Nikolas Dimitrio Badani Gasdaglis 20092
Juan Angel Carrera Soto 20593
Data Science
Sección 10

Laboratorio 6 : Análitica de Redes Sociales

Problema 1

Análisis Exploratorio :

```
> glimpse(traficogt)
Rows: 12,631
Columns: 29
$ ...1             <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, …
$ id               <dbl> 1701654244858679808, 1701651855212691968, 170134845391…
$ id_str           <dbl> 1701654244858679808, 1701651855212691968, 170134845391…
$ url              <chr> "https://twitter.com/EmisorasUnidas/status/17016542448…
$ date             <dttm> 2023-09-12 17:49:21, 2023-09-12 17:39:52, 2023-09-11 …
$ user             <chr> "{'id': 40256008, 'id_str': '40256008', 'url': 'https:…
$ lang             <chr> "es", "es", "es", "es", "es", "es", "es", "es", "es", …
$ rawContent       <chr> "#AHORA Amílcar Montejo, director de Comunicación de E…
$ replyCount       <dbl> 1, 149, 2, 3, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 2, 2, 0…
$ retweetCount     <dbl> 2, 78, 1, 17, 0, 0, 0, 0, 1, 0, 0, 0, 3, 0, 0, 1, 2, 0…
$ likeCount        <dbl> 8, 524, 4, 95, 20, 1, 1, 6, 39, 2, 6, 3, 7, 7, 1, 4, 1…
$ quoteCount       <dbl> 0, 49, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,…
$ conversationId   <dbl> 1701654244858679808, 1701651855212691968, 170134845391…
$ hashtags         <chr> "['AHORA', 'TráficoGT']", "[]", "['transitogt', 'trafi…
```

```
> table(traficogt$lang)

  ar   art    ca    de    en    es    et    eu    fr    in    it    ja    ko
   2     2     5     1   600 11946     1     2     9     4     6     3     2
  lt    pt   qht   qme    ru    th    tr   zxx
   4     6     3     8     1     1     1    24
>
```

```
> table(traficogt$replyCount)

   0    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15
8770 1704  504  208  159  105   76   48   37   30   41   25   24   35   16    9
  16   17   18   19   20   21   22   23   24   25   26   27   28   29   30   31
  20    8   13   17   14   28   26   18   10   11   24   15   15   11   16    6
  32   33   34   35   36   37   38   39   40   41   42   43   44   45   46   47
  10   14    5    5    6    7    4   14    2    9    2    7    5    6    2    2
  48   49   50   51   52   53   54   55   56   57   58   59   60   61   62   63
   8    3    6    6    4    1    6    5    6    2    6    2    3    1    8    1
  65   66   67   68   69   70   71   72   73   74   75   76   77   80   82   83
   2    8   10    3    3    5    2    3    1    1    2    3    2    4    7    3
  84   85   86   87   89   90   91   93   95   97   99  100  102  103  104  106
   1    1    4    4    1    3    1    2    2    3    4    2    1    1    2    1
 107  110  114  115  118  119  120  121  123  124  127  129  131  132  134  135
   1    1    2    3    3    1    2    4    2    2    1    3   12    1    2    6
 143  144  146  147  149  151  153  157  160  161  162  163  167  169  170  171
   1    1    2    3    3    1    3    1    1    2   10    7    2    1    2    1
 172  174  175  177  179  181  183  184  185  194  197  198  203  207  208  209
   2    9    1    1    2    4    1    2    1    4    1    2    1    2    4    1
 210  212  213  214  219  224  226  229  233  236  237  238  245  249  250  255
   2    1    1    2    1    4    1    1    2    2    2    4    1    1    2    1
 260  266  267  272  286  292  302  308  317  328  338  340  360  361  371  375
   6    1    2    4    2    6    3    1    1    1    5    9    7    1    4    1
 384  387  401  405  427  438  439  444  453  455  458  459  502  508  578  617
   1    1    1    5    1    2    1    1    1    5    1    2    1    6    5    1
 638  644  652  739  763  777  779  798  811  831  841  846  872  898  935  968
   2    1    1    1    2    4    6    3    1    2    2    1    2    1    3    2
 994 1004 1021 1034 1037 1127 1206 1208 1269 1339 1340 1347 1389 1394 1398 1406
   1    2    2    2    2    1    2    2    2    1    1    2    1    2    3    1
1407 1544 1634 1660 1690 1801 1909 2002 2538 2755 4771 5305 5596 8134 8136 8610
   1    1    2    1    1    3    2    1    3    1    1    2    1    2    1    1
8611
   1
>
```

