



ONLINE RETAIL CASE STUDY

CONTENTS

Introduction and Assumptions

- Business Case
- Dataset Overview

1

Data Analysis

2

Data Exploration and Preparation

- Identified Issues
- Approaches to resolve
- Limitations

3

Conclusion

4

Findings and Recommendations

5

INTRODUCTION

The OnlineRetail dataset is from a UK-based retailer specialising in gift and party items. The business serves customers worldwide, including wholesalers.

The dataset consists of the transaction records from 1st Dec 2010 – 9th Dec 2011.

This report explores trends and insights from the dataset, providing recommendations to help drive business growth.



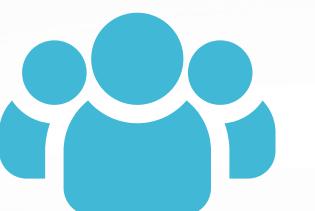
1 YEAR
DATA RANGE



38
COUNTRIES



20K+
ORDERS



4K+
CUSTOMERS

OVERVIEW OF THE DATASET

(BASED ON ASSUMPTIONS)

DATASET SIZE: 8 x Columns
541909 x Rows

USABILITY SCORE: 4.12/10 (Kaggle)

InvoiceNo: 6 digit integer uniquely assigned to each transaction

StockCode: Product item code uniquely assigned to each distinct product

Description: Product name

Quantities: Quantities of each product per transaction

InvoiceDate: Date & Time of the transaction

UnitPrice: Price per unit for a product

CustomerID: 5 digit integer uniquely assigned to each customer

Country: Name of the country the order is from

3

IDENTIFIERS

2

CATEGORICAL

2

QUANTITATIVE

1

DATE-TIME

CASE STUDY ASSUMPTIONS

InvoiceNo:

- Should always uniquely refer to a single transaction.
- Each can have only one associated CustomerID
- All records of an unique InvoiceNo must have the same Country

StockCode:

- Must always refer to a product item uniquely

Description:

- Must always refer to a product name uniquely
- Ideally should be only one for each StockCode

UnitPrice:

- Must always be positive or zero
- Zero unit price refers to free products on offer

Business:

- A transaction must always be positive (no return or refunds)

CustomerID

- Should uniquely refer to a single customer
- If a CustomerID is missing that refers to a guest customer who hasn't signed up

Quantities

- Must always be positive (according to business assumption)

Country:

- A customer can be travelling and ordering from different countries
- An order/InvoiceNo must belong to a single country only

DATA EXPLORATION

(TABLE LEVEL)

RAW Summary:

	Non-NullCount	NullCount	Null(%)	UniqueValues	DataType
InvoiceNo	541909	0	0.00	25900	object
StockCode	541909	0	0.00	4070	object
Description	540455	1454	0.27	4223	object
Quantity	541909	0	0.00	722	int64
InvoiceDate	541909	0	0.00	23260	object
UnitPrice	541909	0	0.00	1630	float64
CustomerID	406829	135080	24.93	4372	float64
Country	541909	0	0.00	38	object

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
count	541909	541909	540455	541909.000000	541909	541909.000000	406829.000000	541909
unique	25900	4070	4223	NaN	23260	NaN	NaN	38
top	573585	85123A	WHITE HANGING HEART T-LIGHT HOLDER	NaN	10/31/2011 14:41	NaN	NaN	United Kingdom
freq	1114	2313	2369	NaN	1114	NaN	NaN	495478
mean	NaN	NaN	NaN	9.552250	NaN	4.611114	15287.690570	NaN
std	NaN	NaN	NaN	218.081158	NaN	96.759853	1713.600303	NaN
min	NaN	NaN	NaN	-80995.000000	NaN	-11062.060000	12346.000000	NaN
25%	NaN	NaN	NaN	1.000000	NaN	1.250000	13953.000000	NaN
50%	NaN	NaN	NaN	3.000000	NaN	2.080000	15152.000000	NaN
75%	NaN	NaN	NaN	10.000000	NaN	4.130000	16791.000000	NaN
max	NaN	NaN	NaN	80995.000000	NaN	38970.000000	18287.000000	NaN

