

ATRIBUCIÓN MULTICANAL DE CONVERSIONES EN MARKETING DIGITAL

Autor: David Blanco Casasola

Programa: Certificado Big Data & IA



ÍNDICE

1. CRISP-DM • COMPRENSIÓN DEL NEGOCIO.....	6
1.1. Contexto del problema.....	6
1.2. Objetivo del proyecto.....	7
1.3. Alcance y limitaciones.....	7
1.4. Criterios de éxito.....	7
1.5. Supuestos y restricciones.....	7
2. CRISP-DM • ENTENDIMIENTO DE LOS DATOS (síntesis)....	7
3. CRISP-DM • PREPARACIÓN DE DATOS (<i>Feature Engineering</i>)... ..	8
3.1. Objetivo de la fase.....	8
3.2. Construcción de recorridos de usuario (<i>journeys</i>).....	8
3.3. Ingeniería de variables orientada a atribución.....	8
3.4. Justificación metodológica.....	8
4. CRISP-DM • MODELADO.....	9
4.1. Objetivo de la fase.....	9
4.2. Modelos de atribución considerados.....	9
4.2.1. Modelo baseline: Last Click.....	9
4.2.2. Modelo secuencial: Cadenas de Markov.....	9
4.2.3. Modelado de reparto de contribución: Shapley simplificado.....	9
4.3. Estrategia de comparación.....	10
4.4. Criterio profesional: supuestos fuertes y contextos no recomendados... ..	10
4.4.1. Modelo baseline: Last Click.....	10
4.4.2. Modelo secuencial: Markov (análisis estructural de transiciones).....	10
4.4.3. Modelo de reparto: Shapley simplificado.....	10
4.4.4. Tabla resumen de riesgo y contexto no recomendado.....	11
4.5. Riesgos de uso operativo (mala interpretación).....	11
4.5.1. Last Click (baseline).....	11
4.5.2. Markov (estructura de transiciones, sin <i>removal effect</i>).....	11
4.5.3. Shapley simplificado.....	11
4.5.4. Conclusión operativa.....	12
5. CRISP-DM • EVALUACIÓN ORIENTADA A NEGOCIO.....	12
5.1. Enfoque de evaluación.....	12
5.2. Comparación frente al baseline (Last Click).....	12

5.3. Robustez de los modelos ante perturbaciones.....	12
5.3.1. Objetivo del análisis de robustez.....	12
5.3.2. Escenarios de <i>stress test</i> considerados.....	13
5.3.3. Resultados comparativos.....	13
5.3.4. Interpretación de negocio.....	13
5.3.5. Estabilidad del ranking de canales.....	14
5.3.6. Sensibilidad al cambio de recorridos.....	14
5.4. Contexto económico relativo por canal (coste normalizado).....	14
5.4.1. Motivación.....	14
5.4.2. Definición del coste relativo por canal.....	14
5.4.3. Métrica contribución / coste.....	14
5.4.4. Integración con la simulación de presupuesto.....	15
5.5. Simulación de redistribución de presupuesto.....	15
6. CRISP-DM • CONCLUSIONES Y DEPLOYMENT CONCEPTUAL.....	15
6.1. Conclusiones del proyecto.....	15
6.2. Implicaciones para la toma de decisiones.....	15
6.3. Deployment conceptual.....	16
6.4. Limitaciones y líneas futuras.....	16
7. ANEXO TÉCNICO: NOTEBOOK REPRODUCIBLE.....	16
8. RESUMEN EJECUTIVO DEL PROYECTO.....	16

PRIMEROS PASOS / ALCANCE

Tema: Atribución multicanal de conversiones.

Objetivo: Estimar contribuciones relativas por canal bajo distintos supuestos de atribución y analizar su impacto en simulaciones de decisión de presupuesto.

DATASET

Fuente: Kaggle

Tipo: Customer Journey / Marketing Attribution

Formato esperado:

- User ID
- Timestamp
- Channel
- Campaign
- Eventos de interacción por canal, con marca de conversión asociada al recorrido del usuario.
- Conversion (Yes / No).

Condiciones del conjunto de datos:

- Datos reales sin atribución previa.
- Presencia de recorridos incompletos y ruido propio del *tracking* de eventos.
- Ausencia de una verdad de referencia sobre la contribución real de los canales.

El conjunto de datos no proporciona una verdad absoluta de atribución, sino observaciones parciales del recorrido de usuario, reflejando condiciones habituales en entornos reales de marketing digital.

MODELOS CONSIDERADOS

1. Modelo baseline – Last Click

- Enfoque determinista que asigna el 100% del mérito al último canal del recorrido.
- Utilizado como referencia mínima, al representar una práctica habitual en entornos reales de marketing digital.
- No incorpora técnicas de *Machine Learning*, pero resulta necesario como línea de base de comparación.

2. Modelo Secuencial – Cadenas de Markov

- Estados: canales + conversiones.
- Probabilidades de transición.

Lo que demuestra:

- Comprensión de secuencias.
- Dependencia del orden de las interacciones.
- Lógica causal básica del recorrido del usuario.

3. Modelo de Reparto - Shapley simplificado

- Contribución media marginal por canal.
- Aproximado, no exhaustivo.
- Sin combinación salvaje.