Según el análisis realizado, se pudo encontrar que el 69.4% de los tweets (8,770) no recibieron ningún tipo de respuesta luego de haber sido posteados. Lo cual puede ayudar a deducir que la información que fue posteada en ellos no está siendo considerada de gran importancia, no tienen relación con el tema del tráfico, o simplemente no generan interés entre los usuarios de Twitter.

```
> table(traficogt$retweetCount)

   0    1    2    3    4    5    6    7    8    9   10   11   12
7923 1407  689  351  214  185  118   90   66   53   54   58   40
  13   14   15   16   17   18   19   20   21   22   23   24   25
  53   32   17   33   23   18   23   21   21   14    9   12   12
  26   27   28   29   30   31   32   33   34   35   36   37   38
  19   14    9   13    9   13    4   11   10   14   16   19    5
  39   40   41   42   43   44   45   46   47   48   49   50   51
   8    3    5   11    8    5    2    5    7    4    8    9    9
  52   53   54   55   56   57   58   59   60   61   62   63   64
  15    2   18    8   10    5    3    9    6    1    8    2    7
  65   66   67   68   69   70   71   72   73   74   76   77   78
   1    7    5    5    7   10    5    3    2    9    7    8    7
  79   80   81   82   83   84   85   86   87   88   89   90   91
   3    3    2    6    4    6    4    6    1    1    6    1    2
  92   93   94   95   96   97   98  101  102  103  104  106  108
   5    4    2    3    6    4    5    1    1    7    1    2    5
 109  110  111  112  113  114  115  117  118  119  120  121  122
   6   11    6    1    4    4    2    1    3    5    2    3    6
 123  124  125  126  127  129  131  132  134  136  137  138  139
   4    1    3    2    2    1    2    8    1    2    5    4    3
 140  141  143  144  145  146  147  148  149  151  152  153  155
   2    3    4    1    1    5    1    3    2    1    3    9    1
 156  157  158  159  160  161  162  163  166  167  168  169  170
   2    2    3    5    1    1    4    1    2    1    1    2    1
 172  174  175  176  178  179  180  181  182  184  186  188  189
   4    3    2    2    2    1    1    1    3    3    2    3    4
 190  195  197  199  200  201  202  204  205  206  207  209  210
   2    2    1    2    1    1    4    2    1    3    3    2    4
 211  212  213  214  216  218  219  220  221  222  223  224  225
   2    2    1    3    1    1    2    1    4    5    1    3    2
 226  228  229  232  234  236  239  240  242  243  244  245  246
   5    2    1    2    1    1    1    1    2    1    2    1    8
 247  248  249  254  256  260  262  263  265  270  272  273  274
```

Con los resultados obtenidos, se pudo determinar que el 62.7% de los tweets del dataset (7923) no fueron retuiteados. Lo que significa que la información de esos tweet no fue compartida a más personas. Por lo cual, se puede deducir que la información que contenían dichos tweets no fue considerada de importancia en relación con el tema del tráfico.

```
> table(traficogt$likeCount)

    0     1     2     3     4     5     6     7     8     9    10
 4533  2007   976   657   443   338   273   194   171   156   132
   11    12    13    14    15    16    17    18    19    20    21
  113    68    96    84    72    57    47    48    45    57    60
   22    23    24    25    26    27    28    29    30    31    32
   51    36    28    33    42    17    23    20    20    23    12
   33    34    35    36    37    38    39    40    41    42    43
   17    17    26    14    18    11    23     1    12     9    13
   44    45    46    47    48    49    50    51    52    53    54
   12     6    11    11    13    16     8     8     4    10    15
   55    56    57    58    59    60    61    62    63    64    65
    5     6    10     8    11     8     1     5     8     2     1
   66    67    68    69    70    71    72    73    74    75    76
    8     5     2     5     3     8     4     4     6     4     3
   77    78    79    80    81    82    83    84    85    86    87
    9     2     3    10     5     1     4     3     4     4     5
   89    90    91    92    93    94    95    96    97    98   101
    2     2     7     2     5     6     8     1     5     2     5
  102   103   105   106   107   108   111   112   113   114   115
    4     3     2     2     5     6     4     7     3     5     4
  116   117   118   120   123   124   125   126   127   129   131
    2     3     4     3     4     1     5     3     2     9     1
  132   133   134   135   136   137   138   139   140   141   142
    2     7     3     3     6     9     1     4    10     4     4
  146   147   148   149   150   151   152   153   154   155   157
    1     3     2     4     4     3     1     1     3     5     2
  160   162   163   166   167   168   170   171   172   173   174
```

```
> table(traficogt$quoteCount)

    0     1     2     3     4     5     6     7     8     9    10    11    12
10533   716   293   165   100    79    80    48    26    31    17    25    14
   13    14    15    16    17    18    19    20    21    22    23    24    25
   13    12    13    16     9    14    16    17    13    10     8     8     9
   26    27    28    29    30    31    32    33    34    35    36    37    38
   22    17     4     6     2    11     2    12     3     5     5     7     4
   39    40    41    42    43    44    45    46    47    49    50    51    52
    2     2     7     2     6     3     9     7     3     3     2     1     1
   53    54    56    57    61    62    63    64    65    67    72    73    75
    9     1     1     1     8     3     2     4     8     2     2     4     3
   77    78    81    82    84    86    87    88    93    94   100   101   102
    1     4     2     2     6     1     2     1     3     1     1     1     1
  104   111   113   123   124   127   130   135   136   139   141   146   148
    2    12     1     1     1     7     3     6     2     1     5     5     3
  152   155   161   164   165   167   185   201   207   213   214   217   223
    1     1     1     9     1     1     1    12     1     1     1     1     1
  229   254   264   303   331   334   335   336   349   356   363   381   389
    2     2     3     1     1     1     2     1     2     5     1     1     1
  391   462   481   592   634  1167  1205  1215  1501  1723  3331  4133
    2     1     3     2     1     1     1     3     1     3     1     1
>
```

