ETL Process, Daily Reports and Data Analysis for Sales Data

by Claudya Evangelica Putri Soleman

Objectives

Managing simple ETL pipelines, transforming and analyzing data from daily sales CSV files to produce daily reports and provide insights to support the management's business decisions.

Scope

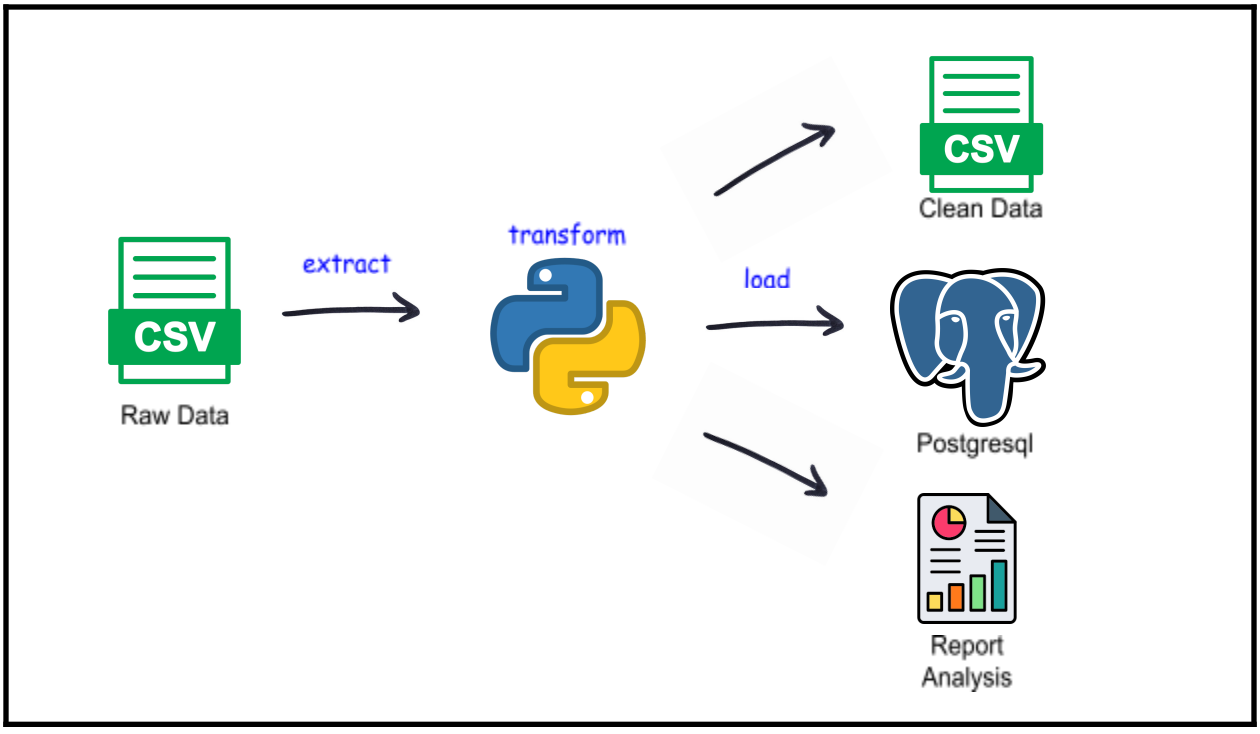
Requirements :

- Clean the data, delete transactions with an empty Price, and add a new column Total_Revenue = Quantity * Price. Then insert data into a separate table daily_sales_cleaned
- Load data into database and calculate total daily revenue from a table daily_sales_cleaned
- Explore any business analysis other than the above that can be used by management in making business decisions

Deliverables :

- Clean data in .csv file
- Query total daily revenue
- Data Analysis

Process Flow



The ETL (Extract, Transform, Load) process in this project use Python as the core tool for handling and processing raw data.

