



Big Data untuk Penelitian Sosial era COVID-19

- MD. Enyat M -

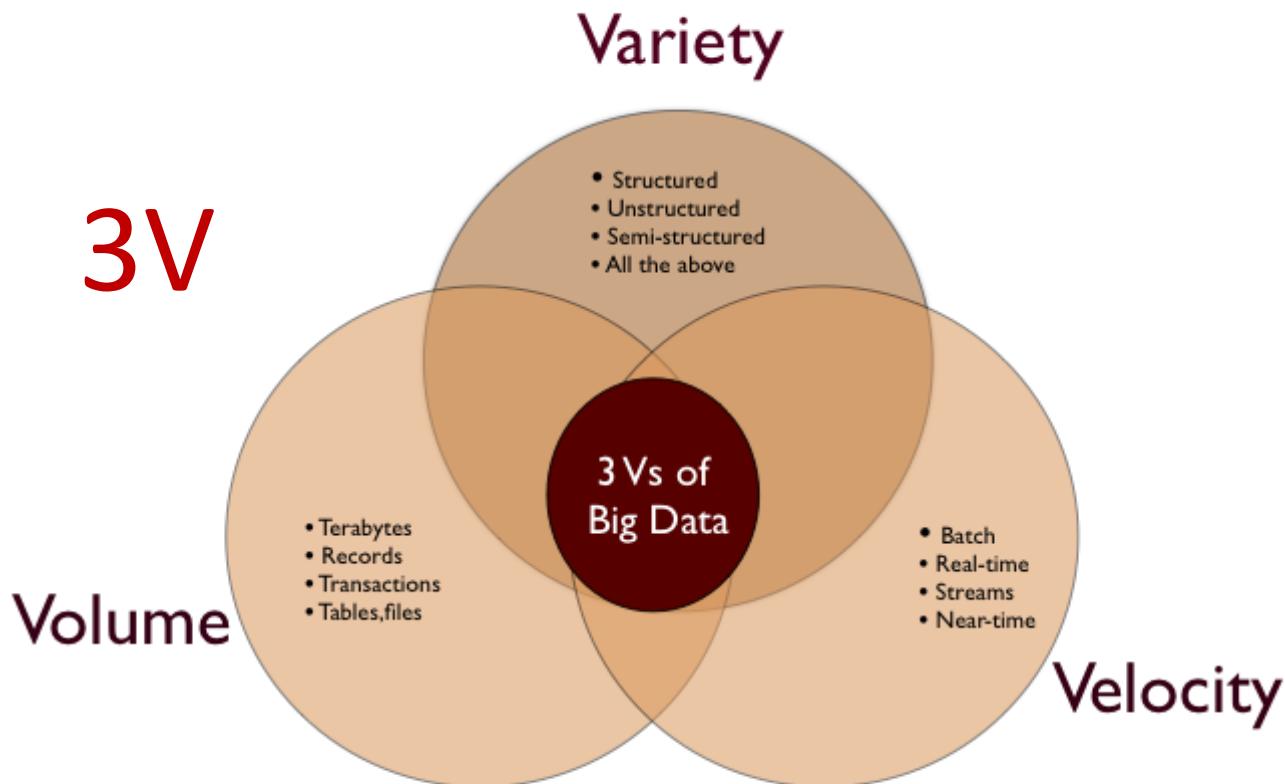
A dense cloud of small, semi-transparent text labels in various colors (purple, yellow, blue, red, green, etc.) is arranged in a circular pattern at the bottom of the slide. Some of the visible words include 'transformation', 'human', 'right', 'new my', 'declarator', 'articles', and 'author seva'. These words likely represent themes or hashtags related to the topic of the presentation.



Latar Belakang

1. The Fall of Mobility, The Rise of Stay @ Home
2. **Online-Shopping** Widening+Deepening: From Wants to Needs
3. **Food Delivery**: From “Indulgence” to “Utility”
4. The Comeback of **Home Cooking**
5. **Frozen Food**: Convenience Solution
6. **Going Omni**
7. **Subscription Model** Matters
8. **TV Strikes Back**
9. **DIY & Self-Care @ Home**
10. **Zoomable Workplace @ Home**
11. “Work-Live-Play” Balance: Well-Being Revolution
12. The Century of **Self Distancing**
13. Contact-Free Lifestyle
14. Low-Trust Society
15. Constantly-Fear Customers
16. Jamu Is the New Espresso
17. Halal (Thoyyiban) Becomes Mainstream
18. **Paylater Solution**
19. **The Future of Traveling**
20. Virtual Experience Is the Nex Big Thing
21. The Emerging **VirSocial**
22. Flexible Working Hours: From “9-to-5” to “3-to-2”
23. The Birth of **Zoom Generation**
24. Cloud Lifestyle
25. Telemedicine: from Visit to **Virtual**
26. **Online+Home-Schooling**
27. **Ibadah Virtual**
28. The Rise of Empathy and Solidarity
29. From Drone Parenting to **Positive Parenting**
30. More Suffering, More Religious

Apa itu BigData

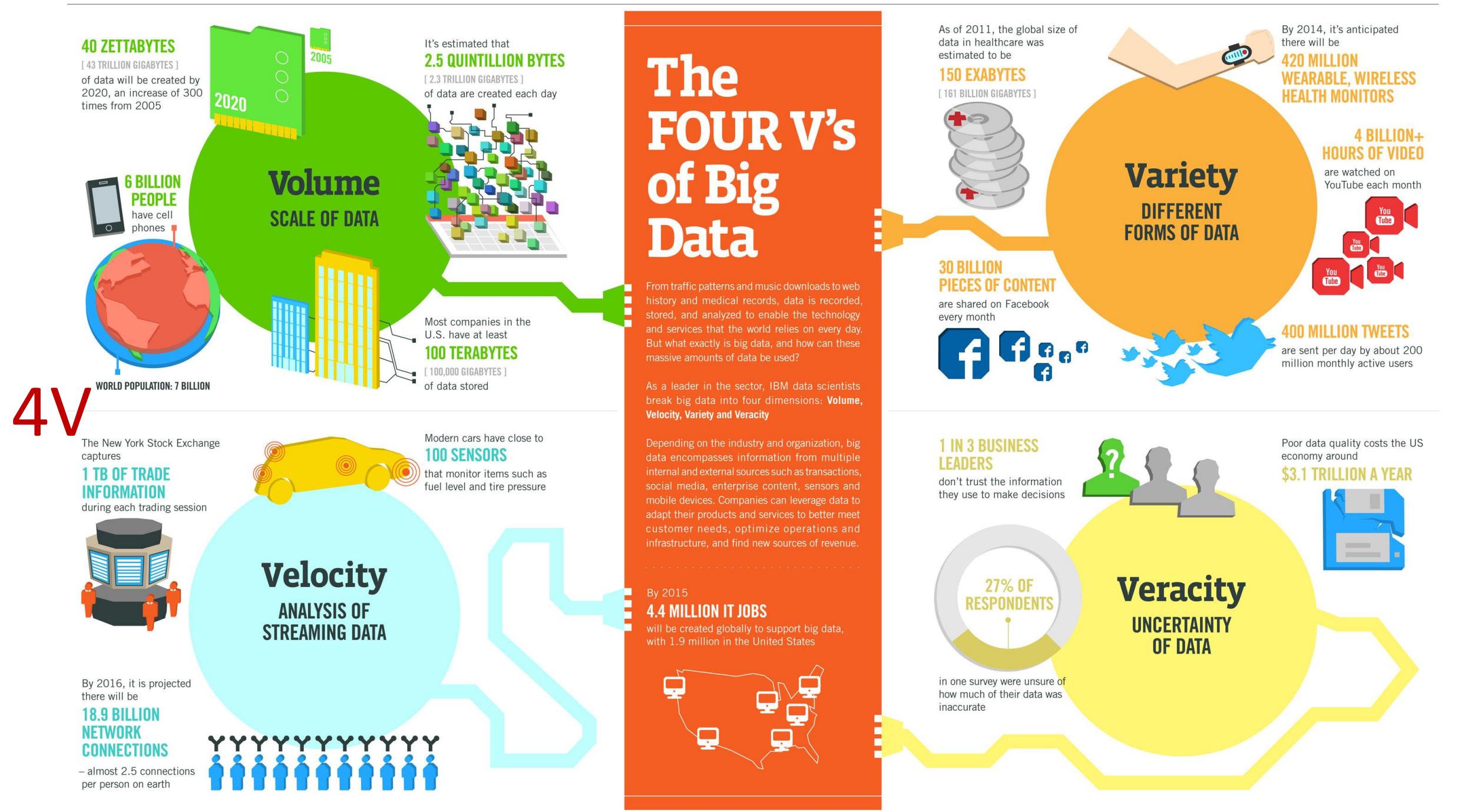


The Definition of Big Data
What exactly is big data?

To really understand big data, it's helpful to have some historical background. Here is Gartner's definition, circa 2001 (which is still the go-to definition): **Big data is data that contains greater variety arriving in increasing volumes and with ever-higher velocity. This is known as the three Vs.**

*Put simply, **big data is larger, more complex data sets, especially from new data sources**.* These data sets are so voluminous that traditional data processing software just can't manage them. But these massive volumes of data can be used to address business problems you wouldn't have been able to tackle before.

-Oracle, 2019-





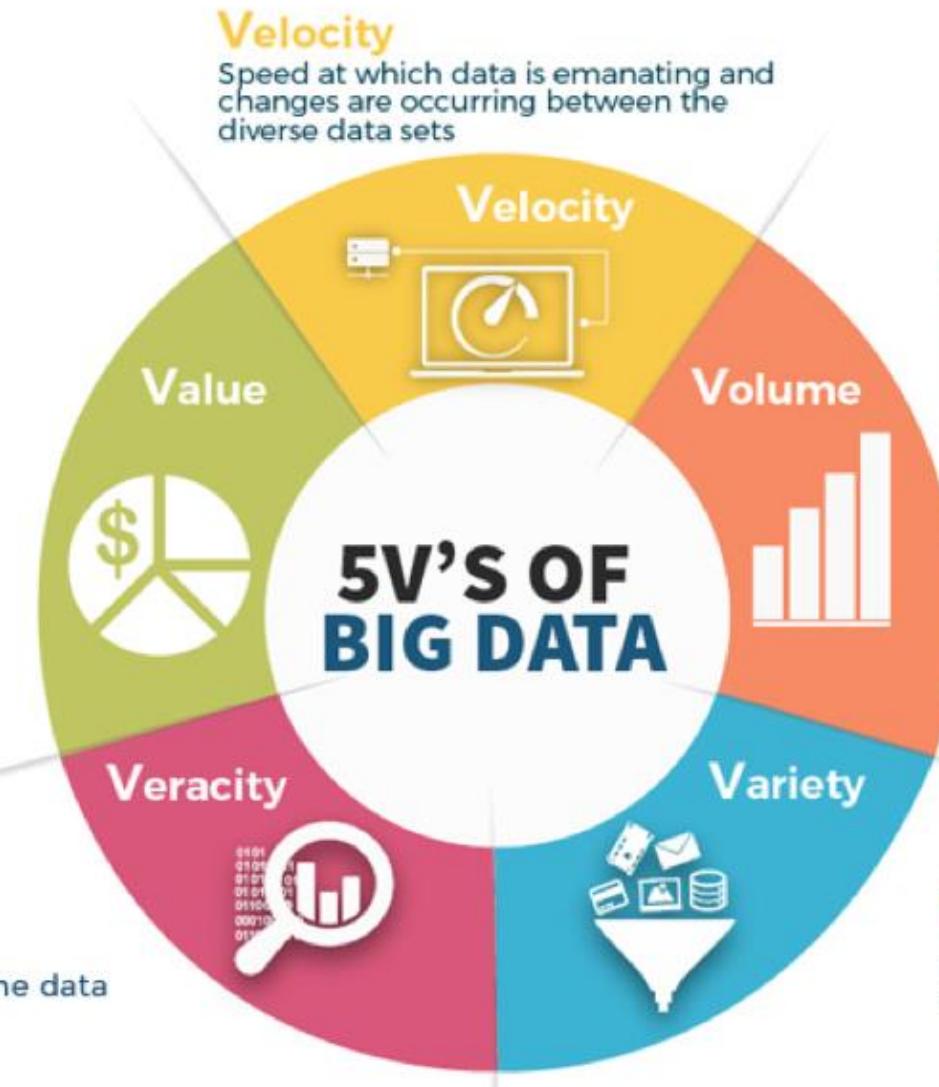
5V

Value

Having access to big data is all well and good but that's only useful if we can turn it into a value.

