Defining Objectives and Questions

Pens and Printers was founded in 1984 and provides high quality office products to large organizations. We are a trusted provider of everything from pens and notebooks to desk chairs and monitors. We don't produce our own products but sell those made by other companies. We have built long lasting relationships with our customers and they trust us to provide them with the best products for them. As the way in which consumers buy products is changing, our sales tactics have to change too. Launching a new product line is expensive and we need to make sure we are using the best techniques to sell the new product effectively. The best approach may vary for each new product so we need to learn quickly what works and what doesn't.

New Product Sales Method

Six weeks ago we launched a new line of office stationery. Despite the world becoming increasingly digital, there is still demand for notebooks, pens and sticky notes. Our focus has been on selling products to enable our customers to be more creative, focused on tools for brainstorming. We have tested three different sales strategies for this, targeted email and phone calls, as well as combining the two.

- **Email:** Customers in this group received an email when the product line was launched, and a further email three weeks later. This required very little work for the team.
- **Call:** Customers in this group were called by a member of the sales team. On average members of the team were on the phone for around thirty minutes per customer.
- **Email and call:** Customers in this group were first sent the product information email, then called a week later by the sales team to talk about their needs and how this new product may support their work. The email required little work from the team, the call was around ten minutes per customer.

Questions

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

Data collection

We are using data from the sales teams that includes sales from the past 6 weeks. The information that we have is:

| Column Name	Details
week	Week sale was made, counted as weeks since product launch
sales_method	Character, which of the three sales methods were used for that customer
customer_id	Character, unique identifier for the customer
nb_sold	Numeric, number of new products sold
revenue	Numeric, revenue from the sales, rounded to 2 decimal places.
years_as_customer	Numeric, number of years customer has been buying from us (company founded in 1984)
nb_site_visits	Numeric, number of times the customer has visited our website in the last 6 months
state	Character, location of the customer i.e. where orders are shipped

Data Cleaning and Validation

Summary

- **week:** no cleaning needed
- **sales_method:** there were 5 unique values ('Email', 'Email + Call', 'Call', 'em + call', 'email') when there should have only be 3 ('Email', 'Email + Call', 'Call')
- **customer_id:** every id is unique, no cleaning needed
- **nb_sold:** no cleaning needed
- **revenue:** there were 1074 missing values which were imputed based on the median when grouped by nb_sold and sales_method (the two columns with strong relationships with revenue)
- **years_as_customer:** there were 2 impossible values higher than the number of years since the company was founded in 1984, which were replaced with the maximum possible value (41)
- **nb_site_visits:** no cleaning needed
- **state:** all 50 states are represented in the data, confirmed by checking for typos

Exploratory Analysis

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

The email only method accounted for nearly half of all customers (7466), followed by the call only method with roughly a third of the customers (4962), and the combination method attracted the least customers (2572).
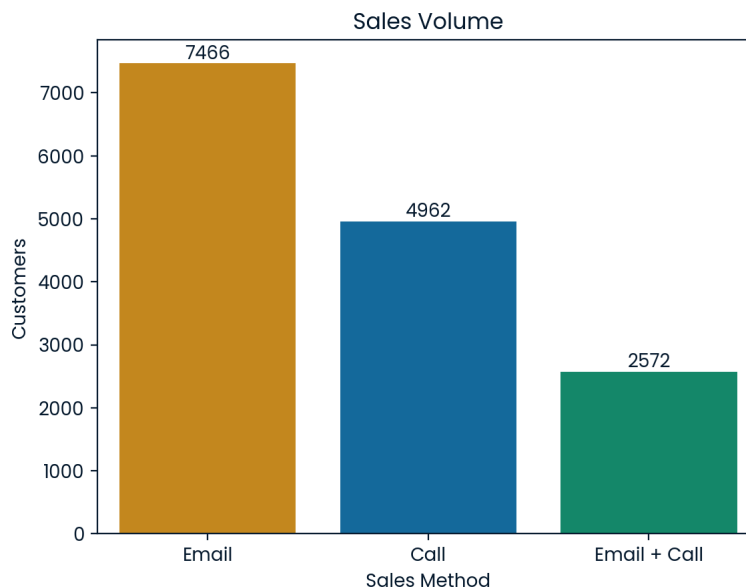
2. What does the spread of the revenue look like overall? And for each method?

