



Khiara Nicol Chavez Cartagena
Data Analytics

PROYECTO FINAL

INTRODUCCIÓN.....	3
1.1. Descripción del Dataset:.....	3
1.2. Objetivo:.....	3
1.3. Hipótesis:.....	4
1.4. Alcance:.....	4
1.4.1. Aspectos Incluidos:.....	4
1.4.2. Aspectos Excluidos:.....	5
1.5. Usuario Final.....	5
1.7. Diagrama de Entidad Relación:.....	8
1.8. Listado de las tablas:.....	9
1.9. Listado de las columnas de cada tabla.....	12
1.10. Maqueta Dashboard Inicial - Power BI.....	14
1.11. Power BI - Documentación del Proyecto.....	14
1.12. Power BI - Solapas Finales.....	24
1. Portada.....	24
2. Glosario.....	24
3. Análisis General.....	25
4. Análisis por Empleados.....	26
5. Desempeño por Categoría de Producto.....	26
6. Relación entre Antigüedad y Categorías de Producto.....	27
1.13. Futuras Líneas y Posibles Iniciativas para Complementar el Proyecto.....	28

INTRODUCCIÓN

La temática a analizar en el presente Dataset será el registro de las ventas y sus dimensiones (producto, clientes, empleados, tiendas y tipo de venta).

1.1. Descripción del Dataset:

El dataset proporcionado contiene la siguiente información:

1. Ventas: Registra transacciones de ventas con detalles como fecha de transacción, tienda, empleado, cliente, tipo de venta, producto vendido, cantidad y precio unitario.
2. Productos: Contiene información sobre los productos, organizados por grupos de productos.
3. Clientes: Incluye información de los clientes, como su tienda de origen, nombre, correo electrónico, fecha de nacimiento y número de tarjeta de fidelidad.
4. Empleados: Lista a los empleados con su información básica, como nombre, cargo y ubicación.
5. Tiendas: Describe las tiendas, con detalles sobre el tipo de tienda, dirección, ciudad, estado, código postal y coordenadas geográficas.
6. Tipo venta: Define los diferentes tipos de ventas, como en línea o en tienda física.

1.2. Objetivo:

Analizar la relación entre el rendimiento de los empleados y las ventas de diferentes productos, identificando si ciertos empleados o categorías de productos generan mayores ingresos.

1.3. Hipótesis:

Los empleados con mayor antigüedad generan más ventas de productos de mayor valor, como las categorías de "Whole Bean/Teas", en comparación con empleados con menos experiencia o que se encargan de ventas de productos de menor valor, como "Beverages".

1.4. Alcance:

El presente dataset abarca la información necesaria para analizar la relación entre el rendimiento de los empleados y las ventas de productos en diferentes categorías. El alcance del análisis incluye la identificación de patrones que correlacionen la antigüedad y el desempeño de los empleados con las ventas, permitiendo determinar si ciertos productos o categorías generan mayores ingresos y si estos están vinculados a la experiencia del personal.

1.4.1. Aspectos Incluidos:

- Ventas: Se incluirá información sobre cada transacción realizada, detallando aspectos como la fecha de la venta, el tipo de producto vendido, la cantidad y el precio unitario, así como la existencia de promociones.
- Empleados: Se incluirán los datos de los empleados, específicamente en su identificación, antigüedad, posición y ubicación, lo que permitirá establecer correlaciones entre la experiencia y el volumen o valor de ventas que generan.
- Productos: El dataset contempla las diferentes categorías de productos, permitiendo clasificar las ventas según el tipo de producto (por ejemplo, "Whole Bean/Teas" o "Beverages") para analizar qué categorías generan mayores ingresos y cómo estos se distribuyen entre diferentes empleados.

- Clientes y Tiendas: La relación entre los clientes y las tiendas también será incluida, analizando cómo las características del cliente y la tienda (como su ubicación o tipo de tienda) influyen en el patrón de ventas y en el desempeño de los empleados en diferentes contextos.

1.4.2. Aspectos Excluidos:

- El análisis no incluirá aspectos relacionados con costos operativos ni análisis financieros detallados que no estén vinculados directamente a las ventas.
- Tampoco se considerará el análisis de la satisfacción del cliente u otras métricas cualitativas relacionadas con la percepción de los productos o del servicio de los empleados.

1.5. Usuario Final

El usuario final de este dataset estará compuesto por varios grupos clave dentro de la organización, cada uno con necesidades específicas que se verán beneficiadas directamente del análisis de los datos:

- Área de Ventas:
El equipo de ventas podrá identificar qué empleados están logrando los mejores resultados en términos de ventas y en qué productos o categorías destacan. Esto les permitirá establecer metas más precisas, definir estrategias de incentivos para los empleados más productivos y diseñar capacitaciones específicas para aquellos con menor rendimiento.
- Área de Recursos Humanos:
Los datos relacionados con la antigüedad, posición y ubicación de los empleados permitirán al equipo de recursos humanos evaluar el impacto de

la experiencia en el rendimiento de ventas. Esto puede ser clave para tomar decisiones sobre contrataciones, promociones y planificación de carreras.

