

國立東華大學應用數學系  
統計軟體與實務應用成果報告書

掌握 Instagram 流量的密碼

指導教授：曹振海

學生：應數四 張秣穎（410711316）

應數三 吳雨瑾（410811304）

應數三 鍾宜玲（41087A008）

中 華 民 國 1 1 1 年 1 月 3 日

## 摘要

我們以 Kaggle 裡 Instagram 資料，選取了一些可能和得讚數相關的因素來討論。主以散佈圖討論相關性，輔以文字雲作為延伸，得出了熱門關鍵字才是流量的關鍵，也就是我們所謂的「密碼」。

## 大綱

摘要.....	1
大綱.....	1
1. 介紹.....	2
1.1 研究背景與動機.....	2
1.2 研究目的.....	2
1.3 數據.....	2
1.4 分析變量.....	2
2. 資料整理與觀察.....	3
2.1 分割標籤.....	3
2.2 加上每篇文章的標籤個數.....	4
3. 資料分析.....	4
3.1 追蹤人數與得讚數的關係.....	4
3.2 標籤數量與得讚數關係.....	5
3.3 文字雲.....	6
3.4 發佈後幾小時內追蹤人數與得讚數關係.....	6
3.5 迴歸分析.....	7
4. 結論.....	8
5. 工作分配.....	8
參考文獻.....	9

## 1. 介紹

### 1.1 研究背景與動機

在手機普及之前，擁有明星夢的人往往只能藉由電視節目才可能得到大眾關注。現今智慧型手機普及率高，人人都有自己的手機與社群帳號，動動手指就可以擁有展現自己的機會，無需尋求電視台或是節目曝光。至於如何在眾多素人中脫穎而出，就是本研究的動機。我們發現正確的標籤可以讓文章擁有更多曝光度，從而達到表現自己的機會。

### 1.2 研究目的

對於剛起步的素人來說，首要目標就是推廣自己、進入大眾視野，能達到此目的的重要關鍵就是「圖文」與「標籤」。運用正確且效率高的標籤可以大大增加曝光度；發文的時間則關係到貼文的得讚數與留言數，妥善利用上列因素便可輕易達到預期的曝光度及觸擊率。此研究目的正是提供素人流量的密碼，達到與藝人明星不相上下的曝光度，也給自己的夢想一個更好的機會。

### 1.3 數據

#### i. 來源

Instagram Reach

(<https://www.kaggle.com/rxsraghavagrawal/instagram-reach>)

#### ii. 焦點問題

- 自媒體與個人品牌是現今的創新產業，正因如此社群的流量與活動顯得更加重要，如何發文與如何創造流量是創造商機的關鍵。
- 如何在正確的時間發文？==>時間與得讚數的相關性？
- 標籤數量越多，得讚數就越多嗎？

### 1.4 分析變量

- 得讚數
- 追蹤人數
- 標籤(Hashtags)數量
- 文章發布後時長

## 2. 資料整理與觀察

### 2.1 分割標籤

S.No	Hashtags
1	#MachineLearning #AI #DataAnalytics #DataScienc #DataLake
2	?#deck? #mac?#macintosh#sayhello?#apple?#stevejobs?#ai?#evolution#artificialintelligence?#machinelearning#terminator?#illbeback? #technology#computerevolution?#computerscience#sciencefiction#computersciencetosciencefiction#tomorrowstechnology?#vr?#ar? #robot#robots?#2?#businessdeck?#businessslides#illustration?#sketches?#drawing
3	#whoiswho?#aitrading?#ai?#aitradingteam#instateam?#instapeople?#ai?#trading#artificialintelligence?#crypto#cryptocurrency?#blockchain? #tradingforex#forex?#fiatmoney?#coins?#machinelearning#userexperience?#instamachinelearning#instabigdata? #instamarketing#artificialintelligence?#deeplearning#datascience?#industry?#marketing#bigdata?#datascience?#machinelearning#ml
4	#iot?#cre#workplace?#CDO?#bigdata?#technology#tech?#work?#computerscience?#blockchain#softwaredeveloper? #machinelearning#smartcity?#artificialintelligence#futureofwork?#opensource?#cybersecurity#ai?#softwareengineer?#engineering#smartcities? #developer?#autonomous#singularity?#webdeveloper#cryptocurrency?#programmers?#crypto#iot?#robot
5	#instamachinelearning?#instabigdata#instamarketing?#artificialintelligence#industry?#deeplearning?#datascience#marketing?#machinelearning? #bigdata#artificialintelligence?#instadaily?#tech#software?#softwareengineer?#technology#crea.vision?#techy?#artificial? #artificiale#pictureoftheday?#marketingdigital?#digital#live?#better?#enjoy

