Sales Data Analysis

Alina Siddiqui 26555
Rafsha 26639
Muhammad Ahsan 27134

1. Problem Identification and Dataset Background

Problem Statement

In today's fast-paced retail world, businesses operate across an intricate network of sales channels, regions, warehouses, and customer segments. As online and in-store commerce converge, companies face a difficult balancing act: drive revenue growth through promotions and global expansion — without losing profits to returns, fulfillment inefficiencies, or poorly optimized campaigns.

This complexity is precisely what our analysis seeks to untangle. By leveraging business intelligence tools, we aim to create a unified view of the company's sales, returns, discounts, and logistics — to answer a critical business question:

> **"How can we optimize sales performance and reduce returns across countries and channels in an online retail business?"**

This guiding statement encapsulates the broader pain point: the company is generating substantial revenue, but performance is **fragmented** — and decisions are often made in silos. For example:

- While **mid-level discounts (10–30%)** drive strong sales, higher discounting does **not** yield proportional benefits.
- **Return rates** hover around 10% overall — but **spike** in high-priority orders, specific product categories (e.g., *Furniture*), and certain countries (*Australia*, *Germany*).
- **Shipping costs** vary widely across providers and countries, and **do not correlate** directly with sales output — hinting at inefficiencies in logistics planning.

Without a centralized BI solution to track these KPIs across dimensions like region, category, and channel, the company risks losing both **profitability** and **strategic clarity**.

Our solution uses interactive dashboards to consolidate and visualize this complexity — empowering managers to not only monitor trends but also make data-driven decisions across pricing, operations, and marketing.

Main Business Questions

To ensure our dashboards provide actionable insights, we structured our analysis around four key questions:

1. **Which countries and sales channels deliver the highest revenue and growth potential?**

2. **What discount ranges maximize sales volume while minimizing returns and margin erosion?**

3. **Are return rates linked to specific payment methods, product categories, or customer segments?**

4. **How can we reduce shipping costs while maintaining fulfillment quality and customer satisfaction?**

5. **What emerging sales trends and risks should we anticipate for the coming year?**

Each of these guided our BI dashboard structure and KPI selection.

Dataset Overview

Dataset Name: Online Sales Dataset

Description:

The original dataset consists of **49,782 transactional records** from a global retail company, spanning a 5-year period from **January 2020 to May 2025**. It captures sales made through both **in-store and online** channels, and includes a mix of product, customer, transaction, shipping, and return-related attributes.

However, as expected with real-world data, the raw dataset contains a number of issues such as:

- **Missing values** in key fields like ShippingCost, CustomerID, and WarehouseLocation
- **Negative or invalid entries** in UnitPrice and Quantity
- **Inconsistent formatting** or incomplete entries in some categorical fields
- Lack of engineered fields needed for deeper analysis, such as TotalSales or time breakdowns

This version served as the **starting point** for our business intelligence project. Before performing any analysis or dashboard development, a thorough **data cleaning and transformation** process was necessary to ensure accuracy and usability.

Industry Context

The challenges faced in this dataset reflect broader struggles in the **global retail and e-commerce industry**:

- **Returns** cost retailers billions annually, especially when concentrated in certain SKUs or regions.
- **Shipping and logistics** are often managed independently from sales, leading to cost-revenue mismatches.
- **Promotions**, if not optimized, can drive short-term sales but reduce margins or increase return risk.

In this context, Business Intelligence dashboards provide the **visibility, flexibility, and decision support** necessary to stay competitive. Our solution is designed to align the company's operations with real-time performance insights — guiding better promotions, smarter shipping strategies, and more precise inventory targeting.

Dashboard Structure

Dashboard Name	Focus
1. Executive Overview	KPIs (Sales, Returns, AOV), Time trends, Sales by Country/Channel
2. Returns Breakdown	Return rates by Product, Channel, Country, and Payment Method
3. Discount Analysis	Discount vs Sales trends, Return behavior across discount bands/categories
4. Shipping Cost Impact	Shipping costs by region and provider, correlation with sales

5. Category Performance	Category-wise sales, return rates, and sales per SKU
6. Recommendations	Data-driven actions based on combined insights from all dashboards (e.g., cap high discounts, optimize shipping, reduce returns in target segments)

2. Data Cleaning and Exploratory Data Analysis (EDA)

This section describes the preprocessing and analytical steps taken to transform the raw online sales dataset into a refined, analysis-ready format. The goal was to ensure data quality, extract useful features, and uncover patterns that would inform the development of actionable business intelligence dashboards.

Data Quality Assessment and Cleaning

Missing Values

Three key columns were found to contain missing values:

- **ShippingCost**:
  Missing in a sizable portion of records. These were imputed using **K-Nearest Neighbors (KNN)** imputation with k=5, leveraging related numerical features such as Quantity, UnitPrice, and Discount to estimate likely shipping values.

- **WarehouseLocation**:
  Missing values were predicted using a **Random Forest Classifier** trained on Country, ShippingCost, and TotalSales (engineered later). This helped infer warehouse patterns based on regional logistics and order value.

- **CustomerID**:
  Instead of imputing, missing values were labeled as **"Guest"** to preserve anonymous, valid transactions without distorting customer-based patterns.

Invalid Values and Outlier Treatment

- **Negative UnitPrice values** were treated as data entry errors and the affected rows were removed.

- **Negative Quantity values** were interpreted as **product returns** rather than errors. These were retained to support analysis of return behavior.

- **Outlier Detection** was conducted using boxplots and **Interquartile Range (IQR)** analysis. No extreme outliers were found beyond the invalid negatives, so no further removal was necessary.

Feature Engineering

To support deeper insights and trend-based analysis, several new features were derived:

Feature Name	Description
TotalSales	Computed as Quantity × UnitPrice × (1 - Discount)
InvoiceMonth	Extracted from InvoiceDate for monthly aggregation
InvoiceYear	Extracted for year-over-year comparisons
DayOfWeek	Helps identify weekday vs. weekend sales behavior
DiscountBand	Binned into **Low (0–10%)**, **Medium (10–30%)**, and **High (30%+)**

These engineered variables enabled us to conduct temporal analyses, customer behavior studies, and grouped insights across pricing segments.

Univariate Analysis

Numerical Features:

- **Quantity** and **TotalSales** showed **right-skewed distributions**, with most orders being low in volume and moderate in value.
- **Discount** was mostly concentrated in the **0–30%** range, which aligns with common promotional practices.
- **ShippingCost** varied widely, indicating potential logistics cost inefficiencies.

Categorical Features:

- **SalesChannel** was well balanced between **online** and **in-store**.
- **ReturnStatus** showed an average return rate of ~10%.
- **PaymentMethod** and **OrderPriority** were also fairly distributed.

Bivariate and Multivariate Analysis

Correlation Heatmap:

- **TotalSales** was strongly correlated with **Quantity**, but less so with Discount or ShippingCost.
- No strong multicollinearity was detected.

Discount vs Sales:

- Sales were highest in the **10–30% discount band**.
- Higher discounts (**>30%**) did **not** yield proportional increases in revenue, indicating **diminishing returns** from aggressive promotions.

Country-wise Shipping Cost:

- **Shipping costs** varied significantly by country.
- Some regions with high shipping costs (**e.g., Portugal**) did not generate high sales — suggesting **inefficient fulfillment strategies**.

Statistical Testing

To validate group-wise patterns, we applied formal hypothesis tests:

Test	Comparison	Result
ANOVA	TotalSales across OrderPriority	No significant differences found
Independent T-Test	TotalSales between SalesChannel	No significant difference
Chi-Square Test	ReturnStatus vs PaymentMethod	No dependency observed

These results indicate that while visual trends are observable, they are not always statistically significant — supporting the use of **combined** rather than overly segmented dashboards.

Output and Readiness

The cleaned dataset — saved as online_sales_dataset_cleaned.csv — is now fully ready for dashboarding. It includes:

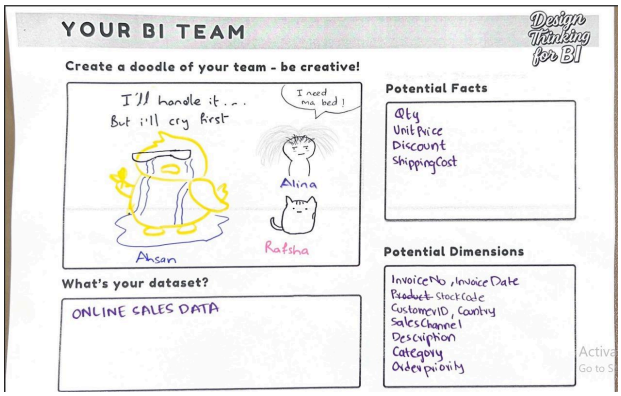
- Cleaned and imputed values for all critical fields
- Derived features to support advanced slicing, trend analysis, and segmentation
- Retained return-related data and high-granularity time stamps
- Verified statistical integrity

This foundational work allowed us to build interactive dashboards with accurate KPIs, supporting decision-making across pricing, logistics, and product strategy.

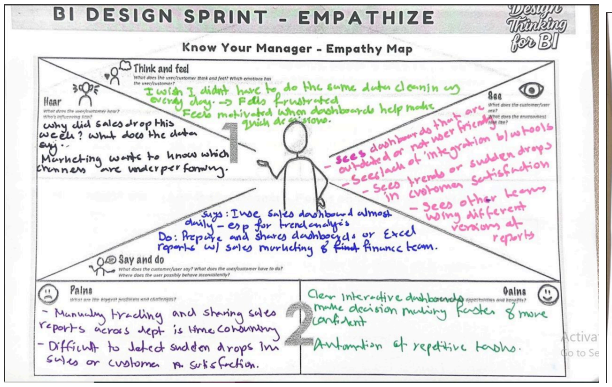
3. Design Thinking Framework Implementation

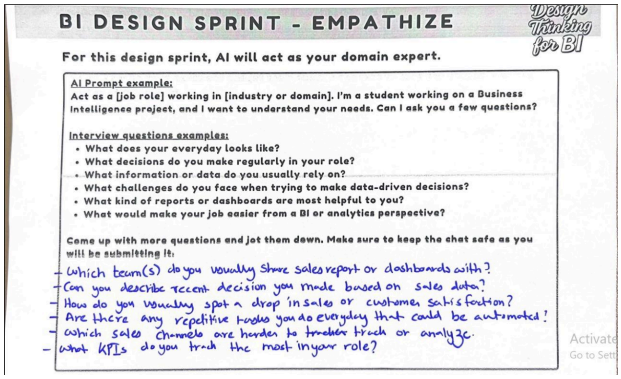
In this project, we followed the **Design Thinking framework** to build a business intelligence solution grounded in real user needs, creativity, and rapid iteration. This human-centered process helped us transform a vague problem into a focused, insight-driven dashboard suite.

> **Note:**
> The Design Thinking Sprint activities were conducted **early in the project timeline**, prior to finalizing our BI problem statement and dashboard structure. While some of the ideas or sketches captured during the sprint may differ slightly from the final solution, this process was instrumental in helping us **clarify our direction**, **identify user pain points**, and **streamline our approach**. It enabled us to ground our dashboard in real-world needs and build a more focused, effective BI solution.



Empathize: Understanding the User





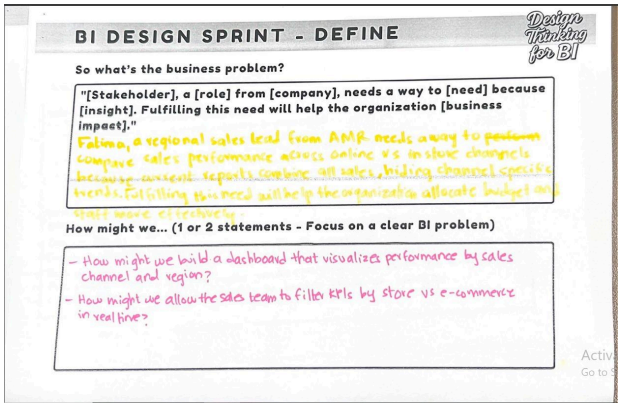
From the empathy map, we discovered:

- **Frustrations**:

    - Repetitive daily tasks (e.g., cleaning data, manually sharing reports)
    - Poor integration between tools and inconsistent reports across departments
    - Difficulty detecting **sudden drops** in sales or customer satisfaction
- **Needs & Goals**:
    - Clean, **interactive dashboards** that aid decision-making
    - Ability to **filter KPIs by channel or region in real-time**
    - Automated tracking of critical KPIs like returns, revenue, and discount performance
- **Frequent Questions She Hears**:
    - “Why did sales drop this week?”
    - “Which channels are underperforming?”
    - “Can we track performance by product category or payment method?”

This stage shaped our entire direction by emphasizing the need for **clear, comparative insights** across regions, product types, and sales channels.

