

BIG DATA

PROYECTO FINAL

Jonathan Perez Sedova



Informe metodológico – Proyecto de Business Intelligence

Contexto

Durante el curso de Business Intelligence se nos propuso elaborar un proyecto final relacionado con el impacto mundial del COVID-19. El objetivo era generar informes interactivos en Power BI que combinan datos de casos y muertes con información de ventas y clientes para extraer conclusiones de negocio. A continuación, se describen los pasos realizados para desarrollar el proyecto.

1. Obtención y preparación de fuentes de datos

1.1 Descarga de datos de COVID-19

La fuente principal del proyecto es **Our World in Data (University of Oxford)**. Descargué el archivo desde el repositorio oficial (GitHub) y dejé anotada la fecha hasta la que llegan los datos para saber con qué versión estoy trabajando y evitar confusiones al comparar resultados

Fuente: <https://github.com/owid/covid-19-data/tree/master/public/data> (archivo usado: *cases_deaths.csv*).

1.2 Descarga de datos de ventas

Para la parte de negocio (ventas y clientes) utilicé el dataset '**Proyecto 3: Segmentación de Clientes en ecommerce**' descargado de **Kaggle**, que contiene ventas por factura (campos como Fecha de factura, País, ID Cliente, Monto, Cantidad, etc.). En mi caso trabajé con el archivo *ventas-por-factura.csv*.

Link: <https://www.kaggle.com/datasets/datalabatoria/proyecto-3-segmentacion-de-clientes-en-ecommerce>.



1.3 Carga de datos en Power BI

En Power BI Desktop se seleccionó Obtener datos → Texto/CSV y se importaron ambos ficheros. Se renombraron las tablas como COVID_OWID y Ventas_Facturas para facilitar su identificación.

2. Transformación y limpieza de los datos

2.1 Ajuste de tipos de datos

En el editor de Power Query se revisaron los tipos de cada columna:

- date y Fecha de factura se convirtieron en **Fecha**.

*En Ventas_Facturas, la columna **Fecha de factura** venía en formato Fecha/Hora, y en Power Query la transformé a **Fecha** para que el calendario y las medidas por periodo (mes, año, acumulados) funcionaran correctamente.*

- Las columnas numéricas (new_cases, total_cases, Monto, Cantidad) se convirtieron en **Número decimal**.
- Las columnas de texto (country, País, ID Cliente) se mantuvieron como **Texto**.

2.2 Filtro de filas no necesarias

Para evitar sumas duplicadas en las medidas DAX se eliminaron los registros que representaban agregados (continentes y grupos de ingresos): World, Asia, Europe, High-income countries, Upper-middle-income countries, Lower-middle-income countries, etc. Se mantuvieron sólo las filas correspondientes a países reales.

2.3 Creación de una tabla de calendario

Mediante DAX se generó una tabla Calendario con la función **CALENDARAUTO()** y se añadieron columnas derivadas:

- Año = **YEAR(Calendario[Fecha])**
- Mes = **FORMAT(Calendario[Fecha], "mmmm")**
- MesNum = **MONTH(Calendario[Fecha])**

El campo Mes se ordenó por MesNum para mantener el orden cronológico



2.4 Creación de relaciones

Se establecieron relaciones de muchos a uno entre:

- COVID_OWID[date] → Calendario[Fecha]
- Ventas_Facturas[Fecha de factura] → Calendario[Fecha]

Estas relaciones permiten filtrar por período todas las tablas a través de un único slicer de fechas.

3. Definición de medidas DAX

Se calcularon varias medidas para analizar los datos:

3.1 Medidas de COVID

- Casos nuevos = `SUM(COVID_OWID[new_cases])`
- Muertes nuevas = `SUM(COVID_OWID[new_deaths])`
- Casos totales = `MAX(COVID_OWID[total_cases])`
- Muertes totales = `MAX(COVID_OWID[total_deaths])`
- Variación casos % =
`DIVIDE(`
 `[Casos nuevos] - [Casos período anterior],`
 `[Casos período anterior],`
 `0`
`)`

3.2 Medidas de ventas

- Ventas totales = `SUM(Ventas_Facturas[Monto])`
- Ventas período anterior =
`CALCULATE(`
 `[Ventas totales],`
 `DATEADD(Calendario[Fecha], -1, MONTH)`
`)`
- Ventas acumuladas =
`CALCULATE (`
 `[Ventas totales],`
 `FILTER(`
 `ALL(Calendario[Fecha]),`
 `Calendario[Fecha] <= MAX (Calendario[Fecha])`
`)`
`)`



