

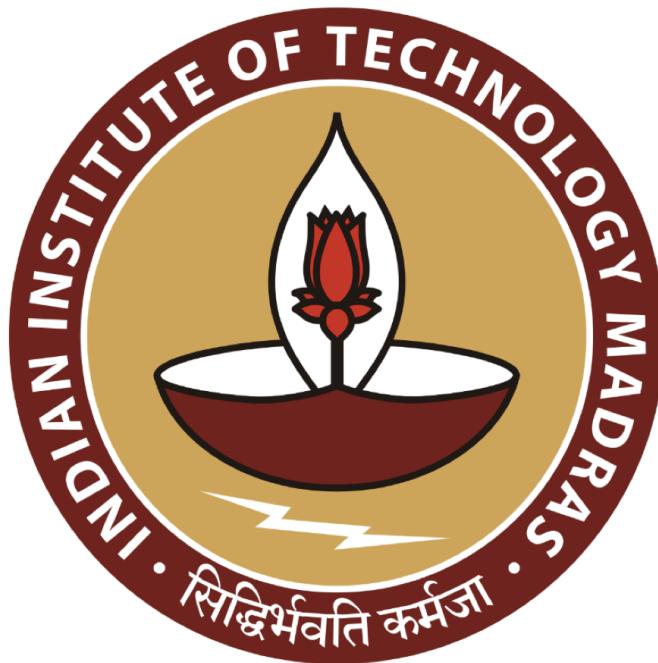
Optimizing Retail Operations at VeggieBell Exotics through Data-Driven Insights

A Mid-term report for the BDM capstone Project

Submitted by

Name: Soni Dhriti Harshadkumar

Roll number: 22f3001203



IITM Online BS Degree Program,
Indian Institute of Technology, Madras, Chennai
Tamil Nadu, India, 600036

Executive Summary:

VeggieBell is a homegrown retail outlet based in Ahmedabad that specializes in fresh, exotic, and locally sourced vegetables. In addition to individual produce sales, it offers pre-prepared salad bowls and monthly vegetable subscription boxes, catering to health-conscious urban consumers. With a loyal customer base and minimal reliance on external marketing, VeggieBell has grown steadily through quality service and community engagement. However, as operations expand, the store is beginning to face challenges related to team coordination and inventory mismanagement.

The objective of this project is to analyze VeggieBell's transactional sales data to derive insights that can inform both strategic and operational decisions. Key problem areas include identifying peak sales periods (by month and weekday), understanding how discounting influences final purchase value, and evaluating customer payment preferences. In addition, we aim to uncover inefficiencies and patterns that might reflect underlying inventory or staffing issues, such as irregular sales spikes or mismatched discount timing.

To address these issues, we applied exploratory data analysis (EDA) using Python libraries including pandas, seaborn, and matplotlib. The analysis includes monthly and weekly sales trends, a deep dive into the relationship between discounts and final sales, and visualizations of payment method usage. The expected outcome is a practical set of recommendations to improve pricing strategy, streamline operations, and better align inventory with demand cycles — helping VeggieBell transition from intuition-based to insight-driven decision-making.

Proof of Originality:

Details :

Business Name : VeggieBell Exotics

Type of Organisation : B2C (Business-to-Consumer)

Shop's Address : VeggieBell Exotics, Shop 1, SKY CITY ARCADE, B Block, Club O7 Rd, Shela, Gujarat 380054

Business Owner's Name : Ms. Kesha Solanki

Business website: www.veggiebell.in

Working Time : 09:00 AM to 8:00 PM

Interview link:  video1137773424.mp4

Letterhead from Veggiebell Exotics



Some photos from the visit and meeting:



Metadata

Below is the metadata constructed based on the dataset shared:

Field Name	Data Type	Description
Sr.No	Integer	Serial number of the transaction
Sales Date	Date	Date of transaction
Gross Amount	Float	Gross billing amount
Tax Amount	Float	Total tax charged on the bill
CGST	Float	Central GST applicable
SGST	Float	State GST applicable
IGST	Float	Integrated GST for interstate sales
5%	Float	Amount taxed at 5% slab
12%	Float	Amount taxed at 12% slab
18%	Float	Amount taxed at 18% slab

28%	Float	Amount taxed at 28% slab
CESS	Float	Additional tax applicable on some goods
Discount	Float	Manual discount given to the customer
Bank	Float	Payment received via Bank Transfer
Cash	Float	Payment received via Cash
Credit Note	Float	Store credit or credit note applied
Coupon Discount	Float	Discount received via promo/coupon
Additional Charge	Float	Any packaging or service charges
Total	Float	Final total payable by the customer after all charges

Descriptive Statistics

- **Total Sales:** ₹731,614
- **Average Daily Sales:** ₹4,782
- **Max Sale Day:** 2025-02-13 (₹290,350)
- **Min Sale Day:** 2025-05-22 (₹21)
- **Cash Payment Share:** ~82%
- **Bank Payment Share:** ~18%
- **Highest Monthly Sales:** February
- **Highest Average Sales Day:** Thursday

Dataset link: [Daily sales summary 1.1.25 to 20.6.25.xlsx](#)

Analysis Process and Method Justification

The analysis adopts an **Exploratory Data Analysis (EDA)** approach to understand the sales dynamics of VeggieBell, a retail brand offering fresh vegetables, salad bowls, and subscription services. The EDA methodology is appropriate for this midterm stage, where the primary goal is to uncover trends, outliers, and hidden patterns in the available transactional data. The dataset includes fields such as sales dates, final billed amount, discounts, and payment mode (cash or bank transfer).

Methods Used:

1. Data Preparation:

- Columns were renamed and standardized.
- Missing values and inconsistencies in discount and tax fields were resolved.
- Datatypes were converted (e.g., sales date to datetime) for time-based grouping.

2. Temporal Sales Analysis:

We grouped data by **month** and **weekday** to identify seasonal and weekly sales behaviors. Bar plots revealed that:

- **February** had the highest total sales. This aligns with qualitative inputs from the founder that strong New Year resolution campaigns in January likely influenced customer habits into February.

- **Thursday** emerged as the weekday with the highest average sales, indicating possible consumer behavior patterns, such as bulk purchases before weekends or mid-week restocking needs.

These time-based trends were visualized using `pandas` for grouping and `matplotlib` for clear, labeled bar charts.

3. Payment Behavior Insights:

We aggregated sales based on payment methods (Cash and Bank). A pie chart showed that:

- **82% of sales occurred via Cash**, while **18% were through Bank transfers**. This suggests that despite growing digital adoption in general markets, VeggieBell's customer base still heavily relies on cash, possibly due to doorstep delivery culture or tech accessibility gaps.

4. Discount Effectiveness on Sales:

Two levels of analysis were done here:

- At the **transaction level**, we used a scatter plot of individual discounts vs. final sales amounts. This indicated a **non-linear** relationship, where even small discounts could accompany high sales, but with significant variability.
- To enhance visibility and reduce skewness due to outliers, we plotted **daily aggregated discount vs. daily sales** with a **logarithmic scale on the Y-axis (sales)**. This helped manage wide variations—ranging from under ₹100 to over ₹100,000—allowing us to uncover subtler trends.

Key observations:

- **A mild positive correlation** was noted: higher daily discounts were often associated with higher daily sales, but not consistently.
- **Concentration at low discount levels** suggests discounts are most often small, likely used as nudges rather than aggressive marketing.
- **Outlier days** with high discounts and massive sales were identified (e.g., sales of ₹290,350 with ~₹1100 discount), potentially pointing to successful promotional campaigns or bulk order days.

