CITY UNIVERSITY
MALAYSIA

FINAL PROJECT – REPORT

"E-COMMERCE DATA ANALYSIS USING BUSINESS INTELLIGENCE TOOLS"

BIT2053

FUNDAMENTALS OF MODERN DATA

(202505 WEDNESDAY 1400-1700)

Prepared By,

No	Name	STUDENT ID
1.	SAM HAROLD KAHCHUNTHIEN	202309010370
2.	JANWELL EDMAXRON JAMES	202309010383
3.	MUHAMMAD SYAHIR BIN ISMAIL	202309010539
4.	AHMAAD ADAM BIN JIKON	202309010389

GitHub Repository :

online_retail_data_analysis

Lecturer :

MR. NAZMIRUL IZZAD BIN NASSIR

Submission Date :

3RD SEPTEMER 2025

TABLE OF CONTENTS

1.0 INTRODUCTION

In the era of digital commerce, organizations generate vast amounts of transaction data daily. Leveraging this data through modern Business Intelligence (BI) tools allows businesses to improve decision-making, optimize operations, and enhance customer relationships.

This project analyses an e-commerce retail dataset to uncover insights into sales performance, customer segmentation, and geographic distribution. By applying data preprocessing, statistical analysis, and visualization techniques, we aim to answer critical business questions and recommend strategies for sustainable growth.

The following business questions were formulated to guide the analysis :

1. Which products drive the highest revenue and sales volume, and how can inventory be optimized to support demand?
2. How do customers differ in terms of value, and which segments should be prioritized for retention and marketing campaigns?
3. Which countries contribute the most to overall revenue, and where should the company focus for potential market expansion?

These questions provided a clear direction for the following data analysis and visualization, ensuring that the insights generated were aligned with business decision-making needs.

2.0 DATASET OVERVIEW

The dataset was sourced from the UCI Machine Learning Repository (also available on Kaggle). It contains records of transactions from a UK-based online retailer between December 2009 and December 2011.

Characteristic	Value
Total Records	1,048,575
Columns	8 (Invoice, StockCode, Description, Quantity, InvoiceDate, Price, Customer ID, Country)
Countries Covered	41
Final Cleaned Records	737,097 (after removing duplicates & handling missing values)

Table 1 : Dataset Characteristics

Preprocessing Steps :

1. Removed 34,150 duplicate entries.
2. Handled missing descriptions and customer IDs.
3. Converted InvoiceDate into datetime format.
4. Derived new features: Total Revenue, Average Order Value and Product Category.

The final cleaned dataset was exported as *"processed_ecommerce_data.csv"* for BI integration.

3.0 DATA ANALYSIS

3.1 REVENUE ANALYSIS

- Total Revenue : £16,273,396.94
- Total Orders : 34,945
- Average Order Value (AOV) : £465.69
- Unique Customers : 5,771
- Unique Products : 4,509

Rank	Product Description	Revenue (£)	Quantity
1	Regency Cakestand 3 Tier	273,694	23,780
2	White Hanging Heart T-Light Holder	229,063	84,966
3	Manual (Gift-related)	149,143	8,560
4	Jumbo Bag Red Retrospot	133,467	73,806
5	Postage	119,624	5,017

Table 2 : Top Products by Revenue

What we saw in table 2 is, a small set of decorative items like cake stands, holders, ornaments, and bags, contributes disproportionately to revenue.

3.2 CUSTOMER ANALYSIS

Customer segmentation was performed using the RFM (Recency, Frequency and Monetary) model, categorizing customers into distinct groups based on their purchasing behaviour. This approach helps identify high-value segments for targeted strategies.

Key metrics from the analysis :

- 5,771 Total Unique Customers.
- 1263 High-Value Customers (RFM 444 and above, classified as "Champions").

Based on the findings, the company has a strong base of repeat customers with high purchase frequency, particularly in the Champions and Loyal Customers segments.

3.3 TIME-BASED ANALYSIS

In our observation of monthly revenue trends from 2010–2011.

