Blinkit Sales Data Analysis

Name: Annanya Mojumder
College Name:Guru Nanak Institute Of Technology
Department : Computer Science and Engineering
Semester : 7th

Internship Under: Classroom

Phone: 8981838547
Website: https://classroomtech.in
LinkedIn: https://www.linkedin.com/company/classroom-tech/

Duration: 1st April 2025 – 30th June 2025

Acknowledgement

I would like to express my heartfelt gratitude to Classroom for providing me with this valuable internship opportunity.

I am especially thankful to Satyaki Das for his constant support, feedback, and guidance throughout the project.

Lastly, I am grateful to my peers and teammates who contributed to discussions and enriched my understanding.

Table of Contents

Objective / Problem Statement

- **Goal**: Analyze Blinkit sales to identify trends and improve strategies.

- **Questions:**
  - What are the most sold product categories?
  - When are peak order times?
  - Which cities highest sales?

Dataset Overview

Dataset Source: Kaggle

Number of rows and columns for each dataset:

- blinkit_customer_feedback.csv → 5000 rows × 8 columns
- blinkit_customer.csv → 2500 rows × 11 columns
- blinkit_delivery_performance.csv → 5000 rows × 8 columns
- blinkit_inventory.csv → 75172 rows × 4 columns
- blinkit_inventoryNew.csv → 18105 rows × 4 columns
- blinkit_marketing_performance.csv → 5400 rows × 11 columns
- blinkit_order_items.csv → 5000 rows × 4 columns
- blinkit_order.csv → 5000 rows × 10 columns
- blinkit_products.csv → 268 rows × 10 columns

Tools Used

- Power BI – For data modeling and visual storytelling
- GitHub – For version control and project sharing

Methodology

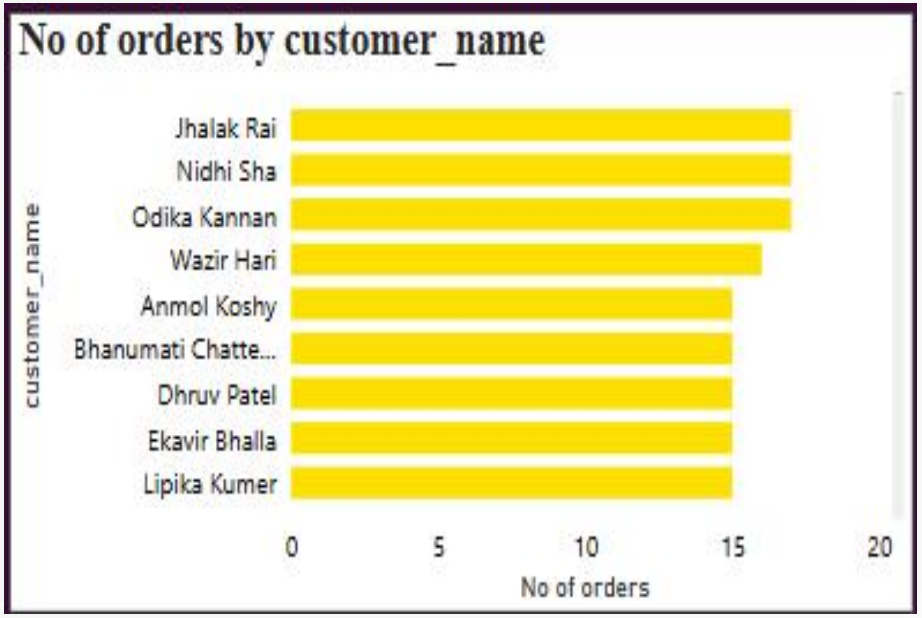
• Data Cleaning:

1. Removed nulls, handled outliers
2. Removed zero/negative price entries.
3. Removed outliers in stock, price, and delivery time to ensure accurate analysis.
4. Merged relevant tables using relationships

• Visualization:

1. Created KPI Cards to highlight key metrics like Total Revenue, Average Order Value, ROAS.
2. Used Bar & Column Charts to compare product sales, stock trends, and order quantities.
3. Applied Pie Charts to display customer segments and feedback categories.
4. Designed Scatter Plots & Maps to analyze delivery efficiency and order locations.
5. Implemented Funnel Charts & Decomposition Trees for campaign and sales breakdowns.

Visualization : Orders per Customer



Insights:

•Jhalak Rai, Nidhi Sha, and Odika Kannan are the top three repeat customers.

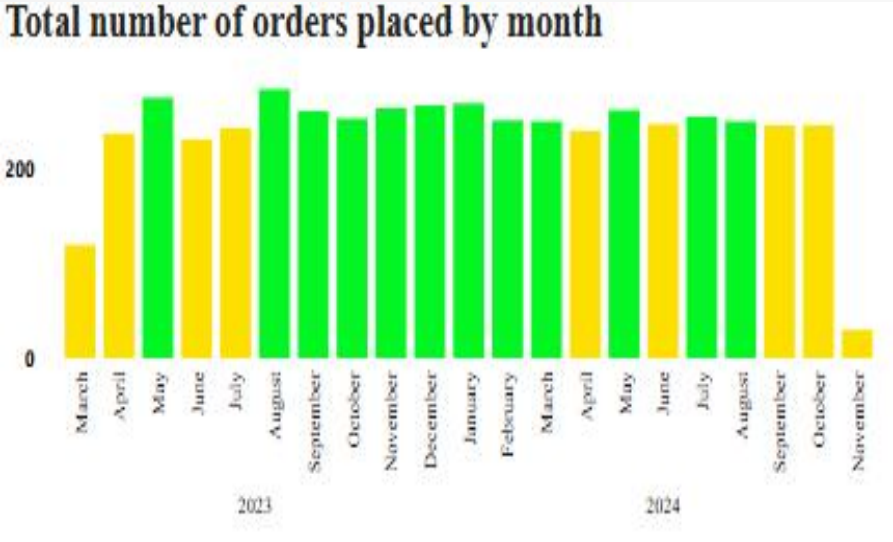
•Most top customers placed 15–19 orders over the period.

•Indicates strong customer loyalty and potential for targeted retention programs.

Purpose:

To identify which customers placed the highest number of orders.

Visualization : Monthly Order Volume



Insights:

•May and August 2023 had the highest order volumes (around 250+).

