Product Demand Forecasting and Sales Trend Analysis for an Online Grocery Delivery Platform

A Final report for the BDM capstone Project (Analysis with Secondary Data)

Submitted by

Name: HARSH VARDHAN KUMAR

Roll number: 24DS2000078

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1) Executive Summary:

Blinkit is an online grocery delivery platform in India that operates in a fast-paced and high-demanding environment, where consumer purchasing behavior exhibits high variability across various product categories, time periods, and price fluctuations. One of the key operational challenges of Blinkit is its capability to forecast product demand, an aspect that has a direct impact on product inventory management, customer satisfaction, and the operational efficiency of the company.

This project uses historical sales data (2023-2024, ~5,000 records) to develop a product-level demand forecasting model and perform time series analysis.

This project followed a structured analytical approach, starting with loading all relevant data files into DataFrames. The next step involved data cleaning, addressing missing values, duplicate records, inconsistent date formats, and potential outliers to ensure reliable analysis. A detailed Exploratory Data Analysis (EDA) was conducted to uncover key patterns in customer orders, including trend analysis, seasonality detection, category-wise demand variations, and identification of anomalies. The analysis highlighted recurring weekly cycles, notable differences in behaviour across product categories, and a recent decline in sales that may reflect operational changes or partial data capture.

To support inventory planning and operational decisions, two time-series forecasting models were applied: ARIMA *(AutoRegressive Integrated Moving Average)* and Exponential Smoothing *(Holt-Winters)*. Between the two, the Holt-Winters model provided more stable and seasonally consistent forecasts for this dataset. These forecasts offer actionable insights to anticipate future demand, optimize stock allocation, and reduce the risk of overstocking or stockouts, thereby supporting better overall operational planning.

2) Proof of Originality:

Source: Kaggle (www.kaggle.com)

Dataset Name: Blinkit Sales Dataset

Dataset Link: https://www.kaggle.com/datasets/akxiit/blinkit-sales-dataset

Collaborator: Akshit Vaghasiya (owner)

License: MIT

Analysis Notebook Link (Google Colab): https://colab.research.google.com/drive/19q3mD7cUyQh7NRhPtp61dALKvZc6dDup?usp=sharing

GitHub Repository Link: https://github.com/harshvk25/bdmproject2025

3) Metadata and Descriptive Statistics:

Metadata: The full dataset has 9 csv files and 2 excel files.

DATASET	ROWS, COLUMNS	KEY COLUMNS	PURPOSE
blinkit_customers.csv	(2500,11)	customer_id, total_orders	Customer profiles
blinkit_orders.csv	(5000,10)	order_id, order_total, delivery_partner_id	Order & delivery performance
blinkit_order_items.csv	(5000,4)	order_id, product_id, quantity	Order-level details
blinkit_products.csv	(268,10)	product_id, price, category	Product metadata
blinkit_customer_feedback.csv	(5000,8)	feedback_id, rating	Customer feedback
blinkit_inventory.csv	(75172,4)	product_id, stock_received	Stock tracking
blinkit_inventoryNew.csv	(18105,4)	product_id, stock_received	Updated stock levels
blinkit_marketing_performance.csv	(5400,11)	campaign_id, roas	Marketing analytics
blinkit_delivery_performance.csv	(5000,8)	order_id, delivery_time_minutes	Delivery speed & delays

Table 1: Metadata

Descriptive Statistics:

- **SHAPE OF THE DATASET:**
  1. Total no. of rows: 5000
  2. Total no. of columns/features: 54

After applying **describe() method** in **bt_master** dataframe, we have the following information:

DESCRIPTIVE STATISTICS FOR NUMERICAL COLUMNS

COLUMN NAME	MEAN	MEDIAN	STANDARD DEVIATION
order_id	5029129224.19	5074377551.0	2863532520.89
customer_id	50096845.22	49978078.5	29190820.31
order_total	2201.86	2100.69	1303.02
delivery_partner_id	50050.32	50262.5	28802.28
store_id	4999.69	4987.0	2886.09
phone	915072800563.19	915023401479.0	2921351067.66
pincode	497579.18	499186.0	281174.2
total_orders	10.48	10.0	5.8
avg_order_value	1092.34	1114.5	523.06
product_id	509974.94	540618.0	293678.31
quantity	2.01	2.0	0.82
unit_price	493.16	448.16	298.08
price	493.16	448.16	298.08
mrp	685.91	643.82	418.23
margin_percentage	27.71	30.0	7.43
shelf_life_days	235.24	365.0	151.26
min_stock_level	20.49	21.0	5.98
max_stock_level	75.12	74.0	14.5
delivery_time_minutes	4.44	2.0	8.06
distance_km	2.72	2.69	1.29
rating	3.34	4.0	1.19
order_hour	11.52	12.0	6.95
order_quarter	2.56	3.0	1.01
day_of_month	15.78	16.0	8.78
week_of_year	27.06	27.0	13.3
delivery_delay	4.44	2.0	8.06
order_value_per_customer	11494.44	9162.75	8992.22
stock_received	68.99	71.0	9.39
damaged_stock	3.81	3.0	1.83
total_item_value	994.48	795.36	774.16
available_stock	65.18	67.0	9.12
stock_sufficiency_ratio	39.76	33.5	19.53

