

Fastbook 06

Medición de Eficacia Publicitaria

Modelización e interpretación
de resultados (MMM)



06. Modelización e interpretación de resultados (MMM)



En los fastbooks anteriores hemos visto las variables que se suelen emplear en modelos de MMM. También hemos hablado sobre los puntos a cubrir en el análisis descriptivo inicial para entender mejor los datos. Recordemos que el análisis descriptivo tiene que servir también para detectar errores e incongruencias en los datos, por lo que se tendrá que dedicar tiempo suficiente a realizarlo.

Una vez que hemos explorado los datos y confiamos en la información disponible, es la hora de **comenzar con la modelización**. En este **fastbook veremos un ejemplo detallado de modelo de MMM a partir de regresión múltiple e interpretaremos los resultados**. Hay modelos más sencillos y complejos, pero la finalidad de todos ellos es similar: **medir el retorno de las inversiones publicitarias**. Por esta razón, dedicaremos una lección de este fastbook a interpretar los resultados y comparar el retorno de las inversiones.

Autores: Carlos Real Ugena y Beatriz Asenjo

[Introducción al ejemplo de modelización](#)

[Análisis descriptivo de la información](#)

[Modelización de ventas](#)

[Interpretación de resultados](#)

[Retorno publicitario](#)

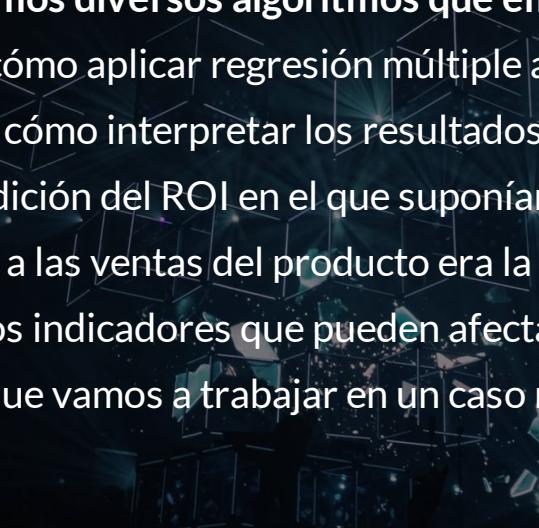
Lesson 1 of 5

Introducción al ejemplo de modelización

X Edix Educación

Por fin hemos llegado a la lección de modelización. Como se ha repetido en varias ocasiones, es importante haber completado las fases previas antes de comenzar con la modelización. La modelización corresponde a las fases de **selección de la tarea de modelización, elección del algoritmo, aplicación del algoritmo y evaluación** según el framework KDD.

Como ya sabemos, en el enfoque de MMM, **utilizaremos modelos de regresión y tendremos diversos algoritmos que emplear**. Para empezar, veremos cómo aplicar regresión múltiple a nuestro conjunto de datos y veremos cómo interpretar los resultados. Ya vimos un caso muy sencillo de medición del ROI en el que suponíamos que el único factor que afectaba a las ventas del producto era la publicidad. Ahora ya conocemos todos los indicadores que pueden afectar al negocio de un anunciante, por lo que vamos a trabajar en un caso más cercano a la realidad.



Supongamos que trabajamos para una famosa marca de **bebidas refrescantes** y nos solicita **desarrollar un modelo de MMM para estimar la contribución de la publicidad en las ventas**. El objetivo final es estimar el retorno invertido en cada uno de los dos medios que ha empleado: televisión y online.

A partir de ahora nos centraremos en este ejemplo. Para ello, se proporciona un fichero en Excel de apoyo con los datos que se analizarán y con la resolución del problema. Es un ejemplo inventado para ilustrar mejor los conceptos de este fastbook.



Descárgate el siguiente fichero de [ejemplo de modelización](#).

El Excel tiene cuatro pestañas:

BBDD inicial



Contiene la información relevante para la modelización.

Ad-stock



Contiene las transformaciones de las series de medios y sus correlaciones.

Salida del modelo de regresión

Detalles del modelo de regresión.

Outputs del modelo

Modelo de regresión final para explicar las ventas.

Lo primero que necesitamos es **tener algo de contexto sobre el negocio**. En gran consumo, el brand manager (responsable de la marca) nos podrá trasladar su conocimiento de la marca para poder tenerlo en cuenta al desarrollar la modelización. En nuestro caso, supongamos que nos ha proporcionado esta información inicial sobre el negocio y las variables disponibles (pestaña BBDD inicial):

1

El **producto a modelizar (bebida refrescante)** se vende en supermercados e hipermercados en España.

2

Se dispone de **80 meses de histórico** que van desde enero del año 1 hasta agosto del año 7. Los datos son mensuales.

3

Las ventas son en miles de litros y el resto de las variables disponibles son las siguientes:

- **Distribución ponderada:** recoge el % de tiendas en el que se vende el producto. Cabe destacar que, a partir de junio del año 5, la marca deja de venderse en una importante cadena de supermercados y esto provoca una bajada en la distribución.
- **Viajeros:** es una marca muy consumida en países de los que España recibe gran cantidad de turistas. Por este motivo, se estima que la presencia de turistas puede fomentar el incremento de las ventas. Además, al ser una bebida refrescante, esta variable recoge también el efecto de la temperatura: a más calor, más ventas.
- **Intensidad promocional (en %):** es el porcentaje de ventas que se realizan en promoción.
- **Precio (€/l):** es el precio por volumen de la marca. Recordemos que preferimos utilizar precio por unidad de volumen que precio por unidad al construir modelos de MMM.
- **GRPs TV 20”:** el anunciante ha tenido cuatro campañas de TV a lo largo del histórico. Los GRPs ya están transformados a 20” para que los resultados sean comparables. A la hora de estimar el ROI, supondremos que el coste por GRP ha sido de 1200€, es decir, ha habido una inversión de 4,332 millones de euros en todo el histórico.
- **Inversión online:** el otro medio empleado por el anunciante ha sido online. En este caso, la inversión total ha sido de 2,1 millones de euros en todo el histórico.

