Analysis of Women's March Tweets data

The Women's March was a worldwide protest on January 21, 2017, to advocate legislation and policies regarding human rights and other issues, including women's rights, immigration reform, healthcare reform, reproductive rights, the natural environment, LGBTQ rights, racial equality, freedom of religion and workers' rights. The tweets of this protest are analysed in this project.

In [1]:

```
import pandas as pd
import numpy as np
import re
import string
```

In [2]:

```
Filename
='/Users/sivaut/Documents/ramawork/Womensmarch/womensmarchtweets.csv'

tweets_df = pd.read_csv(Filename)
```

In [3]:

```
tweets_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3100 entries, 0 to 3099
Data columns (total 5 columns):
id             3100 non-null float64
text           3100 non-null object
source         3100 non-null object
created_at     3100 non-null object
place          308 non-null object
dtypes: float64(1), object(4)
memory usage: 121.2+ KB
```

The Women's March tweets data has 3100 entries with 5 columns. There are no missing values in 4 columns whereas the column named "place" has 308 known values(or 2792 missing values).

In [4]:

```
tweets_df.head()
```

Out[4]:

	id	text	source	created_at	place
0	8.230000e+17	#WomensMarch Can someone send me a pussy-hat ?...	Twitter for iPhone	1/21/17 23:59	NaN
1	8.230000e+17	_ôëŸo melhor protesto da histí_ria que voce ad...	Twitter for iPhone	1/21/17 23:59	NaN
2	8.230000e+17	I thank God that we have each other! #WomensMa...	Twitter for iPhone	1/21/17 23:59	NaN

	id	text	source	created_at	place
3	8.230000e+17	#WomensMarch https://t.co/otODcBgLUO	Twitter for iPhone	1/21/17 23:59	NaN
4	8.230000e+17	#WomensMarch #imwithher https://t.co/wDspd2eN3p	Twitter for Android	1/21/17 23:59	NaN

The first five rows of the dataframe are shown. The column "place" has missing values for all the 5 rows.

In [5]:

```
tweets_df.dropna().head()
```

Out[5]:

	id	text	source	created_at	place
7	8.230000e+17	Reppin' for all the badass babes out there in ...	Instagram	1/21/17 23:59	Seattle, WA
10	8.230000e+17	äĺš #cepostaperte\r\näĺŠ #MilanNapoli\r\näĺ_ #...	Tendenze Italia	1/21/17 23:59	Rome, Lazio
17	8.230000e+17	äĺš #WomensMarch\r\näĺŠ Juan Pablo Montoya\r\n...	Es Tendencia en Colombia	1/21/17 23:59	Bogot\xe1, D.C., Colombia
19	8.230000e+17	Have no idea what to expect but the #WomensMar...	Twitter for iPhone	1/21/17 23:59	Queens, NY
27	8.230000e+17	äĺš #WomensMarch\r\näĺŠ VICICONTE SOS ĺ_NICA E...	Es Tendencia en Argentina	1/21/17 23:59	Ciudad Aut\xf3noma de Buenos Aires, Argentina

The first five rows having non missing values are shown.

Analysis by id

In [6]:

```
tweets_df.groupby('id').size().reset_index(name='count').sort_values(['count'], ascending=False)
```

Out[6]:

	id	count
0	8.230000e+17	3100

All the entries have the same id. The ids aren't unique. Also, the id looks unusual.

Analysis by source

In [7]:

```
tweets_df.groupby('source').size().reset_index(name='count').sort_values(['
```

```
tweets_df.groupby('source').size().reset_index(name='count').sort_values(['
count'], ascending=False)
```

