Data Validation:

Sales Method:

I used the Unique function to print the unique values in a column.

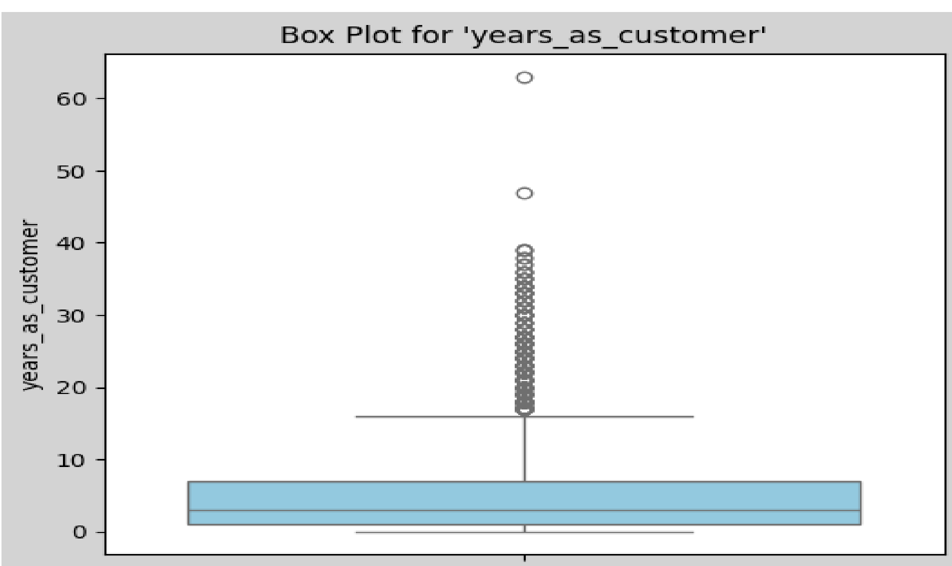
After exploring, there were inconsistencies in the column. for example, Email value and email value, both of them are the same but was inconsistent in entry phase where one of them written with uppercase for the first letter and the other is not. In addition, (Email + Call) and (em + call), it is a must to match similar values to each other since they mean the same.

```
Before
 ['Email' 'Email + Call' 'Call' 'em + call' 'email']
After
 ['Email' 'Email + Call' 'Call']
```

Number of Products Sold / Number Of Site Visits /Years As Customers:

I used Box plot to detect outliers.

It is valid to remove outliers but sometimes it is not, it depends on the case that we are working with. In this case, the number of the products sold, and site visits are not outliers because they are real data points that was made by customers. Therefore, nothing will be done to these columns. However, the box plot of years as customer column detected customers more than 40 years. It is a mistake in data entry because the company was founded in 1984, and we are in 2024. the maximum reach as customers must be 40 years.



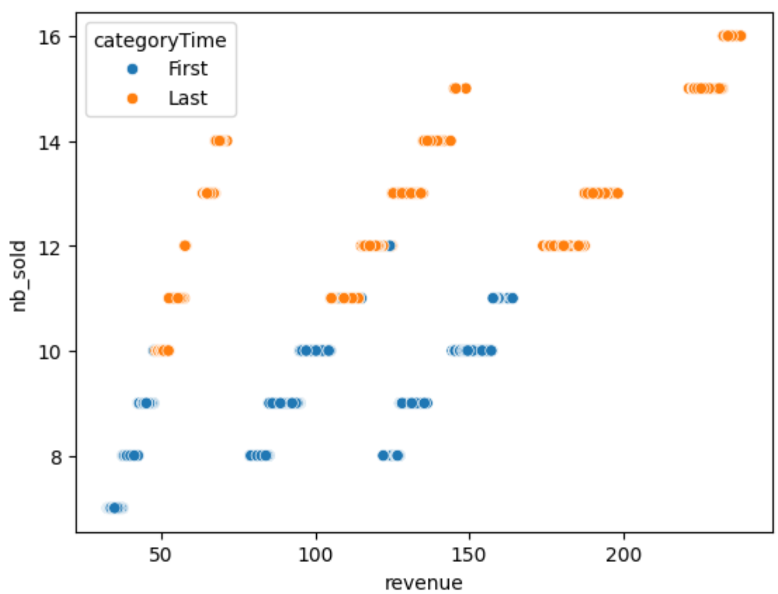
	week	sales_method	customer_id	nb_sold	revenue	years_as_customer	nb_site_visits	state
13741	2	Email	18919515-a618-430c-9a05-2c7d8fea96af	10	97.22	63	24	California
13800	4	Call	2ea97d34-571d-4e1b-95be-fea1c404649f	10	50.47	47	27	California

At first, I tried to do filtering to find common attributes between the data and the outliers records. But unfortunately, I could not detect a clear and logical relationship between the similar attributes that the outliers records share with other records. Therefore, I have decided to delete these two rows since it is only 2 out of 15000. Lastly, I created a new feature that categorizes customers into Long time customers (more than 10 years), medium time customers (between 5 to 10), short time customers (to 5 years).