Observations:

- Inappropriate data type for CustomerID, InvoiceDate,
- Null values for Description (0.27%), CustomerID (24.93%)
- Unique values in StockCode are lower than unique values in Description meaning some StockCode(s) may have multiple descriptions.
- There are negative Quantity and UnitPrices
- There are 5,268 (0.97%) duplicate records

Actions:

- Dropped duplicates
- Converted InvoiceDate to datetime64 type
- Converted CustomerID to object type
- Added a TotalPrice column by multiplying UnitPrice with Quantity

DATA PREPERATION

(Column LEVEL)

InvoiceNo:

- Checked standard length & format. Dropped anomalies after confirming they're adjustment records

StockCode:

- Checked for overlapping descriptions (same description belonging to multiple StockCode).
- Capitalised all records after confirming the StockCode is case insensitive which reduced the number of unique StockCode values

Description:

- Standardised string format by removing whitespaces, capitalising and removing fullstops.
- Reduced the missing values to 0.08% by imputing from other records with one non-null Description value associated to the same StockCode
- To reduce multiple Description for the same StockCode, used frequency mode method to keep one value that appears the most times.

Quantity:

- Removed outliers after record level investigation
- Removed negative and zero quantities after confirming they're adjustment records

Quantity:

- Checked for any outliers
- Removed negative unit prices after confirming they're adjustment records

CustomerID:

- Unsuccessful attempt to impute missing values by searching for non-null CustomerID value for associated InvoiceNo. (No available CustomerID found)
- Checked and confirmed standard CustomerID length and format
- Imputed the null values with unique dummy IDs (00001, 00002....) for each Country

Country:

- Capitalised, removed whitespaces and replaced abbreviations with full country name
- Confirmed the assumption of a single CustomerID ordering from multiple countries

PREPARED DATASET

	Non-NullCount	NullCount	Null(%)	UniqueValues	DataType
InvoiceNo	526049	0	0.0	20723	object
StockCode	526049	0	0.0	3827	object
Description	526033	16	0.0	3787	object
Quantity	526049	0	0.0	391	int64
InvoiceDate	526049	0	0.0	19048	datetime64[ns]
UnitPrice	526049	0	0.0	1291	float64
CustomerID	526049	0	0.0	4347	object
Country	526049	0	0.0	38	object
TotalPrice	526049	0	0.0	4594	float64

Statistical summary of quantitative columns:						
	Quantity	InvoiceDate	UnitPrice	TotalPrice		
count	526049.000000	526049	526049.000000	526049.000000		
mean	10.435921	2011-07-04 13:30:22.008253952	3.892806	19.742259		
min	1.000000	2010-12-01 08:26:00	0.000000	0.000000		
25%	1.000000	2011-03-28 11:36:00	1.250000	3.900000		
50%	4.000000	2011-07-19 17:17:00	2.080000	9.920000		
75%	11.000000	2011-10-19 11:13:00	4.130000	17.700000		
max	12540.000000	2011-12-09 12:50:00	13541.330000	38970.000000		
std	43.596885		NaN	32.670864	90.349122	

