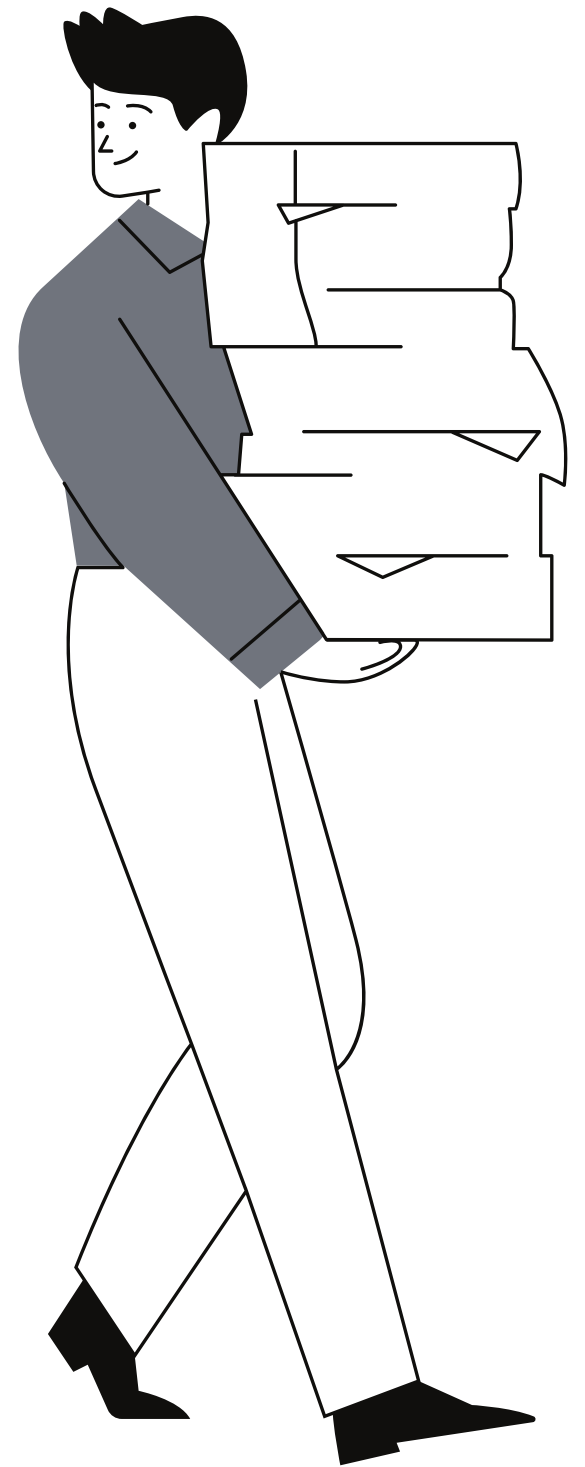


Big Data Analysis Course Project

Rumour Spreading Analysis on Twitter



ขั้นตอนการทำงาน

1

เลือกหัวข้อข่าว , นำเข้า sourcetweets และ retweet

2

Update sourcetweets ด้วย annotation

3

วิเคราะห์คำและความถี่โดยใช้ Word Cloud

4

วิเคราะห์แต่ละทวิตได้รับการรีทวิตกี่ครั้งบนโซเชียลมีเดีย (ทั้งข่าวลือและไม่ใช่ข่าวลือ)

