

การศึกษาปัจจัยที่เกี่ยวข้อง กับการเข้าดูพอดแคสต์ ของคนไทยด้วยหลักการ ทางสถิติ

Presented by :

6530182121 ธนันพ

6530368621 วันพิชิต

6532170021 คิวพร

The Secret Sauce

Overview

01

บทนำ

02

วิธีการดำเนินงาน

03

ผลการดำเนินงาน

04

การวิเคราะห์ทางสถิติ

05

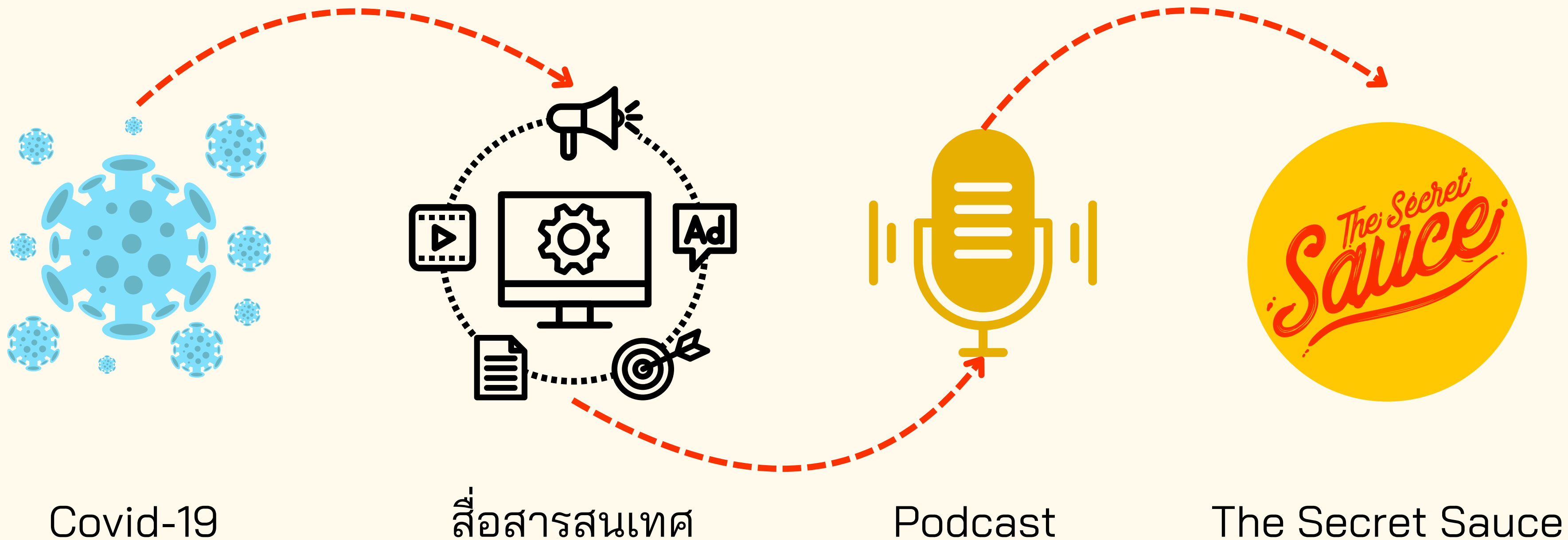
สรุปผล

บทที่ 1

บทนำ

- ที่มาและความสำคัญ
- วัตถุประสงค์
- ขอบเขตของการศึกษา
- เครื่องมือ

ที่มาและความสำคัญ



วัตถุประสงค์

เพื่อวิเคราะห์และหาความสัมพันธ์ระหว่าง
หมวดหมู่ในการฟัง podcast กับ จำนวนคนดู
ผ่านทาง Youtube ใน 채널 The Secret
Sauce โดยใช้เครื่องมือทางสถิติ

1

เพื่อศึกษาพฤติกรรมและรูปแบบการเข้า
ฟัง Podcast ของคนไทยที่ส่งผลต่อ
จำนวนคนดูของ 채널 The Secret
Sauce บน YouTube

2

เพื่อวิเคราะห์และหาความสัมพันธ์ระหว่าง
ปัจจัยต่าง ๆ ที่เกี่ยวข้องในการฟัง Podcast
ของคนไทยกับจำนวนคนดูและจำนวนคนกด
ถูกใจบนวิดีโอ Podcast ของ 채널 The
Secret Sauce โดยใช้เครื่องมือทางสถิติ

3

เพื่อนำเสนอและวิเคราะห์ข้อมูลในรูปแบบภาพ
ระหว่างปัจจัยต่าง ๆ ที่เกี่ยวข้องในการฟัง
Podcast ของคนไทยกับจำนวนคนดูและจำนวน
คนกดถูกใจบนวิดีโอ Podcast ของ 채널
The Secret Sauce

ขอบเขตของการศึกษา

ข้อมูลการฟังการ Podcast จาก
YouTube บน채นแนล
The Secret Sauce
ตั้งแต่วันที่ 21 กุมภาพันธ์ พ.ศ. 2561
จนถึง 21 มีนาคม พ.ศ. 2566
จำนวน **1039** วิดีโอ
เป็นเวลา 5 ปี 28 วัน

องค์ความรู้จากรายวิชา

2104253 สถิติเชิงวิศวกรรม 1

2104254 สถิติเชิงวิศวกรรม 2

เครื่องมือ

1. การดำเนินการศึกษาหลัก :
Microsoft Excel และ Microsoft word
2. เก็บข้อมูลและทำการศึกษาย่อย :
Google Drive และ Google Sheet
3. การติดต่อสื่อสาร : Discord และ Line
4. การศึกษาเนื้อหาในการทำงาน :
Mycourseville
5. การวิเคราะห์ข้อมูลขนาดใหญ่ :
Python และ R-Studio

