	0.00
5	2x Excise Tax (Fuel) *	300.00	0.00
6	2x Passenger Service Charge *	560.00	0.00

Big Data Ecosystem

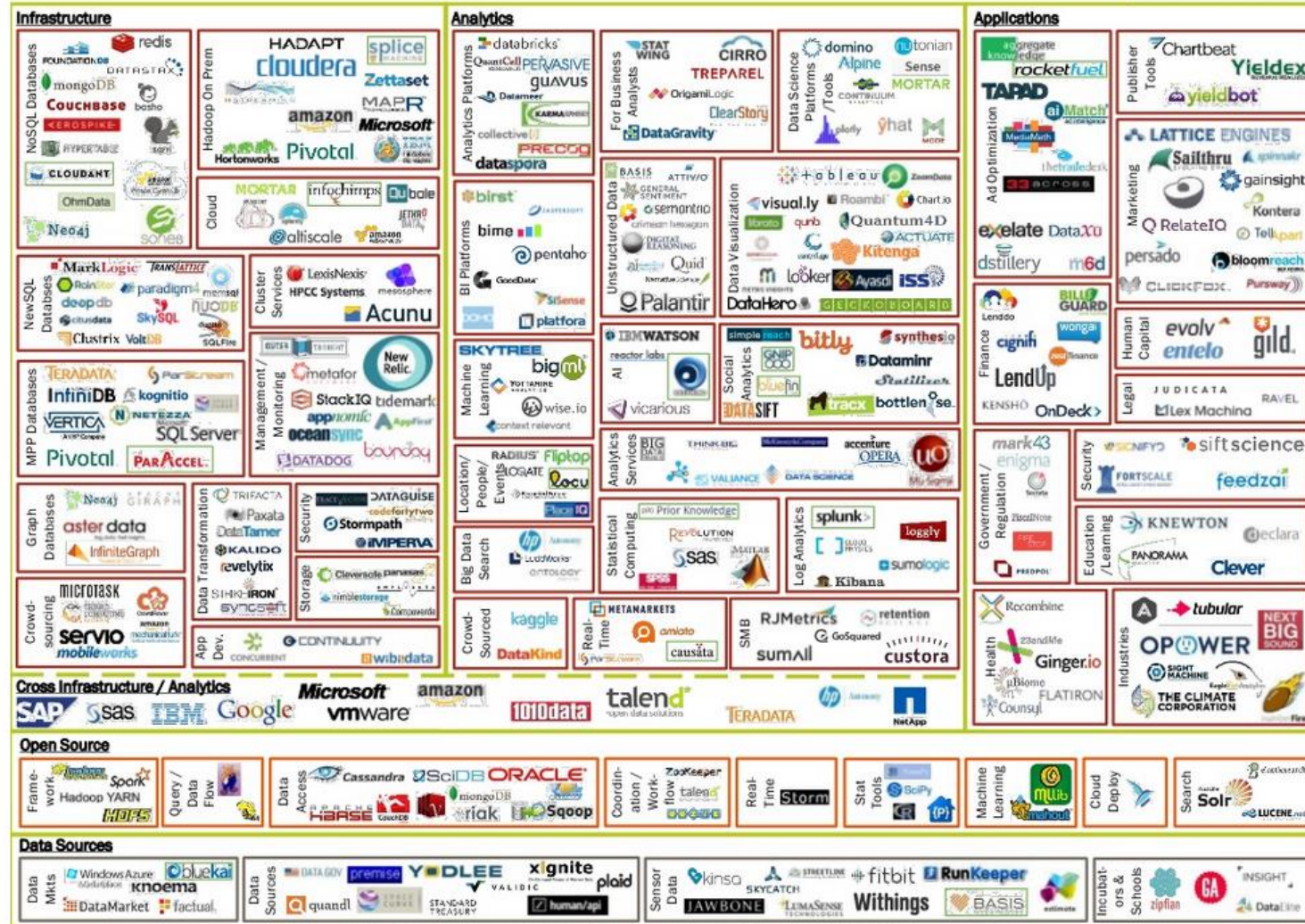
- การรวบรวมกระบวนการวิเคราะห์และรายงานข้อมูล
- โครงสร้างพื้นฐานและเครื่องมือสำหรับกระบวนการ / การวิเคราะห์
- รายงาน

กระบวนการ Big Data

1. Ingest (การจัดเก็บรวบรวมข้อมูล)
2. Process (การประมวลผล)
3. Publish (การเผยแพร่)
4. Report (รายงาน)

BIG DATA LANDSCAPE, VERSION 3.0

Exited: Acquisition or IPO



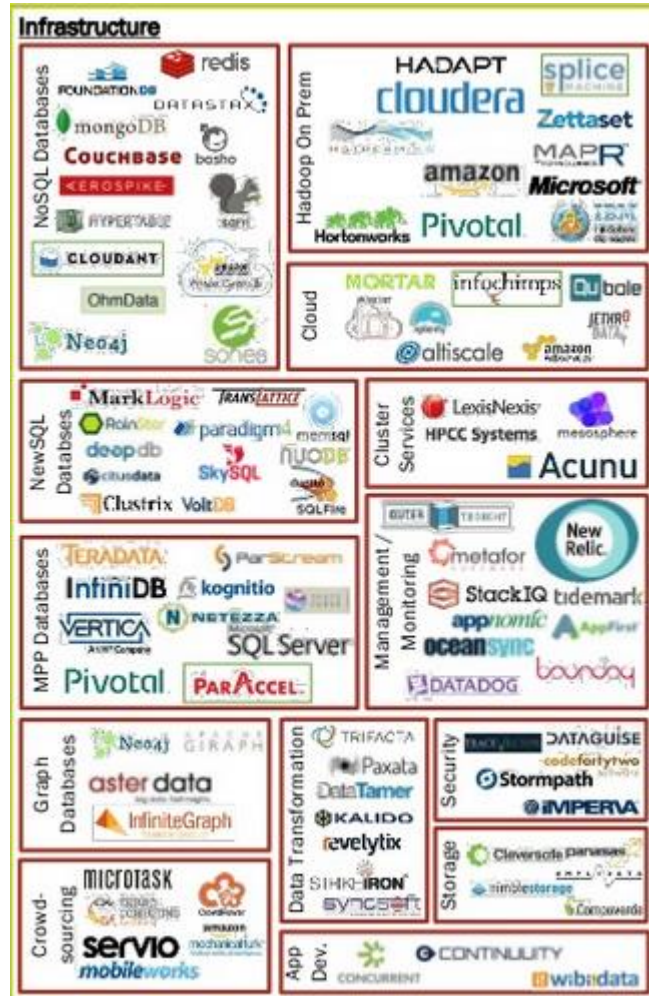
© Matt Turck (@mattturck), Sutian Dong (@sutiandong) & FirstMark Capital (@firstmarkcap)

INFRASTRUCTURE



- โครงสร้างพื้นฐานเป็นส่วนสำคัญของระบบนิเวศข้อมูลขนาดใหญ่
- ในการประมวลผลจัดเก็บและวิเคราะห์ข้อมูล ในอดีตจะใช้ฐานข้อมูลเชิงสัมพันธ์
- เมื่อปริมาณข้อมูล ความเร็วและข้อมูลที่มีความหลากหลายมากขึ้น และในการจัดการข้อมูลขนาดใหญ่ มีความซับซ้อน การเพิ่มขึ้นของข้อมูลที่ไม่มีโครงสร้าง ทำให้การจัดเก็บในฐานข้อมูลเชิงสัมพันธ์ขาดประสิทธิภาพ
- ทำให้ต้องใช้เทคโนโลยีโครงสร้างพื้นฐานแบบใหม่ เพื่อรองรับข้อมูลรูปแบบใหม่และข้อมูลขนาดหลายพันเทราไบต์

INFRASTRUCTURE



- ตัวอย่างของเทคโนโลยีของ Infrastructure
- Hadoop – เป็นเทคโนโลยีทั้งหมดของระบบ ที่ออกแบบมา เพื่อจัดเก็บ ประมวลผล และวิเคราะห์ข้อมูล โดยหลักการสำคัญของเทคโนโลยี Hadoop คือ การทำงานบนหลักการของแยกและกระจายข้อมูลเป็นส่วนๆ
- NoSQL – ใช้สำหรับทำงานในการประมวลผลข้อมูลจำนวนมาก ๆ ของข้อมูล ที่มีโครงสร้างหลากหลาย โดยฐานข้อมูล NoSQL ส่วนใหญ่นั้นจะทำงานกับ ข้อมูลที่ไม่ต่อเนื่อง กับข้อมูลที่มีโครงสร้างหลากหลาย และบางตัวยังสามารถ ทำงานกับ Hadoop ได้อีกด้วย
- Massively Parallel Processing (MPP) Databases – ฐานข้อมูล MPP ทำ กงานกับข้อมูลที่เป็นส่วน ๆ ข้ามโหนดต่าง ๆ และประมวลผลข้อมูลแบบ ขนาน และใช้ SQL โดย MPP จะทำงานบนฮาร์ดแวร์ที่ราคาแพง

Analytics/Process



- แม้ว่าเทคโนโลยีโครงสร้างพื้นฐานจะรวมการวิเคราะห์ข้อมูล แต่ก็มีเทคโนโลยีเฉพาะที่ออกแบบมาพร้อมความสามารถในการวิเคราะห์ข้อมูล โดยจะมีส่วนย่อยของงานในส่วนนี้ดังนี้
- Analytics Platforms – ทำหน้าที่รวบรวมและวิเคราะห์ข้อมูลไปเป็นรายงาน เพื่อช่วยในการตัดสินใจ เน้นเฉพาะเรื่องที่เราสนใจและนำเสนอข้อมูลเชิงลึกเพื่อให้ผู้ใช้ปลายทางสามารถใช้งานได้อย่างทันที่ว่างที่สุด
- Visualization Platform – เพื่อออกแบบนำเสนอในรูปแบบที่ซับซ้อน จากข้อมูลดิบ เพื่ออธิบายข้อมูลได้
- Business Intelligence (BI) Platforms – ใช้การบูรณาการและวิเคราะห์ข้อมูลทางธุรกิจ ซึ่ง BI Platform ใช้วิเคราะห์ข้อมูลจากหลายแหล่ง เพื่อนำส่งบริการ ทั้ง Business Intelligence Report, Dashboard และ Visualization
- Machine Learning - เป็นส่วนของการเรียนรู้จากข้อมูล ความสัมพันธ์ต่าง ๆ ของข้อมูลที่เกิดขึ้น ด้วยตัวเอง เพื่อใช้ในการตอบคำถามต่าง ๆ ในอนาคต

Application



- Application เป็นเครื่องมือที่จะช่วยให้ผู้ใช้งาน สามารถวิเคราะห์ Big Data เพื่อให้ได้มาซึ่งข้อมูลเชิงลึก ซึ่งจะมีตัวอย่าง Application ในด้านต่าง ๆ ดังนี้
- ด้านสุขภาพ Application Minitabs ที่สรุปของการสแกนสมองแบบ 3 มิติและข้อมูลเกี่ยวกับระบบประสาทที่สามารถเข้าถึงได้โดยนักประสาทวิทยาจากทั่วโลกและช่วยในการวินิจฉัยการพยากรณ์โรคและการรักษาผู้ป่วยโรคสมองเสื่อม
- ด้านค้าปลีก Avansera สามารถเอาแอปพลิเคชันการช้อปปิ้งบนมือถือที่นำเสนอข้อมูลเชิงลึกสำหรับ บริษัท ผู้ผลิตอาหารเป็นตัวแปรที่มีผลต่อการซื้ออาหารเช่นความถี่ของแบรนต์และความยืดหยุ่นของราคา
- ด้านพลังงาน AutoGrid ใช้ข้อมูลจากสมาร์ทเมตรระบบการจัดการอาคารตัวควบคุมแรงดันไฟฟ้าและเครื่องควบคุมอุณหภูมิเพื่อช่วยให้ผู้บริโภคดีดตามและลดการใช้พลังงานลดของเสียสมดุลระบบกริดปรับปรุงการดำเนินงานของระบบและคาดการณ์การบริโภคในอนาคตได้

Big Data: Technology

- Storage และ Analytics

Storage

Analytics

Big Data: Technology

- Storage และ Analytics

Storage

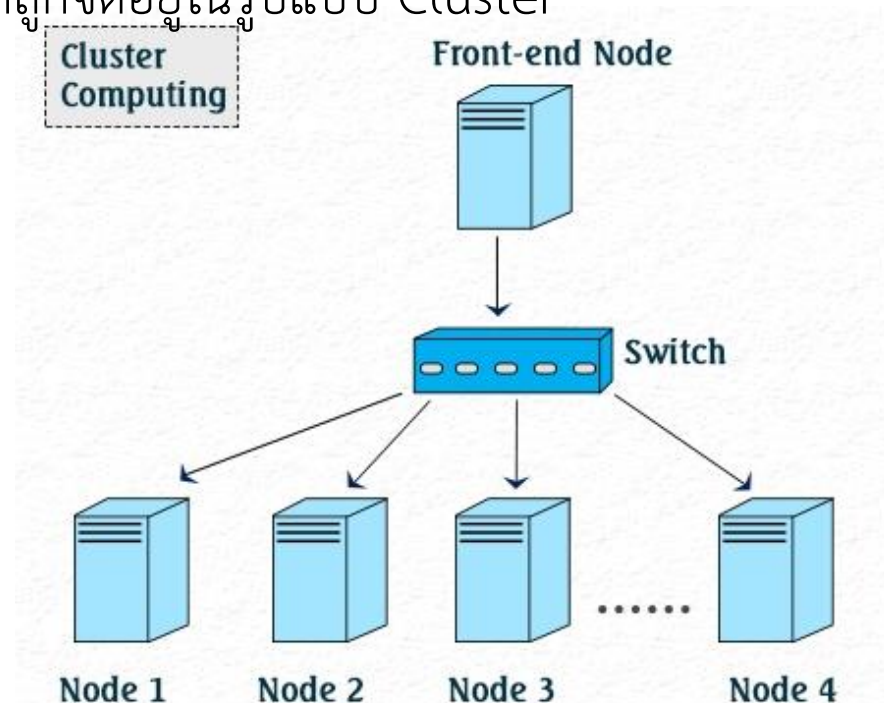
Analytics

เทคโนโลยีสำหรับ Big Data

- Storage
- ฐานข้อมูล RDBMS แบบเดิม ก็ยังเป็นเทคโนโลยีที่เหมาะสมที่สุดในการเก็บข้อมูลแบบ Structure แต่ถ้าข้อมูลมีขนาดใหญ่มากก็จะเจอปัญหาเรื่องต้นทุนที่สูง และหากข้อมูลมีจำนวนเป็น PetaByte ก็คงยากที่จะเก็บ ถึงแม้ในปัจจุบันจะมี MPP Database อย่าง Oracle ExaDta หรือ SAP HANA แต่มีค่าใช้จ่ายสูง
- Hadoop HDFS เป็นเทคโนโลยีที่มีเริ่มใช้งาน หน่วยงานส่วนใหญ่เริ่มวางแผนจะต้องใช้ในอนาคต เพราะมีความต้องการเก็บข้อมูลขนาดใหญ่ทั้งที่เป็น Unstructure Data หรือนำข้อมูลที่เป็น structure มาเก็บไว้ โดยสามารถจะเก็บข้อมูลได้เป็น PetaByte ทั้งนี้ขึ้นอยู่กับจำนวนเครื่องที่มีอยู่ในลักษณะ scale-out ที่สำคัญ Hadoop มีต้นทุนที่ค่อนข้างต่ำเมื่อเทียบกับเทคโนโลยีการเก็บข้อมูลแบบอื่น

Hadoop

- ซอฟต์แวร์ประเภท open source ที่จัดทำขึ้นเพื่อเป็นแพลตฟอร์มในการจัดเก็บข้อมูล ซึ่งมีกรอบการทำงานเพื่อใช้ในการจัดเก็บข้อมูลและประมวลผลข้อมูลที่มีขนาดใหญ่ สามารถปรับขยาย ยืดหยุ่น เพื่อรองรับข้อมูลที่มีจำนวนมากมายมหาศาลได้ ทั้งนี้ก็เพราะมันมีการกระบวนการประมวลผลที่แข็งแกร่งมากซึ่งเป็นผลมาจากการประมวลผลข้อมูลแบบกระจายผ่านเครื่องคอมพิวเตอร์ที่ถูกจัดอยู่ในรูปแบบ Cluster



เทคโนโลยีสำหรับ Big Data

- Storage
- NoSQL เป็นเทคโนโลยีที่ต้องการเก็บข้อมูลจำนวนมากกว่าของ RDBMS ในลักษณะ scale-out เป็นจำนวนหลาย TeraByte แต่อาจไม่ได้เน้นเรื่องความมั่นคงของข้อมูลมากนัก เหมาะกับ Application บางประเภท ทั้งนี้เราสามารถจะแบ่งเทคโนโลยี NoSQL ออกไปได้สี่กลุ่มคือ Column Oriented, Document Oriented, Key-Value และ Graph
- Cloud Storage ข้อมูลขนาดใหญ่ขององค์กรบางส่วนอาจต้องเก็บไว้ใน Public Cloud Storage เช่น Amazon S3 โดยเฉพาะข้อมูลภายนอกอาทิเช่น Social Media Data หรือข้อมูลที่เป็น Archiving ที่ไม่ได้มีความสำคัญมาก เพราะ Cloud Storage จะมีราคาในการเก็บที่ถูกสุด และสามารถที่จะเก็บได้โดยมีขนาดไม่จำกัด แต่ข้อเสียคือเรื่องความปลอดภัยและความเร็วในการถ่ายโอนข้อมูล

เทคโนโลยีสำหรับ Big Data

- การประมวลผลข้อมูลที่เป็น Big Data จะมีทั้งการวิเคราะห์ข้อมูลที่เป็น business intelligence (BI) เพื่อที่จะดึงข้อมูลมานำเสนอ หรือการทำ Predictive Analytics
- SQL ก็เป็นภาษาที่ยังต้องใช้ในการประมวลผลข้อมูลโดยเฉพาะ Structure Data ที่เก็บอยู่ใน RDBMS และสามารถประมวลผลแบบ RealTime ได้
- APIs ข้อมูลที่เก็บอยู่ใน Storage ต่าง ๆ ที่กล่าวมาข้างต้นเช่น NoSQL หรือ Cloud Storage อาจต้องพัฒนาโปรแกรมด้วยภาษาคอมพิวเตอร์ต่าง ๆ ในการประมวลผลข้อมูลโดยใช้ APIs ในการเข้าถึงข้อมูล

เทคโนโลยีสำหรับ Big Data

- MapReduce เป็นเทคโนโลยีที่พัฒนาโดย Google ในการประมวลผลข้อมูลที่อยู่ใน HDFS โดยใช้ภาษาคอมพิวเตอร์อย่าง Java ในการพัฒนาโปรแกรม โดยจะประมวลผลแบบ Batch และเป็นวิธีการประมวลผลที่มากับเทคโนโลยี Hadoop
- Hive หรือ Pig เป็นภาษาคัดลอก SQL หรือ Scripting ที่ทำให้เราสามารถประมวลผลข้อมูลที่อยู่ใน Hadoop HDFS ได้โดยไม่ต้องพัฒนาโปรแกรม MapReduce แต่ทั้งนี้ข้อมูลจะต้องอยู่ในรูปแบบที่เหมาะสมเช่น ไฟล์ csv หรือ ไฟล์ข้อความบางประเภท

เทคโนโลยีสำหรับ Big Data

- Impala เป็นภาษาคัดลอก SQL ที่ทำให้เราสามารถประมวลผลข้อมูลที่อยู่ใน Hadoop HDFS ได้ โดยทำงานได้รวดเร็วกว่า Hive มาก แต่มีข้อเสียคือเป็นภาษาที่เป็น proprietary ของ Cloudera
- Spark เป็นเทคโนโลยีที่สามารถประมวลผลข้อมูลขนาดใหญ่แบบ Real-time โดยอาจมี Data Source มาจากหลากหลายแหล่งเช่น RDBMS, Cloud Storage, NoSQL หรือ Hadoop ซึ่งสามารถเขียนโปรแกรมโดยใช้ภาษา Scala, Java, Python หรือจะเขียนโดยใช้ภาษาคัดลอก SQL ก็ได้ และมี Library สำหรับการทำให้ Data Science คือ MLlib เป็นเทคโนโลยีที่น่าสนใจมากอันหนึ่ง

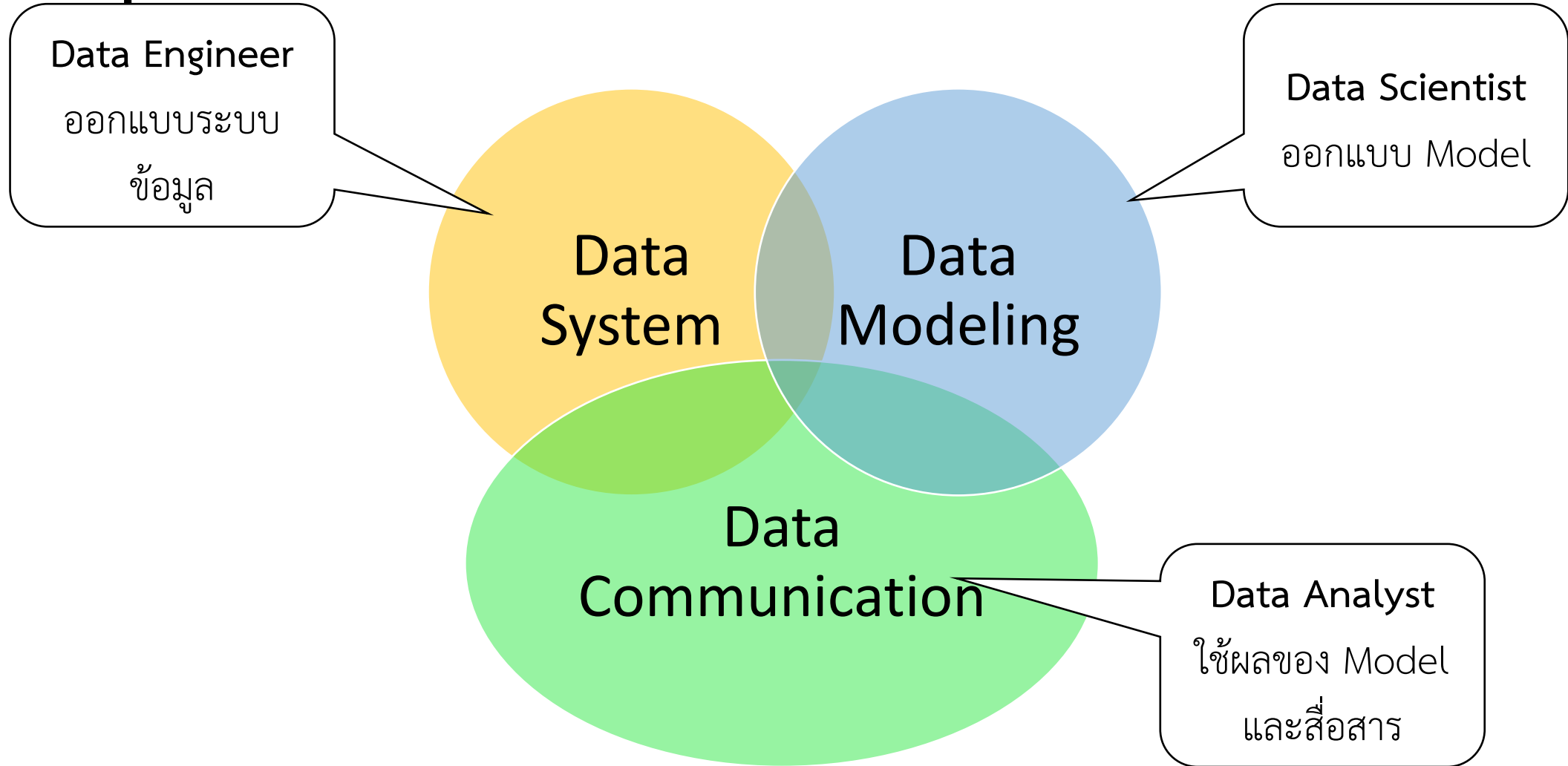
เทคโนโลยีสำหรับ Big Data

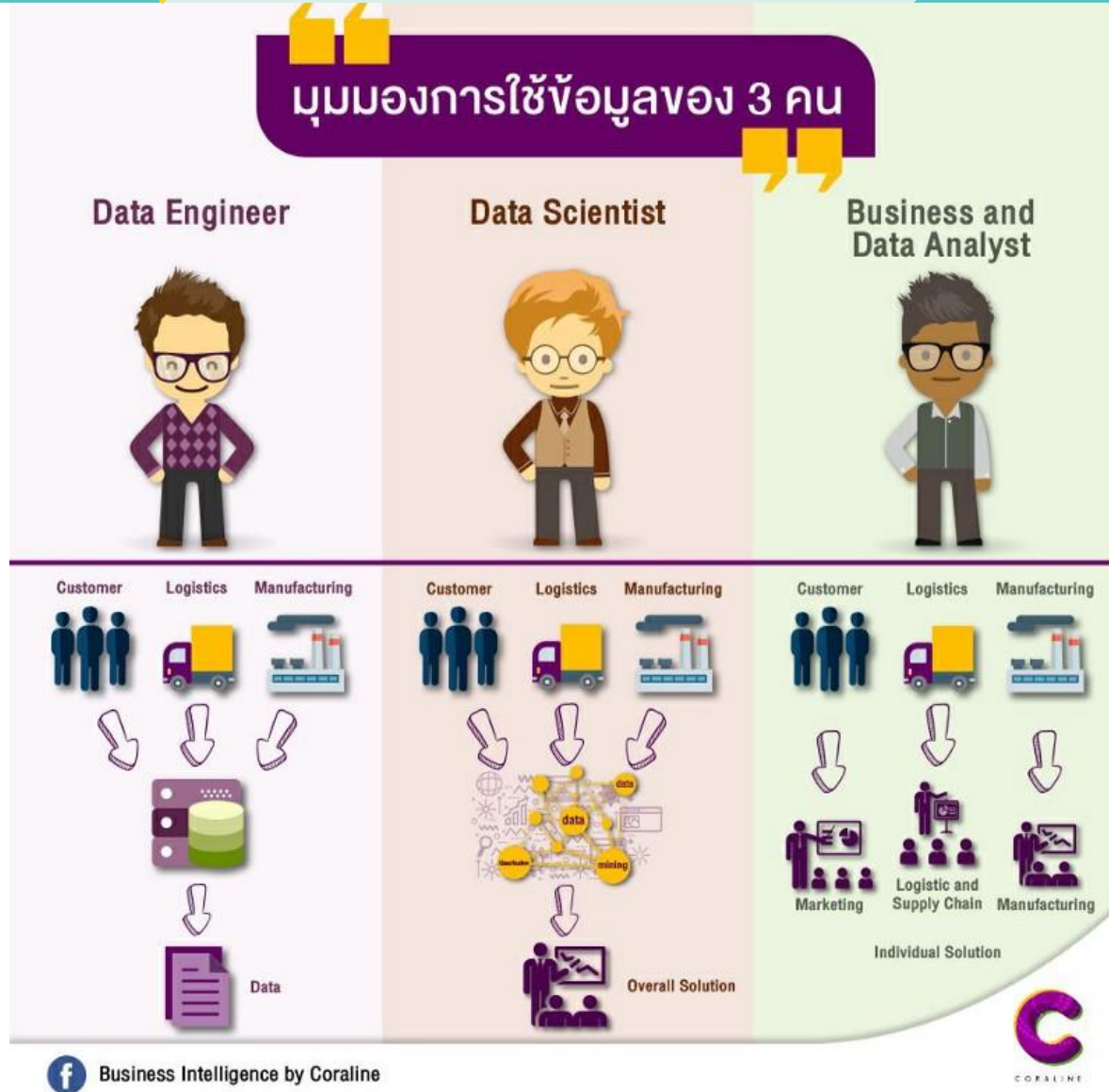
- ภาษาและเทคโนโลยีในการทำ Machine Learning ซึ่งก็จะมีหลากหลายทั้ง R Hadoop, Mahout, Azure Machine Learning หรือ AWS ML
- เทคโนโลยีสำหรับการทำ Data Visualisation และ BI อาทิเช่น Tableau, Pentaho, SaS, Excel และอื่น ๆ

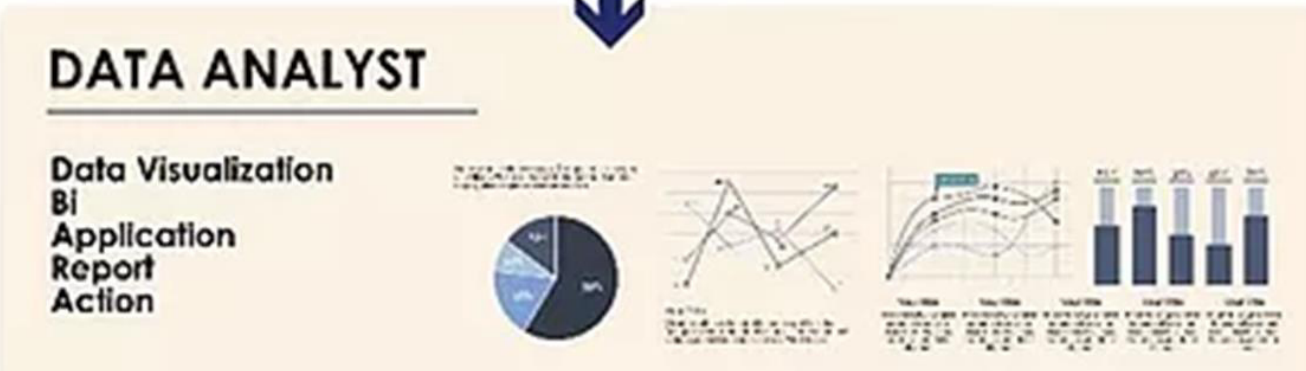
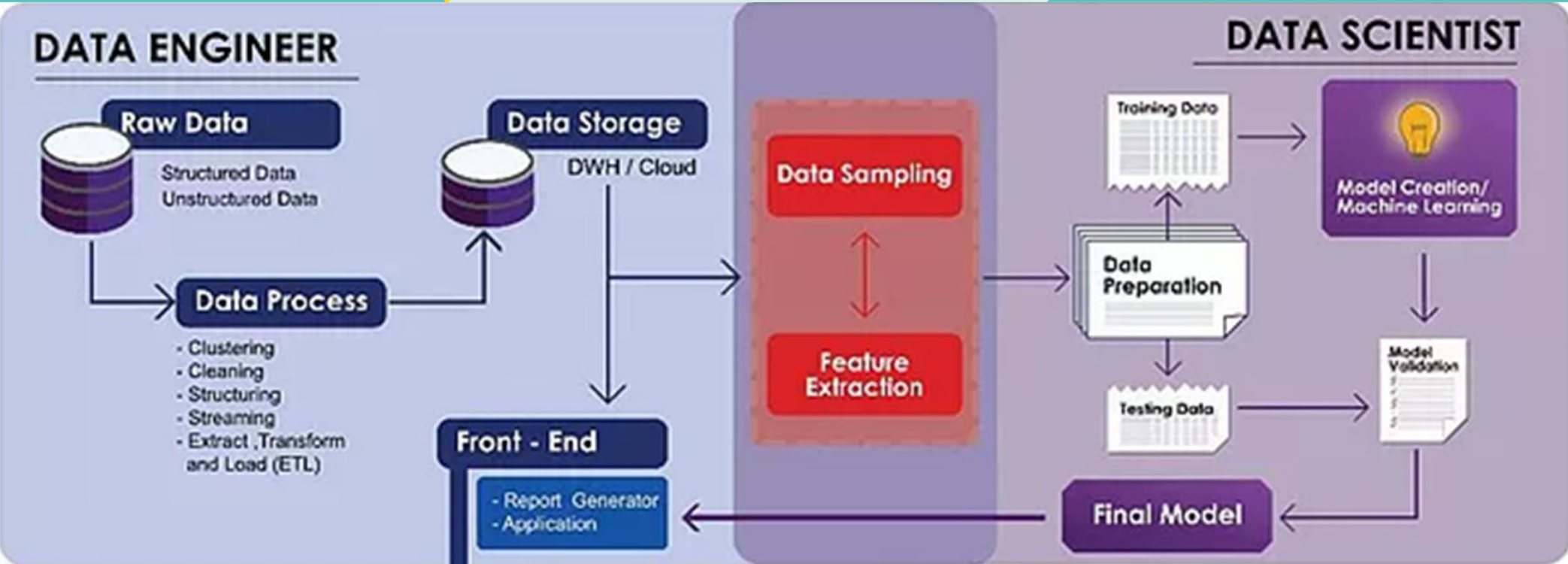
บุคลากรในการจัดการข้อมูลขนาดใหญ่

- ในการทำงานที่เกี่ยวข้องกับข้อมูลขนาดใหญ่ หรือการสร้างระบบนิเวศน์ของข้อมูลขนาดใหญ่ ประกอบไปด้วย 3 ผู้เชี่ยวชาญหลัก (Data Expert) 3 กลุ่มดังนี้
 1. Data System กลุ่มงานนี้จะมี Data Engineer ทำหน้าที่ในการออกแบบระบบการไหลของข้อมูล การจัดโครงสร้าง การจัดเก็บข้อมูล และช่องทางการนำผลลัพธ์ที่ได้จากการประมวลผลไปแสดงผล
 2. Data Modeling กลุ่มงานนี้จะมี Data Scientist ทำหน้าที่สร้างแบบจำลอง (Model) ที่เหมาะสม โดยการนำข้อมูลขนาดใหญ่เป็นต้นแบบ และมองหาผลลัพธ์เชิงลึก (Insight) จากการสร้าง Data Model
 3. Data Communication กลุ่มงานนี้จะมี Data Analyst ทำหน้าที่ใช้ข้อมูลผลลัพธ์ที่ได้จาก Model ไปใช้งานและสื่อสารต่อ และยังเป็นผู้ประเมินปัญหา และประเมินผลลัพธ์จาก Model

People







Career Path ของ Data Expert

1. **Data Engineer** -> Chief Technology Officer (CTO) หรือ ผู้บริหารระดับสูงด้านเทคโนโลยี
2. **Data Scientist** -> Specialist -> Chief Information Officer (CIO) หรือ ผู้บริหารเทคโนโลยีสารสนเทศระดับสูง
3. **Data Analyst** -> Business Analyst -> Chief Marketing Officer (CMO)
Chief Financial Officer (CFO)
Chief Operating Officer (COO)
Chief Executive Officer (CEO)

Skills

	Data Analyst	Machine Learning Engineer	Data Engineer	Data Scientist
Programming Tools	Very important	Very important	Very important	Very important
Data Visualization and Communication	Very important	Somewhat important	Somewhat important	Very important
Data Intuition	Somewhat important	Very important	Somewhat important	Very important
Statistics	Somewhat important	Very important	Somewhat important	Very important
Data Wrangling	Not that important	Not that important	Very important	Very important
Machine Learning	Not that important	Very important	Not that important	Very important
Software Engineering	Not that important	Somewhat important	Very important	Somewhat important
Multivariable Calculus and Linear Algebra	Not that important	Very important	Not that important	Somewhat important

Not that important
 Somewhat important
 Very important

ตัวอย่างการนำไปใช้

Big Data for Organization Data Disruption

- ภาคเอกชนและภาครัฐ

กรณีศึกษาตัวอย่างของการใช้ข้อมูลขนาดใหญ่

- จับผู้ร้ายด้วย Big Data:
- Beware ระบบประเมินความเสี่ยงการก่ออาชญากรรมรายบุคคล
- Beware ถูกพัฒนาโดยบริษัท Intrado ด้วยแนวคิดในการรวบรวมประวัติรายบุคคลของประชากรทั้งหมด เข้ามาผูกกับข้อมูลที่มีในฐานข้อมูลต่าง ๆ, ข้อมูลบน Web Site รวมถึง Social Network และทำการประเมินความเสี่ยงในการก่ออาชญากรรมด้วยระบบ Scoring เพื่อให้คะแนนแต่ละคนจากประวัติ, ข้อมูล และความสัมพันธ์ระหว่างบุคคลต่าง ๆ ได้แบบ Real-time ทำให้เจ้าหน้าที่ที่รับเรื่องด่วนผ่าน 911 หรือเจ้าหน้าที่ที่กำลังไปปฏิบัติการสามารถเข้าใจในสถานการณ์และความเสี่ยงได้อย่างรวดเร็วจากการวิเคราะห์ Big Data เหล่านี้

กรณีศึกษาตัวอย่างของการใช้ข้อมูลขนาดใหญ่

96

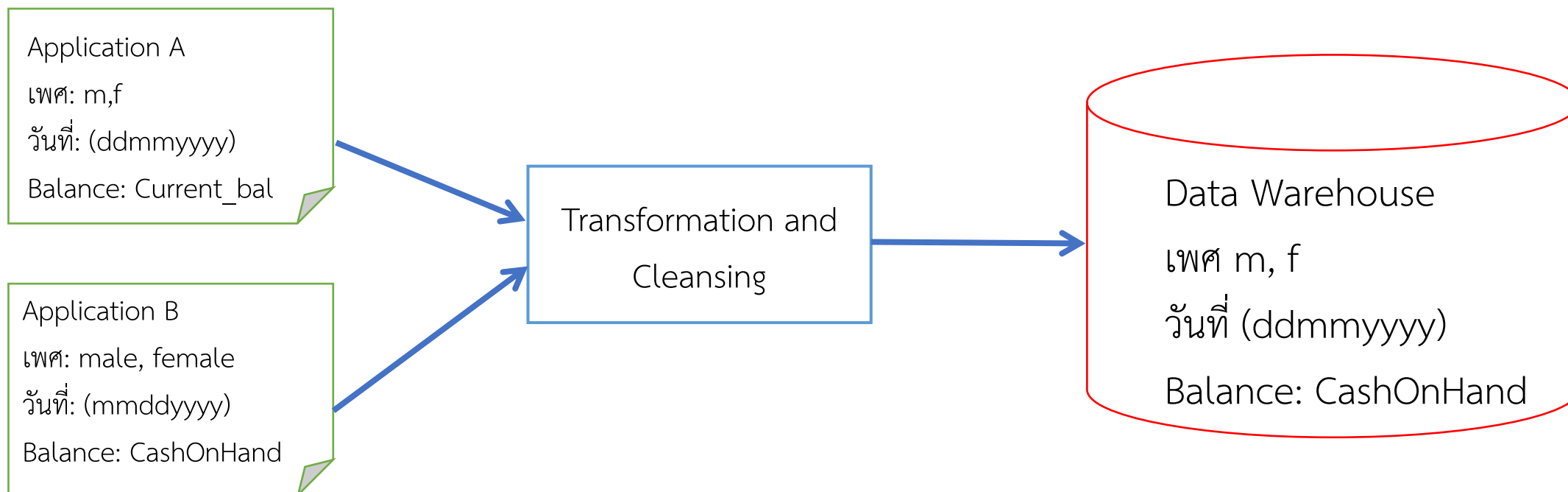
- จับผู้ร้ายด้วย Big Data:
- ShotSpotter ระบบระบุจุดเกิดเหตุยิงปืน
- ด้วยการติดตั้งไมโครโฟนเอาไว้ทั่วเมือง และมีระบบวิเคราะห์เพื่อระบุเสียงปืนที่ได้ยินจากไมโครโฟนแต่ละตัว และทำการค้นหาตำแหน่งจากไมโครโฟนที่ได้ยินเสียงในระดับและเวลาที่แตกต่างกันจากหลายๆ จุดนั้น ก็ทำให้ตำรวจสามารถสืบเสาะหาจุดเกิดเหตุยิงปืนได้อย่างรวดเร็วยิ่งขึ้นได้แล้ว

กรณีศึกษาตัวอย่างของการใช้ข้อมูลขนาดใหญ่

- จับผู้ร้ายด้วย Big Data:
- Media Sonar ระบบวิเคราะห์หาความเสี่ยงการเกิดอาชญากรรมจาก Social Media
- ตำรวจสามารถติดตามข้อความต่าง ๆ ที่ประชาชนโพสต์อยู่บน Social Media เพื่อประเมินความเสี่ยงในการเกิดอาชญากรรมหรือค้นหาเหตุอาชญากรรมที่กำลังเกิดขึ้นได้ทันที โดยเฉพาะเหตุต่าง ๆ ที่อาจเกิดขึ้นกับสถาบันการศึกษา และข้อความจากสมาชิกแก๊งโจรต่าง ๆ ในเมืองได้แบบ Real-time

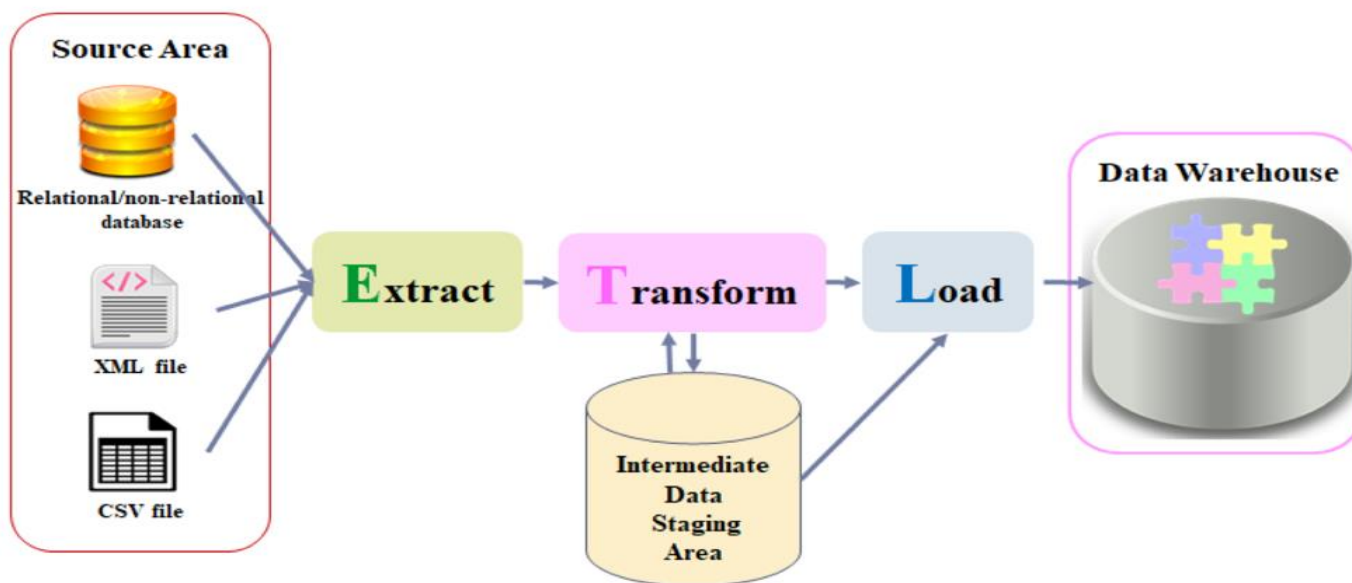
ETL: Extract, Transform, Load

- เป็นการนำข้อมูล (Extract) จากหลากหลายแหล่งข้อมูล เช่น จากฝ่าย Marketing, จากฝ่าย Sale, จากฝ่าย Customer Service ฯลฯ มาแปลงรูปแบบให้เหมาะสม (Transform) แล้วเก็บ (Load) เข้าไปใน Data Warehouse หรือ Data Lake

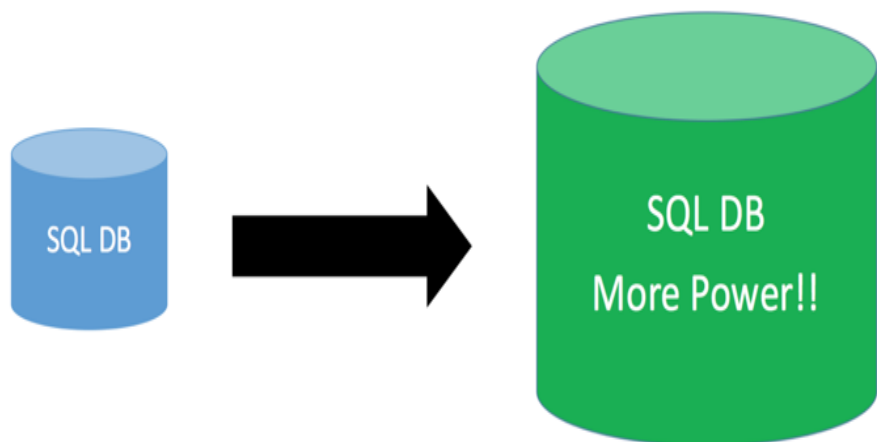


ETL: Extract, Transform, Load

- สามารถเอาข้อมูลชุดนี้ไปให้ทีม Analytics ประมวลผลต่อ หรือเอาไปทำ Dashboard สรุปให้ผู้บริหาร
- หรือนำข้อมูลระบบเก่า (Legacy System) มารวมกับระบบใหม่



Scaling up

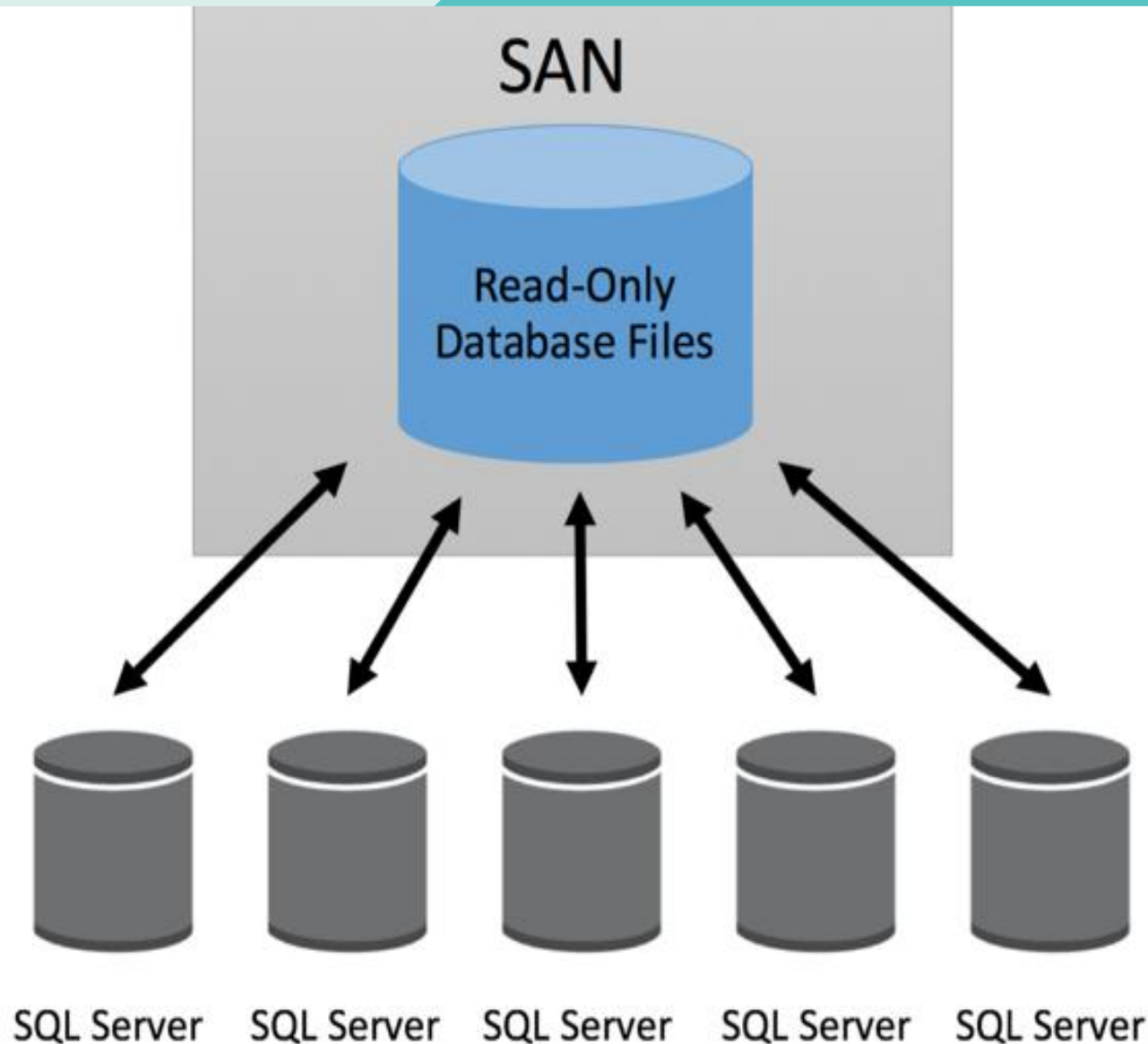


- To scale up the system
- เพิ่มความเร็วการทำงาน เพิ่มประสิทธิภาพ ETL
- เลือกใช้ HDD ที่มีประสิทธิภาพอ่านเขียนข้อมูลเร็ว
- เพิ่มRAM CPU Network เพื่อเพิ่มประสิทธิภาพการประมวลผล

Scaling Out

- Scaling out แนวทางการดำเนินการจะตรงข้ามกับ Scaling up
- จะสร้างระบบที่ใหญ่ขึ้นเป้าหมายคือการกระจายภาระการประมวลผลระหว่างระบบขนาดเล็กจำนวนมากแทน
- วิธีที่ง่ายที่สุดในการปรับขนาดฐานข้อมูล SQL คือการใช้เครือข่ายพื้นที่เก็บข้อมูล (SAN) เพื่อจัดเก็บข้อมูล
- จากนั้นใช้เซิร์ฟเวอร์ SQL ค้นหาข้อมูลโดยใช้ภาษาแบบสอบถาม วิธีนี้จะกระจายโหลดระหว่างเซิร์ฟเวอร์ทั้งหมดที่ทำงานร่วมกัน

Scaling out a SQL Database



What is ELT?

- ELT หมายถึง “extract, load, and transform”
- The processes a data pipeline uses to replicate data from a source system into a target system such as a cloud data warehouse.

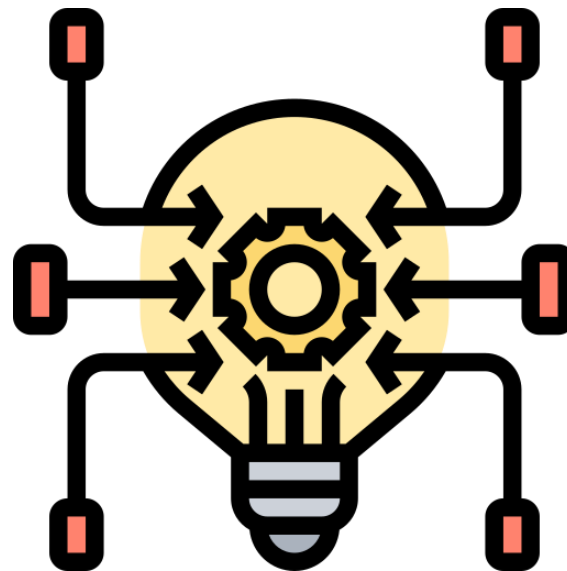
ELT

- **Extraction:** This first step involves copying data from the source system.
- **Loading:** During the loading step, the pipeline replicates data from the source into the target system, which might be a data warehouse or data lake.
- **Transformation:** Once the data is in the target system, organizations can run whatever transformations they need. Often organizations will transform raw data in different ways for use with different tools or business processes.

ELT — the next generation of ETL

- ELT is a modern variation on the older process of extract, transform, and load (ETL), in which transformations take place before the data is loaded.
- Running transformations before the load phase results in a more complex data replication process.

3.2 บทที่ 2 : Business Intelligence



Business Intelligence

8 - 9 กันยายน 2563

ผศ.ดร.สิริกร สันติโรจนกุล

อาจารย์ภูษิต ก้อนสุรินทร์

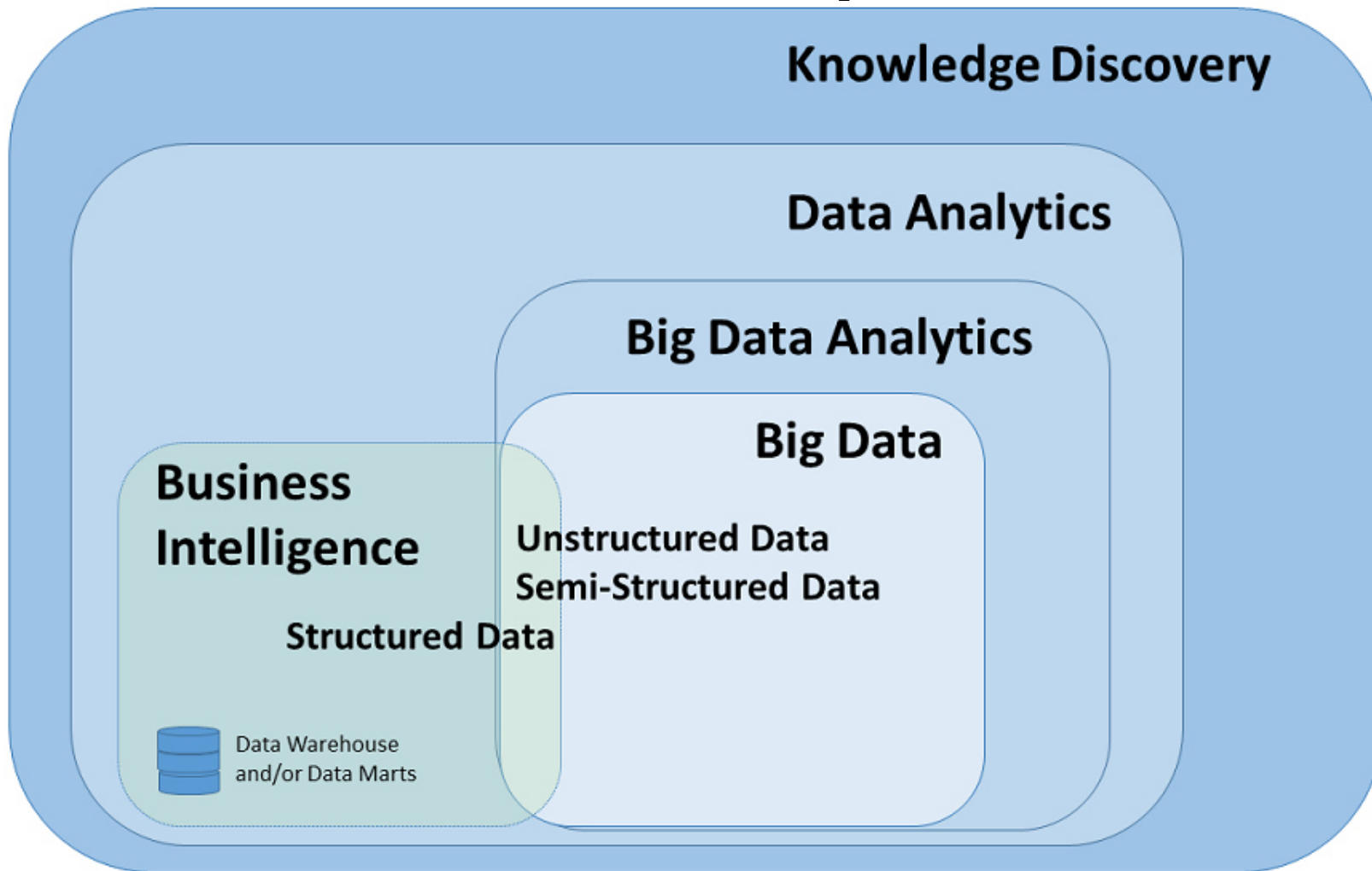
อาจารย์ ดร.เอกสิทธิ์ เทียมแก้ว

Digital Marketing Solution Providers Center

เพจ @digitalSPC



Big Data and Data Analytic

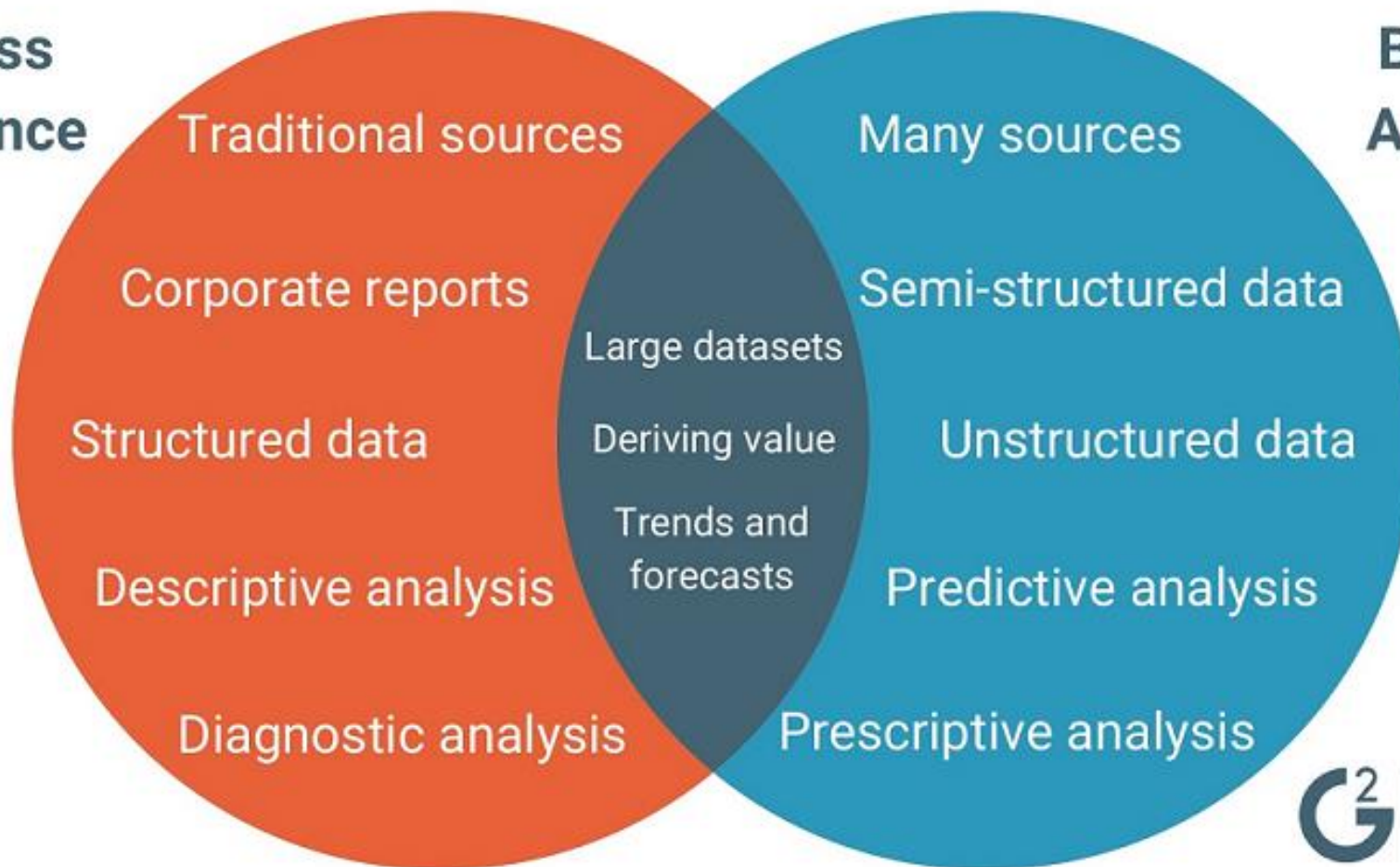


Source: e-salzburg.com

● BI and Big Data

Business Intelligence

Big Data Analytics



Customer Lifetime Value (มูลค่าตลอดชีวิตลูกค้า)

- Big Data & Data Analytics

Customer Segmentation (การแบ่งกลุ่มลูกค้า)

- Big Data & Data Analytics

Up-Sell & Cross-Sell (การขายเพิ่มให้กับลูกค้า)

- Big Data & Data Analytics

Next Best Action (ต่อไปควรทำอะไรที่สุด)

- Big Data & Data Analytics

Propensity to Buy (โอกาสที่จะซื้อ)

- Big Data & Data Analytics

Churn Prevention (ป้องกันการยกเลิก)

- Big Data & Data Analytics

Fraud Detection (สังเกตเหตุผิดปกติ)

- Big Data & Data Analytics

Risk Management (บริหารความเสี่ยง)

- Big Data & Data Analytics

Demand Forecast (พยากรณ์ความต้องการ)

- Big Data & Data Analytics

Price Optimization (ปรับราคาให้เหมาะสมที่สุด)

- Big Data & Data Analytics

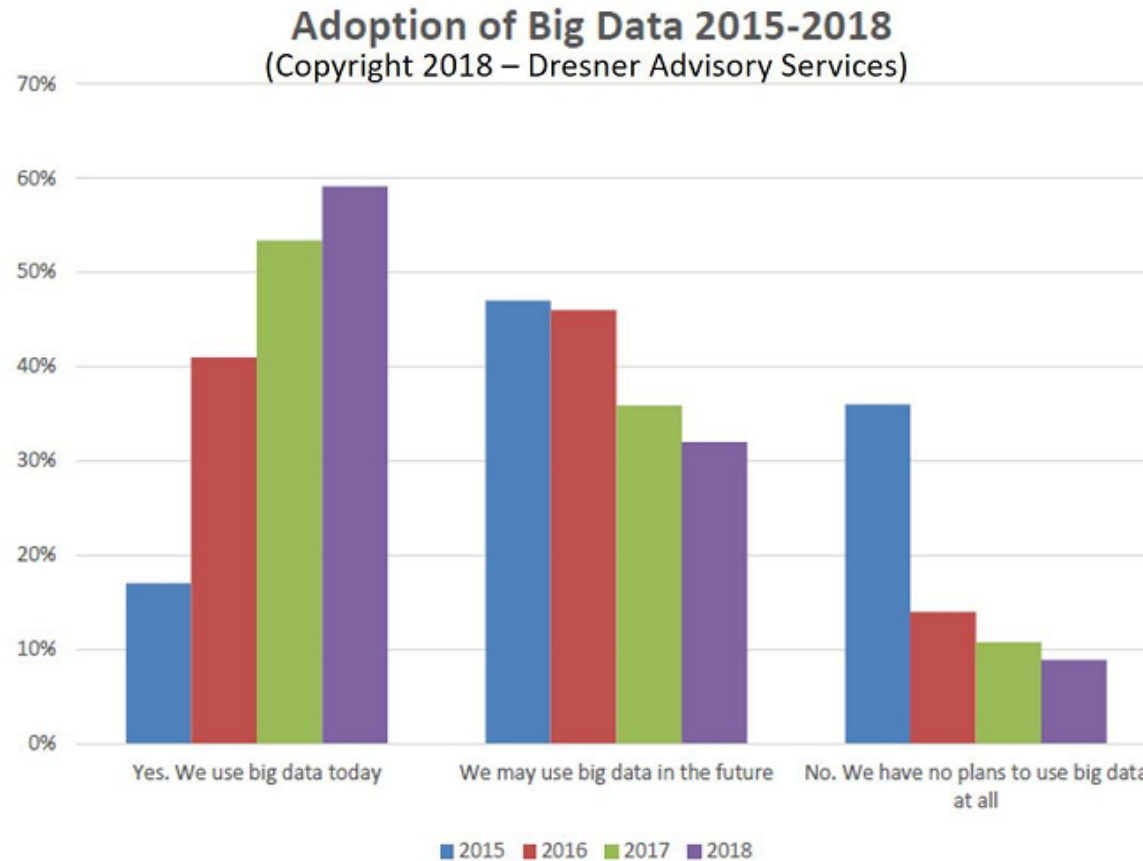
Quality Assurance (ประกันคุณภาพ)

- Big Data & Data Analytics

Predictive Maintenance (พยากรณ์การซ่อมบำรุง)

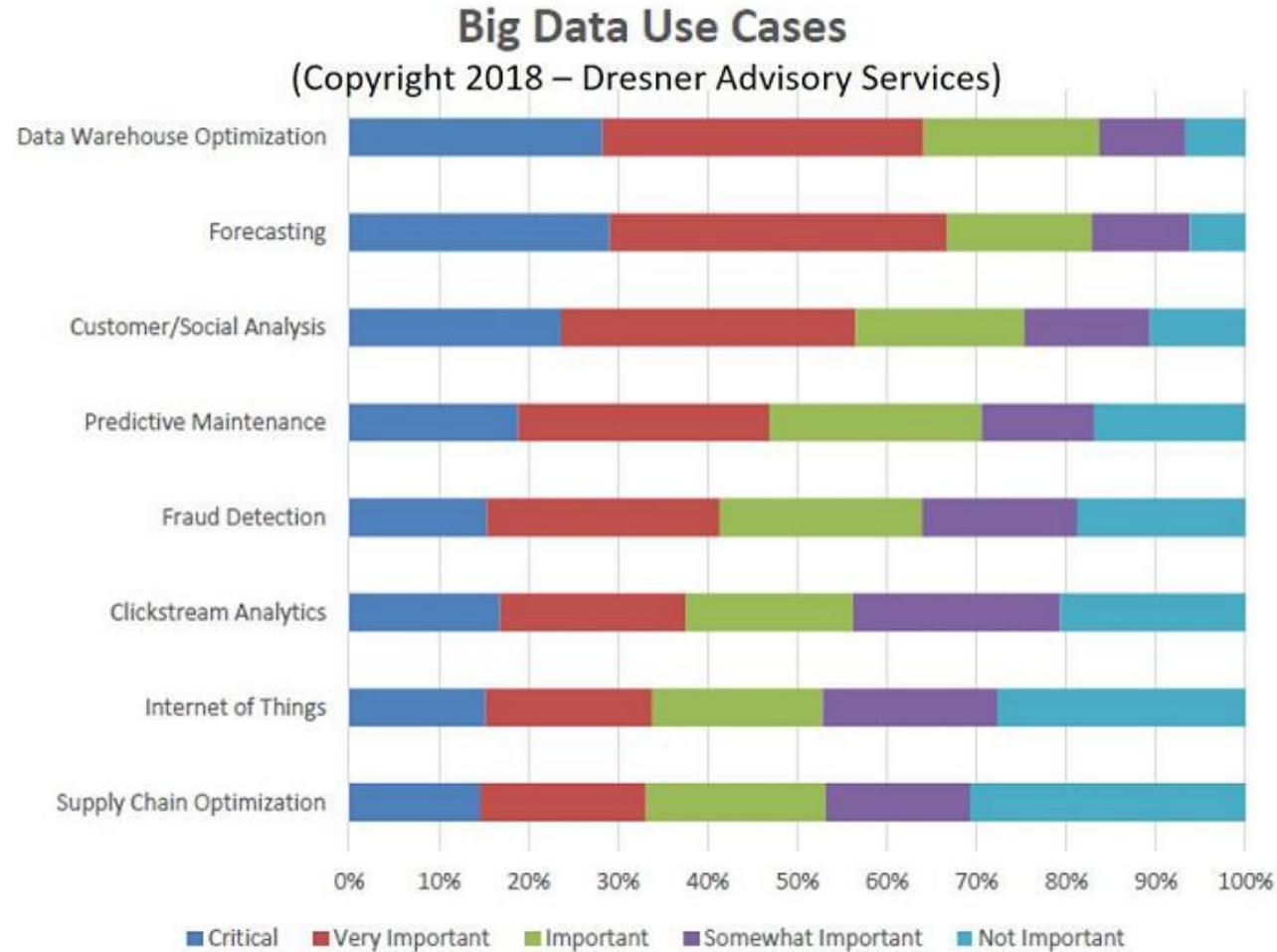
- Big Data & Data Analytics

Big Data Adoption



Source: Louis Columbus, Forbes

Big Data Adoption



Source: Louis Columbus, Forbes

Big Data Adoption Process

- Workshop: Understand & harness what you have
- Pre-adoption: Get instant answers (even with unstructured sources)
- Early adoption: Integrate legacy solution. Executive buy-in.
Phased approach. Technologies. Tools. Skills.
- Corporate adoption: Data security. Coordinate with other initiatives.
- Mature/Visionary: Analytics closer to the source.
Enhanced decision-making process.
Better cost-effective approach.