5

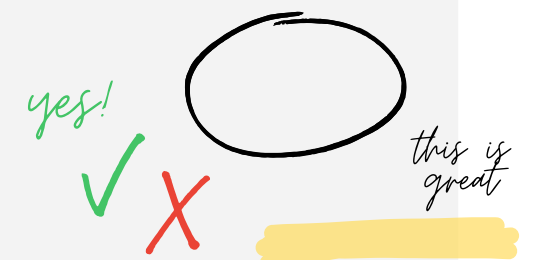
วิเคราะห์แต่ละทวิตเร็วแค่ไหน

6

วิเคราะห์ว่าใครทวิต และแต่ละทวิตถูกรีทวิตโดยใคร

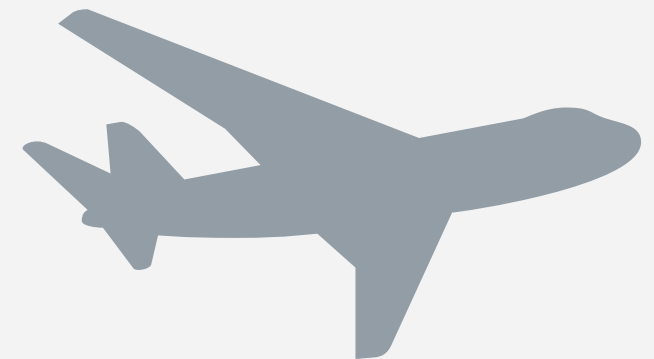
7

วิเคราะห์เกี่ยวกับตำแหน่งที่ทวิตแต่ละทวิต



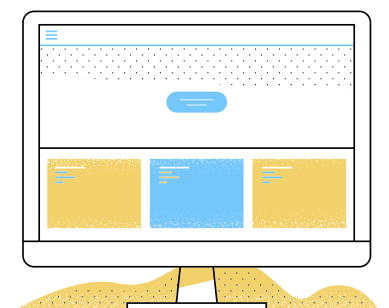
Data is a set of rumour tweets

germanwings-crash



STEP 1

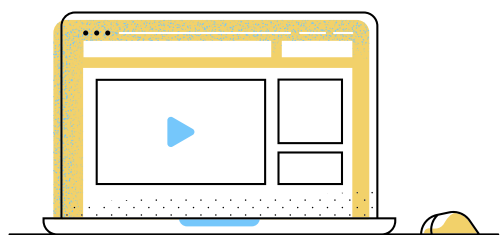
ทำการ Import source-tweets
data ลงในฐานข้อมูล MongoDB



```
1 # insert source tweet
2
3 with open(r'PATH') as file:
4     f = json.load(file)
5     if isinstance(f, list):
6         Collection.insert_many(f)
7         print("many")
8     else:
9         Collection.insert_one(f)
10        print("one")
```

STEP 1

ทำการ Import retweets data
ลงในฐานข้อมูล MongoDB



```
1  # insert retweet
2
3  with open(r'PATH') as file:
4      f = [json.loads(line) for line in file]
5  if isinstance(f, list):
6      Collection.insert_many(f)
7      print("many")
8  else:
9      Collection.insert_one(f)
10     print("one")
```

STEP 2

UPDATE anotation.json file into sourcetweet collection



```
1 // Update annotation to sourcetweet *Step 2*  
2  
3 db.sourcetweets.updateMany({"id" : 580319078155468800}, {"$set": {"annotation": {"is_rumour": "rumour"  
4
```



STEP 3

วิเคราะห์เกี่ยวกับคำสำคัญ
และความถี่โดยใช้ Word Cloud
เพื่อดูว่ามีคำอะไรบ้างและความถี่
แค่ไหนในข้อมูลที่เป็นข่าวลือ
และข้อมูลที่ไม่ใช่ข่าวลือ

Misinformation : "1"



