

Customer-Related Columns

1. Customer_First_Order_Date

- **Description:** Earliest Order Date for the customer.
- **Purpose:** Tracks when the customer first engaged, used for RFM analysis and tenure calculations.

2. Customer_Last_Order_Date

- **Description:** Most recent Order Date for the customer.
- **Purpose:** Supports recency calculations and churn analysis.

3. Customer_Tenure

- **Description:** Number of days between Customer_First_Order_Date and Customer_Last_Order_Date .
- **Purpose:** Measures customer loyalty and lifecycle duration.

4. Customer_No_of_Orders

- **Description:** Count of distinct Order ID values per Customer ID .
- **Purpose:** Quantifies purchase frequency for RFM and segmentation.

5. Customer_Total_Quantity

- **Description:** Total Quantity across all orders for the customer.
- **Purpose:** Assesses purchase volume for customer value analysis.

6. Customer_75th_Percentile_Order_Quantity

- **Description:** 75th percentile of Quantity per order for the customer.
- **Purpose:** Identifies typical high-end order sizes for inventory planning.

7. Customer_Average_Order_Value

- **Description:** Average Total Price (Discount Price * Quantity) per order for the customer.
- **Purpose:** Measures spending behavior for segmentation.

8. Customer_75th_Percentile_Order_Value

- **Description:** 75th percentile of Total Price per order for the customer.

- **Purpose:** Highlights high-value orders for targeting.

9. Customer_Historical_CLV

- **Description:** Sum of Total Price across all orders for the customer.
- **Purpose:** Estimates lifetime value for profitability analysis.

10. Customer_Diversity_Score

- **Description:** Herfindahl-Hirschman Index (HHI) based on proportion of Total Price across Product Category for the customer ($1 - \sum((\text{category_sales}/\text{total_sales})^2)$).
- **Purpose:** Measures diversity of purchases across product categories, lower HHI = more diverse.

11. Customer_Total_Revenue

- **Description:** Same as Customer_Historical_CLV (sum of Total Price), included for clarity as requested.
- **Purpose:** Quantifies total revenue from the customer.

12. Customer_Is_Repeat_Customer

- **Description:** Binary flag (1 if Customer_No_of_Orders > 1, else 0).
- **Purpose:** Segments one-time vs. repeat customers.

13. Customer_Churn_Probability

- **Description:** Simple heuristic probability based on recency and frequency (e.g., $1 - \exp(-\text{recency}/365) * (1/\text{No_of_Orders})$).
- **Purpose:** Estimates likelihood of customer churn for retention strategies.

14. Customer_Avg_Days_Between_Orders

- **Description:** Average days between consecutive Order Date values for the customer.
- **Purpose:** Measures purchase regularity for engagement analysis.

15. Customer_Return_Count

- **Description:** Count of orders with negative Quantity (assumed as returns; adjust if you have explicit return data).
- **Purpose:** Tracks return frequency for customer reliability.

16. Customer_Return_Value

- **Description:** Sum of absolute Total Price for orders with negative Quantity .
- **Purpose:** Quantifies financial impact of returns.

17. Customer_Avg_Days_Between_Order_and_Return

- **Description:** Average days between Order Date and assumed return date (using negative Quantity orders).
- **Purpose:** Evaluates return timeliness (requires return date data or proxy).

18. Customer_First_Order_Product_Count

- **Description:** Number of distinct Product Name values in the customer's first Order ID .
- **Purpose:** Assesses initial purchase complexity.

19. Customer_First_Order_Value

- **Description:** Sum of Total Price for the customer's first Order ID .
- **Purpose:** Measures initial purchase value for acquisition analysis.

20. Customer_Recency

- **Description:** Days between Customer_Last_Order_Date and current date (May 20, 2025).
- **Purpose:** Key metric for RFM analysis and churn prediction.

Logistics-Related Columns

21. Revenue_Per_Shipment_Type

- **Description:** Sum of Total Price per Ship Mode .
- **Purpose:** Evaluates revenue contribution by shipping method.

22. Days_Between_Shipment_and_Order

- **Description:** Days between Ship Date and Order Date for each order.
- **Purpose:** Measures fulfillment speed for logistics efficiency.

23. Customer_Count_Per_Shipment_Type

- **Description:** Number of distinct Customer ID values per Ship Mode .
- **Purpose:** Identifies customer preferences for shipping methods.

24. Avg_Quantity_Per_Shipment_Type

- **Description:** Average Quantity per order for each Ship Mode .
- **Purpose:** Assesses typical order sizes by shipping method.

25. Order_Count_Per_Shipment_Type

- **Description:** Count of distinct Order ID values per Ship Mode .
- **Purpose:** Measures shipping method popularity.

26. Shipping_Cost_Revenue_Per_Type

- **Description:** Sum of Shipping Cost per Ship Mode .
- **Purpose:** Quantifies shipping revenue for cost analysis.

Region-Related Columns

27. Region_Total_Revenue

- **Description:** Sum of Total Price per Region .
- **Purpose:** Identifies high-revenue regions for market focus.

28. Region_Order_Count

- **Description:** Count of distinct Order ID values per Region .
- **Purpose:** Measures order volume by region.

29. Region_Customer_Density

- **Description:** Number of distinct Customer ID values divided by distinct City values per Region .
- **Purpose:** Evaluates customer concentration for market penetration.

30. Region_Avg_Order_Value

- **Description:** Average Total Price per order in each Region .

- **Purpose:** Compares purchasing power across regions.

