

UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE INGENIERÍA

**75.06 Organización de Datos Trabajo Práctico 1**

**Segundo Cuatrimestre de 2018**

Grupo 36

|  |  |
| --- | --- |
| Franchi, Eugenia | 100855 |
| Sanjines, Aaron | 99176 |
| Zugna, Federico | 95758 |
| Delgadillo, Cesar | 89835 |

**Link de GitHub:** [**https://github.com/EugeniaFranchi/TP-Datos**](https://github.com/EugeniaFranchi/TP-Datos) **(Privado)**

# Índice

# Introducción

Este informe se encarga de analizar los datos sobre un conjunto de eventos de web analytics de usuarios que visitaron [www.trocafone.com](http://www.trocafone.com/), su plataforma de ecommerce de Brasil. Trocafone es un side to side Marketplace para la compra y venta de dispositivos electrónicos que se encuentra actualmente operando en Brasil y Argentina.

La empresa realiza distintas actividades que van desde la implementación de plataformas de trade-in (conocidos en la Argentina como Plan Canje), logística directa y reversa, reparación y recertificación de dispositivos (refurbishing) y venta de productos recertificados por múltiples canales (ecommerce, marketplace y tiendas físicas).Amablemente se nos proporcionó acceso a los datos, en unos archivos .CSV, en los cuales se encontraba la siguiente información:

* **timestamp:** Fecha y hora cuando ocurrió el evento. (considerar BRT/ART).
* **event:** Tipo de evento
* **person:** Identificador de cliente que realizó el evento.
* **url:** Url visitada por el usuario.
* **sku:** Identificador de producto relacionado al evento.
* **model:** Nombre descriptivo del producto incluyendo marca y modelo.
* **condition:** Condición de venta del producto
* **storage:** Cantidad de almacenamiento del producto.
* **color:** Color del producto
* **skus:** Identificadores de productos visualizados en el evento.
* **search\_term:** Términos de búsqueda utilizados en el evento.
* **staticpage:** Identificador de página estática visitada
* **campaign\_source:** Origen de campaña, si el tráfico se originó de una campaña de marketing
* **search\_engine:** Motor de búsqueda desde donde se originó el evento, si aplica.
* **channel:** Tipo de canal desde donde se originó el evento.
* **new\_vs\_returning:** Indicador de si el evento fue generado por un usuario nuevo (New) o por un usuario que previamente había visitado el sitio (Returning) según el motor de analytics.
* **city:** Ciudad desde donde se originó el evento
* **region:** Región desde donde se originó el evento.
* **country:** País desde donde se originó el evento.
* **device\_type:** Tipo de dispositivo desde donde se genero el evento.
* **screen\_resolution:** Resolución de pantalla que se está utilizando en el dispositivo desde donde se genero el evento.
* **operating\_system\_version:** Version de sistema operativo desde donde se origino el evento.
* **browser\_version:** Versión del browser utilizado en el evento

Por otro lado, los siguientes tipos de eventos se encuentran disponibles (en el campo event) sobre los cuales se brinda una breve descripción:

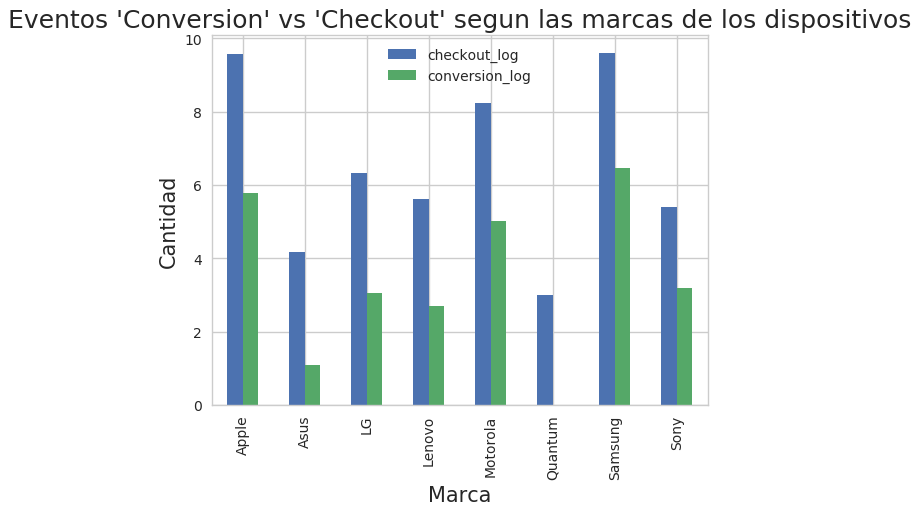
* **“viewed product”:** El usuario visita una página de producto.
* **“brand listing”:** El usuario visita un listado específico de una marca viendo un conjunto de productos.
* **“visited site”:** El usuario ingresa al sitio a una determinada url.
* **“ad campaign hit”:** El usuario ingresa al sitio mediante una campana de marketing online.
* **“generic listing”:** El usuario visita la homepage.
* **“searched products”:** El usuario realiza una búsqueda de productos en la interfaz de búsqueda del site.
* **“search engine hit”:** El usuario ingresa al sitio mediante un motor de búsqueda web.
* **“checkout”:** El usuario ingresa al checkout de compra de un producto.
* **“staticpage”:** El usuario visita una página
* **“conversion”:** El usuario realiza una conversión, comprando un producto.
* **“lead”:** El usuario se registra para recibir una notificación de disponibilidad de stock, para un producto que no se encontraba disponible en ese momento.

