

การฝึกอบรมผ่าน Clinic Online สำหรับสำนักงานสถิติจังหวัด
โครงการจัดทำระบบบัญชีข้อมูลภาครัฐ (Government Data Catalog)
ประจำปีงบประมาณ พ.ศ. 2566

การวิเคราะห์โจทย์เพื่อการนำเสนอ

วันที่ 9 พฤษภาคม 2566

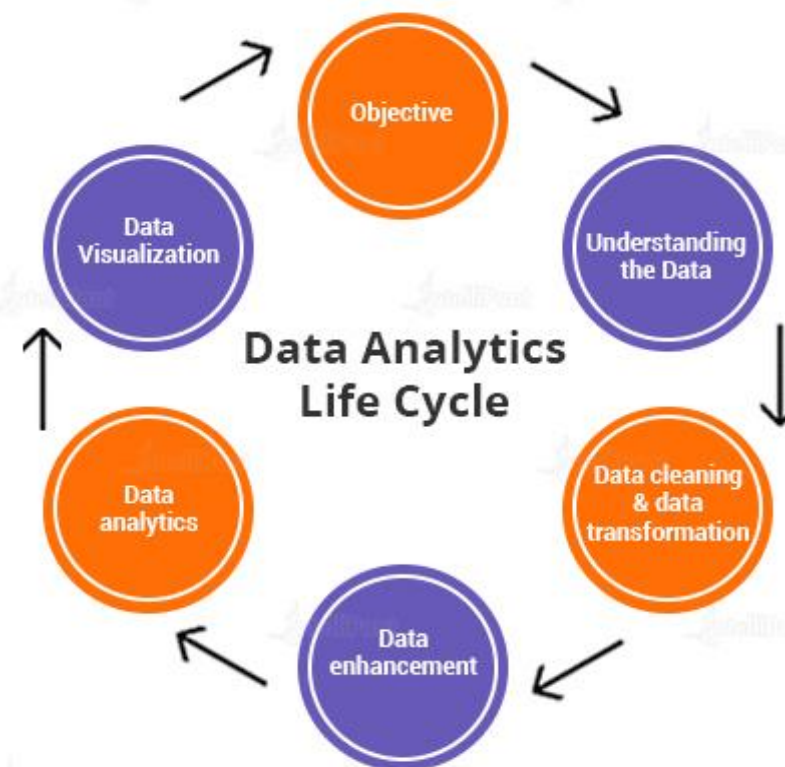
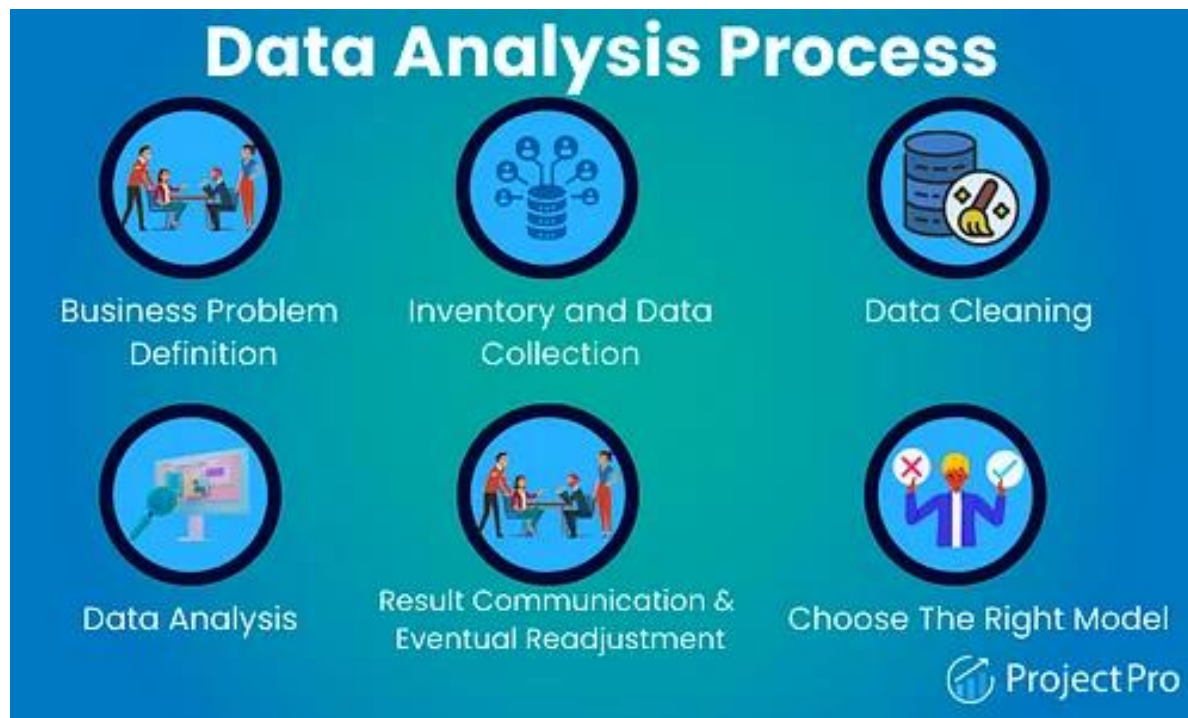


Clinic Online | **GOVERNMENT
DATA
CATALOG
2023**

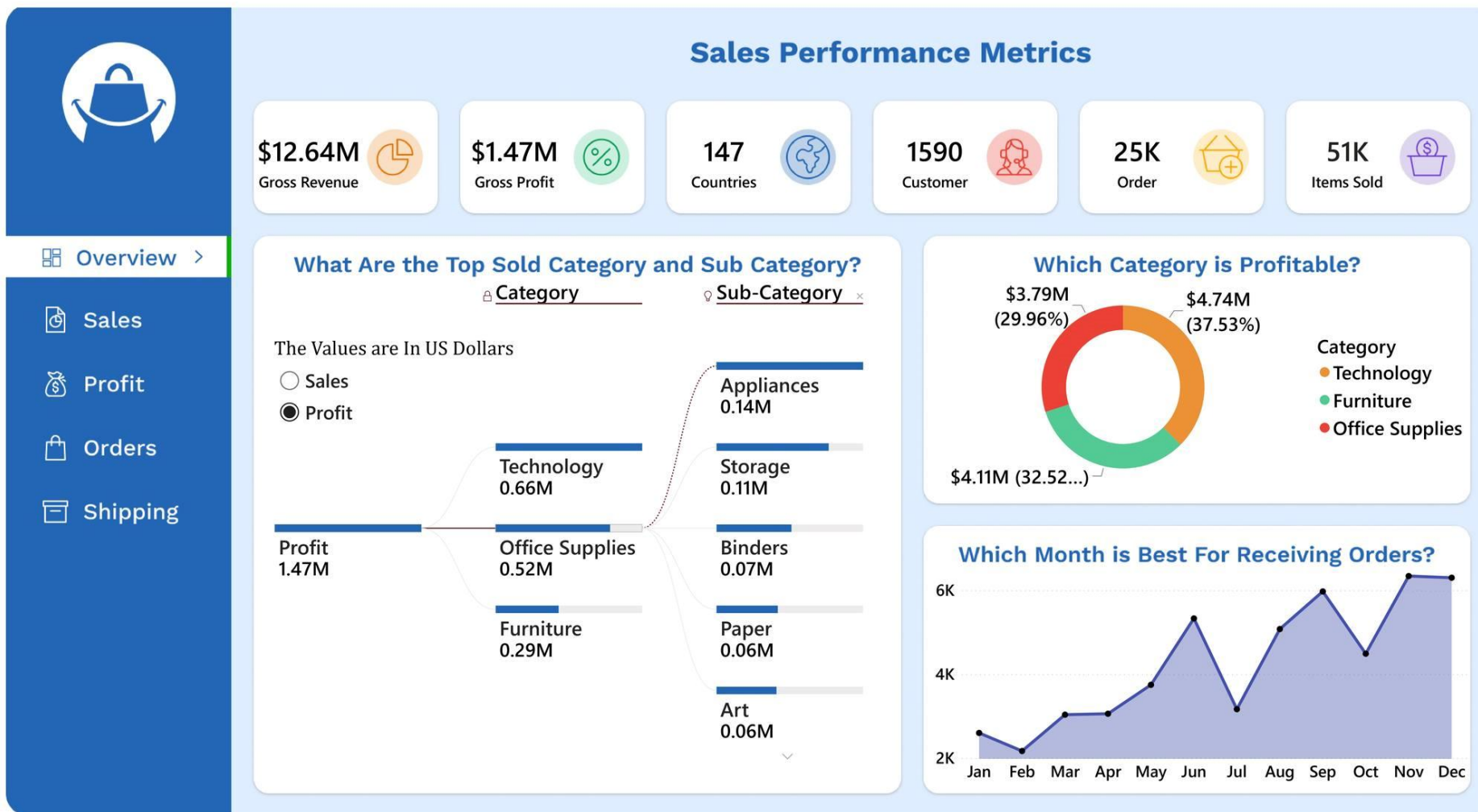
นายทรงพล ใหม่อาลี
ผู้อำนวยการศูนย์สารสนเทศศาสตร์ภาครัฐ
สำนักงานสถิติแห่งชาติ