Out[7]:

	source	count
44	Twitter for iPhone	1521
38	Twitter for Android	596
37	Twitter Web Client	460
15	Instagram	182
43	Twitter for iPad	79
19	Mobile Web (M5)	38
10	Facebook	28
32	TweetDeck	28
14	IFTTT	27
23	Put your button on any page!	26
20	Paper.li	17
34	Tweetbot for i‘ôS	12
42	Twitter for Windows Phone	10
48	config001 via...	10
13	Hootsuite	6
3	Echofon	6
31	TweetCaster for Android	6
41	Twitter for Windows	4
22	PlumeŒæforŒæAndroid	4
24	SocialFlow	4
18	Mobile Web (M2)	3
0	Buffer	2
21	Periscope	2
39	Twitter for BlackBerry	2
45	UberSocial Professional	1
46	Untappd	1
47	WordPress.com	1
36	Twitter Dashboard for iPhone	1
35	Tweetlogix	1
49	iOS	1
33	Tweetbot for Mac	1
40	Twitter for BlackBerryŒ¬	1

	source	count
25	Talon (Plus)	1
30	Trends from India	1
29	Trendinalia Mí©xico	1
2	Deutschland in Trends	1
4	Es Tendencia en Argentina	1
5	Es Tendencia en Chile	1
6	Es Tendencia en Colombia	1
7	Es Tendencia en Dominicana	1
8	Es Tendencia en Espaí±a	1
9	Es Tendencia en Venezuela	1
11	Fenix for Android	1
12	Google	1
16	Medium	1
17	Meet Edgar	1
1	Dadaist robot	1
26	Tendenze Italia	1
27	TendíÈncias Brasil	1
28	Test This Again	1
50	twicca	1

51 different sources were used to post the tweets. Most people used "Twitter for iphone" as their source, followed by "Twitter for Android" and "Twitter webclient". "Test This Again" and "Put your button on any page" are also found as sources in the data.

Analysis by created_at

In [8]:

```
tweets_df.groupby('created_at').size().reset_index(name='count').sort_value
s(['count'], ascending=False)
```

Out[8]:

	created_at	count
3	1/21/17 23:59	906
1	1/21/17 23:57	900
2	1/21/17 23:58	833
0	1/21/17 23:56	461

In [9]:

```
tweets_df.dropna().head()
```

```
tweets_df.dropna().head()
```

Out[9]:

	id	text	source	created_at	place
7	8.230000e+17	Reppin' for all the badass babes out there in ...	Instagram	1/21/17 23:59	Seattle, WA
10	8.230000e+17	äĺš #cepostaperte\r\näĺŠ #MilanNapoli\r\näĺ_ #...	Tendenze Italia	1/21/17 23:59	Rome, Lazio
17	8.230000e+17	äĺš #WomensMarch\r\näĺŠ Juan Pablo Montoya\r\n...	Es Tendencia en Colombia	1/21/17 23:59	Bogot\xe1, D.C., Colombia
19	8.230000e+17	Have no idea what to expect but the #WomensMar...	Twitter for iPhone	1/21/17 23:59	Queens, NY
27	8.230000e+17	äĺš #WomensMarch\r\näĺŠ VICICONTE SOS í_NICA E...	Es Tendencia en Argentina	1/21/17 23:59	Ciudad Aut\xf3noma de Buenos Aires, Argentina

The tweets were created in different countries having different timezones and in different languages. Surprisingly, all the tweets were created between the time 23:56 and 23:59 on January 21, 2017. Most likely, the tweets data were taken for the day January 21, 2017 hence it stops at 23:59. This tweets data might be a slice of a big data.

Analysis by place

In [10]:

```
File ='/Users/sivaut/Documents/ramawork/Womensmarch/state.csv'

state_df = pd.read_csv(File)
```

In [11]:

```
state_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 2 columns):
State            51 non-null object
Abbreviation     51 non-null object
dtypes: object(2)
memory usage: 896.0+ bytes
```

In [12]:

```
state_df.head()
```

Out[12]:

	State	Abbreviation
0	Alabama	AL
1	Alaska	AK
2	Arizona	AZ

	State	Abbreviation
3	Arkansas	AR
4	California	CA

The Women's March happened in various countries. The tweets include both abbreviated and full name of the US states. The state and its abbreviation are read from the csv file and stored in the state_df dataframe, to process the data.

