**Pens and Printers Sales Data Analysis and Business Recommendations**

1. **Introduction**

The primary objective of this analysis is to inform and optimize our sales strategies for the recently launched office stationery line at Pens and Printers. The analysis will address critical questions posed by the executive team, aiming to provide actionable insights within a tight timeframe. A brief overview of the analysis goals and the dataset is given below:

1. Customer Distribution by Sales Approach: Understand the distribution of customers among different sales approaches, namely Email, Call, and Email and Call combined.
2. Revenue Spread: Investigate the overall spread of revenue and dissect it for each sales method to identify patterns and potential outliers.
3. Revenue Over Time: Explore whether there are discernible differences in revenue over time for each sales method, enabling us to refine our ongoing strategies.
4. Method Recommendation: Based on the data, recommend the most effective sales method. Consideration will be given to the balance between results and the time investment required from our dedicated team.
5. Customer Group Differences: Uncover potential differences between customer groups that extend beyond revenue metrics. This exploration aims to provide valuable context to what went well with each sales approach.

Pens and Printers, established in 1984, is a trusted provider of high-quality office products to large organizations. We have built enduring relationships with our customers, ensuring their trust in our ability to provide the best-suited solutions. As consumer buying behaviour evolves, our sales tactics must adapt accordingly. There are three new product sales methods:

1. Email: Customers received targeted emails at the product line launch and follow-up, requiring minimal team effort.
2. Call: Sales team members conducted phone calls, investing approximately thirty minutes per customer.
3. Email and Call: A combined approach starting with an informational email followed by a shorter call a week later.

The urgency of this analysis is underscored by the need to report to the executive team within four weeks. The insights derived from this analysis will play a pivotal role in shaping our future sales strategies.

1. **Data Validation**
2. **week:** There are 15000 entries in this column. The entries are of integer data type and between 1, 2, 3, 4, 5, and 6 in the table. There is no missing value in the column. No transformation was done to this column.
3. **sales\_method:** There are 5 unique entries (Email, Call, Email + Call, em + call, and email) in the column instead of the 3 (Email, Call, Email + Call) discussed in the project description. They are of the object data type. There was no missing value in the column. The ‘em + call’ and ‘email’ entries were replaced with ‘Email’ and ‘Email + Call’.
4. **customer\_id:** There are 15000 unique entries in this column. The column entries are of the object data type. There was no duplicates or missing values in this column. No transformation was done to this column.
5. **nb\_sold:** There are 15000 entries in this column. The entries range between 7, 8, 9, 10, 11, 12, 14, 15, and 16. They are of the integer data type. There was no missing value or outliers. No transformation was done to this column.
6. **revenue:** There are 13926 entries in this column of which 6743 are unique entries in the column. There are 1074 missing values (NaN) in this column. The column is of the float data type. There were no outliers in this column. The missing values were replaced with the median value of the column data because the revenue values were skewed and did not follow a normal distribution.
7. **years\_as\_customer:** There are 15000 entries in this column of which 42 are unique entries. The entries range between 0 and 39, 47 and 63. They are of the integer data type. There was no missing value. No transformation was done to this column.
8. **nb\_site\_visits:** There are 15000 entries in this column of which 27 are unique entries. The entries range between 12 and 37, and 41. They are of the integer data type. There was no missing value. No transformation was done to this column.
9. **state:** There are 15000 unique entries in this column of which 50 are unique entries. The column entries are of the object data type. There was no duplicates or missing values in this column. No transformation was done to this column.
10. **Exploratory Analysis**
    1. **Histogram of Revenue with missing values**

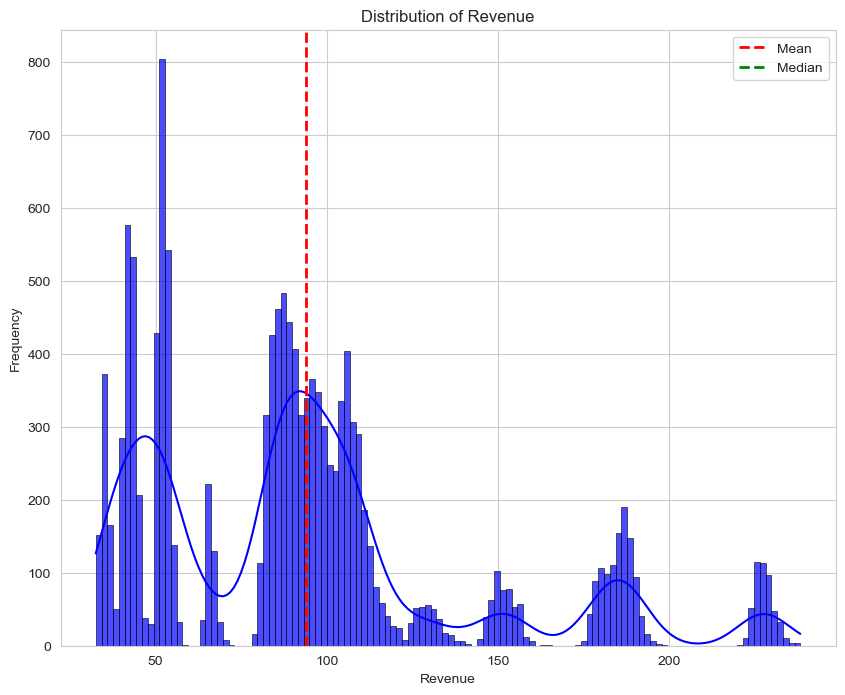


Figure 1: Histogram of Revenue before filling missing values with Median value

As seen from the figure above, the histogram has more than two distinct peaks; as such, it has multiple modes, and the tail is longer on the right side, indicating that it is positively skewed. From the analysis, the revenue data has a median value of 89.5 and a mean value of 93.935, as seen by the red dashed lines on the plot. It therefore means that the mean is on the right of the median, indicating the effect of higher values on the right-hand side of the plot.

The presence of multiple modes indicates that the data is heterogeneous and contains distinct groups or clusters. Each mode represents a concentration of data points.

The right skewness implies that there are some fewer but higher values that are pulling the mean to the right of the median, with the majority of the data points concentrated on the left side of the distribution.

Combined, the multimodal nature and skewness of the revenue histogram at Pens and Printers are essential for decision-making as they are indicative of the different market conditions and the behaviour of the customers.

* 1. **Histogram of Revenue with missing values filled with the Median.**

As seen from the figure above, the histogram still maintained its numerous peaks, and the tail remains longer on the right side, indicating that it is positively skewed. From the analysis, the data has a median value of 89.5 and a mean value of 93.62, as seen by the green and red dashed lines on the plot.

This suggests that the central tendency of the data hasn't been greatly affected. This could mean that the missing values were not concentrated in a way that would significantly shift the central location.

It also suggests that the overall structure of the data is maintained. If the shape remains similar after imputation, it could indicate that the imputed values did not introduce a significant amount of variability into the dataset.

