PENS AND PRINTERS

Sales Approaches' Analysis

Context

Six weeks ago, **Pens and Printers** launched a new line of office stationery. They used three different sales strategies for this: targeted email and phone calls, as well as combining the two. This report helps them understand how the sales performed during these six weeks.

1. Data validation

The dataset underwent rigorous validation and cleaning to ensure accuracy. Table 1 shows the original database as it was received — some null values for revenue can be observed.

Table 1. The original database file first ten rows.

	week	sales_method	customer_id	nb_sold	revenue	years_as_customer	nb_site_visits	state
0	2	Email	2e72d641-95ac-497b-bbf8-4861764a7097	10	NaN	0	24	Arizona
1	6	Email + Call	3998a98d-70f5-44f7-942e-789bb8ad2fe7	15	225.47	1	28	Kansas
2	5	Call	d1de9884-8059-4065-b10f-86eef57e4a44	11	52.55	6	26	Wisconsin
3	4	Email	78aa75a4-ffeb-4817-b1d0-2f030783c5d7	11	NaN	3	25	Indiana
4	3	Email	10e6d446-10a5-42e5-8210-1b5438f70922	9	90.49	0	28	Illinois
5	6	Call	6489e678-40f2-4fed-a48e-d0dff9c09205	13	65.01	10	24	Mississippi
6	4	Email	eb6bd5f1-f115-4e4b-80a6-5e67fcfbfb94	11	113.38	9	28	Georgia
7	1	Email	047df079-071b-4380-9012-2bfe9bce45d5	10	99.94	1	22	Oklahoma
8	5	Email	771586bd-7b64-40be-87df-afe884d2af9e	11	108.34	10	31	Massachusetts
9	5	Call	56491dae-bbe7-49f0-a651-b823a01103d8	11	53.82	7	23	Missouri

Approaching the quality of the data by investigating its unique and null values is common practice, and with this, in Table 2 it is possible to observe the points that should be maintained. In Table 3 one can observe the transformed data, which has been modified with the following manipulations:

- **Missing values**:
    - **revenue**: 1,074 missing values were removed by filtering non-null entries.
    - The other columns had no missing values.
- **Regex adjustments & typos**:
    - **sales_method**: Standardized entries (e.g., "Email" → "email", "em + call" → "email + call").
    - **state**: Validated against U.S. state names and converted to lowercase.

- **Case adjustment**:
    - **customer_id**, **state**, and **sales_method** converted to lowercase for consistency.
- **Handling outliers:**
    - 733 values were calculated as outliers, hence dropped from the database. Worth mentioning one value from the field “years as customer”, was 64 years. This value is notably wrong, given that **Pens and Printers** only exist for 41 years. A boxplot (Figure 1) served as visualization to first infer about outliers.

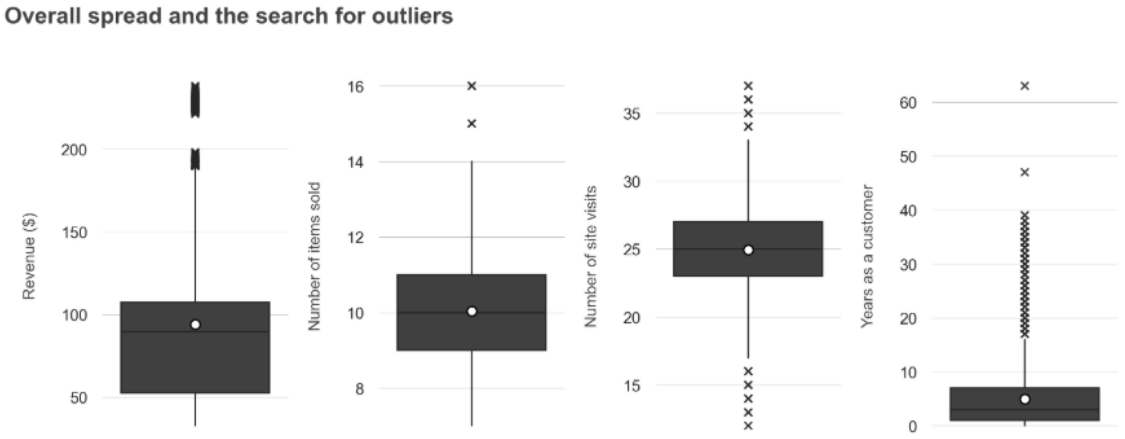

Figure 1. First look at outliers.

