

PROYECTO FINAL | SUPER MERCADO



Universidad Galileo	Guatemala, 27 de marzo de 2,023
Instituto en Investigación de Operaciones	Diseño y Construcción de Data Warehouse
Edgar Eduardo Mazariegos Godoy	23000966
Rodrigo Aragon Rodriguez	23008159
Josué Daniel King Calderón	23000712
José Ricardo Mérida López	23006385
Fabián Antonio Hernández Ventura	18003299
Elí Javier Medrano Yat	23006806
Profesor: Jose Rolando Lucero Morataya	Asistente de Cátedra: Héctor Armando Asturias

Índice

1	CASO DE ESTUDIO	2
2	DISEÑO DIMENSIONAL	4
2.1	Selección del Proceso de Negocio	4
2.2	Identificación de la Granularidad	5
2.3	Identificación Slow Changing Dimensions	6
2.3.1	DIM_CUSTOMER (Cliente)	6
2.3.2	DIM_PRODUCT (Producto)	6
2.3.3	DIM_DATE (Fecha)	6
2.4	Identificación de Dimensiones	8
2.5	Identificación de tabla de Hechos	14
3	EXPLORACIÓN DE DATOS	19
3.1	¿Cuáles son los meses de mayor venta por cada región?	19

3.2	¿Cuál es el porcentaje de venta por región del país?	19
3.3	¿Qué categorías y subcategorías de productos presentan las mayores ganancias para la compañía?	20
3.4	¿Qué tipo de envío presenta mayores ventas durante los fines de semana, en un periodo del 2016 al 2018?	21
3.5	La agrupación de la exploración	22
4	CONCLUSIONES	23

1. CASO DE ESTUDIO

Un supermercado con múltiples sucursales en diferentes ubicaciones, ha estado recopilando datos de ventas de sus productos en cada sucursal a través de su sistema de punto de venta. Sin embargo, la empresa no ha estado aprovechando estos datos de manera efectiva y no ha estado obteniendo información valiosa para la toma de decisiones. La empresa ha decidido implementar un data warehouse para analizar estos datos y obtener información útil para mejorar la rentabilidad y la eficiencia de sus operaciones.

La empresa ha establecido los siguientes requerimientos para el desarrollo del data warehouse:

1. El data warehouse debe integrar la información de ventas de todas las sucursales de la empresa.
2. La información debe estar actualizada diariamente para permitir un análisis en tiempo real.
3. La información debe ser fácilmente accesible y presentada en un formato intuitivo y fácil de entender.
4. La información debe ser detallada por categoría de productos, por sucursal y por fecha.
5. Se deben desarrollar paneles de control personalizados para los gerentes de cada sucursal, permitiéndoles acceder a información específica sobre sus ventas y operaciones.

El proceso de desarrollo del data warehouse para el supermercado se llevó a cabo en varias etapas:

1. Análisis de requisitos: En esta etapa, se analizan los datos para identificar la granularidad e identificar y describir las dimensiones del modelo para el data

warehouse. También se identificaron los datos necesarios para la integración en el data warehouse, incluyendo información sobre productos, clientes, ventas, regiones tipos de envío y transacciones.

2. Diseño: En esta etapa, se diseñó la arquitectura del data warehouse, incluyendo la definición de las tablas de datos y la identificación de los sistemas necesarios para la integración de los datos. También se establecieron los procesos de extracción, transformación y carga (ETL) necesarios para garantizar que la información se actualizara diariamente.
3. Implementación: En esta etapa, se desarrolló el data warehouse y se integraron los datos de ventas de todas las sucursales de la empresa. También se desarrollaron los paneles de control personalizados para los gerentes de cada sucursal.
4. Pruebas: En esta etapa, se llevaron a cabo pruebas para asegurarse de que la información integrada en el data warehouse fuera precisa y estuviera actualizada diariamente.

2. DISEÑO DIMENSIONAL

2.1. Selección del Proceso de Negocio

Las ventas diarias son la fuente principal de datos para crear el data warehouse. Para creación del data warehouse del supermercado utilizando las ventas diarias, se siguió el proceso de diseño dimensional.

1. Se identifican las dimensiones que se utilizarán para analizar los datos de ventas..
2. Se identifican las medidas que se utilizarán para analizar los datos de ventas.
3. Se crea una estructura de tablas para almacenar los datos de ventas. Esta estructura incluirá una tabla de hechos, que contendrá las medidas de las ventas diarias, y una serie de tablas dimensionales, que contendrán los valores únicos de cada una de las dimensiones identificadas.
4. Se extraen los datos de ventas diarias de los documentos dados para la elaboración de proyecto y se transforman para adaptarlos a la estructura de tablas del data warehouse. Durante esta fase de transformación, se pueden realizar tareas como la limpieza de datos, la agregación de datos y la eliminación de duplicados.
5. Se cargan los datos transformados en las tablas correspondientes del data warehouse. Este proceso puede ser realizado por medio de la creación de scripts de carga, que extraen datos de la fuente, los transforman y cargan en el data warehouse. También es posible utilizar herramientas ETL (Extracción, Transformación y Carga) especializadas para automatizar este proceso.

Una vez que se ha creado el data warehouse, se pueden realizar consultas y análisis para obtener información valiosa sobre las ventas diarias del supermercado. Por ejemplo,

pueden analizar las ventas por producto, por fecha, por sucursal, por cliente y por tipos de envío para identificar tendencias y oportunidades de mejora. También se pueden realizar análisis de rentabilidad y pronósticos de ventas para tomar decisiones más informadas en el futuro.

