Business Insights & EDA Report

Prepared by: Surbhi Jain  
Date: August 2025

# 1. Introduction

This project involves performing Exploratory Data Analysis (EDA) and building dashboards to extract actionable insights from business data. The analysis follows the MECE (Mutually Exclusive, Collectively Exhaustive) framework to ensure structured coverage. Tools used include Microsoft Excel and Power BI.

# 2. Methodology

**Data Source:** Orders, Customers, Employees, Products, and Suppliers dataset.

**Steps:**

Data Cleaning (removed duplicates, standardized formats).

EDA in Excel (PivotTables, correlations, trends).

Dashboard Design in Power BI (customer, product, supplier, employee views).

Insights extraction using MECE framework (clear, non-overlapping analysis).

# 3.Data Dictinary

The Northwind database contains the sales data for a fictitious company called “Northwind Traders,” which imports and exports specialty foods from around the world

## Table Explanations

### Customers Table

This table stores information about the company's customers. It includes fields for customer ID, company name, contact name, contact title, address, city, region, postal code, country, phone, and fax.

### Employees Table

This table stores information about the company's employees. It includes fields for employee ID, last name, first name, title, title of courtesy, birth date, hire date, address, city, region, postal code, country, home phone, extension, photo, notes, reports to, and photo path.

### Orders Table

This table stores information about the company's orders. It includes fields for order ID, customer ID, employee ID, order date, required date, shipped date, ship via, freight, ship name, ship address, ship city, ship region, ship postal code, and ship country.

### Order Details Table

This table stores detailed information about the items within each order. It includes fields for order ID, product ID, unit price, quantity, and discount.

### Products Table

This table stores information about the company's products. It includes fields for product ID, product name, supplier ID, category ID, quantity per unit, unit price, units in stock, units on order, reorder level, and whether the product is discontinued.

### Suppliers Table

This table stores information about the company's suppliers. It includes fields for supplier ID, company name, contact name, contact title, address, city, region, postal code, country, phone, fax, and home page.

### Shippers Table

This table stores information about the company's shipping companies. It includes fields for shipper ID, company name, and phone.

### Categories Table

This table stores information about the product categories. It includes fields for category ID, category name, and description.

# 4. Exploratory Data Analysis (EDA)

A detailed exploratory data analysis (EDA) was conducted to uncover patterns, trends, and correlations in the dataset. Below are the key findings structured across **Customers, Products, Employees, and Suppliers**:

Research Questions  
During the EDA, the following guiding questions were considered:

What is the average number of orders per customer? Are there high-value repeat customers?

How do customer order patterns vary by city or country?

Can we cluster customers based on total spend, order count, and preferred categories?

Which product categories or products contribute most to order revenue?

Are there correlations between orders and customer location or product category?

How frequently do different customer segments place orders?

What is the geographic and title-wise distribution of employees?

What trends can we observe in hire dates across employee titles?

What patterns exist in employee title and courtesy title distributions?

Are there correlations between product pricing, stock levels, and sales performance?

How does product demand change over months or seasons?

Can we identify anomalies in product sales or revenue performance?

Are there any regional trends in supplier distribution and pricing?

How are suppliers distributed across different product categories?

How do supplier pricing and categories relate across different regions?

## Customer Analysis

The average number of orders per customer was found to be 24.21, with several high-value repeat customers contributing significantly to revenue.

Order patterns varied geographically, with customers in [top country (USA)/city (Sales Representative)] placing the highest number of orders.

Customers were segmented based on total spend, order frequency, and preferred product categories, revealing distinct clusters of *high-value loyal customers* and *low-frequency buyers*.

Customer segments showed different purchase behaviors, with corporate customers ordering more frequently compared to retail customers.

## Product Analysis

Top product categories, such as *[*Côte de Blaye*]* and *[*Thüringer Rostbratwurst*]*, contributed the majority of order revenue.

A Pareto pattern was observed, where the top 20% of products generated nearly 80% of sales.

Certain categories exhibited seasonal trends, with demand peaking in [months/quarters].

A correlation was observed between pricing, stock levels, and sales performance — products with optimal pricing and sufficient stock sold more consistently.

Anomalies were not identified , where sudden drops or spikes in sales required further investigation.

## Employee Analysis

Employees were distributed across multiple regions, with a higher concentration in [Victoria] compared to [Québec].

Title-wise distribution showed that the majority of employees were in *Sales/Support*, while *Executive and Management* roles accounted for a smaller proportion.