Table 2: Descriptive Statistics for Numerical columns

DESCRIPTIVE STATISTICS FOR CATEGORICAL COLUMNS

COLUMN NAME	COUNT	UNIQUE	TOP	FREQUENCY
order_delivery_status	5000	3	On Time	3470
payment_method	5000	4	Card	1285
customer_name	5000	2166	Nidha Sha	9
email	5000	2169	ewali@example.org	9
address	5000	2172	H.No. 174, Saraf Chowk, Deoghar-906819	9
area	5000	316	Orai	44
customer_segment	5000	4	Regular	1320
product_name	5000	51	Pet Treats	233
category	5000	11	Dairy & Breakfast	566
brand	5000	267	Karnik PLC	34
sentiment	5000	3	Neutral	1738
partner_delivery_status	5000	3	On Time	3470
reasons_if_delayed	5000	2	Traffic	3098
order_day	5000	7	Wednesday	744

Table 3: Descriptive Statistics for Categorical columns

4) Detailed explanation of Analysis process:

4.1) Data Preparation:

Objectives of Data Preparation:

- Handle missing values and duplicates.
- Convert date columns to datetime format.
- Ensure consistent data types across datasets.
- Create new calculated fields (feature engineering).
- Merge related datasets to create master dataset for analysis

4.1.1) Data Import and Inspection:

LOADING THE BLINKIT DATASET

```
bt_custfeed = pd.read_csv("blinkit_customer_feedback.csv")
bt_cust = pd.read_csv("blinkit_customers.csv")
bt_delp = pd.read_csv("blinkit_delivery_performance.csv")
bt_invt = pd.read_csv("blinkit_inventory.csv")
bt_invtnew = pd.read_csv("blinkit_inventoryNew.csv")
bt_markper = pd.read_csv("blinkit_marketing_performance.csv")
bt_orditm = pd.read_csv("blinkit_order_items.csv")
bt_ordr = pd.read_csv("blinkit_orders.csv")
bt_prod = pd.read_csv("blinkit_products.csv")
bt_caticon = pd.read_excel("Category_Icons.xlsx")
bt_raticon = pd.read_excel("Rating_Icon.xlsx")
```

Figure 1: Loading all files to create DataFrames in Google Colab (using pandas package)

4.1.2) Data Cleaning:

- Checking missing values for all dataframes using **isna()** method
- Handling missing values, formatting dates and removing duplicates

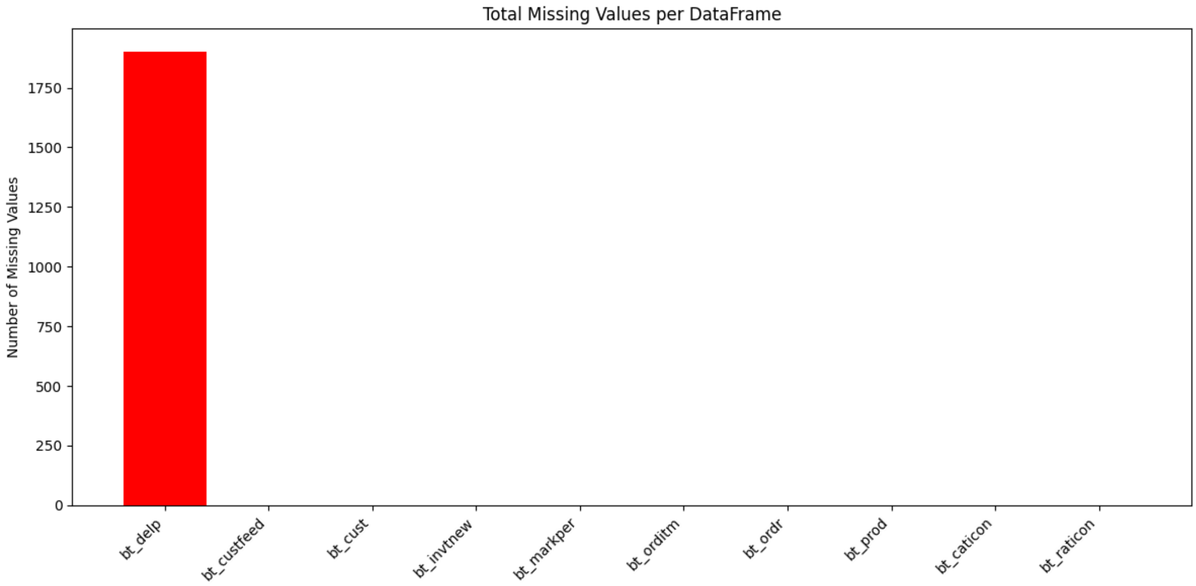

Figure 2: Visualizing no. of missing values in each DataFrame

4.1.3) Data Integration and Feature Engineering:

- Merged dataframes on common keys (i.e. **order_id**, **product_id**, **customer_id**).
- Create derived features for analytical insights and future modelling
- Created a master dataframe for analysis (**bt_master**)

MASTER DATAFRAME INFORMATION

S.NO	COLUMN	DATATYPE	TOTAL NO. OF ROWS
1	order_id	int64	5000
2	customer_id	int64	5000
3	order_date	datetime64	5000
4	promised_delivery_time	datetime64	5000
5	actual_delivery_time	datetime64	5000
6	order_delivery_status	object	5000
7	order_total	float64	5000
8	payment_method	object	5000
9	delivery_partner_id	int64	5000
10	store_id	int64	5000
11	customer_name	object	5000
12	email	object	5000
13	phone	int64	5000

