

Data Analyst Professional Practical Exam Submission

You can use any tool that you want to do your analysis and create visualizations. Use this template to write up your summary for submission.

You can use any markdown formatting you wish. If you are not familiar with Markdown, read the [Markdown Guide](#) before you start.

Task List

Your written report should include written text summaries and graphics of the following:

- Data validation:
 - Describe validation and cleaning steps for every column in the data
- Exploratory Analysis:
 - Include two different graphics showing single variables only to demonstrate the characteristics of data
 - Include at least one graphic showing two or more variables to represent the relationship between features
 - Describe your findings
- Definition of a metric for the business to monitor
 - How should the business use the metric to monitor the business problem
 - Can you estimate initial value(s) for the metric based on the current data
- Final summary including recommendations that the business should undertake

Report

1. **data validation and cleaning steps** : 1.1 check for null values and found revenue have some. Those were replaced by mean value. 1.2 check for years_as_customer. Give the company started 1984, largest value of years_as_customer should be 2024-1984 = 40. All value exceeds that were replaced by 40. 1.3 standardized sales method. em + call should be email + call. And upper cases were all standardized to lower case.

2. **Exploratory Analysis:**
 - 2.1 two graphics showing single variables only to demonstrate the characteristics of data

2.2 one graphic showing two or more variables to represent the relationship between features

2.3 Findings

- Revenue have wide-range distribution. From the investigation, each customer have different characteristic in buying, which leads to different number of products buying and different revenues.
- number of site visit (nb_visit) have an approximately normal distribution.
- Total revenues by sales method of email is the largest among 3 methods. However, average revenue per customer by sales method of email+call is the largest. It can be interpreted that email+call are most likely to contribute to largest amount of revenues per customer, on average.

```
total revenue
- call      244,565.73
- email     723,418.47
- email+call 441,040.01
```

 Copy

```
Average revenue per customer
- call      49.29
- email     96.90
- email+call 171.48
```

```
Total sample customers
- call      4962
- email     7466
- email+call 2572
```

- Average revenue per customers by nb_visit is the best when nb_visit is around 37 times, after that it dropped. Therefore, we can interpret that customers who visit website around 37 times are most likely to contribute to high revenues.

3. Metrics for business to monitor

- Average revenue per customer by sales method
- Average revenue per customer by number of sites visit

4. Initial value of average:

4.1 Average revenue per customer by sales method:

- call = 49.29

- email = 96.90
- email+call = 171.48

4.2 Site Visits Revenue Statistics:

nb_site_visits	revenue
12	40.95
13	51.33
14	72.24
15	70.99
16	61.67
17	67.19
18	66.23
19	71.78
20	74.57
21	77.52
22	82.42
23	85.37
24	89.71
25	91.97
26	96.69
27	101.84
28	104.27
29	112.09
30	115.21
31	120.54
32	128.08
33	122.26
34	139.15
35	149.87
36	156.49
37	167.78
41	93.94

Copy

4.3 Reason of choosing these metrics: These 2 metrics should be monitored to optimize the process to get higher revenues. For example, if we found that the Average revenue per customer by call is really low, we may investigate further what's wrong and improve it. Or in the future if we have various sales methods we may decide to cut down some methods that don't bring high revenues. Also, investigating average revenue per customer by number of sites visit can help us to know how many times customers take until buying the products. If we found some customers visit website a lot of time but turn out buying less, we can investigate further and make improvements as well. For example, customers who visit the website 41 times contribute to around -40% in revenues comparing to 37 times visitors. We may investigate further what's wrong.