▲ 圖 1 Hashtag 資料分割(取前五項)

No	words
1	machinelearning
1	ai
1	dataanalytics
1	datascienc
1	datalake
2	deck
2	mac
2	macintosh
2	sayhello
2	apple
2	stevejobs
2	ai
2	evolution
2	artificialintelligence
2	machinelearning

▲ 圖 2 資料分割後(僅以部份表示)

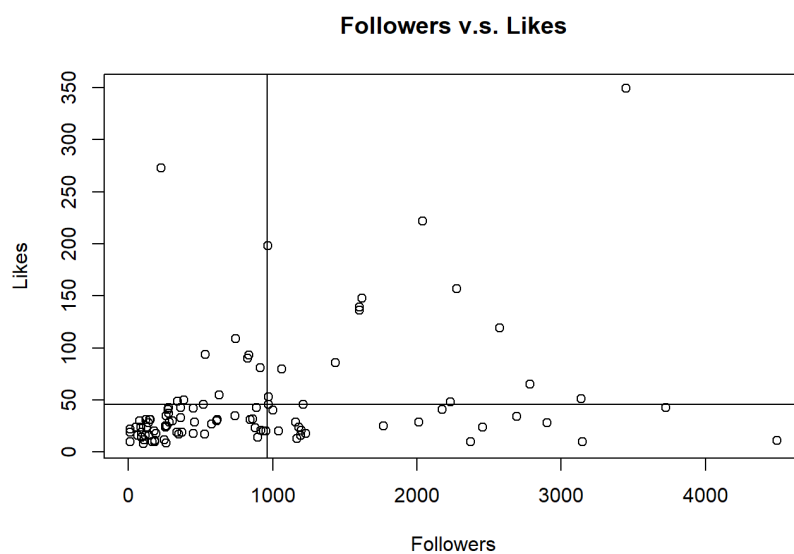
## 2.2 加上每篇文章的標籤個數

No	Followers	Time	Hashtags	Likes
1	1600	11	5	139
2	880	2	28	23
3	255	2	30	25
4	340	3	30	49
5	304	3	26	30

