

## Caso de Uso: “Modelización del Tráfico Web y Atribución de Inversión Publicitaria en Retail de Gran Consumo”

### 1.- Contexto de Negocio

El sector del retail de gran consumo, especialmente en moda y hogar, se caracteriza por una alta dependencia de la estacionalidad y las tendencias del mercado. Las grandes marcas operan con una estrategia omnicanal donde las ventas en tiendas físicas y el tráfico online están estrechamente relacionados con campañas publicitarias, estrategias de precios y eventos comerciales.

La empresa en estudio es una compañía internacional que se enfoca en ofrecer productos de moda y hogar a precios accesibles. Su estrategia de marketing combina inversión en medios tradicionales (prensa, radio, cine) y digitales (redes sociales, publicidad online, plataformas de video). La atribución del impacto de estas inversiones en las ventas, es clave para la optimización del presupuesto publicitario y la maximización de ingresos.

Según los datos disponibles en el archivo, los productos se pueden agrupar en las siguientes categorías:

1. **Ropa:** Hombres, Mujeres, Interior
2. **Complementos:** Accesorios de moda
3. **Calzado:** Zapatos
4. **Hogar:** Productos de decoración y artículos para el hogar
5. **Otros:** Cualquier producto no categorizado

Dado que la empresa sigue una estrategia **Click & Collect**, la atribución del impacto de los canales de marketing debe considerar no solo las ventas directas online y en tienda, sino también las conversiones híbridas, donde los clientes compran en línea pero recogen en tienda. Esto implica que el modelo de atribución de marketing mix (MMM) debe implicar:

- **Click & Collect** (compra online, recogida en tienda o punto de recogida).
- **Ventas offline puras** (compra en tienda sin interacción digital previa).

### 2.- Descripción de los Datos

Los datos disponibles están organizados en diferentes pestañas que contienen información clave para el análisis:

- **WEB:** Contiene datos del tráfico web, incluyendo visitas únicas, descargas de catálogos, uso de configuradores de productos y fuentes de tráfico (redes sociales, búsqueda orgánica, pago por clic, etc.).
- **VISIT:** Representa el tráfico de visitas por categoría de producto en el sitio web.
- **OFFLINE:** Muestra información sobre visitas a tiendas, ventas, ticket promedio y distribución de compras por categoría.

- **TIME:** Incluye datos temporales como días hábiles, fines de semana y eventos comerciales (Semana Santa, por ejemplo).
- **INV:** Registra la inversión en diferentes canales publicitarios, incluyendo medios digitales y tradicionales.

### 3.- Objetivo del Caso de Uso

#### Objetivo 1

Se construirá un modelo de series temporales SARIMA para predecir el tráfico web a 6 meses. Se utilizarán las siguientes variables de la pestaña **WEB**:

- **Variable a predecir:** Unique\_visitors (visitantes únicos en la web).
- **Variables explicativas:** Se considerarán factores como días hábiles, fines de semana y eventos comerciales (de la pestaña TIME), además de tendencias observadas en los últimos periodos.

El modelo permitirá prever la evolución del tráfico en función de patrones históricos, identificando tendencias estacionales y variaciones anuales.

#### Objetivo2

El modelo de atribución analizará la relación entre la inversión publicitaria y los resultados comerciales, optimizando tanto el tráfico físico a las tiendas como la tasa de conversión de visitas en ventas. Dado que la estrategia de la empresa se basa en un enfoque Click & Collect, el modelo también debe identificar qué canales impulsan la combinación de tráfico digital y físico. El objetivo es responder a preguntas clave como:

- ¿Qué canales de inversión publicitaria generan más tráfico web y físico?
- ¿Cuáles contribuyen más a la conversión de visitas en ventas?
- ¿Cómo redistribuir la inversión para maximizar la rentabilidad?

### 4.- Variables del modelo

**Variables de inversión publicitaria (INV):** Estos son los diferentes medios en los que se distribuye el presupuesto de marketing:

- **Digitales:** Internet (publicidad online), plataformas de video, redes sociales.
- **Tradicionales:** Cine, exterior (carteles publicitarios), prensa, radio, revistas.
- **Otros:** Producción de materiales y gastos diversos en marketing.

