

社群媒體資料分析實務

PRACTICE OF SOCIAL MEDIA

M10515040 王 云

- 1. 建構個人資料分析平台。
- **a.**請說明所安裝的作業系統以及硬體資源



- 1.建構個人資料分析平台。

- **b.** 請簡述安裝步驟以及平台架構

- 下載網址：<https://docs.docker.com/engine/installation/mac/>

- 下載 Docker for mac

安裝完成就會在右上角看到一隻鯨魚



- Container 技術建構分析平台：

OS內的核心系統層來打造虛擬執行環境，

透過共用Host OS的作法，取代一個一個Guest OS的功用。

Container也因此被稱為是OS層的虛擬化技術。

- 1.建構個人資料分析平台。
- **c.**請簡述所遇到的困難以及解決方式：
照著老師的、PPT、以及、Docker、的引導步驟
即可順利完成安裝
- **d.**若有額外安裝其它套件,請說明：
無

2.請下載已抓好的TWITTER ES檔案,選定十組以內關特定主題之關鍵字,對於關鍵字組進行蒐集,並選定蒐集時間範圍,並請回答下述問題:

十組 資安相關 關鍵字 :

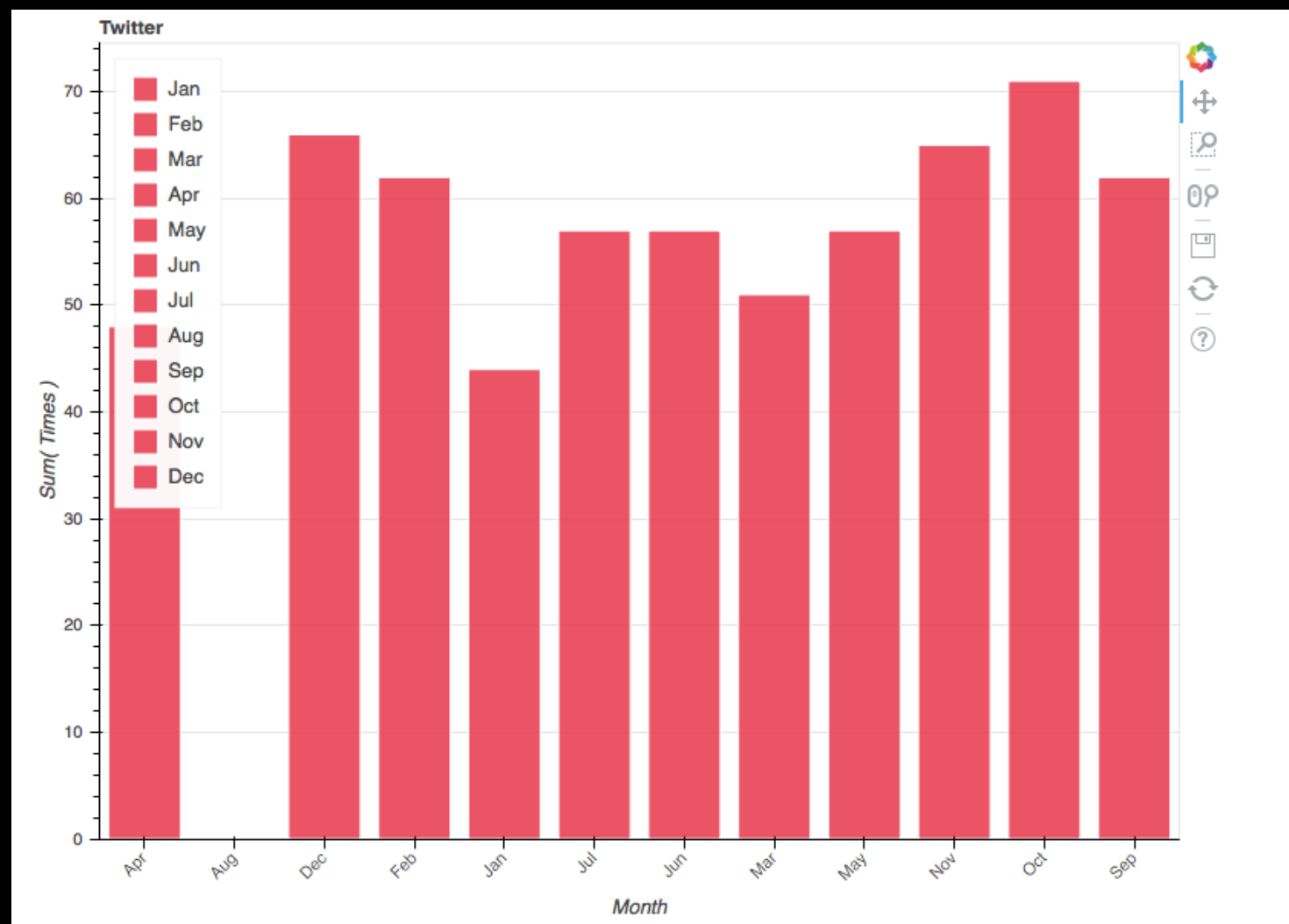
**cve 、 ddos 、 vulnerability 、 malware 、 exploit 、
cryptolocker 、 apt 、 anonymous 、 attack 、 hack**

