Retail Sales Data Analysis using SQL

Project Overview

This project analyzes retail sales data using SQL. It includes database creation, data cleaning, exploration, and analysis to generate meaningful business insights.

1. Database Setup

```
CREATE DATABASE project;
SHOW DATABASES;
USE project;

SELECT * FROM project.retail_sales;
SELECT * FROM project.retail_sales LIMIT 10;
SELECT COUNT(*) FROM project.retail_sales;
```

2. Data Cleaning

```
DELETE FROM project.retail_sales
WHERE transactions_id IS NULL
   OR sale_date IS NULL
   OR sale_time IS NULL
   OR gender IS NULL
   OR category IS NULL
   OR quantity IS NULL
   OR cogs IS NULL
   OR total_sale IS NULL;
```

3. Data Analysis Queries

```
1. Total Sales
SELECT COUNT(*) AS total_sales FROM retail_sales;

2. Unique Customers
SELECT COUNT(DISTINCT customer_id) FROM retail_sales;

3. Sales on specific date
SELECT * FROM retail_sales WHERE sale_date='2022-11-05';

4. Total sales per category
SELECT category, SUM(total_sale)
FROM retail_sales GROUP BY category;
```

```
5. Clothing quantity >3 in Nov 2022
SELECT * FROM retail_sales
WHERE category='Clothing'
AND quantity>3
AND sale_date BETWEEN '2022-11-01' AND '2022-11-30';

6. Average age (Beauty)
SELECT ROUND(AVG(age),2)
FROM retail_sales WHERE category='Beauty';

7. Transactions above 1000
SELECT * FROM retail_sales WHERE total_sale>1000;

8. Transactions by gender & category
SELECT category, gender, COUNT(*)
FROM retail_sales GROUP BY category, gender;

9. Top 5 customers
SELECT customer_id, SUM(total_sale)
FROM retail_sales
GROUP BY customer_id
ORDER BY SUM(total_sale) DESC LIMIT 5;

10. Unique customers per category
SELECT category, COUNT(DISTINCT customer_id)
FROM retail_sales GROUP BY category;
```