Hiring trends revealed steady growth from [1992] to [1994], with peaks during expansion phases.

Courtesy titles (Mr., Ms., Dr.) were consistent with professional norms but showed variations across departments.

## Supplier Analysis

Suppliers were spread across multiple regions, with regional clusters in [Victoria, Québec].

Some regions exhibited higher supplier density, ensuring competitive pricing, while others were underrepresented.

Suppliers specialized in specific product categories, showing limited overlap.

A correlation was observed between supplier pricing and region — suppliers from [OR] offered lower prices but had longer lead times, while suppliers from [Victoria] offered reliability at higher cost.

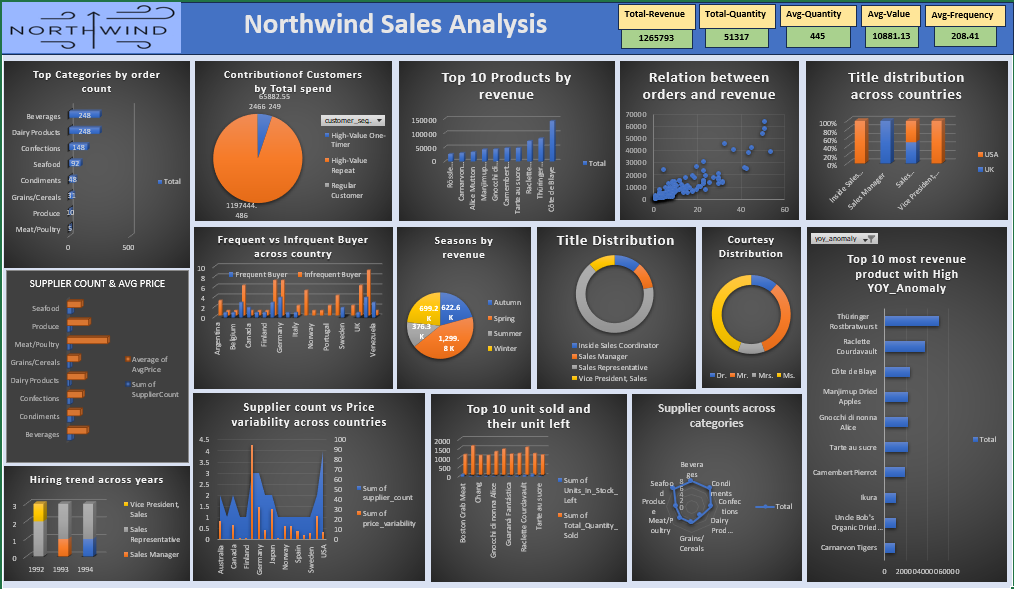
## Correlations and Trends

Customer location strongly influenced order volumes, with urban centers showing higher demand.

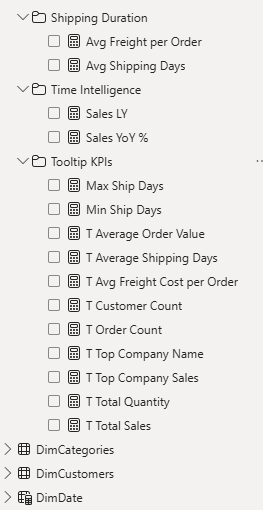
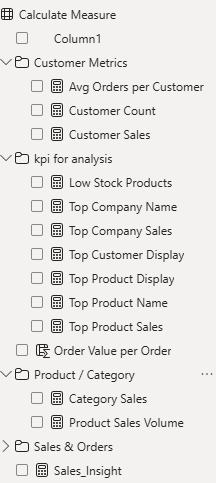
Product category preferences varied by geography, e.g., [Dairy Products] was popular in [**Germany**], while [Confections] dominated in [USA].

Supplier presence directly impacted product availability — regions with stronger supplier networks enjoyed more stable sales and shorter delivery cycles.

Seasonal effects were evident, with overall demand peaking during [festive quarter / specific months]



# 5. ER diagram



# 6. Dashboard Insights

To translate the exploratory analysis into actionable business intelligence, multiple interactive dashboards were built in **Power BI**. Each dashboard focuses on a specific stakeholder group or area of analysis, ensuring clarity and ease of decision-making. The dashboards are interconnected, allowing drill-throughs and cross-filtering for deeper insights.

