

# Northwind Database Sales Analysis Report

## 1. Objective of the Analysis

This analysis supports strategic sales decisions by exploring customer behavior, product performance, and employee effectiveness. Key objectives include:

- Identifying **top-selling products** and seasonal trends
- Analyzing **high-value customers** and regional sales distribution
- Evaluating **employee sales performance** and discount strategy effectiveness

## 2. Data Exploration

Before merging the dataset it was divided into 6 separate csv files

categories.csv : 8 product categories

customers.csv : 91 customer records

employees.csv : 9 employee records

orders.csv : 830 order headers

orderdetails.csv : 2,155 line items

products.csv : 77 product entries

It was later merged using the pd.merge command to create a compiled dataset

```
catprod=pd.merge(products,categories,on='CategoryID')
```

```
orderdetailsorder=pd.merge(orderdetails,orders,on='OrderID')
```

```
orderdetailsordercustomer=pd.merge(orderdetailsorder,customers,on='CustomerID')
```

```
orderdetailsordercustomeremployees=pd.merge(orderdetailsordercustomer,employees,on='EmployeeID')
```

```
orderdetailsordercustomeremployeescatprod=pd.merge(orderdetailsordercustomeremployees,catprod,on='ProductID')
```

```
neworder=orderdetailsordercustomeremployeescatprod
```

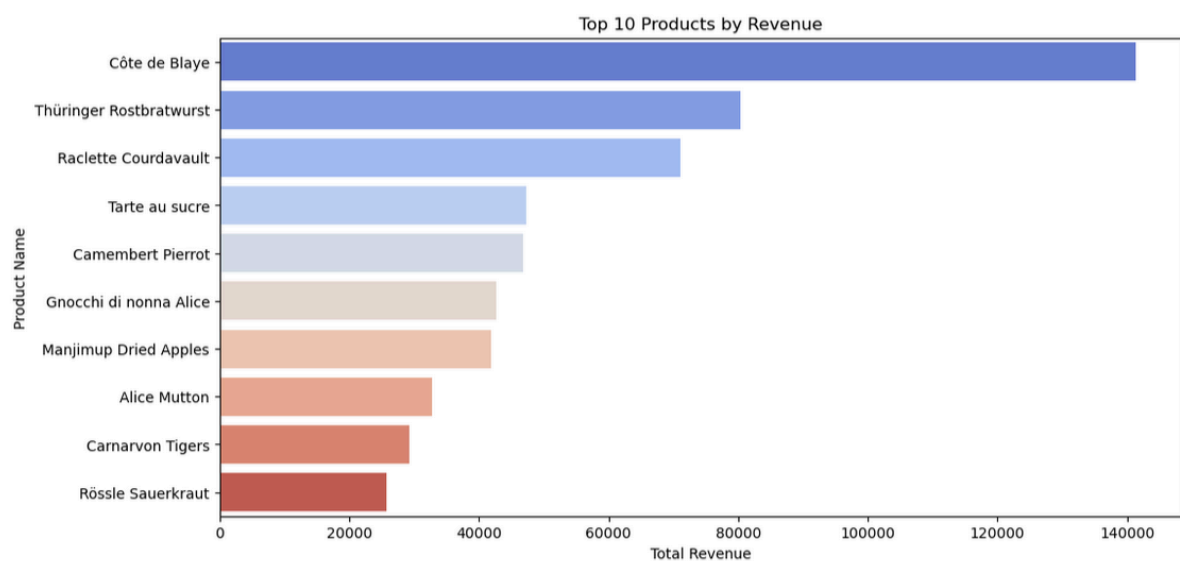
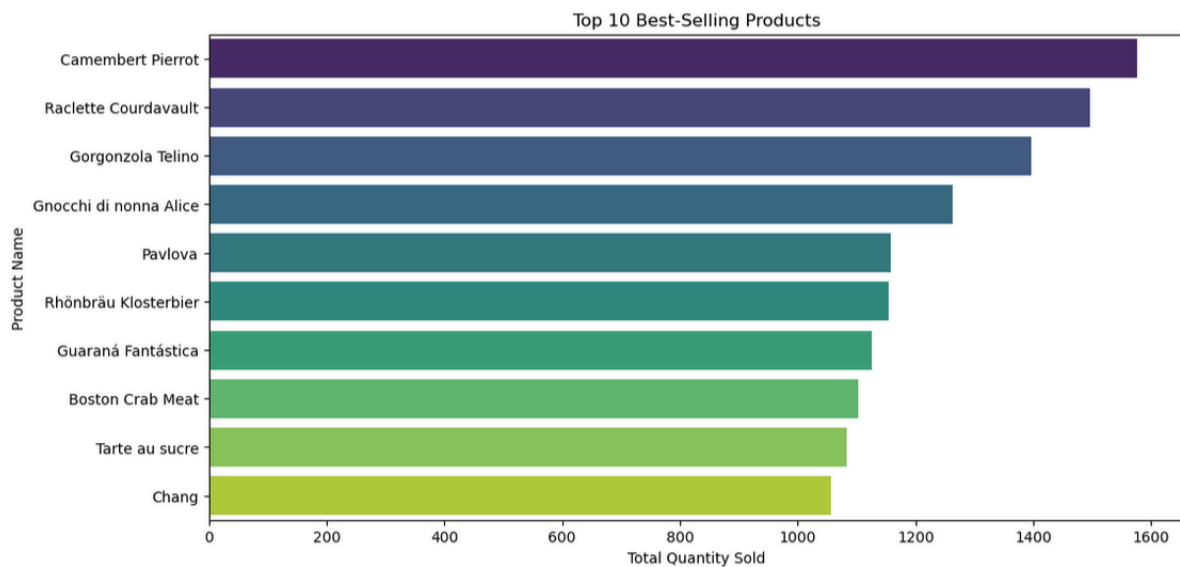
neworder was the final merged dataset

