Pens and Printers

Data analysis of three different selling methods for the New product line

3 Sales methods for new product line

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**.

General business need: 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**.

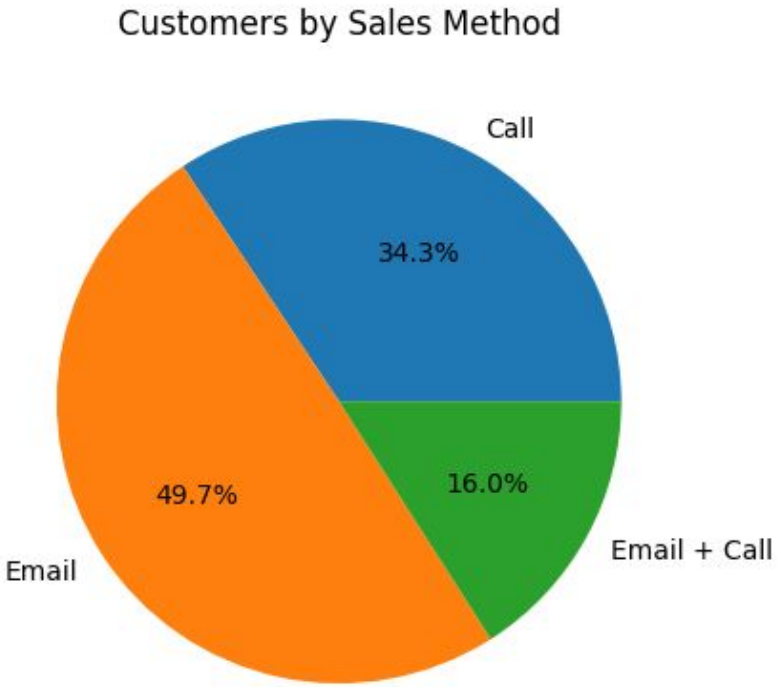
Data set and data cleaning

Column Name	Details	Validation of Data Set and Cleaning Steps if Necessary
week	Week sale was made, counted as weeks since product launch	No cleaning of data was necessary, the original data was valid. 15000 data points no-null, int64.
sales_method	Character, which of the three sales methods were used for that customer	15000 data points no-null, object. I replace 23 values "em + call" for "Email + Call" and 10 "email" values for "Email"
customer_id	Character, unique identifier for the customer	No cleaning of data was necessary, the original data was valid. 15000 data points no-null, object.
nb_sold	Numeric, number of new products sold	No cleaning of data was necessary, the original data was valid. 15000 data points no-null, int64.
revenue	Numeric, revenue from the sales, rounded to 2 decimal places.	13926 data points no-null, float64, 1074 missing values. I will not replace the missing values because I just want to analyse the original revenue data, I dont want to replace with the mean or other metric, also its less than 10% of the data that is nor available, I will work with the 13926 data points available, wich is more than enough to understand and propose solutions to the problem.
years_as_customer	Numeric, number of years customer has been buying from us	15000 data points no-null, int64. There are 2 non valid data points of customers that have been clients before 1984, i replace this values for the mean.
nb_site_visits	Numeric, number of times the customer has visited our website	No cleaning of data was necessary, the original data was valid. 15000 data points no-null, int64.
state	Character, location of the customer i.e. where orders are shipped	No cleaning of data was necessary, the original data was valid. 15000 data points no-null, object.

How many customers were there for each approach?

- Total of **15000** customers in the given data.
- **13929** customers that have revenue reported.
- Data quantity is statistically significant, so we can draw meaningful conclusions.
- Email is the easier method, so it has more customers reached.

sales_method	Count	percentage
Call	4781	34%
Email	6922	50%
Email + Call	2223	16%

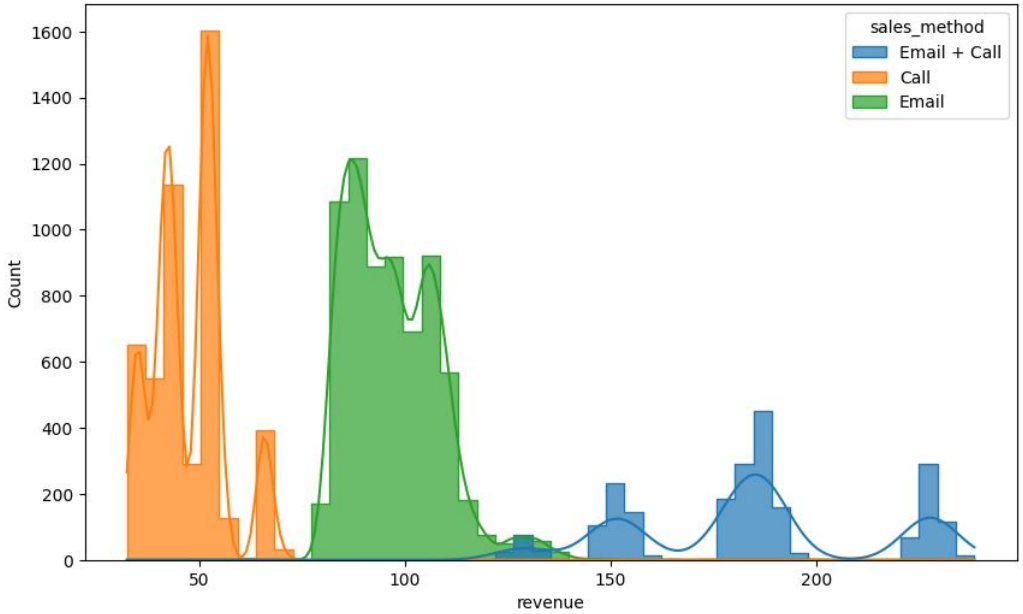


What does the spread of the revenue look like overall? And for each method?

