TABLE OF CONTENTS

TABLE OF FIGURES

In today's highly competitive e-commerce landscape, leveraging data to drive business decisions is crucial for sustainable growth. Seller A achieved 10 billion VND in gross revenue in July and has now set an ambitious goal of doubling this figure to 20 billion VND in August. Meeting this target requires not only increasing the volume of orders but also improving conversion efficiency, customer value, and overall operational performance.

By examining historical order data from the Order Management System together with traffic performance data from Google Analytics, we will identify key findings, uncover growth opportunities, and propose a concrete action plan to achieve the revenue target.

In order to solve this case, I would divide it into 4 main phrases:

1. Overview

Before starting the analysis and working toward the target of achieving 20 B VND revenue in August, I defined a set of Business Questions. These questions guide the exploration of the dataset and help shape actionable recommendations.

With Order data:

- What was the company's business performance last month in terms of total revenue, average order value (AOV), and cancellation rate?
- How did the number of customers, revenue, and AOV change on a daily basis over the past month?
- How do discounts affect the number of orders? Do deeper discounts actually drive more transactions?
- Is revenue growth primarily driven by an increase in average order value (AOV) or by higher order volume?

With Traffic data:

- Which source is generating many new users but has a low purchase rate?
- Which small medium is showing outstanding revenue-per-session growth?
- Who is driving growth?
- Which source is wasting its potential?
- Where are there high quality visitors who haven't converted yet?

2. Data preparation

Before identifying key findings, I first needed to understand the dataset through Exploratory Data Analysis (EDA). During this process, I discovered several issues that could distort the results.

2.1 Historical sales performance data

2.1.1 Duplicate values

Some rows in the dataset were duplicated. To ensure data accuracy, I removed all duplicate records.

2.1.2 Selling price

There were instances where the selling price was recorded as 0, which is not valid. It is just a small part (0.122%) so these records were treated as errors and I decided to remove them all from the dataset.

2.1.3 Shipping fee

After reviewing the data, I noticed an unusual pattern in many orders. The shipping fee was much higher than the selling price. This is an abnormal point that needs attention and proper handling to avoid distorting the analysis results.

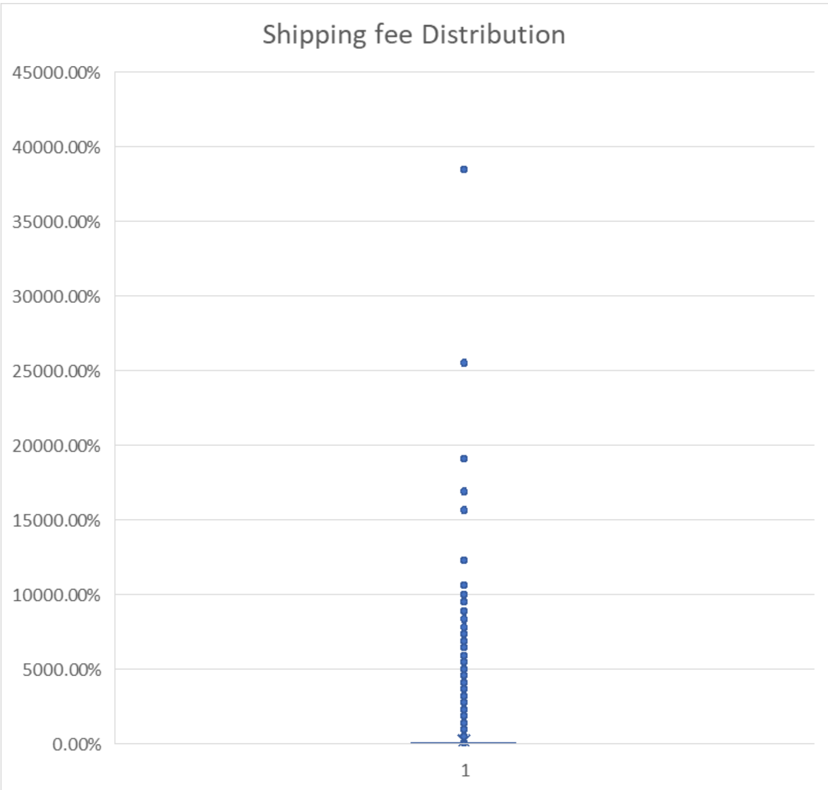

Figure 1: Boxplot of the ratio between Shipping Fee and Selling Price

The chart shows that there are many outliers with very large values, indicating unusually high shipping costs.

Figure 2: Boxplot with the scale narrowed to 100%

It reveals that 75% of the values are below approximately 16% (Q3 = 16%) of the Shipping Fee/Selling Price ratio, while the upper threshold calculated using Q3 + 1.5 * IQR is 39.77%. This means any ratio exceeding 39.77% is likely an outlier. Therefore, I decided to apply a winsorization approach:

- Keep the original shipping fee for ratios below 39.77%.
- For values exceeding the threshold, cap the ratio at 39.77%.

Here is the result before and after I apply:

Figure 3: Distribution of the Shipping Fee before adjustment

Figure 4: Distribution of the Shipping Fee after adjustment

After processing, the maximum value dropped significantly (from nearly 400,000 to below 200,000). Although the distribution remains slightly right-skewed and a few outliers still exist, the data has become cleaner and more reliable for results.