- Revenue peaks in October–December, indicating strong seasonality like Christmas and holiday shopping with countless discount offers.
- For example, October 2010 and 2011 revenue exceeded £1.03M, and November 2010 and 2011 revenue both exceeded £1.16M.

Seasonal campaigns, such as Black Friday and Christmas significantly drive sales. Inventory planning must anticipate these peaks.

3.4 GEOGRAPHIC ANALYSIS

Rank	Country	Revenue (£)
1	United Kingdom	13.40M
2	EIRE (Ireland)	592K
3	Netherlands	527K
4	Germany	408K
5	France	337K
6	Australia	169K
7	Spain	100K
8	Switzerland	99K
9	Sweden	91K
10	Denmark	67K

Table 3 : Top 10 Countries by Revenue

The UK dominates sales (84.9%), but European markets such as Netherlands, Germany, and France show strong potential for growth.

4.0 BUSINESS INTELLIGENCE DASHBOARD

An interactive dashboard was developed in Google Looker Studio. See Dashboard Here.

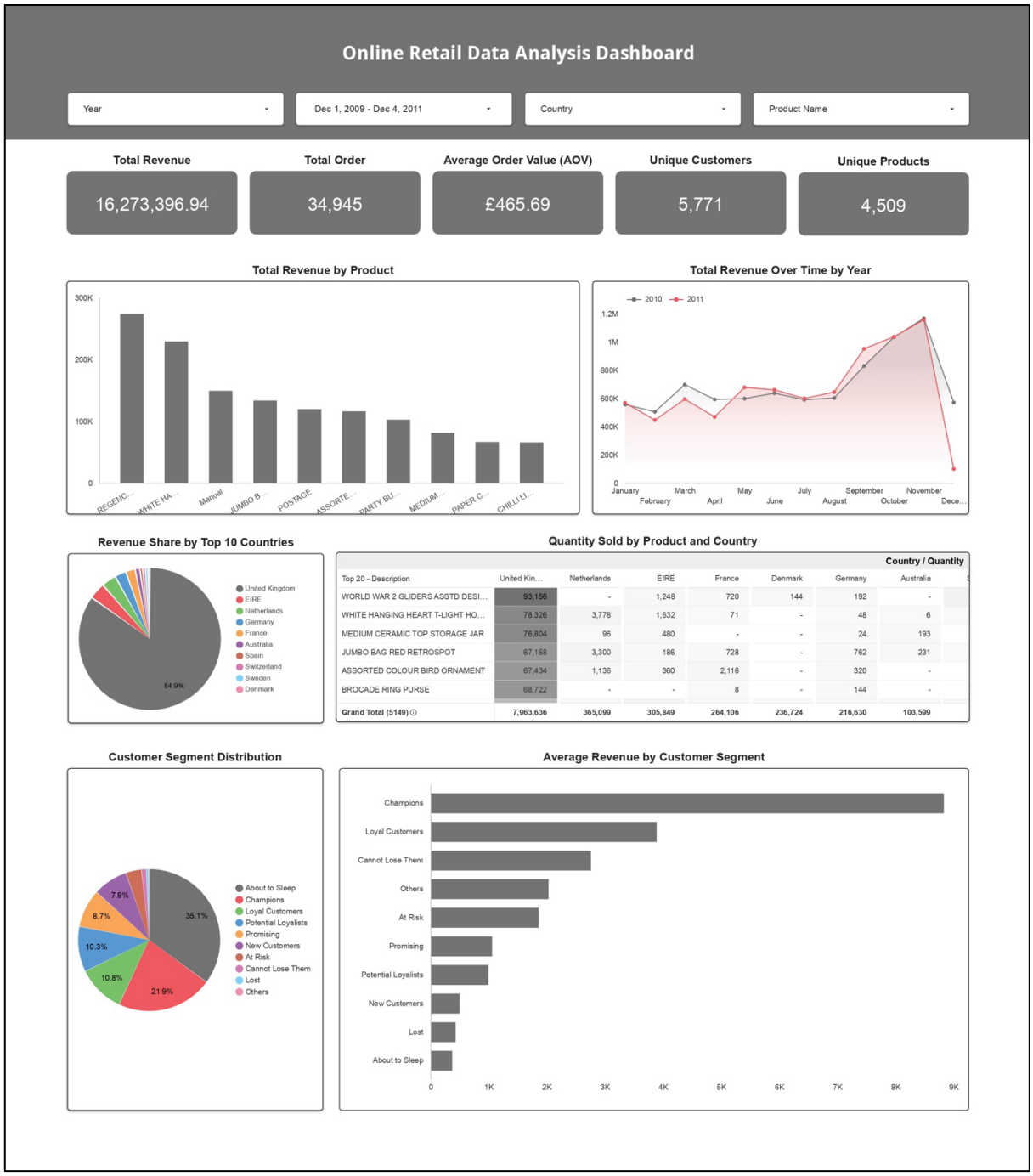

Figure 1 : Online Retail Data Analysis Dashboard

This dashboard enables managers to explore sales trends dynamically and make evidence-based decisions.

5.0 BUSINESS INSIGHTS & RECOMMENDATIONS

The analysis of the e-commerce dataset has revealed several critical insights that can guide the company's strategic decisions. First, the revenue distribution across products shows that a small number of items, such as the Regency Cake stand, the White Hanging Heart T-Light Holder, and the Jumbo Bag Red Retro spot, contribute disproportionately to overall revenue. This suggests that the company should focus on inventory optimization by ensuring that these high-performing products are consistently available, especially during periods of peak demand. Seasonal analysis further supports this, as sales volumes rise sharply in the final quarter of the year, highlighting the importance of aligning inventory management with holiday-driven shopping patterns.

Customer segmentation using the RFM model provides another key area for improvement. The results show a diverse distribution, with "About to Sleep" comprising the largest group at 35.1%, indicating many inactive customers who could be re-engaged. In contrast, there are 414 high-value customers in the "Champions" segment (21.9% of total), who score the maximum across recency, frequency, and monetary value and generate the highest average revenue (around £8,000–£9,000 per customer). These