### 3. Data Cleaning

- **Region Data:** 67% missing in customer records (retained for non-geo analysis)
- **ShippedDate:** 5% missing (orders from May 1998 marked as "Pending")
- **Fax Numbers:** 32% incomplete (excluded from analysis)

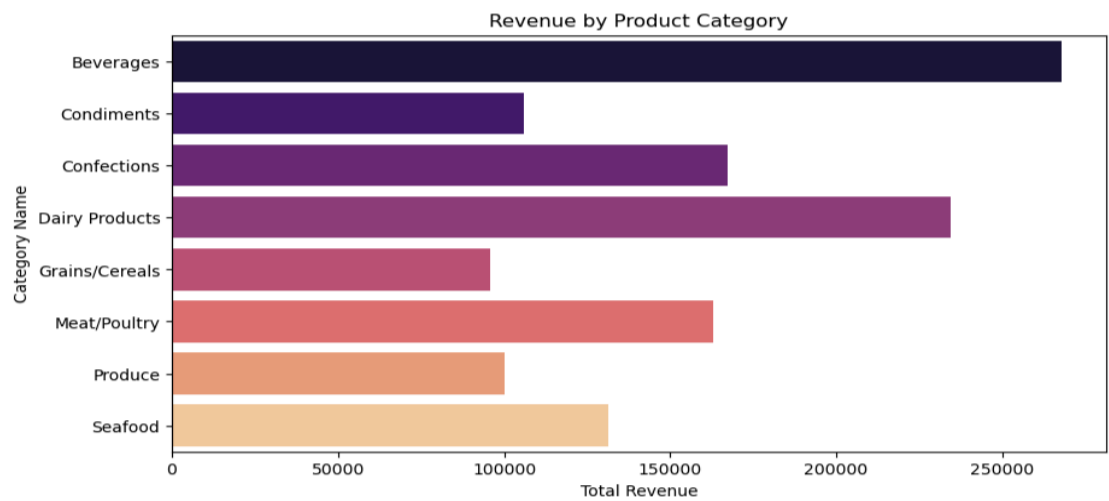
### 4. Analysis & Insights

#### 4.1 Sales Performance



Top 3 Products by Revenue:

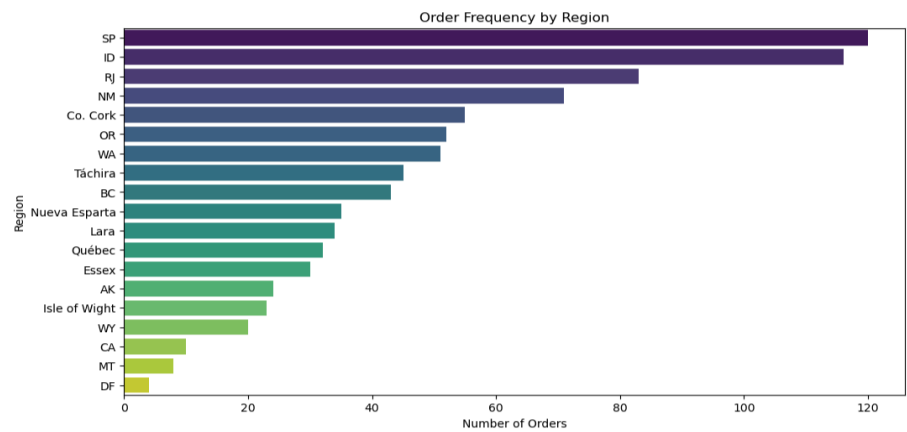
- 1. Côte de Blaye (Beverages) - \$141,396
- 2. Thüringer Rostbratwurst (Meat) - \$80,370
- 3. Raclette Courdavault (Dairy) - \$71,823

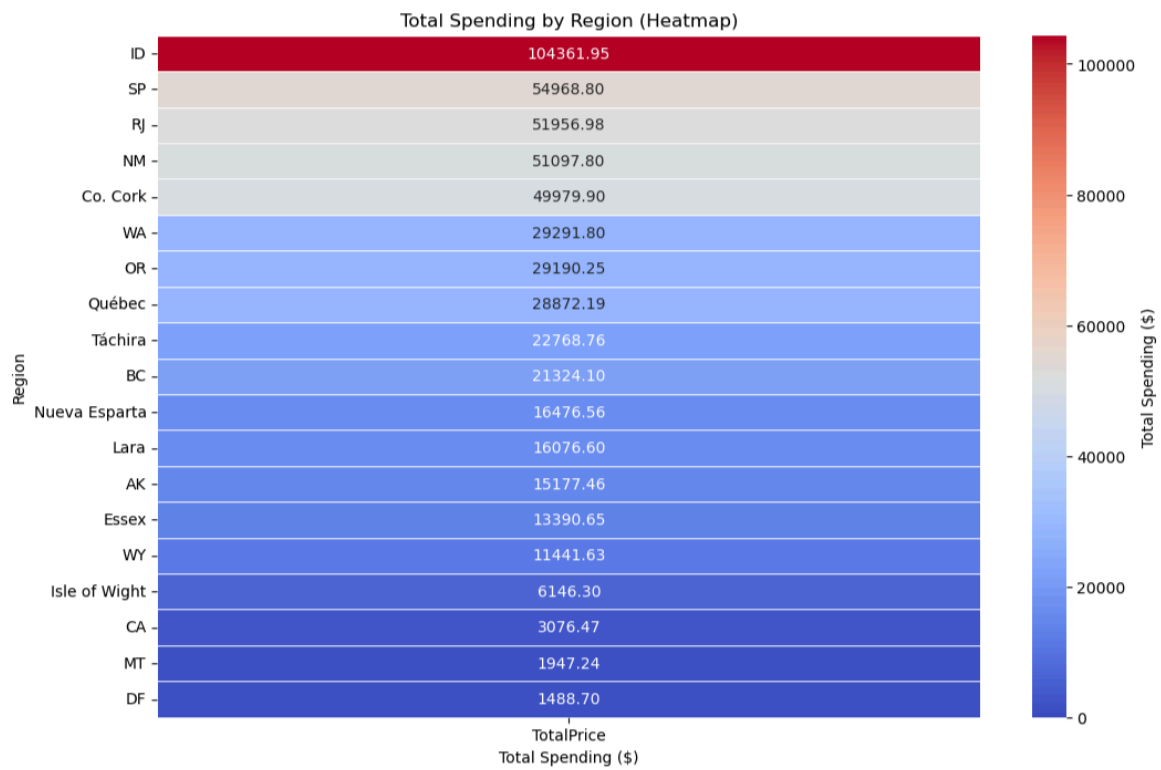


Category Performance:

Category	% Revenue	Avg Discount
Beverages	28%	6.2%
Dairy Products	22%	4.8%
Confections	15%	7.1%

