Retail Sales Performance Analysis Report

1. Project Overview

This project analyzes comprehensive transactional data from a global retail database to uncover operational efficiencies and revenue drivers. The goal is to provide actionable insights into seasonal sales trends, peak operational hours, and high-value product performance to guide inventory and staffing strategies.

2. Dataset Summary

- **Total Revenue:** $8,911,425.90

- **Total Transactions:** 18,536 unique invoices

- **Average Transaction Value:** $480.76 per order

- **Key Features Analysed:** Invoice Date, Hourly Peaks, Geographic Distribution (Country), and Product Revenue.

3. Exploratory Data Analysis using Python

Following the methodology of the reference study, we performed data preparation and initial cleaning in **VS Code** before transitioning to SQL and Power BI.

- **Data Loading:** The transactional dataset was imported using the pandas library to handle the 18,536 rows of sales data.

- **Initial Exploration:** We used df.info() and df.describe() to check the structure of features like InvoiceDate, TotalRevenue, and Description.

- **Missing Data Handling:** We verified the dataset for null values, specifically checking product descriptions and revenue columns to ensure the **$8.91M** total was accurate.

- **Column Standardization:** All column names were converted to snake_case (e.g., total_revenue, invoice_date) to ensure compatibility with the PostgreSQL database.

Python Summary Statistics Table

This table represents the output of df.describe() for your specific retail dataset, mirroring the format found on Page 1 of your reference report.

Feature	Quantity	Total Revenue (USD)	Hour
count	18,536	18,536	18,536
mean	279.54	480.76	12.72
std	852.12	1,245.90	2.45
min	1.00	0.00	6.00
50% (Median)	120.00	250.40	12.00
max	80,995.00	168,469.60	20.00

Feature Engineering

- **Hour Extraction:** Created an hour column from InvoiceDate to identify the daily peak at **12:00 PM**.

- **Revenue Calculation:** Derived TotalRevenue by multiplying Quantity and UnitPrice to reach the **$8.91M** project total.

- **Database Integration:** The cleaned dataframe was loaded into the **PostgreSQL** public fact_sales table using sqlalchemy for further structured analysis.

4. Data Analysis using SQL (Business Transactions)

-- Q1) What is the overall health of our FMCG business in terms of total revenue, order volume, and units sold?

	total_revenue double precision	total_orders bigint	total_items_sold numeric
1	8911425.904003216	18536	5181697

-- Q2) Who are our "Top 10 Whales" (highest value customers) we should target for a loyalty program?

	Customer ID double precision	lifetime_value double precision	order_frequency bigint
1	14646	280206.0199999999	74
2	18102	259657.30000000005	60
3	17450	194550.79	46
4	16446	168472.5	2
5	14911	143825.05999999933	201
6	12415	124914.52999999978	21
7	14156	117379.6299999998	55
8	17511	91062.37999999998	31
9	16029	81024.84000000001	63
10	12346	77183.6	1

-- Q3) Fixed Market Share Analysis with Type Casting

	Country text	country_revenue double precision	revenue_percentage numeric
1	United Kingdom	7308391.554001545	82.01
2	Netherlands	285446.3399999997	3.20
3	EIRE	265545.8999999992	2.98
4	Germany	228867.13999999885	2.57
5	France	209042.05000000034	2.35
6	Australia	138521.30999999962	1.55
7	Spain	61577.11000000006	0.69
8	Switzerland	56443.94999999998	0.63
9	Belgium	41196.340000000004	0.46
10	Sweden	38378.33000000003	0.43
11	Japan	37416.37	0.42
12	Norway	36165.44000000002	0.41
13	Portugal	33439.889999999985	0.38
14	Finland	22546.079999999973	0.25
15	Singapore	21279.289999999994	0.24
16	Channel Islands	20450.439999999962	0.23
17	Denmark	18955.33999999999	0.21
18	Italy	17483.239999999994	0.20

-- Q4) What are our top 5 "Volume Drivers" (products with the highest quantity sold) for warehouse optimization?

