# DATA SEO

E-Book



Del SEO tradicional al futuro: la importancia de integrar el análisis de datos en tus estrategias.

Alan Preiti Antoine Eripret

# Introducción

### Motivación para este e-book

- ¿Por qué este e-book es necesario?
- Nuestra experiencia

### Panorama actual

- Identificando desafíos en el uso de data
- Limitaciones actuales

# Objetivos y alcance del e-book

# Desafíos en el Uso de Data

### Limitaciones técnicas

- Uso ineficiente de scripts
- Uso de múltiples plataformas o herramientas
- Análisis efectivo de la competencia

### Optimización de procesos y recursos

- Liderazgo en la era de la data
- Superación de la dependencia de otros equipos (BI/IT)
- Gestión de grandes volúmenes de datos

### Escalabilidad y automatización

- Repetición de tareas
- Crecimiento y escalabilidad de servicios

# Oportunidades con Data en la industria de SEO

¿Necesito Python para hacer un buen SEO?

**Diversificando la carrera SEO** 

**Oportunidades Profesionales** 

# Casos de Uso [Estructura]

### Cómo resolver un problema con datos.

- Una explicación general de la problemática.
- Punto de dolor al que nos enfrentamos.
- Resolución de la problemática anterior.
- Aplicación de análisis de datos con Python.



# Introducción

# Motivación para este e-book

¡Bienvenidos al maravilloso mundo de Data SEO! En este eBook, nos enfocaremos en concientizar a los especialistas de la industria sobre la importancia del análisis de datos y cómo puede ser la clave para resolver los puntos de dolor más comunes que enfrentamos en nuestra vida diaria.

Desde el año 2018, empleamos SQL y Python para cuestiones de SEO. Aunque inicialmente enfrentamos dificultades, llegamos a entender cómo estos dos lenguajes, cuando se utilizan adecuadamente, pueden agilizar y mejorar nuestro trabajo. Aunque Excel sigue siendo parte de nuestra rutina diaria, nuestros insights más valiosos ya no provienen de esta herramienta.

Entrevistamos a más de cien especialistas en SEO de diversos campos (tecnología, contenido, gestión, entre otros) con el fin de recopilar sus opiniones y abordar los problemas presentes en el mercado, y decidimos compartir nuestros conocimientos y experiencias por varias razones.

En primer lugar, nuestro propósito es ayudar a otros individuos a evolucionar como profesionales, proporcionándoles las herramientas y el conocimiento esenciales para destacarse en nuestra área. En segundo lugar, construir un puente sólido entre los datos y el SEO. Creemos que la combinación de ambos es una fórmula exitosa. Finalmente, en un mundo cada vez más complejo y saturado de información, es fundamental contar con espacios donde los datos sean sintetizados y depurados para su correcta aplicación. Esperamos aportarles el valor necesario para que superen el miedo y se sumerjan de lleno en el apasionante mundo de los datos.

# Panorama actual

Existe una brecha en la industria entre la percepción de la importancia de los datos y la realidad. ¿Por qué? Porque la mayoría de las personas carece de los conocimientos necesarios para extraer ideas valiosas a partir de los datos que tenemos a nuestra disposición.

A medida que el valor de la programación en nuestro campo comienza a ser reconocido, surge una concepción errónea acerca de lo que realmente importa: si bien crear un script puede parecer sencillo, utilizar esta destreza para generar ideas impactantes en un proyecto es mucho más desafiante. Y aquí es donde radica la diferencia entre una habilidad común y una habilidad altamente remunerada.

# Objetivos y alcance del e-book

Esta primera versión se centrará en enumerar los diversos desafíos de la industria y proponer soluciones a través del análisis de datos. No nos adentraremos en la complejidad de códigos de programación complejos. En su lugar, nuestro enfoque será proporcionarles las herramientas y conceptos fundamentales que necesitarán para sacar el máximo provecho del análisis de datos en sus estrategias de SEO.

Aunque el potencial para SEO es enorme, la falta de recursos específicos ha impedido que muchas personas den el salto, a menudo debido al temor de aprender algo que perciben como poco útil. Nuestro objetivo es que se sientan capacitados y confiados para enfrentar estos desafíos, incluso si carecen de experiencia previa en programación.

# ¿Para quién es este e-book?

Para todas aquellas personas que están buscando adentrarse en el mundo de la Data aplicada a SEO, sin importar si cuentan con conocimientos previos en análisis de datos o en algún lenguaje de programación, y se sientan identificados con algunos de los puntos de dolor que describiremos más adelante.

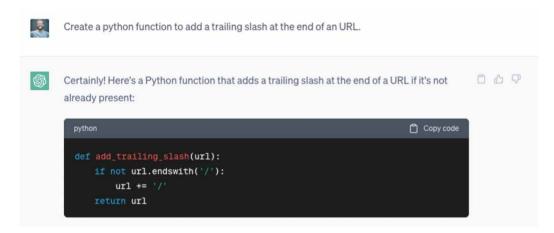


# Desafíos en el Uso de Data

# Limitaciones técnicas

# Uso ineficiente de scripts

El problema surge cuando buscamos soluciones rápidas sin comprender plenamente el proceso de aprendizaje. En lugar de buscar personas con experiencia y resultados comprobados para entender su camino, optamos por el enfoque de "copiar y pegar" con el fin de economizar tiempo. En la era actual, con el progreso de la IA, es seductor confiar en soluciones predefinidas sin adquirir el conocimiento necesario ya que la gran mayoría de las operaciones simples pueden ser resueltas sin dificultad. Por ejemplo:



No obstante, si nuestra única habilidad es copiar un script sin comprender su verdadero funcionamiento, podríamos enfrentar los siguientes problemas:

- Limitar nuestra capacidad de escalar soluciones.
- No ser capaces de resolver problemas cuando surgen o de replicar estrategias exitosas.
- Implementar algo que no sabemos manejar.

Es vital comprender que el código es una herramienta con un propósito específico y que debe ajustarse según los problemas y desafíos que se presenten en el negocio.

Una calculadora por sí sola carece de valor.