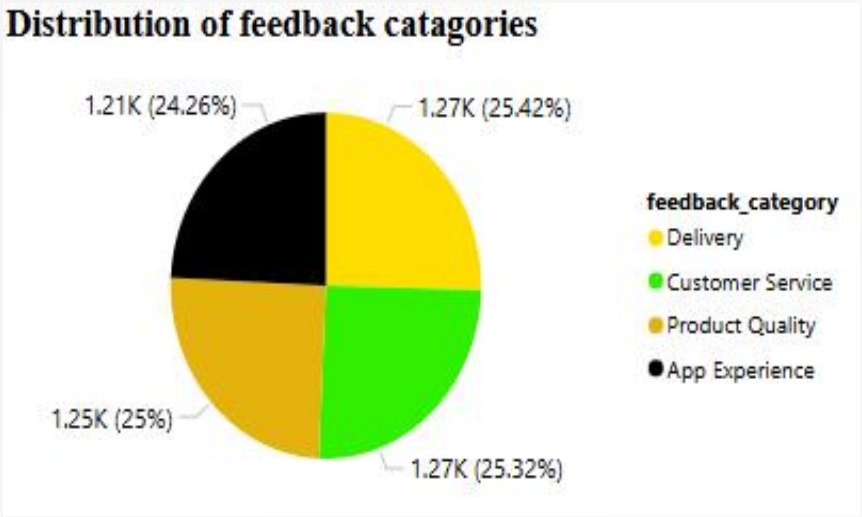
•A consistent order rate is seen from September 2023 to April 2024.

•Noticeable drop in orders in November 2024, likely due to incomplete data or seasonal decline.

Purpose:

To analyze order trends over time and identify high-performing months.

Visualization : Feedback Distribution



Insights:
Feedback is evenly distributed: Delivery, Customer Service, Product Quality, and App Experience are all around 25%.

Indicates balanced attention required in all areas of customer experience.

No major category dominates, showing diverse concerns among users.

Purpose:

To visualize how customer feedback is distributed across different categories.

Visualization : Customer Table

customer_id	customer_name	email
95912198	Bhavna Gara	zwagle@example.com
47577193	Chaaya Wable	zvenkatesh@example.net
73807097	Manbir Manda	zroy@example.org
77661491	Jyoti Srinivas	zramesh@example.org
74761384	Banjeet Gara	zpau@example.net
35049420	Udarsh Mangal	zpatil@example.org
22020776	Manan Banerjee	zpal@example.net
76441843	Neelima Chander	znadkarni@example.org
43243913	Rushil Iyer	zkrishnan@example.net
97492493	Alexander Ray	zkalla@example.com
27038630	Megha Thakkar	zinalsarna@example.com
90246977	Yuvraj Chacko	zinal84@example.org
49152878	Patrick Sandhu	zehaanmahajan@example.org
24000329	Aarush Dey	zdewan@example.net
61486140	Jeremiah Chand	zdeep@example.com
19080489	Reva Sha	zchad@example.com
38694209	Harinakshi Patla	zbhagat@example.com
38808614	Xavier Shetty	zbandi@example.net
79135318	Amaira Issac	zayyankala@example.org
65692224	Aahana Buch	zayyan92@example.com
20940403	Tanvi Kala	zayyan13@example.com
62971413	Baljiwan Agrawal	zayanmenon@example.org
17681645	Maanav Bhat	zayan80@example.net
16359370	Harsh Gaba	zayan66@example.org
46025543	Rachit Chatterjee	zayan61@example.org
13122744	Yashvi Dalal	zashilkaran@example.com

Insights:

Useful for segmentation, personalized campaigns, or communication.

Data includes over 20 customers with structured contact details.

Purpose:

To display customer IDs, names, and emails for analysis or reference.

Visualization : Orders by Delivery Status

order_id	delivery_status
60465	On Time
2237858	On Time
3101265	On Time
7550508	On Time
9408428	On Time
10161194	On Time
10448052	On Time
15642223	On Time
16878685	On Time
22830983	Significantly Delayed
23158044	On Time
27296314	Significantly Delayed
32604190	On Time
32613017	On Time
35135941	On Time
38408517	Significantly Delayed
39791095	On Time
42332130	On Time
47467466	On Time
47953423	On Time

Insights:

Majority of orders are marked as On Time.

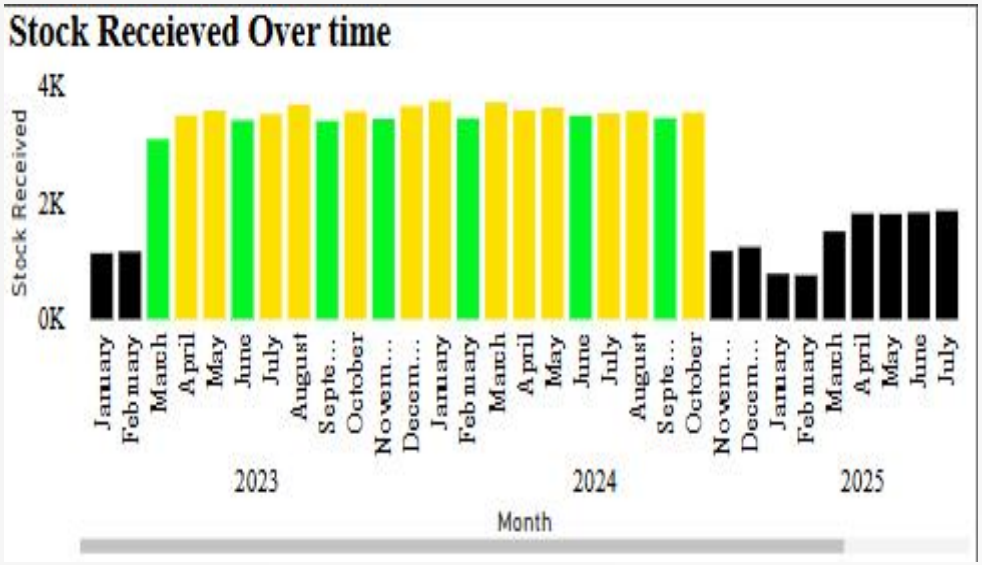
Only a small portion of orders are Significantly Delayed, indicating efficient delivery service.

Can be used to track performance of logistics over time.

Purpose:

To filter and view orders based on their delivery status (e.g., On Time, Significantly Delayed).

Visualization :Stock Received Over Time



Insights:

Peak stock receipts observed during mid-2023 to mid-2024.

A drop in stock is seen in late 2024, followed by a rise again in early 2025.

Helps identify inventory gaps and procurement patterns.

Purpose:

To analyze how much stock was received monthly from January 2023 to July 2025.

