

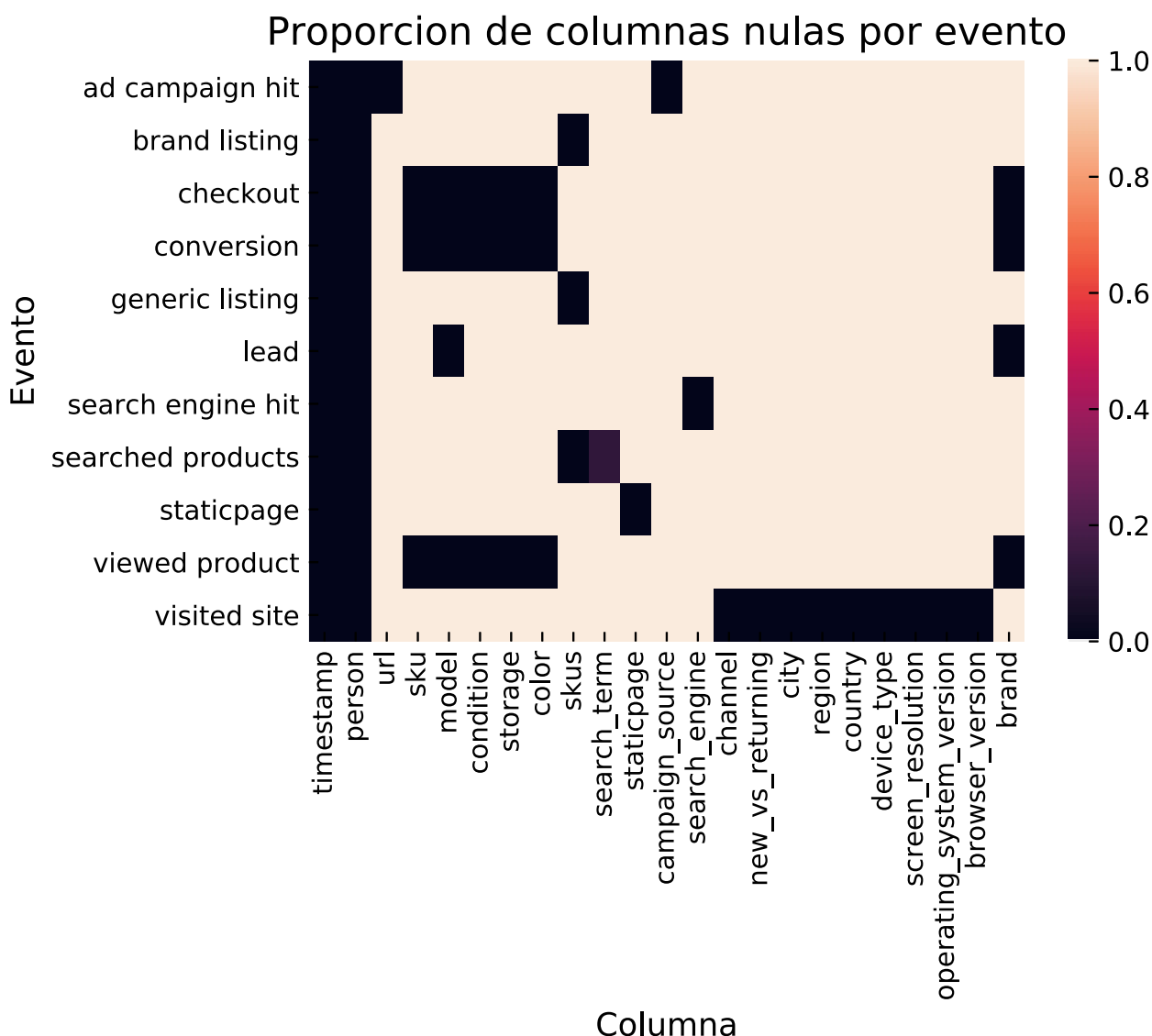
# **Informe TP1**

## **Grupo null**

- Carlos Talavera
- Federico Jure
- Juan Pablo Capurro

# Introducción

Al estar todos los eventos en un mismo dataframe, creimos que los eventos iban a tener una cantidad importante de columnas nulas dependiendo del tipo de evento.



Efectivamente, dependiendo del tipo de evento, las columnas son nulas o no en un 100%, con excepción del campo search\_term en el que hay una pequeña proporción de nulos.