- Variación ventas % =
`DIVIDE(
 [Ventas totales] - [Ventas período anterior],
 [Ventas período anterior],
 0
)`
- Unidades vendidas = `SUM(Ventas_Facturas[Cantidad])`
- Clientes únicos = `DISTINCTCOUNT(Ventas_Facturas[ID Cliente])`
- Ticket medio =
`DIVIDE(
 [Ventas totales],
 [Unidades vendidas],
 0
)`

3.3 Medidas cruzadas

Para comparar ventas y casos sólo en los países presentes en las ventas se definió:

Casos nuevos (países ventas) =
`CALCULATE(
 [Casos nuevos],
 TREATAS(VALUES (Ventas_Facturas[País]), COVID_OWID[country])
)`

Y se añadieron métricas derivadas:

- Ratio Ventas/Casos = `DIVIDE([Ventas totales], [Casos nuevos (países ventas)], 0)`
- Ventas por 1000 casos = `DIVIDE([Ventas totales], [Casos nuevos (países ventas)], 0) * 1000`

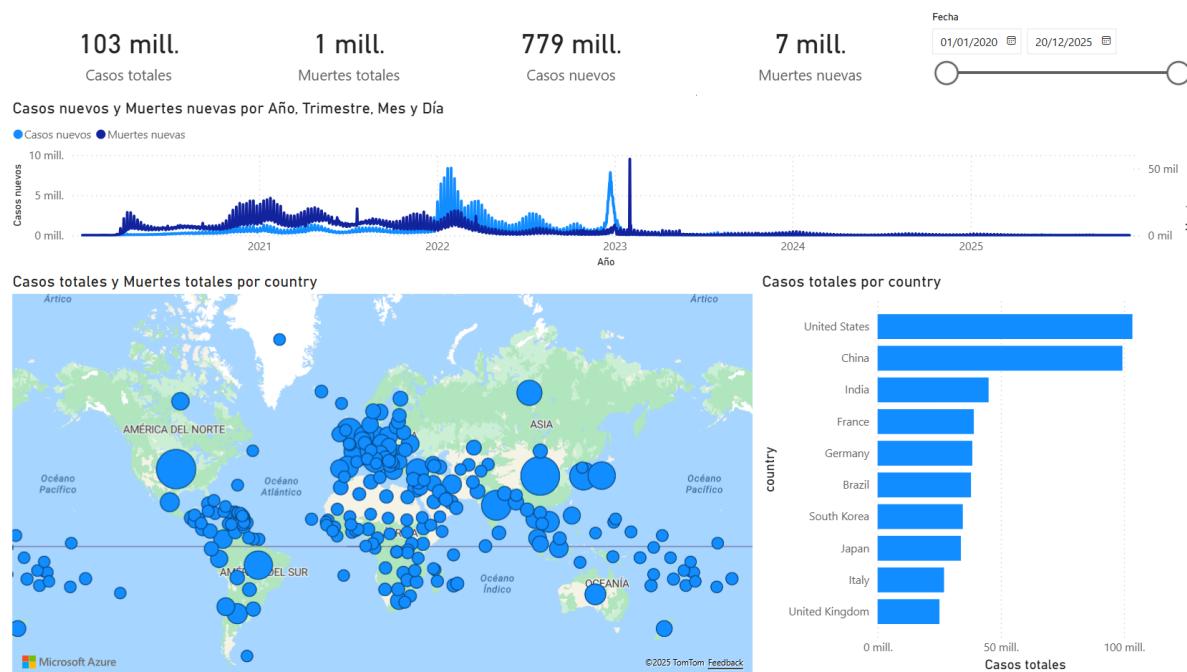


4. Diseño de informes en Power BI

Se crearon cuatro páginas principales:

4.1 Resumen mundial

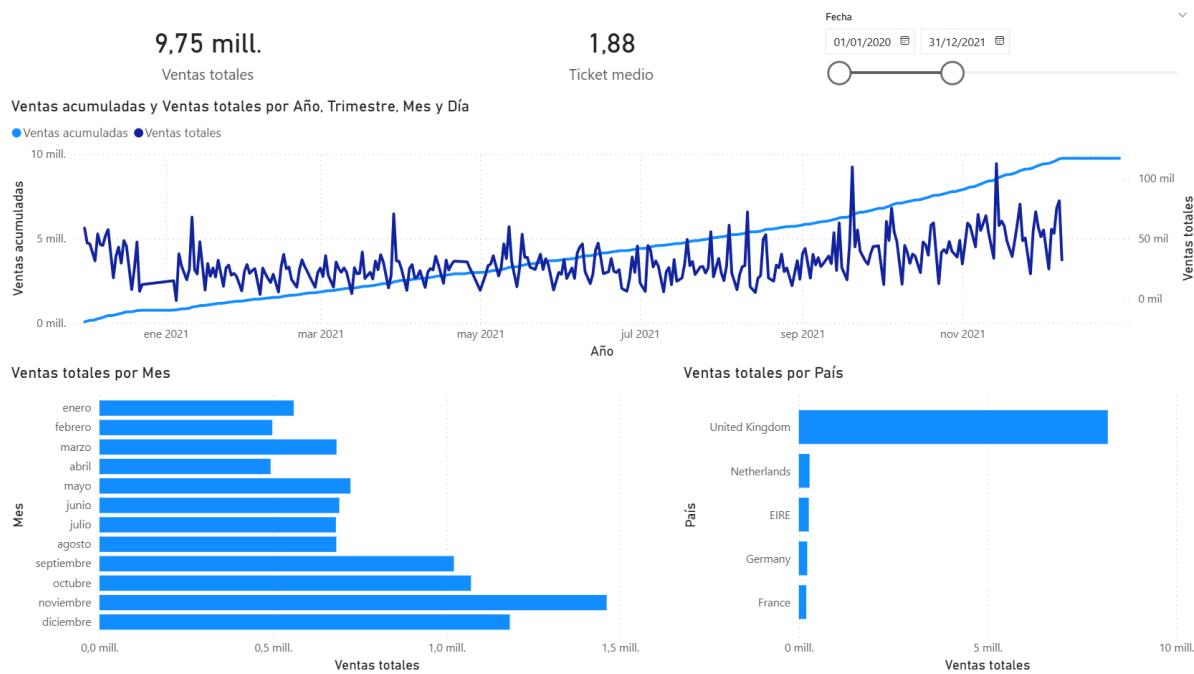
- Tarjetas con **Casos totales, Muertes totales, Casos nuevos y Muertes nuevas**.
- Gráfico de líneas combinadas para **Casos nuevos y Muertes nuevas** por fecha.
- Mapa con burbujas mostrando **Casos totales por país** y información sobre herramientas con muertes totales.
- Gráfico de barras Top 10 países por casos totales.





4.2 Análisis de Ventas

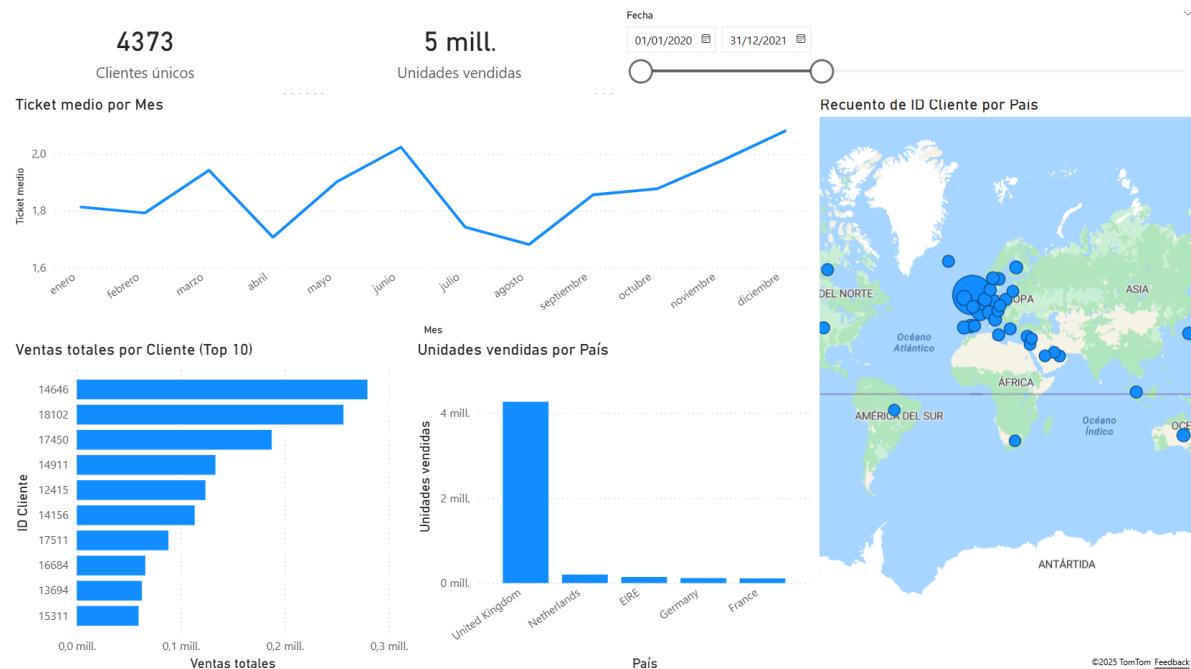
- Tarjetas con **Ventas totales**, **Ticket medio**.
- Gráfico de líneas para **Ventas totales y Ventas acumuladas**.
- Barras horizontales de **Ventas totales por Mes** y **Top países por ventas**.





