

# 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

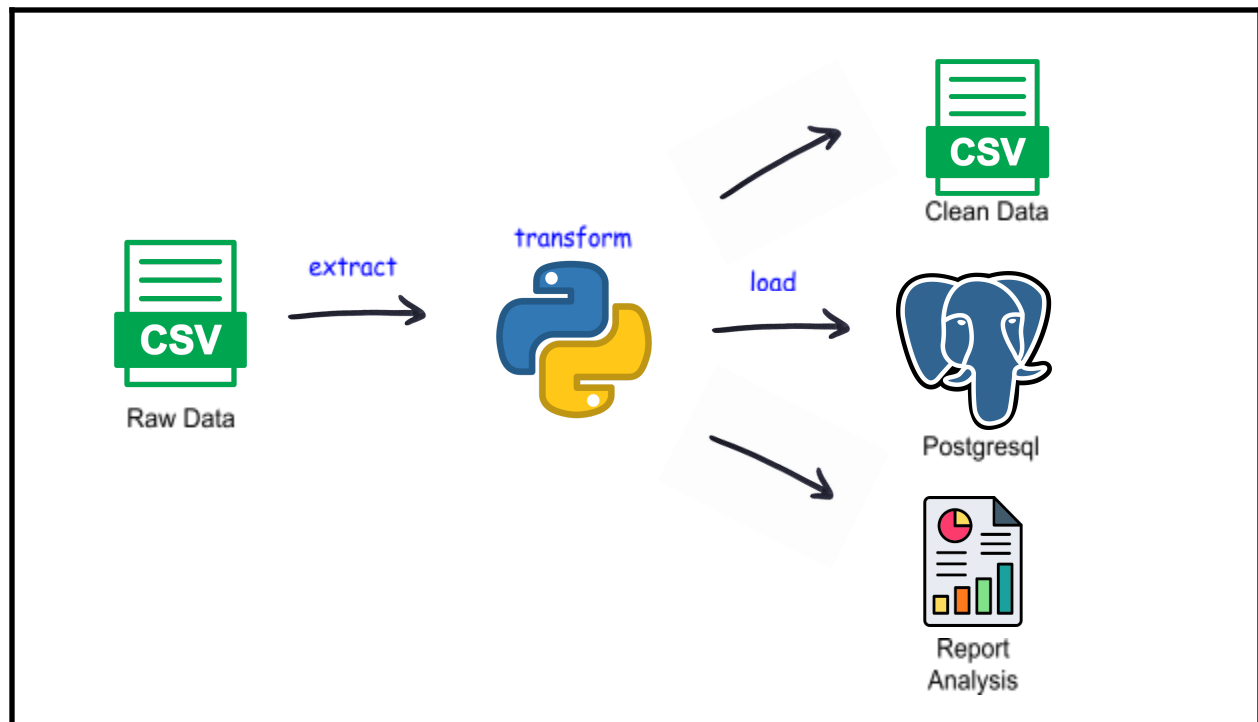
Requirements :

- 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.

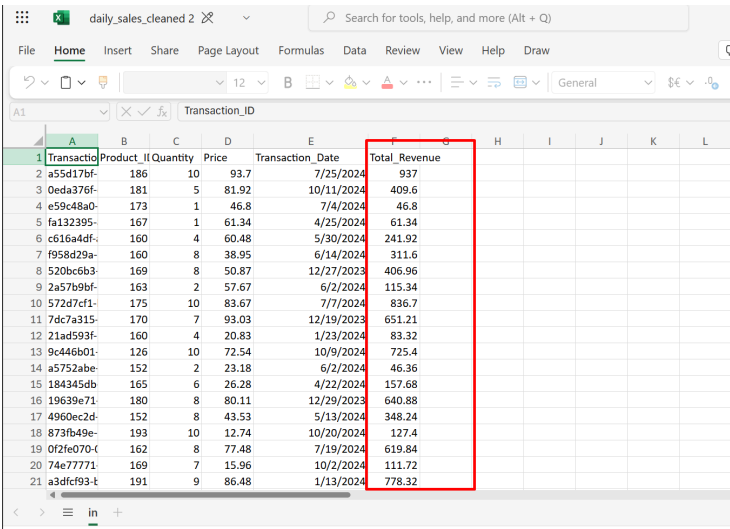
Extract : Raw data in CSV format is imported into Python for initial processing.

