

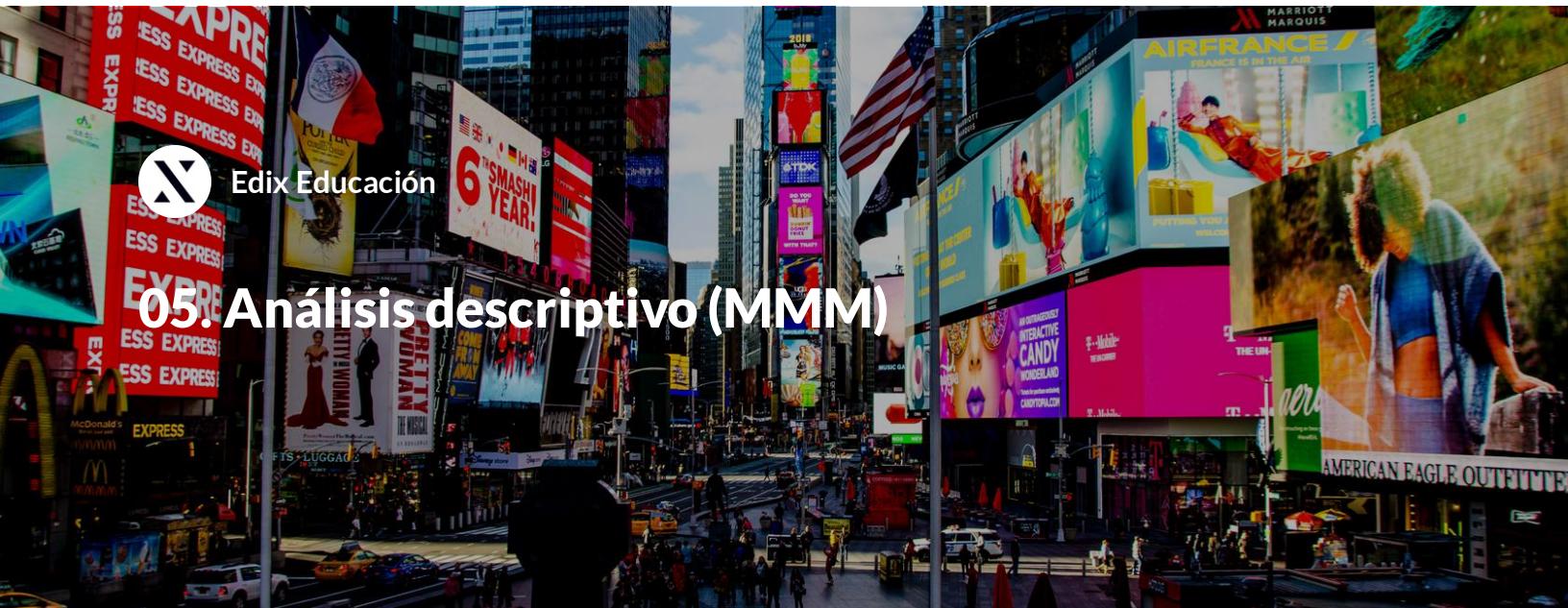
# Fastbook 05

## Medición de Eficacia Publicitaria

Análisis descriptivo (MMM)



## 05. Análisis descriptivo (MMM)



En los fastbooks anteriores hemos hablado sobre variables, tanto dependientes como independientes, que se utilizan al desarrollar modelos de MMM. Para ello, hemos revisado varios ejemplos de distintos tipos de anunciantes.

Ahora vamos a ver qué hacer con esta información. Una vez que tenemos todos los datos recopilados y ordenados, es necesario realizar un análisis descriptivo minucioso para entender patrones de comportamiento en los datos. Cuando conozcamos bien la información con la que vamos a trabajar, podremos empezar a desarrollar los modelos de regresión.

En este fastbook vamos a profundizar en el desarrollo del análisis descriptivo repasando varios puntos que hay que cubrir.