Lo que demuestra:

- *Fairness*
- Pensamiento cooperativo
- Interpretabilidad

OUTPUT FINAL

1. Tabla de contribución por canal para cada modelo de atribución.
2. Simulación de escenarios de reasignación de presupuesto a partir de las contribuciones estimadas:
 - variaciones porcentuales en la inversión por canal,
 - estimación del impacto relativo en conversiones,
 - análisis hipotético sin optimización automática ni carácter prescriptivo.

LO QUE NO FORMA PARTE DEL ALCANCE DEL PROYECTO

- Modelos de *Deep Learning*.
- Arquitecturas *LSTM* o basadas en *Transformers*.
- Optimización matemática automática de presupuesto.
- Procesamiento en tiempo real o arquitecturas de *Streaming*.
- Desarrollo de dashboards avanzados de visualización.
- Evaluación de un número superior a tres modelos de atribución.

Estas exclusiones responden al objetivo del proyecto, centrado en la interpretación de la contribución de los canales y en el análisis comparativo de hipótesis de atribución. Cualquier elemento que no contribuya directamente con este fin queda fuera del alcance definido.

ESTRUCTURA LÓGICA DEL PROYECTO

1. **Problema de negocio**
2. **Limitaciones de los modelos de atribución clásicos.**
3. **Descripción del conjunto de datos y sus restricciones.**
4. **Modelos de atribución empleados y justificación metodológica.**
5. **Comparativa de resultados entre enfoques.**
6. **Simulación de escenarios de decisión de inversión.**
7. **Conclusiones y limitaciones del análisis.**

Este proyecto se desarrolla siguiendo la **metodología CRISP-DM**, estructurando el análisis desde la comprensión del problema de negocio hasta la evaluación de los modelos y su posible aplicación.

1. CRISP-DM • COMPRENSIÓN DEL NEGOCIO

1.1. Contexto del problema

En el entorno del marketing digital, los usuarios interactúan con una marca a través de múltiples canales a lo largo del tiempo, como búsqueda orgánica, publicidad de pago, email marketing, redes sociales o display. Estas interacciones no tienen el mismo propósito ni el mismo impacto, ya que corresponden a distintas fases del proceso de decisión del usuario: **descubrimiento**, **consideración** y **conversión**.

Sin embargo, muchas organizaciones continúan utilizando modelos de atribución simplificados, siendo el más habitual el modelo **Last Click**, que asigna todo el mérito de la conversión al último canal con el que el usuario interactúa antes de comprar. Este enfoque introduce una distorsión sistemática en la interpretación de los datos, al sobrevalorar los canales de cierre e infravalorar los canales que contribuyen en fases tempranas del recorrido.

Como consecuencia, las decisiones de inversión en marketing se apoyan en una visión parcial del **recorrido del usuario**, lo que puede derivar en una asignación ineficiente del presupuesto y en la eliminación de canales que, aun no cerrando conversiones directamente, desempeñan un papel clave en la generación de demanda.

1.2. Objetivo del proyecto

El objetivo de este trabajo es estimar la contribución relativa de los distintos canales de marketing en la generación de conversiones, utilizando datos de recorridos de usuario (*customer journeys*) y comparando distintos supuestos de atribución.

El proyecto no persigue identificar una verdad absoluta de atribución, ya que en datos reales no existe una etiqueta objetiva que indique la contribución real de cada canal. En su lugar, se adopta un enfoque comparativo, analizando cómo varían las conclusiones cuando se aplican diferentes modelos de atribución sobre el mismo conjunto de datos.

1.3. Alcance y limitaciones

El proyecto se centra en la **atribución multicanal** a partir de eventos ordenados temporalmente, comparando tres enfoques:

- un modelo determinista clásico (**Last Click**),
- un modelo secuencial probabilístico basado en **cadenas de Markov**,
- un modelo de reparto de contribución basado en impacto marginal (**Shapley simplificado**).

Como resultado, se obtienen estimaciones de contribución por canal y simulaciones sencillas en escenarios de reasignación de presupuesto.

Quedan fuera del alcance del proyecto modelos de *Deep Learning*, optimización automática de presupuestos, arquitecturas en tiempo real, integraciones con plataformas externas y desarrollos de visualización avanzados. Esta delimitación permite mantener el foco de interpretación y la trazabilidad de los resultados.

1.4. Criterios de éxito

Dado que no existe una verdad de referencia para la atribución, el éxito del proyecto se evalúa en función de:

- la coherencia lógica de los resultados obtenidos,
- la comparabilidad entre modelos aplicados sobre los mismos datos,

- la estabilidad de los rankings de canales,
- y la utilidad de los resultados como apoyo a la toma de **decisiones de inversión**.

1.5. Supuestos y restricciones

Se asume que los datos disponibles representan una parte significativa del recorrido del usuario, aunque puedan existir interacciones no registradas. Asimismo, se considera normal la presencia de ruido, usuarios mal identificados y recorridos incompletos, reflejando condiciones habituales en entornos reales de marketing digital.

Conclusión de la fase: se establece el marco conceptual y de negocio del proyecto, delimitando el problema, los objetivos y las restricciones que condicionan el resto del análisis y la interpretación de los resultados.

