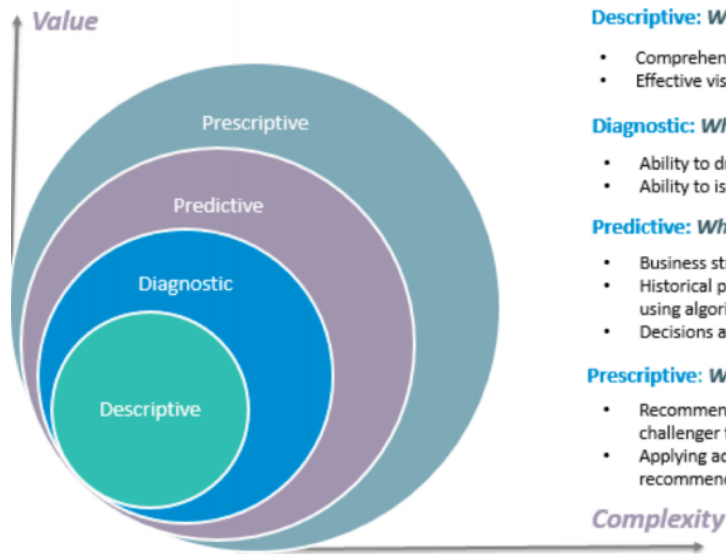


4 types of Data Analytics



What is the data telling you?

Descriptive: What's happening in my business?

- Comprehensive, accurate and live data
- Effective visualisation

Diagnostic: Why is it happening?

- Ability to drill down to the root-cause
- Ability to isolate all confounding information

Predictive: What's likely to happen?

- Business strategies have remained fairly consistent over time
- Historical patterns being used to predict specific outcomes using algorithms
- Decisions are automated using algorithms and technology

Prescriptive: What do I need to do?

- Recommended actions and strategies based on champion / challenger testing strategy outcomes
- Applying advanced analytical techniques to make specific recommendations

Discriptive เจริญพัฒนา - อะไรขายดี

Diagnostic พอรู้ว่าอะไรขายดี -ทำไมถึงขายดี เพราะอะไร เช่นไอติม ขายดีเพราะ หน้าร้อนหรือเปล่า

