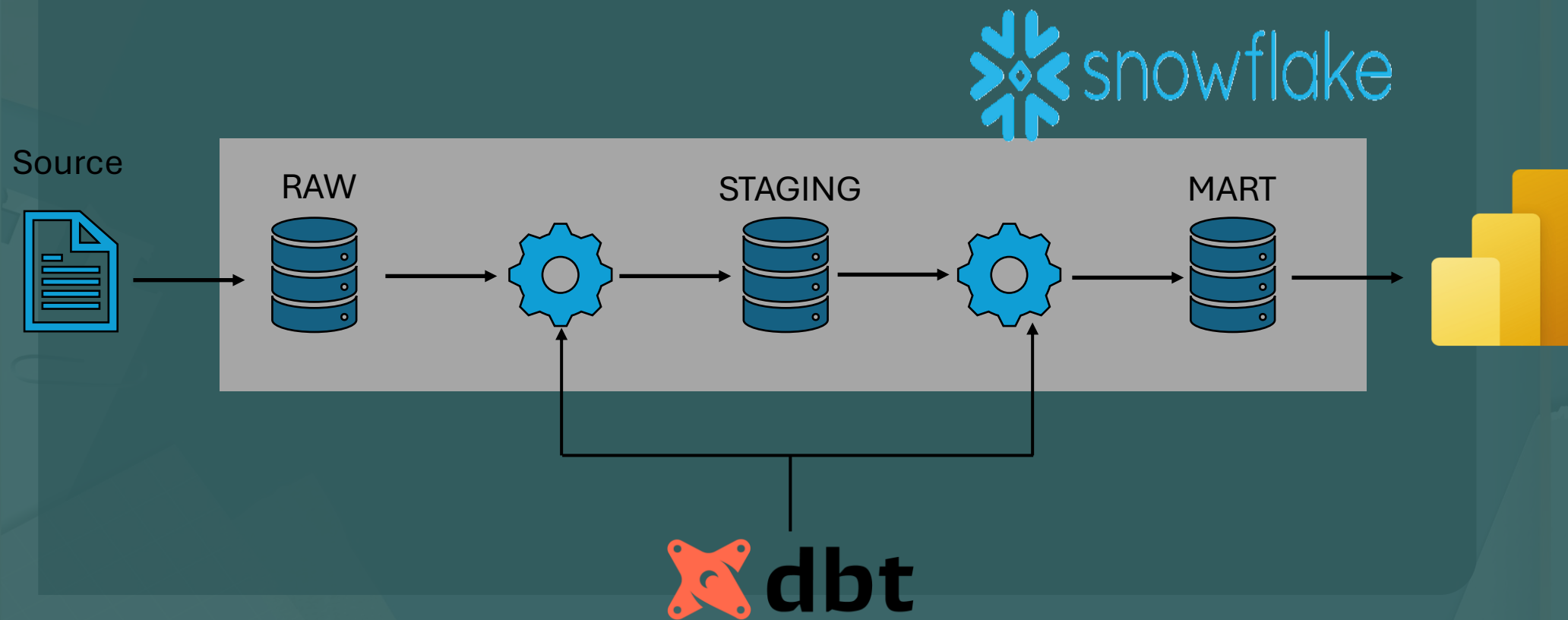# Online Retail Project

BY/ Mahmoud Abdelmoty

Project Overview

Data Set

This Online Retail II data set contains all the transactions occurring for a UK-based and registered, non-store online retail between 01/12/2010 and 09/12/2011.The company mainly sells unique all-occasion gift-ware. Many customers of the company are wholesalers.

Content

Attribute Information:

- Invoice_No: Invoice number. Nominal. A 6-digit integral number uniquely assigned to each transaction. If this code starts with the letter 'c', it indicates a cancellation.
- Stock_Code: Product (item) code. Nominal. A 5-digit integral number uniquely assigned to each distinct product.
- Description: Product (item) name. Nominal.
- Quantity: The quantities of each product (item) per transaction. Numeric.
- Invoice_Date: Invice date and time. Numeric. The day and time when a transaction was generated.
- UnitPrice: Unit price. Numeric. Product price per unit in sterling (Â£).
- CustomerID: Customer number. Nominal. A 5-digit integral number uniquely assigned to each customer.
- Country: Country name. Nominal. The name of the country where a customer resides.