Veracity

Data reliability and trust.
Verifying and validating the data



Volume

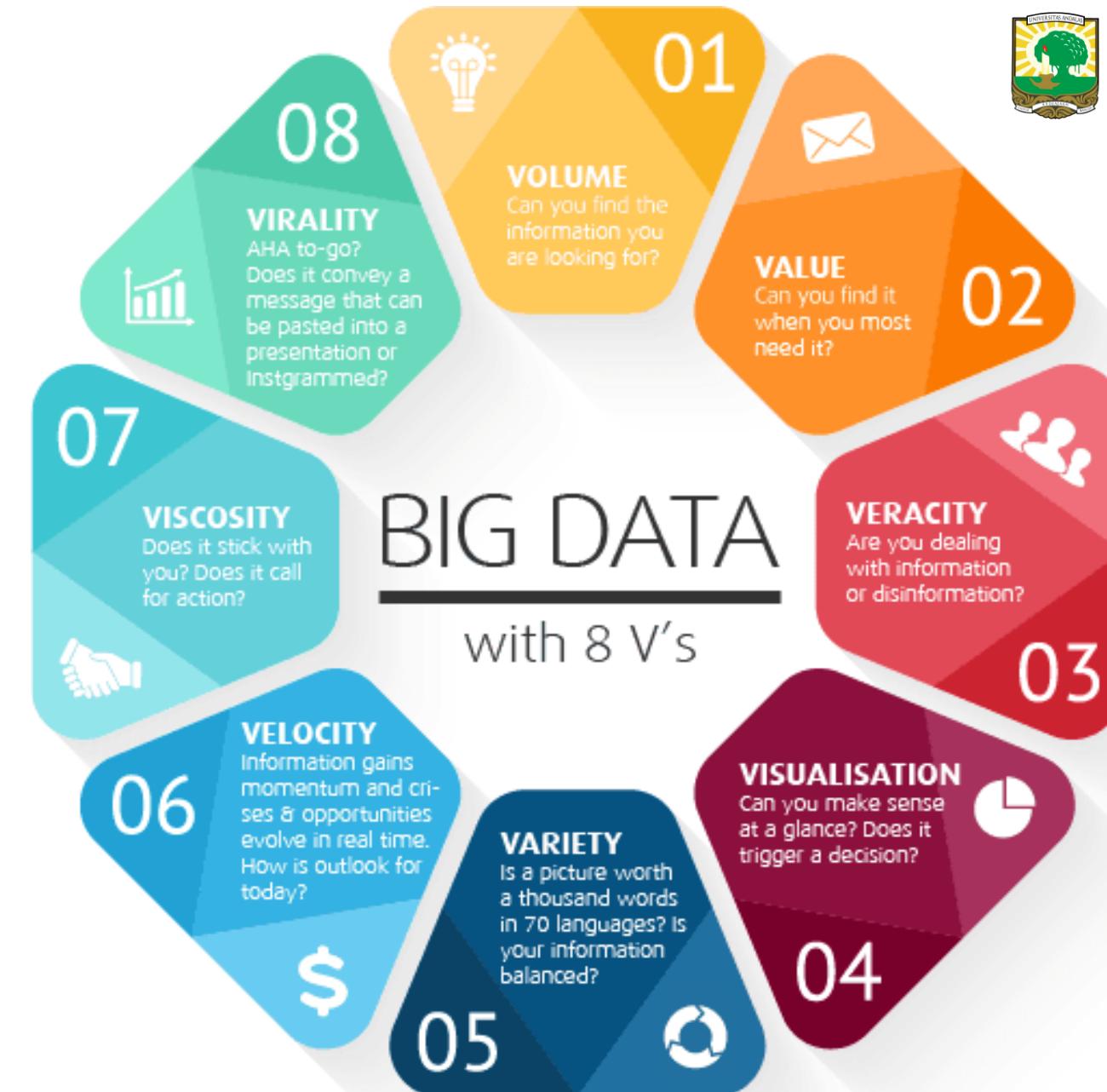
This refers to the sheer volume of data being generated every second.

Variety

Can use structured as well as unstructured data.



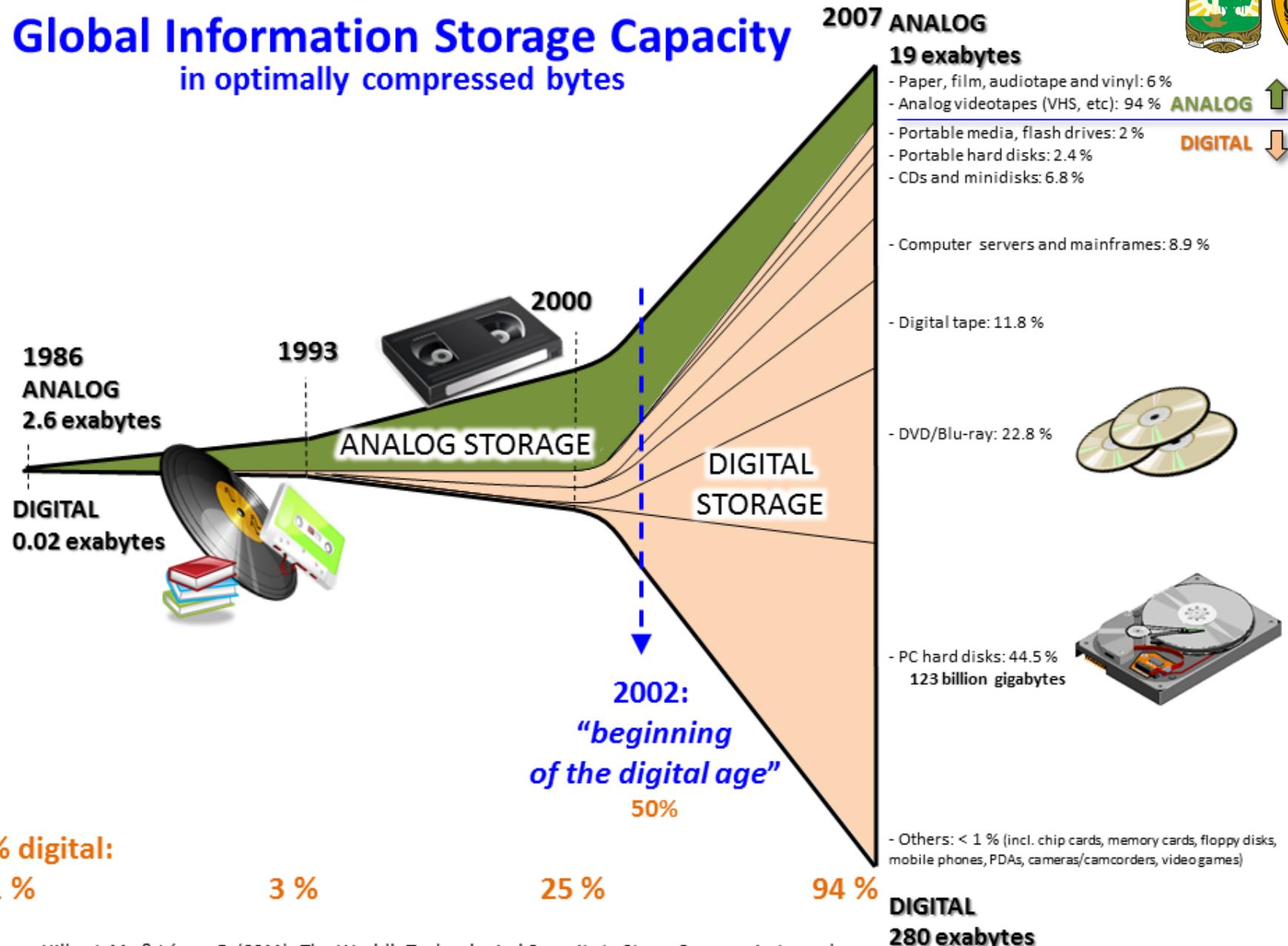
8V





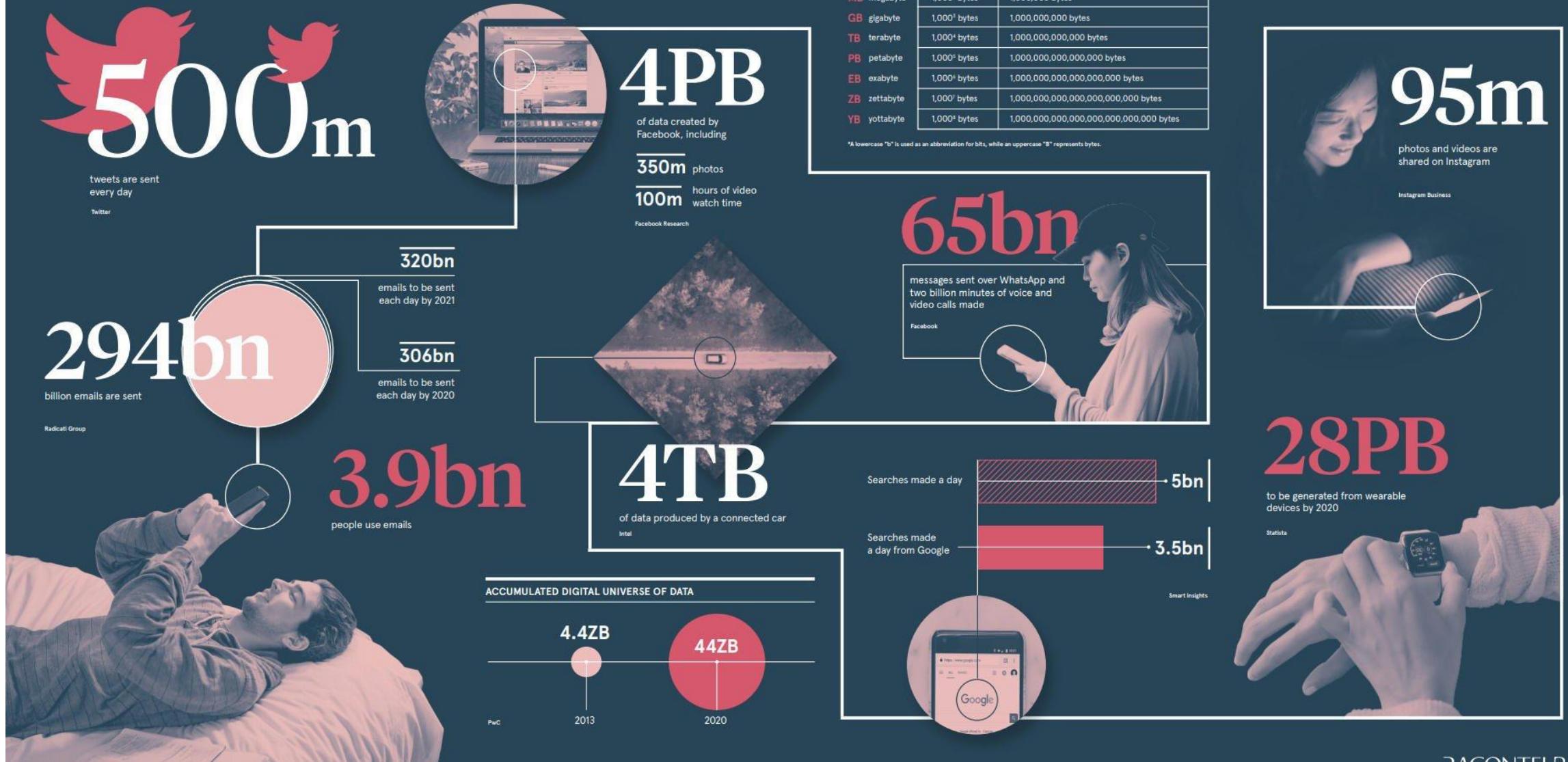
Global Information Storage Capacity

in optimally compressed bytes



A DAY IN DATA

The exponential growth of data is undisputed, but the numbers behind this explosion – fuelled by internet of things and the use of connected devices – are hard to comprehend, particularly when looked at in the context of one day