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	<p>Exported data <b>daily_sales_cleaned.csv</b></p> <p>Column Total Revenue</p>  <p>Total Rows plus header = 9003</p>

8995	3a8e04a-	151	1	54	8/21/2024	3/8
8997	261f0aa6-	123	3	31.99	7/5/2024	95.97
8998	75f92b13-	193	8	10.91	6/27/2024	87.28
8999	e900f17e-	196	6	96.56	1/8/2024	579.36
9000	b5647a7d	152	6	71.77	11/20/2024	430.62
9001	4a9e8260-	194	2	11.04	1/17/2024	22.08
9002	bb93444-	157	8	77.47	3/16/2024	619.76
9003	d6b9b5dd-	199	10	13.51	12/9/2023	135.1
9004						
9005						
9006						
9007						
9008						
9009						

Successfully loaded the data to Postgresql and calculated daily revenue

PostgreSQL Script ×

```

-- Validasi jumlah data yang telah diload
select * from daily_sales_cleaned;
select count(*) total from daily_sales_cleaned;

```

daily\_sales\_cleaned 1 ×

Enter a SQL expression to filter results (use Ctrl+Space)

	transaction_id	product_id	quantity	price	transaction_date	total_revenue
1	a5d517bf-b4dd-4279-88dd-6e87f668b654	186	10	93.7	2024-07-25	937
2	0ed3a7bf-d960-4c4b-b79f-c8f30f5666	181	5	81.92	2024-10-11	409.6
3	e59c48a0-b233-414e-b246-a85244ab9878	173	1	46.8	2024-07-04	46.8
4	fa132395-8bbe-485a-9b24-cd7d6d4670b5	167	1	61.34	2024-04-25	61.34
5	c616ad4f-a904-480a-ba46-19e44933b04a	160	4	60.48	2024-05-30	241.92
6	f95d29a4-b5fa-4743-9746-27a55712b72e	160	8	38.95	2024-06-14	311.6
7	520b06b3-13b7-4d3b-3720-182c12edab6a	169	8	50.87	2023-12-27	406.96
8	2a576bf3-747f-43a6-8073-67f0f2eaddb3	162	2	57.67	2024-06-02	115.34
9	572d7cf1-b5a0-4688-b3a6-18306c3d2d2e	175	10	83.67	2024-07-07	836.7
10	72d7a35c-e274-4a36-908b-8800e0a7c05f	170	7	80.83	2023-12-19	565.81
11	21e4d5f3-374e-4e27-b4bb-c90c27d0204a	160	4	20.83	2024-01-23	83.32
12	9d464601-154c-4d8f-898b-878e0b6a0072	126	10	72.54	2024-10-09	725.4
13	a3752abe-a611-4d0c-8cfd-c505f0288cfa	152	2	23.18	2024-06-02	46.36
14	1843435d-7bf6-4e27-9b1d-0249a42758ff	165	6	26.28	2024-04-22	157.68
15	19639f71-5713-45a7-99d1-527b90c0d79b	180	8	80.11	2023-12-29	640.88
16	4950c2d-02b8-4966-840c-61655a7a5145	152	8	43.53	2024-05-13	348.24
17	873ba49c-1732-496c-9819-9f144be47eb7	193	10	12.74	2024-10-20	127.4
18	0f2f6f70-01e3-46e0-bc23-32a41cb29c3b	162	8	77.48	2024-07-19	619.84
19	74e777f1-1797-4b34-ad1e-5cd6ed62ee27	169	7	15.96	2024-10-02	111.72
20	a3ddfc93-b97e-4bfa-a2ee-ed4ad7749a25	191	9	86.48	2024-01-13	778.32
21	60d23a5c-d741-485b-a93c-27f46c2c78ff	106	10	80.66	2024-01-25	806.6
22	2cd82c6b-6044-406c-982d-21b6387ec905	107	4	23.35	2024-09-26	93.4