Week:

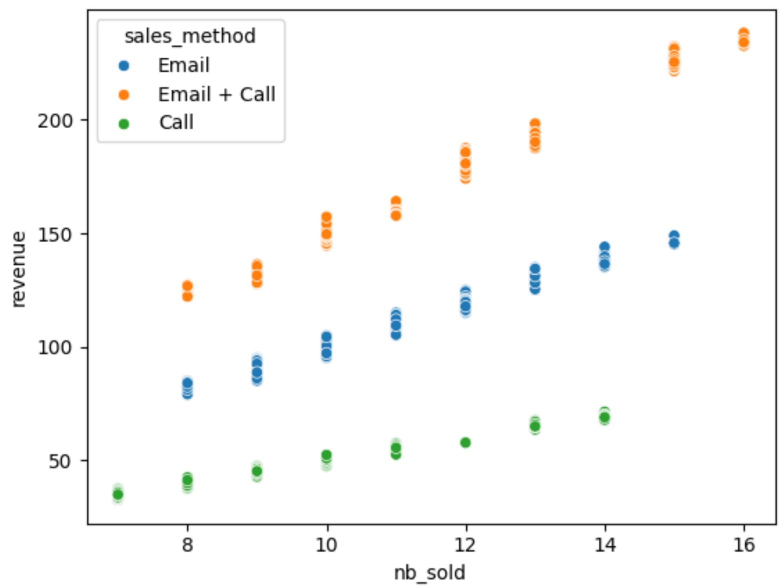
For visualization purposes, I created a new feature that has two values, First (for values that has week equals to [1,2,3]). Last (for values that have week equal to [4,5,6]).

Revenue:



From this plot, there is a relationship between (Week of Purchase After Launching, Revenue and Number Of Products Sold).

Most likely, when there are more than 12 purchases for a product, it would be after launching the product for 3 weeks.



From this plot, there is a relationship between (Sales method, Revenue and Number Of Products Sold).

Therefore, I made a function that will fill nan values of revenue by calculating the average revenue based on the group of (sales method, week, number of products sold). The missing value will be replaced by the proper average value based on the conditions (the conditions are what the sales method used, what week was the missing revenue row in, and how many products sold were there).

Customer ID:

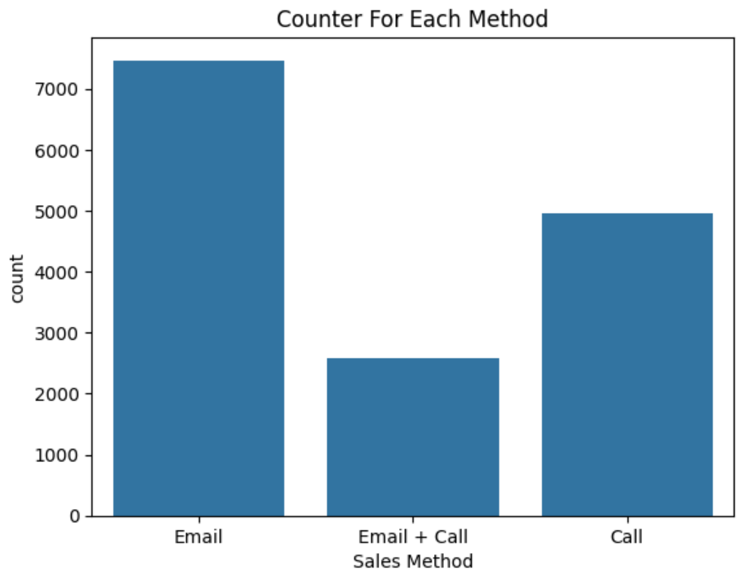
There was no usage of customer ID since the number of rows we have in the data is 15000 and the unique values of customers are 15000. Which indicates that there is no valid use to do with a column that all of its values are unique.

State:

No validation was done for the State column.

