

Hands-on Data Science and Machine Learning

1

Introduction to Data Science and Analytic Thinking

Dr. Parisut Jitpakdee

ดร.ปริสุทธิ์ จิตต์ภักดี

Senior Data Scientist and Project Manager

ผู้เชี่ยวชาญงานพิเศษด้านข้อมูลขนาดใหญ่

สถาบันข้อมูลขนาดใหญ่ (องค์การมหาชน)

**BIG DATA
INSTITUTE**

Other Positions:

- Technical Committee on Government Data Management Standards
คณะกรรมการเทคนิคด้านมาตรฐานการบริหารจัดการข้อมูลภาครัฐ

• Education



- Ph.D., Engineering and Technology (Computer Science)
Sirindhorn International Institute of Technology (SIIT), Thammasat University



- M.Sc., Information Technology (Information Science)
King Mongkut's Institute of Technology Ladkrabang



- B.Sc., Computer Science
Chiang Mai University



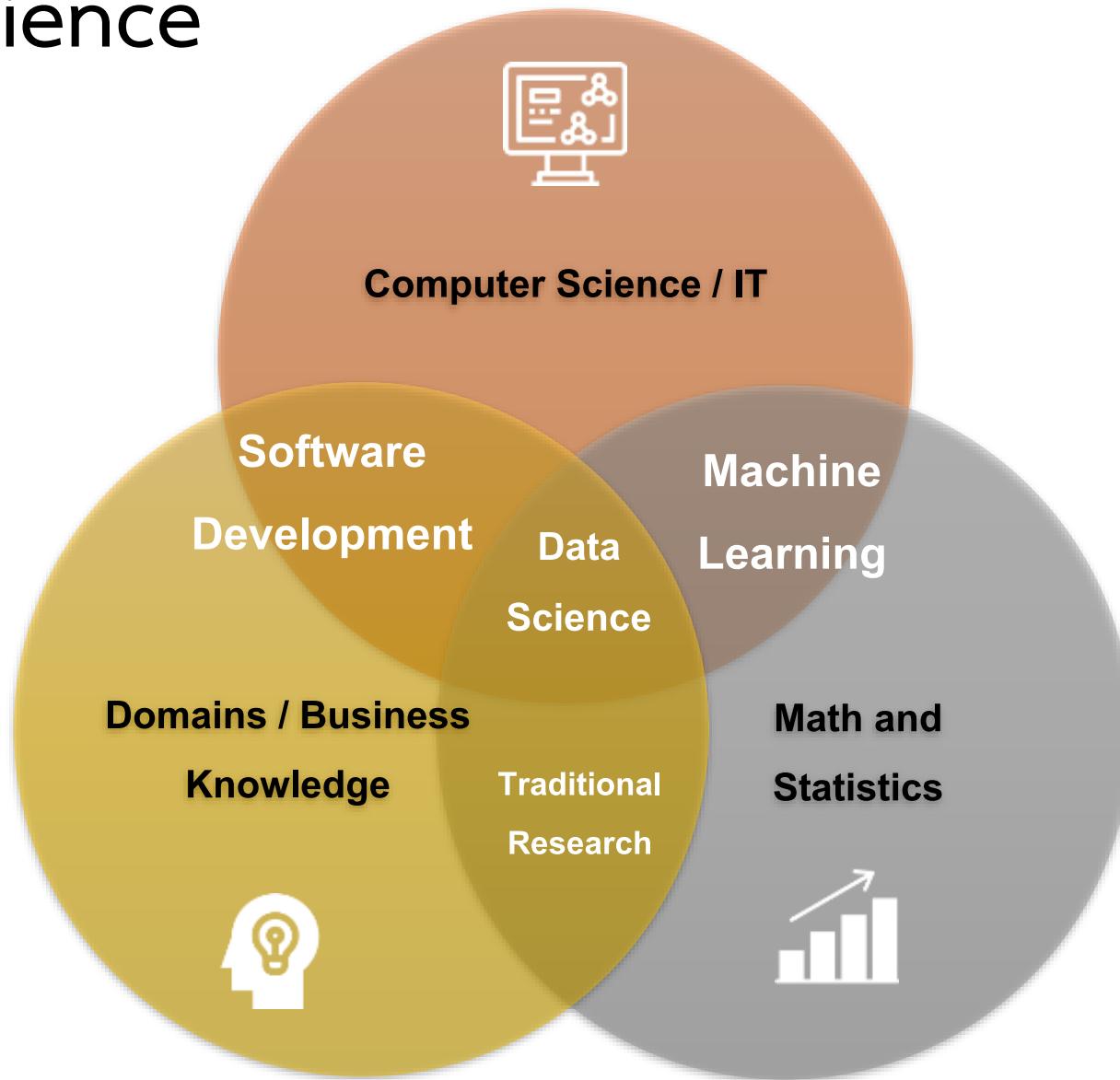
Introduction to Course: Course Outline

1. Introduction to data science and analytic thinking / Project inception
2. Introduction to statistics
3. Data science with basic Python programming
4. Exploratory data analysis
5. Basic machine learning

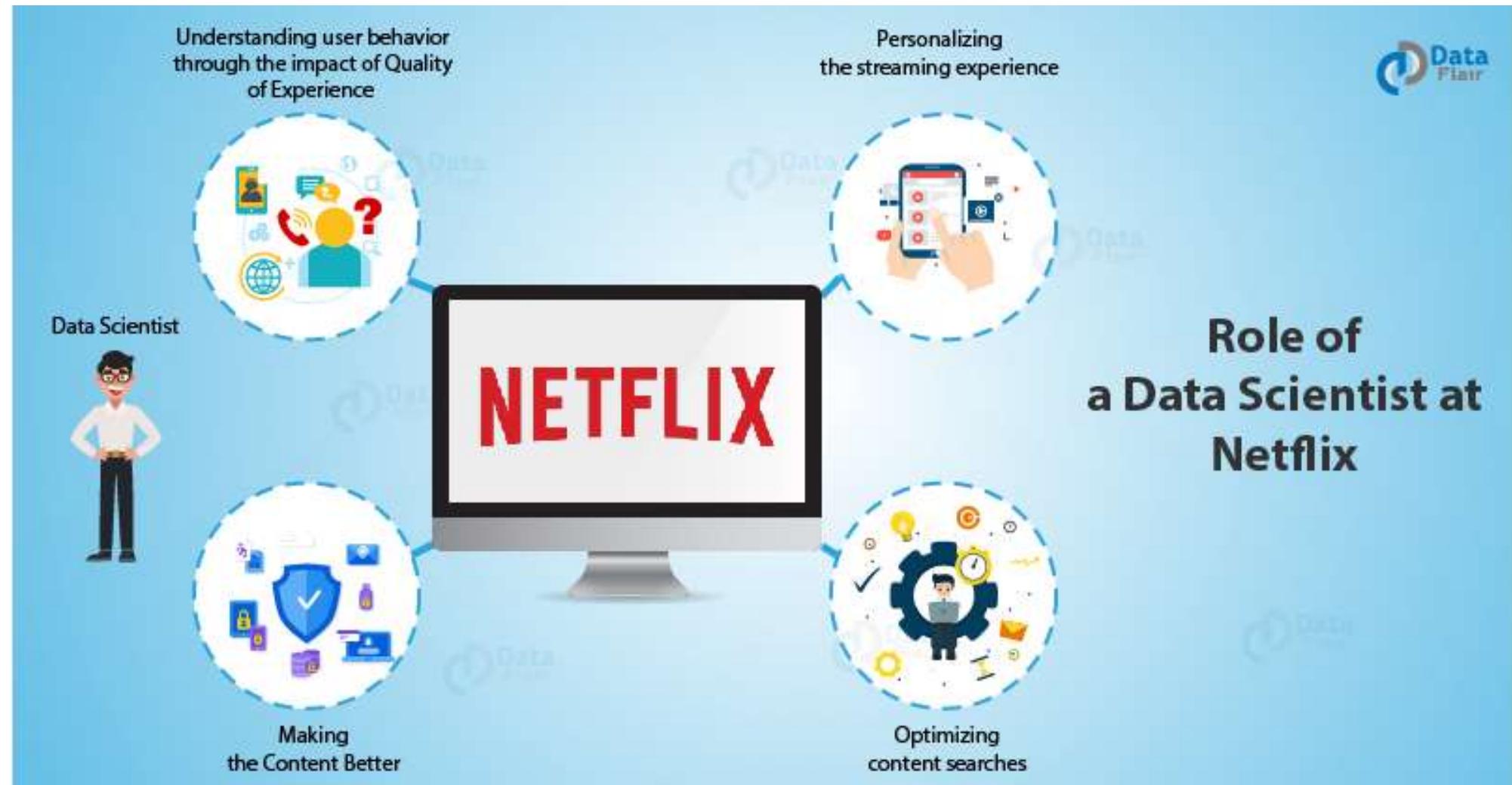
Introduction to Data Science

- What is Data Science?
- What is Big Data?
- Data Science Process
- Application of Data Science
- Big Data Analytics
- Example

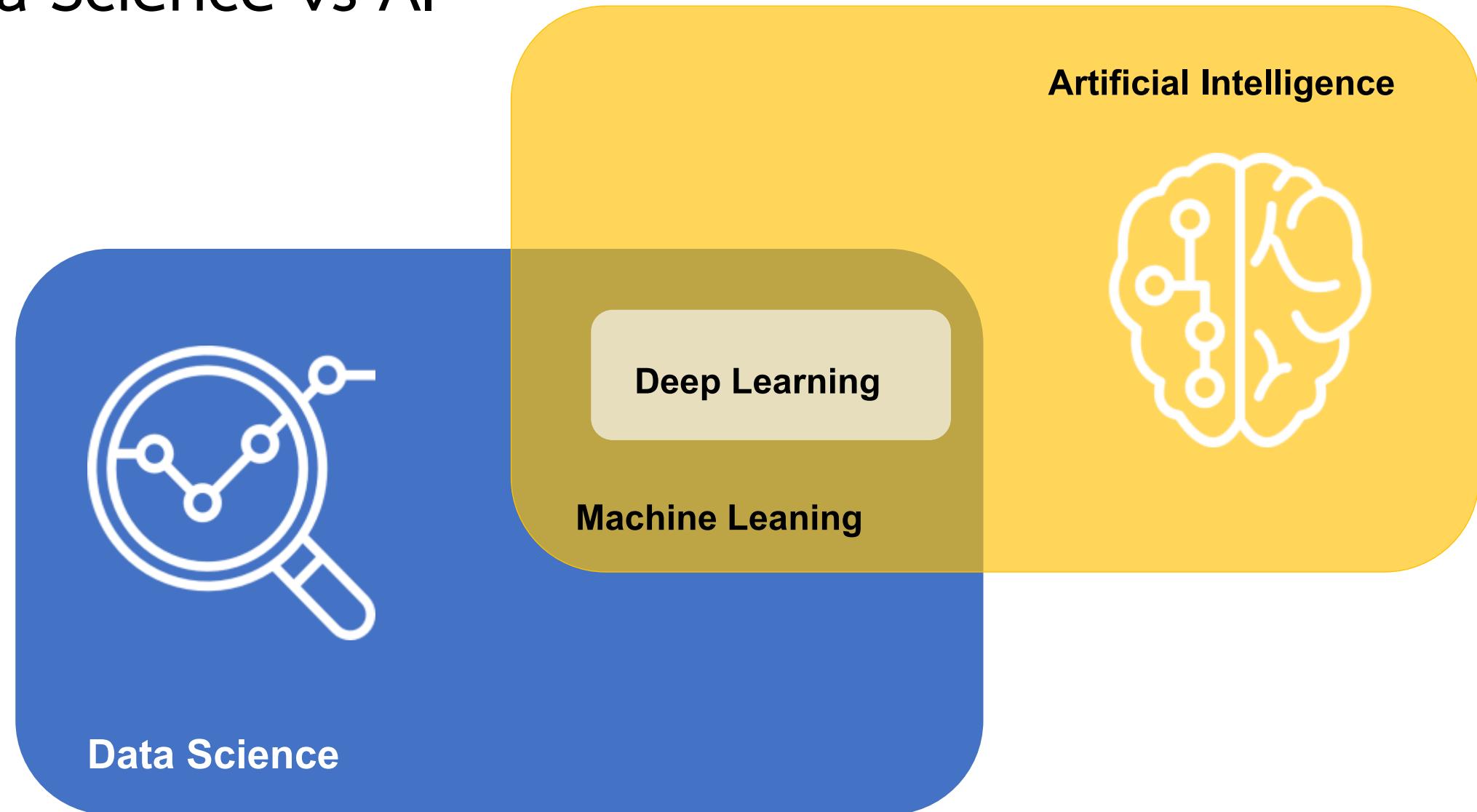
What is Data Science



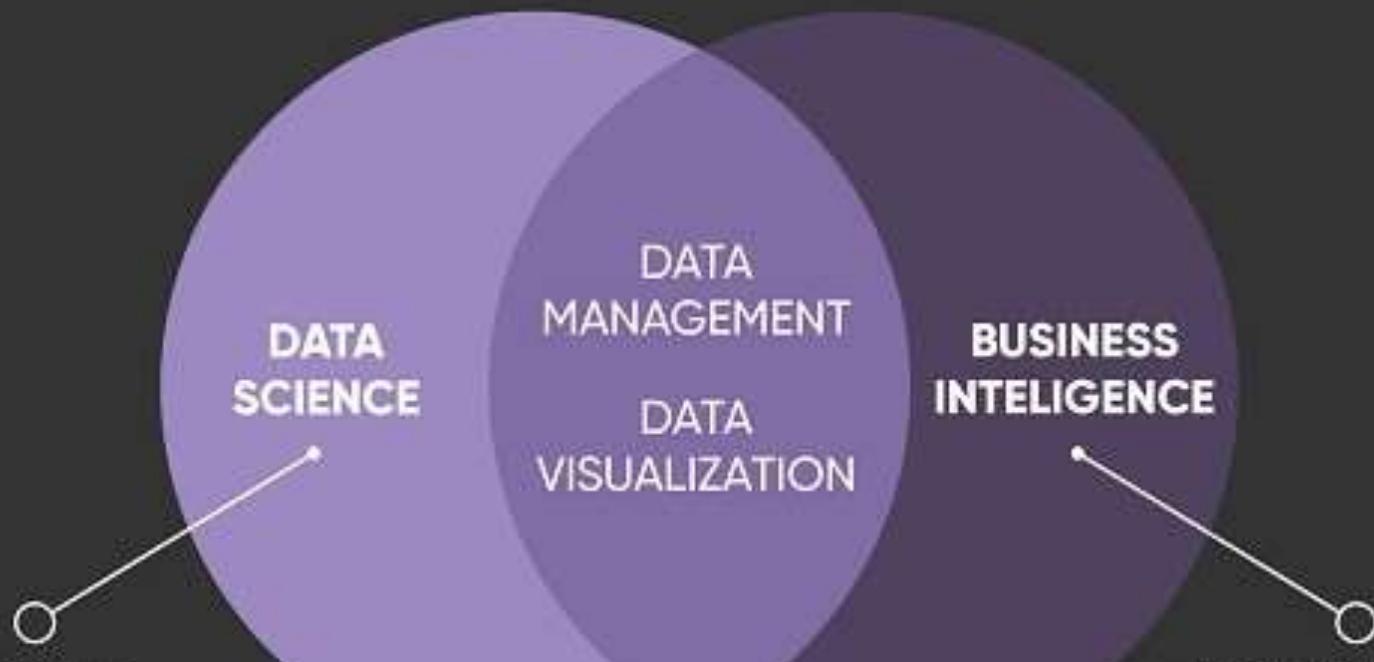
Case Study: Data Science at Netflix



Data Science vs AI



Data Science vs. Business Intelligence



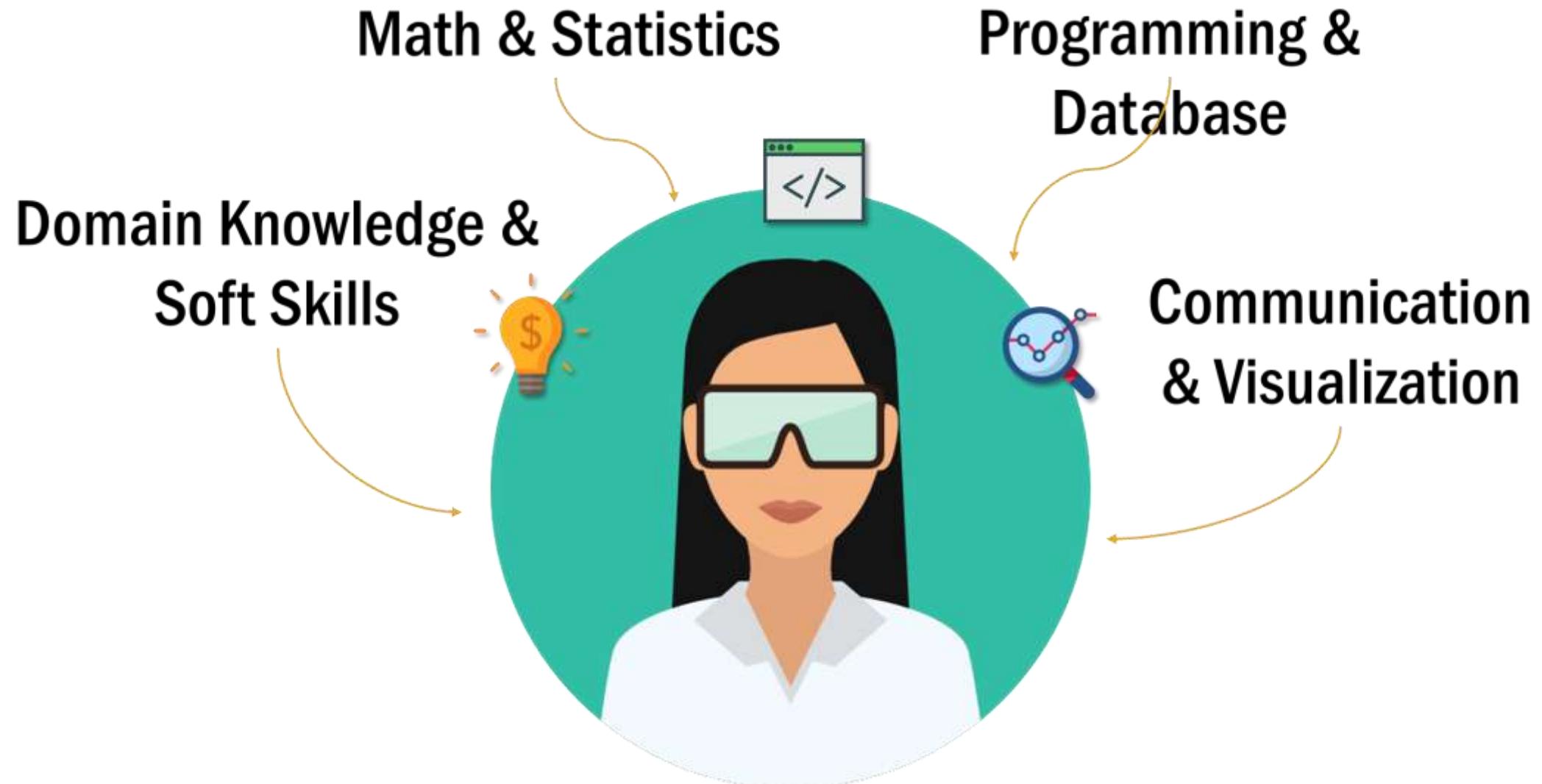
Time Focus: Future
Data Approach: Explorative
Deliverable: Statistical Models
Automation: Low
Data Types: Unstructured/Structured
Business Driver: Planning
Business Value: Hypothesis Testing

Time Focus: Historic
Data Approach: Descriptive
Deliverable: Reports, Ad-hoc views
Automation: High
Data Types: Structured
Business Driver: Decision Support
Business Value: Trend Identification

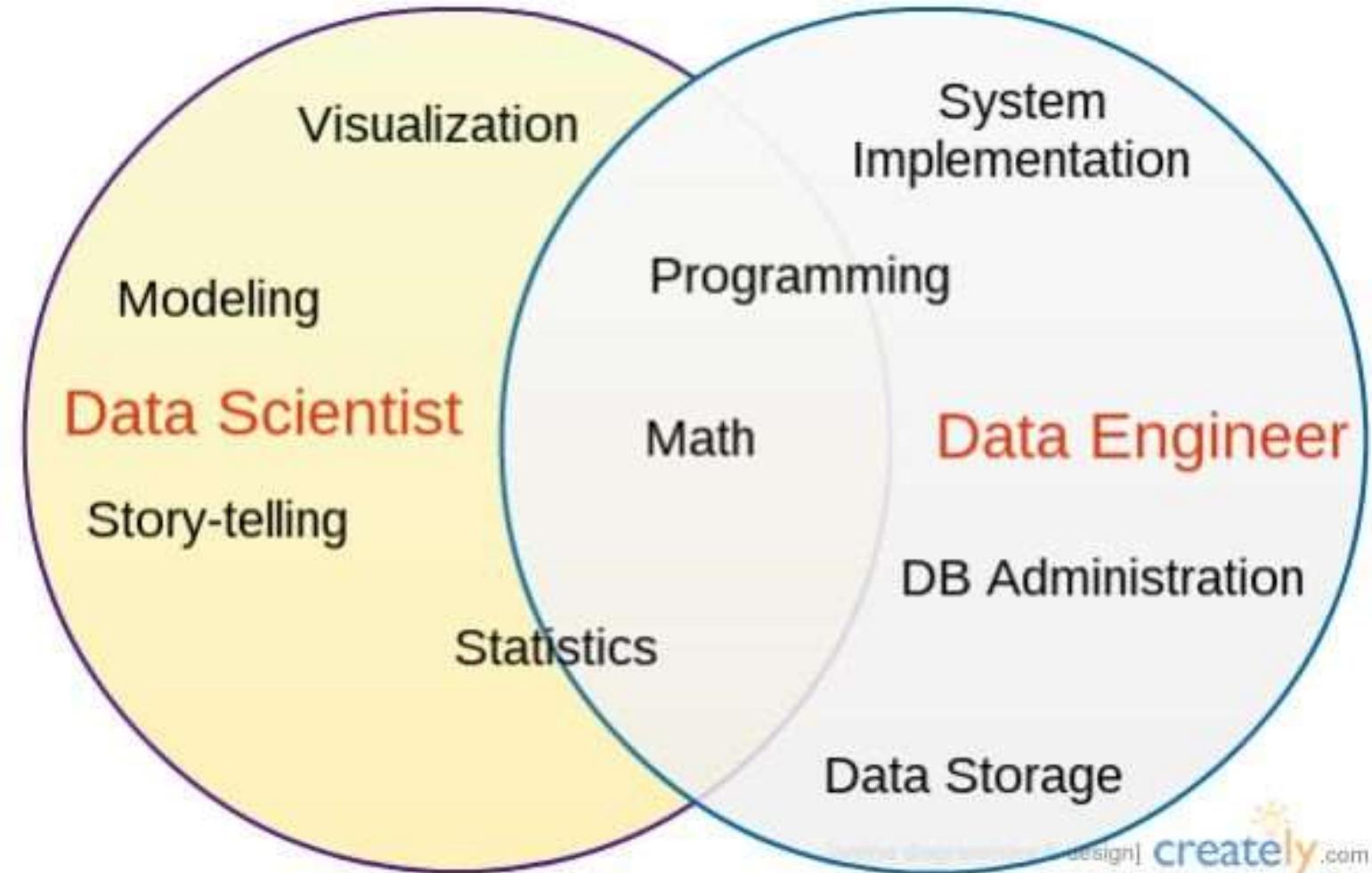
Data Science vs Data Analysis

	Data Science	Data Analysis
Scope	Macro	Micro
Goal	To ask the right questions	Find actionable data
Major Fields	Machine learning, AI, search engine engineering, corporate analytics	Healthcare, gaming, travel, industries with immediate data needs

