

HIGH LEVEL DOCUMENT

Overview

This repository presents a detailed case study on retail data analysis, implemented entirely using SQL. The project demonstrates a systematic approach to transforming raw transactional data into actionable business intelligence, covering everything from data integrity to advanced customer behavior patterns. It's designed to showcase robust SQL skills for data cleaning, exploratory analysis, and deriving strategic insights for retail operations.

Technologies Used

- **Database:** Assumed to be a relational database (queries written with MySQL syntax, e.g., DATEDIFF, DATE_FORMAT, YEAR, HOUR, DAYNAME).
- **Language:** SQL

Objectives

- To utilize SQL queries for data cleaning and exploratory data analysis to ensure data quality and gain initial insights.
- To identify high and low sales products to optimize inventory and tailor marketing efforts.
- To segment customers based on their purchasing behavior for targeted marketing campaigns.
- Create Customer segments

Total Number of Orders Customer Segment

0	No orders
1 - 10	Low
10 - 30	Mid
> 30	High Value

Datasets

- **Sales Transactions Dataset:** Records of sales transactions, including transaction ID, customer ID, product ID, quantity purchased, transaction date, and price.

- **Customer Profiles Dataset:** Information on customers, including customer ID, age, gender, location, and join date.

- **Product Inventory Dataset:** Data on product inventory, including product ID, product name, category, stock level, and price.

Data Cleaning

Remove duplicate transactions and correct any discrepancies in product prices between sales transactions and product inventory. Also, take care of null values.

Exploratory Data Analysis (EDA)

Perform basic product performance overview, customer purchase frequency analysis, and product categories performance evaluation.