UNIVERSIDAD AUSTRAL | INGENIERÍA

Tweets COVID-19: Análisis de Sentimiento

Del Villar, Javier – Otrino, Facundo - Pistoya, Haydee - Rojas, Mariano - Sorza, Andrés - Vaillard, Leandro

Maestría en Explotación de Datos y Gestión del Conocimiento

Índice

UNIVERSIDAD AUSTRAL

Planteo del Problema

- Realizar un análisis de sentimiento sobre la temática del COVID-19 en Argentina
- Determinar si hubieron cambios de ánimos en la población respecto del COVID-19 durante el período Enero 2020 hasta Abril 2021.
- Detectar las temáticas de mayor ocurrencia durante el período mencionado anteriormente.

Enfoque Metodológico

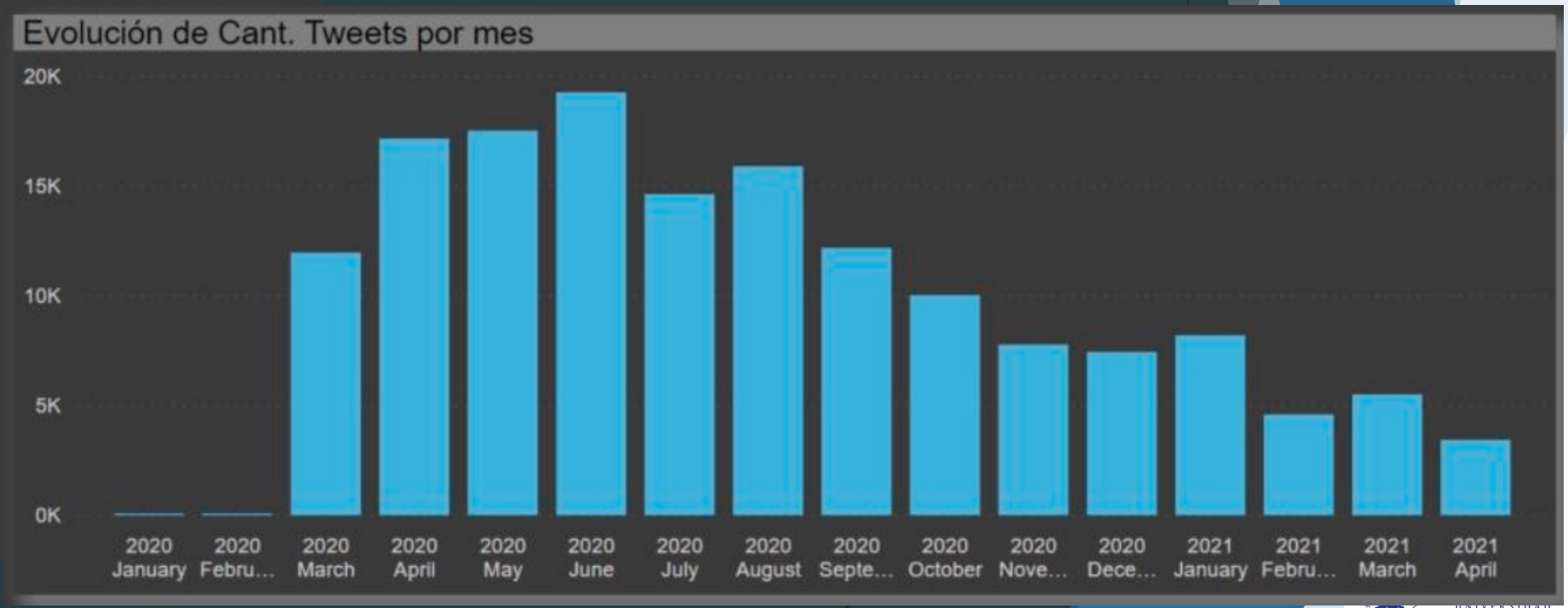
- Obtención de los datos:
  - Utilizar un repositorio de GitHub que incluye los identificadores de *tweets* (Tweet ID) relacionados a la temática de COVID-19 (base total a nivel global: 1.000 millones de tweets) (https://github.com/thepanacealab/covid19_twitter)
    - Filtrar por país: Argentina
    - Filtrar por lenguaje: Español
  - Utilizar la API de Twitter para descargar la base de *tweets*
    - Más de 150.000 *tweets* fueron descargados
- Pre-Procesamiento
  - Determinar en base a la metadata los campos a emplear.
  - Realizar la clasificación manual de aprox. 4.000 *tweets*
  - Aplicar técnicas de *feature engineering*.

