**Data Cleaning Documentation**

**Steps (using Power Query):**

1. **Split Order Date into Year, Month, Quarter**
   * Use Power Query to extract the year, month, and quarter from the order date column.
2. **Filter Non-Gift Products**
   * Exclude products that are labeled as gifts from the dataset.
3. **Removed Duplicated data**

**DAX Formulas (Using Power BI to visualize):**

1. **Gross Merchandise Value (GMV): #** This formula calculates the total gross merchandise value by summing up the total order value in VND.

|  |
| --- |
| GMV =  CALCULATE(  SUM('Analyze data'[Tổng giá trị đơn hàng (VND)])  ) |

1. **Net Merchandise Value (NMV):** This formula calculates the net merchandise value by subtracting shop and Shopee discounts from the total order value, excluding canceled orders.

|  |
| --- |
| NMV =  SUMX(  'Analyze data',  IF(  'Analyze data'[Trạng Thái Đơn Hàng] <> "Đã hủy",  'Analyze data'[Tổng giá trị đơn hàng (VND)] - 'Analyze data'[Mã giảm giá của Shop] - 'Analyze data'[Mã giảm giá của Shopee],  0  )  ) |

1. **Cancellation Rate: #** This formula calculates the cancellation rate by dividing the number of canceled orders by the total number of orders.

|  |
| --- |
| Cancellation Rate =  DIVIDE(  CALCULATE(  COUNTROWS('Analyze data'),  'Analyze data'[Trạng Thái Đơn Hàng] = "Đã hủy"  ),  COUNTROWS('Analyze data'),  0  ) |

1. **Total Discounts:** This formula calculates the total discounts given by summing up both shop and Shopee discount values.

|  |
| --- |
| Total Discounts =  SUMX(  'Analyze data',  'Analyze data'[Mã giảm giá của Shop] + 'Analyze data'[Mã giảm giá của Shopee]  ) |

1. **Total Orders: #** This formula calculates the total number of distinct orders, excluding canceled orders.

|  |
| --- |
| TotalOrders =  CALCULATE(  DISTINCTCOUNT('Analyze data'[Mã đơn hàng]),  'Analyze data'[Trạng thái đơn hàng] <> "Đã hủy"  ) |

1. **Average Order Value (AOV) for All Orders: #** This formula calculates the average order value by dividing the total order value by the total number of orders, including canceled orders.

|  |
| --- |
| AOV All Orders =  DIVIDE(  SUM('Analyze data'[Tổng giá trị đơn hàng (VND)]),  COUNTROWS('Analyze data')  ) |

1. **Orders with Discounts:** # Calculate the number of orders with discounts applied (either shop or Shopee discount codes) Orders with Discounts

|  |
| --- |
| = CALCULATE( COUNTROWS('Analyze data'), // Count the rows in the 'Analyze data' table 'Analyze data'[Mã giảm giá của Shop] > 0 || 'Analyze data'[Mã giảm giá của Shopee] > 0 // Filter where either the shop discount code or the Shopee discount code is greater than 0 ) |

1. **Orders without Discounts:** # Calculate the number of orders without any discounts applied (both shop and Shopee discount codes are 0) Orders without Discounts

|  |
| --- |
| = CALCULATE( COUNTROWS('Analyze data'), // Count the rows in the 'Analyze data' table 'Analyze data'[Mã giảm giá của Shop] = 0 && 'Analyze data'[Mã giảm giá của Shopee] = 0 // Filter where both the shop discount code and the Shopee discount code are 0 ) |

1. **Total Cancelled Orders**: #Calculate the total number of cancelled orders by counting distinct order IDs with a status of "Đã hủy" (Cancelled) Total Cancelled Orders

|  |
| --- |
| = CALCULATE( DISTINCTCOUNT('Analyze data'[Mã đơn hàng]), // Count the distinct order IDs in the 'Analyze data' table 'Analyze data'[Trạng thái đơn hàng] == "Đã hủy" // Filter where the order status is "Đã hủy" (Cancelled) ) |

1. **Total Completed Orders**: #Calculate the total number of completed orders by counting distinct order IDs with a status other than "Đã hủy" (Cancelled) Total Completed Orders

|  |
| --- |
| = CALCULATE( DISTINCTCOUNT('Analyze data'[Mã đơn hàng]), // Count the distinct order IDs in the 'Analyze data' table 'Analyze data'[Trạng thái đơn hàng] <> "Đã hủy" // Filter where the order status is not "Đã hủy" (Cancelled) ) |