



## **BASE DE DATOS AVANZADAS**

Proyecto 2P

### **Integrantes:**

Darinka Townsend

Eddo Alvarado

Danny Loor

### **Docente:**

Shirley Coloma

PAO 2023 – 1T

Tabla de contenido

**REPOSITORIO.....3**

**DEFINICIÓN DEL PROBLEMA A RESOLVER .....3**

**DISEÑO .....3**

**ARQUITECTURA .....5**

**PROCESO DE CARGA DE DATOS.....6**

**DISEÑO DEL SISTEMA .....7**

**RESULTADOS OBTENIDOS .....9**

**CONCLUSIONES .....12**

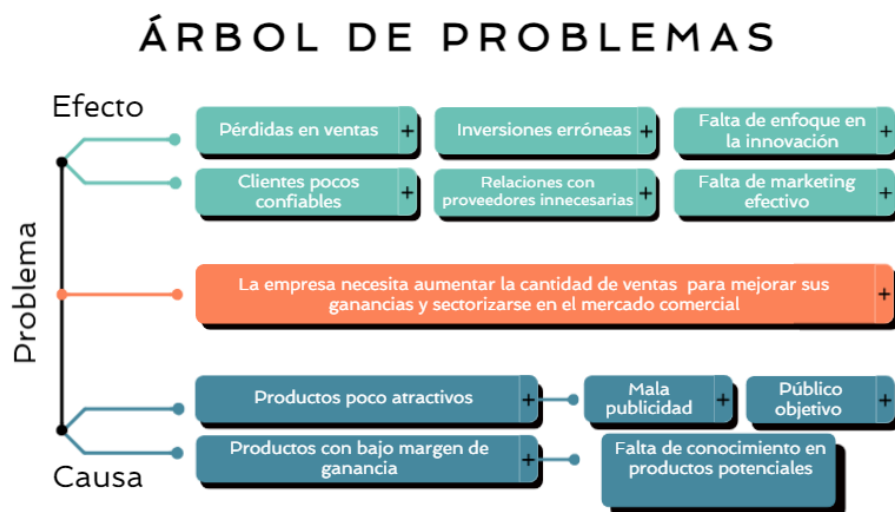
**RECOMENDACIONES.....12**

## Repositorio

<https://github.com/DarinkaTownsend/DataMart-AdventureWorks2019.git>

## Definición del problema a resolver

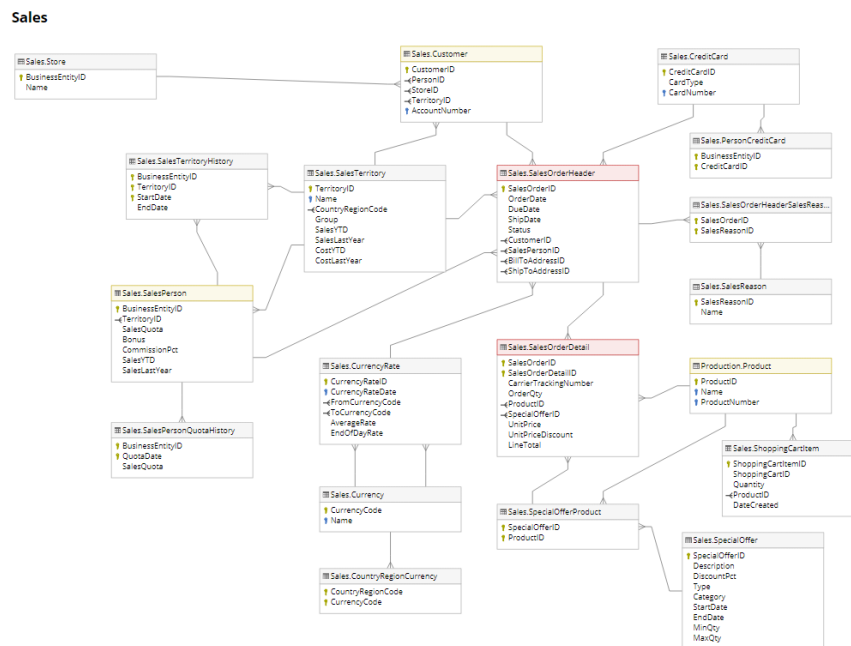
La empresa Adventure atraviesa una etapa en donde se ha observado una disminución significativa de ventas por internet. Aunque han logrado atraer tráfico a su sitio web, la cantidad de ventas realizadas es mucho menor de lo que esperaban. Esto está afectando negativamente sus ingresos y su capacidad para establecerse en el mercado del comercio electrónico. El equipo está buscando entender las razones detrás de estas pocas ventas en línea y desarrollar estrategias efectivas para aumentar la conversión de visitantes en compradores y revitalizar su presencia en el mercado digital.



