Session 03

Data Visualization with Gemini in Colab

Stat on Campus: Data innovators

Turning data into impact for public sector

Taweesak Samanchuen, Ph.D.

Mahidol University

วัตถุประสงค์ของ Session

เมื่อจบช่วงนี้ ผู้เข้าอบรมสามารถ:

1. เลือกประเภทกราฟให้เหมาะกับชนิดข้อมูลและเป้าหมาย
2. ใช้ Gemini in Colab ช่วยออกแบบและสร้างกราฟได้รวดเร็วขึ้น
3. ตีความผลจากกราฟอย่างถูกต้องและชัดเจน
4. สื่อสารข้อมูลให้ผู้บริหารและผู้เกี่ยวข้องเข้าใจได้ง่าย

ภาพรวมเนื้อหา

1. หลักการเลือกประเภทกราฟ
2. แนวคิดการสื่อสารข้อมูลด้วย Visualization
3. การใช้ AI ช่วยสร้างกราฟจากข้อมูล
4. การตีความและเล่าเรื่องจากกราฟ
5. Workshop 3

แพลตฟอร์มที่ใช้ในชั้นเรียน

1. Gemini in Colab: ช่วยคิด วิเคราะห์ และเขียนโค้ดในโน้ตบุ๊กเดียว
2. GitHub Raw CSV: ใช้เป็นแหล่งข้อมูลกลางให้ทุกคนโหลดข้อมูลชุดเดียวกัน
3. Visualization Libraries: ใช้ `pandas`, `matplotlib`, `seaborn`

