การศึกษาและหาปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย

Investigating the Factors Influencing Behavior and Interest in Selecting Content Related to Children's Products on Social Media

นาย สุวดล	โกกุล	6304800001
นางสาว พลอยชมพู	ไกรเจริญ	6304800012

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์

มหาวิทยาลัยสยาม

ปีการศึกษา 2566

หัวข้อปริญญานิพนธ์	การศึกษาและหาปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย Investigating the Factors Influencing Behavior and Interest in Selecting Content Related to Children's Products on Social Media		
หน่วยกิตของปริญญานิพนธ์	3 หน่วยกิต		
รายชื่อคณะผู้จัดทำ	นาย สุวดล	โกกุล	6304800001
	นางสาว พลอยชมพู	ไกรเจริญ	6304800012
อาจารย์ที่ปรึกษา	อาจารย์ ธนาภรณ์	รอดชีวิต	
ระดับการศึกษา	วิทยาศาสตรบัณฑิต		
ภาควิชา	วิทยาการคอมพิวเตอร์		
ปีการศึกษา	2566		

อนุมัติให้ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

คณะกรรมการสอบปริญญานิพนธ์

.....ประธานกรรมการ
(พล.อ.ท.ผศ.ดร. พาห์รณ สงวนโภคัย)

.....กรรมการ
(อาจารย์จรรยา แหยมเจริญ)

.....อาจารย์ที่ปรึกษา
(อาจารย์ธนาภรณ์ รอดชีวิต)

หัวข้อปริญญานิพนธ์	การศึกษาและหาปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย		
หน่วยกิตของปริญญานิพนธ์	3 หน่วยกิต		
รายชื่อคณะผู้จัดทำ	นาย สุวดล	โกกุล	6304800001
	นางสาว พลอยชมพู	ไกรเจริญ	6304800012
อาจารย์ที่ปรึกษา	อาจารย์ ธนาภรณ์	รอดชีวิต	
ระดับการศึกษา	วิทยาศาสตรบัณฑิต		
ภาควิชา	วิทยาการคอมพิวเตอร์		
ปีการศึกษา	2566		

บทคัดย่อ

วัตถุประสงค์ในการจัดทำปริญญานิพนธ์นี้เพื่อศึกษาและหาปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย โดยปัจจุบันการสร้างคอนเทนต์ที่น่าสนใจและโดดเด่นเป็นความท้าทาย โดยเฉพาะผลิตภัณฑ์สำหรับเด็กซึ่งความน่าเชื่อถือและคุณภาพเป็นสิ่งสำคัญ ทางคณะผู้จัดทำจึงได้ทำการศึกษาและหาปัจจัยจากหลายแพลตฟอร์ม เช่น Facebook, Twitter(X), Instagram, YouTube, TikTok, และ Forum โดยใช้เครื่องมือ Zocial Eye ในการรวบรวมข้อมูล และโปรแกรม Visual Studio Code สำหรับการทำความสะอาดและจัดกลุ่มข้อมูล รวมถึงโปรแกรม Tableau สำหรับการนำเสนอข้อมูล จากผลการศึกษาพบว่าปัจจัยที่มีผลต่อความสนใจ ได้แก่ 1) พรีเซนเตอร์หรือบุคคลที่มีชื่อเสียงที่มีลูกน้อยจะดึงดูดความสนใจได้มากขึ้น เพราะความน่ารักของเด็กๆ ทำให้เกิดความสนใจ 2) แฮชแท็ก (#) ช่วยดึงดูดให้คนอยากคลิกเข้าไปดูรายละเอียดว่าแฮชแท็กนี้เกี่ยวกับอะไร 3) การใช้ภาพหรือวีดีโอที่สะดุดตาจะดึงดูดสายตาผู้คนมากขึ้น ทำให้มีคนสนใจและคลิกเข้าไปดูมากขึ้น 4) ความคิดเห็นและปฏิสัมพันธ์ของคนอื่นๆ เช่น ยอดไลค์หรือโพสต์ที่มีคอมเมนต์เยอะๆ จะช่วยดึงดูดให้ผู้คนสงสัยและสนใจว่าทำไมถึงมีการ Interactive สูง 5) การจัดแคมเปญหรือกิจกรรมที่มีของแจกฟรีหรือของรางวัลจาก Influencer ยิ่งทำให้เกิดความน่าสนใจและการมีส่วนร่วมมากขึ้น ปัจจัยเหล่านี้ช่วยให้แบรนด์หรือผู้ประกอบการเข้าใจวิธีสร้างคอนเทนต์ที่น่าสนใจในยุคโซเชียลมีเดียและนำไปปรับใช้ในการวางแผนการตลาดและการสื่อสารกับกลุ่มเป้าหมายได้อย่างมีประสิทธิภาพ

คำสำคัญ: ผลิตภัณฑ์เด็ก, พฤติกรรม, คอนเทนต์

Project title	Investigating the Factors Influencing Behavior and Interest in Selecting Content Related to Children's Products on Social Media		
Project credits	3 Units		
By	Mr. Suwadol	Gogul	6304800001
	Ms. Ploychomphoo	Kraicharoen	6304800012
Advisor	Miss Thanaporn	Rodcheewit	
Degree	Bachelor of Science		
Major	Computer Science		
Faculty	Science		
Academic year	2023		

Abstract

The objective of this thesis is to study and identify factors that influence behavior and interest in selecting content related to children's products on social media. Creating engaging and standout content has become a challenge, especially for children's products where credibility and quality are paramount. Therefore, the research team conducted a study to determine these factors across various platforms such as Facebook, Twitter (X), Instagram, YouTube, TikTok, and forums. The data was collected using Zocial Eye, cleaned and categorized with Visual Studio Code, and presented using Tableau. The study's findings revealed several key factors influencing interest: 1) Presenters or celebrities with young children are particularly effective at attracting attention, as the appeal of children generates interest. 2) Hashtags (#) encourage users to click and explore the content associated with them. 3) Eye-catching images or videos draw more attention and increase the likelihood of clicks. 4) Comments and interactions, such as posts with many likes or comments, arouse curiosity and interest as people wonder why there is high interaction. 5) Campaigns or activities offering free gifts or prizes from influencers significantly increase interest and engagement. These factors help brands and businesses understand how to create engaging content in the social media era and apply these insights to marketing strategies and communication with target audiences effectively.

Keywords: children's products, behavior, content

Thanaporn Rodcheewit
(Project-Advisor)

Approved by

กิตติกรรมประกาศ

(Acknowledgment)

การจัดทำปริญญานิพนธ์ฉบับนี้สำเร็จได้นั้น คณะผู้จัดทำได้รับความกรุณาจาก อาจารย์ผู้สอนทุกท่านที่ให้ข้อมูลต่าง ๆ ส่งผลให้คณะผู้จัดทำได้รับความรู้และประสบการณ์ต่าง ๆ ที่มีค่ามากมายสำหรับปริญญานิพนธ์ฉบับนี้สำเร็จลงได้ด้วยดีจากความร่วมมือและสนับสนุนจากหลายฝ่ายดังนี้

1. อาจารย์ ธนาภรณ์ รอดชีวิต อาจารย์ที่ปรึกษา
2. คุณอนันต์ ตีระบูรณะพงษ์ ผู้อำนวยการแผนก Data & Innovation บริษัท ดาต้า เฟิร์ส จำกัด

คณะผู้จัดทำใคร่ขอขอบพระคุณคณะกรรมการสอบปริญญานิพนธ์ ที่ได้ให้คำแนะนำสำคัญเพื่อให้การสอบปริญญานิพนธ์ฉบับนี้ และผู้มีส่วนร่วมทุกท่าน รวมทั้งผู้ที่ไม่ได้กล่าวนาม ที่มีส่วนร่วมในการให้ข้อมูลให้ความช่วยเหลือ และเป็นที่ปรึกษาให้คำแนะนำต่าง ๆ จนทำให้งานทุกอย่างประสบความสำเร็จไปด้วยดี และทำรายงานฉบับนี้จนเสร็จสมบูรณ์ ซึ่งคณะผู้จัดทำขอขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้ด้วย

คณะผู้จัดทำ

นายสุวดล โกกุล

นางสาวพลอยชมพู ไกรเจริญ

สารบัญ

หน้า

บทคัดย่อ..... ก

Abstract..... ข

กิตติกรรมประกาศ..... ค

บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญของปัญหา..... 1

1.2 วัตถุประสงค์ของปริญญานิพนธ์..... 1

1.3 ขอบเขตของปริญญานิพนธ์..... 2

1.4 ประโยชน์ที่คาดว่าจะได้รับ..... 2

1.5 ขั้นตอนและวิธีดำเนินงานปริญญานิพนธ์..... 3

1.6 ขั้นตอนและระยะเวลาในการดำเนินปริญญานิพนธ์..... 4

1.7 อุปกรณ์และเครื่องมือที่ใช้ในการพัฒนาระบบ..... 5

1.8 อุปกรณ์และเครื่องมือที่รองรับ..... 5

บทที่ 2 การทบทวนเอกสารและวรรณกรรมที่เกี่ยวข้อง

2.1 Data Analytics..... 6

2.2 Clustering..... 7

2.3 Data Preparation..... 9

2.4 Data Visualization..... 9

2.5 Tableau..... 12

2.6 Visual studio code..... 13

2.7 Social Listening..... 14

2.8 ZOCIAL EYE..... 14

2.9 PCA(Principal component analysis) 17

2.10 Word Cloud..... 19

2.11 Isolation Forest..... 19

บทที่ 3 การวิเคราะห์และออกแบบระบบ

3.1 รายละเอียดของปริญญานิพนธ์..... 22

3.2 ขั้นตอนการวิเคราะห์ข้อมูล..... 23

3.2.1 กำหนดวัตถุประสงค์ในการวิเคราะห์ข้อมูล (Research Understanding) 23

3.2.2 รวบรวมข้อมูล (Data Understanding) 23

สารบัญ (ต่อ)

หน้า

3.2.3 เตรียมข้อมูล (Data Preparation) 23

3.2.4 จัดกลุ่มข้อมูล (Data Clustering) 34

บทที่ 4 การนำเสนอแผนภาพข้อมูล

4.1 การวิเคราะห์ภาพรวมสัดส่วน Engagement ตามแพลตฟอร์มต่างๆ..... 48

4.2 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อทั้งหมดทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์..... 49

4.3 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Product ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์..... 52

4.4 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Promotion ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์..... 55

4.5 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Activity ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์..... 57

บทที่ 5 สรุปผลและข้อเสนอแนะ

5.1 สรุปผลปริญญานิพนธ์..... 58

5.2 ข้อเสนอแนะ..... 58

บรรณานุกรม..... 60

สารบัญตาราง

หน้า

ตารางที่ 1.1 ขั้นตอนและระยะเวลาในการดำเนินงาน..... 4

สารบัญรูปภาพ

หน้า

รูปที่ 2.1 ตัวอย่างการแบ่งกลุ่มแบบ K-means..... 8
รูปที่ 2.2 ตัวอย่างลักษณะของการ clustering มีความคล้ายกับต้นไม้..... 8
รูปที่ 2.3 ตัวอย่างกระบวนการ data cleaning..... 9
รูปที่ 2.4 ตัวอย่าง Data Visualization แบบแผนภูมิ (Charts)..... 10
รูปที่ 2.5 ตัวอย่าง Data Visualization แบบกราฟ (Graphs)..... 10
รูปที่ 2.6 ตัวอย่าง Data Visualization แบบตาราง (Tables)..... 11
รูปที่ 2.7 ตัวอย่าง Data Visualization แบบแผนที่ (Maps)..... 11
รูปที่ 2.8 ตัวอย่าง Data Visualization แบบแดชบอร์ด (Dashboards)..... 12
รูปที่ 2.9 ตัวอย่างโปรแกรม Tableau..... 13
รูปที่ 2.10 ตัวอย่างหน้าจอโปรแกรม Visual Studio Code ที่ใช้ในการพัฒนา..... 14
รูปที่ 2.11ตัวอย่างฟีเจอร์ Campaign View ของ ZOCIAL EYE..... 15
รูปที่ 2.12 ตัวอย่างฟีเจอร์ Comparison View ของ ZOCIAL EYE..... 16
รูปที่ 2.13 ตัวอย่างฟีเจอร์ Trend Views ของ ZOCIAL EYE..... 17
รูปที่ 2.14 สมการการทำ Standardization..... 17
รูปที่ 2.15 Word Cloud..... 19
รูปที่ 2.16 ตัวอย่างการแบ่งข้อมูลปกติ [ซ้าย] และข้อมูลผิดปกติ [ขวา] 20
รูปที่ 2.17 ตัวอย่าง ต้นไม้ตัดสินใจ (Decision Tree) 20
รูปที่ 2.18 ตัวอย่างการใช้ฟังก์ชัน Isolation forest ในการหาOutlier..... 21
รูปที่ 3.1 ขั้นตอนการศึกษาและหาปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจ..... 23
รูปที่ 3.2 ข้อมูลที่นำออกมาจาก Zocial eye..... 24
รูปที่ 3.3 แสดงส่วนหนึ่งของการ Labeling แบรนด์ผลิตภัณฑ์เด็ก..... 24
รูปที่ 3.4 เป็นการสร้างคอลัมน์ใหม่ที่ชื่อว่า Brand เพื่อเก็บชื่อของแบรนด์สินค้า..... 26
รูปที่ 3.5 แสดงส่วนหนึ่งของการ Labeling คอนเทนต์ที่เกี่ยวกับ Activity..... 27
รูปที่ 3.6 แสดงส่วนหนึ่งของการ Labeling คอนเทนต์ที่เกี่ยวกับ Presenter..... 28
รูปที่ 3.7 แสดงส่วนหนึ่งของการ Labeling คอนเทนต์ที่เกี่ยวกับ Promotion..... 28

สารบัญรูปภาพ (ต่อ)

หน้า

รูปที่ 3.8 แสดงส่วนหนึ่งของการ Labeling คอนเทนต์ที่เกี่ยวกับ Product..... 29
รูปที่ 3.9 แสดงส่วนหนึ่งของการ Labeling คอนเทนต์ที่เกี่ยวกับ Other..... 30
รูปที่ 3.10 การสร้างคอลัมน์ใหม่ที่ชื่อว่า TOPIC..... 31
รูปที่ 3.22 การสร้างคอลัมน์ใหม่ที่ชื่อว่า Post type..... 32
รูปที่ 3.12 แสดงคอลัมน์ที่ Labeling เสร็จแล้ว..... 32
รูปที่ 3.13 แสดงการกรองข้อมูลและทำการ One-Hot Encoding..... 33
รูปที่ 3.14 แสดงการตรวจจับ Outlier ด้วย Isolation Forest..... 33
รูปที่ 3.15 แสดงการปรับขนาดของข้อมูลให้มีขนาดเดียวกัน..... 35
รูปที่ 3.16 แสดงฟังก์ชัน Scree plot เพื่อหาว่าควรใช้ Components เท่าไหร่ในการลดมิติ.... 36
รูปที่ 3.17 แสดงกราฟ Scree plot..... 38
รูปที่ 3.18 แสดงฟังก์ชันการทำงานของ Clustering..... 38
รูปที่ 3.19 แสดงกราฟ Scatter plot ที่ได้จากการทำ Clustering..... 39
รูปที่ 3.20 แสดงการเปลี่ยนชื่อและสีกลุ่มของ Cluster..... 40
รูปที่ 3.21 แสดงกราฟ Scatter plot ที่ได้จากการเปลี่ยนชื่อกลุ่มและสีของ Cluster..... 40
รูปที่ 3.22 แสดงการกรองข้อมูลและทำการ One-Hot Encoding Product..... 40
รูปที่ 3.23 แสดงกราฟ Scree plot..... 41
รูปที่ 3.24 แสดงกราฟ Scatter plot ของ Topic : Product..... 41
รูปที่ 3.25 แสดงการกรองข้อมูลและทำการ One-Hot Encoding..... 42
รูปที่ 3.26 แสดงกราฟ Scree plot..... 43
รูปที่ 3.27 แสดงกราฟ Scatter plot ของ Topic : Promotion..... 44
รูปที่ 3.28 แสดงการกรองข้อมูลและทำการ One-Hot Encoding..... 45
รูปที่ 3.29 แสดงกราฟ Scree plot..... 46
รูปที่ 3.30 แสดงกราฟ Scatter plot ของ Topic : Activity..... 46
รูปที่ 4.1 แสดงการวิเคราะห์ภาพรวมสัดส่วน Engagement ตามแพลตฟอร์มต่างๆ..... 48
รูปที่ 4.2 แสดงการวิเคราะห์การจัดกลุ่มตาม Engagement..... 49
รูปที่ 4.3 แสดงการเปรียบเทียบจำนวนโพสต์และจำนวน Engagement..... 50
รูปที่ 4.4 แสดงการเปรียบเทียบจำนวนโพสต์และจำนวน Engagement Twitter (X)..... 50
รูปที่ 4.5 ประเภทของบัญชีโซเชียลมีเดียที่คนไทยติตามมากที่สุด..... 51

สารบัญรูปภาพ (ต่อ)

หน้า

รูปที่ 4.6 แสดงการวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Promotion..... 52
รูปที่ 4.7 แสดงการเปรียบเทียบจำนวนโพสต์และจำนวน Engagement หัวข้อ Product..... 53
รูปที่ 4.8 แสดงการเปรียบเทียบจำนวนโพสต์และจำนวน Engagement Tiktok..... 53
รูปที่ 4.9 ประเภทของบัญชีโซเชียลมีเดียที่คนไทยติตามมากที่สุด..... 54
รูปที่ 4.10 แสดงการวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Promotion..... 55
รูปที่ 4.11 แสดงการเปรียบเทียบยอดผู้ติดตามเพจโปรโมชัน..... 56
รูปที่ 4.12 แสดงการวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Activity..... 57

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ในยุคที่โซเชียลมีเดียเป็นส่วนสำคัญในชีวิตประจำวันของผู้คนการวิเคราะห์พฤติกรรมและความสนใจของผู้บริโภคบนโซเชียลมีเดียจึงกลายเป็นเครื่องมือที่มีประสิทธิภาพสำหรับแบรนด์ในการทำความเข้าใจลูกค้าและพัฒนากลยุทธ์ทางการตลาดที่มีประสิทธิภาพ การเติบโตของโซเชียลมีเดียได้สร้างช่องทางใหม่ๆ ให้กับแบรนด์ในการสื่อสารและเข้าถึงผู้บริโภคได้ง่ายและรวดเร็วกว่าเมื่อก่อน สำหรับกลุ่มผลิตภัณฑ์เด็กเป็นกลุ่มหนึ่งที่ได้รับการกล่าวถึงอย่างมากในโซเชียลมีเดีย ผู้ปกครองมักใช้โซเชียลมีเดียในการหาข้อมูลเกี่ยวกับผลิตภัณฑ์ แบ่งปันประสบการณ์ และแลกเปลี่ยนความคิดเห็นเกี่ยวกับสินค้าต่างๆ เนื่องจากผู้ปกครองมักต้องการผลิตภัณฑ์ที่มีความน่าเชื่อถือสำหรับลูกของพวกเขา การวิเคราะห์พฤติกรรมและความสนใจของผู้บริโภคที่กล่าวถึงผลิตภัณฑ์สำหรับเด็กบนโซเชียลมีเดียสามารถให้ข้อมูลเชิงลึกที่มีคุณค่าเกี่ยวกับการตอบสนองของผู้บริโภคต่อผลิตภัณฑ์และแคมเปญต่างๆของแบรนด์ได้

การทำความเข้าใจพฤติกรรมและความสนใจของผู้บริโภคบนโซเชียลมีเดียเกี่ยวกับผลิตภัณฑ์สำหรับเด็กเป็นเรื่องที่สำคัญต่อการพัฒนากลยุทธ์ทางการตลาดที่มีประสิทธิภาพ สำหรับแบรนด์ที่จำหน่ายผลิตภัณฑ์สำหรับเด็ก การวิเคราะห์ดังกล่าวสามารถช่วยให้แบรนด์สามารถนำไปกำหนดกลยุทธ์ทางการตลาดในอนาคต

ปริญญานิพนธ์นี้มีวัตถุประสงค์เพื่อการศึกษาและหาปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย เพื่อให้ได้ข้อมูลเชิงลึกที่สามารถนำมาใช้ในการพัฒนาและปรับปรุงกลยุทธ์ทางการตลาดของแบรนด์ต่อไป อีกทั้งยังช่วยให้แบรนด์ต่างๆ มองเห็นโอกาสในการปรับปรุงผลิตภัณฑ์และบริการ และสามารถใช้ข้อมูลเชิงลึกเหล่านี้ในการพัฒนากลยุทธ์ทางการตลาดที่มีประสิทธิภาพ เพื่อสร้างความพึงพอใจให้กับลูกค้าและเพิ่มความสามารถในการแข่งขันในตลาด

1.2 วัตถุประสงค์ของปริญญานิพนธ์

เพื่อศึกษาและหาปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย

1.3 ขอบเขตปริญญานิพนธ์

1.3.1 รวบรวมและวิเคราะห์ข้อมูลจากโซเชียลมีเดียโดยใช้แพลตฟอร์ม Zocial Eye เป็นเครื่องมือที่ใช้สำหรับดึงข้อมูล Social listening บนโซเชียลมีเดีย

1.3.2 การวิเคราะห์ข้อมูลแบ่งออกเป็น 4 ส่วน ได้แก่

1.3.2.1 Data Understanding

รวบรวมและทำความเข้าใจกับข้อมูลที่มีอยู่ รวมถึงการระบุแหล่งข้อมูลที่เกี่ยวข้องและการตั้งคำถามที่ต้องการหาคำตอบผ่านการวิเคราะห์ข้อมูล โดยจะใช้ Zocial Eye ในการรวบรวมข้อมูลจากโซเชียลมีเดีย

1.3.2.2 Data Preparation

การจัดเตรียมข้อมูลให้พร้อมสำหรับการวิเคราะห์ ซึ่งรวมถึงการทำความสะอาดข้อมูล การจัดการกับข้อมูลที่ขาดหายหรือไม่ครบถ้วน และการปรับแต่งข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการวิเคราะห์

1.3.2.3 Data Analytics

ใช้เทคนิคและเครื่องมือในการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ และการใช้โมเดลทางสถิติหรือการเรียนรู้ของเครื่อง (Machine Learning) เพื่อจำแนกกลุ่มข้อมูล (Clustering)

1.3.2.4 Data Visualization

การนำผลลัพธ์จากการวิเคราะห์ข้อมูลมาแสดงในรูปแบบที่เข้าใจง่าย เช่น กราฟ แผนภาพ หรือแผนภูมิต่างๆ เพื่อที่จะสามารถมองเห็นข้อมูลในมุมมองที่ชัดเจนและเข้าใจได้ง่าย

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ข้อมูลที่ได้จากการวิเคราะห์สามารถช่วยให้ผู้ประกอบการและนักการตลาดเข้าใจปัจจัยที่มีผลต่อการรับชมคอนเทนต์ของผู้รับชม และสามารถสร้างคอนเทนต์ที่ตรงกับความต้องการและความสนใจของผู้รับชมได้มากขึ้น

1.4.2 การรู้จักปัจจัยที่มีผลต่อพฤติกรรมและความสนใจจะช่วยให้สามารถปรับปรุงวิธีการสื่อสารและการนำเสนอคอนเทนต์ให้มีประสิทธิภาพและน่าสนใจมากขึ้น

1.4.3 เมื่อคอนเทนต์สอดคล้องกับความสนใจของผู้รับชม จะช่วยเพิ่มการมีส่วนร่วม เช่น การกดไลค์ การแชร์ และการแสดงความคิดเห็น ทำให้แบรนด์มีการรับรู้ที่ดีขึ้นในสังคมออนไลน์

1.5 ขั้นตอนและวิธีดำเนินงานปริญญานิพนธ์

1.5.1 การศึกษาและรวบรวมข้อมูล (Data Collection)

ทำการศึกษาและรวบรวมข้อมูลจากการเข้าไปติด Keyword บน Zocial eye ซึ่งเป็นเครื่องมือในการดึงข้อมูล Social listening จากแพลตฟอร์มต่างๆ บนโลกโซเชียลออนไลน์

1.5.2 การตั้งเป้าหมาย (Goal)

ตั้งโจทย์หรือคำถามเพื่อหาแนวทางในการวิเคราะห์ข้อมูลว่าสามารถนำไปวิเคราะห์เพื่อหาอะไร และนำไปต่อยอดด้านใดได้บ้าง โดยคณะผู้จัดทำตั้งโจทย์จากการใช้โซเชียลในแต่ละวันที่ถูกถาโถมด้วยคอนเทนต์มากมายในแต่ละวัน แล้วผลิตภัณฑ์เด็กเป็นผลิตภัณฑ์ที่เฉพาะกลุ่ม ส่วนใหญ่เป็นกลุ่มคุณแม่หรือเด็กๆ หากไม่ได้อยู่ในวัยที่ต้องใช้ผลิตภัณฑ์เหล่านี้ ก็แทบจะไม่เคยเห็นผ่านตาบนโลกโซเชียลเลย ถ้าหากแบรนด์สินค้าอยากจะทำการตลาดให้คนทั่วไปได้เห็นบ้างจะต้องลงแพลตฟอร์มไหนหรือใช้อะไรเข้ามาดึงดูดให้คนสนใจ เนื่องจากผลิตภัณฑ์สำหรับเด็กไม่ได้มีแค่คุณแม่หรือลูกๆ เท่านั้นที่สนใจแต่บุคคลที่มีผิวแพ้ง่ายหรือสนใจในกลิ่นที่เป็นเอกลักษณ์ของผลิตภัณฑ์เหล่านี้ พวกเขาก็ต้องการเหมือนกัน คณะผู้จัดทำจึงศึกษาและหาปัจจัยที่จะสามารถดึงดูดความสนใจจากบุคคลทั่วไป โดยสมมติฐานที่คาดไว้นั้นคือ หากมีพรีเซนเตอร์หรือบุคคลที่มีชื่อเสียงโพสต์สินค้าหรือกิจกรรมที่น่าสนใจ จะเป็นสิ่งที่ดึงดูดให้แวะชมคอนเทนต์หรือมีปฏิสัมพันธ์ร่วมกับคอนเทนต์นี้ การจัดกิจกรรมจากทางแบรนด์สินค้าให้มีความน่าสนใจก็อาจจะเป็นสิ่งที่ผู้ชมคอนเทนต์ให้ความสนใจเหมือนกัน นอกจากนี้ยังมีเรื่องของการใช้ แฮชแท็ก (#) เพื่อให้เกิดการค้นหาได้ง่ายจากบุคคลทั่วไป

1.5.3 การเตรียมข้อมูล (Data Preparation)

