



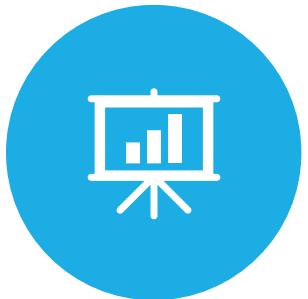
แนวคิดพื้นฐานของ วิทยาการข้อมูลเบื้องต้น

วิทยาการข้อมูลเบื้องต้น
(Fundamentals of Data Science)

ผศ.ดร.ธิติพร ชาญศิริวัฒน์



DATA SCIENCE
คืออะไร



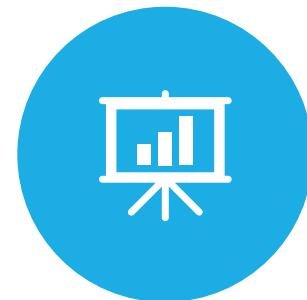
องค์ประกอบของ
DATA SCIENCE



เครื่องมือที่ใช้ในงาน
DATA SCIENCE



ขั้นตอนการทำ
DATA SCIENCE



สายงานใน
DATA SCIENCE

Data Science គីវេតា



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.



Use Science to Understand Data

Data Science

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

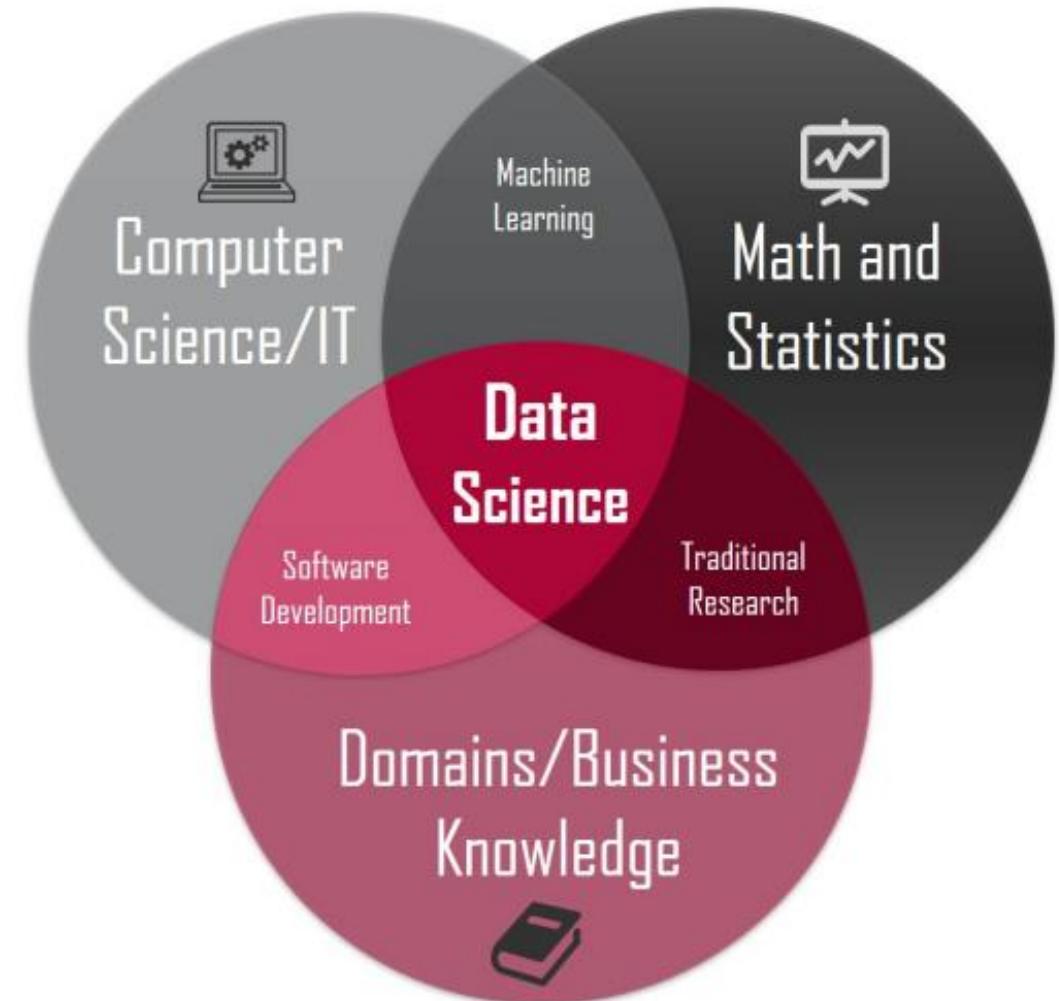
Data Science คืออะไร

- Data Science หรือ วิทยาการข้อมูล หรือ วิทยาศาสตร์ข้อมูล คือ ศาสตร์ กี่ว่าด้วยการนำเอาข้อมูลมาใช้ให้เกิดประโยชน์ในงานด้านต่างๆ
- ศาสตร์ที่เกี่ยวกับการจัดการ จัดเก็บ รวม ตรวจสอบ วิเคราะห์ วิจัย ข้อมูล 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 consisting of:

