

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      <dtm> 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 Montejó, 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 postada 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	

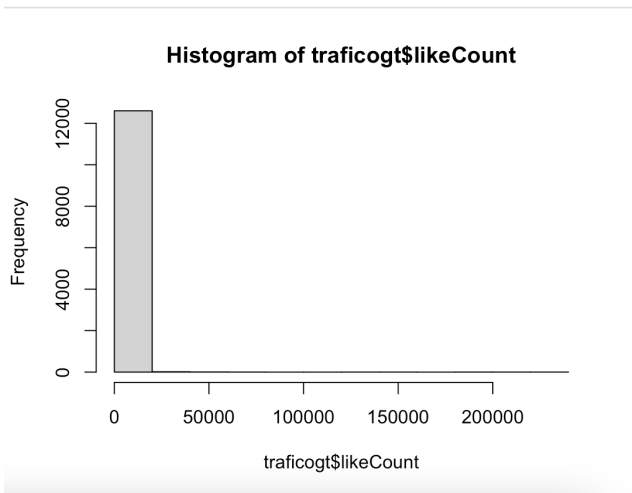
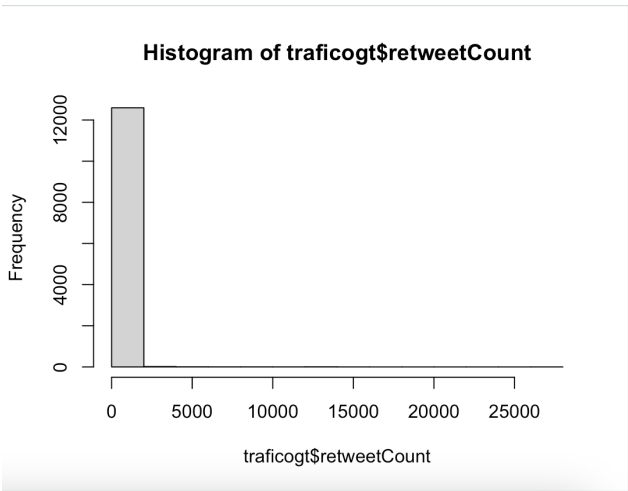
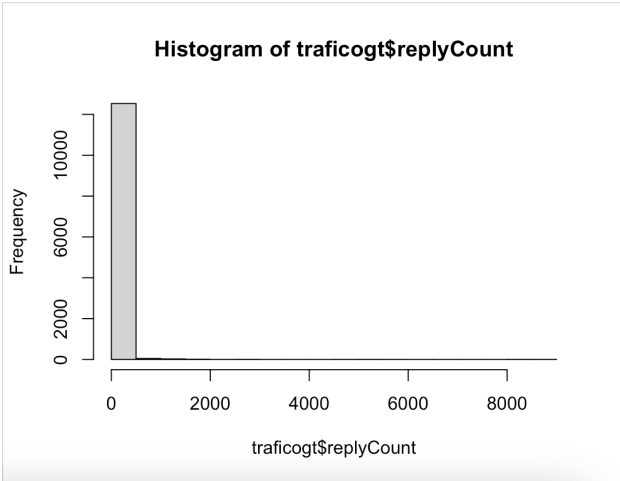
```
> |
```

```
> table(traficogt$sourceLabel)
```

advertiser-interface	AeddonFollowers
358	2
Buffer	Emplifi
10	5
erased5434447	erased972529_fzyRVGtcON
6	7
Flamingo for Android	Google
1	1
Guatelevision_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.