my
CourseVille



Discord



บทที่ 2

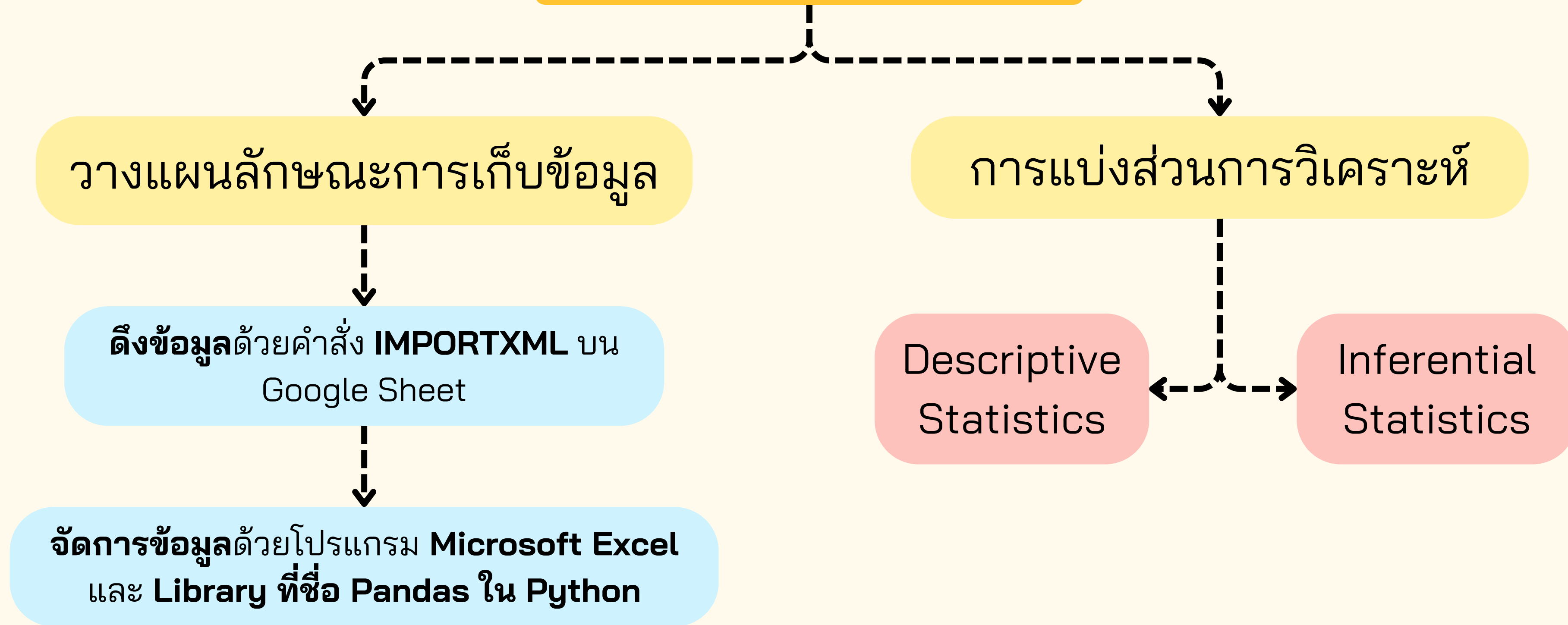
วิธีการ ดำเนินงาน

- Timeline
- การวิเคราะห์ข้อมูล

Timeline

ลำดับที่	รายการ	กุมภาพันธ์				มีนาคม				เมษายน			
		1	2	3	4	1	2	3	4	1	2	3	4
1	กำหนดหัวข้อโครงการ												
2	รวบรวมข้อมูลและจัดการข้อมูล												
3	วิเคราะห์สถิติเชิงพรรณนา												
4	วิเคราะห์สถิติเชิงอนุมาน												
5	เขียนรูปเล่มรายงาน												
6	ทำเอกสารประกอบการนำเสนอ												

การวิเคราะห์ข้อมูล



บทที่ 3

ขั้นตอนการ ดำเนินงาน

- การเก็บข้อมูล
- การจัดการข้อมูล
- การออกแบบการวิเคราะห์ข้อมูล

การเก็บข้อมูล

ดึงข้อมูลด้วยคำสั่ง **IMPORTXML** บน
Google Sheet

importxml เป็นการนำเข้าข้อมูลจากภายนอกเข้ามาใน google sheet โดยใช้ url

URL



=importxml("url")



=importxml("url")

เก็บและเรียบเรียงข้อมูลใน google
sheets

การจัดการข้อมูล

จัดการข้อมูลด้วยโปรแกรม Microsoft Excel
และ Library ที่ชื่อ Pandas ใน Python

ใช้ Library Pandas ในการตัดข้อความส่วนเกินออก เช่น
views และ Like this video along with ... people

ใช้ microsoft excel เพื่อสำหรับในการจัดเก็บและจัดเรียงข้อมูล

ตัวอย่างข้อมูล

Viewcount	Likecount	Duration	View	Like	Duration (Min)
10000 views	Like this video along with 100 other people	40:30	10000	100	40.5

จัดประเภทและหมวดหมู่ของวิดีโอ

Business

Food delivery
ประกัน
ธุรกิจค้าปลีก
ธุรกิจเครื่องดื่ม
SMEs
E-commerce

Economic & Logistic

ธนาคารและการเงิน
Logistics
สินเชื่อ

Education & Life style

การศึกษา
Entertainment
Health & Wellness
Life style
Fashion
การท่องเที่ยว

Government

Product & Service

รถยนต์
เอเจนซี่และสื่อโฆษณา
สินค้าอุปโภคบริโภค
โทรคมนาคม
พลังงาน
และอื่นๆ

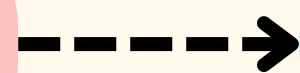
Others

การออกแบบการ วิเคราะห์ข้อมูล

สถิติเชิงพรรณนา
(Descriptive Statistics)

สถิติเชิงอนุมาน
(Inferential Statistics)

สถิติเชิงพรรณนา



มีทั้งหมด 8 เรื่องดังต่อไปนี้

จำนวนคลิป

- จำนวนคลิปที่ถูกอัปโหลดในแต่ละปี ตลอดช่วงปี ค.ศ. 2018 ถึง ค.ศ. 2023
- จำนวนคลิปแต่ละหมวดหมู่ที่ถูกอัปโหลดในแต่ละปี

จำนวนคนดู

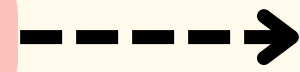
- ค่าเฉลี่ยของจำนวนคนดูโดยแบ่งตามหมวดหมู่*ของ podcast
- ค่าเฉลี่ยของคนดูโดยแบ่งตามประเภทสื่อ
- ค่าเฉลี่ยของคนดูโดยแบ่งตาม Keyword ที่พบบ่อยในชื่อคลิป 9 อันดับแรก
- การกระจายตัวระหว่างจำนวนคนดูและจำนวนคนกดถูกใจ หรือ สัดส่วนการกดถูกใจ

ความยาวคลิป

- สัดส่วนความยาวคลิปต่อไตรมาส
- การกระจายตัวของความยาวคลิปตลอดปี ค.ศ. 2018 ถึง ค.ศ. 2023

หมวดหมู่เป็นไปตามบทที่ 3*

สถิติเชิงอนุมาน



มีทั้งหมด 3 เรื่อง

- การทดสอบความเป็นอิสระต่อกันระหว่างหมวดหมู่*ของ podcast และความยาวคลิป
- การวิเคราะห์ความแปรปรวนระหว่างหมวดหมู่ podcast ในแต่ละปี
- สมการ **regression** เพื่อแสดงความสัมพันธ์ระหว่างจำนวนคนดูและคนกดถูกใจ