*Autor: Carlos Real Ugena*

Análisis descriptivo

Análisis tendencial

Análisis gráfico

Análisis estacional

Análisis de correlaciones

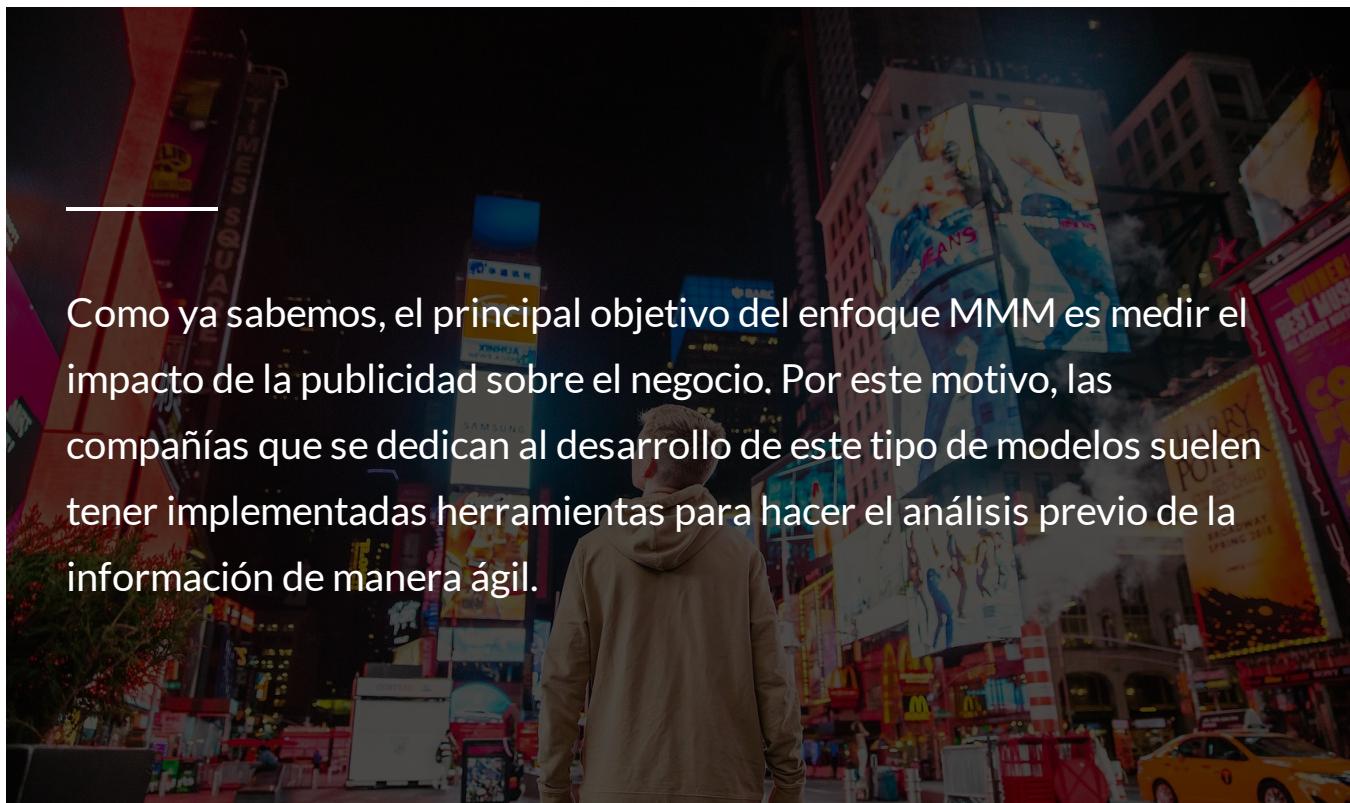
Otros análisis

# Análisis descriptivo

X Edix Educación

---

Antes de comenzar la modelización, es preciso realizar un análisis descriptivo (exploratorio) de los datos para **conocer mejor las variables objetivo a modelizar, las variables explicativas a testar y las relaciones que hay entre ellas.**



Como ya sabemos, el principal objetivo del enfoque MMM es medir el impacto de la publicidad sobre el negocio. Por este motivo, las compañías que se dedican al desarrollo de este tipo de modelos suelen tener implementadas herramientas para hacer el análisis previo de la información de manera ágil.

---

En el caso de que no se tengan, se podrá utilizar R, Python o cualquier software que sirva para hacer análisis estadístico de datos.

---

Pero ¿qué tipo de análisis descriptivos vamos a llevar a cabo? A continuación, se muestran los ejemplos más habituales al explorar series temporales.

Lesson 2 of 6

## Análisis gráfico

X Edix Educación

Lo primero que haremos es graficar las series más relevantes del estudio. Así, podremos contrastar la información que hayamos recopilado. Por ejemplo, ¿han crecido las ventas a lo largo del tiempo?, ¿hay estacionalidad del mercado?, ¿cuánto se está invirtiendo en publicidad respecto a la competencia?

Veamos un ejemplo de serie temporal de ventas semanales:



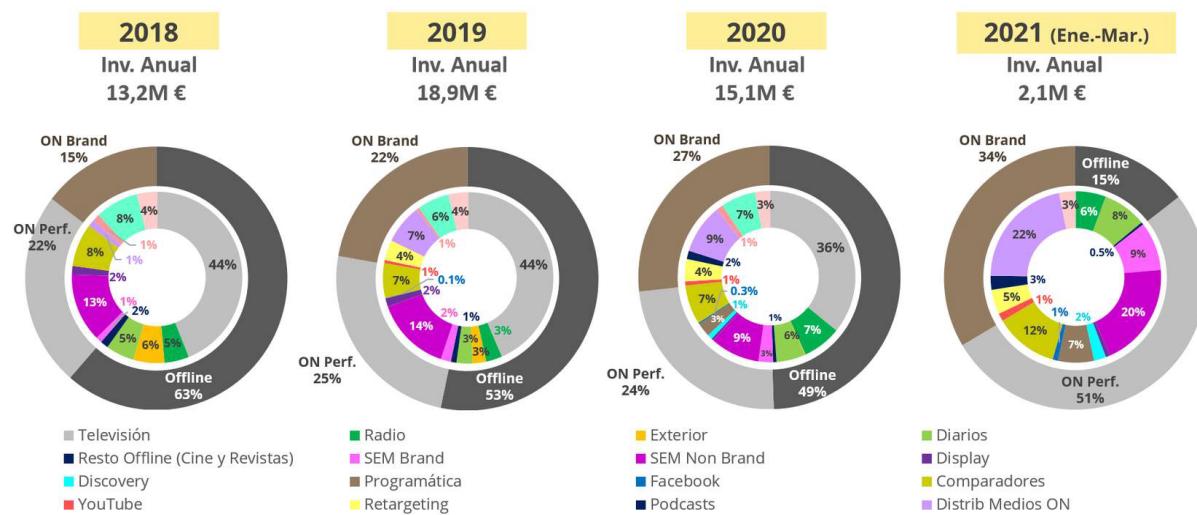
Con un simple gráfico podemos obtener una serie de conclusiones interesantes:

- Se observa una ligera tendencia decreciente en la serie de ventas, en 2020, es el mayor decremento (-37%) debido a un inicio de año (enero-marzo) un -15% más bajo que el anterior y la posterior bajada por el COVID (marzo-abril). A partir de mayo de 2020, vemos una recuperación, pero sin llegar al nivel inicial de ventas.
- Se ve una clara estacionalidad negativa en: Semana Santa, puente de mayo, puente de agosto, Navidad, Nochevieja y Reyes. Es decir, son momentos en los que sistemáticamente se vende menos que lo habitual.

**Este mismo análisis lo podemos replicar con todas las series que tengamos disponibles y anticipar comportamientos que tener en cuenta en la modelización.** Lo primero que evaluaremos es la tendencia y estacionalidad de las series temporales bajo análisis para crear indicadores que testar en los modelos.

- 1 En el ejemplo anterior, tendremos que construir variables que identifiquen las estacionalidades negativas.
- 2 En el caso de Semana Santa, crearemos una variable que tome el valor 1 durante la semana de Semana Santa de cada año y 0 durante el resto de las semanas.
- 3 De manera análoga, necesitaremos variables para el resto de las estacionalidades.
- 4 En algunos casos, será necesario dividir estas variables por años o por agregaciones de años si se estima que el impacto ha sido distinto en cada uno de ellos.
- 5 Cuando estemos analizando series temporales, lo ideal es emplear gráficos de líneas para poder observar visualmente cuándo se producen los incrementos y los decrementos.
- 6 Además, graficar los datos de medios será muy útil. Para ello, utilizaremos normalmente dos tipos de gráficos: gráficos de tartas y gráficos de áreas apiladas.

A continuación, vemos un ejemplo de representación gráfica de la evolución de inversión anual en medios:



En el gráfico anterior, podemos ver cuál ha sido la estrategia de inversión anual por medios del anunciante y extraer las primeras conclusiones:

- La inversión se reduce en 2020 significativamente.
- Cada año las inversiones online van incrementando su cuota en el mix de medios, sobre todo, la publicidad de marca.
- Televisión es el medio con mayor volumen de inversión entre 2018 y 2020.

