



# บทที่ 5

บทนำสู่การวิเคราะห์ความรู้สึก

# การวิเคราะห์ความรู้สึกคืออะไร

- การวิเคราะห์ความรู้สึก หรือที่เรียกว่า **การขุดค้นความคิดเห็น** เป็นกระบวนการทำความเข้าใจความคิดเห็นของผู้เขียนเกี่ยวกับหัวข้อใดหัวข้อหนึ่ง

# ระบบวิเคราะห์ความรู้สึกประกอบด้วยอะไรบ้าง?

- 3 องค์ประกอบในระบบการวิเคราะห์ความรู้สึกซึ่งโดยปกติจะขึ้นอยู่กับบริบท:

- อันดับแรก: ความคิดเห็น / อารมณ์

- ความคิดเห็น (ข้อตรงข้าม) อาจจะเป็นเชิงบวก เป็นกลาง หรือเชิงลบก็ได้



- อารมณ์อาจเป็นเชิงคุณภาพ (เช่น ความสุข ความประหลาดใจ หรือความโกรธ) หรือเชิงปริมาณ (เช่น การให้คะแนน ภาพยนตร์ในระดับจาก 1 ถึง 10)



# ระบบวิเคราะห์ความรู้สึกประกอบด้วยอะไรบ้าง?

- ที่สอง: หัวข้อ
  - หัวข้อที่กำลังพูดถึงคืออะไร? • เช่น หนังสือ ภาพยนตร์ หรือผลิตภัณฑ์ • ตัวอย่างเช่น: "กล้องของโทรศัพท์รุ่นนี้ดีมาก แต่แบตเตอรี่กลับหมดเร็ว"
- ที่สาม: ผู้ถือความเห็น (บุคคลหรือองค์กรที่อ้างความเห็น)
  - โดยใคร?

# เหตุใดจึงต้องวิเคราะห์ความรู้สึก?

- การวิเคราะห์ความรู้สึกมีการประยุกต์ใช้ในทางปฏิบัติมากมาย เช่น:
  - การติดตามโซเชียลมีเดีย:
    - เราไม่เพียงแต่ต้องการรู้ว่าผู้คนกำลังพูดถึงแบรนด์หรือไม่ แต่เราต้องการทราบว่าพวกเขาพูดถึงแบรนด์อย่างไร • โซเชียลมีเดียไม่เพียงแต่เป็นแหล่งข้อมูลเท่านั้น เรายังสามารถค้นหาความคิดเห็นจากบล็อกและข่าวสารได้อีกด้วย • การติดตามแบรนด์:
      - แบนด์ส่วนใหญ่วิเคราะห์แหล่งข้อมูลทั้งหมดเหล่านี้เพื่อเพิ่มความเข้าใจของพวกเขา
        - ลูกค้าโต้ตอบกับแบรนด์ของตนอย่างไร • พวกเขาพอใจหรือไม่พอใจเกี่ยวกับอะไร
  - การวิเคราะห์ความรู้สึกจึงมีความสำคัญอย่างยิ่งในการติดตามแบรนด์ และในสาขาต่างๆ เช่น การวิเคราะห์ลูกค้าและผลิตภัณฑ์ การวิจัยและวิเคราะห์ตลาด

# มาดูรีวิวนั่งกันดีกว่า!

<https://www.imdb.com/>

- ชุดข้อมูล: ตัวอย่างบทวิจารณ์ภาพยนตร์จาก IMDB ที่มี 2 คอลัมน์ โดยคอลัมน์หนึ่งเป็นข้อความของบทวิจารณ์ และอีกคอลัมน์หนึ่งเรียกว่า "ป้ายกำกับ" ซึ่งแสดงความรู้สึกโดยรวม:
  - หมวดยกหรือขึ้น 1 หมายถึงเชิงบวก และ 0 หมายถึงเชิงลบ

## ป้ายข้อความ (รีวิว)

ข้อมูล.หัว()

This short spoof can be found on Elite's Mille...	0
A singularly unfunny musical comedy that artif...	0
An excellent series, masterfully acted and dir...	1
The master of movie spectacle Cecil B. De Mill...	1
I was gifted with this movie as it had such a ...	0