2.1.4 Create some new columns for the analysis:

To support the insight analysis, I created several new columns:

- shipping_fee_cleaned: the adjusted shipping fee after treating extreme outliers.
- discount: the price difference between onsite_original_price and selling_price. (discount = onsite_original_price − selling_price)

- discount_rate: the percentage of discount compared to the original price. (discount_rate = discount * 100 / onsite_original_price)
- revenue: the revenue of each order, calculated as: (selling_price * item_quantity + shipping_fee - voucher_platform - voucher_seller), the revenue includes both completed and cancelled orders.
- revenue_processed: similar to revenue but uses shipping_fee_cleaned instead of the original shipping_fee.

2.2 Traffic source performance data

2.2.1 Finding invalid values

In both the traffic and order datasets, the first step was to identify invalid or unreasonable values. In the Bounce Rate column, there were 24 rows with a value of 0. This is unrealistic because, no matter how good a website is, there will always be some bounce rate. This issue may stem from various reasons, such as tracking errors, misconfigured tags, or a very small number of sessions that don't reflect actual behavior. In addition, these rows also had Revenue = 0, so they did not provide any valuable information for revenue analysis. Therefore, I decided to remove these rows from the dataset.

Another special point was that some pages had an Average Session Duration as high as 3,379.36 seconds (about 56 minutes). This is quite unreasonable compared to normal user behavior.

2.2.2 Create new columns and calculations (in PowerBI)

To make the analysis easier, I created an additional column Medium by splitting it from the Source / Medium column. This allowed me to group data by intermediary channels (e.g., cpc, social, organic, email) and more easily compare and evaluate the performance of different traffic sources.

I also created New Users Rate (New Users Rate = New Users / Users), which reflects the proportion of new users among total users for each channel, helping to identify which channels are most effective at attracting new customers.

Finally, I added Revenue per Session (RPS) (RPS = Revenue / Sessions), which shows the average revenue generated per session. This metric helps compare the true value of each channel independently of traffic volume.

3. Data insights and Key findings:

3.1 Historical sales performance data

3.1.1 Company's performance

In July, the company achieved total revenue of nearly 11 billion VND, serving approximately 30.5 thousand customers and fulfilling around 37.84 thousand unique orders, equivalent to an average of 1.24 orders per customer.

The Average Order Value (AOV) was about 275,000, reflecting customers' average spending per transaction. The order completion rate remained at 80% (30.27 thousand orders), while the cancellation rate reached 20% (7.56 thousand orders). This high cancellation rate is a significant concern, as it reduces overall revenue. The company should investigate the root causes behind this situation and implement measures to reduce cancellations.

3.1.2 Trend

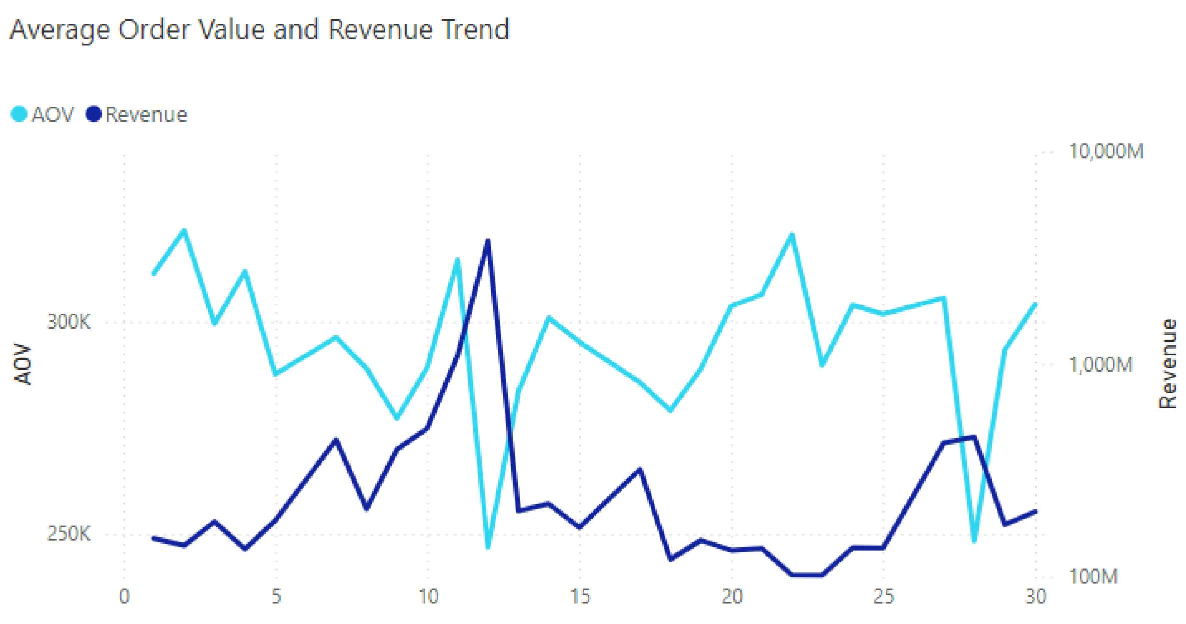

Figure 5: OV and Revenue Trend

- The Average Order Value (AOV) changed a lot last month, especially on the 12th and 28th, when it almost reached the lowest point (about 247,000), around 15–20% lower than on other days.
- Revenue was more stable. It went up from the 1st to the 12th, then dropped and stayed steady for about 10 days, and finally started to rise again in the last week of the month.

