Annual Grocery Sales Analysis of the Prestige Group of Companies (2014)

Prepared by: Adaobi Cynthia Okonkwo

Tools Used: Excel, Data Visualization Techniques

Category: Sales Data Analytics

1. Outline

This report follows a structured roadmap to present the analysis clearly and logically:

o   Introduction

o   Story of Data

o   Data Splitting and Preprocessing

o   Pre-Analysis

o   In-Analysis

o   Post-Analysis and Insights

o   Data Visualizations & Charts

o   Recommendations and Observations

o   Conclusion

o   References & Appendices

2. Introduction

This project analyzes sales performance for the Prestige Group of Companies in 2014, focusing on understanding the factors that influence revenue generation and customer purchasing behavior. The goal is to identify sales trends, highlight top-performing products and regions, and provide insights that can enhance business strategy and profitability.

The analysis specifically examines sales outcomes across product categories, customer demographics, regions, and payment types. The core problem being addressed is optimizing sales performance and uncovering patterns that can inform marketing, inventory, and sales planning decisions.

A structured sales dataset was used, and Microsoft Excel served as the primary analytical tool. Key Excel methods applied include data cleaning, pivot tables, charts, sorting, filtering, and descriptive analysis.

3. Story of Data

o   **Data Source:** The data originates from the 2014 sales records of Prestige Group of Companies, representing real-world transactions across multiple products, customers, and regions.

o   **Data Collection:** The data was compiled from the company's internal sales and order management systems. Each record captures details of customer purchases, product information, and sales outcomes.

- o **Data Structure:**

  - Rows: Each row represents a single sales transaction.

  - Columns: Variables include Order ID, Order Date, Customer ID, Customer Name, Address, City, State, ZIP/Postal Code, Country/Region, Salesperson, Region, Shipped Date, Shipper Name, Ship Name, Ship Address, Ship City, Ship State, Ship ZIP/Postal Code, Ship Country/Region, Payment Type, Product Name, Category, Unit Price, Quantity, Revenue, and Shipping Fee.

- o **Important Features and Their Significance:**

  - Product Name & Category: Helps identify top-selling products and high-performing categories.
  - Revenue & Unit Price: Measures profitability and sales contribution.
  - Region & Salesperson: Provides insights into geographical and staff performance trends.
  - Payment Type & Shipping Fee: Highlights customer preferences and cost implications.

Additional engineered features, such as High Value Shipment and Weight category, were added to enrich the pattern

- o **Limitations:** Potential limitations include missing customer or shipping information, occasional inconsistencies in transaction records, and uneven distribution of sales across regions or products.

4. Data Splitting and Preprocessing

- o **Data Cleaning:** Steps included removing duplicate transactions, correcting errors in customer and product information, and ensuring consistency in formats (e.g., currencies).
- o **Handling Missing Values:** Missing entries in customer details, shipping addresses, and product information were addressed using Excel's built-in functions. In cases where data was critical and missing, affected rows were reviewed and corrected where possible.
- o **Data Transformations:** No normalization or standardization was necessary for this descriptive analysis.
- o **Data Splitting:** Variables were separated into:

- Independent Variables: Order ID, Order Date, Customer ID, Customer Name, Address, City, State, Country, Salesperson, Region, Shipped Date, Shipper Name, Ship State, Ship Country, Payment Type, Product Name, Category.

- Dependent Variable: Unit Price, Quantity, Revenue, Shipping Fee.

This separation ensures the analysis clearly identifies which factors influence revenue and overall sales performance.

o **Industry Context:** The data belongs to the retail sector, particularly e-commerce and product distribution. Understanding sales patterns is crucial for inventory management, marketing strategies, and revenue optimization.

o **Stakeholders:** Key stakeholders include the sales and marketing teams, inventory managers, and senior management, all of whom can use the insights to make informed business decisions.

o **Value to the Industry:** Insights from this analysis can help retail companies optimize product offerings, target high-value customers, allocate resources efficiently, and ultimately increase profitability and operational efficiency.