- Computer Science
- Mathematics and Statistics
- Business Knowledge



AI + Big Data + Data Science

Big Data

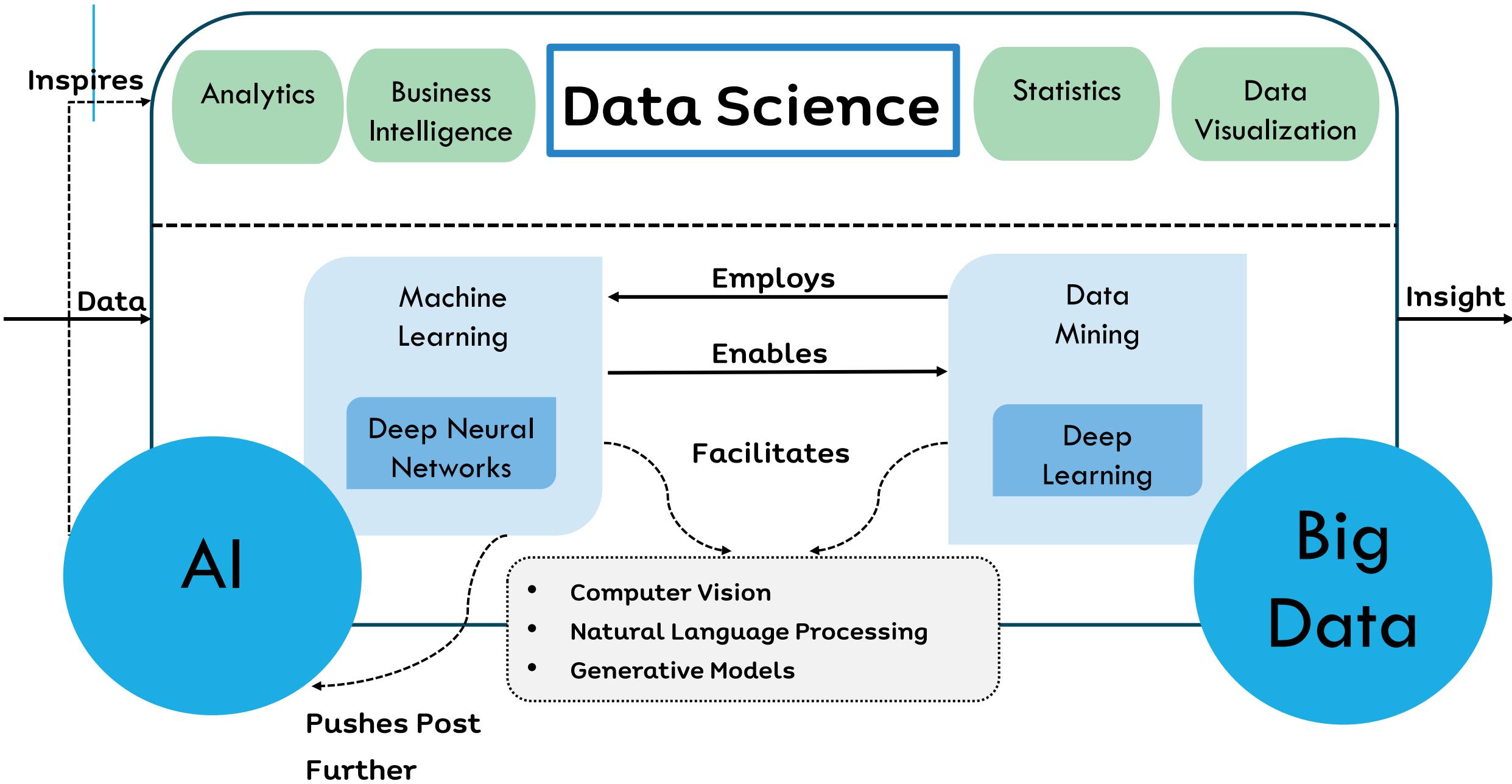
ข้อมูลที่มีโครงสร้าง และไม่มีโครงสร้างจำนวนมาก จากแหล่งที่มาต่างๆ เช่น Social Media, Sensor, ธุรกรรม สำหรับใช้เป็นวัตถุดิบสำหรับ Data Science

Data Science

วิทยาการด้านข้อมูลที่เกี่ยวข้องกับการใช้วิธีการทางสถิติและการคำนวณ เพื่อสกัดข้อมูลเชิงลึก และความรู้ออกมาจาก Big Data

AI

เป็นโครงสร้าง Machine/Computer อัจฉริยะที่เลียนแบบพฤติกรรม/การทำงานของมนุษย์ สร้างกระบวนการเรียนรู้เพื่อให้สามารถทำงานแทนมนุษย์ได้