Define: Stating the Problem



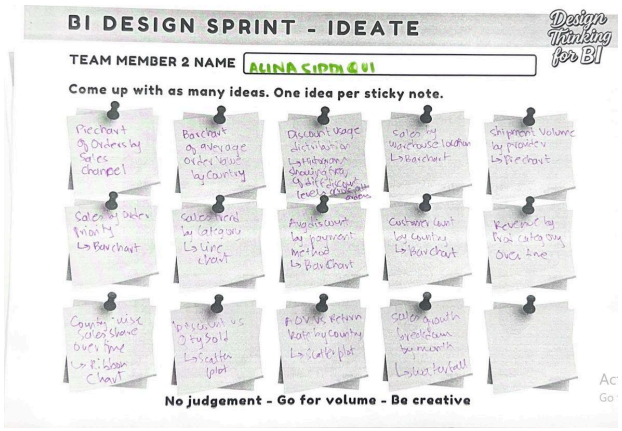
Based on our empathy work, we clearly defined our business problem:

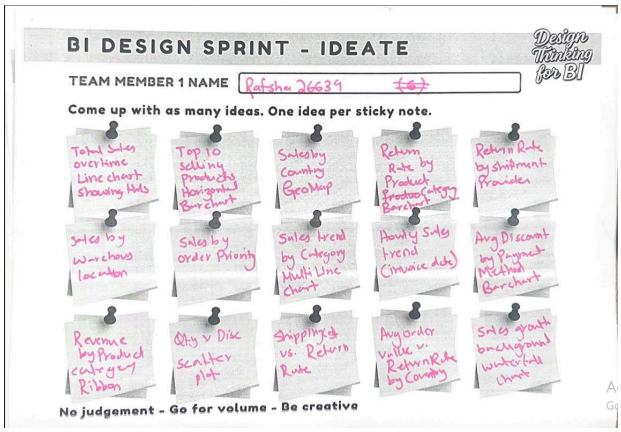
> **"Fatima, a regional sales lead from AMR, needs a way to compare sales performance across online vs in-store channels because current reports combine all sales, hiding channel-specific trends. Fulfilling this need will help the organization allocate budget and staff more effectively."**

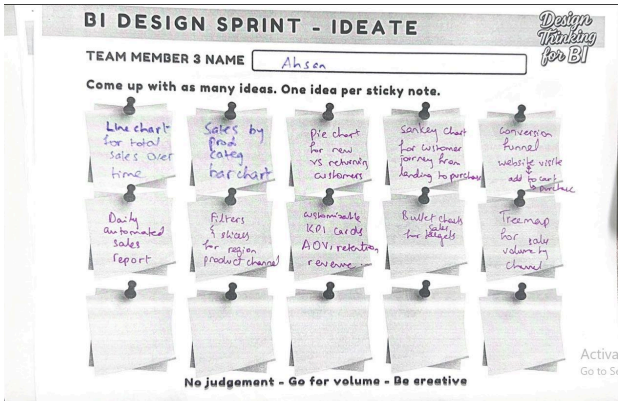
We followed up with two **"How Might We"** statements to focus our scope:

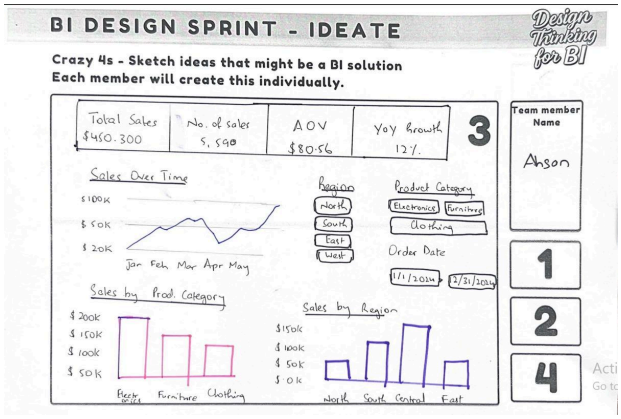
- *How might we build a dashboard that visualizes performance by sales channel and region?*
- *How might we allow the sales team to filter KPIs by store vs e-commerce in real time?*

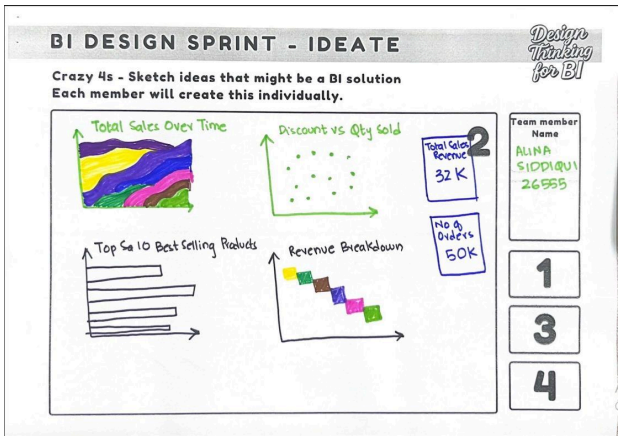
Ideate: Generating Creative BI Ideas

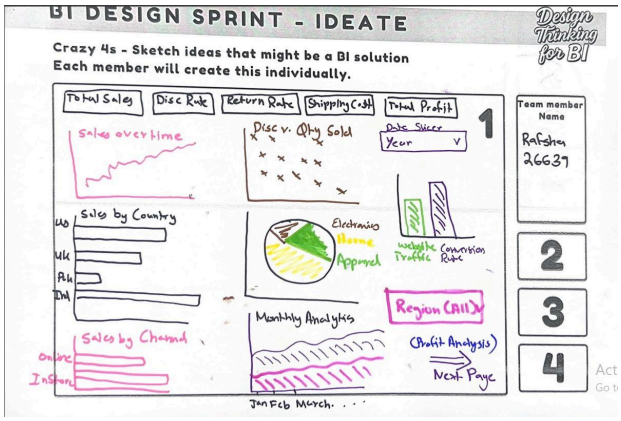










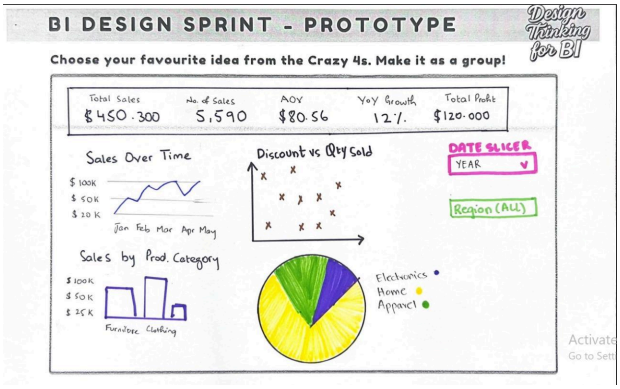


All three team members — Alina, Rafsha, and Ahsan — brainstormed chart ideas through sticky notes. Ideas spanned multiple perspectives:

- **Alina** suggested bar charts by sales channel, discount usage visualizations, and category-level breakdowns
- **Rafsha** focused on geo-maps, return rate analysis, and trends by order priority
- **Ahsan** proposed funnel charts, KPI cards, automation suggestions, and treemaps for product volume

This ideation phase gave us a **broad canvas of visualization options**, which we narrowed down in the next phase.

Prototype: Creating Our Unified Dashboard Sketch

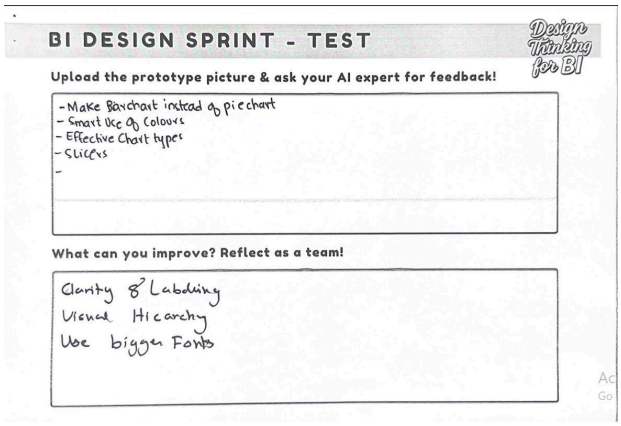


As a team, we combined our favorite ideas into a low-fidelity dashboard prototype. The layout included:

- **Top KPI Cards**: Total Sales, No. of Orders, AOV, YoY Growth, Total Profit
- **Line Chart**: Sales Over Time
- **Bar Chart**: Sales by Product Category
- **Scatter Plot**: Discount vs Quantity Sold
- **Pie Chart**: Sales Share by Category
- **Slicers**: By Year and Region

This sketch served as our **visual blueprint** and laid the foundation for our final Power BI dashboards.

Test: Feedback and Improvements



During the testing phase, we presented the prototype to peers and our AI mentor for feedback. Key takeaways included:

- **Replace Pie Chart** with a **Bar Chart** for better clarity
- Improve **labeling**, **font sizes**, and **color hierarchy**
- Ensure slicers are **clearly visible and intuitive**

These insights were incorporated directly into our Power BI implementation, where we used consistent font sizes, slicers for all dashboards, and avoided pie charts unless the segment count was small.

Summary of Our Design Thinking Framework Implementation

Stage	What We Did
Empathize	Created an empathy map and interviewed a fictional sales lead (Fatima) to uncover real frustrations and BI needs
Define	Articulated a clear business problem and framed two key "How might we..." questions
Ideate	Used sticky-note brainstorming and Crazy 4s to generate a wide variety of creative dashboard ideas
Prototype	Built a sketched layout combining key ideas from all three members into a unified visual plan
Test	Received structured feedback and reflected as a team to improve chart type, clarity, and usability before final dashboarding

This framework helped us **stay focused**, **design with empathy**, and **deliver insights that truly matter** to decision-makers in a retail analytics setting.

Interview

Name: Muhammad Naveed

Company: Jubilee Insurance

Position: Manager - Digital Sales

LinkedIn: linkedin.com/in/muhammad-naveed-b0771a76

1. What key performance indicators (KPIs) do you typically track to measure sales success?

The sales team closely monitors a set of core KPIs that provide a comprehensive view of both ongoing performance and growth potential. These include:

- **Existing Portfolio Performance:**
  Tracking the sales revenues generated from the current portfolio of products or services is fundamental. This helps assess how well established offerings continue to perform in the market, ensuring that steady revenue streams are maintained.

- **New Acquisitions:**
  Monitoring the number and quality of new customer acquisitions or contracts secured is another critical KPI. This metric indicates the effectiveness of sales outreach and market expansion efforts, highlighting growth beyond existing accounts.

- **Sales Revenue Trends:**
  Comparing current sales revenue figures with those of the previous month and the same period last year allows the team to evaluate both short-term momentum and long-term growth patterns. These comparisons help identify seasonal effects, emerging trends, or areas needing intervention

By tracking these KPIs, the sales team gains actionable insights into the balance between maintaining steady revenue from existing customers and driving expansion through new business, supporting informed decision-making and strategic planning.

2. What types of sales data are most critical for your decision-making? (e.g., product categories, regions, sales channels)

The sales team places high importance on **tracking month-over-month (MoM) growth**, as well as comparing current sales figures with those of previous months and the corresponding period last year. This temporal comparison provides a clear view of sales momentum and growth trends. They emphasize monitoring the **performance of the sales team on a regional basis**, focusing on how different geographic markets contribute toward overall sales targets. This includes measuring **target versus actual sales** to identify which regions are lagging behind expectations and where additional efforts or resources

might be required. Understanding these dynamics allows the team to pinpoint areas that need strategic attention and growth focus, enabling them to allocate sales resources more effectively.

3. Are there any specific sales challenges or pain points you face regularly?

One of the most significant challenges faced by the sales team is the **lack of training and advocacy among retail agents**, who act as the primary distribution channel. Unlike direct sales to a single retailer, sales here happen through a **group or network of retail agents**, making it difficult to maintain consistent product knowledge and motivation across all touchpoints. The team highlighted that these agents often lack a deep understanding of the products they sell, which hampers their ability to communicate benefits or upsell effectively to customers. Therefore, **retail advocacy**—providing proper training, product education, and clear incentives—is critical. By empowering retail agents with the right knowledge and motivation, companies can significantly enhance product visibility and sales effectiveness in these dispersed retail networks.

4. What types of reports or dashboards do you currently use or wish you had?

The sales team currently relies on **internally developed sales dashboards** tailored to their specific needs. These dashboards help them track and visualize key sales metrics, but much of their analysis and reporting work still happens in **Excel spreadsheets**. Excel remains a core tool due to its flexibility, ease of use, and familiarity among team members. However, this may also point to potential limitations in scalability or real-time data integration, suggesting an opportunity for more advanced BI tools to automate and enhance reporting capabilities. This blend of custom dashboards and Excel-based analysis allows the team to generate actionable reports but might benefit from integrating more dynamic and interactive BI platforms.

5. What insights from sales data have been most useful for improving performance or strategy?

By continuously analyzing **sales flow, trends, and patterns**, the team gains critical insights into the market's evolving dynamics. They closely monitor **where sales are increasing or decreasing**, which helps them understand the effectiveness of sales strategies and market responses. A key takeaway from their approach is identifying **"low hanging fruits"**—segments or opportunities where sales can be quickly and easily boosted without major resource investment. These quick wins are prioritized and targeted through focused campaigns or operational changes, providing immediate performance improvements. The insights generated are shared through reports distributed across the sales channels, ensuring alignment and coordinated action to drive sustained growth.