5. Summary Statistics and Performance Metrics:

We computed important KPIs like:

- **Total Sales:** ₹731,614

- **Average Daily Sales:** ₹4,782
- **Max Sale Day:** Feb 13, ₹290,350
- **Min Sale Day:** May 22, ₹21

These statistics establish the magnitude and variability of operations, offering a baseline for planning targets and identifying underperforming days.

Tools Used:

- **pandas:** Data cleaning, transformation, and grouping
- **matplotlib:** Standard plots and charts for visual clarity
- **seaborn:** Advanced plots like scatter plots with regression and distribution insights

Justification of Methods

Exploratory Data Analysis (EDA) was the most suitable approach for this phase because it helped uncover patterns in sales without needing complex models. Using EDA, we identified:

- **Monthly and weekday trends**, such as February having the highest sales and Thursday showing peak average sales.
- **Customer payment behavior**, where 82% of transactions were in cash and 18% via bank transfers.
- **Impact of discounts**, where higher daily discounts often aligned with higher daily sales, though not always proportionally.
- **Sales outliers**, such as one day with over ₹290,000 in sales and around ₹1100 in discount, visible clearly using logarithmic scaling.

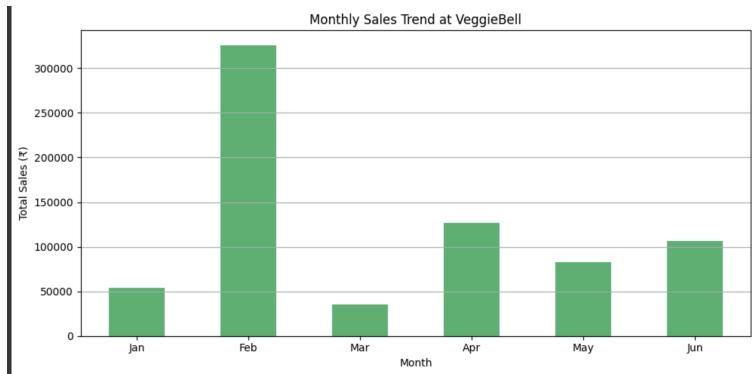
Visualizations helped clearly communicate these insights, even to non-technical stakeholders. Overall, EDA gave us a strong understanding of customer behavior, discount strategy impact, and sales timing, forming a foundation for future improvements in operations and marketing.

Results and Findings:

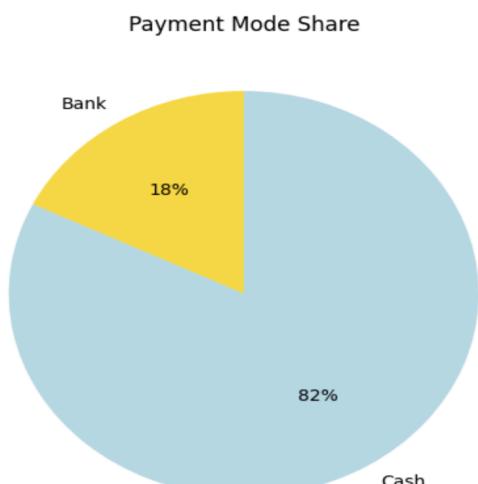
- **Monthly Sales Trend:** February leads in total sales, likely due to targeted marketing around New Year resolutions.
- **Weekday Sales:** Thursdays show the highest average sales, indicating higher mid-week customer footfall.
- **Payment Modes:** Cash dominates (82%), but digital payments via bank transfers (18%) are significant, showing some shift towards cashless transactions.

- **Discount Impact:** Scatter plot analysis suggests a **positive but weak correlation** between daily total discounts and sales. Higher discounts generally coincide with higher sales, but sales vary widely at each discount level, indicating other influencing factors.
- **Sales Variability:** Even zero-discount days sometimes record high sales, meaning discounts are not the sole driver.
- **Diminishing Returns:** Increasing discounts beyond moderate levels does not guarantee proportionate sales growth, highlighting the need for balanced discounting.