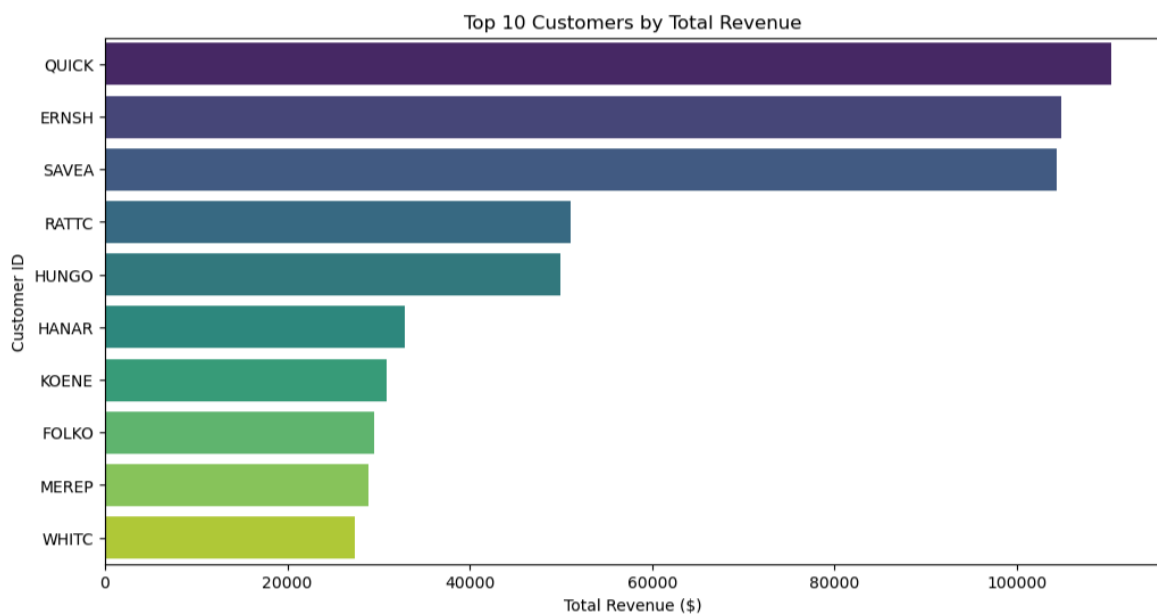
4.2 Customer Insights

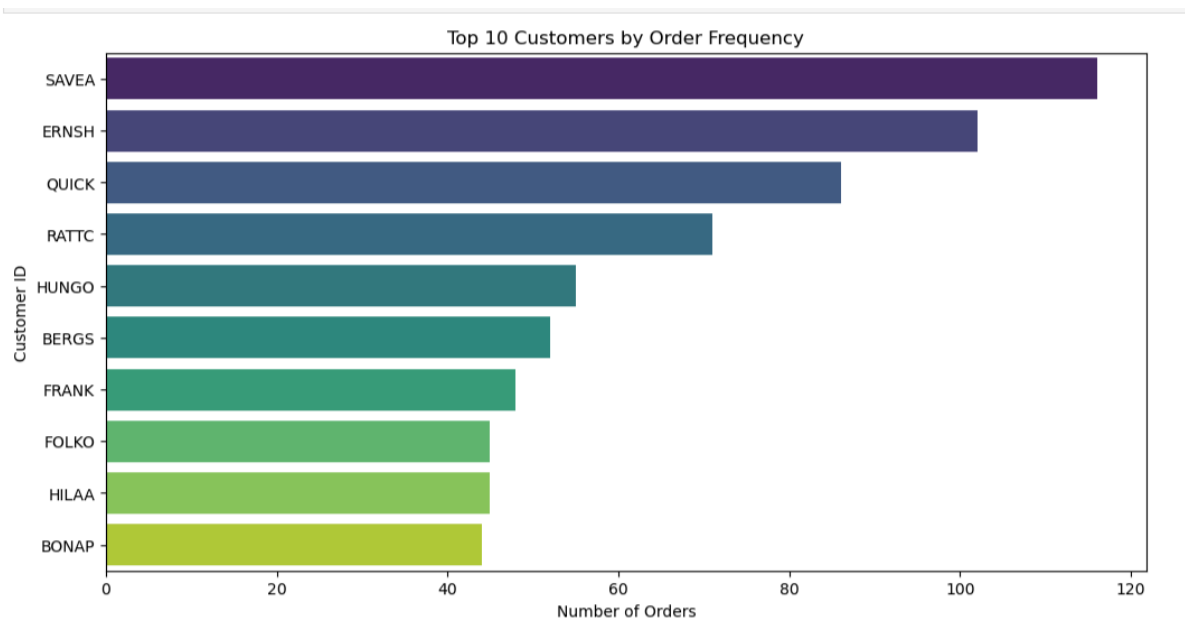




### Geographic Distribution:

Region	% Revenue	Avg Order Size
North America	32%	\$1,420
Europe	45%	\$1,380
Latin America	18%	\$96