- **Área de Marketing:**

El análisis permitirá al equipo de marketing entender qué productos tienen mayor aceptación entre los clientes y cómo el perfil de los empleados y las tiendas afecta estas ventas. A partir de esta información, podrán desarrollar campañas de marketing más efectivas, enfocadas en los productos y categorías que generan mayores ingresos.

- **Alta Dirección:**

La alta dirección podrá utilizar este análisis para tomar decisiones estratégicas a nivel macro sobre la asignación de recursos, incentivos y formación del personal. Además, podrán evaluar si es necesario ajustar la estructura de ventas o el enfoque comercial en función de los resultados obtenidos.

Los usuarios finales esperan que el análisis proporcione una visión clara y basada en datos sobre el rendimiento de ventas, destacando posibles áreas de mejora y ofreciendo recomendaciones específicas. El objetivo es que, a través del análisis de este dataset, puedan tomar decisiones informadas que impulsen el crecimiento de los ingresos y la optimización de recursos, tanto a nivel de empleados como de productos.

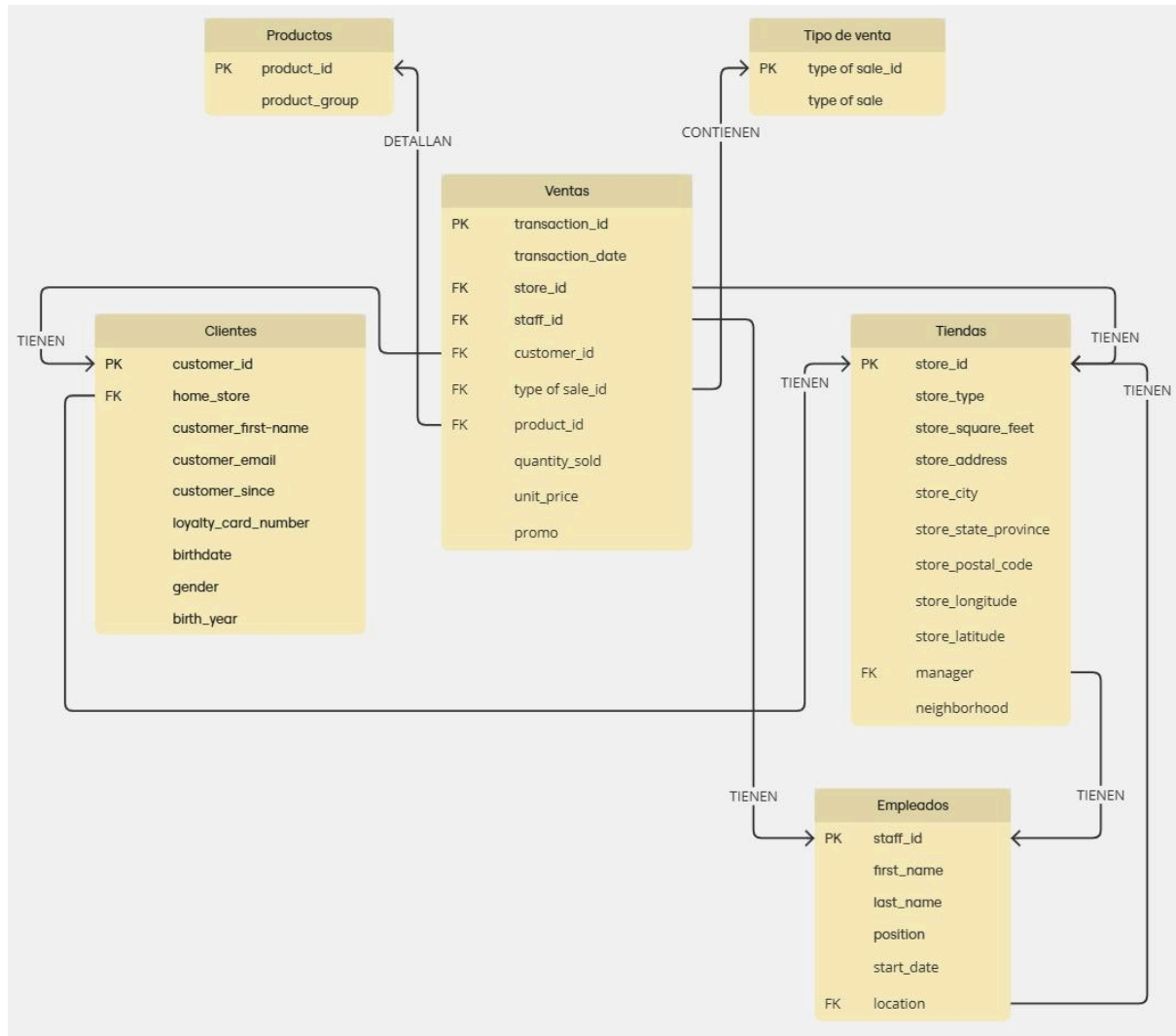
1.6. Nivel de aplicación del análisis:

El análisis de este dataset se sitúa en el nivel táctico, ya que está orientado a la toma de decisiones a mediano plazo que impactan áreas específicas de la organización, como el departamento de ventas, marketing y recursos humanos. A través de este análisis, los gerentes y ejecutivos podrán optimizar procesos dentro

de sus equipos, implementando estrategias para mejorar el rendimiento de los empleados y aumentar las ventas de productos clave.