- From the 9th to the 11th, both AOV and revenue went up a lot and showed a clear positive correlation.
- Around the 30th, AOV rose a lot, but revenue did not grow the same way. This may mean the number of orders went down while the value per order went up, or revenue was limited by other things, such as discounts or not enough customers.
- One special point is that on the 12th, AOV fell sharply to 250.000, while revenue went up to nearly 4 billion. This may mean there was a special event that day, which made customers buy more low value products, such as a fixed low price promotion.

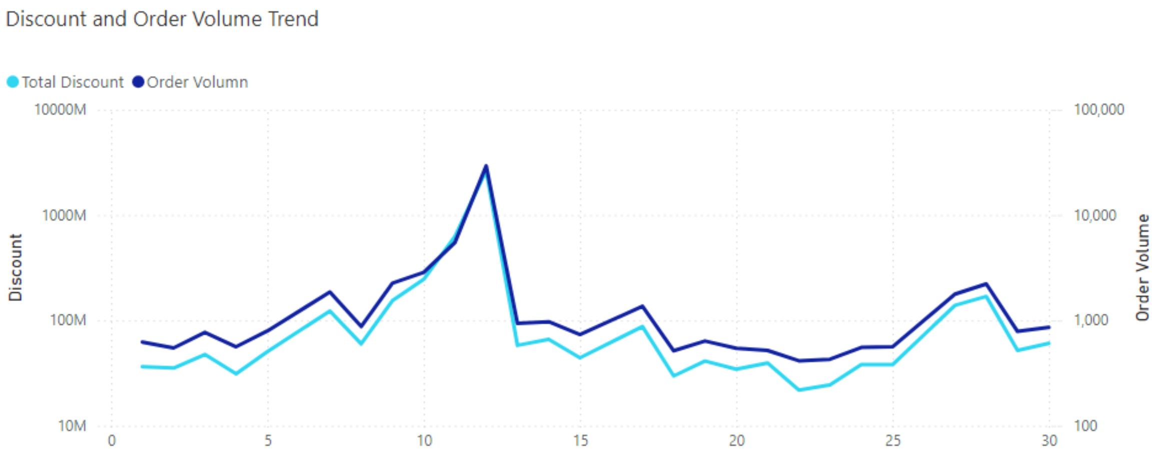

Figure 6: Discount and Order Volume Trend

- Both discounts and order volume moved in the same way. They went up at the start of the month and reached their peak on the 12th, suggesting that a big discount campaign may have pushed the number of orders up sharply. After that peak, both indicators went down together, showing that the effect of discounts might have been only temporary.