Data Scientist



Data Scientist vs. Data Engineer





https://www.scnsoft.com/blog-pictures/business-intelligence/what_is_big_data-01_1.png

BIG DATA

An umbrella term for all sorts of data

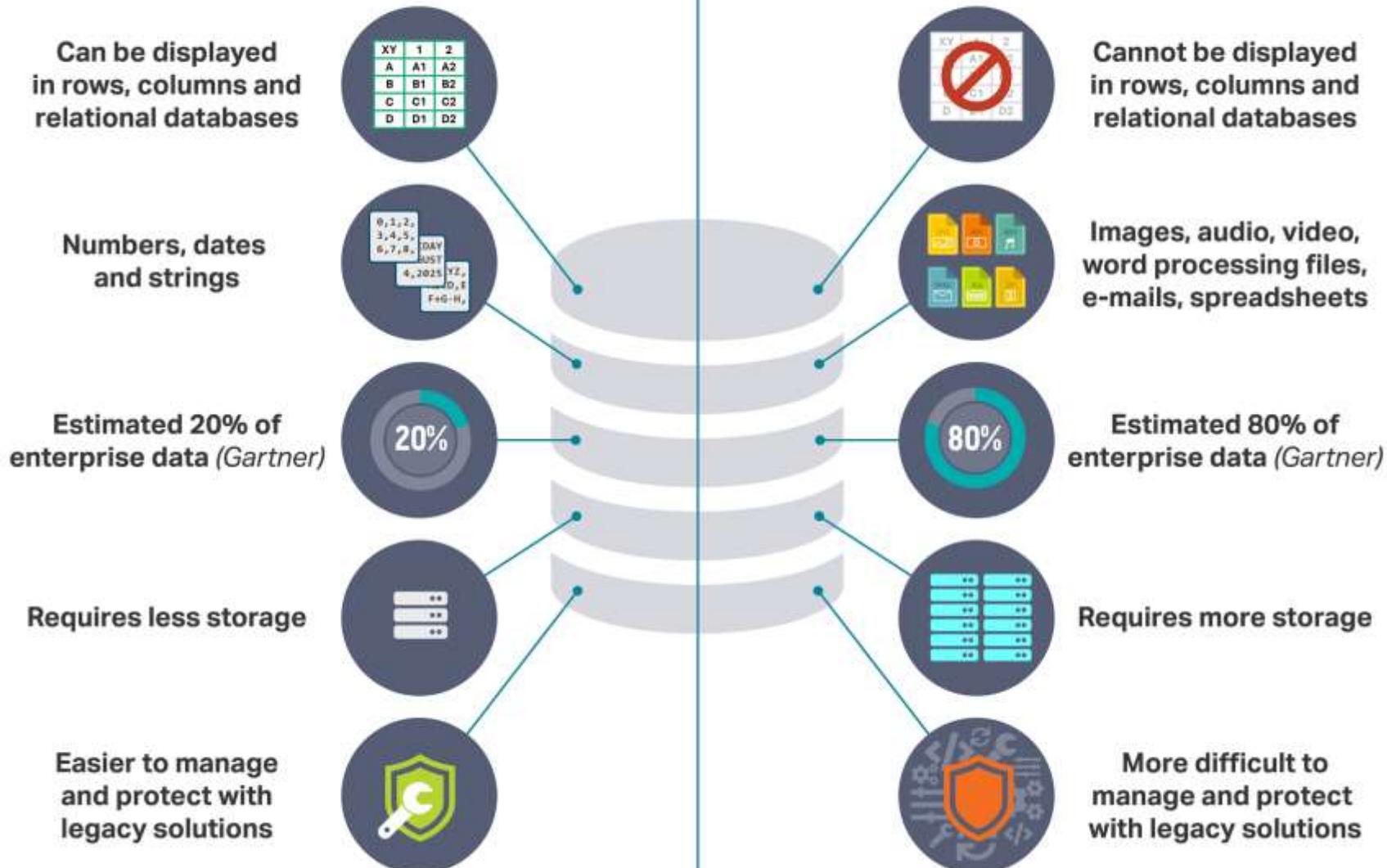


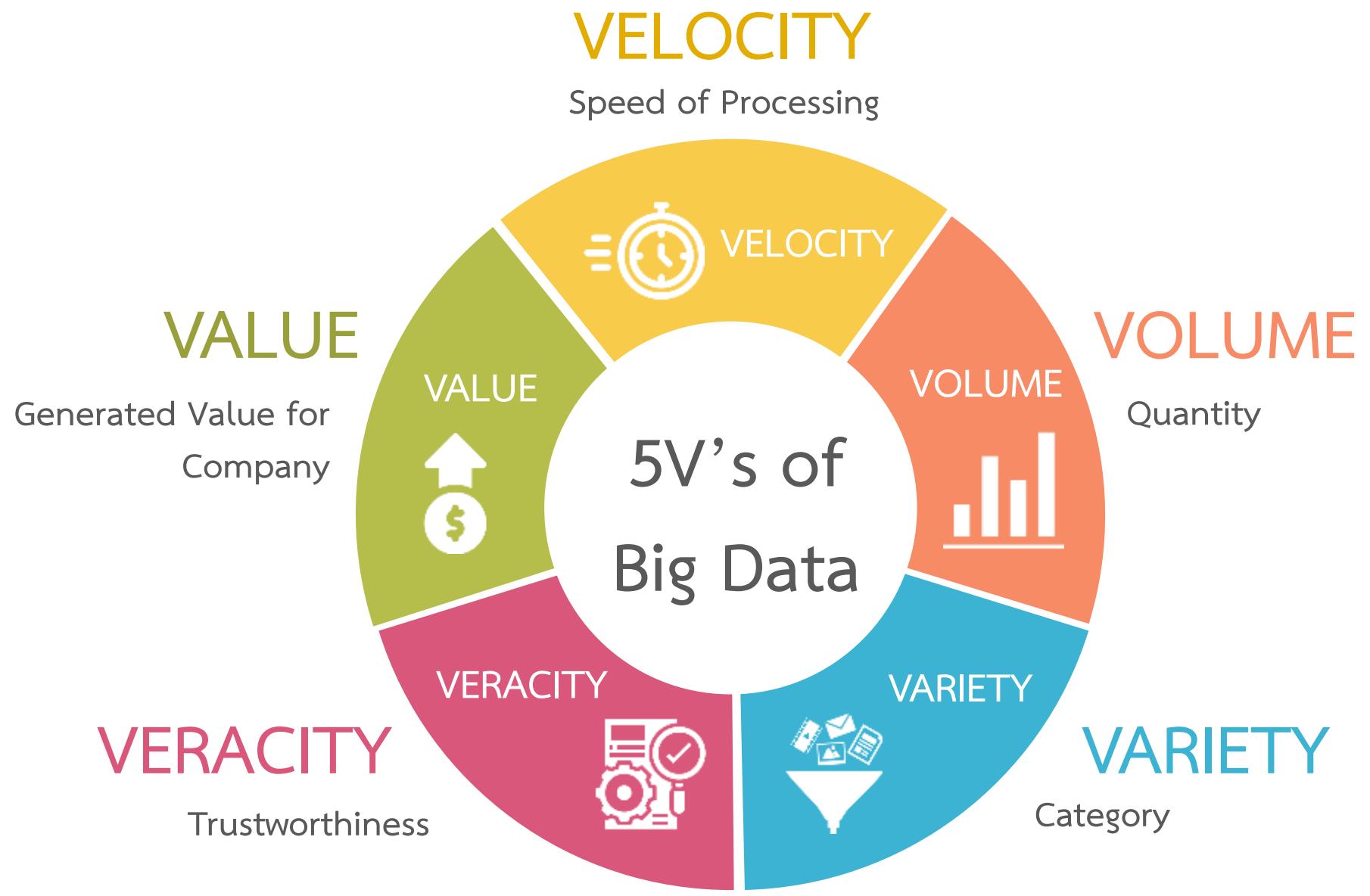
Structured



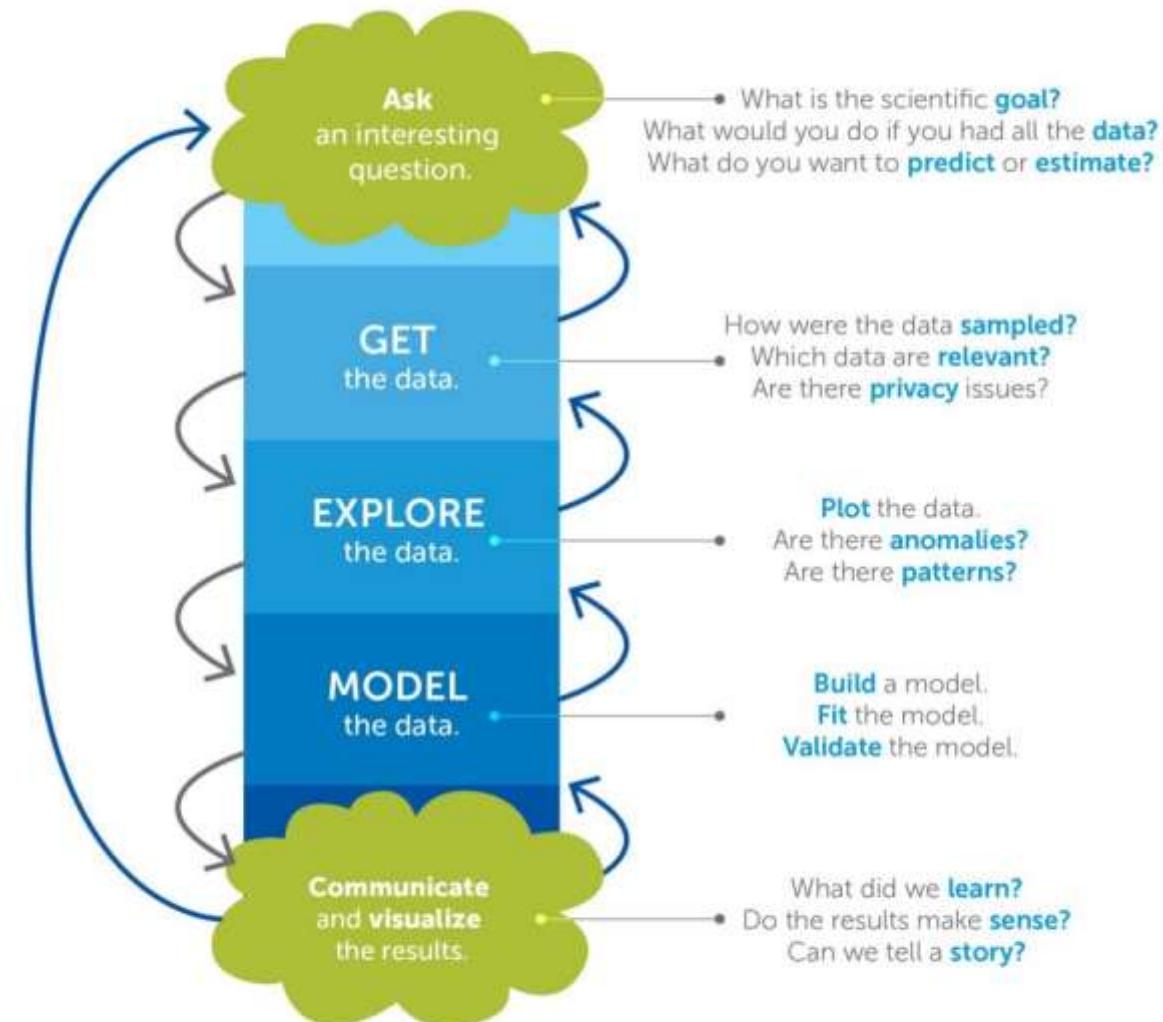
Unstructured

Structured Data vs Unstructured Data



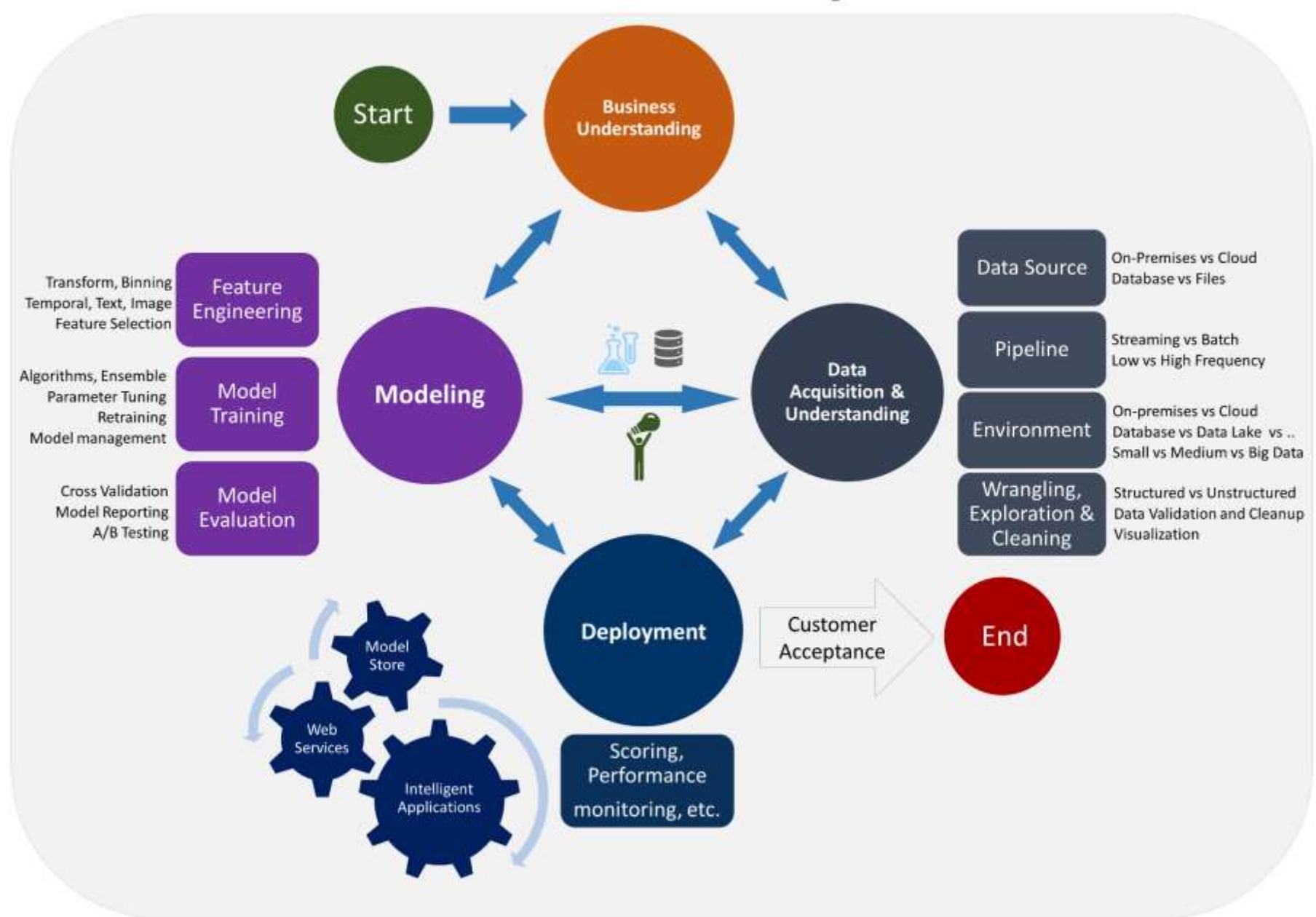


The Data Science Process

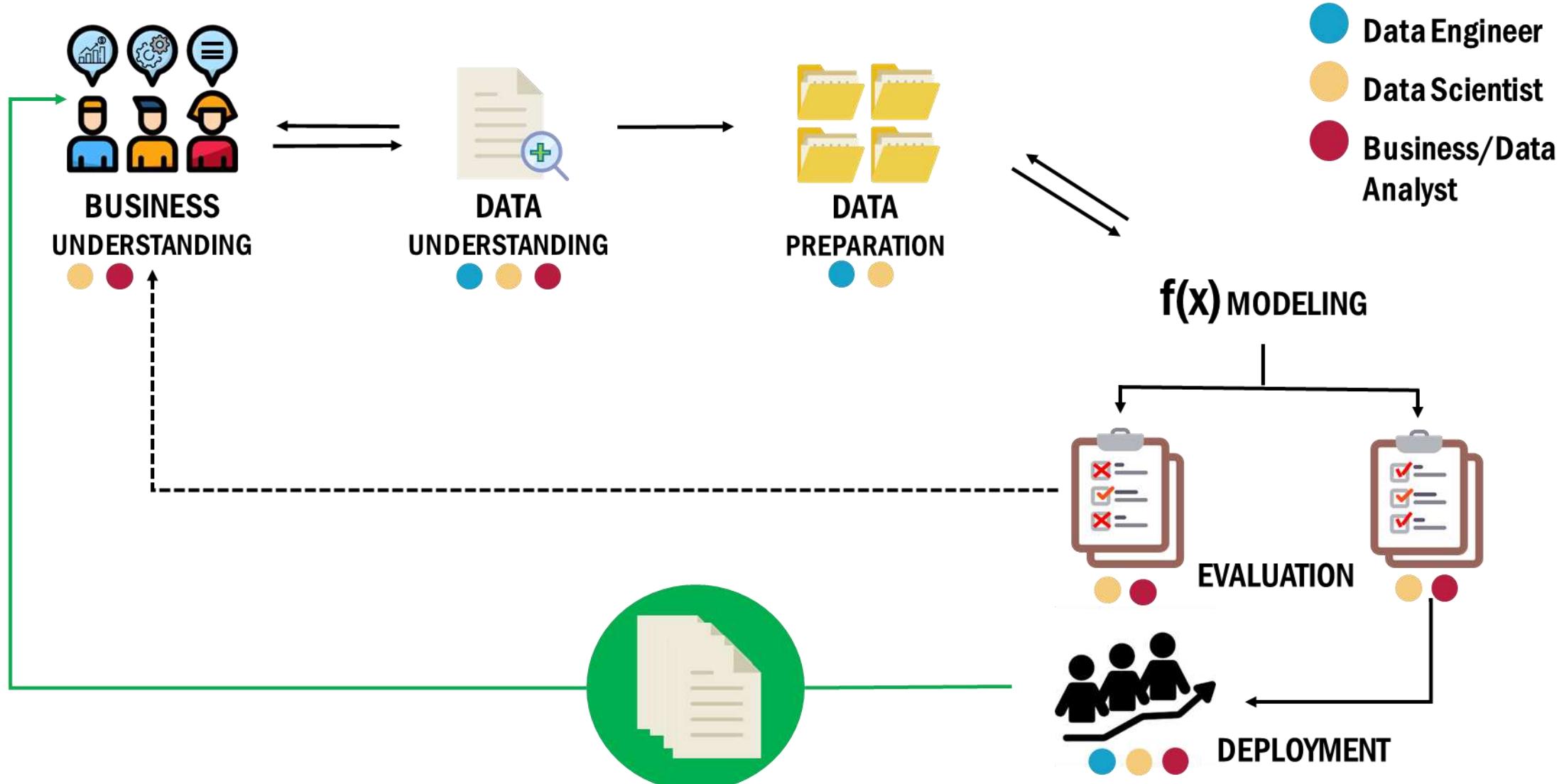


Derived from the work of Joe Blitzstein and Hanspeter Pfister,
originally created for the Harvard data science course <http://cs109.org/>.

Data Science Lifecycle



DATA SCIENCE PROCESS CYCLE



Why Big Data Analytics?

1. Making Smarter and More Efficient Organization



<https://www.edureka.co/blog/big-data-analytics>

Why Big Data Analytics?

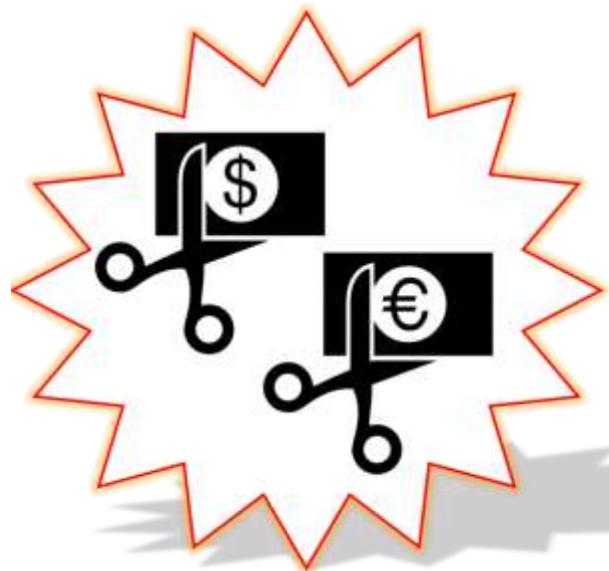
2. Optimize Business Operations by Analyzing Customer Behavior



<https://www.edureka.co/blog/big-data-analytics>

Why Big Data Analytics?

3. Cost Reduction



Parkland Hospital uses analytics and predictive modelling to identify high-risk patients and predict likely outcomes once patients are sent home. As a result, Parkland reduced 30-day readmissions for patients with heart failure, by 31 percent, saving \$500,000 annually.



<https://www.edureka.co/blog/big-data-analytics>

Why Big Data Analytics?