5. Pre-Analysis

The pre-analysis phase provides an initial understanding of the dataset and uncovers early trends before deeper statistical evaluation.

o **Identify Key Trends:** Early exploration of the 2014 sales data revealed several notable patterns:

- Top-Performing Product Categories: Certain categories consistently generated higher revenue and sales volume compared to others.

- Regional Sales Differences: Some regions showed stronger sales performance, suggesting differences in customer demand or market penetration.

- Salesperson Impact: A few sales representatives appear repeatedly in high-revenue transactions, indicating potential variations in sales effectiveness.

- Seasonal Purchasing Behavior: Certain months reflected spikes in customer purchases, hinting at seasonal demand or promotional periods.

o **Potential Correlations:** Initial observations suggest possible relationships between key variables:

- Unit Price vs. Revenue: Higher-priced products frequently contribute more to total revenue, even when sold in smaller quantities.

- Category vs. Quantity Sold: Some product categories sell in higher volumes despite lower prices, indicating strong customer preference.

- Region vs. Shipping Fee: Certain regions appear to incur higher shipping fees, possibly due to distance or logistics constraints.

- Payment Type vs. Purchase Frequency: Specific payment methods seem more common in high-value transactions.

o **Initial Insights:**

- Early analysis indicates that product category, region, unit price, and salesperson performance are likely influential factors in overall revenue generation.

- Some lower-priced items generate significant revenue through sheer volume, while premium products contribute disproportionately through high unit prices.

These emerging patterns highlight key areas such as regional demand, product performance, and customer purchasing behavior that warrant deeper investigation.

6. In-Analysis

This stage represents the core of the analytical process, where deeper exploration of the dataset reveals verified and emerging patterns. Using Excel-based techniques such as PivotTables, correlation checks, conditional formatting, and lookup functions, several insights were uncovered.

o **Unconfirmed Insights (Hypotheses Under Evaluation):** During initial exploration, several patterns appeared promising but require further validation with statistical tools:

- Shipping Fee vs. Revenue Relationship: Higher-value orders appear to be associated with higher shipping fees. This may indicate that heavier or bulkier products cost more to ship, or that high-revenue orders require premium shipping services.

- Region and Sales Performance: Certain regions show consistently higher total revenue, potentially due to stronger customer bases, better logistics accessibility, or higher product demand in those locations.

- Category Influence on Sales: Products in specific categories (e.g., electronics, furniture) seem to generate significantly higher revenue than others. The combination of unit price and quantity ordered may be shaping this trend.

- Order Date vs. Shipped Date: Some orders appear to have longer processing times before shipping. This could indicate stock issues or regional efficiency variations.

- Payment Type Patterns: Payments made through certain methods (e.g., card vs. bank transfer) may correlate with larger orders, though this requires deeper verification.

o **Recommendations (Preliminary and Based on Emerging Patterns):** Although the analysis is ongoing, early patterns allow for initial recommendations:

- Improve Fulfillment Speed: Review regions where the gap between order date and shipped date is longer. Streamlining processing will improve delivery timelines and customer satisfaction.

- Optimize Shipping Cost Structure: If high shipping fees are linked to specific product categories, consider negotiating better rates with shipping providers or adjusting pricing strategies.

- Strengthen High-Revenue Regions: Regions consistently generating high sales should be targeted for increased marketing investment, stock availability, and premium delivery options.

- Reassess Product Mix: If particular categories dominate revenue, the business may benefit from expanding product lines in those categories.

- Analyze Payment Preferences: Understanding the payment types associated with high-value orders can guide promotional strategies or checkout improvements.

These recommendations will be refined once hypotheses are fully validated.

o **Analysis Techniques Used in Excel:** Excel served as the primary tool for exploring and validating relationships in the dataset. Key functionalities included:

- Pivot Tables: Used extensively to summarize revenue by region, product category, salesperson, and customer segment.

