

Customer Sales Data Analysis

A. Revenue and Margin Analysis

1. Discount Impact on Margin: Calculate the percentage difference in Average Purchase Amount between transactions with a discount and those without.

```
SELECT
  -- Undiscounted AOV
  (SELECT AVG(CASE WHEN Discount_Applied = 'No' THEN `Purchase_Amount_(USD)` END) FROM
final_table) AS Undiscounted_AOV,

  -- Discounted AOV
  (SELECT AVG(CASE WHEN Discount_Applied = 'Yes' THEN `Purchase_Amount_(USD)` END) FROM
final_table) AS Discounted_AOV,

  -- Calculating Margin Erosion Percentage by nesting the averages
  ROUND(
    (
      (SELECT AVG(CASE WHEN Discount_Applied = 'No' THEN `Purchase_Amount_(USD)` END) FROM
final_table)
      - (SELECT AVG(CASE WHEN Discount_Applied = 'Yes' THEN `Purchase_Amount_(USD)` END)
FROM final_table)
    )
    / (SELECT AVG(CASE WHEN Discount_Applied = 'No' THEN `Purchase_Amount_(USD)` END) FROM
final_table)
    * 100,
  2) AS Margin_Erosion_Pct;
```

Undiscounted_AOV	Discounted_AOV	Margin_Erosion_Pct
60.1305	59.2791	1.42

2. Total Revenue Lost to Discounts: Estimate the total dollar amount lost to promotions (assuming items would have been bought at the average full price).

```
SELECT
  (COUNT(CASE WHEN Discount_Applied = 'Yes' THEN 1 END) * (SELECT
AVG(`Purchase_Amount_(USD)`)
  FROM final_table
  WHERE Discount_Applied = 'No')) - SUM(CASE WHEN Discount_Applied = 'Yes' THEN
`Purchase_Amount_(USD)` END)
AS Total_Revenue_Lost
FROM final_table;
```

Result Grid		Filter Row
	Total_Revenue_Lost	
	1427.7719	

3. Revenue per Payment Method: Identify which methods generate the most revenue (relevant for payment processing Opex).

```
SELECT
    Payment_Method,
    SUM(`Purchase_Amount_(USD)`)
AS total_revenue
FROM final_table
GROUP BY Payment_Method
ORDER BY total_revenue DESC;
```

	Payment_Method	total_revenue
▶	Credit Card	40310
	PayPal	40109
	Cash	40002
	Debit Card	38742
	Venmo	37374
	Bank Transfer	36544

4. Highest Discount Dependency: Find the top 5 product categories where discounts are most frequently used.

```
SELECT
    Category,
    COUNT(CASE WHEN Discount_Applied = 'Yes' THEN 1 END)
AS total_discounted_purchases
FROM final_table
GROUP BY Category
ORDER BY total_discounted_purchases DESC LIMIT 5;
```

	Category	total_discounted_purchases
▶	Clothing	731
	Accessories	543
	Footwear	259
	Outerwear	144

5. AOV by Shipping Type: Compare Average Order Value for different shipping types (informs shipping Opex decisions).

```
SELECT
    Shipping_Type,
    ROUND(AVG(`Purchase_Amount_(USD)`), 2)
AS average_order_value
FROM final_table
GROUP BY Shipping_Type;
```

	Shipping_Type	average_order_value
▶	Express	60.48
	Free Shipping	60.41
	Next Day Air	58.63
	Standard	58.46
	2-Day Shipping	60.73
	Store Pickup	59.89

- # B. Customer Segmentation & Value
- # 6. Average CLV Proxy by Age Group:
- # Find the average Previous_Purchases (a CLV proxy) across 10-year age bands.

```
SELECT
    FLOOR(Age / 10) * 10 AS age_group,
    AVG(Previous_Purchases) AS avg_clv_proxy
FROM final_table
GROUP BY age_group
ORDER BY age_group;
```

	age_group	avg_clv_proxy
▶	10	22.8933
	20	25.0659
	30	24.8771
	40	24.9432
	50	26.2088
	60	26.2080
	70	24.5075

- # 7. Customer Count by Frequency:
- # Count how many customers fall into each purchase frequency category.

```
SELECT
    Frequency_of_Purchases,
    COUNT(Customer_ID) AS customer_count
FROM final_table
GROUP BY Frequency_of_Purchases
ORDER BY customer_count DESC;
```

	Frequency_of_Purchases	customer_count
▶		2266
	1.0	553
	2.0	542
	4.0	539

- # 8. Retention Rate by Subscription:
- # Compare the average Previous_Purchases between subscribers and non-subscribers.

```
SELECT
    Subscription_Status,
    AVG(Previous_Purchases) AS avg_retention_proxy
FROM final_table
GROUP BY Subscription_Status;
```

	Subscription_Status	avg_retention_proxy
▶	Yes	26.0845
	No	25.0804

9. Top Spending Customers:

List the 10 customers with the highest total spend.

```
SELECT
    Customer_ID,
    SUM(`Purchase_Amount_(USD)`) AS total_spend
FROM final_table
GROUP BY Customer_ID
ORDER BY total_spend DESC LIMIT 10;
```

Customer_ID	total_spend
205	100
456	100
43	100
194	100
582	100
96	100
249	100
244	100
519	100
616	100

10. High-Value Item Sizing:

Find the most common size purchased for the highest revenue category (e.g., 'Clothing').

```
SELECT
    Size,
    COUNT(Customer_ID) AS
    purchase_count
FROM final_table
WHERE Category = 'Clothing'
GROUP BY Size
ORDER BY purchase_count DESC;
```

