

Tercera entrega del PF

“E-Commerce Worldwide”



- Alumna *Gimena Cruz*
- Comisión *534485*
- Profesor *Sebastián Kuperman*
- Tutor *Fernando Rubin*

DESCRIPCION DE LA DATOS TEMATICA DE LOS OBTENIDOS

Para el desarrollo de este Proyecto Final se eligieron datos y aspectos relacionados con el comercio electrónico a nivel mundial (moda y accesorios), para lo cual se tomo referencia y se extrajo la informacion del sitio <https://www.kaggle.com/datasets/mustafakeser4/looker-ecommerce-bigquery-dataset>, de un trabajo titulado “Looker Ecommerce BigQuery Dataset “ (CSV version of BigQuery Looker Ecommerce Dataset”), que es una Base de Datos brindada por Looker, Plataforma de BI unificada.-

Se hizo necesaria la incorporación de una base de datos de geolocalización de países desde el sitio https://developers.google.com/public-data/docs/canonical/countries_csv?hl=es-419.-

DATASET

Se adjunta en formato xls y csv bases de datos organizada a los fines de normalizarla y optimizarla en Power BI.-

METODOLOGIA DE TRABAJO

Para el análisis de los mismos, se adoptó una metodología de trabajo en donde el Objetivo concreto (Specific) es determinar consumos, tendencias, geolocalización y otros aspectos con la finalidad de potenciar y mejorar las ofertas a las demandas mundiales en el comercio electrónico particularmente moda y accesorios.-

Podemos determinar (Mesurable) cantidades, porcentajes, variaciones interanuales y diferentes valores para una comparación a nivel global y particular en cada area que precisemos.-

Con los datos concretos historicos (Attainable) se facilita la investigacion y proporciona las respuestas necesarias para evacuar las preguntas que se presentan.-

Asimismo todos estos datos son de fundamental importancia (Relevant) a la hora de optimizar los recursos para incrementar las ventas, proposito que buscamos de forma agil y rápida (Timely) para corregir inmediatamente fallas o errores y evitar la perdida de recursos.-

OBJETIVO del PROYECTO

El Objetivo concreto (Specific) del Proyecto Final es confirmar o desestimar la hipótesis establecida y determinar:

- 1) diferentes localizaciones de transacciones comerciales para hacer distintos estudios basados en su geolocalización como por ej. abrir, cerrar y/o unificar centros de distribución;
- 2) identificar zonas de mayor y menor consumos, para establecer las estrategias de ventas acordes a cada caso;
- 3) Establecer por características y categoría del producto sus tendencias;
- 4) establecer el “ estado” en las ordenes de transacciones comerciales para determinar por zona su optimización como “shipped” o “completed”;
- 5) otros aspectos que puedan surgir con la finalidad de potenciar y mejorar en el mundo el comercio electrónico de estas compañías en particular.-

