

## Product Sales Data Analysis: Optimizing Sales Strategies

 Felix Odhiambo

 Contact: [otienofelix@live.com](mailto:otienofelix@live.com)

 LinkedIn: <https://www.linkedin.com/in/felixoodhiambo/>

## Data Validation

The dataset contains 15,000 rows and 8 columns before cleaning and validation. I have validated all the columns against the criteria in the dataset table:

- week: 6 weeks indicating the week sale was made. No cleaning is needed.
- sales \_ method: Character values but contains four sales methods instead of three. 23 em + call replaced with 23 Email + Call, leaving 3 sales methods used by Pens and Printers. 7 "email" values were also replaced by "Email."
- customer\_id: Character values, with 15,000 unique ID for each customer. No cleaning is needed.
- nb\_sold: Numeric values representing the number of new products sold. No cleaning is needed.
- revenue: Numeric and Character values. Numeric values are rounded to two decimal places. Character values (NA) were removed.
- years\_as\_customer: Numeric values. Two values are more than the years the company has existed and were removed.
- nb\_site\_visits: Numeric values. No cleaning is needed.
- state: Character values representing customer location. No cleaning is needed.

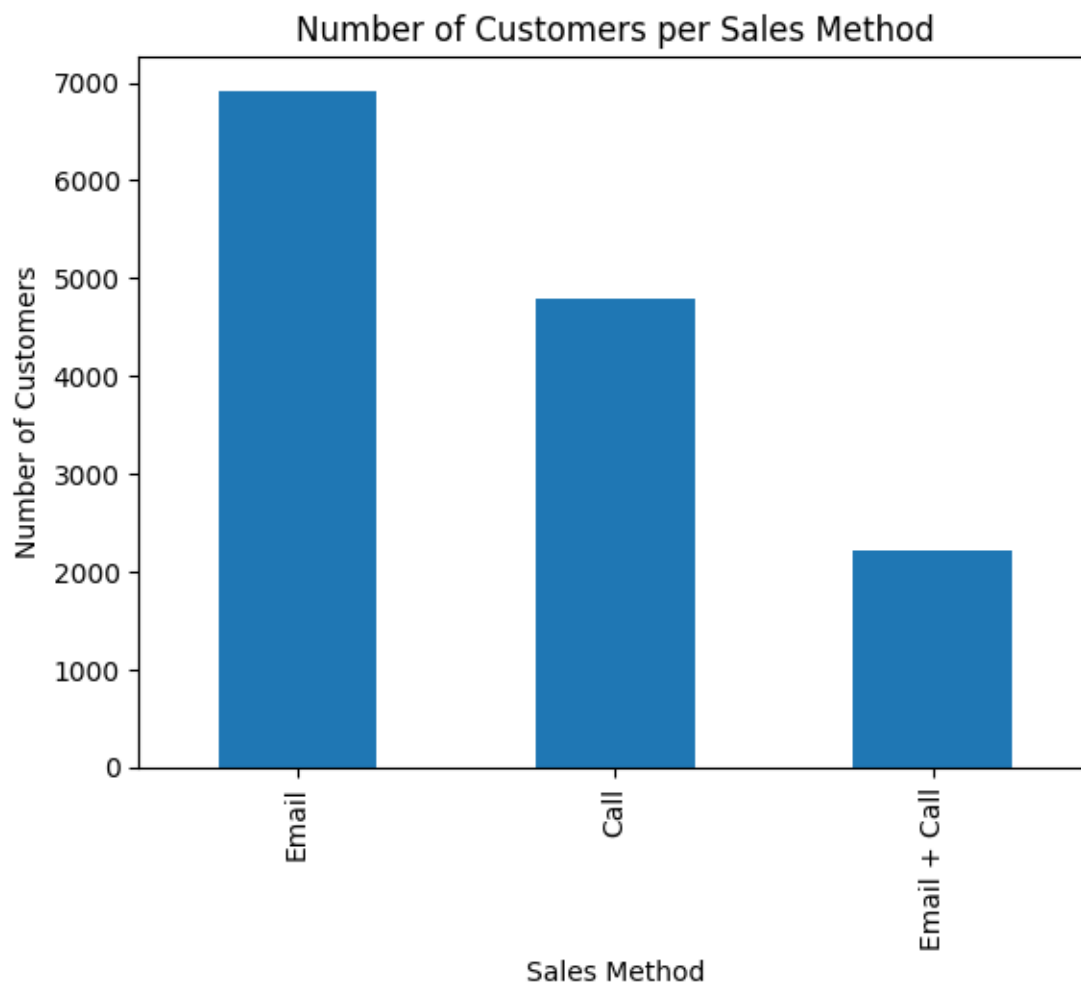
After the data validation, the dataset contains 13,924 rows and 8 columns.