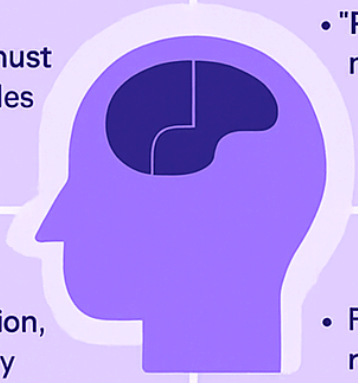
Empathy Map

SAYS

- "We need better training for our retall agents to boost product advocacy."
- "Discounts must be carefully managed to avoid margin erosion."
- "Excel and custom dashboards get the job done, but can be limiting."
- "Marketing support must align with regional sales performance."

THINKS

- Tracking MoM growth and regional sales is crucial to hitting targets.
- Returns are a big challenge, expecially in categories like Furniture and Apparel.
- We should focus on quick wins and low hanging fruits first.
- "Retail is complex due to multiple agents and inconsistent product knowledge.



DOES

- Monitors sales by region, channel, and category regularly
- Compares current sales data with past months and years.
- Creates reports and dashboards mostly using Excel and in-house tools.
- Prioritizes regions and categories fot targeted marketing and sales efforts.

FEELS

- Frustrated by inconsistent retall advocacy and lack of agent training.
- Concerned about rising returns and their impact on profitability.
- Motivated by identifying actionable insights that can improve performance.
- Pressured to meet targets despito operational chalenges and growth.

PAIN POINTS

- Lack of uniform retail agent training and product knowledge.
- Compares current sales by margins, especially in specific categories.

GAINS

- Improved sales visibiility by region and category through Bi dashboards,
- Ability to identify and act on low hanging fruit opportunities.

4. Custom Measures & Calculations

Before creating the visualizations, we defined several DAX measures and columns to drive accurate analysis and insights across key metrics like sales, returns, discounts, and shipping.

Here's a list of all the measures and columns we created:

1. Average Sales per SKU
2. AvgDiscount
3. DiscountBand
4. DiscountedOrders
5. HighReturnCountries
6. LowestReturnCategory
7. MostExpensiveCarrier
8. Returned Orders
9. ReturnRate
10. ReturnRevenueLossPct
11. Sales from Discounted Orders
12. Sales from Returns
13. Sales Target
14. Sales_YoY_Growth
15. Sales.Good
16. Sales.Max
17. Sales.NI
18. Sales.Satis
19. Sales.VGood
20. Shipping Cost %
21. ShippingCost
22. TargetAchievement
23. Top_Discounted_Category
24. TopCategory
25. TopReturnCountry
26. TopShippingCostCountry
27. Total Shipping Cost
28. TotalSales
29. UnitPrice
30. Total Shipments

Some of the measures are as follows:

```
HighReturnCountries =
VAR CountryReturnTable =
    ADDCOLUMNS(
        VALUES('cleaned_online_sales'[Country]),
        "ReturnRate",
            DIVIDE(
                CALCULATE(COUNTROWS('cleaned_online_sales'), 'cleaned_online_sales'[ReturnStatus] = "Returned"),
                CALCULATE(COUNTROWS('cleaned_online_sales'))
            )
    )
RETURN
COUNTROWS(
    FILTER(CountryReturnTable, [ReturnRate] > 0.11)
)

```

Counts the number of countries where the return rate exceeds 11%

```
LowestReturnCategory =
CALCULATE (
    SELECTEDVALUE('cleaned_online_sales'[Category]),
    TOPN (
        1,
        ADDCOLUMNS (
            VALUES('cleaned_online_sales'[Category]),
            "ReturnRateCalc",
                DIVIDE(
                    CALCULATE(COUNTROWS(FILTER('cleaned_online_sales', 'cleaned_online_sales'[ReturnStatus] = "Returned"))),
                    CALCULATE(COUNTROWS('cleaned_online_sales'))
                )
        ),
        [ReturnRateCalc], ASC
    )
)

```

Identifies the product category with the lowest return rate.

```
MostExpensiveCarrier =
CALCULATE (
    SELECTEDVALUE('cleaned_online_sales'[ShipmentProvider]),
    TOPN (
        1,
        SUMMARIZE (
            'cleaned_online_sales',
            'cleaned_online_sales'[ShipmentProvider],
            "TotalShippingCost", SUM('cleaned_online_sales'[ShippingCost])
        ),
        [TotalShippingCost], DESC
    )
)

```

Determines the shipment provider with the highest total shipping cost.

```
ReturnRate =
DIVIDE(
    CALCULATE(
        COUNTROWS('cleaned_online_sales'),
        'cleaned_online_sales'[ReturnStatus] = "Returned"
    ),
    COUNTROWS('cleaned_online_sales'),
    0
)
```

Calculates the proportion of orders that have been returned out of total orders.

```
Top_Discounted_Category =
CALCULATE (
    MAX ( 'cleaned_online_sales'[Category] ),
    TOPN (
        1,
        VALUES ( 'cleaned_online_sales'[Category] ),
        CALCULATE ( AVERAGE ( 'cleaned_online_sales'[Discount] ) ),
        DESC
    )
)
```

Identifies the product category with the highest average discount.

```
Top_Risky_Discount_Category =
CALCULATE (
    MAX ( 'cleaned_online_sales'[Category] ),
    TOPN (
        1,
        VALUES ( 'cleaned_online_sales'[Category] ),
        CALCULATE ( [ReturnRate] ),
        DESC
    )
)
```

Identifies the product category with the highest return rate.

```
TopCategory =
CALCULATE (
    MAX ( 'cleaned_online_sales'[Category] ),
    TOPN (
        1,
        VALUES ( 'cleaned_online_sales'[Category] ),
        CALCULATE ( SUM ( 'cleaned_online_sales'[TotalSales] ) ),
        DESC
    )
)
```

Determines the product category with the highest total sales.

```
TopShippingCostCountry =
CALCULATE (
    SELECTEDVALUE('cleaned_online_sales'[Country]),
    TOPN (
        1,
        SUMMARIZE (
            'cleaned_online_sales',
            'cleaned_online_sales'[Country],
            "AvgShipping", AVERAGE('cleaned_online_sales'[ShippingCost])
        ),
        [AvgShipping], DESC
    )
)

```

Finds the country with the highest average shipping cost.

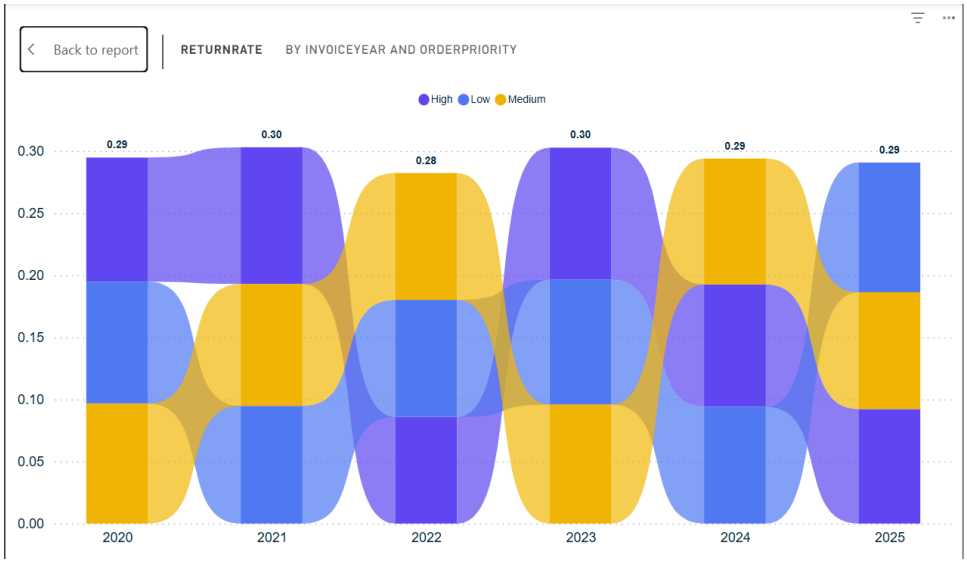
Additionally, we also created a calculated column:

```
DiscountBand =
SWITCH(TRUE(),
    'cleaned_online_sales'[Discount] < 0.1, "Low (0-10%)",
    'cleaned_online_sales'[Discount] < 0.3, "Medium (10-30%)",
    "High (30%+)"
)
```

Categorizes each order's discount into 'Low (0-10%)', 'Medium (10-30%)', or 'High (30%+)' discount bands.

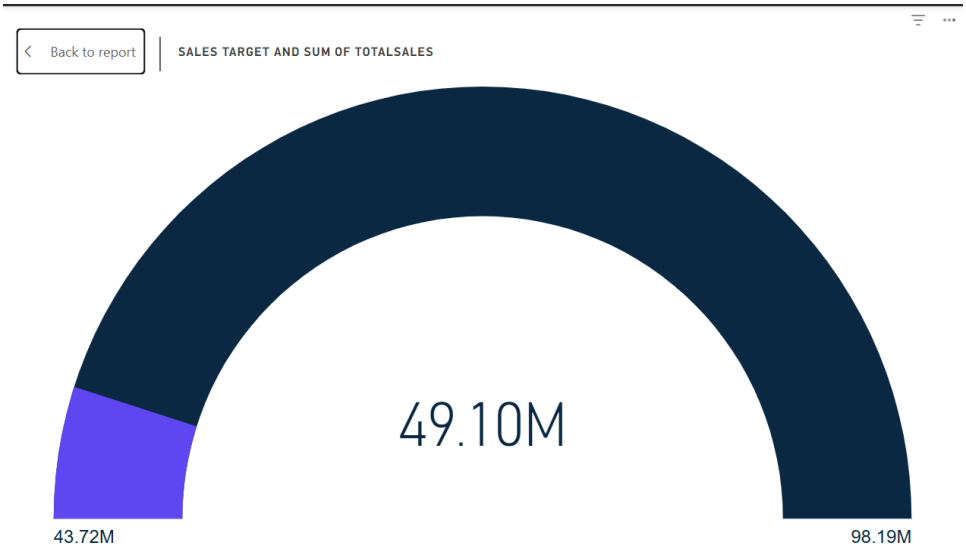
5. Individual Charts & Insights

Executive Summary Charts



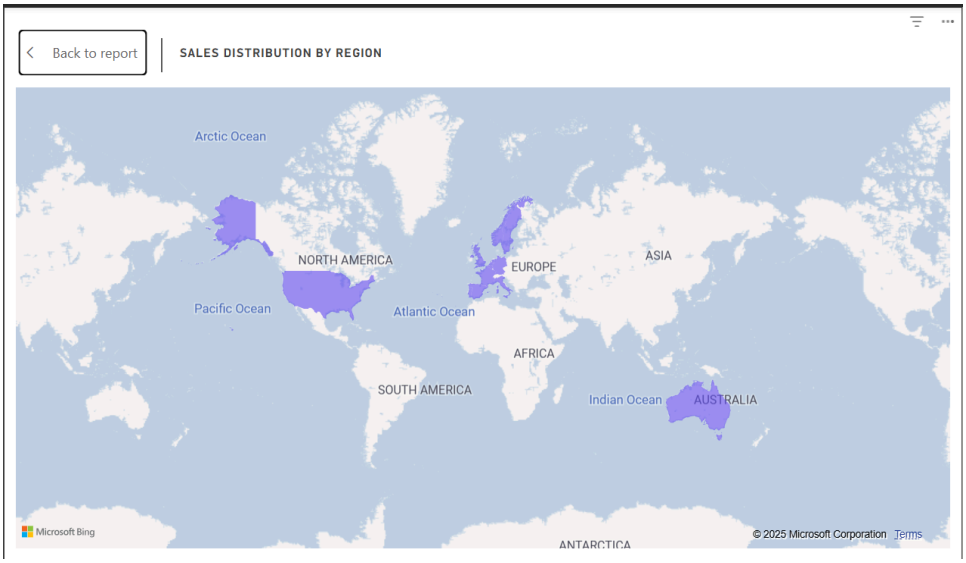
Return Rate by Invoice Year and Order Priority (Ribbon Chart)

This ribbon chart shows how return rates vary across years and order priorities, revealing that high-priority orders consistently face slightly higher return rates (~0.30), indicating potential fulfillment or expectation gaps.



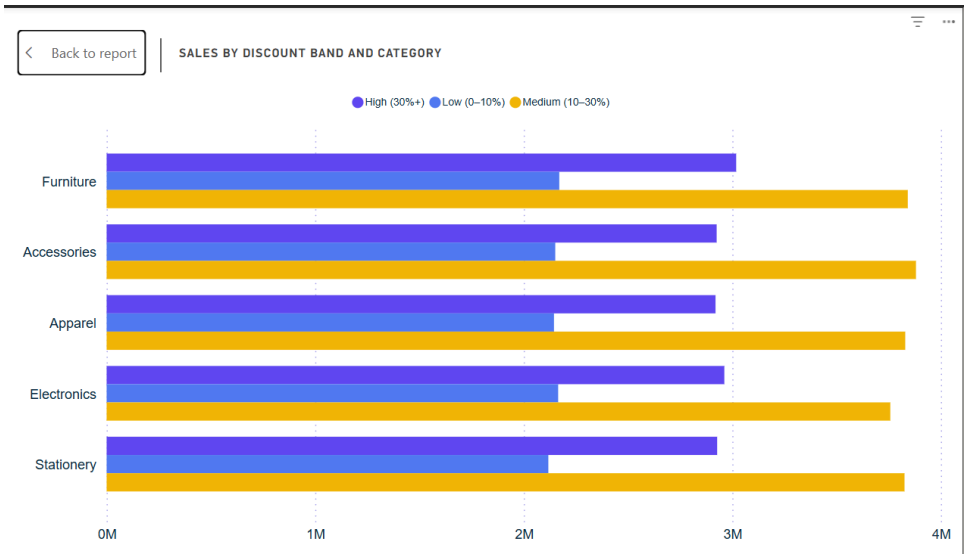
Sales Target vs Total Sales (Gauge Chart)

The gauge shows that the current sales (43.72M) are progressing steadily toward the 49.10M target, indicating good momentum but still short of goal.



