Printers and Pens Report

By Thaddeus Lee

New Product Sales Methods

Six weeks ago Printers and Pens 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.

Report Objective

The objective of this report is to validate and analyse the data on the sale of a new line of stationery recently launched By Printers and Pens.
This report will outline the steps taken to tidy and validate data in each column of the data set. The report will document the Exploratory Data Analysis performed to gain insight into the effectiveness of each sales method at generating revenue.

From the Exloratory Data Analysis, this report will recommend a business metric for Printers and Pens to use for effectively monitoring and evaluating sales performance and provide a recommendation on which of the tested sales method/s the company should use.

Data Validation:

The sales rep has pulled some data on the New Products sold. The data set consists of **15,000 rows and 8 columns**. Before we go into the steps of validating each column, description of each column is given below.

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

First step in validating our data set we use the **.info()** method to list each column, its data type and **non-null values**. A description of how each column is validated is provided below:

Week:- Integer data type. No changes made.

Sales_method:- Object data type. Using **unique()** method, we get the following values:

- 'Email'
- 'Email + Call'
- 'Call'
- 'em + call'
- 'email'

We use the **str.replace()** method to correct this to the expected 3 sales method values of:

- 'Email'
- 'Email + Call'
- 'Call'

Customer_id:- Object data type. No changes made. Using **nunique()** method shows 15000 unique values for individual customers.

nb_sold:- Integer data type. We check value counts and sum to check for irregularities and missing data. No changes made.

revenue:- Float data type. We see that there are 1074 missing values in the revenue column. Comparing the percentage of missing values for each **sales_method** we find:

- Call 0.036485
- Email 0.072873
- Email + Call 0.135692

As revenue for each **sales-method** is below 15% and the large amount of data available in our dataset, we can remove rows with missing values in the **revenue** column.

years_as_customer:- Integer data type. Identified two rows where customer has been with company for 47 and 63 years, but the company has only been operating since 1984. Rows dropped

nb_site_visits:- Integer data type. We look at value counts to check for irregularities but no changes need to be made.

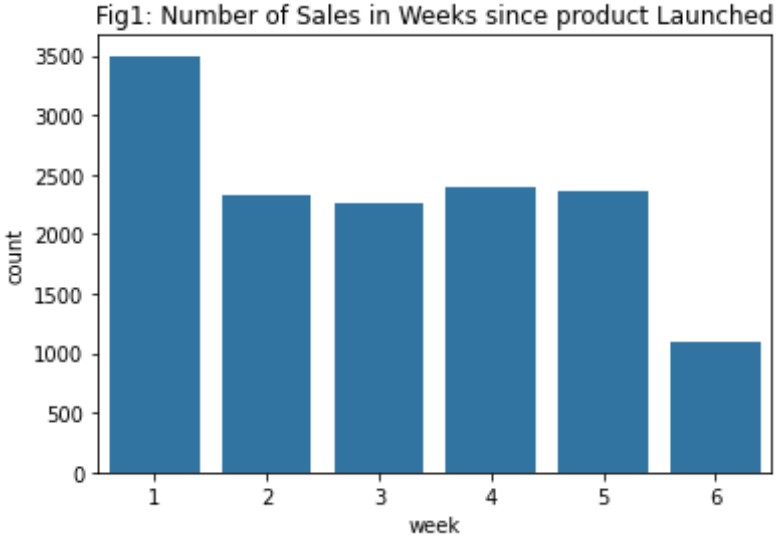
state:- Object data type. We use use **unique()** method to show each unique value in the column. There are no spelling mistakes or any other irregularities. We use **nunique()** method to confirm there are 50 unique values.

After tidying the columns and dropping the missing values from the data set, our tidy data set has **13,924 rows** and **8 columns**.