Table 2. Unique and null values for the original database.

	Column	Top 5 Unique Values	Unique Values Count	Null Values Count
0	week	{1: 3721, 4: 2575, 5: 2574, 2: 2491, 3: 2411}	6	0
1	sales_method	{'Email': 7456, 'Call': 4962, 'Email + Call': 2549, 'em + call': 23, 'email': 10}	5	0
2	customer_id	{'00019f95-cd18-4a2a-aa62-512cc6b17ac5': 1, 'a907f16b-2481-4ac6-95bd-6161e15cbe08...	15000	0
3	nb_sold	{10: 3677, 9: 3535, 11: 2439, 8: 1941, 12: 1088}	10	0
4	revenue	{51.86: 11, 53.25: 10, 52.51: 10, 98.42: 9, 51.58: 9}	6744	1074
5	years_as_customer	{1: 2504, 2: 1987, 3: 1619, 0: 1471, 4: 1338}	42	0
6	nb_site_visits	{25: 1688, 26: 1645, 24: 1601, 23: 1478, 27: 1440}	27	0
7	state	{'California': 1872, 'Texas': 1187, 'New York': 965, 'Florida': 904, 'Illinois': ...	50	0

Table 3. Unique and null values after the transformation.

	Column	Top 5 Unique Values	Unique Values Count	Null Values Count
0	week	{1: 3295, 4: 2300, 5: 2261, 2: 2224, 3: 2151}	6	0
1	sales_method	{'email': 6523, 'call': 4579, 'email + call': 2091}	3	0
2	customer_id	{'00020b38-1ebb-427f-88d1-c60f3c426cfe': 1, 'a9620439-f29c-4e96-905a-2b2deb9ea535...	13193	0
3	nb_sold	{10: 3275, 9: 3184, 11: 2194, 8: 1719, 12: 918}	10	0
4	revenue	{52.51: 10, 51.86: 10, 51.91: 9, 53.25: 9, 53.69: 9}	6498	0
5	years_as_customer	{1: 2302, 2: 1809, 3: 1482, 0: 1332, 4: 1209}	17	0
6	nb_site_visits	{25: 1491, 26: 1456, 24: 1414, 23: 1316, 27: 1289}	22	0
7	state	{'california': 1645, 'texas': 1044, 'new york': 849, 'florida': 770, 'illinois': ...	50	0

Table 4, A and B, illustrate the descriptive statistics of the fields. It is possible to observe that the data transformation discussed previously has minor effects on measures of spread.

Table 4. Descriptive statistics.

	week	nb_sold	revenue	years_as_customer	nb_site_visits
count	15000.0	15000.0	13926.0	15000.0	15000.0
mean	3.1	10.1	93.9	5.0	25.0
std	1.7	1.8	47.4	5.0	3.5
min	1.0	7.0	32.5	0.0	12.0
25%	2.0	9.0	52.5	1.0	23.0
50%	3.0	10.0	89.5	3.0	25.0
75%	5.0	11.0	107.3	7.0	27.0
max	6.0	16.0	238.3	63.0	41.0

A) Original data.

	week	nb_sold	revenue	years_as_customer	nb_site_visits
count	13193.0	13193.0	13193.0	13193.0	13193.0
mean	3.1	10.0	93.5	4.3	25.0
std	1.6	1.8	47.4	3.8	3.4
min	1.0	7.0	32.5	0.0	15.0
25%	2.0	9.0	52.4	1.0	23.0
50%	3.0	10.0	89.3	3.0	25.0
75%	4.0	11.0	106.8	6.0	27.0
max	6.0	16.0	238.3	16.0	36.0

B) Transformed data.

2. Exploratory analysis

Answering customer questions and showing findings.

The original demand had four driven questions to answer:

- Q1: “How many customers were there for each approach?”,
- Q2: “What does the spread of the revenue look like overall? And for each method?”,
- Q3: “Was there any difference in revenue over time for each of the methods?”, and
- Q4: Based on the data, which method would you recommend we continue to use? Some of these methods take more time from the team so they may not be the best for us to use if the results are similar.

The data analysis made it possible to answer the questions as follows.

Q1: “How many customers were there for each approach?”

- **Key insight/answer**: Email was the most common method (6,523 customers), followed by call (4,579) and email + call (2,091), as one can observe in Figure 2.
- **Other thoughts**: Simpler methods (email) reach more customers, while combined methods (email + call) are less used.