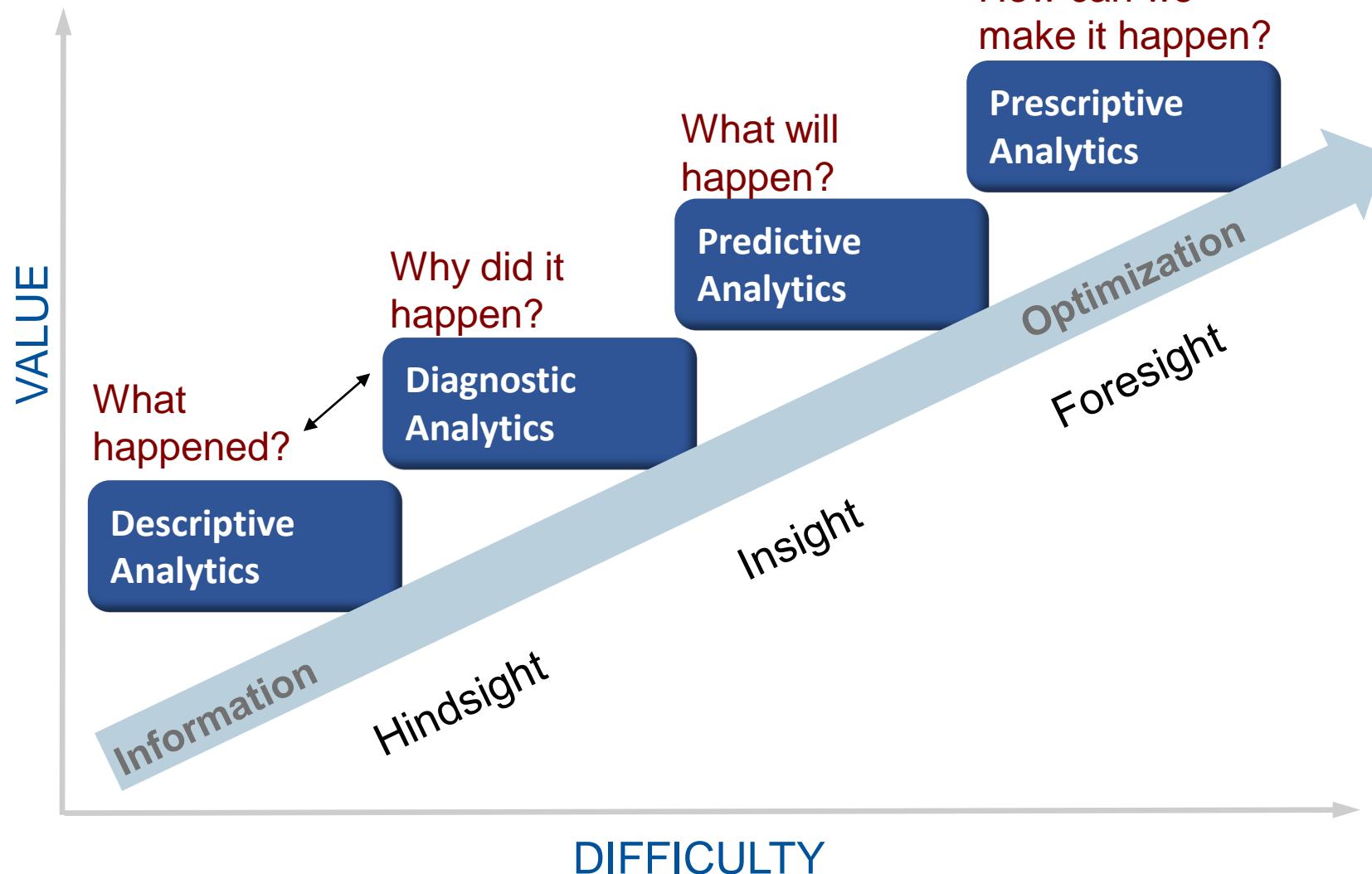
463EB

of data will be created every day by 2025

IDC



Analytics Models



APR
2020

FACEBOOK REACH RANKINGS

COUNTRIES AND TERRITORIES* WITH THE GREATEST POTENTIAL FACEBOOK ADVERTISING REACH

#	COUNTRY	REACH	▲ QOQ	▲ QOQ
01	INDIA	280,000,000	+8%	+20,000,000
02	U.S.A.	190,000,000	+6%	+10,000,000
03	INDONESIA	130,000,000	0%	[UNCHANGED]
04	BRAZIL	120,000,000	0%	[UNCHANGED]
05	MEXICO	86,000,000	+2%	+2,000,000
06	PHILIPPINES	72,000,000	+3%	+2,000,000
07	VIETNAM	63,000,000	+3%	+2,000,000
08	THAILAND	48,000,000	+2%	+1,000,000
09	EGYPT	41,000,000	+8%	+3,000,000
10	BANGLADESH	37,000,000	+9%	+3,000,000

52

SOURCE: EXTRAPOLATIONS OF DATA FROM FACEBOOK'S SELF-SERVICE ADVERTISING TOOLS (APRIL 2020). *NOTE: ONLY INCLUDES COUNTRIES AND TERRITORIES WITH POPULATIONS OF AT LEAST 50,000 PEOPLE. COMPARABILITY ADVISORY: BASE CHANGES.

APR
2020

TWITTER REACH RANKINGS

COUNTRIES AND TERRITORIES* WITH THE GREATEST POTENTIAL TWITTER ADVERTISING REACH

#	COUNTRY	REACH	▲ QOQ	▲ QOQ
01	U.S.A.	64,200,000	+8%	+4,850,000
02	JAPAN	48,450,000	+6%	+2,700,000
03	RUSSIA	23,550,000	+149%	+14,090,000
04	U.K.	17,750,000	+6%	+1,050,000
05	SAUDI ARABIA	15,000,000	+5%	+650,000
06	BRAZIL	14,350,000	+18%	+2,200,000
07	TURKEY	13,650,000	+16%	+1,850,000
08	INDIA	13,150,000	+15%	+1,700,000
09	INDONESIA	11,800,000	+11%	+1,155,000
10	MEXICO	10,200,000	+8%	+750,000

86

SOURCE: EXTRAPOLATIONS OF DATA FROM TWITTER'S SELF-SERVICE ADVERTISING TOOLS (APRIL 2020). *NOTE: ONLY INCLUDES COUNTRIES AND TERRITORIES WITH POPULATIONS OF AT LEAST 50,000 PEOPLE. COMPARABILITY ADVISORY: BASE CHANGES.

APR
2020

INSTAGRAM REACH RANKINGS

COUNTRIES AND TERRITORIES* WITH THE GREATEST POTENTIAL INSTAGRAM ADVERTISING REACH

#	COUNTRY	REACH	▲ QOQ	▲ QOQ		#	COUNTRY	REACH	▲ QOQ	▲ QOQ	
01	U.S.A.	120,000,000	0%	[UNCHANGED]		13	ARGENTINA	18,000,000	+6%	+1,000,000	
02	INDIA	88,000,000	+10%	+8,000,000		14	SPAIN	17,000,000	+6%	+1,000,000	
03	BRAZIL	82,000,000	+6%	+5,000,000		15=	CANADA	13,000,000	0%	[UNCHANGED]	
04	INDONESIA	64,000,000	+2%	+1,000,000		15=	SOUTH KOREA	13,000,000	0%	[UNCHANGED]	
05	RUSSIA	46,000,000	+5%	+2,000,000		15=	SAUDI ARABIA	13,000,000	+8%	+1,000,000	
06	TURKEY	37,000,000	+12%	+4,000,000							+1,000,000
07	JAPAN	37,000,000	0%	[UNCHANGED]							[UNCHANGED]
08	MEXICO	32,000,000	+3%	+1,000,000							+1,000,000
09	U.K.	30,000,000	+3%	+1,000,000							[UNCHANGED]
10	GERMANY	28,000,000	0%	[UNCHANGED]							[UNCHANGED]
11	ITALY	26,000,000	+8%	+2,000,000							+500,000
12	FRANCE	23,000,000	+10%	+2,000,000							

68

SOURCE: EXTRAPOLATIONS OF DATA FROM INSTAGRAM'S SELF-SERVICE ADVERTISING TOOLS (APRIL 2020). *NOTE: ONLY INCLUDES COUNTRIES AND TERRITORIES WITH POPULATIONS OF AT LEAST 50,000 PEOPLE. COMPARABILITY ADVISORY: BASE CHANGES.

Social media activities on select social networks by social media users in the United States in February 2019

Instagram Facebook Snapchat Pinterest Twitter

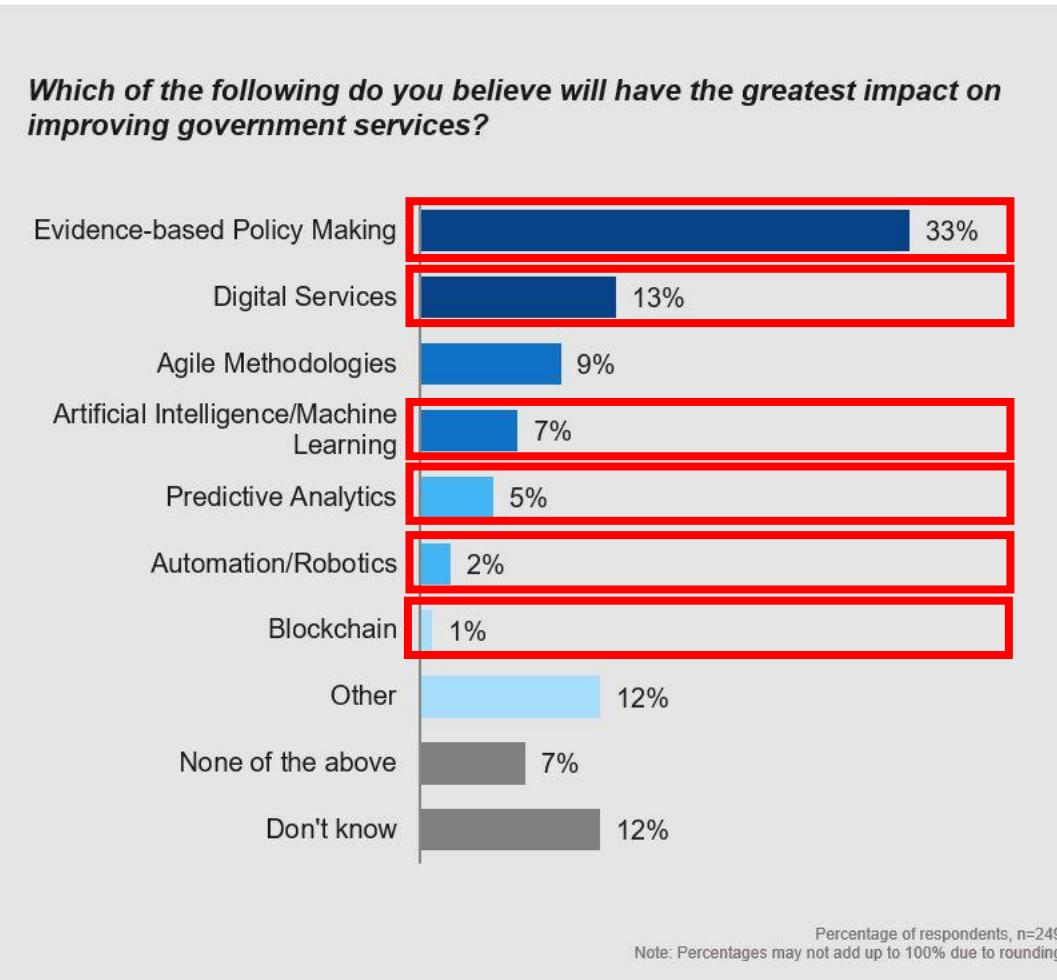


News	18%	38%	17%	9%	56%
Viewing photos	77%	65%	64%	59%	42%
Watching videos	51%	46%	50%	21%	32%
Sharing content with everyone	45%	57%	46%	21%	32%
Networking	23%	33%	21%	10%	26%
Sharing content one to one	31%	43%	45%	12%	20%
Finding/shopping for products	11%	15%	5%	47%	7%
Promoting my business	9%	7%	6%	5%	7%

Source: Statista



Latar Belakang



The survey received 245 responses from employees representing a number of federal agencies. The findings were conclusive — among the options presented, respondents identified **evidence-based policy making** as most likely to have the greatest impact on the improvement of government services.

Which of the following do you believe will have the greatest impact on improving government services?

IT Based Decision Making :

$$33\% + 13\% + 7\% + 5\% + 2\% + 1\% = \\ 61\%$$



Covid-19 → The New Normal di segala aspek...

1. The Fall of Mobility, The Rise of Stay @ Home
2. **Online-Shopping**
Widening+Deepening: From Wants to Needs
3. **Food Delivery**: From “Inconvenience” to “Utility”
4. The Comeback of **Home Cooking**
5. **Frozen Food**: Convenience Solution
6. Going Omni
7. **Subscription Model** Matters
8. TV Strikes Back
9. DIY & Self-Care @ Home
10. Zoomable Workplace @ Home
11. “Work-Live-Play” Balance: Well-Being Revolution
12. The Century of **Self Disinfecting**
13. Contact-Free Life Style
14. Low-Touch Society
15. Instantaneous Customers
16. Japan’s the New Espresso
17. Halal (and Urban) Becomes Mainstream
18. Paylater Solution
19. The Future of Traveling
20. Virtual Experience Is the Next Big Thing
21. The Emerging **VirSocial**
22. Flexible Working Hours: From “9-to-5” to “3-to-2”
23. The Birth of **Zoom Generation**
24. Cloud Lifestyle
25. Telemedicine: from Visit to **Virtual**
26. **Online+Home-Schooling**
27. Ibadah Virtual
28. The Rise of Empathy and Solidarity
29. From Drone Parenting to **Positive Parenting**
30. More Suffering, More Religious

WFH



Sumber Data



Sumber Data...

