

The background of the slide features a complex network diagram. It consists of numerous nodes of varying sizes, some solid (dark blue, light blue, grey) and some hollow (white with dark blue or light blue centers). These nodes are interconnected by a web of thin, light grey lines, creating a sense of connectivity and data flow. The overall aesthetic is clean and modern, typical of data science or technology presentations.

แนวคิดพื้นฐานของ วิทยาการข้อมูลเบื้องต้น

วิทยาการข้อมูลเบื้องต้น
(Fundamentals of Data Science)

ผศ.ดร.ธิติพร ชาญศิริวัฒน์



DATA SCIENCE
คืออะไร



องค์ประกอบของ
DATA SCIENCE



เครื่องมือที่ใช้ในงาน
DATA SCIENCE



ขั้นตอนการทำ
DATA SCIENCE



สายงานใน
DATA SCIENCE

Data Science คืออะไร

Data



- Facts and statistics collected together for reference or analysis.
- A set of values of subjects with respect to qualitative or quantitative variables.

Science



A systematic enterprise that builds and organizes knowledge in the form of testable explanations and predictions about the universe.



Data Science

Uses of scientific methods, processes, algorithms and systems to extract knowledge and insights from data.

Use Science to Understand Data

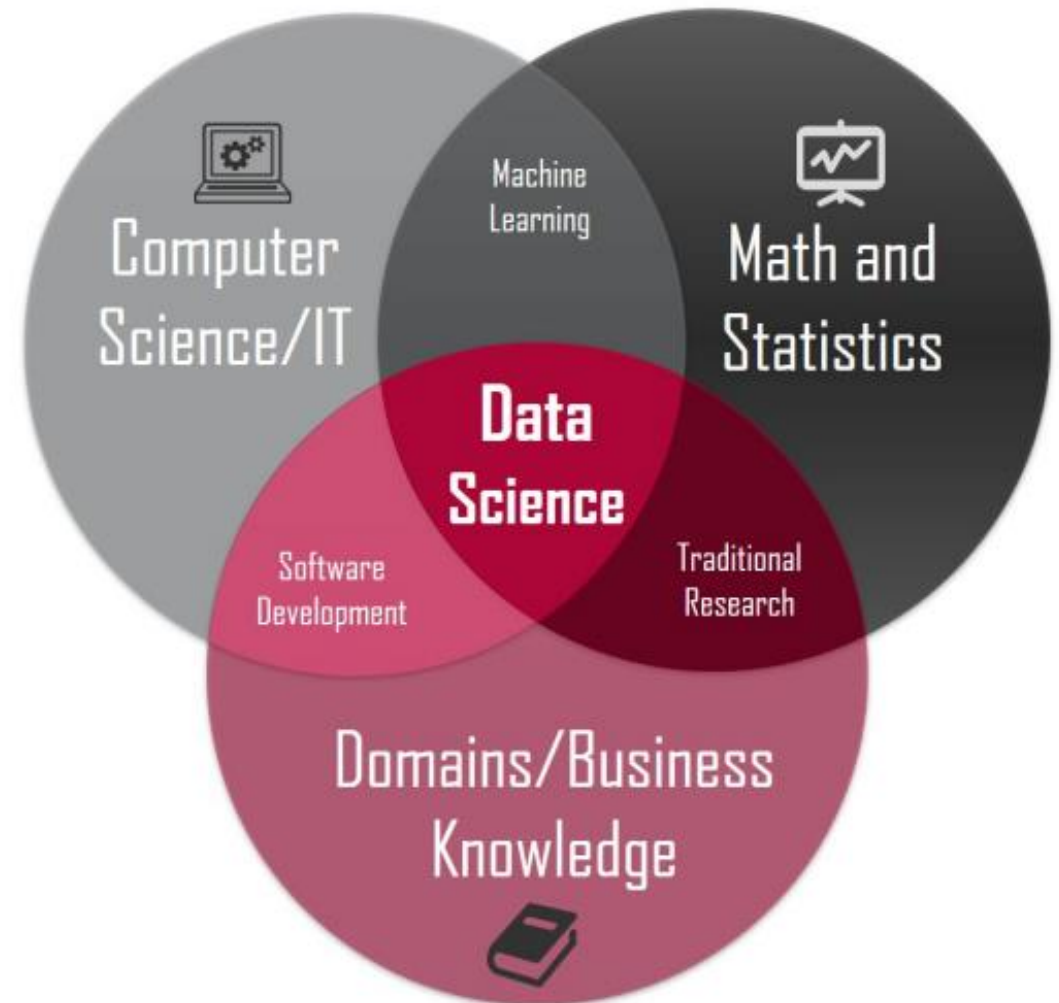
Data Science คืออะไร

- Data Science หรือ **วิทยาการข้อมูล** หรือ **วิทยาศาสตร์ข้อมูล** คือ ศาสตร์ที่ว่าด้วยการนำเอาข้อมูลมาใช้ให้เกิดประโยชน์ในงานด้านต่างๆ
- ศาสตร์ที่เกี่ยวกับการจัดการ จัดเก็บ รวบรวม ตรวจสอบ วิเคราะห์ วิจัยข้อมูล และนำเสนอผลการวิเคราะห์
- สร้างขึ้นร่วมกันระหว่างความรู้ทางด้านเทคโนโลยี สถิติ และธุรกิจ เพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ หรือ Big Data
- ประโยชน์ในงานด้านต่างๆ เช่น ธุรกิจ โลจิสติกส์ เศรษฐศาสตร์ การเงิน การแพทย์ วิศวกรรม สารสนเทศ สาธารณสุข เป็นต้น