4. New Generation Products

Big Data tools are used to operate Google's Self Driving Cars. The Toyota Prius is fitted with cameras, GPS as well as powerful computers and sensors to safely drive on the road without the intervention of human beings.



Netflix launched the seasons of its TV show House of Cards based on the user reviews, ratings and viewership.

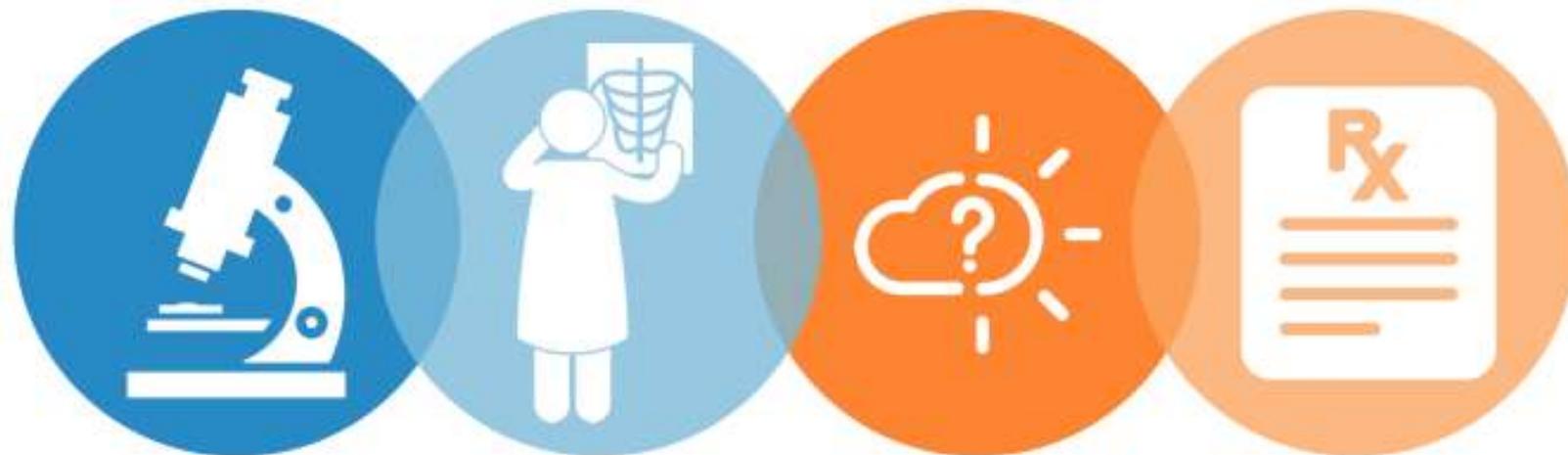


A smart yoga mat has sensors embedded in the mat will be able to provide feedback on your postures, score your practice, and even guide you through an at-home practice.



<https://www.edureka.co/blog/big-data-analytics>

Types of Big Data Analytics



Descriptive

Explains what happened.

Diagnostic

Explains why it happened.

Predictive

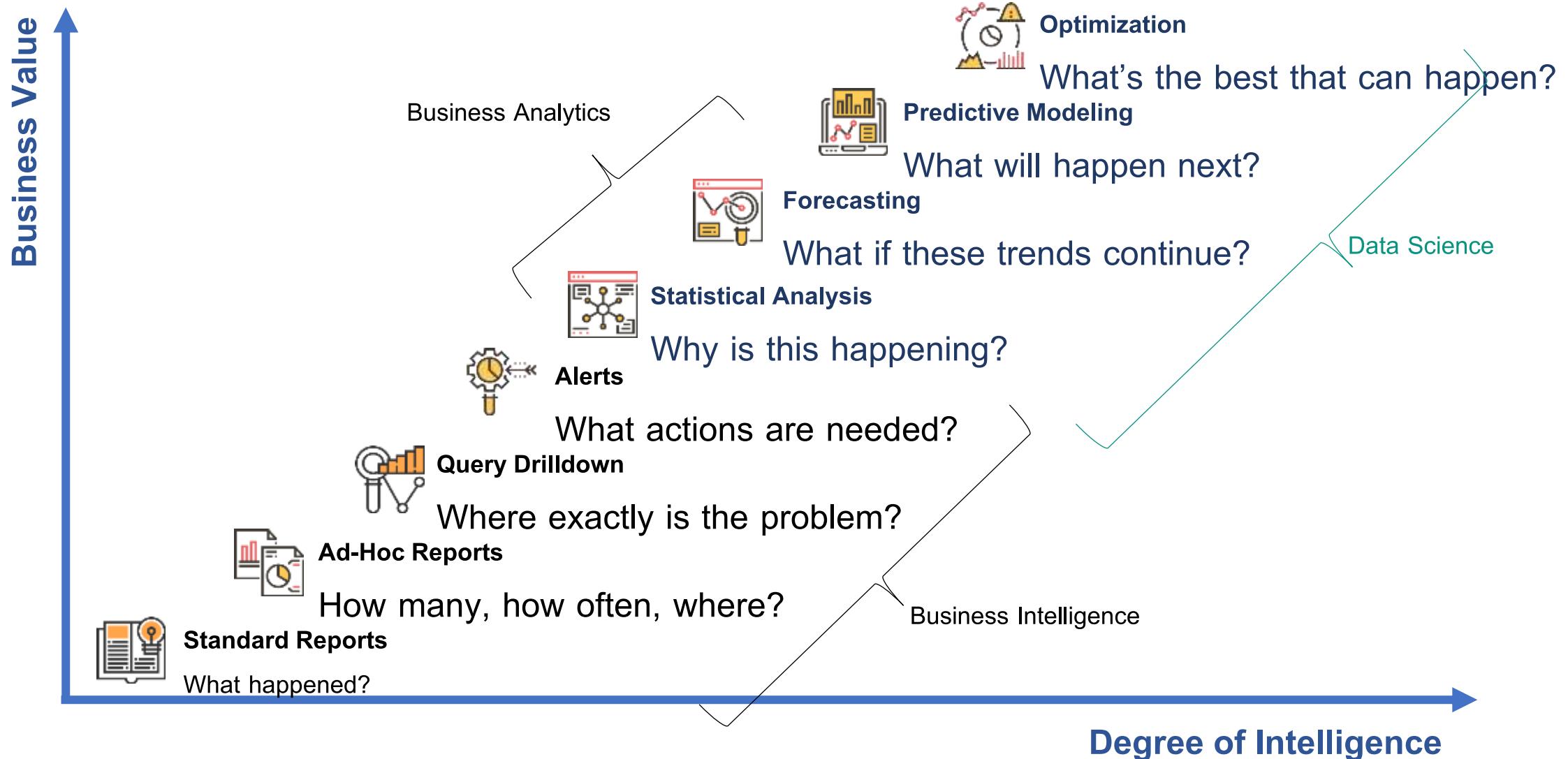
Forecasts what might happen.

Prescriptive

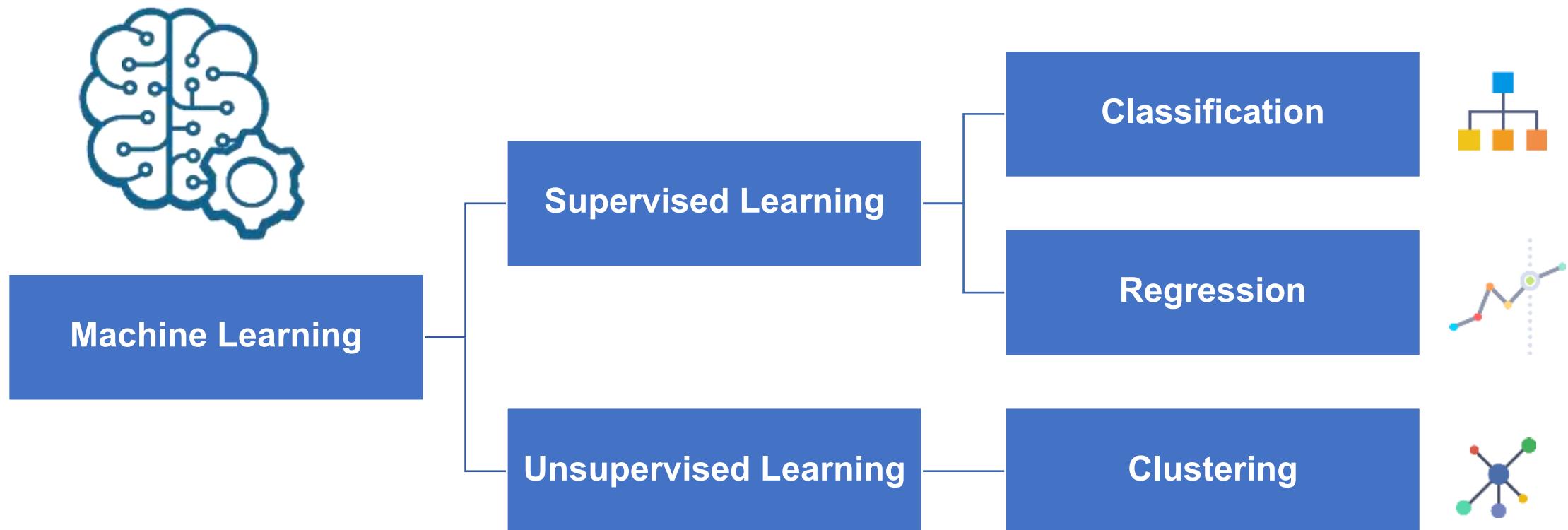
Recommends an action based on the forecast.

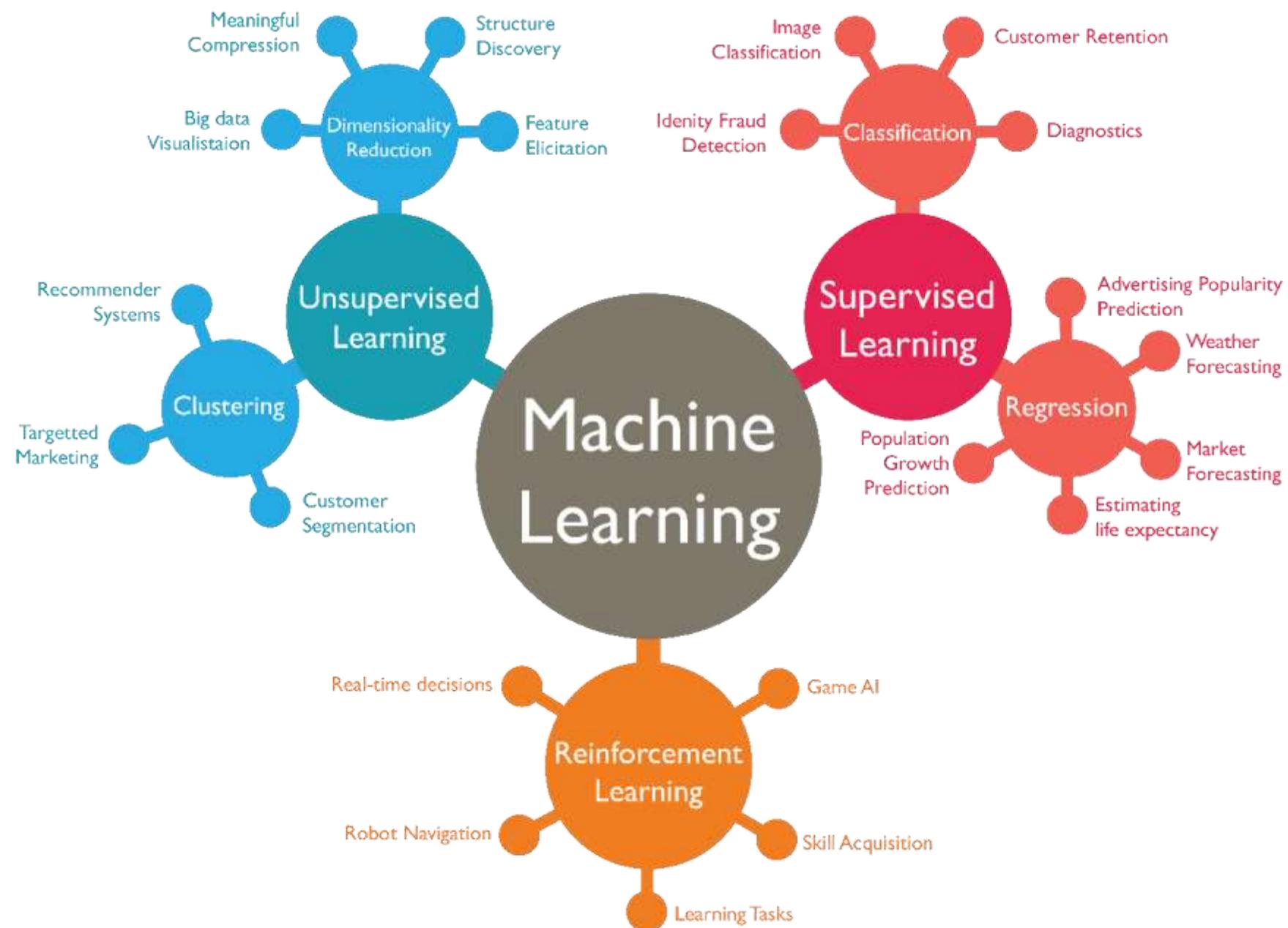
<https://www.analyticsinsight.net/four-types-of-business-analytics-to-know/>

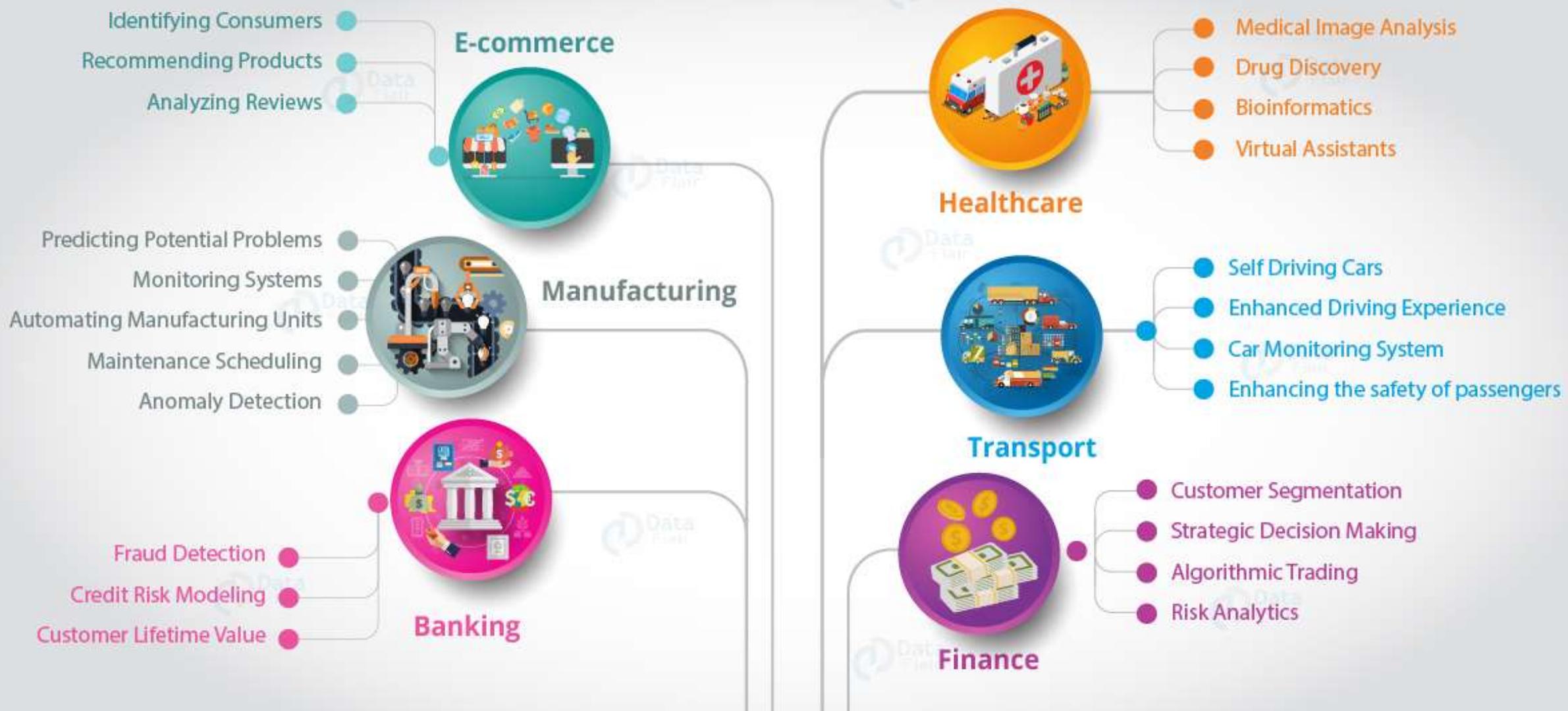
8-LEVELS of ANALYTICS



ML MACHINE LEARNING







Data Science Applications

Data Science Use Cases



SENTIMENT ANALYSIS

- Use text analytics to gather citizen concerns and sentiments
- Assess current situations to create a set of keywords to gather social media post related to topics of interest
- Determine the sentiment with respect to topics or the overall contextual polarity of a post/comment



CREDIT SCORING



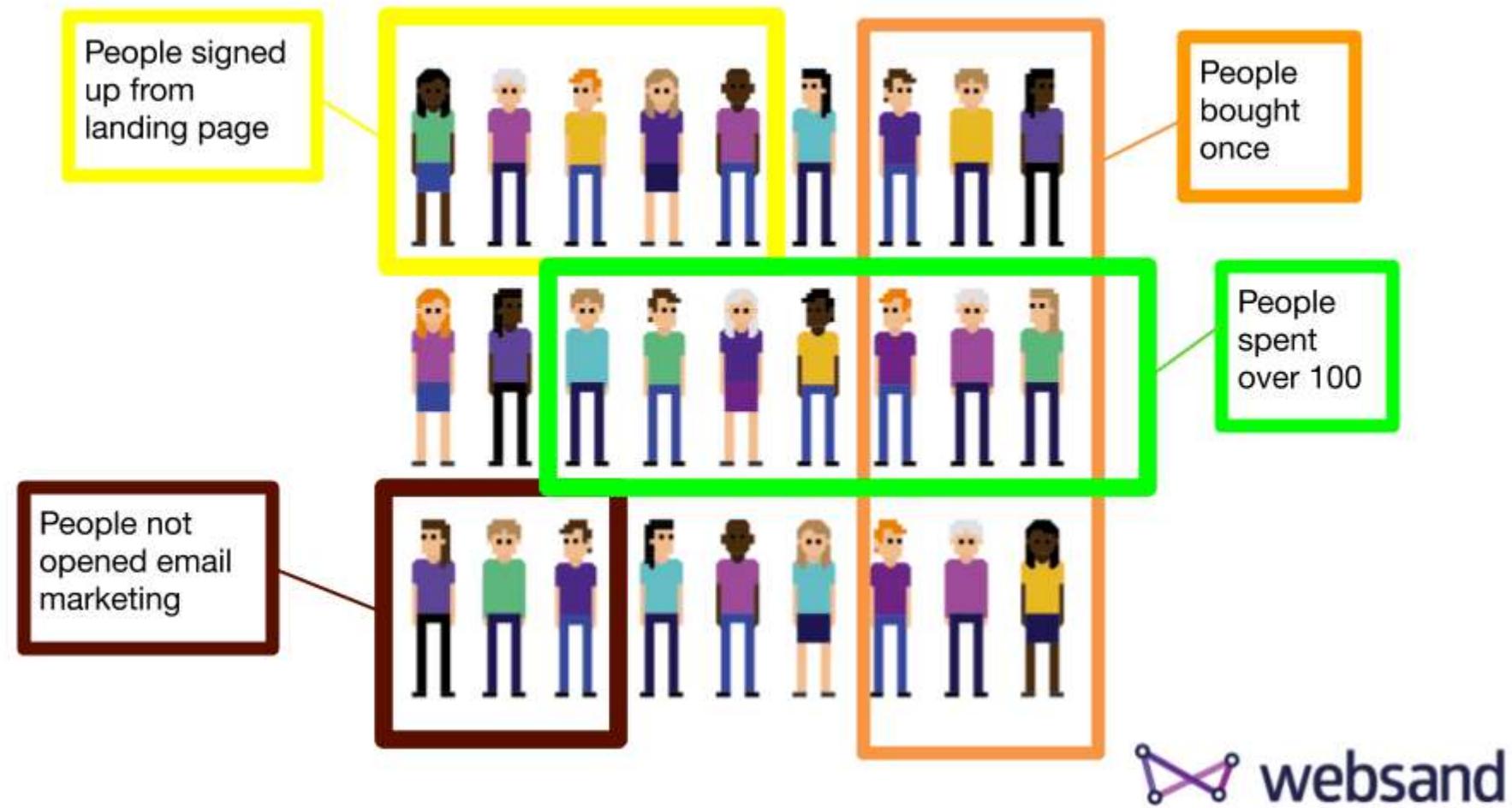
- First Union Bank deployed a value predicting system that assigns green / yellow / red flag to each customer, based on their predicted lifetime value.
- Service representatives were instructed to waive fee for green customers, and not waive for red customers. For yellow customers, they can make their own judgement.
- This strategy generated over \$100 million in incremental revenue.