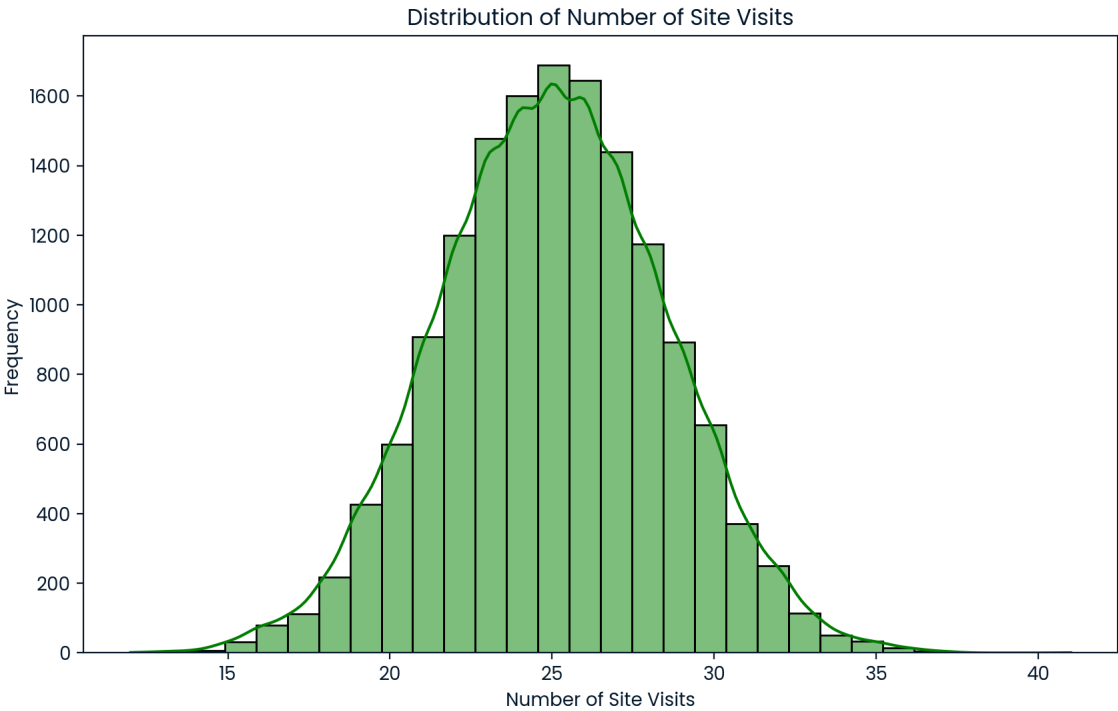
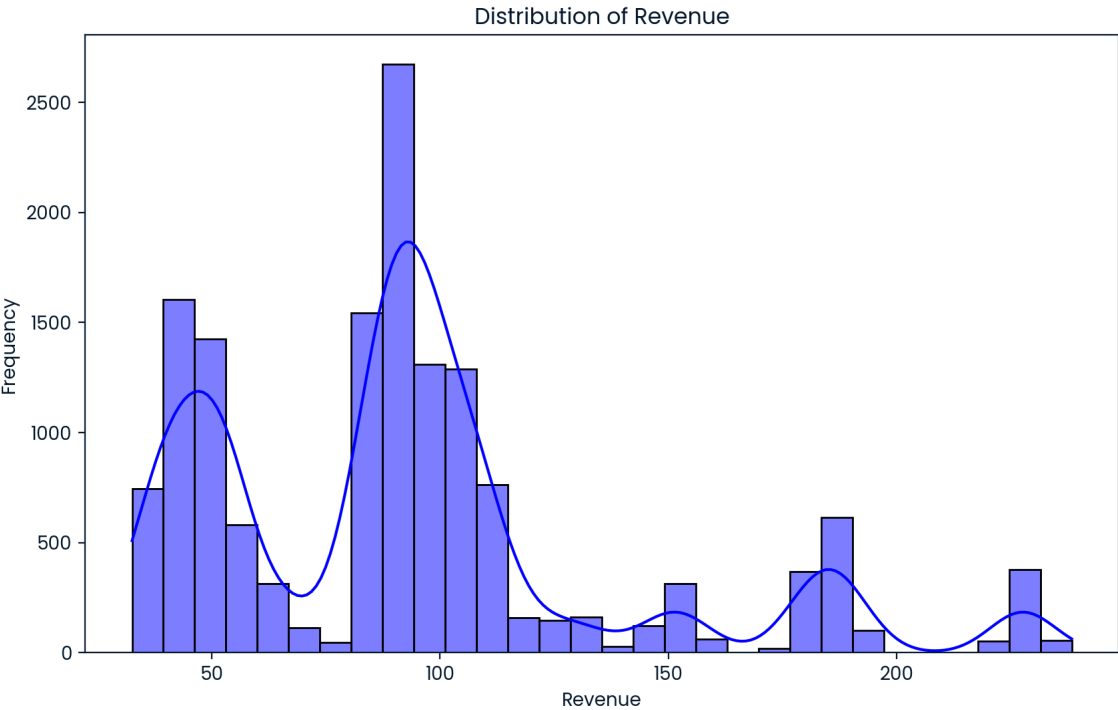
5. **Final summary and recommendations:** Business should focus more on email+call as a sales method. From the data, it can be seen that this method bring the highest amount of revenues on average, even though the sample size is lowest. Also, business could investigate more on customers who visit the website more than 37 times on why they contribute to less revenues.

