**Retail Sales Analysis SQL Project**

**Project Overview**

**Project Title**: Retail Sales Analysis

This project is designed to demonstrate SQL skills and techniques typically used by data analysts to explore, clean, and analyze retail sales data. The project involves setting up a retail sales database, performing exploratory data analysis (EDA), and answering specific business questions through SQL queries. This project is ideal for those who are starting their journey in data analysis and want to build a solid foundation in SQL.

**Objectives**

1. **Set up a retail sales database**: Create and populate a retail sales database with the provided sales data.
2. **Data Cleaning**: Identify and remove any records with missing or null values.
3. **Exploratory Data Analysis (EDA)**: Perform basic exploratory data analysis to understand the dataset.
4. **Business Analysis**: Use SQL to answer specific business questions and derive insights from the sales data.

**Project Structure**

**1. Database Setup**

* **Database Creation**: The project starts by creating a database named p1\_retail\_db.
* **Table Creation**: A table named retail\_sales is created to store the sales data. The table structure includes columns for transaction ID, sale date, sale time, customer ID, gender, age, product category, quantity sold, price per unit, cost of goods sold (COGS), and total sale amount.

**2. Data Exploration & Cleaning**

* **Record Count**: Determine the total number of records in the dataset.
* **Customer Count**: Find out how many unique customers are in the dataset.
* **Category Count**: Identify all unique product categories in the dataset.
* **Null Value Check**: Check for any null values in the dataset and delete records with missing data.

**Business Questions Addressed**

The following business questions were answered using SQL queries:

1. Retrieve all columns for sales made on '2022-11-05'.
2. Retrieve all transactions where the category is 'Clothing' and the quantity sold is more than 4 in the month of Nov-2022.
3. Calculate the total sales (total\_sale) for each category.
4. Find the average age of customers who purchased items from the 'Beauty' category.
5. Find all transactions where the total\_sale is greater than 1000.
6. Find the total number of transactions made by each gender in each category.
7. Calculate the average sale for each month and find the best-selling month for each year.
8. Find the top 5 customers based on the highest total sales.
9. Find the number of unique customers who purchased items from each category.
10. Determine the number of orders for each shift (Morning, Afternoon, Evening).

**Conclusion**

This project serves as a comprehensive introduction to SQL for data analysts, covering database setup, data cleaning, exploratory data analysis, and business-driven SQL queries. The findings from this project can help drive business decisions by understanding sales patterns, customer behavior, and product performance.

**Acknowledgments**

The dataset used for this analysis was provided by the YouTube channel [**Zero Analyst**](https://www.youtube.com/@zero_analyst) from one of their project videos. While the dataset originates from this source, all SQL queries for data cleaning, exploration, and analysis were written independently for this project.

**Author**

* **Parveen Jalwal**
* **LinkedIn**: <https://www.linkedin.com/in/parveen-jalwal-201a2a302>