2. CRISP-DM • ENTENDIMIENTO DE LOS DATOS (síntesis)

El conjunto de datos utilizado está compuesto por registros de eventos que describen interacciones de los usuarios con distintos canales de marketing a lo largo del tiempo. Aunque los datos se presentan a nivel de evento, **la unidad de análisis relevante para la atribución es el *journey***. Definido como la secuencia ordenada de interacciones de un usuario dentro de una ventana temporal, que puede finalizar o no en conversión.

En esta fase se ha realizado un análisis exploratorio orientado a comprender la estructura de los recorridos y detectar posibles sesgos de medición. Se ha analizado la distribución de canales, la frecuencia de conversión, la longitud de los *journeys*, los retardos temporales entre interacciones y la dominancia de determinados canales en primera y última posición del recorrido.

Este análisis permite identificar patrones relevantes y justifica la necesidad de emplear modelos de atribución que tengan en cuenta tanto la secuencia de interacciones como la contribución marginal de cada canal.

Las conclusiones obtenidas en esta fase condicionan directamente las decisiones adoptadas en la preparación de los datos y en la selección de los modelos de atribución empleados, el análisis exploratorio completo y reproducible se encuentra documentado en el **Anexo Técnico (Notebook)**.

Conclusión de la fase: el análisis exploratorio confirma la existencia de patrones secuenciales y sesgos inherentes a los datos que hacen insuficientes los modelos de atribución simplificados basados únicamente en el último contacto.

Nota técnica: en el EDA inicial el timestamp se inspecciona como texto; la conversión a formato temporal y el cálculo de retardos se formalizan en la fase 3.

3. CRISP-DM • PREPARACIÓN DE DATOS (*Feature Engineering*)

3.1. Objetivo de la fase

El objetivo de esta fase es **transformar los datos de eventos en una representación adecuada del recorrido del usuario**, incorporando información temporal y secuencial que permita aplicar modelos de **atribución multicanal**.

A diferencia de un problema de predicción clásica, en atribución multicanal la preparación de los datos no consiste únicamente en limpiar y escalar variables. En su lugar se centra en reconstruir el contexto del *journey* y capturar el papel relativo de cada interacción dentro del proceso de conversión.

3.2. Construcción de recorridos de usuario (*journeys*)

Los datos originales se presentan a nivel de evento individual. Para su uso en modelos de atribución, es necesario:

- ordenar las interacciones por usuario y tiempo,
- definir una secuencia coherente de eventos,
- y establecer una relación explícita entre cada interacción y la conversión final (si existe).

En esta fase se asume que cada usuario puede generar uno o varios recorridos, y que la atribución se realiza sobre la secuencia completa de interacciones previa a la conversión.

3.3. Ingeniería de variables orientada a atribución

La ingeniería de variables se ha diseñado con un enfoque causal y temporal, evitando técnicas de representación complejas y priorizando la interpretabilidad.

Las variables construidas incluyen:

- **Posición del *touchpoint*** dentro del recorrido.
- **Tiempo transcurrido desde el impacto anterior.**
- **Número de interacciones previas por canal.**
- **Indicador de canal final frente a canal asistente.**
- **Decaimiento temporal simple**, que reduce el peso de interacciones lejanas en el tiempo.

Estas variables permiten capturar no solo la presencia de un canal, sino su rol y contexto dentro del recorrido del usuario.

3.4. Justificación metodológica

No se emplean técnicas de *Deep Learning*, *embeddings* ni representaciones latentes complejas. Esta decisión responde a dos criterios:

- La necesidad de **interpretabilidad**, clave en problemas de atribución.
- El objetivo del proyecto, centrado en comparar **hipótesis de atribución**, no en maximizar capacidad predictiva.

La preparación de datos realizada permite aplicar modelos secuenciales y de reparto de contribución manteniendo trazabilidad y coherencia con el problema de negocio.

El proceso completo de transformación y generación de variables se documenta de forma reproducible en el **Anexo Técnico (Notebook)**, garantizando la trazabilidad del análisis.

Conclusión de la fase: la preparación de los datos permite reconstruir recorridos de usuario coherentes e incorporar información temporal y contextual necesaria para aplicar modelos de atribución multicanal interpretables.

4. CRISP-DM • MODELADO

4.1. Objetivo de la fase

El objetivo de esta fase es **comparar distintos enfoques de atribución multicanal** aplicados al mismo conjunto de recorridos de usuario, con el fin de analizar cómo varían las conclusiones según los supuestos de cada modelo.

No se persigue identificar un modelo “correcto” en términos absolutos, sino **evaluar distintas hipótesis razonables** sobre la contribución de los canales a la conversión.

Cada modelo representa una hipótesis distinta sobre como contribuyen los canales.

4.2. Modelos de atribución considerados

4.2.1. Modelo baseline: Last Click

El modelo **Last Click** asigna el 100% del mérito de la conversión al último canal del recorrido del usuario. Este enfoque se utiliza como línea base, dado que representa una práctica común en entornos reales de marketing digital.

Su principal limitación es que ignora completamente las interacciones previas del recorrido, lo que puede sobrevalorar los canales de cierre e infravalorar aquellos que actúan como asistencia o descubrimiento.

