# Informe Final: Sistema Data Warehouse Kimball

## Resumen Ejecutivo

Este proyecto tiene como objetivo construir un sistema de análisis de ventas basado en un Data Warehouse con modelo Kimball. Se integró un scraper para obtener información real de productos en tiendas online y se simularon usuarios y ventas para alimentar el sistema.

Se logró un flujo completamente automatizado de generación de datos, con tiempos de respuesta óptimos y un diseño de base de datos enfocado en consultas eficientes. El sistema ofrece una base sólida para el análisis de comportamiento de clientes, efectividad de promociones y rendimiento de productos.

## Arquitectura del Sistema

El sistema está dividido en tres grandes componentes:

- Scraper de tiendas online (Yapo, MercadoLibre, Paris, Falabella).

- Generador de usuarios y simulador de ventas realistas.

- Data Warehouse en PostgreSQL utilizando el modelo estrella.

Cada parte se conecta a través de scripts organizados que permiten la ejecución de todo el proceso desde una interfaz por consola.

## Modelo de Datos Kimball

La base de datos se estructura en un modelo estrella con una tabla de hechos (ventas) y cinco tablas de dimensiones: usuarios, productos, tiendas, promociones y tiempo.

Esto permite consultas rápidas y análisis multidimensionales, facilitando operaciones como segmentación de clientes, comparación entre tiendas, y evaluación del impacto de promociones.

## Funcionalidades Clave

- Scraping automatizado con detección de promociones.

- Generación de usuarios con datos realistas.

- Simulación de ventas basada en rangos etarios y categorías.

- Exportación de datos en CSV y Excel.

- Carga automática de datos a PostgreSQL.

- Consultas SQL avanzadas preconfiguradas.

## Consultas y Análisis

El sistema incluye ejemplos de análisis como:

- Segmentación por edad y género.

- Ventas por trimestre.

- Ranking de productos por tienda.

- Evaluación de descuentos aplicados.

Todos los datos están estructurados para realizar análisis en tiempo real de manera eficiente.

# README - Modo de Uso

## Instalación y Requisitos

- Python 3.8+

- PostgreSQL 12+

- Conexión a Internet

Instala dependencias con:

pip install -r requirements.txt

Configura el archivo `.env` con los datos de tu base de datos:

DB\_USER=...

DB\_PASSWORD=...

DB\_NAME=...

DB\_HOST=...

## Ejecución del Sistema

Desde consola, ejecuta:

python main.py

Menú disponible:

1. Generar usuarios

2. Scraping de productos

3. Scraper completo (usuarios + productos + ventas)

4. Cargar archivo CSV a PostgreSQL

5. Ver archivos generados

6. Consultar tablas en la base de datos

## Flujo Recomendado

1. Usar opción 3: Scraper completo.

2. Usar opción 4: Cargar archivo completo generado.

3. Consultar los datos (opción 6 o usando pgAdmin/SQL).

## Archivos Clave

- `main.py`: Menú principal.

- `scrapy\_integrado.py`: Scraping completo.

- `models.py`: Definición del modelo de datos.

- `cargar\_csv.py`: Carga de datos a PostgreSQL.

- `archivos/`: Carpeta con los CSV y Excel generados.