หมวดหมู่เป็นไปตามบทที่ 3*

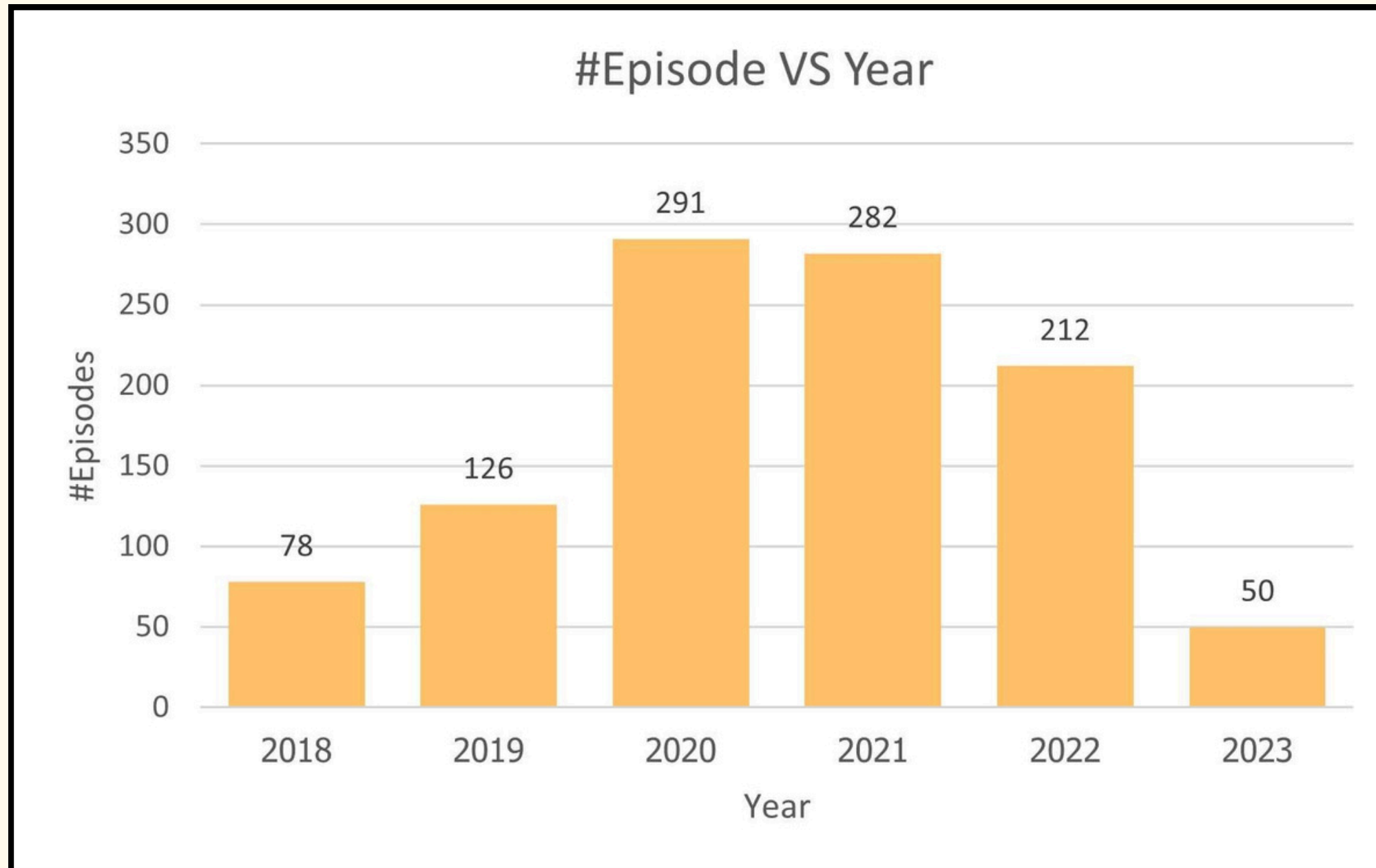
บทที่ 4

การวิเคราะห์ ทางสถิติ

- การวิเคราะห์ด้วยสถิติเชิงพรรณนา
- การวิเคราะห์ด้วยสถิติเชิงอนุมาน

การวิเคราะห์ด้วยสถิติเชิงพรรณนา

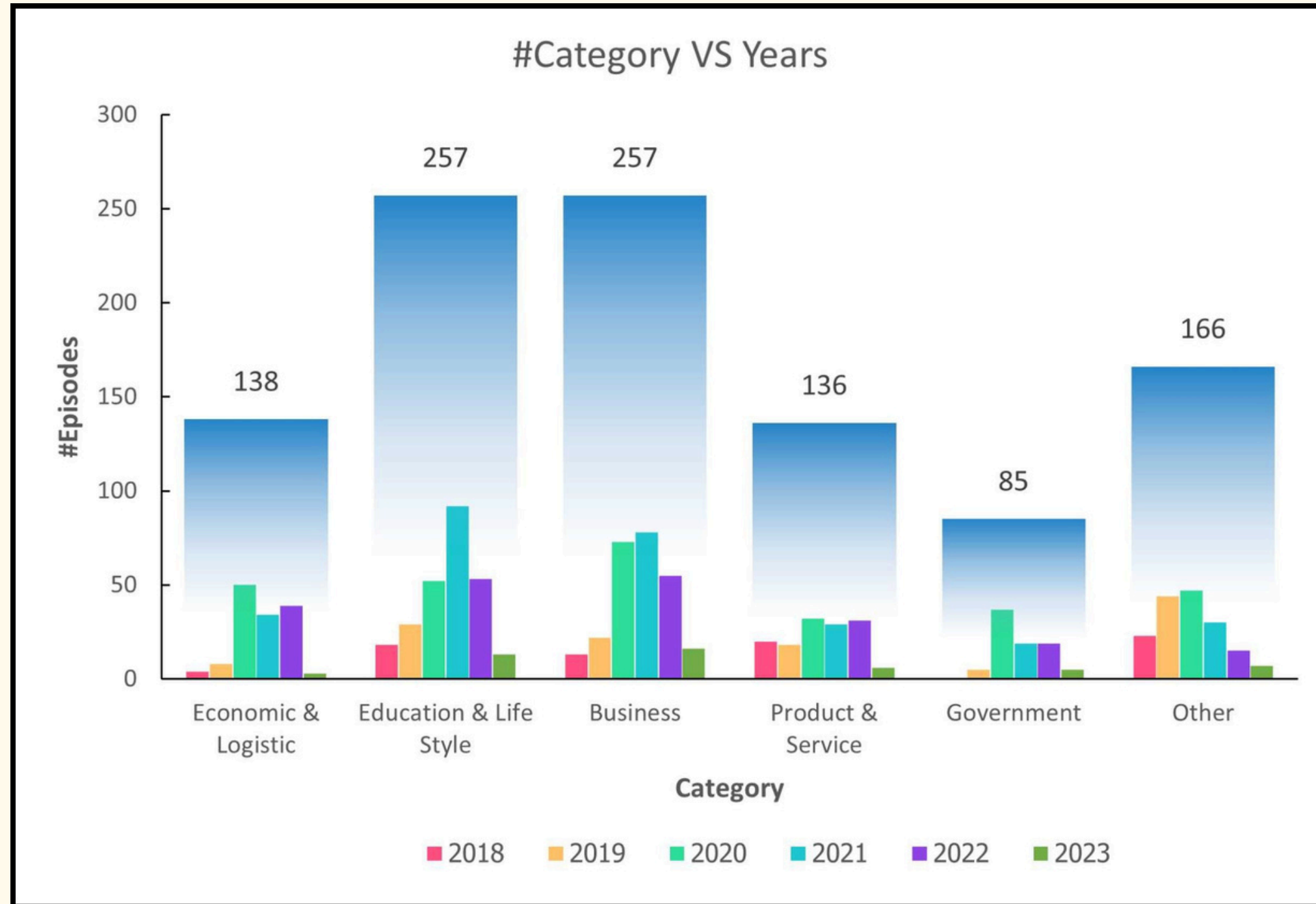
จำนวนคลิปที่ถูกอัปโหลดในแต่ละปีตลอดปี ค.ศ. 2018 ถึง ค.ศ. 2023



จากแผนภูมิแท่งสังเกตได้ว่า **จำนวนที่อัปโหลดจะค่อยๆเพิ่มมากขึ้นในช่วงหลัง** โดยทางผู้จัดทำคาดว่าในช่วงปีค.ศ. 2020 และ 2021 นั้นเป็นช่วงที่เกิดการแพร่ระบาดของไวรัส COVID-19 ทำให้การอัปโหลดคลิปเพื่อนำเสนอข้อมูลが多มากขึ้น

