Blinkit Business Intelligence Report

Prepared by: Toviyah Williams
Date: June 3, 2025

1. Executive Summary

This report delivers a comprehensive performance review of Blinkit's marketing, sales, and operations using a combination of SQL-based analysis and Python-driven exploration. The data was structured into a relational PostgreSQL database and visualized in Tableau. The key findings reveal opportunities to optimize marketing spend, reduce delivery inefficiencies, better understand customer behavior, and enhance overall business performance. The analysis combines descriptive statistics, distribution visualizations, and performance evaluations. The recommendations provided are data-backed and immediately actionable.

2. Business Objectives

The goal of this analysis was to:

- Evaluate the effectiveness of marketing campaigns
- Identify top-performing and underperforming products
- Analyze revenue trends across timeframes
- Understand customer acquisition, retention, and behavior
- Explore delivery efficiency and inventory stock trends
- Assess sentiment and satisfaction from customer feedback

3. Data Sources and Methodology

Data Sources:

- 8 normalized tables built from Blinkit’s internal CSVs:
    - Orders, Customers, Products, Inventory, Order Items, Delivery, Marketing, Feedback
- Derived views in PostgreSQL:

    - kpi_summary, daily_sales_summary, campaign_performance_summary, etc.

Methodology:

- Python (Pandas, Seaborn, Matplotlib) used for data cleaning, aggregation, EDA, and sentiment analysis
- SQL used for data joining and extended insight generation
- Tableau used to design an interactive dashboard, including KPI banners, trends, and key campaign, customer, and sales insights
- All monetary values were converted from INR to USD using a fixed rate of ₹1 = $0.012

4. Key Findings and Recommendations

4.1 Marketing Performance

- Average weighted **ROAS is healthy (above 2.0)**. However, around **one-third of campaigns had ROAS at or below 1.5**. Several campaigns underperform significantly, especially when unweighted by spend.
- Friday- Sunday had slightly higher average conversion rates.

Recommendation: Pause or reevaluate campaigns with ROAS below 2.0. Consider reallocating spending to channels or audiences with higher weighted ROAS.

- CPC (Cost Per Conversion) across all campaigns averaged around **$0.66**, indicating cost-effective conversion.

Recommendation: Continue to leverage high-conversion, low-cost channels. However, monitor if lower CPC campaigns are also delivering quality customers or just volume.

- Audience type, channel, and campaign group breakdowns revealed **flat performance across most groups**, with minor but consistent variation.

Recommendation: Segment more granularly within audience types to uncover more in-depth behavioral trends and retarget based on engagement segments

- A heatmap of impressions, clicks, conversions, spend, and revenue shows a weak linear correlation. This suggests non-linear behavior such as diminishing returns and more complex audience responses.

Recommendation: Use non-linear modeling techniques to uncover more accurate patterns. Consider A/B testing budget increments to identify points where performance levels out.

- Average ROAS across Channels is uniform, but Engagement Varies.
- Engagement Rate** is highest for: Email (14.85%) and SMS(14.27%).

- Click-Through Rate (CTR) follows a similar trend.
- Audience Groups Perform Similarly.

Recommendation: Even though ROAS is flat, specific combinations of higher-performing segments may offer better engagement leverage. Test deeper segmentation or personalization for high-engagement channels.

4.2 Sales and Revenue Trends

- A significant margin gap exists between MRP (Maximum Retail Price) and actual selling prices, especially in lower-cost categories.

- Products that are significantly discounted (30-40%) do not necessarily sell more units.

- The most frequently ordered items and the items generating the most revenue do not fully overlap.

Recommendation: Conduct a pricing strategy review to ensure competitive pricing doesn't excessively cut into margins. Prioritize marketing for products that are both high-frequency and high-revenue. Review pricing or bundling strategies for high-volume but low-revenue items.

- Average order values by area vary greatly, ~ $9.60 at minimum, and $39.00 at maximum.
- Average order value is $11.73.
- Customers typically purchase 2 items per order, with low variance

Recommendation: Introduce offers like 'buy 2, get 1 free' or small discounts for orders containing 3+ items to increase average order size and value. Tailor discounts and promotions to high-spend and low-spend areas to increase promotion engagement levels.

4.3 Customer Behavior

- Overall retention rate is strong at ~94%.
- Monthly registrations are stable, peaking in March 2024, followed by a plateau.

Recommendation: Investigate what marketing activities were run during the March peak and replicate those tactics where possible.

- Negative feedback most frequently mentioned delivery delays, damaged products, and pricing dissatisfaction.

Recommendation: Use this feedback to inform improvements to packaging, pricing transparency, and delivery expectations. Run A/B tests for customer communications around delayed orders.

- Sentiment trends are steady, with no major fluctuations.

Recommendation: Maintain consistency in service and product quality, but explore targeted initiatives to turn neutral or mildly negative customers into brand advocates

- The majority (~35%) of feedback is neutral. Negative is at 32.84%, and Positive is at 32.40%.

Recommendation: Because negative sentiment is nearly equal to positive, focus on closing the gap by identifying common customer pain points and systematically addressing them through operations and messaging.