2.2. Identificación de la Granularidad

Las ventas diarias son la mejor granularidad para la construcción de un data warehouse de un supermercado por varias razones:

Representan la unidad de medida más detallada: Las ventas diarias proporcionan una granularidad fina que permite un análisis detallado de los patrones de compra de los clientes y la performance de las sucursales. Al tener información de las ventas diarias, se pueden identificar tendencias a lo largo del tiempo y comparar los resultados de cada día para detectar patrones de consumo.

Son relevantes para la toma de decisiones: Las ventas diarias son una medida clave en la toma de decisiones de negocios, ya que permiten a los gerentes y analistas comprender cómo se está desempeñando el negocio día a día. La información de ventas diarias es esencial para evaluar el éxito de una promoción o para determinar si una estrategia de marketing está funcionando o no.

Permiten un análisis detallado a nivel de sucursal: Las ventas diarias se pueden analizar por sucursal, lo que permite comparar el rendimiento de diferentes sucursales y detectar posibles problemas. Además, esto permite identificar cuáles son los productos más vendidos en cada sucursal y hacer una gestión de inventario más eficiente.

Son una fuente confiable de información: Las ventas diarias son una fuente de información confiable y precisa, ya que los datos se registran en la base de datos transaccional en línea en tiempo real. Esto asegura que la información sea coherente y precisa, lo que es esencial para tomar decisiones informadas.

En resumen, la granularidad de ventas diarias es la mejor opción para la construc-

ción del data warehouse del supermercado porque proporciona una granularidad fina, es relevante para la toma de decisiones, permite un análisis detallado a nivel de sucursal, y es una fuente confiable de información.

2.3. Identificación Slow Changing Dimensions

2.3.1. DIM_CUSTOMER (Cliente)

Esta dimensión representa los clientes del supermercado y puede sufrir cambios con el tiempo como, por ejemplo, el cambio del segmento de un cliente. Se sugiere tratar esta dimensión como una SCD (Slowly Changing Dimension) del Tipo 2 en donde se mantenga un registro/historial de esos cambios. Esto permitiría analizar si el aumento de ventas de un segmento en específico viene dado por un alcance mayor a nuevos clientes de ese segmento o si se ve influenciado por el cambio de clientes de un segmento a otro. Al tener un historial, es posible determinar qué porcentaje representa los clientes que han cambiado de segmento con respecto a las ventas de un segmento en específico. Además, permite identificar patrones de compra y comportamiento del cliente a lo largo del tiempo.

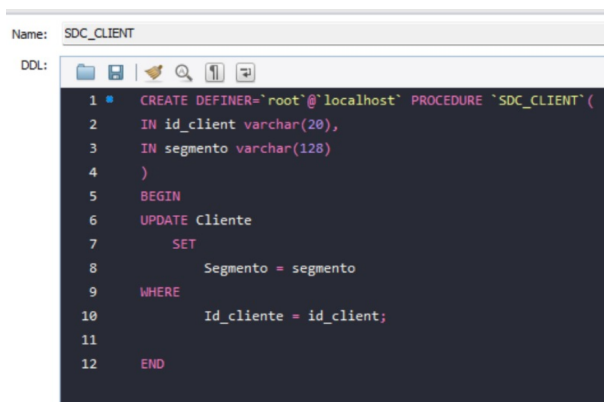
2.3.2. DIM_PRODUCT (Producto)

Esta dimensión representa los productos que ofrece el supermercado y puede sufrir cambios en su nombre, categoría o subcategoría. Considerando que estos cambios en un producto no suelen ser comunes y que normalmente son discutidos por el negocio previo a su cambio, se sugiere tratar esta dimensión como una SCD del tipo 1, lo que implica sobre escribir la información existente.

2.3.3. DIM_DATE (Fecha)

Esta dimensión representa las fechas de las transacciones de compra y puede sufrir cambios en caso de ser necesario actualizar información como la fecha de shipping o cuan-

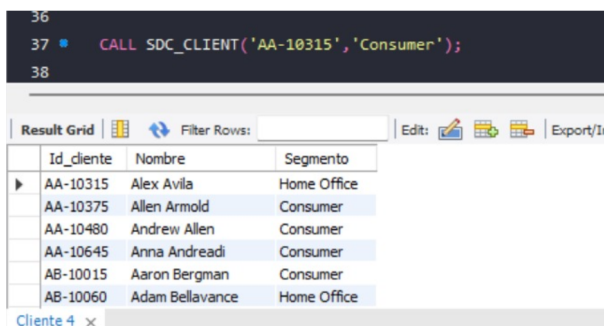
do existe una devolución y es necesario crear una columna para registrar esa devolución. Se sugiere tratar esta dimensión como una SCD del tipo 3 en donde se creen columnas para registrar estas posibles acciones y mantener el registro de lo que sucede con la compra a lo largo del tiempo.



```
DDL:
1 CREATE DEFINER='root'@'localhost' PROCEDURE `SDC_CLIENT` (
2   IN id_client varchar(20),
3   IN segmento varchar(128)
4 )
5 BEGIN
6   UPDATE Cliente
7     SET
8       Segmento = segmento
9   WHERE
10     Id_cliente = id_client;
11 END
12
```

Figura 1: Fuente: *El autor*.

En la figura anterior se muestra como crea un procedimiento para implementar una slow changing dimensión, en este caso aplica para el atributo de segmento de la tabla dimensional de cliente, esto dado que observamos que es un campo que puede variar con el tiempo, dado que es una categoría que no influirá a mayor escala en los análisis de forma histórica se implemento un tipo 1 donde se sobrescribe el registro mediante el comando de update, en este caso en el procedimiento recibimos la llave(ID del cliente) y el valor del segmento al que se desea cambiar.



```
36
37 CALL SDC_CLIENT('AA-10315', 'Consumer');
38
```

Id_cliente	Nombre	Segmento
AA-10315	Alex Avila	Home Office
AA-10375	Allen Arnold	Consumer
AA-10480	Andrew Allen	Consumer
AA-10645	Anna Andreadi	Consumer
AB-10015	Aaron Bergman	Consumer
AB-10060	Adam Bellavance	Home Office

Cliente 4 x

Figura 2: Fuente: *El autor*.