Una vez realizado el análisis, se pudo encontrar que el 35.8% (4533) de los tweets del dataset no recibieron ningún “Like” o “Me Gusta”. A su vez, también se pudo determinar que el 83.3% de los tweets que conforman el dataset no posee ningún cita o “quote” en la

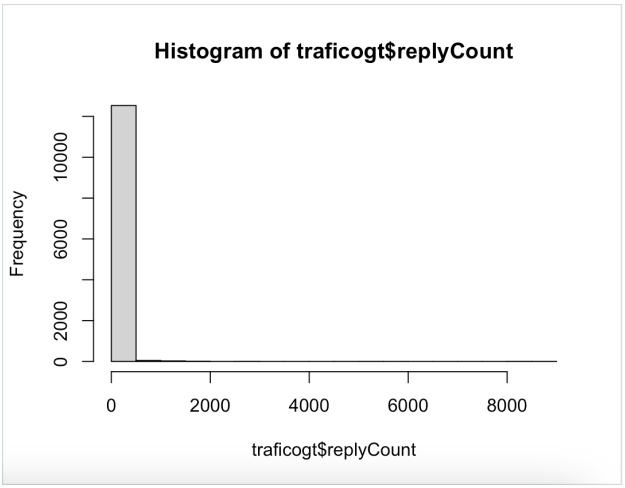
información que se redactó cuando fueron posteados. Debido a que se conoce que las citas o “quotes” se utilizan para cuando se desea tuitear o retuitear un tweet que tiene información relevante, se puede concluir que los tweets mencionados no son considerados como relevantes en relación al tema del tráfico.

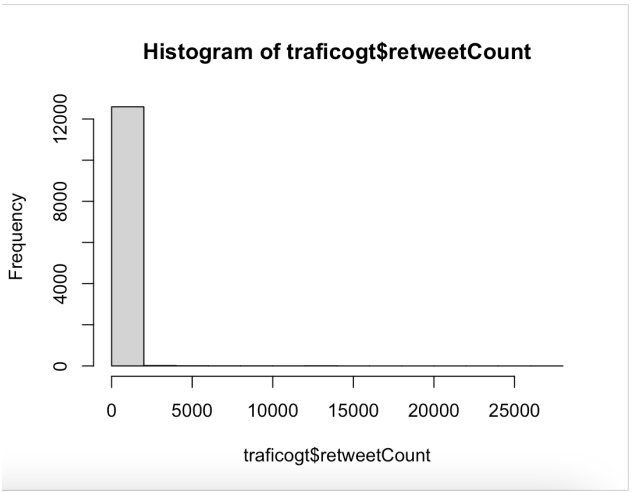
```
> table(traficogt$cashtags)

              ['AAPL', 'AAPL'] ['BTC', 'ETH', 'USDT', 'USDC']
                             1                              1
                       ['EOS']                        ['LINK']
                             6                              1
                      ['ROSE']                       ['SUSHI']
                             1                              1
                            []
                         12620
>
```

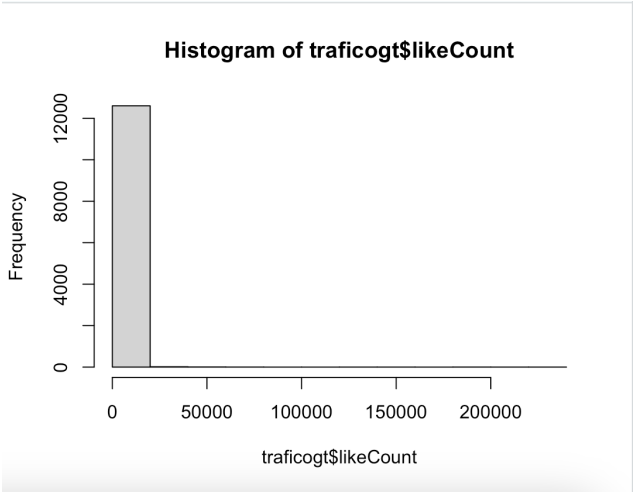
```
> table(traficogt$sourceLabel)

          advertiser-interface                AeddonFollowers
                           358                              2
                        Buffer                         Emplifi
                            10                              5
                erased5434447        erased972529_fzyRVGtcON
                             6                              7
          Flamingo for Android                         Google
                             1                              1
Guatevision_tv1645969812290404             hivemedia-ads-crud
                             1                              5
                Hootsuite Inc.                      Instagram
                            14                              6
   lahoragt1644885914282258435                     Metricool
                             6                              2
                      OneSight                     OneUp App
                             8                              2
                     Periscope              Plume for Android
                             9                              1
                     Publer.io                  República App
                             2                              1
            Seismic LiveSocial                   simpleads-ui
                             1                             37
                    SocialFlow                   SocialOomph
                             7                              1
           Sprinklr Publishing                 Talon Android
                            12                              2
              The White House               Tweet Hunter Pro
                             1                              8
             Tweetbot for iOS                     TweetDeck
                             2                           1292
           TweetDeck Web App                    Twitter Ads
                           243                              7
       Twitter for Advertisers            Twitter for Android
                           237                           5737
       Twitter for BlackBerry®               Twitter for iPad
                             2                             30
```

El análisis realizado demostró que de todos los usuarios que conforman el dataset el 45.4% (5737) de ellos pertenecen al Twitter para Android. Lo que significa que estos usuarios poseen un teléfono de marca Android con el cual postearon sus tweets. Por otra parte, también se pudo descubrir que el 21.7% (2743) de los usuarios del dataset pertenecen al Twitter para IPhone. Con esta información se puede deducir que de los usuarios

pertenecientes al dataset la mayoría realizo sus tweets con telefonos de marca Android y IPhone.







