**Informe del Proyecto Integrador M0**

**Descubriendo el Éxito Comercial**

**Alumno: Christian Daniel Lara Larios**

El presente proyecto tiene como objetivo analizar los datos históricos de ventas de 45 tiendas pertenecientes a una empresa de retail, distribuidas en distintas regiones. A través de procesos de limpieza, transformación y exploración de datos, se busca obtener información clave que contribuya a una mejor toma de decisiones estratégicas, incluso frente a la limitación del historial disponible.

Para el análisis se trabajó con tres conjuntos de datos:

* Stores: Contiene información anónima de cada tienda, incluyendo su tipo y tamaño.
* Features: Incluye variables externas como temperatura regional, precio del combustible e indicadores económicos.
* Sales: Registra las ventas semanales por departamento en cada tienda, desde el 5 de febrero de 2010 hasta el 1 de noviembre de 2012.

**Carga y consolidación de los Datos**

La primera etapa del proyecto consistió en integrar la información contenida en tres datasets:

* stores.csv: información general de cada tienda (Store, Type, Size)
* sales.csv: registros históricos de ventas semanales por tienda y departamento.
* features.csv: datos adicionales por semana como clima, precios de combustible, markdowns, CPI, desempleo y feriados.

**Limpieza y normalización de los Datos**

Se aplicaron los siguientes procesos sobre el dataset stores.csv:

* Conversión de la columna Type a valores numéricos, mediante una hoja auxiliar llamada Type\_Num, en la que se definió:
  + A = 1
  + B = 2
  + C = 3
* Normalización de la columna Size, para poder comparar tamaños de tiendas de forma equitativa. Se creó una nueva columna Size\_Normalized, en donde cada valor fue escalado entre 0 y 1 utilizando la fórmula:

=(A2 - MIN($A$2:$A$46)) / (MAX($A$2:$A$46) - MIN($A$2:$A$46))

Esto permite comparar tamaños relativos independientemente de las unidades exactas.

Justificación

* Convertir Type a números permite agrupar y analizar de manera más eficiente
* Normalizar Size facilita la comparación entre tiendas grandes, medianas y pequeñas sin dejarse llevar por diferencias.

Se aplicaron los siguientes procesos sobre el dataset features.csv:

Agregar filtros a todas las columnas

* Se activó la función de filtro (Data -> Crear filtro) sobre la fila de encabezados, de modo que cada columna muestra su ícono de embudo para ordenar o filtrar sus valores

¿Por qué?

* + Facilita la exploración: Con filtros en cada columna, puedes rápidamente aislar un rango de fechas, ver ciertos valores en Store o identificar si hay celdas vacías
  + Detectar inconsistencias: Filtrar nos ayuda a encontrar huecos

Quitar decimales en la columna “Temperature” para mejorar lectura y limpieza

* + Simplificación de la tabla: Al quitar esas columnas, reducimos el ancho de la hoja y eliminamos datos que no vamos a utilizar para un análisis posterior

Eliminar decimales y aplicar formato de número en Fuel\_Price

* + Mayor claridad para el usuario final: Mostrar “2976” en lugar de “2976.00” hace que sea mas sencillo leer rápidamente tendencias generales en costos de combustible sin “ruido” decimal
  + Consistencia visual: Al aplicar formato “Número” sin decimales, la columna Fuel\_Price mantiene una alineación uniforme y ocupa menos espacio visual. Facilitando la comparación directa entre tiendas o fechas.

Se aplicaron los siguientes procesos sobre el dataset stores.csv

* Se habilitó la funcionde filtro en os encabezados de todas las columnas (Store, Type y Size).
* Se aplicó un formato numérico personalizado en la columna Size para que los valores aparezcan con separador de miles y dos decimales (por ejemplo, “153.15” en lugar de “1513.15”)

Esto mejora la legibilidad inmediata de los datos especialmente cuando las cifras exceden los tres dígitos. Facilita la comparación visual de tamaños entre tiendas, destacando de un vistazo cuáles superan el umbral de 1,000.

**Creación de la Base según el DER**

* Se verificó que cada tabla respeta sus PK y FK.
* **Stores:** columnas Store, Type\_Num, Size\_Normalized, Location.
* **Features:** Store, Date, Temperature, Fuel\_Price, CPI, Unemployment, IsHoliday.
* **Sales:** ID\_Sales, Store, ID\_Dept, Date, Weekly\_Sales.
* No hay registros de Sales sin su correspondiente Store en Stores.