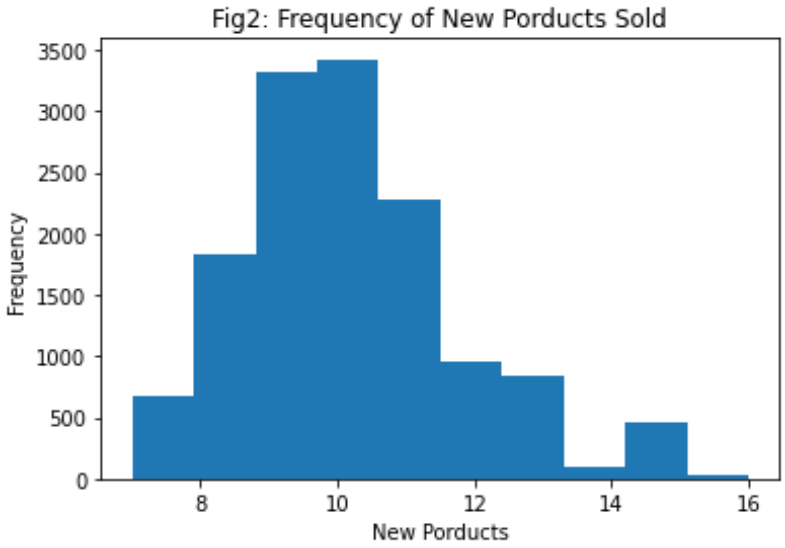
Exploratory Data Analysis

Exploring data in each column:

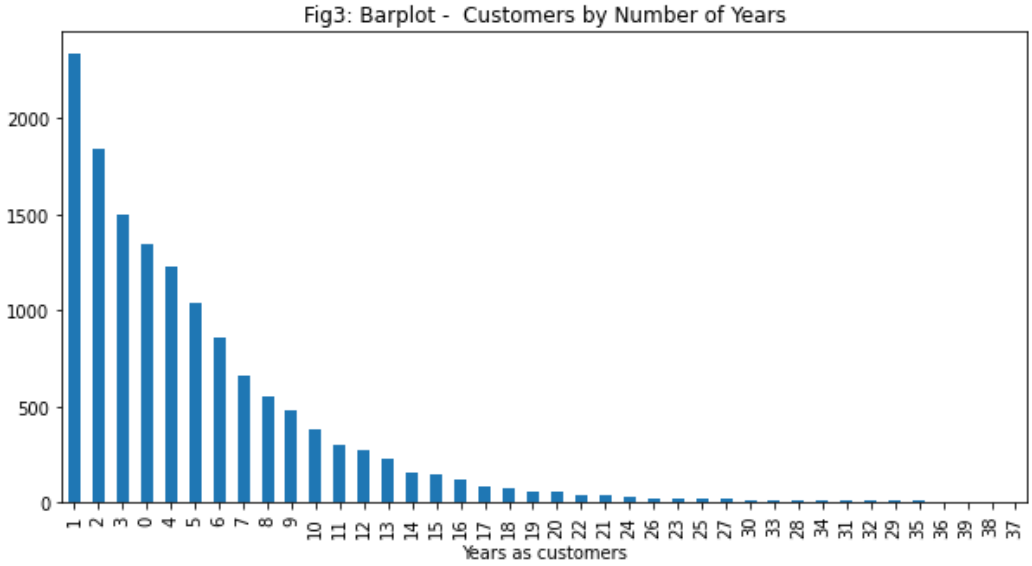
Week Column: Referring to **Fig 1** below, a bar plot showing the number of sales per week since product was launched six weeks ago. We can observe that sales are highest in the first week, decreasing slightly, and stabilising between weeks 2 to 5, then declining in week 6.



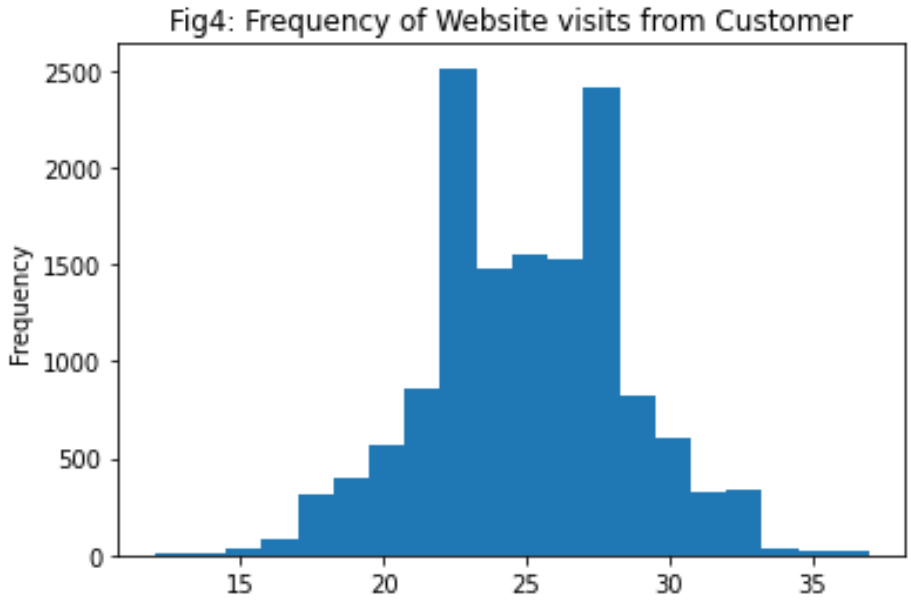
Nb_sold column: In **Fig2** we shows the frequency of the number of new products sold with each sale. The distribution resembles that of a normal distribution. We can observe that most products are purchased together in groups of 8 and 12. Most customers purchase products in groups of 10.