Visualization :Damaged Stock Percentage

product_id	Damaged Stock %
903336	32.69%
490602	33.41%
287070	33.45%
120532	33.95%
837480	34.52%
298985	34.54%
52729	34.97%
274207	35.75%
760684	35.81%
757274	35.90%
937542	36.20%
82484	37.26%
432617	37.47%
354409	37.48%
677169	37.93%
726958	38.18%
613942	38.40%
913428	38.74%
240179	38.85%
123983	39.04%
566052	39.83%
120385	40.16%

Insights:

All products listed have damaged stock percentages above 30%, indicating a quality control issue.

Highest damaged stock % is around 41.26%, which may need urgent supplier review.

Overall damage rate is 54.41%, a significant concern for operational costs.

Purpose:

To calculate and rank products based on the percentage of damaged stock.

Visualization :Campaign Performance

campaign_id	campaign_name	spend	revenue_generated
957924	App Push Notification	1,000.63	6,444.48
509838	Email Campaign	1,001.64	5,810.57
105015	Category Promotion	1,001.72	2,793.26
142246	Festival Offer	1,001.85	9,085.03
923137	Membership Drive	1,003.02	6,460.01
465005	Email Campaign	1,003.74	2,537.58
331583	Weekend Special	1,003.75	5,794.51
380639	Festival Offer	1,004.12	5,988.75
917018	Referral Program	1,005.00	4,625.85
821056	New User Discount	1,005.23	7,465.65
404539	App Push Notification	1,006.50	2,238.69
49757	Email Campaign	1,006.57	8,402.50
552750	App Push Notification	1,006.73	9,745.83
972675	Category Promotion	1,008.16	3,780.37
400626	Membership Drive	1,008.70	3,524.89
138944	Email Campaign	1,009.36	4,797.79
341661	App Push Notification	1,010.36	9,099.22

Insights:

Campaigns like Festival Offer and Referral Program have the highest revenue returns.

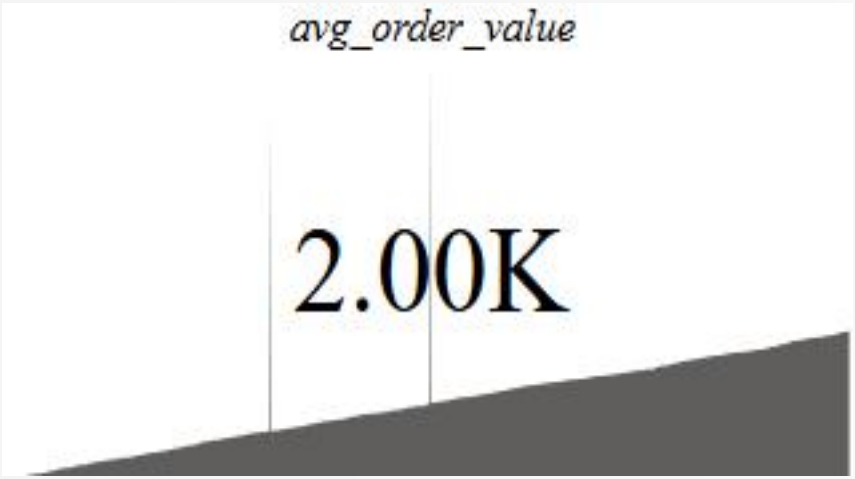
Category Promotion campaigns also show efficient revenue vs spend ratios.

Useful for planning future marketing budgets.

Purpose:

To review the ROI of different marketing campaigns..

Visualization : Average Order Value (KPI Card)



Insights:

The average order value is ₹2,000, suggesting customers typically spend a moderate amount per transaction.

Can help set promotional thresholds or free delivery minimums.

Purpose:

To display the average monetary value of each order across all transactions.

Visualization : Total Revenue from All Campaigns

Insights:

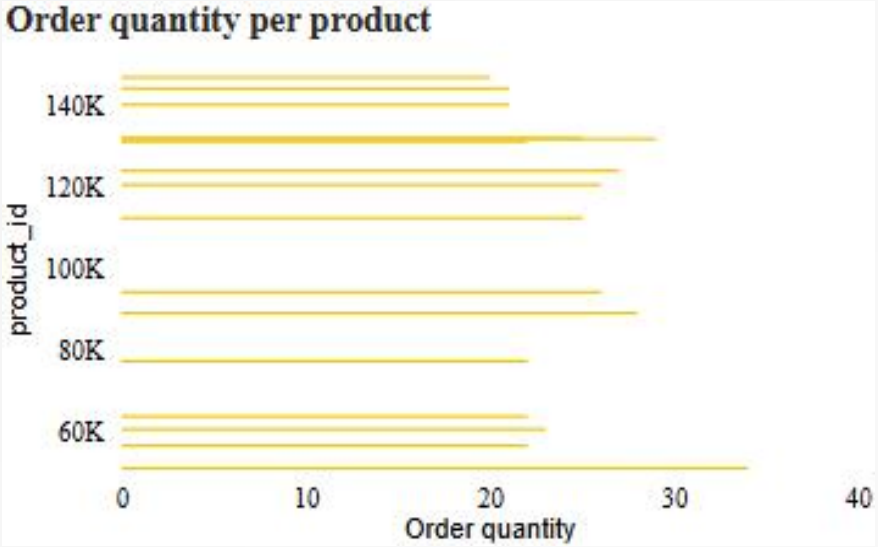
The cumulative revenue is ₹4.97 million, showcasing a successful overall campaign performance.

Can be used as a benchmark for future targets.

Purpose:

To display the total revenue generated from all sales or marketing campaigns.

Visualization : Product-wise Order Quantity



Insights:

Products like 992178 and 991443 have the highest order quantities.

A few products show relatively low order numbers, indicating low demand or visibility.

Purpose:

To compare how many times each product was ordered.

Visualization :Sales Revenue by Product

product_id	Total Sales Revenue
4452	6,625.32
6405	8,041.34
9436	3,890.92
11422	3,433.32
14145	25,233.60
15314	18,523.20
18035	46,509.12
26060	14,060.54
33797	7,993.08
33955	6,326.72
34186	56,464.65
34200	14,071.20
36412	19,088.64
39154	37,874.40
41853	29,134.08
51036	65,212.70

Insights:

Product 51036 alone generated over ₹65K in revenue.

Top 5 products contribute a significant portion of overall revenue, indicating their importance in inventory decisions.

Purpose:

To calculate and display revenue generated by individual products.

