

notebook

October 6, 2020

```
[1]: from config import *
     from utils import *
     import pandas as pd
```

1 Leer pickle

```
[2]: df = pd.read_pickle(pkl_file)
     df.head()
```

```
[2]:
```

	tweet_id	tweet_date	tweet_screen_name	\
0	1096977737930158080	2019-02-17	gabrielguerronv	
1	1096977884848082944	2019-02-17	jhonav00	
2	1096981630881882113	2019-02-17	Cristhyn3000	
3	1096982892759130112	2019-02-17	PelotonFusilero	
4	1096983080072556544	2019-02-17	carosu_92	

	tweet_text
0	@AlexTonello5RC @CompromisoRC5 @PaolaPabonC @M...
1	@PaolaPabonC @CompromisoRC5 @MashiRafael @Alex...
2	Que importante ha sido sentir el calor de los ...
3	Mi campaña no tiene pactos políticos ni aporte...
4	¡Lo más avergonzarte que he visto cómo mujer! ...

2 Top 10 candidatos más activos

```
[3]: candidates_tweets, candidates_lastnames = top_candidates(df,
    ↪only_manifestos=False)
```

Total tweets: 964439

Total tweets from candidates: 10046

The top 10 most active candidates are:

tweet_screen_name

HernanUlloa	1183
abenavidesgol1	1070
wgomezr	804
jimmyjairala	662

```
LuisaMaldonadoM      574
LoroHomero            509
VickyDesintonio      477
CynthiaViteri6       357
fcojimenez21         335
davidroserow         305
Name: count, dtype: int64
```

3 Top 10 palabras por candidato

```
[4]: candidates_words, tf_idf, vocabulary, top_words_set =   
      ↪top_words(candidates_tweets)
```

The entire vocabulary has size 10852

Top 10 words for candidates:

```
abenavidesgol1 ['triunfo', 'copa', 'liga', 'empate', 'quito', 'ecuador',  
'casa', 'ligapro', 'independiente', 'encuentro']
```

```
VickyDesintonio ['participación', 'consejo', 'ciudadanos', 'marzo', 'ciudadana',  
'voto', 'ecuatorianos', 'violencia', 'control', 'vamos']
```

```
jimmyjairala    ['guayaquil', 'lee', 'nota', 'ciudad', 'alcaldía', 'hoy',  
'vamos', 'multas', 'mira', 'ser']
```

```
HernanUlloa     ['voto', 'candidato', 'casillero', 'ser', 'mejor', 'candidatos',  
'ulloa', 'cpccs', 'votar', 'gracias']
```

```
wgomezr ['pueblo', 'prov', 'voto', 'voluntad', 'debe', 'ser', 'entrevista',  
'apoyo', 'ecuador', 'gracias']
```

```
CynthiaViteri6  ['guayaquil', 'ciudad', 'cada', 'progreso', 'sueños', 'obras',  
'seguiremos', 'bienestar', 'vamos', 'hoy']
```

```
LoroHomero      ['alcalde', 'jorge', 'yunda', 'quito', 'gracias', 'ciudad',  
'hoy', 'trabajo', 'sur', 'campaña']
```

```
fcojimenez21    ['vamos', 'guayaquil', 'marzo', 'ciudad', 'vía', 'municipal',  
'hoy', 'vivo', 'junto', 'obra']
```

```
LuisaMaldonadoM ['vamos', 'ciudad', 'gracias', 'pueblo', 'recuperar', 'lista',  
'hoy', 'alcaldía', 'vecinos', 'quito']
```

```
davidroserow    ['voto', 'david', 'gracias', 'apoyo', 'vamos', 'candidato',  
'rosero', 'consejero', 'participación', 'consejo']
```