🔍 <https://gdcatalog.go.th> ✕

Data Analytics Process



ตัวอย่างของ Report-1



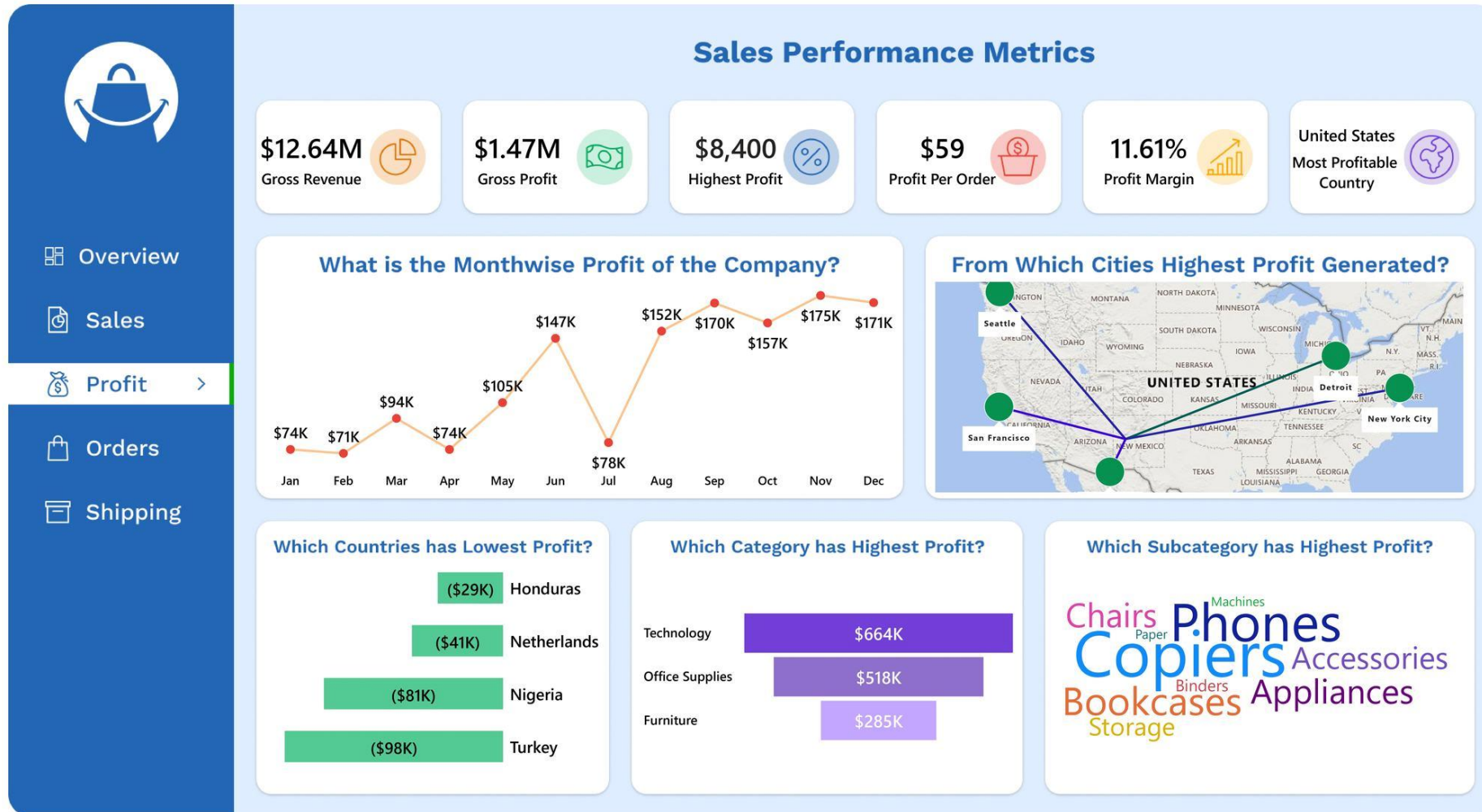


ตัวอย่างของ Report-2



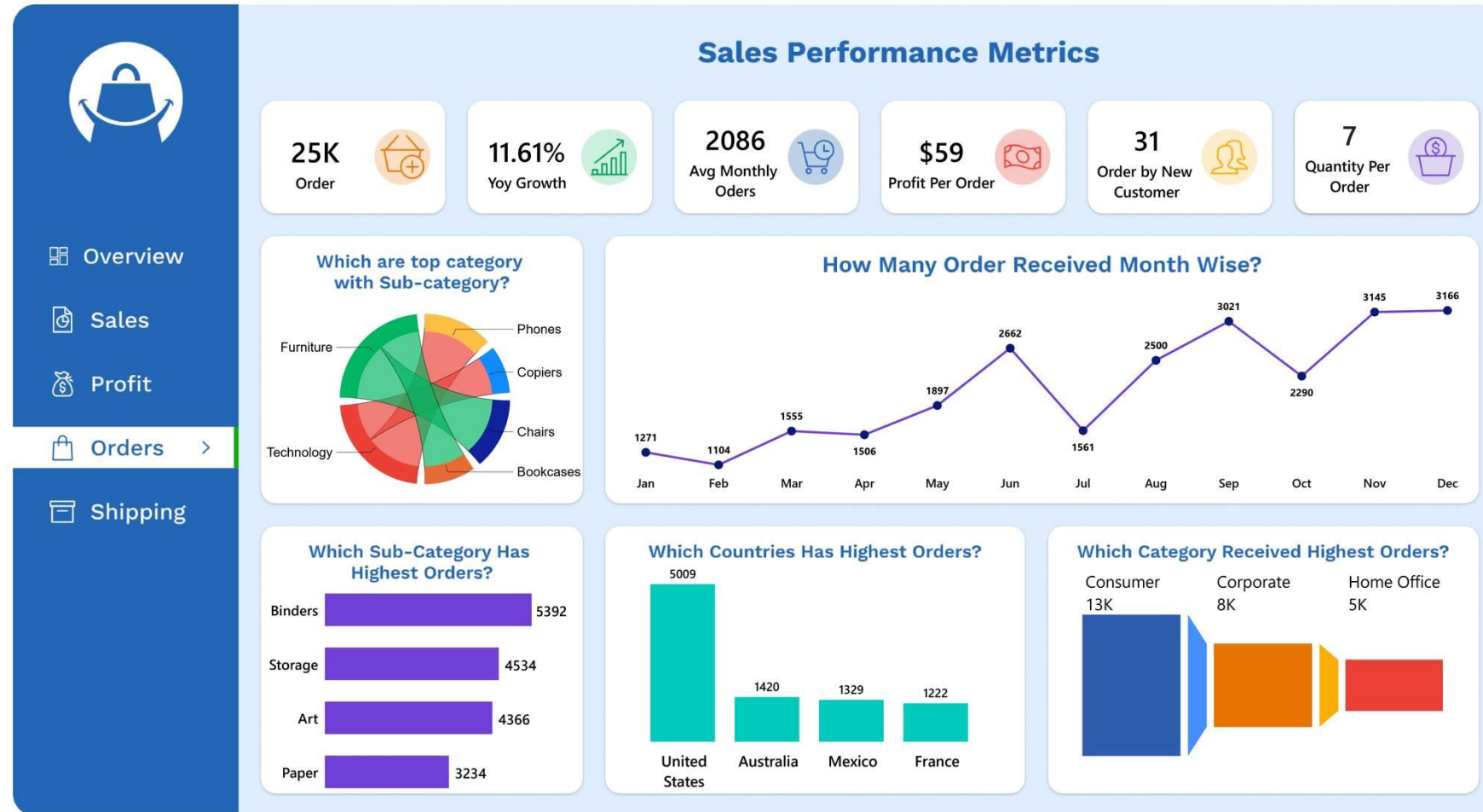


ตัวอย่างของ Report-3



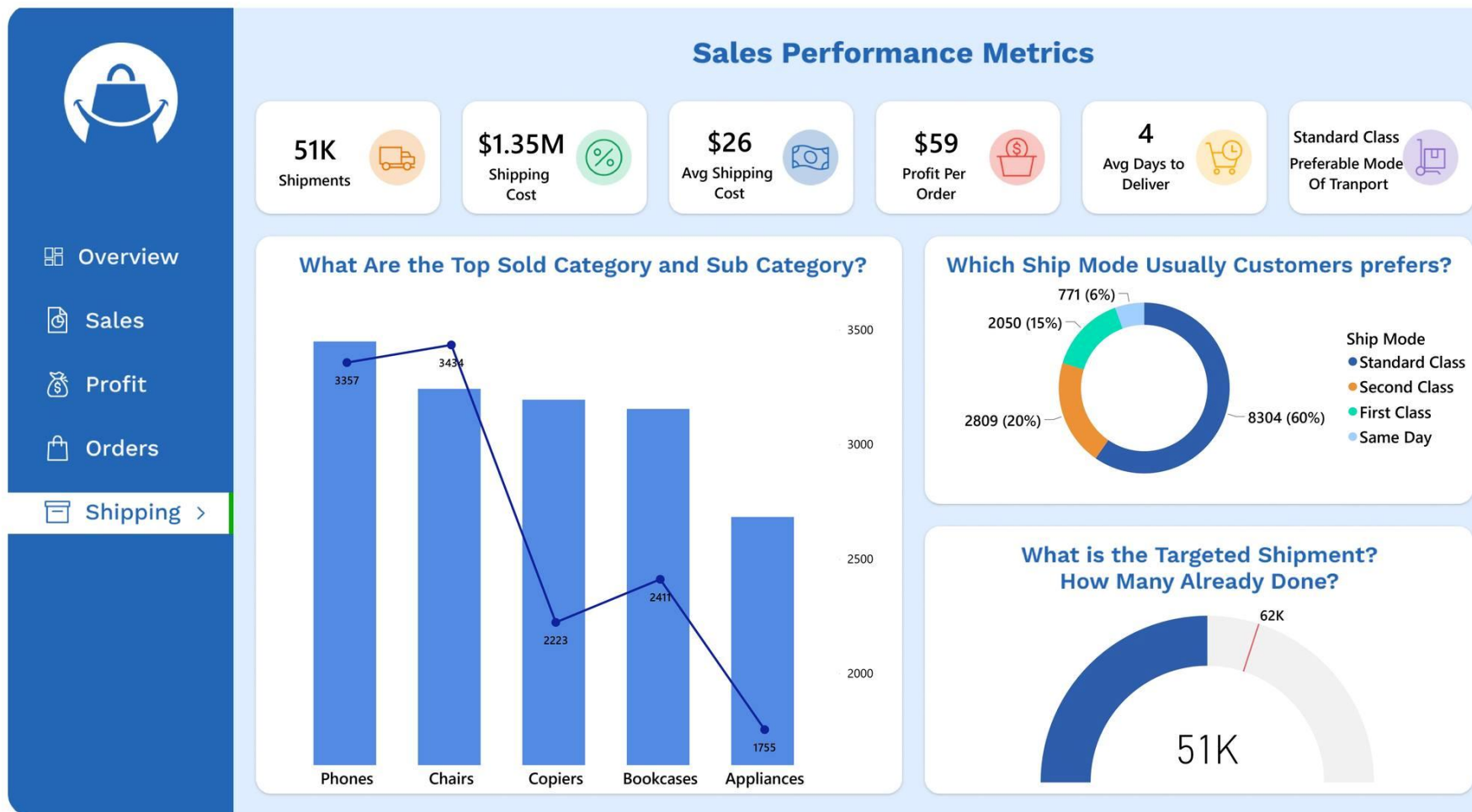


ตัวอย่างของ Report-4





ตัวอย่างของ Report-5



คำถามที่ต้องการคำตอบด้วยสารสนเทศ

- การทำความเข้าใจกับเรื่องให้เข้าใจ
 - อ่านเล่มเอกสารการวิเคราะห์ (ระบุปัญหา) และ BCM (ระบุคำถามไว้ที่คำถาม ?)
 - ทบทวนปรับปรุงเล่มเอกสารให้สมบูรณ์มากยิ่งขึ้นโดยเฉพาะบทที่
- สารสนเทศที่จัดทำขึ้นนำไปใช้ในการแก้ปัญหาได้ก็คำถาม
 - ต้องสามารถบอกได้ว่าปัญหาที่กำหนดใช้สารสนเทศใดแก้ปัญหา ?
 - สารสนเทศที่สร้างขึ้น(ผังภาพ)ตอบประเด็นที่ต้องการแก้ปัญหาใดแล้วหรือไม่ ?

ดูเอกสารการวิเคราะห์ ฯ ที่เกี่ยวข้อง

- ดูคำถามที่ต้องการคำตอบมีกี่คำถาม (หาจากเอกสาร BCM)
- ชุดข้อมูลใด ?
- จากทรัพยากรข้อมูลใด ?
- ตัวแปรใดบ้าง ?

ระบบบัญชีข้อมูลหน่วยงานของตนเองมีทรัพยากรข้อมูลพร้อม

ประเด็นที่ ๑

.....

?

ประเด็นที่ ๒

.....