- Pivot Charts: Visual comparisons enabled quick identification of top-performing categories, sales trends over time, and variation in shipping fees

- Sorting and Filtering: Helpful in spotting high-revenue transactions, delayed shipments, and customers with frequent purchases.

- Conditional Formatting: Used to highlight orders with unusually high shipping fees or low revenue contribution.

- Charts and Graphs: Used to visualize revenue trends, shipping fee distribution, and regional performance.

7. Post-Analysis and Insights

This stage consolidates the findings from the full analytical process, confirming or rejecting the initial hypotheses, interpreting the patterns observed within the dataset, and providing actionable insights into sales performance, shipping behavior, and customer trends.

o **Key Findings:** Based on the completed analysis, several significant insights emerged:

- Top-performing and low-performing salespersons show a significant disparity in revenue contribution, highlighting the impact of individual performance on overall sales.

- Company D is the top-performing customer, while Company K contributes the least, highlighting significant differences in revenue generation across customers.

- The North region leads in revenue performance, while the West region ranks lowest, showing variation in regional sales contributions.

- New York is the most profitable state, while Nevada ranks lowest, highlighting a significant revenue gap of over $50,000 between top and bottom states.

- New York leads as the top-performing city, significantly outpacing others, with a revenue gap of over $26,000 compared to mid-tier cities like Chicago.

- A small number of top shipping cities led by New York, Portland, and Miami drive the majority of revenue, while most other cities contribute significantly less, highlighting the concentration of sales in key locations.

- Most transactions fall within the lower revenue ranges, with a clear decline in frequency as transaction amounts increase, indicating that smaller-value sales dominate overall activity.

- Beverages led as the top-performing product category, while Canned Meat ranks lowest, revealing a substantial revenue gap of over $85,000 between the highest and lowest categories.

- Sales peak in December and Q4 overall, while February is the lowest-performing month, revealing strong seasonal trends with notable mid-year and late-year declines.

o **Comparison with Initial Findings**

Comparing the validated results to initial expectations provides both alignment and unexpected discoveries:

- Validated Insights: Regions initially identified as strong revenue contributors remained top performers after full analysis. Product categories with higher prices indeed produced the most revenue, confirming early assumptions. Salespersons suspected to be top performers were validated through their revenue contribution and transaction frequency.

- Surprising or Counter-Intuitive Results: Some regions with relatively few transactions still generated high revenue due to premium product purchases. A few lower-priced categories produced unexpected revenue surges due to high purchase frequency. Some shippers that handled lower volumes had faster delivery speeds than high-volume shippers, suggesting operational efficiency rather than workload determines shipping performance.

- Insights: The final analysis confirms that revenue performance is shaped by a combination of product category, region, salesperson activity, and shipping dynamics. Targeted marketing in high-performing regions, better support for underperforming territories, closer evaluation of shipping partners, and nurturing repeat customers can significantly enhance future sales growth.

8. Data Visualizations & Charts

Visual representations were created in Microsoft Excel to simplify complex data relationships and highlight key patterns related to the sales performance. A summary of the visualizations and their interpretations is presented below.

o **Visuals**

- **Sales Performance Analysis by Reps (Bar Chart)**

A Bar chart was used to show the sales performance of the company's salesperson.

Insight: Nancy Freehafer leads in total sales and therefore has the highest revenue contributions, while Jan Kotas generates the lowest, making them key indicators of overall sales performance, with other salespersons contributing moderately in between.

- **Sales Performance Analysis by Region (Pie Chart)**

This chart visualizes the sales performance across regions.

Insight: The North region leads in revenue performance, while the West region lags, making these regions key indicators of overall regional sales performance.

- **Top Five Customers (Bar Chart)**

This visualization categorizes the top five companies' customers by sales performance.

Insight: Company D, followed by H, BB, F, and A, contribute the largest share of total revenue, making them the core drivers of customer performance. Meanwhile, Company K sits at the bottom of the customer list, highlighting a clear gap between the highest- and lowest-performing customers.

- **Sales Trend Report (Line Chart)**

Order date (month) was plotted against revenue.