En la 2 imagen se observa como se va a llamar al procedimiento de SDC_CLIENT

mandando como argumentos el ID del primer cliente y el segmento al que se desea cambiar que en este caso seria de Home Office a Consumer.

```

30 CREATE TABLE Cliente (
31     Id_cliente varchar(20) unique NOT NULL,
32     Nombre varchar(128) NOT NULL,
33     Segmento varchar(128) NOT NULL,
34     CONSTRAINT Id_clientePK PRIMARY KEY (Id_cliente)
35 );
36
37 CALL SDC_CLIENT('AA-10315','Consumer');
38

```

Id_cliente	Nombre	Segmento
AA-10315	Alex Avila	Consumer
AA-10375	Allen Arnold	Consumer
AA-10480	Andrew Allen	Consumer
AA-10645	Anna Andreadi	Consumer
AB-10015	Aaron Bergman	Consumer
AB-10060	Adam Bellavance	Home Office

Output:

#	Time	Action
9	13:26:50	CALL SDC_CLIENT('AA-10315','Consumer')
10	13:26:56	SELECT * FROM Cliente LIMIT 0, 1000

Figura 3: Fuente: *El autor*.

Finalmente se observa como se ha ejecutado el procedimiento de forma exitosa y también como es que ha cambiado el atributo de segmento para Alex Avila siendo ahora Consumer

2.4. Identificación de Dimensiones

El uso de las tablas dimensionales en un data warehouse de ventas de un supermercado es fundamental para proporcionar un contexto y una comprensión más profunda de los datos de ventas. En este caso en particular, las siguientes tablas dimensionales seran útiles para la construcción de un data warehouse de ventas de un supermercado:

Tabla dimensional de sucursal (Región): Esta tabla contiene información del nombre de la sucursal. Esta dimensión permitiría a los analistas de datos desglosar las ventas por sucursal, lo que puede ser útil para evaluar la rentabilidad de cada una de ellas.

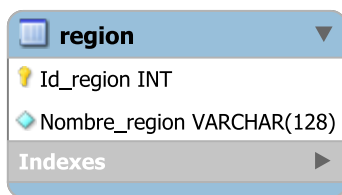


Figura 4: Fuente: *El autor*.

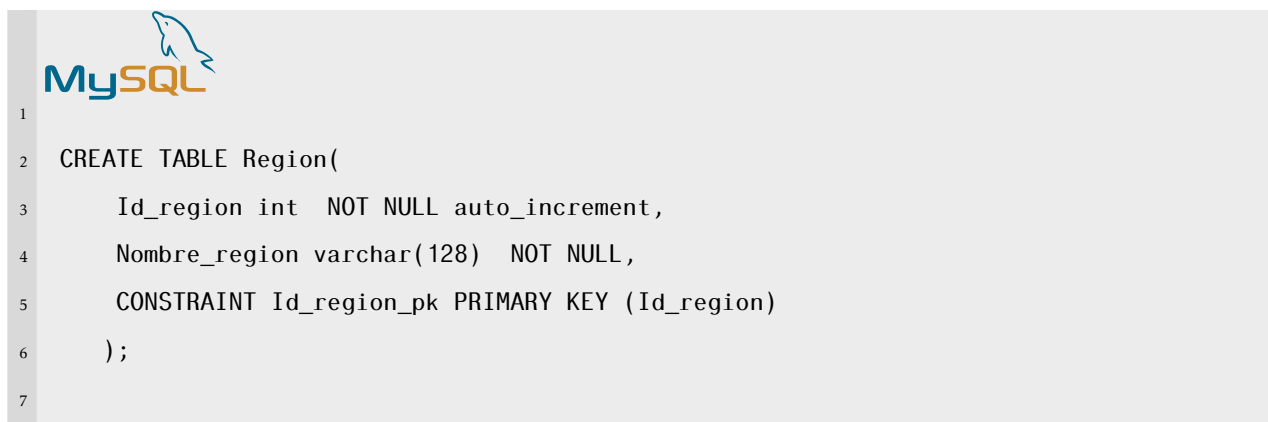


Tabla dimensional de ubicación: Esta tabla contiene información detallada sobre la ubicación geográfica de cada envío, como el código postal, la ciudad, el estado y el país. Esta dimensión permitiría realizar el análisis de las ventas por ubicación geográfica y determinar patrones y tendencias en las ventas por región.

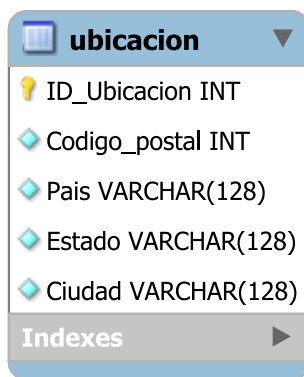


Figura 5: Fuente: *El autor*.



```
1
2 CREATE TABLE Ubicacion (
3     ID_Ubicacion int auto_increment,
4     Codigo_postal int NOT NULL,
5     Pais varchar(128) NOT NULL,
6     Estado varchar(128) NOT NULL,
7     Ciudad varchar(128) NOT NULL,
8     CONSTRAINT ID_UbicacionPK PRIMARY KEY (ID_Ubicacion)
9 );
10
```

Tabla dimensional de fecha de envío: Esta tabla contiene información detallada sobre la fecha en que se envió cada pedido, lo que permite analizar eficacia de las operaciones de envío y logística.

