Data Engineering Case Study

Nirosh Kumar R• AA.SC.P2MCA2107478

Project link https://github.com/NiroshKumarR/etlretailsales

Introduction

- This project focuses on performing Extract, Transform, and Load (ETL) operations on a Retail Sales dataset.
- The objective is to process the raw data, transform it into a suitable format for analysis, and store it in a PostgreSQL database.

Data Transformation

- The raw data was analyzed and transformed using Python and the Pandas library.
- Data transformation steps included cleaning, filtering, aggregating, and structuring the data.
- The transformed data provides valuable insights into sales trends, popular products, and category performance.

Data Storage and Retrieval:

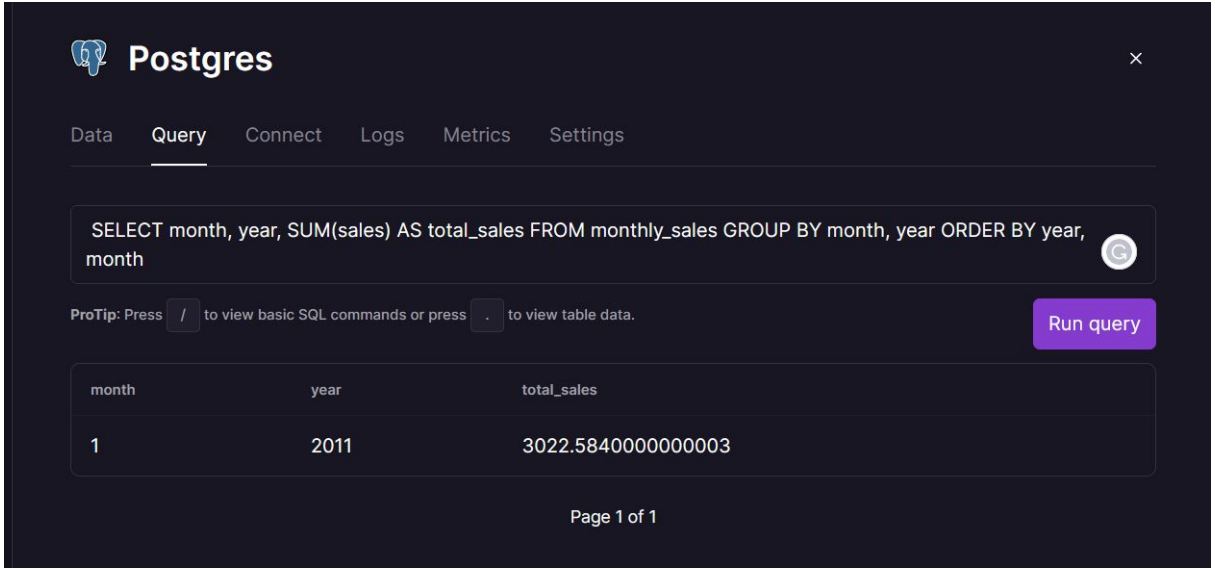
- The transformed data is stored in a PostgreSQL database for efficient storage and retrieval.
- The psycopg2 library is used to establish a connection to the database and execute SQL queries.
- The database schema is designed to accommodate the transformed data, ensuring easy access for analysis.

Data Analysis and Insights:

- The transformed data enables comprehensive analysis of retail sales performance.
- Key insights include monthly sales trends, top-selling products, and revenue by region.
- Visualizations and charts are used to present the findings and aid in understanding the data.