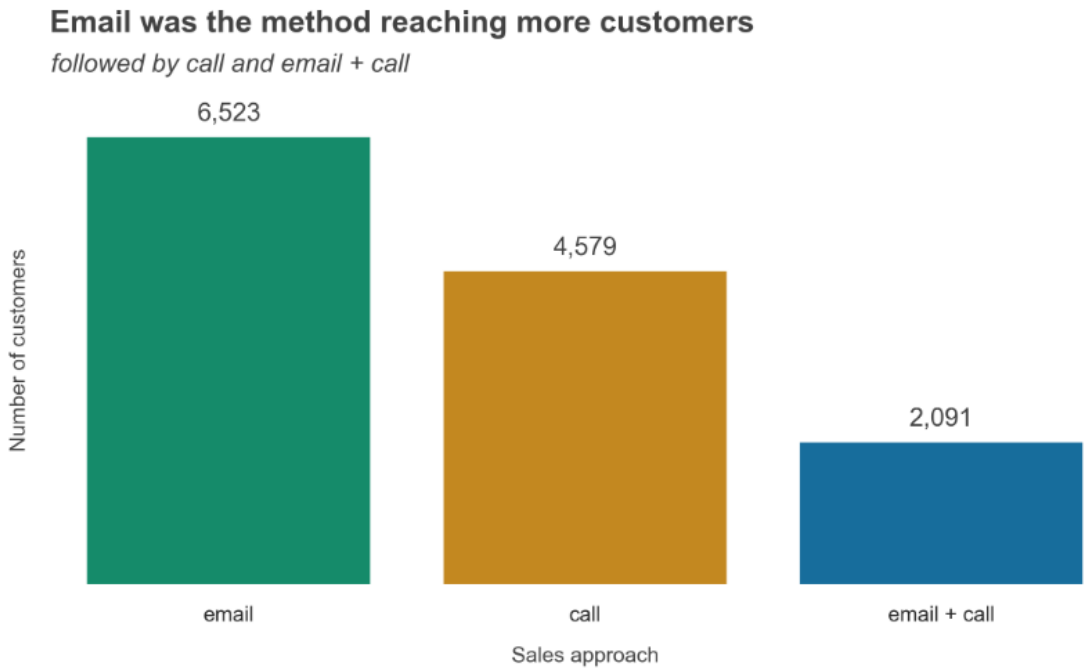

Figure 2. The number of customers reached by each method.

Q2: "What does the spread of the revenue look like overall? And for each method?"

- **Key insight/answer**: Email + call has a wider spread of revenue and number of units sold and a higher average revenue, as is seen in Figure 3. Also, the approach of email + call has significantly higher median revenue than the others — Figure 4[1]. Another way of looking to spread is with a violin plot (Figure 5). One can see the density of the approach "email + call" is closer to the median.