Source: microfocus.com

Big Data Technologies



Big Data Technologies

Data Storage	Hadoop, MongoDB (NoSQL), RainStor, Hunk
Data Mining	Presto, Rapid Miner, Elasticsearch
Data Analytics	Kafka, Splunk, KNIME, Spark, R
Data Visualization	Tableau, Plotly
Others	TensorFlow, Beam, Docker, etc

Source: edureka.co

3.3 บทที่ 3 : Big Data Planning



Big Data Planning

8 - 9 กันยายน 2563

ผศ.ดร.สิริกร สันติโรจนกุล

อาจารย์ภูษิต ก้อนสุรินทร์

อาจารย์ ดร.เอกสิทธิ์ เทียมแก้ว

Digital Marketing Solution Providers Center

เพจ @digitalSPC

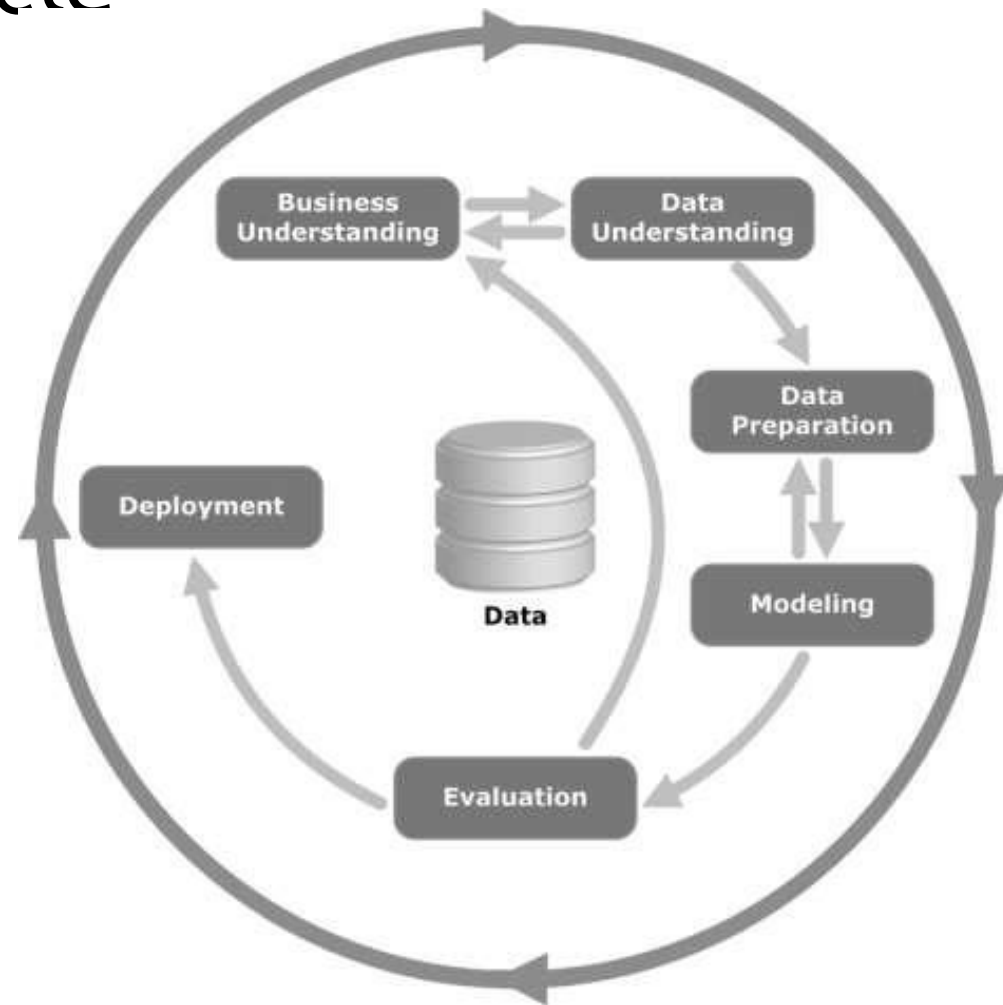


Topics

- Big Data Analytics Lifecycle
- Business Case Evaluation
- Data Identification
- Data Acquisition & Filtering
- Data Extraction
- Data Validation & Cleansing
- Data Aggregation & Representation
- Data Analysis
- Data Visualization

Traditional Data Mining Lifecycle

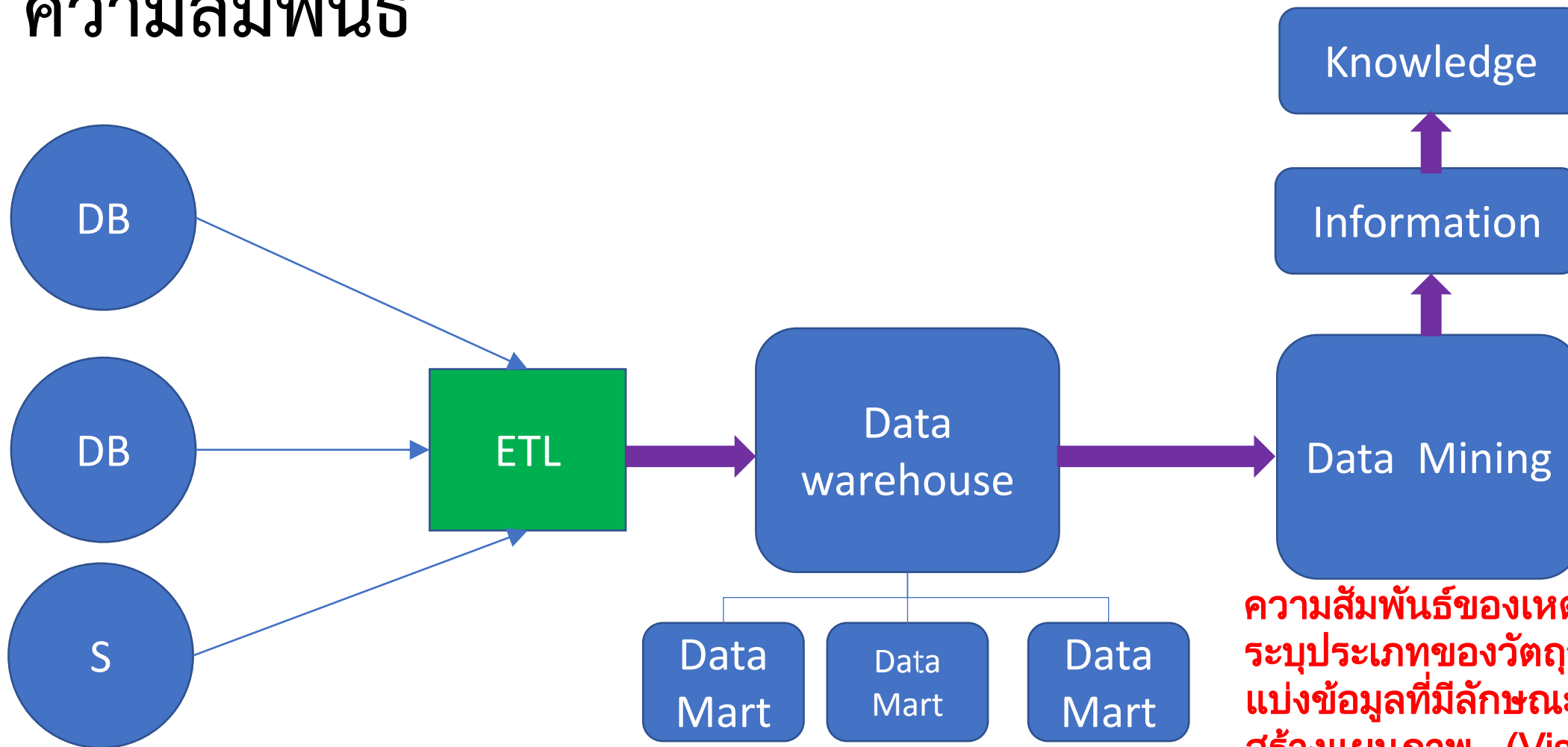
- CRISP-DM methodology
- Cross Industry Standard Process for Data Mining



Data Mining ?

- Database : ฐานข้อมูลที่ใช้ในการเก็บข้อมูล สามารถจัดเก็บ แก้ไข ปรับปรุงและลบข้อมูลได้
- Data warehouse : คลังข้อมูลที่เกิดจากการรวบรวมหลายๆ ฐานข้อมูลมีความเกี่ยวข้องกัน นำมาใช้ในการแสดงรายงาน
- Data Mining เป็นการวิเคราะห์ข้อมูลเพื่อค้นหาความสัมพันธ์ หรือ รูปแบบของข้อมูลที่เกี่ยวข้องกัน เพื่อนำมาใช้ให้เกิดประโยชน์ต่อองค์กร

ความสัมพันธ์



ความสัมพันธ์ของเหตุการณ์หรือวัตถุ
ระบุประเภทของวัตถุจากคุณสมบัติของ
แบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็น
สร้างแผนภาพ (Visualization)

ตัวอย่างการนำเอา Data Mining ไปใช้งาน

- ด้านการตลาด

- การทำนายผลการตอบสนองกับการเปิดตัวสินค้าใหม่
- การทำนายยอดขายเมื่อมีการลดราคาสินค้า
- การทำนายกลุ่มลูกค้าที่น่าจะใช้สินค้าของเรา

- ด้านการเงินการธนาคาร

- การคาดการณ์ถึงโอกาสในการชำระหนี้ของลูกค้าว่าสูงเท่าไร?
- การค้นหาลูกค้าขาดคุณภาพ เพื่อหลีกเลี่ยงความเสี่ยงในการปล่อยกู้
- การค้นหาลูกค้าชั้นดี เพื่อเสนอการปล่อยกู้
- การทำนายแนวโน้มของพฤติกรรมการใช้บัตรเครดิต

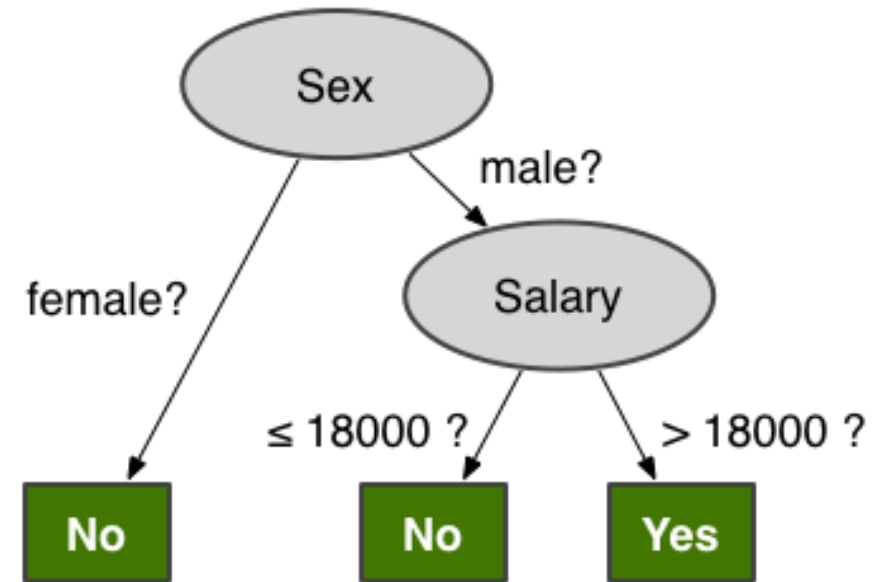
ตัวอย่าง Data Mining

Classification Target

Attributes

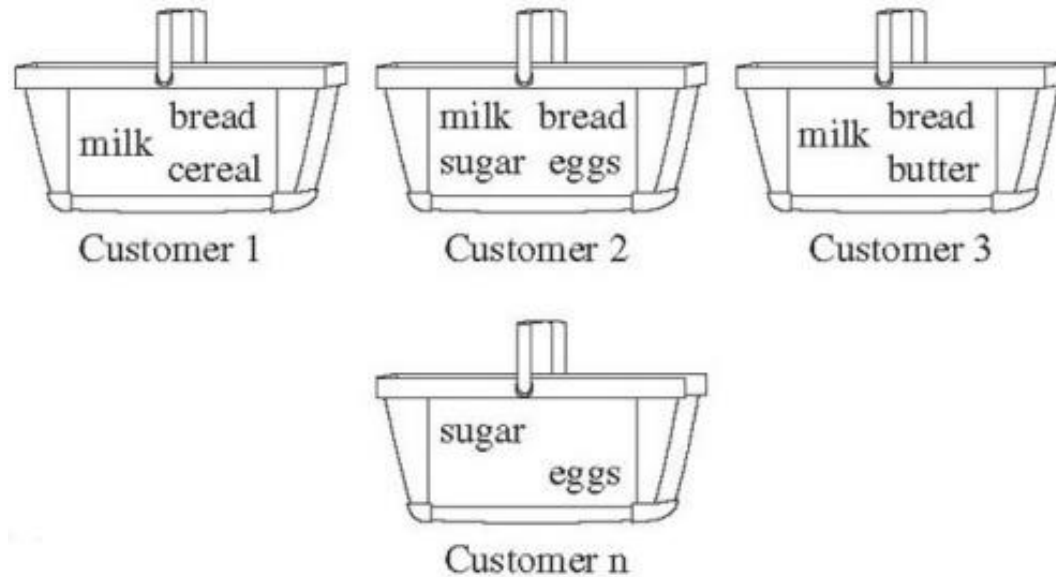
Instances

Name	Salary	Sex	Age	Buy widget
Bloggs	15000	male	19	No
Jones	25000	male	33	Yes
Smit	23000	female	50	No
Smit	16000	male	40	No
...				
Smit	200	male	10	No
Patel	30000	female	30	No
Steel	25000	male	23	Yes
Higgs	18000	female	55	No
Puggs	50000	male	57	Yes
Puggs	51000	female	57	No

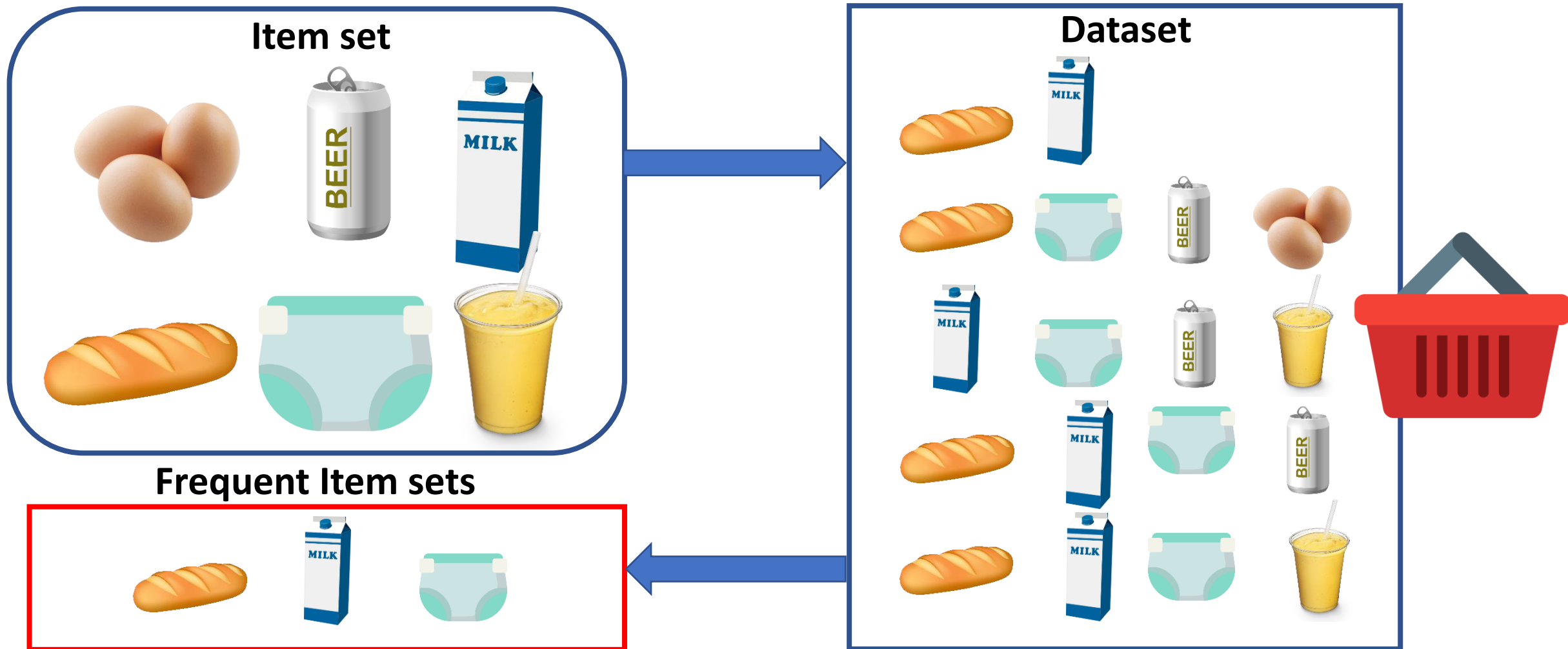


ตัวอย่าง Data Mining : Market Basket Analysis

- เทคนิคในการค้นหากลุ่มสินค้าที่ซื้อพร้อมกันในร้านค้า เป็นการวิเคราะห์ธุรกรรม โดยจะแสดงรูปแบบความสัมพันธ์ต่างๆ เช่น สินค้าที่ลูกค้าซื้อพร้อมกันบ่อย ๆ เช่น ขนมปังและเนย หรือรายการใดที่มียอดขายสูงขึ้นในบางวันเช่น ยอดขายเบียร์ในวันศุกร์
- ข้อมูลนี้ช่วยในการวางแผนการจัดวางร้านค้าโดยเสนอส่วนลดพิเศษให้กับสินค้าที่มีความต้องการน้อยสร้างข้อเสนอเช่น “ซื้อ 2 แถม 1” หรือ “รับ 50% สำหรับการซื้อครั้งที่สอง” เป็นต้น



ตัวอย่าง Data Mining : Market Basket Analysis



Amazon กับ Data Mining

- ใช้เทคนิค Data Mining ไปประยุกต์ใช้กับงาน ได้แก่
- ระบบแนะนำหนังสือให้กับลูกค้า ข้อมูลการสั่งซื้อทั้งหมดจะถูกนำมาประมวลผลเพื่อหาความสัมพันธ์ของข้อมูล เช่น
- ลูกค้าที่ซื้อหนังสือเล่มหนึ่ง ๆ มักจะซื้อหนังสือเล่มใดพร้อมกันด้วยเสมอ ใช้ความสัมพันธ์ที่ได้จะสามารถนำไปใช้คาดเดาได้ว่าควรแนะนำหนังสือเล่มใดให้กับลูกค้าที่เพิ่งซื้อหนังสือจากร้าน
- Amazon เรียกว่า Product Recommendations ทำให้ได้ส่วนแบ่งจากการทำการแนะนำสินค้าแบบนี้กว่าร้อยละ 35 จากยอดขายทั้งหมด” (Siegel, 2013)

Amazon's Product Recommendations

Data Science for Marketing Analytics
\$25.99

Buy **\$25.99**
Digital List Price: ~~\$34.99~~
Print List Price: ~~\$39.99~~ - Save \$14.00 (35%)

eBook features:

- Highlight, take notes, and search in the book
- In this edition, page numbers are just like the physical edition
- Length: 422 pages
- Enhanced Typesetting: Enabled
- Page Flip: Enabled
- Due to its large file size, this book may take longer to download

Read with the free Kindle apps (available on iOS, Android, PC & Mac), Kindle E-readers and on Fire Tablet devices. See all supported devices

Sold by: Amazon.com Services LLC

More Buying Choices New (1) from \$25.99

Key Features

- Study new techniques such as clustering and regression for marketing analytics
- Explore the applications of machine learning to power your marketing engines

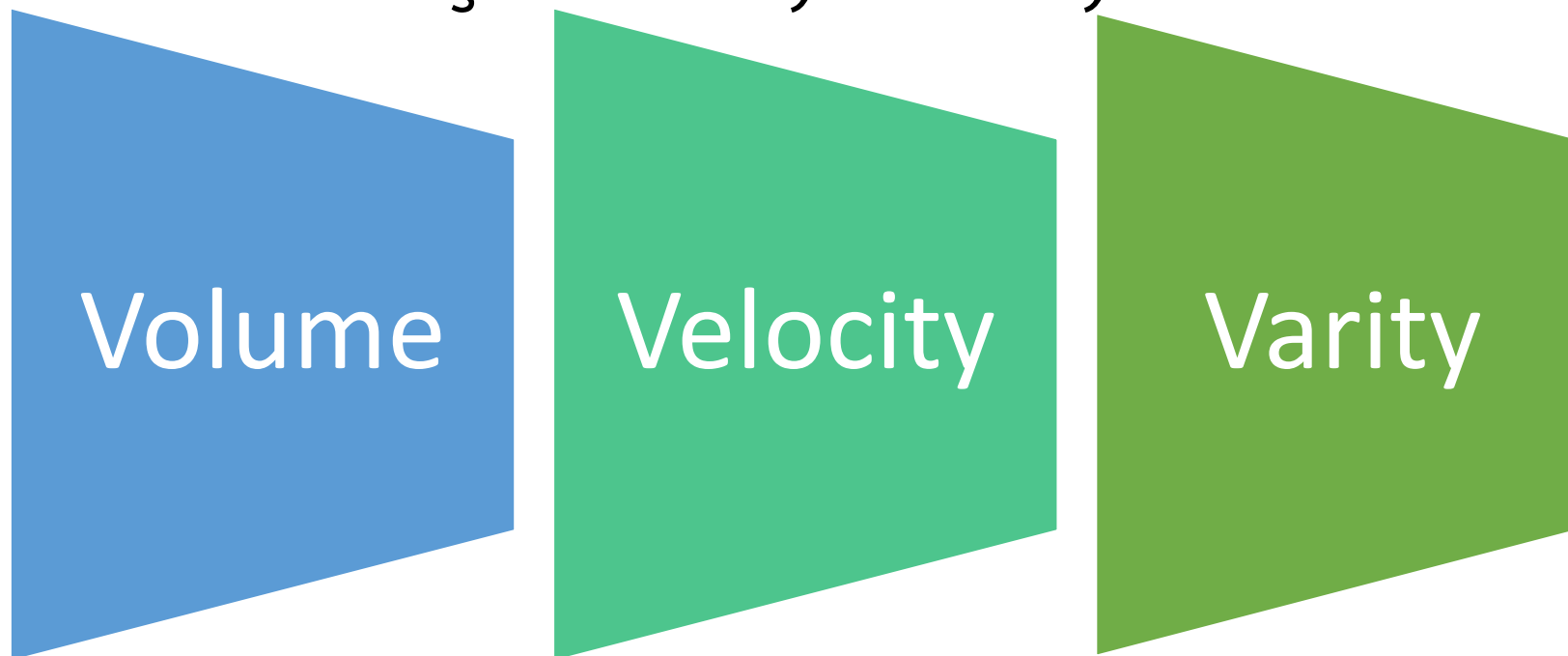
Customers who bought this item also bought

- Predictive Marketing: Easy Ways Every Marketer Can Use Customer Analytics...** by Omer Artun (38 reviews)
- Marketing Data Science: Modeling Techniques in Predictive Analytics with...** by Miller Thomas W. (41 reviews)
- Data-Driven Marketing: The 15 Metrics Everyone in Marketing Should Know** by Mark Jeffery (112 reviews)
- Beyond The Prototype: A roadmap for navigating the fuzzy area between...** by Douglas Ferguson (58 reviews)
- Lean Analytics: Use Data to Build a Better Startup Faster** by Alistair Croll (223 reviews)
- Making Websites Win** by Karl Blanks (150 reviews)
- Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn,...** by Matthew A. Russell (11 reviews)

Traditional Data Mining Lifecycle

V.S.

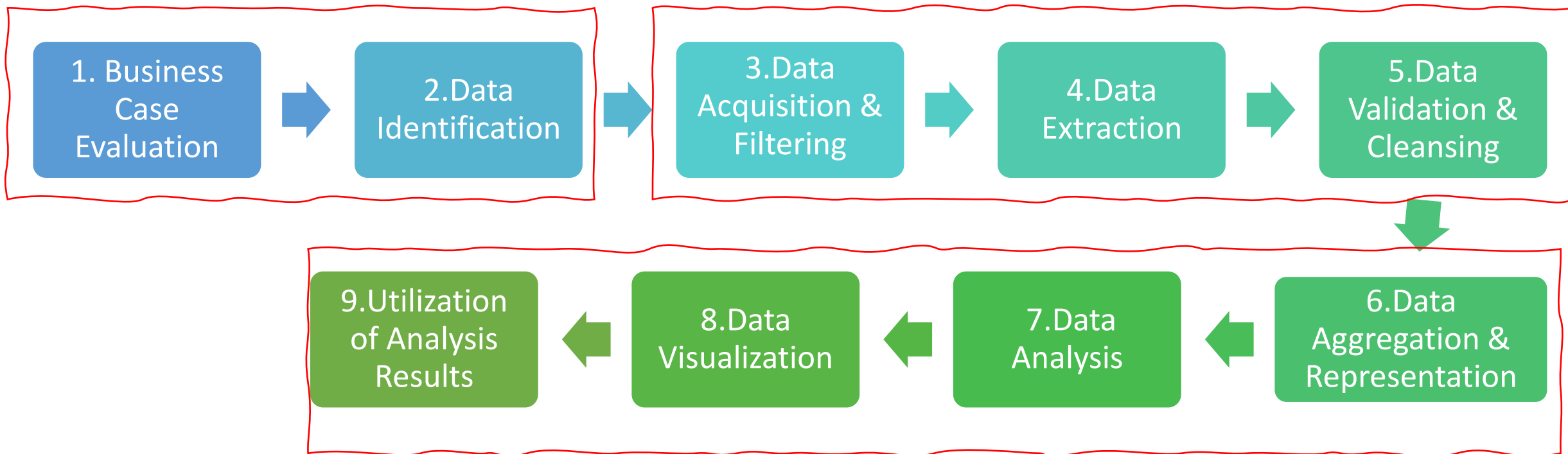
Big Data Analytics Lifecycle



Big Data Analytics Lifecycle

Business Domain Tasks

Data Engineer Tasks

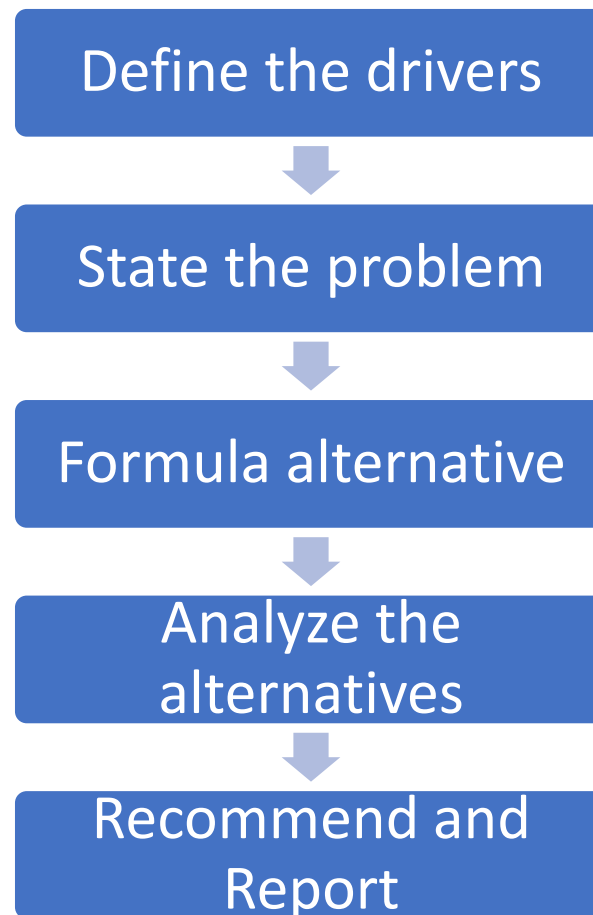


Data Analyst Tasks

Business Case Evaluation (BCE)

- การประเมินธุรกิจ : เป็นการประเมินการดำเนินการในกรณีต่างๆของธุรกิจที่กำหนดไว้อย่างชัดเจน โดยสร้างความเข้าใจในธุรกิจก่อนการวิเคราะห์ข้อมูล ทำให้ทราบวิสัยทัศน์ พันธกิจ เป้าหมาย กลยุทธ์ กิจกรรม กระบวนการ ทรัพยากร ความท้าทาย ความเสี่ยง เป็นต้น

แนวทางในการทำ BCE



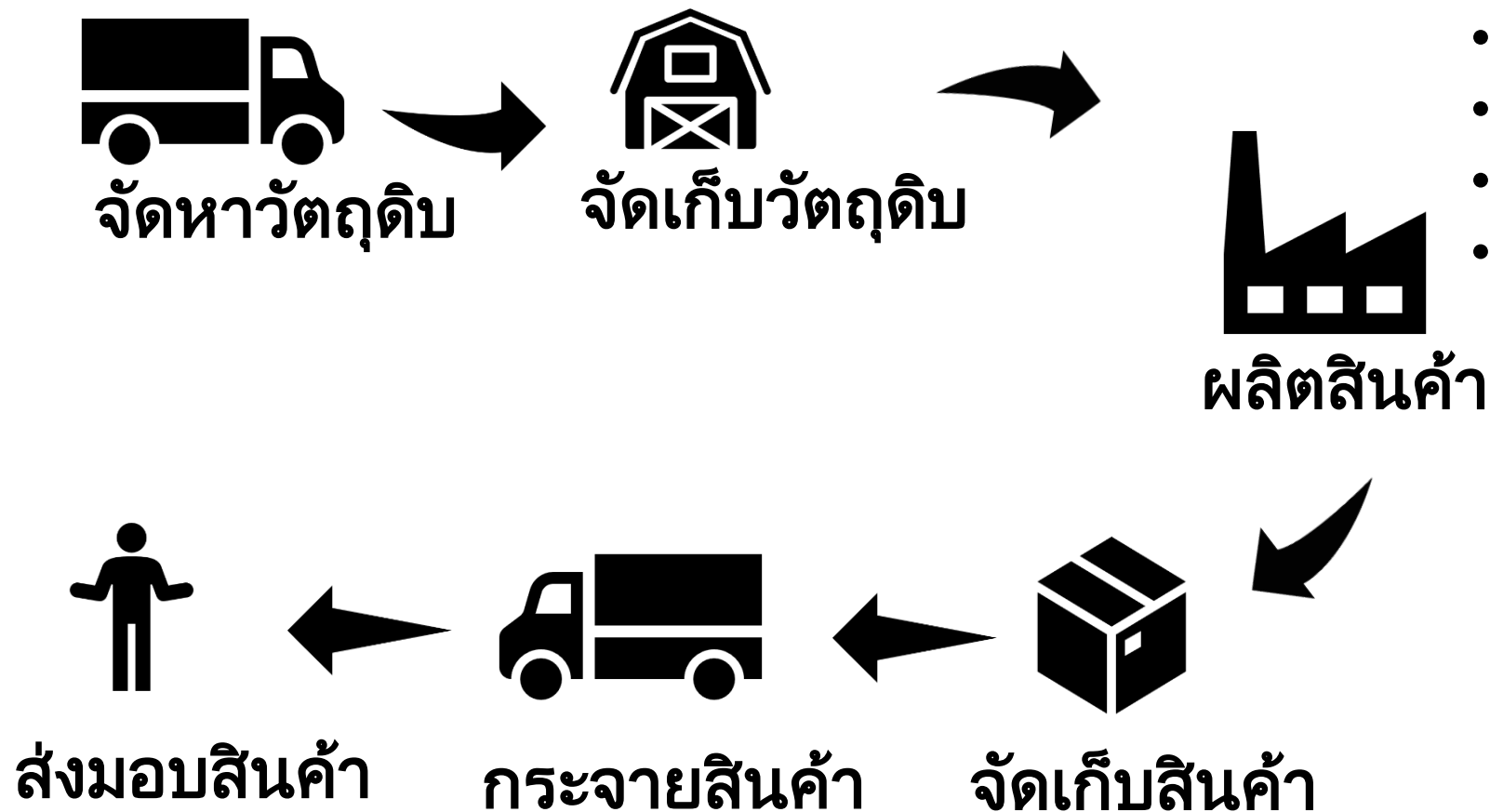
Big Data Analytics Project Understanding

- สร้างความเข้าใจในการทำโครงการวิเคราะห์ข้อมูล โดยจะต้องทราบในประเด็นต่อไปนี้
- เป้าหมายของโครงการ
 - วัตถุประสงค์
 - ตัวชี้วัด
- ขอบเขตและข้อจำกัดของโครงการ
- ทรัพยากรของโครงการ
 - ทรัพยากรบุคคล
 - เทคโนโลยี
 - งบประมาณ

วิเคราะห์กระบวนการทำงาน

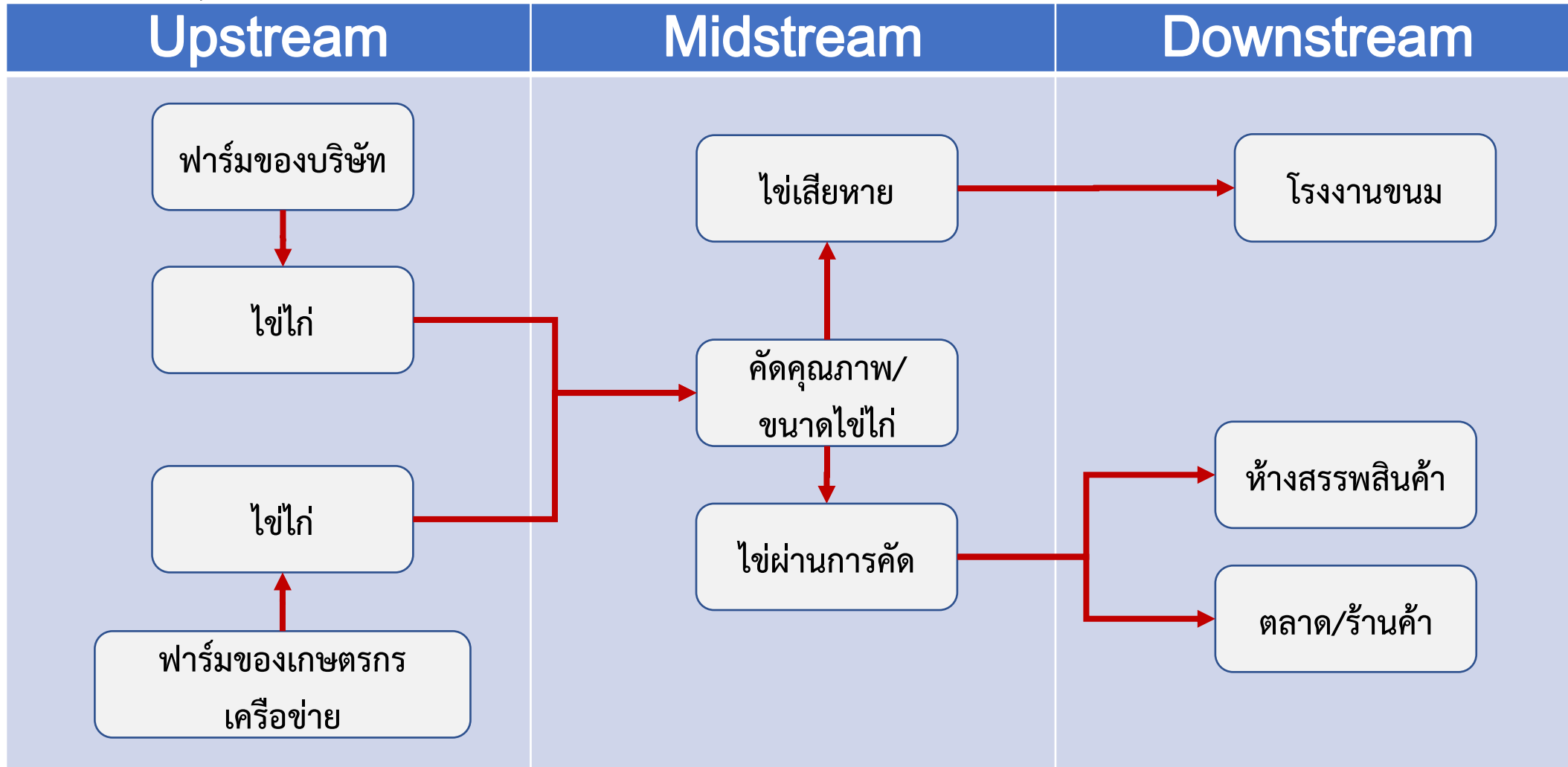
- เป็นการวิเคราะห์กระบวนการทางกิจกรรมต่างๆ ที่สอดคล้องกัน โดยแบ่งออก 3 ส่วนคือ
 1. **ต้นน้ำ (Upstream)** แหล่งที่มาของวัตถุดิบ แหล่งทุน แรงงานมาจากไหน ซัพพลายเออร์ ผู้ผลิต หรือวัตถุดิบต่างๆ
 2. **กลางน้ำ (Midstream)** กระบวนการ การจัดการ การผลิต เพื่อให้เกิดสินค้า หรือ บริการ
 3. **ปลายน้ำ (Downstream)** กระบวนการ จัดเก็บ, ขนส่ง และ ส่งมอบสินค้าหรือ บริการนั้น ไปถึงมือลูกค้า ที่ต้องใช้สินค้าของเรา

Supply Chain หรือห่วงโซ่อุปทาน



- Material Flow
(การไหลของวัตถุดิบ)
- Information Flow
(การไหลของข้อมูล)
- Financial Flow
(การไหลของเงินทุน)

Supply Chain



วิเคราะห์ SWOT



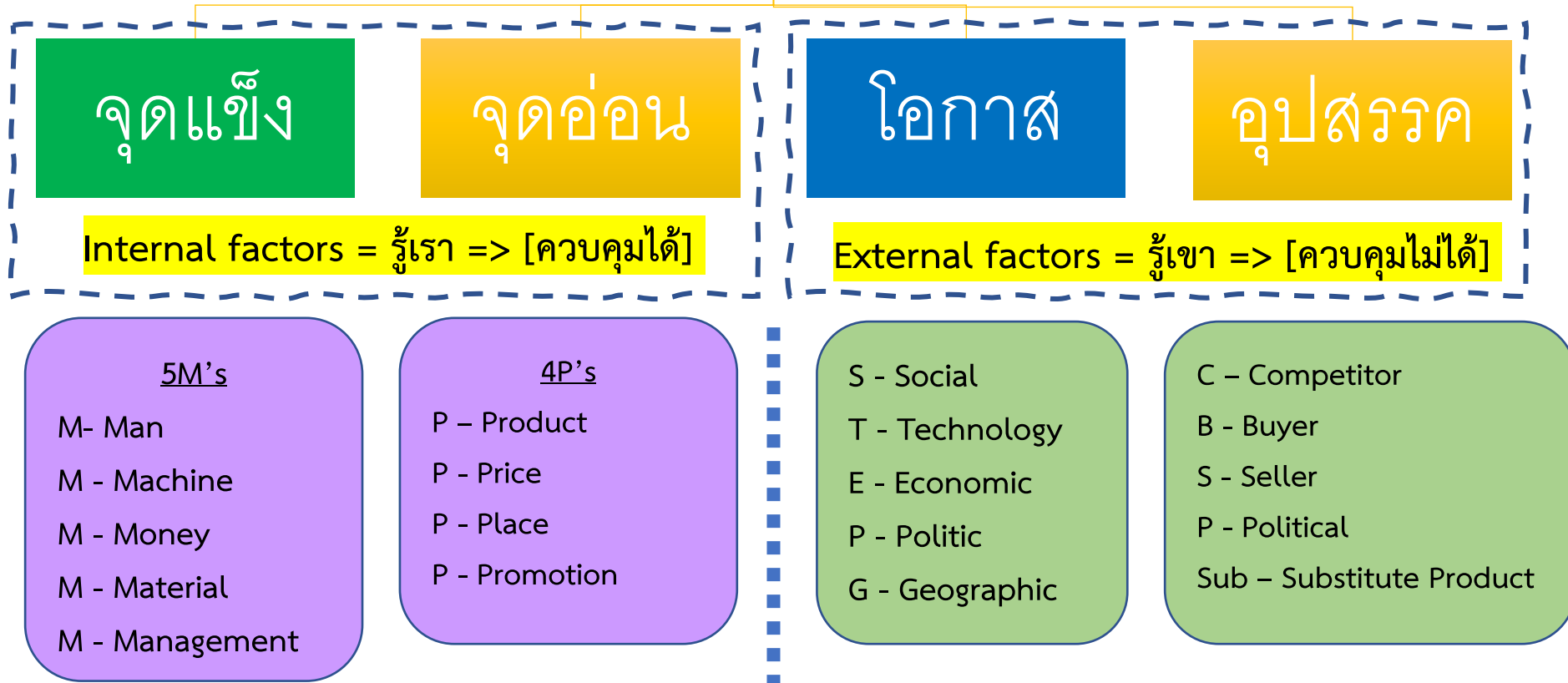
“SWOT analysis is a powerful way to evaluate your company or project”

Strengths Weaknesses Opportunities Threats

(จุดแข็ง จุดอ่อน โอกาส อุปสรรค)

การวิเคราะห์จากปัจจัยภายใน-ภายนอกองค์กร

SWOT



การสร้างกลยุทธ์ด้วย TOWS Matrix



ตัวอย่างการวิเคราะห์ SWOT ธุรกิจโรงแรม

<p>จุดแข็ง (Strengths)</p> <ul style="list-style-type: none"> • โรงแรมได้รับรางวัล Best Choice 9.8 คะแนน • โรงแรมตั้งอยู่ในทำเลที่ดีที่สุด ใกล้การคมนาคม และแหล่งท่องเที่ยว ช้อปปี้ง • มีฐานลูกค้าเหนียวแน่น • มีห้องประชุม สัมมนา ที่ทันสมัย 	<p>จุดอ่อน (Weaknesses)</p> <ul style="list-style-type: none"> • ไรท์ทำตลาดออนไลน์ • ลูกค้ารายใหม่มีน้อย • การจัดการระบบห้องประชุมสัมมนา ยังไม่เป็นระบบ ทำให้การให้บริการมีข้อบกพร่อง
<p>โอกาส (Opportunities)</p> <ul style="list-style-type: none"> • รัฐบาลมีนโยบายคูปองส่งเสริมการท่องเที่ยว สำหรับการท่องเที่ยวในประเทศ 	<p>อุปสรรค (Threats)</p> <ul style="list-style-type: none"> • เกิดสถานการณ์ COVID-19 ทำให้ไม่มีลูกค้าจากนักท่องเที่ยวต่างชาติ • มีโรงแรมใหม่ๆเกิดขึ้นจำนวนมาก • พฤติกรรมของลูกค้าเปลี่ยนไป

กลยุทธ์กลยุทธ์เชิงรุก (SO) ใช้จุดแข็ง+โอกาส

จุดแข็ง (Strengths)

- โรงแรมได้รับรางวัล Best Choice 9.8 คะแนน
- โรงแรมตั้งอยู่ในทำเลที่ดีที่สุด ใกล้การคมนาคม และแหล่งท่องเที่ยว ช้อปปิ้ง
- มีฐานลูกค้าเหนียวแน่น
- มีห้องประชุม สัมมนา ที่ทันสมัย

โอกาส(Opportunities)

- รัฐบาลมีนโยบายคุ้มครองส่งเสริมการท่องเที่ยว สำหรับการท่องเที่ยวในประเทศ

กลยุทธ์เชิงแก้ไข/พัฒนา (WO) ใช้โอกาสลดจุดอ่อน

จุดอ่อน (Weaknesses)

- การทำตลาดออนไลน์ไม่เก่ง
- ลูกค้ารายใหม่มีน้อย
- การจัดการระบบห้องประชุมสัมมนา ยังไม่เป็นระบบ ทำให้การให้บริการมีข้อบกพร่อง

โอกาส (Opportunities)

- รัฐบาลมีนโยบายคุ้มครองส่งเสริมการท่องเที่ยว สำหรับการท่องเที่ยวในประเทศ

กลยุทธ์เชิงรับ (ST) ใช้จุดแข็งรับมืออุปสรรค

จุดแข็ง (Strengths)

- โรงแรมได้รับรางวัล Best Choice 9.8 คะแนน
- โรงแรมตั้งอยู่ในทำเลที่ดีที่สุด ใกล้ การคมนาคม และแหล่งท่องเที่ยว ช้อปปิ้ง
- มีฐานลูกค้าเหนียวแน่น
- มีห้องประชุม สัมมนา ที่ทันสมัย

อุปสรรค (Threats)

- เกิดสถานการณ์ COVID-19 ทำให้ ไม่มีลูกค้าจากนักท่องเที่ยวต่างชาติ
- มีโรงแรมใหม่ๆ เกิดขึ้นจำนวนมาก
- พฤติกรรมของลูกค้าเปลี่ยนแปลงไป

กลยุทธ์เชิงป้องกัน (ถอย) (WT) แก้ไขจุดอ่อน+เลี่ยงอุปสรรค

จุดอ่อน (Weaknesses)

- การทำตลาดออนไลน์ไม่เก่ง
- ลูกค้ารายใหม่มีน้อย
- การจัดการระบบห้องประชุมสัมมนา ยังไม่เป็นระบบ ทำให้การให้บริการมีข้อบกพร่อง

อุปสรรค (Threats)

- เกิดสถานการณ์ COVID-19 ทำให้ไม่มีลูกค้าจากนักท่องเที่ยวต่างชาติ
- มีโรงแรมใหม่ๆเกิดขึ้นจำนวนมาก
- พฤติกรรมของลูกค้าเปลี่ยนแปลงไป

Data Identification

External Data Source

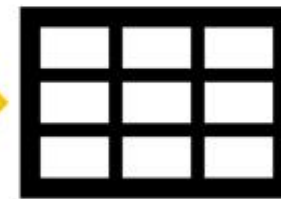
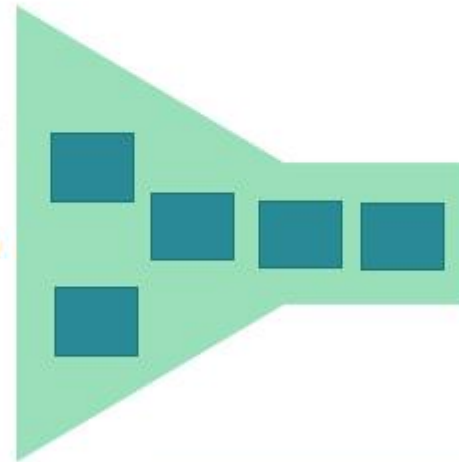


- Business Scope
- Business problem
- Business Case Evaluation



Internal Data Source

- Business Scope
- Business problem
- Business Case Evaluation



Datasets



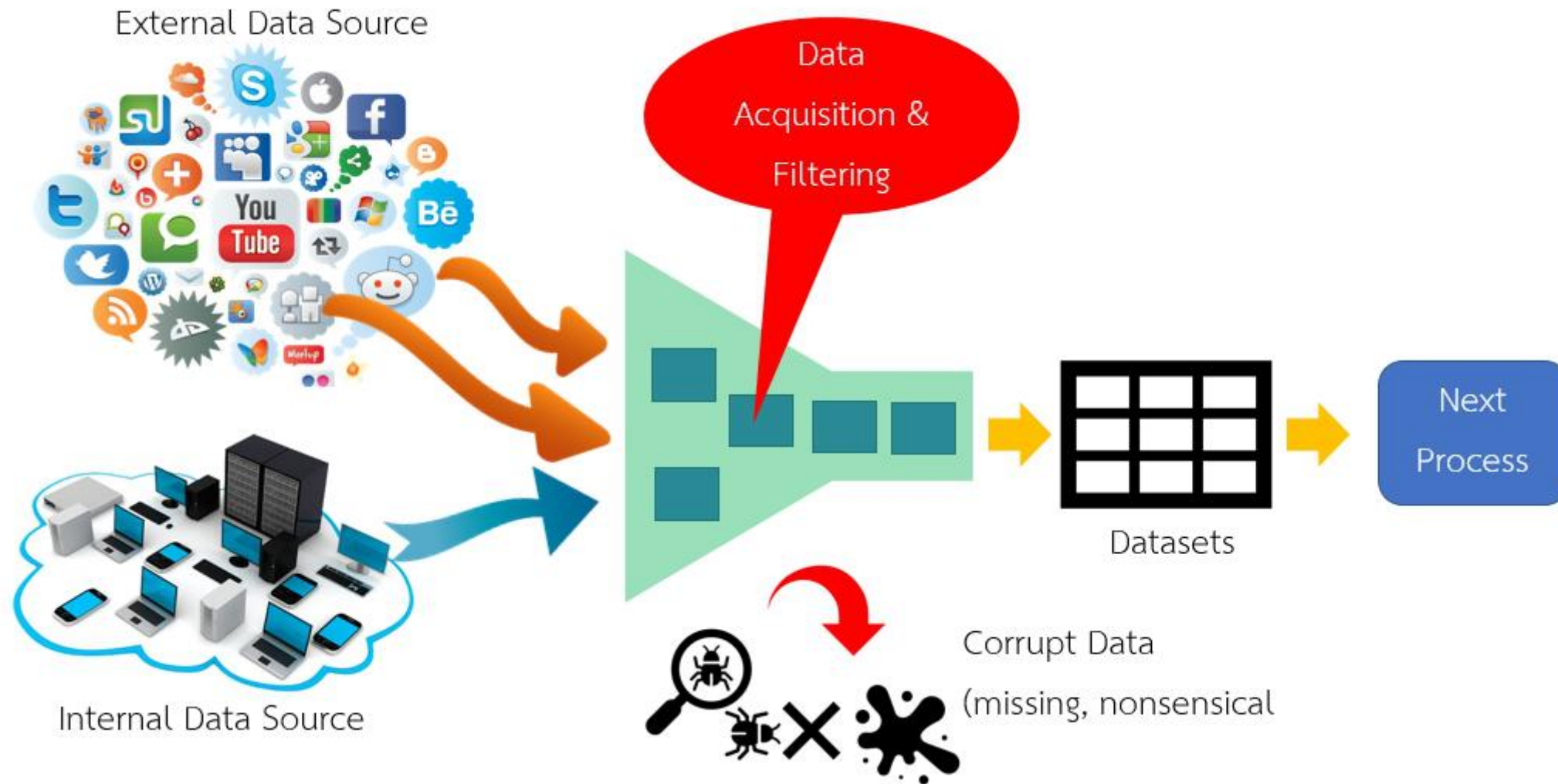
Next Process

Data Acquisition & Filtering

- ข้อมูลจะถูกรวบรวมจากแหล่งข้อมูลทั้งหมดที่ระบุในขั้นตอนก่อนหน้านี้ จากนั้นข้อมูลที่ได้มาจะถูกกรองโดยอัตโนมัติ
- การกรองข้อมูลที่เสียหายหรือข้อมูลที่ไม่มีคุณค่าในการวิเคราะห์ออกไป



Data Acquisition & Filtering



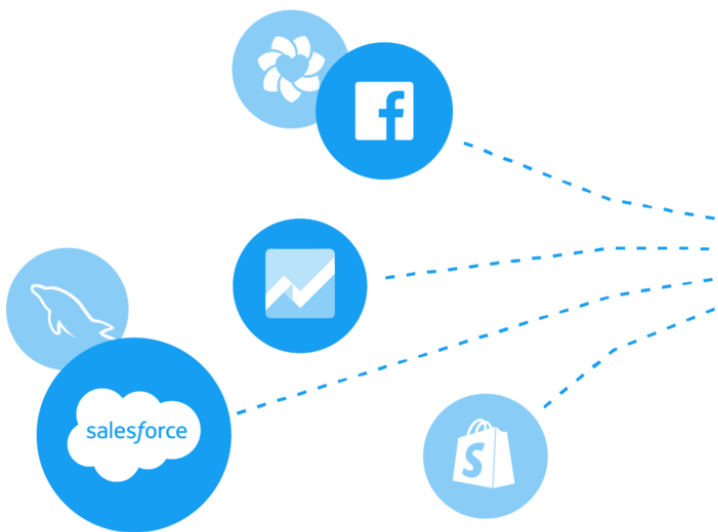
Data Extraction

- ข้อมูลบางส่วนที่ระบุว่าป็นอินพุตสำหรับการวิเคราะห์อาจอยู่ในรูปแบบที่ไม่สอดคล้องกับโซลูชันของ Big Data
- ดังนั้นมีความจำเป็นอย่างยิ่งในการระบุประเภทของข้อมูลที่แตกต่างกันเมื่อต้องใช้ข้อมูลจากแหล่งภายนอก
- จึงต้องใช้การสกัดข้อมูลเพื่อแยกข้อมูลที่แตกต่างกันและแปลงเป็นรูปแบบที่สามารถนำมาใช้ในการวิเคราะห์ข้อมูล



