

# **Grocery Store Database Analysis Project Report**

**By**

**Selemani Mziray**

**Submitted on:**

October 29, 2025

**Sponsored by:**

**Data Science Infinity (DSI)**

**Submitted in Partial Fulfillment of:**

Certificate in Data Science Analytics

## **Executive Summary**

The grocery store maintains a comprehensive transaction database covering both food and non-food products purchased by customers. As the data analyst, the primary responsibility is to analyze this data to identify which product areas are performing well or poorly and to uncover spending patterns among top customers. This analysis will provide actionable insights to guide inventory management, pricing strategies, and targeted promotions.

The assignment begins with exploring the database structure by reviewing the first five transactions and counting the total number of records to confirm scope and integrity. Using SQL, key metrics such as minimum, average, and maximum sales cost per product area are calculated, along with total sales by product area and overall store level. Additional queries summarize sales by customer and product category, identify customers with the highest spending in each category, and rank top customers by total sales.

To communicate findings effectively, visualizations are created using Excel and smart (GPT-5), including charts and histograms that show sales by product area versus customer ID, overall product area performance, and top five customers. Comparative visuals combine histograms and pie charts for two scenarios: total sales per customer across all product areas and sales per customer by specific product categories. In addition, Python-AI was used to create an interactive dashboard for better visualization from the same dashboard.

Analysis of these scenarios reveals that non-food products dominate overall sales, consistently contributing to the largest share of revenue. Among food categories, meat products lead, followed by fruits, dairy, and vegetables. Customer spending patterns vary significantly with indication that customer ID 224 favors non-food items; customer ID 593 leads in meat purchases, and customer ID 150 ranks highest in vegetable sales. These insights highlight opportunities to strengthen performance in lower-contributing categories such as dairy and vegetables through targeted promotions.

The outcome of this analysis is a clear, data driven view of product performance and customer behavior, enabling management to make informed decisions that optimize sales and enhance customer satisfaction.

## Table of Contents

<b>1. Introduction</b> .....	4
<b>2. Database Exploration</b>	
2.1. Review the first five transactions .....	5
2.2. Count the total number of transactions.....	5
<b>3. SQL Analysis</b>	
3.1. Determine minimum sales cost for each product area.....	5
3.2. Determine maximum sales cost for each product area.....	5
3.3. Determine average sales costs for each product area.....	6
3.4. Calculate the sum of sales costs for each product area.....	6
3.5. Compute the overall sales cost for all product area .....	7
3.6. Identify five customers with the highest sales cost/product area.....	7
3.7. Calculate the total sales cost for each customer/all product area.....	8
<b>4. Data Visualization (Excel and GPT-5)</b>	
4.1. A graph showing the total sales cost for the top five customers.....	9
4.2. A graph showing top sales cost for each product area.....	10
4.3. Histograms and pie charts for scenarios 4.1 and 4.2 .....	11
4.4. A table with total sales cost from each product area .....	13
4.5. A histogram/pie chart showing total sales cost for each product area .....	14
4.6. Interactive Dashboards Created by Python-AI.....	15
<b>5. Conclusions</b> .....	20
<b>Acknowledgment</b> .....	21

## 1. Introduction

The grocery store operates a comprehensive transaction database that records purchases across both food and non-food product categories. This project focuses on analyzing these records to identify which product areas are performing well or poorly and to uncover spending patterns among top customers. Understanding these trends is critical for optimizing inventory management, pricing strategies, and targeted promotions that enhance overall profitability and customer satisfaction.

The analysis begins by reviewing the database structure and confirming its integrity. This includes examining the first few transactions to understand the layout and counting the total number of records to establish scope. Once the structure is validated, SQL queries are employed to calculate key performance metrics such as minimum, average, and maximum sales cost per product area. Additional computations include total sales by product category and overall store level, as well as aggregated sales by customers. These queries help identify top-spending customers and highlight category-specific preferences, providing a granular view of purchasing behavior.