ชนิดของข้อมูล

▪ Structured Data

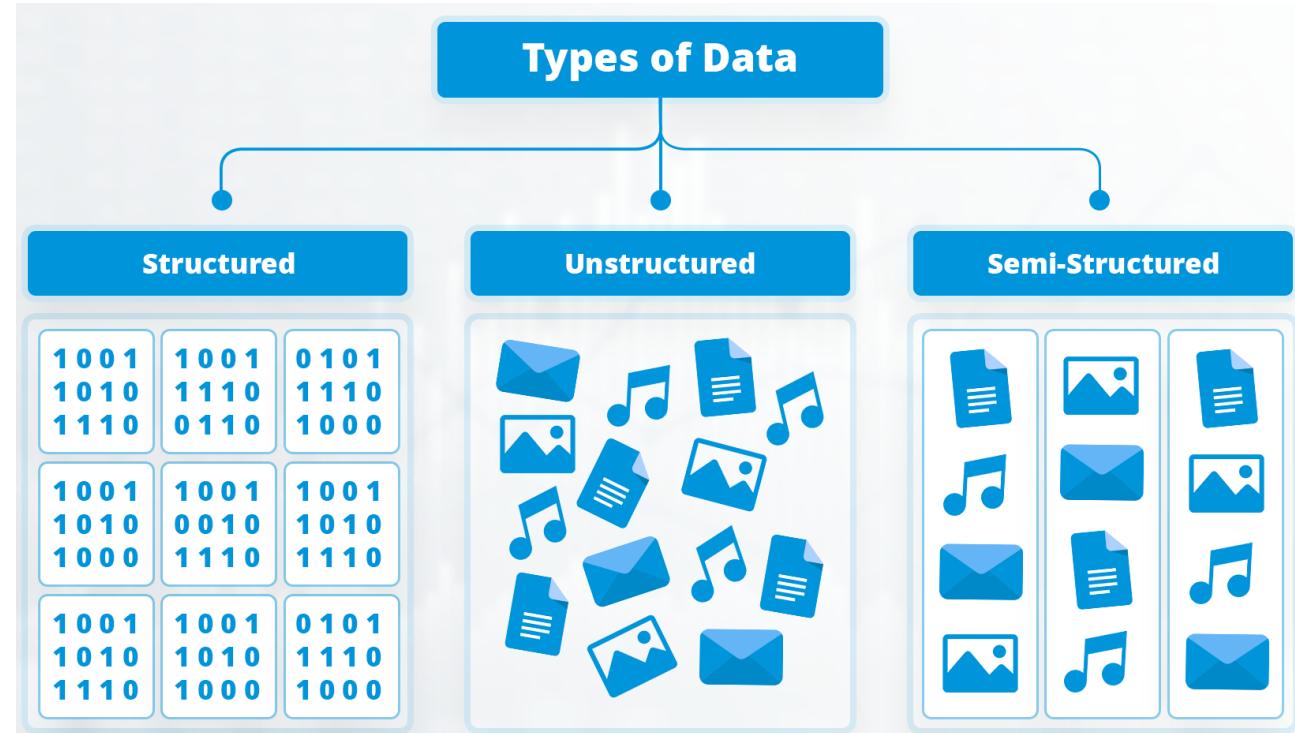
- ข้อมูลที่เก็บอยู่ในรูปแบบของตาราง (Table) ซึ่งประกอบด้วยแถว (Row) และ คอลัมน์ (Column)

▪ Unstructured Data

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

▪ Semi-Structured Data

- ข้อมูลที่มีโครงสร้างข้อมูลที่มีความยืดหยุ่น สามารถขยายโครงสร้างข้อมูลได้ในอนาคต และเรียกใช้ได้รวดเร็ว



เครื่องมือที่ใช้ในงาน Data Science



RAPIDMINER



ขั้นตอนการทำ Data Science

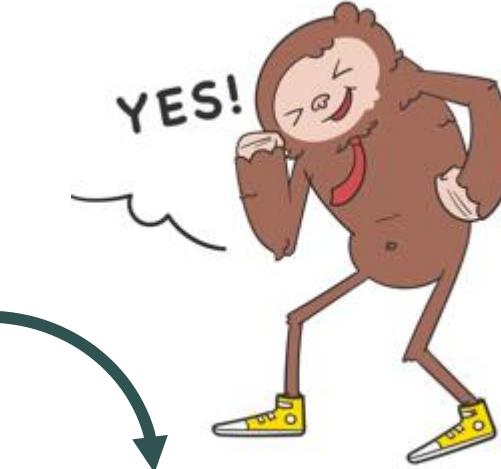
1

Business Problem

Business Questions



Understand



Define Objectives

ทำความเข้าใจปัญหา และแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูล และวางแผนการดำเนินการคร่าวๆ