### Solución

La solución radica en adquirir los conocimientos necesarios en análisis de datos, sin necesidad de ser un programador experto, pero con la capacidad de comprender y comunicarse de manera efectiva en el lenguaje del análisis de datos. Al hacerlo, evitaremos futuros problemas al contar con una comunicación fluida y una comprensión clara de los desafíos que afrontamos. Aprender los fundamentos del análisis de datos nos permitirá utilizar el código como una herramienta potente y adaptarla a nuestras necesidades comerciales, en lugar de depender únicamente de soluciones predefinidas.



# Uso de múltiples plataformas o herramientas

### Gestión Unificada

Como bien sabemos, una gran parte de nuestro trabajo como SEO implica el análisis de datos. Sin embargo, estos datos pueden provenir de diversas fuentes, como: Google Analytics, Google Search Console o Bing Webmaster Tools, Screaming Frog o Oncrawl, Ahrefs o Semrush, entre otras herramientas. Cuando deseamos realizar un análisis basándonos en varias fuentes, es posible que perdamos tiempo en la extracción y la fusión de los datos. Aunque existen extensiones como *Search Analytics* que facilitan esta tarea, esto no siempre es aplicable.

### Solución

En su mayoría, estas herramientas cuentan con una API o la opción de exportar datos en formato CSV. En ambos casos, podemos integrar la información de diversas fuentes, procesarla, manipularla y, posteriormente, generar resultados con tan solo unas pocas líneas de código.

### Ventajas:

- Velocidad del proceso una vez esté bien diseñado
- Escalabilidad: podemos agregar nuevas fuentes sin problema

### **Costos**

Si nuestros competidores están utilizando herramientas que nosotros no tenemos para simplificar sus tareas diarias, nos encontramos en una desventaja significativa. El mercado actual está saturado de herramientas que buscan abordar y resolver problemáticas similares. Además, muchas de estas herramientas tienen un costo elevado que no siempre podemos asumir, ya sea porque somos autónomos o porque nuestros clientes no cuentan con el presupuesto necesario.

Lo más frustrante es cuando utilizamos herramientas de terceros para medir nuestros propios datos debido a que no sabemos cómo acceder a ellos. Sin embargo, mediante el uso de técnicas como el web scraping, la API de Search Console y el procesamiento de lenguaje natural (NLP), podemos obtener los datos requeridos sin depender por completo de herramientas externas.

Esta situación resulta beneficiosa tanto para nuestros clientes como para nosotros, ya que podemos ofrecer servicios más accesibles y eficientes sin sacrificar la calidad de nuestros resultados. Al tomar decisiones sobre la inversión en herramientas, podemos evaluar el presupuesto disponible y determinar si la solución proporcionada por Python es suficiente para satisfacer nuestras necesidades.

Es importante aclarar que la elección de pagar por una herramienta dependerá del presupuesto disponible, la solución que brinda y su potencial para escalar. Es decir, hay situaciones en las cuales puede ser conveniente invertir en estas herramientas.

# Análisis efectivo de la competencia

Cuando se trata de posicionar un sitio web, no podemos pasar por alto las acciones de nuestros competidores. Ya sea para un e-commerce o un blog, la estrategia implica identificar qué hacen, cómo lo hacen y por qué lo hacen. Esto es especialmente importante si sus resultados no son óptimos (nuestro objetivo es evitar replicar sus fallas).

Sin embargo, realizar un análisis de esta naturaleza puede ser un proceso laborioso si carecemos de las herramientas adecuadas. De hecho, muchos especialistas descartan realizar un análisis exhaustivo precisamente por este motivo.



Algunos ejemplos de análisis incluyen:

Analizar el sitemap de nuestros competidores para clasificar y evaluar la frecuencia de su contenido a lo largo del tiempo.

Obtener los precios de los productos para las palabras clave en las que mejor nos posicionamos.

### Solución

Para abordar este problema, empleamos la técnica de web scraping. Esto nos permite trabajar a gran escala y automatizar una parte sustancial del proceso. Por ejemplo, podemos:

Recuperar las URL de los contenidos del blog de un competidor utilizando su sitemap.

Analizar las fechas de publicación y obtener conclusiones sobre la frecuencia de sus publicaciones/actualizaciones, lo que nos brinda un contexto más claro sobre su estrategia.

100	Traffic	
		lastmod
2.0%	0.0%	2017
4.8%	0.4%	2018
3.6%	1.2%	2019
45.5%	1.5%	2020
6.7%	7.8%	2021
28.9%	75.8%	2022
8.4%	13.3%	2023

# Optimización de procesos y recursos

# Liderazgo en la era de la data

Un desafío común que a menudo enfrentan los managers y fundadores de agencias de SEO es la falta de conocimiento en análisis de datos. Esta limitación puede convertirse en un obstáculo significativo, ya que conlleva una faltma de decisiones estratégicas acertadas.

¿Cómo podría pedirle a un colaborador que almacene los datos históricos de Search Console para evitar perderlos en 16 meses, si no tengo conocimiento de la existencia de una API que permite hacerlo?

### Solución

La solución dependerá de cada situación particular y del costo de oportunidad asociado al tiempo invertido en la formación. Es decir, en ciertos casos, un fundador podría delegar la tarea a un head SEO para que adquiera los conocimientos mínimos y necesarios. Por otro lado, un líder o manager debería estar dispuesto y capacitado para invertir tiempo en adquirir estos conocimientos, especialmente considerando que su relación con los analistas es mucho más cercana. Tenemos certeza de que dentro de las organizaciones debe existir un rol que asuma el liderazgo para coordinar mejor los equipos y ofrecer un valor adicional a los clientes.

# Superación de la dependencia de otros equipos (BI/IT)

Depender de otras personas o equipos puede ralentizar los procesos debido a la burocracia y ser un verdadero problema. Solicitar análisis específicos al equipo de BI/IT debido a la falta de acceso a bases de datos o a la incapacidad para crear análisis y gráficos, genera dependencia, pérdida de tiempo e ineficiencia. Es curioso que en la mayoría de los casos, estas solicitudes podrían resolverse en cuestión de minutos. Esta dinámica es la que finalmente puede causar tensiones entre los diferentes miembros de un proyecto.