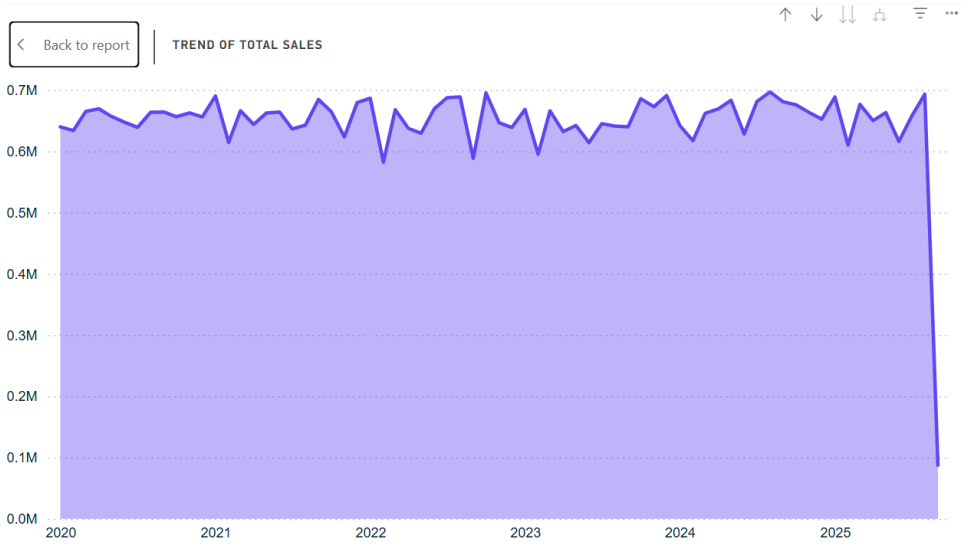
Sales Distribution by Region (Filled Map)

The map highlights Europe, North America, and Australia as the strongest contributing regions in terms of sales, with regional strategies needed based on geographic performance.



Sales by Discount Band and Category (Clustered Bar Chart)

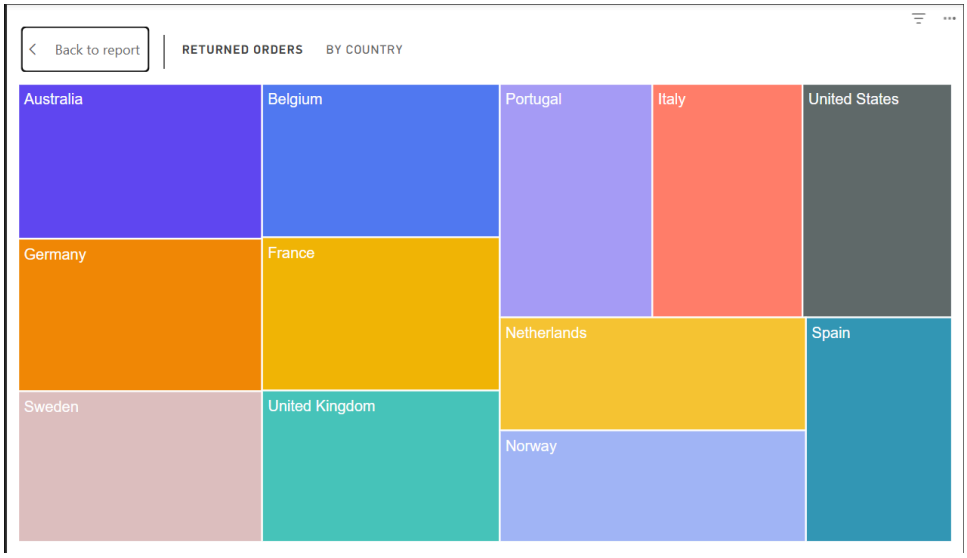
The bar chart indicates that most categories perform best within the medium discount band (10–30%), with Furniture and Accessories heavily utilizing higher discounts.



Trend of Total Sales Over Time (Area Chart)

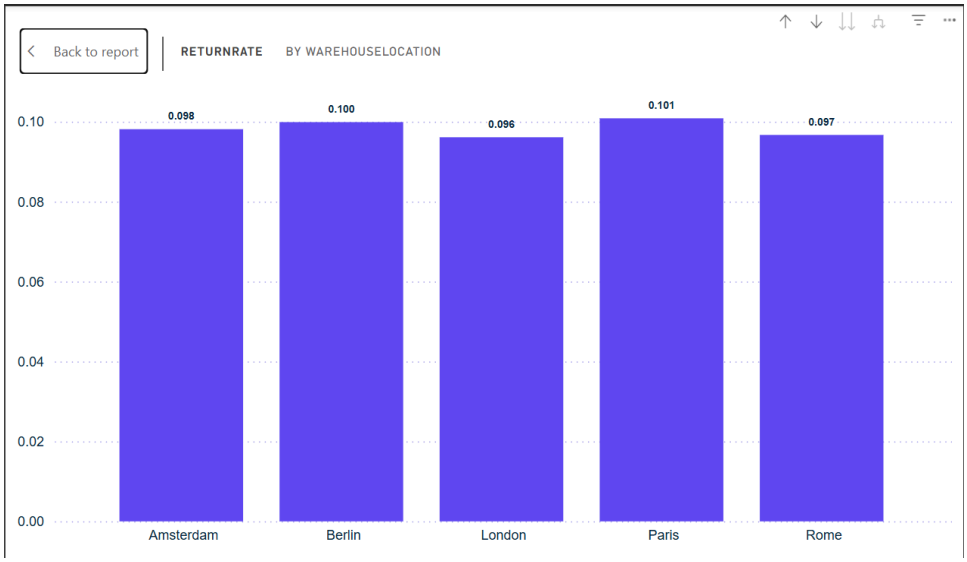
A consistent monthly sales trend is visible until 2025, where a steep drop raises concern—potentially linked to operational or demand-side issues.

Returns Breakdowns Charts



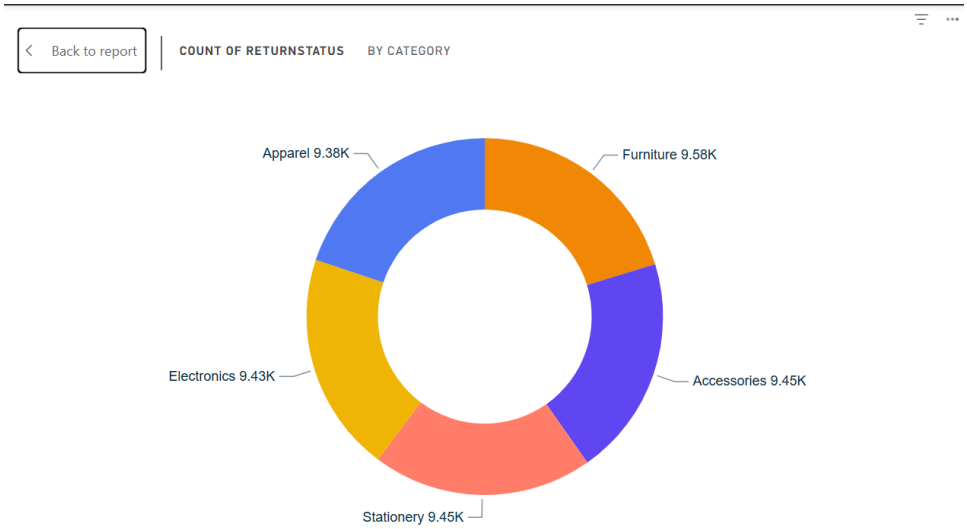
Returned Orders by Country (Treemap)

This treemap highlights which countries contribute the most to returned orders. Australia, Germany, and Belgium show the highest volume, indicating potential issues in fulfillment or customer expectations in these regions



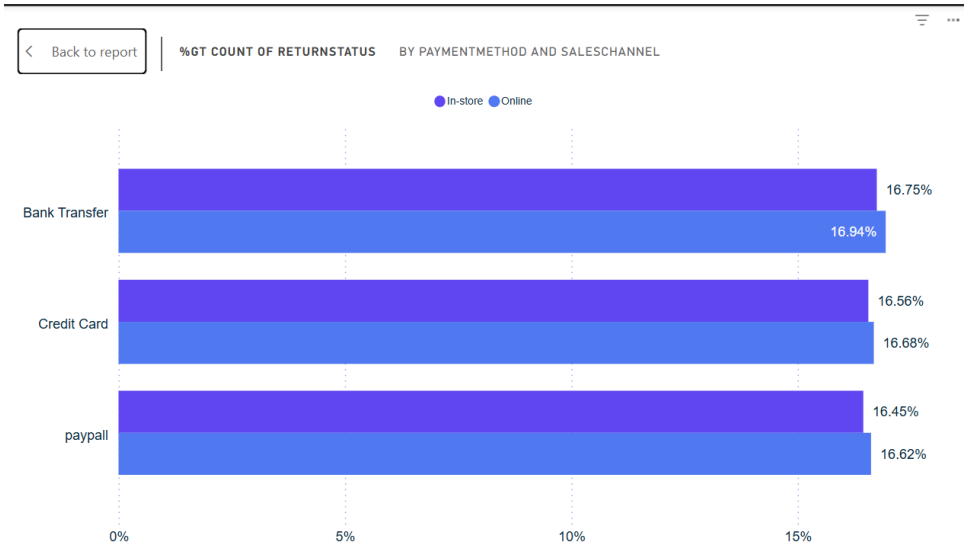
Return Rate by Warehouse Location (Column Chart)

This chart shows return rates by warehouse. Paris (0.101) and Berlin (0.100) have slightly elevated return rates, suggesting process inefficiencies or regional logistic challenges at these hubs.



Count of Returns by Category (Donut Chart)

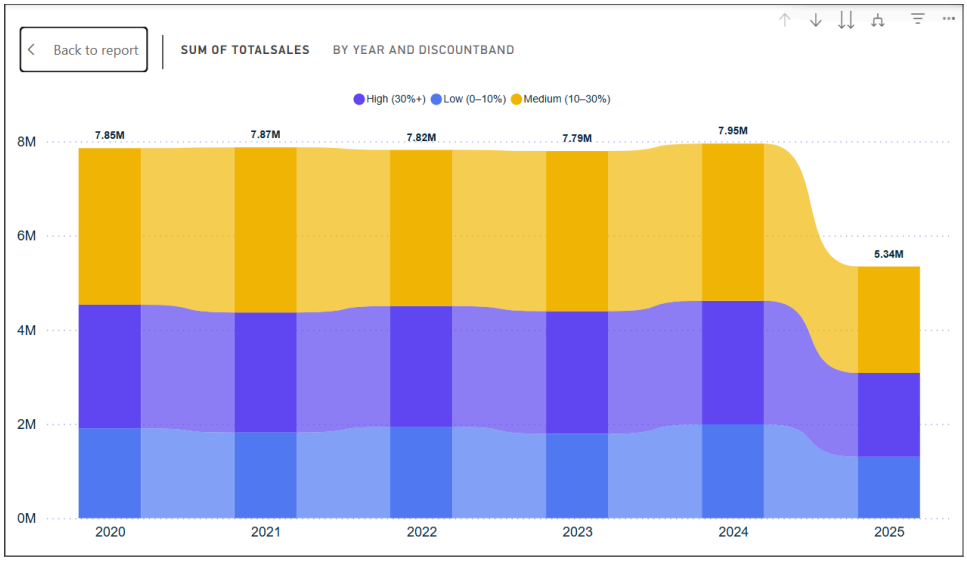
Furniture and Accessories have the highest number of returned items, indicating product-type-specific return patterns that could guide quality control or return policy refinements.