...	↑↓	week	...	↑↓	nb_sold	...	↑↓	revenue	...	↑↓	years_as_cus...	...	↑↓	nb_site_...	...	↑↓	
count		15000			15000			13926			15000			15000			
mean		3.0982666667			10.0846666667			93.9349425535			4.9659333333			24.9908666667			
std		1.6564198071			1.8122133327			47.4353122457			5.0449515589			3.5009142152			
min		1			7			32.54			0			12			
25%		2			9			52.47			1			23			
50%		3			10			89.5			3			25			
75%		5			11			107.3275			7			27			
max		6			16			238.32			63			41			

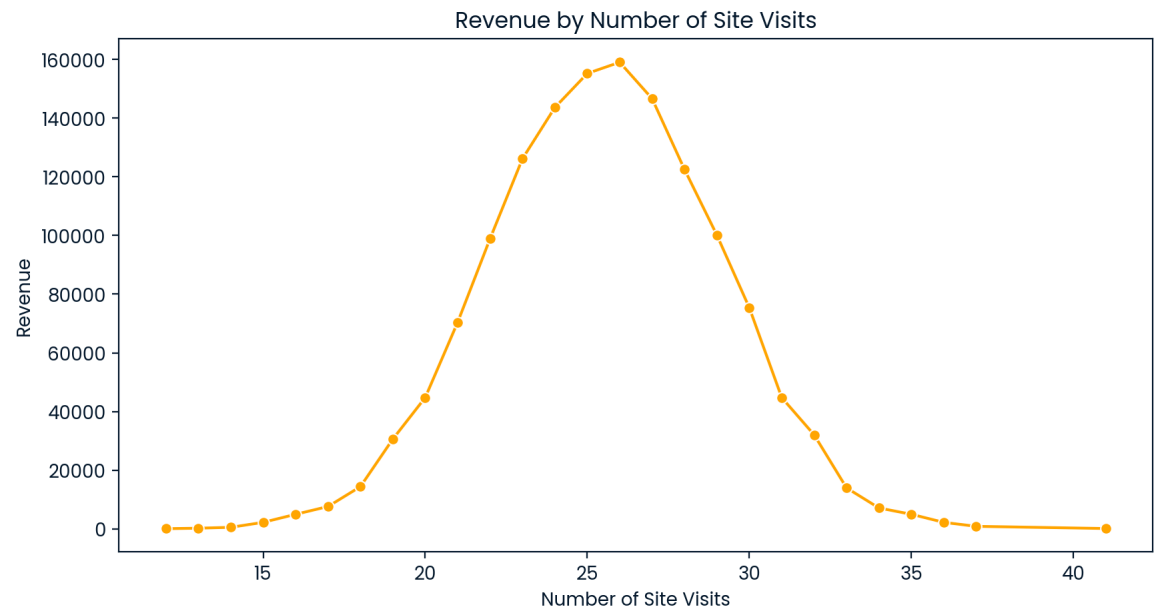
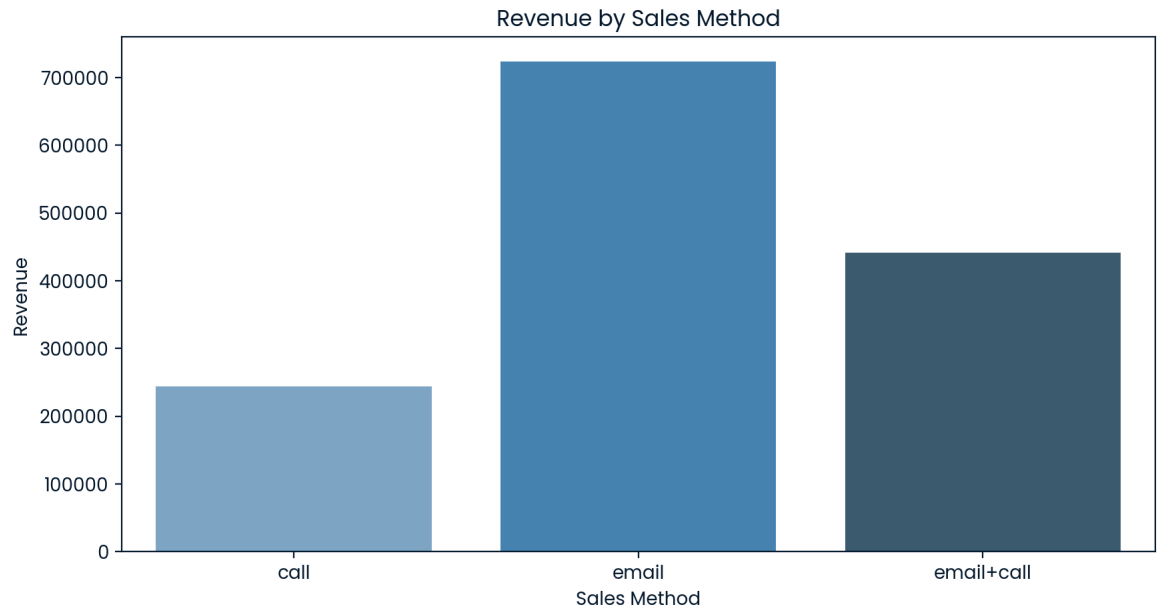
Rows: 8 ↓

