

SENTIMENT ANALYSIS FOR PRODUCT MANAGEMENT



MEMBERS:

Devika Bhatt
Divya Krishnan
Yu-Cheng Liu
Yu-Jui Chen
Chih-Hsiang Lai
Katie YuXian Jin
Fiona D'Souza

Social Media Analytics



PINTEREST

**SOCIAL SITE
THAT IS ALL ABOUT
DISCOVERY**

**LARGEST
OPPORTUNITIES**



USERS ARE:

17%
MALE
83%
FEMALE

**20
MILLION
ACTIVE USERS**



TWITTER

**MICRO BLOGGING
SOCIAL SITE
THAT LIMITS EACH
POST TO 140
CHARACTERS**

**LARGEST
PENETRATION**



BUT SPREADING
SLOWLY AND STEADILY

**5,700 TWEETS
HAPPEN
EVERY SECOND**

**241
MILLION
ACTIVE USERS**



FACEBOOK

**SOCIAL SHARING
SITE THAT HAS
1+ BILLION
USERS WORLDWIDE**

**LARGEST
OPPORTUNITIES**



**COMMUNICATING WITH
CONSUMERS
IN A NON-OBTRUSIVE WAY**

**USERS
SHARE
1 MILLION LINKS
EVERY 20 MINUTES**

**f
1+
BILLION
ACTIVE USERS**



INSTAGRAM

**SOCIAL SHARING
SITE ALL AROUND
PICTURES
AND NOW 15 SECOND
VIDEOS**

**MANY BRANDS
ARE PARTICIPATING
THROUGH THE USE OF**

HASHTAGS

**AND POSTING
PICTURES
CONSUMERS
CAN RELATE TO**

**MOST FOLLOWED
BRAND IS
NATIONAL
GEOGRAPHIC**

**200
MILLION
ACTIVE USERS**



GOOGLE+

**SOCIAL NETWORK
BUILT BY GOOGLE
THAT ALLOWS FOR
BRANDS
AND USERS
TO BUILD CIRCLES**

**CO
NOT AS MANY
BRANDS
ACTIVE,
BUT THE ONES THAT ARE
TEND TO BE A
GOOD FIT
WITH A
GREAT FOLLOWING**

**25-35 YEAR
OLDS
ARE THE MOST
ACTIVE**

**g+
540
MILLION
ACTIVE USERS**



LINKEDIN

**BUSINESS
ORIENTED
SOCIAL NETWORKING SITE**

**BRANDS THAT ARE
PARTICIPATING
ARE CORPORATE
BRANDS
GIVING POTENTIAL AND
CURRENT ASSOCIATES
A PLACE TO NETWORK
& CONNECT**



**POWERS
50% OF THE
WORLD'S HIRES**

**in
300
MILLION
USERS**



30 Billion Pieces of Content are shared on FaceBook every month

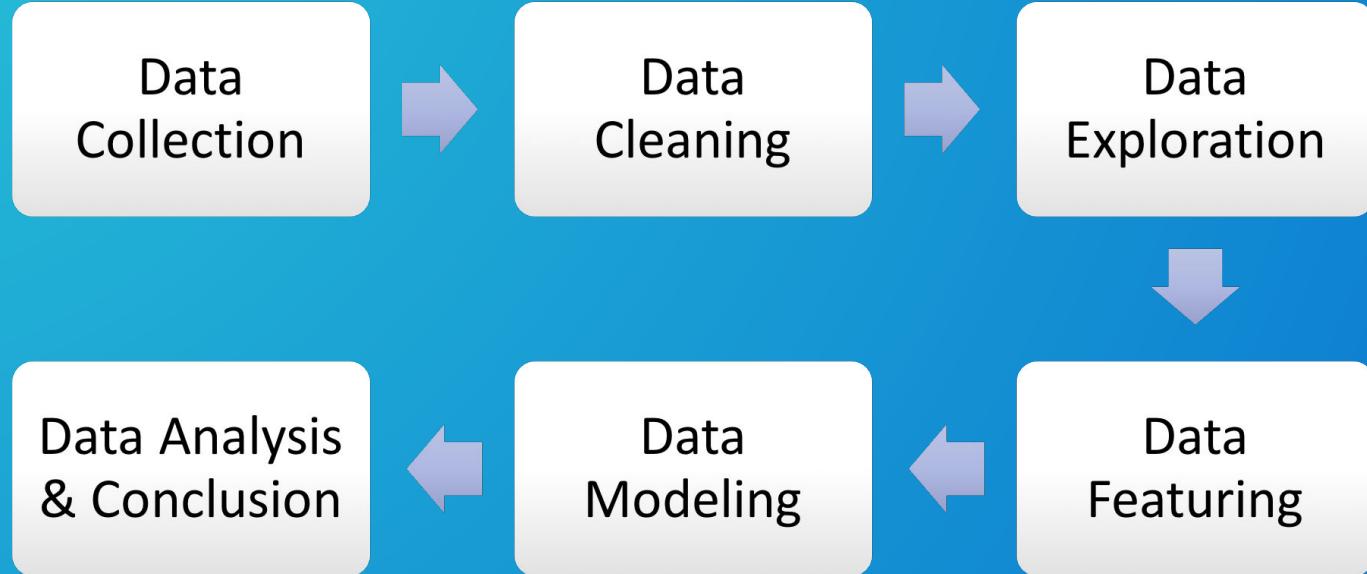


400 Million Tweets are sent per day



Over 4 Billion+ Hours of Video are watched on Youtube per month

Overall Scheme of Analysis



Data Collection

- Internal Data of Companies
- Purchasing Data
- Open Source: Kaggle, Github, Data.gov, and etc
- Open API: Tweepy API

Per User or Per Application

Rate limiting of the standard API is primarily on a per-user basis — or more accurately described, per user access token. If a method allows for 15 requests per rate limit window, then it allows 15 requests per window per access token.