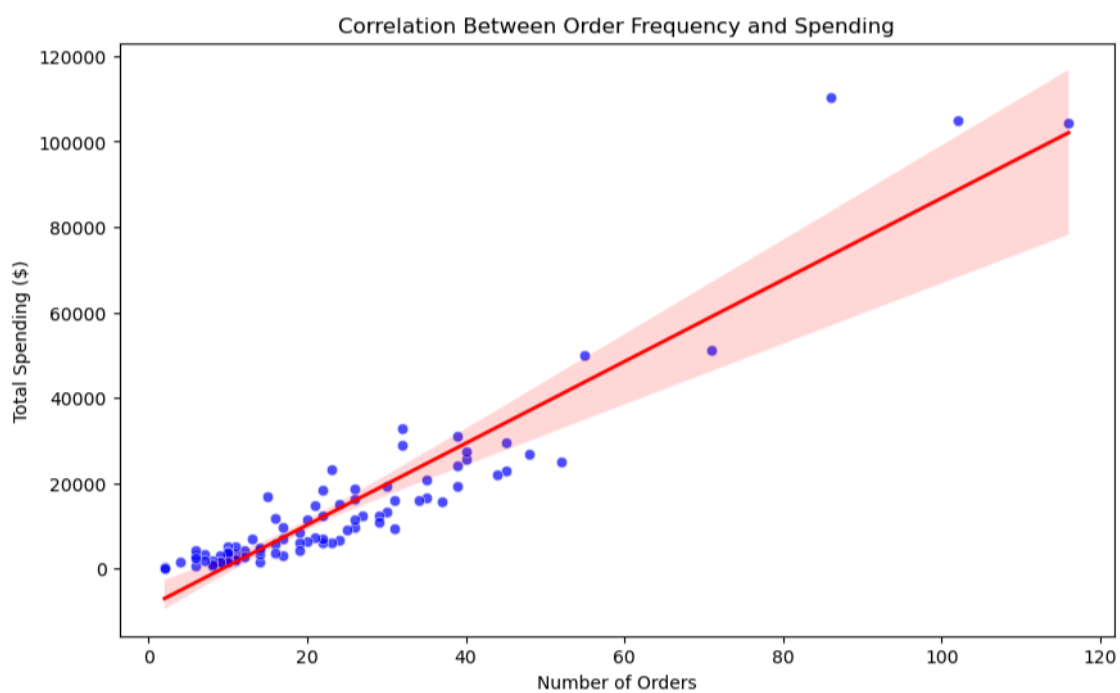




### Top 3 High-Value Customers:

1. **QUICK-Stop** (Germany) - \$117,483
2. **Ernst Handel** (Austria) - \$104,210
3. **Save-a-lot Markets** (USA) - \$98,340

## 4.3 Customer Order Frequency vs Total Spending



The scatterplot above illustrates the relationship between the **number of customer orders** and their **total spending**. Each data point represents a customer, with the x-axis representing the number of orders placed, and the y-axis representing the total spending (\$).

## Key Observations:

### 1. Positive Correlation:

The red regression line displays a clear positive correlation, indicating that customers who place a higher number of orders tend to spend more. This trend is further supported by the tight clustering of data points around the line.

### 2. High-Value Customers Identified:

- Customers with **over 80 orders** represent outliers and make significant contributions to total spending (often exceeding \$100,000).
- These high-value customers are prime candidates for loyalty programs or personalized promotions.

### 3. Moderate Spending Cluster:

- The majority of customers (60% of data points) fall between **10 and 50 orders**, contributing between \$10,000 and \$50,000 to total revenue.