3.1.3 Frequency

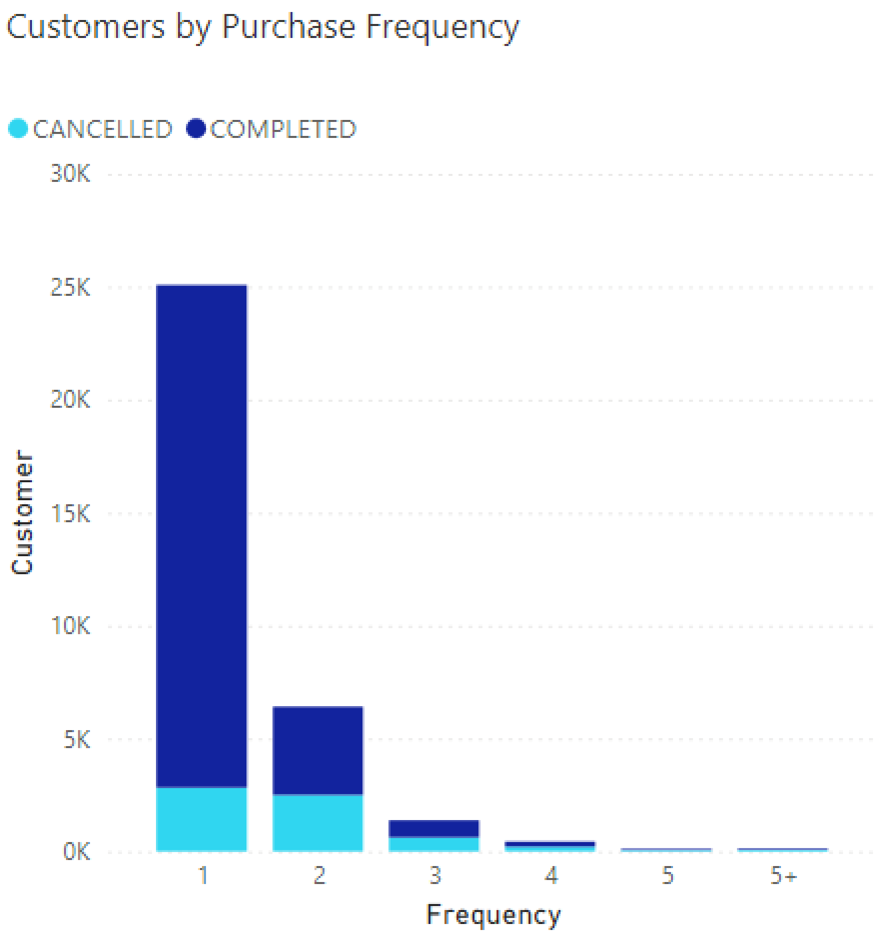

Figure 7: Customers by Purchase Frequency

Most customers make only one purchase in a month, accounting for about 80%. Around 16% of customers buy twice. This may be because the product or service quality after the first order does not meet their expectations, leading them to cancel when they try to make a second purchase. However, it is important to note that the cancellation rate for the second purchase is at an alarming rate, about 73%. The company should review possible causes such as promotions, shipping policies, and other factors to address this issue.

3.1.4 Relation between Order Value and Discount

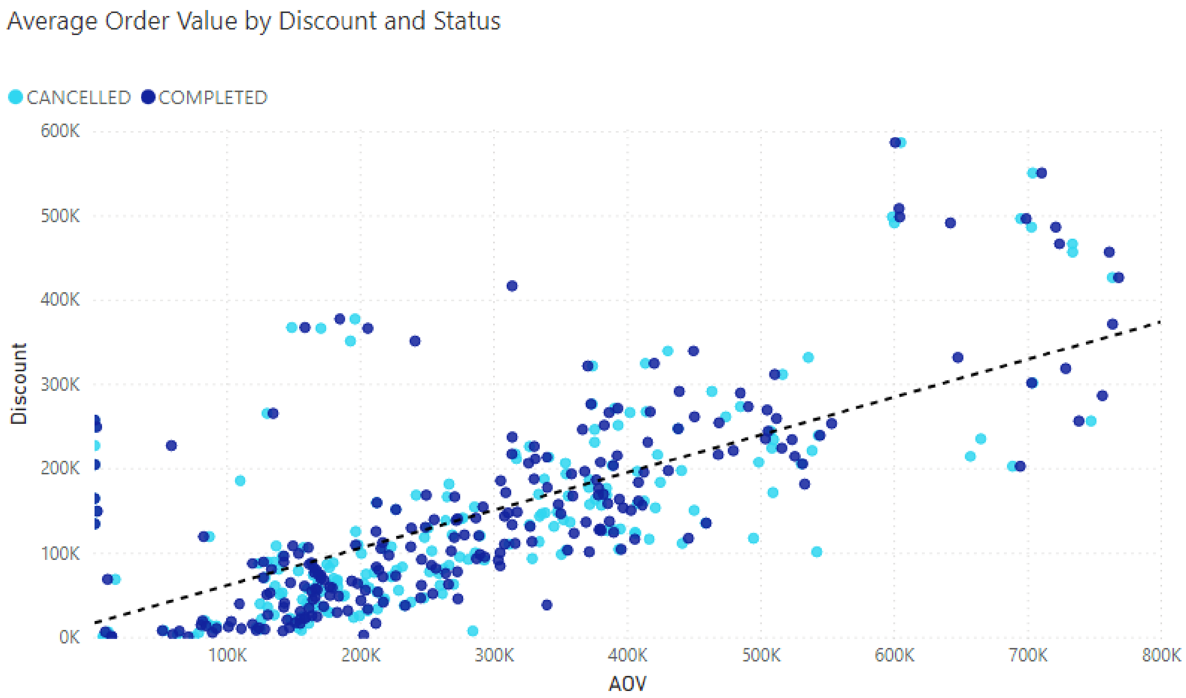

Figure 8: Average Order Value by Discount and Status

- The upward trend line indicates that as AOV increases, the discount level also tends to rise for both order statuses. Customers buying higher value orders often receive greater discounts.
- Completed orders are higher than Cancelled ones, especially at higher AOV levels. This suggests that high value orders are more likely to be successfully fulfilled. Conversely, cancelled orders mainly cluster around unusually high discounts or low AOV, reflecting higher risk in that segment.
- Most data points fall below the trend line, meaning actual discounts are often lower than the model's predicted values. This implies that the business generally keeps discount levels moderate and many customers are willing to pay higher prices without heavy promotions.

3.2 Traffic source performance data

3.2.1 Company's performance

In July, the company recorded 10 billion in revenue with more than 10,000 transactions completed. There were a total of 1.26 million sessions and 657,000 users, of which 529,670 were new users, accounting for 80.62%. This is a positive result, indicating that the business is attracting a large number of new customers.

Regarding website performance, the Bounce Rate reached 68.14%, showing that many visitors left the site immediately after entering, suggesting a need to improve content and the initial user experience. Pages per session reached 2.85, meaning that on average, users viewed nearly three pages per visit reflecting some interest, but not yet deep engagement. The average session duration was 141 seconds (around 2 minutes 20 seconds), showing that a group of users spent time exploring, this represents an opportunity for the company to guide them further along the purchase journey.