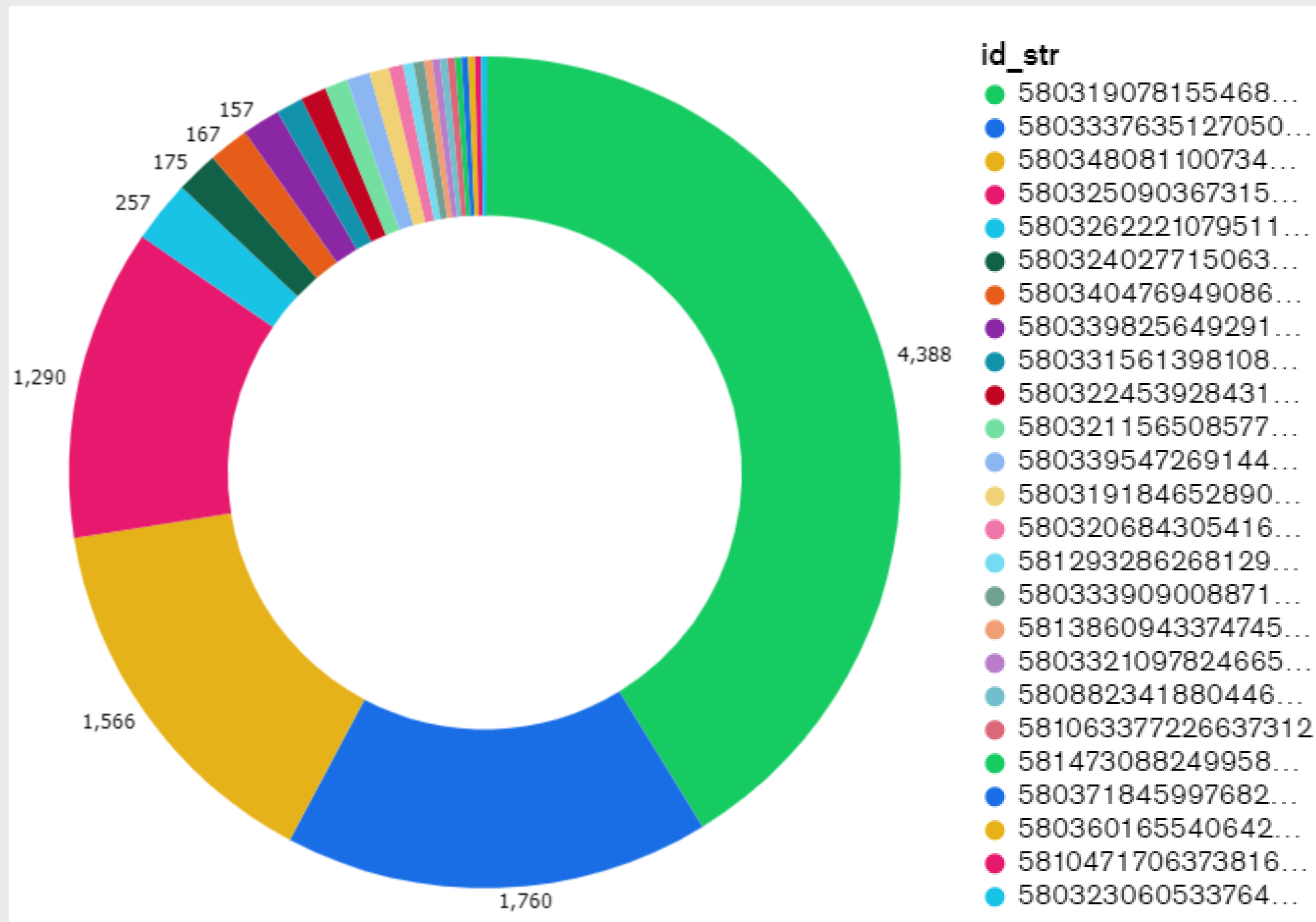
Misinformation : "0"