To communicate findings effectively, visualizations are created using Excel, Python-AI, and smart (GPT-5). These include charts and histograms that illustrate product area performance, customer-level sales distribution, and comparative scenarios. Two key scenarios are analyzed: total sales per customer across all product areas and sales per customer by specific product categories. Insights from these visualizations reveal that non-food products dominate overall revenue, while meat products lead among food categories, followed by fruits, dairy, and vegetables. Customer preferences vary significantly; some prioritize non-food items, while others favor specific food categories. These patterns highlight opportunities to improve performance in lower contributing segments such as dairy and vegetables through targeted strategies.

This project establishes a data driven foundation for optimizing sales performance and enhancing customer satisfaction. By integrating SQL database exploration with advanced visualization tools, it aims to uncover actionable insights that empower informed decision making and strategic business growth. The analysis begins with comprehensive data extraction and exploration using SQL, followed by the creation of dynamic visualizations in Excel and Smart (GPT-5). Together, these tools enable deeper understanding, effective management, and more accurate business forecasting.

### 2. 1. Methodology of Study 1: Review the first five transactions

#This query retrieves the first five rows from the transactions table:

```
SELECT *  
FROM grocery_db.transactions  
ORDER BY transaction_id
```

LIMIT 5;

**Results:**

customer_id	transaction_date	transaction_id	product_area_id	num_items	sales_cost
224	7/22/2020	4.36683E+11	1	27	669.34
27	8/31/2020	4.37086E+11	1	27	639.22
10	9/29/2020	4.37372E+11	1	26	630.88
593	9/4/2020	4.37127E+11	1	24	611.83
731	4/29/2020	4.35847E+11	1	27	600.48

**2.2. Methodology Study 2: Count the total number of transactions**

#Counts all rows in the transactions table to confirm number of rows:

```
SELECT COUNT (*) AS total_transactions
FROM grocery_db.transactions;
```

**Results: Count = 38506**

**3.1. Methodology Study 3: Determine Minimum sales cost for each product area**

# Minimum sales cost per product area:

```
SELECT
product_area_id,
MIN (sales_cost) AS min_sales_cost
FROM grocery_db.transactions
GROUP BY product_area_id
ORDER BY product_area_id
LIMIT 5;
```

**Results:**

product_area_id	min_sales_cost
1	10.02
2	0
3	1.01
4	2
5	5

**3.2. Methodology of Study 4: Determine maximum sales cost for each product area**

#Average sales cost per product area:

```
SELECT
product_area_id,
AVG (sales_cost) AS avg_sales_cost
FROM grocery_db.transactions
GROUP BY product_area_id
ORDER BY product_area_id
LIMIT 5;
```

**Results:**

product_area_id	avg_sales_cost
1	95.98
2	15.8
3	28.97
4	23.88
5	38.92

**3.3. Methodology of Study 5: Determine average sales costs for each product area**

#Maximum sales cost per product area:

```
SELECT  
product_area_id,  
MAX (sales_cost) AS max_sales_cost  
FROM grocery_db.transactions  
GROUP BY product_area_id  
ORDER BY product_area_id  
LIMIT 5;
```

**Results:**

product_area_id	max_sales_cost
1	669.34
2	557.73
3	288.35
4	309.25
5	392.62

**3.4. Methodology of Study 6: Calculate the sum of sales costs for each product area**

#Total sales cost per product area:

```
SELECT  
product_area_id,  
SUM (sales_cost) AS total_sales_cost  
FROM grocery_db.transactions  
GROUP BY product_area_id  
ORDER BY product_area_id;
```

## Results:

product_area_id	total_sales_cost
1	747129.45
2	133879.25
3	252033.53
4	175792.77
5	240892.82

- *The overall total sales cost for all product areas is the summation of sales obtained in each individual sales cost which is equal to \$747129 + \$133879.25 + \$252033.53 + \$175792.77 + \$240892.82 = **\$1549727.82***

### **3.5. Methodology of Study 7: Compute the overall sales cost for all product area**

#Overall total sales cost for all product area:

```
SELECT
SUM (sales_cost) AS overall_total_sales_cost,
AVG (sales_cost) AS avg_sales_cost,
MIN (sales_cost) AS min_sales_cost,
MAX (sales_cost) AS max_sales_cost,
COUNT (*) AS num_rows,
COUNT (distinct transaction_id) as number_transaction
FROM grocery_db.transactions;
```