The overall distribution of revenue has a very wide spread with strong right skew. The range is 205.78 with a mean of 95.72 and a median of 90. When we use sales method as a color overlay, we can see 3 fairly distinct distributions within the overall distribution with the call only method encompassing revenues below ~75, the email only method encompassing revenues between ~75 and ~140, and the combination method encompassing revenues above ~130, with some overlap between the email only and combination distributions.

Sales Method ··· ↑↓	Revenue Mean ··· ↑↓	Revenue Median ··· ↑↓	Revenue Variance
Call	47.64	49.26	
Email	97.17	95.79	
Email + Call	184.23	184.5	

Rows: 3   ⤢ Expand Table

Overall Revenue Distribution

Revenue Distribution by Sales Method
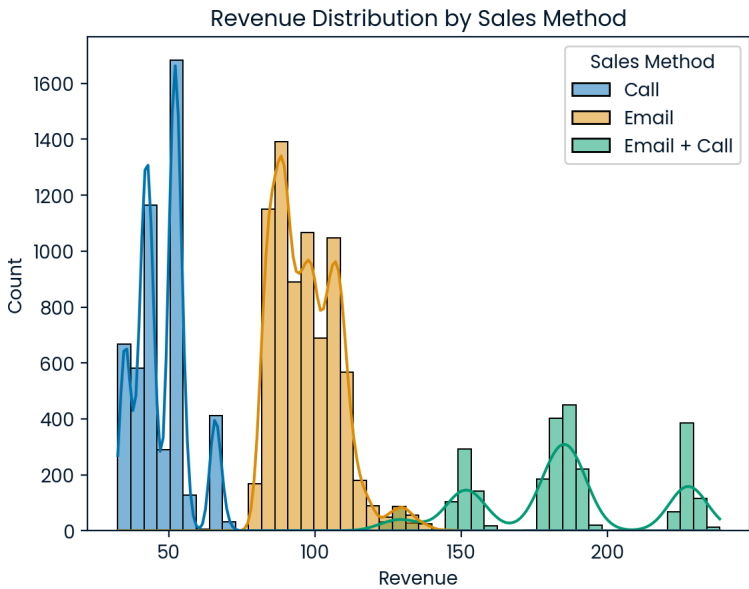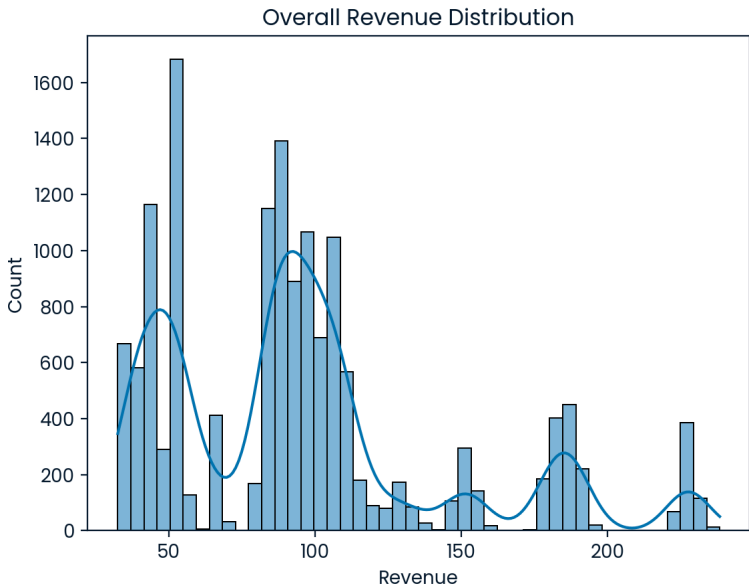
3. Was there any difference in revenue over time for each of the methods?

Important weeks:

- **Email:** Customers received a second email in *Week 3*.
- **Call:** No further action after Week 1 call.
- **Email + Call:** Customers were called in *Week 1*.

We can see that revenue from the email-only method began extremely high in week 1 ($246,357.88). This was followed by a precipitous decline with only week 4 showing a positive change in revenue from the previous week, probably due to the second email being sent. In contrast, both the call only and combination methods accounted for small initial revenues ($26,796.43 and $19,077.76, respectively) followed by steady positive growths, although the combination method showed more positive growth (>60% increase each week). All three methods had their largest negative growths in week 6 with the combination method declining the least.

