

Data Analysis Report

1. Data Validation

1.1 Validation and Cleaning Steps

Column: week

- *Validation:* The week column should contain integer values between 1 and 6, as the product was launched 6 weeks ago.
- *Cleaning:* Verified that all values in the week column fall within the expected range. No cleaning was necessary as all data points were correct.

Column: sales_method

- *Validation:* The sales_method column should contain only three unique values: "Email", "Call" and "Email + Call".
- *Cleaning:* Discovered five unique values, including typos: "em + call" and "email." Corrected these typos to standardize the values to "Email + Call" and "Email".

Column: customer_id

- *Validation:* The customer_id should contain unique string identifiers for each customer.
- *Cleaning:* Verified that there are 15,000 unique customer_id values, confirming no duplicates or missing values. No cleaning required.

Column: nb_sold

- *Validation:* The nb_sold column should contain integer values greater than or equal to 0.
- *Cleaning:* Confirmed that all values are non-negative, as expected. No cleaning was necessary.

Column: revenue

- *Validation:* The revenue column should contain non-negative float values.
- *Cleaning:* Identified 1,074 missing values, which exceed the 5% threshold of total entries and therefore require imputation. Since the distribution of revenue was skewed, the median revenue for each combination of sales_method and nb_sold was calculated and used to fill in the missing values.

Column: years_as_customer

- *Validation:* This column should contain integer values between 0 and 40, as the company was founded in 1984 (40 years ago).

- *Cleaning*: Detected and removed entries where years_as_customer exceeded 40 years.

Column: nb_site_visits

- *Validation*: The nb_site_visits column should contain non-negative integer values.
- *Cleaning*: Confirmed that all values are non-negative, as expected. No cleaning was necessary.

Column: state

- *Validation*: The state column should contain fewer than 50 unique values, corresponding to the U.S. states.
- *Cleaning*: Verified that all state abbreviations are valid U.S. states, with 50 unique values detected. No cleaning was necessary.

2. Exploratory Analysis

This section focuses on understanding the characteristics of the data through various visualizations and statistical summaries. The key questions addressed are:

1. **How many customers were there for each approach?**
2. **What does the spread of the revenue look like overall? And for each method?**
3. **Was there any difference in revenue over time for each of the methods?**
4. **Based on the data, which method would you recommend we continue to use?**

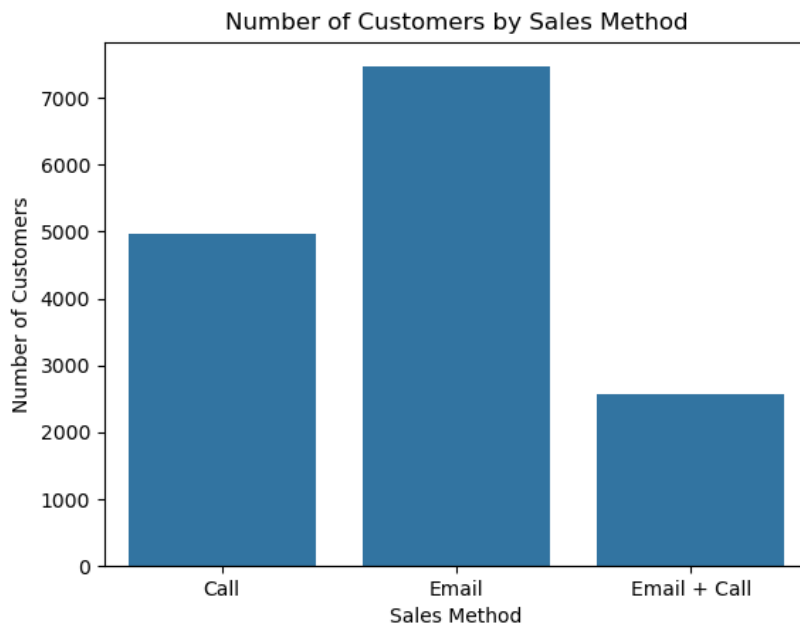
2.1 Number of Customers by Sales Method

To understand the distribution of customers across different sales methods, I calculated the number of unique customers for each method (Call, Email, Email + Call).

- **Findings:**
 - The Email method reached the highest number of customers (7,465).
 - The Call method reached 4,961 customers.
 - The combined method (Email + Call) was used for 2,572 customers.