Enfoque Metodológico

- Cantidad de Tweets

Year	Month	Count of id
2020	January	9
2020	February	44
2020	March	11,943
2020	April	17,141
2020	May	17,497
2020	June	19,238
2020	July	14,606
2020	August	15,874
2020	September	12,162
2020	October	10,011
2020	November	7,761
2020	December	7,423
2021	January	8,187
2021	February	4,582
2021	March	5,496
2021	April	3,425
Total		**155,399**

Year	Count of id
⊞ 2020	133,709
⊞ 2021	21,690
Total	**155,399**



Enfoque Metodológico

- Selección del modelo
  - Probar distintos modelos y seleccionar el que arrojase los resultados más satisfactorios
    - Aplicar un modelo donde se cambiaban los emojis por cadenas de texto
      - E.j. ☺ se convirtió en: “:cara_sonriente:”
    - Eliminar caracteres de puntuación
      - “:cara_sonriente:” se convirtió en “cara sonriente”
- Aplicación del Modelo
  - Aplicar modelo tipo BERT para clasificar los *tweets* en positivo, negativo, neutral
  - Comparar la clasificación del modelo con la realizada manualmente

Validación del Modelo

- Clasificación Manual de Tweets

	Negativo	Neutral	Positivo	Total
Negativo	1536	86	252	**1874**
Neutral	423	41	146	**610**
Positivo	418	42	376	**836**
Total	**2377**	**169**	**774**	**3320**

Resultado	Conteo	Porcentaje
Igual	1953	59%
Distinto	1367	41%
Total	3320	100%

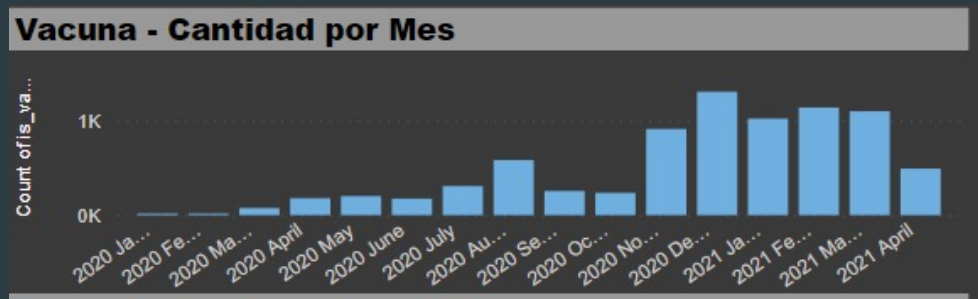
UNIVERSIDAD AUSTRAL

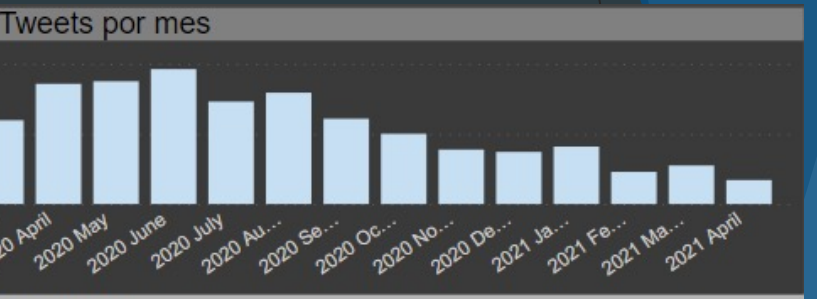
Resultados

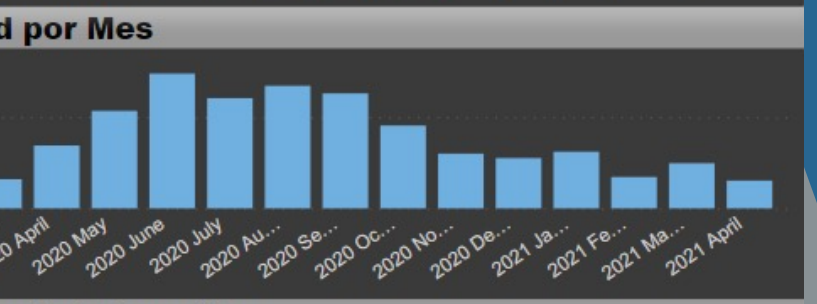
Year	Negativo (1&2)	Neutral (3)	Positivo (4 & 5)	Total
⊟ 2020	71%	5%	24%	100%
⊞ Qtr 1	68%	6%	27%	100%
⊞ Qtr 2	70%	5%	25%	100%
⊞ Qtr 3	72%	5%	23%	100%
⊞ Qtr 4	74%	4%	22%	100%
⊟ 2021	73%	4%	23%	100%
⊞ Qtr 1	73%	4%	23%	100%
⊞ Qtr 2	72%	6%	23%	100%
Total	71%	5%	24%	100%

