INDIVIDUAL WRITING ASSIGNMENT #2

INTRODUCTION

This report evaluates the retail grocery sales on a transactional leve to inform the launch of a new neighborhood store in Rye, East Sussex (UK). The task, using a cleaned data mart derived from the original multi-country retail dataset, is to analyze this data. Some important metrics to gather would be how spending is distributed across countries and customers, which products derive revenue, and how sales evolve over time. The goal is to translate patterns into recommendations for the Rye store to develop further.

The dataset must first be cleaned to remove invalid or irrelevant records and to construct new measures for analysis. Second, there will be several analytical techniques applied using Excel which uses the data provided to create customer-value segmentation by geography, product mix analysis, and time-series trend analysis. These results must be interpreted to provide decisions for launch, merchandising, and stock in Rye.

PROBLEMS (Marketing)

The core problems to bring up involve:

1. Market Focus: Which geographies and customer types will drive spending today and what is needed for the Rye store?

2. Merchandising: Which products must be carried because they generate the most units and revenue and which higher-value items justify promotion?

3. Demand: When does demand spike or dip (taking into account seasonal patterns) and how should inventory and promotions align with these cycles?

DATA DESCRIPTION

a) Source & Fields

The original dataset had the following columns for comparison: InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country. The key fields are the Description (Product), Quantity, Country, TotalPrice (added: Quantity * Unit Price), and CustomerID, StockCode (descending order of importance).

b) Data mart cleaning

To ensure valid analysis, rows were removed: rows with Quantity ≤ 0 (returns/cancellations), UnitPrice ≤ 0, missing CustomerID, Description, or InvoiceNo, and rows with “Unspecified” country. Then, converted InvoiceDate to Excel Date/Time and added Total Price (Quantity * Unit Price). Complete duplicates across all columns were removed as they were most likely due to system error. Approximately 140,000 dirty rows were removed from ~540,000, yielding a consistent data mart on a separate Excel sheet (named ‘Cleaned’).

c) Business Questions

- Which countries contribute the most revenue, and how concentrated is spend by customer within the United Kingdom**?**
- Which products sell most frequently and generate the highest revenue (overall and in the UK), and what share of sales do the "vital few" contribute?
- How does total sales change over time (trend/seasonality), and what timing implications exist for inventory and promotions?

ANALYTICAL TECHNIQUES

Technique 1 – Customer Segmentation by Geography

a) Technique: Using Excel PivotTables to aggregate Sum of Quantity by Country to solve which countries are the biggest buyers. Then, grabbing each CustomerID by Quantity within the United Kingdom (top spender by far) by quantity purchased and amount spent.

b) Why: This identifies where demand is the strong and within the UK, which customer segments drive spending and to see which individual customers are important to satisfy.

c) "Customer Segmentation"

Technique 2 – Product Mix Analysis

a) Technique: By analyzing the product frequency, by description, quantity, and total price to rank the top purchased items and which ones are important to total revenue.

b) Why. This classification guides core merchandising items that must be stocked at all times to reduced revenue loss by seeing the demand drive.

c) "Product Frequency"

Technique 3 – Time-Series Trend Analysis

a) Using another PivotTable that divides the time into quarter, month, and year and Total Price to evaluate the sales over the year period of time.

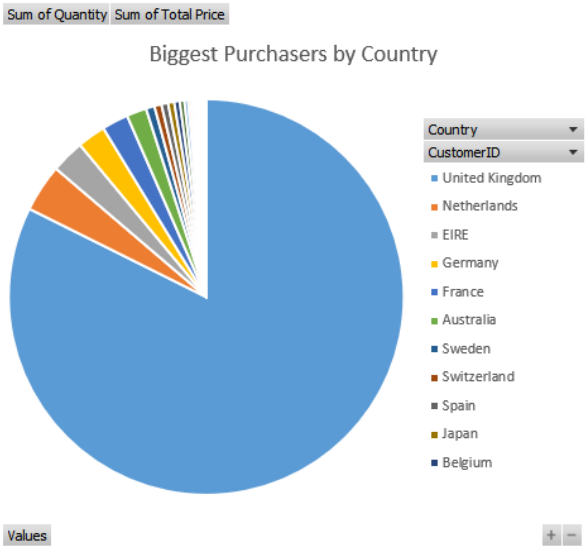
b) Why. Trend and seasonality inform timing of inventory buys and promotions (pre-holiday ramp, weekday peaks)

c) "Time Series Analysis"

