

Customer Funnel Intelligence

Análisis de Conversión y Retención.

Desarrollado por: **Ori Di Ciaccio.**
Data Analyst | Business Intelligence



*Este proyecto analiza el embudo de conversión desde el primer contacto hasta la conversión a cliente utilizando datos reales del CRM. A través de análisis de embudo, cohortes y matrices de transición, se identifican patrones clave en la **eficiencia del proceso**, el **tiempo de conversión** y las **oportunidades de optimización**, impulsando una **toma de decisiones basada en datos**.*

Flujo de trabajo:

- Se explican **terminologías clave**, asegurando que los hallazgos sean interpretables para usuarios técnicos y no técnicos.
- Se utilizó **Python** para la exploración inicial garantizando una comprensión profunda del dataset y generando insights generales.
- Se emplea **BigQuery**, para resolver preguntas a través de consultas SQL avanzadas.
- Los resultados se presentan en **Looker Studio**, facilitando la visualización de métricas clave, tendencias y flujos de conversión de manera intuitiva para stakeholders.

Conceptos Clave para el Análisis:

Para alinear conceptos y dar contexto al análisis, se definen algunos términos que utilizaremos a lo largo del estudio:

- **Lead:** contacto que muestra interés inicial en el producto o servicio, aunque no necesariamente se convierte en cliente.
- **Cohorte:** grupo de contactos que comparten una característica común en el tiempo (semana de creación). Analizar cohortes permite entender el comportamiento y la evolución de leads de forma comparativa.
- **Conversión:** paso en el que un lead avanza en el embudo de ventas, por ejemplo, desde contacto a cliente. La métrica principal es la **tasa de conversión**, calculada como:

$$(Numerototaldeleads / Numerodeconversiones) \times 100 = \%$$

Análisis exploratorio y descriptivo de datos

Objetivo: Comprender en detalle las características de los datos disponibles, detectar patrones, inconsistencias y posibles sesgos, con el fin de orientar la limpieza, transformación y posterior modelado.

- **Descripción de los datos:**

Dataset 1: `df_contacts` (518.666 registros)

Columna	Tipo de dato	Descripción/ Observacion
<code>contact_id</code>	int	Identificador único del contacto.
<code>created_at</code>	object (fecha)	Fecha y hora en la que se creó el contacto.
<code>utm_source</code>	object	Fuente de adquisición de marketing . Valores únicos → 8 <code>customer</code> , <code>instagram</code> , <code>web</code> , <code>google</code> , <code>facebook</code> , <code>tiktok</code> , <code>affiliate</code> , <code>call</code> , <code>employee</code> , <code>[nan]</code>
<code>object_source</code>	object	Fuente técnica por la cual se creó el contacto (interfaz CRM, formulario, API). Muy útil para

Columna	Tipo de dato	Descripción/ Observacion
		segmentar origen. Valores unicos → 13 <code>crm_ui</code> , <code>meetings</code> , 0 <code>form</code> , <code>integration</code> , <code>conversations</code> , <code>email_integration</code> , <code>payments</code> , <code>extension</code> , <code>import</code> , <code>presentations</code>
<code>utm_medium</code>	object	Medio de la campaña (campañas de pago, email, orgánico). Valores unicos → 9 <code>referral</code> , <code>direct</code> , <code>paid</code> , <code>partner</code> , <code>outbound</code> , <code>organic</code> , <code>inbound</code> , <code>social</code> , <code>ppc</code> , <code>[nan]</code>

Dataset 2: `df_events` (23.866 registros)

Columna	Tipo de dato	Descripción / Observacion
<code>hs_object_id</code>	int	Identificador único de cada contacto. Menor que el total de filas → un mismo contacto puede tener múltiples registros (diferentes modificaciones).
<code>lastmodified_ts</code>	object (fecha)	Última fecha de modificación del contacto. casi único por fila.
<code>lifecyclestage</code>	object	Categoría clave del análisis. Valores unicos → 5 <code>customer</code> , <code>lead</code> , <code>marketingqualifiedlead</code> , <code>subscriber</code> , <code>155548387</code> , <code>[nan]</code> Hay un valor anómalo <code>"155548387"</code> .

