



PROYECTO INTEGRADOR M1

Carrera: Data Analytics

Cohorte: Data Analytics Part-Time 09

Alumno: Hillebrand, Francisco Javier

12 de Junio de 2025

Índice

Limpieza de Datos.....	2
Paso a Paso de la Limpieza.....	2
Hoja Master.....	2
Columnas Eliminadas.....	4
Hoja Products.....	4
Descripción de Variables.....	5
Hoja Master.....	5
Hoja Products.....	6
Análisis Dinámico con Fórmulas y Tablas.....	7
Dimensiones Identificadas.....	7
Mencionadas por el Cliente.....	7
Extras.....	7
Métricas.....	7
Comportamiento del Negocio.....	7
¿El Negocio es rentable?.....	7
¿Cuál es la categoría más rentable?.....	8
Proporción de las Ventas para cada Segmento.....	8
¿Cuál es la Subcategoría más Rentable?.....	9
Rentabilidad de los Mercado en que vende sus productos.....	10
¿Qué método de envío es el más rentable?.....	10
¿Cuál es el año más rentable?.....	10
¿Cuál es la región más rentable?.....	11
Rentabilidad por País.....	13
¿Cómo impacta la prioridad de los pedidos en las ventas?.....	14
Dashboard Interactivo.....	16
Extra-Credit.....	17
Guía de uso del Dashboard.....	17
Funcionalidades relevantes a futuro.....	17

Limpieza de Datos

Paso a Paso de la Limpieza

Antes de adentrarnos en la transformación de cada una de las columnas de cada hoja, es necesario hacer dos aclaraciones generales:

- La configuración regional del archivo se cambió a la opción Argentina, y además se colocó la zona horario de Buenos Aires (GMT - 03:00).
- Para los números se utilizó la coma como separador decimal.
- Los datos originales, están en la hoja de cálculo llamada **DATOS ORIGINALES**, en la hoja **DATOS PROCESADOS**, utilice la fórmula **QUERY¹** con la fórmula **IMPORTRANGE** anidada para importar los datos sin filas vacías en la hoja “DATOSSINVACIOS”, y así poder mantener actualizada la información en caso de que se haga una modificación en los datos originales.
- En la hoja “MASTER” encontrará los datos de las ventas realizadas entre 01/01/2013 y 31/12/2014 (como lo pedía la consigna), para ello he usado la siguiente fórmula **=QUERY(DATOSSINVACIOS!A1:A; "SELECT A WHERE A IS NOT NULL AND A >= date'2013-01-01' AND A <= date'2014-12-31";1)**.² **En esta hoja se encuentran los datos que se utilizarán para el análisis.**

Hoja Master

Las transformaciones fueron las siguientes:

Columna	Formato	Aclaración
Sales Date	Fecha	
City	Texto Sin Formato	
Country	Texto Sin Formato	
Customer ID	Texto Sin Formato	
Customer Name	Texto Sin Formato	
Discount	Porcentaje	Cree una nueva columna nueva y en ella dividí los datos por 100 para poder expresarlos en porcentaje
Order Date	Fecha	Si bien sus datos son iguales a la columna Sales

¹ Ejemplo:
=QUERY(IMPORTRANGE("https://docs.google.com/spreadsheets/d/110TAAOOQd0ajki6ZjHwk9Aql1fUdAI49Px5FdxdXxro";"MASTER!C:C");"SELECT Col1 WHERE Col1 IS NOT NULL";0). Así traigo la columna C de la hoja MASTER del archivo “DATOS ORIGINALES” sin las filas vacías.

² Seguí controlando que no devuelva celdas vacía por si se ingresaban en la hoja “DATOSSINVACIOS” en un futuro

Columna	Formato	Aclaración
		Date ³ , decidí mantener la columna por si en el futuro se puede ordenar y confirmar la venta en días distintos.
Order ID	Texto Sin Formato	
Order Priority	Texto Sin Formato	
Product ID	Texto Sin Formato	
Total net profit	Moneda (Dólar Estadounidense)	
Quantity	Número	
Region	Texto Sin Formato	
Unit Cost	Moneda (Dólar Estadounidense)	
Segment	Texto Sin Formato	
Ship Date	Fecha	
Ship Mode	Texto Sin Formato	
Shipping Cost	Moneda (Dólar Estadounidense)	
State	Texto Sin Formato	
Year	Texto Sin Formato	Al ser solo el año, decidí no tratarlo como un número ya que no se realizarán operaciones matemáticas con él. No pude tratarlo como fecha ya que este formato toma el 2013 como un número y lo convierte en otra fecha con el formato dd/mm/aaaa.
Market	Texto Sin Formato	
weeknum	Texto Sin Formato	Al ser solo el número de la semana en el año en la cual se hizo la venta/orden, decidí no tratarlo como un número ya que no se realizarán operaciones

³ Lo controlé con la función =IF(A2=G2;"IGUALES";"NO IGUALES")

Columna	Formato	Aclaración
		matemáticas con él.