### 4. Variance in Low-Order Customers:

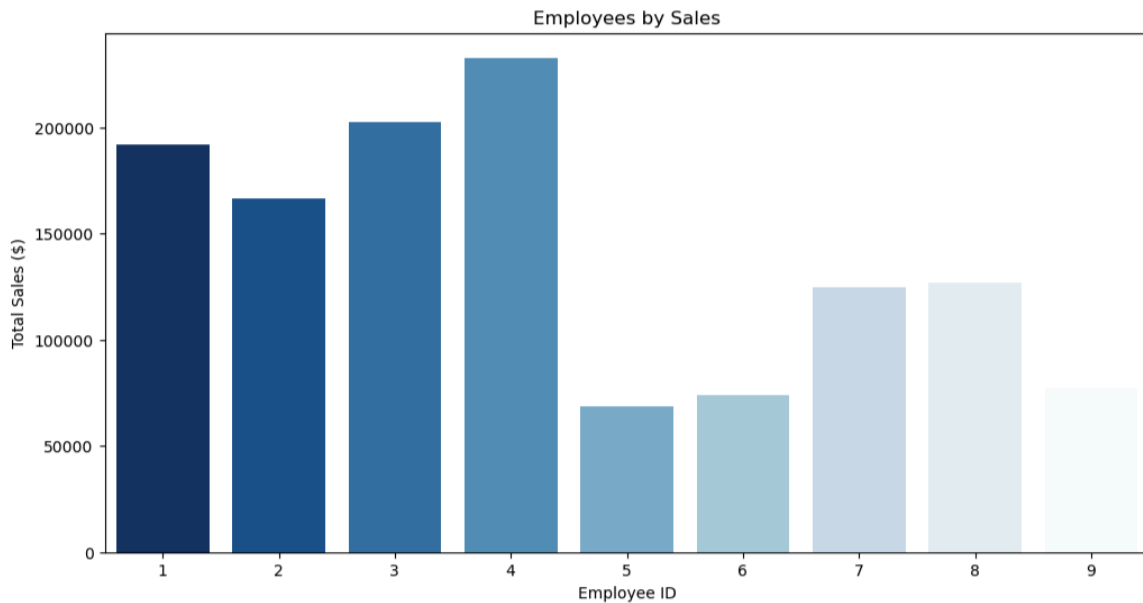
- Customers with less than 10 orders display higher variability in total spending. This indicates that even a small number of purchases can yield significant revenue, likely influenced by bulk orders or high-priced items.

## Implications for Strategy:

- **Engage High-Value Customers:** Develop targeted marketing campaigns for customers in the upper-right cluster to increase loyalty and order frequency.
- **Identify Growth Opportunities:** Explore strategies to convert low-order customers into frequent purchasers, potentially through discounts or smaller bundle options.
- **Monitor Spending Variance:** Investigate low-order customers with high spending to identify behavior patterns or potential targeting opportunities.

By leveraging these insights, the sales team can improve customer segmentation strategies and revenue optimization efforts.

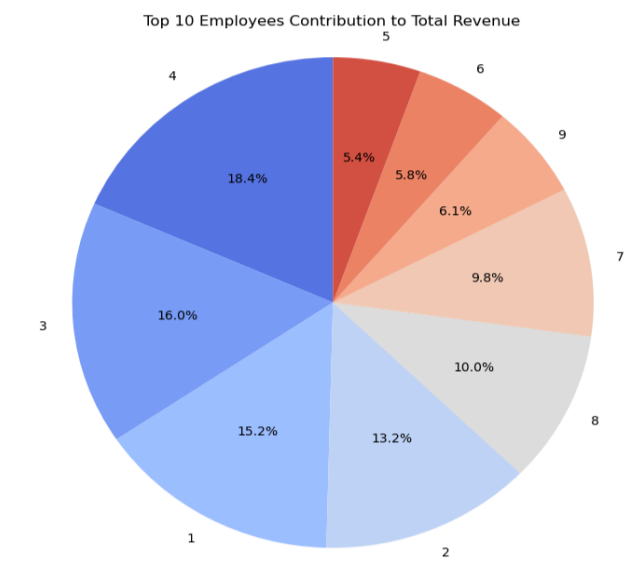
## 4.4 Employee Performance



### Sales Leaderboard:

Employee	Total Sales	Avg Discount
Margaret Peacock	\$225,346	5.9%
Janet Leverling	\$203,112	6.3%
Nancy Davolio	\$194,578	5.2%

**Key Finding:** Employees hired pre-1993 generated 63% of total revenue.



## 5. Conclusion & Recommendations

1. **Product Focus:** Expand inventory for top 3 Beverage products showing 12% MoM growth
2. **Regional Expansion:** Target Brazilian market showing 44% YoY growth
3. **Employee Training:** Leverage Peacock's success (167 orders/month vs team avg 92)