**Data Spark: Illuminating Insights for Global Electronics**

# 1. Introduction

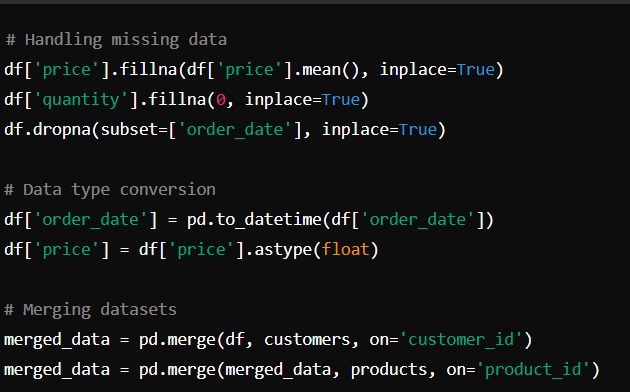
This project involves building an interactive sales dashboard using Streamlit, SQL, and Python. The goal is to extract, clean, and analyze sales data to gain insights into business performance, including sales trends, product performance, and customer behavior. The final deliverable is a Streamlit-based web app that connects to an SQL database and provides interactive visualizations of key business metrics.

#### ****2. Objectives****

* **Data Cleaning and Preparation**: Identify and handle missing data, ensure proper data types, and merge different datasets (sales, customers, products) for comprehensive analysis.
* **SQL Data Loading**: Preprocessed data will be loaded into an SQL database, ensuring a well-structured relational schema.
* **Interactive Visualization**: Build a Streamlit web app that connects to the SQL database and displays interactive dashboards.
* **SQL Queries for Insights**: Develop a set of SQL queries that extract important insights from the data, addressing key business questions

#### ****3. Data Cleaning and Preparation****

* **Datasets Used**:
  + sales csv : containing details of transactions (sales, prices, quantities, dates).
  + Customers csv : containing customer details (e.g., region, customer ID).
  + Product .csv : containing product details (e.g., product category, stock levels).
* **Steps Taken**:
  + **Missing Data Handling**: Imputed missing values by filling average values for numerical fields such as price and setting default values for missing quantities. Dropped rows where critical information (e.g., order\_date) was missing.
  + **Data Type Conversion**: Converted order\_date to a datetime format, and ensured that price and quantity were numerical values.
  + **Merging Datasets**: The datasets were merged using customer\_id and product\_id to create a unified dataset that allowed for better analysis of sales in context with customer and product information.

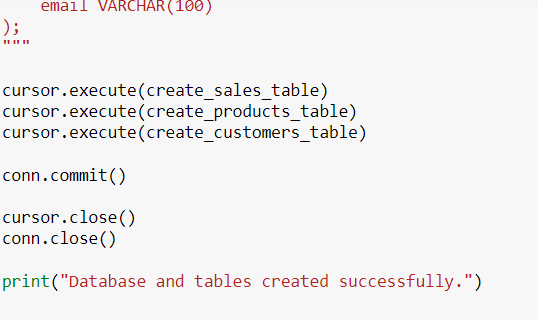


#### ****4. Data Loading into SQL****

Once the data was cleaned, it was loaded into an SQL database. We used MySQL as the database solution.

* **Steps Taken**:
  1. Created a table structure in MySQL for sales, customers, and products.
  2. Loaded the pre processed data into the corresponding tables using SQL Alchemy in Python.





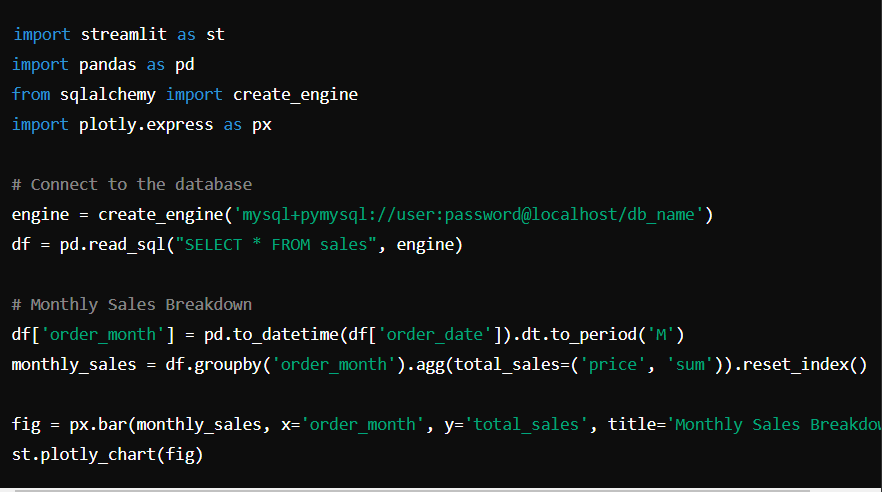
#### ****5. Streamlit Dashboard Development****

