



UNIVERSIDAD PRIVADA DEL VALLE

FACULTAD: FACULTAD DE INFORMÁTICA Y ELECTRÓNICA

CARRERA: INGENIERÍA DE SISTEMAS INFORMÁTICOS

MATERIA: DATA WAREHOUSING

Proyecto Final:

Sistema de Análisis de Ventas y Producción para Pastelería

INTEGRANTES:

- Antequera Paco Robert Ruben
- Villarroel Galarza Vivian Mabel

MATERIA: Data Warehousing

DOCENTE: Ing. Zeballos Orellana Fernando

FECHA PRESENTACIÓN INFORME: 11/12/2025

1. INTRODUCCIÓN

Este proyecto académico implementa un Data Warehouse para analizar operaciones de ventas y producción de una pastelería. Siguiendo la metodología Kimball, se desarrolla un sistema completo que incluye diseño dimensional, procesos ETL, y dashboards de visualización. El objetivo principal es demostrar la aplicación práctica de conceptos de data warehousing en un escenario empresarial simulado.

2. OBJETIVOS

Objetivo General

Diseñar e implementar un Data Warehouse dimensional para analizar ventas y producción de pastelería, aplicando la metodología Kimball en un proyecto académico.

Objetivos Específicos

- Definir requerimientos analíticos del negocio simulado
- Diseñar modelo en estrella con 2 hechos y 5 dimensiones
- Desarrollar procesos ETL con Talend
- Crear dashboards en Power BI con los gráficos solicitados
- Documentar todo el proceso según estructura requerida

3. DEFINICIÓN DE REQUISITOS DEL NEGOCIO

3.1 Análisis Solicitados

Para Dashboard de Ventas:

- Total ventas por año
- Total ventas por tipo cliente
- Total ventas por ciudad
- Total de ventas por productos
- Total de ventas por categoría
- Total ventas por sabor
- Descuento total por tipo de cliente
- Cantidad vendida por sucursal
- Cantidad vendida por tamaño

Para Dashboard de Producción:

- Cantidad producida por producto
- Merma total por año
- Tiempo total de producción por empleado
- Cantidad producida por ciudad
- Merma total por producto
- Total producción por antigüedad de empleados
- Costo materia prima por producto
- Cantidad producida por turno
- Merma total por turno
- Cantidad producida por sucursal

3.2 Preguntas Clave a Responder

- ¿Cómo han evolucionado las ventas anualmente?
- ¿Qué tipo de cliente genera más ingresos?
- ¿Qué productos y categorías son más populares?
- ¿Qué sucursales tienen mejor desempeño?
- ¿Cómo se comportan las mermas en el tiempo?
- ¿Qué empleados son más eficientes?
- ¿Qué turnos tienen mejor rendimiento?

4. SELECCIÓN DE HERRAMIENTAS Y PRODUCTOS

4.1 Stack Tecnológico

Base de Datos

- Oracle Database: Para almacenamiento del data warehouse
- Justificación: Requerimiento académico, robustez para modelos dimensionales

Integración de Datos

- Talend Open Studio: Para procesos ETL
- Justificación: Interfaz gráfica intuitiva, amplia conectividad

Visualización

- Power BI Desktop: Para creación de dashboards
- Justificación: Facilidad de uso, amplias capacidades visuales

Control de Versiones

- GitHub: Para gestión del código
- Justificación: Colaboración en equipo, histórico de cambios

5. MODELO DIMENSIONAL

5.1 Esquema en Estrella

Tablas de Hechos

1. fac_ventas (Ventas)
 - Medidas: cantidad, total_venta, descuento
 - Dimensiones: producto, cliente, tiempo, sucursal
2. fac_produccion (Producción)
 - Medidas: cantidad_prod, costo_mp, tiempo_produccion, merma
 - Dimensiones: producto, tiempo, sucursal, empleado

Tablas de Dimensiones

1. dim_producto (15 productos)
2. dim_cliente (15 clientes)
3. dim_tiempo (período 2019-2021)
4. dim_empleado (15 empleados)
5. dim_sucursal (10 sucursales)

6. DISEÑO FÍSICO

6.1 Estructura de Base de Datos

Usuarios Creados

1. stPasteleria: Para operaciones de stage
2. dmPasteleria: Para data mart dimensional

Tablas Dimensiones

- Claves primarias en todas las dimensiones
- Atributos descriptivos para análisis
- Índices en columnas frecuentemente filtradas

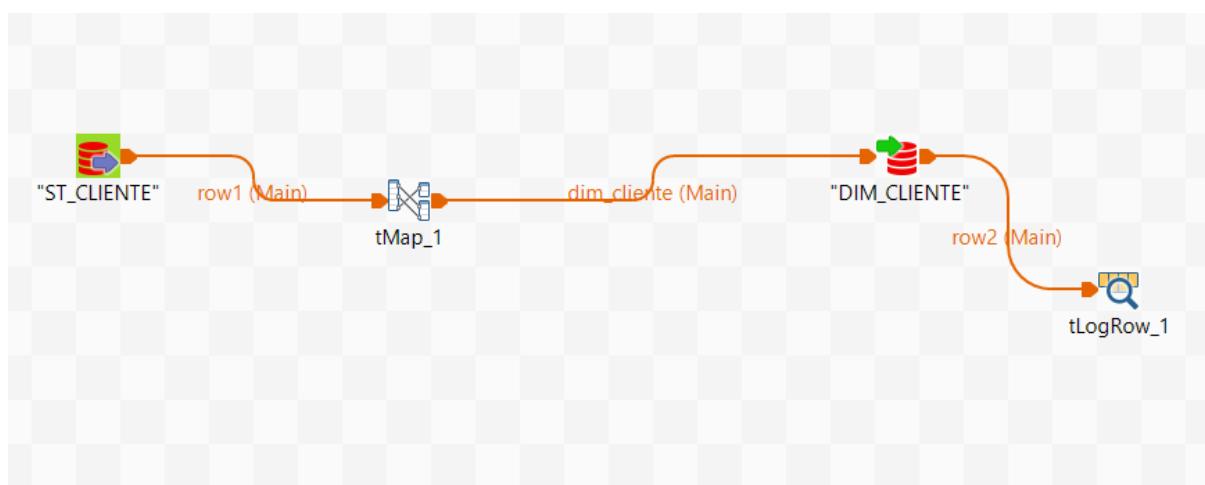
Tablas Hechos

- Claves foráneas a dimensiones
- Medidas con tipos de datos apropiados
- Particionamiento por tiempo para optimización