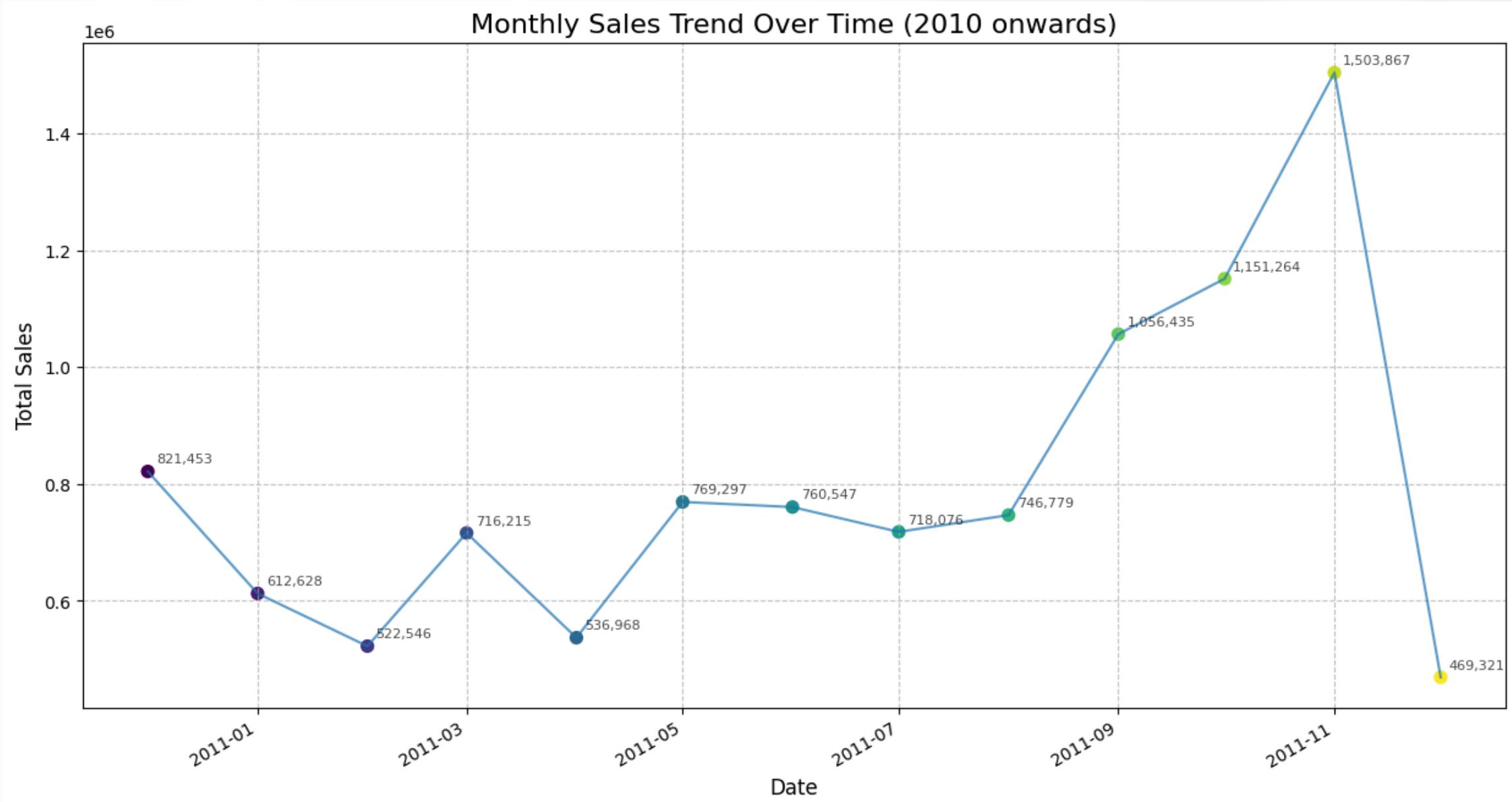
Dataset size: 526049 rows x 9 columns									
	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	TotalPrice
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850	UNITED KINGDOM	15.30
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850	UNITED KINGDOM	20.34
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	2011-12-09 12:50:00	4.15	12680	FRANCE	16.60
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	2011-12-09 12:50:00	4.95	12680	FRANCE	14.85

Limitations:

- There are significant number of special records with alphabetic StockCode (M, D, POST, AMZN, DOT etc) with associated Descriptions such as (DOTCOM, POSTAGE, AMAZONFEE, MANUAL) etc. These indicate transaction information rather than product information. However, as they all account for positive revenues, they weren't removed.
- For the above reason, any product level analysis may introduce a specific transaction type than product type.
- Because of the imputed CustomerIDs, all null CustomerID records will be treated as one customer for each country (Although they can actually belong to multiple customers). They need to be exempted from any customer level analysis to avoid skewed results. Hence, the analysis will not be entirely accurate.
- There are occurrences of same description belonging to multiple StockCode(s) with suffixes (i.e. 123A, 123B). This may indicate product variation across region.
- The NULL quantity and unit price may actually mean return or cancelled orders. But most of them had Descriptions such as "Missing", 'Lost', 'Wet' etc which indicates stock adjustments than actual transactions.

DATA ANALYSIS

```
SELECT
    EXTRACT(YEAR FROM InvoiceDate) AS Year,
    EXTRACT(MONTH FROM InvoiceDate) AS Month,
    ROUND(SUM(TotalPrice), 2) AS TotalRevenue
FROM `custom-blade-407700.online_retail.online-retail-cleaned` OnlineRetail
GROUP BY Year, Month
ORDER BY Year, Month
```

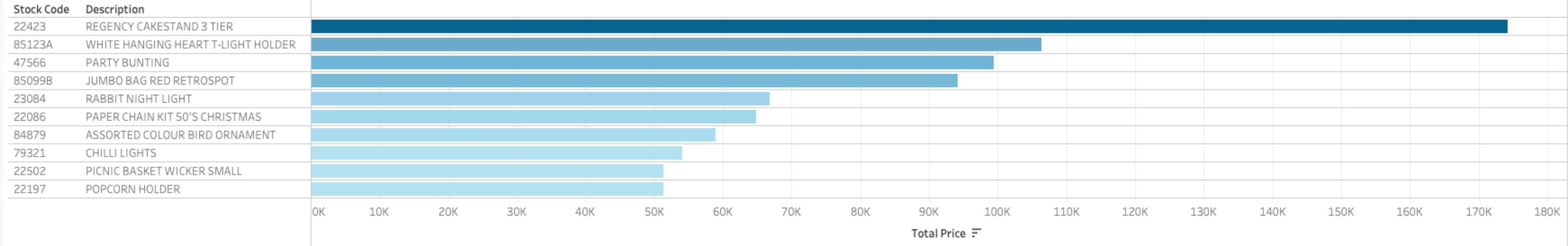


- The **lowest** monthly sale records was in **Feb 2011** with a total revenue **522546** and the **highest** was in **Nov 2011** with a total revenue of **15038667**
- The month on month change from Dec to Nov is the highest recorded percentage. This is most likely due to the lack of data in December. (Data available till 9th of December only)