200 rows fetched - 0.015s (0.015s fetch, on 2024-10-16 at 09:49:37)

The screenshot shows a PostgreSQL IDE interface. The top pane displays a SQL script with a comment and a query:

```
-- Validasi jumlah data yang telah diload
select * from daily_sales_cleaned;
select count(*) total from daily_sales_cleaned;
```

The bottom pane shows the results of the second query. A table with one column 'total' and one row containing the value '9,002' is displayed.

	total
1	9,002

PostgreSQL Client Interface:

```

-- Total pending transaction
SELECT to_char(transaction_date, 'YYYY-MM-DD') "Date", SUM(total_revenue) as Total_Revenue FROM daily_sales_cleaned
group by to_char(transaction_date, 'YYYY-MM-DD') order by to_char(transaction_date, 'YYYY-MM-DD');

```

Results 1 x

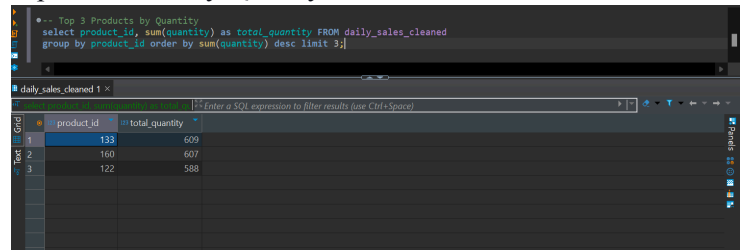
Enter a SQL expression to filter results (see Ctrl+Space)

Grp	Date	total_revenue
1	2023-12-02	8,711.78
2	2023-12-03	7,676.6
3	2023-12-04	5,391.81
4	2023-12-05	6,685.45
5	2023-12-06	8,200.31
6	2023-12-07	7,692.36
7	2023-12-08	5,536.12
8	2023-12-09	7,406.62
9	2023-12-10	7,121.48
10	2023-12-11	6,594.76
11	2023-12-12	4,685.35
12	2023-12-13	8,911.04
13	2023-12-14	9,270.61
14	2023-12-15	8,118.49
15	2023-12-16	7,948.13
16	2023-12-17	7,478.66
17	2023-12-18	5,809.6
18	2023-12-19	11,173.51
19	2023-12-20	5,695.67
20	2023-12-21	9,367.23

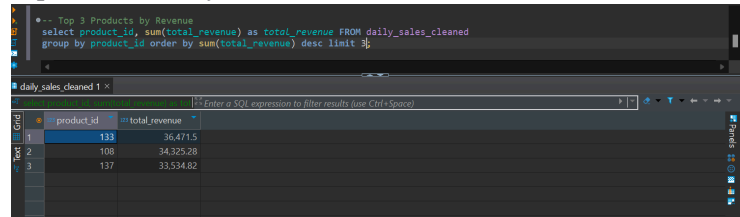
3

Succeeded in analyzing available data and creating reports for supporting business decision making