Todas las partes del negocio resultan afectadas. Por ejemplo:

**Analistas SEO:** Se ven limitados en la construcción de informes y la implementación de estrategias.

**Equipo de BI:** Reciben solicitudes constantes y deben asignar tiempo a tareas que consideran poco relevantes.

**Founder:** Se generan problemas entre los equipos internos y se afecta la entrega del servicio al cliente.

Clientes: Experimentan resultados más lentos debido a lo mencionado anteriormente.

### Solución

Una solución a este problema es que los profesionales de SEO adquieran conocimientos en análisis de datos a través del uso de herramientas específicas como Python o SQL. Esto les permitirá ganar velocidad e independencia en la resolución de problemas diarios, liberando la carga de trabajo de los equipos de BI, añadiendo un mayor valor al negocio y, por lo tanto, obteniendo resultados más rápidos para los clientes.

# Gestión de grandes volúmenes de datos

Vivimos en un mundo en el que los datos están ampliamente disponibles. En el ámbito de SEO, podemos fácilmente obtener datos de rastreo (propios o logs), de posicionamiento (GSC o extrayendo datos de las SERP) y de resultados (GSC, GA y propios), por mencionar solo algunos ejemplos evidentes.

Frente a esta abundancia, surgen varios desafíos con los que debemos lidiar:

**Parálisis por análisis:** Debemos ser capaces de gestionar esta cantidad de datos para tomar decisiones relevantes y evitar sentirnos abrumados por ellos.

**Gestión de los datos:** Las herramientas como Excel / Sheets tienen limitaciones que nos impide trabajar únicamente con ellas. Abrir archivos con más de 100.000 líneas puede ralentizar significativamente nuestro trabajo.

# Escalabilidad y automatización

# Repetición de tareas

Aunque nos guste o no, cualquier trabajo conlleva una serie de tareas repetitivas. En ocasiones, no somos conscientes de que realizar el mismo informe o llevar a cabo tareas monótonas no solo es aburrido, sino que también agota nuestros recursos cognitivos y consume tiempo.

En el ámbito de una agencia, por ejemplo, una parte de muchos procesos (como un keyword research) se puede automatizar para garantizar un nivel uniforme de calidad para todos los clientes. ¿Todas las tareas se deben automatizar? No necesariamente, pero si logras identificar cuáles sí pueden ser automatizadas, podrías disponer de más tiempo para hacer lo que quieras, ya sea pasar tiempo con tu familia o asumir más proyectos.

Un ejemplo sencillo: si tu equipo invierte cuatro horas cada mes extrayendo datos de diversas fuentes para crear un informe de seguimiento de resultados, y el costo asociado es de 50€ por hora, automatizar esta tarea representaría un ahorro de 2400€ al año.



### Solución

El gran desafío de la automatización es entender cuándo resulta beneficioso y cuándo no lo es. Si utilizamos el ejemplo del keyword research, no es posible automatizar todo el proceso dado que esto implica comprender el mercado, el negocio, y no se limita a elegir palabras clave únicamente basadas en el volumen de búsqueda. Sin embargo, sí podemos automatizar la agrupación de palabras clave por intención de búsqueda, la extracción de tendencias y el posicionamiento actual del cliente.

# Crecimiento y escalabilidad de servicios

Uno de los desafíos tanto para agencias como para profesionales autónomos o freelance es expandir su negocio y lograr una economía de escala. Si vendes únicamente tus horas, te enfrentas a un límite de ingresos que está determinado por las horas disponibles y el precio por hora. Por ejemplo, si tienes 20 horas productivas a la semana (excluyendo tareas administrativas, por ejemplo) y cobras €50 por hora, tus ingresos mensuales no podrían superar los 4000 €. Aquí solo estamos hablando de ingresos, sin considerar beneficios.

Este modelo es sostenible hasta cierto punto, pero tener conocimientos en análisis de datos te permitirá trabajar de manera más eficiente, lo que se traduce en hacer menos trabajo activo para ganar lo mismo o incluso más. La decisión dependerá de tus objetivos personales.

Como analista de datos con conocimientos en SEO, aportas un valor agregado en proyectos que demandan estas habilidades, y los clientes pueden estar dispuestos a pagar más por tu experiencia. Trabajar en un proyecto con millones de URLs no es lo mismo que posicionar un nicho con 10 páginas, ya que requiere habilidades distintas.

### Solución

Existen diversas opciones que puedes elegir:

**Especialización:** Como analista de datos con conocimientos de SEO aportarás un valor agregado a proyectos que requieren estas habilidades y los clientes estarán estarán dispuestos a pagar más por tu experiencia. Trabajar en un proyecto con millones de URLs no es lo mismo que posicionar un nicho con 10 páginas, ya que requiere habilidades distintas.

**Automatización:** Si logras identificar y automatizar las partes de tu proceso que consumen más tiempo, podrás proporcionar el mismo valor a tus clientes en menos tiempo.

**Productivización:** Existen ejemplos de servicios que han evolucionado hacia productos, como las soluciones en formato SaaS (Software como Servicio), lo que puede abrir nuevas oportunidades de ingresos y escalabilidad.



# Oportunidades con Data en la industria de SEO

# ¿Necesito Python para hacer un buen SEO?

El SEO, como disciplina, requiere de una variedad de habilidades. De hecho, hay especialistas que optan por enfocarse en algunas ramas específicas, como por ejemplo, el SEO técnico. Sin embargo, existen numerosos casos en los que se realiza un buen trabajo sin un conocimiento profundo en ciertas áreas, como la programación.

Si estás recién comenzando en el campo del SEO, nuestra sugerencia es que adquieras conocimientos generales para comprender cómo funcionan los motores de búsqueda. No sirve de nada saber HTML, CSS, Python o SQL si no hay un marco de referencia para su aplicación.

Entonces, ¿por qué existe un debate entre quienes consideran esencial saber programar y analizar datos y aquellos que opinan que no es necesario? Creemos que tiene que ver con el tamaño del proyecto. Si trabajas con un e-commerce pequeño como <a href="https://www.vitaminajota.com/">https://www.vitaminajota.com/</a>, es posible que consideres que no es imprescindible. Sin embargo, si estás involucrado con un sitio como <a href="https://www.amazon.es/">https://www.amazon.es/</a>, sabes que es una necesidad debido a la enorme cantidad de datos.