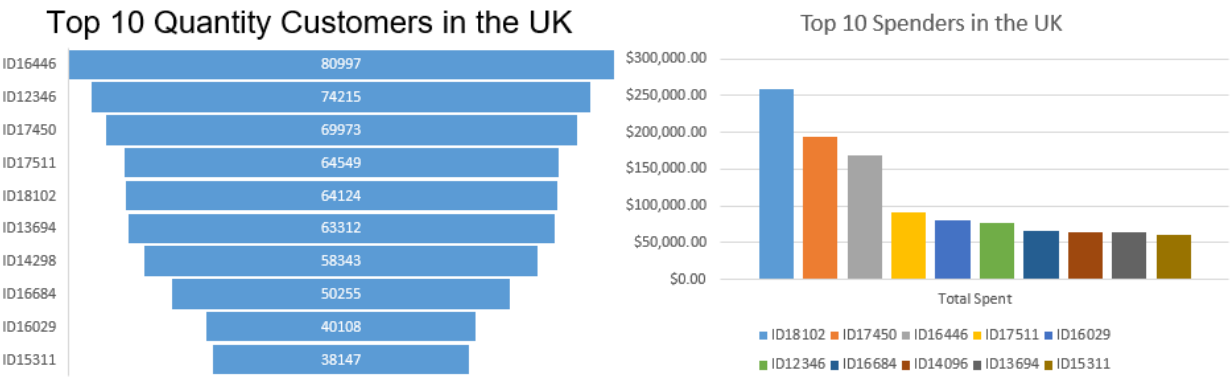
ANALYSIS RESULTS

Technique 1 – Customer Segmentation by Geography

Quantity Total = 5,150,217 | Total $$$ = $8,884,548

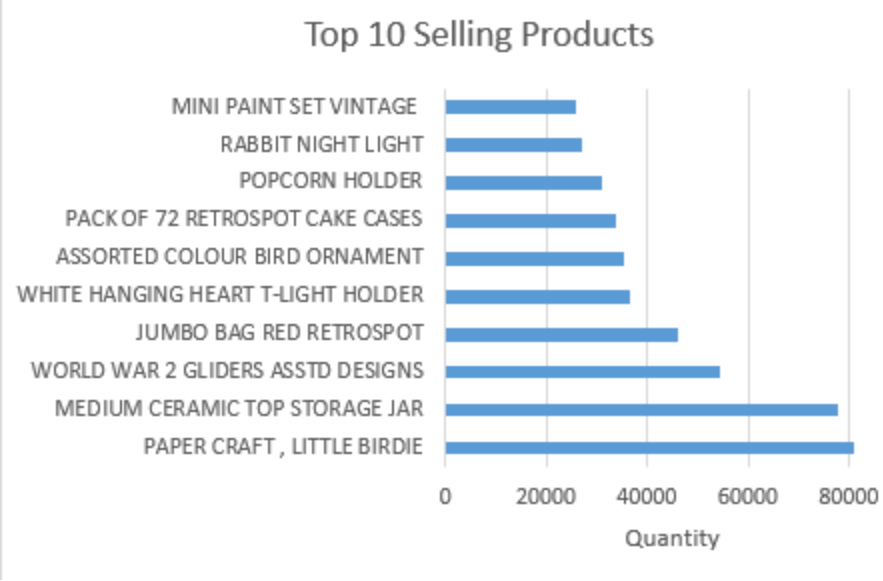


This first chart represents Sum of Quantity by Country, which shows the United Kingdom having >80% of purchases along with several other countries in this pie chart!