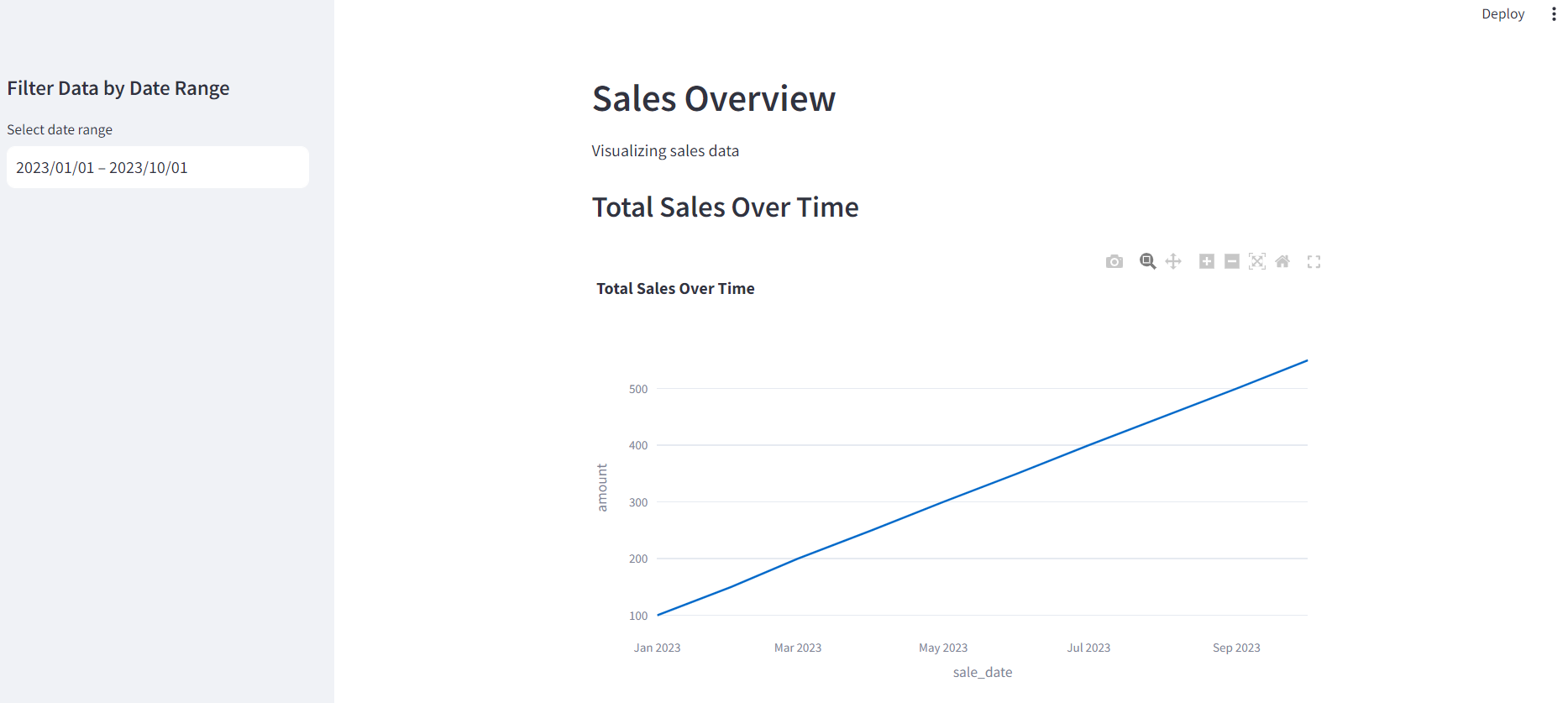
A Streamlit web application was developed to visualize key metrics from the sales data.