This distribution indicates that the Email method is the most widely used approach, followed by Call, and lastly, the Email + Call approach.

Figure 1:



2.2 Revenue Distribution Analysis

I analyzed the revenue distribution across the entire dataset and separately for each sales method.

- **Overall Revenue Distribution:**
 - The distribution is slightly right skewed with most revenues concentrated between 50 and 150.
 - There are outliers with revenues exceeding 200.

Figure 2:

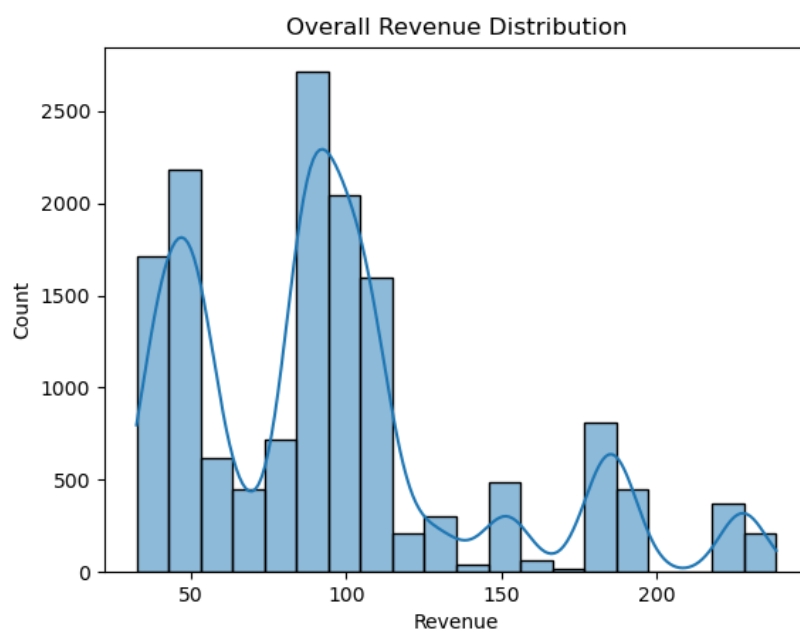
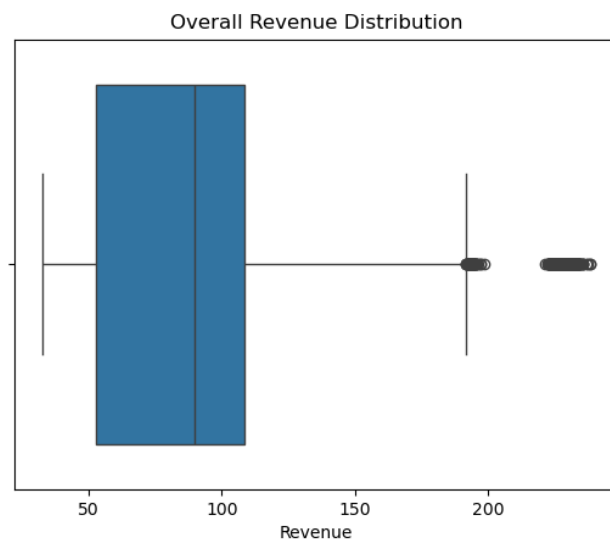


Figure 3:



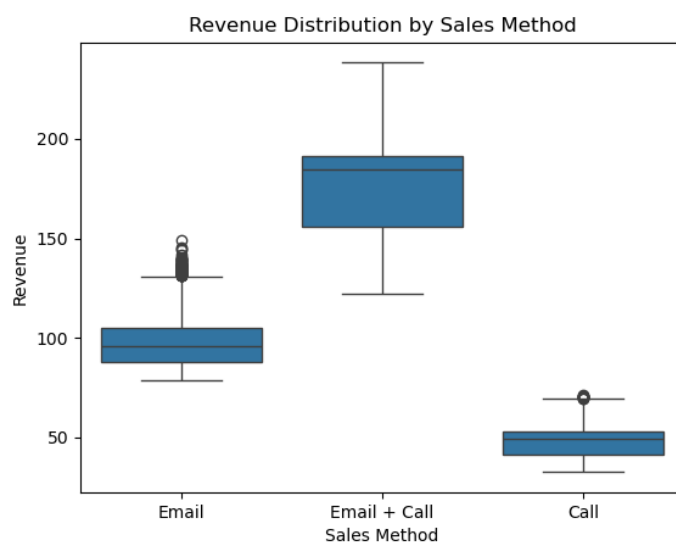
- **Revenue by Sales Method:**

- Email + Call has the highest average revenue per sale, followed by Email, with Call having the lowest.
- The variability in revenue is much higher for the Email + Call method.