The column named "place" holds the location from where the tweets are published. It has 308 known values(or 2792 missing values). The analysis was carried out with the known values.It was split into "city" and "State" to facilitate the analysis.

In [13]:

```
mydict = dict(zip(state_df.State, state_df.Abbreviation))
```

In [14]:

```
list_state = state_df.values.T.tolist()
```

In [15]:

```
tweets_df['city'], tweets_df['State'] = tweets_df['place'].str.split(',', 1
).str
```

In [16]:

```
tweets_df.dropna().tail()
```

Out[16]:

	id	text	source	created_at	place	city	State
3049	8.230000e+17	#math + #solidarity = #womensmarch https://t.c...	Twitter for Android	1/21/17 23:56	Manhattan, NY	Manhattan	NY
3060	8.230000e+17	I haven't seen the city this energized about s...	Twitter for iPhone	1/21/17 23:56	San Francisco, CA	San Francisco	CA
3077	8.230000e+17	#nofilter #missingobama #nevertrump #armyoflov...	Instagram	1/21/17 23:56	East Potomac Park, Washington	East Potomac Park	Washington
3089	8.230000e+17	@BitterSaltiness I actually went to the #Women...	Twitter for iPhone	1/21/17 23:56	Wisconsin, USA	Wisconsin	USA
3091	8.230000e+17	Women are such beautiful and wonderful creatur...	Twitter for iPhone	1/21/17 23:56	Kansas, USA	Kansas	USA

In the above output, the non missing values of the column named "State" is either a country name or an abbreviated/full names of an US state.

In [17]:

```
tweets_df['State']= tweets_df['State'].str.strip()
```

In [18]:

```
tweets_df['State'] = np.where(tweets_df['State']=='USA' ,tweets_df['city'],
tweets_df['State'])
```

In [19]:

```
tweets_df['city'] = np.where(tweets_df['State'].isin(list_state[0]), tweets
_df['State'], tweets_df['city'])
```

In [20]:

```
tweets_df['country'] = np.where(tweets_df['State'].isin(list_state[1]) , 'U
SA', tweets_df['State'])
```

In [21]:

```
tweets_df['country'] = np.where(tweets_df['State'].isin(list_state[0]), 'US
A', tweets_df['country'])
```

In [22]:

```
tweets_df['Abbreviated'] = tweets_df['State']
```

In [23]:

```
tweets_df['State'] = np.where(tweets_df['State'].isin(list_state[0]), 'Abb'
, tweets_df['State'])
```

In [24]:

```
tweets_df['State'] = np.where(tweets_df['State']=='USA' ,tweets_df['city'],
tweets_df['State'])
```

In [25]:

```
tweets_df['Abbreviated'] = np.where(tweets_df['State']=='Abb' ,tweets_df['c
ity'].map(mydict), 'NaN')
```

In [26]:

```
tweets_df['State'] = np.where(tweets_df['State']=='Abb' ,tweets_df['Abbrevi
ated'], tweets_df['State'])
```

In [27]:

```
del tweets_df['Abbreviated']
```

In [28]:

```
tweets_df.dropna().head()
```

Out[28]:

	id	text	source	created_at	place	city	Stat
7	8.230000e+17	Reppin' for all the badass babes out there in ...	Instagram	1/21/17 23:59	Seattle, WA	Seattle	WA
10	8.230000e+17	äíš #cepostaperte\r\näÍŠ #MilanNapoli\r\näÍ_ #...	Tendenze Italia	1/21/17 23:59	Rome, Lazio	Rome	Lazi
17	8.230000e+17	äíš #WomensMarch\r\näÍŠ Juan Pablo Montoya\r\n...	Es Tendencia en Colombia	1/21/17 23:59	Bogot\xe1, D.C., Colombia	Bogot\xe1	D.C. Colo
19	8.230000e+17	Have no idea what to expect but the #WomensMar...	Twitter for iPhone	1/21/17 23:59	Queens, NY	Queens	NY
27	8.230000e+17	äíš #WomensMarch\r\näÍŠ VICICONTE SOS í_NICA E...	Es Tendencia en Argentina	1/21/17 23:59	Ciudad Aut\xf3noma de Buenos Aires, Argentina	Ciudad Aut\xf3noma de Buenos Aires	Arge

The dataframe was processed in detail to represent all the US states in an abbreviated form in the column named "State" and USA as their country in the column named "country". The Non-USA locations are unchanged. This facilitates the analysis by country and states within USA.