# มีวิธีวิวเชิงบวกและเชิงลบที่รายการ?

- เรียกใช้เมธอด `.value_counts()` ในคอลัมน์ "label":
  - ข้อมูล.ป้าย.ค่า\_นับ()
    - 0: จำนวนบทความเชิงลบ • 1: จำนวนบทความเชิงบวก

```
จำนวนการวิจัยเชิงบวกและเชิงลบ: 0 20019
```

```
1 19981
```

```
ชื่อ: ป้าย, dtype: int64
```

# เปอร์เซ็นต์ของบทความเชิงบวกและเชิงลบ

- เพื่อดูจำนวนผลบวกและผลลบเป็นเปอร์เซ็นต์

- หาวิธีพจน์ด้วยจำนวนแถวที่ได้รับด้วย

วิธี `len()` •

`data.label.value.count() / len(data)`

```
สัดส่วนของบทความเชิงบวกและเชิงลบ: 0 0.500475
```

```
1 0.499525
```

```
ชื่อ: ป้าย, dtype: float64
```



# บทวิจารณ์ที่ยาวที่สุดใช้เวลาานที่สุดเท่าไร?

- เลือกคอลัมน์การตรวจสอบของชุดข้อมูล ตามด้วย `.str.len()`

- `str` เป็นคำย่อของสตริง • เรียก

ใช้ฟังก์ชันสตริงเพื่อแปลงชุดบทวิจารณ์ให้เป็นสตริง

ความยาว\_บทวิจารณ์ = ภาพยนตร์.ข้อความ.str.len()

ประเภท(ความยาว\_บทวิจารณ์)

```
<คลาส 'pandas.core.series.Series'>
```

- ค้นหาข้อความวิจารณ์ที่มีความยาวสูงสุด

สูงสุด(ความยาว\_รีวิว)

การหาความยาวของรีวิวมีความสำคัญอย่างไร?

มาฝึกกันเถอะ!

# มีบทวิจารณ์เชิงบวกและเชิงลบกี่จำนวน?

- เป็นขั้นตอนแรกของงานวิเคราะห์ความรู้สึก คล้ายกับงานวิทยาศาสตร์ข้อมูลอื่นๆ  
ปัญหา • เพื่อ  
สำรวจชุดข้อมูลในรายละเอียดเพิ่มเติม

- ขั้นแรก โหลดชุดข้อมูล

นำเข้า แพนด้า เป็น pd

ภาพยนตร์ = pd.read\_csv('.\ch5\train.csv')

# มีบทวิจารณ์เชิงบวกและเชิงลบกี่จำนวน?

- ค้นหาจำนวนบทวิจารณ์เชิงบวกและเชิงลบในชุดข้อมูลภาพยนตร์

```
# ค้นหาจำนวนบทวิจารณ์เชิงบวกและเชิงลบ print('จำนวนบทวิจารณ์เชิงบวกและ  
เชิงลบ: ', movies.label.value_counts())
```

- ค้นหาเปอร์เซ็นต์ของบทวิจารณ์เชิงบวกและเชิงลบในชุดข้อมูล

```
# ค้นหาสัดส่วนของการวิจารณ์เชิงบวกและเชิงลบ print('Proportion of positive and  
negative reviews: ', movies.label.value_counts() / len(movies))
```

# รีวิวที่ยาวที่สุดและสั้นที่สุด

- ตอนนี้ งานของคุณคือการสำรวจคอลัมน์บทวิจารณ์อย่างละเอียดมากขึ้น • ใช้คอลัมน์ข้อความของชุดข้อมูลภาพยนตร์เพื่อค้นหาความยาวของบทวิจารณ์ที่ยาวที่สุด

ความยาว\_บทวิจารณ์ = ภาพยนตร์.ข้อความ.str.len()