Primary Data

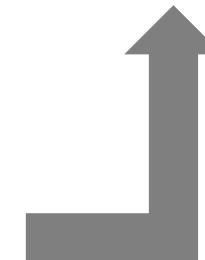
is **data that is collected by a researcher from first-hand sources**, using methods like surveys, interviews, or experiments. It is collected with the research project in mind, directly from primary sources.

Secondary data

is **data gathered from studies, surveys, or experiments** that have been run by other people or for other research.



Typically, a researcher will begin a project by working with secondary data. This allows time to formulate questions and gain an understanding of the issues being dealt with before the more costly and time consuming operation of collecting primary data.





Sumber Data Sekunder...

Website

- [Bps.go.id](https://bps.go.id)
- <https://www.kaggle.com>
- <http://hdr.undp.org/en/data>
- <https://datasetsearch.research.google.com>
- <https://covid19-public.digitalservice.id/api/v1/>
- https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv
- <https://catalog.data.gov/dataset>
- FiveThirtyEight
- BuzzFeed
- ProPublica
- AWS Public Data sets
- UCI Machine Learning Repository
- Quandl
- Dsb... dsb...

Media Sosial

- Twitter
- Facebook
- Instagram
- Whatsapp
- Dsb.. Dsb...

Note :

- *Mayoritas Data tersebut tidak siap pakai (API),*
- *Butuh proses download data → third-party apps..*
- *Butuh proses penerjemahan format*
- *Butuh analisa*
- *Dst.. Dsb...*



jakarta.csv

Open with Microsoft Excel



```
response_1588900572650.json
},
{
  "tanggal": "2020-03-15",
  "kode_prov": "32",
  "nama_prov": "Jawa Barat",
  "odp": 76,
  "odp_proses": 47,
  "odp_selesai": 29,
  "pdp": 1,
  "pdp_proses": 1,
  "pdp_selesai": 0,
  "positif": 7,
  "sembuh": 0,
  "meninggal": 2
},
{
  "tanggal": "2020-03-16",
  "kode_prov": "32",
  "nama_prov": "Jawa Barat",
  "odp": 76,
  "odp_proses": 0,
  "odp_selesai": 86,
  "pdp": 6,
  "pdp_proses": 5,
  "pdp_selesai": 1,
  "positif": 1,
  "sembuh": 2,
  "meninggal": 0
},
{
  "tanggal": "2020-03-17",
  "kode_prov": "32",
  "nama_prov": "Jawa Barat",
  "odp": 187,
  "odp_proses": 103,
  "odp_selesai": 84,
  "pdp": 16,
  "pdp_proses": 22,
  "pdp_selesai": 0,
  "positif": 0,
  "sembuh": 0,
  "meninggal": 0
},
{
  "tanggal": "2020-03-18",
  "kode_prov": "32",
  "nama_prov": "Jawa Barat",
  "odp": 53,
  "odp_proses": 13,
  "odp_selesai": 40,
  "pdp": 20,
  "pdp_proses": 18,
  "pdp_selesai": 2,
  "positif": 12,
  "sembuh": 0,
  "meninggal": 0
},
{
  "tanggal": "2020-03-19",
  "kode_prov": "32",
  "nama_prov": "Jawa Barat",
  "odp": 168,
  "odp_proses": 17,
  "odp_selesai": 151,
```

Format file
*.json
300-an baris lebih...

Parameters

Name	Description
tanggal_update <small>string (query)</small>	Tanggal update kasus. Nilai yang tersedia: YYYY-mm-dd. Nilai default: null . tanggal_update - Tanggal update kasus. Nilai
tanggal_konfirmasi <small>string (query)</small>	Tanggal konfirmasi kasus. Nilai yang tersedia: YYYY-mm-dd. Nilai default: null . tanggal_konfirmasi - Tanggal konfirmasi kasu

Execute

Responses

Code	Description
200	Success

GET /sebaran/jabar/faskes

Sebaran fasilitas kesehatan di Jawa Barat

Parameters

Name	Description
------	-------------

Jenis data yang didapatkan...

Dari Media Sosial...

Crawling data di **Twitter**
menggunakan aplikasi
Orange



orange
DATA MINING

Corpus Viewer

Info

Documents: 200
Preprocessed: False
◦ Tokens: n/a
◦ Types: n/a
POS tagged: False
N-grams range: 1-1
Matching: 200/200

Search features

- Author
- Content
- Date
- Language
- Location
- Number of Likes
- Number of Retweets
- In Reply To

Display features

- Author
- Content
- Date
- Language
- Location
- Number of Likes
- Number of Retweets

Show Tokens & Tags

Auto send is on

RegExp Filter:

139 RT @masplonget: @SKamil55 si Retno bilang ...
140 RT @kingsoka01: @SKamil55 Geram ...
141 RT @BobbySurya85: @SKamil55 Perlu langkah...
142 RT @L0YOR: #DPRDibawahKendaliRezim...
143 RT @ReoSR484: @SKamil55 perlu penanganan...
144 RT @hamdJr666: @SKamil55 satukan visi Misi...
145 RT @semut_ireng98: @SKamil55 ...
146 Apa kita memang masih punya daulat?...
147 RT @Sangk4k4l4: Lucuuu...emang... Tapi itula...
148 RT @alexander_murfi: Idealisme PKI Menyusup...
149 RT @boco_aluih_ang: https://t.co/Op3eY724pE...
150 RT @alexander_murfi: LPSK Pernah Temukan ...
151 RT @LoveniaHutama: @Altan_Sidiq ...
152 RT @Altan_Sidiq: @LoveniaHutama ...
153 RT @kingsoka01: @LoveniaHutama Tolak tapi ...
154 RT @LoveniaHutama: @ArsyVenita Penting kan...
155 RT @OO1972419: Klu DPR hanya utk legitimasi ...
156 RT @Q_Rei_Yoona: @LoveniaHutama See, OBL ...
157 RT @_JackCombat: @LoveniaHutama Hou ...
158 RT @FiyaHumayra: @LoveniaHutama Nyimak...

Author: @adalahu
Content: RT @L0YOR: #DPRDibawahKendaliRezim
Date: 2020-05-08 14:53:47
Language: in
Location: ?
Number of Likes: 0
Number of Retweets: 62
In Reply To: ?
Author Name: Pras_manan
Author Description: GDP growth using massive debt is like build big mansion on the quick sand #DeathOfInterest Everything goes up must come down and every bubble must pop up!
Author Statuses Count: 23010
Author Favourites Count: 12841
Author Friends Count: 540
Author Followers Count: 437

Wabah Corona Masih Melanda RI, Bappenas: Proyek Ibu Kota Baru Jalan Terus <https://t.co/OzPjVlfK06>

英勇

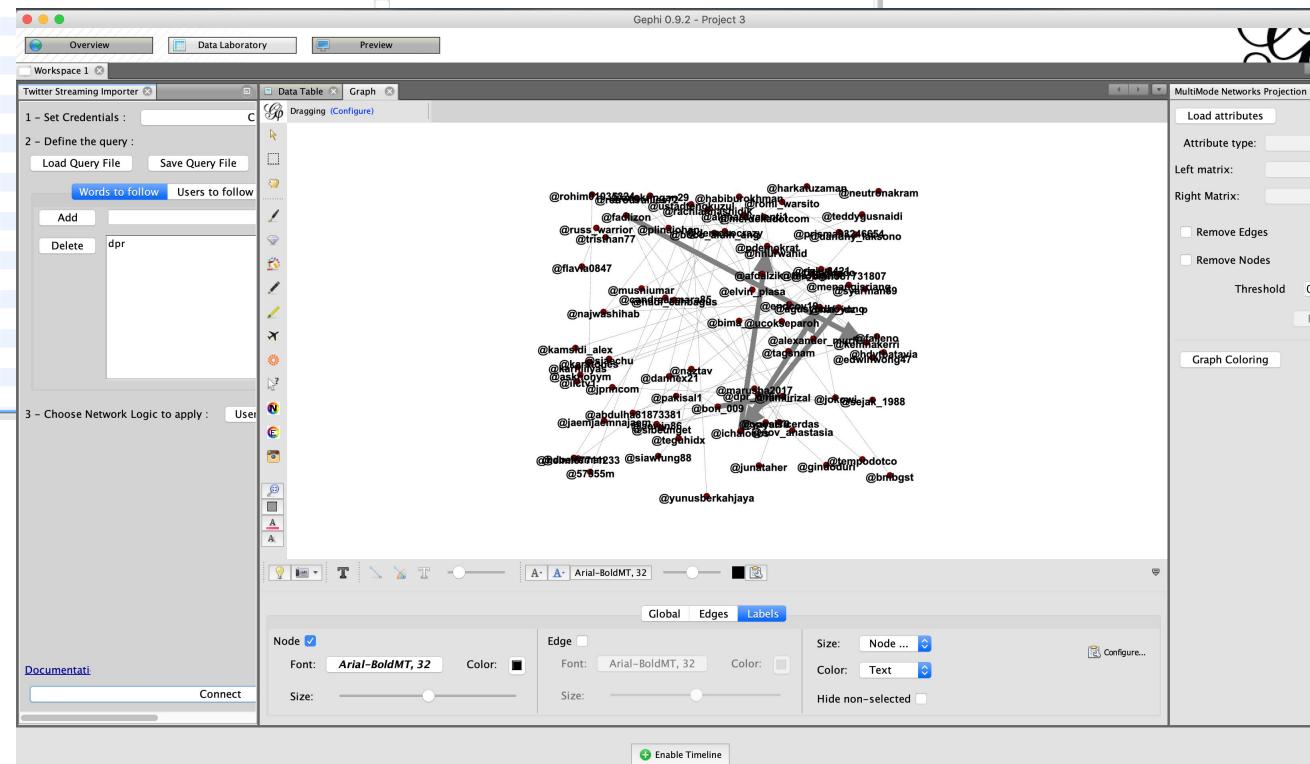


Jenis data yang didapatkan...

Dari Media Sosial...

Crawling data di **Twitter**
menggunakan aplikasi
Gephi

The figure shows a dual-interface view of Gephi 0.9.2. The top half is a Data Table with columns for Nodes, Edges, Configuration, Add node, Add edge, Search/Replace, Import Spreadsheet, Export table, More actions, Filter, and Id. The bottom half shows a network graph with nodes representing users and edges representing connections. A central workspace window contains a Twitter Streaming Importer configuration panel with tabs for Overview, Data Laboratory, Preview, Workspace 1, and a sub-panel for Dragging (Configure). The sub-panel includes sections for Set Credentials, Define the query, and a Words to follow/Users to follow tab with fields for Load Query File, Save Query File, Add, and Delete.



Tools Pengolah Data

date	odp_process	odp_end	odp_total	pdp_process	pdp_end	pdp_total	j
3/1/2020	14	115	129	9	30	39	
3/2/2020	14	115	129	9	30	39	
3/3/2020	94	121	215	24	30	54	
3/4/2020	120	121	241	26	30	56	
3/5/2020	127	140	267	49	34	83	
3/6/2020	83	221	304	44	44	88	
3/7/2020	94	262	356	44	54	98	
3/8/2020	64	298	362	57	70	127	
3/9/2020	68	310	378	87	79	166	
3/10/2020	70	331	401	97	100	197	
3/11/2020	98	347	445	103	116	219	
3/12/2020	112	374	486	120	118	238	
3/13/2020	144	442	586	126	135	261	
3/14/2020	174	486	660	152	137	289	

Olah dengan :

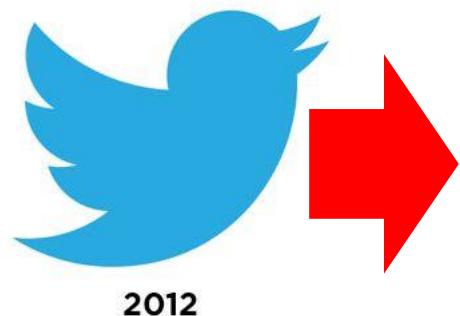
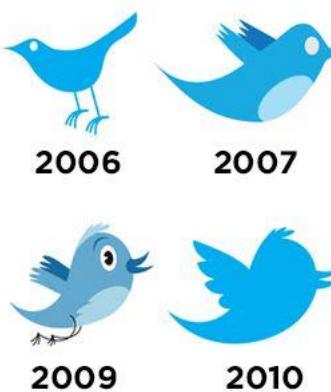
- Microsoft Excel
- Python
- Smartpls
- Lisrel
- Dsb dsb..