Problem Definition

The Online Retail dataset represents **real-world transactional data** from an online UK-based giftware store operating between December 2010 and December 2011. Despite being rich and useful, the dataset contains **several data quality challenges** that can directly affect analytics and business insights:

• Transactions with:
  • Missing CustomerID → cannot identify who made the purchase
  • Cancelled invoices (InvoiceNo starts with 'C') → affect true sales numbers
  • Quantity=0 or UnitPrice=0 → invalid or incomplete transactions
• Inconsistent product descriptions (mixed casing, typos, or unwanted text like "damaged".
• Duplicate or inconsistent product codes (StockCode)
• Need to distinguish between **completed orders** and **returned / cancelled orders**

Data Warehouse Creation

1. Create Stage – Upload raw dataset for access
2. Schemas:
   - Raw – Store raw data
   - Staging – Cleaned & transformed data
   - Mart – Final star schema (facts & dimensions)
3. Process:
   - Clone data from stage to `Raw` schema
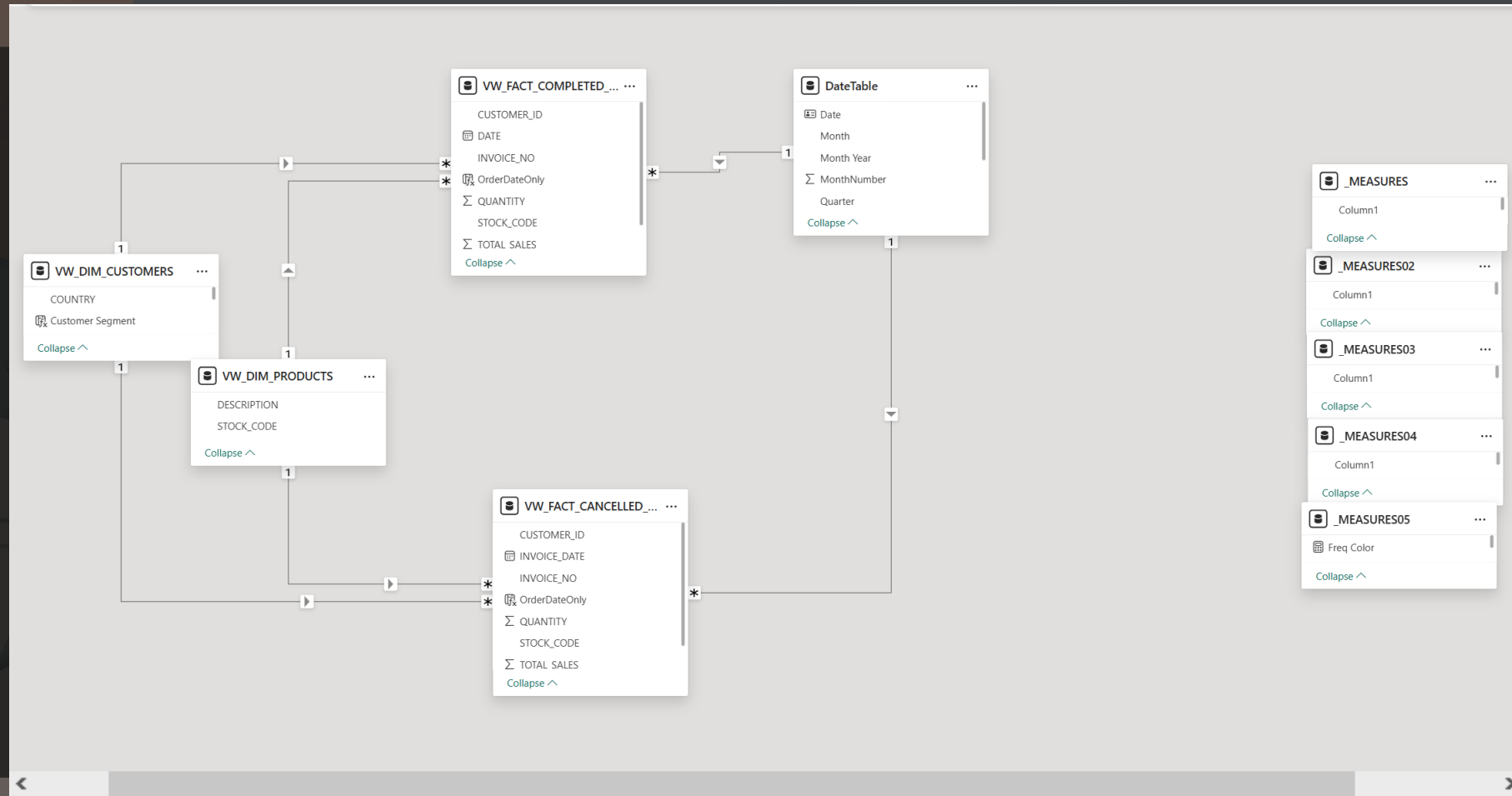   - Connect dbt to Snowflake for transformations