4 TF-IDF de una palabra

```
[5]: tf_idf[tf_idf['word'] == 'quito']
```

```
[5]:
```

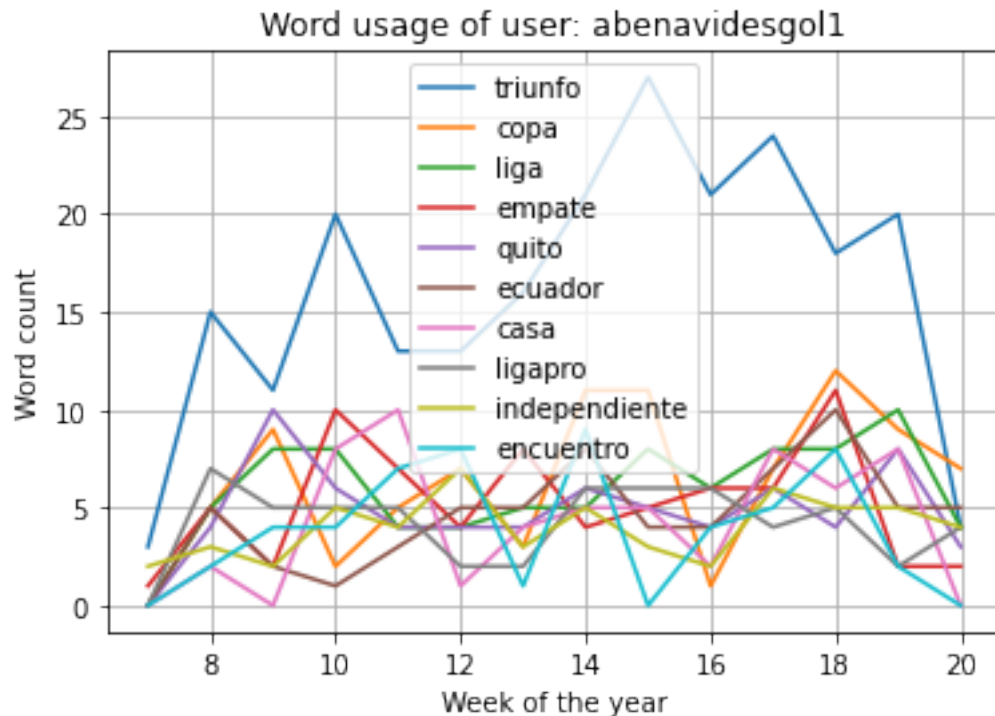
	word	tweet_id	tf	df	idf	tf-idf
119244	quito	1116187589227184128	0	21	5.699966	0.0
119245	quito	1101234475579723777	0	21	5.699966	0.0
119246	quito	1098678459084230658	0	21	5.699966	0.0
119247	quito	1105329867040665601	0	21	5.699966	0.0
119248	quito	1097897607265955841	0	21	5.699966	0.0
119249	quito	1104019174031060993	0	21	5.699966	0.0
119250	quito	1101300571980603392	0	21	5.699966	0.0
119251	quito	1114521333432508418	0	21	5.699966	0.0
119252	quito	1108462928376877056	0	21	5.699966	0.0
119253	quito	1106253053815148546	0	21	5.699966	0.0
119254	quito	1108194179593748480	0	21	5.699966	0.0
119255	quito	1106545126279651329	0	21	5.699966	0.0
119256	quito	1100094126136139776	0	21	5.699966	0.0
119257	quito	1099459461129875457	0	21	5.699966	0.0
119258	quito	1120881979807928321	0	21	5.699966	0.0
119259	quito	1098590236601389060	0	21	5.699966	0.0
119260	quito	1100257022396846081	0	21	5.699966	0.0
119261	quito	1116786807352561664	0	21	5.699966	0.0
119262	quito	1110616081780166658	0	21	5.699966	0.0
119263	quito	1129042280189448194	0	21	5.699966	0.0
119264	quito	1101264587960320001	0	21	5.699966	0.0
119265	quito	1127012287900995585	0	21	5.699966	0.0
119266	quito	1100617960077643776	0	21	5.699966	0.0
119267	quito	1108193852924538882	0	21	5.699966	0.0
119268	quito	1113278052740620288	0	21	5.699966	0.0
119269	quito	1104186773314433026	0	21	5.699966	0.0
119270	quito	1100635886956486657	0	21	5.699966	0.0
119271	quito	1111338080651354112	0	21	5.699966	0.0
119272	quito	1101533134405939203	0	21	5.699966	0.0
119273	quito	1108848921474928641	0	21	5.699966	0.0
...
125490	quito	1098206796219645953	0	21	5.699966	0.0
125491	quito	1098092958153363456	0	21	5.699966	0.0
125492	quito	1112923448400121856	0	21	5.699966	0.0
125493	quito	1101150955045834752	0	21	5.699966	0.0
125494	quito	1100740029905846273	0	21	5.699966	0.0
125495	quito	1112139634472075264	0	21	5.699966	0.0
125496	quito	1100813215934623744	0	21	5.699966	0.0
125497	quito	1102018338925015041	0	21	5.699966	0.0
125498	quito	1105299903583084545	0	21	5.699966	0.0
125499	quito	1120026046106148865	0	21	5.699966	0.0
125500	quito	1100909171090882560	0	21	5.699966	0.0