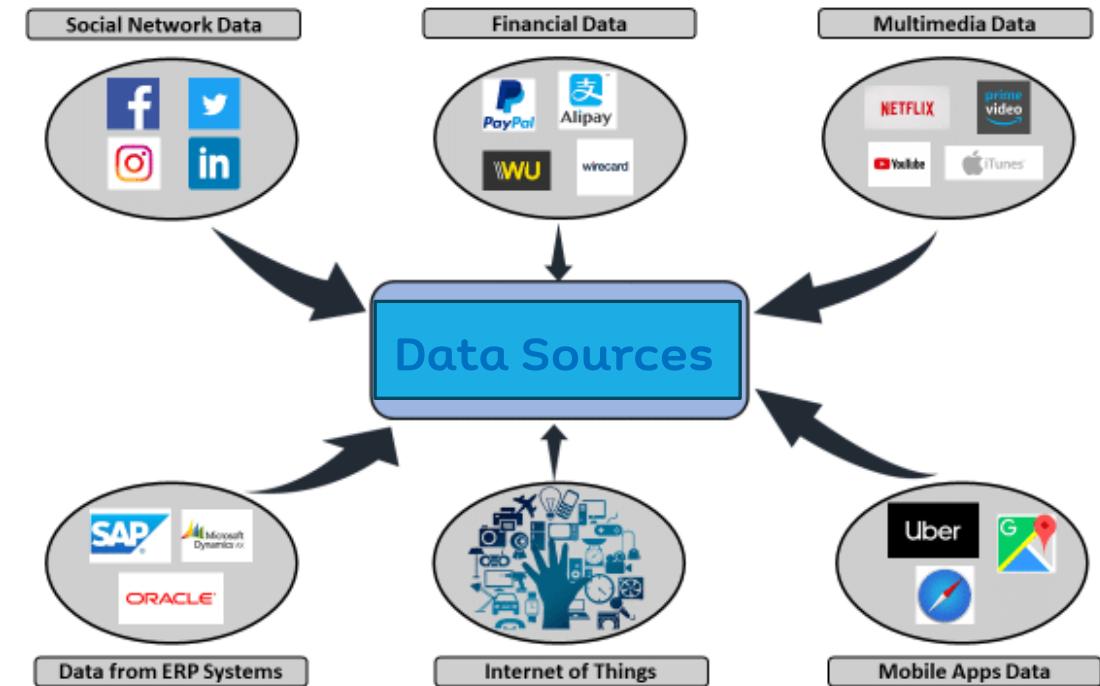
ขั้นตอนการทำData Science

2

Data Acquisition

Data Sources

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



การเก็บรวบรวมข้อมูล

ขั้นตอนการทำ Data Science

3

Data Preparation

Data Selection



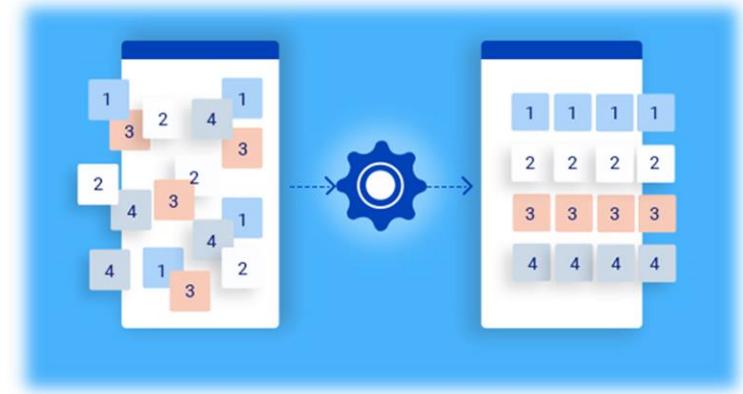
เลือกข้อมูล
ตัวแปรที่จะใช้งาน

Data Cleaning



หาสิ่งผิดปกติของข้อมูล
noise, missing value,
outliner

Data Transformation



แปลงข้อมูลให้อยู่ใน
รูปแบบที่พร้อมนำไปใช้ใน
การวิเคราะห์

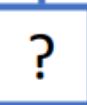
ขั้นตอนการทำ Data Science

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