- มีชุดข้อมูลระเบียบจำนวน ... ชุด
- มีชุดข้อมูลสถิติจำนวน ... ชุด
- มีชุดข้อมูลภูมิศาสตร์สารสนเทศ ... ชุด
- มีชุดข้อมูลอื่น ๆ ... ชุด
- มีชุดข้อมูลหลากหลาย ... ชุด

- มีชุดข้อมูลระเบียบจำนวน ... ชุด
- มีชุดข้อมูลสถิติจำนวน ... ชุด
- มีชุดข้อมูลภูมิศาสตร์สารสนเทศ ... ชุด
- มีชุดข้อมูลอื่น ๆ ... ชุด
- มีชุดข้อมูลหลากหลาย ... ชุด



ชุดข้อมูลที่แจ้งลงทะเบียน

- ชุดไหนจะใช้กับผู้มีส่วนได้เสียตรง
 -
- ชุดไหนจะใช้กับเจ้าหน้าที่รัฐ
 -
- ชุดไหนจะใช้กับผู้บริหารในจังหวัด
 -

ประเมินทรัพยากรข้อมูลที่จะใช้

- แบบยากมาก
 - pdf Image (Un search Able)
 - ต้องใช้คนแปลงข้อมูล
- แบบยาก
 - Excel Report
 - PDF Search Able
 - ต้องใช้คนปรับปรุง
- แบบง่าย
 - CSV & Structure & Complete

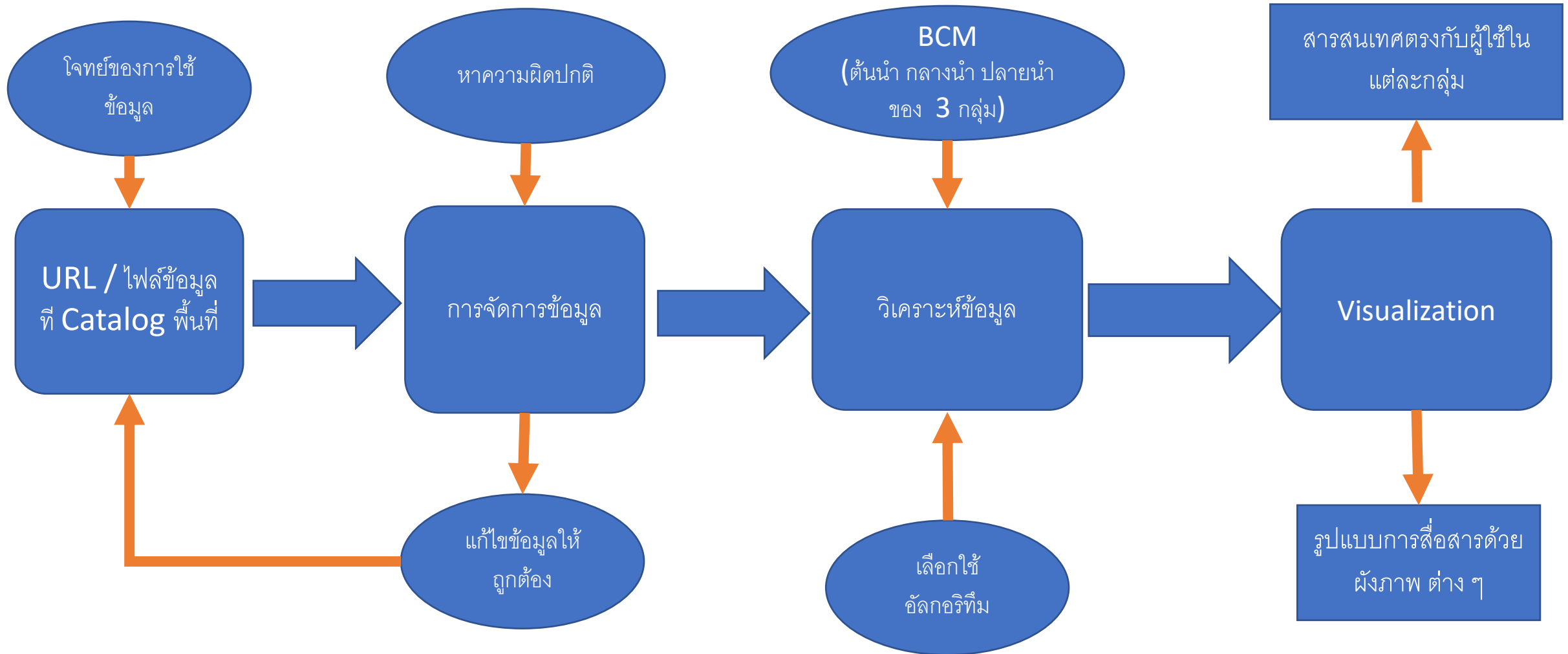
ดึงข้อมูลมาแล้ว

- ทำความเข้าใจกับตารางข้อมูล และ Fields ข้อมูล
- ต้อง Clean Data : 100 %
- มีตัวแปรที่ต้องการใช้งานครบถ้วน
- เป็นแบบโครงสร้าง

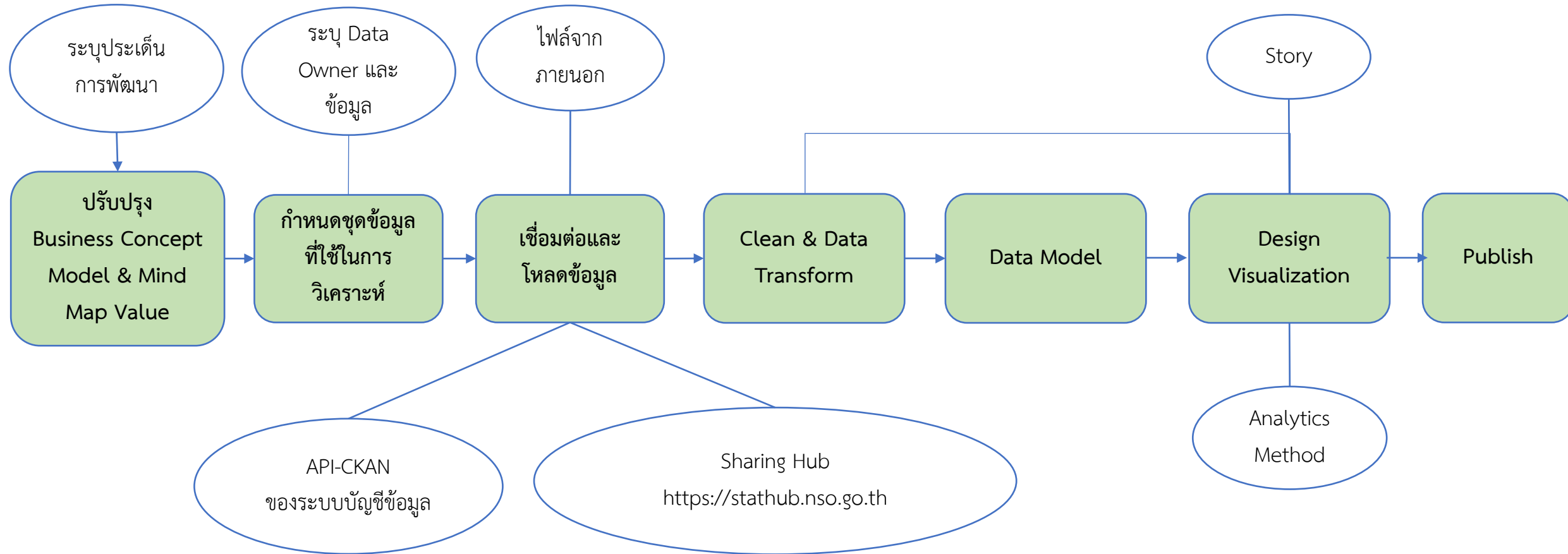
ค่าข้อมูลในทรัพยากร

- แบบนามธรรม (ชอบ, ไม่ชอบ) : ใช้ความรู้สึก เช่น ชอบ ไม่ชอบ พอใจ ไม่พอใจ
- แบบไม่ต่อเนื่อง (จำนวน นักท่องเที่ยว) : ตัวเลขจำนวนเต็ม เช่น จำนวน นักท่องเที่ยว
- แบบต่อเนื่อง (อุณหภูมิ , เกรดเฉลี่ย) : ตัวเลขทศนิยม เช่น อุณหภูมิ
- แบบช่วงชั้น (ช่วงอายุ) : ค่าข้อมูลแบ่งเป็นกลุ่มชั้น เช่น ช่วงอายุ 16-20 ปี

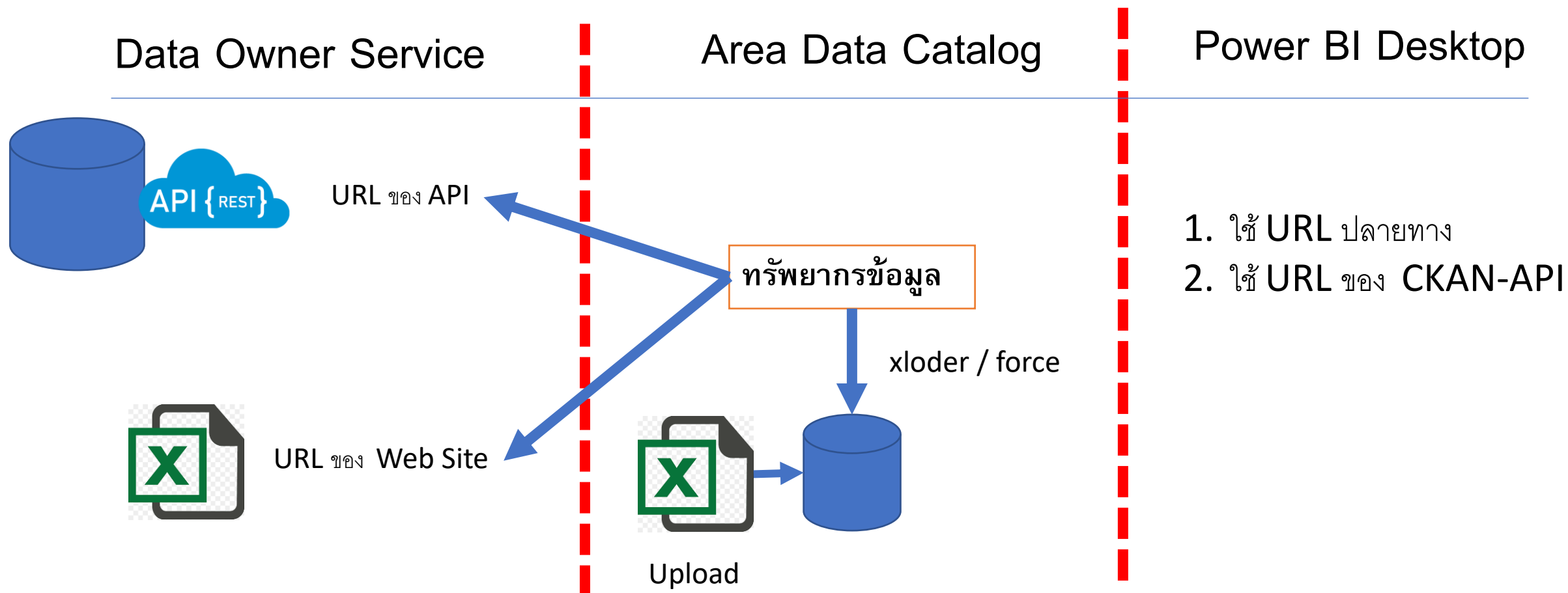
การวิเคราะห์ข้อมูลเพื่อจัดทำสารสนเทศ



ขั้นตอน.การทำ Visualization ด้วย Power BI.