Insight: Sales follow a clear seasonal pattern, with December delivering the strongest performance and February the weakest. Sharp declines occur immediately after major peak months like June and October, showing rapid shifts in demand. Overall, Q4 is the strongest quarter, contributing the largest share of total revenue.

- **Product Category Performance over Time (Bar Chart)**

This visualization shows the proportions of product categories across sales performance.

Insight: Beverages led the product categories by a wide margin, far outperforming all others and creating a substantial gap between the top and lowest performers. Sauces and Jams/Preserves form a strong mid-tier, while Dairy Products and Dried Fruit & Nuts contribute moderately. Canned Meat sits at the bottom, reinforcing a significant performance divide within the product portfolio.

- **Transaction by Amount (Column Chart)**

This visualization shows the proportions of grouped transaction amounts by the count of transactions of each order.

Insight: Lower-value transactions dominate the dataset, with the 0–1000 range recording the highest activity. As transaction values increase, the number of occurrences falls sharply, indicating that high-value transactions are significantly less frequent and contribute minimally to overall transaction volume.

- **Performance by State (Map Chart)**

This visualization shows the sales performance of each state.

Insight: New York is the top-performing state, while Nevada ranks lowest, revealing a revenue gap of over $50,000 and highlighting significant differences in state-level performance.

- **Top Five Cities by Revenue (Column Chart)**

This visualization shows the proportions of sales performance across cities.

Insight: New York stands out clearly as the city with the highest revenue, far ahead of all others. Portland and Miami follow behind as strong performers, while Memphis and Chicago fall into the mid-tier range. Overall, the revenue gap, especially the more than $26,000 difference between New York and Chicago, highlights New York's dominant sales performance.

- **Top 6 Ship Cities by Revenue (Bar Chart)**

This visualization shows the proportions of sales performance across ship cities.

Insight: New York leads as the most profitable shipping city, followed closely by Portland and Miami, which form the top-performing cluster. After these three, revenue drops sharply, with Milwaukee ranking lowest among the top six. The substantial gap between New York and Milwaukee underscores how strongly the top cities outperform the rest.

o **Prestige Group of Companies' Sales Report for the Year 2014 (Excel Dashboard)**

A consolidated dashboard was developed using slicers, pivot charts, and KPI indicators to allow interactive exploration of: Sales Performance Analysis by Reps, Sales Performance Analysis by Region,

Top Five Customers, Sales Trend Report, Product Category Performance over Time, Transaction by Amount, Performance by State, and Top Six Ship Cities by Revenue.

Dashboard Insights: The dashboard offers a clear, consolidated view of how products, customers, regions, and shipping patterns contribute to overall revenue performance. It allows stakeholders to quickly spot top-performing products, identify areas of decline, and understand the key factors influencing sales trends across the company.

9. Recommendations and Observations

This section translates the findings from the sales analysis into actionable recommendations that can help improve revenue growth, product performance, and market reach.

- **Actionable Insights**
  - Beverages dominate revenue performance, creating a large gap between the top category and the least-performing ones. Strengthen inventory planning, promotional campaigns, and product placement for high-performing categories while reassessing the strategy for weaker ones like Canned Meat.
  - The North region leads by a wide margin, while the West performs the lowest. Prioritize marketing efforts, distribution support, and customer engagement initiatives in weaker regions to balance overall performance.
  - New York is the strongest city by revenue, with a steep drop after the top three cities. Consider targeted expansion strategies in mid-tier and low-performing cities to reduce overdependence on a few locations.
  - Most transactions fall within the lower revenue brackets (0–1000), indicating many small purchases. Introduce bundling, volume discounts, or cross-selling strategies to increase average transaction value.
  - Sales performance fluctuates sharply across months, with peaks in June, October, and December and significant drops afterward. Develop seasonal sales plans, inventory forecasting, and promotional timing that align with these cyclical patterns.
  - A few customers contribute significantly more revenue than others, creating a dependency on top accounts. Strengthen relationships with high-value customers while identifying ways to grow revenue from low-performing accounts.