Columnas Eliminadas

Columna	Justificación
Columna1	Al no aparecer en el glosario que me otorgó el cliente, y tener como valores 0 o nulos, creo que no es una columna que aporte datos relevantes para el análisis.
Columna2	Al no aparecer en el glosario que me otorgó el cliente, y tener como valores 0 o nulos, creo que no es una columna que aporte datos relevantes para el análisis.
Category	<p>Este mismo dato del producto lo podemos encontrar en la hoja “PRODUCTS”, por ende es redundante que también esté en esta hoja.</p> <p>Si necesitamos el dato lo buscaremos con una fórmula de búsqueda como XLOOKUP, utilizando el ID del producto como valor de búsqueda.</p>
Sub-Category	<p>Este mismo dato del producto lo podemos encontrar en la hoja “PRODUCTS”, por ende es redundante que también esté en esta hoja.</p> <p>Si necesitamos el dato lo buscaremos con una fórmula de búsqueda como XLOOKUP, utilizando el ID del producto como valor de búsqueda.</p>
Customer	Es una concatenación de las columnas Customer ID y Customer Name, por lo que esta columna generaría datos duplicados.
Records	Esta columna no tiene datos relevantes para el análisis, ya que sus valores son vacíos o 1.

Hoja Products

Las transformaciones fueron las siguientes:

Columna	Formato	Aclaración
Product ID	Texto Sin Formato	Al ser un campo alfanumérico, decidí tratarlo como texto, ya que no se realizarán operaciones matemáticas sobre él
Product.Name	Texto Sin Formato	
Category	Texto Sin Formato	
Sub.Category	Texto Sin Formato	

Descripción de Variables

La descripción de las variables también la puede ver como notas en las respectivas columnas.

Hoja Master

- Sales Date: Fecha en que se realiza la venta. Solo tendrá fechas entre 01/01/2013 y 31/12/2014.
- City: Ciudad a la cual se envía el producto.
- Country: País al cual se envía el producto.
- Customer ID: ID alfanumérico y único de cada cliente.
- Customer Name: Nombre del Cliente.
- Discount: Descuento aplicado al cliente. Se resaltó en rojo aquellos descuentos mayores o iguales al 100%, ya que estos datos me parecen erróneos, sin embargo, no los eliminé ya que necesitaría más información de parte del cliente.
- Order Date: Fecha en la que el cliente realizó la orden del producto comprado.
- Order ID: ID alfanumérico y único de cada orden realizada.
- Priority: Prioridad de la orden, puede ser: Critical, High, Low y Medium.
- Product ID: ID alfanumérico y único de cada producto.
- Total net profit: Total de ingresos netos por cada venta.
- Quantity: Cantidad del producto vendido en cada orden.
- Region: Región del mundo a la cual se envió el producto.
- Unit Cost: Costo unitario del producto.
- Segment: Segmento al cual pertenece cada cliente.
- Ship Date: Fecha en que se envió el producto.
- Ship Mode: Modo de envío de la orden.
- Shipping Cost: Costo de envío de la orden.
- State: Estado en el que se encuentra la ciudad a la cual fue enviada la orden.
- Year: Año de la venta/orden.
- Market: Segmento del mercado al cual se le vendió el producto.
- weeknum: Número de la semana del año en que se realizó la venta/orden.

Hoja Products

- Product ID: ID alfanumérico y único de cada producto.
- Product.Name: Nombre del producto, junto con una/s característica/s.
- Category: Categoría a la que pertenece el producto.
- Sub.Category: Subcategoría a la que pertenece el producto. Cada subcategoría pertenece a una sola categoría.

Análisis Dinámico con Fórmulas y Tablas

Para este análisis dinámico, se calcularon las métricas pedidas por el cliente y además se identificaron 3 dimensiones más con respecto a las que se mencionaron en la lista de requerimientos del cliente.

Todas las dimensiones especificadas se verán en la hoja “DIMENSIONES”, con los datos en esta hoja se han realizado las tablas dinámicas necesarias para calcular las métricas.

Dimensiones Identificadas

Mencionadas por el Cliente

- Categoría del Producto
- Subcategoría del producto
- Región
- Mercado
- Fecha
- Método de Envío
- Prioridad
- Segmento

Extras

- País

Métricas

En el presente análisis, se han calculado las siguientes métricas para cada dimensión.

- Total de Ventas: Es igual a la columna Total net profit, la cual tenía los ingresos netos por cada venta.
- Total de Costos⁴: En este caso se han tenido en cuenta los costos unitarios, la cantidad del producto vendida y el costo de envío (decidí tenerlo en cuenta suponiendo que la empresa paga el envío y luego lo recupera cuando el cliente paga su pedido). En la hoja de cálculo verá una nota con el cálculo realizado $((cantidadProducto * costoUnitarioProducto) + costoEnvioOrden)$.
- Rentabilidad: $\frac{\text{Ingresos Netos}}{\text{Total de Costos}} * 100$.

