

## Test Técnico Marketing Science

### Ejercicio 1:

TABLE_NAME: BT_CRYPTO_TRANSACTION_HISTORY					
SITE_ID	USER_ID	PURCHASE_DATE	CRYPTO_TYPE	PURCHASE_PRICE	PURCHASE_UNITS
ARGENTINA	100000	1/1/2023	BTC	20,000	2
ARGENTINA	100012	1/1/2023	BTC	20,010	1.4
BRASIL	200234	1/1/2023	BTC	20,200	0.5
BRASIL	200234	1/2/2023	ETH	1,200	1.3
ARGENTINA	105013	2/1/2023	USDC	1	4000
ARGENTINA	116821	2/1/2023	ETH	1,350	0.5
ARGENTINA	143159	2/1/2023	BTC	19,983	5
ARGENTINA	132169	2/1/2023	ETH	1,246	1.2
MEXICO	315951	3/1/2023	BTC	20,200	0.5
MEXICO	356479	14/1/2023	ETH	1,200	1.3
BRASIL	200234	1/1/2023	USDC	1	5500
BRASIL	200234	1/2/2023	ETH	1,200	1.3

Se requiere extraer registros de la tabla **BT\_CRYPTO\_TRANSACTION\_HISTORY** y cargarlos en una nueva tabla **BT\_CRYPTO\_EVENTS**, aplicando las siguientes condiciones:

1. **Cálculo de PURCHASE\_VALUE:** La tabla **BT\_CRYPTO\_EVENTS** debe incluir una columna **PURCHASE\_VALUE**, calculada como **PURCHASE\_PRICE \* PURCHASE\_UNITS**.
2. **Conversión de PURCHASE\_DATE:** La columna **PURCHASE\_DATE** en **BT\_CRYPTO\_TRANSACTION\_HISTORY** es de tipo STRING (VARCHAR) y debe convertirse a tipo DATE en **BT\_CRYPTO\_EVENTS**.

**Responder de forma clara las siguientes preguntas, si considera necesario puede armar un SQL de apoyo:**

- 1- ¿Es necesaria una tabla adicional para un proceso ETL que cargue registros en **BT\_CRYPTO\_EVENTS** bajo las condiciones especificadas?
- 2- Describa un proceso ETL de al menos 4 pasos para la carga de registros.
- 3- Tras la carga inicial, el proceso ETL debe actualizar diariamente los datos correspondientes a los últimos cinco días. Esto incluye la adición de nuevos registros, la modificación de los existentes y la desactivación de aquellos que hayan sido eliminados. Es fundamental que el ETL garantice la integridad de los datos antes de su despliegue en producción en la tabla **BT\_CRYPTO\_EVENTS**. Describa el proceso ETL propuesto para abordar esta casuística.

### Ejercicio 2:

**Se busca el diseño de un pipeline integral para un modelo de predicción de Churn. El**

**proyecto debe estar organizado siguiendo las mejores prácticas de MLOps, y se requiere el desarrollo de los scripts necesarios para cada etapa.**

Cabe destacar que el rendimiento del modelo no será evaluado, por lo que se recomienda implementar una regresión simple. El enfoque y las opciones consideradas serán discutidos durante la entrevista.

## **Descripción del Problema y Modelo Propuesto**

El objetivo es desarrollar un modelo predictivo que determine la probabilidad de que un usuario de Mercado Pago (MP) en Argentina realice nuevas transacciones en los próximos dos meses. Para ello, se cuenta con los siguientes conjuntos de datos:

1. **PAYMENTS.csv:** Historial transaccional de MP de los últimos seis meses (segmento Wallet). Incluye:
  - **FECHA:** Fecha de la transacción exitosa.
  - **CUS\_CUST\_ID\_SEL:** ID del vendedor.
  - **CUS\_CUST\_ID\_BUY:** ID del comprador.
  - **SPENT:** Monto gastado.
  - **TPV\_SEGMENT\_DETAIL:** Flujo de uso (ej., Instore, Transport, Utilities, transferencias, recargas).
  - **DESCUENTO:** Descuento aplicado por MP.
2. **ACTIVE\_USER.csv:** Datos sobre la actividad del usuario en las aplicaciones. Incluye:
  - **CUS\_CUST\_ID\_BUY:** ID del comprador.
  - **MAU\_MP\_1:** Cantidad de accesos a la aplicación de Mercado Pago (azul) en el último mes.
  - **MAU\_ML\_1:** Cantidad de accesos a la aplicación de Mercado Libre (amarilla) en el último mes.
  - **MAU\_ML\_2:** Cantidad de accesos a la aplicación de Mercado Libre (amarilla) en el penúltimo mes.
  - **last\_login\_mp\_date\_1:** Fecha del último inicio de sesión en la aplicación de Mercado Pago en el mes -1.
3. **DEMOGRAFICOS.csv:** Información demográfica (edad, sexo, tipo de tarjeta, lugar de residencia).
4. **DINERO\_CUENTA.csv:** Datos sobre el saldo de la cuenta y el estado de inversión del usuario. Incluye:
  - **PLATA\_CUENTA\_1:** Saldo máximo en cuenta en el último mes.
  - **PLATA\_CUENTA\_2:** Saldo máximo en cuenta en el penúltimo mes.
  - **INVERSION:** Estado de inversión (Investing, Elegible, Warmup, Pending).
5. **MARKETPLACE\_DATA.csv:** Resumen de la actividad del usuario en Mercado Libre (ML) durante los últimos seis meses. Incluye:
  - **SPENT\_ML:** Dinero gastado en Mercado Libre.
  - **REGENCY\_ML:** Fecha de la última compra en Mercado Libre.
  - **FREQUENCY\_ML:** Cantidad de días con compras.