Este nivel táctico permite monitorear el progreso hacia los objetivos estratégicos de la empresa, identificando oportunidades de mejora o problemas en tiempo real. Además, al analizar la relación entre la antigüedad de los empleados y su impacto en la venta de productos de mayor valor, las áreas de ventas y marketing podrán ajustar sus enfoques de capacitación, incentivos y distribución de productos. Recursos Humanos, por su parte, podrá utilizar esta información para tomar decisiones relacionadas con la formación y gestión del personal.

1.7. Diagrama de Entidad Relación:



1.8. Listado de las tablas:

➤ Ventas:

Esta tabla almacena información sobre cada transacción de venta realizada, incluyendo datos sobre el producto vendido, el empleado que la realizó, el cliente, y la tienda.

- PK: transaction_id (Identificador único de cada transacción).
- FK:
 - store_id: Relacionado con la tabla Tiendas.
 - staff_id: Relacionado con la tabla Empleados.
 - customer_id: Relacionado con la tabla Clientes.
 - type of sale_id: Relacionado con la tabla Tipo de Venta.
 - product_id: Relacionado con la tabla Productos.

Relación con otras tablas:

Ventas se relaciona con Tiendas, Empleados, Clientes, Productos y Tipo de Venta a través de claves foráneas, almacenando datos de las transacciones que implican cada entidad.

➤ Productos:

Esta tabla almacena información sobre los productos disponibles para la venta, incluyendo su grupo o categoría.

- PK: product_id (Identificador único de cada producto).

Relación con otras tablas:

Productos se relaciona con Ventas a través de la clave foránea product_id.

➤ Clientes:

Contiene los detalles de cada cliente, como su nombre, correo electrónico, y tienda de origen.

- PK: customer_id (Identificador único de cada cliente).
- FK:
 - home_store: Relacionado con la tabla Tiendas.

Relación con otras tablas:

Cientes se relaciona con Ventas a través de la clave foránea customer_id, y con Tiendas a través de home_store.

➤ **Empleados:**

Almacena los datos de los empleados, como sus nombres, posición y la tienda en la que trabajan.

- PK: staff_id (Identificador único de cada empleado).
- FK:
 - location: Relacionado con la tabla Tiendas.

Relación con otras tablas:

Empleados se relaciona con Ventas a través de la clave foránea staff_id, y con Tiendas a través de manager.

➤ **Tiendas:**

Esta tabla almacena información sobre las tiendas, como su dirección, tipo, y el gerente a cargo.

- PK: store_id (Identificador único de cada tienda).
- FK:
 - manager: Relacionado con la tabla Empleados.

Relación con otras tablas:

Tiendas se relaciona con Ventas, Clientes, y Empleados a través de las claves foráneas store_id, home_store y location, respectivamente.

➤ **Tipo de Venta:**

Almacena los diferentes tipos de venta que se pueden realizar.

- PK: type of sale_id (Identificador único del tipo de venta).

Relación con otras tablas:

Tipo de Venta se relaciona con Ventas a través de la clave foránea type of sale_id

1.9. Listado de las columnas de cada tabla

Tabla 1: Ventas

Tipo de clave	Campo	Tipo de Campo
PK - index	transaction_id	int
-	transaction_date	date
FK	store_id	int
FK	staff_id	int
FK	customer_id	int
FK	type of sale_id	int
FK	product_id	int
-	quantity_sold	int
-	unit_price	int
-	promo	boolean

Tabla 2: Productos

Tipo de clave	Campo	Tipo de Campo
PK - index	product_id	int
-	product_group	nvarchar(50)

Tabla 3: Clientes

Tipo de clave	Campo	Tipo de Campo
PK - index	customer_id	int
FK	home_store	int
-	customer_first-name	nvarchar(50)
-	customer_email	nvarchar(100)
-	customer_since	date
-	loyalty_card_number	nvarchar(50)
-	birthdate	date
-	gender	char(1)

-	birth_year	int
---	------------	-----

Tabla 4: Empleados

Tipo de clave	Campo	Tipo de Campo
PK - index	staff_id	int
-	first_name	nvarchar(50)
-	last_name	nvarchar(60)
-	position	nvarchar(15)
-	start_date	date
FK	location	int

Tabla 5: Tiendas

Tipo de clave	Campo	Tipo de Campo
PK - index	store_id	int
-	store_type	nvarchar(10)
-	store_square_feet	int
-	store_address	nvarchar(100)
-	store_city	nvarchar(50)
-	store_state_province	char(2)
-	store_postal_code	nvarchar(20)
-	store_longitude	decimal(9,6)
-	store_latitude	decimal(9,6)
FK	manager	int
-	Neighborhood	nvarchar(50)

Tabla 6: Tipo venta

Tipo de clave	Campo	Tipo de Campo
PK - index	type of sale_id	int

-	Type of sale	nvarchar(20)
---	--------------	--------------

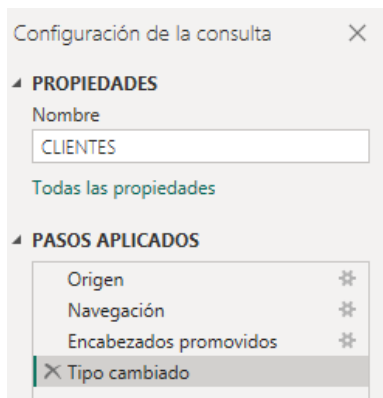
1.10. Power BI - Documentación del Proyecto