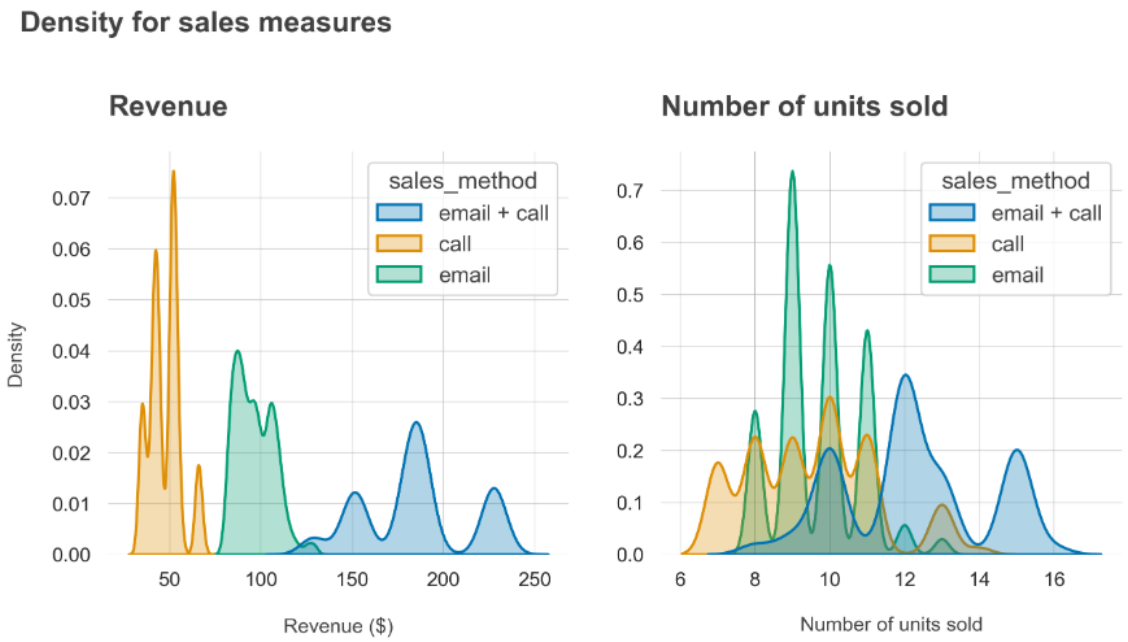

Figure 3. Density for key figures of the sales: revenue and number of units sold, grouped by the sales approach.

[1] In Figure 4 one can see outliers are still showing up, despite the handling made previously. However, it is not a good practice to remove outliers a second time, without a deeper understanding of the database.

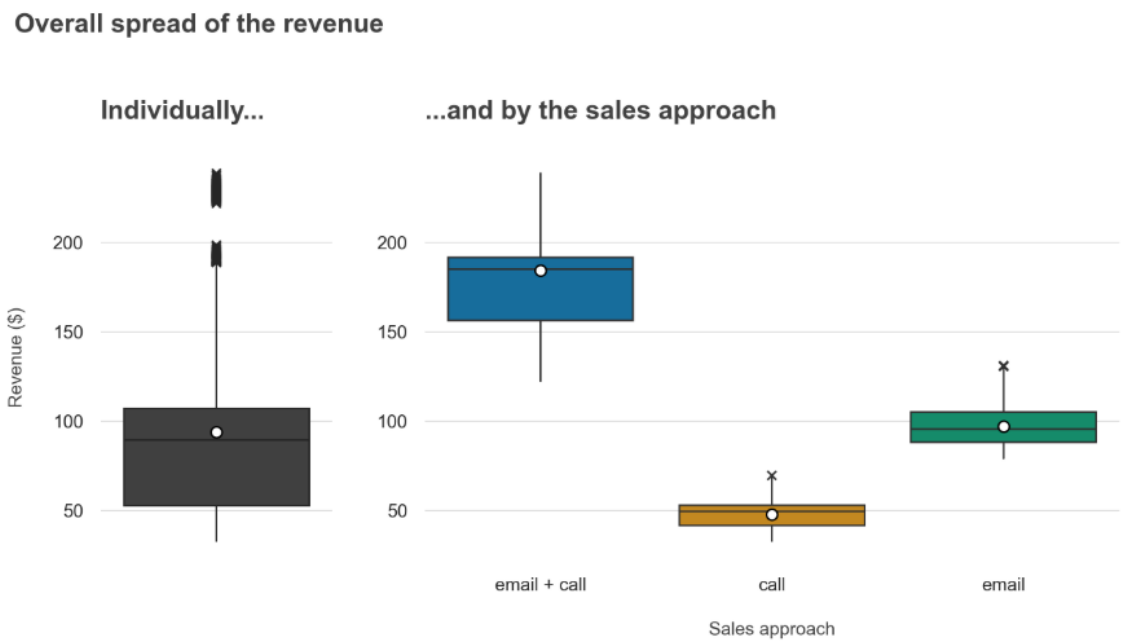

Figure 4. Spread of the revenue — alone and by each sales approach.