...	↑↓	...	↑↓	sales...	...	↑↓	customer_id	...	↑↓	...	↑↓	...	↑↓	years_as_cus...	...	↑↓	nb_site_...	...	↑↓	state	...
	0		2	Email			2e72d641-95ac-497b-bbf8-4861764a7097				10					0		24		Arizona	
	3		4	Email			78aa75a4-ffeb-4817-b1d0-2f030783c5d7				11					3		25		Indiana	
	16		2	Email			0f744f79-1588-4e0c-8865-fdaecc7f6dd4				10					6		30		Pennsylvania	
	17		6	Email + Call			d10690f0-6f63-409f-a1da-8ab0e5388390				15					0		24		Wisconsin	
	28		5	Email			f64f8fd5-e9b7-4326-9f5d-ef283f14d7ad				12					4		32		Florida	
	35		1	Email			844db26b-028e-4251-ad72-b7fde9cf8ca0				9					1		26		Arizona	
	41		5	Email			027f0947-4c19-4bf4-a5f4-150f9deae826				10					1		32		New Mexico	
	108		3	Email			29895ef8-9b50-4422-8a18-598efedb3803				11					6		31		Texas	
	110		2	Email + Call			c21cd0cf-c3ab-4e39-8ea0-da55f8f09bc5				11					3		25		Louisiana	
	138		5	Email + Call			44275854-d501-4b59-89fe-52a013f09624				12					6		25		Indiana	
	139		3	Email + Call			2b68ef02-8229-498d-b9f1-79fca3145453				10					2		29		Illinois	