Comportamiento del Negocio

¿El Negocio es rentable?

Viendo las métricas calculadas en las tarjetas, podemos decir que la empresa fue rentable en el periodo 2013/2014, de hecho la rentabilidad fue casi del 50%.

⁴ No se ha tenido en cuenta el descuento ya que presentaba datos atípicos, para los cuales necesitaría más información de parte del cliente para obtener los datos correctos. No me pareció correcto tener en cuenta el descuento para algunas ventas y para otras no.

Total Vendido

\$149.107.205,00

Costos

\$306.938.796,00

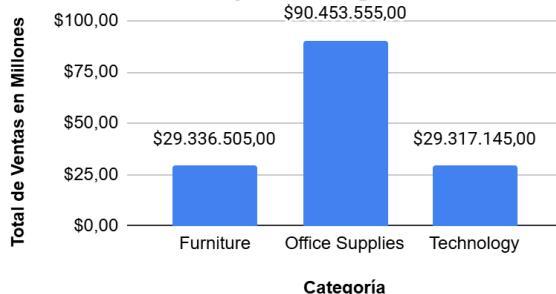
Rentabilidad

48,58%

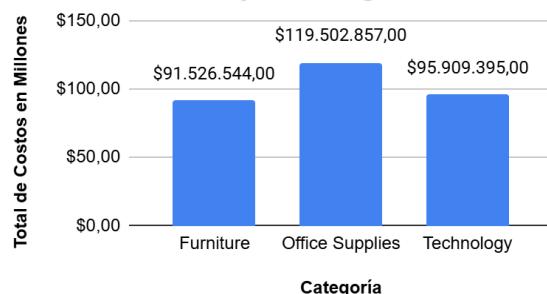
¿Cuál es la categoría más rentable?

Vemos que la categoría más rentable es con diferencia “Office Supplies”, ya que su rentabilidad alcanza el 75%. Esto justifica los costos realizados en esta categoría, ya que es la categoría en la que más gasta la empresa.

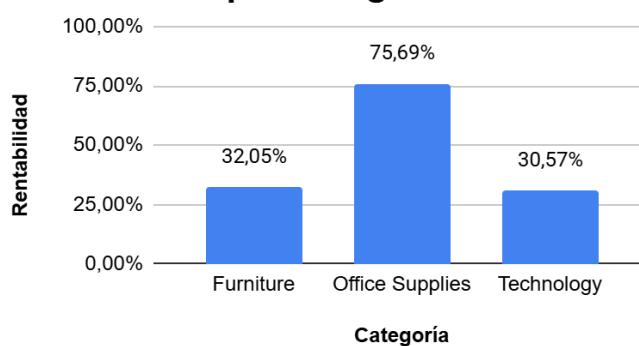
Total de Ventas por Categoría



Total de Costos por Categoría



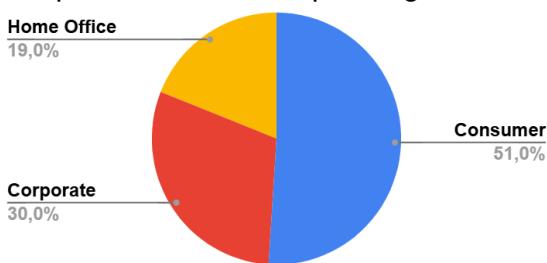
Rentabilidad por Categoría



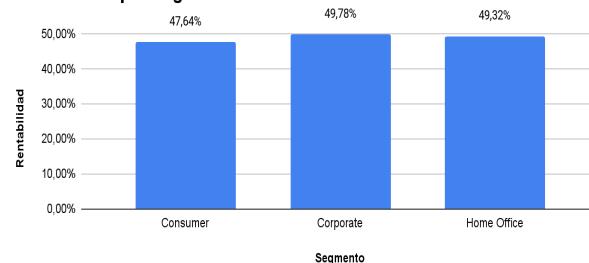
Proporción de las Ventas para cada Segmento

En este caso sucede algo muy interesante, si bien el segmento que les genera mayores ganancias son los consumidores finales, los segmentos, Corporativo y Home Office, son más rentables, aunque la diferencia es mínima. De hecho, la rentabilidad de los segmentos es similar