# การพิมพ์รีวิวจะยาวที่สุดนานแค่ไหน (max(length\_reviews))

# ประเภทและแนวทางการวิเคราะห์ความรู้สึก

# ระดับความละเอียด

- ระดับเอกสาร

- ดูวิธีผลิตภัณฑ์ทั้งหมด • ระดับประโยชน์

- หมายถึงการพิจารณาว่าความเห็นที่แสดงในแต่ละประโยชน์เป็นไป  
บวก ลบ หรือเป็นกลาง

- ระดับลักษณะ

- หมายถึงการแสดงความคิดเห็นเกี่ยวกับคุณสมบัติต่างๆ ของผลิตภัณฑ์ • ลองนึกถึงประโยชน์:

- "กล้องในโทรศัพท์รุ่นนี้ค่อนข้างดี แต่แบตเตอรี่

- ชีวิตมันน่าผิดหวัง" • มันแสดงความ

- คิดเห็นทั้งเชิงบวกและเชิงลบเกี่ยวกับโทรศัพท์และต้องการที่จะสามารถพูดได้  
คุณสมบัติใดของผลิตภัณฑ์ที่ถูกคำชอบและไม่ชอบ

# ประเภทของอัลกอริทึมการวิเคราะห์ความรู้สึก

- ตามกฎหรือคำศัพท์

- วิธีการนี้โดยทั่วไปจะมีรายการคำที่กำหนดไว้ล่วงหน้าพร้อมคะแนนความแรง ตัวอย่างเช่น ดีอาจเป็น +2, ดี +1, แย่มาก -3 และอื่นๆ
- อัลกอริทึมจะจับคู่คำจากพจนานุกรมกับคำในข้อความ แล้วรวมหรือเฉลี่ยคะแนน ตัวอย่างเช่น "วันนี้เป็นวันที่ดี" วันนี้: 0, เคยเป็น: 0, a: 0, ดี: +1, วัน: 0

วาเลนซ์รวม: +1

- แต่ละคำจะได้รับคะแนน และจะได้รับค่าความแรงรวม ดังนั้นเราจะได้ประโยคที่เป็นบวก
- ระบบอัตโนมัติซึ่งใช้การเรียนรู้ของเครื่องจักร
  - โดยทั่วไปจะจำลองเป็นปัญหาการจำแนกประเภทโดยใช้ข้อมูลทางประวัติศาสตร์บางส่วนที่มีความรู้สึกที่ทราบ • ทำนายความรู้สึกของข้อความชิ้นใหม่



# ประโยคมีค่าความยิ่งใหญ่เท่าไร?

- คำนวณคะแนนความสำคัญของข้อความโดยใช้ไลบรารี TextBlob ของ Python
- วัตถุ TextBlob เป็นเหมือนสตริง Python ซึ่งได้รับการรับรองมาตรฐานบางอย่าง  
ทักษะการประมวลผลภาษา

ข้อความ = 'วันนี้เป็นวันที่ดี'

my\_valence = TextBlob(ข้อความ)

ความรู้สึกของฉัน

ความรู้สึก (ขั้วตรงข้าม = 0.7, ทัศนคติ = 0.6000000000000001)

- คุณสมบัติความรู้สึกส่งคืนทUPLE: • องค์ประกอบแรกคือขั้ว (ขั้วความรู้สึกลับ) • วัดจาก [-1.0 ถึง 1.0] โดยที่ -1.0 เป็นค่าลบมาก 0 เป็นกลาง และ +1.0 เป็นค่าบวกมาก •

ตัวอย่างของเรา 'วันนี้เป็นวันที่ดี' มีอารมณ์เชิงบวกและจะมีความรู้สึกเชิงบวก

คะแนนขั้ว: 0.7 • องค์ประกอบที่

สองคือความเป็นอัตวิสัย

- measured from [0.0 to 1.0] where 0.0 is very objective (ความเห็นที่อยู่บนข้อเท็จจริงที่ทุกคนเห็นเหมือนกัน) and 1.0 is very subjective (ความเห็นในมุมมองของตนเองเป็นหลัก). So this example is subjective.