% Count of Return Status by Payment Method and Sales Channel (Grouped Bar Chart)

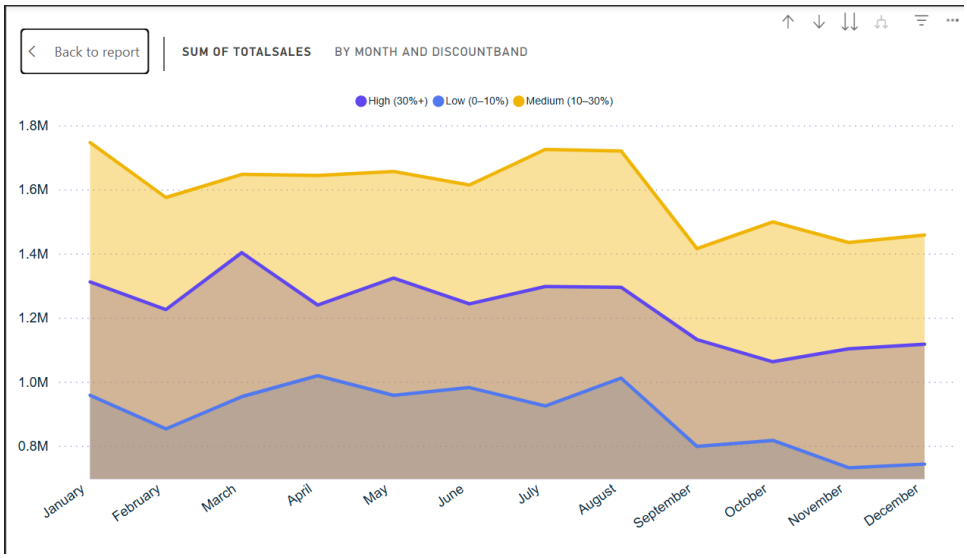
This chart compares return percentages across payment methods and channels. Bank Transfers show the highest return rates (~16.9%) in online channels, hinting at potential trust or delivery issues with these transactions.

Discount Analysis Charts



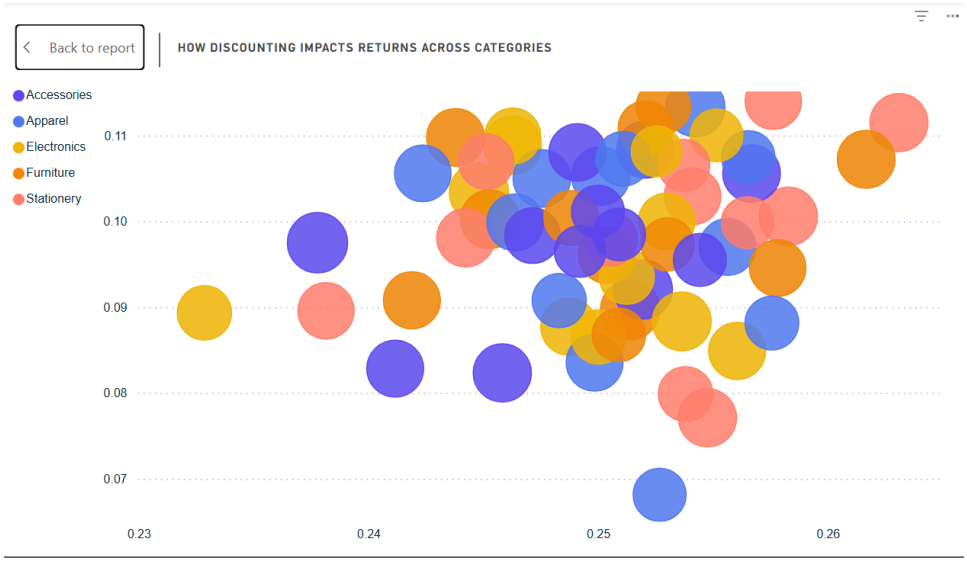
Sum of TotalSales by Year and Discount Band (Ribbon Chart)

This ribbon chart illustrates that the 10–30% discount band consistently generates the highest annual sales, with a sharp decline in overall sales in 2025 across all bands.



Sum of TotalSales by Month and Discount Band (Area Chart)

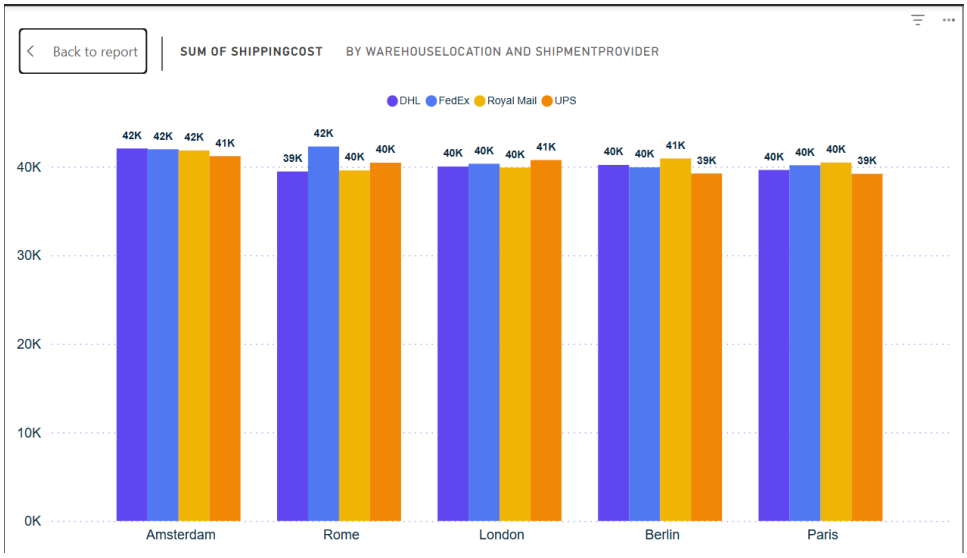
Shows monthly trends, reinforcing that medium discounts yield stable performance while high discounts have limited impact.



How Discounting Impacts Returns Across Categories (Scatter/Bubble Chart)

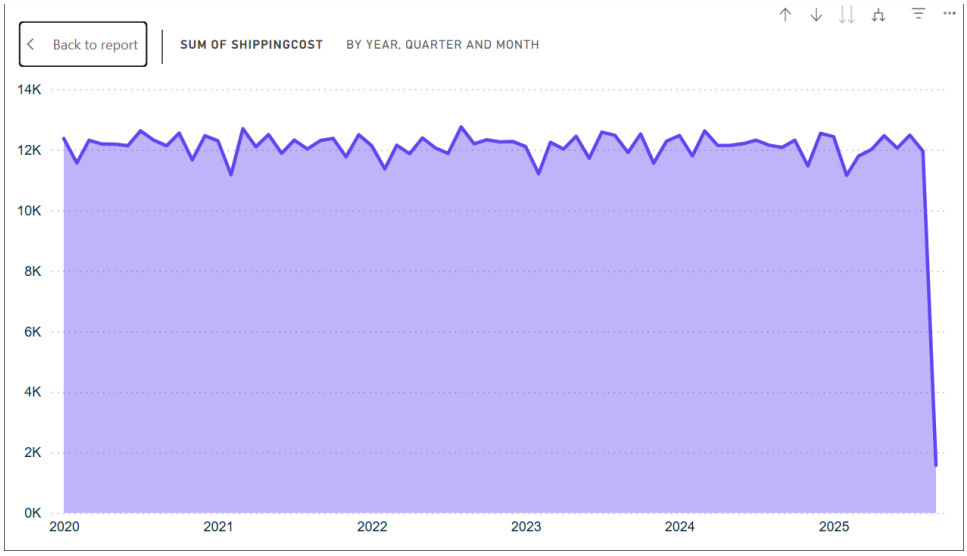
Reveals that while most discounts cluster around 0.25, categories like Furniture show slightly elevated return rates, indicating varying risk levels.

Shipping Cost Impact Charts



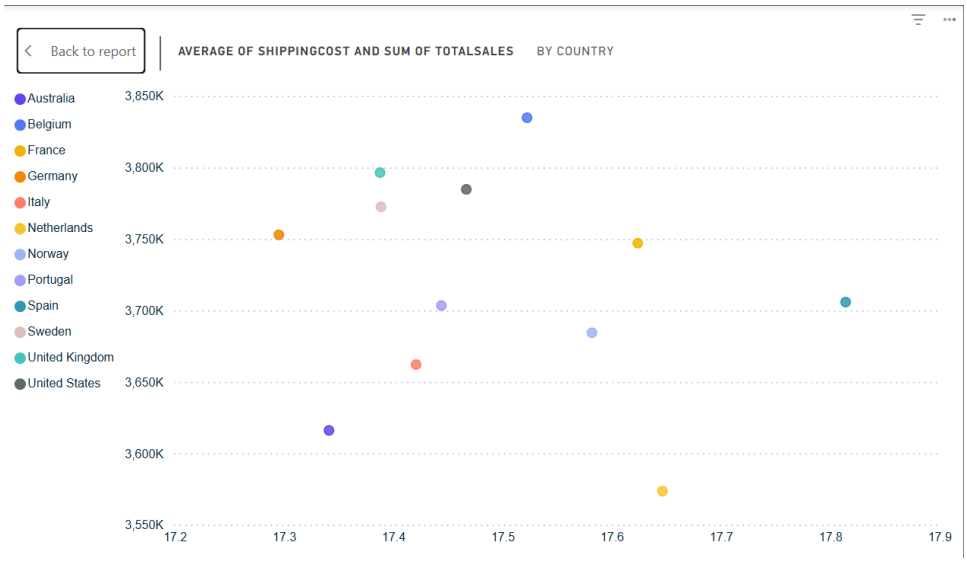
Shipping Cost by Warehouse and Provider (Clustered Bar Chart)

Shipping costs remain consistent across all warehouse locations and providers. No single carrier significantly dominates in cost or usage volume.



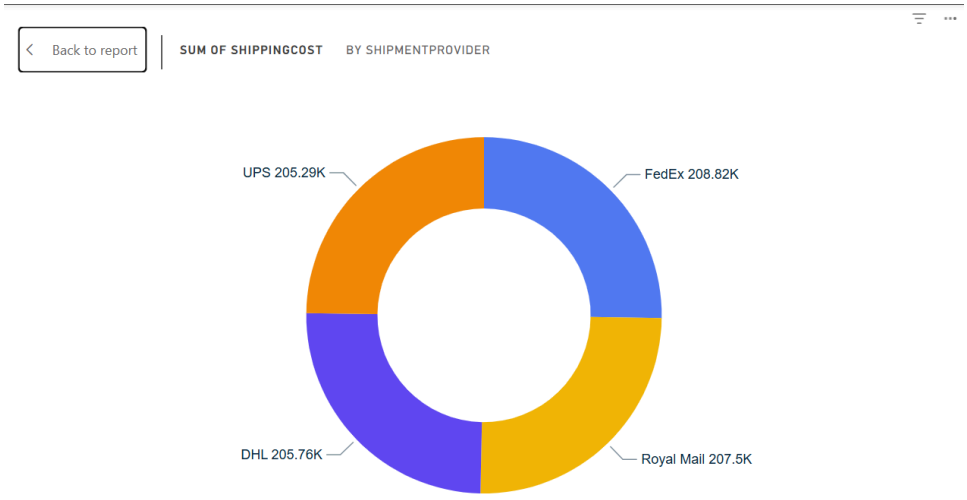
Shipping Cost Trend Over Time (Area Chart)

Shipping expenses stay stable across years until a sharp drop in 2025. This decline may indicate a system gap or operational shift needing investigation.



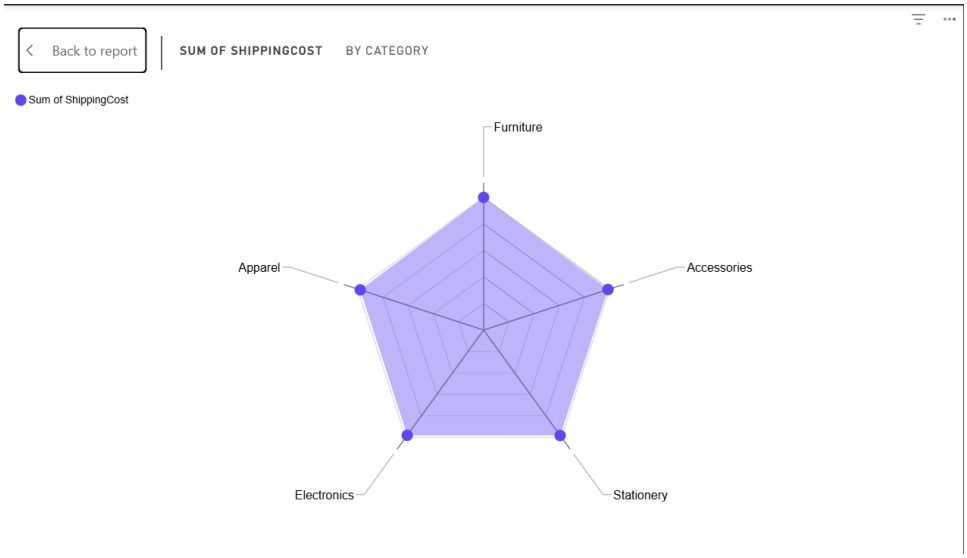
Avg Shipping Cost vs Total Sales by Country (Scatter Plot)

Some countries like Portugal have high shipping costs but lower sales output. This indicates inefficiencies and opportunities to optimize logistics spending.



Shipping Cost by Provider (Donut Chart)

Shipping costs are almost evenly split among FedEx, DHL, UPS, and Royal Mail. Balanced distribution suggests no over-dependence on any one provider.