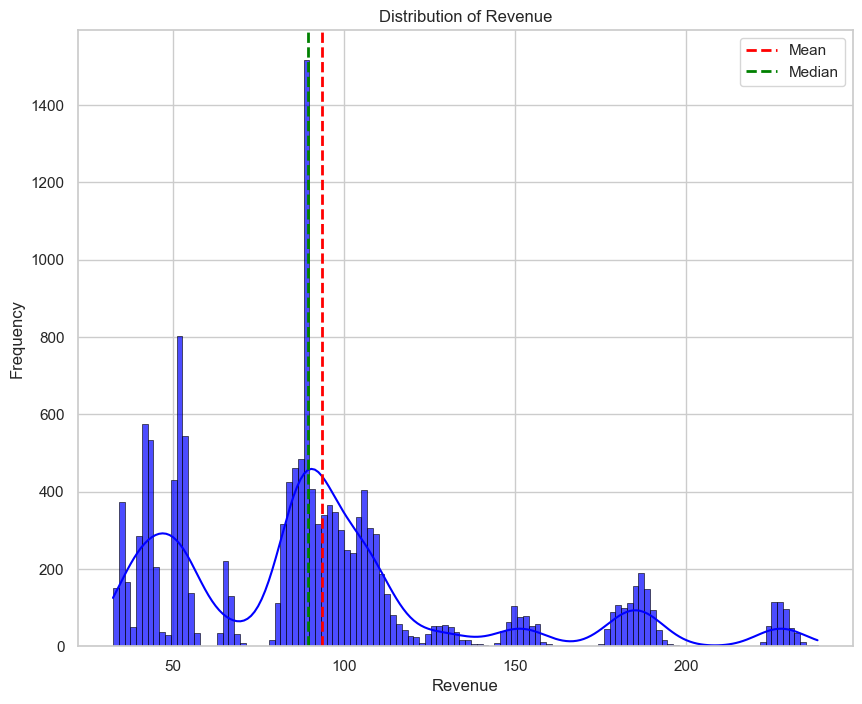


Figure 2: Histogram of Revenue after filling missing values with Median value

* 1. **How many customers were there for each approach?**

The sales method distribution reveals that Email emerges as the most prevalent sales method, constituting the highest count of 7466 and percentage of 49.77. The Call approach demonstrates substantial engagement, with a count of 4962 customers accounting for 33.08 percent of the total customers. The Email + Call approach had a lower count of 2572 and a percentage of 17.15 percent, representing a distinct segment of engaged customers.

Optimising and refining our email strategy could yield continued success given the large customer base of this approach. The significant engagement through the Call approach indicates the effectiveness of direct communication. Exploring ways to enhance this strategy could further boost customer interaction. The Email + Call approach represents a unique set of engaged customers, and applying it strategically can yield targeted benefits.

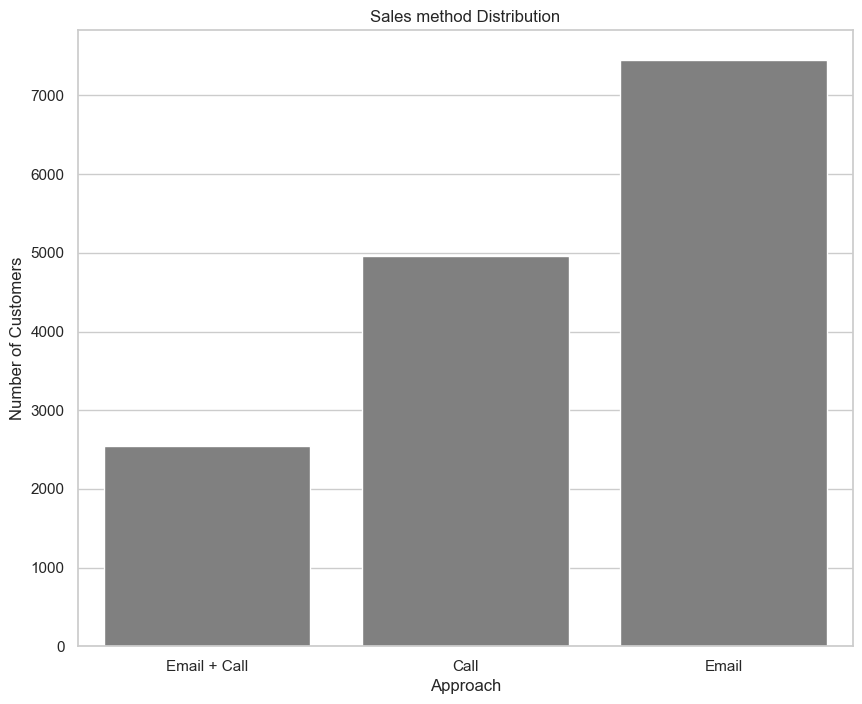


Figure 3: Bar Plot of Sales Method Distribution

Understanding the distribution of customers across different sales methods is instrumental in refining our sales strategies. By focusing on the insights provided, we can tailor our approaches to maximise customer engagement and ultimately drive long-term success.

It is recommended to conduct a detailed analysis of customer behaviour within each sales method and implement A/B testing to refine email content and call scripts. explore personalised strategies for the Email + Call segment, and continuously monitor and adapt sales approaches based on evolving customer preferences.

* 1. **What does the spread of the revenue look like overall?**

The median value of $89.50 suggests that approximately 50% of the revenue data falls below this point when the dataset is arranged in ascending order, indicating a central concentration of values around this median. The Lower Quartile (25%) value of $53.04 represents the value below which 25% of the revenue data falls; it indicates that a quarter of the revenue values are relatively lower, while the Upper Quartile (75%) value of $106.07 signifies the upper range of revenue values. The interquartile range (IQR) of approximately $53.03 illustrates the spread of the middle 50% of the revenue data. This provides insights into the variability and dispersion of values within this range.