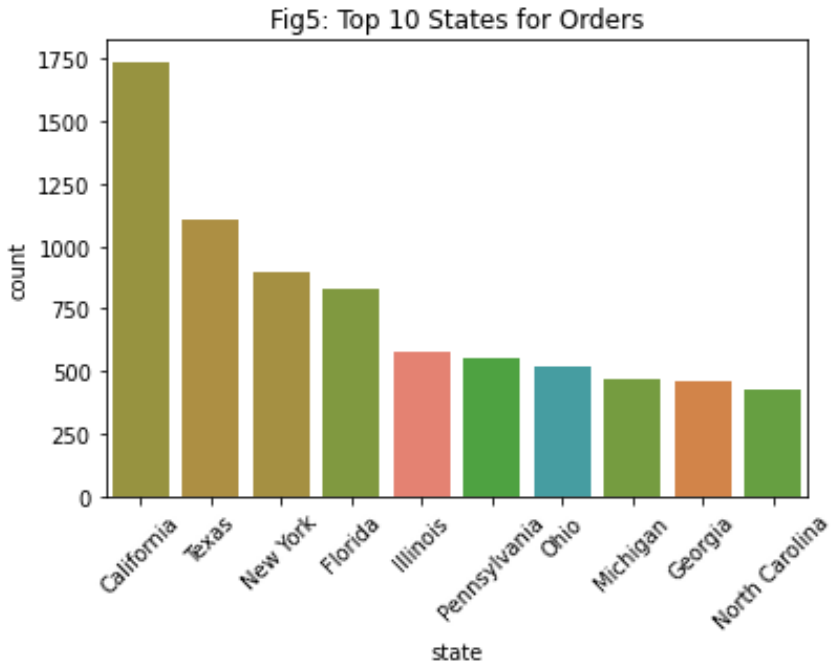
Years_as_customer column: Fig 3 below shows the duration in years that customers have been purchasing from Pens and Printers. We can see that many customers are new, and that Pens and Printers are successful in attracting new customers.



Nb_site_visits column: In **Fig4** we can see the frequency in the number of times a customer visits our website. The distribution is close in appearance to a normal distribution. We can see that a high proportion of customers has visited the Pens and Printer's website between 25 and 28 times.



State column: Fig5 shows the top 10 states where orders come from.



Breaking down the number of orders from each state by sales method, we can see the top 5 states per sales method:

Sales Method	State	Orders
Call	California	620
	Texas	357
	New York	330
	Florida	287
	Illiois	203
Email	California	854
	Texas	561
	New York	432
	Florida	408
Email + Call	California	263
	Texas	191
	New York	137
	Florida	131
	Pennsylvania	81

Business Questions:

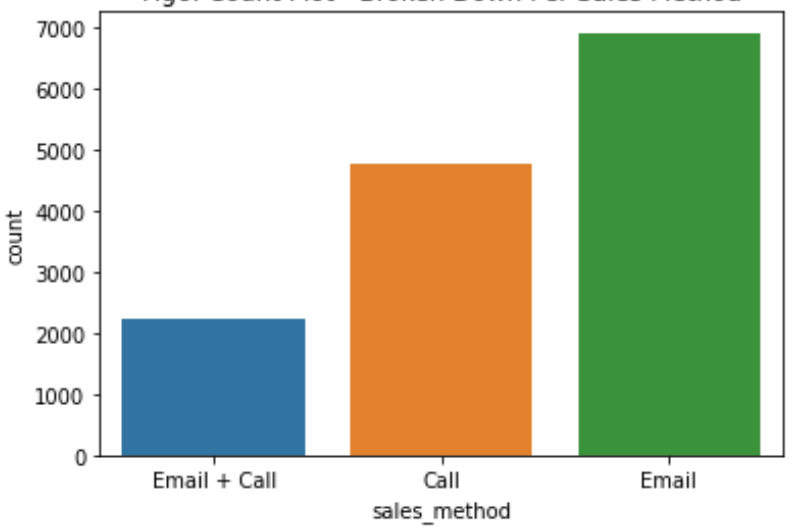
How many customers were there for each approach?

By counting the unique values in the **'sales_method'** column we can get a breakdown of the number of customers across each of the 3 sales methods (**see Fig6**):

Total: 13,924

- **Email:** 6,921
- **Call:** 4,780
- **Email + Call:** 2,223

Fig6: Count Plot - Broken Down Per Sales Method

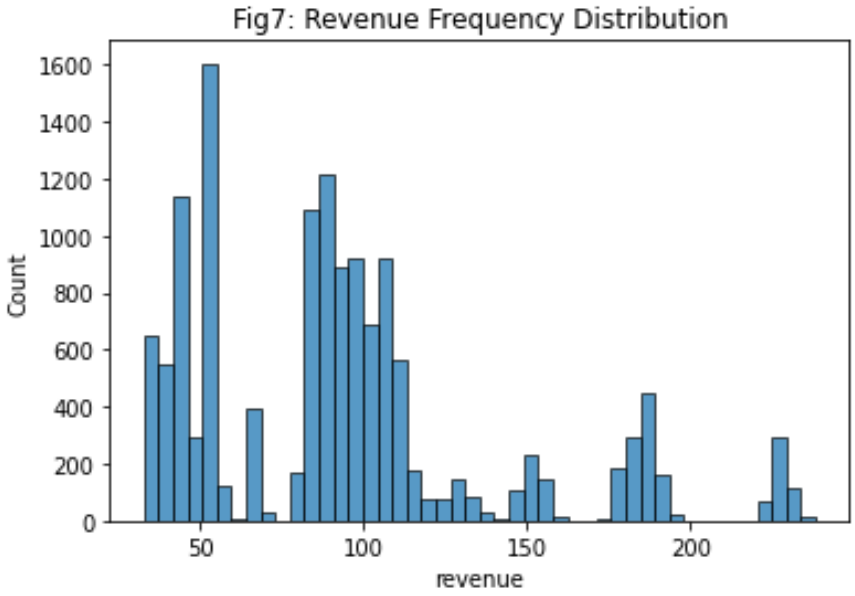


What does the spread of the revenue look like overall? And for each method?

Overall Revenue:

The overall spread of revenue appears to be of a non-normal distribution with 5 peaks and is skewed to the right (**see Fig7**).

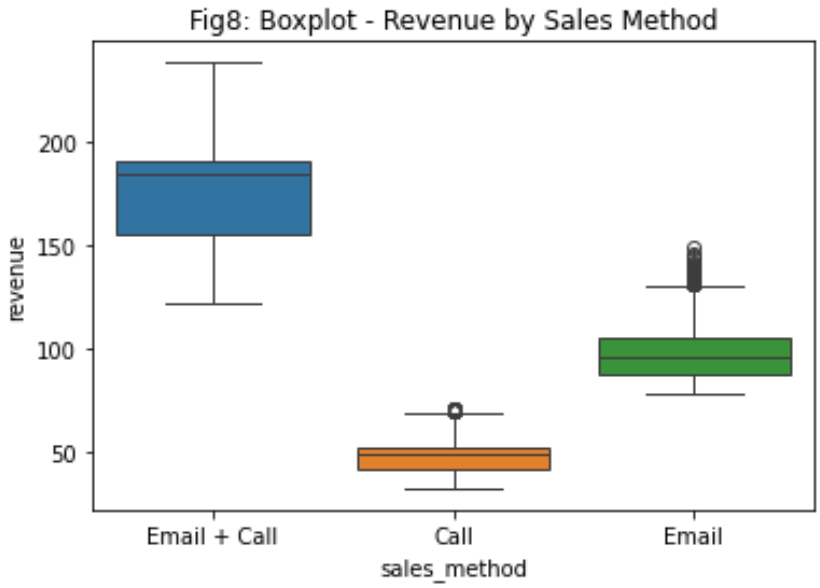
Fig7: Revenue Frequency Distribution



Overall revenue has a min value of $32.54, max value of $238.32, mean value of $93.93, a median value of $89.50 and an interquartile range (spread) of $54.86. From Fig7, we can see