Olah dengan :

- Tableau
- R Language
- Python
- Dsb.. Dsb..

Olah dengan :

- Orange
- Gephi
- Tableau
- R Language
- Python
- Dsb..dsb...

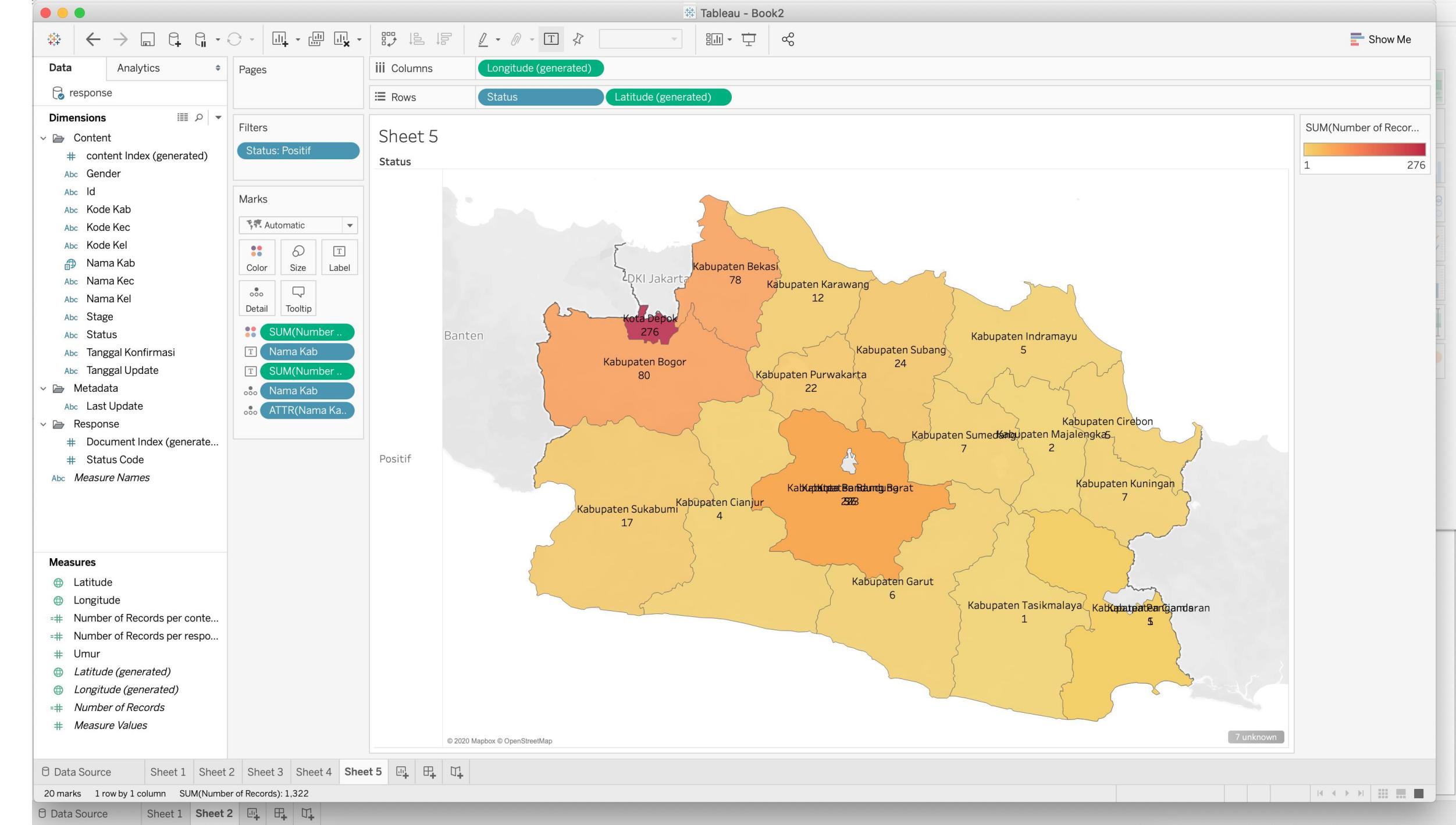


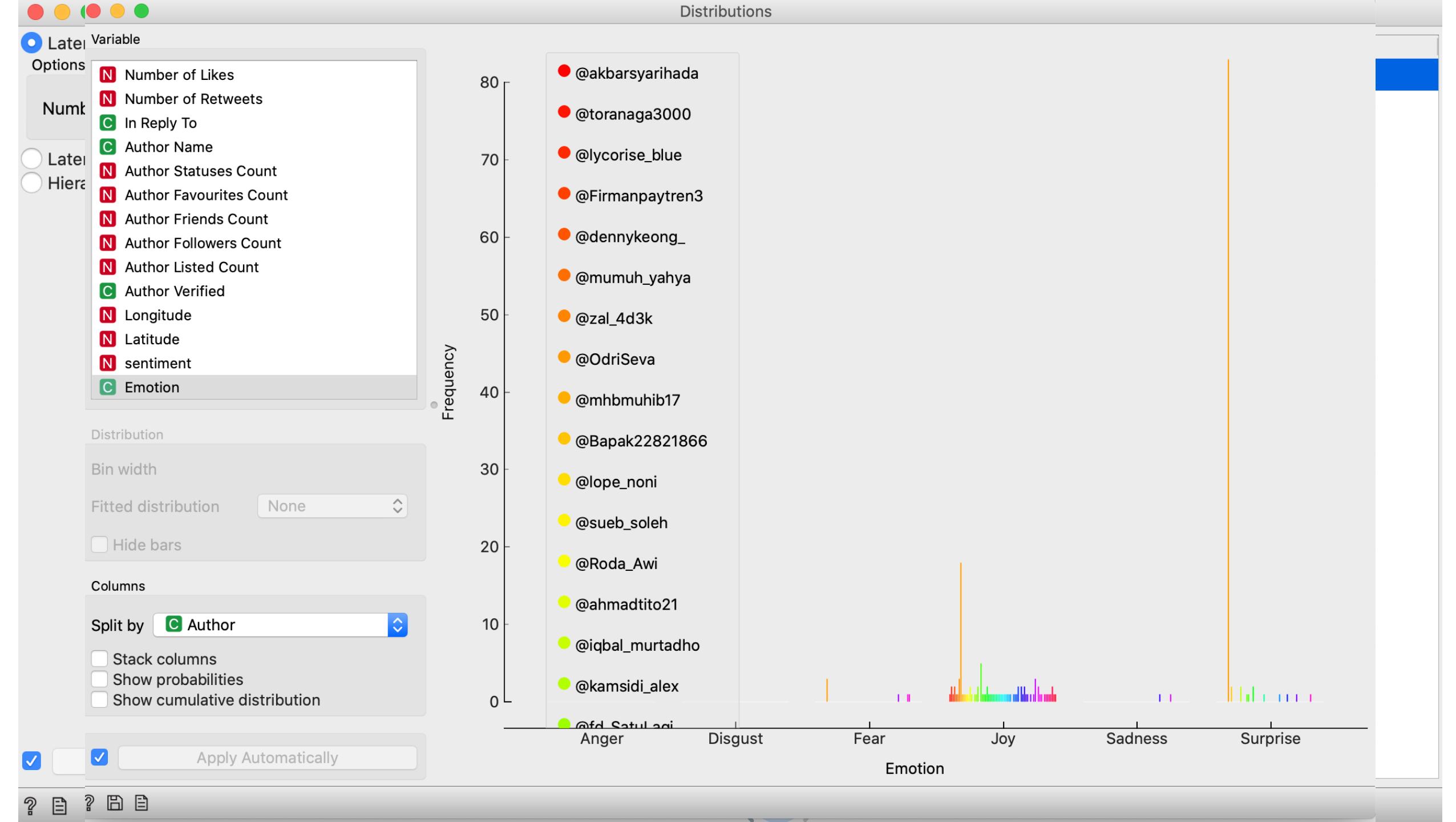
```
response_16889065/2650.json
},
{
  "tanggal": "2020-03-15",
  "kode_prov": "32",
  "nama_prov": "Jawa Barat",
  "odp": 76,
  "odp_proses": 47,
  "odp_selesai": 29,
  "pdp": 1,
  "pdp_proses": 1,
  "pdp_selesai": 0,
  "positif": 7,
  "sembuh": 0,
  "meninggal": 2
},
{
  "tanggal": "2020-03-16",
  "kode_prov": "32",
  "nama_prov": "Jawa Barat",
  "odp": 106,
  "odp_proses": 0,
  "odp_selesai": 86,
  "pdp": 6,
  "pdp_proses": 5,
  "pdp_selesai": 1,
  "positif": 1,
  "sembuh": 2,
  "meninggal": 0
},
{
  "tanggal": "2020-03-17",
  "kode_prov": "32",
  "nama_prov": "Jawa Barat",
  "odp": 107,
  "odp_proses": 103,
  "odp_selesai": 84,
  "pdp": 1,
  "pdp_proses": 22,
  "pdp_selesai": 0,
  "positif": 0,
  "sembuh": 0,
  "meninggal": 0
},
{
  "tanggal": "2020-03-18",
  "kode_prov": "32",
  "nama_prov": "Jawa Barat",
  "odp": 53,
  "odp_proses": 13,
  "odp_selesai": 40,
  "pdp": 20,
  "pdp_proses": 18,
  "pdp_selesai": 2,
  "positif": 7,
  "sembuh": 0,
  "meninggal": 0
},
{
  "tanggal": "2020-03-19",
  "kode_prov": "32",
  "nama_prov": "Jawa Barat",
  "odp": 168,
  "odp_proses": 17,
  "odp_selesai": 151,
```