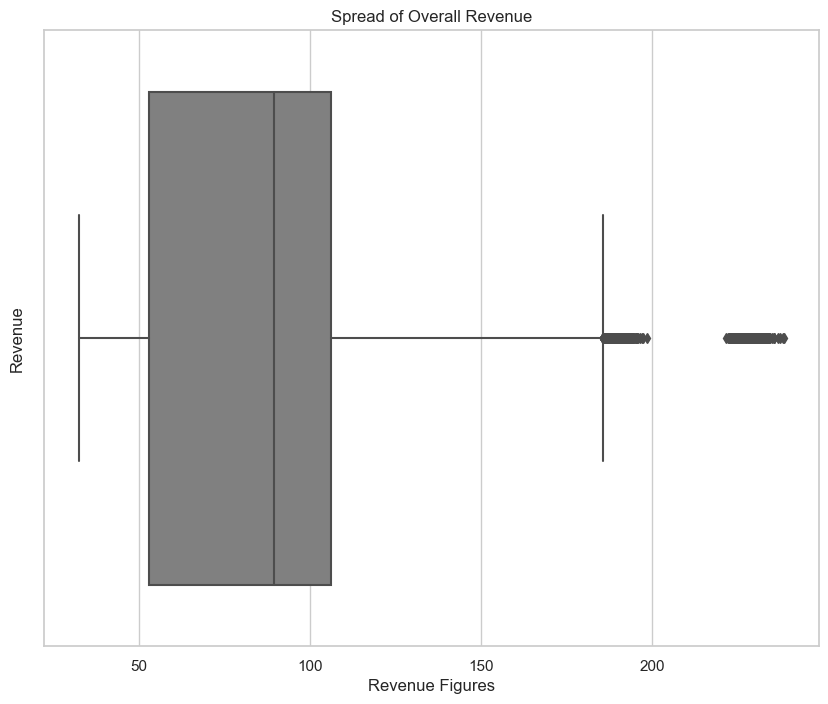


Figure 4: Spread of Revenue for each Sales Method

Higher values may require further investigation. Understanding the spread of revenue data allows for targeted marketing strategies. For example, customers falling below the median value can be targeted with promotions or incentives. Insights into quartiles and the spread of revenue values can inform pricing strategies and guide decisions on product offerings.

It is recommended to regularly monitor revenue data for shifts and trends, investigate potential outliers for anomalies or opportunities, and consider targeted marketing initiatives based on quartile insights.

* 1. **What does the spread of the revenue for each sales method look like?**

The Call method demonstrates a relatively narrow spread of revenue values, with a median of $49.94 and an IQR of $11.35. The Email method exhibits a wider spread, reflected in the higher IQR of $16.1 and a median revenue of $94.2750. The Email + Call method shows the broadest spread among the three, with a substantial IQR of $39.71; the median value is $182.14.

The Call method, with its narrower spread, indicates consistent revenue generation. Strategies to optimise this method may focus on efficiency and targeted customer engagement. The wider spread of the Email method suggests diverse revenue outcomes. Strategies can be tailored for different customer segments to maximize revenue potential. The Email + Call method, while having a broader spread, shows a significant median. Focused efforts to nurture high-value customers within this method can yield substantial returns.

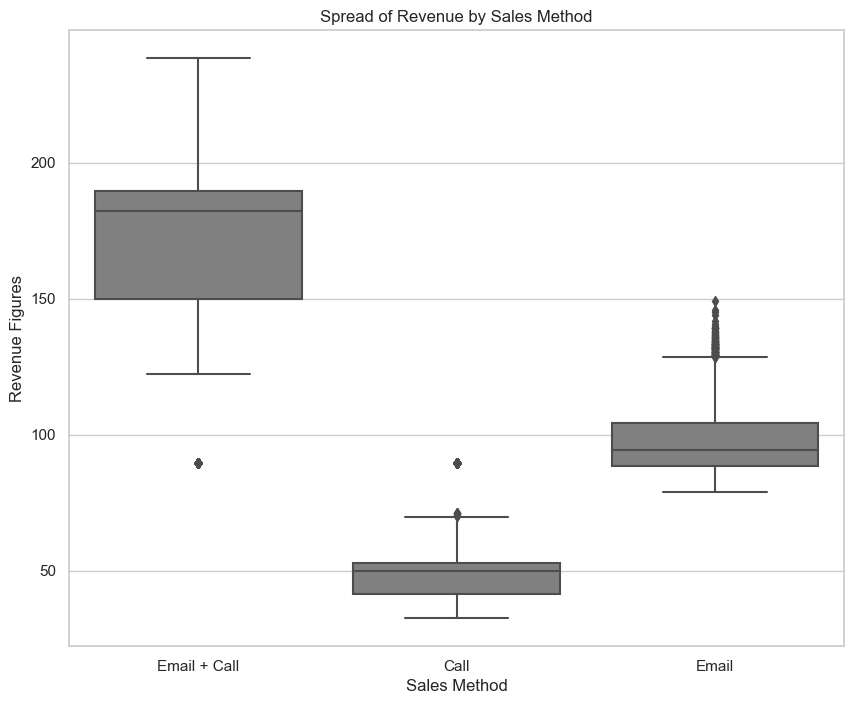


Figure 5: Spread of Revenue for each Sales Method

The spread of revenue data for different sales methods is pivotal for strategic decision-making. It is recommended to regularly monitor revenue trends for each sales method, conduct in-depth analyses to identify factors influencing revenue variability between sales methods, and implement targeted strategies for customer segments within each sales method.

* 1. **Was there any difference in revenue over time for each of the methods?**

The provided data reveals the weekly variations in revenue and highlights the differential impact of each sales method. Each sales method exhibits varying levels of revenue over the six-week period.

The Email method consistently shows the highest revenue in the initial weeks following product launch and declines as the number of weeks increases, followed by Email + Call and Call. Fluctuations in revenue are observed across all methods, indicating dynamic market responses. Email + Call demonstrates a notable surge in the last week after a slow start in the initial weeks following the product launch, contributing to a significant difference.

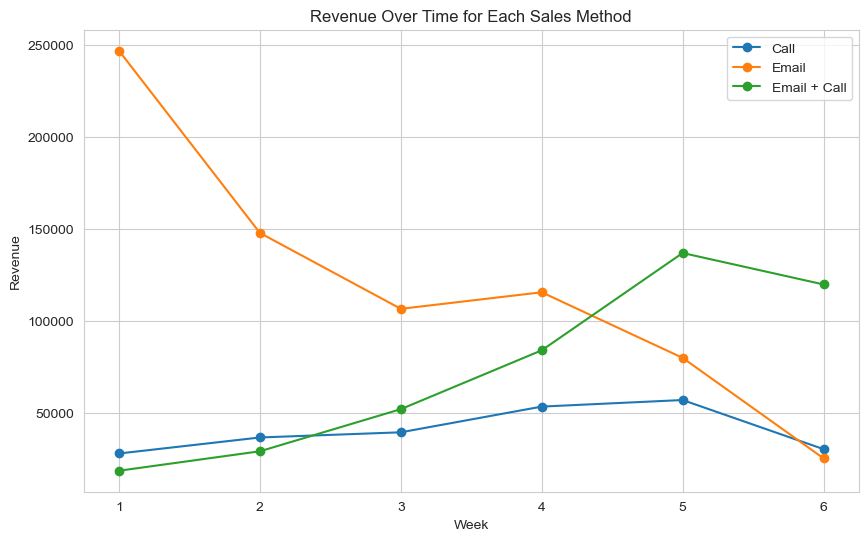


Figure 6: Line Plot of Revenue Over Time by Sales Method

Understanding the differences in revenue over time helps identify weeks where specific sales methods had a pronounced impact. Insights into weekly revenue differences guide the optimisation of sales and marketing strategies for each method.

It is recommended to implement agile marketing strategies to capitalise on weeks with high revenue differences, regularly monitor revenue trends, and adapt sales approaches based on evolving market dynamics.