# Comparación por marca

Para comenzar, primero se realizará un análisis comparando por la marca de cada celular, teniendo en cuenta que tanto el iPhone como el iPad son productos correspondientes a la marca Apple.

## Eventos según marcas

Lo primero que se realiza es la construcción de una tabla para poder visualizar los eventos “checkout” y “conversión” según las marcas de los dispositivos.

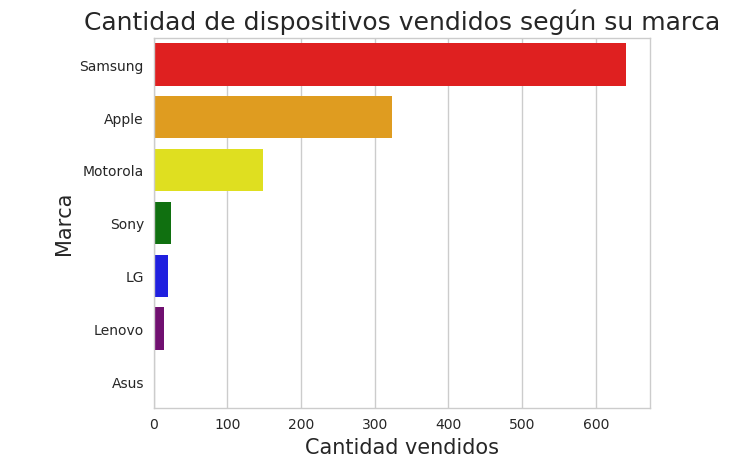


*Figura 2.1 Eventos según las marcas*

Como se puede observar en el gráfico, el usuario realiza muchos más checkouts que conversiones. Las dos marcas que lideran esta tabla son las dos marcas líderes del mercado como Apple y Samsung seguidos de cerca por Motorola. Puede observarse ademas que los registros de checkout son proporcionales a los de conversión.

## Cantidad de dispositivos vendidos según la marca

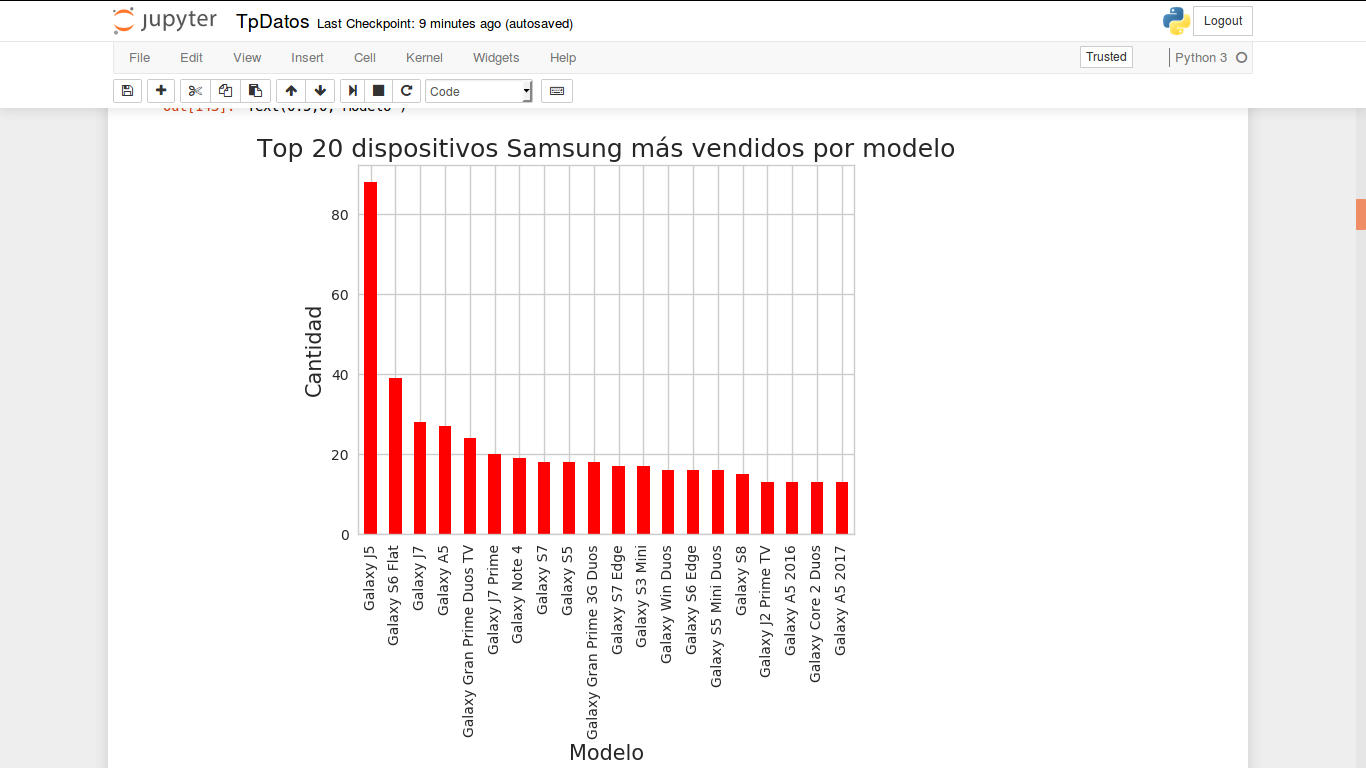
Ahora veremos la cantidad de dispositivos vendidos según la marca de cada uno.