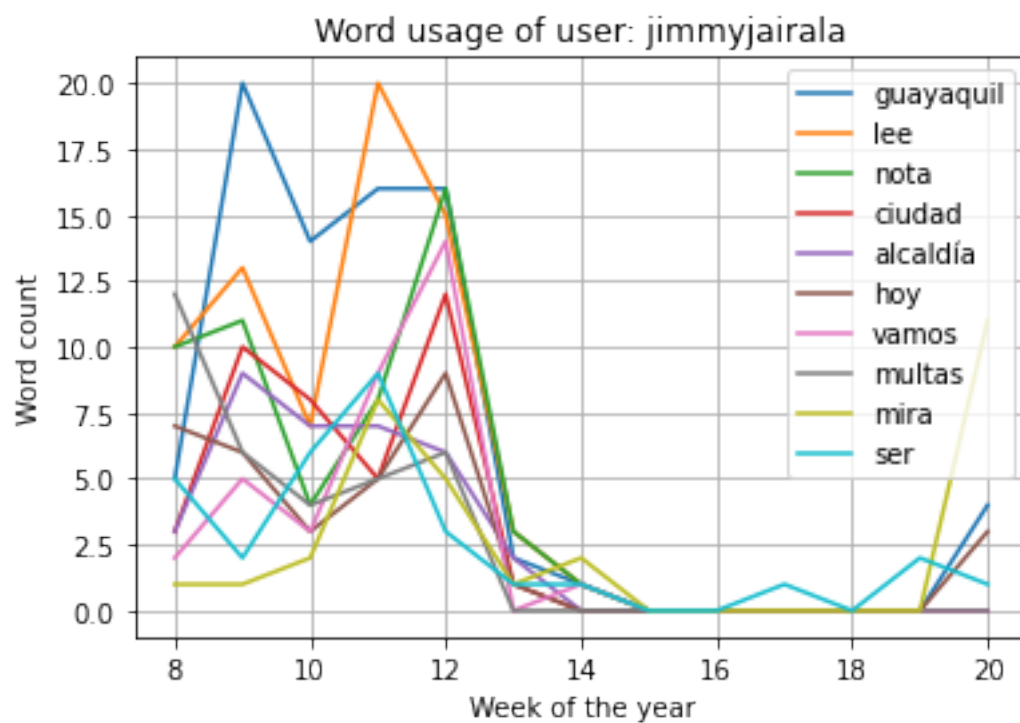
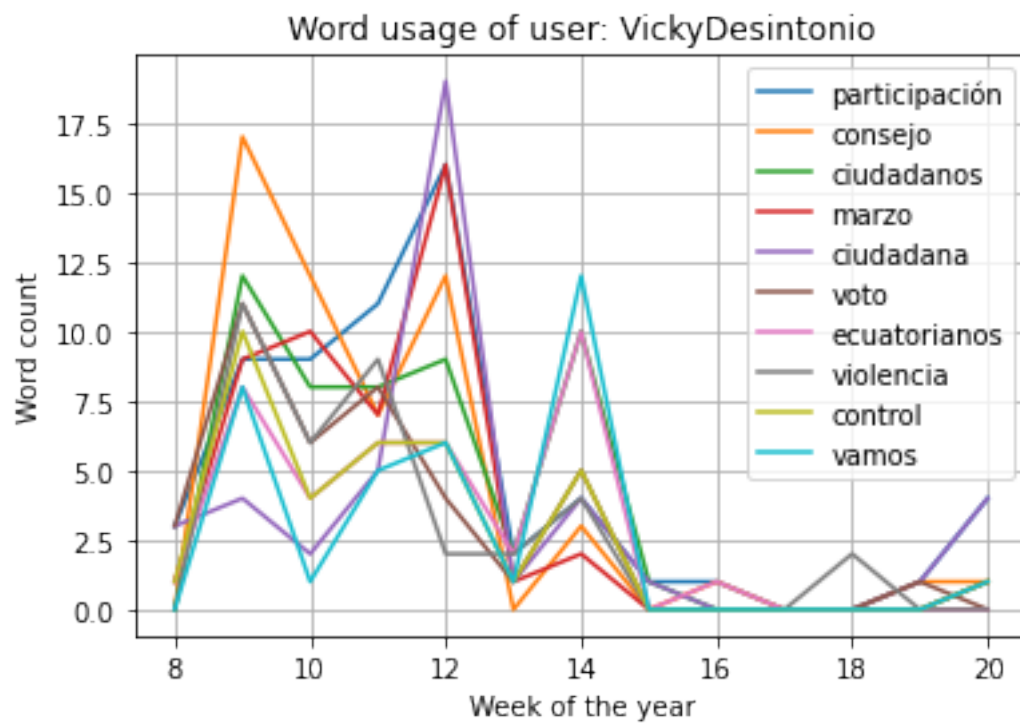
125501	quito	1099369548954955778	0	21	5.699966	0.0
125502	quito	1099895837893775360	0	21	5.699966	0.0
125503	quito	1109674060210802688	0	21	5.699966	0.0
125504	quito	1097307536724840449	0	21	5.699966	0.0
125505	quito	1104422650766655490	0	21	5.699966	0.0
125506	quito	1097716750857179136	0	21	5.699966	0.0
125507	quito	1104043026769473537	0	21	5.699966	0.0
125508	quito	1106380301679382529	0	21	5.699966	0.0
125509	quito	1101123056066576390	0	21	5.699966	0.0
125510	quito	1102337333129551872	0	21	5.699966	0.0
125511	quito	1109117350786318336	0	21	5.699966	0.0
125512	quito	1101494010642878474	0	21	5.699966	0.0
125513	quito	1108001826249625603	0	21	5.699966	0.0
125514	quito	1117909615100874752	0	21	5.699966	0.0
125515	quito	1109825429345615873	0	21	5.699966	0.0
125516	quito	1098405815499874304	0	21	5.699966	0.0
125517	quito	1117471827863973888	0	21	5.699966	0.0
125518	quito	1111377502839013379	0	21	5.699966	0.0
125519	quito	1103628553919303680	0	21	5.699966	0.0