Proporción de Ventas por Segmentos



Rentabilidad por Segmento



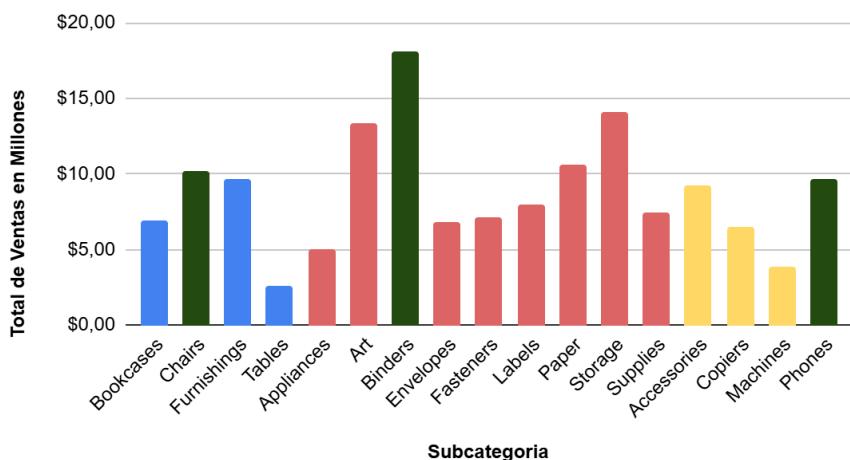
¿Cuál es la Subcategoría más Rentable?

Como vemos en el gráfico la subcategoría “Binders” es la que más vende, sin embargo, no es la más rentable, ya que es “Labels” la más rentable. Igualmente estos resultados confirman que la categoría “Office Supplies” es la categoría que más vende y la que mayor rendimiento genera.

En tecnología son los celulares los que generan mayores ingresos, pero los accesorios son los más rentables.

Entre los muebles son las sillas las que más ingresos netos generan, pero la subcategoría “Furnishings” es la más rentable.

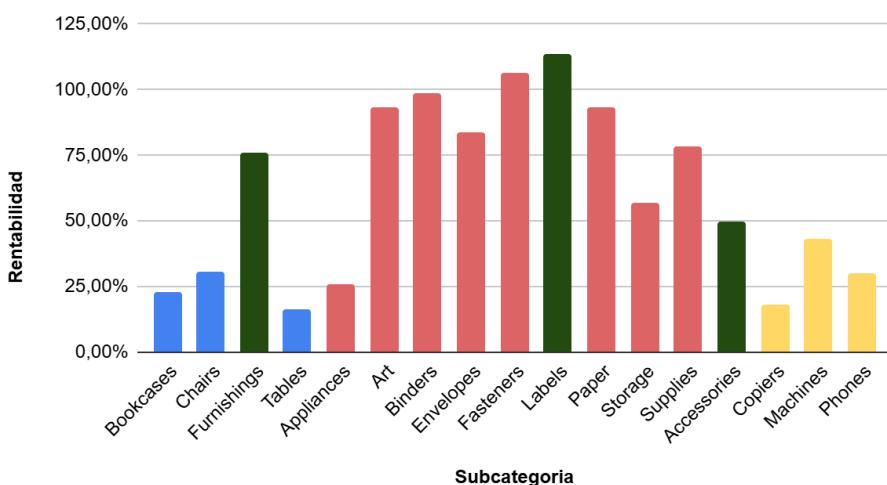
Total de Ventas por Subcategoría



Leyenda del gráfico:

- Verde: Subcategorías con el mayor valor de ventas dentro de cada categoría.
- Azul: Subcategorías que pertenecen a la categoría “Furniture”.
- Rojo: Subcategorías que pertenecen a la categoría “Office Supplies”.
- Amarillo: Subcategorías que pertenecen a la categoría “Technology”.

Rentabilidad por Subcategoría



Leyenda del gráfico:

- Verde: Subcategoría con el mayor rendimiento dentro de cada categoría.
- Azul: Subcategorías que pertenecen a la categoría "Furniture".
- Rojo: Subcategorías que pertenecen a la categoría "Office Supplies".
- Amarillo: Subcategorías que pertenecen a la categoría "Technology".

Rentabilidad de los Mercado en que vende sus productos

El Mercado Asía Pacífico es el que más ingresos le genera a la empresa, sin embargo, EMEA (Europa, Medio Oriente y África) es el mercado más rentable, siendo además el segundo mercado que menos costos le genera a la empresa, detrás del mercado africano. Es importante destacar además que Latinoamérica es el mercado menos rentable con una amplia diferencia, mientras que los demás mercados manejan rentabilidades parecidas.

Market	Total de Ventas	Costos	Rentabilidad
Africa	\$9.154.260,00	\$11.785.928,00	77,67%
APAC	\$34.672.152,00	\$46.695.582,00	74,25%
EMEA	\$9.900.059,00	\$12.193.872,00	81,19%
EU	\$31.524.248,00	\$40.474.143,00	77,89%
LATAM	\$32.605.169,00	\$154.464.067,00	21,11%
North America	\$31.251.317,00	\$41.325.204,00	75,62%

¿Qué método de envío es el más rentable?