*Figura 2.2 Cantidad de dispositivos vendidos según marca*

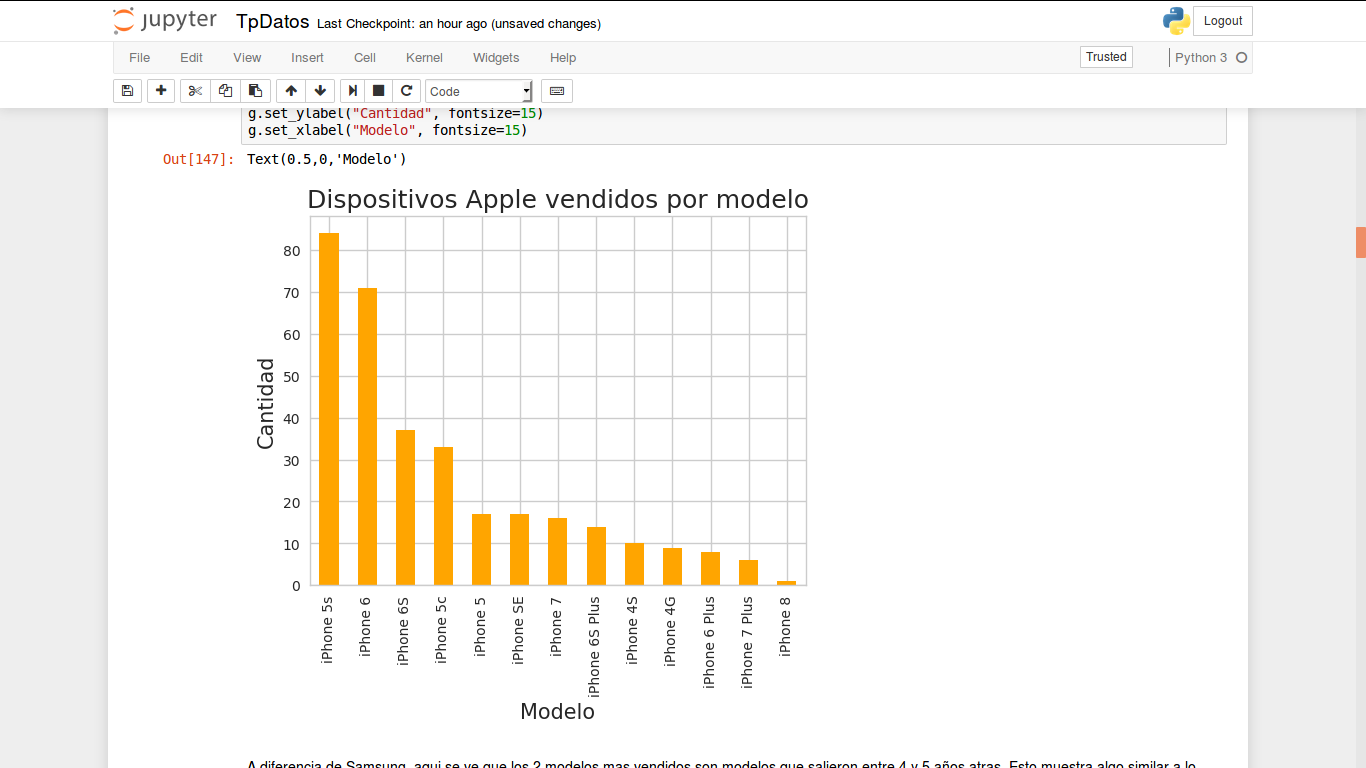
Se observa en el gráfico que los dispositivos de marca Samsung son los más vendidos junto con los Apple y los Motorola, habiendo mucha diferencia con el siguiente que es Sony. La amplia diferencia que hay entre Samsung y Apple puede ser por la alta cantidad de modelos de celulares activos que tiene en el mercado.

## Dispositivos Samsung más vendidos

******

Se puede ver que los celulares Samsung más vendidos no son los más nuevos, sino que son celulares que fueron lanzados al mercado hace 3 años aproximadamente. Esto demuestra que los usuarios tienen tendencias a comprar dispositivos que no están discontinuados pero que por otro lado ya han sido sucedidos por varios modelos nuevos.

## Dispositivos Apple más vendidos

A diferencia de Samsung, aquí se ve que los 2 modelos más vendidos son modelos que salieron entre 4 y 5 años atrás. Esto muestra algo similar a lo sucedido con Samsung, demostrando que los dispositivos de Apple más nuevos suelen ser muy caros y por lo que la gente que recurre a la página está más interesada en celulares viejos pero que todavía están funcionales.