14	address	object	5000
15	area	object	5000
16	pincode	int64	5000
17	registration_date	datetime	5000
18	customer_segment	object	5000
19	total_orders	int64	5000
20	avg_order_value	float64	5000
21	product_id	int64	5000
22	quantity	int64	5000
23	unit_price	float64	5000
24	product_name	object	5000
25	category	object	5000
26	brand	object	5000
27	price	float64	5000
28	mrp	float64	5000
29	margin_percentage	float64	5000
30	shelf_life_days	int64	5000
31	min_stock_level	int64	5000
32	max_stock_level	int64	5000
33	delivery_time_minutes	float64	5000
34	distance_km	float64	5000
35	promised_time	datetime	5000
36	actual_time	datetime	5000
37	rating	int64	5000
38	sentiment	object	5000
39	partner_delivery_status	object	5000
40	reasons_if_delayed	object	5000
41	month_year	period[M]	5000
42	order_day	object	5000
43	order_hour	int32	5000
44	is_weekend	bool	5000
45	order_quarter	int32	5000
46	day_of_month	int32	5000
47	week_of_year	int32	5000
48	delivery_delay	float64	5000
49	order_value_per_customer	float64	5000
50	stock_recieved	int64	5000
51	damaged_stock	int64	5000
52	total_item_value	float64	5000
53	available_stock	int64	5000
54	stock_sufficiency_ratio	float64	5000

Table 4: bt_master DataFrame information

4.2) Sales Trend Analysis:

Objective of Sales Trend Analysis:

- To identify patterns and insights across Sales trends, Product performance, Customer segments, and Damaged stocks to guide strategic business decisions.

Analysis Methodologies:

1. **Monthly sales trend**

   Objective: To understand how total sales fluctuate month-to-month and identify seasonality, growth patterns, or demand shifts.

2. **Top Performing products & brands**

   Objective: To identify which products and brands contribute most to revenue, order frequency, and customer preference.

3. **Weekly sales trend**

   Objective: To observe short-term sales variation and operational performance at a weekly level.

4. **Customer Segmentation Analysis**

   Objective: To group customers based on purchasing behavior and analyze the profitability of each segment.

5. **Damaged Stock Analysis**

   Objective: To measure how much stock is damaged and assess its impact on profitability and inventory efficiency.

4.3) Time Series Forecasting (Predictive Modeling):

Objective of Time Series Forecasting:

- The main goal is to **forecast future product demand (weekly basis)** for better demand planning.

Methodologies:

1. **ARIMA Model (AutoRegressive Integrated Moving Average):**
   ARIMA is a classical statistical forecasting model that captures autocorrelation, trends, and patterns in time-dependent data.
2. **Exponential Smoothing Model (Holt-Winters)**: Holt-Winters is ideal for time series with trend and seasonality. It applies exponentially decreasing weights to past observations, allowing more emphasis on recent data.

5) Results and findings:

5.1) Results and findings of Sales Trend Analysis:

5.1.1) Monthly Sales Trend:

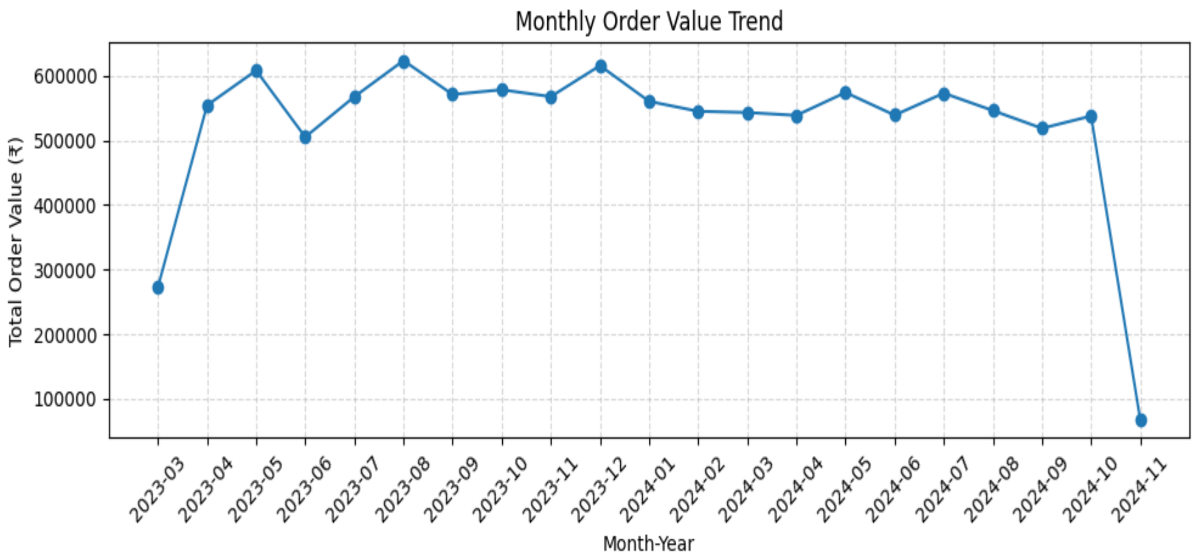

Figure 3: Monthly Sales Trends Analysis

Results & Findings:

- **Early Growth Phase (March–April 2023):**
  1. The total monthly order value rises sharply during this period, increasing from roughly ₹270,000 to more than ₹600,000.
     This early surge may reflect the effect of a strong launch phase, successful promotional efforts, or a seasonal boost in customer activity.

- **Period of Consistency (Mid-2023 to Mid-2024):**
  1. Between May 2023 and August 2024, monthly sales remain relatively steady, generally falling between ₹520,000 and ₹620,000.

- **Drop in Recent Months (Late 2024):**
  1. A noticeable decline appears toward the end of the timeline, where monthly sales fall below ₹100,000.
  2. This dip may be due to either:
     a. Incomplete data for the final recorded month, or
     b. A real decline resulting from reduced demand, fewer marketing activities, or possible supply-side challenges.

- **Limited Evidence of Seasonality:**
  1. Apart from the end-period drop, the series does not show recurring peaks or dips across years.
  2. The absence of strong seasonal patterns indicates stable customer demand throughout the year, which is advantageous for budgeting, procurement, and inventory planning.

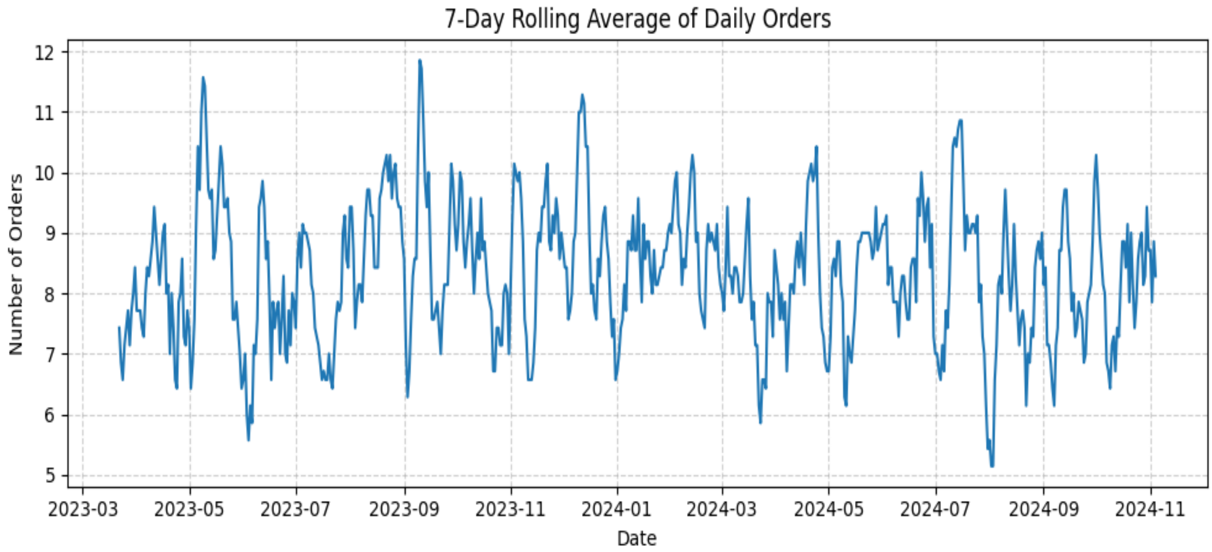

Figure 4: Weekly moving average

Results & Findings:

- **Overall Trend:**
  1. From March 2023 to November 2024, the smoothed trend shows a generally steady demand pattern, even though the raw daily values fluctuate.
  2. The 7-day rolling average helps filter out day-to-day noise and reveals the underlying weekly behaviour.
  3. On average, the platform receives 7 to 10 orders per day, indicating consistent customer engagement across the period.

- **Peaks and Dips:**
  1. The chart displays recurring spikes, with order volumes occasionally rising to 11–12 orders per day.
  2. Brief downward movements, sometimes falling to around 5 orders per day, may reflect quieter periods, operational delays, or natural short-term demand drops.

- **Demand Stability:**
  1. There is no strong long-term upward or downward movement, suggesting that demand remains stable over the entire timeframe.
  2. These mild cycles make the data well-suited for time-series models like **ARIMA** or **Holt–Winters**, which capture recurring patterns effectively.

5.1.2) Top-Performing Products & Brands:

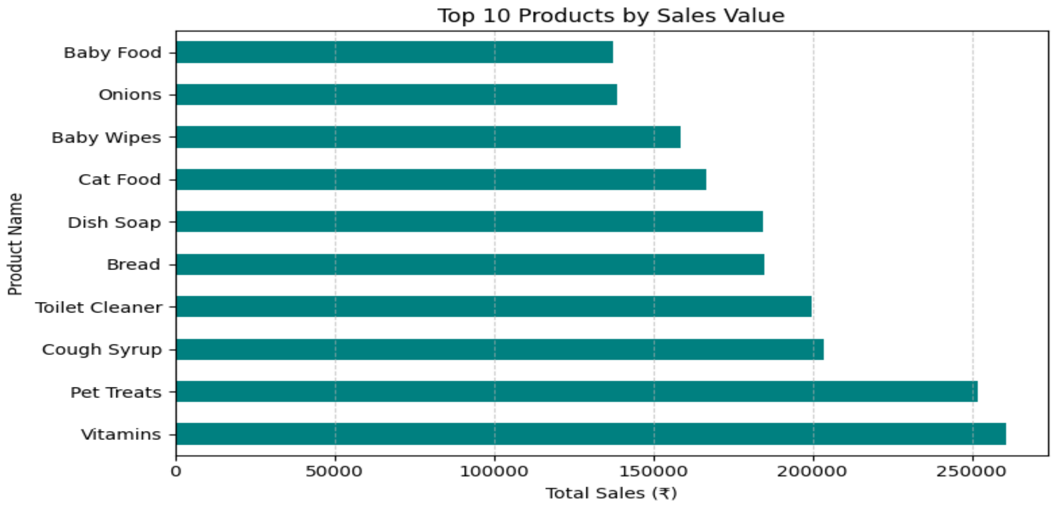

Figure 5: Top Performing products by revenue

Results & Findings:

- **Top Performers:**
  1. Vitamins lead the chart with total sales exceeding ₹260,000, making it the highest-selling category.
  2. Pet Treats follow closely at around ₹250,000, highlighting strong and consistent spending on pet-related products.
  3. Both categories represent non-essential but high-priority purchases, indicating that customers are willing to spend on wellness and pet care despite these not being daily-need items.

