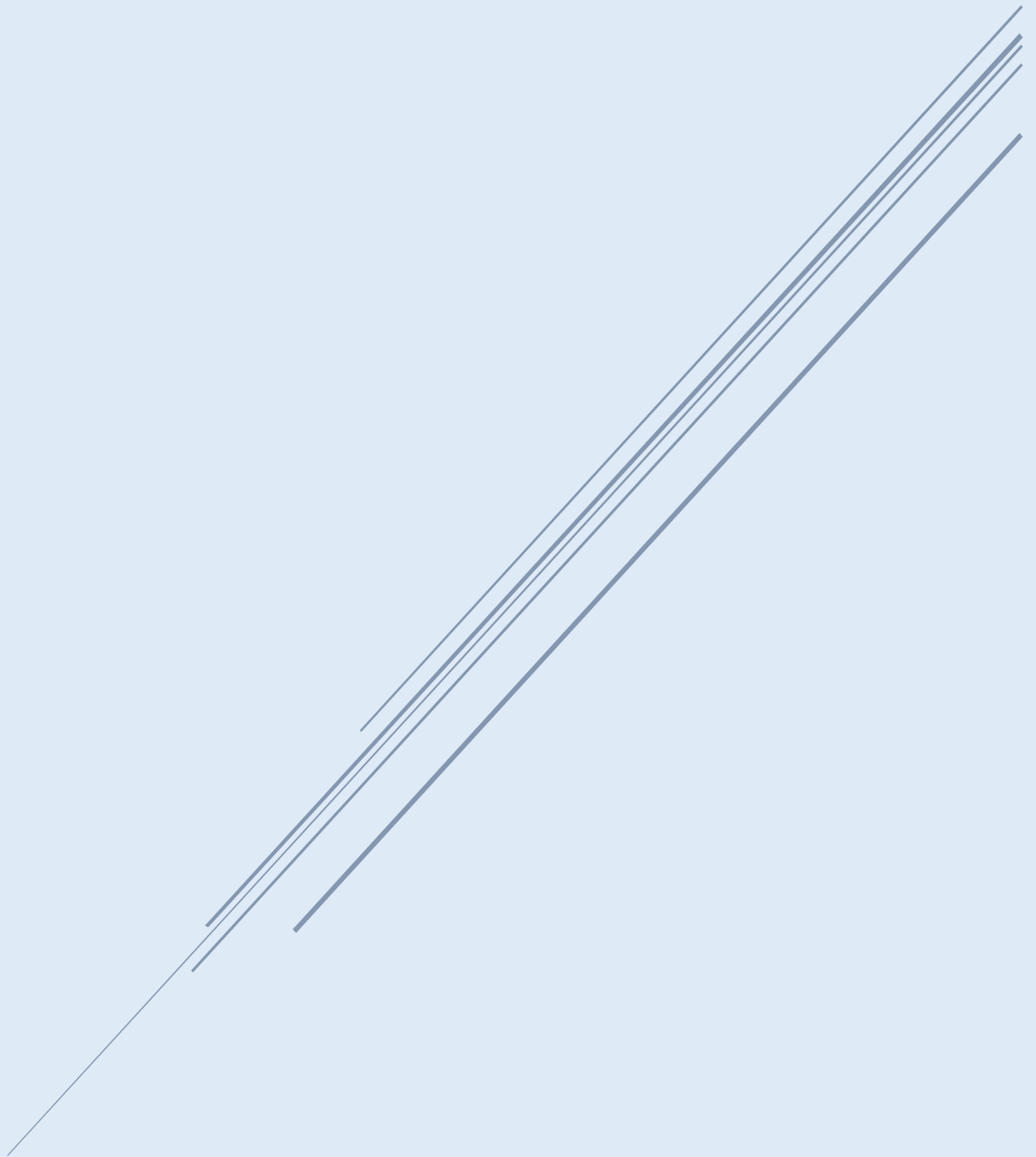


ANALISIS DE DATOS EN MYSQL SERVER

Este proceso forma parte del proyecto de exploración y análisis de datos de un negocio ecommerce de productos electrónicos.

[Link del proyecto](#)



Paul Chipana Muñiz

Análisis de datos en MySQL Server

Luego de garantizar la calidad de los datos con los procesos de Modelado y Transformación de datos, buscaremos responder las preguntas clave detalladas anteriormente interactuando con el lenguaje MySQL. Este análisis proporciona información esencial para tomar decisiones informadas y estratégicas, permitiendo comprender mejor el rendimiento y el compromiso de los clientes en línea del ecommerce de productos electrónicos.