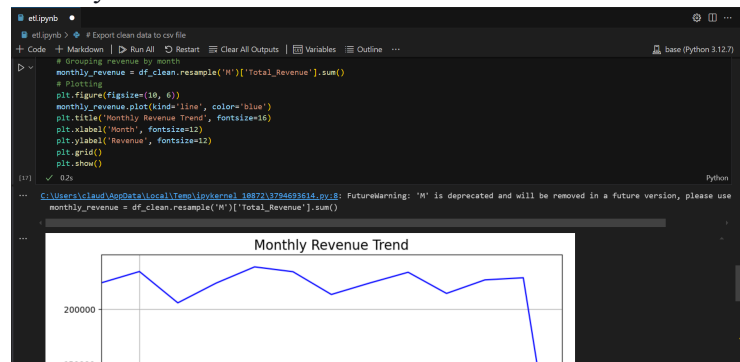
### Top 3 Products by Quantity



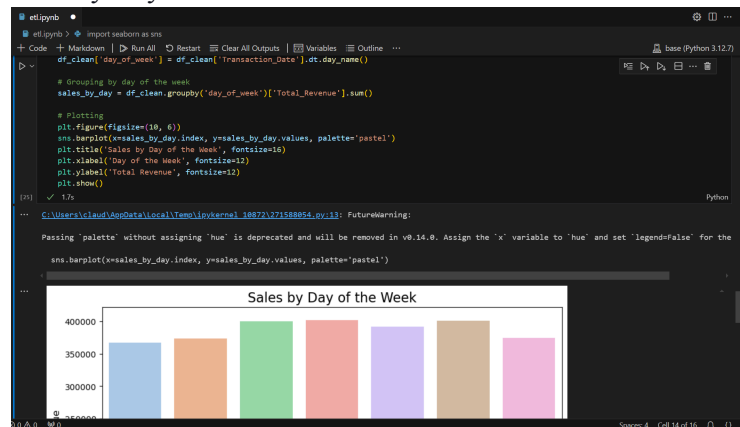
### Top 3 Products by Revenue



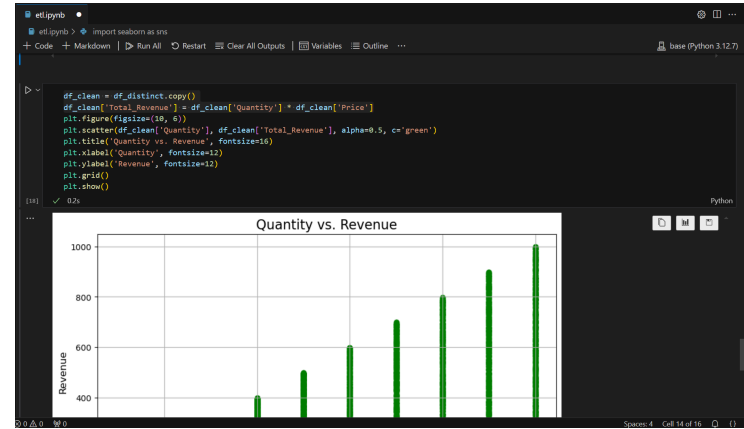
### Monthly Revenue Trend



### 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.

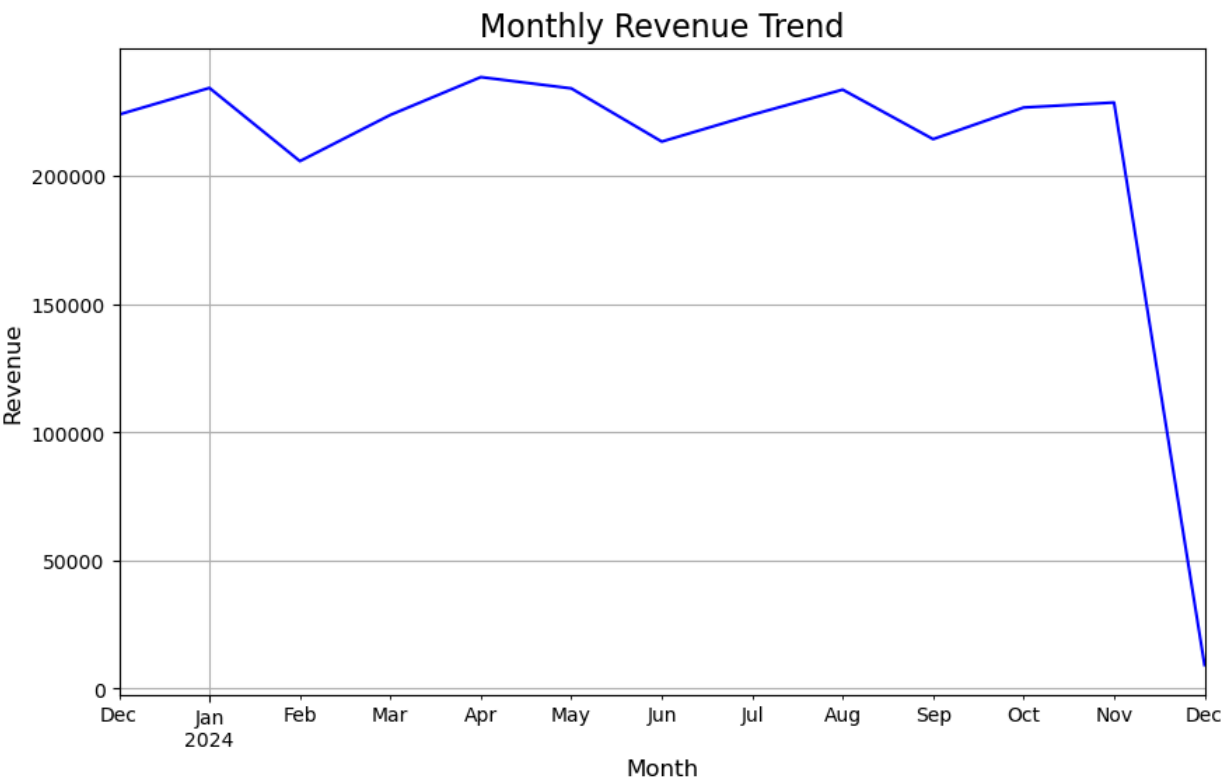
### Top Selling Products by Revenue

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.