Estructura Inicial de Datos:

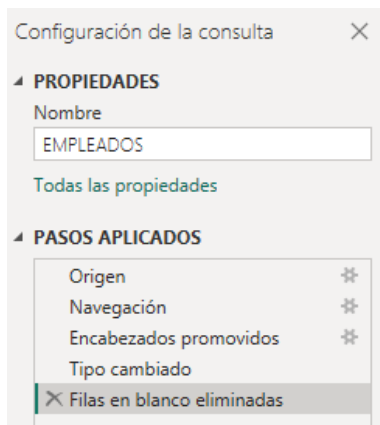
- Conectamos Power BI desktop a la base de datos que será la fuente de datos en el proyecto final.
- Aplicamos el conector de bases de datos de archivo plano requerido, utilizando un documento de extensión “Hoja de cálculo de Google” como fuente de origen de datos. (Hoja de cálculo de Google: https://docs.google.com/spreadsheets/d/1Wp_7bjUnt3zxIS6boDQAEuDNcwn8SnGg/edit?usp=sharing&ouid=113847633781930237516&rtpof=true&sd=true)

Transformaciones realizadas:

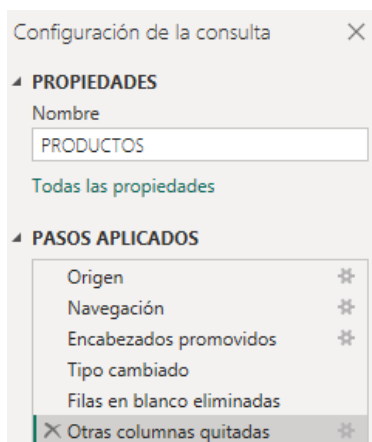
- Se establecen los siguientes cambios en la tabla “Clientes”:



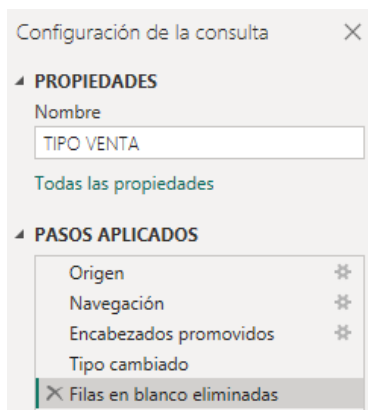
- Se establecen los siguientes cambios en la tabla “Empleados”:



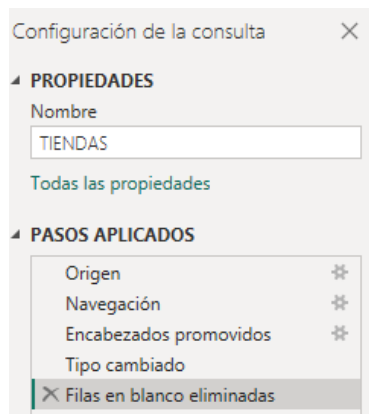
- Se establecen los siguientes cambios en la tabla “Productos”:



- Se establecen los siguientes cambios en la tabla “Tipo Venta”:



- Se establecen los siguientes cambios en la tabla “Tiendas”:



- Se establecen los siguientes cambios en la tabla “Ventas”:

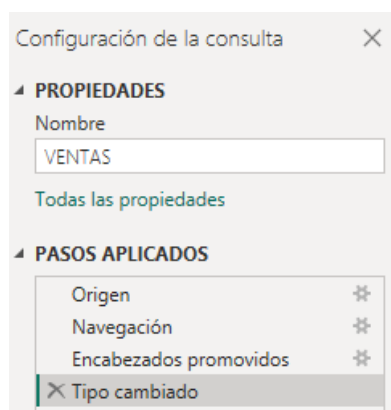
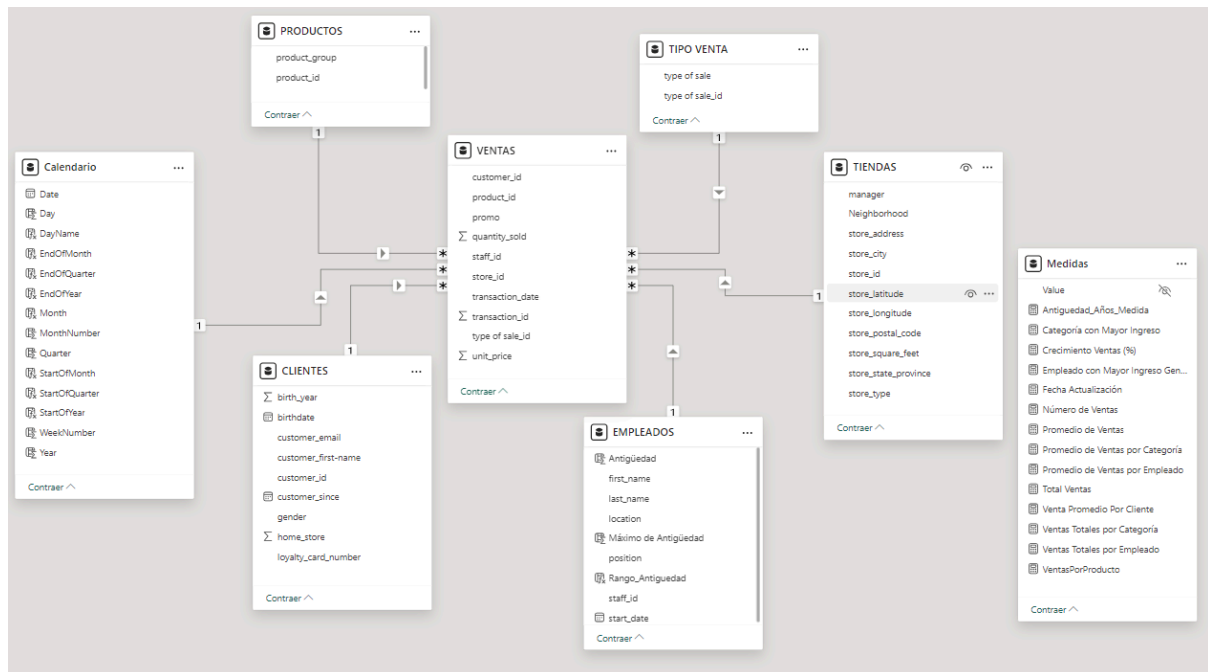