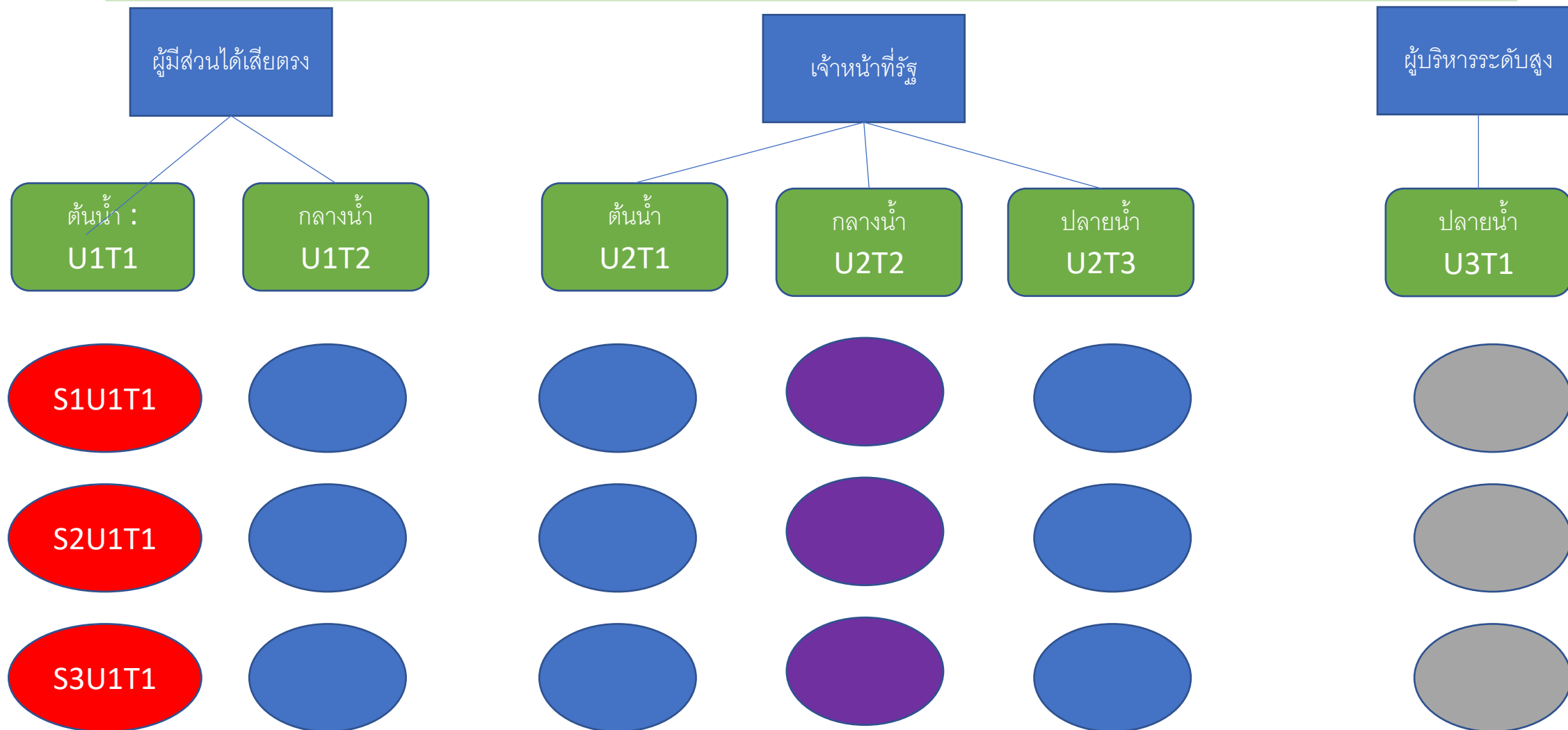
การไหลทรัพยากรข้อมูล



เตรียมการเล่าเรื่องอย่างเป็นระบบ

- ความต้องการบริโภค :: ผลผลิตที่ได้
- เหตุ :: ผล
- การแบ่งแยก :: ต้นน้ำ กลางน้ำ ปลายน้ำ
- ป้องกัน :: แก้ไข
- เป้าหมาย :: ผล
- มีอยู่ :: ไม่พอเพียง
- ต่ำกว่า : สูงกว่า

กำหนดขอบเขตการเล่าเรื่อง.



สารสนเทศมุ่งหมายให้กลุ่มผู้ใช้



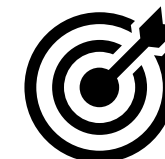
- สภาพปัจจุบัน
- ปัญหาที่สำคัญ
- ความต้องการ

ผู้มีส่วนได้เสียตรง



- ความก้าวหน้าในการดำเนินงาน
- ปัญหาที่เกิดขึ้น
- การสนับสนุนและส่งเสริม

เจ้าหน้าที่รัฐ



- บรรลุเป้าหมายที่กำหนดไว้
- แสดงถึงการเปลี่ยนแปลงในทิศทางใด

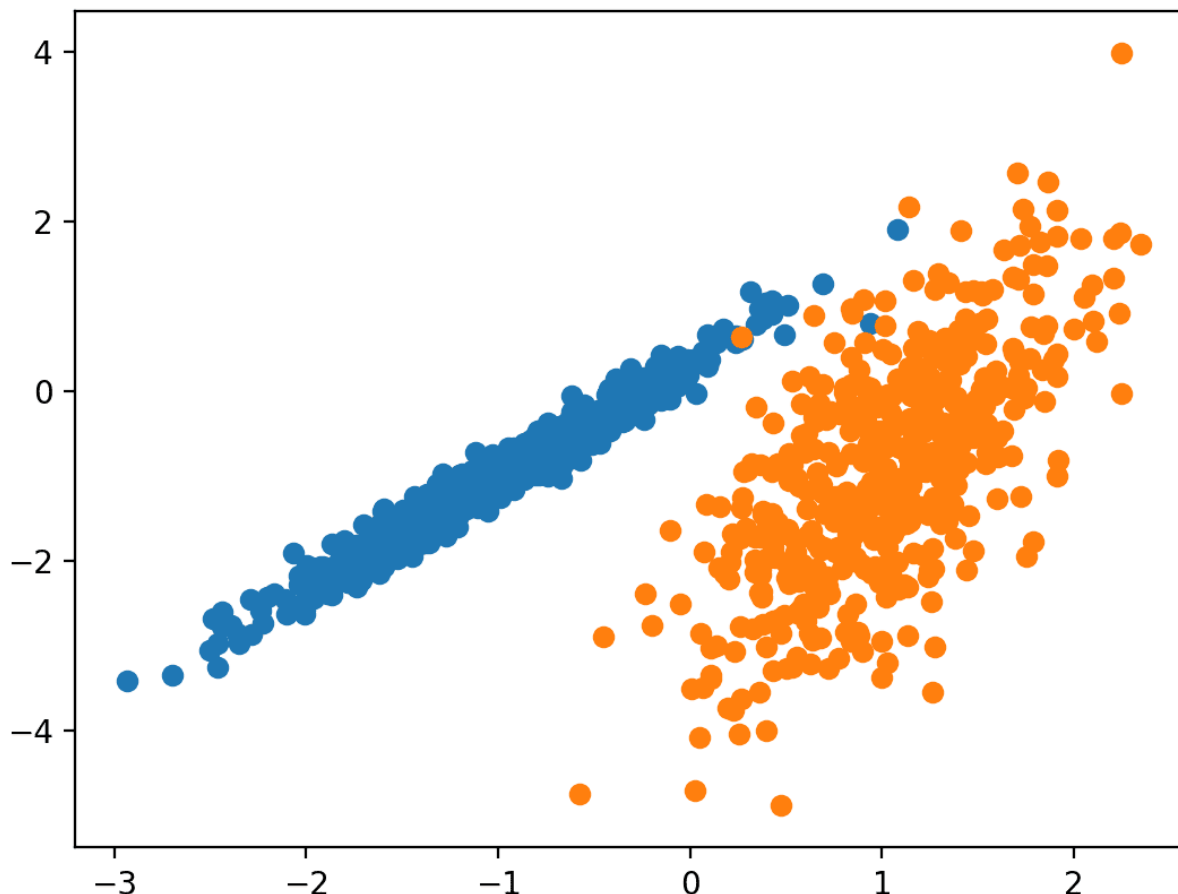
ผู้บริหารระดับสูง

วิธีการวิเคราะห์เพื่อตอบความต้องการ

- การประมาณค่า (Estimation)
- การพรรณนา (Descriptive)
- การทำนาย (Prediction)
- การจำแนกกลุ่ม (Classification)
- การจัดกลุ่ม (Clustering)
- การหาความสัมพันธ์ (Association)

Clustering Algorithms

- Affinity Propagation
- Agglomerative Clustering
- BIRCH
- DBSCAN
- K-Means
- Mini-Batch K-Means
- Mean Shift
- OPTICS
- Spectral Clustering
- Mixture of Gaussians



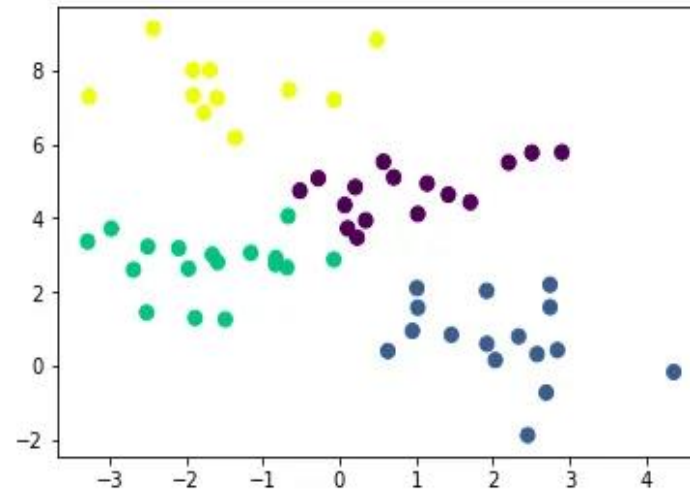
Clustering Algorithms

- Logistics regression
- Random Forest Classifier
- Nearest Neighbours Classification
- SVM Classifier
- Gradient Boosting Classifier

```
from sklearn import svm
svm_clf = svm.SVC(kernel="linear").fit(X_train, y_train)
svm_clf.score(X_test, y_test)
```

0.9833333333333333

```
plt.scatter(X_test[:, 0], X_test[:, 1], c=svm_clf.predict(X_test), s=50);
```

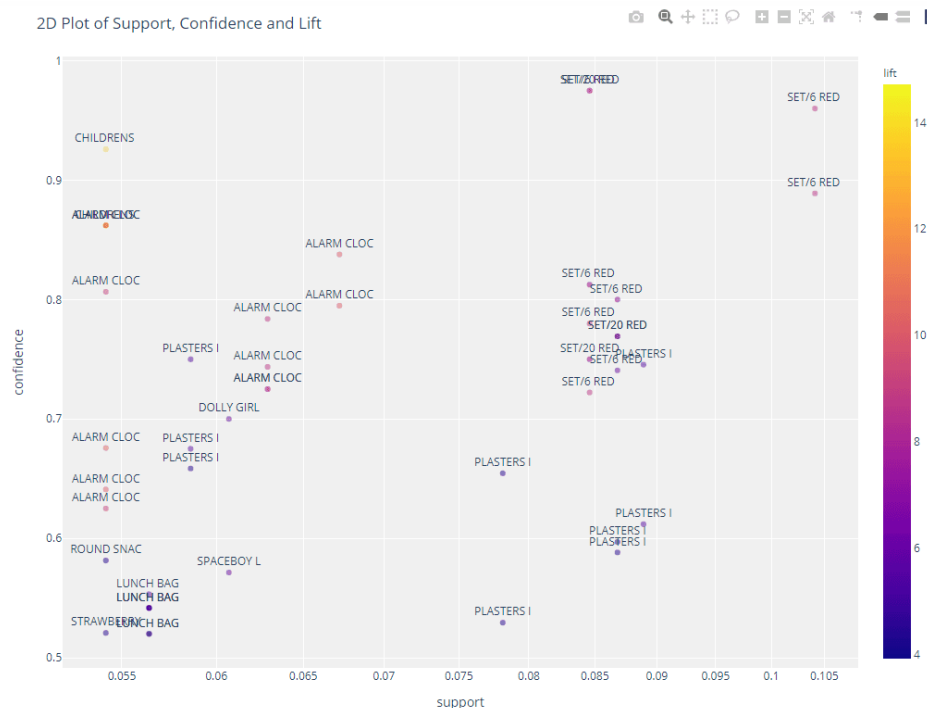


Association

$$\text{Support}(\{X\} \rightarrow \{Y\}) = \frac{\text{Transactions containing both } X \text{ and } Y}{\text{Total number of transactions}}$$

$$\text{Confidence}(\{X\} \rightarrow \{Y\}) = \frac{\text{Transactions containing both } X \text{ and } Y}{\text{Transactions containing } X}$$