Visualization : Total Delivery Time(Card)

Insights:

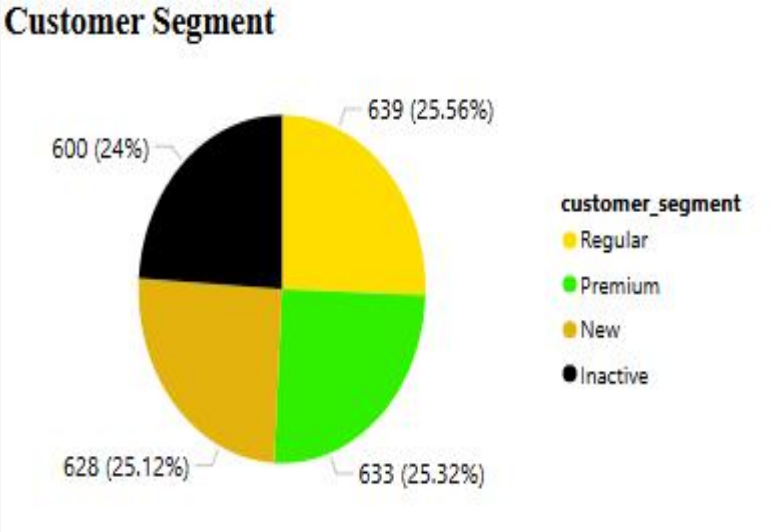
Total delivery time logged is 10 hours and 15 minutes.

Useful in analyzing efficiency and identifying time delays in logistics.

Purpose:

To calculate the cumulative delivery time using actual_time - promised_time.

Visualization : Customer Segment Distribution



Insights:

Segments are fairly balanced: Regular (25.56%), Inactive (24%),Premium(25.32%), New(25.12%).

The Premium segment is almost equal in size to Regular and New.

Helps in tailoring marketing strategies for each group.

Purpose:

To visualize how customers are segmented based on behavior or purchase history.

Visualization : Pincode Heatmap

Pincode	Frequency of Orders
1489	5
2246	2
3787	2
6749	1
6805	2
6864	3
6891	2
7186	1
7202	2
8262	4
8504	3
8953	2
9157	2
9792	2

Insights:

Pincode 1489 has the highest number of orders (5).

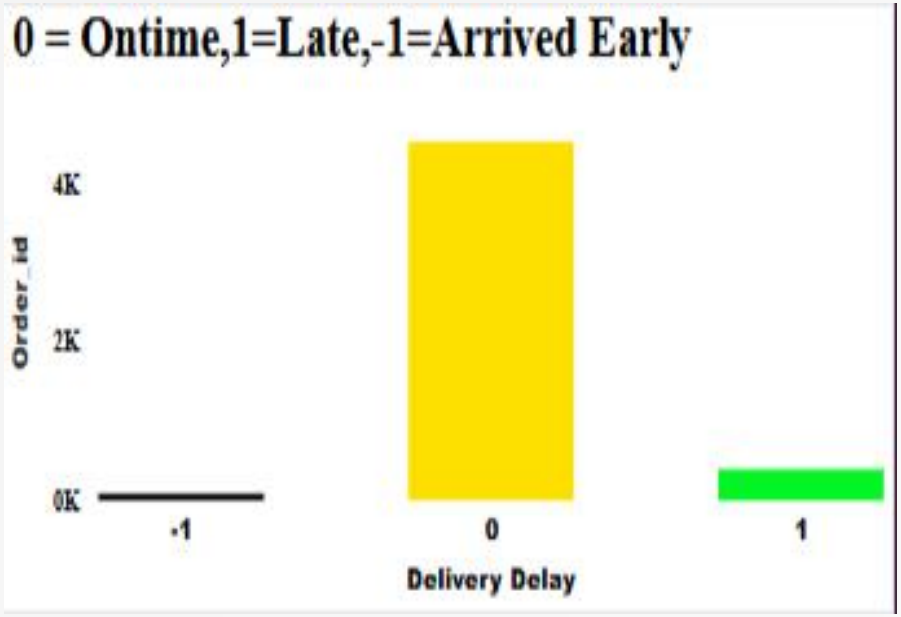
Most pincodes have 2–3 orders, indicating distributed demand.

Helps target areas with higher engagement for local promotions.

Purpose:

To analyze which locations (pincodes) have the most customer orders.

Visualization : Delivery Delay Analysis



Insights:

Majority of deliveries are on time.

A small fraction arrived early, while late deliveries are rare.

Demonstrates strong operational efficiency.

Purpose:

To visualize and categorize orders based on their delivery status: 0 = On Time, 1 = Late, -1 = Arrived Early.

Visualization : Campaign ROAS Comparison

campaign_name	ROAS1
App Push Notification	1.99
Category Promotion	1.94
Email Campaign	1.99
Festival Offer	1.95
Flash Sale	1.94
Membership Drive	1.97
New User Discount	1.97
Referral Program	2.03
Weekend Special	1.98
Total	1.97

Insights:

Referral Program has the highest ROAS (2.03), indicating great efficiency.

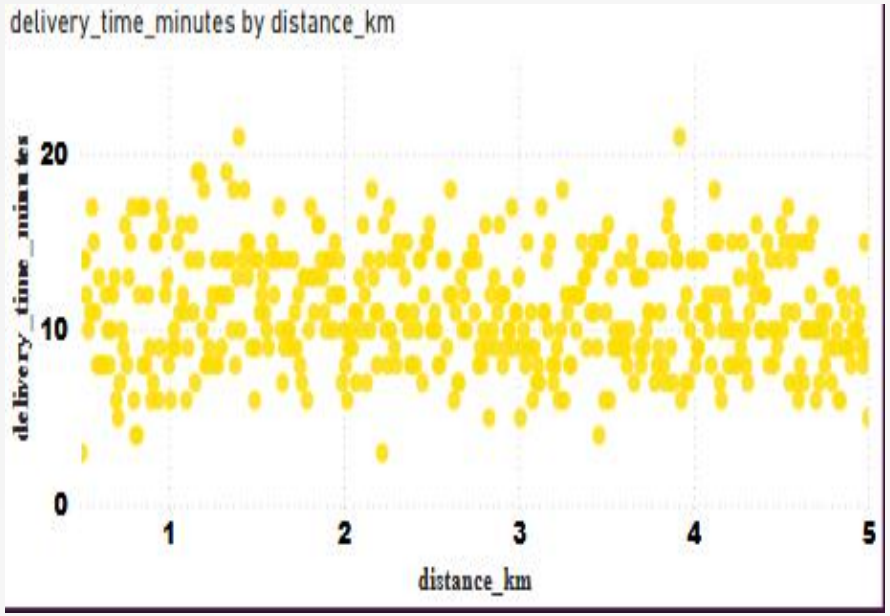
Most campaigns have ROAS around 1.95–2.00, reflecting consistent performance.

Can be used to optimize budget allocation in future campaigns.

Purpose:

To measure and compare the Return on Ad Spend across multiple marketing campaigns.

