**Documentación del Proyecto: ETL y Análisis en Power BI**

**Introducción**

Este documento detalla el proceso de trabajo para la transformación y análisis de datos en el proyecto final. Tras haber conversado con Silvia y obtener su aprobación, se ha definido la siguiente estrategia de trabajo.

Estrategia de Trabajo

El proyecto se divide en tres fases principales:

**1. ETL en Python**

El primer paso consiste en la Extracción, Transformación y Carga (ETL) de los datos utilizando Python. Se trabajará con siete bases de datos en formato CSV, que serán limpiadas y transformadas antes de ser importadas en Power BI.

**Bases de Datos en el Proyecto:**

**📂 Fuente de los datos: Los datos han sido extraídos del siguiente enlace:** [**Grocery Sales Dataset - Kaggle**](https://www.kaggle.com/datasets/andrexibiza/grocery-sales-dataset?select=customers.csv)

**Diccionario de Datos**

A continuación, se presenta un listado de las columnas de cada archivo y su significado. Se han marcado con negrita las claves primarias:

| Archivo | Columna | Descripción |
| --- | --- | --- |
| categories.csv | CategoryID | Identificador único de la categoría de producto. |
|  | CategoryName | Nombre de la categoría del producto. |
| cities.csv | CityID | Identificador único de la ciudad. |
|  | CityName | Nombre de la ciudad. |
|  | Zipcode | Código postal de la ciudad. |
|  | CountryID | Identificador del país al que pertenece la ciudad. |
| countries.csv | CountryID | Identificador único del país. |
|  | CountryName | Nombre del país. |
|  | CountryCode | Código del país en formato de dos letras. |
| customers.csv | CustomerID | Identificador único del cliente. |
|  | FirstName | Nombre del cliente. |
|  | LastName | Apellido del cliente. |
|  | CityID | Identificador de la ciudad del cliente. |
|  | Address | Dirección del cliente. |
| employees.csv | EmployeeID | Identificador único del empleado. |
|  | FirstName | Nombre del empleado. |
|  | LastName | Apellido del empleado. |
|  | BirthDate | Fecha de nacimiento del empleado. |
|  | Gender | Género del empleado. |
|  | CityID | Identificador de la ciudad del empleado. |
|  | HireDate | Fecha de contratación del empleado. |
| products.csv | ProductID | Identificador único del producto. |
|  | ProductName | Nombre del producto. |
|  | Price | Precio del producto. |
|  | CategoryID | Identificador de la categoría del producto. |
|  | Class | Clasificación del producto. |
|  | ModifyDate | Fecha de la última modificación. |
| sales.csv | SalesID | Identificador único de la venta. |
|  | SalesPersonID | Identificador del empleado que realizó la venta. |
|  | CustomerID | Identificador del cliente que realizó la compra. |
|  | ProductID | Identificador del producto vendido. |
|  | Quantity | Cantidad de unidades vendidas. |
|  | Discount | Descuento aplicado a la venta. |
|  | TotalPrice | Precio total de la venta después del descuento. |
|  | SalesDate | Fecha y hora de la venta. |
|  | TransactionNumber | Número de transacción único asociado a la venta. |

**2. Transformación del Modelo de Datos (Copo de Nieve → Estrella)**

Actualmente, los datos están organizados en un modelo en copo de nieve, lo que implica una mayor cantidad de relaciones y normalización en las tablas.

El objetivo es transformar este esquema en un modelo en estrella, que facilite el análisis y optimice el rendimiento en Power BI.

**Acciones clave en esta fase:**

* Definir la tabla de hechos (ventas).
* Definir y estructurar las tablas de dimensiones (clientes, productos, empleados, etc.).
* Crear relaciones optimizadas en Power BI.

**3. Análisis de Datos y Creación de Dashboards en Power BI**

Una vez completada la ETL y la transformación del modelo de datos, se procederá con el análisis exploratorio (EDA) y la creación de dashboards en Power BI.

Objetivos de esta fase:

* Importar los datos limpios en Power BI.
* Realizar un análisis exploratorio de los datos para identificar tendencias y patrones.
* Crear dashboards interactivos para visualizar los insights más relevantes.

Este documento se actualizará a medida que avance el proyecto.