Data Transformation (dbt)
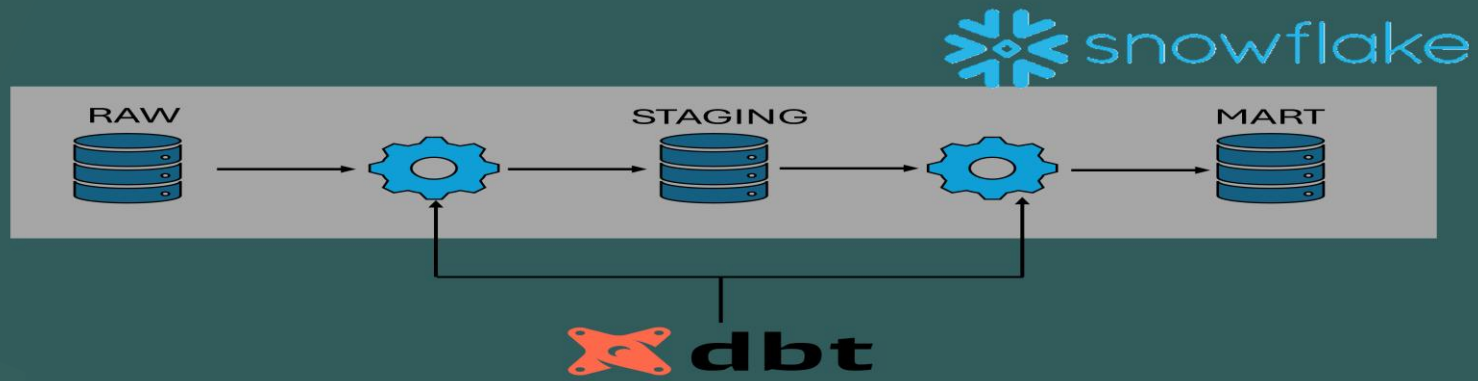
Power BI Dashboard

Data Modeling

Completed Orders Dashboard

- KPIs:
  - Number of Customers
  - Number of Completed Orders
  - Total Sales
  - Number of Sold Products
- Visuals:
  - MTD Completed Orders (line chart)
  - MTD Sold Quantity (line chart)
  - Top 10 sold products & % of total price (table)
  - Map with:
    - Total sales by country
    - Top 5 products per country (tooltip)
    - Unique products sold, total quantity per country

Online Retail Dashboard

4,359
#NoCustomer

23,139
#NTransaction

£10M
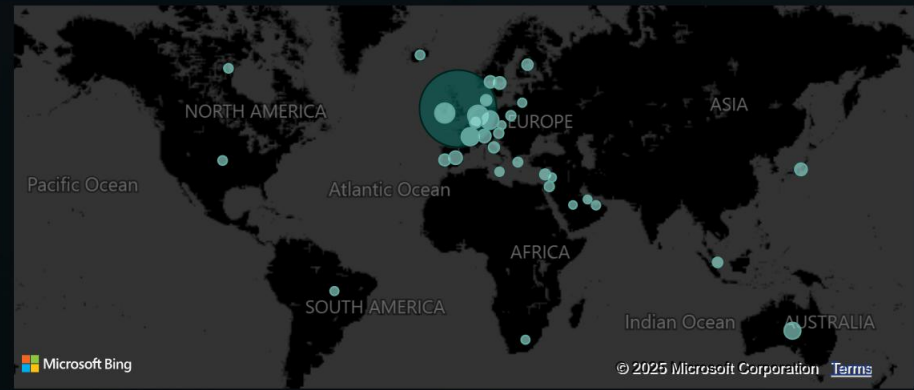totalCompletedSales

3,799
#NoProduct

14.6%
CancellationRate

CANCELLED

N Completed Orders MTD

41,237
34,097
26,911
35,562
28,902
35,853
35,716
38,160
34,253
48,932
59,010
82,992
25,020