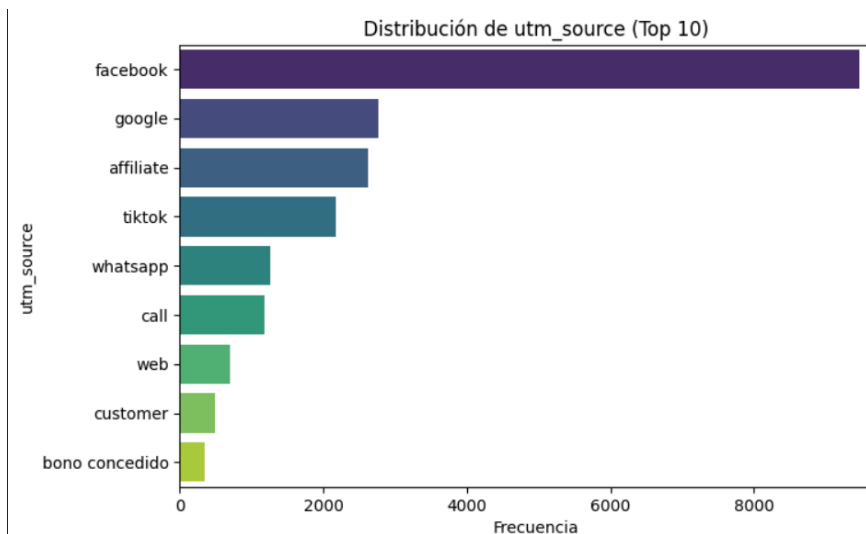
• **Análisis de Valores Nulos**

Columna	Nº de valores nulos	Observación
<code>hs_object_id</code>	0	Sin valores nulos, identificador fiable.
<code>lastmodified_ts</code>	0	Fechas completas para todos los registros.
<code>lifecyclestage</code>	10.217	2% de los contactos no tienen etapa definida.
<code>contact_id</code>	0	Sin valores nulos, confiable.
<code>created_at</code>	0	Fechas completas.
<code>utm_source</code>	2.334	Casi un 10% sin fuente de campaña, puede afectar análisis de adquisición.
<code>object_source</code>	1	Prácticamente completo.
<code>utm_medium</code>	2.006	8.40%. Faltan datos en medios de campaña.

- Duplicados NO presentes

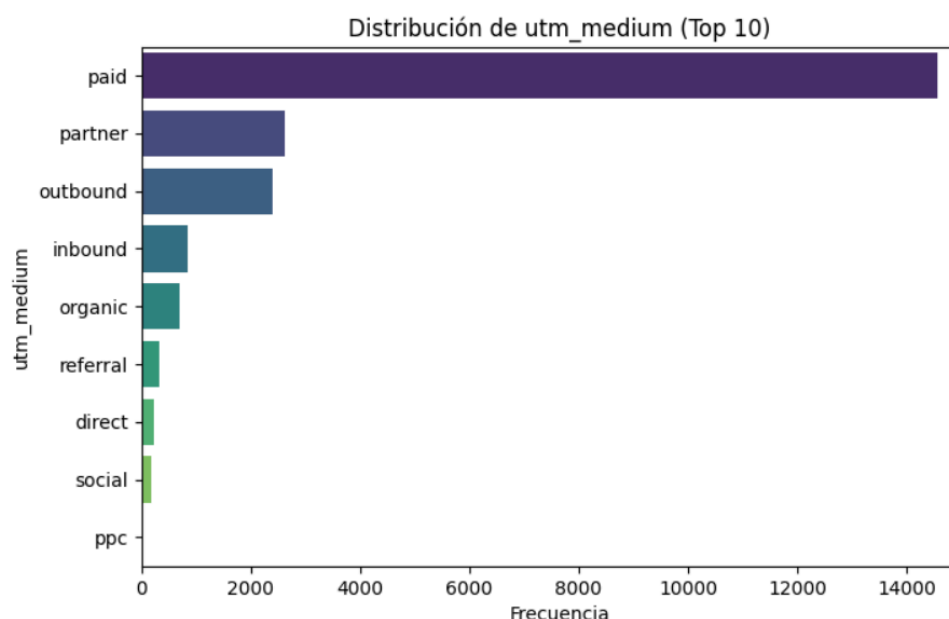
Análisis de distribución → variables `utm_source` , `utm_medium` , `object_source` & `lifecyclestage`