Lo normal será que un anunciante emplee más medios publicitarios, pero mantenemos únicamente dos en este ejemplo para simplificar la interpretación de resultados. Si tuviéramos más medios, se trabajaría de manera análoga a como lo vamos a hacer.

Lesson 2 of 5

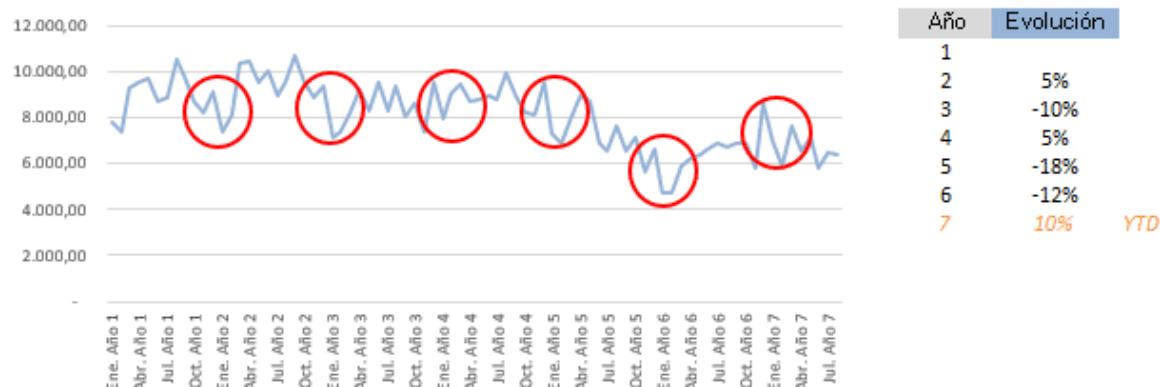
Análisis descriptivo de la información

X Edix Educación

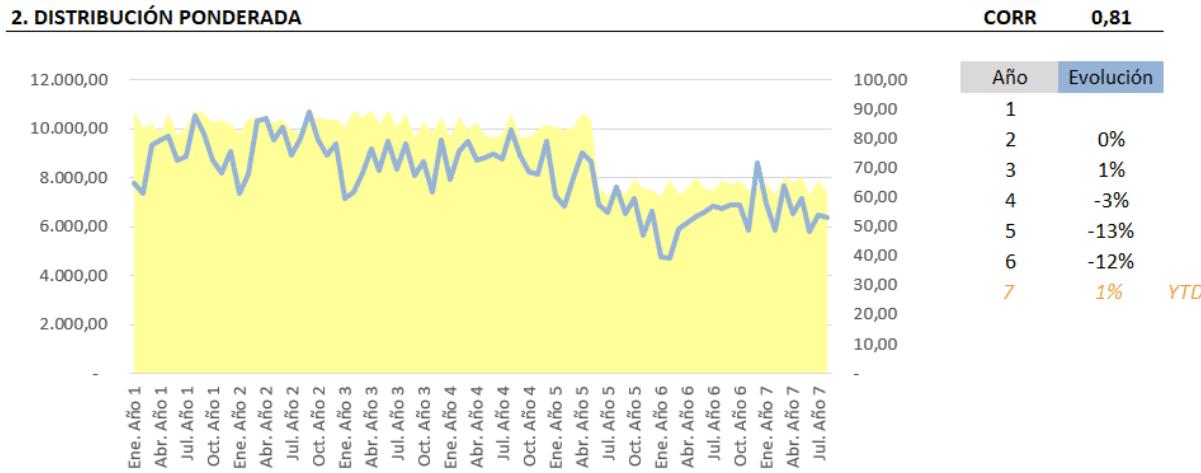
Antes de desarrollar el modelo, vamos a cubrir un **análisis descriptivo** muy sencillo basándonos en lo aprendido en el fastbook anterior.

Lo primero que observamos es una tendencia decreciente en las ventas con una caída más pronunciada en el año 5. Además, parece existir un comportamiento similar con aumentos en los meses de diciembre, seguidos por una caída en enero y febrero.

1. EVOLUCIÓN DE LAS VENTAS



En los modelos de gran consumo la distribución ponderada suele funcionar como nivel base, debido a su correlación con las ventas 0,81. Como sabemos, a partir del año 5 la marca deja de estar presente en una importante cadena de supermercados, lo que ha afectado de manera negativa a las ventas.

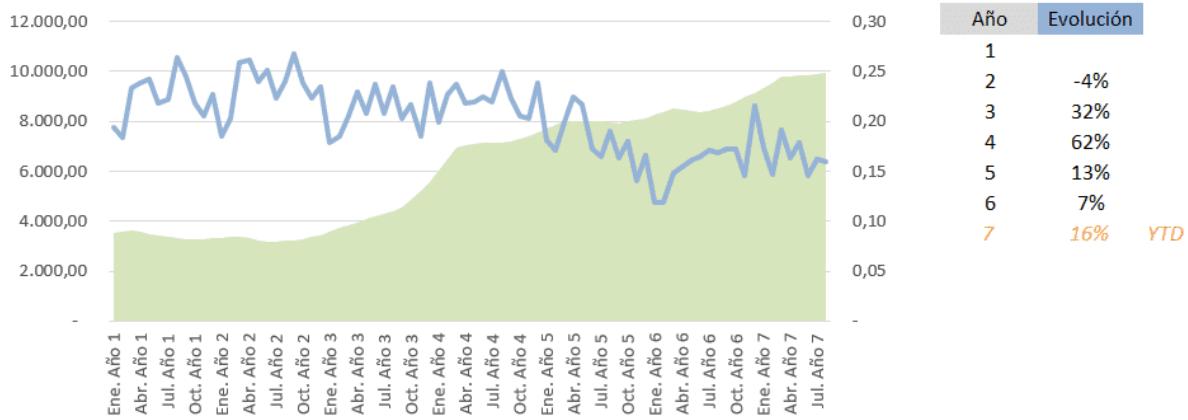


La intensidad promocional no tiene el signo esperado. La correlación es -0,7 y nos querría decir que afecta de manera negativa a las ventas.