Imagen modelo de Power BI:



- Se establecen las relaciones “Varios a Uno” entre tablas (Clientes, Productos, Tiendas, Empleados, Tipo de Venta) a la tabla Ventas.

Tabla de calendario:

←

Editar relación

×

Permite seleccionar tablas y columnas relacionadas.

Desde la tabla

VENTAS

customer_id	product_id	promo	quantity_sold	staff_id	store_id	transaction_date
8600	24	NO	1	42	8	12/01/2017
8600	24	NO	1	42	8	14/01/2017
8600	24	NO	1	42	8	19/01/2017

A la tabla

Calendario

Date	Day	DayName	EndOfMonth	EndOfQuarter	EndOfYear	Month
01/01/2017	1	domingo	31/01/2017 0...	31/03/2017 0...	31/12/2017 0...	enero
02/01/2017	2	lunes	31/01/2017 0...	31/03/2017 0...	31/12/2017 0...	enero
03/01/2017	3	martes	31/01/2017 0...	31/03/2017 0...	31/12/2017 0...	enero

Cardinality

Varios a uno (*:1)

Dirección de filtro cruzado

Único

☒ Activar esta relación
 ☐ Aplicar filtro de seguridad en ambas direcciones

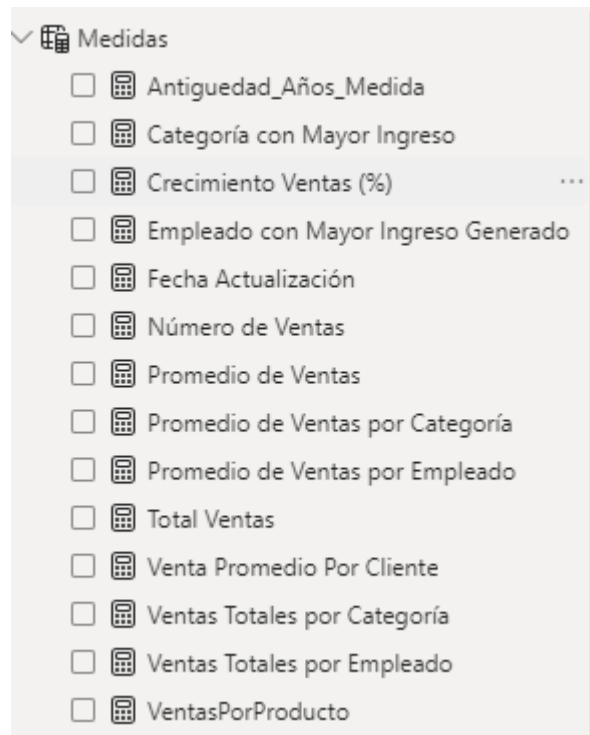
☐ Asumir integridad referencial

Guardar

Cancelar

- Se establece la relación “Varios a Uno” entre la tabla de Calendario a la tabla Ventas.

Medidas calculadas



- Antigüedad_Años_Medida

```
1 Antigüedad_Años_Medida =  
2 DATEDIFF(  
3     MIN(EMPLEADOS[start_date]),  
4     MAX(CALENDARIO[Date]),  
5     YEAR  
6 )
```

- Categoría con Mayor Ingreso

```
1 Categoría con Mayor Ingreso =  
2 MAXX(  
3     TOPN(  
4         1,  
5         SUMMARIZE(  
6             VENTAS,  
7             PRODUCTOS[product_group],  
8             "VentasCategoría", SUMX(VENTAS, VENTAS[quantity_sold] * VENTAS[unit_price])  
9         ),  
10        [VentasCategoría], DESC  
11    ),  
12    PRODUCTOS[product_group]  
13 )  
14
```