องค์ประกอบของ Data Science

■ is a multi-disciplinary field comt

- Computer Science
- Mathematics and Statistics
- Business Knowledge



AI + Big Data + Data Science

Big Data

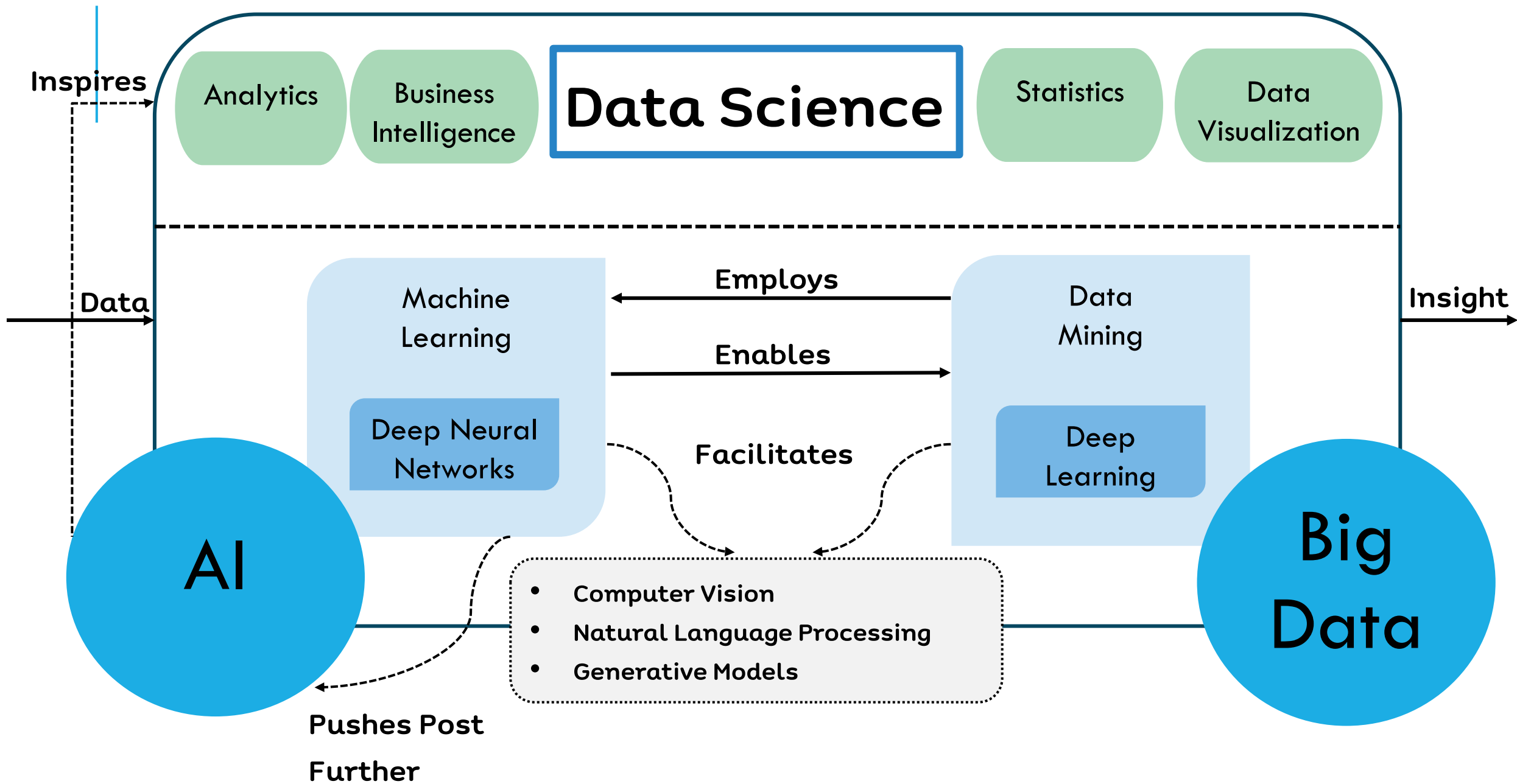
ข้อมูลที่มีโครงสร้าง และไม่มีโครงสร้างจำนวนมากจากแหล่งที่มาต่างๆ เช่น Social Media, Sensor, อุปกรณ์ สำหรับใช้เป็นวัตถุดิบสำหรับ Data Science

Data Science

วิทยาการด้านข้อมูลที่เกี่ยวข้องกับการใช้วิธีการทางสถิติและการคำนวณ เพื่อสกัดข้อมูลเชิงลึกและความรู้จาก Big Data

AI

เป็นโครงสร้าง Machine/Computer อัจฉริยะที่เลียนแบบพฤติกรรม/การทำงานของมนุษย์ สร้างกระบวนการเรียนรู้เพื่อให้สามารถทำงานแทนมนุษย์ได้



ชนิดของข้อมูล

■ Structured Data

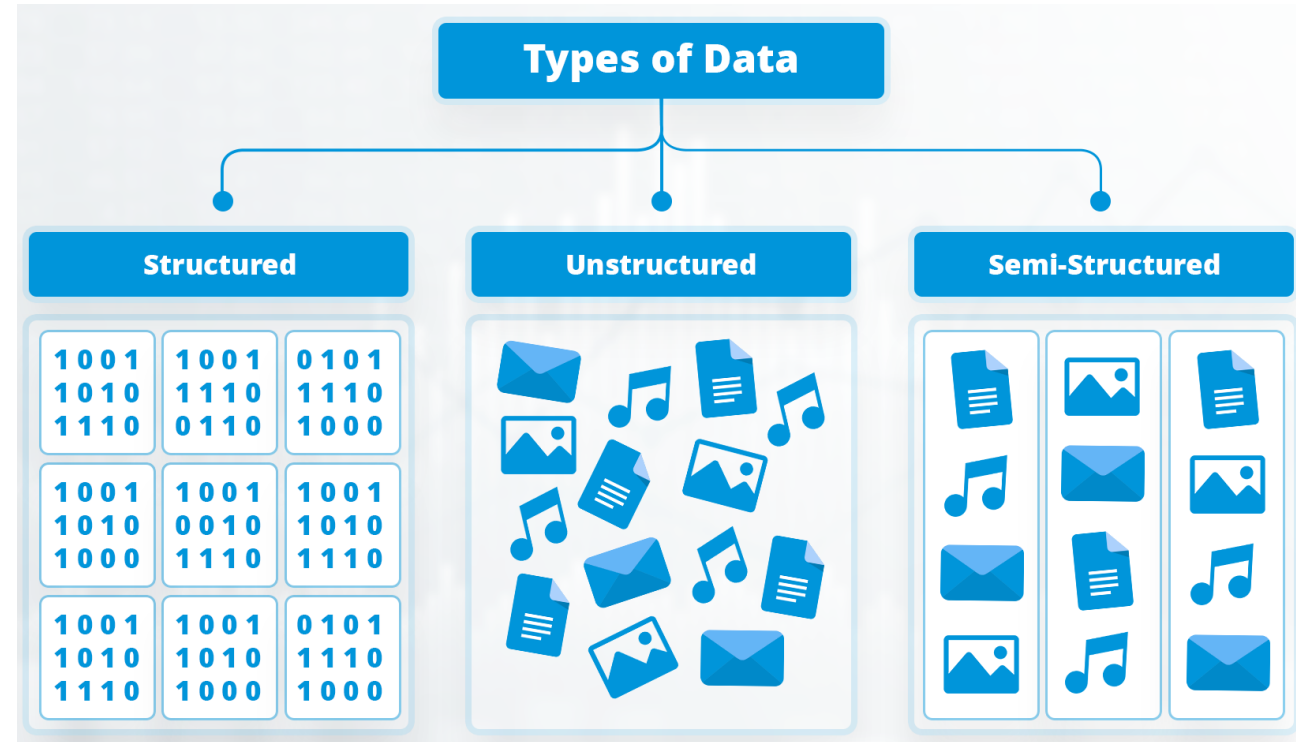
- ข้อมูลที่เก็บอยู่ในรูปแบบของตาราง (Table) ซึ่งประกอบด้วยแถว (Row) และคอลัมน์ (Column)