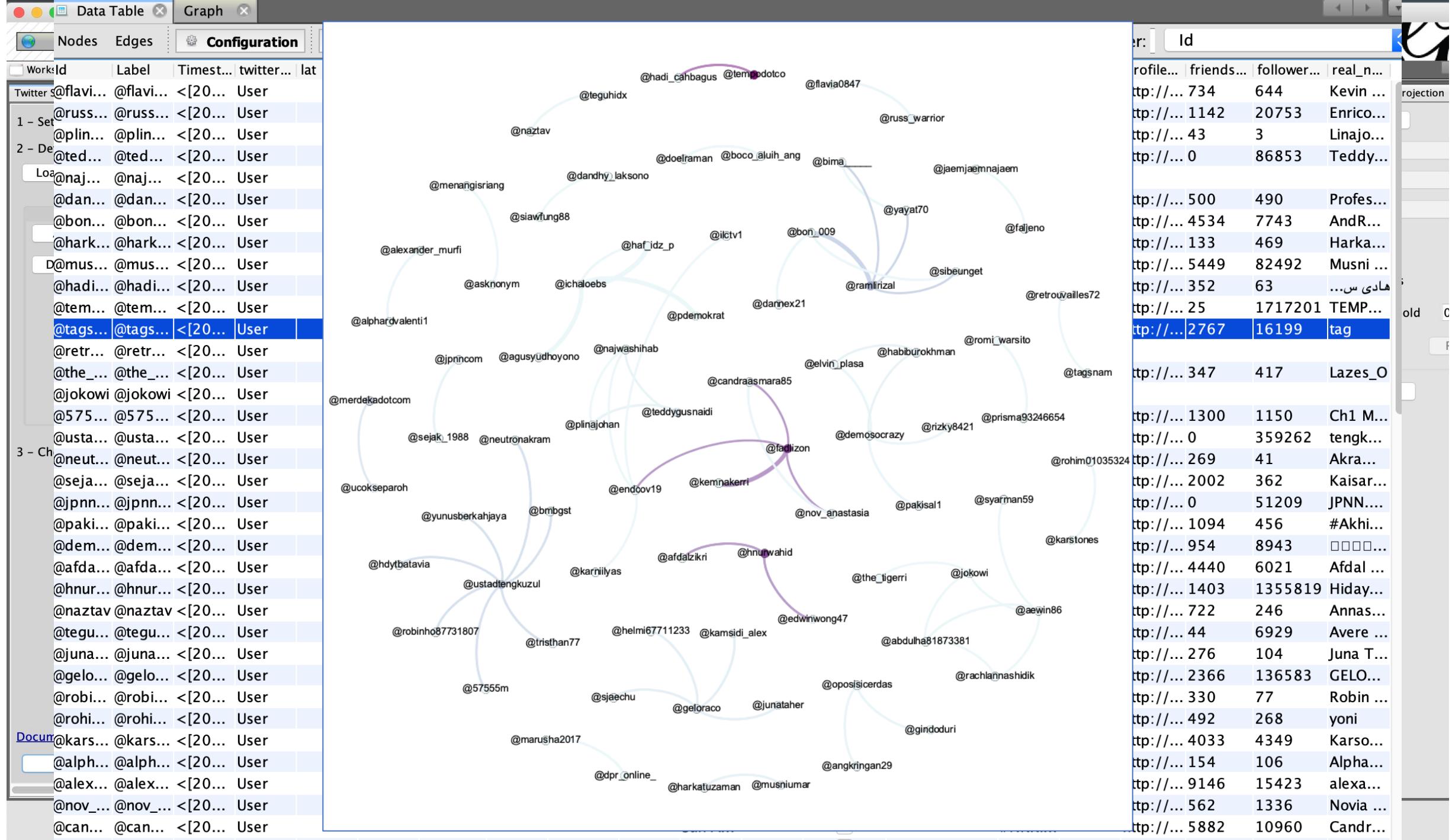


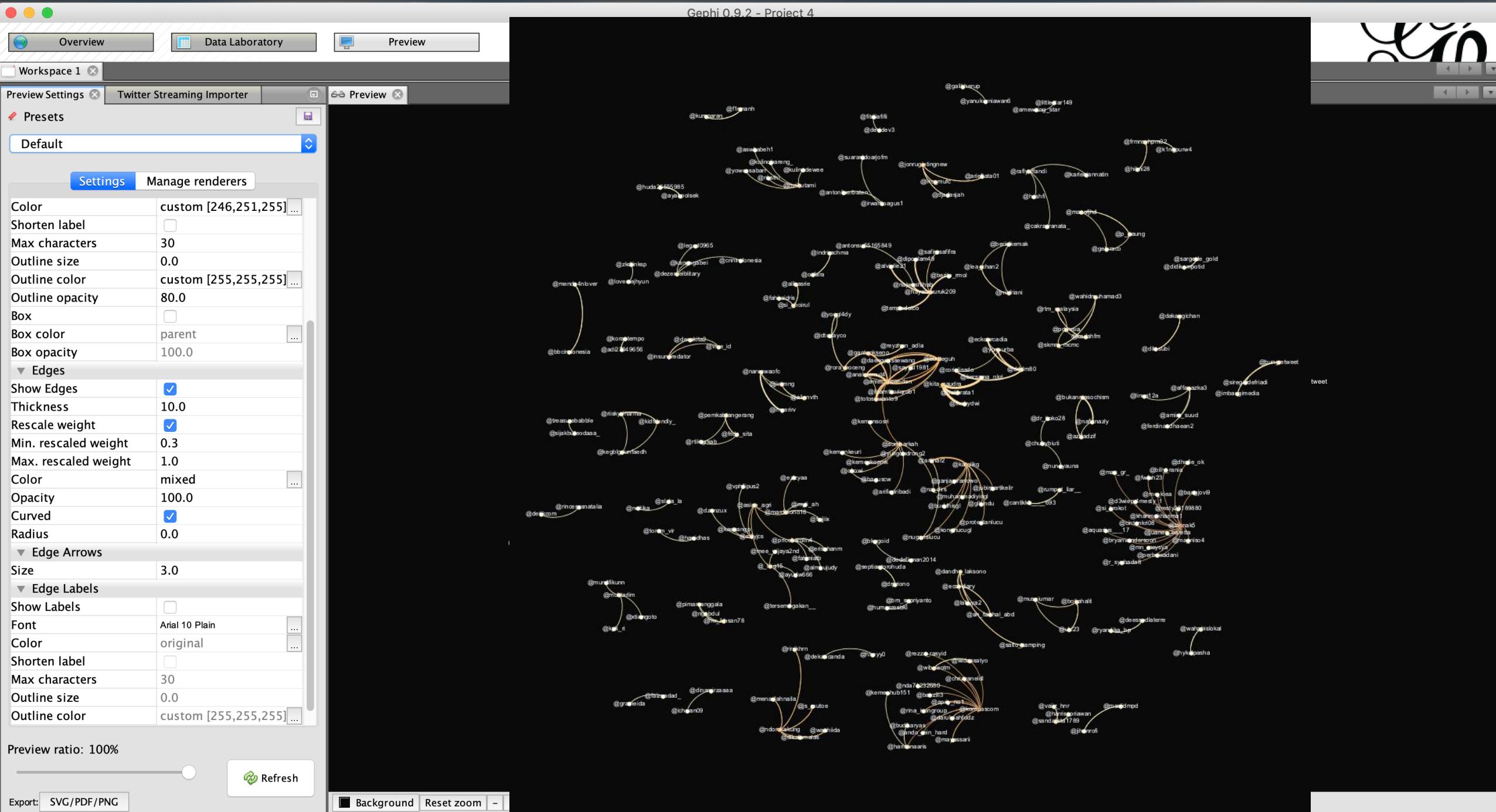
Analisa Data...

1. Tableau
2. Orange
3. Gephi



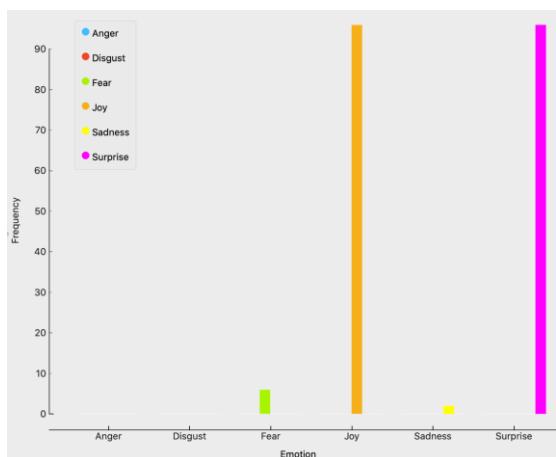






Publish Data

Topic Topic keywords
1 dprdibawahkendalirezim, the, no, my, articles, healthy, seva, human, find, right
2 tka, tolak, cina, alexander_murfi, dewan, china, air, indonesia, masuk, tanah
3 dprdibawahkendalirezim, g4knyer4hbr0___, rakyat, mati, dpr, seva, new, human, odri, therighttobehealthy
4 dpr, rakyat, perppu, ri, corona, dprdibawahkendalirezim, dakwah, , , pki
5 up, must, rakyat, wni, diperlakukan, dpr, quick, build, deathofinterest, sand
6 turunkan, lantunan, mbuktikan, penelitian, psikologi, quran, meningkatkn, ugm, al, hasil
7 dakwah, rakyat, perppu, corona, , , , mati, hidup, must, up
8 perppu, corona, adil, rakyat, tuk, keadilan, dholim, sweety_mu_20, pks, uu
9 anggota, mayoritas, dpr, rakyat, fiyahumayra, masplonget, 2019 ,2024 , , ,
10 mayoritas, kingsoka01, ri, dakwah, wni, diperlakukan, dunia, masplonget, fiyahumayra, lucius

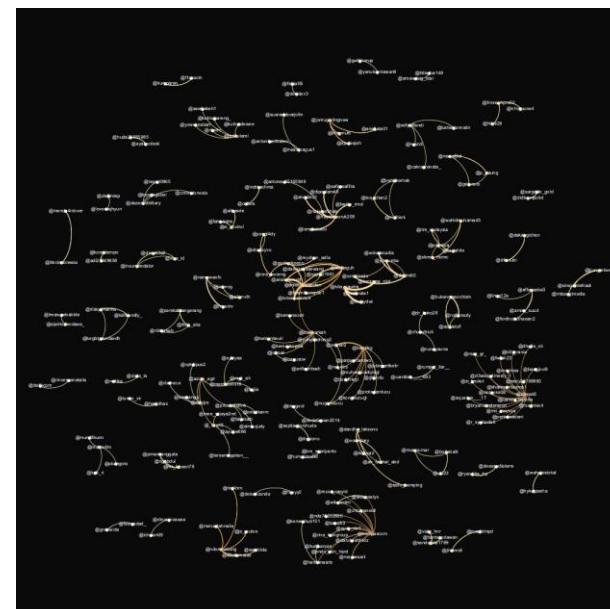
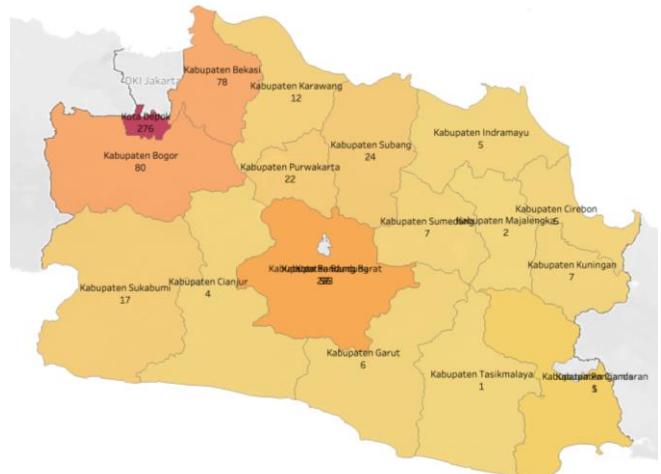


Isu yang paling banyak dibicarakan adalah #DPRdibawahKendaliRezim, Save, Human, Right

Rerata netizen surprise dan cenderung terkonsentrasi di isu ini.

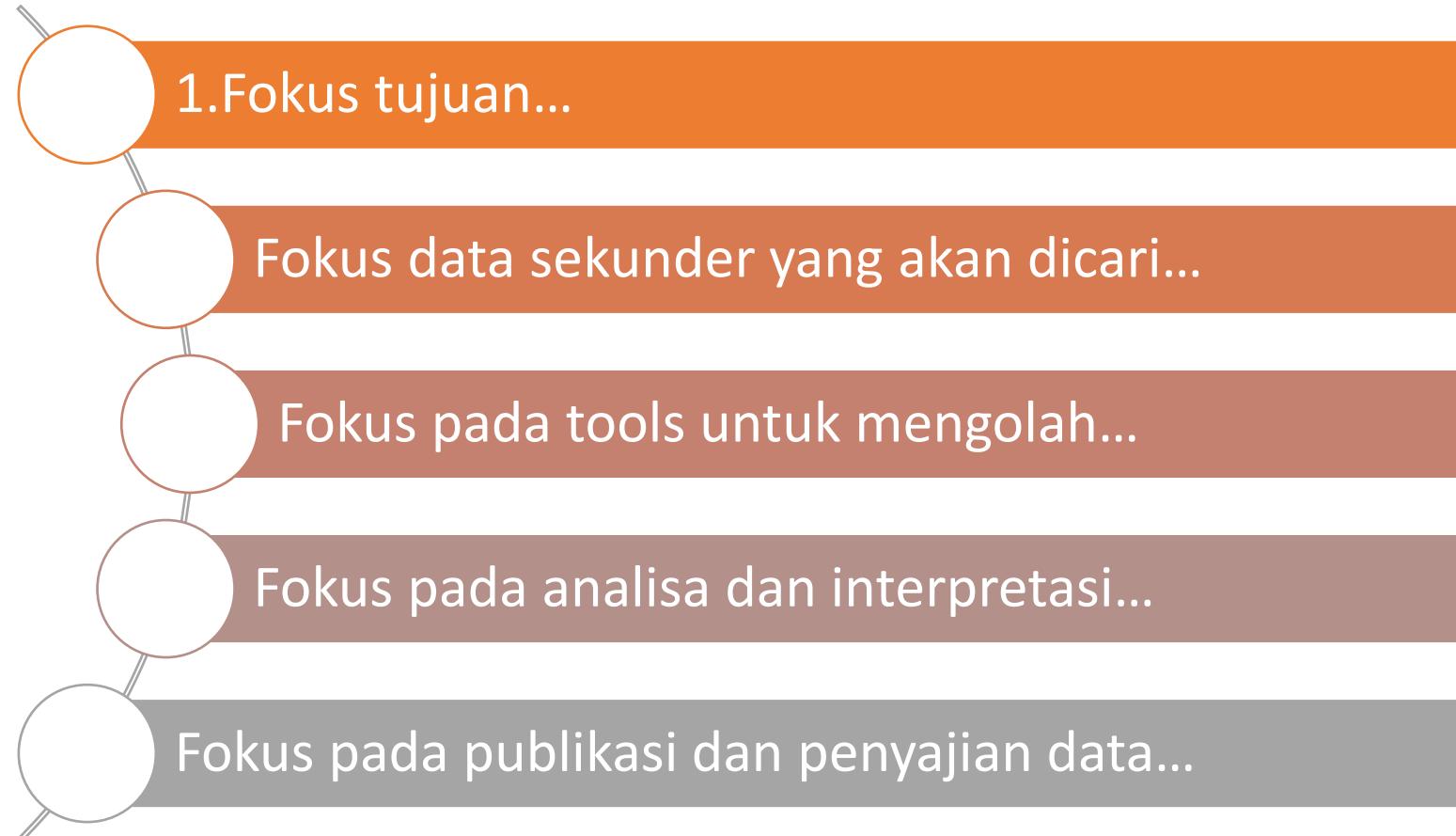
Sementara beberapa tokoh nasional ikut meramaikan isu ini diantaranya adalah....