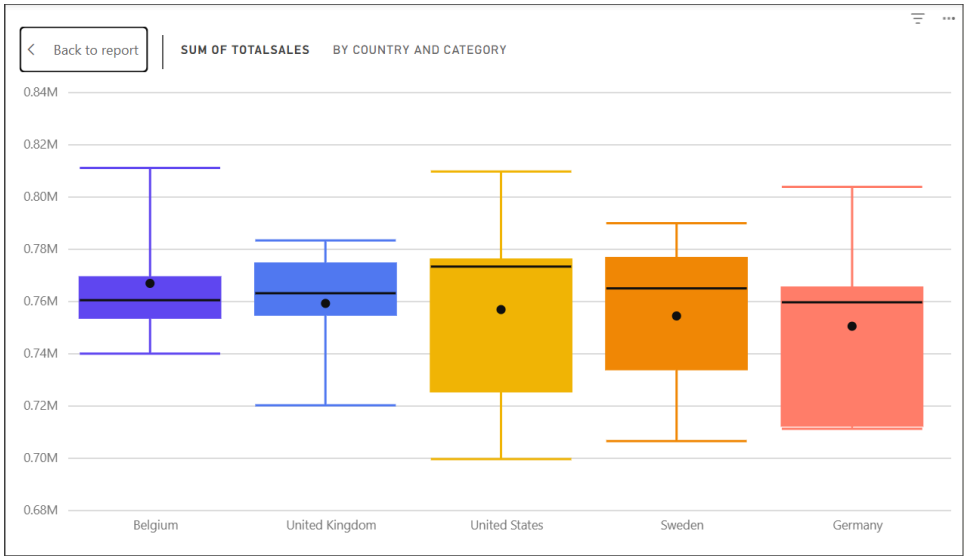
Shipping Cost by Category (Radar Chart)

Furniture leads in shipping cost, followed by Accessories and Apparel.
Heavier or bulkier items likely drive the higher shipping spend.

Category Performance Charts

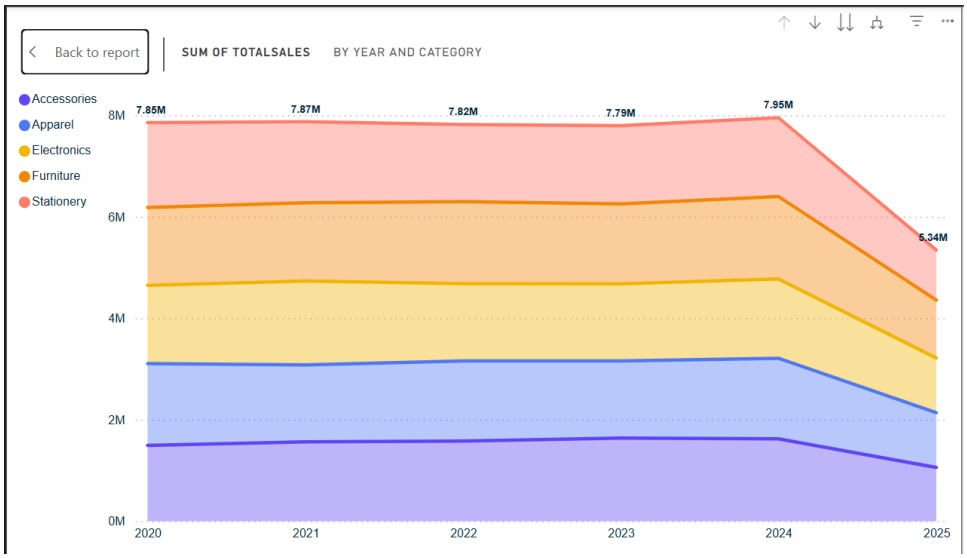
Category Target Performance Status (Bullet Chart)

Displays progress of each product category against predefined sales targets using color bands.
All categories except Stationery are deep into the green zone, indicating strong performance.



Sum of Total Sales by Country and Category (Box Plot)

Highlights sales variability across countries with category-wise distribution.
U.S. and Germany show wider IQRs, hinting at inconsistent category performance compared to Belgium and the UK.



Sum of Total Sales by Year and Category (Area Chart)

Shows annual sales distribution across five categories over time.
A drop is seen in 2025 across all categories, possibly indicating external disruptions like demand dip or supply issues.

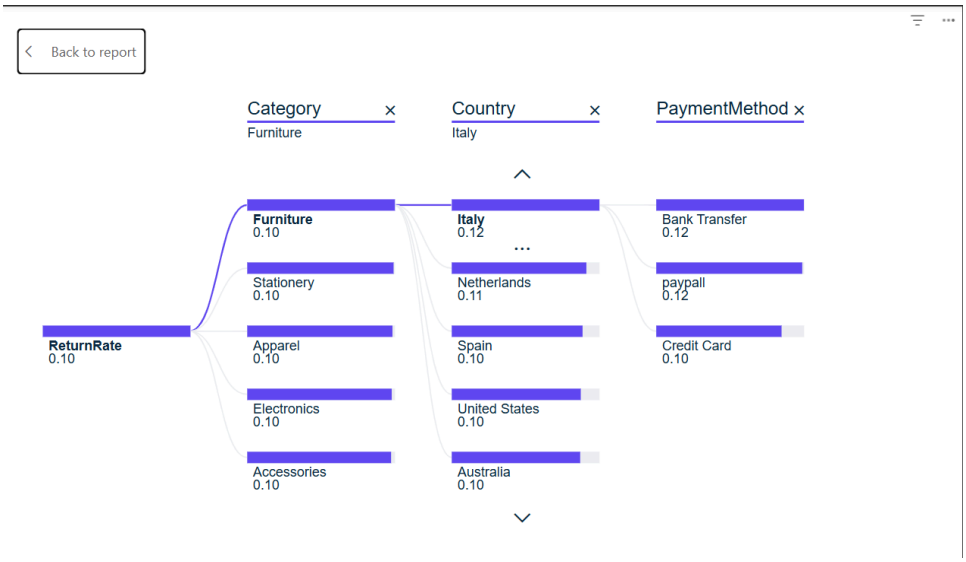
Back to report | SALES BY COUNTRY AND PRODUCT CATEGORY

Country	Accessories	Apparel	Electronics	Furniture	Stationery	Total
Total	**8,956,126.00**	**8,893,156.05**	**8,883,923.59**	**9,029,878.54**	**8,870,610.28**	**44,633,694.47**
Belgium	740,054.12	753,812.10	811,167.26	760,532.07	769,260.56	**3,834,826.11**
United Kingdom	774,575.81	783,397.64	755,023.32	720,284.88	763,194.25	**3,796,475.89**
United States	773,396.50	725,662.75	809,809.35	776,120.24	699,718.24	**3,784,707.07**
Sweden	765,061.32	776,652.10	706,613.61	790,032.08	734,109.88	**3,772,469.00**
Germany	759,729.09	712,506.57	711,313.11	803,919.48	765,368.69	**3,752,836.92**
France	697,649.81	783,254.57	756,167.08	750,502.37	759,486.32	**3,747,060.16**
Spain	696,836.57	757,719.10	714,942.53	779,313.86	756,994.53	**3,705,806.58**
Portugal	773,584.67	735,952.12	761,659.51	737,214.60	694,976.97	**3,703,387.87**
Norway	748,992.09	696,294.00	780,649.70	733,395.04	725,038.37	**3,684,369.20**
Italy	756,356.14	739,048.37	707,850.60	701,700.72	757,220.05	**3,662,175.87**
Australia	733,234.73	700,916.15	706,573.96	740,922.06	734,415.15	**3,616,062.04**
Netherlands	736,655.17	727,940.60	662,153.57	735,941.13	710,827.28	**3,573,517.75**

Sales by Country and Product Category (Matrix Table)

A detailed breakdown of sales figures for each product category across countries. Belgium and the UK lead in total sales, with consistent performance across all five categories.

Recommendations Charts



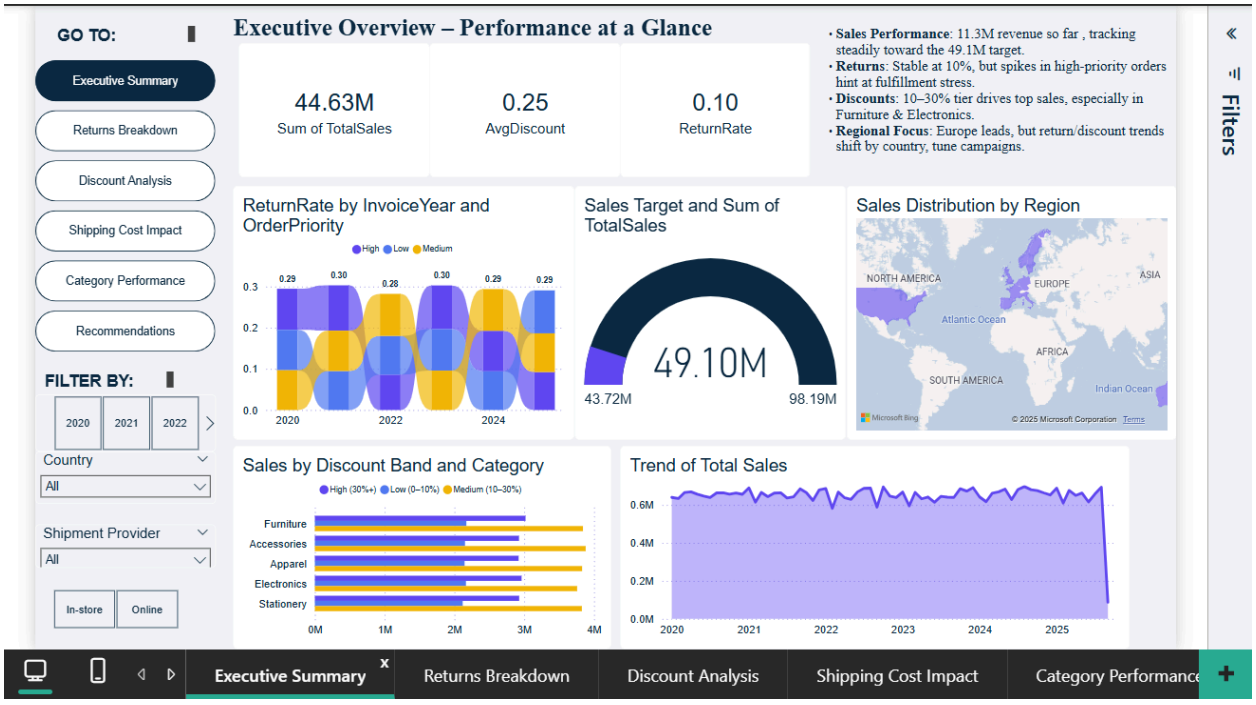
Return Rate Analysis (Decomposition Tree)

This decomposition tree breaks down the return rate by category, country, and payment method. It highlights that Furniture returns in Italy are higher (0.12), especially with Bank Transfer and PayPal payments.

6. Dashboard Overview & Insights

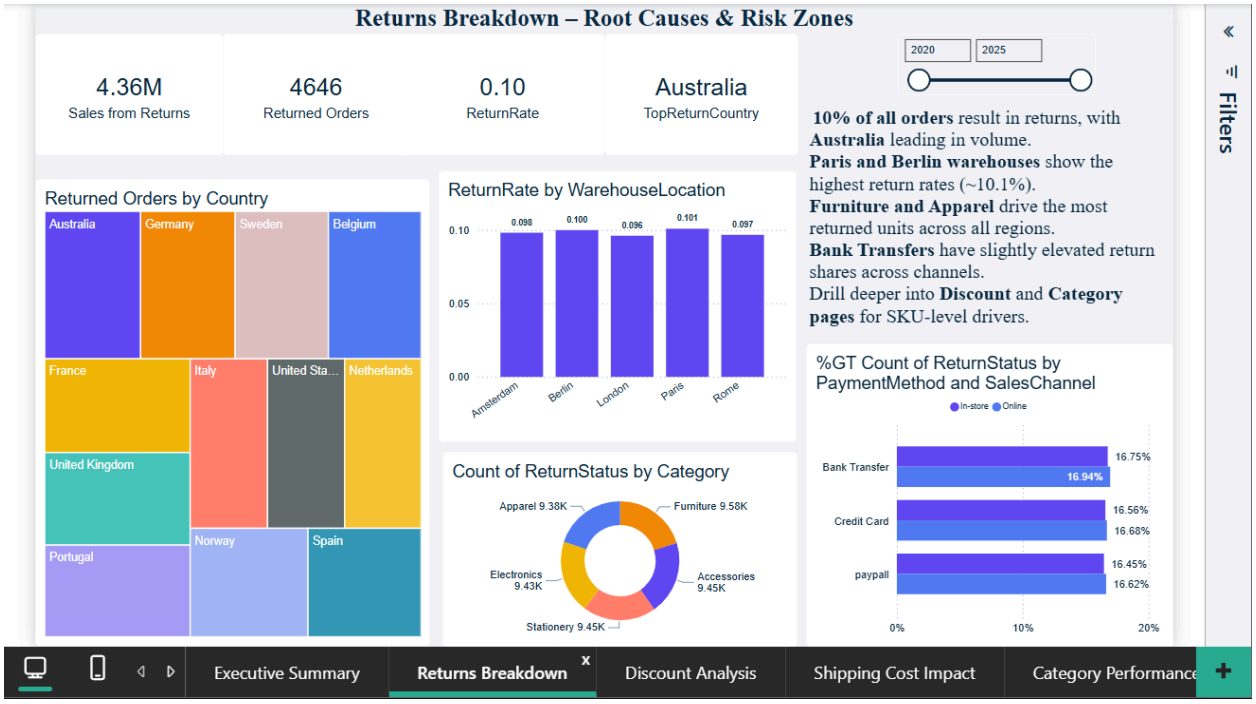
Each dashboard page below captures a unique aspect of our business problem — from sales trends and return analysis to shipping impact and strategic recommendations — offering actionable insights through clear, data-driven visuals.