7. DISEÑO DE PROCESOS ETL

Dimension

Cliente:



Designer | Code

Trabajo(Job dim_cliente 0.1) Contexts(dim_cliente) Component Run (Trabajo dim_cliente)

Trabajo dim_cliente

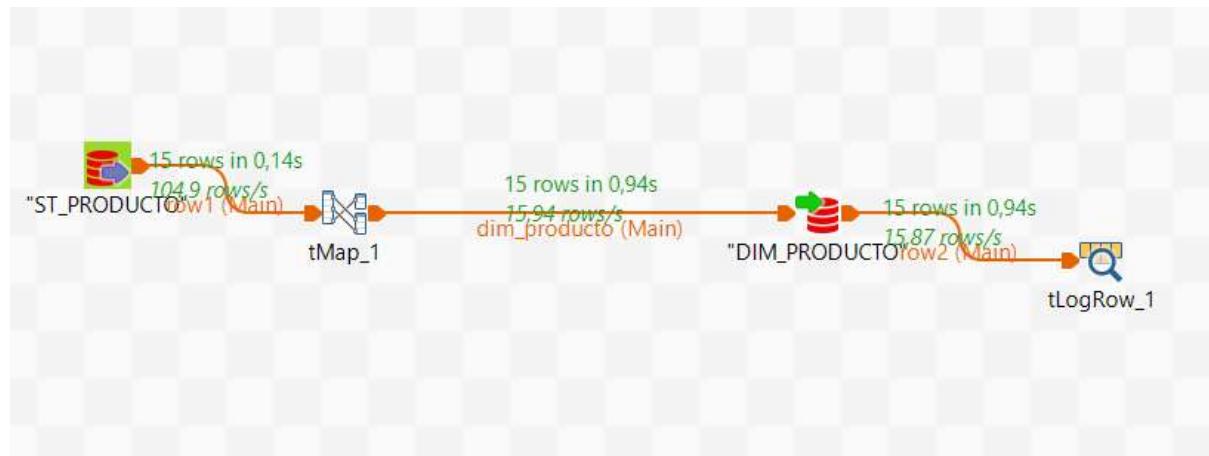
Basic Run

Execution

Run Kill Clear

```
Starting job dim_cliente at 12:43 11/12/2025.
[statistics] connecting to socket on port 3770
[statistics] connected
+-----+
| tLogRow_1 |
+-----+
| ID_CLIENTE | NOMBRE | TIPO_CLIENTE | CIUDAD |
+-----+
| 1 | Luis | Minorista | La Paz |
| 2 | Maria | Mayorista | Santa Cruz |
| 3 | Paola | Minorista | Cochabamba |
| 4 | Andres | Mayorista | La Paz |
| 5 | Jenny | Minorista | El Alto |
| 6 | Oscar | Minorista | Tarija |
| 7 | Miguel | Mayorista | Oruro |
| 8 | Claudia | Minorista | Pando |
| 9 | Sofia | Mayorista | Beni |
| 10 | Renato | Minorista | Sucre |
| 11 | Diego | Minorista | La Paz |
| 12 | Lucia | Mayorista | Cochabamba |
| 13 | Victor | Minorista | Santa Cruz |
| 14 | Laura | Mayorista | Tarija |
| 15 | Camila | Minorista | La Paz |
+-----+
[statistics] disconnected
Job dim_cliente ended at 12:43 11/12/2025. [Exit code = 0]
```

Dimension Producto



Execution

Run **Kill** **Clear**

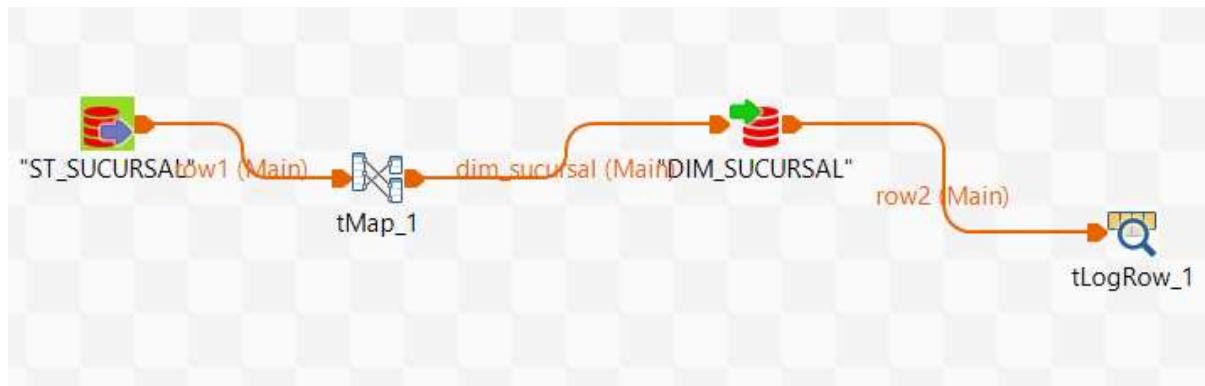
Starting job dim_producto at 12:45 11/12/2025.
[statistics] connecting to socket on port 3767
[statistics] connected