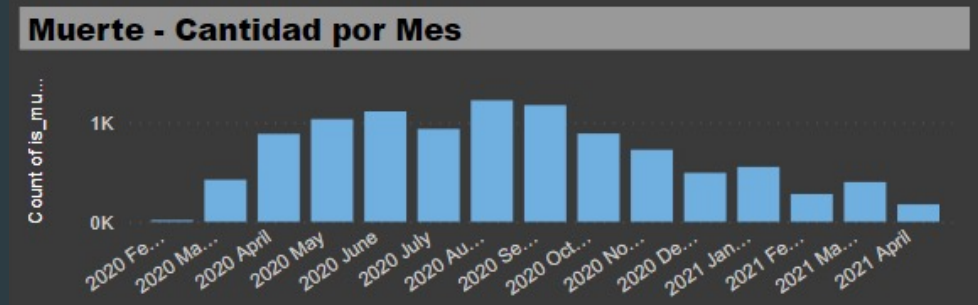
UNIVERSIDAD AUSTRAL

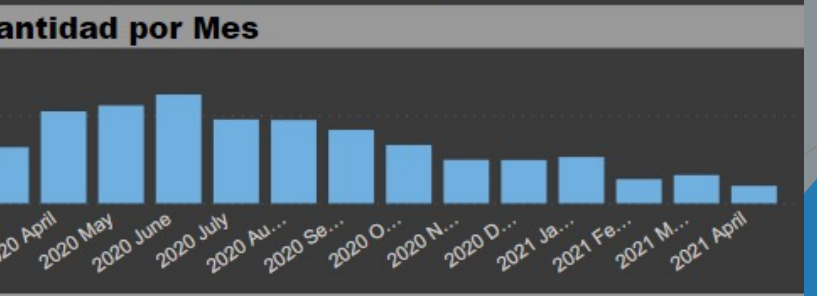
Análisis de Resultados

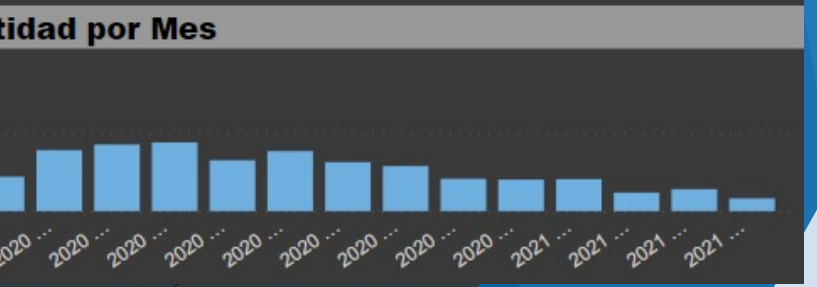








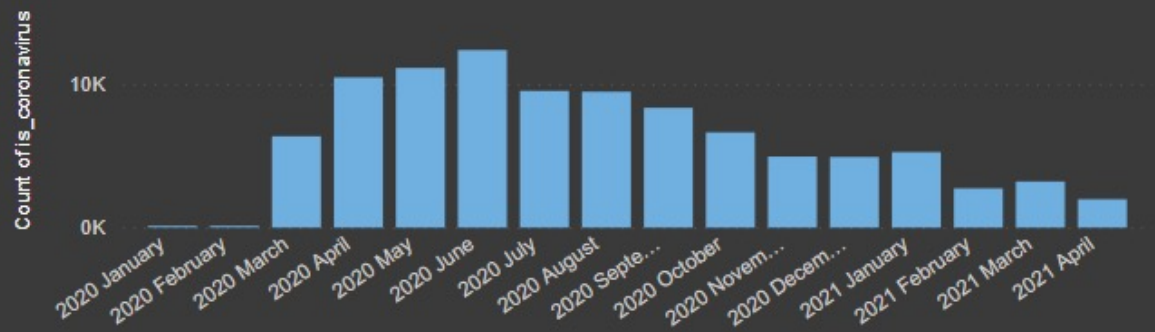




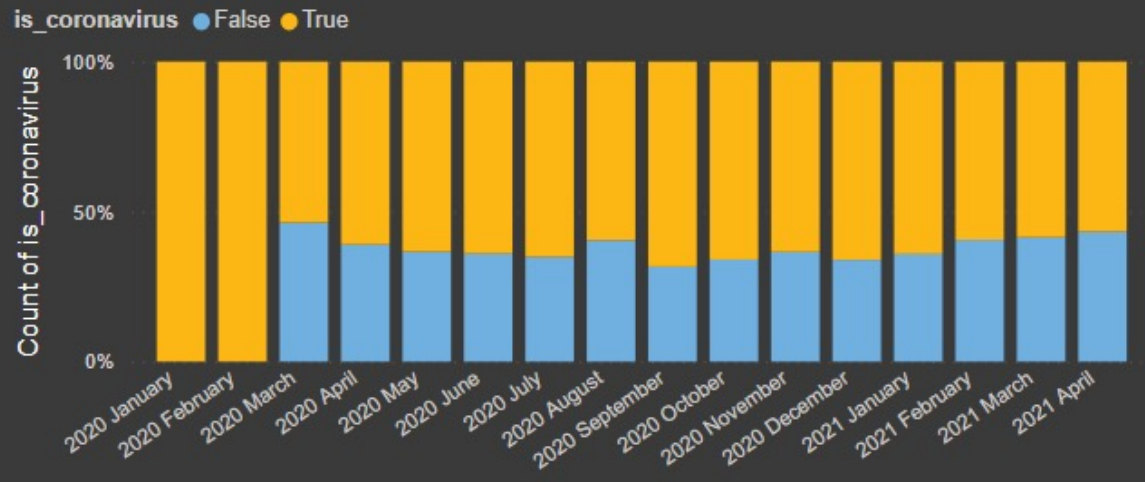
Análisis de Resultados

Keyword: Coronavirus

Cantidad por Mes



Representatividad por mes



Is coronavirus

Year	False	True	Total
2020	37%	63%	100%
January		100%	100%
February		100%	100%
March	46%	54%	100%
April	39%	61%	100%
May	37%	63%	100%
June	36%	64%	100%
July	35%	65%	100%
August	40%	60%	100%
September	32%	68%	100%
October	34%	66%	100%
November	37%	63%	100%
December	34%	66%	100%