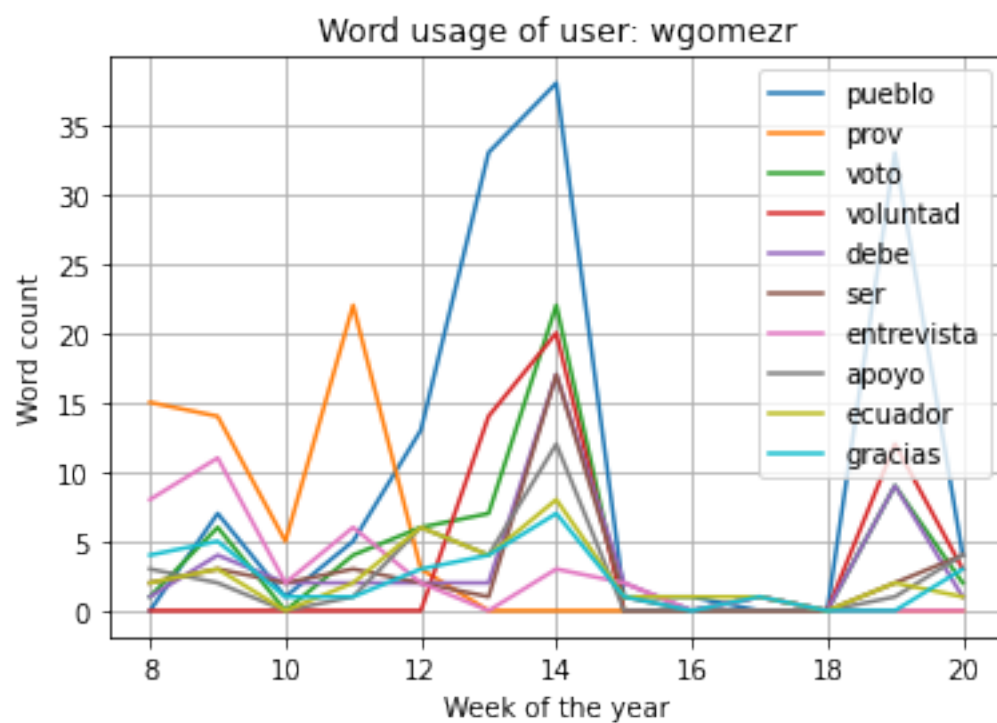
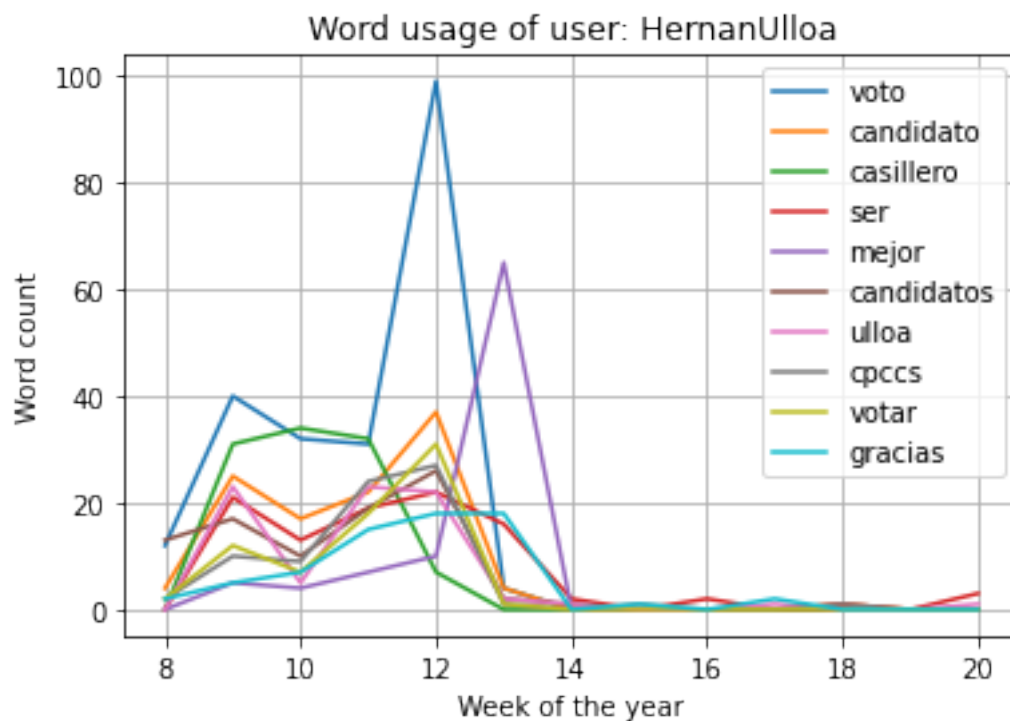
[6276 rows x 6 columns]