Analisa untuk daerah merah dan orange... Yellow cenderung masih aman... dst dst...





Tips and Trick Mengolah Data era COVID-19

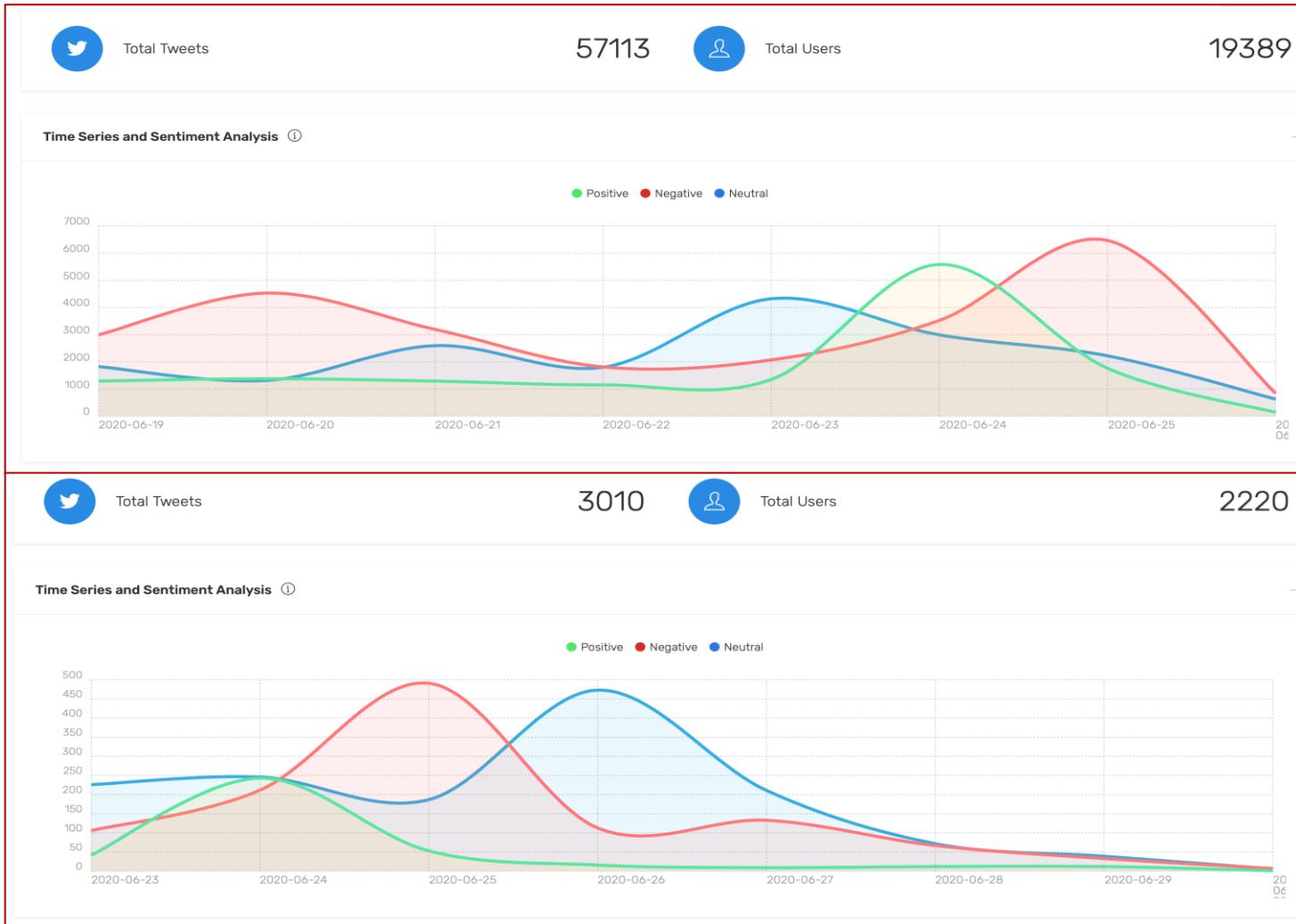




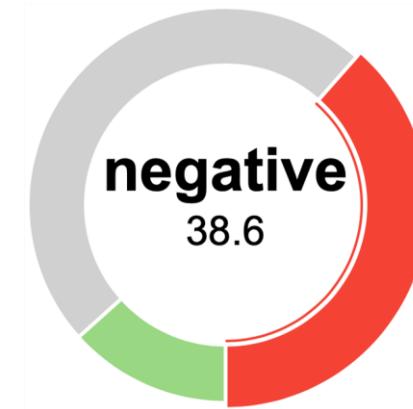
Contoh Kasus...



RUU HIP (twitter)

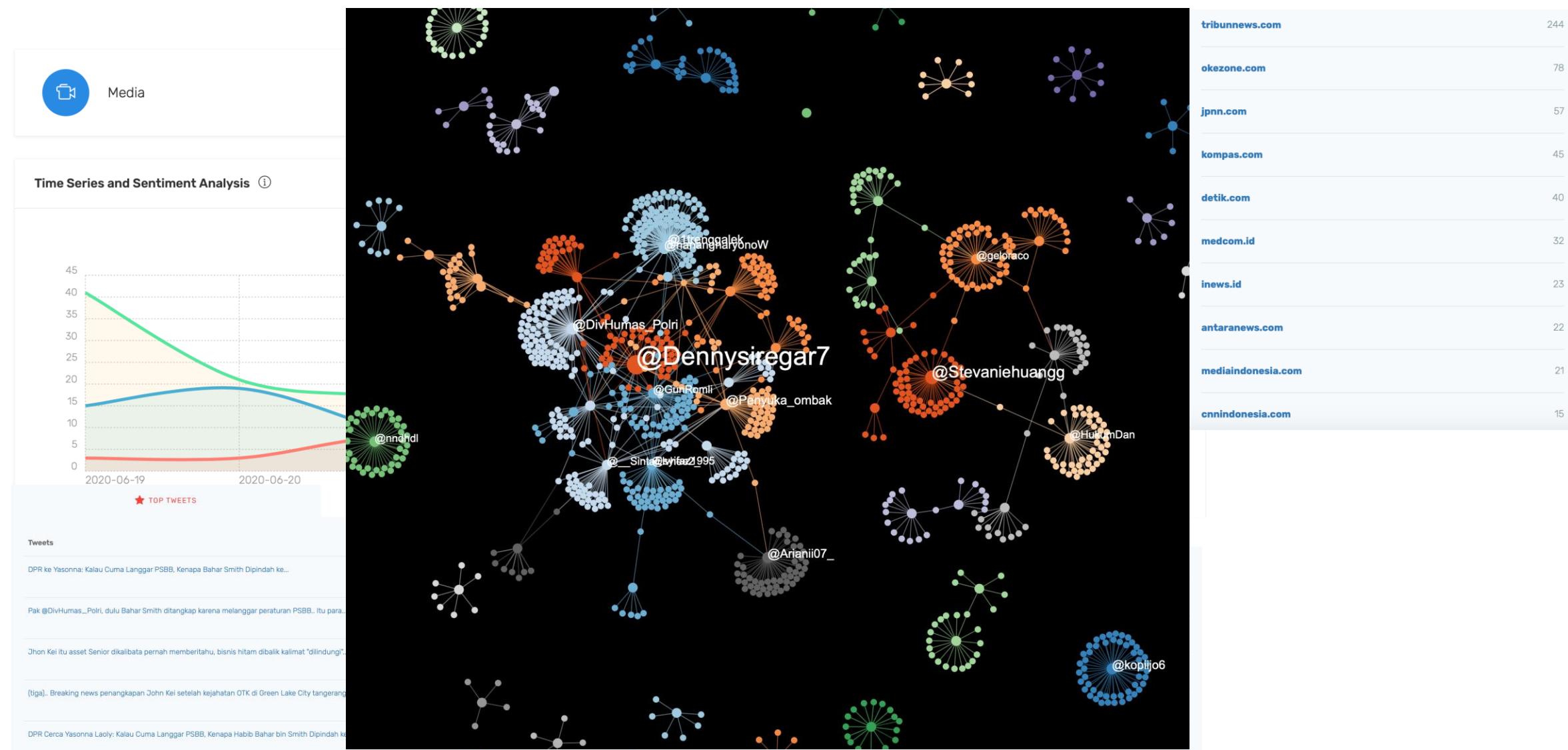


sentiment



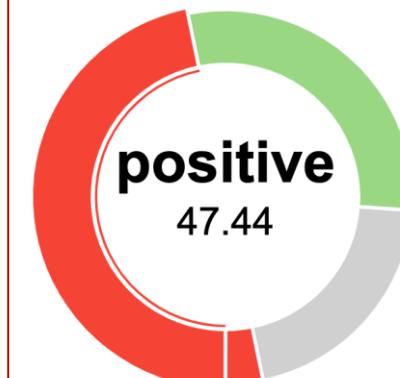
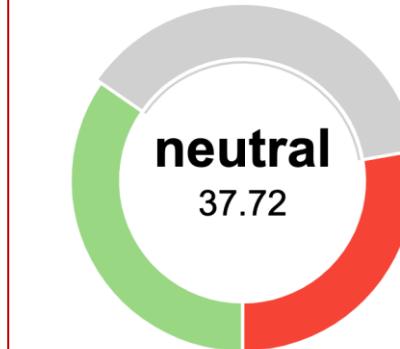
dukungruuhip
ruuhip
rakyatbengkulu
pancasilaabadi
jokowi BPIP Jokowi
PDIP **RUUHIP**
RUUHIP Usulan DPR
Kami Tidak Takut Banteng
Masyarakat **pancasilasakti likeforlikes**

John Kei vs Habib Bahar



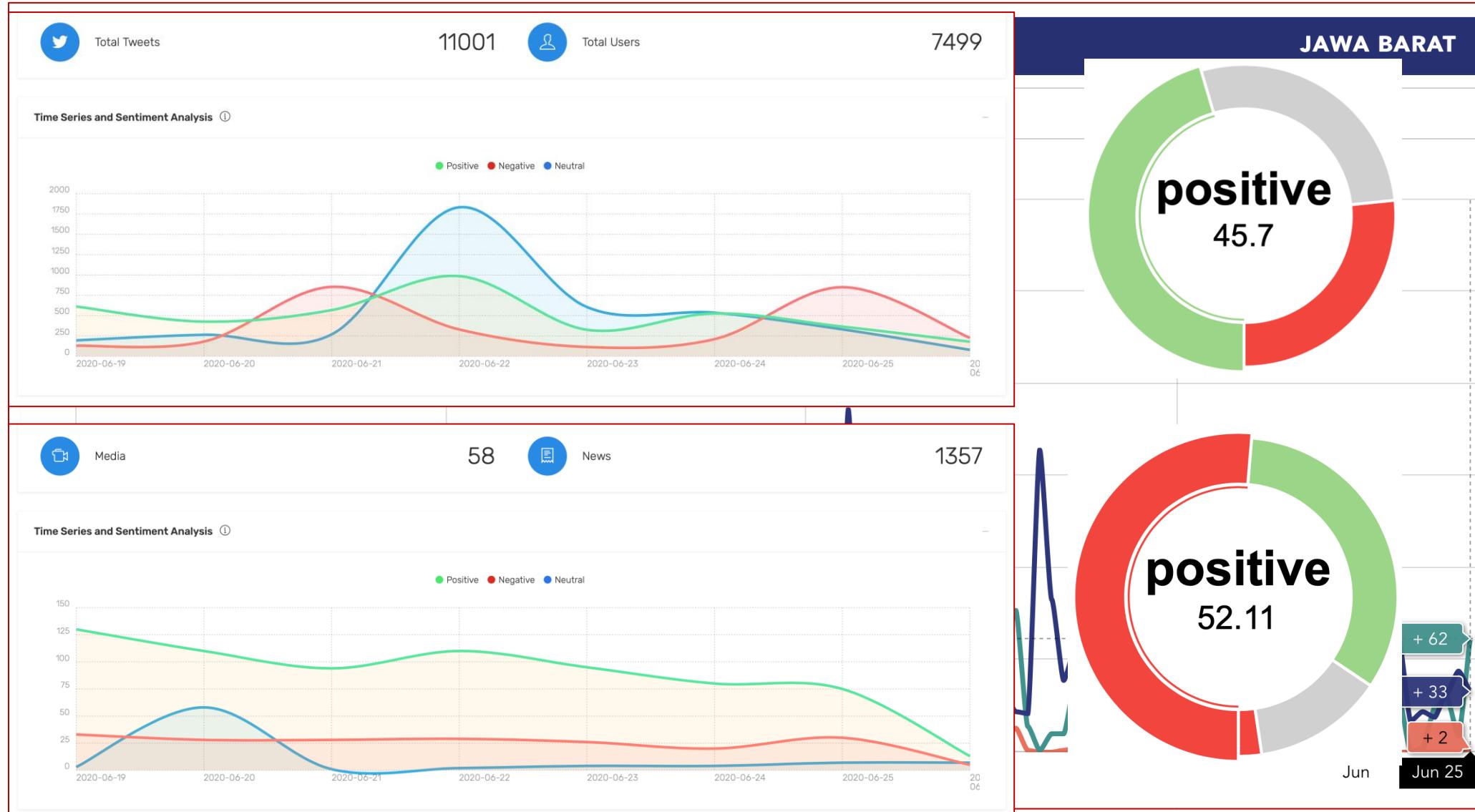


Tri Rismaharini



BeritaTerkini
TriRismaharini
BuRisma Risma Surabaya
Korona hotel88 GanyangPKI
VirusCorona Redaksi dirumahaja COVID