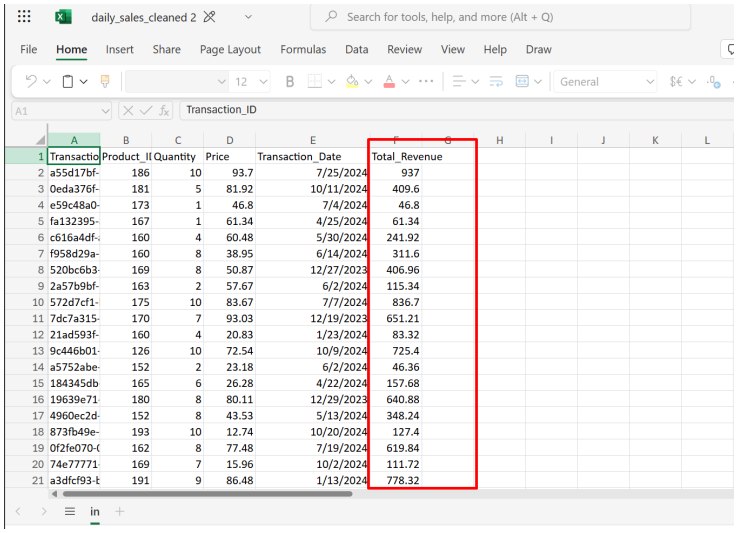
Extract : Raw data in CSV format is imported into Python for initial processing.

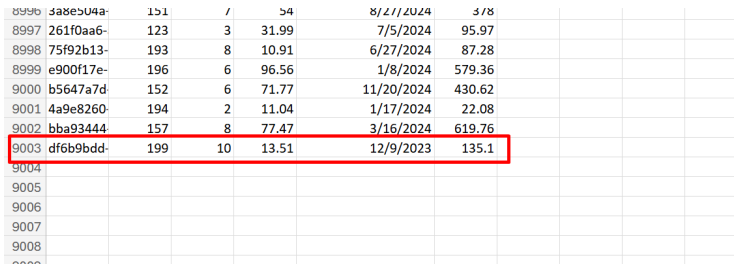
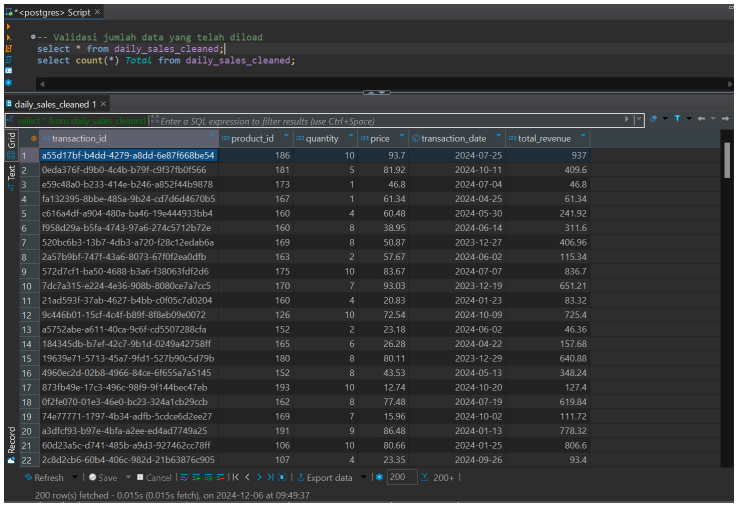
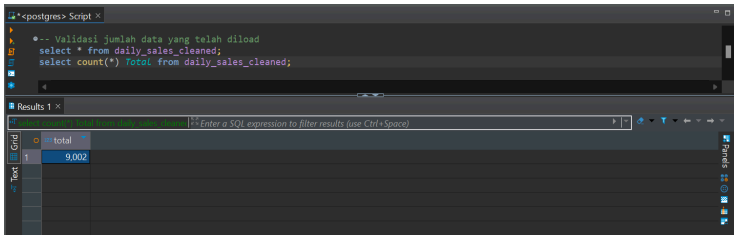
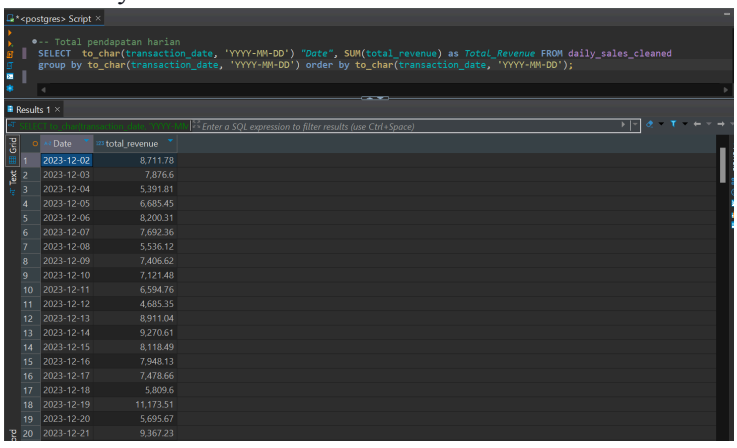
Transform : Python is used to clean, manipulate, and prepare the data.