Data Extraction

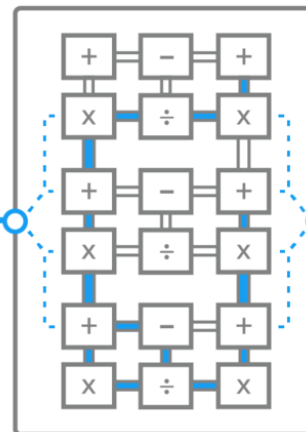
Extract



Load



Transform



Analyze

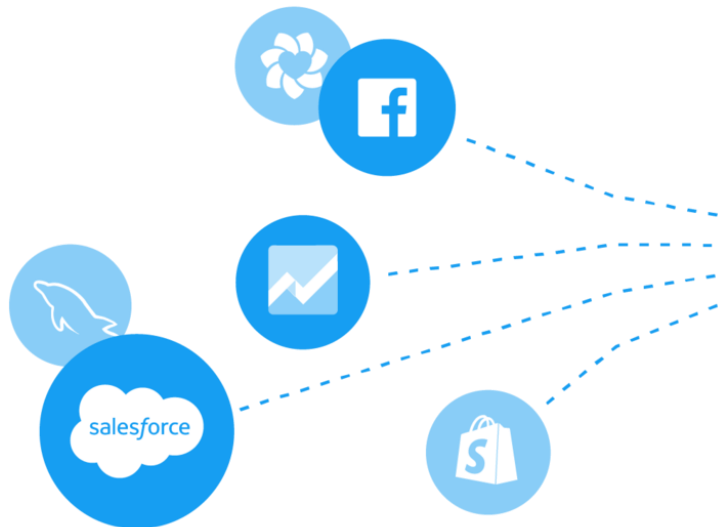


Analytics Warehouse

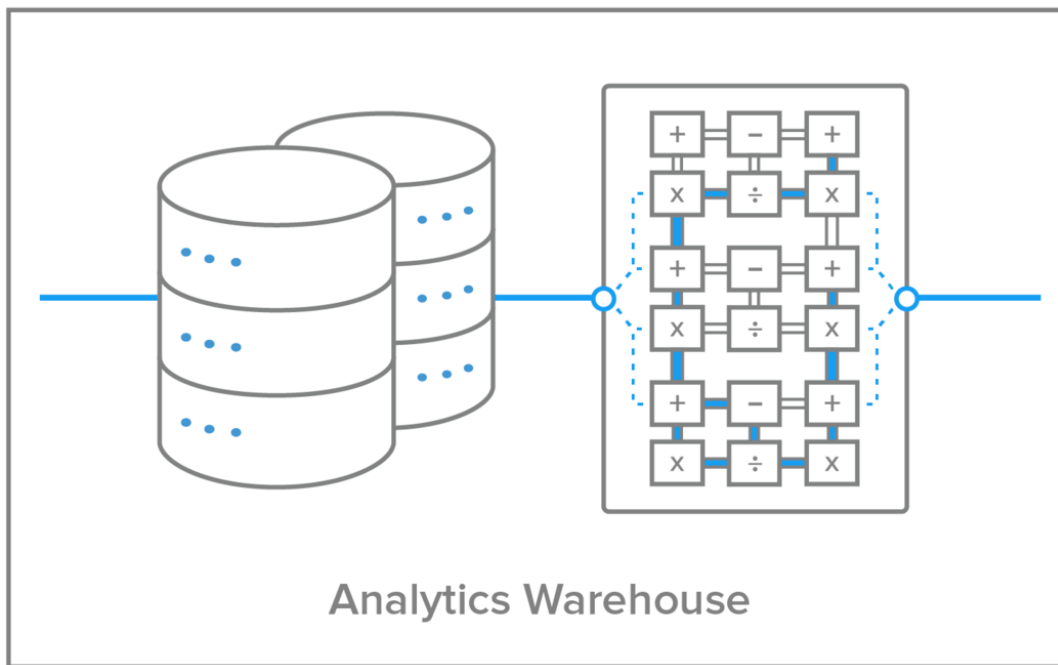
Data Extraction Software



Extract



Load



Transform

Analyze



Extracting XML data into a Structured data (table)

```

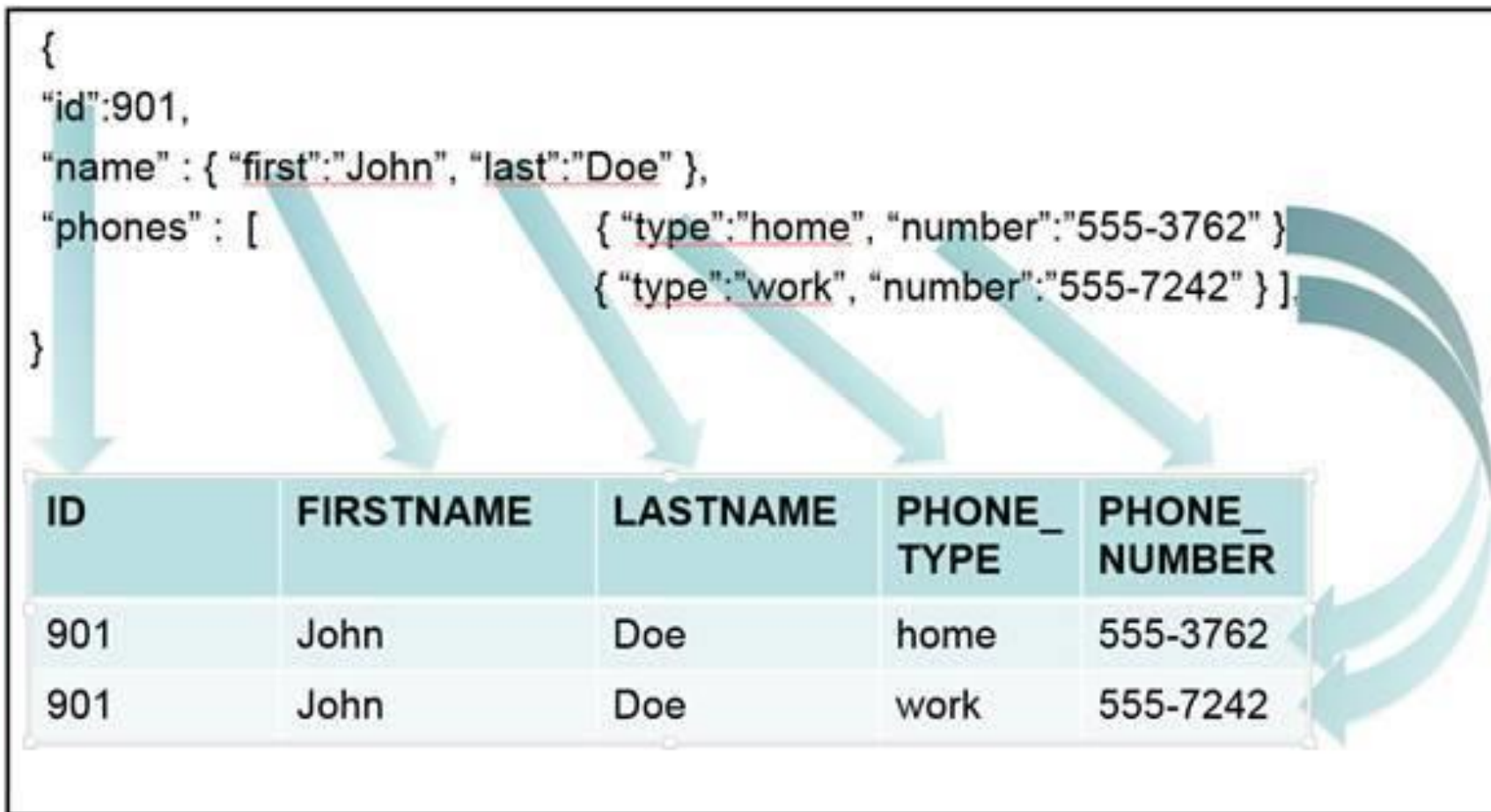
1 <?xml version="1.0"?>
2 <catalog>
3   <book id="bk101">
4     <author>Gambardella, Matthew</author>
5     <title>XML Developer's Guide</title>
6     <genre>Computer</genre>
7     <price>44.95</price>
8     <publish_date>2000-10-01</publish_date>
9     <description>An in-depth look at creating applications
10    with XML.</description>
11  </book>
12  <book id="bk102">
13    <author>Ralls, Kim</author>
14    <title>Midnight Rain</title>
15    <genre>Fantasy</genre>
16    <price>5.95</price>
17    <publish_date>2000-12-16</publish_date>
18    <description>A former architect battles corporate zombies,
19    an evil sorceress, and her own childhood to become queen
20    of the world.</description>
21  </book>
22  <book id="bk103">
23    <author>Corets, Eva</author>
24    <title>Maeve Ascendant</title>
25    <genre>Fantasy</genre>

```



book_id	author	title	genre	price	publish_date	description
bk101	Gambardella, Matthew	XML Developer's Guide	Computer	44.95	2000-10-01	An in-depth look at cr
bk102	Ralls, Kim	Midnight Rain	Fantasy	5.95	2000-12-16	A former architect bat
bk103	Corets, Eva	Maeve Ascendant	Fantasy	5.95	2000-11-17	After the collapse of a
bk104	Corets, Eva	Oberon's Legacy	Fantasy	5.95	2001-03-10	In post-apocalypse En
bk105	Corets, Eva	The Sundered Grail	Fantasy	5.95	2001-09-10	The two daughters of

Extracting JSON data into a Structured data (table)



Web Scraping

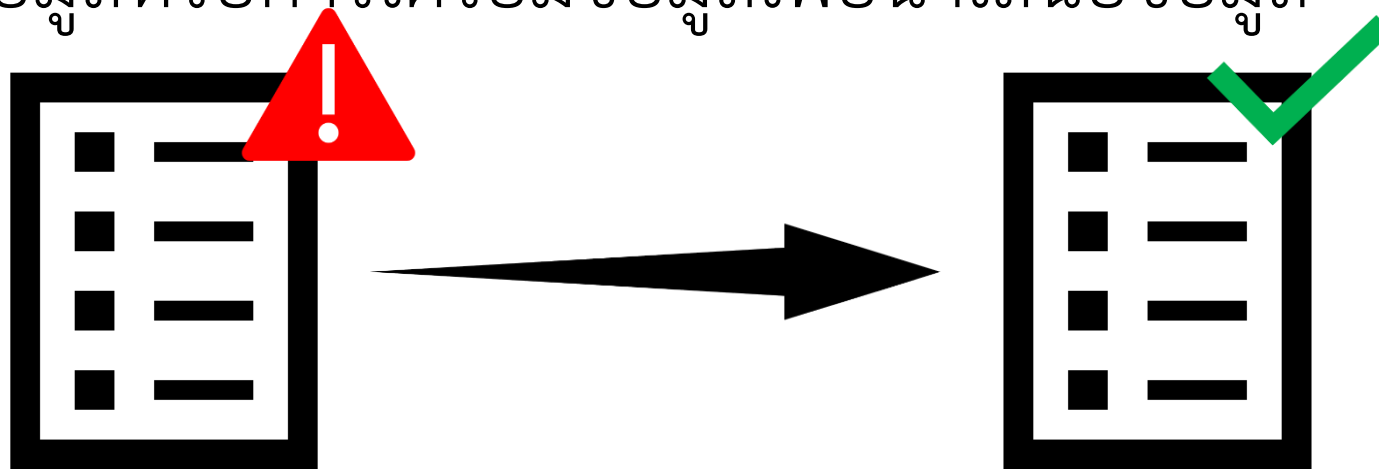


Data Validation & Cleansing

- Big Data มักจะมีการซ้ำซ้อนในชุดข้อมูลต่างๆ ความซ้ำซ้อนนี้สามารถ
ใช้เพื่อสำรวจชุดข้อมูลที่เชื่อมต่อกันเพื่อประกอบพารามิเตอร์การ
ตรวจสอบความถูกต้องและกรอกข้อมูลที่ถูกต้องที่ขาดหายไป
- ข้อมูลที่ไม่ถูกต้องอาจทำให้ผลการวิเคราะห์ผิดพลาดได้

Data Validation?

- “การตรวจสอบข้อมูล”
- เป็นส่วนสำคัญของงานจัดการข้อมูล ทั้งการรวบรวมข้อมูล การวิเคราะห์ข้อมูลหรือการเตรียมข้อมูลเพื่อนำเสนอข้อมูล



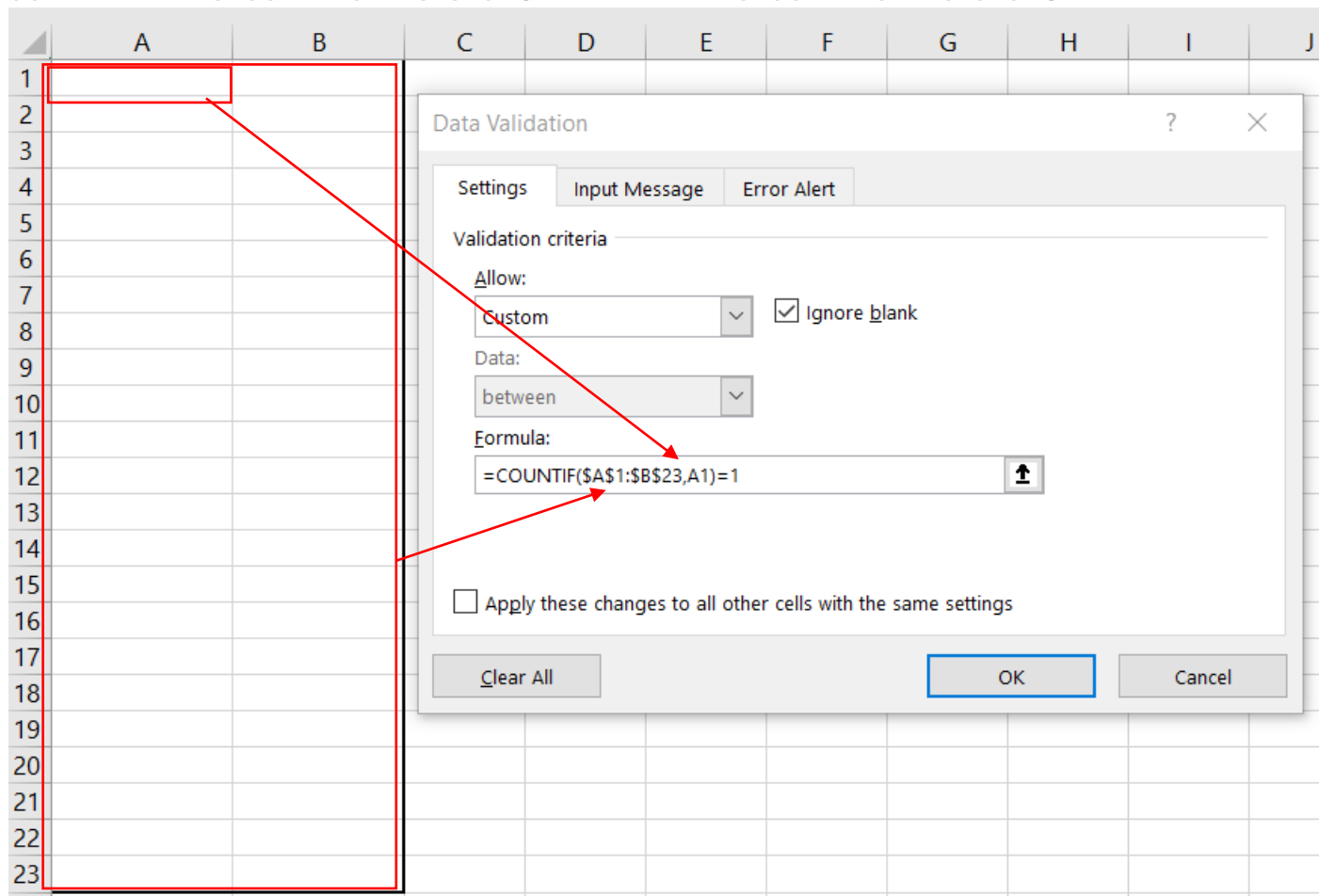
แนวทางของ Data Validation

167

- แนวทางที่ใช้ในการตรวจสอบความถูกต้องของข้อมูลคือแนวทางที่รับประกันความถูกต้องและความสมบูรณ์ของข้อมูล เช่น ตรวจสอบการสะกด การตรวจสอบข้อมูล. ความยาวรหัสผ่านขั้นต่ำ? การตรวจสอบข้อมูล เป็นต้น
- แต่ละองค์กรที่ใช้ข้อมูลจะต้องมีแนวทางของตนเองในการตรวจสอบข้อมูลเพื่อใช้งานข้อมูลอย่างมีประสิทธิภาพ ได้แก่
- ตรวจสอบชนิดข้อมูล (Data type) เช่น integer, float, string
- ตรวจสอบช่วงของข้อมูล (Range เช่น ช่วงอายุ ระหว่าง 35-40)
- ตรวจสอบข้อมูลเอกลักษณ์ (Uniqueness เช่น รหัสบุคลากร รหัสไปรษณีย์)
- ตรวจสอบความสอดคล้องของข้อมูล (Consistent expressions เช่น ชม. แทนเชียงใหม่)
- ตรวจสอบข้อมูลต้องไม่มีค่าว่าง (No null values)

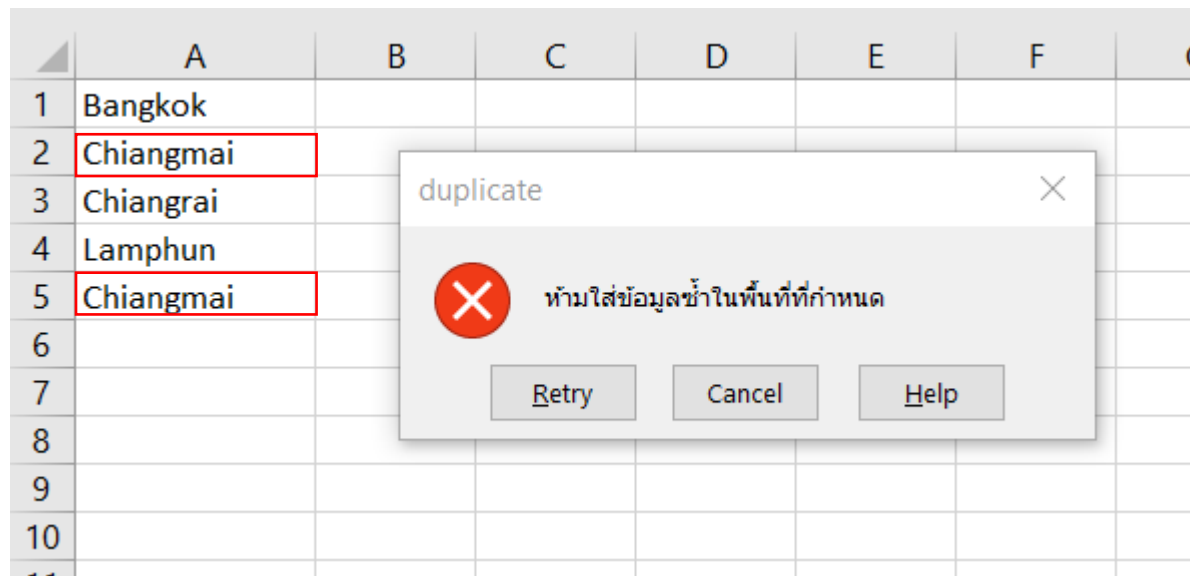
Data Validation ด้วย Excel

- Ribbon Data => Data Validation => Data Validation...



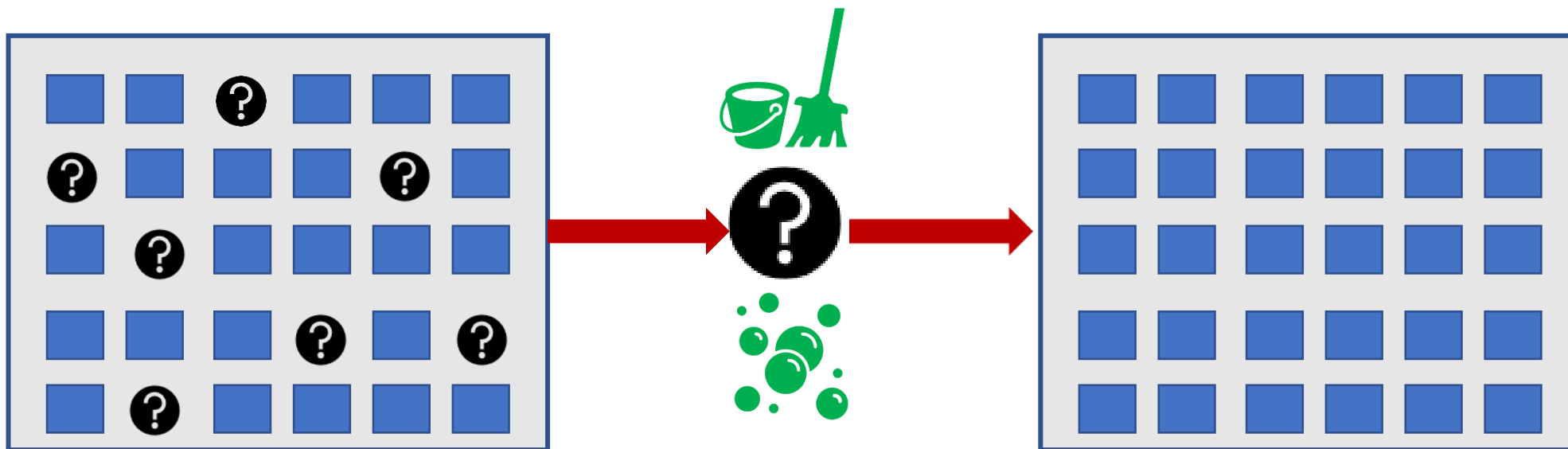
Data Validation ด้วย Excel

- Ribbon Data => Data Validation => Data Validation...



Data Cleansing?

- Data cleansing หรือ data cleaning หรือ data scrubbing เป็นการทำความสะอาดข้อมูล เป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล




Data Aggregation & Representation


- เป็นกระบวนการรวบรวมข้อมูลและนำเสนอในรูปแบบสรุป
- เป็นขั้นตอนที่สำคัญเนื่องจากความแม่นยำของข้อมูลเชิงลึกที่ได้จากการวิเคราะห์ข้อมูลขึ้นอยู่กับปริมาณและคุณภาพของข้อมูลที่ใช้
- ข้อมูลอาจถูกรวบรวมจากหลายแหล่งเพื่อจุดประสงค์ในการรวมแหล่งข้อมูลเหล่านี้เป็นสรุปสำหรับการวิเคราะห์ข้อมูล
- ประโยชน์สำหรับทุกอย่างตั้งแต่การตัดสินใจทางการเงินหรือกลยุทธ์ทางธุรกิจ ไปจนถึงการกำหนดราคาผลิตภัณฑ์การดำเนินงานและกลยุทธ์ทางการตลาด

ตัวอย่าง Data Aggregation

Location	Bear	Date	Quantity
Canada	Polar Bear	11/4/1982 12:00:00 AM	5.00
Canada	Brown Bear	1/6/1990 12:00:00 AM	10.00
USA	Black Bear	1/6/1990 12:00:00 AM	15.00
USA	Polar Bear	3/2/1990 12:00:00 AM	20.00
Mexico	Grizzly Bear	8/4/2018 12:00:00 AM	25.00

Location	Bear	Date	Quantity	Location
Canada	Brown Bear	1/6/1990 12:00:00 AM	10.00	Canada
Canada	Polar Bear	11/4/1982 12:00:00 AM	5.00	Mexico
Mexico	Grizzly Bear	8/4/2018 12:00:00 AM	25.00	USA
USA	Black Bear	1/6/1990 12:00:00 AM	15.00	
USA	Polar Bear	3/2/1990 12:00:00 AM	20.00	



Location	Bear	Date	Quantity		Location	Bear	
Canada	Brown Bear	1/6/1990 12:00:00 AM	10.00		Canada	Brown Bear	
Canada	Polar Bear	11/4/1982 12:00:00 AM	5.00			Polar Bear	
Mexico	Grizzly Bear	8/4/2018 12:00:00 AM	25.00			Mexico	Grizzly Bear
USA	Black Bear	1/6/1990 12:00:00 AM	15.00			USA	Black Bear
USA	Polar Bear	3/2/1990 12:00:00 AM	20.00				Polar Bear

Data Visualization

DATA



SORTED



ARRANGED



PRESENTED
VISUALLY



What is Data Visualization?

- การสร้างภาพข้อมูล (Data Visualization) คือการสร้างและศึกษาการแสดงภาพข้อมูล บางครั้งเรียกว่าการสื่อสารด้วยภาพหรือสถิติเชิงพรรณนาและรวมถึงเทคนิคในการนำเสนอข้อมูลในลักษณะที่เป็นภาพเพื่อการสื่อสารข้อมูลที่ชัดเจนและกระตุ้นการมีส่วนร่วมและความสนใจของผู้ชม
- การทำ Data Visualization แล้ว อาจทำให้เราไม่สามารถค้นพบนัยยะของข้อมูลในแง่ของแนวโน้ม, รูปแบบพฤติกรรม, และความสัมพันธ์เชื่อมโยงได้

ลักษณะของ Data Visualization

- **ใช้งานง่าย (Intuitive)** : เนื่องจากการนำเสนอภาพข้อมูล สามารถสื่อสารภาพได้เข้าใจได้ทันทีทำให้รู้สึกถึงแม้กับคนที่ไม่เคยทำงานกับกราฟมาก่อน
- **เข้าใจได้รวดเร็ว (Fast)** : ด้วยการสร้างภาพ ทำให้สามารถมองเห็นแนวโน้มและข้อมูลต่างๆ ได้อย่างมีประสิทธิภาพ
- **ความยืดหยุ่น (Flexible)** : มีความยืดหยุ่น ตามลักษณะของข้อมูล
- **ข้อมูลเชิงลึก (Insightful)** : มีความชาญฉลาด: ภาพข้อมูลจะช่วยให้ผู้ที่จะได้รับความรู้เพิ่มเติมในเชิงลึกเข้าใจบริบทและถามคำถามมากขึ้นเมื่อเทียบกับการสร้างภาพแบบคงที่หรือข้อมูลดิบ

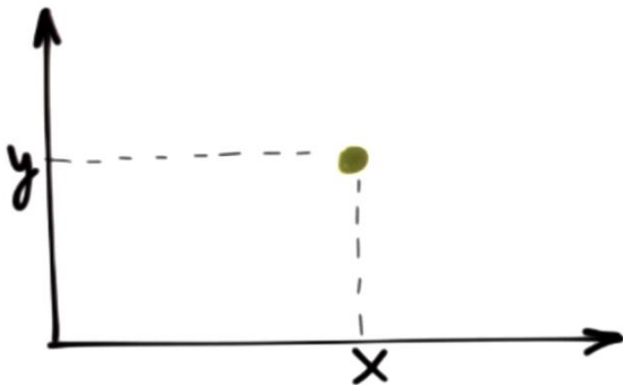
ปัจจัยประสิทธิผลในการทำ Visualization

- **Visual cues / Visual encoding:** การแปลงข้อมูลให้แสดงในลักษณะของรูปร่าง, สี, หรือขนาด
- **Coordinate systems:** หมายถึงระบบพิกัดที่ช่วยทำให้เกิดพื้นที่ที่มีโครงสร้าง สำหรับใส่ข้อมูลที่ต้องการแสดง รูปแบบพื้นฐานที่เราเห็นได้ทั่วไปคือ ระบบพิกัดที่มีแกน X แกน

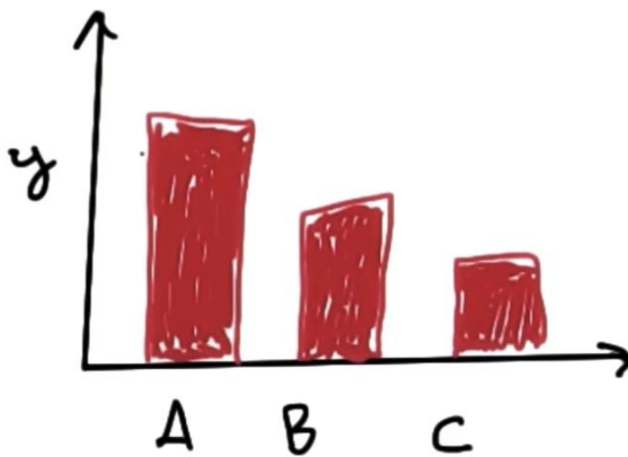
ปัจจัยประสิทธิผลในการทำ Visualization

- **Scale / Data types:** จะต้องสอดคล้องกับระบบพิกัดข้างต้น โดย Data types พื้นฐานนั้นมีอยู่ 3 ประเภทได้แก่ ข้อมูลเชิงตัวเลข (Numeric), ข้อมูลที่เป็นหมวดหมู่ (Categorical), และ
- **Context:** ในกรณีที่ผู้รับสารไม่คุ้นเคยกับข้อมูล เป็นหน้าที่ของ Data Scientist ที่จะต้องระบุบริบทของข้อมูลให้ชัดเจน และอธิบายว่าควรจะอ่านข้อมูลจากกราฟอย่างไร อย่างเช่น การตั้งชื่อกราฟ, ระบุแกน X แกน Y ว่าคืออะไร, หรือการระบุเหตุการณ์บางอย่างที่สำคัญลงบนแกนลำดับเวลา เพื่อให้ผู้อ่านสามารถตีความและเข้าใจได้มากขึ้นข้อมูลตามลำดับเวลา (Time Series)

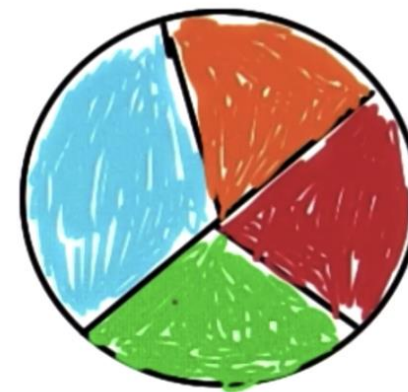
Visual Encoding



Position



Length



Angle

Visual Encoding

- Position:** การระบุตำแหน่งของข้อมูลแต่ละจุด ทำให้สามารถเปรียบเทียบข้อมูลแต่ละจุดได้บนระบบพิกัด (Coordinate System) เดียวกัน การแปลงข้อมูลให้อยู่ในรูปแบบของตำแหน่งนั้นใช้พื้นที่น้อยเมื่อเทียบกับการเข้ารหัสข้อมูลแบบอื่น ๆ และเนื่องด้วยที่จุดแต่ละจุดมีขนาดเท่ากัน ทำให้สามารถสังเกตเห็นแนวโน้ม (Trend), การเกาะกลุ่ม (Cluster), หรือจุดที่แปลกแยก (Outlier) ของข้อมูลได้ง่าย อย่างไรก็ตาม หากมีจุดข้อมูลมากเกินไปอาจทำให้ไม่สามารถเข้าใจความหมายของแต่ละจุดได้
- Length:** การเข้ารหัสข้อมูลพื้นฐานอีกแบบหนึ่งคือการใช้ความยาว ซึ่งมักจะเห็นได้ในรูปแบบของกราฟแท่ง ความยาวของกราฟแท่งระบุปริมาณของข้อมูล

Visual Encoding

- **Angle:** รูปแบบนี้ก็สามารถพบเห็นได้ทั่วไปเช่นเดียวกัน ตัวอย่างที่เห็นได้ชัดก็คือ กราฟวงกลมหรือ Pie Chart นั่นเอง ข้อมูลจะอยู่ในช่วงของมุม 0 - 360 องศา ซึ่งวิธีนี้จะถูกใช้แสดงข้อมูลในแง่ของอัตราส่วน แต่ข้อเสียอย่างหนึ่งของรูปแบบนี้คือ การสังเกตเห็นความแตกต่างของมุมนั้นทำได้ยากจากการมองด้วยตา เช่น เราอาจไม่สามารถสังเกตเห็นความแตกต่างระหว่างมุม 25 องศา กับมุม 30 องศาได้เลย ดังนั้น จึงควรหลีกเลี่ยงวิธีนี้หากต้องการแสดงความแตกต่างเล็กน้อยของข้อมูล

Visual Encoding



Direction



Shape

Area / Volume

Visual Encoding

- **Direction:** การใช้ทิศทาง จะทำให้สามารถสังเกตความชันและรู้ได้อย่างง่ายดายว่า ข้อมูลมีการเพิ่มขึ้น ลดลง หรือมีความผันผวนอย่างไร การเข้ารหัสข้อมูลโดยใช้ทิศทางมีความคล้ายกับการใช้มุม (Angle) โดยเป็นไปได้ยากที่จะแยกความแตกต่างระหว่างเส้นที่มีความชันใกล้เคียงกัน
- **Shape:** รูปร่างหรือสัญลักษณ์สามารถใช้ในการจัดหมวดหมู่ให้กับข้อมูล หรือแยกข้อมูล 2 ชุดเพื่อเปรียบเทียบกันได้
- **Area / Volume:** เรายังสามารถใช้พื้นที่ หรือขนาดในการเข้ารหัสข้อมูลได้ ขนาดที่ใหญ่กว่าแสดงถึงค่าที่มากกว่า

Visual Encoding

- **Color:** เราสามารถเข้ารหัสข้อมูลโดยใช้สีได้ ซึ่งสามารถใช้ได้ทั้ง Hue และ Saturation โดยสามารถใช้อันใดอันหนึ่งหรือใช้ทั้งคู่ประกอบกันก็ได้



Data Types: ประเภทข้อมูลพื้นฐาน

- **Numeric:** ใช้ในการแสดงข้อมูลเชิงปริมาณ (Quantitative) อย่างไม่อย่างหนึ่ง อาจจะเป็นความสูง, น้ำหนัก, หรือการนับจำนวนโฮมรัน โดยเป็นได้ทั้ง Discrete และ Continuous ตัวอย่างเช่น จำนวนโฮมรันนับเป็น Discrete เพราะแสดงจำนวนตัวเลขนั้น ๆ โดยเฉพาะ (10, 34, 25) แต่อัตราการตีได้ (Hit Rate) นับเป็น Continuous เพราะแทนค่าตัวเลขในช่วงนั้น ๆ (.250, .357, .511)

Data Types: ประเภทข้อมูลพื้นฐาน

- **Categorical:** ใช้ในการแสดงข้อมูลคุณลักษณะ เช่น ตำแหน่งการเล่นในทีมเบสบอล, ทีมเบสบอลที่สังกัด, บ้านเกิด, มือข้างที่ถนัด เป็นต้น มีข้อมูลประเภทหนึ่งที่ถูกเรียกว่า Ordinal Data ซึ่งเหมือนการผสมระหว่าง Numeric และ Categorical แต่ถูกจัดไว้ในประเภทของ Categorical Data หมายถึงข้อมูลที่มีลำดับอย่างใดอย่างหนึ่ง ตัวอย่างของข้อมูลประเภทนี้ ได้แก่ ความแรงในการตีลูกเบสบอลอาจเรียงได้จาก Very Low > Low > Average > High > Very High หรือการให้คะแนนภาพยนตร์เรื่องหนึ่งๆ ตั้งแต่ 1 ดาว ไปจนถึง 5 ดาว

Data Types: ประเภทข้อมูลพื้นฐาน

- **Time Series:** หมายถึงข้อมูลที่ถูกเก็บตามลำดับเวลา อย่างเช่น การเก็บข้อมูลจำนวนโฮมรันที่เกิดขึ้นในแต่ละปี จะเห็นได้ว่าจริง ๆ แล้วข้อมูลที่เก็บนับเป็น Numeric Data แต่การเก็บข้อมูลตามลำดับเวลาทำให้ข้อมูลเหล่านั้นเกิดลำดับขึ้นมาอย่างหนึ่ง ซึ่งนั่นก็คือ ลำดับเวลา

ชนิดของ Visualizations

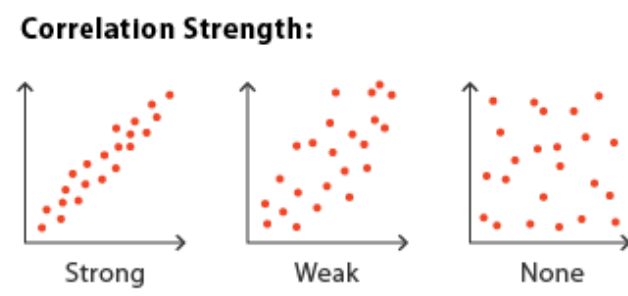
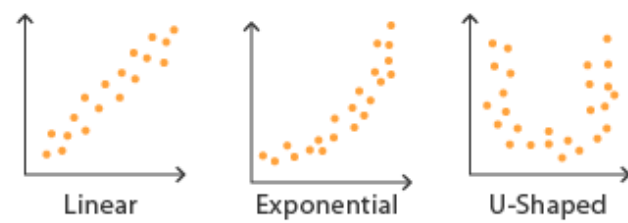
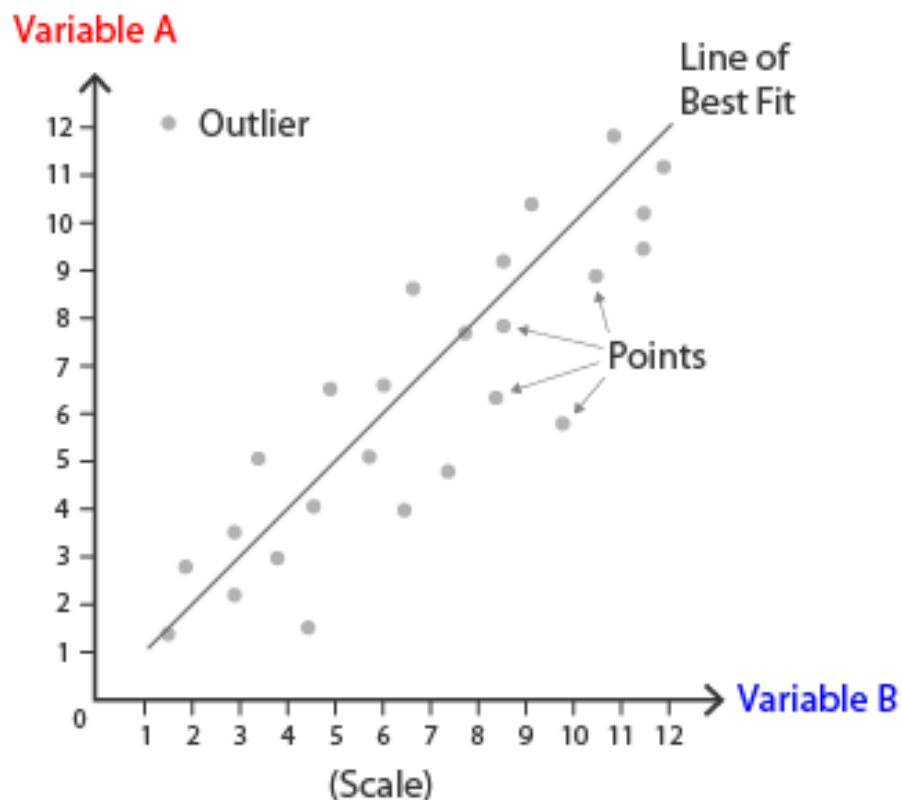
1. Temporal
2. Hierarchical
3. Network
4. Multidimensional
5. Geospatial

Temporal

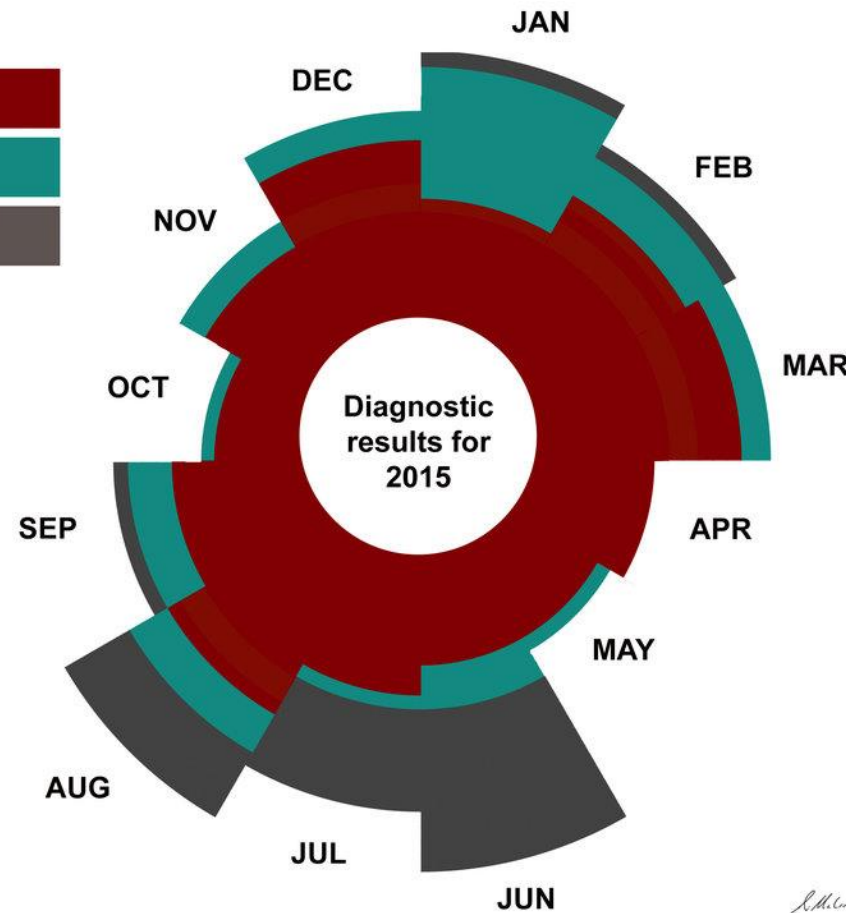
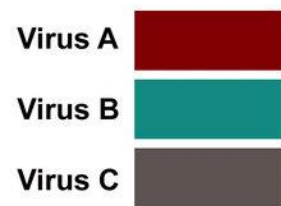
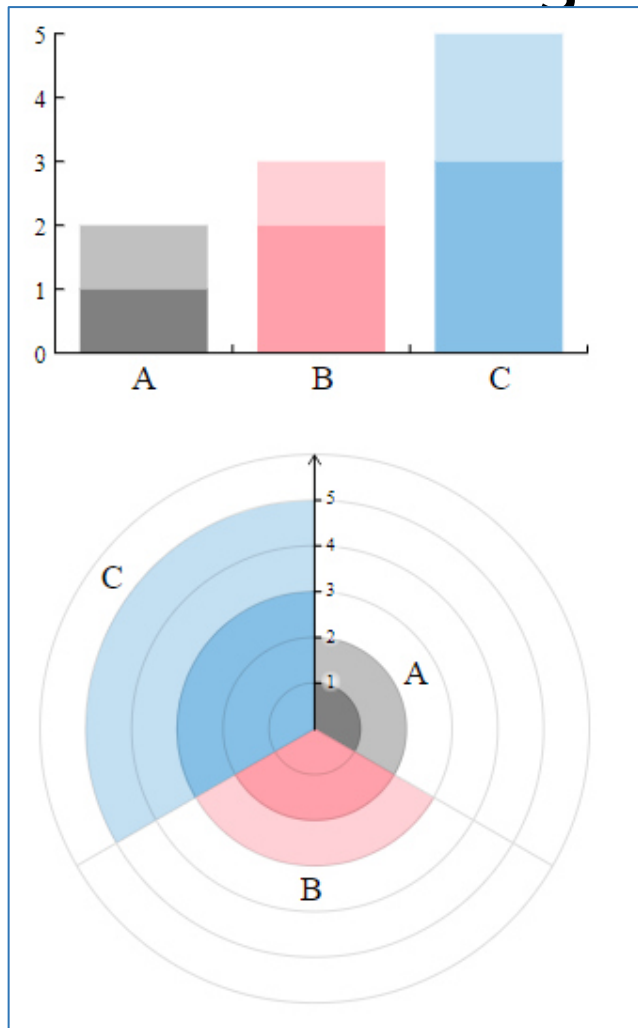
- การสร้างภาพเชิงเวลา (หรือไทม์ไลน์) คล้ายกับการสร้างภาพข้อมูลเชิงเส้นแบบหนึ่งมิติ เนื่องจากไทม์ไลน์มีการใช้กันอย่างแพร่หลายและมีความสำคัญ เช่น ข้อมูลเวชระเบียน การจัดการโครงการ และการนำเสนอข้อมูลในอดีต ถือว่าเป็นประเภทข้อมูลแยก
- ข้อมูลเชิงเวลามีลักษณะรายการที่มีเวลาเริ่มต้นและสิ้นสุดและรายการอาจทับซ้อนกัน การแสดงข้อมูลเส้นเวลามักจะรวมเหตุการณ์ทั้งหมดก่อนหลังหรือระหว่างช่วงเวลาหรือช่วงเวลา
- เช่น
 - Scatter plots
 - Polar area diagrams
 - Time series sequences
 - Timelines
 - Line graphs

Scatter plots

- แผนภาพกระจาย(scatter plot) จะใช้แสดงภาพตัวแปรสองตัว พร้อมกัน เช่น อาจจะศึกษาความสัมพันธ์ของตัวแปรสองตัว หรือสามตัว ข้อมูลที่เราได้จากการศึกษาก็จะมีสองข้อมูล (Rivariate data)

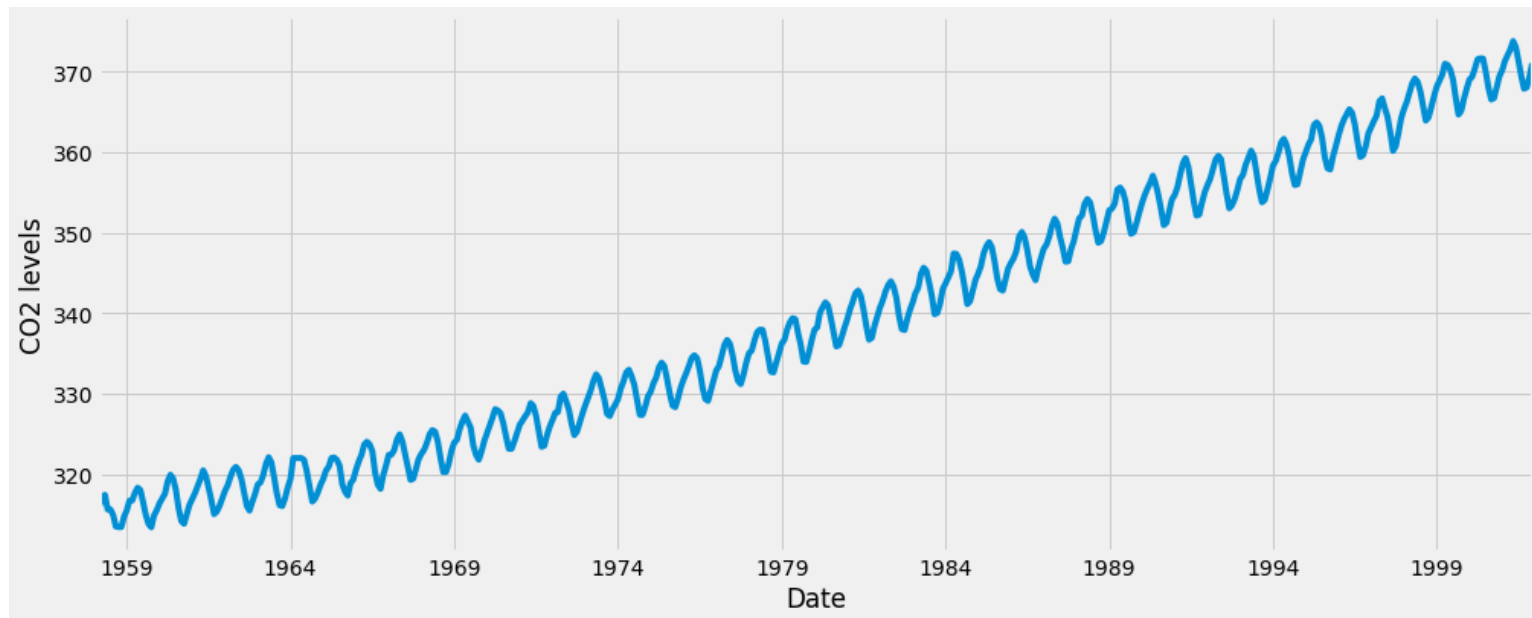
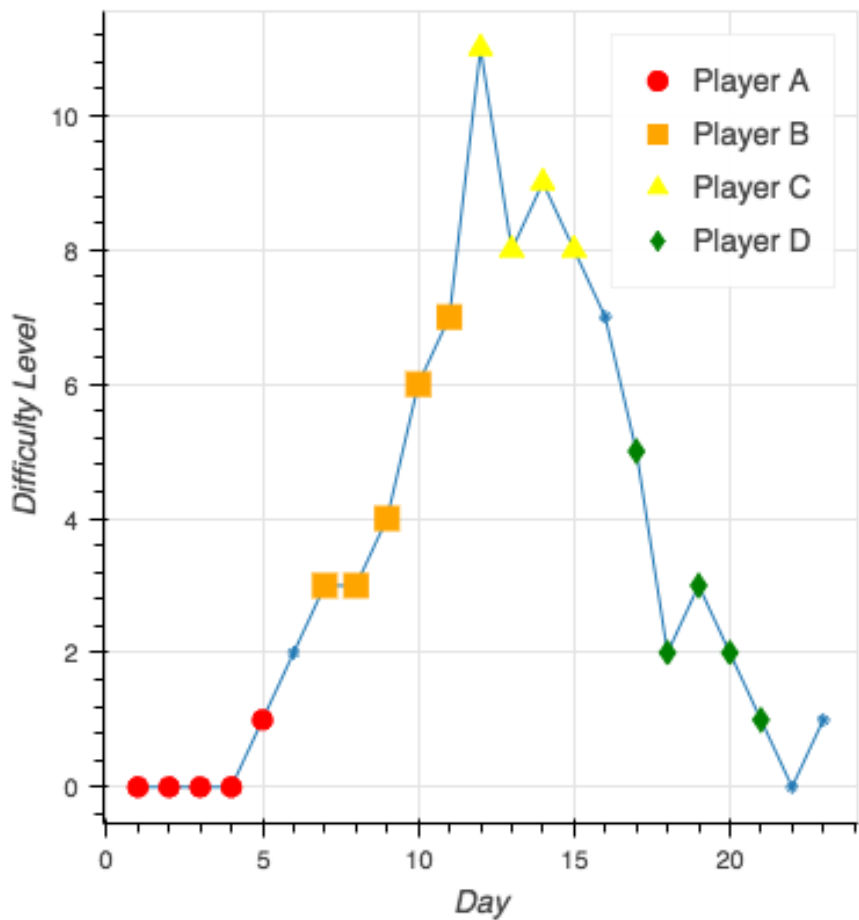


Polar area diagrams



S.M. Co.

Time series sequences

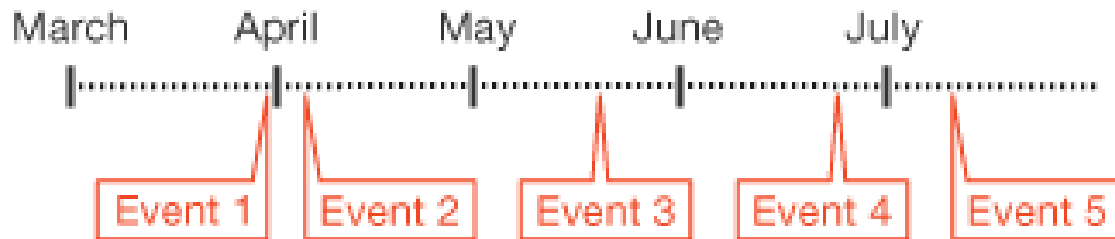


Timelines

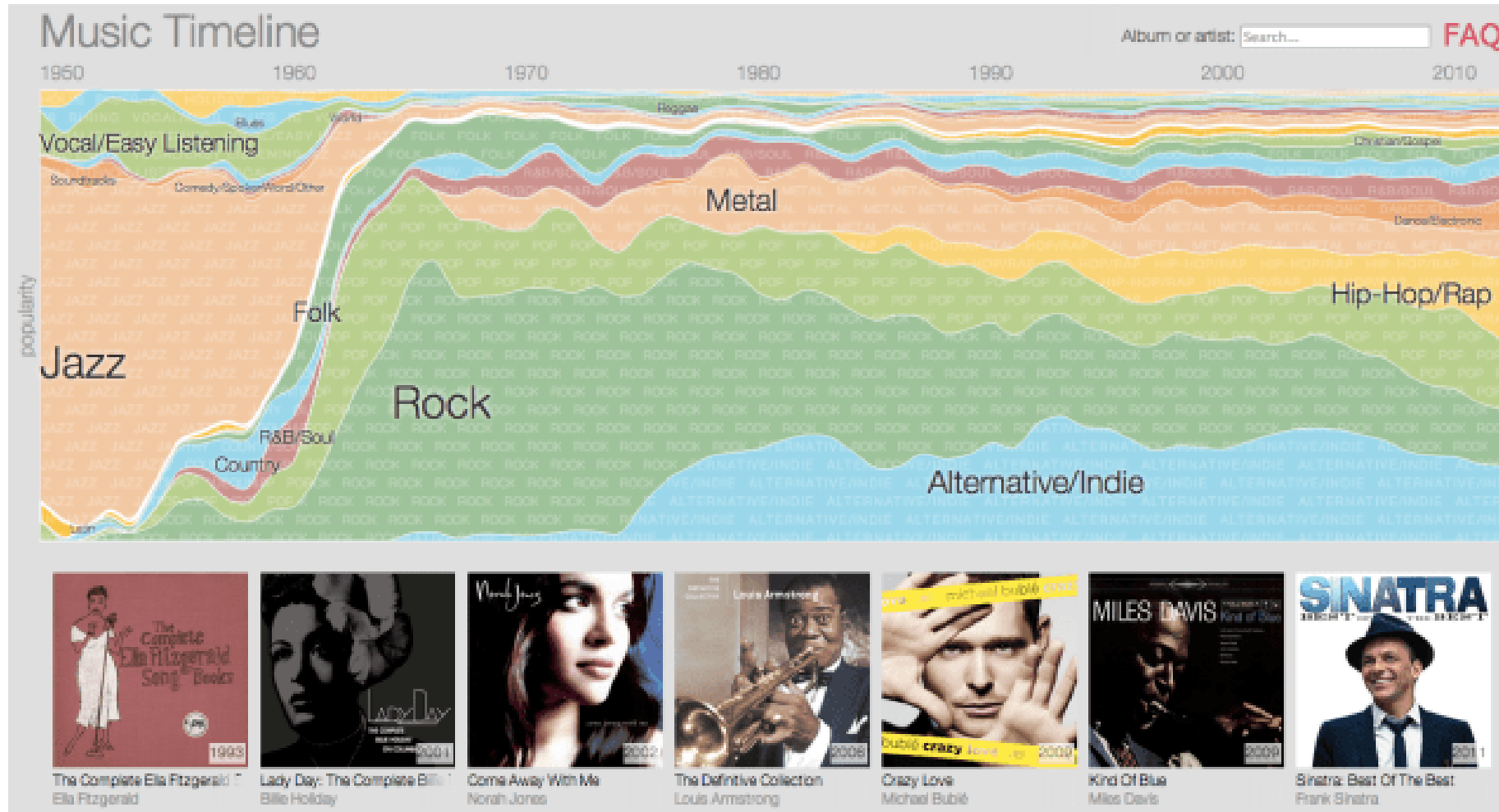
Sequential Timeline:



Scaled Timeline:

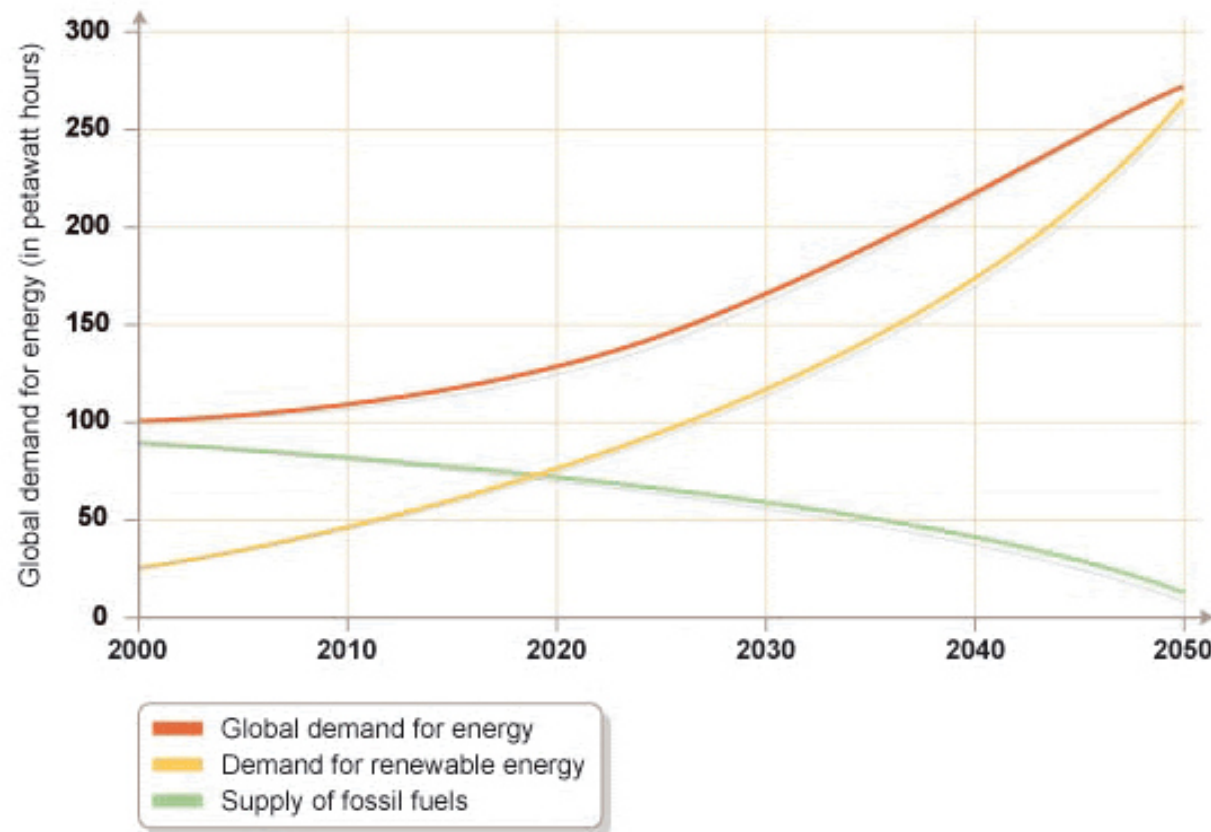
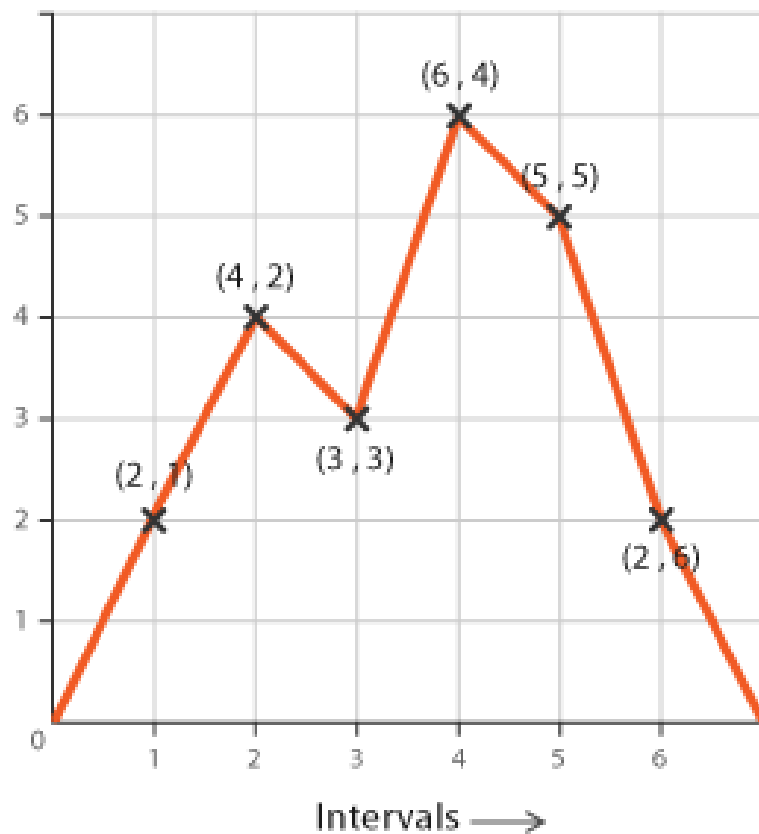


