Presentation

Intro:
When we came across this competition, there were few options to choose. like mobile security, cloud security, Internet of things, Forensics, etc. We chose cloud as it is the top choice of every organization these days to store their data and operate on it. Nowadays, almost every big or small organization is moving their business on cloud. So, Cloud Security is of utmost priority since the personal and company's data is at stake. But these organizations have encountered major breaches in recent years which affected millions of users and also caused huge financial loss.
It brought our focus to the level of awareness among people and what are their sentiments with respect to the major incidents.
So, keeping cloud security in mind, we have researched for certain major incidences that has happened in the past and by using Python language, we did Twitter sentiment analysis and got some insights from the analysis.

Workflow:

1. Identifying incidents of interest (scope of interest) - based on the amount of users affected and financial and personal data loss.

Affected organizations
Facebook - 87 million users (70 million US users)
Typeform -
Equifax - 143 million in US
Dropbox - login and password pairs for more than 68 million Dropbox accounts
LinkedIn - hackers had obtained the passwords to 6.5 million
Yahoo - breaches knocked an estimated $350 million off Yahoo's sale price, 500 million user accounts had been hacked in 2014
Twitter

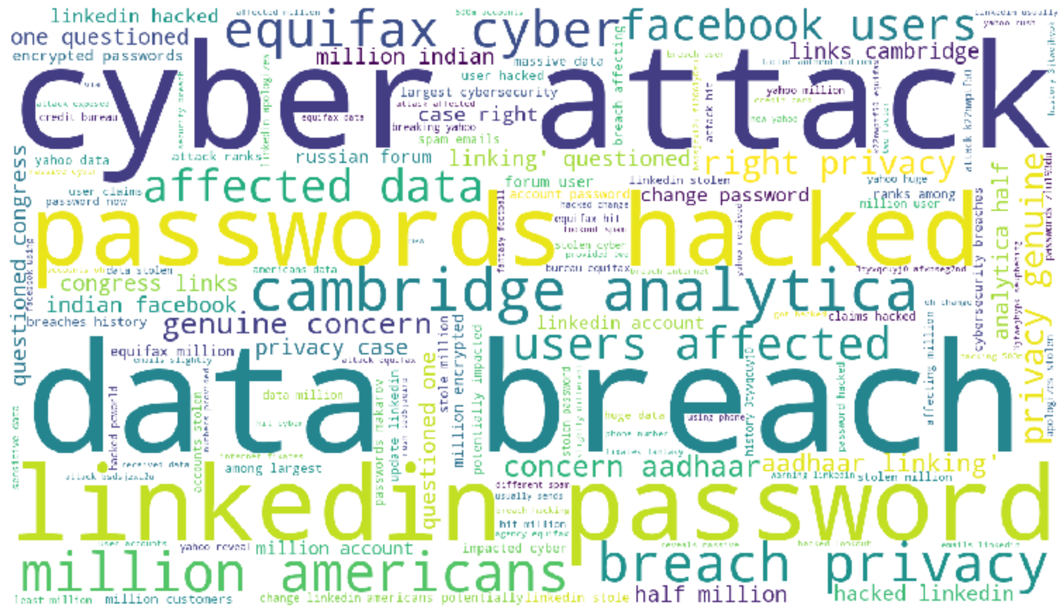
2. Selecting appropriate *key words* for each incident and *time* the incident happened. Time - the date of announcement of the incident. Key words - manually chosen from browsing news and articles.