1) หลักการเลือกประเภทกราฟ

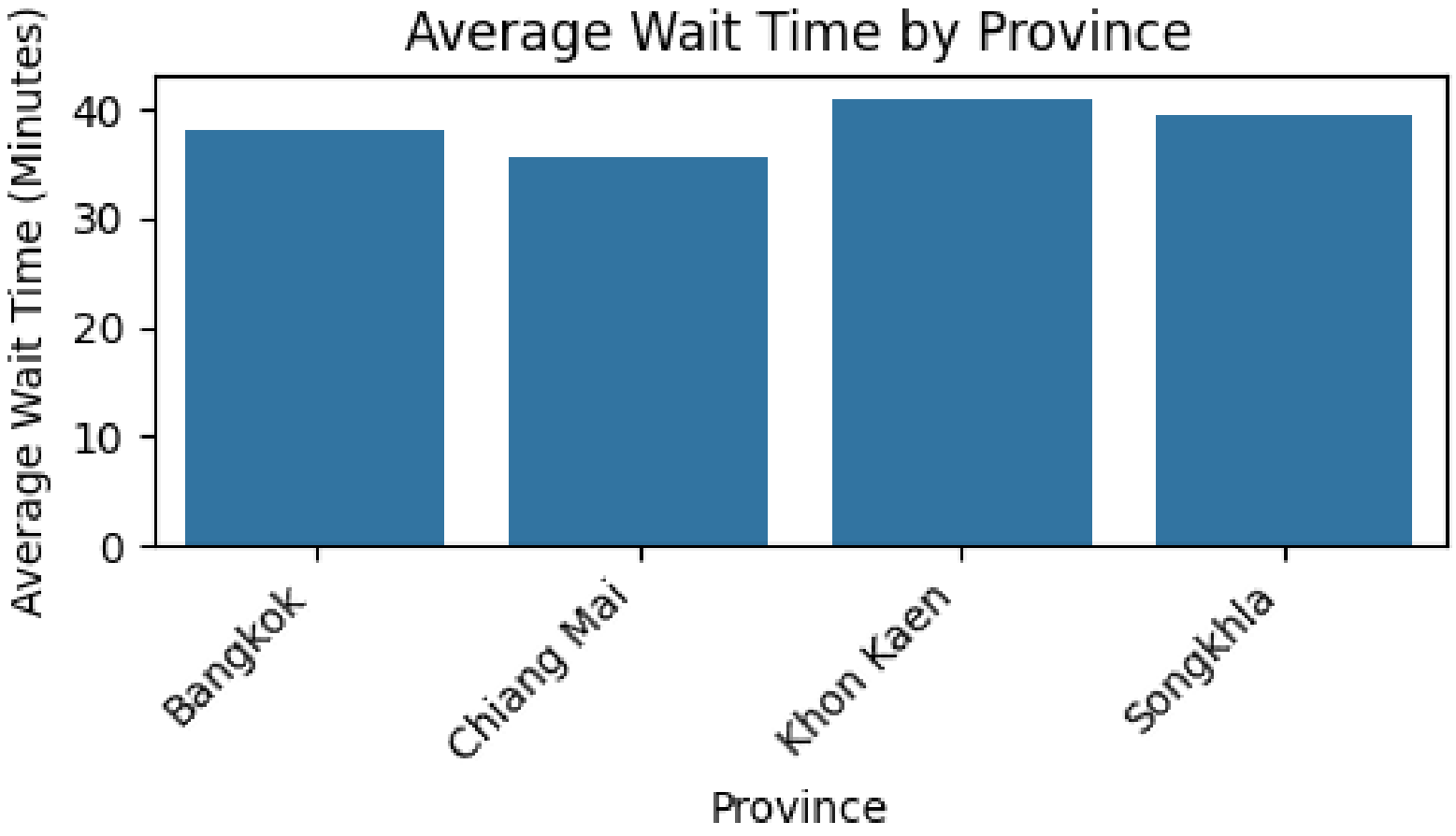
คำถามก่อนเลือกกราฟ

- ต้องการเปรียบเทียบ, แสดงแนวโน้ม, หรือแสดงสัดส่วน?
- ข้อมูลเป็นหมวดหมู่, ตัวเลขต่อเนื่อง, หรือเวลา?
- ผู้ชมต้องการตัดสินใจเรื่องใดจากกราฟนี้?