El gráfico nos desvela el motivo de la correlación negativa. El porcentaje de promoción se ha ido incrementando a lo largo del tiempo. En cambio, las ventas han bajado, provocando una correlación negativa. Sin embargo, ya vimos previamente que las ventas se veían muy afectadas por la salida de la marca de una cadena de supermercados y es el motivo principal de la bajada en ventas, no porque se realice más promoción.

3. INTENSIDAD PROMOCIONAL

CORR -0,67

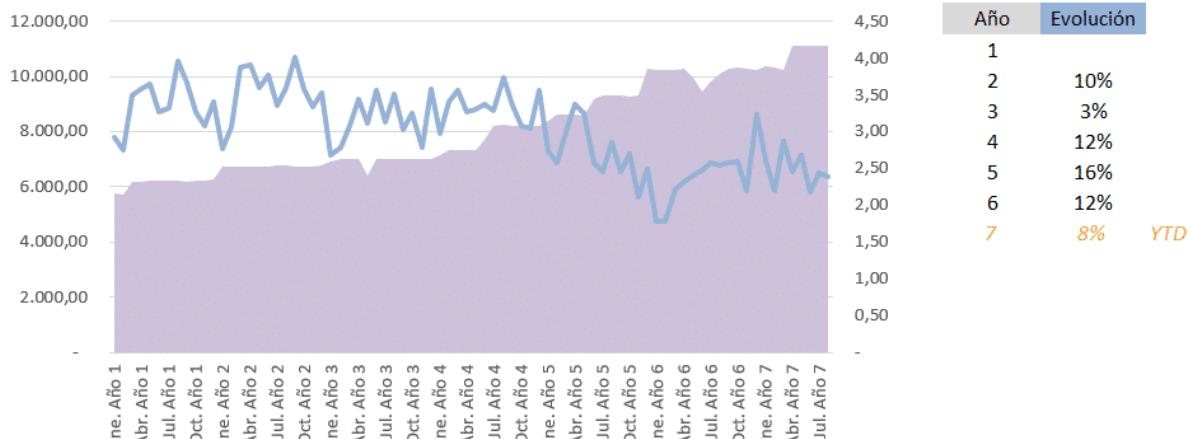


En el modelo testaremos ambas variables (distribución ponderada y % ventas en promoción) para ver cuál es el resultado de los coeficientes y p-valores.

Este es un ejemplo que ilustra la necesidad de entender y graficar los datos antes de comenzar la modelización. Si no supiéramos que la distribución del producto se ha reducido, podríamos pensar que la promoción no genera un incremental de ventas.

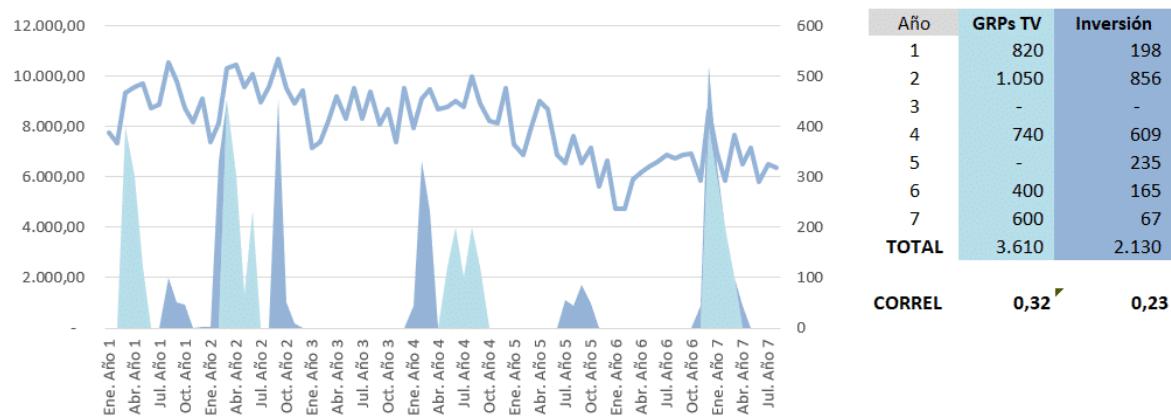
4. PRECIO

CORR -0,74



Cabe destacar que tiene sentido que la correlación entre las ventas y el precio sea negativa. Esto quiere decir que cuando hay incrementos de precio coincide con decrementos en el volumen de ventas.

5. MEDIOS



Por último, la correlación de las ventas y la publicidad es positiva. Es una correlación menor que para otras variables como la distribución o el precio. Esto es normal porque, como podemos ver, los GRPs TV o inversión en online no son continuas en el tiempo y toman valores o en bastantes meses del histórico.

En general, la correlación entre ventas y publicidad se situará entre +0,15 y + 0,35. No debemos preocuparnos si son cercanas a 0, ya que, al aplicar el efecto recuerdo y al combinarlas con otras variables, pueden resultar significativas en nuestro modelo.

| | Ventas (miles de litros) |
|--------------------------|--------------------------|
| Ventas (miles de litros) | 1,000 |
| GRPs TV 20 ¹¹ | 0,318 |
| Inversión Online | 0,226 |
| GRPS_ad_40 | 0,335 |
| GRPS_ad_45 | 0,339 |
| Inversión Online_ad_15 | 0,253 |
| Inversión Online_ad_20 | 0,262 |

En ambos casos, la correlación aumenta al aplicar el ad-stock, por lo que nos decantaremos por TV con un ad-stock del 45% e inversión online con un ad-stock del 20%.

¡Cuidado! A partir de ahora trabajaremos con transformadas de las variables originales. Esto debemos tenerlo en cuenta al calcular el ROI de las inversiones.