31. Region_Shipping_Efficiency

- **Description:** Average Days_Between_Shipment_and_Order per Region .
- **Purpose:** Assesses logistics performance by region.

32. Region_Return_Rate

- **Description:** Ratio of orders with negative Quantity to total orders per Region .
 - **Purpose:** Measures return frequency by region.
-

Product-Related Columns

33. Product_Total_Revenue

- **Description:** Sum of Total Price per Product Name .
- **Purpose:** Identifies top-performing products.

34. Product_Order_Count

- **Description:** Count of distinct Order ID values per Product Name .
- **Purpose:** Measures product popularity.

35. Product_Avg_Quantity

- **Description:** Average Quantity per order for each Product Name .
- **Purpose:** Assesses typical purchase volume per product.

36. Product_Discount_Frequency

- **Description:** Ratio of orders with Discount > 0 to total orders per Product Name .
- **Purpose:** Identifies frequently discounted products.

37. Product_Category_Revenue_Share

- **Description:** Percentage of total Total Price contributed by each Product Category .
- **Purpose:** Highlights dominant product categories.

38. Product_Return_Probability

- **Description:** Ratio of orders with negative Quantity to total orders per Product Name .
- **Purpose:** Flags products prone to returns.

Discount-Focused Columns

39. Customer_Discount_Junkie_Score

- **Description:** Ratio of orders where Discount > 0 to total orders for the customer.
- **Purpose:** Identifies customers who rely heavily on discounts (e.g., score > 0.8 = "Discount Junkie").

40. Customer_Avg_Discount_Rate

- **Description:** Average Discount value per order for the customer.
- **Purpose:** Measures discount dependency (high avg = price-sensitive).

41. Customer_Discount_Order_Value_Ratio

- **Description:** (Total Price with discounts) / (Total Price without discounts).
- **Purpose:** Quantifies revenue impact of discounts for the customer.

42. Customer_Non_Discounted_Order_Count

- **Description:** Count of orders with Discount = 0 .
- **Purpose:** Segments customers who buy at full price.

43. Customer_Discount_Threshold_Behavior

- **Description:** Binary flag (1 if customer orders only when Discount > X% , else 0).
- **Purpose:** Flags customers who wait for steep discounts.

Return-Focused Columns

44. Customer_Returnholic_Flag

- **Description:** Binary flag (1 if `Customer_Return_Count` > median return count across customers, else 0).
- **Purpose:** Identifies habitual returners ("Returnholics").

45. **Customer_Return_Rate**

- **Description:** `Customer_Return_Count / Customer_No_of_Orders`.
- **Purpose:** Measures return frequency (high rate = risky customer).

46. **Customer_Net_Spend_After_Returns**

- **Description:** `Customer_Historical_CLV - Customer_Return_Value`.
- **Purpose:** Adjusts CLV for return costs.

47. **Customer_Return_Product_Categories**

- **Description:** Top 3 `Product Category` values in returned orders.
- **Purpose:** Identifies categories prone to returns for the customer.

48. **Customer_Return_Time_StdDev**

- **Description:** Standard deviation of days between order and return.
- **Purpose:** Measures consistency in return timing (high = unpredictable).

Loyalty & Engagement Columns

49. **Customer_Loyalty_Tier**

- **Description:** Tier (Bronze/Silver/Gold) based on `Customer_Historical_CLV` quartiles.
- **Purpose:** Simplifies segmentation for marketing.

50. **Customer_Engagement_Score**

- **Description:** Weighted sum of (recency, frequency, diversity, returns).
- **Purpose:** Holistic measure of engagement (e.g., $(0.4 * \text{No_of_Orders}) - (0.3 * \text{Return_Rate})$).

51. **Customer_Purchase_Stickiness**

- **Description:** Ratio of orders in the customer's top 3 Product Categories to total orders.
- **Purpose:** Measures category loyalty (high = niche buyer).

52. Customer_Cross_Category_Ratio

- **Description:** Number of unique Product Categories / Customer_No_of_Orders .
 - **Purpose:** Low ratio = loyal to few categories; high = explorer.
-

Seasonal/Time-Based Columns

53. Customer_Seasonality_Index

- **Description:** StdDev of monthly order counts (high = seasonal buyer).
- **Purpose:** Identifies seasonal purchase patterns.

54. Customer_Days_Since_Last_Purchase

- **Description:** Days between Customer_Last_Order_Date and current date.
- **Purpose:** Simplifies recency for retention campaigns.

55. Customer_Black_Friday_Hunter

- **Description:** Binary flag (1 if $\geq 50\%$ of orders occur in November/December).
 - **Purpose:** Flags holiday-driven customers.
-

Experimental/Advanced Columns

56. Customer_Predictive_CLV

- **Description:** Machine learning-predicted future CLV based on RFM + returns.
- **Purpose:** Forward-looking value estimation.

57. Customer_Next_Purchase_Category

- **Description:** Predicted Product Category for next order (ML/NLP).
- **Purpose:** Personalizes recommendations.

58. Customer_Sentiment_Score

- **Description:** Aggregated sentiment from reviews/emails (if data exists).
- **Purpose:** Quantifies satisfaction beyond purchases.

59. Customer_Referral_Flag

- **Description:** Binary (1 if customer acquired via referral).
- **Purpose:** Tracks referral program effectiveness.

60. Customer_Price_Elasticity

- **Description:** Estimated sensitivity of order count to price changes.
- **Purpose:** Informs dynamic pricing strategies.