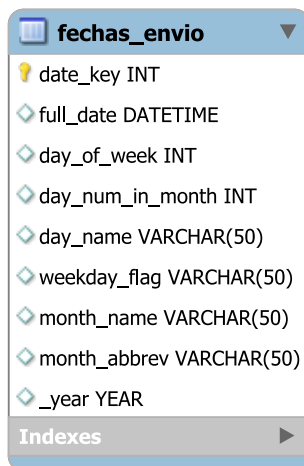


Figura 6: Fuente: *El autor*.



```
1
2 CREATE TABLE Fechas_envio(
3     date_key int,
4     full_date datetime,
5     day_of_week int,
```

```

6      day_num_in_month int,
7      day_name varchar(50),
8      weekday_flag varchar(50),
9      month_name varchar(50),
10     month_abbrev varchar(50),
11     _year year,
12
13     CONSTRAINT Id_dateenvioPK PRIMARY KEY (date_key)
14 );
15

```

Tabla dimensional de fecha de venta (orden): Esta tabla contiene información detallada sobre la fecha en que se realizó cada venta, lo que permite desglosar las ventas por día, semana, mes o año y evaluar la tendencia de ventas.

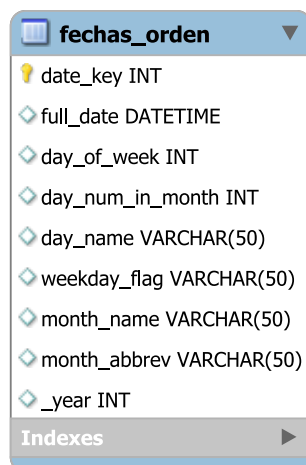



Figura 7: Fuente: *El autor*.



```

1
2
3  CREATE TABLE Fechas_orden(
4      date_key int,
5      full_date datetime,
6      day_of_week int,

```

```

7      day_num_in_month int,
8      day_name varchar(50),
9      weekday_flag varchar(50),
10     month_name varchar(50),
11     month_abbrev varchar(50),
12     _year int,
13     CONSTRAINT Id_dateordenPK PRIMARY KEY (date_key)
14 );
15

```

Tabla dimensional de tipo de envío: Esta tabla contiene información detallada sobre el tipo de envío utilizado para cada pedido. Esta dimensión permitiría evaluar la efectividad de diferentes tipos de envío y tomar decisiones informadas sobre cómo optimizar los procesos de envío y logística.

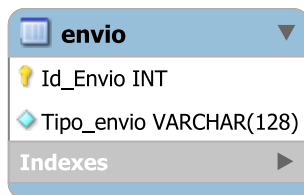



Figura 8: Fuente: *El autor*.



```

1
2  CREATE TABLE Envio (
3      Id_Envio int auto_increment,
4      Tipo_envio varchar(128) NOT NULL,
5      CONSTRAINT Id_ordenPK PRIMARY KEY (Id_Envio)
6  );
7

```

Tabla dimensional de cliente: Esta tabla contiene información detallada sobre los clientes, como su nombre, segmento al que pertenece y otra información relevante. Esta dimensión permitiría analizar el comportamiento de compra de los clientes, como las categorías de productos que compran con más frecuencia o la frecuencia de sus compras.

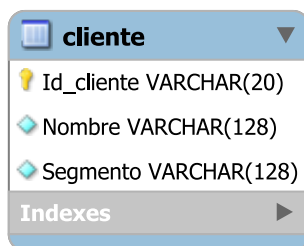


Figura 9: Fuente: *El autor*.

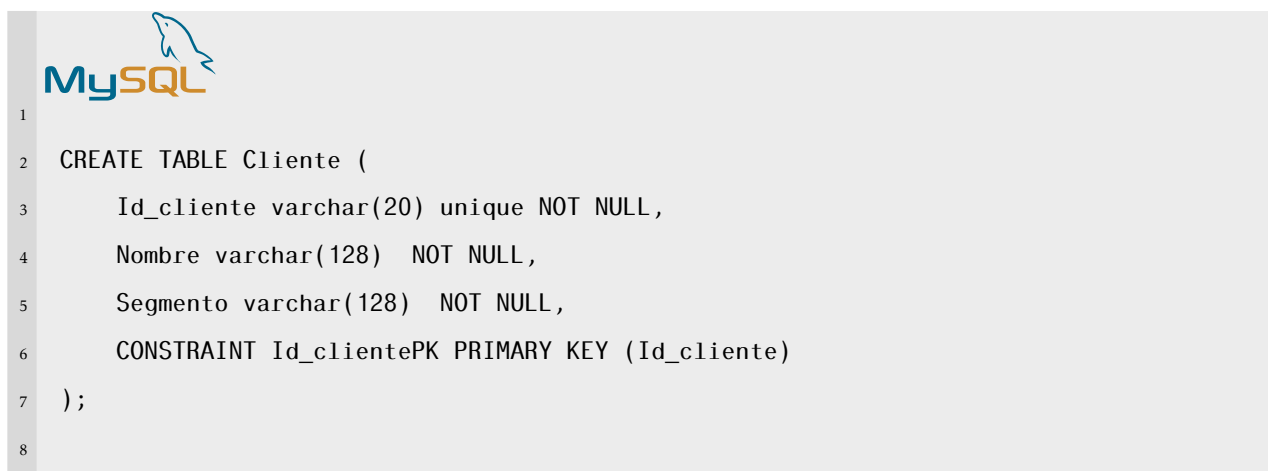


Tabla dimensional de producto: Esta tabla contiene información detallada sobre los productos, como su nombre y categoría. Esta dimensión permitiría desglosar las ventas por producto y analizar la rentabilidad de cada producto.