■ Unstructured Data

- เช่น ข้อความใน Website, Twitter, FB ไฟล์ รูปภาพ เสียง วิดีโอ

■ Semi-Structured Data

- ข้อมูลที่มีโครงสร้างข้อมูลที่มีความยืดหยุ่น สามารถขยายโครงสร้างข้อมูลได้ในอนาคต และเรียกใช้ได้รวดเร็ว



เครื่องมือที่ใช้ในงาน Data Science



ขั้นตอนการทำData Science

1

Business Problem

Business Questions



Understand



Define Objectives

ทำความเข้าใจปัญหา และแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูล และวางแผนการดำเนินการคร่าวๆ

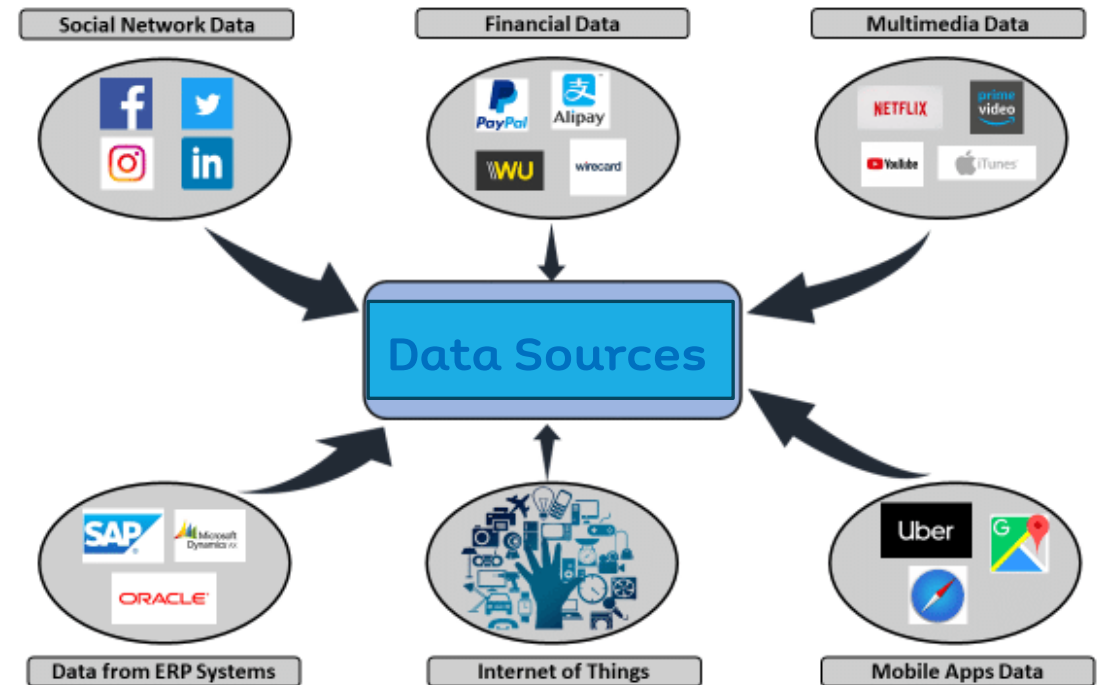
ขั้นตอนการทำ Data Science

2

Data Acquisition

Data Sources

- Questionnaires
- Web servers
- Web services (API)
- Database
- Logs
- Online repositories



การเก็บรวบรวมข้อมูล

ขั้นตอนการทำData Science

3

Data Preparation

Data Selection



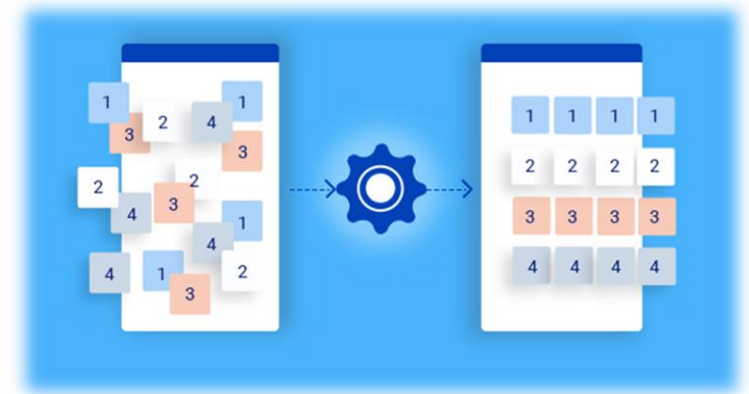
เลือกข้อมูล
ตัวแปรที่จะใช้งาน

Data Cleaning



หาสิ่งผิดปกติของข้อมูล
noise, missing value,
outliner

Data Transformation



แปลงข้อมูลให้อยู่ใน
รูปแบบที่พร้อมนำไปใช้ใน
การวิเคราะห์

ขั้นตอนการทำData Science

4

Exploratory Data Analysis

name	id	align	eye	hair	gender	alive	appearances	first_appear	publisher
Spider-Man (Peter Parker)	Secret	Good	Hazel Eyes	Brown Hair	Male	Living Characters	4043	Aug-62	marvel
Captain America (Steven Rogers)	Public	Good	Blue Eyes	White Hair	Male	Living Characters	3360	Mar-41	marvel
...
Natalia Romanova (Earth-616)	Public	Good	Green Eyes	Red Hair	Female	Living Characters	1050	Apr-64	marvel