Lesson 3 of 5

Modelización de ventas

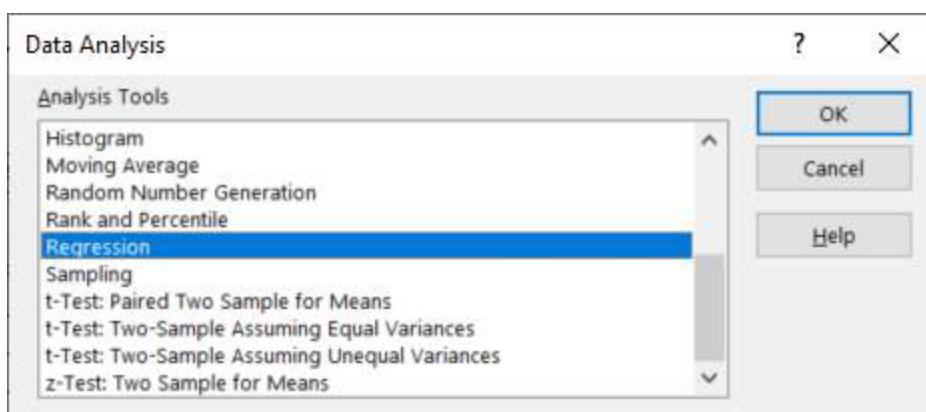
X Edix Educación

Pasamos ahora a la modelización de las ventas en la pestaña ‘Salida del modelo de regresión’.

Para la técnica de MMM, recordemos que necesitamos aplicar un modelo de regresión. En nuestro caso aplicaremos un modelo de regresión múltiple. Este tipo de modelo lo podemos ejecutar directamente en Excel, pero si quisiéramos aplicar uno más complejo, tendremos que emplear software estadístico como R o Python.

1

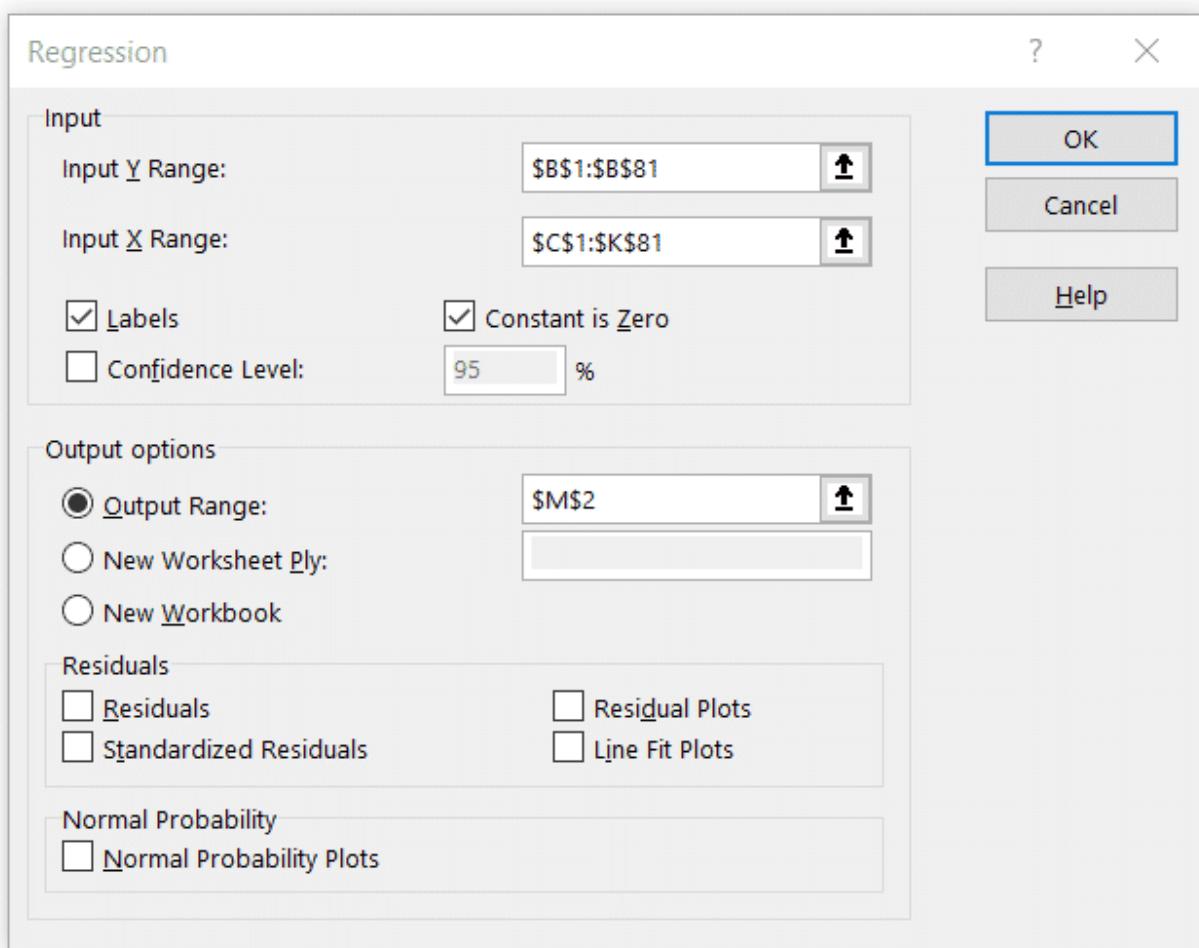
Seleccionamos ‘Regression’ (regresión) del menú de data analysis:



2

Seleccionamos la columna ventas en volumen como variable Y y el resto de las variables explicativas con X. En nuestro caso marcaremos la casilla 'Constant is

Zero' (ya que la distribución ponderada hará de baseline) y 'Labels' (ya que los nombres de las variables están en la primera fila). Mostraremos los resultados en la misma pestaña a la derecha de los datos:



3

Ahora podemos analizar los resultados.