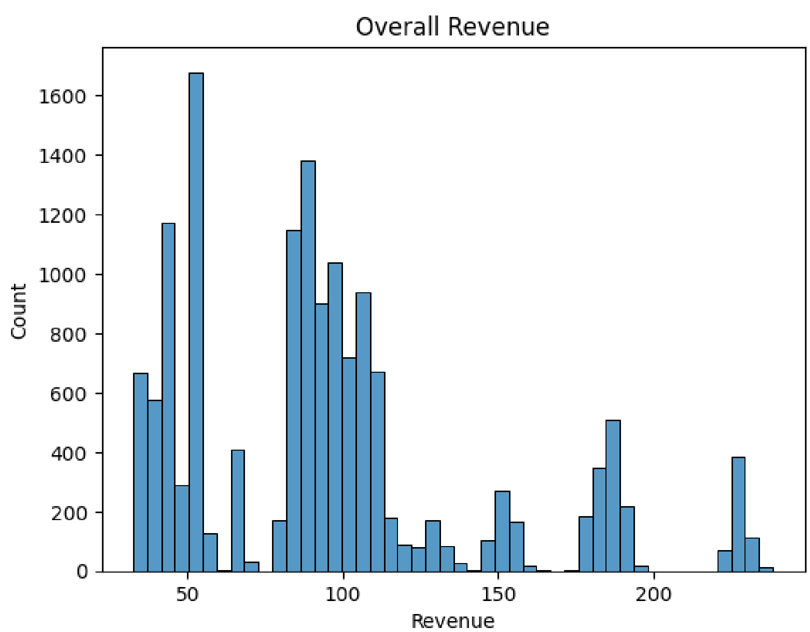
Exploratory Data Analysis:

How many customers were there for each approach?

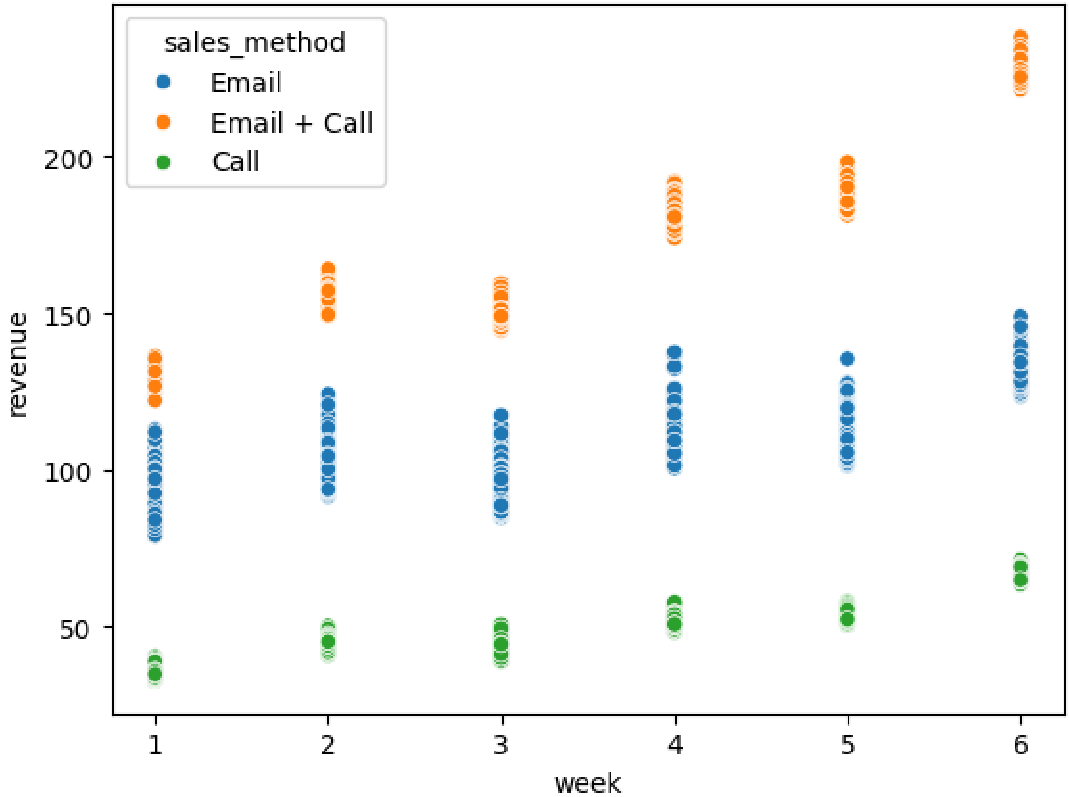


It shows that the most frequent approach used is by sending emails to customers. Then by calling customers. And less than 3000 customers got an email and a call.

does the spread of the revenue look like overall?