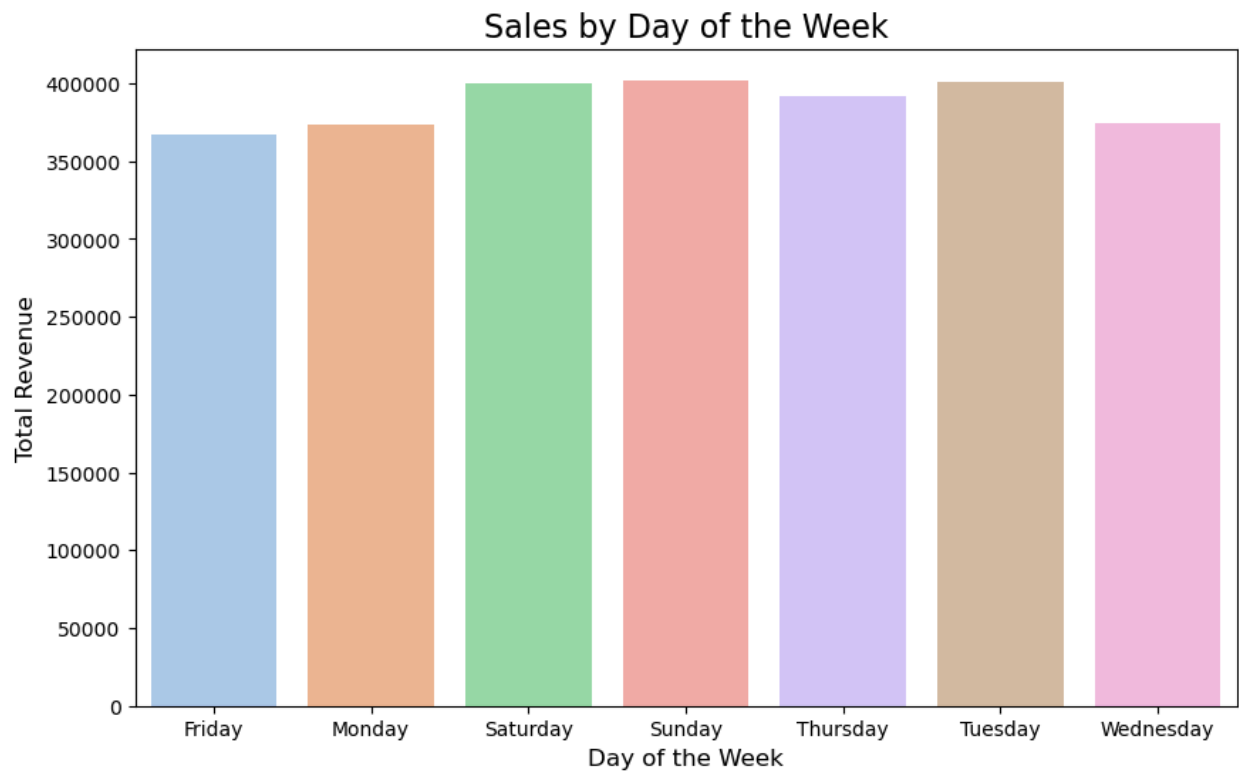
These figures indicate that these three products are significant contributors to the company's revenue stream, suggesting their high demand or premium pricing. To further analyze their performance, it would be useful to compare their sales volume, profit margins, and customer segments.