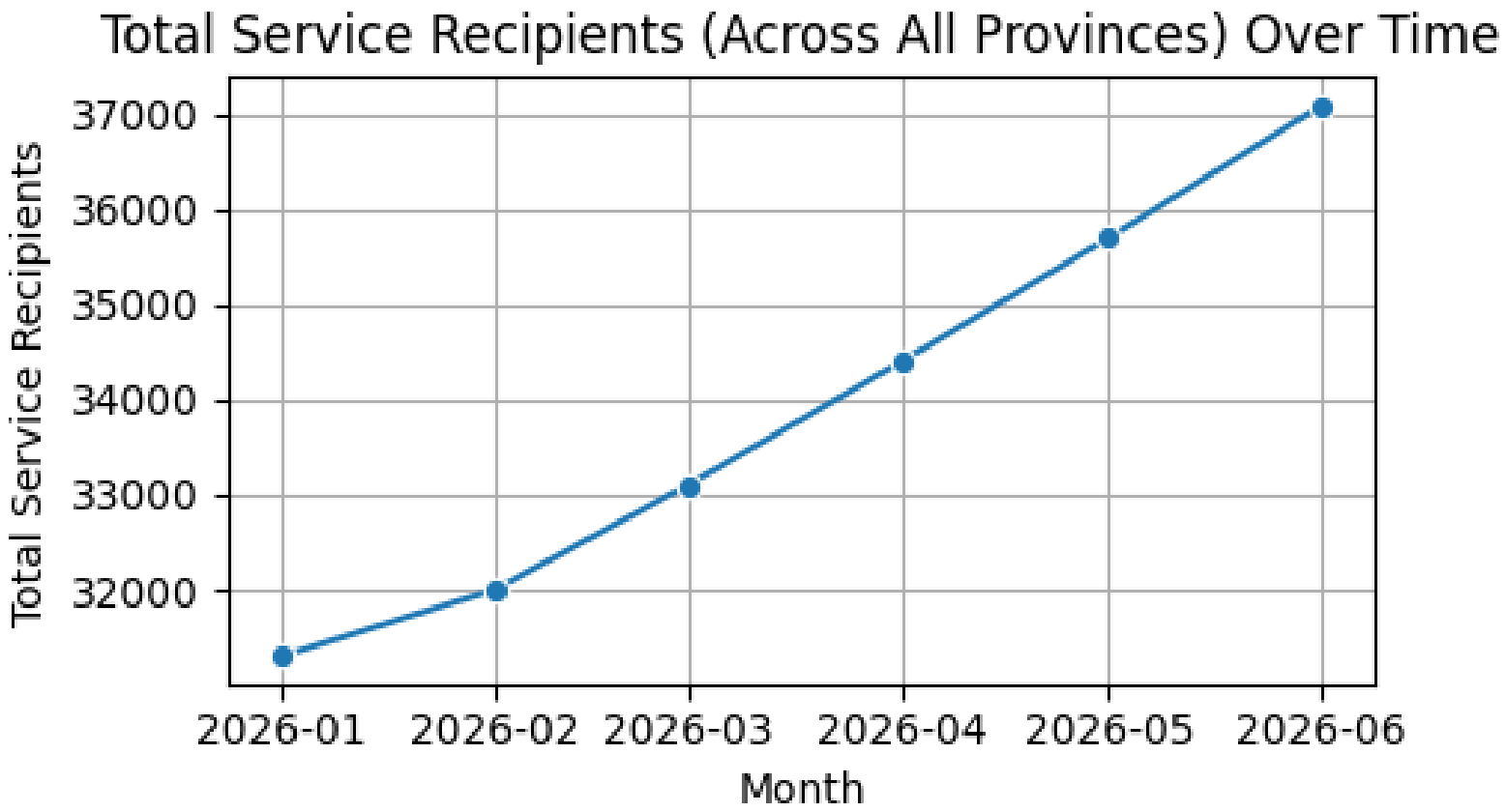
แนวทางใช้งาน

- Bar chart: เปรียบเทียบค่าระหว่างกลุ่ม
- Line chart: แนวโน้มตามเวลา
- Scatter plot: ความสัมพันธ์ระหว่างตัวแปร
- Heatmap: รูปแบบ/ความเข้มข้นในหลายมิติ

