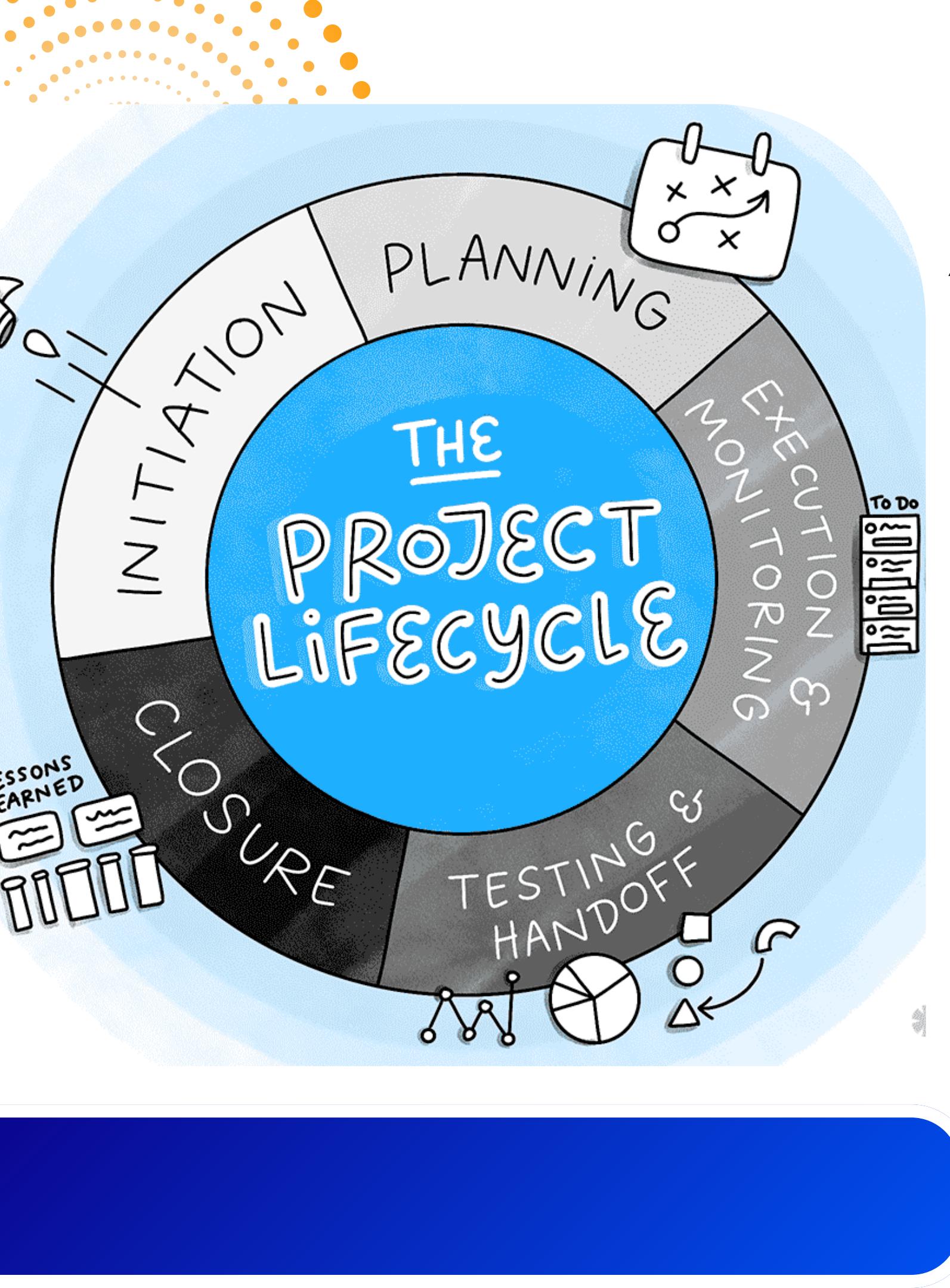




PRIMECART E- COMMERCE DATA ANALYSIS

Capstone Project

Presented by: Ajay Malik



About Project

As a Data Analyst working at PrimeCart E-Commerce, a global e-commerce platform that processes thousands of customer orders daily. The company wants to gain better insights into user behavior, order trends, and inventory management to optimize business performance. The objective is to analyze the given dataset using SQL queries and Power BI.



Phase 1:
SQL Analysis -
Advanced Queries

1

Retrieve the number of unique customers who placed an order in the last 6 months.

```
SELECT
    COUNT(DISTINCT o.user_id) AS unique_customers
FROM
    orders o
    JOIN
    users u ON o.user_id = u.id
WHERE
    o.created_at >= DATE_SUB(NOW(), INTERVAL 6 MONTH);
```

2 Calculate the total revenue generated per month for the last 12 months.

```
SELECT
    DATE_FORMAT(oi.created_at, '%Y-%m') AS month,
    ROUND(SUM(oi.sale_price), 2) AS total_revenue
FROM
    order_items oi
WHERE
    oi.created_at >= DATE_SUB(CURDATE(), INTERVAL 12 MONTH)
GROUP BY DATE_FORMAT(oi.created_at, '%Y-%m')
ORDER BY month;
```

3

Identify the top 5 most sold products based on quantity ordered.

```
SELECT
    p.name AS product_name, COUNT(oi.id) AS total_quantity_sold
FROM
    order_items oi
    JOIN
    products p ON oi.product_id = p.id
GROUP BY p.name
ORDER BY total_quantity_sold DESC
LIMIT 5;
```

4

Find customers who have placed more than 3 orders in the last year.

```
SELECT
    u.id,
    CONCAT(u.first_name, u.last_name) AS customer_name,
    COUNT(o.order_id) AS total_orders
FROM
    orders o
        JOIN
    users u ON o.user_id = u.id
WHERE
    o.created_at >= DATE_SUB(CURDATE(), INTERVAL 1 YEAR)
GROUP BY u.id , u.first_name , u.last_name
HAVING COUNT(o.order_id) > 3
ORDER BY total_orders DESC;
```

5

Calculate the average delivery time (difference between order date and delivery date) for each distribution center.

```
SELECT
    dc.name AS distribution_center,
    ROUND(AVG(TIMESTAMPDIFF(DAY,
        o.created_at,
        oi.delivered_at)),
    2) AS avg_delivery_time_days
FROM
    orders o
    JOIN
    order_items oi ON o.order_id = oi.order_id
    JOIN
    products p ON oi.product_id = p.id
    JOIN
    distribution_centers dc ON p.distribution_center_id = dc.id
WHERE
    oi.delivered_at IS NOT NULL
GROUP BY dc.name
ORDER BY avg_delivery_time_days;
```

6

Determine the top 3 cities that generated the highest revenue.

```
SELECT
    u.city, ROUND(SUM(oi.sale_price), 2) AS total_revenue
FROM
    order_items oi
    JOIN
    orders o ON oi.order_id = o.order_id
    JOIN
    users u ON o.user_id = u.id
GROUP BY u.city
ORDER BY total_revenue DESC
LIMIT 3;
```

7

Identify products that have been out of stock for more than 15 days in the last 3 months.

```
SELECT
    p.name AS product_name,
    p.id AS product_id,
    MAX(i.created_at) AS last_stock_date,
    DATEDIFF(CURDATE(), MAX(i.created_at)) AS days_out_of_stock
FROM
    inventory i
        JOIN
    products p ON i.product_id = p.id
WHERE
    i.sold_at IS NULL
        AND i.created_at >= DATE_SUB(CURDATE(), INTERVAL 3 MONTH)
GROUP BY p.id , p.name
HAVING days_out_of_stock > 15
ORDER BY days_out_of_stock DESC;
```

8

Count the number of cancelled orders per month and find the month with the highest cancellation rate.

```
WITH MonthlyOrders AS (
  SELECT
    DATE_FORMAT(created_at, '%Y-%m') AS month,
    COUNT(order_id) AS total_orders,
    SUM(CASE WHEN status = 'cancelled' THEN 1 ELSE 0 END) AS cancelled_orders
  FROM
    orders
  WHERE
    created_at >= DATE_SUB(CURDATE(), INTERVAL 12 MONTH)
  GROUP BY
    DATE_FORMAT(created_at, '%Y-%m')
)
SELECT
  month,
  total_orders,
  cancelled_orders,
  round((cancelled_orders / total_orders) * 100, 2) AS cancellation_rate
FROM
  MonthlyOrders
ORDER BY
  cancellation_rate DESC
LIMIT 1;
```

9

Identify the most common traffic source (e.g., Adwords, Email, Organic) that leads to a successful purchase.

```
SELECT
    u.traffic_source, COUNT(*) AS purchase_count
FROM
    orders o
        JOIN
    users u ON o.user_id = u.id
WHERE
    o.status IN ('Shipped' , 'Processing', 'Complete')
GROUP BY u.traffic_source
ORDER BY purchase_count DESC
LIMIT 1;
```

10

Find the conversion rate of users.

```
SELECT
    ROUND((COUNT(DISTINCT o.user_id) * 100.0 / COUNT(DISTINCT u.id)),
          2) AS conversion_rate
FROM
    users u
    LEFT JOIN
    orders o ON u.id = o.user_id;
```



Power BI

Phase 2:
Data Visualization
– Power BI

Dashboard 1

Sales & Revenue Analysis

Total Revenue

\$10.82M

Total Orders

125K

AOV

86.31

Gross Profit Margin (%)

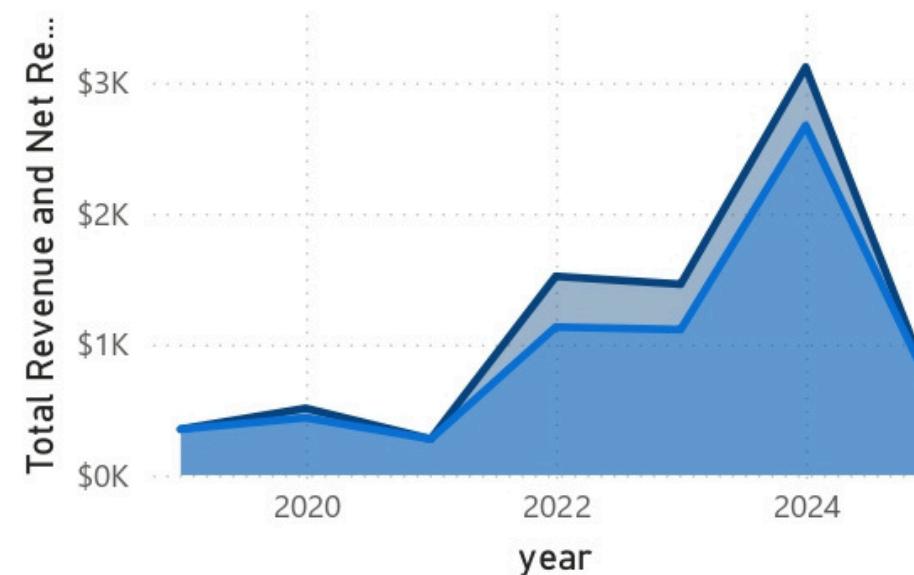
51.92

Net Revenue

\$9.73M

Total Revenue and Net Revenue by year

Total Revenue Net Revenue



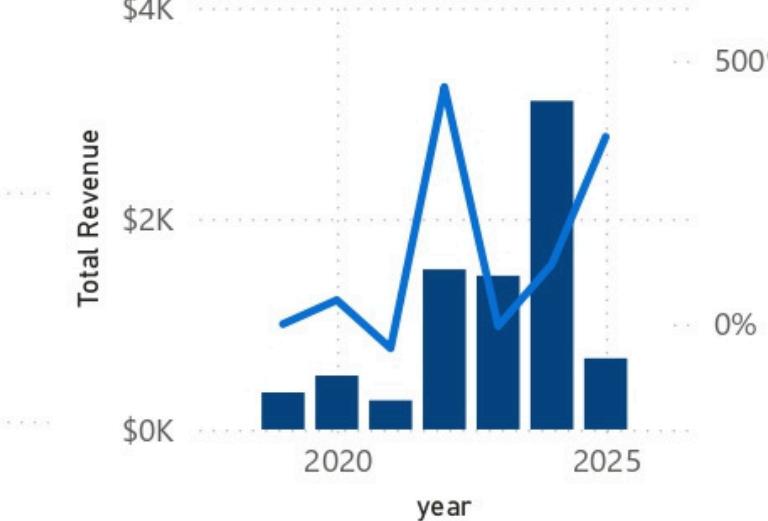
Total Revenue by month and year

year 2019 2020 2021 2022 2023 2024



Total Revenue and YoY Revenue Growth (%) by year

Total Revenue YoY Revenue Growth (%)



YoY Revenue Growth (%) by year

355.25%

MoM Revenue Growth (%) by year

-54.30%

category

All

brand

All

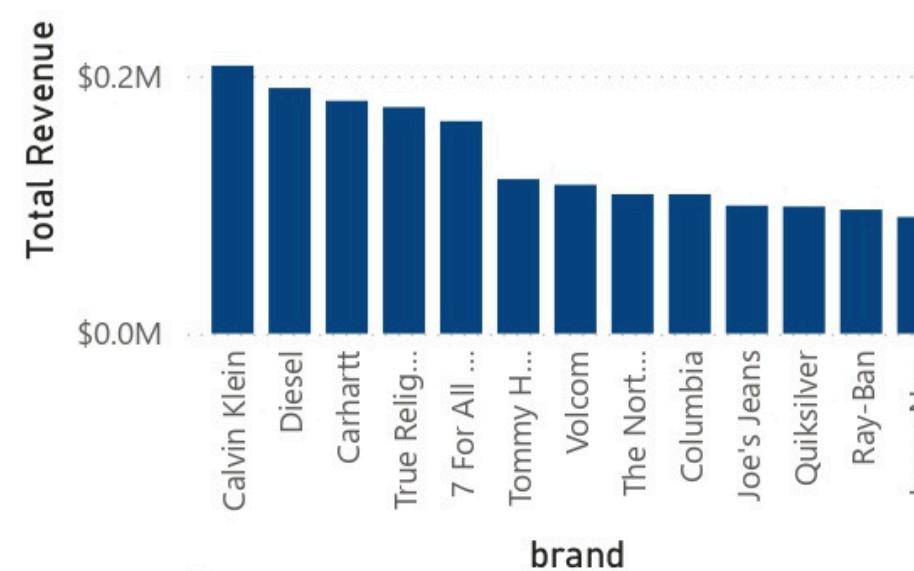
year

2025	2023	2021
2024	2022	2020

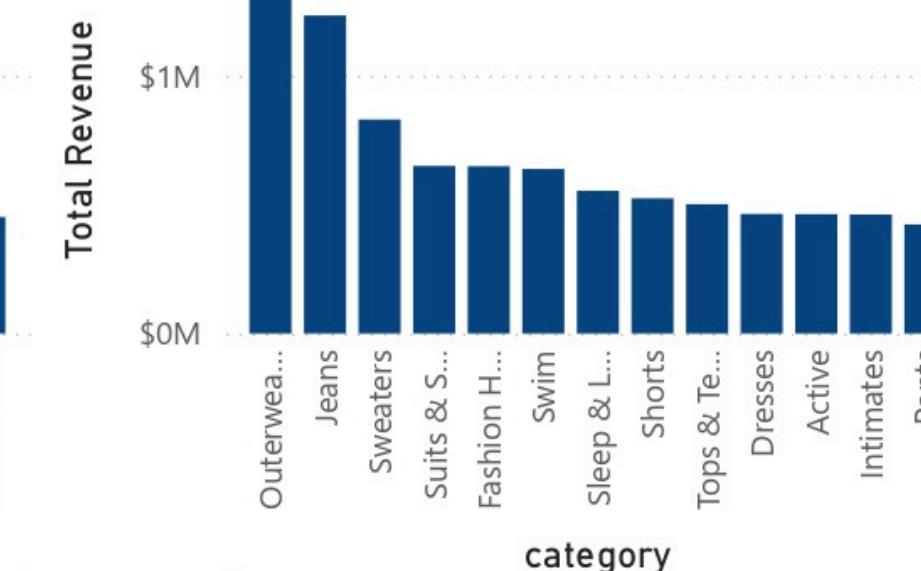
month

All

Total Revenue by brand



Total Revenue by category



Total Revenue by state and category

category Accessories Active Blazers ...



© 2025 TomTom, © 2025 Microsoft Corporation,
© OpenStreetMap

Dashboard 2

Customer Analytics

Total Customers

100K

New Customers

80K

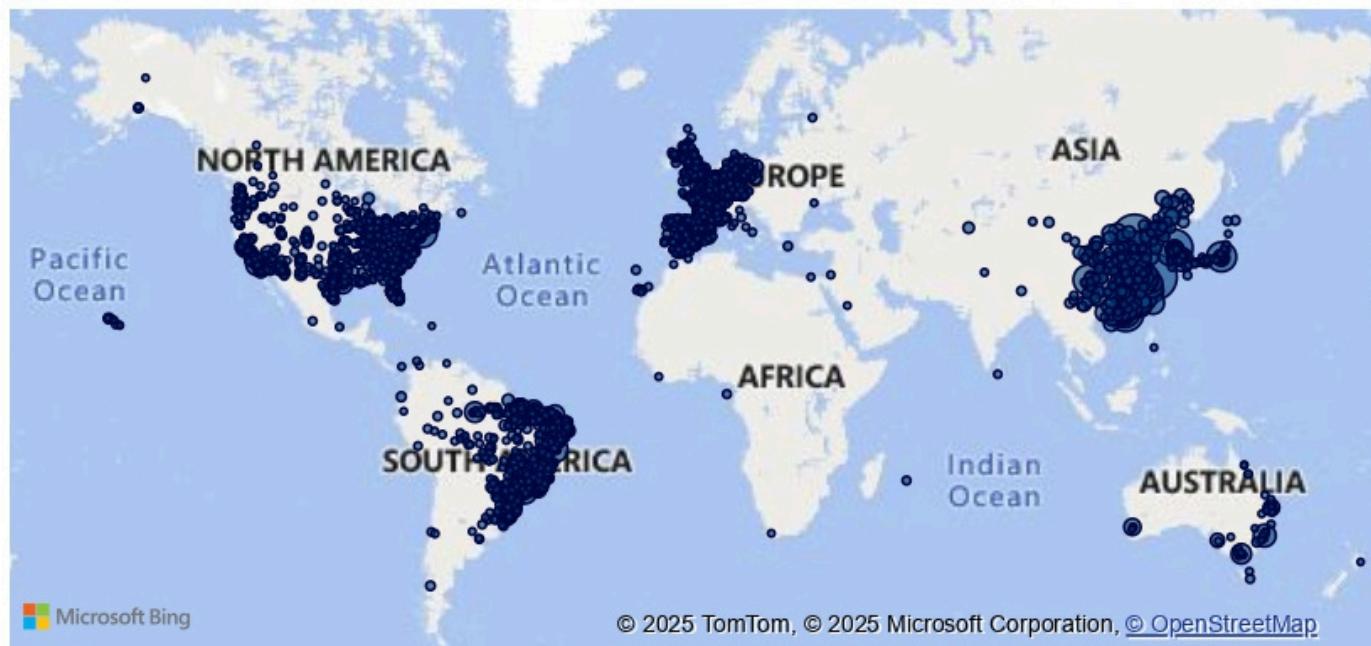
CLV Based on Revenue generated

135.41

Conversion Rate (Session based) (%)

15.09

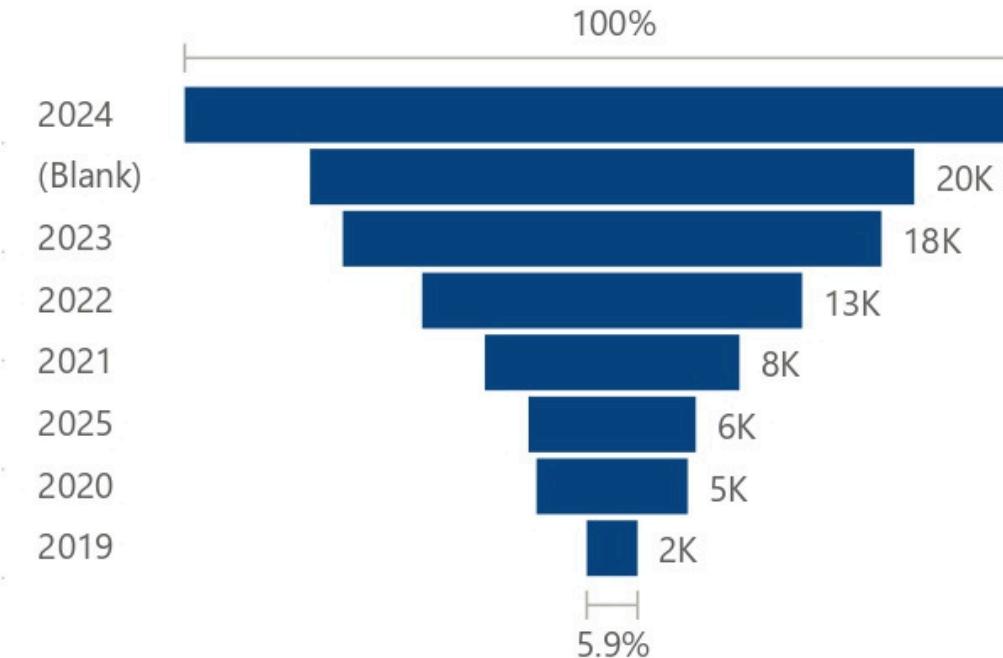
Total Customers by city



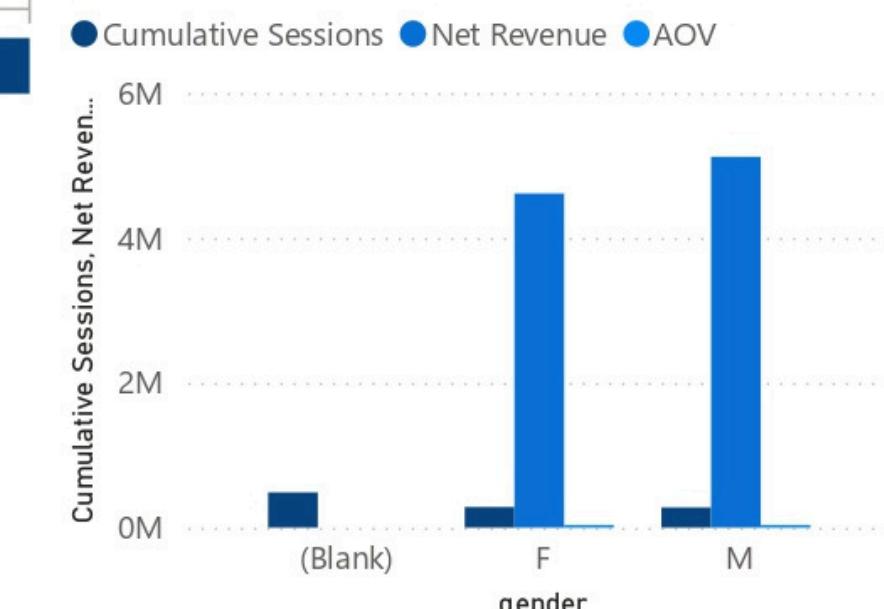
Returning Customers and New Customers by month



New Customers and Returning Customers by year



Cumulative Sessions, Net Revenue and AOV by gender



Year

(Blank)	2020	2022
2019	2021	2023

month

All

distribution_center_id

All

category

All

brand

(Blank)
lit Jeans
[BLANKNYC]
007Lingerie
10 Deep
106Shades

Dashboard 3

Product & Inventory Performance

Revenue Rank by Product Name

1

Sell-Through Rate (%)

37.08

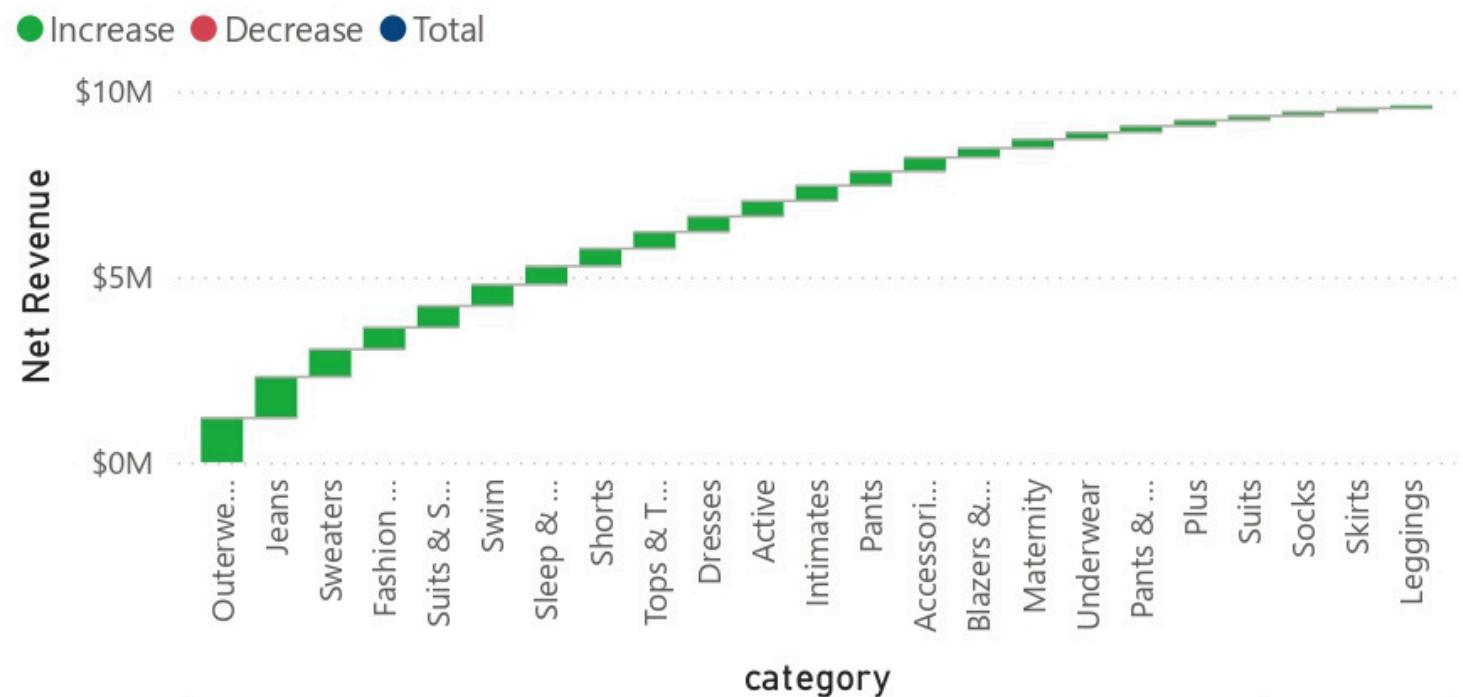
Inventory Turnover Ratio

1.06

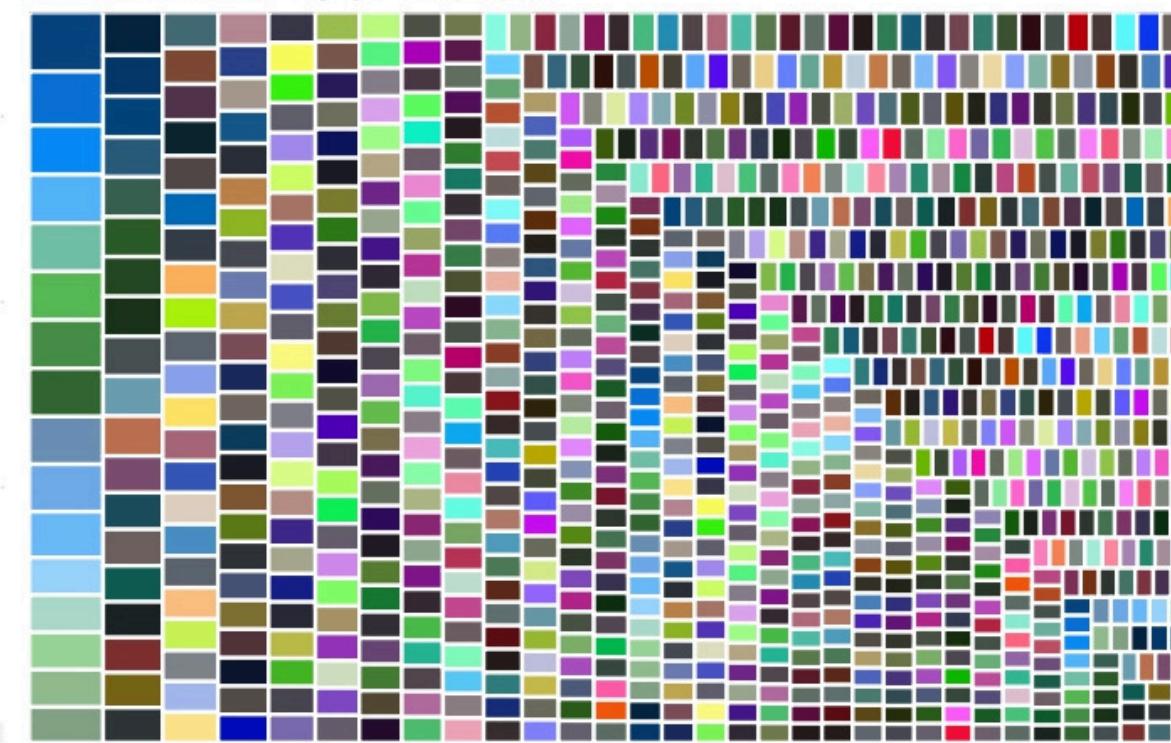
Stock Ratio (%)

62.92

Net Revenue by category



Net Revenue by product_id



Year

(Blank)	2020	2022	2024
2019	2021	2023	2025

category

All

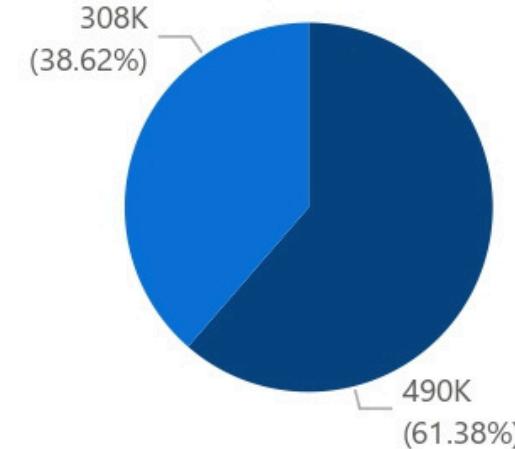
distribution_center_id

All

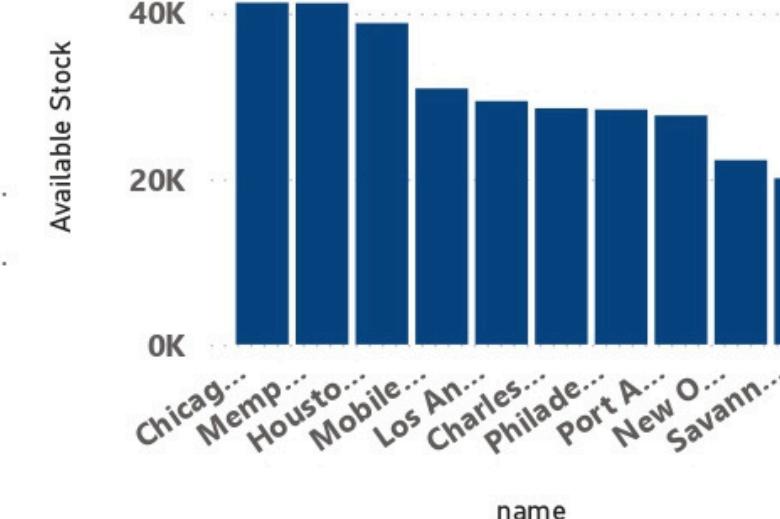
name

(Blank)
!iT Jeans Junior's Billy Shorts
!iT Jeans Women's Bonaroo Sl...
!iT Jeans Women's Bureau Tro...
!iT Jeans Women's Coachella ...
!iT Jeans Women's Coachella ...
!iT Jeans Women's Coachella ...

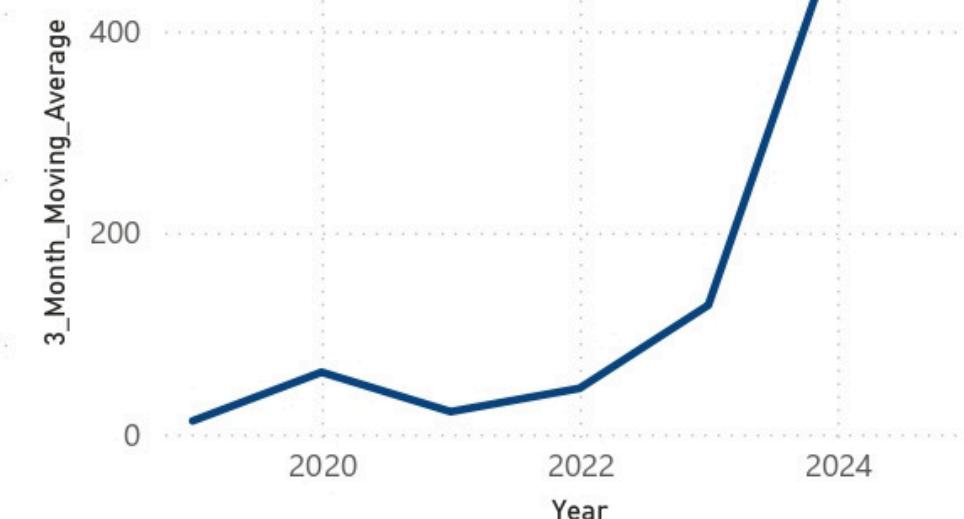
Total Stock Received and Available Stock



Available Stock by name



3_Month_Moving_Average by Year



Dashboard 4

Order Fulfillment & Logistics

On-Time Delivery Rate (%) assuming...

90.42

Return Rate (%)

10.10

Distribution Center Efficiency Score

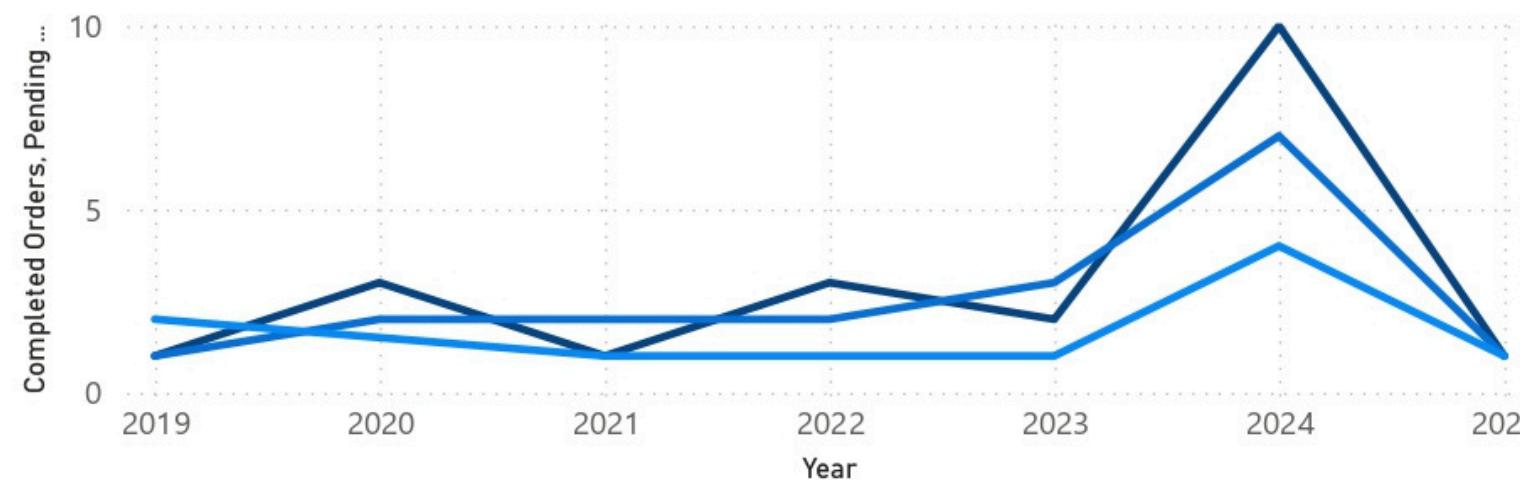
69.06

Average Order Processing Time (Days)

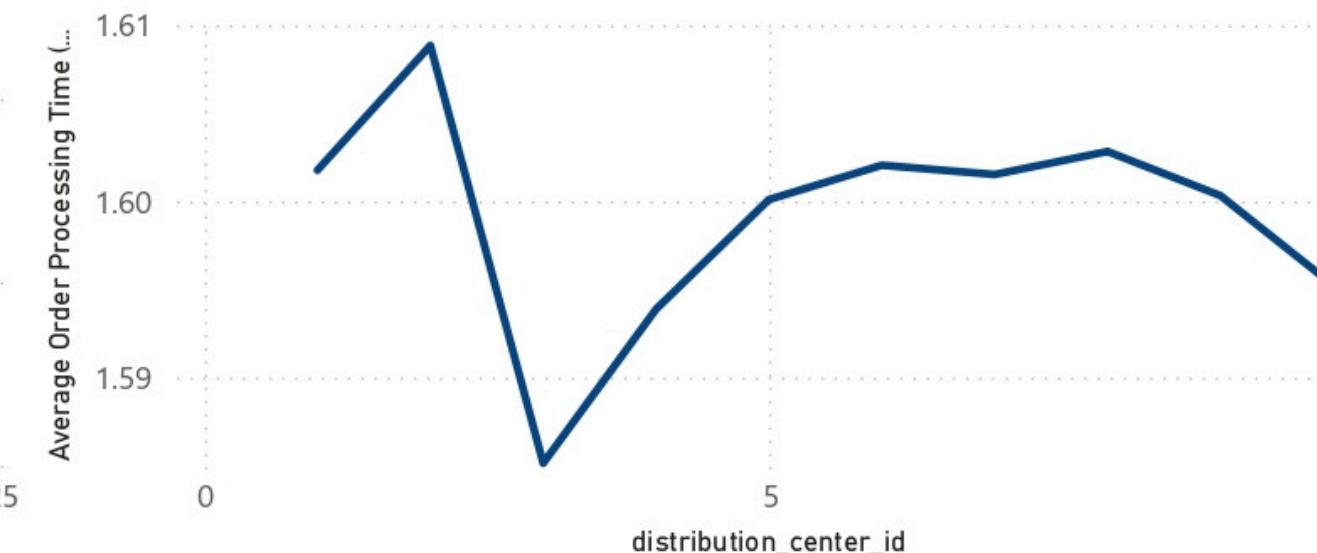
1.60

Completed Orders, Pending Orders and Cancelled Orders by Year

● Completed Orders ● Pending Orders ● Cancelled Orders



Average Order Processing Time (Days) by distribution_center_id



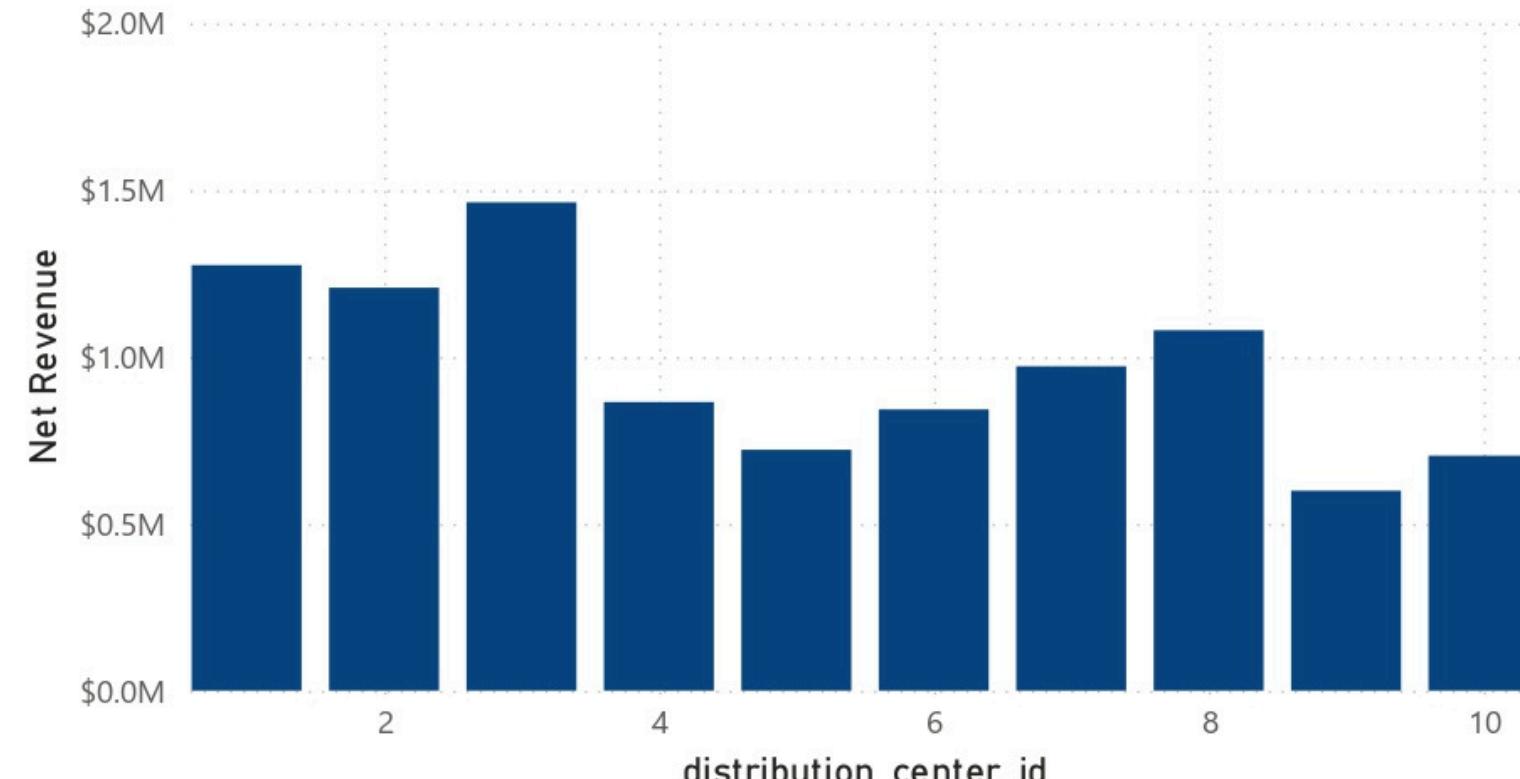
distribution_center_id

All

category

- (Blank)
- Accessories
- Active
- Blazers & Jackets

Net Revenue by distribution_center_id



Distribution Center Efficiency Score by latitude and longitude



Completed Orders

31K

Cancelled Orders

19K

Pending Orders

25K

Dashboard 5

Advanced Financial Reporting

Net Profit Margin (%)

51.92

Total Revenue per Customer

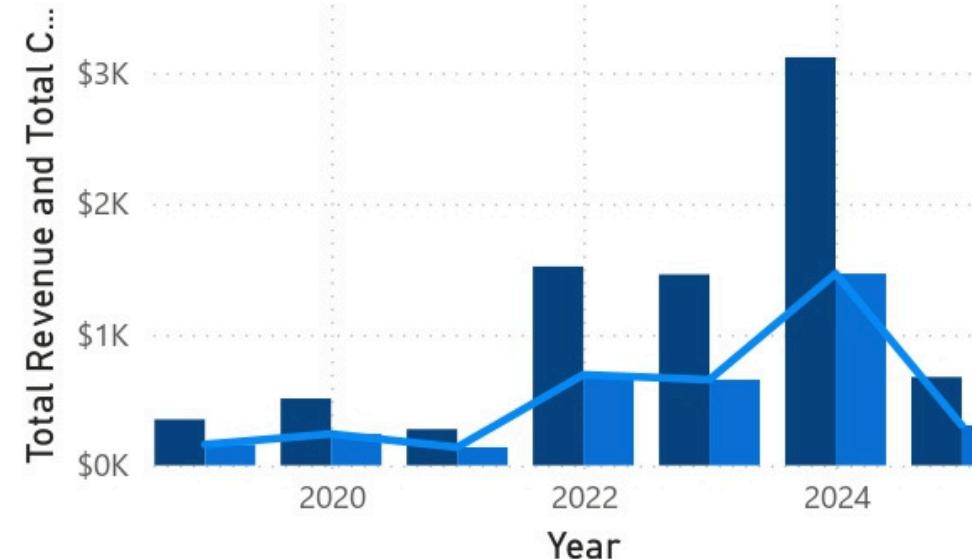
9.73M

On-Time Delivery Rate (%) assuming SLA(4)

90.42

Total Revenue, Total Costs and Total Costs by Year

Total Revenue Total Costs Total Costs



Net Profit Margin (%) by product_category

Total Revenue Total Costs Total Costs



Total Units Sold by country



year

(Blank)	2022
2019	2023
2020	2024
2021	2025

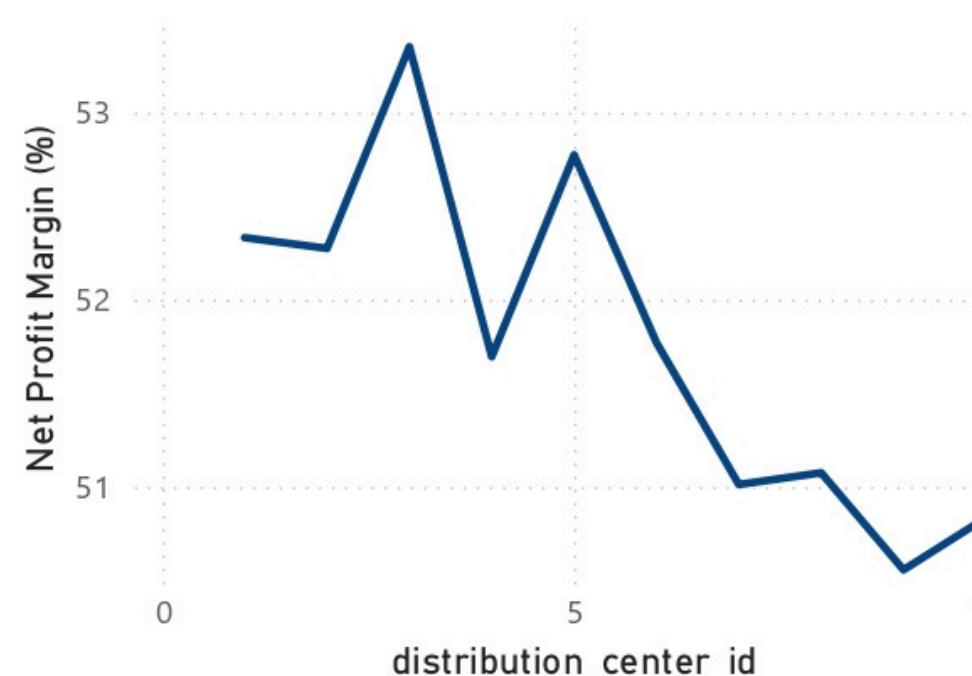
month

All

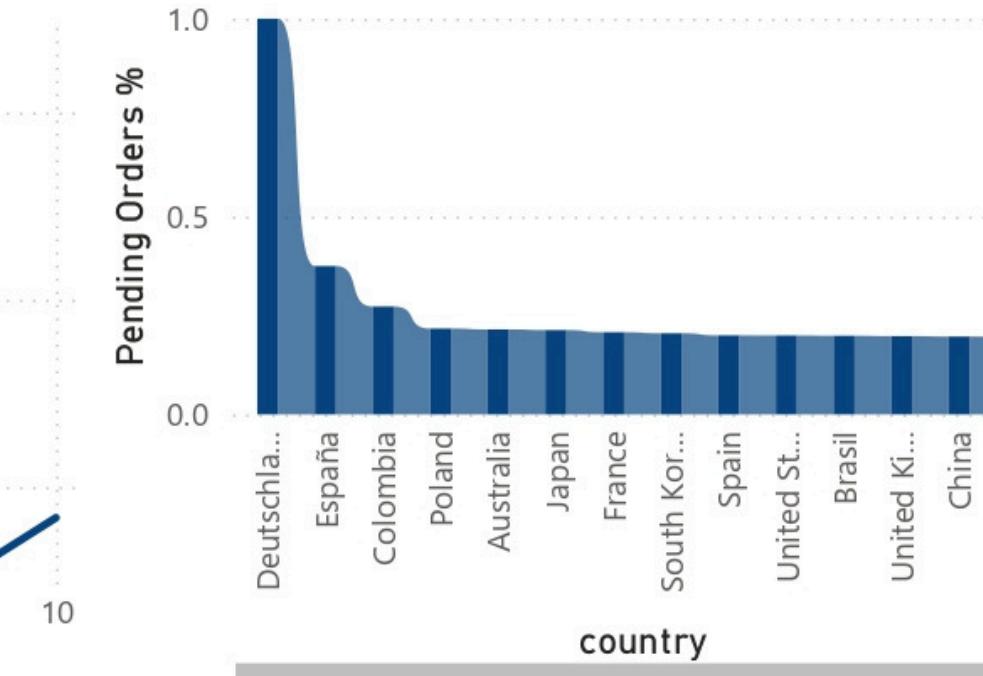
user_id

<input type="checkbox"/> (Blank)
<input type="checkbox"/> 1
<input type="checkbox"/> 2
<input type="checkbox"/> 3
<input type="checkbox"/> 4
<input type="checkbox"/> 5
<input type="checkbox"/> 6
<input type="checkbox"/> 7
<input type="checkbox"/> 8

Net Profit Margin (%) by distribution_center_id



Pending Orders % by country



Inventory Turnover Ratio by country





Phase 3: PrimeCart E- Commerce Insights Report

Insights from SQL Queries and Results:

1. Customer Base:

- 26,188 unique customers placed orders in the last 6 months, indicating a healthy customer base.
- Retention Opportunity: 14 customers placed 4+ orders in the last year. A loyalty program could further boost retention.

2. Revenue Growth:

- Total revenue grew steadily from 187k (Feb2024) to 931k (Jan 2025), peaking in December (\$592k) likely due to holiday shopping.
- January's spike suggests potential post-holiday promotions or New Year campaigns.

3. Top Products:

- Wrangler Men's jeans dominated sales (55 units), followed by True Religion jeans and Puma socks. Quantities sold are relatively low, possibly indicating premium pricing or niche markets.

4. Delivery Efficiency:

- All distribution centers averaged 3.46–3.52 days for delivery, with Philadelphia being the fastest. Minimal variation suggests a well-optimized supply chain.

5. Regional Sales:

- Shanghai, Beijing, and Seoul generated the highest revenue, highlighting strong market presence in Asia. Consider expanding inventory or marketing in these regions.

6. Inventory Issues:

- 9 products were out of stock for 92 days (e.g., Ed Hardy apparel). Proactive restocking or discontinuation is critical to avoid revenue loss.

7. Cancellation Trends:

- February 2024 had the highest cancellation rate (16.75%). Investigate causes (e.g., post-holiday returns, payment issues).

8. Traffic Source:

- Search traffic (66,030 purchases) drove the most conversions. Prioritize SEO and paid search ads to maintain this lead.

9. Conversion Rate for website in terms of new users:

- 79.93% conversion rate is exceptional, suggesting effective user experience or compelling product offerings. Evaluate if discounts are necessary or if organic demand drives this.

Insights from Sales & Revenue:

1. Revenue Performance Highlights

Total Revenue: 10.82M, Net Revenue: 9.73M (low deductions, strong cost control).

YoY Revenue Growth: +355.25% which is in year 2025, maxed out +449.79% at year 2022, lowest for year 2021 which is -45.92%.

Top brand in terms of revenue changes every year, but top category by revenue fluctuate between Outwears & coats and Blazers & Jackets.

2. Order Metrics

Total Orders: 125K | Implied AOV: \$86.56 (calculated as Net Revenue / Total Orders).

Maximum AOV is in year 2025 which is 135.01 followed by 113.16 in year 2022 and lowest in year 2021 which was 55.26.

Revenue Distribution: Graphs suggest revenue peak is in 2024.

3. Category & Geographic Focus

Top Categories(in terms of revenue): Outwear and Coats in year 2019, Jeans in year 2020, Maternity in year 2021, Blazers & Jackets in year 2022 and 2023 followed by Sweaters and Siuts in year 2024 and Dresses in year 2025.

4. Profit Margin

Maximum profit margins were in year 2023 which were 55.94%

Lowest profit margins were in year 2021 which were 49.92%.

This is considered a high-profit margin, often seen in luxury goods, software, or specialized services.

Insights from Customer Analytics:

1. Customer Overview

Total Customers: 100K

New Customers: 80K

2. Conversion Metrics

Conversion Rate (based on Sessions): 15.09%

A relatively strong conversion rate, suggesting effective marketing and website optimization.

Customer Lifetime Value (CLV) Based on Revenue: 135.41

Indicates the revenue expected from an average customer over their lifetime.

3. Order & Customer Distribution

Total Orders per Customer: 125K

Total Customers: 100K

This implies an average of 1.25 orders per customer, suggesting many customers are making only a single purchase.

4. Geographic Insights

The heatmap shows that the majority of customers are concentrated in North America, Europe, and parts of Asia & Australia.

5. Customer Trends Over Time

New Customers and Returning Customers by Year:

The number of new customers has been increasing, with 2024 having the highest number (~20K).

Returning customers are absent, reinforcing a retention issue.

6. Gender-Based Insights

Cumulative Sessions, Net Revenue, and AOV by Gender:

Both male (M) and female (F) customers contribute significantly to revenue, but male customers seem to have slightly higher engagement

7. Brand & Category Filter Options

Users can filter by year, month, distribution center, category, and brand, suggesting this dashboard is interactive for deeper analysis.

Key Takeaways & Recommendations

Strengths:

Strong new customer acquisition

Significant global reach

Opportunities for Improvement: Customer Retention

Increase Repeat Purchases: The average order per customer is low—consider implementing upselling, subscription models, or incentives for repeat purchases.

Insights from Product & Inventory Performance:

1. Sales & Revenue Performance

- **Top Revenue-Generating Product:** Product with product ID 24201 generates most revenue.
- **Net Revenue by Category:**
 - Outerwear, Jeans, and Sweaters seem to contribute the most to revenue.
 - Jumpsuits and clothing sets generate the least revenue.

2. Inventory Metrics

- **Sell-Through Rate: 37.08%**
 - Indicates that about 37% of stocked inventory has been sold.
 - A moderate rate—opportunity to optimize inventory turnover.
- **Inventory Turnover Ratio: 1.06**
 - Suggests that inventory is turning over slowly (i.e., stock is sold and replaced only slightly more than once per year).
 - Could indicate overstocking issues or slow-moving products.
- **Stock Ratio: 62.92%**
 - Indicates that about 63% of stock remains available.
 - Suggests potential overstocking or slower sales in some categories.

3. Stock & Distribution Analysis

- **Total Stock vs. Available Stock:**
 - Total Stock: 490K units
 - Available Stock: 308K units (~39%)
 - A significant portion of inventory remains unsold.
- **Available Stock by Location:**
 - Chicago, Houston, and Memphis have the highest available stock.
 - Some locations may be struggling with sales, requiring redistribution of inventory.

4. Trends & Forecasting

- **3-Month Moving Average by Year:**
 - Sales performance appears to be increasing over time, with a steep rise in 2024.
 - If the trend continues, demand might increase, requiring better inventory management.

Opportunities for Improvement:

- **Increase Inventory Turnover:** A turnover ratio of 1.06 is low—consider promotions or clearance sales for slow-moving items.
- **Optimize Stock Allocation:** Some locations (e.g., Chicago & Houston) have high available stock—redistribute excess inventory.
- **Boost Sell-Through Rate:** At 37.08%, there's room to improve by offering discounts, bundling products, or running targeted ads.
- **Analyze Underperforming Categories:** Leggings, Socks, and Skirts contribute the least revenue—consider discontinuation or strategy shifts.

Insights from Order Fulfillment & Logistics:

1. Order Processing & Delivery Efficiency

- **Average Order Processing Time: 1.6 days**
 - This is relatively fast, suggesting an efficient order handling system.
- **On-Time Delivery Rate: 90.42% (assuming SLA is)**
 - A strong performance, indicating that most deliveries are arriving as scheduled.
- **Distribution Center Efficiency Score: 69.06**
 - This suggests moderate efficiency, with room for improvement in fulfillment processes.

2. Order Completion & Cancellation Trends

- **Completed Orders: 31K**
- **Pending Orders: 25K (High)**
- **Cancelled Orders: 19K (Very High)**
 - A high cancellation rate (~38% of total orders) could indicate issues like stockouts, long processing times, or customer dissatisfaction.
 - Opportunity: Investigate causes of cancellations—whether due to payment failures, product availability, or order processing delays.

3. Performance by Year & Distribution Centers

- **Order Trends Over Time:**
 - Completed orders increased in 2023 and peaked in 2024, then slightly dropped in 2025.
 - Pending and cancelled orders have also fluctuated, increasing in 2024, which might indicate operational challenges.
- **Distribution Centers & Revenue:**
 - Some centers generate over \$1.5M in revenue, while others are below \$0.5M, showing performance inconsistencies.
 - Opportunity: Improve logistics at underperforming centers to balance order fulfillment speed and efficiency.

4. Areas for Improvement

Strengths:

- Fast processing time (1.6 days).
- High on-time delivery rate (90.42%).
- Certain distribution centers perform well in revenue and efficiency.

Challenges & Recommendations:

- **High Cancellation Rate (19K orders)** → Investigate reasons (e.g., inventory issues, delayed shipments).
- **High Pending Orders (25K orders)** → May indicate backlog issues—consider increasing processing capacity.
- **Distribution Center Efficiency Variability** → Optimize fulfillment operations for lower-performing locations.
- **Customer Experience Focus** → A return rate of 10.1% suggests potential product quality or expectation mismatch—address product descriptions and quality control.

Insights from Advanced Financial Reporting:

1. Profitability & Revenue Metrics

- **Net Profit Margin: 51.92%**
 - A strong profitability indicator, meaning the company is maintaining healthy margins after costs.
- **Revenue & Costs Over Time:**
 - There is a notable revenue increase in 2024, but costs have also risen, affecting profitability.
 - **Recommendation:** Analyze what drove the revenue surge in 2024 and whether it's sustainable.

2. Performance by Distribution Centers & Countries

- **Net Profit Margin by Distribution Centers:**
 - Some distribution centers are highly profitable, while others show a declining trend.
 - **Opportunity:** Investigate inefficiencies in low-performing centers and optimize operations.
- **Total Units Sold by Country:**
 - High sales activity in Europe, North America, and Asia, indicating strong market demand.
- **Inventory Turnover Ratio by Country:**
 - Higher turnover in Europe and South America, suggesting good inventory management.
 - Lower turnover in other regions may indicate overstocking issues or slow-moving products.

3. Pending Orders & Operational Challenges

- **Pending Orders % by Country:**
 - Germany, Spain, and Colombia have the highest percentage of pending orders, potentially affecting customer satisfaction.
 - **Recommendation:** Investigate logistics and fulfillment issues in these regions to reduce order backlogs.

4. Areas for Optimization

Strengths:

- Strong Net Profit Margin (51.92%)
- Efficient inventory turnover in key regions

Challenges & Recommendations:

- **Rising Costs in 2024** → Need to control operational expenses without affecting revenue growth.
- **Pending Orders in Key Markets** → Improve order fulfillment in Germany, Spain, and Colombia to reduce delays.
- **Distribution Center Inefficiencies** → Focus on optimizing low-performing centers to increase profitability.



Action Plan

Actionable Next Steps:

1. Reduce Inventory Overstocking → Improve stock planning based on demand trends.
2. Optimize Order Fulfillment → Address delays in Germany, Spain, and Colombia to improve delivery speed.
3. Control Rising Costs → Review operational expenses to sustain high-profit margins.
4. Enhance Distribution Center Performance → Focus on underperforming centers to improve efficiency.

By addressing these challenges, the business can achieve higher profitability, improved customer satisfaction, and optimized operations.

Thank you

Ajay Malik

Phone : 9888510256

Email : *ajaymalik021@gmail.com*

LinkedIn **GitHub**