

Project Document G.2

รายวิชา: Data Analytics and Business Intelligence

ชุดข้อมูล: Customer Shopping Trends Dataset - kaggle

วัตถุประสงค์: ประยุกต์ใช้ทักษะการวิเคราะห์ข้อมูลแบบครบวงจร เพื่อสกัดข้อมูลเชิงลึกจากชุดข้อมูลที่เลือกมา ฝึกทักษะการสร้าง visualization ใน Tableau และสื่อสารผลผ่านวิดีโอและรายงานใน GitHub

1. Project Canvas (10%)

➔ กำหนด ปัญหา (Problem Statement), วัตถุประสงค์ทางธุรกิจ (Business Objectives), ผู้มีส่วนได้ส่วนเสีย (Key Stakeholders) และ ตัวชี้วัดความสำเร็จ (Success Metrics)

◆ ปัญหา (Problem Statement)

- ธุรกิจค้าปลีกออนไลน์มีข้อมูลการสั่งซื้อจำนวนมากจากลูกค้าหลากหลายกลุ่ม แต่ยังขาดการวิเคราะห์เชิงลึกเพื่อเข้าใจพฤติกรรม การซื้อ ความพึงพอใจของลูกค้า และปัจจัยที่มีผลต่อรายได้และกำไร ทำให้ไม่สามารถใช้ข้อมูลตัดสินใจเชิงกลยุทธ์อย่างมีประสิทธิภาพ

◆ วัตถุประสงค์ทางธุรกิจ (Business Objectives)

- วิเคราะห์แนวโน้มการซื้อสินค้าตามเพศ อายุ และกลุ่มผลิตภัณฑ์
- วิเคราะห์ผลกระทบของส่วนลดและโปรโมชั่น
- ระบุประเภทลูกค้าที่มีแนวโน้มซื้อซ้ำหรือมียอดใช้จ่ายสูง เพื่อใช้ในแผนส่งเสริมการขายแบบเฉพาะกลุ่ม
- สำรวจความสัมพันธ์ระหว่างวิธีการจัดส่งและระดับความพึงพอใจของลูกค้า เพื่อปรับปรุงประสบการณ์ของลูกค้า
- สร้าง Dashboard เพื่อสรุปภาพรวมธุรกิจให้เห็นผลประกอบการได้ง่ายขึ้น

◆ ผู้มีส่วนได้ส่วนเสีย (Key Stakeholders)

- ผู้บริหารธุรกิจ (Executives)
- ฝ่ายการตลาด (Marketing Team)
- นักวิเคราะห์ข้อมูล (Data Analysts)
- ฝ่ายจัดส่งสินค้า (Logistics Team)

◆ ตัวชี้วัดความสำเร็จ (Success Metrics)

- ยอดขายต่อคำสั่งซื้อ
- ลูกค้ามีการกลับมาซื้อซ้ำมากขึ้น
- คะแนนความพึงพอใจเพิ่มขึ้น
- มีลูกค้าที่เข้าร่วมใช้โปรโมชั่นมากขึ้น

➔ Project Canvas : กลุ่มเป้าหมาย ทรัพยากร (ข้อมูล/เครื่องมือ) สมมติฐาน ข้อจำกัด และไทม์ไลน์

◆ กลุ่มเป้าหมาย (Target Audience)

- ผู้บริหารธุรกิจ (Executives)

- ฝ่ายการตลาด (Marketing Team)
 - นักวิเคราะห์ข้อมูล (Data Analysts)
 - ฝ่ายจัดส่งสินค้า (Logistics Team)
- ◆ ทรัพยากร (Resources)
- ชุดข้อมูล: Customer Shopping Trends Dataset
- ◆ สมมติฐานเบื้องต้น (Initial Hypotheses)
- ลูกค้ากลุ่มวัยรุ่น (อายุ 18 - 25 ปี) มีแนวโน้มจะสั่งซื้อสินค้ามากที่สุด
 - ลูกค้าที่ใช้รหัสโปรโมชั่นมีแนวโน้มสั่งซื้อของบ่อย
 - ลูกค้าผู้หญิงมีแนวโน้มซื้อสินค้า มากกว่าผู้ชาย
 - ในฤดูร้อน (Summer) มีคำสั่งซื้อ มากกว่าฤดูหนาว (Winter)
 - ลูกค้าส่วนใหญ่ชำระเงินผ่านบัตรเครดิต (Credit Card)
- ◆ ข้อจำกัด (Constraints)
- ไม่มีข้อมูลวันที่และเวลาสั่งซื้อ → ไม่สามารถวิเคราะห์ตามช่วงเวลาได้ว่าช่วงไหนที่มีการสั่งซื้อสินค้าเยอะกว่ากัน
 - ไม่มีข้อมูลคำสั่งซื้อของแบบหลายชิ้น → ไม่สามารถแยกสินค้าหลายประเภทในคำสั่งซื้อเดียว
 - ไม่มีรหัสลูกค้า → ไม่สามารถติดตามพฤติกรรมของลูกค้าแต่ละคนในระยะยาวหรือวิเคราะห์ Customer Loyalty ได้
- ◆ ไทม์ไลน์ (Timeline)
- สัปดาห์ที่ 1 กำหนดปัญหา, สร้าง Project Canvas และทำการเตรียมข้อมูล (Data Preparation)
 - สัปดาห์ที่ 2 ทำ EDA ด้วย Tableau หา insight และวิเคราะห์เชิงลึก (In-Depth Analysis)
 - สัปดาห์ที่ 3 ปรับปรุงเนื้อหางานให้มีความถูกต้องเรียบร้อย และสร้าง Visualization ใน Tableau
 - สัปดาห์ที่ 4 ทำวิดีโอนำเสนอ และเขียนรายงาน GitHub