a peak of 1600 customers as well as two groups of customers with higher value spend per purchase or around $180 and 230 per purchase.

Revenue per Sales Method:
For each sales method, we observe the following ranges in revenue(**see Fig8**):

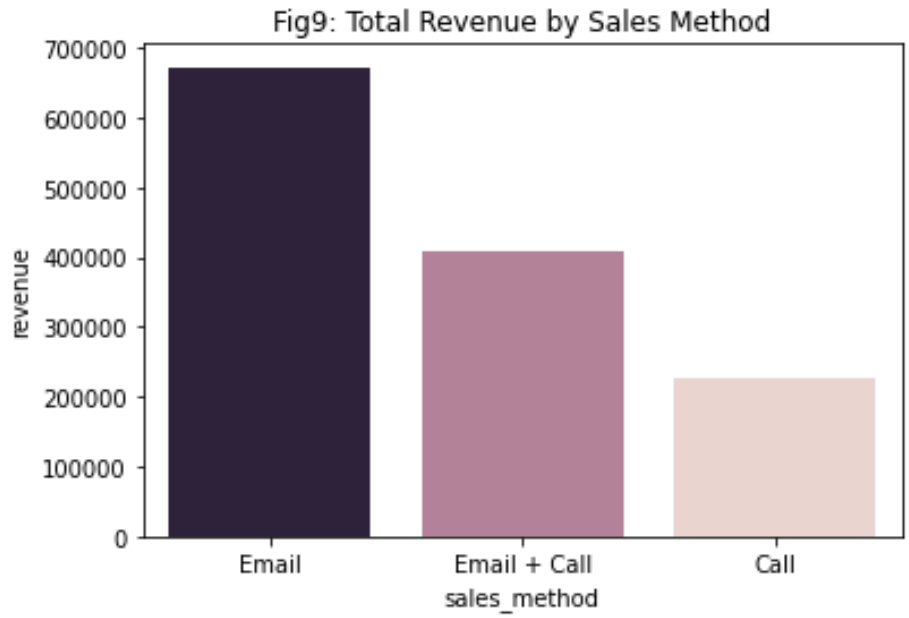


Call: Has the smallest median and spread. 'Call' has a min value of $32.54, a max value of $71.36, mean value of $47.60, a median value of $49.05 and an interquartile range of $11.21.

Email: Has the second smallest median and spread. It has a min value of $78.83, a max value of $148.97, mean value of $97.12, a median value of $95.58 and an interquartile range of $17.29.

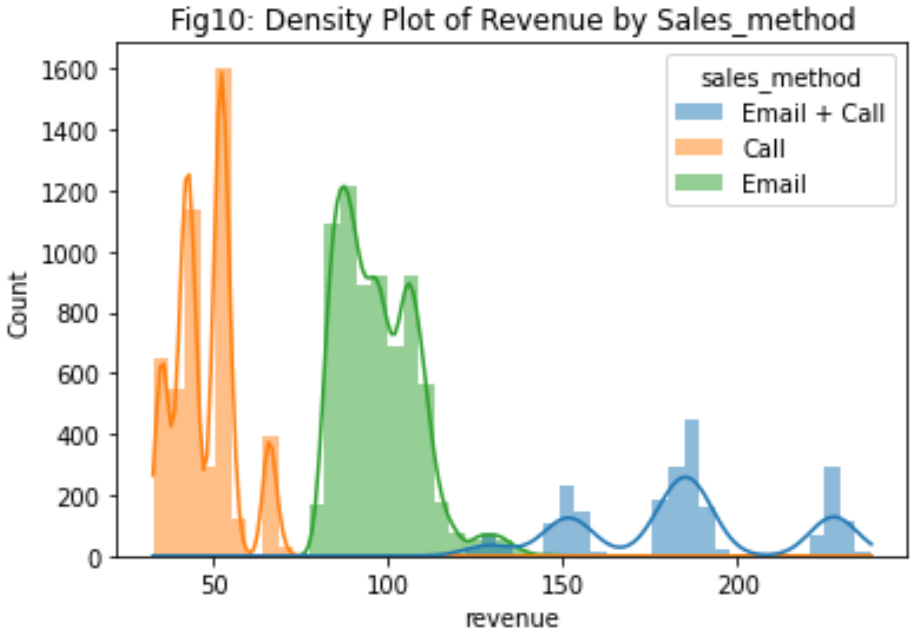
Email + Call: Has the largest median and spread. It has a min value of $122.11 a max value of $238.32 mean value of $183.65, a median value of $184.74 and an interquartile range of $35.34