* 1. **Based on the data, which method would you recommend we continue to use?**
     1. **Total Revenue and Number of Sales by Sales Methods**

The scatter plot clearly illustrates the influence of sales methods on the revenue generated from sales, given the same number of products sold. Each point on the plot represents a combination of the number of sales and the corresponding revenue, with different colours indicating distinct sales methods.

The most noticeable trend is that the Email + Call sales method consistently outperforms other methods in terms of revenue for a given number of products sold. Points associated with the Email + Call method are clustered in the upper-right region of the plot, indicating higher revenue compared to other methods.

Points from different sales methods are not evenly distributed, signifying that the choice of sales method has a substantial impact on the revenue generated. It is evident that revenue varies significantly even when the number of products sold remains constant, underscoring the importance of selecting an effective sales strategy.

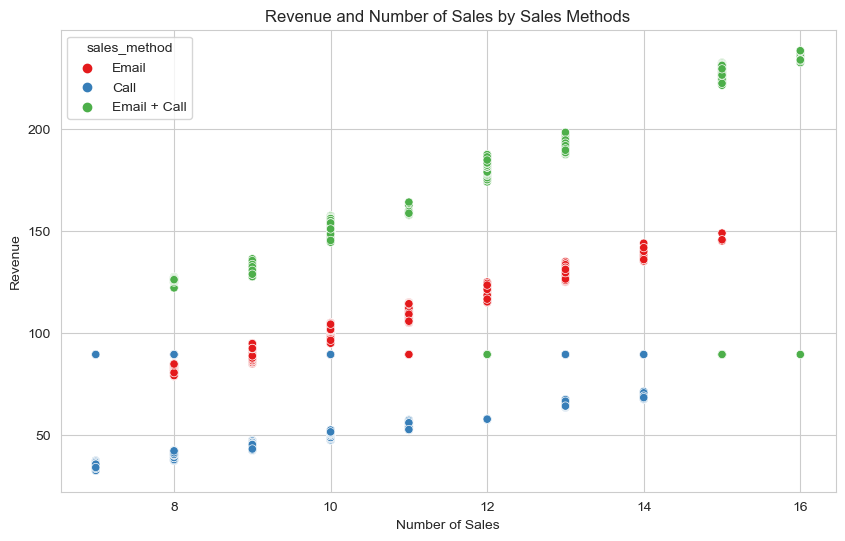
****

Figure 7: Total Revenue and Number of Sales by Sales Methods

Based on the analysis, the Email + Call sales method has consistently proven to be more effective in generating revenue. Consider allocating more resources and focus to this method to capitalize on its success. Explore the specific attributes of the Email + Call strategy that contribute to its success. This may include analysing customer interactions, timing, or other factors that differentiate it from less successful methods.

* + 1. **Total Revenue and Number of Sales by Sales Methods**

The average revenue values and conversion rates provide insights into the effectiveness of each method in generating revenue and converting leads into customers. The Call method demonstrates a lower average revenue of $49.13 but maintains a reasonable conversion rate of 0.389. This indicates that, while individual transactions may be smaller, the method is effective in converting leads.

The Email method boasts a higher average revenue of $96.57, coupled with a conversion rate of 0.393. This suggests that the Email method generates higher revenue per transaction and maintains a competitive conversion rate. The Email + Call method stands out with the highest average revenue of $170.88 and a conversion rate of 0.457. This indicates a successful combination of email communication and personalised calls, resulting in both high revenue and conversion rates.

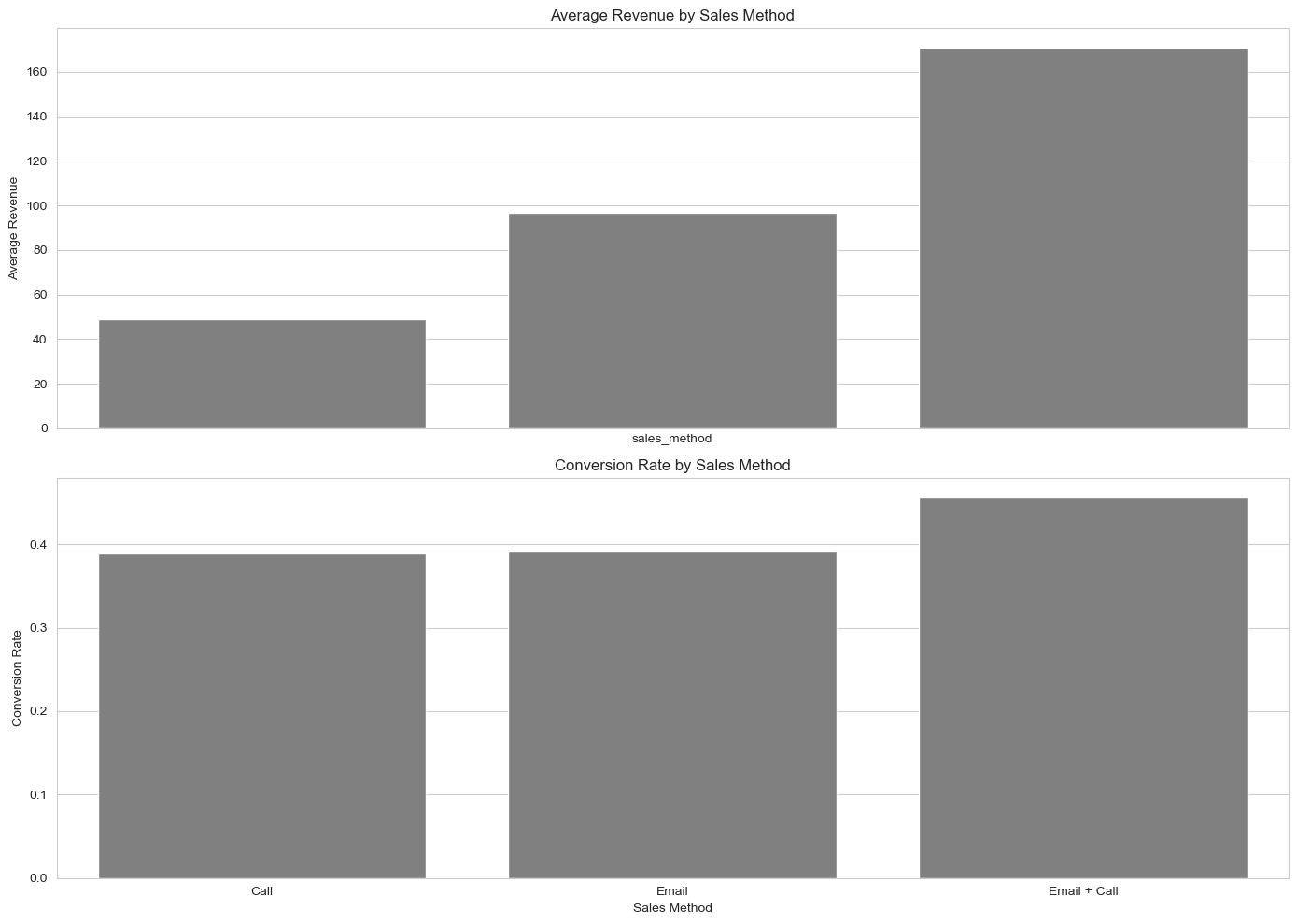


Figure 8: Average revenue and conversion rate for each sales method.