การนำข้อมูลที่รวบรวมมานำมาทำการทำความสะอาดข้อมูล (Data Cleaning) หรือการเพิ่มคอลัมน์ใหม่สำหรับค่าที่ไม่มีอยู่ โดยมีการทำ Labeling หรือ Clustering เพื่อให้ได้ข้อมูลตามวัตถุประสงค์ที่ต้องการจะวิเคราะห์ ซึ่งเป็นสิ่งสำคัญในขั้นตอนของการทำความสะอาดข้อมูลเพื่อให้ได้ข้อมูลที่ถูกต้องและแม่นยำ

1.5.4 การวิเคราะห์ข้อมูล (Data Analytics)

นำข้อมูลที่ได้จากการทำ Data Preparation มาทำการแบ่งกลุ่ม (Clustering) ตาม Engagement เพื่อหาปัจจัยที่คาดไว้ โดยจะมีการวิเคราะห์ดังนี้

1.5.4.1 การวิเคราะห์ภาพรวมสัดส่วน Engagement ตามแพลตฟอร์มต่างๆ

1.5.4.2 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อทั้งหมดทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์

1.5.4.3 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Product ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์

1.5.4.4 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Promotion ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์

1.5.4.5 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Activity ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์

1.5.5 การนำเสนอการวิเคราะห์ (Data Visualization)

สรุปผลและแสดงข้อมูลในรูปของแผนภาพ มานำเสนอในรูปแบบที่เข้าใจง่ายผ่านการมองเห็น โดยการเลือกใช้กราฟ Scatter plot เพื่อแบ่งกลุ่ม Engagement ในแต่ละหัวข้อคอนเทนต์ทำให้ผู้อ่านเข้าใจได้ง่ายยิ่งขึ้น

1.5.6 จัดทำเอกสารประกอบปริญญานิพนธ์ (Documentation)

เป็นการจัดทำเอกสารแนวทางประกอบปริญญานิพนธ์ วิธีการและขั้นตอนการดำเนินปริญญานิพนธ์ เพื่อนำเสนอรายงานต่ออาจารย์ที่ปรึกษา และคณะกรรมการสอบปริญญานิพนธ์ เพื่อขอรับคำแนะนำ และนำไปใช้เป็นเอกสารอ้างอิงในอนาคต

1.6 ขั้นตอนและระยะเวลาในการดำเนินงานปริญญานิพนธ์

ตารางที่ 1.1 ขั้นตอนและระยะเวลาในการดำเนินงาน

ขั้นตอนการดำเนินงาน	2567						
	ม.ค.	ก.พ.	มี.ค.	เม.ย	พ.ค.	มิ.ย.	ก.ค.
1. รวบรวมข้อมูล	←	→					
2. การเตรียมข้อมูล		←	—	→			
3. การวิเคราะห์ข้อมูล					←	→	
4. การนำเสนอข้อมูลด้วยแผนภาพ					←	→	
5. จัดทำเอกสารปริญญานิพนธ์						←	→

1.7 อุปกรณ์และเครื่องมือที่ใช้ในการพัฒนา

1.7.1 ฮาร์ดแวร์ (Hardware)

1.7.1.1 เครื่องคอมพิวเตอร์ โน๊ตบุ๊ค Macbook Pro M1

- Apple M1 chip
- 16 GB Ram
- 512 GB SSD
- MacOS Sonoma14.2.1

1.7.1.1 เครื่องคอมพิวเตอร์ โน๊ตบุ๊ค Dell Inspiron 3520

- Intel Core i7-1255U
- 8 GB DDR4 2666 MHz
- 512 GB M.2 PCle NVme SSD
- Windows 11

1.7.2 ซอฟต์แวร์ (Software)

1.7.2.1 โปรแกรม Virtual Studio Code

1.7.2.2 โปรแกรม Tableau

1.7.2.3 Zocial Eye

1.7.2.4 โปรแกรม Microsoft excel

บทที่ 2
การทบทวนวรรณกรรมที่เกี่ยวข้อง

ในกระบวนการจัดทำปริญญานิพนธ์นี้ คณะผู้จัดทำได้ศึกษาค้นคว้าแนวคิด ทฤษฎี และเทคโนโลยีที่เกี่ยวข้อง เพื่อวิเคราะห์ข้อมูลและหาปัจจัยที่ส่งผลต่อพฤติกรรมและความสนใจในการเลือกชมคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กบนโซเชียลมีเดีย พร้อมทั้งนำเสนอแผนภาพจากข้อมูลที่ได้ทำการวิเคราะห์หัวข้อโดยมีรายละเอียดดังต่อไปนี้

2.1 Data Analytics[1]

Data Analytics คือ การนำข้อมูลที่มีอยู่มาเข้าสู่กระบวนการวิเคราะห์ เพื่อนำข้อมูลไปใช้ประโยชน์ในด้านต่าง ๆ ที่ครอบคลุมในทุกอุตสาหกรรม ทั้งภาครัฐและภาคเอกชน ซึ่งจะต้องอาศัยเทคโนโลยีหรือซอฟต์แวร์เฉพาะด้านที่เข้ามาช่วยให้การประมวลผลและวิเคราะห์ข้อมูลเป็นไปอย่างแม่นยำและนำไปใช้งานได้อย่างมีประสิทธิภาพ โดยสามารถแบ่งประเภทของการวิเคราะห์ข้อมูล ได้ 4 รูปแบบ คือ

Descriptive Analytics

การวิเคราะห์ข้อมูลแบบพื้นฐาน เพื่อแสดงผลที่เกิดขึ้น หรือกำลังจะเกิดขึ้น จากข้อมูลในอดีต ในลักษณะที่เข้าใจง่ายสามารถสร้างขึ้นได้ด้วยตนเอง เช่น รายงาน แผนภูมิ กราฟ ตาราง เป็นต้น

Diagnostic Analytics

การวิเคราะห์เชิงวินิจฉัย ซึ่งเป็นรูปแบบหนึ่งของการวิเคราะห์ขั้นสูงแบบเจาะลึก โดยการวิเคราะห์ข้อมูลที่ได้จากการทำ Descriptive Analytics เพื่อหาคำตอบว่าทำไมจึงเกิดสิ่งนั้น ๆ หรืออธิบายปัจจัยและตัวแปรที่เป็นสาเหตุของการเกิดสิ่งนั้น ๆ ขึ้น ซึ่งจะต้องอาศัยเทคนิคต่าง ๆ เข้ามาช่วย เช่น การทำ Data discovery หรือ Data mining เป็นต้น

Predictive Analytics

การวิเคราะห์ข้อมูลทั้งข้อมูลในอดีตและปัจจุบันออกมาในเชิงคาดการณ์ ทำนาย หรือการพยากรณ์ เพื่อหาแนวโน้มที่จะเกิดสิ่งต่าง ๆ ขึ้นตามวัตถุประสงค์ที่กำหนด โดยการสร้างแบบจำลองทางสถิติ บวกกับการนำเทคโนโลยีปัญญาประดิษฐ์มาใช้ ซึ่งสามารถสร้างประโยชน์ได้มากมายในหลายแง่มุม การทำ Predictive Analytics ที่ถูกต้องและแม่นยำนั้น ขึ้นอยู่กับคุณภาพของข้อมูล

---

[1] https://www.blendata.co/ทำความรู้จัก-data-analytics-ศาสตร์แห่/

โดยการเตรียมข้อมูลให้มีคุณภาพที่ดีและเหมาะสม ก่อนนำไปใช้วิเคราะห์เพื่อให้เกิดผลลัพธ์ที่มีประสิทธิภาพ

Prescriptive Analytics

การวิเคราะห์แบบให้คำแนะนำ เป็นการวิเคราะห์ที่มีความซับซ้อนมากที่สุด ต่อเนื่องจากการทำ Predictive Analytics กล่าวคือ เมื่อได้ข้อมูลแนวโน้มที่จะเกิดบางสิ่งขึ้นแล้ว การทำ Prescriptive Analytics จะช่วยแนะนำแนวทางการดำเนินการในขั้นตอนต่อไปที่เหมาะสมที่สุด และวิเคราะห์ไปถึงผลที่จะเกิดขึ้นถ้าหากเลือกปฏิบัติตามแนวทางนั้น ๆ หรือแม้แต่แนะนำแนวทางในการรับมือและแก้ไขปัญหา การวิเคราะห์แบบให้คำแนะนำจึงถือเป็นเครื่องมือที่สำคัญอย่างมากสำหรับการตัดสินใจที่ขับเคลื่อนด้วยข้อมูล การวิเคราะห์แบบให้คำแนะนำคือทำงานร่วมกันระหว่าง Big data อัลกอริทึมของ Machine learning และเทคโนโลยี AI เพื่อช่วยในการวิเคราะห์ข้อมูลจำนวนมหาศาลที่มีความซับซ้อนเกินกว่าที่มนุษย์จะทำได้ การทำ Prescriptive Analytics ยังช่วยเพิ่มประสิทธิภาพในการตัดสินใจด้านต่าง ๆ

2.2 Clustering[2]

Clustering Model คือ Machine Learning Model ประเภท Unsupervised ที่ไม่มี Target หรือไม่มีต้นแบบของผลลัพธ์ ซึ่งเป็น Model ที่นำไปใช้ในการจัดกลุ่มของข้อมูลที่ไม่เคยมีการจัดกลุ่มมาก่อน โดยจะแบ่งกลุ่มข้อมูลจากความคล้าย Algorithms ที่ใช้ในการทำ Clustering Model ได้แก่

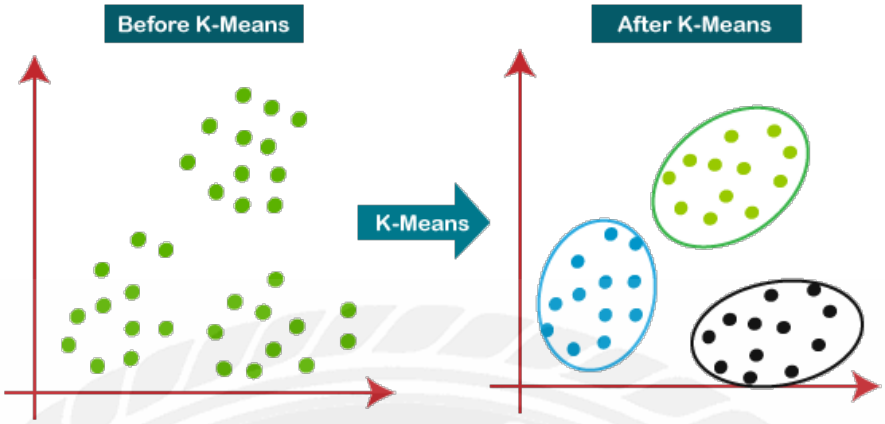
K-Mean Clustering[3]

K-means คือ การแบ่งกลุ่ม แบบ Clustering การแบ่งกลุ่มในลักษณะนี้จะใช้พื้นฐานทางสถิติ หน้าที่ของ clustering คือ การจับกลุ่มของข้อมูลที่มีลักษณะใกล้เคียงกันเป็นกลุ่มเดียวกัน เช่น ในการศึกษาข้อมูลและหาปัจจัยที่ส่งผลต่อพฤติกรรมและความสนใจในการเลือกชมคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กบนโซเชียลมีเดีย ได้มีการนำข้อมูลมาแบ่งกลุ่มตาม Engagement ให้ได้ออกมา 3 Clustering คือ Low Engagement, Medium Engagement และ High Engagement

---

[2] https://medium.com/tni-university/การทำ-machine-learning-ด้วย-clustering-model-2a3c392e7faa

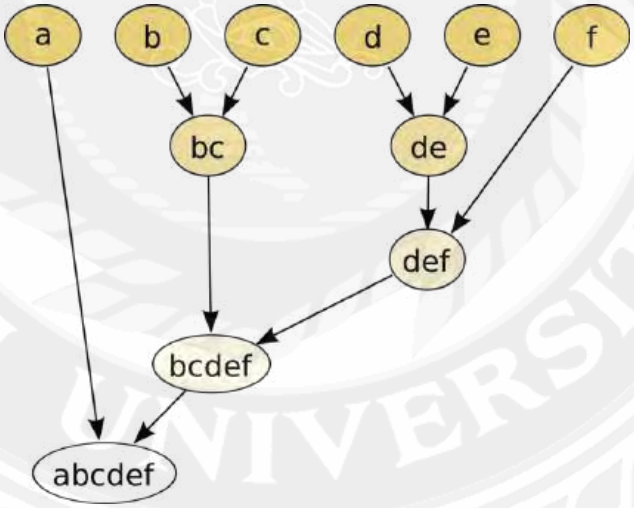
[3] https://medium.com/@Jutharath.Thankittikoon/what-is-k-means-clustering-ee36ab6f7638



รูปที่ 2.1 ตัวอย่างการแบ่งกลุ่มแบบ K-means

Hierarchical Clustering[4]

Hierarchical Clustering คือ การวิเคราะห์กลุ่มแบบลำดับชั้นโดยขั้นตอนในการ cluster นั้นจะมีการแบ่งกลุ่มออกเป็น 2 กลุ่มโดยกลุ่มแรกคือ agglomerative clustering หรือ Divisive clustering (การแบ่งกลุ่มแบบแยก) และ Hierarchical Clustering จะมีการเชื่อมโยงข้อมูลคล้ายกับโครงสร้างต้นไม้



รูปที่ 2.2 ตัวอย่างลักษณะของการ clustering มีความคล้ายกับต้นไม้

[4] https://www.mindphp.com/คู่มือ/73-คืออะไร/6851-what-is-hierarchical-clustering.html

2.3 Data Preparation[5]

Data Preparation หรือการเตรียมข้อมูล หมายถึง กระบวนการใดๆ ที่จำเป็นต้องทำกับข้อมูลดิบ (raw data) ที่ได้รับมา เพื่อปรับเปลี่ยนข้อมูลให้อยู่ในรูปแบบที่เหมาะสม ที่จะนำไปโหลดเข้าฐานข้อมูล หรือนำไปวิเคราะห์ต่อไป โดยคณะผู้จัดทำได้ทำการเขียนชุดคำสั่งในโปรแกรม Visual Studio Code ในการทำ Data Labeling เพื่อแบ่งประเภทของโพสต์ออกเป็นหัวข้อต่างๆ เช่น รีวิวสินค้า การจัดโปรโมชัน การร่วมสนุกหรือจัดกิจกรรม และพรีเซนเตอร์ นอกจากนี้ยังแบ่งประเภทผู้โพสต์ออกเป็น บุคคลทั่วไป แบรนด์ Influencer สำนักข่าว และร้านค้า

รูปที่ 2.3 ตัวอย่างกระบวนการ data cleaning

2.4 Data Visualization[6]

Data Visualization คือ การนำข้อมูลหรือ Data ที่ได้มาจากแหล่งข้อมูลต่างๆ มาวิเคราะห์ประมวลผลแล้วนำเสนอออกมาในรูปแบบที่มองเห็นและทำความเข้าใจได้ด้วยตา เช่น แผนภูมิ รูปภาพ แผนที่ กราฟแสดงเทรนด์ ตาราง วิดีโอ อินโฟกราฟิก (Infographic) แดชบอร์ด (dashboard) ฯลฯ เพื่อการนำเสนอข้อมูลให้เข้าใจง่าย ผู้อ่านข้อมูลสามารถเข้าใจได้ทันทีว่าตัวชิ้นงาน (media) ต้องการสื่อสารอะไร ชี้จุดสำคัญของเนื้อหา และชี้ Insight ข้อเปรียบเทียบให้เห็นอย่างชัดเจน ช่วยให้สังเกตเห็นจุดที่น่าสนใจของข้อมูลได้ง่ายขึ้น

Data Visualization มีหลากหลายรูปแบบและไม่จำกัดว่าต้องใช้รูปแบบต่อไปนี้ในการนำเสนอข้อมูลเท่านั้น เพราะแต่ละรูปแบบก็มีฟังก์ชันเฉพาะของการนำเสนอข้อมูล บางรูปแบบใช้เปรียบเทียบข้อมูลแต่ละชุดได้ดี บางรูปแบบช่วยให้มองเห็นเทรนด์ได้ง่าย บางรูปแบบช่วยเล่าข้อมูลที่ไกลตัวให้เข้าใจได้ง่ายโดยการเปรียบเทียบให้สอดคล้องกับสิ่งที่คุ้นเคยในชีวิตประจำวัน และ 6 รูปแบบการทำ Data Visualization ต่อไปนี้ คือ รูปแบบพื้นฐาน

แผนภูมิ (Charts)

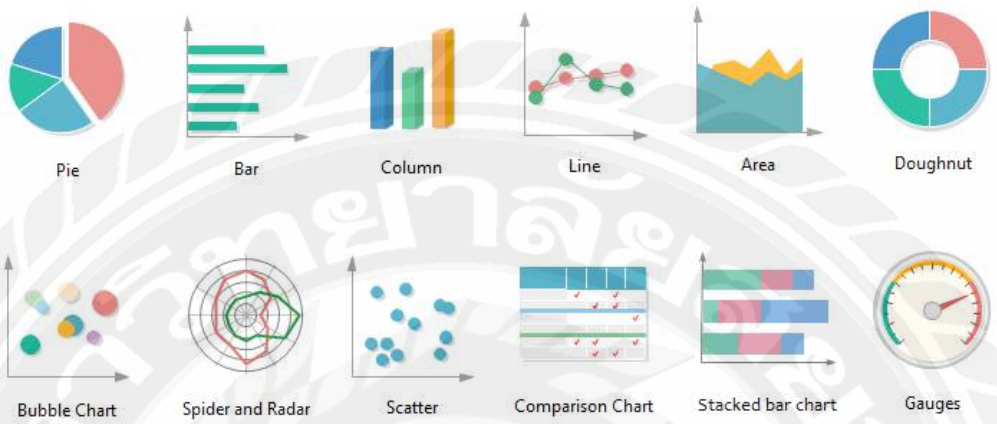
Data Visualization รูปแบบแรก คือ แผนภูมิ (Charts) ซึ่งเป็นรูปแบบที่น่าจะคุ้นเคยกันมากที่สุด และเป็นรูปแบบที่มีหลากหลายชนิดที่เหมาะกับการนำเสนอข้อมูลที่แตกต่างกันไปตาม

---

[5] https://bzinsight.wordpress.com/2014/06/11/การทำ-data-preparation-อย่างมืออาชีพ/

[6] https://1stcraft.com/what-is-data-visualization/

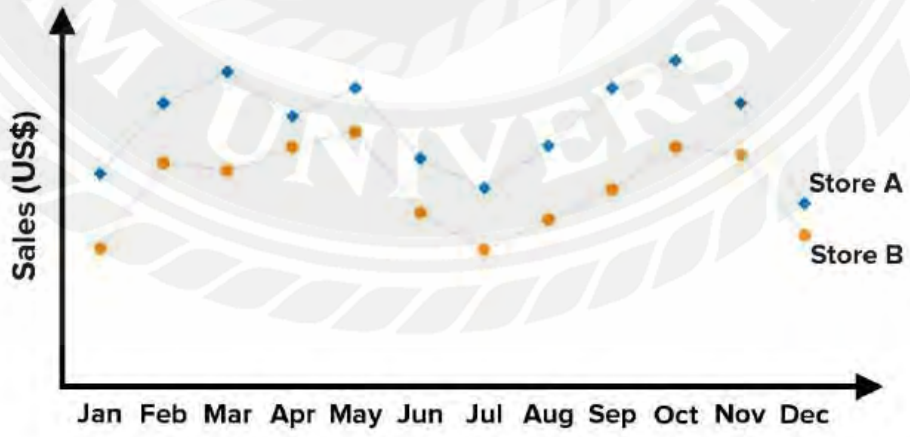
วัตถุประสงค์ เช่น Pie chart จะช่วยให้เราเห็นปริมาณความแตกต่างได้ชัดเจน, Comparison chart เหมาะสำหรับการเปรียบเทียบคุณสมบัติหลายๆ ข้อ, มาตรวัด (Gauges) จะช่วยให้เห็นความเข้มข้น ความรุนแรง หรือน้ำหนัก เช่น แผนภูมิแท่งที่แสดงถึง Brand ผลิตภัณฑ์สำหรับเด็กที่มีจำนวน Engagement สูงที่สุด 5 Brand



รูปที่ 2.4 ตัวอย่าง Data Visualization แบบแผนภูมิ (Charts)

กราฟ (Graphs)

กราฟ (Graphs) คือ subset หรือประเภทหนึ่งของแผนภูมิ โดยกราฟจะทำหน้าที่แสดงความสัมพันธ์ระหว่างข้อมูล 2 ตัวแปร ผ่านแกนแนวนอน (แกน X) และแกนแนวตั้ง (แกน Y) ช่วยให้เห็นเทรนด์สถานการณ์ประกอบกับบริบทได้เป็นอย่างดี เช่น กราฟ Brand ผลิตภัณฑ์สำหรับเด็กที่มีจำนวน Engagement ในระยะเวลา 6 เดือน



รูปที่ 2.5 ตัวอย่าง Data Visualization แบบกราฟ (Graphs)

ตาราง (Tables)

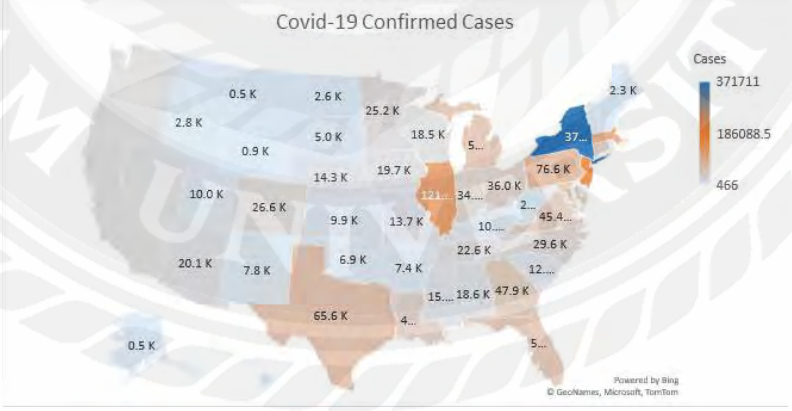
ตาราง (Tables) เป็นอีกรูปแบบเพื่อนำเสนอข้อมูลให้ออกมาดูง่าย ตารางประกอบไปด้วย 2 ส่วน ได้แก่ คอลัมน์และแถว ซึ่งช่วยจัดการข้อมูลให้เรียบร้อย ช่วยให้มองเห็นบริบทและความสัมพันธ์ของข้อมูลหลายๆ ชุดได้อย่างง่ายดาย

Marks	Number of Students		Total
	Males	Females	
30 – 40	8	6	14
40 – 50	16	10	26
50 – 60	14	16	30
60 – 70	12	8	20
70 – 80	6	4	10
Total	56	44	100

รูปที่ 2.6 ตัวอย่าง Data Visualization แบบตาราง (Tables)

แผนที่ (Maps)

แผนที่ (Maps) เป็นการนำเสนอข้อมูลบนแผนที่เพื่อแสดงข้อมูลเกี่ยวกับพื้นที่ต่างๆ ยกตัวอย่างเช่น การนำเสนอข้อมูลยอดผู้ติดเชื้อ Covid-19 ในแต่ละรัฐของประเทศสหรัฐอเมริกา ซึ่งนอกจากการใส่ข้อมูลลงไปยังพื้นที่ต่างๆ แล้ว ยังสามารถใช้สีสันเพื่อบอกช่วงปริมาณหรือความหนาแน่นของผู้ติดเชื้ออีกด้วย



รูปที่ 2.7 ตัวอย่าง Data Visualization แบบแผนที่ (Maps)

อินโฟกราฟิก (Infographic)

อินโฟกราฟิก (Infographic) การนำเสนอสารสนเทศ (Info: information) ด้วยภาพกราฟิก (Graphic) เป็นรูปแบบการนำเสนอข้อมูลที่ใช้ภาพสื่อแทน ทำให้ผู้อ่านข้อมูลเข้าใจข้อมูลได้ง่าย หรือสามารถทำความเข้าใจผ่านภาพแทนที่คุ้นเคย

แดชบอร์ด (Dashboards)

แดชบอร์ด (Dashboards) การนำข้อมูลต่างๆ มาเรียบเรียงและสรุปเป็นภาพ โดยใช้แผนภูมิและกราฟต่างๆ มาใช้นำเสนอ ปัจจุบันแดชบอร์ดเป็น Data Visualization ที่นิยมใช้กับการนำเสนอข้อมูลแบบ Real-time ผ่านซอฟต์แวร์หรือเครื่องมือจัดการ และวิเคราะห์ข้อมูลต่างๆ

รูปที่ 2.8 ตัวอย่าง Data Visualization แบบแดชบอร์ด (Dashboards)

2.5 Tableau[7]

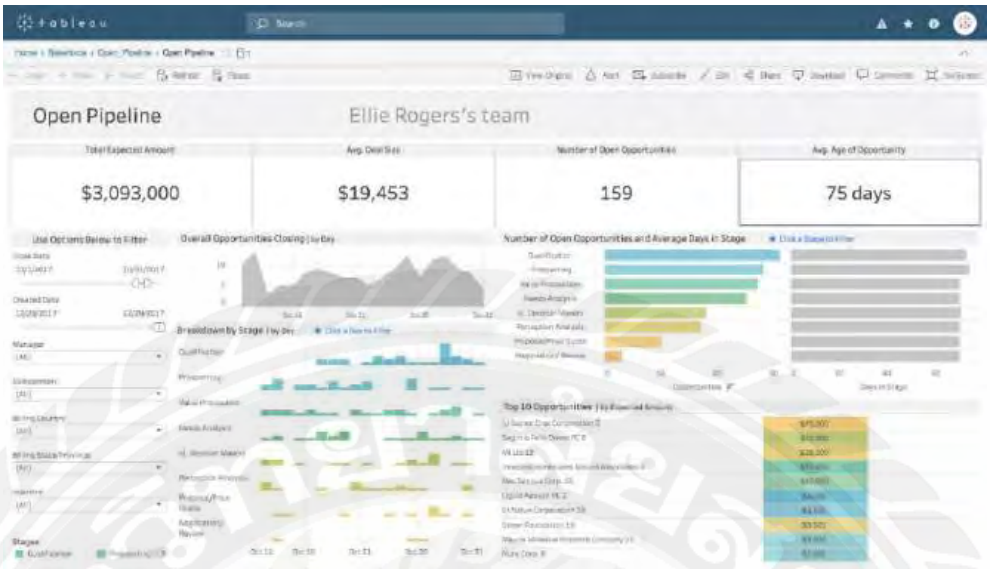
Tableau เป็นโปรแกรมสำหรับการวิเคราะห์ข้อมูล และหาคำตอบในเชิงธุรกิจอย่างรวดเร็ว โดยการใช้ Tableau จะเป็นการช่วยให้ผู้ใช้งานมีความเข้าใจในข้อมูลมากขึ้น รวมถึงผู้ใช้งานสามารถนำข้อมูลที่มีอยู่มาวิเคราะห์และแสดงผลอย่างมีประสิทธิภาพ Tableau สามารถนำข้อมูลที่มีอยู่ มาสร้างเป็น Visualization และสามารถสร้างเป็น Dashboard ได้ และสามารถ Share Content ได้อย่างรวดเร็วด้วย Tableau Server