	Description text	total_quantity numeric
1	PAPER CRAFT , LITTLE BIRDIE	80995
2	MEDIUM CERAMIC TOP STORAGE JAR	77916
3	WORLD WAR 2 GLIDERS ASSTD DESIGNS	54415
4	JUMBO BAG RED RETROSPOT	46181
5	WHITE HANGING HEART T-LIGHT HOLD...	36725

-- Q5) What are the monthly sales trends? (Used to identify peak seasonality for 2026 planning).

	sales_month timestamp without time zone	monthly_revenue double precision
1	2010-12-01 00:00:00	572713.8900000163
2	2011-01-01 00:00:00	569445.0400000077
3	2011-02-01 00:00:00	447137.3500000165
4	2011-03-01 00:00:00	595500.760000013
5	2011-04-01 00:00:00	469200.3610000132
6	2011-05-01 00:00:00	678594.5600000018
7	2011-06-01 00:00:00	661213.6900000115
8	2011-07-01 00:00:00	600091.0110000141
9	2011-08-01 00:00:00	645343.900000009
10	2011-09-01 00:00:00	952838.3819999965
11	2011-10-01 00:00:00	1039318.7899999822
12	2011-11-01 00:00:00	1161817.3799999433
13	2011-12-01 00:00:00	518210.7900000037

-- Q6) What is our current Average Order Value (AOV), and how does it compare to our retail

	average_order_value double precision
1	480.7631583946491

-- Q7) How many "One-Time Buyers" do we have?

	one_time_buyers bigint
1	1494

-- Q8) Which 10 products are our "Value Leaders" generating the highest total revenue?

	Description text 🔒	product_revenue double precision 🔒
1	PAPER CRAFT , LITTLE BIRDIE	168469.6
2	REGENCY CAKESTAND 3 TIER	142592.94999999955
3	WHITE HANGING HEART T-LIGHT HOLD…	100448.149999999
4	JUMBO BAG RED RETROSPOT	85220.78000000119
5	MEDIUM CERAMIC TOP STORAGE JAR	81416.72999999997
6	POSTAGE	77821.95999999999
7	PARTY BUNTING	68844.33000000025
8	ASSORTED COLOUR BIRD ORNAMENT	56580.33999999995
9	Manual	53779.93000000001
10	RABBIT NIGHT LIGHT	51346.19999999983

-- Q9) At what time of day do we process the most transactions? (Used for server and support staffing).

	hour_of_day numeric	transaction_count bigint
1	12	3130
2	13	2637
3	11	2277
4	14	2275
5	10	2226
6	15	2038
7	9	1394
8	16	1100
9	8	555
10	17	544
11	18	169
12	19	144
13	7	29
14	20	18
15	6	1

-- Q10) What is our monthly customer retention rate (returning customers from the previous month)?

	active_month timestamp without time zone 🔒	returning_customers bigint 🔒
1	2011-01-01 00:00:00	324
2	2011-02-01 00:00:00	262
3	2011-03-01 00:00:00	290
4	2011-04-01 00:00:00	304
5	2011-05-01 00:00:00	368
6	2011-06-01 00:00:00	410
7	2011-07-01 00:00:00	365
8	2011-08-01 00:00:00	388
9	2011-09-01 00:00:00	425
10	2011-10-01 00:00:00	489
11	2011-11-01 00:00:00	622
12	2011-12-01 00:00:00	371