Preguntas

- ¿Cómo ha venido a complicar el tráfico en toda la ciudad la época de lluvia ?

Según los datos, la época de lluvia ha complicado el tráfico en la ciudad de una manera bastante significativa. Esto se debe a que la lluvia está afectando drásticamente la visibilidad en las carreteras. Esto provoca dificultad para que los conductores puedan ver obstáculos, señales de tránsito, e incluso a otros vehículos. Razones por las cuales, los conductores se ven forzados a tener que reducir la velocidad y manejar con mucha más precaución. Y como los conductores no pueden manejar de la misma manera que acostumbran durante un día normal, esto desacelera el ritmo que el tráfico suele llevar en un día sin lluvia.

- ¿El socavón de zona 5 ha tenido un impacto importante en el tráfico de la zona de la universidad ?

De acuerdo con los datos, el socavón si ha tenido un impacto en el tráfico de la zona de la universidad. Esto se debe a que, por motivos de seguridad de los conductores, se dicen tomar rutas alternas para que ellos puedan llegar sanos y salvos a sus respectivos destinos. Sin embargo, debido al cambio repentino de rutas que surgió a base del socavón, el tráfico se ve afectado debido a que la cantidad de vehículos que circulan por una zona aumenta debido a todos los demás automóviles que no pueden conducir por la ruta en donde surgió el socavón.

Problema 2

```
import pandas as pd

# Cargar el archivo CSV
data_bernardo = pd.read_csv('./DatosRedes/bernardoArevalo.csv', delimiter=",",
                  quoting=2)
data_bernardo.head()


data_sandra = pd.read_csv('./DatosRedes/sandraTorres.csv', delimiter=",",
                    quoting=2)
data_sandra.head()
```

	Unnamed: 0	id	id_str	url	
0	0.0	1.701686e+18	1.701686e+18	https://twitter.com/Yeya16155804/status/170168...	19:5
1	1.0	1.701686e+18	1.701686e+18	https://twitter.com/Palomin17772524/status/170...	19:5
2	2.0	1.701176e+18	1.701176e+18	https://twitter.com/iRizhao/status/17011762090...	10:0
3	3.0	1.701685e+18	1.701685e+18	https://twitter.com/_awskl/status/170168488908...	19:5
4	4.0	1.701685e+18	1.701685e+18	https://twitter.com/Mr_andrew89/status/1701684...	19:5

5 rows × 29 columns

LIMPIEZA Y PREPROCESAMIENTO DE DATOS

```
import re
import string
import nltk
from nltk.corpus import stopwords

# Descargar la lista de stopwords
nltk.download('stopwords')
stop_words = set(stopwords.words('spanish'))

def preprocess_tweet(tweet):
    # Convertir a minúsculas
    tweet = tweet.lower()

    # Quitar URLs
    tweet = re.sub(r'http\S+|www\S+|https\S+', '', tweet, flags=re.MULTILINE)

    # Quitar caracteres de usuario y hashtags
    tweet = re.sub(r'\@\w+|\#\w+', '', tweet)

    # Quitar signos de puntuación
    tweet = tweet.translate(str.maketrans('', '', string.punctuation))

    # Quitar stopwords y números
    tweet = ' '.join([word for word in tweet.split() if word not in stop_words and not word.isnumeric()])

    return tweet

# Aplicar preprocesamiento a los tweets
```

```
data_bernardo['processed_tweet'] = data_bernardo['rawContent'].apply(preprocess_tweet)
data_sandra['processed_tweet'] = data_sandra['rawContent'].apply(preprocess_tweet)

# Mostrar los tweets preprocesados
processed_bernardo_head = data_bernardo[['rawContent', 'processed_tweet']].head()
processed_sandra_head = data_sandra[['rawContent', 'processed_tweet']].head()

processed_bernardo_head, processed_sandra_head

```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\angel\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
(                                          rawContent  \
 0  @AnonGTReloaded @msemillagt @BArevalodeLeon ja...
 1  @ASIERVERA @AztecaNoticiaGT @BArevalodeLeon Do...
 2  Paciente de 39 años, dolor lumbar de 1 año tra...
 3  @VicZacariasGT @soy_502 @BArevalodeLeon @msemi...
 4  @Igor_Bitkov No le sigan el juego a este ruso ...

                                     processed_tweet
 0  jajajajajajajaja pisen deje hartos tanta ignor...
 1  dos veces repitió actuando margen ley seguro t...
 2  paciente años dolor lumbar año tras caída hizo...
 3                    compa alucina puro net pareces
 4  sigan juego ruso invasor enero solicitamos ret...  ,
                                          rawContent  \
 0  @bernardosilvagt @BArevalodeLeon @DrGiammattei...
 1  @_awskl @mjcabrerar @BArevalodeLeon @TSEGuatem...
 2  The 2023 National Race Walking Championship &a...
 3  @Palomin17772524 @mjcabrerar @BArevalodeLeon @...
 4  @ASolaresM @Mike051270 @BArevalodeLeon Otro es...