---

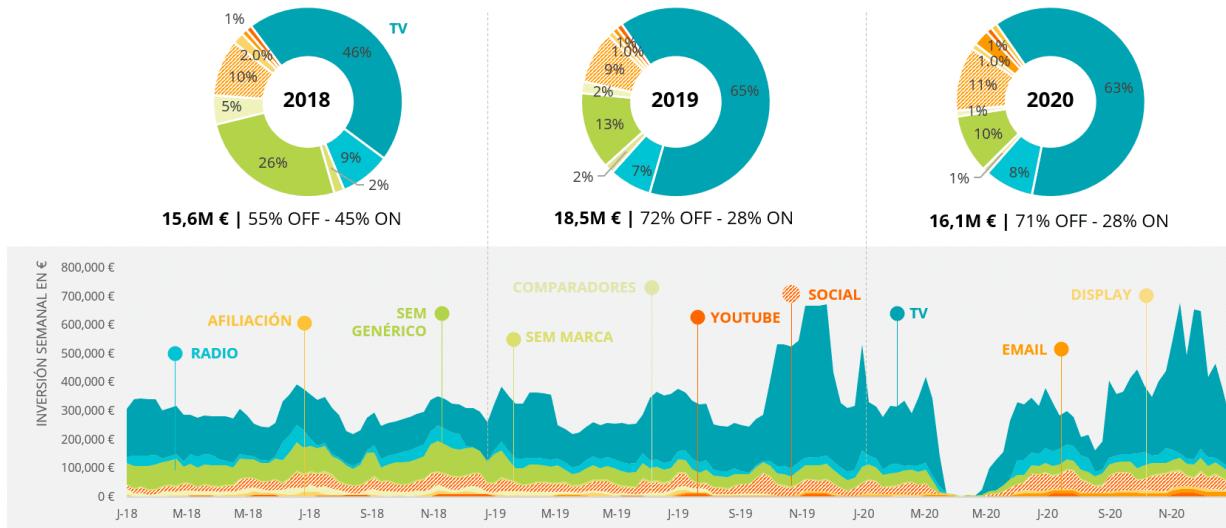
**Los gráficos de tartas pueden resultar confusos y se podrían sustituir por gráficos de barras apilados. Es una decisión que tomaremos en función de los datos disponibles y de la audiencia a la que se quiera mostrar los resultados.**

El resumen anterior está bien para saber cuál ha sido la estrategia anual.

¿Qué ha ocurrido con la distribución de las inversiones?

Podemos utilizar un gráfico de áreas apiladas para visualizar dicha información.

Veamos otro ejemplo de representación gráfica:



Estos gráficos nos muestran lo siguiente:

- La estrategia de medios ha cambiado en los últimos años, dando más importancia a los medios offline, y especialmente a la televisión, que ha pasado del 46% del presupuesto en 2018 al 63% en 2020. SEM Genérico en cambio ha bajado del 26% al 10 % en 2020.
- El inicio de la expansión del covid-19 por Europa ha hecho que la inversión se reduzca significativamente durante los primeros meses de 2020.

Lesson 3 of 6

# Análisis de correlaciones

 Edix Educación

---

Otro punto importante es el **estudio de correlaciones**.

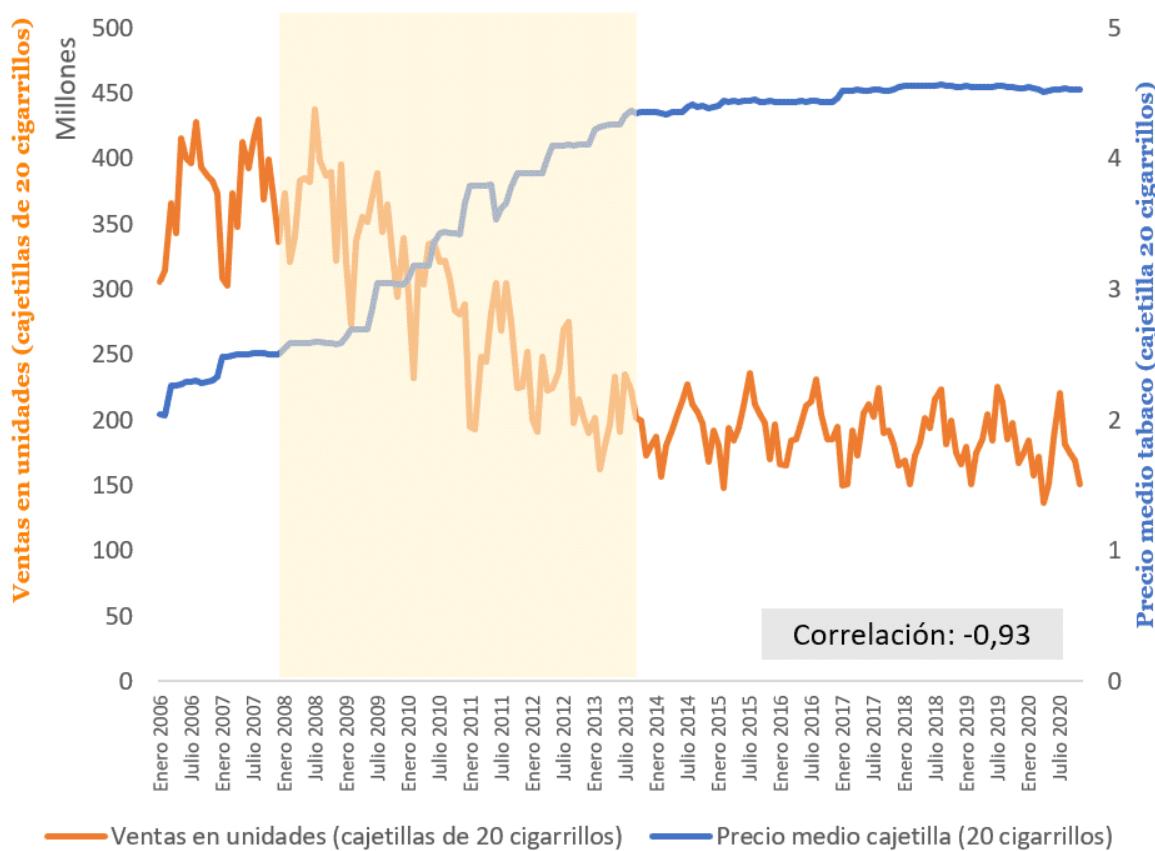
---

La correlación indica la fuerza y la dirección de una relación lineal y proporcionalidad entre dos variables estadísticas y puede ir de -1 a +1.

En nuestro caso, va a servir como primer indicio de impacto de las variables explicativas en la serie a modelizar. Lo recomendable es **acompañar el análisis de correlaciones con gráficos** que sirvan para entender mejor los datos.

A continuación, un ejemplo de correlación entre las ventas y el precio medio del tabaco durante los últimos años:

## Ventas de tabaco vs Precio tabaco



El gráfico no deja lugar a dudas. El consumo de tabaco ha disminuido de manera significativa entre 2008 y 2013 coincidiendo con la subida de precios y la crisis en España. Podemos observar que el comportamiento es inverso, cuánto más alto es el precio, menos se vende.