4.2.2. Modelo secuencial: Cadenas de Markov (con removal effect)

El modelo basado en **cadena de Markov** representa el recorrido del usuario como una secuencia de estados: $Start \rightarrow canales \rightarrow \{Conversion, Null\}$. A partir de los recorridos observados se estiman probabilidades de transición y se calcula la probabilidad base de conversión.

La contribución de cada canal se estima mediante *removal effect*: se elimina el canal de grafo, se renormalizan las transiciones entrantes y se mide la caída en la probabilidad de conversión respecto al modelo base. Las contribuciones se normalizan para obtener un reparto relativo comparable con el resto de modelos.

Este resultado no debe interpretarse como causalidad ni ROI: es una estimación relativa bajo supuestos probabilísticos y dependiente de la estructura observada en los journeys.

4.2.3. Modelado de reparto de contribución: Shapley simplificado

El enfoque basado en **valores de Shapley** distribuye el mérito de la conversión entre los canales en función de su contribución marginal promedio a lo largo de los distintos recorridos.

En este proyecto se emplea una versión simplificada y computacionalmente viable del método, priorizando la interpretabilidad y la comparación con los otros modelos frente a una implementación académica compleja.

A modo de síntesis, esta tabla resume los supuestos, el tipo de medición y la aportación conceptual de cada modelo de atribución considerado.

Modelo	Qué asume	Qué mide	Qué aporta
Last Click	Último impacto	Cierre	Línea base de comparación
Markov	Secuencia	Contribución marginal	Dependencia del orden
Shapley	Cooperación	Contribución media	Reparto equitativo (<i>fairness</i>)

4.3. Estrategia de comparación

Los tres modelos se aplican sobre el mismo conjunto de recorridos preparados en la fase anterior, permitiendo comparar:

- la contribución asignada a cada canal,
- los rankings resultantes,
- y las diferencias conceptuales entre enfoques.

Esta comparación permite analizar cómo distintos supuestos de atribución conducen a conclusiones distintas, con implicaciones directas en la toma de decisiones de inversión. El objetivo de la comparación no es identificar un modelo superior en términos absolutos, sino observar cómo varía la lectura del negocio cuando se adoptan hipótesis de atribución diferentes sobre los mismos datos.

4.4. Criterio profesional: supuestos fuertes y contextos no recomendados

En atribución multicanal no existe una “verdad” observable de contribución real, por lo que cada modelo debe interpretarse como una **hipótesis** sobre el comportamiento del usuario y el papel de los canales. En consecuencia, la evaluación profesional del modelo no se limita a sus resultados numéricos, sino a **cuándo sus supuestos son razonables** y cuándo su uso puede inducir decisiones frágiles.

4.4.1. Modelo baseline: Last Click

- **Supuestos fuertes:** el último impacto es el responsable principal de la conversión; el resto de interacciones previas tienen valor despreciable.
- **Cuándo funciona bien:** cuando el ciclo de decisión es corto, el *funnel* es directo, o el objetivo es *reporting* operativo simple; también como baseline mínimo para comparar hipótesis más ricas.
- **Cuándo NO debería usarse:** cuando existen recorridos largos con canales asistentes relevantes (*awareness*/consideración), o cuando hay campañas *upper-funnel* que no cierran pero sí habilitan.

4.4.2. Modelo secuencial: Markov (removal effect)

- **Supuestos fuertes:** el journey puede modelarse como cadena de estados; al eliminar un canal se redistribuyen transiciones observadas mediante renormalización.
- **Cuándo funciona bien:** cuando se busca una estimación marginal secuencial interpretable (removal effect) y comparación de contribuciones bajo supuestos probabilísticos.
- **Cuándo NO debería usarse:** tracking incompleto/ruidoso o taxonomía inestable; y, sobre todo, cuando se necesita inferencia causal/incremental o ROI real.

4.4.3. Modelo de reparto: Shapley simplificado

- **Supuestos fuertes:** los canales contribuyen “cooperativamente” y el mérito se reparte por contribución marginal media; en versión simplificada se asume una aproximación computacional, no exhaustiva.
- **Cuándo funciona bien:** cuando se busca un reparto interpretable y relativamente estable del mérito entre canales presentes en conversiones, con lectura comparativa frente a baseline.
- **Cuándo NO debería usarse:** cuando se requiere rigor académico completo (Shapley exacto), cuando la granularidad de canales es demasiado fina (explosión de combinaciones) o cuando se necesita incorporar explícitamente costes, saturación o efectos causales (no observables aquí).

4.4.4. Tabla resumen de riesgo y contexto no recomendado

Modelo	Riesgo principal	Contexto no recomendado
Last Click	Sesgo a canales de cierre; infravalora asistencia/ <i>awareness</i>	<i>Journeys</i> largos, <i>funnels</i> con asistencia real, campañas <i>upper-funnel</i>
Markov(Removal Effect)	Sensible a estructura/taxonomía + lectura causal indebida	ROI real, lift incremental, decisiones prescriptivas sin validación
Shapley simplificado	Aproximación; depende de presencia en conversiones; no modela secuencia	Necesidad de Shapley exacto, efectos temporales fuertes, análisis causal/ROI real

4.5. Riesgos de uso operativo (mala interpretación)

En contextos reales, los modelos de atribución suele utilizarse para **reporting** y como insumo para **decisiones de presupuesto**. El riesgo no es el modelo en sí, sino **cómo se convierte una estimación relativa en una orden operativa** (“sube/baja inversión”, “cierra canal”, “duplica campaña”). Por ello se identifican los principales riesgos de mala interpretación por modelo y su impacto potencial en decisiones.