2021	39%	61%	100%
January	36%	64%	100%
February	41%	59%	100%
March	41%	59%	100%
April	43%	57%	100%
Total	37%	63%	100%

Is coronavirus

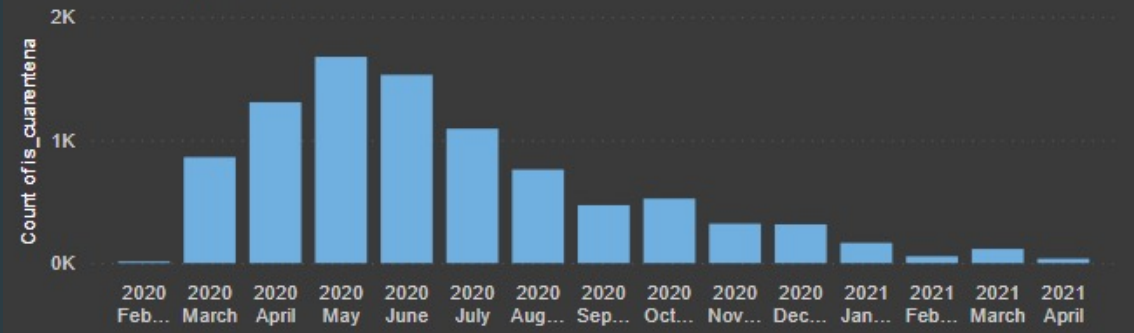
Year	False	True	Total
2020	49,668	84,015	133,683
January		9	9
February		44	44
March	5,482	6,349	11,831
April	6,682	10,437	17,119
May	6,418	11,097	17,515
June	6,942	12,340	19,282
July	5,093	9,505	14,598
August	6,416	9,446	15,862
September	3,857	8,326	12,183
October	3,413	6,618	10,031
November	2,844	4,935	7,779
December	2,521	4,909	7,430
2021	8,571	13,145	21,716
January	2,937	5,245	8,182
February	1,865	2,728	4,593
March	2,265	3,201	5,466
April	1,504	1,971	3,475
Total	58,239	97,160	155,399

