

Práctica 2: Análisis de tuits sobre Cyberpunk

Autores: Mario Carrilero Sánchez, Mario Gallego Hernández, Diego Linares Espíldora, Álvaro Llera Calderón

Archivo Analizado

Ruta del archivo: cyberpunk.csv

Procedimiento

1. Cargamos los datos desde uno de los archivos CSV (cyberpunk.csv) proporcionados por el profesor desde el Campus Virtual.
2. Filtramos los tuits con al menos 10 retweets para enfocarnos en los de mayor relevancia.
3. Seleccionamos una muestra representativa de hasta 1000 registros.
4. Procesamos el contenido textual eliminando puntuación, convirtiendo a minúsculas y eliminando stopwords.
5. Analizamos los términos más frecuentes para identificar patrones en el contenido.
6. Exploramos las relaciones clave mediante gráficos, como la longitud del texto y los retweets, tendencias temporales, palabras clave y la distribución de retweets.

Términos Más Frecuentes

Término	Frecuencia
rt	988
cyberpunk	768
cyberpunkcortes	161
2077	160
boys	156
amp	87
game	78
farevalee9s	78
genshinimpact	78
\u539F\u795E	78

Análisis Adicionales

Relación entre longitud del texto y retweets

Este gráfico muestra cómo la longitud de un tuit afecta a su popularidad, medida en retweets.

Cantidad de tuits a lo largo del tiempo

Este gráfico refleja la distribución temporal de los tuits, destacando picos de actividad.

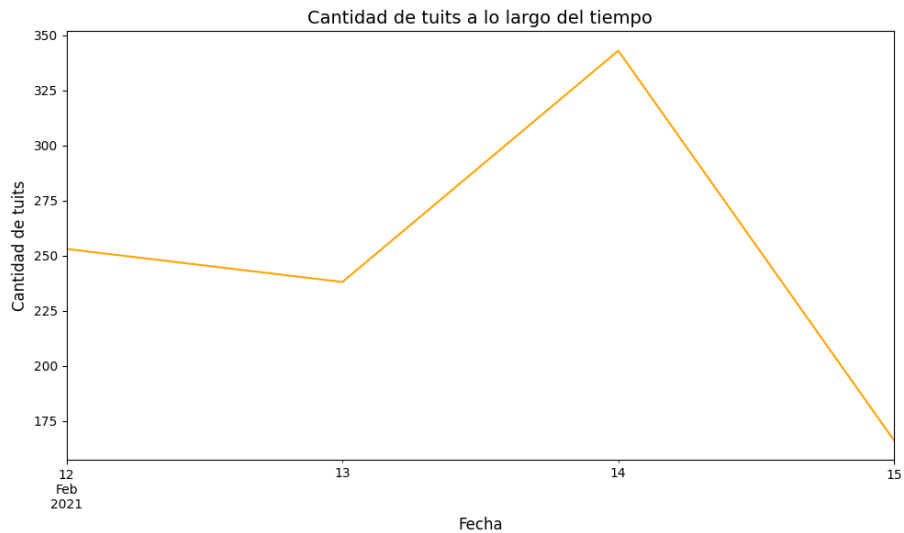


Figure 3: Cantidad de tuits a lo largo del tiempo

Promedio de retweets según palabras clave

Aquí se comparan las palabras clave más relevantes según su influencia en los retweets.

Distribución de retweets

Este gráfico ilustra cómo están distribuidos los retweets en el conjunto de datos.

Dispositivos donde se realizan los tweets

En este grafico analizamos los dispositivos en los que más se twiteó, habiendo un gran reparto entre la propia aplicación de Twiter, iPhone y Android, siendo la que más en la aplicación de twiter, pero sin sacar gran ventaja a las otras dos. Cabe destacar que este análisis los hicimos con un gráfico circular.

Distribucion de tweets por horas

Este grafico muestra las horas en las que más se han publicado tweets.