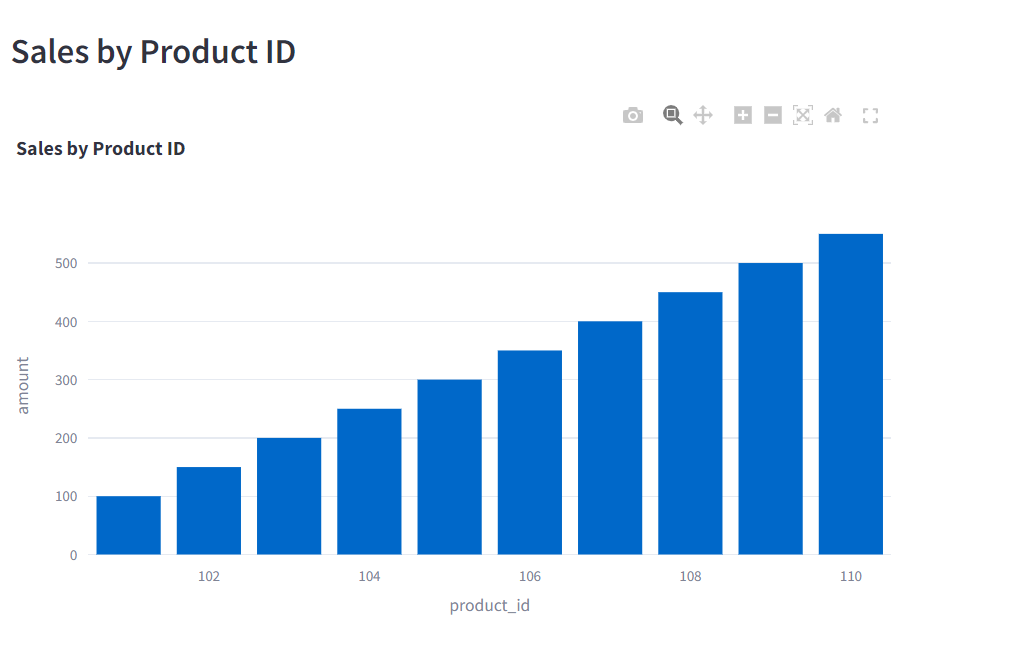
* **Steps Taken**:
  1. Established a connection between Streamlit and the MySQL database using SQLAlchemy.
  2. Created interactive visualizations using Plotly for various metrics, such as sales trends, top-selling products, and regional sales distribution.
  3. The dashboard includes filtering options to view sales for specific time periods, products, or regions.

**Key Features of the Dashboard**:

* **Monthly Sales Breakdown**: A bar chart showing total sales for each month.
* **Top Products**: A bar chart showing the top 5 best-selling products by quantity.
* **Regional Sales Distribution**: A pie chart showing sales distribution across different regions.