Referring to **Fig9** below, we can see the sum of revenue across each sales method:



Below in **Fig10**, we recreate the histogram for revenue in **Fig7**, this time breaking down the revenue by sales method and overlaying KDE to show density between different sales methods.

We can see that the revenue for each sales method is distinct from each other, with the **'Email + Call'** method generating the highest amounts of revenue. The graph shows that sales methods **'Email + Call'** and **'Email'** attract higher revenue amounts compared to **'Call'** sales method.



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

The below table shows the mean and median revenue of each sales method per week.

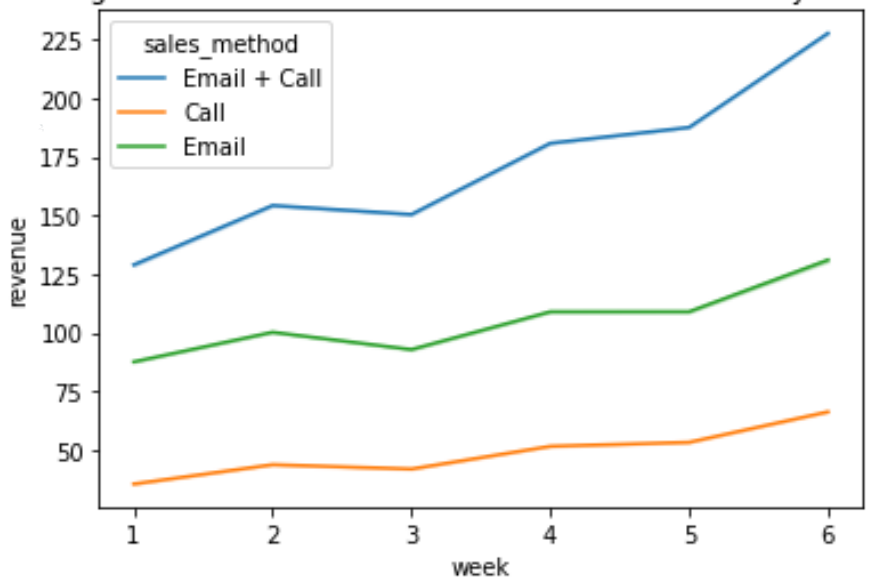
sales_method / week	mean			median		
	Call	Email	Email + Call	Call	Email	Email + Call
1	35.3502	87.4964	128.896	35.195	86.535	128.72
2	43.6026	100.141	154.247	43.46	99.25	154.17
3	41.75953	92.7626	150.423	41.47	91.8	149.93
4	51.4471	108.877	180.823	51.27	108.09	180.55
5	53.1461	108.892	187.588	52.97	108.02	187.195
6	66.1661	130.977	227.771	65.97	130.45	227.405

In **Fig11** we can see how revenue increases over time across sales methods. While revenue for each sales method increases over time, we can see that revenue for **'Call'** is the lowest across with mean revenue of $35.35 in week 1 compared to $66.16 in week 6.

The **'Email'** sales method follows a similar trend to the **'Call'** sales method. However, its mean revenue in week 1 is $87.49 and in week 6 the mean revenue is $130.97.

For the **'Email + Call'** sales method. **Fig 11** shows, that not only is the average revenue is higher than the other methods, but mean sales per week tend to increase more compared to the other methods. We can see a huge difference from week 1 where mean revenue was $128.90 compared to week 6 where mean revenue was $227.77.