The Call method may benefit from strategies that focus on increasing the average transaction value while maintaining its effective conversion rate. Strategies for the Email method should capitalise on its higher average revenue, potentially by targeting high-value customer segments to further increase revenue and conversion rates. The Email + Call method showcases exceptional performance in both average revenue and conversion rate. Efforts should be directed towards scaling this method and identifying factors contributing to its success.

It is recommended to conduct a detailed analysis of customer segments for each sales method, implement targeted marketing campaigns to maximise conversion rates, and explore opportunities to scale the Email + Call method based on its outstanding performance.

* + 1. **Efficiency by Revenue per Site Visit for Each Sales Method**

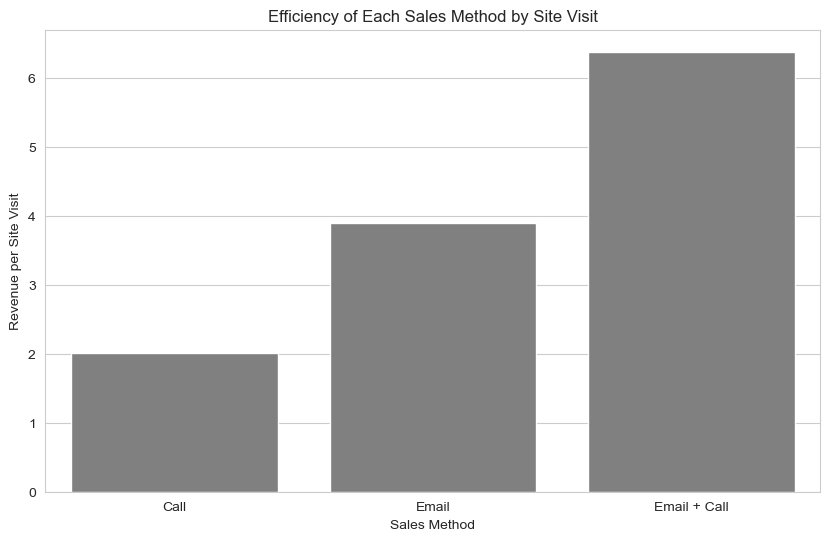


Figure 9: Efficiency by Revenue per Site Visit for Each Sales Method

The Revenue per Site Visit values offer insights into the effectiveness of each method in generating revenue in relation to the number of site visits.

The Call method demonstrates a lower average revenue per site visit of $2.01. The method still generates revenue efficiently, indicating a moderate return on investment for each site visit. The Email method boasts a higher revenue per site visit of $3.90, suggesting that each site visit results in a more substantial revenue contribution. This aligns with the method's higher average revenue. The Email + Call method stands out with the highest revenue per site visit at $6.38, indicating a significant return on each interaction. This method proves to be particularly effective in converting site visits into revenue.

There may be opportunities to optimise and increase the effectiveness of each interaction to enhance the metrics of the Call method. Strategies for the Email method should focus on maintaining its higher revenue per site visit, potentially by targeting high-value customer segments and optimising email content. The Email + Call method showcases exceptional performance in revenue per site visit. Efforts should be directed towards scaling this method and understanding the factors contributing to its success.

* + 1. **Recommendation of Sales Method to Adopt**

The recommendation on which sales method to continue using depends on the specific priorities and goals of Pens and Printers, taking into account the effort put in by the sales team, the revenue and conversion metrics. Let's consider the key insights from the provided data.

1. Email Method: this requires very little work for the team. It shows a competitive average revenue, conversion rate, and Revenue per Site Visit. The Email method is efficient and may be suitable for targeting a broad audience with minimal resource investment.
2. Call Method: this involves direct communication, with team members spending around thirty minutes per customer. While the average revenue is lower compared to Email + Call, the Call method maintains a reasonable conversion rate. It's effective in converting leads and might be suitable for personalized engagement.
3. Email and Call Method: this Combines the efficiency of an initial email with a follow-up call to discuss customer needs. The call is shorter, around ten minutes per customer. This method exhibits the highest average revenue, conversion rate, and Revenue per Site Visit. It shows significant potential for both revenue generation and effective conversion.
4. Recommendation: Considering the comprehensive analysis and the emphasis on revenue, conversion rates, and efficiency, the "Email and Call" method appears to be the most promising. It strikes a balance between the efficiency of email communication and the effectiveness of personalized calls, resulting in both high revenue and conversion rates. Efforts should be directed towards optimizing and scaling this combined approach.
   1. **Other difference between customer groups.**
      1. **Customer Behaviour Analysis**

This offers valuable information on the effectiveness of each method in terms of both transaction volume and generated revenue.

The Email + Call method demonstrates the highest average number of items sold at 12.23, suggesting that this approach is particularly effective in encouraging customers to purchase more items per transaction. The Call and Email methods have comparable average numbers of items sold at 9.51 and 9.73, indicating a similar transaction volume.

The Email + Call method stands out with the highest average revenue of $170.88, suggesting that not only does it sell more items, but each transaction yields a higher total revenue. The Email method follows with a considerable average revenue of $96.57, emphasising its effectiveness in generating substantial transaction values. The Call method has a lower average revenue per transaction at $49.13 but maintains a competitive position.

Given the higher average number of items sold and average revenue, there is an opportunity to further optimise and scale the Email + Call method. The Email method's higher average revenue indicates potential for targeted marketing strategies to maximise transaction values. While the Call method has a lower average revenue, its transaction volume and efficiency in generating revenue should be considered in the overall strategy.