De acuerdo con el experto Andrew Ng en su charla TED sobre IA, esta perspectiva podría estar equivocada. No es necesario contar con millones de datos para proporcionar valor a un negocio pequeño. Citando sus palabras: "Para el dueño de una pequeña pizzería, tener algo que le permita aumentar sus ingresos en miles de dólares al año, definitivamente valdría la pena". Traduciendo esto al ámbito SEO, cada acción que tomemos (positiva o negativa) tiene un impacto significativo en su negocio sin importar el tamaño del cliente.

Imagina por un momento que el canal de SEO representa el 30% de las ventas de Vitaminajota. Si por alguna razón se pierden todos los datos históricos de Search Console, esto tendría consecuencias graves en nuestra estrategia y, por ende, en el negocio del cliente.

Nuestra conclusión es que la adopción de herramientas que minimicen el impacto de las consecuencias negativas o maximicen las positivas no depende del tamaño o sector de la industria.

# **Diversificando la carrera SEO**

¿Alguna vez has considerado qué sucedería si en el futuro te sientes cansado de dedicarte al SEO? Es una posibilidad completamente normal, especialmente en un mundo tan cambiante como el actual. La vida no sigue una trayectoria lineal y es posible que en algún momento te aburras o desees cambiar de dirección.

La pregunta sería: ¿con los conocimientos que ya poseo, es posible cambiar hacia otras industrias? Es posible que, si solo te has especializado en SEO, las opciones sean limitadas.



Sin embargo, la buena noticia es que si tienes habilidades en Python y análisis de datos, cuentas con una herramienta sumamente poderosa que te permite incursionar en prácticamente cualquier otra disciplina. Podrías trabajar en un banco, en el ámbito farmacéutico, en clubes de fútbol o en cualquier otro campo que te interese.

De hecho, cada sector tiene sus propias sub-disciplinas. Podrías trabajar en supply-chain para mejorar el rendimiento de atletas. Este es otro motivo importante por el cual aprender a analizar datos es tan valioso. Te brinda seguridad y libertad para explorar y cambiar de industrias.

# **Oportunidades Profesionales**

Es posible que te haya sucedido: encuentras una oferta de trabajo para un puesto de especialista SEO en una empresa importante y reconocida. Revisas la lista de requisitos y te das cuenta de que cumples con todos, excepto uno: "Se requieren conocimientos en Python, SQL, R o análisis de datos". Las grandes empresas no tienen alternativa y necesitan perfiles con estas habilidades. Incluso en ocasiones están dispuestas a sacrificar experiencia en SEO con tal de tener a alguien capaz de manejar grandes volúmenes de información.

Esto nos enfrenta a una brecha, una oportunidad en la que podrías posicionarte con confianza. Alguien podría argumentar: "Eso solo ocurre en las grandes empresas, las medianas y pequeñas no lo requieren". Sin embargo, ya discutimos en la sección anterior cómo el análisis de datos aporta valor en cualquier tipo de negocio, sin importar su tamaño. Más temprano que tarde, esta habilidad será un requisito. Hace una década, los negocios pequeños y medianos no se involucraban en SEO como lo hacen ahora. Pero esa tendencia ha cambiado. Y si por algún motivo crees que esta tendencia no evolucionará, considera estas dos preguntas:

¿Cuánto valor adicional puedo aportar a una agencia mediana o pequeña si combino mi conocimiento en Python y análisis de datos con mis estrategias de SEO?

¿Estoy seguro de que en pocos años no será un requisito saber analizar datos para trabajar en una agencia, y que podré seguir compitiendo en el mercado laboral?

¡Tarea para reflexionar!



# Casos de Uso [Estructura]

A continuación expondremos 2 casos de uso con la siguiente dinámica:

**Contexto:** Sobre qué industria hacemos referencia y una explicación general de la problemática.

Problema: Punto de dolor al que nos enfrentamos.

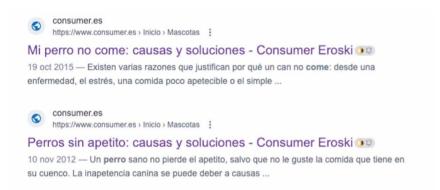
Solución: Resolución de la problemática anterior.

**Ejemplo:** Aplicación de análisis de datos con Python.

# Número 1

### Contexto

La canibalización de palabras clave es una situación en la que un mismo proyecto presenta múltiples contenidos que compiten por la misma intención de búsqueda. **Para una mejor comprensión, te proporcionamos el siguiente ejemplo:** 



Como se puede apreciar, aunque los contenidos tengan títulos y enfoques distintos, ambos intentan abordar la inquietud de un propietario preocupado porque su perro no está comiendo. En este escenario, es poco probable que ambos contenidos puedan lograr un buen posicionamiento, por lo que sería más beneficioso fusionarlos.

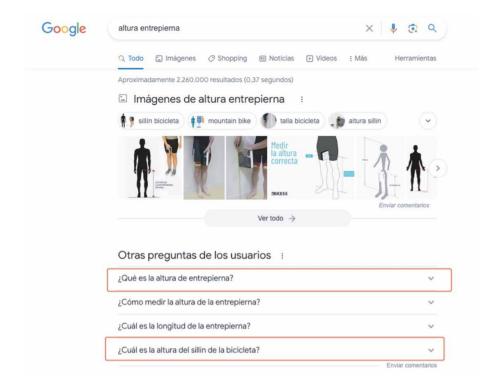
### **Problema**

Tener múltiples contenidos que compiten por una misma palabra clave no siempre es negativo, especialmente si la intención de búsqueda está fragmentada. Por ejemplo:

"alquiler casa Barcelona": Esta palabra clave tiene una intención de búsqueda bastante clara: los usuarios buscan una lista de ofertas para alquilar una casa. Si tienes varias páginas que apuntan a esta misma intención, sería recomendable unificarlas.

"altura entrepierna": En este caso, la intención de búsqueda es menos evidente. Los usuarios podrían estar buscando una explicación sobre cómo se mide la altura entrepierna o cómo se utiliza para ajustar el sillín de una bicicleta. La página de resultados de búsqueda (SERP) refleja esta dualidad de intenciones.