## Dispositivos Motorola más vendidos

## 

## Conclusión

## Comparación por fecha

En esta etapa del análisis nos centramos en la actividad de los usuarios dentro de la pagina de Trocafone comparados por los días de cada mes.

## Cantidad de Actividad por combinación de Mes y Día

## En este punto se analizó la actividad de cada usuario en la página de la empresa gracias a este mapa de calor el cual nos indicará que a mayor actividad mayor intensidad en el color.

## En el mes de Mayo es donde se registra más actividad dentro del site de Trocafone siguiendo la tendencia en el mes de Junio.

## 

## Cantidad de Compras por combinación de Mes y Día

## En este punto se analizó la cantidad de compras de cada usuario en la página de la empresa gracias a este mapa de calor el cual nos indicará que a mayor actividad mayor intensidad en el color.

## Al igual que en el punto anterior, la mayor cantidad de compras se realizan durante el mes de Mayo, pero los días en los que la intensidad del color es más fuerte es previo a los días de alta actividad en la página.

## Cantidad de Checkout por combinación de Mes y Día

## En este punto se analizó la cantidad de checkout de cada usuario en la página de la empresa gracias a este mapa de calor el cual nos indicará que a mayor actividad mayor intensidad en el color.

## En este caso, la mayor cantidad de checkout coincide casi en su totalidad con la actividad total del sitio.

## Conclusión

## Se nota que la mayor actividad del sitio corresponde a los checkouts de los usuarios, que es cuando los usuarios ingresan al checkout de compra de algún producto, pero no siempre hacen efectiva la compra correspondiente. Esto se debe a que, según el mapa de calor, las compras están distribuidas por los meses y no se estaciona fuertemente en un período específico como pasa en el otro caso.

## Comparación por ubicación

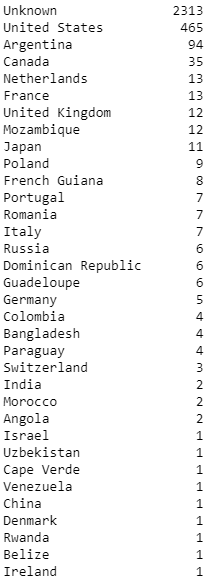
## Aquí se verá los que los accesos al sitio desde cada país del mundo.

## Cantidad de ingresos a urls especificas por país

## Casi en su totalidad, los ingresos corresponden al país Brasil, lugar donde opera activamente Trocafone juntos con Argentina.

## 

Dentro del 3.5% correspondiente a otros países del mundo tenemos a los siguientes:



Donde la mayoría no informa desde donde está consultando el site,

seguido por Estados unidos, uno de los países líderes en tecnología,

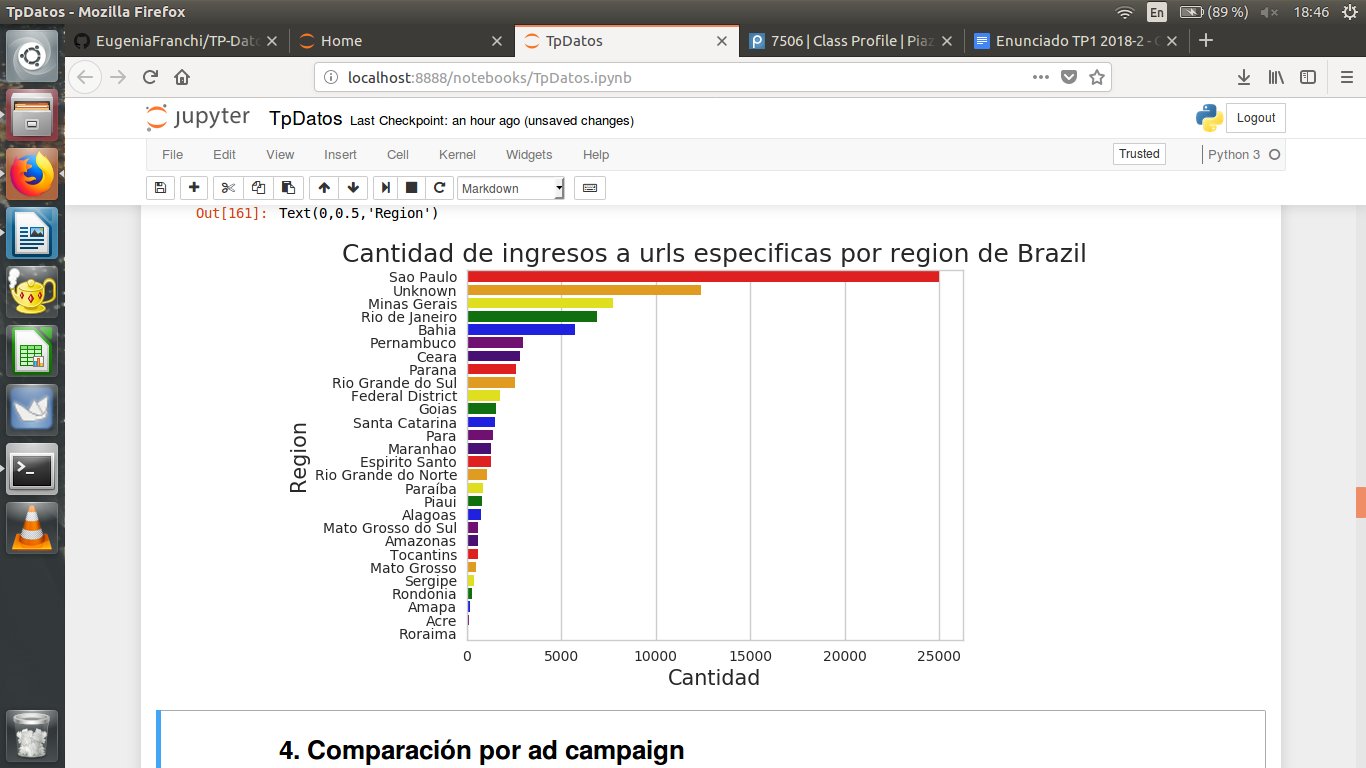
y tercero Argentina, país donde opera la empresa Trocafone.

