



INTRODUCTION

in...	...	↑↓	w...	...	↑↓	sales_method	...	↑↓	customer_id	...	↑↓	n...	...	↑↓	r...	...	↑↓
		0			2	Email			2e72d641-95ac-497b-bbf8-4861764a7097			10		NA			
		1			6	Email + Call			3998a98d-70f5-44f7-942e-789bb8ad2fe7			15		225.47			
		2			5	Call			d1de9884-8059-4065-b10f-86eef57e4a44			11		52.55			
		3			4	Email			78aa75a4-ffeb-4817-b1d0-2f030783c5d7			11		NA			
		4			3	Email			10e6d446-10a5-42e5-8210-1b5438f70922			9		90.49			
		5			6	Call			6489e678-40f2-4fed-a48e-d0dff9c09205			13		65.01			
		6			4	Email			eb6bd5f1-f115-4e4b-80a6-5e67fcfbfb94			11		113.38			
		7			1	Email			047df079-071b-4380-9012-2bfe9bce45d5			10		99.94			
		8			5	Email			771586bd-7b64-40be-87df-afe884d2af9e			11		108.34			
		9			5	Call			56491dae-bbe7-49f0-a651-b823a01103d8			11		53.82			
		10			3	Email			c40f2602-8a7c-429e-bf13-cb1ec9e5f92f			9		89.49			
		11			2	Call			c20ab049-cbac-4ba7-8868-310aa89e0549			9		45.42			
		12			5	Call			0b026b91-fe12-4af0-86f9-387ba81c8fdb			11		53.42			
		13			2	Email			6103bcac-9da6-4000-a0ce-fa2615cce846			10		101.54			
		14			5	Call			96c8b5b8-cb81-4c75-a284-0e0026a03be8			10		51.87			
		15			4	Email			189d4f1b-9e76-4f64-9e71-7bd9b133a2d1			10		104.22			
		16			2	Email			0f744f70-1588-4e0c-8865-fd9acc7f6dda			10		NA			

Rows: 12,500 ⚠ truncated from 15 000 rows

in...	...	↑↓	we...	...	↑↓	sales_method	...	↑↓	customer_id	...	↑↓	nb...	...	↑↓	re...
		0			2	Email			2e72d641-95ac-497b-bbf8-4861764a7097					10	NA
		1			6	Email + Call			3998a98d-70f5-44f7-942e-789bb8ad2fe7					15	225.47
		2			5	Call			d1de9884-8059-4065-b10f-86eef57e4a44					11	52.55
		3			4	Email			78aa75a4-ffeb-4817-b1d0-2f030783c5d7					11	NA
		4			3	Email			10e6d446-10a5-42e5-8210-1b5438f70922					9	90.49

Rows: 5

in...	...	↑↓	w...	...	↑↓	sales_method	...	↑↓	customer_id	...	↑↓	n...	...	↑↓	r...	...	↑↓
		0			2	Email			2e72d641-95ac-497b-bbf8-4861764a7097			10		NA			
		1			6	Email + Call			3998a98d-70f5-44f7-942e-789bb8ad2fe7			15		225.47			
		2			5	Call			d1de9884-8059-4065-b10f-86eef57e4a44			11		52.55			
		3			4	Email			78aa75a4-ffeb-4817-b1d0-2f030783c5d7			11		NA			
		4			3	Email			10e6d446-10a5-42e5-8210-1b5438f70922			9		90.49			
		5			6	Call			6489e678-40f2-4fed-a48e-d0dff9c09205			13		65.01			
		6			4	Email			eb6bd5f1-f115-4e4b-80a6-5e67fcfbfb94			11		113.38			
		7			1	Email			047df079-071b-4380-9012-2bfe9bce45d5			10		99.94			
		8			5	Email			771586bd-7b64-40be-87df-afe884d2af9e			11		108.34			
		9			5	Call			56491dae-bbe7-49f0-a651-b823a01103d8			11		53.82			
		10			3	Email			c40f2602-8a7c-429e-bf13-cb1ec9e5f92f			9		89.49			
		11			2	Call			c20ab049-cbac-4ba7-8868-310aa89e0549			9		45.42			
		12			5	Call			0b026b91-fe12-4af0-86f9-387ba81c8fdb			11		53.42			
		13			2	Email			6103bcac-9da6-4000-a0ce-fa2615cce846			10		101.54			
		14			5	Call			96c8b5b8-cb81-4c75-a284-0e0026a03be8			10		51.87			
		15			4	Email			189d4f1b-9e76-4f64-9e71-7bd9b133a2d1			10		104.22			
		16			2	Email			0f744f70-1588-4e0c-8865-fd9acc7f6dda			10		NA			

Rows: 100

In this report, I will outline the process and findings from my analysis of the sales data provided. The report will cover several key areas to address the questions posed by the sales team and provide actionable insights.

First, I will perform **data validation** by detailing the steps I took to clean and validate each column in the dataset, ensuring its accuracy and reliability for analysis.

