

# ANÁLISIS EXPLORATORIO DE DATOS – IVÁN SANTAMARÍA

## 1. Elegir temática

La temática de este EDA se centra analizar la publicidad en Amazon Ads España del libro La Erupción Zafiro de I.M. Redwright. Se trata de una novela de fantasía que es el primer volumen de la saga La Decisión de la Espada.

## 2. Objetivo

Con este análisis exploratorio se pretende optimizar las campañas de publicidad realizadas en Amazon Ads en base a conclusiones basadas en datos históricos. De esta manera, se pretende aumentar el número de ventas a través de una optimización del presupuesto dedicado a estas campañas de publicidad.

## 3. Obtención de datos

La obtención de los datos objeto de nuestro análisis se realiza a través del mismo portal de Ads (<https://advertising.amazon.es>) ya que incluye un botón de exportación de datos. De este modo, hemos seleccionado las 4 campañas que queremos analizar y hemos extraído sus datos.

Debido a la forma en la que se extraen los

## 4. Definición de hipótesis

Nuestro análisis se divide en 3 subapartados:

1. Concordancias: las palabras clave de una campaña pueden ser seleccionadas como concordancia amplia, frase, o exacta. De acuerdo a la teoría, la que más rendimiento ofrece es la concordancia exacta, si bien también es la que resulta más costosa (su CPC es más elevado), comprobaremos si efectivamente es así.
2. Palabras clave: se pretende detectar las palabras clave más eficientes, establecer cuánto cuesta obtener una venta tanto en cuestión de clicks como también en coste. Adicionalmente, se espera que exista correlación de las ventas con alguna variable para así establecer algún modelo que nos permita establecer una predicción.
3. Ineficiencias: se prevé que existan palabras clave que consuman presupuesto sin resultar en ventas, del mismo modo, también puede haber keywords que sí que conviertan pero que lo hagan tras un elevado número de clicks (o un coste alto debido a un CPC superior a la media).

## 5. Preprocesado:

Se ha obtenido datos de 4 campañas, siendo las siguientes: Broad, Phase, Kindle y Frase. Estos 4 CSV tienen una estructura de columnas idéntica, por lo que se agregaron para obtener así un data set completo.

## 6. Limpieza de los datos:

Existía un número significativo de columnas que no eran necesarias para nuestro estudio, de manera que se ha procedido a eliminarlas. Adicionalmente, para agilizar el proceso de análisis, se ha cambiado el nombre a varias variables de tal modo que fueran más fácilmente “llamables”.

Del mismo modo, se han añadido columnas adicionales que permitieran extraer conclusiones, en concreto, se han añadido 2 columnas principalmente:

Clicks para obtener una venta: Dividimos los clicks de cada keyword (Clicks) por el número de ventas (Pedidos).

Coste para obtener una venta: : Dividimos el coste total (gasto) por el número de ventas (Pedidos)

Adicionalmente, debido a la naturaleza exploratoria de nuestro estudio, se trabajó con dos data sets principales.

Datos\_raw: recoge los datos ya agregados y debidamente tratados incluyendo la limpieza.

Datos\_con\_ventas: se trata de un data set filtrado del anterior, donde únicamente tiene en cuenta aquellos valores en los que haya habido ventas (útil para determinados cálculos concretos).

## 7. Exploratorio:

Con todo lo anterior, se ha realizado una serie de comprobaciones, agrupación de datos, gráficos, etc para poder sustentar el estudio y llegar a conclusiones.