It shows that revenue has segments, where there is a group of revenue focused in 50, and another group in 100, another group in 150, another group in almost 180, and another group in 220. I could conduct there is a specific relationship on why the distribution of revenue looks like it has different mean points.



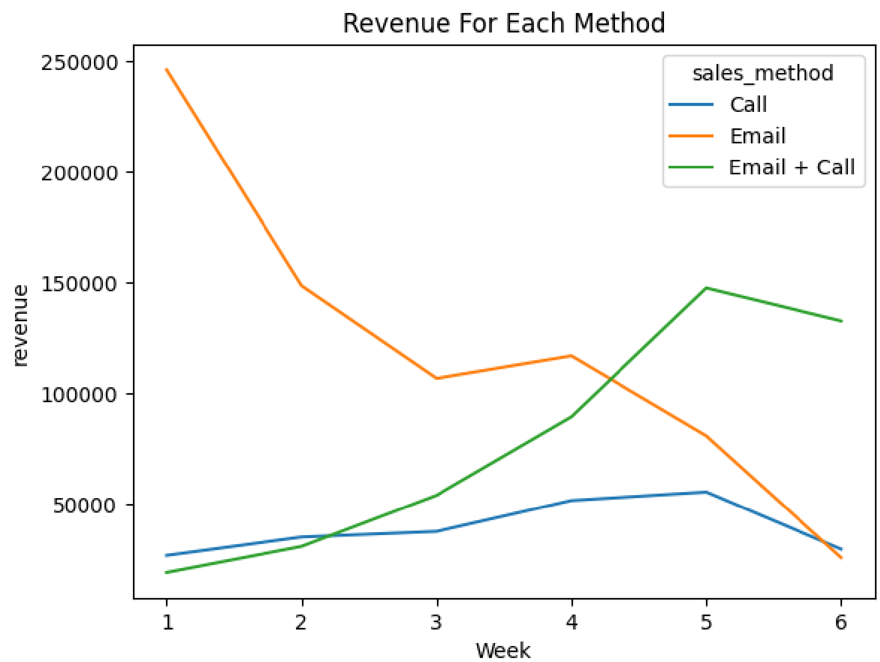

```
Minimum Revenue For Call: 32.54
Maximum Revenue For Call: 71.36
Mean Of Revenue For Call Method: 47.59686610878661
-------------------------------------------------------
Minimum Revenue For Email: 78.83
Maximum Revenue For Email: 148.97
Mean Of Revenue For Email Method: 97.1276708568126
-------------------------------------------------------
Maximum Revenue For Email + Call: 122.11
Maximum Revenue For Email + Call: 238.32
Mean Of Revenue For Email + Call Methods: 183.651232568601
```

As shown, there is a pattern that:

- When the sale method is by Call, the revenue is between 32 to 72.
- When the sale method is by Email, the revenue is between 78 to 149.
- When the sale method is by Call And Email, the revenue is between 122 to 239.

This explains the multi mean points in the distribution of the revenue, where each method has a specific and a different range of revenue.