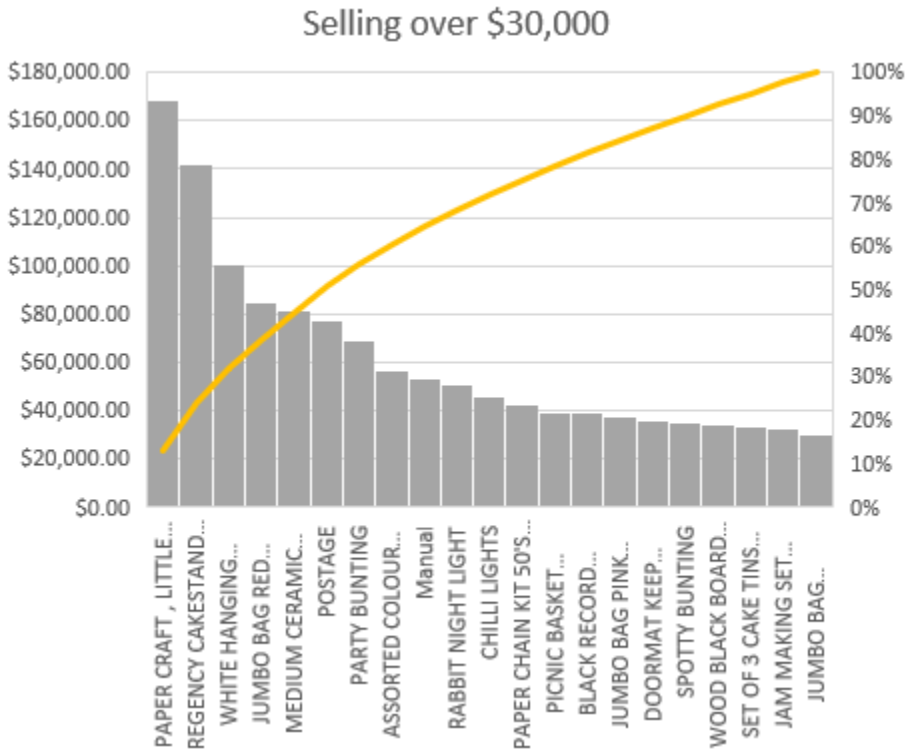


These two charts do a deeper dive into the individual customers within the United Kingdom to see which ones are ordering the most quantity from the country by CustomerID using a Funnel chart. On the right hand side are the top spenders in the UK by CustomerID to see which ones will spend the most using a clustered column chart.

Technique 2 – Product Mix Analysis

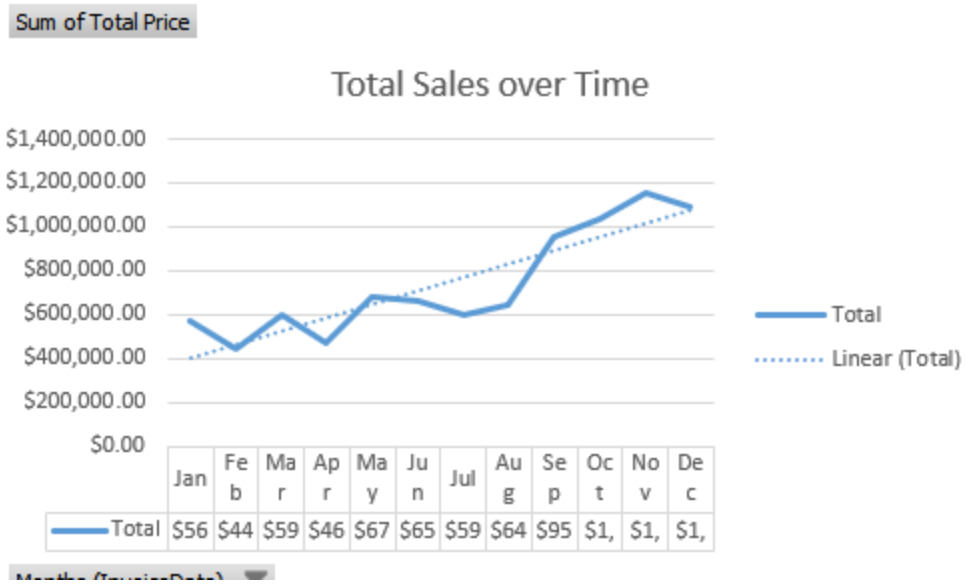


This chart which is using a clustered bar chart and is comparing Description to Quantity to evaluate the top selling products overall in the data.



This chart is a Pareto histogram that grabs the Description and Sum of Total Price and I chose a small sample of products selling over $30,000 to see which ones are earning the most, and on the right is a percentage of that just for a little extra information.

Technique 3 – Time Series Analysis



This is a line chart that is comparing the Months from the InvoiceDate of all products and the Sum of Total price for each of those months. This will help evaluate the Total Sales over the period of data to see seasonality and trends of sales. In addition, it can help show that total sales are growing over time which is very helpful to know when opening a new store.

Month	Sum of Total Price
Jan	$568,101.31
Feb	$446,084.92
Mar	$594,081.76
Apr	$468,075.23
May	$676,502.47
Jun	$659,860.27
Jul	$598,170.72
Aug	$643,520.01
Sep	$950,690.20
Oct	$1,035,642.45
Nov	$1,156,205.61
Dec	$1,087,613.17
Grand Total	**$8,884,548.12**

CONCLUSION

This report transforms raw multi-country transaction data into an actionable plan for opening a grocery store in Rye, UK. After rigorous cleaning (removing invalid quantities/prices, missing keys, unspecified countries, and exact duplicates), three complementary analyses were performed in Excel: customer-segmentation by geography with UK drill-down, product mix

using quantity and revenue, and time-series trend of total sales. Together, these reveal where spending concentrates, which products are essential to carry, and when demand peaks. The recommendations prioritizing top demanded items, focusing promotions on high-value UK customers, and aligning inventory with demand cycles directly support a lower-risk, data-driven launch in Rye.