SUMMARY OUTPUT

| <i>Regression Statistics</i> | | | | | | | | |
|------------------------------|--|----------|--|--|--|--|--|--|
| Multiple R | | 0,999414 | | | | | | |
| R Square | | 0,998829 | | | | | | |
| Adjusted R Square | | 0,984612 | | | | | | |
| Standard Error | | 297,0496 | | | | | | |
| Observations | | 80 | | | | | | |

| <i>ANOVA</i> | | | | | | |
|--------------|-----------|-----------|-----------|-------------|-----------------------|--|
| | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Significance F</i> | |
| Regression | 9 | 5,34E+09 | 5,94E+08 | 6727,957658 | 4,1E-99 | |
| Residual | 71 | 6264930 | 88238,45 | | | |
| Total | 80 | 5,35E+09 | | | | |

| | <i>Coefficients</i> | <i>Standard Err</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i> | <i>Upper 95%</i> | <i>Lower 95% J</i> | <i>Upper 95% J</i> |
|-------------------------------|---------------------|---------------------|---------------|----------------|------------------|------------------|--------------------|--------------------|
| Intercept | 0 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| Distribución ponderada | 99,44963 | 2,003267 | 49,64373 | 7,18106E-57 | 95,45523 | 103,444 | 95,45523 | 103,444 |
| Intensidad Promocional (en %) | 2509,508 | 1556,184 | 1,612603 | 0,111267267 | -593,4355 | 5612,451 | -593,4355 | 5612,451 |
| Precio (€/l) | -363,4912 | 126,9215 | -2,863906 | 0,005498181 | -616,5655 | -110,4169 | -616,5655 | -110,4169 |
| GRPS_ad_45 | 3,012382 | 0,233054 | 12,92569 | 2,48665E-20 | 2,547686 | 3,477078 | 2,547686 | 3,477078 |
| Inversión Online_ad_20 | 4,337565 | 0,425816 | 10,18649 | 1,57247E-15 | 3,488512 | 5,186617 | 3,488512 | 5,186617 |
| Viajeros | 3,86E-05 | 5,3E-06 | 7,270071 | 3,74462E-10 | 2,8E-05 | 4,91E-05 | 2,8E-05 | 4,91E-05 |
| Diciembre | 1080,889 | 144,6078 | 7,474627 | 1,57047E-10 | 792,5494 | 1369,229 | 792,5494 | 1369,229 |
| Enero | -580,9304 | 138,9082 | -4,182116 | 8,13236E-05 | -857,9056 | -303,9553 | -857,9056 | -303,9553 |
| Febrero | -1137,857 | 137,2628 | -8,289625 | 4,85455E-12 | -1411,551 | -864,1629 | -1411,551 | -864,1629 |

El R^2 no es válido al no estar incluyendo 'Intercept' (constante) en el modelo. Cualquier software estadístico espera que el modelo tenga constante, por lo que el R^2 calculado no será correcto. En la celda AJ5 de la pestaña 'Modelo MMM resuelto' está el R^2 real calculado a partir de la fórmula.

En nuestro caso, tenemos un R^2 del 95,9%. Esto quiere decir que estamos ajustando bien el comportamiento de la serie de ventas.

Lo siguiente es analizar los coeficientes. Todos tienen sentido de negocio. Podemos clasificar las variables en positivas o negativas en función del coeficiente obtenido:

- Variables con coeficientes positivos: distribución ponderada, viajeros, intensidad promocional, diciembre, GRPs TV e inversión online.
- Variables con coeficientes negativos: precio (€/l), enero y febrero.



Cabe destacar que el coeficiente de la variable de 'Intensidad promocional' ha resultado ser positivo. Recordemos que la correlación era negativa debido a la bajada de distribución.

Y terminamos con los **p-valores**. Los p-valores nos indican la probabilidad de cometer error al suponer que las variables afectan a las ventas, por lo que buscaremos que sean lo más bajos posible. Excepto la variable de 'Intensidad promocional', todas tienen p-valores muy bajos que nos indican la significatividad de las variables.

4

Pero, ¿debemos eliminar la 'Intensidad promocional' de nuestro modelo? Aquí habrá opiniones de todo tipo. Desde el punto de vista teórico, habitualmente se buscará que los p-valores sean menores a 0,05, pero esto a veces resulta complicado en la práctica. En nuestro caso, estamos ante un p-valor de 0,10 que no es muy alto, por lo que mi recomendación sería analizar el motivo de este p-valor.

5

Si nos fijamos en la **correlación de la intensidad promocional con el resto de las variables**, comprenderemos el motivo. La correlación entre la intensidad promocional y la distribución ponderada es alta. Es -0,8, es decir, muy próxima a -1. Como ya vimos previamente, la distribución ha bajado y la promoción se ha incrementado. Cuando hay variables correladas, es normal que alguna de ellas tenga un p-valor mayor.

6

Teniendo en cuenta lo anterior y que el coeficiente de la promoción es positivo, vamos a dejar la variable en el modelo. Sin duda, desde un punto de vista de negocio, la promoción debería afectar a las ventas, por lo que decidimos no eliminar la variable. Para practicar, se recomienda lanzar el modelo de nuevo sin esta variable, analizar los resultados y ver las diferencias.