➔ ส่งมอบ Canvas แบบหน้ากระดาษเดียว (PowerPoint หรือ PDF)

2. การเตรียมข้อมูล (Data Preparation) (15%)

2.1 การทำความสะอาดข้อมูล (Data Cleaning) ในขั้นตอนนี้ ได้มีการตรวจสอบและจัดการกับปัญหาที่อาจส่งผลกระทบต่อการวิเคราะห์ข้อมูล ได้แก่:

- ค่าที่ขาดหาย (Missing Values): จากการตรวจสอบข้อมูลไม่พบค่าที่ขาดหายใน dataset

```
1 #ตรวจสอบข้อมูลที่ขาดหายว่ามีไหม
2 print("Missing Values - shopping_trends.csv")
3 print(df1.isnull().sum())
4
5 print("\n Missing Values - shopping_trends_updated.csv")
6 print(df2.isnull().sum())
7
```

- แปลงพีเจอร์แบบ Yes/No ให้เป็นตัวเลข เพราะให้เหมาะสมแก่การใช้งานมากขึ้น

```

1 #แปลงพีเจอรแบบ Yes/No ให้เป็นตัวเลข
2 df1['Discount Applied'] = df1['Discount Applied'].map({'Yes': 1, 'No': 0})
3 df1['Promo Code Used'] = df1['Promo Code Used'].map({'Yes': 1, 'No': 0})

```

- ข้อมูลซ้ำ (Duplicate Data): มีการตรวจสอบข้อมูลซ้ำ และลบแถวที่ซ้ำซ้อนออก

```

1 #ตรวจสอบข้อมูลซ้ำ
2 duplicates_df1 = df1.duplicated()
3 duplicates_df2 = df2.duplicated()

```

2.2 การสร้างคุณลักษณะใหม่ (Feature Engineering) มีการสร้างพีเจอรใหม่หลายรายการเพื่อเพิ่มความสามารถในการวิเคราะห์ข้อมูล โดยมีรายละเอียดดังนี้:

- Cost: คำนวณต้นทุนโดยสมมติให้ต้นทุนเป็น 70% ของยอดซื้อ ($\text{Cost} = \text{Purchase Amount} \times 0.7$)
- Profit: กำไร = ยอดซื้อ - ต้นทุน ($\text{Profit} = \text{Purchase Amount} - \text{Cost}$)
- Profit Margin: อัตรากำไร = กำไร / ยอดซื้อ ($\text{Profit Margin} = \text{Profit} / \text{Purchase Amount}$)
- Age Group: จำแนกกลุ่มอายุตามช่วง เช่น Teen, Young Adult, Adult, Mid Age, และ Senior

```

1 #สมมติ cost = 70%ของราคาขาย
2 df1['Cost'] = df1['Purchase Amount (USD)'] * 0.7
3
4 #คำนวณกำไร(Profit)
5 df1['Profit'] = df1['Purchase Amount (USD)'] - df1['Cost']
6
7 #คำนวณอัตรากำไร(Profit Margin)
8 df1['Profit Margin'] = df1['Profit'] / df1['Purchase Amount (USD)']
9

```

```

1 #สร้างกลุ่มอายุ(Age Group)
2 df1['Age Group'] = pd.cut(df1['Age'],
3                             bins=[0, 18, 30, 45, 60, 100],
4                             labels=['Teen', 'Young Adult', 'Adult', 'Mid Age', 'Senior'])

```

- High Spender: ผู้ที่มียอดใช้จ่ายมากกว่า 100 ดอลลาร์สหรัฐ
- Loyal Customer: ลูกค้าที่เคยซื้อซ้ำตั้งแต่ 5 ครั้งขึ้นไป
- Discount Applied & Promo Code Used: แปลงค่าประเภท Yes/No ให้เป็นค่าตัวเลข (1 = Yes, 0 = No)
- Preferred Payment Match: ตรวจสอบว่าวิธีการชำระเงินตรงกับวิธีที่ลูกค้าต้องการหรือไม่ (1 = ตรง, 0 = ไม่ตรง)