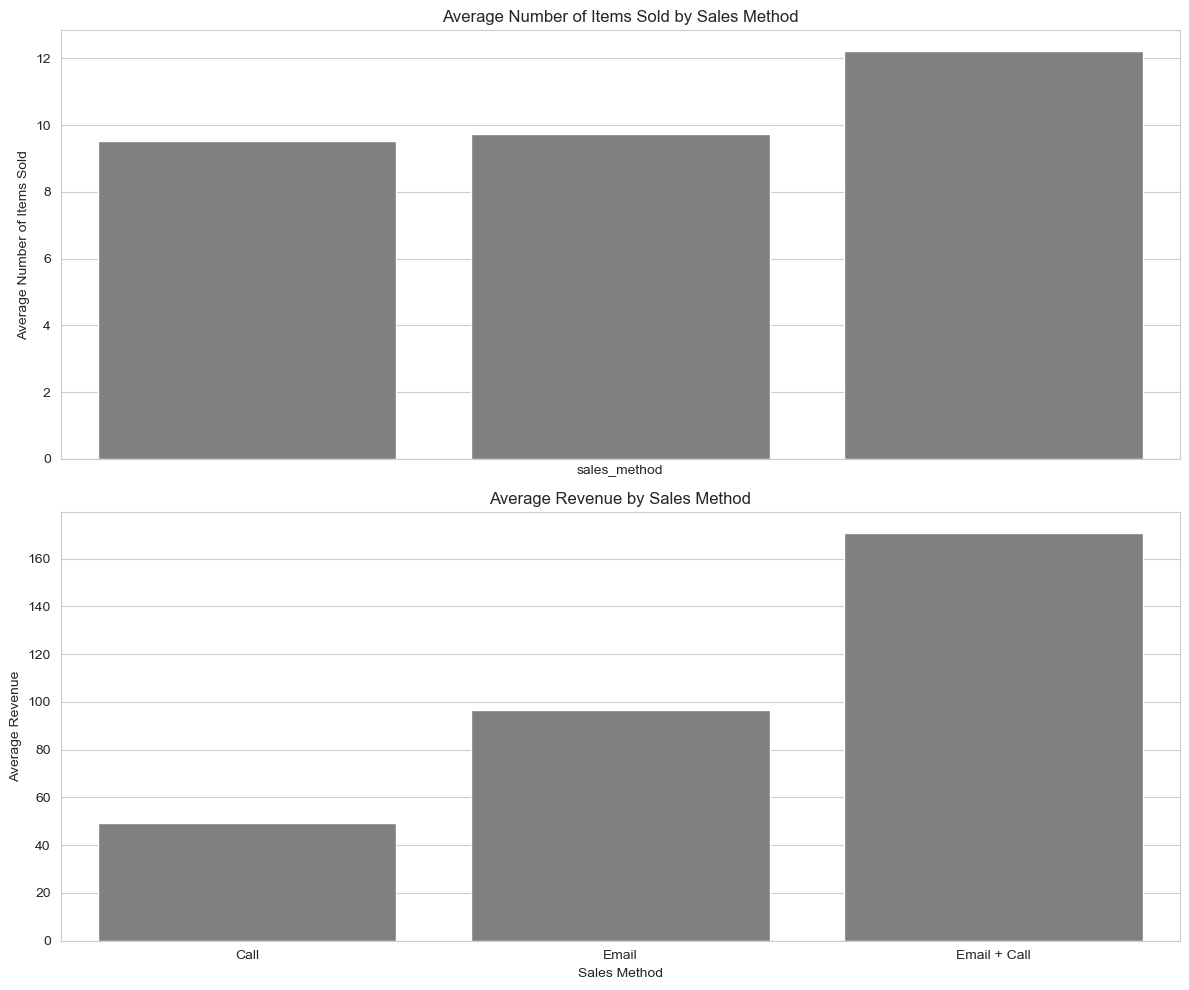


Figure 10: Average Number of Items Sold and Average Revenue by Sales Method

It is recommended to scale the Email + Call method based on its success in both average number of items sold and average revenue, implement targeted marketing campaigns for the Email method to enhance transaction values, and consider optimising the Call method for increased efficiency in generating revenue.

* + 1. **Customer Tenure and Site Visits Analysis**

This provides valuable evidence on customer loyalty, tenure, and engagement with each sales method.

The Call method has the highest average years as a customer at 5.18, indicating that customers using this method have been with the company for a longer period of time. The Email method follows closely, with an average of 4.98 years as a customer, reflecting a strong level of customer loyalty. The Email + Call method has the lowest average years as a customer at 4.51, suggesting a relatively newer customer base.

The Email + Call method has the highest average number of site visits at 26.77, indicating active and engaged customers who visit the site frequently. Both the Call and Email methods have comparable average numbers of site visits at 24.42 and 24.75, suggesting a similar level of customer engagement.

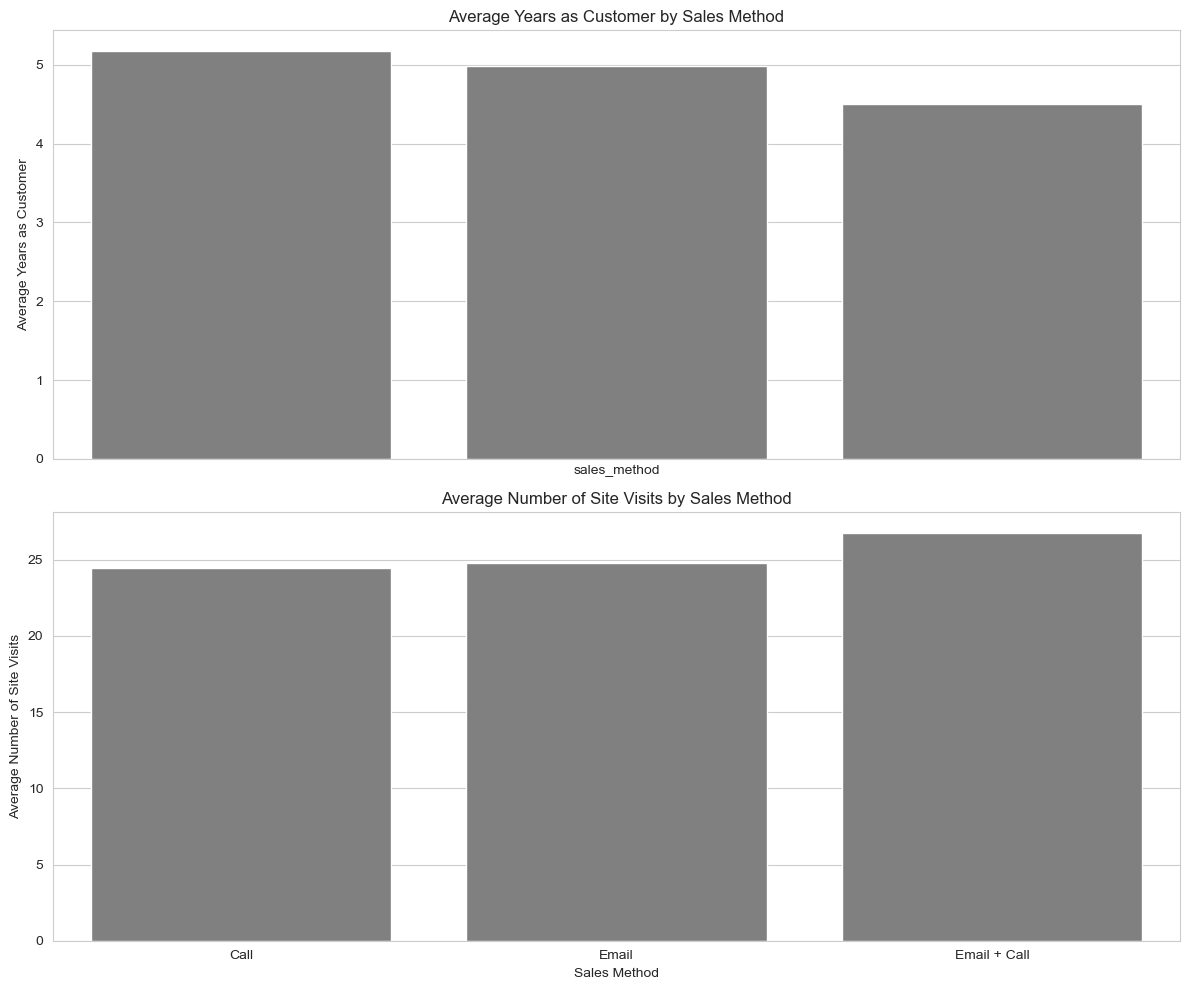


Figure 11: Average Years as Customer and Average Number of Site Visits by Sales Method