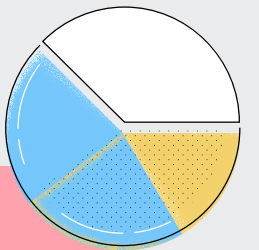
Misinformation : "1" --> ข่าวลือ

Misinformation : "0" --> ข่าวจริง

STEP 4

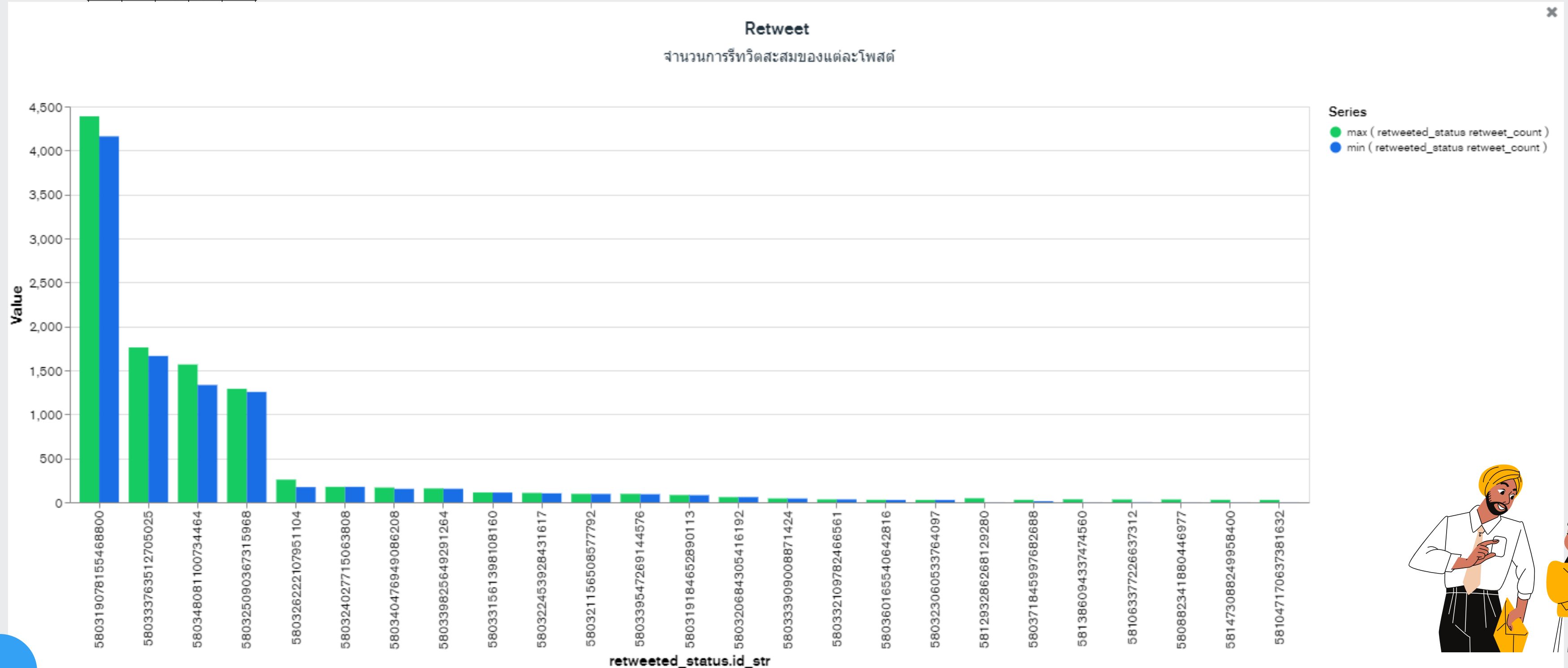
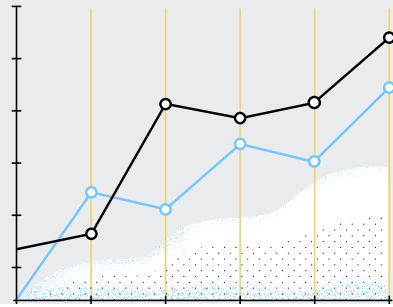


วิเคราะห์
เพื่อดูว่าแต่ละทวีตมีกี่ครั้ง
(ทั้งที่เป็นข่าวลือและที่ไม่ใช่ข่าวลือ)
ที่ได้รับการรีทวีตบนโซเชียลมีเดีย



STEP 5

ทำการวิเคราะห์เพื่อดูว่าแต่ละทวีต
ถูกรีทวีตบนโซเชียลมีเดียโดยการนับจำนวนทวีตที่สะสมตลอดเวลา



