

## Introducción

Se ha realizado un streaming usando la api tweepy durante 24 horas. Como parámetros de filtrado se han usado:

```
⊕ track = ['Curie', 'Planck', 'Einstein', 'Bohr', 'Fleming', 'Higgs']  
⊕ languages = ['en']
```

En cada cambio de hora, se han exportado los tweets recogidos a un archivo *json* y guardado en un directorio separado ('./jsons'). Las imágenes también se han guardado en su propio directorio ('./images').

La clase *MyListener* que se presenta en el ejercicio enviado, recupera el número de tweets que solicita el usuario a través de un *input*:

```
user_input = raw_input(  
    'Cuántos tweets desea recuperar con el stream?\n')  
try:  
    num_tweets = int(user_input)  
    break  
except ValueError:  
    print("El valor introducido no es un número entero.\n")
```

En el archivo adjunto *my\_listener.py* se muestra la clase tal y como se usó para recuperar los tweets y crear los *jsons* cada cambio de hora.

Total de tweets recogidos: **11.385**

## Distribución horaria

La distribución horaria de los tweets creados es la siguiente:

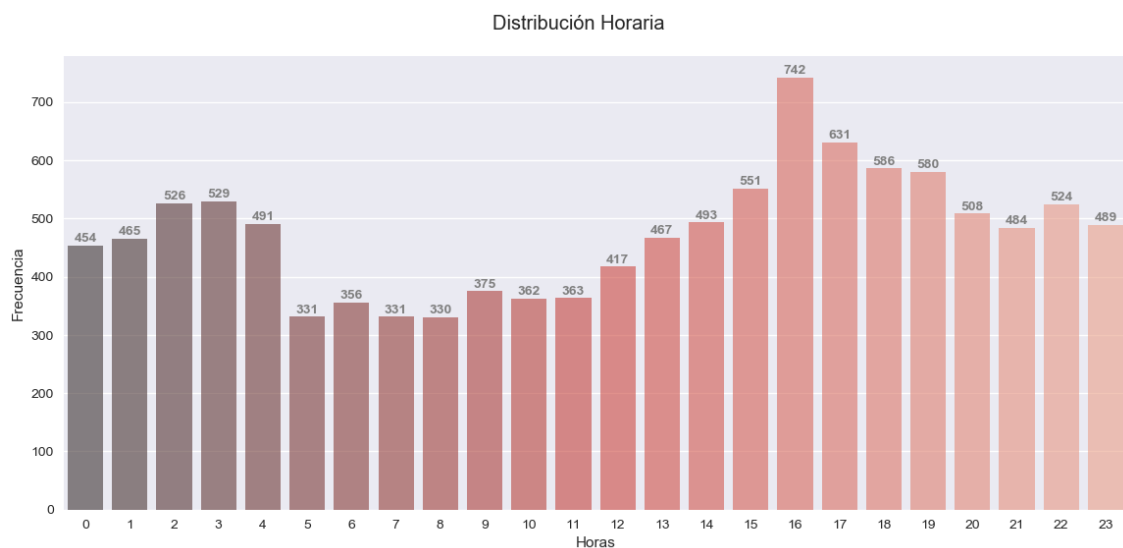


Figura 1

## Frecuencia de elementos del Query

En el siguiente gráfico mostramos el número de ocurrencias de cada uno de los elementos que conforman el *query* que se ha pasado como parámetro de filtrado.