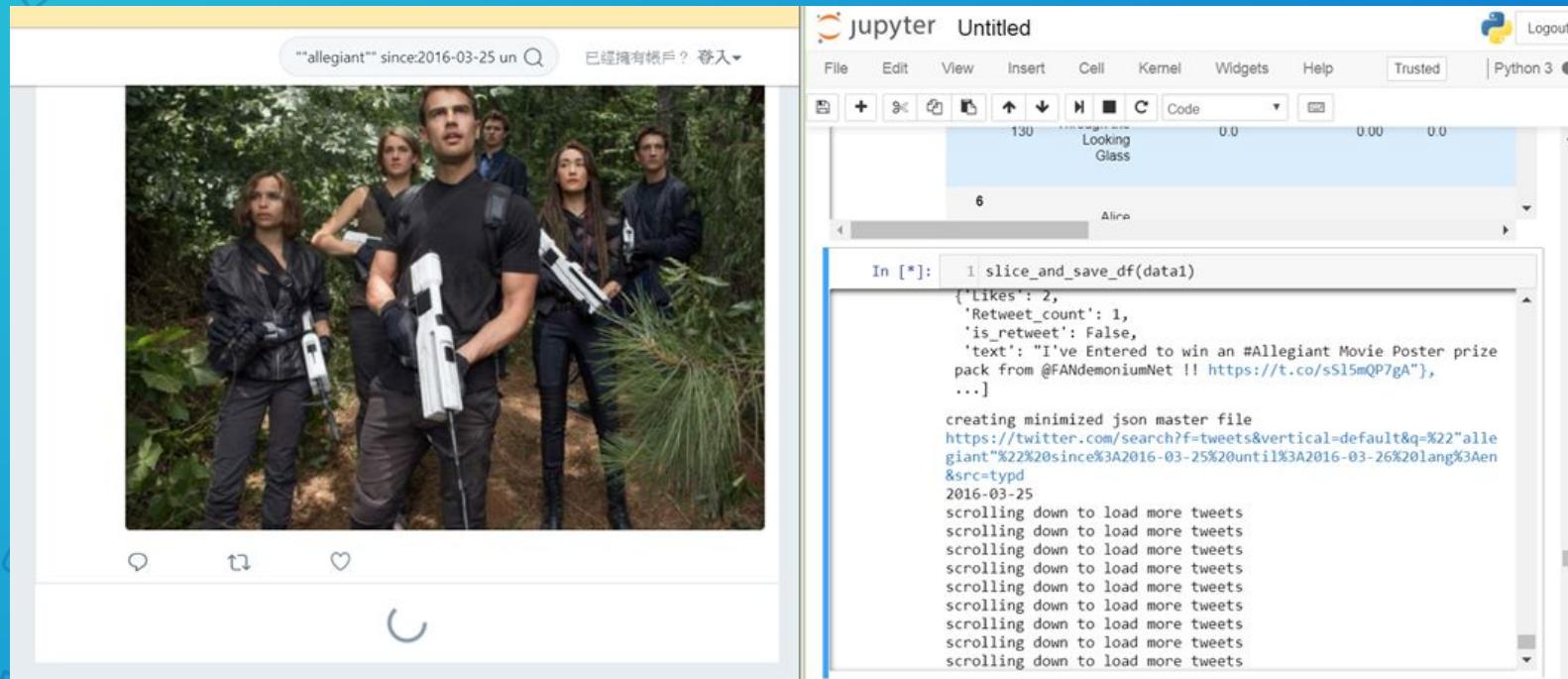
When using [application-only authentication](#), rate limits are determined globally for the entire application. If a method allows for 15 requests per rate limit window, then it allows you to make 15 requests per window — on behalf of your application. This limit is considered completely separately from per-user limits.

15 Minute Windows

Rate limits are divided into 15 minute intervals. All endpoints require authentication, so there is no concept of unauthenticated calls and rate limits.

There are two initial buckets available for GET requests: 15 calls every 15 minutes, and 180 calls every 15 minutes.