4.5.1. Last Click (baseline)

- **Riesgo: penalización de canales de descubrimiento (*early funnel*).** Al concentrar el mérito en el último impacto, tiende a “castigar” canales que generan demanda pero no cierran.
- **Decisión real afectada:** recorte de inversión en *upper-funnel* (display/social/email de asistencia), dejando al embudo sin abastecimiento.
- **Efecto típico de reporting:** aparente “eficiencia” de canales de cierre (*search/direct*), que se sobrevaloran por definición.

4.5.2. Markov (*removal effect*)

- **Riesgo:** confundir contribución marginal del modelo (*removal effect*) con causalidad/ROI. Además, el método es sensible a ruido, tracking y cambios en la taxonomía de canales (porque la renormalización redistribuye masa de probabilidad).
- **Decisión real afectada:** reasignaciones agresivas por variaciones pequeñas en la estructura de datos o redefinición de canales.
- **Efecto típico en reporting:** cambios aparentes de relevancia por cambios del dato, no necesariamente por cambios reales de comportamiento.

4.5.3. Shapley simplificado

- **Riesgo: falsa sensación de “justicia” o rigor.** Al repartir mérito de forma “equitativa”, puede percibirse como una verdad objetiva, pese a ser una aproximación que no modela orden ni tiempo.
- **Decisión real afectada:** “aplanar” el presupuesto entre canales por percepción de *fairness*, incluso cuando el negocio requiere priorizar cierre o captación según objetivos.
- **Efecto típico en reporting:** lectura excesivamente estable o “neutral”, que puede ocultar la dependencia real de canales de cierre.

4.5.4. Conclusión operativa

Ninguno de los enfoques debe interpretarse como ROI ni como causalidad. El uso profesional consiste en tratarlos como **hipótesis comparativas**, combinando:

- baseline (para detectar sesgos),
- lectura secuencial (para entender *journeys*),
- reparto marginal aproximado (para comparación interpretativa), y validando estabilidad mediante *stress testing* antes de usar resultados como soporte de decisión.

En este sentido, el valor del análisis no reside en el valor numérico de la contribución estimada, sino en la capacidad de contrastar decisiones bajo supuestos explícitos y evaluar su estabilidad antes de su uso operativo.

La implementación detallada de los modelos y los resultados obtenidos se documentan en el **Anexo Técnico (Notebook)**.

Conclusión de la fase: la aplicación de distintos modelos de atribución evidencia cómo diferentes supuestos metodológicos conducen a lecturas divergentes sobre la contribución de los canales.

5. CRISP-DM • EVALUACIÓN ORIENTADA A NEGOCIO

5.1. Enfoque de evaluación

Dado que el objetivo del proyecto no es la predicción de conversiones, sino la **atribución de su contribución**, no se emplean métricas clásicas de clasificación como *accuracy*, *precisión* o *recall*.

La evaluación se realiza desde una **perspectiva de negocio**, analizando cómo los distintos modelos de atribución afectan a:

- la distribución de la inversión entre canales,
- la estabilidad de los rankings obtenidos,
- y la sensibilidad de los resultados ante cambios en los recorridos de usuario.

Este enfoque es coherente con el uso real de los modelos de atribución en entornos de toma de decisiones.

La calidad de un modelo de atribución se mide por su utilidad para decidir, no por su capacidad predictiva.

5.2. Comparación frente al baseline (Last Click)

El modelo **Last Click** se utiliza como referencia, dado que representa una práctica común en entornos reales de marketing digital.

Los modelos secuenciales y de reparto marginal se evalúan comparando cómo redistribuyen la contribución de los canales respecto a este baseline, identificando qué canales ganan o pierden peso cuando se consideran hipótesis alternativas de atribución.

En los resultados obtenidos se observa que los modelos alternativos al **Last Click** redistribuyen parte del mérito hacia canales que participan en fases tempranas o intermedias del recorrido, mientras que en los canales de cierre tienden a perder peso relativo.

5.3. Robustez de los modelos ante perturbaciones

5.3.1. Objetivo del análisis de robustez

En problemas de atribución multicanal no existe una verdad de referencia que permita validar de forma objetiva la contribución real de cada canal. Las estimaciones obtenidas dependen necesariamente de los supuestos metodológicos adoptados y de la estructura de datos observados. En este contexto, la calidad del análisis no se evalúa en términos de precisión predictiva, sino en la coherencia y estabilidad de las conclusiones obtenidas.

Por este motivo, se incorpora un análisis de robustez cuyo objetivo es evaluar hasta qué punto los resultados y rankings de canales se mantienen estables ante modificaciones razonables en los datos de entrada. Este enfoque permite identificar modelos excesivamente sensibles a pequeñas perturbaciones y distinguir conclusiones estructurales de aquellas que dependen de configuraciones específicas del *dataset*. El análisis no persigue optimizar métricas ni validar modelos en sentido predictivo, sino reforzar la fiabilidad de la interpretación desde una perspectiva de negocio.