PSBB Jabar



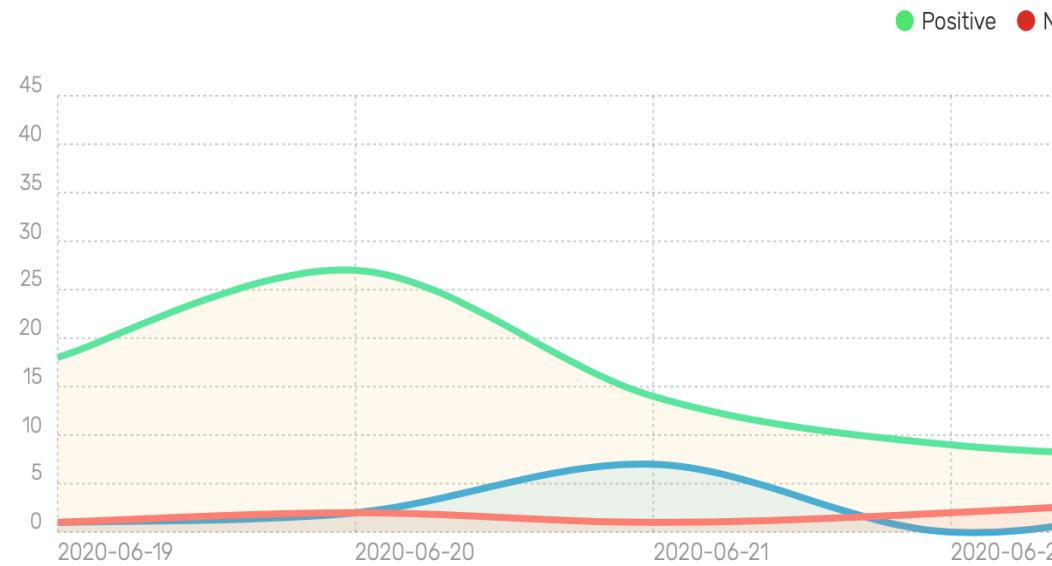
Khofifah – Jawa Timur



Media

32

Time Series and Sentiment Analysis (i)



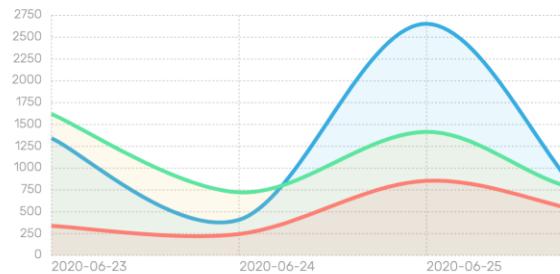


PPDB DKI



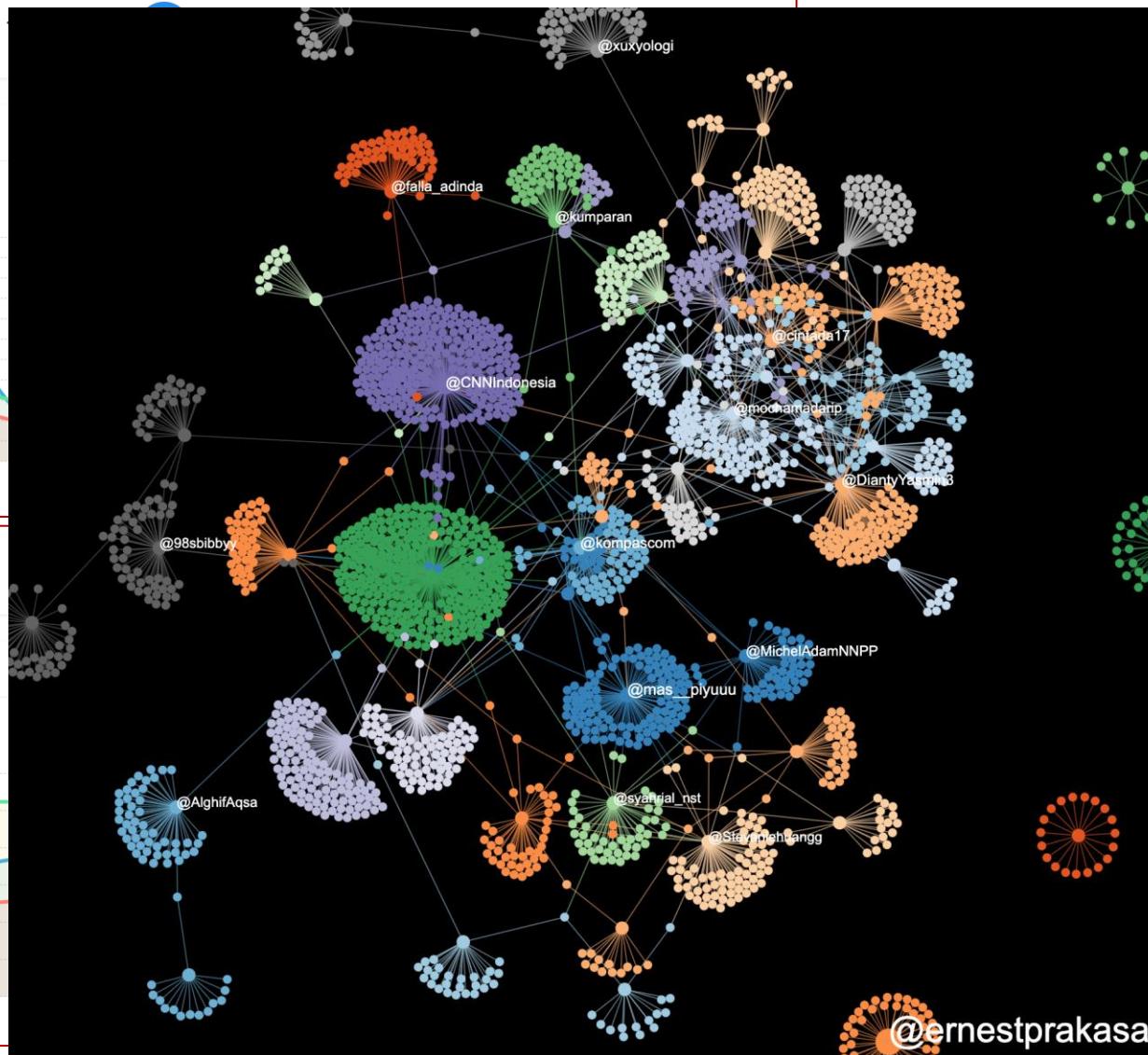
Total Tweets

Time Series and Sentiment Analysis ①



Media

Time Series and Sentiment Analysis ①



[tribunnews.com](#)

[mediaindonesia.com](#)

[kompas.com](#)

[okezone.com](#)

[medcom.id](#)

[kontan.co.id](#)

[solopos.com](#)

[detik.com](#)

[jpnn.com](#)

@ernestprakasa

SATU DATA UNTUK LAYANAN PUBLIK

OFFERING :

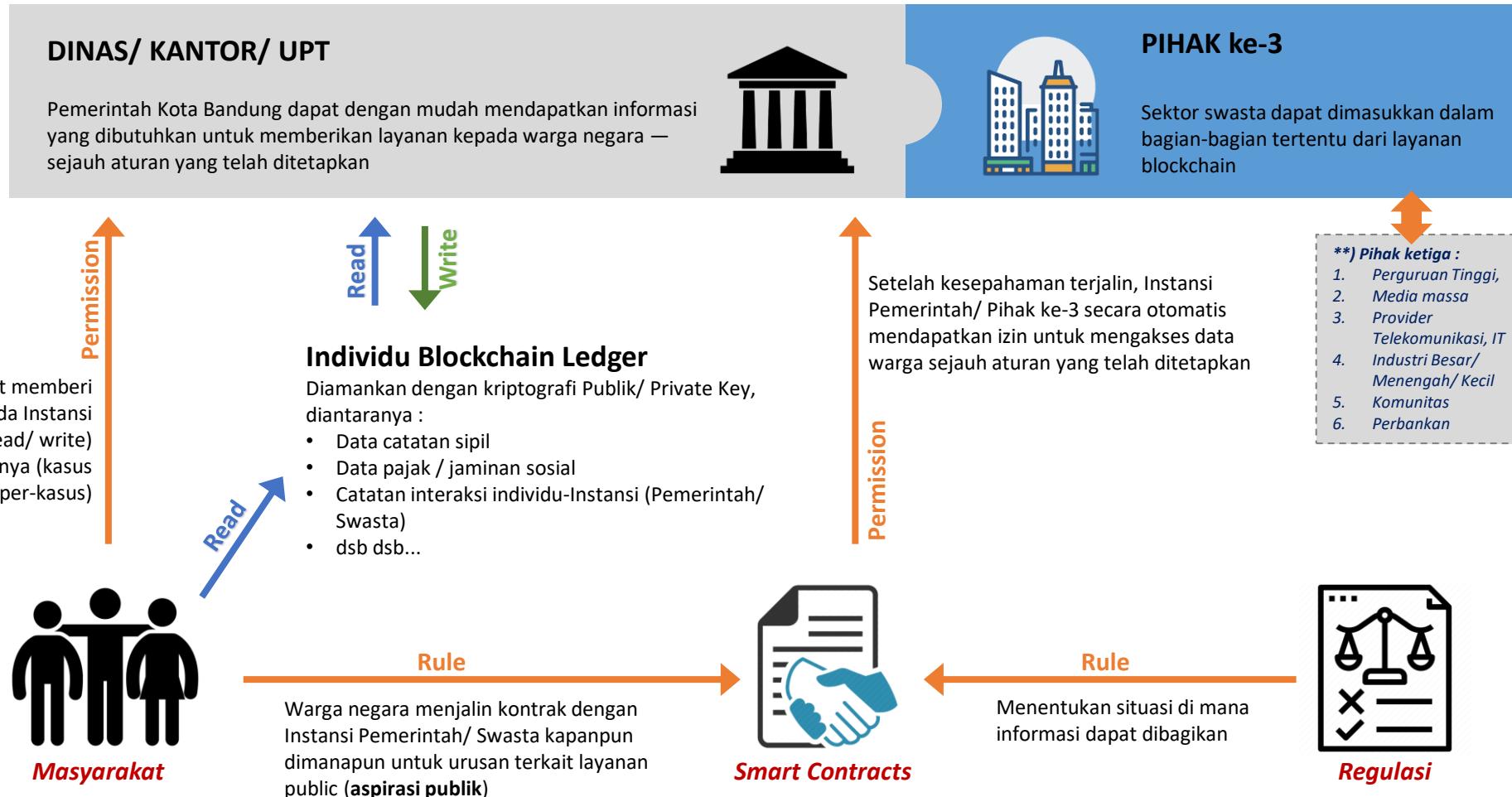
- 1 Voting
- 2 Combating Corruption
- 3 Improving public services
- 4 Transparang Budgeting
- 5 ease of supervision
- 6 Paperless based systems

ADVANTAGES :

- ✓ No major changes to your existing infrastructure
- ✓ An information relay layer where are stakeholders publish bonafide information
- ✓ Increase interoperability among stakeholders
- ✓ Decentralized, permissioned data ledger
- ✓ Certified and immutable history of all citizens



BIG PICTURE





Terima Kasih...