โปรแกรม Tableau เป็นโปรแกรมที่ใช้งานง่าย ถูกสร้างขึ้นมาโดยผู้เชี่ยวชาญด้าน Computer Graphic, Database และ Human-Computer Interaction จึงทำให้โปรแกรมเป็นมิตรกับทั้งทาง Business User และทาง IT

โดยคณะผู้จัดทำได้นำ Tableau มาใช้ในการแสดง Visualization เพื่อให้ง่ายต่อการวิเคราะห์ข้อมูลและเห็นภาพชัดเจน โดยผู้คณะผู้จัดทำนำ Tableau มาประยุกต์ใช้ทำในส่วนของการวิเคราะห์ข้อมูลเปรียบเทียบข้อมูลที่มีอยู่ และสร้างแผนภาพให้เกิดการเข้าใจง่ายต่อผู้รับฟังในการนำเสนอ

---

[7] https://www.bac.co.th/web/products-and-services/tableau-software/



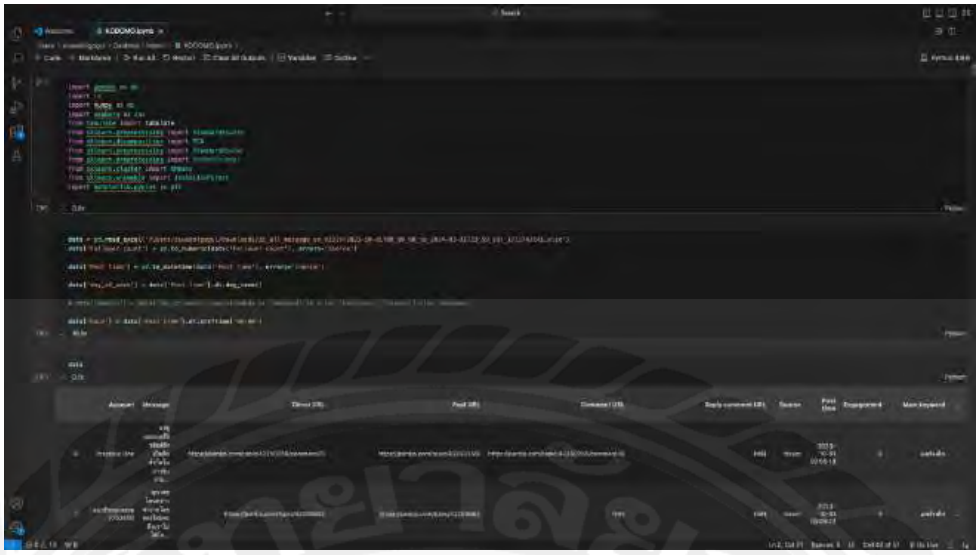
รูปที่ 2.9 ตัวอย่างโปรแกรม Tableau

2.6 Visual studio code [8]

VS Code (Visual studio code) คือ โปรแกรมประเภท Editor หรือเครื่องมือในการเขียนโปรแกรม โดยจะใช้ในการแก้ไข Code ที่มีขนาดเล็ก แต่มีประสิทธิภาพสูง เหมาะสำหรับนักพัฒนาโปรแกรมในทุกระดับ ไม่ว่าจะอยู่ในระดับเริ่มต้นยันมืออาชีพเลย รองรับการใช้งานทั้ง Windows, MacOS และ Linux รวมทั้งรองรับได้หลายภาษาไม่ว่าจะเป็น JavaScript, TypeScript, Python, C++ และ อื่นๆ สามารถนำมาใช้งานได้ง่ายไม่ซับซ้อน ซึ่งมีส่วนขยายหรือเครื่องมืออำนวยความสะดวกที่ให้เลือกใช้อยู่มาก

โดยทางคณะผู้จัดทำได้ใช้ Visual studio code ในการทำความสะอาดข้อมูลและการแบ่งกลุ่มข้อมูลในการหา Cluster ด้วยภาษา Python

[8] https://codingonblog.com/what-is-vscode-codingonblog/



รูปที่ 2.10 ตัวอย่างหน้าจอโปรแกรม Visual Studio Code ที่ใช้ในการพัฒนา

2.7 Social Listening [9]

Social Listening คือ การฟังเสียงโซเชียลมีเดียจากทุกแพลตฟอร์มที่ครอบคลุมข้อความทุกประเภท เช่น โพสต์สาธารณะบนเพจ Facebook, Twitter, Instagram และแพลตฟอร์มโซเชียลมีเดียอื่นๆ, การคอมเมนต์บนโพสต์ต่างๆ รวมถึงการตอบ Reply บน Twitter และคอมเมนต์วิดีโอบน YouTube, ข้อความที่ติดแฮชแท็ก (#Hashtag), การแท็กเพื่อนโดยการเมนชัน (@mention) ฯลฯ โดยนำข้อความเหล่านี้มาสกัดออกมาเป็นข้อมูลเชิงลึก (Insight) ที่สามารถนำไปประยุกต์ใช้กับธุรกิจหรือหน่วยงานได้ โดยปัจจุบัน Social Listening จะมีหลักการทำงานอยู่ 2 รูปแบบ คือ

- การหาประเด็นหรือหัวข้อที่สนใจจากคำ Keyword เช่น หากนักการตลาดสนใจคำว่าสบู่ ก็จะเก็บข้อมูลจากทุกช่องทางโซเชียลมีเดียที่พูดถึงคำว่า ‘สบู่’
- การหาประเด็นหรือหัวข้อที่สนใจจากการติดตามบัญชีผู้ใช้ (Track By Account) เช่น ตั้งค่าให้ติดตามบัญชีของคู่แข่ง เพื่อดูว่าคู่แข่งกำลังพูดถึงเรื่องอะไรบ้าง

2.8 ZOCIAL EYE [10]

ZOCIAL EYE คือ เครื่องมือวิเคราะห์ข้อมูลโซเชียลมีเดีย ที่สามารถดึงข้อมูลจากแหล่งโซเชียลมีเดียต่างๆ ไม่ว่าจะเป็น Facebook, Twitter, Instagram, TikTok, Forum (Pantip), Youtube, News, Blog และสรุปออกมาเป็น Insight หรือข้อมูลเชิงลึก ผ่านแดชบอร์ดที่มีหลากหลายตัวชี้วัด (Metric) เช่น การดูสัดส่วนของข้อความที่เกิดขึ้นบนช่องทางต่างๆ (Share of Voice), ผลสรุปข้อมูล

---

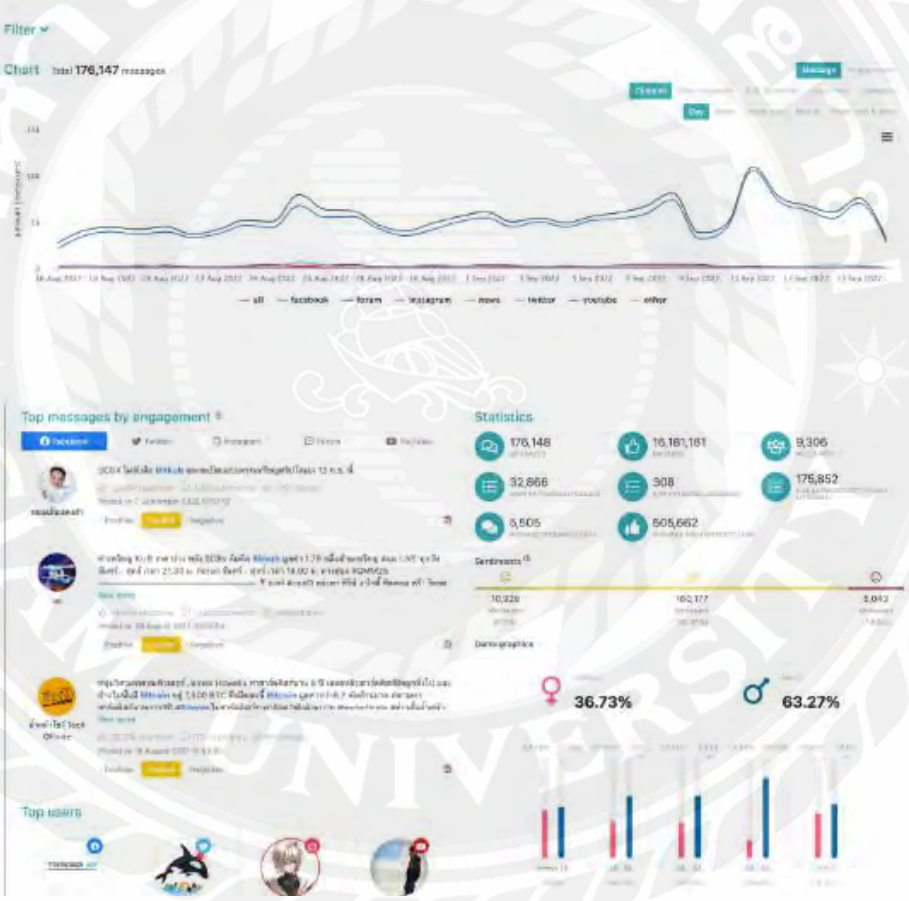
[9] https://wisesight.com/th/news/social-listening-2/

[10] https://wisesight.com/th/news/zocial-eye/

การมีส่วนร่วมของผู้บริโภค (Engagement), ผลสรุปของประเภท ‘เสียง’ ของผู้บริโภค เช่น เป็นเสียงเชิงบวกหรือเชิงลบ(Sentiment), ผลสรุปของกลุ่มคำที่ผู้บริโภคใช้ในการพูดถึงแคมเปญหรือแบรนด์ (Word Cloud) เป็นต้น สำหรับมุมมองของการดูข้อมูลของ ZOCIAL EYE สามารถดูได้ใน 3 รูปแบบ คือ

Campaign View

Campaign View ทำหน้าที่แสดงผลในภาพรวมที่เข้าใจง่าย และเข้าถึงข้อมูลแบบเชิงลึกได้อย่างครบถ้วน ใช้สำหรับดูหัวข้อที่ต้องการตามช่วงเวลาของแคมเปญการตลาด (Marketing Campaign) ในทุกๆ แพลตฟอร์มของโซเชียลมีเดีย โดยสามารถใช้ดูกราฟของเสียงผู้บริโภคเกี่ยวกับแคมเปญที่ทำการปล่อยออกไปว่ามีการพูดถึงอย่างไร มีภาพสรุปของอินฟลูเอนเซอร์ที่เกี่ยวข้องกับแคมเปญ รวมถึง ใช้เพื่อทำวิจัยตลาด (Market Research) ดูความต้องการ หรือรสนิยมผู้บริโภคได้



รูปที่ 2.11 ตัวอย่างฟีเจอร์ Campaign View ของ ZOCIAL EYE

Comparison View

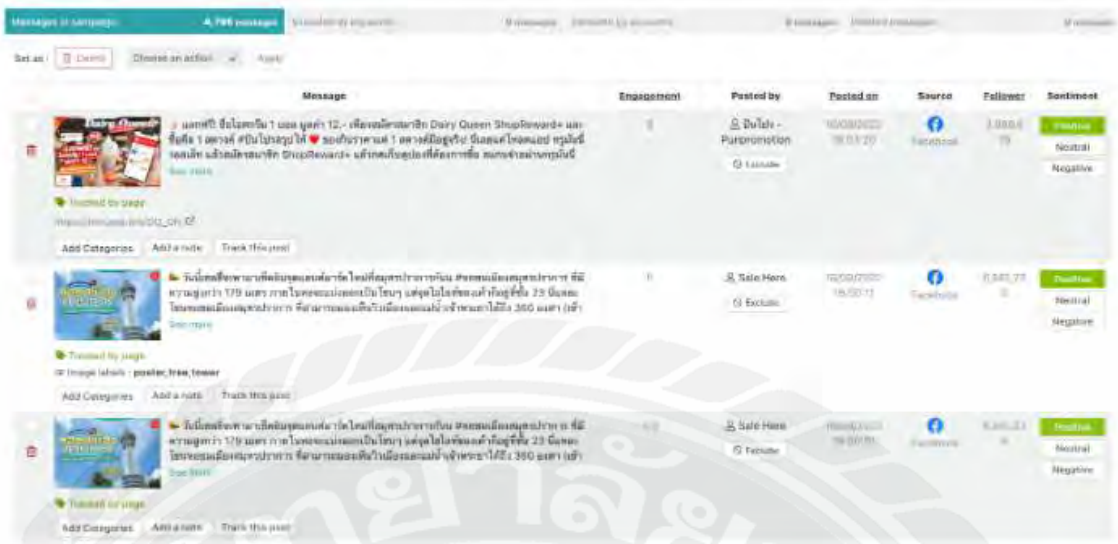
Comparison View เป็นหน้าที่นำมาใช้เปรียบเทียบเรื่องต่างๆ ที่เกิดขึ้นบนโซเชียลมีเดียตามที่ผู้ใช้งานต้องการ เช่น ใช้เพื่อวิเคราะห์และเปรียบเทียบคู่แข่ง (Competitive Analysis),

เปรียบเทียบว่าเดือนไหนมีจำนวนข้อความมากหรือน้อยกว่ากัน, เปรียบเทียบประสิทธิภาพของช่องทางโซเชียลมีเดียต่างๆ, เปรียบเทียบสัดส่วนคำค้นหา (Keyword) ที่ต้องการทราบ เป็นต้น นอกจากนี้ยังใช้เพื่อการวิเคราะห์ภาพรวมตลาด (Market Analysis) และการวิเคราะห์ผลิตภัณฑ์ (Product/ Service Analysis) ได้

รูปที่ 2.12 ตัวอย่างฟีเจอร์ Comparison View ของ ZOCIAL EYE

Trend Views

Trend Views เป็นมุมมองที่ใช้สำหรับการค้นหาเทรนด์หรือกระแสที่เป็นที่พูดถึงหรือคนสนใจ สามารถทำการฟิลเตอร์เพื่อค้นหาคอนเทนต์ที่สนใจ ใช้ในการหากระแสที่คนกำลังพูดถึงแบบ Real Time หรือจะใช้เพื่อดูแนวโน้มของเทรนด์ก็ได้เช่นกัน เช่น สามารถฟิลเตอร์เรียงลำดับจากคอนเทนต์ที่ได้รับเอ็นเกจเมนต์ (Engagement) สูงสุดได้ เป็นต้น



รูปที่ 2.13 ตัวอย่างฟีเจอร์ Trend Views ของ ZOCIAL EYE

2.9 PCA (Principal component analysis)[11]

PCA เป็นวิธีการลด Dimension ของ Dataset ที่มีขนาดใหญ่ ด้วยการแปลง Variables ที่มีจำนวนมาก ให้มีจำนวนน้อยลงแต่ยัง Contains ข้อมูลส่วนใหญ่ของชุดข้อมูลไว้ได้ การลดจำนวน Variables ของชุดข้อมูลย่อมแลกมาด้วยการสูญเสียความแม่นยำเล็กน้อย อย่างไรก็ตามการลด Dimension ของข้อมูลจะช่วยให้การวิเคราะห์ง่ายและสะดวกมากขึ้น เนื่องจากชุดข้อมูลที่มีขนาดเล็กกว่านั้นง่ายต่อการ Explore และ Visualize การวิเคราะห์ข้อมูลจึงรวดเร็วมากขึ้นสำหรับ Machine Learning Algorithms โดยไม่ต้องประมวลผล Variables จำนวนมาก

ขั้นตอนของ PCA

- **Standardization**

  จุดมุ่งหมายของขั้นตอนนี้คือการ Standardize Range ของ Continuous Variables ใน Dataset เพื่อให้แต่ละ Variables มีผลต่อการวิเคราะห์เท่าๆ กัน วิธีการทำ Standardization คือ นำค่าของแต่ละตัวแปร ลบด้วยค่าเฉลี่ย แล้วหารด้วย Standard Deviation ดังสมการด้านล่าง

$$x_{\text{stand}} = \frac{x - \text{mean}(x)}{\text{standard deviation } (x)}$$

รูปที่ 2.14 สมการการทำ Standardization

[11] https://medium.com/kbtg-life/principal-component-analysis-pca-ฉบับละเอียดและเข้าใจง่าย-d0c514e6fa0a

หลังจากที่ทำการ Standardize เสร็จ ตัวแปรทั้งหมดจะถูกแปลงเป็น Scale เดียวกัน

- **Covariance Matrix Computation**

ขั้นตอนนี้ คือ การดูว่าแต่ละตัวแปรมีความสัมพันธ์กันหรือไม่ เนื่องจากบางครั้งตัวแปรมีความสัมพันธ์กันสูงในลักษณะที่มีข้อมูลซ้ำซ้อนกัน เพื่อระบุความสัมพันธ์เหล่านี้ เราจึงจำเป็นต้องคำนวณ Covariance Matrix ขึ้นมา

Covariance Matrix คือเมทริกซ์สมมาตร p × p (โดยที่ p คือจำนวน Dimension) ที่มี Covariance ของคู่ที่เป็นไปได้ทั้งหมดของตัวแปรในชุดข้อมูล เนื่องจาก Covariance ของตัวแปรใดตัวแปรหนึ่งกับตัวมันเอง เท่ากับ Variance ของตัวแปรนั้นๆ ก็คือ Variance ของแต่ละตัวแปรนั่นเอง และเนื่องจาก Covariance เป็น Commutative Covariance Matrix จึงมีความสมมาตรเมื่อเทียบกับเส้นทแยงมุมหลัก ซึ่งหมายความว่า Upper กับ Lower Triangle มีค่าเท่ากัน หลังจากที่เราคำนวณ Covariance Matrix เสร็จ สิ่งที่สนใจคือ Sign ของ Covariance แต่ละตัวใน Matrix ถ้าเป็นค่าบวก ตัวแปรทั้งสองเพิ่มขึ้น หรือลดลงพร้อมกัน (Correlated) ถ้าเป็นค่าลบ ตัวแปรหนึ่งเพิ่มขึ้น เมื่ออีกตัวแปรลดลง (Inversely Correlated)

- **Compute the Eigenvectors and Eigenvalues of the Covariance Matrix**

Eigenvectors และ Eigenvalues เป็น Linear Algebra Concepts ที่จำเป็นต้องคำนวณจาก Covariance Matrix เพื่อกำหนด Principal Components ของข้อมูล

Principal Components คือตัวแปรใหม่ที่สร้างขึ้นเป็น Linear Combinations หรือ Mixtures ของตัวแปรตั้งต้นจากชุดข้อมูล โดย Combinations เหล่านี้จะไม่มีความสัมพันธ์กัน และข้อมูลส่วนใหญ่ภายในตัวแปรเริ่มต้นจะถูก Squeeze หรือ Compress ลงใน First Component แนวคิดคือ 10-Dimensional Data มี 10 Principal Components แต่ PCA จะพยายามใส่ข้อมูลที่เป็นไปได้สูงสุดใน First Component จากนั้นข้อมูลสูงสุดที่เหลืออยู่ใน Second Component และทำแบบนี้ต่อไปเรื่อยๆ

- **Feature Vector**

ขั้นตอนที่แล้ว การคำนวณ Eigenvectors และเรียงลำดับตาม Eigenvalues ในลำดับจากมากไปน้อย ทำให้สามารถค้นหา Principal Components ตามลำดับความสำคัญได้ สิ่งที่ทำในขั้นตอนนี้คือเลือกว่าจะเก็บ Components เหล่านี้ทั้งหมดหรือทิ้ง Components ที่มีนัยสำคัญน้อยกว่าไป (ที่มี Eigenvalues ต่ำ) และสร้างด้วย Components ที่เหลือเป็น Matrix ของเวกเตอร์ที่เราเรียกว่า Feature Vector

- **Recast the Data along the Principal Components Axes**

ขั้นตอนสุดท้าย จุดมุ่งหมายคือการใช้ Feature Vector ที่สร้างขึ้นโดยใช้ Eigenvectors ของ Covariance Matrix เพื่อปรับทิศทางข้อมูลจากแกนเดิมไปยังแกนที่แสดงโดย Principal Components (ด้วยเหตุนี้จึงตั้งชื่อว่า Principal Components Analysis) ซึ่งทำได้โดยการคูณทรานสโพสของข้อมูลเดิมที่กำหนด ด้วยทรานสโพสของ Feature Vector นั่นเอง

คณะผู้จัดทำได้นำวิธีนี้มาใช้ในการลดมิติของ Feature ลงเพื่อให้การวิเคราะห์ข้อมูลง่ายยิ่งขึ้น

2.10 Word Cloud[12]

รูปที่ 2.15 Word Cloud

Word Cloud คือ การจับกลุ่มคำโดยเรียงลำดับจากคำที่มีมากที่สุดไปน้อยที่สุด โดยแสดงเป็นตัวหนังสือขนาดต่างๆ ตามความถี่ที่พบจากกลุ่มคำที่ซ้ำกัน ยิ่งคำซ้ำกันมากขนาดของตัวหนังสือยิ่งใหญ่ขึ้น ทั้งนี้เพื่อให้มองเห็นคำที่ถูกใช้มากที่สุดได้ง่ายขึ้น

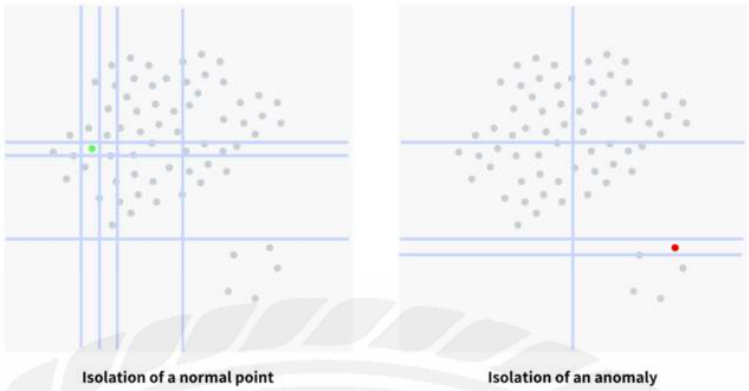
2.11 Isolation Forest[13]

Isolation Forest มีรากฐานมาจาก Decision Tree มีเป้าหมาย คือ การแบ่งข้อมูลไปเรื่อยๆ จนกระทั่งข้อมูลนั้นไม่สามารถแบ่งให้เล็กกว่านี้ได้อีกแล้ว หรือก็คือแบ่งจนกว่าข้อมูลแต่ละตัวจะแยกจากกันโดยสมบูรณ์นั่นเอง

---

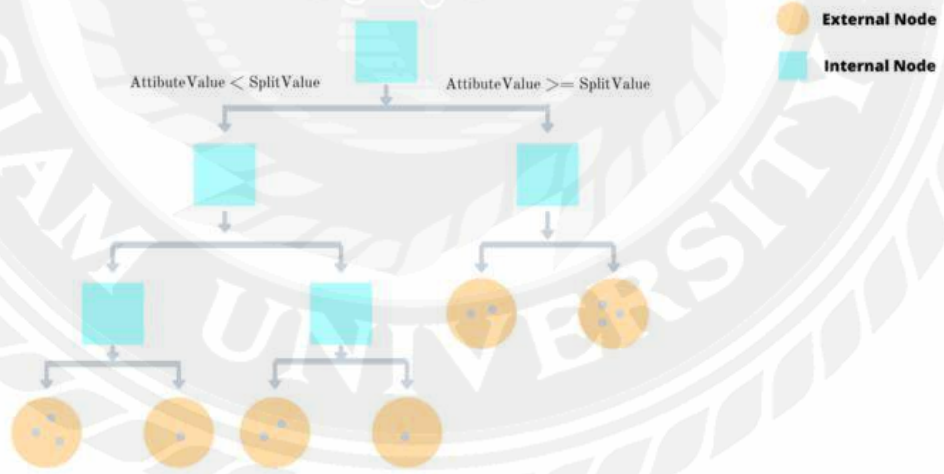
[12] https://jirat-boonphun.medium.com/word-cloud-สำหรับวันวาเลนไทน์-23bc88af4f3c

[13] https://bigdataexperience.org/anomaly-detection-with-isolation-forest-แยกข้อมูลผิดปกติง/



รูปที่ 2.15 ตัวอย่างการแบ่งข้อมูลปกติ [ซ้าย] และข้อมูลผิดปกติ [ขวา]

ต้นไม้ตัดสินใจ (Decision Tree) จะได้ว่าเส้นแบ่งแต่ละเส้นก็คือเส้นที่ตัดแบ่งข้อมูลทั้งหมดออกเป็นกิ่งซ้าย และกิ่งขวา แล้วทำการแบ่งไปจนกระทั่งถึงจุดที่ทุกอย่างแยกออกจากกัน ทำให้ได้ต้นไม้ที่มีกิ่งแยกข้อมูลออกจากกัน หลังจากนั้น จะพบว่าสามารถแยกข้อมูลผิดปกติออกจากข้อมูลปกติได้ด้วย “ความลึก” ของต้นไม้ เนื่องจากข้อมูลปกติจะใช้จำนวนชั้นของต้นไม้ที่ลึกมากในการแบ่งข้อมูลให้เป็นอิสระจากกัน แต่ข้อมูลผิดปกติจะถูกกรองตั้งแต่ชั้นแรกๆ ของต้นไม้แล้วทำให้ความลึกของข้อมูลที่ผิดปกติจะตื้นกว่าข้อมูลอื่นๆ



รูปที่ 2.16 ตัวอย่าง ต้นไม้ตัดสินใจ (Decision Tree)

เมื่อทราบความลึกของต้นไม้แล้วสามารถคำนวณเป็นคะแนนความผิดปกติ (Anomaly score) เพื่อใช้ในการแยกประเภทของข้อมูลได้ ซึ่งคะแนนความความผิดปกตินั้นจะมีค่าตั้งแต่ 0 ถึง 1 โดยค่าที่เข้าใกล้ 1 จะหมายถึงข้อมูลที่มีแนวโน้มผิดปกติ ส่วนข้อมูลที่มีค่าน้อยกว่า 0.5 ลงไปจะ

ถือว่าเป็นข้อมูลทั่วไปที่ไม่มีความผิดปกติ เพียงเท่านี้เราก็สามารถแยกข้อมูลผิดปกติด้วย Isolation Forest