Timelines



Line graphs

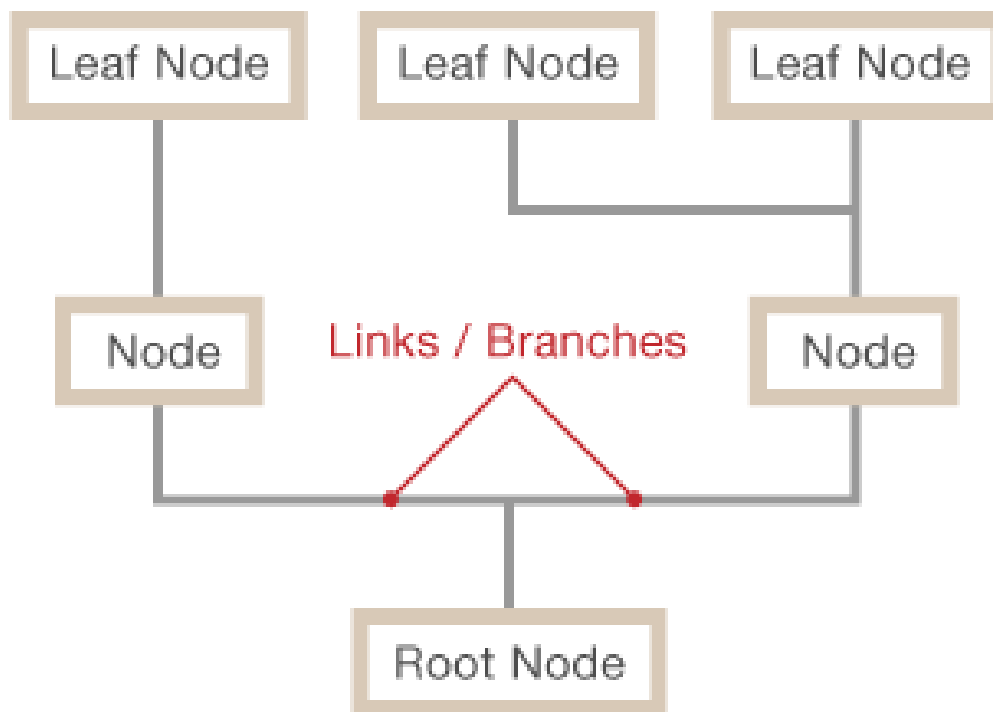
Value Scale



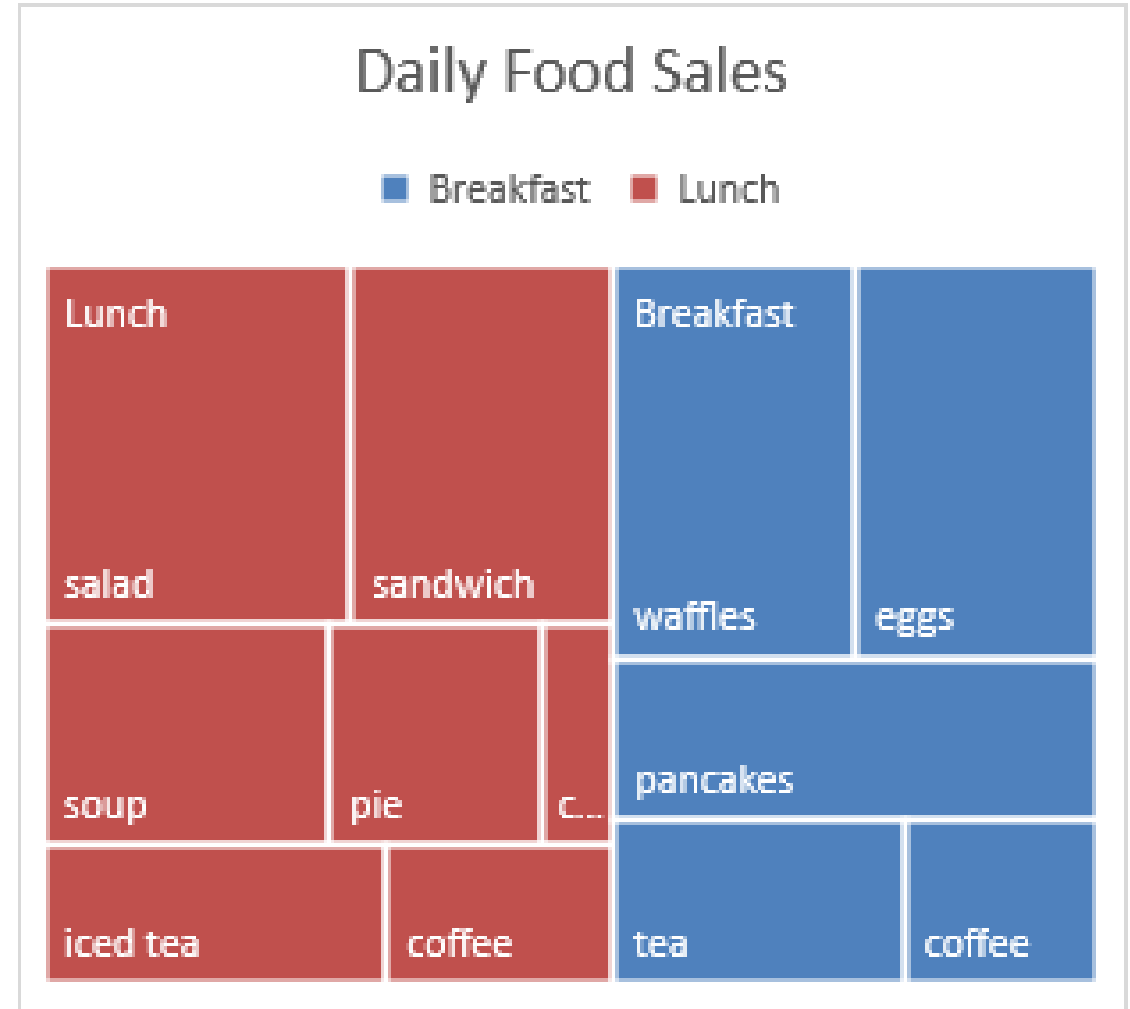
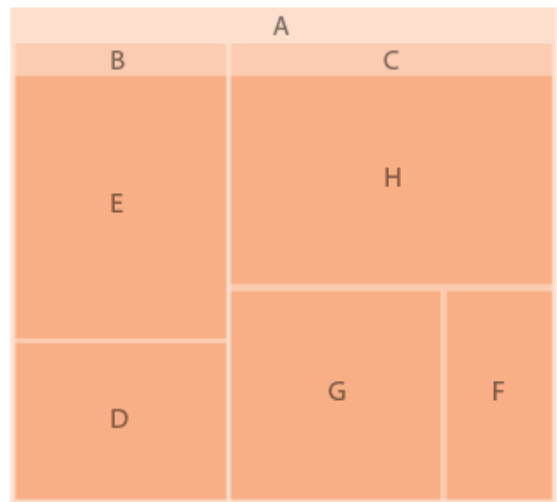
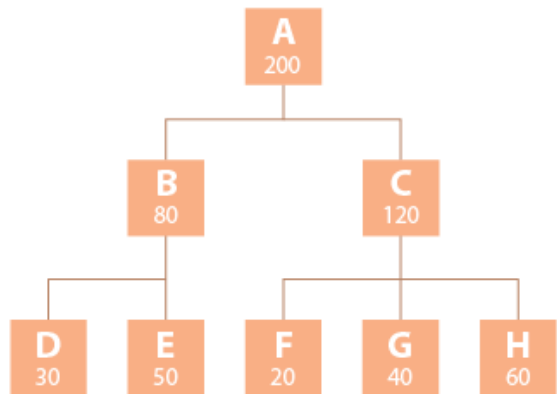
Hierarchical

- การสร้างภาพข้อมูล ประเภทลำดับชั้น (Hierarchical) จะใช้กับข้อมูลที่สามารถจัดเรียงในรูปแบบของต้นไม้ ข้อมูลแต่ละรายการจะกำหนดโหนดในทรีและแต่ละโหนดอาจมีการรวบรวมโหนดอื่น ๆ เป็นโหนดย่อย ความสัมพันธ์ระหว่างโหนดพารেন্টและโหนดย่อยเป็นเครือข่ายแบบต้นไม้
- เป็นเทคนิคการสร้างภาพข้อมูลสำหรับชุดข้อมูลที่มีลำดับชั้นขนาดใหญ่ มักจะเลือกใช้แผนภาพต้นไม้ (Treemap) เนื่องจากเป็นวิธีที่ง่ายที่สุดในการติดตามจากเส้นทางเชิงเส้น
- เช่น
 - Tree diagrams (Treemap)
 - Ring charts
 - Sunburst diagrams
- ข้อด้อยของกราฟเหล่านี้คือพวกเขาจะมีความซับซ้อนและอ่านยาก

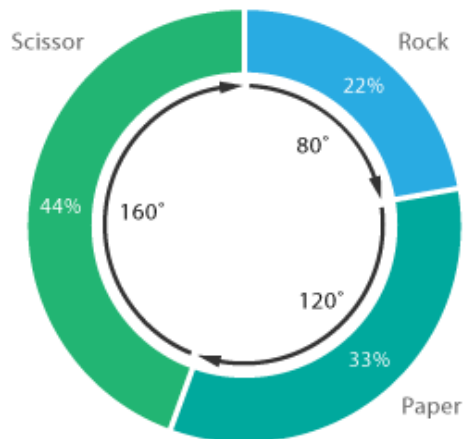
Hierarchical



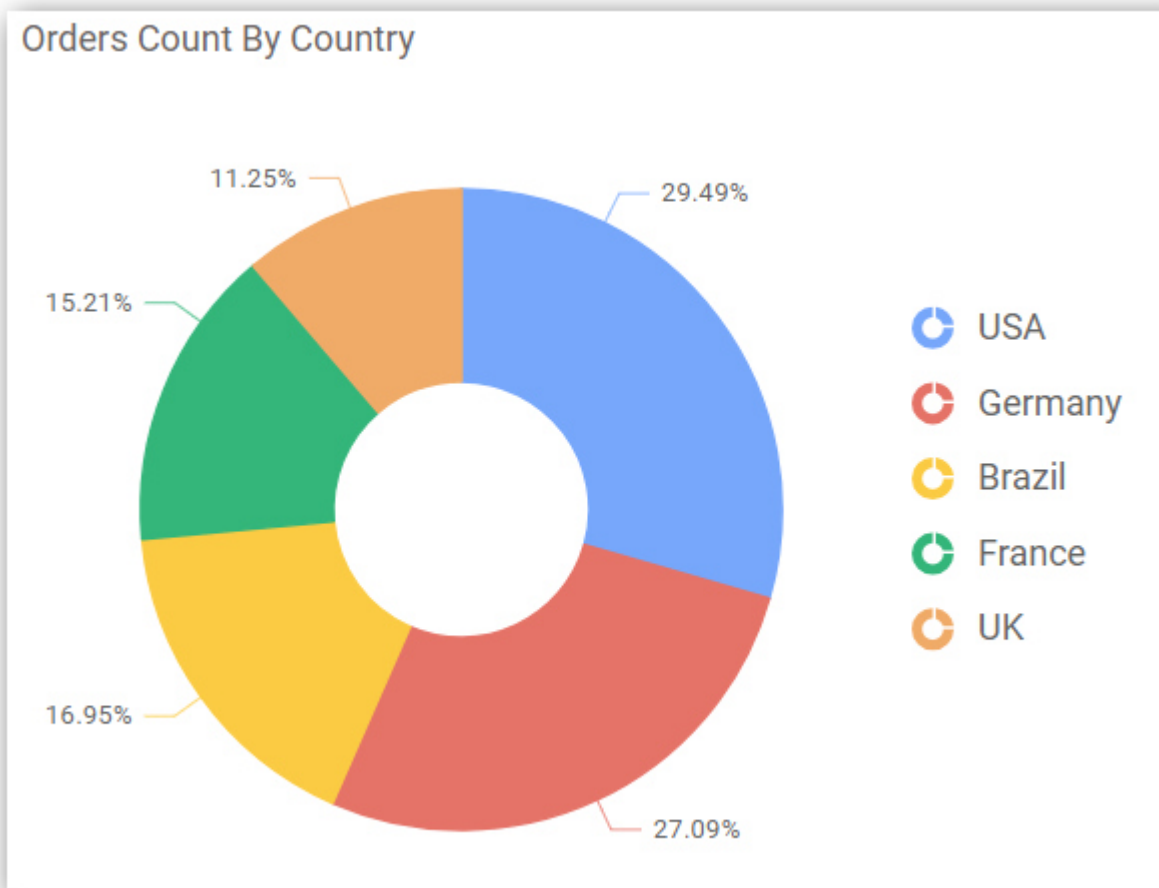
Tree diagrams (Treemap)



Ring charts /Donut Chart

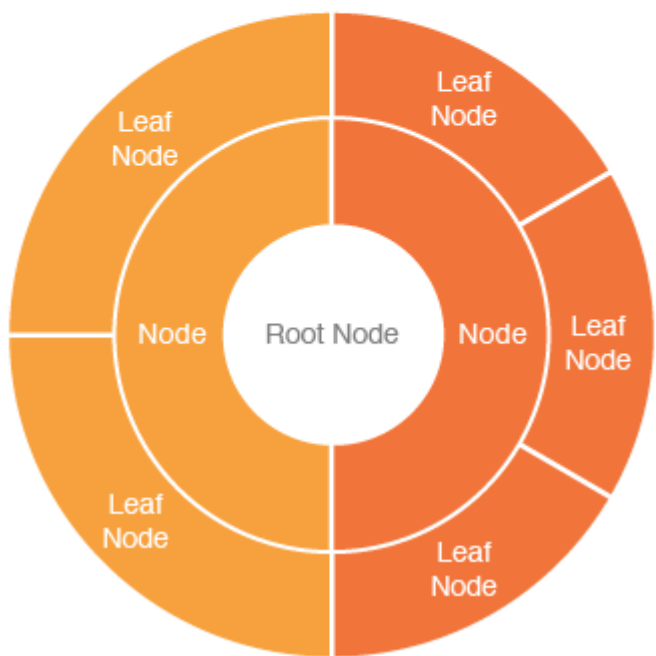
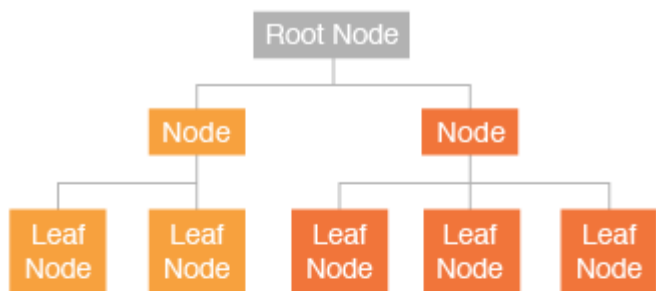


Data			
Rock	Paper	Scissor	TOTAL
2	3	4	9
To calculate percentages			
$2/9 = 22\%$	$3/9 = 33\%$	$4/9 = 44\%$	100%
Degrees for each "donut slice"			
$(2/9) \times 360 = 80^\circ$	$(3/9) \times 360 = 120^\circ$	$(4/9) \times 360 = 160^\circ$	360°



- USA
- Germany
- Brazil
- France
- UK

Sunburst diagrams

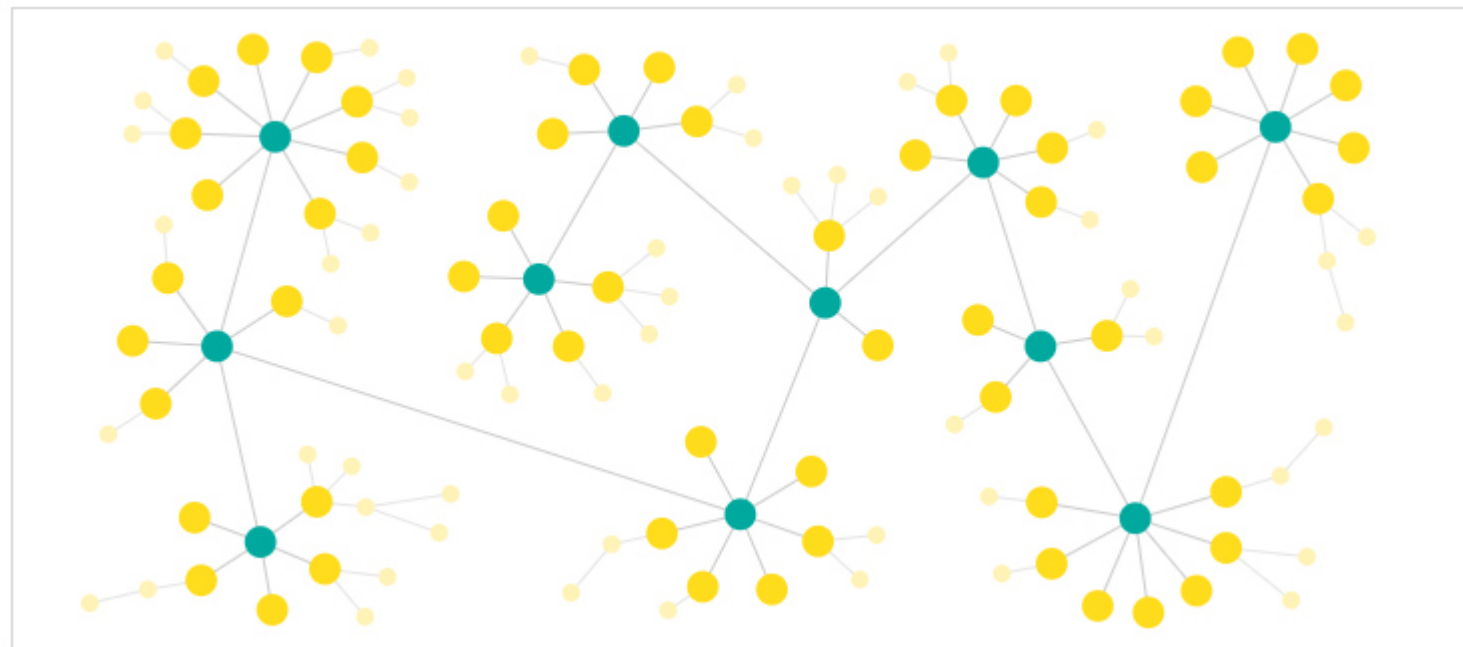
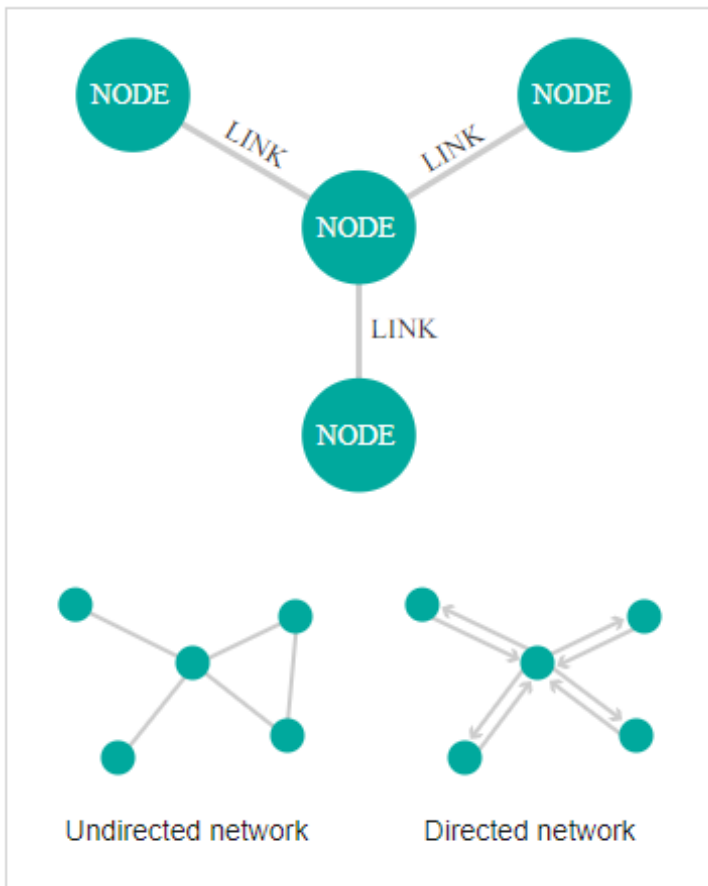


- USA
- India
- Germany
- UK
- France



Network Data Visualization

- เป็นกระบวนการของการนำเสนอเครือข่ายที่เชื่อมต่อกันด้วยการมองเห็นเป็นลิงค์และโหนด

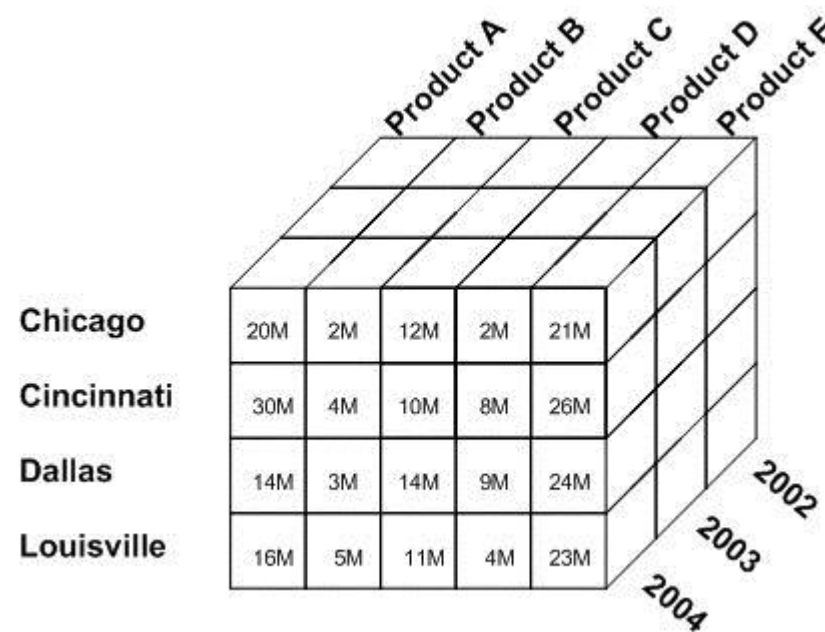
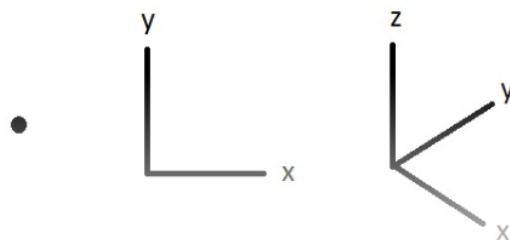


Multidimensional

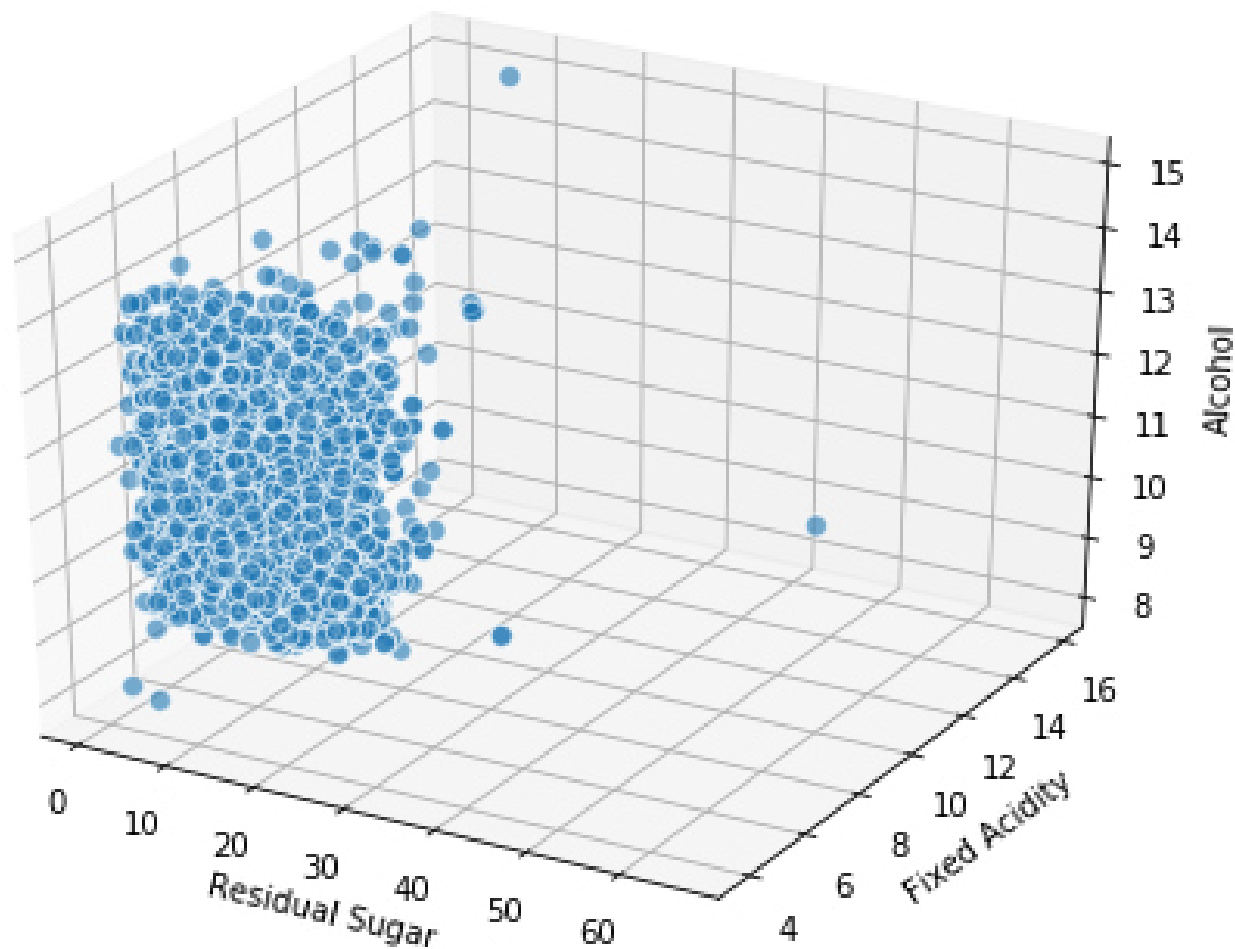
• Multidimensional จะสัมพันธ์กับมิติของตัวแปรข้อมูลที่เราสนใจ เช่น 3 ตัวแปร ก็จะใช้ภาพแบบ 3 มิติ

• เช่น

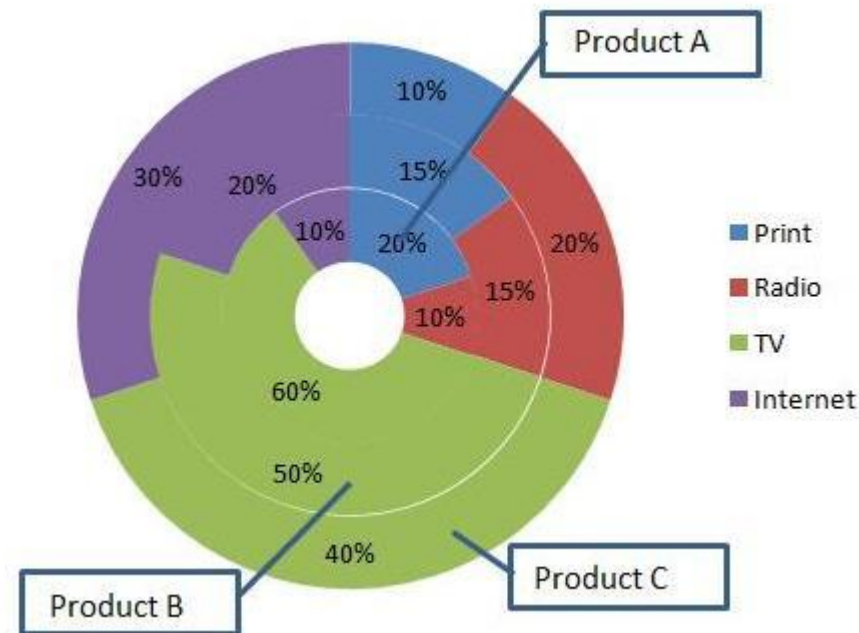
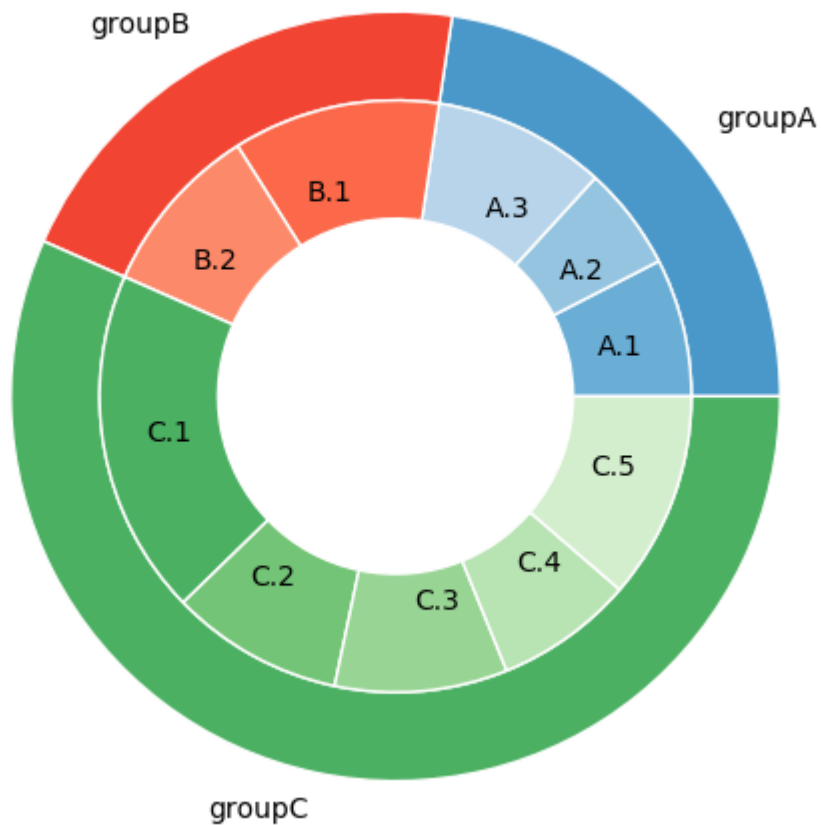
- Scatter plots
- Pie charts
- Venn diagrams
- Stacked bar graphs
- Histograms



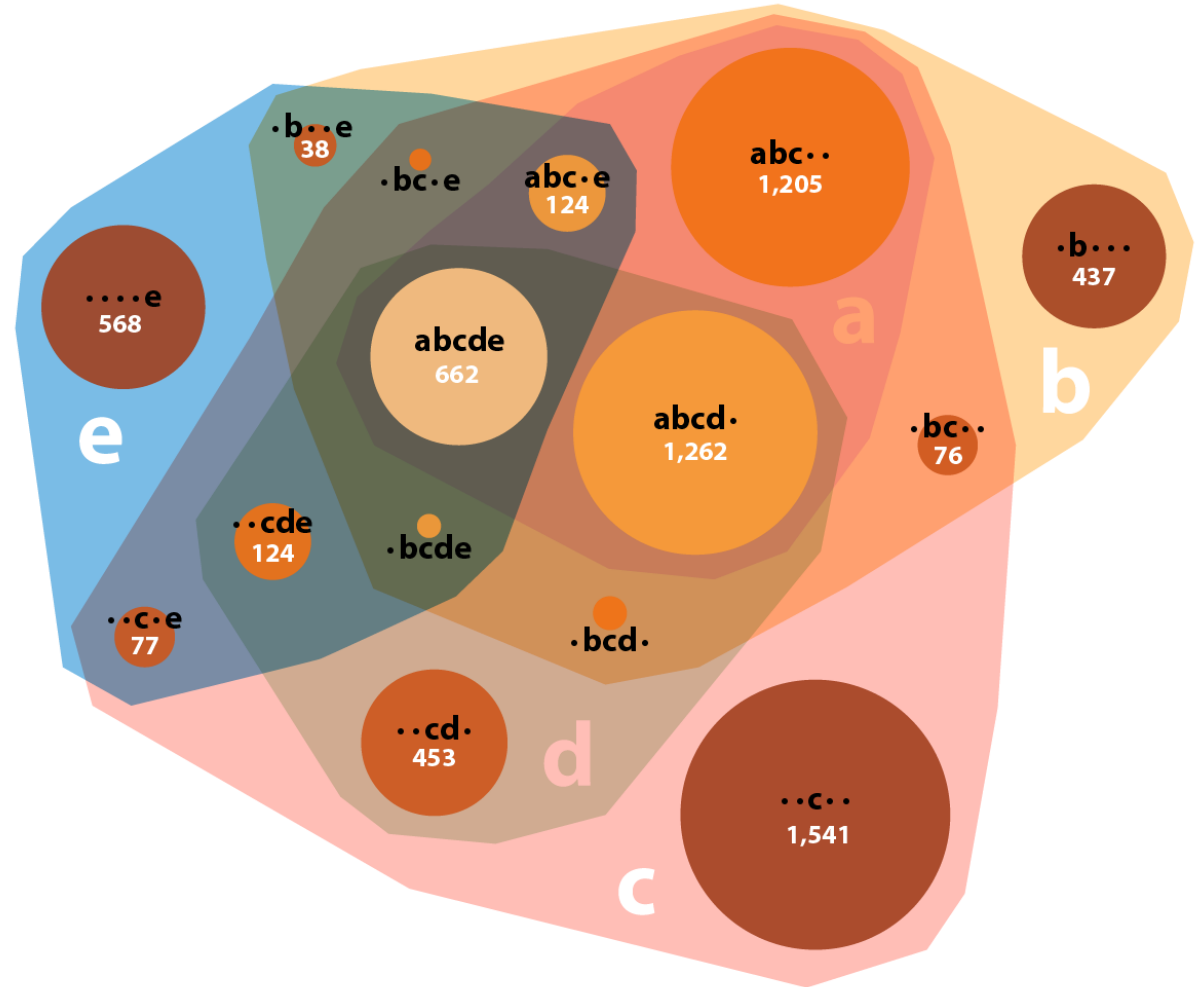
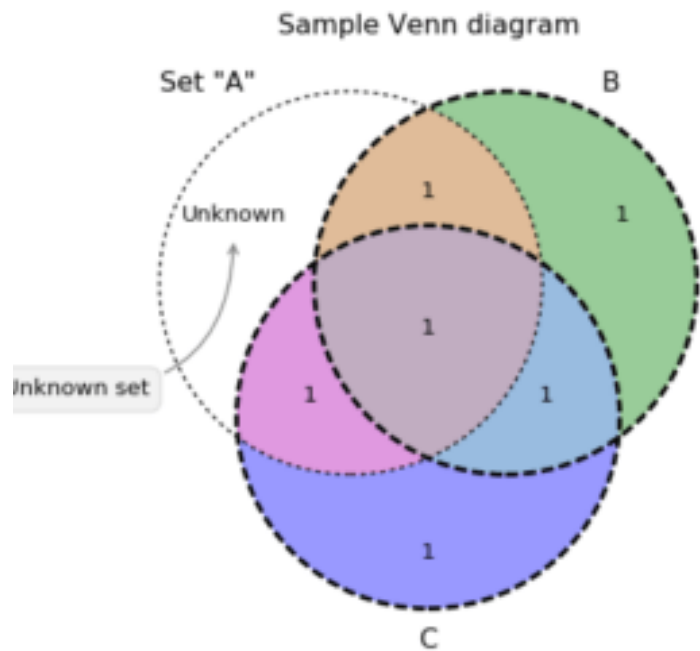
Scatter plots Multidimensional



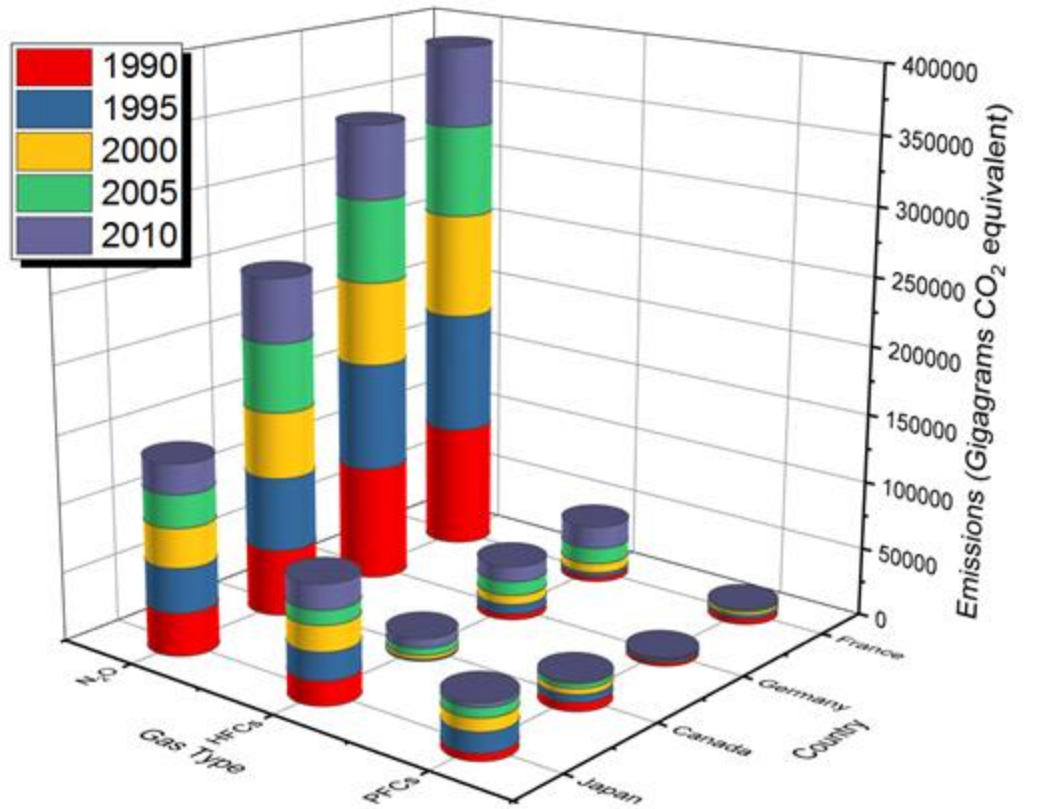
Pie charts Multidimensional (Multilevel)



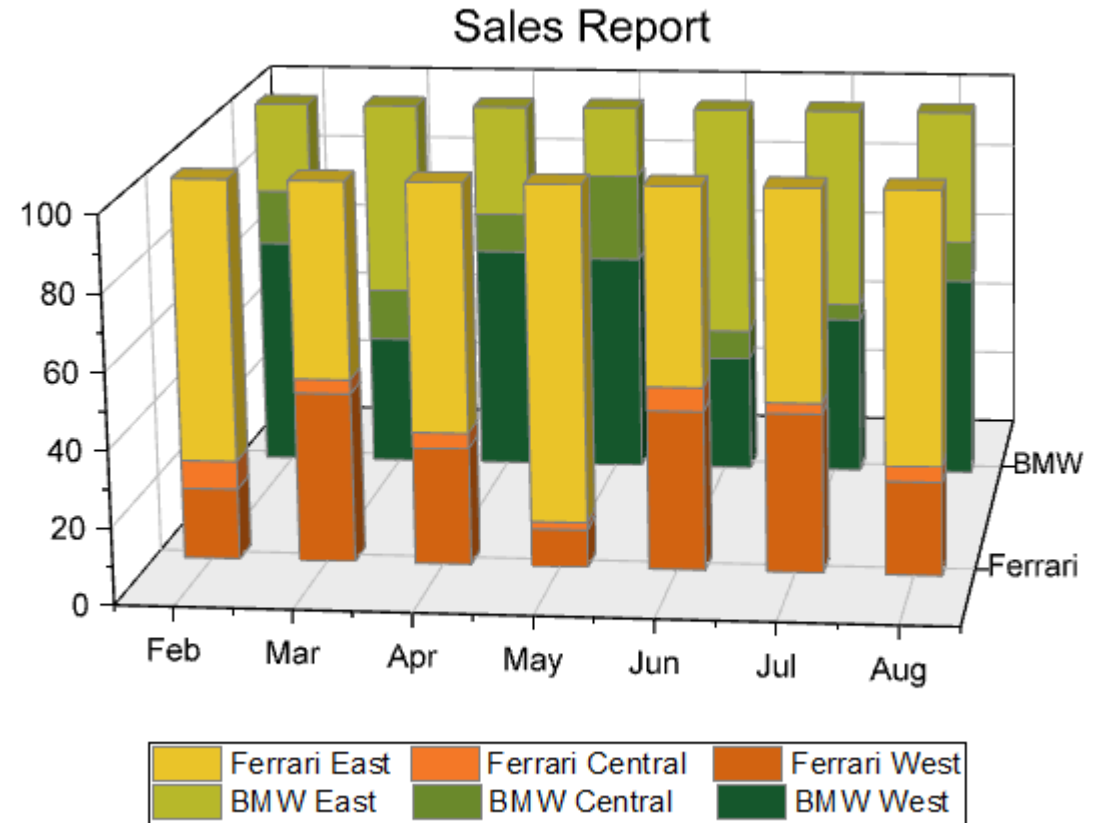
Venn diagrams Multidimensional



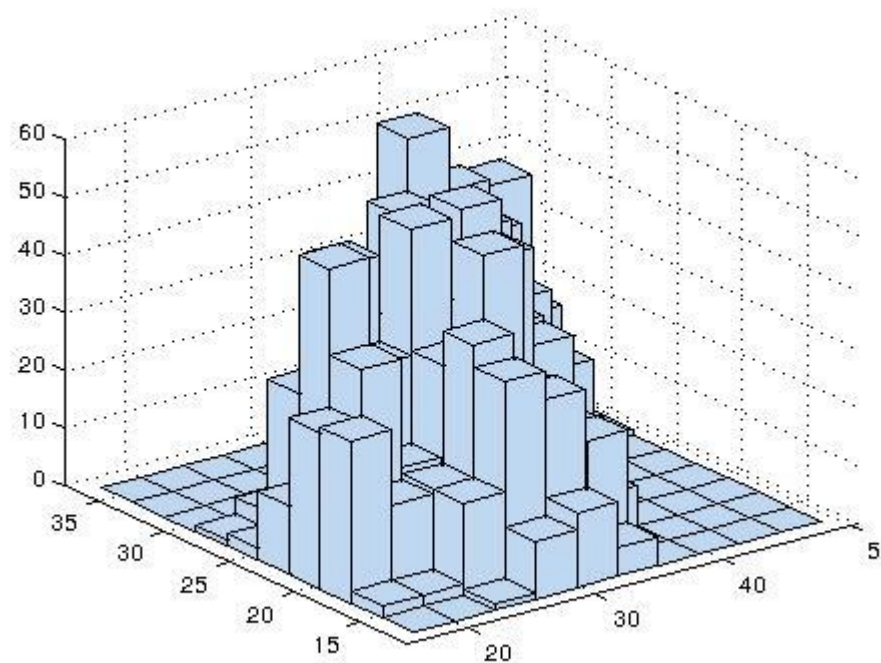
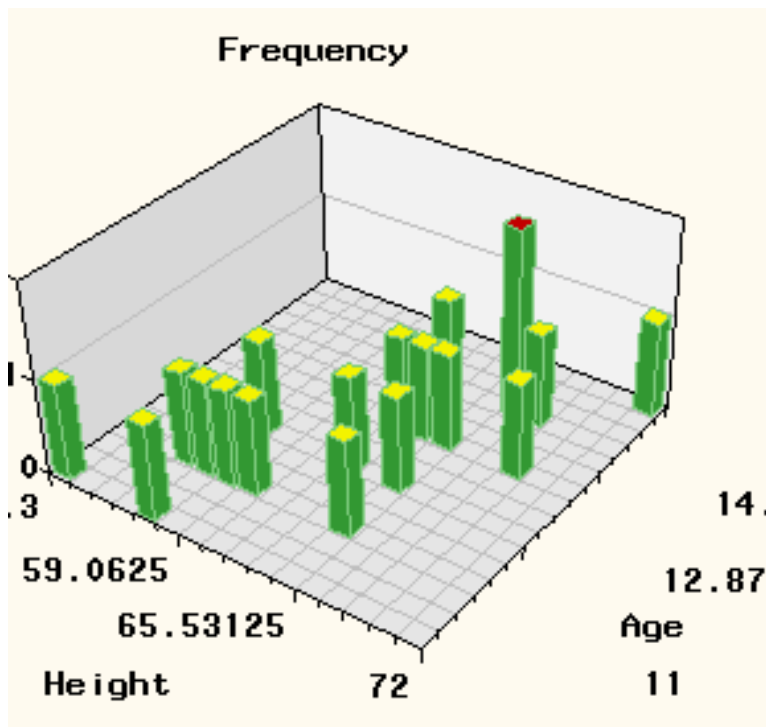
Stacked bar graphs Multidimensional



Source - <http://data.un.org>

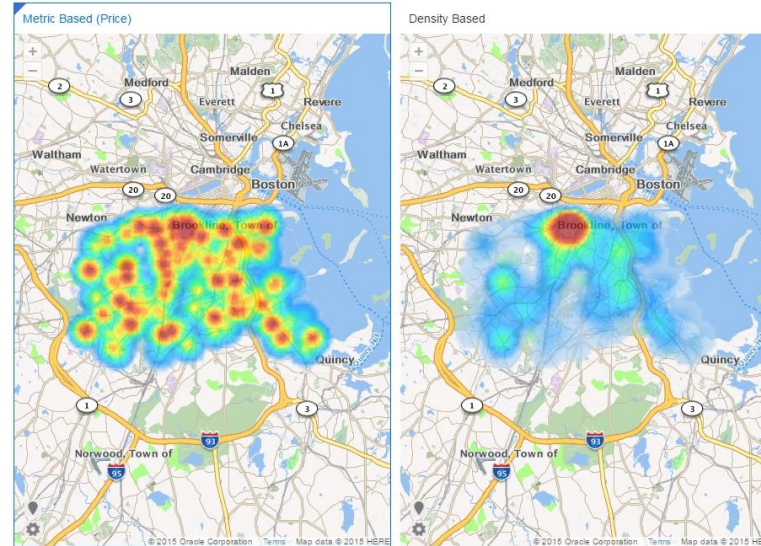


Histograms Multidimensional

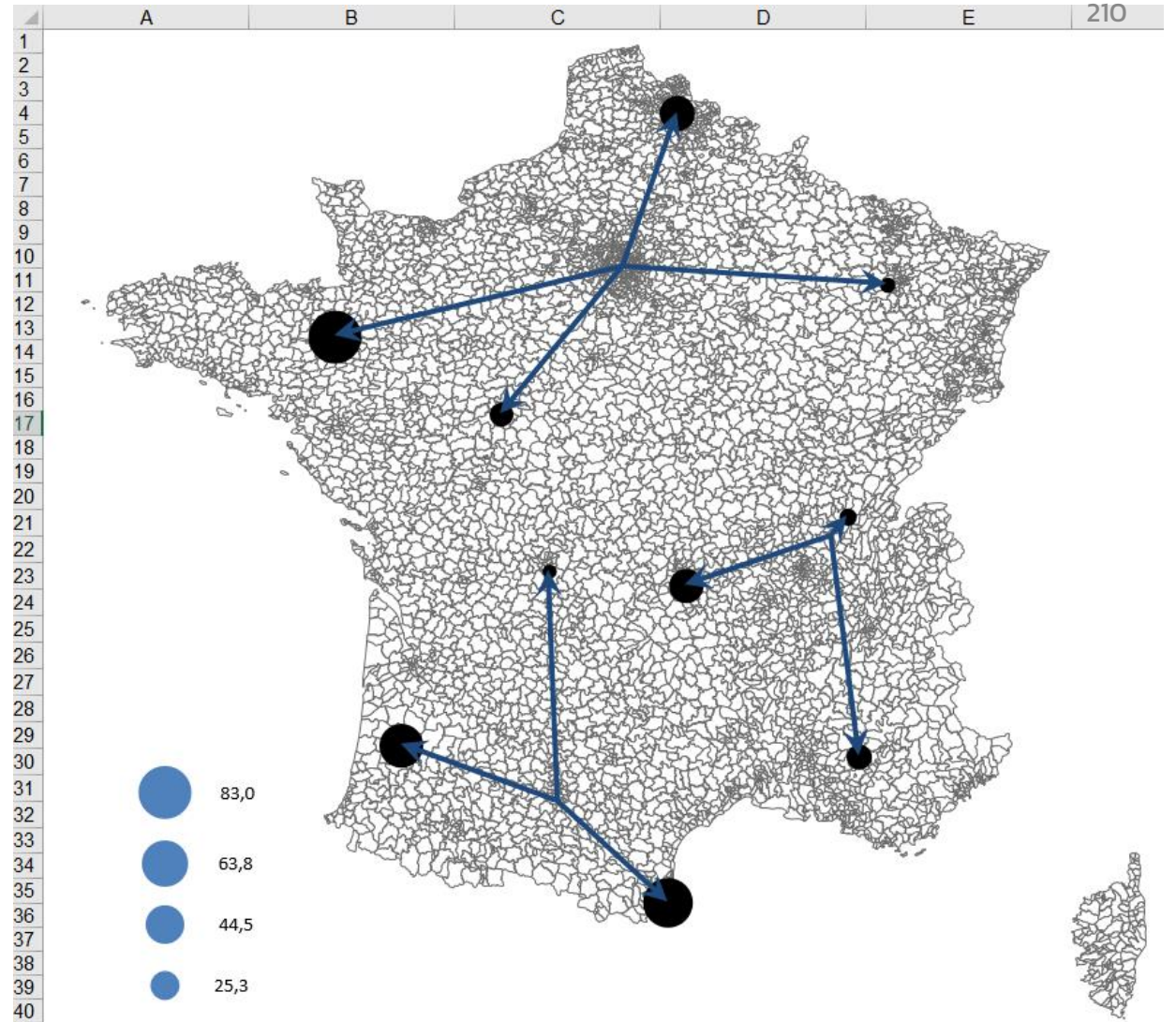
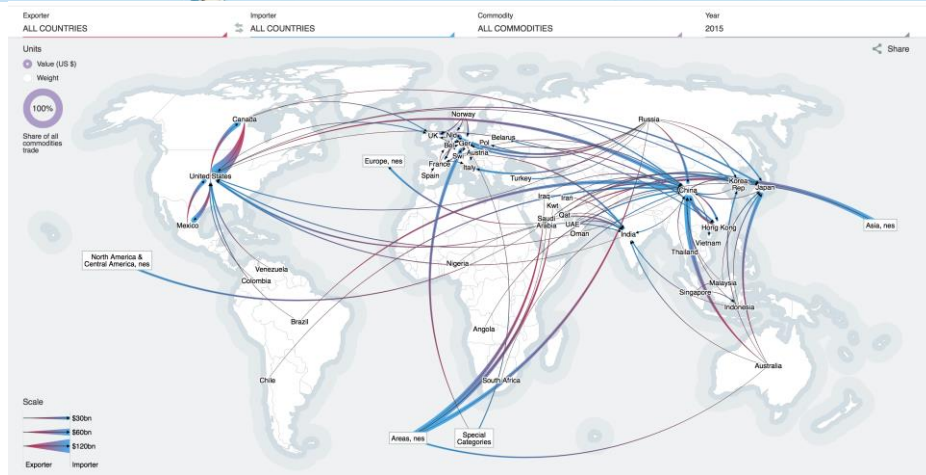


Geospatial

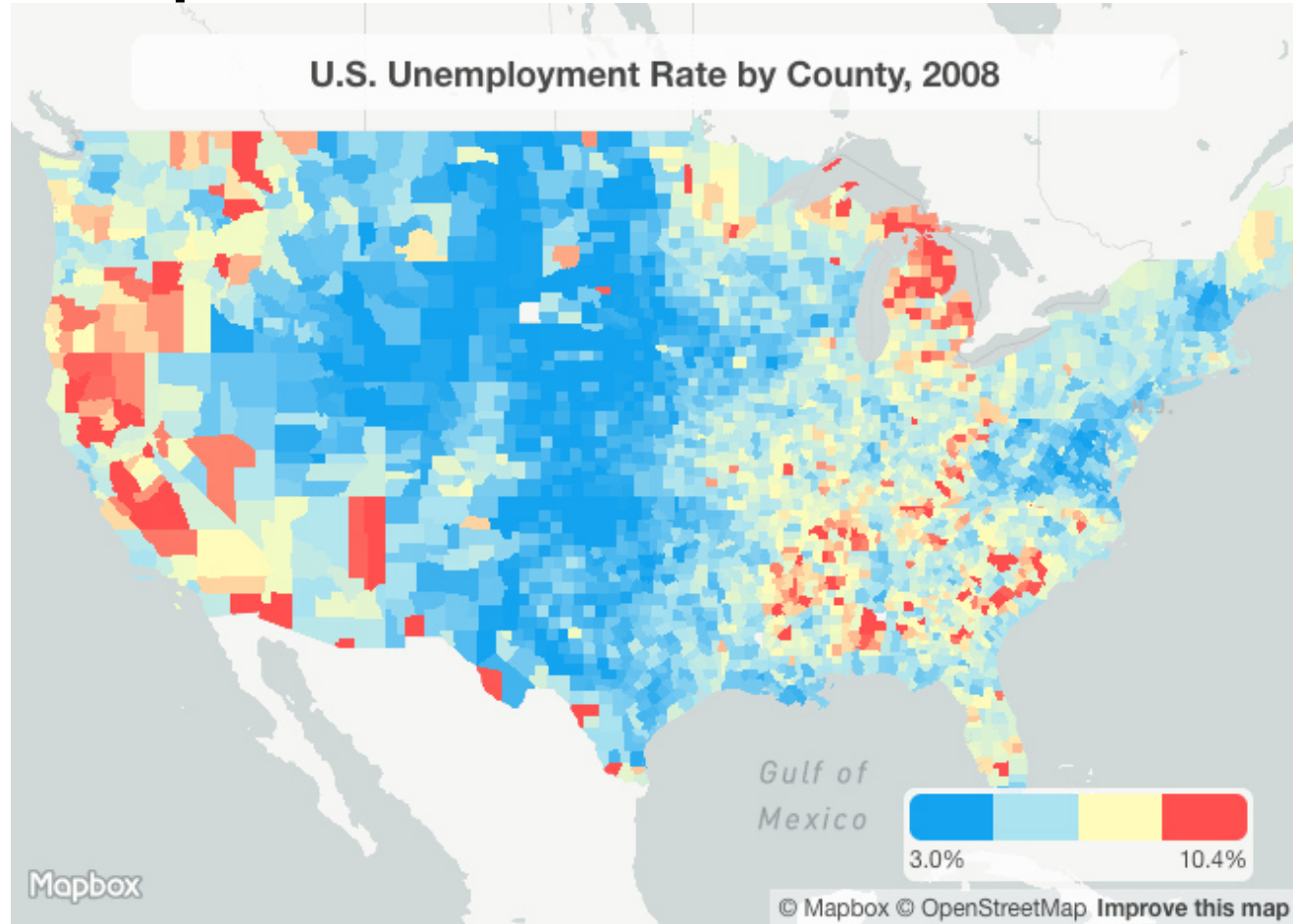
- การสร้างภาพข้อมูลเชิงพื้นที่ จะเกี่ยวข้องกับสถานที่ทางกายภาพในชีวิตจริงซ้อนทับแผนที่ โดยมีจุดข้อมูลที่แตกต่างกัน
- การแสดงข้อมูลด้วยภาพเหล่านี้มักใช้เพื่อแสดงยอดขายหรือการได้มาในช่วงเวลาหนึ่งและสามารถเป็นที่รู้จักมากที่สุดสำหรับการใช้ในแคมเปญทางการเมืองหรือเพื่อแสดงส่วนแบ่งการตลาดใน บริษัท ข้ามชาติ เป็นต้น
- เช่น
 - Flow map
 - Density map
 - Cartogram
 - Heat map



Flow map

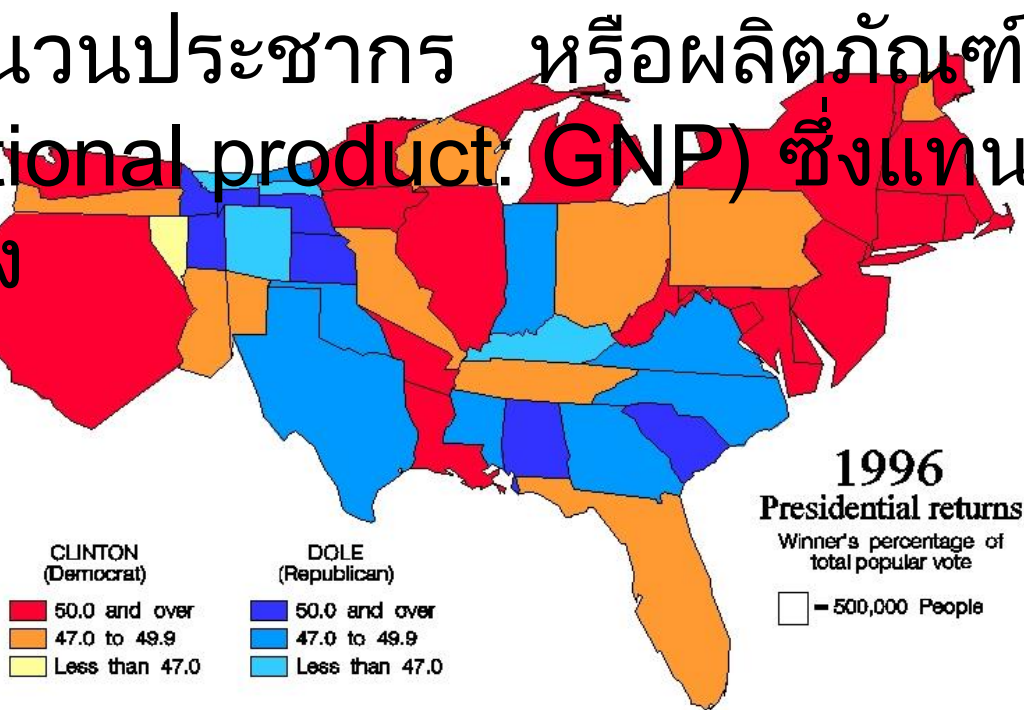


Density map



Cartogram

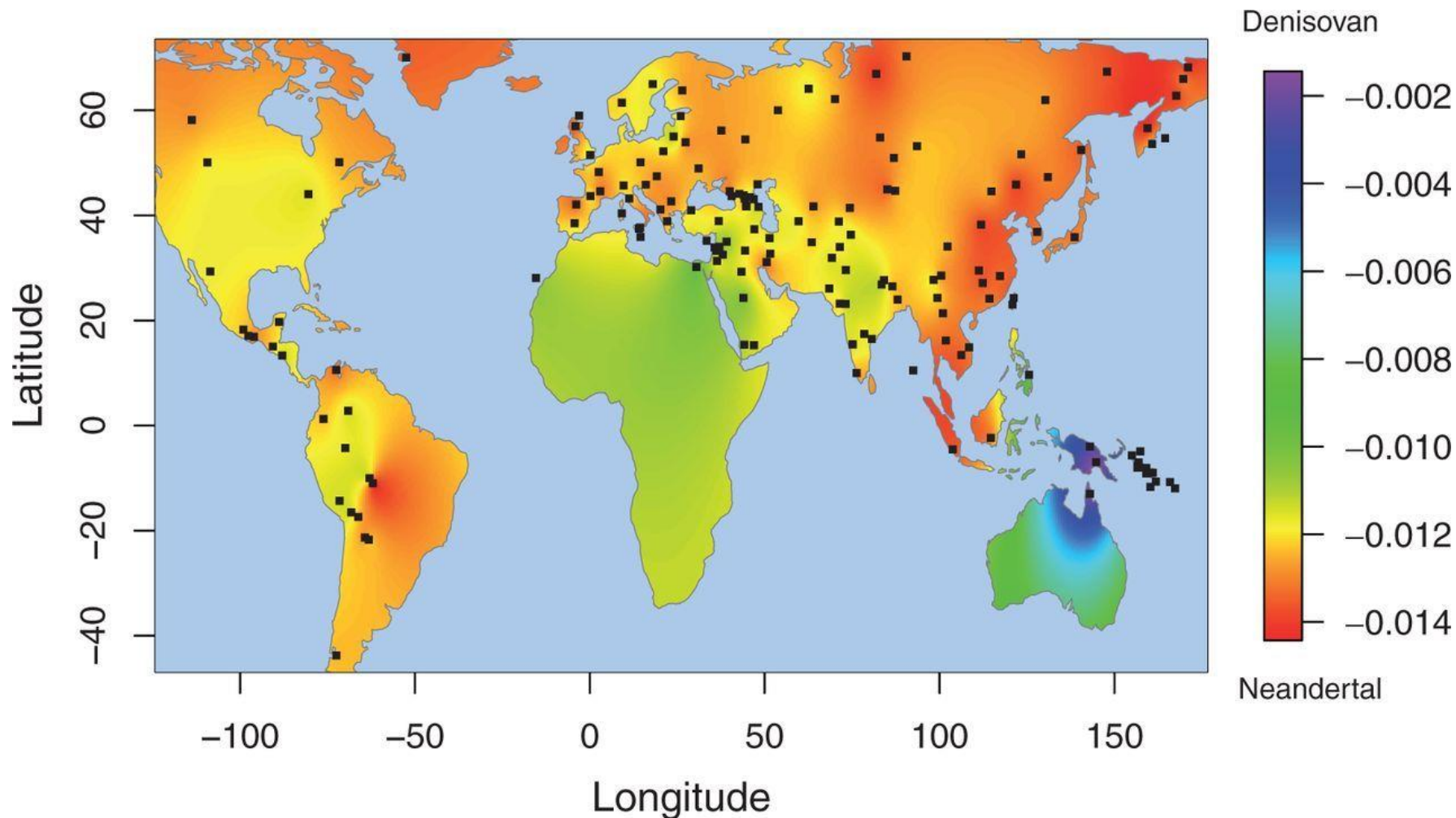
การนำเสนอข้อมูลสถิติในรูปแบบของแผนที่ซึ่งสามารถเปลี่ยนแปลง
 รูปทรงหรือพื้นที่ว่างได้ เพื่อที่จะแสดงให้เห็นความแปรปรวนของ
 ข้อมูล *นำเสนอโดยใช้ค่าทางสถิติ* เช่น เวลาที่ใช้ในการเดินทาง
 จำนวนประชากร หรือผลิตภัณฑ์มวลรวมประชาชาติ (Gross
 national product: GNP) ซึ่งแทนที่ด้วยขนาดพื้นที่ หรือระยะความ
 ห่าง