La correlación entre las ventas y el precio del  $-0,93$ , valor muy cercano a  $-1$ . Es muy alta si la comparamos con lo que sucede en otros mercados. Por lo tanto, si queremos modelizar las ventas de tabaco, necesitaremos incluir el precio como variable explicativa de dichas ventas.

Adicionalmente, podemos construir tablas de correlaciones que nos marquen las variables que están más y menos correlacionadas en nuestro conjunto de datos:

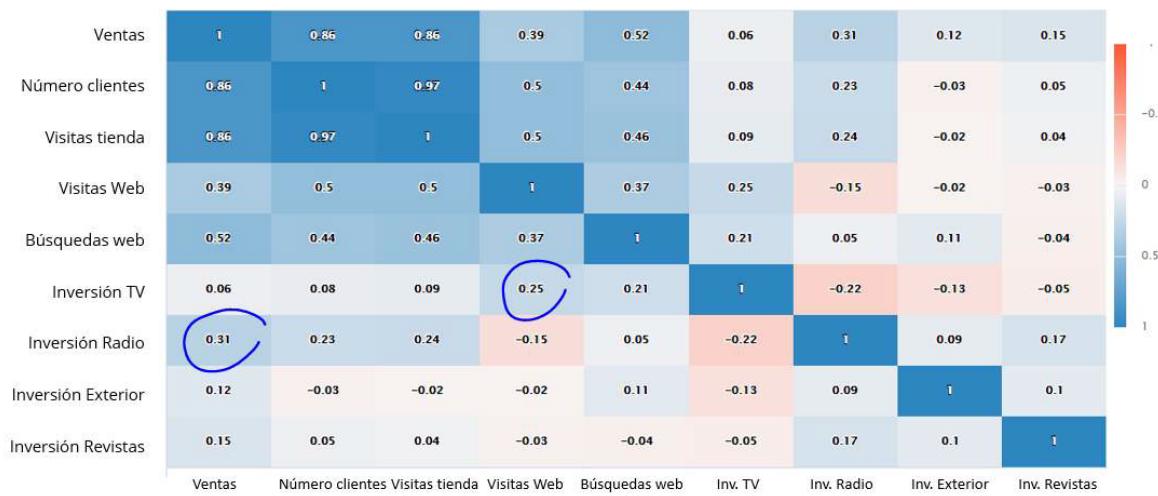


Tabla de correlaciones entre variables.

---

En el ejemplo anterior, se puede observar que, en general, las correlaciones son positivas. En particular, la inversión en radio es la que presenta una correlación mayor con las ventas y la televisión con el número de visitas web. El modelo será el que determine la relevancia de cada medio, pero, a priori los datos confirman que existe correlación positiva entre las variables bajo estudio.

## Ad-stock de los medios

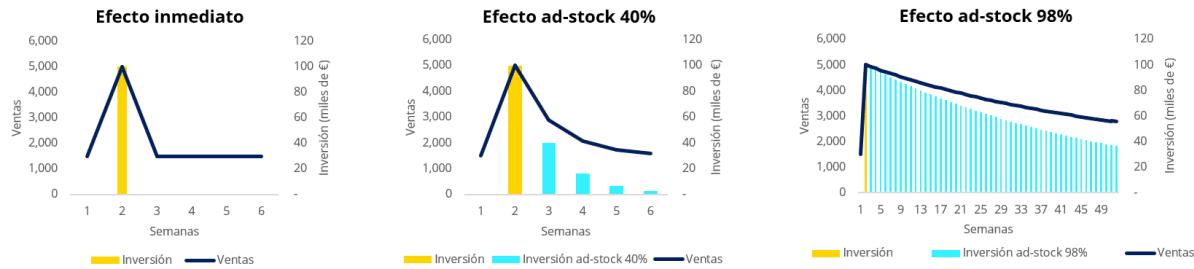
**Los medios publicitarios tienen un efecto sobre el negocio que no cae de forma brusca cuando se deja de hacer campaña, sino que se prolonga durante cierto periodo de tiempo.**

---

Este efecto se estudia con las variables de ad-stock, dichas variables se calculan teniendo en cuenta el recuerdo de la publicidad en los días siguientes, el cual es un porcentaje del impacto original.

---

A continuación, se muestran tres ejemplos:



1

**Efecto inmediato:** el efecto de la publicidad es inmediato. Este ejemplo ilustraría negocios de respuesta directa en el que la acción de los consumidores (visitar la web, comprar, etc.) se realiza de manera inmediata.

2

**Efecto ad-stock 40%:** en este caso, la semana siguiente a la inversión publicitaria tendríamos el 40% del efecto de la semana anterior. Es decir, si la semana 1 tenemos una inversión de 100.000€ en radio, la semana siguiente es como si tuviéramos 40.000€ de inversión sin hacer nada. Este efecto será acumulativo si se va realizando inversiones de manera constante.

3

**Efecto ad-stock 98%:** caso similar al anterior, pero en el que el efecto de la publicidad en la semana posterior es del 98%. Este efecto recoge el impacto a más largo plazo de la publicidad.

En resumen, la fórmula del ad-stock será como sigue:

$$m_{ADx}[i] = m[i] + m_{ADx}[i - 1] \times \frac{x}{100}$$

En donde 'm' es el medio, 'i', el instante de tiempo y 'x' es el ad-stock. Es una fórmula recursiva.

La definición del ad-stock no es algo trivial. Habitualmente se decidirá teniendo en cuenta:

- Significatividad de las variables en el modelo o correlación con la serie de negocio.
- Benchmark de comportamiento de la publicidad.

Hay múltiples estudios que muestran que los medios audiovisuales (televisión, vídeo online...) generan mayor efecto de recuerdo. Por lo tanto, lo normal será que el ad-stock incluido para estos medios sea más alto.

Entonces, ¿qué ad-stock es el más común para cada medio? Para medir los efectos del corto plazo en un modelo semanal, los niveles habituales de ad-stock son:

- **Televisión y vídeo online:** 40%-50%.
- **Otros medios offline:** 15%-25%.
- **Otros medios online:** 0%-20%.

---

Para elegir el más adecuado, podemos utilizar aquel que maximice la correlación entre la variable de medios con ad-stock y las variables de negocio a modelizar.

---

Por otra parte, para medir los efectos en el largo plazo, habitualmente se construye una variable de inversión total en medios con niveles de ad-stock entre 90%-98%.

En este punto, también tenemos que elegir el valor de ad-stock inicial.