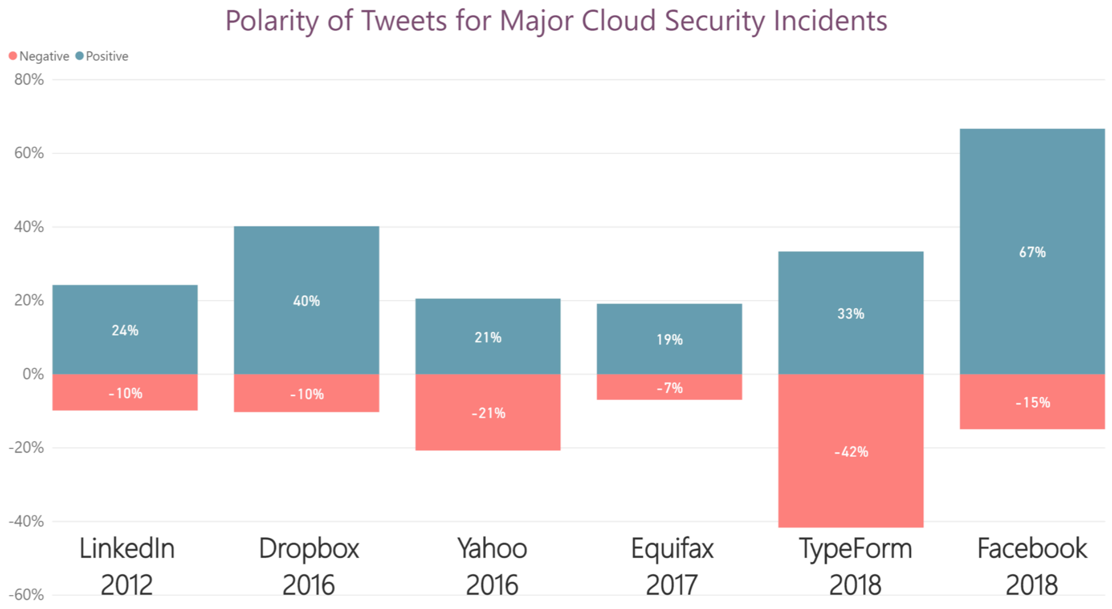
3. Fetching tweets using python and Twitter Historical search of tweets services (libraries used for fetching historical data - *search tweets, tweepy, twython*)

- Problems with fetching historical data due to limitations on free twitter developer account, such as 50 requests per month is available, 100 rows per request, not all fields are available (location is hidden)

4. Running sentiment analysis for each incident dataset. Identifying the proportion of negative/positive/neutral tweets for each incidents. (library - *TextBlob*)
5. Running analysis for most common words/phrases - World Cloud. (library - *WordCloud*) Iterative cleaning for non-words, meaningless words. (Ex. https, .co, //, says…)

Conclusion


cyber attack
data breach
passwords hacked
linkedin password
cambridge analytica
million americans
breach privacy
affected data
users affected
facebook users
equifax cyber
right privacy
privacy genuine
genuine concern
concern aadhaar
aadhaar linking'
questioned congress
questioned one
analytica half
linkedin hacked
one questioned
encrypted passwords
million indian
links cambridge
largest cybersecurity
case right
russian forum
linking' questioned
change password
congress links
indian facebook
privacy case
linkedin account
breaches history
equifax million
among largest
million account
half million
hacked linkedin
million customers
yahoo data
credit bureau
user claims
password now
user hacked
massive data
spam emails
breaking yahoo
forum user
account password
linkedin stolen
ranks among
million user
yahoo huge
equifax hit
stolen cyber
bureau equifax
claims hacked
data million
huge data
stolen million
hit million
impacted cyber
yahoo reveal
least million
change linkedin
americans potentially
linkedin stole
attack ranks
breach affecting
cybersecurity breaches
affecting million
password hacked
history
million encrypted
potentially impacted
stolen password
passwords makarov
update linkedin
stole million
accounts stolen
yahoo million
attack hit
data stolen
americans data
different spam
usually sends
breach hacking
agency equifax
phone number
using phone
reveals massive
emails linkedin


Polarity of Tweets for Major Cloud Security Incidents

1. The awareness for LinkedIn, Dropbox, Equifax was high immediately after the announcement, whereas TypeForm incident was less mentioned in the Tweets. (that is based on the amount of tweets we were able to get.
    1. LinkedIn - people tweeted approximately *15 tweets per minute*
    2. Dropbox - people tweeted approximately *2 tweets per hour*
    3. Yahoo - people tweeted *15-20 tweets per minute*
    4. Equifax - people tweeted *1-2 tweets per minute*
    5. TypeForm - *Very low awareness* - 4 tweets within 1 day.
    6. Facebook - people tweeted approximately *10 tweets per hour*

2. Negative polarity is high for *Yahoo* data leak. TypeForm negative polarity is high, however the *awareness* was comparatively *low.*
3. Unexpectedly a positive polarity for the latest Facebook breach is considerably high (67% of all tweets) - from observing analysis - people are not angry, even if they have concerns.
4. Mostly used words for incidents spreading are: data breach,

password, hack, cyber attack, affected data and privacy.

5. The date and time has not influenced the awareness of the security threads, which means that awareness has no trend over time.

———————————

Limitations

Twitter limitations that influenced our study:

- Historical tweets fetching is available only for premium account
- 50 requests of historical data per month is available for Sandbox subscription (free)
- 100 rows per request is a max (free)
- Restrictions on fields of data (Ex: location is hidden)

Limitations of the study:

- Twitter incident in 2016 was eliminated from the scope of study due to twitter request limitations (run out of data requests)

#Fall2018/conference