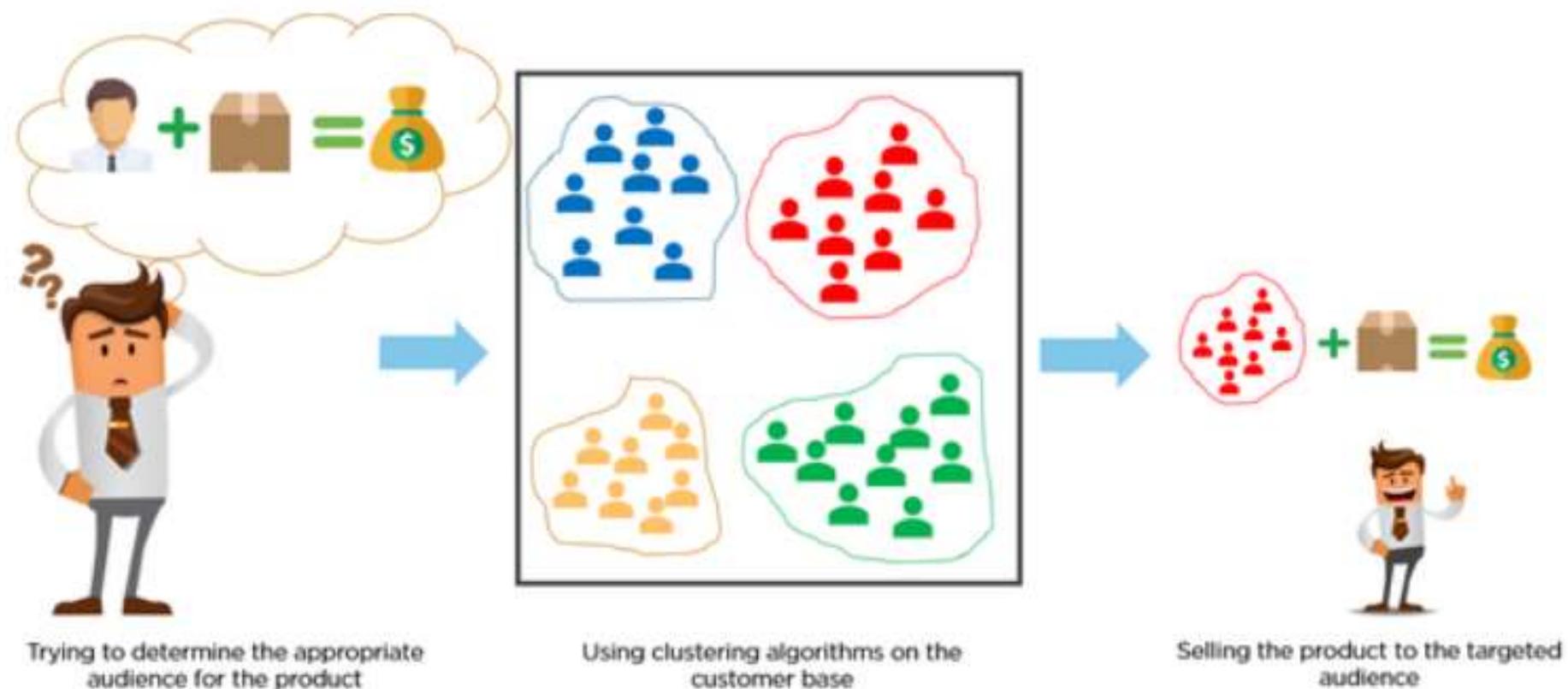


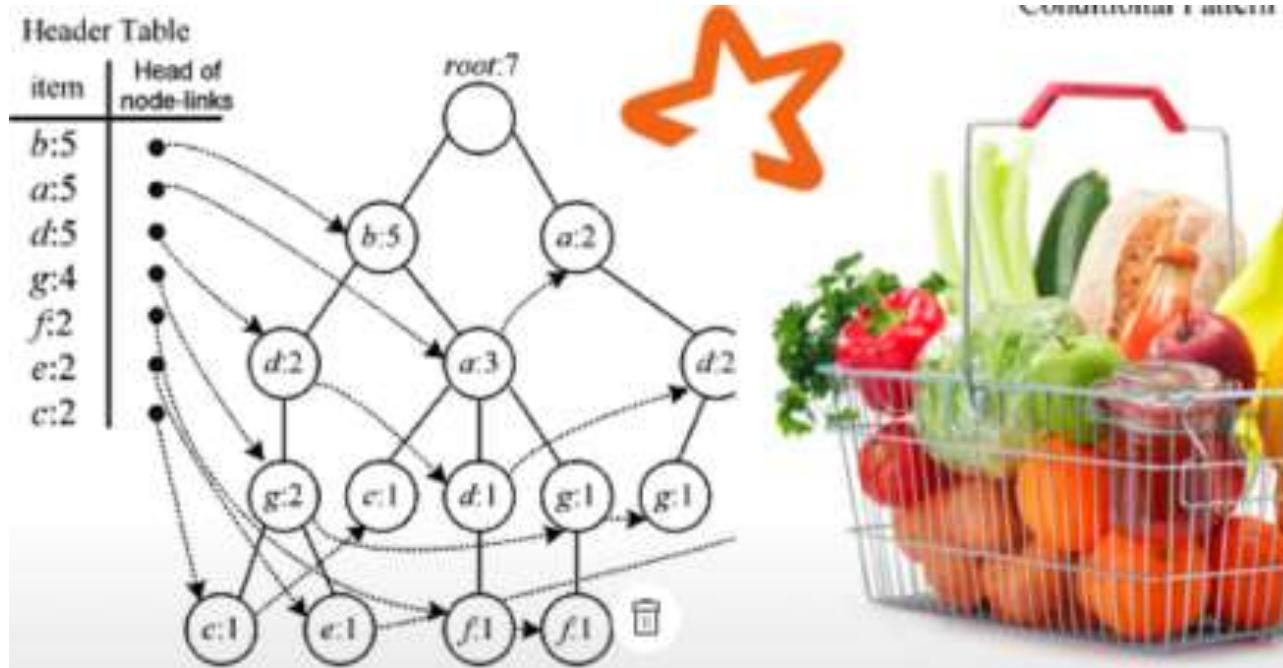
INFLUENCER ANALYSIS

- An influencer is an individual who has above-average impact on a specific niche process.
- On the social network, an influencer can be referred to as the most shaping a discussion about a topic.

Customer Segmentation







#AssociationRules, #MarketBasketAnalysis

MARKETING

- What product are often purchased together ?
- Does product brands matter ?
- How are the demographics of neighborhoods effecting what customers are buying ?
- Where should each product be shelved to maximize the sales ?



STOCK THE RIGHT PRODUCTS

Specifically to each community

Predict purchasing trends and
Manage efficient stock

- What do each customer buy ?
- Are purchasing patterns fall into behavior groups ?
- Are there purchasing patterns and trends over time ?

RECOMMENDER ENGINE FOR CREDIT CARD USERS



Predictive Marketing Analytics

- Predict shop preference from previous behaviors.
- Swipe time prediction.
- Win back inactive customers by predicting deal preference from demographics.

MICRO-LOAN

From government to local convenient stores

Predict collection results and segment customers

- Why can't they pay?
- How late will they pay in the upcoming cycle?
- What actions have been taken and how to maximize collection volume?



#classification

PEOPLE ANALYTIC

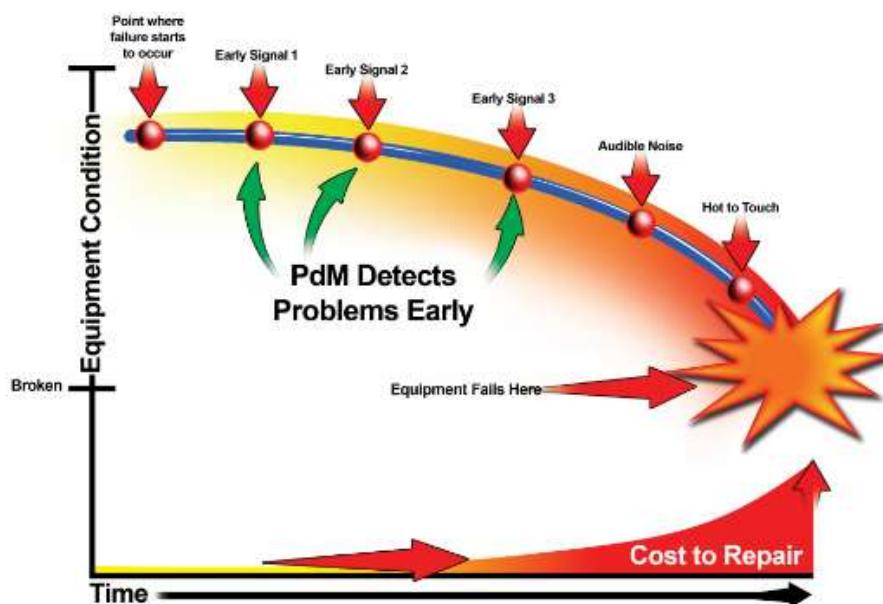
- Predictive analytics is used for talent acquisition, retention, placement, promotion, compensation, or workforce and succession planning.
- Analyzing the skills and attributes of high performers in the present, then build a template with quality hiring factors for future hires.
- Non-traditional data gathering sources
 - Social media channels where prospective candidates usually leave their digital '*thought prints*'.
- Statistical analysis of productivity and turnover
 - The data showed that old indicators (such as GPA and education) were far less critical to performance and retention. Factors like experience is much more important.



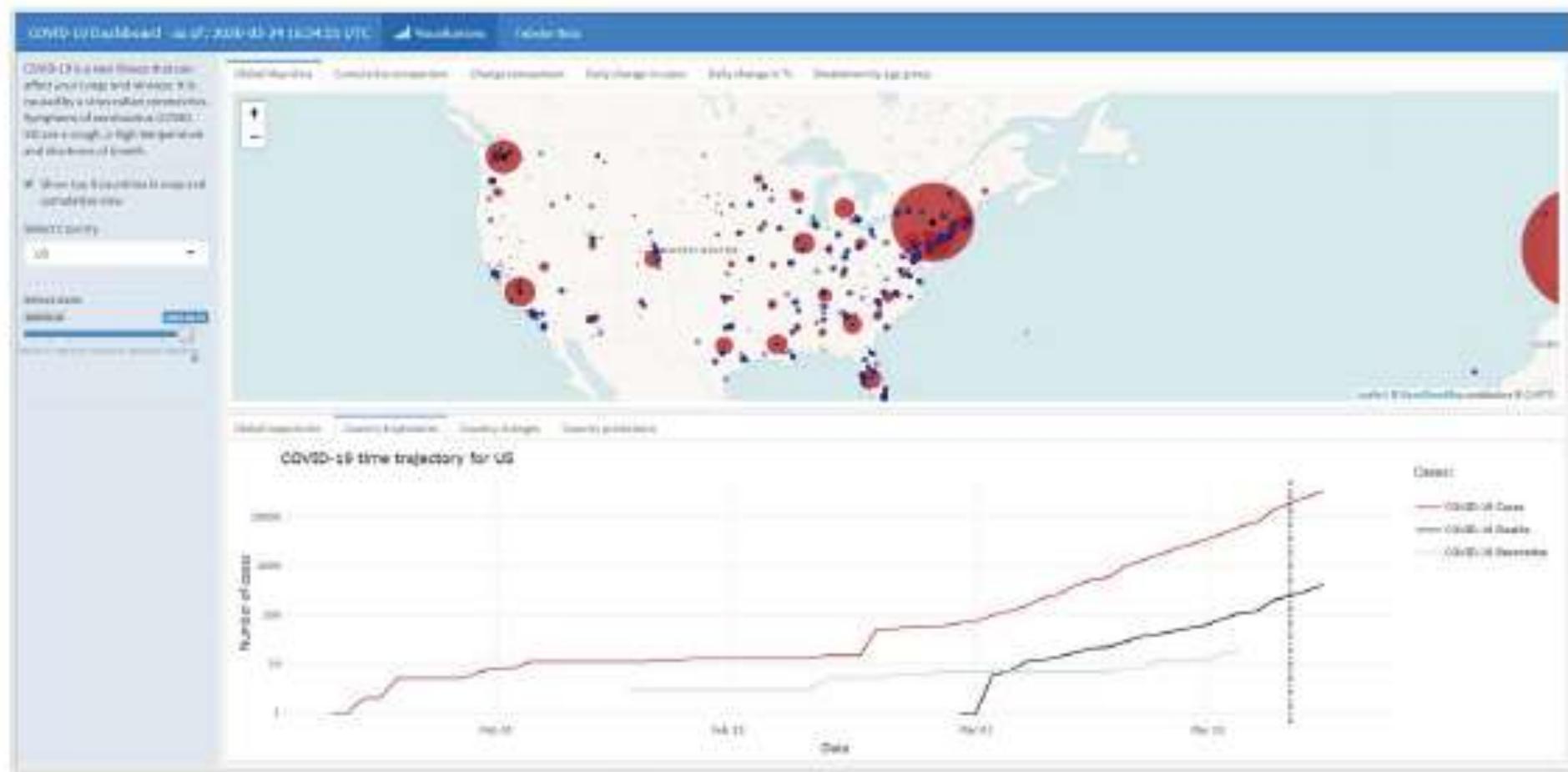
Ref: Forbe

PREDICTIVE MAINTENANCE (PdM)

- Determine the condition of in-service equipment in order to predict when the maintenance should be performed
- Cost savings over routine or time-based preventive maintenance
- Fault Tree Analysis,
- Time Series Analysis



COVID-19 Prediction Dashboard



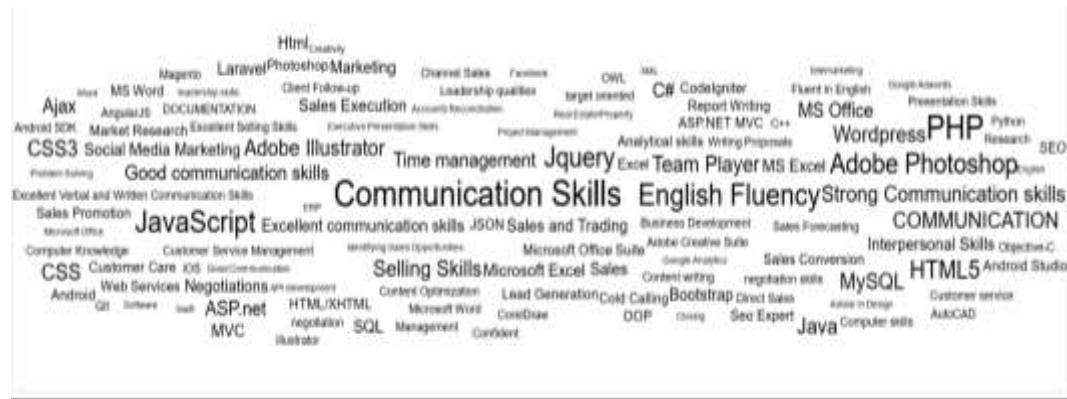
Source: Arthur D. Little analysis

The Case of a Pakistani Online Job Portal

Study labor market conditions and skill demand and supply in Pakistan based on data from a leading online job portal, Rozee.pk.

- Data Used
 - Job title
 - Job Description
 - Job Seeker Profile
 - International Standard Classification of Occupations 08 (ISCO08)

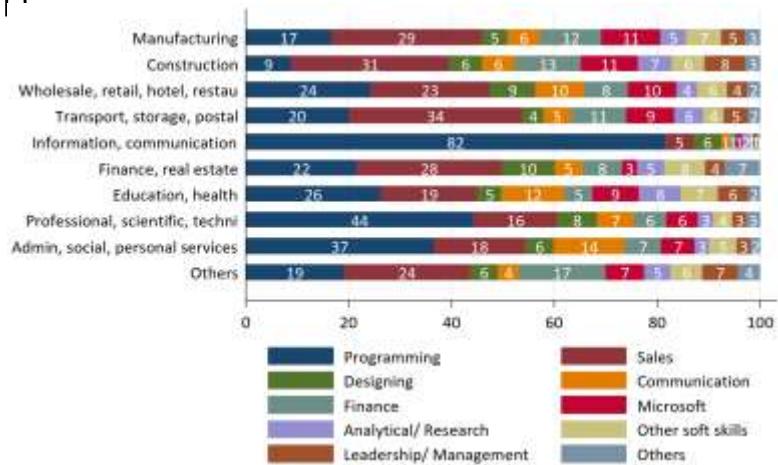
Method



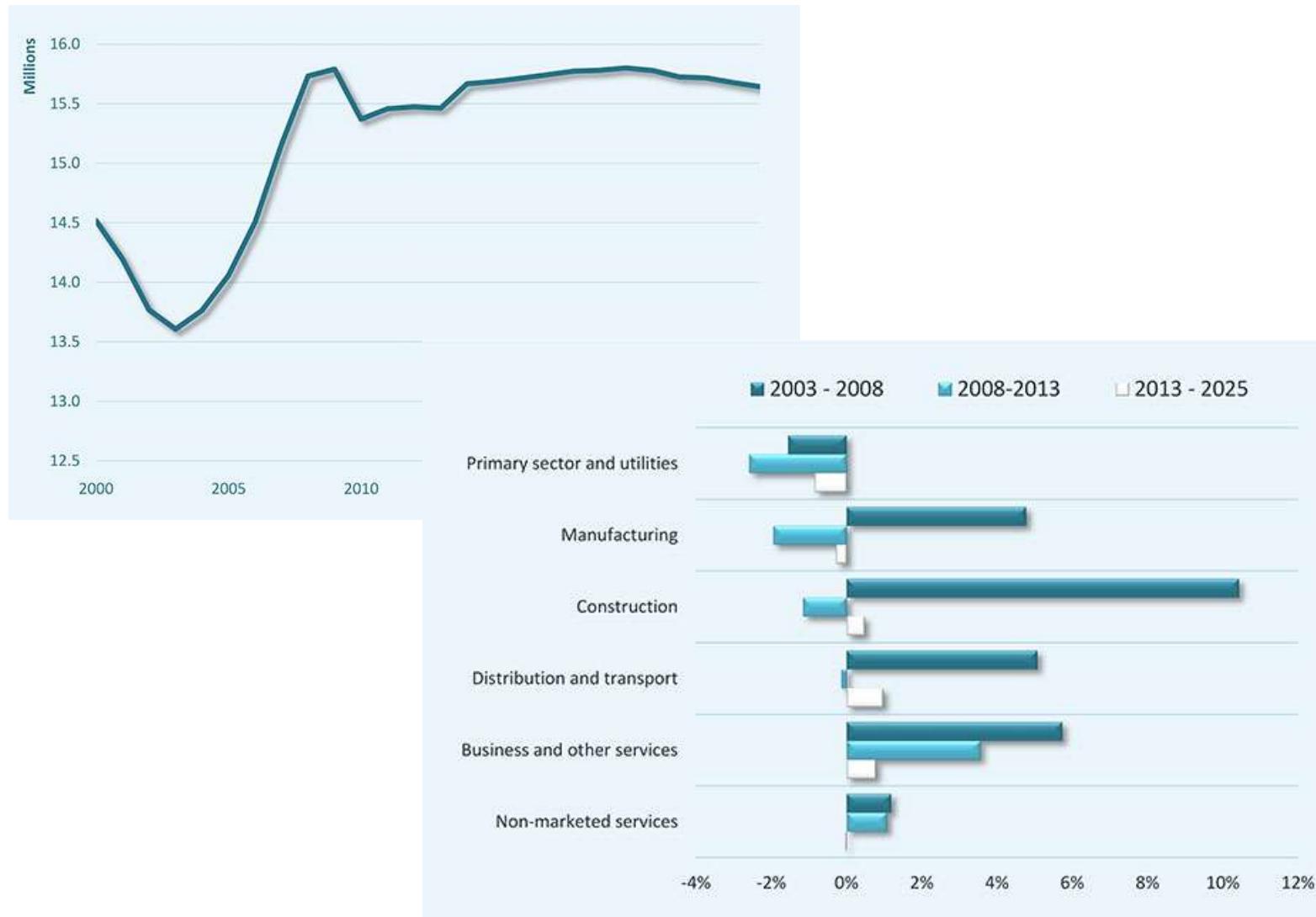
Text Analysis Technique, Categorization

ISCO-08	Occupation titles	Demand	Supply
1	Managers	312,768	230,938
2	Professionals	439,365	309,331
3	Technicians and associate professionals	324,564	303,325
4	Clerical support workers	46,362	69,733
5	Service and sales workers	118,887	75,340
6	Skilled agricultural, forestry and fishery workers	2,043	4,865
7	Craft and related trades workers	19,334	30,456
8	Plant and machine operators, and assemblers	16,955	25,914
9	Elementary occupations	21,588	15,447

- Output
 - Demand and Supply Matching by Job title and skill



LABOR FORECASTING

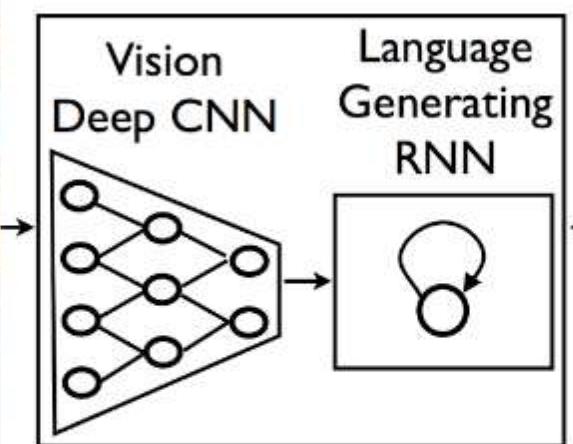


Predict demand and supply of labor

- By skills
- By Industries



State of the Art Deep Learning



Multi-modal
Processing

credit: gigaom.com

State of the Art Deep Learning

Generate high-resolution images based on text.