#### Tráfico web (WEB):

- **Visitantes únicos:** Número total de usuarios que acceden al sitio web.
- **Fuentes de tráfico:** Social networks, búsqueda pagada, búsqueda orgánica, email marketing, otros referidos.
- **Interacciones:** Descargas de catálogos, uso del configurador de productos.

**Tráfico físico y conversión a ventas (*OFFLINE*):**

- **Visitantes en tienda:** Número de personas que ingresan a la tienda física.
- **Ventas totales:** Cantidad de transacciones realizadas en tienda.
- **Categorías de compra:** Ropa, calzado, complementos, hogar, entre otros.
- **Ticket medio:** Valor promedio de compra por cliente.

**Factores temporales (*TIME*)**

- **Días hábiles vs. fines de semana:** Influencia del calendario laboral y festivo en la demanda.
- **Eventos comerciales:** Semana Santa, rebajas de temporada, Black Friday, entre otros.

**Visitas web en el showroom (*VISIT*):** contiene información sobre el tráfico web segmentado por categorías de producto. Las variables disponibles incluyen:

- **RopaHombre:** Número de visitas a productos de ropa para hombre.
- **RopaMujer:** Número de visitas a productos de ropa para mujer.
- **Complementos:** Accesorios y complementos de moda.
- **Zapatos:** Visitas a la sección de calzado.
- **Home:** Visitas a la sección de productos para el hogar.
- **Interior:** Visitas a la sección de ropa interior.
- **Otros:** Visitas a productos que no entran en las categorías anteriores.
- **Total:** Total de visitas web sumando todas las categorías.

**5.- Metodología de atribución**

El modelo asignará pesos a cada canal de comunicación basado en su impacto en las ventas utilizando una regresión múltiple ajustada. Se emplearán técnicas de Series Temporales, Machine Learning, regresión de Ridge o modelos bayesianos para manejar la multicolinealidad entre canales de inversión. El presupuesto publicitario para la simulación tomará como base el presupuesto de 2024 y se ajustará de la siguiente forma:

- **Primeros 6 meses:** Incremento del **15%** sobre el presupuesto de 2024.
- **Siguientes 6 meses:** Incremento del **21%**.
- **Año siguiente:** Incremento del **25%** sobre el total de 2024.

Este modelo permitirá simular diferentes escenarios de inversión y analizar cómo impactaría en las ventas y el tráfico web.

## 6.- Tablas en Azure

Todas las tablas están en Azure y empiezan con MMM y la cadena de conexión se adjunta a continuación. Se recomienda cargarlas en un entorno SQL para poder hacer consultas. El ID que une las tablas es el ID\_date.

```
DATAEX.MMM01_WEB
DATAEX.MMM02_VISIT
DATAEX.MMM03_OFFLINE
DATAEX.MMM04_TIME
DATAEX.MMM05_INV
```

```
# Configuración de conexión a Azure SQL con AAD
SERVER = 'uaxmathfis.database.windows.net'
DATABASE = 'usecases'
DRIVER = '{ODBC Driver 18 for SQL Server}'
AZURE_SCHEMA='DATAEX'

# Usar autenticación interactiva de Azure Active Directory (abre una
ventana de login)
conn_str =
f"DRIVER={DRIVER};SERVER={SERVER};DATABASE={DATABASE};Authentication=ActiveDirectoryInteractive"
conn = pyodbc.connect(conn_str)
cursor = conn.cursor()
```

Consulta para sacar la información de los datos:

```
SELECT * FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_SCHEMA = 'DATAEX' AND TABLE_NAME in ('MMM01_WEB',
'MMM02_VISIT', 'MMM03_OFFLINE', 'MMM04_TIME', 'MMM05_INV')
```

Requisitos para la realización del trabajo:

- Grupos de máximo 1 personas
- Se puede usar cualquier plataforma con python
- Se debe de presentar un documento de máximo 5 diapositivas (sin incluir portadas). Si se presenta más de 5 hojas el trabajo no será evaluado.
- El trabajo debe de contener una solución viable. No intensificar en la parte técnica.
- El plagio será sancionado con el suspenso automático de la tarea.