Visualization : Delivery Efficiency Analysis (Scatter Plot)



Insights:

No strong correlation: delivery time fluctuates regardless of distance.

Indicates other factors (traffic, time of day, logistics) may influence delivery time more than distance.

Useful for performance optimization.

Purpose:

To analyze the relationship between delivery distance and delivery time.

Visualization : Customer Retention Rate

Insights:

Retention rate is 94.20%, indicating excellent customer loyalty.

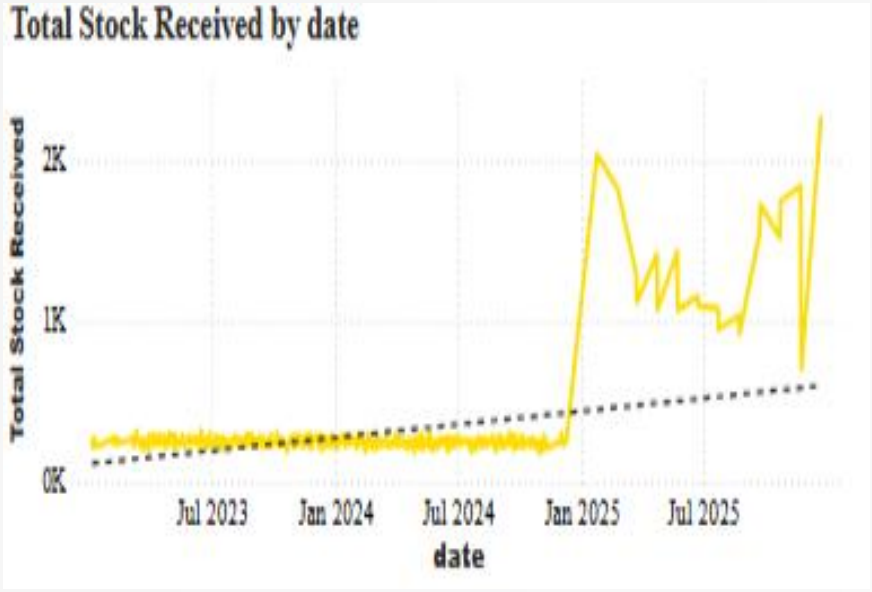
Suggests successful user experience and repeat engagement.

Useful for projecting future revenue and customer lifetime value.

Purpose:

To measure how many customers return to place another order.

Visualization : Total Stock Received Over Time with Forecast



Insights:

Stock received increases sharply from early 2025.

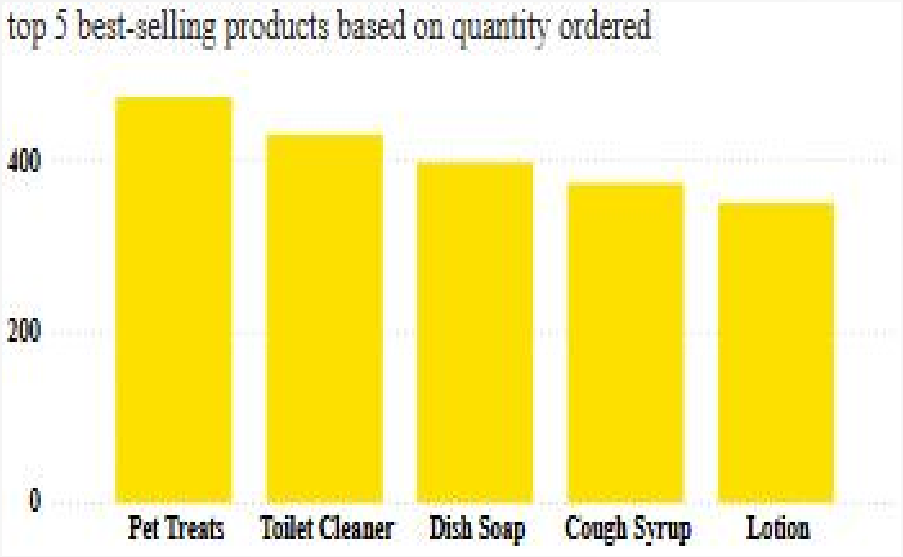
The dotted trendline suggests a continued upward movement in demand.

Important for procurement and warehouse planning.

Purpose:

To forecast future stock needs using trends from past received stock data.

Visualization : Top 5 Best-Selling Products



Insights:

Pet Treats and Toilet Cleaner are top sellers.

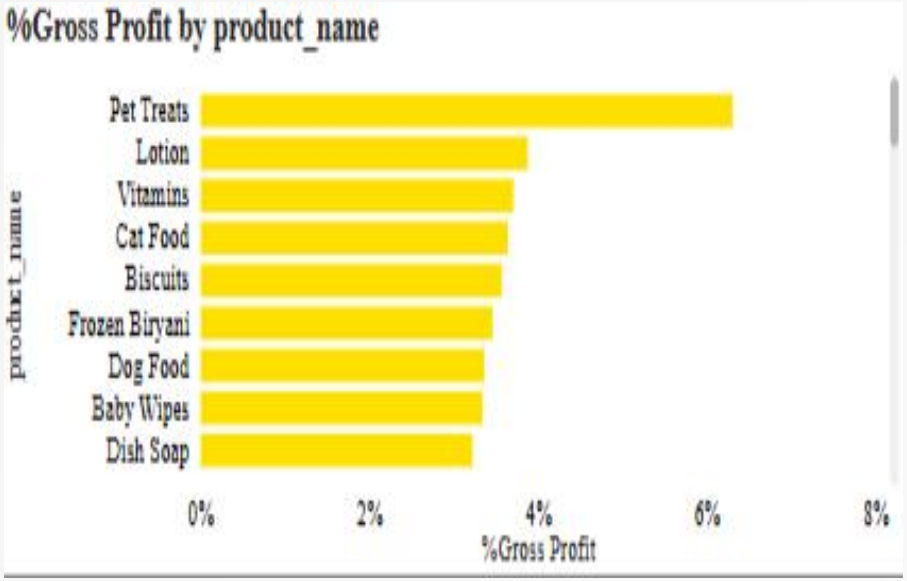
All top 5 products have over 350 units sold.

Indicates where inventory priority and promotions should focus.

Purpose:

To identify the most frequently ordered products.

Visualization : Gross Profit by Product



Insights:

Pet Treats has the highest gross profit (~8%).