Selection of feature variable that will be used in the model development

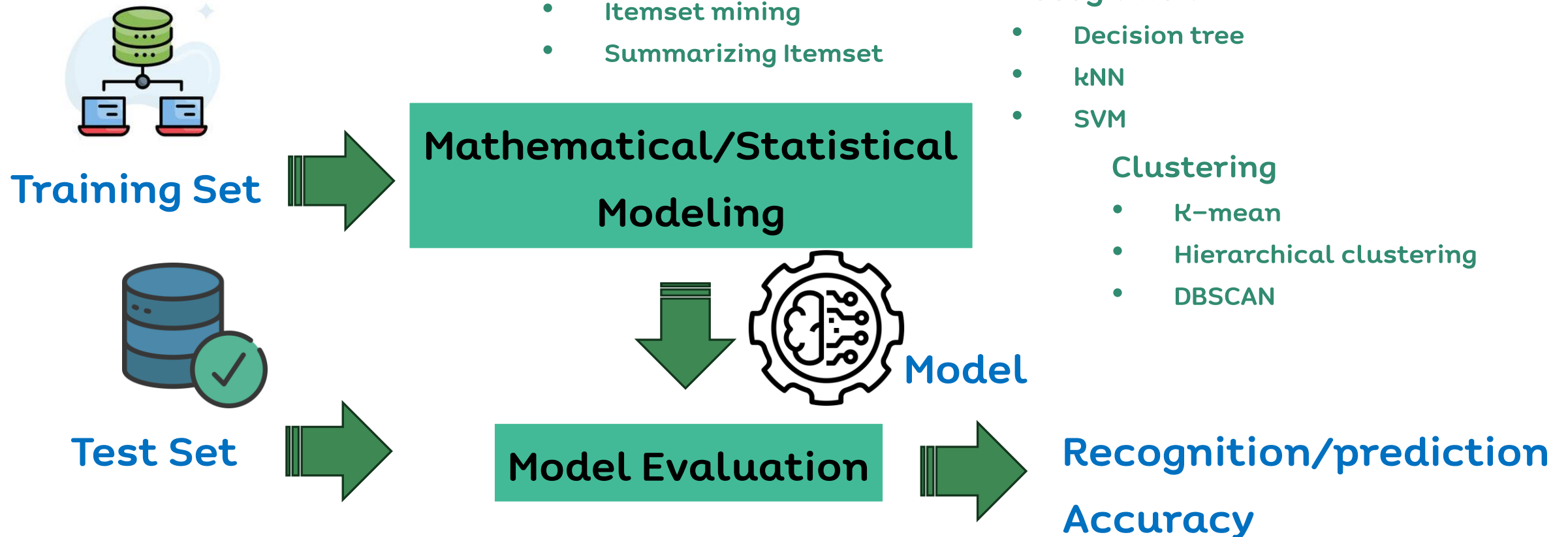
?

สำรวจ และสรุปสิ่งที่อยู่ในข้อมูล

ขั้นตอนการทำData Science

5

Data Modeling

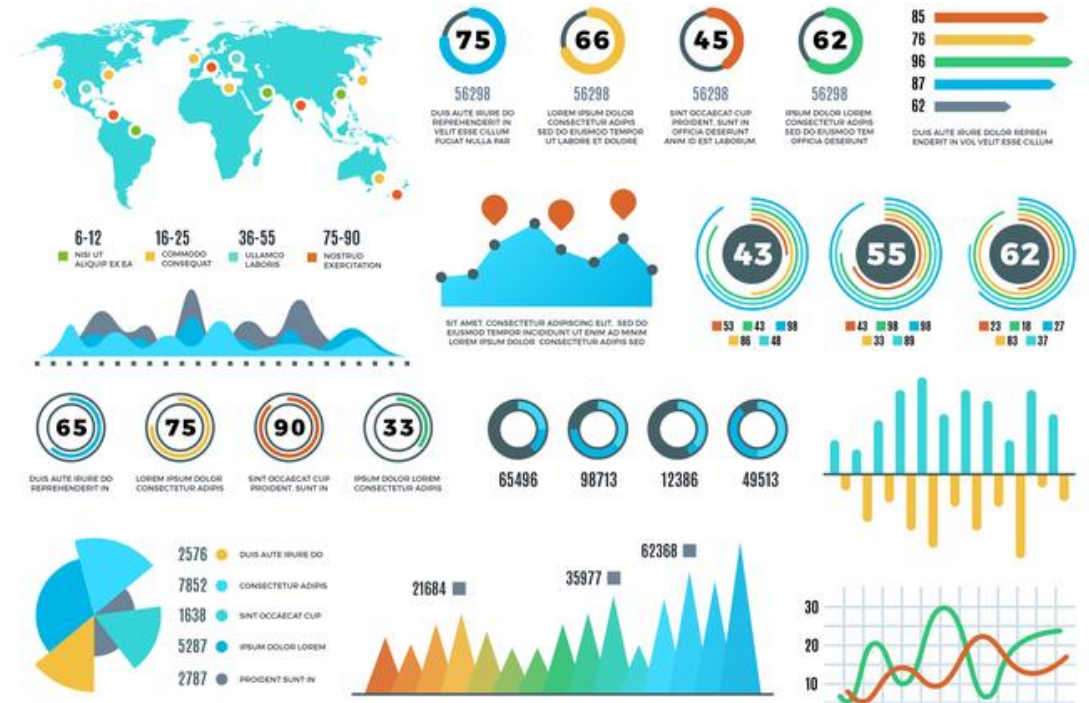


ขั้นตอนการทำData Science

6

Visualization & Communication

- Create powerful reports and dashboards
- Communicate business finding to convince the stakeholders