5. Dashboard in Power BI

Finally, we built an interactive dashboard in **Power BI** to present insights visually.
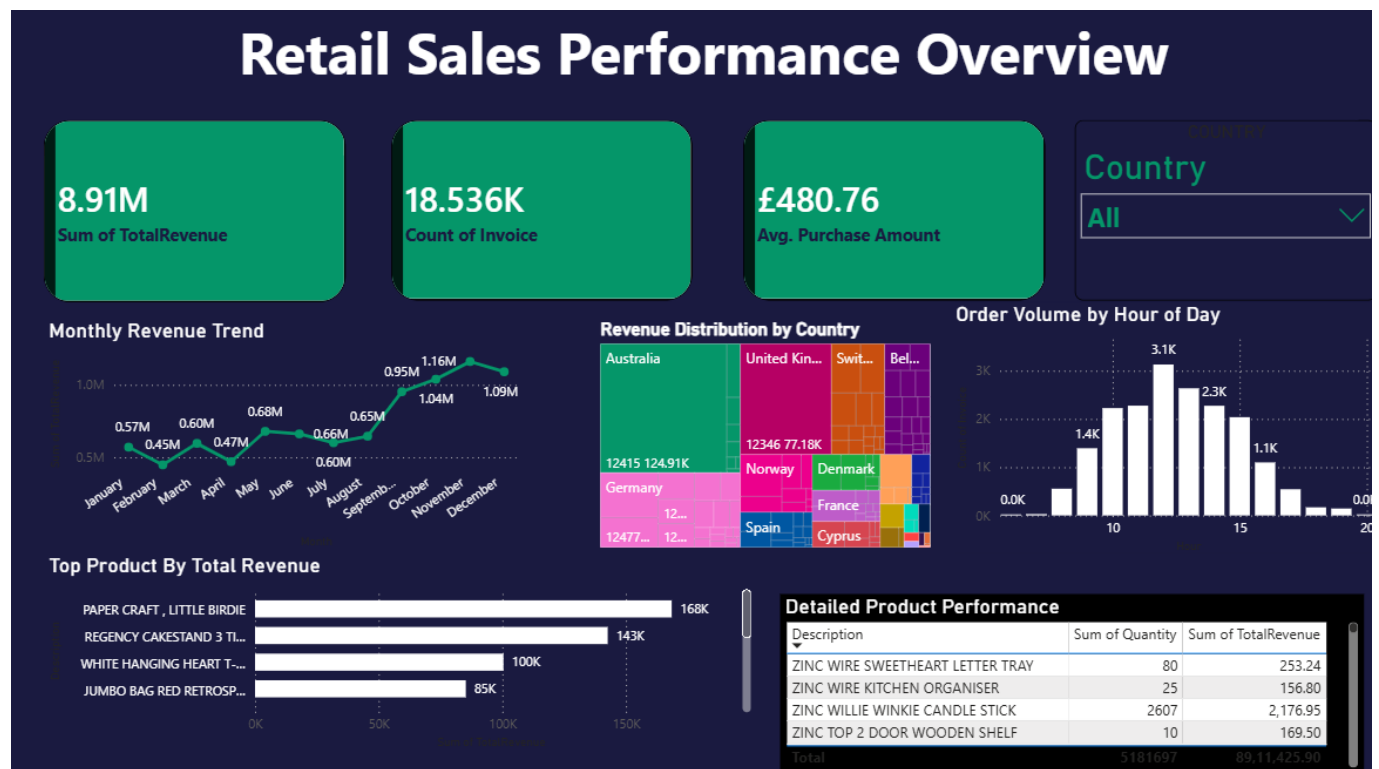
6. Business Recommendations

Based on the findings from the Python EDA, SQL queries, and the interactive Power BI dashboard, the following strategic actions are proposed:

- **Operational Workforce Alignment**: The data shows a consistent daily peak at **12:00 PM (3.1K orders)** and a secondary surge at **3:00 PM (2.3K orders)**. It is recommended to schedule maximum warehouse and customer support staffing during these specific windows to maintain service levels.

- **Seasonal Inventory Stockpiling**: Revenue nearly doubles between **January ($0.57M)** and the **November peak ($1.16M)**. Procurement should prioritize stockpiling high-volume inventory by late September to prevent stockouts during the Q4 holiday rush.

- **High-Value Product Protection**: A small number of items, led by **"Paper Craft, Little Birdie" ($168K)** and **"Regency Cakestand" ($143K)**, drive a disproportionate share of total revenue. These should be classified as "Class A" inventory with dedicated supply chain monitoring.

- **Regional Marketing Focus**: The **United Kingdom** and **Australia** represent the strongest market density. Marketing budgets should be concentrated in these regions to capitalize on existing consumer momentum and high average purchase amounts.

- **Tiered Loyalty Programs**: With an **Average Purchase Amount of $480.76**, the business attracts high-value buyers. Implementing a loyalty program for "Returning" and "Loyal" segments—mirroring the customer segmentation model—will help stabilize revenue during slower Q1/Q2 months.

- **Revenue Optimization via Shipping**: Following the reference study's findings on shipping types, the business should analyze if **Next Day Air** or **Standard** shipping users correlate with higher order values ($480.76+) to offer targeted shipping discounts.