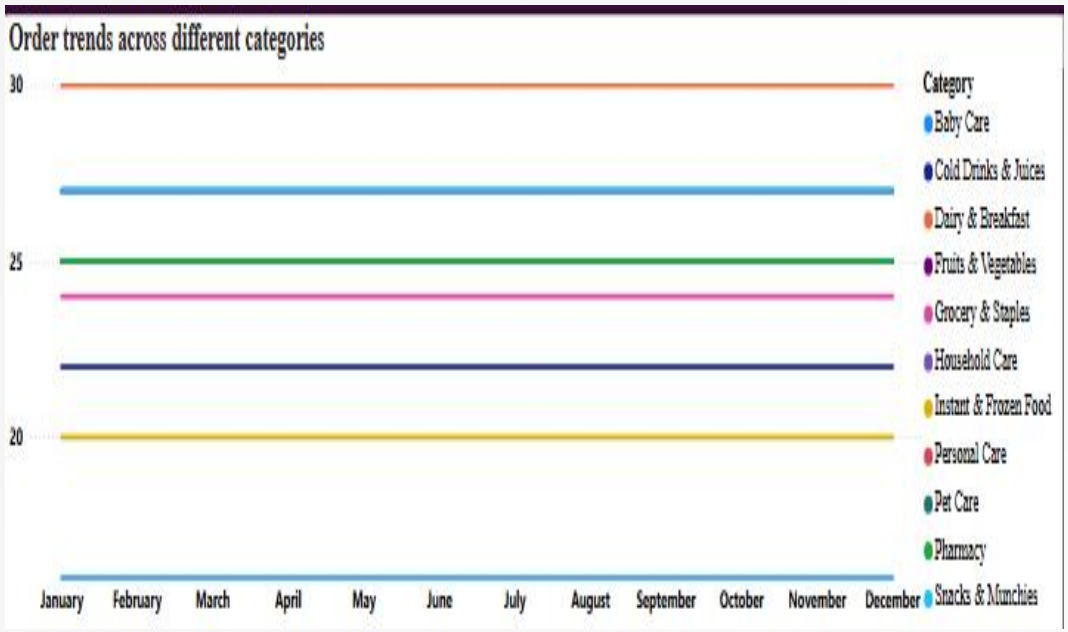
Dish Soap and Baby Wipes have the lowest margins (~2%).

Helps optimize pricing and marketing around high-margin items.

Purpose:

To calculate and compare gross profit margins for individual products

Visualization : Order Trends Across Different Categories



Insights:

Baby Care consistently has the highest order volume (~30 orders monthly).

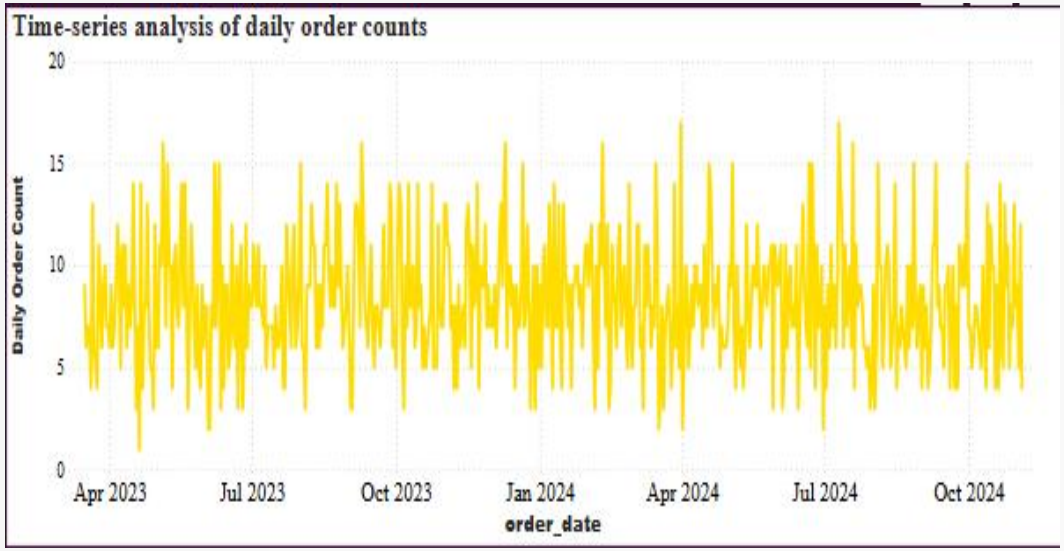
Pharmacy and Instant & Frozen Food see relatively lower order volumes.

Trends remain stable over the year—suggesting consistent customer preference.

Purpose:

To analyze how different product categories perform over the months.

Visualization : Time-Series Order Analysis



Insights:
Orders remain steady, with minor peaks and troughs.

Some short-term spikes might align with campaigns or holidays.

No major long-term increase/decrease—indicates market stability.

Purpose:

To observe daily order volume fluctuations and trends over 2023–2024.

Visualization : Product Category vs Order Quantity

Product_category	Order_quantity
Dairy & Breakfast	566
Household Care	509
Pet Care	501
Fruits & Vegetables	492
Snacks & Munchies	483
Pharmacy	481
Personal Care	454
Grocery & Staples	449
Cold Drinks & Juices	375
Instant & Frozen Food	356
Baby Care	334
Total	5000

Insights:

Dairy & Breakfast leads with 566 units, followed by Household Care and Pet Care.

Lower quantities in Baby Care and Cold Drinks indicate seasonal or niche demand.

Ideal for inventory planning and category-specific promotions.

Purpose:

To display which categories have the highest cumulative product orders.

Visualization : Customer Lifetime Value

Customer_id	Customer_Lifetime_Value
75732571	39634
90666073	39389
35237846	39373
25221501	38717
82374625	38473
49769532	38468
3893903	38454
72415263	38284
42366724	37925
1342184	37877
2361575	37869
17521263	37828
43256928	37706
7288562	37673
81183415	37633
96109948	37583
39261715	37512
93522848	37505
43761120	37141

Insights:

Top customers show lifetime values over ₹39,000, with a long-term revenue potential.