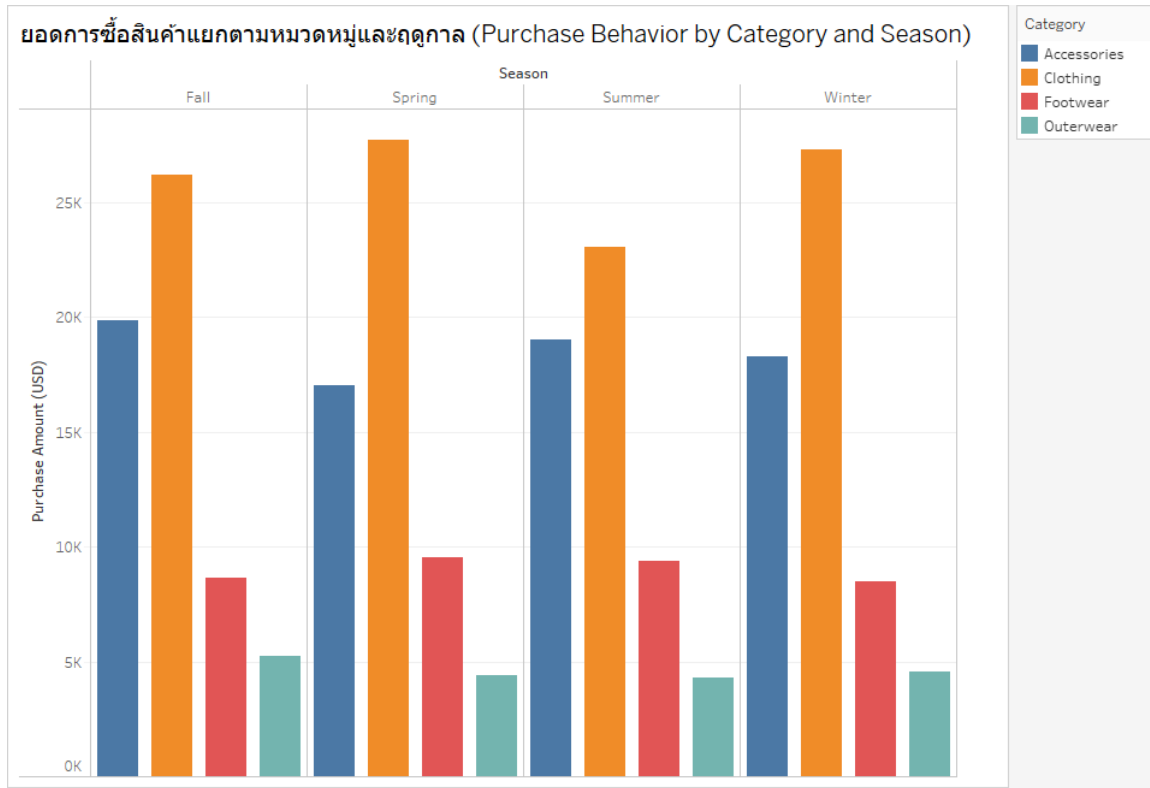
2.3 เอกสารประกอบ (Documentation) ขั้นตอนการเตรียมข้อมูลทั้งหมดถูกจัดทำใน Jupyter Notebook อย่างเป็นระบบ โดยประกอบด้วย:

- โค้ด Python พร้อมคำอธิบายเป็น Markdown
- ขั้นตอนการทำความสะอาดข้อมูล และสร้างคุณลักษณะใหม่

- ใช้ library เช่น pandas, numpy, และ matplotlib เพื่อการจัดการข้อมูลและแสดงผล

สามารถอ้างอิงได้จากไฟล์แนบ Shopping_Trends_Data_Preparation.ipynb

3. การสำรวจข้อมูลเบื้องต้น (Exploratory Data Analysis, EDA) (20%)



3.1 กราฟ: ยอดการซื้อสินค้าแยกตามหมวดหมู่และฤดูกาล

กราฟนี้แสดงยอดการซื้อสินค้ารวม (Purchase Amount) แยกตามหมวดหมู่สินค้า ได้แก่ Accessories, Clothing, Footwear และ Outerwear ภายใต้นี้แต่ละฤดูกาล คือ Fall, Spring, Summer และ Winter

ในส่วนของ Key Patterns

- Clothing เป็นหมวดหมู่ที่ทำยอดขายสูงสุดในทุกฤดูกาลโดยเฉพาะในฤดูใบไม้ผลิ (Spring) และฤดูหนาว (Winter) ที่ยอดขายเกิน 27,000 USD แสดงให้เห็นว่าผู้บริโภคให้ความสำคัญกับการซื้อเสื้อผ้าในช่วงเปลี่ยนฤดูและฤดูที่อากาศเย็น
- Accessories ทำยอดขายได้อย่างสม่ำเสมอโดยเฉพาะในฤดูใบไม้ร่วง (Fall) และฤดูร้อน (Summer) สะท้อนให้เห็นว่าเครื่องประดับเป็นสินค้าที่ได้รับความนิยมตลอดทั้งปี
- Footwear และ Outerwear มียอดขายต่ำที่สุดในทุกฤดูกาล Outerwear มีแนวโน้มยอดขายต่ำมากในฤดูร้อน (ต่ำกว่า 5,000 USD) สอดคล้องกับสภาพอากาศที่ไม่จำเป็นต้องใช้เสื้อกันหนาว

ในส่วนของ Irregularities

- i. Outerwear ในฤดูใบไม้ร่วง (Fall) มียอดขายต่ำผิดปกติ แม้อากาศจะเริ่มเย็น
- ii. Footwear ในฤดูหนาว (Winter) มียอดขายต่ำกว่าที่คาดไว้ แม้จะเป็นฤดูที่รองเท้าอุ่นน่าจะขายดี
- iii. ช่องว่างระหว่างหมวด Clothing กับหมวดอื่นๆ ค่อนข้างมาก ในบางฤดู เช่น Spring ที่ Clothing มียอดขายเกือบ 28,000 USD แต่ Outerwear ไม่ถึง 5,000 USD