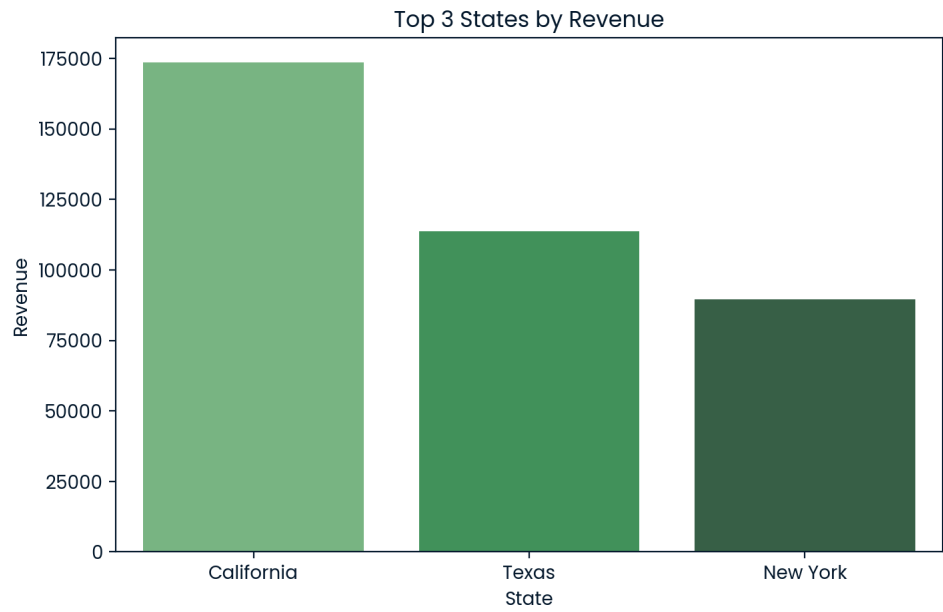
Rows: 1,074



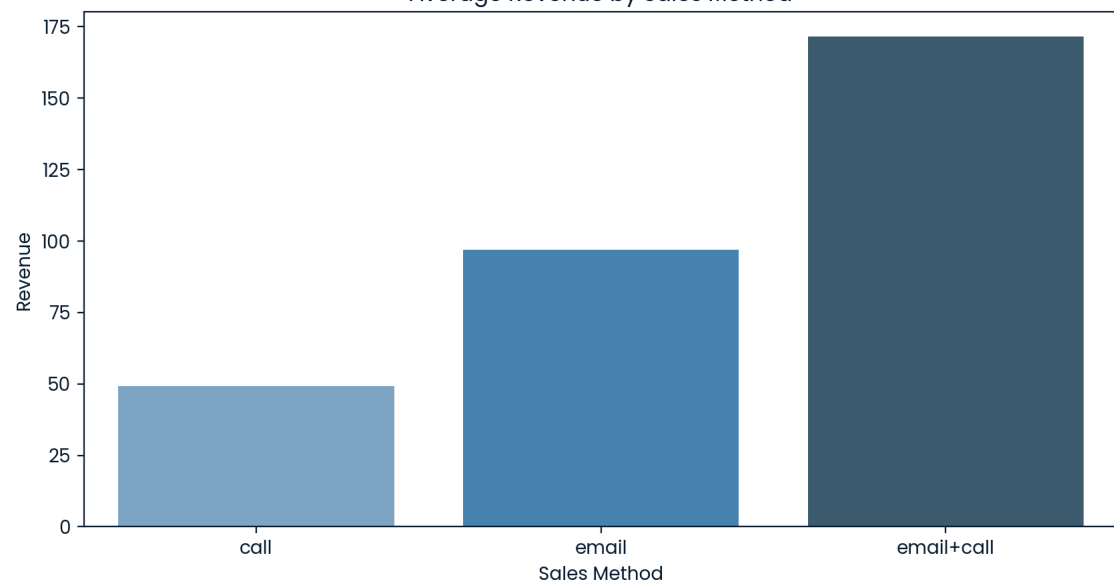
Unique sales methods after standardization: ['email' 'email+call' 'call']