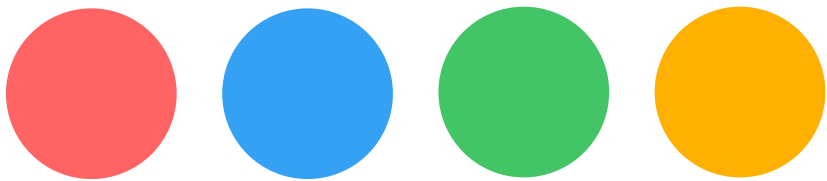
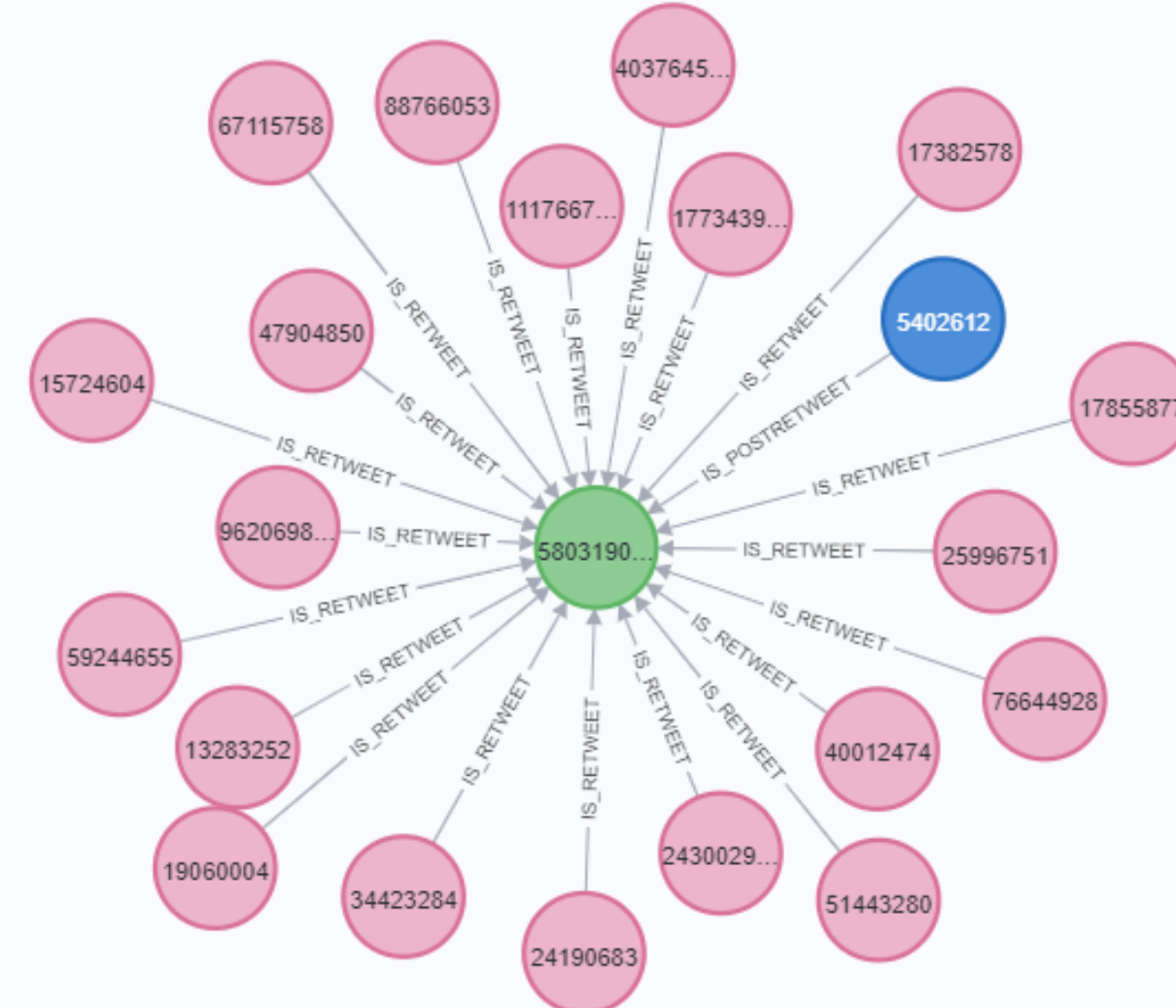
STEP 6

analysis on who tweeted
the tweets,
each of them was
retweeted by whom

NEO4J

*(22) Userretweet(20) Userpost(1) Post(1)
*(21) IS_RETWEET(20) IS_POSTRETWEET(1)

```
1 call apoc.load.json("file:/retweets.json")
2 YIELD value
3 UNWIND value as R
4 merge (r:Userretweet {id:R.user.id})
5 merge (p:Userpost{id:R.retweeted_status.user.id})
6 merge (w:Post {id:R.retweeted_status.id_str})
7 merge (r)-[rl:IS_RETWEET]→(w)
8 merge (p)-[rl2:IS_POSTRETWEET]→(w)
9 return r,p,w
10 limit 20
```



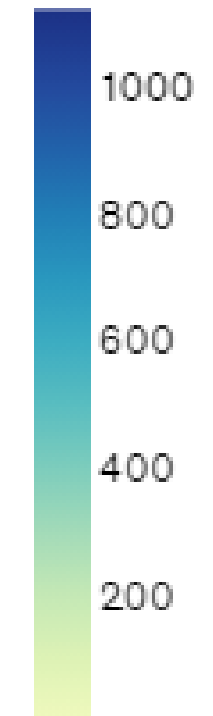
STEP 7

วิเคราะห์เกี่ยวกับตำแหน่งที่ทวีตแต่ละทวีต

Map



count (id_str)



รายชื่อสมาชิก

61050322 นางสาวอรอุษา ทองม้วน

61050941 นางสาวจณิสตา ธารอำมฤต

61050952 นายพีรพล แสงสว่าง

61050956 นางสาววรดา ศิริเนตร์



THANK YOU

Have a great
day ahead