Analysis by Country

In [29]:

```
tweets_df.groupby('country').size().reset_index(name='count').sort_values([
'count'], ascending=False)
```

Out[29]:

	country	count
21	USA	249
17	Ontario	5
3	Brasil	4
4	Brazil	4
8	Distrito Federal	3
0	Argentina	2
27	Wales	2

	country	count
27	Wales	2
10	England	2
15	London	2
2	Belgium	2
5	British Columbia	2
7	D.C., Colombia	2
9	Dominican Republic	1
20	Spain	1
28	Western Australia	1
26	Victoria	1
25	Venezuela	1
24	Veneto	1
23	Uruguay	1
22	United Kingdom	1
19	South Africa	1
18	Portugal	1
6	Chile	1
16	Nederland	1
1	Austria	1
14	Lazio	1
13	Ireland	1
12	India	1
11	Germany	1
29	attributes={}, id=u	1

The tweets came from 30 different countries. From the known location tweets, it is found that the major tweets came from USA. In Portugese, "Brazil" is represented as "Brasil". Erroneous value like "attributes={}, id=u" is also found in the data.

Analysis by US states

In [30]:

```
tweets_df.loc[tweets_df['country'] == 'USA'].groupby('State').size().reset_
index(name='count').sort_values(['count'], ascending=False)
```

Out[30]:

	State	count
4	CA	57
7	DC	35

	State	count
7	DC	[illegible]
27	NY	24
35	TX	18
39	WA	16
30	OR	14
9	FL	10
37	VA	9
18	MI	6
10	GA	5
22	NC	4
12	IL	4
31	PA	4
40	WI	3
16	MA	3
28	OH	3
13	KS	3
17	MD	2
29	OK	2
19	MN	2
24	NJ	2
25	NM	2
26	NV	2
14	KY	2
33	SC	1
32	RI	1
34	TN	1
36	UT	1
38	VT	1
0	AK	1
23	NH	1
21	MT	1
1	AL	1
15	LA	1
11	HI	1
8	DE	1
6	CT	1

	State	count
5	CO	1
3	AZ	1
2	AR	1
20	MO	1

There were tweets from 41 US states. The state CA(California) has the major tweets, followed by DC(The District of Columbia) and NY(Newyork).

Analysis by Text

In [31]:

```
tweets_df['text']= tweets_df['text'].astype(str)
tweets_df['text'].head()
```

Out[31]:

```
0     #WomensMarch Can someone send me a pussy-hat ?...
1     _ôëŸo melhor protesto da histí_ria que voce ad...
2     I thank God that we have each other! #WomensMa...
3                  #WomensMarch https://t.co/otODcBgLUO
4       #WomensMarch #imwithher https://t.co/wDspd2eN3p
Name: text, dtype: object
```

There are 3100 tweets in the data. The tweets include hashtags, text messages, links, tags and emojis or some symbols. The tweets are in different languages.