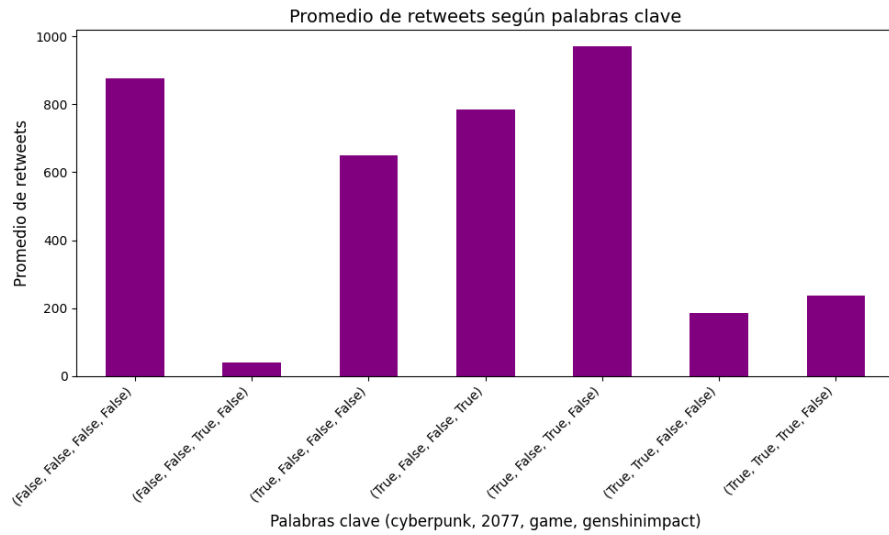


Figure 4: Promedio de retweets según palabras clave

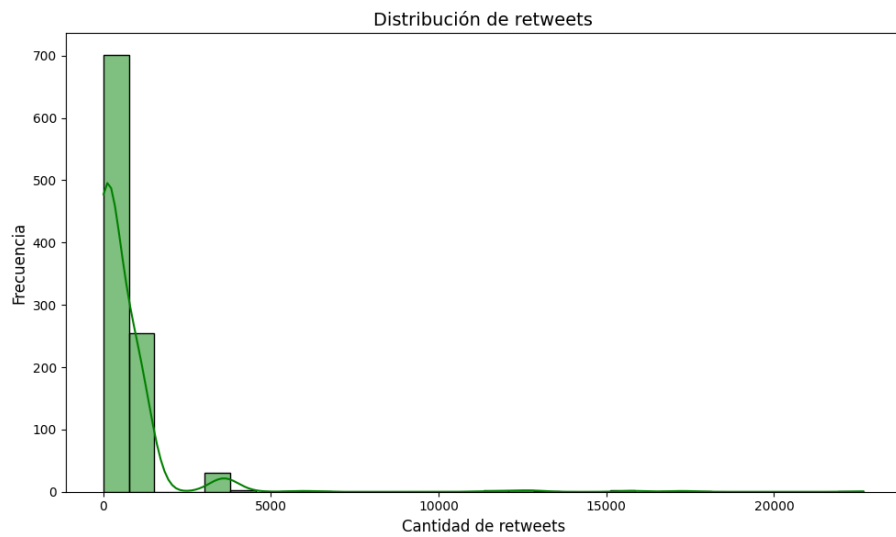


Figure 5: Distribución de retweets

Distribución de tuits por dispositivo (agrupados los menores al 1.5%)

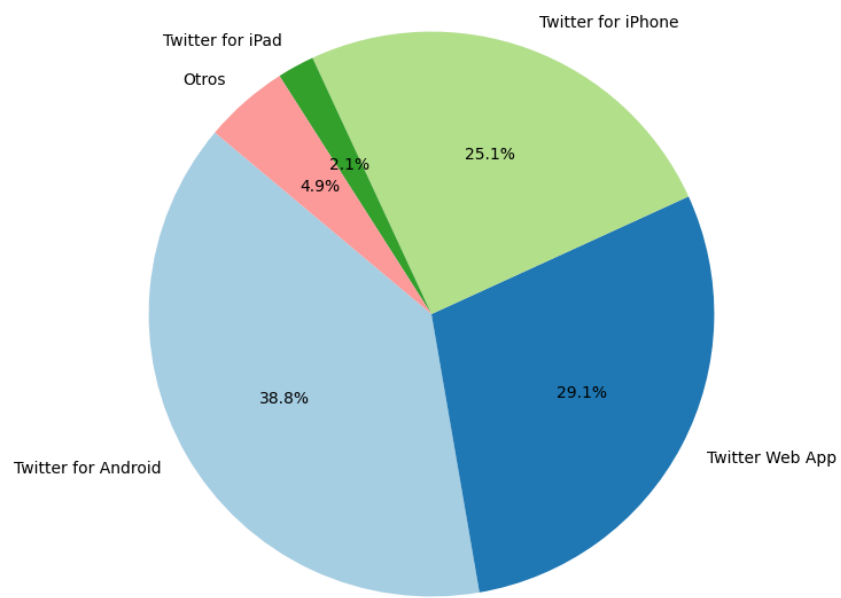


Figure 6: Dispositivos donde se realizan los tweets

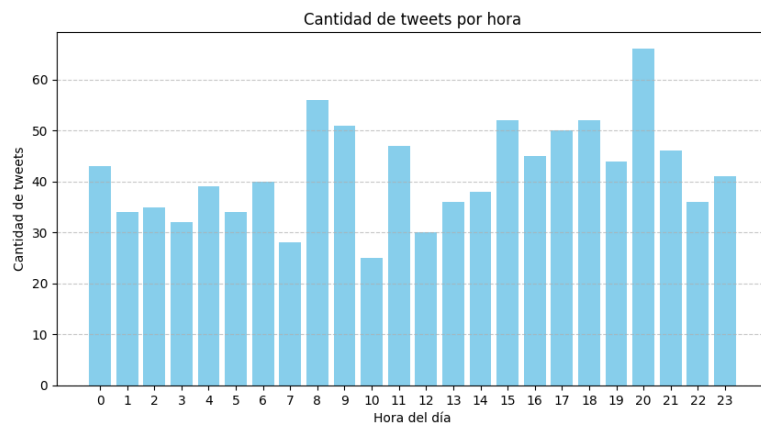


Figure 7: Distribucion de tweets por horas