▲ 圖 3 加上標籤個數後的資料

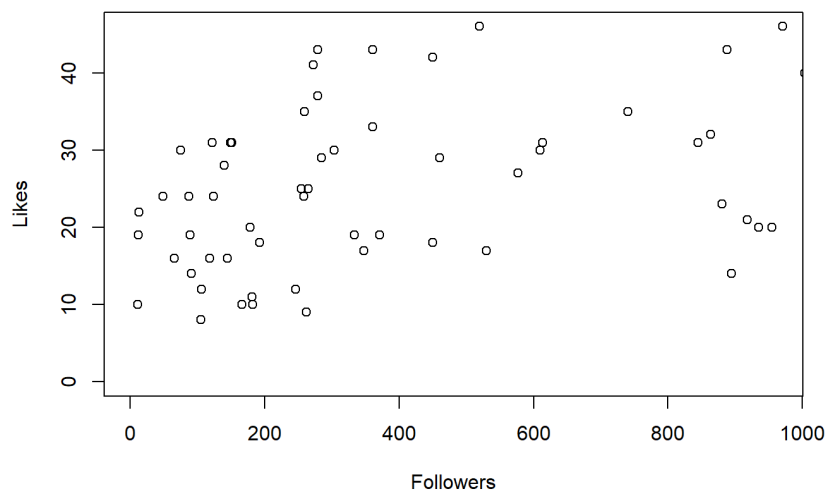
## 3. 資料分析

### 3.1 追蹤人數與得讚數的關係



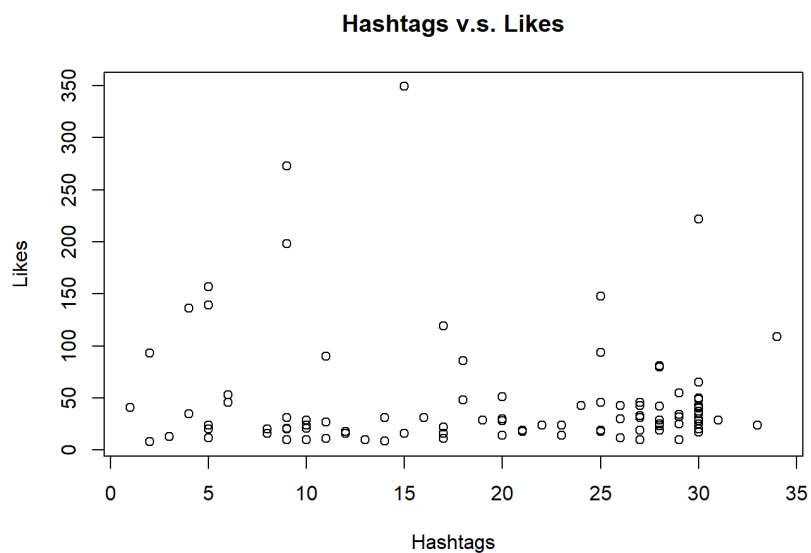
▲ 圖 4 追蹤人數 vs. 得讚數散佈圖

▶ 發現左下角資料筆數較多，將進一步觀察該區域



- ▲ 圖 5 追蹤人數 vs. 得讚數散佈圖 - 聚焦左下角
  - ▶ 可看出粉絲人數在 1000 人以內，得讚數與追蹤者數量成正相關。

### 3.2 標籤數量與得讚數關係



- ▲ 圖 6 標籤數量 vs. 得讚數散佈圖
  - ▶ 由於得讚數在 0 到 50 之間的資料無法判斷變化，且我們認為標籤是 Instagram 一項非常強大的功能，也是用戶的共同語言，因此我們做了標籤的文字雲。



由於資料筆數少，我們無法得出發文後時長及按讚數的相關性。於是我們分別將各項時長（2、3、4 小時以及 5 小時及以上）得讚數最高的資料抓出來查看，發現每筆所標記的標籤至少包含一項熱門標記（”ai”、”blockchain”、”startup”）。