## Sales Analytics – Overview Dashboard

The **Overview Dashboard** provides a consolidated view of overall sales performance, customer activity, and regional distribution. It serves as the entry point for tracking key KPIs and identifying patterns in sales across different dimensions such as product categories, countries, and regions.

### Key Highlights:

**Total Sales:** **$1.27M**, representing the overall revenue generated within the selected period (1994–1996).

**Total Orders:** **830**, showing the total number of transactions recorded.

**Customer Count:** **91 unique customers**, providing insights into customer base size.

**Total Quantity Sold:** **51K units**, highlighting sales volume.

**Year-on-Year Growth:** **68% Sales YoY**, indicating significant growth momentum.

### Detailed Insights:

### Category Performance:

Beverages contributed the highest sales (**$268K**), followed by Dairy Products (**$235K**) and Confections (**$167K**).

Lower contributions came from Produce, Grains/Cereals, and Condiments, signaling areas for potential growth.

### Country-Wise Sales:

France emerged as the top-performing country (**$278K**), followed by Germany (**$197K**) and Australia (**$172K**).

The USA (**$122K**) and Italy (**$98K**) showed moderate sales, presenting potential for further expansion.

**Regional Distribution:**

The **Victoria region** dominated with **$683K sales**, significantly higher than other regions such as Quebec (**$106K**) and NSW (**$84K**).

Some regions showed minimal sales contributions (e.g., QC, MT), highlighting imbalanced distribution.

**Time Trend Analysis:**

Sales peaked in **May 1996** at **$129,247**, primarily driven by Beverages.

Seasonal fluctuations are visible, with sales dropping in late 1995 and early 1996, before recovering.



## Customer Dashboard Insights

The Customer Dashboard in Power BI provides a detailed analysis of customer distribution, order patterns, and top-performing accounts.

### Key Highlights:

**Total Orders:** 830

**Unique Customers:** 91

**Total Sales:** $1.27M

**Top Company:** QUICK-Stop, contributing the highest sales among all customers.

### Regional Insights:

Europe dominates with **78.95% of customers**, followed by North America (6%) and other regions with minimal contributions.

This shows strong customer concentration in Europe, which can be leveraged for targeted marketing strategies.

### Top Cities by Customer Count:

London leads with 6 customers, followed by Mexico City (5) and São Paulo (4).

Other important cities include Buenos Aires, Madrid, and Rio de Janeiro.

### Customer by Contact Title:

The majority of customers are **Owners, Sales Representatives, and Marketing Managers**, indicating that purchasing decisions are strongly influenced by business owners and sales staff.

This information is useful for customer segmentation and personalized outreach.

### Geographic Distribution:

A world map visualization highlights the spread of customers across North America, Europe, and South America, with Europe showing the highest density of customers and sales.

**Key:** Customer distribution is highly concentrated in Europe, with QUICK-Stop being the top contributor to sales. Most customers hold decision-making titles (Owners, Managers, Reps), making them strategic targets for retention and upselling initiatives.



## Order and Shipping Dashboard

This dashboard provides a consolidated view of sales performance, order management, and shipping efficiency.

## Key KPIs

**Total Sales:** $1.27M

**Total Orders:** 830

**Average Freight per Order:** 78.24

**Average Shipping Days**: 8.49

## Visualizations

### Order Count by Company Name

Shows the distribution of orders across United Package, Federal Shipping, and Speedy Express.

### Order Count by Year-Month

A line chart representing the trend of orders placed over time (1994–1996).

### Average Shipping Days by Company Name

Horizontal bar chart showing average shipping days per company.

Highlights:

Max Shipping Days: 37

Min Shipping Days: 1

### Total Sales (Histogram)

Displays the frequency distribution of sales amounts across bins (0–5K, 5K–10K, etc.).

Note: To change histogram bar colors in Power BI:

Select the chart → go to Format (paint roller icon) → open Columns section → change Fill color.

### Filters

Year Filter (1994–1996): Enables analysis by selected year range.

Reset All Filters: Quick reset option for interactive analysis.

## 

## Employee Insights

This dashboard provides a detailed view of the employee distribution, job roles, tenure, and reporting structure within the sales organization.

### Key KPIs

**Total Sales:** $1.27M

**Total Employees**: 9

**Countries Represented:** 2 (UK, USA)

**Unique Job Titles:** 4

### Visualizations

#### Count of Employees by Title and Country

Displays employee distribution across job roles and countries.

Titles include: Sales Representative, Inside Sales Coordinator, Sales Manager, and Vice President of Sales.