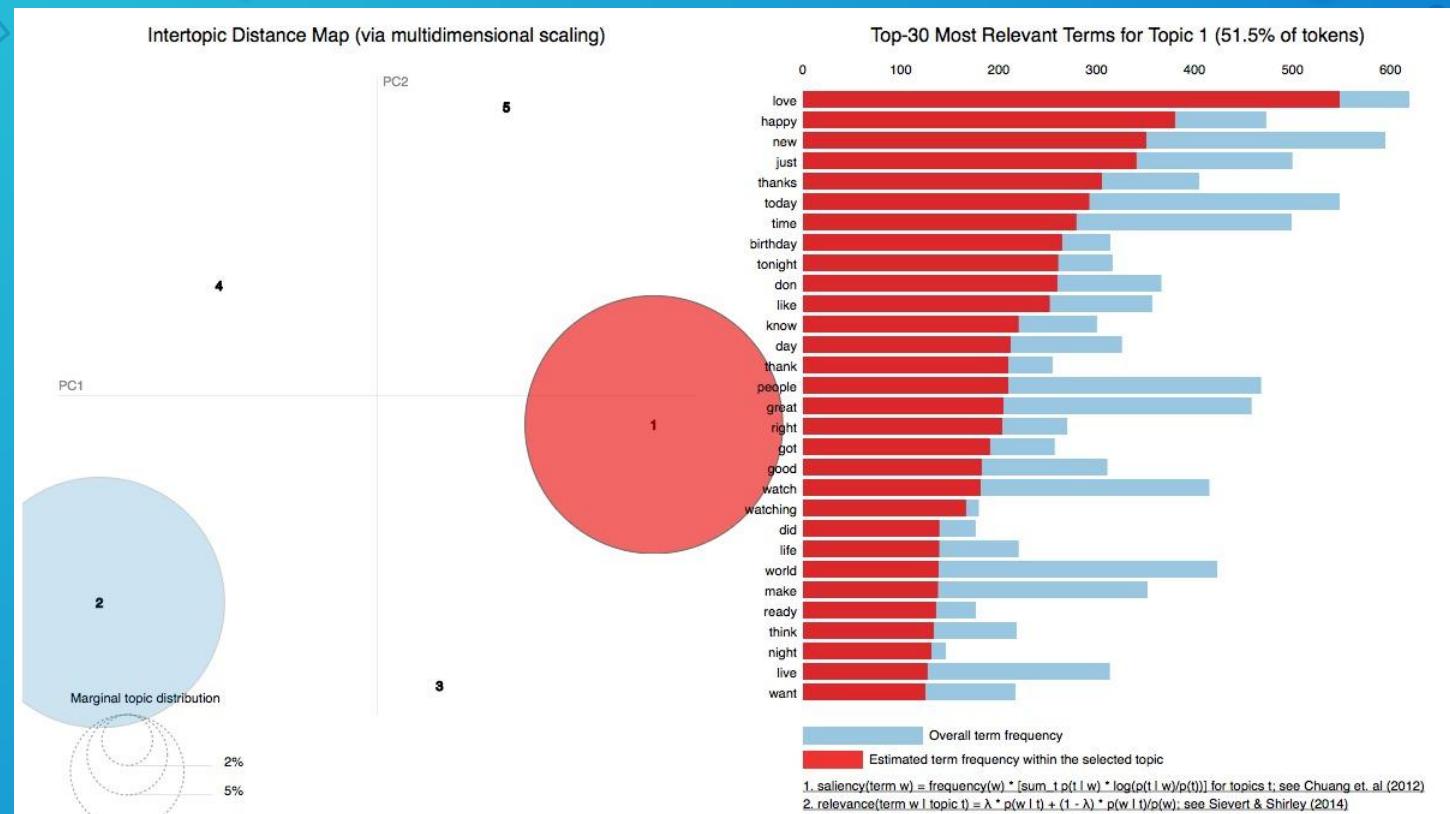
Data Collection - Web Crawler



Data Cleaning

" RT @Alesso: My babies @lelepons @HannahStocking <https://t.co/HoFsQCgV7s> RT @Georgia64377763: Are We Frenemies?! | @lelepons, @HannahStocking & ; @anwar_cyzLHA via @YouTube RT @erniesomblingo: @lelepons ARE WE FRENEMIES @HannahStocking?! <https://t.co/Lt6Jo3sPTw> <https://t.co/5Ewumm9AOY> 😂😂😂😂 i love it RT 'I hate when that happens' hahahahaha @lelepons <https://t.co/MYqqjIAAxm> RT @BradandBran: #AreWeFrenemies @lelepons @HannahStocking 😂😂😂 | <https://t.co/fKS1ZFeeFM> RT @loxspinosa: Are We Frenemies?! | @lelepons, @HannahStocking & ; @anwar <https://t.co/zZ50CwDaDo> via @YouTube RT @shots: ARE WE FRENEMIES with @HannahStocking & ; @anwar via Shots Studios: <https://t.co/kdnvSXfzJm> <https://t.co/CUic2iGdFU> RT @JavaddDrew0: Are We Frenemies?! | @lelepons, @HannahStocking & ; @anwar <https://t.co/vX3mV6GYmc> via @YouTube ARE WE FRENEMIES @HannahStocking?! <https://t.co/XSqb8JP7Q0> <https://t.co/KiYOpn4Sd4> 9 million youtube subscribers so much 😂😂😂 LOVE YOU!! <https://t.co/Frg3a2yEht> LA PIZZA COÑO!!!!!! 😊 <https://t.co/hXVwNaD14W> New music!!! <https://t.co/DpgZFV3Fc5> RT @atcxmoran: @lelepons new video from @lelepons 🎉🎉 <https://t.co/DPG0HPRJuN> RT @nataliegarcia00: @lelepons 😂 so did you end up staying at the university of hair removal 🎉🎉 y? | Lele Pons <https://t.co/iWynlQTGSF> @lelepons ❤️❤️❤️ WATCH THIS VIDEO!!!!!!❤️❤️❤️ RT @LilaLoveBeauty: I'm in a Sorority? <https://t.co/4wNRQLPSME> via @YouTube @lelepons RT @shots: I'M IN A SORORITY? by @lelepons via Shots Studios: <https://t.co/rbxBmqc8GU> <https://t.co/mPPFH61Bma> I'M 😂 GO check out my new YouTube video!! <https://t.co/s9HqoKYiuV> <https://t.co/GlvtjswwQy> My baby @HannahStocking's new video for @poobear and @justinbieber <https://t.co/bzxfDtkDw> Praying for my YouTube family. My heart is with everyone at @YouTube. Also praying for the day when this violence ends. <https://t.co/Wrj> <https://t.co/T2jiARLYqB> RT @garcesjinky13: Funny scenes 😂 i love you so bad lele bunny!!!! @lelepons <https://t.co/6kh5dEsMDX> <https://t.co/ChLvIKoF6> RT @atcx ATCH NEW VIDEO @lelepons <https://t.co/R46TA149By> RT @seokjincabeyoo: I'm the Easter Bunny? | @lelepons !!! \n<https://t.co/WFD6K0dpDL> RT @serrano_dinny: Love unny 😊\nNew video of @lelepons Go watch it noowww!! <https://t.co/Y2zW3NjpRr> RT @KaitlynGarnick: @lelepons is the Easter bunny 🐰 😊 <https://t.co/2CYHvc0GCJ> AZWaak RT @GeralMateoMus: @lelepons I laughed a lot watching this video 😂😂😂 ❤️ <https://t.co/8XL1T78fyAF> RT @marouabieber94: Where's the golden egg? 😊 @lelepons RT @pTR1TWhE2j RT @shots: I'M THE EASTER BUNNY? by @lelepons via Shots Studios: <https://t.co/TrcjXvilWT> <https://t.co/NkuFO1M32m> RT @Ka_Macas: Find the bunny! \n<https://t.co/5yoUAJvw23> <https://t.co/kUIMlxI18V> RT @EonlineLatino: Conoce todos los detalles de la increíble fiesta de cumpleaños que @lelepons le dio a @lelepons <https://t.co/nMOZc...> RT @justjaredjr: .@lelepons threw a surprise party for @ElJuanpaZurita's birthday last night!\n\n<https://t.co/jctytZONPG> RT @mikecostello: g custom michael Costello "Lele" gown last night she looked so elegant . Hair... <https://t.co/uOrXrpYs56> I'M THE EASTER BUNNY?? 🎉 Go check it out!! <https://t.co/mkCRpcjBLG> RT @lelemybae: HAPPY BIRTHDAY JUANPA!!! 😊❤️🎈🎊🎈 <https://t.co/imm9Z9Y1Bc> RT @HannahStocking: Happy ita 🎉🎉 <https://t.co/9ADQRFBQCI> RT @HannahStocking: Love you @john @lelepons 🎉 <https://t.co/847Pz70XIX> UPDATED MY PLAYLIST!!! @Anitta YOUR NEW SONG IS AMAZING <https://t.co/0bVWY1aI5w> Thank you Billboard! Watch my 5 tips!!! <https://t.co/H5wz7S0VYV> RT @NoteToSelf: Who is the real @lelepons? The Youtube & ; Instagram megas open with @manoushz about how she constructs her d.. Updated!! Send a screenshot if you're following my playlist for a follow back!! ❤️ [@manoushz @lelepons @NoteToSelf "\[My followers\] remember before I had my nose job and then after my nose job. They've been with... RT @matthunter123: Tenemos eado....pero muy muy grande....@lelepons ❤️ <https://t.co/zYCbosYPOB> RT @TheCut: .@manoushz interviews @lelepons on finding viral fame at a young age, the rays in her videos, and "editing \[h... I talked to @manoushz and @TheCut about why I edit all my own videos, for their new series No Filter: Women Owning It Online. To check it out!! <https://t.co/T5xzyE31ly> RT @nymagPR: This week the Cut is teaming up with @NoteToSelf on a series called No Filter: Women Owning It Online. To be sure... RT @TheCut: A day in the life of YouTube's reigning teen queen, @lelepons <https://t.co/ycp47WAB1z> RT @billboarddance: Here's how @lelepons gets the per... 🎉 <https://t.co/fpSXk2Sgut> RT @JS_JZ_MB_MS: LA FIESTA NO TERMINA NUNCA\n@ElJuanpaZurita @lelepons SON UNOS DUROS <https://t.co/yQDp2NqsUq> Me at #UltraMusicDay 😊 <https://t.co/00brEu1Qea> RT @JS_JZ_MB_MS: LA VIDA SIN FIESTA NO ES VIDA\n@ElJuanpaZurita @lelepons <https://t.co/y6FAtuvWIp> RT @negy19: ULTRA !!!! BEAUTIFUL @lelepons ❤️ <https://t.co/fsfCrdcv4y> RT @_salazart: Fiesta!!!! @ElJuanpaZurita @lelepons @Twankuyper <https://t.co/ItvvXmh3id> RT @KingBach: My guys are killing the management game! Proud of them! <https://t.co/3uKd6cbPmM> RT @Ka_Macas: THIS VIDEO WAS VERY FUNNY @lelepons @anwar 😂😂😂 I loved!!\n\n<https://t.co/00brEu1Qea>](https://t.co/9yioQ)