### 3.5 迴歸分析

- 基本設定

$Y$  = 得讚數

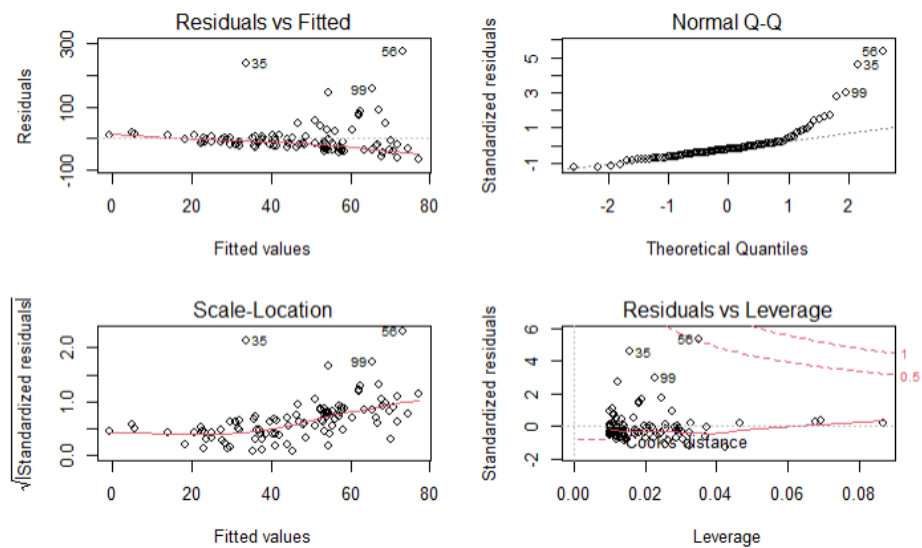
$X_1$  = 追蹤人數

$X_2$  = 標籤(Hashtags)數量

$X_3$  = 文章發布後時長

- 模型

$$Y = \log(X_1 + X_2 + X_3)$$



▲ 圖 9 線性迴歸模型分析

- ▶ 可以看到在 Residuals vs Fitted 圖以及 Normal Q-Q plot 表現都不佳。

除了這次模型分析以外，我們也做了許多嘗試來處理這筆資料，例如： $\sqrt{Y} = X_1 + X_2 + X_3$ 、 $\log Y = X_1 + X_2 + X_3$  等等。因為目前迴歸方面的知識還不夠熟悉，模型改善後的表現都不如預期，所以僅展示此次分析結果。



## 4. 結論

- 現今自媒體種類繁多，自我行銷的方式隨著社群不同而有所差異。例如 YouTube 上吸引人氣的主要為 10 至 20 分鐘的影片；Facebook 上的 15 秒短影片近期也成為眾人目光聚焦之處；而對於 Instagram，除了吸睛的圖文外，由於其特殊的搜尋以及追蹤機制，是否標記到熱門的 Hashtags(標籤)也是創造流量及商機的「密碼」之一。
- 由於數據筆數不足，因此在迴歸方面我們無法得出時間與得讚數是否具有顯著相關性的結論，但在不同的貼文發布時長之下，我們發現得讚數高的依然是標記到熱門關鍵字的貼文。
- 我們發現標籤數量多並不是得讚數的重點，有時標記太多反而讓文章失去重點，適得其反。

## 5. 工作分配

- 資料蒐集：吳雨瑾、鍾宜玲、張秭穎
- 資料處理：吳雨瑾
- 資料分析：吳雨瑾、鍾宜玲
- 迴歸分析：鍾宜玲
- 成果發表：張秭穎

## 參考文獻

1. 數據來源：Instagram Reach

[<https://www.kaggle.com/rxsraghavagrawal/instagram-reach>]

2. 不花錢也能讓 IG 粉絲變多？IG「最火熱門標籤 Top10」加上#hashtag 讓你人氣爆增！

[<https://www.bella.tw/articles/design&gadget/23551>]

3. 社群什麼時候發文最多人看？FB、YouTube、IG 最佳發文時間！

[<https://www.marketersgo.com/marketing-strategies-tactics/202001/ig-fb-yt-social-media/#IG>]

4. 2019 IG 必學 4 招全攻略，貼文觸及翻倍，從零開始打造品牌成為人氣王！

[<https://www.cyberbiz.io/blog/2019-ig-%E5%BF%85%E5%AD%B8-4-%E6%8B%9B%E5%85%A8%E6%94%BB%E7%95%A5%E7%BC%8C%E8%B2%BC%E6%96%87%E8%A7%B8%E5%8F%8A%E7%BF%BB%E5%80%8D%E7%BC%8C%E6%89%93%E9%80%A0%E4%BA%BA%E6%B0%A3%E7%8E%8B%E5%93%81%E7%89%8C/>]

5. 此報告之 RMD 檔案（GitHub 網站）

