

UNIVERSIDAD DE CÓRDOBA

ESCUELA POLITÉCNICA SUPERIOR DE CÓRDOBA

MÁSTER UNIVERSITARIO EN INTELIGENCIA COMPUTACIONAL
E INTERNET DE LAS COSAS

ANALÍTICA WEB (AW)

Cálculo de métricas de un sitio web

Estudiante:

Mabrouka Salmi

< z12salsm@uco.es >



UNIVERSIDAD DE CÓRDOBA

09 de abril de 2024

Índice

1	Introducción	1
2	Cálculo de Métricas del sitio web de Dataiku	1
2.1	Número de visitas	1
2.2	Número de visitantes únicos	1
2.3	Número medio de páginas/visitas	2
2.4	Tasa de rebote	2
2.5	Tasa de salida para cada página	3
2.6	Tráfico directo	3
2.7	Tráfico de búsqueda	4
2.8	Tráfico referido	4
3	Conclusión	4

Listings

1	Cálculo de Número de visitas	1
2	Cálculo de Número de visitantes únicos	1
3	Cálculo de Número medio de páginas/visitas	2
4	Cálculo de Tasa de rebote	2
5	Cálculo de Tasa de salida para cada página	3
6	Cálculo de Tráfico directo	3
7	Cálculo de Tráfico de búsqueda	4
8	Cálculo de Tráfico referido	4

1 Introducción

Esta tarea de Analítica Web tiene como objetivo calcular un conjunto de métricas para el conjunto de datos de registros del sitio web de Dataiku. Explicamos cada métrica y mostramos el código utilizado para calcular cada una y su resultado. Sin embargo, todo el código está disponible en este [Cuaderno de Colab](#).

2 Cálculo de Mátricas del sitio web de Dataiku

2.1 Número de visitas

Esta métrica se calcula contando el número de sesiones únicas en el dataset. Cada sesión está representada por un valor único en la columna "session_id". Se cuenta el número de valores únicos en esta columna para obtener el número total de sesiones. Aquí en el código a continuación 1 calculamos el Número de visitas:

Listing 1: Cálculo de Número de visitas

```
1 # numero de visitas es el total de visitas al sitio web
2 # el calculo: se calcula usando el "utmb" que se guarda usando
  los cookies
3 # Que vamos a usar "session_id": id de la sesión/visita (~
  _utmb)
4 # La operación es count
5 N_visitas = len(np.unique(df['session_id'])) # metodo 1
6 #N_visitas = df['session_id'].nunique() # metodo 2
7 print("Número de visitas (sesiones que ha tenido el sitio web)
  = ", N_visitas)
```

La cantidad de sesiones o visitas al sitio web de Dataiku durante el período dado es de 3946, como se menciona en la salida del código de cálculo que se muestra a continuación.

```
Número de visitas (sesiones que ha tenido el sitio web) = 3946
```

2.2 Número de visitantes únicos

Para calcular esta métrica, se cuenta el número de usuarios únicos que han visitado el sitio web. Cada usuario está representado por un valor único en la columna "visitor_id". Se cuentan los valores únicos en esta columna para obtener el número total de usuarios únicos. El código 2 se utiliza para calcular el número de visitantes únicos.

Listing 2: Cálculo de Número de visitantes únicos

```
1 # numero de visitantes unicos, se fija sobre las visitantes
  unicos
2 # el calculo: se calcula usando el "utma" que se guarda tras
  los cookies
3 # Que vamos a usar -> "visitor_id"
4 # la operación es count
```

```
5 #N_VisitantesUnicos=len(np.unique(df['visitor_id']))
6 N_VisitantesUnicos = df['visitor_id'].nunique()
7 print("Número de usuarios diferentes que han visitado el sitio
web", N_VisitantesUnicos)
```

Como se puede ver en la salida del código a continuación, el número de visitantes únicos es 2537.

```
Número de usuarios diferentes que han visitado el sitio web 2537
```

2.3 Número medio de páginas/visitas

Esta métrica se calcula encontrando el promedio del número de páginas visitadas por sesión. Se agrupa el dataset por sesión utilizando la columna “session_id” y se cuenta el número de páginas visitadas para cada sesión. Luego se calcula la media de estas cuentas para obtener el número medio de páginas por sesión. La siguiente pieza de código 3 es la que utilizamos para calcular el número promedio de páginas visitadas por sesión.