In terms of purchasing performance, Revenue per Session reached nearly 8,000, while the Conversion Rate was 0.45%. These figures highlight that although traffic volume is high, the conversion rate remains relatively low, emphasizing the need to further optimize the shopping experience and customer journey to increase revenue.

3.2.2 Channel Performance & Revenue

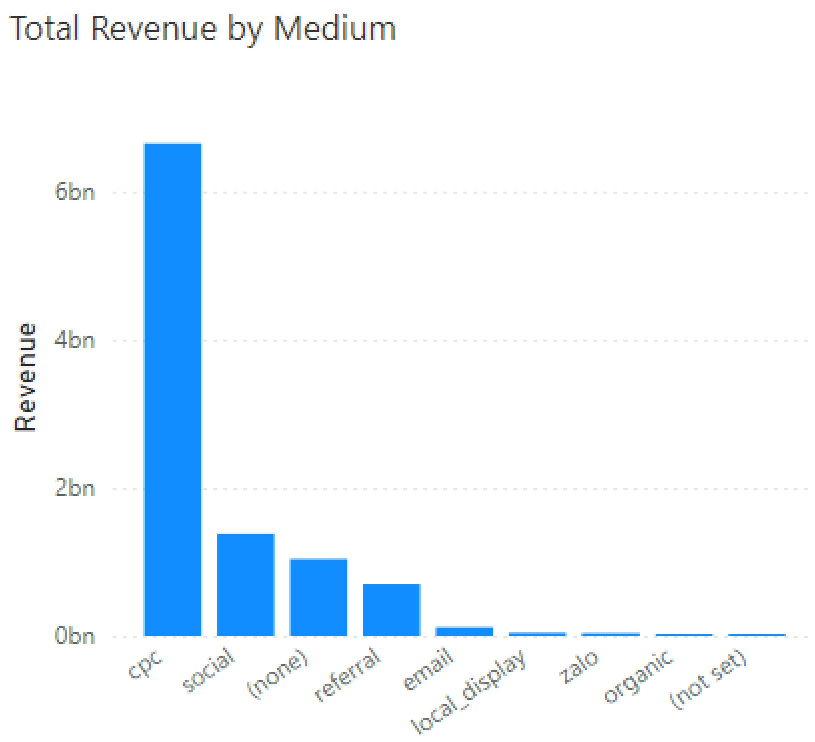

Figure 9: Total Revenue by Medium

It is evident that CPC is the Medium generating the highest revenue, with over 6.6 billion (66%), far surpassing the other mediums. Following CPC are Social (13%), None (known as Direct) (10%), and Referral (7%). The remaining mediums account for only a very small proportion, around 3–4% of total revenue.

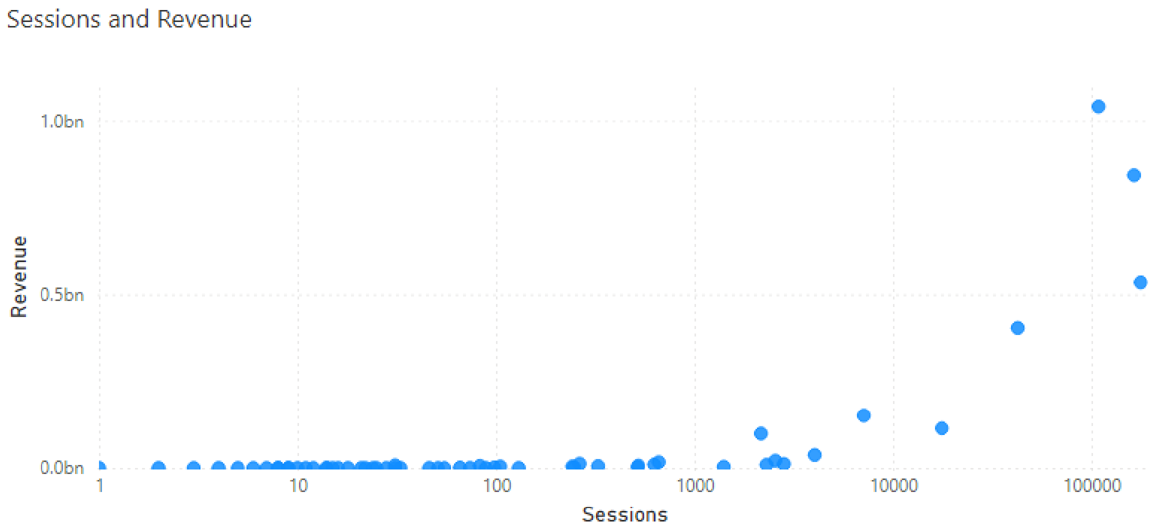

Figure 10: Relation between Sessions and Revenue

Most data points are in the area with low sessions and almost no revenue, meaning many traffic sources bring little real value. Only a few with very high sessions generate strong revenue (about 0.5 to 1 billion), showing that just a small number of channels drive most of the sales. In addition, it can be observed that several sources are underperforming in terms of potential. Some sources record a relatively solid number of sessions, even as high as around 17,000 and 42,000, yet their revenue remains unsatisfactory. Some sources have high traffic but generate very little revenue, showing their potential is being wasted. The company should focus on optimizing underperforming traffic sources and shifting resources to channels that generate higher revenue.