ความสัมพันธ์ระหว่างส่วนลดกับยอดซื้อ (Discount vs. Purchase Amount)



3.2 กราฟ: ความสัมพันธ์ระหว่างส่วนลดกับยอดซื้อ

กราฟนี้เป็น Scatter Plot แสดงความสัมพันธ์ระหว่าง ยอดซื้อ (Purchase Amount) บนแกน X และ ส่วนลดที่ใช้ (Discount Applied) บนแกน Y โดยแบ่งตามกลุ่มอายุ (Age Group) และแสดงขนาดเสื้อผ้า (Size) ที่จุดข้อมูลแต่ละจุด

ในส่วนของ Key Patterns

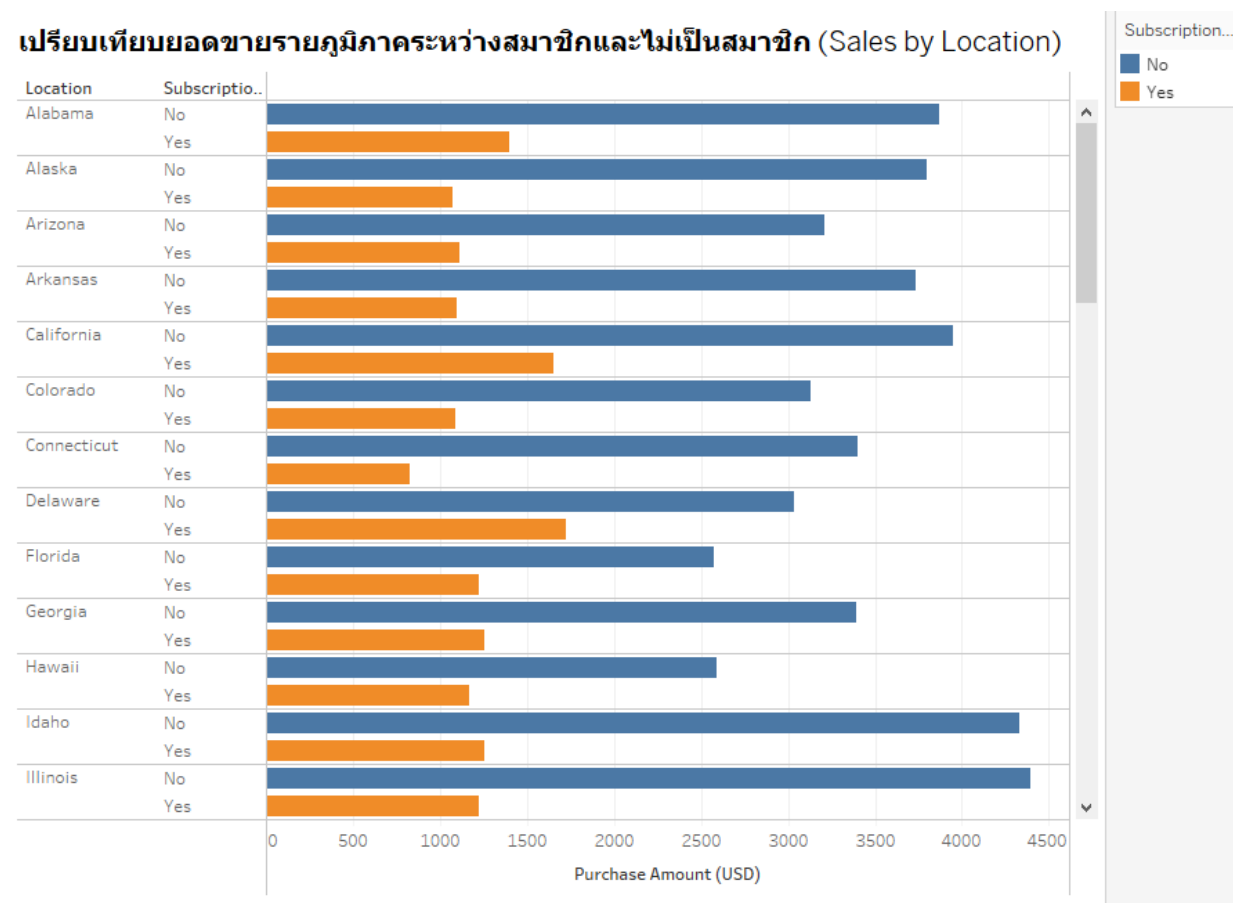
- i. มีความสัมพันธ์เชิงบวกระหว่างยอดซื้อและส่วนลดโดยทั่วไปแล้ว เมื่อยอดซื้อเพิ่มขึ้น ส่วนลดที่ได้รับก็มักจะสูงขึ้นตาม ซึ่งเป็นแนวโน้มที่สมเหตุสมผลในเชิงกลยุทธ์การตลาด