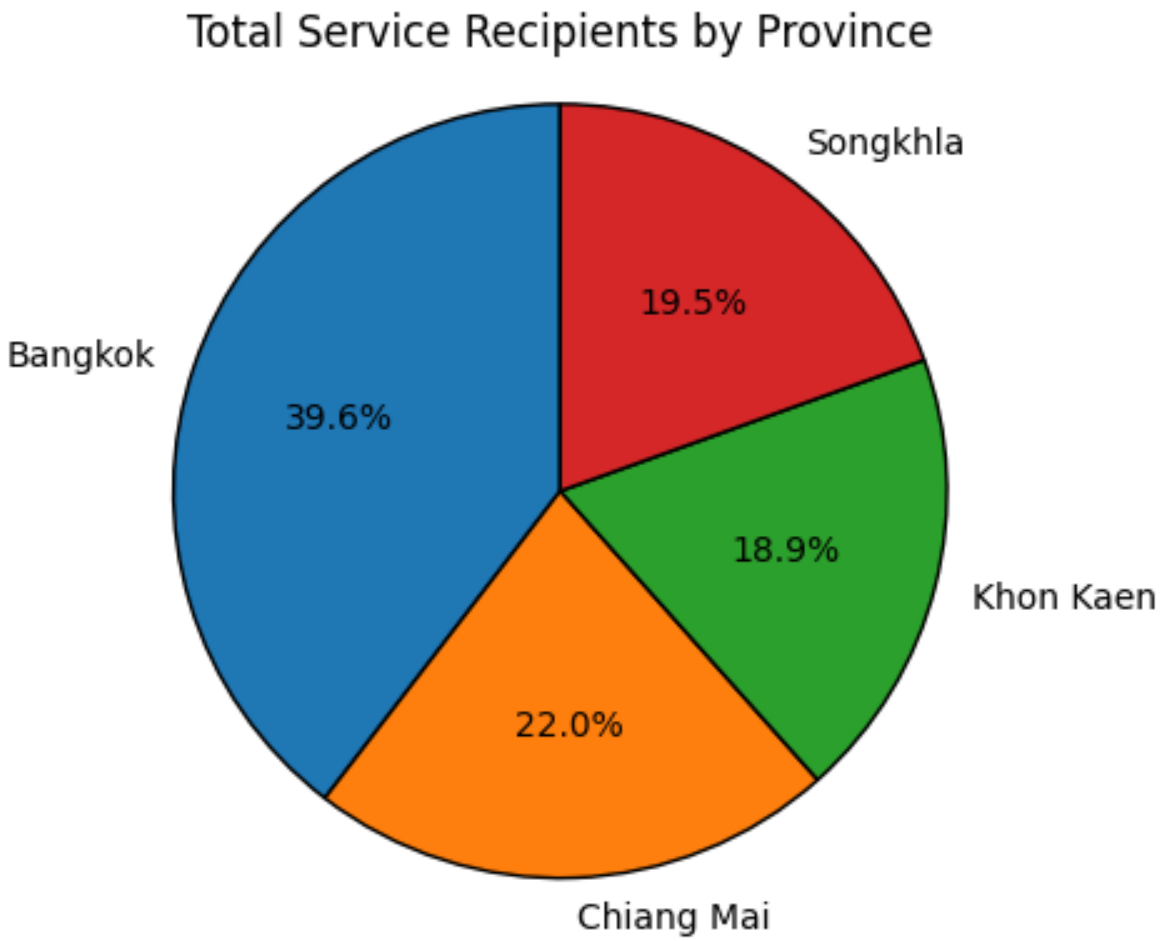
Bar chart: เปรียบเทียบค่าระหว่างกลุ่ม



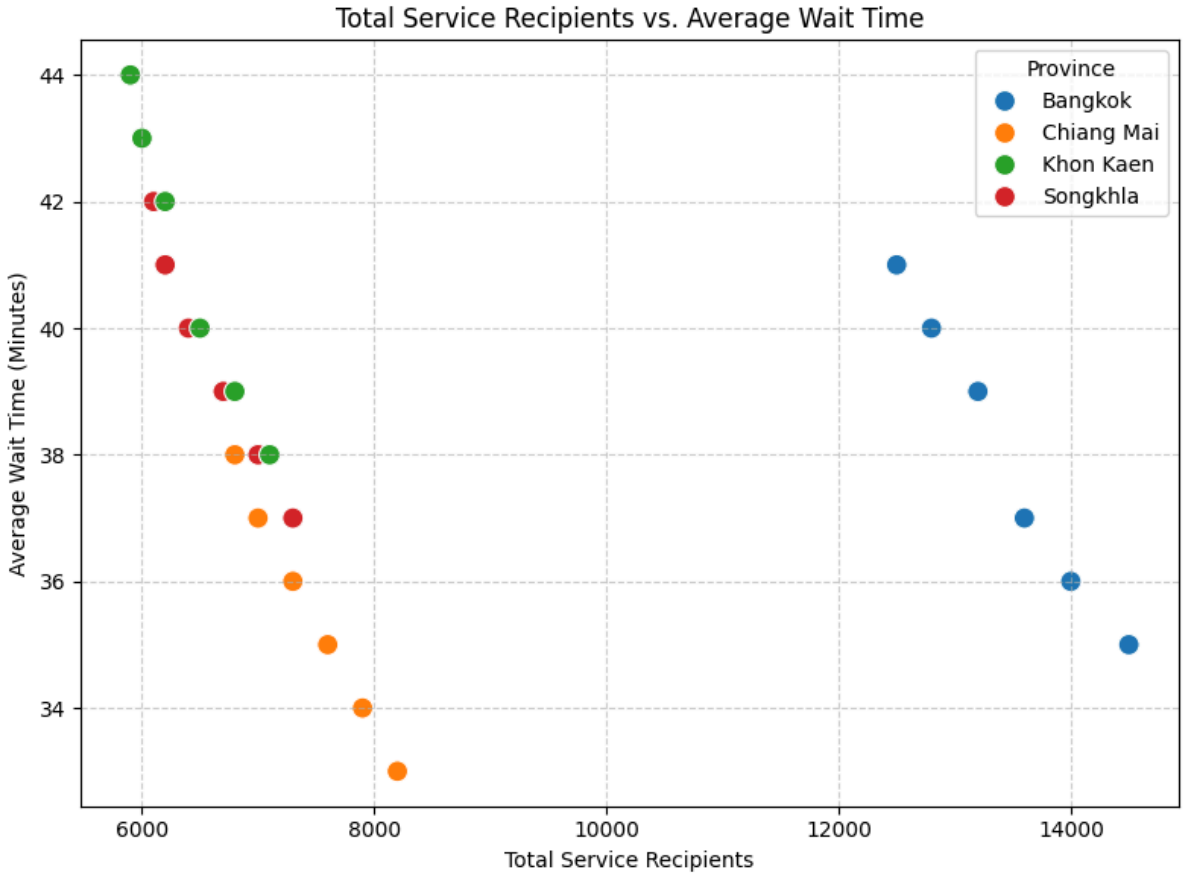
Line chart: แนวโน้มตามเวลา



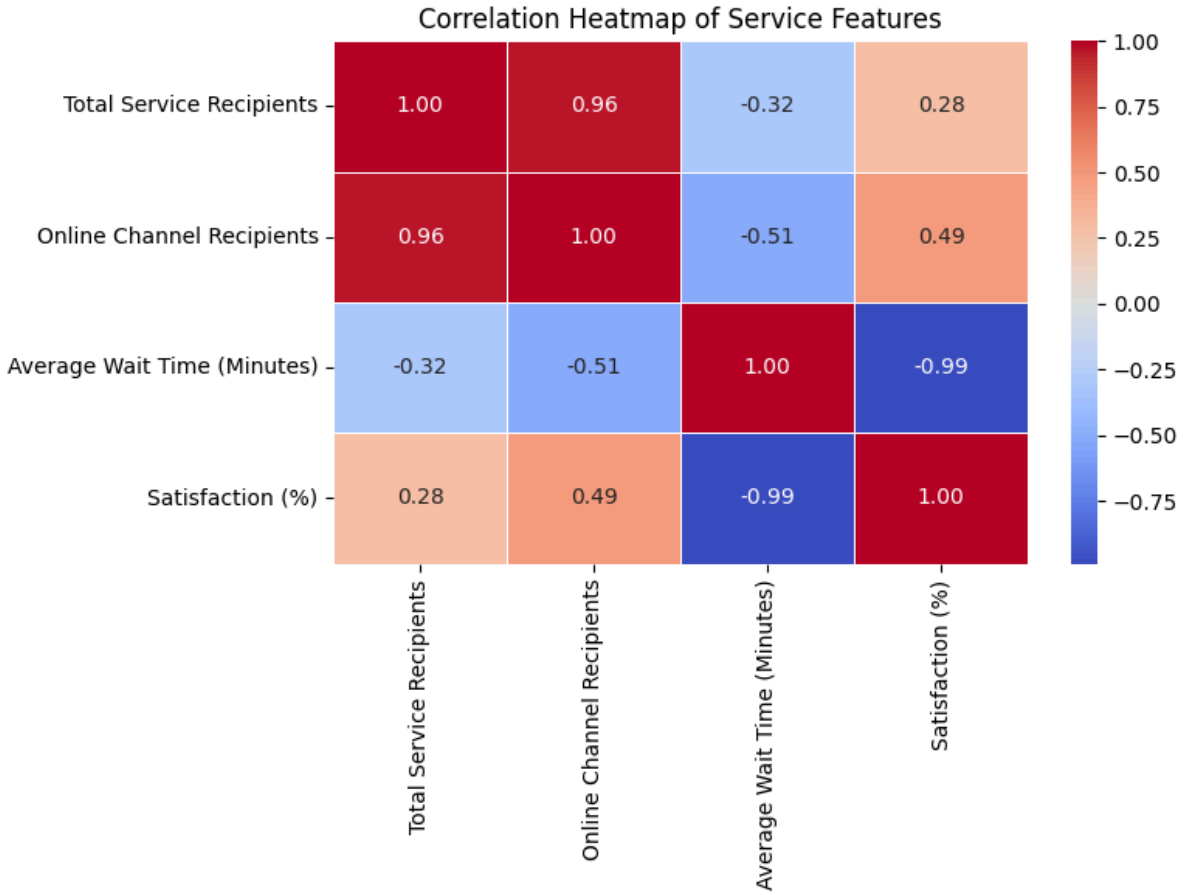
Pie chart: แสดงสัดส่วนของแต่ละกลุ่มในภาพรวม



Scatter plot: ความสัมพันธ์ระหว่างตัวแปร



Heatmap: รูปแบบ/ความเข้มข้นในหลายมิติ



ตัวอย่าง: คำถามแบบไหน ใช้กราฟอะไร

ใช้ชุดข้อมูลผู้รับบริการภาครัฐรายเดือน

คำถาม	กราฟที่เหมาะ	สิ่งที่อยากเห็น
จังหวัดไหนมีผู้รับบริการมากที่สุดในเดือนล่าสุด	Bar chart	เปรียบเทียบขนาดระหว่างจังหวัด
จำนวนผู้รับบริการเพิ่มหรือลดลงอย่างไรใน 6 เดือน	Line chart	แนวโน้มและจังหวะการเปลี่ยนแปลง
เวลารอที่ลดลง ทำให้ความพึงพอใจสูงขึ้นหรือไม่	Scatter plot	ความสัมพันธ์ระหว่าง 2 ตัวแปร