3.2.3 Conversion and Efficiency

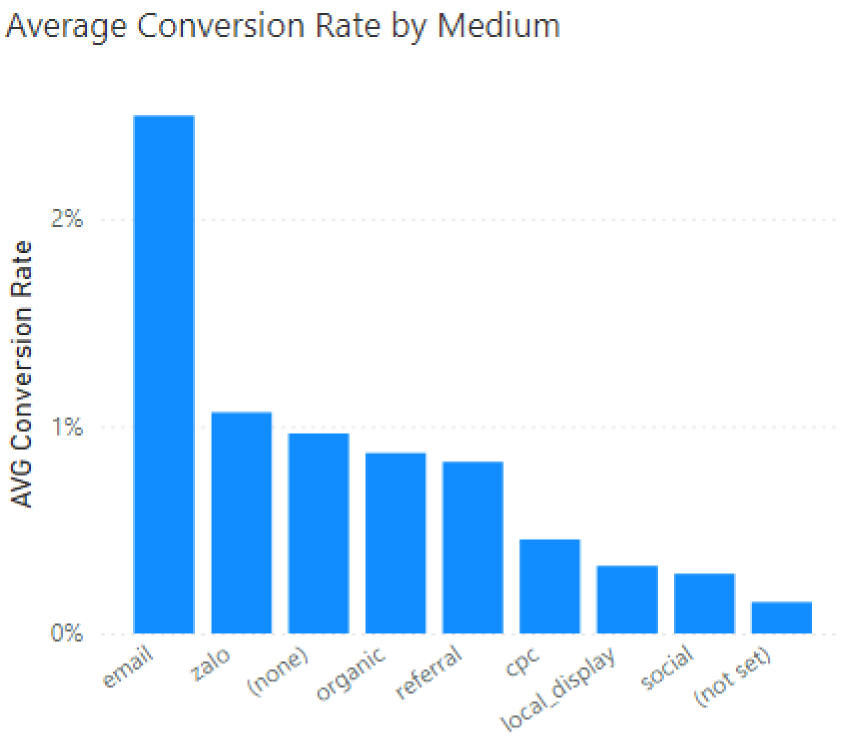

Figure 11: Average Conversion Rate by Medium

The chart shows that Email is the channel with the highest conversion rate, standing out from the others at over 2.5%. This demonstrates that email marketing is highly effective in driving purchase behavior. Although Email is not the channel that generates the highest revenue, it still achieves the best conversion rate. This indicates that customers coming from Email often already have a clear interest or demand, making them more likely to complete a purchase. In other words, Email is a focused and efficient channel for turning potential customers into actual buyers, even though its traffic volume is not as large as other channels. Following Email are Zalo, Direct (none), Organic, and Referral, with conversion rates around 1%, suggesting these are mid-level effective channels that still hold potential for optimization. On the other hand, Mediums such as CPC, Display, and Social show significantly lower conversion rates, reflecting their limitations in turning traffic into revenue.

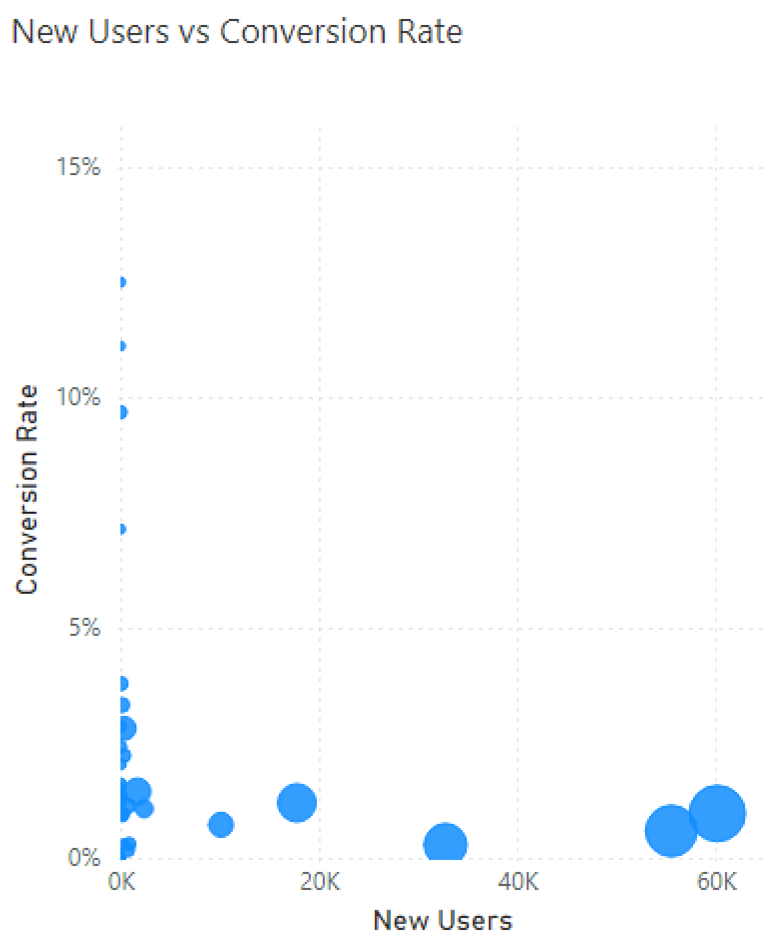

Figure 12: New Users vs Conversion Rate (Bubble Size: Revenue)

It can be observed that there are several Sources with relatively high Conversion rates, however, the number of New Users and the revenue generated remain quite low. This could stem from reasons such as these Sources targeting existing customers or those already ready to purchase, without focusing on attracting New Users, resulting in a lack of new user acquisition. On the other side, there are a few Sources that attract a large number of New Users and also generate significant revenue, but their Conversion rates are relatively low.