Isolation forest สามารถทำให้การตรวจสอบความผิดปกติในงานต่างๆเป็นเรื่องที่ง่ายขึ้น ซึ่งใช้เวลาในการประมวลผลค่อนข้างน้อย อีกทั้งยังให้ผลลัพธ์ที่แม่นยำในระดับที่ยอมรับได้ แม้ว่ามีข้อจำกัดบางอย่าง เช่น ไม่สามารถใช้กับข้อมูลที่ไม่ใช่ตัวเลข เป็นต้น แต่ก็สามารถประยุกต์ใช้วิธีนี้ได้ ถ้าหากหาความเชื่อมโยงระหว่างข้อมูลชนิดนั้นๆ กับตัวเลขได้

โดยคณะผู้จัดทำได้นำฟังชันก์ Isolation forest มาทำการกำจัด Outlier ซึ่งจะได้คอลัมน์ที่ชื่อว่า “anomaly” โดยมีค่า 1 (ไม่ใช่ Outlier) และ -1 (เป็น Outlier) เป็นค่าในแต่ละแถวเพื่อบ่งบอกว่าแถวไหนมีค่า Outlier และแถวไหนไม่มีค่า Outlier

```
iso_forest = IsolationForest(contamination=0.01, random_state=42)
data_encoded['anomaly'] = iso_forest.fit_predict(data_encoded[['Engagement']])

data_no_outliers = data_encoded[data_encoded['anomaly'] != -1]
```

รูปที่ 2.17 ตัวอย่างการใช้ฟังก์ชัน Isolation forest ในการหา Outlier

บทที่ 3
การวิเคราะห์ข้อมูล

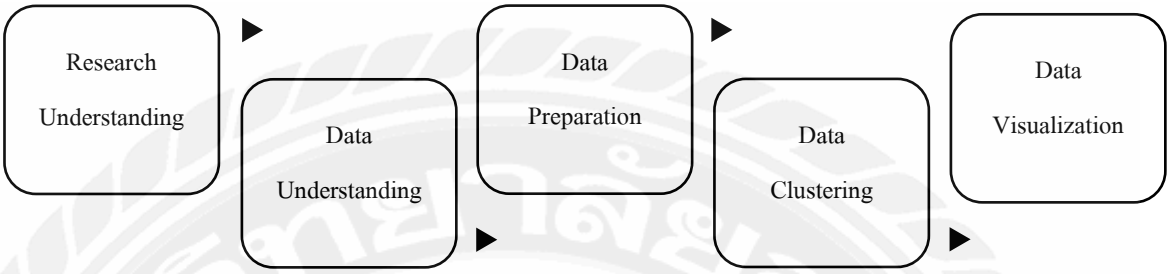
3.1 รายละเอียดของปริญญานิพนธ์

ในการศึกษาและวิเคราะห์หาปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย โดยข้อมูลตามโซเชียลมีเดียที่มีจำนวนมากถาโถมเข้ามาผ่านสายตามากมายในแต่ละวัน การเป็นหนึ่งที่ผู้ชมจะเลือกและหยุดดูเนื้อหาที่ถูกสร้างจากผู้คนธรรมดา หรือจากผู้มีชื่อเสียงจากแบรนด์เองหรือ Influencer ได้นั้นต้องเกิดจากหลากหลายปัจจัยที่ทำให้ผู้ชมสนใจ คณะผู้จัดทำจึงได้ทำการรวบรวมข้อมูลจาก Zocial eye ซึ่งเป็นเครื่องมือ Social listening ในการไปดึงข้อมูลตามแพลทฟอร์มของโซเชียลมีเดียต่างๆ เช่น Facebook, Instagram, Twitter (X), Youtube, Tiktok และ Forum ที่ได้เข้าไปติด Keyword ไว้ และนำมา Labeling เพื่อแยกว่าข้อความที่เกิดขึ้นจากผู้โพสต์ต่างๆ พูดถึงอะไร ใครเป็นคนโพสต์ โดยแบ่งออกเป็นประเภทเนื้อหา ได้แก่ สินค้า โปรโมชัน พรีเซนเตอร์ และแคมเปญหรือกิจกรรรม ต่อมาแบ่งออกเป็นประเภทผู้โพสต์ ได้แก่ บุคคลทั่วไป ดาราหรือผู้มีชื่อเสียง แบรนด์เป็นคนโพสต์เองและสำนักข่าวต่างๆ โดยทาง Zocial eye จะมีค่าของ Engagement หรือ การมีส่วนร่วมในโพสต์เป็นจำนวนการคลิกไลค์ แชร์และคอมเมนต์ มาเป็นตัวแบ่งกลุ่มความสนใจของผู้รับชมเนื้อหา

หลังจากที่ได้รวบรวมข้อมูลจากแหล่งต่างๆ รวมถึงการ Labeling ข้อมูลเพิ่มเติมเข้าไปแล้ว จะนำมาทำการตรวจสอบข้อมูลว่าพร้อมสำหรับนำไปวิเคราะห์ข้อมูลหรือไม่ โดยทำความสะอาดข้อมูล (Data cleansing) ซึ่งหลักๆ จะมีการหาค่าที่ขาดหายไป (Missing value), กำจัดข้อมูลที่ซ้ำซ้อน (Duplicated data) และกำจัดข้อมูลที่ไม่เกี่ยวข้อง (Irrelevant data) เพื่อให้การวิเคราะห์ข้อมูลได้ผลลัพธ์ที่มีความแม่นยำมากที่สุดในการจัดกลุ่มข้อมูล(Clustering) และหาปัจจัยที่เกี่ยวข้อง จากนั้นนำเสนอด้วยแผนภาพของข้อมูล (Data visualization) โดยใช้โปรแกรม Visual Studio Code และ Tableau ควบคู่กัน

3.2 ขั้นตอนการวิเคราะห์ข้อมูล

เนื่องจากมีการใช้ข้อมูลจริงไม่สามารถนำข้อมูลออกมาเผยแพร่ได้ จึงสามารถแสดงตัวอย่างงานได้บางส่วนเท่านั้น ขั้นตอนในการศึกษาและหาปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย มีดังนี้



รูปที่ 3.1 ขั้นตอนการศึกษาและหาปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย

3.2.1 กำหนดวัตถุประสงค์ในการวิเคราะห์ข้อมูล (Research Understanding)

ตั้งสมมติฐานการวิเคราะห์ที่มีผลต่อปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย ไว้ด้วยกัน 4 ปัจจัย ได้แก่

1) พรีเซนเตอร์หรือบุคคลที่มีชื่อเสียง
2) การใช้แพลทฟอร์มที่เหมาะสมกับเนื้อหา
3) แคมเปญหรือกิจกรรม
4) การใช้แฮชแท็ก(#)

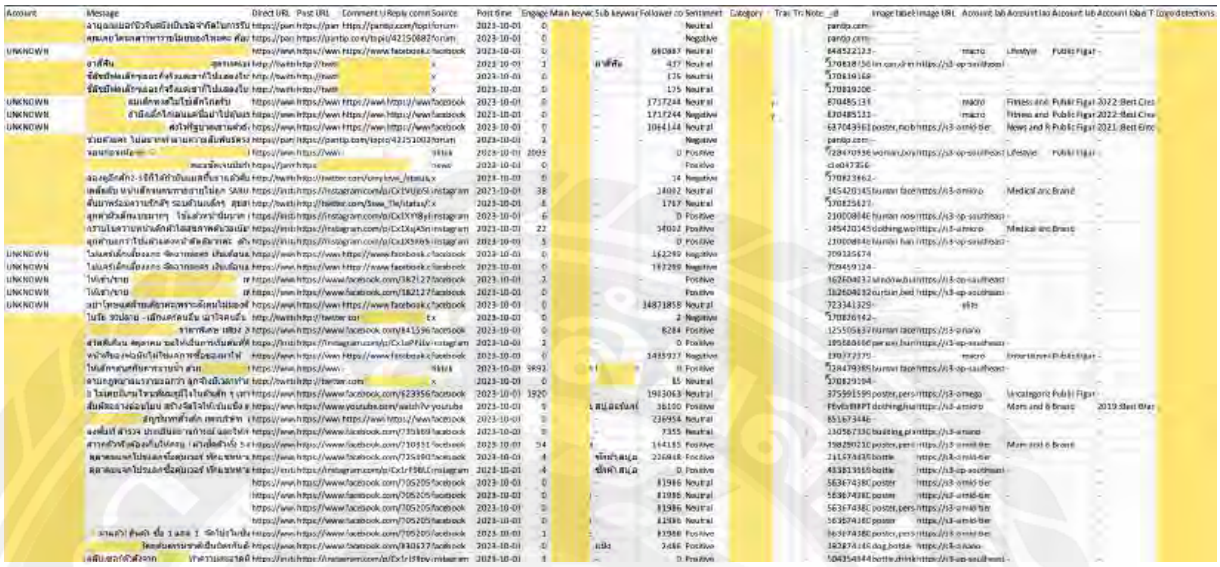
3.2.2 รวบรวมข้อมูล (Data Understanding)

ในการรวบรวมข้อมูลเพื่อนำมาใช้ในการวิเคราะห์ ทางคณะผู้ตัดทำได้ทำการรวบรวมจากเครื่องมือสำหรับ Social listening อย่าง Zocial eye โดยการเข้าไปติด Keyword เพื่อให้ได้ข้อมูลของผลิตภัณฑ์สำหรับเด็ก เช่น ชื่อแบรนด์สินค้าสำหรับเด็กและอุปกรณ์ต่างๆ เช่น สบู่ แปรงสีฟัน ยาสีฟัน แชมพู แป้ง อาบน้ำ ซักผ้า สระผม เป็นต้น และนำข้อมูลตั้งแต่วันที่ 1 ตุลาคม 2567 – 31 มีนาคม 2567 ออกมาเป็นไฟล์ Excel

3.2.3 เตรียมข้อมูล (Data Preparation)

เมื่อได้ข้อมูลมาครบแล้ว ทางคณะผู้จัดทำได้ทำการทำความสะอาดข้อมูล โดยปกติแล้วข้อมูลที่ได้จาก Zocial eye มักจะมีเนื้อหาที่ไม่ได้เกี่ยวข้องกับการนำมาวิเคราะห์ข้อมูล เช่น ข้อมูลที่มีคำพ้องรูป พ้องเสียงกับ Keyword ที่นำไปติด แต่ไม่ได้เกี่ยวกับเนื้อสินค้าผลิตภัณฑ์เด็ก

คณะผู้จัดทำจึงต้องใช้เทคนิคการ Labeling เข้ามาช่วย โดยใช้โปรแกรม Visual Studio Code โดยมีรายละเอียดในการเตรียมข้อมูลดังนี้



รูปที่ 3.2 ข้อมูลที่นำออกมาจาก Zocial eye โดยมีข้อมูลทั้งหมด 61,508 แถว 24 คอลัมน์ โดยเป็นข้อมูลตั้งแต่วันที่ 1 ตุลาคม 2566 – 31 มีนาคม 2567

จากรูปที่ 3.2 เป็นข้อมูลจาก Zocial eye โดยข้อมูลที่นำไปวิเคราะห์จะมีคอลัมน์ Message, และ Engagement ส่วนคอลัมน์อื่นๆ จะเพิ่มมาอีกทีตอนทำ Labeling ซึ่งตอนนี้จะเห็นว่าคอลัมน์ Message มีค่า Missing Value ค่าตรงนั้นหากคลิกเข้าไปดูต้นทางของลิงก์ในคอลัมน์ Direct URL จะพบว่าเป็นรูปภาพที่เกิดจากการโพสต์โดยไม่ได้มีแคปชันใดๆ ไว้ ซึ่งปัญหานี้จะถูกการ Labeling ในขั้นตอนถัดไป

รูปที่ 3.3 แสดงส่วนหนึ่งของการ Labeling แบรนด์ผลิตภัณฑ์เด็ก

จากรูปที่ 3.3 แสดงการ Labeling คอลัมน์ Message โดยใช้ฟังก์ชัน str.contains เพื่อแยกออกมาว่า Message นี้มีการพูดถึงแบรนด์อะไร โดยการสร้างคอลัมน์ใหม่และตั้งชื่อคอลัมน์ใหม่ด้วยชื่อแบรนด์สินค้า ภายใต้เงื่อนไขหลายประการเพื่อตรวจสอบการกล่าวถึงของแบรนด์ หากเงื่อนไขเป็นจริงจะให้ค่าเป็น 1 (ใช่) หากไม่เป็นจริงจะให้ค่าเป็น 0 (ไม่ใช่)

เงื่อนไขหลัก

- ค้นหาคำในคอลัมน์ Message ด้วยชื่อแบรนด์ที่เป็นภาษาไทยหรืออังกฤษ
- ค้นหาคำในคอลัมน์ Category ด้วยชื่อแบรนด์ที่เป็นภาษาไทยหรืออังกฤษ ซึ่งคอลัมน์นี้ Zocial eye ได้ทำการ Labeling มาให้ส่วนหนึ่งแล้ว แต่ความแม่นยำยังไม่มากพอ จึงต้องมีการตรวจสอบซ้ำ
- ค้นหาคำในคอลัมน์ Account ด้วยชื่อช่องทางโซเชียลมีเดียของแบรนด์หากเงื่อนไขใดเงื่อนไขหนึ่งเป็นจริงจะไปตรวจสอบขั้นต่อไปในเงื่อนไขรองต่อไปดังนี้

เงื่อนไขรอง

- ค้นหาคำในคอลัมน์ Message ด้วยชื่อผลิตภัณฑ์สินค้ารวมถึงสรรพคุณต่างๆ ที่สินค้านั้นจะมี หรืออาจจะเป็นค่าว่างเพื่อให้ครอบคลุมค่าว่างในคอลัมน์ Message ที่เป็นรูปภาพ
- ค้นหาลิงก์เฉพาะในคอลัมน์ Post URL หากเงื่อนหลักและเงื่อนไขรองเป็นจริงแล้ว จะต้องตรวจสอบเงื่อนไขยกเว้นต่อไปดังนี้

เงื่อนไขยกเว้น

- หากเจอชื่อผู้ใช้งานที่กำหนดในคอลัมน์ Account จะทำให้เป็นโมฆะ
- หากเจอลิงก์ที่กำหนดในคอลัมน์ Direct URL จะทำให้เป็นโมฆะ
- หากเจอลิงก์ที่กำหนดในคอลัมน์ Post URL จะทำให้เป็นโมฆะ
- หากเจอคำที่กำหนดในคอลัมน์ Message จะทำให้เป็นโมฆะทำกับแบรนด์ทั้งหมดของผลิตภัณฑ์สำหรับเด็กโดยเปลี่ยนคำที่ต้องการให้ค้นหาหรือยกเว้นในกรณีที่ไม่เกี่ยวข้อง

รูปที่ 3.4 เป็นการสร้างคอลัมน์ใหม่ที่ชื่อว่า Brand เพื่อเก็บชื่อของแบรนด์สินค้าที่ได้ทำการ Labeling มาก่อนหน้านี้

จากรูปที่ 3.4 แสดงการใช้ฟังก์ชัน for loop ในการกำหนดค่าให้กับคอลัมน์ Brand โดยกำหนดรายชื่อแบรนด์สินค้าไว้ที่ตัวแปร brand จากนั้นนำไปวนลูปโดยใช้ฟังก์ชัน iterrows() เพื่อวนลูปในแต่ละแถวและนับจำนวนแบรนด์โดยใช้ sum(row[brand]) เพื่อคำนวณแบรนด์ที่ปรากฏในแต่ละแถวโดยเช็คค่าของคอลัมน์ที่มีชื่อแบรนด์ทั้งหมด ที่ได้จากการทำ Labeling 0,1 ไว้ก่อนหน้านี้ภายใต้เงื่อนไขดังนี้

เงื่อนไข

- หากในแถวมีมากกว่า 1 แบรนด์ หรือบวกกันแล้วได้มากกว่า 1 จะกำหนดค่าในคอลัมน์ Brand เป็น Non-specific
- หากในแถวมีแค่ 1 แบรนด์ หรือบวกกันแล้วเท่ากับ 1 จะทำการวนลูปผ่านลิสต์ตัวแปร brand เพื่อตรวจสอบว่าแถวมีค่า 1 ที่แบรนด์ใด และกำหนดค่าในคอลัมน์ Brand เป็นชื่อแบรนด์นั้น
- หากในแถวไม่มีแบรนด์ใดเลย หรือ บวกกันแล้วได้เท่ากับ 0 จะกำหนดค่าในคอลัมน์ Brand เป็น – ซึ่งหมายความว่า คอนเทนต์นี้ไม่เกี่ยวข้องผลิตภัณฑ์สำหรับเด็ก

รูปที่ 3.5 แสดงส่วนหนึ่งของการ Labeling คอนเทนต์ที่เกี่ยวกับ Activity

จากรูปที่ 3.5 แสดงการ Labeling คอนเทนต์ที่เกี่ยวกับกิจกรรมที่ทางแบรนด์ เพจต่างๆ ดาราหรือผู้มีชื่อเสียง แม้กระทั่งบุคคลทั่วไป จัดกิจกรรมแจกของรางวัลต่างๆ คณะผู้จัดทำจึงได้ทำการ Labeling คอนเทนต์นี้เพื่อนำไปวิเคราะห์ความสนใจของผู้รับชมคอนเทนต์ โดยขั้นตอนการทำจะคล้ายๆ กับการ Labeling ชื่อแบรนด์ ซึ่งคำที่เลือกมาจับคู่จะมีมากกว่านี้แต่นี่เป็นเพียงตัวอย่างเพื่อให้เห็นภาพรวมของคำที่นำไปค้นหา หากเงื่อนไขเป็นจริงจะให้ค่าเป็น 1 (ใช่) หากไม่เป็นจริงจะให้ค่าเป็น 0 (ไม่ใช่)

เงื่อนไขหลัก

- ในคอลัมน์ Brand จะต้องไม่เป็น – หรือไม่มีชื่อแบรนด์ปรากฏหรือเป็นค่า Non-specific

เงื่อนไขรอง

- ค้นหาคำในคอลัมน์ Message ด้วยคำที่มักจะปรากฏในโพสต์ของแคมเปญหรือกิจกรรม
- ค้นหาลิงก์เฉพาะในคอลัมน์ Post URL หากเงื่อนหลักและเงื่อนไขรองเป็นจริงแล้ว จะต้องตรวจสอบเงื่อนไขยกเว้นต่อไปดังนี้

เงื่อนไขยกเว้น

- หากเจอลิงก์ที่กำหนดในคอลัมน์ Direct URL จะทำให้เป็นโมฆะ
- หากเจอชื่อผู้ใช้งานที่กำหนดในคอลัมน์ Account จะทำให้เป็นโมฆะ
- หากเจอคำที่กำหนดในคอลัมน์ Message จะทำให้เป็นโมฆะ

รูปที่ 3.6 แสดงส่วนหนึ่งของการ Labeling คอนเทนต์ที่เกี่ยวกับ Presenter

จากรูปที่ 3.6 แสดงการ Labeling คอนเทนต์ที่เกี่ยวกับพรีเซนเตอร์ของทางแบรนด์ ไม่ว่าจะเป็นการกล่าวถึง กล่าวชม รวมถึงการรีวิวสินค้าโดยพรีเซนเตอร์ ซึ่งขั้นตอนการทำจะคล้ายกับการ Labeling คอนเทนต์แคมเปญหรือกิจกรรม หากเงื่อนไขเป็นจริงจะให้ค่าเป็น 1 (ใช่) หากไม่เป็นจริงจะให้ค่าเป็น 0 (ไม่ใช่)

เงื่อนไขหลัก

- ในคอลัมน์ Brand จะต้องไม่เป็น – หรือไม่มีชื่อแบรนด์ปรากฏ หรือเป็นค่า Non-specific

เงื่อนไขรอง

- ค้นหาคำในคอลัมน์ Message ด้วยชื่อพรีเซนเตอร์
- ค้นหาชื่อผู้ใช้งานของพรีเซนเตอร์ในคอลัมน์ Account ทุกแพลตฟอร์มหากเงื่อนหลักและเงื่อนไขรองเป็นจริงแล้ว จะต้องตรวจสอบเงื่อนไขยกเว้นต่อไปดังนี้

เงื่อนไขยกเว้น

- หากเจอชื่อผู้ใช้งานที่กำหนดในคอลัมน์ Account จะทำให้เป็นโมฆะ
- หากเจอคำที่กำหนดในคอลัมน์ Message จะทำให้เป็นโมฆะ

รูปที่ 3.7 แสดงส่วนหนึ่งของการ Labeling คอนเทนต์ที่เกี่ยวกับ Promotion

จากรูปที่ 3.7 แสดงการ Labeling คอนเทนต์ที่เกี่ยวกับโปรโมชันสินค้า โดยขั้นตอนการทำจะคล้ายๆ กับการ Labeling คอนเทนต์พรีเซนเตอร์ ซึ่งคำที่เลือกมาจับคู่จะมีมากกว่านี้แต่นี่เป็นเพียงตัวอย่างเพื่อให้เห็นภาพรวมของคำที่นำไปค้นหา หากเงื่อนไขเป็นจริงจะให้ค่าเป็น 1 (ใช่) หากไม่เป็นจริงจะให้ค่าเป็น 0 (ไม่ใช่)

เงื่อนไขหลัก

- ในคอลัมน์ Brand จะต้องไม่เป็น – หรือไม่มีชื่อแบรนด์ปรากฏ หรือเป็นค่า Non-specific

เงื่อนไขรอง

- ค้นหาคำในคอลัมน์ Message ด้วยคำที่มักจะปรากฏอยู่ในโพสต์โปรโมชั่น
- ค้นหาลิงก์เฉพาะในคอลัมน์ Post URL
- ค้นหาชื่อผู้ใช้งานของพรีเซนเตอร์ในคอลัมน์ Account ทุกแพลตฟอร์มหากเงื่อนหลักและเงื่อนไขรองเป็นจริงแล้ว จะต้องตรวจสอบเงื่อนไขยกเว้นต่อไปดังนี้

เงื่อนไขยกเว้น

- หากเจอคำที่กำหนดในคอลัมน์ Message จะทำให้เป็นโมฆะ

รูปที่ 3.8 แสดงส่วนหนึ่งของการ Labeling คอนเทนต์ที่เกี่ยวกับ Product

จากรูปที่ 3.8 เป็นการ Labeling คอนเทนต์ที่เกี่ยวกับการเปิดตัวผลิตภัณฑ์ใหม่ หรือมีการพูดถึงผลิตภัณฑ์ในด้านต่างๆ โดยขั้นตอนการทำจะคล้ายกับการ Labeling คอนเทนต์พรีเซนเตอร์ ซึ่งคำที่เลือกมาจับคู่จะมีมากกว่านี้แต่นี่เป็นเพียงตัวอย่างเพื่อให้เห็นภาพรวมของคำที่นำไปค้นหา หากเงื่อนไขเป็นจริงจะให้ค่าเป็น 1 (ใช่) หากไม่เป็นจริงจะให้ค่าเป็น 0 (ไม่ใช่)

เงื่อนไขหลัก

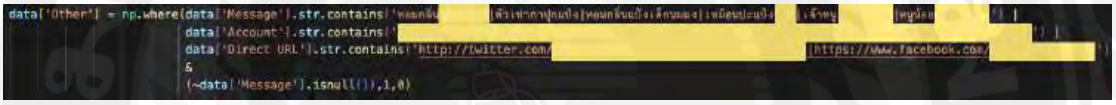
- ในคอลัมน์ Brand จะต้องไม่เป็น – หรือไม่มีชื่อแบรนด์ปรากฏ หรือเป็นค่า Non-specific

เงื่อนไขรอง

- ค้นหาคำที่กำหนดในคอลัมน์ Message
- ค้นหาชื่อผู้ใช้งานที่กำหนดในคอลัมน์ Account ทุกแพลตฟอร์ม
- ค้นหาลิงก์เฉพาะในคอลัมน์ Direct URL
- ค้นหาลิงก์เฉพาะในคอลัมน์ Post URL หากเงื่อนหลักและเงื่อนไขรองเป็นจริงแล้วจะต้องตรวจสอบเงื่อนไขยกเว้นต่อไป ดังนี้

เงื่อนไขยกเว้น

- หากเจอชื่อผู้ใช้งานที่กำหนดในคอลัมน์ Account จะทำให้เป็นโมฆะ
- หากเจอคำที่กำหนดในคอลัมน์ Message จะทำให้เป็นโมฆะ
- หากเจอลิงก์ที่กำหนดในคอลัมน์ Post URL จะทำให้เป็นโมฆะ



รูปที่ 3.9 แสดงส่วนหนึ่งของการ Labeling คอนเทนต์ที่เกี่ยวกับ Other

จากรูปที่ 3.9 แสดงการ Labeling คอนเทนต์ที่ไม่ได้เกี่ยวกับผลิตภัณฑ์หรือแบรนด์สินค้าโดยตรง แต่เป็นการนำชื่อแบรนด์ไปใช้ในจุดประสงค์อื่น เช่น การชมศิลปินที่ชื่นชอบ โดยแทนชื่อแบรนด์เป็นคำว่า น่ารัก หอม ฯลฯ โดยขั้นตอนการทำคล้ายกับการ Labeling คอนเทนต์อื่นๆ ซึ่งคำที่เลือกมาจับคู่จะมีมากกว่านี้แต่นี่เป็นเพียงตัวอย่างเพื่อให้เห็นภาพรวมของคำที่นำไปค้นหา หากเงื่อนไขเป็นจริงจะให้ค่าเป็น 1 (ใช่) หากไม่เป็นจริงจะให้ค่าเป็น 0 (ไม่ใช่)

เงื่อนไขหลัก

- ค้นหาชื่อผู้ใช้งานที่กำหนดในคอลัมน์ Account ทุกแพลตฟอร์ม
- ค้นหาคำที่กำหนดในคอลัมน์ Message
- ค้นหาลิงก์เฉพาะในคอลัมน์ Direct URL

เงื่อนไขยกเว้น

- คอลัมน์ Message จะต้องไม่เป็นค่าว่าง

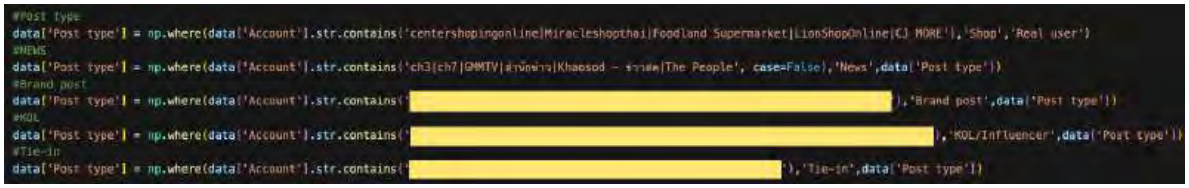
```
Topic = ['Activity', 'Promotion', 'Presenter', 'Other','Product']

for index, row in data.iterrows():
    Topic_count = sum(row[Topic])
    if Topic_count > 1:
        for topic in Topic:
            if row[topic] == 1:
                data.at[index, 'TOPIC'] = topic
                break
    elif Topic_count == 1:
        for topic in Topic:
            if row[topic] == 1:
                data.at[index, 'TOPIC'] = topic
                break
    else:
        data.at[index, 'TOPIC'] = '-'
```