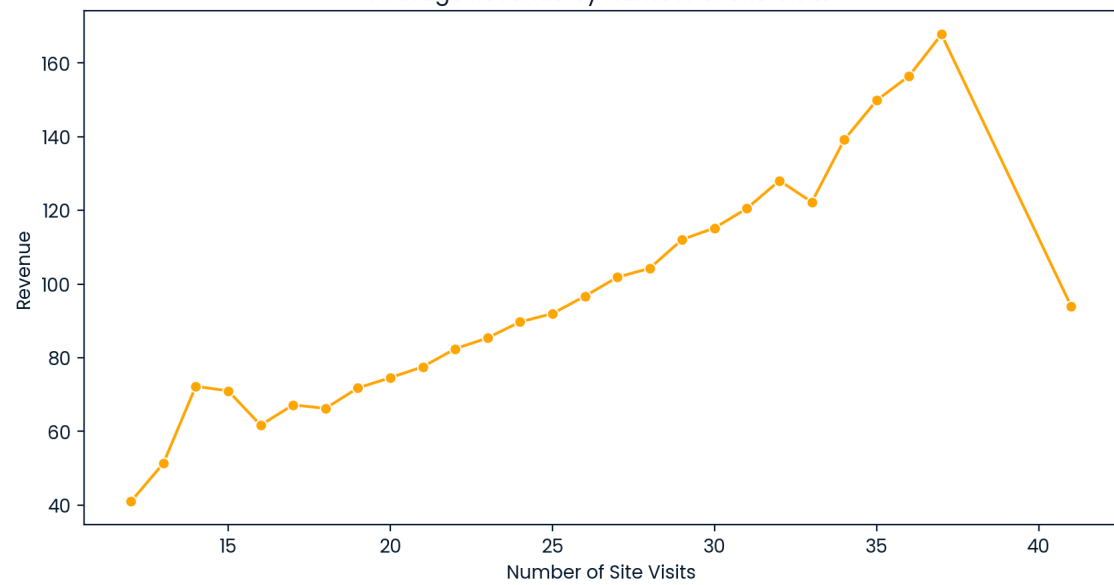
The best sales method is email with revenue of 723418.47.



Average Revenue by Sales Method

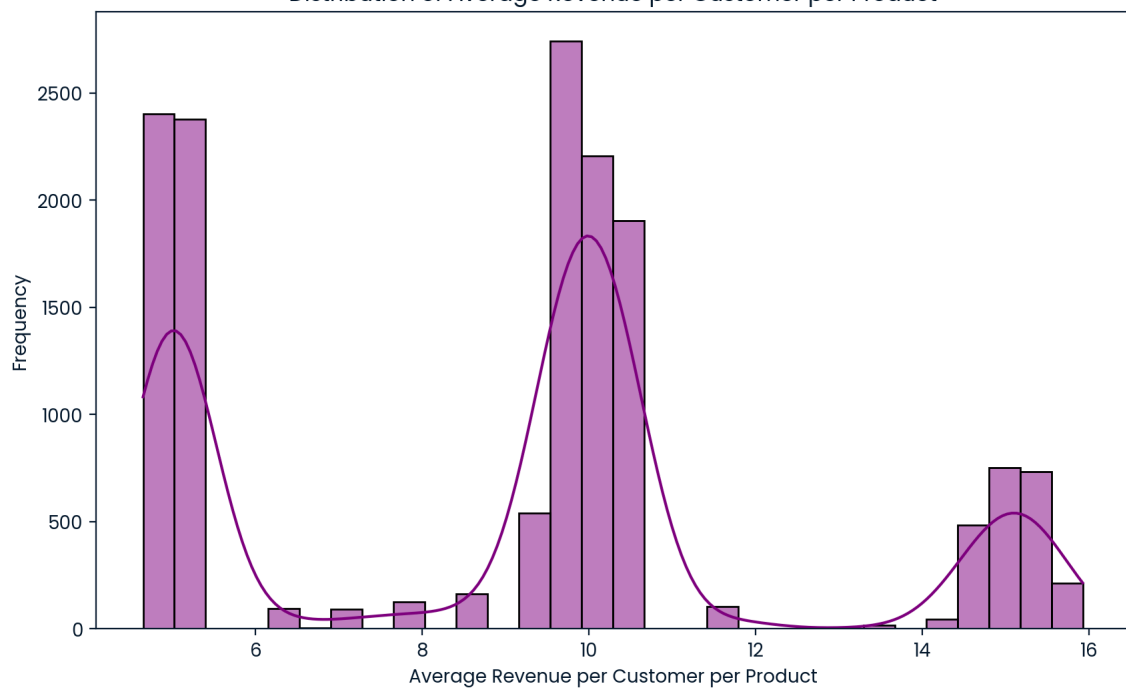


Average Revenue by Number of Site Visits



The best sales method is email+call with revenue of 171.47745139968896.

Distribution of Average Revenue per Customer per Product



Box Plot of Average Revenue per Customer per Product