## Exploracion

### Exploración de tipos de evento por separado

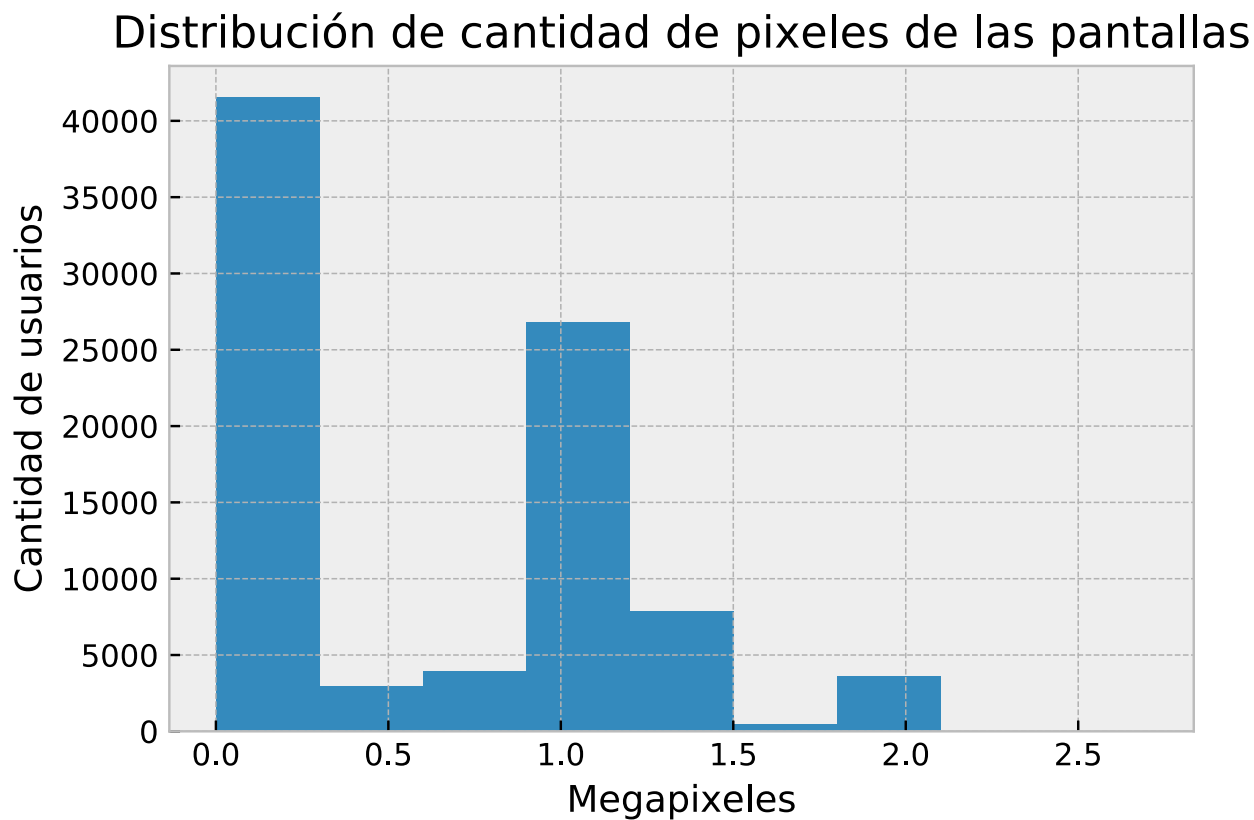
#### Eventos de búsqueda

Procedemos a ver si hay registros inválidos de búsquedas, y si amerita dropear registros. Por un lado, hay una proporción importante (7k nulos en 56k total) de eventos de búsqueda que tienen NaN como search\_term, pero tienen distintas listas de skus, por lo que podemos suponer que hay otros factores que afectan a la búsqueda.