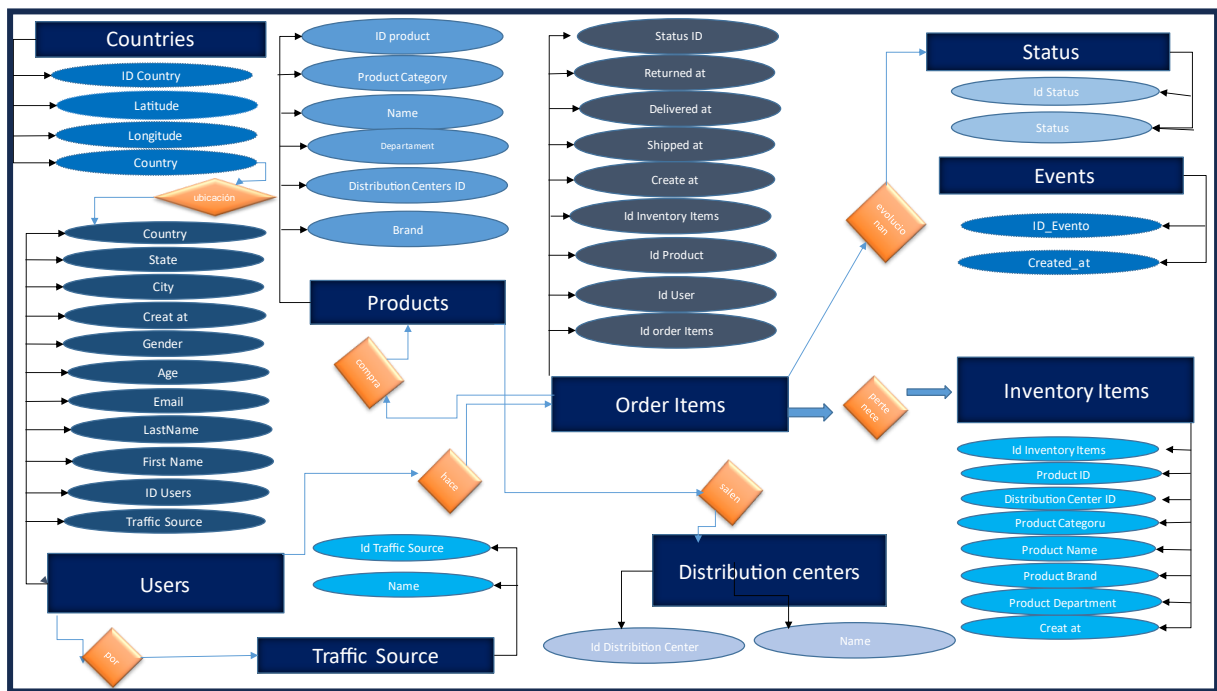
Junto con el equipo de trabajo analizaremos e interpretaremos los datos los cuales tendrán un destino dirigido al Area Estratégica y Táctica responsables de gestionar todo el E-commerce.-

HIPOTESIS

Hipótesis Primera: Según distintas encuestas realizadas en Agosto 2022 sobre las relaciones de transacciones comerciales via Internet (e-commerce) en el ámbito moda y accesorios en particular, se determina un porcentaje mayor al 80% de mujeres que compran moda online, en comparación con un porcentaje en torno al 75% de hombres que compran online.- Hipótesis segunda: La implementación de Eventos promotores de Marketing Digital, incrementa en forma directamente proporcional la participación de Users y Transacciones comerciales digitales.-

Además de esta comprobación o rechazo de hipótesis estableceremos como Objetivos del Proyecto, con una base de datos Descriptiva, dotar de toda la información necesaria para poder determinar con los distintos analisis, que áreas geolocalizadas pueden ser de potenciales clientes, donde esta el mayor mercado de transacciones comerciales para fortalecerlo y acrecentarlo aun mas, trasladar ese tipo de actividad promocional a zonas potenciales y asi ampliar y reforzar el e-commerce en zonas de interés.-

DIAGRAMA ENTIDAD -RELACION PROYECTO FINAL



LISTADO de TABLAS

A continuación se realiza una breve descripción de cada Tabla con sus respectivas claves primarias y foráneas : 1) USERS contiene los datos de los usuarios (potenciales compradores) como nombre, ciudad, fecha de registración, etc. y sus claves son: PK=ID Users, FK=Country Name.

2) PRODUCTS contiene nombre y características del producto, sus claves son : PK=ID Producto, FK=Distribucion Center ID.

3) ORDER ITEMS contiene el detalle de los pedidos de productos y sus características como fecha, status, etc. y sus claves son : PK=ID Order Item y FK= ID User, ID Product, ID Inventory Items, ID-

4) INVENTORY ITEMS contiene características de todos los productos potenciales a ser vendidos y sus centros de distribución. Sus claves son : PK=ID Inventory Item y FK= Producto ID y Distribution Center ID.-

5) STATUS nos indica el estado de la transacción comercial. La clave es PK=IDStatus .-

6) DISTRIBUTION CENTER nos indica el centro que distribuye los productos comercializados. Su clave PK=ID Distribution Center.-

7) COUNTRIES LATITUDE LONGITUDE nos indica la localización exacta del país a los fines de poder ubicarlo en el mapa interactivo. Su clave PK=Country Name.