TOP 10 PRODUCTS

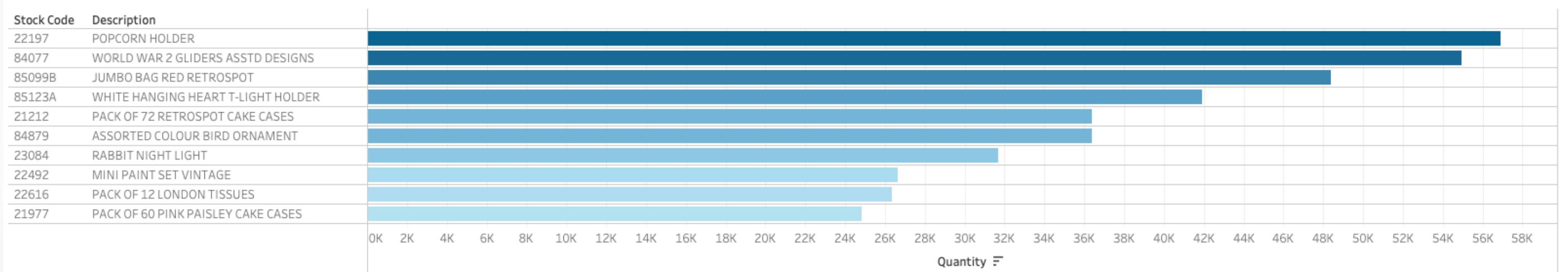
BY REVENUE

Top 10 Products by Revenue



BY QUANTITY SOLD

Top 10 Products by Quantity Sold



AOV PER COUNTRY

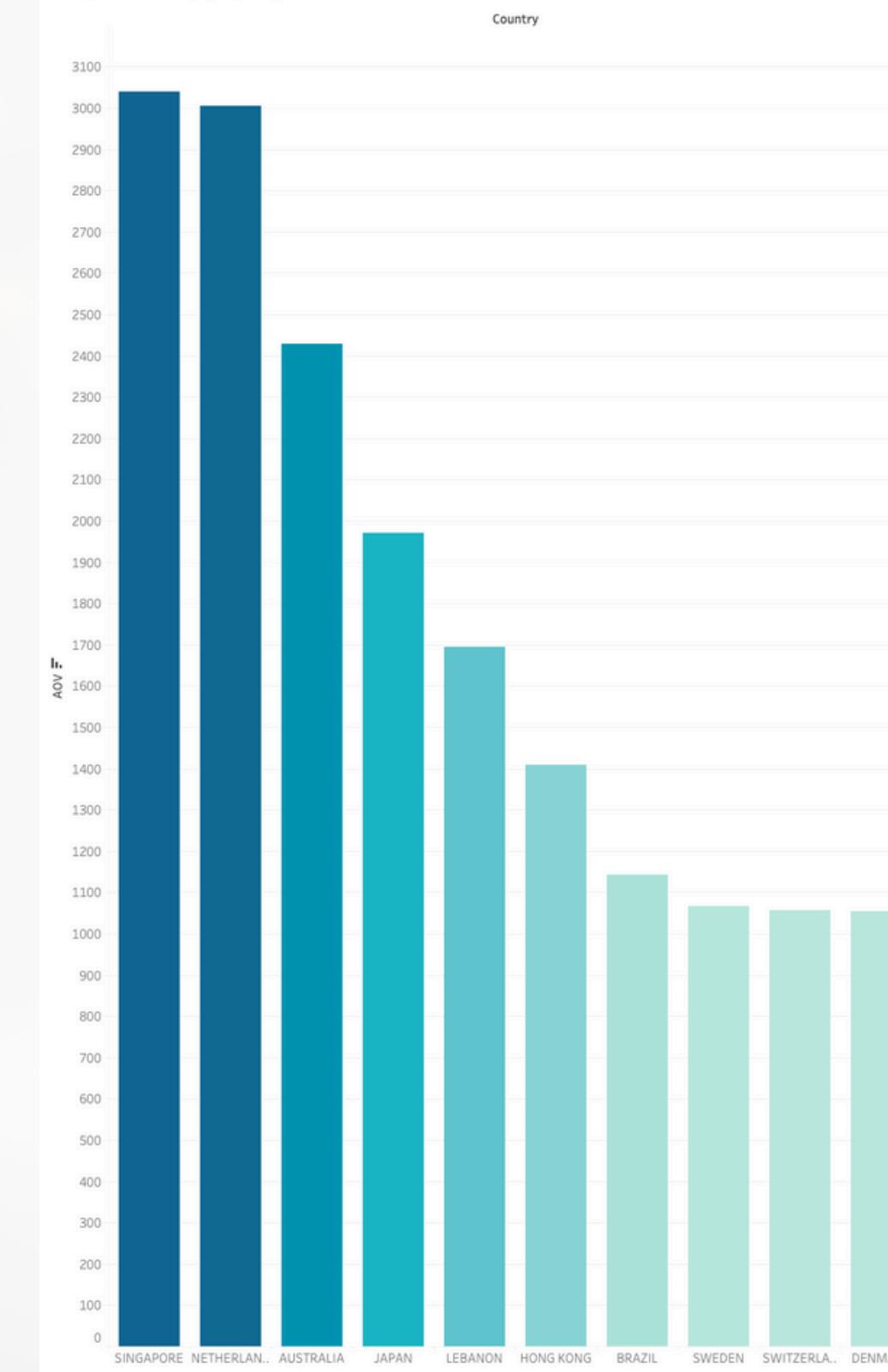
(Average Order Value)

```
--Method 2 (without CTE):
SELECT
    Country,
    ROUND(SUM(TotalPrice)/COUNT(DISTINCT InvoiceNo), 2) AS AOV
FROM `custom-blade-407700.online_retail.online-retail-cleaned` OnlineRetail
GROUP BY Country
ORDER BY AOV DESC
```