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

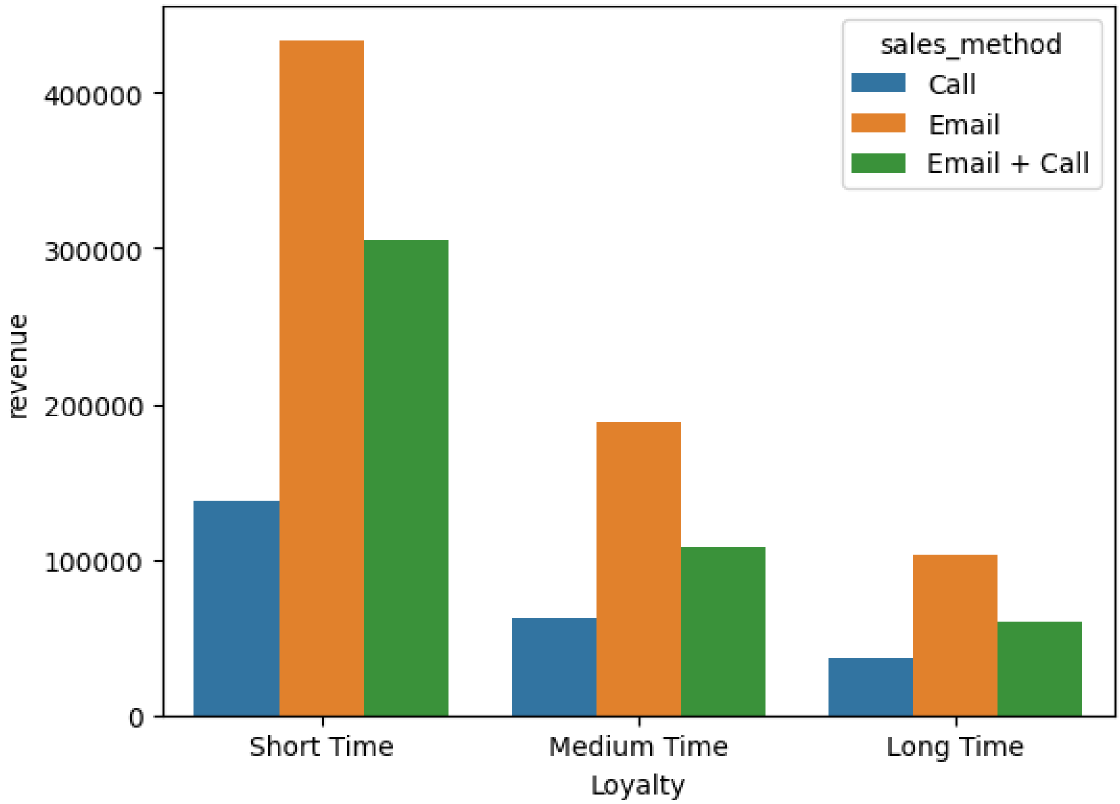


It shows the following:

- Email: email method made a huge revenue for the first 4 weeks and half when launching. This indicates that email method is a great approach to follow for almost a month. Then gradually decrease with the time goes on, it seems that the launched product will be known enough after a month where emails will be neglected.

- Email + Call: The email and call method made a great revenue but after one month when the product launched. It is a good approach when reminding customers of the launched product after a month of its release. However, it seems that it is only beneficial for two weeks after a month of launching. The data is still not enough to detect, however after week five, the revenue is in a decreasing line.

- Call: The call method made less revenue, and it seems it will not grow more even as time goes on. Which indicates that calling is not a valid approach for making revenue and consuming the team efforts for nothing. However, using this approach for high loyalty customers is preferable to show a sign of appreciation.



The efforts in terms of calling are almost for nothing, it is preferable to use it wisely especially for long time customers to gain loyalty and intangible revenue.

Business Metric:

The business can monitor their goals by tracking weekly revenue trends across sales methods.

week	sales_method	revenue
1	Call	26794
1	Email	246262
1	Email + Call	19081
2	Call	35104
2	Email	148693
2	Email + Call	30861
3	Call	37661
3	Email	106815
3	Email + Call	53998
4	Call	51651
4	Email	117030
4	Email + Call	89521
5	Call	55494
5	Email	80887
5	Email + Call	147644
6	Call	29642
6	Email	25792
6	Email + Call	132757

We can detect the total revenue of each method for each week, which can help us identify trends and ensure that we are on the right track. As shown, the initial values of the metric will be the first week, where call revenue was 26794$, email revenue was 246262$ and email + call revenue was 19081$. We can see that in the second week, the call revenue and email + call revenues increased. However, email revenue significantly decreased. To compare, the total revenue of the first week was 292137$, while the second week revenue was 214658$. The difference equals to -77479$. Even though two approaches revenue increased in the second week.

Business Recommendations:

What I believe from the data is we should use more of Email Approach for a month when launching a new product. Then, after a month, we should add the Call Approach for two weeks while keep sending emails. Which means using Email + Call Approach. In addition, adding only calls approach on high customer loyalty for those who have been customers for many years. It is a great sign of appreciation for long-term customers to remember them with a special call. In this way, we are balancing methods and how to use them properly. In addition, balancing in terms of efforts that team members make.

Since there is no data that can help us decide what to do next after 6 weeks of launching a new product, we better get more data by experimenting with the same methods or add another sales method for marketing.

For example, launching a YouTube video and advertising it for targeting wider and different segments of customers. By that, we can detect what customers we should focus on to gain more revenue while consuming less effort.