Predictive แล้วคาดว่าจะเกิดอะไรขึ้น

Prescriptive อะไรเราต้องทำอะไร ทำยังไงกับมัน

Discriptive เกิดอะไรขึ้น

Diagnostic เกิดขึ้นได้อย่างไร

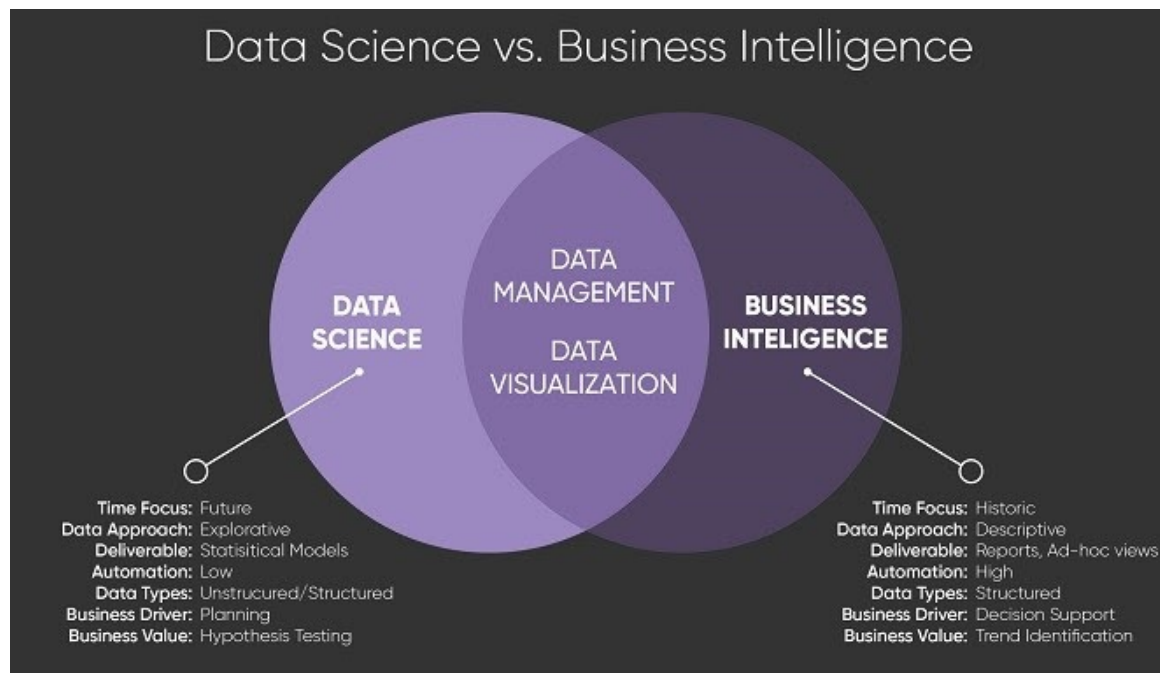
Predictive แล้วคาดว่าจะเกิดอะไรขึ้น

Prescriptive

EDA(Exploratory Data Analysis) = Data Understanding

ใส่ใจด้าน data source เพราะถ้าข้อมูลไม่เพียงพอต่อสิ่งที่เราสนใจมันจะมีปัญหา

Business Intelligence vs Data Science



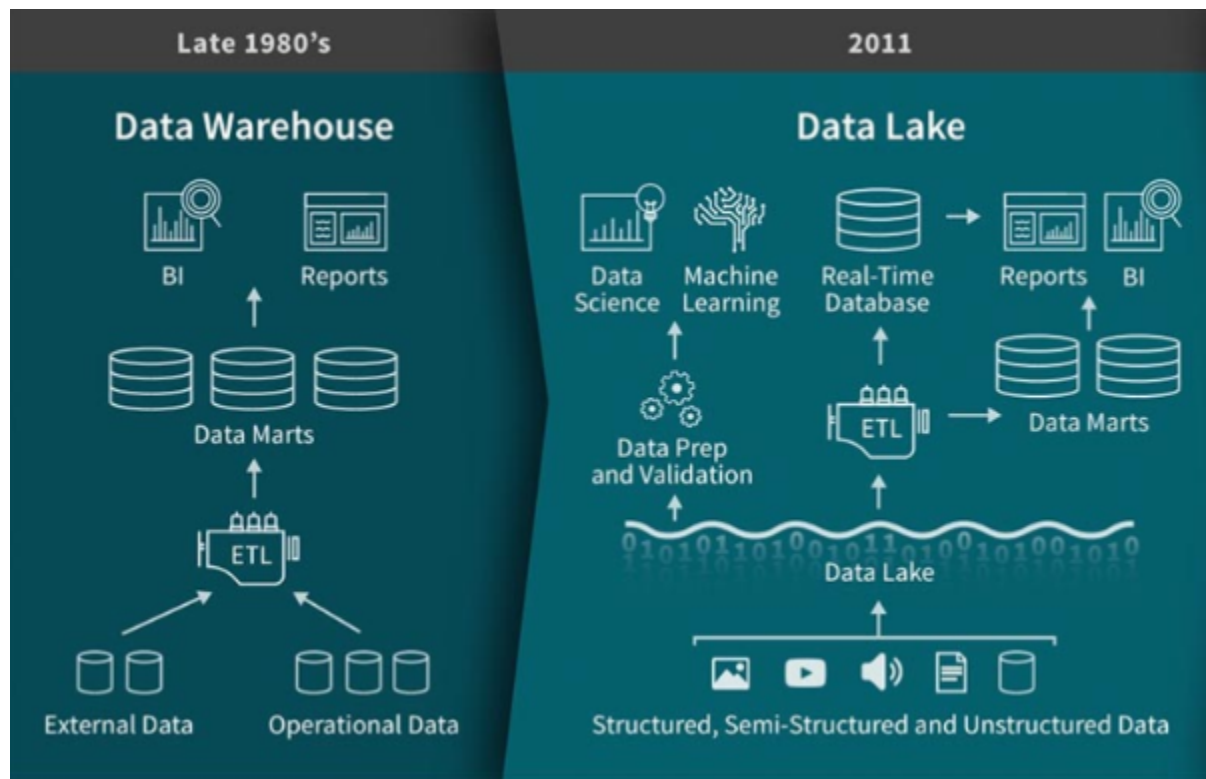
	BI	Data Science
	การใช้ข้อมูลมาช่วยในการขับเคลื่อนธุรกิจ	
	ข้อมูลเก่าเอามาสรุป แล้วนำมาวิเคราะห์ข้อมูลในแง่มุมต่างๆ	
Data Sources	แบบมีโครงสร้าง (Usually SQL, often Data Warehouse)	ข้อมูลต้องเป็นแบบมีโครงสร้างและ ไม่มีโครงสร้าง (logs, cloud data, SQL, NoSQL, text)
Approach	Statistics and Visualization	Statistics, Machine Learning, Graph Analysis, Neuro-linguistic Programming (NLP)
Focus	อดีตและปัจจุบัน	โฟกัสปัจจุบันในอนาคต
Tools	Pentaho, Microsoft BI, QlikView, R	RapidMiner, BigML, Weka, R