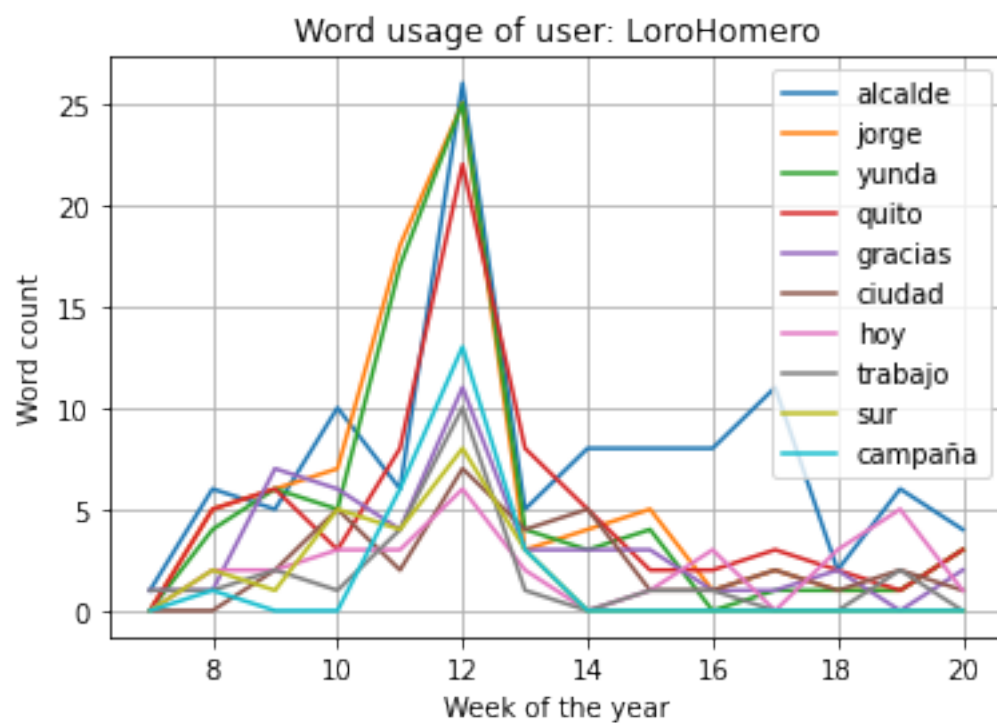
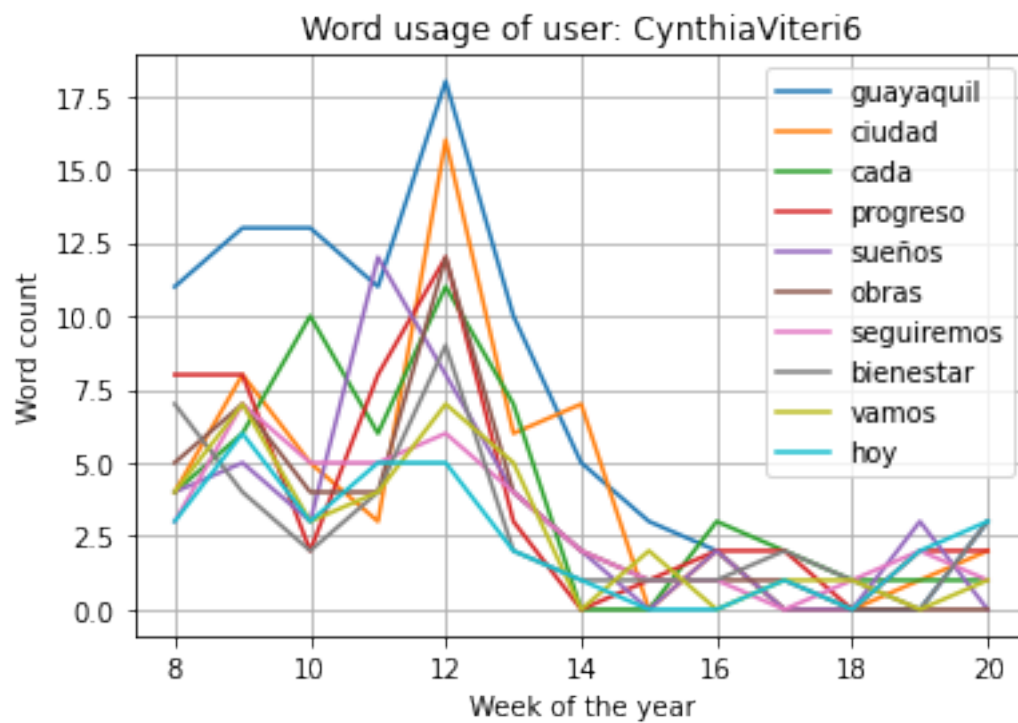
5 Timeline de palabras por candidato