In [32]:

```
hash_list = []
def splithash(s):
    mylist = [i  for i in s.split() if i.startswith("#") ]

    hash_list.extend(mylist)
    return ', '.join(mylist)
tweets_df['Hashtags'] =  tweets_df.text.apply(splithash)
```

In [33]:

```
def splitlink(s):
    mylink = [i  for i in s.split() if i.startswith("http") ]

    return ', '.join(mylink)
tweets_df['Links'] =  tweets_df.text.apply(splitlink)
```

In [34]:

```
Tags = []
def splittags(s):
    mytags = [i  for i in s.split() if i.startswith("@") ]
    Tags.extend(mytags)
    return ', '.join(mytags)
tweets_df['Tags'] = tweets_df.text.apply(splittags)
```

In [35]:

```
tweets_df.dropna().tail()
```

Out[35]:

	id	text	source	created_at	place	city	State	cou
3049	8.230000e+17	#math + #solidarity = #womensmarch https://t.c...	Twitter for Android	1/21/17 23:56	Manhattan, NY	Manhattan	NY	US/
3060	8.230000e+17	I haven't seen the city this energized about s...	Twitter for iPhone	1/21/17 23:56	San Francisco, CA	San Francisco	CA	US/
3077	8.230000e+17	#nofilter #missingobama #nevertrump #armyoflov...	Instagram	1/21/17 23:56	East Potomac Park, Washington	Washington	WA	US/
3089	8.230000e+17	@BitterSaltiness I actually went to the #Women...	Twitter for iPhone	1/21/17 23:56	Wisconsin, USA	Wisconsin	WI	US/
3091	8.230000e+17	Women are such beautiful and wonderful creatur...	Twitter for iPhone	1/21/17 23:56	Kansas, USA	Kansas	KS	US/

To analyse the tweets, the hashtags, links and tags were separated.

Analysis by Tags

In [36]:

```
tweets_df.groupby('Tags').size().reset_index(name='count').sort_values(['co
unt'], ascending=False).head(8)
```

Out[36]:

	Tags	count
0		2497
1	@	66
233	@c0nvey	21
388	@womensmarch	20
334	@realDonaldTrump	16
208	@WhiteHouse, @realDonaldTrump, @POTUS	16
140	@POTUS	13

353	@seanspicer	8
	Tags	count

Analysis by Tags was carried out to find the most used tags. 399 rows were found. The most tweets doesn't have a tag. some rows have more than one tags. The tags in each row are split and counted separately.

In [37]:

```
tags_df = pd.DataFrame({'Tags': Tags})
```

In [38]:

```
tags_df.groupby('Tags').size().reset_index(name='count').sort_values(['coun
t'], ascending=False).head(7)
```

Out[38]:

	Tags	count
0	@	74
470	@realDonaldTrump	48
192	@POTUS	40
525	@womensmarch	30
324	@c0nvey	23
489	@seanspicer	19
283	@WhiteHouse	17

There were 532 different tags found. Top 7 tags were shown here.

Analysis by Links

In [39]:

```
tweets_df.groupby('Links').size().reset_index(name='count').sort_values(['c
ount'], ascending=False).head(7)
```

Out[39]:

	Links	count
0		1295
285	https://t.co/9By5oAmkSG	2
982	https://t.co/XHAB8WVt5m	2
1533	https://t.co/qwszhpfZq6	2
1431	https://t.co/nPBIl0RuXE	2
851	https://t.co/SqJPYKAIDP	2
918	https://t.co/VE5bib4Zzv	2

Analysis by links was carried out to find the most tweeted links. 1789 rows were found. 1295 tweets did not include any links.

Analysis by hashtags

In [40]:

```
hash_df = pd.DataFrame({'Hashtags_raw': hash_list})
```

In [41]:

```
hash_df.groupby('Hashtags_raw').size().reset_index(name='count').sort_value
s(['count'], ascending=False).head()
```

Out[41]:

	Hashtags_raw	count
507	#WomensMarch	2053
1107	#womensmarch	708
538	#WomensMarchOnWashington	65
515	#WomensMarch.	25
191	#Inauguration	20

There were 1201 hashtags found. The hashtag #WomensMarch was the most used one. some hashtags like #WomensMarch and #WomensMarch. differs only by the punctuation marks. When the punctuation marks were removed, the number of hashtags gets reduced and the count increases.