Objetivo: Identificar cómo se distribuyen las variables categóricas y temporales, evaluando su frecuencia y relevancia en el dataset.



1. **Gráfico 1: Distribución de `utm_source` → Cantidad de contactos según la fuente de adquisición (tráfico o canal de marketing).**
 - **Facebook** es, con diferencia, la fuente principal (10.000 registros).
 - Google y affiliate también aportan bastante volumen (3.000).
 - Canales como tiktok, whatsapp y call tienen relevancia media.

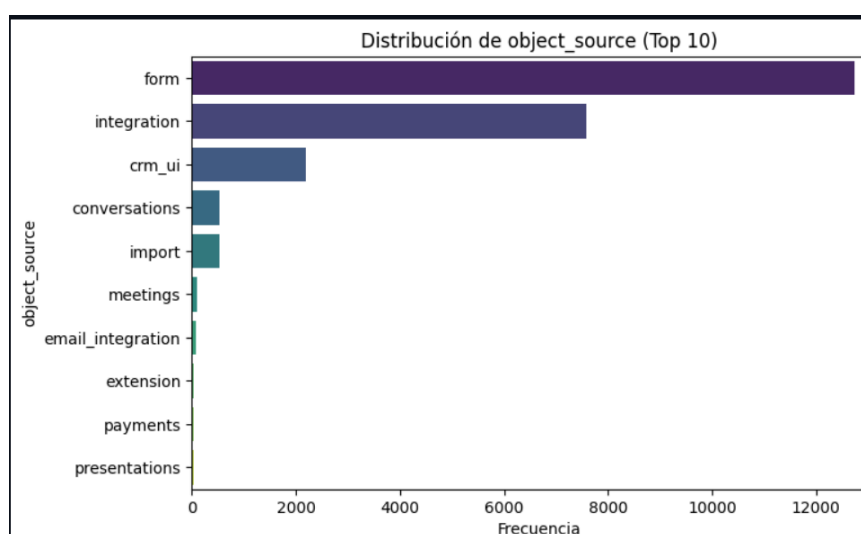
Conclusión: La captación está fuertemente concentrada en Facebook, lo que puede indicar dependencia de un solo canal.



2. **Gráfico 2: Distribución de** `utm_medium` **→ Medio publicitario o de captación.**

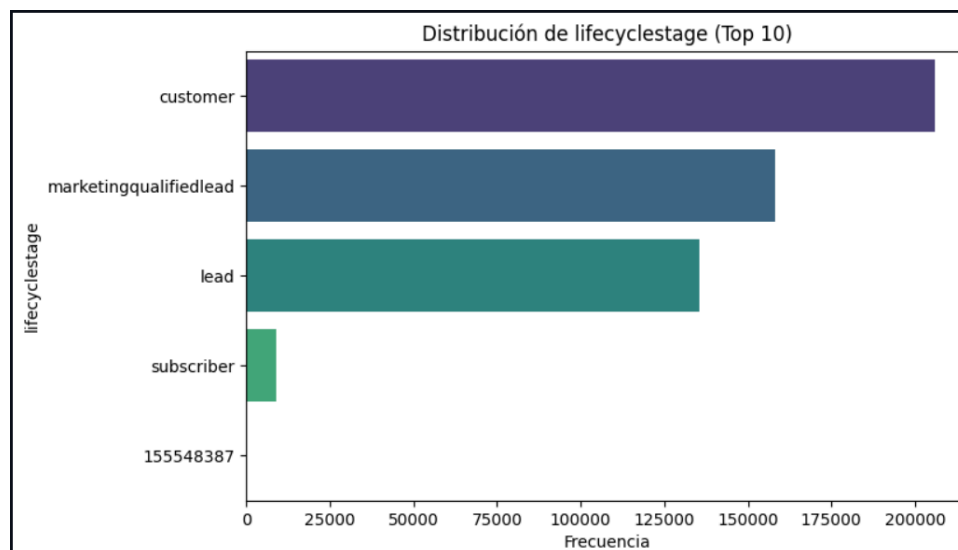
- Paid domina claramente con más de 14.000 registros, lo que refleja una fuerte apuesta por campañas pagadas.
- Partner y outbound (4.000 cada uno) tienen presencia secundaria, pero significativa.