Size	purchase_count
M	778
L	481
S	284
XL	194

C. Product Performance and Inventory

11 Top 5 Revenue-Generating Items:

Identify the products driving the most sales.

```
SELECT
    Item_Purchased,
    SUM(`Purchase_Amount_(USD)`) AS total_revenue
FROM final_table
GROUP BY Item_Purchased
ORDER BY total_revenue DESC LIMIT 5;
```

Item_Purchased	total_revenue
Blouse	10410
Shirt	10332
Dress	10320
Pants	10090
Jewelry	10010

12. Lowest Rated Products:

Find the 5 products with the lowest average review rating.

```
SELECT
    Item_Purchased,
    AVG(Review_Rating) AS avg_rating
FROM final_table
GROUP BY Item_Purchased
ORDER BY avg_rating ASC LIMIT 5;
```

Item_Purchased	avg_rating
Shirt	3.6224852071005915
Jeans	3.6483870967741954
Blouse	3.680701754385968
Scarf	3.7070063694267517
Shorts	3.714012738853503

13. Seasonal Demand by Category:

Calculate the total revenue per season for a specific category (e.g., 'Footwear').

```
SELECT
    Season,
    SUM(`Purchase_Amount_(USD)`)
AS footwear_revenue
FROM final_table
WHERE Category = 'Footwear'
GROUP BY Season
ORDER BY footwear_revenue DESC;
```

Season	footwear_revenue
Spring	9555
Summer	9393
Fall	8665
Winter	8480

14. Rating Distribution:

Count the number of purchases for each discrete Review_Rating (1.0 to 5.0).

```
SELECT
    Review_Rating,
    COUNT(*) AS rating_count
FROM final_table
GROUP BY Review_Rating
ORDER BY Review_Rating DESC;
```

15. Most Popular Item Color:

Find the color purchased most often across all items.

```
SELECT
    Color,
    COUNT(*) AS purchase_count
FROM final_table
GROUP BY Color
ORDER BY purchase_count DESC LIMIT 1;
```

Color	purchase_count
Olive	177

D. Retention and Subscriptions

16. Subscriber vs. Non-Subscriber Demographics:

Find the average age and gender split for subscribers vs. non-subscribers.

```
SELECT
    Subscription_Status,
    ROUND(AVG(Age), 1) AS avg_age,
    COUNT(CASE WHEN Gender = 'Male' THEN 1 END) AS male_count,
    COUNT(CASE WHEN Gender = 'Female' THEN 1 END) AS female_count
FROM final_table
GROUP BY Subscription_Status;
```

Subscription_Status	avg_age	male_count	female_count
Yes	44.2	1053	0
No	44.0	1599	1248

17. Conversion by Location:

Find the top 5 locations with the highest percentage of subscribers.

```
SELECT
    Location,
    (COUNT(CASE WHEN Subscription_Status = 'Yes' THEN 1 END) * 100.0 / COUNT(Customer_ID))
    AS subscription_rate_pct
FROM final_table
GROUP BY Location
ORDER BY subscription_rate_pct DESC LIMIT 5;
```

Location	subscription_rate_pct
South Carolina	35.52632
West Virginia	34.56790
Nevada	34.48276
Oklahoma	33.33333
Missouri	33.33333

18. Average Rating for Subscribers:

Check if subscribers give better reviews (high satisfaction).

```
SELECT
Subscription_Status,
AVG(Review_Rating) AS avg_rating
FROM final_table
GROUP BY Subscription_Status;
```

Location	subscription_rate_pct
South Carolina	35.52632
West Virginia	34.56790
Nevada	34.48276
Oklahoma	33.33333
Missouri	33.33333

19. Loyalty Level vs. Subscription:

Check if high-loyalty customers (Previous_Purchases > 15) are more likely to subscribe.

```
SELECT
CASE
  WHEN Previous_Purchases >= 15 THEN 'High Loyalty'
  ELSE 'Low Loyalty'
END AS loyalty_level,
COUNT(CASE WHEN Subscription_Status = 'Yes' THEN 1 END) AS subscriber_count
FROM final_table
GROUP BY loyalty_level;
```

loyalty_level	subscriber_count
Low Loyalty	274
High Loyalty	779

20. Average Age of High-Frequency Buyers:

Target marketing Opex towards these specific, active cohorts.

```
SELECT
  Frequency_of_Purchases,
  ROUND(AVG(Age), 1) AS avg_age
FROM final_table
WHERE
  Frequency_of_Purchases IN ('Daily', 'Weekly')
GROUP BY Frequency_of_Purchases;
```

Frequency_of_Purchases	avg_age
------------------------	---------

E. Demographic/Geographic Analysis

21. Top 5 Revenue-Generating States:

Identify the core geographical areas driving total sales.

```
SELECT
    Location,
    SUM(`Purchase_Amount_(USD)`) AS total_revenue
FROM final_table
GROUP BY Location
ORDER BY total_revenue DESC LIMIT 5;
```

Location	total_revenue
Montana	5784
Illinois	5617
California	5605
Idaho	5587
Nevada	5514

22. Average Spend by Age Group:

(Simple breakdown of total spend by decade).

```
SELECT
    FLOOR(Age / 10) * 10 AS age_group,
    ROUND(AVG(`Purchase_Amount_(USD)`), 2) AS avg_spend
FROM your_data_table
GROUP BY FLOOR(Age / 10) * 10
ORDER BY age_group;
```

Location	total_revenue
Montana	5784
Illinois	5617
California	5605
Idaho	5587
Nevada	5514

23. Gender Split of Discount Use:

Compare how often each gender uses a discount.

```
SELECT
    Gender,
    COUNT(CASE WHEN Discount_Applied = 'Yes' THEN 1 END) AS discount_user_count
FROM your_data_table
GROUP BY Gender;
```

Location	total_revenue
Montana	5784
Illinois	5617
California	5605
Idaho	5587
Nevada	5514