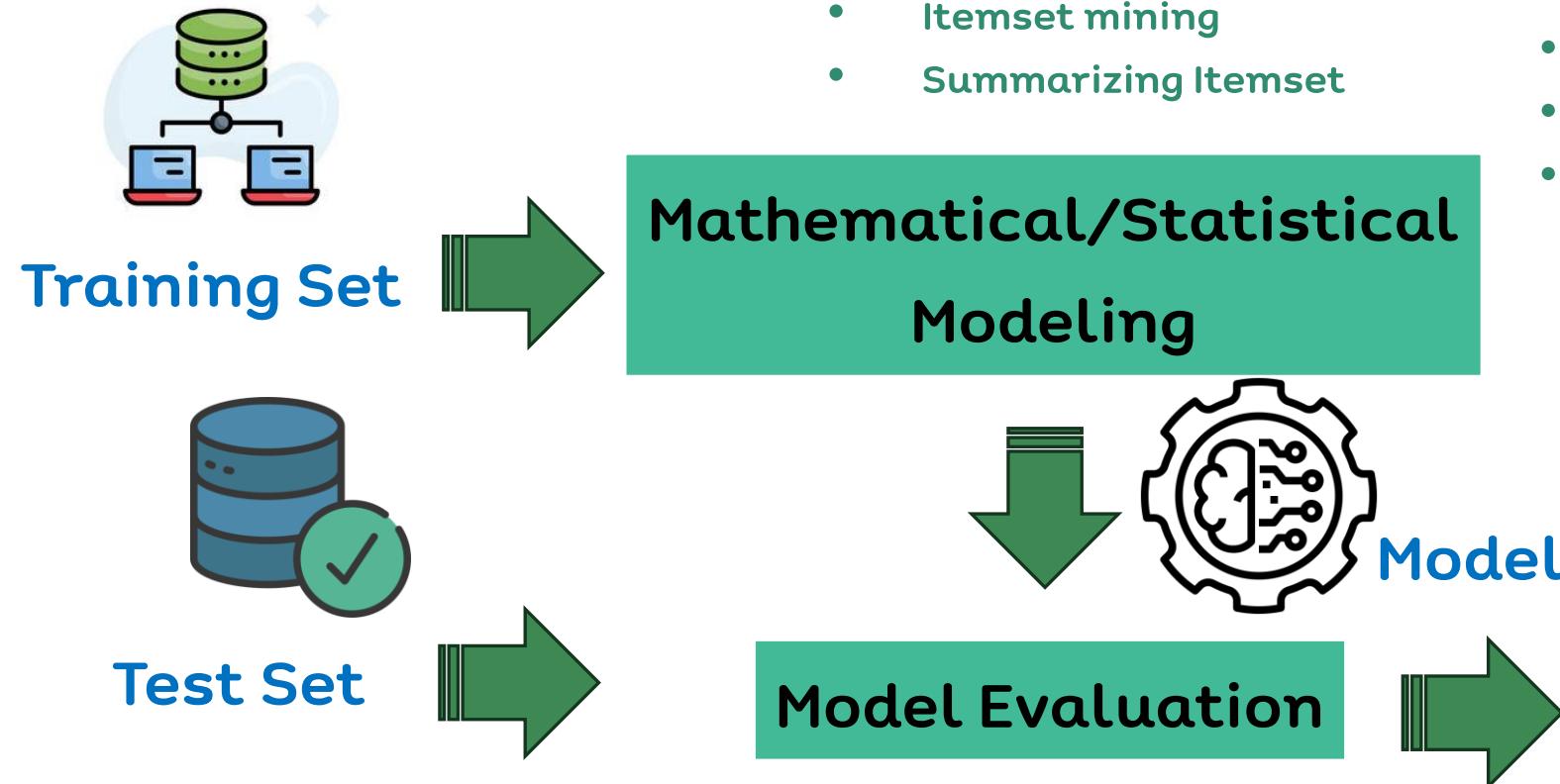


สำรวจ และสรุปสิ่งที่อยู่ในข้อมูล

ขั้นตอนการทำ Data Science

5

Data Modeling



Association

- Itemset mining
- Summarizing Itemset

Regression

- Linear regression
- Polynomial regression

Recognition

- Decision tree
- kNN
- SVM

Clustering

- K-mean
- Hierarchical clustering
- DBSCAN

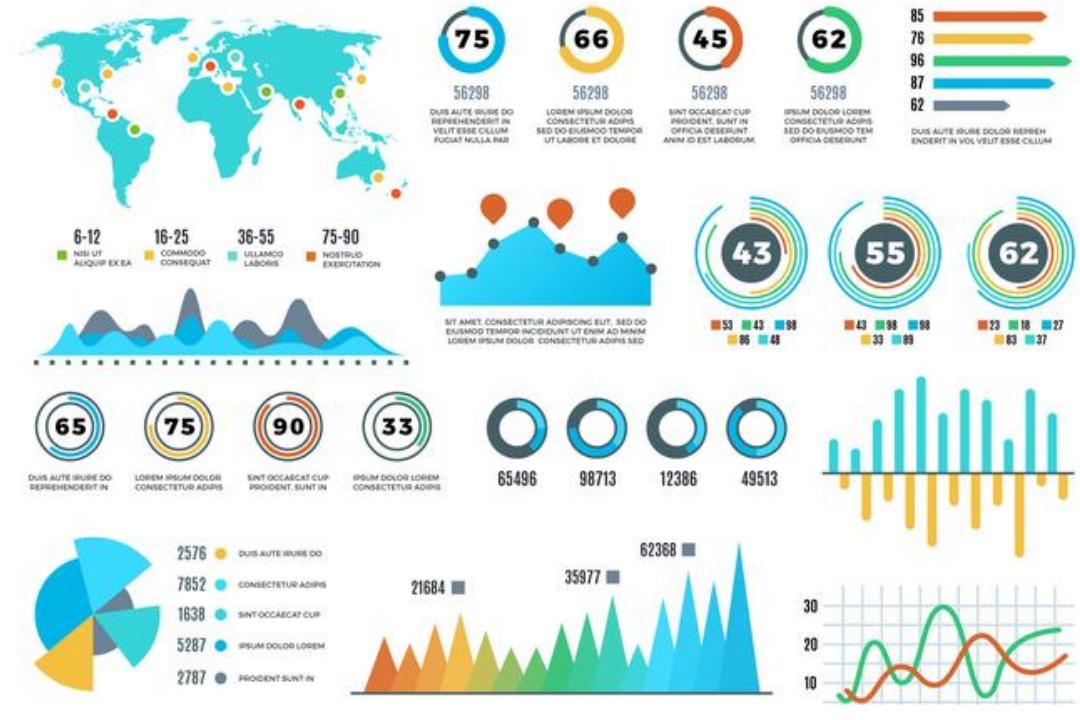
Recognition/prediction
Accuracy

ขั้นตอนการทำ Data Science

6

Visualization & Communication

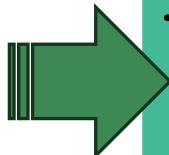
- Create powerful reports and dashboards
- Communicate business finding to convince the stakeholders