จังหวัดใดมีจำนวนผู้รับบริการสูงต่อเนื่องหลายเดือน	Heatmap	จุดที่เข้ม/อ่อนในหลายมิติ

ตัวอย่างการตีความจากข้อมูลเดียวกัน

ตัวอย่างข้อความสรุปที่ “ดีพอสำหรับนำเสนอ”

- Bar chart: กรุงเทพมหานครมีผู้รับบริการสูงสุดในเดือนล่าสุด จึงอาจต้องจัดทรัพยากรหน้าจุดบริการมากกว่าจังหวัดอื่น
- Line chart: ทุกจังหวัดมีแนวโน้มผู้รับบริการเพิ่มขึ้นต่อเนื่อง สะท้อนว่าความต้องการบริการยังขยายตัว
- Scatter plot: เมื่อเวลารอลดลง ความพึงพอใจมีแนวโน้มสูงขึ้น แต่ยังไม่พอจะสรุปว่าเป็นเหตุโดยตรง
- Heatmap: เชียงใหม่และกรุงเทพมหานครมีความเข้มสูงหลายเดือนต่อเนื่อง จึงควรติดตามกำลังคนและช่องทางออนไลน์ควบคู่กัน

สิ่งที่ต้องระวัง

- อย่าสรุปว่า “เวลารอลดลงทำให้ความพึงพอใจสูงขึ้นแน่นอน” โดยไม่มีปัจจัยอื่นประกอบ
- อย่าดูเฉพาะจังหวัดที่ค่าสูงสุด แต่ให้ดูอัตราการเปลี่ยนแปลงและบริบทของพื้นที่ด้วย