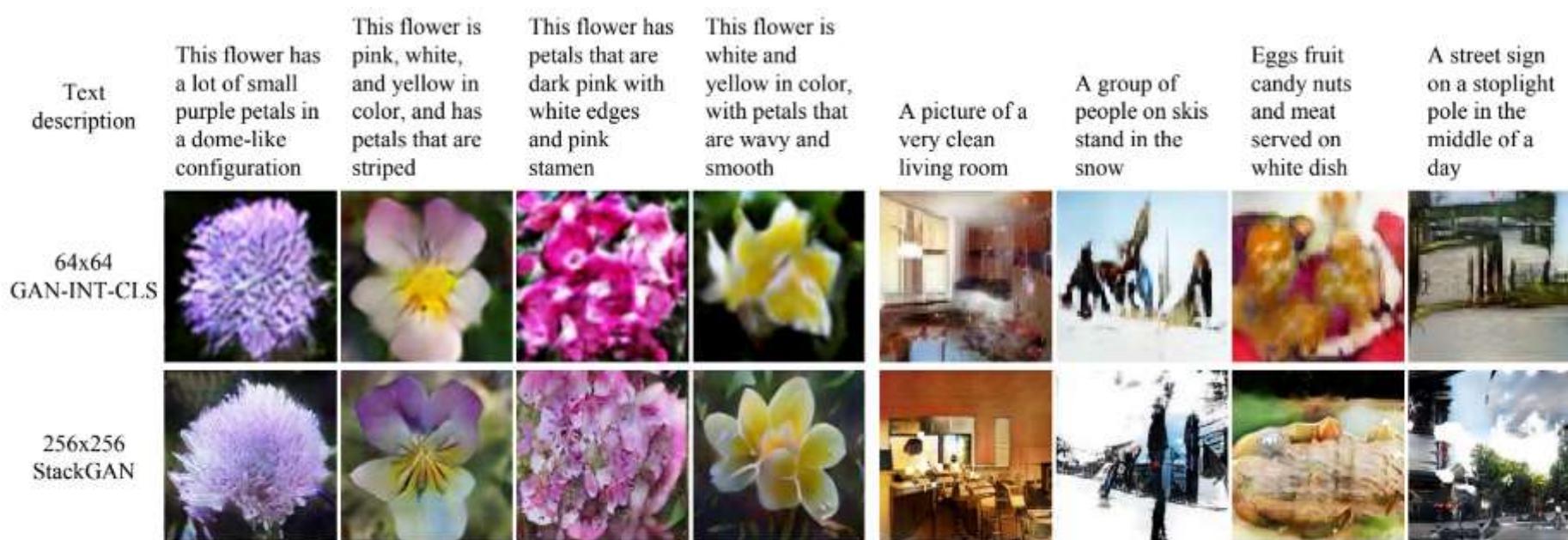
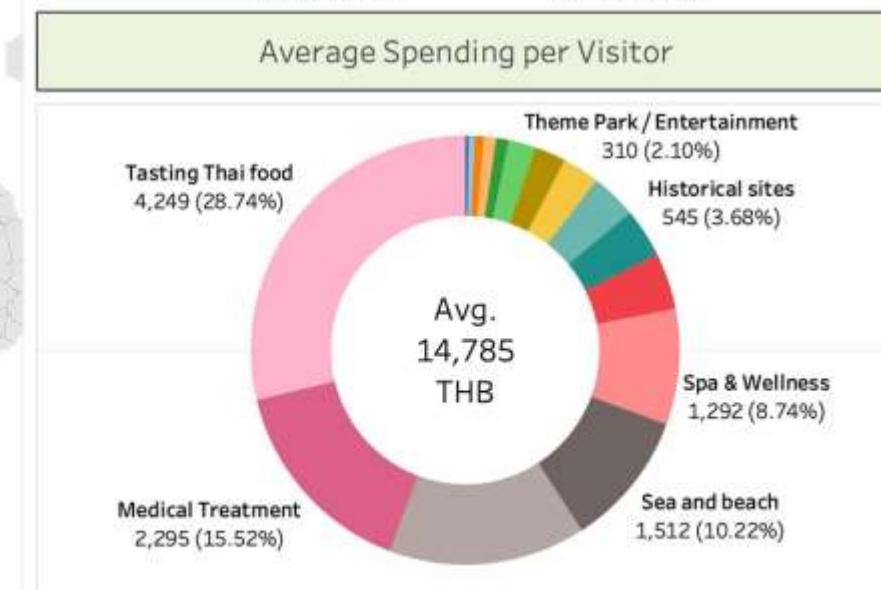
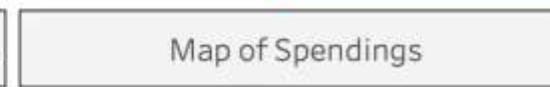
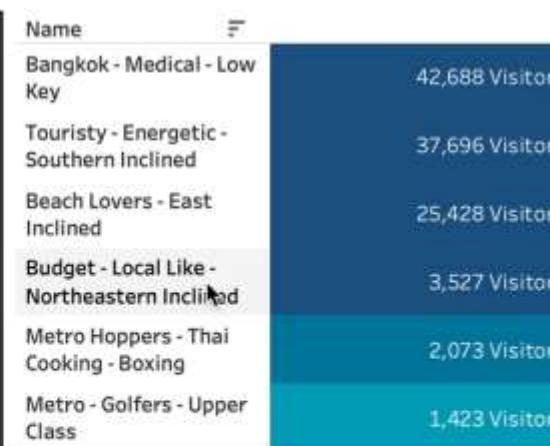
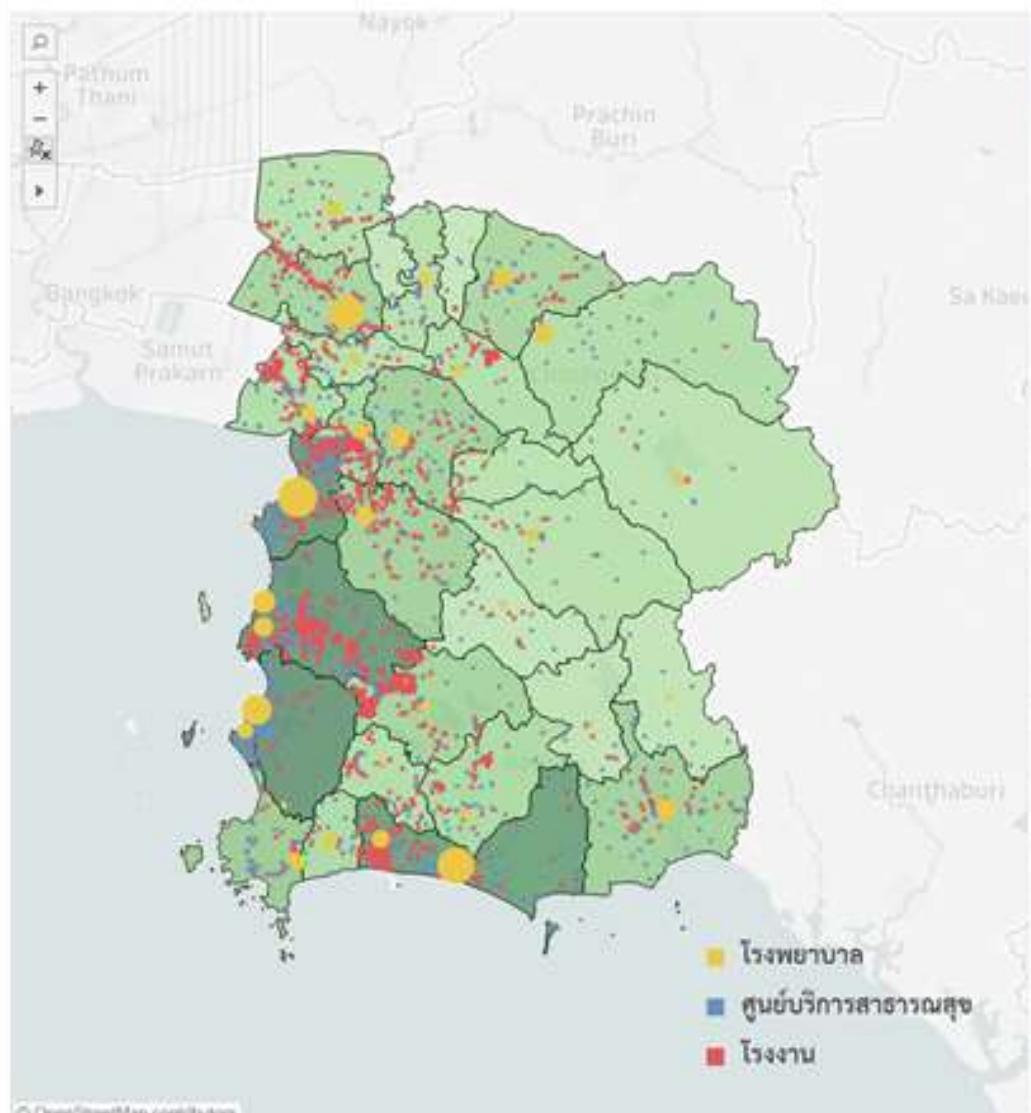


Figure 4. Example results by our StackGAN and GAN-INT-CLS [26] conditioned on text descriptions from Oxford-102 test set (leftmost four columns) and COCO validation set (rightmost four columns).

Big Data Case Study



ภาพรวมการกระจายตัวในเขต EEC - ที่ตั้งรพ.และโรงพยาบาล



กราฟแท่งที่

ที่ตั้งโรงพยาบาล

จำนวนสถานบริการสาธารณสุขในพื้นที่

2,288

จำนวนโรงพยาบาลในพื้นที่

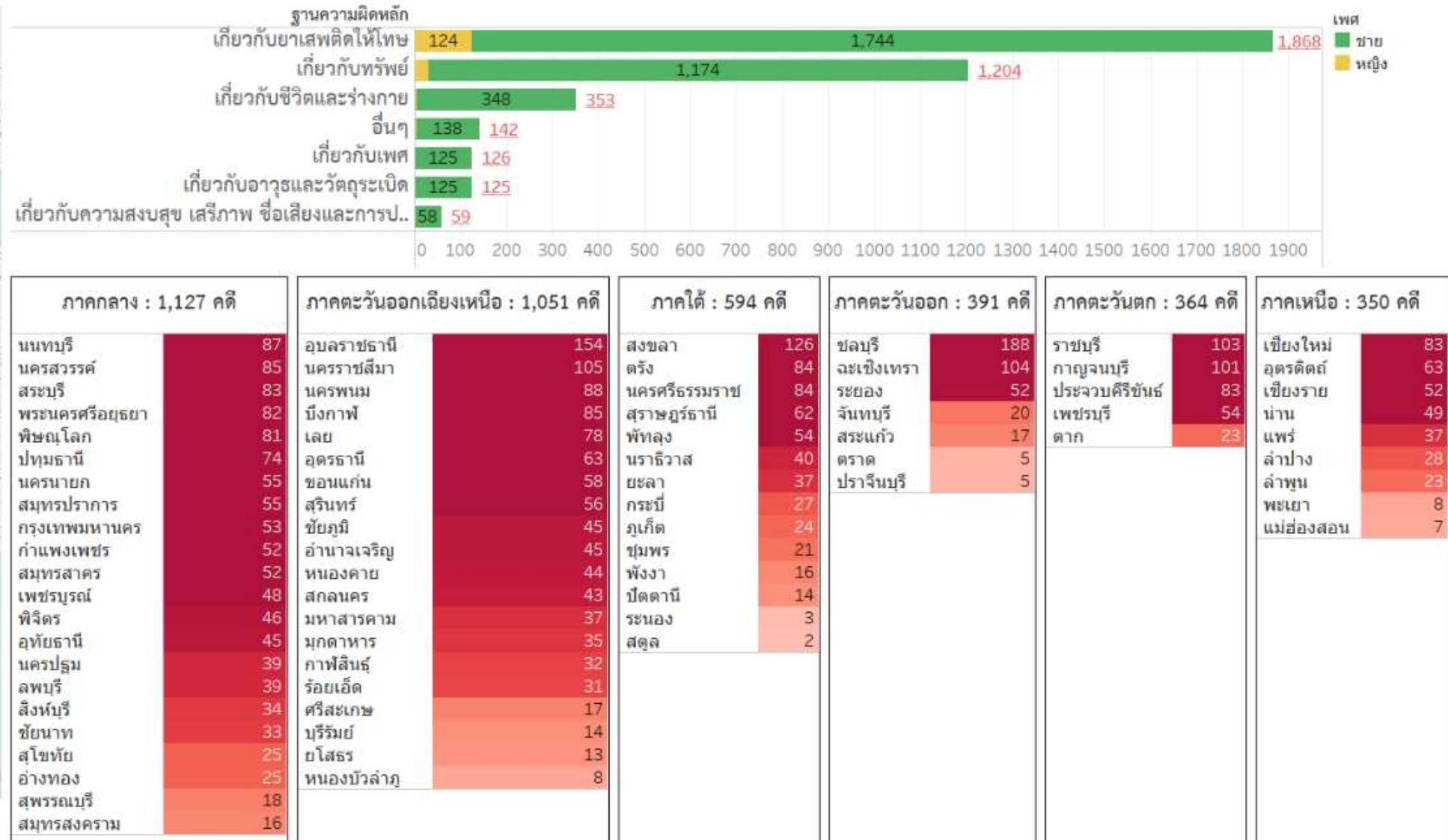
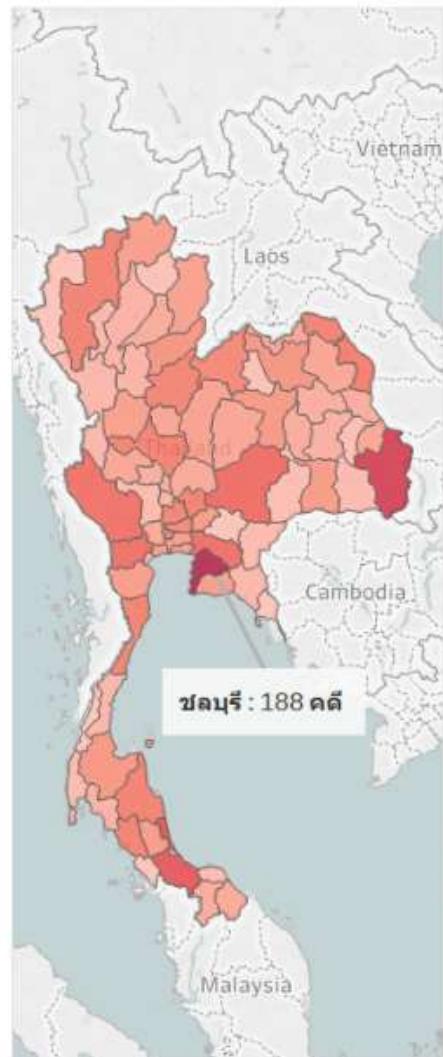
3,373

ข้อมูลผู้ป่วยใน (IPD)

การให้บริการ (ครั้ง)	ค่าใช้จ่ายค่าครัวเรือนเดือน (บาท)	จำนวนคืนที่พักเฉลี่ย (คืน)	จำนวนเตียงนอนผู้ป่วย
โรงพยาบาลสกลบุรี	101,339	26,638	5
โรงพยาบาลราชยอ	86,805	19,237	5
โรงพยาบาลสุทธิโภธ	75,550	18,090	5
โรงพยาบาลศรีบังคิรเมือง	58,846	10,582	4
โรงพยาบาลสุรศรีพัฒน์เจพระร่วง...	34,503	12,477	4
โรงพยาบาลสกลนคร	30,017	11,111	4
โรงพยาบาลสันติคุณ	27,472	11,242	4
โรงพยาบาลสหกรณ์ฉะบัง	24,199	11,515	4
โรงพยาบาลสันบ้านบึง	23,115	6,828	3
โรงพยาบาลสุรศรีพัฒน์การค้า	22,489	8,042	4



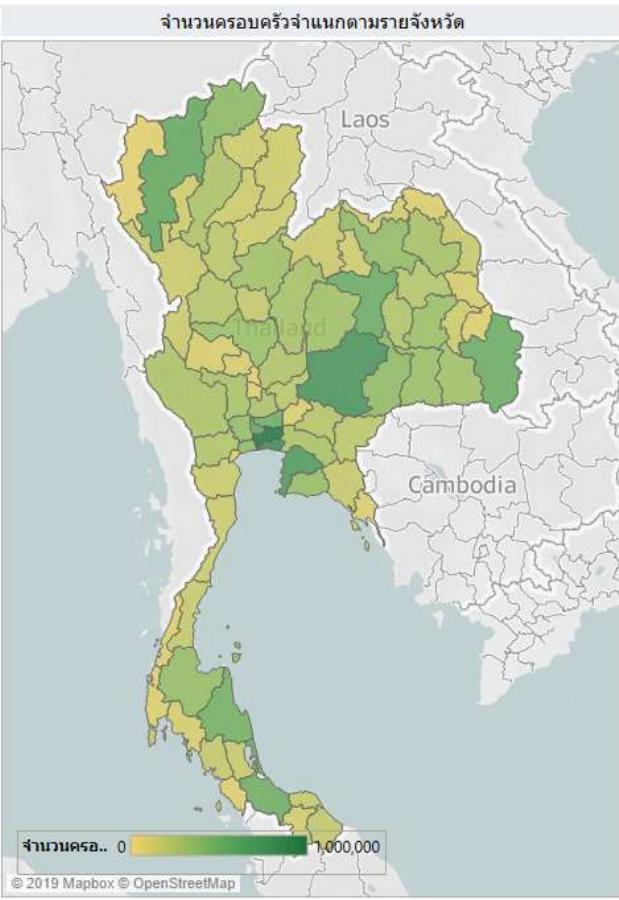
จำนวนคดีในแต่ละภูมิภาค



ข้อมูลสถิติครอบครัว

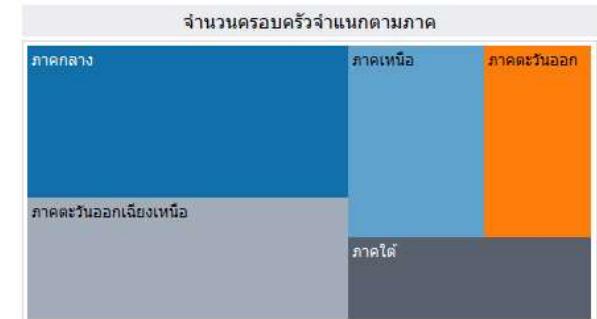
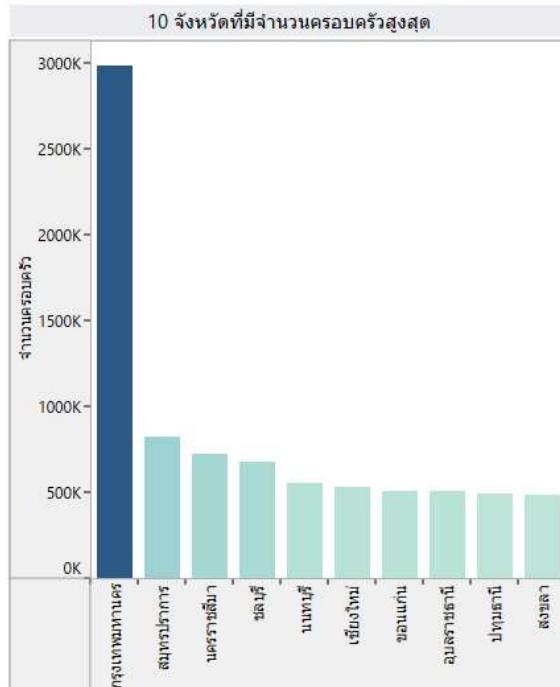
...

จำนวนครอบครัวรายจังหวัด



รวม	พ่อ/แม่เลี้ยงเดียว	สามี-ภรรยา	พ่อแม่ลูก
21,514,737	1,379,860	3,920,008	4,503,363
ผู้ไทย-หลาน	ขยาย	ล่าพัง	ไม่ใช่ญาติ

407,505 7,089,220 4,128,923 85,859



จำนวนโครงการและจำนวนเงินทุนเฉลี่ยในแต่ละสาขา



- ປີ
- (All)
 - 2551
 - 2552
 - 2553
 - 2554
 - 2555
 - 2556
 - 2557
 - 2558
 - 2559
 - 2560
- สาขาวิชา
- (All)
 - No info
 - สาขาวิชาศึกษา
 - สาขาวิชาเกษตรศาสตร์และชีววิทยา
 - สาขาวิชาเทคโนโลยีสารสนเทศและนิเทศศาสตร์
 - สาขาวิชานิติศาสตร์
 - สาขาวิชาปรัชญา
 - สาขาวิชารัฐศาสตร์และรัฐประศาสนศาสตร์
 - สาขาวิทยาศาสตร์กายภาพและคณิตศาสตร์
 - สาขาวิทยาศาสตร์การแพทย์
 - สาขาวิทยาศาสตร์เคมีและเกลือ
 - สาขาวิศวกรรมศาสตร์และอุตสาหกรรมวิจัย
 - สาขาวิชารัฐศาสตร์
 - สาขาวิชสังคมวิทยา

การกระจายตัวของเงินทุนโครงการในแต่ละสาขา





Data Platform



Co-Link:

- Covid Data Consolidation and Linkage Platform (MOPH, MDES, Bangkok, etc.)