*Ilustración 1 Árbol de Causa y Efecto*

## Diseño

Se ha recibido una copia de respaldo de la base de datos correspondiente a Adventure Works del año 2019. Dicha base de datos engloba información integral de la empresa, incluyendo datos sobre ventas, productos, personal, clientes, proveedores, entre otros aspectos relevantes. La decisión tomada consiste en enfocar el análisis en el área de ventas con el propósito de evaluar si se han alcanzado los objetivos de ventas establecidos para el año en cuestión. Además, se pretende utilizar esta información para realizar proyecciones y establecer nuevas metas comerciales para los años venideros. Este enfoque se lleva a cabo con el fin de optimizar el desempeño y la planificación estratégica de Adventure Works en el ámbito de sus operaciones comerciales.



*Ilustración 2: dataware House Original*

Para el datamart definimos que serán necesarias las dimensiones de Ventas, los detalles de la venta, el cliente y el tiempo de la venta, para la creación del datamart se deben seguir los siguientes pasos:

1. En la tabla SalesOrderHeader, se extrajeron los campos SalesOrderId, OrderDate, Status, OnlineOrderFlag, TerritoryId, TotalDue y se desnormalizo la tabla Sales.Territory, para obtener el nombre y continente.
2. En la tabla SalerOrderDetails, se extrajeron SalesOrderID, SalesDetailID, OrderQty, ProductId, UnitPrice y se desnormalizo la tabla Producto para obtener el productName, StandardCost.
3. En la tabla Customer, se extrajeron las columnas CustomerID, StoreID, y se desnormalizo la tabla Person para obtener el PersonType, EmailPromotion, Title.
4. Se creo una nueva dimensión Time, en base al OrdenDate, separando el año y los meses.
5. Se crearon los campos de la tabla hechos, entre los cuales tenemos los que la conectan con las dimensiones los cálculos para dar respuesta a las preguntas.
  - a. Cantidad de ventas por territorio
  - b. Cantidad de ventas por genero
  - c. Cantidad de ventas por año
  - d. Cantidad de ventas por mes