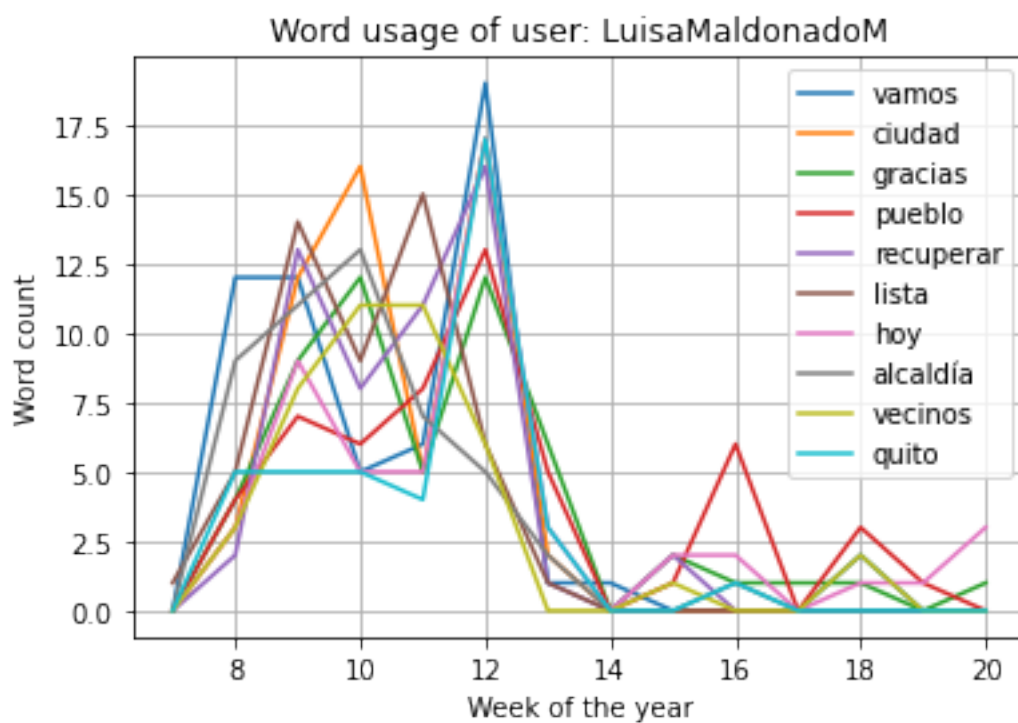
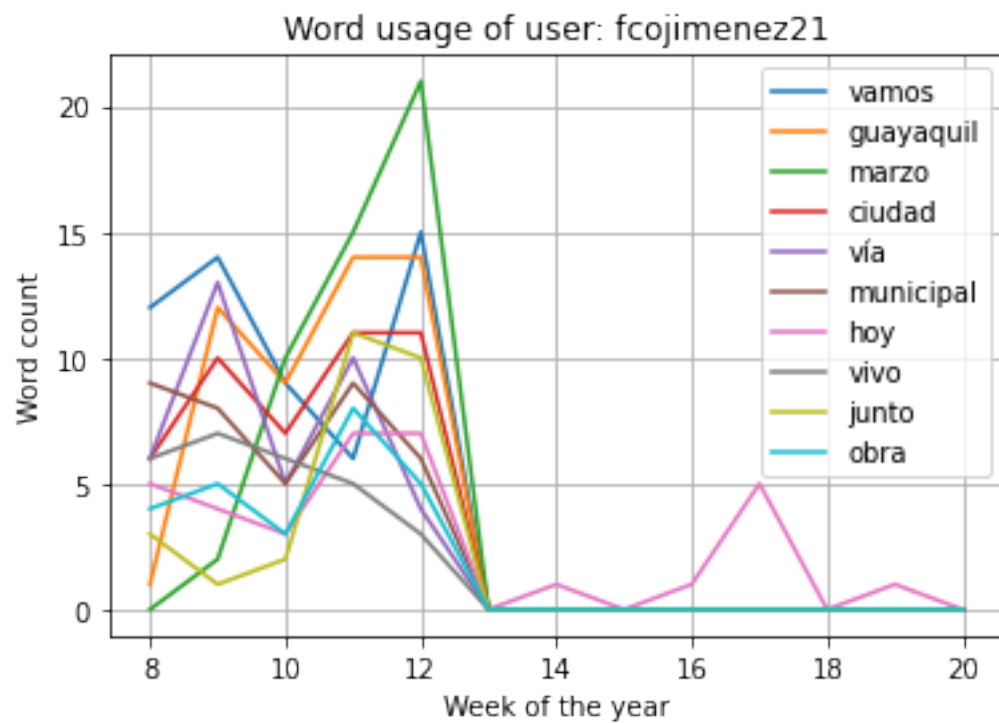
```
[6]: time_line(candidates_words, candidates_tweets)
```