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.

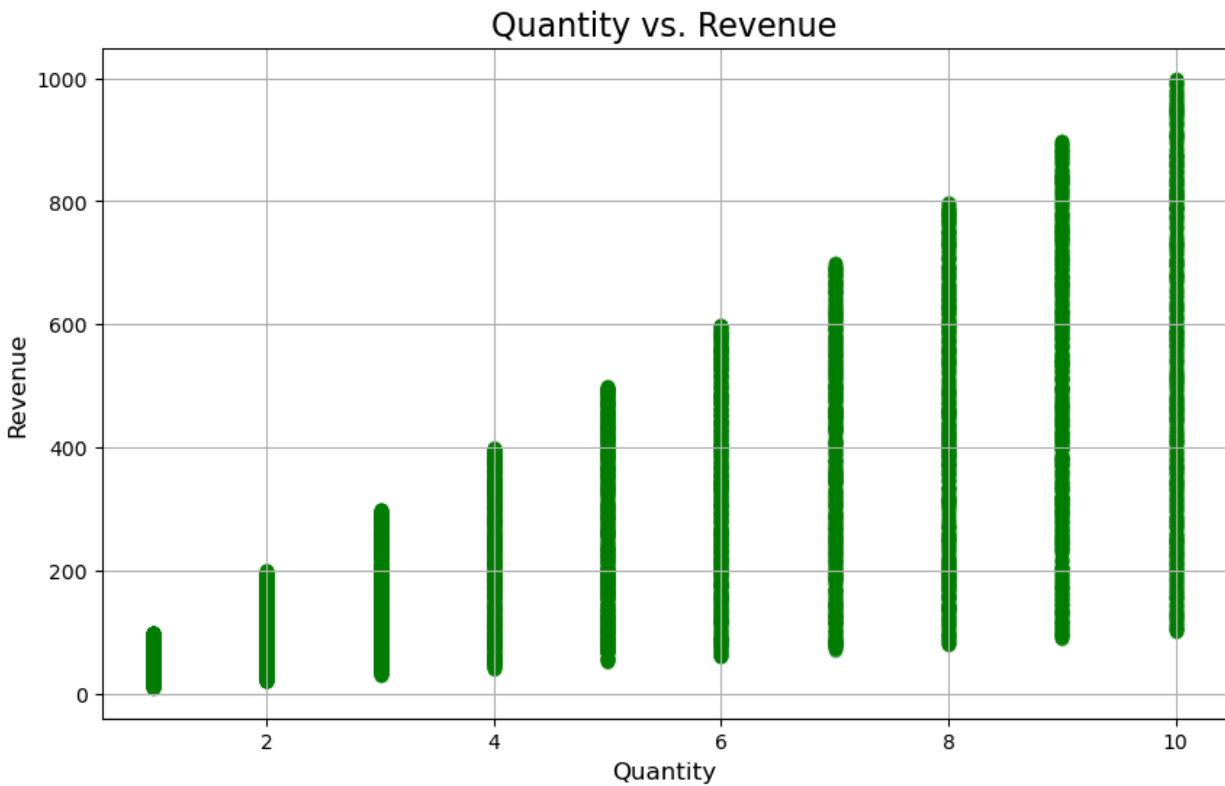
Sales by Day of the Week



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.