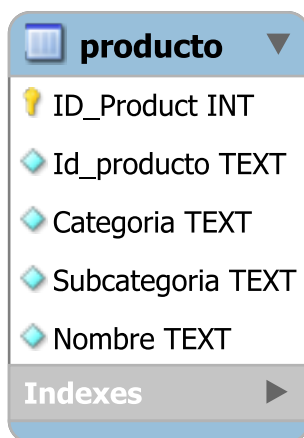


Figura 10: Fuente: *El autor*.



```
1
2 CREATE TABLE Producto (
3     ID_Product int auto_increment,
4     Id_producto text NOT NULL,
5     Categoria text NOT NULL,
6     Subcategoria text NOT NULL,
7     Nombre text NOT NULL,
8     CONSTRAINT Id_productoPK PRIMARY KEY (ID_Product)
9 );
10
```

El uso de las tablas dimensionales de sucursal, ubicación, fecha de envío, fecha de venta, tipo de envío, cliente, producto y categoría de producto sería beneficioso para la construcción de un data warehouse de ventas de un supermercado. Estas dimensiones proporcionarían un contexto adicional para analizar los datos de ventas y permitirían a los analistas de datos realizar análisis detallados

2.5. Identificación de tabla de Hechos

La tabla de hechos en este caso estaría compuesta por información relacionada con cada venta diaria, y las dimensiones proporcionarían información adicional para analizar los datos. La tabla de hechos en un data warehouse de ventas diarias de un supermercado estaría compuesta por información detallada sobre las ventas diarias, con las dimensiones que proporcionan contexto adicional para el análisis.

las ventas diarias, la cantidad de productos, la ganancia y los descuentos como métricas en la tabla de hechos de un data warehouse de ventas de un supermercado es apropiado y relevante por varias razones:

Las ventas diarias proporcionan información detallada sobre la cantidad de productos vendidos en un período de tiempo específico, lo que permite al analista de datos determinar cuáles son los días de mayor y menor venta. Esta información es importante para

el pronóstico de la demanda futura y para la planificación de la reposición de inventario.

La cantidad de productos vendidos es una métrica clave para la rentabilidad del supermercado. La cantidad de productos vendidos junto con el precio de venta unitario se puede utilizar para calcular los ingresos totales y, junto con el costo de los productos, se puede calcular la ganancia bruta. Estas métricas permiten a los analistas evaluar la rentabilidad de los productos y tomar decisiones estratégicas en consecuencia.

Los descuentos son una métrica importante porque permiten a los analistas de datos determinar cuántos productos se vendieron con descuento y en qué medida afectan a la rentabilidad del supermercado. Los descuentos pueden ser utilizados como una estrategia de marketing para atraer a los clientes, pero si se utilizan en exceso, pueden reducir la rentabilidad del negocio.

La ganancia es una métrica importante para evaluar el desempeño financiero del supermercado. Permite a los analistas de datos evaluar la eficacia de las decisiones estratégicas en términos de rentabilidad.

Estas métricas permiten a los analistas de datos realizar análisis detallados sobre el rendimiento financiero del supermercado y tomar decisiones informadas para mejorar la rentabilidad del negocio.

A continuación se muestra el diagrama relacional entre las tablas y el código con las llaves foráneas que permite la relación para construir la tabla de hechos.

fact_super	
sk_super_fact	INT
Fecha_orden	INT
Fecha_envio	INT
Id_orden	VARCHAR(14)
Id_region	INT
Id_ubicacion	INT
Id_cliente	VARCHAR(20)
Id_producto	INT
Id_envio	INT
Venta	FLOAT
Cantidad	INT
Descuento	FLOAT
Ganancia	FLOAT
Indexes	

Figura 11: Fuente: *El autor*.



```

1
2 CREATE TABLE Fact_super (
3   sk_super_fact int auto_increment,
4   Fecha_orden int,
5   Fecha_envio int,
6   Id_orden varchar(14) not null,
7   Id_region int NOT NULL,
8   Id_ubicacion int NOT NULL,
9   Id_cliente varchar(20) NOT NULL,
10  Id_producto int NOT NULL,
11  Id_envio int NOT NULL,
12  Venta float NOT NULL,
13  Cantidad int NOT NULL,
14  Descuento float,
15  Ganancia float NOT NULL,
16  CONSTRAINT Id_ventaPK PRIMARY KEY (sk_super_fact)
17 );
18