individuals represent the company's most loyal and profitable customers, making them essential to long-term business sustainability. Developing targeted retention strategies, such as loyalty programs, personalized promotions, and exclusive product offers, will help strengthen their engagement while encouraging repeat purchases. At the same time, medium-value segments like "Loyal Customers" (10.8%, ~£6,000–£7,000 average revenue) and "Potential Loyalists" (10.3%, ~£1,500–£2,500) can be nurtured through campaigns that incentivize increased spending, gradually moving them toward high-value status. Lower segments, such as "At Risk" and "Lost," should receive win-back initiatives to prevent further churn.

Geographic analysis indicates that while the United Kingdom is the dominant market, accounting for over 80% of total revenue, there are promising opportunities for international growth. Countries such as the Netherlands, Germany, and France already generate substantial sales and could serve as natural expansion targets. By tailoring marketing campaigns to these regions and optimizing logistics, the company can capture a greater share of international demand. Furthermore, localized promotions and regional partnerships could strengthen brand presence and customer loyalty in these markets.

Together, these insights demonstrate that the company's growth strategy should balance short-term operational improvements with long-term strategic investments. Inventory management, customer relationship programs, and international expansion initiatives will not only sustain profitability but also build resilience against market fluctuations. By adopting a data-driven approach to decision-making, the retailer can ensure that it maximizes opportunities from both existing and emerging markets.

6.0 CONCLUSION

This project demonstrates how modern data analysis and BI tools can transform raw transaction data into actionable insights. By focusing on product performance, customer segmentation, and international markets, the retailer can optimize inventory, strengthen customer loyalty, and expand globally.

The findings highlight the importance of data-driven decision-making in today's competitive e-commerce landscape. Future work may include advanced predictive modelling, such as demand forecasting and churn prediction, to further enhance strategic planning.