



DATA CAMP

Data Analyst Certification Project: Analyzing Product Sales Data

D. S. A. Aashiqur Reza

December, 2024

Agenda

- Data Cleaning and Validation
- Data Exploration
- Definition of Performance Metric
- Conclusion

Data Cleaning and Validation

15K ROWS, 8 COLUMNS

Data contains 8 columns including unique customer ID, and other relevant variables.

1074 MISSING VALUES

Revenue column contains 7% missing values.

UNWANTED DATA

Data entry error noticed on the

Summarizing missing values in the data:

```
week          0
sales_method  0
customer_id    0
nb_sold        0
revenue       1074
years_as_customer  0
nb_site_visits  0
state          0
dtype: int64
```

Count of unique values:

```
week          6
sales_method   5
customer_id   15000
nb_sold        10
revenue       6743
years_as_customer  42
nb_site_visits  27
state         50
dtype: int64
```

Unique values in sales method column:

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

Years as customer more than 40:

```
13741    63
13800    47
```

Data Cleaning and Validation

- Missing values in the revenue column are imputed with the median of each sales method.
- Typos occurred in sales_method column: email, and em + call. These are replaced by Email and Email + Call

Data columns (total 8 columns):

#	Column	Non-Null Count	Dtype
0	week	15000 non-null	int64
1	sales_method	15000 non-null	object
2	customer_id	15000 non-null	object
3	nb_sold	15000 non-null	int64
4	revenue	15000 non-null	float64
5	years_as_customer	15000 non-null	int64
6	nb_site_visits	15000 non-null	int64
7	state	15000 non-null	object

Count of unique values:

week	6
sales_method	3
customer_id	15000
nb_sold	10
revenue	6743
years_as_customer	41
nb_site_visits	27
state	50

dtype: int64

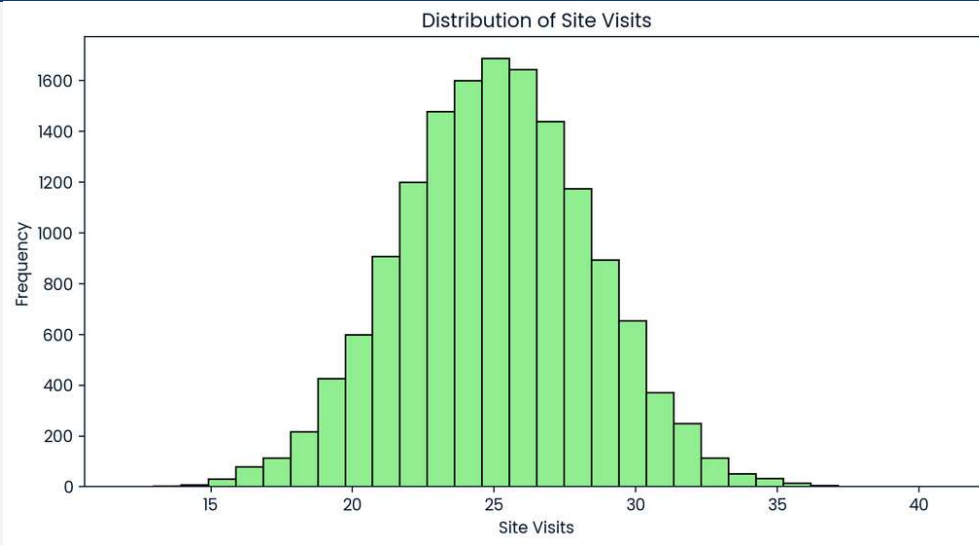
Unique values in sales method column:

['Email' 'Email + Call' 'Call']