UNIVERSIDAD AUSTRAL

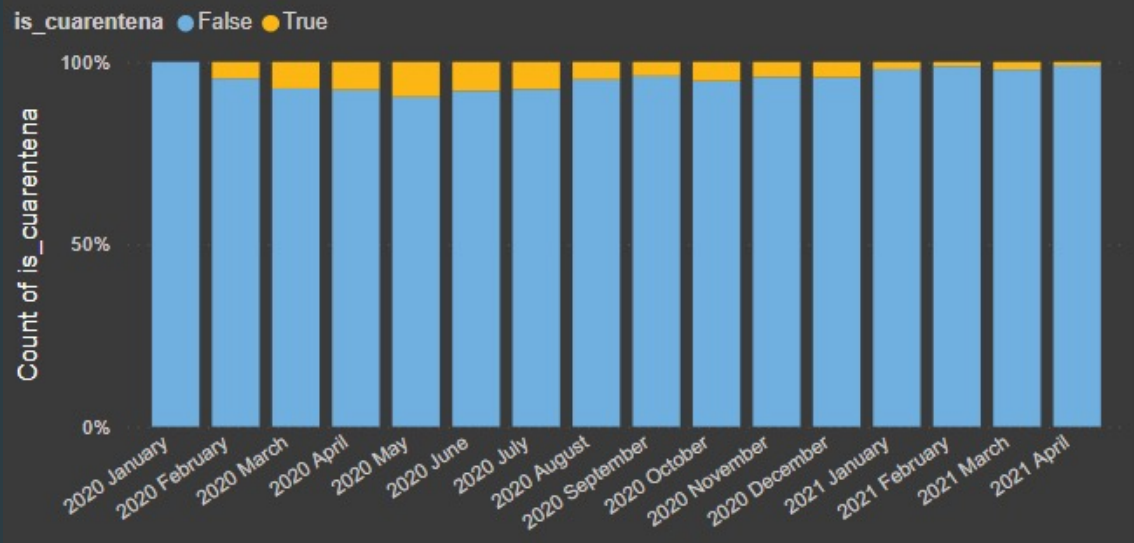
Análisis de Resultados

Keyword: Cuarentena

Cantidad por Mes



Representatividad por mes



Is cuarenten

Year	False	True	Total
⊟ 2020	93%	7%	100%
January	100%		100%
February	95%	5%	100%
March	93%	7%	100%
April	92%	8%	100%
May	90%	10%	100%
June	92%	8%	100%
July	93%	7%	100%
August	95%	5%	100%
September	96%	4%	100%
October	95%	5%	100%
November	96%	4%	100%
December	96%	4%	100%
⊟ 2021	98%	2%	100%
January	98%	2%	100%
February	99%	1%	100%
March	98%	2%	100%
April	99%	1%	100%
Total	94%	6%	100%

Is cuarenten

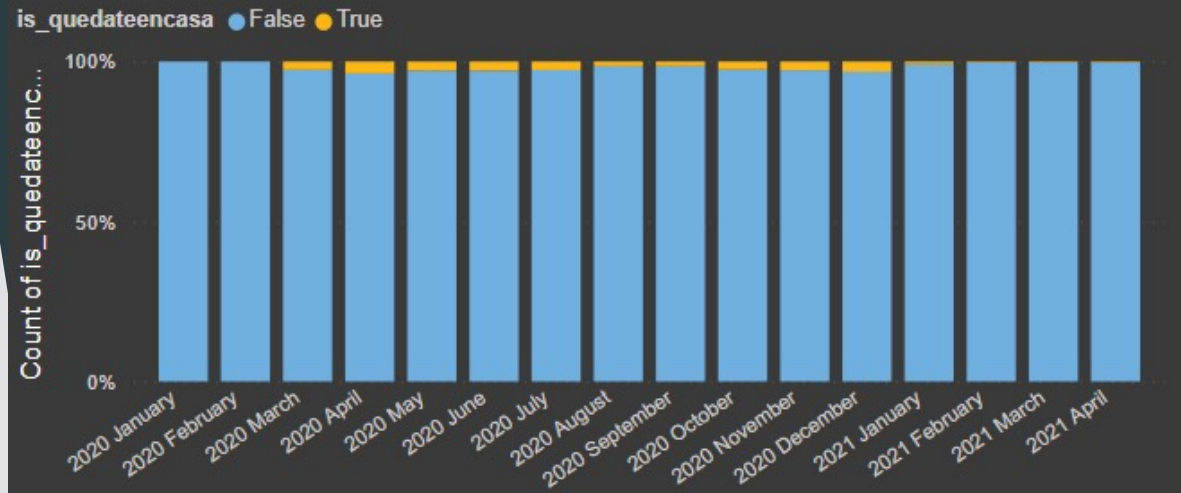
Year	False	True	Total
⊟ 2020	124,825	8,858	133,683
January	9		9
February	42	2	44
March	10,970	861	11,831
April	15,812	1,307	17,119
May	15,839	1,676	17,515
June	17,752	1,530	19,282
July	13,506	1,092	14,598
August	15,102	760	15,862
September	11,712	471	12,183
October	9,507	524	10,031
November	7,459	320	7,779
December	7,115	315	7,430
⊟ 2021	21,346	370	21,716
January	8,018	164	8,182
February	4,538	55	4,593
March	5,351	115	5,466
April	3,439	36	3,475
Total	146,171	9,228	155,399