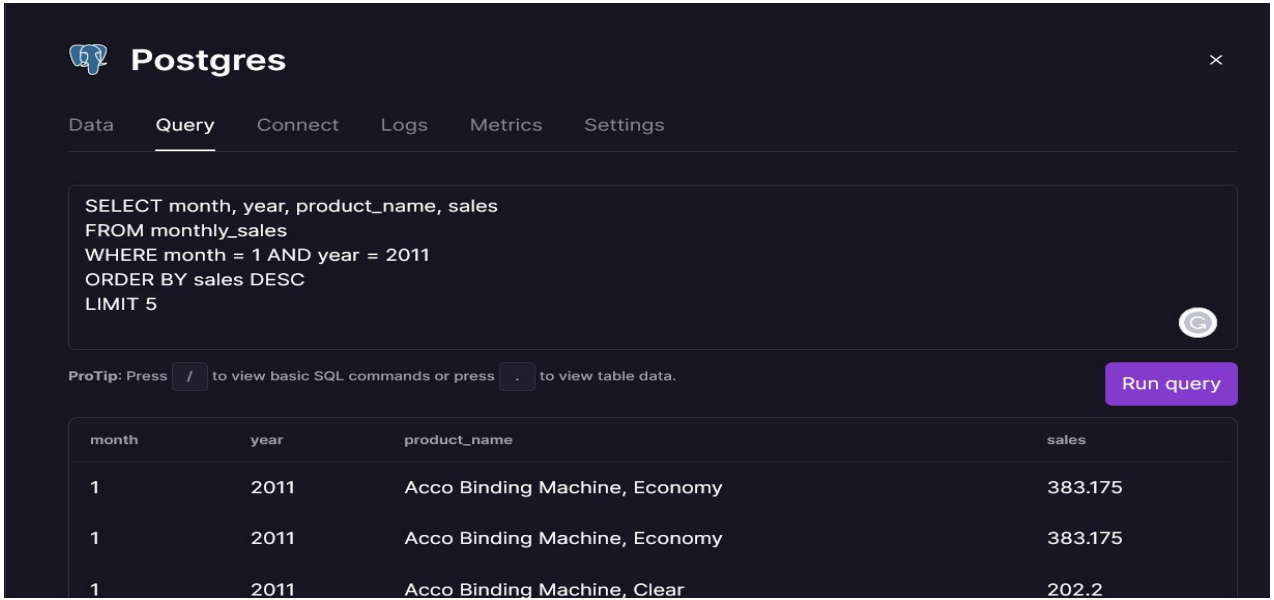
Business Questions:

- Monthly sales trends over the years



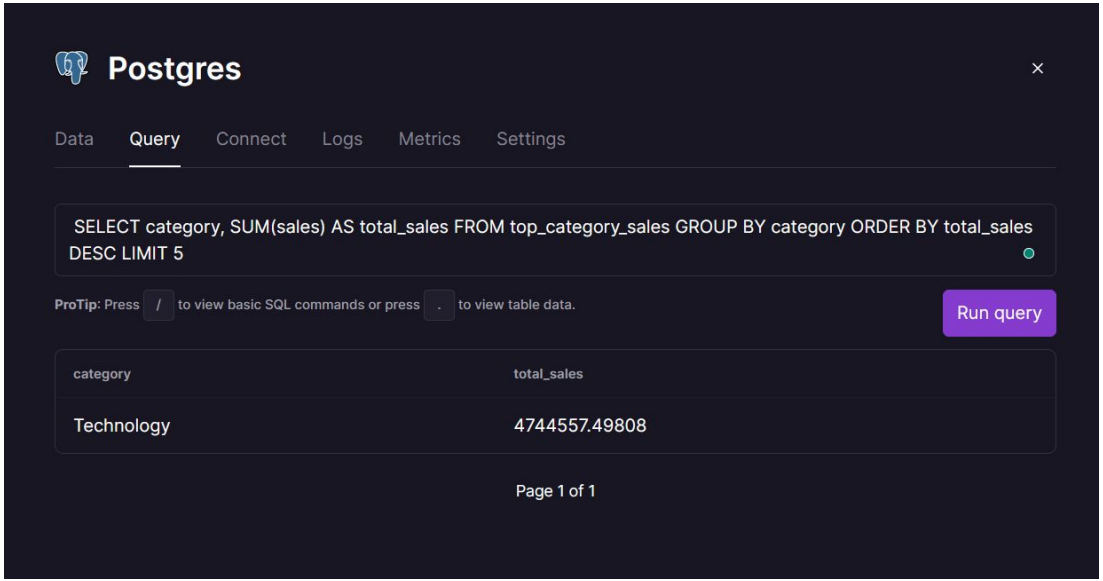
Business Questions:

- Top-selling products in a given month and year



Business Questions:

- What are the top-selling product categories overall?



Conclusion:

- The ETL Retail Sales Data Engineering Project demonstrates the importance of data preparation for meaningful analysis.
- The project highlights the use of Python, Pandas, and PostgreSQL for efficient data processing and storage.
- The insights gained from the transformed data can drive informed business decisions and identify areas for improvement.

Thank You