```

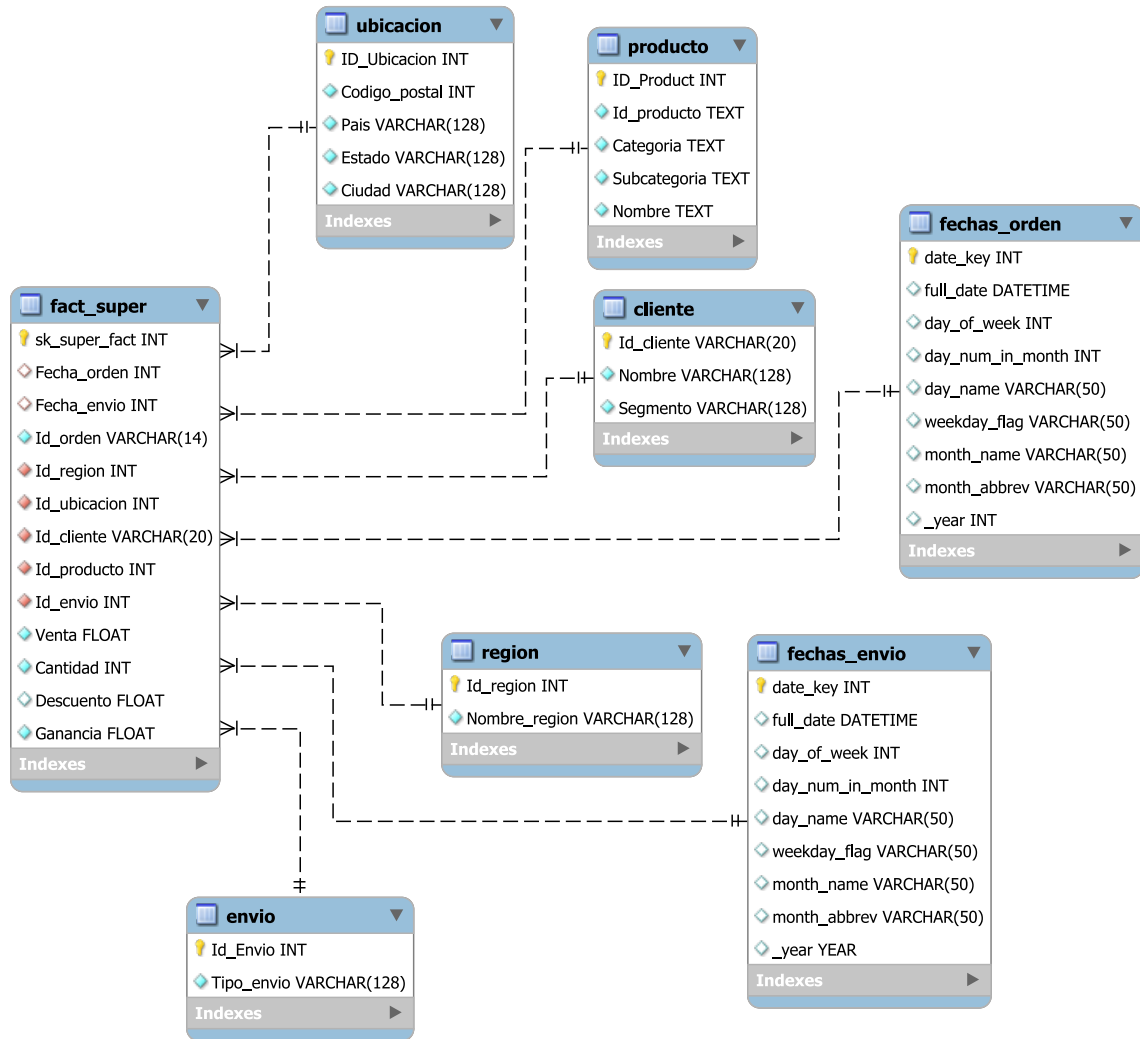


Figura 12: Fuente: *El autor*.



```

1
2 ALTER TABLE Fact_super ADD CONSTRAINT fact_Fechas_orden
3     FOREIGN KEY (Fecha_orden)
4     REFERENCES Fechas_orden (date_key);
5
6 ALTER TABLE Fact_super ADD CONSTRAINT fact_Fechas_envio
7     FOREIGN KEY (Fecha_envio)
8     REFERENCES Fechas_envio (date_key);
9
    
```

```
10 ALTER TABLE Fact_super ADD CONSTRAINT fact_region
11     FOREIGN KEY (Id_region)
12     REFERENCES Region (Id_region);
13
14 ALTER TABLE Fact_super ADD CONSTRAINT fact_Ubicacion
15     FOREIGN KEY (Id_ubicacion)
16     REFERENCES Ubicacion (ID_Ubicacion);
17
18 ALTER TABLE Fact_super ADD CONSTRAINT fact_cliente
19     FOREIGN KEY (Id_cliente)
20     REFERENCES Cliente (Id_cliente);
21
22 ALTER TABLE Fact_super ADD CONSTRAINT fact_producto
23     FOREIGN KEY (Id_producto)
24     REFERENCES Producto (ID_Product);
25
26 ALTER TABLE Fact_super ADD CONSTRAINT fact_orden
27     FOREIGN KEY (Id_envio)
28     REFERENCES Envio (Id_Envio);
29
30
```