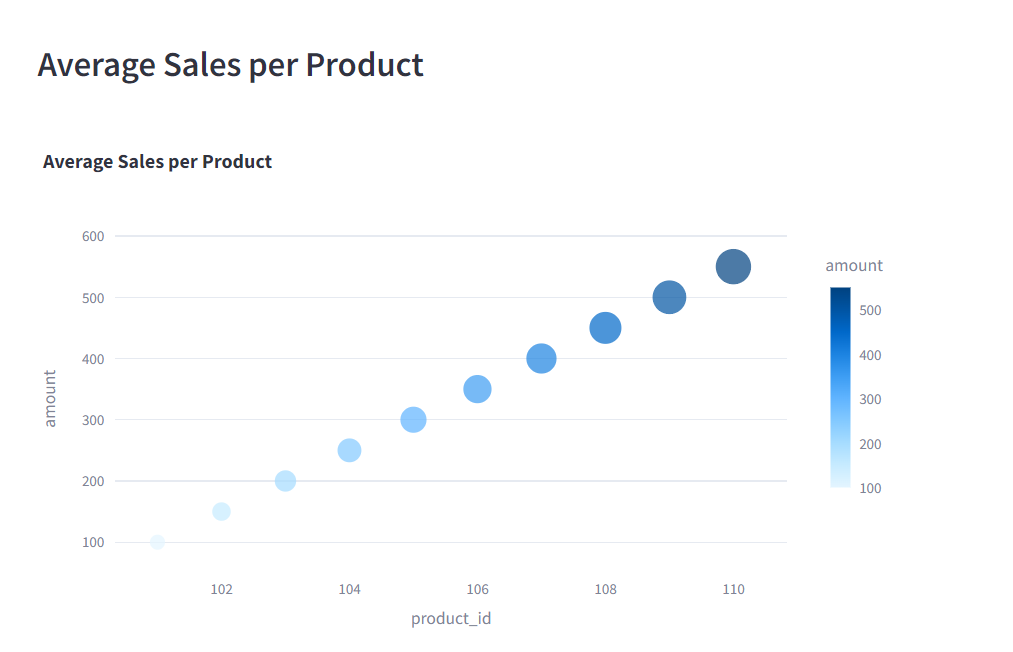


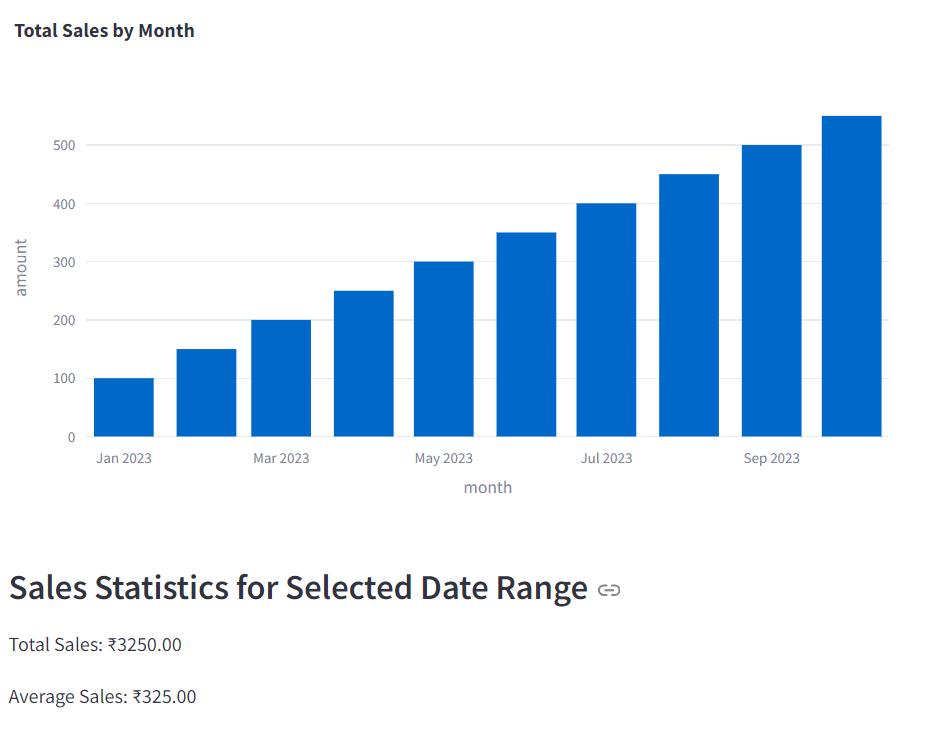










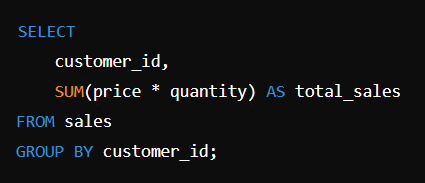


#### ****6. SQL Queries for Insights****

The following SQL queries were developed to extract key business insights:

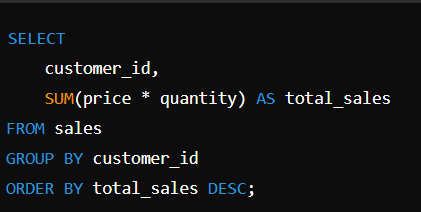
### 1. ****Sales by Customer****

This query retrieves the total sales amount per customer.



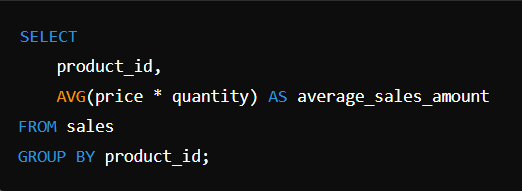
### 2. ****Total Sales by Customer****

This query retrieves the total amount of sales (sum of the sales amounts) for each customer.



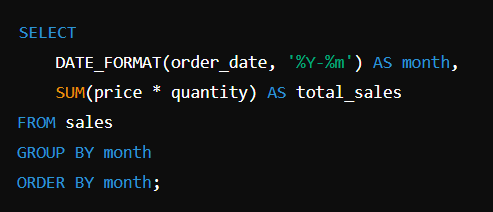
### 3. ****Average Sales Amount by Product****

This query calculates the average sales amount per product.