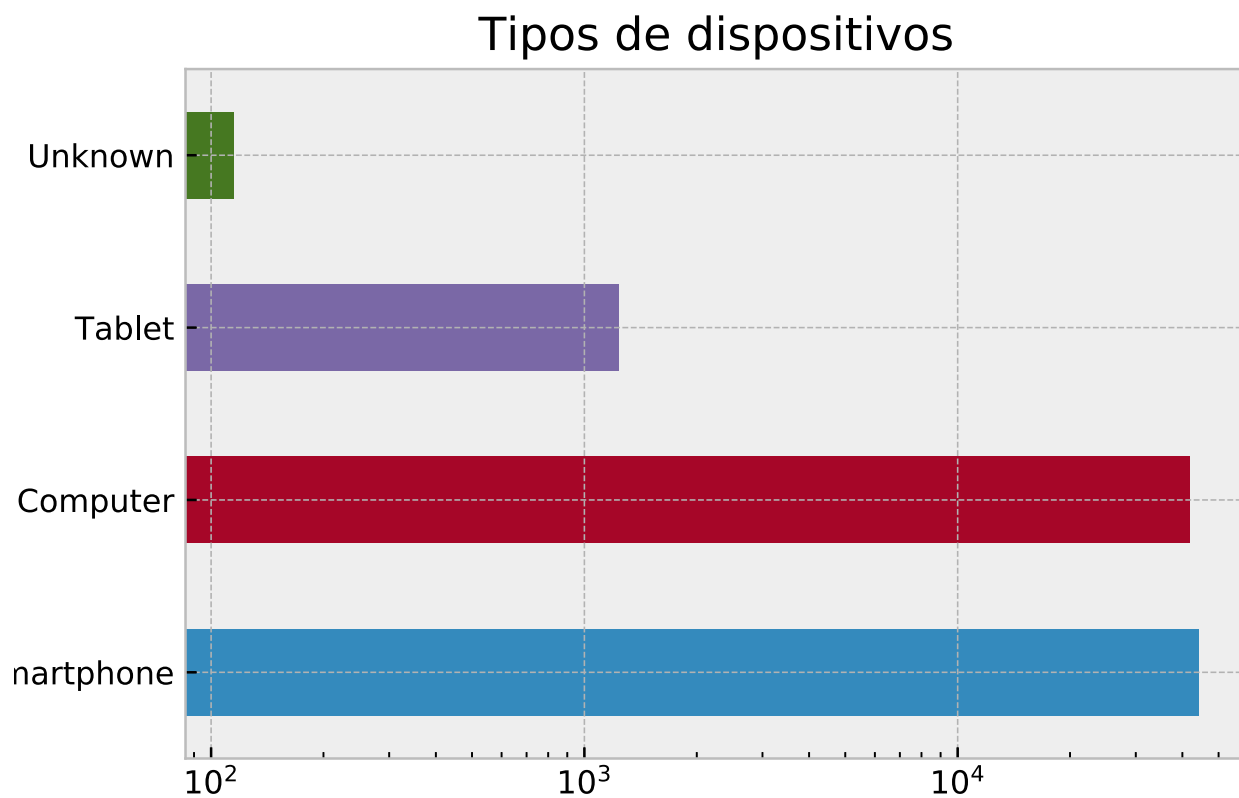
#### Eventos de visita de sitio

Consideramos la resolución de pantalla una forma de ver qué poder adquisitivo tienen las personas que visitan el sitio. Medimos la cantidad de píxeles de las pantallas, porque hay muchas variantes de

resoluciones y solo nos importa el tamaño.

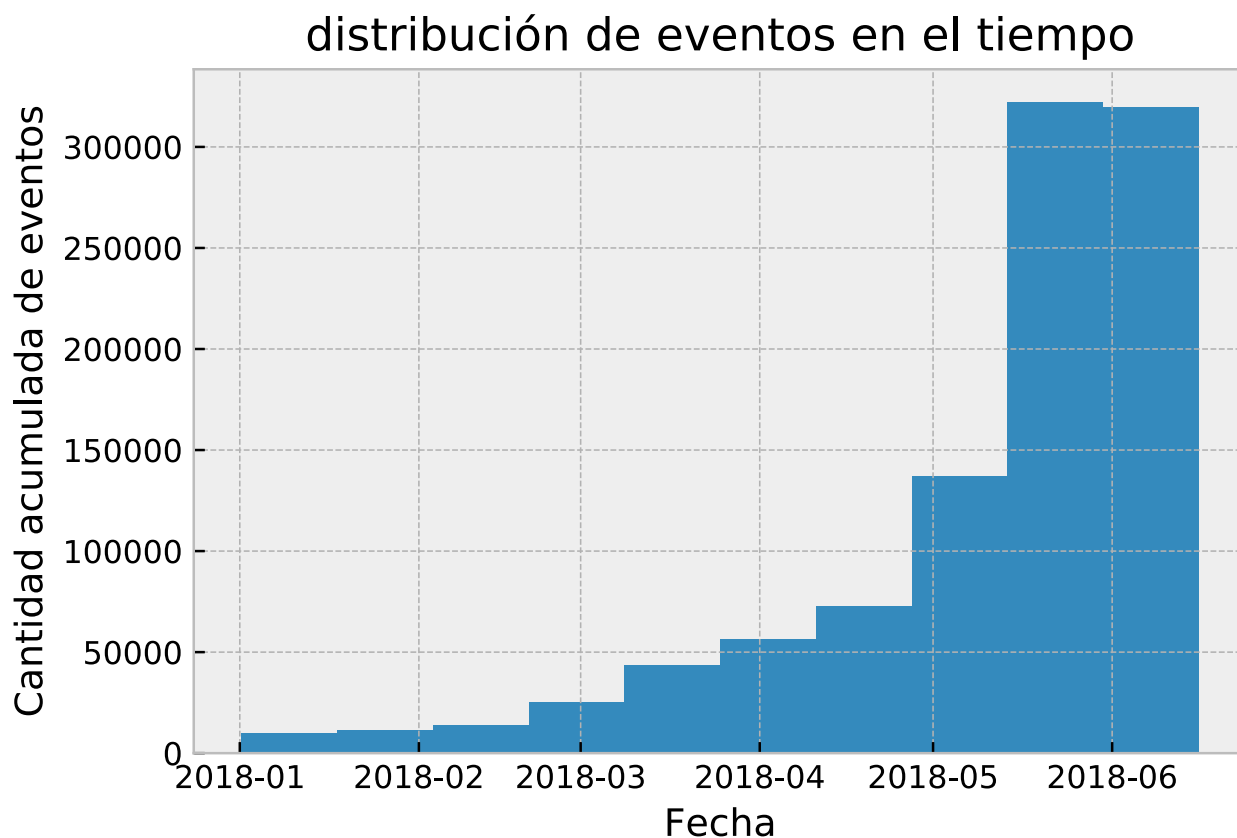


Nos interesó también que proporción de los usuarios accedían desde mobile y cuántos desde desktop