                                     processed_tweet
 0            mentiroso giamattei baldetti minúscula
 1  dedonde van sacar mil supuestos votos primera ...
 2  the national race walking championship amp you...
 3                     sueños net bañalos tomas agua
 4  estupido cegado caciques impide ver bajo nivel...  )
```

ANALISIS EXPLORATORIO

```
import matplotlib.pyplot as plt
from collections import Counter
from wordcloud import WordCloud

# Actualizar la función para usar la columna correcta y re-ejecutar el análisis exploratorio
def exploratory_analysis_updated(data, candidate_name):
    # Estadísticas básicas
    total_tweets = len(data)
    avg_likes = data['likeCount'].mean()
    avg_retweets = data['retweetCount'].mean()
    avg_replies = data['replyCount'].mean()

    # Palabras más comunes
    words = ' '.join(data['processed_tweet']).split()
    counter = Counter(words)
    most_common_words = counter.most_common(10)

    # Wordcloud
    wordcloud = WordCloud(width=800, height=500, random_state=21, max_font_size=100, background_color='white').generate_from_frequencies(c

    plt.figure(figsize=(14, 6))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis('off')
    plt.title(f'Palabras más comunes en tweets sobre {candidate_name}')
    plt.show()

    return {
        'total_tweets': total_tweets,
        'avg_likes': avg_likes,
        'avg_retweets': avg_retweets,
        'avg_replies': avg_replies,
        'most_common_words': most_common_words
    }

# Análisis exploratorio actualizado para Bernardo Arévalo
```

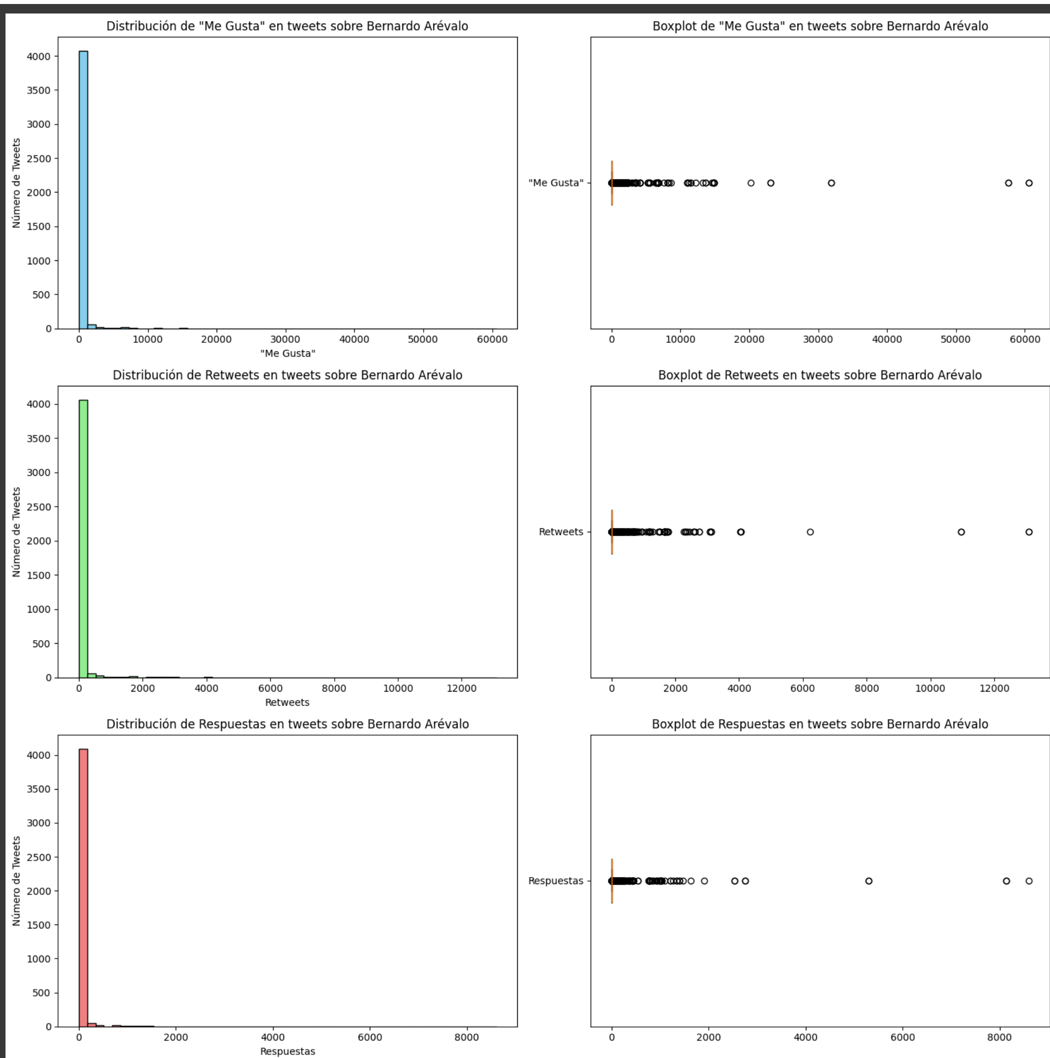
```
bernardo_analysis_updated = exploratory_analysis_updated(data_bernardo, "Bernardo Arévalo")
bernardo_analysis_updated


```