We know that the email-only method captured the highest total sales volume over the 6 weeks, but when examining sales volume week-by-week, we can see that sales volume, similar to revenue totals fell dramatically week after week, ending as the method with the lowest sales volume in week 6. As a reminder, both the email-only and combination method customers received a product launch email in week 0, which means that until the week 1 phone call, those customers experienced the exact same sales method. Despite the identical week 0 sales methods, the email-only sales method strangely captured 5x as many sales as the combination method. One explanation could be that the sales methods were not evenly applied to the same number of customers.

Controlling for sales volume by calculating weekly averages, we can see that **all** weekly revenue averages actually increased week over week, except for week 3, when all three sales methods experienced a decline. We can also see that despite the lower sales volume, the combination method experienced the highest initial revenue per sale ($128.90) with email-only in second ($87.52) and call-only in last ($35.35). The combination method also experienced the most positive average revenue throughout the 6 weeks.

Growth in Weekly Revenue Total

⋯ ↑↓	Call … ⋯ ↑↓	Email … ⋯ ↑↓	Email + Call Reven… ⋯ ↑↓	Call Revenue % Cha… ⋯ ↑↓	Email Revenue % Cha… ⋯ ↑↓	Email + Call Re
1	26796.43	246357.88	19077.76			
2	35095.39	148764.06	30794.85	30.97	-39.61	
3	37666.24	106781.46	54031.24	7.33	-28.22	
4	51716.34	116924.715	89692.34	37.3	9.5	
5	55479.73	80823.22	147493.3	7.28	-30.88	
6	29641.215	25788.49	132748.56	-46.57	-68.09	

Rows: 6   ⤢ Expand Table

Weekly Revenue Totals
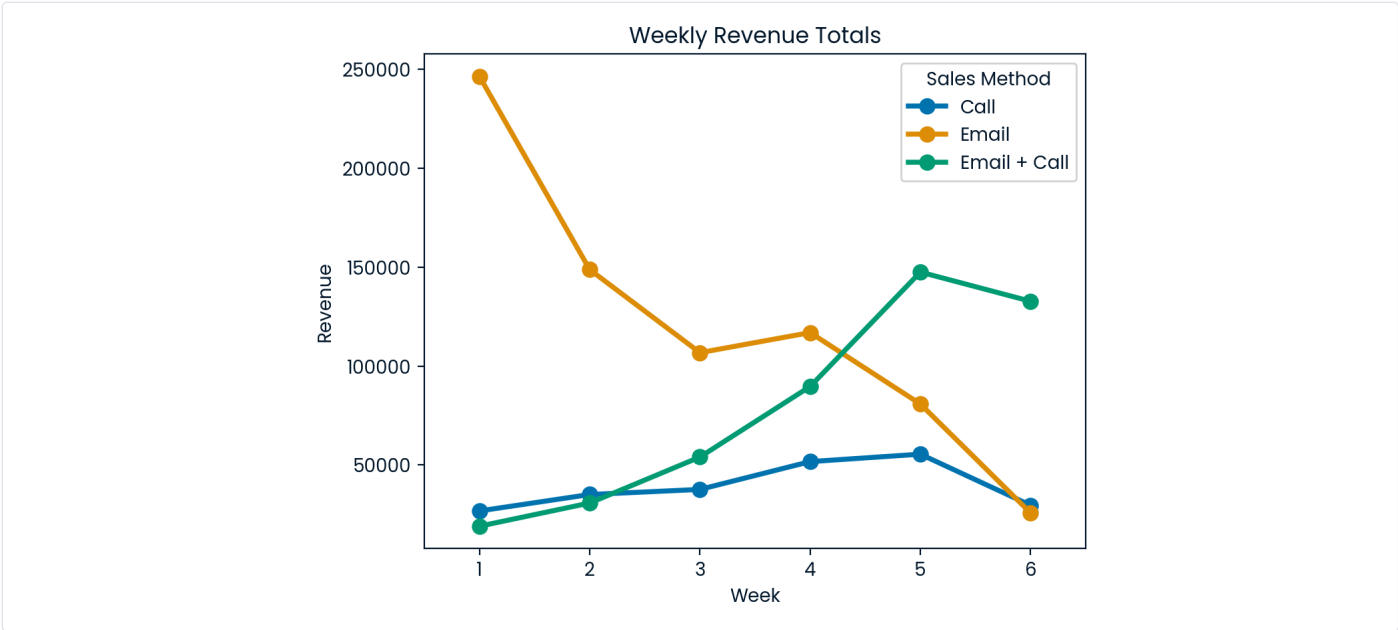
Growth in Weekly Revenue Volume