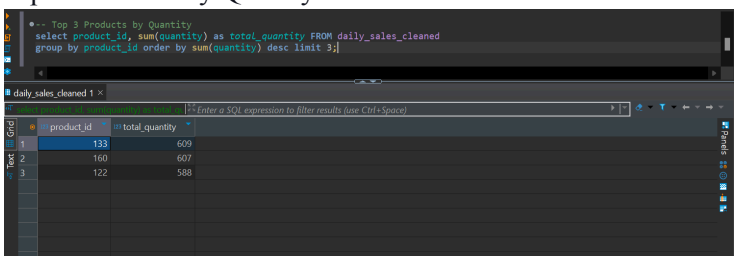
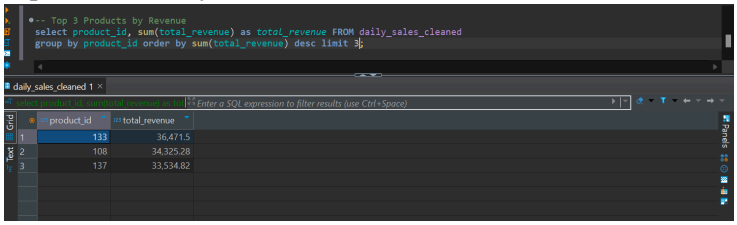
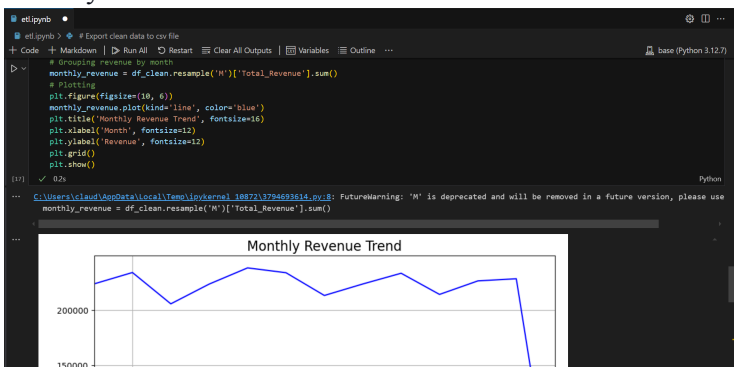
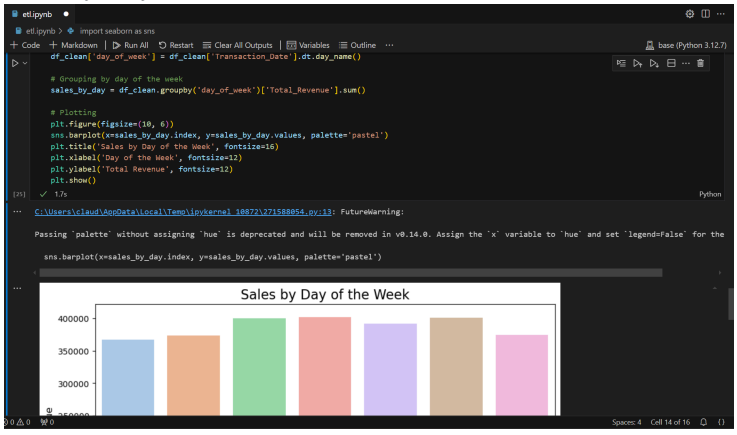
Load : The transformed data is output to:

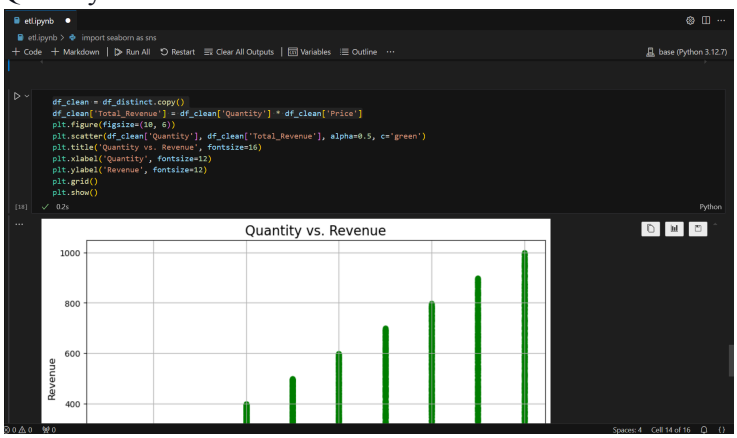
- A cleaned CSV file for external use or sharing.
- A PostgreSQL database for structured storage and query execution.
- A report or analysis, providing visual insights and actionable results.