En algunos casos, tener múltiples contenidos que aborden distintos aspectos de una misma palabra clave puede ser útil para capturar una variedad de intenciones de búsqueda y brindar información más completa a los usuarios. La canibalización de palabras clave no es necesariamente un problema si la intención de búsqueda es diferente o fragmentada. Sin embargo, esta situación no es la norma. A menudo, una canibalización puede ser perjudicial.

### Solución

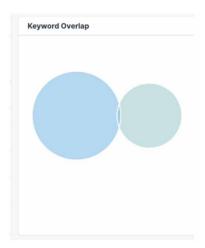
Para esta tarea, nuestro objetivo es identificar los contenidos que están compitiendo entre sí. Queremos crear una tabla con posibles casos que luego revisaremos manualmente. Aunque no podemos automatizar todo el proceso, podemos agilizarlo significativamente siguiendo estos pasos:

- 1. Accede al "Keyword Gap" en SEMRUSH (o su equivalente en otras herramientas).
- 2. Ingresa las URL que deseas analizar. Incluso puedes agregar más de dos URLs si es necesario.





3 El resultado mostrará si las URLs están compitiendo entre sí.



Sin embargo, este proceso presenta algunas limitaciones:

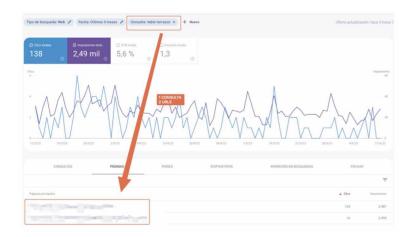
Es un proceso manual y efectivo únicamente si tenemos unas pocas URLs y sabemos qué comparar. ¿Qué ocurre si tenemos miles de URLs?

Proporciona una visión segmentada que no considera la dinámica de Google. Por ejemplo, SEMRUSH podría indicar que dos URLs no compiten, pero en la realidad, Google podría alternar sus posiciones. Una URL podría dominar la palabra clave principal un día y la otra al día siguiente, lo cual no se reflejaría en SEMRUSH.

No aprovechamos nuestros propios datos. Para proyectos con un flujo de tráfico constante, Google Search Console (GSC) o Bing Webmaster Tools (BWT) son fuentes más completas en comparación con las bases de datos de herramientas externas.

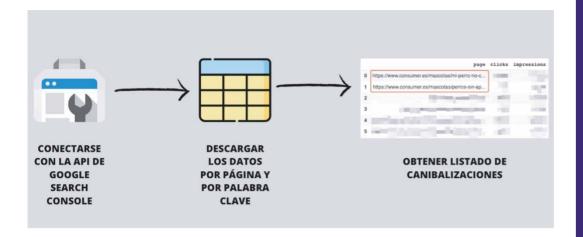
# Ejemplo

Cuando empleamos la interfaz de Google Search Console (GSC), es posible identificar un posible caso de canibalización al filtrar el informe por una palabra clave específica y observar si aparecen más de una URL en la sección "Páginas". Por ejemplo:





Podemos replicar este análisis manual a un nivel más amplio utilizando nuestros propios datos: buscamos palabras clave para las cuales existen más de una URL compitiendo. Podemos ejecutar este proceso de manera repetitiva mediante una función sencilla. En resumen, el proceso que estamos tratando de agilizar consta de varios pasos:



La gran ventaja de este proceso es que es escalable y puede aplicarse a diferentes proyectos sin necesidad de ajustes significativos. Una vez que has desarrollado y afinado este proceso, puedes usarlo de manera consistente en múltiples proyectos donde la identificación de canibalización de palabras clave sea relevante.

### **Pasos**

1- Inicialmente, nos conectamos a la API de Google Search Console, empleando una librería que facilita el proceso.

```
#cargamos pandas
import pandas as pd

#cargamos la librería que nos permite conectarnos a la API de
GSC
import searchconsole

#cargamos los datos

df = get_gsc_data('<https://www.midominio.es/>')
```



2-Por supuesto, es necesario añadir más código para definir la función get\_gsc\_data, pero no debería resultar intimidante si ya estás familiarizado con la interfaz de Google Search Console. Una vez que hayas creado esta función, podrás utilizarla en múltiples proyectos sin preocuparte, ya que solo necesitas escribirla una vez.

```
Python
def get_gsc_data(property):
     #nos conectamos a la API de GSC
     account = (
   searchconsole
    .authenticate(
     flow="console",
     client_config='client-secret.json',
     credentials='credentials.json'
return (
   #datos de la propiedad que nos interesa
   account[property]
   .query
   #fechas
    .range('2023-01-01', 'today')
   #dimensiones
    .dimension('page', 'query')
    .get()
   #convertir en tabla para pandas
    .to_dataframe()
```



3- Obtendremos una tabla de datos similar a la exhibida a continuación. Podrá contener muchas o pocas líneas de acuerdo al proyecto que estés analizando.

	query	page	clicks	impressions
0	The second second		54335	322304
1			21620	174897
2	The second secon		16848	48760
3			13566	239086
4	Control of the later of the lat	-	8365	91486

4- Continuamos filtrando esta tabla para incluir únicamente aquellas palabras clave que han generado al menos una impresión en dos o más URLs. En otras palabras, aplicamos una definición estricta de canibalización.

```
#cargamos pandas
import pandas as pd
#cargamos la librería que nos permite conectarnos a la API de
GSC
import searchconsole

#cargamos los datos
df = get_gsc_data('<https://www.midominio.es/>')
#buscamos casos de canibalización
**df = find_cannibalized_queries(df)**
```

Aquí proporcionamos el código que se usa detrás de la función find\_cannibalized\_queries. Cabe destacar que se trata de una tabla dinámica con algunas modificaciones adicionales.



5- En este punto del proceso, contamos con una tabla similar a la previa, pero esta vez está filtrada y muestra el número de URLs que están posicionadas por palabra clave.



Continuamos con el razonamiento: buscamos identificar páginas que comparten un número significativo de impresiones de las mismas palabras clave entre ellas. Sin embargo, este proceso no es perfecto, ya que en el caso de que una URL aparezca con enlaces de sitio, podría entrar dentro de nuestra definición de canibalización. Somos conscientes de esta limitación, por lo que no es posible automatizar este proceso al 100%.