Conclusión: Existe **alta dependencia del canal de pago**, lo que puede implicar mayor costo de adquisición, Los canales partner y outbound pueden representar oportunidades de crecimiento si se optimizan, dado que ya aportan volumen relevante.



3. Gráfico 3: Distribución de `object_source` → Describe el origen

- **Form** aporta la mayor cantidad (12.000), lo que indica que la mayoría de contactos entran a través de formularios.
- Integration (8.000) también es fuerte, lo que refleja uso de integraciones externas (CRM, APIs).



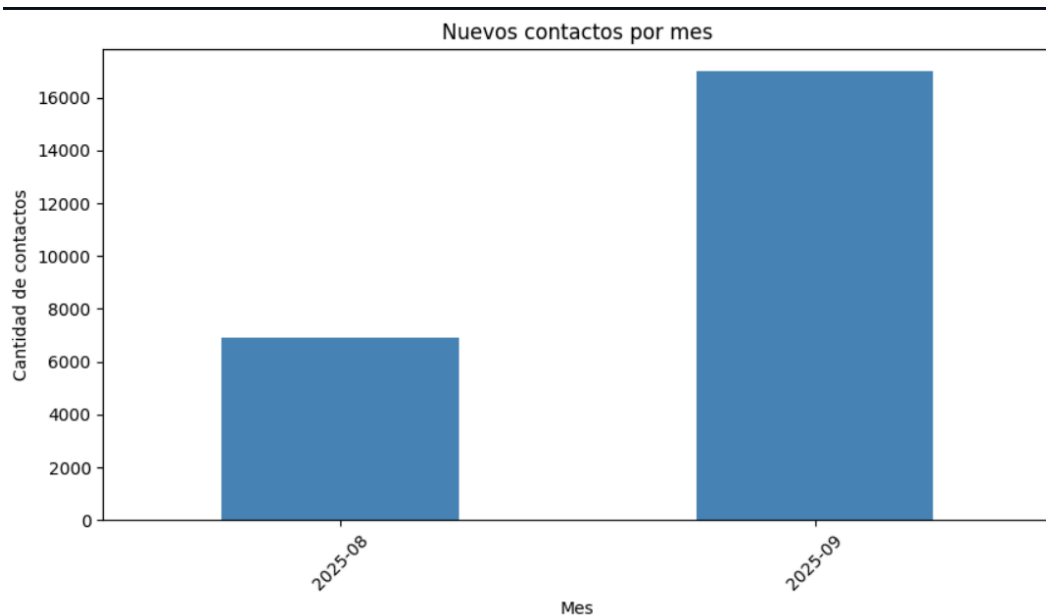
4. Gráfico 4: Distribución de `lifecyclestage` → Contactos según su etapa en el ciclo de vida del cliente.

- **Customer** (200.000) es la categoría dominante. Le siguen marketing qualified lead (MQL) y lead con grandes volúmenes (150.000 y 140.000).
- Existe un valor atípico **155548387**, que parece un error de datos o mal codificación con 202 registros.