Results

No	Result	Validation and Evidence
1	Successfully cleaned the data, add column Total Revenue and saved it into a CSV file	Exported data **daily_sales_cleaned.csv** Colum Total Revenue  Total Rows plus header = 9003

		
2	Successfully loaded the data to Postgresql and calculated daily revenue	Loaded Data  Total Count Rows  Total Daily Revenue 

3	Succeeded in analyzing available data and creating reports for supporting business decision making	Top 3 Products by Quanity  Top 3 Products by Revenue  Monthly Revenue Trend  Sales by Day of the Week 

		Quantity vs Revenue 

Analysis Results

Top Selling Products by Quantity

No	Product ID	Total Quantity
1	133	609
2	160	607
3	122	588

Table Top Selling Products by Quantity highlights the top three products sold in the highest quantities. The narrow difference in quantities sold suggests that these products enjoy similar levels of popularity and consistent demand among customers. This insight underscores the importance of maintaining adequate inventory for these items to avoid stockouts and ensure customer satisfaction.

Additionally, these top-performing products present opportunities for targeted marketing strategies, such as bundling, discounts, or promotions, to further enhance sales. Furthermore, these items may reflect broader customer trends or preferences, which could be explored through additional analysis of customer demographics or purchasing behavior.

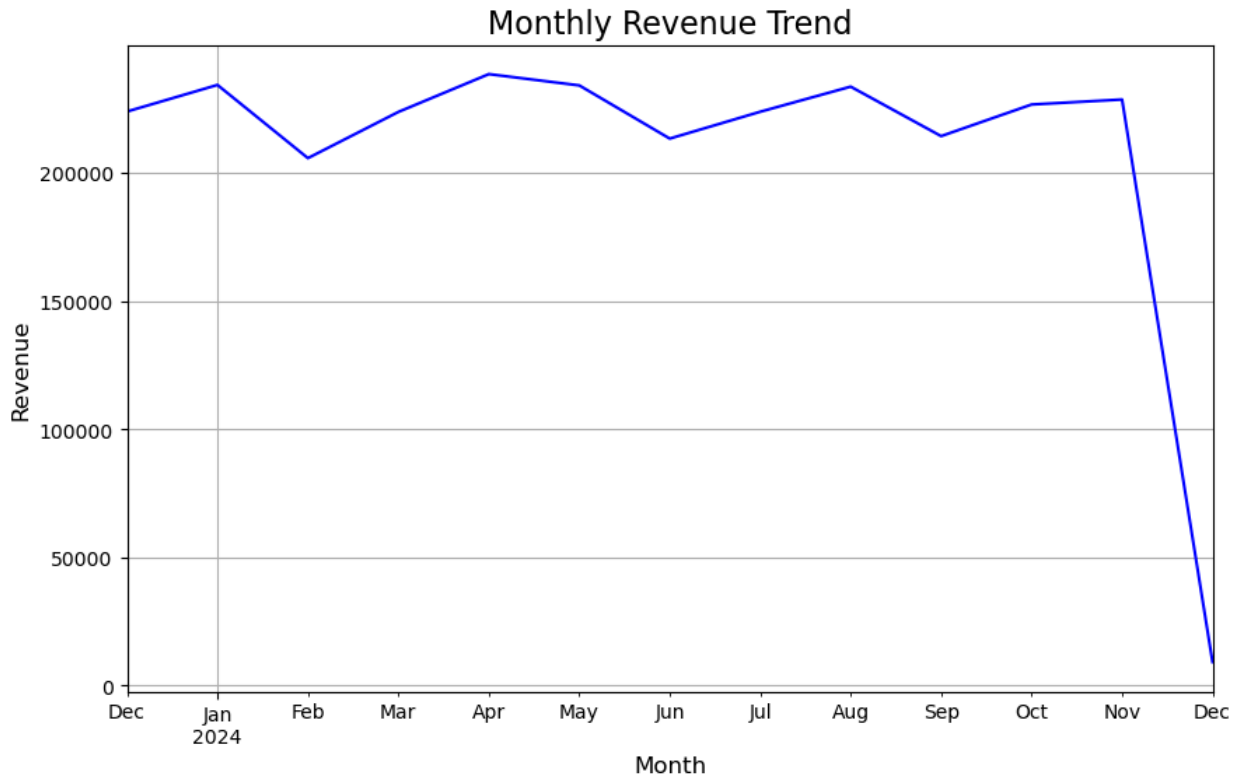
Top Selling Products by Revenue

No	Product ID	Total Revenue
1	133	36471.50
2	108	34325.28
3	137	33534.82

This table highlights the top-selling products by revenue, providing valuable insight into the company's most financially impactful offerings. Product with ID 133 ranks first, generating the highest total revenue, followed by Product with ID 108 and 137.

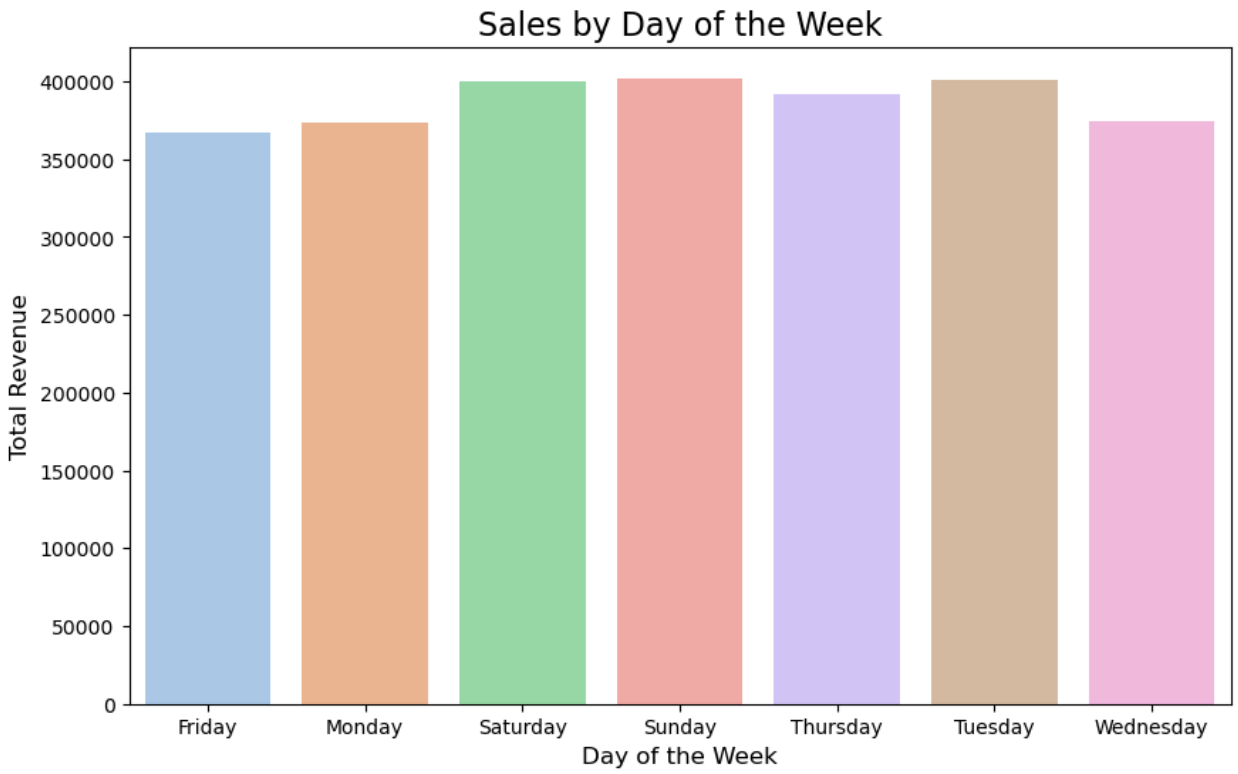
These figures indicate that these three products are significant contributors to the company's revenue stream, suggesting their high demand or premium pricing. To further analyze their performance, it would be useful to compare their sales volume, profit margins, and customer segments.