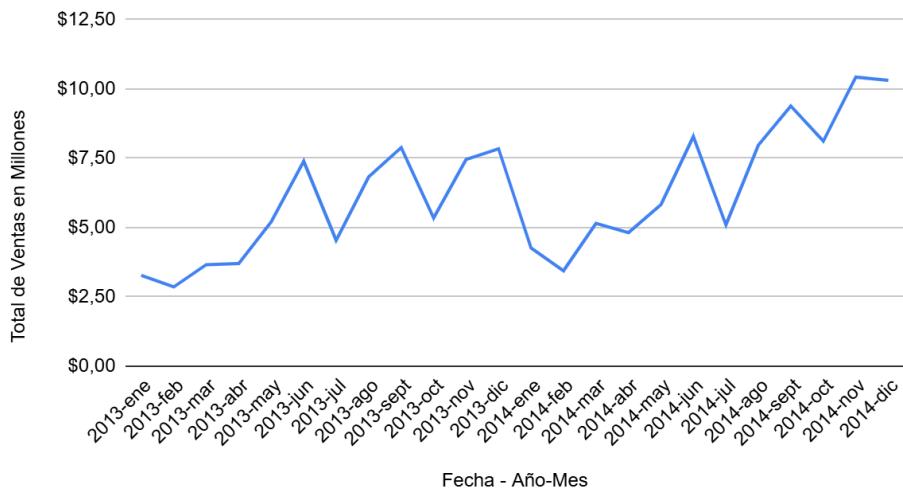
En este caso, el método más usado, el que más ingresos netos genera y además el de mejor rentabilidad.

Ship Mode	Nº de veces que se utilizó	Total de Ventas	Costos	Rentabilidad
First Class	4561	\$21.686.312,00	\$63.574.300,00	34,11%
Same Day	1673	\$7.821.255,00	\$21.979.950,00	35,58%
Second Class	6403	\$30.393.861,00	\$70.466.230,00	43,13%
Standard Class	18693	\$89.205.777,00	\$150.918.316,00	59,11%

¿Cuál es el año más rentable?

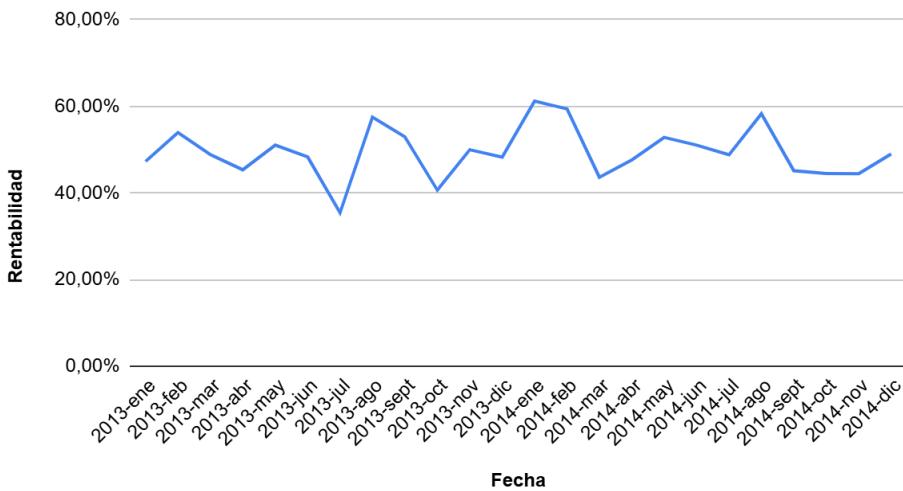
Con respecto a los ingresos totales, noto una baja en los primeros meses del año en comparación a los últimos meses del año. Al parecer se vende más sobre el final del año.

Ventas en el período 2013/2014



Sin embargo, la rentabilidad tuvo un aumento en los primeros meses del año 2014, al igual que en agosto y septiembre del año 2013 y 2014.