Health Link:

- Linking patients' data across hospitals in Thailand (MOPH, MDES, Private hospitals)

Travel Link:

- Thai tourism data and AI service platforms (MOTS, MDES, TAT, Phuket Sandbox, Startups)

Youth Link:

- Linking Children data (Education, Health, Welfare)

SPM Link:

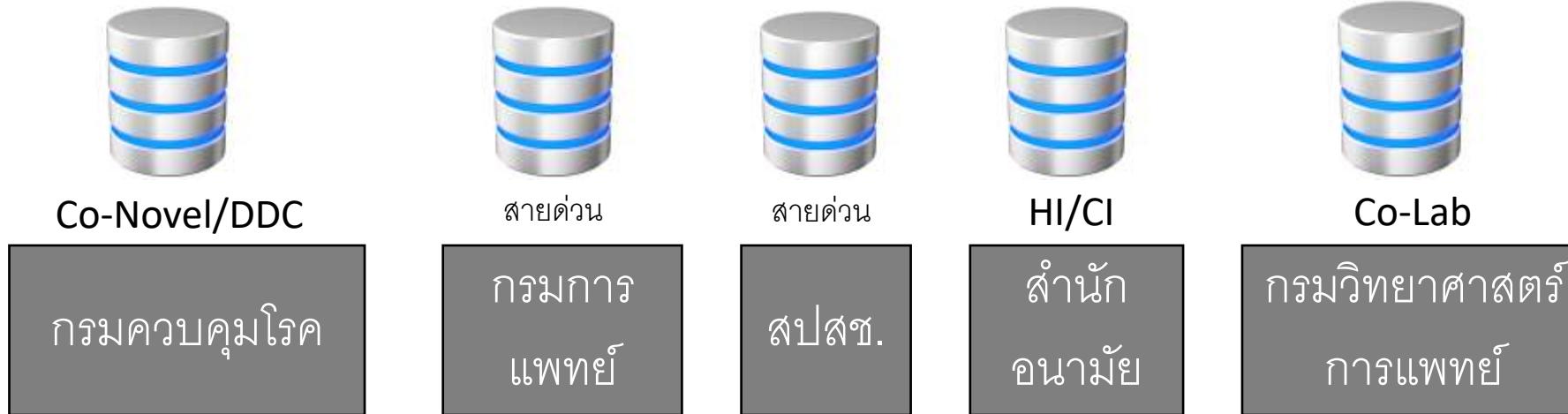
- Government Project Tracking Platform



CO-Link:

Covid Data Platform

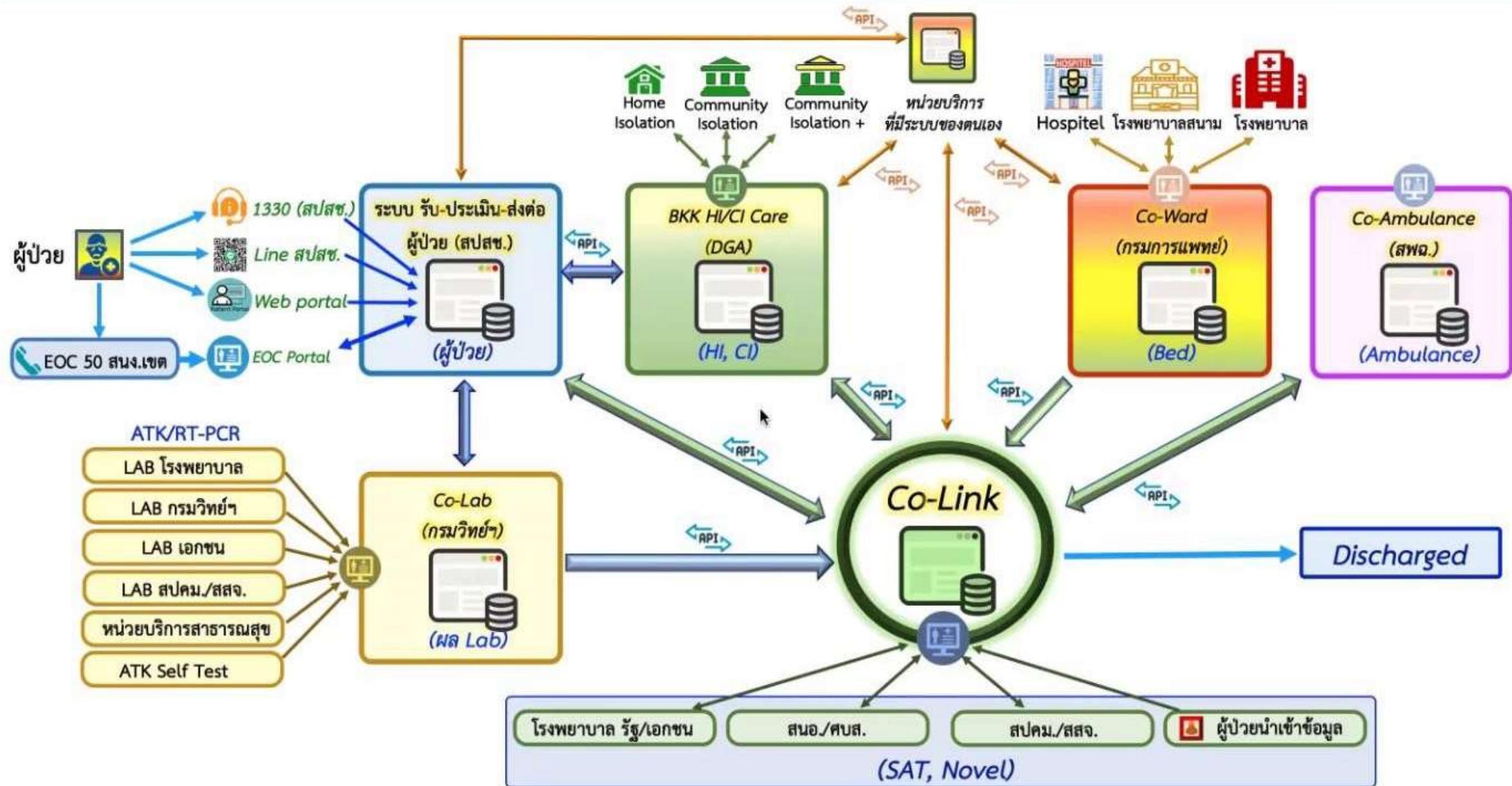
ตัวอย่างความจำเป็นของการมี Single Portal เพื่อบริหารและบูรณาการข้อมูล



โรงพยาบาลต้องการ

- ไม่ต้องค้นหา lab ซ้ำ ไม่ต้องซักอาการซ้ำ ไม่ต้องหาที่ดูแลผู้ป่วยซ้ำ ไม่ต้องตามคนไปรับส่งผู้ป่วยซ้ำ ไม่ต้องถามว่าเตียงว่างใหม่ อย่างดูจากระบบเดียวได้ เพราะหน้างานยุ่งมากอยู่แล้ว
- รับ-ส่งผู้ป่วย / ย้ายผู้ป่วย จากเบ้าไปหนัก หรือ จากหนักไปเบ้า ได้สะดวก
- ให้มีทีมคอยช่วยเหลือโรงพยาบาลที่ต้องการ รับ-ส่ง ข้อมูลเพื่อบริการผู้ป่วยทั้ง Journey ผ่าน APIs
- ให้เอกชนที่อยากรเข้ามาช่วย (ที่ได้รับการรับรองจาก สธ) ทั้งปัจจุบันและอนาคต สามารถนำข้อมูลนี้ไปช่วยกันทำงานต่อให้โรงพยาบาลได้

ตัวอย่างความจำเป็นของการมี Single Portal เพื่อบริหารและบูรณาการข้อมูล



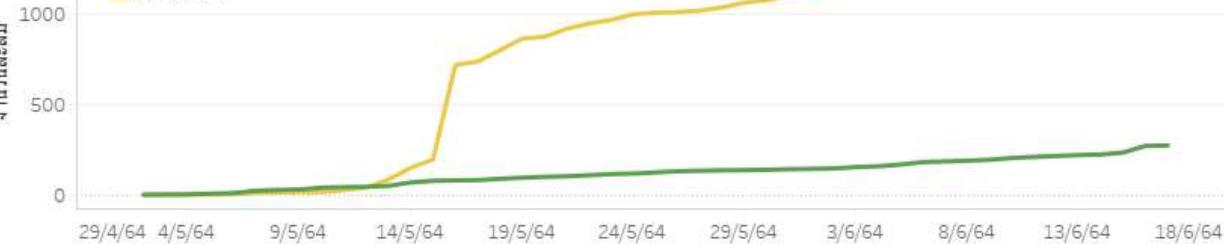
รายงานการจัดสรรเตียงของผู้ป่วย COVID-19

ประมวลผลข้อมูลล่าสุด
17/6/2564 14:28:37

- เตียงหนึ่งตัวที่ติดเชื้อ
- เตียงหนึ่งตัวที่รักษา
- ปิดเตียงแล้ว



- Home/Community isolation & ปฏิเสธการรักษา
- ติดต่อไม่ได้



ยอด Admit รายวัน (ตามวันที่ได้รับการรักษา)



ตัวกรองข้อมูล

จากสปปค

- (All)
- เฉพาะที่มีในระบบ
- เฉพาะที่ไม่มีในระบบ

จาก Hotline

- (All)
- เฉพาะที่มีในระบบ
- เฉพาะที่ไม่มีในระบบ

จากสายด่วนอท

- (All)
- เฉพาะที่มีในระบบ
- เฉพาะที่ไม่มีในระบบ

มีเลข 13 หลัก

- (All)
- นี่
- ไม่นี่

วันเริ่มต้น

1/5/2021

วันสิ้นสุด

28/6/2021

ผู้ป่วยสั่งสม

53,348
(วันที่ 0)

เคลสที่ Admit แล้ว

48,982

ปิดเคลสอีก ๑

1,227

ผู้ป่วยรอเตียง (14 วันจากการตรวจ)

693

แดง

1
(เกิน ๑ วัน: 1 ราย)

เหลือง

98
(เกิน ๒ วัน: 98 ราย)

เขียว

403
(เกิน ๓ วัน: 403 ราย)

รอการคัดกรองความรุนแรง

191

Home / Community Isolation

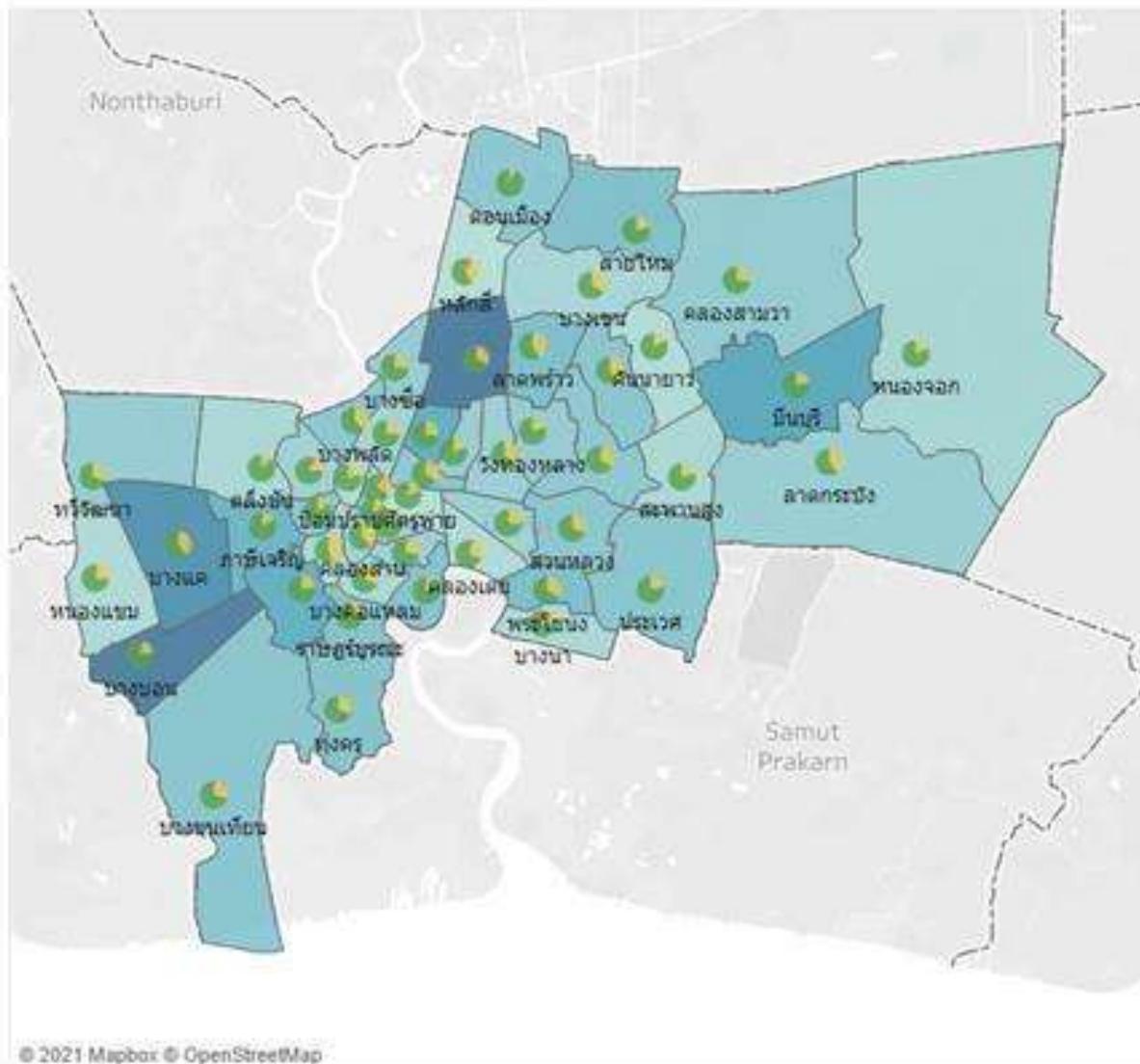
ปฏิเสธการรักษา

37

ติดต่อไม่ได้

19

จำนวนเคสที่รอเดียงและระหว่างการส่งต่อ คลิกเพื่อดูข้อมูลในแต่ละเขต



จำนวนรวมของเคสที่รอเดียง/รอส่งต่อ ในรายที่.. จำนวนเคสที่รอเดียงและระหว่างส่งต่อ

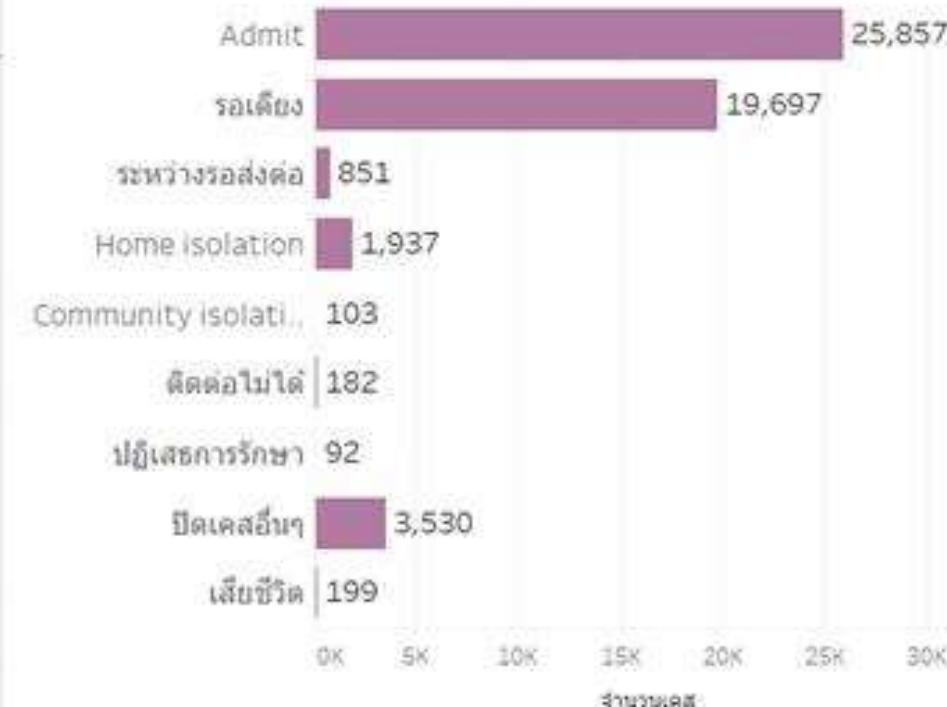
■ รอเดียง ■ เหลืออยู่ ■ เชี่ยว

0 1,000

วันเริ่มดำเนิน 7/7/2021 วันสิ้นสุด 22/7/2021

GBDI

จำนวนเคสแบ่งตามสถานการรักษา



จำนวนเคสที่รอเดียงและรอส่งต่อ แบ่งตามความรุนแรง





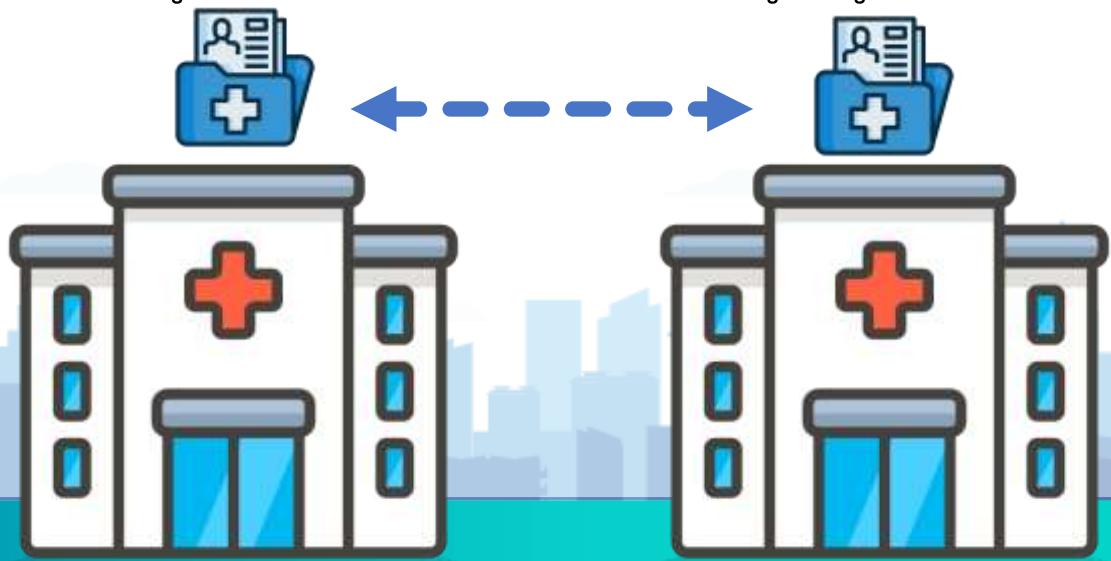
HEALTH LINK

The Health Information
Exchange System (HIE)



วัตถุประสงค์

1. เชื่อมโยงข้อมูลประวัติการรักษาข้ามโรงพยาบาล
(รพ. มหาวิทยาลัย, รพ. กระทรวงสาธารณสุข, รพ. ในสังกัดกองทัพและตำรวจ และ รพ. เอกชน)
2. ใช้ในการรักษาผู้ป่วยในการณีชุกเฉิน (ผู้ป่วยหมวดสติ)
3. เพื่อเพิ่มประสิทธิภาพการรักษาผู้ป่วยในการณีปกติ และการส่งต่อ (ผู้ป่วยให้ความยินยอมแก่แพทย์ดูข้อมูล)



អ្នកចូលរួម

ເລີນບັດປະຊາບ ທີ່ຈົດແລະບານຄຸກ

www.ສມັກ.ກອອສອບ
ໂຮງພາບາລຄີຣາຈ

Overview

Allergy

Vaccine

Dx

Diagnosis

Medication

Procedure

Lab

นาย ສາມຈາຍ ກົດສອບ
39 ປີ

Citizen ID: 1-2300-00000-00-0
Mobile: 9999999999
Birthday: 09 March 1982

Allergy (4/4)

Visit (4/4)

Reset All

Expand All

hepatitis B virus vaccine
Criticality: high

Dicloxacillin
Criticality: high

Penicillin G
Criticality: high

PENICILLINS1.0
Criticality: low

Visit Details

24/03/2021 ໂຮງພາບາລສຸກປາກ
OPD Blood

25/01/2021 ໂຮງພາບາລກົມພັດວັດຍ
IPD Cardiothoracic Surgery

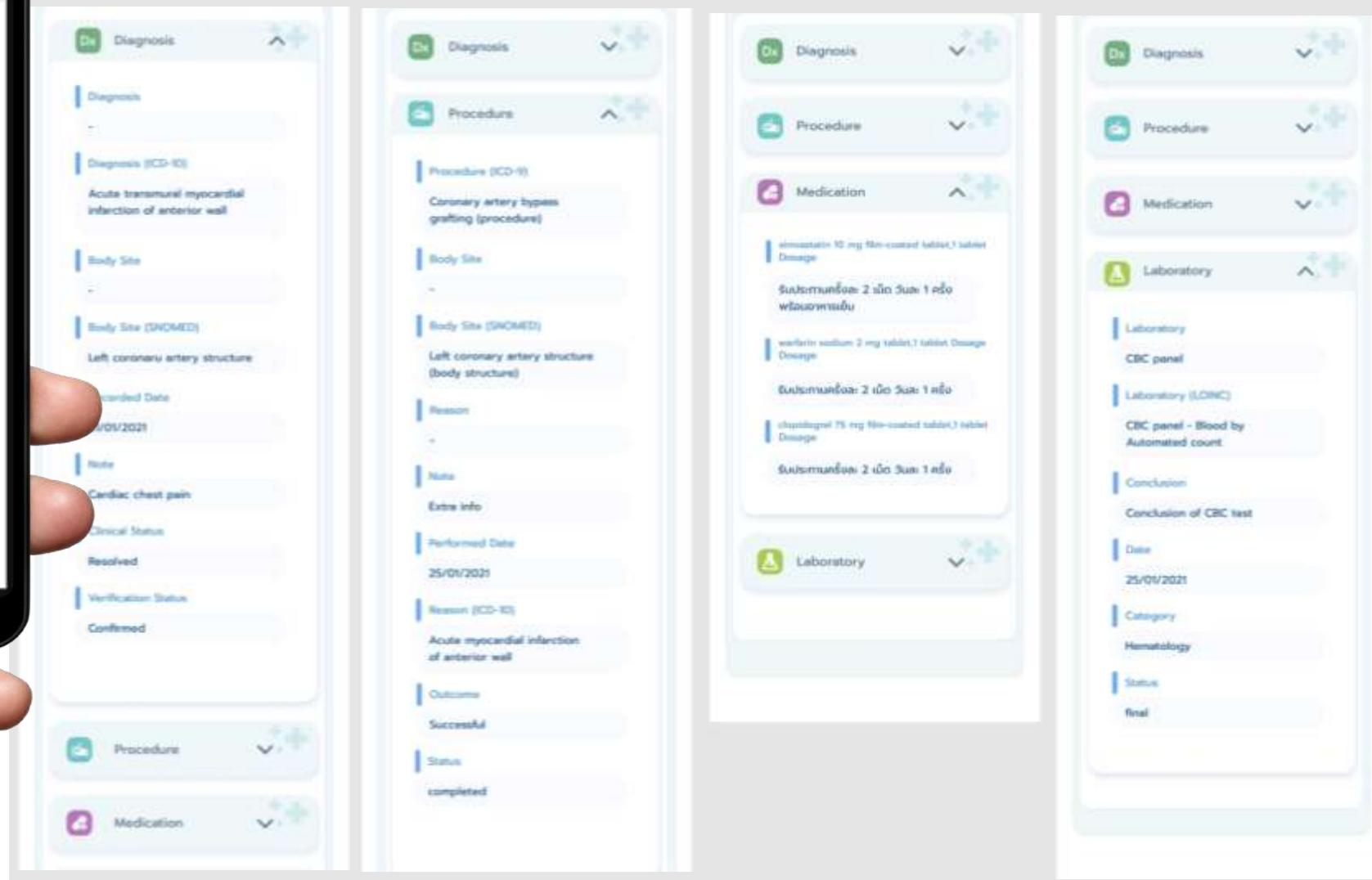
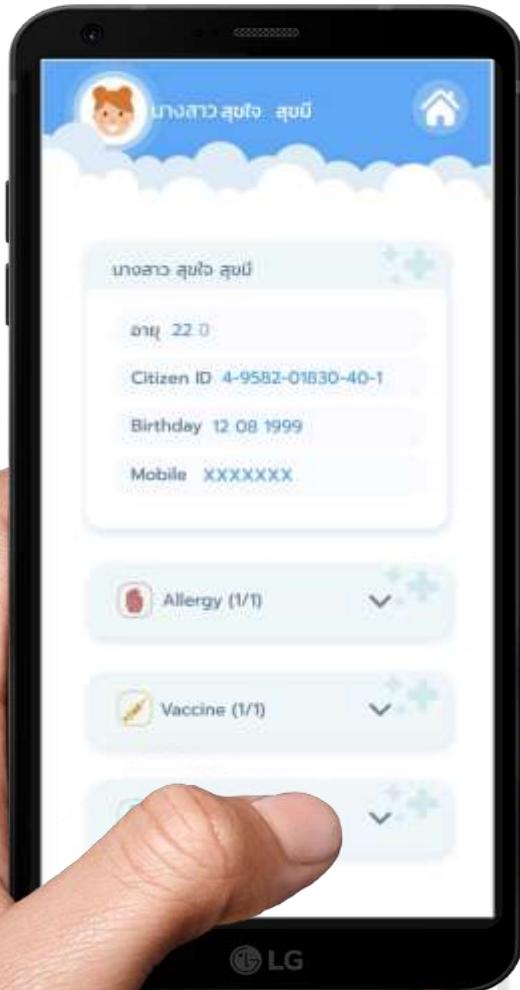
20/06/2020 ໂຮງພາບາລຈຸ່າລັດກຣນ
OPD ສູນເລືອຂອ່ວຍຕາງຝ່າ