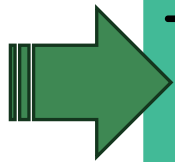
ขั้นตอนการทำData Science

7

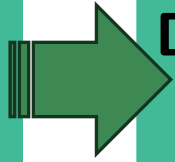
Deploy & Maintenance



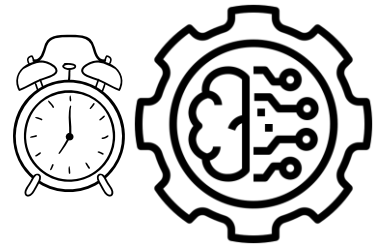
Model



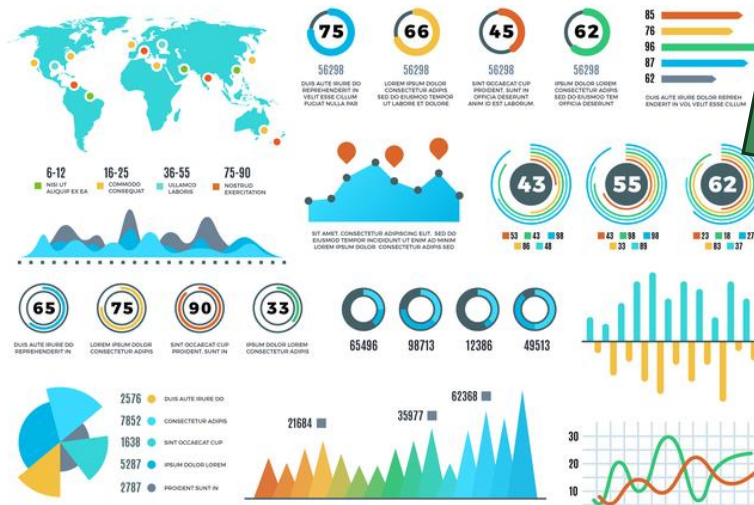
Test in pre-production environment



Deploy in production environment



Running Model



Real-time Analytics

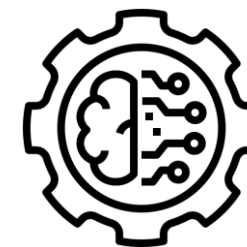
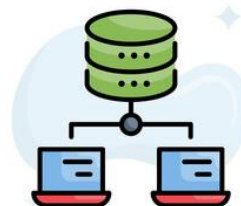
Train/Build Model

ข้อมูลในอดีต

Age	Gender	Income	Class
25	Male	70K	Yes
28	Male	20K	No
..			..
47	Female	100K	Yes



Data Science



Model

Deployment

Age	Gender	Income	Class
37	Female	65K	?

ข้อมูลใหม่ที่ไม่ทราบ Class



Model

Deployment



Class = Yes
ความน่าจะเป็น 0.9

สายงานใน Data Science

Top 10 วงการยุคใหม่ ที่ให้ความสำคัญกับการนำ Data Analytics มาใช้



**Banking and
Securities**



Healthcare



Manufacturing



**Consumer
Trade**



Energy



**Communications
& Media**



Education



Insurance



Transportation



Sports

ที่มา <https://marutitech.com/big-data-analytics-need-business/>

สายงานใน Data Science

Data Analyst

นักวิเคราะห์ข้อมูล

ทำหน้าที่ในการนำ Big Data มาวิเคราะห์เพื่อแก้ไขปัญหาที่องค์กรกำลังเผชิญหรือวิเคราะห์ประสิทธิภาพของผลิตภัณฑ์และการให้บริการ

- ทักษะการใช้เครื่องมือ
- ทักษะการคิดวิเคราะห์และความเข้าใจธุรกิจ
- ทักษะการนำเสนอ

Data Scientist

นักวิทยาศาสตร์ข้อมูล

ทำหน้าที่ในการวิเคราะห์ข้อมูลและสร้างเป็นโมเดลเพื่อทำนายแนวโน้มและกำหนดทิศทางธุรกิจในอนาคต

- ทักษะทางคณิตศาสตร์และสถิติ
- ทักษะการเขียนโปรแกรม

































Data Engineer

วิศวกรข้อมูล

ทำหน้าที่ในการรวบรวมข้อมูลจากหลากหลายแหล่ง ไปเก็บไว้ใน Database, Data Warehouse หรือ Data Lake เพื่ออำนวยความสะดวกให้กับผู้ที่จะนำข้อมูลไปใช้งาน

- ทักษะด้าน Database
- ทักษะการใช้แพลตฟอร์ม

สายงานและทักษะที่เกี่ยวข้องกับ Data Science

ทักษะที่เกี่ยวข้อง	DATA ANALYST	MACHINE LEARNING ENGINEER	DATA ENGINEER	DATA SCIENTIST
ใช้เครื่องมือโปรแกรมมิ่งเป็น				
แปลงข้อมูลเป็นภาพและสื่อสารให้เข้าใจง่าย				
เข้าใจข้อมูลได้เร็วและถูกต้อง				
สถิติและคณิตศาสตร์ที่เกี่ยวข้อง				
การจัดการข้อมูลให้มีคุณภาพ				
เข้าใจ Machine Learning				
มีพื้นฐานด้านวิศวกรรมซอฟต์แวร์				
เข้าใจแคลคูลัสของฟังก์ชันหลายตัวแปร และพีชคณิตเชิงเส้น				



สำคัญน้อย



สำคัญ



สำคัญมาก

แหล่งที่มา: <https://blog.udacity.com/2014/11/data-science-job-skills.html>

ตัวอย่างการใช้งาน Data Science

Netflix

แพลตฟอร์มวิดีโอสตรีมมิง ที่มีข้อมูลของผู้ใช้อยู่ในมือ และใช้ Big Data ในการดำเนินธุรกิจ กำหนดรูปแบบ วิดีโอ ไม่ว่าจะเป็นทั้งภาพยนตร์ หรือซีรีส์ได้ตรงกับ กลุ่มลูกค้า

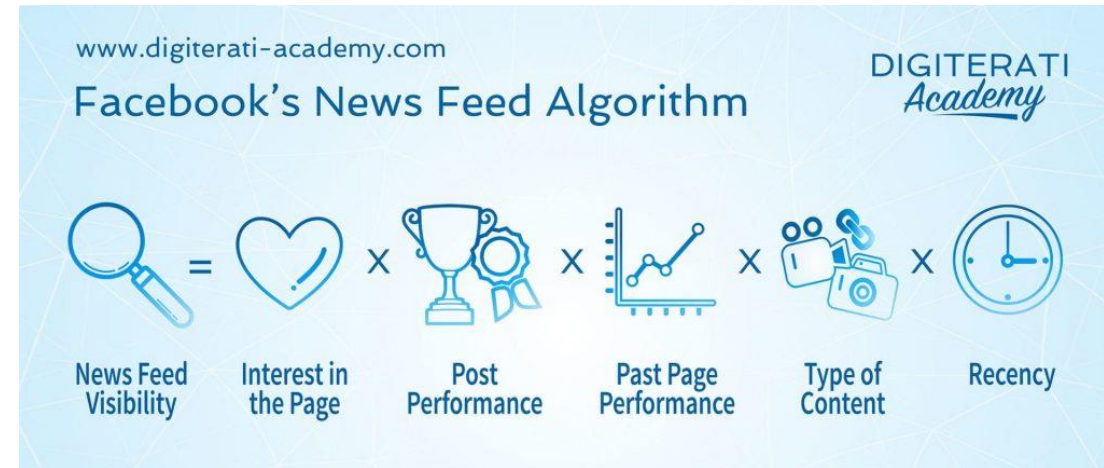
โดยนำข้อมูลอายุผู้ใช้ รูปแบบวิดีโอที่ชอบรับชม ระยะเวลาในการรับชม และอื่นๆ มาวิเคราะห์ พฤติกรรมผู้ชมทั่วโลก รวมถึงการไปซื้อลิขสิทธิ์ ภาพยนตร์หรือซีรีส์ที่ตอบโจทยความต้องการของลูกค้า