Importance of Data Cleaning



Word Cloud

BENEFITS:

- Provides insights into qualitative data
- Helps to identify trends & patterns
- Easy to understand
- Easy to share

SHORTCOMINGS:

- Gives only a high level picture of what lies inside the data
- Need to self-analyze the data
- Not effective for exploring complex topics

2003



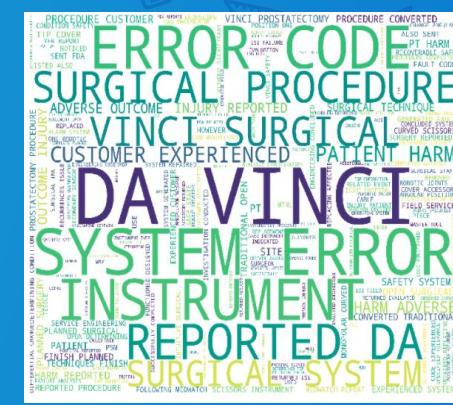
2004



2005



2007



CONCLUSION RESTARTED ALIVE
CAUTERY DEVICE CAUSED
ADVERSELY OUTCOME MATCH RIGHT TAG
INSTRUMENT ALL EGED OCCURRED
INSTRUMENT BASED VETMENT VERSUS
EVENT RELATED RECEIPT PERMANENT COMPONENT
CASE ARTERY PLASTIC FAULTED FAULTED
INTUITIVE PROVIDED EXCHANGE FAULTED
COMPLETED PROVIDED RECEIVED FAULTED
INJURY TOTALLY FAULTED
ADVERSELY RECEIVED FAULTED
PROCEDURE HOOK REMOVED CURRENTLY APPAROSCOPIC PT CONVERTED THROTTLED
REPORT ROUND UNABLE REPORT 5MM INCISION INSTRUMENT
REMOVED 5MM INCISION INSTRUMENT
PERFORMED MONORAIL SURGERY
REMOVED REPORT SURGERY