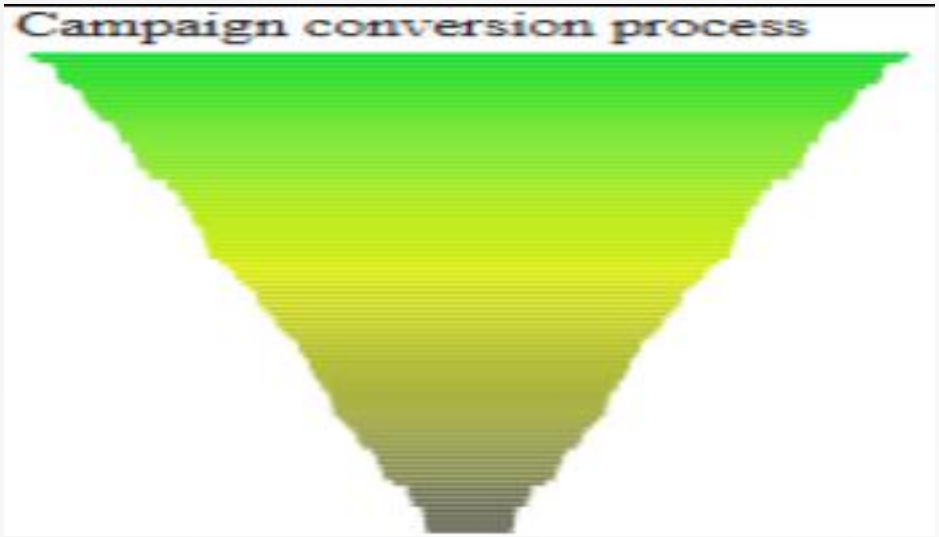
LTV distribution helps segment premium vs casual buyers.

Crucial for targeted loyalty campaigns.

Purpose:

To calculate lifetime value based on avg_order_value × total_orders.

Visualization : Campaign Conversion Funnel



Insights:

Sharp drop at each stage shows potential leaks in the funnel.

Indicates high drop-off before final conversion.

Helps refine campaign messaging and targeting at each level.

Purpose:

To visualize how users progress through stages of a marketing campaign.

Visualization : Filter Orders by Delivery Status(Using Slicer)

Insights:

Interactive slicer allows quick filtering by delivery status.

Current view shows all selected orders as On Time.

Great tool for real-time logistics performance tracking.

Purpose:

To dynamically filter and display orders based on delivery performance (On Time, Slightly Delayed, Significantly Delayed).

Visualization : Merged Stock Report

product_id	Sum of damaged_stock	Total Stock Received
4452	306	769
6405	283	660
9436	289	402
11422	300	599
14145	296	659
15314	285	597
18035	309	607
26060	307	853
33797	305	829
33955	317	663
34186	312	581
34200	276	634
36412	287	551
39154	322	656
41853	317	592
51036	298	509
52729	266	875
53833	330	639
56589	307	869
57405	304	702
60619	298	578
63769	303	620
67197	290	551
70969	327	633

Insights:

Products like 6405 and 9436 have damage counts over 280 units.

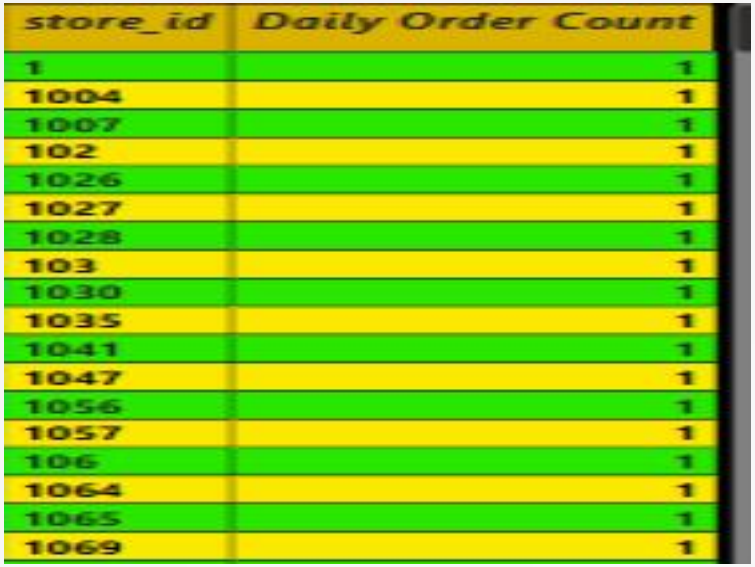
Total Stock Received ranges between 400 and 900 per product.

Useful for identifying suppliers or SKUs with consistent quality issues.

Purpose:

To analyze and compare total received stock and damaged stock per product.

Visualization : Daily Order Count by Store



store_id	Daily Order Count
1	1
1004	1
1007	1
102	1
1026	1
1027	1
1028	1
103	1
1030	1
1035	1
1041	1
1047	1
1056	1
1057	1
106	1
1064	1
1065	1
1069	1

Insights:

All listed stores have exactly 1 order in the dataset (current filtered view).

Total of 3,470 orders processed—shows high store engagement.

Ideal for evaluating store-wise order fulfillment performance.

Purpose:

To track the number of orders processed per store location.

Visualization : Geographic Order Distribution

Insights:

High order density in metro cities and Tier-1 zones.

North and South India show more coverage than the Northeast.

Supports regional expansion and logistics resource allocation.

Purpose:

To visualize which regions place the most orders.

Visualization : Product-wise Discount Percentage

product_name	Discount %
Baby Food	30.00
Baby Wipes	30.00
Bananas	25.00
Biscuits	35.00
Bread	20.00
Butter	20.00
Carrots	25.00
Cat Food	35.00
Cereal	20.00
Cheese	20.00
Chips	35.00
Chocolates	35.00
Cola	30.00
Cookies	35.00

Insights:

.Products like Chocolates, Cookies, and Chips offer up to 35% discount.

Majority of items fall in the 20%–30% discount range.

Useful for identifying top-discounted products to promote.

Purpose:

To show the calculated discount percentages by comparing Price and MRP.

Visualization : Live Order Delivery Status

order_id	customer_id	delivery_status
4054828941	99893898	Significantly Delayed
6331115638	99729547	On Time
2458349464	99714944	On Time
9608676386	99714944	On Time
1918416669	99695529	On Time
199903168	99627097	On Time
239830094	99627097	On Time
3907403745	99490786	On Time
5840290569	99490786	On Time
2732162016	99361865	On Time
7913167260	98307019	Significantly Delayed
1803155516	98088303	On Time
3748957111	98088303	Slightly Delayed
4423613746	98088303	Slightly Delayed
8310385249	97315135	On Time

Insights:

Most deliveries are On Time, with a few marked as Slightly Delayed.