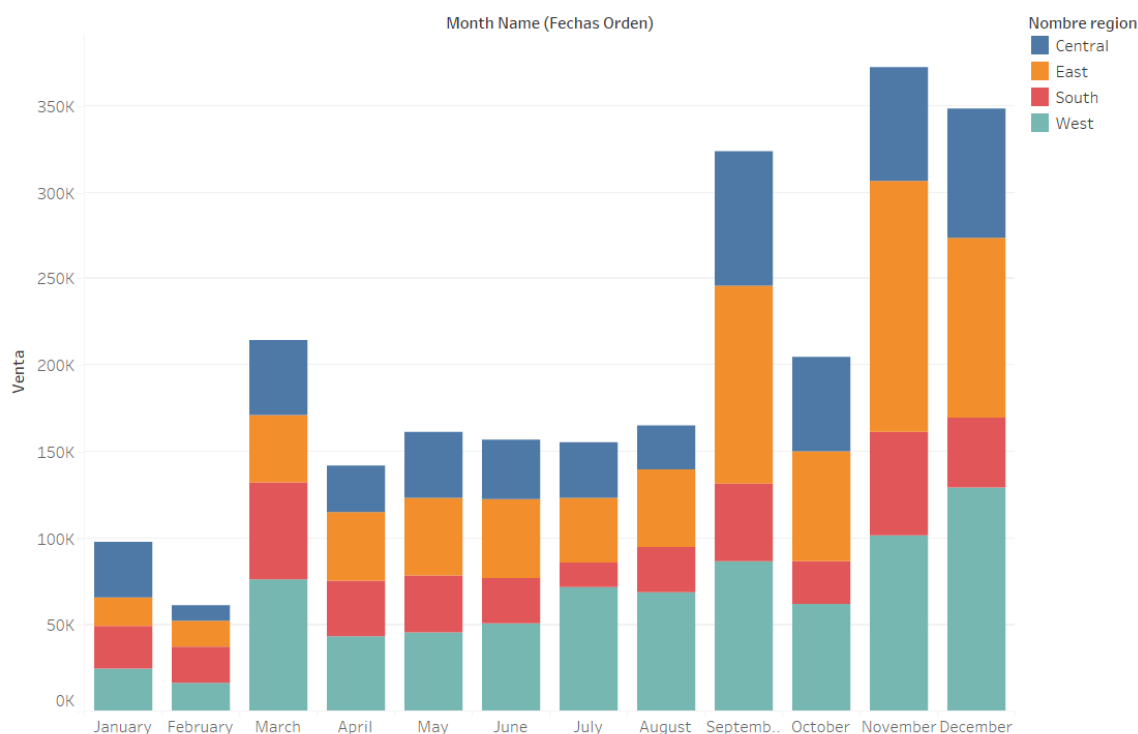
3. EXPLORACIÓN DE DATOS

La exploración de datos se hará contestando las siguientes preguntas.

3.1. ¿Cuáles son los meses de mayor venta por cada región?

Los resultados de la exploración de datos queda de la siguiente forma:

Ventas mensuales por región del país



Suma de Venta para cada Month Name (Fechas Orden). El color muestra detalles acerca de Nombre region. Los datos se filtran en Acción (Categoría, Subcategoría) y Acción (Nombre region). El filtro Acción (Categoría, Subcategoría) conserva 17 miembros. El filtro Acción (Nombre region) conserva 4 miembros.

Figura 13: Fuente: *El autor*.

3.2. ¿Cuál es el porcentaje de venta por región del país?

Los resultados de la exploración de datos queda de la siguiente forma:

Porcentaje de ventas por región del país

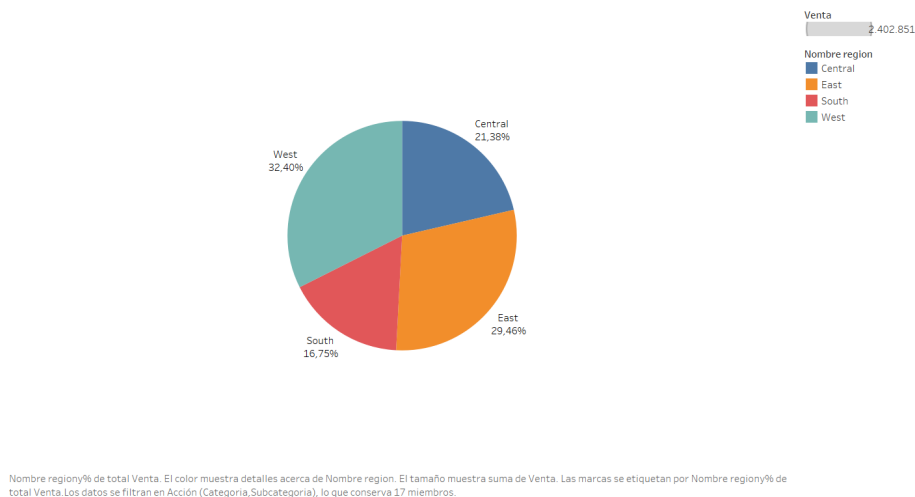


Figura 14: Fuente: *El autor*.

3.3. ¿Qué categorías y subcategorías de productos presentan las mayores ganancias para la compañía?

Los resultados de la exploración de datos queda de la siguiente forma:

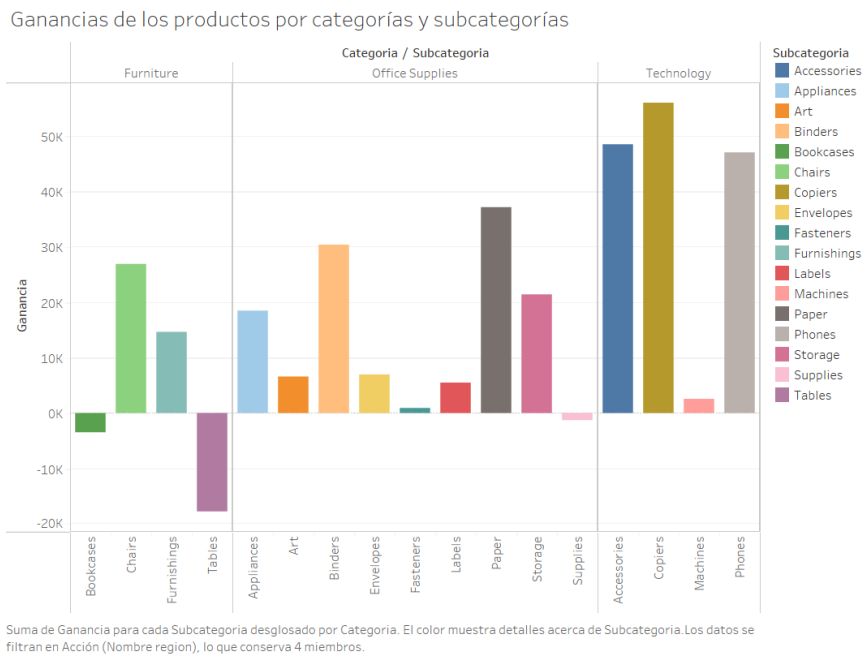


Figura 15: Fuente: *El autor*.

3.4. ¿Qué tipo de envío presenta mayores ventas durante los fines de semana, en un periodo del 2016 al 2018?