รูปที่ 3.10 เป็นการสร้างคอลัมน์ใหม่ที่ชื่อว่า “TOPIC” เพื่อเก็บค่าของหมวดหมู่คอนเทนต์ที่ได้ทำการ Labeling มาก่อนหน้านี้

จากรูปที่ 3.10 แสดงการใช้ฟังก์ชัน for loop ในการกำหนดค่าให้กับคอลัมน์ TOPIC โดยกำหนดรายชื่อหมวดหมู่คอนเทนต์ไว้ที่ตัวแปร Topic จากนั้นนำไปวนลูปโดยใช้ฟังก์ชัน iterrows() เพื่อวนลูปในแต่ละแถวและนับจำนวนหมวดหมู่คอนเทนต์โดยใช้ sum(row[brand]) เพื่อคำนวณหมวดหมู่คอนเทนต์ที่ปรากฏในแต่ละแถวโดยเช็คค่าของคอลัมน์ที่มีชื่อหมวดหมู่คอนเทนต์ทั้งหมดที่ได้จากการทำ Labeling 0, 1 ไว้ก่อนหน้านี้ ภายใต้เงื่อนไขดังนี้

เงื่อนไข

- หากผลรวมมีหมวดหมู่มากกว่า 1 จะทำการวนลูปผ่านลิสต์ Topic และเมื่อเจอหมวดหมู่ที่มีค่าเป็น 1 คอลัมน์แรกที่เจอจะถูกกำหนดให้กับคอลัมน์ TOPIC สำหรับแถวนั้นก่อน และเมื่อเจอแล้วจะไม่ตรวจสอบคอลัมน์ที่เหลือ
- หากผลรวมมีหมวดหมู่เท่ากับ 1 จะหาหมวดหมู่ที่มีค่าเป็น 1 แล้วกำหนดค่าในคอลัมน์ TOPIC สำหรับแถวนั้น
- หากผลรวมหมวดหมู่เท่ากับ 0 จะถูกกำหนดให้กับคอลัมน์ TOPIC มีค่าเป็น – หรือไม่พบหมวดหมู่



รูปที่ 3.11 เป็นการสร้างคอลัมน์ใหม่ที่ชื่อว่า “Post type” เพื่อทำการ Labeling ประเภทของผู้โพสต์บนโลกโซเชียลออนไลน์

จากรูปที่ 3.11 แสดงการ Labeling คอลัมน์ Account โดยใช้ฟังก์ชัน str.contains เพื่อแยกออกมาว่าผู้ใช้นี้ โพสต์โดยใคร เช่น ร้านค้า สำนักข่าว แบรนด์ ดาราหรือคนดัง การ Tie in จากรายการ หรือบุคคลทั่วไปเป็นคนโพสต์ โดยใช้เงื่อนไขการตรวจสอบค่าของคอลัมน์ Account และกำหนดค่าของ Post type ตามผลลัพธ์ที่ได้

ร้านค้า

- ตรวจสอบว่าคอลัมน์ Account มีค่าตรงกับชื่อผู้ใช้ที่กำหนดไว้หรือไม่
- หากเข้าเงื่อนไข ให้กำหนด Post type เป็น Shop
- หากไม่ตรงกับเงื่อนไขใดๆ ให้กำหนด Post type เป็น Real user

สำนักข่าว

- ตรวจสอบว่าคอลัมน์ Account มีค่าตรงกับชื่อผู้ใช้ที่กำหนดไว้หรือไม่
- หากเข้าเงื่อนไข ให้กำหนด Post type เป็น News
- หากไม่ตรงกับเงื่อนไขใดๆ ให้คงค่า Post type ตามที่มีอยู่ในคอลัมน์ Post type

ดาราหรือคนดัง

- ตรวจสอบว่าคอลัมน์ Account มีค่าตรงกับชื่อผู้ใช้ที่กำหนดไว้หรือไม่
- หากเข้าเงื่อนไข ให้กำหนด Post type เป็น KOL/ Influencer
- หากไม่ตรงกับเงื่อนไขใดๆ ให้คงค่า Post type ตามที่มีอยู่ในคอลัมน์ Post type

การ Tie in จากรายการ

- ตรวจสอบว่าคอลัมน์ Account มีค่าตรงกับชื่อผู้ใช้ที่กำหนดไว้หรือไม่
- หากเข้าเงื่อนไข ให้กำหนด Post type เป็น Tie-in
- หากไม่ตรงกับเงื่อนไขใดๆ ให้คงค่า Post type ตามที่มีอยู่ในคอลัมน์ Post type

						Brand	Activity	Presenter	Promotion	Other	Product	TOPIC	Post type
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	1	Product	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	1	[redacted]	0	0	1	0	1	Promotion	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	1	Product	Real user
0	0	0	0	0	0	-	0	0	0	0	1	Product	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	News
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	1	Product	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	1	Product	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	1	0	1	Promotion	Real user
0	0	0	0	0	0	-	0	0	1	0	1	Promotion	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	1	0	1	Promotion	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	[redacted]	0	0	0	0	1	Product	Real user
0	0	0	0	0	0	-	0	0	0	0	0	-	Real user
0	0	0	0	0	0	-	0	0	1	0	1	Promotion	Real user
0	0	0	0	0	1	[redacted]	0	0	0	0	1	Product	Brand post
0	0	0	0	0	0	-	0	0	0	0	0	-	KOL/Influe
0	0	0	0	0	0	-	0	0	0	0	1	Product	Real user
0	0	0	0	0	0	[redacted]	0	0	1	0	1	Promotion	Brand post
0	0	0	0	0	0	Non-specifi	0	0	1	0	1	Promotion	Real user
0	0	0	0	0	0	Non-specifi	0	0	1	0	1	Promotion	Real user
0	0	0	0	0	0	[redacted]	0	0	0	0	0	-	Shop
0	0	0	0	0	0	[redacted]	0	0	0	0	0	-	Shop
0	0	0	0	0	0	[redacted]	0	0	0	0	0	-	Shop
0	0	0	0	0	0	[redacted]	0	0	0	0	0	-	Shop
0	0	0	0	0	0	[redacted]	0	0	1	0	1	Promotion	Shop
0	0	0	0	0	0	[redacted]	0	0	1	1	1	Promotion	Real user
0	0	0	0	0	0	-	0	0	1	0	1	Promotion	Real user
0	0	0	0	0	0	[redacted]	0	0	1	1	1	Promotion	Real user

รูปที่ 3.12 แสดงคอลัมน์ที่ Labeling เรียบร้อย

จากรูปที่ 3.12 แสดงการ Labeling ที่เสร็จแล้วโดยมีคอลัมน์ Brand ที่ภายในมีชื่อแบรนด์กำหนดอยู่ในแต่ละแถว มีคอลัมน์ TOPIC ที่บ่งบอกว่าแถวนี้พูดเกี่ยวกับประเด็นไหน และยังมี Post type ที่แสดงให้เห็นว่าแต่ละแถวข้อความนี้ใครเป็นคนโพสต์

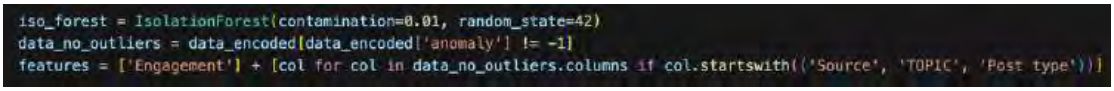
3.2.3 จัดกลุ่มข้อมูล (Data Clustering)

นำข้อมูลที่ได้จากการเตรียมไว้มาทำการจัดกลุ่มข้อมูล (Data Clustering) โดยใช้โปรแกรม Visual Studio Code ในการแบ่งกลุ่มตามจำนวน Engagement หรือ การมีปฏิสัมพันธ์กับโพสต์ เพื่อดูว่าโพสต์แบบไหนที่ได้รับความสนใจจากผู้รับชมคอนเทนต์มากที่สุด และมีปัจจัยใดที่กระตุ้นความสนใจของผู้รับชมคอนเทนต์บ้าง

```
brands_to_include = ['[redacted]', '[redacted]', '[redacted]', '[redacted]', '[redacted]']
data_filtered = data[(data['Engagement'] > 0) & (data['TOPIC'] != '-') & (data['Brand'].isin(brands_to_include))]


data_encoded = pd.get_dummies(data_filtered, columns=['Source', 'TOPIC', 'Post type'], prefix=['Source', 'TOPIC','Post type'])

boolean_columns = data_encoded.select_dtypes(include=['bool']).columns

data_encoded[boolean_columns] = data_encoded[boolean_columns].astype(int)

iso_forest = IsolationForest(contamination=0.01, random_state=42)
data_encoded['anomaly'] = iso_forest.fit_predict(data_encoded[['Engagement']])
```

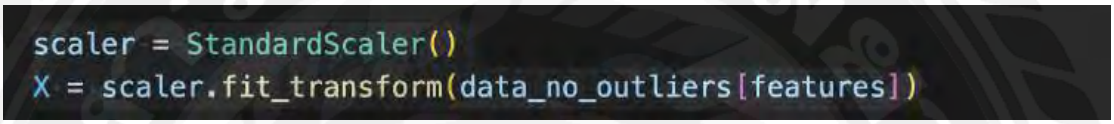
รูปที่ 3.13 แสดงการกรองข้อมูลและทำการ One-Hot Encoding สำหรับการจัดกลุ่มข้อมูลที่เป็น Category ที่ไม่ได้เป็นข้อมูลตัวเลข ใช้ในการวิเคราะห์การจัดกลุ่มทุกหัวข้อ

จากรูปที่ 3.13 แสดงการจัดกลุ่ม Engagement สำหรับหัวข้อคอนเทนต์ทั้งหมด (Activity, Presenter, Promotion, Production และ Other) และได้มีการเลือกแบรนด์ผลิตภัณฑ์เด็ก 5 แบรนด์ที่มีสินค้าใกล้เคียงกันและเป็นที่นิยมในตลาด โดยกำหนดชื่อแบรนด์ที่ต้องการในลิสต์ของตัวแปร brand_to_include โดยเลือกเฉพาะที่มีค่า Engagement มากกว่า 0 และค่าในคอลัมน์ Brand ตรงกับแบรนด์ที่อยู่ใน brands_to_include และให้เก็บผลลัพธ์ไว้ที่ตัวแปร data_filtered หลังจากนั้นใช้ฟังก์ชัน pd.get_dummies() เพื่อแปลงข้อมูลที่ Category (Source, TOPIC, และ Post type) เป็นค่าตัวเลขแบบ One-Hot Encoding (0,1) ในที่นี้ข้อมูลจะเป็น Boolean ซึ่งจะทำการแปลงเป็น (0,1) ในภายหลัง และคอลัมน์ใหม่จะมี prefix เป็น Source, TOPIC และ Post type ตามลำดับ ผลลัพธ์จะถูกเก็บในตัวแปร data_encoded จากนั้นทำการแปลงข้อมูล Boolean ที่ได้จากการทำ One-Hot Encoding เป็นข้อมูลตัวเลข (0,1) โดยเลือกคอลัมน์ที่มีประเภทข้อมูลเป็น bool (Boolean) แปลงคอลัมน์เหล่านี้เป็นประเภทข้อมูล int (Integer) เพื่อให้เข้ากันได้กับการวิเคราะห์ข้อมูลในขั้นตอนต่อไป



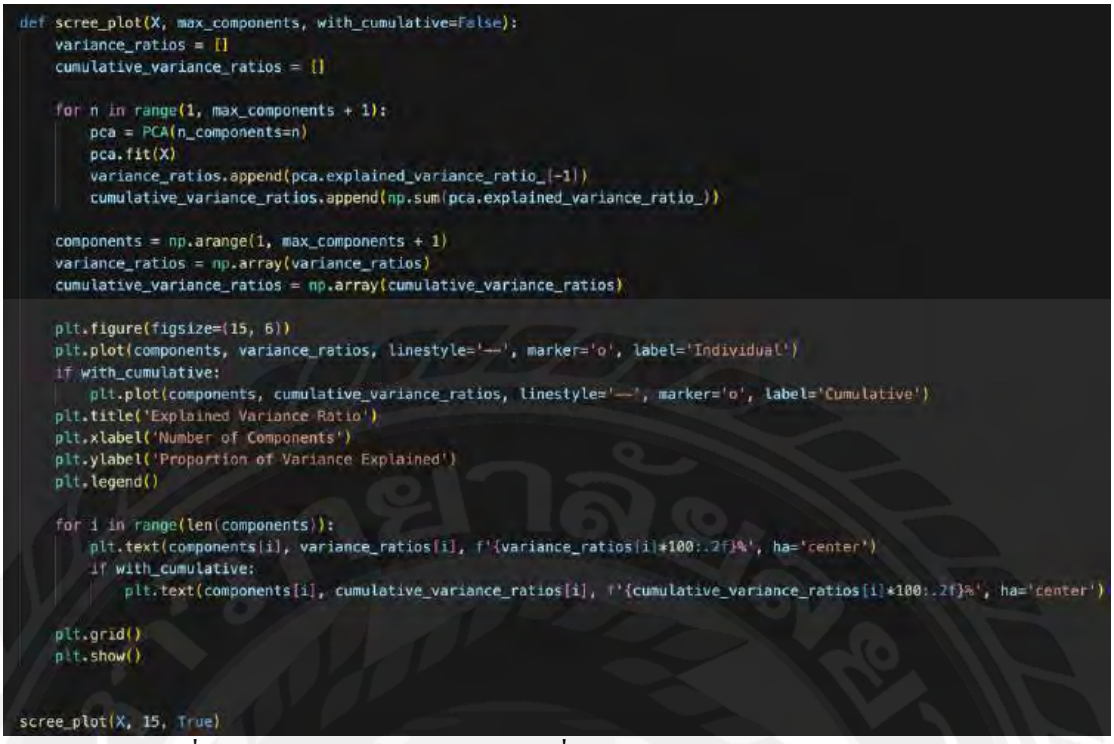
รูปที่ 3.14 แสดงการตรวจจับ Outlier ด้วย Isolation Forest

จากรูปที่ 3.14 แสดงการตรวจจับ Outlier ด้วย Isolation Forest โดยใช้โมเดล Isolation Forest เพื่อตรวจจับค่า Outliers ในข้อมูล ซึ่งตั้งค่าพารามิเตอร์ contamination เป็น 0.01 หมายถึงจะระบุให้ 1% ของข้อมูลเป็น Outliers และคอลัมน์ anomaly จะถูกเพิ่มเข้าไปใน data_encoded โดยมีค่า 1 หมายถึงข้อมูลปกติ และ -1 หมายถึง Outliers หลังจากนั้นกรองข้อมูลโดยลบแถวที่มีค่า anomaly เป็น -1 ออก ผลลัพธ์จะถูกเก็บในตัวแปร data_no_outliers เพื่อนำไปวิเคราะห์เฉพาะค่าปกติ จากนั้นจะทำการเลือกเฉพาะคอลัมน์ที่จะนำไปจัดกลุ่มโดย สร้างลิสต์ features โดยรวมคอลัมน์ Engagement และคอลัมน์ทั้งหมดที่มี prefix เป็น Source, TOPIC, และ Post type



รูปที่ 3.15 แสดงการปรับขนาดของข้อมูลให้มีขนาดเดียวกัน

จากรูปที่ 3.15 แสดงการปรับขนาดของข้อมูลทำให้ข้อมูลมีมาตรฐานเดียวกันช่วยให้การเปรียบเทียบฟีเจอร์ต่างๆ แม่นยำขึ้นโดยสร้างอ็อบเจ็กต์ StandardScaler จากไลบรารี sklearn.preprocessing เพื่อใช้สำหรับปรับขนาดข้อมูล และใช้ fit_transform จากอ็อบเจ็กต์ scaler เพื่อปรับขนาดข้อมูลใน DataFrame data_no_outliers ที่มีคอลัมน์ตามที่กำหนดในตัวแปร features จากนั้นใช้ฟังก์ชัน fit_transform ทำการปรับขนาดข้อมูลโดยการคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของข้อมูล แล้วนำมาปรับข้อมูลให้มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ผลลัพธ์ที่ได้คือ DataFrame ที่มีข้อมูลถูกปรับขนาดแล้ว เก็บในตัวแปร X



```python
def scree_plot(X, max_components, with_cumulative=False):
    variance_ratios = []
    cumulative_variance_ratios = []

    for n in range(1, max_components + 1):
        pca = PCA(n_components=n)
        pca.fit(X)
        variance_ratios.append(pca.explained_variance_ratio_[-1])
        cumulative_variance_ratios.append(np.sum(pca.explained_variance_ratio_))

    components = np.arange(1, max_components + 1)
    variance_ratios = np.array(variance_ratios)
    cumulative_variance_ratios = np.array(cumulative_variance_ratios)

    plt.figure(figsize=(15, 6))
    plt.plot(components, variance_ratios, linestyle='--', marker='o', label='Individual')
    if with_cumulative:
        plt.plot(components, cumulative_variance_ratios, linestyle='--', marker='o', label='Cumulative')
    plt.title('Explained Variance Ratio')
    plt.xlabel('Number of Components')
    plt.ylabel('Proportion of Variance Explained')
    plt.legend()

    for i in range(len(components)):
        plt.text(components[i], variance_ratios[i], f'{variance_ratios[i]*100:.2f}%', ha='center')
        if with_cumulative:
            plt.text(components[i], cumulative_variance_ratios[i], f'{cumulative_variance_ratios[i]*100:.2f}%', ha='center')

    plt.grid()
    plt.show()

scree_plot(X, 15, True)
```

รูปที่ 3.16 แสดงฟังก์ชัน Scree plot เพื่อหาว่าควรใช้ Components เท่าไหร่ในการลดมิติของ Features

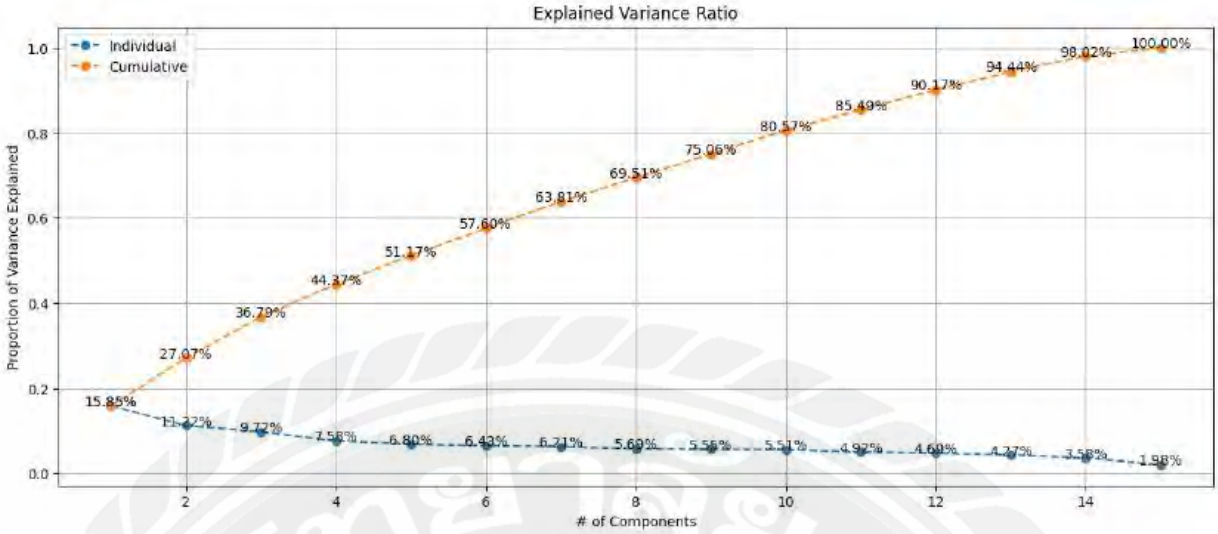
จากรูปที่ 3.16 แสดงฟังก์ชันการใช้งาน Scree Plot เพื่อดูว่าควรใช้ Components เท่าไหร่ในการลดมิติของ Features ลง โดยกำหนดฟังก์ชัน scree_plot รับพารามิเตอร์ 3 ตัว

- X: ข้อมูลที่ต้องการวิเคราะห์ด้วย PCA
- max_components: จำนวนองค์ประกอบสูงสุดที่ต้องการพิจารณา
- with_cumulative: ตัวเลือกในการแสดงกราฟความแปรปรวนสะสม (ค่าเริ่มต้นเป็น False)

ต่อไปทำการสร้างลิสต์เปล่า เพื่อเก็บอัตราส่วนความแปรปรวนไว้ที่ลิสต์ตัวแปร variance_ratios และเก็บความแปรปรวนสะสมไว้ที่ลิสต์ตัวแปร cumulative_variance_ratios จากนั้นใช้ลูปเพื่อคำนวณ PCA สำหรับจำนวนองค์ประกอบตั้งแต่ 1 ถึง max_components และสร้างโมเดล PCA ด้วย n_components=n โดยสร้างออบเจ็กต์ PCA โดยกำหนดจำนวนองค์ประกอบหลัก (Principal Components) ที่ต้องการ (n_components=n) จากนั้นฝึกโมเดล PCA บนข้อมูล X ด้วยคำสั่ง pca.fit(X) จากนั้นเก็บค่าอัตราส่วนความแปรปรวนของแต่ละองค์ประกอบใน variance_ratios และเก็บค่าอัตราส่วนความแปรปรวนสะสมใน cumulative_variance_ratios ต่อจากนั้นทำการแปลงลิสต์ components, variance_ratios, และ cumulative_variance_ratios เป็น numpy array เพื่อความสะดวกในการใช้งาน ต่อมาจะทำการ plot กราฟให้ออกมาเป็นรูปภาพ

- สร้างกราฟขนาด 15x6 นิ้ว ด้วย plt.figure(figsize=(15, 6))
- plot เส้นสำหรับอัตราส่วนความแปรปรวน (variance_ratios) โดยให้ค่าบนแกน X เป็นจำนวน components (1 ถึง max_components) และให้ค่าบนแกน Y เป็นอัตราส่วนความแปรปรวน (variance_ratios)
- กำหนดรูปแบบของเส้นด้วยเส้นประ linestyle= -- และ จุดมาร์คเป็นวงกลม marker = o
- ทำการกำหนดป้ายชื่อให้กับเส้นด้วย Label = Individual
- ตรวจสอบว่าตัวแปร with_cumulative เป็น True หรือไม่ ถ้าเงื่อนไขเป็น True จะ plot กราฟเส้นสำหรับความแปรปรวนสะสม(cumulative_variance_ratios)
- การตั้งค่าพารามิเตอร์เหมือนกับการ plot กราฟเส้นของ variance_ratios โดยมีการใช้ label='Cumulative' เพื่อบอกว่ากราฟเส้นนี้แสดงความแปรปรวนสะสม
- plt.title('Explained Variance Ratio') ตั้งชื่อกราฟว่า Explained Variance Ratio
- plt.xlabel('Number of Components') ตั้งชื่อแกน X ว่า จำนวนขององค์ประกอบ (Number of Components)
- plt.ylabel('Proportion of Variance Explained') ตั้งชื่อแกน Y ว่า สัดส่วนของความแปรปรวนที่อธิบายได้ (Proportion of Variance Explained)
- plt.legend() ทำการแสดงป้ายชื่อ

จากนั้นทำการวนลูปผ่านทุก component ที่มีในกราฟ เพื่อเพิ่มข้อความที่แสดงอัตราส่วนความแปรปรวน (เป็นเปอร์เซ็นต์) บนกราฟ หากเลือกให้แสดงความแปรปรวนสะสม (with_cumulative เป็น True) จะเพิ่มข้อความแสดงความแปรปรวนสะสมบนกราฟด้วยและแสดงตารางบนกราฟพร้อมกับกราฟ สุดท้ายเรียกใช้ฟังก์ชัน scree_plot โดยส่งข้อมูลที่ถูกปรับขนาด (X), จำนวน component สูงสุด 15 และเลือกให้แสดงความแปรปรวนสะสม(True)



รูปที่ 3.17 แสดงกราฟ Scree plot

จากรูปที่ 3.17 แสดงกราฟ Scree plot โดยแกน X แสดงจำนวนขององค์ประกอบ (Principal Components) ซึ่งมีตั้งแต่ 1 ถึง 15 Component และแกน Y แสดงอัตราส่วนของความแปรปรวนที่อธิบายได้ (Proportion of Variance Explained) ซึ่งเป็นเปอร์เซ็นต์ มีเส้นสีน้ำเงินแสดงอัตราส่วนของความแปรปรวนที่แต่ละองค์ประกอบอธิบาย และเส้นสีส้มแสดงอัตราส่วนของความแปรปรวนสะสมที่อธิบายได้โดยรวมจากหลาย Component จากกราฟนี้ Component ที่ 10-12 จะอธิบายความแปรปรวนสะสมได้ประมาณ 80-90% ซึ่งคณะผู้จัดทำจึงเลือกใช้ PCA ที่จำนวน 9 Component