การวิเคราะห์ด้วยสถิติเชิงพรรณนา

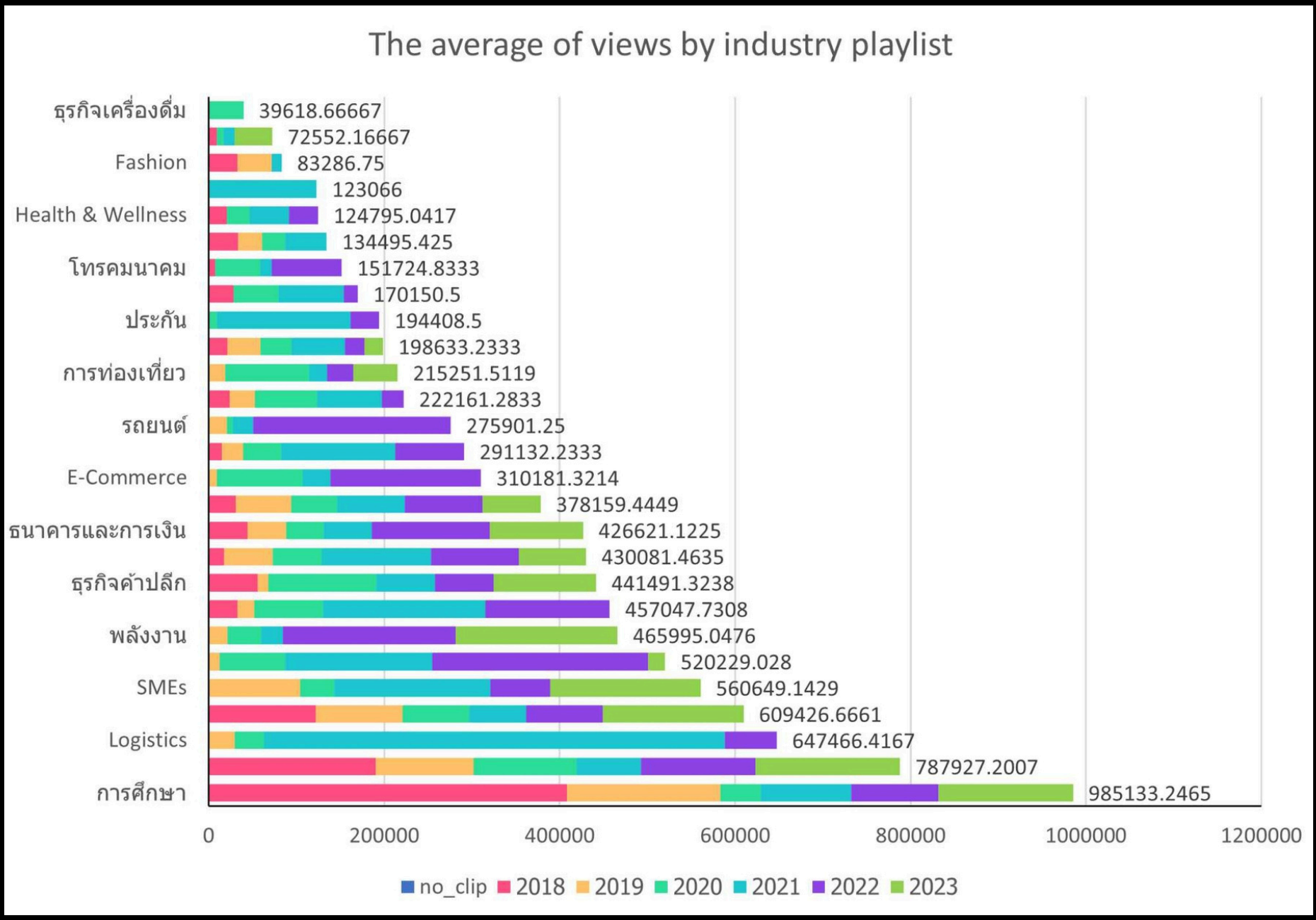
จำนวนคลิปแต่ละหมวดหมู่*ที่ถูกอัปโหลดมาในแต่ละปี



จากแผนภูมิแท่งสังเกตได้ว่า **หมวดหมู่ของวิดีโอมีการเปลี่ยนแปลงตลอดในทุกๆปี** เนื่องจากความนิยม หรือกระแสของเนื้อหาในช่วงนั้น โดยเนื้อหาที่มีจำนวนเยอะที่สุด ได้แก่ Education & Life Style และ Business ซึ่งก็สอดคล้องตามจุดประสงค์ของรายการ The Secret Sauce

การวิเคราะห์ด้วยสถิติเชิงพรรณนา

ค่าเฉลี่ยของจำนวนคนดูแบ่งตามประเภทสื่อ*

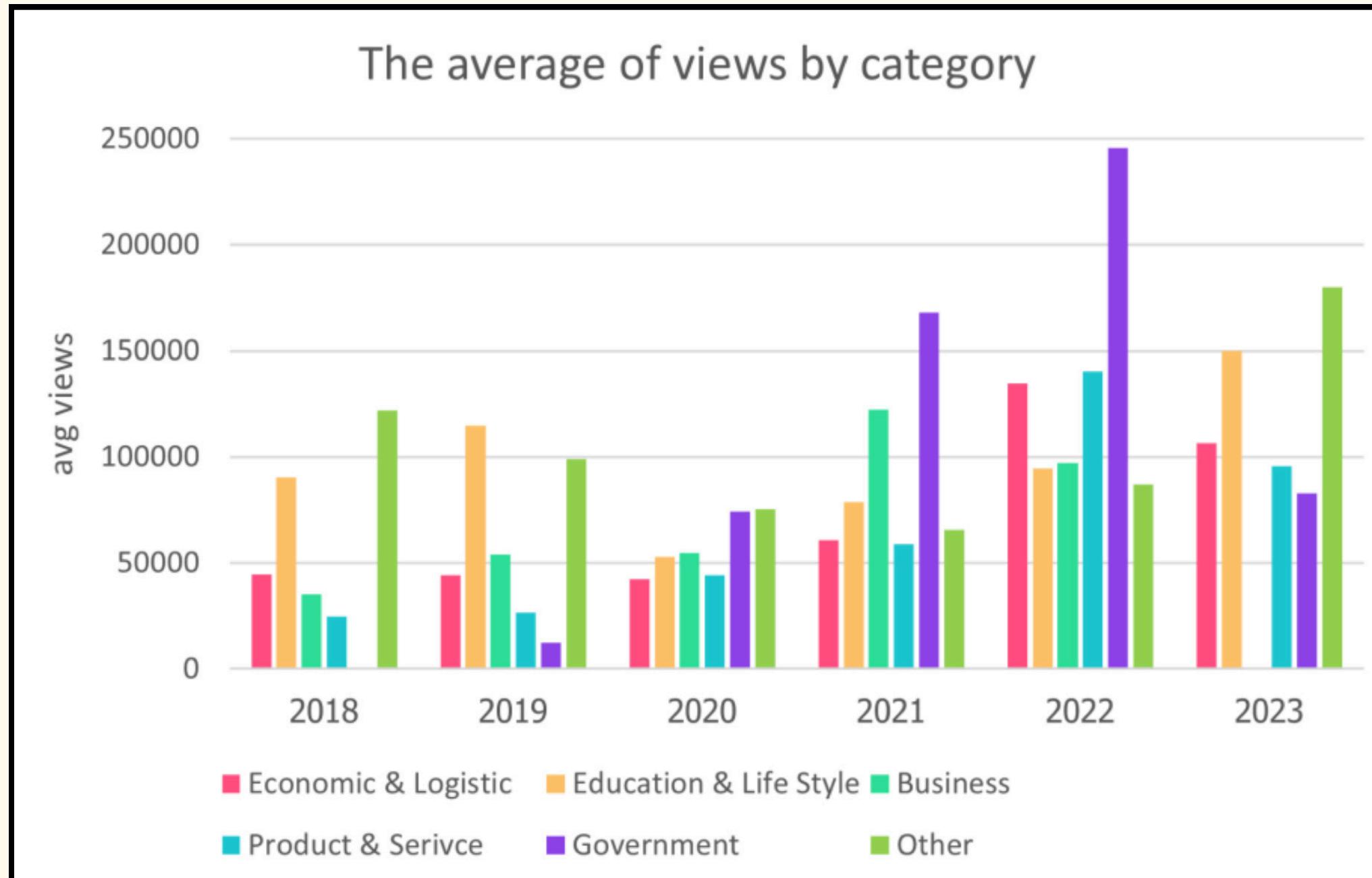


จากแผนภูมิแท่งสังเกตได้ว่า หมวดหมู่ที่มีจำนวนค่าเฉลี่ยของคนดูมากที่สุดคือ “การศึกษา” และสามารถแสดงให้เห็นได้ว่า **ในแต่ละปีจำนวนเฉลี่ยของคนดูในแต่ละหมวดค่อนข้างไม่คงที่** ทั้งนี้เป็นเพราะกระแสและความนิยมในช่วงนั้นด้วย