a. 共計有多少Twitter參與這類關鍵字議題討論? 1387

b. 共計有多少Tweets? 6124

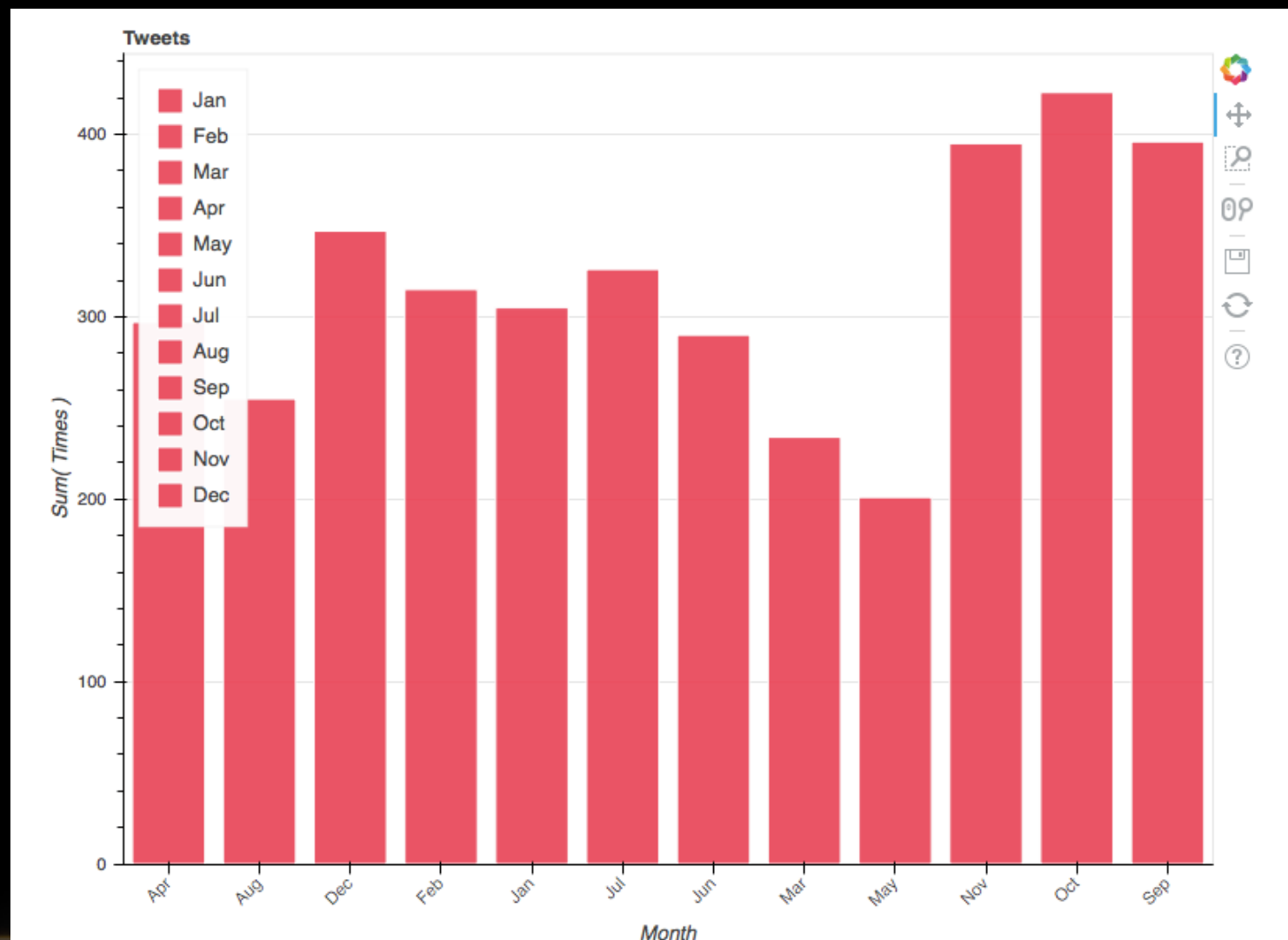
2.請下載已抓好的TWITTER ES檔案,選定十組以內關特定主題之關鍵字,對於關鍵字組進行蒐集,並選定蒐集時間範圍,並請回答下述問題:

C.每個月參與的Twitter數量長條圖



2.請下載已抓好的TWITTER ES檔案,選定十組以內關特定主題之關鍵字,對於關鍵字組進行蒐集,並選定蒐集時間範圍,並請回答下述問題:

d.每個月參與的Tweets數量長條圖

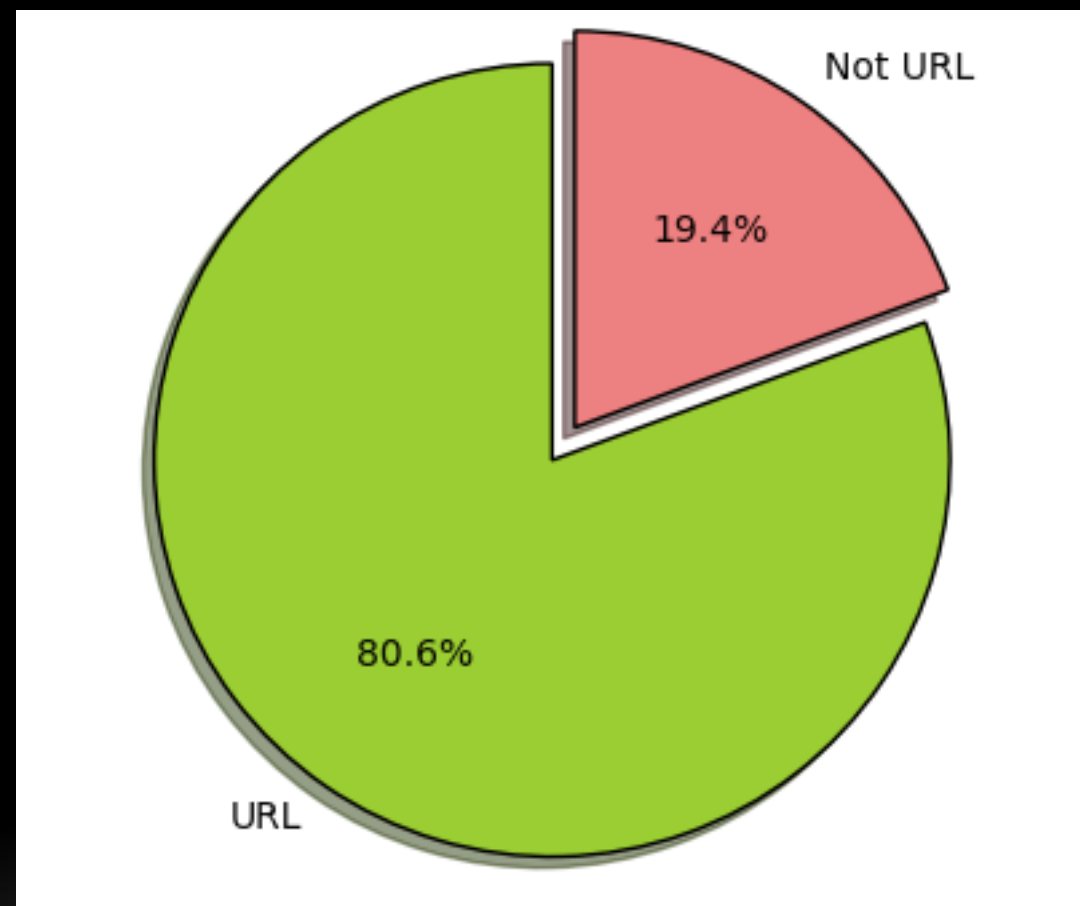


2.請下載已抓好的TWITTER ES檔案,選定十組以內關特定主題之關鍵字,對於關鍵字組進行蒐集,並選定蒐集時間範圍,並請回答下述問題:

e.其中有URL與無URL的Tweets的比例圓餅圖

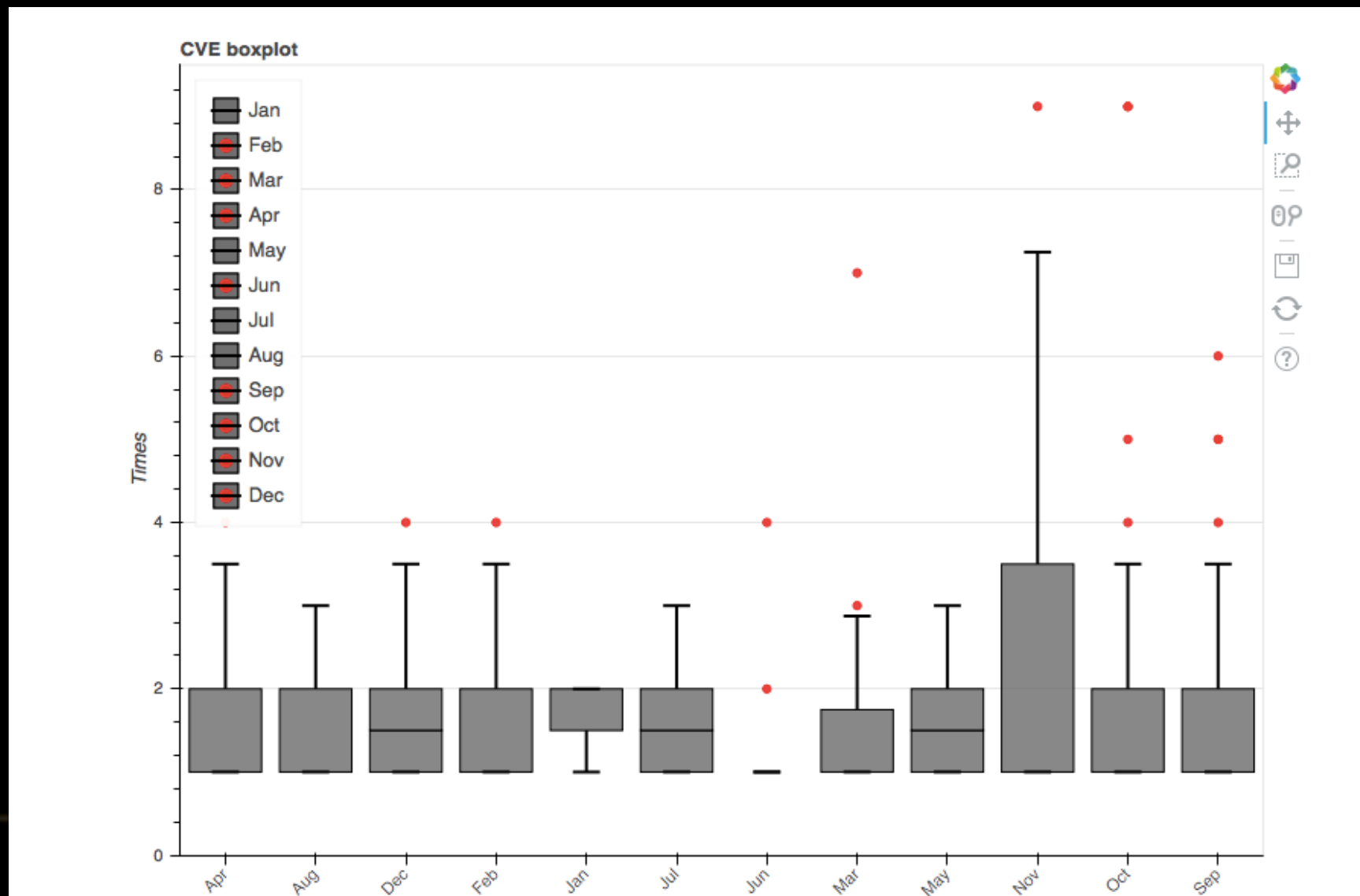
('url:',3055)

('noturl:',733)



2.請下載已抓好的TWITTER ES檔案,選定十組以內關特定主題之關鍵字,對於關鍵字組進行蒐集,並選定蒐集時間範圍,並請回答下述問題:

f.請依照每個月繪製各Twitter所提到CVE的箱型圖(每月最多發表帳號的次數、每月平均發表的次數、每月至少發表的次數、第一分位數、第三分位數)



2.請下載已抓好的TWITTER ES檔案,選定十組以內關特定主題之關鍵字,對於關鍵字組進行蒐集,並選定蒐集時間範圍,並請回答下述問題:

g.能否分析資安弱點分析指標帳號?如何分析?

是,

看哪些帳號 發相關文章的次數最多

3.請選擇20個以上感興趣的TWITTER帳號,定義三類分析帳號標籤,並分析TWITTER所討論的TWEET內容,萃取關鍵字為屬性,透過決策術、演算法取得分類規則,請呈現決策術並以HEATMAP對於決策樹的參數進行分析。

挑選的20個TWITTER帳號：

- 1778983, 11891512, 118019149,198365324,132778684,
- 16730420,198365324,132778684,16730420,198365324,
- 297856522,22790881,18252252,198365324,39262054,
- 8236572,161038997,18789893,18252252,14090906

Label：

0：公司企業

1：路人

2：了解資安相關知識的人

np.array attribute：

1: text內容包含“cryptolocker”

2: text內容包含“cve”

3: 半年內內發文超過5篇

4: 近三個月內有過發文章

5: text內容包含“ddos”

3.請選擇20個以上感興趣的TWITTER帳號,定義三類分析帳號標籤,並分析TWITTER所討論的TWEET內容,萃取關鍵字為屬性,透過決策術、演算法取得分類規則,請呈現決策術並以HEATMAP對於決策樹的參數進行分析。