| ID_PRODUCTO | NOMBRE | CATEGORIA | SABOR | TAMANO |
|-------------|-------------------|-----------|-----------|---------|
| 1 | Torta Chocolate | Torta | Chocolate | Mediana |
| 2 | Cupcake Vainilla | Cupcake | Vainilla | PequeÑo |
| 3 | Cheesecake | Postre | Frutilla | Grande |
| 4 | Brownie | Postre | Chocolate | PequeÑo |
| 5 | Gelatina Fresa | Postre | Fresa | Mediana |
| 6 | Torta Tres Leches | Torta | Crema | Grande |
| 7 | Tarta Durazno | Tarta | Durazno | Mediana |
| 8 | Galletas Chispas | Galletas | Chocolate | PequeÑo |
| 9 | Roll Canela | Postre | Canela | Mediano |
| 10 | Cupcake Oreo | Cupcake | Chocolate | PequeÑo |
| 11 | Torta Selva Negra | Torta | Chocolate | Grande |
| 12 | Flan Vainilla | Postre | Vainilla | Mediano |
| 13 | Pie Manzana | Tarta | Manzana | Grande |
| 14 | Galletas Coco | Galletas | Coco | PequeÑo |
| 15 | Torta Zanahoria | Torta | Zanahoria | Mediana |

[statistics] disconnected

Job dim_producto ended at 12:45 11/12/2025. [Exit code = 0]

Dimension Sucursal



| tLogRow_1 | | |
|-------------|------------------------|------------|
| ID_SUCURSAL | NOMBRE | CIUDAD |
| 1 | Sucursal Central | La Paz |
| 2 | Sucursal Sur | Santa Cruz |
| 3 | Sucursal Norte | Cochabamba |
| 4 | Sucursal Este | El Alto |
| 5 | Sucursal Oeste | Tarija |
| 6 | Sucursal Andina | Oruro |
| 7 | Sucursal Tropical | Beni |
| 8 | Sucursal AmazÃ³nica | Pando |
| 9 | Sucursal Real | Sucre |
| 10 | Sucursal Metropolitana | La Paz |

[statistics] disconnected

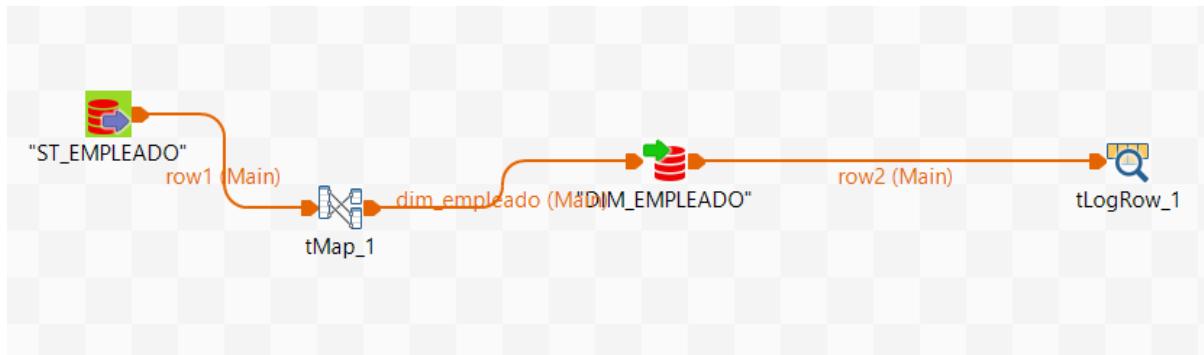
Job dim_sucursal ended at 12:46 11/12/2025. [Exit code = 0]

Dimension Tiempo



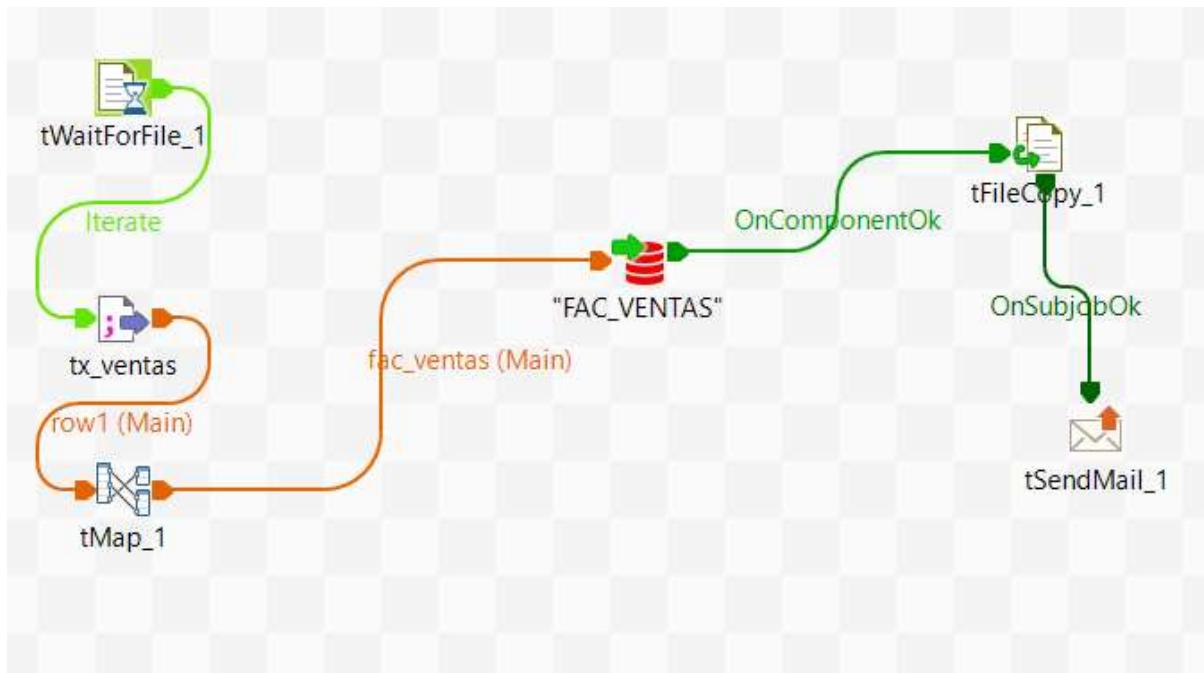
| tLogRow_1 | | | |
|-----------|-----|-----|------|
| ID TIEMPO | DIA | MES | ANIO |
| 1 | 01 | 01 | 2019 |
| 2 | 15 | 03 | 2019 |
| 3 | 22 | 05 | 2019 |
| 4 | 10 | 08 | 2019 |
| 5 | 12 | 10 | 2019 |
| 6 | 03 | 12 | 2019 |
| 7 | 07 | 02 | 2020 |
| 8 | 14 | 04 | 2020 |
| 9 | 29 | 07 | 2020 |
| 10 | 05 | 11 | 2020 |
| 11 | 06 | 01 | 2021 |
| 12 | 18 | 03 | 2021 |
| 13 | 25 | 06 | 2021 |
| 14 | 10 | 09 | 2021 |
| 15 | 22 | 12 | 2021 |