En nuestra experiencia, al eliminar cualquier URL que no comparta al menos el 5% de las impresiones, logramos eliminar la mayoría de los resultados falsos positivos. Es importante tener en cuenta que el porcentaje se ajustará según cada proyecto. Esta es la razón por la que no se puede automatizar completamente todo el proceso: aún es necesario realizar ajustes específicos.

6- Aquí el código completo para ejecutar. La función remove\_false\_positive podría considerarse como la sección más intrincada. No te preocupes si no lo comprendes completamente, incluso con los comentarios; el propósito es que eventualmente puedas llegar a comprenderlo.

```
#cargamos pandas
import pandas as pd

#cargamos la librería que nos permite conectarnos a la API de
GSC
import searchconsole

#cargamos los datos

df = get_gsc_data('<https://www.midominio.es/>')

#buscamos casos de canibalización

df = find_cannibalized_queries(df)

#quitamos los falsos positivos

**df = remove_false_positive(df)**
```



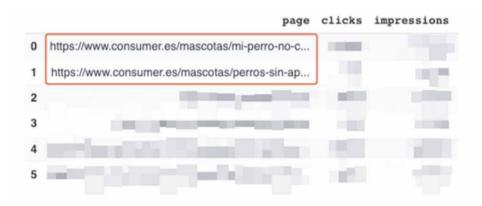
```
#ordenado Z-A por clicks
     #quitamos las palabras cuya suma acumulado es superior
al 75% para eliminar
     #palabras clave sin importancia
     .query('cumsum < 0.75')
     ['query']
   #candidatos
   pages = (
     #únicamente si la palabra está en el DF que hicimos
antes
     df[df['query'].isin(queries)]
     #quitamos columnas inútiles
     .drop('n_pages', axis=1)
     #agrupamos por página
     .groupby('page').sum()
     #ordenamos Z-A
      .sort_values(by='clicks', ascending=False)
     #asignamos un % del total a cada URL
     .assign(imp=lambda df_:100*df_.clicks/df_.clicks.sum())
     #quitamos cualquier URL que tenga menos de 5% de las
impresions
     .query('imp > 4')
```



```
.assign(id=i)
)

#seguimos con el bucle
i+=1
x = pd.concat([x,pages])
return result
```

7- En este punto, habríamos obtenido lo que buscábamos: una lista de URLs que podrían estar compitiendo entre sí (aunque no necesariamente). Esta lista requerirá ser revisada manualmente, pero nos ahorra una cantidad significativa de tiempo.



# Número 2

# Contexto

Trabajamos en un proyecto de e-commerce y hemos detectado una marcada disminución en el tráfico durante los últimos meses. Nuestro objetivo es identificar las categorías y subcategorías que presentan un elevado número de impresiones pero un bajo nivel de clics, con el propósito de optimizar su tasa de clics (CTR).

## **Problema**

Nos enfrentamos a dos problemas claramente delineados:

# 1-Limitación de la disponibilidad de datos

La limitación inherente en la interfaz de Search Console nos restringe a solamente 1,000 registros. Esto resulta insuficiente para nuestro análisis, ya que necesitamos un acceso más detallado a la información para poder generar nuevas columnas basadas en las reglas de negocio.

Si deseas saber más sobre todas las diferencias entre la interfaz normal (UI) y la API de Search Console, visita el siguiente artículo: <a href="https://www.aeripret.com/es/datos-gsc/">https://www.aeripret.com/es/datos-gsc/</a>



### 2- Impacto económico de la ejecución de la estrategia

Nuestro sitio cuenta con 7 verticales, 34 categorías y más de 100 subcategorías. Realizar una validación manual de su rendimiento requeriría semanas de trabajo, lo cual se traduciría en horas y, en última instancia, en una pérdida económica. Se trata de un costo de oportunidad, ya que podríamos estar empleando ese tiempo en tareas de mayor valor.

Adicionalmente, si necesitamos repetir esta estrategia para el mismo cliente (u otros), nos enfrentaríamos nuevamente a estos mismos costos. Por esa razón buscamos automatizar la mayor parte del proceso con el objetivo de reducir una tarea que normalmente tomaría semanas a apenas minutos.



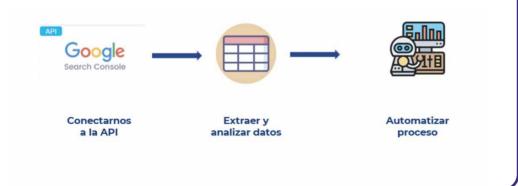
### Solución

Utilizaremos Python como un medio para:

Acceder a la API de Search Console y gestionar grandes volúmenes de datos en cuestión de segundo ((G-sheet, te apreciamos, pero aquí no tienes lugar).

Aplicar múltiples filtros para identificar las áreas de enfoque necesarias.

Automatizar todos los aspectos del proceso que impliquen tareas repetitivas (por ejemplo, programación constante) y así ganar tiempo para los análisis que llevaremos a cabo en el futuro.





# Ejemplo

En el ejemplo anterior ya explicamos como conectarnos a Search Console por lo que no creemos neceasrio volver a hacerlo. De todas formas si le quedó alguna duda, hay un curso gratuito en la comunidad donde explico cómo conectarse a Search Console y realizar múltiples análisis.

El objetivo de cualquier analista de datos debe ser aprender a resolver problemas mediante el razonamiento (no copiar y pegar scripts). Si no estás acostumbrado a trabajar con datos, quizás al comienzo te resulte un desafío (estás saliendo de tu zona de confort). La buena noticia es que la forma de abordar problemas ha sido la misma durante 2.000 años, lo que significa que una vez que hayas adquirido el "mindset" necesario, podrás aplicarla no solo a otros problemas de SEO, sino también a situaciones de tu vida cotidiana.