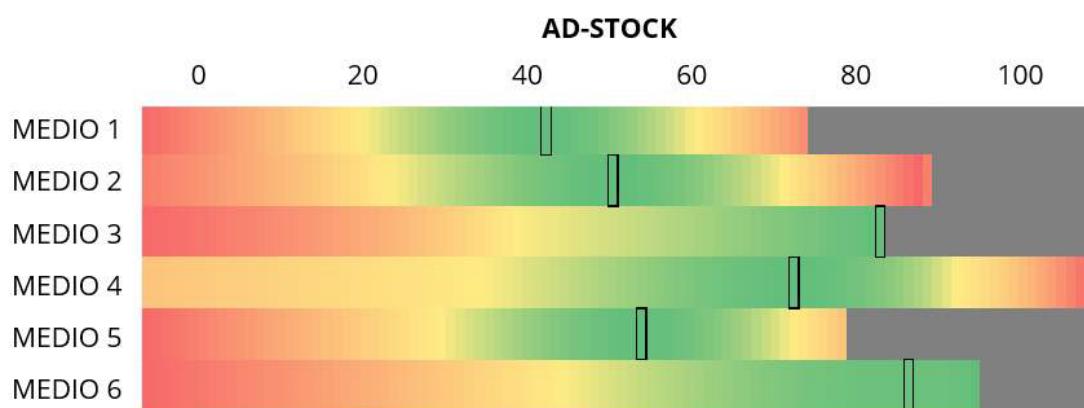
Si tuviéramos la inversión de años anteriores al periodo de análisis podemos estimar el ad-stock inicial.

Si no disponemos de esa información (que es lo habitual), podemos estimar el ad-stock inicial estimando el nivel de inversión previo.

A continuación, un ejemplo de estudio para entender las diferentes correlaciones obtenidas a partir de múltiples ad-stocks:

## CORRELACIÓN ENTRE LOS MEDIOS Y LAS VENTAS

(cuánto más verde más positiva)



A la vista de los resultados, el medio 1 es el que presenta una correlación mayor con un ad-stock más bajo. En cambio, los medios 3 y 6 a priori están generando mayor recuerdo en el consumidor y la máxima correlación se obtiene por encima del 80% de ad-stock.

# Análisis tendencial



Como vimos anteriormente, aislar la tendencia de la serie a modelar será fundamental.

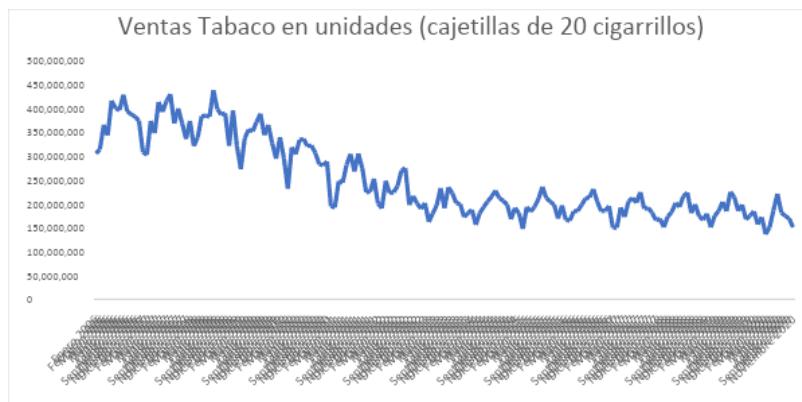
Principalmente, tenemos dos opciones:

- Trabajar directamente con variables tendenciales exógenas.
- Encontrar variables que nos expliquen la tendencia de la serie a modelizar.

Lo mejor será la segunda opción. Sin embargo, en muchas ocasiones tendremos que construir nosotros una variable de tendencia porque no habrá ninguna variable que explique el comportamiento del negocio.

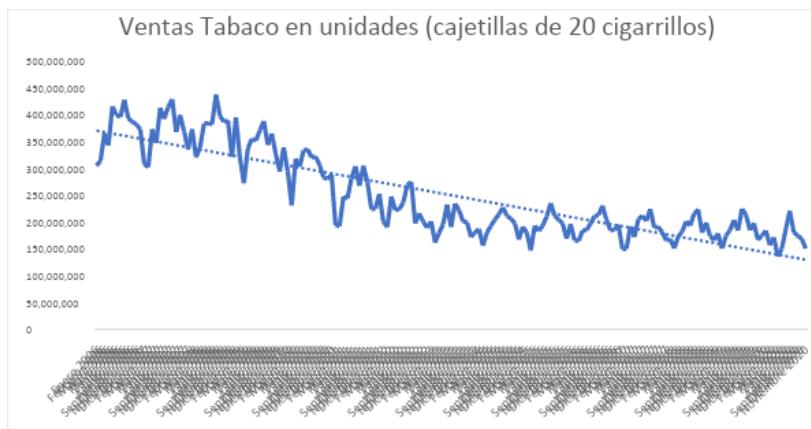
Algunos de los tipos de tendencias que podemos emplear de manera ‘artificial’ son tendencias lineales, logarítmicas o exponenciales.

Cualquier software estadístico podrá generar estas variables. Veamos un ejemplo en Excel a partir de la serie real de ventas mensuales de tabaco en España entre enero 2006 y noviembre 2020:



1

Primero, obtendremos la tendencia lineal. Pinchando en el botón derecho del ratón sobre la serie (en el gráfico) se puede añadir la tendencia 'Add Trendline'. Seleccionamos también la opción de ver la ecuación y el R2 en el gráfico:

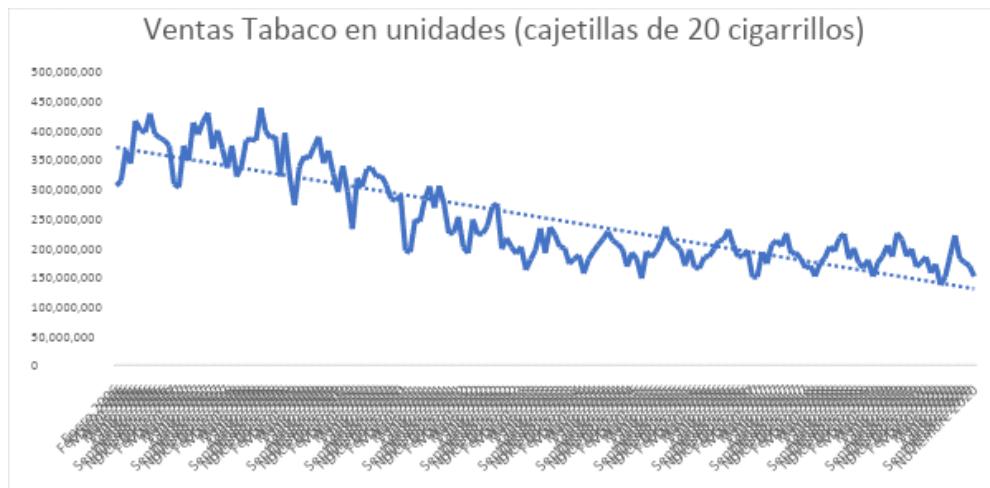


¿Qué podemos observar en el gráfico anterior?