ประเภทสื่อเป็นไปตามที่ทาง The Secret Sauce เป็นคนแยก*

การวิเคราะห์ด้วยสถิติเชิงพรรณนา

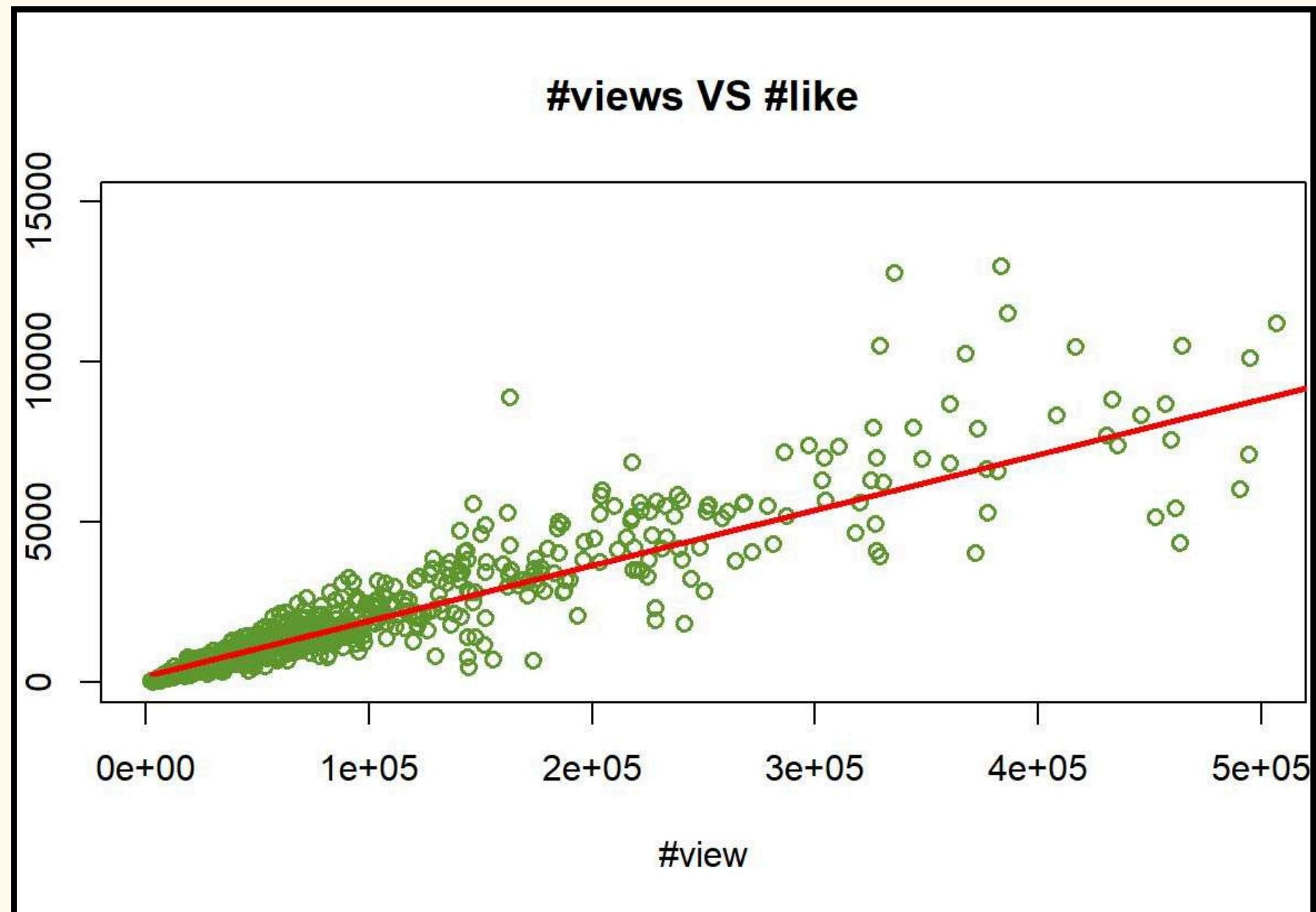
ค่าเฉลี่ยของจำนวนคนดูแบ่งตามหมวดหมู่*



จากแผนภูมิแท่งสังเกตได้ว่า **จำนวนคนดูวิดีโอโดยเฉลี่ยในแต่ละหมวดหมู่มีแนวโน้มไปในทิศทางเดียวกัน**