# อัตโนมัติหรือตามกฎหมาย?

การเรียนรู้อัตโนมัติ/ของเครื่องจักร

- ต้องอาศัยการมีข้อมูลประวัติที่ติดป้ายกำกับ • อาจต้องใช้  
เวลาสักพักในการฝึก • อาจมีประสิทธิภาพ  
มาก

ตามกฎหมาย/พจนานุกรม • อาศัยการสร้างกฎ

หรือพจนานุกรมด้วยตนเอง • คำต่างๆ อาจมีขั้วที่แตกต่างกันในบริบทที่  
แตกต่างกัน • สามารถทำได้

ค่อนข้างเร็ว

มาฝึกกันนะครับ

# การตรวจจับความรู้สึก

- เพื่อตรวจจับความรู้สึก รวมถึงข้อตรงข้ามและความคิดเห็นส่วนตัวของสตริงที่กำหนดโดยใช้วิธีการตามกฎเกณฑ์ดังกล่าวและไลบรารี TextBlob ใน Python
- นำเข้าแพ็คเกจที่จำเป็น
- สร้างวัตถุลบข้อความจากสตริงข้อความ
- พิมพ์ข้อและความเป็นอัตวิสัยออกมา

```
# นำเข้าแพ็คเกจที่จำเป็น
```

```
จาก textblob นำเข้า TextBlob
```

```
ข้อความ = "คุณสวยมาก"
```

```
# สร้างวัตถุ textblob blob_two_cities =  
TextBlob(text)
```

```
# พิมพ์ค่าอวยws print(blob_two_cities.sentiment)
```

```
ความรู้สึก (ข้อตรงข้าม = 0.85, อัตวิสัย = 1.0)
```

# การวิจารณ์หนังมีอารมณ์อย่างไร?

- นำเข้าฟังก์ชันที่จำเป็น

```
# นำเข้าแพ็คเกจที่จำเป็น
จาก textblob นำเข้า TextBlob
```

- อ่าน/โหลดบทวิจารณ์ภาพยนตร์ไททานิค  
(titanic.txt)

```
#อ่านไฟล์ TXT
f = เปิด(".\ch5\titanic.txt", "r")
ไททานิค = f.read()
```

- สร้างวัตถุเก็บข้อความจากสตริงไททานิค

```
# สร้างวัตถุ textblob blob_titanic =
TextBlob(titanic)
```

- พิมพ์ผลลัพธ์ของคุณสมบัติความรู้สึก  
ออกมา

```
# พิมพ์ค่ามันสัญญา print(blob_titanic.sentiment)
```

ความรู้สึก (ขั้วตรงข้าม = 0.2024748060772906, ัฒวิสัย = 0.4518248900857597)

# มาฝึกกัน II

# นำเข้าโมดูลสำหรับการวิเคราะห์ความรู้สึก

- โมดูล re จัดให้มีการดำเนินการสำหรับการจับคู่นิพจน์ทั่วไป ซึ่งมีประโยชน์สำหรับการค้นหารูปแบบและสตริง
- pandas เป็นหนึ่งในเครื่องมือโอเพ่นซอร์สที่ใช้กันอย่างแพร่หลายที่สุดสำหรับการจัดการข้อมูลและการวิเคราะห์
- matplotlib เป็นไลบรารีที่ใช้กันอย่างง่าย เป็นที่นิยม และครอบคลุมใน Python สำหรับการสร้างภาพข้อมูล
- nltk เป็นแพลตฟอร์มโอเพ่นซอร์สที่ครอบคลุมสำหรับการสร้างแอปพลิเคชันเพื่อประมวลผลข้อมูลภาษาของมนุษย์ มาพร้อมกับไลบรารีการประมวลผลข้อความอันทรงพลังสำหรับงานการประมวลผลภาษาธรรมชาติ (NLP) ทั่วไป เช่น การทำความเข้าใจ การแยกวิเคราะห์ การสร้างสเต็ม การแท็ก การสร้างไวยากรณ์ การจำแนกประเภท การให้เหตุผลเชิงความหมาย เป็นต้น
- พจนานุกรม VADER พร้อมคลาส SentimentIntensityAnalyzer ของ NLTK ใช้เพื่อกำหนดคะแนนความรู้สึกต่อความคิดเห็นแต่ละข้อในชุดข้อมูลสาริต  
  - Valence Aware Dictionary and Sentiment Reasoner (VADER) เป็นพจนานุกรมและกฎเกณฑ์ชุดเครื่องมือวิเคราะห์ความรู้สึกที่เน้นไปที่ความรู้สึกที่มีอยู่ในแอปพลิเคชันข้อความทั่วไป เช่น ความคิดเห็นออนไลน์ โพสต์บนโซเชียลมีเดีย และการตอบแบบสำรวจ