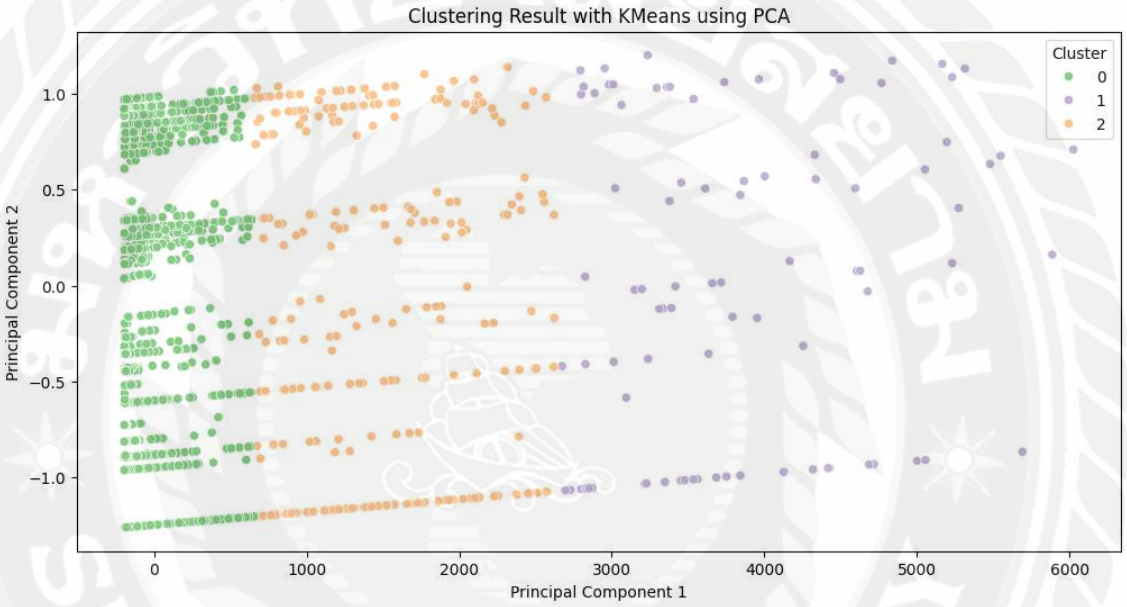
```
pca = PCA(n_components=10)
principal_components = pca.fit_transform(data_no_outliers[features])


data_no_outliers_pca = pd.DataFrame(principal_components, columns=[f'PC{i+1}' for i in range(10)])


kmeans = KMeans(n_clusters=3, random_state=42)
data_no_outliers_pca['cluster'] = kmeans.fit_predict(data_no_outliers_pca)



plt.figure(figsize=(12, 6))
sns.scatterplot(x='PC1', y='PC2', hue='cluster', data=data_no_outliers_pca, palette='Accent')
plt.title('Clustering Result with KMeans using PCA')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.legend(title='Cluster')
plt.show()
```

รูปที่ 3.18 แสดงฟังก์ชันการทำงานของ Clustering

จากรูปที่ 3.18 แสดงฟังก์ชันการหา Clustering โดยกำหนดจำนวน Components ซึ่งในขั้นตอนนี้ได้กำหนด PCA ลดมิติของข้อมูลเหลือ 10 มิติหลัก จาก 15 มิติ จากนั้นทำการลดมิติด้วย PCA โดยใช้ฟังก์ชัน fit_transform แล้วเก็บผลลัพธ์ไว้ที่ตัวแปร principal_components จากนั้นสร้าง

DataFrame ใหม่ที่ชื่อ data_no_outliers_pca จาก principal_components ด้วย pd.DataFrame จากนั้นใช้ list comprehension เพื่อสร้างคอลัมน์ที่มีชื่อว่า 'PC1', 'PC2', ..., 'PC10' และ f'PC{i+1}' เป็นการใช้ format สตริงเพื่อสร้างชื่อคอลัมน์ โดย i+1 เป็นการเพิ่มตัวเลขตามลำดับ

หลังจากนั้นจะทำการ Clustering ด้วย KMeans โดยใช้ KMeans clustering ในการแบ่งกลุ่มข้อมูลออกเป็น 3 Cluster (n_clusters=3) และกำหนดให้คอลัมน์ 'cluster' ใน DataFrame แสดงผลลัพธ์ของการแบ่งกลุ่ม จากนั้นสร้างกราฟ Scatter plot โดยใช้ Seaborn กำหนดให้แกน x เป็น PC1 และแกน y เป็น PC2 ซึ่งเป็นสอง Principal Components แรก มีการใช้สี (hue) เพื่อแสดงคลัสเตอร์ต่างๆ ที่ได้จาก KMeans จากนั้นเพิ่มชื่อกราฟและป้ายกำกับแกน x และ y



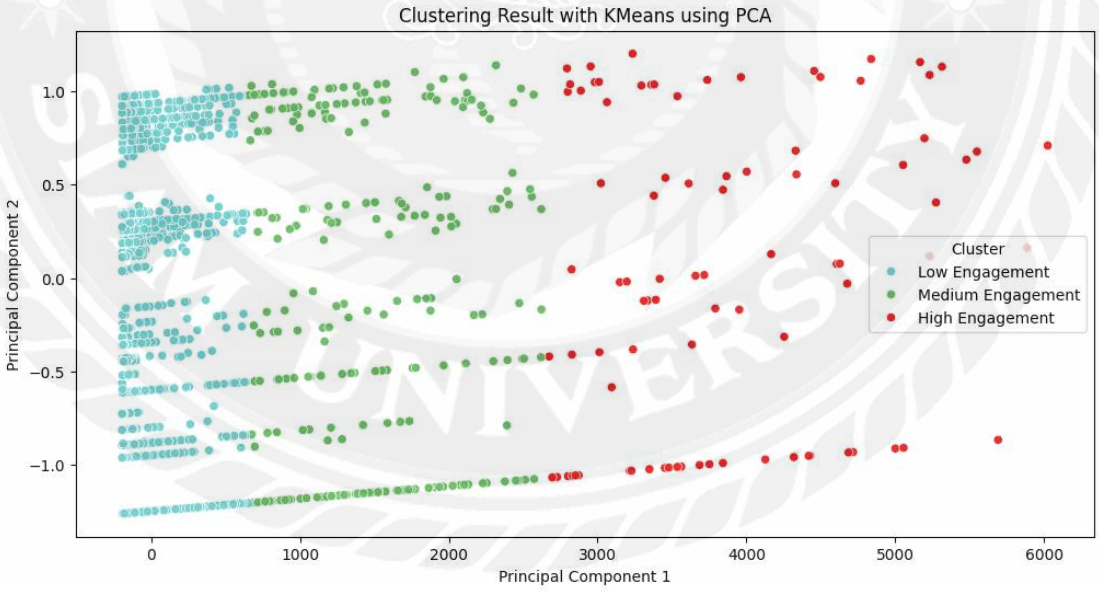
รูปที่ 3.19 แสดงกราฟ Scatter plot ที่ได้จากการทำ Clustering

จากรูปที่ 3.19 แสดงกราฟ Scatter plot ที่ได้จากการทำ Clustering โดยแบ่งออกเป็น 3 โดยที่แกน X มีค่ามากกว่าหรือใกล้ศูนย์ (เช่นเป็นหลักพัน) และแกน Y มีค่าระหว่าง -1 ถึง 1 บ่งบอกถึงความแปรปรวนที่แตกต่างกันในแต่ละแนวแกน จุดที่อยู่ใน cluster สีเขียวอาจบ่งบอกถึงข้อมูลที่มี Engagement ต่ำ ส่วนจุดใน cluster สีม่วงอาจบ่งบอกถึงข้อมูลที่มี Engagement สูง ซึ่งแกน X และ Y สามารถอธิบายได้ว่ามีการเปลี่ยนแปลงหลัก ๆ ที่ PC1 และการเปลี่ยนแปลงรองที่ PC2

```
cluster_names = {0: 'Low Engagement', 2: 'Medium Engagement', 1: 'High Engagement'}
data_no_outliers_pca['cluster_name'] = data_no_outliers_pca['cluster'].map(cluster_names)

custom_palette = {
    'Low Engagement': '#6fcdcd',
    'Medium Engagement': '#68bb63',
    'High Engagement': '#e42021'
}

plt.figure(figsize=(12, 6))
sns.scatterplot(x='PC1', y='PC2', hue='cluster_name', data=data_no_outliers_pca, palette=custom_palette)
plt.title('Clustering Result with KMeans using PCA')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.legend(title='Cluster')
plt.show()
```

รูปที่ 3.20 แสดงการเปลี่ยนชื่อและสีกลุ่มของ Cluster

จากรูปที่ 3.20 แสดงการเปลี่ยนชื่อและสีกลุ่มของ Cluster โดยสร้าง Dictionary เพื่อแม็พ (map) เลขที่แทนกลุ่ม (cluster) กับชื่อกลุ่มที่มีความหมาย (Low Engagement, Medium Engagement, High Engagement) และเพิ่มคอลัมน์ cluster_name ใน DataFrame data_no_outliers_pca โดยใช้ค่าในคอลัมน์ cluster และแทนที่ด้วยค่าจาก Dictionary ของ cluster_names. จากนั้นกำหนดพาเลทสีสำหรับแต่ละกลุ่ม (cluster_name) เพื่อใช้ในการวาดกราฟ ซึ่งได้กำหนดให้ Low Engagement กำหนดให้มีสีฟ้า และ Medium Engagement กำหนดให้มีสีเขียว ส่วน High Engagement กำหนดให้มีสีแดง



รูปที่ 3.21 แสดงกราฟ Scatter plot ที่ได้จากการเปลี่ยนชื่อกลุ่มและสีของ Cluster

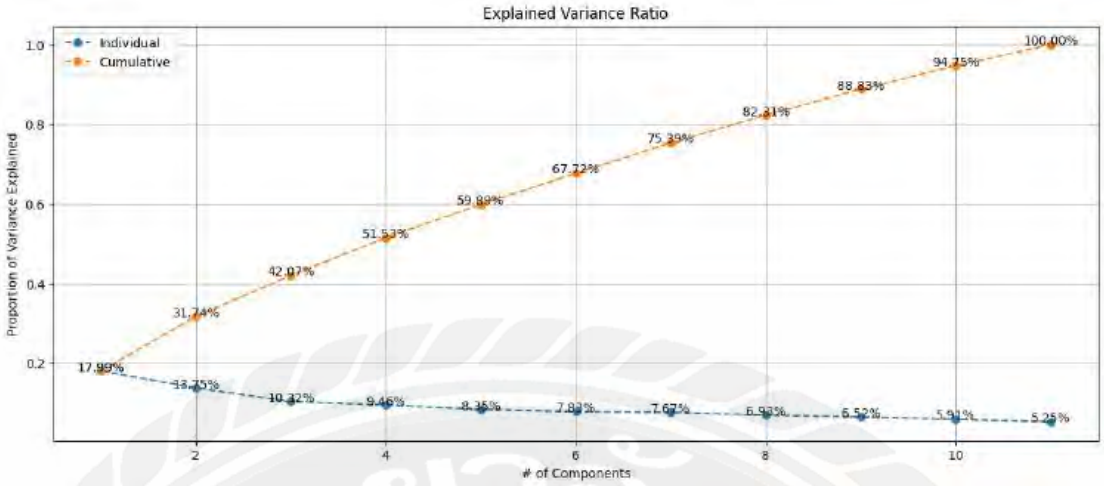
จากรูปที่ 3.21 แสดงกราฟ Scatter plot ที่ได้จากการเปลี่ยนชื่อกลุ่มและสีของ Cluster โดยมี 3 กลุ่ม คือ Low Engagement สีฟ้า, Medium Engagement สีเขียว และ High Engagement สีแดง ซึ่งการ Clustering นี้มี Feature ที่นำมาแบ่งกลุ่มคือ Topic(Product, Promotion, Activity, Presenter,

Other), Source(Facebook, Twitter, Youtube, Tiktok, Instagram, Forum), Post type(Real user, Brand post, KOL/Influencer, News, Shop) และ Engagement

การวิเคราะห์ต่อจากนี้จะเป็นการแบ่งกลุ่มในแต่ละหัวข้อ โดยหัวข้อที่จะแบ่งกลุ่มต่อคือ Product, Promotion และ Activity โดยการเขียนชุดคำสั่งจะเหมือนๆ กับการวิเคราะห์หัวข้อทั้งหมด แต่เพียงเปลี่ยนการฟิลเตอร์หัวข้อและจำนวน PCA

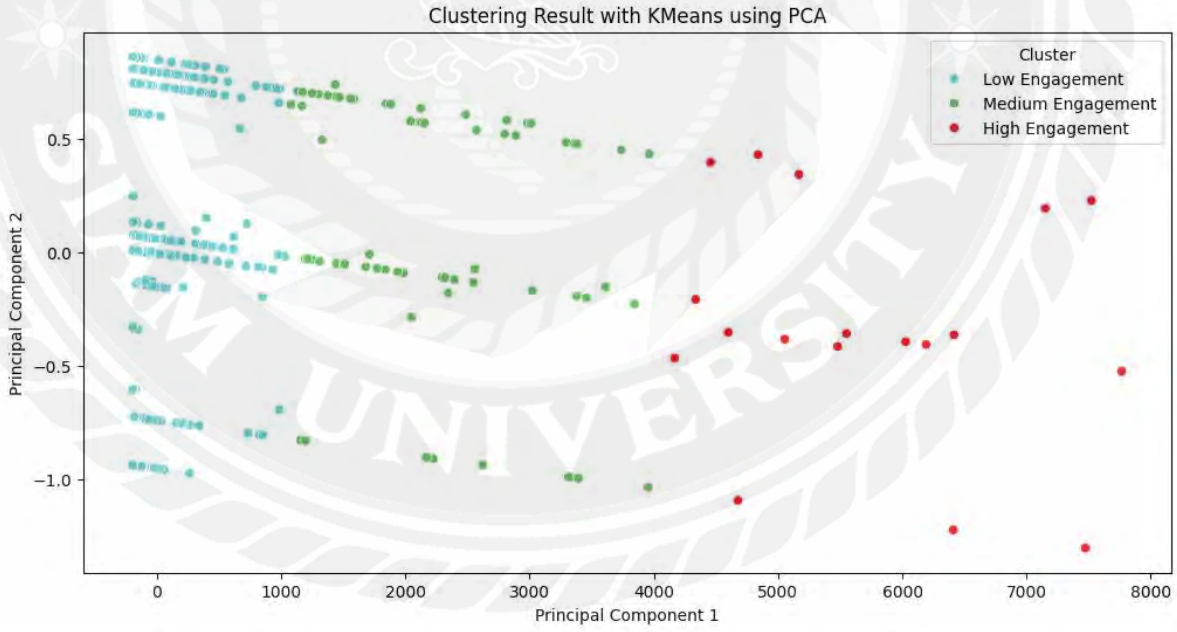
รูปที่ 3.22 แสดงการกรองข้อมูลและทำการ One-Hot Encoding สำหรับการจัดกลุ่มข้อมูลที่เป็น Category ที่ไม่ได้เป็นข้อมูลตัวเลข ใช้ในการวิเคราะห์การจัดกลุ่มหัวข้อ Product

จากรูปที่ 3.22 แสดงการจัดกลุ่ม Engagement สำหรับหมวดหมู่คอนเทนต์เฉพาะหัวข้อ Product และได้มีการเลือกแบรนด์ผลิตภัณฑ์เด็ก 5 แบรนด์ที่มีสินค้าใกล้เคียงกันและเป็นที่นิยมในตลาด โดยกำหนดชื่อแบรนด์ที่ต้องการในลิสต์ของตัวแปร brand_to_include โดยเลือกเฉพาะที่มีค่า Engagement มากกว่า 0 และค่าในคอลัมน์ Brand ตรงกับแบรนด์ที่อยู่ใน brands_to_include และให้เก็บผลลัพธ์ไว้ที่ตัวแปร data_filtered หลังจากนั้นใช้ฟังก์ชัน pd.get_dummies() เพื่อแปลงข้อมูลที่ Category (Source, TOPIC, และ Post type) เป็นค่าตัวเลขแบบ One-Hot Encoding (0,1) ในที่นี้ข้อมูลจะเป็น Boolean ซึ่งจะทำการแปลงเป็น (0,1) ในภายหลัง และคอลัมน์ใหม่จะมี prefix เป็น Source, TOPIC และ Post type ตามลำดับ ผลลัพธ์จะถูกเก็บในตัวแปร data_encoded จากนั้นทำการแปลงข้อมูล Boolean ที่ได้จากการทำ One-Hot Encoding เป็นข้อมูลตัวเลข (0,1) โดยเลือกคอลัมน์ที่มีประเภทข้อมูลเป็น bool (Boolean) แปลงคอลัมน์เหล่านี้เป็นประเภทข้อมูล int (Integer) เพื่อให้เข้ากันได้กับการวิเคราะห์ข้อมูลในขั้นตอนต่อไป



รูปที่ 3.23 แสดงกราฟ Scree plot

จากรูปที่ 3.23 แสดงกราฟ Scree plot โดยแกน X แสดงจำนวนขององค์ประกอบ (Principal Components) ซึ่งมีตั้งแต่ 1 ถึง 15 Component และแกน Y แสดงอัตราส่วนของความแปรปรวนที่อธิบายได้ (Proportion of Variance Explained) ซึ่งเป็นเปอร์เซ็นต์ มีเส้นสีน้ำเงินแสดงอัตราส่วนของความแปรปรวนที่แต่ละองค์ประกอบอธิบาย และเส้นสีส้มแสดงอัตราส่วนของความแปรปรวนสะสมที่อธิบายได้โดยรวมจากหลาย Component จากกราฟนี้ Component ที่ 8-10 จะอธิบายความแปรปรวนสะสมได้ประมาณ 80-90% ซึ่งคณะผู้จัดทำจึงเลือกใช้ PCA ที่จำนวน 9 Component



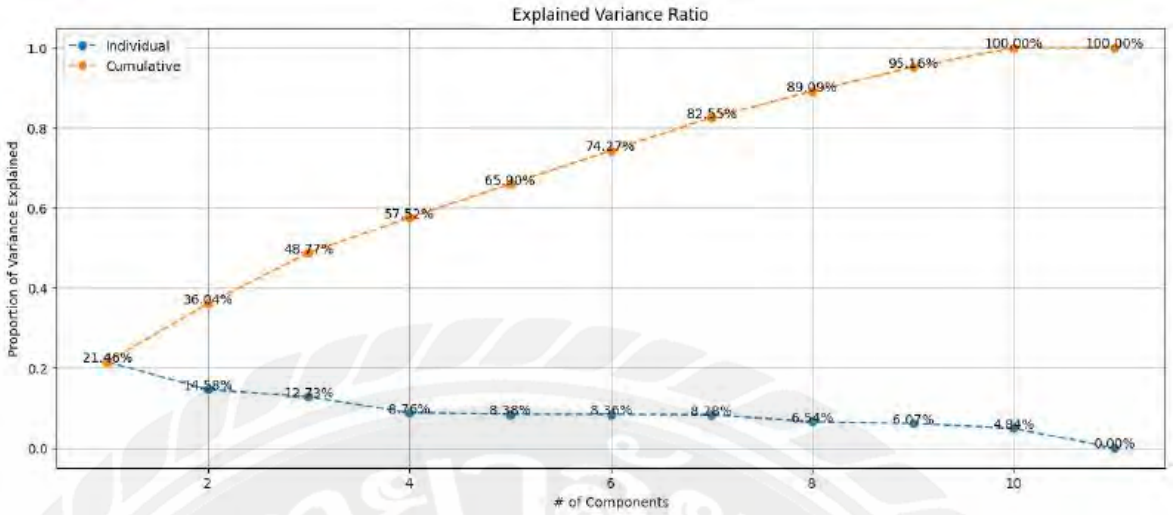
รูปที่ 3.24 แสดงกราฟ Scatter plot ของ Topic : Product

จากรูปที่ 3.24 แสดงกราฟ Scatter plot ที่ได้จากการเปลี่ยนชื่อกลุ่มและสีของ Cluster โดยมี 3 กลุ่ม คือ Low Engagement สีฟ้า Medium Engagement สีเขียว และ High Engagement สีแดง ซึ่งการ Clustering นี้มี Feature ที่นำมาแบ่งกลุ่มคือ Topic(Product), Source(Facebook, Twitter, Youtube, Tiktok, Instagram, Forum), Post type(Real user, Brand post, KOL/ Influencer, News, Shop) และ Engagement

```
brands_to_include = ['[redacted]', '[redacted]', '[redacted]', '[redacted]', '[redacted]']
data_filtered = data[(data['Engagement'] > 0) & (data['TOPIC'] == 'Promotion') & (data['Brand'].isin(brands_to_include))]


data_encoded = pd.get_dummies(data_filtered, columns=['Source', 'TOPIC', 'Post type'], prefix=['Source', 'TOPIC','Post type'])

boolean_columns = data_encoded.select_dtypes(include=['bool']).columns

data_encoded[boolean_columns] = data_encoded[boolean_columns].astype(int)

iso_forest = IsolationForest(contamination=0.01, random_state=42)
data_encoded['anomaly'] = iso_forest.fit_predict(data_encoded[['Engagement']])
```

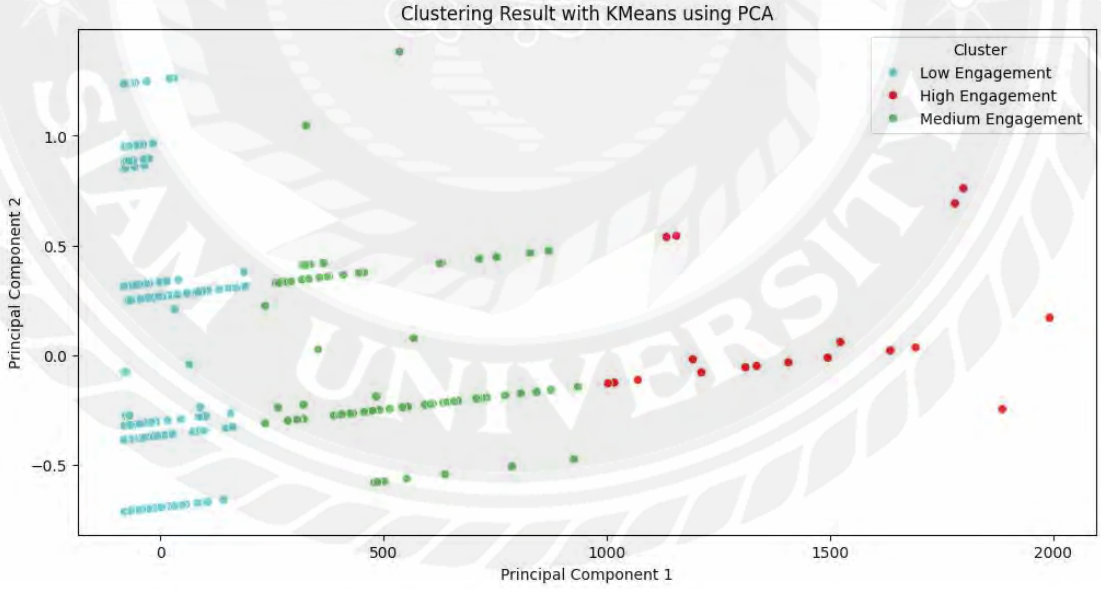
รูปที่ 3.25 แสดงการกรองข้อมูลและทำการ One-Hot Encoding สำหรับการจัดกลุ่มข้อมูลที่เป็น Category ที่ไม่ได้เป็นข้อมูลตัวเลข ใช้ในการวิเคราะห์การจัดกลุ่มหัวข้อ Promotion

จากรูปที่ 3.25 แสดงการจัดกลุ่ม Engagement สำหรับหมวดหมู่คอนเทนต์เฉพาะหัวข้อ Promotion และได้มีการเลือกแบรนด์ผลิตภัณฑ์เด็ก 5 แบรนด์ที่มีสินค้าใกล้เคียงกันและเป็นที่นิยมในตลาด โดยกำหนดชื่อแบรนด์ที่ต้องการในลิสต์ของตัวแปร brand_to_include โดยเลือกเฉพาะที่มีค่า Engagement มากกว่า 0 และค่าในคอลัมน์ Brand ตรงกับแบรนด์ที่อยู่ใน brands_to_include และให้เก็บผลลัพธ์ไว้ที่ตัวแปร data_filtered หลังจากนั้นใช้ฟังก์ชัน pd.get_dummies() เพื่อแปลงข้อมูลที่ Category (Source, TOPIC, และ Post type) เป็นค่าตัวเลขแบบ One-Hot Encoding (0,1) ในที่นี้ข้อมูลจะเป็น Boolean ซึ่งจะทำการแปลงเป็น (0,1) ในภายหลัง และคอลัมน์ใหม่จะมี prefix เป็น Source, TOPIC และ Post type ตามลำดับ ผลลัพธ์จะถูกเก็บในตัวแปร data_encoded จากนั้นทำการแปลงข้อมูล Boolean ที่ได้จากการทำ One-Hot Encoding เป็นข้อมูลตัวเลข (0,1) โดยเลือกคอลัมน์ที่มีประเภทข้อมูลเป็น bool (Boolean) แปลงคอลัมน์เหล่านี้เป็นประเภทข้อมูล int (Integer) เพื่อให้เข้ากันได้กับการวิเคราะห์ข้อมูลในขั้นตอนต่อไป



รูปที่ 3.26 แสดงกราฟ Scree plot

จากรูปที่ 3.26 แสดงกราฟ Scree plot โดยแกน X แสดงจำนวนขององค์ประกอบ (Principal Components) ซึ่งมีตั้งแต่ 1 ถึง 15 Component และแกน Y แสดงอัตราส่วนของความแปรปรวนที่อธิบายได้ (Proportion of Variance Explained) ซึ่งเป็นเปอร์เซ็นต์ มีเส้นสีน้ำเงินแสดงอัตราส่วนของความแปรปรวนที่แต่ละองค์ประกอบอธิบาย และเส้นสีส้มแสดงอัตราส่วนของความแปรปรวนสะสมที่อธิบายได้โดยรวมจากหลาย Component จากกราฟนี้ Component ที่ 7-9 จะอธิบายความแปรปรวนสะสมได้ประมาณ 80-90% ซึ่งคณะผู้จัดทำจึงเลือกใช้ PCA ที่จำนวน 7 Component



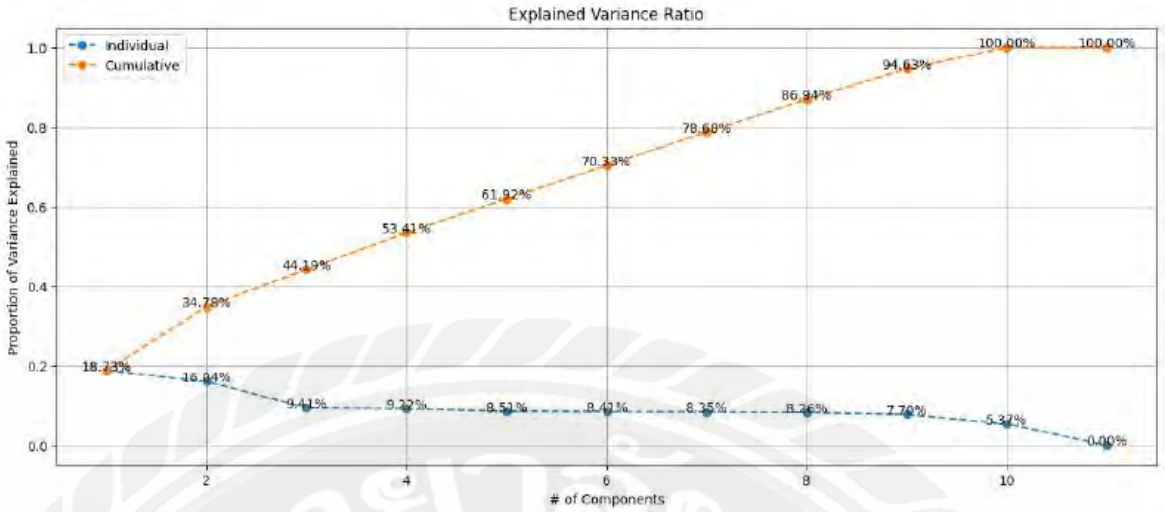
รูปที่ 3.27 แสดงกราฟ Scatter plot ของ Topic : Promotion