- La tendencia (trendline) que hemos añadido es lineal y decreciente. Por defecto, Excel selecciona la tendencia lineal que podremos cambiar posteriormente.
- La fórmula nos permite construir una serie temporal que testar en el modelo como variable explicativa:  $y = -1E+06x + 4E+08$ .
- El  $R^2$  es 73,7%. Este valor lo compararemos con otras tendencias para ver cuál se ajusta mejor a la serie de ventas.

2

Ahora, cambiaremos la tendencia por una logarítmica. Para ello, seleccionamos el formato ‘Format Trendline’ y cambiamos la tendencia lineal ‘Linear’ por una logarítmica ‘Logarithmic’:

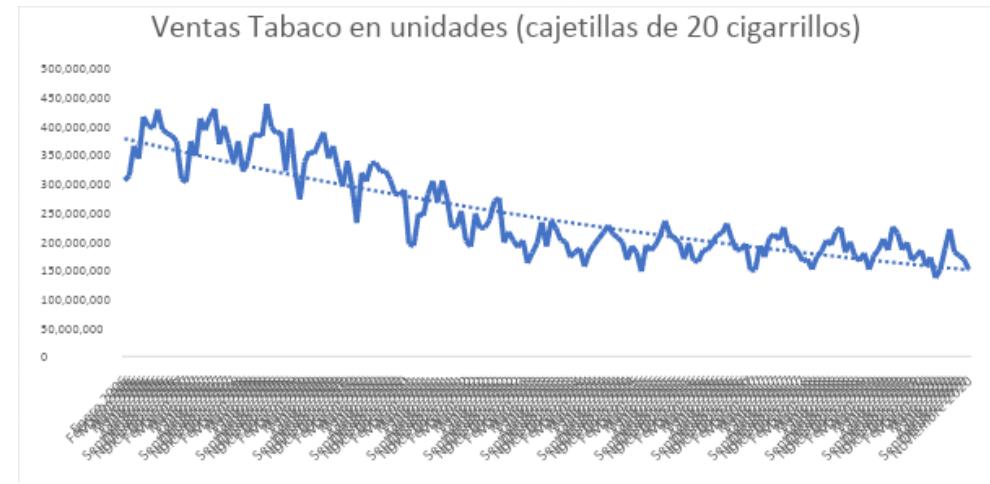


3

En este caso obtenemos un peor ajuste, por lo que preferiremos la primera opción con la tendencia lineal.

4

Para terminar, probaremos con una tendencia exponencial. Para ello, seleccionamos el formato ‘Format Trendline’ y seleccionamos la opción de tendencia exponencial ‘Exponential’:



Como vemos, el  $R^2$  ha aumentado y es el más alto de los tres casos ( $R^2 = 75,1\%$ ). En este caso, también tenemos nuestra fórmula para construir una variable que evaluar en los modelos:  $y = 4E+08e^{-0.005x}$ .

En resumen, podremos testar cualquiera de estas tres tendencias en nuestro modelo.

¿Algún problema?

Que al construir modelos de MMM queremos explicar el comportamiento del negocio.

**Si ha habido un crecimiento o decrecimiento paulatino a lo largo del tiempo, seguro que hay un factor que ha afectado.** Por tanto, testar este tipo de variables debe ser nuestra última prioridad. En el caso anterior del tabaco, solo necesitaremos un poco de tiempo para averiguar que la bajada en ventas se debe a varios motivos: subida de precio, subida del IVA, incremento de la tasa de paro, etc.

Lesson 5 of 6

# Análisis estacional

**X** Edix Educación

---

Además de la tendencia, habrá que determinar la **estacionalidad de la serie a modelizar**.

Podemos construir variables que recojan momentos específicos del tiempo a través de 0 y 1.

¿Existe alguna manera más sofisticada de hacer este trabajo?

Sí, podemos aislar la estacionalidad a través de componentes de Fourier.

Lo importante de este bloque es entender el concepto que hay detrás, ya que lo habitual es que las compañías cuenten con **perfiles técnicos** que entiendan todo el detalle matemático.

## Construcción de series de Fourier

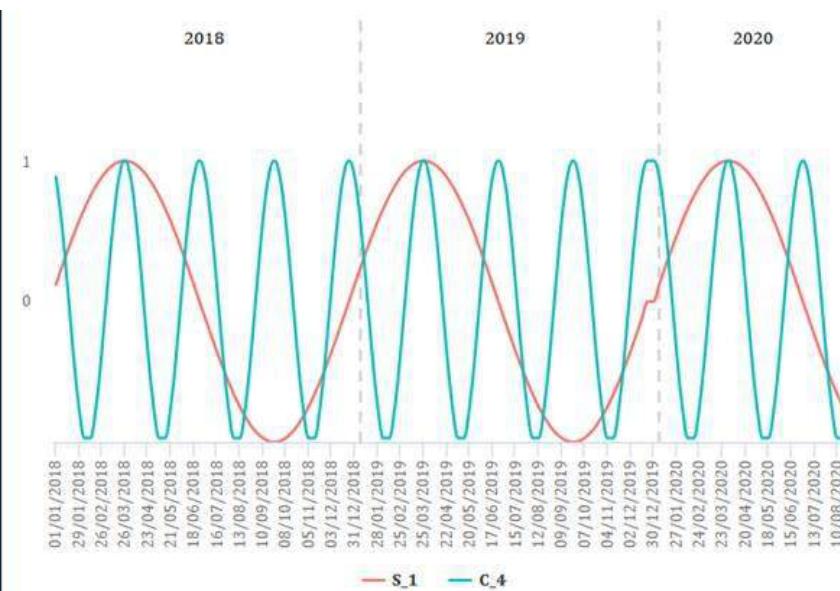
El objetivo de las series de Fourier es detectar patrones estacionales que no puedan ser explicados por las variables incluidas en el modelo.

Para la construcción de dichas series, primero se crean las **componentes de Fourier**, series de senos y cosenos (donde: 'i' varía entre 1 y 12 meses del año y 'semana' es una semana del año) con la siguiente forma:

- $\cos(2 * \pi * i * \text{semana} / 52)$ .
- $\sin(2 * \pi * i * \text{semana} / 52)$ .

A partir de las ecuaciones anteriores se pueden generar las variables C\_1, C\_2, ..., C\_12, S\_1, S\_2, ..., S12.

Imagen: Ejemplo de componentes de Fourier.