แบ่งเป็น 2 ส่วน

1. Transform Data to information
2. Transform Information to knowledge

BI คือ ฐานข้อมูลที่พร้อมนำไปใช้งานแล้ว

Data warehouse VS Data Lake



	Data warehouse (1980 s)	Data lake (2011)
ความหมาย	คือ ฐานข้อมูลที่พร้อมนำไปใช้งานแล้ว	
วัตถุประสงค์	สร้างข้อมูลเชิงกลยุทธ์เพื่อใช้ในการสร้างการตัดสินใจ	
ข้อมูลเป็นแบบ	มีโครงสร้าง	ไม่ได้สนใจ แบบไหนก็ได้ เป็นข้อมูลดิบ ทั้งแบบมีโครงสร้าง กึ่งโครงสร้าง และ ไม่มีโครงสร้าง
สอดคล้องกับ	วัตถุประสงค์ของธุรกิจ เช่น <ol style="list-style-type: none"> 1. กำไร 2. เพิ่มยอดขาย 3. ลดต้นทุน 	
สิ่งที่ต้องมี	Business metrics มาวัดความสำเร็จ มักจะเป็นตัวเลขหรือปริมาณ	ต้องมีที่เก็บข้อมูลที่ใหญ่
การเก็บข้อมูล	แบ่งตามหัวเรื่อง หัวเรื่อง ยอดขาย จำนวนเงิน,จำนวนชิ้น,จำนวนลูกค้า Factor → Business dimension	

	ETL Extract: ดึง Transform: Load: โหลด	ELT Extract: Load: โหลด Transform:
สิ่งที่ได้	Discriptive - อะไรขายดี Diagnostic พอรู้ว่าอะไรขายดี - ทำไมถึงขายดี เพราะอะไร เช่น ไอติม ขายดีเพราะ หน้าร้อนหรือเปล่า	Discriptive - อะไรขายดี Diagnostic พอรู้ว่าอะไรขายดี - ทำไมถึงขายดี เพราะอะไร เช่น ไอติม ขายดีเพราะ หน้าร้อนหรือเปล่า Predictive แล้วคาดว่าจะเกิดอะไรขึ้น Prescriptive อะไรเราต้องทำอะไร ทำยังไงกับมัน

OLTP

Online Transaction Processing หรือเรียกว่า Operational system

คือ ระบบฐานข้อมูลที่สนับสนุนแบบวันต่อวัน ข้อมูลเป็นแบบมีโครงสร้างเช่นกัน มักคู่กันกับ Data warehouse

Data Governance

Data Governance ไม่ใช่แค่เอกสารที่เป็นกฎระเบียบขององค์กรเท่านั้น แต่ต้องมีการปรับเปลี่ยนระบบ Infrastructure ใหม่เสียด้วย

Data Governance จะเป็นการปรับเปลี่ยนวิธีการทำงานขององค์กรทั้งหมด เพื่อรองรับการทำงานแบบ Data-Driven Business นั่นเอง

OSEM

Data mining เริ่มจากเรามีฐานข้อมูลอยู่แล้ว จึงค่อยเอามาหาประโยชน์

Data science ทำความเข้าใจธุรกิจ แล้วหา Data ที่อาจจะไม่มีหรือไม่มี เพื่อมาสร้างโมเดลที่เป็นประโยชน์ กรณีไม่มีข้อมูลก็อาจจะต้องสร้างวิธีการเก็บข้อมูล