2) Data Visualization เพื่อการสื่อสาร

หลักการสำคัญ

- ชัดเจน: ลดองค์ประกอบที่ไม่จำเป็น
- ถูกต้อง: ไม่บิดเบือนสเกลหรือบริบท
- มีสาร: ทุกกราฟต้องตอบคำถามหลักได้

โครงสร้างการเล่าเรื่องข้อมูล

1. บริบทของปัญหา
2. หลักฐานจากข้อมูล
3. Insight และข้อเสนอแนะ

กรณีศึกษาประสิทธิภาพการบริการ

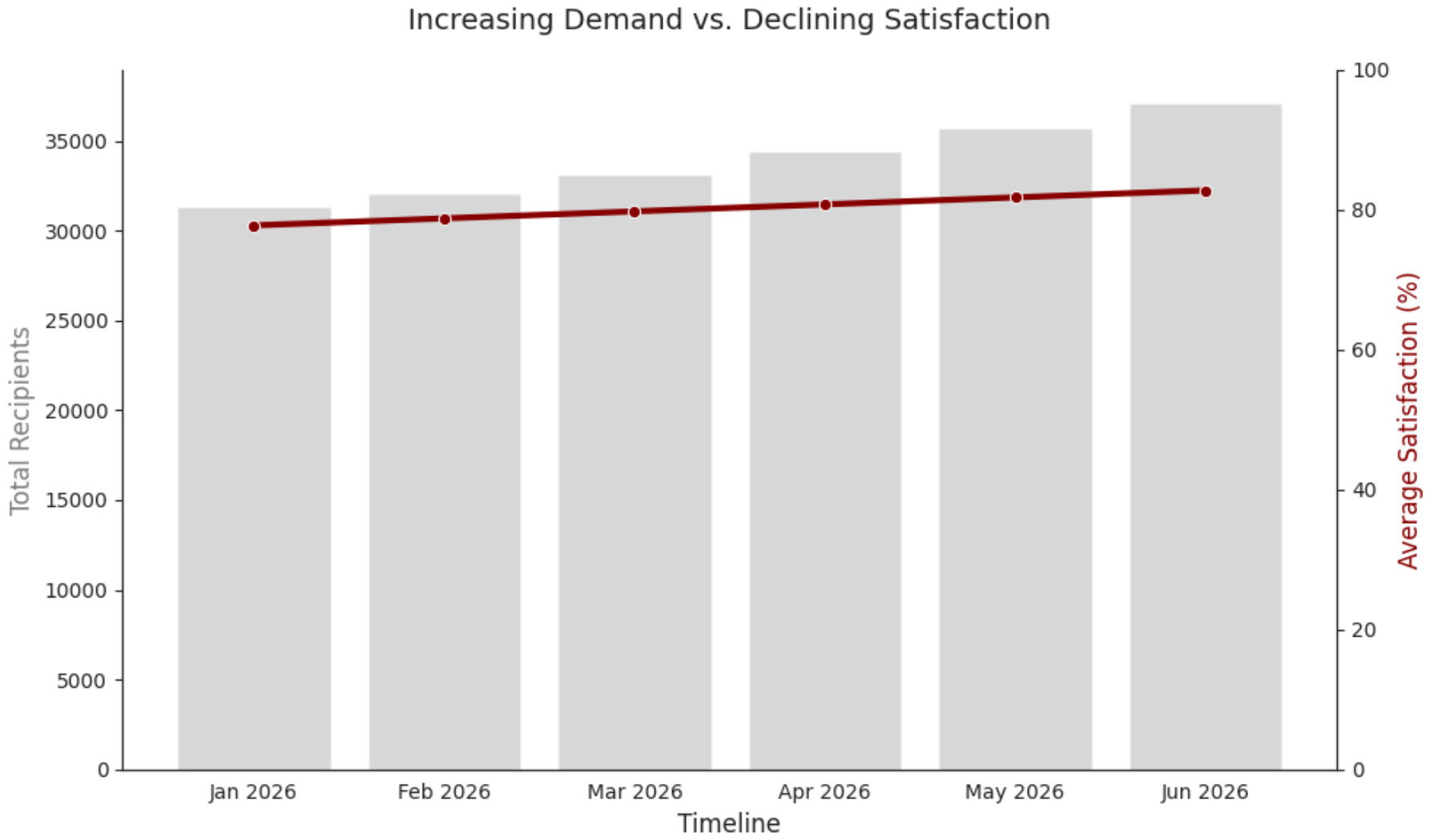
1. บริบทของปัญหา (Context)

เราต้องการตรวจสอบว่า 'จำนวนผู้รับบริการที่เพิ่มขึ้น ส่งผลกระทบต่อคุณภาพการบริการและความพึงพอใจหรือไม่' เพื่อวางแผนการจัดการทรัพยากรในอนาคต