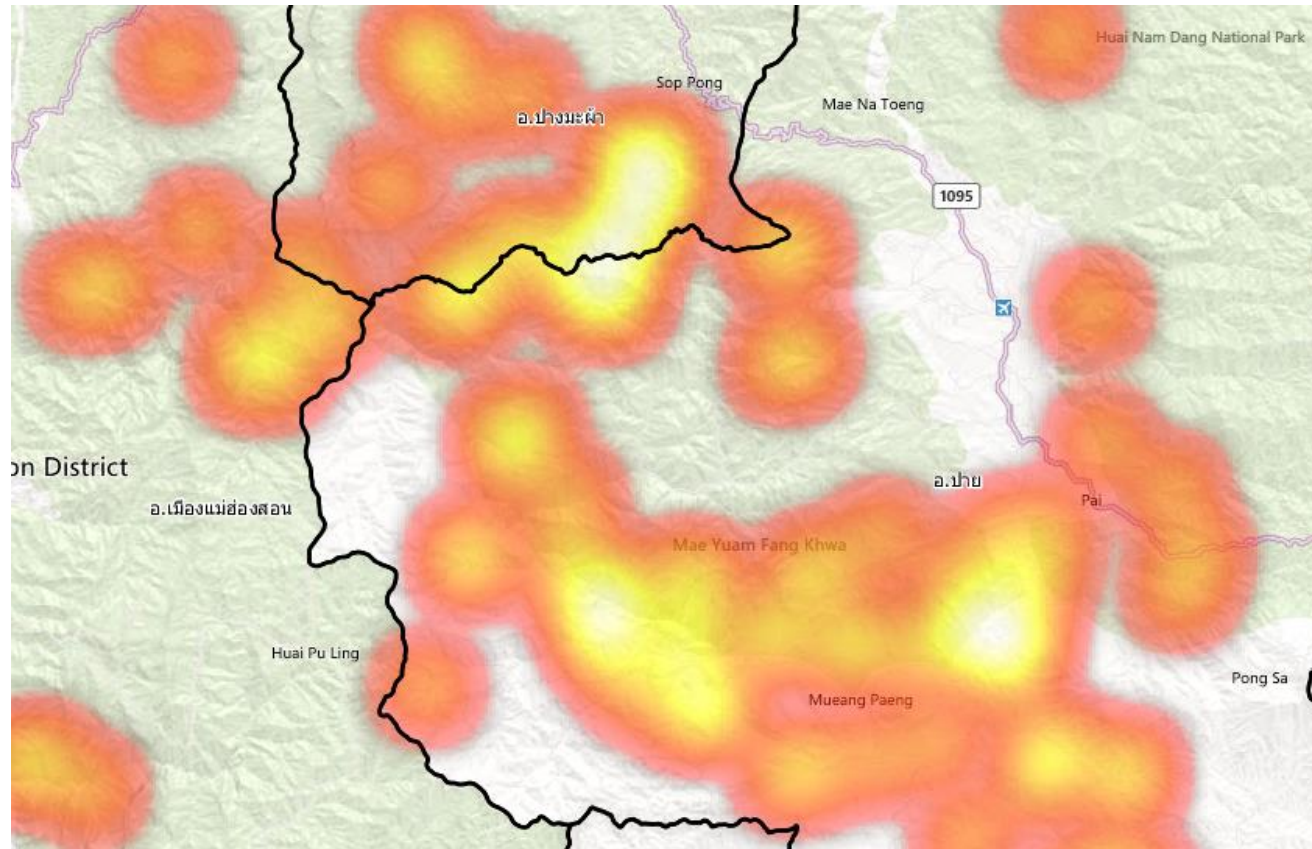
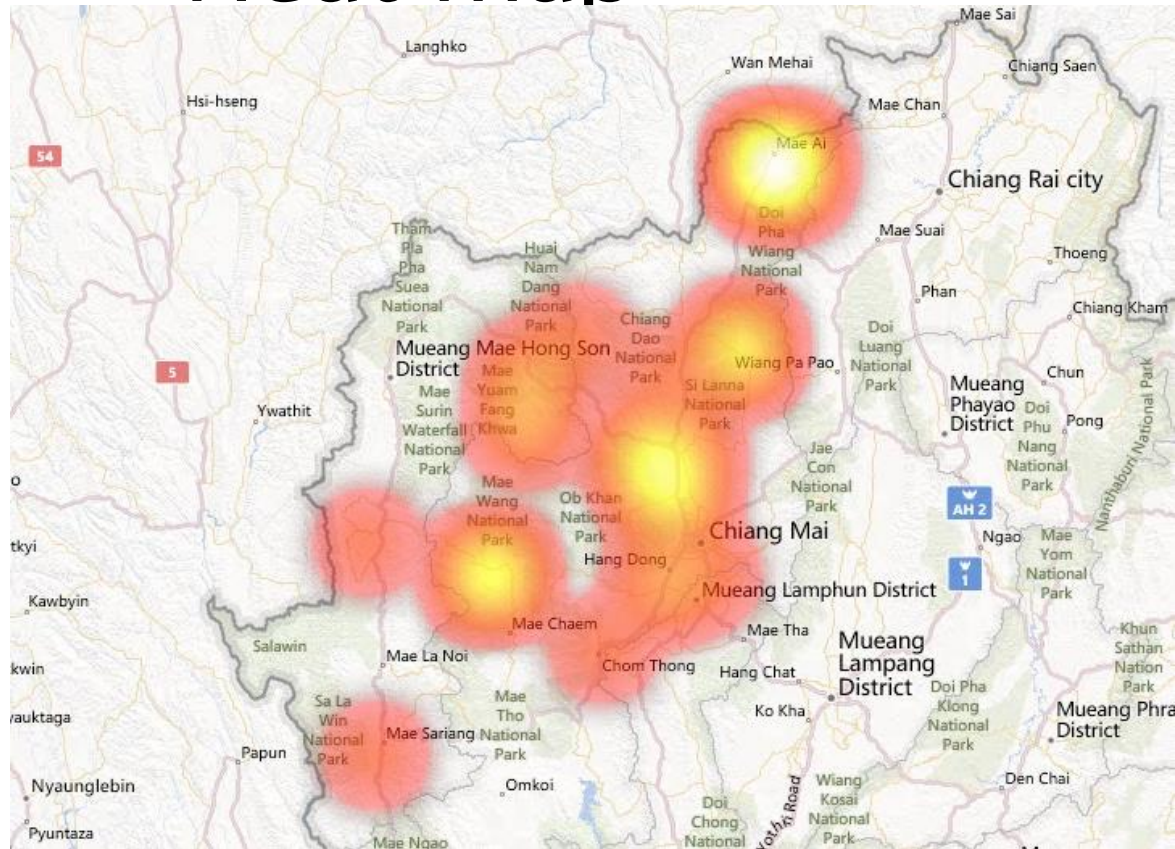
Heat map

- การแสดงภาพของข้อมูล โดยจะแสดงออกมาในรูปแบบของ “ สี ” ซึ่งแต่ละสีจะบ่งบอกถึงระดับความถี่ บ่อย ของพฤติกรรม หรือข้อมูลที่น่าสนใจอื่นๆ
- อาจจะมีการนำเอาแผนที่มาใช้ร่วมกับการวิเคราะห์ข้อมูลทางประชากรศาสตร์ (Demographic data) อาทิ ตำแหน่งสถานที่สำคัญ และนำมาผ่านกระบวนการวิเคราะห์ เพื่อนำผลการวิเคราะห์นั้น มาประกอบการตัดสินใจในการดำเนินธุรกิจ การวางกลยุทธ์ขององค์กรในด้านต่างๆ ไม่ว่าจะเป็นการขยายธุรกิจ การใช้เครื่องมือทางการตลาด และการวิเคราะห์ความสามารถของบริษัทฯ ในการเข้าถึงกลุ่มลูกค้าเป้าหมาย ผลการวิเคราะห์นี้สามารถนำไปประยุกต์ใช้ได้กับลูกค้าหลากหลายธุรกิจที่ไม่ว่าจะเป็น ร้านสะดวกซื้อ, สถาบันการเงิน, บริษัทโฆษณาทำการตลาด,

Heat map



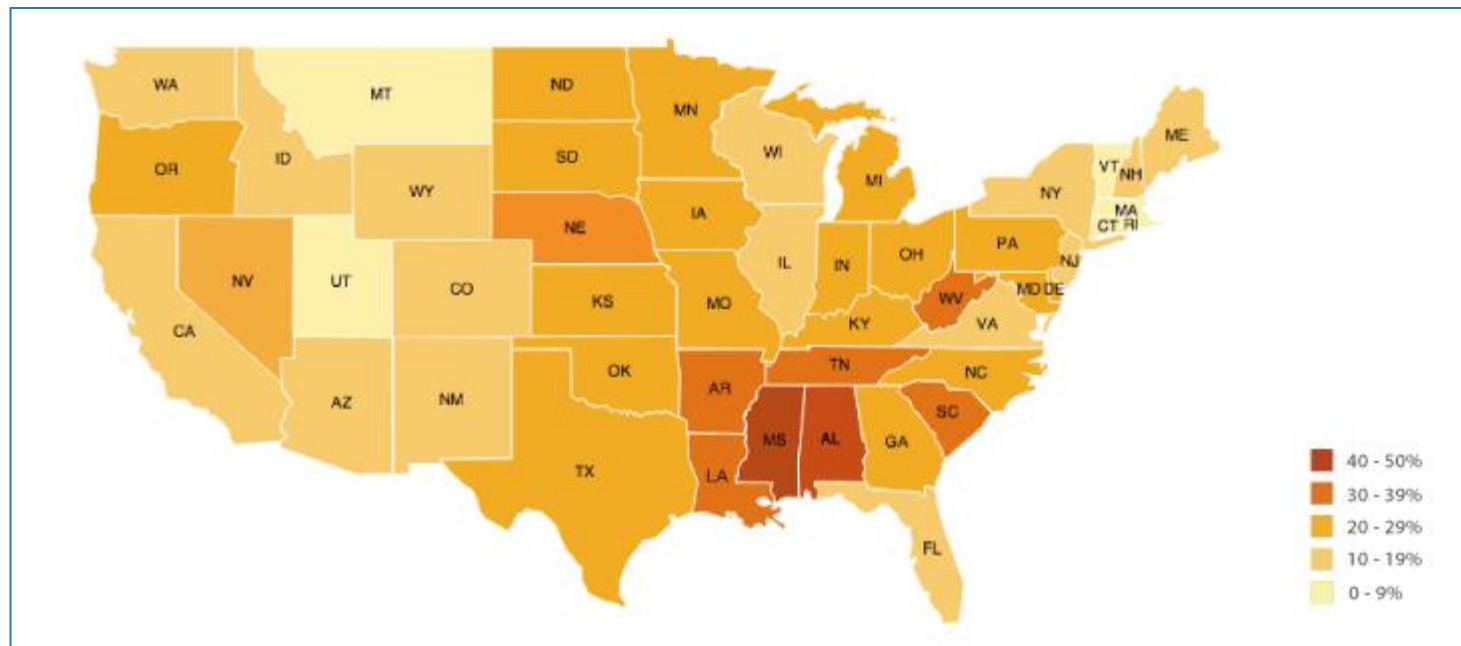
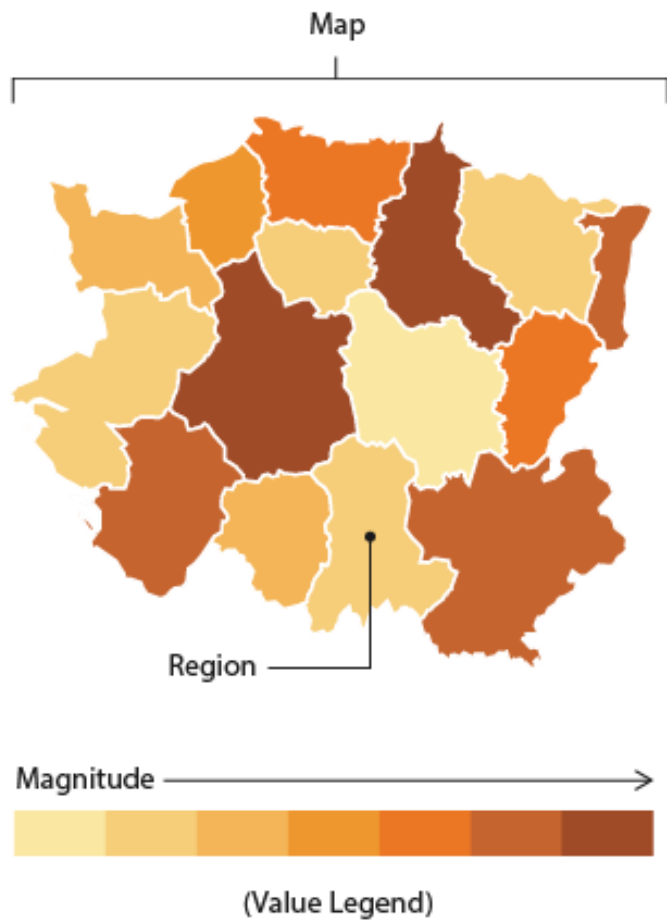
Heat map



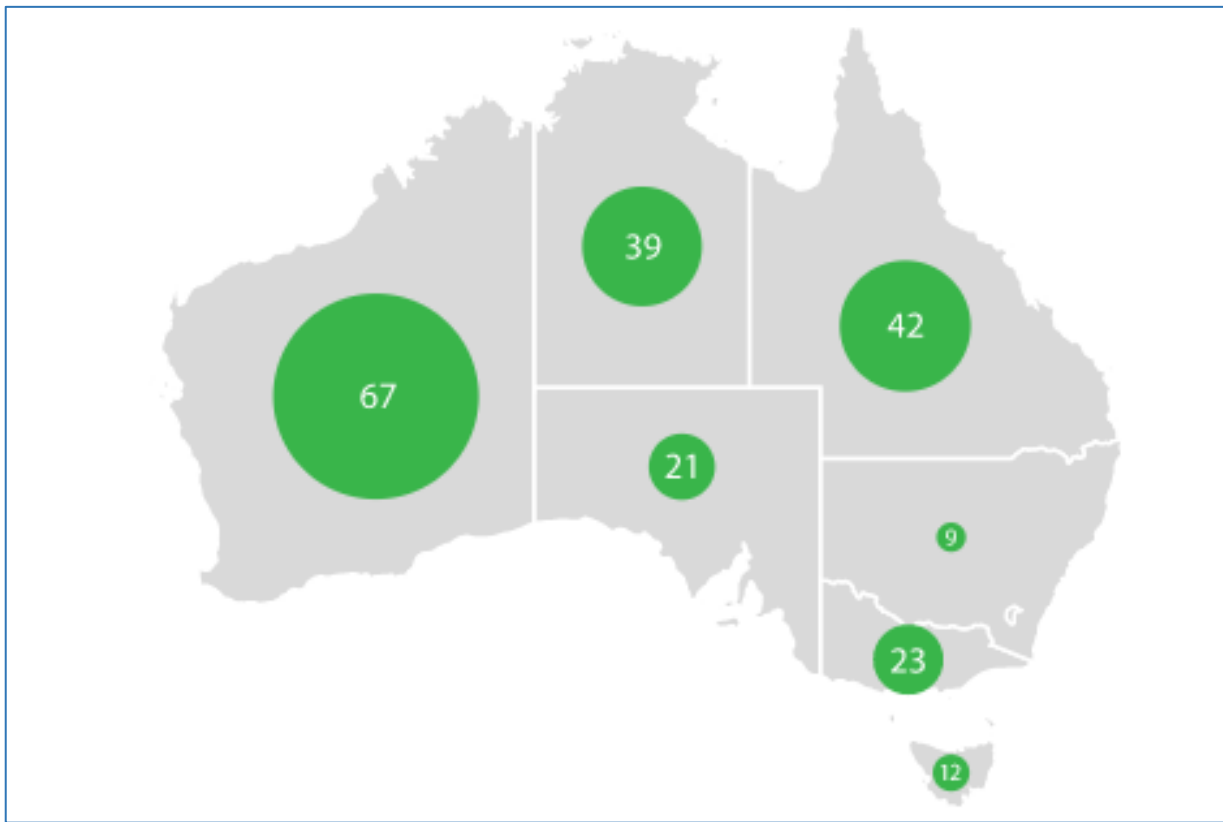
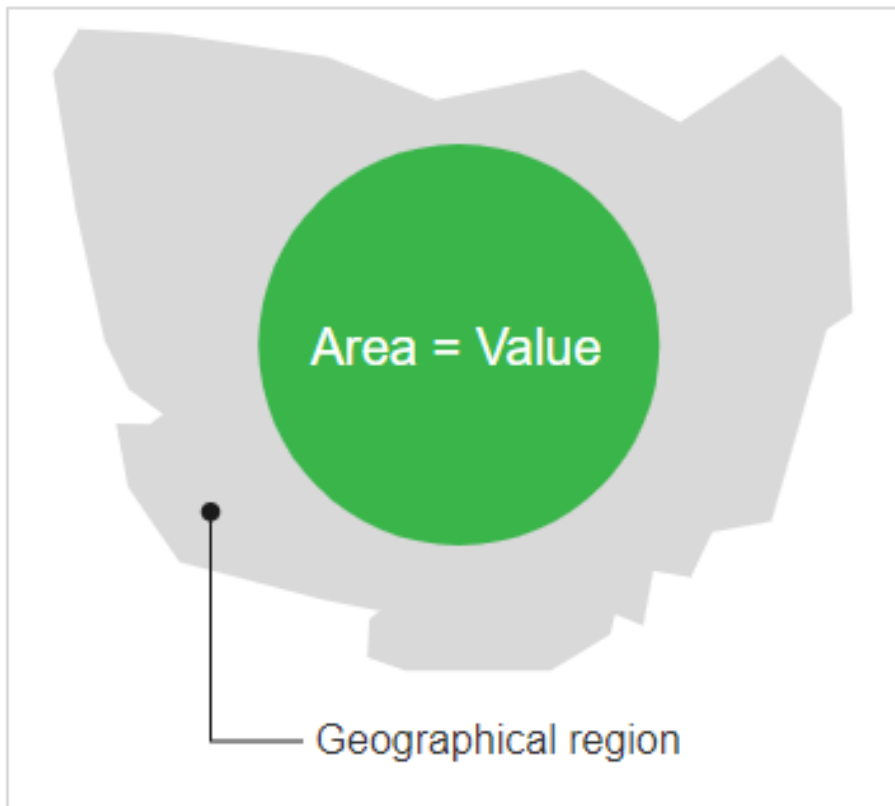
Choropleths

- แผนที่โคโรเพลท (Choropleth Map) เป็นแผนที่ที่ใช้แสดงความแตกต่างของข้อมูลด้วยสีหรือความอ่อนเข้มของสี หรือลวดลายในสัดส่วนที่เป็นตัววัดของตัวแปรทางสถิติที่มี การแสดงบนแผนที่ เช่น ความหนาแน่นของประชากรแผนที่โคโรเพลทอยู่ในประเภทของแผนที่เฉพาะเรื่องเพราะเป็นแผนที่ที่แสดงเฉพาะสิ่งที่สนใจในบริเวณพื้นที่ภูมิศาสตร์ที่สนใจเท่านั้น

Choropleths



Graduated Circles/ Bubble Map

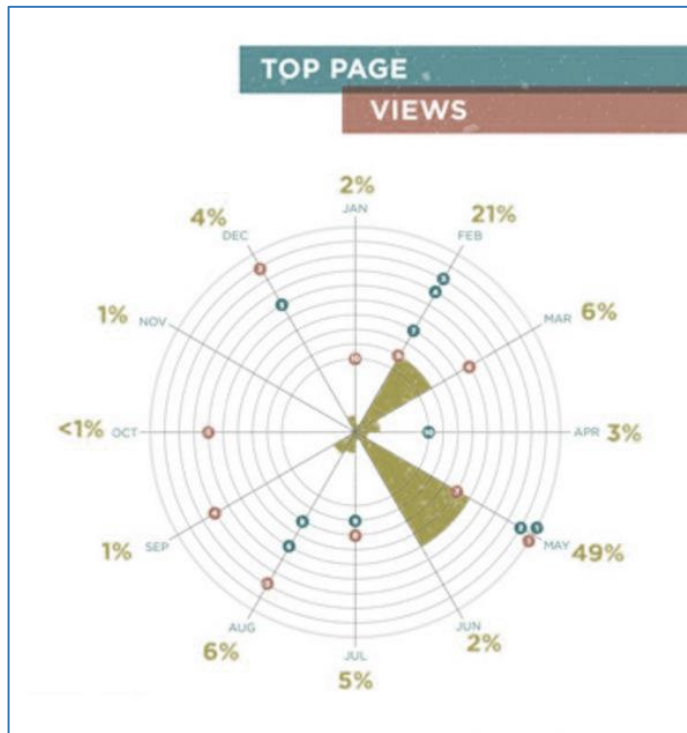


คุณค่าของการแสดงภาพข้อมูล (The Value of Visualization)

- การสร้างภาพข้อมูลเป็นวิธีการหนึ่งที่ทำให้ผู้ใช้สามารถได้รับข้อมูลเชิงลึกของประเด็นที่สนใจ ซึ่งการแสดงผลข้อมูลในรูปแบบของตารางต้องใช้เวลาในการทำ ความเข้าใจค่อนข้างนานและเข้าใจยาก ดังนั้นคุณค่าของการแสดงภาพข้อมูลสามารถจำแนกได้ดังนี้
 1. การสร้างภาพข้อมูลช่วยให้ผู้ชมทำความเข้าใจชุดข้อมูลขนาดใหญ่ได้ง่ายขึ้น
 2. การสร้างภาพข้อมูลช่วยให้ผู้ใช้เข้าใจข้อมูลในระดับที่ใช้งานง่ายขึ้น
 3. การสร้างภาพข้อมูลช่วยให้ผู้ใช้เห็นความสัมพันธ์ระหว่างชุดข้อมูล
 4. การสร้างภาพข้อมูลสามารถช่วยเปลี่ยนมุมมองหรือพฤติกรรมของเราได้

1. การสร้างภาพข้อมูลช่วยให้ทำความเข้าใจชุดข้อมูลขนาดใหญ่ได้ง่ายขึ้น

ข้อมูลจำนวนมากจะถูกเก็บรวบรวมจัดเรียงและนำเสนอในการสร้างภาพข้อมูล โดยช่วยในการบอกเล่าเรื่องราวและให้ผู้ชมภาพรวมของสิ่งที่ข้อมูลที่จะบอกเรา



ภาพนี้เป็นรายงานประจำปี 2012 ของ
บ. Lemonly โดยสามารถดูได้ทันทีว่า
เดือนใดที่มีจำนวนการดูหน้าเว็บมากที่สุด
- พฤษภาคมและกุมภาพันธ์

2. การสร้างภาพข้อมูลช่วยให้ผู้ใช้เข้าใจข้อมูลในระดับที่ใช้งานง่ายขึ้น

เนื่องจากการสร้างภาพข้อมูลช่วยให้เราสามารถมุ่งเน้นไปที่ข้อมูลที่เกี่ยวข้องและมีความสำคัญซึ่งแสดงถึงข้อมูลเพื่อให้เราเข้าใจได้ทันทีว่าเกิดอะไรขึ้น

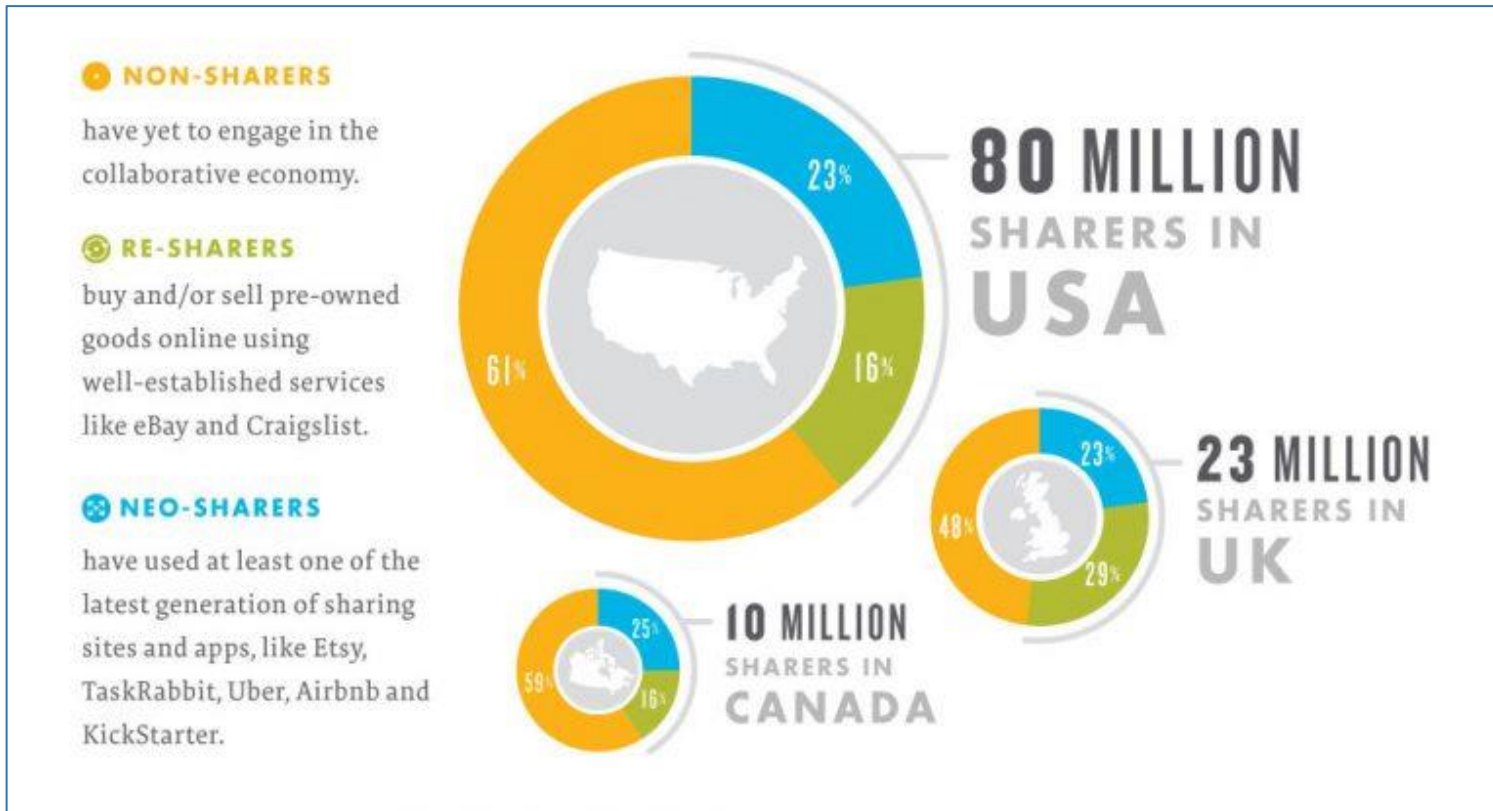


จากภาพจะเห็นข้อมูลสองระดับชั้น โดยมี การใช้สีเน้นสิ่งที่จะมุ่งเน้น เพื่อแสดงว่า

1. ผู้เล่น MLS มาจากประเทศใด
2. และจำนวนผู้เล่นจากแต่ละประเทศ

3. การสร้างภาพข้อมูลช่วยให้ผู้ใช้เห็นความสัมพันธ์ระหว่างชุดข้อมูล

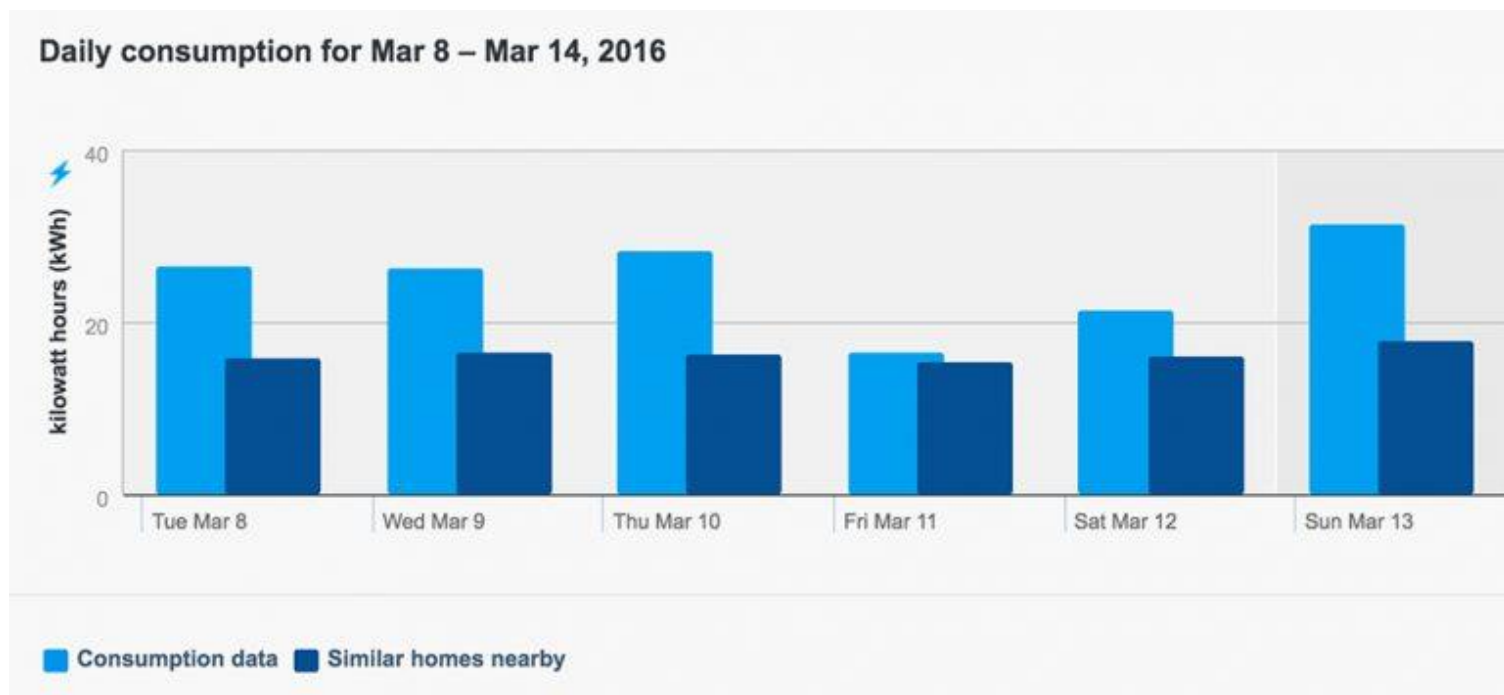
ผู้ใช้สามารถทำการเชื่อมโยงและสามารถระบุรูปแบบข้อมูล ซึ่งจะไม่สามารถทำได้ในข้อมูลแบบปกติ



ภาพนี้แสดงถึง จำนวนคนของแต่ละประเทศด้วยขนาดของกราฟวงกลมในขณะที่ใช้สิ่ระบบประเภทความแตกต่างของ ลูกค้า 3 กลุ่ม โดยเราสามารถสังเกตเห็นความสัมพันธ์ที่แตกต่างกันหลายประการในข้อมูล เช่น - มีผู้ใช้ร่วมกันในสหรัฐอเมริกามากกว่าแคนาดาและสหราชอาณาจักร

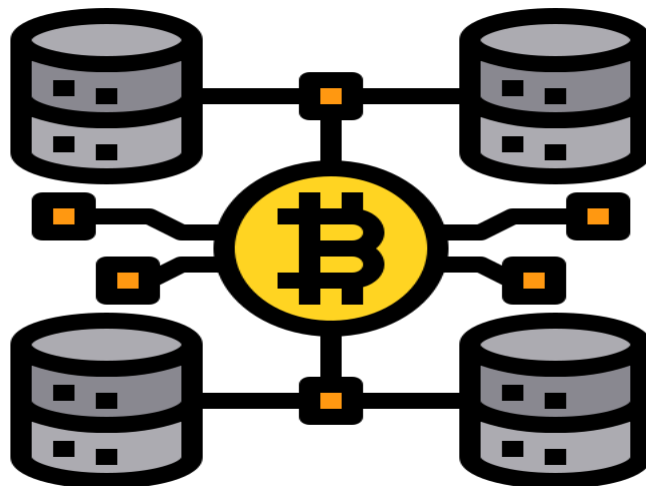
4. การสร้างภาพข้อมูลสามารถช่วยเปลี่ยนมุมมองหรือพฤติกรรมของเราได้

เนื่องจากจะสามารถเข้าใจข้อมูลได้ชัดเจนยิ่งขึ้นเมื่อมีการแสดงข้อมูลเป็นภาพ และสามารถช่วยเราตัดสินใจได้อย่างชาญฉลาดมากขึ้น



- จากภาพเราสามารถดูปริมาณการใช้ไฟฟ้า เมื่อเทียบกับบ้านใกล้เคียงที่คล้ายกัน เนื่องจากมีบริบทของว่า ได้ใช้ไฟฟ้ามากกว่าเมื่อเทียบกับคนอื่น ๆ ในละแวกบ้าน
- เมื่อผู้ใช้ข้อมูลเริ่มเปลี่ยนแปลงพฤติกรรมเพื่อ

3.4 บทที่ 4 : Big Data Analytic



Big Data Analytic

Workshop
10-11 September 2021

What is data analytics?

“Data analytics is the technique of analyzing raw data from various data sources and to arrive at meaningful insights.”

“Keep reading to get a complete understanding of data analytics”

Definition of data analytics

The volume, velocity, and veracity of data generated today is incomparable—and extracting the maximum value out of it is paramount. The significance of analytics is higher now than ever.

Data analytics is the business of deriving meaningful insights in the form of patterns, relationships, and trends, from diverse data sets. It involves the application of both quantitative and qualitative methodologies.

Data analytics software enables collecting, cleansing, storing, analyzing, and reporting of data on any given scale.

Difference Between Data Analysis, Data Mining & Data Modeling

Data analysis is done with the purpose of finding answers to specific questions. Data analytics techniques are similar to business analytics and business intelligence.

Data Mining is about finding the different patterns in data. For this, various mathematical and computational algorithms are applied to data and new data will get generated.

Data Modeling is about how companies organize or manage the data. Here, various methodologies and techniques are applied to data. Data analysis is required for data modeling.

The business case for data analytics

The key business objectives of any organization are to reduce costs, increase profitability, and improve risk management.

A successful business can achieve all three with timely insights on identifying new growth opportunities, business process improvement, segmenting target markets, and planning future roadmaps.

Robust data analytics software that serves these insights quickly and accurately is a major enabler of high business performance and growth.

What can you do with Data Analytics?

Strategize in real time

Actively plan your campaigns, operations, and budgeting with up-to-date data.

Easily collaborate across all levels for continuous data collection, preparation, analysis, and reporting.

Understand your actual performance, course correct deviations, and predict scenarios for more effective business continuity.

What can you do with Data Analytics?

Understand the actual results

“What were your sales last quarter?”

“Who utilized customer service the most?”

Data analytics allows you to constantly reflect on "what's happened" and "what's happening" by measuring and monitoring KPIs using a variety of self-service BI visualizations.

What can you do with Data Analytics?

Plan your future roadmap

Identify trends and anomalies from past performance.

Gather context by analyzing data collected from disparate sources.

Drill down even further to identify drivers that impact business outcomes.

Use them to predict future events via smart forecasting.

Who uses data analytics?

Leadership

The shift to data-native culture is pushing C-level executives to reinvent their strategic and business outlooks.

Integrating data analytics across enterprise functions can give them unprecedented visibility into enhancing operational efficiency.

These analytic insights also drive the proactiveness needed to address gaps effectively and subsequently future-proof the company.

Who uses data analytics?

Marketing & sales

Analytics provides a complete view of prospects' journeys—transcending the old linear concept.

Marketers can track customers all the way from initial interest to final purchase.

The detailed insights available on consumer behavior and preferences help in creating personalized interactions with them.

Who uses data analytics?

Human resources

Analytics-driven organizations can be proactive, rather than reactionary, in understanding their employees.

Data insights empower them to hire the right talent, optimize benefits for employee retention, and identify areas for skill development.

It also helps HR to seamlessly collaborate with other departments for workforce planning and monitoring.

Who uses data analytics?

Finance

Finance teams understand client payment behavior, track cash flow, handle profitability and margin analysis, and run revenue forecasting.

In a regulatory and economic environment that's more challenging than ever, analytics empowers organizations to perform highly accurate risk assessment and management.

Who uses data analytics?

Customer service

To achieve higher customer satisfaction and loyalty, analytics can assist in tailoring customer experiences by predicting their needs, picking up on friction at customer touchpoints, and personalize incentives for retention.

Insights on support agent performance can be useful in reducing latency and bridging skill gaps.

Capabilities of a data analytics software

Data collection and management

Data which is available in any form (structured, unstructured, semi-structured) and from anywhere (files, feeds, local and cloud storage, business apps, local and cloud databases) can be blended and merged, at any scale and complexity.

The collected data is further organized in a standard format, to make sense out of it easily.

It's also imperative to exercise fine-grained access control, to ensure data privacy and security.

Capabilities of a data analytics software

Data analysis and reporting

Useful insights can be discovered through in-depth data exploration.

They can then be meaningfully presented for consumption in the form of visually rich reports and dashboard.

Capabilities of a data analytics software

Data sharing and collaboration

The value of insights gained is only realised by sharing them with appropriate business stakeholders in multiple ways—like email alerts or embedded dashboards.

Facilitating real-time collaboration between them drives quick and informed decision-making.

Benefits of data analytics

- Creating fail-proof business strategies based on detailed insights and highly accurate predictions
- Identifying growth opportunities that would be almost impossible through legacy methods
- Holding on to competitive advantages through constant market awareness
- Proactively spotting and fixing gaps for seamless business continuity
- Widening the scope for innovation by integrating with emerging technologies

Best practices

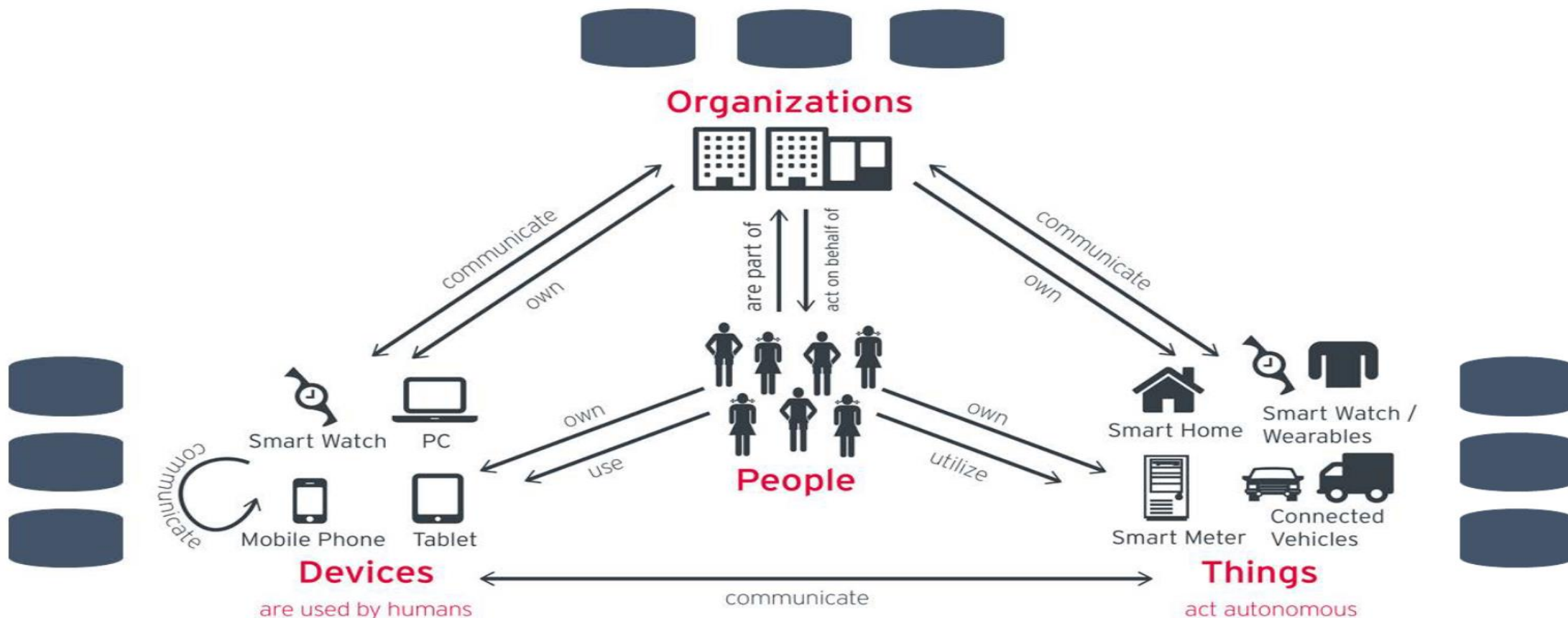
- Embrace analytics throughout your organization's culture
- Re-engineer legacy methods and systems within the latest frameworks
- Stay in tune with advanced analytics features and capabilities
- Work with suitable vendors and focus more on your core business
- Thoroughly evaluate the expertise of potential partners
- Continuously refine and redefine your business' approach to analytics

What is Big Data

Big data analytics is the often complex process of examining big data to uncover information -- such as hidden patterns, correlations, market trends and customer preferences -- that can help organizations make informed business decisions.

On a broad scale, data analytics technologies and techniques provide a means to analyze data sets and take away new information—which can help organizations make informed business decisions. Business intelligence (BI) queries answer basic questions about business operations and performance.

What Data From?



Case Study

Walmart



Walmart is the largest retailer in the world and the world's largest company by revenue, with more than 2 million employees and 20000 stores in 28 countries. It started making use of big data analytics much before the word Big Data came into the picture.

Walmart uses **Data Mining** to discover patterns that can be used to provide product recommendations to the user, based on which products were brought together. WalMart by applying effective Data Mining has increased its conversion rate of customers. It has been speeding along big data analysis to provide best-in-class e-commerce technologies with a motive to deliver superior customer experience. The main objective of holding big data at Walmart is to optimize the shopping experience of customers when they are in a Walmart store. Big data solutions at Walmart are developed with the intent of redesigning global websites and building innovative applications to customize the shopping experience for customers whilst increasing logistics efficiency. **Hadoop** and NoSQL technologies are used to provide internal customers with access to real-time data collected from different sources and centralized for effective use.

Case Study

Uber



Uber is the first choice for people around the world when they think of moving people and making deliveries. It uses the personal data of the user to closely monitor which features of the service are mostly used, to analyze usage patterns and to determine where the services should be more focused. Uber focuses on the supply and demand of the services due to which the prices of the services provided changes. Therefore one of Uber's biggest uses of data is surge pricing. For instance, if you are running late for an appointment and you book a cab in a crowded place then you must be ready to pay twice the amount.

For example, On New Year's Eve, the price for driving for one mile can go from 200 to 1000. In the short term, surge pricing affects the rate of demand, while long term use could be the key to retaining or losing customers. **Machine learning algorithms** are considered to determine where the demand is strong.

Case Study

Netflix



It is the most loved American entertainment company specializing in online on-demand streaming video for its customers. Netflix has been determined to be able to predict what exactly its customers will enjoy watching with Big Data. As such, Big Data analytics is the fuel that fires the ‘recommendation engine’ designed to serve this purpose. More recently, Netflix started positioning itself as a content creator, not just a distribution method. Unsurprisingly, this strategy has been firmly driven by data. Netflix’s recommendation engines and new content decisions are fed by data points such as what titles customers watch, how often playback stopped, ratings are given, etc. The company’s data structure includes Hadoop, Hive and Pig with much other traditional business intelligence.

Netflix shows us that knowing exactly what customers want is easy to understand if the companies just don’t go with the assumptions and make decisions based on Big Data.

What is Business Intelligence (BI)

Business intelligence (BI) is an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance.

BI software applies to the "infrastructure, tools and apps" part of the above definition. It's the software that lets you analyze data and information, and in so doing, helps you improve and optimize the business decisions you make.

Who should use BI software?

Bi software can be used by anyone who wants to analyze data.

However, it's the management side of an organization which uses it the most.

From managers all the way up to the CEO and founders of a company, BI software helps users take informed decisions based on data.

Across Business Type

1. Small and Medium Businesses

There was a time when BI software was expensive, and was used only by large enterprises. But Cloud Business Intelligence (BI) software solutions today are very much affordable for small and medium businesses (SMBs). SMBs increasingly take advantage of this fact, and use BI solutions on the cloud

2. Large Businesses

As the saying goes, "data is the new oil." Large enterprises use Business Intelligence (BI) analytics software to continuously mine data from across the applications and databases they use.

Across Department

1. Sales

How is the sales pipeline looking for the next few months? Who are my top and bottom performing sales reps? Which regions and products should I concentrate on to maximize profits? Sales managers can answer these questions with BI software solutions.

2. Marketing

Marketing professionals run many marketing campaigns across many channels. How effective each of these campaigns are, how marketing metrics like page views to visitors to sign ups is doing, the top five performing regions, products, and services—these are all questions that marketers want answers for.

3. Finance

Beyond looking at the balance sheet, cash flow, and profit and loss statements, smart finance professionals right to the CFO level, now ask deep questions of their financial data.

Across Department

4. Human Resources & Recruitment

HR managers need to be on top of the number of people being recruited, how many people are leaving the organization, effective workforce utilization, productivity trends, and more. Recruitment agencies need to monitor metrics like the average time for filling job positions, selection ratio, cost incurred per hire, offer acceptance rate, and which sourcing funnels are being effective.

5. Operations

Operational—or admin—expenses form an important part of any organization's balance sheet. CFOs need to always keep a tab on admin expenses, as any leakage here will harm the company's performance.

Benefits of using BI software

Crunch large quantities of data

Whether it is gigabytes or petabytes, crunch your raw data into interpretable graphic reports.

Visualize any data

Create visualizations for any type of data

Single version of truth

Get one complete view of your data, instead of multiple segmented spreadsheets floating around

Actionable insight

Pinpoint the cause of problems, and find the subsequent actions you have to take to fix them.

Benefits of using BI software

Optimize business operations

Optimize your business methods and processes, based on the visualizations you generate.

Grow your business

Identify growth areas for your business—which geographical regions to concentrate on, which products or categories can bring in more growth, and more.

Cut expenses

Identify where unnecessary expenses are making your business bleed, and stop them.

Essential features of BI software

1. Data sources

Make sure the BI software you choose supports analyzing data from a wide variety of sources. Data can be anywhere—in the form of files, web feed URLs, databases (both on-premise and on the cloud), business apps, and more. The business intelligence software you choose should connect with data from any source

2. Visualization options

Your BI software should allow you to visualize data in myriad ways. There should be many options to choose from—chart types, color palettes, widgets, themes, and fonts, among others.

3. Self-service

Data is seen as a democratized commodity across organizations now. While a department can own the data, everyone across an organization needs access to it, to create their own visualizations. Users feel empowered when they create the reports they need on their own. Your BI software should be easy to use for any business user.

Essential features of BI software

4. Sharing and collaboration

Your BI software should allow you to collaborate with your peers and clients easily. It should allow fine-grained access control—who can see, and do what—and be able to hold conversations about the visualizations created.

5. AI-augmented analytics

Till recently, AI wasn't an essential feature for BI software. But AI, assisted by technologies like machine learning (ML) and natural language processing (NLP), is becoming a must-have, as it assists users to generate reports and insights more readily.

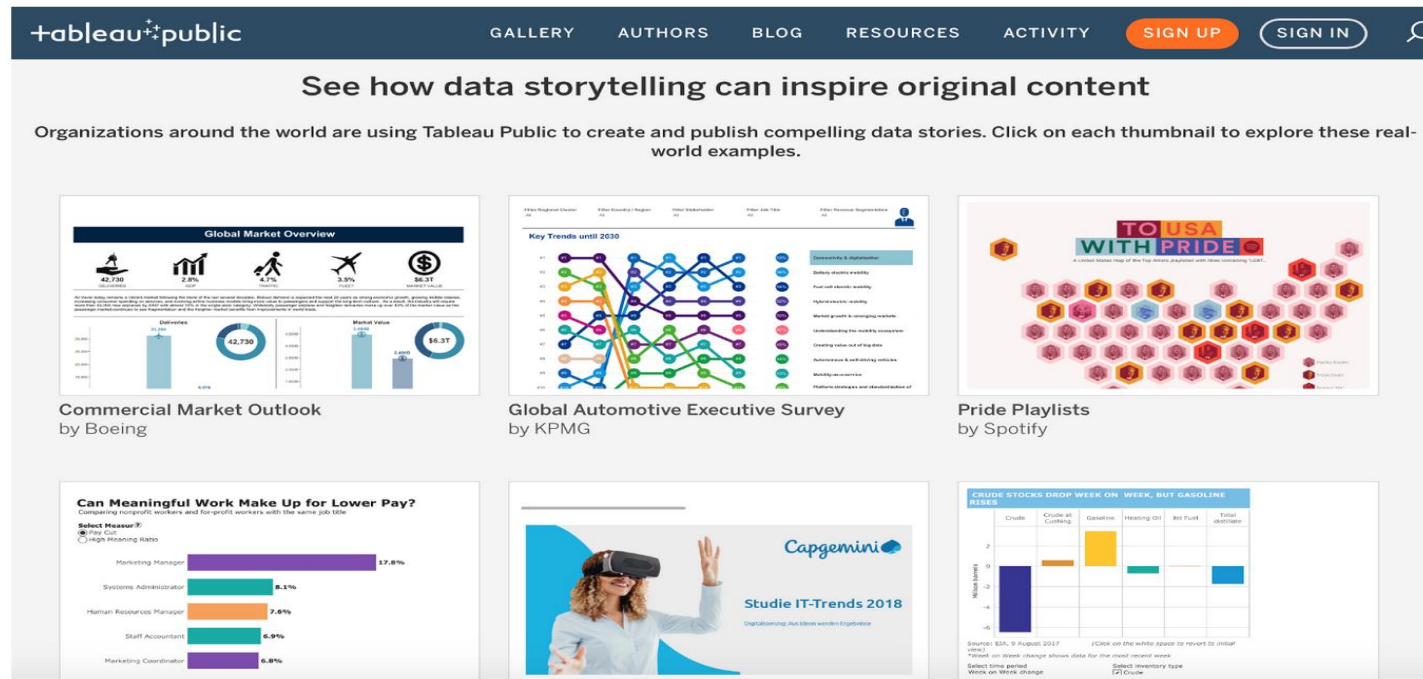
6. Accessibility

Accessing your BI software from any device is a necessity. This allows you to be up-to-date, wherever you are. Having the reports be interactive on mobile devices is an added bonus.

Data Analytics Tool

Tableau

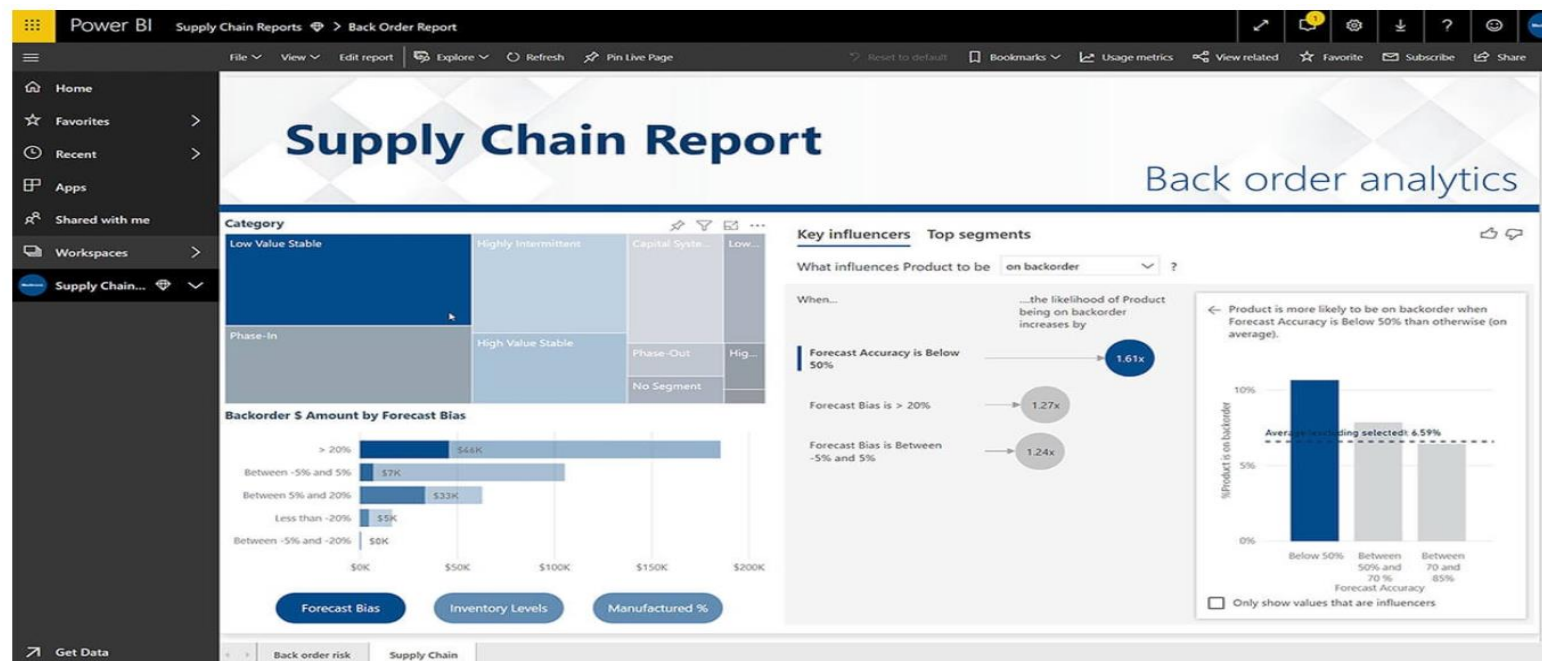
The core essence of Tableau is actually the PivotTable and PivotChart of Excel. It can be said that Tableau is keenly aware of this feature of Excel. It entered the BI market earlier and carried forward this core value.



Data Analytics Tool

Power BI

Power BI was previously a plug-in for Excel, and its development was not ideal. So it got out of Excel and developed into a BI tool. As a latecomer, Power BI has iterative updates every month and catches up very fast.

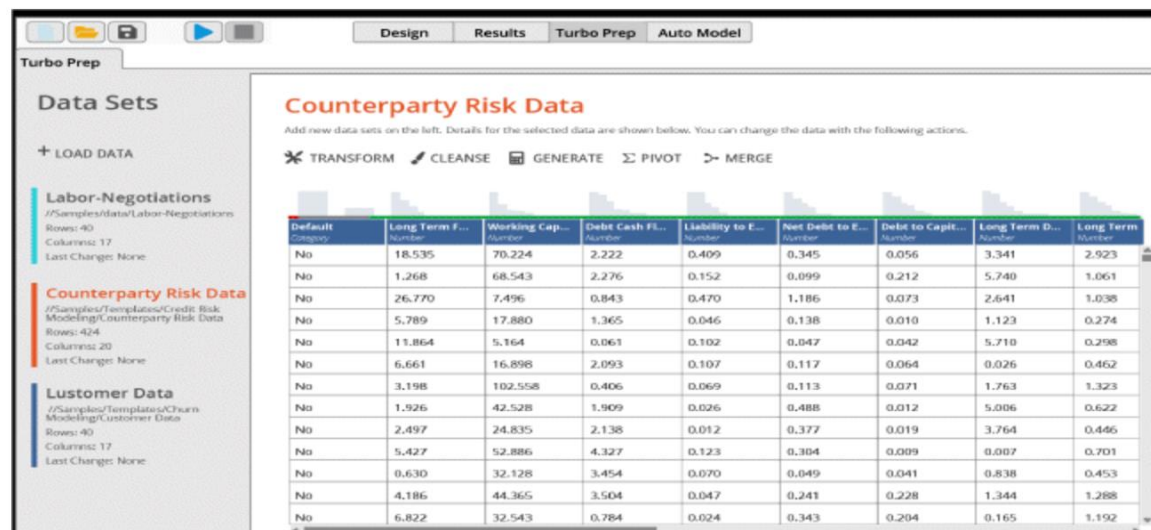


Data Analytics Tool

RapidMiner

RapidMiner is a software platform for data preparation, machine learning, deep learning, text mining, and predictive model deployment. It provides all data prep capabilities.

The tool will help data scientists and analysts in improving their productivity through automated machine learning. You will not have to write the code, to do the data analysis with the help of RapidMiner Radoop.



Counterparty Risk Data

Add new data sets on the left. Details for the selected data are shown below. You can change the data with the following actions.

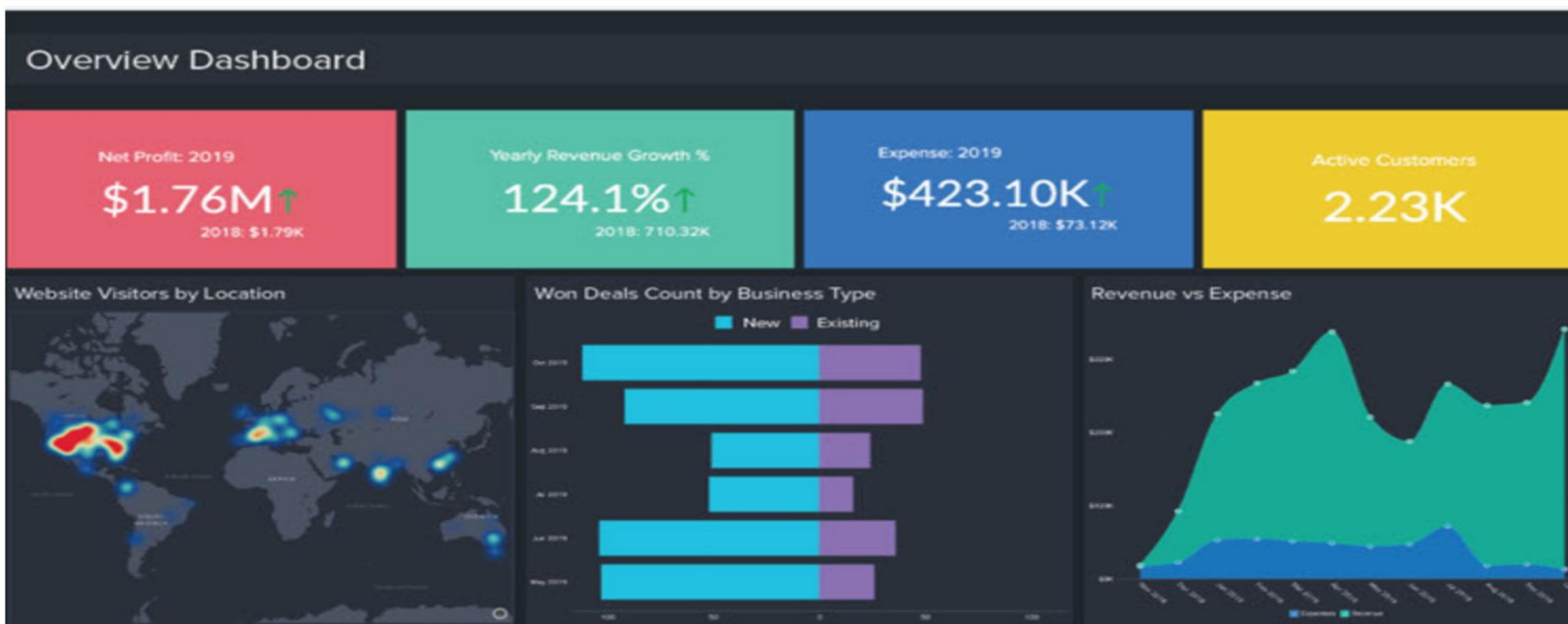
✖ TRANSFORM ✂ CLEANSE 📄 GENERATE ∑ PIVOT ➔ MERGE

Default Status	Long Term F... Number	Working Cap... Number	Debt Cash FL... Number	Liability to E... Number	Net Debt to E... Number	Debt to Capit... Number	Long Term D... Number	Long Term Number
No	18.535	70.224	2.222	0.409	0.345	0.056	3.341	2.923
No	1.268	68.543	2.276	0.152	0.099	0.212	5.740	1.061
No	26.770	7.496	0.843	0.470	1.186	0.073	2.641	1.038
No	5.789	17.880	1.365	0.046	0.138	0.010	1.123	0.274
No	11.864	5.164	0.061	0.102	0.047	0.042	5.710	0.298
No	6.661	16.898	2.093	0.107	0.117	0.064	0.026	0.462
No	3.198	102.558	0.406	0.069	0.113	0.071	1.763	1.323
No	1.926	42.528	1.909	0.026	0.488	0.012	5.006	0.622
No	2.497	24.835	2.138	0.012	0.377	0.019	3.764	0.446
No	5.427	52.886	4.327	0.123	0.304	0.009	0.007	0.701
No	0.630	32.128	3.454	0.070	0.049	0.041	0.838	0.453
No	4.186	44.365	3.504	0.047	0.241	0.228	1.344	1.288
No	6.822	32.543	0.784	0.024	0.343	0.204	0.165	1.192

Data Analytics Tool

Zoho Analytics

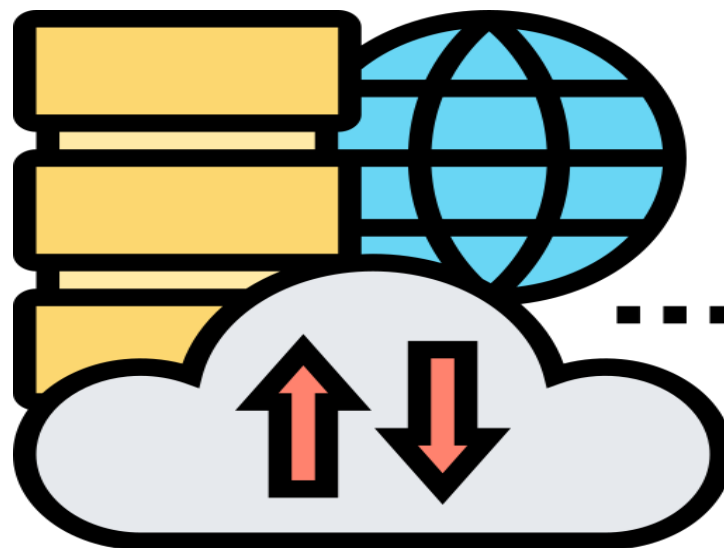
Zoho Analytics is a self-service BI and analytics platform. It allows users to create insightful dashboards and visually analyze any data. It features an AI-powered assistant that allows users to ask questions and get intelligent answers in the form of meaningful reports.