The data for all the revenue is **more spread** than the data of the revenue filtered for each Sales Method individually.

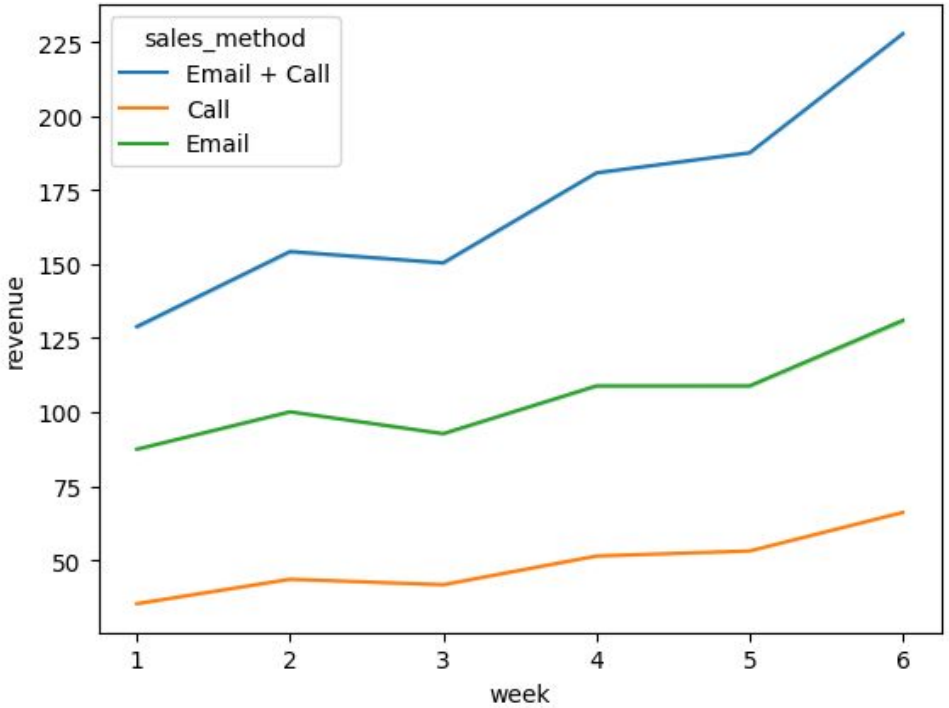
Revenue data for the "Call" method has the **less spread.**

Revenue data for the "Email+Call" method **has the more spread.**



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

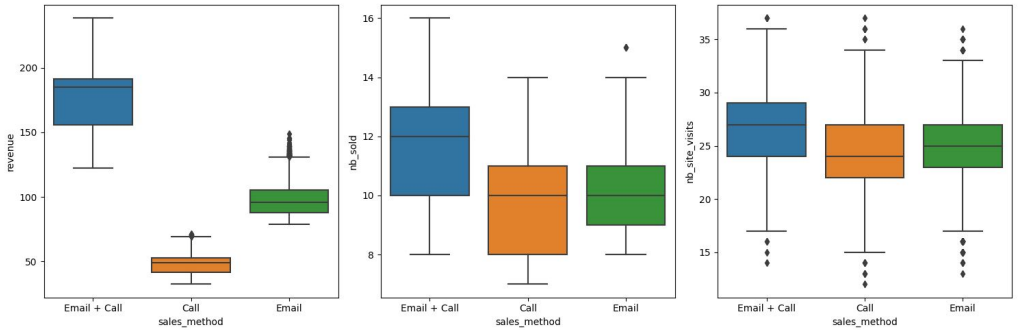
In general for all the methods the revenue **increases** as more time passes, but the sales method that has the **more relative increase over time is "Email+Call".**



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

I would recommend the sales method "Email+Call" for this reasons:

- More revenue (2x-4x).
- More products sold.
- More web page visits.
- Revenue increases with time more.
- Time investment of 10 minutes a call from sales team.



Metric: Average revenue per customer (ARPC)

If the ARPC is low for a particular sales method with respect to the general ARPC, the business may want to focus on the sales methods that has a better ARPC.

The initial value for the general ARPC is **93.93**.

Best performing method from this metric is “Email+Call”

sales_method	revenue per customer
Call	47.59
Email	97.12
Email + Call	183.65

Final remarks

---

1. From the data receive of sales of the new **product line** for the first 6 weeks of sales, it can be conclude that the sales team should focus on the **“Email+Call”** sales method.
2. We propose a metric called **Average Revenue Per Customer (ARPC)** in order to evaluate each sales method, currently the best performing method with this metric is “Email+Call”..