- ii. กลุ่มอายุ Young Adult และ Teen มียอดซื้อสูงและได้รับส่วนลดสูงจุดข้อมูลที่มีส่วนลดมากกว่า 150 และยอดซื้อสูงกว่า 20,000 USD ส่วนใหญ่มาจากกลุ่ม Young Adult (สีเขียว) และ Teen (สีแดง) ซึ่งอาจสะท้อนถึงการมีกำลังซื้อสูง หรือได้รับโปรโมชั่นเฉพาะกลุ่ม
- iii. ขนาดเสื้อผ้าที่ขายดีในยอดซื้อสูงมักอยู่ที่ไซส์ M และ L จุดที่มียอดซื้อสูงสุดและส่วนลดสูงแสดงขนาดเสื้อผ้า M และ L เป็นหลัก
- iv. กลุ่ม Adult และ Mid Age มีการกระจายอยู่ในช่วงยอดซื้อต่ำถึงกลางซึ่งอาจบ่งบอกว่ากลุ่มวัยทำงานมีแนวโน้มซื้อสินค้าบ่อยครั้งแต่ยอดซื้อแต่ละครั้งไม่สูงมากนัก

ในส่วนของ Irregularities

- i. มีบางจุดที่ได้ส่วนลดสูงแต่ยอดซื้อต่ำ (เช่น จุดกลุ่ม Teen ที่ยอดซื้อราว 8,000 แต่ได้ส่วนลดเกิน 80) ซึ่งอาจเป็นผลจากโปรโมชั่นเฉพาะกลุ่ม หรือความผิดปกติในการกำหนดส่วนลด
- ii. มีบางจุดที่ได้ส่วนลดน้อยมาก แม้ยอดซื้อค่อนข้างสูง อาจเป็นไปได้ว่าลูกค้ากลุ่มนั้นไม่เข้าร่วมแคมเปญส่วนลด หรือซื้อสินค้าที่ไม่มีโปรโมชั่น

เปรียบเทียบยอดขายรายภูมิภาคระหว่างสมาชิกและไม่เป็นสมาชิก (Sales by Location)



3.3 กราฟ: เปรียบเทียบยอดขายรายภูมิภาคระหว่างสมาชิกและไม่เป็นสมาชิก

กราฟนี้แสดงยอดขาย (Purchase Amount) ในแต่ละรัฐของสหรัฐอเมริกา โดยแยกตามสถานะการสมัครสมาชิก (Subscription Status)

- สีส้ม: ลูกค้าที่ สมัครสมาชิก (Yes)

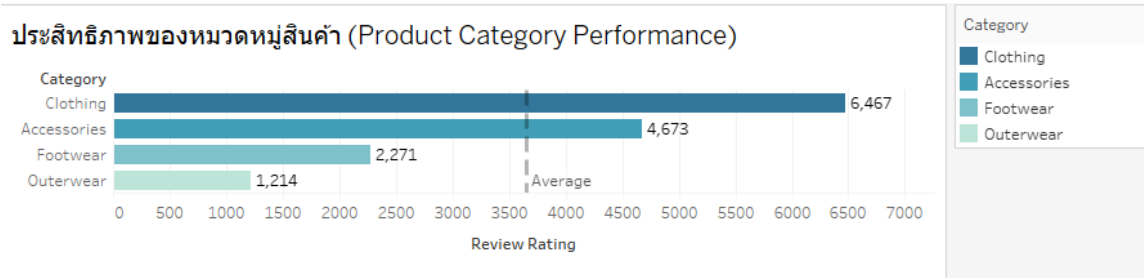
- สิ้นน้ำเงิน: ลูกค้ำที่ ไม่ได้สมัครสมาชิก (No)

ในส่วนของ Key Patterns

- i. ก้ำไรส่วนใหญ่มาจากลูกค้ำที่ ไม่ได้สมัครสมาชิก ในเกือบทุกรัฐ/พื้นที่ ก้ำไรส่วนใหญ่ (แห่งสีแดง) มาจากลูกค้ำที่ "ไม่ได้สมัครสมาชิก" (Subscription Status: No) ซึ่งบ่งชี้ว่าฐานลูกค้ำหลักที่สร้างก้ำไรยังคงเป็นลูกค้ำทั่วไป

ในส่วนของ Irregularities

- i. ก้ำไรจากลูกค้ำที่สมัครสมาชิกค่อนข้างน้อยในภาพรวม แม้ว่าจะมีลูกค้ำที่สมัครสมาชิกในหลายรัฐ แต่สัดส่วนก้ำไรที่มาจากกลุ่มนี้ยังน้อยมาก เมื่อเทียบกับลูกค้ำที่ ไม่ได้สมัครสมาชิก นี้อาจเป็นความผิดปกติหากธุรกิจมีเป้าหมายที่จะเพิ่มรายได้และก้ำไรจากฐานลูกค้ำสมาชิก ซึ่งหมายความว่าอาจมีโอก้ำในการปรับปรุงกลยุทธ์การสมัครสมาชิกหรือการมอบสิทธิประโยชน์เพื่อดึงดูดการซื้อจากกลุ่มลูกค้ำเหล่านี้ให้มากขึ้น