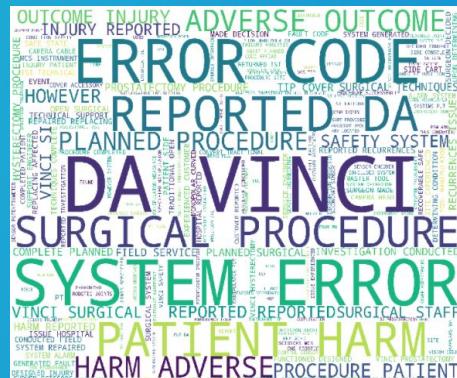
DA VINCI LEFT SYSTEM USED ALLEGED
ROBOTIC TECHNIQUES JAMES FUCHS PT REPAIR
DUE CONVERTED FAILURE LEAFLET
VALVE USE PISTON NOTED DAY REMOVAL
RETRIEVED CYLINDER SCHEDULED SINCE SURGEON
PROSTATECTOMY INJURIES
LENDING PHYSICIAN TERMINALIZED
PIECE MISSING INVENTORY REMOVING
ONE BROKE PESTLED BLADE EVENT
MISSING REMAINING
SURGERY OPEN SCAR NEW
WENT BROKE LISTING ADOLTE SHOCK
ABDOMEN OPENLY EXPLORING INFORMED MATCHED
ADDITIONALLY PROCEDURE BYPASS PROBLEMS
YESTERDAY FILTERED CATHETER TODAY INTERVENTION
NOTICED FIELD CISTER
MALFUNCTION INSTRUMENT INVENTIVE
EXACERBATED PERIODIC FLUOROSCOPY INVOLVING
CLOGGED CLOGGED INVESTIGATE
ADVERSE ERECTILE DYSFUNCTION



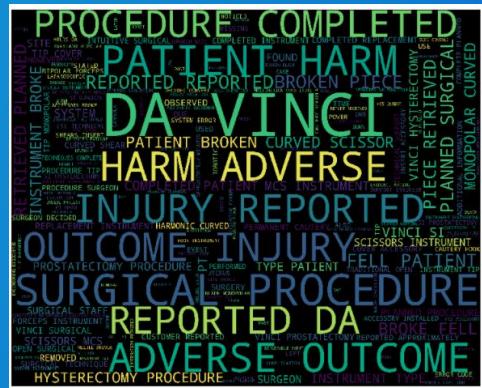
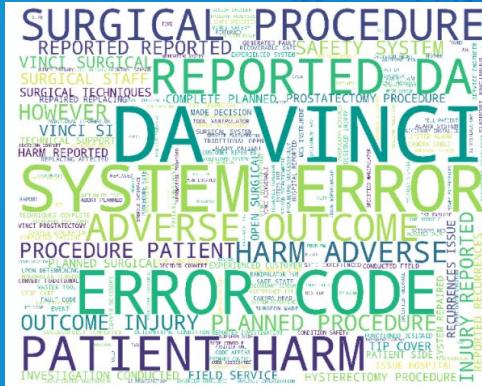
2008



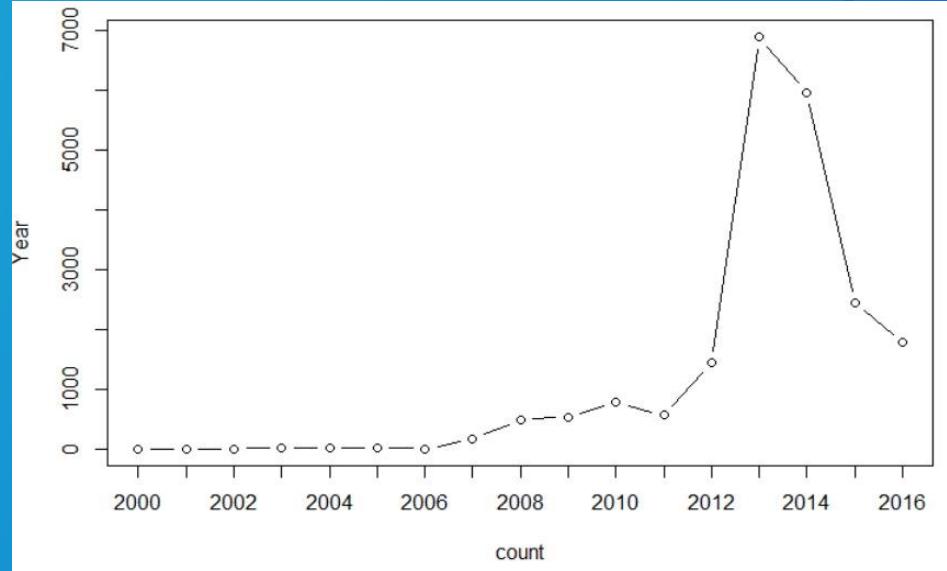
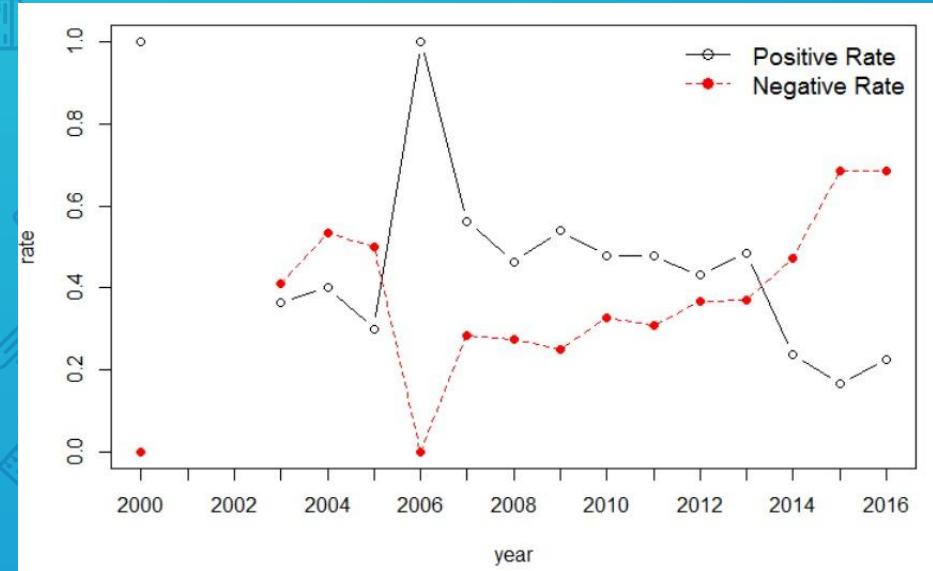
2009