Listing 3: Cálculo de Número medio de páginas/visitas

```
1 # Numero medio de paginas/visitas
2 # para cada session se calcula el numero medio de paginas se
  han visitado
3 # primero calculamos el numero de paginas para cada sesion y
  despues hacemos 'mean'
4 num_paginas_por_visita = df.groupby('session_id')['location'].
  count().mean()
5 print("Numero medio de paginas/visitas:",
  num_paginas_por_visita)
```

Alrededor de 2.75 páginas son visitadas por sesión en el sitio web de Dataiku, según la salida del código a continuación:

```
Número medio de páginas/visitas: 2.749366447034972
```

2.4 Tasa de rebote

La tasa de rebote se define como el porcentaje de sesiones que solo accedieron a una página antes de salir del sitio. Para calcularla, se agrupa el dataset por sesión utilizando la columna “session_id” y se cuenta el número de páginas visitadas para cada sesión. Luego se calcula el porcentaje de sesiones en las que el número de páginas visitadas es igual a 1. La siguiente código 4 se utiliza para calcular la tasa de rebote:

Listing 4: Cálculo de Tasa de rebote

```
1 # Tasa de rebote
2 # para cada session consideramos solo las que solo han tenido
  acceso a una pagina
3 # y despues hacemos el 'mean' que es la tasa de rebote
```

```
4 tasa_rebote = (df.groupby('session_id')['location'].count() ==  
1).mean()  
5 print("Tasa de rebote:", tasa_rebote)
```

La tasa de rebote es del 54.8 %, lo que significa que en el 54.8 % de las visitas/sesiones, el visitante ha visto una sola página.

Tasa de rebote: 0.5486568677141409

2.5 Tasa de salida para cada página

Esta métrica se refiere al porcentaje de veces que cada página del sitio web ha sido la última página visitada en una sesión. Se agrupa el dataset por la URL de la página utilizando la columna “location” y se calcula el porcentaje de veces que la columna “referer” está vacía (“”) para cada página. La línea de código 5 muestra cómo calculamos la tasa de salida para cada página utilizando el conjunto de datos.

Listing 5: Cálculo de Tasa de salida para cada página

```
1 # Tasa de salida para cada p gina  
2 tasa_salida_por_pagina = df.groupby('location')['referer'].  
apply(lambda x: (x == "") | (x.isna())).mean()  
3 print("Tasa de salida para cada pagina", tasa_salida_por_pagina  
)
```

The code output is below; we found that around 11.3 %

Tasa de salida para cada página 0.11300580698681906

2.6 Tráfico directo

Se refiere al número de sesiones que provienen de los usuarios que han escrito la URL del sitio web directamente en la barra de direcciones del navegador. Para calcularlo, se filtran las sesiones donde la columna “referer” está vacía (“”). La siguiente línea de código 6 fue empleada para calcular el tráfico directo al sitio web.

Listing 6: Cálculo de Tráfico directo

```
1 # Trafico directo  
2 # si referer es NAN (null) significa que la visita proviene de  
escribir la url directamente  
3 trafico_directo = df['referer'].isnull().sum()  
4 print("Trafico directo:", trafico_directo)
```

Como se menciona a continuación en la salida del código, el tráfico directo al sitio web de Dataiku es de 1226, lo que significa que el sitio web fue accedido directamente 1226 veces al hacer clic en la URL del sitio web.

Tráfico directo: 1226

2.7 Tráfico de búsqueda

Esta métrica cuenta el número de sesiones que provienen de usuarios que accedieron al sitio web a través de un motor de búsqueda. Se filtran las sesiones donde la columna `referer` contiene la palabra “search” (indicando que el usuario llegó al sitio web a través de una búsqueda). Aquí está el código 7 utilizado para calcular el tráfico de búsqueda.

Listing 7: Cálculo de Tráfico de búsqueda

```
1 # Trafico de busqueda
2 # es la suma de referer que provienen de pinchar en una
  busqueda significa referer que contiene la palabra "search"
3 trafico_busqueda = df['referer'].str.contains("search", na=
  False).sum()
4
5 print("Trafico de busqueda:", trafico_busqueda)
```

In 120 times the website was accessed following a search.

Tráfico de búsqueda: 120

2.8 Tráfico referido

Se refiere al número de sesiones que provienen de usuarios que llegaron al sitio web a través de un enlace desde otra página web. Se filtran las sesiones donde la columna “referer” no está vacía (“”) ni contiene la palabra “search”, lo que indica que el tráfico no es directo ni de búsqueda. La siguiente código 8 muestra el cálculo del tráfico referido del conjunto de datos de registros web.

Listing 8: Cálculo de Tráfico referido

```
1 # Trafico referido
2 # total trafico menos trafico directo y trafico de busqueda
3 referral_traffic = df.shape[0] - trafico_busqueda -
  trafico_directo
4
5 print("Trafico referido:", referral_traffic)
```

La página web fue referida en 9503 ocasiones, como se muestra en la salida a continuación:

Tráfico referido: 9503

3 Conclusión

Hemos calculado métricas en el conjunto de datos de registros de Dataiku en Python. Hemos utilizado agrupaciones y otros métodos cuando ha sido necesario, obtenido los resultados para cada métrica e interpretado los resultados. Sin embargo, todavía se necesita más contexto para saber si los valores de las métricas son buenos o malos, como estudios de referencia y objetivos del sitio web.