การวิเคราะห์ด้วยสถิติเชิงพรรณนา

การกระจายระหว่างจำนวนคนดูและจำนวนคนกดถูกใจ



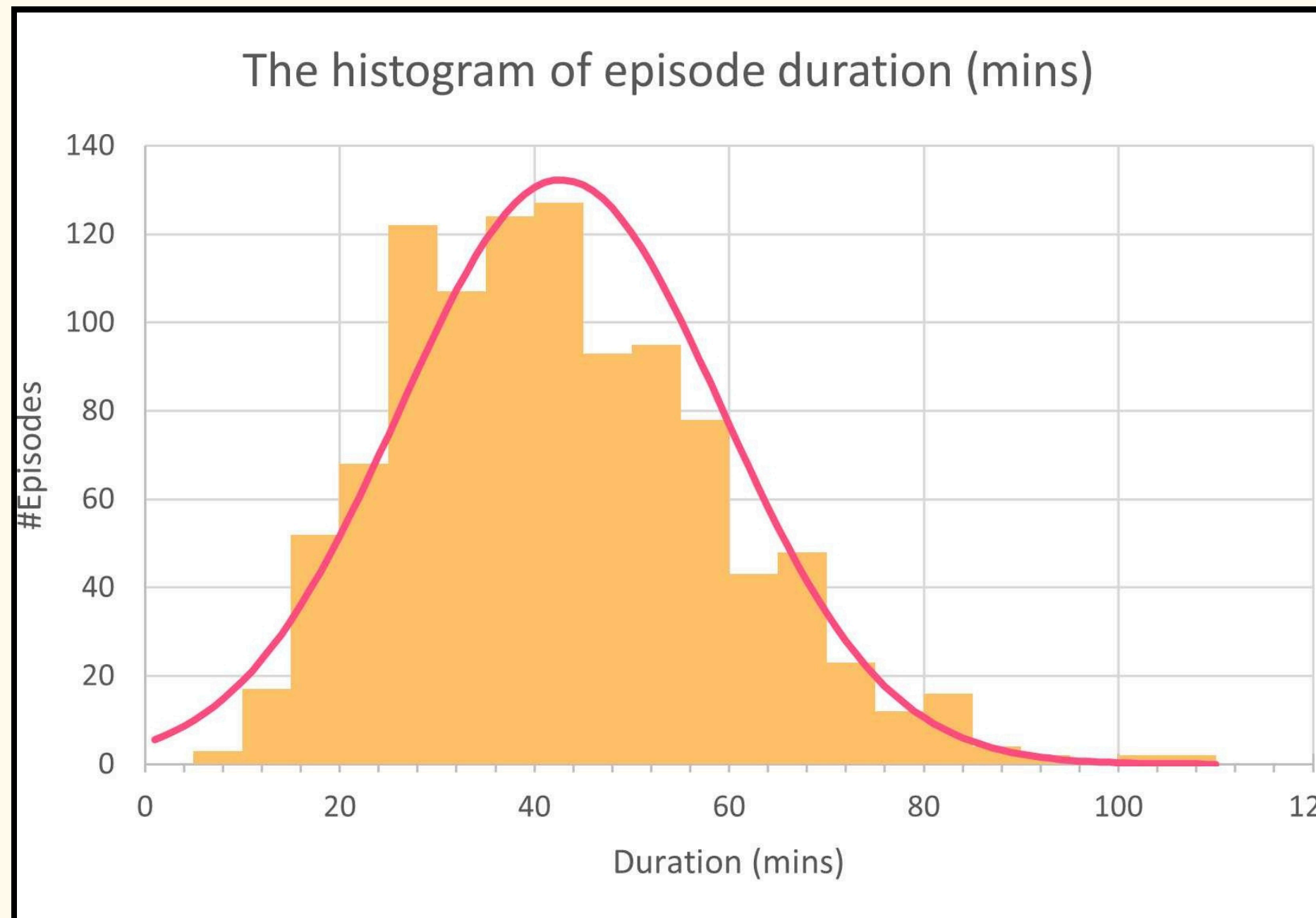
จาก Scatter plot จะสังเกตได้ว่า

ช่วง 0-200,000 คนดูแรกจะมีความสัมพันธ์
ในทางบวก

แต่ถ้า มากกว่า 200,000 คนนั้นจะมี
การกระจายที่ไม่เป็นเส้นตรง

การวิเคราะห์ด้วยสถิติเชิงพรรณนา

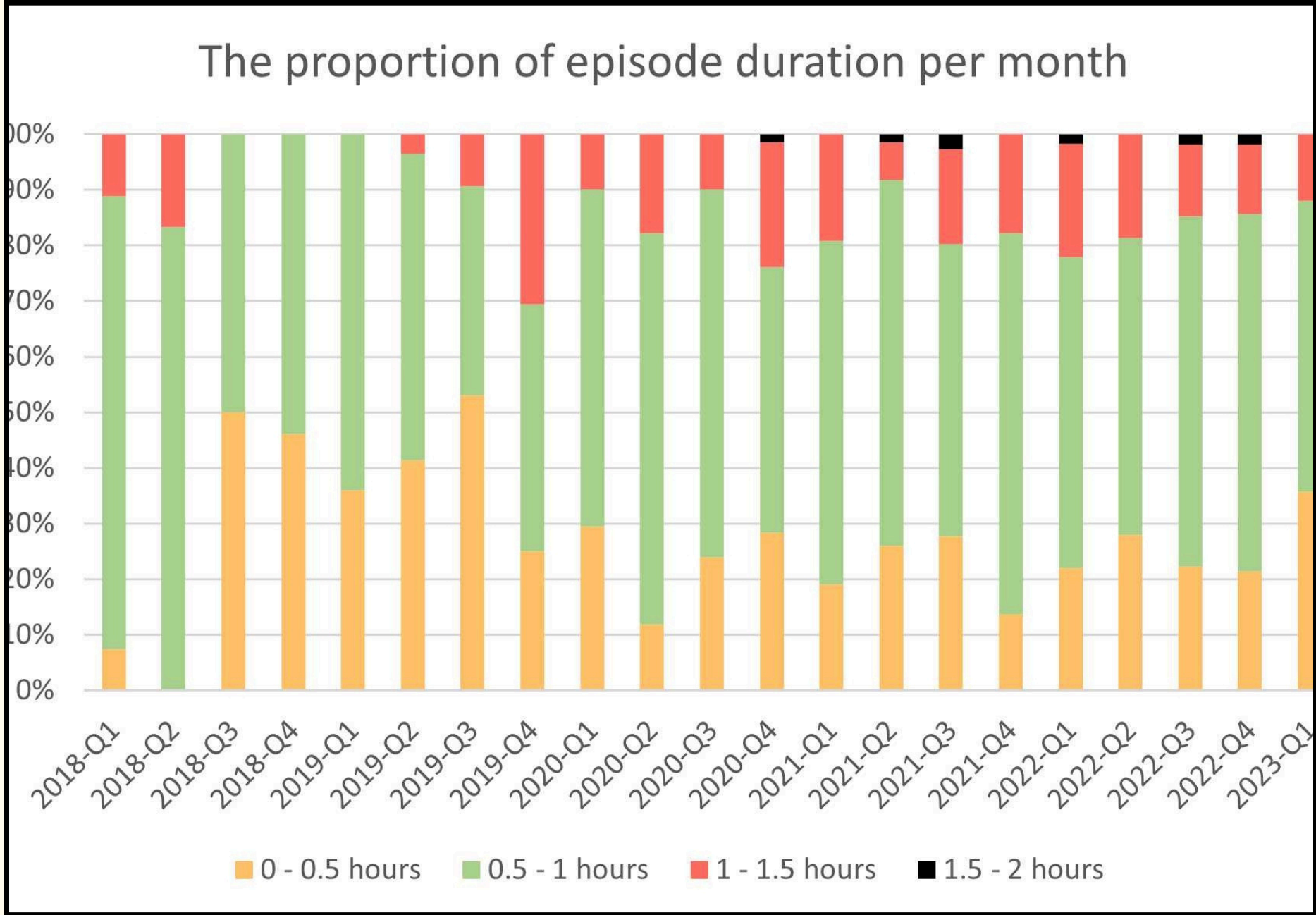
การกระจายตัวของความยาววิดีโอทั้งหมด



จากแผนภูมิ Histogram สังเกตได้ว่า ข้อมูล
มีลักษณะเบ้ขวา โดยพบจำนวนวิดีโอช่วง
30-45 นาทีเป็นส่วนมาก

การวิเคราะห์ด้วยสถิติเชิงพรรณนา

สัดส่วนความยาววิดีโอต่อไตรมาส



จากแผนภูมิแท่งสังเกตได้ว่าวิดีโอส่วนใหญ่
ในช่องจะมี**ความยาววิดีโอในช่วงความยาว
30-60 นาทีมากที่สุด** และได้ความยาว
เฉลี่ยวิดีโอคือ 42.45 นาที

การวิเคราะห์ด้วยสถิติเชิงอนุมาน

การทดสอบความเป็นอิสระต่อกันระหว่าง หมวดหมู่ podcast และความยาวคลิป

ทดสอบด้วย Chi-Square (Test for independence with Chi-Square Test)

Duration Range	Category						Total _i
	เศรษฐกิจ	Life Style	ธุรกิจและการค้า	สินค้าและบริการ	การเมือง	อื่น ๆ	
0.5 - 1 hours	80.8874	150.6381	150.6381	79.7151	49.8219	97.2993	609
0 - 0.5 hours	33.2050	61.8383	61.8383	32.7238	20.4524	39.9423	250
1 - 1.5 hours	15.5399	28.9403	28.9403	15.3147	9.5717	18.6930	117
1.5 - 2 hours	8.3677	15.5833	15.5833	8.2464	5.1540	10.0654	63
Total _j	138	257	257	136	85	166	1039

H0 : หมวดหมู่ podcast ไม่มีความสัมพันธ์
กับความยาวคลิป
Ha : หมวดหมู่ podcast มีความสัมพันธ์กับ
ความยาวคลิป
alpha = 0.05
Test Statistic = 39.6336
p-value = 0.000515 < 0.05

REJECT H0

สรุปได้ว่า ข้อมูลระหว่างหมวดหมู่ podcast และความยาวคลิปมีความสัมพันธ์กัน

การวิเคราะห์ด้วยสถิติเชิงอนุมาน

การวิเคราะห์ความแปรปรวนระหว่าง หมวดหมู่ podcast ในแต่ละปี

ใช้ ANOVA : Two ways without Replication ผ่าน Data analysis ใน Microsoft excel

สรุปข้อมูล จำนวนคนดูโดยเฉลี่ยในรูปแบบของตารางตามหมวดหมู่ ได้ต่อไปนี้

Year	Category					
	เศรษฐกิจ	Life Style	ธุรกิจ	สินค้า	การเมือง	อื่น ๆ
			และการค้า	และบริการ		
2018	44439.75	90310.89	35296.38	24706.20	0.00	121893.61
2019	44093.43	114938.83	54039.95	26648.28	12474.67	99039.98
2020	42410.36	52849.45	54842.54	44262.52	74284.73	75540.48
2021	60689.53	78599.79	122352.09	58778.70	168037.21	65434.67
2022	134593.49	94336.11	96957.07	140157.48	245826.42	86904.27
2023	106431.00	150040.54	95540.31	82697.50	19606.00	160613.67

SUMMARY		Count	Sum	Average	Variance	
	2018	6	316646.8	52774.47	20268723	
	2019	6	351235.1	58539.19	16378253	
	2020	6	344190.1	57365.01	20773892	
	2021	6	553892	92315.33	19343185	
	2022	6	798774.8	133129.1	35404550	
	2023	6	614929	102488.2	25929489	
วิชาชีพ		6	432657.6	72109.59	15290367	
Life Style		6	581075.6	96845.94	10949273	
ธุรกิจและการค้า		6	459028.3	76504.72	11109353	
สินค้าและบริการ		6	377250.7	62875.11	18995901	
การเมือง		6	520229	86704.84	99237986	
อื่น ๆ		6	609426.7	101571.1	12207973	
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Year (B)	3.09E+10	5	6.18E+09	2.913768841	0.033095	2.602987
Category (T)	6.69E+09	5	1.34E+09	0.631514222	0.677454	2.602987
Error	5.3E+10	25	2.12E+09			
Total	9.06E+10	35				