# นำเข้าโมดูลสำหรับการวิเคราะห์ความรู้สึก

- โหลดโมดูลที่จำเป็น • ดาวน์โหลด 'vader\_lexicon'

เพื่อวิเคราะห์ความรู้สึก • สร้าง pd สำหรับอ้างอิง pandas • สร้าง plt สำหรับอ้างอิงโมดูล

```
matplotlib import re import pandas as pd import
```

```
matplotlib.pyplot as plt import nltk from nltk.sentiment.vader
```

```
import
```

```
SentimentIntensityAnalyzer
```

```
nltk.download('vader_lexicon')
```



# นำเข้าไฟล์ข้อมูลสาริตและข้อความก่อนการประมวลผล

- ใช้เมธอด `read_excel` จาก `pandas` เพื่อโหลดไฟล์ข้อมูลอินพุตสาริตลงในข้อมูลเฟรมแพนด้า
- เพิ่มฟิลด์ `row_id` ใหม่ลงในข้อมูลเฟรมนี้โดยเพิ่มฟิลด์ดัชนีในตัว ฟิลด์ `row_id` นี้ทำหน้าที่เป็นคีย์เฉพาะสำหรับชุดข้อมูลนี้เพื่อระบุแถวเฉพาะ และจะใช้ในภายหลังในโค้ดสำหรับการผสานข้อมูลเฟรมสองเฟรม

```
#df = pd.read_excel (r'เส้นทางที่เก็บไฟล์ Excel\ชื่อไฟล์.xlsx') df = pd.read_excel (r".\\ch5\\TeamHealthRawDataForDemo.xlsx") # เพิ่มไฟล์
row_id ลงในข้อมูลเฟรม ซึ่งจะมีประโยชน์ในการรวมข้อมูลในภายหลัง df["row_id"] = df.index + 1 # พิมพ์ 10 แถวแรก พิมพ์
(df.head(10))
```

	Period	Manager	Team	Response	row_id
0	2019-Q1	Mgr 1	Team 1	We're a fun team that works well together and ...	1
1	2019-Q1	Mgr 1	Team 1	we have a sound and collaborative team focused...	2
2	2019-Q1	Mgr 1	Team 1	we work well as a team, we have fun together, ...	3
3	2019-Q1	Mgr 1	Team 1	I fell pretty good about the health of our tea...	4
4	2019-Q1	Mgr 1	Team 1	happy with team's overall health and good dyna...	5
5	2019-Q1	Mgr 1	Team 1	Solid	6
6	2019-Q1	Mgr 1	Team 1	The Team 2 team is a collaborative group prod...	7
7	2019-Q1	Mgr 1	Team 1	We have great teamwork. We have a lot of fun....	8
8	2019-Q1	Mgr 1	Team 1	We feel good about our teamwork, process, tech...	9
9	2019-Q1	Mgr 1	Team 2	A <u>blast</u> ! Always working towards delivering mo...	10

# นำเข้าไฟล์ข้อมูลสาริตและข้อความก่อนการประมวลผล

- แบ่งย่อยฟิลด์ row\_id และ Response ลงในข้อมูลเฟรมใหม่ ซึ่งเป็นรูปแบบอินพุตที่จำเป็นสำหรับคลาส SentimentIntensityAnalyzer • ล้างข้อมูลข้อความโดยลบอักขระที่ไม่ใช่ตัวอักษรทั้งหมดและแปลงข้อความทั้งหมดเป็นตัวพิมพ์เล็ก

```
#สร้างกรอบข้อมูลใหม่ด้วยฟิลด์ "id" และ "comment" df_subset = df[['row_id', 'Response']].copy() #การล้างข้อมูล #ลบอักขระ
ที่ไม่ใช่ตัวอักษรทั้งหมด df_subset['Response'] = df_subset['Response'].str.replace("[^a-zA-Z#]", "
") #แปลงเป็นตัวพิมพ์เล็ก
df_subset['Response'] = df_subset['Response'].str.casefold() พิมพ์
(df_subset.head(10))
```