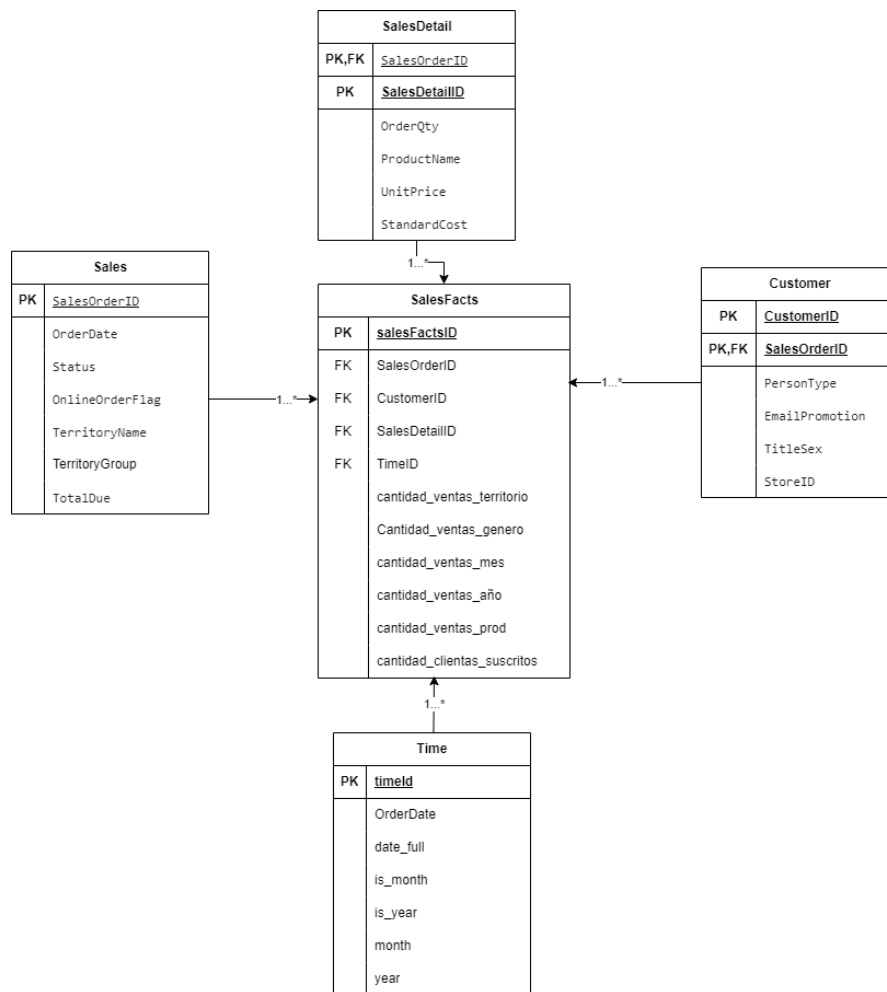


Ilustración 3: DataMart despues

## Arquitectura

La arquitectura utilizada para el proyecto se centra en la restauración de la base de datos de AdverntureWorks con el archivo de respaldo del 2019, para este proceso se alojó dicho archivo de respaldo en un contenedor en los servicios de AWS S3 para proceder con la restauración en los servicios de AWS RDS, ya que proveen herramientas fáciles de utilizar y dando un acceso rápido y eficiente a la base en cualquier lugar al estar publicado en la nube.

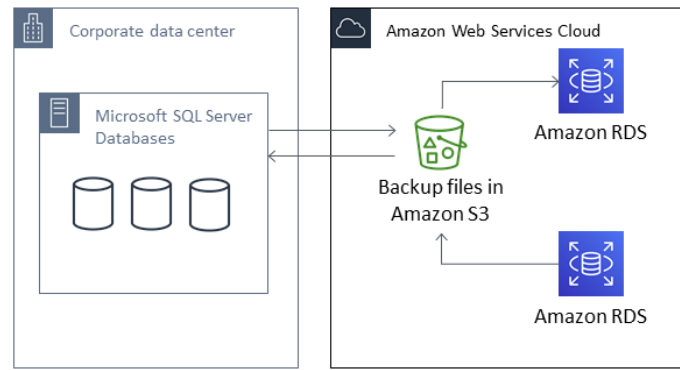


Ilustración 4: Arquitectura de restauración

## Proceso de carga de datos

La arquitectura permite que AWS Glue DataBrew acceda a Amazon RDS de manera segura. DataBrew extrae, transforma y exporta datos entre ellos, facilitando la preparación de datos para análisis y aplicaciones empresariales. La comunicación es segura y eficiente.

Con AWS Glue DataBrew se crearon las distintas dimensiones con las recetas mencionadas anteriormente en la Ilustración 5: Dimensión Sales, con los pasos necesarios para llegar a esa vista.