La serie de Fourier se compondrá de la suma de varias componentes de Fourier. Por ejemplo:  $C_{11} + C_2 + S_6 + S_{11} + C_7$ . Esto nos permitirá medir ciclos y estacionalidades dentro de nuestra serie temporal.

Lesson 6 of 6

# Otros análisis

X Edix Educación

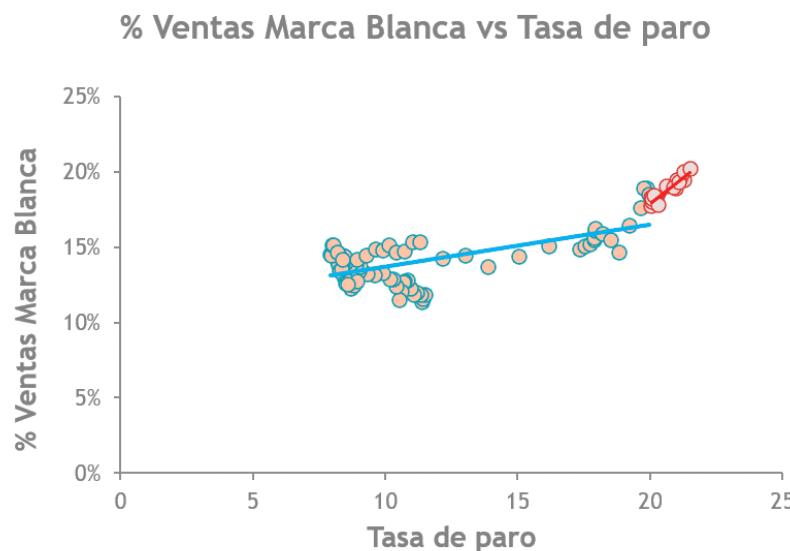
1

## Gráficos de dispersión

Hasta ahora, hemos visto los gráficos de series temporales. A veces será necesario representar los datos con un gráfico de dispersión para entender relaciones más complejas.

**Recordemos que un gráfico de dispersión consiste en la representación gráfica de dos variables para un conjunto de datos.**

Veamos un ejemplo de dispersión en el que cada punto es un mes y representa la tasa de paro frente al % de ventas de la marca blanca:

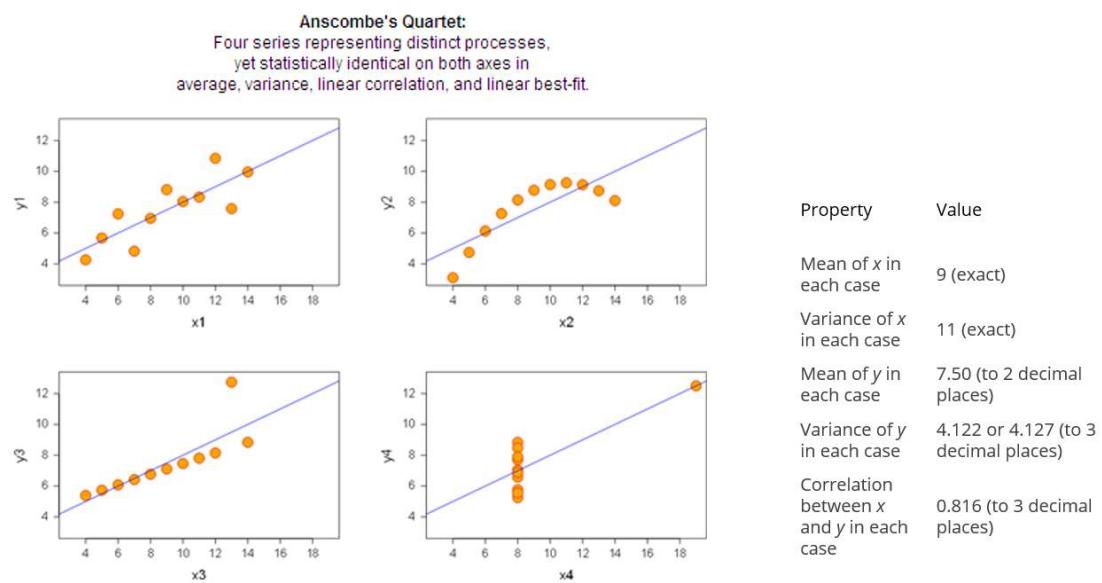


Como era de esperar, el % de ventas de la marca blanca aumenta cuando lo hace la tasa de paro. Además, podemos ver claramente que, cuando la tasa de paro traspasa la barrera del 20%, el % de ventas se dispara. Por lo tanto, tendrá sentido analizar muy bien este comportamiento al modelizar. Posiblemente será necesario modelizar los dos períodos por separado o dividir la variable de tasa de paro para capturar correctamente el impacto que tiene en el negocio.

---

Los gráficos de dispersión son útiles también para no cometer errores de interpretación de las correlaciones.

En el ejemplo siguiente, que recibe el nombre de **cuarteto de Anscombe**, podemos observar que, en los cuatro casos, la correlación es exactamente la misma (0,816), pero que las relaciones entre 'x' e 'y' son totalmente distintas:



## 2

## Análisis de outliers

El descriptivo también nos debe servir para **detectar comportamientos atípicos (outliers)** que puedan desvirtuar los modelos. Los gráficos nos ayudarán a entender si hay algún posible error en los datos o si ha habido algún comportamiento atípico debido a un factor exógeno.

**Los outliers no tienen que ser exclusivamente ‘cuantitativos’, pueden ser también ‘cualitativos’** (a la hora de diseñar y analizar una encuesta,

debemos de tener cuidado a la hora de seleccionar muestras homogéneas). En la literatura, existen múltiples métodos para la detección de estas observaciones anómalas: **test de Tukey, criterio de Chauvenet, rango intercuartílico, criterio MEDA, Test Q de Dixon...**

El método más impartido académicamente por su sencillez y resultados es el test de Tukey, que toma como referencia la diferencia entre el primer cuartil ( $Q_1$ ) y el tercer cuartil ( $Q_3$ ), o rango intercuartílico. En un diagrama de caja (boxplot) se considera un valor atípico el que se encuentra 1,5 veces esa distancia de uno de esos cuartiles (atípico leve) o a 3 veces esa distancia (atípico extremo).