8) TRAFFIC SOURCE nos indica el origen, la fuente del trafico de búsqueda, es decir a través de que motor de búsqueda llega el cliente a nuestro sitio web. Sus claves son PK=ID Traffic Source.-

9) EVENT nos indica eventos promocionales de Marketing Digital.- Su PK = ID Event.-

Podemos determinar el modo en que las tablas se relacionan entre si, siendo a mi entender ORDER ITEMS la principal, debido a que contiene casi a todas las demás tablas en sus FK.- Con los datos de cada pedido de producto (ORDER ITEMS), tenemos la referencia del comprador, del producto, del estado del producto; con los datos de cada producto(PRODUCTS) tenemos la referencia del usuario y del centro de distribución.- Asimismo otras tablas contienen otras relaciones importantes a la hora de definir motores de búsqueda (Tabla Traffic Source con Users) y demás relaciones de interés que surjan a posteriori.-

A continuación, se detallan todas las Tablas con sus respectivos Campos, Tipo de Datos y Tipo de Claves.-

TABLA USERS		
CAMPO	TIPO CAMPO	TIPO CLAVE
ID Users	INT	PK
First Name	VARCHAR	
Last Name	VARCHAR	
Email	VARCHAR	
Age	INT	
Gender	VARCHAR	
Creat at	DATETIME	
City	VARCHAR	
State	VARCHAR	
Country	VARCHAR	FK
ID Traffic Source	VARCHAR	FK

TABLA_PRODUCTOS

CAMPO	TIPO CAMPO	TIPO CLAVE
Id Producto	INT	PK
Prod.Category	VARCHAR	
Name	VARCHAR	
Brand	VARCHAR	
Department	VARCHAR	
Distr Center ID	INT	FK

TABLA_COUNTRIES_LATITUDE_LONGITUDE

CAMPO	TIPO CAMPO	TIPO CLAVE
ID Countries	VARCHAR	
Latitude	VARCHAR	
Longitude	VARCHAR	
CountryName	VARCHAR	PK

TABLA_STATUS

CAMPO	TIPO CAMPO	TIPO CLAVE
ID Sttatus	INT	PK
Status	VARCHAR	

TABLA DISTRIBUTION CENTERS

TIPO CAMPO	TIPO CAMPO	TIPO CLAVE
ID Distribution		
Center	INT	PK
Name	VARCHAR	

TABLA INVENTORY ITEMS

TIPO CAMPO	TIPO CAMPO	TIPO CLAVE
ID Inventory Item	INT	PK
Product ID	INT	FK
Creat at	DATETIME	
Product Category	VARCHAR	
Product Name	VARCHAR	
Producto Brand	VARCHAR	
ID Distri Center	INT	FK
Product department	VARCHAR	

TABLA ORDER ITEMS

TIPO CAMPO	TIPO CAMPO	TIPO CLAVE
ID Order Item	INT	PK
User ID	INT	FK
Product ID	INT	FK
Inventory Item ID	INT	FK

Status ID	VARCHAR	FK
Created at	DATETIME	
Shipped at	DATETIME	
Delivered at	DATETIME	
Returned at	DATETIME	

TABLA TRAFFIC SOURCE

CAMPO	TIPO CAMPO	TIPO CLAVE
ID Traffic Source	INT	PK
Name	VARCHAR	

TABLA EVENTS

CAMPO	TIPO CAMPO	TIPO CLAVE
Event_Type		
ID	INT	PK
Created_at	DATE	

Mockup

E-COMMERCE WORLDWIDE

PROYECTO FINAL (Dashboard)