1. Executive Overview – Performance at a Glance



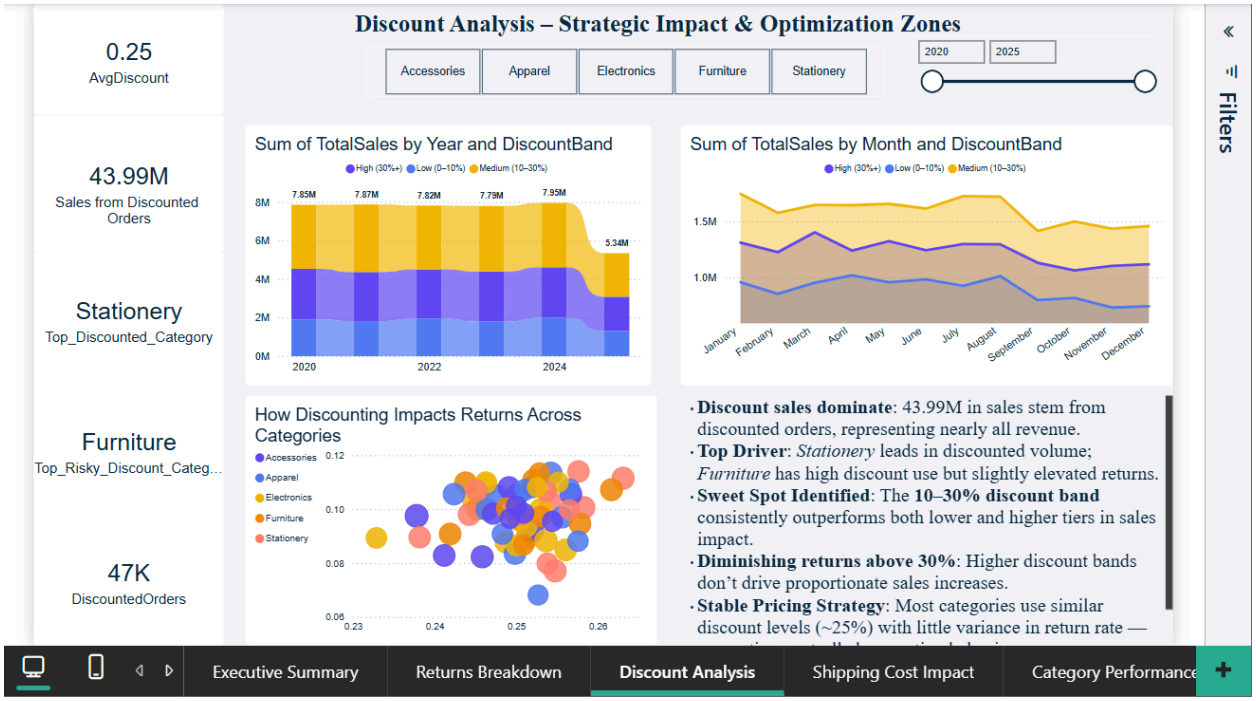
This page provides a high-level summary of key metrics. It shows that the business has achieved **44.63M in total sales**, with a **return rate of 10%** and an **average discount of 25%**. The donut chart confirms that performance is on track toward the **49.1M target**. Regional trends and order priority data highlight performance differences and help plan focused strategies by area.

2. Returns Breakdown – Root Causes & Risk Zones



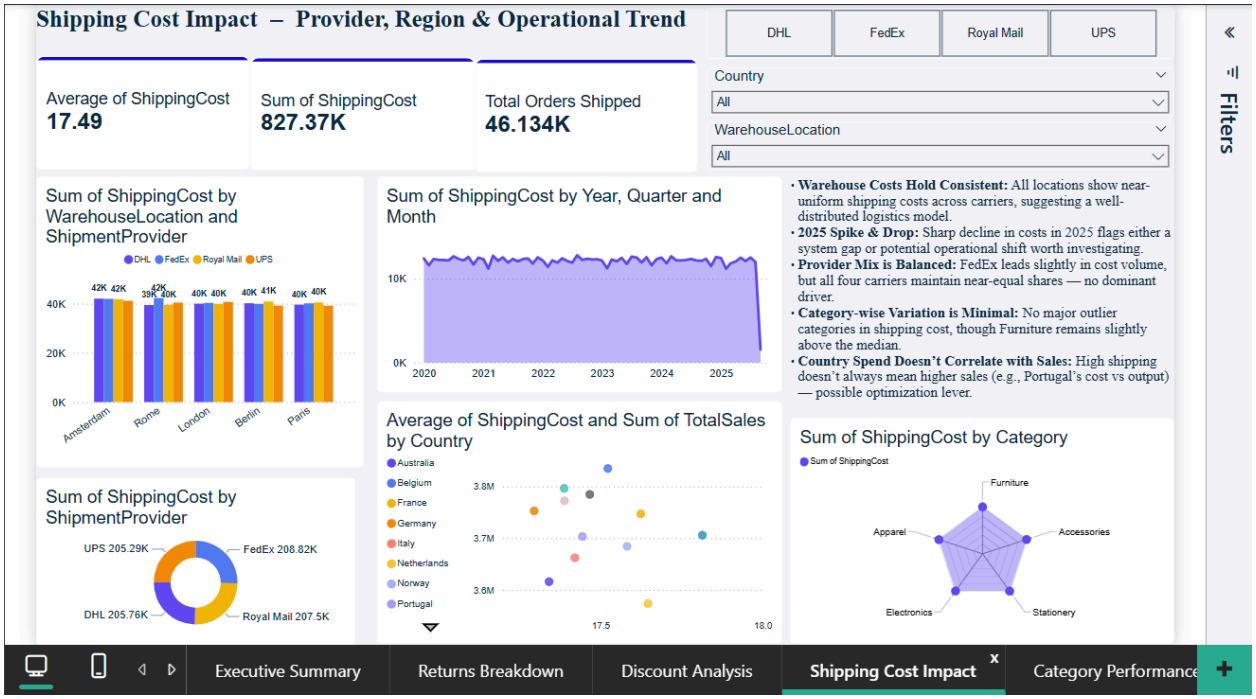
This dashboard drills down into return behavior. **Australia** has the highest return volume, while **Paris and Berlin warehouses** show elevated return rates. **Furniture and Apparel** are the most returned categories. The breakdown by payment methods shows **Bank Transfers** have slightly higher return rates, especially in **online orders**.

3. Discount Analysis – Strategic Impact & Optimization Zones



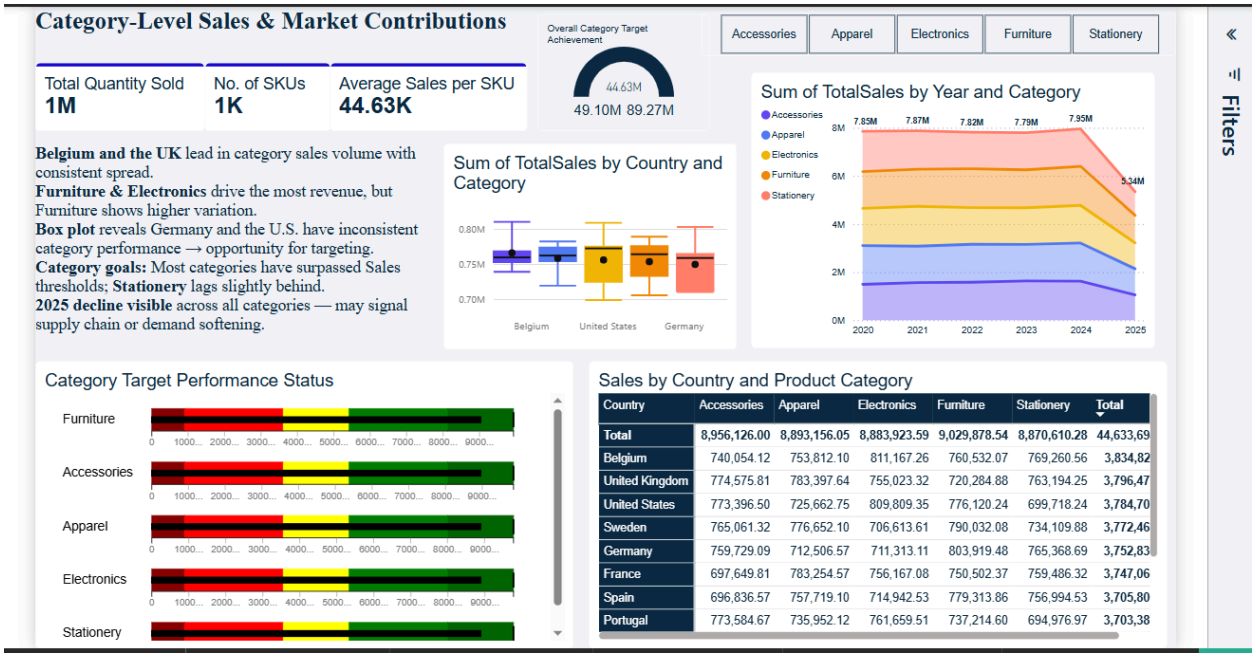
This page highlights how discount bands impact sales. A massive **43.99M in sales** come from discounted orders. **Stationery** leads in discounted volume, while **Furniture** has a risky high discount + return pattern. The scatter plot shows the **10–30% discount band** is the sweet spot—driving high sales with moderate returns. Higher discounts above 30% offer diminishing returns.

4. Shipping Cost Impact – Provider, Region & Operational Trend



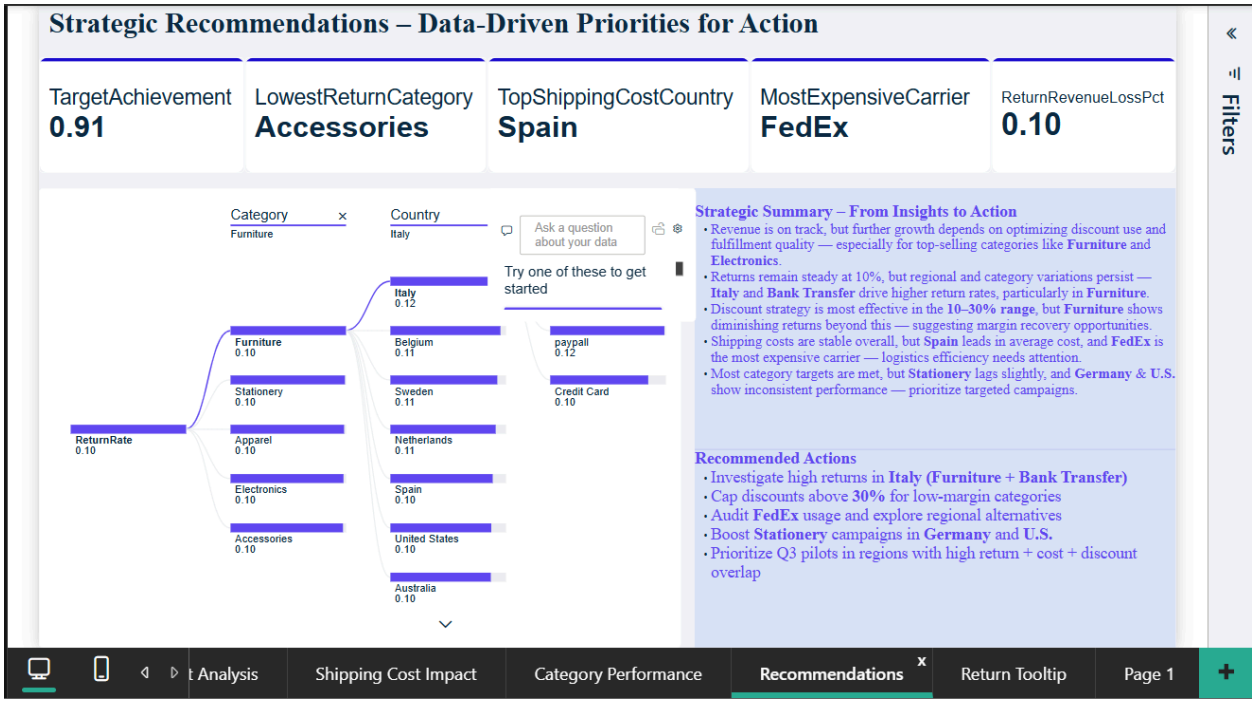
Shipping costs average **17.49 per order** with FedEx slightly leading in cost volume. Warehouse locations have **consistent costs**, suggesting an efficient logistics model. However, the sharp **drop in 2025** raises a flag for system or process issues. The scatterplot shows that **high shipping costs don't necessarily mean high sales**, offering an optimization opportunity.