Dimension Empleado



| tLogRow_1 | | | | |
|-------------|----------|------------|------------|--------|
| ID_EMPLEADO | NOMBRE | CARGO | ANTIGUEDAD | TURNO |
| 1 | Carlos | Pastelero | 5 aÑos | Mañana |
| 2 | Ana | Decoradora | 2 aÑos | Tarde |
| 3 | Luis | Pastelero | 7 aÑos | Noche |
| 4 | Marta | Ayudante | 1 aÑo | Mañana |
| 5 | Cesar | Cajero | 3 aÑos | Tarde |
| 6 | Rosa | Decoradora | 4 aÑos | Mañana |
| 7 | David | Logística | 5 aÑos | Noche |
| 8 | Elena | Pastelera | 2 aÑos | Mañana |
| 9 | Mario | Panadero | 8 aÑos | Noche |
| 10 | Julia | Cajera | 1 aÑo | Tarde |
| 11 | Patricia | Pastelera | 6 aÑos | Mañana |
| 12 | Pedro | Decorador | 3 aÑos | Noche |
| 13 | Hector | Ayudante | 2 aÑos | Tarde |
| 14 | Natalia | Pastelera | 1 aÑo | Mañana |
| 15 | Rodrigo | Panadero | 10 aÑos | Noche |

Fact Ventas





vivianvillarroel5... 12:51
para mí, carbon ▾



Parece que este mensaje está en
inglés
[Traducir al español](#)



Cargado correctamente



vivianvillarroel5... 12:51
para mí, carbon ▾



Parece que este mensaje está en
inglés
[Traducir al español](#)



Cargado correctamente



vivianvillarroel5... 12:51
para mí, carbon ▾



Parece que este mensaje está en
inglés
[Traducir al español](#)



Cargado correctamente

Fact Produccion



vivianvillarroel5... 12:54

para mí, carbon ▾



Parece que este mensaje está en
inglés

[Traducir al español](#)



Archivo cargado correctamente



vivianvillarroel5... 12:54

para mí, carbon ▾



Parece que este mensaje está en
inglés

[Traducir al español](#)



Archivo cargado correctamente



vivianvillarroel5... 12:54

para mí, carbon ▾



Parece que este mensaje está en
inglés

[Traducir al español](#)