ตั้งสมมติฐาน

สำหรับหมวดหมู่

$H_0 : u_{1t} = u_{2t} = u_{3t} = u_{4t} = u_{5t} = u_{6t}$
 $H_a : H_0$ เป็นเท็จ

สำหรับปี

$H_0 : u_{1b} = u_{2b} = u_{3b} = u_{4b} = u_{5b} = u_{6b}$
 $H_a : H_0$ เป็นเท็จ

alpha = 0.05

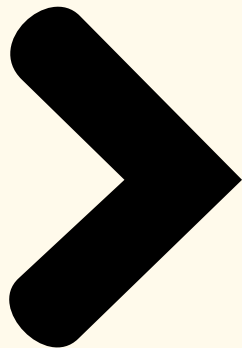
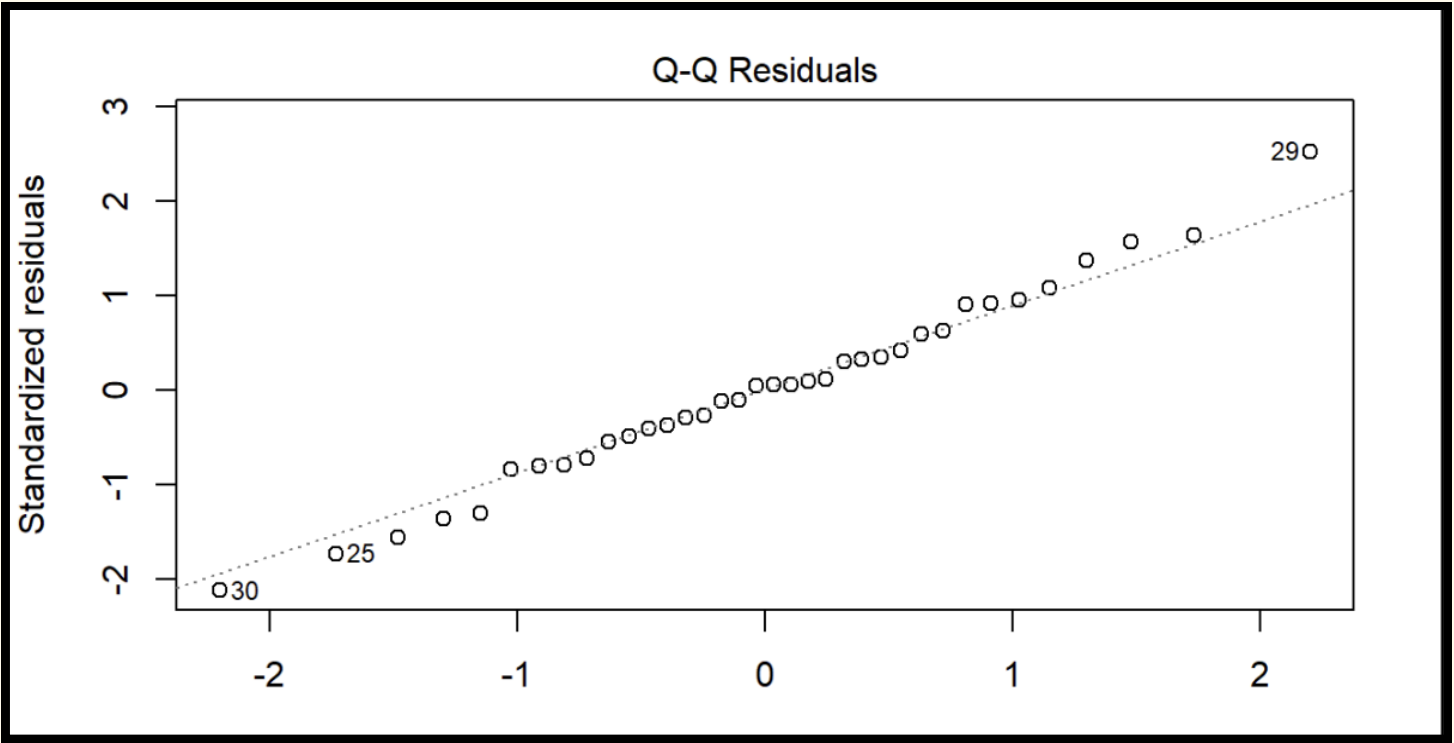
p-value(หมวดหมู่) = 0.033095 < 0.05

p-value(ปี) = 0.677454 > 0.05

สรุปได้ว่า หมวดหมู่ของ podcast ไม่ส่งผลต่อจำนวนคนดูโดยเฉลี่ย

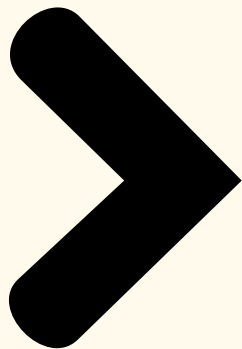
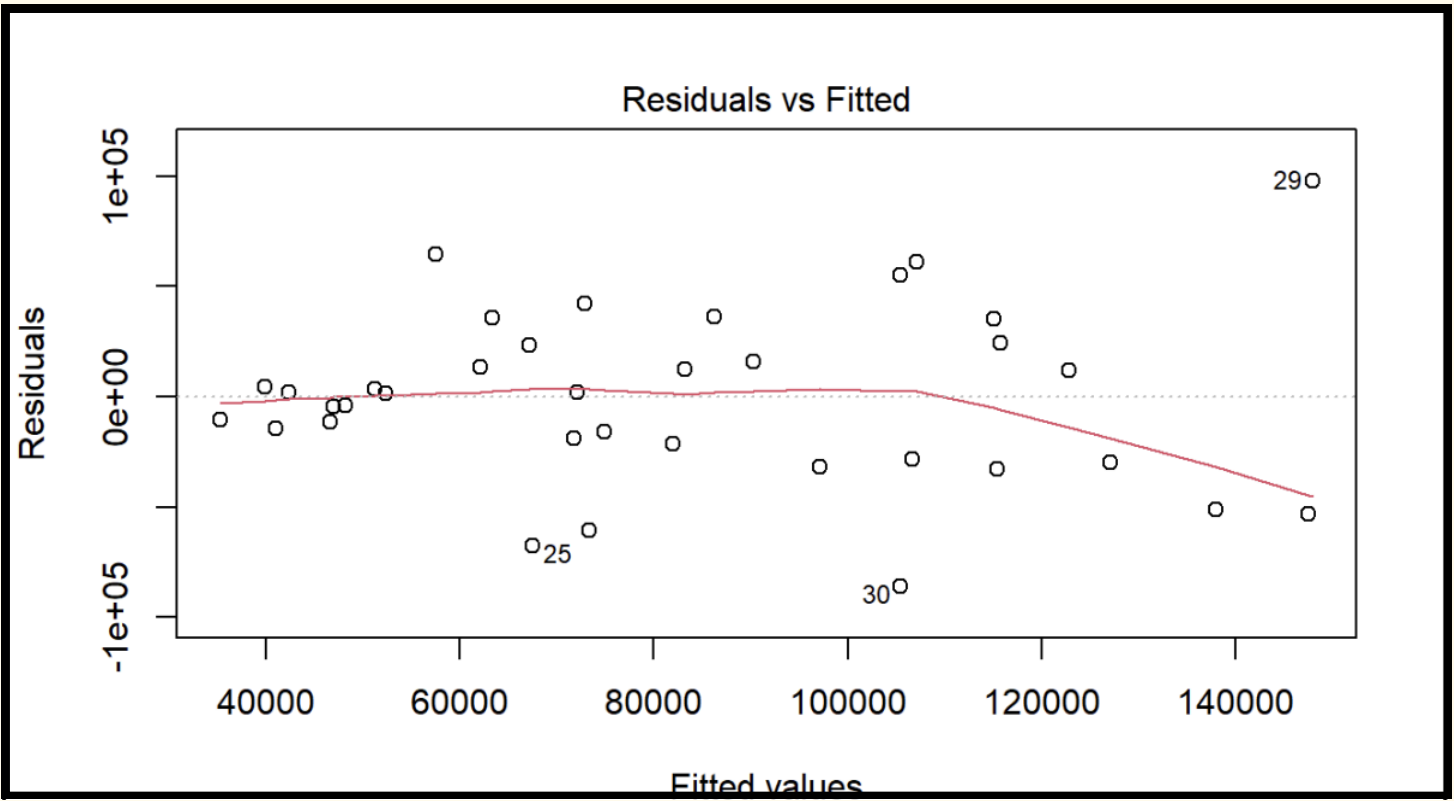
ทดสอบข้อบังคับ ANOVA