- **Strong Performing Tier:**
  1. Categories such as Cough Syrup, Toilet Cleaner, Bread, Dish Soap, and Cat Food fall in the ₹180,000–₹210,000 range.
  2. These items cover a mix of healthcare essentials and routine household consumption, which naturally maintain steady demand.

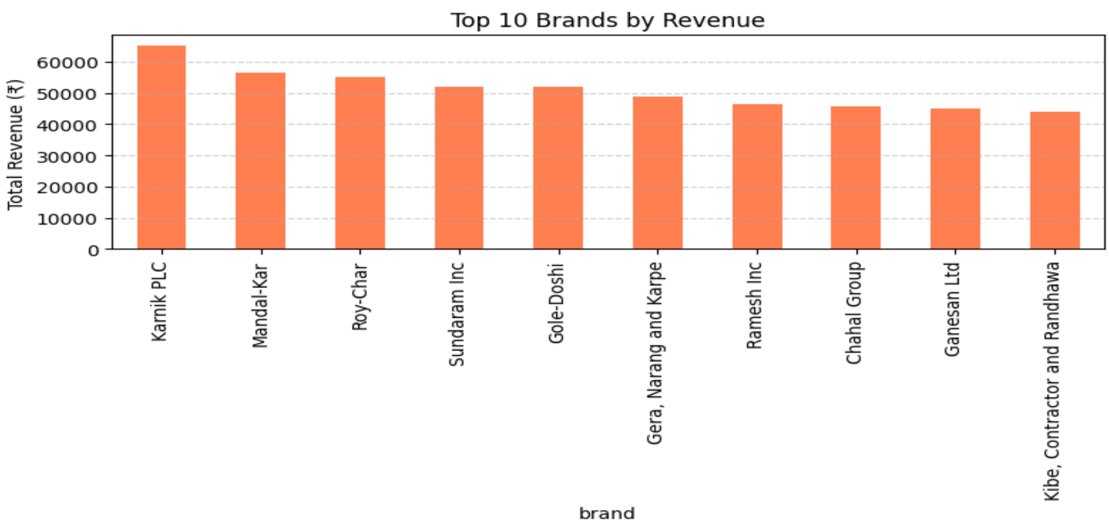

Figure 6: Top Performing brands by revenue

Results & Findings:

- **Leading Brand Performance:**
    1. Karnik PLC emerges as the strongest performer, generating revenue of ₹65,000+, making it the top brand in the dataset.

- **Mid-Tier Brands**:
    1. Brands such as Mandal-Kar, Roy-Char, Sundaram Inc, and Gole-Doshi fall within a tight and highly competitive band of ₹46,000–₹50,000.
    2. The similarity in performance indicates a competitive segment, where multiple brands hold comparable market share.

5.1.3) Weekly Sales Trend:

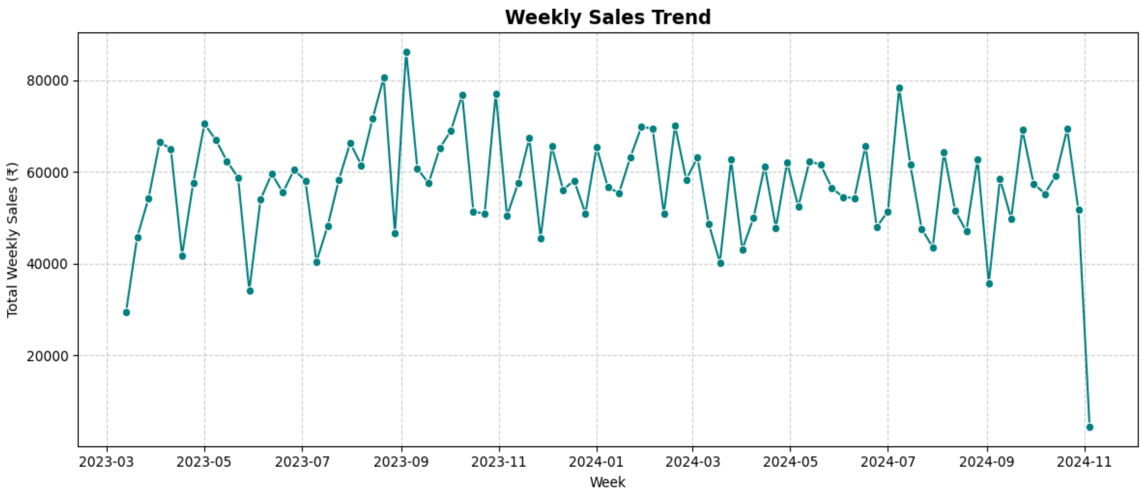

Figure 7: Weekly Sales Trend Analysis

Results & Findings:

- The weekly pattern shows noticeable ups and downs, reflecting how customer demand shifts frequently over short intervals. This kind of movement is expected in fast-paced retail environments.
- Toward the end of the timeline, the chart shows a sharp drop. This is likely because the final week is only partially recorded, although a real decline cannot be ruled out.
- Compared to monthly analysis, the weekly view offers a much clearer picture of short-term changes. It helps uncover brief spikes in demand as well as quick downturns that might be missed in broader monthly trends.

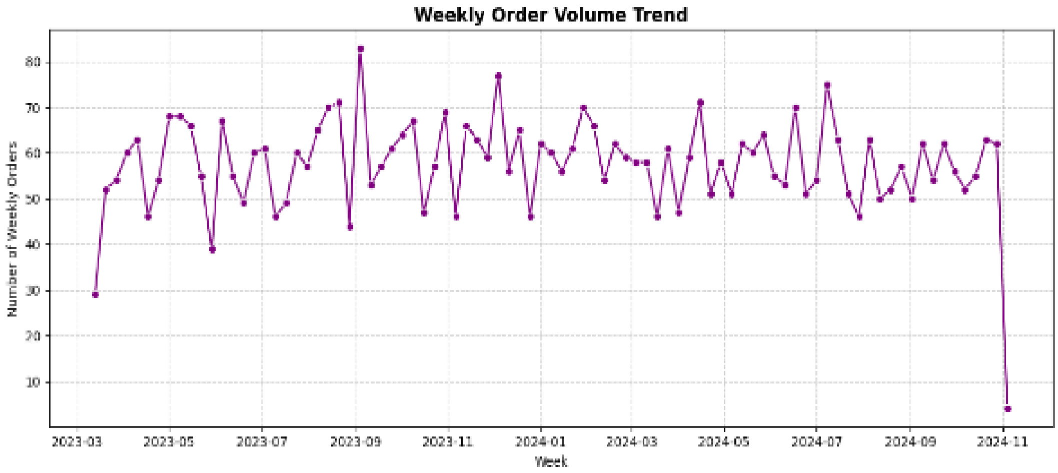

Figure 8: Weekly Order Volume Trend

Results & Findings:

- Weekly orders generally move within a band of 40 to 80 orders, showing a moderate level of variation from week to week.
- A few clear spikes appear in the upper range (above 70–80 orders). These surges are typically associated with occasions such as festivals, local events, or short promotional campaigns that temporarily boost customer activity.
- At the lower end, some weeks fall below 40 orders. These dips could indicate quieter periods in the sales cycle or could be linked to operational issues such as supply shortages, delays, or reduced marketing activity.

5.1.4) Customer Segmentation Analysis:

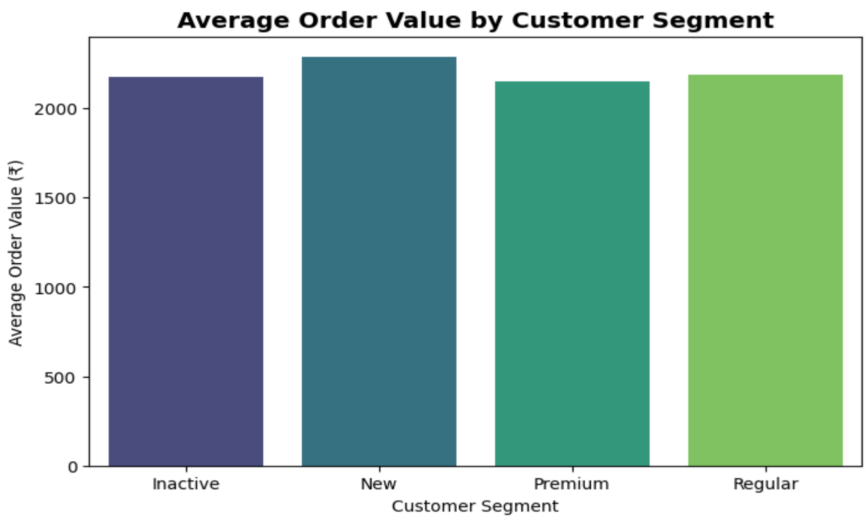

Figure 9: Average Order Value (AOV) by customer segment

Results & Findings:

- **High-Value Customer Segments:**
  1. *New Customers* record the highest average order value (around ₹2,300). This suggests that first-time buyers tend to make larger purchases, possibly influenced by introductory offers or an initial stock-up behaviour.
  2. *Inactive Customers* also display a relatively high AOV (about ₹2,180). Although they order less frequently, the data indicates that when they do return, their baskets are noticeably larger.
- Across all segments, the difference in AOV is not very wide. This points to consistent pricing and similar purchasing patterns overall. However, the stronger spending from New Customers makes them a promising segment for early-stage revenue growth and targeted activation campaigns.

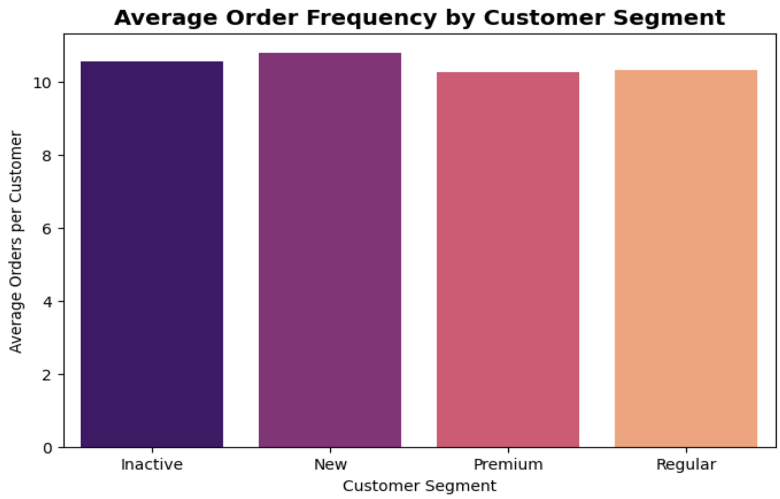

Figure 10: Average Order Frequency (AOF) by customer segment

Results & Findings:

- **Most Frequent Buyers:**
  1. *New Customers* show the highest average order frequency (around 11 orders), indicating that once they join, they tend to stay active and explore the platform more regularly.
  2. *Inactive Customers* are close behind (about 10.7 orders). Although they are not consistent buyers, their behaviour shows that once they return, they tend to make multiple purchases in a short span
- Contrary to expectations, *Premium* customers do not place significantly more orders than *Regular* customers. This suggests that current loyalty or premium benefits may not be compelling enough to drive higher purchase frequency

5.1.5) Damaged Stock Analysis:

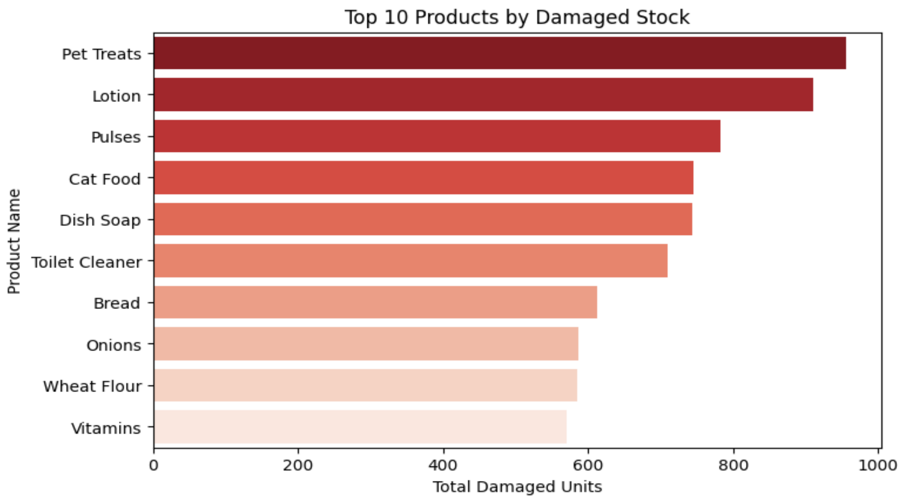

Figure 11: Top products by damaged stocks

Results & Findings:

- Products more prone to damage:
  1. *Pet Treats* show the highest number of damaged units. This may point to issues with how these items are packed, stored, or transported, especially if the packaging is light or easily crushed.
  2. *Lotion* also appears frequently in the damaged stock list. Liquid items are generally more vulnerable because of bottle leakage, fragile caps, or sensitivity to heat and pressure during transit.
- The damage pattern is mostly seen across pet products, liquid items, and bulk food staples. This indicates potential weaknesses in the current supply chain flow.

5.2) Results and Findings of Time Series Forecasting (Predictive Modelling)

5.2.1) ARIMA Model:

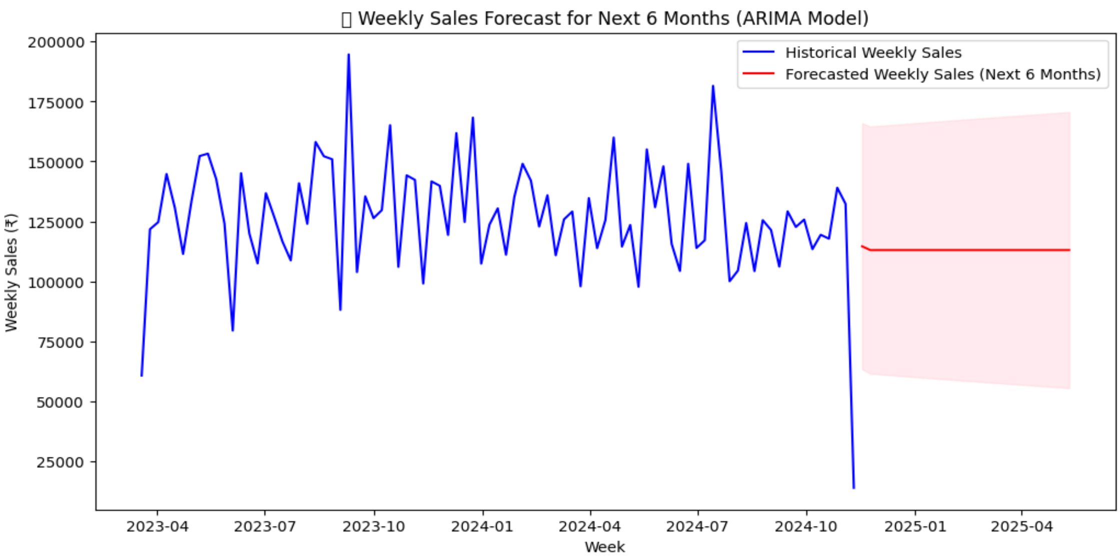

Figure 12: Weekly Sales Forecast using ARIMA

Results & Findings:

- **Historical Trend (Blue Line)**
  1. The weekly sales pattern shows noticeable ups and downs, reflecting the kind of short-term fluctuations typical in grocery delivery—often driven by promotions, holidays, or sudden demand surges.
  2. Even with this movement, most weeks fall somewhere between ₹100,000 and ₹150,000, suggesting a relatively stable demand level without a clear long-term rise or fall.
  3. A few sharp peaks appear close to ₹200,000, which may correspond to special campaigns or festival periods where orders usually spike.
  4. Toward the later part of the timeline (around late 2024), the data shows a mild downturn before the forecasting period begins.