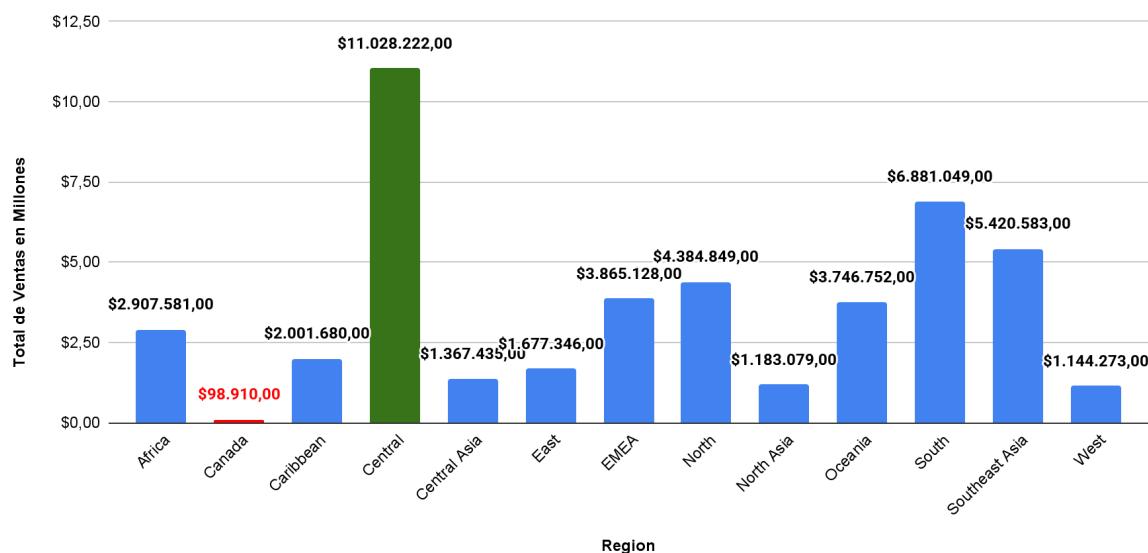
Rentabilidad en el periodo 2013/2014



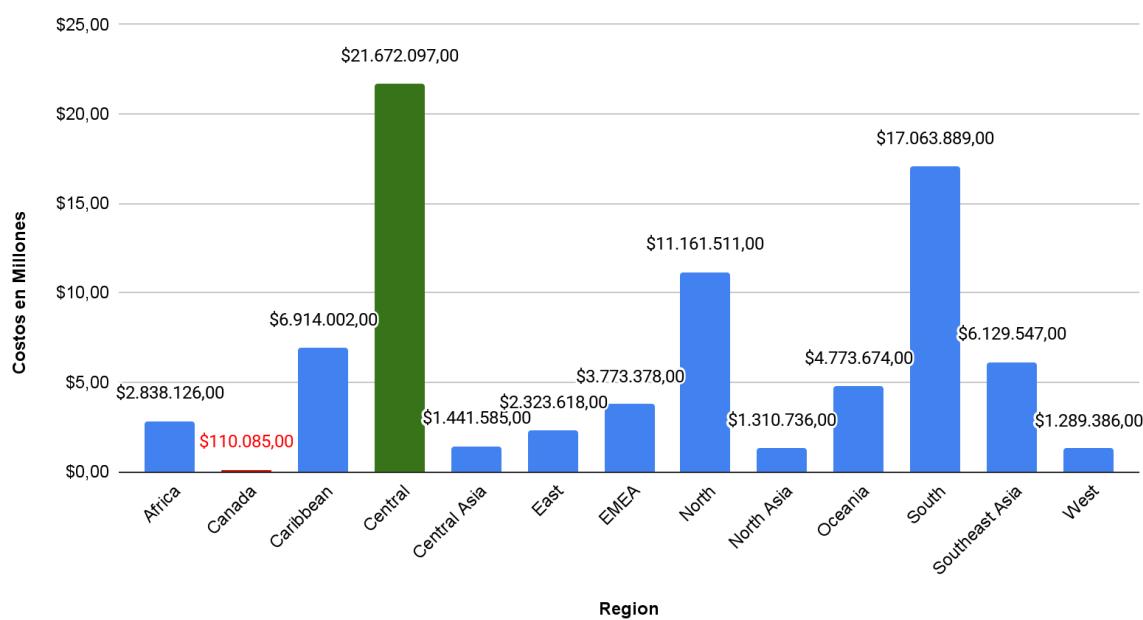
¿Cuál es la región más rentable?

La región Central es aquella que más ganancias genera, lo cual justifica la inversión realizada en ella, ya que es la que más costos le genera a Master Global. También es importante destacar que la región Canadiense es a la que menos se venden los productos de la empresa.