## Cantidad de ingresos a urls especificas por Region de Brazil

## Dentro de Brasil tenemos los siguientes accesos según las regiones de este país:

## 

Como se observa en la tabla, la ciudad con más accesos al sitio es Sao Pablo, la cual es una ciudad con mucha densidad de población ya que es donde se centran las grandes corporaciones.

El siguiente dato muestra que gran cantidad de usuarios no declara desde que cuidad está ingresando al sitio. Esto mismo puede verse en el siguiente gráfico.

## Aquí también se observa que en Sao Pablo tiene una amplia ventaja con respecto al segundo del gráfico, lo cual al no saber de donde es, podría ser una suma a esta ciudad.

## Conclusión

## El mayor tráfico del sitio Trocafone es en Brasil, país donde Trocafone opera activamente juntos a Argentina. También un número alto de visitas corresponde a usuarios que no declaran el lugar de origen por el cual están ingresando al sitio. Estados Unidos, una de las tantas potencias en dispositivos electrónicos, es el tercer país con más ingresos a la página.

## Dentro de Brasil, la región con más ingresos es Sao Pablo, una de las capitales empresariales del país, por lo que hace que haya más densidad de población que en otra ciudad. Mismo caso que el anterior, el segundo es “desconocido”, lo que hace que los usuarios no compartieron su ubicación. Esto podría ayudar a las otras ciudades a crecer en los números. Luego, las tres regiones siguientes, son regiones con grandes cantidades de población.

## Comparación por Ad Campaign

## La empresa Trocafone se promociona mediante campañas de marketing online. En este punto se verá el numero de ingresos de usuarios se realizan gracias a esta vía.

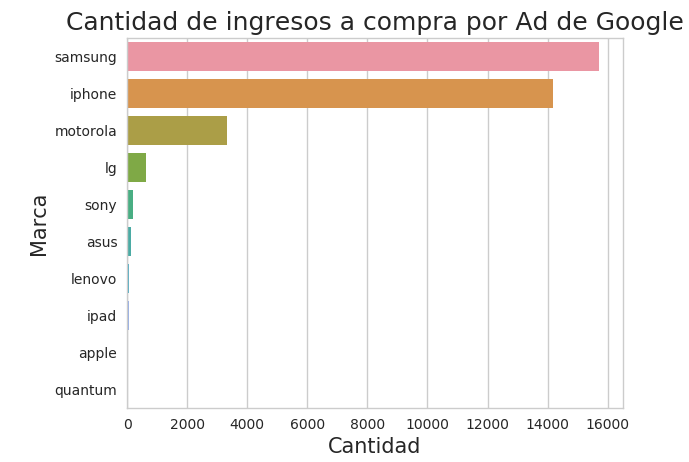
## 

## Se observa que la mayor cantidad de ingresos es por las campañas en Google, ya que está es el buscador más usado de la web y que las Ads de Google abundan en los sites de internet.

## 

## Cantidad de ingresos a compra por Ad de Google

Gracias a las Ads de Google, hay muchos accesos a la web de Trocafone. Cuando el usuario ingresa al site, gracias a esta campaña de marketing online, muchas veces lo hace directo para comprar algún dispositivo.