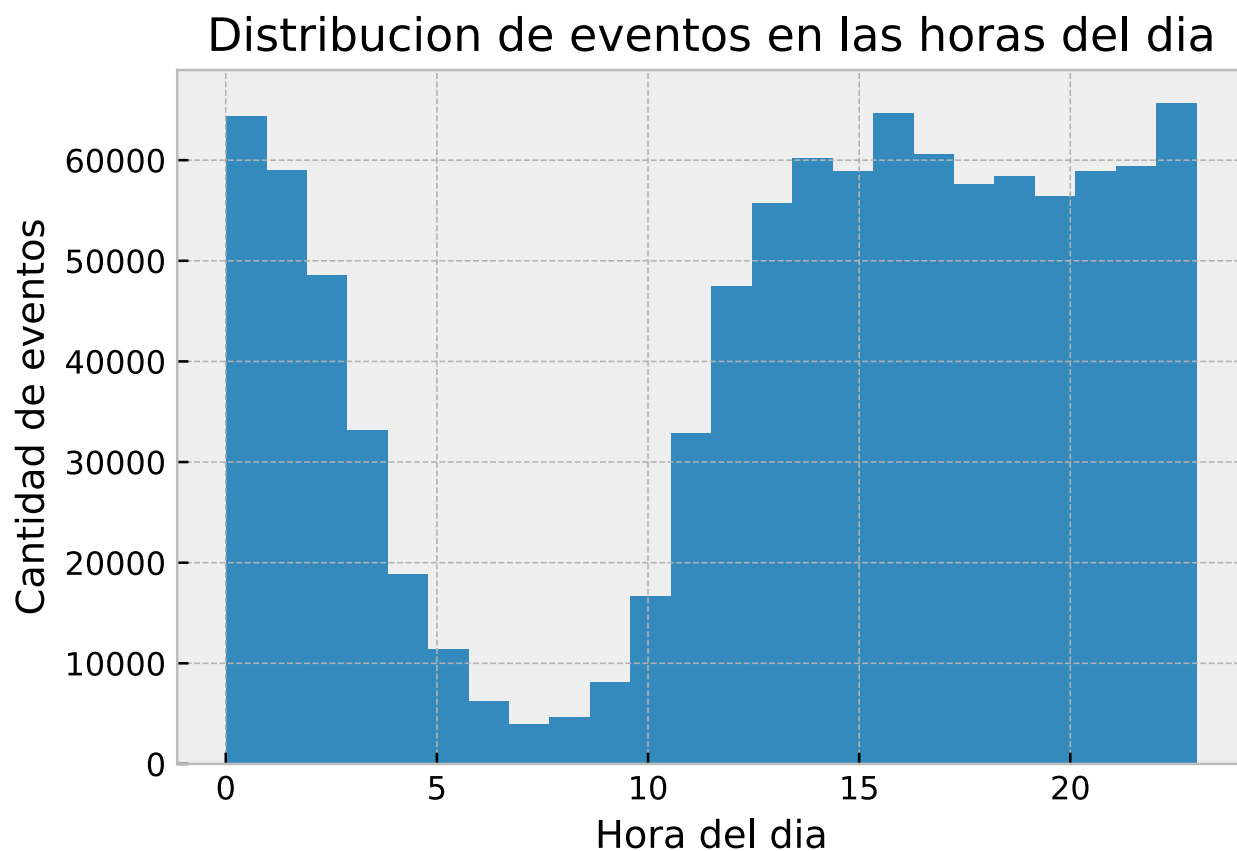


**Exploracion de los eventos en conjunto**

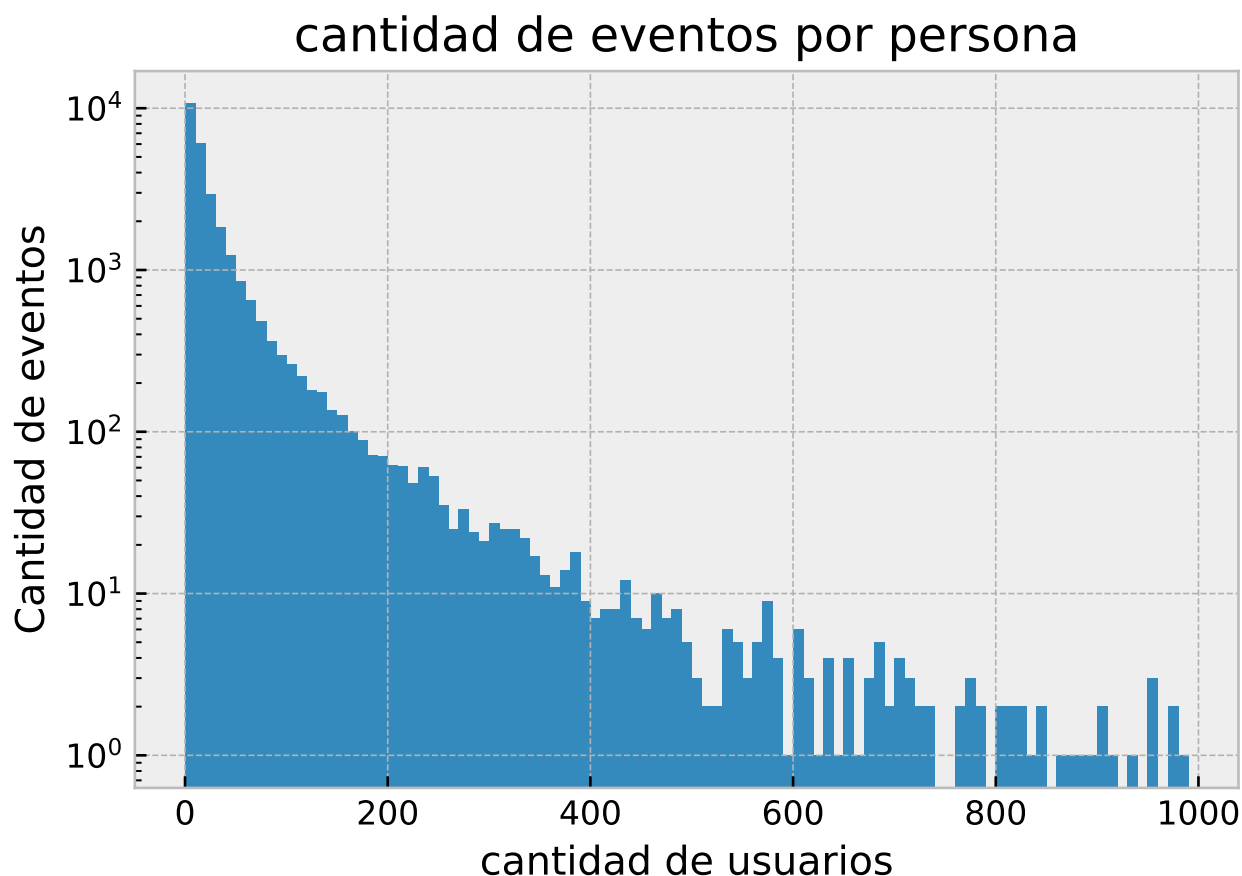
El uso de la plataforma aumentó enormemente a lo largo de los últimos meses:



Y la distribución de los eventos a lo largo del día no nos da muchas sorpresas:



Los usuarios pueden tener una cantidad variable de eventos, y es usual que tengan algunos cientos.



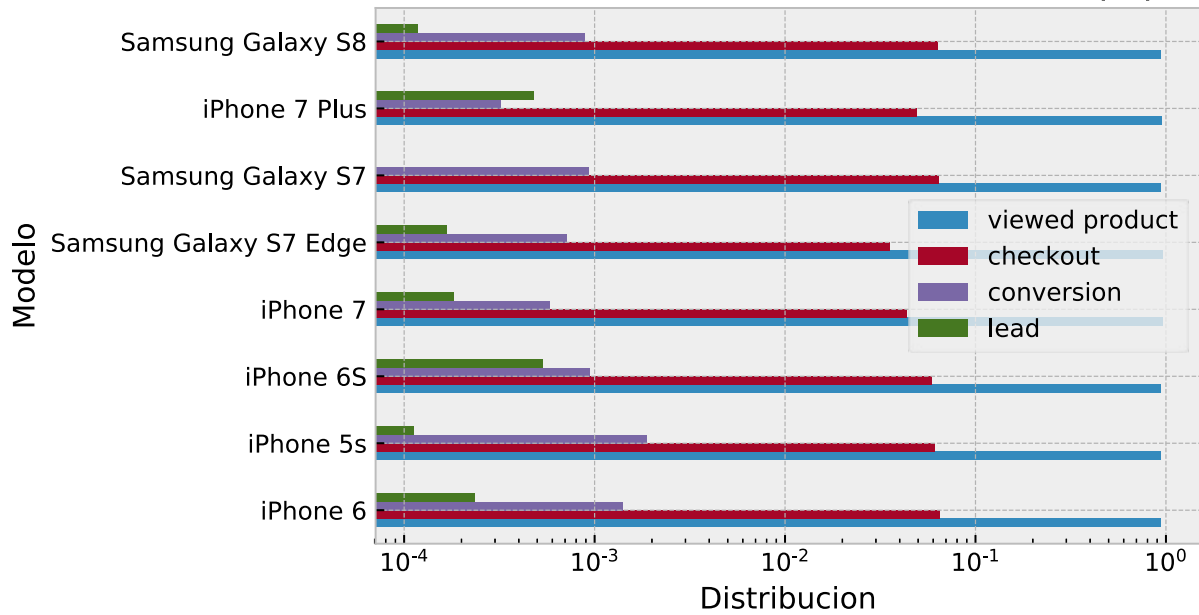
No encontramos relación entre el tiempo que pasa un usuario en total en el sitio y la cantidad de eventos que genera de ningún tipo