In [42]:

```
def remove_p(s):
    exclude = set(string.punctuation)
    s = ''.join(ch for ch in s if ch not in exclude)
    return "#"+s
```

In [43]:

```
hash_df['Hashtags_punc_removed'] =  hash_df.Hashtags_raw.apply(remove_p)
```

In [44]:

```
hash_df.groupby('Hashtags_punc_removed').size().reset_index(name='count').s
ort_values(['count'], ascending=False).head(15)
```

Out[44]:

	Hashtags_punc_removed	count
492	#WomensMarch	2114
1074	#womensmarch	723
509	#WomensMarchOnWashington	65
189	#Inauguration	21
420	#Trump	19

	Hashtags_punc_removed	count
538	#Womensmarch	18
1114	#womensmarchonwashington	18
1130	#womensmarchäó	17
229	#LoveTrumpsHate	16
1138	#womensrightsarehumanrights	15
477	#WhyIMarch	15
1052	#whyimarch	14
489	#WomenWhoHaveInspiredMe	14
527	#WomensMarchäó	14
915	#resist	14

When the punctuation was removed, the number of hashtags reduced to 1160 and the count increases in some of the hashtags. 'Womensmarch', 'WomensMarch', 'womensMarch', 'womensmarch' and 'WOMENSMARCH' all refer to the same word if it is not case sensitive.

In [45]:

```
hash_df['Hashtags_lower']= hash_df['Hashtags_punc_removed'].str.lower()
```

In [46]:

```
hash_df.groupby('Hashtags_lower').size().reset_index(name='count').sort_val
ues(['count'], ascending=False).head(10)
```

Out[46]:

	Hashtags_lower	count
918	#womensmarch	2875
966	#womensmarchonwashington	83
886	#whyimarch	42
990	#womensmarchäó	31
805	#trump	26
347	#inauguration	24
641	#resist	22
944	#womensmarchla	22
1003	#womensrights	20
432	#lovetrumpshate	19

The hashtags are converted to lower case and then they are grouped together. Now the number of hastags reduced to 1039. But the count has increased.