TRANSACTIONS

USERS

MEMBERS

ORDERS

ITEMS

UPDATES

Logos

Links

YEARS

GENDER

COUNTRY

DISTRIB.CENTER

STATUS

TRAFFIC SOURCE

✓ Alumna Gimena Cruz

✓ Prof. Sebastián Kuperman

✓ Tutor Claudia Courau

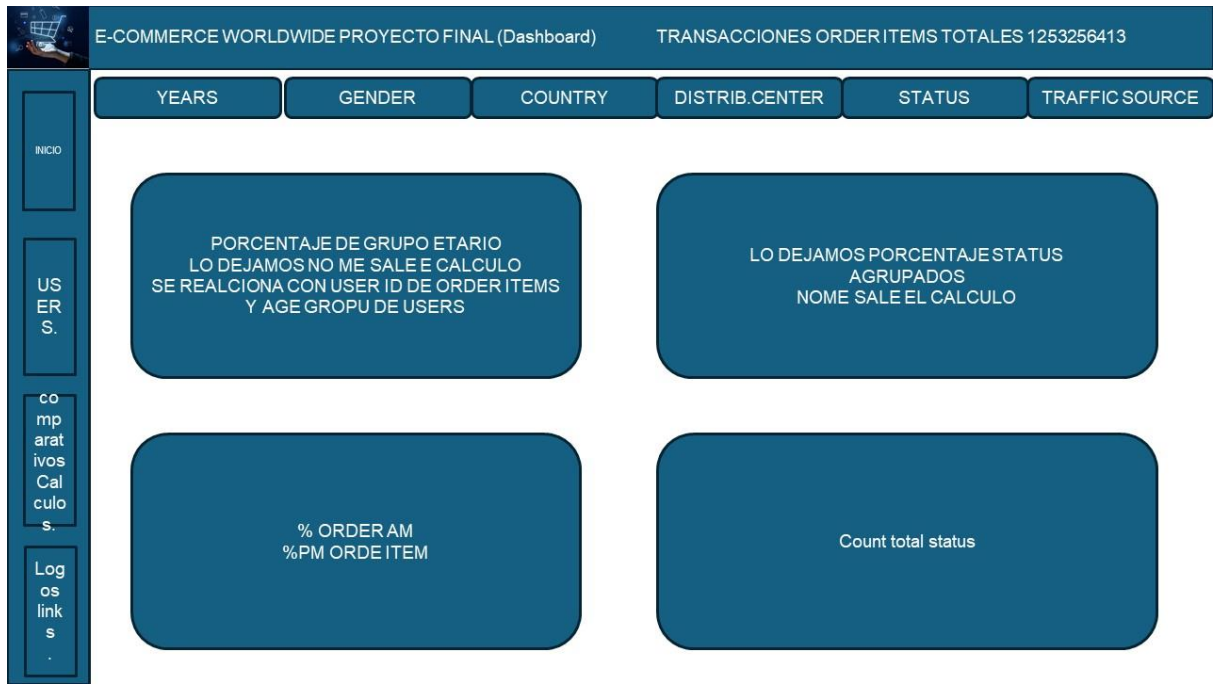
✓ Fecha Actualiz. 22/04/2024

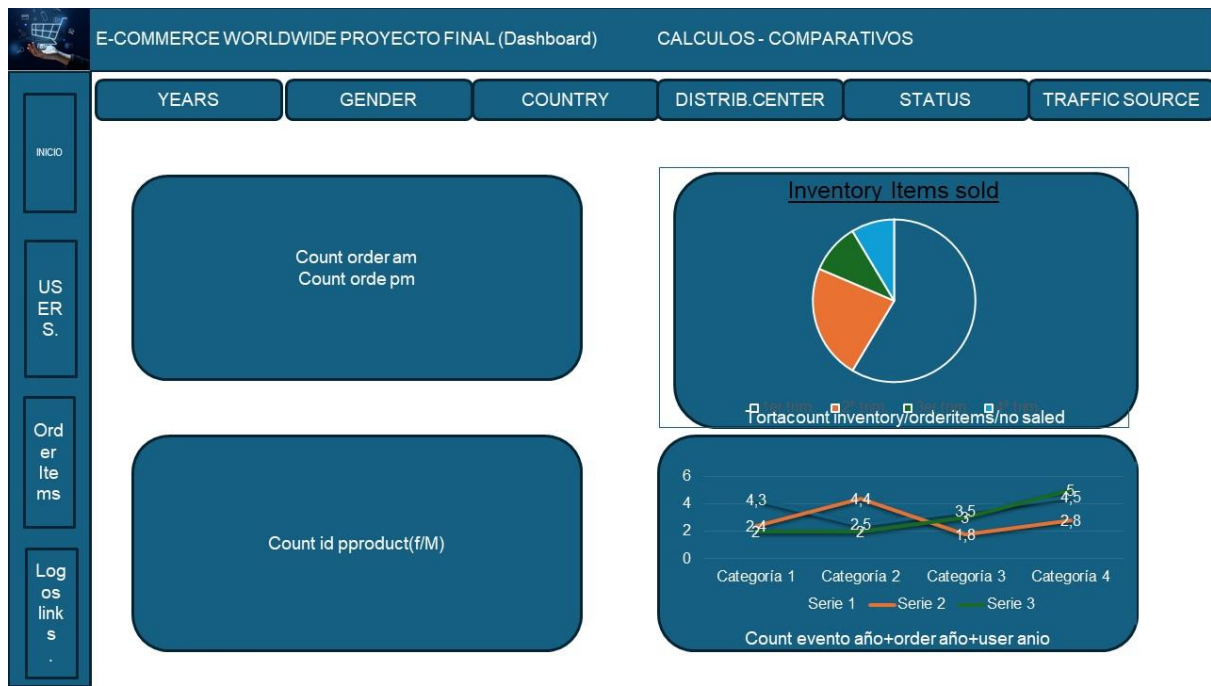
TRANSACCIONES TOTALES