## Exploraciones de los distintos modelos

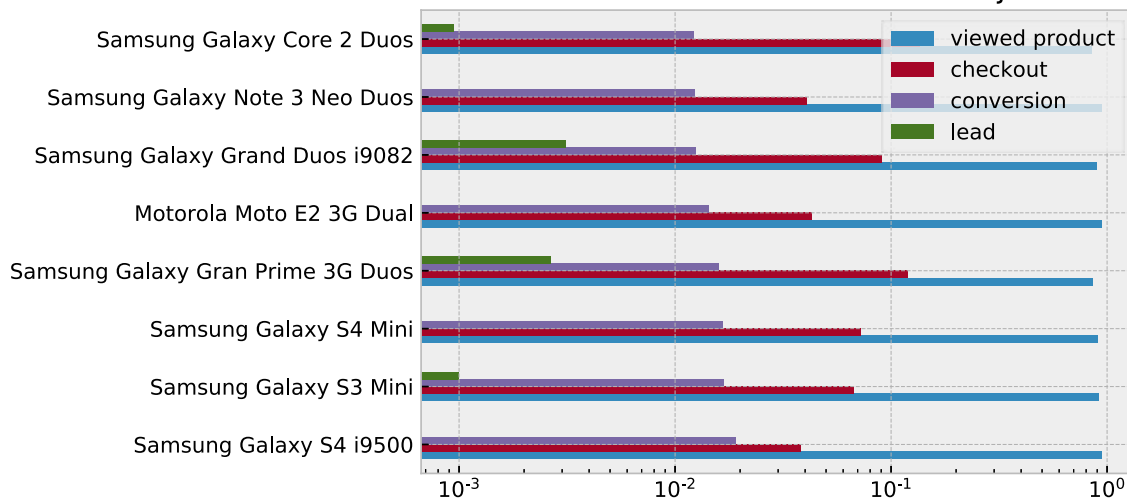
Encontramos que incluso filtrando aquellos modelos con menos de 100 eventos, buscar cuáles presentaban mejor ratio de leads introducía bastante ruido. Por ejemplo, aparecían modelos sin conversiones, por lo que consideramos esta columna relativamente desestimable.

Por otro lado, encontramos que los modelos con más vistas no overlapean mucho con los que tienen mayor ratio de conversiones:

## Distribucion de eventos en los modelos mas populares

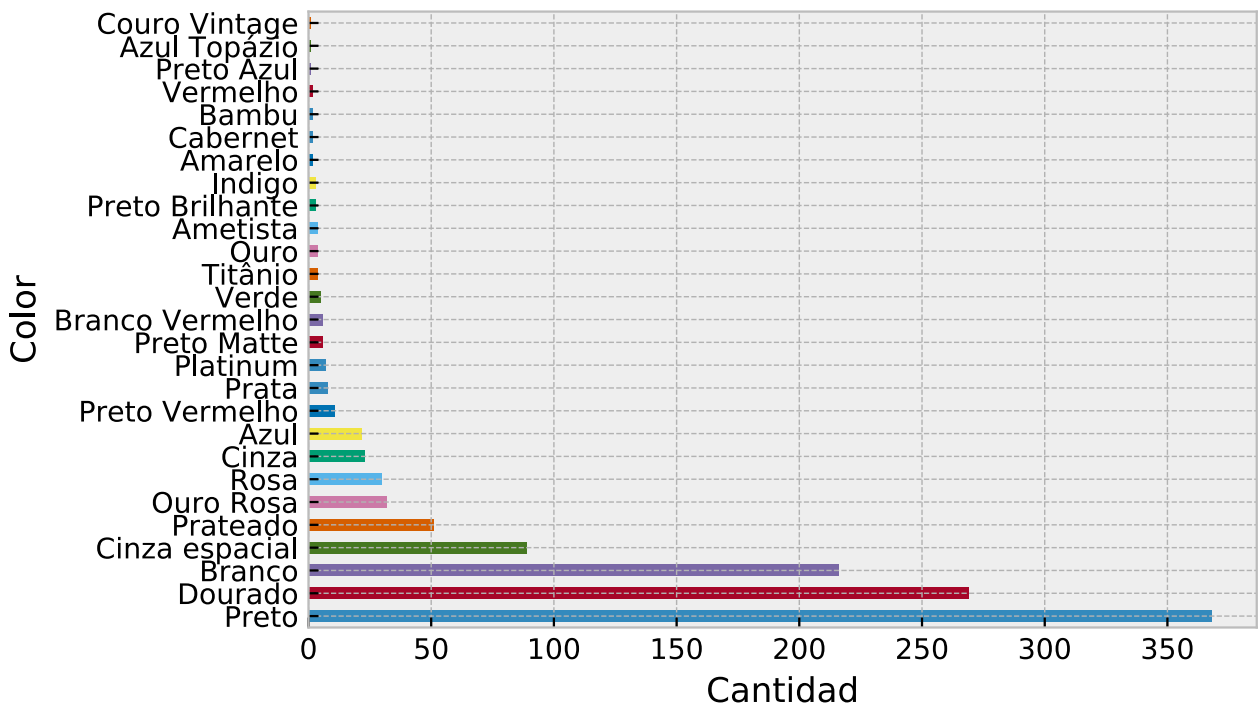


## Distribucion de eventos en los modelos con mejor conversion rate



El color 'negro' es por lejos el mas vendido

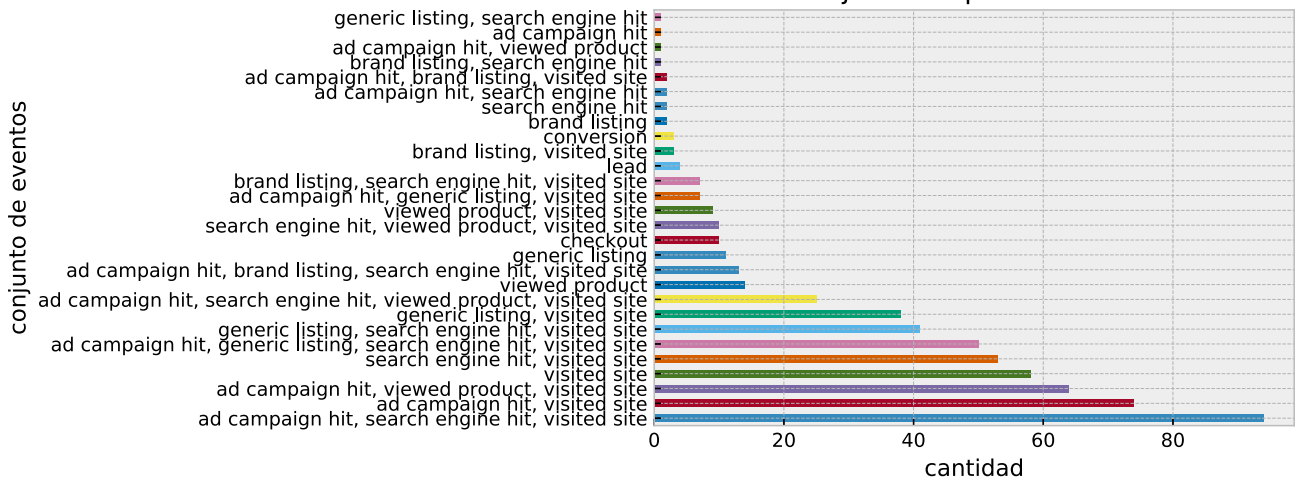
## Colores de telefonos mas vendidos



## Exploración sobre los usuarios

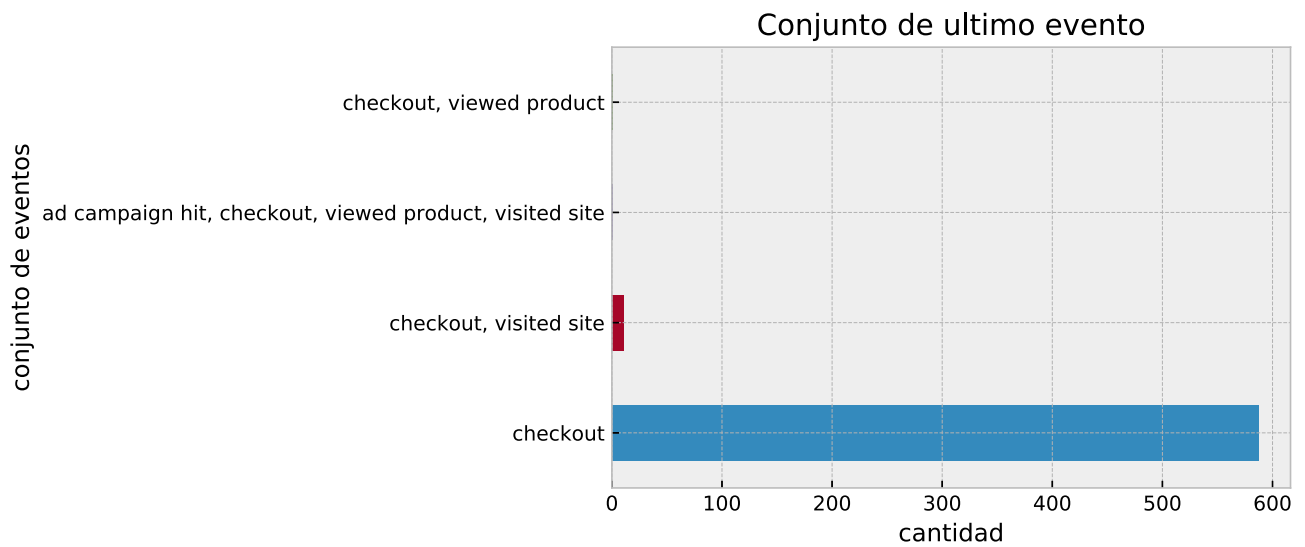
Pudimos observar que en el caso del primer evento del usuario, se hallan en el mismo segundo varios otros eventos de tipos relacionados, que refieren a la misma accion pero desde distintos puntos de vista:

### Conjunto de primer evento



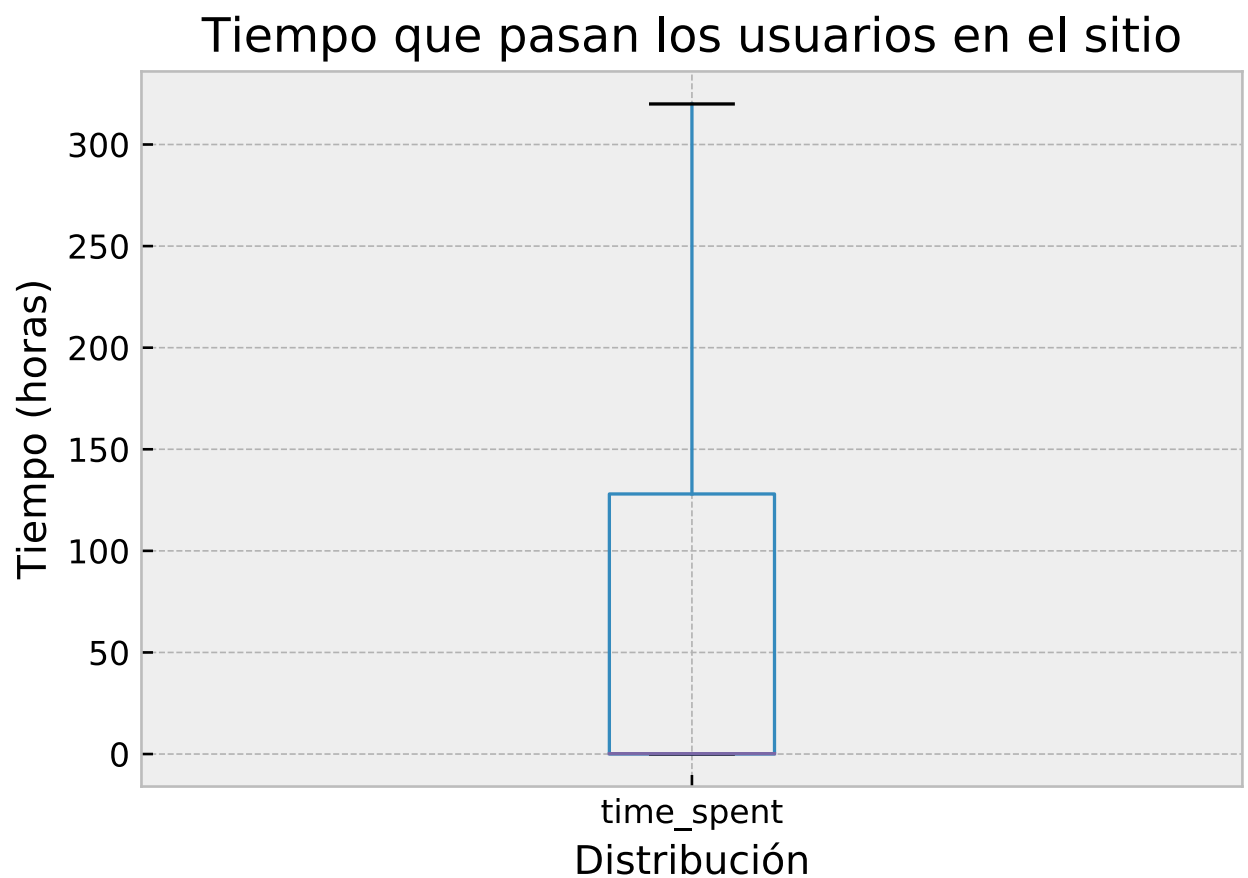
Por ejemplo, lo más usual es que se llegue a visitar el sitio por un ad campaign hit desde un search engine.

En cuanto a los ultimos eventos de un usuario, estos no suelen aparecer en grupos



### Distribución temporal de los eventos

Los usuarios pasan cantidades de tiempo muy variadas en la plataforma, y se puede destacar que hay una proporción alta de outliers que tienen eventos separados por varios miles de horas. Estos no son mostrados en el gráfico porque lo volverían ilegible.



Por otro lado, no se encontró relación entre el tiempo que un usuario lleva usando el sitio y la cantidad de eventos de algún tipo en particular que genera.

## Insights