

# Contexto

Se cuenta con un archivo de Excel por cada sucursal.

Todos los archivos tienen el mismo formato y los mismos campos.

La cantidad de sucursales puede variar: se pueden agregar o quitar sin previo aviso.

Cada archivo contiene una hoja por mes.

En cualquier momento se pueden agregar o quitar meses de información.

El precio de los productos está expresado en ARS.

---

# Objetivo

Construir un dataset único que permita responder las siguientes preguntas:

## 1. Exploración y limpieza

- Analizar y depurar los datos: valores faltantes, duplicados, inconsistencias y outliers.

Repuesta:

En los scripts: ventas\_ingest\_raw\_to\_bronze.py, ventas\_transform\_bronze\_to\_silver.py y ventas\_aggregate\_silver\_to\_gold.py

## 2. Conversión de moneda

- Llevar las ventas y costos a USD utilizando el valor de cierre del mes correspondiente a la venta.

Respuesta:

En el script ventas\_transform\_bronze\_to\_silver para las tablas en silver

## 3. Análisis de margen

- Determinar el producto que mayor margen deja.

Respuesta:

En la tabla mailamericas\_gold.ventas

```
SELECT desc_articulo, SUM(total_margen_ars) AS margen_total
FROM mailamericas_gold.ventas
GROUP BY desc_articulo
ORDER BY margen_total DESC
LIMIT 1;
```

#### 4. Estacionalidad de ventas

- Identificar en qué momento del mes mejoran las ventas.
- Identificar en qué momento de la semana mejoran las ventas.
- Detectar si existen sucursales que no sigan esa tendencia.

Respuesta:

En la tabla mailamericas\_gold.ventas

```
SELECT sucursal, dia_mes_top_ventas, SUM(total_venta_ars) AS total
FROM mailamericas_gold.ventas
GROUP BY sucursal, dia_mes_top_ventas
ORDER BY total DESC;
```

```
SELECT sucursal, dia_semana_top_ventas
FROM mailamericas_gold.ventas
GROUP BY sucursal, dia_semana_top_ventas;
```

#### 5. Cumplimiento de objetivos

- Clasificar las ventas de cada sucursal según su objetivo mensual con las categorías: superó, igualó o no alcanzó.
- 

Respuesta:

En la tabla mailamericas\_gold.ventas

```
SELECT sucursal, cumplimiento_objetivo, COUNT(*) AS total_meses
FROM mailamericas_gold.ventas
GROUP BY sucursal, cumplimiento_objetivo;
```

#### 6. Calidad de datos

- Proponer reglas de control de calidad: qué validar, cómo y en qué etapa del proceso aplicarlo.

Repuesta:

En los scripts: ventas\_ingest\_raw\_to\_bronze.py, ventas\_transform\_bronze\_to\_silver.py y ventas\_aggregate\_silver\_to\_gold.py

---

## Métricas

- Ventas: `VALOR_ARTICULO * CANTIDAD_VENDIDA`
  - Costos: `COSTO_ARTICULO * CANTIDAD_VENDIDA`
  - Margen: `VENTAS – COSTOS`
- 

## Bonus Track

Imaginando un escenario con millones de registros por día:

1. Proponer un diagrama de arquitectura escalable que incluya:
  - Origen de los datos: S3
  - Proceso de ingesta: Glue
  - Transformaciones necesarias: Glue
  - Almacenamiento: Redshift
  - Integración con la capa de visualización: Quicksight o Power BI
2. Indicar en qué parte del flujo intervendrá el proceso orquestador y de qué manera.

Al momento de hacer la ingesta y todos los procesos de transformaciones

3. Definir una arquitectura de un chatbot de AI.
  - PLUS OPCIONAL entregar un chatbot de AI que pueda consultar esta data.

La mencionada y que el chatbot se alimente de la ultima capa (gold)