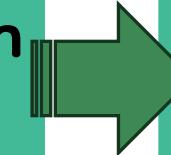
ขั้นตอนการทำ Data Science

7

Deploy & Maintenance



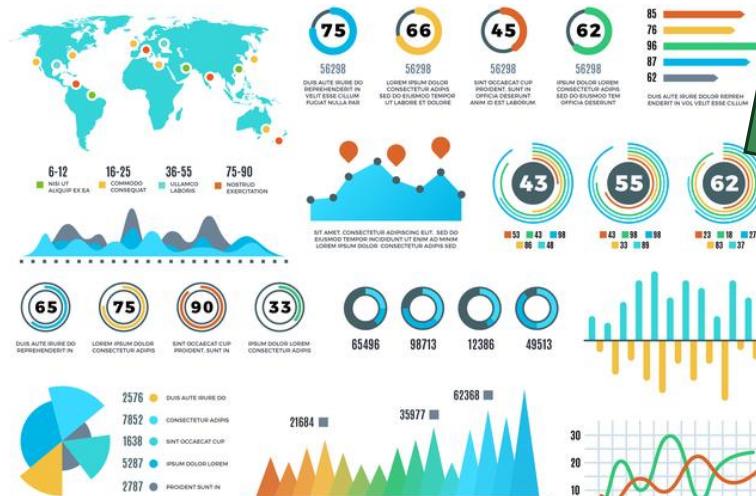
Test in pre-production environment



Deploy in production environment



Model



Running Model

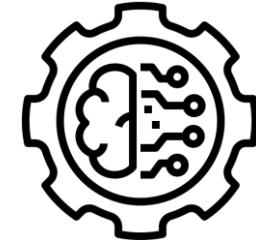
Real-time Analytics

Train/Build Model

ข้อมูลในอดีต

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

Data Science

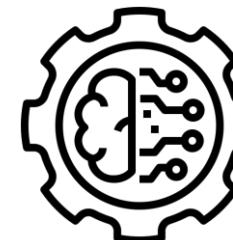


Model

Deployment

Age	Gender	Income	Class
37	Female	65K	?

ข้อมูลใหม่ที่ไม่ทราบ Class



Model

Deployment



Class = Yes
ความน่าจะเป็น 0.9

สายงานใน Data Science

Top 10 วงการยุคใหม่ ที่ใช้ความสำคัญกับการนำ 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

กำหนดที่ในการนำ Big Data มาวิเคราะห์เพื่อแก้ไขปัญหาที่องค์กรกำลังเผชิญหรือวิเคราะห์ประสิทธิภาพของผลิตภัณฑ์และการให้บริการ

- ทักษะการใช้เครื่องมือ
- ทักษะการคิดวิเคราะห์และความเข้าใจธุรกิจ
- ทักษะการนำเสนอ

Data Scientist

นักวิทยาศาสตร์ข้อมูล

กำหนดที่ในการวิเคราะห์ข้อมูลและสร้างเป็นโมเดลเพื่อทำนายแนวโน้มและกำหนดทิศทางธุรกิจในอนาคต

- ทักษะทางคณิตศาสตร์และสถิติ
- ทักษะการเขียนโปรแกรม

Data Engineer

วิศวกรข้อมูล

กำหนดที่ในการรวบรวมข้อมูลจากหลากหลายแหล่ง 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	🟡	🟡	🟡	🟡
เข้าใจกลไกการทำงานของฟังก์ชันหลายตัวแปร และพิษคณิตเชิงเส้น	🟡	🟡	🟡	🟡



สำคัญน้อย



สำคัญ



สำคัญมาก

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

ตัวอย่างการใช้งาน Data Science

Netflix

แพลตฟอร์มวิดีโอสตรีมมิ่ง ที่มีข้อมูลของผู้ใช้อยู่ในเมือง และใช้ Big Data ในการดำเนินธุรกิจ กำหนดรูปแบบ วิดีโอด้วยว่าจะเป็นกังกาพยนตร์ หรือซีรีย์ได้ตรงใจกับกลุ่มลูกค้า

