blinkit

Blinkit Data Analytics

Problem Statement:

Blinkit, a leading quick-commerce grocery delivery platform, seeks to optimize its end-to-end business operations by analyzing customer purchasing behavior, delivery performance, inventory management, marketing effectiveness, and customer satisfaction metrics. The goal is to identify growth levers, improve operational efficiency, personalize customer engagement, and build predictive models for delivery time estimation, customer segmentation, and sales forecasting to drive profitable and scalable growth.

DATASET OVERVIEW

The Blinkit dataset captures detailed operational data of an online grocery delivery business. It includes customer demographics and order history, customer feedback with sentiment, delivery performance metrics, inventory stock records, marketing campaign data, and detailed product information. This dataset helps analyze customer behavior, delivery efficiency, product management, and marketing effectiveness to optimize business decisions.

ANALYSIS WITH SQL POWER BI & PYTHON

PART 0.1

MySQL Assignment — Data Analysis & Queries

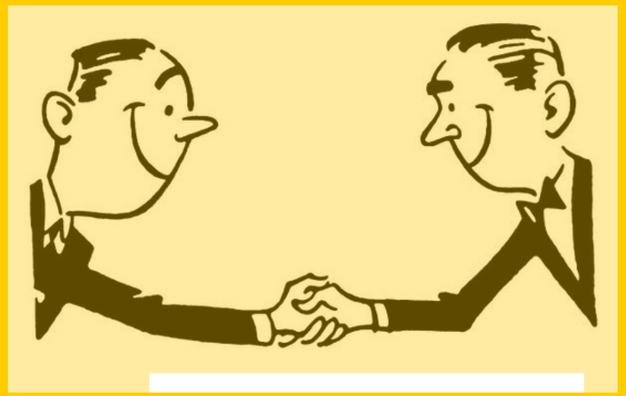
1. CUSTOMER INSIGHTS

- Identify the top 10 customers by total order value.

```sql
select
orders.customer_id,
round(sum(order_items.quantity*order_items.unit_price),2) As total_orde
from orders
join order_items on orders.order_id = order_items.order_id
group by orders.customer_id
order by total_order_value desc
limit 10;
```

- **COUNT THE NUMBER OF CUSTOMERS IN EACH SEGMENT (PREMIUM, REGULAR, INACTIVE, NEW).**

```sql
select customer_segment,count(*) as customer_count
from customers
group by customer_segment;
```



- **FIND CUSTOMERS WITH AN AVERAGE ORDER VALUE ABOVE 500 AND MORE THAN 10 ORDERS.**

```
select * from customers
where avg_order_value >500 and total_orders >10;
```

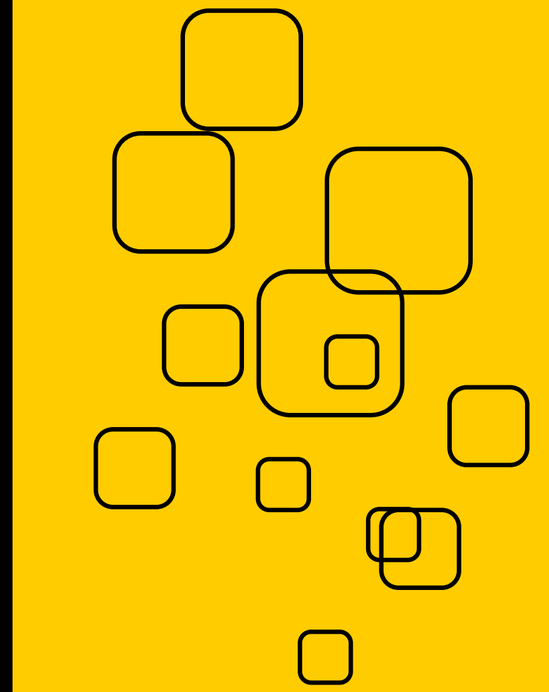
2 Order and delivery performance

- **List orders that were delivered late along with reasons for delay.**

```
select order_id,delivery_status,reasons_if_delayed
From blinkit.delivery_performance
where delivery_status like '%Delay%' and reasons_if_delayed is not null;
```

- **CALCULATE THE AVERAGE DELIVERY TIME (IN MINUTES) PER DELIVERY PARTNER.**

```sql
SELECT
    delivery_partner_id,
    ROUND(AVG(delivery_time_minutes), 2) AS avg_delivery_time_minutes
FROM
    delivery_performance
GROUP BY
    delivery_partner_id
ORDER BY
    avg_delivery_time_minutes DESC;
```



- **FIND THE TOP 5 STORES BY TOTAL ORDER REVENUE.**

```sql
Select* from orders;
SELECT
    store_id,round(SUM(order_total),2) AS total_order_revenue
FROM
    orders
GROUP BY store_id
ORDER BY total_order_revenue DESC
LIMIT 5;
```