Lesson 4 of 5

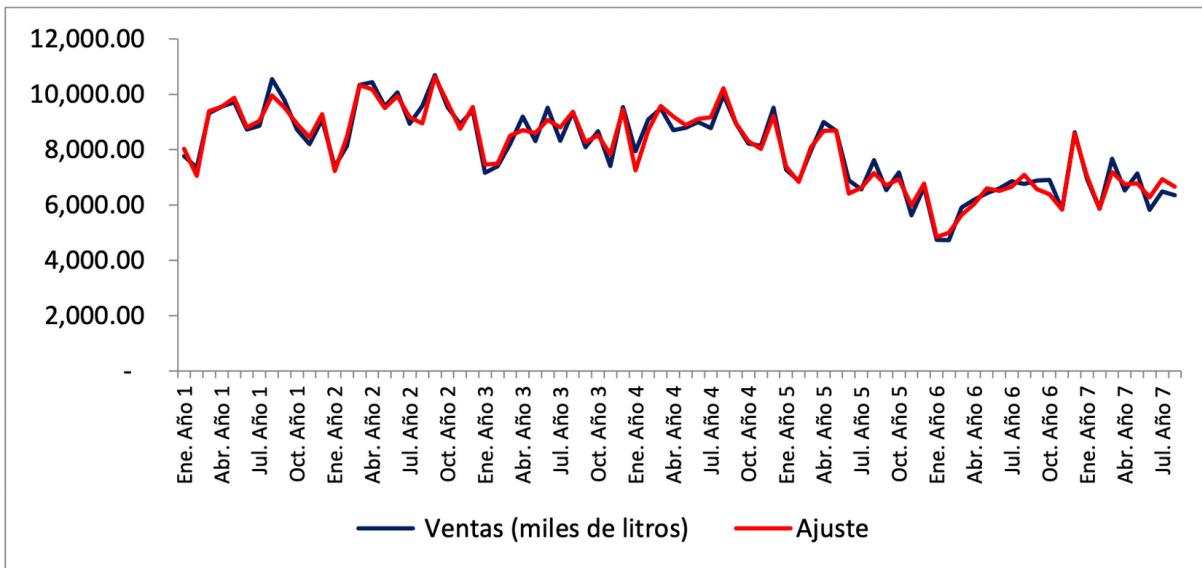
Interpretación de resultados

X Edix Educación

Ahora vamos a utilizar la pestaña ‘Modelo MMM resuelto’ para explotar e interpretar los resultados del modelo. Lo primero que hacemos es pegar los coeficientes transpuestos a partir de la celda ‘G7’. A partir de la columna ‘W’ tendremos los aportes de cada una de las variables.

Estos aportes, también llamados contribuciones, se obtienen al multiplicar cada variable real por el coeficiente obtenido en el modelo. De esta manera tenemos para cada mes del histórico un valor de aporte de cada variable.

Si sumamos los aportes (columnas ‘W’ a ‘AE’), obtendremos el ajuste del modelo que guardamos en la columna ‘AI’ a partir de la fila 9. Esto nos da lugar al gráfico de ajuste:

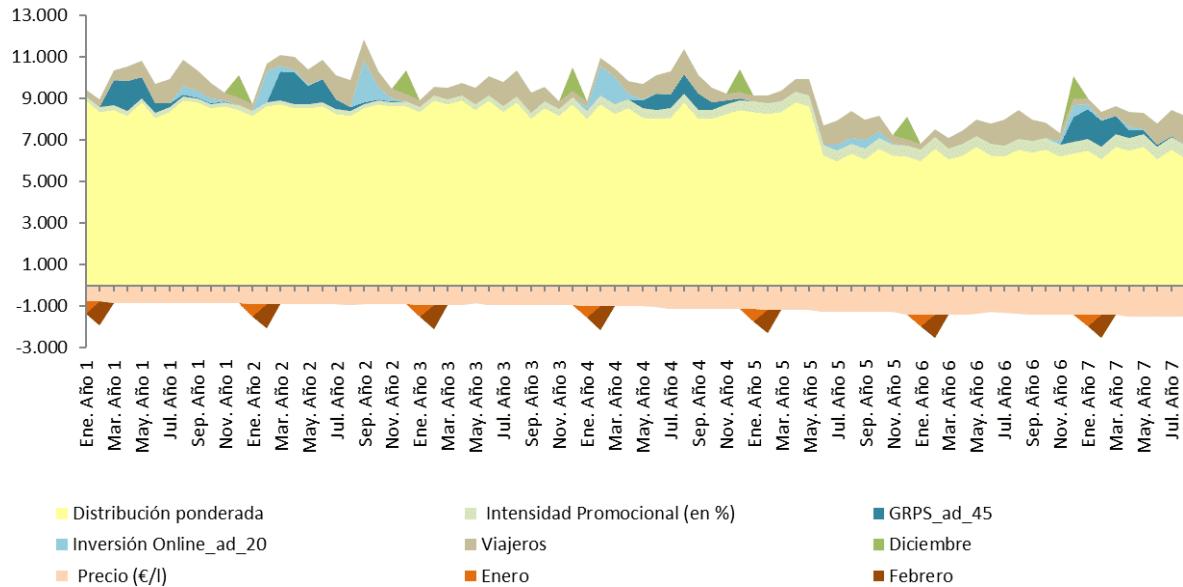


Como podemos ver, el ajuste es bueno y obtenemos un R² de 95,9% (celda AK5). Si el ajuste fuera peor, podríamos seguir testando variables que nos explicaran el comportamiento de las ventas.

En general, un ajuste por encima del 80% se considerará bueno. Dependiendo del mercado, será más fácil o difícil alcanzar dicho ajuste. En el caso de gran consumo que estamos analizando, será relativamente fácil. Dispondremos de las 4p's de MMM bien medidas (distribución, precio, promoción y publicidad) y esto será una gran ventaja respecto a otros mercados en los que resulta más difícil recopilar la información.

Además, no hemos necesitado crear ninguna tendencia exógena como vimos en el fastbook anterior. Estamos ante la mejor situación dado que hemos sido capaces de explicar la bajada en ventas con la información disponible.