3.2.4 Engagement Impact on Conversion

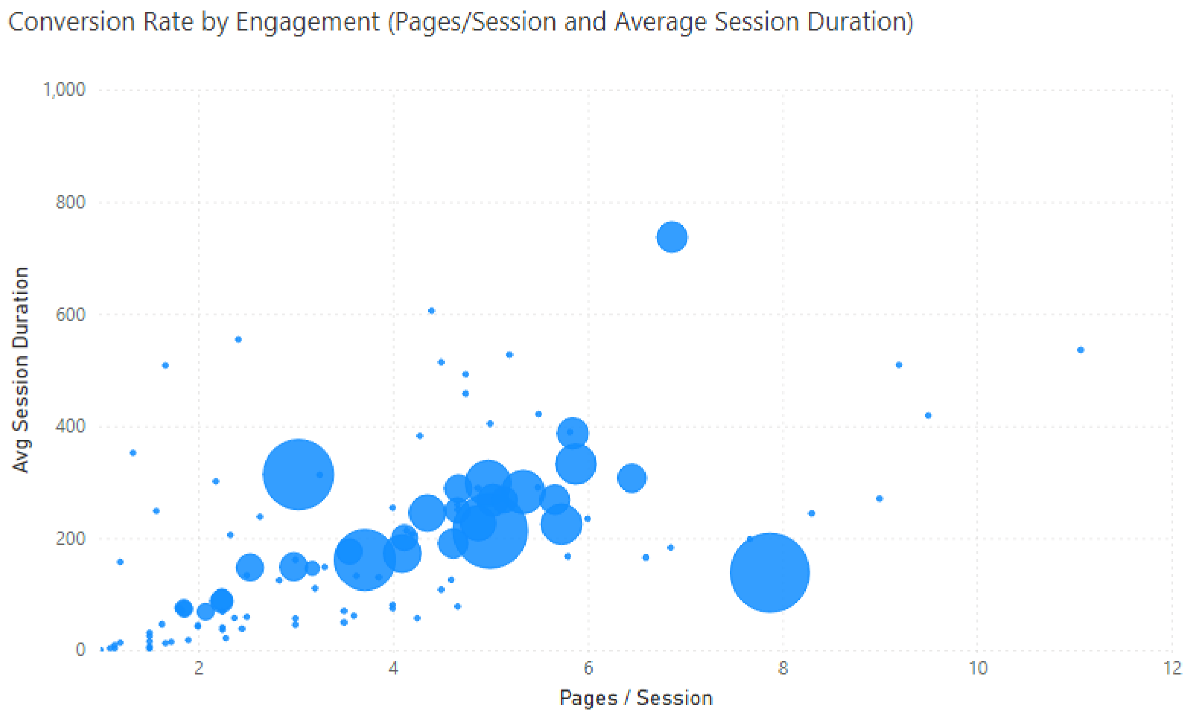

Figure 13: Conversion Rate by Pages/Session and Average Session Duration

From the chart, it can be observed that most sources generating high revenue are concentrated around 4–5 Pages/Session and 200 – 400 seconds for Avg Session Duration. This indicates that customers tend to interact deeply with the website before making a purchase decision, meaning the traffic quality is high and their interest in the products is significant. Additionally, there are quite a few sources that attract high-quality visitors who spend a long time on the website and view multiple pages but have not yet converted. This suggests there is potential to increase revenue by optimizing user experience or conversion strategies.