จากรูปที่ 3.27 แสดงกราฟ Scatter plot ที่ได้จากการเปลี่ยนชื่อกลุ่มและสีของ Cluster โดยมี 3 กลุ่ม คือ Low Engagement สีฟ้า, Medium Engagement สีเขียว และ High Engagement สีแดง ซึ่งการ Clustering นี้มี Feature ที่นำมาแบ่งกลุ่มคือ Topic(Promotion), Source(Facebook, Twitter, Youtube, Tiktok, Instagram, Forum), Post type(Real user, Brand post, KOL/ Influencer, News, Shop) และ Engagement

```
brands_to_include = ['[redacted]', '[redacted]', '[redacted]', '[redacted]', '[redacted]']
data_filtered = data[(data['Engagement'] > 0) & (data['TOPIC'] == 'Activity') & (data['Brand'].isin(brands_to_include))]


data_encoded = pd.get_dummies(data_filtered, columns=['Source', 'TOPIC', 'Post type'], prefix=['Source', 'TOPIC','Post type'])

boolean_columns = data_encoded.select_dtypes(include=['bool']).columns

data_encoded[boolean_columns] = data_encoded[boolean_columns].astype(int)

iso_forest = IsolationForest(contamination=0.01, random_state=42)
data_encoded['anomaly'] = iso_forest.fit_predict(data_encoded[['Engagement']])
```

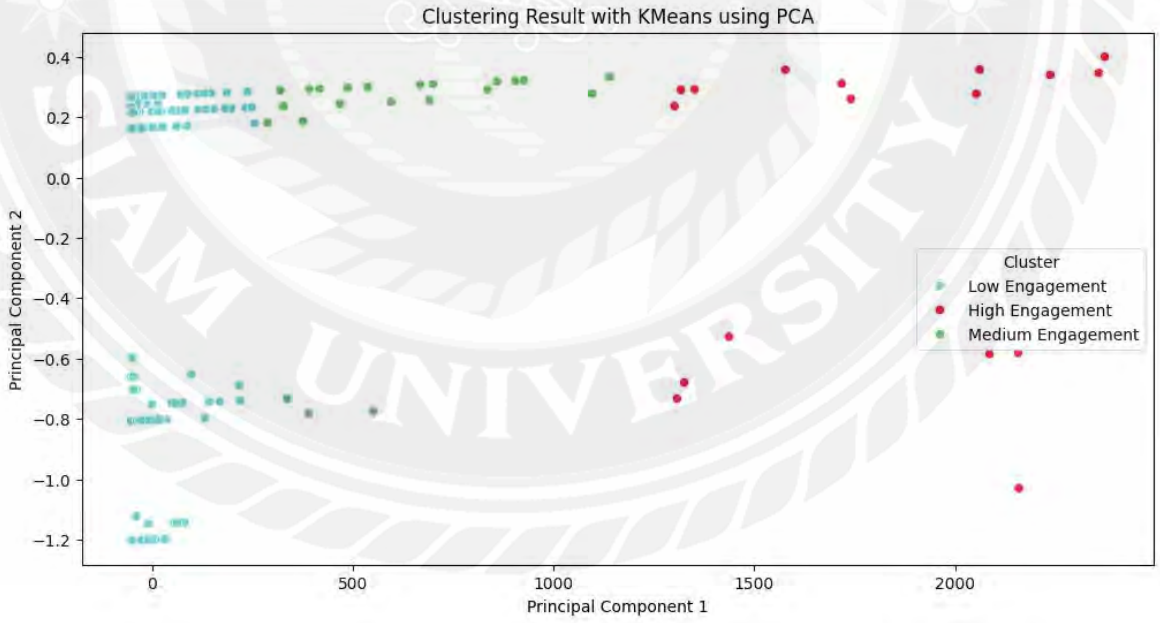
รูปที่ 3.28 แสดงการกรองข้อมูลและทำการ One-Hot Encoding สำหรับการจัดกลุ่มข้อมูลที่เป็น Category ที่ไม่ได้เป็นข้อมูลตัวเลข ใช้ในการวิเคราะห์การจัดกลุ่มหัวข้อ Activity

จากรูปที่ 3.28 แสดงการจัดกลุ่ม Engagement สำหรับหมวดหมู่คอนเทนต์เฉพาะหัวข้อ Activity และได้มีการเลือกแบรนด์ผลิตภัณฑ์เด็ก 5 แบรนด์ที่มีสินค้าใกล้เคียงกันและเป็นที่นิยมในตลาด โดยกำหนดชื่อแบรนด์ที่ต้องการในลิสต์ของตัวแปร brand_to_include โดยเลือกเฉพาะที่มีค่า Engagement มากกว่า 0 และค่าในคอลัมน์ Brand ตรงกับแบรนด์ที่อยู่ใน brands_to_include และให้เก็บผลลัพธ์ไว้ที่ตัวแปร data_filtered หลังจากนั้นใช้ฟังก์ชัน pd.get_dummies() เพื่อแปลงข้อมูลที่ Category (Source, TOPIC, และ Post type) เป็นค่าตัวเลขแบบ One-Hot Encoding (0,1) ในที่นี้ข้อมูลจะเป็น Boolean ซึ่งจะทำการแปลงเป็น (0,1) ในภายหลัง และคอลัมน์ใหม่จะมี prefix เป็น Source, TOPIC และ Post type ตามลำดับ ผลลัพธ์จะถูกเก็บในตัวแปร data_encoded จากนั้นทำการแปลงข้อมูล Boolean ที่ได้จากการทำ One-Hot Encoding เป็นข้อมูลตัวเลข (0,1) โดยเลือกคอลัมน์ที่มีประเภทข้อมูลเป็น bool (Boolean) แปลงคอลัมน์เหล่านี้เป็นประเภทข้อมูล int (Integer) เพื่อให้เข้ากันได้กับการวิเคราะห์ข้อมูลในขั้นตอนต่อไป



รูปที่ 3.29 แสดงกราฟ Scree plot

จากรูปที่ 3.29 แสดงกราฟ Scree plot โดยแกน X แสดงจำนวนขององค์ประกอบ (Principal Components) ซึ่งมีตั้งแต่ 1 ถึง 15 Component และแกน Y แสดงอัตราส่วนของความแปรปรวนที่อธิบายได้ (Proportion of Variance Explained) ซึ่งเป็นเปอร์เซ็นต์ มีเส้นสีน้ำเงินแสดงอัตราส่วนของความแปรปรวนที่แต่ละองค์ประกอบอธิบาย และเส้นสีส้มแสดงอัตราส่วนของความแปรปรวนสะสมที่อธิบายได้โดยรวมจากหลาย Component จากกราฟนี้ Component ที่ 8-9 จะอธิบายความแปรปรวนสะสมได้ประมาณ 80-90% ซึ่งคณะผู้จัดทำจึงเลือกใช้ PCA ที่จำนวน 8 Component



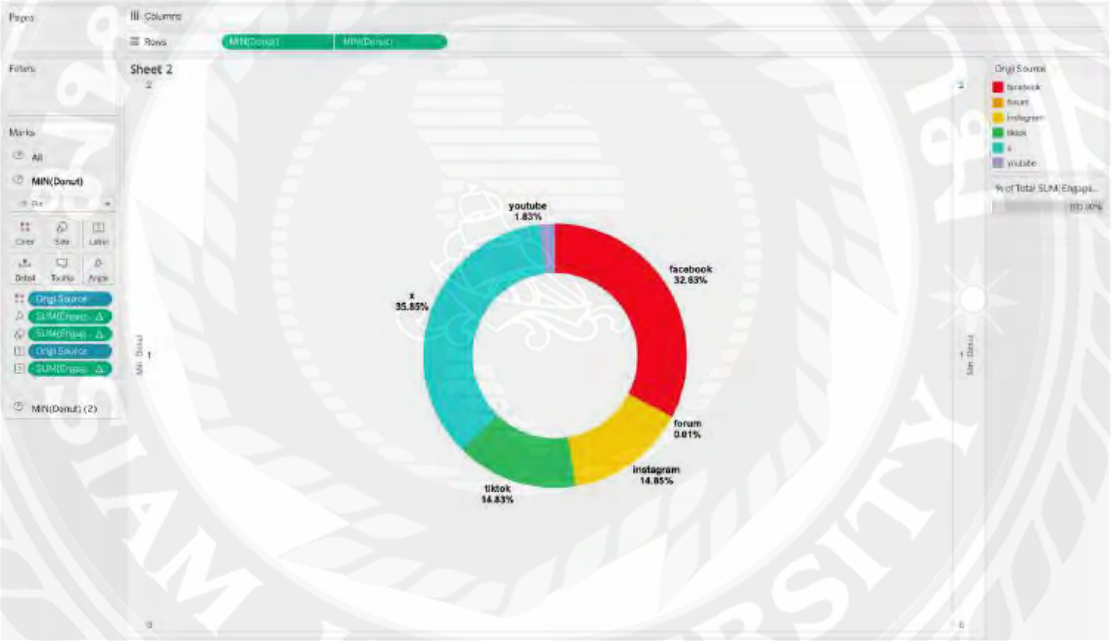
รูปที่ 3.30 แสดงกราฟ Scatter plot ของ Topic : Activity

จากรูปที่ 3.30 แสดงกราฟ Scatter plot ที่ได้จากการเปลี่ยนชื่อกลุ่มและสีของ Cluster โดยมี 3 กลุ่ม คือ Low Engagement สีฟ้า, Medium Engagement สีเขียว และ High Engagement สีแดง ซึ่งการ Clustering นี้มี Feature ที่นำมาแบ่งกลุ่มคือ Topic(Activity), Source(Facebook, Twitter, Youtube, Tiktok, Instagram, Forum), Post type(Real user, Brand post, KOL/ Influencer, News, Shop) และ Engagement

บทที่ 4
การนำเสนอแผนภาพของข้อมูล

หลังจากที่ได้ทำการเตรียมข้อมูล (Data Preparation) จนได้ข้อมูลที่สามารถไปทำการวิเคราะห์เพื่อจัดกลุ่ม (Data Clustering) ตามเงื่อนไขที่ได้ตั้งไว้แล้ว ทางคณะผู้จัดทำจะนำข้อมูลเข้าไปยังโปรแกรม Tableau เพื่อแสดงแผนภาพเพื่อวิเคราะห์ข้อมูลเพิ่มเติม และทำการ Drill down แต่ละ Cluster เพื่อให้ทราบถึงรายละเอียดการแบ่งกลุ่มตาม Engagement โดยกราฟที่ได้นำมาใช้ในการวิเคราะห์ครั้งนี้ เช่น กราฟแท่ง (Bar Chart) เพื่อแสดงการเปรียบเทียบ กราฟการกระจาย (Scatter Plot) เพื่อดูการแบ่งกลุ่ม และกราฟวงกลม (Pie Chart) เพื่อดูสัดส่วนของข้อมูล เป็นต้น

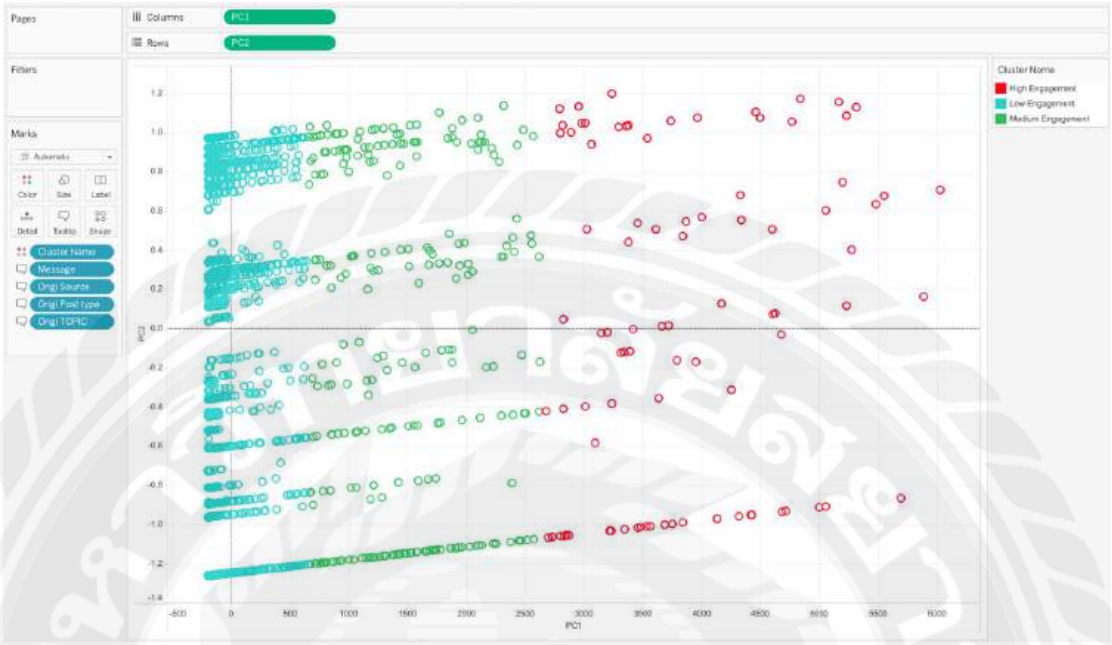
4.1 การวิเคราะห์ภาพรวมสัดส่วน Engagement ตามแพลตฟอร์มต่างๆ



รูปที่ 4.1 แสดงการวิเคราะห์ภาพรวมสัดส่วน Engagement ตามแพลตฟอร์มต่างๆ

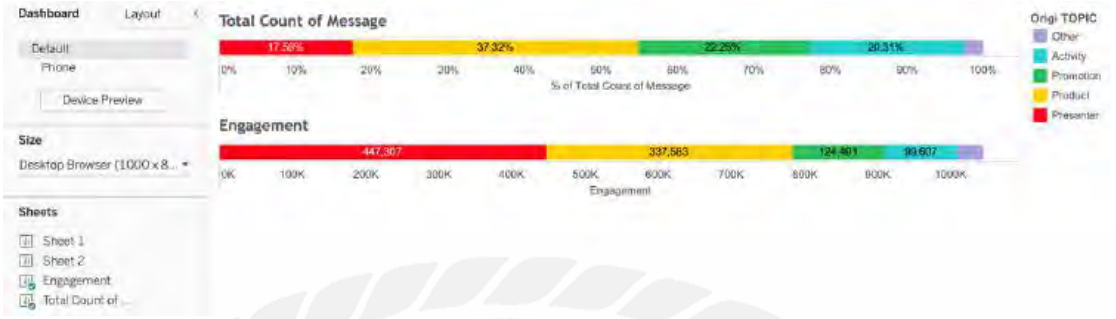
จากรูปที่ 4.1 แสดงการวิเคราะห์ภาพรวมสัดส่วน Engagement ตามแพลตฟอร์มต่างๆ โดยจะเห็นว่า Engagement ส่วนใหญ่มาจากแพลตฟอร์ม Twitter เป็นส่วนใหญ่ รองลงมาคือ Facebook, Instagram, Tiktok, Youtube และ Forum ตามลำดับ

4.2 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อทั้งหมดทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์



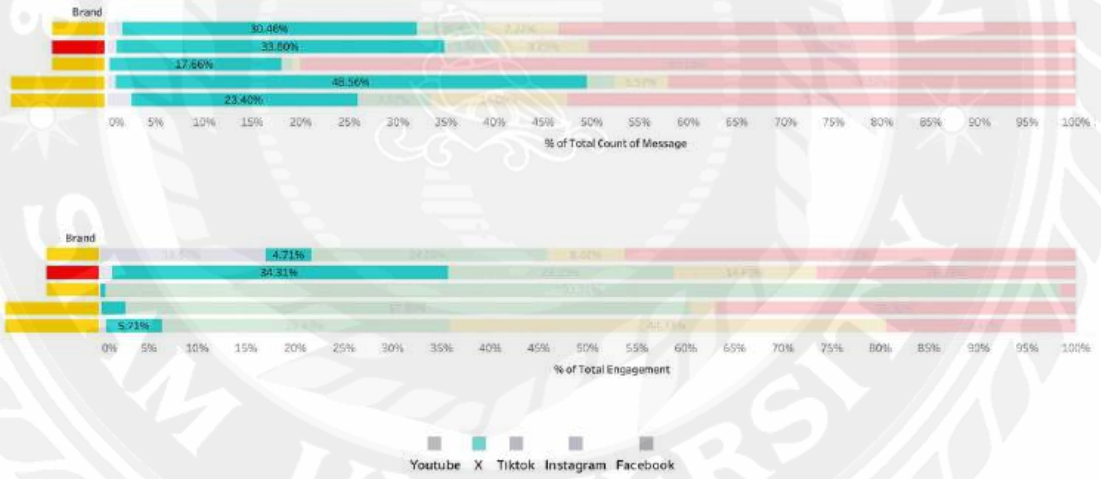
รูปที่ 4.2 แสดงการวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อทั้งหมดทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์

จากรูปที่ 4.2 แสดงการการจัดกลุ่มตาม Engagement ของหัวข้อทั้งหมดทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์ แบ่งกลุ่มออกได้ 3 กลุ่ม (Cluster) คือ 1) เครือข่ายคุณแม่ (Low Engagement) สีฟ้า ข้อมูลส่วนใหญ่เป็นการโพสต์หรือคอมเมนต์ลงบน Facebook เกี่ยวกับการถามหาสินค้าหรือพูดคุยแลกเปลี่ยนการใช้สินค้าโดยบุคคลทั่วไป 2) ผู้ชื่นชมดาราใน Twitter(X) (Medium Engagement) สีเขียว ข้อมูลส่วนใหญ่เป็นการโพสต์เกี่ยวกับการชื่มชมพรีเซนเตอร์บน Twitter (X) ที่เป็นคู่จิ้นจากซีรีส์วายโดยบุคคลทั่วไป ลักษณะการโพสต์เป็นลักษณะการชื่นชมด้วยข้อความที่เป็นเชิงบวก 3) ผู้สนับสนุนตัวยงใน Twitter (High Engagement) สีแดง ข้อมูลส่วนใหญ่เป็นการโพสต์เกี่ยวกับการชื่มชมพรีเซนเตอร์บน Twitter (X) ที่เป็นคู่จิ้นจากซีรีส์วายและโพสต์โดยบุคคลทั่วไป ลักษณะของข้อความจะเป็นการนำคำพูดของคู่จิ้นที่เป็นโมเม้นต์ระหว่างคู่มาโพสต์



รูปที่ 4.3 แสดงการเปรียบเทียบระหว่างจำนวนโพสต์และจำนวน Engagement ตามหัวข้อต่างๆ

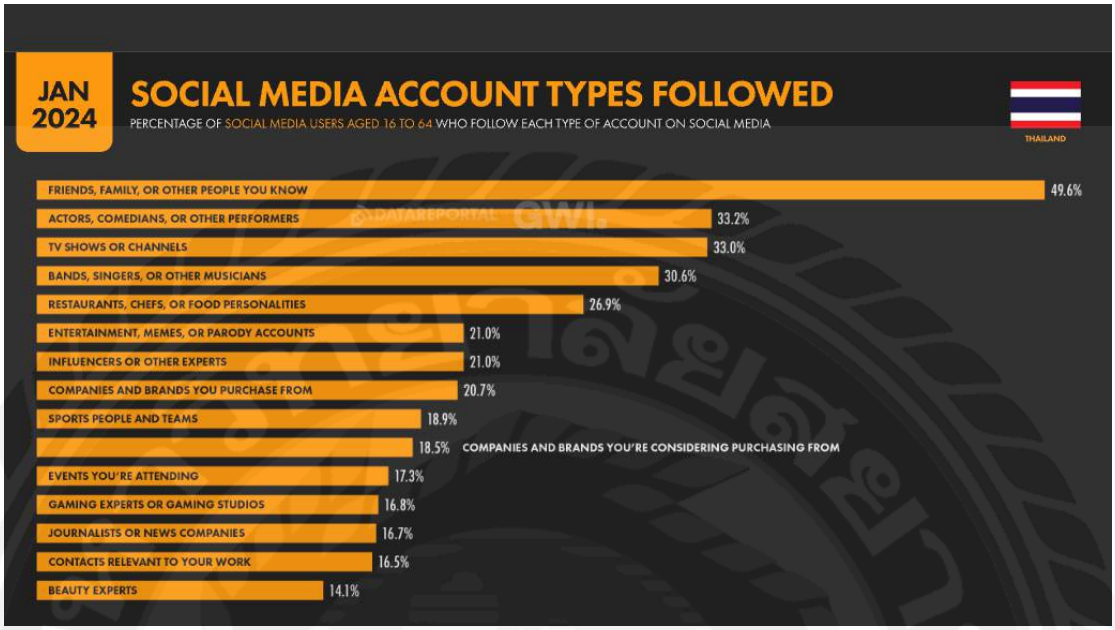
จากรูปที่ 4.3 แสดงให้เห็นว่ากราฟด้านบนจำนวนโพสต์ที่เกี่ยวกับ Product มีจำนวนสูงสุด รองลงมาคือ Promotion, Activity, Presenter และ Other ตามลำดับ ส่วนกราฟด้านล่าง Engagement ที่เกี่ยวกับ Presenter ได้รับความสนใจสูงมาก รองลงมาคือ Product, Promotion, Activity และ Other ตามลำดับ เมื่อดูควบคู่กันทำให้เห็นว่า หัวข้อที่เกี่ยวกับ Presenter เป็นที่ให้ความสนใจของผู้ชมคอนเทนต์ทำให้มีสัดส่วนการมีส่วนร่วมสูงที่สุด แม้ว่าจะมีสัดส่วนของจำนวนข้อความต่ำกว่าหัวข้ออื่นๆ



รูปที่ 4.4 แสดงการเปรียบเทียบระหว่างจำนวนโพสต์และจำนวน Engagement ของแบรนด์ต่างๆ โดยแพลตฟอร์ม Twitter (X)

จากรูปที่ 4.4 แสดงให้เห็นว่ากราฟด้านบนจำนวนโพสต์ที่มีการกล่าวถึงแบรนด์ต่างๆ มีจำนวนใกล้เคียงกัน แต่กราฟด้านล่างในส่วนของ Engagement ทำให้เห็นว่าแบรนด์สีแดงมีการกล่าวถึงบนทวิตเตอร์ค่อนข้างมากและ Engagement ที่ได้รับก็สูงตามเมื่อเทียบกับแบรนด์อื่นๆ ซึ่งแบรนด์สีแดงเป็นแบรนด์ที่มีการทำแคมเปญร่วมกับพรีเซ็นเตอร์จากซีรีส์วาย Twitter (X) เป็น

แพลตฟอร์มที่การปั่นยอด Engagement ทำได้โดยการทวิตและรีทวิตพร้อมกับติด แฮชแท็ก เพื่อให้ติดเทรนด์ทวิต หากติดเทรนด์ทวิตแล้วจะทำให้เกิดการมองเห็นเป็นวงกว้างมากขึ้น

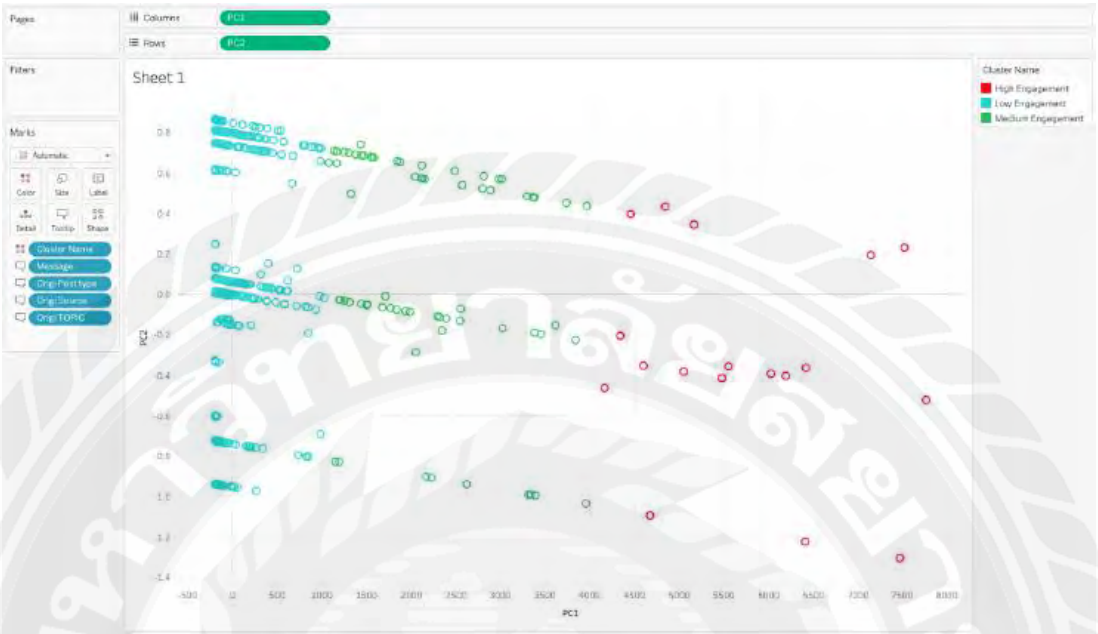


รูปที่ 4.5 ประเภทของบัญชีโซเชียลมีเดียที่คนไทยติดตามมากที่สุด

ที่มา: https://wearesocial.com/uk/blog/2023/04/the-global-state-of-digital-in-april-2023/

จากรูปที่ 4.5 แสดงข้อมูลจาก DataReportal แสดงให้เห็นว่าประเภทของบัญชีโซเชียลมีเดียที่คนไทยติดตามมากที่สุด คือเพื่อนหรือครอบครัว ส่วนอันดับรองลงมาคือ นักแสดงหรือบุคคลที่มีชื่อเสียงทำให้เห็นว่า ผู้คนให้ความสนใจในนักแสดงหรือบุคคลที่มีชื่อเสียงรองจากบุคคลใกล้ชิด

4.3 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Product ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์