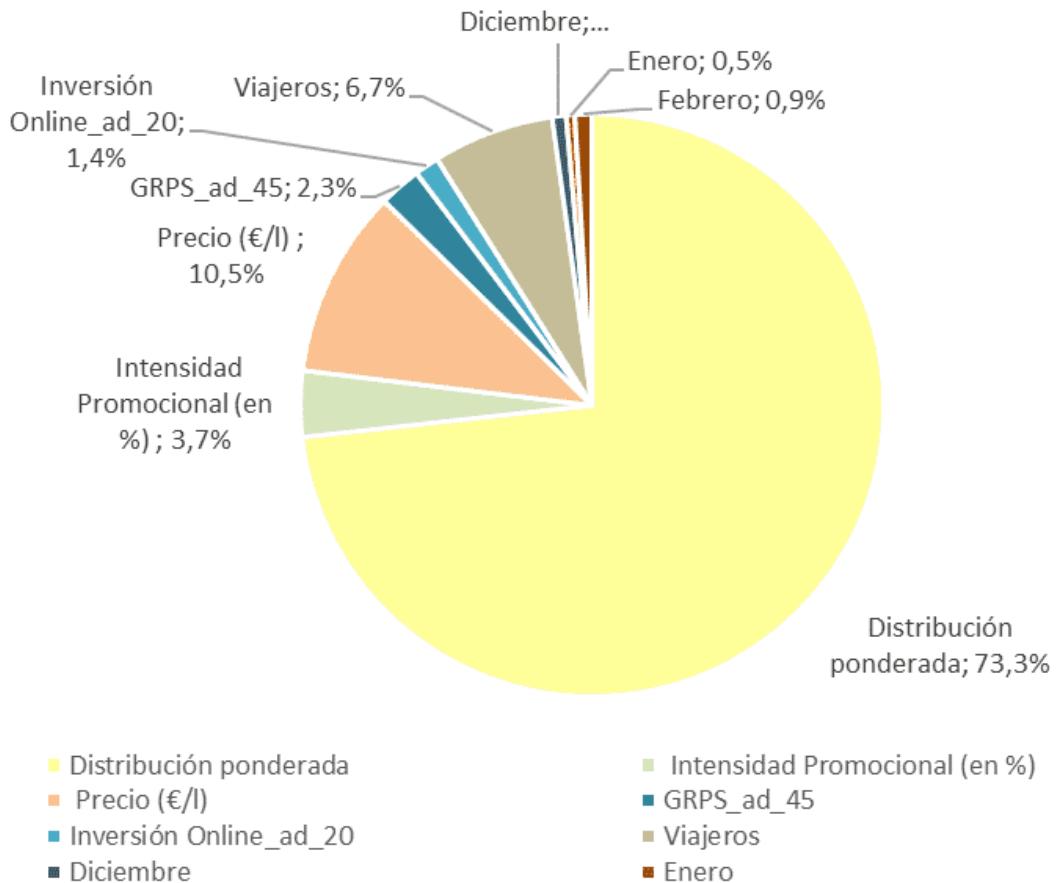
Una vez que hemos analizado los coeficientes, p-valores y ajustes, toca el turno del **gráfico de aportes**. Este gráfico representa a lo largo del tiempo la contribución de cada variable:



Visualmente ya podemos ver las variables más importantes. La distribución ponderada funciona como el nivel base de la serie y podemos observar el impacto de cada variable a lo largo del tiempo. **Las variables de publicidad en televisión y online aportan ventas en los períodos que se han realizado inversiones.**

¿Qué podemos hacer con estas áreas?

Podemos representar la influencia de cada variable en un gráfico de tartas.



Por tanto, tenemos nuestro primer gran resultado: el peso de la publicidad en las ventas es del 3,7%. Además, ya sabemos el desglose por medios:

- La televisión aporta el 2,3% de las ventas.
- El online aporta el 1,4% de las ventas.

Lesson 5 of 5

Retorno publicitario

X Edix Educación

De acuerdo con el gráfico de tartas anterior, alguien podría estar tentado a pensar que el ROI de la televisión es mayor que online, pero recordemos que la inversión en televisión había sido mayor que online en el histórico analizado.

Siempre que queramos **calcular el ROI** a través de un modelo de MMM o de otro enfoque necesitaremos tener dos datos:

- Inversión publicitaria.
 - Aporte en ventas.
-

La inversión publicitaria es directa porque es un dato que tendrá internamente recogido el anunciante. En cambio, el aporte costará más obtenerlo.

En nuestro caso, hemos construido un modelo de MMM de ventas a partir de diversas variables con el objetivo de aislar el efecto de cada una de ellas y ‘desgranar’ la contribución de la publicidad.

| | Inversión | Aporte en ventas (en miles de litros) | ROI (por 1.000€) |
|--------|-------------|--|------------------|
| TV | 4.332.000 € | 19.759 | 4,6 |
| Online | 2.130.000 € | 11.549 | 5,4 |

Como tenemos la contribución de cada medio (televisión y online), podemos calcular la contribución y el ROI de cada uno.

Esta información está recogida a partir de la columna 'AO' de la pestaña 'Modelo MMMresuelto' que hemos estado trabajando.

En resumen, podemos decir lo siguiente:

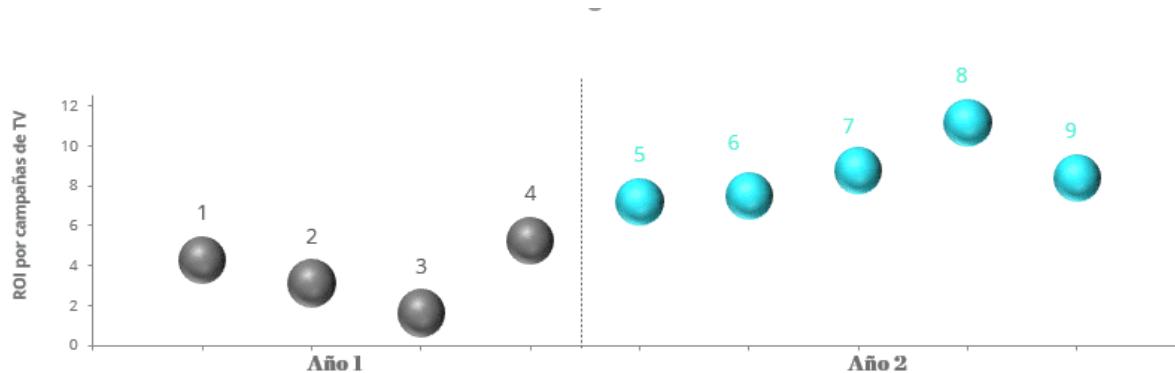
- **La televisión ha sido el medio que más ventas ha generado a lo largo del tiempo.** En particular, ha generado 19,7k vs. 11,5k ventas.
- **Sin embargo, online ha sido capaz de generar más ventas por cada € invertido y, por tanto, ha alcanzado un mayor ROI.** El ROI de online ha sido de 5,4, mientras que el de televisión ha sido de 4,6.