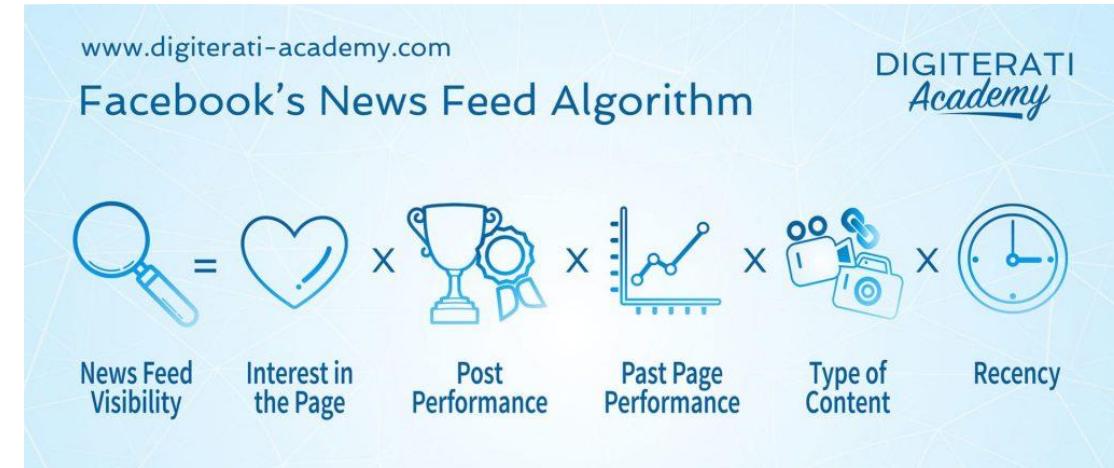
โดยนำข้อมูลอายุผู้ใช้ รูปแบบวิดีโอที่ชอบรับชม ระยะเวลาในการรับชม และอื่นๆ มาวิเคราะห์ พฤติกรรมผู้ชมทั่วโลก รวมถึงการไปเชื้อลิขสิทธิ์ ภาพยนตร์หรือซีรีย์ที่ตอบโจทย์ความต้องการของลูกค้า



ตัวอย่างการใช้งาน Data Science

Facebook

ผู้นำด้านโซเชียลมีเดียของโลก ที่ได้ใช้ เทคนิคขั้นสูงในศาสตร์ของ Data Science เพื่อศึกษาพฤติกรรมของผู้ใช้ และหาข้อมูลเชิงลึก ด้วย Deep learning



Fb ใช้การจัดลำดับหน้าและการวิเคราะห์ข้อความในการจัดลำดับหน้า ใช้ เครือข่ายประสาทเทียมที่ทรงพลังเพื่อจำแนกใบหน้าในรูปถ่าย ใช้เว็บจีนการ เข้าใจข้อความของตนเองที่เรียกว่า "DeepText" เพื่อกำหนดความเข้าใจประโยชน์ ของผู้ใช้ ใช้การเรียนรู้เชิงลึกสำหรับโฆษณาที่ตรงเป้าหมาย