Normal Probability Plot of Residuals



สามารถอนุมานได้ว่า **เป็นไปตาม Normal**
เนื่องจากข้อมูลเกาะบริเวณเส้น

Plot of Residuals vs Fitted value

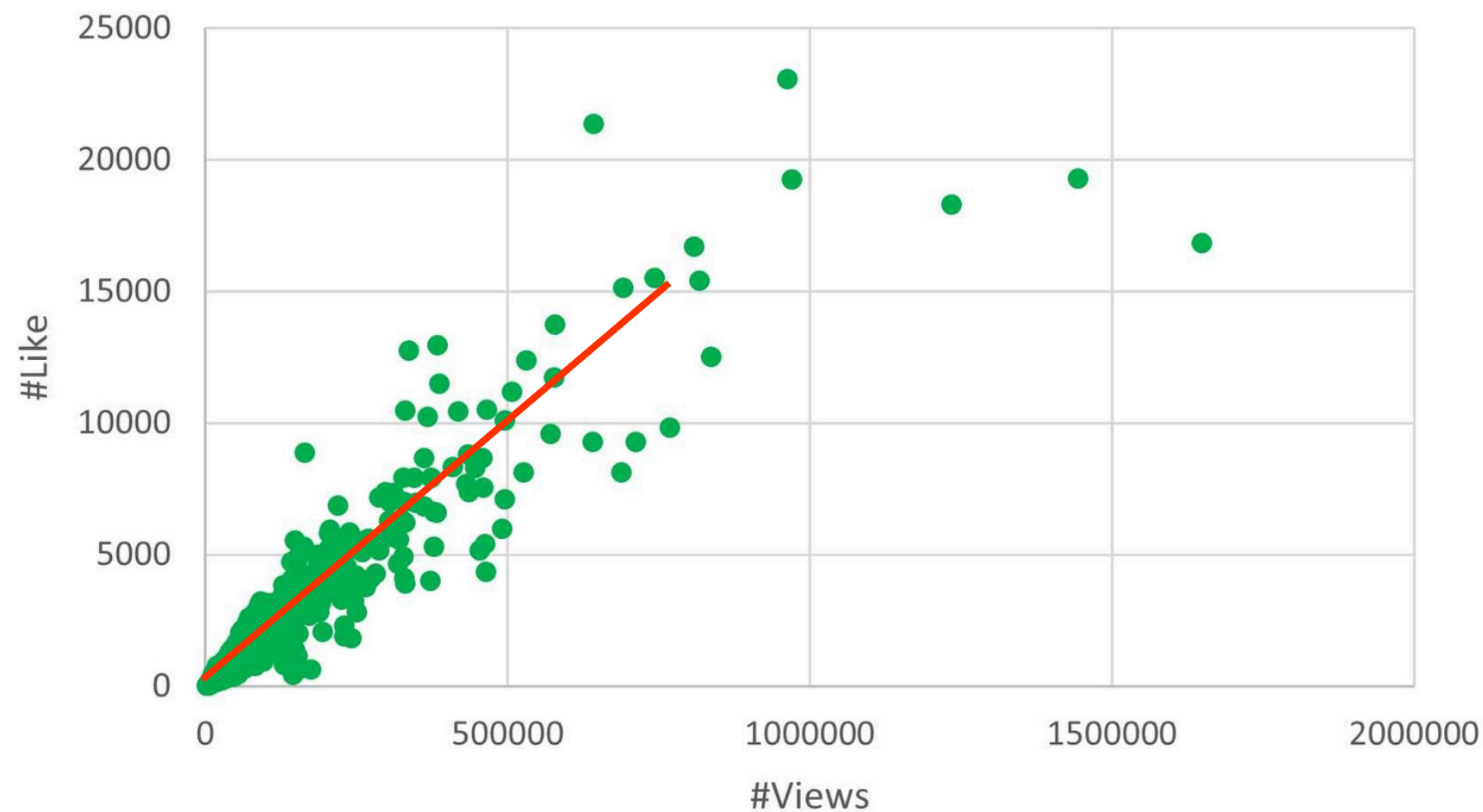


มีค่า Error เฉลี่ยไม่เป็น 0
เนื่องจากเส้นมีการลดลงในช่วงหลัง

การวิเคราะห์ด้วยสถิติเชิงอนุมาน

สมการ Regression ระหว่าง จำนวนคนดูกับจำนวนคนกดถูกใจ

Scatter plot of views vs. likes



$$y = Ax + B$$

กำหนดให้ y = จำนวนคนกดถูกใจต่อคลิป

x = จำนวนคนดูวิดีโอต่อคลิป

A, B เป็นค่าคงที่

คำนวณหา A, B ด้วย LSM

$$A = 207.199$$

$$B = 0.017268$$

จะได้สมการ Regression คือ $y = 207.199 + 0.017268x$

INDIVIDUAL TEST A , B (TWO TAILED TEST)

A

$H_0 : \alpha = 0$

$H_a : H_0$ เป็นเท็จ

significant level = 5%

$t_{-ob} = 5.9697$

$df = 1037$

$p\text{-value} = 3.26 \times 10^{-9} < 0.05$

Reject H_0

B

$H_0 : \beta = 0$

$H_a : H_0$ เป็นเท็จ

significant level = 5%

$t_{-ob} = 5.9697$

$df = 1037$

$p\text{-value} = 0 < 0.05$

Reject H_0

เนื่องจาก α และ β ไม่เท่ากับ 0 เพราะฉะนั้นแบบจำลองมีความเหมาะสม

บทที่ 5

สรุปผล

สรุปและอภิปรายผล

2

จากการทดสอบไคสแควร์ระหว่าง**หมวดหมู่วิดีโอและความยาววิดีโอ** พบว่าทั้งสอง**มีความสัมพันธ์กัน**

4

หมวดหมู่ในแต่ละหมวดหมู่**ไม่ได้ส่งผลต่อจำนวนคนดู**โดยเฉลี่ย และพบว่าในแต่ละหมวดหมู่จะมีทิศทางหรือแนวโน้มของจำนวนคนดูโดยเฉลี่ยไปในทิศทางเดียวกันในแต่ละปี

1

จำนวนคนดูวิดีโอและจำนวนคนกดไลค์มีลักษณะไปในทิศทางเดียวกัน เนื่องจากสมการถดถอยเชิงเส้นอย่างง่ายที่แสดงถึงความสัมพันธ์ดังกล่าว มีโอกาสเป็นเส้นตรงที่ R-square มีค่า 86.72%

3

ความยาววิดีโออยู่ในช่วง 30 ถึง 60 นาทีเป็นส่วนใหญ่ ถัดลง. มาเป็นความยาว 0 ถึง 30 นาที โดยมี**ค่าเฉลี่ย**ของความยาววิดีโอ **คิดเป็น 42.45 นาที**

5

คนไทยส่วนมากให้ความสนใจกับเนื้อหาประเภท **การศึกษา Life Style** และ **Logistics** ตามลำดับ

A large, solid yellow circle is centered on the page, set against a solid orange background. The circle is partially obscured by the text 'Thank you' which is written across it.

Thank you