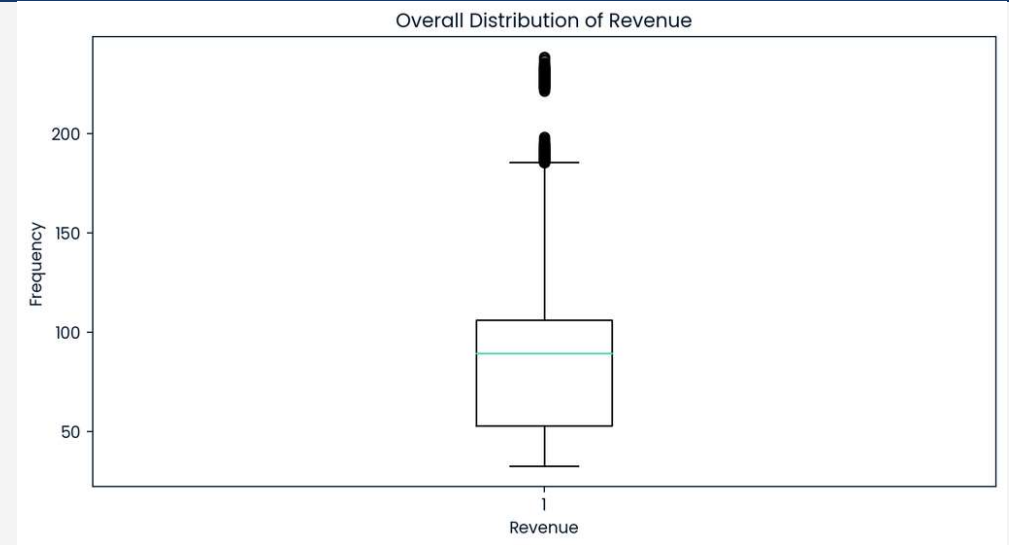
Years as customer more than 40:

Series([], Name: years_as_customer, dtype: int64)

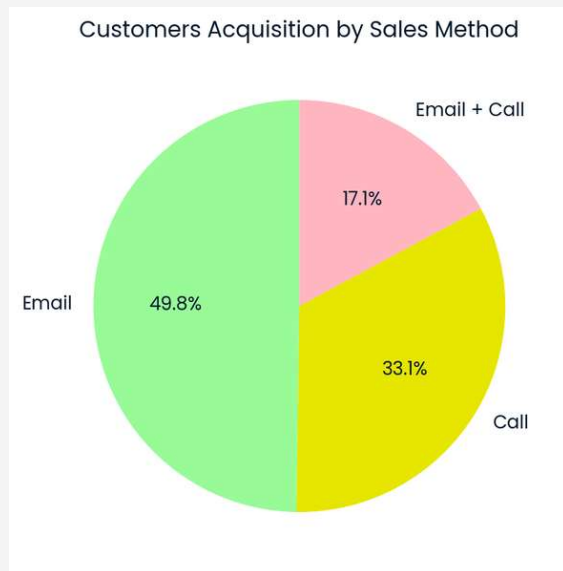
Data Exploration



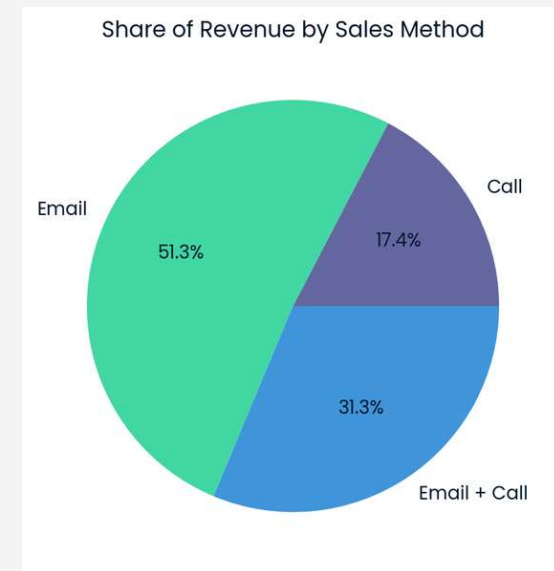
◆ Around 25 web visits per customer, ranging between 12-41.