- Crecimiento Ventas (%)

```

1 Crecimiento Ventas (%) = DIVIDE(
2   [Total Ventas] - CALCULATE([Total Ventas], DATEADD('Calendario'[Date], -1, YEAR)),
3   CALCULATE([Total Ventas], DATEADD('Calendario'[Date], -1, YEAR))
4 )
5

```

- Empleado con Mayor Ingreso Generado

```

1 Empleado con Mayor Ingreso Generado =
2 VAR Tabla_Ventas =
3   SUMMARIZE(
4     VENTAS,
5     VENTAS[staff_id],
6     "IngresoTotal", SUMX(VENTAS, VENTAS[unit_price] * VENTAS[quantity_sold])
7   )
8 VAR Mayor_Ingreso =
9   MAXX(Tabla_Ventas, [IngresoTotal])
10 VAR ID_Empleado =
11   MAXX(
12     FILTER(
13       Tabla_Ventas,
14       [IngresoTotal] = Mayor_Ingreso
15     ),
16     VENTAS[staff_id]
17   )
18 RETURN
19   LOOKUPVALUE(
20     EMPLEADOS[first_name],
21     EMPLEADOS[staff_id],
22     ID_Empleado
23   )
24

```

- Fecha Actualización

```

1 Fecha Actualización = NOW()

```

- Promedio de Ventas:

```

1 Promedio de Ventas = [Total Ventas] / [Número de Ventas]

```

- Total Ventas

```

1 Total Ventas = SUMX(Ventas, Ventas[unit_price] * Ventas[quantity_sold])

```

- Número de Ventas

```

1 Número de Ventas = COUNT(Ventas[transaction_id])

```

- Promedio de Ventas por Categoría

```

1 Promedio de Ventas por Categoría =
2 AVERAGEX(
3     SUMMARIZE(
4         VENTAS,
5         PRODUCTOS[product_group],
6         "VentasCategoría", SUMX(VENTAS, VENTAS[quantity_sold] * VENTAS[unit_price])
7     ),
8     [VentasCategoría]
9 )

```

- Promedio de Ventas por Empleado

```

1 Promedio de Ventas por Empleado =
2 AVERAGEX(
3     SUMMARIZE(
4         VENTAS,
5         VENTAS[staff_id],
6         "IngresoEmpleado", SUMX(VENTAS, VENTAS[quantity_sold] * VENTAS[unit_price])
7     ),
8     [IngresoEmpleado]
9 )

```

- Venta Promedio por Cliente

```

1 Venta Promedio Por Cliente = DIVIDE([Total Ventas], DISTINCTCOUNT(VENTAS[customer_id]))

```

- Venta Promedio por Categoría

```

1 Ventas Totales por Categoría =
2 SUMX(VENTAS, VENTAS[quantity_sold] * VENTAS[unit_price])

```

- Ventas Totales por Empleado

```

1 Ventas Totales por Empleado =
2 SUMX(
3     VENTAS,
4     VENTAS[quantity_sold] * VENTAS[unit_price]
5 )

```

- Ventas por Producto

```

1 VentasPorProducto = SUMX(Ventas, Ventas[quantity_sold] * Ventas[unit_price])

```

EMPLEADOS	
<input type="checkbox"/>	Antigüedad
<input type="checkbox"/>	first_name
<input type="checkbox"/>	last_name
<input type="checkbox"/>	location
<input type="checkbox"/>	Máximo de Antigüedad
<input type="checkbox"/>	position
<input type="checkbox"/>	Rango_Antigüedad
<input type="checkbox"/>	staff_id
<input type="checkbox"/>	start_date

- Antigüedad

```

1 Antigüedad =
2 DATEDIFF(
3     EMPLEADOS[start_date],
4     DATE(2018, 12, 31),
5     YEAR
6 )

```

- Máximo de Antigüedad

```

1 Máximo de Antigüedad =
2 MAXX(
3     FILTER(
4         VENTAS,
5         VENTAS[staff_id] IN VALUES(EMPLEADOS[staff_id])
6     ),
7     EMPLEADOS[Antigüedad]
8 )
9

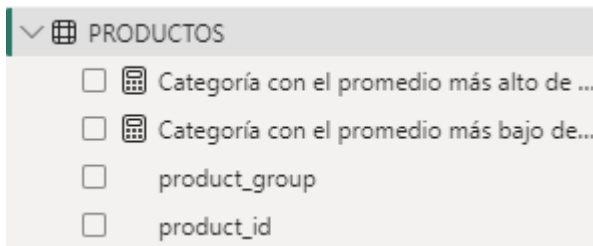
```

- Rango_Antigüedad

```

1 Rango_Antigüedad =
2 SWITCH(
3     TRUE(),
4     EMPLEADOS[Antigüedad] <= 3, "1-3 años",
5     EMPLEADOS[Antigüedad] <= 5, "4-5 años",
6     EMPLEADOS[Antigüedad] > 5, "6+ años",
7     "Sin antigüedad"
8 )

```



- Categoría con el promedio más alto de antigüedad de empleados

```

1 Categoría con el promedio más alto de antigüedad de empleados =
2 MAXX(
3     TOPN(
4         1,
5         SUMMARIZE(
6             VENTAS,
7             PRODUCTOS[product_group],
8             "PromedioAntigüedad",
9             AVERAGEX(
10                VENTAS,
11                RELATED(EMPLEADOS[Antigüedad])
12            )
13         ),
14         [PromedioAntigüedad], DESC
15     ),
16     PRODUCTOS[product_group]
17 )
18