| Dec 2010 | Jan 2011 | Feb 2011 | Mar 2011 | Apr 2011 | May 2011 | Jun 2011 | Jul 2011 | Aug 2011 | Sep 2011 | Oct 2011 | Nov 2011 | Dec 2011 |

total qty Consumed by Month Year

358,481
387,167
283,062
376,867
308,093
394,269
388,548
399,702
420,977
569,367
621,150
749,257
313,205

| Dec 2010 | Jan 2011 | Feb 2011 | Mar 2011 | Apr 2011 | May 2011 | Jun 2011 | Jul 2011 | Aug 2011 | Sep 2011 | Oct 2011 | Nov 2011 | Dec 2011 |

STOCK_CODE	totalCompletedSales	Rank Product	%of Product From Sales
22423	£174,459	1	1.7%
23843	£168,470	2	1.6%
85123A	£106,471	3	1.0%
47566	£99,504	4	1.0%
85099B	£94,340	5	0.9%
23166	£81,701	6	0.8%
23084	£66,938	7	0.7%
22086	£64,952	8	0.6%
84879	£59,058	9	0.6%
79321	£54,118	10	0.5%

Microsoft Bing

NORTH AMERICA
EUROPE
ASIA
Pacific Ocean
Atlantic Ocean
AFRICA
SOUTH AMERICA
Indian Ocean
AUSTRALIA

Cancelled Orders Dashboard

- KPIs:
  - Number of Units Returned
  - Number of Cancelled Orders
  - Total Cancelled Price
  - Number of Unique Returned Products
- Visuals:
  - MTD Cancelled Orders (line chart)
  - MTD Returned Quantity (line chart)
  - Top 10 returned products & % of total price (table)
  - Map with:
    - Total returned sales by country
    - Top 5 returned products per country (tooltip)
    - Unique returned products, total units returned