## Exploratory Analysis

### How many customers were there for each approach?

From the last 6 weeks' sales, the number of customers was highest for the Email approach and lowest for the Email + Call approach. We can conclude that the Email approach can be used to continue reaching more customers. Since the number of customers reached using Call was more than half of those reached by Email, it can also be used (Figure 1).

Figure 1: Number of Customers per Sales Method



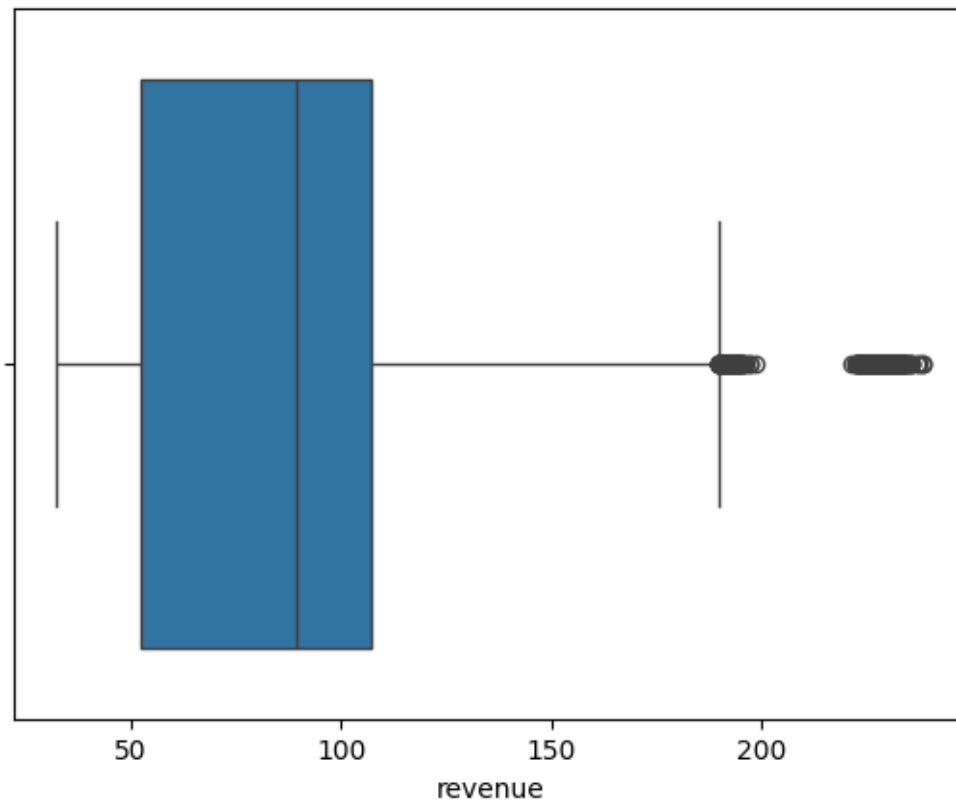
What does the spread of the revenue look like overall? And for each method?

The box plot reveals that the median revenue is 89.5. The upper and lower quartiles are 107.33 and 52.47. The interquartile range is 54.86, meaning the revenue values vary significantly. There are also whiskers extending from the box plot, suggesting potential outliers in the

### Overall Revenue Distribution

The overall revenue boxplot shows a median of 89.50, with an interquartile range (IQR) between 52.47 (Q1) and 107.33 (Q3). The revenue values range from 32.54 (minimum) to 238.32 (maximum), with potential outliers extending beyond the whiskers. The distribution suggests significant variability in revenue across all sales methods.