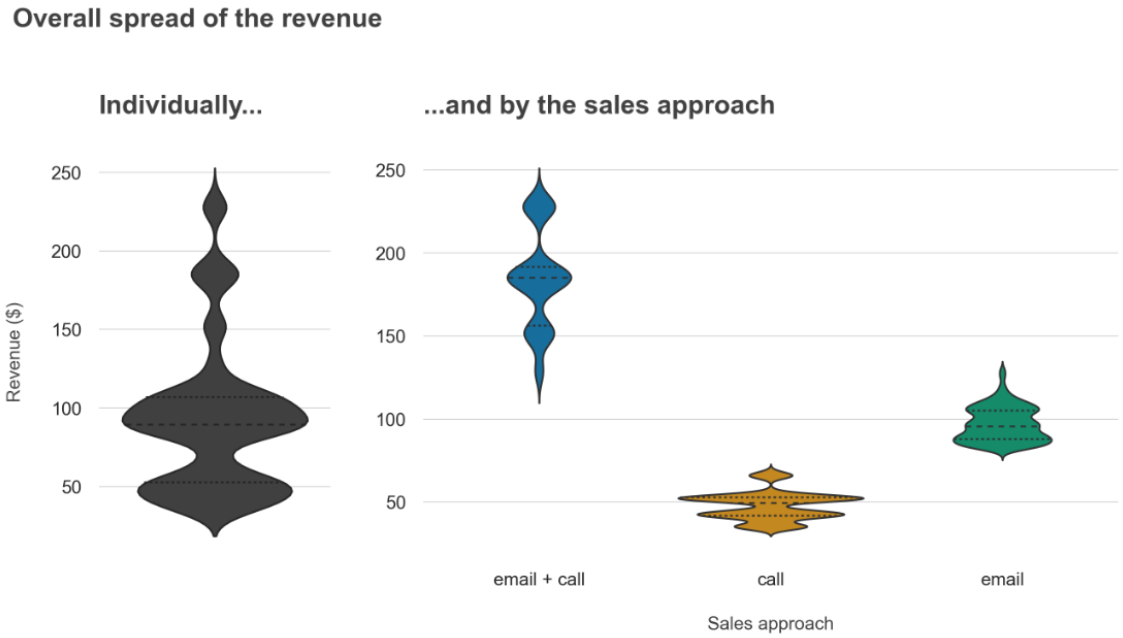

Figure 5. A violin plot to see spread and density in the same plot.

Q3: "Was there any difference in revenue over time for each of the methods?"

- **Key insight/answer**: The average revenue from the approach of email + call surpasses the other methods since the first week, as seen in Figure 6. It is also the approach that has more growth in terms of average revenue over time.
- **Other thoughts**: Combined methods (email and call) drive long-term revenue growth.

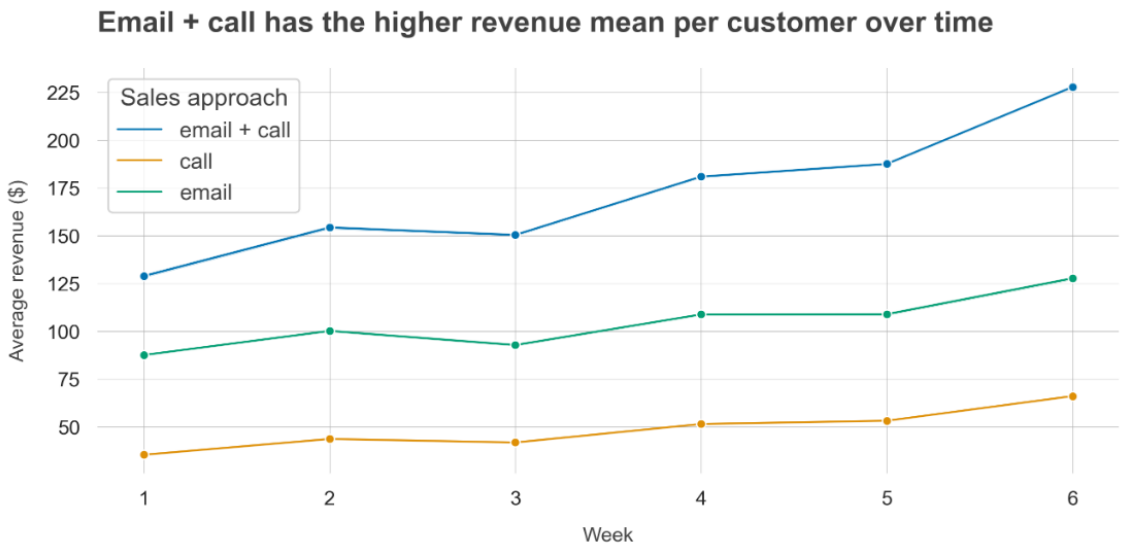

Figure 6. Mean revenue for each sales approach.

3. Monitoring strategy

The marketing team at **Pens and Papers** should have a "monitoring strategy". For that, the first thing is to choose a metric. One possibility is to track the **Average Revenue per Customer (ARPC)** by sales approach. Tracking ARPC helps prioritize high-value methods.

The value of ARPC for each sales approach in the first week is as follows:

- **Call**: $35.36,
- **Email**: $87.56,
- **Email + call**: $128.86.