ตัวอย่างการใช้งาน Data Science

Facebook

ผู้นำด้านโซเชียลมีเดียของโลก ที่ได้ใช้เทคนิคขั้นสูงในศาสตร์ของ Data Science เพื่อศึกษาพฤติกรรมของผู้ใช้ และหาข้อมูลเชิงลึก ด้วย Deep learning

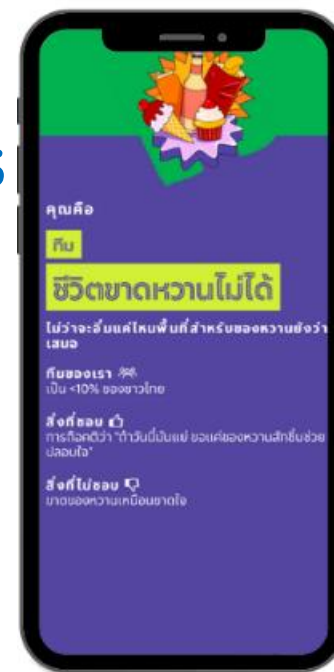


Fb ใช้การจดจำใบหน้าและการวิเคราะห์ข้อความ ในการจดจำใบหน้า ใช้เครือข่ายประสาทเทียมที่ทรงพลังเพื่อจำแนกใบหน้าในรูปถ่าย ใช้เอ็นจินการเข้าใจข้อความของตนเองที่เรียกว่า "DeepText" เพื่อทำความเข้าใจประโยคของผู้ใช้ ใช้การเรียนรู้เชิงลึกสำหรับโฆษณาที่ตรงเป้าหมาย