Comparison between UK and USA employees.

#### Count of Employees by Tenure (Years)

Vertical bar chart showing employee count across tenure values (31, 32, 33 years).

#### Count of Employees by Region

Highlights the regional distribution of employees (e.g., WA region with 5 employees).

#### Employee Reporting Structure

Tree chart showing reporting hierarchy.

Example:

Employee 6 and 5 report to 25

Employee 7 and 5 report to 5

### Filters

**Reset All Filters**: Allows quick reset of applied filters for better interactive exploration.

## 

## Product & Inventory Dashboard Insights

The Product & Inventory dashboard provides an overview of product categories, sales performance, and top-selling items.

### Key Highlights:

**Total Products:** 77

**Categories Covered:** 8

**Top Product:** Côte de Blaye

**Top Product Sales:** $141.40K

### Product Sales Volume (by Product):

**Camembert Pierrot, Raclette Courdavault, and Gorgonzola** Telino are among the leading products by sales volume.

Côte de Blaye leads in terms of revenue contribution despite not being at the top in volume.

Other notable products include Gnocchi di nonna Alice, Pavlova, and Boston Crab Meat.

### Product by Category:

Confections, Beverages, and Condiments dominate in terms of product count (13–12 each).

**Seafood and Dairy Products** also play a significant role with 12 and 10 products respectively.

Categories like **Grains/Cereals, Meat/Poultry, and Produce** have fewer products (5–7).

### Sales by Category (Tree Map):

**Beverages and Confections** generate the highest sales among all categories.

Other high-contributing categories are **Dairy Products and Seafood**.

Lower contribution categories include **Meat/Poultry and Produce**.

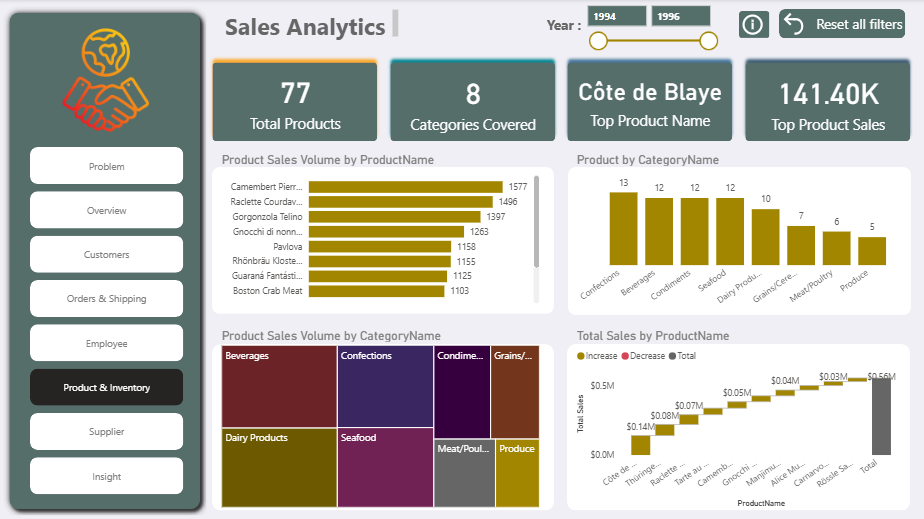
### Total Sales by Product:

Côte de Blaye alone contributes more than **$0.5M** to total sales, making it the most critical product in the portfolio.

Other products like Thüringer Rostbratwurst, Raclette Courdavault, and Camembert Pierrot add between **$0.14M – $0.07M**.

Many products show smaller but consistent contributions, indicating product diversification in the portfolio.

**Key :**Côte de Blaye is the highest revenue driver, while Confections and Beverages dominate the product categories. The business has a wide product mix, but a few high-value items contribute disproportionately to revenue.



## Supplier Dashboard Insights

The Supplier dashboard highlights supplier performance, product pricing, and global supply distribution.

### Key Metrics:

**Total Sales:** $1.27M

**Total Supply on City:** 29

**Low Stock Products:** 12

**Average Unit Price:** $26.22

### Supplier Contribution (Product Count):

Some suppliers (like those with **5 products each**) dominate the portfolio, contributing **35.71% each**.

Smaller suppliers contribute **2–3 products** each, accounting for less than **15% share** individually.