Conclusión: La base de datos contiene una gran proporción de clientes activos, lo cual es bueno para estudios de retención.

Análisis temporal general → variables `created_at` & `lastmodified_ts`

Objetivo: Permite evaluar la **antigüedad de los registros** (transformación de fechas en variables temporales), lo que permite segmentar cohortes y analizar comportamiento reciente vs. histórico.

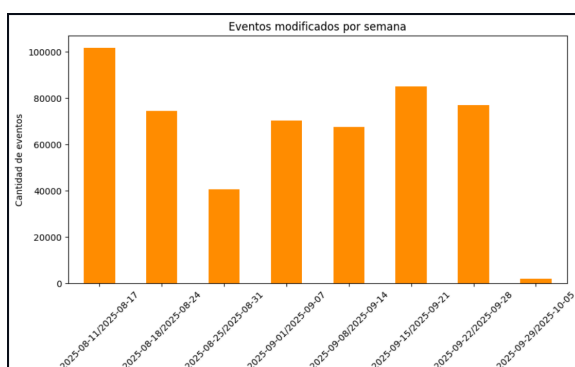


Interpretación: Hay un aumento **muy significativo** de nuevos contactos de un mes al siguiente (más del doble).

Esto puede indicar:

- Una **campaña de captación efectiva** en septiembre.
- Incremento en la actividad comercial.
- Posible estacionalidad o eventos especiales que impulsaron el registro de contactos.

Interpretación: Picos irregulares:



Interpretación: Se observa un **incremento de 35%** en eventos modificados de agosto a septiembre.

Esto puede reflejar:

- Mayor actividad de actualización o seguimiento de los eventos existentes.

- El histograma no es suave → algunas semanas se modificaron muchos contactos, otros menos. Esto puede deberse a campañas de captación, promociones o actividades comerciales específicas.
- Optimización de datos por parte del equipo (por ejemplo, completando información faltante).



Análisis de conversión semanal

Objetivo del análisis:

El análisis semanal de conversión permite **monitorizar la eficiencia del embudo de adquisición** desde el contacto inicial hasta la conversión a cliente.

Medir la tasa de conversión por semana ayuda a:

- Detectar tendencias y variaciones temporales en el rendimiento comercial.
- Identificar picos de conversión asociados a campañas, mejoras del proceso o cambios en el producto.
- Evaluar la efectividad de las estrategias de captación y priorizar acciones que aumenten la conversión.

En resumen, esta información es esencial para **optimizar el proceso de conversión** y mejorar la planificación de recursos y estrategias de marketing.

1. Crearemos una tabla que muestre el número total de contactos y la tasa de conversión a cliente por semana.


```
-- creacion de las semanas
WITH contacts_weekly AS (
  SELECT
    contact_id,
    FORMAT_DATE('%Y-%W', DATE(created_at)) AS year_week
  FROM `bigdataarchitecture-453018.453018.contacts`
)
,
-- Uso de customer como conversión = Booleano
conversion_status AS (
  SELECT
    hs_object_id AS contact_id,
    MAX(CASE WHEN lifecyclestage = 'customer' THEN 1 ELSE 0 END) AS is_customer
  FROM `bigdataarchitecture-453018.453018.events`
  GROUP BY contact_id
)
SELECT
  contacts_weekly.year_week,
  COUNT(contacts_weekly.contact_id) AS total_contacts,
  SUM(conversion_status.is_customer) AS total_customers,
  SAFE_DIVIDE(SUM(conversion_status.is_customer), COUNT(contacts_weekly.contact_id)) * 100 AS conversion_rate
FROM contacts_weekly
LEFT JOIN conversion_status
  ON contacts_weekly.contact_id = conversion_status.contact_id
GROUP BY year_week
ORDER BY conversion_rate DESC;
```

Resultados de la consulta

Información del trabajo		Resultados	Visualización	JSON	Detalles de la ejec	
Fila	year_week	total_contacts	total_customers	conversion_rate		
1	2025-35	3562	698	19.59573273441...		
2	2025-32	1802	348	19.31187569367...		
3	2025-37	4093	778	19.00806254580...		
4	2025-34	2779	473	17.02051097517...		
5	2025-33	2309	390	16.89042875703...		
6	2025-36	4842	810	16.72862453531...		
7	2025-38	4479	624	13.93168117883...		