Una vez dentro de la conexión con MySQL server, procedemos a responder las preguntas clave:

```
-----  
-- DATA ANALYSIS PROCESS --  
-----
```

```
-----  
-- PREGUNTAS A RESPONDER --  
-----
```

```
-- 1. Cuanto se vendió cada mes ( septiembre 2020 - febrero 2021)  
-- 2. A qué hora del día se vende más ( septiembre 2020 - febrero 2021)  
-- 3. Qué porcentaje (%) del total de visitantes web se quedaron en los  
siguientes tipos de eventos:{solo visita, carrito, remover carrito, compra}  
-- 4. Cuáles son las 5 categoría mas vendida (calcular en VENTAS TOTALES)  
durante los 2 ultimos meses ( Enero y Febrero 2021)
```

```
-----  
-- RESPONDIENDO PREGUNTAS --  
-----
```

```
-- 1. Cuanto se vendió cada mes ( septiembre 2020 - febrero 2021)
```

```
-- Luego de transformar los datos, quitando UTC al final de los datos del campo  
event_time PARA CAMBIAR EL DATATYPE de VARCHAR a DATETIME  
-- Ahora podemos usar funciones como 'MONTH' o 'HOUR' para extraer el número de  
mes de cada registro con formato de fecha y hora (datetime format)
```

```
-- Calculando Ventas por mes
```

```
USE electronics_store_project;  
SELECT event_type, event_time, MONTH(event_time) AS mes_numero, user_id, price  
FROM ecommerce_events_history_in_electronic_store  
WHERE event_type = "purchase";  
SELECT  
    event_type,  
    YEAR(event_time) AS año_numero,  
    MONTH(event_time) AS mes_numero,  
    COUNT(user_id) AS total_de_compradores,  
    COUNT(user_session) AS total_de_sesiones_finalizadas_en_compra,  
    SUM(price) AS Ventas_mensuales  
FROM ecommerce_events_history_in_electronic_store  
WHERE event_type = "purchase"  
GROUP BY año_numero, mes_numero;
```

	event_type	año_numero	mes_numero	total_de_compradores	total_de_sesiones_finalizadas_en_compra	Ventas_mensuales
▶	purchase	2020	9	1007	1007	96353.01000000017
	purchase	2020	10	6276	6276	562590.38999999916
	purchase	2020	11	7574	7574	787886.23999999866
	purchase	2020	12	6531	6531	813339.86999999908
	purchase	2021	1	8315	8315	1488410.57000000024
	purchase	2021	2	7643	7643	1376815.5400000105

-- 2. A qué hora del día se vende más (septiembre 2020 - febrero 2021)

```

SELECT
    event_type,
    COUNT(user_id) AS total_de_compradores,
    COUNT(user_session) AS total_de_sesiones_finalizadas_en_compra,
    HOUR(event_time) AS hora_del_dia_numero,
    SUM(price) AS Ventas_totales
FROM ecommerce_events_history_in_electronic_store
WHERE event_type = "purchase"
GROUP BY hora_del_dia_numero
ORDER BY SUM(price) desc;

```

	event_type	total_de_compradores	total_de_sesiones_finalizadas_en_compra	hora_del_dia_numero	Ventas_totales
▶	purchase	2496	2496	11	339988.7
	purchase	2535	2535	10	325458.00999999925
	purchase	2282	2282	12	312623.31999999996
	purchase	2085	2085	17	303283.17999999997
	purchase	2474	2474	9	302129.85999999999
	purchase	2160	2160	14	298496.95000000003
	purchase	2171	2171	13	298443.909999999974
	purchase	2019	2019	16	280653.470000000015
	purchase	1934	1934	19	278956.37999999997
	purchase	1957	1957	15	273321.05000000005
	purchase	2215	2215	8	273187.94000000006
	purchase	1950	1950	18	265216.23000000002
	purchase	2000	2000	7	258770.500000000015
	purchase	1691	1691	6	226600.579999999978
	purchase	1533	1533	20	215278.770000000046
	purchase	1041	1041	21	163331.36999999999
	purchase	1165	1165	5	155351.430000000014
	purchase	701	701	22	110509.860000000004
	purchase	788	788	4	99933.660000000003
	purchase	575	575	3	86117.000000000001
	purchase	482	482	23	82359.580000000016
	purchase	361	361	0	71119.270000000005
	purchase	403	403	2	55409.470000000002
	purchase	328	328	1	48855.1300000000056

- Como puedo observarse, durante el rango de 9am a 12pm son las horas en las que más se vende
- Asimismo, es importante resaltar la poca diferencia en volumen de ventas que presentan los 11 volúmenes más altos del ranking, que comprende el rango entre 09hrs a las 19hrs 59m, excluyendo el rango de 18hrs - 18hrs 59m.