Table 1:

sales_method	count	mean
Call	4961.0	47.64
Email	7465.0	97.17
Email + Call	2572.0	184.23

Figure 4:

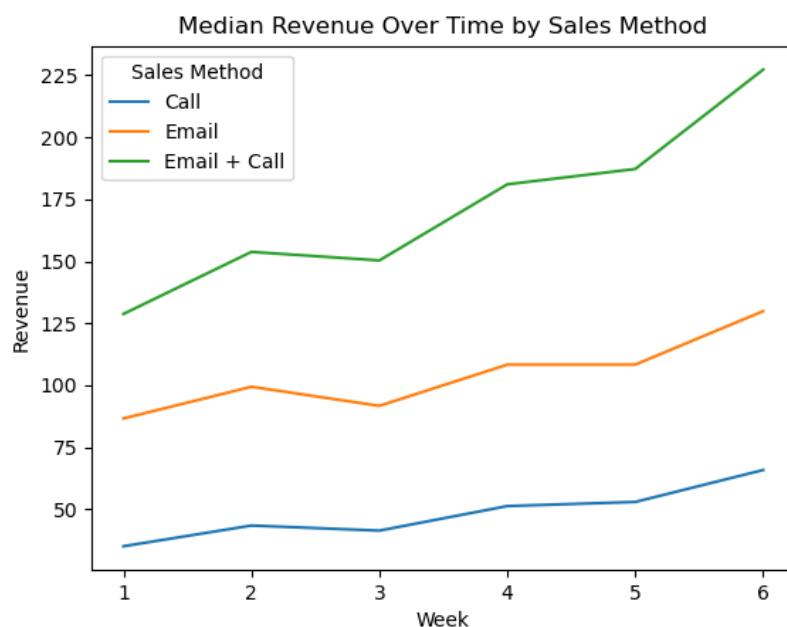


2.3 Revenue Over Time by Sales Method

To understand the revenue over time for different sales methods, I calculated the median revenue for each method and week.

- **Findings:**
 - Email + Call consistently generates higher median revenues compared to the other two methods.
 - The revenue for all methods tends to increase over time with a slight drop at week 3.
 - All methods have the highest revenue at week 6
 - The Call method yields the lowest median revenues during all weeks.

Figure 5:



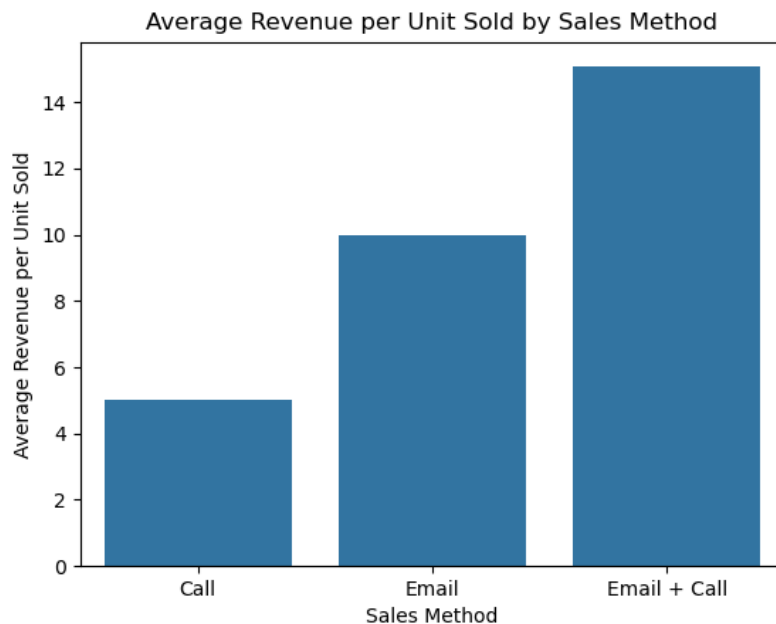
2.4 Revenue per unit sold

Apart from the analysis above, a new column was created to calculate revenue per unit sold, followed by a grouped summary by sales_method, as shown in Table 2.

Table 2:

sales_method	total_revenue	median_revenue	avg_revenue_per_unit	nb_customers
Call	236344.88	49.26	5.01	4961
Email	725343.35	95.79	10.00	7465
Email + Call	473838.05	184.5	15.07	2572

Figure 6:



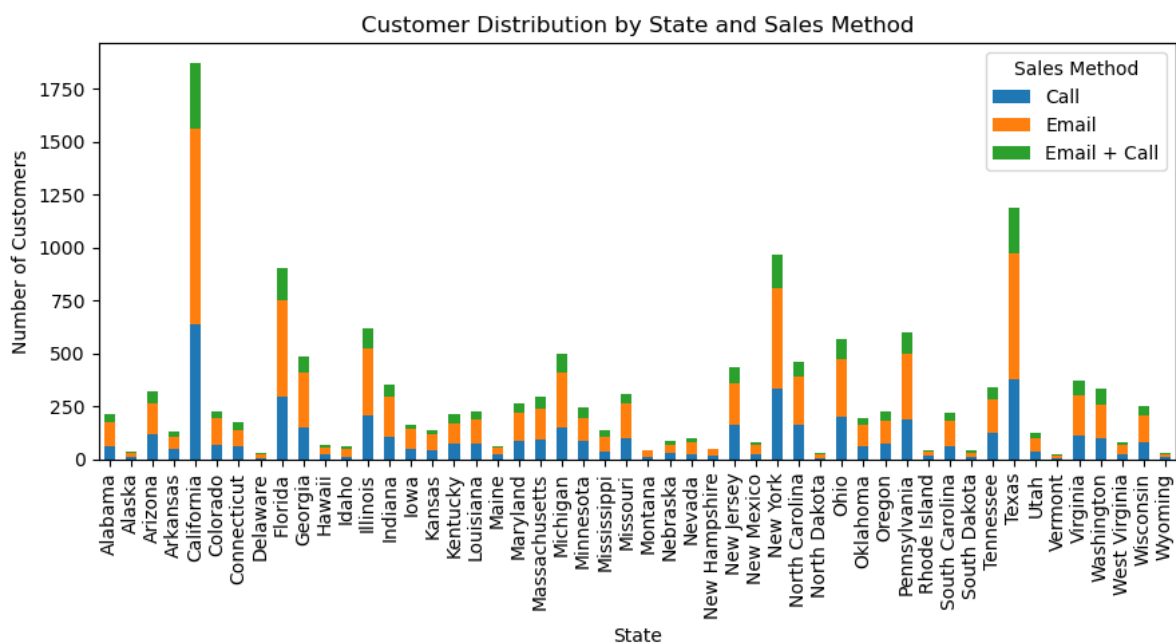
Based on the analysis, the Email + Call method stands out for its higher average revenue per unit sold. Even though it targets fewer customers compared to the Email method, its effectiveness in generating significantly higher revenue makes it a valuable approach for more targeted sales efforts.

2.5 Other Customers Insights

To see the concentration of customers, `customer_id` was grouped by state and `sales_method`. The stacked bar plot allows business to see how customers in each state are distributed across the three sales methods: Call, Email, and Email + Call.

- **Findings:**
 - **California** stands out with the highest total customer count, followed by **New York, Texas, and Florida**. This suggest that these states are key regions for the business.
 - The Email method tends to have the highest proportion of customers among US states.
 - The Email + Call method plays important role in states such as **California, Florida, Texas and Illinois**. This suggests that a multi-channel sales strategy works well in these regions.

Figure 7:



3. Metric Definition for Business Monitoring

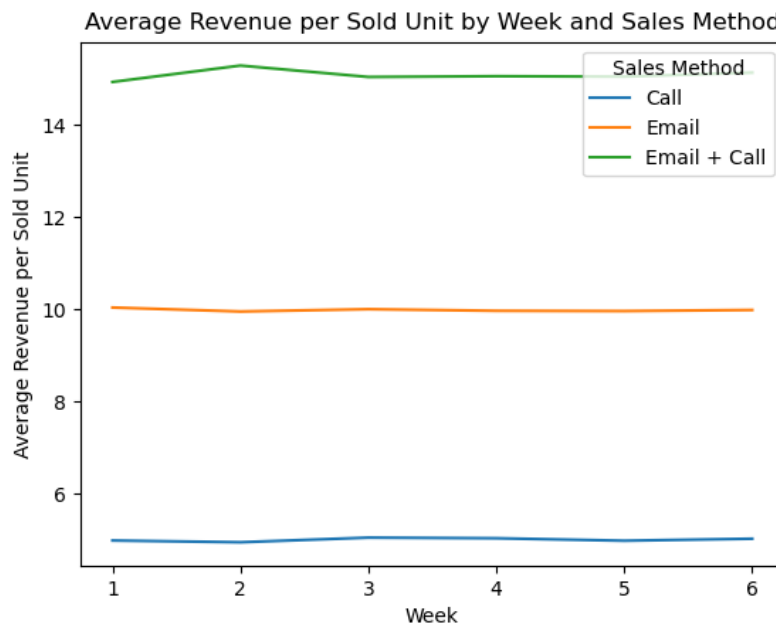
3.1 Metric Definition

- **Metric Name:** Average Revenue per Unit Sold
- **Description:** The chosen metric should give business insight into the effectiveness and profitability of each sales method. **Revenue per Unit Sold** helps the business monitor how much value they are getting from each product sold, which is critical for understanding which sales method delivers the most profit per unit sold, allowing the company to focus resources on the most effective methods.
- **Calculation:** $\text{revenue} / \text{nb_sold}$

3.2 Metric Application

- **How should the business use this metric:**
 1. **Track Over Time:** Measure this metric at regular intervals (e.g. weekly) to observe trends and changes.
 2. **Compare by Sales Method:** Break down the metric by sales_method to understand which method yields the highest return.

Figure 8:



- **Initial Value Estimate:** To estimate the initial values, we can calculate the average of revenue_per_unit for each sales_method:

Table 3:

sales_method	avg_revenue_per_unit
Call	5.01
Email	10.00
Email + Call	15.07

These initial values can serve as a baseline for the business to track how the revenue per unit evolves over time as they optimize their sales tactics

4. Final Summary and Recommendations

4.1 Summary of Key Findings

- **Revenue Differences by Method:** The analysis shows significant differences in revenue generation between the sales methods. The Email + Call method yields the highest average revenue per unit sold, indicating that combining both approaches is more effective for generating higher revenue. The Email method, while less effective in terms of revenue per unit, reaches the largest number of customers and is cost-efficient. The Call method, despite its higher personal engagement, results in the lowest average revenue per unit sold and is the least cost-effective.

- **Customer Distribution Insights:** The Email method has the broadest reach, while the Email + Call method is notably more effective in regions like California, Florida, Texas, and Illinois, suggesting that a multi-channel approach is particularly successful in these areas.
- **Revenue Trends Over Time:** Revenue generally increases over time for all methods, with the highest revenue achieved by week 6. The Email + Call method consistently shows higher median revenue, though there is a slight drop in revenue around week 3 for all methods, which may require further investigation.

4.2 Recommendations

- **Focus on the Email + Call method** for higher-value customers or when a more personalized sales approach is required. This method generates the highest average revenue per unit sold, and the time investment of 10 minutes per customer call is justified by the significantly better returns.
- **Continue prioritizing the Email method** for mass outreach. It reaches the largest number of customers with minimal effort from the sales team while still generating a reasonable average revenue per unit sold. This method is cost-effective for maintaining broad customer engagement.
- **Phase out the Call method** as it is not cost-effective. The 30-minute time investment per customer results in the lowest average revenue per unit sold. Instead, shift these efforts toward the Email + Call method, which produces nearly three times the revenue with only one-third of the time investment.

4.3 Next Steps for the Business

- **Implement the Recommended Strategies:** Allocate resources to enhance the Email + Call method, especially in regions where it has shown strong performance. Adjust outreach plans to leverage the Email method for broader customer engagement while reducing reliance on the Call method.
- **Monitor Performance:** Establish a regular review process to track the Average Revenue per Unit Sold metric.
- **Investigate Revenue Trends:** Explore the reasons behind the revenue drop in week 3 and identify any external factors that may have influenced this trend.