**Insights**

1. Ventas Totales y Promedio Semanal

* Ventas totales: $4 448 332 867 (suma de “Weekly\_Sales” en toda la hoja “sales”).
* Semanas únicas: 95 semanas distintas registradas.
* Promedio semanal: $46 824 556

(ventas totales ÷ 95 semanas).

*Por qué:* Esto me permite ver cuánto vende el negocio en conjunto cada semana y valorar si una semana determinada estuvo por encima o por debajo del promedio.

2. Impacto de Semanas Festivas

* Promedio en festivos: $17 055
* Promedio en no festivos: $15 754
* Diferencia: +8 % en ventas cuando hay feriado.

*Por qué:* Comparé las semanas con “IsHoliday = TRUE” y “FALSE” para entender si los feriados realmente aumentan las ventas.

3. Departamentos que Más Venden

| ID\_Dept | Venta total (aprox.) | % sobre total |
| --- | --- | --- |
| 1 | $844 138 818 | 19 % |
| 3 | $717 767 236 | 16 % |
| 7 | $630 102 997 | 14 % |

* Los 5 primeros departamentos (1, 3, 7, 5, 8) concentran más del 75 % de las ventas.

*Por qué:* Al agrupar las ventas por “ID\_Dept” (con SUMAR.SI), veo qué categorías son las más importantes para enfocar inventario y promociones.

4. Tiendas con Mayor Aporte

| Store | Venta total (aprox.) | % sobre total |
| --- | --- | --- |
| 36 | $230 347 258 | 5.2 % |
| 40 | $225 001 464 | 5.1 % |
| 14 | $223 178 197 | 5.0 % |

* Las 5 tiendas líderes suman casi el 25 % de las ventas totales.

*Por qué:* Me ayuda a identificar qué locales están rindiendo mejor y cuáles necesitan apoyo (marketing, stock, etc.).

5. Tendencia Mensual

* Mes con más ventas: Diciembre 2011 con $288 078 102.
* Otros picos importantes: Diciembre 2012 ($275 322 890), Diciembre 2013 ($265 418 746).

*Por qué:* Extraje año/mes de la columna “Date” y usé una tabla dinámica para sumar “Weekly\_Sales” por mes. Confirma que la temporada de fin de año tiene el mayor volumen.

6. Crecimiento Semana a Semana (WoW)

* Calculé, solo donde la misma tienda y departamento aparecían con 7 días de diferencia, el cambio porcentual (venta\_actual − venta\_anterior) / venta\_anterior.
* Dept 7 tuvo el mayor crecimiento promedio: +7,5 %.
* Dept 3 siguió con +6,9 %.

**KPIs (números clave)**

1. Ventas Totales
   * Fórmula: =SUMA(sales!Weekly\_Sales)
   * Resultado: $4 448 332 867
2. Semanas Únicas
   * Fórmula: =CONTAR.UNICOS(sales!Date)
   * Resultado: 95
3. Promedio Semanal
   * Fórmula: =Ventas\_Totales / Semanas\_Únicas
   * Resultado: $46 824 556
4. Promedio Festivos vs No Festivos
   * Promedio Festivos   = PROMEDIO.SI(sales!IsHoliday; VERDADERO; sales!Weekly\_Sales) → $17 055
   * Promedio No Festivos = PROMEDIO.SI(sales!IsHoliday; FALSO;   sales!Weekly\_Sales) → $15 754
   * Boost Factor (%)    = (Promedio Festivos – Promedio No Festivos) / Promedio No Festivos \* 100 → 8 %

**Conclusión**

1. Depto 1 es el más fuerte (19 % del total), así que conviene tener suficiente stock y promos ahí.
2. Tienda 36 lidera en ventas (5,2 % del total). Sería bueno ver qué la hace especial y replicarlo en otras tiendas.
3. Diciembre es el mes con más ventas cada año (pilas para campaña navideña).
4. Feriados suben las ventas un 8 % en promedio, útil para planificar promos especiales.
5. WoW Growth mayor en los deptos 7 y 3, lo que indica que esas categorías están creciendo rápido y vale la pena monitorearlas de cerca.