Ilustración 5: Dimensión Sales

Una vez ya creadas todas las dimensiones se crean Jobs, que se encargan de sacar la información y exportarla a archivos .csv para ser consultados por nuestros sistemas de análisis estadísticos. Dentro de la interfaz se realiza la conexión por medio del bucket.

```
# Descarga el .csv de la tabla de hechos de AWS

bucket = 'adventuretransformado'
object = 'FACT-AdventureWorks_part00000.csv'
save_as = 'FACT-AdventureWorks.csv'
download_s3_object(bucket, object, save_as)
```

Ilustración 6 Conexión AWS

Pasos de la receta (56)
1. Eliminar columna RevisionNumber, DueDate, ShipDate, Status, PurchaseOrderNumber, AccountNumber, BillToAddressID, ShipToAddressID, ShipMethodID, CreditCardID, CreditCardApprovalCode, CurrencyRateID, SubTotal, TaxAmt, Freight, Comment, rowguid, ModifiedDate
2. Unión interna AdventureWorks-Details
3. Unión interna AdventureWorks-Product
4. Unión interna AdventureWorks-Territory
5. Eliminar columna OnlineOrderFlag
6. Eliminar columna TABLE_B_SalesOrderID
7. Cambiar nombre Name a ProductName
8. Eliminar columna TABLE_B_ProductID
9. Eliminar columna TABLE_B_TerritoryID
10. Cambiar nombre TABLE_B_Name a TerritoryName
11. Cambiar nombre CountryRegionCode a Country
12. Cambiar nombre Group a Continent
13. Eliminar columna SalesPersonID
14. Crear columna Month uso de Función dateTime MONTH
15. Crear columna Año uso de Función dateTime YEAR
16. Unión izquierda AdventureWorks-Customer
17. Ordenar valores por PersonID por Ascendente
18. Ordenar valores por PersonID por Ascendente
19. Eliminar columna TABLE_B_TerritoryID
20. Eliminar columna TABLE_B_CustomerID, rowguid, ModifiedDate, AccountNumber
21. Ordenar valores por PersonID por Ascendente
22. Ordenar valores por PersonID por Ascendente
23. Unión izquierda Person-person
24. Unión izquierda AdventureWorks-Stores
25. Mover OrderDate a inicio de la tabla
26. Mover StandardCost a inicio de la tabla
27. Eliminar columna TABLE_B_BusinessEntityID
28. Mover SalesOrderID a Índice 0
29. Mover CustomerID a Índice 1
30. Mover SalesOrderDetailID a Índice 2
31. Mover OrderDate a Índice 3
32. Eliminar columna SalesOrderNumber
33. Mover Año a Índice 4
34. Mover Month a Índice 5
35. Eliminar columna TerritoryID, ProductID, PersonID, StoreID, BusinessEntityID
36. Cambiar nombre Name a PersonName
37. Crear columna MULTIPLY Columna 1 uso de función matemática MULTIPLY
38. Cambiar nombre MULTIPLY Columna 1 a Total_Fabric_Cost
39. Crear columna SUBTRACT Columna 1 uso de función matemática SUBTRACT
40. Cambiar nombre SUBTRACT Columna 1 a Total_Earns
41. Rellenar valores faltantes con no registrado en PersonType
42. Rellenar valores faltantes con no registrado en Title
43. Cambiar nombre Title a Gender
44. Rellenar valores faltantes con 0 en EmailPromotion
45. Rellenar valores faltantes con no registrado en PersonName
46. Ordenar valores por Año por Descendente
47. Cambiar formato de OrderDate a dd'month/yyyy
48. Eliminar columna Month
49. Crear columna OrderDate_extracted uso de Función dateTime MONTH
50. Eliminar columna OrderDate_extracted
51. Crear columna Month uso de Función dateTime MONTH_NAME
52. Crear columna EmailPromotion2 uso de Funciones lógicas CASE
53. Eliminar columna EmailPromotion
54. Cambiar nombre EmailPromotion2 a EmailPromotion
55. Crear columna Gender2 uso de Funciones lógicas CASE
56. Eliminar columna Gender

Ilustración 7 Transformaciones realizadas en AWS

## Diseño del sistema

Se desarrollo el sistema usando como lenguaje base Python y las librerías de streamlit para facilitar la creación de la interfaz. Se usan los filtros para determinar que grafico se va a mostrar, conforme se va llenando los filtros.