Figure 2: Overall Revenue

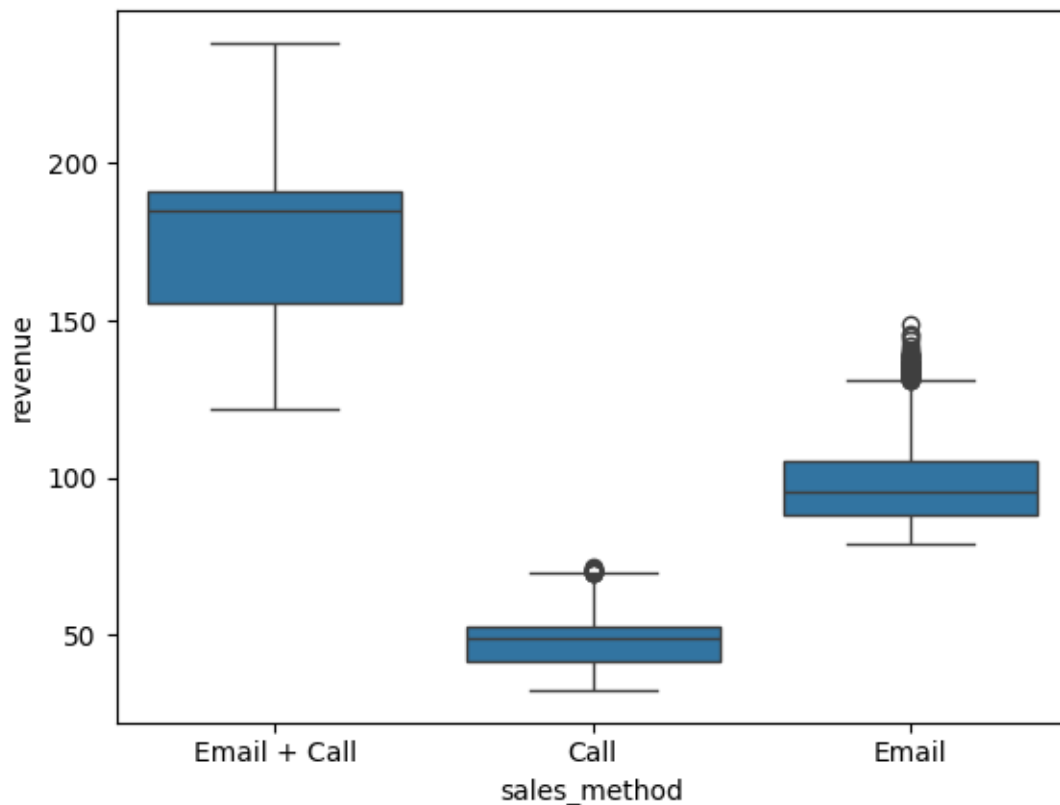


### Revenue Distribution by Sales Method

- Call Method:
  - Median Revenue: 49.05
  - IQR: 41.47 (Q1) to 52.68 (Q3)
  - Range: 32.54 (min) to 71.36 (max)
  - Interpretation: Revenue for this method is lower and more tightly clustered, showing low variation.
- Email Method:
  - Median Revenue: 95.58
  - IQR: 87.88 (Q1) to 105.17 (Q3)
  - Range: 78.83 (min) to 148.97 (max)
  - Interpretation: Email generates higher revenue than Call, with moderate variability.
- Email + Call Method:
  - Median Revenue: 184.74

- IQR: 155.77 (Q1) to 191.11 (Q3)
- Range: 122.11 (min) to 238.32 (max)
- Interpretation: This method produces the highest revenue, but the wider IQR suggests greater variability.

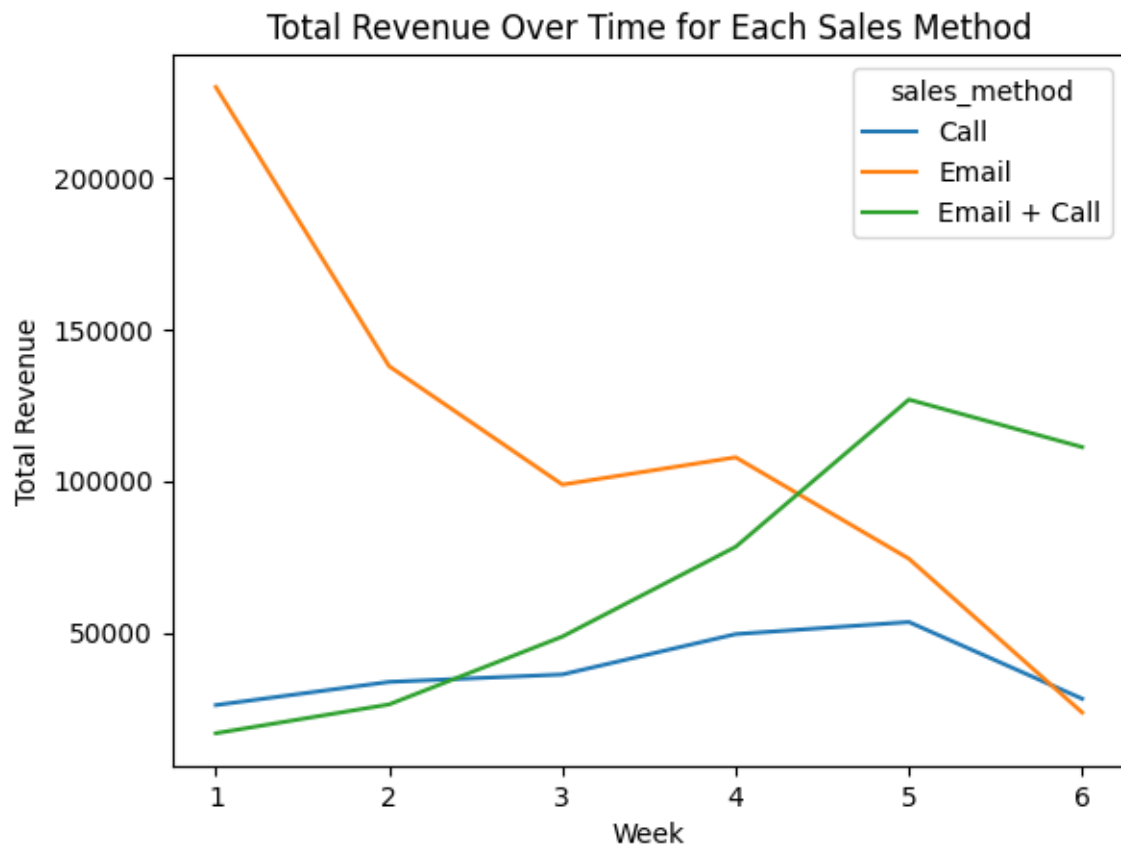
Figure 3: Revenue by Sales Method



### Was there any difference in revenue over time for each of the methods?

The total revenue for the call method showed a steady increase in the first few weeks, reaching its peak in week 5, but dropped significantly in week 6. The email method began with higher revenue in week 1 and declined significantly over the 6 weeks. The Email + Call method had moderate revenue in week 1 but showed a substantial increase, reaching its peak in week 5 and slightly dropping in week 6.

Figure 4: Revenue Trend by Sales Method



## Business Metrics

Given the information about the sales approaches and revenue spread, I recommend using the “Customer Acquisition Rate” and “Revenue Growth” to monitor. Customer Acquisition Rate will help assess the effectiveness of each approach in reaching and converting customers. The Revenue Growth metric will track the total revenue generated by each sales method over time to provide insights into the financial impact of each approach.

Based on the 6 weeks' data, the Email approach reached most customers. Email also generated the highest revenue. However, the financial impact suggests that the Email + Call approach generates more revenue with time. Therefore, the Email approach is the best approach to increase the number of customers. However, the primary purpose of a business is to generate revenue, making the Email + Call approach viable if its variability is addressed.

## Recommendations

1. Using key metrics, monitor whether the increased customer acquisition reflects on the revenue generated. Analyze trends to identify the approach that performs better consistently with customer acquisition and revenue generation.
2. Focus on top performers. Since the Email approach has consistently acquired more customers and revenue, and Email + Call showed a steady increase in revenue, prioritize these methods for customer acquisition and revenue growth.
3. Evaluate how each method performs in terms of revenue generation and conversion rates. Ensure that an increase in the number of customers aligns with revenue growth.
4. Data collection for in-depth analysis: Expand data collection to understand customer behavior better. Consider collecting data on customer preferences, response times, and engagement patterns to help refine sales approaches.
5. Refine email content and call scripts for Email and Email + Call approaches to maximize engagement, conversions, and revenue.