2. หลักฐานจากข้อมูล (Evidence)

ใช้กราฟที่เน้นความ ชัดเจน โดยลด Grid lines ที่ไม่จำเป็น และเน้นความ ถูกต้อง ของสเกล

Case Study: การสื่อสารข้อมูลด้วย Visualization



3) ใช้ Gemini in Colab ช่วยสร้างกราฟ

ตัวอย่างการใช้งาน

- ช่วยร่างโค้ดสร้างกราฟจากตารางข้อมูล
- แนะนำกราฟที่เหมาะกับตัวแปรแต่ละชุด
- ช่วยปรับแต่ง label/title/annotation ให้ชัดเจน
- สร้างสรุปข้อความประกอบกราฟอัตโนมัติ

Prompt ที่แนะนำ

Prompt ที่แนะนำ

จากข้อมูลนี้ ช่วยเสนอ 3 กราฟที่เหมาะสม พร้อมเหตุผล และตัวอย่างโค้ด

Prompt ตัวอย่างที่ชัดขึ้น

ข้อมูลนี้เป็นจำนวนผู้รับบริการภาครัฐรายเดือน แยกตามจังหวัด

ช่วยเลือกกราฟที่เหมาะ 4 แบบ โดยอธิบายว่ากราฟแต่ละแบบตอบคำถามผู้บริหารอะไรได้

และเขียนโค้ด Python ที่พร้อมรันใน Colab"

4) การตีความกราฟอย่างมีคุณภาพ

เช็กลิสต์ก่อนสรุปผล

- ความแตกต่างที่เห็นมีนัยสำคัญเชิงปฏิบัติหรือไม่
- มีปัจจัยแทรกซ้อนที่ยังไม่ถูกควบคุมหรือไม่
- ข้อสรุปสอดคล้องกับนิยามตัวชี้วัดหรือไม่
- มีข้อจำกัดข้อมูลที่ต้องแจ้งหรือไม่

ข้อผิดพลาดที่พบบ่อย

- สรุปเหตุ-ผลจากความสัมพันธ์เพียงอย่างเดียว
- ใช้กราฟ 3D หรือเอฟเฟกต์ที่ทำให้ตีความคลาดเคลื่อน

Workshop 4

กิจกรรมฝึกปฏิบัติ

1. เลือกคำถามวิเคราะห์จาก Session 02
2. ออกแบบกราฟอย่างน้อย 3 แบบ
3. ใช้ Gemini in Colab ช่วยสร้าง/ปรับโค้ด
4. สรุปข้อความ insight ต่อกราฟละ 1-2 ประโยค

ผลลัพธ์ที่คาดหวัง

- ชุดกราฟที่พร้อมนำเสนอ
- ข้อความอธิบายที่กระชับและใช้งานได้จริง

ติดตั้ง font ภาษาไทยใน Colab

Prompt

```
ช่วยเขียนโค้ด Python ใน Colab เพื่อดาวน์โหลดและติดตั้งฟอนต์ภาษาไทยที่เหมาะกับการสร้างกราฟใน Matplotlib
และทดสอบการแสดงผลด้วยข้อความภาษาไทย
```

- เลือก model ที่มีเก่งในการทำเรื่องนี้เพราะมีความซับซ้อนในการจัดการฟอนต์และการแสดงผลภาษาไทยในกราฟ ซึ่งอาจต้องใช้ความรู้เฉพาะทางในการตั้งค่าฟอนต์ใน Matplotlib และการจัดการกับภาษาไทยใน Python โดยตรง
- ให้เลือกฟอนต์ที่นิยมใช้สำหรับการนำเสนอในภาษาไทย เช่น "TH Sarabun New" หรือ "Kanit" และช่วยเขียนโค้ดที่พร้อมรันได้เลยใน Colab เพื่อให้ผู้เรียนสามารถใช้ฟอนต์นี้ในการสร้างกราฟและแสดงผลภาษาไทยได้อย่างถูกต้อง

ติดตั้ง font ภาษาไทยใน Colab

```
# ดาวน์โหลดฟอนต์ TH Sarabun New
!wget -q https://github.com/Phonbopit/sarabun-webfont/raw/master\
/fonts/thsarabunnew-webfont.ttf
import matplotlib as mpl
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm

# เพิ่มฟอนต์เข้าระบบ Matplotlib
fm.fontManager.addfont('thsarabunnew-webfont.ttf')
# ตั้งค่า default font ให้เป็น TH Sarabun New
mpl.rc('font', family='TH Sarabun New')
```