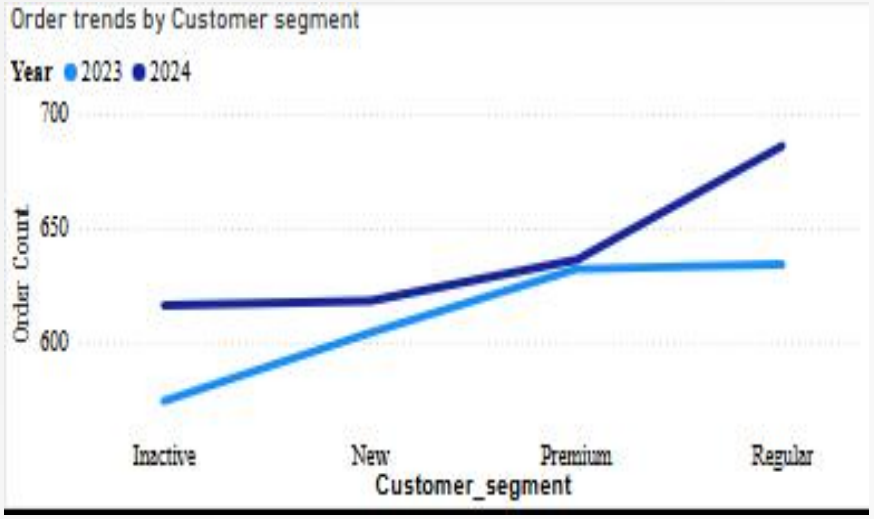
Helps customer service teams monitor exceptions in real time.

Useful for immediate operational decision-making.

Purpose:

To track the live status of orders and delivery outcomes.

Visualization : Order Trends by Customer Segment (Drill-down)



Insights:

Regular and Premium customers show an upward order trend into 2024.

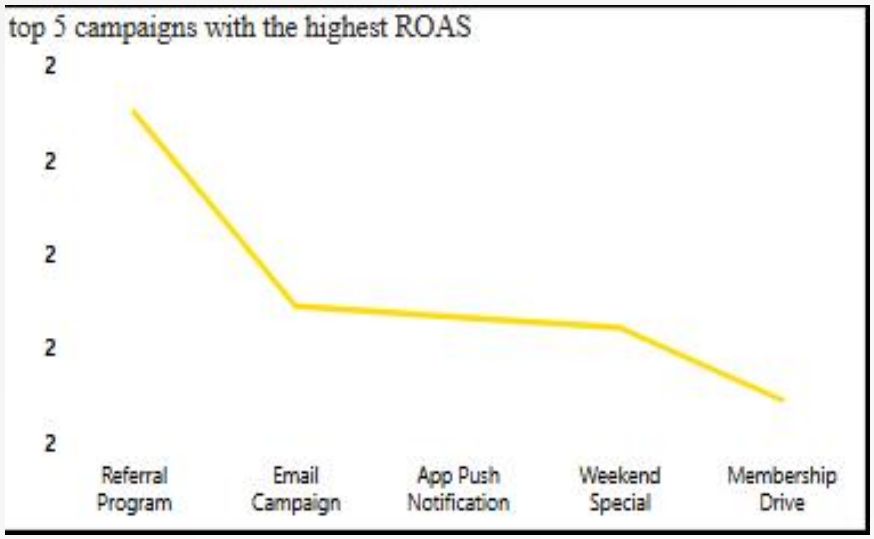
Inactive segment shows a flat trend, indicating limited re-engagement.

Drill-down can further reveal trends by region, campaign, or product.

Purpose:

To analyze how order volume trends across segments (Inactive, New, Premium, Regular) over time.

Visualization : Top 5 ROAS-Performing Campaigns



Insights:

Referral Program delivers the best return.

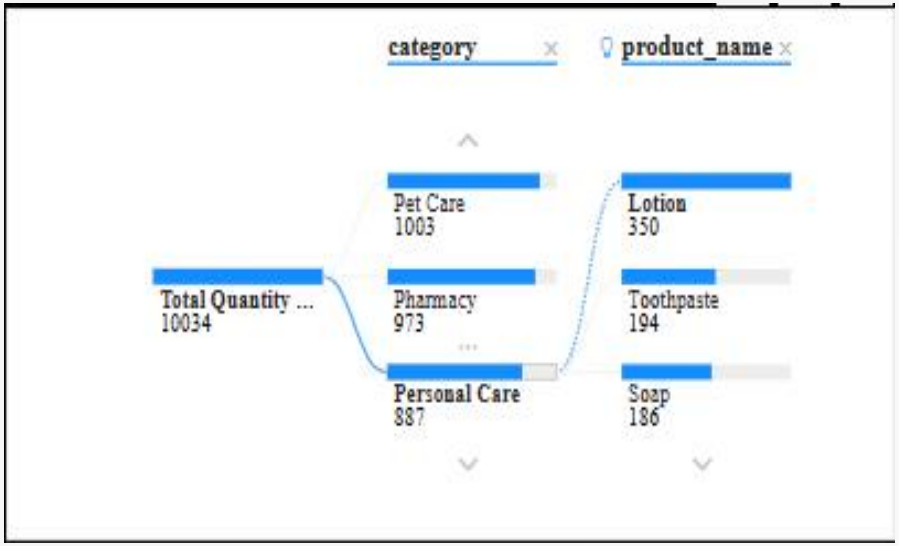
Campaigns like Member Drive underperform in comparison.

Can inform future ad budget prioritization.

Purpose:

To identify the most efficient marketing campaigns using ROAS metric.

Visualization : Decomposition of Product Sales by Category



Insights:

Pharmacy and Pet Care are top categories based on quantity sold.

Within Personal Care, items like Toothpaste and Soap perform well.

Great for multi-level drill-down analysis and exploring sales hierarchies.

Purpose:

To visually break down total sales into category-wise and product-wise contributions.

Key Findings

1. Customer retention is high at 94.20%, with Regular and Premium segments placing the most orders.

2. Pet Treats is both the best-selling and highest-margin product.

3. The Referral Program delivers the highest ROAS among all campaigns.

4. Most deliveries are on time, but some products show recurring damaged stock.

5. Sales trends, order heatmaps, and store-level data support smarter regional planning.

Conclusion

Through this analysis, I learned how to use Power BI to clean data, build interactive dashboards, and extract meaningful business insights. It helped me understand customer behavior, sales trends, and campaign performance using real-world datasets. This project also strengthened my skills in data visualization, DAX, and report storytelling. The insights generated can guide decision-makers in optimizing product offerings, improving delivery efficiency, and allocating marketing budgets more effectively. Overall, this analysis supports data-driven strategies to enhance customer satisfaction and business growth.

References

Dataset Source: [Kaggle – Blinkit Sales Dataset]

Visualization Tool: Microsoft Power BI

Report Hosting: GitHub (Your project repository link)

Internship Platform: Classroom Tech

Official LinkedIn: Classroom Tech – https://www.linkedin.com/company/classroom-tech/

Thank You

Name : Annanya Mojumder

Email : annanyamajumder22@gmail.com

LinkedIn : https://www.linkedin.com/in/annanya-mojumder-3734a81a4/