Total de Ventas por Región



Costos por Región

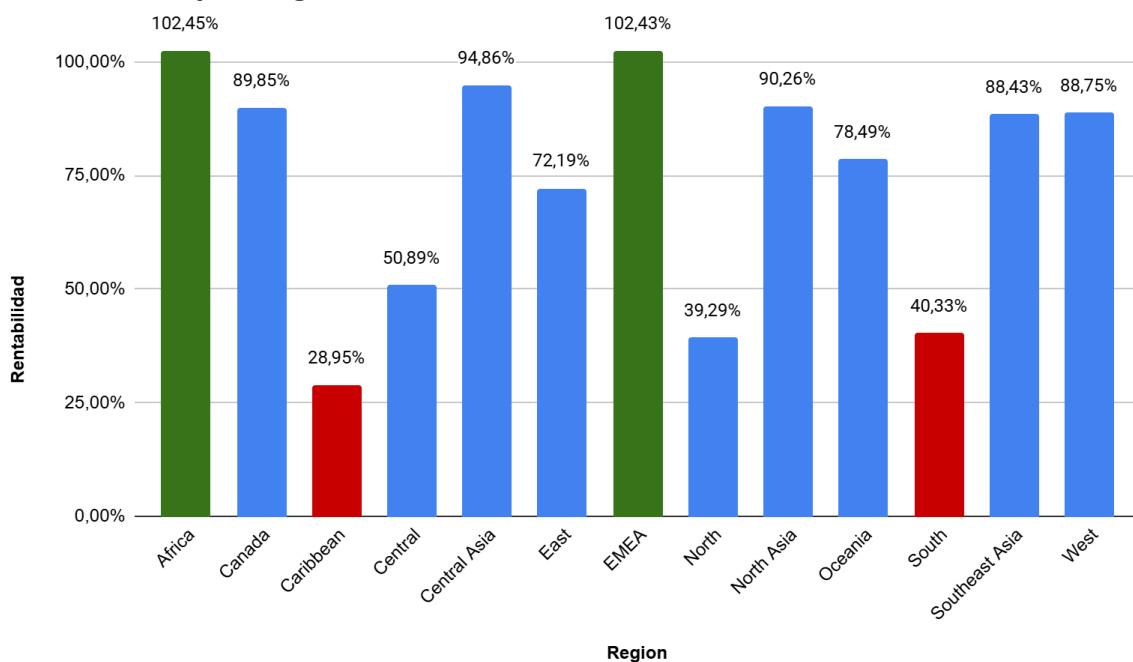


Con respecto a la rentabilidad vemos que el EMEA se destaca nuevamente, junto con la región africana, región la cual también pertenece al EMEA. A mi parecer, esto demuestra que el continente africano es bastante rentable para Global Master.

Además vemos que la región sur⁵ y Caribe, son de las menos rentables, esto ayuda a explicar por qué el mercado LATAM es también de los menos rentables.

⁵ Asumo que se refiere al sur del mundo

Rentabilidad por Región

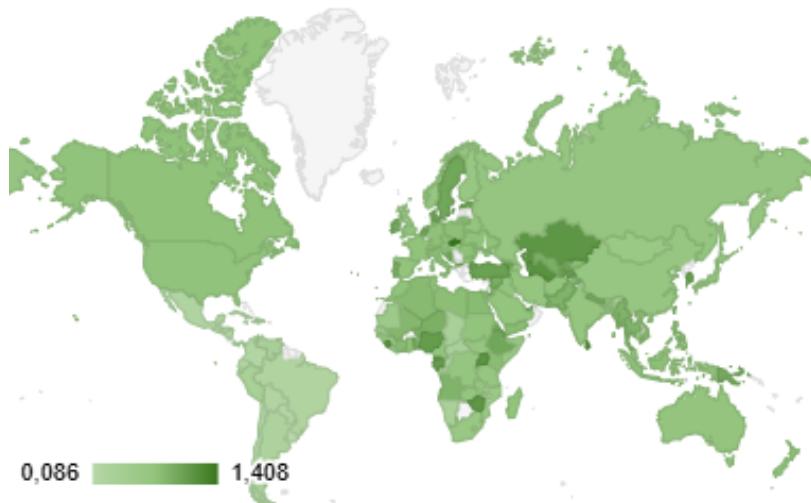


Rentabilidad por País

Si analizamos la rentabilidad por país, vemos que nuevamente en África se encuentran algunos de los países más rentables para la empresa. El mercado africano es muy rentable para Global Store.