3.4 กราฟ: ประสิทธิภาพของหมวดหมู่สินค้า

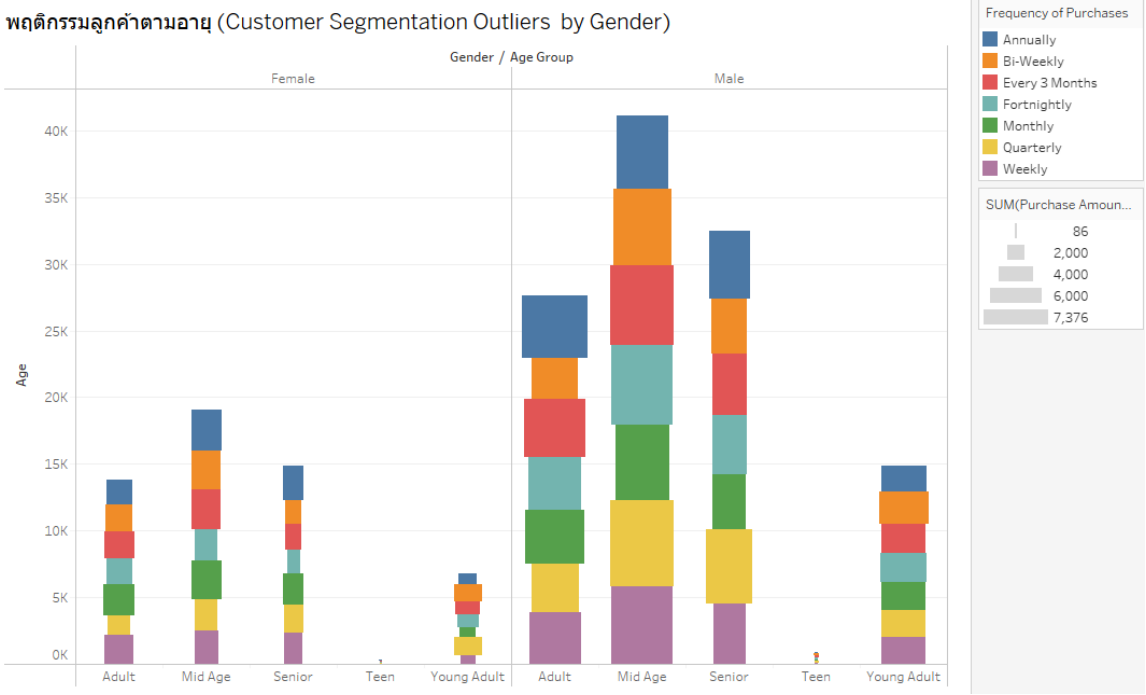
กราฟนี้เป็นแผนภูมิแท่งแนวนอน (Horizontal Bar Chart) แสดงผลรวมคะแนนรีวิว (Review Rating) แยกตามหมวดหมู่สินค้า ได้แก่ Clothing, Accessories, Footwear, และ Outerwear โดยความยาวของแท่งแสดงถึงยอดรวมของคะแนนรีวิวแต่ละกลุ่ม พร้อมเส้นค่าเฉลี่ย (Average) เพื่อใช้เปรียบเทียบภาพรวม

ในส่วนของ Key Patterns

- i. Clothing มีคะแนนรีวิวรวมสูงสุด (6,467 คะแนน) สะท้อนความนิยมของสินค้ากลุ่มนี้
- ii. Accessories และ Footwear อยู่ในระดับกลาง โดย Accessories สูงกว่าค่าเฉลี่ย
- iii. Outerwear มีคะแนนร่วมน้อยที่สุด (1,214 คะแนน) ต่ำกว่าค่าเฉลี่ยมาก อาจเกิดจากการเป็นสินค้าตามฤดูกาล หรือยอดขายต่ำ
- iv. ความแตกต่างระหว่างหมวดหมู่มีนัยสำคัญ โดย Clothing ได้คะแนนมากกว่า Outerwear ถึง 5 เท่า

ในส่วนขอ Irregularities

- i. สินค้าบางกลุ่ม เช่น Outerwear มีคะแนนรีวิวต่ำกว่าค่าเฉลี่ยอย่างเห็นได้ชัด อาจต้องพิจารณาปรับกลยุทธ์การตลาดหรือโปรโมชั่น
- ii. ยังไม่มีการแบ่งช่วงเวลา (Time Series) หรือ Sub-category ซึ่งอาจช่วยในการวิเคราะห์เชิงลึก



3.5 กราฟ: พฤติกรรมลูกค้าตามอายุและเพศ

กราฟนี้เป็น Stacked Bar Chart แสดงพฤติกรรมการซื้อของลูกค้า แยกตามกลุ่มอายุ (Age Group) และเพศ (Gender) โดยใช้ ยอดซื้อรวม (Total Purchase Amount) เป็นแกน Y และจำแนกความถี่ในการซื้อ (Frequency of Purchases) ด้วยสีในแต่ละแท่ง โดยมีการแยกฝั่ง ชาย (Male) และ หญิง (Female)

ในส่วนของ Key Patterns

- i. เพศชายในกลุ่ม Mid Age (วัยกลางคน) มียอดซื้อรวมสูงที่สุดจากทุกกลุ่ม แสดงถึงการเป็นกลุ่มลูกค้าที่มีกำลังซื้อสูงและมีความถี่ในการซื้อหลากหลาย
- ii. กลุ่มชายสูงวัย (Senior Male) และ ชายวัยทำงาน (Adult Male) ก็มียอดซื้อสูงเช่นกัน โดยเฉพาะการซื้อแบบรายสัปดาห์ (Weekly) และ รายเดือน (Monthly) สะท้อนพฤติกรรมการซื้ออย่างสม่ำเสมอ
- iii. เพศหญิงในทุกช่วงอายุ มียอดซื้อน้อยกว่าฝั่งชายในทุกกลุ่มอายุเดียวกัน โดยเฉพาะในกลุ่ม Adult และ Senior
- iv. กลุ่มวัยรุ่น (Teen) ทั้งชายและหญิง มียอดซื้อน้อยที่สุด ซึ่งอาจเกิดจากการกำลังซื้อจำกัด หรือไม่มีอำนาจในการตัดสินใจซื้อสินค้าเอง