Cancelled Orders Dashboard

3,381
N Cancelled Orders

269,352
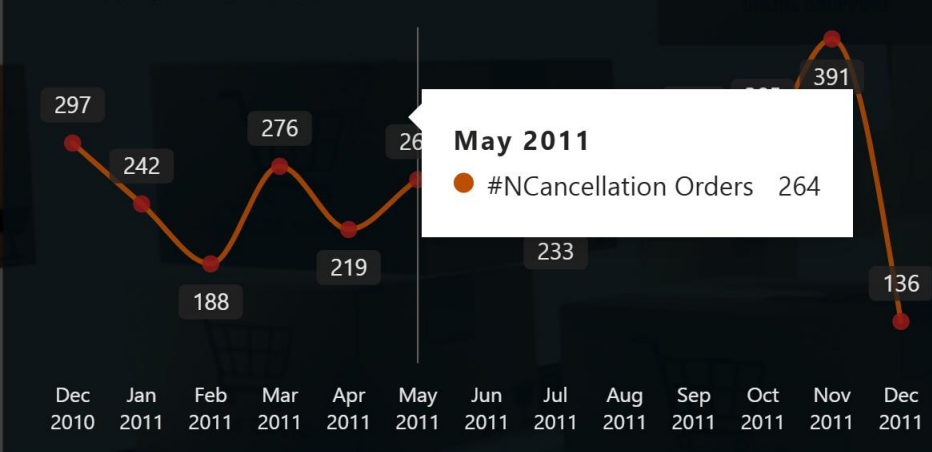Total Units Returned

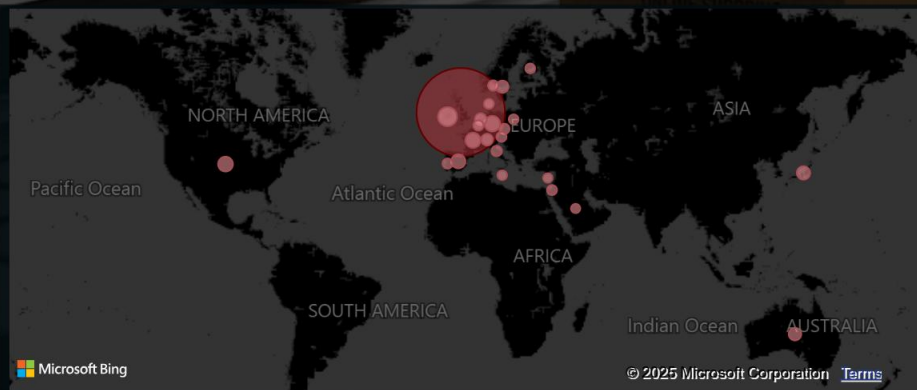**£473,746**
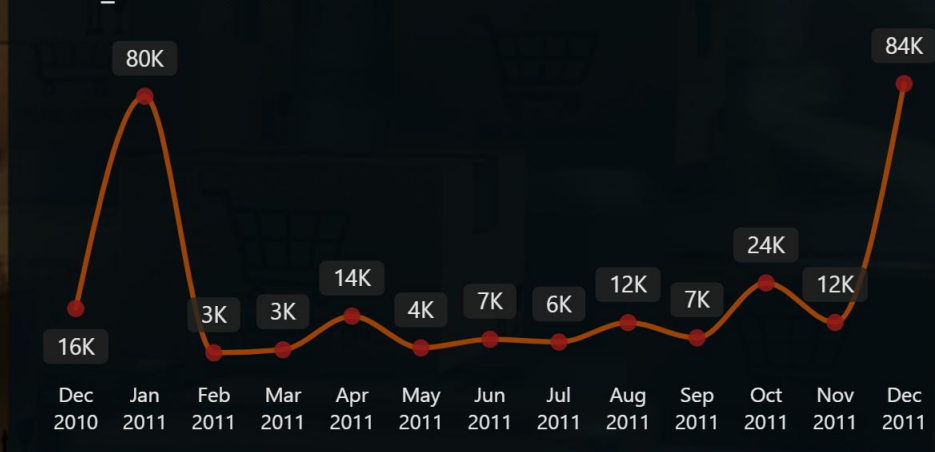Total Cancelled Sales

1,912
N Unique Product

4.42%
Returned Value %

MTD NCancellation Orders

297
242
276
26
188
219
233
391
136

May 2011
● #NCancellation Orders   264

Dec 2010 | Jan 2011 | Feb 2011 | Mar 2011 | Apr 2011 | May 2011 | Jun 2011 | Jul 2011 | Aug 2011 | Sep 2011 | Oct 2011 | Nov 2011 | Dec 2011

MTD N_Unit Returned

80K
84K
16K
3K
3K
14K
4K
7K
6K
12K
7K
24K
12K

Dec 2010 | Jan 2011 | Feb 2011 | Mar 2011 | Apr 2011 | May 2011 | Jun 2011 | Jul 2011 | Aug 2011 | Sep 2011 | Oct 2011 | Nov 2011 | Dec 2011

STOCK_CODE	totalCancelledSales	Product C RNK	%Cancelled Product from Sales
23843	£168,470	1	35.56%
23166	£77,480	2	16.35%
22423	£9,723	3	2.05%
85123A	£6,624	4	1.40%
21108	£6,591	5	1.39%
23113	£4,803	6	1.01%
48185	£4,555	7	0.96%
21175	£3,775	8	0.80%
47566B	£3,693	9	0.78%
22273	£3,513	10	0.74%

NORTH AMERICA | EUROPE | ASIA | Pacific Ocean | Atlantic Ocean | AFRICA | SOUTH AMERICA | Indian Ocean | AUSTRALIA

RFM Analysis Dashboard

- KPIs:
  - Number of Customers
  - Number of Champion Customers
  - Number of At Risk Customers
  - At Risk Revenue
- Visuals:
  - Waterfall chart (Monetary by segment)
  - Bar chart (Number of customers by segment)
  - Histogram (Customer distribution by Recency)
  - Histogram (Distribution by RFM Score)