12/11/2012 ໂຮງພາບາລກົມພັດວັດຍ
OPD Immunization

Influenza virus vaccine
Date: 12/11/2020

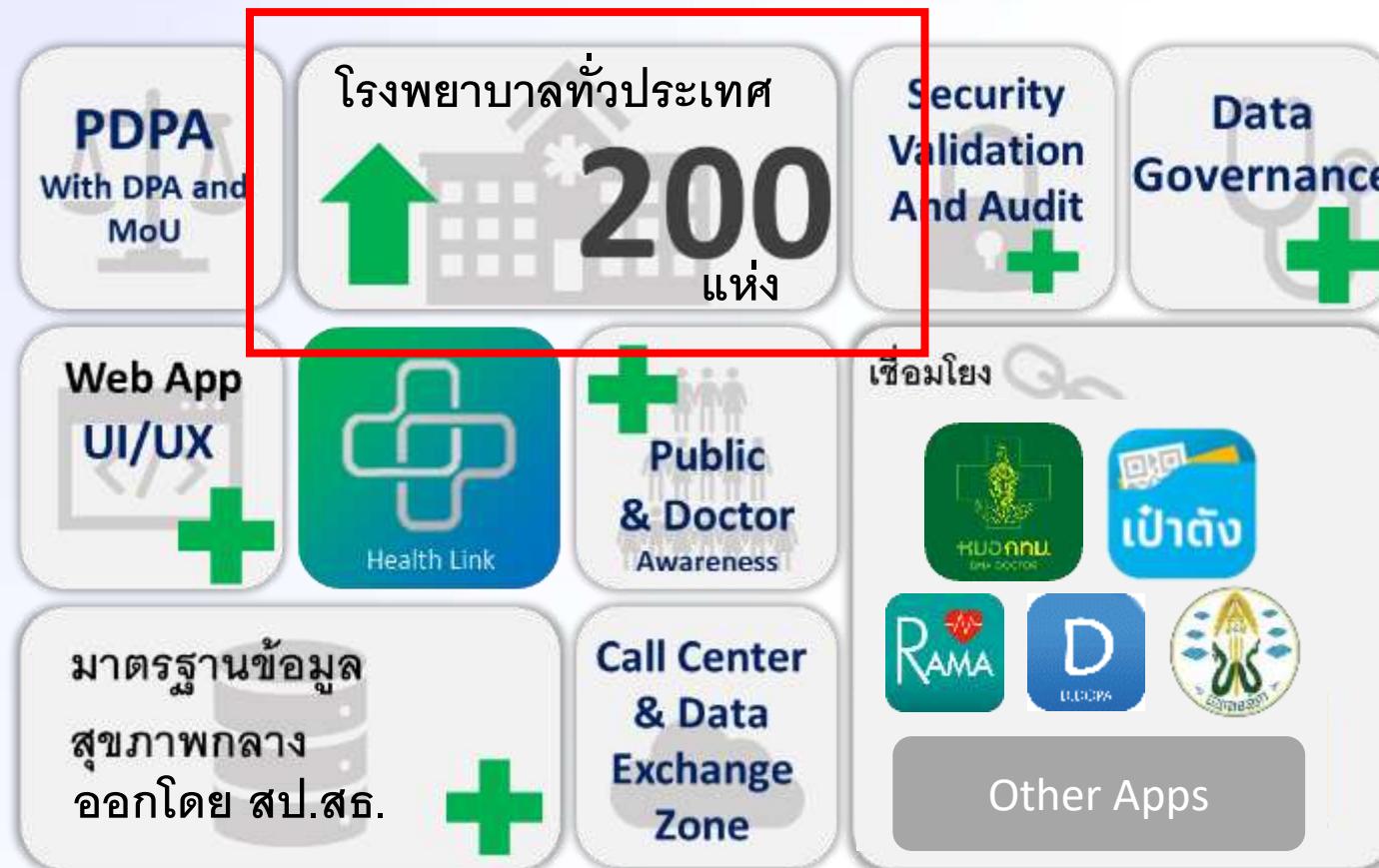


เตรียมข้อมูลประวัติการรักษาให้ ปชช.สามารถเข้าถึงได้





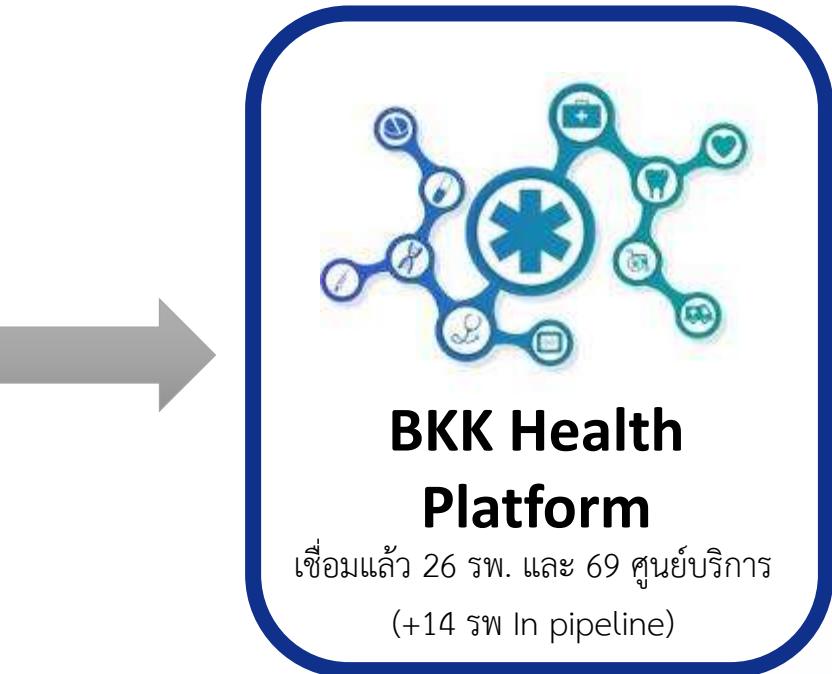
เชื่อมโยงข้อมูลประวัติการรักษา ของผู้ป่วยไว้ในที่เดียวกัน



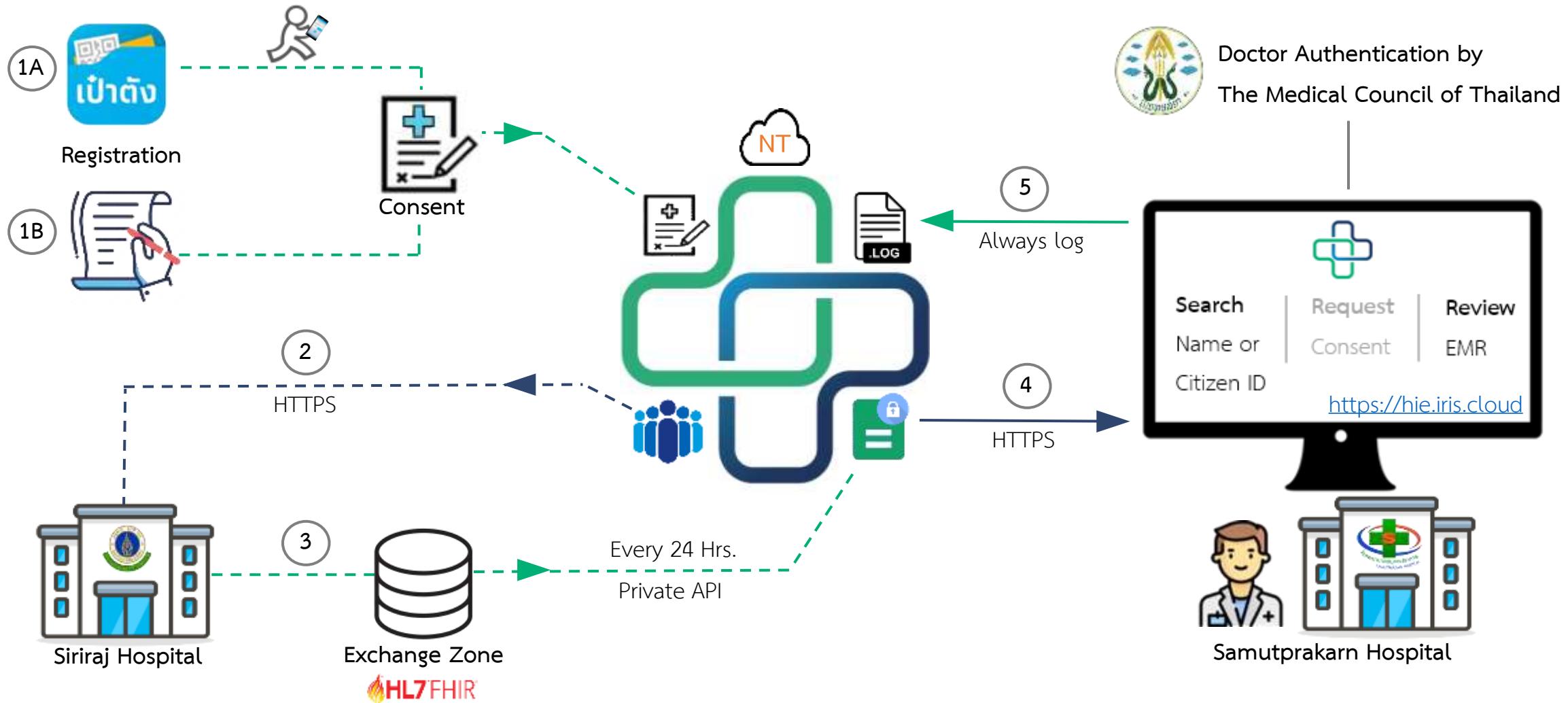
- มีการเชื่อมโยงข้อมูลประวัติการรักษาเสร็จสิ้นพร้อมใช้งาน **167 สถานพยาบาล**
 - ✓ สังกัด รพ ที่เข้าร่วมโครงการ Health Link: กลาโหม, อวน., กรมการแพทย์, กรมสุขภาพจิต, กทม., เอกชน, กระทรวงสาธารณสุข, หน่วยงานอิสระ
- มีสถานพยาบาลที่เพิ่มแพทย์เข้าสู่ระบบ Health Link และสามารถเปิดดูข้อมูลได้แล้ว **225 สถานพยาบาล**
- สถานพยาบาลที่แจ้งความจำนงเข้าร่วมและกำลังทยอยประสานงานและดำเนินการอีก **182 แห่ง** โดยตั้งเป้าจะดำเนินการให้แล้วเสร็จอีก **33 แห่ง** ในปี 2566 (ตามเป้าหมาย 200 สถานพยาบาล)



เชื่อมโยงข้อมูลประวัติการรักษา ของผู้ป่วยไว้ในที่เดียวกัน



รพ.สังกัด อวvn กรมการแพทย์ กลาโหม
สภาพาชัด กทม. เอกชน ฯลฯ ร่วมกันเปิดบริการ
ในกรุงเทพฯ (ปรับตาม UX ของโรงพยาบาลสังกัด
สำนักการแพทย์)

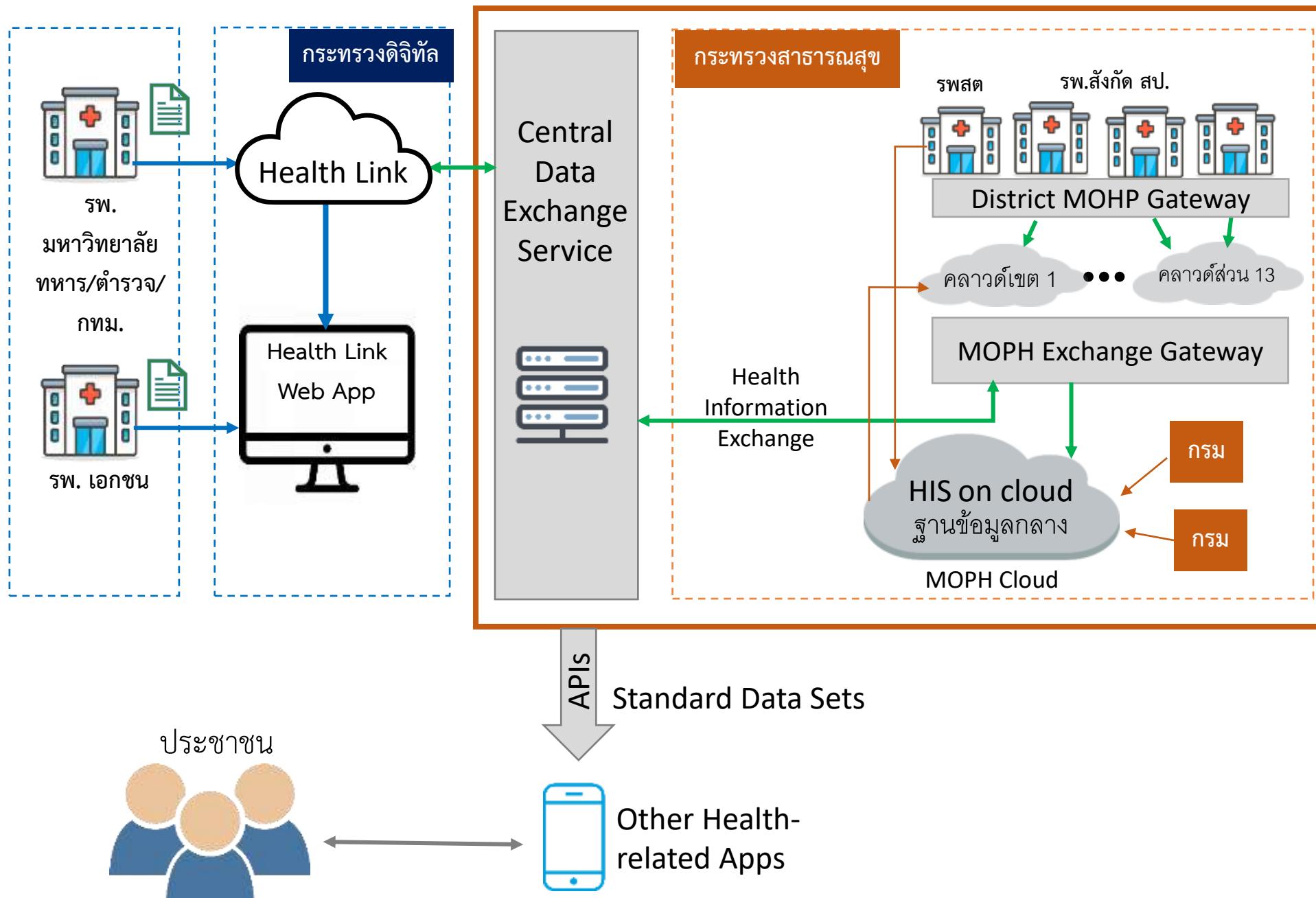


บุคคลกี่แลกเปลี่ยนในระบบ

ข้อมูลในระบบปัจจุบัน	ปัจจุบัน
ข้อมูลส่วนตัว	PERSON
ข้อมูลการแพ้ยา	DRUGALLERGY
ข้อมูลโรควินิจฉัย	DIAGNOSIS
ข้อมูลการจ่ายยา	DRUG
ข้อมูลวัคซีน	EPI
ข้อมูลหัตถการ	PROCEDURE
ข้อมูลผลตรวจห้องปฏิบัติ	LABFU
ข้อมูลการเข้ารับบริการ (ราย Visit)	SERVICE

ข้อมูลระหว่างการพัฒนา	ปัจจุบัน
ข้อมูลบุคคลที่ติดต่อได้กรณีฉุกเฉิน	
ข้อมูลโรคประจำตัว	
ข้อมูลอาการของผู้ป่วย	SERVICE
ข้อมูลการแพ้	DRUGALLERGY
ข้อมูลสัญญาณชีพ	SERVICE
ข้อมูลการตรวจทางพยาธิวิทยา	LABFU
ข้อมูลผลสรุปภาพถ่ายทางการแพทย์	

National Health Information Platform (Private Cloud)



สป.สธ. ออก Standard Data Set
สำหรับประวัติคนไข้เรียกว่า
“ข้อมูลปฐมภูมิ 1”

สดช. ดศ. ให้บริการ Cloud
Service บน
GDCC
Government Data Center and Cloud Service



TRAVEL LINK

Thai Tourism
Data Platform

Travel Link The Thai Tourism Data Platform

ขับเคลื่อนเศรษฐกิจการท่องเที่ยวด้วยข้อมูลเปิดภาครัฐ (Open Government Data)

- ชุดข้อมูลสถานที่ท่องเที่ยว (สป.ก.ท่องเที่ยว, ททท, กรมการท่องเที่ยว, อพท, ก.วัฒนธรรม, ฯลฯ)
- ชุดข้อมูลโรงแรม ที่พัก (กรมการปกครอง และ สำนักงานสถิติแห่งชาติ)
- ชุดข้อมูลการเดินทางโดยสารสาธารณะ
- ชุดข้อมูลพฤติกรรมนักท่องเที่ยว



Provided by
MoU 22 หน่วยงาน

- ชุดข้อมูลที่เกี่ยวข้องกับการท่องเที่ยว จากแหล่งต่าง ๆ เช่น Google, Reddit, Pantip.com, Social Media

Data Collection and Integration

มาตรฐานข้อมูล (Data Standard)

Data Catalog

Open Data Service APIs

AI Service APIs

- ระบบค้นหาสถานที่ท่องเที่ยว
- ระบบแนะนำสถานที่ท่องเที่ยว

Data Analytics Dashboards

Data Analytics Models

Data Sources

Data Integration and Analytics Services

หน่วยงานภาครัฐ: มีเครื่องมือสนับสนุนการขับเคลื่อน (Implement) นโยบาย



ภาคเอกชน: มีข้อมูลและระบบปัญญาประดิษฐ์ช่วยกระตุ้นธุรกิจ



ประชาชน: ได้รับความสะดวกสบายจากการใช้ข้อมูลและระบบปัญญาประดิษฐ์



* Developed by GBDi

Travel Link Website



- Data Catalog
- Data and AI service API
- Dashboard ข้อมูลท่องเที่ยว



3rd Party Applications

(หน่วยงานภายใต้ MOU, SMEs, Startups)

Presentations

Collaboration: Gathering government data from tourism supply chain to serve tourism sandbox