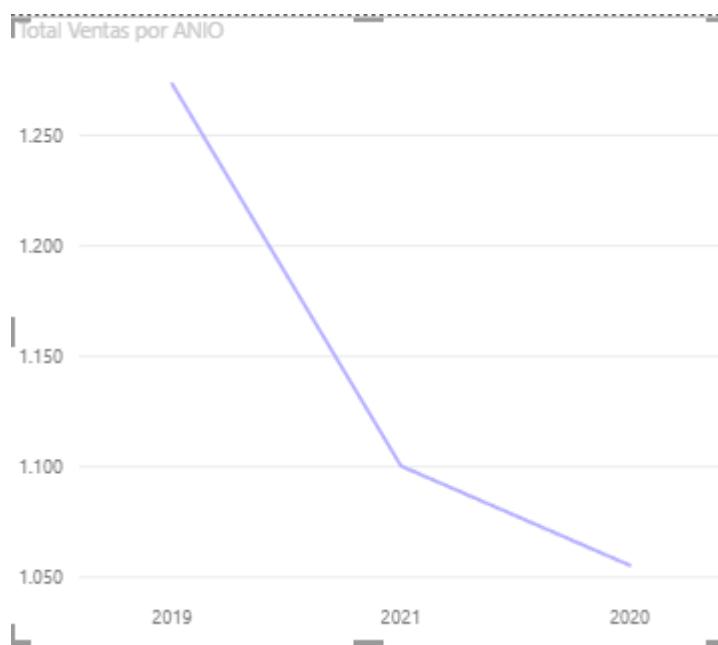


8. DESARROLLO DE HERRAMIENTA DE EXPLORACIÓN

8.1 Dashboards en Power BI

Dashboard de Ventas - Gráficos Implementados

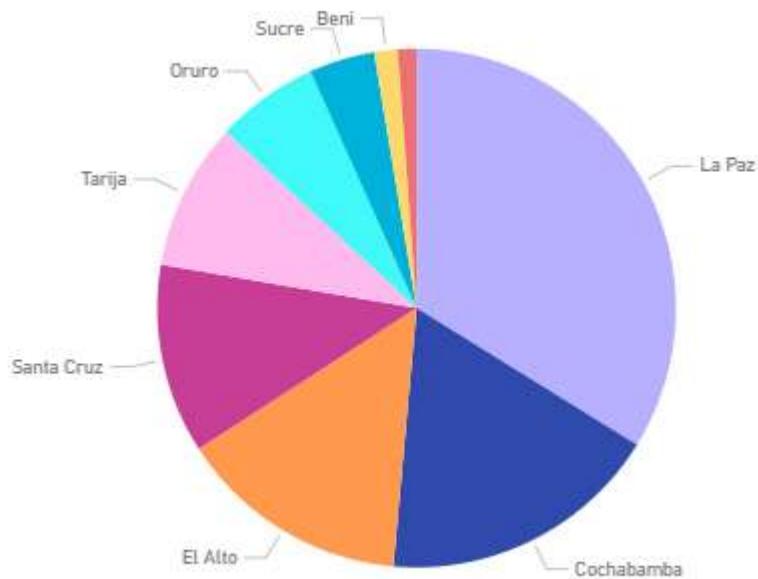
1. Total Ventas por Año



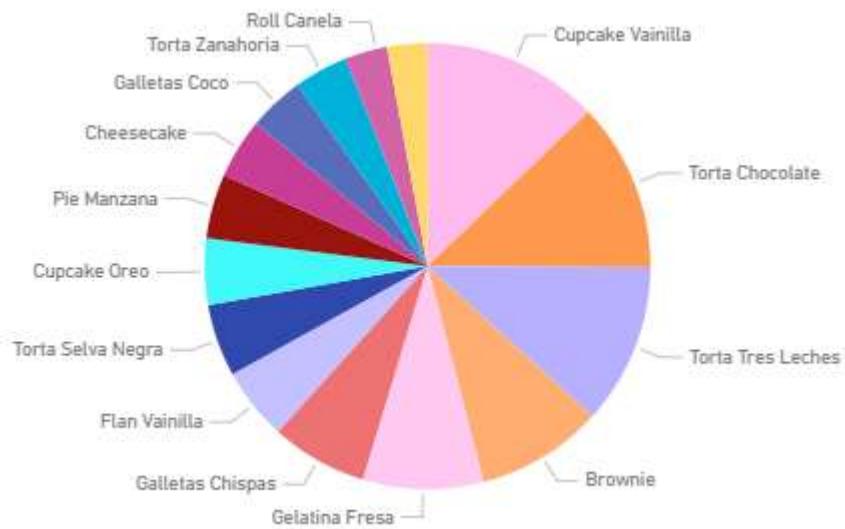
2. Total Ventas por Tipo Cliente



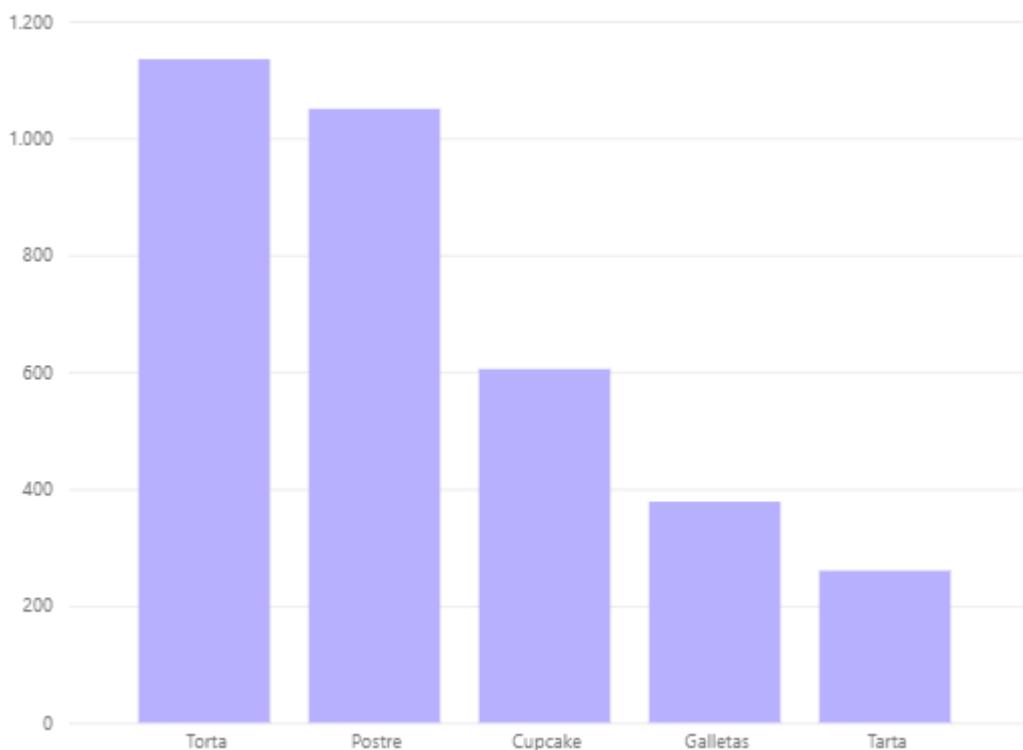
3 Total Ventas por Ciudad