- o **Optimizations or Business Decisions**
  - Focus on low-performing regions, cities, and product categories when allocating marketing budgets, sales resources, and promotional intensity.
  - Use predictive sales planning to prepare for expected high-demand months and mitigate seasonal revenue dips.
  - Develop a customer segmentation strategy to tailor offers, pricing, and communication to high-, mid-, and low-value customer groups.
  - Leverage dashboards to monitor real-time sales metrics such as product trends, region performance, and monthly revenue swings to enable timely decision-making.
- o **Unexpected Outcomes**
  - The revenue gap between Beverages and the lowest product category is unusually large, suggesting consumers strongly prefer certain categories. This may indicate market saturation or insufficient visibility for lower-performing products.
  - Several mid-level states and cities show steady but not exceptional performance, revealing untapped growth potential that might be unlocked with better distribution, localized campaigns, or partnerships.
  - Sales spikes appear to follow a six-month cycle, which is an uncommon pattern. This may be driven by promotional schedules, festival seasons, or customer buying behaviors and should be further investigated.

10. Conclusion

This section summarizes the key findings, acknowledges limitations, and outlines directions for future analysis, providing closure to the report while highlighting actionable insights.

- o **Key Learnings**
  - Regional and city-level performance varies significantly, with the North region and New York City emerging as the strongest contributors, while the West region and several smaller cities lag behind.
  - Product categories show sharp contrasts, with Beverages generating exceptionally high revenue compared to others, revealing clear customer preferences.

- Customer performance is uneven, with a few top customers accounting for a substantial share of total revenue and many others contributing modestly.

- Monthly revenue trends display cyclical fluctuations, including predictable peaks in June, October, and December, followed by noticeable declines.

- Most transactions fall within lower revenue ranges, indicating that the business processes a high volume of small purchases rather than large orders.

- **Limitations**

  - Data completeness and granularity may restrict deeper exploration. Additional variables such as customer demographics, marketing efforts, or operational costs could provide a fuller picture of performance drivers.

  - Temporal scope may limit the ability to detect long-term trends. A multi-year dataset would strengthen forecasting and trend analysis.

  - Missing contextual data, such as promotions, product availability issues, or external economic factors, may influence sales patterns but are not captured in the dataset.

  - Data quality constraints, if present, may affect the accuracy of insights, particularly if records were incomplete, duplicated, or inconsistently formatted before cleaning.

- **Future Research**

  - Customer behavioral analysis, exploring purchase patterns, retention rates, and lifetime value to improve targeting and loyalty strategies.

  - Price sensitivity and demand modeling, to understand how pricing changes impact sales across different product categories and regions.

  - Promotional impact assessment, identifying how discounts, campaigns, and seasonal offers influence revenue trends.

  - Predictive analytics, such as forecasting monthly revenue, identifying at-risk customers, or predicting future top-performing products.

  - Regional market expansion analysis, especially for mid-tier cities and states that show growth potential.

11. References

This section provides supporting materials and resources for readers who want to explore the analysis in more detail. It ensures transparency, reproducibility and gives credit to sources and tools used.

- Dataset Source: Internal company sales dataset (cleaned, structured, and analyzed for this report).

- Tools Used: Microsoft Excel (Pivot Tables, Charts, Dashboard).

- External Resources:

  o Research on customer satisfaction and sales prioritization strategies.

  o Online tutorials and documentation for Excel analytical functions.

o **Appendices**

- Additional Charts and Tables: Detailed visualizations for salesperson, state, regional performance, customer purchase performance, and shipping location impacts.

- Raw Data Samples: Subsets of the dataset used for analysis.

- Formulas and Functions**:** Step-by-step explanations of key Excel formulas used, including:

  Pivot Table setups for aggregating sales performance by region, salesperson, state, and product category.

- Dashboard Screenshots: Illustrations of the consolidated dashboard displaying KPIs and critical metrics for stakeholders.