```

- Categoría con el promedio más bajo de antigüedad de empleados

```

1 Categoría con el promedio más bajo de antigüedad de empleados =
2 MAXX(
3     TOPN(
4         1,
5         SUMMARIZE(
6             VENTAS,
7             PRODUCTOS[product_group],
8             "PromedioAntigüedad",
9             AVERAGEX(
10                VENTAS,
11                RELATED(EMPLEADOS[Antigüedad])
12            )
13         ),
14         [PromedioAntigüedad], ASC
15     ),
16     PRODUCTOS[product_group]
17 )
18

```

1.12. Power BI - Solapas Finales

1. Portada

- **Descripción:** Es la página de introducción del dashboard. Contiene el título del proyecto, botones de navegación entre las diferentes solapas y la fecha de actualización del reporte.
- **Propósito:** Sirve como una bienvenida y guía para navegar por las diferentes secciones del dashboard.



2. Glosario

- **Descripción:** Lista de los conceptos clave utilizados en el dashboard, con descripciones funcionales y enriquecidas con emojis para facilitar la comprensión.
- **Propósito:** Ayuda a los usuarios a entender las métricas y términos técnicos utilizados en el análisis, asegurando claridad en la interpretación de los datos.

Glosario del Dashboard
Portada
Glosario
Análisis General
Empleados
Categoría de Producto
Antigüedad y Categorías de

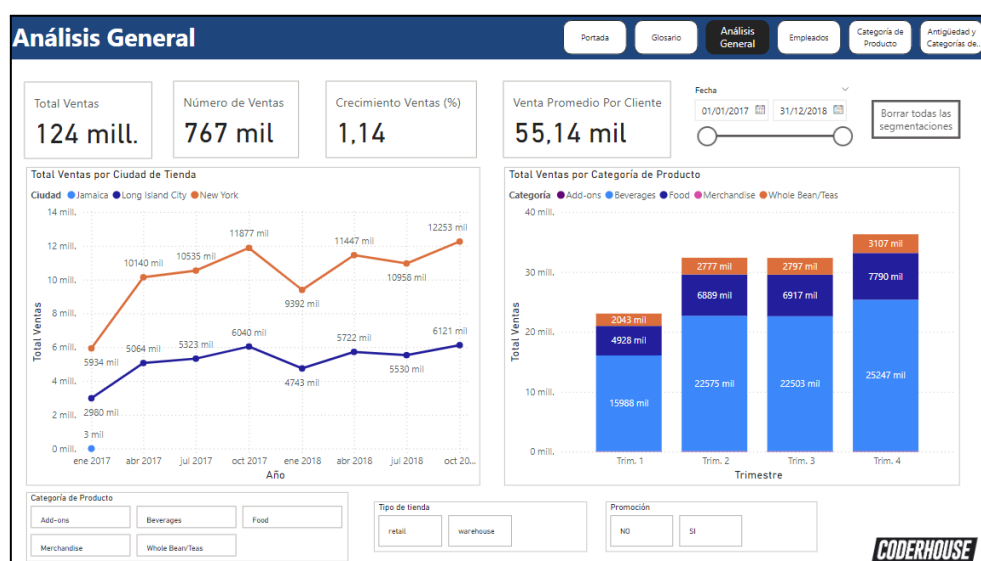
Glosario

Promoción: Indica si una transacción incluyó descuentos o campañas promocionales activas.
Tiendas: Ubicaciones físicas donde se realizaron las transacciones, clasificadas por ciudad y tipo.
Clientes: Información sobre los compradores, diferenciando entre clientes únicos y recurrentes.
Ventas: Total de ingresos generados por las transacciones, considerando unidades y precios.
Crecimiento: Incremento o decremento porcentual de las ventas en comparación con el período anterior.
Ventas por Producto: Total de ingresos por producto, calculado como cantidad vendida x precio unitario.
Productos: Detalle de los bienes vendidos, incluyendo su categoría, precio unitario y cantidades.
Cantidad de Ventas: Número total de transacciones realizadas en un período específico.
Tipo de Venta: Clasificación de las transacciones según su canal: en línea, tienda física o almacén.