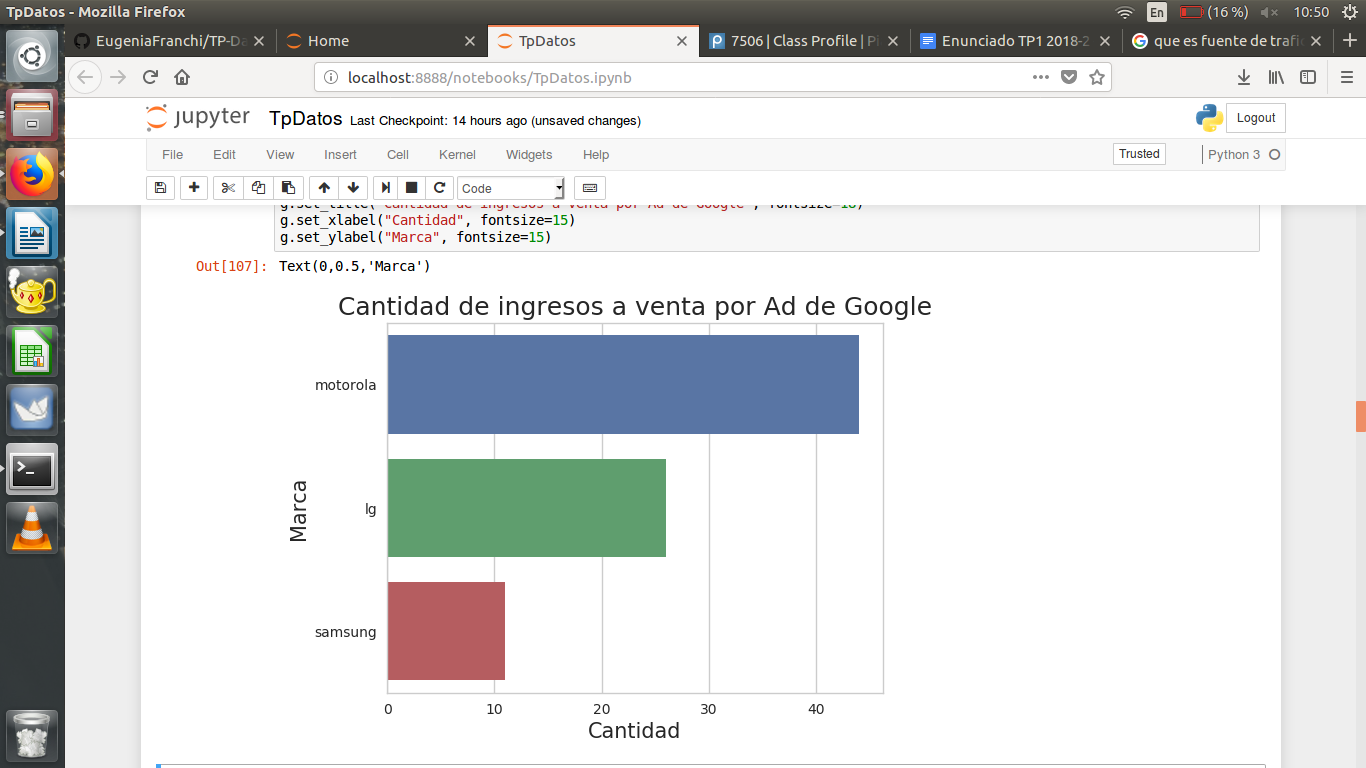


Esta vez no mostramos los valores de manera logarítmica, ya que de esta forma puede apreciarse la gran diferencia entre dispositivos samsung y apple frente a los demás.

~~Los mayores ingresos a compra son por los celulares Samsung o Iphone.~~ Esto se debe a que son las dos marcas líderes del mercado en cuando a telefonía celular.

( REVISAR!)

## Cantidad de Ingresos a Venta por Ad de Google



***HAY QUE AGREGAR DESRIPCION***

## Conclusión

## Clasificación de celulares por apariencia

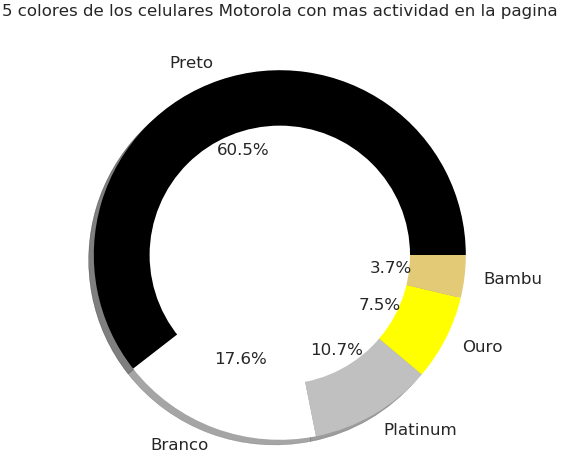
Todos los celulares se fabrican con distintos colores. Aquí se verá los diferentes colores de cada celular que son parte de las operaciones.

## Cinco colores de Celular con más actividad

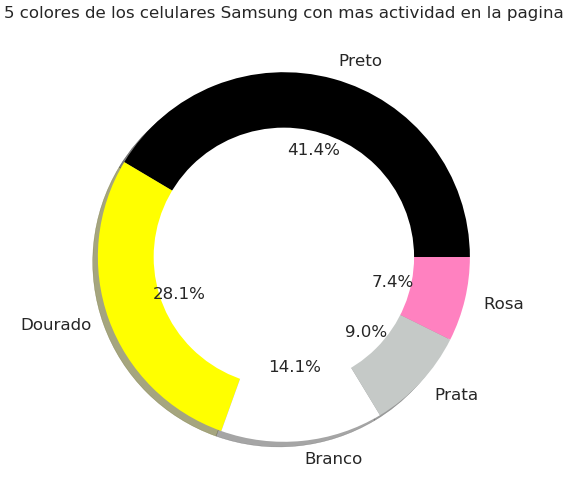
## Primero se observará los 5 colores más activos en la página diferenciándolos por marca. Observar que no se toman en cuenta los colores que no pertenezcan a los 5 mas vendidos:

## 

## Para el caso de Apple, el color que predomina es el gris ya que es el color más básico de la marca, seguido por el dorado y el plateado.