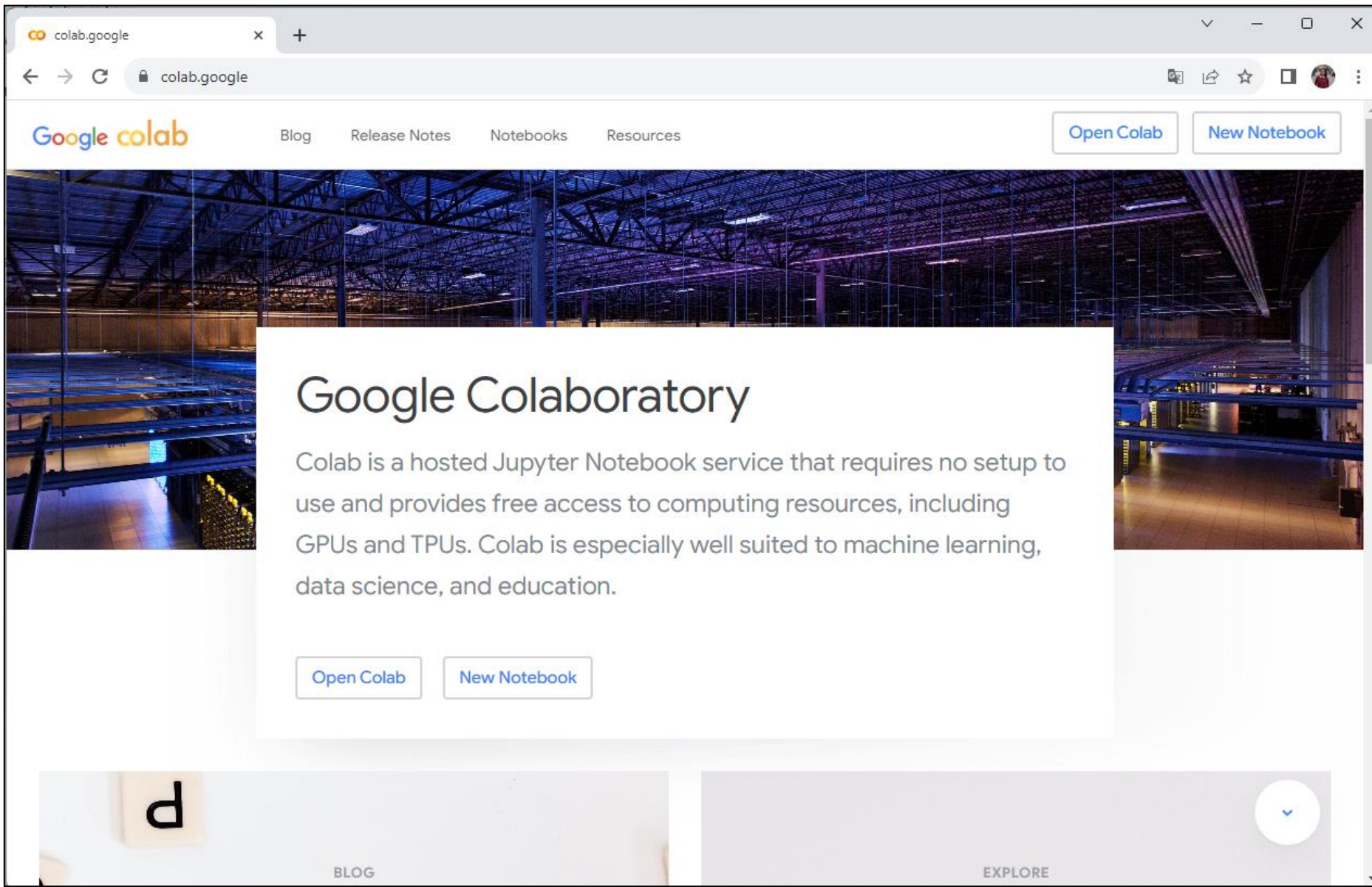
ตัวอย่างการใช้งาน Data Science

Grab Food Thailand

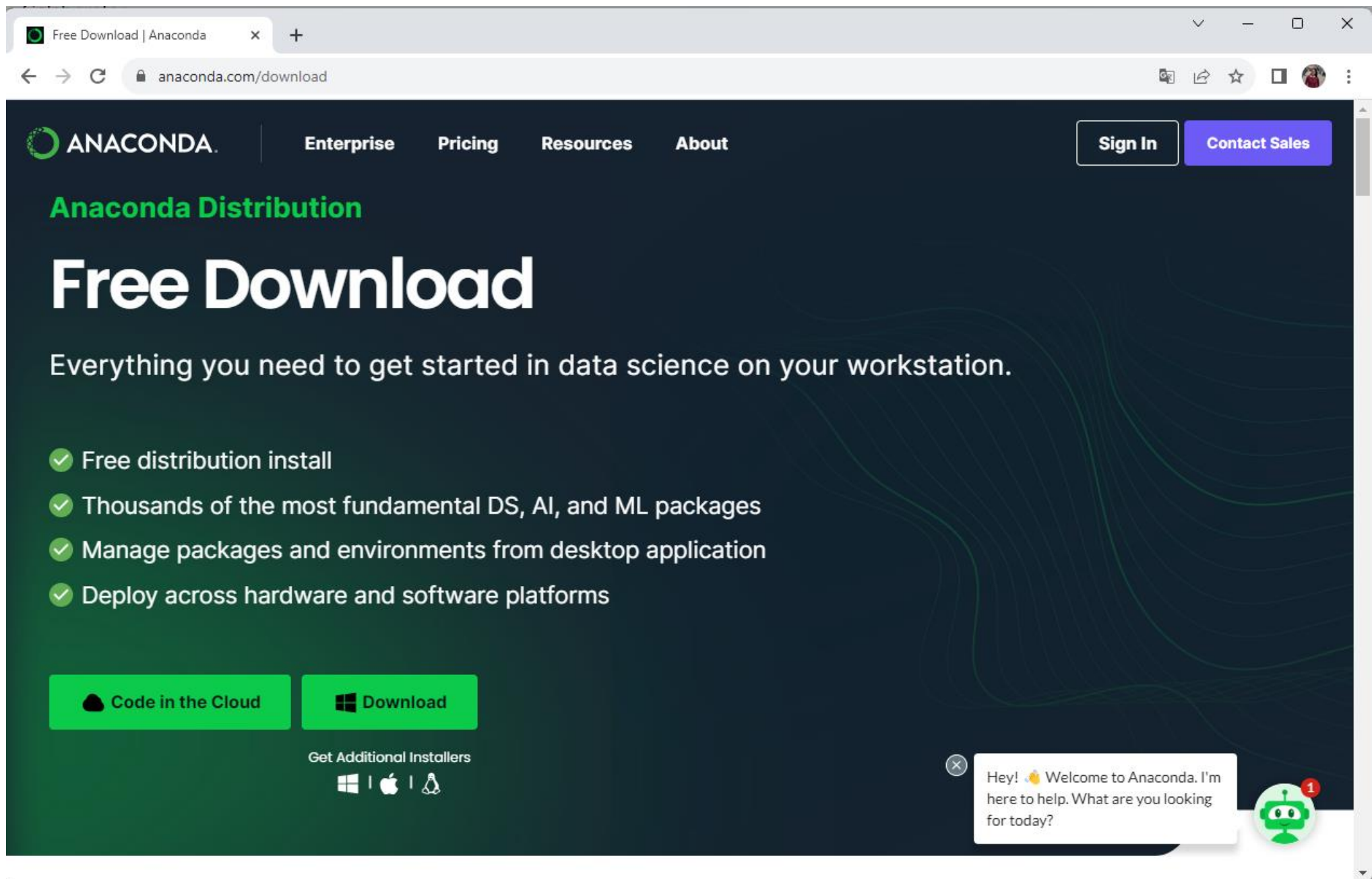
Grab บ้านเราก็มีการใช้กลยุทธ์ Data-Driven Marketing เช่นกัน อย่างแคมเปญช่วงต้นปี 2022 ที่ผ่านมาก็ได้มีการนำข้อมูลส่วนตัวของ User มาสรุปให้ผู้ใช้งานได้ย้อนดูอย่างสนุกสนานว่าปีนี้ตัวเองสั่งอะไรกินบ้าง เป็นสายไหน และปิดท้ายด้วยการเสนอ Code ส่วนลดร้านอาหารประเภทนั้นๆ ให้ลูกค้าได้นำไปใช้ต่อ



ติดตั้งใช้งานโปรแกรม



<https://colab.google/>



The screenshot shows the Anaconda website's download page. The browser's address bar displays 'anaconda.com/download'. The page features a dark blue background with a green wavy pattern on the right. At the top, there is a navigation bar with the Anaconda logo, links for 'Enterprise', 'Pricing', 'Resources', and 'About', and buttons for 'Sign In' and 'Contact Sales'. The main heading is 'Anaconda Distribution' in green, followed by 'Free Download' in large white text. Below this, a subheading reads 'Everything you need to get started in data science on your workstation.' A list of four benefits is provided, each with a green checkmark: 'Free distribution install', 'Thousands of the most fundamental DS, AI, and ML packages', 'Manage packages and environments from desktop application', and 'Deploy across hardware and software platforms'. Two large green buttons are present: 'Code in the Cloud' with a cloud icon and 'Download' with a Windows icon. Below the 'Download' button, it says 'Get Additional Installers' with icons for Windows, macOS, and Linux. A chatbot bubble in the bottom right corner says 'Hey! Welcome to Anaconda. I'm here to help. What are you looking for today?' with a red notification badge showing '1'.

Free Download | Anaconda

anaconda.com/download

ANACONDA

Enterprise Pricing Resources About

Sign In Contact Sales

Anaconda Distribution

Free Download

Everything you need to get started in data science on your workstation.