5. Category Performance – Sales & Market Contributions



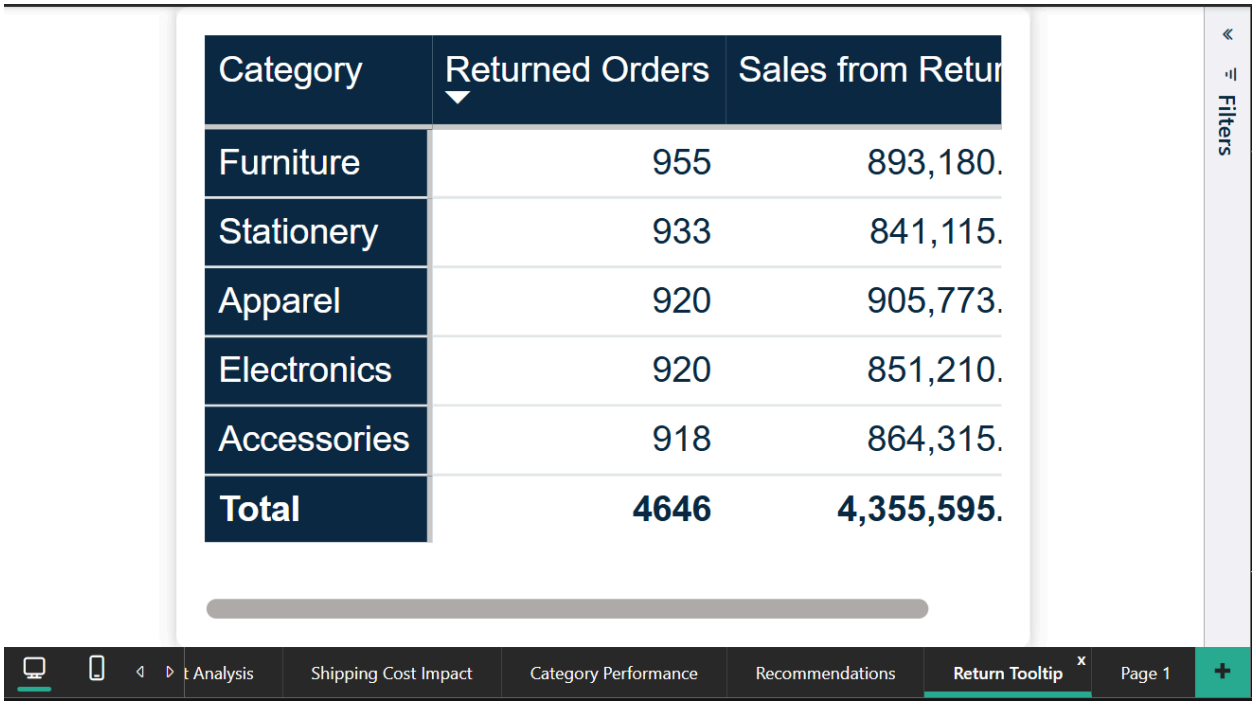
This dashboard focuses on SKU-level performance. **Belgium and the UK** lead in consistent category performance. **Furniture and Electronics** generate high revenue, but **Furniture** shows variation across regions. The **box plot** indicates inconsistency in **Germany and the U.S.**, making them ideal targets for optimization campaigns. **Stationery** slightly lags behind category targets

6. Recommendations – Data-Driven Priorities for Action



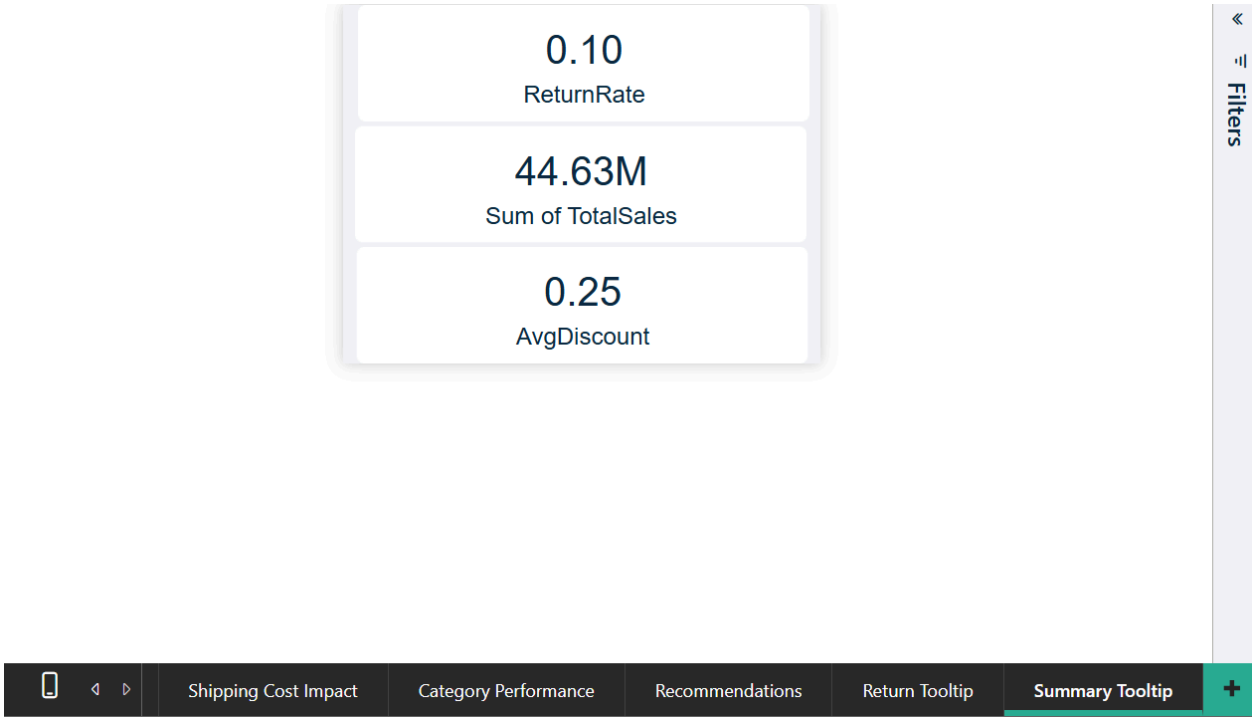
This page summarizes all insights into actionable steps. It identifies **Italy + Bank Transfers** as high-return zones, **Spain** as the most expensive shipping country, and **FedEx** as the costliest carrier. Suggestions include capping high discounts, auditing logistics costs, and targeting **Germany/U.S.** for performance improvement.

7. Return Tooltip (Table)



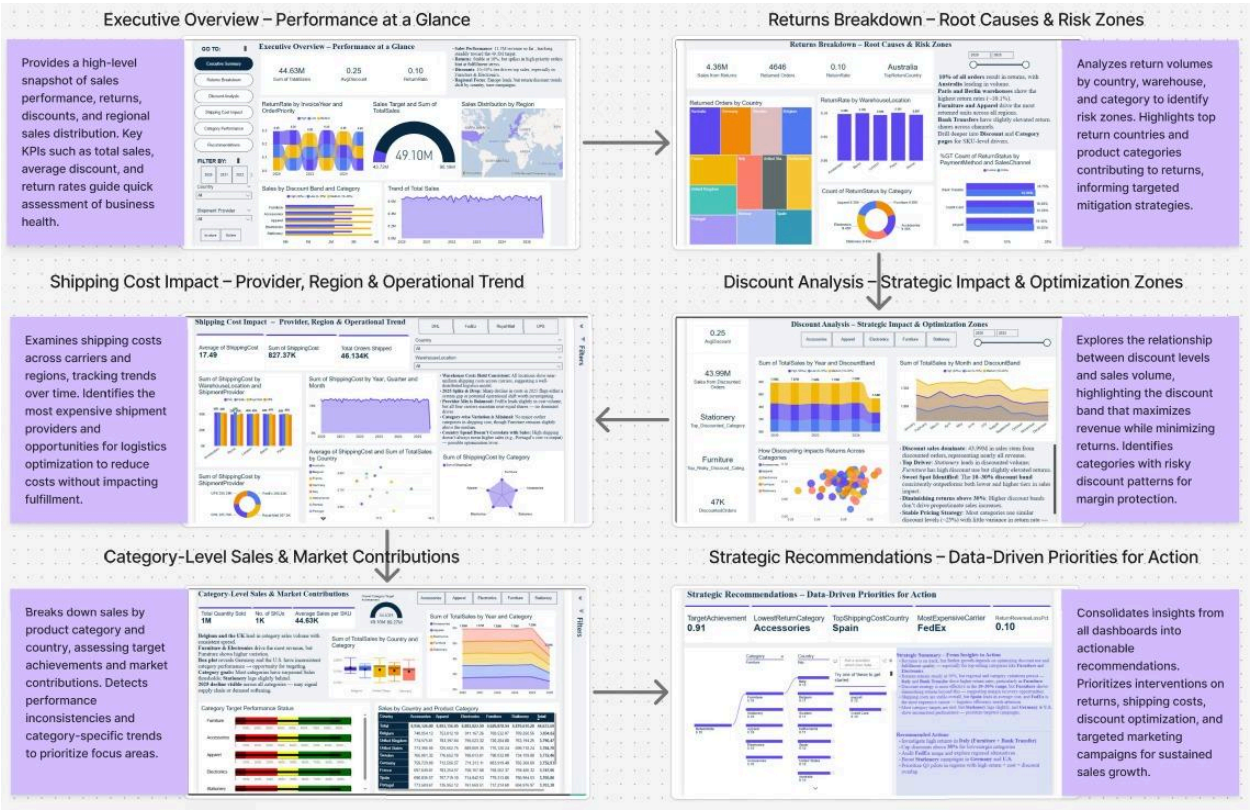
This is not a standalone page but a **tooltip** that appears when hovering over category-specific return visuals. It provides a clear breakdown of **Returned Orders** and **Sales from Returns** by product category, highlighting that **Furniture** has the highest return volume both in count and revenue.

8. Return Summary Tooltip (KPI Cards)



This tooltip offers quick contextual KPIs — **Return Rate**, **Total Sales**, and **Average Discount** — which appear when hovering over return-related metrics. It's designed to provide **at-a-glance context** without navigating away from the main dashboard.

7. Data Story



Problem Context

The business faced three core challenges:

1. Revenue optimization – understanding where sales are underperforming relative to targets
2. Return rate management – identifying what's driving high returns in certain products, locations, or payment methods.
3. Operational inefficiencies – especially around discount strategies and shipping costs that may erode profit margins

How the Data Story Solves It

1. Executive Overview – Setting the Baseline

This page establishes the foundation of your business performance by summarizing key metrics: total revenue, average discount, and return rate. It visualizes trends over time and compares performance to

targets (like the 49.1M goal). It helps identify that sales are progressing steadily but return rate remains constant and significant.

2. Returns Breakdown – Pinpointing Pain Areas

This page dives deeper into return data. It shows which countries, categories, and warehouses are most affected by returns. For instance, Australia has the highest return volume, while Furniture and Apparel lead in returned units. This isolates key contributors to loss and flags fulfillment concerns like those in Paris and Berlin warehouses.

3. Discount Analysis – Finding the Profit Sweet Spot

Here, you explore how different discount bands impact total sales and returns. A 10–30% discount range emerges as the most effective for driving sales with manageable return rates. It solves the problem of blanket discounting by identifying which discount levels are beneficial and where they backfire (e.g., returns rising with very high discounts).

4. Shipping Cost Impact – Operational Cost Control

This section breaks down shipping expenses across providers and locations. While FedEx is the costliest, there is minimal variation between carriers. Spain's shipping cost vs. sales mismatch is also revealed. This analysis helps uncover potential cost-saving opportunities and regional inefficiencies that don't align with revenue output.

5. Category-Level Sales & Market Contribution – Zooming into Product Performance

You evaluate how well each product category is contributing to overall sales, using box plots, area charts, and performance gauges. This flags that most categories have met or exceeded sales thresholds, but Stationery lags behind and Furniture shows higher variance – guiding where to double down or adjust.

6. Strategic Recommendations – Connecting All Insights into Action

The final page pulls everything together. Through a decomposition tree and clearly stated actions, it guides stakeholders toward impactful decisions:

- Limit deep discounting on high-return items
- Audit FedEx for cost optimization
- Prioritize campaigns for low-performing categories/regions (like Germany & U.S. in Stationery)

  This resolves the original problems by directly addressing inefficiencies, cost leakages, and sales underperformance through targeted, data-backed recommendations.

Final Outcome

The flow of this dashboard empowers decision-makers to move from diagnosing business pain points to implementing strategic solutions using clear, connected visuals. Each page builds logically on the previous one to create a seamless narrative that solves real operational and financial challenges.

8. Work Contribution

Rafsha cleaned the dataset, formatted the overall dashboard theme, and developed the pages for the *Executive Summary* and *Returns Breakdown*. She also ensured consistency in visual design across the dashboards.

Alina prepared the final report, summarizing the analysis, design thinking process, and data story with annotated visuals. She also developed the dashboard pages for *Discount Analysis* and *Shipping Cost Impact*.

Ahsan conducted the domain expert interview and completed the *Knowledge Excel Sheet*. He also created the dashboard pages for *Category Performance* and *Recommendations*, aligning insights with business objectives.