The higher average years as a customer in the Call and Email methods highlight strong customer loyalty. Strategies should focus on leveraging this loyalty for repeat business and upselling. While the Email + Call method has newer customers on average, its higher engagement with more site visits provides an opportunity to enhance customer relationships and potentially increase customer loyalty over time.

It is recommended to implement targeted loyalty programmes for the Call and Email methods to capitalise on the existing customer base, develop engagement strategies for the Email + Call method to foster long-term customer relationships, and regularly monitor customer behaviour and adjust strategies based on evolving trends and preferences.

* + 1. **Geographic Analysis**

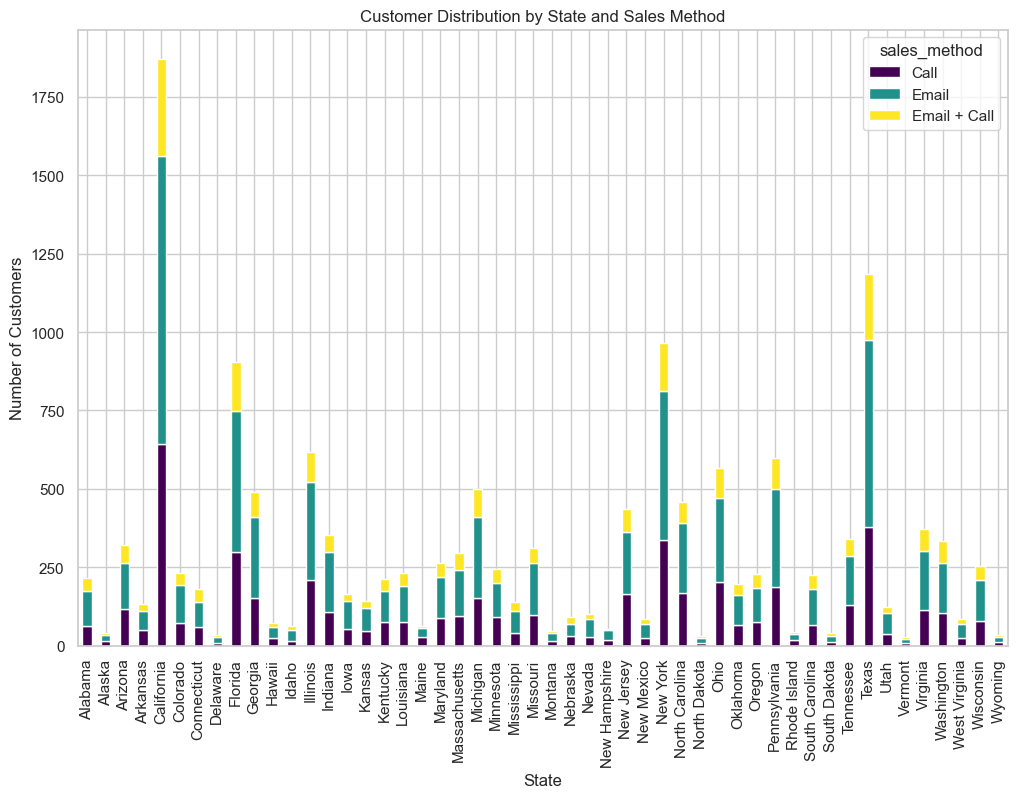


Figure 12: Customer distribution by state and sales method

The data offers insights into the geographical distribution of customers for each sales method. The provided data presents the customer distribution by state for each sales method, with values indicating the number of customers in each state for the Call, Email, and Email + Call methods.

Customer distribution varies across states for each sales method, reflecting regional preferences and market dynamics. Larger customer concentrations are observed in states with higher populations, such as California and Texas.

Certain states may exhibit preferences for specific sales methods, influencing the overall distribution. Analysis of state-level data can inform targeted marketing strategies based on regional preferences.

Understanding customer distribution by state allows for targeted regional marketing campaigns, tailoring messages to specific state preferences. States with higher customer concentrations can be prioritised for more extensive marketing efforts and resource allocation.

It is recommended to implement state-specific marketing strategies to cater to regional preferences, prioritise resource allocation and sales efforts in states with higher customer concentrations, and regularly update and analyse state-level data to adapt strategies based on evolving customer distribution patterns.

1. **Defining of a Metric for Business Monitoring**

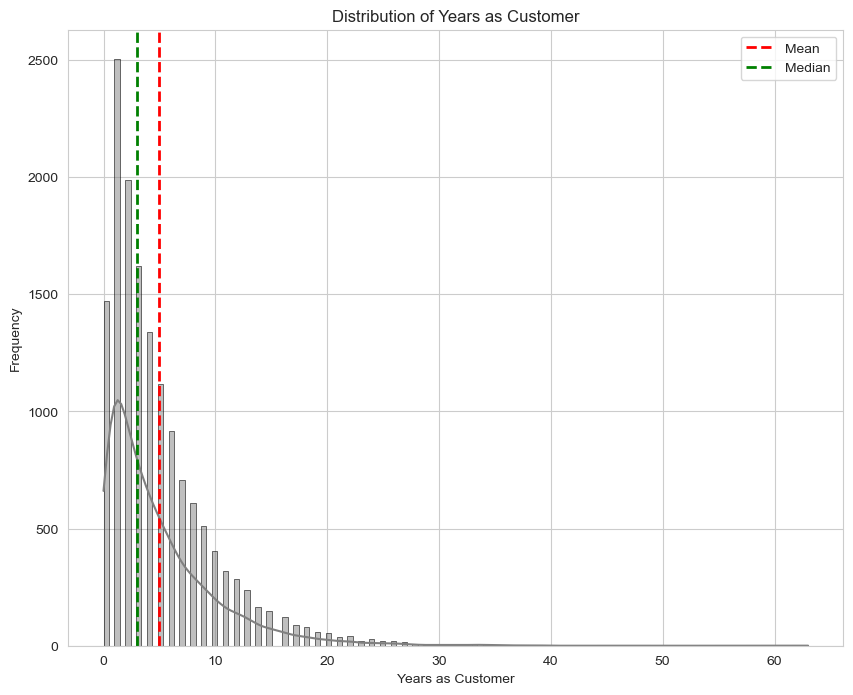


Figure 13: Distribution of Years as Customer

The calculated Average Customer Tenure and Customer Retention Rate provide essential insights into the longevity and loyalty of the customer base. With a median tenure of 3.0 years and a mean tenure of 4.97 years, the average customer commitment stands at 4.97 years. This underscores the effectiveness of retention strategies, as a longer tenure correlates with heightened customer loyalty, satisfaction, and stable revenue.