UNIVERSIDAD AUSTRAL

Análisis de Resultados

Cantidad por Mes

Representatividad por mes



Keyword: Quedate en casa

Is quedate en casa

Year	False	True	Total
⊟ 2020	97%	3%	100%
January	100%		100%
February	100%		100%
March	97%	3%	100%
April	96%	4%	100%
May	97%	3%	100%
June	97%	3%	100%
July	97%	3%	100%
August	99%	1%	100%
September	99%	1%	100%
October	97%	3%	100%
November	97%	3%	100%
December	97%	3%	100%
⊟ 2021	99%	1%	100%
January	99%	1%	100%
February	100%	0%	100%
March	100%	0%	100%
April	100%	0%	100%
Total	98%	2%	100%

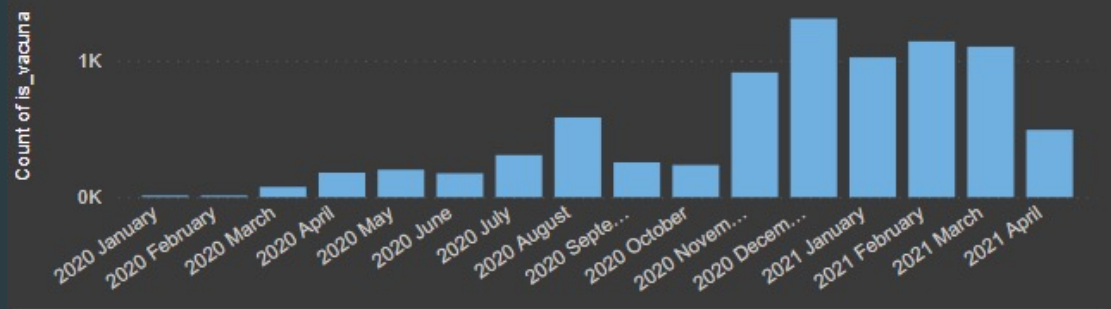
Is quedate en casa

Year	False	True	Total
⊟ 2020	130,186	3,497	133,683
January	9		9
February	44		44
March	11,528	303	11,831
April	16,480	639	17,119
May	17,010	505	17,515
June	18,728	554	19,282
July	14,211	387	14,598
August	15,630	232	15,862
September	12,008	175	12,183
October	9,780	251	10,031
November	7,566	213	7,779
December	7,192	238	7,430
⊟ 2021	21,604	112	21,716
January	8,114	68	8,182
February	4,578	15	4,593
March	5,448	18	5,466
April	3,464	11	3,475
Total	151,790	3,609	155,399