- v. กลุ่ม Young Adult (20–30 ปี) ทั้งสองเพศ มีการซื้อค่อนข้างสม่ำเสมอ แม้ยอดรวมจะไม่สูงเท่ากลุ่มอื่น แต่มีความถี่แบบรายสัปดาห์/รายเดือนเด่นชัด

ในส่วนของ Irregularities

- i. กลุ่มผู้หญิงวัยกลางคน (Mid Age Female) มีความถี่ในการซื้อหลายประเภท แต่ยอดรวมยังน้อยกว่าชายกลุ่มเดียวกันมาก อาจมีข้อจำกัดด้านงบประมาณหรือสินค้าไม่ตรงกลุ่มเป้าหมาย
 - ii. Teen Female และ Teen Male มีการซื้อแบบ “รายเดือน/ราย 3 เดือน” แทนที่จะเป็นรายสัปดาห์ ซึ่งอาจเกิดจากการรับสินค้าในช่วงโปรโมชั่นเท่านั้น
- ฝังภาพหน้าจอและอธิบายในรายงาน GitHub

4. การวิเคราะห์เชิงลึก (In-Depth Analysis) (20%)

4.1 ปัจจัยใดส่งผลต่อความแปรปรวนของกำไร (Profit) มากที่สุด

```

1 import pandas as pd
2 from sklearn.linear_model import LinearRegression
3 from sklearn.model_selection import train_test_split
4 from sklearn.metrics import r2_score
5 from sklearn.preprocessing import LabelEncoder
6
7 # โหลดข้อมูล
8 df = pd.read_csv("shopping_trends_updated.csv")
9
10 # Feature Engineering
11 df['Cost'] = df['Purchase Amount (USD)'] * 0.7
12 df['Profit'] = df['Purchase Amount (USD)'] - df['Cost']
13 df['Profit Margin'] = df['Profit'] / df['Purchase Amount (USD)']
14
15 # Label Encoding สำหรับคอสัมพันธ์ที่เป็นข้อความ
16 le = LabelEncoder()
17
18 # แปลงค่าที่เป็นข้อความให้เป็นตัวเลข
19 df['Frequency of Purchases'] = le.fit_transform(df['Frequency of Purchases'])
20 df['Promo Code Used'] = le.fit_transform(df['Promo Code Used'])
21 df['Discount Applied'] = le.fit_transform(df['Discount Applied'])
22
23 # สร้างตัวแปรสำหรับโมเดล
24 X = df[['Age', 'Frequency of Purchases', 'Discount Applied', 'Promo Code Used']]
25 y = df['Profit']
26
27 # แบ่งข้อมูล
28 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
29
30 # เทรนโมเดล
31 model = LinearRegression()
32 model.fit(X_train, y_train)
33
34 # ทำนายและประเมินผล
35 y_pred = model.predict(X_test)
36
37 # แสดงค่า R-squared
38 print(f"R-squared: {r2_score(y_test, y_pred):.4f}\n")
39
40 # แสดง Coefficients แยกบรรทัด
41 print("Coefficients:")
42 for name, coef in zip(X.columns, model.coef_):
43     print(f" {name}: {coef:.4f}")

```

ผลลัพธ์การวิเคราะห์เชิงลึก ข้อ 4.1

- ค่า R^2 : -0.0088
แสดงให้เห็นว่า ตัวแปรเหล่านี้ ไม่ใช่ตัวทำนายที่ดี สำหรับกำไร ในชุดข้อมูลนี้

ค่าสัมประสิทธิ์ (Coefficients):

- Age : -0.0060
→ อายุสูงขึ้น 1 ปี กำไรลดลงเฉลี่ยประมาณ 0.006 USD (น้อยมากจนไม่นับสำคัญ)
- Frequency of Purchases : -0.1068
→ ยิ่งซื้อถี่ กำไรเฉลี่ยต่อคำสั่งซื้อลดลง
อาจเป็นเพราะลูกค้าที่ซื้อบ่อยซื้อของราคาต่ำหรือใช้ส่วนลดบ่อย

- Discount Applied : -0.0514
→ การใช้ส่วนลดมีผลลดกำไรลงเล็กน้อย
- Promo Code Used : -0.0514
→ เช่นเดียวกัน การใช้โค้ดโปรโมชั่นก็ลดกำไรเล็กน้อย

4.2 กลุ่มลูกค้าแบบใดมีแนวโน้มซื้อซ้ำ (Loyal Customers) มากที่สุด

```
1 from sklearn.linear_model import LogisticRegression
2 from sklearn.preprocessing import LabelEncoder
3
4 # เตรียมข้อมูล
5 df['Loyal Customer'] = df['Previous Purchases'] >= 5
6 df['Loyal Customer'] = df['Loyal Customer'].astype(int)
7
8 # Encode ตัวแปรกลุ่มค่า
9 df['Subscription Status'] = LabelEncoder().fit_transform(df['Subscription Status'])
10 df['Gender'] = LabelEncoder().fit_transform(df['Gender'])
11
12 X = df[['Age', 'Subscription Status', 'Gender', 'Discount Applied']]
13 y = df['Loyal Customer']
14
15 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
16
17 model = LogisticRegression()
18 model.fit(X_train, y_train)
19
20 print(f"Accuracy: {model.score(X_test, y_test):.4f}")
21 print("Coefficients:")
22 for feature, coef in zip(X.columns, model.coef_[0]):
23     print(f" - {feature:}: {coef:.4f}")
```

