**Prueba técnica - Ingeniero de datos**

Fecha: 14 de Noviembre de 2024

Nombre: Vladimir Ortega Botero

Cédula: 1037589731

1. **Introduccion**

En el presente informe se detalla la implementación de una aplicación de ETL (Extracción, Transformación y Carga de datos) en Python, diseñada para cargar datos sobre la serie Los Simpsons en una base de datos MySQL. Se utilizó Programación Orientada a Objetos (POO) para modularizar el código y facilitar la escalabilidad, además de incorporar observabilidad mediante registros de eventos o logs. La aplicación emplea PySpark, un potente framework de procesamiento de datos, para manejar grandes volúmenes de datos y realizar transformaciones eficientes. Esta arquitectura POO permite además extender y modificar la aplicación con facilidad.

1. **Analisis Exploratorio de los datos**

Se realizó un análisis exploratorio de los datos de un sistema de alquiler de películas. La base de datos contiene información relacionada con tiendas, clientes, inventarios, alquileres, y películas. Las siguientes observaciones resumen los datos y estadísticas clave de cada tabla analizada:

1. **Tabla store:**
   * Contiene detalles de las tiendas, como la dirección y el ID del gerente de cada una.
   * **Observación:** La cantidad de tiendas es limitada (solo dos), lo cual permite un análisis detallado de cada una en términos de rendimiento y ubicación.
2. **Tabla customer:**
   * Contiene información de cada cliente, como su nombre, dirección, fecha de creación y estado de actividad.
   * **Observación:** La tabla revela la frecuencia de actualización de datos (última actualización) y muestra clientes inactivos y activos, lo cual es clave para analizar la retención de clientes y evaluar la base de clientes leales.
3. **Tabla rental:**
   * Registra detalles sobre cada alquiler, como fecha de alquiler, fecha de devolución, y los IDs de inventario y cliente.
   * **Observación:** La tabla permite analizar patrones de alquiler (días de mayor demanda) y tendencias en la devolución de películas. Las relaciones con *customer* e *inventory* facilitan la obtención de perfiles de cliente y tendencias de popularidad de inventario.
4. **Tabla inventory:**
   * Contiene información sobre la disponibilidad de cada película en cada tienda.
   * **Observación:** Esta tabla es esencial para calcular la rotación de inventario y la disponibilidad de títulos, lo cual puede influir en el tiempo de espera y satisfacción del cliente.
5. **Tabla film:**
   * Detalla cada película, incluyendo el título, duración de alquiler, costo de reemplazo y características especiales.
   * **Observación:** La información sobre costos y características especiales permite clasificar las películas y entender la disposición de los clientes a pagar por características adicionales (ej. escenas eliminadas, comentarios).

**2. Preguntas de Negocio Propuestas**

A partir del análisis exploratorio, se han formulado cinco preguntas clave de negocio a las cuales los datos procesados pueden dar respuesta:

1. **¿Cuál es el perfil de los clientes más rentables y cómo podemos fidelizarlos?**
   1. Los datos de *customer* y *rental* permiten identificar a los clientes frecuentes y aquellos con un gasto elevado. Este perfil ayudaría a diseñar programas de fidelización específicos.
2. **¿Qué factores influyen en la rotación del inventario y cómo se puede optimizar?**
   1. Utilizando *inventory* y *rental*, es posible observar patrones de demanda en cada tienda. Esta información ayudaría a ajustar la cantidad de copias de películas en inventario y optimizar los costos.
3. **¿Cuáles son los días y horas de mayor actividad en las tiendas?**
   1. Analizando los *rental\_date* y *return\_date* en la tabla *rental*, se puede identificar los momentos de mayor afluencia, lo cual es útil para ajustar la cantidad de personal y optimizar el servicio.
4. **¿Existen categorías o características de películas que sean más populares?**
   1. Al observar las características especiales y el rating en la tabla film, se puede entender cuáles son las preferencias de los clientes y ajustar el inventario en consecuencia, destacando las categorías más demandadas.
5. **¿Cuál es el índice de inactividad de los clientes y qué estrategias se pueden implementar para reducirlo?**
   1. La tabla *customer* permite identificar clientes inactivos, lo cual podría dar paso a estrategias de retención y reactivación de clientes, como promociones para clientes inactivos.
6. **Conclusiones**

El análisis exploratorio de los datos ha revelado información valiosa sobre la operación y la demanda de un sistema de alquiler de películas. Las tablas analizadas permiten responder a preguntas críticas de negocio que abarcan desde la optimización de inventario hasta la fidelización de clientes. Por ejemplo:

* **Optimización del Inventario:** Se han identificado patrones de demanda que podrían optimizarse, ajustando la cantidad de copias de películas populares y minimizando la de aquellas menos rentadas.
* **Fidelización de Clientes:** Existe una oportunidad significativa para identificar y retener a los clientes de mayor rentabilidad mediante programas de fidelización.
* **Ajustes Operativos:** La identificación de días y horas pico facilita decisiones estratégicas en la asignación de personal y recursos, mejorando la eficiencia operativa de las tiendas.
* **Preferencias de los Clientes:** Las características de las películas (ej. escenas adicionales, comentarios) influyen en la preferencia de los clientes, lo cual puede orientar futuras adquisiciones de títulos en inventario.