- La **semana con mayor tasa de conversión es la 2025-35 (19,6%)**.
- La **mínima tasa se da en la semana 2025-38 (13,9%)**, a pesar de ser una semana con alto volumen de contactos.
- La semana 2025-32 (19.31%), presenta buena conversión(comparativamente) con contactos y customers bajos.
- La **semana con mayor volumen no fue la mas eficiente 2025-36 (16.73%)**, sugiere campañas que atrajeron contactos de menos calidad

Hipótesis para tasa de conversión mayor:

- **Campañas efectivas:** puede coincidir con el lanzamiento de una campaña de marketing o promoción que atrajo clientes con mayor intención de compra.
- **Acciones comerciales:** quizás hubo refuerzo en follow-ups del equipo de ventas.

Hipótesis para tasa de conversión menor:

- **Efecto estacional o temporal:** semana con festivales que afectan la disposición de compra
- **Capacidad de gestión:** el equipo de ventas pudo verse saturado con tantos contactos y no dar seguimiento adecuado.

Análisis de cohortes

Objetivo del análisis:

Analizar cómo evolucionan las conversiones a cliente según la semana de creación del contacto (cohorte).

Esto permite responder a:

- ¿Qué tan rápido convierten los nuevos contactos?

- ¿Qué cohortes (semanas de adquisición) tienen mejor rendimiento inicial (7 y 14 días)?
 - ¿Existen patrones de mejora o deterioro en la conversión temprana?
1. Crear una tabla de análisis de cohortes que muestre la conversión a cliente e identifica en qué semana la tasa de conversión a los 7 días es mayor y en qué semana la tasa de conversión a

```
-- preparar los cohortes por semana
WITH contacts_cohort AS (
  SELECT
    contact_id,
    DATE(created_at) AS created_date,
    FORMAT_DATE('%Y-%W', DATE(created_at)) AS cohort_week
  FROM `bigdataarchitecture-453018.453018.contacts`
)
```

```

-- conversion desde primera vez que aparecen como customer
conversion AS (
  SELECT
    hs_object_id AS contact_id,
    MIN(DATE(lastmodified_ts)) AS conversion_date
  FROM `bigdataarchitecture-453018.events`
  WHERE lifecyclestage = 'customer'
  GROUP BY contact_id
),

-- diferencias entre dias de creacion y conversion
cohort_analysis AS (
  SELECT
    contacts_cohort.cohort_week,
    contacts_cohort.contact_id,
    contacts_cohort.created_date,
    conversion.conversion_date,
    DATE_DIFF(conversion.conversion_date, contacts_cohort.created_date, DAY) AS days_to_convert
  FROM contacts_cohort
  LEFT JOIN conversion
  ON contacts_cohort.contact_id = conversion.contact_id
)

SELECT
  cohort_week,
  COUNT(DISTINCT contact_id) AS total_contacts,
  -- primeros 7 dias
  COUNT(CASE WHEN days_to_convert BETWEEN 0 AND 7 THEN contact_id END) AS converted_7d,
  -- primeros 14 dias
  COUNT(CASE WHEN days_to_convert BETWEEN 0 AND 14 THEN contact_id END) AS converted_14d,
  ROUND(SAFE_DIVIDE(COUNT(CASE WHEN days_to_convert BETWEEN 0 AND 7 THEN contact_id END),
    COUNT(contact_id)) * 100, 2) AS conversion_rate_7d,
  ROUND(SAFE_DIVIDE(COUNT(CASE WHEN days_to_convert BETWEEN 0 AND 14 THEN contact_id END),
    COUNT(contact_id)) * 100, 2) AS conversion_rate_14d
FROM cohort_analysis
GROUP BY cohort_week
ORDER BY cohort_week
;