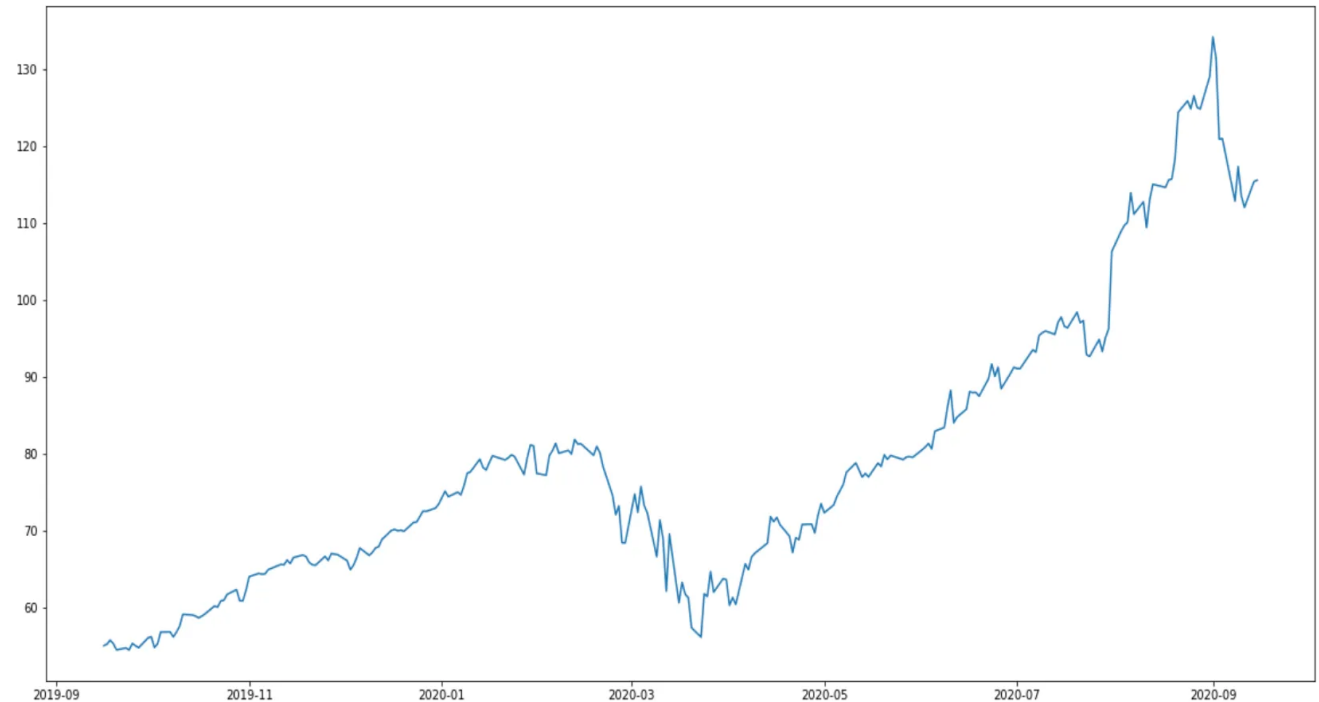
$$\text{Lift}(\{X\} \rightarrow \{Y\}) = \frac{(\text{Transactions containing both } X \text{ and } Y) / (\text{Transactions containing } X)}{\text{Fraction of transactions containing } Y}$$



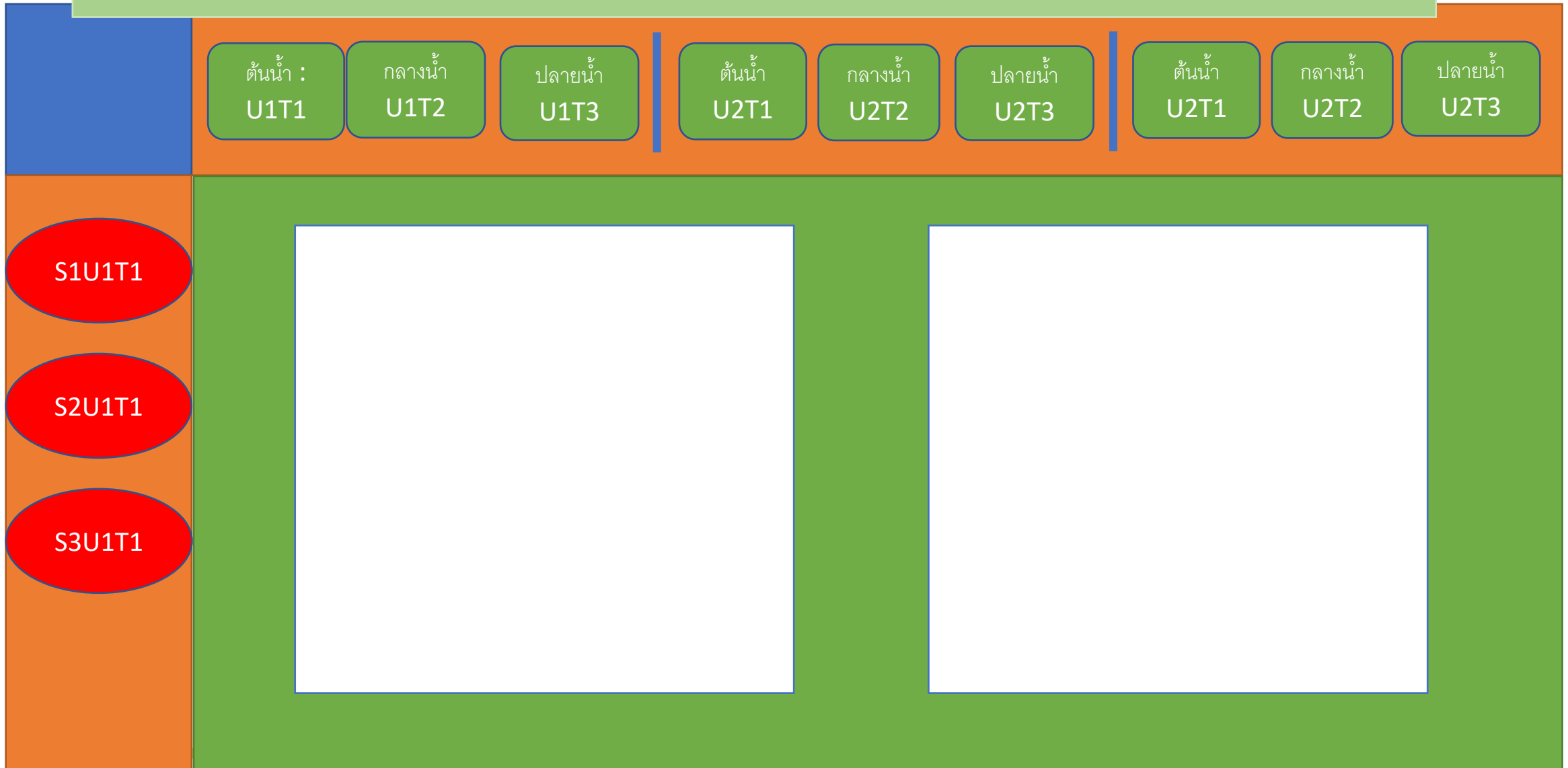
Prediction

- LSTM Prediction Model

```
In [19]: plt.plot(df["Close"],label='AAPL Close Price history')
Out[19]: [<matplotlib.lines.Line2D at 0x14b306c50>]
```



กรอบแนวคิดของเลเอาท์ในการเล่าเรื่อง

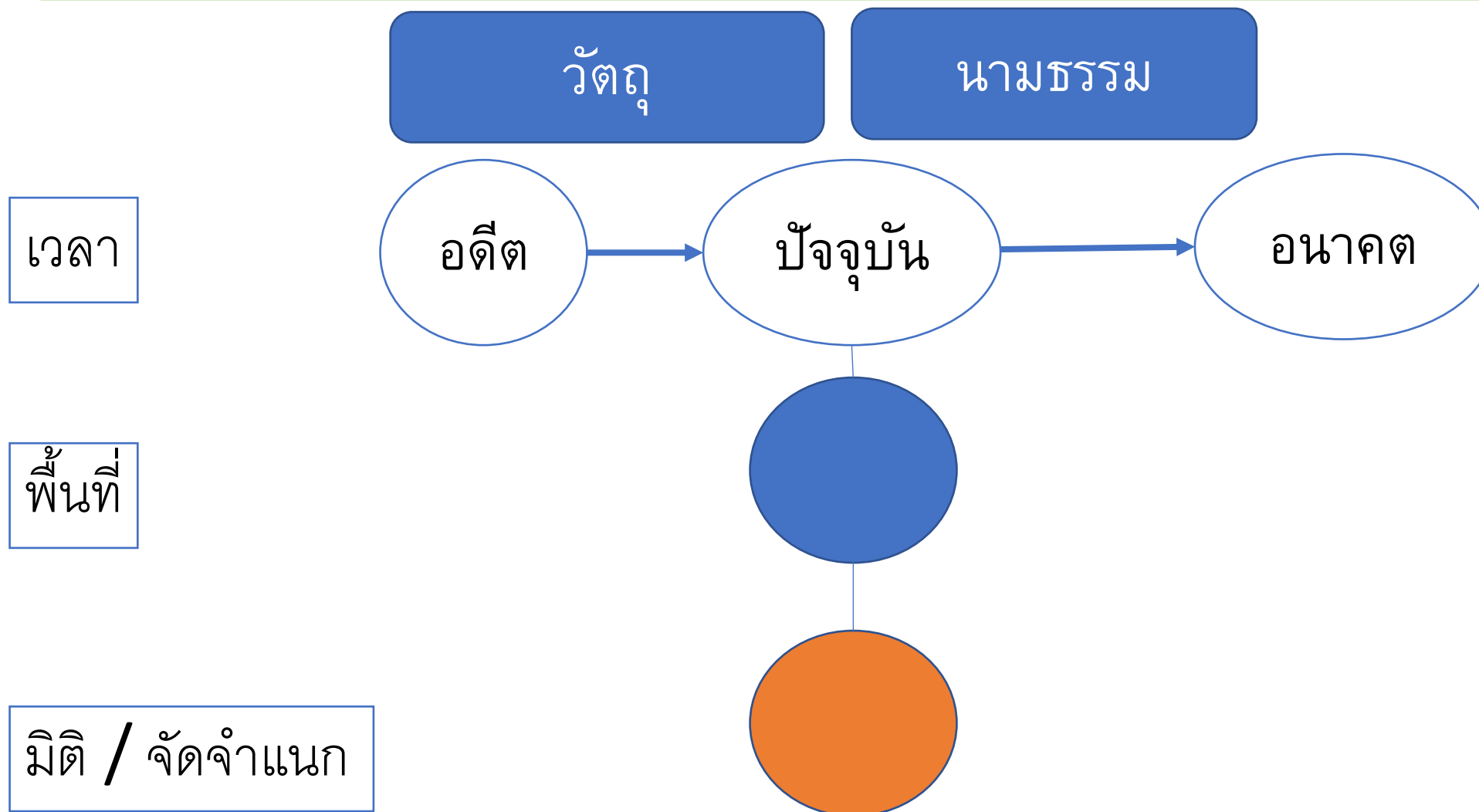


เครื่องมือการสื่อสาร

- ใช้ข้อความ
 - รูปแบบตัวหนังสือ
- ใช้รูปภาพ
 - สี
 - รูป
- ใช้สื่อผสม
 - สี
 - รูป
 - เคลื่อนไหว



ความสัมพันธ์พื้นฐานของข้อมูล



ตัวอย่างประเด็น T1:S1

ต้องการบอกอะไร ?