AOV per Country

Country	AOV
SINGAPORE	3,040
NETHERLANDS	3,005
AUSTRALIA	2,429
JAPAN	1,969
LEBANON	1,694
HONG KONG	1,408
BRAZIL	1,144
SWEDEN	1,066
SWITZERLAND	1,057
DENMARK	1,053
ISRAEL	1,016
NORWAY	1,005
SOUTH AFRICA	1,002
IRELAND	983
GREECE	952
CYPRUS	844
CHANNEL ISLANDS	786
UNITED STATES	716
SPAIN	684
UNITED ARAB EMIRATES	634
ICELAND	616
CANADA	611
AUSTRIA	600
PORTUGAL	581
FINLAND	550
MALTA	545
FRANCE	535
GERMANY	500
UNITED KINGDOM	466
ITALY	460
BELGIUM	420
LITHUANIA	415
CZECH REPUBLIC	413
POLAND	386
UNSPECIFIED	365
EUROPEAN COMMUNITY	325
BAHRAIN	251
SAUDI ARABIA	146

AOV per Country (Top 10)

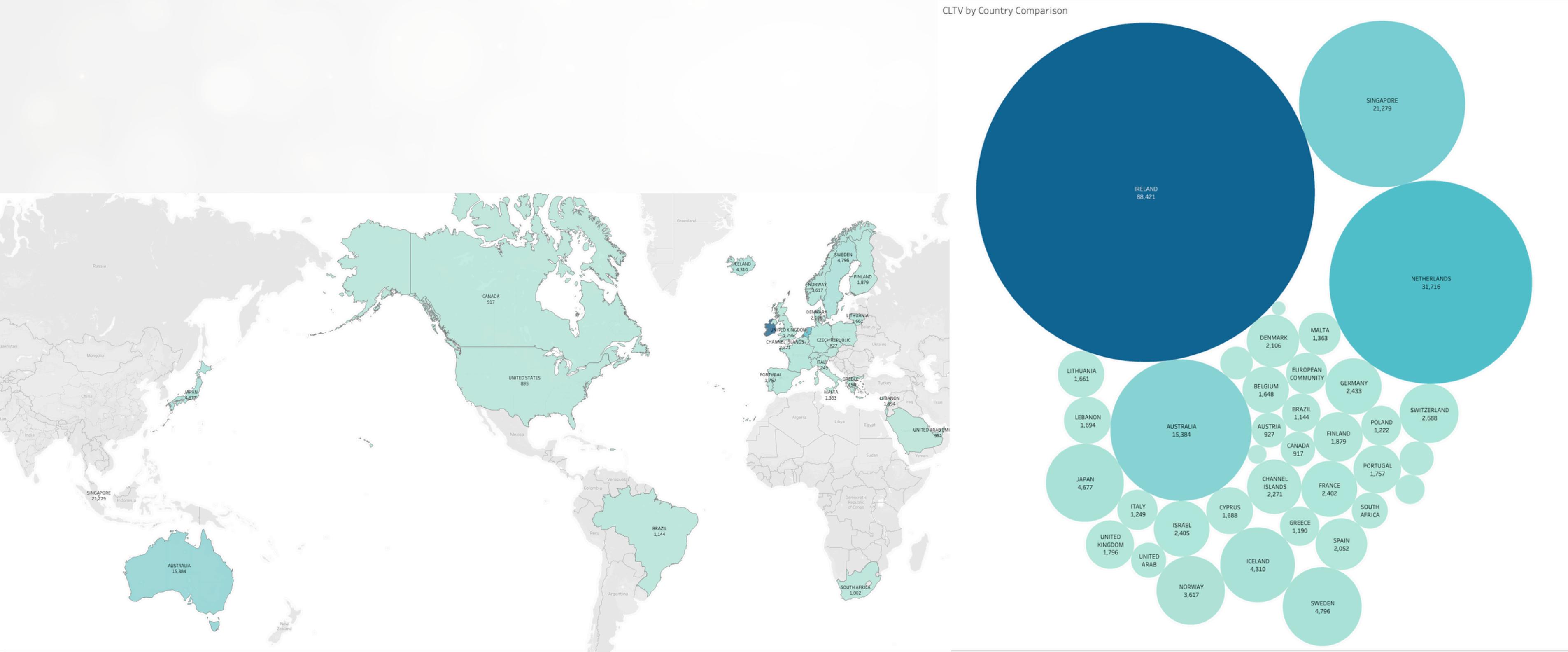


AOV per Country

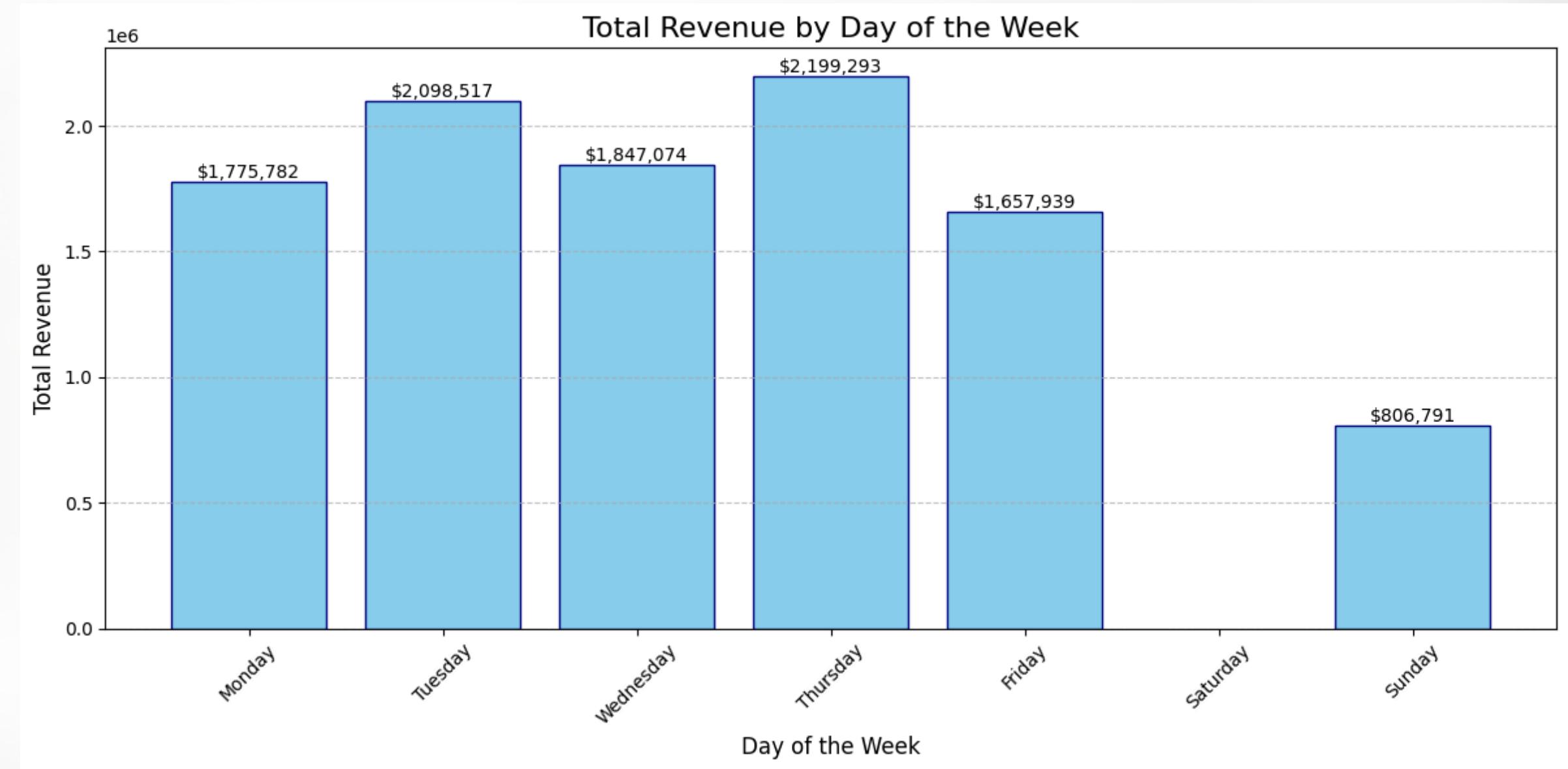


CLTV PER COUNTRY

(Customer Lifetime Value)



DAY OF THE WEEK TOTAL REVENUE



INSIGHTS

Seasonal Trends

- November consistently emerges as the peak sales month, likely due to holiday shopping.
- Strong performance in October and September indicates a prolonged pre-holiday sales period.
- Early months of the year, particularly February, show a consistent dip in sales.
- A noticeable sales decline is observed in April, following the first quarter.

Product Performance