```

Resultados de la consulta

Información del trabajo		Resultados	Visualización	JSON	Detalles de la ejecución		Gráfico de ejecución
Fila	cohort_week	total_contacts	converted_7d	converted_14d	conversion_rate_7d	conversion_rate_14d	
1	2025-32	1802	249	294	13.82	16.32	
2	2025-33	2309	292	335	12.65	14.51	
3	2025-34	2779	377	434	13.57	15.62	
4	2025-35	3562	554	641	15.55	18.0	
5	2025-36	4842	707	789	14.6	16.29	
6	2025-37	4093	735	778	17.96	19.01	
7	2025-38	4479	624	624	13.93	13.93	



- La semana 37 destaca como la más eficiente en términos de rapidez y volumen de conversiones.
- **Analizar en profundidad la Semana 37** . ¿Qué campañas estaban activas? ¿Qué canales de adquisición se utilizaron? ¿Hubo ofertas o promociones especiales?
- **Volatilidad en la Calidad:** A pesar de tener el mayor volumen (Semana 36), esta cohorte no fue la más rápida en convertirse (14,60% en 7 días).
- **Anomalía en Semana 38:** Las tasas de conversión a 7 y 14 días son **idénticos (13,93%)** .análisis incompleto, aun está ocurriendo
- **Velocidad de la Conversión:** La mayoría de las cohortes logran la **mayor parte de sus conversiones en los primeros 7 días** .

El porcentaje promedio de incremento en la conversión de clientes de semana a semana es aproximadamente del 13.79%



indicador directo del rendimiento del equipo de ventas.

Matriz de transición (número de contactos)

Objetivo del análisis:

Analizar el flujo de contactos a través de las etapas del ciclo de vida (**lifecyclestage**) para entender **cuántos contactos avanzan, se estancan o se pierden** en cada punto del embudo.

Esto ayuda a:

- Identificar **cuellos de botella** o puntos de fricción en el funnel.
- Evaluar la efectividad de conversión.
- Visualizar la dinámica completa de los leads a clientes.



- **Filas (`from_stage`)** → etapa de origen.
- **Columnas (`lead` , `marketingqualifiedlead` , `subscriber` , `customer`)** → etapa de destino.
- Cada celda → número de contactos que pasaron de la etapa de origen a la etapa de destino.
- `NULL` significa que no había etapa anterior registrada para estos contactos.
- La matriz cuenta solo los contactos que tuvieron un cambio de etapa registrado en el periodo analizado.

```
WITH ordered_events AS (  
  SELECT  
    hs_object_id,  
    lifecyclestage AS from_stage,  
    lastmodified_ts,  
    LEAD(lifecyclestage) OVER (PARTITION BY hs_object_id ORDER BY lastmodified_ts) AS to_stage  
  FROM `bigdataarchitecture-453018.453018.events`  
)  
  
SELECT *  
FROM (  
  SELECT  
    from_stage,  
    to_stage,  
    COUNT(hs_object_id) AS num_contacts  
  FROM ordered_events  
  WHERE to_stage IS NOT NULL  
  GROUP BY from_stage, to_stage  
)  
PIVOT(  
  SUM(num_contacts) FOR to_stage IN  
    ('lead', 'marketingqualifiedlead', 'subscriber', 'customer'))  
ORDER BY from_stage  
:
```

Resultados de la consulta



Información del trabajo	Resultados	Visualización	JSON	Detalles de la ejecución	Grá
Fila	from_stage	lead	marketingqualifie...	subscriber	customer
1	null	136	534	23	39
2	155548387	null	null	null	null
3	customer	1	null	null	5674
4	lead	15260	5802	null	51
5	marketingqualifiedlead	174	11919	null	4007
6	subscriber	1386	37	1358	6

- contactos se iniciaron directamente como **subscriber** y 3 contactos se iniciaron directamente como **customer**. Campañas que saltan en las primeras etapa.