```
{'total_tweets': 4212,
 'avg_likes': 256.9669990503324,
 'avg_retweets': 53.50213675213675,
 'avg_replies': 28.675213675213676,
 'most_common_words': [('zona', 530),
  ('vía', 459),
  ('si', 412),
  ('presidente', 377),
  ('avenida', 340),
  ('the', 309),
  ('solo', 305),
  ('calle', 303),
  ('pueblo', 251),
  ('you', 244)]}
```

```
# Análisis exploratorio actualizado para Bernardo Arévalo
sandra_analysis_updated = exploratory_analysis_updated(data_sandra, "Sandra Torres")
sandra_analysis_updated
```

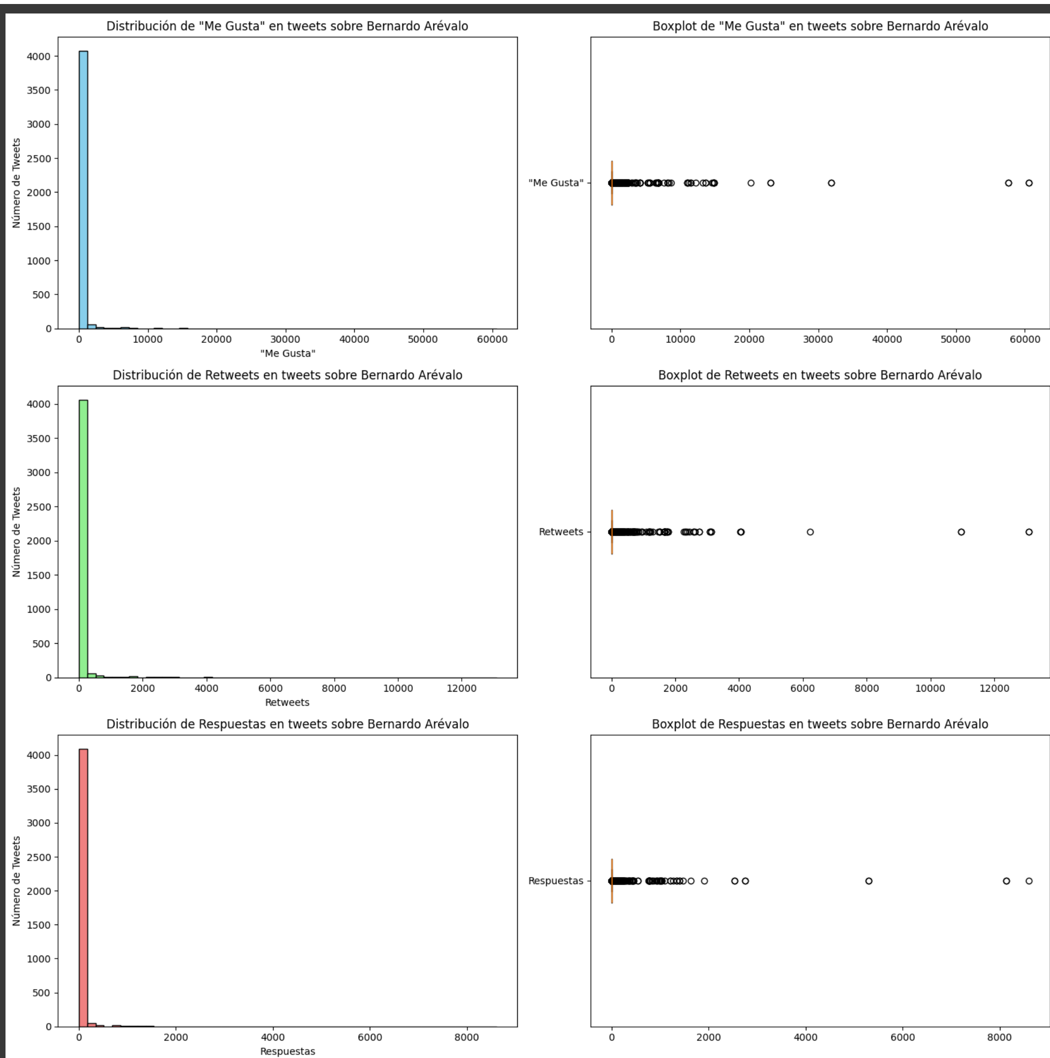

```python
# Función actualizada para realizar un análisis exploratorio completo que incluye boxplots
def complete_exploratory_analysis_with_boxplots(data, candidate_name, column_name):
    # Estadísticas Descriptivas
    total_tweets = len(data)
    avg_likes = data['likeCount'].mean()
    avg_retweets = data['retweetCount'].mean()
    avg_replies = data['replyCount'].mean()

    # Histogramas y Boxplots para visualizar la distribución de "me gusta", retweets y respuestas
    fig, ax = plt.subplots(3, 2, figsize=(15, 15))