3. PRODUCT AND INVENTORY ANALYSIS

- **IDENTIFY PRODUCTS THAT HAVE HAD DAMAGED STOCK MORE THAN 5 TIMES IN TOTAL.**

```
51 •    SELECT product_id, SUM(damaged_stock) AS total_damaged_stock
52      FROM (
53          SELECT product_id, damaged_stock FROM inventory
54          UNION ALL
55          SELECT product_id, damaged_stock FROM inventorynew
56      ) AS combined_inventory
57      GROUP BY product_id
58      HAVING SUM(damaged_stock) > 5;
```

- **CALCULATE THE TOTAL QUANTITY ORDERED FOR EACH PRODUCT.**

```
60    SELECT products.product_name,
61    SUM(order_items.quantity) AS total_quantity_ordered
62    FROM order_items
63    join products on
64    products.product_id = order_items.product_id
65    group by products.product_name
66    order by total_quantity_ordered;
```

- **FIND PRODUCTS THAT OFTEN FALL BELOW THE MINIMUM STOCK LEVEL (COMPARE CURRENT STOCK WITH MIN STOCK).**

```
69 •    select products.product_id,products.product_name,
70      sum(order_items.quantity) as current_quantity, sum(products.min_stock_level) as min_stock
71      from products left join order_items on
72      products.product_id = order_items.product_id
73      group by products.product_id, products.product_name
74      having current_quantity < min_stock ;
75
```

4. MARKETING CAMPAIGN EFFECTIVENESS

- CALCULATE TOTAL REVENUE GENERATED AND ROAS FOR EACH MARKETING CAMPAIGN.

```sql
SELECT
    campaign_name,
    campaign_id,
    SUM(revenue_generated) AS total_revenue,
    round(SUM(revenue_generated) / NULLIF(SUM(spend), 0),2) AS calculated_roas
FROM marketing_performance
GROUP BY campaign_name, campaign_id;
```

- **FIND THE CAMPAIGN WITH THE HIGHEST CONVERSION RATE (CONVERSIONS/IMPRESSIONS).**

```sql
#11 Find the campaign with the highest conversion rate (conversion /impression)
SELECT
campaign_id,
campaign_name,
  SUM(conversions) AS total_conversion,
  SUM(impressions) AS total_impression,
  (SUM(conversions) * 1.0 / NULLIF(SUM(impressions), 0)) AS conversion_rate
FROM marketing_performance
GROUP BY campaign_id,campaign_name
ORDER BY conversion_rate DESC
limit 1;
```

- **LIST ALL CAMPAIGNS TARGETED AT PREMIUM CUSTOMERS AND THEIR PERFORMANCE METRICS.**

```sql
SELECT
  campaign_id,
  campaign_name,
  SUM(impressions) AS total_impressions,
  SUM(clicks) AS total_clicks,
  SUM(conversions) AS total_conversions,
  ROUND(SUM(clicks) * 1.0 / NULLIF(SUM(impressions), 0), 4) AS ctr,
  ROUND(SUM(conversions) * 1.0 / NULLIF(SUM(impressions), 0), 4) AS conversion_rate
FROM marketing_performance
WHERE target_audience = 'premium'
GROUP BY campaign_id, campaign_name;
```

- **COUNT FEEDBACK ENTRIES BY SENTIMENT (POSITIVE, NEUTRAL, NEGATIVE).**

```
 9 ●    select count(customer_feedback.feedback_id) as feedback_count,
10     case
11         when blinkit_customer_feedback.sentiment = "Neutral" then "Neutral"
12         when blinkit_customer_feedback.sentiment = "Negative" then "Negative"
13         else "Positive"
14     end as sentiment_type
15     from  blinkit.customer_feedback
16     group by sentiment_type;
```

- **LIST CUSTOMERS WHO GAVE NEGATIVE FEEDBACK AND THEIR CORRESPONDING ORDERS.**

```
18 ●    select customer_feedback.customer_id, customer_feedback.order_id
19      from customer_feedback
20      where customer_feedback.sentiment = "Negative";
```

ANALYSIS WITH POWER BI

PART 0.2

POWER BI

Dashboard Assignment — Visualization & Insights

Dashboard 1: Customer Behavior & Segmentation

- Total customers by segment
- Orders and average order value by customer segment
- Registration trends over time
- Customer distribution by area and pincode
- High-value customers identification (filterable)

Dashboard 1: Customer Behavior & Segmentation

blinkit

sentiment
- Negative
- Neutral
- Positive

feedback_cat...
- App Experience
- Customer Servi...
- Delivery
- Product Quality

Year
- 2023
- 2024

Next Page

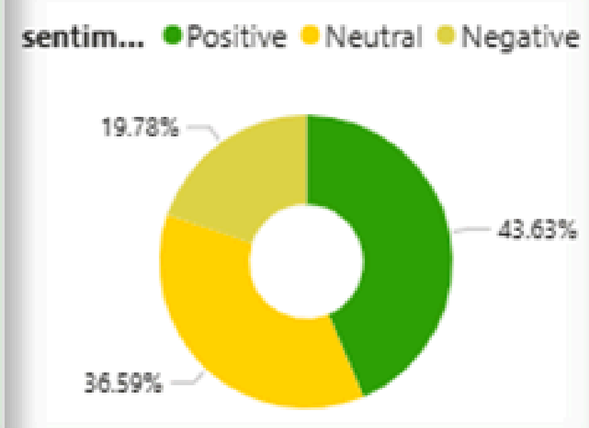
4.97M Total sales

3K Total Customers

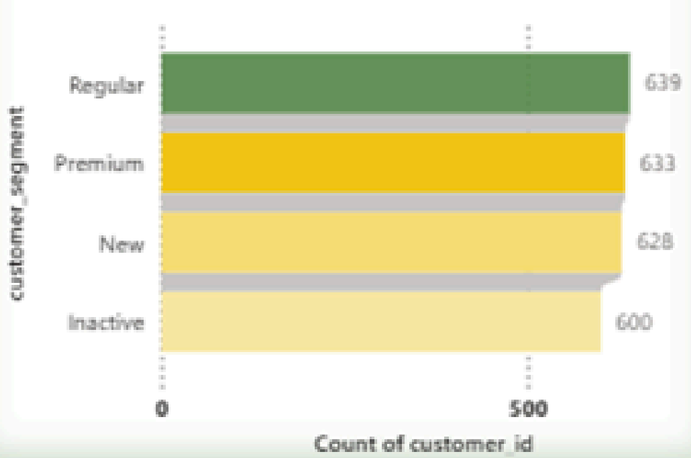
5K Total_Orders

Registrations trend

Rating by sentiment



Count of customer_id by customer_segment



customer_segment	Sum of avg_order_value	Sum of total_orders
Regular	696,169.44	6598
Premium	697,233.02	6452
New	701,247.13	6829
Inactive	661,290.00	6350
Total	2,755,939.59	26229

Count of customer_id by area and pincode

Total Orders and High-Value Customer

Dashboard 2: Order and Delivery Performance

- Total orders vs. late deliveries over time
- Average delivery time by delivery partner and area
- Delivery delay reasons breakdown
- On-time vs. delayed delivery rate KPIs

Dashboard 2: Order and Delivery Performance

blinkit

sentiment
- Negative
- Neutral
- Positive

feedback_cat...
- App Experience
- Customer Servi...
- Delivery
- Product Quality

Year
- 2023
- 2024

Previous Page

Next Page

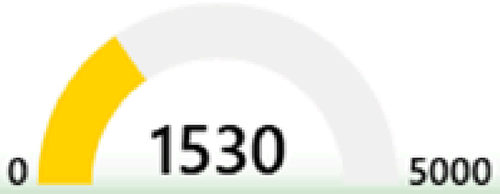
3.34
Average Rating

0.69
On-Time Delivery...

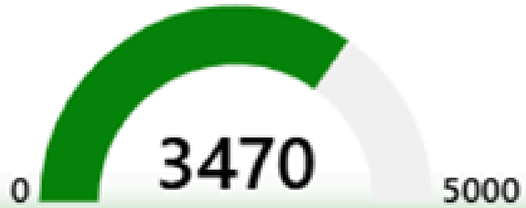
0.31
Delayed Delivery...

-2.12
Average Delivery Time

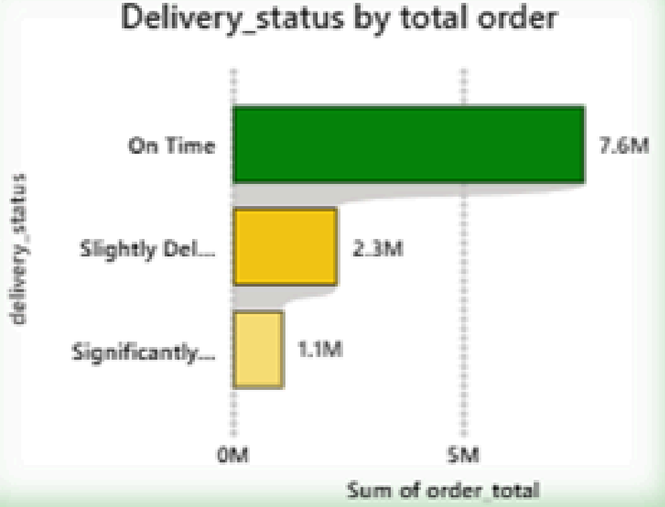
Late Deliveries Vs Total Deliveries



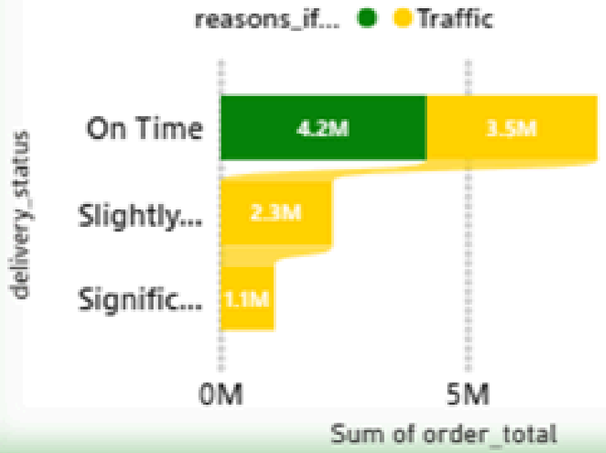
On-Time Deliveries Vs. Total Deliveries



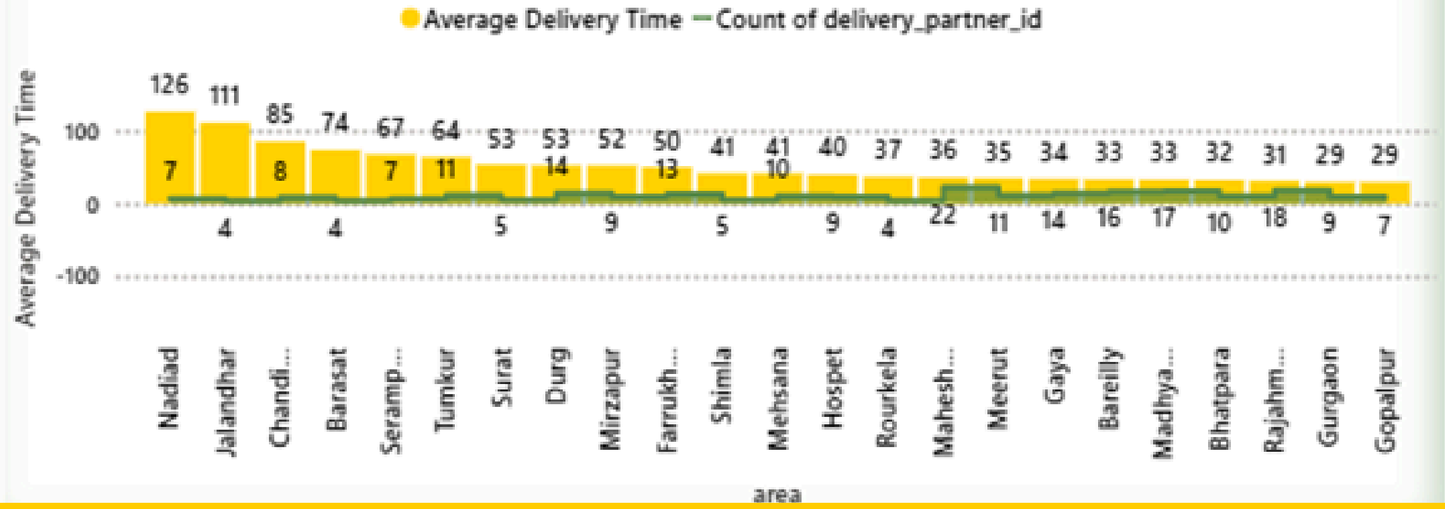
Delivery_status by total order



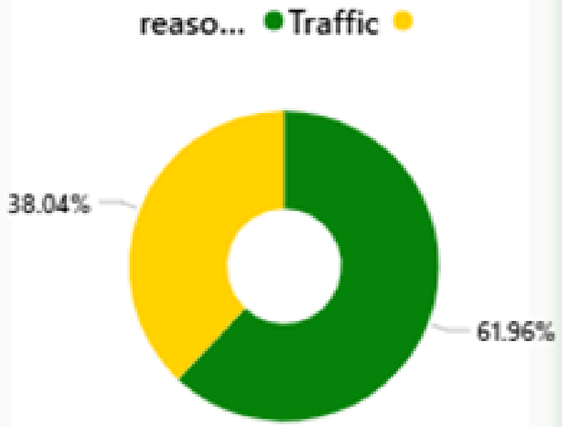
Delivery_status by reasons_of_delayed



Average Delivery Time Count of delivery_partner by area



Count of order_time by reasons_if_delayed



Dashboard 3: Product & Inventory Management

- Stock received vs. damaged stock trends by product and date
- Products below minimum stock level
- Top-selling products by quantity and revenue
- Margin percentage vs. price analysis
- Shelf life and stock rotation insights

Dashboard 3: Product & Inventory Management

blinkit

sentiment
- Negative
- Neutral
- Positive

feedback_cat...
- App Experience
- Customer Servi...
- Delivery
- Product Quality

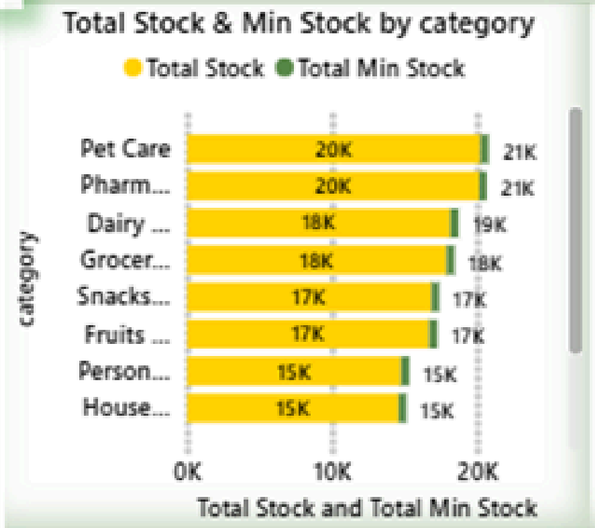
Year
- 2023
- 2024

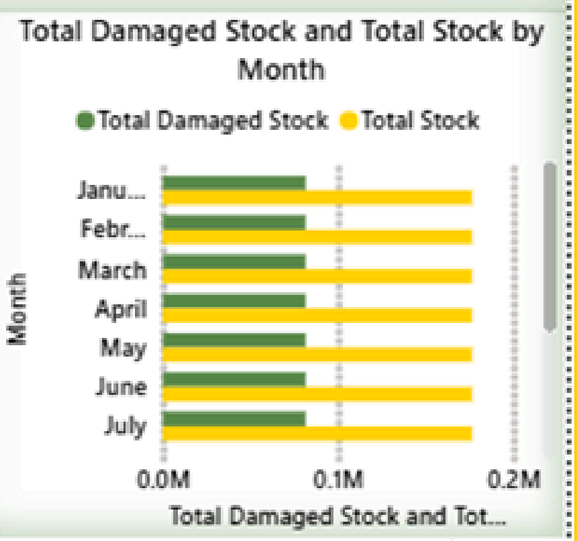
Previous Page

Next Page

268	176K	5K	231.76	1.08M
Total Product	Total Stock	Total Min Stock	Average Shelf life	StockRotatio...

product_name	Sum of shelf_life_days	Averageinventory Stock	Sum of quantity	StockRotationPerYear
Spinach	3	1.97	40	4,012.36
Lemonade	180	1.82	45	3,986.46
Milk	7	2.02	56	3,947.98
Instant Noodles	360	1.95	73	7,610.45
Bananas	9	1.85	85	11,875.03
Cereal	14	1.89	86	7,540.46
Diapers	730	2.00	91	8,078.22
Cookies	270	1.91	97	11,946.26
Tomatoes	9	1.92	97	11,189.28
Ice Cream	540	1.94	98	11,611.48
Total	62111	1.96	10034	1,078,036.37

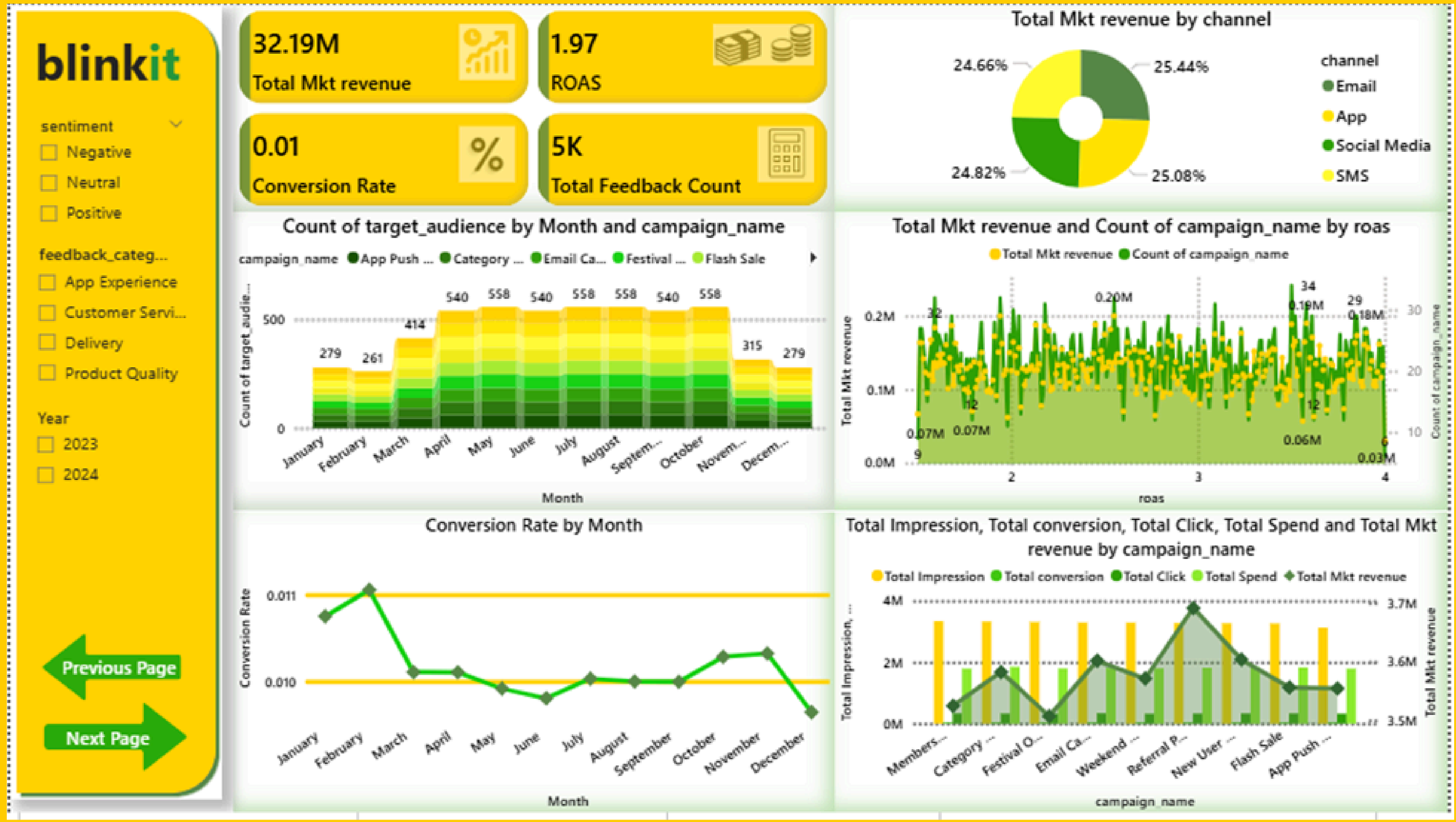




Dashboard 4: Marketing Campaign Performance

- Campaign-wise impressions, clicks, conversions, and spend
- Revenue generated and ROAS by campaign
- Conversion rate trends over time
- Performance by marketing channel (App, Email, SMS)
- Target audience segment-wise campaign effectiveness

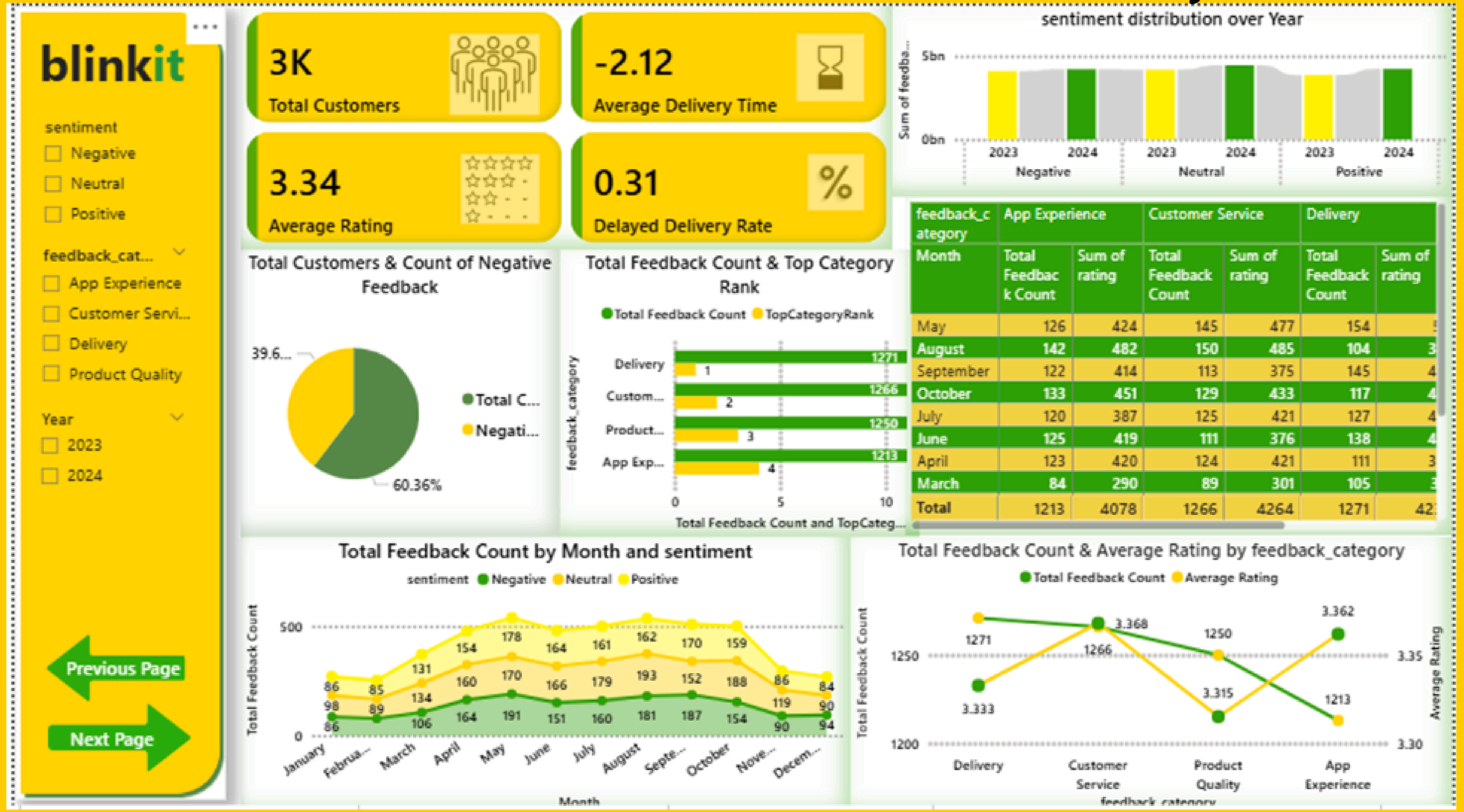
Dashboard 4: Marketing Campaign Performance



Dashboard 5: Customer Feedback & Sentiment Analysis

- Feedback count and average rating by category
- Sentiment distribution over time
- Text analysis word cloud or top feedback themes (optional)
- Link feedback sentiment to delivery performance or order value
- Negative feedback customers and resolution tracking

Dashboard 5: Customer Feedback & Sentiment Analysis

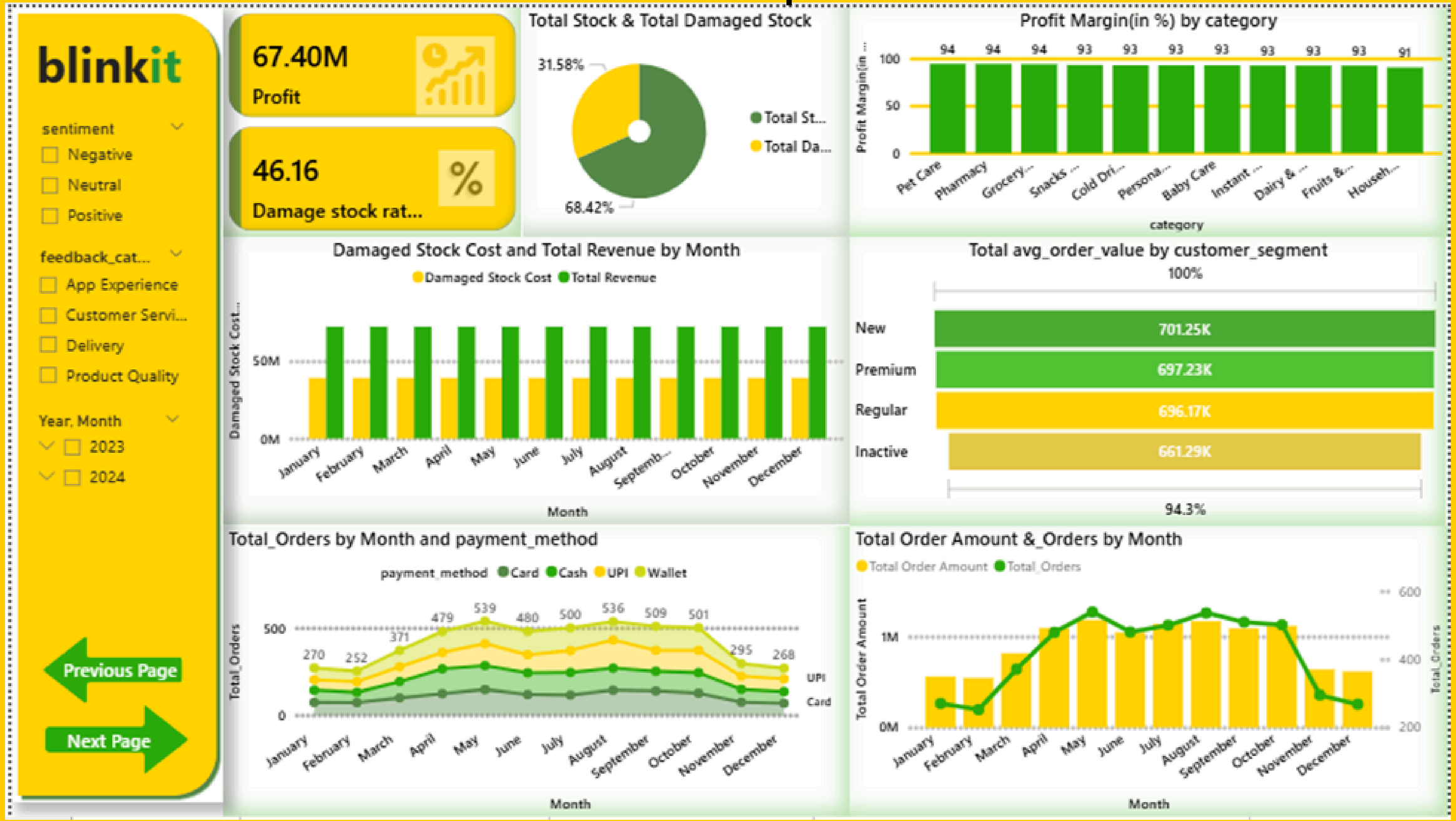


feedback_category	App Experience		Customer Service		Delivery	
Month	Total Feedback Count	Sum of rating	Total Feedback Count	Sum of rating	Total Feedback Count	Sum of rating
May	126	424	145	477	154	
August	142	482	150	485	104	3
September	122	414	113	375	145	4
October	133	451	129	433	117	4
July	120	387	125	421	127	4
June	125	419	111	376	138	4
April	123	420	124	421	111	3
March	84	290	89	301	105	3
Total	1213	4078	1266	4264	1271	42

Dashboard 6: Financial & Operational KPIs

- Total revenue and order totals over time
- Payment method distribution and trends
- Average order value trends by segment
- Cost analysis: damaged stock cost vs. revenue
- Profit margin overview by product category

Dashboard 6: Financial & Operational KPIs



Overall Summary

- **Overall Performance**
- Sales: ₹4.97M
- Customers: 3K
- Orders: 5K (avg ~1.7 per customer)
- Profit: ₹67.4M

🛍️ **Customer Segments & Behavior**

- Counts by segment: Regular (639), Premium (633), New (628), Inactive (600)
- Avg order value:
- New: ₹701K (highest)
- Premium: ₹697K
- Regular: ₹696K
- Inactive: ₹661K
- Orders Distribution: Most of the customers place 2–4 orders & high-value orders are well-performed.

📦 **Delivery Metrics**

- Avg rating: 3.34 / 5
- On-time delivery: 69%
- Delayed delivery: 31%
- Avg delivery time variance: –2.12 days (ahead of schedule on average)
- Delay causes: Traffic is responsible for 62%, other factors 38%.
- Regional performance: Some areas (e.g. Nadiad) have longer delivery times (around 126 days) than others.

🧾 Inventory & Stock

- SKU count: 268
- Total stock: 176K units
- Min stock threshold: 5K
- Avg shelf life: 232 days
- Stock rotation: 1.08M units/year
- Top product shelf-life ranges from 7 days (Milk) to 730 days (Diapers).
- Stock vs damaged stock: 31.6% of stock is damaged—significant loss.
- Damage ratio: 46.2% of stock value is damaged.

💰 Financials by Category

- Profit margins strong: 91–94% across all product categories.
- Revenue trends: Steady growth in the middle of the year & monthly revenue peaks in May–August.
- 📊 Payments & Monthly Orders
- Payment methods: Mostly UPI and Card Wallet usage.
- Monthly trends:
- Order volume peaks in the mid-year (536–540 orders/monthly), decrease in 2nd half of the year (268 orders in Dec).
- Order value tracks similar trend.

📢 Marketing Insights

- Marketing revenue: ₹32.2M
- ROAS: 1.97x
- Conversion rate: 0.01% (low)
- Total campaigns: 5K feedback entries gathered from customer via multiple sources
- Revenue by channel: Email, App, Social & SMS perform equally (25% each)
- Monthly audience size: Gradually increase from 279K in the Jan to 558K mid of the year
- Campaign peaks: Highest revenue months see in 30–34 campaigns & ROAS performed strongest in those periods.

🗣️ Sentiment & Feedback Trends

- Positive: 44%; Neutral: 36.6%; Negative: 19.8%
- Monthly feedback volume: Peaks in mid of the year (190–200 feedbacks), trailing later in 2nd half of the year (90 by Dec).
- Top issues: Delivery, Customer Service, Product Quality, and App Experience
- Average ratings per category: getting avg. rating for each category 3.33–3.36, with App Experience scoring highest.

🔑 Strategic Recommendations

- Focus on New Customers: Upsell and nurture loyalty with personalized engagement with customer.
- Activate Inactive: conduct reactivation campaigns or offers to recoup lost customer.
- Improve Delivery Logistics: Addressed high variance areas, reduced delayed deliveries, and proactively manage traffic related issue.
- Reduce Inventory Loss: Implement better handling & storage to cut damaged stock (currently 31.6%).
- Boost Conversion Rates: Optimize campaigns particularly around middle of the year peaks and improve landing experiences.
- Diversify Payment Options & Monitoring: Track underperforming payment methods (like Wallet or COD), and promote popular ones during slower months.
- Enhance Customer Support: Address the 20% negative sentiments by improving service channels. Work on root cause to reduced negative number.
- Continue Mid-Year Campaign Push: Arrange seasonal trends by launching & targeting campaigns during peak months.

PYTHON ASSIGNMENT - DATA PROCESSING & FORECASTING

PART 0.3

PYTHON

PYTHON

1. Exploratory Data Analysis (EDA)

- Load and clean the dataset. Handle missing or inconsistent data.
- Perform descriptive statistics for key numeric columns (order totals, delivery times, etc.).
- Visualize distributions of order values, delivery times, and customer segments.

2. Sentiment Analysis & Text Processing

- Analyze customer feedback text for common themes or frequent words (using basic text processing).
- Correlate sentiment scores with order values or delivery delays.
- 3. Delivery Time Prediction Model
- Build a regression model to predict delivery time in minutes based on factors like distance, promised time, and delivery partner.
- Evaluate the model’s accuracy and interpret important features.

3. Analysis details find on below GitHub link

https://github.com/Amruta8181/Capstone-project.git

THANK YOU