Next, I will conduct an **exploratory analysis** to answer the key questions raised by the sales team. This will include visualizing the data in two different types of graphs that focus on individual variables and at least one graph that examines the relationship between multiple variables. I will describe my findings from this analysis, offering insights into patterns and trends that emerge.

I will also define a **metric for the business to monitor**, suggesting how this metric should be tracked moving forward to align with business goals. I will estimate the initial value(s) of this metric based on the current data to provide a benchmark for future monitoring.

Finally, I will conclude with a **summary of my findings and recommendations** for the business to consider in order to optimize their strategies based on the data analysis. These recommendations will be grounded in the insights gained throughout the process, with the goal of helping the sales team make data-driven decisions moving forward.

Data Validation

We need to convert the data type of the 'revenue' column to numeric in order to accurately count the number of 'null' values.

Now, after these operations, we can count null values from 'revenue' column.

As you can see there are 1074 null values in 'revenue' column.

index	...	↑↓	week	...	↑↓	nb_sold	...	↑↓	revenue	...	↑↓	↓
count						15000			15000			13926
mean						3.0982666667			10.0846666667			93.9349425535
std						1.6564198071			1.8122133327			47.4353122457
min						1			7			32.54
25%						2			9			52.47
50%						3			10			89.5
75%						5			11			107.3275
max						6			16			238.32
Rows: 8												

Unfortunately, there are some unwanted values such as 'em + call' and 'email'. We need to clean these up.

We will analyze whether the presence of NaN values in the revenue column impacts the mean revenue for each sales method type.

I would like to replace NaN's with the overall mean of the 'revenue' column. This operation could help maintain consistency across the data.

As you can see, there is a significant difference in mean for 'Email + Call' type.

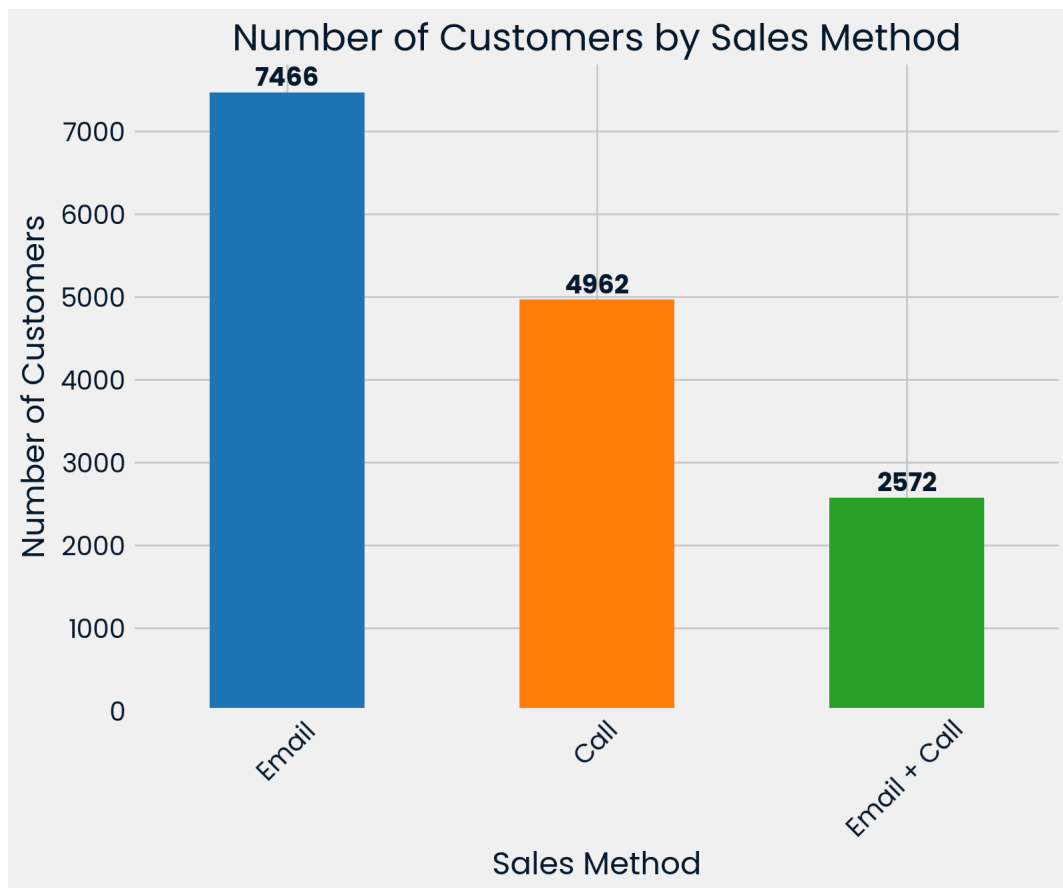
Company was founded in 1984 so there is no chance that some customer has been buying from as for 63 years now.

Your query ran successfully but returned no results.