Fig11: Line Plot - Revenue from Each Sales Method by Week



The below table shows the mean and median new products sold for each sales method per week.

	mean			median		
sales_method week	Call	Email	Email + Call	Call	Email	Email + Call
1	7.07973	8.730388	8.64885496	7	9	9
2	8.803871	10.06323	10.0701754	9	10	10
3	8.268433	9.270423	10.0123457	8	9	10
4	10.20872	10.91919	12.0369515	10	11	12
5	10.65641	10.92679	12.4704142	11	11	12
6	13.1452	13.11602	15.0614754	13	13	15

Similarly to the trend observed in **Fig11** with revenue increasing each week, **Fig12** shows how the number of products sold increases over time. We can observe that the mean products sold for each sales method is distinct, however by week 6 the **'Call'** sales method is slightly higher than **'Email'. Fig12** shows that while the number of sales decreases in week 6, as shown earlier in **Fig1**, in the later week's, customers are buying more products per order.

Fig12: Line Plot - Number of Products Sold from Each Sales Method by Week

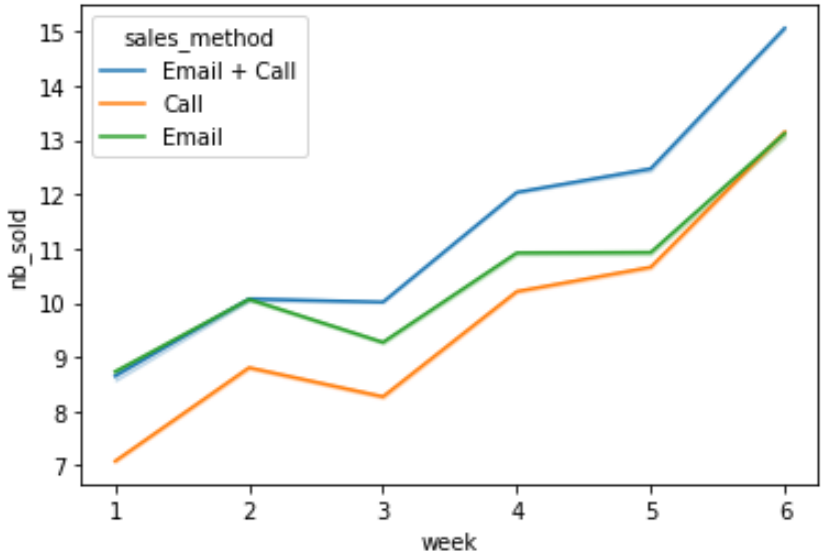
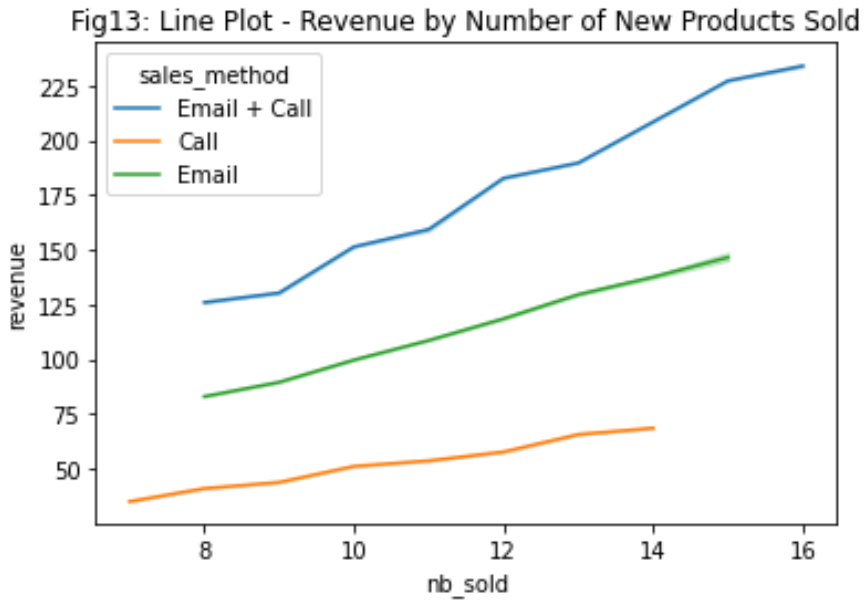
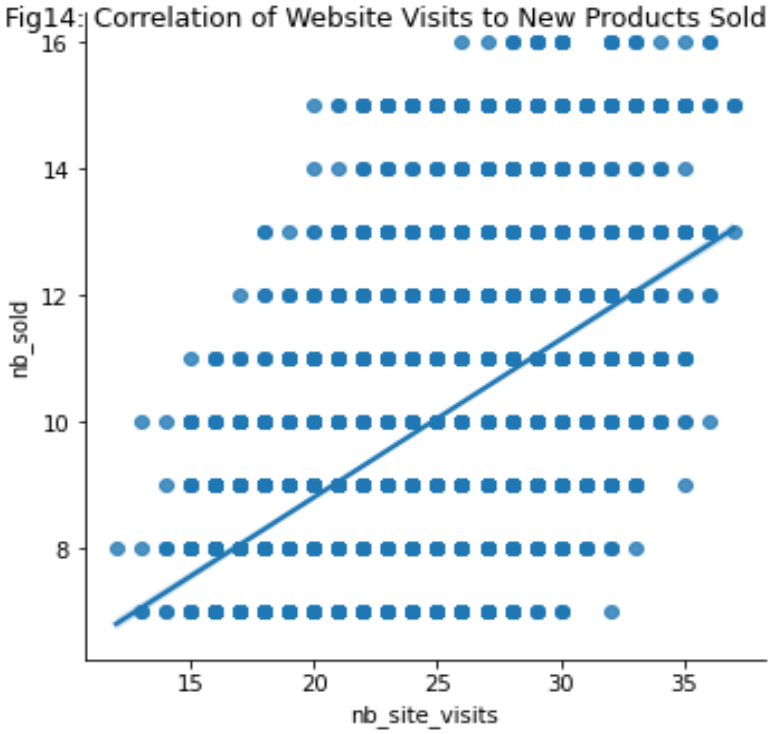