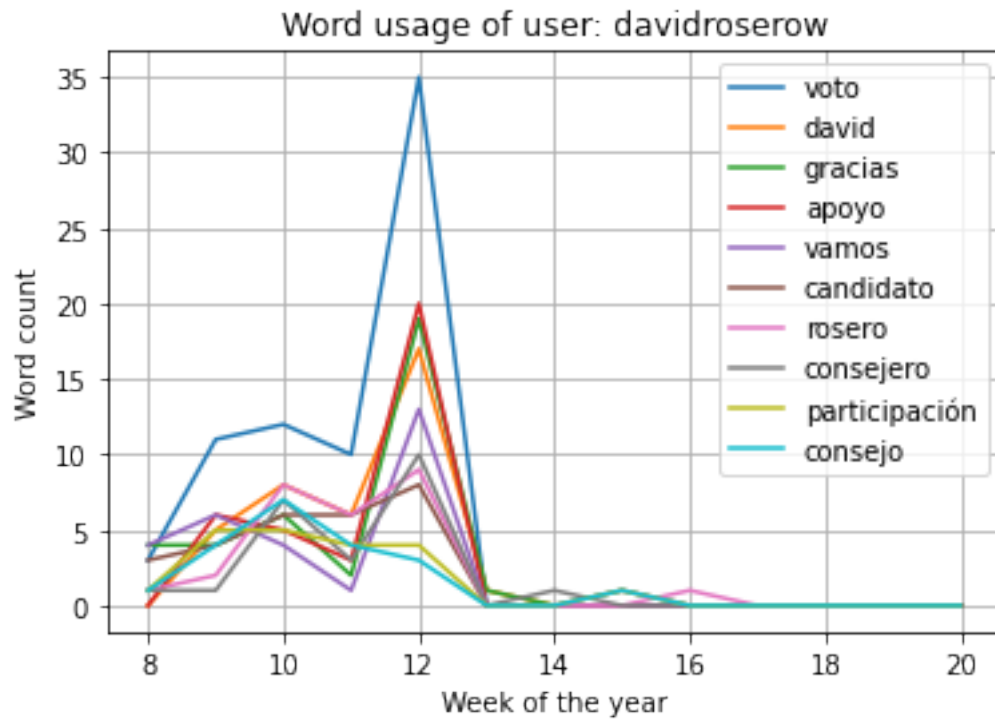












6 Ranking de candidatos por query

```
[7]: total_user_vectors = ranking('alcaldía quito ciudad trabajamos empresa',  
    ↪ vocabulary, candidates_tweets, tf_idf)
```

The query vector is:

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Total score: 1.9232760536479185

Top ranked tweet:

#QuéProponen

El candidato a la alcaldía por la Lista 3, @abenavidesgol1, plantea convertir a Quito como a una ciudad digital y además la creación de paradas inteligentes para el Metro de Quito.