การแบ่งประเภทของลูกค้า ใช้ RFM model

RFM model แบ่งเป็น 3 ประเภทคือ

- **Monetary (M)** ลูกค้าจ่ายเงินหนักหรือไม่หนัก เช่นนี้ถึง เอไอเอสเซอร์เเนด ถ้าจ่ายเงินให้จะได้สิทธิพิเศษเยอะ ยิ่งสูงยิ่งดี
- **Recency (R)** ล่าสุดเมื่อไหร่ เช่นเพิ่งซื้อของเมื่อวาน อาทิตย์ที่แล้วมันก็มีความแตกต่างกันนะ เช่นถ้าโลตัส ซื้อของเมื่อปีที่แล้ว อาจจะคิดว่าเขาไม่ได้เป็นลูกค้าเราแล้ว ยิ่งต่ำยิ่งดี แปลงเป็น 1-Recency ยิ่งสูงยิ่งดี เวลานั้นน้อยๆ ล่าสุดเมื่อวันที่แล้ว

- Frequency (F) บ่อยแค่ไหน ยิ่งสูงยิ่งดี

H = สูง

L = ต่ำ

1. M=H, F=H, R=H
2. M=H, F=H, R=L → สนใจว่าหายไปนานเท่าไรแล้วมีโอกาที่จะกู้คืนกลับมาเป็นลูกค้าอีกมั้ย
3. M=H, F=L, R=H → ซื้อหลายชิ้น แต่ละชิ้นไม่แพง หรือ ซื้อของแพงแต่น้อยชิ้น (Stocking)
4. M=L, F=H, R=H → เหมาะกับการระบายสินค้าที่ไม่แพง (warehouse management)
5. M=H, F=L, R=L → อาจเป็นขจร แบบมาซื้อของเราดูคนเดียวแล้วหายไปเลย
6. M=L, F=H, R=L → อดีตเคยมักกันอยู่แต่ตอนนี้หายไปแล้ว
7. M=L, F=L, R=H → อาจต้องปรับความบ่อยในการซื้อก่อนเพราะการจะปรับเรื่องจ่ายหนักน่าจะยาก
8. M=L, F=L, R=L

ค่าใช้จ่ายระหว่างหาลูกค้าใหม่กับการคงลูกค้าเดิมเป็นอย่างไร

สรุป หาลูกค้าใหม่แพงกว่า เช่นต้องมีโปรโมชั่นคุ้มๆ ขายของต้องถูกกว่าที่อื่น

Churn predictive

1. Trigger move ลูกค้ายกเลิกบริการ เช่นแบบก่อนยกเลิกมีการคอมเพลนกับทางร้านก่อนว่าไม่พอใจ ซึ่งแนวโน้มของลูกค้าเหล่านี้มีโอกาที่จะยกเลิกบริการ ต้องทำการตลาดเพื่อรักษา คงลูกค้าเหล่านี้ไว้
2. Silent move เป็นลูกค้าที่ ยกเลิกบริการแบบ ไปแบบเงียบๆ ดังนั้นต้องมีการทำการศึกษ ทำนายว่าทำไมถึงยกเลิกบริการเรา

การตีความ 'ลูกค้าที่หายไป' มีความแตกต่างเล็กน้อยที่ต้องนำมาพิจารณา:

- 'Absolute' vs. 'Presumed' churn:
- เวลา

- เช่นมีการกำหนดระยะเวลา เช่น ตามร้านทรู ที่มีติดสัญญา 6 - 16 เดือน แล้วหลังจากช่วงนั้นจะเริ่มมีการทำการโทรมาเพื่อต้องการให้ลูกค้ามีการต่อสัญญา
- เช่น จะเห็นได้ว่าการยกเลิกสัญญาต่างๆจะเห็นได้ว่า ไม่สามารถยกเลิกการบริการทางคอลเซ็นเตอร์ได้ ต้องไปที่สาขาเพื่อลดความสะดวสบายในการจะยกเลิกบริการกับทางทรู
- 'Reactive' vs. 'Prospective' churn:
 - อาจเกิดปัญหาต่างๆ เช่น สำหรับธนาคาร ถ้าเราไม่ได้ทำการอัปเดตบัญชีเป็นเวลานาน อาจเกิดเหตุการณ์ที่นายธนาคารอาจทำการโกงเราได้ เช่นมีการทำการเอาเงินเราออกจากธนาคารไปหมุนก่อน เพราะเขาสามารถทำการลบ transaction ได้
 - มี 2 กรณี
 - 'Prospective' trigger จะมีสัญญาณบอกมาก่อนว่าผู้ใช้นั้นจะยกเลิกเช่นการโทรมารายงานปัญหาต่างๆที่เคยเจอ
 - 'Silent' churn จู่ๆก็หายไปแบบไม่ได้บอกสาเหตุ