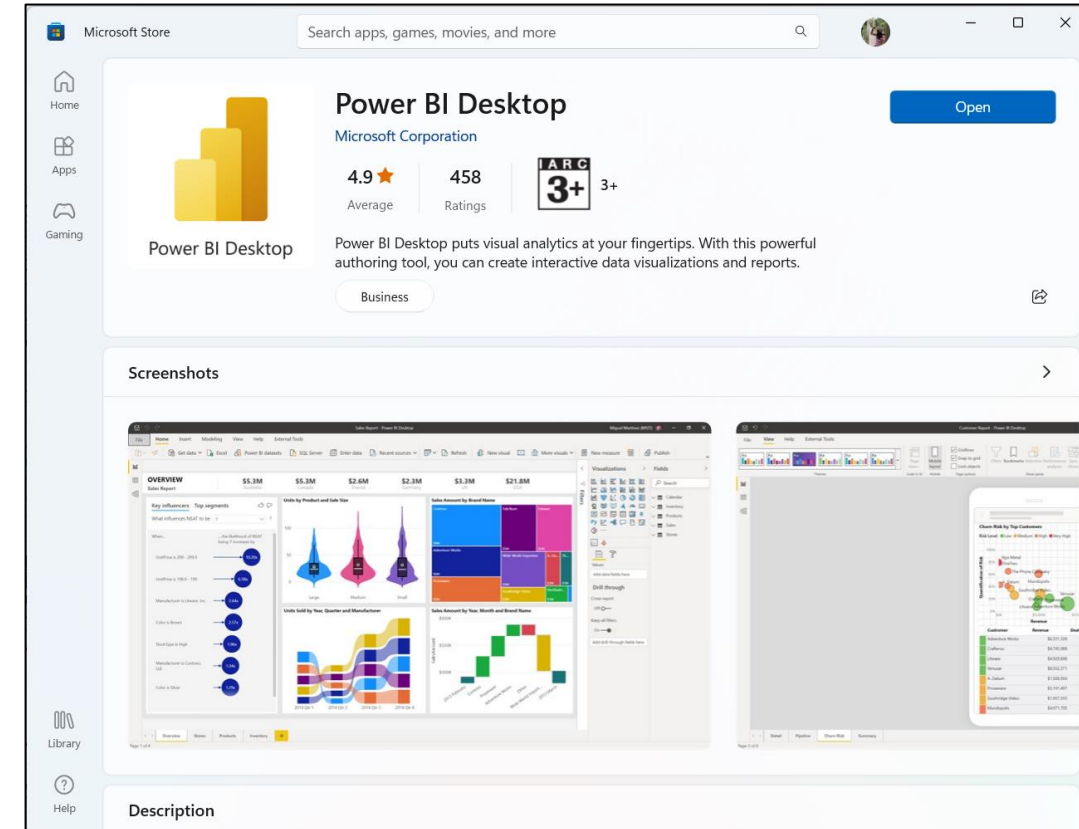
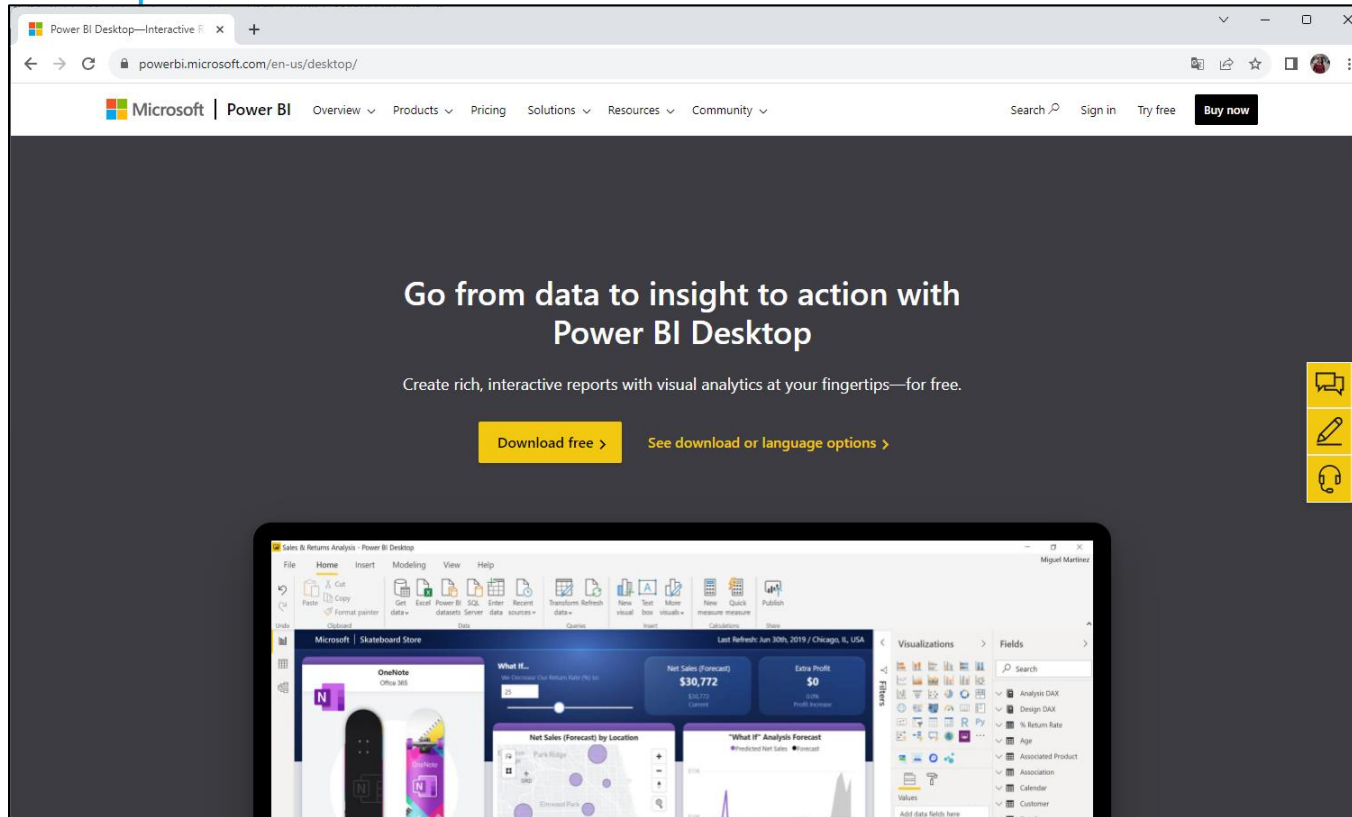
- ✓ Free distribution install
- ✓ Thousands of the most fundamental DS, AI, and ML packages
- ✓ Manage packages and environments from desktop application
- ✓ Deploy across hardware and software platforms

Code in the Cloud Download

Get Additional Installers

Windows macOS Linux

Hey! Welcome to Anaconda. I'm here to help. What are you looking for today?



<https://powerbi.microsoft.com/en-us/desktop/>