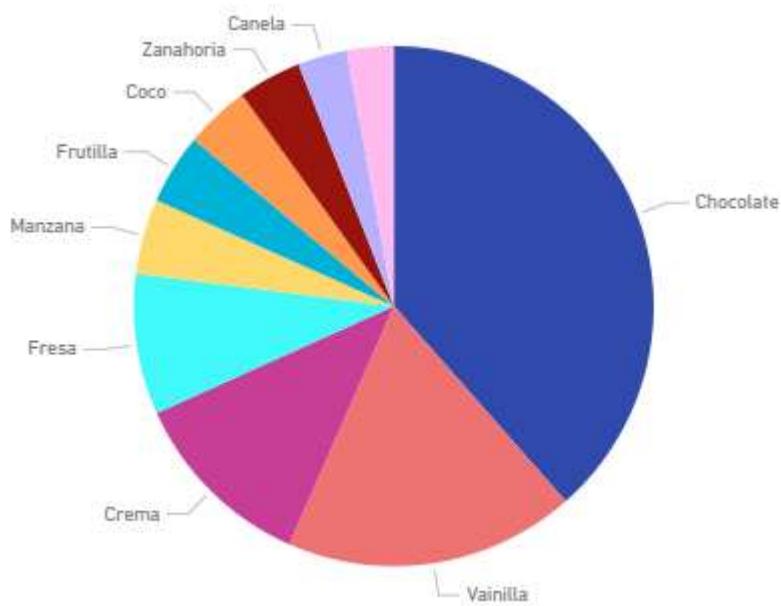
4 Total Ventas por Productos



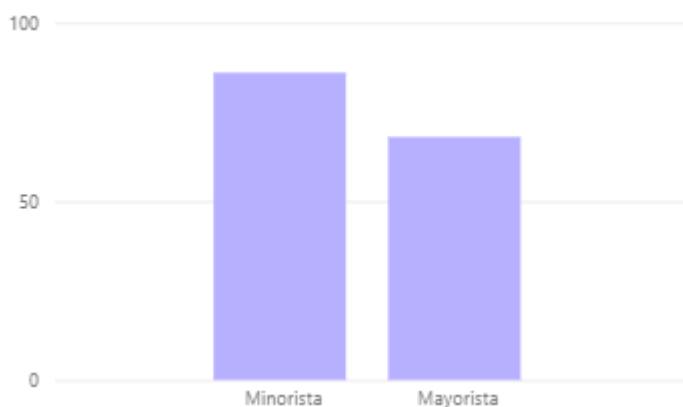
5. Total Ventas por Categoría



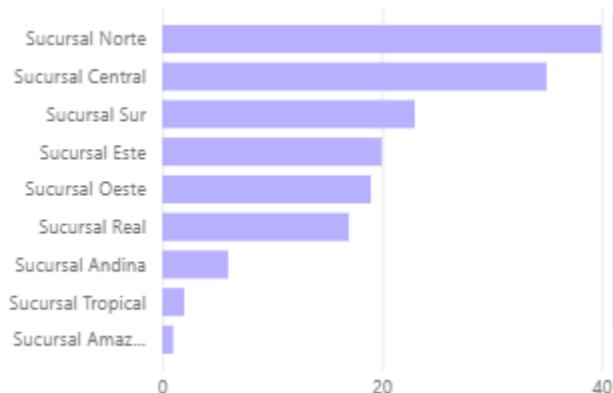
6. Total Ventas por Sabor



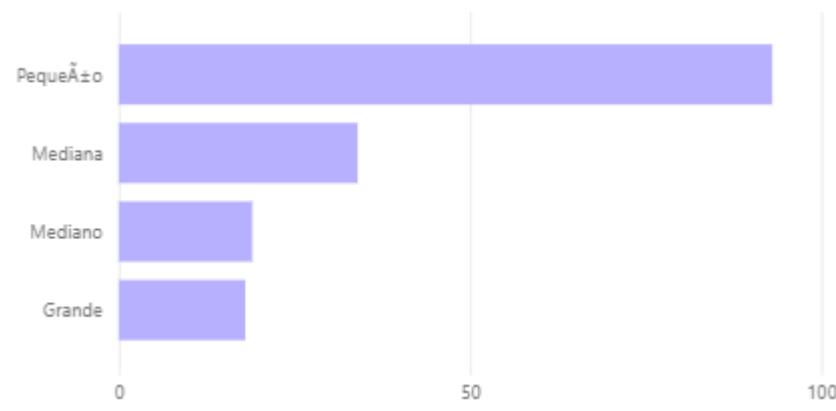
7. Descuento Total por Tipo de Cliente



8. Cantidad Vendida por Sucursal

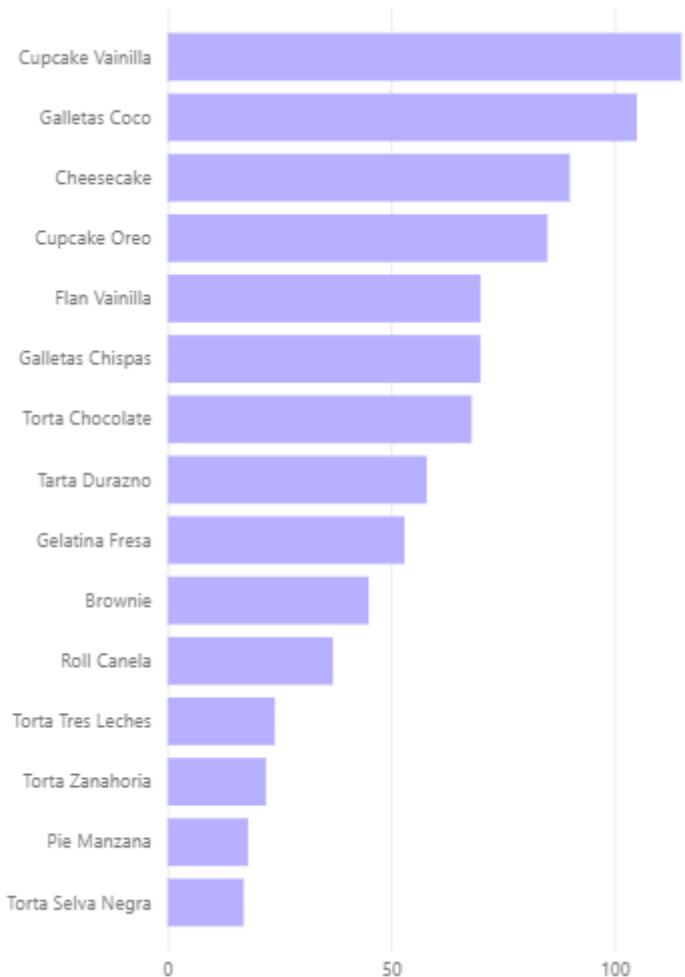