	row_id	Response
0	1	we re a fun team that works well together and ...
1	2	we have a sound and collaborative team focused...
2	3	we work well as a team we have fun together ...
3	4	i fell pretty good about the health of our tea...
4	5	happy with team s overall health and good dyna...
5	6	solid
6	7	the team team is a collaborative group prod...
7	8	we have great teamwork we have a lot of fun ...
8	9	we feel good about our teamwork process tech...
9	10	a blast always working towards delivering mo...

# สร้างคะแนนข้อความรู้สึ

- พจนานุกรมความรู้สึกใน VADER เป็นรายการคุณลักษณะของคำศัพท์ เช่น คำและวลีที่ระบุว่าเชิงบวกหรือเชิงลบตามแนวทางการความหมาย
- แนวทางตามกฎเกณฑ์นั้นเหมาะเป็นอย่างยิ่งสำหรับการตรวจจับความรู้สึกในแอปพลิเคชันทั่วไป เช่น โพสต์โซเชียลมีเดีย บทความผลิตภัณฑ์หรือบริการ และการตอบแบบสำรวจ
- สร้างคะแนนตัวเลขในช่วงตั้งแต่ลบหนึ่ง (-1) ถึงบวกหนึ่ง (+1) เพื่อระบุความเข้มข้นของความรู้สึกเชิงลบหรือเชิงบวก
- สิ่งนี้เรียกว่าคะแนนขั้วและได้รับการนำไปใช้โดยวิธี `polarity_score` ของคลาส `SentimentIntensityAnalyzer`
  - ช่วง -1 ถึง -0.5 หมายถึงความรู้สึกเชิงลบ
  - คะแนนที่มากกว่า -0.5 และน้อยกว่า +0.5 แสดงถึงความรู้สึกเป็นกลาง
  - ช่วง +0.5 ถึง 1 หมายถึงความรู้สึกเชิงบวก

# สร้างคะแนนข้อความรู้สึก

- สร้างดาต้าเฟรมสำหรับการจัดเตรียมเอาต์พุตของ  
วิธีการ `SentimentIntensityAnalyzer.polarity_scores`

```
# ตั้งค่า dataframe ว่างสำหรับแสดงเอาต์พุต df1=pd.DataFrame()  
df1['row_id']=['99999999999']  
df1['sentiment_type']='NA999NA'  
df1['sentiment_score']=0
```

# สร้างคะแนนข้อความรู้สึ

- เกี่ยวข้องกับการสร้างอินสแตนซ์ของวัตถุของคลาส `SentimentIntensityAnalyzer` และรัน for-loop เพื่อวนซ้ำวิธี `polarity_scores` ในแต่ละแถวของข้อมูลอินพุต `df_subset`
- ลูป for อีกอันหนึ่งถูกฝังไว้ในลูปก่อนหน้านี้เพื่อเขียนคะแนนข้อความรู้สึสำหรับประเภทความรู้สึกแต่ละประเภทลงในข้อมูลเฟรมกลาง
- ค่าประเภทความรู้สึกทั้งสามมีดังนี้:
  - **neg** ความรู้สึเชิงลบ
    - neu สำหรับความรู้สึกเป็นกลาง
  - **pos** สำหรับความรู้สึกเชิงบวก
    - **ผสม** คะแนนรวมที่ประกอบด้วยค่าลบ ค่าบวก และค่ากลาง  
ความรู้สึกรวมเป็นคะแนนเดียว