Tools Comparation

Tool	Platform	Description	Price
Tableau	Windows, Mac, Web-based, Android, iOS	Nice tool available for free with good features and functionalities.	Tableau Public: Free Tableau Creator: \$70 per user per month.
Power BI	Window, Mac, Linux, Android, iOS.	The advantage of Power BI lies in its business model and data analysis capabilities.	Desktop: free Pro: 9.99\$/user/m Premium: 4,995\$
RapidMiner	Cross-platform	System is easy to use. Powerful GUI. Five products to choose from.	Free: 10,000 rows. Small: \$2500 Medium: \$5000 Large: \$10000
Zoho Analytic	Web-based, Android, iOS	a self-service BI and analytics platform. Create insightful dashboards and visually analyze any data	Free: 10,000 rows 22\$-445\$ per month On-premise

3.5 บทที่ 5 : Data Preparation



Data Preparation

Data Preparation

Data preparation is the process of cleaning and transforming raw data prior to processing and analysis. It is an important step prior to processing and often involves reformatting data, making corrections to data and the combining of data sets to enrich data.

Data preparation is often a lengthy undertaking for data professionals or business users, but it is essential as a prerequisite to put data in context in order to turn it into insights and eliminate bias resulting from poor data quality.

Benefits of Data Preparation

The data preparation is the worst part of their job, but the efficient, accurate business decisions can only be made with clean data. Data preparation helps:

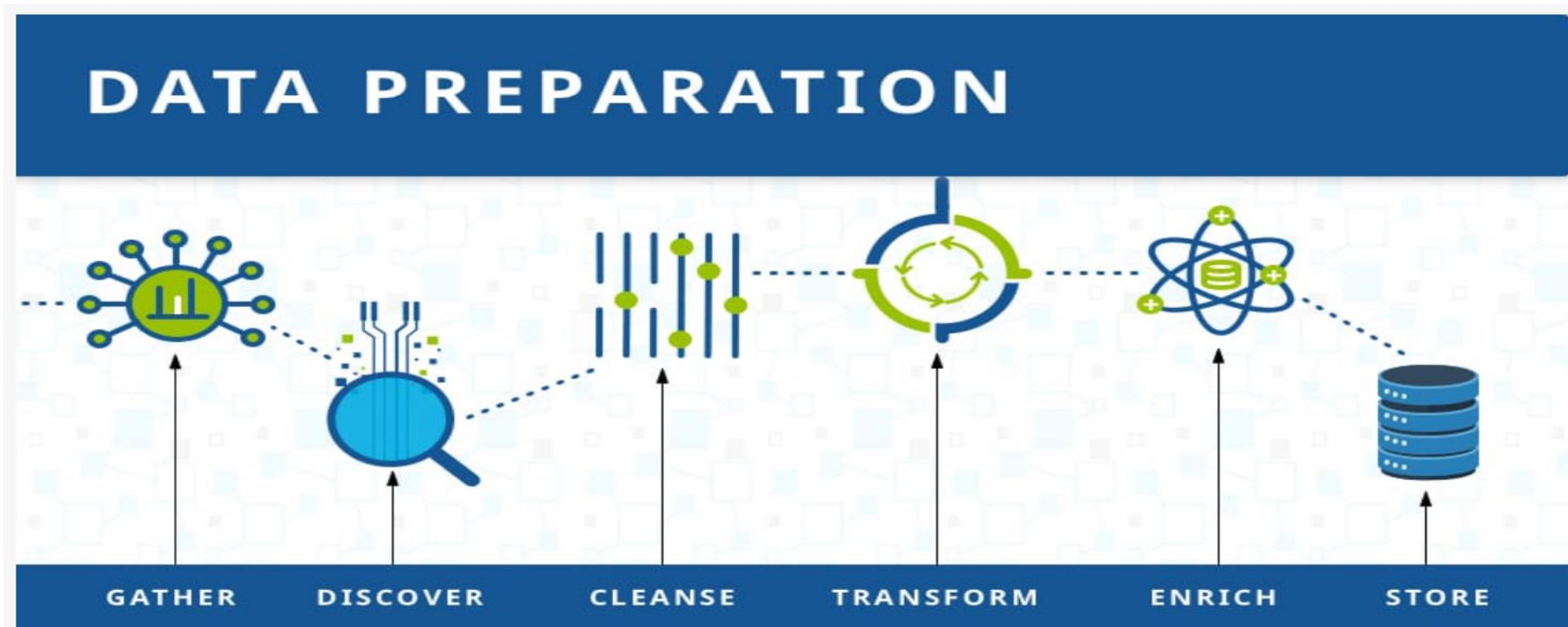
Fix errors quickly — Data preparation helps catch errors before processing. After data has been removed from its original source, these errors become more difficult to understand and correct.

Produce top-quality data — Cleaning and reformatting datasets ensures that all data used in analysis will be high quality.

Make better business decisions — Higher quality data that can be processed and analyzed more quickly and efficiently leads to more timely, efficient and high-quality business decisions.

Data Preparation Steps

The specifics of the data preparation process vary by industry, organization and need, but the framework remains largely the same.



Data Preparation Steps

- 1. Gather data** - The data preparation process begins with finding the right data. This can come from an existing data catalog or can be added ad-hoc.
- 2. Discover and assess data** - After collecting the data, it is important to discover each dataset. This step is about getting to know the data and understanding what has to be done before the data becomes useful in a particular context.
- 3. Cleanse and validate data** - leaning up the data is traditionally the most time consuming part of the data preparation process, but it's crucial for removing faulty data and filling in gaps.
- 4. Transform and enrich data** - Transforming data is the process of updating the format or value entries in order to reach a well-defined outcome, or to make the data more easily understood by a wider audience.
- 5. Store data** - Once prepared, the data can be stored or channeled into a third party application—such as a business intelligence tool—clearing the way for processing and analysis to take place.

Data from Difference Source

Essential Connectors

Files and Feeds

Import data from a variety of file formats and web/URL feeds. Also, you can set up automatic synchronization.

Formats: CSV, XLS, XLSX, JSON, TSV, HTML, XML, etc.

Cloud Storage/Drive

Easily upload data from major cloud storage vendors and even set up automatic synchronization.

Cloud Drive: Zoho Workdrive, Box, Google Drive, Microsoft OneDrive, Dropbox, etc.

Multiple Databases

Effortlessly connect/upload data, not only from popular relational databases but also from NoSQL databases. Your databases can be setup locally (behind firewall) or hosted or in the cloud.

Data from Difference Source

Sale Analytics

Zoho CRM

Visually analyze your Zoho CRM data right away using Zoho Analytics, with this seamless connector. Jump start your analysis with 100+ hand picked reports & dashboards. Empower yourself with the right sales performance indicators.

Salesforce CRM

Slice and dice your Salesforce CRM data the way you want. Instantly track your key performance indicators to proactively manage your sales pipeline. Unleash unlimited reporting potentials with this connector.

Microsoft Dynamics CRM

Dive deep into your Microsoft Dynamics CRM data. Keep track of critical sales indicators, take informed business decisions, and enrich your sales funnel with valuable insights delivered by the Advanced Analytics connector.

Data from Difference Source

Sale Analytics

HubSpot CRM

Analyze and derive key actionable insights from your HubSpot CRM data. Track key sales metrics, enhance performance and maximize your sales conversions.

Analyze other CRM data

Import data from CRM solutions such as Insightly CRM, Pipedrive CRM and Base CRM into Zoho Analytics using Flatly.

Data from Difference Source

271

Marketing Analytics

Zoho Campaigns

Deeply analyze your Zoho Campaigns data. Know how your campaigns perform, understand recipient behaviors, and improve your email marketing strategy for effective conversions.

MailChimp

Perform in-depth email marketing analytics on your MailChimp data. Analyze your open rates and click rates, create conversion funnels, track campaign performance, and do much more.

Google Analytics

SEO/SEM professionals, webmasters, and digital marketers can now slice & dice Google Analytics data the way they want. Analyze key metrics, and optimize SEO/SEM operations, using Zoho Analytics.

Data from Difference Source

Marketing Analytics

Google AdWords

Measure the success of your Google AdWords campaigns. Track your clicks, conversion rates, cost incurred and more. Create insightful reports and make smart business decisions.

Bing Ads

Increase the effectiveness of your Bing Ad campaigns by tracking key performance indicators such as CTR, CPC, and more. Share insights with peers and device powerful SEM strategies together.

SurveyMonkey

Extract meaningful insights from your SurveyMonkey data. Track your survey performance, analyze your survey response rate, completion rate, and do much more.

Data from Difference Source

Marketing Analytics

Facebook

Measure and boost your impact on Facebook with in-depth Facebook Ads analytics. Resonate with your target audience, maximize ROI and optimize your marketing spend.

Twitter

Visually analyze and uncover hidden insights from your Twitter data. Run powerful competitive analytics, increase engagement, and make your Twitter marketing campaigns effective.

YouTube

Grow your YouTube channel with meaningful insights. Understand your subscribers better and find out what type of content resonates with your target audience.

Data from Difference Source

Marketing Analytics

LinkedIn Ads

Convert your visitors into customers with powerful LinkedIn Ads Analytics. Optimize your ad campaigns with meaningful insights on CTR, CPC, engagement, and more.

Eventbrite

Plan and execute your events efficiently with in-depth event analytics. Understand your attendees, analyze your events, track the ticket sales, and make data-driven decisions for future events.

Instagram

Target, engage, and grow your Instagram followers with Advanced Analytics. Identify your best content and track key metrics like impressions, reach, best time to post, and much more - all under one roof.

Data from Difference Source

Financial Analytics

Zoho Books

Power up financial analytics on top of Zoho Books. This connector helps you to create insightful reports & dashboards over your Zoho Books data for in-depth financial analysis

Xero

Derive deep insights from your Xero data using Advanced Analytics. With this connector, track your overall accounting performance, analyze invoices, generate financial reports and do much more.

Sage

Measure your financial health with in-depth Sage Analytics. Uncover hidden insights on your invoices, purchases, orders, sales, and much more. Share insights, make smart decisions, and grow your business.

Data from Difference Source

E-Commerce Analytics

Shopify

Measure, analyze, and run your e-commerce business efficiently with in-depth Shopify analytics. Track key metrics, extract powerful insights and grow your business.

Data from Difference Source

Project Management Analytics

Zoho Projects

The Zoho Projects connector helps you to get a 360° view of your projects data with Project Management Analytics. Get a head start with more than 50+ pre-packaged reports and dashboards.

Teamwork Projects

With the Teamwork Projects advanced analytics connector, extract rich insights on your project management data. Empower yourself with key performance metrics and optimize your project execution.

Data from Difference Source

IT & Helpdesk Analytics

Zoho Desk

Slice and dice your Zoho Desk data the way you want. Analyze and track your crucial help desk metrics. Make informed decisions and gain insights into your customer support activity.

Zendesk

Derive insights from your Zendesk data through the powerful analytical capabilities of Zoho Analytics. Track your important help desk metrics and make informed decisions.

Teamwork Desk

Discover hidden insights from your Teamwork Desk data and drive your business' success. Measure key helpdesk metrics and arrive at the right decisions -faster and with confidence.

Data from Difference Source

HR Analytics

Zoho Recruit

Perform deeper analytics on your Zoho Recruit data using this connector. Transform your recruitment data into valuable insights and optimize your recruitment process.

What is a Table?

Zoho Analytics stores all your tabular data/datasets in entities called **Tables**.

A table contains a set of columns and actual data rows (similar to a spreadsheet).

Each column has a name and a type (data type) associated with it. A table in Zoho Analytics is identified with this icon ...



A Zoho Analytics database could contain a collection of tables. Two tables could be related to each other using lookup columns, thus enabling you to model a relational Workspace

Working with Tables

- Searching
- Filtering
- Sorting
- Find and Replace
- Split Column
- Show/Hide Columns
- Freeze Columns
- Reorder and Resize Columns
- Exporting a Table
- Sharing a Table
- Publishing a Table
- Changing Column Properties
 - Format a Column
 - Conditional Formatting
 - Set Chart Drill Down Path
 - Rename Column
 - Change Data type of a Column
 - Change a Column to a Lookup Column
- Analyze this Column
- Copying a Table

Data Cleansing

Find and Replace

When it comes to maintaining and processing huge amount of data, the major concern that needs to be addressed is data quality.

The quality of the data often degrades over time. This could happen due to a number of reasons.

Some of the prominent reasons could be,

- Data being blended from multiple sources
- Data being keyed in by different people
- Non-standardization of the data stored

Data Cleansing

Find and Replace

For instance, if you take a CRM database, data could be keyed in by several sales reps and each of them will have his/her own way of entering data.

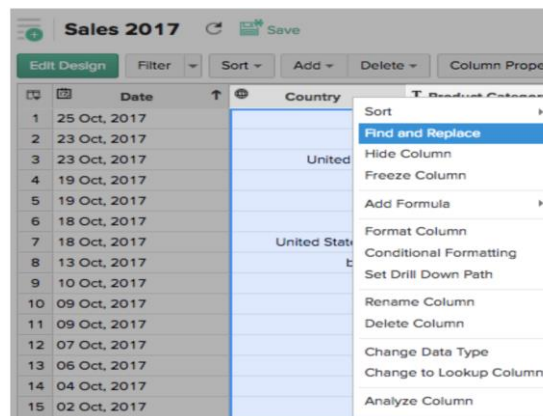
Over a period of time, this inconsistency will affect the quality of data.

One such common case is, some reps might use the abbreviation for a country (CA for Canada) while others might prefer to spell out the complete word.

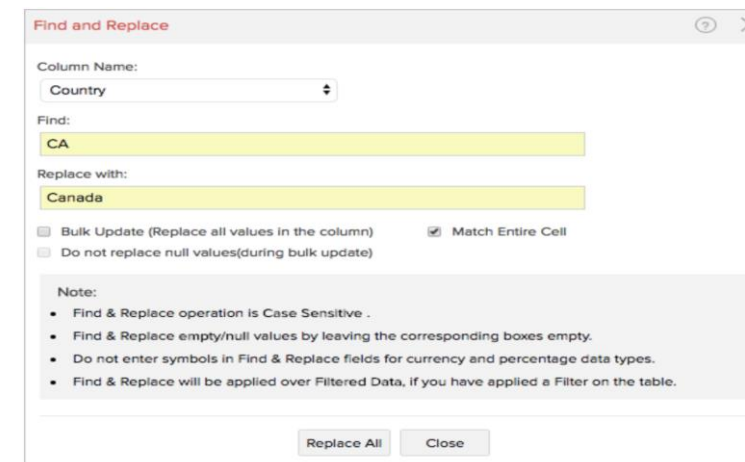
	Date	Country	Product Category	Product	Customer Name	Sales
1	25 Nov, 2017	CA	Grocery	Fruits and Vegetables	Pete Zachriah	\$3,002.46
2	23 Oct, 2017	United States of A	Grocery	Fruits and Vegetables	Santa Cruz Clara	\$2,875.25
3	23 Oct, 2017	USA	Stationery	Specialty Envelopes	John Britto	\$3,882.65
4	19 Oct, 2017	Angola	Stationery	Scissors	Rozario Diego	\$986.38
5	19 Oct, 2017	United Arab Emirates	Stationery	Standard Labels	Neil Seth	\$58.25
6	18 Oct, 2017	Afganistan	Stationery	Binder Clips	Thomas Mondrake	\$189.03
7	18 Oct, 2017	Canada	Grocery	Fruits and Vegetables	Rick Reed	\$360.78

Find and Replace

Right-click on the *Country* column and select **Find and Replace**



In the **Find and Replace** dialog that opens, enter the relevant data for the **Find and Replace** fields. In this example, we are replacing 'CA' with 'Canada'.

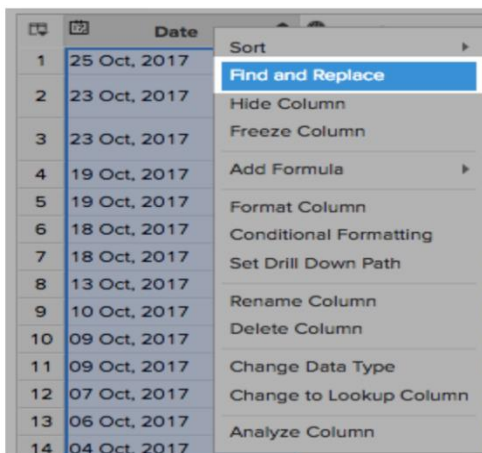


Find and Replace for Date Datatype

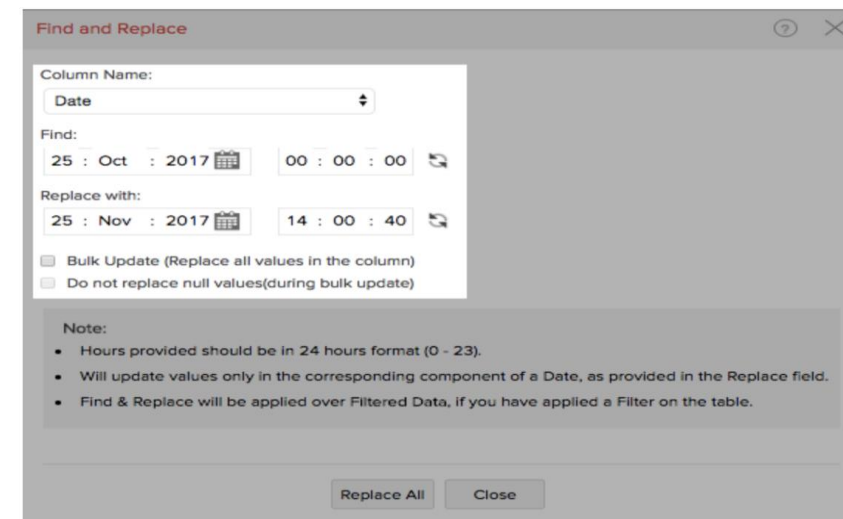
we are going to see how to find and replace a date in the column.

	Date	Country	Product Category	Product	Customer Name	Sales
1	25 Oct, 2017	Canada	Grocery	Fruits and Vegetables	Pete Zachriah	\$3,002.46
2	23 Oct, 2017	United States of America	Grocery	Fruits and Vegetables	Santa Cruz Clara	\$2,875.25
3	23 Oct, 2017	United States of America	Stationery	Specialty Envelopes	John Britto	\$3,882.65
4	19 Oct, 2017	Angola	Stationery	Scissors	Rozario Diego	\$986.38
5	19 Oct, 2017	United Arab Emirates	Stationery	Standard Labels	Neil Seth	\$58.25
6	18 Oct, 2017	Afganistan	Stationery	Binder Clips	Thomas Mondrake	\$189.03
7	18 Oct, 2017	Canada	Grocery	Fruits and Vegetables	Rick Reed	\$360.78

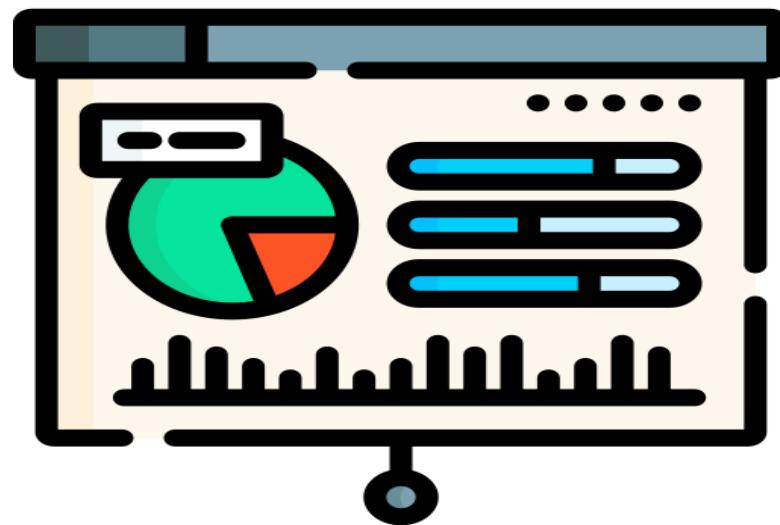
Right-click on the *Date* column and select **Find and Replace**.



the **Find and Replace dialog**, enter the date that you wish to find and replace. Enter the date that you wish to find in the **Find** section and the date to be replaced in the **Replace With** section.



3.6 บทที่ 6 : Data Visualization



Data Visualization

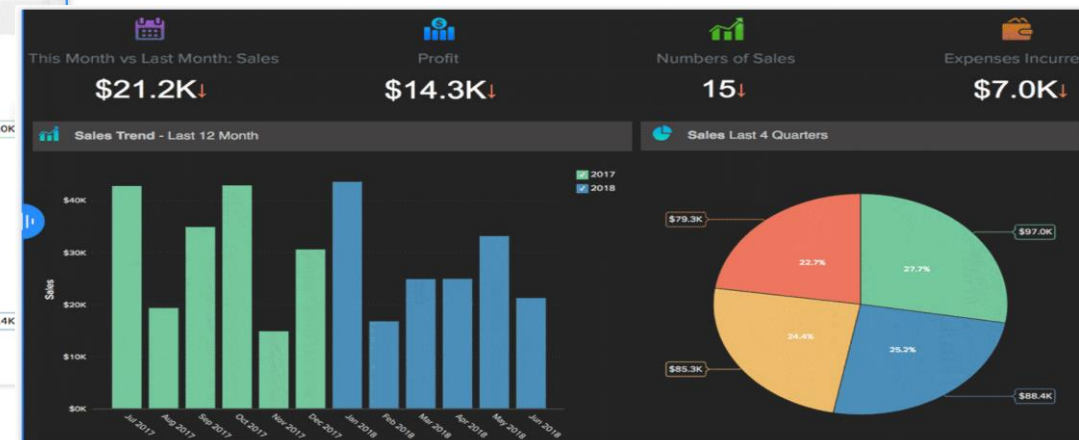
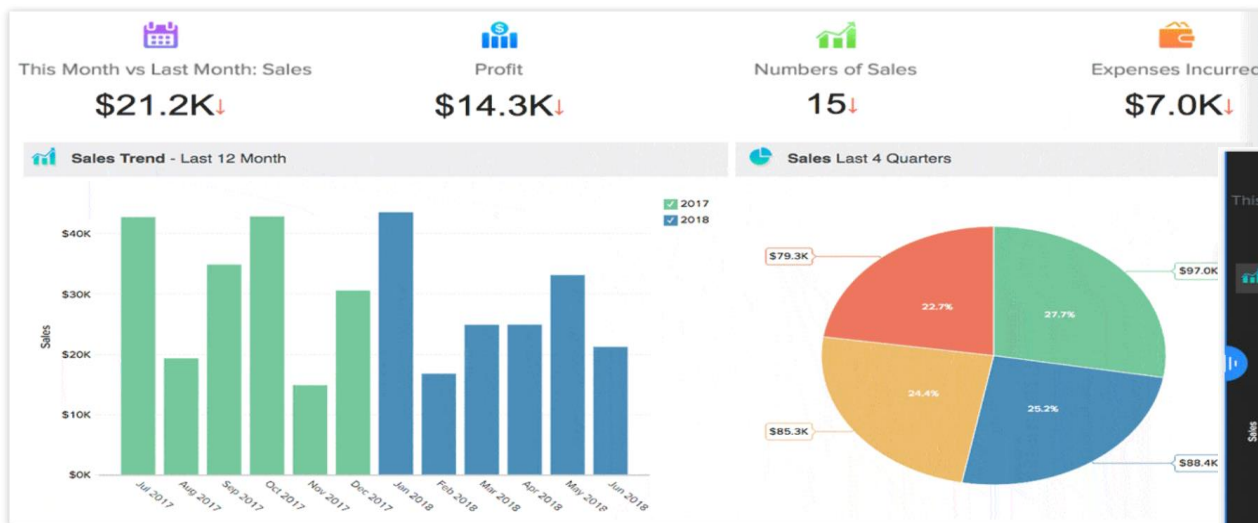
Data Visualization

“Are you getting the real insights you need, when it comes to making business decisions? “

“Data Analytics transforms your data into rich visuals, and helps you arrive at the right decisions faster, and with confidence.”

Visual Analysis

Make use of a variety of charts, widgets, pivot tables and tabular view components to create insightful reports and dashboards with an easy drag-and-drop interface.



Secure online collaboration

Share and collaborate on reports and dashboards with your colleagues. Have full control over what your colleagues or clients can see and do with the reports you share with them.

Embed your reports and dashboards to reach a wider audience.

Export or print reports for offline access.

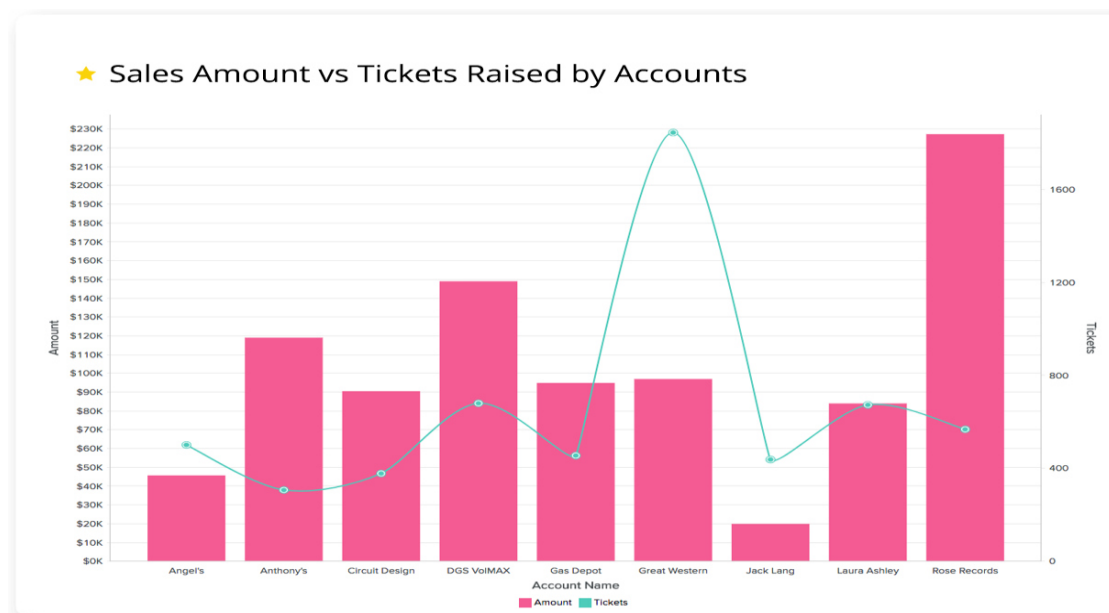
Schedule reports and dashboards to be emailed automatically.



Cross-functional data analytics

Combine data from different sources to create cross-functional reports, and thereby get more insight into your business.

For example, combine your sales (Zoho CRM) data with your helpdesk (Zoho Desk) data to create a 'Sales vs. Tickets Raised' report.



Create Chart

292

Data Visualization

A chart is a visual representation of data which allows you to effectively analyze and interpret data.

Zoho Analytics provides an easy to use drag and drop interface which enables you to create charts easily. It supports a wide range of chart types such as **Area, Line, Bar, Stacked, Pie, Scatter, Combination, Funnel, Web etc.**

Zoho Analytics offers a wide range of customization options to customize your charts like **Titles, Colors, Legends, Tooltips** etc. We also support a comprehensive filtering capability that allows you to extract (filter) the data you desire and portray the chart the way you want.

Zoho Analytics also offers a wide range of interactivity options like **On-Mouse over highlights, Data Drill Down, Legend based filters, User Filters, Chart Exports**, etc., for those who access the chart.

Chart Types

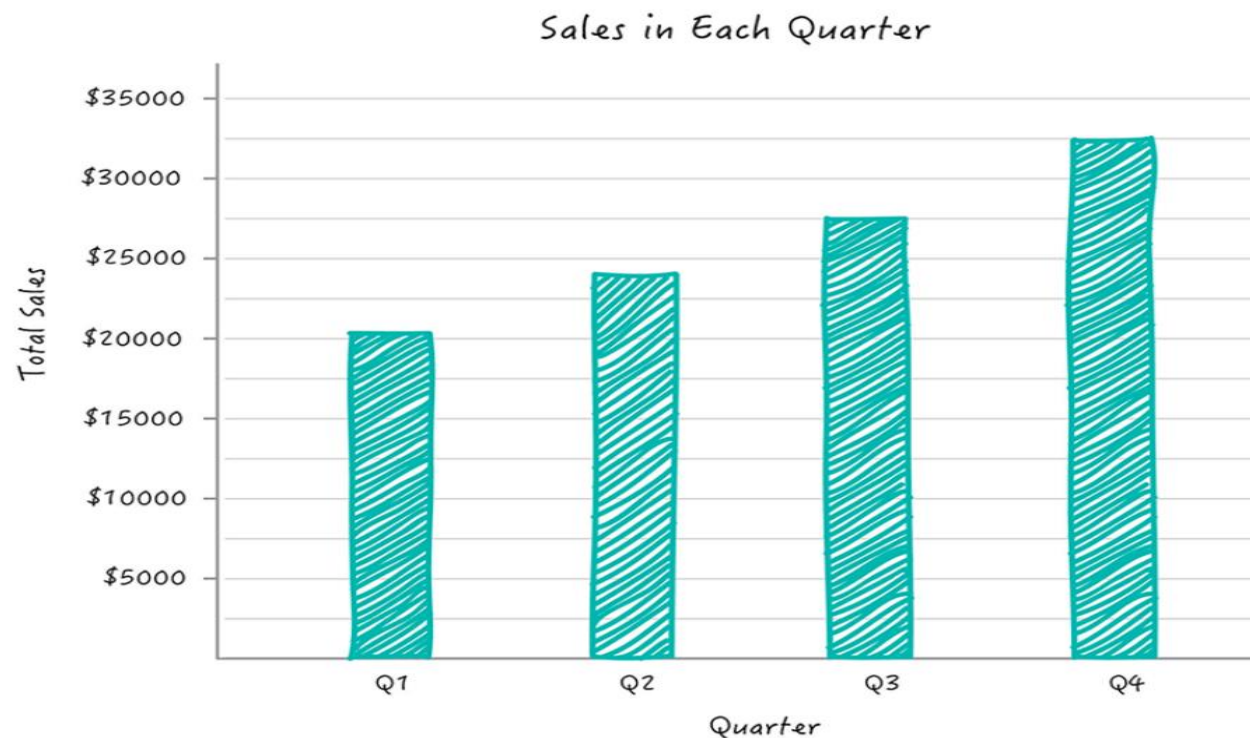
Data Visualization

Charts have the magical ability to transform abstract numbers into beautiful readable visualizations. Often we end up feeling confused in the process of choosing the best chart to visualize our data. To assist users on this, Zoho Analytics offers a unique feature of automatically suggesting the best chart that suits the data you are plotting, while designing a chart.

This chart picker guide is a simple guideline that would enable you to choose the best chart to visualize your data. This guide also offers you a high level insight behind the formula used by Zoho Analytics to auto-suggest the best chart for you.

Column/Bar Chart

Pick this chart to compare data classified into discrete groups. Be it your sales in each quarter, or say scores comparison among teams.



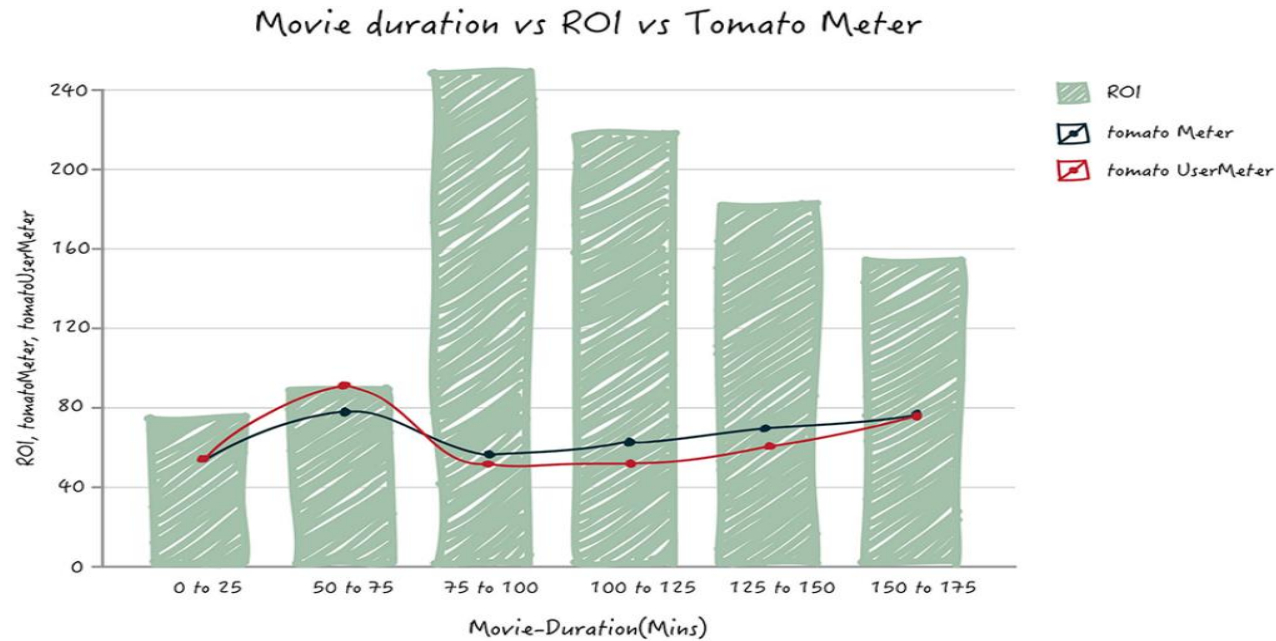
Stacked Column/Bar Chart

This chart shows the contribution of individual stacks that make up the columns. For example, quarter-wise sales (represented as stacks) for each region



Combination Chart

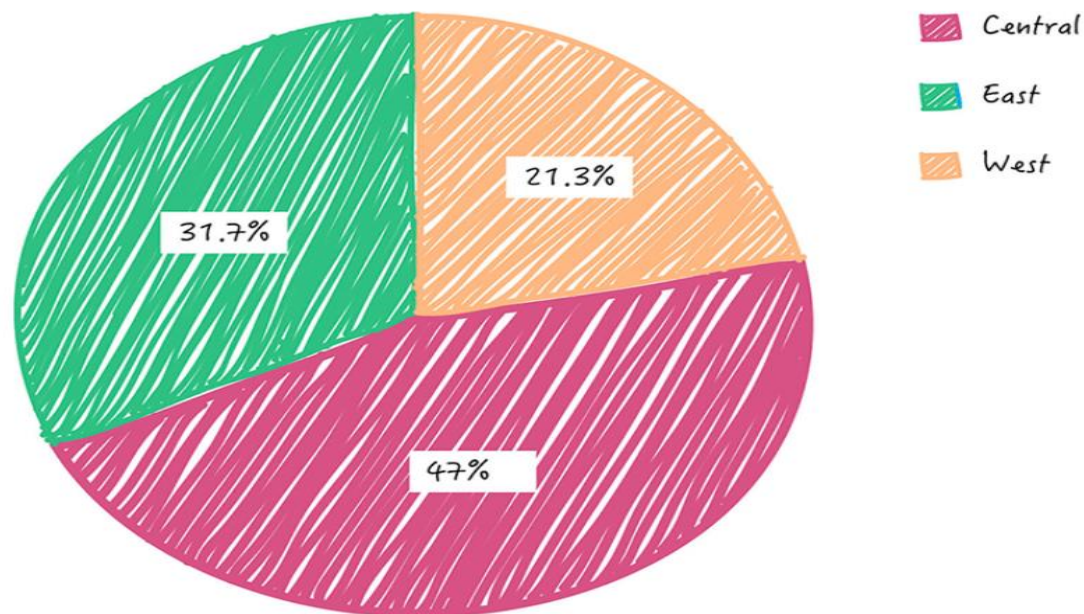
As the name suggests, this is a combination of bar and line charts. Pick this chart when you have mixed types of data series. Or, to represent a complementary data series along with the main data. For example, if total sales across regions is represented in bars, the average sales and ROI can be represented as individual lines in the same chart.



Pie Chart

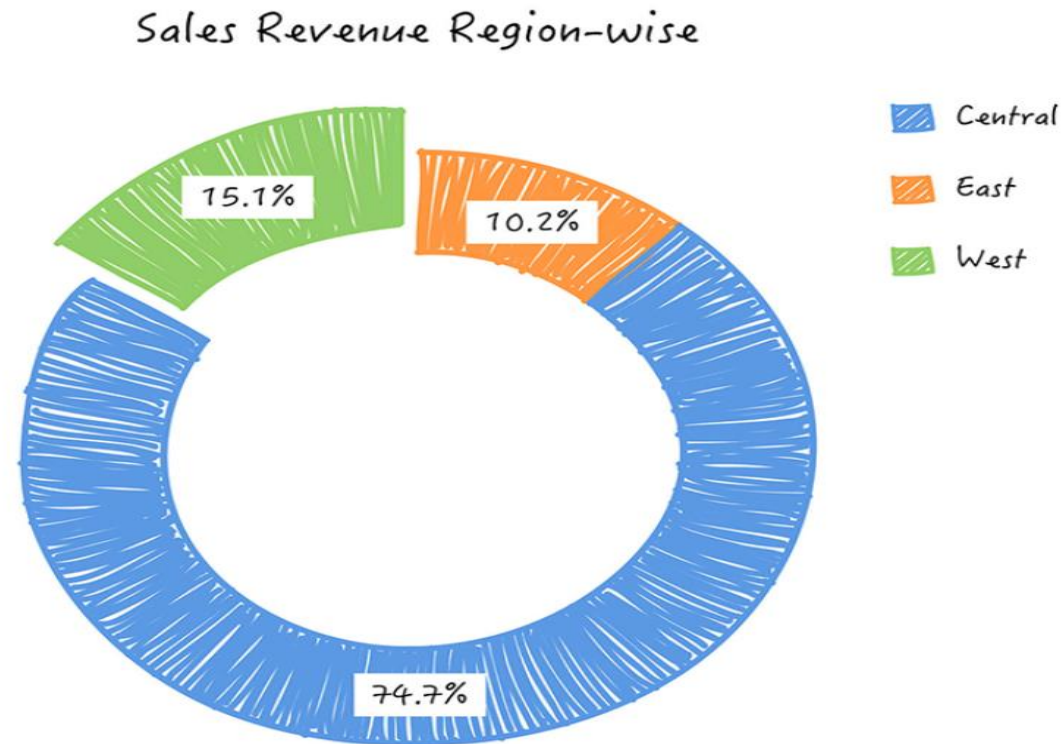
Use this chart to find out the contribution of parts to a whole set. Like which expense this month has eaten up your income-pie the most. Or say, sales distribution across regions.

Sales Revenue Region-wise



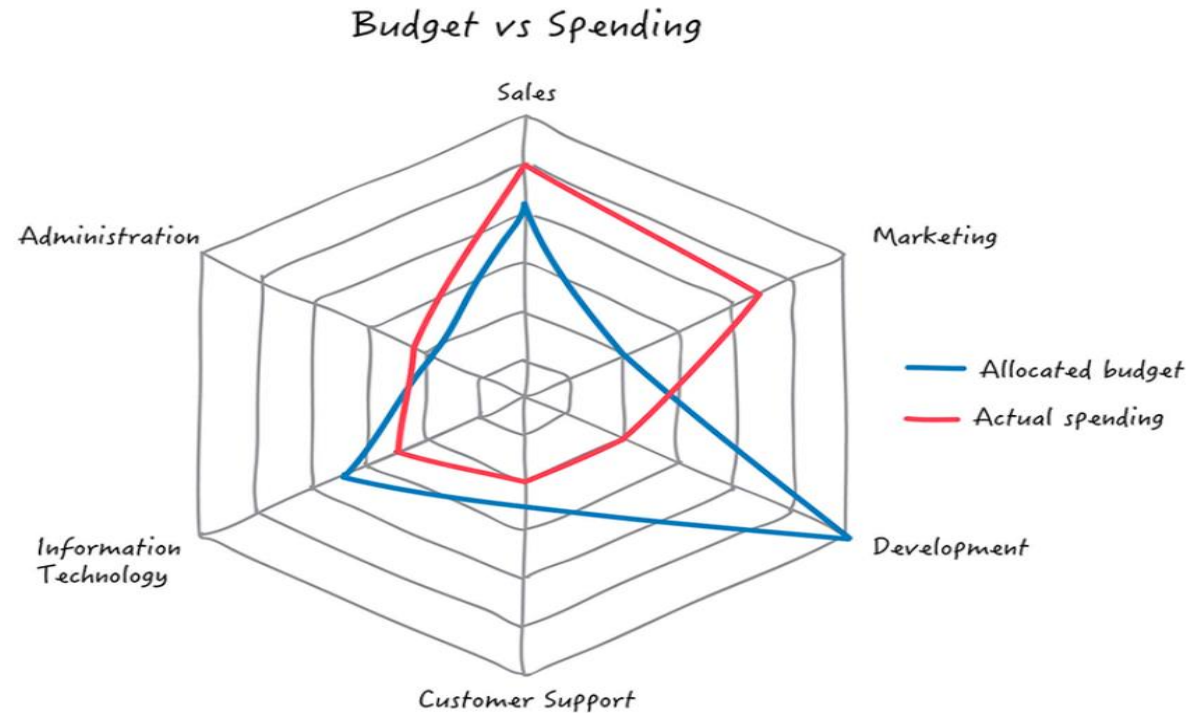
Ring Chart

A doughnut chart is akin to a pie-chart represented in the form of a ring/dough-nut. This chart too displays the contribution of parts to a whole set.



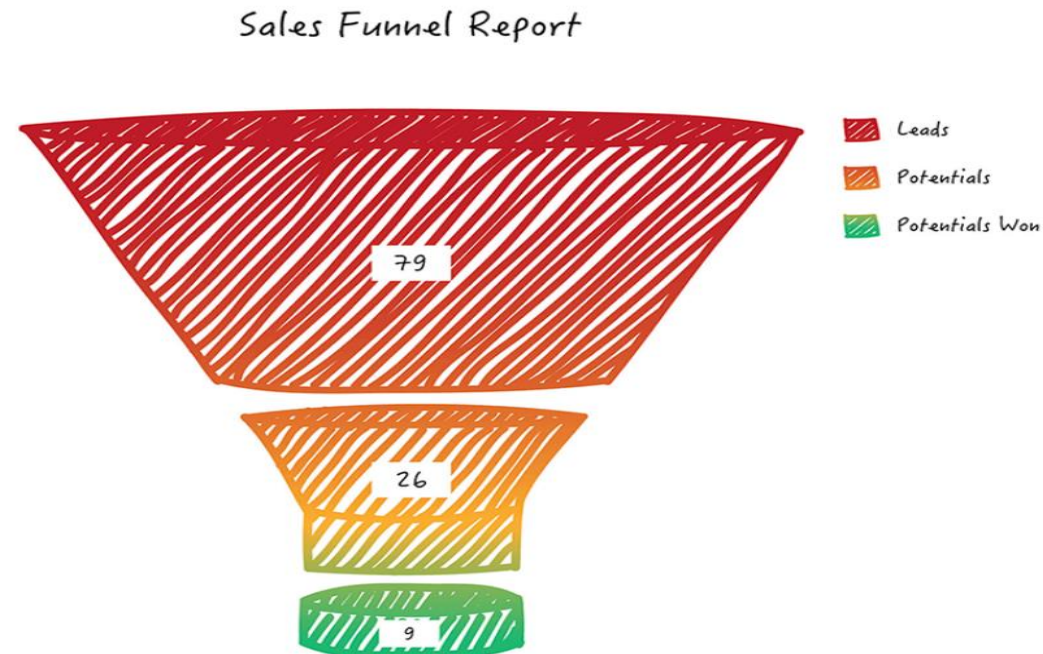
Web Chart

Pick this chart to study the comparison of different data series. A radar chart compares the values of a number of data series represented with data markers, relative to a center point.



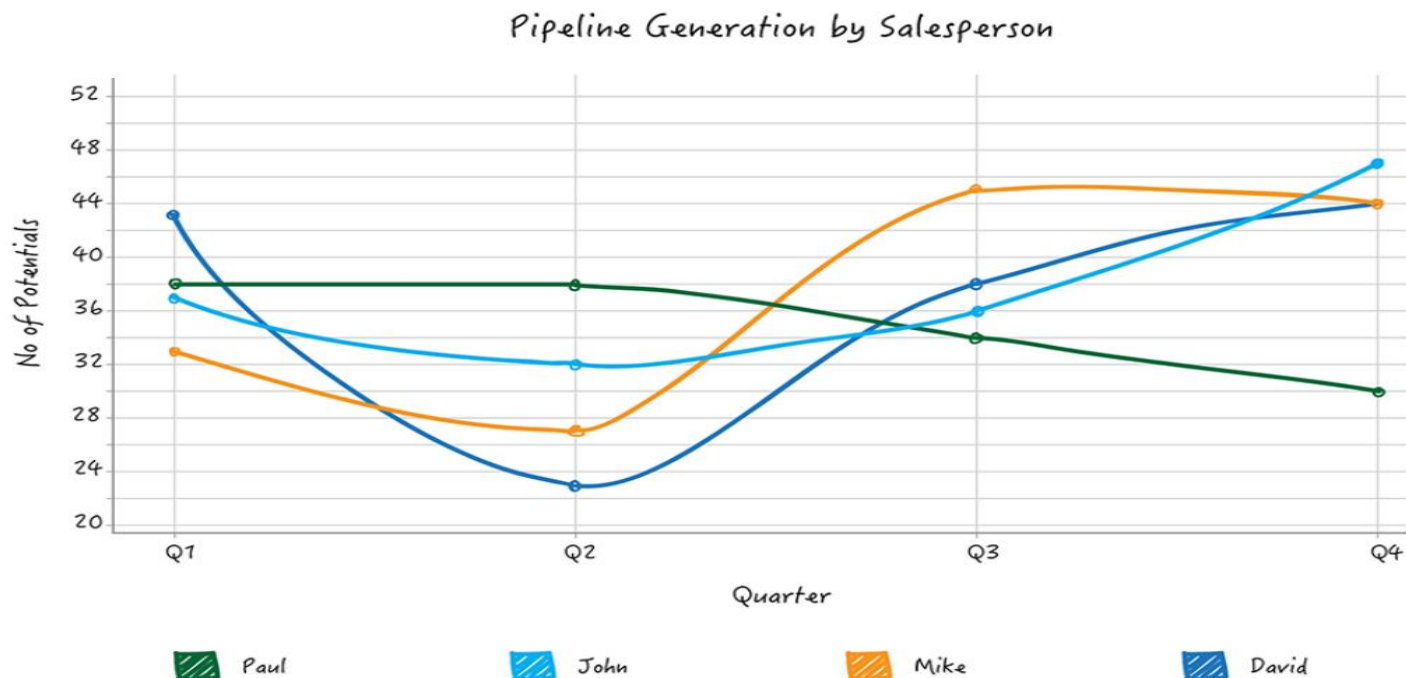
Funnel Chart

Pick this chart to represent progressive flow/reduction of a business metric across phases. For instance, visualize the conversion of your leads into actual sales across stages like marketing qualified leads, sales leads & won leads.



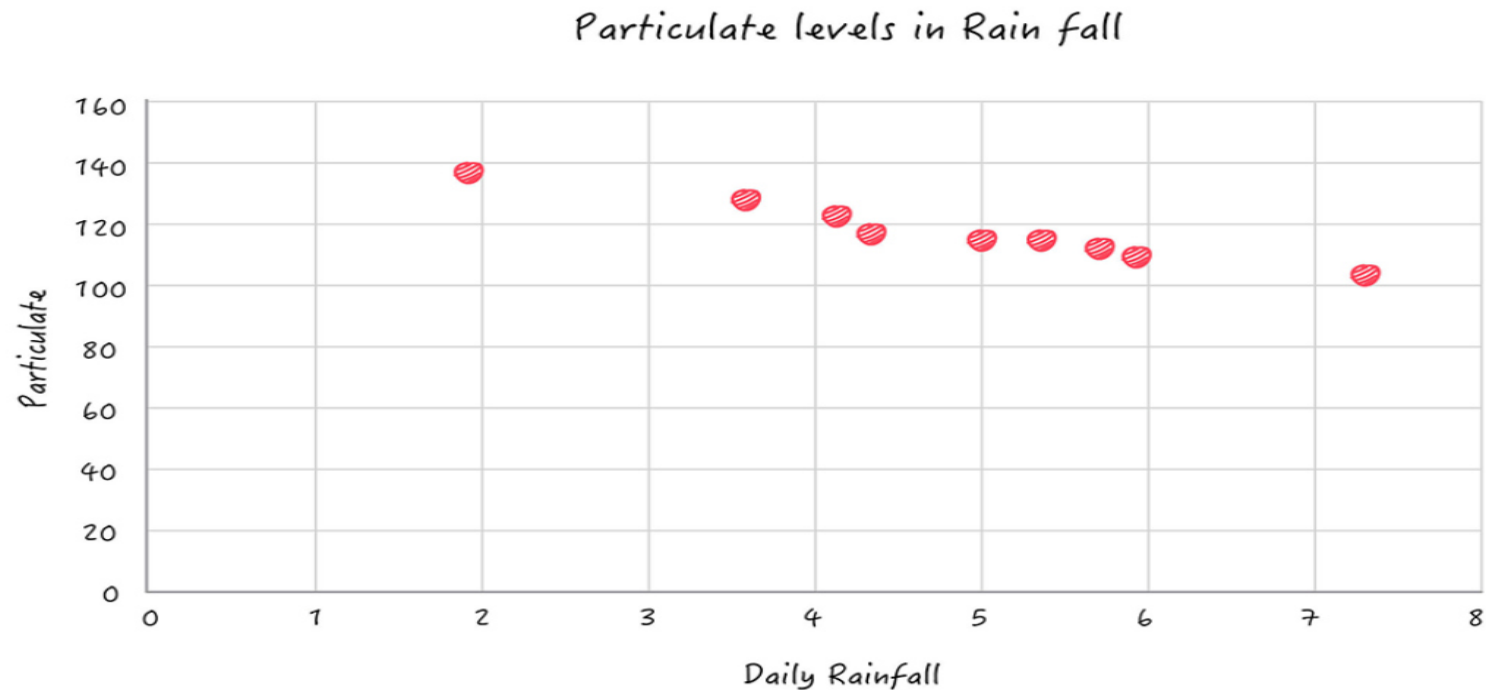
Line Chart

Pick this chart to visualize the trend of any data series across any time period. Just like visualizing the roller-coaster ride that your favorite stock went through in the last quarter.



Scatter Chart

Typical use of the scatter chart is to plot sporadic data with uneven intervals. It is used for the comparison between two numerical axes unlike the line chart, where one axis is never numerical.



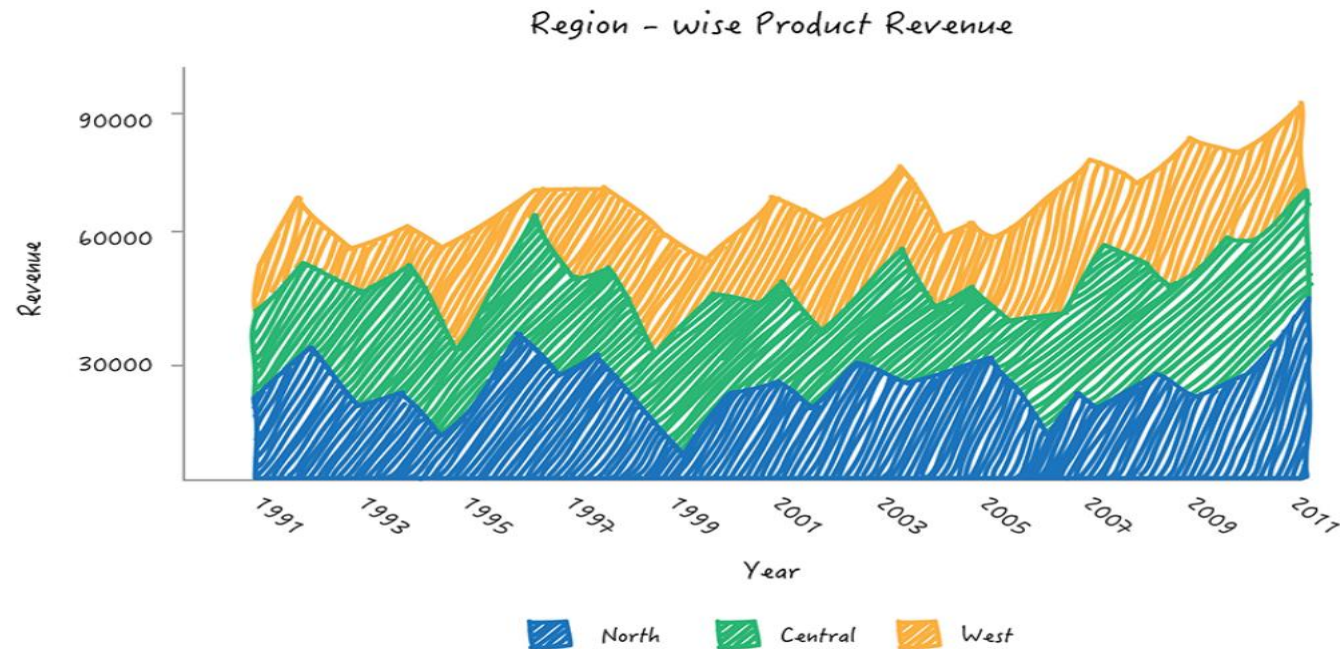
Area Chart

Area charts shade the area beneath the lines and therefore help you more readily compare data magnitudes. They are mainly useful for emphasizing the change in metrics across time. For instance, change in sales figures over a particular time period.



Stacked Area Chart

A stacked area chart shows the relationship of parts to a whole. With this chart, find how much the individual stacks or factors have contributed to the total value, across time.



Bullet Chart

A bullet chart highlights a primary measure and compares it with a target value, with respect to qualitative ranges of performance, such as poor, satisfactory, and good. Bullet charts can be typically used as widgets in dashboards as they pack a lot of information within a smaller space.

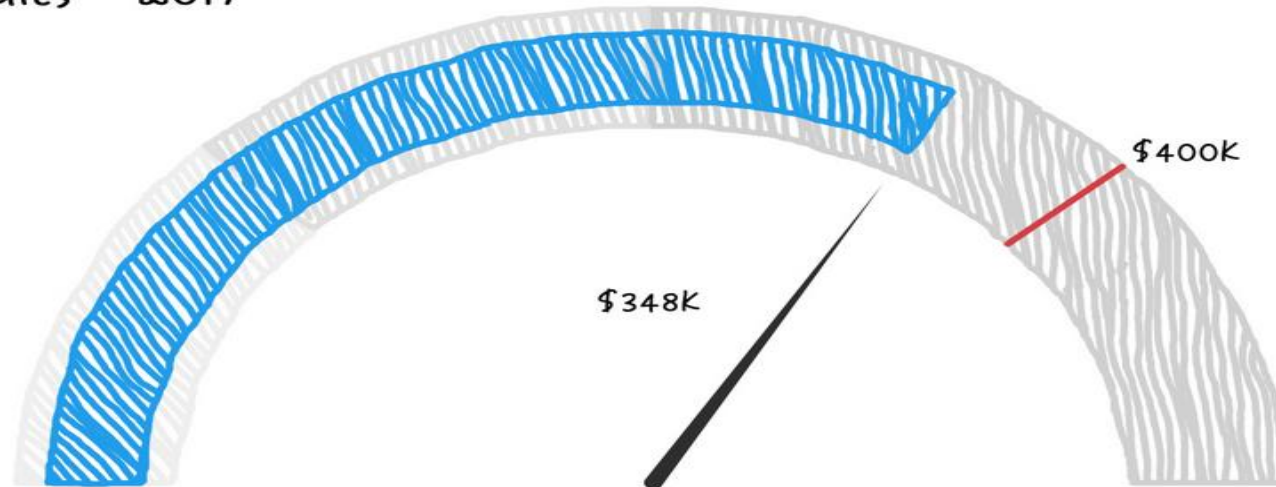
Total Sales



Dial Chart

Pick this chart to indicate the current value in a range. It is similar to a bullet chart, but the values are displayed on a dial. Dial charts are a perfect fit for your business and executive dashboards.

Sales - 2017



Bubble Chart

Pick this chart if you would like to add another magnitude - size to your visualizations. Bubble charts are extremely useful in highlighting the weightage of a data metric.



Packed Bubble Chart

Packed Bubble, as the name suggests displays data in a cluster of circles or bubbles. They are used to display the values disregarding the axes. You can use a packed bubble chart to visualize large amount of data in a small space.



Create Report

Zoho Analytics, as a reporting software, offers a wide variety of reporting options such as Charts, Pivot tables, Summary views and Tabular views.

This enables you to easily analyze your data and derive great insights.

The intuitive drag and drop based designer of Zoho Analytics allows you to easily construct the required reports.

Pivot Table

Pivot Table allows you to dynamically rearrange, group and summarize data for easy analysis of large sets of data.

You can transform data in a table(s) into interactive and meaningful summaries easily by using intuitive drag and drop interface provided by Zoho Analytics.

You can also filter, sort and customize the appearance and content of your Pivot Table the way you want it, by using wide range of options provided by Zoho Analytics.

For users who view/access the pivot tables, Zoho Analytics offers a range of interactivity options like View underlying data, User Filters, Sorting, Export etc.,

Pivot Table

the basic layout of a Pivot Table in **Design Mode**.

Sales for Region and Product Category by Year Save

Total Sales for each Region and Product Category by each Year (based on Date)

Sort Show/Hide Totals Underlying Data Themes

Pivot Filters (2) User Filters Reset All

Columns: Date Year **2** Region Actual **2**

Rows: Product Cate... Actual **1** Product Actual **1**

Data: Cost Sum **3**

Product Category	Product	2018 2			Summary 4
		Central	East	West	Total Cost
Grocery	Fruits and Vegetables	\$6,358.48	\$19,198.06	\$6,322.44	\$31,878.98
	Meat		\$876.85	\$38.42	\$915.27
	Spices	\$6,365.29			\$6,365.29
Grocery		\$12,723.77	\$20,074.91	\$6,360.86	\$39,159.54
Stationery	Air Purifiers		\$1,060.58		\$1,060.58
	Art Supplies	\$7.70	\$1,180.66	\$2.39	\$1,190.75
	Binder Clips		\$79.82		\$79.82
	Binding Supplies			\$173.16	\$173.16
	Computer Paper		\$2,873.96	\$1,789.01	\$4,662.97
	Copy Paper		\$1,673.06		\$1,673.06
	File Labels	\$9.85	\$61.64		\$71.49
	Heavy Duty Binders			\$252.83	\$252.83
	Paper Clips			\$2.66	\$2.66
	Pencils	\$42.07			\$42.07
	Pins and Tacks		\$5.01		\$5.01
	Round Ring Binders		\$30.69		\$30.69
	Scissors			\$569.35	\$569.35
	Specialty Envelopes	\$1,392.25			\$1,392.25
Standard Labels		\$91.25	\$115.77	\$207.02	
Stationery		\$1,451.87	\$7,056.67	\$2,905.17	\$11,413.71
Grand Total:		\$14,175.64	\$27,131.58	\$9,266.03	\$50,573.25

1. Columns from the source table that you assign to a row orientation in a Pivot Table.
2. Values of the column from the source table that you assign to a column orientation in a Pivot Table.
3. Column from the source table that contains values to be summarized.
4. Column/Rows in the Pivot Table that contains grand summary information.

Summary View

Summary View, as the name indicates, is a report type that enables you to view your summarized data in tabular formats. This report type is extremely useful when you need to analyze huge amount of data with logical grouping and appropriate summarizations in a visually intuitive manner. You can also filter, sort, apply conditional formatting and do much more.

	Product Category ↓	Product ↓	Total Sales	Total Profit
1	Furniture	End Tables	\$18,791	\$17,421
2		Meeting Room Tables	\$32,521	\$25,495
3		Office Chairs	\$50,905	\$33,216
4	Grocery	Fruits and Vegetables	\$763,815	\$480,623
5		Meat	\$80,200	\$46,224
6		Spices	\$33,382	\$16,084
7	Stationery	Art Supplies	\$33,590	\$19,304
8		Computer Paper	\$89,292	\$50,758
9		Copy Paper	\$19,993	\$12,371
10		Storage Cabinets	\$17,485	\$15,063

Create Dashboard

Visualize and analyze your data with dashboards.