ทดสอบการแสดงผลภาษาไทย

```
# ทดสอบการพล็อตภาษาไทย
plt.figure(figsize=(6, 4))
plt.text(0.5, 0.5, 'สวัสดีภาษาไทย (Thai Language Test)', fontsize=25, ha='center', va='center')
plt.title('ทดสอบการแสดงผลภาษาไทย')
plt.axis('off')
plt.show()
```

ผลลัพธ์ที่คาดหวัง

ทดสอบการแสดงผลภาษาไทย

สวัสดีภาษาไทย (Thai Language Test)

Workshop 4: โหลดข้อมูลจาก GitHub

Dataset สำหรับฝึก

- Raw CSV: `https://raw.githubusercontent.com/toche7/DataSets/main/session-03-workshop-sample-data.csv`
- แพลตฟอร์มหลัก: Gemini in Colab
- เครื่องมือ: `pandas`, `matplotlib`, `seaborn`
- ใช้ชุดข้อมูลเดียวกันเพื่อสาธิตกราฟ 4 ประเภท

Prompt ที่ให้ผู้เรียนใช้

```
โหลดข้อมูลจากไฟล์ CSV นี้ และช่วยสร้างกราฟ 4 ประเภท ได้แก่ bar chart, line chart, scatter plot
และ heatmap แยก เป็นภาพ ๆ พร้อมอธิบายว่าแต่ละกราฟตอบคำถามเชิงนโยบายอะไรได้บ้าง และช่วยปรับชื่อกราฟ
แกน และสีให้เหมาะกับการนำเสนอผู้บริหาร:
```

python code ตัวอย่างการโหลดข้อมูลและเริ่มต้นวิเคราะห์

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
url = "https://raw.githubusercontent.com/toche7/DataSets/main/session-03-workshop-sample-data
df = pd.read_csv(url)
df["เดือน"] = pd.to_datetime(df["เดือน"])
sns.set_theme(style="whitegrid")
df.head()
```

ตัวอย่าง 1-2: Bar และ Line

Bar chart: เปรียบเทียบจังหวัดในเดือนล่าสุด

```
latest = df[df["เดือน"] == df["เดือน"].max()]
sns.barplot(data=latest, x="จังหวัด", y="จำนวนผู้รับบริการทั้งหมด(คน)")
plt.xticks(rotation=15)
plt.title("จำนวนผู้รับบริการทั้งหมด แยกตามจังหวัด")
plt.show()
```

Line chart: แนวโน้มรายเดือนของผู้รับบริการ

```
sns.lineplot(data=df, x="เดือน", y="จำนวนผู้รับบริการทั้งหมด(คน)", hue="จังหวัด", marker="o")
plt.title("แนวโน้มผู้รับบริการรายเดือน")
plt.show()
```

ตัวอย่าง 3-4: Scatter และ Heatmap

Scatter plot: เวลารอกับความพึงพอใจ

```
sns.scatterplot(data=df, x="เวลารอเฉลี่ย(นาที)", y="ความพึงพอใจ(%)", hue="จังหวัด", s=120)
plt.title("ความสัมพันธ์ระหว่างเวลารอและความพึงพอใจ")
plt.show()
```

Heatmap: ผู้รับบริการตามเดือนและจังหวัด

```
pivot = df.pivot(index="จังหวัด", columns="เดือน", values="จำนวนผู้รับบริการทั้งหมด(คน)")
sns.heatmap(pivot, cmap="YlGnBu", annot=True, fmt=".0f")
plt.title("Heatmap จำนวนผู้รับบริการ")
plt.show()
```

Prompt ให้ Gemini ช่วยต่อยอด

```
> ใช้ Gemini in Colab กับข้อมูลจากไฟล์ CSV นี้เพื่อสร้าง bar chart, line chart, scatter plot
และ heatmap พร้อมอธิบายว่าแต่ละกราฟตอบคำถามเชิงนโยบายอะไรได้บ้าง และช่วยปรับชื่อกราฟ แกน
และสีให้เหมาะกับการนำเสนอผู้บริหาร
```

จุดที่ให้ผู้เรียนสังเกต

- จังหวัดใดมีปริมาณผู้รับบริการสูงสุดอย่างสม่ำเสมอ
- เวลารอที่ลดลงสัมพันธ์กับความพึงพอใจหรือไม่
- จังหวัดใดมีโอกาสเพิ่มสัดส่วนบริการออนไลน์ได้อีก

สรุป Session 03

- กราฟที่ดีต้องตอบคำถามเชิงนโยบายได้
- AI ช่วยเร่งขั้นตอน แต่การตีความยังต้องอาศัยผู้เชี่ยวชาญ
- Visualization คือสะพานจากข้อมูลสู่การตัดสินใจ

Q&A

เตรียมต่อ Session 04: AI-Assisted Reporting and Presentation