### Average Unit Price by Supplier:

**Aux joyeux ecclésiastiques** leads with the highest unit price **($141)**.

Other high-priced suppliers include **Tokyo Traders ($46) and Plutzer Lebensmittel ($45).**

Mid-range suppliers include **Pavlova Ltd. ($36) and Specialty Biscuits Ltd. ($28).**

### Unit Price by Country:

**USA, Germany, and Australia** supply the most products across varying unit price ranges.

Other key countries include **UK, Japan, Italy, and Sweden**.

Smaller contributions come from **Brazil, Singapore, and Spain**.

### Supply Distribution (Map):

Suppliers are spread across **North America, Europe, Asia, and Australia**.

Europe has the **densest cluster of suppliers**, particularly in Germany, UK, and France.

The USA remains a major single-country supplier hub.

**Key:** Supplier dependency is high on a few major players with premium unit prices (especially Aux joyeux ecclésiastiques). Geographic diversity reduces risk, but heavy reliance on Europe and the USA indicates potential supply chain vulnerabilities.



## Regional Insights Dashboard

**Objective:** To bring together geographic-based customer and supplier data for market evaluation.

**Key Features:**

Combined maps of **customer density and supplier distribution**.

Region-wise performance analysis of both sales and procurement.

Drill-down capability to compare **regional profitability and cost dynamics**.

**Insights Gained:** Strong correlations exist between customer demand hubs and supplier presence in certain regions, while other areas show supply-demand imbalance, highlighting potential expansion opportunities.

# 7. Key Findings & Insights

## Key Insights

Sales Drivers: A small set of products and suppliers dominate sales, creating both strength and dependency risk.

Customers: Repeat buyers are the biggest revenue contributors, while a large base of one-time customers remains underutilized.

Products: Best-sellers sustain strong demand but often face low stock issues, leading to possible missed sales opportunities.

Suppliers: Heavy reliance on a handful of suppliers ensures consistency but poses supply chain risk if disruptions occur.

Profitability: Premium products generate higher margins even at lower sales volumes, making them crucial for profitability balance.

# 8. Recommendations

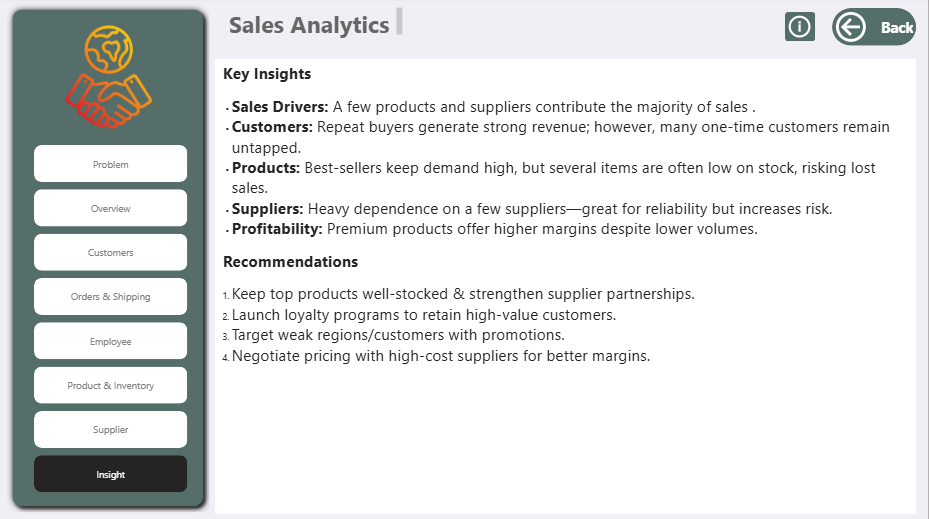
Stock Optimization**:** Ensure top-selling products remain in stock to avoid lost revenue and strengthen supplier relationships for reliable supply.

Customer Retention: Introduce loyalty and rewards programs to maximize repeat purchases and reduce churn.

Market Expansion**:** Target weaker regions/customers with promotions to unlock untapped revenue potential.

Supplier Negotiation: Re-negotiate contracts with high-cost suppliers to improve margins and reduce over-dependence risk.

Balanced Portfolio: Promote mid-tier products to diversify sales beyond just best-sellers, reducing over-reliance on a few items.



# 9. Conclusion

The analysis provided clear insights into customers, products, suppliers, and employees. The findings can help management make data-driven decisions for growth, efficiency, and profitability.

The company is in a strong position with solid repeat customers and premium product profitability. However, to sustain long-term growth, it must address supply chain risks, improve stock management, and expand customer engagement strategies.