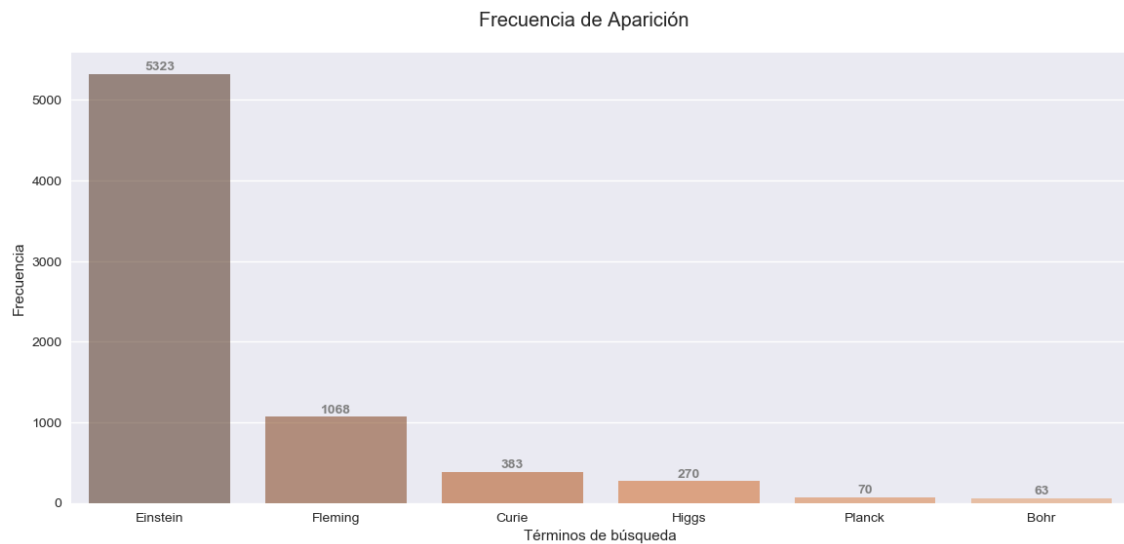


Figura 2

## Distribución de las 10 Zonas Horarias más frecuentes

En este gráfico mostramos la distribución de las 10 zonas horarias con mayor número de *tweets*.

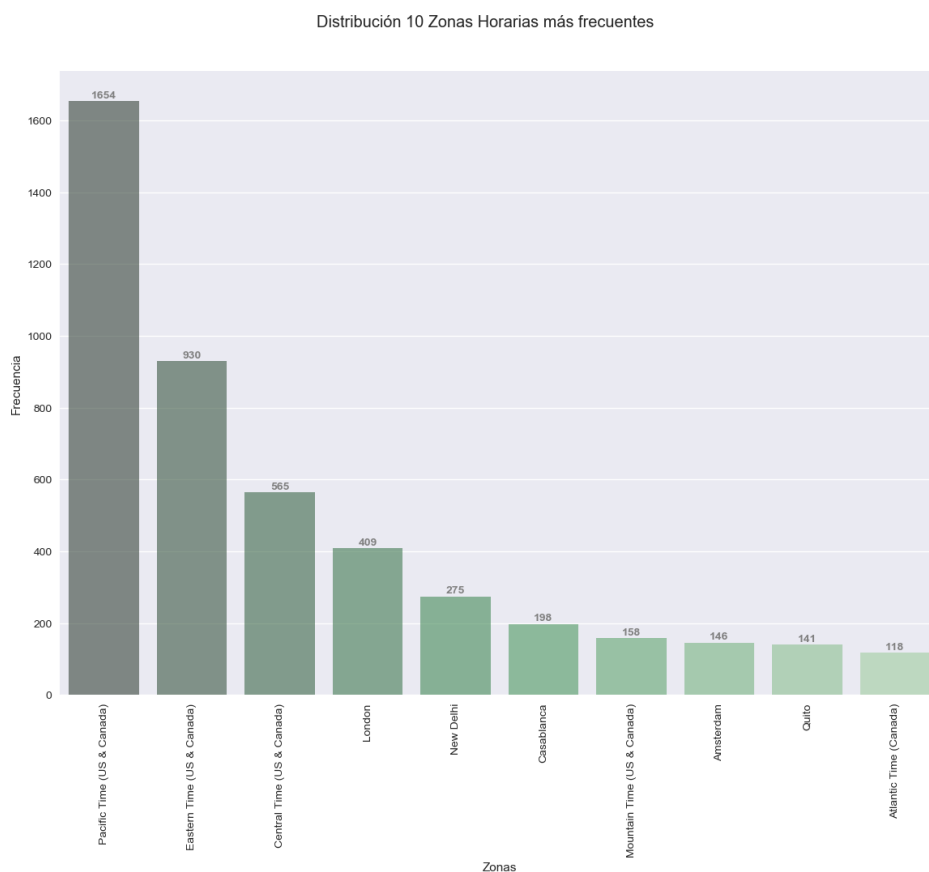


Figura 3

## Distribución de Fuentes (Sources)

Entendemos por Fuente (Sources) de un *tweet*, aquella utilidad que ha sido usada para crearlos como *Twitter for Android*, *Twitter Web Client*, etc. En el gráfico se muestran por separado los cinco porcentajes de mayor valor, el resto se agrupa bajo la etiqueta “Otros”.