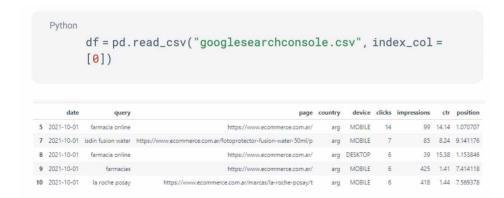
# 1 - Importar librerías

Son paquetes que nos permiten acceder a soluciones previamente generadas por otros (la magia de Python). En vez de ponerte a cocinar algo sumamente complejo, adquirimos la comida que ya está preparada (en este caso, de manera gratuita).

```
#estas librerias sirven para manipular datos
import numpy as np
import pandas as pd
import datetime import datetime, timedelta
```

### 2 - Leer Data Frame

Son paquetes que nos permiten acceder a soluciones previamente generadas por otros (la magia de Python). En vez de ponerte a cocinar algo sumamente complejo, adquirimos la comida que ya está preparada (en este caso, de manera gratuita).





# 3 - Transformar y crear nuevas variables

Necesitamos asegurarnos de que cada tipo de dato esté en su formato adecuado. Esto es similar a lo que hacemos en una hoja de cálculo cuando ajustamos los datos para llevar un seguimiento de nuestras finanzas (¿por qué todos lo hacemos, no?).

### Además, crearemos las siguientes variables:

keyword brand Vertical Category Sub Category

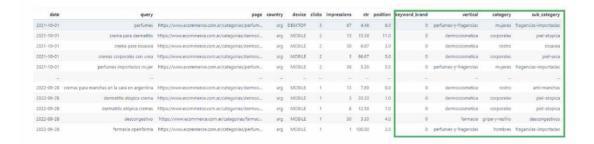
Para lograr esto, utilizaremos la función data\_cleaning que creamos anteriormente.

La analogía de una función en Python se asemeja al proceso de seguir una receta culinaria específica. Una función es un bloque de código que ejecuta una secuencia de acciones predefinidas cuando se invoca. Al igual que seguimos los pasos de una receta en la cocina para lograr un resultado deseado, en Python, una función encapsula un conjunto de instrucciones que pueden ser reutilizadas en diversas partes del programa. Al llamar a una función, las instrucciones internas se ejecutan en un orden específico, lo que permite automatizar tareas y estructurar el código de manera modular.

### Python

#Con esta función limpiamos el data frame y creamos columnas automaticamente.

new\_df = data\_cleanining(df)



# 4 - Filtros Múltiples

Previo al análisis de datos, necesitaremos filtrar la información. Para lograr esto, emplearemos una función que, a partir de parámetros específicos, ejecutará todas las instrucciones deseadas.



### Parámetros:

df: data frame search console.

fecha: 6 (representa últimos 6 meses).

country: arg (Argentina).

position: 10 (selecciona las que se sitúan en la primer página).

**URLs:** 'categorias' (quitamos home y productos).

Keyword brand: Marcamos 0 para eliminar las keyworbrand (1 significa "si" 0

significa "no").

Podríamos considerar también aplicar un filtro según el dispositivo, pero en esta ocasión optaremos por no incluirlo.

```
#Aplicamos una función para filtrar la data donde cada argumento puede ser modificable.

df_filter = data_filter(df,6,'arg',10,'categorias',0)
```

En el cuadro verde, se pueden observar las columnas a las cuales hemos aplicado filtros.

sub_categor	category	vertical	keyword_brand	position	ctr	impressions	clicks	device	country	page	query	date
	barbijos-y-tapabocas	cuidado-de-la-salud	0	7.0	3,49	86	3	desktop	arg	https://www.ecommerce.com.ar/categorias/cuidad	barbijos	022-03-01
fragancias-importada	mujeres	perfumes-y-fragancias	0	5.0	5.00	60	3	mobile	arg	https://www.ecommerce.com.ar/categorias/perfum	perfumes importados de mujer	022-03-01
	barbijos-y-tapabocas	cuidado-de-la-salud	0	10.0	0.39	509	2	mobile	arg	https://www.ecommerce.com.ar/categorias/cuidad	barbijos	022-03-01
fragancias-importada	mujeres	perfumes-y-fragancias	0	8.0	1.46	137	2	mobile	arg	https://www.ecommerce.com.ar/categorias/perfum	perfumes importados	022-03-01
	barbijos-y-tapabocas	cuidado-de-la-salud	0	8.0	50.00	2	1	mobile	arg	https://www.ecommerce.com.ar/categorias/cuidad	bsrbijos	022-03-01
		100	100		-	-		+1			PF.	-
piel-atopic	corporales	dermocosmetica	0	1.0	12.50	8	- 1	mobile	arg	https://www.ecommerce.com.ar/categorias/dermoc	dermatitis atopica cremas	1022-09-28
descongestivo	gripe-y-resfrio	farmacia	0	4.0	3.33	30	1	mobile	arg	https://www.ecommerce.com.ar/categorias/farmac	descongestivo	022-09-28
Nat	t	dermocosmetica	0	1.0	50.00	2	1	mobile	arg	https://www.ecommerce.com.ar/categorias/dermoc	descuentos en cremas dermatologicas	022-09-28
fragancias-importada	hombres	perfumes-y-fragancias	0	2.0	100.00		1	mobile	arg	https://www.ecommerce.com.ar/categorias/perfum	farmacia openfarma	022-09-28
Nati	t t	dermocosmetica	0	5.0	0.64	156	1	mobile	arg	https://www.ecommerce.com.ar/categorias/dermoc	farmacias	022-09-28

# 5 - Análisis exploratorio - Optimización

En esta fase, planteamos preguntas en relación a los datos que hemos construido anteriormente, siguiendo un enfoque desde lo general hasta lo más específico. El objetivo es tomar decisiones basadas en los volúmenes de búsqueda y clicks. Comenzamos por comprender las verticales, luego las categorías, subcategorías y, por último, las consultas individuales. Con el fin de evitar la repetición con el código que proporcionaremos al final, presentaremos un ejemplo de cómo realizar el análisis específicamente para las verticales:



```
clicks_sum=('clicks', 'sum'))
    .reset_index()
    .assign(ctr=lambda df: round(df['clicks_sum'] /
df['impressions_sum'],4)*100)
    .sort_values(['impressions_sum', 'clicks_sum'],
ascending=[False, False])
)
```

### **Verticales**

Seleccionamos dermocosmética

	vertical	impressions_sum	clicks_sum	ctr
7	dermocosmetica	106013	4487	4.23
3	perfumes-y-fragancias	104196	5692	5.46
1	cuidado-de-la-salud	44913	549	1.22
5	farmacia	17338	712	4.11
6	maquillaje	3469	71	2.05
8	suplementos-y-nutricion	251	25	9.96
2	cuidado-personal	98	39	39.80
0	bebes-y-maternidad	6	3	50.00
4	electro	4	3	75.00

# Categorías

Dentro de la vertical de dermocosmética, nos enfocaremos en la categoría **protección** solar.

category	impressions_sum	clicks_sum	ctr
proteccion-solar	70060	1191	1.70
rostro	20701	2415	11.67
corporales	8950	1952	21.81
ť	4469	122	2.73
capilares	9	6	66.67
accesorios	7	6	85.71

# Sub categorías

Dentro de la categoría de protección solar, encontramos "rostro" y los productos no asignados (t).