Inspiration



관광
관광
관광



CAAT
ສໍານັກນາມບົດບັນເຊີຍກົມປະກາດ
The Civil Aviation Authority of Thailand



Attraction



Tour

Hotel



Purchase



Safety Supporter



Catalog Service



RAW

Data Service



Dashboard Service

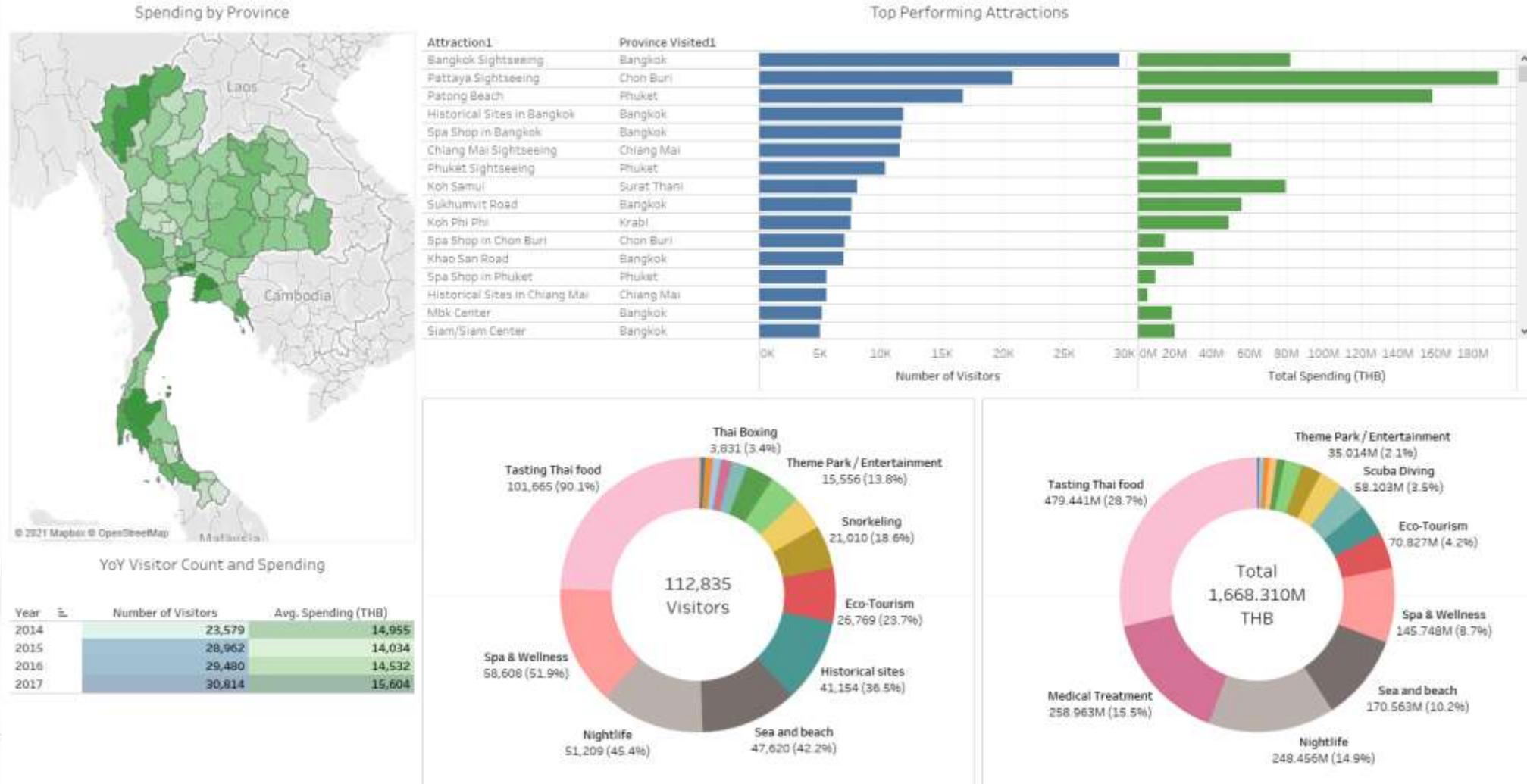


AI-based Service





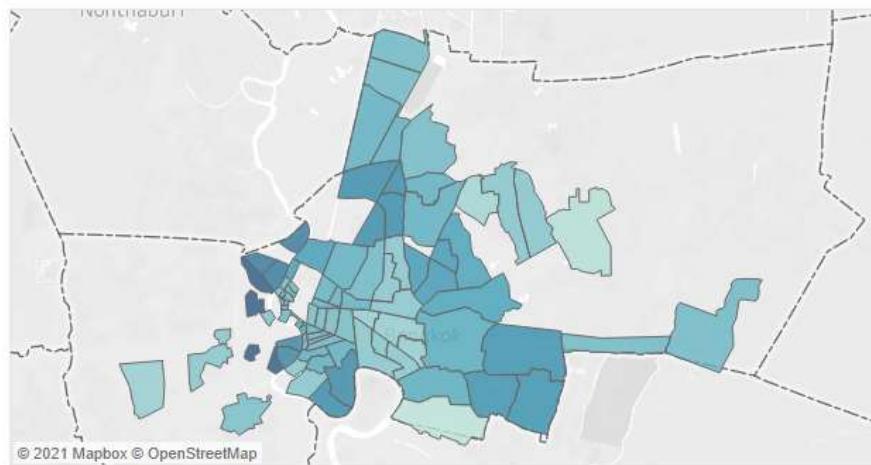
Dashboard : Tourist behaviors



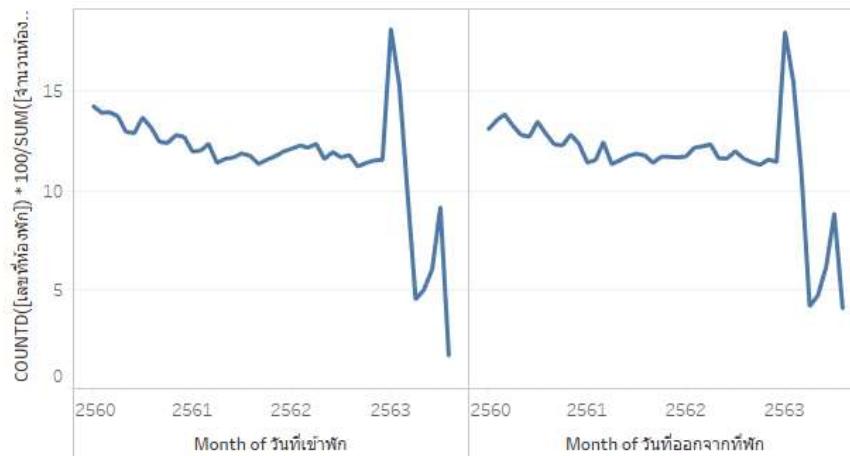


Dashboard : Hotel occupancy

Occupancy map



Occupancy Stats

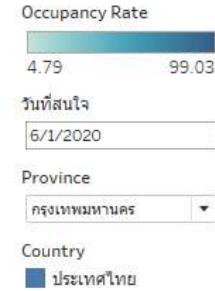


Overall Occupancy Rate

10.49 %

Hotels Involved

540 แห่ง

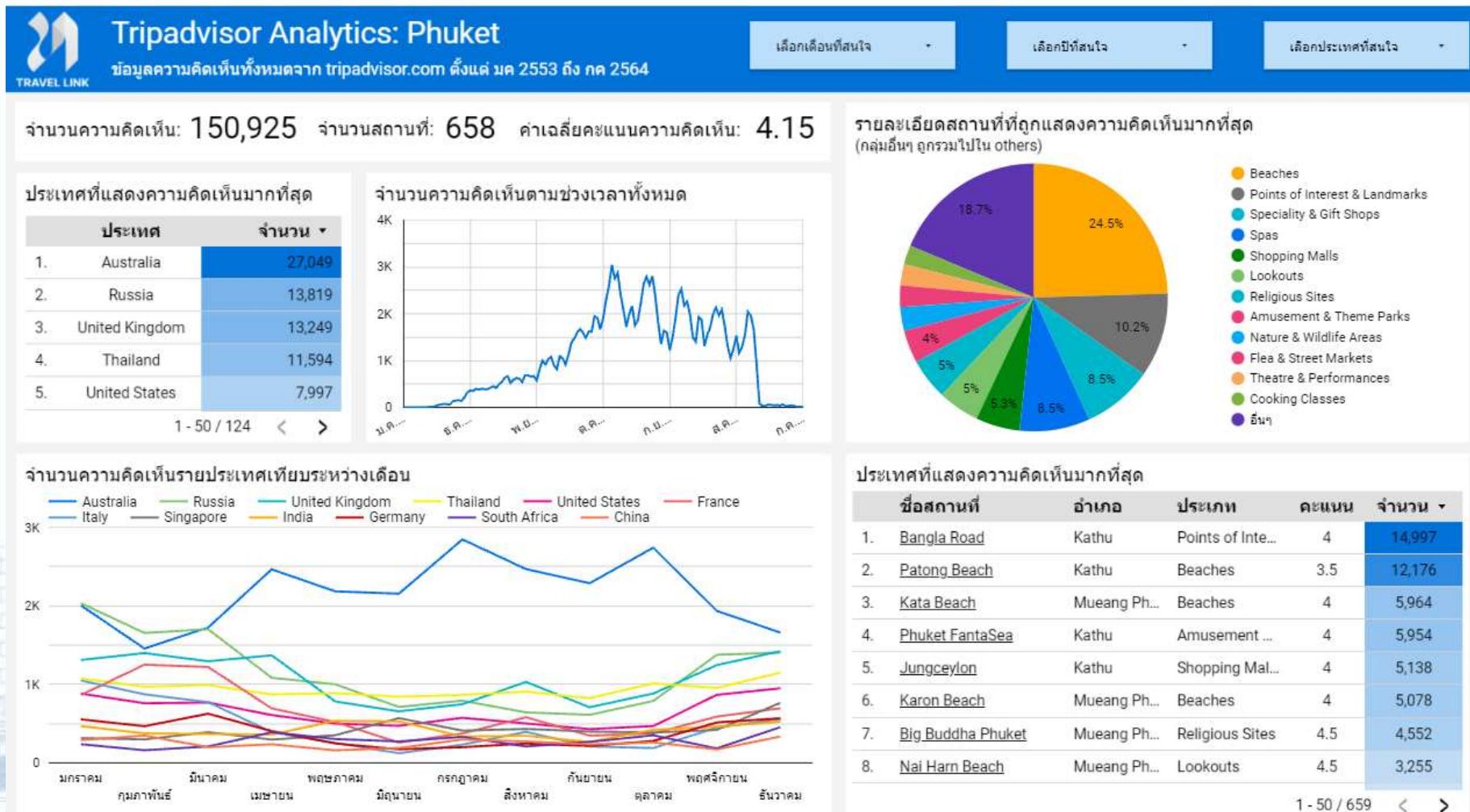


Occupancy By Hotel

ชื่อสถานที่ประกอบการ	OR	จำนวนห้อง
แกรนด์ เซ็นเตอร์ พอยต์ สุขุมวิท 55	56	246
สกาโน่โนม รีซิเดนซ์	56	30
โรงแรม "เพรส" ชีน่าฮอลล์ สเตชั่น กรุงเทพฯ	55	83
โรงแรมแอล พาราไดส์ กรุงเทพฯ	55	23
อนาร์ รีสอร์ฟเดนซ์ กรุงเทพฯ	55	70
แบงก์คอร์กไอกล็อกส์ สุขุมวิท	54	122
สีลตัน สุขุมวิท กรุงเทพฯ	54	152
เชอราตัน แกรนด์ สุขุมวิท	54	241
บลีสตัน สุวรรณ พาร์ค วิว	53	81
โรงแรม 48 วิลล์ ดอนเมือง แอร์พอร์ต	51	23
สวานา แบงคอก โซเทล	51	28
ศรีนิล สีลม	51	42
โรงแรม แอนปี กลาส ลอดจ์	50	5
โรงแรม เดอะพาร์ค โนนี สุวรรณภูมิ	50	60
โรงแรม เดอะ พาโนรัม	50	14
โรงแรม ชีรินา สาทร สวีท	50	4
โรงแรม ชัยญา	49	67
โรงแรมอคิด้า แทส สุขุมวิท	49	26
สลีโนว์สอร์ท	49	119
โนโวเทล กรุงเทพ พิมิเก็ช สีลม	49	105
บีวีเน็พลี	48	79
โรงแรม เอป้า เอ็กซ์	48	24
โรงแรม แกรนด์พรีเมี่ยม รีสอร์ท	48	95



Dashboard : Social media analytics



Youth Link

การเชื่อมโยงข้อมูลเด็ก
(เด็กปฐมวัย)



Youth Link The Data Platform on Thai Children

พัฒนาคุณภาพชีวิตเด็กด้วยการเชื่อมโยงข้อมูลข้ามหน่วยงาน

Data Sources

ข้อมูลรายบุคคล (ระบุเลขประจำตัวประชาชน)

- ชุดข้อมูลเด็กแรกเกิด
- ชุดข้อมูลสุขภาพเด็ก
- ชุดข้อมูลเด็กที่ประสบปัญหา/ต้องการความช่วยเหลือ
- ชุดข้อมูลเด็กในสถานศึกษา



ข้อมูลรายสติติที่เกี่ยวข้องกับตัวชี้วัดด้านการพัฒนาคุณภาพชีวิต สวัสดิการเด็ก

เด็กปฐมวัย



หน่วยงานภาครัฐ: มีเครื่องมือสนับสนุนการขับเคลื่อน (Implement) นโยบายการพัฒนาคน



ประชาชน: สามารถสืบค้นข้อมูลของตนเองเพื่อตรวจสอบสิทธิและสวัสดิการที่พึงได้รับจากรัฐ

หน่วยงานหลักที่เกี่ยวข้องกับการเชื่อมโยงข้อมูลเด็กปฐมวัย

- สำนักงานปลัดกระทรวงศึกษาธิการ
- สำนักงานปลัดกระทรวงสาธารณสุข
- กรมกิจการเด็กและเยาวชน (พม.)
- กรมส่งเสริมการปกครองท้องถิ่น (มหา.)
- กรมการปกครอง (มหา.)
- กรุงเทพมหานคร
- กรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ (พม.)
- ศูนย์อำนวยการบริหารจัดการจังหวัดชายแดนภาคใต้ (ศอ.บต.)

Youth Link The Data Platform on Thai Children

พัฒนาคุณภาพชีวิตเด็กด้วยการเชื่อมโยงข้อมูลข้ามหน่วยงาน

Data Sources

ข้อมูลรายบุคคล (ระบุเลขประจำตัวประชาชน)

- ชุดข้อมูลเด็กแรกเกิด
- ชุดข้อมูลสุขภาพเด็ก
- ชุดข้อมูลเด็กที่ประสบปัญหา/ต้องการความช่วยเหลือ
- ชุดข้อมูลเด็กในสถานศึกษา



ภายใต้ความร่วมมือหน่วยงานผู้ดูแลข้อมูล

ข้อมูลรายสถิติที่เกี่ยวข้องกับตัวชี้วัดด้านการพัฒนาคุณภาพชีวิต สวัสดิการเด็ก

Action Plan

ทบทวนคำนิยาม ว่าด้วยเด็กปฐมวัย

ทบทวน กฎหมาย ระเบียบ ข้อบังคับ ว่าด้วยพัฒนาการ การศึกษา สวัสดิการเด็ก

ระบบบัญชีข้อมูล (Data Catalog)

API Data Services

- สืบค้นข้อมูลรายบุคคล ผ่านระบบยืนยันตัวบุคคลเพื่อตรวจสอบสิทธิ์การเข้าถึงข้อมูล

แดชบอร์ดและโมเดลวิเคราะห์ข้อมูล
(Analytic Dashboards and Models)

เด็กปฐมวัย



หน่วยงานภาครัฐ: มีเครื่องมือสนับสนุนการขับเคลื่อน (Implement) นโยบายการพัฒนาคน



ประชาชน: สามารถสืบค้นข้อมูลของตนเองเพื่อตรวจสอบสิทธิ์และสวัสดิการที่พึงได้รับจากรัฐ

Expected Outcomes



เว็บไซต์บริการข้อมูล

- ระบบบัญชีข้อมูล
- API Data Services
- แดชบอร์ดข้อมูลเด็ก



แอปพลิเคชัน

ตรวจสอบ
สิทธิที่พึงได้รับ

แพลตฟอร์มบริการข้อมูลด้านเด็ก

Big Data Project

Project Inception

Problem Analysis

Exploratory Data Analysis

Predictive Analytic

Visualization Dashboard

Implementation & Deployment

Data Scientists และคณะทำงาน Big Data ของแต่ละกลุ่มงาน ร่วมกำหนดโจทย์ที่เหมาะสมและตั้งโครงการพัฒนาและทดสอบโมเดลคณิตศาสตร์นำร่องที่เหมาะสม

ทีม Data Scientists สำรวจข้อมูลที่มีอยู่ในปัจจุบันตามโจทย์นำร่องที่กำหนด เพื่อประเมินความพร้อมและแปลงจากความต้องการในเชิงปัญหาให้เป็นข้อกำหนดในเชิงข้อมูลและระบบ

สำรวจการกระจายตัวของข้อมูลเพื่อทำความเข้าใจข้อมูล และหาความสัมพันธ์ระหว่างตัวแปรในข้อมูล ในขั้นตอนนี้ทีมงานจะต้องการตัวอย่างข้อมูลจริง ระบุข้อมูลที่ต้องการเพิ่มเติมและเริ่มเตรียมข้อมูลจริง

ทีมงานนำข้อมูลที่จัดเตรียมในเบื้องต้นมาใช้ในการสร้างแบบจำลองหรือโมเดลทางคณิตศาสตร์เพื่อการนำร่อง โดยใช้เทคนิคและอัลกอริズึมต่างๆ และทดสอบความแม่นยำของโมเดลคณิตศาสตร์

ออกแบบวิธีการแสดงผลโดยเลือกมิติของข้อมูลที่เหมาะสมบน Interactive Dashboard เพื่อให้คณะทำงานทดลองใช้และสื่อสารกับทีมผู้บริหาร และผู้ปฏิบัติ ให้สามารถนำความเข้าใจดังกล่าวไปแปลงเป็นแผนการพัฒนาต่ออยอด

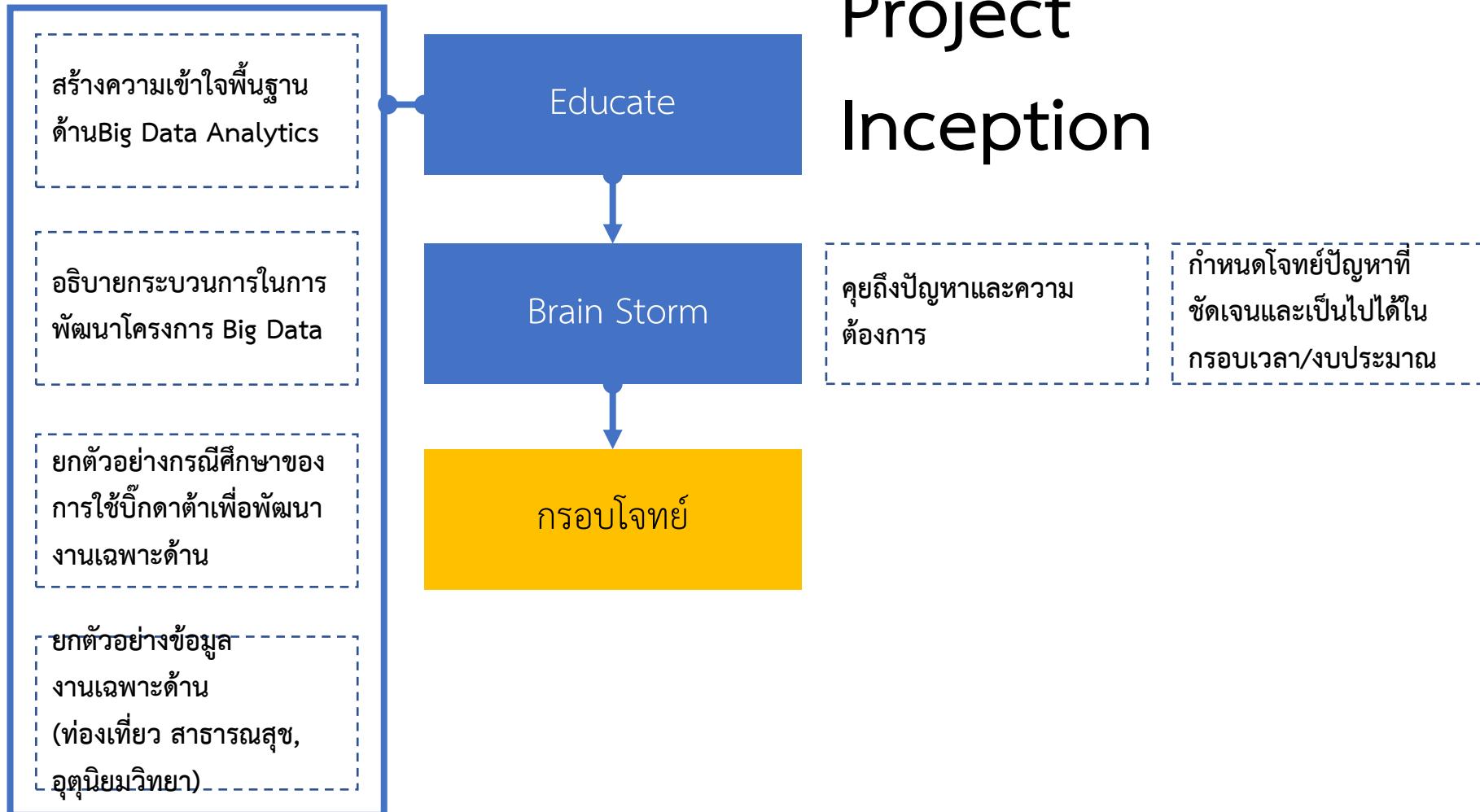
หลังจากผลลัพธ์เป็นที่พอใจแล้ว นักพัฒนาระบบเริ่มพัฒนาโปรแกรมตามรูปแบบของโมเดลคณิตศาสตร์ที่วางไว้ และตั้งค่าให้โปรแกรมให้ประมวลผล โมเดลแบบอัตโนมัติตามความถี่ที่วางแผนไว้ จากนั้นติดตั้งระบบซอฟต์แวร์เพื่อการใช้งานจริง

กลุ่มผู้ใช้ข้อมูล

กลุ่มผู้วิเคราะห์

กลุ่มผู้สร้างและพัฒนาระบบ

Project Inception





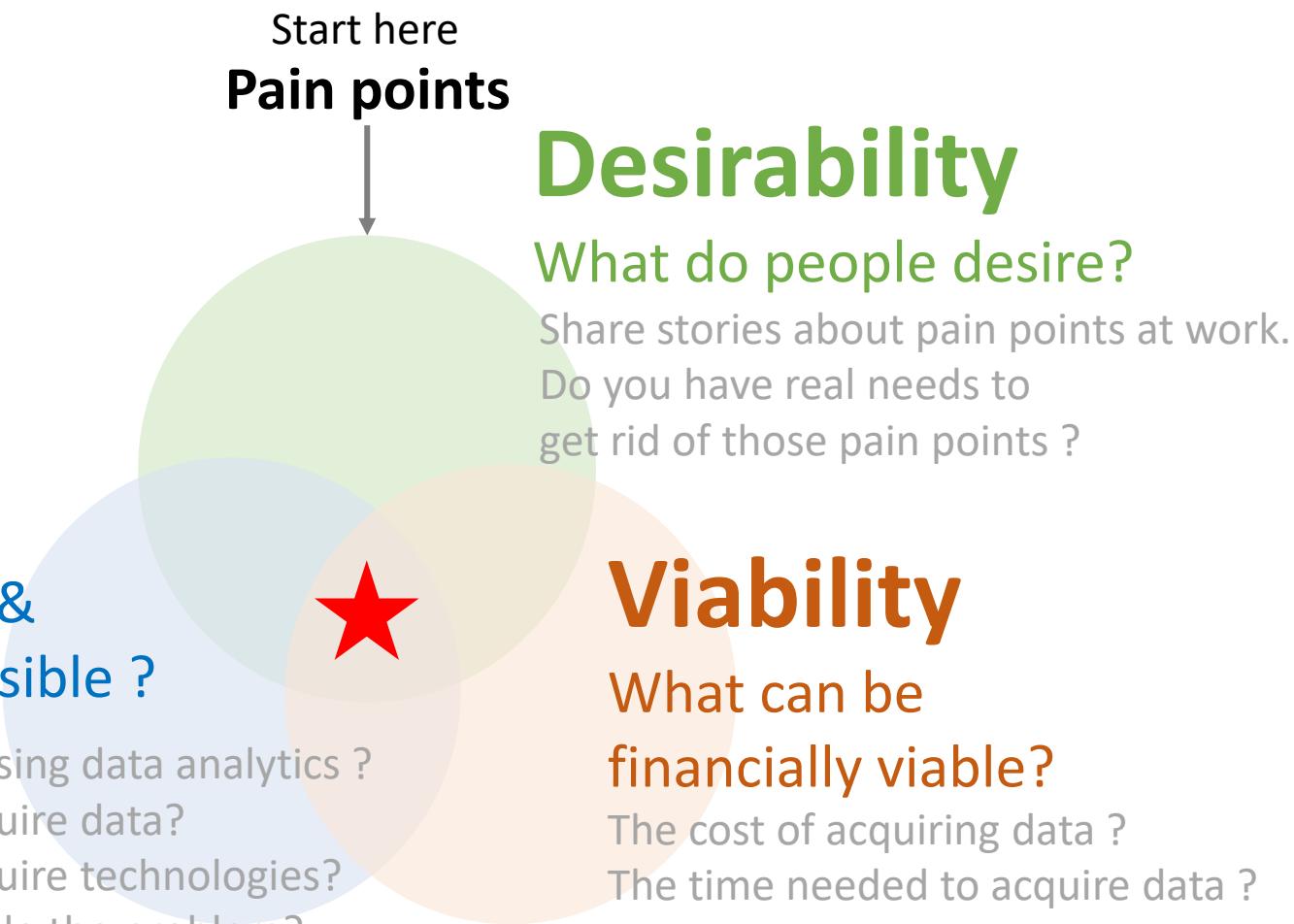
30 mins

Ref: Design Thinking from SEE Program at Babson College

Feasibility

What is technically & organizationally feasible ?

- Can those needs be met using data analytics ?
- Do we have or can we acquire data?
- Do we have or can we acquire technologies?
- Do we have people to tackle the problem?



*The analytic solutions that emerge should hit the overlap of these three lenses:
They need to be Desirable, Feasible and Viable*