Los filtros seleccionados son:

- Territorio
- País

- Continente
- Sexo
- Año
- Mes
- Aceptación de Mails
- Nombre del Producto

Choose your filter:

Selecciona el territorio

Southwest xGermany xCentral x

Selecciona el país

US xCA x

Selecciona el continente

North America xPacific x

Selecciona el genero

F xM x

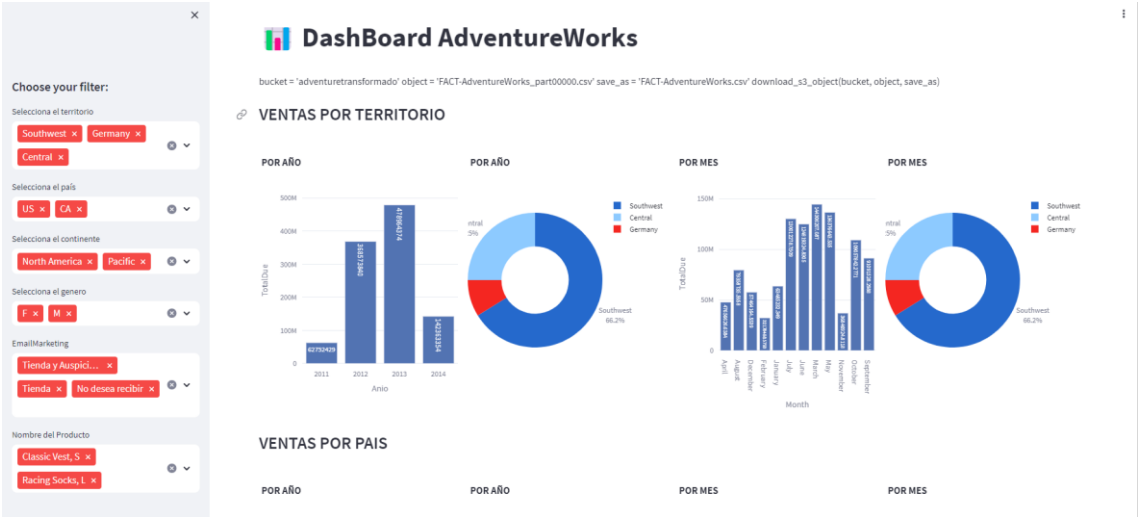
EmailMarketing

Tienda y Auspici... xTienda xNo desea recibir x

Nombre del Producto

Classic Vest, S xRacing Socks, L x

Ilustración 8 Filtros Aplicados





# Resultados obtenidos

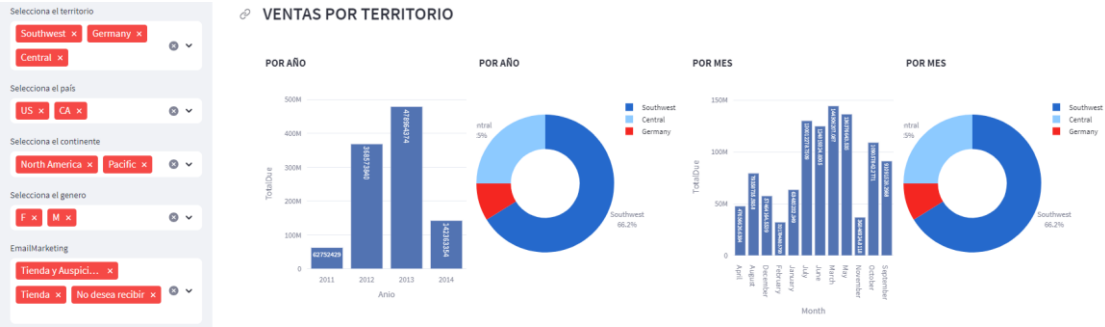


Ilustración 10 Resultados por filtro Territorio

Se puede observar que el año que mayores ingresos se obtuvieron fue el 2013 y el territorio que más ingresos se obtuvo fue Soutwest con un margen de 23.8%, y los meses con mayores ventas son los meses de marzo a julio, por lo que se debería tomar la decisión de dar mayor prioridad a esos meses de baja demanda.