3.2.5 Value per Visit

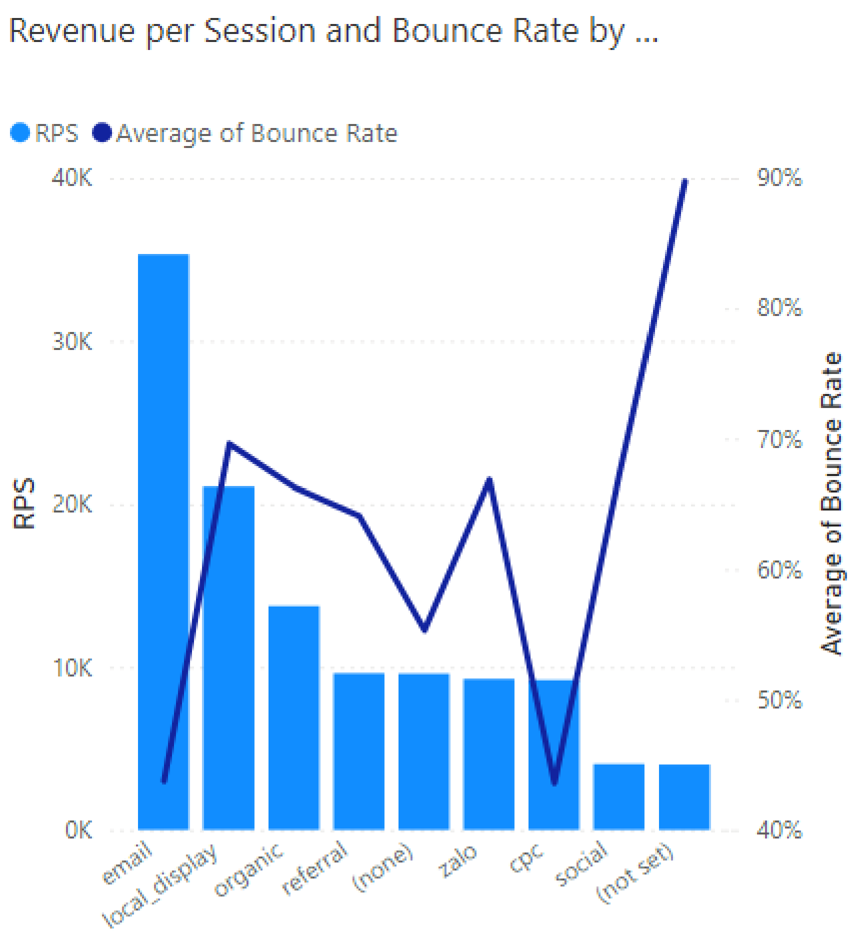

Figure 14: Revenue per Session (RPS) and Bounce Rate by Medium

The chart shows that when viewed by Medium, the groups differ quite clearly in terms of RPS and Bounce Rate. For email, RPS is high while Bounce Rate is low, indicating that traffic from this channel is of good quality and users tend to engage more deeply. Direct and CPC also show fairly good levels of RPS and Bounce Rate. In contrast, social has a high Bounce Rate and low RPS, suggesting that users from this channel often leave quickly and create little value.

4. Recommendation:

4.1 With Sales performance

Based on the previous analysis, in July the company reached about 11 billion VND in revenue with more than 37,000 orders. However, the cancellation rate was still high at 20%, and for the second purchase the cancellation rate even reached 73%. This shows problems in customer experience, promotions, or delivery operations. The recommendation is to review the fulfillment process to reduce cancellations, and at the same time create special offers or loyalty programs for returning customers. This will both lower cancellations and increase repeat purchases.

Second, the Average Order Value (AOV) in July changed a lot and was strongly influenced by promotions. On July 12th, the AOV dropped sharply while revenue jumped, showing that promotions had a strong effect but only for a short time. The recommendation is to redesign promotions to be tiered or spread across different periods instead of focusing on just a few days.

High-value orders often received bigger discounts and had a higher completion rate, while cancellations mostly came from orders with very deep discounts or low AOV. This means high-value orders are more stable and better quality, while orders pushed by heavy discounts are more risky. The company should focus on encouraging customers to increase their order value instead of offering very deep discounts. This can be done by upselling (e.g., buy 2 products and get 1 accessory, bundle offers), or giving benefits when spending above a certain amount. At the same time, limit very deep discounts for low-value segments to avoid cancellations, and focus resources on customers willing to spend more, who bring in more stable revenue.

However, it's also important to note that if the company increases discounts to push AOV but does not improve after-sales services (like delivery and support), it could backfire and cause more cancellations, as seen in the cancelled orders.

4.2 With Traffic performance

Based on the analysis, in July the company had 80% new users, showing strong ability to attract first-time visitors. However, the Bounce Rate was as high as 68% and the Conversion Rate only 0.45%, which indicates that although traffic was high, the conversion efficiency was still low. The recommendation here is to improve landing page experience, optimize page loading speed, and add clearer CTAs to reduce bounce and increase conversions.

By channel, CPC contributed the majority of revenue (66%), but the purchase rate from this channel was low, meaning the company is paying for large traffic volumes without proportional effectiveness. On the other hand, Email generated less revenue but had the highest conversion rate (over 2.5%), proving to be a high-quality channel since visitors from Email often already have clear purchase intent. Meanwhile, Social had a high Bounce Rate and low RPS, reflecting poor traffic quality. Therefore, the company should adjust budget allocation: for CPC, focus only on ads that deliver strong results and cut wasteful ones; for Email, expand the mailing list through sign-ups and promotions, while deploying automated campaigns such as cart abandonment reminders or cross-sell suggestions; and for Social, shift toward retargeting campaigns aimed at visitors who already engaged with the website or products, and prioritize best-selling products with special offers to boost conversions.

In addition, user behavior shows that customers spending 200–400 seconds on the website and viewing 4–5 pages are more likely to purchase, but there are also segments with high browsing activity but no conversion. This suggests potential for revenue growth by optimizing the checkout process, shortening steps, and adding policies such as free shipping or threshold-based promotions.

In summary, by reducing bounce and increasing conversion, optimizing CPC performance, expanding email marketing, leveraging untapped quality traffic, improving Social efficiency, and raising AOV, the company can significantly grow revenue and move closer to doubling its performance in the coming month.