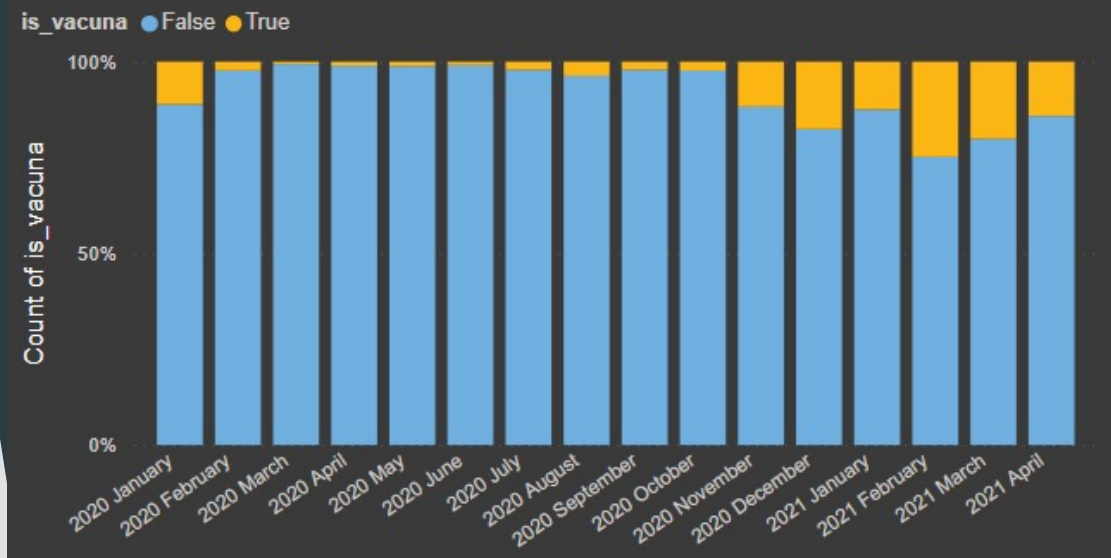
UNIVERSIDAD AUSTRAL

Análisis de Resultados

Cantidad por Mes



Representatividad por mes



Keyword: Vacuna

Is Vacuna

Year	False	True	Total
⊟ 2020	97%	3%	100%
January	89%	11%	100%
February	98%	2%	100%
March	99%	1%	100%
April	99%	1%	100%
May	99%	1%	100%
June	99%	1%	100%
July	98%	2%	100%
August	96%	4%	100%
September	98%	2%	100%
October	98%	2%	100%
November	88%	12%	100%
December	82%	18%	100%
⊟ 2021	83%	17%	100%
January	88%	12%	100%
February	75%	25%	100%
March	80%	20%	100%
April	86%	14%	100%
Total	95%	5%	100%

Is Vacuna

Year	False	True	Total
⊟ 2020	129,453	4,230	133,683
January	8	1	9
February	43	1	44
March	11,756	75	11,831
April	16,940	179	17,119
May	17,314	201	17,515
June	19,108	174	19,282
July	14,290	308	14,598
August	15,279	583	15,862
September	11,928	255	12,183
October	9,796	235	10,031
November	6,867	912	7,779
December	6,124	1,306	7,430
⊟ 2021	17,963	3,753	21,716
January	7,160	1,022	8,182
February	3,454	1,139	4,593
March	4,366	1,100	5,466
April	2,983	492	3,475
Total	147,416	7,983	155,399

UNIVERSIDAD AUSTRAL

Análisis de Resultados

Is coronavirus

Clasif.	Cant. Coronavirus	Cant Tweets Totales	%
neg	72,496	111,393	65.08%
neu	3,398	7,291	46.61%
pos	21,266	36,715	57.92%
Total	**97,160**	**155,399**	**62.52%**

Is Cuarentena

Clasif.	Cant. Cuarentena	Cant Tweets Totales	%
neg	6,975	111,393	6.26%
neu	289	7,291	3.96%
pos	1,964	36,715	5.35%
Total	**9,228**	**155,399**	**5.94%**

Is Quedate en Casa

Clasif.	Cant. Quedate en Casa	Cant Tweets Totales	%
neg	2,801	111,393	2.51%
neu	134	7,291	1.84%
pos	674	36,715	1.84%
Total	**3,609**	**155,399**	**2.32%**

Is Casos

Clasif.	Cant Casos	Cant Tweets Totales	%
neg	18886	111,393	16.95%
neu	597	7,291	8.19%
pos	2905	36,715	7.91%
Total	**22388**	**155,399**	**14.41%**

Is Muerte

Clasif.	Cant. Muerte	Cant Tweets Totales	%
neg	9,651	111,393	8.66%
neu	116	7,291	1.59%
pos	532	36,715	1.45%
Total	**10,299**	**155,399**	**6.63%**