CLIENTES TOTALES

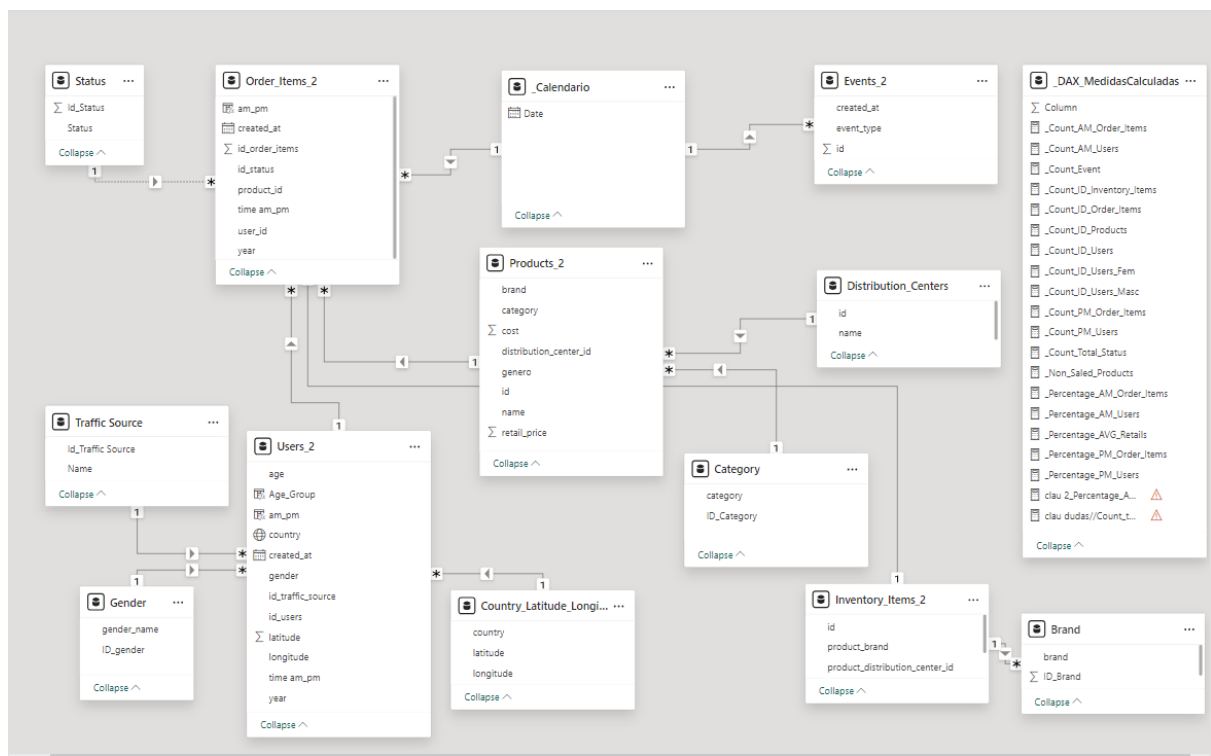
AVG % REATIL

TENDENCIA Clientes Transacciones Por year





RELACIONE POWE BI



Transformaciones

Se cambian el tipo de archivo de csv a xls para hacer su transformacion a tablas.- Se crean nuevas tablas (Countries_Latitude_Longitude, Distribution Centers, Status) desconociendo