CODERHOUSE

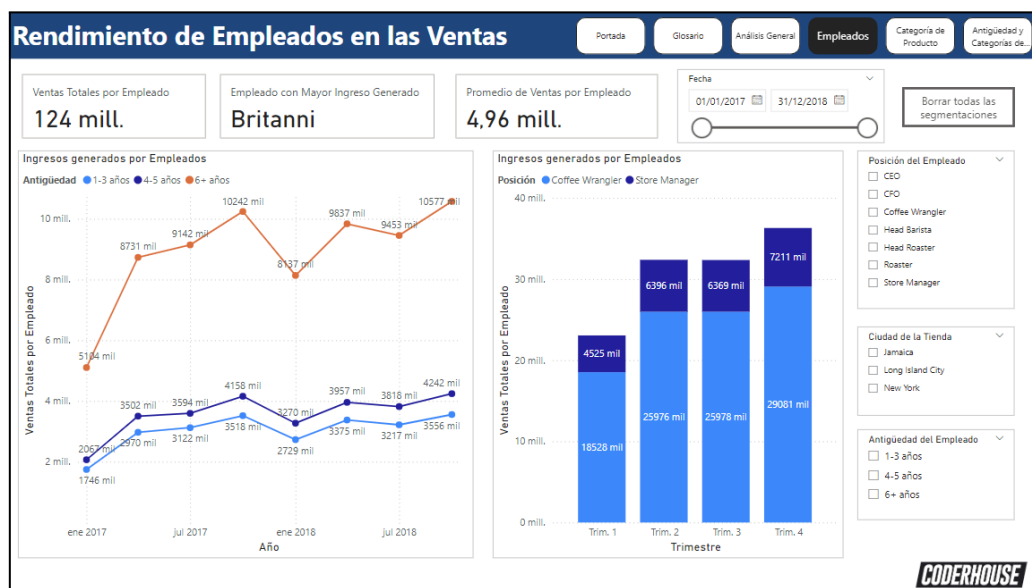
3. Análisis General

- Descripción:** Muestra un resumen global de las ventas, incluyendo métricas principales como ventas totales, número de ventas, y el crecimiento de ventas (%). Además, se presentan gráficos que muestran la evolución temporal de ventas por ciudad y por categoría de producto.
- Propósito:** Proporciona una vista de alto nivel de las ventas, identificando tendencias generales y las ciudades o categorías que generan mayores ingresos.



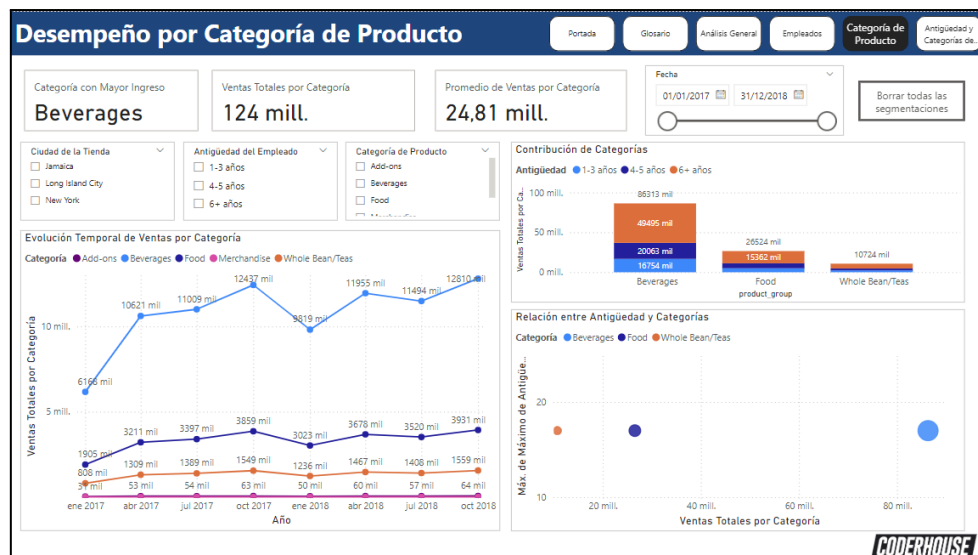
4. Análisis por Empleados

- **Descripción:** Examina el impacto del rendimiento de los empleados en las ventas totales. Incluye indicadores como el empleado con mayor ingreso generado, gráficos de evolución temporal de ventas por antigüedad y distribución por posición.
- **Propósito:** Permite analizar cómo la experiencia y posición de los empleados influyen en las ventas y determinar patrones relacionados con el desempeño.



5. Desempeño por Categoría de Producto

- **Descripción:** Presenta un análisis detallado de las ventas por categoría de producto. Incluye gráficos de evolución temporal, una distribución de categorías según antigüedad y un gráfico de dispersión que relaciona las ventas con la experiencia.
- **Propósito:** Ayuda a identificar qué categorías son más rentables y cómo están relacionadas con la antigüedad de los empleados responsables de esas ventas.



6. Relación entre Antigüedad y Categorías de Producto

- Descripción:** Explora cómo la antigüedad de los empleados impacta en las categorías de productos. Muestra las categorías con mayor y menor promedio de antigüedad, la evolución temporal de ventas por antigüedad y un gráfico de impacto por categoría.
- Propósito:** Determina si ciertas categorías dependen de empleados con más experiencia, brindando información valiosa para asignaciones y estrategias de capacitación.



1.13. Futuras Líneas y Posibles Iniciativas para Complementar el Proyecto

- Modelos Predictivos: Anticipar tendencias de ventas y rendimiento de empleados para optimizar la planificación estratégica.
- Análisis de Rentabilidad: Incorporar márgenes de ganancia para identificar productos o categorías más rentables.
- Segmentación de Clientes: Analizar perfiles y comportamientos de clientes para personalizar estrategias de marketing.
- Impacto de Promociones: Evaluar la efectividad de campañas promocionales en las ventas totales y por categorías.
- Optimización de Empleados: Diseñar capacitaciones para empleados menos experimentados según su desempeño.
- Expansión Geográfica: Analizar oportunidades en nuevas regiones para planificar estrategias de crecimiento.