-- 3. Qué porcentaje (%) del total de visitantes web se quedaron en los siguientes tipos de eventos:{solo visita, carrito, remover carrito, compra}

```
SELECT
    event_type,
    COUNT(user_session) AS Recurrencia_por_tipo_de_evento,
    COUNT(event_type) / (SELECT count(event_type) FROM
ecommerce_events_history_in_electronic_store)*100 AS
Porcentaje_segun_eventos_totales
FROM ecommerce_events_history_in_electronic_store
GROUP BY event_type;
```

	event_type	Recurrencia_por_tipo_de_evento	Porcentaje_segun_eventos_totales
▶	view	793748	89.6760
	cart	54035	6.1048
	purchase	37346	4.2193

-- 4. Cuáles son las 5 categoría mas vendida (calcular en VENTAS TOTALES) durante los 2 ultimos meses (Enero y Febrero 2021)

-- NOTA: Usaremos category_id y no category_code, porque el 2do campo contiene demasiados valores en blanco (según análisis exploratorio)

```
SELECT
    event_type,
    category_id,
    category_code,
    COUNT(user_id),
    SUM(price) AS Ventas_totales
FROM ecommerce_events_history_in_electronic_store
WHERE event_type = "purchase" and YEAR(event_time) = "2021"and MONTH(event_time)
in (1,2)
GROUP BY category_id,category_code -- al agrupar primero por category_id (campo
sin nulos y blancos), evitamos resultados sesgados debido a la gran cantidad de
blancos encontrados en category_code
ORDER BY SUM(price) desc
limit 5;
```

	event_type	category_id	category_code	cantidad_de_usuarios	Ventas_totales
▶	purchase	2144415922427789416	computers.components.videocards	5081	1955753.6800000283
	purchase	2144415923107266682	computers.peripherals.printer	612	105423.1400000002
	purchase	2144415926966026460	computers.components.cpu	434	73466.71000000015
	purchase	2144415922050302046		122	42793.31999999999
	purchase	2144415924491387038	computers.components.motherboard	492	39722.670000000064

- La ventas de la categoría "computers.components.videocards" representa casi 20 veces el volumen de la 2da categoría con volúmenes de venta más altos "computers.peripherals.printer"
- Asimismo, es importante resaltar la diferencia en la cantidad de usuarios compradores de "computers.components.videocards" y la cantidad de usuarios compradores de "computers.peripherals.printer", el primero tiene 5081 usuarios compradores y el segundo, 612. Es importante calcular el ticket promedio por usuario para poder complementar la información que actualmente tenemos sobre las volúmenes de venta de las 5 categorías mas vendidas

-- Añadiendo información sobre ticket promedio por usuario comprador por cada categoría, esto nos brinda información relevante sobre la categoría de producto con mayor volumen de venta por comprador.

```
SELECT
    event_type,
    category_id,
    category_code,
    COUNT(user_id),
    SUM(price) AS Ventas_totales,
    SUM(price)/COUNT(user_id) as ticket_promedio_por_usuario
FROM ecommerce_events_history_in_electronic_store
WHERE event_type = "purchase" and YEAR(event_time) = "2021"and MONTH(event_time)
in (1,2)
GROUP BY category_id,category_code -- al agrupar primero por category_id (campo
sin nulos y blancos), evitamos resultados sesgados debido a la gran cantidad de
blancos encontrados en category_code
ORDER BY SUM(price) desc
limit 5;
```

	event_type	category_id	category_code	cantidad_de_usuarios	Ventas_totales	ticket_promedio_por_usuario
►	purchase	2144415922427789416	computers.components.videocards	5081	1955753.6800000283	384.91511119858853
	purchase	2144415923107266682	computers.peripherals.printer	612	105423.1400000002	172.26003267973888
	purchase	2144415926966026460	computers.components.cpu	434	73466.71000000015	169.27813364055334
	purchase	2144415922050302046		122	42793.319999999999	350.7649180327868
	purchase	2144415924491387038	computers.components.motherboard	492	39722.670000000064	80.73713414634159

- Como puede observarse, tenemos una categoría con ID(2144415922050302046) , pero sin nombre, que representa el 2do ticket más alto de volumen de venta por usuario comprador (350.76) dentro de las 5 categorías con volúmenes más altos de venta.