RFM Analysis Dashboard

COUNTRY

All ▾

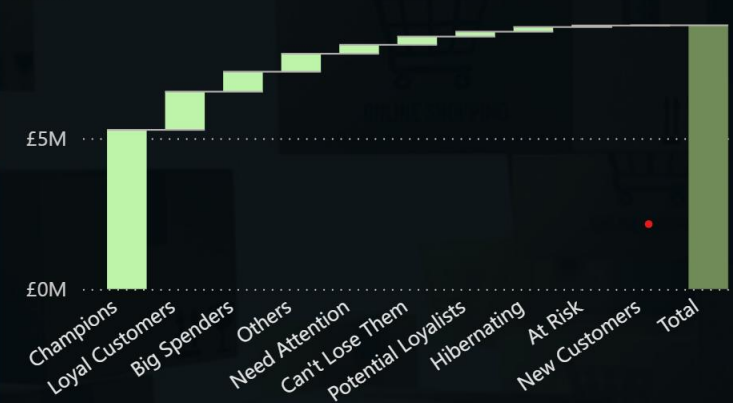**657**
Champion Customers

£49,785
At Risk Revenue

31
At Risk Customers

4,358
N Customers

Monetary by Customer Segment

£10M
£5M
£0M

Champions, Loyal Customers, Big Spenders, Others, Need Attention, Can't Lose Them, Potential Loyalists, Hibernating, At Risk, New Customers, Total

N Customers by Customer Segment

1,097
764
678
657
565
268
210
67
31
21

Need Atte..., Others, Hibernating, Champions, Loyal Customers, Big Spenders, Potential Loyalists, Can't Lose Them, At Risk, New Customers

Distribution of Customers by Recency

600
400
200
0

0   100   200   300   400

#NoCustomer by RFM Score

400
200
0

0   200   400   600

Have more than 180 days not make a purchases
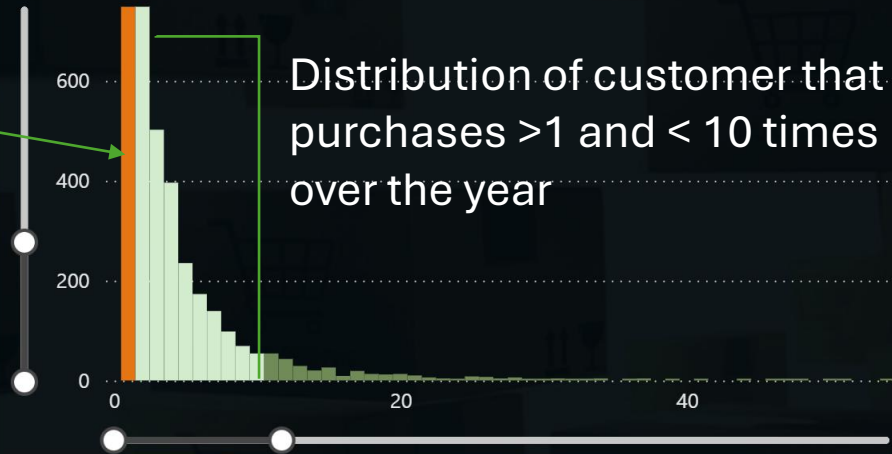
Customer Distribution Dashboard