5.3.2. Escenarios de *stress test* considerados

Para evaluar la robustez de los modelos de atribución analizados, se han definidos tres escenarios de perturbación que reflejan situaciones plausibles en entornos reales de marketing digital:

- **Eliminación de un canal dominante**, con el objetivo de analizar cómo se redistribuye la contribución cuando se retira un canal con alto peso en los resultados base.
- **Agrupación de canales funcionalmente similares**, evaluando la sensibilidad de los modelos a cambios en la taxonomía de canales y a decisiones de categorización habituales en la práctica.
- **Limitación de recorridos de usuario excesivamente largos**, con el fin de comprobar si las conclusiones depende de *journeys* atípicos o poco frecuentes.

Estos escenarios permiten analizar la estabilidad de los modelos frente a alteraciones estructurales sin introducir supuestos artificiales ni modificar el objetivo original del análisis.

En los escenarios evaluados se recalculan los repartos de contribución (Shapley) y, en el caso de Markov, se vuelve a estimar el modelo con removal effect para observar cambios en la contribución marginal relativa y estructura de transiciones.

5.3.3. Resultados comparativos

Los resultados obtenidos muestran diferencias relevantes en el comportamiento de los modelos ante las perturbaciones introducidas. En términos generales, el modelo basado en **cadena de Markov** presenta una mayor sensibilidad a cambios en la estructura de los recorridos, especialmente en escenarios donde se elimina un canal dominante o se modifica la longitud de los *journeys*. Esta sensibilidad se refleja en variaciones apreciables en la estructura del recorrido y en la lectura relativa a los canales y en cambios en el orden relativo del ranking.

Por su parte, el modelo de reparto basado en **valores de Shapley** simplificados tiende a mostrar una mayor estabilidad relativa, manteniendo con mayor frecuencia la jerarquía general de los canales, aunque también presenta ajustes en la magnitud de las contribuciones cuando se altera la composición de los recorridos. En ambos enfoques, se identifican canales cuya contribución resulta particularmente inestable, lo que sugiere una dependencia elevada del contexto del *journey* o de la presencia de canales específicos.

En este análisis, los canales asociados a medios de pago (*Search Ads* y *Display Ads*) muestran mayor variación relativa bajo los escenarios evaluados, lo que sugiere sensibilidad a la definición de taxonomía y a la longitud del recorrido.

5.3.4. Interpretación de negocio

Desde una perspectiva de negocio, los resultados del análisis de robustez ponen de manifiesto que la elección del modelo de atribución puede tener implicaciones relevantes en la toma de decisiones cuando los rankings de canales no son estables ante perturbaciones razonables de los datos. Un modelo excesivamente sensible puede conducir a reasignaciones de presupuesto basadas en patrones frágiles, aumentando el riesgo de decisiones subóptimas.

La robustez de los resultados adquiere, por tanto, una importancia superior al valor exacto de la contribución estimada para cada canal. Rankings consistentes bajo distintos escenarios proporcionan mayor confianza para apoyar decisiones estratégicas, mientras que la inestabilidad debe interpretarse como una señal de cautela. Este análisis refuerza la necesidad de interpretar los modelos de atribución como hipótesis analíticas sobre el comportamiento del usuario, y no como representaciones exactas o verdades absolutas de la contribución real de los canales.

5.3.5. Estabilidad del ranking de canales

Se analiza la estabilidad del ranking de canales entre modelos, observando:

- coincidencias en los canales más relevantes,
- cambios significativos de posición,
- y posibles señales de sobrevaloración o infravaloración sistemática.

La estabilidad del ranking se interpreta como un indicador de robustez del modelo frente a variaciones razonables en los supuestos de atribución.

Este análisis permite identificar hasta qué punto las conclusiones sobre la relevancia de los canales dependen del modelo de atribución adoptado.

5.3.6. Sensibilidad al cambio de recorridos

Finalmente, se evalúa de forma exploratoria la sensibilidad de los modelos ante cambios en los recorridos de usuario, como la eliminación de ciertos canales o la modificación de la composición de los *journeys*.

Este análisis permite comprobar si los resultados dependen de patrones aislados o reflejan comportamientos consistentes del conjunto de datos, lo que refuerza la interpretación de los modelos desde una perspectiva de negocio.

Como salida adicional, se identifican canales “inestables” (mayor variación agregada entre escenarios), que requieren cautela si se pretende usar rankings como soporte directo de decisión.

5.4. Contexto económico relativo por canal (coste normalizado)

5.4.1. Motivación

En atribución multicanal, comparar únicamente contribuciones puede inducir decisiones incompletas si no se incorpora el coste relativo de activación de cada canal. Dos canales con contribuciones similares pueden tener implicaciones económicas distintas si su coste de adquisición o exposición difiere de una forma significativa. Por este motivo, se introduce un ajuste de contexto económico basado en un índice de coste relativo por canal, con el objetivo de complementar la interpretación del reparto del mérito de los modelos.

5.4.2. Definición del coste relativo por canal

Dado que el *dataset* no incluye información de inversión (*spend*), **CPM** o **CPC**, se asigna un **coste relativo sintético** por canal (índice normalizado), inspirado en prácticas habituales de marketing digital. El índice no pretende reproducir costes reales del caso, sino proporcionar una referencia comparativa que permita analizar cómo cambia la lectura del “mérito” cuando se considera el esfuerzo económico aproximado asociado a cada canal.