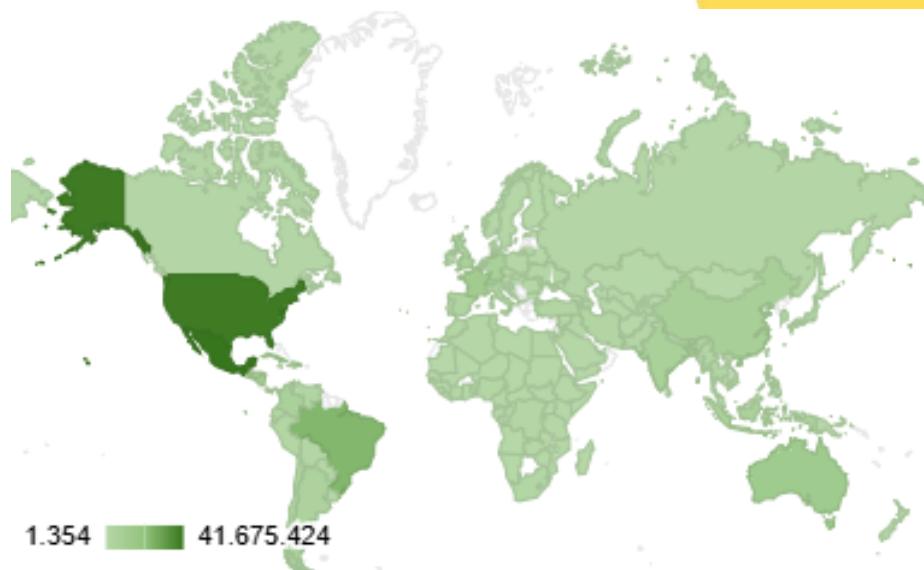
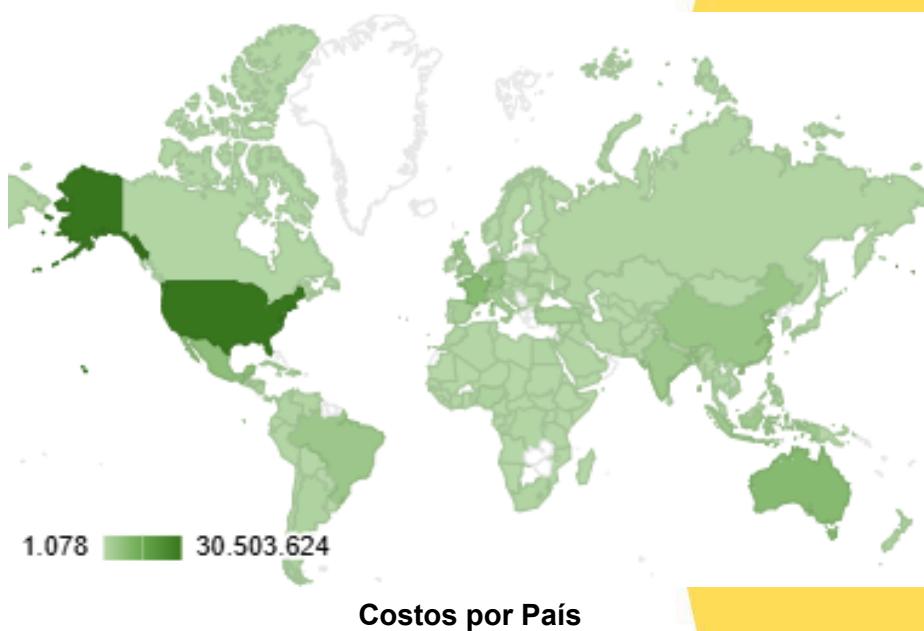
También se vuelve a destacar el EMEA, ya que hay países de Medio Oriente (Turquía con más del 100%) y Europa (Eslovaquia con un poco más de 140%) con alta rentabilidad.

Además, este gráfico también confirma la baja rentabilidad de todos los países de Latinoamérica.



Igualmente, el país que más ingresos y costos genera es por diferencia Estados Unidos, acompañado por México en cuanto a costos generados.

Ventas Totales por País



¿Cómo impacta la prioridad de los pedidos en las ventas?

Lo primero que se puede resaltar, es que los pedidos con mayor prioridad (Crítica), son los que menos rentabilidad le generan a la empresa, quizás se le da la prioridad incorrecta a ciertos pedidos. Creo que sería mejor darle más prioridad a los envíos más rentables.

Justamente, los pedidos más rentables son aquellos con una prioridad media, incluso son los que más ingresos generan.

Recomendaría a la empresa reevar el proceso que le da prioridad a los distintos pedidos, ya que aquellos con menor prioridad, son los más rentables.

Order Priority	Ventas Totales	Costos	Rentabilidad
Critical	\$11.247.871,00	\$43.873.707,00	25,64%
High	\$44.556.965,00	\$105.864.320,00	42,09%
Low	\$7.228.453,00	\$15.909.468,00	45,43%

Order Priority	Ventas Totales	Costos	Rentabilidad
Medium	\$86.073.916,00	\$141.291.301,00	60,92%

Dashboard Interactivo

El Dashboard lo puede ver en la hoja de cálculo “ DATOS PROCESADOS”, se presenta de la siguiente manera:

Un Dashboard General, con un resumen de los resultados, desde el cual puede acceder a los demás tableros de:

- Ventas
- Costos
- Rentabilidad

Extra-Credit

Guía de uso del Dashboard

En el archivo  GUÍA DASHBOARD/MOCKUP encontrará un mockup de cómo se desarrollaron los dashboard, lo cual pretende ser una guía para el uso del dashboard.

Funcionalidades relevantes a futuro

- En la hoja “PRODUCTS”, en la columna “Product.Name”, encontramos el nombre del producto junto con algunas características. Si bien en este análisis no fue relevante diferenciar el nombre del producto y sus características, a futuro, para tener un mejor modelo de datos, sería importante dividir en dos columnas diferentes al nombre del producto y su característica.
- Esclarecer el formato con el cual se debe trabajar en la columna de Descuentos, de esta manera se agregarían a los costos, lo cuál puede desencadenar nuevos insights.
- Se pudo observar que en la región EMEA y en el continente africano, la empresa está obteniendo buenas rentabilidades, sería importante analizar en detalle la información de las ventas en estas zonas para replicarlas en otras regiones, como en latinoamérica donde las rentabilidades eran más bajas.