que podíamos crearlas desde Power Bi.-Se efectuaron modificaciones minimas en tablas, desconociendo que podíamos hacerlo desde PowerBi, las cuales volveran a revisarse por completo desde PowerBi (normalizacion de formatos de datos, estandarizar nombres de columnas, particularmente Gender que posee campos distintos F/Women, M/Men). En formato de fechas se debera controlar la adecuada Date/Time? Y CREAR NUEVA TABLA DE TRAFFIC SOURCE CON ID PARA COUNT. NO ME COINCIDEN LAT LONG DE USERS C/COUNTRIES LAT LONG, se creara una tabla Country Lat Long con datos de Tabla Users.

Se confecciona MOCKUP con seleccion imagenes <https://thelogisticsworld.com/logistica-comercio-electronico/10-tendencias-de-e-commerce-que-marcaran-el-segundo-semester-de-2023/>.- El DIAGRAMA ER con Fuente Calibri 11 campo/atributos, Fuente Calibri 18 nombre tablas, se agregan columna Latitud/Longitud.- Se normaliza nombre (Arial) y tamaño de fuentes de títulos (40), subtítulos (28), saltos de página(12), texto(15).

En POWER BI se cargan tablas (formato csv de Products y Users), por indicacion del profesor se vuelven a cargar con formato excel tabla porque daba error. Se incorpora otra tabla a la base de datos para incrementar la informacion y fortalecer la hipotesis que se amplio agregando otra variable mas.-De formato CSV se pasa a xls y con editor se eliminan nulls, vacios porque el archivo es muy grande . Se eliminan columnas .

En POWER BI/Power Query se comienzan a normalizar, generar y modelar tablas.-Se crea tabla calendario y tabla Dax Medidas calculadas.- Tabla Inventory Items, eliminamos NULLS, created at, product category . Tabla order Items se crea nueva columna YEAR, Transform DATE en year, se crea columna time AM/PM, se eliminan tablas csv que estan duplicadas, se eliminan columnas shipped, returned y delivered at, inventory items ID, order ID. Se creo tabla Genero con ID para simplificar.Se reemplazan valores columna genero Tabla Products, se elimina columna sku. Tabla Users se eliminan columnas de nombres, telefonos y mails, se reemplazan valores de genero, se eliminan columnas de direccion, estado, codigo postal y ciudad, se crean columnas year y Time am/pm. Tabla Events se eliminan columnas SEQUECY, SESSION, city, state, postal code, browser, trafic source, uri, se eliminan cancelados.Se crean medidas calculadas que se detallaran abajo. Se crean tablas de dimension BRAND, CaTEGORY para relacionarlas con calculos de productos.-Se vuelve a la tabla origen de Products para modificar columnas de cost y retail_price formato numero sin decimales, luego se modificara en powerbi la cantidad de numeros para darle coherencia.-El error se mantiene, se tratan de modificar los numeros desde powerbi.con transform.Calculated Absolute Value Base 10 logarithm y luego Fixed decimal number con 2 decimales.-La tabla creada Country Lat-Long, se modifica en power bi sacando los datos desde la tabla Users con campos country, latitud, longitud.-En tabla Users_2 se crea nueva columna con calculo de edades para grupos etarios. Se crea columna nueva en Users Age_Group .- Se reemplazan valores de traffic source por ID en Tabla Users y nombre columna id_traffic_source.-se reemplazan valores de status por ID EN Tabla Order Items y nombre columna id_status.-

Formulas Dax

1)_Count_ID_Inventory_Items =

//creamos cantidad Inventario de productos

```
COUNT(  
    Inventory_Items_2[id]  
)
```