9. Cantidad Vendida por Tamaño

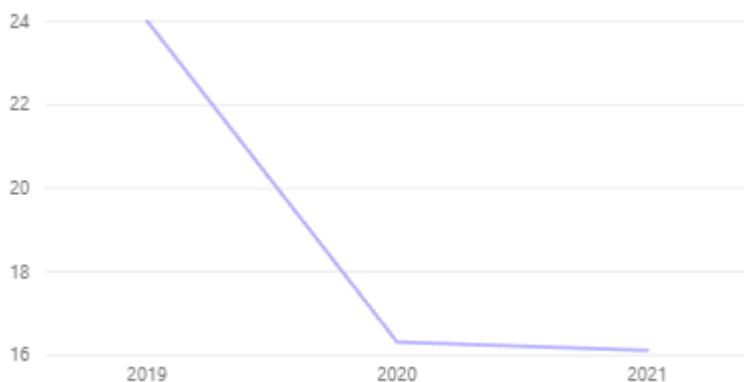


Dashboard de Producción - Gráficos Implementados

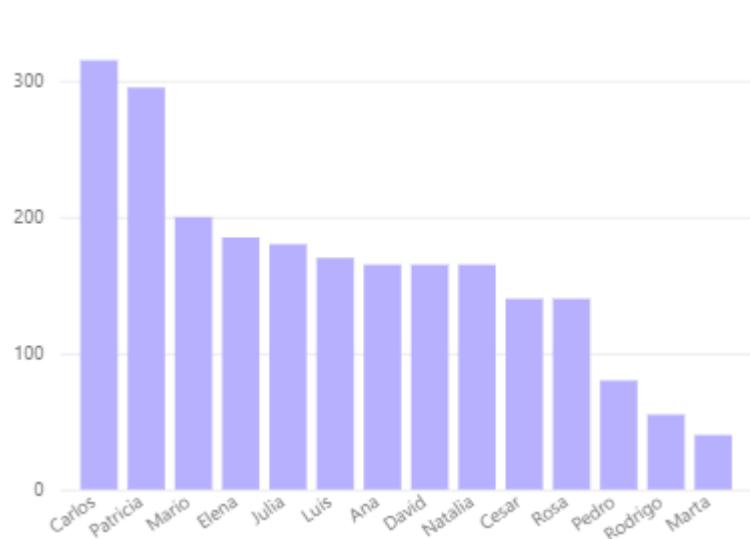
1 Cantidad Producida por Producto



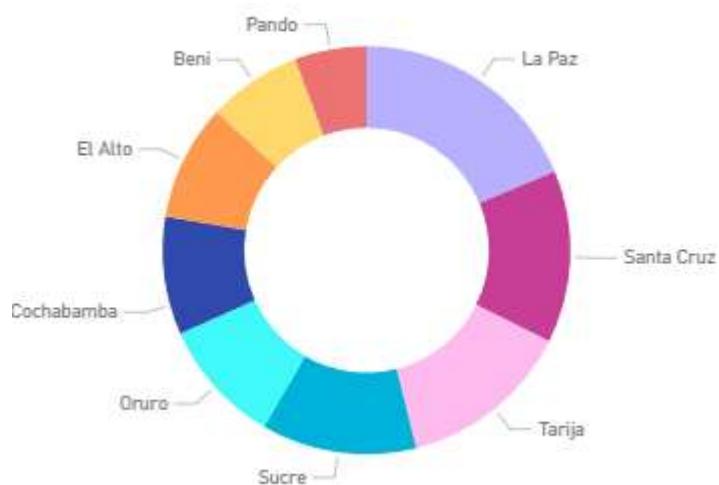
2 Merma Total por Año



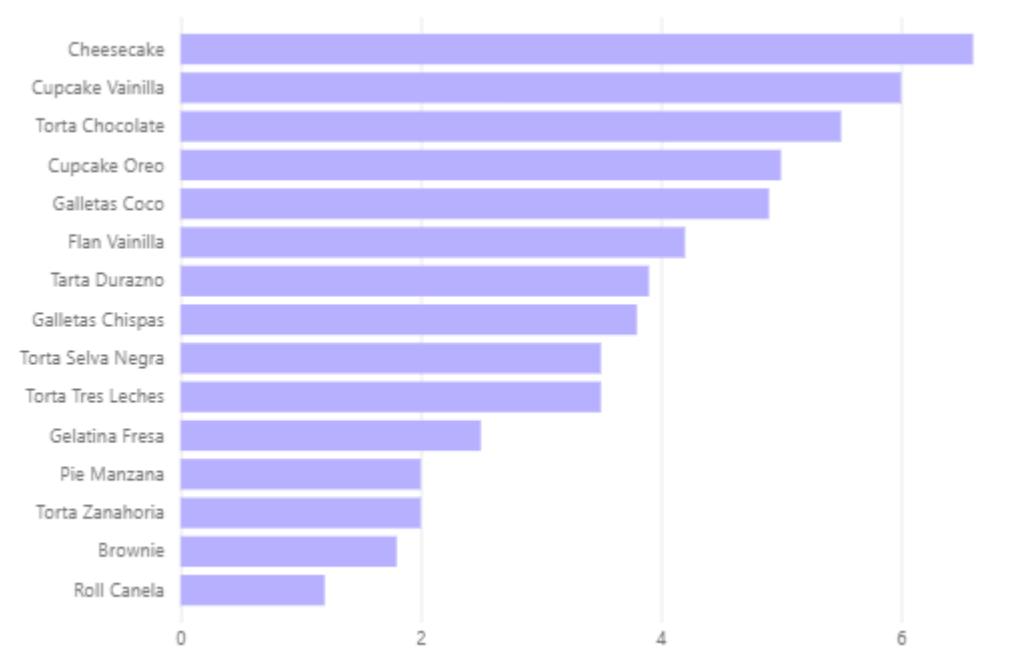
3 Tiempo Total de Producción por Empleado