CHURN Example :

Telecom churn

กรณีลูกค้า ทรู ที่มีติดสัญญา 6 - 16 เดือน แล้วหลังจากช่วงนั้นจะเริ่มมีการทำการโทรมาเพื่อต้องการให้ลูกค้ามีการต่อสัญญา

Telecom : <https://www.kaggle.com/blastchar/telco-customer-churn>

Telecom : <https://www.kaggle.com/barun2104/telecom-churn>

Telecom : <https://www.kaggle.com/sourabhsabharwal/mobile-company-customers-retention>

Network churn

กรณี

Network provider : <https://www.kaggle.com/lokeshkum/network-provider-customer-churn-data>

Bank churn

กรณีธนาคาร

Bank : <https://www.kaggle.com/shrutimechlearn/churn-modelling>

Employee churn

กรณีการลาออกจากงาน เพราะแบบบางบริษัทรับคนเข้ามาพอมาทำการฝึกฝน เสร็จก็ออกจากงาน ทำให้บริษัทตระหนักในเรื่องนี้เพราะต้องมีการเสียค่าใช้จ่ายในการฝึกพนักงานให้มีความสามารถที่มากขึ้น บริษัทจำเป็นต้องมีการดึงพนักงานที่เก่งๆไว้

Employee : <https://www.kaggle.com/HRAntalyticRepository/employee-attribution-data>

Employee : <https://www.kaggle.com/quynquyentm/employeechurn>

Employee : <https://www.kaggle.com/rwzhang/retain-rate>

Sell churn

กรณี ที่ลูกค้ามาซื้อของเราแล้วต้องการให้ซื้อต่อไป ไม่หยุดซื้อ ไม่หยุดใช้บริการ คือเขาอาจเจอสิ่งที่ยอมรับไม่ได้ในสินค้าหรือบริการของเรา หรือเขาอาจเจอสิ่งที่ดีกว่า

Sell : <https://www.kaggle.com/hassanamin/customer-churn>

Sell : <https://www.kaggle.com/huzaiftila/customer-churn-prediction-analysis>

Newspaper churn

กรณี หนังสือพิมพ์ บล็อกความรู้ เช่น medium

Newspaper churn : <https://www.kaggle.com/leiviting01/newspaper-churn?>

Club churn

กรณี ที่นักได้เช่น ฟิตเนสที่มีการสมัครคอร์สทุกๆเดือน

Club : <https://www.kaggle.com/sonannguyenngoc/club-data-set>

Employee Churn พนักงานจะออกจากองค์กรไหม

1. Personal
 - 1.1. Salary
 - 1.2. Department
2. Business
 - 2.1. Satisfaction_level ถ้าค่าออกมาน้อย อาจหมายความว่าคนนั้นอาจจะไปจากธุรกิจของเรา
 - 2.2. Work accident ทำงานผิดพลาดหรือไม่ หรืออาจใช้คำว่า Complain/Notice
 - 2.3. Time_spend_company เวลาที่ใช้ในการเดินทาง
 - 2.4. Average_monthly_hours เวลาที่ต้องทำงานในแต่ละเดือน
 - 2.5. Number_project เขาได้รับงานเยอะไหม
 - 2.6. Last_evaluation

Bank churn

- กรณีธนาคารสามารถดูข้อมูลในเรื่องการตัดเงินตามบัญชี เช่น ค่าไฟ ทางธนาคารสามารถดูได้เลยว่า เจ้าของบัญชีนี้โอกาสที่จะปิดบัญชีนี้น้อยแค่ไหน