2010



Word Cloud Analysis



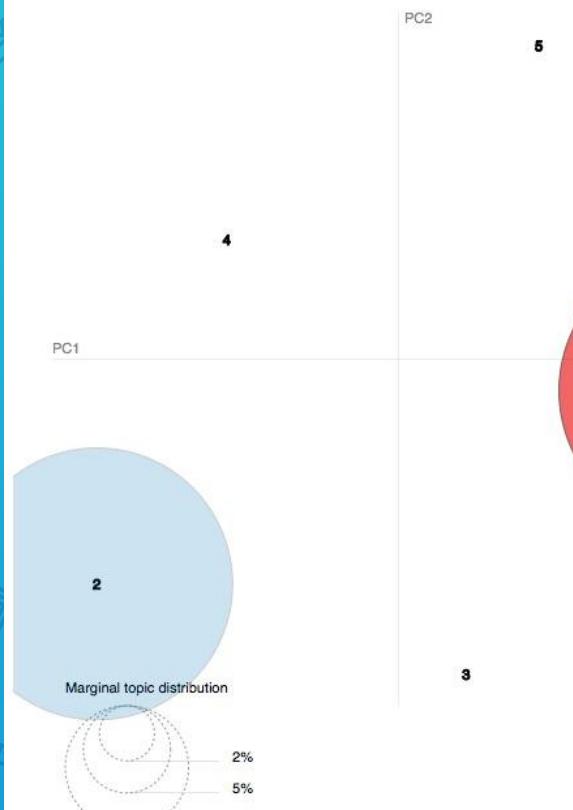
Selected Topic: 0 Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:(2)

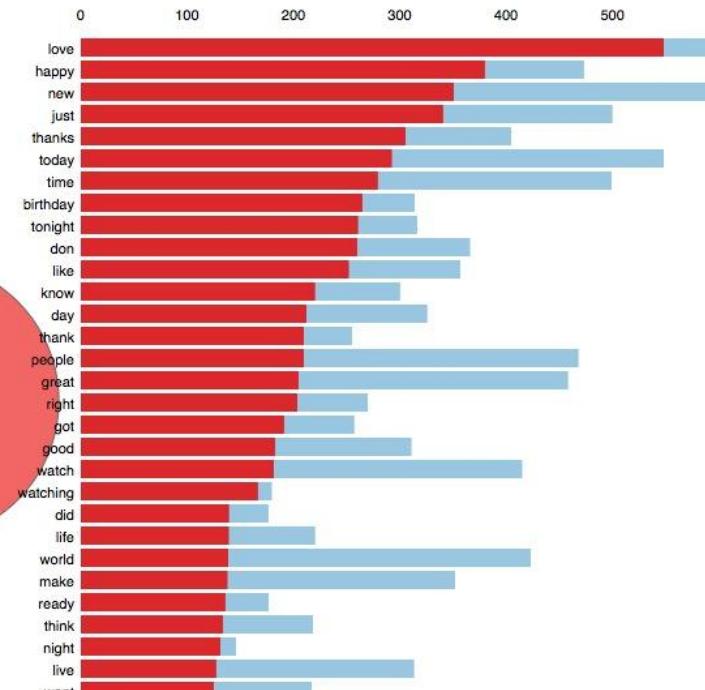
$\lambda = 1$



Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 1 (51.5% of tokens)



Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t|w) * log(p(t|w)/p(t))] for topics t; see Chuang et. al (2012)

2. relevance(term w | topic t) = $\lambda \cdot p(w|t) + (1 - \lambda) \cdot p(w|t)/p(w)$; see Sievert & Shirley (2014)

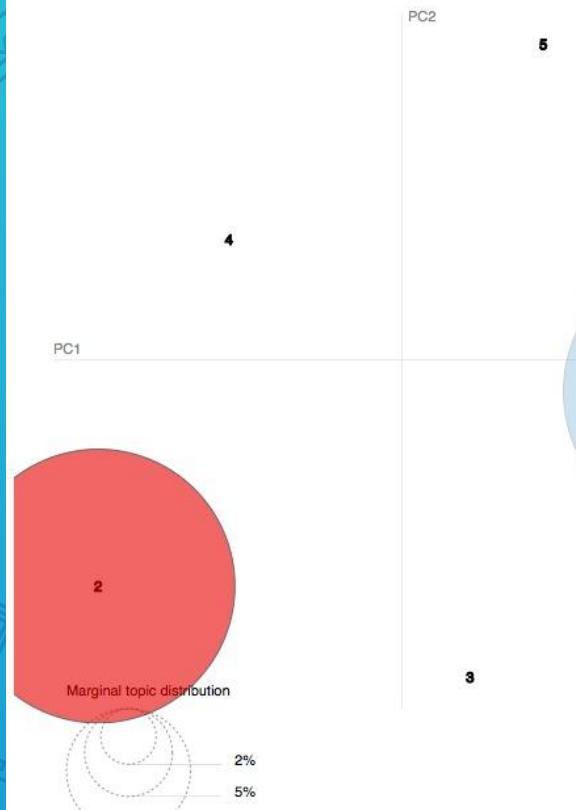
Selected Topic: 0 | Previous Topic | Next Topic | Clear Topic

Slide to adjust relevance metric:(2)