(Foto de: @elcomerciocom) <https://t.co/HCPgDifTiD>

Tweet score: 0.5979506707338352

7. VickyDesintonio

Total score: 1.3574040735259947

Top ranked tweet:

#ManabíEsVictoria || Resumen de actividades: Participé de Conversatorio con Organizaciones de la Sociedad Civil, al que fui invitada en la ciudad de #Portoviejo

#VictoriaDeLaPatria

#EleccionesCPCCS #CPCCS <https://t.co/KAAnoYEtrb>

Tweet score: 0.27351753365908077

8. wgozmezr

Total score: 1.0436255160504972

Top ranked tweet:

En Playita Mía conversando con los líderes de las Asociaciones de comerciantes pesqueros de la ciudad de Manta #2marzo #FelizSabado #2Mar #Carnaval2019

<https://t.co/2NYxD3qKv3>

Tweet score: 0.27351753365908077

9. HernanUlloa

Total score: 0.3374345828047828

Top ranked tweet:

@HernanUlloa @HectorYepezM @CPCCS @AsambleaEcuador En eso t apoyo piden voto nulo para el consejo y x no propongo eliminar alcaldía prefectura asambleísta si en todos lados se ve corrupción nos quiere engañar.

Tweet score: 0.3374345828047828

10. davidroserow

Total score: 0.0

Top ranked tweet:

<https://t.co/6dv547YsPn>

Tweet score: 0.0

7 Similaridad de actividad entre candidatos

```
[8]: similarity(total_user_vectors)
```

```
[8]:      abenavidesgol1 VickyDesintonio jimmyjairala HernanUlloa \
abenavidesgol1      1      0.0429547      0.0503325      0.039187
VickyDesintonio    0.0429547      1      0.433083      0.309011
jimmyjairala      0.0503325      0.433083      1      0.316765
HernanUlloa      0.039187      0.309011      0.316765      1
wgomezr          0.0303906      0.485281      0.273165      0.375044
CynthiaViteri6    0.0400067      0.192576      0.368372      0.163848
LoroHomero        0.0715069      0.25897      0.546419      0.390514
fcojimenez21      0.0471326      0.302195      0.628873      0.166514
LuisaMaldonadoM   0.0596091      0.506937      0.541197      0.308394
davidroserow      0.0323243      0.396988      0.237703      0.557377

      wgomezr CynthiaViteri6 LoroHomero fcojimenez21 \
abenavidesgol1    0.0303906      0.0400067      0.0715069      0.0471326
VickyDesintonio  0.485281      0.192576      0.25897      0.302195
jimmyjairala     0.273165      0.368372      0.546419      0.628873
HernanUlloa     0.375044      0.163848      0.390514      0.166514
wgomezr         1      0.131742      0.341901      0.15873
CynthiaViteri6  0.131742      1      0.263352      0.339707
LoroHomero      0.341901      0.263352      1      0.361067
fcojimenez21    0.15873      0.339707      0.361067      1
LuisaMaldonadoM 0.457227      0.31868      0.579546      0.456204
davidroserow    0.417479      0.157873      0.362314      0.255705

      LuisaMaldonadoM davidroserow
abenavidesgol1    0.0596091      0.0323243
VickyDesintonio  0.506937      0.396988
jimmyjairala     0.541197      0.237703
HernanUlloa     0.308394      0.557377
wgomezr         0.457227      0.417479
CynthiaViteri6  0.31868      0.157873
LoroHomero      0.579546      0.362314
fcojimenez21    0.456204      0.255705
LuisaMaldonadoM 1      0.378554
davidroserow    0.378554      1
```

8 Similaridad entre actividad de candidatos con sus manifiestos

```
[9]: deviation(candidates_words, candidates_lastnames, vocabulary,
→candidates_tweets, total_user_vectors, tf_idf)
```

```
Manifiesto for desintonio not found.
Manifiesto for jairala not found.
Manifiesto for ulloa not found.
Manifiesto for gomez not found.
Manifiesto for viteri not found.
Manifiesto for jimenez not found.
```

Manifiesto for rosero not found.

Euclidean distance of reduced vocabulary vectors of tweets vs. manifestos