2)_Count_ID_Order_Items =

//creamos contar ordenes de productos

```
COUNT(  
    Order_Items_2[id_order_items]  
)
```

3)_Count_ID_Products =

//contamos cantidad productos

```
COUNT(  
    Products_2[id]  
)
```

4)_Count_ID_Users =

//contamos cantid usuarios

```
COUNT(  
    Users_2[id_users]  
)
```

5)_Count_ID_Users_Fem =

//contamos cant. fem

```
CALCULATE(  
    COUNT(  
        Users_2[id_users]),  
        Users_2[gender]="1"  
    )  
)
```

6)_Count_ID_Users_Masc =

//contamos cant. masc

```
CALCULATE(  
    COUNT(  
        Users_2[id_users],  
        Users_2[gender]="2"  
    )
```

```
7)_Count_AM_Users =
```

```
//contamos users creados AM
```

```
CALCULATE(  
    COUNT(  
        Users_2[id_users],  
        Users_2[am_pm]="AM"  
    )
```

```
8)_Count_PM_Users =
```

```
//contamos users creados PM
```

```
CALCULATE(  
    COUNT(  
        Users_2[id_users],  
        Users_2[am_pm]="PM"  
    )
```

```
9)_Count_PM_Order_Items =
```

```
//contamos pedidos creados PM
```

```
CALCULATE(  
    COUNT(  
        Order_Items_2[id_order_items],  
        Order_Items_2[am_pm]="PM"  
    )
```

```
10)
```

```
_Count_AM_Order_Items =
```

```
//contamos pedidos creados AM
```

```

CALCULATE(
    COUNT(
        Order_Items_2[id_order_items]),
        Order_Items_2[am_pm]="AM"
    )
11)
_Percentage_AM_Users =
//calculamos porcentaje creacion AM Users
DIVIDE(
    _DAX_MedidasCalculadas[_Count_AM_Users],
    _DAX_MedidasCalculadas[_Count_ID_Users],BLANK()

)
12)
_Percentage_PM_Users =
//calculamos porcentaje creacion PM Users
DIVIDE(
    _DAX_MedidasCalculadas[_Count_PM_Users],
    _DAX_MedidasCalculadas[_Count_ID_Users],BLANK()

)
13)
_Non_Saled_Products =
//se crea medida para calcular cant prod no vendidos
_DAX_MedidasCalculadas[_Count_ID_Inventory_Items] -
_DAX_MedidasCalculadas[_Count_ID_Order_Items]

14)
_Count_Event =

```

//CONTAMOS EVENTOS por la fecha, no se pudo hacer id de eventos por la gran cantidad y no es relevante

```
COUNT(  
    Events_2[created_at]  
)  
15)  
_Percentage_AVG_Retails =
```

//calculamos porcentaje de los promedios de retail y costos

```
DIVIDE(  
    AVERAGE(  
        Products_2[retail_price]),  
    AVERAGE(  
        Products_2[cost]  
    ),BLANK()  
)
```

Creacion de Tablas y/o columnas nuevas y/o calculadas

1)Age_Group =

//creo columna grupo etario para mejorar el calculo por edades

```
IF(  
    Users_2[age]<=20,"0-20",  
    IF(  
        Users_2[age]>20 && Users_2[age] <=30,"21-30",  
        IF(  
            Users_2[age]>30 && Users_2[age] <=40,"31-40",  
            IF(  
                Users_2[age]>40 && Users_2[age] <=50,"41-50", "+ 60"  
            )  
        )  
    )
```

)))

2)_Calendario =

//CREO TABLA CALENDARIO

```
CALENDAR(  
  MIN('Order_Items_2'[created_at]),  
  MAX('Order_Items_2'[created_at])  
  
)
```

3)am_pm =

//se crea columna am/pm para determinar horario creacion usuarios

```
IF(Users_2[time am_pm]<="12","AM","PM"  
)
```

4)am_pm =

//se crea columna am/pm para determinar horario creacion order items

```
IF(  
  Order_Items_2[time am_pm]<="12","AM","PM"  
)
```

5) se hace un join/merged en tabla products para poder relacionarla con la tabla Brand pero no me da correcto