- **Forecasted Period (Red Line + Pink Band)**
  1. The ARIMA model projects weekly sales to remain steady, mostly hovering around ₹115,000 to ₹120,000
  2. The confidence interval gradually widens as the forecast extends further, which is expected as uncertainty increases with time.
  3. Since the model produces a nearly flat forecast, it suggests that **ARIMA does not detect strong upward or downward momentum in recent sales patterns and expects demand to stay close to its historical average.**

5.2.2) Exponential Smoothing Model:

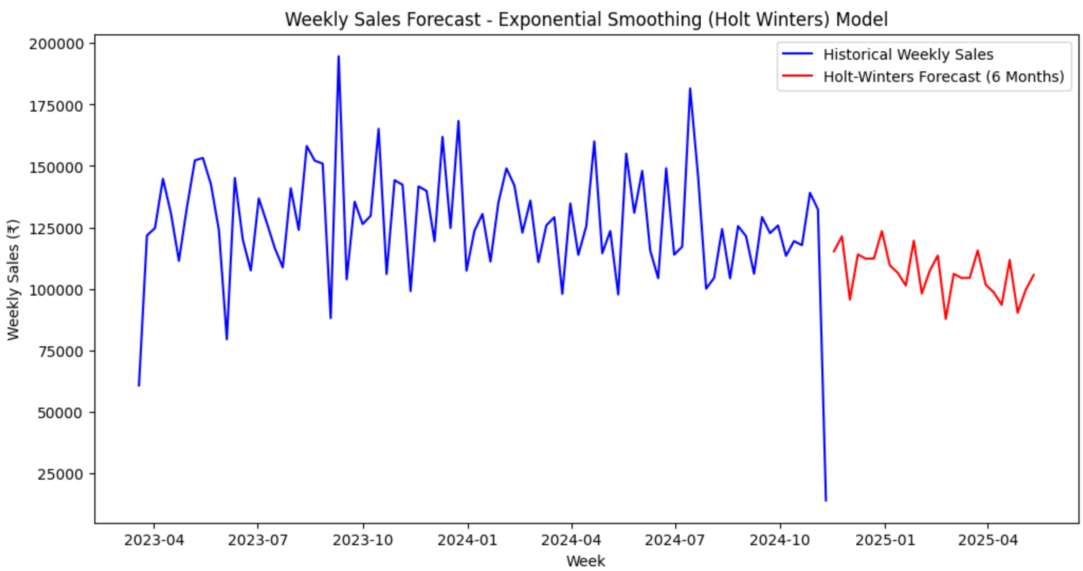

Figure 13: Weekly Sales Forecast using Exponential Smoothing (Holt Winters) Model

Results & Findings:

❖ **Historical Trend (Blue Line)**

1. The weekly sales pattern shows noticeable ups and downs, which is common in grocery delivery where short promotions, festive weeks, or sudden shifts in customer needs influence order volume.
2. For most of the timeline, weekly sales stay in the range of ₹110K to ₹140K, indicating a fairly steady underlying demand level.
3. From the middle of 2024 onward, the data begins to show a gentle downward movement, which the model naturally incorporates into its forecast.

❖ **Forecasted Period (Red Line)**

1. The forecast displays more natural week-to-week variations compared to the flatter ARIMA output, giving a more realistic representation of short-term sales behaviour.
2. The projection indicates a slight downward drift over the next six months, which may reflect a seasonal cooling-off period or simply the model responding to weaker recent sales levels.
3. Most predicted values fall between ₹90K and ₹120K per week, suggesting that sales may stabilize at a slightly lower level than earlier months.

Conclusion:

The Holt-Winters method aligns better with the sales dynamics of a grocery platform. Its ability to react to gradual changes and capture recurring patterns makes its forecasts more believable and useful for demand planning.

6) Interpretation of the results:

1. Market Stability & Customer Behavior:

Analysis of daily and weekly sales shows that Blinkit operates in a relatively stable environment. Order volumes remain fairly consistent, with no major seasonal spikes or dips. This indicates that customers rely on the platform regularly rather than only during specific seasons or events.

2. Demand Concentration and Product Trends:

A detailed look at product-level performance highlights that a few categories such as Household items, Pet Care products, and Health essentials contribute a large share of total sales. These are everyday needs, which explains why customers purchase them frequently and consistently.

3. Insights from Forecasting Models:

- The **ARIMA model** projects a mostly stable demand pattern, with no significant upward or downward shifts.
- The **Holt-Winters model** shows a slight softening in demand, possibly indicating mild saturation in the current customer base or seasonal cooling.

4. Inventory and Operational Observations:

Examining inventory-related metrics like damaged stock, stock sufficiency, and available stock revealed actionable patterns:

- Certain items exhibit higher damage counts, suggesting the need for improved handling or packaging.
- Stock sufficiency varies across products; some SKUs are overstocked while others run low during peak demand periods.

5. Recommendations for Business Decisions

Based on the patterns observed across customers, sales behaviour, and forecasted demand, the following actions are advisable:

a. Customer Engagement and Marketing

- Run targeted promotions during weeks with historically slower sales.
- Provide personalized recommendations based on past purchase behavior.
- Strengthen loyalty programs to encourage repeat orders.

b. Product and Inventory Strategy

- Prioritize high-selling and high-margin products.
- Reduce orders for slow-moving SKUs to free up storage space..
- Bundle frequently purchased items to increase average order value.

c. Forecast-Based Planning

- Use weekly forecasts to refine procurement schedules
- Adjust warehouse stocking levels to align supply with expected demand.

d. Continuous Monitoring

- Re-train forecasting models periodically to reflect evolving customer behavior and marketing effects.