VistaCart Retail Sales Data Analysis Report

Sales Performance, Product Insights & Outlet Analysis

Executive Summary

This project presents an end-to-end analysis of retail sales data for VistaCart, a retail company, with the objective of uncovering actionable insights to support data-driven business decisions.

Using SQL for data cleaning and analysis, Excel for supplementary analysis, and Power BI for interactive dashboards, the project evaluates overall sales performance, product trends, outlet effectiveness, and customer behavior. Key performance indicators such as total sales, average sales, item counts, and customer ratings were analyzed to identify top-performing outlets and products, uncover revenue drivers, and highlight areas for operational improvement.

The findings provide practical recommendations for inventory optimization, outlet performance management, and strategic planning, demonstrating how structured data analysis can directly support business outcomes.

Business Objectives

The analysis was guided by the following business questions:

- What is the overall sales performance of VistaCart?

- Which outlets and outlet types generate the highest revenue?

- Which product categories contribute most to total sales?

- How do outlet size, location, and establishment year influence performance?

- Is there a relationship between item visibility, ratings, and sales?

- Which products and outlets should be prioritized for strategic focus?

Dataset Overview

The dataset represents simulated retail operations for VistaCart and includes information on:

- Product attributes (item type, fat content, weight, visibility)

- Outlet characteristics (type, size, location tier, establishment year)

- Sales metrics (total sales per item)

- Customer ratings

After data cleaning, the dataset contained 7,060 records, representing transactions across multiple outlets and product categories.

Data Cleaning & Preparation

Data quality was ensured before analysis through the following steps:

- Identification and handling of missing values (e.g., item weight, ratings)

- Removal of duplicate records to prevent skewed results

- Standardization of categorical variables (e.g., consolidating variations of Low Fat and Regular)

- Trimming inconsistent text values

- Validation of numerical fields such as sales and ratings

All data cleaning steps were implemented using SQL and documented in a dedicated data cleaning script.
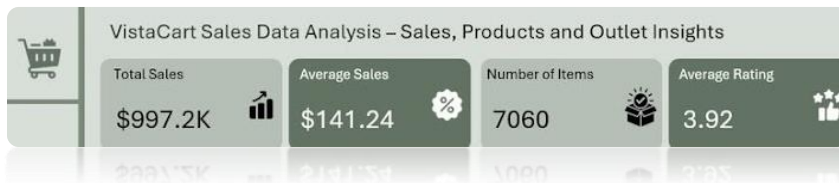
```
9    -----------------------------------------
10   -- DATA CLEANING
11   -----------------------------------------
12
13   -- 1. CHECK FOR MISSING VALUES
14   SELECT
15       SUM(CASE WHEN Item_Weight IS NULL THEN 1 ELSE 0 END) AS Missing_Item_Weight,
16       SUM(CASE WHEN Item_Fat_Content IS NULL THEN 1 ELSE 0 END) AS Missing_Item_Fat_Content,
17       SUM(CASE WHEN Item_Type IS NULL THEN 1 ELSE 0 END) AS Missing_Item_Type,
18       SUM(CASE WHEN Outlet_Size IS NULL THEN 1 ELSE 0 END) AS Missing_Outlet_Size,
19       SUM(CASE WHEN Item_Visibility IS NULL THEN 1 ELSE 0 END) AS Missing_Item_Visibility,
20       SUM(CASE WHEN Total_Sales IS NULL THEN 1 ELSE 0 END) AS Missing_Total_Sales,
21       SUM(CASE WHEN Rating IS NULL THEN 1 ELSE 0 END) AS Missing_Rating
22   FROM vistacart;
23
24   -- Remove rows with missing Item_Weight
25   DELETE FROM vistacart
26   WHERE Item_Weight IS NULL;
27
28   -----------------------------------------
29   -- 2. STANDARDIZE CATEGORICAL VALUES (Item_Fat_Content)
30   -----------------------------------------
31
32   UPDATE vistacart
33   SET Item_Fat_Content = 'Low Fat'
34   WHERE Item_Fat_Content IN ('LF', 'low fat', 'Low Fat');
35
36   UPDATE vistacart
37   SET Item_Fat_Content = 'Regular'
38   WHERE Item_Fat_Content = 'reg';
```

```
40   -----------------------------------------
41   -- 3. REMOVE DUPLICATE RECORDS
     -----------------------------------------

     WITH CTE AS (
         SELECT *,
             ROW_NUMBER() OVER (
                 PARTITION BY Item_Identifier,
                             Outlet_Identifier,
                             Item_Type,
                             Total_Sales
                 ORDER BY Item_Identifier
             ) AS rn
         FROM vistacart
     )
     DELETE FROM CTE
     WHERE rn > 1;
     -- 0 ROWS AFFECTED, NO DUPLICATES

     -----------------------------------------
     -- 4. TRIM EXTRA SPACES
     -----------------------------------------

     UPDATE vistacart
     SET
         Item_Type = LTRIM(RTRIM(Item_Type)),
         Outlet_Location_Type = LTRIM(RTRIM(Outlet_Location_Type)),
         Outlet_Size = LTRIM(RTRIM(Outlet_Size)),
         Outlet_Type = LTRIM(RTRIM(Outlet_Type));
```

Key Performance Indicators (KPIs)

The following KPIs were calculated to assess overall business performance:

- **Total Sales:** $997.2K

- **Average Sales per Item:** $141.24

- **Number of Items Sold:** 7,060

- **Average Customer Rating:** 3.92

These metrics provide a high-level snapshot of VistaCart's performance and form the foundation for deeper analysis.