Fig13 below shows the relationship between number of products sold and revenue across each of the sales methods. We can observe that the **'Call'** Sales method has the lowest revenue and products sold compared to the other methods. **'Email'** has higher revenue compared to **'Call'** and we can see that customers in the 'Email' category purchase more products. **'Email + Call'** not only has the highest revenue, but it looks like customers in this category buy more products compared to other sales methods.



Other Relationships in the data

Looking at **Fig14** below, we can see that there is a positive relationship between the number of products sold and number of site visits. The more frequent a customer visits the Printers and Pens website, the more products sold. Correlation Co-efficient is 48.76%.



Conclusions

Business Metrics:

The company's goal has been on selling products which enable our customers to be more creative, focused on tools for brainstorming. I would recommend that the sales team use the average revenue amount of their chosen sales method for the last 6 weeks as their metric.

This would account for changes in the number of sales and products sold over time.

Based on data for new products sold, **Email + Call** method has the highest sale revenue average ($183.65). The average revenue for the **Email** sales method is $97.12, and the average revenue for the **Call** sales method is $47.60.

For example, the Printers and Pens sales team can monitor whether they are achieving objectives with the **Email + Call** method by comparing the average revenue of the last 6 weeks to the current baseline average of $183.65. Averages above or below this figure would indicate whether or not they were meeting expectations.

Recommendations:

- Of the sales methods tested, I recommend that the company continues to use the **Email + Call** method since it captures the most revenue and sales with the potential to significantly increase over time (**see Figs 11 and 12**).

- As the **Email** sales method requires little work from the team and is thus less costly, it is recommended that Printers and Pens also continue to use this method.

- Due to having the lowest revenue in comparison to the other sales methods, as well as the higher cost and effort to service customers (average of 30 minutes on phone to customers), it is recommended that Printers and Pens discontinue using the **'Call'** method.

- Given the positive relationship observed between site visits and products sold in Fig14, more effort can be placed into marketing new products, and offering customers promotions and discounts to enhance sales. An example of this would be a loyalty program.

- I would recommend that the company uses the average revenue respective of the sales methods recommended above as the metric to monitor for strong and increasing sale revenues.

- Improve data collection for enhanced analysis and quality:
    - The company should investigate how there are 1074 missing values in the revenue column and prevent loss of data in future.
    - Data on the types of products purchased by customers. For example pens, printers, notepads and other types of stationery for detailed analysis.