Los resultados de la exploración de datos queda de la siguiente forma:

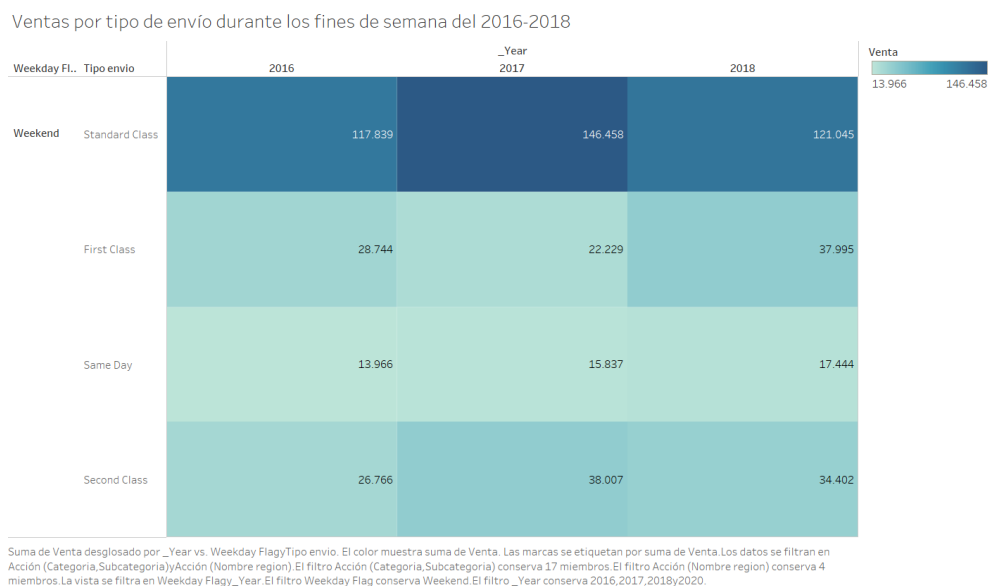
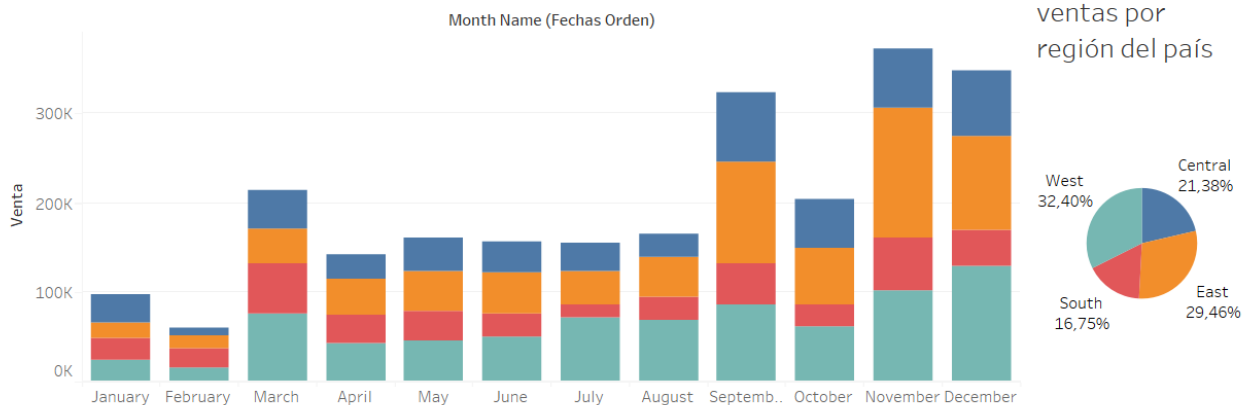


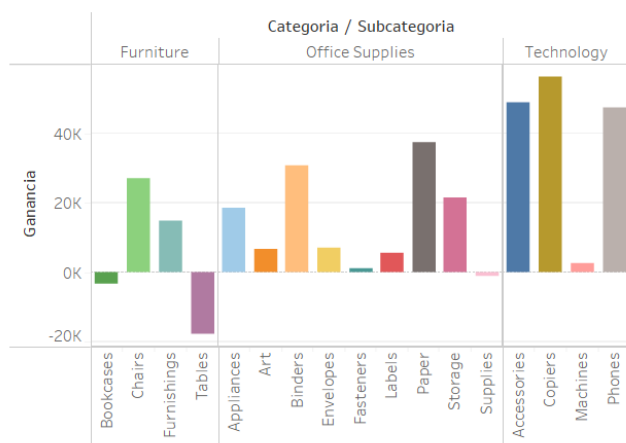
Figura 16: Fuente: *El autor*.

3.5. La agrupación de la exploración

Ventas mensuales por región del país



Ganancias de los productos por categorías y subcategorías



Ventas por tipo de envío durante los fines de semana del 2016-2018

Weekday Fl.	Tipo envío	2016	_Year 2017	2018
Weekend	Standard Class	117.839	146.458	121.045
	First Class	28.744	22.229	37.995
	Same Day	13.966	15.837	17.444
	Second Class	26.766	38.007	34.402

Figura 17: Fuente: *El autor*.

4. CONCLUSIONES

Al realizar el desarrollo de un data warehouse basado en la información de ventas de un supermercado, es posible identificar patrones y tendencias en las ventas, conocer cuáles son los productos más vendidos en cada sucursal y en cada categoría de producto, detectar oportunidades de mejora en el proceso de envío y entrega de productos, y tomar decisiones informadas para mejorar la rentabilidad del negocio.

En resumen, la construcción de un data warehouse para un supermercado con granularidad en las ventas diarias y tablas dimensionales resulta en una herramienta útil y valiosa para la toma de decisiones basadas en datos, que puede ayudar a mejorar la eficiencia y la rentabilidad de la empresa.