Pens and Paper Sales Method Analysis Project Report

Data Validation

The dataset contained **15000 rows and 8 columns** before cleaning and validation. I have validated the entire dataset. The following describes how each column was validated:

- **week:** There are **six weeks** in the data set, matching the number of weeks since the new product line was launched. There are no missing values in the week data set.No cleaning is needed.
- **sales_method:** There are no missing values in the sales_method dataset. Validated by checking for character data type, same as the description. To clean, I converted all characters to lower case. I then converted "em + call" to "email + call" throughout the column. The dataset now has **three** sales methods **categories**: "**email**", "**call**", and "**email + call**". `product_sales$sales_method <- tolower(product_sales$sales_method) product_sales$sales_method[product_sales$sales_method == "em + call"] <- 'email + call'`
- **customer_id:** There are **15000 unique values** in the customer_id dataset. Validated by checking character type and unique rows. Data matches detail requirements. There are no missing values in the customer_id dataset. No cleaning is needed.
- **nb_sold:** Validated by checking for numeric data type. The nb_sold data set has no missing values and matches the detail requirements. No cleaning is needed.
- **revenue:** Validated by checking for numeric data type. There were **1074 missing values** in the revenue data set. This was cleaned by taking the mean of the revenue by nb_sold groups and applying the mean revenue of each nb_sold group to the appropriate missing values. I chose to apply the mean of each nb_sold group, rather than the entire data set, as the revenue was strongly correlated with increasing nb_sold values. I then rounded the dataset to two decimal places. There are now no missing values in the revenue data set. `product_sales <- product_sales %>% group_by(nb_sold) %>% mutate_at(vars(revenue), ~replace_na(., mean(., na.rm = TRUE))) product_sales$revenue <- round(product_sales$revenue, 2)`
- **years_as_customer:** The years_as_customer data set was validated by checking for numeric datatype. I removed the two rows of data that exceeded the maximum years one could be a customer. The data set now contains only values from **0 to 39**. The entire data set now has **14998** rows. `product_sales <- product_sales %> % filter(years_as_customer < 40)`
- **nb_site_visits:** The nb_site_visits data was validated by checking for numeric datatype. The nb_site_visits data set has no missing values and matches the detail requirements. No cleaning is needed.

- **state:** The state data set was validated by checking for character datatype. The state data set has no missing values and matches the detail requirements as there are **50** different **states** represented in the dataset. No cleaning is needed.

After data validation, the data set conatains **14998 rows and 8 columns** without missing values.

How many customers were there for each approach?

There were three sales appoaches used for the product launch. There were 7,465 "e-mail" only customers, accounting for 49.77% of all customers. There were 4,961 "call" only customers, accounting for 33.08% of all customers. There were 2,572 "email + call" customers, accounting for 17.15% of all customers. #1

What does the spread of revenue look like?

The spread of revenue shows a distribution of revenue from a minimum of 32.54 to a maximum of 238.32, with multimodal distribution. The box plot shows that 50% of revenue ranged between 53.00 and 107.60, with a median revenue of 89.70. #2.a

#2.b (2) (1)

What does the spread of revenue look like for each sales approach?

The multimodal distribution of revenue is explained easily by the sales method used. The "call" sales method is associated with the lower end spread of revenue, having a median of 49.9. The "e-mail" sales method is associated with the average spread of revenue as the median of the "e-mail" sales method of 94.4 is close to the median of the entire spread of revenue at 89.7. The "e-mail + call" sales method is associated with the upper end spread of revenue, having a median of 183.7. #2.a (2)

#2.b (3)

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

All three sales methods differed significantly in regards to their difference of revenue across the 6 weeks the data was collected.

- The "call" sales method started week 1 with 26,820 in revenue, accounting for 9% of total revenue in week 1 and it ended week 6 with 30,793 in revenue, accounting for 16% of total revenue in week 6. The "call" sales method saw revenue increase by 14% through the first six weeks of the product launch.

- The "e-mail" sales method started week 1 with 243,881 in revenue, accounting for 84% of total revenue in week 1 and it ended week 6 with 25,686 in revenue, accounting for 13% of total revenue in week 6. The "e-mail" sales method saw revenue decrease by 89% through the first six weeks of the product launch.
- The "e-mail + call" sales method started week 1 with 18,158 in revenue, accounting for 6% of total revenue in week 1 and it ended week 6 with 132,708 in revenue, accounting for 70% of total revenue in week 6. The "e-mail + call" sales method saw revenue increase by 630% through the first six weeks of the product launch. rotsm

Differences between customers in each group:

Differences between average revenue and site visits by sales method:

For every number of site visits:

- the "call" sales method saw an increase of 1.60 in average revenue.
- the "e-mail" sales method saw an increase of 0.99 in average revenue.
- the "e-mail + call" sales method saw an increase of 4.16 in average revenue. #4.i (2)

Differences between number of customers through first six weeks by sales method:

Through the first six weeks:

- the "call" sales method saw a slight drop in customers, losing 20 customers per week.
- the "e-mail" sales method saw a sharp drop in customers, losing 439 customers per week.
- the "e-mail + call" sales method was the only group to see an increase in customers, gaining 116 customers per week.

metrics (4)

Business Metrics

Since our goal is to use the best sales approach to ensure the new product line is a success, I would recommend we use a **6 week revenue growth rate percentage** as our metric.

Based on the last 6 weeks of data, we had a 6 week revenue growth rate percentage of -34.5%. Therefore, if the recommended sales method is successful, the 6 week revenue growth rate percentage should increase, and could turn positive.

Recommendation

- Based on the data, my recommendation would be to continue using the **"e-mail + call"** sales approach. The data tells us that the **"e-mail + call"** sales approach:
-
  - has the most positive revenue growth.
  - is the only sales approach that increases in customers weekly.
  - brings in more revenue per customer site visit.
  - brings in more revenue per customer order.
  - accounted for 32% of total revenue despite this approach being used on only 17% of the customers.
- For initial launch, "e-mail" only sales approach is great, but it is not sustainable. For continued sales success and revenue growth, the "e-mail + call" sales approach is the recommended sales approach.
- I would recommend targeting customers that have been with us 6 years or less, as they spend more and accounted for about 75% of all of our revenue.
- Further data collection of success / fail of sale for customers contacted could provide further insights.