- Visuals:
  - Histogram (Customer distribution by Frequency)
  - Box plot (Frequency distribution & outliers)
  - Histogram (Customer distribution by Monetary)
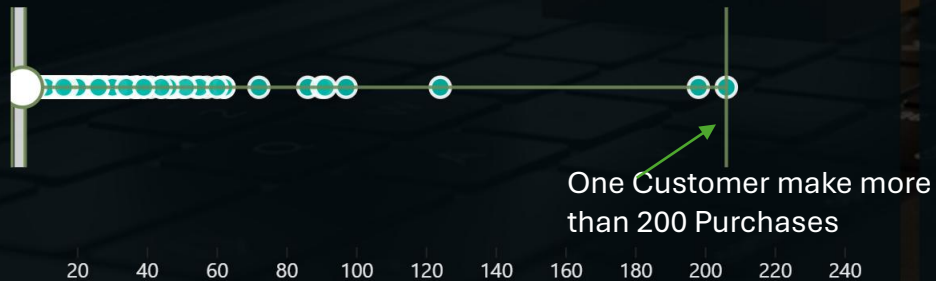 - Box plot (Monetary distribution & outliers)
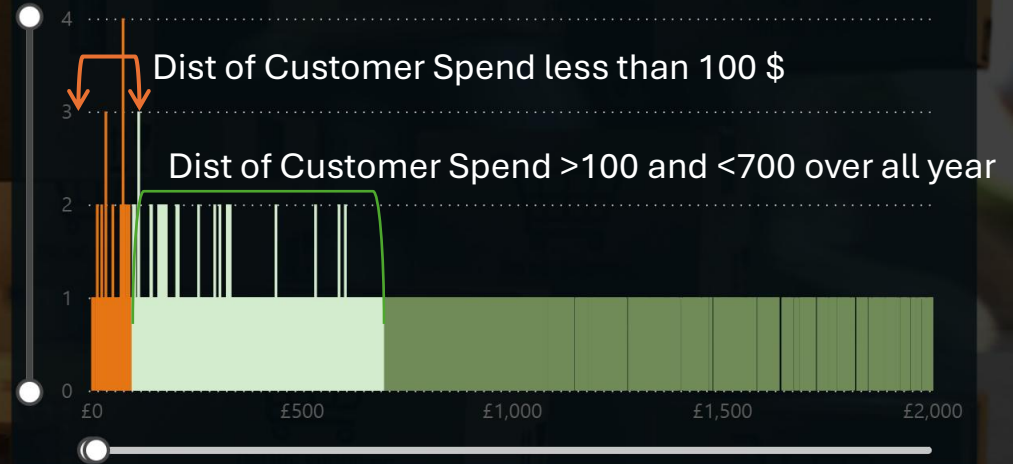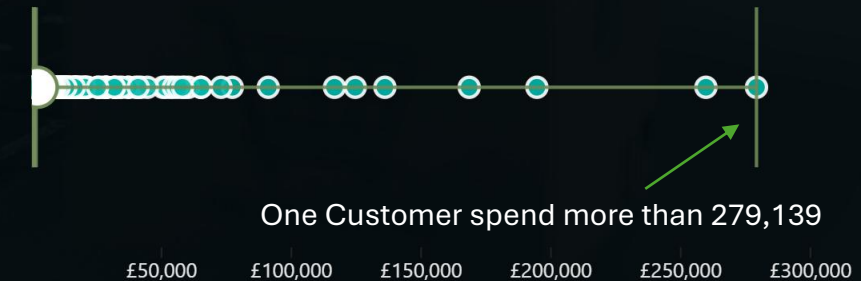
Recommendation

Segment	Who are they?	Goal	Recommended Actions
Champion	Buy frequently, spent a lot, purchased recently	Keep them loyal & make them feel special	• VIP loyalty program • Early access to new products • Invite to exclusive events or beta features • Thank you messages
Loyal	Buy often, but not the top spenders	Encourage to spend a bit more	• Cross-sell & upsell related products • Small loyalty discounts • Personalized product recommendations
Big Spender	Spent a lot, maybe not recent or frequent	Increase recency/frequency	• Special offers valid for short time • Appreciation messages highlighting total spend • Premium customer support
Potential Loyalist	Purchased recently, not frequent yet	Convert to repeat buyers	• Welcome journey emails • Discounts for next purchase • How-to-use content or product tips
New Customers	Just made first purchase	Encourage second purchase quickly	• Welcome discount or free shipping on next order • Thank you email introducing brand story • Highlight best-sellers or customer favorites
Need Attention	Used to buy, recency and/or frequency dropping	Bring them back before churn	• Personalized "We miss you" campaigns • Win-back offers • Feedback survey to know why
Can't Lose	Used to be very good customers, now inactive	Prevent churn	• Strong reactivation campaigns • Larger loyalty reward or special incentive • Personal call from sales/account team
At Risk	High monetary but haven't bought recently	Reactivate them	• Limited time discount • Reminder of what they used to buy • Show new arrivals similar to previous purchases
Other	Low recency, frequency, monetary	See if can activate or clean from list	• Occasional broad campaigns • Consider excluding from expensive campaigns • Run survey to know why they didn't engage

Thank you