KPI summary from Power BI dashboard

Sales & Outlet Analysis

Sales by Outlet Type and Size

- Medium and large outlets generated higher total sales compared to smaller outlets.

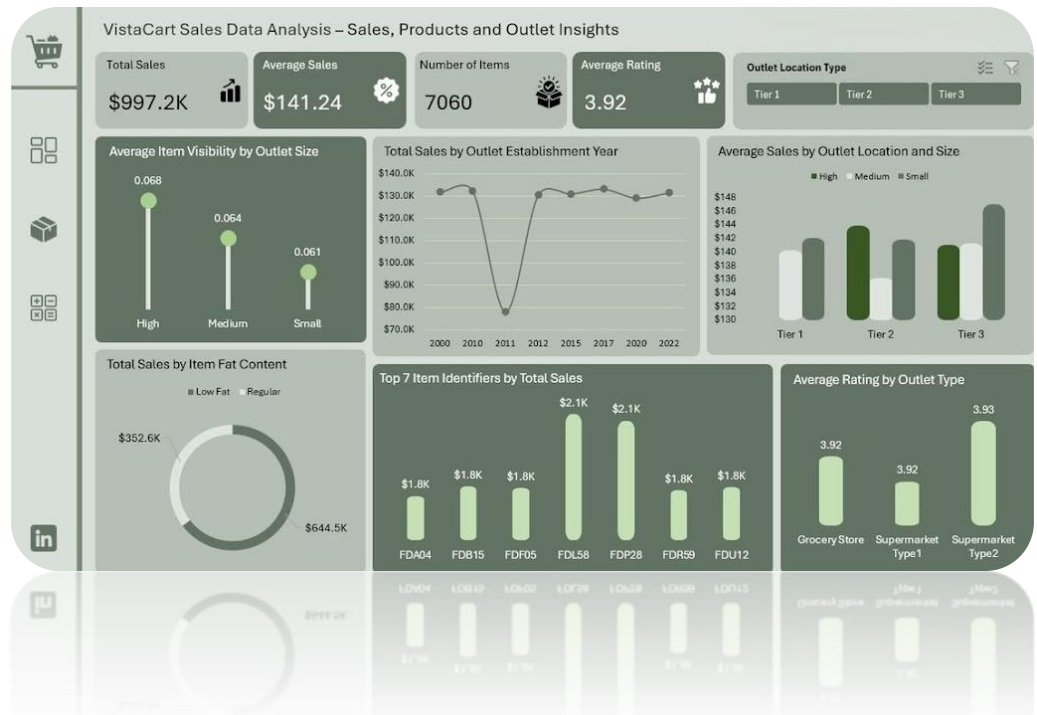- Supermarket-type outlets outperformed grocery stores in revenue contribution.

Sales by Outlet Location Tier

- Tier 2 locations recorded the highest overall sales, followed by Tier 3.

- Tier 1 outlets showed lower revenue, indicating potential location-based demand differences.

Outlet Establishment Year

- Older outlets generally performed more consistently, suggesting operational maturity contributes to revenue stability.

Power BI visuals showing sales by outlet type, size, location, and establishment year

Product & Category Insights

Top-Performing Product Categories

- Fruits & Vegetables and Snack Foods emerged as the highest revenue-generating categories.

- Certain categories contributed disproportionately to total sales, highlighting opportunities for focused inventory planning.

Item Fat Content Analysis

- Regular fat products generated a larger share of revenue compared to low-fat alternatives.

- This insight can support targeted marketing and assortment decisions.

Visibility, Ratings & Sales Relationship

- Items with moderate to high visibility generally achieved higher average sales.

- Higher customer ratings correlated with improved sales performance, reinforcing the importance of product quality and customer satisfaction.

- These relationships suggest that improving product placement and maintaining quality standards can directly influence revenue.

Average Item Visibility by Outlet Size

High 0.068, Medium 0.064, Small 0.061

Top 7 Item Identifiers by Total Sales

FDA04 $1.8K, FDB15 $1.8K, FDF05 $1.8K, FDL58 $2.1K, FDP28 $2.1K, FDR59 $1.8K, FDU12 $1.8K

Average Rating by Outlet Type

Grocery Store 3.92, Supermarket Type1 3.92, Supermarket Type2 3.93

Tools & Technologies Used

- **SQL (SQL Server)** – Data cleaning, transformation, and analysis

- **Excel** – Supporting analysis and validation

- **Power BI** – Interactive dashboards and visual storytelling

Key Business Insights & Recommendations

- Focus inventory and promotions on high-performing product categories.

- Prioritize medium and large outlets for expansion and optimization.

- Re-evaluate underperforming locations and outlet types.

- Improve product visibility to enhance sales performance.

- Use customer ratings as a quality signal for decision-making.

Conclusion

This project demonstrates the value of structured data analysis in understanding retail performance and supporting strategic decisions. By combining SQL-based data preparation, analytical querying, and interactive dashboards, the analysis delivers clear insights into sales drivers, customer preferences, and outlet effectiveness.
The approach highlights practical, real-world analytics skills and showcases how data can be transformed into meaningful business value.

by Phemelo Bohlale Sebopelo