- A small number of key items contribute significantly to overall sales, indicating a concentrated revenue stream.
- Shipping-related items feature prominently in top sales, highlighting the importance of e-commerce.
- Home decor, seasonal items, and crafts show strong consumer interest.
- The diverse range of top-selling products suggests a broad appeal across various customer segments.

Channel Insights

- The prominence of postage-related items in sales figures underscores the significance of online sales channels.

Strategic Implications

- The consistent seasonal pattern presents clear opportunities for targeted marketing and inventory management.
- The diversity in product popularity indicates potential for cross-selling and developing product bundles.
- The post-holiday sales dip suggests a need for strategies to stimulate sales during slower periods.
- The strong performance of online channels highlights the importance of optimizing the e-commerce experience.

RECOMMENDATIONS

- November Sales Boost: Leverage November's consistent high sales by ramping up marketing efforts. Design compelling promotions, such as tiered discounts or exclusive bundles, to maximize revenue during this peak season.
- Proactive Inventory Management: Implement a robust inventory forecasting system to ensure adequate stock levels before November's sales surge. This preemptive approach will minimize lost sales opportunities due to stockouts.
- Pre-November Customer Engagement: Develop a multi-channel outreach strategy, focusing on personalized email campaigns and interactive social media content, to build anticipation for November sales events.
- January Recovery Plan: Counter the post-holiday slump by launching innovative "New Year, New Deals" promotions. Consider clearance events or "customer appreciation" sales to stimulate January purchases.
- February Sales Revitalisation: Inject excitement into February by introducing new product lines or limited-edition offerings. Use this as an opportunity to refresh your product catalog and re-engage customers.
- Customer Loyalty Enhancement: Design and implement a dynamic loyalty program that offers escalating rewards. This could include tiered benefits, exclusive access to new products, or personalised discounts to maintain sales momentum throughout the year.
- Push personalised offers on the weekends to drive sales.



THANK YOU