The impressive Customer Retention Rate of 90.19% further emphasizes the organization's ability to retain a significant portion of its customer base, contributing to financial stability and growth. These metrics serve as benchmarks guiding strategies to enhance customer satisfaction, loyalty, and long-term business success. Regular monitoring ensures adaptive approaches aligned with evolving customer needs.

* 1. **Customer Lifetime Value (CLV)**

Customer Lifetime Value (CLV) is a comprehensive metric designed to quantify the long-term value that a customer brings to our business. It goes beyond the immediate revenue generated by a customer and considers multiple dimensions, including revenue, tenure, and site visits. We shall explore these dimensions below:

1. Revenue Contribution: CLV takes into account the total revenue generated by a customer over their entire relationship with Pens and Printers. This includes revenue from the initial purchase of the new office stationery line and any subsequent transactions.
2. Tenure: The duration for which a customer remains engaged with our brand is a crucial factor in CLV. Longer tenure implies sustained loyalty and repeated interactions, contributing to higher overall value.
3. Site Visits: The frequency of a customer's visits to our site reflects their ongoing engagement and interest in our offerings. Site visits can be indicative of potential future purchases and brand advocacy, both of which contribute to CLV.

​

Interpretation: A higher CLV indicates that a customer is not only contributing significant revenue but is also remaining engaged with Pens and Printers over an extended period, making frequent site visits.

Significance for Business Strategy: Monitoring CLV is crucial for strategic decision-making. It helps us identify high-value customers, tailor marketing efforts to retain and engage them, and optimize our sales approaches for long-term success.

In the context of our current business case, CLV will serve as a key metric to evaluate the effectiveness of different sales methods and guide decisions on resource allocation and customer relationship management.

* 1. **Monitoring Strategy**

Monitoring Customer Lifetime Value (CLV) is a continuous and dynamic process that involves regular tracking, analysis, and strategic adjustments. Here's a comprehensive guide on how the business should monitor CLV to achieve its goals:

1. Establish Baseline CLV: Utilize historical data to calculate CLV for different customer segments and sales methods, providing insights into variations.
2. Regular Tracking: Implement a system for real-time or periodic tracking of customer interactions, purchases, and site visits using CRM tools and analytics platforms.
3. Segmentation for Targeted Analysis: Segment customers based on CLV to identify high-value and low-value segments, enabling targeted analysis and tailored strategies for each group.
4. Integration with Marketing Efforts: Align CLV monitoring with marketing efforts, evaluating the impact of campaigns on CLV and adjusting strategies based on observed outcomes.
5. Customer Retention Strategies: Implement retention strategies based on CLV insights, identifying opportunities to enhance customer experience, loyalty programs, and personalized engagement.
6. Cross-Functional Collaboration: Foster collaboration between marketing, sales, and customer service teams, sharing CLV insights across departments to ensure a holistic approach to customer management.
7. Periodic Review and Adjustments: Conduct periodic reviews of CLV metrics, revisiting business strategies and adjusting sales approaches, customer engagement tactics, and resource allocation based on evolving CLV trends.
8. Benchmarking Against Industry Standards: Compare CLV against industry standards to gain a broader perspective on performance, identifying areas for improvement or excellence compared to competitors.
9. Invest in Employee Training: Ensure employees involved in customer interactions are trained to understand and contribute to CLV goals, aligning team incentives with CLV metrics to reinforce a customer-centric approach.

Initial Value estimation based on Current Data

Having values such as

Average Revenue per purchase = 9.28

Average Purchases per year = 2401.11

Average Customer Tenure = 4.97 years

Therefore, estimated value of CLV based on current data is $110,743.03

1. **Summary and Recommendations**
   1. **Summary**

The revenue analysis of Pens and Printers' new office stationery product line reveals a multimodal and right-skewed distribution, indicating distinct customer segments and outliers. With a median revenue of $89.5 and a positive skewness influenced by higher values on the right, the data suggests a heterogeneous market. Customer distribution by sales method highlights Email as the most prevalent (49.77%), followed by Call (33.88%), and Email + Call (17.15%). Quartile analysis provides insights into the spread of revenue, guiding targeted marketing strategies and pricing decisions.

Weekly revenue fluctuations, observed through a line plot, indicate Email consistently leads in revenue generation in the initial weeks following product launch, then declines as the week progresses, while Email + Call revenue rises slowly in the initial weeks following product launch and then surges in the final weeks of product launch. Average revenue and conversion rate analysis identifies the Email + Call method as the most promising, balancing high average revenue, strong conversion rates, and efficient revenue generation per site visit. Average Number of Items Sold and Revenue by Sales Method identifies the Email + Call method with the highest average number of items sold (12.23) and average revenue ($170.88).

Average Years as Customer and Site Visits by Sales Method identifies the Call method as having the highest average years as a customer (5.18) and Email + Call method as having the highest average number of site visits (26.77). Customer geographic analysis indicates regional preferences and market dynamics. Metrics for business monitoring such as average customer tenure of 4.97 years and customer retention rate of 90.19% highlight success in fostering long-term customer relationships. The CLV, estimated at $110,743.03, becomes a pivotal metric, reflecting the long-term value of customers. Recommendations include state-specific marketing strategies, continuous CLV monitoring, and adapting strategies based on data insights for sustained success in the dynamic market landscape.

* 1. **Recommendations**

The recommendations for Pens and Printers encompass several key strategies.

1. Firstly, continuous monitoring of Customer Lifetime Value (CLV) is emphasized as a crucial tool for guiding resource allocation, customer acquisition, and retention strategies. CLV is highlighted as a strategic compass, ensuring decisions align with the goal of maximizing customer lifetime value for sustainable growth.
2. Secondly, optimizing sales methods involves a strategic analysis of revenue and conversion rate performance. This includes prioritizing high-performing methods, refining underperforming ones, and adopting a customer-centric approach with tailored sales methods.
3. Thirdly, customer engagement strategies are recommended to enhance CLV, including personalized promotions, loyalty programs, interactive content, and active social media engagement.
4. Lastly, further investigation and data collection are deemed essential in areas such as customer segmentation, marketing channel effectiveness, customer journey mapping, product performance evaluation, and measuring customer satisfaction. These insights are considered critical for optimizing operations, enhancing customer engagement, and driving sustainable growth in the competitive stationery and printing market. Overall, a data-driven approach is emphasized to empower Pens and Printers to refine strategies, innovate products, and elevate the customer experience for long-term success in the industry.