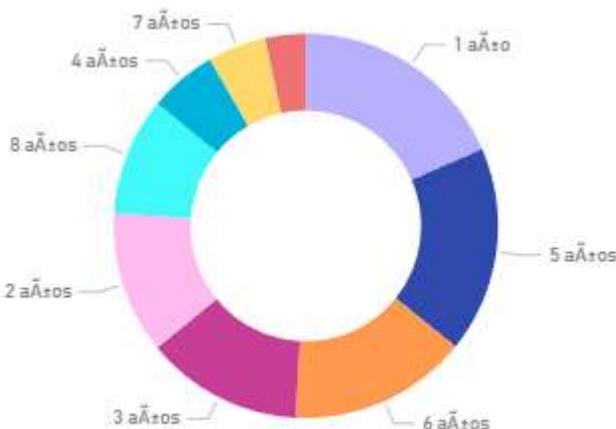
4 Cantidad Producida por Ciudad



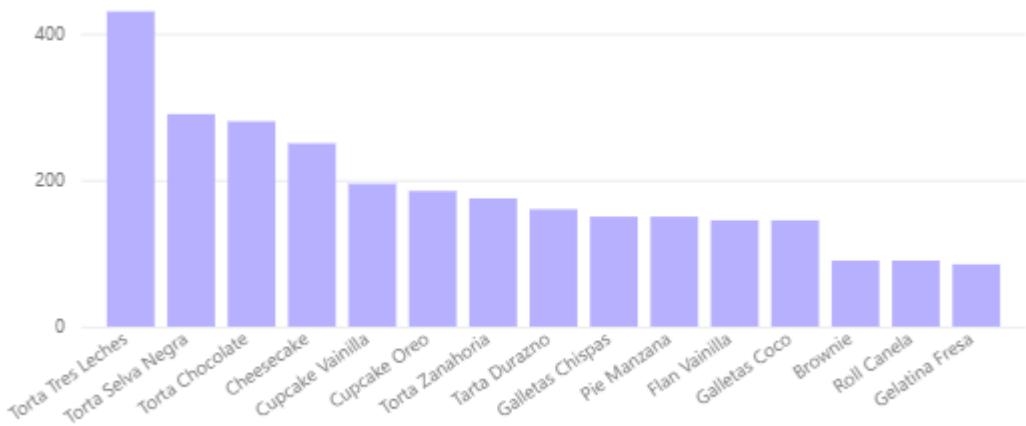
5. Merma Total por Producto



6. Total Producción por Antigüedad de Empleados



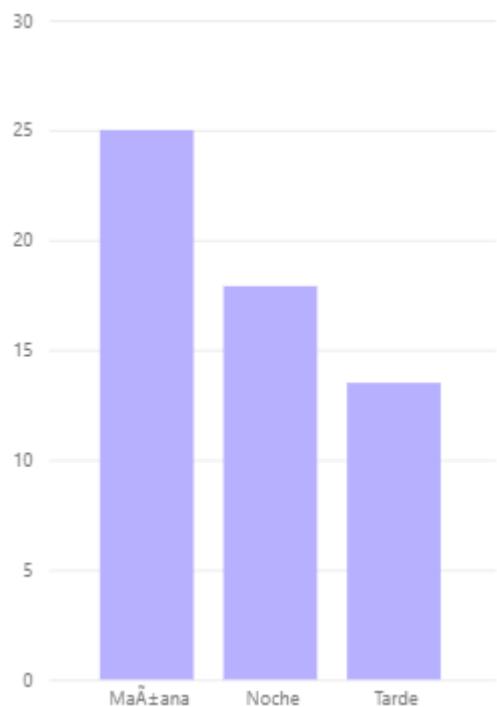
7 Costo Materia Prima por Producto



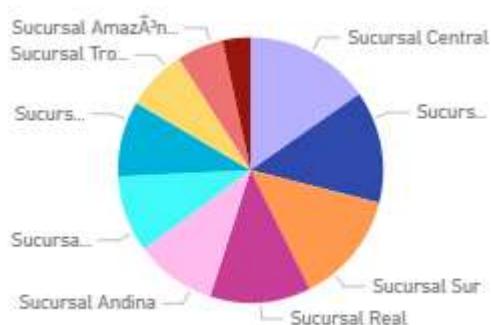
8 Cantidad Producida por Turno



9 Merma Total por Turno



10 Cantidad Producida por Sucursal



8.2 Medidas DAX Implementadas

Principales Medidas

- Total de ventas = $\text{SUM}(\text{fac_ventas}[\text{total_venta}])$
- Cantidad Vendida = $\text{SUM}(\text{fac_ventas}[\text{cantidad}])$
- Descuento Total = $\text{SUM}(\text{fac_ventas}[\text{descuento}])$

- Cantidad Producida = SUM(fac_produccion[cantidad_prod])
- Merma Total = SUM(fac_produccion[merma])
- Tiempo Total Producción = SUM(fac_produccion[tiempo_produccion])
- Costo Materia Prima = SUM(fac_produccion[costo_mp])

9. CONCLUSIONES Y RECOMENDACIONES

9.1 Conclusiones del Proyecto

Logros Alcanzados

- 1 Se implementó exitosamente un modelo en estrella con 2 tablas de hechos y 5 dimensiones, cumpliendo con los requisitos académicos.
- 2 Se desarrollaron procesos ETL completos que manejan extracción, transformación y carga de datos de manera eficiente.
- 3 Se crearon dashboards en Power BI con los 19 gráficos solicitados, organizados en ventas y producción.
- 4 Documentación Completa: Se elaboró el documento requerido con los 9 puntos de la metodología Kimball.

Aprendizajes Obtenidos

- Importancia del diseño dimensional correcto
- Complejidad de los procesos ETL en escenarios reales
- Valor de la visualización efectiva de datos
- Necesidad de documentación detallada

9.2 Recomendaciones

Para Mejoras Técnicas

- 1 Implementar particionamiento más granular en tablas de hechos
- 2 Agregar más validaciones de calidad de datos en ETL
- 3 Crear versiones móviles de los dashboards
- 4 Implementar alertas proactivas para métricas críticas

Para Expansión Funcional

- 1 Agregar análisis de rentabilidad por producto
- 2 Incorporar pronósticos de ventas y producción
- 3 Añadir análisis de estacionalidad detallado
- 4 Incluir benchmarking con estándares de industria