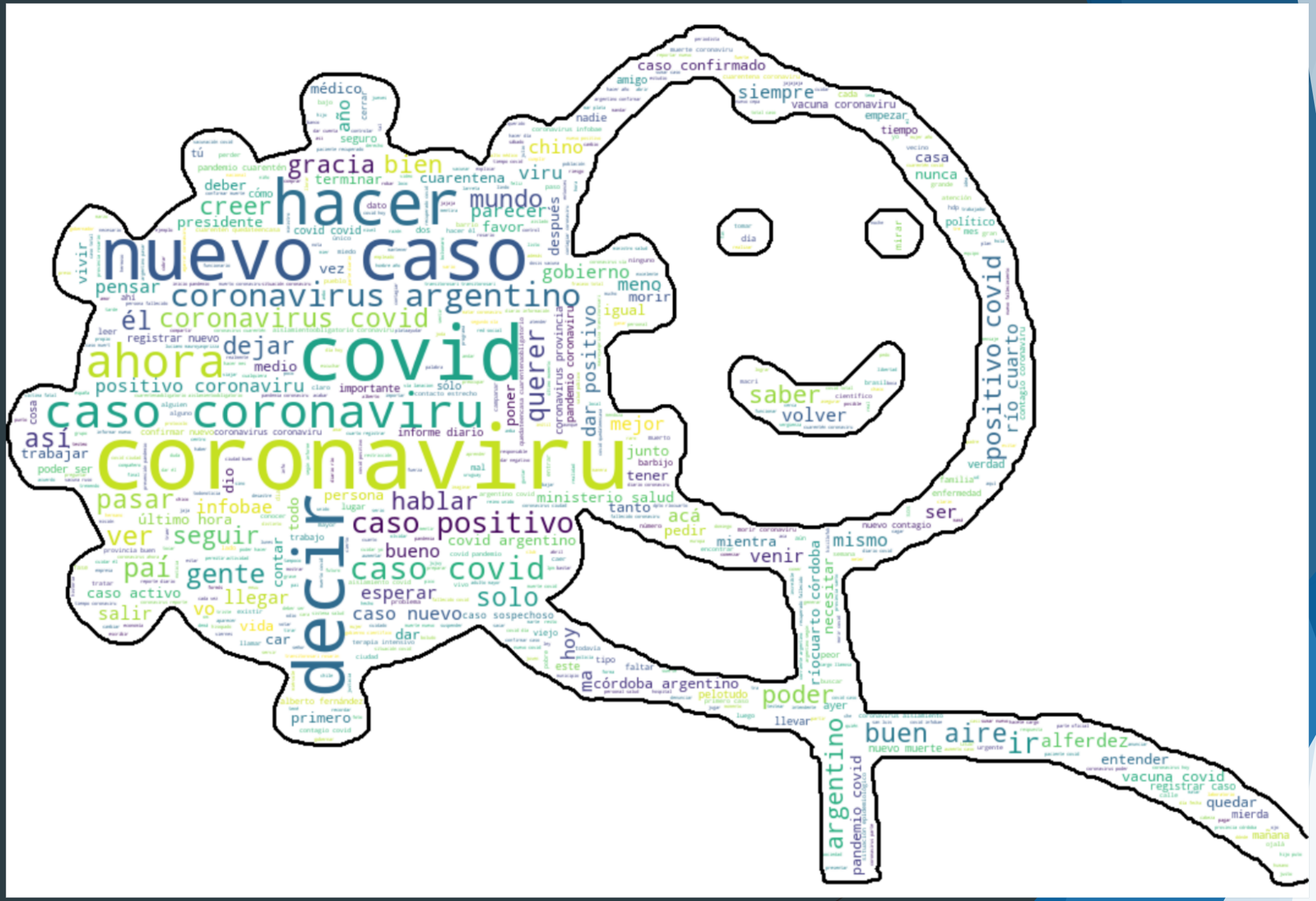
Is Vacuna

Clasif.	Cant, Vacuna	Cant Tweets Totales	%
neg	6,066	111,393	5.45%
neu	204	7,291	2.80%
pos	1,713	36,715	4.67%
Total	**7,983**	**155,399**	**5.14%**

UNIVERSIDAD AUSTRAL

Conclusiones

- Tendencia en cantidad de tweets:
  - Fuerte crecimiento al principio del período
  - Decaimiento con el correr del tiempo
- No hay marcadas diferencias en cuanto al sentimiento sin importar los meses. Resultados promedio:
  - Positivos: 24%
  - Neutrales: 5%
  - Negativos: 71%
- Investigación a posteriori:
  - Interpretación de tweets irónicos
  - Reentrenamiento del modelo
  - Fine-tuning
  - Análisis de tweets principales, re-tweets y respuestas


caso confirmado
amigo
médico
año
cerrar
seguro
siempre
vacuna coronaviru
empezar
tiempo
chino
gracia
bien
tú
pandemio cuarentén
terminar
cuarentena
viru
casa
nunca
deber
cómo
creer
hacer
mundo
parecer
presidente
covid covid
favor
después
político
mes
vivir
nuevo caso
vez
gobierno
pensar
coronavirus
argentino
meno
morir
igual
él
coronavirus covid
positivo covid
río cuarto
contagio coronaviru
registrar nuevo
dejar
covid
querer
ahora
medio
positivo coronaviru
importante
sólo
contacto estrecho
dar positivo
pandemio coronaviru
coronavirus provincia
saber
volver
caso coronaviru
así
trabajar
poder ser
informe diario
mejor
junto
barbijo
tener
verdad
familia
enfermedad
coronaviru
pasar
infobae
último hora
persona
hablar
ministerio salud
tanto
acá
ser
caso positivo
ver
seguir
todo
decir
pedir
mientra
mismo
venir
nuevo contagio
covid argentino
bueno
caso covid
país
gente
contar
solo
esperar
caso activo
llegar
córdoba
necesitar
salir
vo
caso nuevo
caso sospechoso
vida
car
dar
hoy
este
ma
tipo
faltar
peor
córdoba argentino
pelotudo
poder
primero
alberto fernández
ayer
llevar
buen aire
ir
alferdez
nuevo muerte
entender
vacuna covid
registrar caso
quedar
mierda
mañana
argentino
pandemio covid