# สร้างคะแนนชีวความรู้สึก

- เมื่อสิ้นสุดรูป for ให้ทำความสะอาดข้อมูลเอาต์พุตโดย:
  - การลบแถวเสมือนจากข้อมูลเฟรมเอาต์พุต
  - การลบแถวที่ซ้ำกันซึ่งอาจแทรกเข้าไปในข้อมูลเฟรมเอาต์พุตได้
- การกรองข้อมูลเอาต์พุตเพื่อเก็บเฉพาะแถวสำหรับประเภทความรู้สึกของสารประกอบเท่านั้น



```

พิมพ์('กำลังประมวลผลการวิเคราะห์ความรู้สึก...') sid =
SentimentIntensityAnalyzer() t_df = df1 สำหรับ ดัชนี แถว
ใน
df_subset.iterrows(): คะแนน = sid.polarity_scores(row[1]) สำหรับ
คีย์ ค่า ใน scores.items(): temp = [key,value,row[0]]
df1['row_id']=row[0] df1['sentiment_type']=key
df1['sentiment_score']=value
t_df=pd.concat([t_df,df1]) #au
แถวเสมือนที่มี row_id = 9999999999
t_df_cleaned = t_df[t_df.row_id !=
'9999999999'] #au ซ้ำกันถ้ามีอยู่
t_df_cleaned = t_df_cleaned.drop_duplicates() # เก็บเฉพาะแถวที่
sentiment_type = compound t_df_cleaned = t_df[t_df.sentiment_type ==
'compound'] print(t_df_cleaned.head(10))

```

	row_id	sentiment_type	sentiment_score
0	1	compound	0.6597
1	2	compound	0.9287
2	3	compound	0.8122
3	4	compound	0.8225
4	5	compound	0.8271
5	6	compound	0.1531
6	7	compound	0.9382
7	8	compound	0.9381
8	9	compound	0.9468
9	10	compound	0.5519

# สร้างคะแนนชั่วคราว

- รวมข้อมูลเอาต์พุต t\_df\_cleaned กับข้อมูลอินพุต df โดยใช้ฟิลด์ row\_id • การดำเนินการ

การผสานข้อมูลเฟรมใน Python นี้มีแนวคิดคล้ายคลึงกับการดำเนินการรวมตารางฐานข้อมูลสองตารางใน SQL • ข้อมูลเฟรมที่

ผสานกันจะมีฟิลด์ต่อไปนี้:

- ระยะเวลา
- ผู้จัดการ
- ทีม
- การตอบสนอง
- แกลว\_id
- ประเภทความรู้สึก
- คะแนนความรู้สึก

#รวมข้อมูลเพื่อ df\_output

= pd.merge(df, t\_df\_cleaned, on= 'row\_id' , how='inner') print(df\_output.head(10))

	Period	Manager	Team	Response \
0	2019-Q1	Mgr 1	Team 1	We're a fun team that works well together and ...
1	2019-Q1	Mgr 1	Team 1	we have a sound and collaborative team focused...
2	2019-Q1	Mgr 1	Team 1	we work well as a team, we have fun together, ...
3	2019-Q1	Mgr 1	Team 1	I fell pretty good about the health of our tea...
4	2019-Q1	Mgr 1	Team 1	happy with team's overall health and good dyna...
5	2019-Q1	Mgr 1	Team 1	Solid
6	2019-Q1	Mgr 1	Team 1	The Team 2 team is a collaborative group prod...
7	2019-Q1	Mgr 1	Team 1	We have great teamwork. We have a lot of fun....
8	2019-Q1	Mgr 1	Team 1	We feel good about our teamwork, process, tech...
9	2019-Q1	Mgr 1	Team 2	A blast! Always working towards delivering mo...

	row_id	sentiment_type	sentiment_score
0	1	compound	0.6597
1	2	compound	0.9287
2	3	compound	0.8122
3	4	compound	0.8225
4	5	compound	0.8271
5	6	compound	0.1531
6	7	compound	0.9382
7	8	compound	0.9381
8	9	compound	0.9468
9	10	compound	0.5519



# คำถาม

อ้างอิง:

<https://campus.datacamp.com/>

ภาษาไทย: <https://www.red-gate.com/simple-talk/การพัฒนา/การพัฒนาข้อมูลวิทยาศาสตร์/การวิเคราะห์ความรู้สึก-python/?fbclid=IwAR2saZfrAYF3CGuPiMgBongQPuRyrv3olMa7rCR1CDhPQ1Q3kpoYBjCDrME>