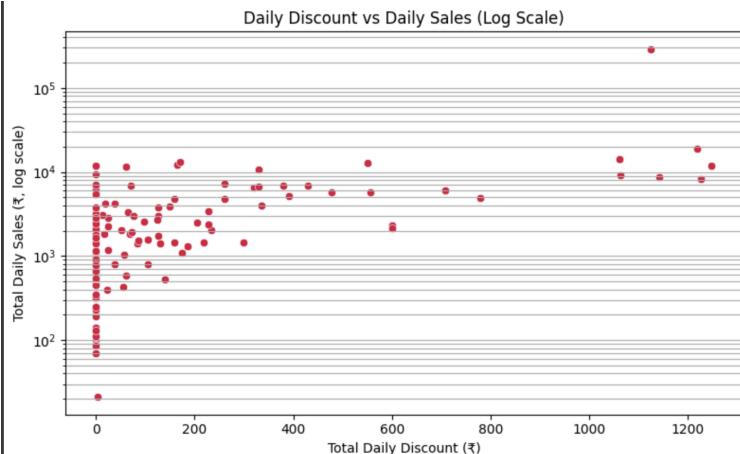
Some Visualizations and their interpretation:



This shows the monthly sales of VeggieBell between January and June. February was observed to have the highest sales. Upon interviewing the founder, it was observed that this was due to heavy social media marketing during January over new year resolutions. The summer season also helped in selling organic mangoes.



The pie chart clearly shows that cash is the dominant payment method, representing 82% of transactions, while bank payments make up only 18%. This significant reliance on cash suggests a need for the store to embrace modernization. Shifting towards digital payment options would enhance customer convenience, improve operational efficiency, and reduce the risks associated with cash handling.



Based on the scatter plot, higher daily discounts generally correlate with increased daily sales, particularly evident in significant sales spikes at very high discount levels. However, sales vary widely for any given discount, indicating other influential factors are at play. To grow the business, consider strategically leveraging targeted high-discount campaigns, while simultaneously

analyzing other drivers of sales to optimize revenue across all discount ranges.

Limitations and Future Scope

While this midterm analysis provides valuable insights into sales trends, discount effectiveness, and payment behaviors, it is important to recognize certain limitations that constrain the depth of conclusions.

Limitations:

- **Lack of Itemized Data:** The current dataset aggregates sales at the transaction level without product-wise breakdowns. As a result, we are unable to assess demand for individual vegetables, salad bowls, or subscription products, limiting the granularity of inventory insights.
- **Absence of Customer Identifiers:** The data lacks unique customer IDs or contact details, preventing Recency-Frequency-Monetary (RFM) analysis and customer segmentation, which are essential for understanding buying behavior and loyalty.
- **No Delivery or Logistics Records:** Since delivery logs, timestamps, or route-level data are not included, evaluating delivery efficiency or logistics performance is not currently feasible.
- **Team and Operational Data Missing:** Although inefficiencies in team management and inventory processes were noted qualitatively, the dataset does not include time-stamped order processing or inventory movement logs that would allow quantifiable analysis.

Future Scope:

- **Inventory-Level Forecasting:** By integrating item-wise sales data and tracking per-SKU stock movement, predictive models can be developed to optimize restocking and reduce waste—especially for perishable goods.
- **Customer Segmentation and RFM Modeling:** Including customer identifiers will enable personalized marketing, loyalty tracking, and segmentation using clustering techniques.
- **Logistics Optimization:** Incorporating delivery and order fulfillment data would allow for evaluation of delays, efficiency of delivery routes, and driver performance metrics.
- **Operational Analytics:** Future data collection can include timestamps for order preparation and staff performance to address inefficiencies in team workflows.

In summary, expanding the data architecture will not only enhance analytical depth but also empower VeggieBell to transition from reactive decisions to a predictive and proactive model.