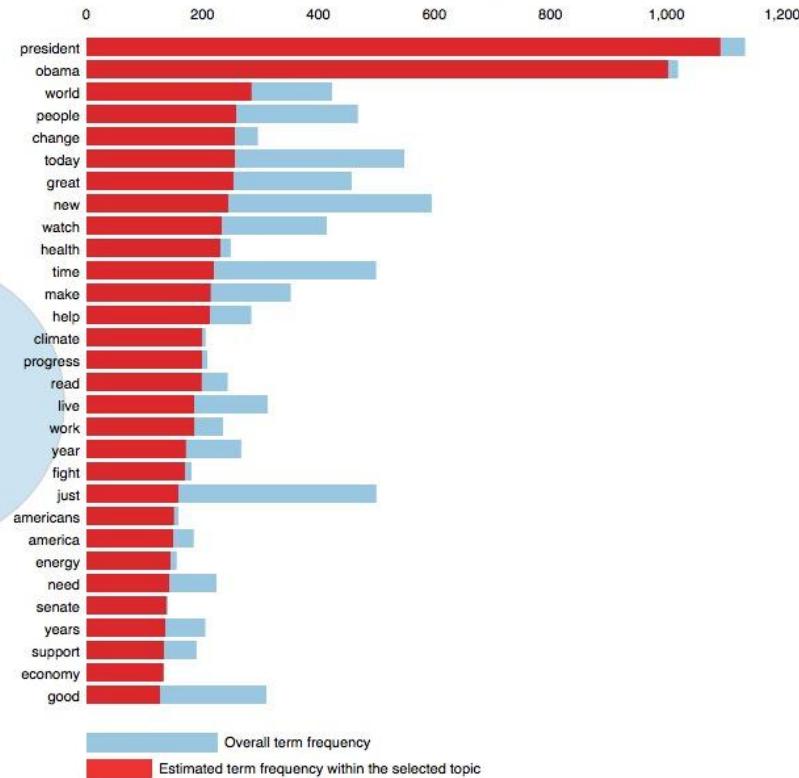
$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 2 (48.5% of tokens)



Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t|w) * log(p(t|w)/p(t))] for topics t; see Chuang et. al (2012)

2. relevance(term w | topic t) = $\lambda * p(w|t) + (1 - \lambda) * p(w|t)/p(w)$; see Sievert & Shirley (2014)

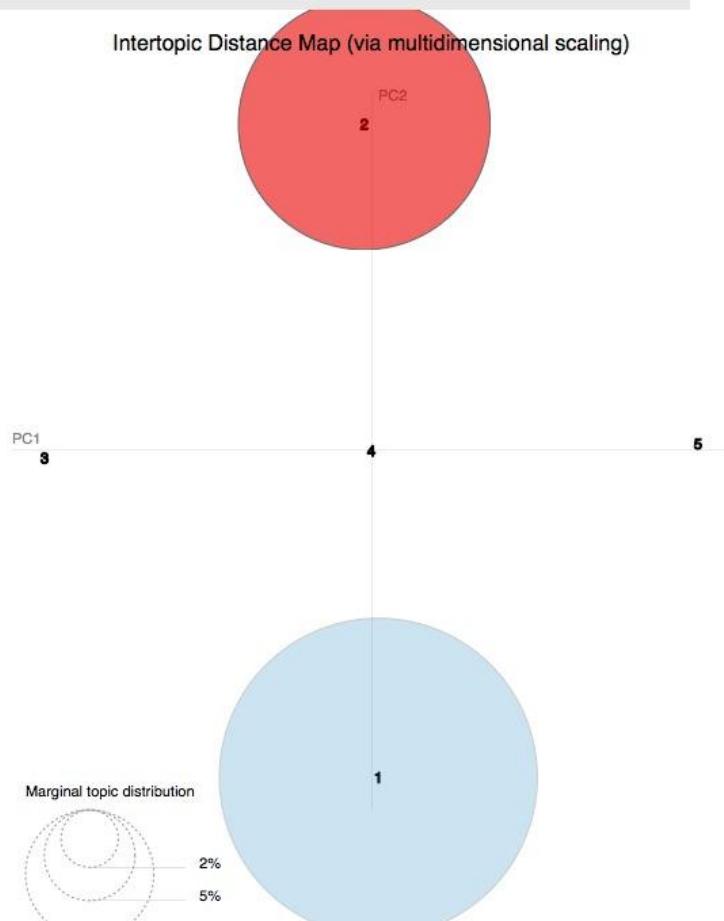
Selected Topic: 0 Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:(2)

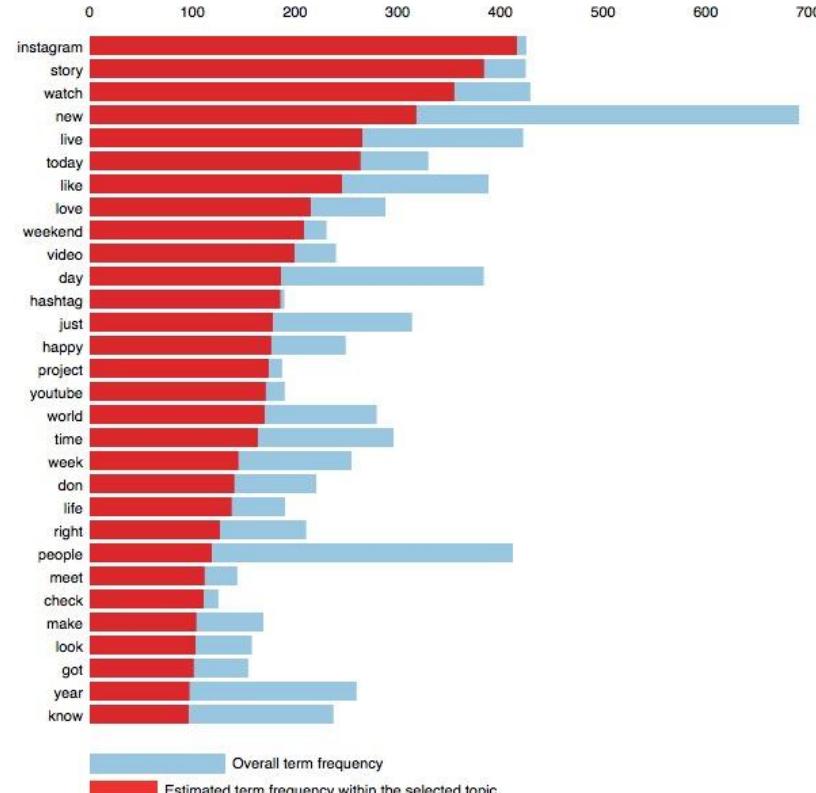
$\lambda = 1$



Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 2 (38.5% of tokens)