	Call Sales Vol...	Email Sales Vol...	Email + Call Sales Volume	Call Sales Volume % Change	Email Sales Volume
1	758	2815	148		
2	805	1486	200	6.2	
3	902	1150	359	12.05	
4	1005	1075	495	11.42	
5	1044	743	787	3.88	
6	448	197	583	-57.09	

Rows: 6   ⤢ Expand Table

Weekly Sales Volume
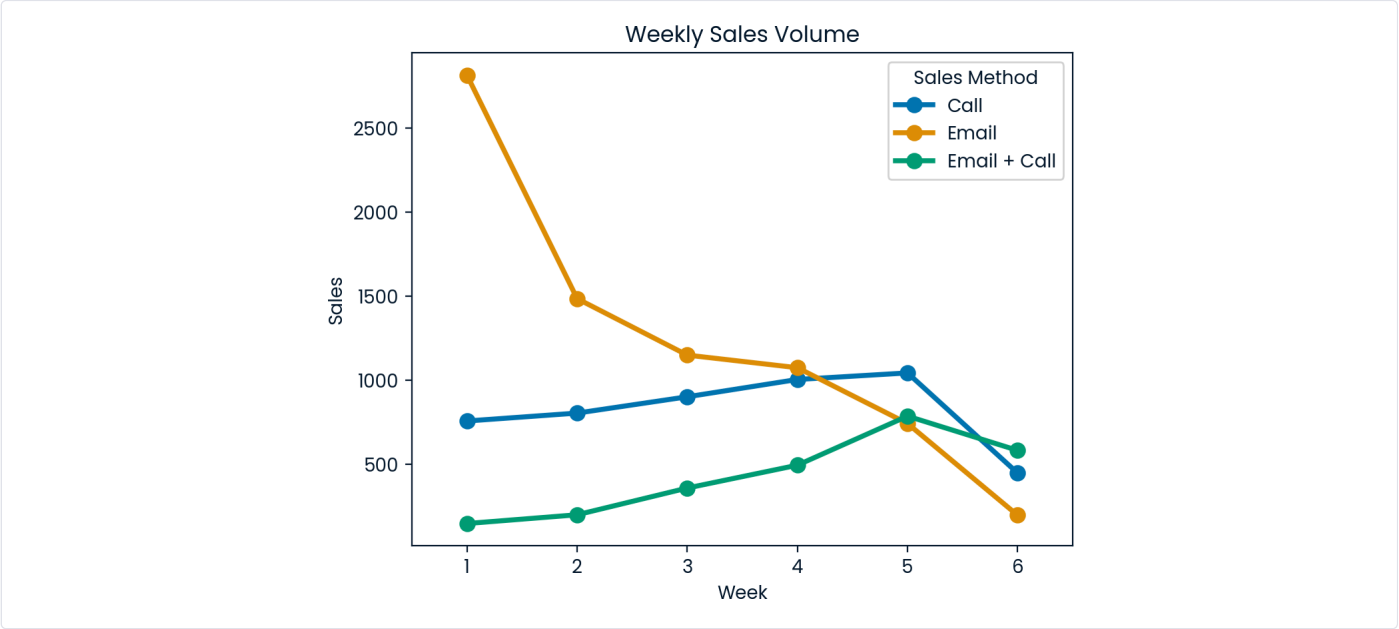
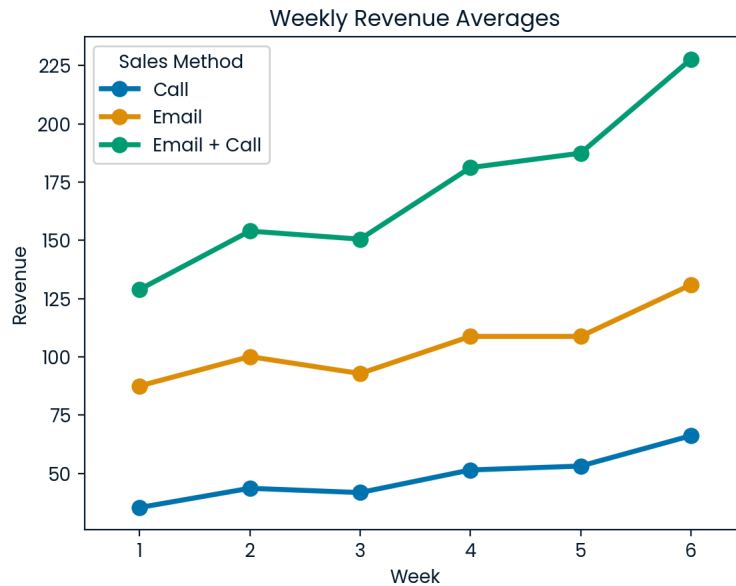### Growth in Weekly Revenue Average