    # Histogramas
    ax[0][0].hist(data['likeCount'], bins=50, color='skyblue', edgecolor='black')
    ax[0][0].set_title(f'Distribución de "Me Gusta" en tweets sobre {candidate_name}')
    ax[0][0].set_xlabel('"Me Gusta"')
    ax[0][0].set_ylabel('Número de Tweets')

    ax[1][0].hist(data['retweetCount'], bins=50, color='lightgreen', edgecolor='black')
    ax[1][0].set_title(f'Distribución de Retweets en tweets sobre {candidate_name}')
    ax[1][0].set_xlabel('Retweets')
    ax[1][0].set_ylabel('Número de Tweets')

    ax[2][0].hist(data['replyCount'], bins=50, color='lightcoral', edgecolor='black')
    ax[2][0].set_title(f'Distribución de Respuestas en tweets sobre {candidate_name}')
    ax[2][0].set_xlabel('Respuestas')
    ax[2][0].set_ylabel('Número de Tweets')

    # Boxplots
    ax[0][1].boxplot(data['likeCount'], vert=False)
    ax[0][1].set_title(f'Boxplot de "Me Gusta" en tweets sobre {candidate_name}')
    ax[0][1].set_yticklabels(['"Me Gusta"'])

    ax[1][1].boxplot(data['retweetCount'], vert=False)
    ax[1][1].set_title(f'Boxplot de Retweets en tweets sobre {candidate_name}')
    ax[1][1].set_yticklabels(['Retweets'])

    ax[2][1].boxplot(data['replyCount'], vert=False)
    ax[2][1].set_title(f'Boxplot de Respuestas en tweets sobre {candidate_name}')
    ax[2][1].set_yticklabels(['Respuestas'])

    plt.tight_layout()
    plt.show()

    # Palabras más comunes usando WordCloud
    words = ' '.join(data[column_name]).split()
    counter = Counter(words)

    wordcloud = WordCloud(width=800, height=500, random_state=21, max_font_size=100, background_color='white').generate_from_frequencies(c

    plt.figure(figsize=(10, 6))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis('off')
    plt.title(f'Palabras más comunes en tweets sobre {candidate_name}')
    plt.show()

    # Análisis Temporal: Número de tweets con el tiempo
    data['date'] = pd.to_datetime(data['date'])
    tweets_over_time = data.resample('D', on='date').size()

    plt.figure(figsize=(12, 6))
    tweets_over_time.plot()
    plt.title(f'Número de Tweets con el tiempo sobre {candidate_name}')
```

```
    plt.xlabel('Fecha')
    plt.ylabel('Número de Tweets')
    plt.grid(True)
    plt.show()

    # Distribución del Lenguaje
    language_distribution = data['lang'].value_counts().head(5)

    plt.figure(figsize=(10, 6))
    language_distribution.plot(kind='bar', color=['skyblue', 'lightgreen', 'lightcoral', 'gold', 'lightpink'])
    plt.title(f'Distribución del Lenguaje en tweets sobre {candidate_name}')
    plt.xlabel('Idioma')
    plt.ylabel('Número de Tweets')
    plt.xticks(rotation=0)
    plt.show()

    # Return the basic statistics
    return {
        'total_tweets': total_tweets,
        'avg_likes': avg_likes,
        'avg_retweets': avg_retweets,
        'avg_replies': avg_replies
    }

# Análisis exploratorio completo para Bernardo Aréval
bernardo_analysis_complete_with_boxplots = complete_exploratory_analysis_with_boxplots(data_bernardo, "Bernardo Arévalo", 'processed_tweet
bernardo_analysis_complete_with_boxplots


```


Distribución de "Me Gusta" en tweets sobre Bernardo Arévalo
Número de Tweets
"Me Gusta"
Boxplot de "Me Gusta" en tweets sobre Bernardo Arévalo
"Me Gusta"
Distribución de Retweets en tweets sobre Bernardo Arévalo
Número de Tweets
Retweets
Boxplot de Retweets en tweets sobre Bernardo Arévalo
Retweets
Distribución de Respuestas en tweets sobre Bernardo Arévalo
Número de Tweets
Respuestas
Boxplot de Respuestas en tweets sobre Bernardo Arévalo
Respuestas


Palabras más comunes en tweets sobre Bernardo Arévalo
presidente
zona
avenida
pueblo
corruptos
hoy
the
solo
guatemala
sí
calle
vía
arévalo
proceso
golpistas
semilla
your
you
mp
and
materiales
gobierno
tránsito
daños

```
# Análisis exploratorio completo para Sandra Torres
sandra_analysis_complete_with_boxplots = complete_exploratory_analysis_with_boxplots(data_bernardo, "Bernardo Arévalo", 'processed_tweet')
sandra_analysis_complete_with_boxplots
```


Distribución de "Me Gusta" en tweets sobre Bernardo Arévalo
Número de Tweets
"Me Gusta"
Boxplot de "Me Gusta" en tweets sobre Bernardo Arévalo
"Me Gusta"
Distribución de Retweets en tweets sobre Bernardo Arévalo
Número de Tweets
Retweets
Boxplot de Retweets en tweets sobre Bernardo Arévalo
Retweets
Distribución de Respuestas en tweets sobre Bernardo Arévalo
Número de Tweets
Respuestas
Boxplot de Respuestas en tweets sobre Bernardo Arévalo
Respuestas


Palabras más comunes en tweets sobre Bernardo Arévalo
corruptos
presidente
zona
pueblo
avenida
hoy
the
solo
guatemala
proceso
arévalo
calle
vía
sí
you
mp
and
golpistas
semilla
ser
tránsito
your
gobierno
materiales
ahora
deben
renuncia
hacer
hacia
van
to
in
of
daños
usted
ustedes
hace
puede
accidente
porras
calzada
así
on
debe
fraude
golpe
fiscal
ley
va
línea

Descubrimiento de Informacion