El paso siguiente sería medir el ROI por campañas e introducir más variables de medios en el modelo. En nuestro caso hemos trabajado con un ejemplo sencillo mensual para que se entienda bien la metodología, pero podríamos complicarlo hasta que queramos.

Cada vez más, se están desarrollando estos modelos a nivel diario y con datos geográficos, dando lugar a insights (aprendizajes) más accionables para los anunciantes.

Por ejemplo, esta tipología de modelos puede servir para determinar el ROI de distintos patrones creativos empleados en las campañas de televisión.

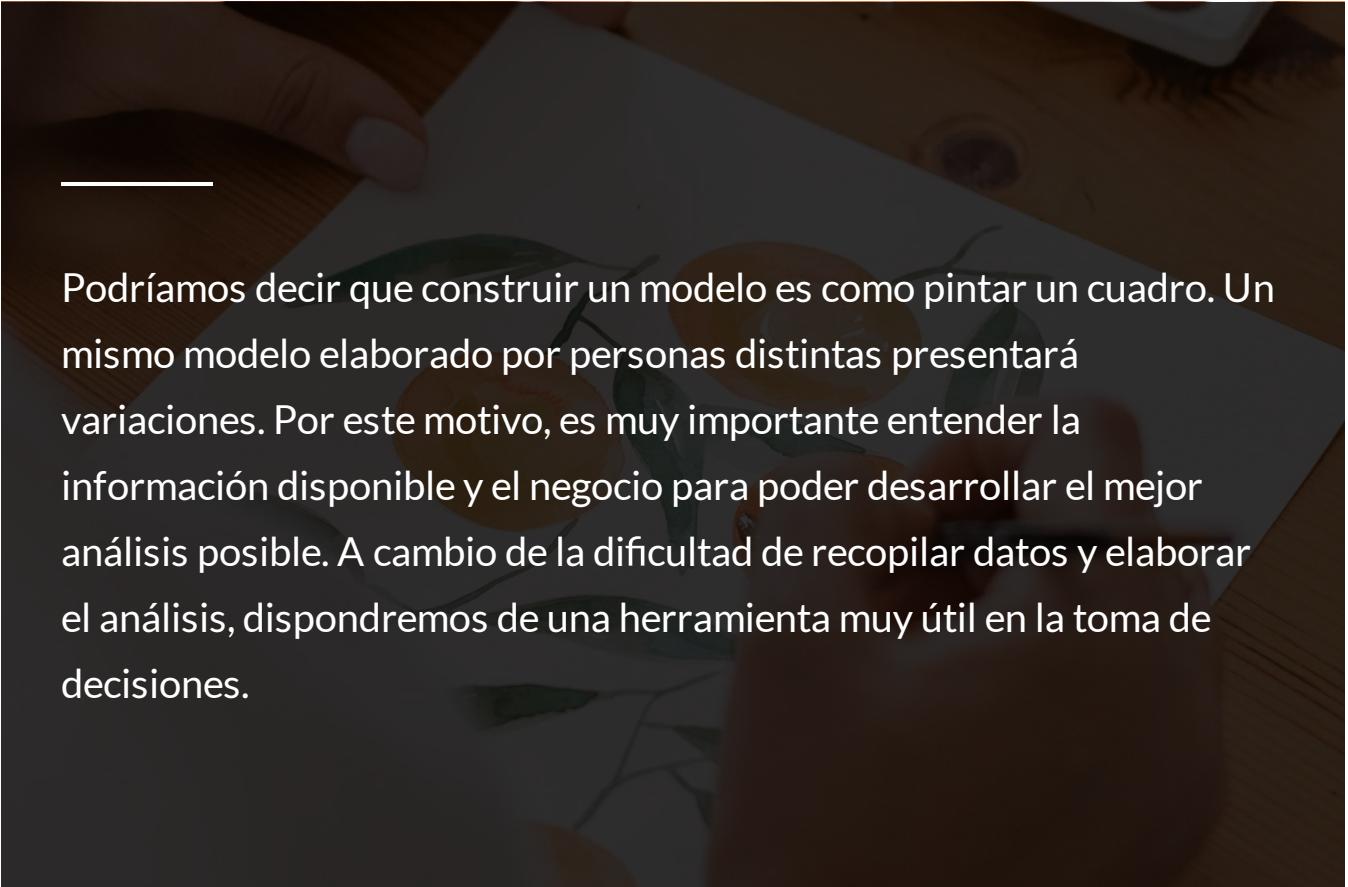
Veamos un ejemplo a continuación:



De las 9 campañas realizadas durante los dos años, la campaña de televisión número 8 es la que ha generado mayor retorno. Estos resultados nos podrían servir para investigar cómo se puede generar mayor ROI con las campañas.

Creatividad, nivel de promoción, aparición de famoso o costes por GRP son algunas de las palancas que habitualmente afectan al ROI.

En resumen, un buen modelo nos dará respuesta a interrogantes sobre el negocio. Modelizar no es sencillo y requiere de conocimientos técnicos, de negocio y experiencia previa.



Podríamos decir que construir un modelo es como pintar un cuadro. Un mismo modelo elaborado por personas distintas presentará variaciones. Por este motivo, es muy importante entender la información disponible y el negocio para poder desarrollar el mejor análisis posible. A cambio de la dificultad de recopilar datos y elaborar el análisis, dispondremos de una herramienta muy útil en la toma de decisiones.

“Sin datos, no es usted más que otra persona con una opinión.”

- W. Edwards Deming.

Los modelos de MMM permiten contrastar esas opiniones y validar si son ciertas, dejando a un lado la opinión o la intuición.

¡Enhорabuena! Fastbook superado

edix

Creamos Digital Workers