ใช้ผังกราฟ
อะไรตอบ ?

สารสนเทศนำไปใช้ได้
หรือไม่ ?

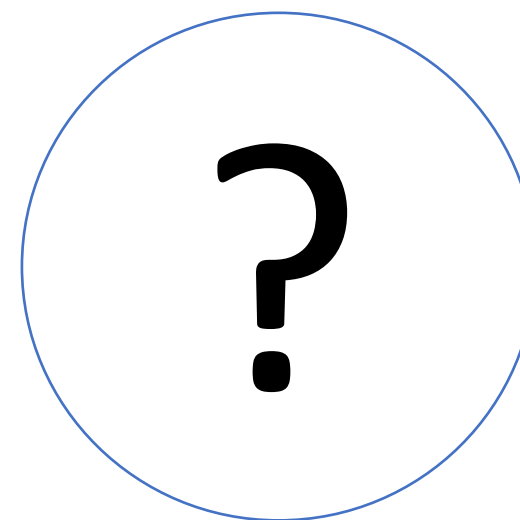
ข้อมูลมีความพร้อม ?

ได้สารสนเทศจากผัง
กราฟหรือไม่ ?

สารสนเทศถูกต้อง ?

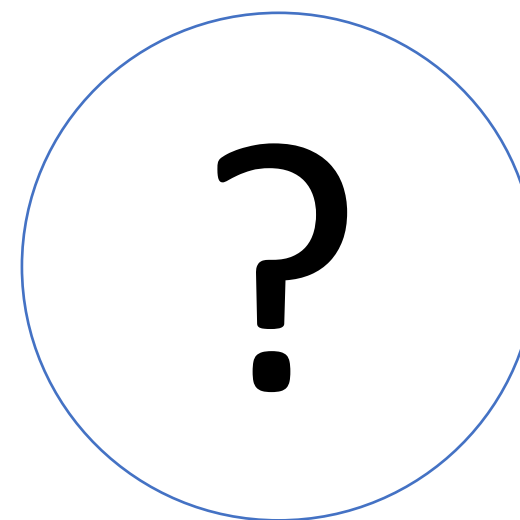
ตัวอย่างปัญหาที่เป็นโจทย์ตั้งต้น: เกษตร

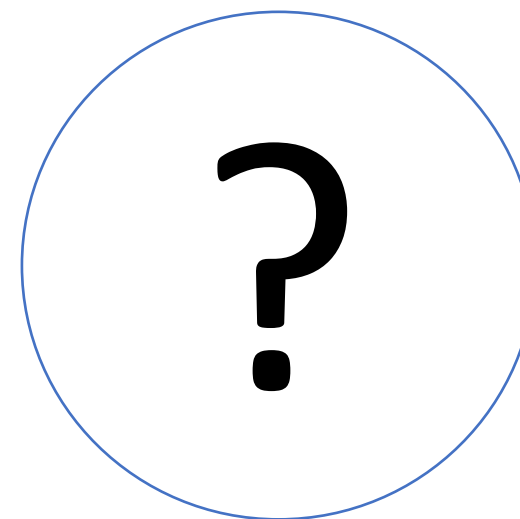
- การส่งเสริมสินค้า**เกษตร (ข้าว)** ให้มีคุณภาพ ปลอดภัยและได้มาตรฐาน เพื่อเพิ่มมูลค่าของผลผลิต อันจะไป**สู่การยกระดับรายได้ของเกษตรกร และผู้ประกอบการแปรรูป**
- ราคาสินค้าและบริการที่สูงขึ้น ส่งผลต่อการบริโภคและการผลิต ทำให้ผู้ผลิตไม่สามารถแบกรับ**ภาระต้นทุนที่สูงขึ้น**ได้ โดยเฉพาะเกษตรกรที่สินค้าเกษตรราคาตกต่ำและไม่แน่นอน เกษตรกรต้องแบกรับความเสี่ยงที่จะขาดทุน
- ภาวะการณ์**เปลี่ยนแปลงของสภาพอากาศ** ส่งผลกระทบต่อการทำเกษตรกรรม เนื่องจากอุณหภูมิที่สูงขึ้น **ปริมาณน้ำฝน**ที่เปลี่ยนแปลงมากขึ้นหรือลดลงกว่าเดิม ความแปรปรวนการกระจายของฝน รวมทั้งการเกิดภัยธรรมชาติที่รุนแรง



ตัวอย่างปัญหาที่เป็นโจทย์ตั้งต้น: การบริหารจัดการน้ำ

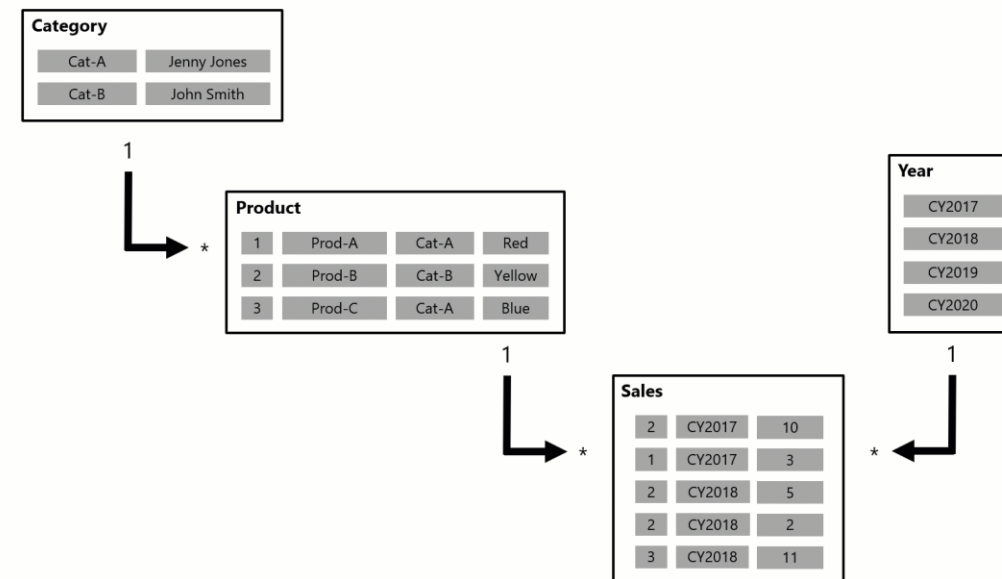
•





ตัวอย่าง ตารางความสัมพันธ์

- ช่วยโยงไปยังวัตถุหรือนามธรรมที่สัมพันธ์กัน
- นำไปสู่การหาคำตอบเชิงสัมพันธ์
- ความเกี่ยวเนื่องต่อกันเป็น เหตุ เป็นผล



Key
(ต้องตรงกัน)

Code &
HCL

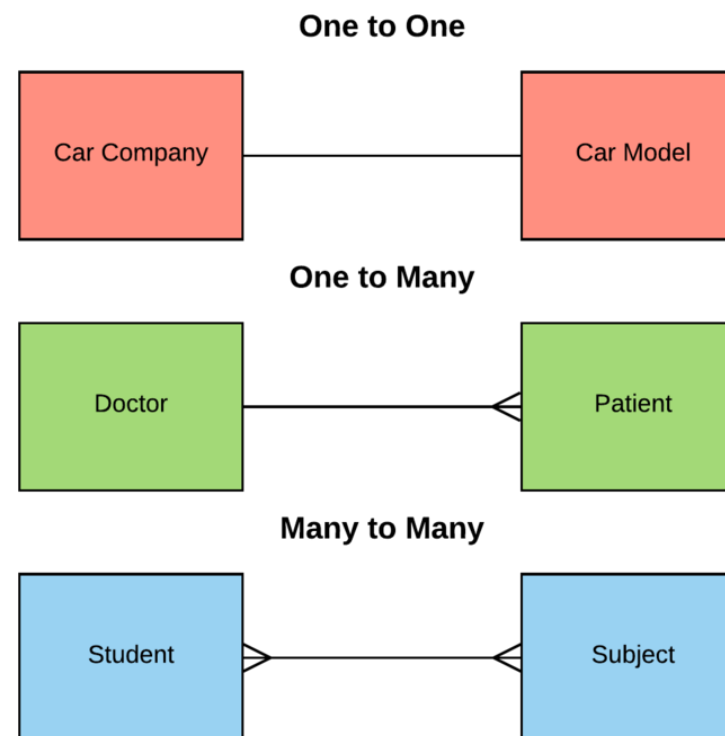
อ้างอิง : <https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-relationships-understand>

ความสัมพันธ์ระหว่างตารางข้อมูล

ตารางข้อมูล

ตัวแปร/คอลัมน์

วัตถุ/นามธรรม



หลักสำคัญของเรื่องท่องเที่ยว

- รายได้ : จำนวนคนที่มาเที่ยว , ค่าบริการ, ต้นทุน
- แหล่งท่องเที่ยว : ประเภทแหล่งท่องเที่ยว
- หาลูกค้าใหม่ : กลุ่มลูกค้า
- การสร้างแหล่งท่องเที่ยวใหม่
- การลดปัญหาที่เป็นอุปสรรค
- การส่งเสริมการท่องเที่ยว