The values evolve as one can see in Figure 6 or in Table 5.

Table 5. Key indicators for the three sales approaches.

		revenue					customer_id
		mean	median	sum	max	min	count
sales_method	week						
call	1	35.36	35.21	24507.65	40.47	32.54	693
	2	43.61	43.47	32492.37	50.19	40.84	745
	3	41.78	41.51	34722.52	50.66	38.97	831
	4	51.45	51.26	48565.54	57.74	48.21	944
	5	53.15	52.97	51606.65	58.01	50.40	971
	6	66.04	65.88	26087.54	69.47	63.54	395
email	1	87.56	86.59	217574.69	112.88	78.83	2485
	2	100.16	99.22	132405.24	124.24	91.56	1322
	3	92.76	91.80	94519.55	117.46	84.82	1019
	4	108.77	108.08	102790.53	126.01	100.59	945
	5	108.88	108.02	71316.23	127.87	101.29	655
	6	127.66	127.83	12383.02	130.85	123.48	97
email + call	1	128.86	128.59	15076.14	136.36	122.11	117
	2	154.27	154.27	24220.75	164.14	149.11	157
	3	150.37	149.84	45260.36	159.55	144.51	301
	4	180.87	180.62	74338.88	192.09	174.07	411
	5	187.55	187.12	119091.98	198.32	181.33	635
	6	227.79	227.40	107060.70	238.32	221.41	470

The following list of items from “a)” to “g)” explore the ARPC metric with more detail.

a) Identifies high-value sales strategies

- **Key insight**: The analysis showed stark differences in ARPC across methods:
  - **Email + call**: $183.65,
  - **Email**: $97.13,
  - **Call**: $47.60.
- **Why it matters**:
  - Email + call generates **3.8x more revenue per customer** than calls and **1.9x more** than email alone,
  - Tracking ARPC helps prioritize methods that deliver the highest value per customer, even if they serve fewer people.

b) Balances volume vs. value trade-offs

- **Observation**:

- **Email** had the **highest customer volume** (6,922) but **moderate ARPC**,
- **Email + call** had the **lowest volume** (2,223) but **the highest ARPC**.

- **Strategic implications**:
  - Businesses can use ARPC to decide whether to focus on scaling high-ARPC methods (e.g., expanding email + call adoption) or improving volume-driven methods (e.g., improving email conversion rates).

c) Guides resource allocation

- **Example**:
  - If email + call requires more effort (e.g., personalized calls), tracking ARPC ensures the extra cost is justified by higher returns,
  - Conversely, low ARPC for calls ($47.60) suggests reallocating resources to more effective methods.

d) Highlights customer segmentation opportunities

- **Insight**:
  - High ARPC methods (e.g., email + call) likely target **high-value customers** who respond to personalized outreach,
  - Low ARPC methods (e.g., call) may attract price-sensitive or transactional buyers.
- **Action**:
  - Use ARPC to tailor strategies: deploy email + call for premium customers, and email for mass outreach.

e) Monitors sales team performance

- **Use case**:
  - If ARPC for email + call declines over time, it could signal issues like:
    - Market saturation,
    - Declining customer quality, or
    - Ineffective sales tactics.
  - Initiative-taking adjustments (e.g., retraining staff) can mitigate risks.

f) Complements other metrics

- **Holistic view**:
  - Pair ARPC with **customer acquisition cost (CAC)** to calculate profitability (e.g., ARPC/CAC ratio),
  - Combine with **customer lifetime value (CLV)** to assess long-term impact.

g) Tracks impact of strategic initiatives

- **Example**:
    - If the business invests in training call teams to upsell, ARPC for calls should rise,
    - If email campaigns are redesigned to target high-value segments, email ARPC should improve.

4. Final summary and recommendations

After broad data analysis, the recommendations for the sales team at Pens and Papers are:

a) Expand Email + Call Adoption:

- Train sales teams to use this method for high-value customers,
- Incentivize customers (e.g., discounts) to opt for combined communication.

b) Optimize email campaigns:

- Improve email content to boost conversion rates, as email serves the largest audience.

c) Monitor ARPC weekly:

- Track ARPC by method to assess strategy effectiveness,
- Aim to increase the customer base that is approached with email + calls while keeping its high ARPC.

Conclusion: Prioritizing email + call sales approach while refining email campaigns can potentially maximize revenue growth. Continuous monitoring of ARPC ensures alignment with business goals.