Exploratory Analysis

Number of customers for each sales method

```
sales_method
Email      7466
Call       4962
Email + Call 2572
Name: count, dtype: int64
```

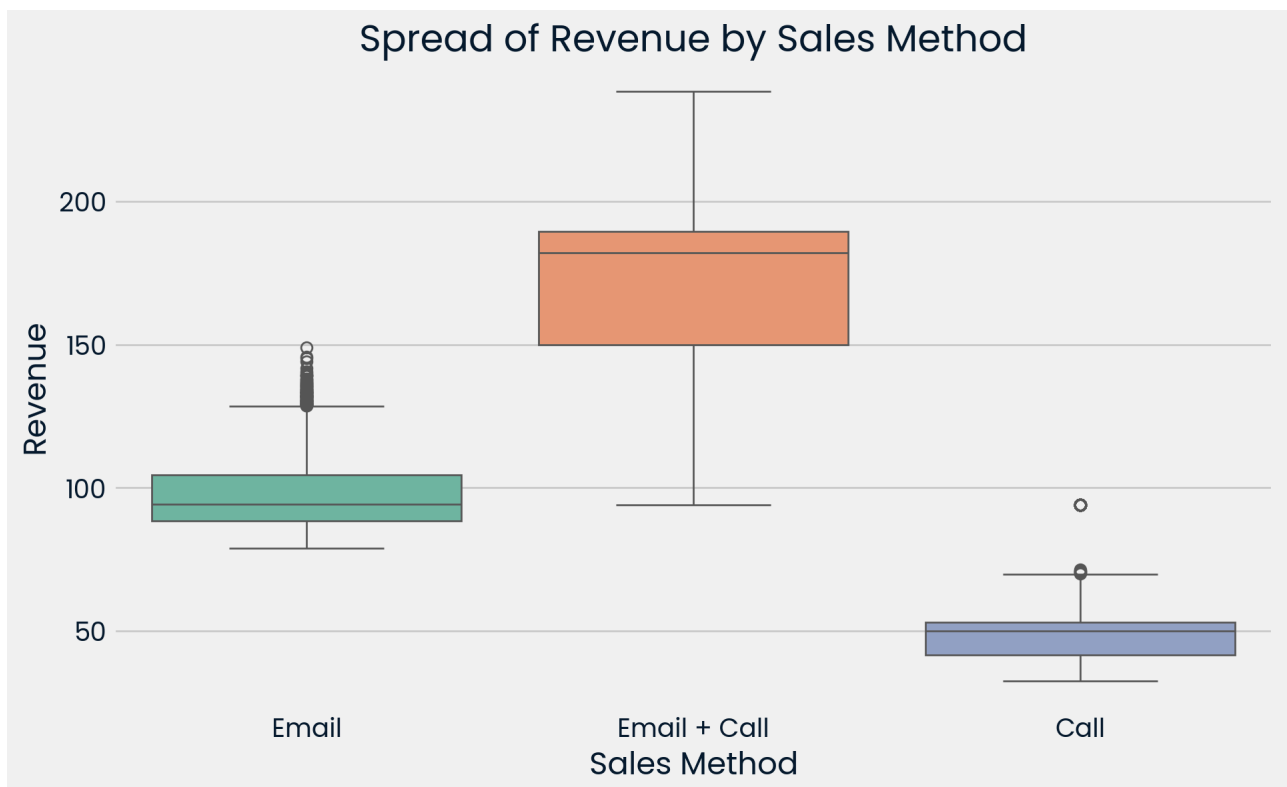
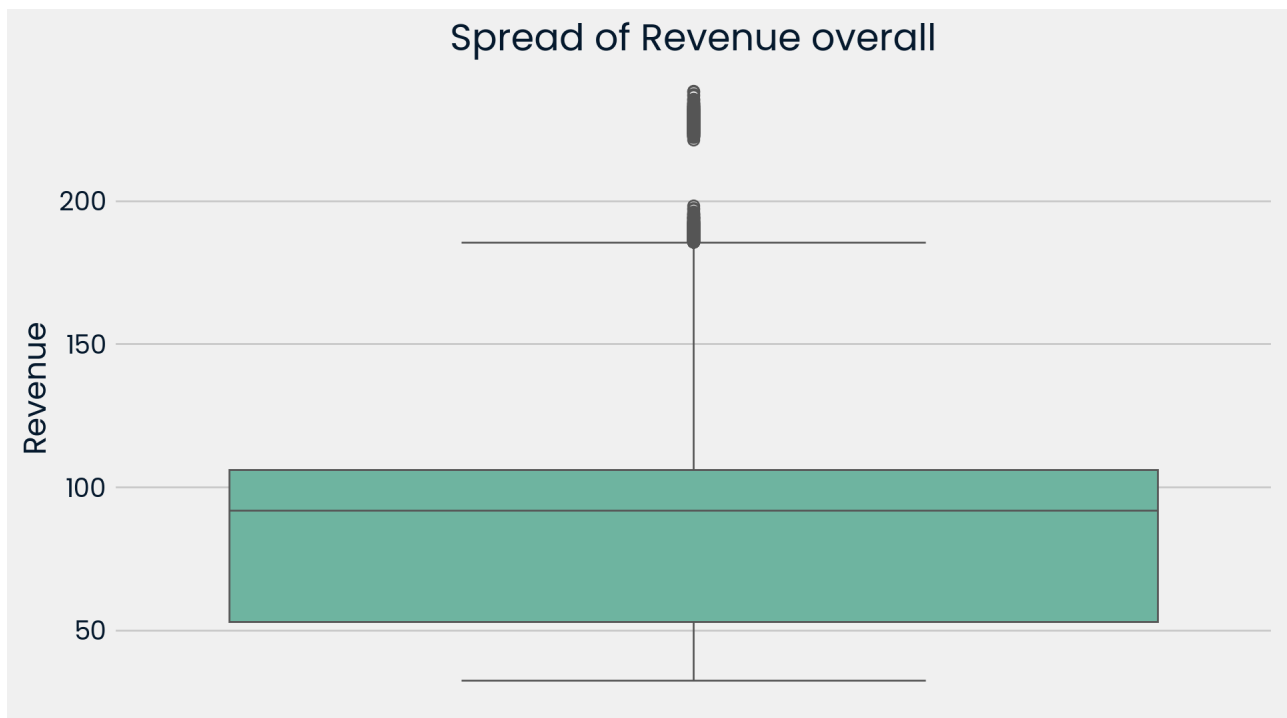


During the 6-week period, the most popular sales method was Email, with a total of 7,466 customers. This was followed by the Call method, which engaged 4,962 customers, and the Email + Call method, which had 2,572 customers.

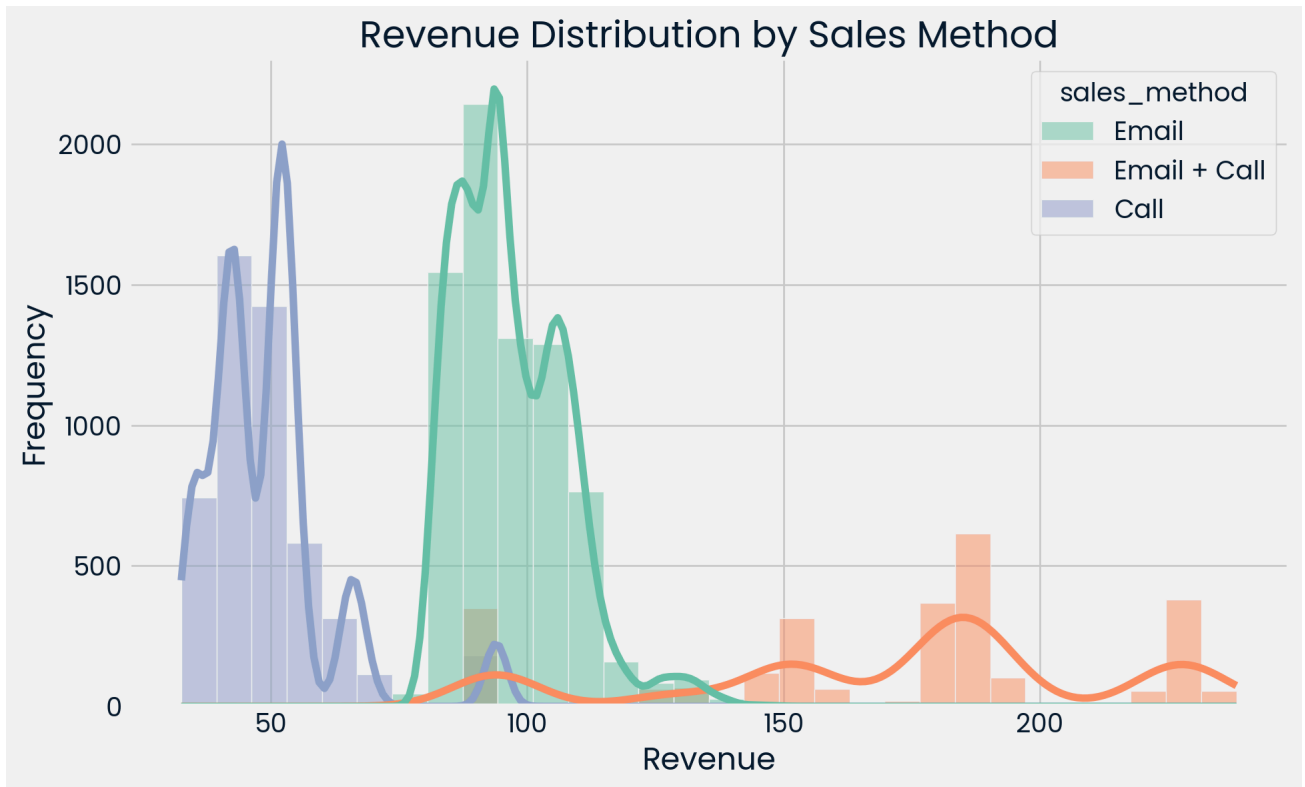
What is the spread of revenue overall and for each sales method?

Calculating the Spread of Revenue Overall and by Sales Method

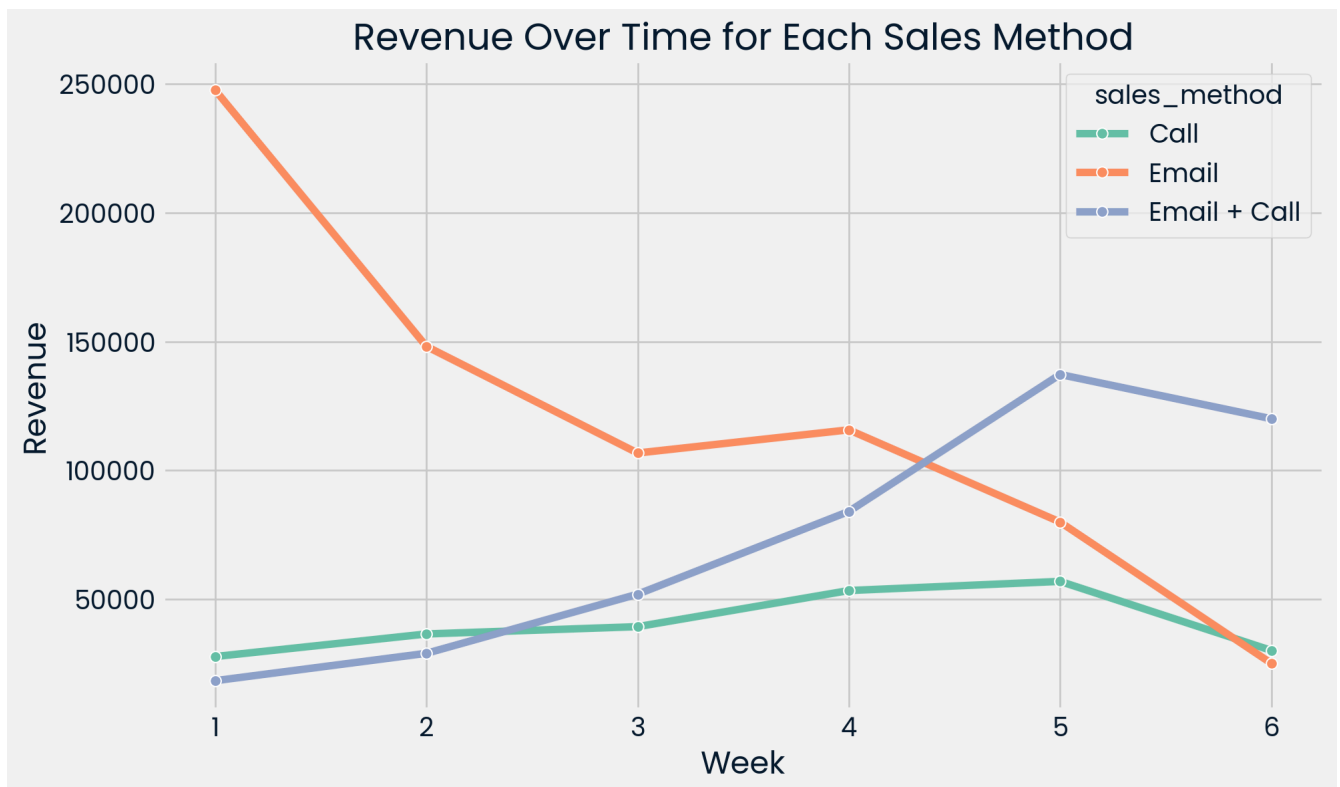
Visualizing the Spread using Box Plot



Visualizing the Spread using Histogram



Revenue over time for each sales method



...	↑↓	Call	...	↑↓	Email	...	↑↓	Email + Call	...	↑↓
1		27850.0089659629			247519.2541426109			18482.2240234094		
2		36610.0682766049			148130.4787383312			29100.3433340514		
3		39441.0580468189			106776.6101170473			52024.8629893724		
4		53445.3226446934			115772.8401170472			84120.1664383168		
5		56993.7028744794			80009.5465532098			137236.4986234382		
6		30225.5537936234			25209.709080856			120075.8895425822		

Rows: 6

Interpretation

The following observations can be made from the table:

In the first week, the 'Email' sales method generated the highest revenue (247519.25), followed by 'Call' (27850.01) and 'Email + Call' (18482.22). It is noteworthy that the 'Email' method required the least effort from the sales team, making it the most efficient choice during week one.

Over the course of the 6-week period, revenue from the 'Email' method steadily declined, suggesting that the effectiveness of this approach may have diminished over time.

In contrast, the 'Call' method exhibited a general upward trend in revenue, with some fluctuations. However, as the 6 weeks progressed, revenue from 'Call' began to decline toward the end of the period.

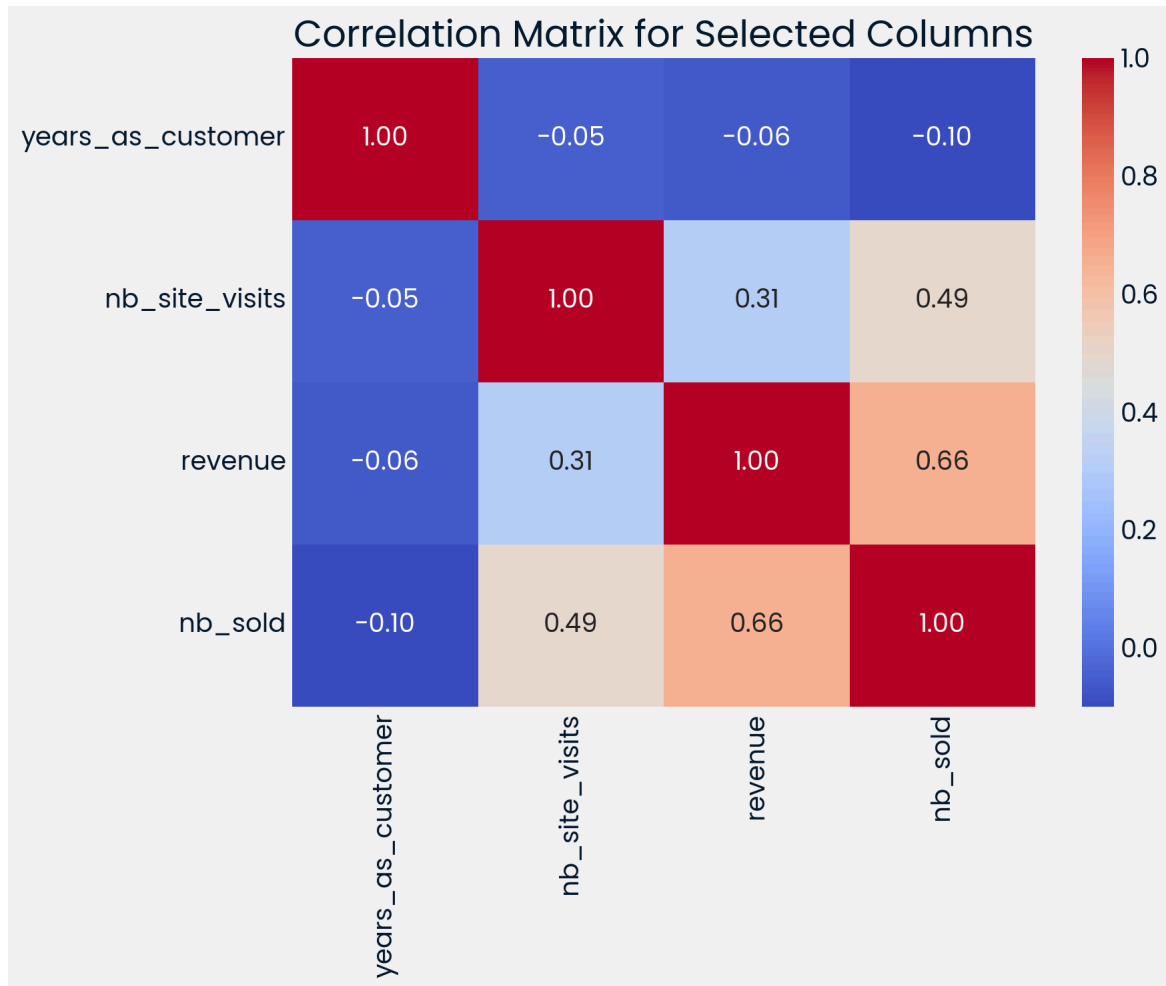
The 'Email + Call' method demonstrated a consistent and healthy increase in revenue throughout the entire 6-week period, indicating its potential for long-term revenue growth compared to the other methods.

In conclusion, the 'Email + Call' method appears to offer the most promise for long-term revenue generation, although it requires more effort from the sales team. While the 'Email' method performs well initially, it shows a steady decline over time, and the 'Call' method, despite an overall upward trend, fluctuates more. When deciding which method to continue using, it is important to weigh the effort required against the potential for sustained revenue growth.

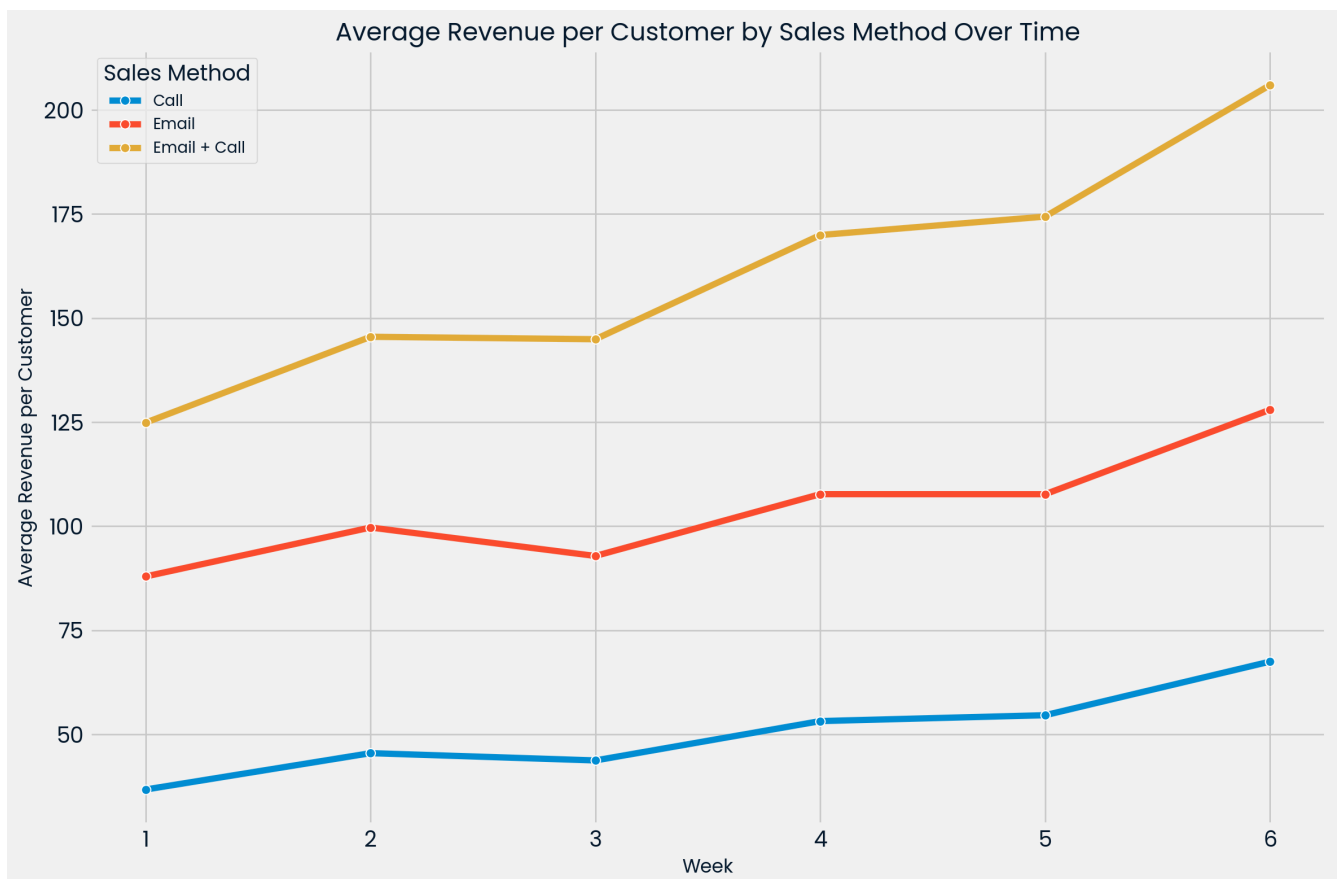
If it were up to me, based on the above observations, I would recommend using the 'Email + Call' method, which appears to strike the optimal balance between the team's effort and the results achieved.

Investigating other differences between customers in each group

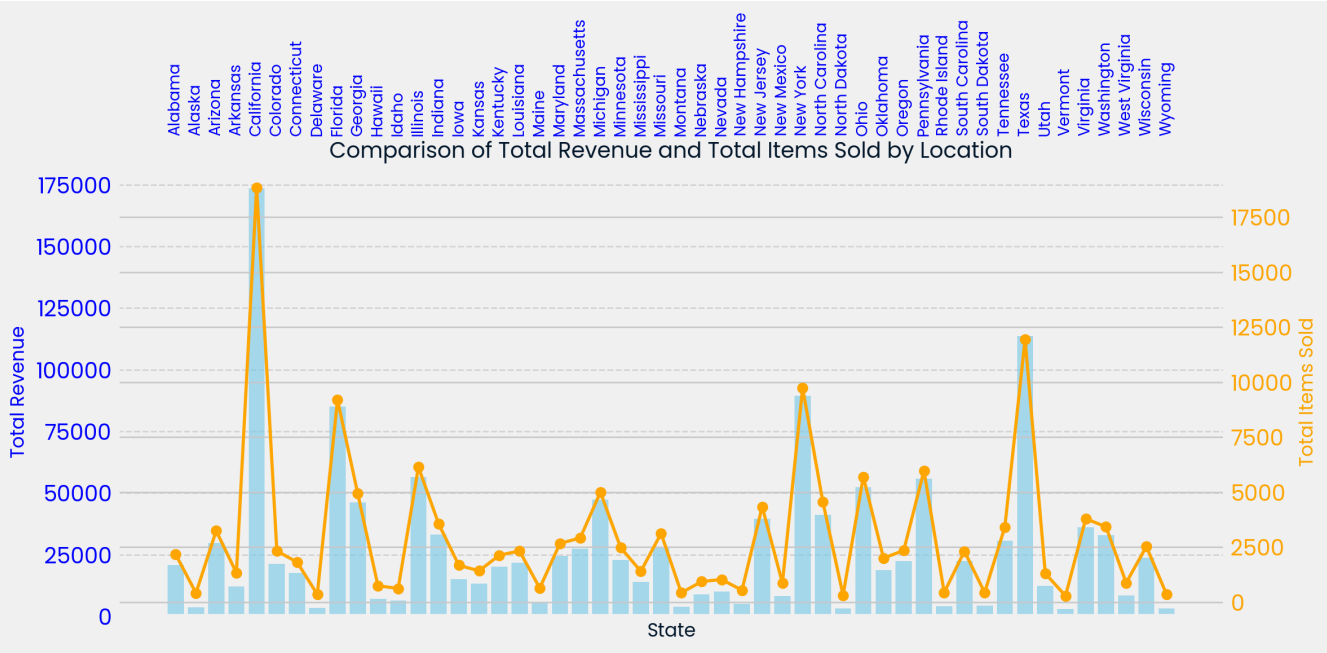
Let's see if any columns are correlated to each other.



As expected, there is a strong correlation between the number of items sold (nb_sold) and revenue, which is natural since an increase in the number of products sold leads to higher revenue. Interestingly, and less obvious, there is also a moderate yet noticeable correlation between the number of items sold (nb_sold) and the number of site visits (nb_site_visits), suggesting that as website traffic increases, there is also a tendency for purchases to rise.



Interestingly, regardless of the sales method chosen, the average revenue per customer within each method shows a similar trend, which is upward.



...	↑↓	state	...	↑↓	total_revenue	...	↑↓	total_...	...	↑↓
0		Alabama			20711.079310642			2161		
1		Alaska			3669.6647127675			412		
2		Arizona			29643.4885063909			3238		
3		Arkansas			12110.469195749			1327		
4		California			173534.2373596151			18859		
5		Colorado			21236.3089659629			2322		
6		Connecticut			17479.869310642			1816		
7		Delaware			3438.3645978745			350		
8		Florida			84978.6755191728			9201		
9		Georgia			46150.5233340514			4930		
10		Hawaii			6995.3247127675			740		

Rows: 50

Average nb_sold and nb_site_visits by sales_method.

Metric for the Business to Monitor

Metric Name: **Revenue per Customer per Unit of Effort RCUE**

Formula: The metric will be Revenue per Customer per Unit of Effort, which incorporates both the revenue generated and the amount of time or effort the sales team spends on each customer.

Effort Factor for each method:

- 1. **Email:** Very little effort. We can assign an effort factor of 0.5.
- 2. **Call:** On average, 30 minutes per customer. We can assign an effort factor of 3 (since calls are more effort-intensive).
- 3. **Email + Call:** Email is quick, but calls take 10 minutes on average. Assign an effort factor of 1, considering that the call still requires time but less than the full call method.

Formula for the Metric:

Revenue per Customer per Unit of Effort =

Total Revenue

Number of Customers × Effort Factor

Where:

- 1. Total Revenue is the sum of revenue for each sales method.
- 2. Number of Customers is the count of unique customers for each sales method.
- 3. Effort Factor is assigned based on the method:

Email = 0.5

Call = 3

Email + Call = 1.

How Should the Business Monitor the Metric?

The business should monitor this metric to evaluate both the revenue generation and the effort required for each sales method. By doing so, they can assess the efficiency of each sales method and allocate resources accordingly.

To calculate the metric, you can use the following steps in Python.

Key Takeaways for the Business:

- 1. Revenue per Customer per Unit of Effort will help the business identify which sales methods are delivering the most value relative to the effort put in.
- 2. Methods that provide a high return on effort are more efficient, and the business may want to focus more resources on these methods.
- 3. The business should track this metric weekly to assess trends and adjust strategies to optimize sales efforts.

As we can see, in the sixth week, there is a significant difference in the RCUE (Revenue per Customer per Unit of Effort) indicator between the Email and Call sales methods, suggesting that the Email method was considerably much more profitable during this time period, however, it is important to note that the Email + Call method recorded a higher average revenue per customer over the 6-week period.

Conclusion and Recommendations

Considering all factors from the analysis, it is recommended to prioritize the **Email + Call** method for the following reasons:

Although the 'Email' method has the highest RCUE, the 'Email + Call' approach generates greater customer engagement, as reflected by higher website visits and an increased average number of items purchased. This engagement could foster stronger customer relationships and support long-term growth.

The 'Email + Call' method has also demonstrated a steady upward trend in revenue generation over the past six weeks, highlighting its potential for sustained success.

In contrast, the 'Call' method is the least efficient in terms of effort and revenue generation, with a declining revenue trend over time.

Additionally, 'Email + Call' has the highest average revenue per customer.

In final conclusion, the 'Email + Call' method should be prioritized for sales efforts, as it offers better customer engagement, consistent revenue growth, and a more sustainable balance between effort and return. Continuous monitoring of key metrics such as customer engagement, website visits, and revenue per customer will be essential to ensure that the chosen sales strategy continues to deliver strong results. By regularly evaluating and refining our approach, we can maintain a competitive edge and drive long-term success.