4.3 Análisis de Clientes

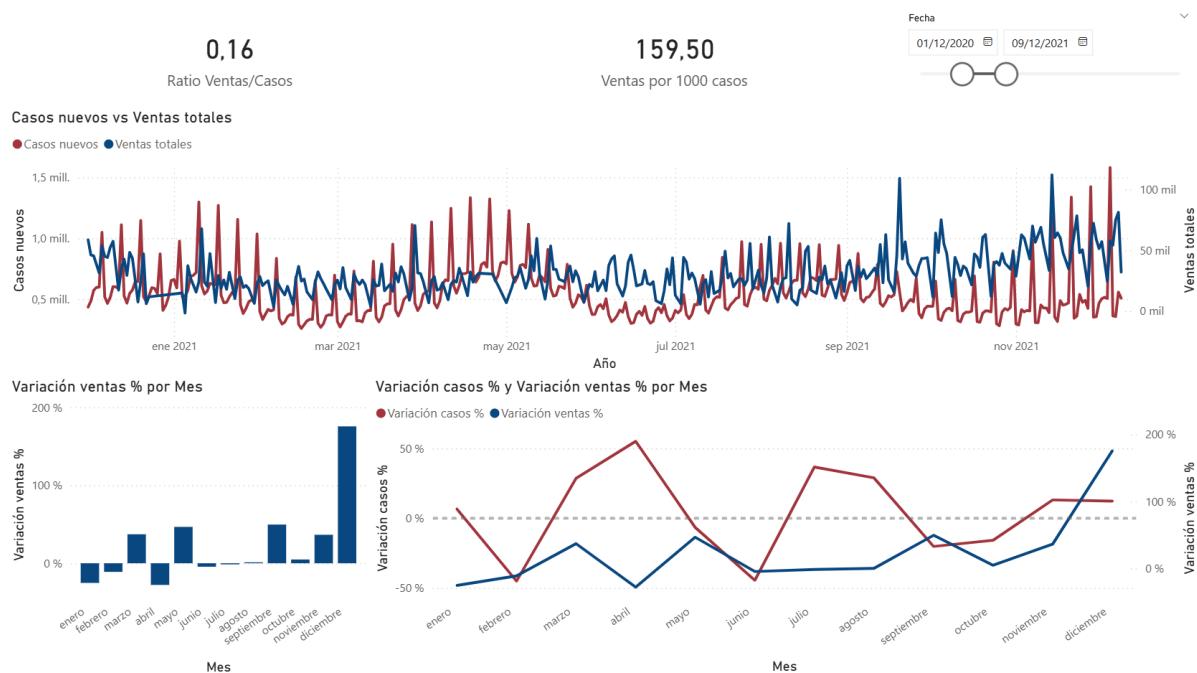
- Tarjetas con **Clients únicos** y **Unidades vendidas**.
- Línea de **Ticket medio por Mes**.
- Barras horizontales de **Ventas totales por cliente (Top 10)**.
- Barras horizontales de **Unidades vendidas por país**.
- Mapa de burbujas con recuento de clientes por país.





4.4 Relación COVID-Ventas

- Tarjetas con **Ratio Ventas/Casos** y **Ventas por 1000 casos**.
- Gráfico de líneas para **Casos nuevos vs Ventas totales**.
- Barras para **Variación ventas % por Mes**.
- Línea combinada para **Variación casos % y Variación ventas % por Mes**.



Se configuraron slicers de fechas en cada página para que el usuario pueda seleccionar intervalos temporales.