|	Call Revenue Aver...	Email Revenue Aver...	Email + Call Revenue Average	Call Revenue Average % Change
1	35.35	87.52	128.9	
2	43.6	100.11	153.97	23.34
3	41.76	92.85	150.5	-4.22
4	51.46	108.77	181.2	23.23
5	53.14	108.78	187.41	3.26
6	66.16	130.91	227.7	24.5

Rows: 6   ⤢ Expand Table

Weekly Revenue Averages

## 4. Based on the data, which method would you recommend we continue to use?

Based on the data, I would recommend that we use the Email + Call combination method. This recommendation is made under the assumption that the three sales methods were not evenly applied to customers. While the email only method captured the highest total number of customers/total revenue (and would otherwise be my recommendation), this is only due to its enormous initial sales volume in week 1. In theory, the combination method is capable of capturing an equivalent initial volume. There should be no logical difference in week 1 between the two email methods as customers from both segments experienced what was effectively the same sales method at product launch (week 0). To demonstrate this, I sampled from each method to create equal populations and it's clear to see that it doesn't take long for the combination method to outpace the email-only method.

The reason I'm recommending the combination method over email-only is the change in revenue over time between the methods. The email-only method exhibited a concerning decline in total revenue growth while the combination method exhibited excellent total revenue growth. The combination method also demonstrated superior average revenue growth as well as a higher initial revenue per sale in week 1.

The call only method exhibited the poorest metrics in volume and growth of both total and average revenue, while also being the most resource intensive, so we can unquivocally dismiss the method.

In order to confirm these results, we must determine if the differences between the email-only and combination methods are statistically significant. We already know that revenue between the sales methods is not normally distributed, so we utilized the non-parametric Mann-Whitney U test to compare distributions. Using a significance level of 0.05, we returned a p-value of 2.81e-143, therefore we can confirm that the difference in results is statistically significant.
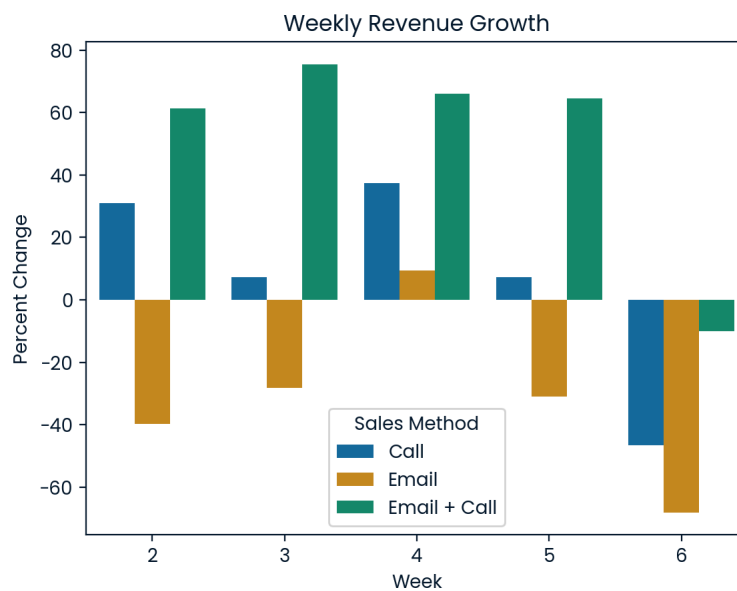