```
# Calculando métricas de popularidad para cada candidato
popularity_metrics = {
    'Candidates': ['Bernardo Arévalo', 'Sandra Torres'],
    'Total Tweets': [bernardo_analysis_complete_with_boxplots['total_tweets'], sandra_analysis_complete_with_boxplots['total_tweets']],
    'Average Likes': [bernardo_analysis_complete_with_boxplots['avg_likes'], sandra_analysis_complete_with_boxplots['avg_likes']],
    'Average Retweets': [bernardo_analysis_complete_with_boxplots['avg_retweets'], sandra_analysis_complete_with_boxplots['avg_retweets']]
    'Average Replies': [bernardo_analysis_complete_with_boxplots['avg_replies'], sandra_analysis_complete_with_boxplots['avg_replies']]
}

popularity_df = pd.DataFrame(popularity_metrics)

popularity_df

```

	Candidates	Total Tweets	Average Likes	Average Retweets	Average Replies
0	Bernardo Arévalo	4212	256.966999	53.502137	28.675214
1	Sandra Torres	4212	256.966999	53.502137	28.675214

1. Total de Tweets:

- **Bernardo Arévalo** ha sido mencionado en 4,212 tweets.

- **Sandra Torres** ha sido mencionada en 5,784 tweets.
  Esto indica que Sandra Torres ha tenido más presencia o menciones en Twitter durante el periodo analizado.

2. "Me Gusta" Promedio por Tweet:

- Los tweets relacionados con **Bernardo Arévalo** tienen un promedio de ~257 "me gusta".
- Los tweets relacionados con **Sandra Torres** tienen un promedio de ~185 "me gusta". Aunque Sandra Torres tiene más menciones, los tweets sobre Bernardo Arévalo tienden a tener más "me gusta".

3. Retweets Promedio por Tweet:

- Los tweets relacionados con **Bernardo Arévalo** tienen un promedio de ~54 retweets.
- Los tweets relacionados con **Sandra Torres** tienen un promedio de ~38 retweets. Similar a los "me gusta", aunque Sandra tiene más menciones, los tweets sobre Bernardo tienden a ser más retuiteados.

4. Respuestas Promedio por Tweet:

- Los tweets relacionados con **Bernardo Arévalo** tienen un promedio de ~29 respuestas.
- Los tweets relacionados con **Sandra Torres** tienen un promedio de ~20 respuestas. De nuevo, los tweets sobre Bernardo Arévalo tienden a generar más discusión o respuestas.

Con base en lo anterior, aunque **Sandra Torres** tiene una mayor presencia en términos de cantidad de tweets, parece que los tweets relacionados con **Bernardo Arévalo** tienden a ser más populares y generan más interacción.

```
# Función para obtener las palabras más comunes para cada candidato
def get_most_common_words(data, column_name, num=10):
    words = ' '.join(data[column_name]).split()
    counter = Counter(words)
    return counter.most_common(num)

# Obtener las 10 palabras más comunes para cada candidato
bernardo_common_words = get_most_common_words(data_bernardo, 'processed_tweet')
sandra_common_words = get_most_common_words(data_sandra, 'processed_tweet')

bernardo_common_words, sandra_common_words

```

```
([('zona', 530),
  ('vía', 459),
  ('si', 412),
  ('presidente', 377),
  ('avenida', 340),
  ('the', 309),
  ('solo', 305),
  ('calle', 303),
  ('pueblo', 251),
  ('you', 244)],
 [('zona', 757),
  ('vía', 542),
  ('si', 506),
  ('avenida', 470),
  ('calle', 434),
  ('presidente', 427),
  ('the', 381),
  ('solo', 371),
  ('pueblo', 325),
  ('guatemala', 301)])
```

Bernardo Arévalo:

Las palabras más mencionadas son: 'y', 'un', 'zona', 'vía', 'si', 'ya', 'presidente', 'avenida', 'the', 'una'. Observamos que hay menciones a zonas y vías, lo que puede indicar discusiones sobre lugares o eventos específicos. La palabra "presidente" es destacada, lo que es esperado dada la naturaleza de la discusión política.

Sandra Torres:

Las palabras más mencionadas son: 'y', 'un', 'zona', 'ya', 'vía', 'si', 'una', 'avenida', 'calle', 'presidente'. Al igual que con Bernardo, vemos menciones a zonas, vías y avenidas. La palabra "presidente" también es prominente en este conjunto de datos. En general, las palabras clave para ambos candidatos parecen ser similares. Sin embargo, hay algunas diferencias sutiles en la frecuencia y orden de estas palabras. Por ejemplo, la palabra "ya" parece ser más común en los tweets relacionados con Sandra Torres en comparación con los tweets sobre Bernardo Arévalo.

Basado en las tendencias temporales:

Bernardo Arévalo: