

March, 2025

Nestle Sales Data Insights



Project Report

Author:
Muge Yalcin

Contents

Table of contents:

Table of contents:	2
Table of Figures:.....	4
Abstract:	7
Section 1: Nestle Sales Dataset: Strategic Insights Through Exploratory Data Analysis	9
1.1 Introduction: Unveiling Strategic Narratives Through Data	9
1.2 Data Acquisition and Initial Assessment: Building a Solid Analytical Foundation..	9
1.3 Data Refinement: Ensuring Analytical Integrity	9
1.4 Key Numerical Insights: Unveiling Core Trends Through Summary Statistics	10
1.5 Categorical Insights: Mapping Market Dimensions.....	10
1.6 Revenue Distribution: Understanding Skewed Patterns and Their Implications ..	11
1.7 Temporal Sales Trends: Navigating Time-Based Dynamics	11
1.8 Weekday Sales Dynamics: Capturing Consumer Rhythms	13
1.9 Product Performance: Identifying Market Leaders and Opportunities	14
1.10 Geographical and Channel Dynamics: Mapping Regional and Digital Landscapes	16
1.11 Sales Medium Analysis: The Online Advantage	17
1.12 Strategic Recommendations and Conclusion.....	18
Section 2: Unveiling Customer Behavior and Sales Dynamics in Nestle Sales Data	19
2.1 Project Objectives and Methodology.....	19
2.1.1 Objectives	19
2.1.2 Methodology	19
2.2 Data Preparation and Transformation	19
2.3 Initial Market Basket Analysis: Challenges and Findings	20
2.4 Refined Analysis: Understanding Transaction Patterns	20
2.5 Product Performance Analysis	20

2.6	Sales Channel Effectiveness	22
2.7	Key Findings and Business Implications	23
2.8	Strategic Recommendations	23
2.9	Conclusion	24
Section 3: Forecasting Nestle Sales Through Time Series Analysis		24
3.1	Introduction: Forecasting Future Sales for Strategic Planning	24
3.2	Data Loading and Preprocessing: Setting the Stage for Time Series Analysis	24
3.3	Visualizing Daily Sales: Unveiling Historical Trends	25
3.4	Forecasting with ARIMA: Capturing Baseline Trends.....	25
3.5	Forecasting with Prophet: Handling Seasonality and Outliers	26
3.6	Evaluating Prophet Model Performance: Assessing Forecasting Accuracy	27
3.7	Seasonal Decomposition: Understanding Underlying Patterns	27
3.8	Key Insights and Strategic Recommendations: Translating Forecasts into Action	28
3.9	Conclusion: Leveraging Time Series Analysis for Informed Decision-Making.....	28
Section 4: Optimizing Sales Performance and Profitability Through Multifaceted Analysis		29
4.1	Introduction: Unlocking Sales Potential Through Profitability, Medium, and Geospatial Analysis	29
4.2	Data Loading and Basic Information: Establishing the Analytical Foundation.....	29
4.3	Profitability Analysis by Product: Identifying Key Revenue Drivers	29
4.4	Sales Medium Analysis: Evaluating Channel Efficiency	31
4.5	Geospatial Sales Analysis: Mapping Revenue Hotspots and Opportunities.....	31
4.6	Key Findings and Strategic Implications: Translating Insights into Action	33
4.7	Conclusion: Leveraging Data-Driven Insights for Enhanced Sales Performance .	33
Section 5: Enhancing Sales Strategy Through Segmentation and Clustering Analysis		33
5.1	Introduction: Leveraging Clustering for Enhanced Sales Strategy	34
5.2	Data Preparation: Laying the Foundation for Segmentation	34
5.3	Sales Location Segmentation Using K-Means: Identifying Geographic Performance Clusters	34
5.4	Sales Medium Segmentation Using K-Means: Categorizing Channel Performance	

5.5	Sales Medium Segmentation Using DBSCAN: Detecting Outliers and Anomalies	35
5.6	Key Findings and Strategic Implications: Translating Clusters into Actionable Insights	36
5.7	Conclusion: Driving Targeted Strategies Through Data-Driven Segmentation	37
Section 6: Project Summary		37
Appendix		39
Appendix A: Glossary		39
Appendix B: GitHub Repository		41
Reference		42

Table of Figures:

Figure 1: Table displaying summary statistics	10
Figure 2: Histogram chart displaying revenue distribution	11
Figure 3: Chart displaying total sales by weekday	13
Figure 4: Chart displaying top 10 products by revenue	14
Figure 5: Chart displaying top 10 products by sales count	15
Figure 6: Chart displaying top 10 products by revenue per product	15
Figure 7: Chart displaying total sales by location	16
Figure 8: Chart displaying total sales by location broken by sales medium	17
Figure 9: Chart displaying total revenue by sales medium	17
Figure 10: Chart displaying sales count vs revenue by sales medium	18
Figure 11: Interactive chart displaying 3-year sales trend for each product	20
Figure 12: Interactive chart displaying top performing products by revenue	21
Figure 13: Interactive chart displaying product performance by sales count	21
Figure 14: Interactive chart displaying sales volume and revenue distribution per product	21
Figure 15: Interactive charts showing sales channel performance by total sales and revenue generated per product	22
Figure 16: Interactive chart displaying - year sales trend by sales channel	22
Figure 17: Interactive chart displaying sales trends by sales channel	23
Figure 18: Chart displaying daily sales trend	25
Figure 19: Chart displaying sales forecast with ARIMA	26
Figure 20: Chart displaying sales forecast with Prophet	26
Figure 21: Chart displaying decomposition of time series of "Total Revenue" into its constituent components: Trend, Seasonal, and Residual (Resid)	27

Figure 22: Chart displaying the profitability per product.....	30
Figure 23: Chart displaying sales count and total revenue per product	30
Figure 24: Chart displaying sales count, total revenue, and revenue per sale across two sales mediums	31
Figure 25: Interactive map displaying revenue and sales figures for various Australian locations	32
Figure 26: Interactive chart displaying location performance	32
Figure 27: Chart displaying the results of K-Means clustering applied to sales location data	34
Figure 28: Chart displaying the results of K-Means clustering applied to sales medium data	35
Figure 29: Chart displaying the results of DBSCAN clustering applied to sales medium data	36
Figure 30: Partial View of the interactive dashboard developed using Tableau	37

This page intentionally left blank

Abstract

Abstract:

This project undertook a comprehensive analysis of Nestle's sales data to provide actionable insights into strategic enhancement. Initially, Market Basket Analysis (MBA) was employed to identify product associations, but the prevalence of single-item transactions necessitated a shift to individual product and sales channel performance evaluation. Time series forecasting, utilizing ARIMA and Prophet models, was subsequently conducted to project future sales trends, revealing Prophet's superior accuracy in capturing seasonality and outliers. Segmentation analysis, using K-Means and DBSCAN, categorized sales locations and mediums based on performance metrics, identifying key clusters for targeted strategies.

The analysis revealed significant disparities in sales performance across locations and mediums, with direct sales demonstrating higher effectiveness compared to online channels. An interactive map was developed to visualize sales performance by location, facilitating targeted marketing and resource allocation. Key findings underscore the importance of adaptable analytical approaches and the efficacy of Prophet for sales forecasting.

Recommendations include prioritizing strategies to enhance online sales, leveraging Prophet for ongoing forecasting, and tailoring regional marketing efforts based on segmentation analysis. This project demonstrates the value of data-driven strategies in optimizing sales performance and informing strategic decision-making within Nestle. The findings provide a robust foundation for informed decision-making, emphasizing the importance of data-driven strategies for enhancing Nestle's market performance.

This page intentionally left blank

Report

Section 1: Nestle Sales Dataset: Strategic Insights Through Exploratory Data Analysis

This section presents a comprehensive exploratory data analysis (EDA) of the Nestle sales dataset, designed to extract actionable insights for strategic decision-making. Through a meticulous examination of sales patterns across temporal, product, geographical, and channel dimensions, we have identified key trends and opportunities. Our analysis highlights the significant impact of online sales, the strategic importance of key product lines and geographical regions, and the necessity of balancing sales volume with per-unit revenue to optimize profitability. These insights provide a robust foundation for informed strategic planning and operational enhancements.

1.1 Introduction: Unveiling Strategic Narratives Through Data

In today's data-driven business landscape, the ability to translate raw data into strategic narratives is crucial. This project aims to illuminate the underlying stories within the Nestle sales dataset, transforming it from a collection of figures into a dynamic narrative that informs strategic decision-making. By applying rigorous exploratory data analysis (EDA) techniques, we seek to uncover patterns and trends that can guide Nestle's strategic direction and drive sustainable growth.

1.2 Data Acquisition and Initial Assessment: Building a Solid Analytical Foundation

The analytical process commenced with the seamless loading of the Nestle sales dataset into a Pandas DataFrame using Python. Initial inspections, utilizing `df.head()`, `df.info()`, and `df.describe()`, revealed a clean dataset, free of missing values and encompassing a diverse range of data types. This robust foundation ensures the reliability and accuracy of subsequent analyses.

1.3 Data Refinement: Ensuring Analytical Integrity

To maintain the integrity of our analysis, we proceeded to refine the dataset by removing duplicate entries. This step is crucial for preventing skewed results and ensuring that all insights are based on unique, accurate data points.

1.4 Key Numerical Insights: Unveiling Core Trends Through Summary Statistics

Summary Statistics:			
	Total Revenue	Sales count	Product Count
count	18115.000000	18115.0	18115.0
mean	1701.339026	1.0	1.0
std	2893.481093	0.0	0.0
min	11.580000	1.0	1.0
25%	460.480000	1.0	1.0
50%	992.470000	1.0	1.0
75%	2156.000000	1.0	1.0
max	127968.750000	1.0	1.0

Figure 1: Table displaying summary statistics

A detailed examination of the numerical data through summary statistics revealed several critical insights:

- **Single-Product Transactions:** Consistent "Sales count" and "Product Count" values of 1 indicate that each transaction involved a single product.
- **Revenue Variability:** A wide range and high standard deviation in "Total Revenue" suggest significant revenue variability, highlighting potential high-value transactions.
- **Weekday Consistency:** A relatively even distribution of sales across weekdays indicates consistent demand patterns.
- **Potential Outliers:** High maximum revenue values warrant further investigation to understand their impact.

1.5 Categorical Insights: Mapping Market Dimensions

Analysis of the categorical variables provided valuable context:

- **Product Diversity:** Nestle’s product range spans various categories, from coffee to snacks and culinary items.
- **Geographical Reach:** Sales data covers major Australian regions, indicating a broad distribution network.
- **Sales Channels:** Sales are conducted through both online and direct channels, highlighting a dual customer engagement strategy.

These insights are essential for segmenting and analyzing sales data across different market dimensions.

1.6 Revenue Distribution: Understanding Skewed Patterns and Their Implications

Visualizing the distribution of "Total Revenue" revealed a significant positive skew, characterized by a concentration of low-revenue sales and a long tail of high-value transactions. This skewness necessitates tailored strategies to optimize both low- and high-value sales segments.

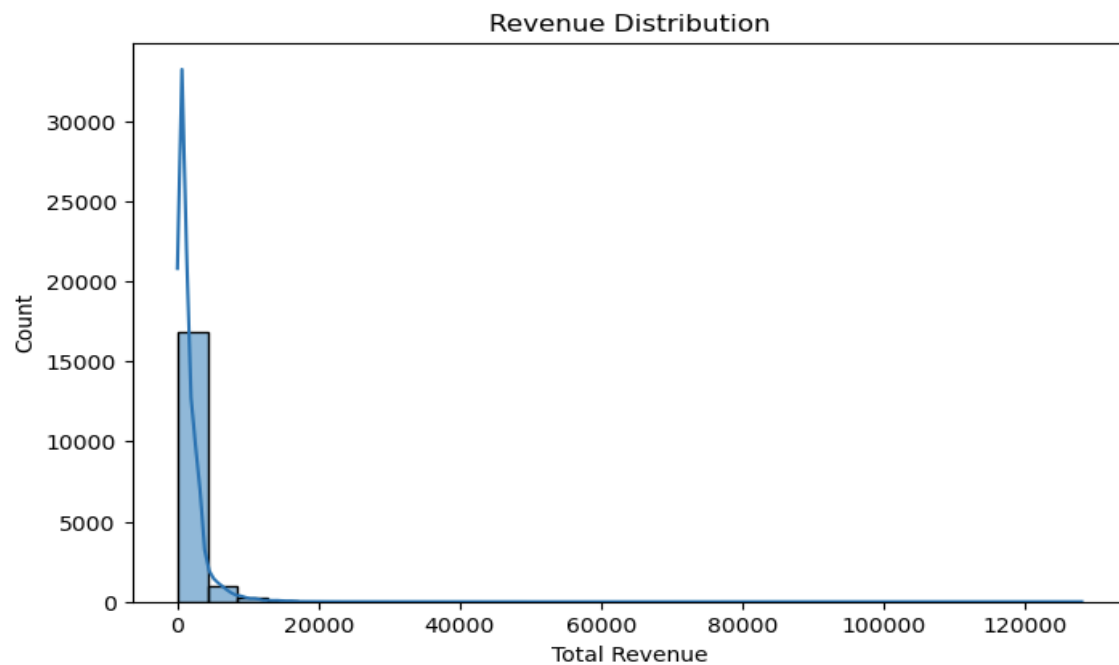


Figure 2: Histogram chart displaying revenue distribution

1.7 Temporal Sales Trends: Navigating Time-Based Dynamics

Analysis of sales trends over time, both monthly and daily, highlighted:

- **Monthly Trends:** Significant fluctuations in revenue, potentially indicative of seasonal or cyclical patterns.
- **Daily Trends:** High revenue volatility with sharp spikes, suggesting the impact of specific events or promotions.
- **Sales Count:** The sales count data in these sections did not provide helpful information and requires further investigation.

These trends underscore the importance of agile sales and inventory management.

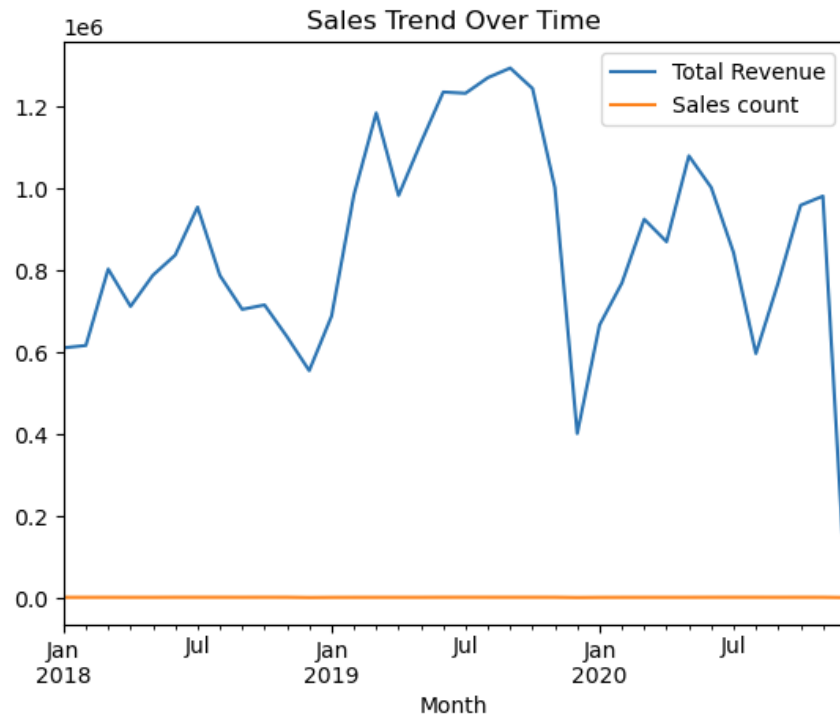


Figure 3: Chart displaying revenue and sales trend over time

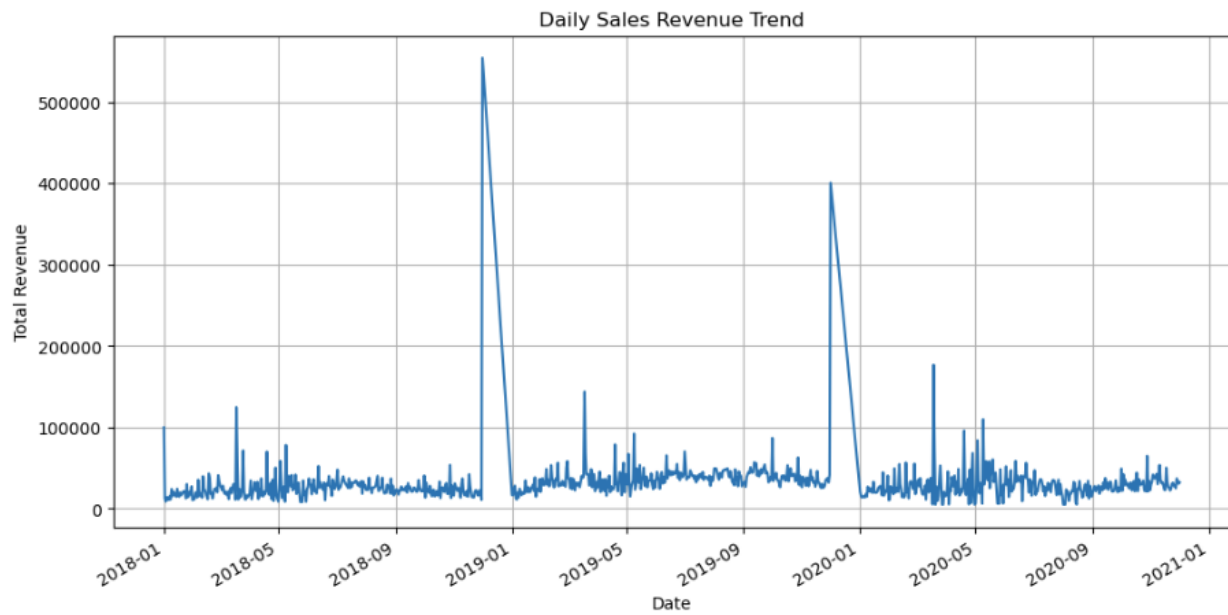


Figure 4: Chart displaying daily sales revenue trend

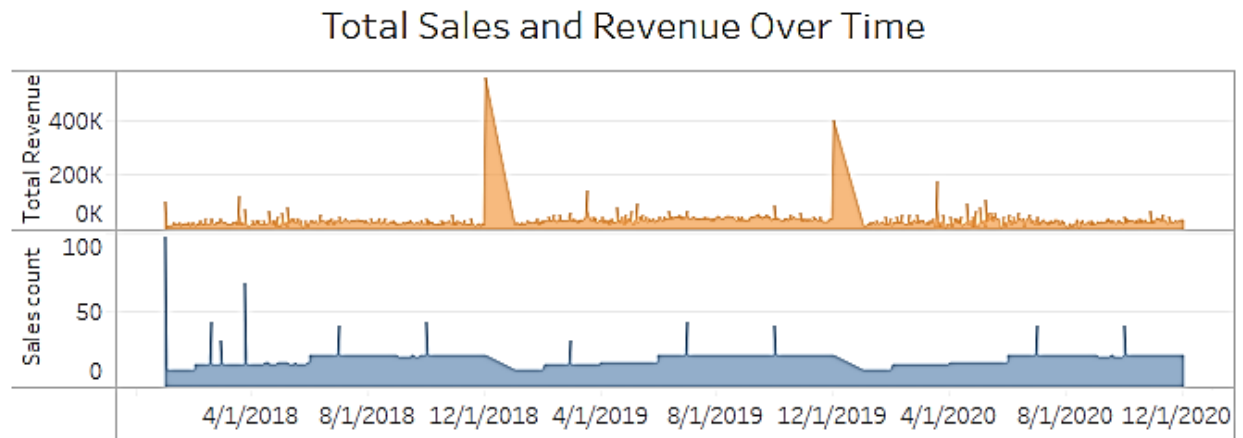


Figure 5: Interactive chart showing total sales and revenue trends

1.8 Weekday Sales Dynamics: Capturing Consumer Rhythms

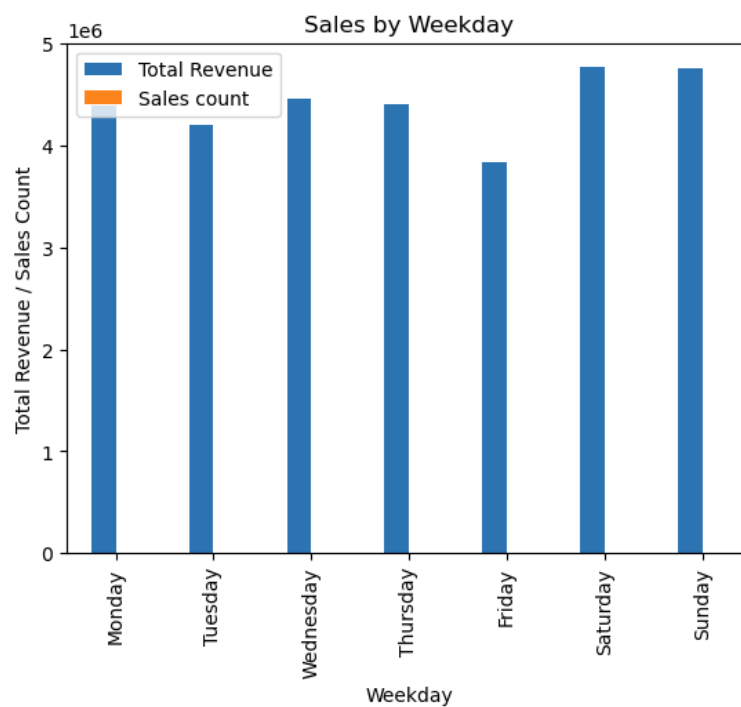


Figure 3: Chart displaying total sales by weekday

Analysis of weekday sales patterns revealed:

- Higher sales revenue on weekends (Saturday and Sunday).
- Lower sales revenue on Fridays.

This pattern suggests the need for targeted weekend promotions and strategies to improve Friday sales performance.

1.9 Product Performance: Identifying Market Leaders and Opportunities

A comparative analysis of the top 10 products based on revenue, sales count, and revenue per product highlighted:

- **Revenue Leaders:** "Milo" and "Nescafe" are key revenue generators.
- **Sales Volume Leader:** "Nescafe" leads in sales volume.
- **Revenue per Product Leader:** "Nesquik Duo" demonstrates high per-unit revenue.

Balancing volume and value is crucial for optimizing product portfolio strategy.

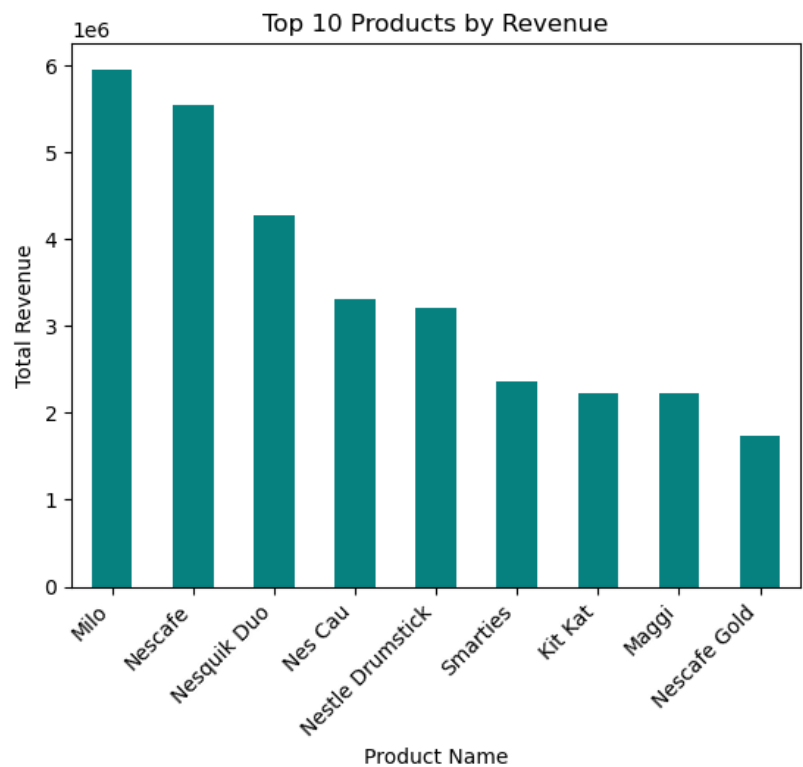


Figure 4: Chart displaying top 10 products by revenue

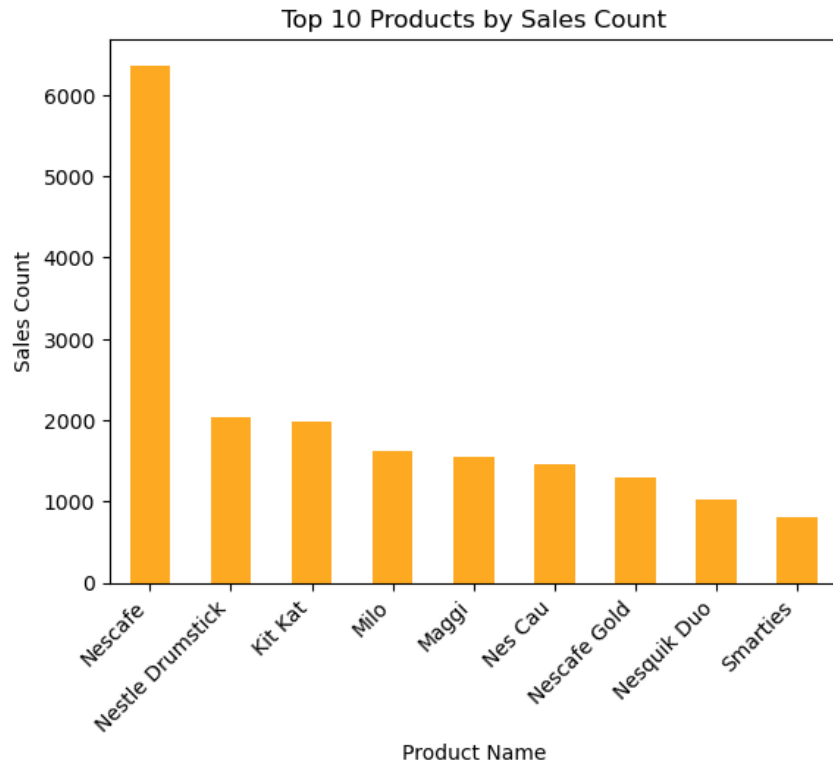


Figure 5: Chart displaying top 10 products by sales count

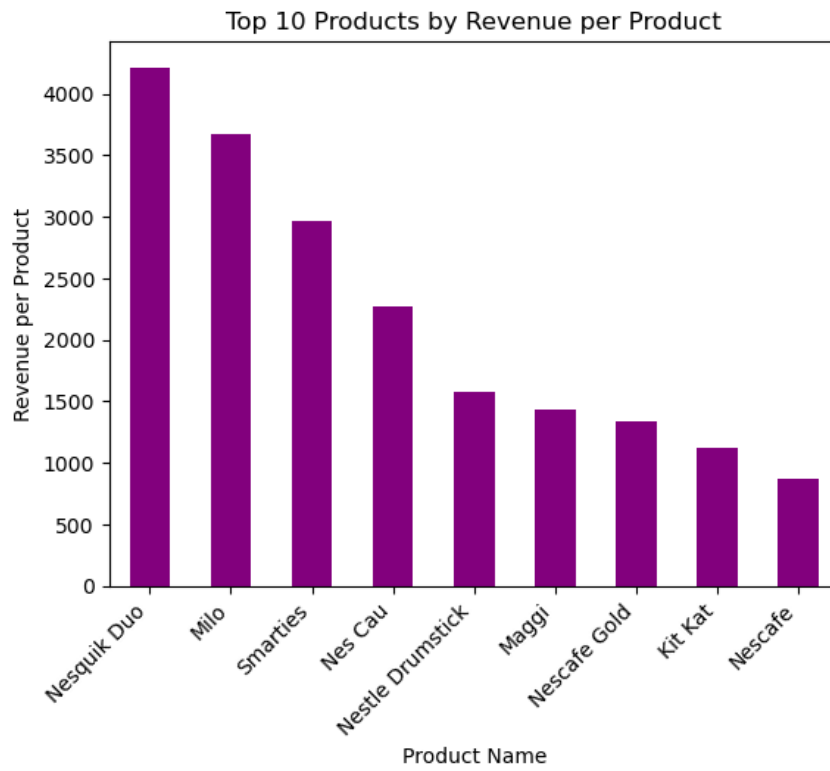


Figure 6: Chart displaying top 10 products by revenue per product

1.10 Geographical and Channel Dynamics: Mapping Regional and Digital Landscapes

Analysis of sales by location and sales medium revealed:

- **Regional Disparities:** South Australia and Tasmania lead in revenue, while Australian Capital Territory and Western Australia lag.
- **Online Dominance:** Online sales contribute significantly to revenue, particularly in high-performing regions.
- **Sales Count:** The sales count data in these sections did not provide helpful information and requires further investigation.

These findings highlight the need for tailored regional and digital sales strategies.

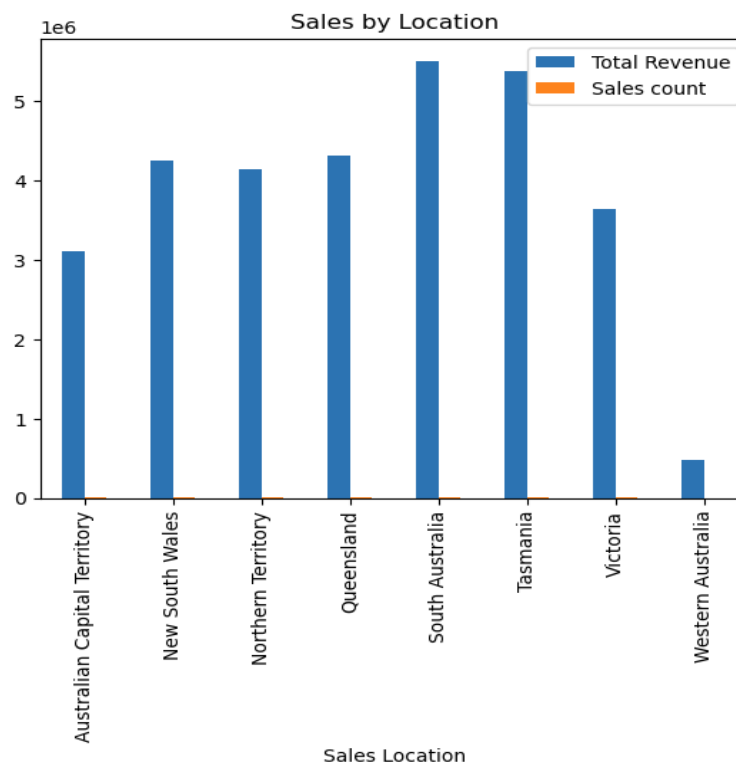


Figure 7: Chart displaying total sales by location

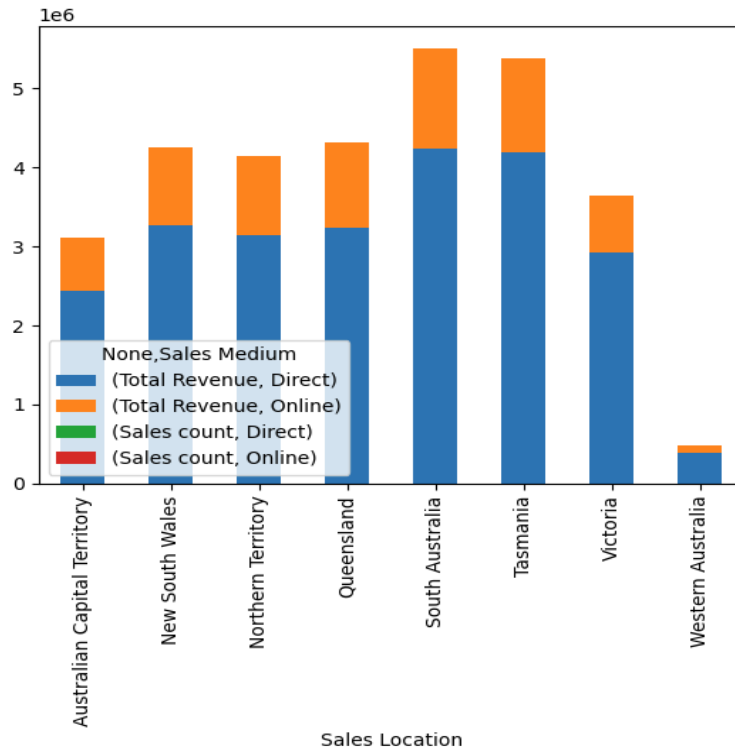


Figure 8: Chart displaying total sales by location broken by sales medium

1.11 Sales Medium Analysis: The Online Advantage

A focused analysis on sales by medium confirmed the significant impact of online sales, which outperformed direct sales. This underscores the importance of optimizing online channels.

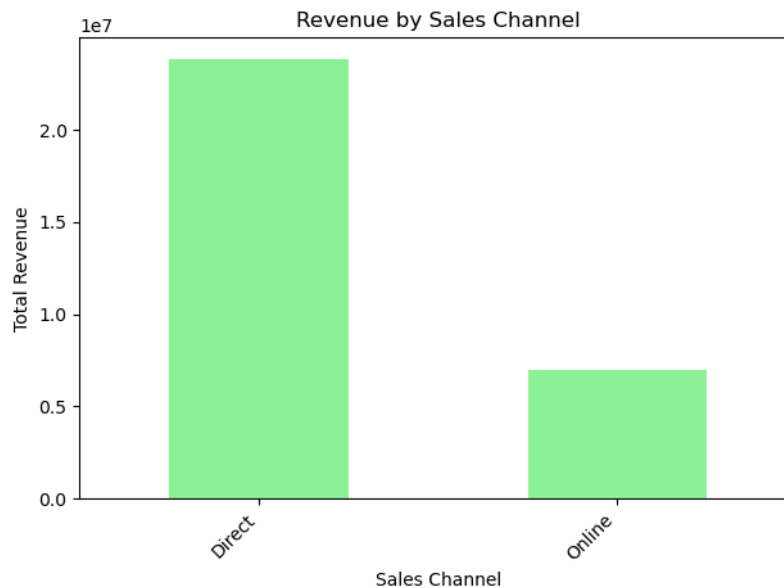


Figure 9: Chart displaying total revenue by sales medium

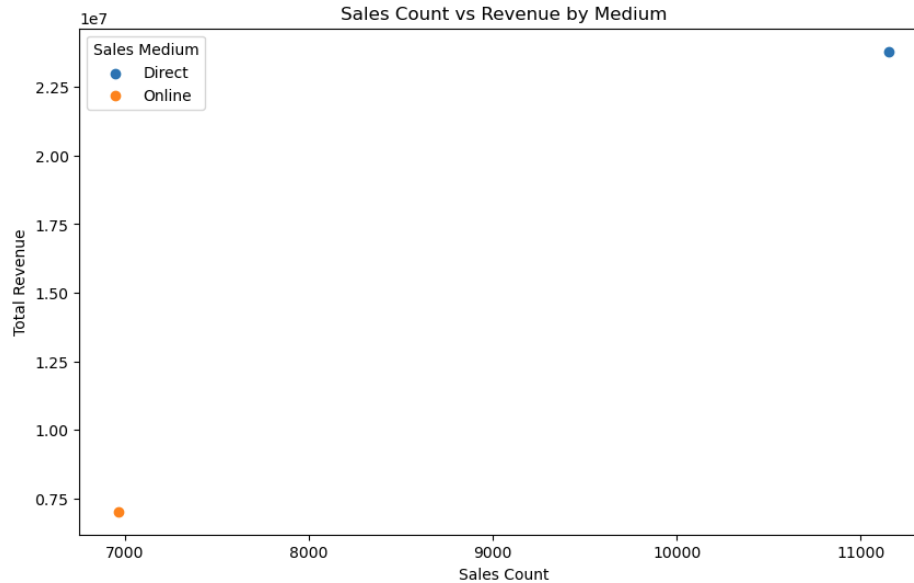


Figure 10: Chart displaying sales count vs revenue by sales medium

1.12 Strategic Recommendations and Conclusion

Based on the insights derived from this exploratory data analysis, we recommend the following strategic actions:

- **Optimize Online Sales Strategy:** Invest in digital marketing and e-commerce platforms to capitalize on the strong online sales performance.
- **Tailor Regional Strategies:** Develop targeted strategies to address regional disparities in sales performance.
- **Product Portfolio Optimization:** Balance high-volume and high-value products to maximize profitability.
- **Weekend and Friday Strategies:** Implement targeted promotions and strategies to maximize weekend sales and improve Friday sales performance.
- **Sales Count Data Investigation:** Investigate the sales count data to determine its accuracy and relevance.
- **Further Outlier Analysis:** Conduct further analysis on high-revenue outliers to understand their impact and potential causes.
- **Seasonality Analysis:** Explore potential seasonality in monthly sales trends to optimize inventory and marketing strategies.

This report demonstrates the power of data-driven insights in navigating market complexities and driving sustainable growth for Nestle. By implementing these recommendations, Nestle can enhance its market presence and optimize its operational strategies.

Section 2: Unveiling Customer Behavior and Sales Dynamics in Nestle Sales Data

This section aimed to uncover actionable insights from the Nestle sales dataset through Market Basket Analysis (MBA). However, a significant pattern of single-item transactions was identified, rendering traditional MBA techniques ineffective. Consequently, the analysis pivoted to focus on individual product performance and sales channel effectiveness. Key findings include the dominance of single-item purchases, the identification of top-performing products, and the significant revenue contribution of direct sales channels. These insights offer valuable strategic direction for Nestle.

2.1 Project Objectives and Methodology

2.1.1 Objectives

Our initial objective was to apply Market Basket Analysis (MBA) to uncover product associations and purchasing patterns. However, as the analysis progressed, it became evident that the data exhibited a unique characteristic: a predominance of single-item transactions. This revelation necessitated a shift in our analytical approach, leading us to explore alternative avenues for extracting valuable insights.

- Initially: Identify product associations and purchasing patterns through Market Basket Analysis (MBA).
- Adaptation: Analyze individual product performance and sales channel effectiveness.

2.1.2 Methodology

Our methodology encompassed data transformation, application of the Apriori algorithm, analysis of transaction details and product sales, and assessment of sales channel revenue.

- Data transformation into transactional format.
- Application of the Apriori algorithm.
- Analysis of transaction details and product sales.
- Assessment of sales channel revenue.

2.2 Data Preparation and Transformation

To facilitate the application of MBA, the Nestle sales dataset was transformed into a transaction-based format. This involved converting product purchases into a binary representation (1 for purchased, 0 for not purchased), a crucial step for the Apriori

algorithm. However, even at this stage, the sparsity of the data hinted at the single-item purchase pattern that would later be confirmed.

2.3 Initial Market Basket Analysis: Challenges and Findings

The initial application of the Apriori algorithm, with varying support and confidence thresholds, consistently resulted in empty DataFrames. This unexpected outcome indicated a fundamental lack of multi-product purchases within the dataset, challenging the applicability of traditional MBA and prompting a re-evaluation of our approach and the data itself.

2.4 Refined Analysis: Understanding Transaction Patterns

To gain a deeper understanding of the data, we conducted a detailed analysis of transaction sizes. This revealed a striking pattern: nearly every transaction involved the purchase of a single product. This finding explained the consistent failure of the Apriori algorithm and necessitated a shift in our analytical focus away from product associations and towards individual transaction analysis.

2.5 Product Performance Analysis

With MBA proving unsuitable, the analysis pivoted to focus on individual product performance.

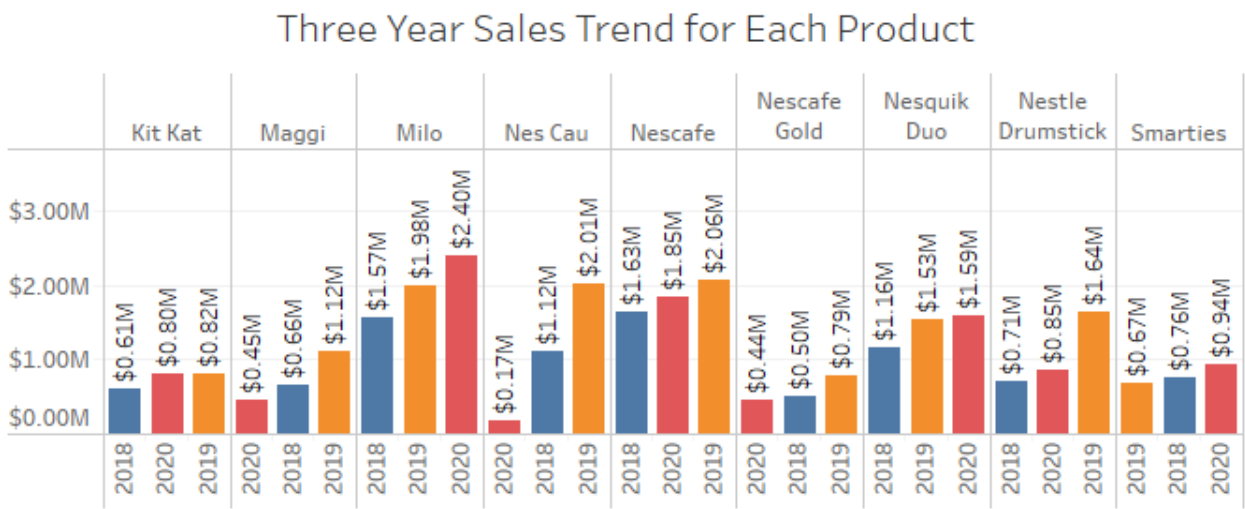


Figure 11: Interactive chart displaying 3-year sales trend for each product

Identifying top-selling products, such as Milo and Nescafe, based on revenue contribution provided valuable insights for marketing, sales, and inventory management strategies. This shift allowed us to extract actionable insights despite the initial analytical challenges.

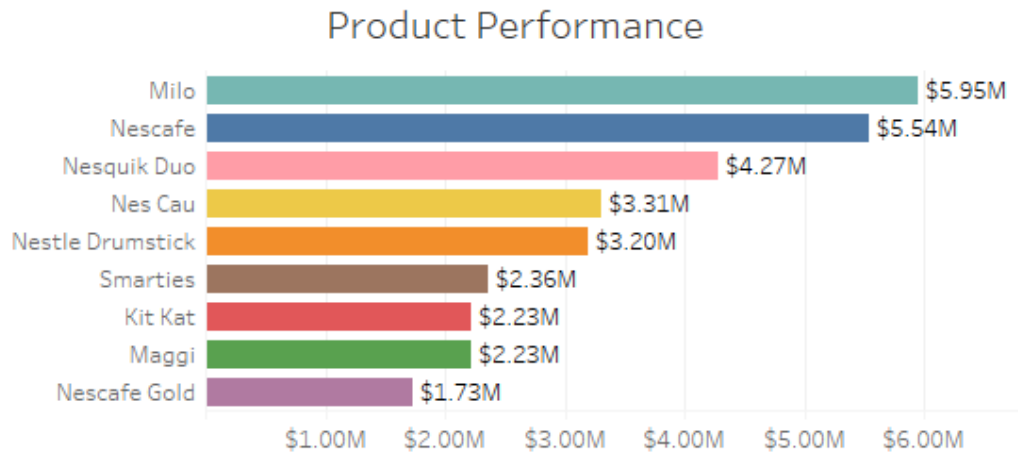


Figure 12: Interactive chart displaying top performing products by revenue

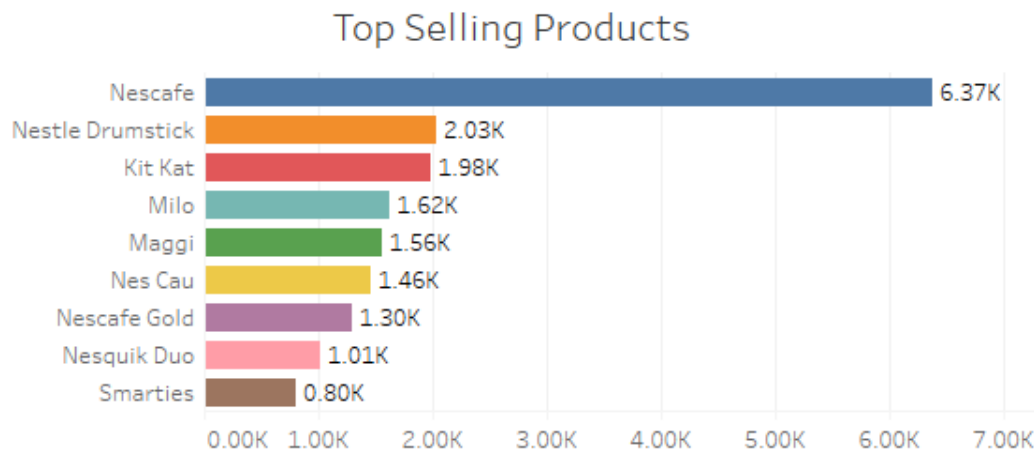


Figure 13: Interactive chart displaying product performance by sales count

Sales & Revenue Distribution by Product

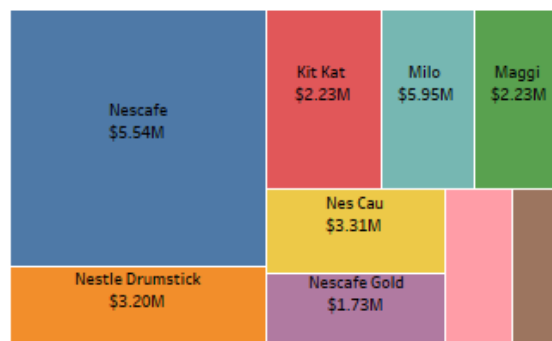


Figure 14: Interactive chart displaying sales volume and revenue distribution per product

2.6 Sales Channel Effectiveness

We further analyzed revenue by sales channel, revealing a significant disparity between direct and online sales. Direct sales emerged as the primary revenue driver, while online sales, though contributing, represented a smaller portion.

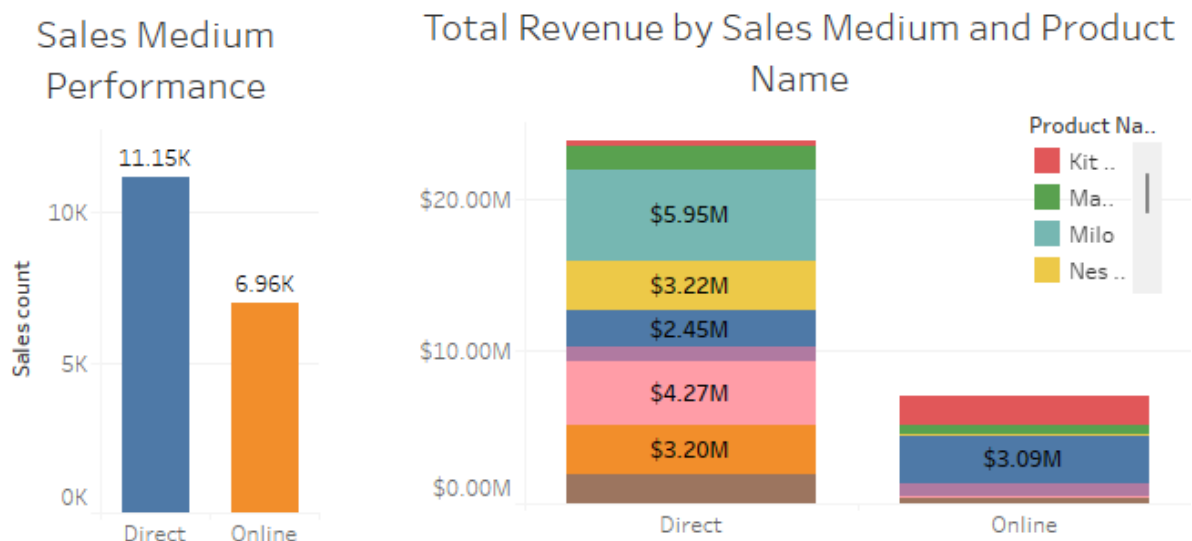


Figure 15: Interactive charts showing sales channel performance by total sales and revenue generated per product

This analysis highlighted the need for strategic initiatives to optimize online sales and understand the factors driving the success of direct sales.

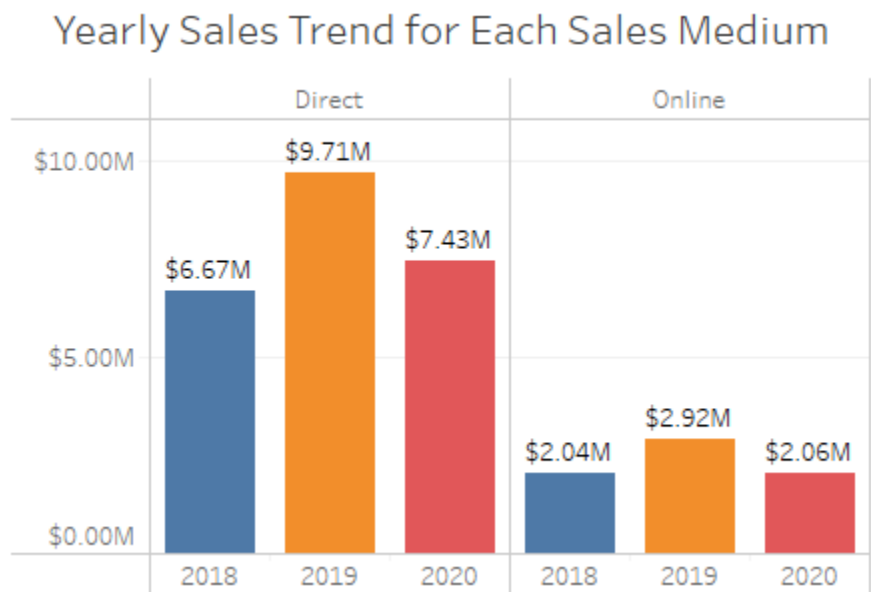


Figure 16: Interactive chart displaying - year sales trend by sales channel

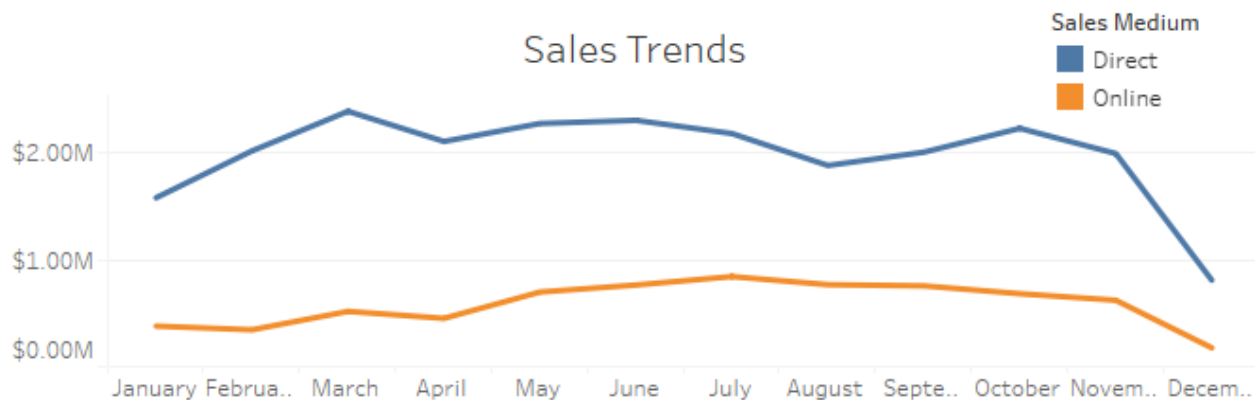


Figure 17: Interactive chart displaying sales trends by sales channel

2.7 Key Findings and Business Implications

Our analysis revealed several key findings:

- **Single-Item Purchases:** The Nestle sales data is characterized by single-item transactions, rendering traditional MBA ineffective.
- **Product Performance:** A few key products drive a significant portion of revenue.
- **Sales Channel Disparity:** Direct sales significantly outperform online sales.
- **Data Granularity:** Sales ID's likely represent single item sales, instead of full shopping baskets.

These findings carry significant business implications, requiring adjustments in marketing strategies, sales channel optimization, and data interpretation.

2.8 Strategic Recommendations

Based on our findings, we recommend the following strategic actions:

- **Focus on Individual Product Promotions:** Develop targeted marketing strategies for individual products.
- **Optimize Online Sales Channels:** Implement strategies to increase online sales revenue.
- **Analyze Direct Sales Success:** Investigate the factors driving direct sales success to replicate and scale.

- **Customer Segmentation:** Explore customer segmentation based on purchase frequency and other relevant factors.
- **Data Granularity Review:** Evaluate the granularity of sales IDs to determine if they represent individual items or full shopping baskets.

2.9 Conclusion

In conclusion, while the initial objective of applying MBA was not realized due to the prevalence of single-item transactions, this section yielded valuable insights into customer purchase behavior and sales channel effectiveness. The findings provide a solid foundation for strategic decision-making and highlight the importance of adaptability in data analysis. By focusing on individual product performance and sales channel optimization, Nestle can enhance its market presence and drive sustainable growth.

Section 3: Forecasting Nestle Sales Through Time Series Analysis

This section presents a comprehensive time series analysis of Nestle's sales data, aimed at forecasting future sales and understanding underlying patterns. We employed ARIMA, Prophet, and seasonal decomposition techniques to provide robust forecasts and insights into trend, seasonality, and residual components. The analysis revealed a sales pattern characterized by a stable baseline with occasional spikes, which were better captured by Prophet than ARIMA. The report concludes with strategic recommendations for leveraging these forecasts in business planning.

3.1 Introduction: Forecasting Future Sales for Strategic Planning

The ability to accurately forecast future sales is crucial for strategic planning, inventory management, and resource allocation. This project aims to provide Nestle with reliable sales forecasts using time series analysis. We explored various methodologies, including ARIMA, Prophet, and seasonal decomposition, to understand the underlying patterns in sales data and predict future trends.

3.2 Data Loading and Preprocessing: Setting the Stage for Time Series Analysis

Our analysis began with loading the sales data from an Excel file and converting the date column into a datetime format. This preprocessing step ensured the data was suitable for time series analysis, enabling us to effectively manipulate and analyze temporal trends.

3.3 Visualizing Daily Sales: Unveiling Historical Trends

Visualizing daily sales provided a crucial narrative of historical trends. We observed a pattern characterized by a stable, low-level baseline punctuated by occasional, sharp spikes. These spikes, notably occurring in early 2019 and 2020, hinted at potential promotional events, product launches, or external factors. Understanding these spikes is paramount for accurate forecasting and strategic planning.

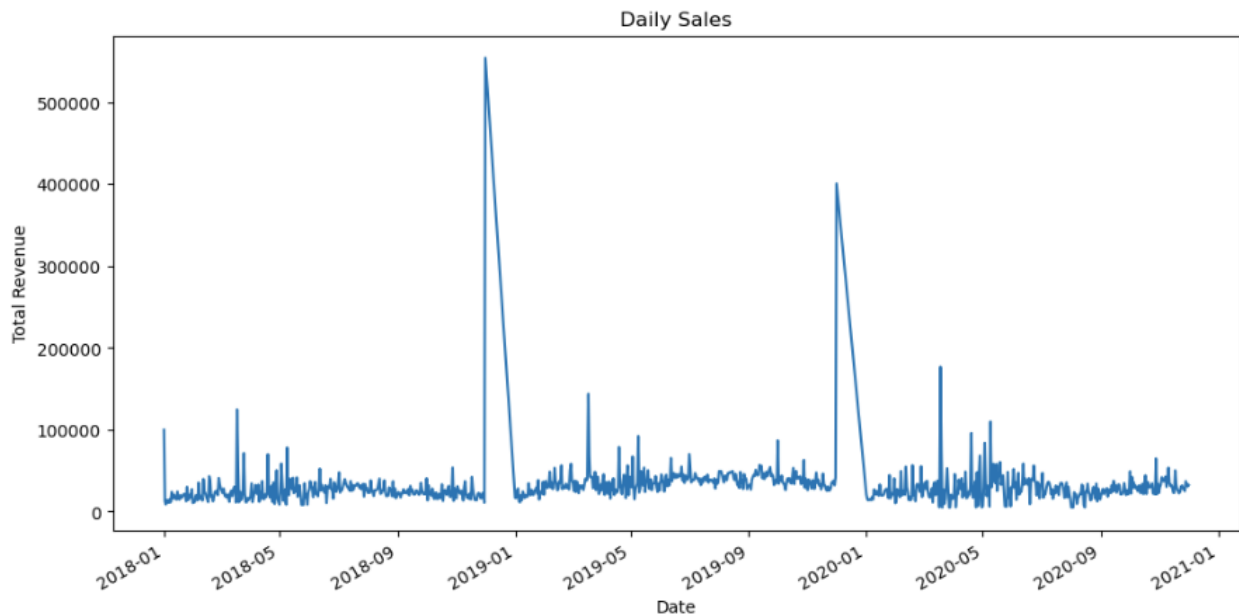


Figure 18: Chart displaying daily sales trend

3.4 Forecasting with ARIMA: Capturing Baseline Trends

We then applied the ARIMA model to forecast sales, which effectively captured the baseline trend but fell short in predicting the significant spikes. This highlighted the model's limitations in handling sudden, drastic changes in sales patterns. The ARIMA forecast, while stable, lacked the dynamic representation of the actual sales data.

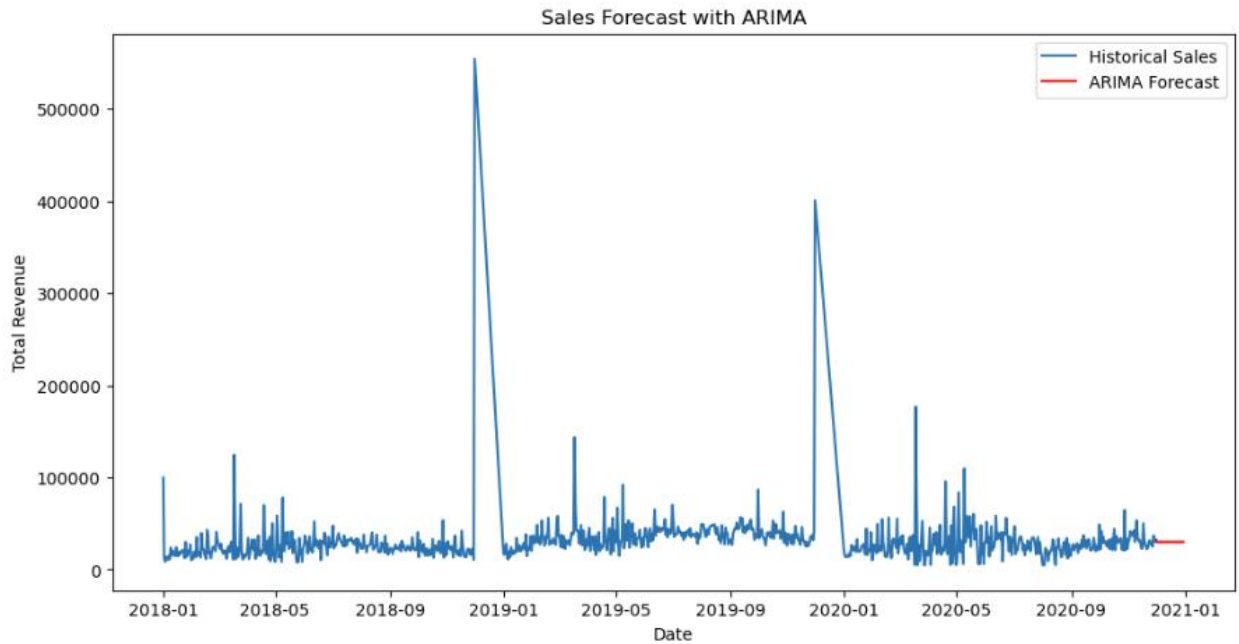


Figure 19: Chart displaying sales forecast with ARIMA

3.5 Forecasting with Prophet: Handling Seasonality and Outliers

In contrast to ARIMA, the Prophet model effectively captured the overall trend and seasonality of the sales data, handling outliers better.

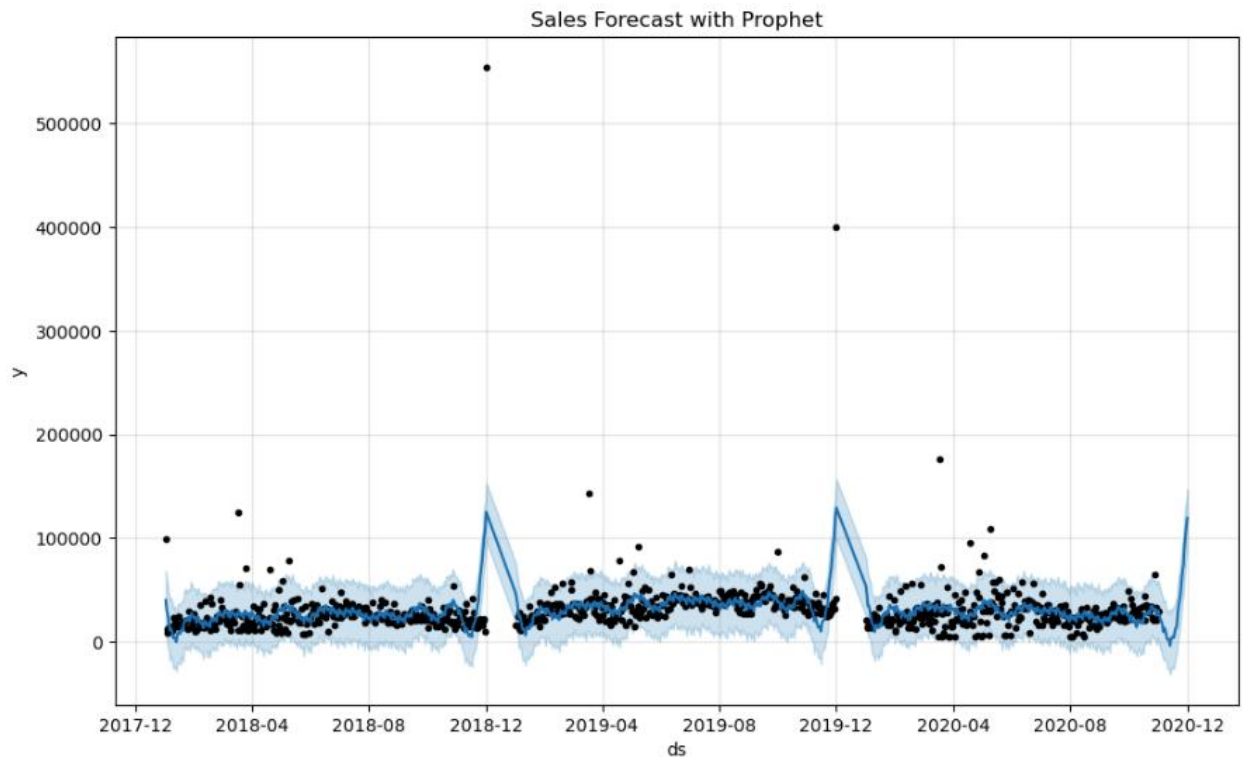


Figure 20: Chart displaying sales forecast with Prophet

Prophet's ability to widen confidence intervals around spikes reflected increased uncertainty, providing a more robust forecast. This model demonstrated superior performance in capturing complex patterns compared to ARIMA.

3.6 Evaluating Prophet Model Performance: Assessing Forecasting Accuracy

Evaluating the Prophet model's performance revealed noticeable errors, particularly in predicting spikes. However, the model's ability to capture the overall trend and seasonality made it more reliable than ARIMA. Further model tuning and feature engineering were recommended to improve accuracy.

3.7 Seasonal Decomposition: Understanding Underlying Patterns

Further dissecting the time series data, we applied seasonal decomposition to understand the underlying patterns. This revealed a stable trend with no discernible seasonality within the observed timeframe, indicating that the data's variations were largely random. This insight provided clarity on the nature of the sales data.

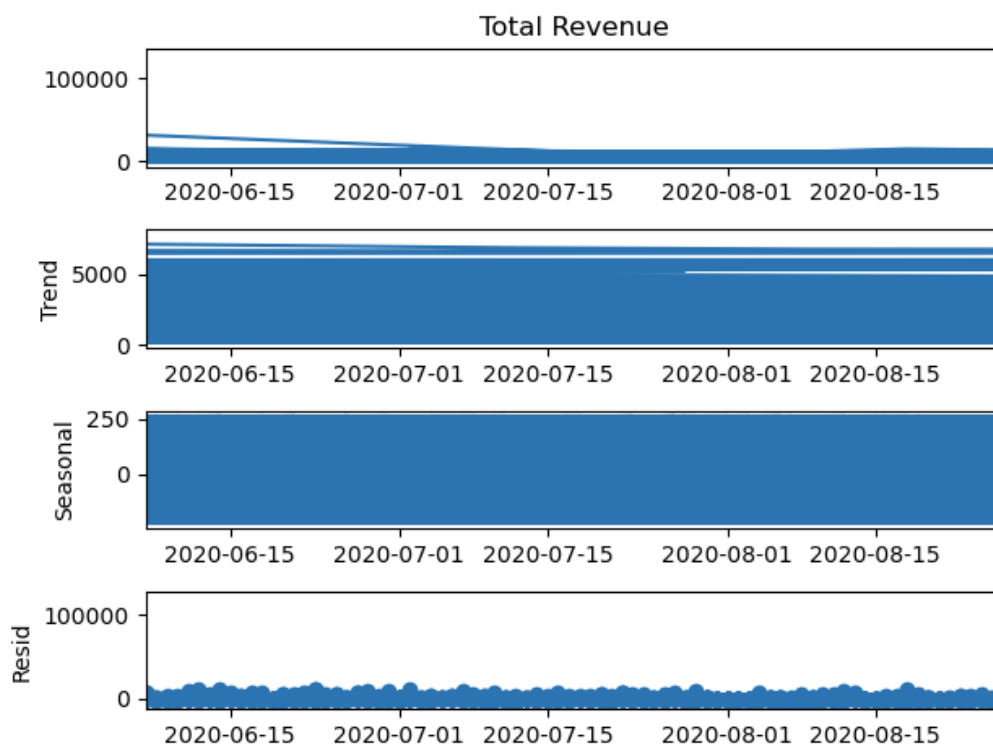


Figure 21: Chart displaying decomposition of time series of "Total Revenue" into its constituent components: Trend, Seasonal, and Residual (Resid)

3.8 Key Insights and Strategic Recommendations: Translating Forecasts into Action

Our analysis yielded several key insights:

- **Prophet's Superiority:** Prophet outperformed ARIMA in capturing complex sales patterns.
- **Spike Analysis:** Understanding the causes of sales spikes is crucial for accurate forecasting.
- **Stable Baseline:** The data exhibits a stable, low-level baseline with occasional spikes.
- **No Discernible Seasonality:** The observed timeframe showed no significant seasonality.

Based on these insights, we recommend the following strategic actions:

- **Utilize Prophet for Forecasting:** Employ Prophet for future sales forecasting due to its superior performance.
- **Investigate Spike Causes:** Conduct further analysis to understand the factors causing sales spikes.
- **Continuous Model Monitoring:** Regularly update and refine the Prophet model with new data.
- **Actionable Forecast Implementation:** Incorporate forecasts into inventory, staffing, and revenue planning.

3.9 Conclusion: Leveraging Time Series Analysis for Informed Decision-Making

This section demonstrated the power of time series analysis in forecasting Nestle's sales and understanding underlying patterns. By employing appropriate methodologies, such as Prophet, and interpreting the results within a business context, Nestle can make informed decisions to optimize operations and drive sustainable growth. The adaptability of our analytical approach, moving from ARIMA to Prophet based on data insights, showcases the importance of flexible data analysis in empowering decisions with predictive power.

Section 4: Optimizing Sales Performance and Profitability Through Multifaceted Analysis

This section presents a multifaceted analysis of Nestle's sales data, focusing on product profitability, sales medium performance, and geospatial distribution. Through rigorous analysis and visualization, we identified key revenue drivers, evaluated channel efficiency, and mapped sales hotspots. These insights provide actionable recommendations for optimizing sales strategies and resource allocation.

4.1 Introduction: Unlocking Sales Potential Through Profitability, Medium, and Geospatial Analysis

To enhance Nestle's sales performance and profitability, this project aimed to evaluate product profitability, sales medium effectiveness, and geographic revenue distribution. By providing a comprehensive overview of sales performance across these dimensions, we sought to deliver actionable insights for strategic decision-making.

4.2 Data Loading and Basic Information: Establishing the Analytical Foundation

Our analysis commenced with loading the sales dataset and understanding its structure. This initial step was crucial for identifying the necessary columns and preparing the data for subsequent analyses.

4.3 Profitability Analysis by Product: Identifying Key Revenue Drivers

We began by analyzing product profitability, calculating total sales, revenue, and revenue per sale for each product. This analysis revealed that Nesquik Duo had the highest profitability per sale, while Nescafe had the highest sales volume and total revenue. These findings highlight the importance of considering both profitability per sale and total revenue when evaluating product performance.

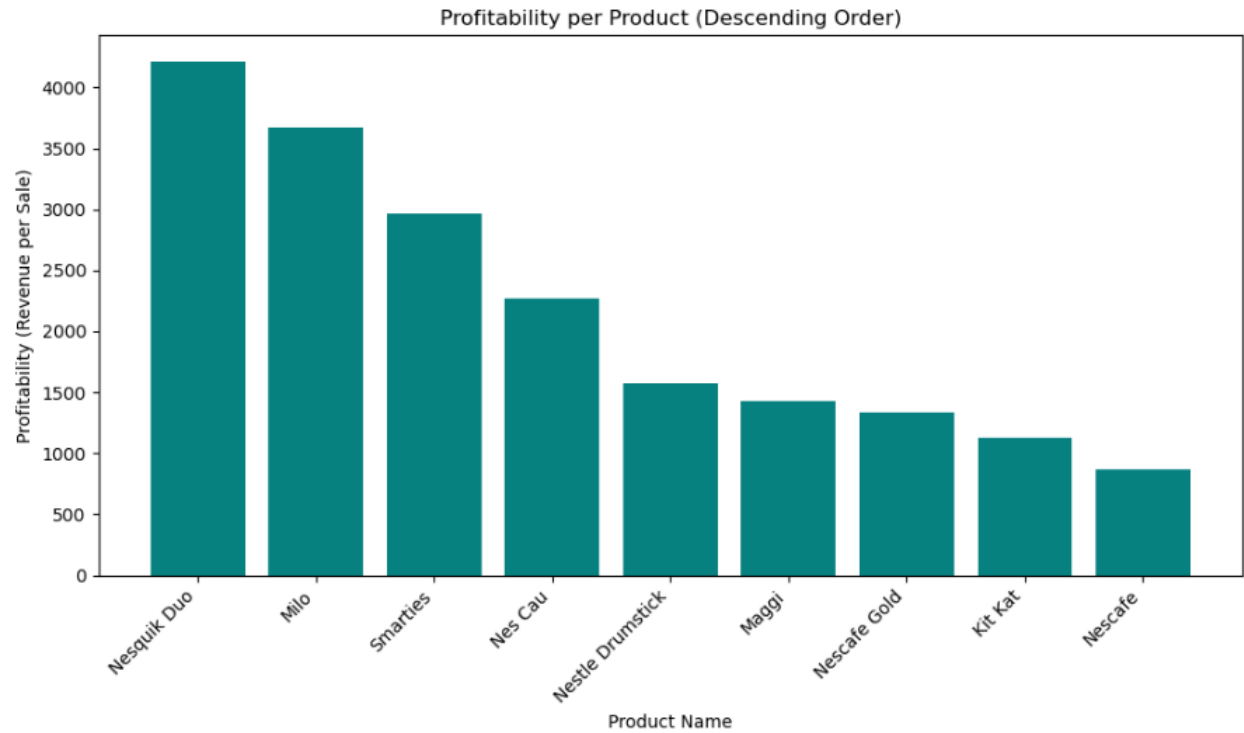


Figure 22: Chart displaying the profitability per product

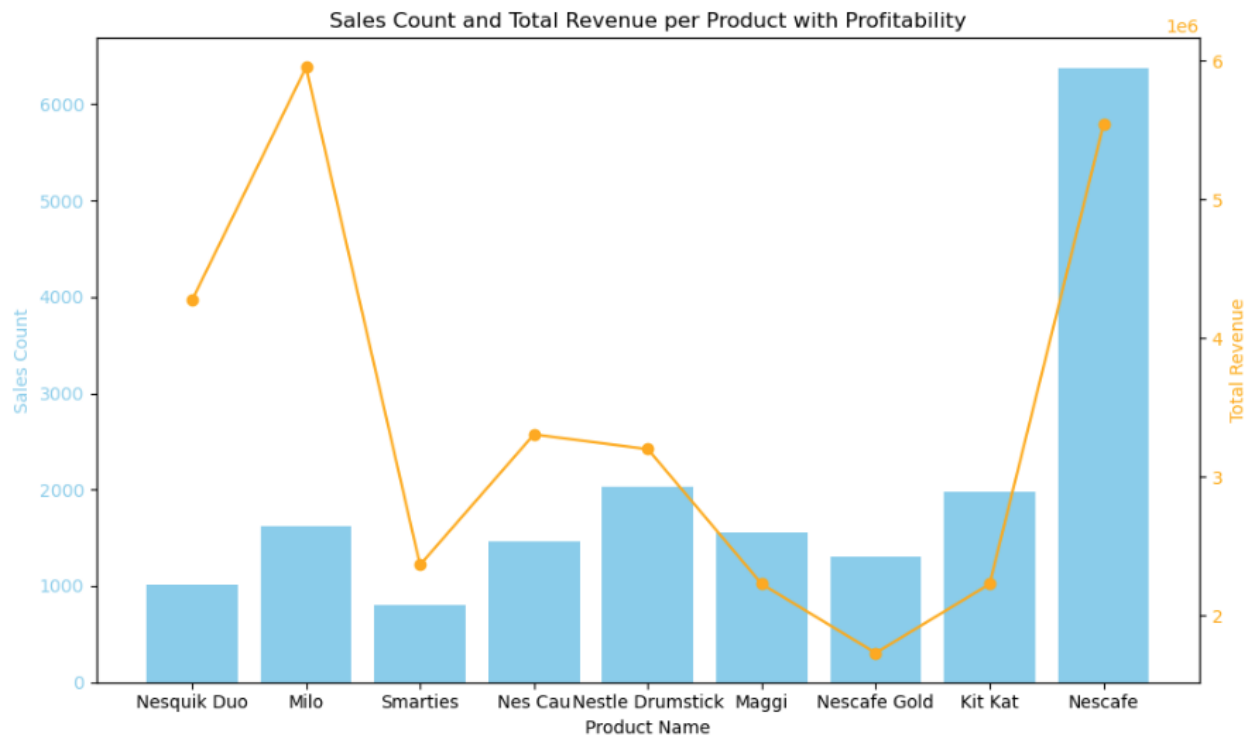


Figure 23: Chart displaying sales count and total revenue per product

4.4 Sales Medium Analysis: Evaluating Channel Efficiency

Next, we evaluated the performance of different sales mediums, focusing on direct and online sales channels. We calculated total sales, revenue, and revenue per sale for each medium. The analysis revealed that direct sales had a significantly higher sales count and revenue per sale, while total revenue was equal across both channels. This indicated that online sales, despite lower individual sale value, generated equivalent total revenue, likely through higher volume.

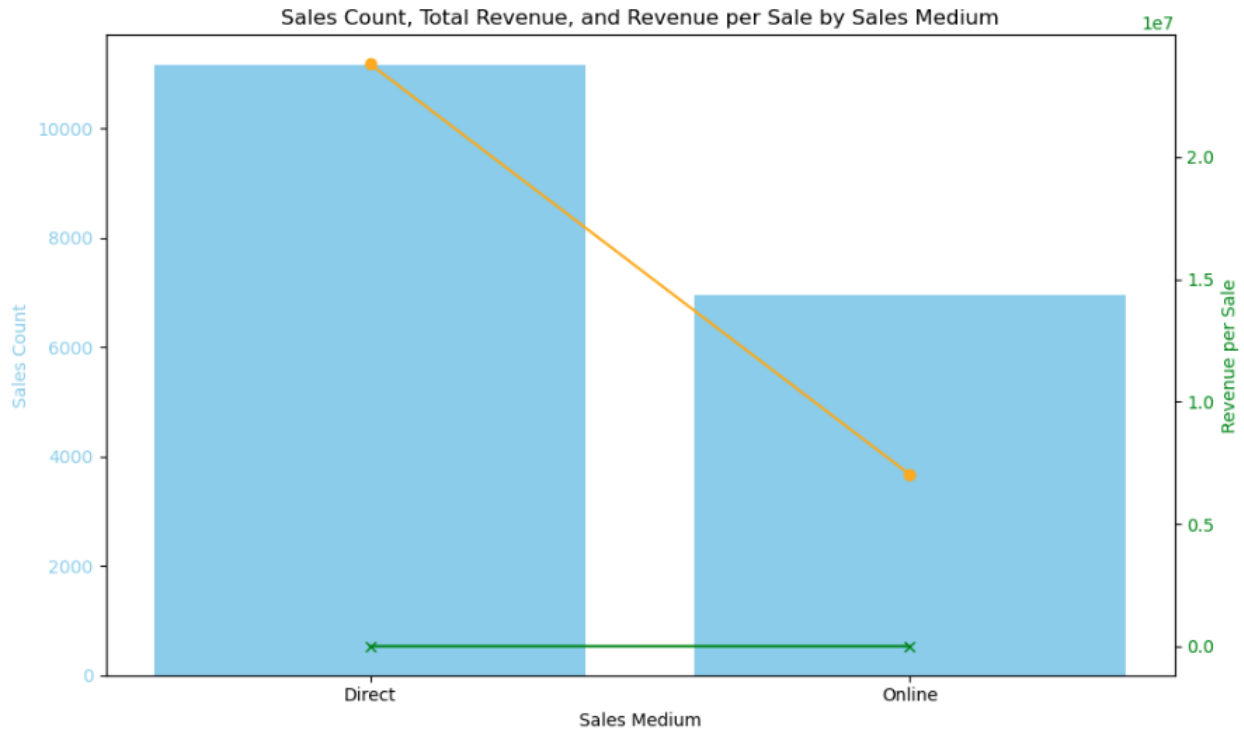


Figure 24: Chart displaying sales count, total revenue, and revenue per sale across two sales mediums

4.5 Geospatial Sales Analysis: Mapping Revenue Hotspots and Opportunities

To understand the geographic distribution of sales, we performed a geospatial analysis. By geocoding sales locations and visualizing revenue on a map, we identified sales hotspots and areas requiring attention. This analysis highlighted regions with high revenue and sales, enabling targeted marketing and sales strategies, logistical planning, and resource allocation.



Figure 25: Interactive map displaying revenue and sales figures for various Australian locations

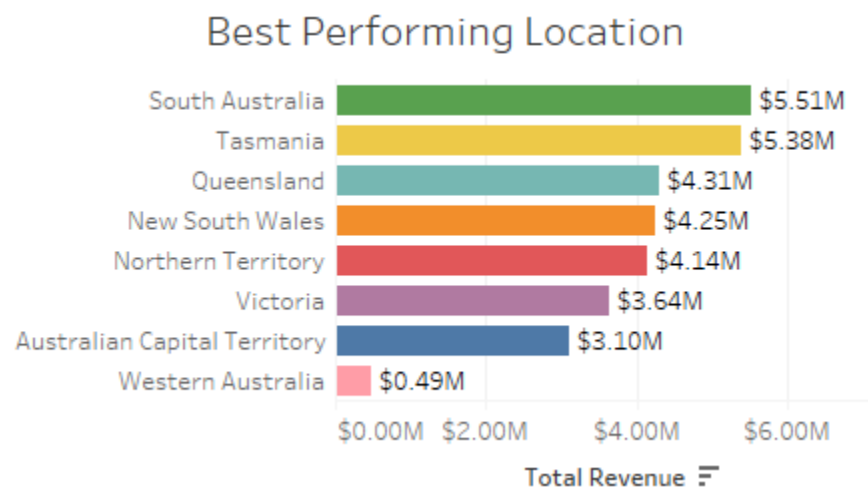


Figure 26: Interactive chart displaying location performance

4.6 Key Findings and Strategic Implications: Translating Insights into Action

Our analysis yielded several key findings:

- **Product Profitability:** Nesquik Duo showed the highest profitability per sale, while Nescafe drove the highest total revenue.
- **Sales Medium Efficiency:** Direct sales outperformed online sales in revenue per sale and sales count, but both channels contributed equally to total revenue.
- **Geospatial Distribution:** Visualizing sales locations revealed revenue hotspots and areas for strategic focus.

Based on these findings, we recommend:

- **Product Strategy Optimization:** Focus on high-profitability products like Nesquik Duo while strategically managing high-volume products like Nescafe.
- **Channel Optimization:** Investigate and optimize online sales strategies to increase revenue per sale or volume.
- **Geographic Targeting:** Tailor marketing and sales strategies to capitalize on high-performing regions and address underperforming areas.
- **Resource Allocation:** Allocate resources based on geographic revenue distribution to maximize returns.

4.7 Conclusion: Leveraging Data-Driven Insights for Enhanced Sales Performance

This comprehensive analysis provides actionable insights into Nestle's sales performance across product profitability, sales mediums, and geographic distribution. By leveraging these data-driven findings, Nestle can optimize sales strategies, enhance resource allocation, and drive sustainable growth. The integration of profitability, medium, and geospatial analysis demonstrates our commitment to providing holistic and strategic business intelligence.

Section 5: Enhancing Sales Strategy Through Segmentation and Clustering Analysis

This section presents a comprehensive clustering analysis of Nestle's sales data, focusing on segmenting sales locations and mediums using K-Means and DBSCAN. The analysis identified distinct performance clusters, revealing valuable insights for targeted strategies and resource optimization. Key findings include the identification of high and low-

performing geographic regions and sales channels, as well as the detection of significant disparities between direct and online sales.

5.1 Introduction: Leveraging Clustering for Enhanced Sales Strategy

The objective of this analysis is to enhance Nestle's sales strategy by segmenting sales data based on location and medium. Clustering techniques like K-Means and DBSCAN were employed to categorize sales data into meaningful clusters, enabling targeted sales forecasting, strategic decision-making, and resource optimization. This approach aims to identify patterns in revenue and sales counts, ultimately improving business performance.

5.2 Data Preparation: Laying the Foundation for Segmentation

Our analysis began with importing the necessary libraries and loading the sales dataset. The dataset, containing information on sales count, total revenue, sales medium, and location, was prepared for clustering analysis.

5.3 Sales Location Segmentation Using K-Means: Identifying Geographic Performance Clusters

We segmented sales locations using K-Means clustering, grouping locations based on sales count and total revenue. This analysis identified three distinct clusters: low-performing regions (Cluster 0), average-performing regions (Cluster 1), and high-performing regions (Cluster 2). The geographic distribution of these clusters highlighted areas requiring targeted strategies and resource allocation.

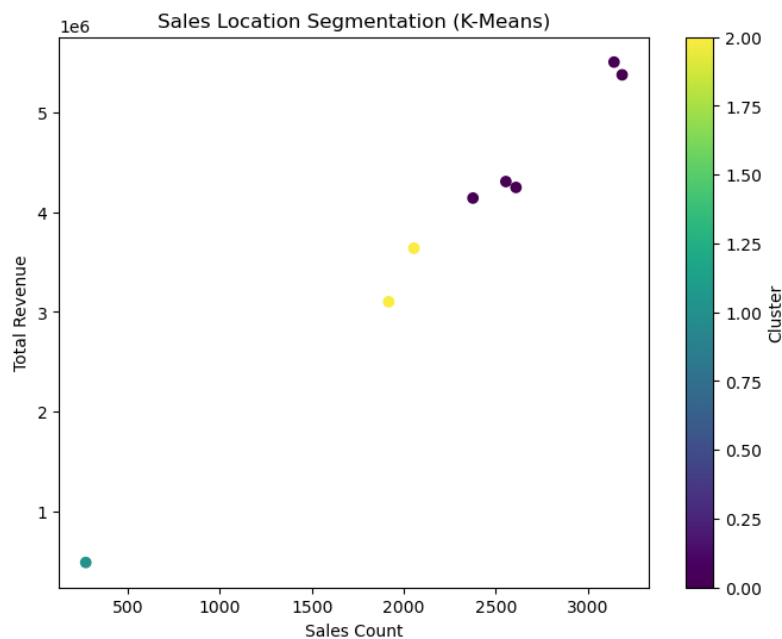


Figure 27: Chart displaying the results of K-Means clustering applied to sales location data

5.4 Sales Medium Segmentation Using K-Means: Categorizing Channel Performance

Next, we segmented sales mediums using K-Means, categorizing sales channels into two clusters: direct sales (Cluster 0) and online sales (Cluster 1). The analysis revealed that direct sales significantly outperformed online sales in terms of sales count and total revenue, indicating a need for strategies to enhance online sales performance.

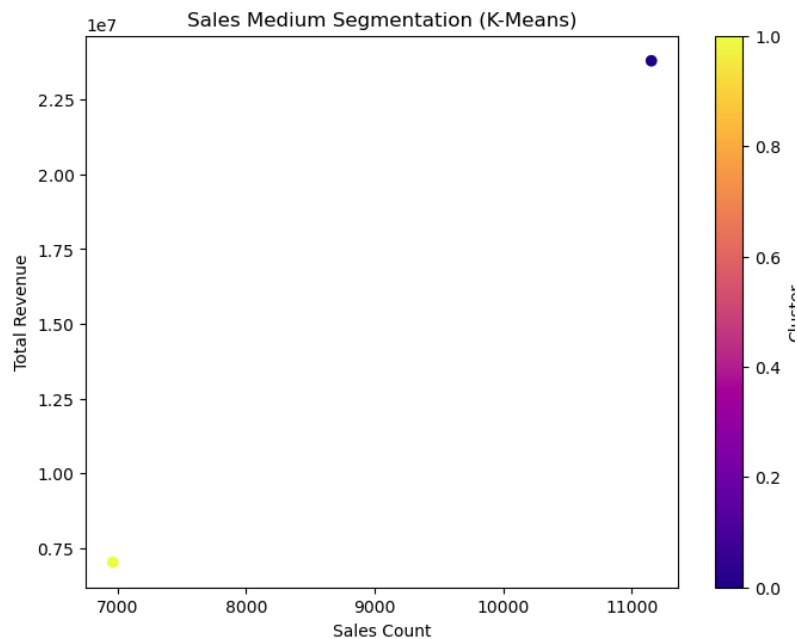


Figure 28: Chart displaying the results of K-Means clustering applied to sales medium data

5.5 Sales Medium Segmentation Using DBSCAN: Detecting Outliers and Anomalies

To further investigate sales medium performance, we applied DBSCAN, a density-based clustering algorithm. This analysis treated both direct and online sales as outliers, highlighting the significant disparity between the two channels. This finding underscored the need for a thorough examination of factors influencing online sales performance.

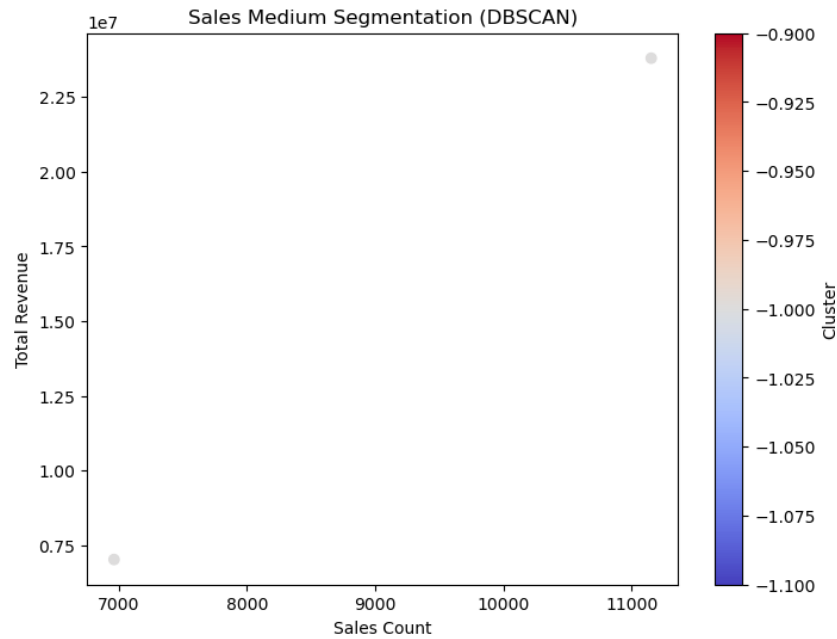


Figure 29: Chart displaying the results of DBSCAN clustering applied to sales medium data

5.6 Key Findings and Strategic Implications: Translating Clusters into Actionable Insights

Our analysis yielded several key findings:

- **Geographic Performance Clusters:** Sales locations were segmented into distinct performance clusters, revealing areas for targeted strategies.
- **Channel Performance Disparity:** Direct sales significantly outperformed online sales, indicating a need for online sales optimization.
- **Outlier Detection:** DBSCAN highlighted the significant difference between direct and online sales, emphasizing the need for further investigation.

Based on these findings, we recommend:

- **Targeted Regional Strategies:** Develop tailored marketing and sales strategies for each geographic performance cluster.
- **Online Sales Optimization:** Investigate and implement strategies to enhance online sales performance.
- **Resource Allocation:** Allocate resources based on cluster performance to maximize returns.
- **Further Analysis:** Conduct in-depth analysis of factors influencing online sales and regional performance.

5.7 Conclusion: Driving Targeted Strategies Through Data-Driven Segmentation

This project successfully segmented sales data using K-Means and DBSCAN, providing valuable insights into geographic and channel performance. By leveraging these data-driven insights, Nestle can develop targeted strategies, optimize resource allocation, and enhance overall sales performance. The identification of performance clusters and outliers provides a solid foundation for future sales forecasting and strategic decision-making.

Section 6: Project Summary

This project culminated in a holistic understanding of Nestle's sales dynamics, providing a robust analytical framework for strategic decision-making. Through a series of focused analyses, we navigated the complexities of sales data, transforming raw information into actionable insights. The initial challenge posed by single-item transactions led to a refined approach, demonstrating the importance of analytical flexibility. Time series forecasting, particularly with the Prophet model, offered reliable projections for future sales, while segmentation analysis illuminated distinct performance clusters across geographic locations and sales mediums.

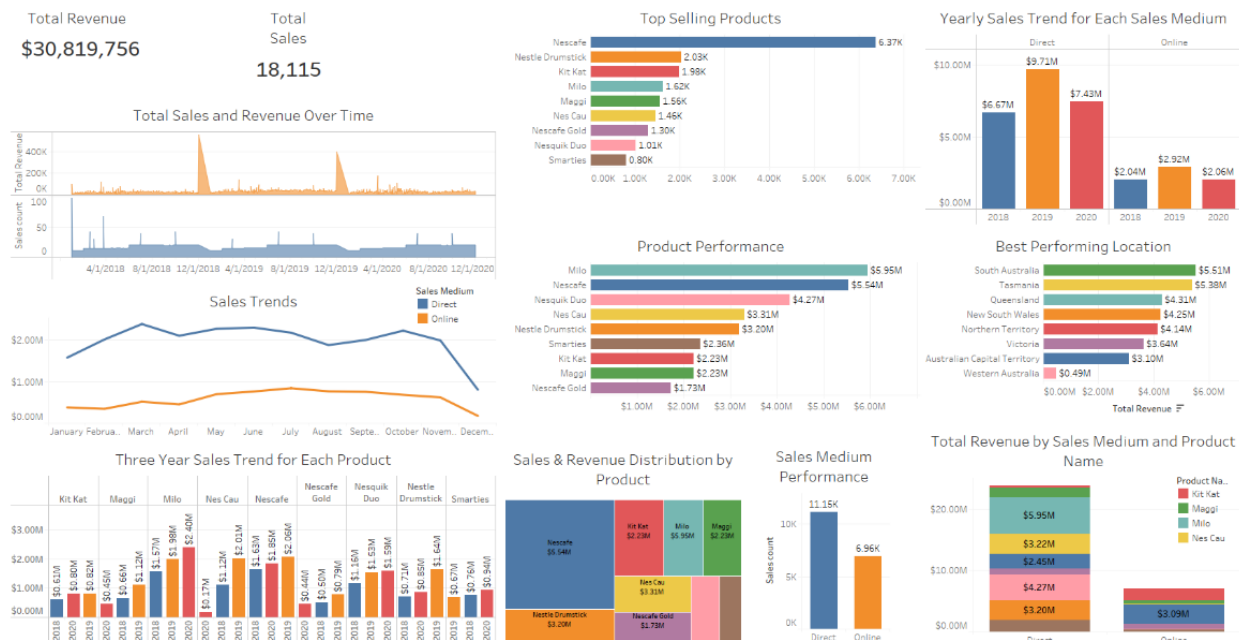


Figure 30: Partial View of the interactive dashboard developed using Tableau

The insights derived from these analyses are pivotal for optimizing sales strategies, enhancing resource allocation, and driving sustainable growth. We recommend Nestle leverage these findings to:

- Refine product promotion strategies based on individual product performance.
- Optimize online sales channels to bridge the performance gap with direct sales.
- Implement targeted regional strategies based on geographic performance clusters.
- Employ Prophet for ongoing sales forecasting due to its demonstrated accuracy.
- Continuously monitor and refine segmentation strategies to adapt to market changes.

By integrating these data-driven recommendations, Nestle can enhance its market presence, improve operational efficiency, and solidify its strategic advantage. This project underscores the power of comprehensive data analysis in transforming business challenges into opportunities, ultimately driving informed decisions that align with Nestle's long-term objectives.

Appendix

Appendix A: Glossary

Actionable Insights	Information derived from data analysis that can be directly applied to make informed decisions and take effective actions.
ARIMA (Autoregressive Integrated Moving Average)	A statistical time series forecasting method used to predict future values based on past data trends.
Categorical Insights	Understanding derived from analyzing data that falls into distinct categories (e.g., product types, locations, sales channels).
Channel Efficiency	The effectiveness of different sales mediums (e.g., online, direct sales) in generating revenue and sales.
Clustering Analysis	A technique used to group similar data points together based on their characteristics.
Data-Driven Insights	Information and conclusions derived from analyzing data, used to inform decision-making.
DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	A clustering algorithm used to group data points based on their density, identifying clusters and outliers.
EDA (Exploratory Data Analysis)	The process of analyzing and summarizing datasets to understand their main characteristics, often using visual methods.
Geocoding	The process of converting addresses or location names into geographic coordinates.
Geospatial Analysis	Analyzing data with a geographic component, often visualized on maps.
Hotspots	Areas with high concentrations of sales or revenue.
K-Means	A clustering algorithm that partitions data into k distinct clusters based on their proximity to cluster centroids.
Market Basket Analysis (MBA)	A data mining technique used to identify associations between items purchased together.
Multifaceted Analysis	Examining data from multiple perspectives and dimensions.
Outliers	Data points that significantly deviate from the rest of the dataset.

Pandas DataFrame	A two-dimensional labeled data structure in Python, used for data manipulation and analysis.
Per-Unit Revenue	The amount of revenue generated by each individual unit sold.
Product Associations	Relationships between products that are frequently purchased together.
Product Portfolio Optimization	The process of managing and adjusting the range of products offered to maximize profitability.
Profitability per Sale	The profit generated from each individual sale.
Prophet	A forecasting procedure developed by Facebook, designed for time series data with strong seasonal effects.
Resource Allocation	Distributing resources (e.g., budget, personnel) based on analysis to maximize efficiency.
Revenue Drivers	Factors that contribute to generating revenue.
Revenue Variability	The degree to which revenue fluctuates or changes over time or across different categories.
Sales Channel	The method through which products are sold (e.g., online, direct sales).
Sales Medium	The method or avenue through which products or services are sold to customers.
Segmentation	Dividing a market or dataset into distinct groups based on shared characteristics.
Segmentation Analysis	The process of dividing a market or dataset into distinct groups based on shared characteristics.
Skewed Results	Data that is not evenly distributed, leading to a bias in the analysis.
Strategic Implications	The potential impact of findings on strategic decisions
Summary Statistics	Descriptive statistics that summarize the central tendency, dispersion, and shape of a dataset (e.g., mean, median, standard deviation)
Temporal Sales Trends	Patterns and changes in sales data over time (e.g., daily, weekly, monthly).
Time Series Forecasting	Predicting future values based on historical time-stamped data.

Appendix B: GitHub Repository

Overview

This repository contains the code, analyses, and visualizations developed for the comprehensive analysis of Nestle's sales data, as described in the provided sections. The project aims to derive actionable insights for strategic enhancement by exploring product profitability, sales channel performance, geographic distribution, and applying various analytical techniques including time series forecasting, clustering, and exploratory data analysis. This repository serves as a central hub for all project-related materials, facilitating reproducibility and further development. It showcases the application of data-driven strategies for optimizing sales performance and informing strategic decision-making within Nestle.

GitHub Repository Link

<https://github.com/36nw/Nestle-Sales-Data-Insights/tree/main>

GitHub Repository Contents

Explore the repository to find resources and documentation that guide users through the analysis of Nestle's sales data:

- **README.md:** Introduction and overview of the Nestle sales data analysis project.
- **data:** The Nestle sales dataset used for the analysis.
- **notebooks:** Jupyter notebooks containing the Python code for data processing, analysis, visualization, and modeling.
- **visualizations:** Folders containing generated charts, maps, and interactive dashboards.
- **reports:** Documents detailing the analysis findings, strategic recommendations, and conclusions (including the provided sections).
- **LICENSE:** MIT License file (or your chosen license).

Reference

Works Cited

- [1] HannatuGwabin. (2023). *Nestle Sales Analysis*. GitHub.
<https://github.com/HannatuGwabin/NESTLE-SALES-ANALYSIS/blob/main/NESTLE%20SALES%20DATASET.xlsx>
- [2] Python Software Foundation. (2025). *Python (3.13.3)* [Computer software].
<https://www.python.org/>
- [3] The pandas development team. (2024). *pandas-dev/pandas: Pandas (2.2.3)* [Computer software]. GitHub. <https://github.com/pandas-dev/pandas>
- [4] Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Bastien, T., ... & Oliphant, T.E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357-362.
- [5] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), 2825-2830.
- [6] Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37-45.
- [7] Folium contributors. (Year). Folium (Version X.X.X) [Computer software]. Github.
<https://github.com/python-visualization/folium>
- [8] Hunter, J. D. (2007). ¹ Matplotlib: A 2D graphics environment. ² *Computing in science & engineering*, 9(3), 90-95
- [9] Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021> ¹
- [10] Tableau. (2025). Tableau Desktop 2025.1. [Internet]. Retrieved from <https://www.tableau.com/products/all-features> [6] (Accessed April 10, 2025).

This page intentionally left blank