Analysis by country and hashtags

In [47]:

```
tweets_df.groupby(['country','Hashtags']).size()
```

Out[47]:

```
country                 Hashtags

Argentina               #WomensMarch
1
                        #WomensMarch, #LaliEnPinamar
1
Austria                 #WomensMarch, #Icantkeepquiet
1
Belgium                 #WomensMarch
1
                        #womensmarch
1
Brasil                  #WomensMarch
3
                        #Womensmarch
1
Brazil                  #WomensMarch
2
                        #WomensMarch, #fcseguindofcs, #cocodegatomelhorquekpop,
#trndnl                                       1
                        #womensmarch
1
British Columbia        #WomensMarch
1
                        #womensmarchonwashington, #yyjwomensmarch, #womensmarch
, #womensmarch2017, #yyj, #yyjwomen        1
Chile                   #WomensMarch, #ColoColoJuegaEnEl13, #bomberosdechile, #
MisterPiernas2017, #FrenteAmplio          1
D.C., Colombia          #WomensMarch, #SUPERLIGAAGUILAxWIN, #trndnl
1
                        #womensmarch, #ungovernable, #antifascist
1
Distrito Federal        #WomensMarch, #ElDineroSeInventí_Para,
#YaEsFinDeSemanaPara, #GayGames2018                      1
                        #americaferrera, #womensmarch
1
                        #womensmarch, #ciudaddemí©xico
1
Dominican Republic      #WomensMarch, #LuceroEnRD2017, #trndnl
1
England                 #TrumpleThinSkin, #womensmarch
1
                        #womensmarch, #womensrightsarehumanrights
1
Germany                 #WomensMarch, #TrashTVTalk, #sportstudio, #trndnl
1
India                   #WomensMarch, #RaeesInDubai, #AAP_ö_æö_„_ö__ö«š, #NewSu
rpriseHKNKJ, #ThinkBIGSundayWithMarsha     1
Ireland                 #WomensMarch, #WomensMarch
1
Lazio                   #cepostaperte, #MilanNapoli, #WomensMarch, #Cavallidiba
ttaglia, #SonoInnocente, #trndnl          1
London                  #WomensMarch
1
```

```
                    #womensmarch
1
Nederland           #WomensMarch
1
Ontario             #WomensMarch
2
                    #WomensMarchäó_
1

..
USA                 #womensmarch, #womensmarchgreenville, #greenvillesc
1
                    #womensmarch, #womensmarchla
1
                    #womensmarch, #womensmarchla, #womensmarchlosangeles, #
whyimarch, #wmla                      1
                    #womensmarch, #womensmarchmiami
1
                    #womensmarch, #womensmarchnyc
1
                    #womensmarch, #womensmarchoakland, #whyimarch, #imarchf
oräó_                                 1
                    #womensmarch, #womensmarchonnyc, #whyimarch
1
                    #womensmarch, #womensmarchonphiladelphia, #nastywoman,
#philly, #feminist, #beyoncí©         1
                    #womensmarch, #womensmarchonwashington, #antitrump, #ri
seandresist, #rar                     1
                    #womensmarch, #womensmarchonwashington,
#inaugurationäó_                                     1
                    #womensmarch, #womensmarchonwashington,
#womensmarch2017, #snl                               1
                    #womensmarch, #womensmarchpdx
1
                    #womensmarch, #womensmarchsanjose
1
                    #womensmarch, #womxnsmarchseattleäó_
1
                    #womensmarch2017, #womensmarch
1
                    #womensmarchnyc, #mypussymychoice, #equalityäó_
1
                    #womensmarchonwashington, #womensmarch, #pussypower
1
                    #womensmarchparkcity, #womensmarch,
#sundancefilmfestival                                    1
                    #womensmarchseattle
1
                    #womensmarchäó_
2
                    #womensmarchŒæ
1
                    #womensrights, #womensmarch, #womensmarchchicago, #femi
nist, #freedom, #choiceäó_            1
United Kingdom      #WomensMarch
1
Uruguay             #WomensMarch, #WashingtonDC
1
Veneto              #womensmarch, #heforshe
1
```

```
Venezuela              #CarnetDeLaMiseria, #WomensMarch, #SacateTuCarnet, 
#GanaSeguidoresVenezuela, #OrquideaSSS        1
Victoria               #Islam, #WomensMarch
1
Wales                  #WomensMarch
2
Western Australia
1
attributes={}, id=u  #WomensMarch, #WomensMarchTO, #whoruntheworldGIRLS
1
dtype: int64
```

The hashtag #WomensMarch, was widely used across various countries. USA has most hashtags.

Analysis by US state and hashtags

In [48]:

```
tweets_df.loc[tweets_df['country'] == 'USA'].groupby(['State','Hashtags']).
size()
```

Out[48]:

```
State  Hashtags

AK     #WhyIMarch, #WomensMarch
1
AL     #WomensMarch
1
AR     #womensmarch, #womensmarcharkansas, #bettertogether
1
AZ     #WomensMarch
1
CA
4
       #Calexit, #WomensMarch
1
       #InTheNameOfLove!, #equality, #WomensMarch, #LoveTrumpsHate, 
#humanrights                            1
       #WomansMarchSanDiego, #resist, #WomensMarch, #sandiego, 
#demonstration, #equalityäó_                  1
       #WomensMarch
15
       #WomensMarch, #CatchyAF
1
       #WomensMarch, #IniteNational, #TheFutureisFemale, #TheFutureIsNow
1
       #WomensMarch, #Oakland, #resist
1
       #WomensMarch, #OaklandWomensMarch
1
       #WomensMarch, #WomenWhoHaveInspiredMe, #Madonna
1
       #WomensMarch, #WomensMarchLA
1
       #WomensMarch, #WomensRights
2
       #WomensMarch, #WomensRightsAreHumanRights
```