! a tomar en cuenta:

- **retroceso/ fuga** de **MQL a Lead**.
- **estancamiento masivo** de **lead a lead**.
- **retroceso significativo** de **subscriber a lead**. Este movimiento podría indicar desinterés o la activación de alguna regla de descalificación por inactividad.
- **customer a customer**: No es una transición sino un evento de actualización o repetición de compra.

✓ rutas exitosas:

- Transición **lead** → **MQL**
- Transición **MQL** → **Customer** principal vía de ingresos.
- **Transición subscriber** → **Customer**: Solo **6** conversiones. El *suscriptor* es una base de bajo rendimiento para la conversión final

Matriz de transición (tiempo de transición)

Objetivo del análisis:

El objetivo de esta matriz es medir la **eficiencia del proceso de conversión** entre etapas del embudo (por ejemplo, de *lead* a *marketing qualified lead*, o de *subscriber* a *customer*).

En lugar de contar cuántos contactos avanzan, analizamos **cuánto tardan en hacerlo**.

Calcular el **percentil 80** del tiempo de transición permite:

- Estimar el **tiempo máximo esperado** en el que la mayoría de los contactos completan su transición (sin verse afectados por valores atípicos).
- Proporcionar una métrica robusta para **evaluar la eficiencia operativa**.
- Cada celda → **percentil 80 del tiempo (en días) que tardaron los contactos en pasar de una etapa a otra**.
- **Definición de percentil 80 → el 80% de los contactos tardó menos o igual que este valor en pasar de una etapa a otra.**

```
WITH ordered_events AS (  
  SELECT  
    hs_object_id,  
    lifecyclestage AS from_stage,  
    lastmodified_ts,  
    LEAD(lifecyclestage) OVER (PARTITION BY hs_object_id ORDER BY lastmodified_ts) AS to_stage,  
    LEAD(lastmodified_ts) OVER (PARTITION BY hs_object_id ORDER BY lastmodified_ts) AS next_ts  
  FROM `bigdataarchitecture-453018.453018.events`  
)  
  
transition_times AS (  
  SELECT  
    from_stage,  
    to_stage,  
    DATE_DIFF(DATE(next_ts), DATE(lastmodified_ts), DAY) AS days_to_next  
  FROM ordered_events  
  WHERE to_stage IS NOT NULL  
)  
  
SELECT *  
FROM (  
  SELECT  
    from_stage,  
    to_stage,  
    APPROX_QUANTILES(days_to_next, 5)[OFFSET(4)] AS p80_days  
  FROM transition_times  
  GROUP BY from_stage, to_stage)  
PIVOT(MAX(p80_days) FOR to_stage IN ('lead', 'marketingqualifiedlead', 'subscriber', 'customer'))  
ORDER BY from_stage;
```


Resultados de la consulta

Información del trabajo	Resultados	Visualización	JSON	Detalles de la ejecución	Gráfi
Fila	from_stage ▾	lead ▾	marketingqualifie...	subscriber ▾	customer ▾
1	null	0	0	2	3
2	155548387	null	null	null	null
3	customer	0	null	null	0
4	lead	0	0	null	1
5	marketingqualifiedlead	3	1	null	0
6	subscriber	0	0	0	2

Para los contactos que comenzaron sin etapa: Tardaron hasta 2 días en llegar a **subscriber** y 3 días en llegar a **customer**.



- **Entrada de Alta Calidad (0 días):** El 80% de los contactos que inician como **lead** o **marketingqualifiedlead** lo hacen **el mismo día de su creación (0 días)**
- **Entrada Ligeramente Retrasada:** El 80% de los contactos que inician como **subscriber** o **customer** tardan 2y3 días, respectivamente.
- **Velocidad Impresionante (**lead** → **MQL**):** El 80% de los **leads** se convierte en MQL el **mismo día (0 días)**

Visualizaciones

Enlace iterativo