Para el caso de Motorola, el que domina es el negro, ya que es el color base de la marca para sus celulares.



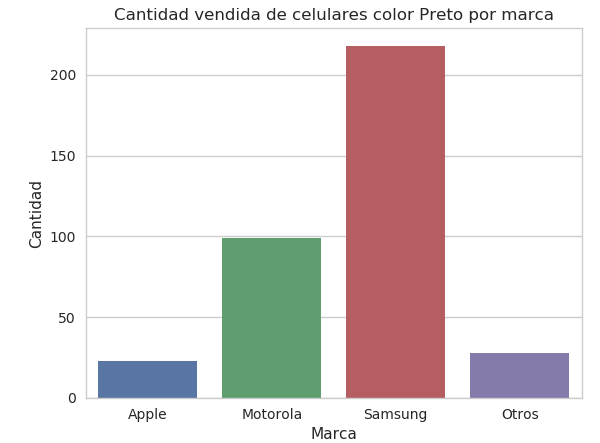
En Samsung, el color con más actividad es el negro, al igual que en Motorola. Pero al tener Samsung muchos más modelos en el mercado hace que el dorado también sea un color fuerte en la actividad de la página.

## Cantidad Vendida Color Preto

## En la siguiente tabla se puede observar los colores más vendidos en el sitio:

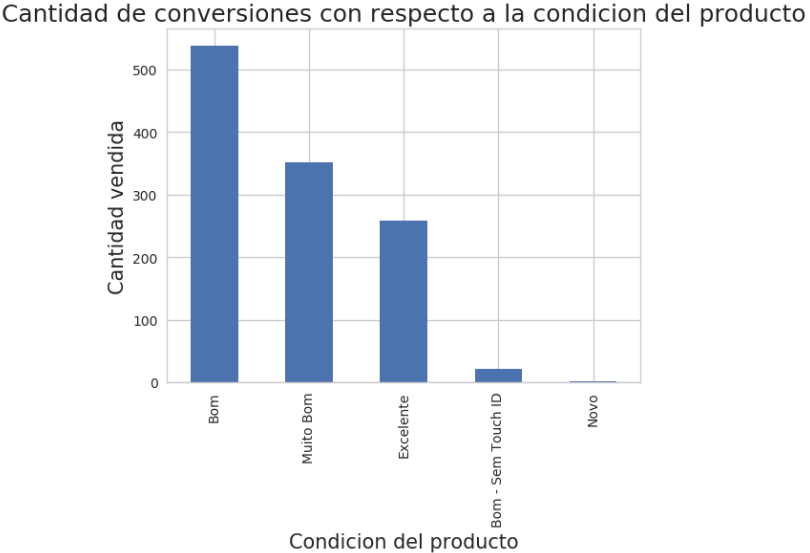
## 

El color más vendido es el negro y los Samsung son los que mayor cantidad vendida tienen, seguido por Motorola, dos de las marcas que tienen como color común el negro.



## Cantidad de Ventas con respecto a la condición del producto

Los celulares aparte del color, al momento de venderse tienen una condición del producto. Según esta clasificación es como se guían los usuarios para realizar la compra.



Se puede apreciar que la mayor cantidad de ventas se realizan para productos en condición buena, donde se entiende, el usuario solo busca comprar un celular en buen estado y que funcione.

## Ventas de los mayores vendedores según condiciones más populares

## Del gráfico anterior podemos concluir que las tres condiciones populares son “Buena”, “Muy Buena” y “Excelente”.

## 

Se usan las tres condiciones populares nombradas anteriormente y en las tres marcas líderes del mercado, se observa que los más vendidos en las estas condiciones son los Samsung seguidos por los Apple.

## Conclusión

Las marcas de celulares tienen colores básicos, colores que usan en la mayoría de sus celulares. Al ser el negro el color usado en la mayoría de celulares hechos por dos líderes del mercado hacen que el color más vendido en el site sea el negro. La marca que más celulares vendió en ese color fue Samsung.

Los celulares también tienen una condición de clasificación. La cantidad más vendida es para la condición de “buena”, lo que nos dice que el usuario busca un celular que funcione y no le da tanta importancia a la condición del producto ya que si está se encuentra en buen estado es condición suficiente para adquirirlo. Para este caso, es también Samsung la marca que más se vende.

## Datos por sku y URL

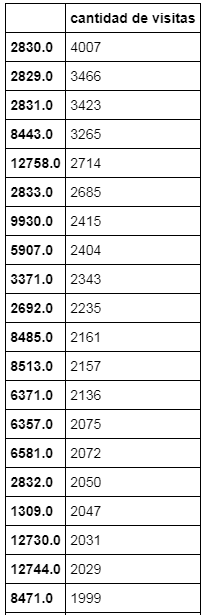
## Cada producto tiene su identificador único visualizado en el evento y una URL asociada a cada producto.