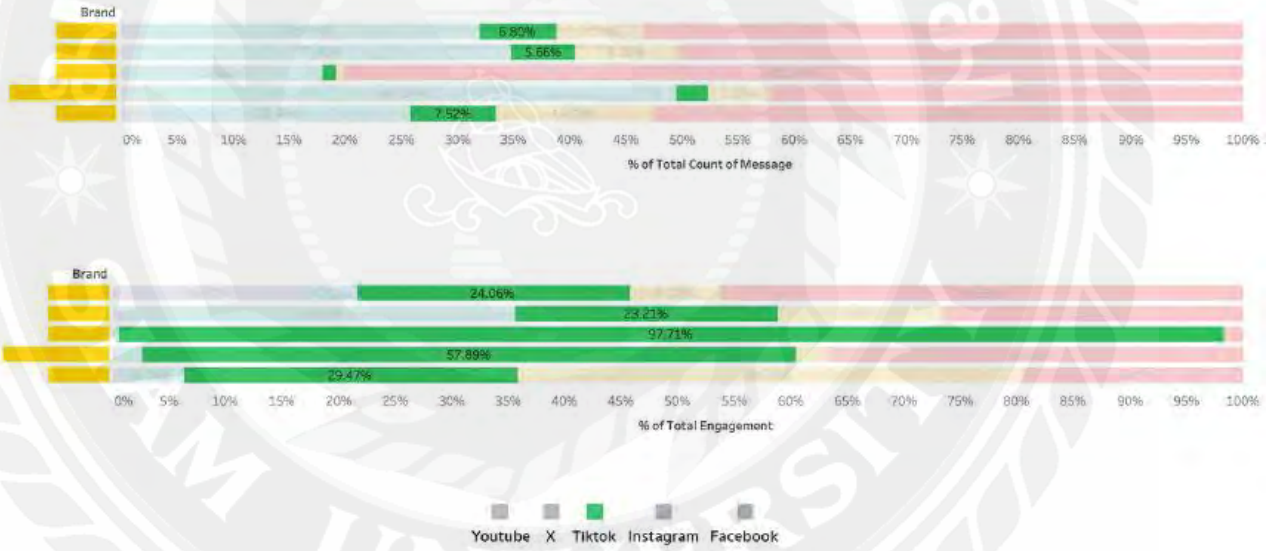


รูปที่ 4.6 แสดงการวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Product ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์

จากรูปที่ 4.6 แสดงการวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Product ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์ สามารถแบ่งกลุ่มออกได้ 3 กลุ่ม (Cluster) คือ 1) เครือข่ายคุณแม่ (Low Engagement) สีฟ้า ข้อมูลส่วนใหญ่เป็นการโพสต์หรือคอมเมนต์ลงบน Facebook เกี่ยวกับการถามหาสินค้าหรือพูดคุยแลกเปลี่ยนการใช้สินค้าโดยบุคคลทั่วไป 2) อินฟลูเอนเซอร์บน Facebook (Medium Engagement) สีเขียว ข้อมูลส่วนใหญ่เป็นการโพสต์เกี่ยวกับการรีวิวสินค้าลงบน Facebook โดย Influencer หรือคนดังที่มีชื่อเสียง 3) อินฟลูเอนเซอร์บน Tiktok (High Engagement) สีแดง ข้อมูลส่วนใหญ่เป็นการโพสต์เกี่ยวกับการรีวิวสินค้าลงบน Tiktok โดย Influencer หรือคนดังที่มีชื่อเสียง

รูปที่ 4.7 แสดงการเปรียบเทียบระหว่างจำนวนโพสต์และจำนวน Engagement สำหรับหัวข้อ Product

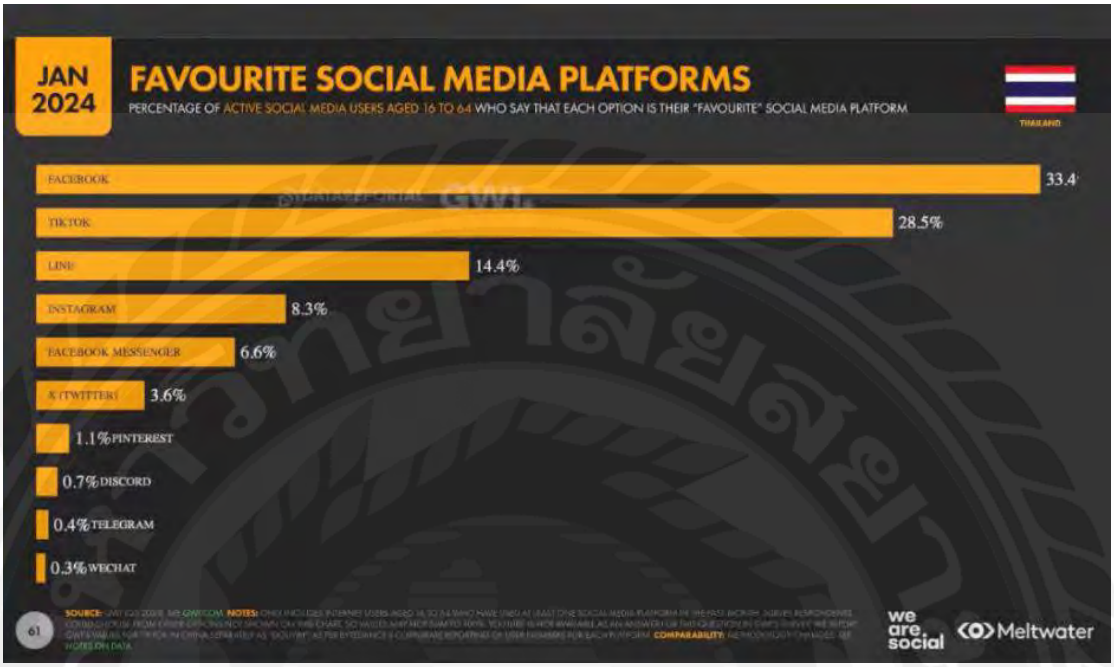
จากรูปที่ 4.7 แสดงให้เห็นว่ากราฟด้านบนจำนวนโพสต์ที่เกี่ยวกับ Product มีจำนวนสูงสุดในแพลตฟอร์ม Facebook รองลงมาคือ Tiktok, Twitter (X), Instagram, Youtube และ Forum ตามลำดับ ส่วนกราฟด้านล่าง Facebook ได้รับความสนใจสูงสุด รองลงมาคือ Tiktok, Instagram, Youtube, Twitter (X) และ Other ตามลำดับ เมื่อดูควบคู่กันทำให้เห็นว่า Tiktok แม้จะมีการพูดถึงน้อยแต่ Engagement ที่ได้รับกลับสูง



จากรูปที่ 4.8 แสดงการเปรียบเทียบระหว่างจำนวนโพสต์และจำนวน Engagement ของแบรนด์ต่างๆ โดยแพลตฟอร์ม Tiktok

จากรูปที่ 4.8 แสดงให้เห็นว่ากราฟด้านบนจำนวนโพสต์ที่มีการกล่าวถึงแบรนด์ต่างๆ ที่ค่อนข้างน้อยมาก แต่กราฟด้านล่างแสดงถึงการมี Engagement ที่สูงทุกแบรนด์ ความได้เปรียบของ Tiktok เป็นแพลตฟอร์มที่การรับชมคอนเทนต์แต่ละคอนเทนต์ได้ไว เนื่องจากเป็นแพลตฟอร์มที่เน้นเป็นคลิปวีดีโอความยาวแต่ละคอนเทนต์ส่วนใหญ่ไม่เกิน 1 นาที และยังมีสิ่งอำนวยความ

สะดวกที่เรียกว่า แผ่นเสียง การคลิกไปที่แผ่นเสียงจะทำให้เห็นคอนเทนต์ที่ใช้แผ่นเสียงนี้ทั้งหมด ซึ่งสะดวกสบายต่อการค้นหาและการรับชมคอนเทนต์ต่างๆ

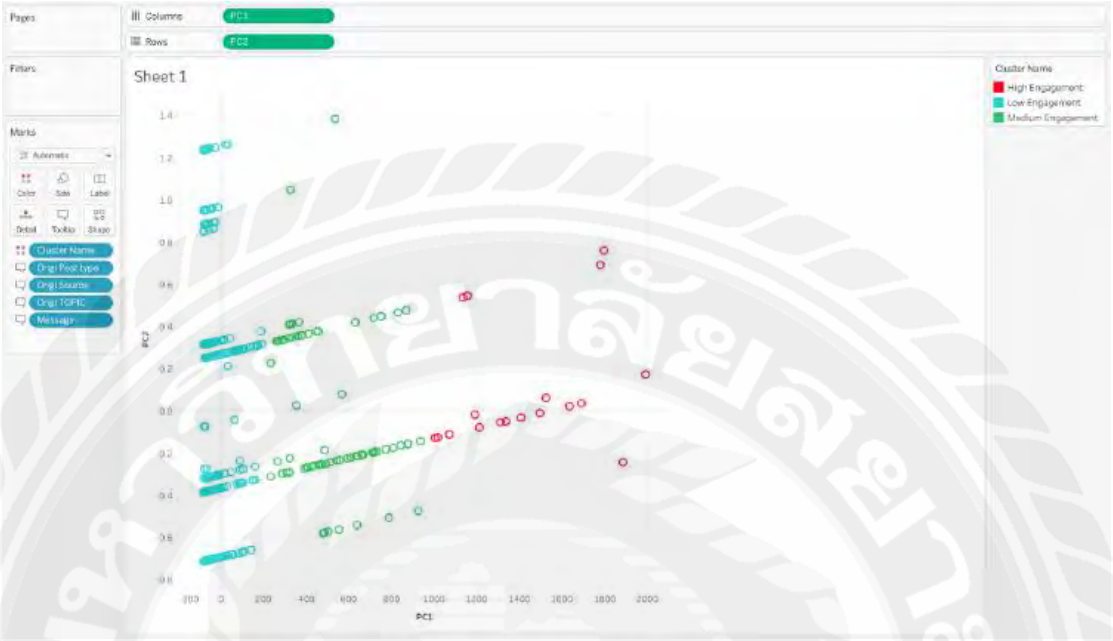


รูปที่ 4.9 ประเภทของบัญชีโซเชียลมีเดียที่คนไทยติดตามมากที่สุด

ที่มา: https://wearesocial.com/uk/blog/2023/04/the-global-state-of-digital-in-april-2023/

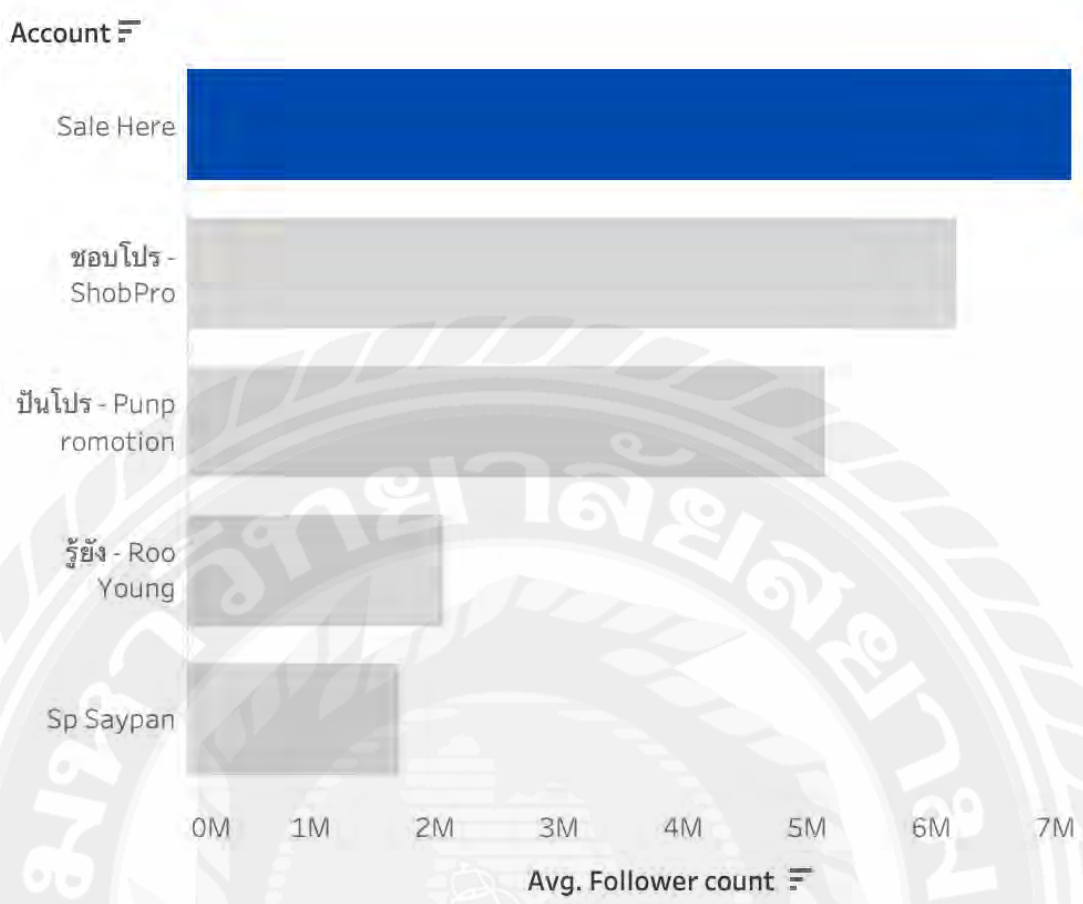
จากรูปที่ 4.9 ข้อมูลจาก DataReportal แสดงให้เห็นว่าประเภทของแพลตฟอร์มโซเชียลที่คนไทยชอบมากที่สุดคือ Facebook และรองลงมาคือ Tiktok ซึ่งทั้ง 2 แพลตฟอร์มเป็นแพลตฟอร์มที่ได้รับความนิยมในปัจจุบันอย่างมาก แบรนด์ต่างๆ เริ่มมาทำการตลาดบน Tiktok มากขึ้น คลิปไหนที่เป็นไวรัลจะยิ่งทำเกิดความสนใจมากยิ่งขึ้น

4.4 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Promotion ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์



รูปที่ 4.6 แสดงการวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Promotion ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์

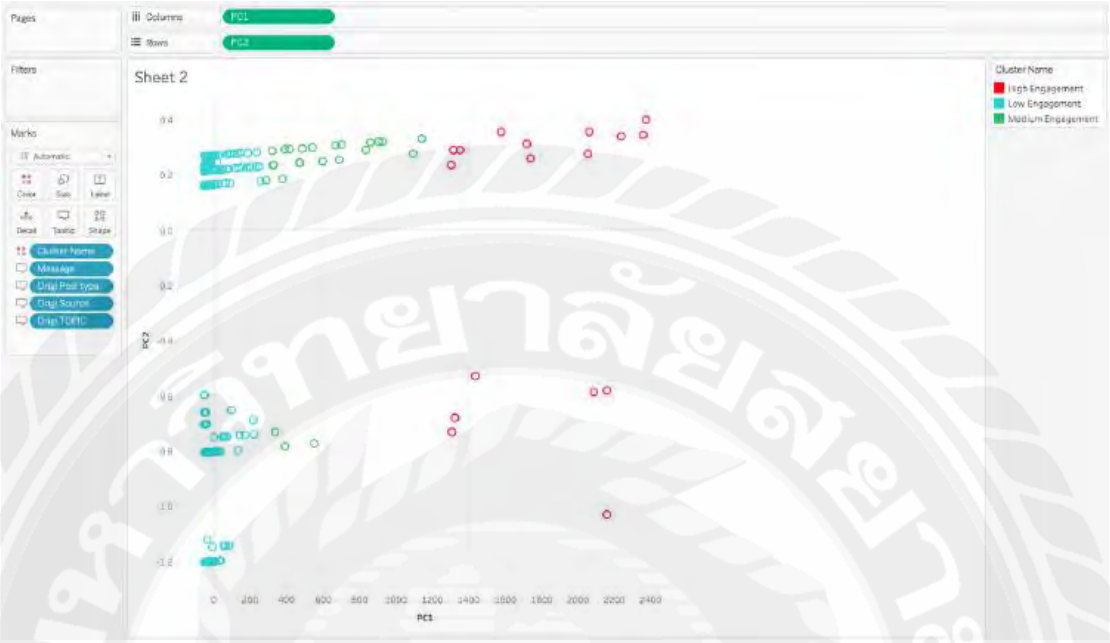
จากรูปที่ 4.6 แสดงการวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Promotion ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์ โดยสามารถแบ่งกลุ่มออกได้ 3 กลุ่ม (Cluster) คือ 1) เพจร้านค้าบน Facebook (Low Engagement) สีฟ้า ข้อมูลส่วนใหญ่เป็นการโพสต์เกี่ยวกับ Promotion สินค้าลงบน Facebook โดยร้านค้าต่างๆ 2) Sale Here วงการโปรโมชัน (Medium Engagement) สีเขียว ข้อมูลส่วนใหญ่มาจากโพสต์ของเพจ Sale Here ซึ่งเป็นเพจโปรโมชันที่โด่งดังของคนไทย เป็นเพจที่จะมีการทำ Artwork ที่ดึงดูดและสะดุดตาอยู่เสมอๆ 3) Sale Here ครองใจนักช้อป (High Engagement) สีแดง ข้อมูลส่วนใหญ่ยังคงมาจากโพสต์ของเพจ Sale Here ซึ่งเป็นเพจโปรโมชันที่โด่งดังของคนไทย เป็นเพจที่จะมีการทำ Artwork ที่ดึงดูดและสะดุดตาอยู่เสมอๆ



รูปที่ 4.7 แสดงการเปรียบเทียบยอดผู้ติดตามเพจโปรโมชัน

จากรูปที่ 4.7 เมื่อเปรียบเทียบยอดผู้ติดตามของเพจโปรโมชัน จะพบว่าเพจที่มาแรงอันดับหนึ่งคือ เพจ Sale Here ยิ่งมีผู้ติดตามเยอะ Engagement จะเยอะตาม หากผู้ติดตามทำการแชร์ออกไปอีกก็ยิ่งเพิ่มการมองเห็นและเพิ่ม Engagement เพิ่มเข้าไปอีก

4.5 การวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Activity ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์



รูปที่ 4.7 แสดงการวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Activity ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์

จากรูปที่ 4.7 แสดงการวิเคราะห์การจัดกลุ่มตาม Engagement ของหัวข้อ Activity ทุกแพลตฟอร์มบนโซเชียลมีเดีย และทุกประเภทของผู้โพสต์ โดยสามารถแบ่งกลุ่มออกได้ 3 กลุ่ม (Cluster) คือ 1) ผู้ร่วมสนุกบนเฟสบุ๊ค (Low Engagement) สีฟ้า ข้อมูลส่วนใหญ่เป็นการตอบคำถามผ่านเพจต่างๆ เพื่อร่วมสนุกกับแคมเปญหรือกิจกรรมของแบรนด์บน Facebook โดยบุคคลทั่วไป 2) กิจกรรมบนเฟสบุ๊คจากแบรนด์ (Medium Engagement) สีเขียว ข้อมูลส่วนใหญ่เป็นการโพสต์เกี่ยวกับ แคมเปญหรือกิจกรรม ลงบน Facebook โดยแบรนด์เป็นคนโพสต์เอง 3) อินฟลูเอนเซอร์แจ้งกิจกรรม (High Engagement) สีแดง ข้อมูลส่วนใหญ่เป็นการโพสต์เกี่ยวกับ แคมเปญหรือกิจกรรม ลงบน Facebook โดย Influencer หรือคนดังที่มีชื่อเสียง

บทที่ 5
สรุปผลและข้อเสนอแนะ

5.1 สรุปผลปริญญานิพนธ์

ผลที่ได้จากการศึกษาและวิเคราะห์ปัจจัยที่ทำให้เกิดพฤติกรรมและความสนใจในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย โดยนำข้อมูล Social listening จาก Zocial eye ตั้งแต่วันที่ 1 ตุลาคม 2566 ถึง 31 มีนาคม 2567 โดยนำเสนอด้วยแผนภาพข้อมูล (Data Visualization) ด้วยโปรแกรม Tableau ซึ่งจากผลการวิเคราะห์ได้ผลลัพธ์ ดังนี้

1) พรีเซนเตอร์หรือบุคคลที่มีชื่อเสียง การใช้คนมีชื่อเสียงที่เป็นกระแสหรือมีลูกน้อยจะดึงดูดความสนใจของผู้ชมมากขึ้น ความน่ารักของเด็กๆ ทำให้เกิดความน่าสนใจได้ อีกทั้งการใช้ พรีเซนเตอร์รุ่นใหม่หรือซีรีส์วายที่กำลังเป็นกระแสจะช่วยให้ได้กลุ่มเป้าหมายใหม่ๆ นอกจากกลุ่มคุณแม่
2) แฮชแท็ก (#) การมีแฮชแท็กปรากฏบนโพสต์ทำให้คนอยากกดเข้าไปดูรายละเอียดว่าทำไมแฮชแท็กนี้ถึงขึ้นมาบนโลกโซเชียล
3) การใช้ภาพหรือวีดีโอภาพและวีดีโอเป็นสิ่งที่ดึงดูดสายตาจากผู้คนมากมาย ยิ่งทำให้สะดุดตามากเท่าไร ยิ่งมีคนสนใจและคลิกเข้าไปดูในการเลือกดูคอนเทนต์เกี่ยวกับผลิตภัณฑ์สำหรับเด็กที่อยู่บนโซเชียลมีเดีย แพลตฟอร์ม Tiktok เป็นแพลตฟอร์มสำหรับวีดีโอ ซึ่งเหมาะสมกับการโปรโมทสินค้าเป็นอย่างมาก หากต้องการเพิ่มความน่าสนใจเข้าไปอีก อาจจะพิจารณาการโปรโมทสินค้าร่วมกับ Influencer
4) ความคิดเห็นและปฏิสัมพันธ์ของคนอื่นๆ ยอดไลค์หรือโพสต์ที่มีคอมเมนต์เยอะๆ จะช่วยดึงดูดให้ผู้คนสงสัยว่าเกิดอะไรขึ้น ทำไมถึงมีการ Interactive สูง
5) การจัดแคมเปญหรือกิจกรรมกิจกรรมที่มีการแจกของหรือของฟรีที่เกิดจากการโพสต์จาก Influencer จะยิ่งทำให้เกิดความน่าสนใจมากขึ้น

5.2 ข้อเสนอแนะ

5.2.1 การทดสอบและปรับปรุงคอนเทนต์ ควรทำการทดสอบรูปแบบและเนื้อหาคอนเทนต์อย่างต่อเนื่อง เพื่อหาวิธีการที่มีประสิทธิภาพสูงสุดในการดึงดูดความสนใจของผู้ชมและเพิ่ม Engagement

5.2.2 การวิเคราะห์เชิงภูมิศาสตร์และประชากรศาสตร์ ควรนำข้อมูลเชิงภูมิศาสตร์และประชากรศาสตร์มาวิเคราะห์ร่วมด้วย เพื่อให้เข้าใจความแตกต่างของกลุ่มเป้าหมายในแต่ละพื้นที่ และสามารถปรับปรุงกลยุทธ์การตลาดให้เหมาะสมกับกลุ่มเป้าหมายในแต่ละภูมิภาค

5.2.3 การวิเคราะห์ในเชิงลึกของเนื้อหา ควรพิจารณาวิเคราะห์เนื้อหา (Content Analysis) อย่างละเอียด เพื่อเข้าใจว่าเนื้อหาประเภทใดที่ดึงดูดความสนใจมากที่สุด เช่น การเล่าเรื่อง การให้ข้อมูลเชิงลึก หรือการใช้ภาษาแบบไหนที่สามารถสร้างความสัมพันธ์กับผู้ชมได้ดี

5.2.4 การปรับปรุงการใช้งานโซเชียลมีเดีย ควรศึกษาเวลาที่เหมาะสมในการโพสต์คอนเทนต์ เพื่อให้เข้าถึงกลุ่มเป้าหมายในช่วงเวลาที่พวกเขาใช้งานโซเชียลมีเดียมากที่สุด ซึ่งจะช่วยเพิ่มโอกาสในการได้รับ Engagement สูงขึ้น

5.2.5 การใช้การตลาดแบบ Influencer พิจารณาใช้กลยุทธ์การตลาดแบบ Influencer ที่มีความเชี่ยวชาญและมีผู้ติดตามที่ตรงกับกลุ่มเป้าหมาย เพื่อเพิ่มความน่าเชื่อถือและสร้างความสนใจในผลิตภัณฑ์

บรรณานุกรม

กาเมเดฟ. (2565, 5 มิถุนายน). *Visual studio code*. [เว็บบล็อก]. https://codingonblog.com/what-is-vscode-codingonblog/

จุฑารัฐ ธังกิตติคุณ. (2562, 3 มกราคม). *K-means Clustering*. Medium. https://medium.com/@Jutharath.Thankittikoon/what-is-k-means-clustering-ee36ab6f7638

จิรัฏฐ์ บุญพันธ์. (2564, 13 กุมภาพันธ์). *Word Cloud*. Medium. https://jirat-boonphun.medium.com/word-cloud-สำหรับวันวาเลนไทน์-23bc88af4f3c

บีเอซี. (2567, 27 กุมภาพันธ์). *โปรแกรม Tableau*. Business Applications. https://www.bac.co.th/web/products-and-services/tableau-software/

เบลนดาต้า. (2567, 25 เมษายน). *Data Analytics*. Blendata. https://www.blendata.co/ทำความรู้จัก-data-analytics-ศาสตร์แห่/

พนา. (2557, 11 มิถุนายน). *Data Preparation*. BzInsight. https://bzinsight.wordpress.com/2014/06/11/การทำ-data-preparation-อย่างมืออาชีพ/

พอล. (2565, 15 กันยายน). *PCA (Principal component analysis)*. Medium. https://medium.com/kbtg-life/principal-component-analysis-pca-ฉบับละเอียดและเข้าใจง่าย-d0c514e6fa0a

ฟิสท์คราฟท์. (2563, 8 สิงหาคม). *Data Visualization*. 1stCraft. https://1stcraft.com/what-is-data-visualization/

มุนลิกา. (2562, 7 ตุลาคม). *Clustering Model*. Medium. https://medium.com/tni-university/การทำ-machine-learning-ด้วย-clustering-model-2a3c392e7faa

มายด์พีเอชพี. (2562, 6 มีนาคม). *Hierarchical Clustering*. Medium. https://www.mindphp.com/คู่มือ/73-คืออะไร/6851-what-is-hierarchical-clustering.html

ไวซ์ไซท์. (2565, 27 กันยายน). *Social Listening*. Wisesight. https://wisesight.com/th/news/social-listening-2/

ไวซ์ไซท์. (2565, 28 กันยายน). *ZOCIAL EYE*. Wisesight. https://wisesight.com/th/news/zocial-eye/

อาณัติชัย เตชะวิเสสชัย. (2565, 11 กุมภาพันธ์). *Isolation Forest*. Big Data Experience Center. https://bigdataexperience.org/anomaly-detection-with-isolation-forest-แยกข้อมูลผิดปกติง/