### ¿Qué acciones tomaremos?

El objetivo de este análisis no es discutir si el tipo de estrategia es la correcta o no. Para eso deberíamos tener más contexto sobre el negocio.

Identificamos una situación en la que una categoría y subcategoría comparten el mismo nombre "rostro", lo que genera competencia entre ellas. Para abordar esto, proponemos realizar las siguientes acciones:

- Eliminar la subcategoría "rostro" de la categoría "protección solar".
- Crear una nueva subcategoría llamada "protector solar facial" (que no existía previamente) dentro de la categoría adecuada, es decir, "rostro".
- Optimizar title, h1 y metadescription basados en esta nueva subcategoría...
- Agregar preguntas frecuentes extraídas de las PAA (Preguntas Frecuentes en los resultados de búsqueda) relacionadas con la keyword "protector solar".

### 6 - Automatización

Cuando hablamos de automatización, muchas veces nos referimos a una "semi-automatización". Es decir, hay ciertas etapas en las que aún mantenemos el control manual. En el contexto del análisis de datos, este approach me parece muy acertado.

De hecho, en este ejemplo utilicé poco código de programación ya que previamente había construido funciones que solo requerían ajustes en el input. Este concepto puede verse en más detalle en el siguiente ejemplo:

```
Python data_filter(df, 6, 'arg', 10, 'categorias', 1)
```

podría modificar los argumentos de la función y, rápidamente, obtener algo de la siguiente índole:

```
Python

data_filter(df,3,'chi',5,'categorias',0)
```

En este ejemplo, estoy aplicando un filtro a la información para considerar los últimos 6 meses restringiendo la búsqueda a Chile. Nos enfocamos en keywords que se sitúan por debajo de la posición 5, que pertenecen al path de categorías y que, además, son términos de marca.

# ¿Puedo usarlo para otros clientes?

La respuesta es un sí rotundo, ya que los datos provienen de la misma API (Search Console) y para la función es exactamente lo mismo.



# **Conclusiones finales**

Sin duda, has notado que analizar datos va mucho más allá de simplemente copiar y pegar un script. Nuestra meta es identificar problemas, descomponerlos en partes y ejecutar diversas estrategias. Python nos permite adentrarnos en la búsqueda de nuevas preguntas y respuestas, sirviendo como una herramienta para explorar rincones que de otra manera serían inaccesibles o, en su defecto, para acelerar este proceso.

### ¿Cuáles han sido los beneficios que hemos obtenido de este e-book?

- **1. El Precio de la Repetitividad:** Ahora somos conscientes de cuánto tiempo y recursos pueden consumir las tareas repetitivas. Esta toma de conciencia nos invita a explorar la automatización y cómo podría traducirse en economías de escala y mayores beneficios.
- **2. Velocidad vs. Eficiencia:** Hemos reflexionado sobre la diferencia entre trabajar rápido y trabajar eficientemente. Mientras que herramientas como Excel tienen su lugar, nos hemos preguntado sobre el potencial que Python podría ofrecer para la agilidad en SEO.
- **3. Profundizando en los Insights:** La revelación sobre la API de Search Console y su capacidad para extraer hasta 50.000 registros diarios nos muestra cuánta información potencialmente valiosa podría estar quedando sin aprovechar.
- **4. El Desafío de la Interdependencia:** Reflexionamos sobre la tensión entre los equipos de negocio y IT. La conciencia de que con el conocimiento adecuado podríamos resolver problemas de manera más autónoma subraya la importancia de la capacitación continua.
- **5. Casos a Considerar:** Presentamos dos ejemplos relacionados con Python y SEO para incitar la curiosidad y el deseo de explorar más sobre las posibles soluciones.



# ¡Gracias!

### ¡Queridos lectores!

Hemos llegado al final de este emocionante viaje por el mundo de los datos y el SEO. Ha sido un auténtico placer compartir con ustedes todo este conocimiento y explorar juntos las oportunidades que nos ofrece este apasionante campo.

Queremos expresar nuestro sincero agradecimiento por su compañía y por seguirnos en nuestras redes sociales. Sus interacciones, preguntas y comentarios han sido una constante fuente de inspiración y motivación para nosotros.

Además, les recordamos que somos una **comunidad gratuita** de datos y SEO, donde seguiremos compartiendo contenido valioso, respondiendo sus preguntas y brindándonos apoyo mutuo. Juntos continuaremos creciendo y aprendiendo.

No queremos dejar de mencionar el exclusivo **programa de Python SEO Data Analytics**, diseñado especialmente por Alan para profesionales de la industria como ustedes. En este programa, nos enfocaremos en enseñarles las técnicas más avanzadas y aplicadas de Python en el ámbito del SEO. Estamos entusiasmados por la oportunidad de trabajar de cerca con cada uno de ustedes y ayudarles a alcanzar nuevos niveles de excelencia en sus estrategias SEO.

Queremos agradecer también a los alumnos que han compartido sus resultados y experiencias con el programa, inspirando la creación de este eBook. Sus testimonios nos motivan a seguir mejorando y ofreciendo contenido de alta calidad.

¡Gracias una vez más por ser parte de esta apasionante comunidad dedicada a los datos y el SEO! Esperamos mantenernos en contacto y seguir explorando juntos las maravillas de este fascinante mundo.

Con gratitud,

**Alan y Antoine** 