## Sku con más conversiones

## 

Dentro de los quince productos más vendidos, estos son los sku con más conversiones que tiene la página.

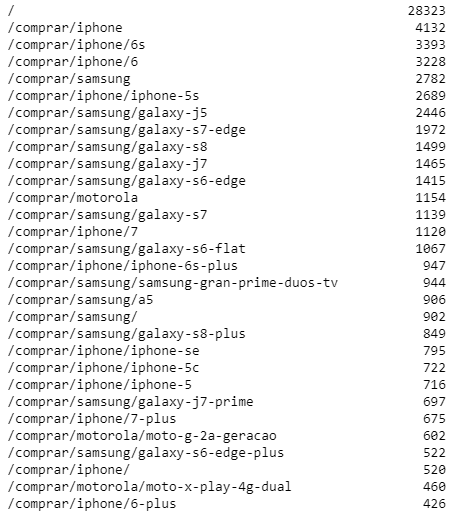
## Sku más visitados



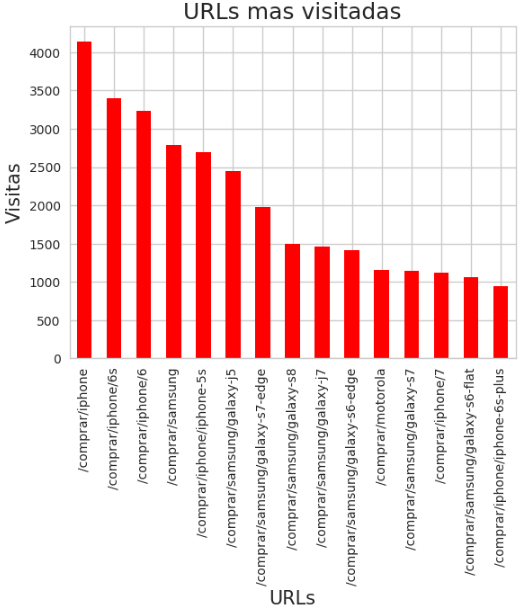
Dentro de todos los productos, los cuales todos tienen un sku diferente dentro del sitio, entre los veinte más visitados, se puede ver que el producto con mayor cantidad de conversiones, no es el que más visitado, sino que recién aparece en la posición número trece.

## URL más visitadas

Cada producto a su vez tiene una URL asociada, por la cual el usuario ingresa a dicho producto.



Como se observa en la lista, la URL más visitada es el home, seguida por URL correspondientes a iPhone.



Filtrando el home del site de Trocafone, se observa que las URLs más visitadas corresponden a Iphone.

## Conclusión

Los productos más visitados no son los más vendidos, por esto entendemos que muchas veces los usuarios visitan los productos y no terminan comprándolo. Esto puede ser porque al entrar al sitio busca la mejor opción o solo por curiosidad de conocer las características de los productos.

En cuanto a las URLs la más visitada es el home, por lo que todos los usuarios siempre pasan por dicha pantalla.

Se filtra de la lista esta URL, y se observa que la URL más visitada es de Iphone, donde también se puede ver que todas las URLs son de celulares viejos, ya que el más nuevo en aparecer en la lista es un Samsung que aparece en la posición número ocho. Esto no puede decir que los usuarios no buscan productos nuevos sino modelos anteriores que aún están activos en el mercado.

## Búsquedas en el sitio

## Los usuarios buscan en el sitio los productos específicos por los cuales quieren informarse

## 

Por la lista podemos ver, al igual que en el punto anterior, los productor más buscado no son los últimos modelos de cada marca sino modelos antiguos pero vigentes en el mercado.

No siempre el usuario busca exacto el producto que quiere ver, hay veces que usa otras palabras o características requeridas para encontrar el producto ideal ante su propia demanda. En el siguiente gráfico observamos las palabras más buscadas por los usuarios en el site.



## Los usuarios también al ingresar al sitio y realizar las búsquedas, puede hacerlo desde cualquier dispositivo con conexión a internet.

## 

## Se puede apreciar, gracias a esta tabla, que las personas ingresan en mayor cantidad de veces a través de celulares y/o computadoras. A su vez, vía estos dos artefactos es donde mayor cantidad de checkouts y conversiones se realizan en el sitio:

## 

## Conclusión

En el sitio los productos más buscados son los modelos antiguos, esto puede ser porque al haber modelos más modernos, estos disminuyen su valor y son más accesibles para cualquier usuario poder adquirirlo.

En las búsquedas, muchas veces las personas no saben exactamente lo querido, entonces buscan según las características requeridas para después elegir la opción que más le convenga a cada uno.

## Conclusión Final

## Samsung es la marca con más unidades vendidas ya que es la que dispone de más cantidad de modelos en el mercado.

## En todas las marcas disponibles en el sitio, siempre la diferencia entre checkouts y conversión es favorable a los checkouts de los productos, por lo que las personas ingresan al checkout de la compra pero no la efectúan.

## Los checkouts son la actividad principal de la página, ya que cuando el tráfico de estos en la página es alto, el tráfico general del sitio también lo es.