Disclaimer explícito: este indicador no es ROI ni estima rentabilidad real. Es un **contexto económico relativo** que permite comparar contribución en relación con un coste aproximado.

5.4.3. Métrica contribución / coste

Para cada modelo que produce un reparto de contribución por canal, se calcula una métrica de eficiencia relativa:

$$\text{Eficiencia relativa} = \frac{\text{Contribución atribuida}}{\text{Índice de coste del canal}}$$

Esta métrica permite identificar canales que, bajo un supuesto de coste relativo, presentan una mejor relación entre contribución estimada y esfuerzo económico aproximado.

5.4.4. Integración con la simulación de presupuesto

La simulación de redistribución de presupuesto se complementa con una segunda lectura “coste-ajustada”, en la que la inversión conceptual se asigna proporcionalmente a la eficiencia relativa (contribución/coste), manteniendo el carácter no prescriptivo del proyecto. De este modo, se ilustra cómo un mismo modelo de atribución puede sugerir decisiones distintas cuando se introduce fricción económica, sin realizar optimización automática ni predicción de resultados futuros.

La simulación genera dos lecturas: (1) asignación proporcional a contribución y (2) asignación proporcional a la eficiencia (contribución/coste). Ambas son ilustrativas y no prescriptivas.

5.5. Simulación de redistribución de presupuesto

A partir de las contribuciones estimadas por cada modelo, se realiza una simulación conceptual de redistribución de presupuesto, asumiendo que la inversión se asigna proporcionalmente al mérito atribuido a cada canal.

Esta simulación se apoya en los resultados del análisis de robustez previamente descrito, asumiendo que los modelos utilizados presentan un nivel de estabilidad suficiente para ilustrar escenarios de decisión.

La evaluación técnica detallada se encuentra documentada en el **Anexo Técnico (Notebook)**.

Conclusión de la fase: la evaluación orientada a negocio muestra que la elección del modelo de atribución tiene un impacto directo en las decisiones potenciales de asignación de presupuesto.

6. CRISP-DM • CONCLUSIONES Y DEPLOYMENT CONCEPTUAL

6.1. Conclusiones del proyecto

El objetivo de este proyecto ha sido analizar la atribución multicanal de conversiones a partir de recorridos de usuario, comparando distintos supuestos de atribución aplicados sobre el mismo conjunto de datos.

Los resultados obtenidos muestran que la contribución asignada a cada canal depende de forma directa del modelo de atribución empleado. Mientras que el modelo **Last Click** concentra el mérito en los canales finales del recorrido, los enfoques secuenciales y de reparto marginal redistribuyen parte de la contribución hacia canales que participan como asistencia o descubrimiento.

Esta variabilidad pone de manifiesto que la atribución no es una cuestión de precisión ni de identificar un modelo “mejor” en términos absolutos, sino de los supuestos metodológicos adoptados y del tipo de decisión que se desea apoyar. Los modelos analizados responden a hipótesis distintas sobre el papel de los canales en el recorrido del usuario, y su elección condiciona directamente a la interpretación del negocio y las decisiones de inversión resultantes, más que la calidad intrínseca del resultado obtenido.

6.2. Implicaciones para la toma de decisiones

Desde una perspectiva de negocio, los resultados del proyecto permiten:

- identificar canales que contribuyen de forma indirecta a la conversión,
- evitar decisiones de inversión basadas únicamente en métricas de cierre,
- y comparar escenarios alternativos de asignación de presupuesto.

El uso de múltiples modelos de atribución proporciona una visión más completa del recorrido del usuario y reduce el riesgo de sobrevalorar o infravalorar determinados canales.

6.3. Deployment conceptual

El objetivo del proyecto no es desplegar un sistema productivo completo, sino **apoyar la toma de decisiones mediante análisis reproducibles e interpretables**.

En un entorno real, el enfoque desarrollado podría integrarse como:

- un pipeline de análisis periódico (*batch*),
- alimentado por datos de eventos de marketing,
- generando informes de contribución por canal y simulaciones de presupuesto.

La simplicidad de los modelos empleados facilita su comprensión por perfiles no técnicos y su adopción en contextos empresariales.

El pipeline conceptual completo del proyecto, desde la ingesta de eventos hasta la generación de resultados para la toma de decisiones, se representa de forma esquemática en el diagrama incluido en el **Anexo Técnico (Notebook)**.

Este diagrama permite visualizar la relación entre las distintas fases del proceso, así como la naturaleza iterativa del análisis, en la que los resultados de evaluación pueden conducir a ajustes metodológicos sin introducir procesos de optimización automática ni prescripción directa.

Como artefactos de salida, se generan tablas normalizadas y ficheros CSV (contribuciones, comparativas y simulaciones) listos para su consumo en un dashboard interactivo en Streamlit.

6.4. Limitaciones y líneas futuras

El proyecto presenta algunas limitaciones inherentes a su alcance y a la naturaleza de los datos utilizados:

- dependencia de la calidad del *tracking* de eventos,
- ausencia de una verdad de referencia objetiva sobre la contribución real de los canales,
- uso de versiones simplificadas de los modelos de atribución, priorizando interpretabilidad frente a rigor académico exhaustivo.