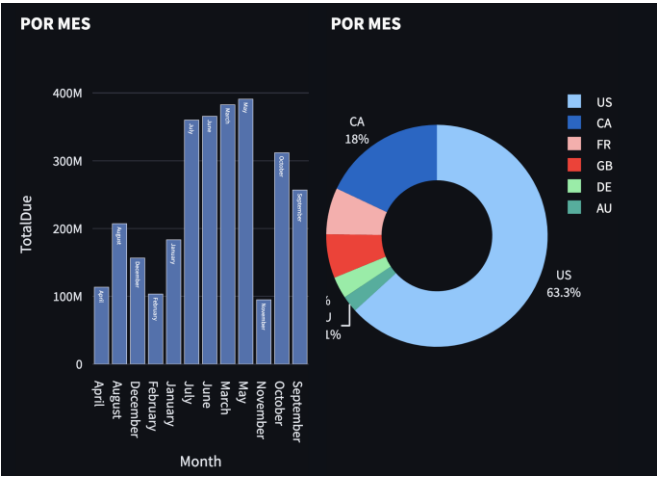


Ilustración 11 Ventas por mes

Se analizan las ventas de la empresa por mes a lo largo de su historia. El mes en donde la empresa tiene el peor rendimiento es en noviembre. Este mes puede ser aprovechado para crear ofertas que puedan ser aprovechadas por los clientes de tal forma que las ventas no se vean afectadas, se espera que no sea alcanzado el valor de ventas del mes de diciembre, pero durante los días de “Black Friday” puede ser interesante una rebaja temporal de precios, haciendo énfasis sobre todo en el país de US.

Por otro lado, si analizamos el país que tiene la menor cantidad de ventas nos damos cuenta que Australia se lleva el puesto, compitiendo con Estados Unidos no es posible realizar una comparación directa debido a la cantidad de habitantes que ambos países tienen.



Ilustración 12 Ventas por Continente

Se puede ver que el continente con mayor presencia tanto por año y mes es Norte América, con un 81,3% dando un gran margen del mercado. Analizando esta información se podría dar una campaña de marketing enfocada más a los continentes con menor participación sin perder el margen de Norte América.

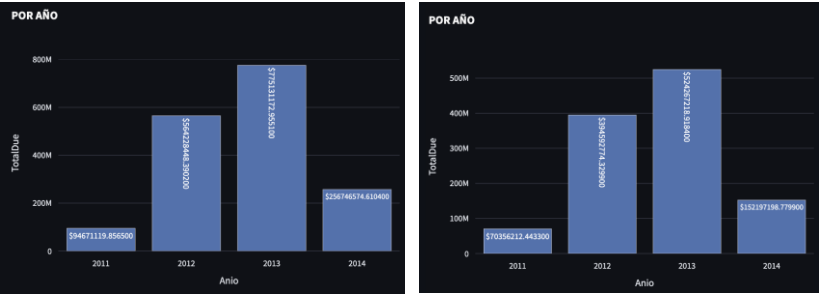
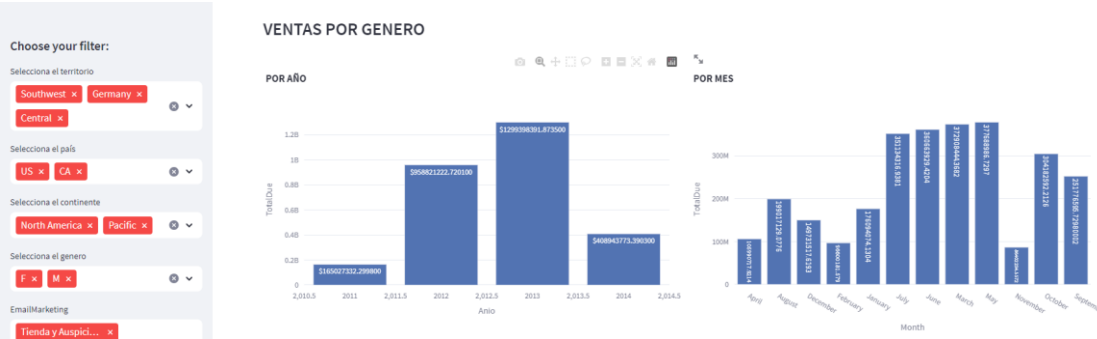