## 3

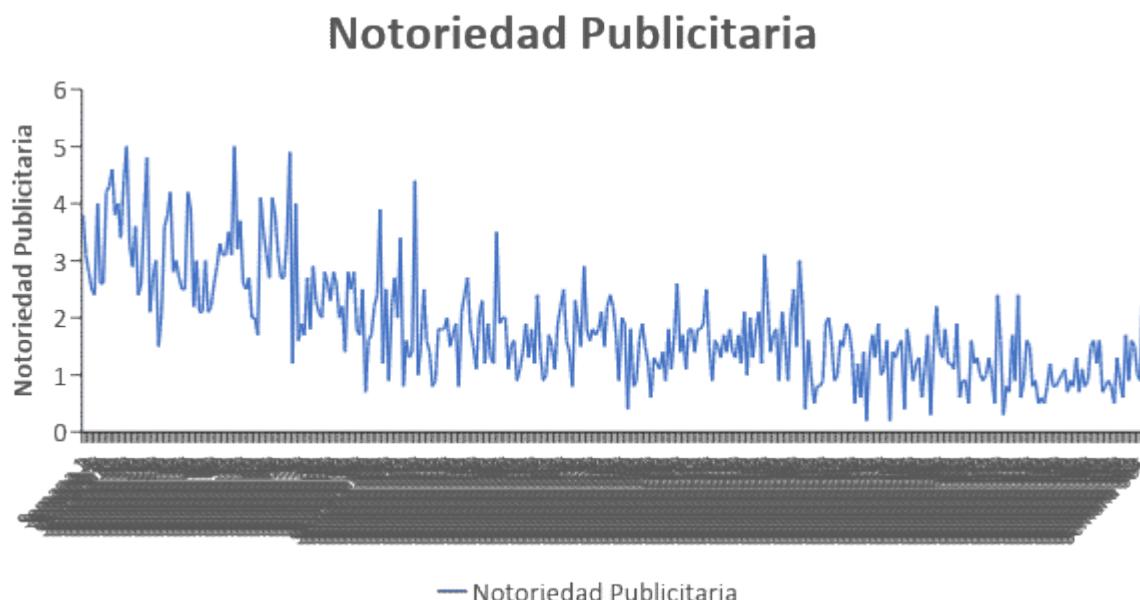
## Suavización de series

Cada fuente de datos es distinta. En algunos casos, el descriptivo nos servirá para determinar cómo tratar o modificar cierta información que presente sesgos.

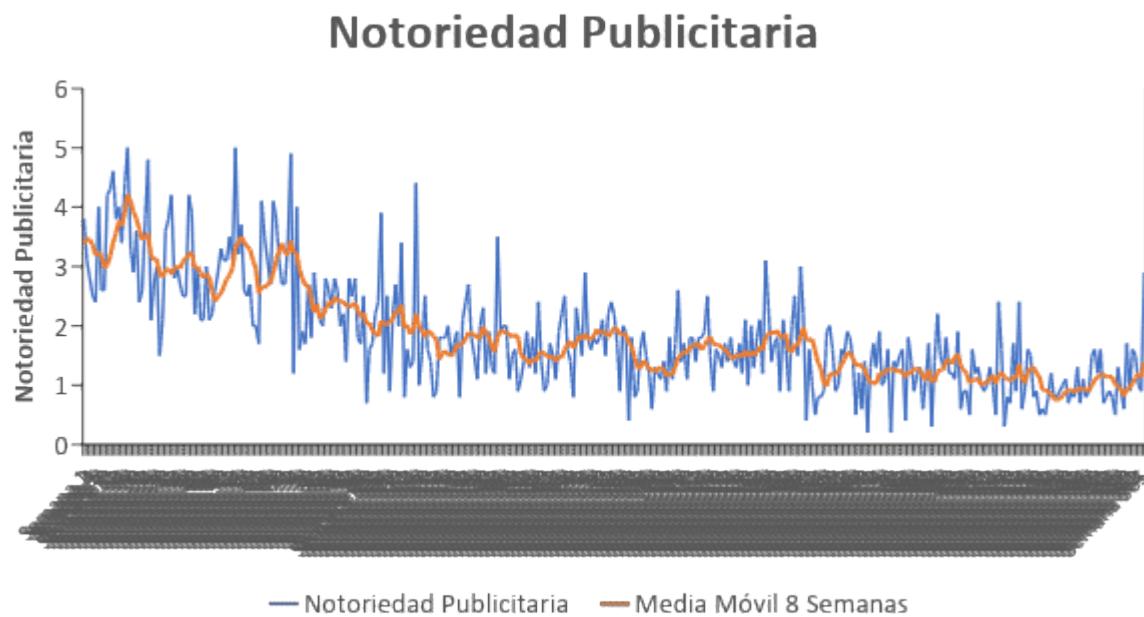
Uno de los procedimientos que se emplean más es la suavización de series temporales.

Pensemos en información que procede de una encuesta. Si la muestra no es muy amplia, es posible que se observen oscilaciones en los datos que no son reales y que, por tanto, tengamos que suavizar para testar en nuestros modelos.

A continuación, un ejemplo de serie temporal de notoriedad publicitaria semanal, es decir, recuerdo de publicidad de una marca declarado por encuestados:



Como se puede observar, la serie presenta fuertes variaciones de una semana a la siguiente. Podemos aplicar una suavización a la serie para testarla en los modelos. Por ejemplo, podemos aplicar una media móvil de las últimas 8 semanas (2 meses):



Este tratamiento de datos realizado durante el análisis descriptivo nos servirá para tener una nueva variable candidata a testar en los modelos.

4

#### Chequeo de datos de fuentes diversas

Al realizar el análisis descriptivo debemos ser capaces de detectar cualquier incoherencia en los datos. **Cuando se tenga información similar de distintas fuentes, se recomienda contrastarla** para comprobar que está alineada. Veamos unos ejemplos:

El anunciante tiene información interna sobre las inversiones reales en medios. Como ya vimos en un fastbook previo, Kantar e Infoadex dan información estimada para cualquier anunciante. Contrastar la fuente interna con estas dos fuentes externas de datos nos servirá para ver si la información está alineada. En algunas ocasiones, se detectan campañas que no han sido almacenadas correctamente de manera interna.

Si se trabaja con información de encuestas, es bueno validar la correspondencia existente entre respuestas a preguntas similares.

El tráfico web se podrá contrastar con fuentes externas como Google Trends, SimilarWeb o similares.

La relación entre las variables a modelizar (ventas, visitas web, notoriedad, etc.) también pueden servir para detectar incongruencias en los datos.

Sin duda, el análisis descriptivo es clave en el desarrollo de medición de ROI a través de MMM. Posiblemente es la fase más importante. Es recomendable dedicar tiempo suficiente a esta fase para no tener que repetir trabajo en la fase de modelización.

Un buen análisis descriptivo nos servirá para descubrir patrones, detectar errores y sugerir nuevas variables a construir. En el análisis descriptivo tenemos que sentirnos como un detective que está buscando pistas para resolver el caso. En MMM, el análisis descriptivo dará todas las ‘pistas’ para, posteriormente, elaborar buenos modelos que se adapten al negocio y a lo que nos muestran los datos.

**¡Enhorabuena! Fastbook superado**

edix

Creamos Digital Workers