Asimismo, los resultados obtenidos deben interpretarse como **estimaciones relativas bajo supuestos explícitos**, no como medidas causales ni de rentabilidad real. La atribución multicanal permite comparar hipótesis sobre el papel de los canales en el recorrido del usuario, pero no estima impacto incremental ni ROI.

Como líneas futuras, podrían explotarse ventanas temporales dinámicas, variantes más completas de los modelos secuenciales o la incorporación de costes reales por canal. Desde una perspectiva de toma de decisiones, apoyar la asignación de inversión en un único modelo de atribución implica asumir sus supuestos como hechos, con el riesgo de sobrevalorar o penalizar determinados canales según la fase del recorrido en la que intervengan. En este sentido, la utilización de múltiples modelos debe entenderse como una medida de **control y contraste de hipótesis**, más que como un aumento innecesario de complejidad analítica.

Finalmente, este enfoque de atribución multicanal no debe interpretarse como una alternativa al **Marketing Mix Modeling (MMM)**, ya que ambos responden a preguntas distintas y operan a niveles de decisión diferentes. Mientras que la atribución analiza la participación relativa de los canales dentro de los recorridos de usuario, el MMM trabaja con datos agregados para estimar el impacto incremental de la inversión sobre los resultados globales del negocio. Ambos enfoques pueden considerarse **complementarios** dentro de un marco analítico más amplio, en el que la atribución aporta comprensión del comportamiento del usuario y el MMM apoya la planificación estratégica de inversión.

7. ANEXO TÉCNICO: NOTEBOOK REPRODUCIBLE

El Anexo Técnico incluye el notebook reproducible desarrollado para este proyecto, en el que se documentan de forma detallada el análisis exploratorio de los datos, la preparación de los recorridos de usuario, la implementación de los modelos de atribución y la generación de resultados y simulaciones.

Este anexo tiene como objetivo garantizar la trazabilidad técnica del proyecto y permitir la revisión completa de los procedimientos y resultados descritos en la Memoria.

8. RESUMEN EJECUTIVO DEL PROYECTO

Este proyecto aborda el problema de la atribución multicanal de conversiones en entornos de marketing digital, donde los usuarios interactúan con una marca a través de múltiples canales antes de completar una conversión. En este contexto, el uso generalizado de modelos de atribución simplificados, como el enfoque **Last Click**, introduce sesgos sistemáticos al concentrar el mérito en los canales finales del recorrido y minimizar el papel de aquellos que contribuyen en fases tempranas o intermedias.

Con el objetivo de analizar cómo distintos supuestos de atribución condicionan la interpretación del rendimiento de los canales, se ha desarrollado un estudio comparativo aplicando tres enfoques sobre un mismo conjunto de recorridos de usuario: un modelo determinista clásico (**Last Click**), un modelo secuencial probabilístico basado en **cadena de Markov** y un modelo de reparto de contribución mediante **valores de Shapley** en una versión simplificada. Todos los modelos se han aplicado sobre datos reales de interacciones multicanal, caracterizados por la ausencia de una verdad de referencia de atribución y por la presencia de recorridos incompletos, reflejando condiciones habituales en entornos reales.

Los modelos analizados no se presentan como estimadores de impacto causal ni como mecanismos de optimización automática de presupuesto. Cada enfoque representa una hipótesis distinta sobre el papel de los canales en el recorrido del usuario, y sus resultados deben interpretarse como estimaciones relativas orientadas a apoyar decisiones, no como recomendaciones prescriptivas.

Los resultados obtenidos muestran que la contribución asignada a cada canal varía de forma significativa en función del modelo empleado. Mientras que el enfoque **Last Click** concentra el mérito en los canales de cierre, los modelos alternativos redistribuyen parte de la contribución hacia canales con un papel asistente o de descubrimiento. Esta variabilidad pone de manifiesto que la atribución no debe interpretarse como un problema de precisión o de identificación de un modelo “correcto”, sino como una cuestión de hipótesis metodológicas que influyen directamente en la lectura del negocio.

A partir de las contribuciones estimadas, se han realizado simulaciones conceptuales de reasignación de presupuesto, evaluando cómo distintos criterios de atribución pueden conducir a decisiones de inversión diferentes. Estas simulaciones no persiguen optimizar ni prescribir automáticamente la asignación de recursos, sino ilustrar el impacto potencial de adoptar un modelo de atribución frente a otro en un contexto de toma de decisiones.

El proyecto no tiene como objetivo predecir conversiones futuras, optimizar presupuestos de forma automática ni desplegar un sistema productivo en tiempo real. Su aportación se centra en proporcionar un marco analítico interpretable y comparativo que permita a los responsables de negocio contrastar hipótesis de atribución y reducir el riesgo de tomar decisiones basadas en un único supuesto metodológico.

En conjunto, el proyecto demuestra que la utilización de múltiples modelos de atribución no introduce complejidad innecesaria, sino que actúa como un mecanismo de control y contraste que mejora la calidad del proceso de toma de decisiones estratégicas en marketing digital, especialmente en ausencia de una verdad objetiva sobre la contribución real de los canales.