ผลลัพธ์การวิเคราะห์เชิงลึก ข้อ4.2

- ค่า Accuracy : 0.9231
โมเดลทำนายได้แม่นยำกว่าใครเป็นลูกค้าประจำ

ค่าสัมประสิทธิ์ (Coefficients):

- Age : 0.0019
→ อายุเพิ่มขึ้น 1 ปี เพิ่มโอกาสเล็กน้อย
- Subscription Status : 0.4126
→ ลูกค้าที่สมัครสมาชิก มีโอกาสเป็นลูกค้าประจำสูงขึ้น
- Gender : 0.4789
→ เพศหญิง มีแนวโน้มรักดีมากกว่ากลุ่มอื่น แปลว่าเพศหญิงซื้อซ้ำมากกว่า
- Discount Applied : 0.5753
→ การใช้ส่วนลดลดโอกาส ที่จะเป็ลูกค้าประจำ

5. ข้อค้นพบและข้อเสนอแนะ (Insights & Recommendations) (15%)

1. Mid Age Male เป็นกลุ่มลูกค้าที่มีกำลังซื้อสูง แต่ กำลังซื้อต่อคำสั่งซื้อลดลงตามความถี่ในการซื้อ จาก EDA พบว่า ลูกค้าชายช่วงวัยกลางคน (Mid Age Male) มี ยอดซื้อรวมสูงสุด แต่จาก In-Depth Analysis พบว่า Frequency of Purchases มีความสัมพันธ์เชิงลบกับกำไร (Coefficient: -0.1068) แสดงให้เห็นว่าลูกค้าที่ซื้อ บ่อยอาจซื้อสินค้ามูลค่าต่ำหรือใช้ส่วนลดบ่อย ทำให้กำไรเฉลี่ยต่อคำสั่งซื้อลดลง

Recommendation: พัฒนา campaign ที่กระตุ้นให้ลูกค้าซื้อเป็น high-value order แทนการซื้อบ่อย เช่น ซื้อครบ \$100 รับ 10% cashback” หรือ “รับของขวัญ ฟรีเมื่อซื้อเกิน \$80 แนวทางนี้ก็จะช่วยเพิ่ม AOV (Average Order Value) และ ยังคงระดับกำไรให้สูงได้

2. Non-member ยังเป็นแหล่งกำไรหลัก แต่สมาชิกมีแนวโน้มที่จะเป็น Loyalty สูงกว่า จาก EDA พบว่า ลูกค้าที่ไม่เป็นสมาชิก (Non-member) สร้างกำไรหลักใน หลายรัฐแต่ In-Depth Analysis แสดงว่า Subscription Status มี coefficient = +0.4126 นั้นแสดงว่าสมาชิกมีแนวโน้มเป็นลูกค้าประจำมากกว่า

Recommendation: ปรับปรุงระบบสมาชิกเป็น Tiered Membership เช่น Basic / Silver / Gold โดยเพิ่ม incentive ให้สมาชิกใช้จ่ายมากขึ้น เช่น ส่งฟรี, สะสม แด้ม, early access ลดการพึ่งพา discount ตรง ๆ แล้วเปลี่ยนมาใช้ value-added benefits

3. การใช้ Discount และ Promo Code มีผลลบต่อกำไรและ Customer Loyalty

ในส่วน In-Depth Analysis พบว่า:

Discount Applied = -0.0514 และ Promo Code Used = -0.0514 กำไรลดลง

Discount Applied = -0.5753 ลูกค้ามีแนวโน้มไม่กลับมาซื้อซ้ำ

ลูกค้าที่ใช้โปรโมชั่นซ้ำอาจไม่ได้เป็น Loyalty จริง และจะซื้อเฉพาะตอนมีดีลเท่านั้น

Recommendation: ใช้โปรโมชั่นแบบ targeted เฉพาะลูกค้าใหม่ หรือใช้เพื่อดึงลูกค้าที่ inactive หลีกเลี่ยงการให้ discount ซ้ำ ๆ กับลูกค้าที่มีแนวโน้มซื้ออยู่แล้ว เปลี่ยนมาใช้ reward system หรือ point-based loyalty program แทนเพื่อรักษา margin

ลิงค์ส่งงาน

1. ไฟล์งานทั้งหมด Canvas, Tableau <https://drive.google.com/drive/folders/1AQEHwT6r3nwPkfwQNj8IGJUY0BRB66Ca?usp=sharing>
2. แบนลิงก์ Tableau Public Dashboard : <https://public.tableau.com/app/profile/pattanapong.suriyawongpree/viz/cp372-NkPsDp/CustomShoppingTrendsOverviewDashboard>
3. แบนลิงก์วิดีโอ YouTube : <https://www.youtube.com/watch?v=jYdEOQhN9Hc>
4. แบน URL GitHub Repository : <https://github.com/TheCopperCoin03/cp372-NkPsDp>