ดูชุดข้อมูลบน
Catalog

ทบทวน
ความสัมพันธ์

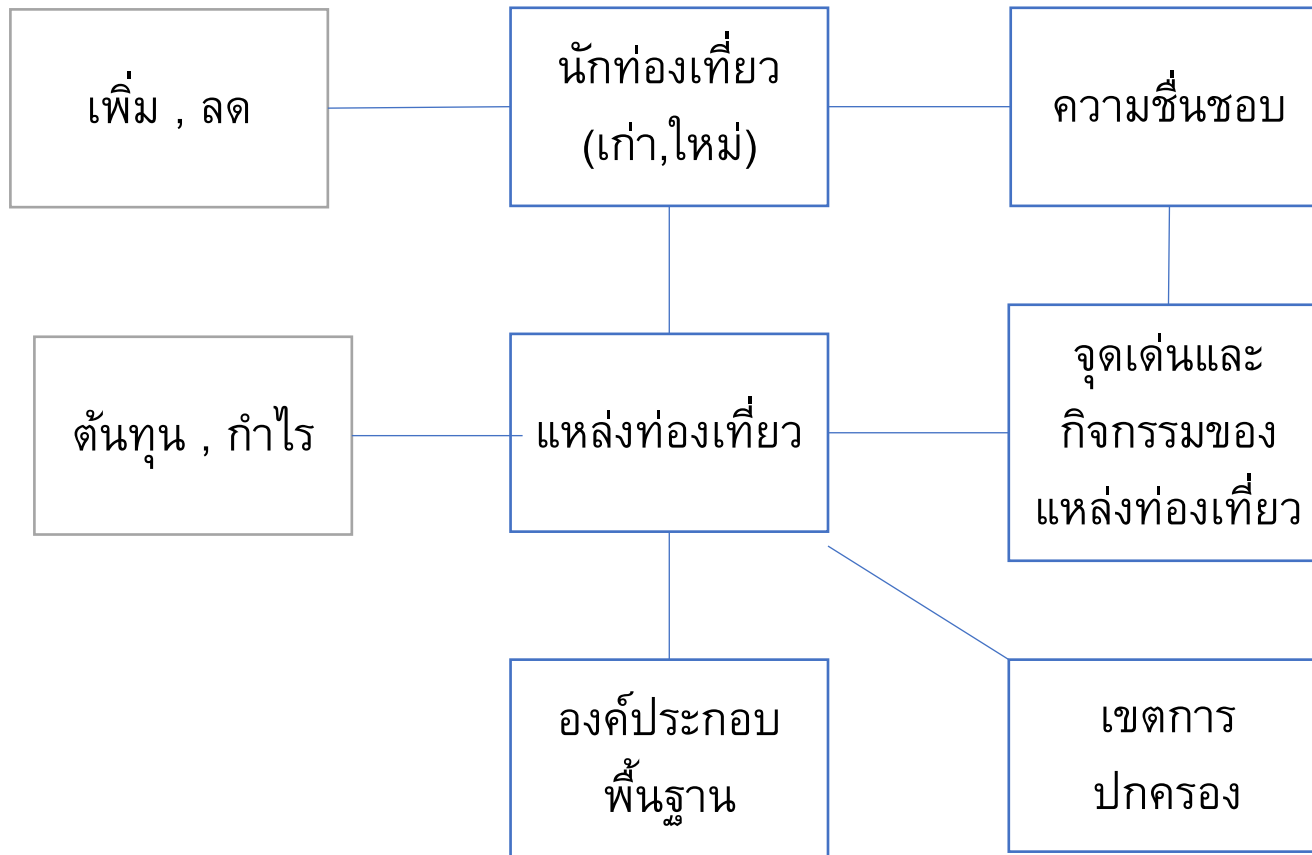
หลักสำคัญของเรื่องการเกษตร

- เกษตรกร :
- ผลิต/การปลูก :
- อัตลักษณ์ของผลผลิต :
- ต้นทุน :
- การแปรรูป :
- รายได้ :
- ตลาดจำหน่าย :
- กลุ่มลูกค้า :
- การลดปัญหาที่เป็นอุปสรรค :
- การส่งเสริม :

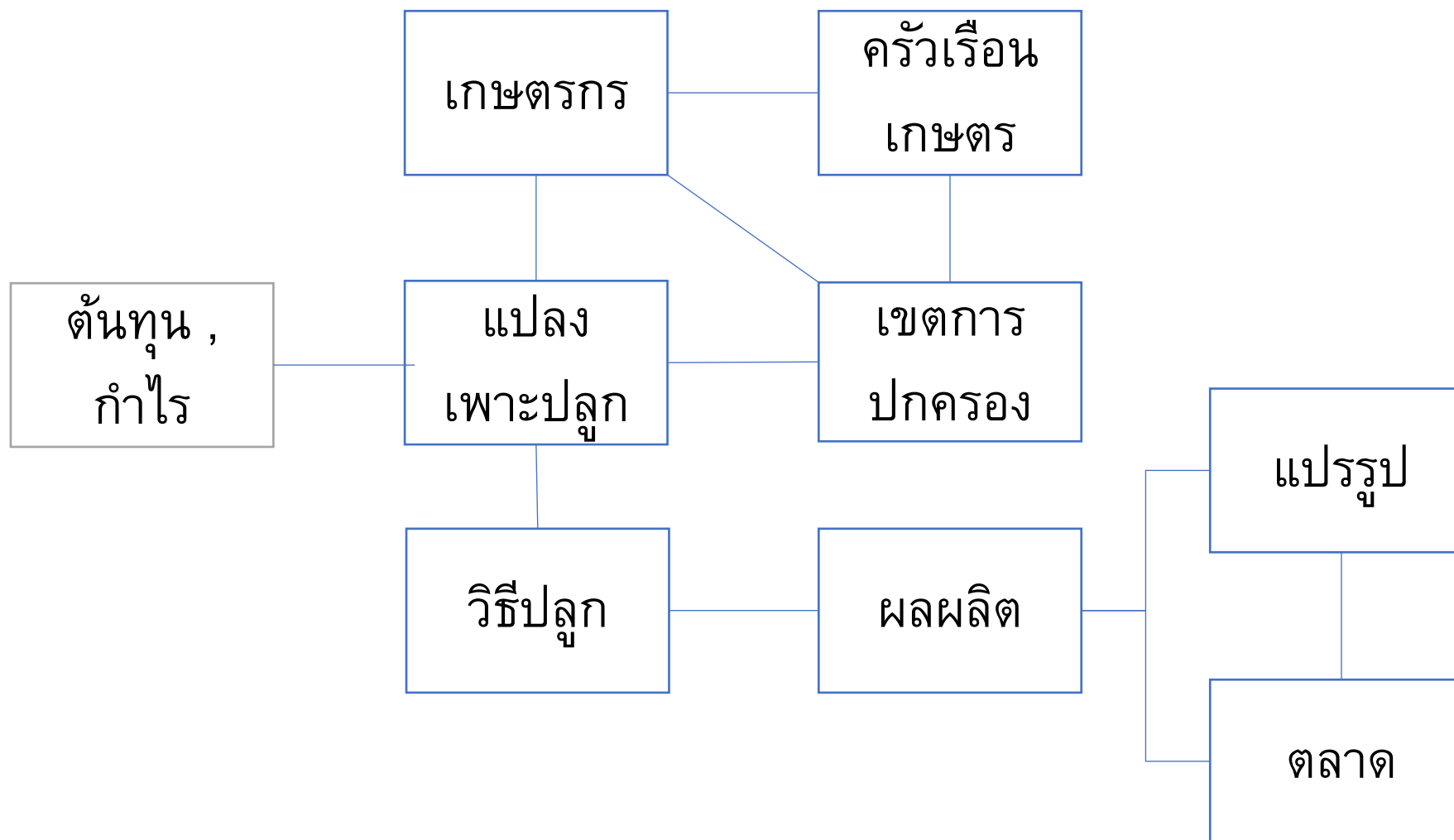
ดูชุดข้อมูลบน
Catalog

ทบทวน
ความสัมพันธ์

ปรับปรุง Data Model : ท้องเที่ยว

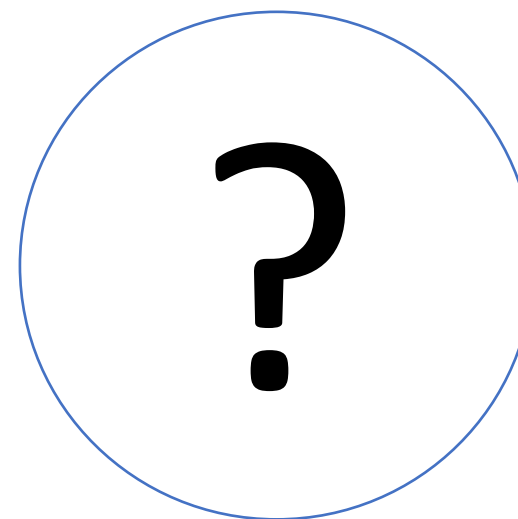


ปรับปรุง Data Model : เกษตร





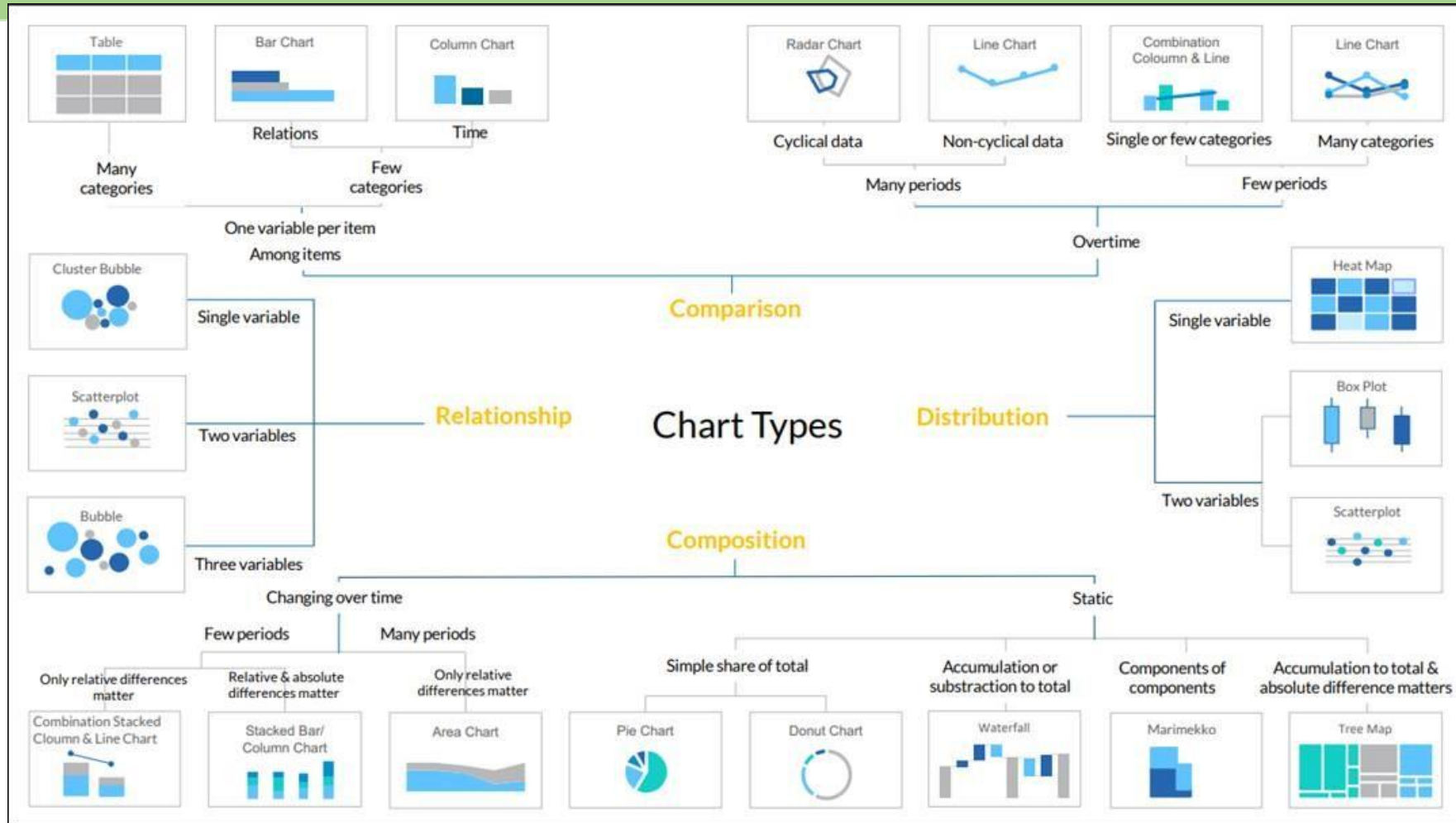
ทำ Data Model :



ภาพ Visualization บอกอะไรได้

- ต้องการสื่อสารบอกถึงลำดับ
- ต้องการสื่อสารการเปลี่ยนแปลง
- ต้องการสื่อสารการจัดกลุ่ม
- ต้องการดูความสัมพันธ์ข้อมูล
- ต้องการดูการกระจายตัวข้อมูล
- การเปลี่ยนแปลงในแต่ละช่วงเวลา
- หาทิศทางหรือแนวโน้ม
- การเปรียบเทียบ
- การอ้างอิงกับพื้นที่

ผังกราฟเพื่อการสื่อสารสารสนเทศ



<https://blogs.sap.com/2020/11/13/sap-successfactors-people-analytics-report-stories-data-visualization-series-selecting-the-right-chart-part-1/>

ตัวแปรที่ใช้ในการวิเคราะห์

- ใช้ตัวแปรเดียว
 - ลำดับ
- ใช้สองตัวแปร
 - ผังกราฟเส้น
- หลายตัวแปร

ตัวอย่างการรายงานประเด็นท่องเที่ยว : ผู้มีส่วนได้เสียโดยตรง

Report-1

จำนวนนักท่องเที่ยว (ไทย / ต่างชาติ)
เที่ยวที่ไหน ?
จำนวนครั้งที่มา ?