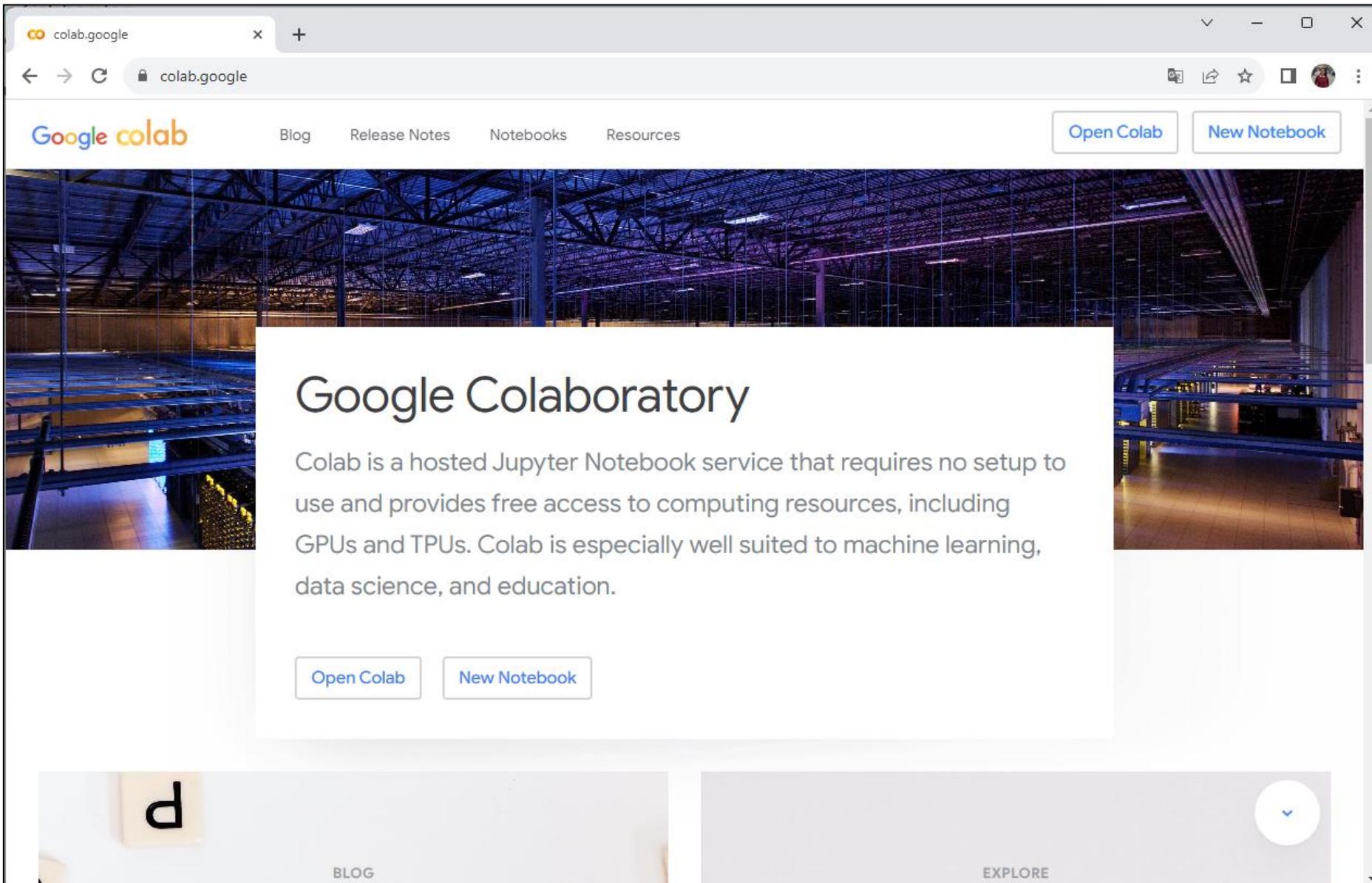
ตัวอย่างการใช้งาน Data Science

Grab Food Thailand

Grab บ้านเราก็มีการใช้กลยุทธ์ Data-Driven Marketing เช่นกัน อย่างแคมเปญช่วงต้นปี 2022 ที่ผ่านมาได้มีการนำข้อมูลส่วนตัวของ User มาสรุปให้ผู้ใช้งานได้ย้อนดูอย่างสนุกสนานว่าปีนี้ตัวเองสั่งอะไรกันบ้าง เป็นสายไหน และปิดก้ายด้วยการเสนอ Code ส่วนลดร้านอาหารประเภทนั้นๆ ให้ลูกค้าได้นำไปใช้ต่อ



ติดตั้ง ใช้งานโปรแกรม



<https://colab.google/>

The screenshot shows a web browser window with the title bar "Free Download | Anaconda". The URL in the address bar is "anaconda.com/download". The page itself has a dark background with green and white text. At the top, there's a navigation bar with the Anaconda logo, links for "Enterprise", "Pricing", "Resources", and "About", and buttons for "Sign In" and "Contact Sales". Below this, the text "Anaconda Distribution" is followed by a large, bold "Free Download" heading. A subtext below it reads "Everything you need to get started in data science on your workstation." To the left of the main content area, there's a bulleted list of benefits: "✓ 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". At the bottom, there are two green buttons: "Code in the Cloud" and "Download". Below these buttons, there's a link "Get Additional Installers" with icons for Windows, Mac, and Linux. A floating chat bubble in the bottom right corner says "Hey! 🌟 Welcome to Anaconda. I'm here to help. What are you looking for today?" with a small green robot icon next to it.

<https://www.anaconda.com/download>

The screenshot shows the Microsoft Power BI Desktop landing page. At the top, there is a navigation bar with links for Overview, Products, Pricing, Solutions, Resources, and Community. Below the navigation bar, a large heading reads "Go from data to insight to action with Power BI Desktop". A subtext below it says "Create rich, interactive reports with visual analytics at your fingertips—for free.". There are two prominent buttons: "Download free >" and "See download or language options >". To the right of these buttons are three yellow icons: a speech bubble, a pencil, and a phone. Below the heading, there is a screenshot of the Power BI desktop application interface, which includes a ribbon menu, a report view with various charts and tables, and a fields pane on the right.

The screenshot shows the Microsoft Store page for Power BI Desktop. The page features a large thumbnail image of the software interface. To the right of the thumbnail, the product name "Power BI Desktop" is displayed along with the developer "Microsoft Corporation". It has a rating of 4.9 stars from 458 reviews and is categorized as a "Business" app. A "Gaming" category is also listed. Below the main image, there is a section titled "Screenshots" showing several screenshots of the software's interface. A "Description" section is present at the bottom of the page.

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

Untitled - Power BI Desktop

File Home Insert Modeling View Optimize Help

Cut Copy Format painter Paste Clipboard Get data workbook OneLake data hub SQL Server Enter data Dataverse Recent sources Transform data New visual Text box More visuals Insert New measure Quick measure Calculations Sensitivity Publish

Search

Sign in Share

Add data to your report

Once loaded, your data will appear in the Data pane.

Import data from Excel Import data from SQL Server Paste data into a blank table Try a sample dataset

Get data from another source →

Visualizations Build visual Filters Data

Values Add data fields here Drill through Cross-report Keep all filters Add drill-through fields here

Page 1 of 1 82%