- Average order values by area vary greatly, ~ $9.60 at minimum, $39.00 at maximum.
- Average order value is $11.73.
- Customers typically purchase 2 items per order, with low variance

Recommendation: Introduce offers like 'buy 2, get 1 free' or small discounts for orders containing 3+ items to increase average order size and value. Tailor discounts and promotions to high-spend and low-spend areas to increase promotion engagement levels.

4.4 Delivery Efficiency

- No correlation found between order total and delivery time, suggesting consistent service and lack of priority or special handling for high-value orders.

Recommendation: Though all customers deserve fast delivery, consider premium delivery perks for loyal or high-spend customers to increase satisfaction and retention.

- **61.96%** of deliveries were delayed.
- Avg. delivery time for all orders: **4.44 min.**
- On-time deliveries averaged just **2.47 min.**

Recommendation (**Limited due to little variation in store and employee delivery data):** Conduct root-cause analysis on logistics delays. Consider:

- Revising delivery staff scheduling/coverage
- Adjusting ETA shown to customers
- Optimize routes for high-traffic areas Though all customers deserve fast delivery, consider premium delivery perks for loyal or high-spend customers to increase satisfaction and retention.

4.5 Stock and Inventory

- Certain product categories show **repeated instances of high damaged stock rates** (as revealed in the inventory analysis and damage spikes over time).
- Damage rate spiked sharply in **three distinct time windows**, suggesting possible vendor or logistics failures.
- August of 2023 saw the highest spike in intake of damaged products at nearly 25%.

Recommendation: Investigate supply chain issues for top-damaged product categories. Consider implementing tighter packaging standards or auditing suppliers for these high-risk SKUs.

- Box plots showed **non-uniform restocking patterns** by month and weekday, with heavier restocks on certain days.

Recommendation: Restructure replenishment scheduling to distribute stock intake more evenly. This can reduce operational bottlenecks and optimize labor allocation.

4.6 Customer Feedback and Sentiment

- Avg. customer rating: **3.34/5**

- Most common negative terms in text feedback revealed the most common reasons for low ratings. These terms suggest issues with delivery, product pricing, and product quality.

Recommendation: Monitor shifts in average rating over time. Target the lowest-rated categories (product quality and delivery) for improvement initiatives — such as packaging revisions, delivery partner audits, or item descriptions.

5. Key Recommendations Summary

- **Consider implementing automated performance monitoring**. Set benchmarks to pause or tweak underperforming campaigns early and focus spending on proven formats/audience mixes. **Monitor campaign effectiveness weekly** via views and dashboards.

- **Reallocate marketing budget** from broad, underperforming campaigns to niche, high-ROAS audiences determined through customer segment restructuring.

- Focus audience **optimization not on broad segments** but on **micro-behaviors** (past order timing, product category interest, churn risk, etc.).

- **Improve delivery partner monitoring** by tracking average delays by store/partner ID.

- **Set sentiment alerts** for flagged keywords in customer feedback to proactively manage customer experience.

- Even out restocking cadence to reduce operational strain and avoid overstocking or stockouts; base timing on historical demand patterns

6. Data Integrity & Methodological Notes

1. **Cost Per Conversion**:
   Due to the absence of direct attribution fields linking marketing campaigns to individual customers, Customer Acquisition Cost (CAC) was replaced with **Cost Per Conversion** (CPC). CPC provides a more reliable metric under these conditions and avoids inflated values caused by approximate joins.

2. **ROAS Recalculation:**
   The ROAS values contained in the original Blinkit Marketing Performance CSV file were **taken at face value**. Upon further investigation, it was discovered that these **values were incorrect**. To keep the data valid, ROAS values were recalculated manually and overwritten in place. This process is documented in the python_analysis.ipynb file. The file containing the corrected values was used in the final Python and SQL-based analysis. The **analysis and report were updated** where appropriate.

3. **Deliveries Data Limitations:**
   The data containing information on deliveries for stores and partners is severely limited. Each unique store has only completed one delivery. Each delivery partner has only completed one delivery. This makes **accurate delivery issues analysis and interpretation impossible, as each store and partner either had a delivery issue or did not**.

4. **SQL vs. Python Logic in Calculations:**
   - In comparing my SQL analysis with the work done using Python, I noticed the ROAS values across campaigns values varied greatly. **SQL-based analysis** naturally calculates **weighted ROAS**, where total revenue and total spending are aggregated before computing the ratio. This provides a true **business-level performance indicator**, emphasizing high-volume campaigns.
   - **Python's initial approach**, using .groupby().mean(), returned **unweighted ROAS**, which treats each campaign equally regardless of size. This was helpful in identifying **outliers or underperformers** that might be hidden in weighted averages.

5. **General Limitations:**
   - **Budget information is not provided**. Making it difficult to execute some core business health tasks.
   - **Customer/Audience type groups are not defined** beyond their basic titles. This stunted an in-depth analysis on customer behavior.
   - **Clear business priorities and expectations were not given**. A general and sweeping analysis was done with generalized business expectations instead.