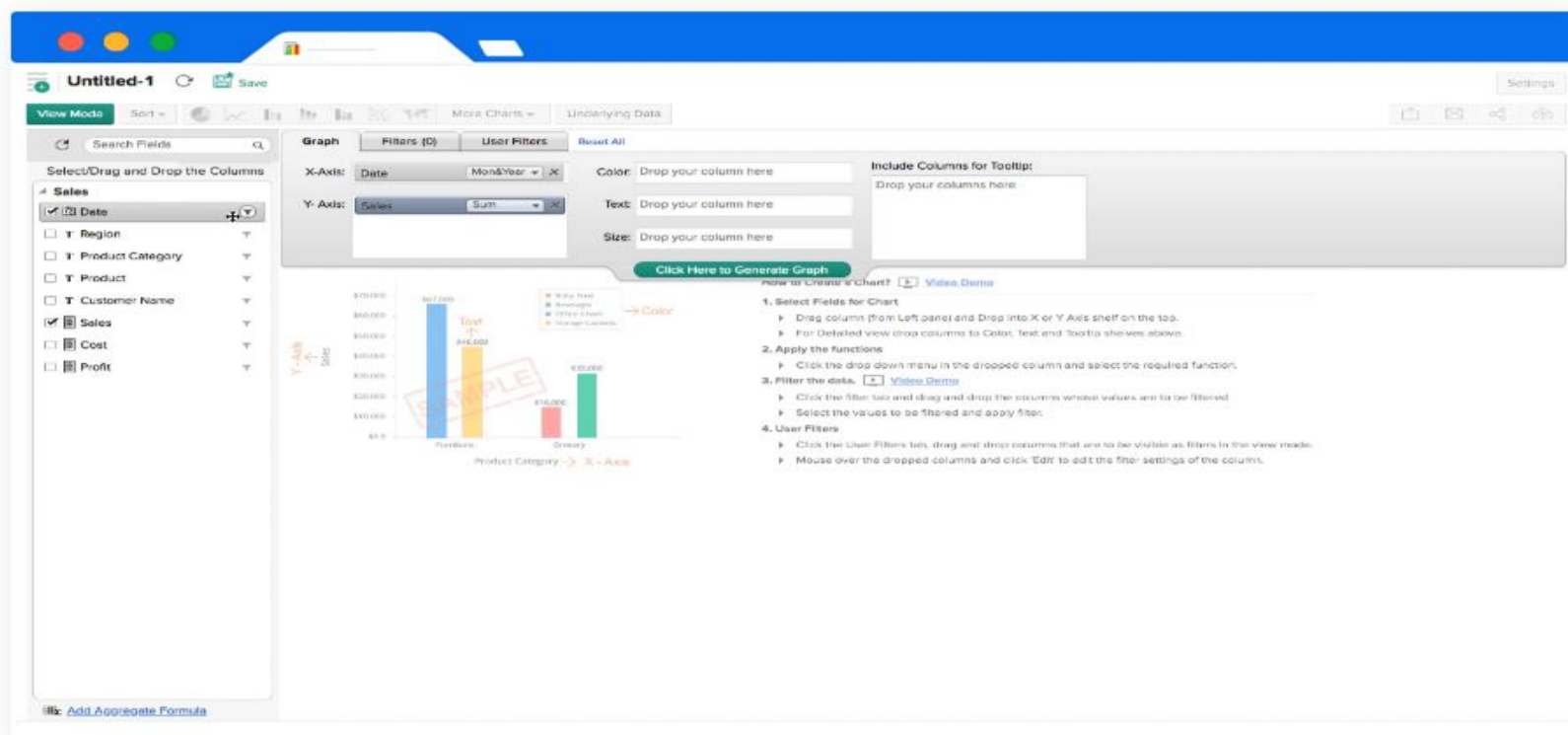
Create insightful business dashboards with our drag-and-drop dashboard creator to track your key performance indicators (KPIs), and share them with your colleagues for easier data analysis and collaboration.

Business Dashboards



Drag-and-drop interface

Build powerful business analytics dashboards in a flash with our easy drag-and-drop interface. No IT help required!



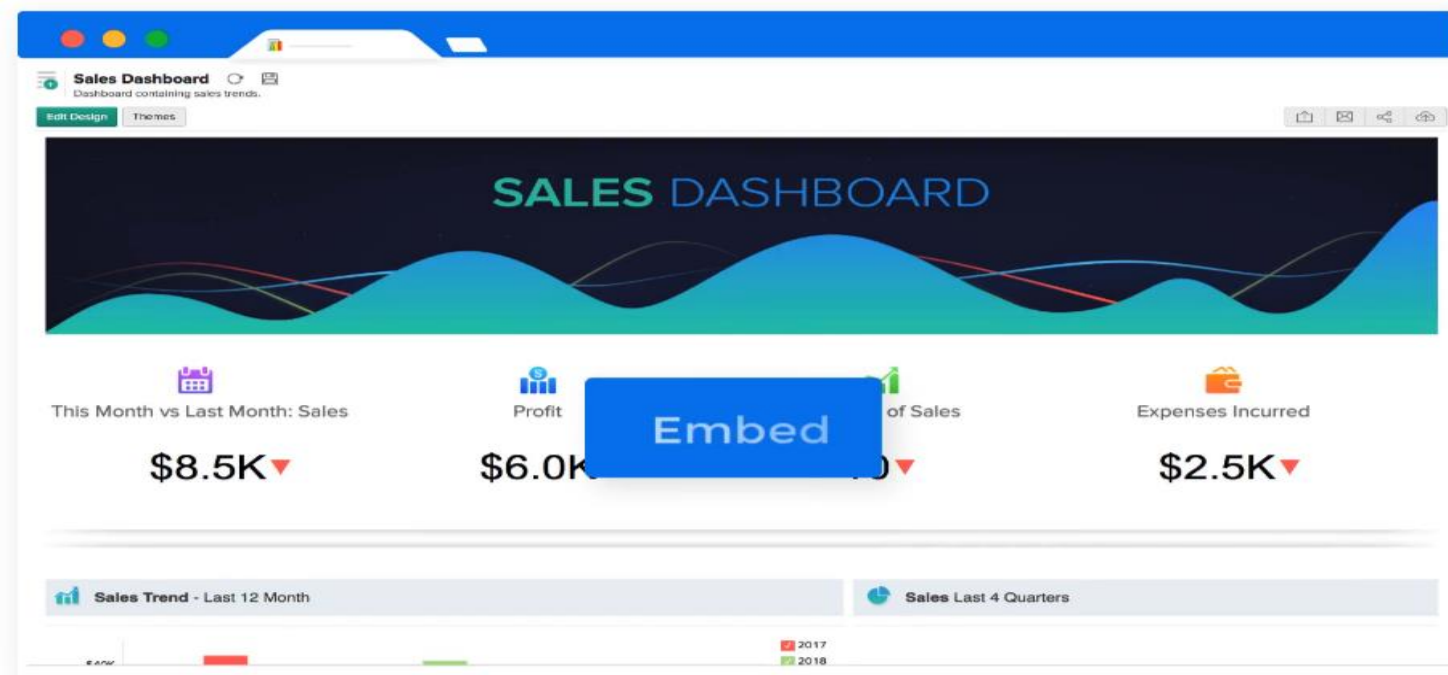
Wide range of reporting components

Use a variety of charts, pivot tables, KPI widgets, and tabular view components to build your dashboards. Drill down into your data for the insights you need.



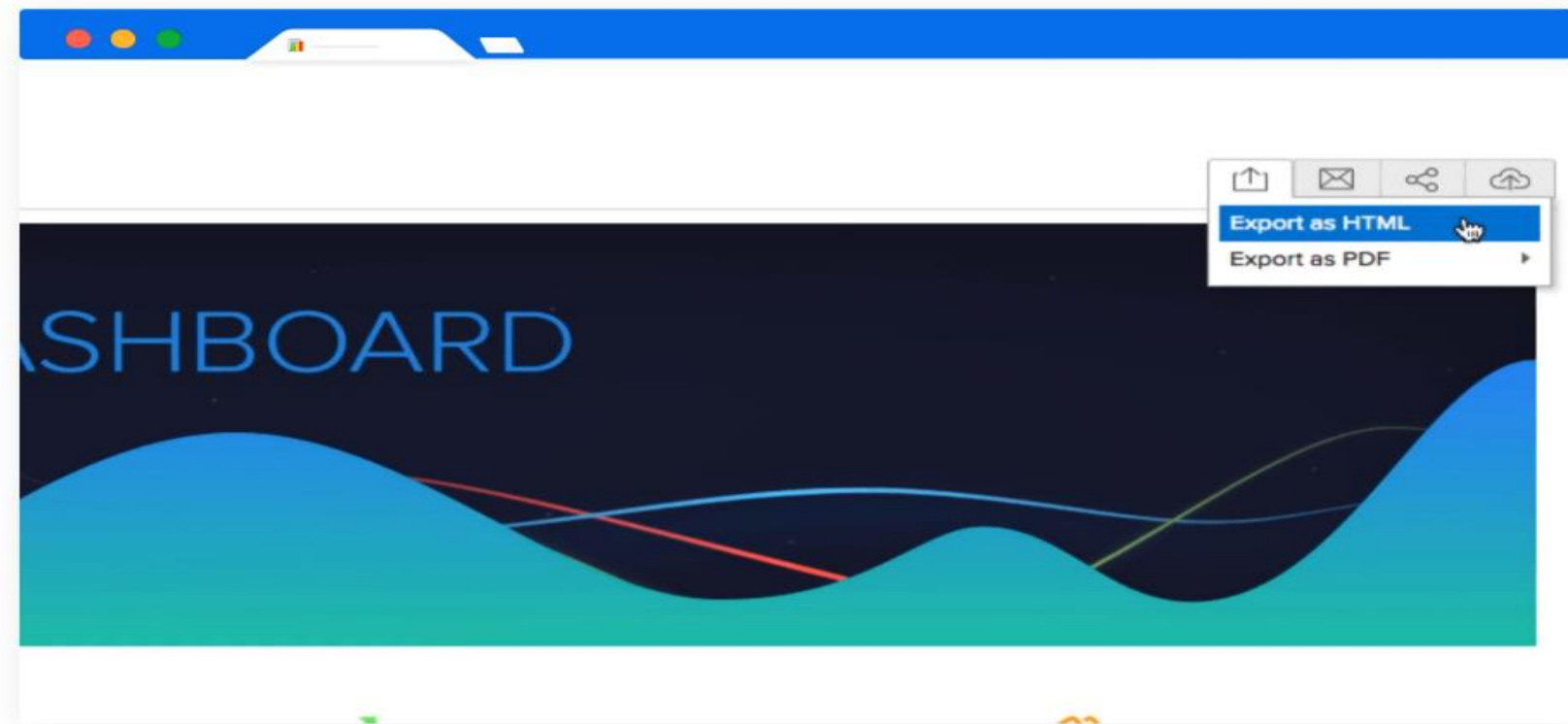
Share and embed

Collaborate on dashboard creation and make more informed decisions as a team, with our business dashboard software. Embed them across various online platforms for mass consumption and provide fine-grained access control.



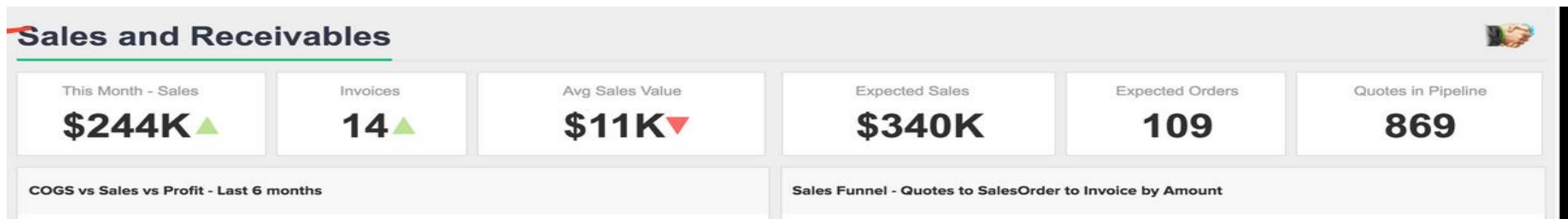
Export and email

Schedule your dashboards to be emailed automatically. Export or print dashboards using our online dashboard creator, for offline access.



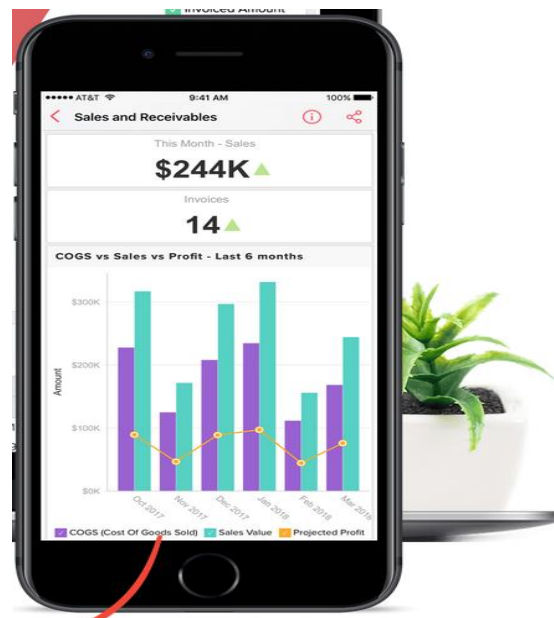
Deeper analytics

Build new calculations with an intuitive formula engine that offers extensive mathematical and statistical functions.



Accessible across all devices

Access the business performance dashboards you create from PCs, laptops, tablets, and mobile phones.



Data Analysis

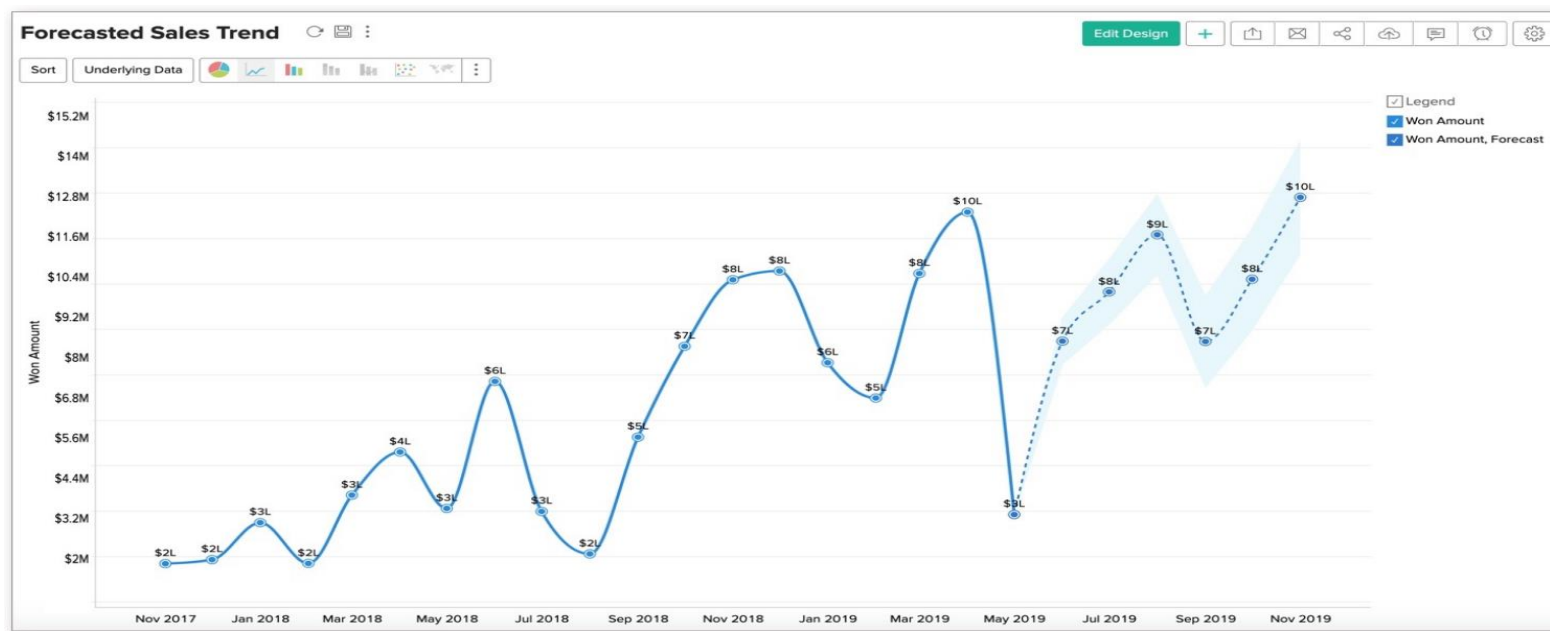
Zoho Analytics enables you to analyze your data with advanced analytical options.

You can easily predict future values or identify the past & future trends in your data with a simple process, without worrying about the underlying complexity.

You can also derive complicated calculations in just a few steps to create powerful reports.

Forecasting

Zoho Analytics enables you to effectively predict your future data trends, using its forecasting feature. Forecasting in Zoho Analytics is based on powerful forecasting algorithms which analyze your past data deeply and come up with the best forecast for the future. You can set up a forecast in a chart using a very simple set up process, without worrying about the underlying complexity.



What is forecasting?

Forecasting is a process of predicting the future based on the past data trend.

Zoho Analytics forecasting is based on powerful forecasting algorithms which analyzes your past data deeply and comes up with the best forecast for the future.

What are the chart types that support Forecasting?

Zoho Analytics supports forecasting for the following chart types.

- Line Chart
- Bar Chart
- Stacked Bar Chart
- Scatter Chart
- Area Chart
- Stacked Area Chart
- Combo Charts (without Bubble Chart).

How Forecasting works in Zoho Analytics?

1. The forecasting engine analyses past data points.
2. Based on the past data, the forecasting engine will identify the periodicity using auto-correlation method.
3. Then it computes the seasonality, trend and randomness using the past data.
4. By iterative processing, the forecasting engine fine-tunes the computed seasonality, trend and randomness.
5. The engine runs linear, logarithmic and exponential regressions and identifies the data series falls under linear, logarithmic or exponential.
6. The accuracy of the predicted results will be verified using Hindcasting. Hindcasting is the process of predicting the past data points and verify the same with the actual points.
7. Once all verifications are done, the forecasting engine produces the final forecasted points

Trend Line

Trend Lines are used to identify the trend present in your data.

Zoho Analytics can plot the past & future trends in your time series or number series data.

Zoho Analytics provides five such models to derive the Trend Line.

What are Trend Lines?

Trend Line are used to identify the trend present in your data.

Zoho Analytics predicts the Trend line for data in time or number series.

How are Trend Lines derived?

Trend Lines can be derived based on various mathematical models. Zoho Analytics provides five models to derive the Trend line. They are Linear, Logarithmic, Exponential, Power, and Polynomial.

- Based on the model selected, along with the data from the chart, possible trend values will be predicted.
- Error square will be identified by calculating the square of the difference between the actual Y and predicted trend Y value.
- The above process will be repeated by changing the parameters and checking whether the error square is reduced.
- The Trend Line will be plotted with the value where the error square is the least.

What are the models supported in Zoho Analytics for predicting the Trend line?

Zoho Analytics supports the following models to predict the Trend Line.

Linear - Choose to apply a Linear model to derive the Trend Line.

Logarithmic - Choose this to apply a Logarithmic model to derive the Trend Line. Not best suited if data points have negative values.

Exponential - Chose this to apply an Exponential model to derive the Trend Line.

Power - Chose this to apply a Power factor model to derive the Trend Line. This is another form of Exponential model where the natural log is applied on factors before applying exponential.

Polynomial - Chose to apply a Polynomial model to derive the Trend Line. The degree value provided is used to derive a polynomial series for the specified degree.

Auto - Choose this to allow Zoho Analytics to select the best Trend Line model based on the given data points.

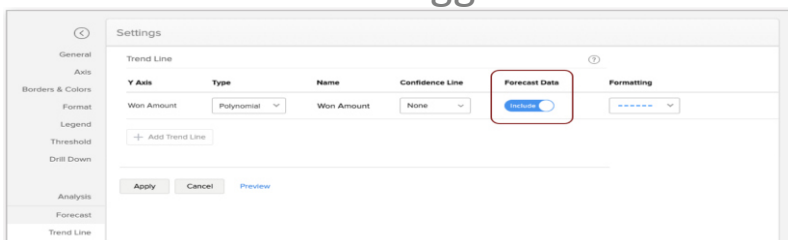
What are the chart types that support the Trend Line?

The following chart types support Trend Line.

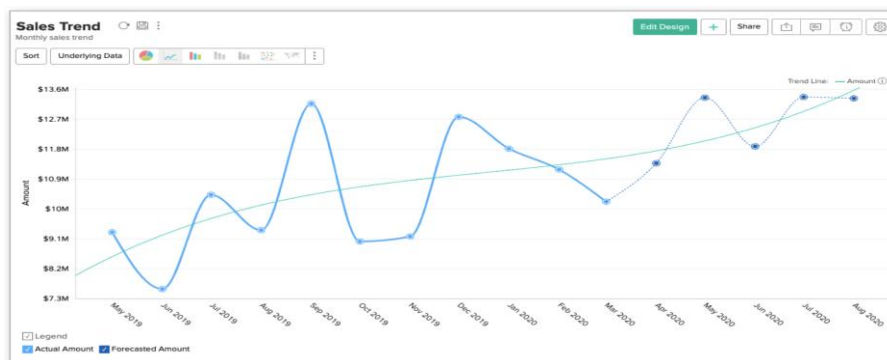
- Line Chart
- Bar Chart
- Stacked Bar Chart
- Scatter Chart
- Area Chart
- Stacked Area Chart
- Bubble Chart (not Bubble Pie and Packed Bubble)
- Combo Charts

Set up a Trend Line for forecasted data

you can predict the Trend Line for forecasted data. To include trend line for forecasted data for Trend Line calculation, set the **Forecast Data** toggle button as **Include**.

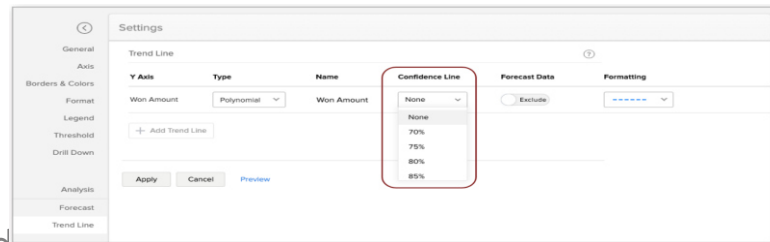


The following illustration shows the Trend Line for forecasted data.

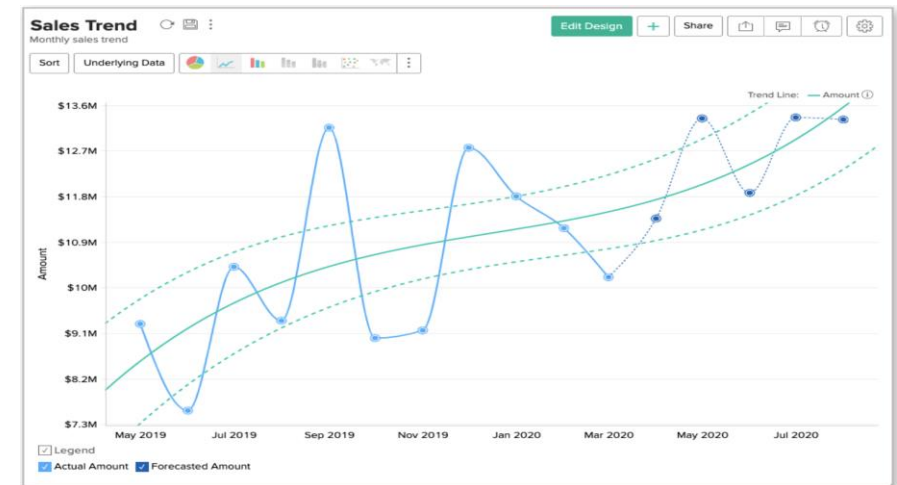


Confidence Line for Trend line

The **Confidence Line** predicts what the Trend Line can be for a similar data set for the given probability. You can select the probability range in the **Confidence Line** from **70% to 95%**.

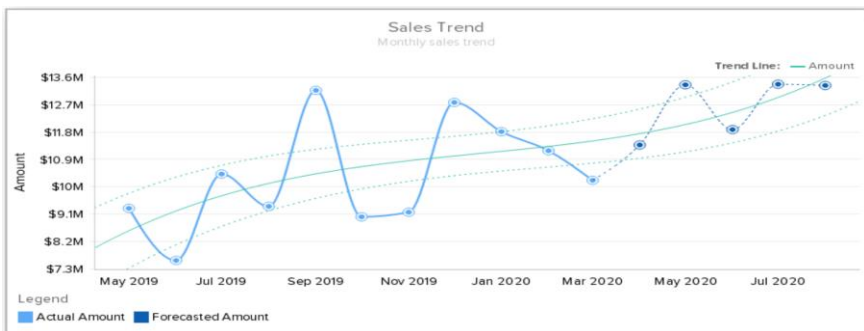


The below chart is drawn with a 90% confidence range. i.e., when you change the underlying data with a similar data set (populated with the same conditions), then the Trend Line will be 90% within this Confidence Line.



Export Trend Line data

you can export Trend Line data along with the chart.



When exporting into a tabular format (CSV, Excel, Zoho Sheet), Trend Line data will be exported as columns.

Date	Amount	Type	Total Amount Trend line	Total Amount Trend line - Lower Confidence Limit	Total Amount Trend line - Upper Confidence Limit
May 2019	\$92,81,700.00	Actual	\$85,46,217.26	\$73,17,141.10	\$97,75,293.42
Jun 2019	\$75,70,200.00	Actual	\$91,94,155.87	\$80,83,386.96	\$1,03,04,924.77
Jul 2019	\$1,04,09,300.00	Actual	\$96,94,548.20	\$86,92,479.35	\$1,06,96,617.05
Aug 2019	\$93,45,900.00	Actual	\$1,01,00,867.29	\$92,02,913.82	\$1,09,98,820.76
Sep 2019	\$1,31,56,800.00	Actual	\$1,04,15,061.07	\$96,09,675.45	\$1,12,20,446.68
Oct 2019	\$90,02,500.00	Actual	\$1,06,50,868.71	\$99,19,921.94	\$1,13,81,815.48
Nov 2019	\$91,54,300.00	Actual	\$1,08,44,124.29	\$1,01,69,137.00	\$1,15,19,111.57
Dec 2019	\$1,27,59,800.00	Actual	\$1,10,01,612.85	\$1,03,55,318.96	\$1,16,47,906.73
Jan 2020	\$1,18,00,111.00	Actual	\$1,11,53,947.18	\$1,05,07,402.87	\$1,18,00,491.49
Feb 2020	\$1,11,73,000.00	Actual	\$1,13,16,150.40	\$1,06,39,031.20	\$1,19,93,269.60
Mar 2020	\$1,02,02,900.00	Actual	\$1,14,95,129.54	\$1,07,65,269.51	\$1,22,24,989.58
Apr 2020	\$1,13,62,690.53	Forecast	\$1,17,35,140.17	\$1,09,28,374.67	\$1,25,41,905.68
May 2020	\$1,33,38,349.88	Forecast	\$1,20,34,352.91	\$1,11,37,994.30	\$1,29,30,711.52
Jun 2020	\$1,18,68,671.88	Forecast	\$1,24,32,828.82	\$1,14,32,512.49	\$1,34,33,145.15
Jul 2020	\$1,33,57,662.37	Forecast	\$1,29,24,112.94	\$1,18,15,207.31	\$1,40,33,018.57
Aug 2020	\$1,33,16,302.91	Forecast	\$1,35,61,071.09	\$1,23,33,942.40	\$1,47,88,199.77

Ask Zia

Amidst the millions of the raw data, lies a huge business opportunity. Bringing out hidden insight is a big challenge. This is where Ask Zia comes in.

Zia, Zoho's smart and Intelligent Assistant based on machine learning (ML) and artificial intelligence (AI) technologies, is now part of Zoho Analytics. Ask Zia provides the capability to ask questions in natural language and get powerful insights quickly.

Ask Zia is driven by natural language processing (NLP). It understands your questions asked in plain English, and gets you powerful insights as answers in the form of cool and relevant visualizations.

What is Ask Zia?

Ask Zia is a search-driven analytical assistant that helps you create reports instantly.

You can ask questions in natural language and Ask Zia generates the most relevant reports as answers to you.

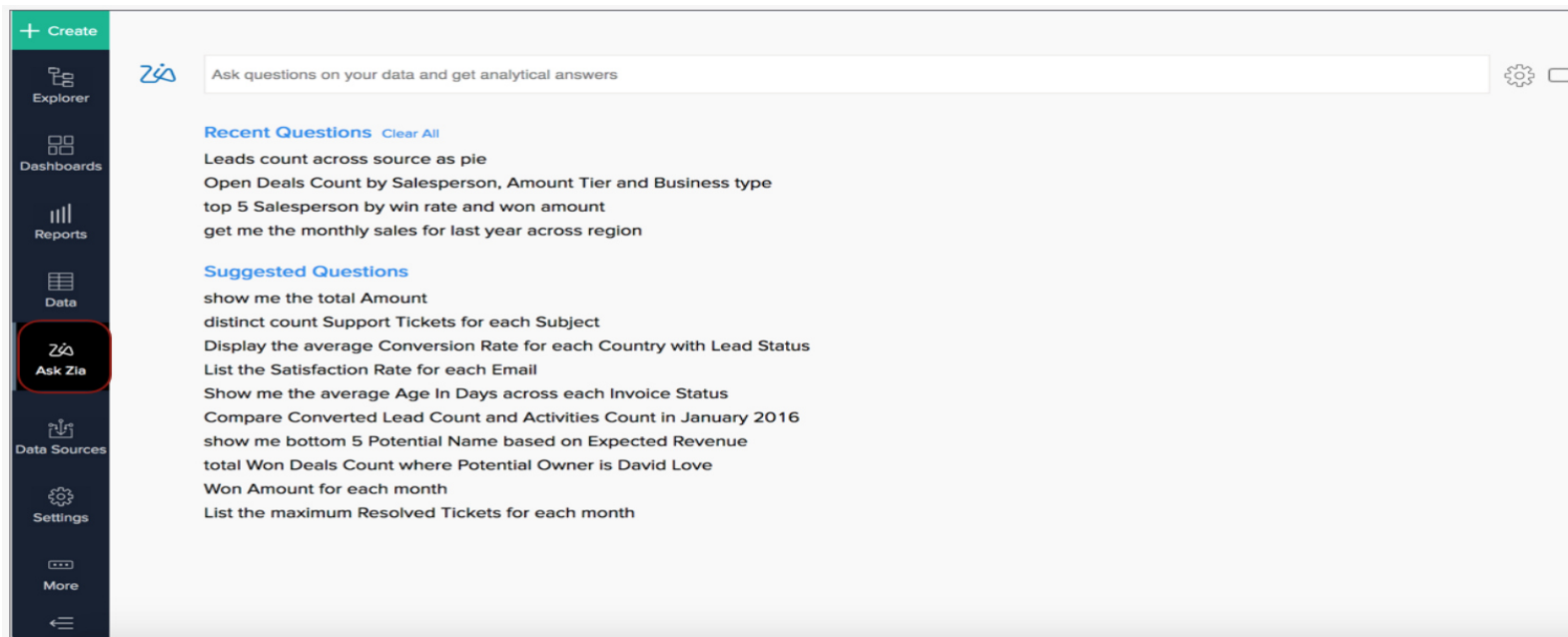
Who can use Ask Zia

To use Ask Zia you need permission to create reports in that Workspace. One of the following user privileges allows you to do this.

- Account Administrator
- Workspace Administrator
- Shared User with access to a data table

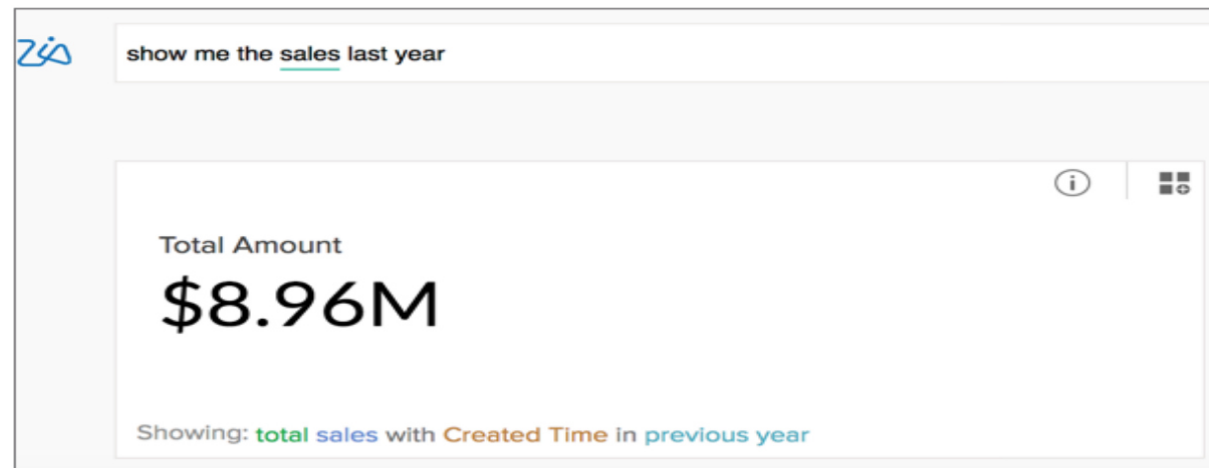
Using Ask Zia

You can easily invoke this by clicking the **Ask Zia** option on the left panel of your workspace. The **Ask Zia** screen will open.



Using Ask Zia

Type your question in the search box provided. Ask Zia will interpret the question and will get the best possible report as an answer.

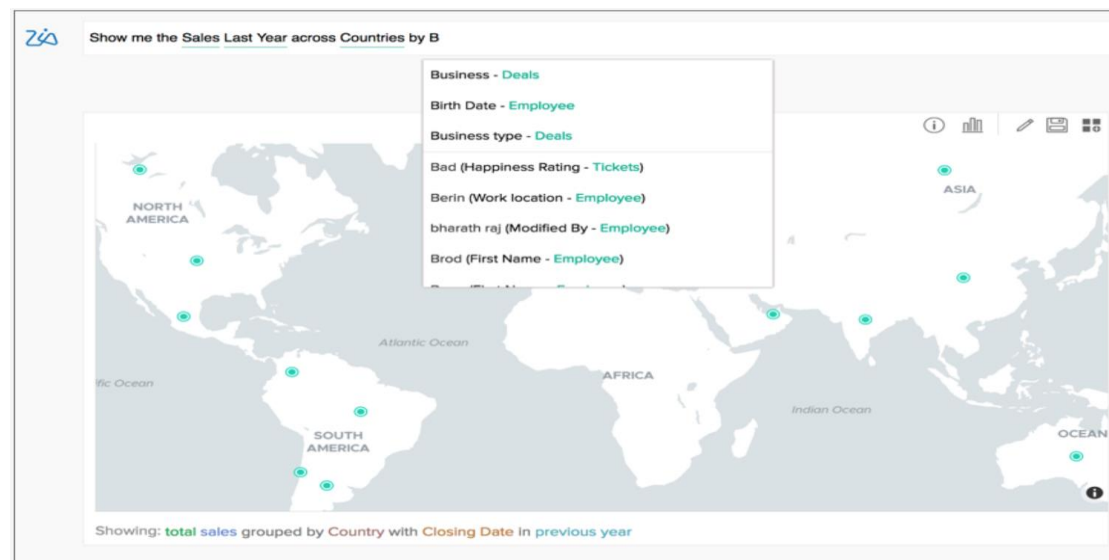


You can add more qualifiers to your question to have Ask Zia generate the most relevant insight for you.

Auto Suggestion

Ask Zia will keep learning from the patterns in your question and provides intelligent suggestions to assist you in drafting the question. These suggestions are highly personalized based on the relationship among the tables, previous history, Ask Zia customization etc.

The Auto Suggestion also helps users by suggesting supporting phrases like Top 10 / Bottom 10, sort options, chart types etc.,

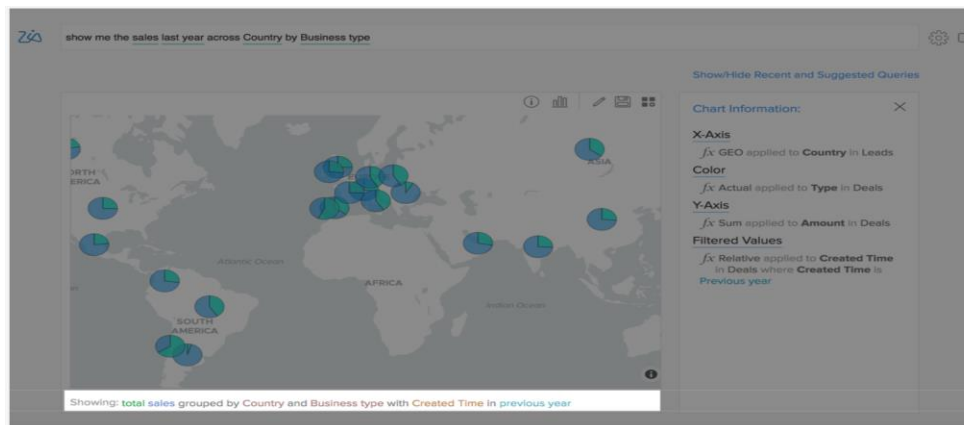


Generated Report

Ask Zia will understand the intent in your question, resolve ambiguities and generates the most appropriate report that matches your question.

The generated report will be displayed as a preview.

The Preview screen displays a detailed description and report information on how the report is plotted. You can verify the report and save them.



Change Chart Type

When Ask Zia generates a chart, you can choose to change the type of chart. Follow the steps given below to do this.

1. Click the **Chart** icon in the contextual menu on top of the Preview screen.
2. Select the required chart type from the drop-down.

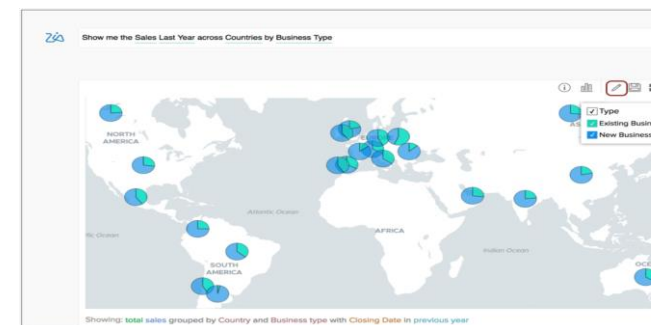


3. Save the report.

Edit the Report

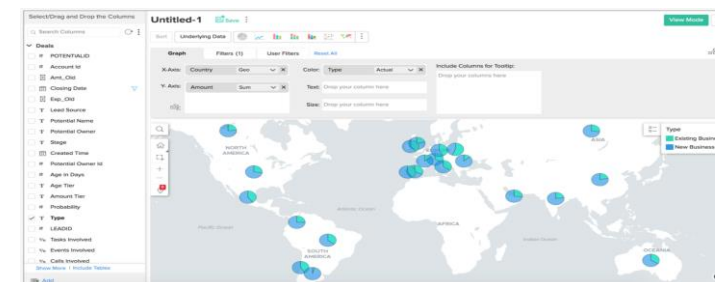
You can easily edit the report that Ask Zia generates. Follow the below steps to edit the report.

1. Click the **Edit** icon in the contextual menu on top of the Preview screen.



2. This will open the report editor with the columns recommended by Ask Zia. You can modify them as needed.

3. Save the report.



Save Report

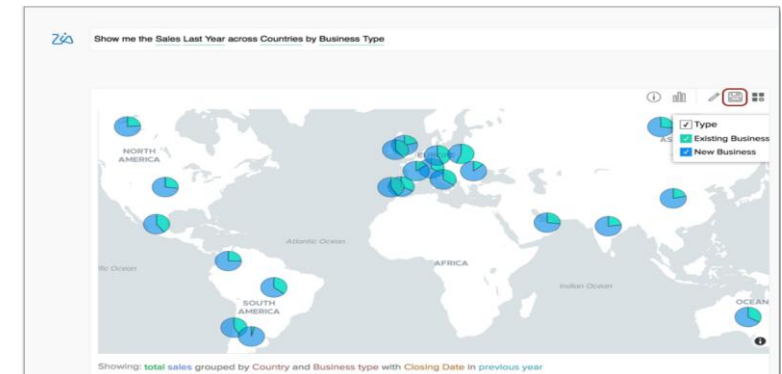
You can save the report permanently by following the below steps:

1. Click the **Save** icon in the contextual menu on top of the Preview screen.

2. The **Save** dialog will open. Specify the **View Name**.

3. Select the **Folder** to save the view.

4. Click **Save** to save the report.

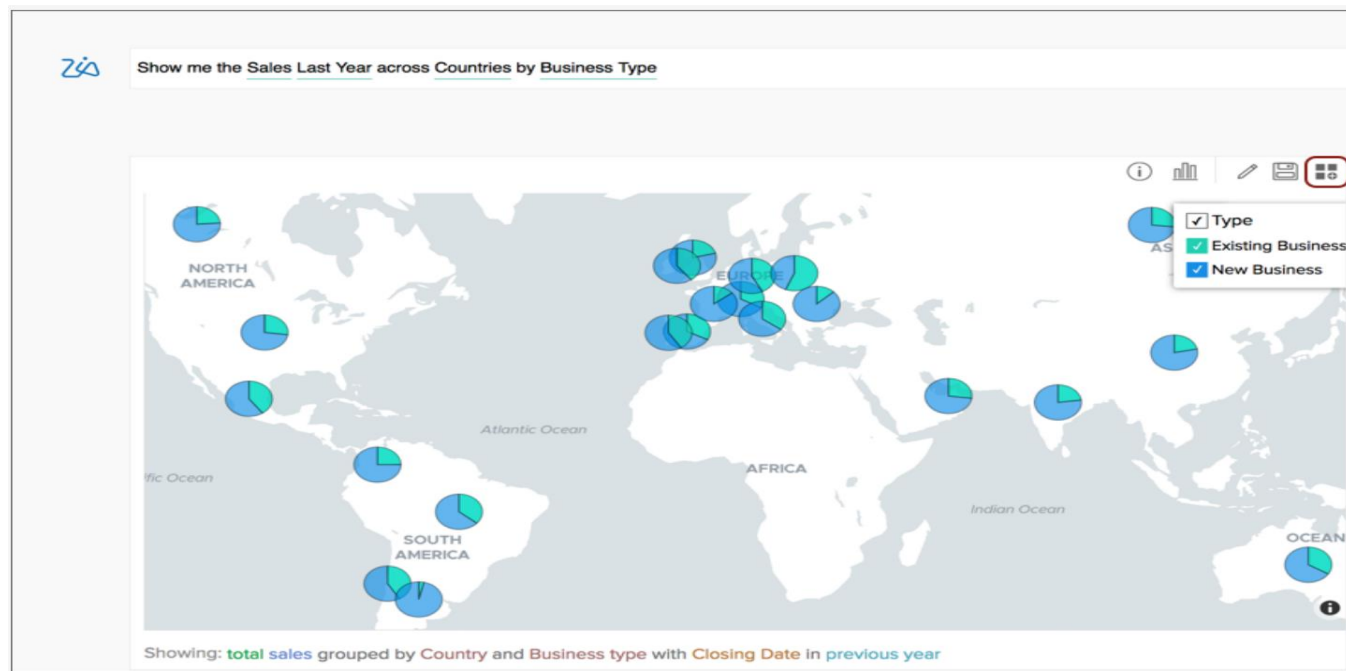


The screenshot shows the same dashboard as above, but with a "Save" dialog box overlaid on the map. The dialog box contains the following fields: "View name:" with the text "total sales grouped by Country and I", "Folder:" with a dropdown menu set to "Custom Reports", and "Description:" with the placeholder text "Enter a description (Optional)". There are "Save" and "Cancel" buttons at the bottom of the dialog box. The text at the bottom of the dashboard remains the same: "Showing: total sales grouped by Country and Business type with Closing Date in previous year".

Add Report to Dashboard

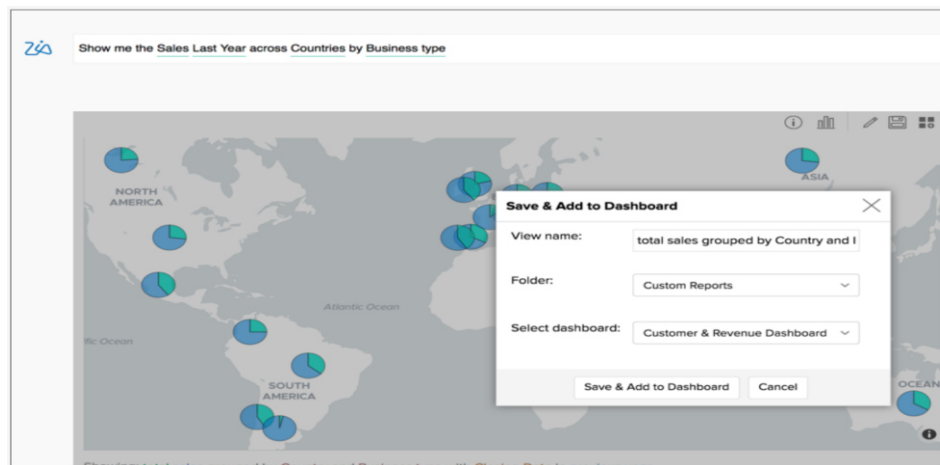
You can also directly add the reports to a dashboard. Follow the below steps to add the report to a dashboard:

1. Click the **dashboard** icon in the contextual menu on top of the preview.



Add Report to Dashboard

2. The **Save & Add to Dashboard** dialog will open. Specify the **View Name**.



3. Select the **Folder** to save the report before adding into the dashboard.
4. Select the **Dashboard** in which you want to add the report.
5. Click **Save & Add to Dashboard** to save the report and add it in the dashboard. The report will be added as the first view in the dashboard. You can realign the dashboard later as needed.

Constructing Questions for Ask Zia

You can ask your question to Zia in plain English.

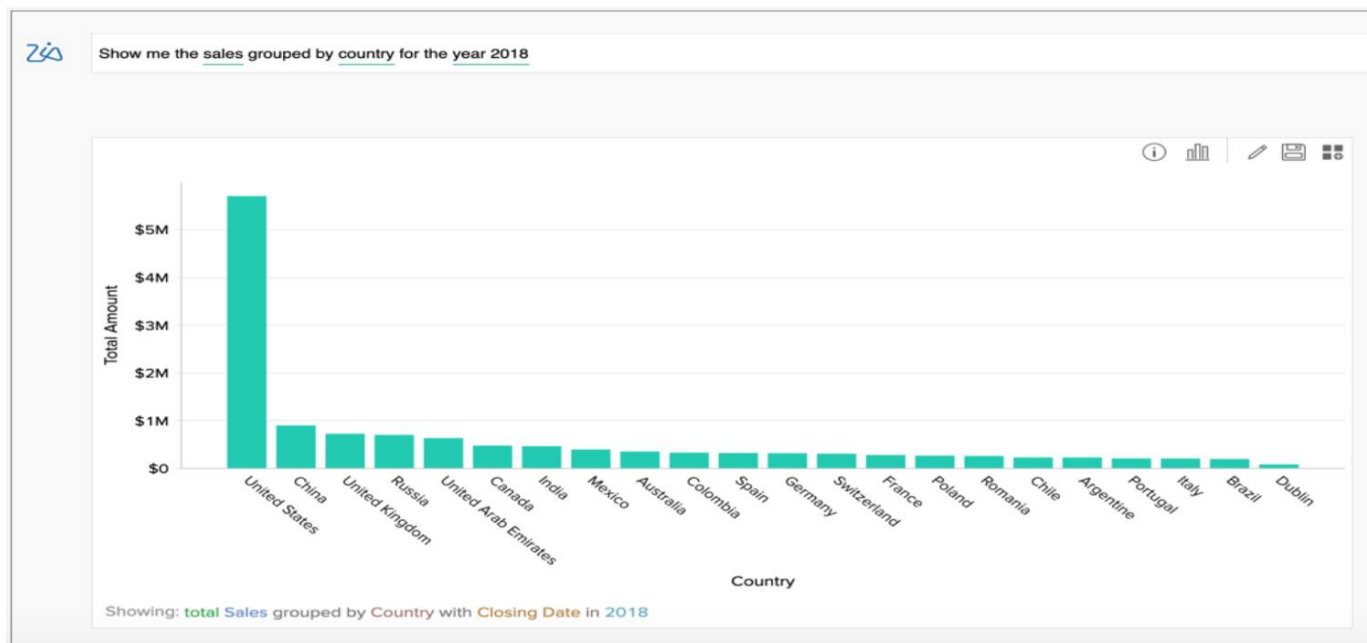
Zia will understand your question's intent and will reply with the most appropriate report that matches your question.

However, there can be some communication gaps like how it exists among people.

Structured Question

Your question can be in a structured language e.g.,

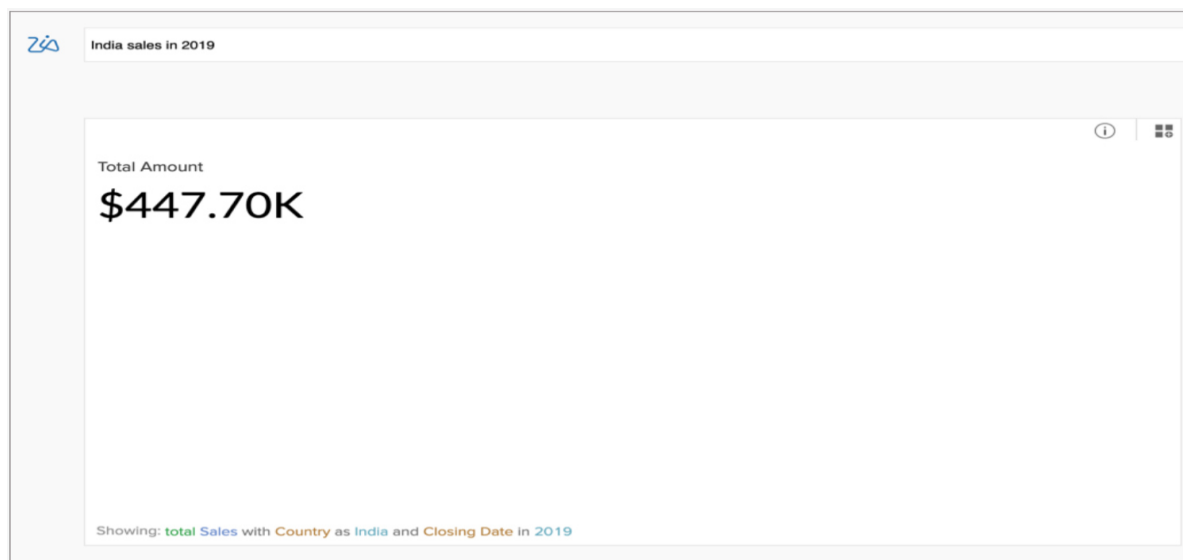
“Show me the sales grouped by country for the year 2018”



Unstructured Question

Your question can be in an unstructured fragment phrase e.g.,

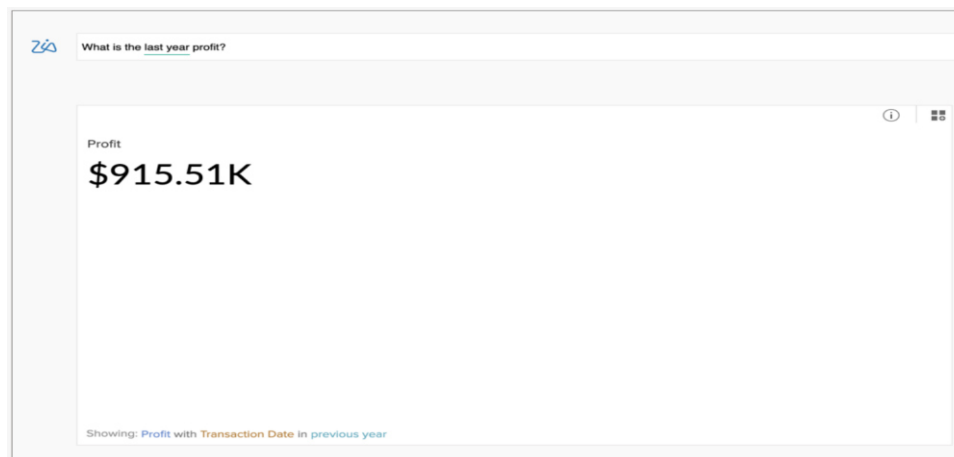
“India sales in 2019”



WH Questions

You can ask a 'WH' question. Ask Zia recognizes the 'WH' questions such as **What, When, who, which** e.g.,

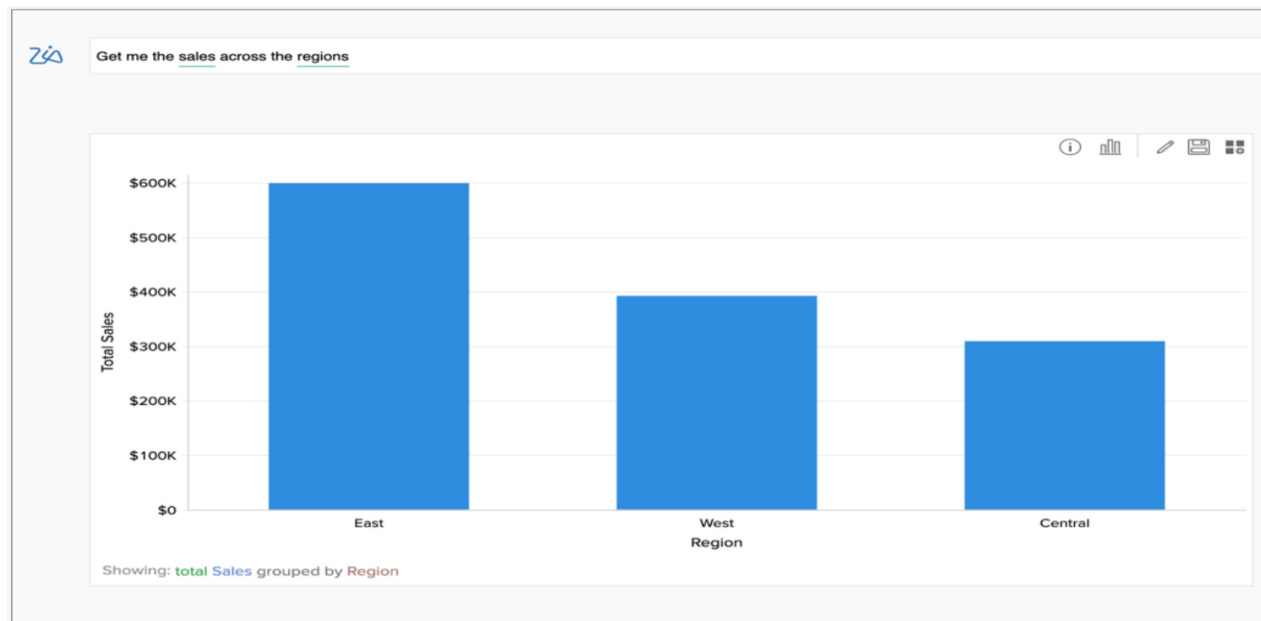
“What is the last year profit?”



Explicit Information

Your question states explicit information in your data. e.g., Let's say you have *Sales* and *Region* columns in your dataset. The below question will get the total sales for each region.

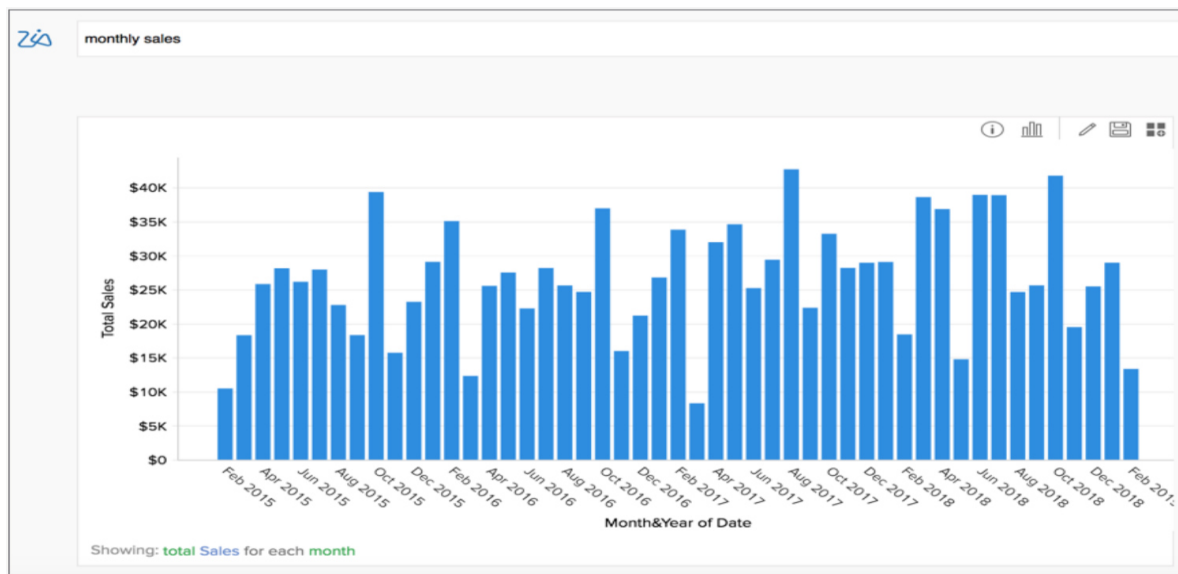
“Get me the sales across the regions”



Implicit Information

Your question contains implicit information. Ask Zia will identify the terms such as *monthly*, *yearly*, *daily*, *quarterly*. e.g., for the below question, Zia will recognize the term *Monthly*, identified a date column in your data set and plots the sales across months.

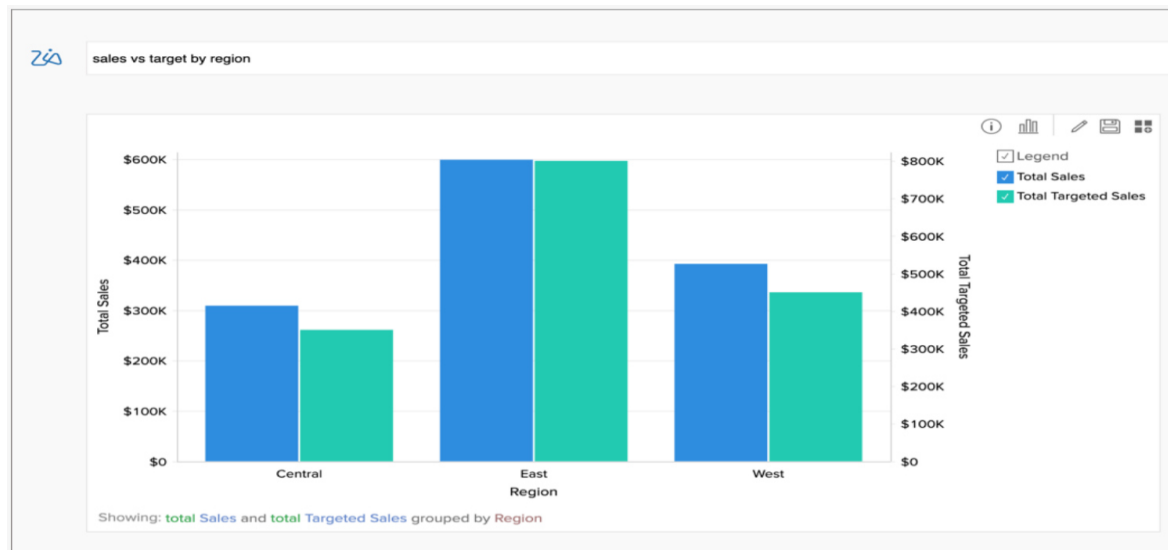
“Monthly Sales”



Across Tables

You can ask for metrics across tables in your dataset. Ask Zia will interpret it based on how you have joined the tables using Lookup column. In this example, the column Sales is from the Sales table and the column Targeted Sales is from Regional Target table.