1
#WomensMarch, #firedupreadytogo_ôî«
1
#WomensMarch, #girlpower
1
#WomensMarch, #mylifematters
1
#WomensMarch, #photo
1
#WomensMarch, #sf
1
#WomensMarch, #vivalamujer
1
#WomensMarch, #welldone
1
#WomensMarch, #whyimarch, #nastywoman, #lovenothate, #democracy
1
#WomensMarchLA, #WomensMarch, #PershingSquare.
1
#aclu, #womensmarch
1
#dtla, #trump, #womensmarch
1
#teamRabinovitz, #sacramento, #womensmarch
1
#voice, #Dtla
1

..
TX #WomensMarch, #StandUpFightBack
1
#WomensMarch, #marchonaustin
2
#WomensMarch, #marchonaustin, #austin
1
#WomensMarch.
1
#WomensMarchOnWashington, #WomensMarch
1
#WomensMarchonATX, #WomensMarch
1
#womensmarch
2
#womensmarch, #ATX
1
#womensmarch, #austin, #womensmarchonaustin, #smashthepatriarchy
1
#womensmarch, #elpaso, #boundlessacrossborders, #streetphotography,
#borderlands, #americanäó_ 1
UT #womensmarchparkcity, #womensmarch, #sundancefilmfestival
1
VA #EthicsMatter, #truth, #WomensMarchOnWashington, #WomensMarch
1
#WhyWeMarch
1
#WomensMarch
2
#WomensMarchOnWashington, #womensmarch
1
#womensmarch

```
3
       #womensmarch, #roanokevirginia
1
VT     #WomensMarch
1
WA     #ImpeachTrump, #WomensMarch
1
       #WomensMarch
6
       #WomensMarch, #Seattle
1
       #WomensMarch, #wearenasty, #standupforyourrights
1
       #WomensMarch, #womxnsmarchseattle
1
       #nofilter, #missingobama, #nevertrump, #armyoflove, #womensmarch, #w
omensmarchonwashington         1
       #thefutureisfemale, #womensmarch, #seattleäó_
1
       #womensmarch
2
       #womensmarch, #womxnsmarchseattleäó_
1
       #womensmarchseattle
1
WI     #WomensMarch
2
       #womensmarchŒæ
1
dtype: int64
```

California has the most hashtags within USA. The hashtag #WomensMarch was the most used one.

Conclusion

The Women's March tweets data has 3100 entries with 5 columns. There are no missing values in 4 columns whereas the column named "place" has 308 known values(or 2792 missing values).

All the entries have the same id. Also, the id looks unusual. 51 different sources were used to post the tweets. Most people used "Twitter for iphone" as their source, followed by "Twitter for Android" and "Twitter webclient".

The tweets were created in different countries having different timezones and in different languages. Surprisingly, all the tweets were created between the time 23:56 and 23:59 on January 21, 2017.

The Women's March happened in various countries. The tweets include both abbreviated and full name of the US states.

The column named "place" holds the location from where the tweets are published. It has 308 known values(or 2792 missing values). The analysis was carried out with the known values.It was split into "city" and "State" to facilitate the analysis.

The dataframe was processed in detail to represent all the US states in an abbreviated form in the column named "State" and USA as their country in the column named "country". The Non-USA locations are unchanged.

The tweets came from 30 different countries. From the known location tweets, it is found that the major tweets came from USA.

There were tweets from 41 US states. The state CA(California) has the major tweets, followed by DC(The District of Columbia) and NY(Newyork). There were 3100 tweets in the data. The tweets include hashtags, text messages, links, tags and emojis or some symbols. The tweets were in different languages.

Analysis by Tags was carried out to find the most used tags. 399 rows were found. The most tweets doesn't have a tag. Some rows have more than one tags. The tags in each row are split and counted separately. There were 532 different tags found.

Analysis by links was carried out to find the most tweeted links. 1789 rows were found. 1295 tweets did not include any links. There were 1201 hashtags found. The hashtag #WomensMarch was the most used one.

The hashtag #WomensMarch, was widely used across various countries. USA has the most hashtags. California has the most hashtags within USA.