Report-4

ระยะเวลาที่พัก (ไทย / ต่างชาติ)
พักที่ไหน ?
จำนวนที่พัก ?

Report-7

ปัญหาพื้นฐานของเมือง

Report-2

แหล่งท่องเที่ยว เที่ยวที่ไหน (ไทย /
ต่างชาติ)
กิจกรรม ? ประเภท ? จำนวน ? การ
ประชาสัมพันธ์

Report-5

การอำนวยความสะดวก และ
ความปลอดภัย

Report-8

ความพึงพอใจของนักท่องเที่ยว
ความชื่นชอบ / ผ่านช่องทาง **social**

Report-3

การเดินทางของนักท่องเที่ยว ?

Report-6

ค่าใช้จ่าย ค่าขนส่ง ค่าอาหาร ค่าที่พัก
ความสะดวก

Report-9

มาตรฐานการท่องเที่ยว การสร้างความ
เชื่อมั่น

ตัวอย่างการรายงานประเด็นท่องเที่ยว : เจ้าหน้าที่รัฐ

Report-1

ปัญหาที่พื้นฐานของเมือง

Report-2

การอำนวยความสะดวกและ
ความปลอดภัย

Report-3

มาตรฐานการท่องเที่ยว
การสร้างเชื่อมั่น

ตัวอย่างการรายงานประเด็นท่องเที่ยว : ผู้บริหาร

Report-1

รายได้จากการท่องเที่ยวเพิ่ม/ลด

Report-2

จำนวนนักท่องเที่ยว เพิ่ม/ลด

Report-3

การประมาณการนักท่องเที่ยว ?

Report-4

มาตรฐานการท่องเที่ยว

Report-5

ค่าใช้จ่าย ค่าขนส่ง ค่าอาหาร
ค่าที่พัก ความสะอาด

Report-6

ความพึงพอใจของนักท่องเที่ยว
ความชื่นชอบ

ตัวอย่างการรายงานประเด็นเกษตร : ผู้มีส่วนได้เสียโดยตรง



Report-1



Report-4



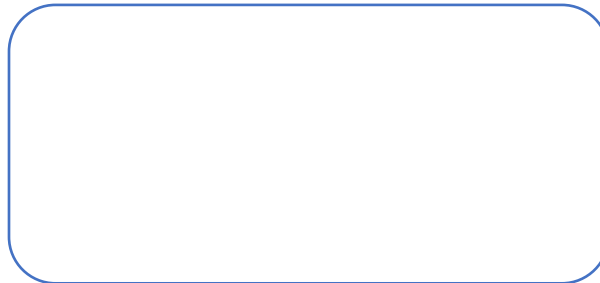
Report-7



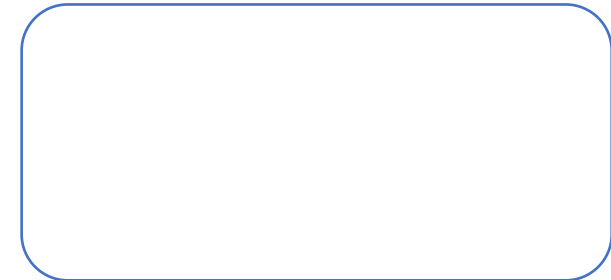
Report-2



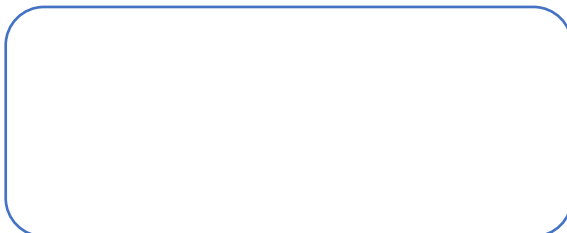
Report-5



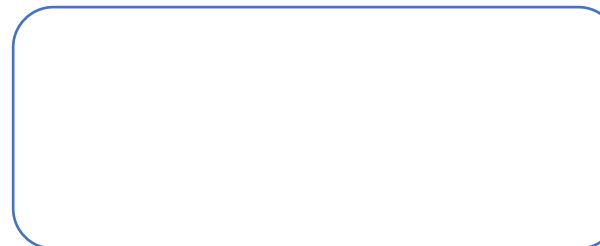
Report-8



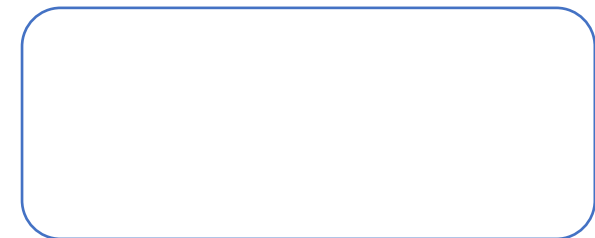
Report-3



Report-6



Report-9



ตัวอย่างการรายงานประเด็นเกษตร : เจ้าหน้าที่รัฐ



Report-1



Report-4



Report-7



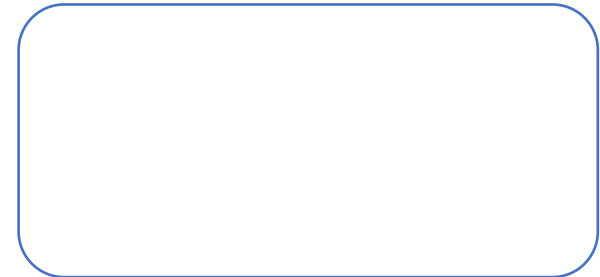
Report-2



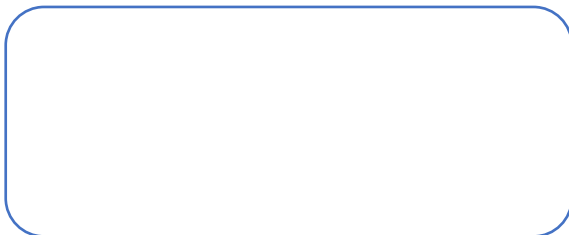
Report-5



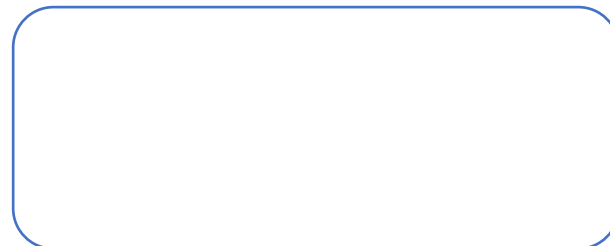
Report-8



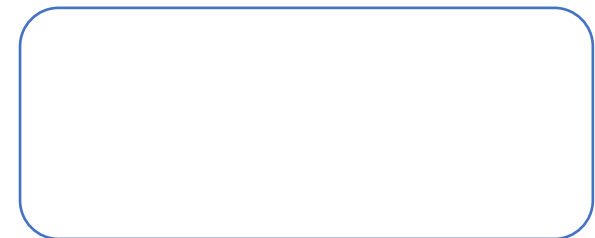
Report-3



Report-6



Report-9



ตัวอย่างการรายงานประเด็นเกษตร : ผู้บริหาร



Report-1



Report-4



Report-7



Report-2



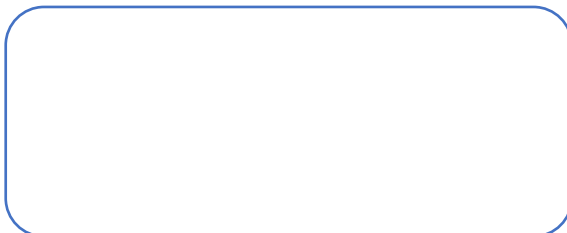
Report-5



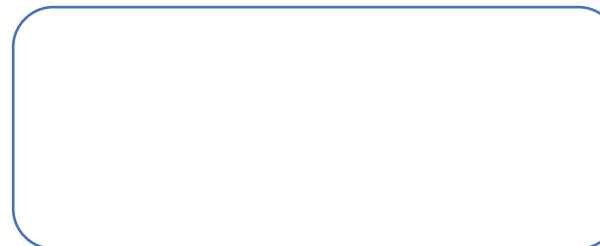
Report-8



Report-3



Report-6



Report-9



ควรมีอะไรบอก

- แหล่งที่มา
- หน่วยนับ
- คำอธิบายแทน
- สเกลแทน



การวางแผนการทำ Report

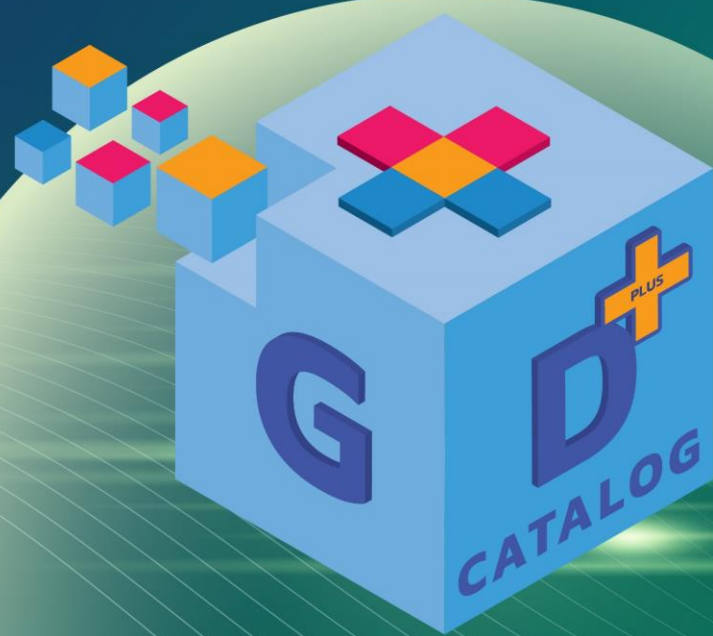
- จำนวน Report มีเท่าไร ?
- คนทำมีกี่คน ?
- ระยะเวลาที่ใช้
- การ Link Report

Create-Project

- province_y66_fp : (หน้ารวม)
- province_y66_p1 : (สำหรับประเด็นที่ 1)
- province_y66_p2 : (สำหรับประเด็นที่ 2)
- วางแผนการทำรายงาน

การเลือกรูปแบบ Visualize

- <https://mindmajix.com/power-bi-visualization-types#what-are-the-different-types>
- <https://s3.amazonaws.com/assets.datacamp.com/email/other/Data+Visualizations+-+DataCamp.pdf>



GOVERNMENT DATA CATALOG