Cumulative Revenue

Cumulative Revenue with Sampling
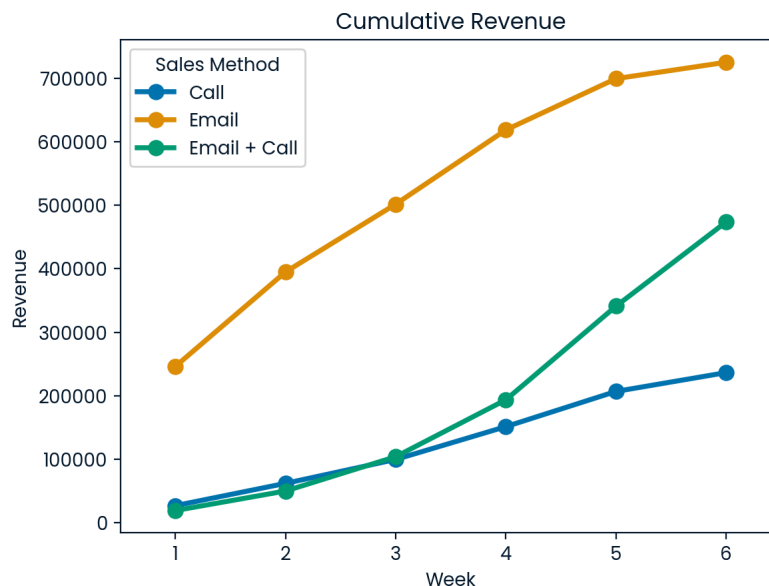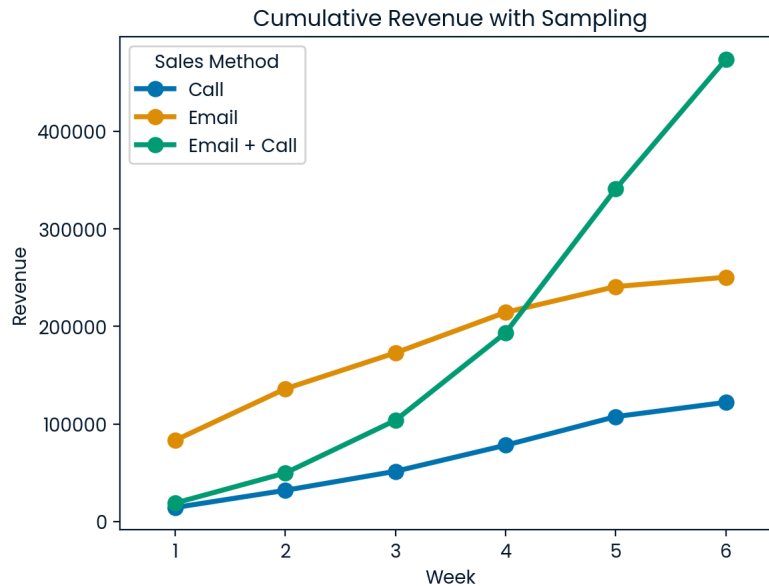
Definition of a Metric for the Business to Monitor

The metric that we should be monitoring is **weekly revenue growth** especially beyond 6 weeks. Once again operating under the assumption that the sales methods were not equally applied, we can assume very little about what the initial sales volume will look like. On the other hand, we do know that weekly revenue growth can forecast sales beyond initial revenue volume and therefore inform business strategy. Initial revenue growth for the recommended combination sales method segment was 61.42% after week 1 and increased from there. We can also see that in week 6, all sales method segments experienced negative growth, but the combination sales method experienced the smallest negative growth by far. By tracking weekly revenue growth, we can also determine the point of diminishing returns from our sales methods. Once we've reached that point, we'll be able to more accurately forecast weekly cumulative revenue.

Initial Weekly Revenue Growth

- Call: +23.34%
- Email: -39.61%
- Email + Call: +61.42%

Weekly Revenue Growth

Final Summary

- Email method <- *Would-be recommendation based solely on cumulative revenue and discounting the count of customers contacted*
  - 50% of customers
  - negative weekly revenue growth
  - middling weekly revenue per order
  - least resource intensive
- Call method
  - 33% of customers
  - middling weekly revenue growth
  - smallest weekly revenue per order
  - most resource intensive
- Email + Call method <- **Recommendation**
  - 17% of customers
  - largest weekly revenue growth
  - largest weekly revenue per order
  - 67% less resource intensive than Call method

Hypothesis testing: There **is** a statistically significant difference between revenue distributions of the email-only and combination methods.

Recommendations:

- Begin to use the Email + Call combination method as the business's primary sales method. Further analysis can be done to find the ideal time delay between the initial email and follow up phone call to maximize revenue.
- The email-only method can continue to be utilized due to its low resource requirement. It can also be transformed into a hybrid method with a phone call after the second email if sales continue to decline.
- Discontinue the call method and reallocate resources to more productive strategies.
- Improve data collection:
  - Continue tracking weekly revenue growth beyond 6 weeks to find out if the week 6 decline in sales growth is part of a larger trend.
  - Include data on how many customers were targeted so that a conversion rate can be established
  - Include individual product information and pricing for further segmentation/analysis