La distribución es la siguiente:

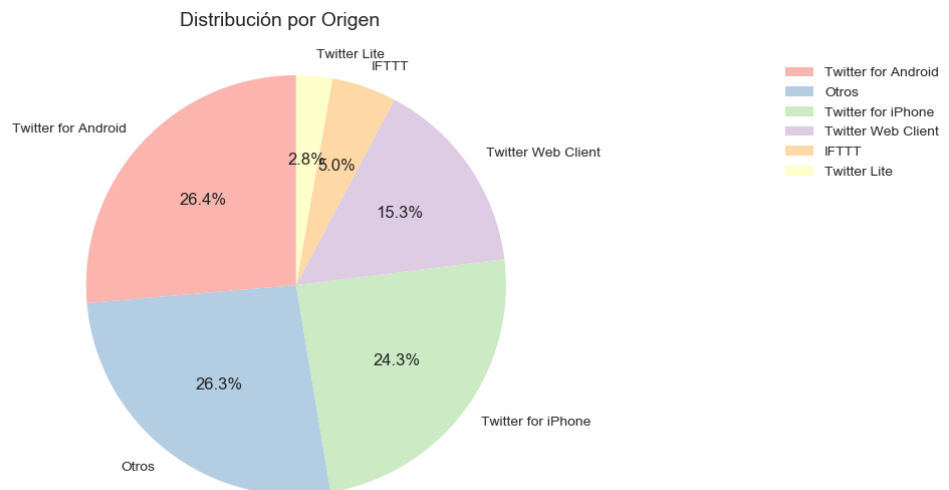


Figura 4

## Sentimental Analysis de los tweets

Haciendo uso de la librería *textblob*, clasificamos los *tweets* por su polaridad y representamos en un gráfico el resultado.

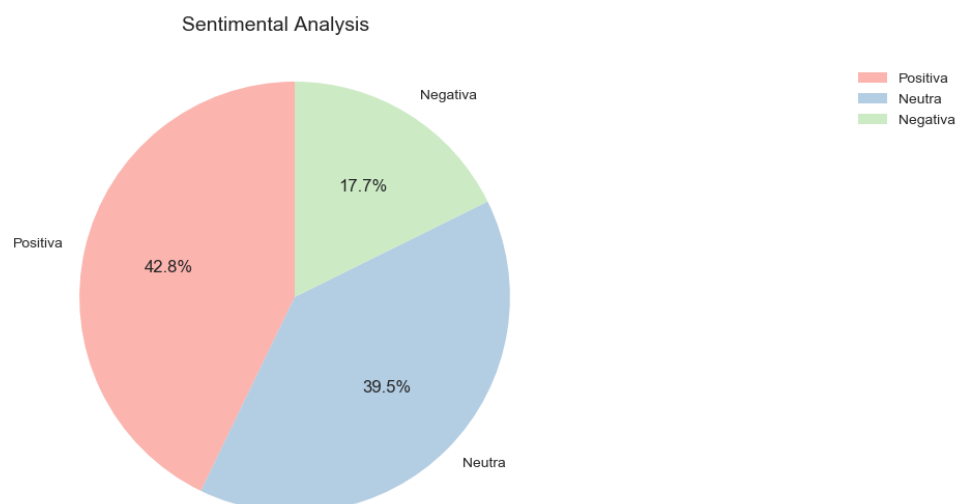


Figura 5

## Longitud media

La longitud media de los tweet es **106** caracteres

### El Tweet más largo<sup>1</sup>

```
⊕ @adidas68jack @doctorbuttons @asb5689 @keigh_see @thedesignclass  
@astronomy_ryan @wiguy45 @captscorch @jimthekingtiger @exsapperbadman  
@badibulgator @alastairjallen1 @thinkytexan @gerbusjames @tbragsdale  
@gavlaar142857 @hplarc12345 @mrbangla @uncastellsmes @ctlahey @bingchemtrails  
@becauseofnow @last_runner_up @liberal_gin @vanitistic @readyornotfory2  
@flatslugbrains @ixoye33ad @limitedview @willgamedesign @badbuc99 @col_sandurz  
@nutt007 @bigfatgit @drichards222i @dimamynedd @tomandsteveshow @flatearthcity  
@jennundercover @secretzchannel @mandokero @matty_lawrence @1984_wsmith  
@kmtildsley @mirelexx @brianhayes39 @shotgunsusie @kittenkoder @bikinatroll  
@jrodck33 no sorry, it is true einstein initially just rewrote the concept of  
the ether, eventually using it to refer to the properties of space-time, but it  
came to be viewed as an unnecessary construct even in einstein's lifetime and  
abandoned. so we concede nothing, sorry jack-ass.
```

Nº de caracteres: 843

### El Tweet más corto

```
⊕ higgs
```

Nº de caracteres: 5

---

<sup>1</sup> En los casos en los que el tweet estaba truncado, se ha cogido como texto el valor *full\_text*