1. saliency(term w) = frequency(w) * [sum_t p(t|w) * log(p(t|w)/p(t))] for topics t; see Chuang et. al (2012)

2. relevance(term w | topic t) = $\lambda * p(w|t) + (1 - \lambda) * p(w|t)/p(w)$; see Sievert & Shirley (2014)

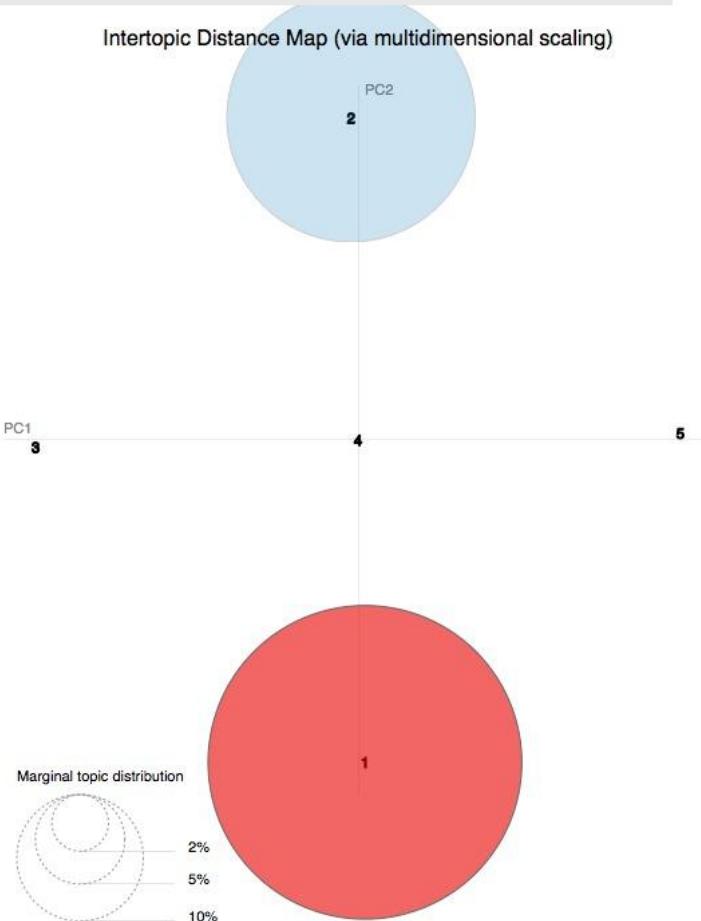
Selected Topic: 0 | Previous Topic | Next Topic | Clear Topic

Slide to adjust relevance metric:(2)

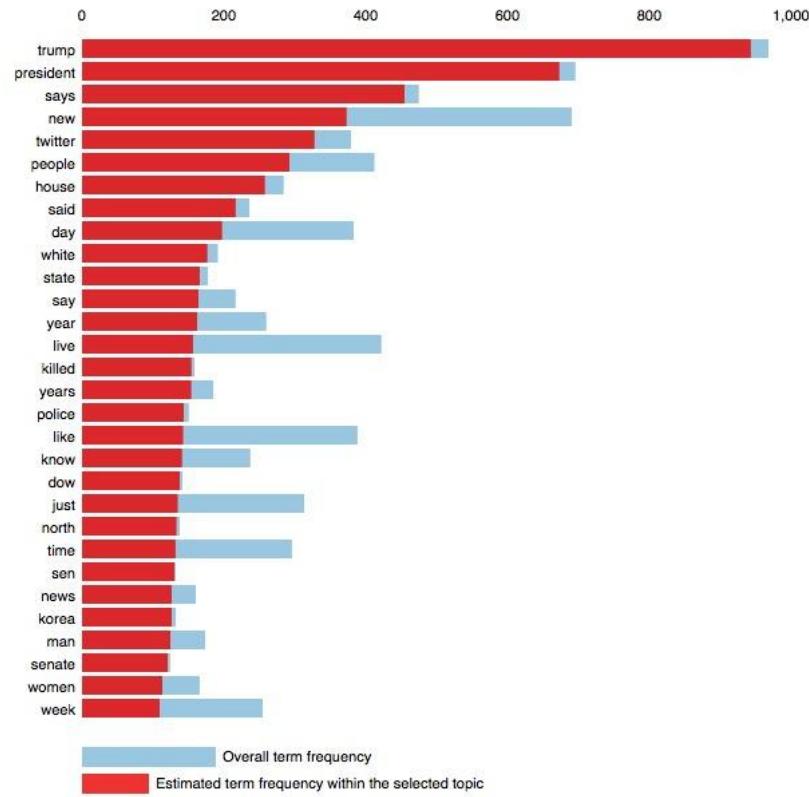
$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 1 (61.5% of tokens)



1. saliency(term w) = frequency(w) * [sum_t p(t|w) * log(p(t|w)/p(t))] for topics t; see Chuang et. al (2012)

2. relevance(term w | topic t) = $\lambda \cdot p(w|t) + (1 - \lambda) \cdot p(w|t)/p(w)$; see Sievert & Shirley (2014)

Conclusion

Purpose of Analysis

Sense and respond to market on a real-time basis

Analyze customer data to reflect on future organization changes

Critical consideration of usage of big data for deriving managerial benefits

Challenge and Limitation

In real practice, data preparation (collection and cleaning) is a lot of time considered the bottleneck

- time consuming
- raise concern in data security

Narrow scope and variety of insights drew from social media or sentiment analysis

Difficult to come up with selection criteria of tools to analyze specific problems

- One tool not for all

Facebook and Cambridge Analytica Data Scandal

(March 2018)

- Increasing negative Tweets
- Almost none “likes”