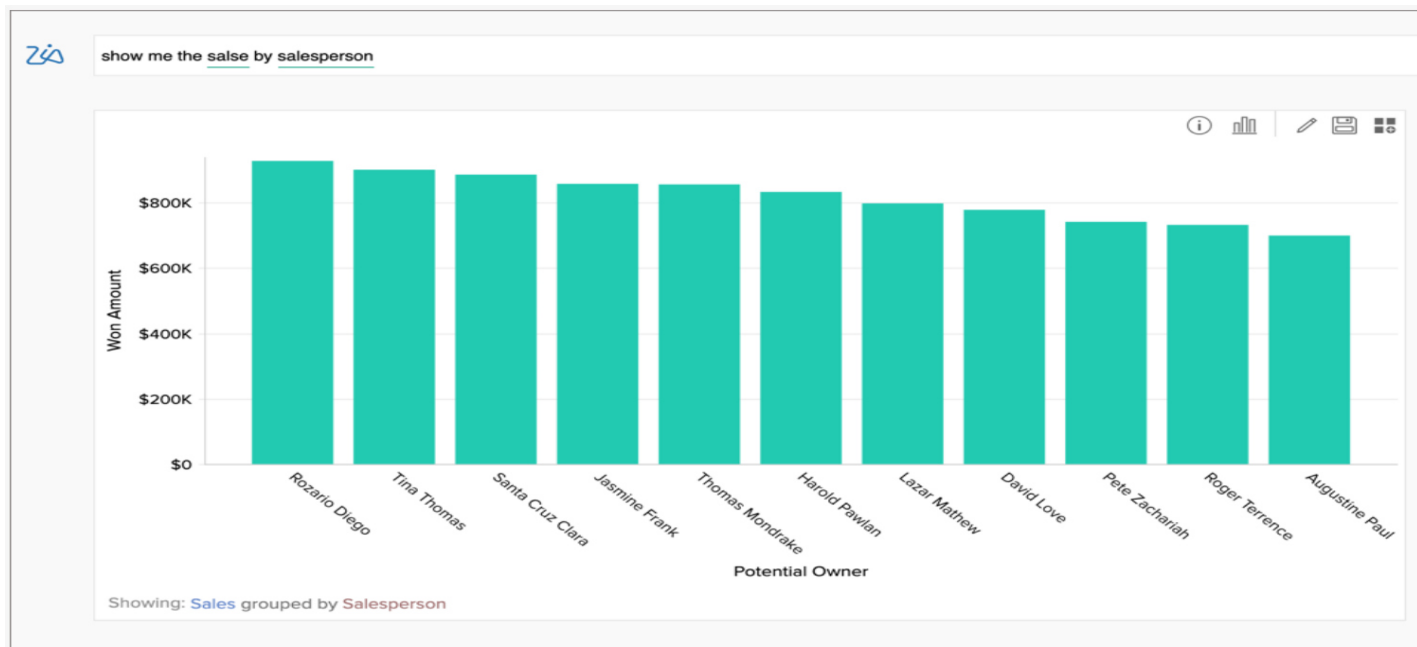
“sales vs target by region”



Typographical Error

Your question can contain a typographical error. Ask Zia can interpret questions with incomplete, incorrect words. In this case, the term 'salse' will be matched with sales

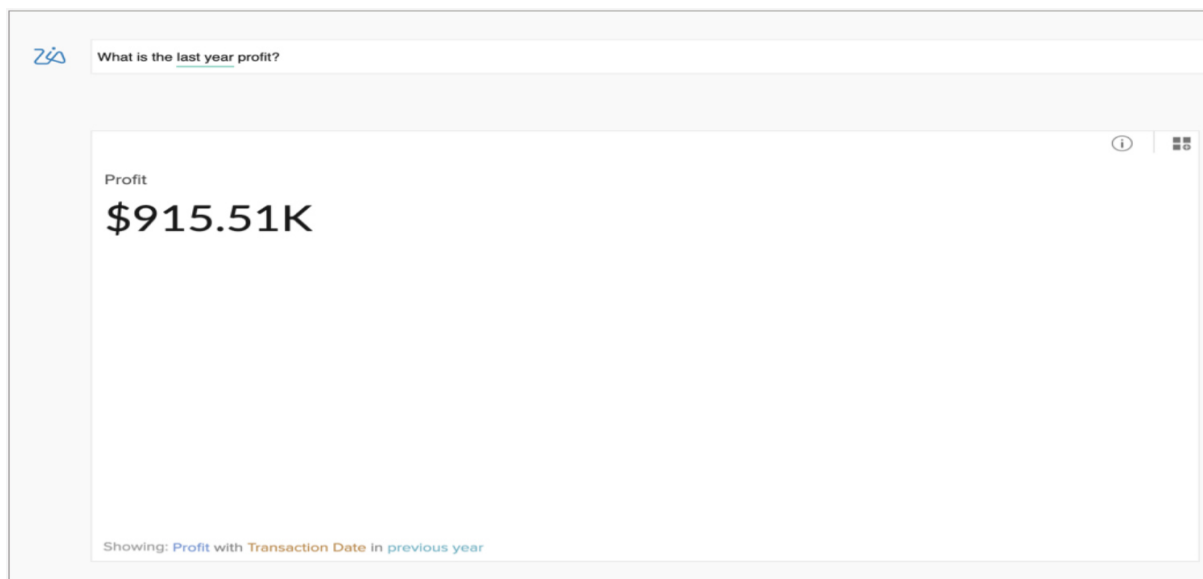
“Show me the salse by salesperson”



Single Metric

You can ask for a single metric in your question. The following question will get the *total profit* for last year as a Numeric KPI widget.

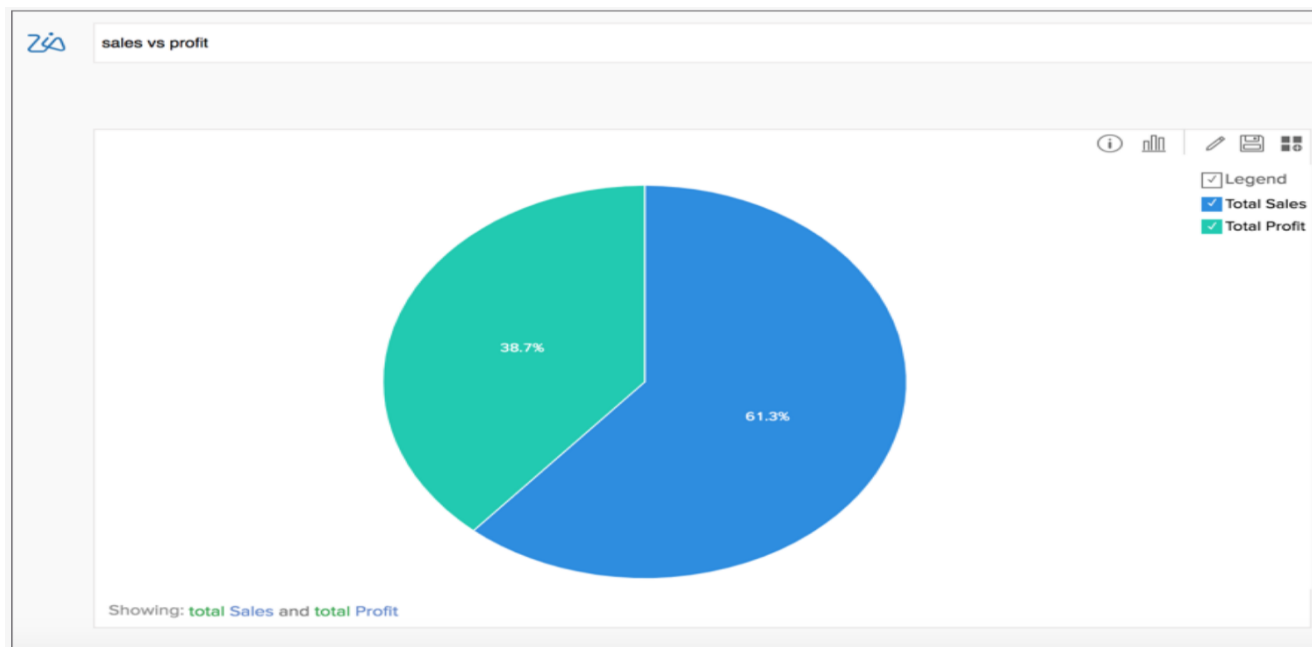
“What is the last year profit?”



Multiple Metrics

You can ask for multiple metrics in a question. The below question gets *Total Sales* and *Total Profit* as a pie chart.

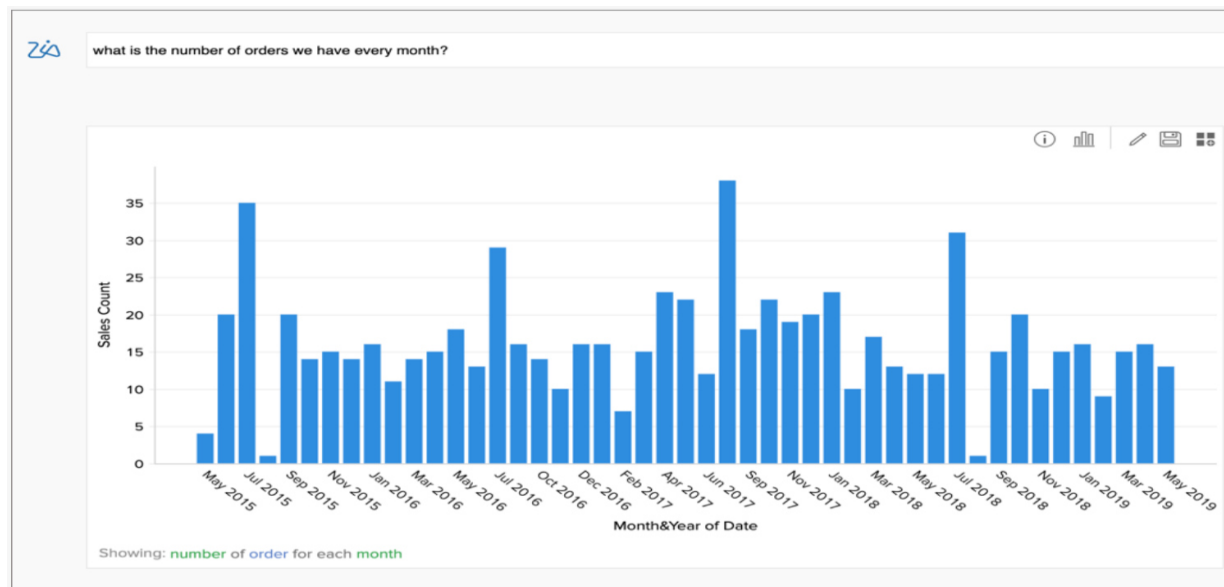
“Sales vs profit”



Summary Function

You can choose to apply the supported summary functions over a metric column by explicitly specifying them in your question. The below example applies count over the sales column and plots it across months.

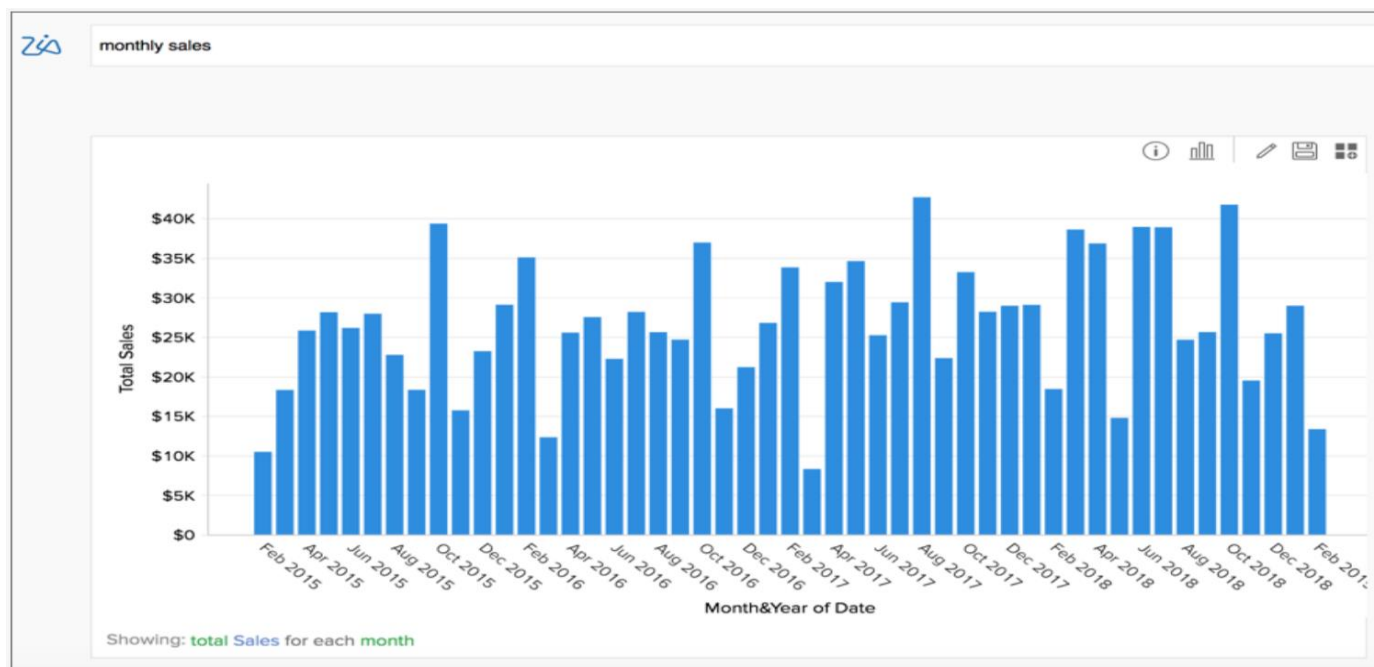
“what is the number of orders we have every month?”



Single Metric by a Single Dimension

In this example, the *Sales* column is grouped into months by the *Date* column.

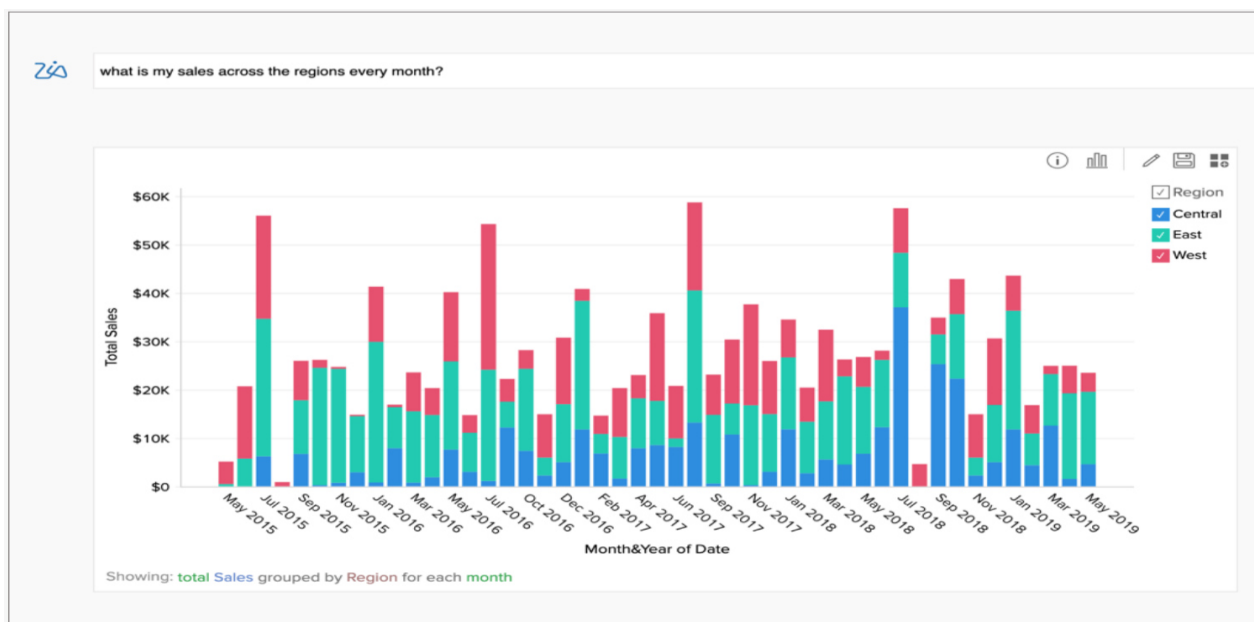
“monthly sales”



Single Metric by Multiple Dimensions

In this example, the *Sales* column is grouped by the *Date* and the *Region* columns.

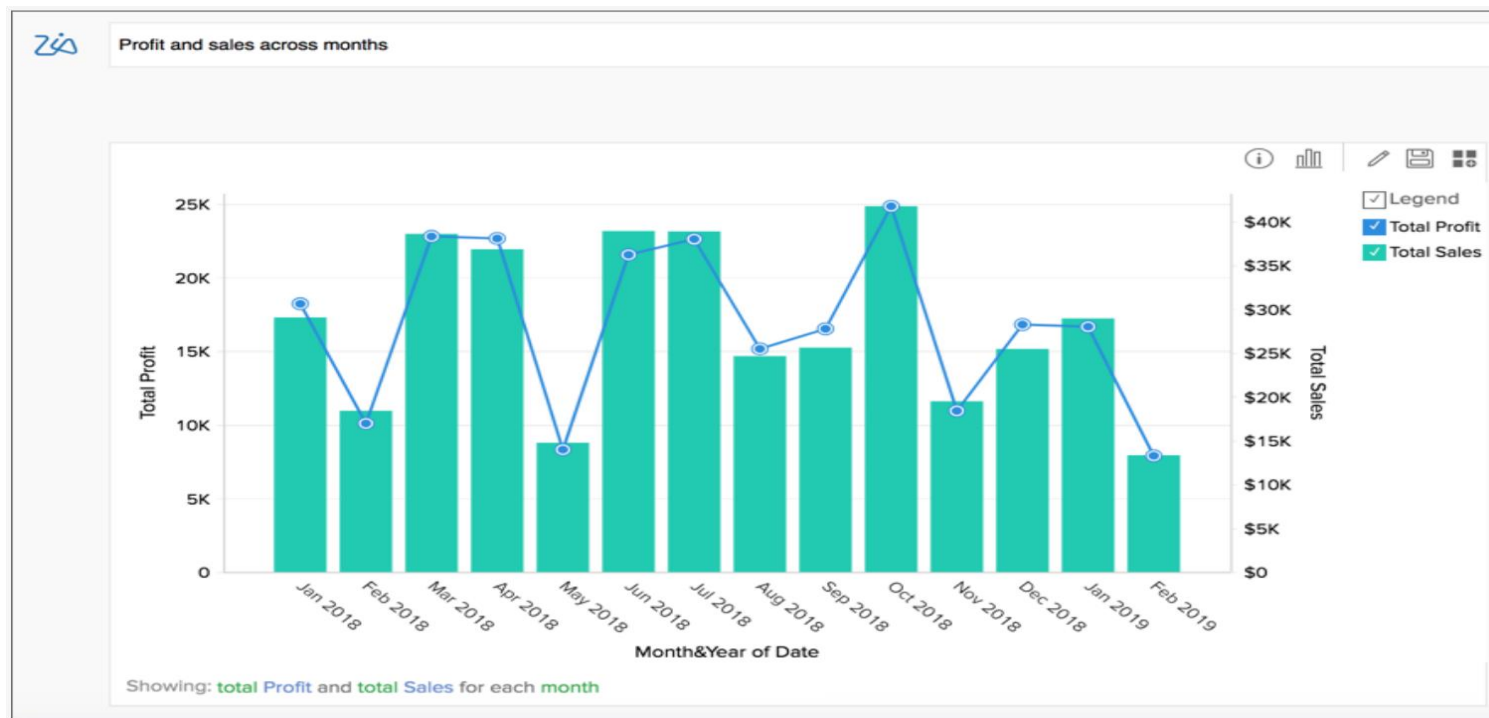
“what is my sales across the regions every month?”



Multiple Metrics by a Single Dimension

In this example, the *Sales* and the *Profit* columns are grouped by the *Date* column.

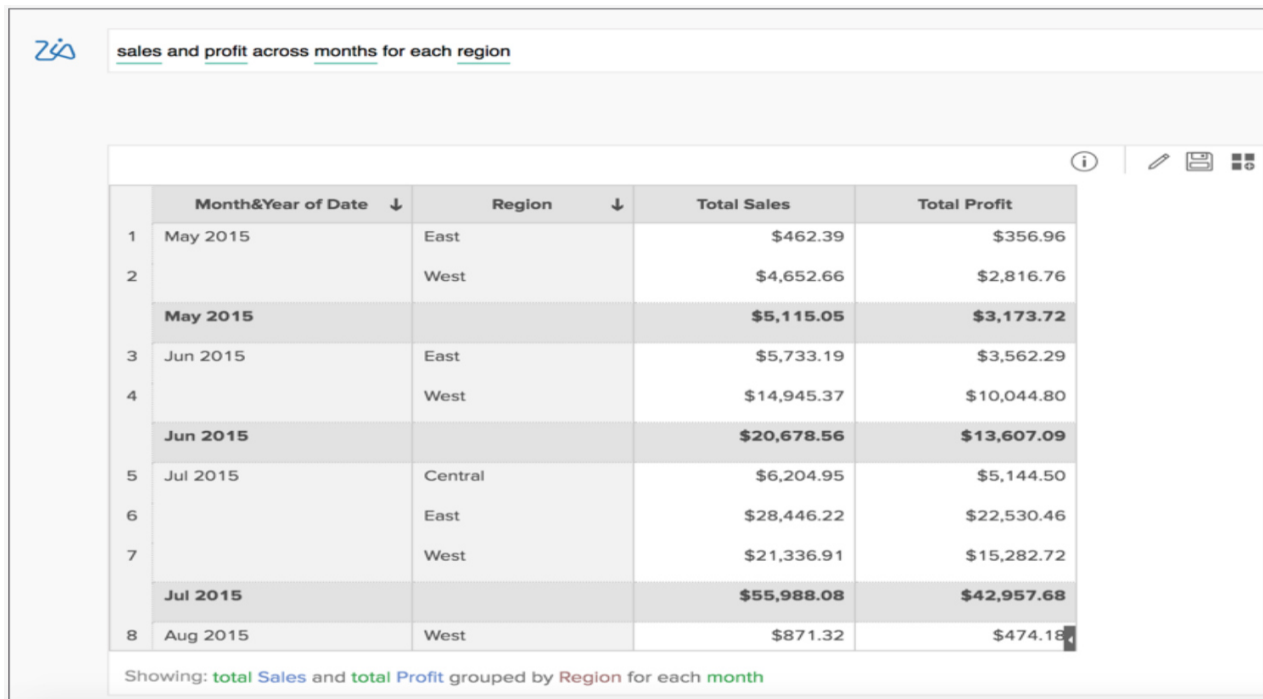
“Profit and sales across months”



Multiple Metrics by Multiple Dimension

In this example, the *Sales* and the *Profit* columns are grouped by the *Date* and the *Region* columns.

“Sales and profit across months for each region”



	Month&Year of Date ↓	Region ↓	Total Sales	Total Profit
1	May 2015	East	\$462.39	\$356.96
2		West	\$4,652.66	\$2,816.76
	May 2015		\$5,115.05	\$3,173.72
3	Jun 2015	East	\$5,733.19	\$3,562.29
4		West	\$14,945.37	\$10,044.80
	Jun 2015		\$20,678.56	\$13,607.09
5	Jul 2015	Central	\$6,204.95	\$5,144.50
6		East	\$28,446.22	\$22,530.46
7		West	\$21,336.91	\$15,282.72
	Jul 2015		\$55,988.08	\$42,957.68
8	Aug 2015	West	\$871.32	\$474.18

Showing: total Sales and total Profit grouped by Region for each month

Apply Filter Criteria

Plain Text

You can choose to include or exclude a value in a text column. In below example, we've filtered *Sales* where the *Region* is *West*.

“Show me the sales across months where the region is west”

Numeric

You can filter a numeric column by a range or by using special criteria such as **greater than 10**, **< 10**, **most**, **highest**, **lowest...** The below question filters the *Deal Size* greater than 12k.

“Salesperson by win rate and won amount where the deal size >12k”

Slice by Range

You can slice data by range using special criteria like **Top 10**, **Bottom 5**, **Highest**, **lowest**, **most**, **recent**. In this example, we have filtered top 10 salespersons.

“My top 10 salespersons by win rate”

Apply Filter Criteria

Exact Dates

You can filter data for a specific date such as 2019, March 2018, 12/3/2019... In this example, we've filtered the tickets escalated on 17th of March 2018.

“Escalated tickets on 17 march 2018”

Date ranges

You can filter data for a specific date range. In this example, the *Revenue* for a specific period i.e., from 3rd March 2018 to 3rd September 2018 is filtered.

“Show me the revenue by salespersons from 3rd Mar 2018 to 3rd Sept 2018”

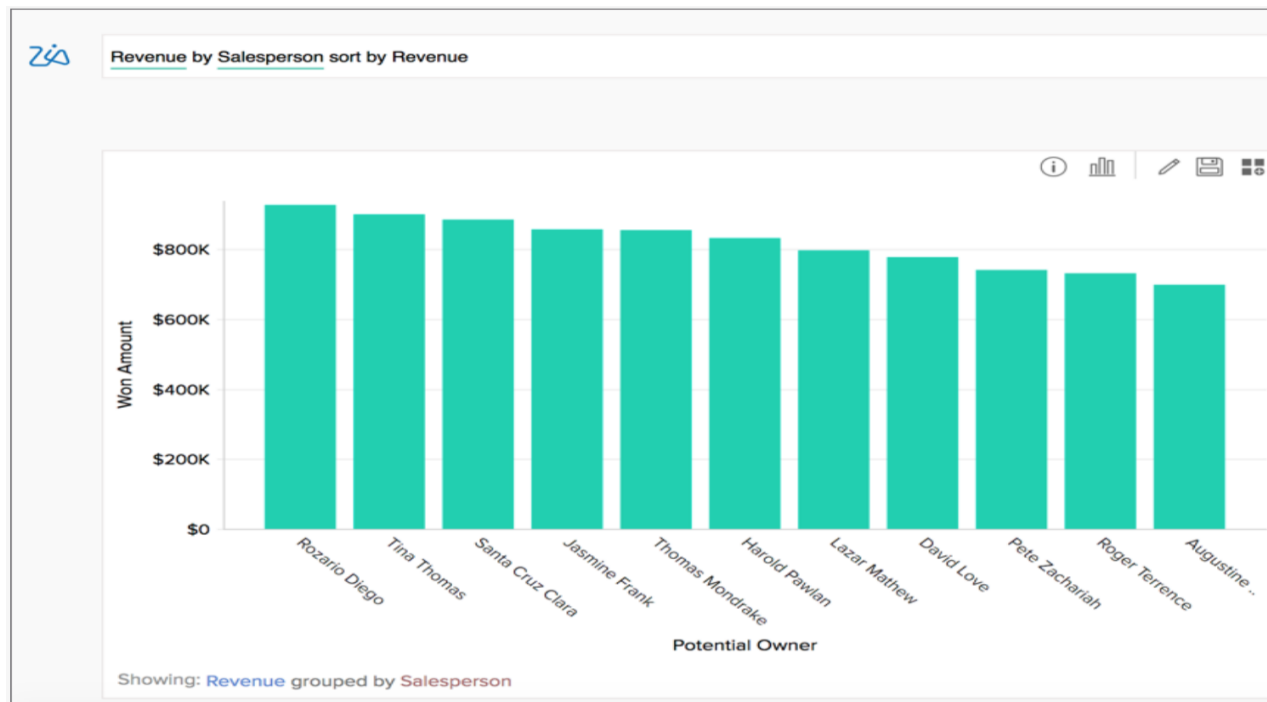
Relative Dates - You can filter data for a relative period such as *This Month, Last Quarter, Today...* In this example, the *Revenue* for the period of *last 1 year* is filtered.

“Show me the revenue across months for the last year”

Sort Data

You can choose to sort the report either by the metric column or by columns to group the metric. In this example, the report is sorted based on the Y-Axis column *Revenue*.

“Revenue by Salespersons sort by Revenue”



Keywords for Ask Zia

The following is the list of keywords that Ask Zia can interpret and create appropriate insights.

Summary Options	
Functions	Keywords
Summary	sum, total, tot
Minimum	minimum, min
Maximum	maximum, max, highest, high
Count	count, number, no, total number, distinct count, unique, unique count
Others	standard deviation, std dev, variance

Keywords for Ask Zia

Advanced Summary Options	
Functions	Keywords
Running Total	running total, running tot., running sum
Percentage of	percentage, total percentage, %, percent, percentage of total, total %, % of total, pct of total, % of sum, sum percentage, sum pct, contribution, contribution of total
Difference From	difference, difference from, diff, growth
% of Difference From	% of difference, pct of difference, difference %, % diff, difference percentage, difference in percentage,
% of Previous Value	percentage of previous, % of previous, pct of previous, prev %, previous value %

Keywords for Ask Zia

Group by Options	
Functions	Keywords
Numeric	actual, range, as range, dimension, as dimension
Date-Actual	<p>year, yearly, year over year, yoy, absyear</p> <p>quarter, quarterly, quarter over quarter, qoq, absquarter</p> <p>month, monthly, month over month, mom, month and year, absmonth</p> <p>week, weekly, week over week, wow, week and year, absweek</p> <p>day, daily, day over day, absday</p> <p>date, hour, hourly, minute, second</p>
Date-Seasonal	seasonal day, seasonal month, seasonal year, seasonal week, seasonal quarter

Keywords for Ask Zia

Filtering Options	
Functions	Keywords
Numeric	>, >=, greater than, above, over, beyond, higher than, more than, exceeds
	<, <=, lesser than, lower, lower than, less than, below, under
	equal to, equals to, equals, =, between
Units and Currency	thousands, k, millions, m, mn, billions, b, bn, Lakhs, L, Crores, C, Rupees, RS, ₹, Dollars, \$
By Range	top 10, top few, top, top most, highest, best
	bottom 10, bottom few, bottom, lowest, worst
Text	starts with, starts with, ends with, contains, is
Date-Actual	2018, 2019, Q1 2019, Q4 of 2018, quarter 1 of 2019
	January 2019, Jan 2019, 1/2019
	15th october , 30 july, nineth july 2017, third may 2016, 2 feb 2001, 2019 mar 4, 17/feb, twenty six october

Keywords for Ask Zia

<p>Date- Seasonal</p>	<p>June, may, jul, oct, Sunday, saturday, thurs, tue</p> <p>1st quarter, 2nd quarter, second quarter, third quarter, fourth quarter, first quarter, quarter 4, quarter1</p>
<p>Date-Relative</p>	<p>this year, previous year, last year, past year, recent year, previous 3 years, next 3 years, last 3 year, recent 3 years, last nine years, previous few years, recent few years, past few years, recent few decades, previous few decades, past few decades, next few decades, last few years, last decade, recent decade</p> <p>this quarter, previous quarter, last quarter, this quarter, recent quarter, previous 9 quarters, last 9 quarters, next 9 quarters, recent 5 quarters, previous five quarters, last three quarters, recent few quarters, next few quarters, previous few quarters, past few quarters, last few quarters</p> <p>this month, previous months, last months, past month, recent months, previous 6 months, next 6 months, last 6 months, recent 6 months, previous thirteen months, recent seventeen months, previous few months, recent few months, previous month, current may, previous november, this apr, past mar</p> <p>this week, last week, past week, recent week, last few weeks</p> <p>todat, yesterday, tomorrow, recent 3 days, past 21 days, recent few days, next few days, next day, recent day</p> <p>recent 22 hours, recent 13 hours, next hour, past hour</p>

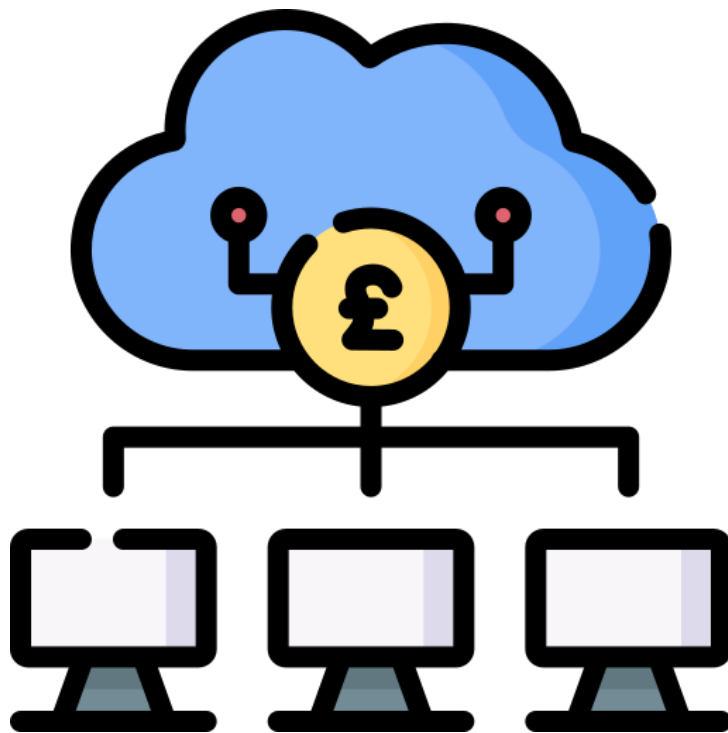
Keywords for Ask Zia

Sorting Options	
Functions	Keywords
Ascending	ascending order, ascending, increasing order, alphabetical order, alphabetical order (a-z)
Descending	descending, descending order, decreasing order, alphabetical order (z-a)
By Column	sorted, sort, sort by column

บทที่ 4 : แบบทดสอบและ ประเมินผลหลังเรียน



4.1 แบบทดสอบหลัง พัฒนาทักษะ (Post-Test)



ส่วนที่ 1 ลงทะเบียน

1. กรุณากรอกชื่อ-นามสกุล.....
2. สถานประกอบการ.....
3. Email.....
4. เบอร์โทร.....

ส่วนที่ 2 แบบทดสอบหลังพัฒนานักทักษะ (Post-Test)

- คำชี้แจง
1. แบบทดสอบฉบับนี้เป็นแบบปรนัยชนิดเลือกตอบ 4 ตัวเลือก จำนวน 16 ข้อ 16 คะแนน
 2. จงเลือกคำตอบที่ถูกต้องที่สุดเพียงข้อเดียว

1. ข้อมูลในข้อใดต่อไปนี้เป็นข้อมูลเชิงคุณภาพ (Qualitative Data)

- เพศ
- สัญชาติ
- ยอดขาย
- ความพึงพอใจ

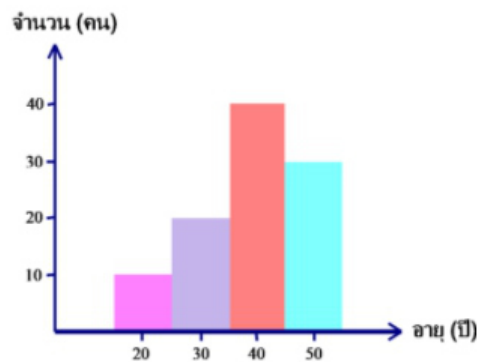
2. ข้อมูลในข้อใดต่อไปนี้เป็นข้อมูลเชิงปริมาณ (Quantitative Data)

- อุณหภูมิ
- ระดับการศึกษา
- ราคาสินค้า
- น้ำหนัก

3. ข้อใดไม่ใช่องค์ประกอบสำคัญของข้อมูลขนาดใหญ่ (Big Data)

- Volume (ข้อมูลที่มีปริมาณมาก)
- Variety (ข้อมูลที่มีหลากหลายรูปแบบ)
- Velocity (ข้อมูลที่มีการเปลี่ยนแปลงอย่างรวดเร็ว)
- Visitor (ข้อมูลจากผู้เข้าเยี่ยมชม)

4. จากกราฟแท่งต่อไปนี้ แสดงการกระจายของอายุประชาชนในหมู่บ้านแห่งหนึ่ง ประชาชนในหมู่บ้านแห่งนี้มีทั้งหมดกี่คน



- 100 คน
- 80 คน
- 70 คน
- 60 คน

5. จากการสำรวจอายุของเด็กจำนวน 100 คนพบว่า เด็กอายุ 4 ขวบ มีจำนวน 20 คน เด็กอายุ 5 ขวบ มีจำนวน 40 คน เด็กอายุ 6 ขวบ มีจำนวน 40 คน จงหาค่าเฉลี่ยของอายุของเด็กกลุ่มนี้

- 5.2 ขวบ
- 5.3 ขวบ
- 5.4 ขวบ
- 5.5 ขวบ

6. ชุดข้อมูลหรือกลุ่มข้อมูลใดที่เกี่ยวข้องกัน

- Datasets
- Data analytics
- Data Visualization
- Data Cleansing

7. ข้อใดเป็นการวิเคราะห์เชิงวินิจฉัยหาสาเหตุ โดยเปรียบเทียบข้อมูลในอดีตเทียบกับข้อมูลอื่นเพื่อหาคำตอบ

- Descriptive Analytics
- Predictive Analytics
- Diagnostic Analytics
- Prescriptive Analytics

8. ข้อใดคือคุณสมบัติของ Big data

- ปริมาณข้อมูลขนาดใหญ่
- ความหลากหลายของกลุ่มข้อมูล
- คุณภาพหรือความน่าเชื่อถือของข้อมูล
- ถูกทุกข้อ

9. แหล่งข้อมูลของBig data ที่ข้อมูลสร้างมาจากโซเชียลมีเดีย Blog posts อีเมล ภาพ ข้อความในออนไลน์ เป็นแหล่งข้อมูลจากข้อใด

- Machine Generated
- Software Generated
- Human-generated data
- Social Network Generated

10. ข้อมูลชนิดวิดีโอคลิป ที่เผยแพร่ หรือแชร์ทาง YouTube.com นั้นเป็นประเภทข้อมูลในข้อใด

- ข้อมูลที่ไม่มีโครงสร้าง
- ข้อมูลกึ่งมีโครงสร้าง
- ข้อมูลเชิงโครงสร้าง
- ข้อมูลโซเชี่ยลเน็ตเวิร์ค

11. กระบวนการ Big data ในข้อใดจัดเรียงตามลำดับได้ถูกต้อง

- การจัดเก็บรวบรวมข้อมูล, การประมวลผล, สร้างรายงาน
- การจัดเก็บรวบรวมข้อมูล, การประมวลผล, การเผยแพร่, สร้างรายงาน
- การประมวลผล, การจัดเก็บรวบรวมข้อมูล, การเผยแพร่
- การประมวลผล, การจัดเก็บรวบรวมข้อมูล, สร้างรายงาน, การเผยแพร่

12. ข้อใดต่อไปนี้ ไม่ใช่การทำ Data Preparation

- Data Visualization
- Data Cleansing
- Data Collection
- Find & Replace Data

13. ข้อใดไม่ใช่สาเหตุของการทำ Data Cleansing

- เพราะแหล่งข้อมูลมาจากหลายที่
- เพราะข้อมูลนำเข้ามามีจำนวนมาก
- เพราะข้อมูลที่นำเข้ามามีไม่มี standard
- เพราะผู้นำเข้าข้อมูลมาหลากหลาย

14. สำนักข่าวต้องการรวบรวมความคิดเห็นสาธารณะในเหตุการณ์หนึ่ง โดยที่มีการพิจารณาถึงกลยุทธิ์ที่แตกต่างกัน 2 กลยุทธิ์

ในการรวบรวมข้อมูลออนไลน์ โดย 2 กลยุทธิ์ดังกล่าว มีดังนี้

กลยุทธิ์ที่ 1:

- ใช้ฐานข้อมูลในการเก็บข้อมูลแบบสอบถามทั้งหมด
- เก็บข้อมูลเป็นข้อความ และ ตัวเลข
- ติดตามข้อมูลเพิ่มเติมเกี่ยวกับผู้รับการสำรวจที่จะไม่ปรากฏต่อสาธารณะ

กลยุทธิ์ที่ 2:

- ใช้ spreadsheet เดียวเพื่อจัดเก็บคำตอบแบบสำรวจทั้งหมด
- เก็บข้อมูลทั้งหมดเป็นตัวเลข
- ไม่มีการติดตามข้อมูลใด ๆ นอกเหนือจากแบบสอบถาม

ข้อความต่อไปนี้เป็นารเปรียบเทียบที่ถูกต้องที่สุดของกลยุทธิ์เหล่านี้

- กลยุทธิ์ที่หนึ่งจะช่วยให้จัดเรียงและกรองข้อมูลได้ง่ายขึ้นในขณะที่กลยุทธิ์ที่สองจะช่วยให้สร้างกราฟข้อมูลได้ง่ายขึ้น
- กลยุทธิ์ที่หนึ่งจะทำให้เกิดปัญหาเนื่องจากประเภทข้อมูลที่หลากหลายในขณะที่กลยุทธิ์ที่สองจะทำให้ง่ายต่อการค้นหาข้อมูลเฉพาะ
- กลยุทธิ์ที่หนึ่งจะช่วยให้หน่วยงานสามารถสรุปเกี่ยวกับความคิดเห็นของประชาชนได้มากขึ้นเนื่องจากติดตามข้อมูล metadata เพิ่มเติมในขณะที่กลยุทธิ์ที่สองจะทำให้ยากที่จะมองแนวโน้มและเข้าถึงข้อมูลบางส่วน
- กลยุทธิ์ที่หนึ่งลดเวลาสำหรับการเตรียมและจัดการข้อมูลให้น้อยลงในขณะที่กลยุทธิ์ที่สองต้องการการคำนวณเพิ่มเติมจำนวนมากเพื่อใช้ข้อมูล

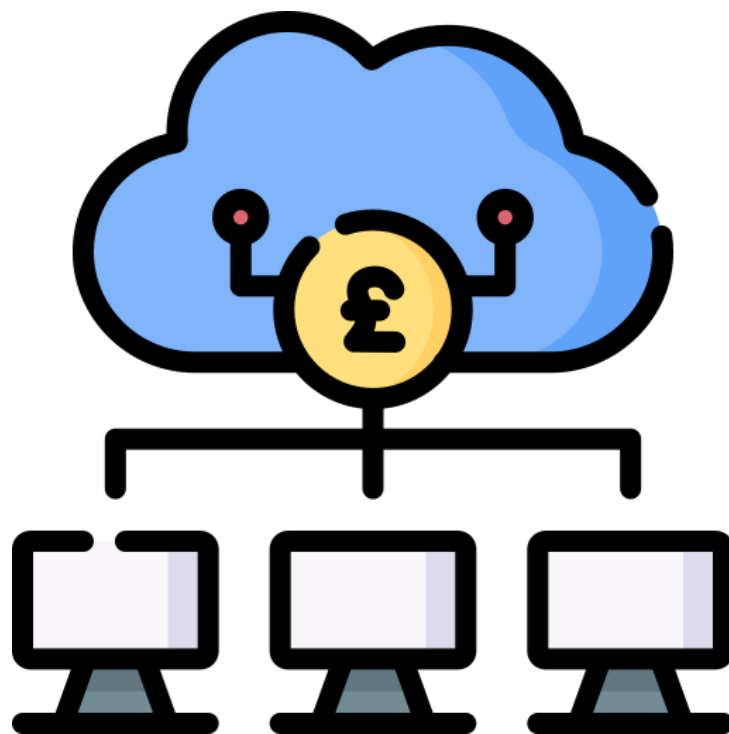
15. เจ้าของเว็บไซต์ต้องการวางแผนการแยกแหล่งที่มาของการเข้าชมเว็บไซต์ โดยมี 4 หมวดหมู่ ได้แก่ การเข้าชมโดยตรง (Direct Traffic), เครื่องมือค้นหา (Search Engines), ไซต์อ้างอิง (Referring Sites) และ อื่น ๆ (Other) แผนภูมิ (Chart) ประเภทใดที่จะเป็นทางเลือกที่ดีที่สุด

- Line Chart
- Dial Chart
- Pie Chart
- Bubble Chart

16. นักวิเคราะห์ข้อมูล สามารถทำนาย (Forecasting) ผลได้จากการอ่าน แผนภูมิ (Chart) ยกเว้น ข้อมูลจาก Chart ประเภทใด

- Line Chart
- Bar Chart
- Pie Chart
- Area Chart

4.2 แบบประเมินทักษะหลังการพัฒนาทักษะ (Post-Embedded Skill)



ส่วนที่ 1 สำหรับ ผู้เรียน

1.1 ข้อมูลทั่วไป

ชื่อ-นามสกุล

ชื่อสถานประกอบการ

1.2 เปรียบเทียบความรู้และทักษะที่ได้รับหลังเข้าร่วมพัฒนาทักษะ กับ พื้นฐานความรู้เดิม

() ได้พัฒนาทักษะใหม่ที่เพิ่มเติมและเป็นประโยชน์ มากกว่าความรู้เดิม () ไม่ได้รับการพัฒนาทักษะ

1.3 ความพึงพอใจต่อหลักสูตรพัฒนาทักษะ

() พึงพอใจมากที่สุด () พึงพอใจมาก () พึงพอใจปานกลาง () พึงพอใจน้อย

ส่วนที่ 2 สำหรับ เจ้าของกิจการ หรือ หัวหน้างาน

2.1 การประเมินผู้เรียน

ความหมายระดับคะแนน

- 0 = Beginner ไม่มีความรู้ ไม่มีทักษะ
- 1 = Learner มีความเข้าใจในทฤษฎีเบื้องต้น
- 2 = Practitioner มีความเข้าใจในทฤษฎีอย่างเต็มที่ มีความรู้ด้านปฏิบัติเล็กน้อย สามารถตอบคำถามหรือแก้ไขปัญหาก็ไม่ซับซ้อนได้
- 3 = Experienced มีความเข้าใจในทฤษฎีและปฏิบัติอย่างเต็มที่ สามารถประยุกต์ใช้ความรู้เพื่อแก้ไขปัญหาซับซ้อนปานกลางได้
- 4 = Embedded เกิดทักษะติดตัว สามารถเชื่อมโยงความรู้ในการแก้ไขปัญหาที่ซับซ้อนมากได้ และสามารถกำหนดแผนเพื่อปรับปรุงและพัฒนาประสิทธิภาพการทำงานในองค์กรได้และนำไปสู่การต่อยอดเพื่อลงมือทำจริง
- 5 = Broaden เกิดทักษะอย่างท่องแท้ในระดับผู้เชี่ยวชาญ และสามารถถ่ายทอดทักษะให้แก่ผู้อื่นได้

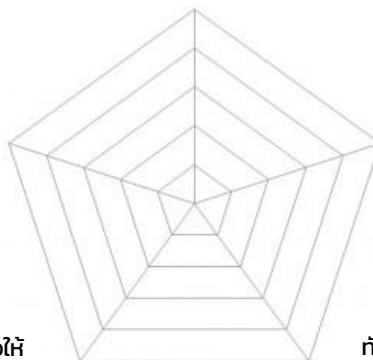
กรุณา (✓) ในช่องระดับคะแนน

ผลลัพธ์ทักษะ	ระดับคะแนน					
	0	1	2	3	4	5
1. ทักษะการกำหนดปัญหาในธุรกิจด้วย Big Data						
2. ทักษะการรวบรวมข้อมูลเพื่อประกอบการใช้เทคนิค Big Data						
3. ทักษะการวิเคราะห์ข้อมูลโดยใช้ Big Data เพื่อสร้างประสิทธิภาพให้กับธุรกิจ						
4. ทักษะการเลือกเครื่องมือและการนำเสนอให้สอดคล้องกับธุรกิจ						
5. ทักษะการประยุกต์ใช้ Big Data						

การวิเคราะห์ผลการพัฒนาทักษะด้วยกราฟเรดาร์ (Radar Chart)

ทักษะการกำหนดปัญหาในธุรกิจด้วย Big Data

ทักษะการประยุกต์ใช้ Big Data

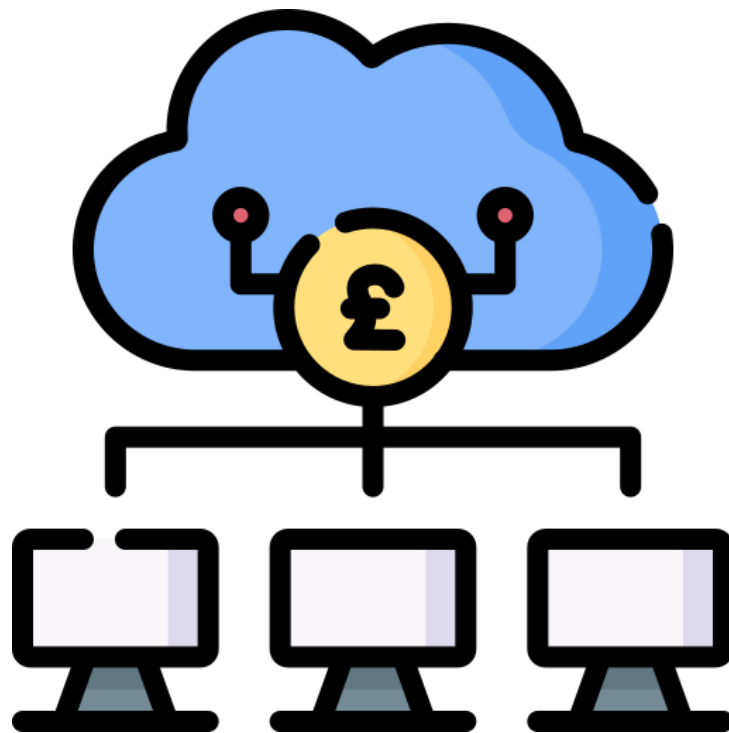


ทักษะการรวบรวมข้อมูลเพื่อประกอบการใช้เทคนิค Big Data

ทักษะการเลือกเครื่องมือและการนำเสนอให้สอดคล้องกับธุรกิจ

ทักษะการวิเคราะห์ข้อมูลโดยใช้ Big Data เพื่อสร้างประสิทธิภาพให้กับธุรกิจ

4.3 เฉลยแบบทดสอบ ก่อน-หลังพัฒนาทักษะ



ส่วนที่ 1 เจเลยแบบทดสอบพัฒนาทักษะ:

1. ข้อมูลในข้อใดต่อไปนี้เป็นข้อมูลเชิงคุณภาพ (Qualitative Data)

- เพศ
- สัญชาติ
- ยอดขาย
- ความพึงพอใจ

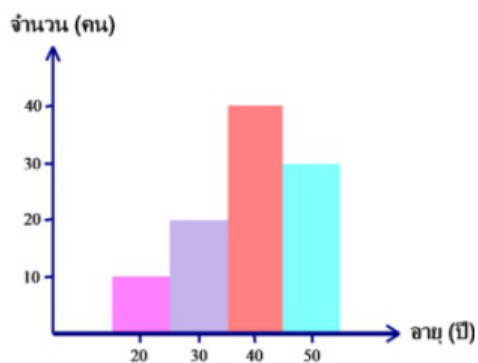
2. ข้อมูลในข้อใดต่อไปนี้เป็นข้อมูลเชิงปริมาณ (Quantitative Data)

- อุณหภูมิ
- ระดับการศึกษา
- ราคาสินค้า
- น้ำหนัก

3. ข้อใดไม่ใช่องค์ประกอบสำคัญของข้อมูลขนาดใหญ่ (Big Data)

- Volume (ข้อมูลที่มีปริมาณมาก)
- Variety (ข้อมูลที่มีหลากหลายรูปแบบ)
- Velocity (ข้อมูลที่มีการเปลี่ยนแปลงอย่างรวดเร็ว)
- Visitor (ข้อมูลจากผู้เยี่ยมชม)

4. จากกราฟแท่งต่อไปนี้ แสดงการกระจายของอายุประชาชนในหมู่บ้านแห่งหนึ่ง ประชาชนในหมู่บ้านแห่งนี้มีทั้งหมดกี่คน



- 100 คน
- 80 คน
- 70 คน
- 60 คน

5. จากการสำรวจอายุของเด็กจำนวน 100 คนพบว่า เด็กอายุ 4 ขวบ มีจำนวน 20 คน เด็กอายุ 5 ขวบ มีจำนวน 40 คน เด็กอายุ 6 ขวบ มีจำนวน 40 คน จงหาค่าเฉลี่ยของอายุของเด็กกลุ่มนี้

- 5.2 ขวบ
- 5.3 ขวบ
- 5.4 ขวบ
- 5.5 ขวบ

6. ชุดข้อมูลหรือกลุ่มข้อมูลใดที่เกี่ยวข้องกัน

- Datasets
- Data analytics
- Data Visualization
- Data Cleansing

7. ข้อใดเป็นการวิเคราะห์เชิงวินิจฉัยหาสาเหตุ โดยเปรียบเทียบข้อมูลในอดีตเทียบกับข้อมูลอื่นเพื่อหาคำตอบ

- Descriptive Analytics
- Predictive Analytics
- Diagnostic Analytics
- Prescriptive Analytics

8. ข้อใดคือคุณสมบัติของ Big data

- ปริมาณข้อมูลขนาดใหญ่
- ความหลากหลายของกลุ่มข้อมูล
- คุณภาพหรือความน่าเชื่อถือของข้อมูล
- ถูกทุกข้อ

9. แหล่งข้อมูลของ Big data ที่ข้อมูลสร้างมาจากโซเชียลมีเดีย Blog posts อีเมล ภาพ ข้อความในออนไลน์ เป็นแหล่งข้อมูลจากข้อใด

- Machine Generated
- Software Generated
- Human-generated data
- Social Network Generated

10. ข้อมูลชนิดวิดีโอคลิป ที่เผยแพร่ หรือแชร์ทาง YouTube.com นั้นเป็นประเภทข้อมูลในข้อใด

- ข้อมูลที่ไม่มีโครงสร้าง
- ข้อมูลกึ่งมีโครงสร้าง
- ข้อมูลเชิงโครงสร้าง
- ข้อมูลโซเชียลเน็ตเวิร์ค

11. กระบวนการ Big data ในข้อใดจัดเรียงตามลำดับได้ถูกต้อง

- การจัดเก็บรวบรวมข้อมูล, การประมวลผล, สร้างรายงาน
- การจัดเก็บรวบรวมข้อมูล, การประมวลผล, การเผยแพร่, สร้างรายงาน
- การประมวลผล, การจัดเก็บรวบรวมข้อมูล, การเผยแพร่
- การประมวลผล, การจัดเก็บรวบรวมข้อมูล, สร้างรายงาน, การเผยแพร่

12. ข้อใดต่อไปนี้ ไม่ใช่การทำ Data Preparation

- Data Visualization
- Data Cleansing
- Data Collection
- Find & Replace Data

13. ข้อใดไม่ใช่สาเหตุของการทำ Data Cleansing

- เพราะแหล่งข้อมูลมาจากหลายที่
- เพราะข้อมูลนำเข้ามามีจำนวนมาก
- เพราะข้อมูลที่นำเข้ามามีไม่มี standard
- เพราะผู้นำเข้าข้อมูลหลากหลาย

14. สำนักข่าวต้องการรวบรวมความคิดเห็นสาธารณะในเหตุการณ์หนึ่ง โดยที่มีการพิจารณาถึงกลยุทธิ์ที่แตกต่างกัน 2 กลยุทธิ์

ในการรวบรวมข้อมูลออนไลน์ โดย 2 กลยุทธิ์ดังกล่าว มีดังนี้

กลยุทธิ์ที่ 1:

- ใช้ฐานข้อมูลในการเก็บข้อมูลแบบสอบถามทั้งหมด
- เก็บข้อมูลเป็นข้อความ และ ตัวเลข
- ติดตามข้อมูลเพิ่มเติมเกี่ยวกับผู้รับการสำรวจที่จะไม่ปรากฏต่อสาธารณะ

กลยุทธิ์ที่ 2:

- ใช้ spreadsheet เดียวเพื่อจัดเก็บคำตอบแบบสำรวจทั้งหมด
- เก็บข้อมูลทั้งหมดเป็นตัวเลข
- ไม่มีการติดตามข้อมูลใด ๆ นอกเหนือจากแบบสอบถาม

ข้อความใดต่อไปนี้เป็นเปรียบเทียบที่ถูกต้องที่สุดของกลยุทธิ์เหล่านี้

- กลยุทธิ์ที่หนึ่งจะช่วยให้จัดเรียงและกรองข้อมูลได้ง่ายขึ้นในขณะที่กลยุทธิ์ที่สองจะช่วยให้สร้างกราฟข้อมูลได้ง่ายขึ้น
- กลยุทธิ์ที่หนึ่งจะทำให้เกิดปัญหาเนื่องจากประเภทข้อมูลที่หลากหลายในขณะที่กลยุทธิ์ที่สองจะทำได้ง่ายต่อการค้นหาข้อมูลเฉพาะ
- กลยุทธิ์ที่หนึ่งจะช่วยให้หน่วยงานสามารถสรุปเกี่ยวกับความคิดเห็นของประชาชนได้มากขึ้นเนื่องจากติดตามข้อมูล metadata เพิ่มเติมในขณะที่กลยุทธิ์ที่สองจะทำให้ยากที่จะมองแนวโน้มและเข้าถึงข้อมูลบางส่วน**
- กลยุทธิ์ที่หนึ่งลดเวลาสำหรับการเตรียมและจัดการข้อมูลให้น้อยลงในขณะที่กลยุทธิ์ที่สองต้องการการคำนวณเพิ่มเติมจำนวนมากเพื่อใช้ข้อมูล

15. เจ้าของเว็บไซต์ต้องการวางแผนการแยกแหล่งที่มาของการเข้าชมเว็บไซต์ โดยมี 4 หมวดหมู่ ได้แก่ การเข้าชมโดยตรง (Direct Traffic), เครื่องมือค้นหา (Search Engines), ไซต์อ้างอิง (Referring Sites) และ อื่น ๆ (Other) แผนภูมิ (Chart) ประเภทใดที่จะเป็นทางเลือกที่ดีที่สุด

- Line Chart
- Dial Chart
- Pie Chart**
- Bubble Chart

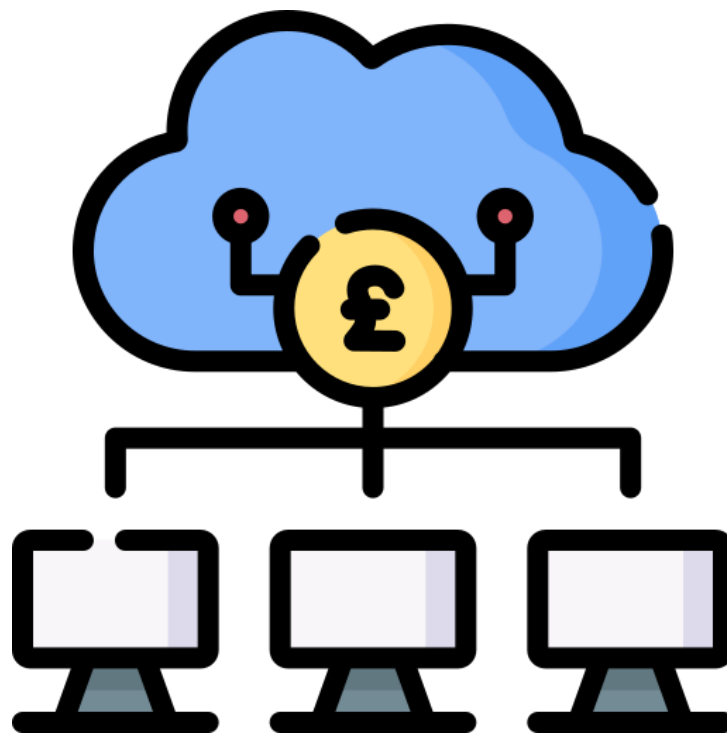
16. นักวิเคราะห์ข้อมูล สามารถทำนาย (Forecasting) ผลได้จากการอ่าน แผนภูมิ (Chart) ยกเว้น
ข้อมูลจาก Chart ประเภทใด

- Line Chart
- Bar Chart
- Pie Chart
- Area Chart

บทที่ 5 : แผนงาน (Action Plan)



5.1 แบบฟอร์มแผนงาน (Action Plan)



ส่วนที่ 1 สำหรับผู้เข้าร่วมพัฒนาทักษะ:

ชื่อ-นามสกุล.....ชื่อสถานประกอบการ.....

ชื่อแผนงาน / ความต้องการ.....

วัตถุประสงค์.....

ที่	เป้าหมาย/ความต้องการ/ ปัญหา	กลยุทธ์/แนวทางการแก้ไข	วิธีการดำเนินงาน (ระบุอย่างละเอียด)	ตัวชี้วัด	ระยะเวลา	ทรัพยากรที่มี
						งบประมาณ

ส่วนที่ 2 สำหรับหัวหน้างาน หรือ เจ้าของกิจการ

พิจารณาแผนที่ผู้เรียนนำเสนอ

.....

.....

แผนการต่อยอดหรือลงทุนจากทักษะที่ได้รับ

ที่	รายการ	พร้อมดำเนินงานทันที	มีแผนการดำเนินงานในอนาคต	โปรดอธิบายเพิ่มเติม	หมายเหตุ
1	ทำนสนใจลงทุนใน เครื่องจักร	<input type="checkbox"/>	<input type="checkbox"/>		
2	ทำนสนใจลงทุนใน กำลังคน เช่น มีการจ้างงานเพิ่มขึ้นเพื่อควบคุมเครื่องจักรที่ได้ลงทุนเพิ่ม	<input type="checkbox"/>	<input type="checkbox"/>		
3	ทำนสนใจต่อยอดและลงทุน ในด้านอื่น ๆ โปรดระบุ.....	<input type="checkbox"/>	<input type="checkbox"/>		

ความพึงพอใจต่อหลักสูตรพัฒนาทักษะ:

() พึงพอใจมากที่สุด () พึงพอใจมาก () พึงพอใจปานกลาง () พึงพอใจน้อย

KNOWLEDGE MANAGEMENT

หลักสูตรทักษะการ จัดการข้อมูลและการ วิเคราะห์ข้อมูล โดยใช้ Big Data

(Big Data Analytics Skill)

ภายใต้แผนงานพัฒนาความสามารถทางเทคโนโลยีของ
บุคลากรภาคอุตสาหกรรม
โครงการสร้างกำลังคนและทักษะแห่งอนาคตในภูมิภาคเพื่อ
ตอบโจทย์การพัฒนานวัตกรรมของประเทศ
ประจำปีงบประมาณ 2563

จัดทำโดย
อุทยานวิทยาศาสตร์ภาคเหนือ