Data Exploration

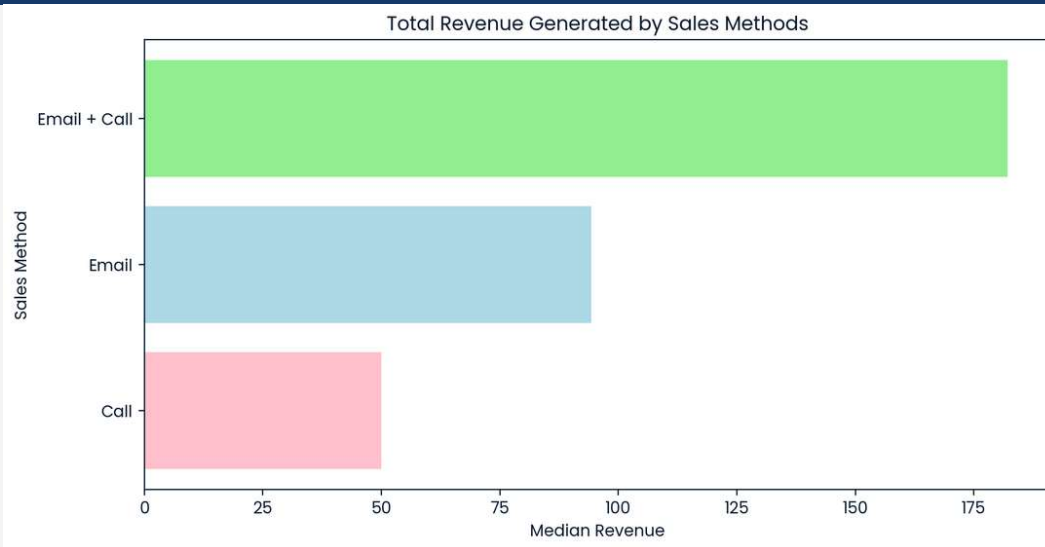


◆ 50% of the total customers were acquired by sending Email only. 17% of total customers are acquired by Email + Call method and the rest by only Call.

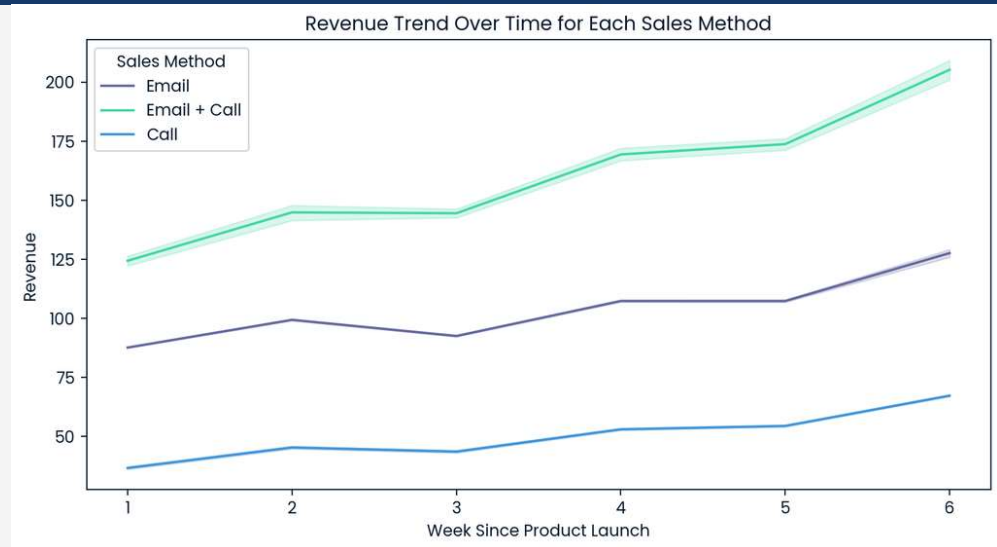


◆ Though the method **Email + Call** is the less used method, it shared 31.3% of the total revenue generated. 51.3% of total revenue came from **Email** sales and rest came from phone calls.

Data Exploration



◆ The median revenue generated by sending Emails to the customer is higher than using Calls only. But the revenue goes much higher combining Emails and Calls.



◆ The Number of items sold increases by week since products are launched by acquiring the **Email + Call** method.

Performance Metric

We are selling our products in three methods: **Email**, **Call**, and **Email + Call**. We are getting site visits for each method and generating revenue. Therefore, to examine the performance of these methods, we set the metric: **Average Revenue per Visit(ARPV)**.

$$ARPV = \frac{\text{Total Revenue}}{\text{Total Visits}}$$

Sales Method	ARPVM
Email	2.01
Call	3.90
Email + Call	6.38

To analyze this metric (Average Revenue per Visits by Sales Method, ARPVM) for each sales method, the formula will be:

$$ARPVM = \frac{\text{Total Revenue (by Sales Method)}}{\text{Total Visits (by Sales Method)}}$$

highest ARPVM for **Email + Call** meaning that it has been the most effective sales method.

Monitor the trend of ARPVM and always seek room for improvement.

A proper benchmarking of revenue per visit according to market standard are helpful to develop the process. Use the metric to evaluate the effectiveness of marketing campaigns, website optimizations, or changes in sales strategies.

Conclusion

Based on the findings, **Email + Call** is the most effective sales method.



Revenue generated with **Email + Call** is 31% whereas only 17% of the customers were reached by this method. Maximizing the usage of this method will generate more revenue for the company.



With more data, further analysis on churn rate, return on investment, customer acquisition cost, average order value - these kind of metrics can be analyzed.

Thank You!