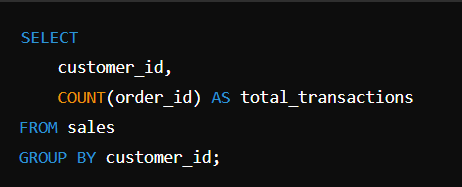
### 4. ****Total Sales by Month****

This query retrieves the total sales amount for each month.



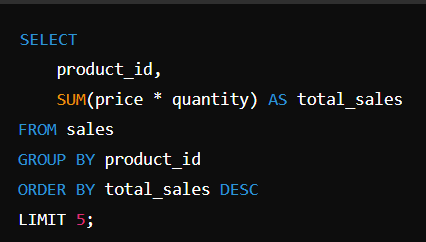
### 5. ****Total Number of Sales Transactions by Customer****

This query counts the total number of sales transactions (orders) for each customer.



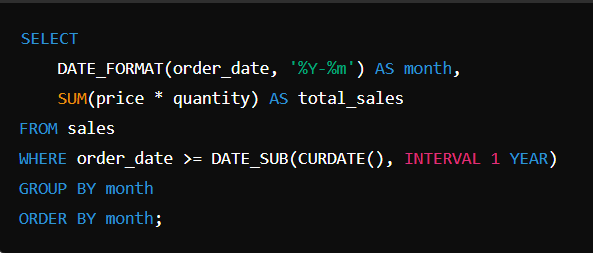
### 6. ****Query to Get the Top 5 Products by Total Sales Amount****

This query retrieves the top 5 products based on the total sales amount.



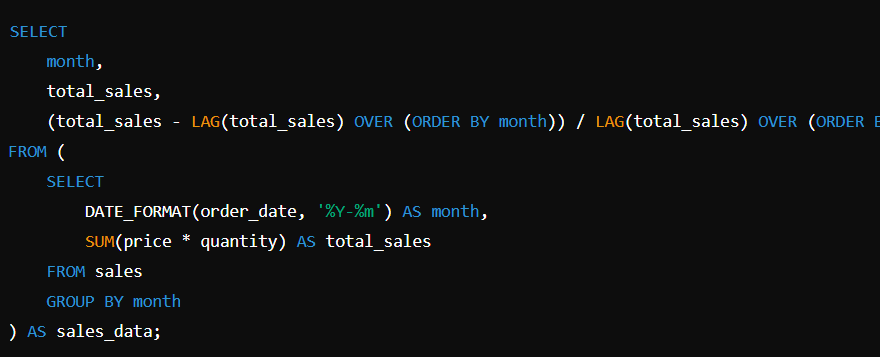
### 7. ****Query to Get Total Sales per Month for the Past Year****

This query retrieves the total sales amount for each month over the past year.



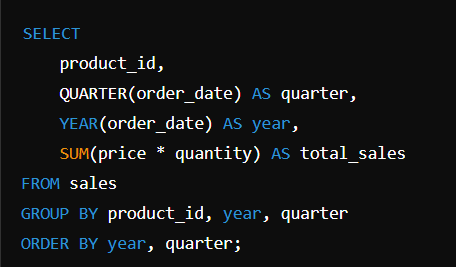
### 8. ****Query to Get Monthly Sales and Calculate Growth Rate****

This query calculates monthly sales and the month-over-month growth rate.



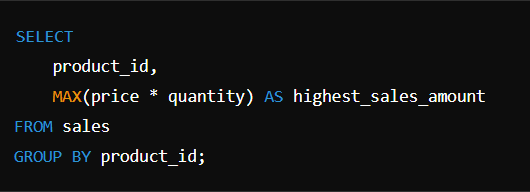
### 9. ****Query to Get Total Sales Amount by Product and Quarter****

This query retrieves the total sales amount by product and quarter.



### 10. ****Query to Get the Highest Sales Amount for Each Product****

This query retrieves the highest single sales amount for each product.



#### ****7. Future Work****

* **Additional Filters**: Add more advanced filtering options (e.g., product categories, specific time frames) for further analysis.
* **Predictive Modeling**: Implement machine learning models for sales forecasting or customer segmentation.
* **Real-Time Data**: Integrate the app with real-time data sources for continuous updates.

#### ****8. Conclusion****

This project successfully demonstrates the process of data cleaning, loading, analysis, and visualization for a sales dataset. By leveraging the power of SQL for data management and Streamlit for visualization, an effective and interactive sales dashboard was created. The dashboard allows for dynamic exploration of sales trends, customer behavior, and product performance, providing valuable insights for business decision-making.