Monthly Revenue Trend



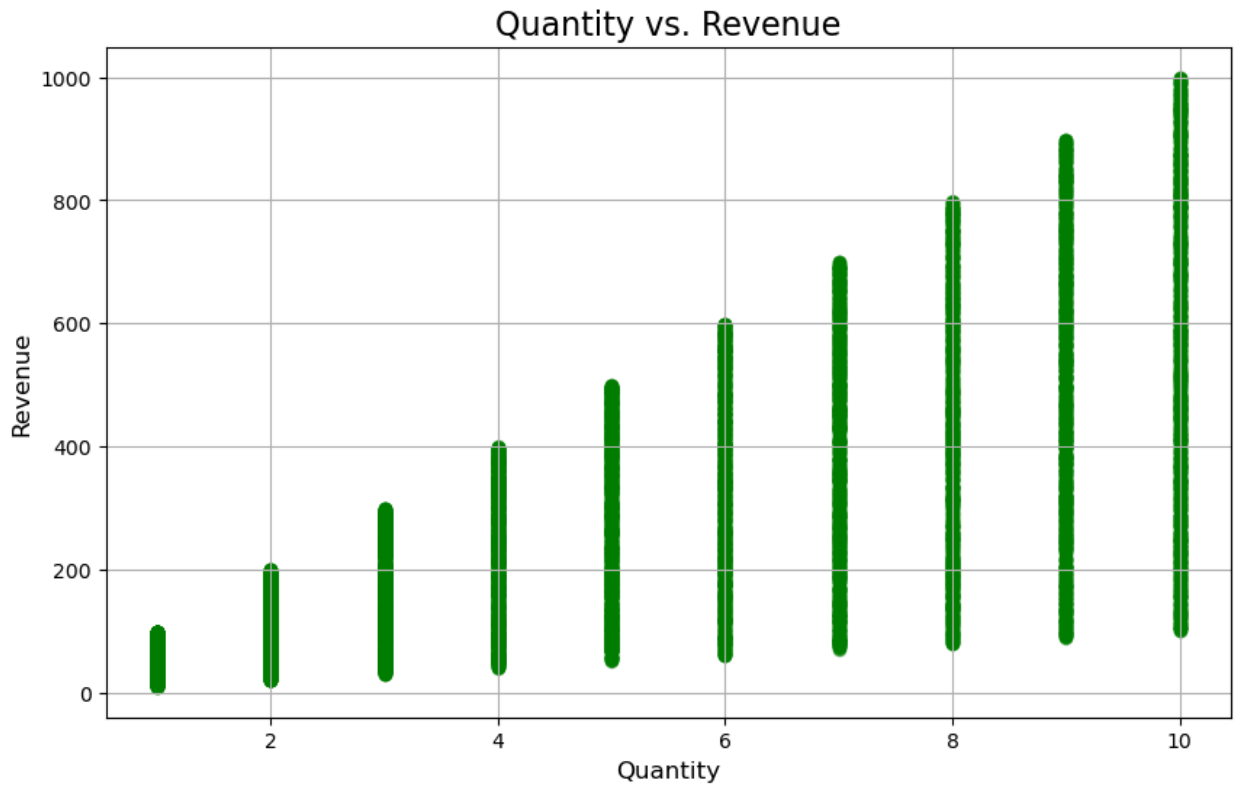
The revenue shows a steady pattern throughout the year, with a rise in some months and dips in others. There is a sharp decline in December, which might indicate incomplete data for December or a significant drop in sales. The performance can be maximized by investigating the factors contributing to the high revenue months, such as promotional campaigns, product launches, or seasonal demand, and replicating these strategies during low-performing periods.

Sales by Day of the Week



The bar chart depicting Sales by Day of the Week provides insight into daily revenue patterns. The sales are relatively consistent, with no significant spikes or declines on any specific day. This uniform distribution indicates that customer purchasing behavior remains steady throughout the week. This even distribution also may imply that marketing efforts, staffing, and inventory levels can be maintained uniformly across all days, without needing to allocate more resources to specific days.

Quantity vs Revenue



The Quantity vs. Revenue scatter plot reveals a positive relationship between the quantity sold and the revenue generated. As the quantity of items sold increases, the corresponding revenue also rises, indicating that higher quantities typically lead to greater total sales. This could point to common purchase patterns or product pricing structures influencing how much customers typically buy. Overall, focusing on strategies to increase the number of higher-quantity sales could drive greater revenue.