Ilustración 13 Venta por género

A nivel global se reporta un mayor nivel de venta de los productos destinados al público masculino, sobre todo en 2013 cuando las ventas de categoría masculina superaron por más de \$200M a las de categoría femenina. El departamento de marketing podría tomar esta información

para estudiar el mercado y proponer campañas que promuevan la compra de los productos por parte de las mujeres.

EMAIL MARKETING

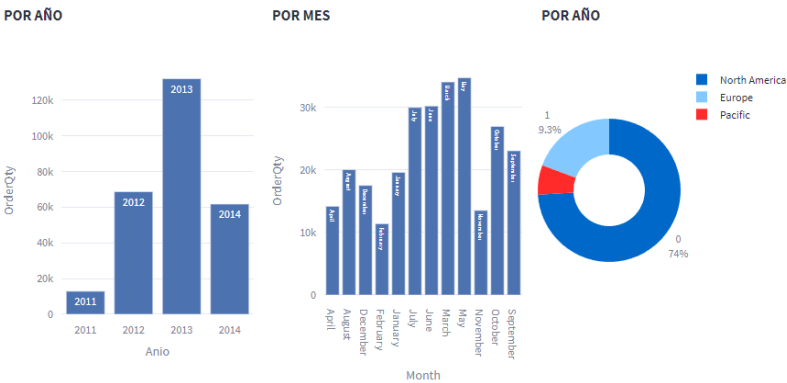


Ilustración 14 Clientes que desean EmailMarketing

Analizando la presencia de marketing que tenemos podemos ver que existe una alta cantidad de clientes que no tienen marcado los emails marketing por lo que se debería hacer un reimpulso para atraer más clientes y fidelizarlos en la empresa.



Ilustración 15 Cantidad de productos vendidos

Se puede observar que, entre los productos seleccionados, el año que más venta ha existido es en el 2013. Mientras que el mes que más venta genera es marzo. Por lo tanto, se pueden aplicar estrategias para empezar a hacer publicidad en esos meses e incrementar la venta de la empresa.

## Conclusiones

- Mediate la creación del datamart se analizaron las ventas de la empresa de una manera más profunda, se identificaron patrones y tendencias que podrían estar contribuyendo a los problemas que tiene Adventure con su disminución de ventas por internet.
- El uso de AWS S3 y RDS para la base de datos permitió un acceso rápido y eficiente a la data de ventas de la compañía, de esta forma se consiguió obtener los datos en tiempo real.
- La información obtenida del datamart generado durante este proyecto brinda a la empresa una base sólida para tomar decisiones sobre todo informadas. Esto permite que sean identificadas áreas de mejora y diseñar estrategias para abordar los problemas reportados en las ventas.

## Recomendaciones

- Las tendencias pueden cambiar conforme avanza el tiempo, por esta razón es importante mantener los datos a consumir actualizados con la información más reciente, es decir se requiere de un constante monitoreo y actualización del datamart.
- Las necesidades de la empresa cambiarán a medida que la empresa progrese, por tanto, es necesario explorar la posibilidad de agregar más dimensiones y métricas al datamart, de esta forma podrá ser alcanzando una mayor comprensión sobre el área de las ventas y sus puntos de mejora.
- A medida que la cantidad de datos aumente en el tiempo será necesario asegurar que el datamart pueda soportar este volumen de información, por tanto, se deben evaluar diferentes pruebas de escalabilidad para mantener un rendimiento adecuado en todo momento.