## Results:

overall_total_sales_cost	avg_sales_cost	min_sales_cost	max_sales_cost	num_rows	num_transaction
1549727.82	40.24639848	0	669.34	38506	18160

### **3.6. Methodology of Study 8: Identify five customers with the highest sales cost/product area**

#Sales cost for each product area versus customer:

```
SELECT customer_id,
SUM (case when product_area_id = 1 THEN sales_cost ELSE 0 end) AS non_food_sales_cost,
SUM (case when product_area_id = 2 THEN sales_cost ELSE 0 end) AS vegetables_sales_cost,
SUM (case when product_area_id = 3 THEN sales_cost ELSE 0 end) AS fruit_sales_cost,
SUM (case when product_area_id = 4 THEN sales_cost ELSE 0 end) AS dairy_sales_cost,
SUM (case when product_area_id = 5 THEN sales_cost ELSE 0 end) AS meat_sales_cost
FROM grocery_db.transactions
GROUP BY customer_id
ORDER BY customer_id
LIMIT 10;
```

## **Results**

customer_id	non_food_sales_cost	vegetable_sales_cost	fruit_sales_cost	dairy_sales_cost	meat_sales_cost
1	2626.25	167.12	276	367.74	543.38
2	1445.09	99.05	120.93	150.63	241.21
3	0	31.5	75.39	46.2	171.13
4	2496.98	142.35	245.36	341.91	272.79
5	3343.87	350.1	815.32	694.71	1405.19
6	2450.22	176.58	328.75	333.94	1011.31
7	1495.43	326.49	349.55	0	0
8	0	14.79	114.5	104.35	121.9
9	0	32.1	146.11	95.4	62.94
10	5183.54	632.57	1116.95	1231.42	1461.54

#This SQL code is copied from objective in section 3.6 to filter the sum of total sales cost when the customer\_id is 224, 593, 193, 150, and 10:

```
SELECT customer_id,  
SUM (case when product_area_id = 1 THEN sales_cost ELSE 0 END) as non_food_sales_cost,  
SUM (case when product_area_id = 2 THEN sales_cost ELSE 0 END) as vegetables_sales_cost,  
SUM (case when product_area_id = 3 THEN sales_cost ELSE 0 END) as fruit_sales_cost,  
SUM (case when product_area_id = 4 THEN sales_cost ELSE 0 END) as dairy_sales_cost,  
SUM (case when product_area_id = 5 THEN sales_cost ELSE 0 END) as meat_sales_cost  
FROM grocery_db.transactions  
WHERE customer_id = 224  
GROUP BY customer_id;
```

## **Results:**

customer_id	non_food_sales_cost	vegetable_sales_cost	fruit_sales_cost	Dairy_sales_cost	meat_sales_cost
224	5656.33	607.88	1016.61	903.03	1694.91
593	4525.52	553.83	1131.91	690.56	2080.37
193	3277.42	403.65	1377.76	829.61	1078.63
150	5049.18	1247.21	973.66	1184.33	1967.94
10	5183.54	632.57	1116.95	1231.42	1461.54

### **3.7. Methodology of Study 9: Calculate the total sales cost for each customer/all product area**

#This SQL code calculate each customer total sales cost of the first 5 customers:

```
SELECT customer_id, sum(sales_cost) AS customer_total_sales_cost  
FROM grocery_db.transactions  
GROUP BY customer_id  
ORDER BY customer_id ILIMIT 5;
```

## **Results:**

Customer_id	customer_total_sales_cost
1	3980.49
2	2056.91
3	324.22
4	3499.39
5	6609.19

#Now we can filter the above SQL query using the customer\_id with the highest sales cost, 224, 150, 193, 593, and 10:

```
#each customer_id is run independently with the SQL code below
SELECT customer_id, sum(sales_cost) as customer_total_sales_cost
FROM grocery_db.transactions
WHERE customer_id = 224
GROUP BY customer_id
ORDER BY customer_id;
```

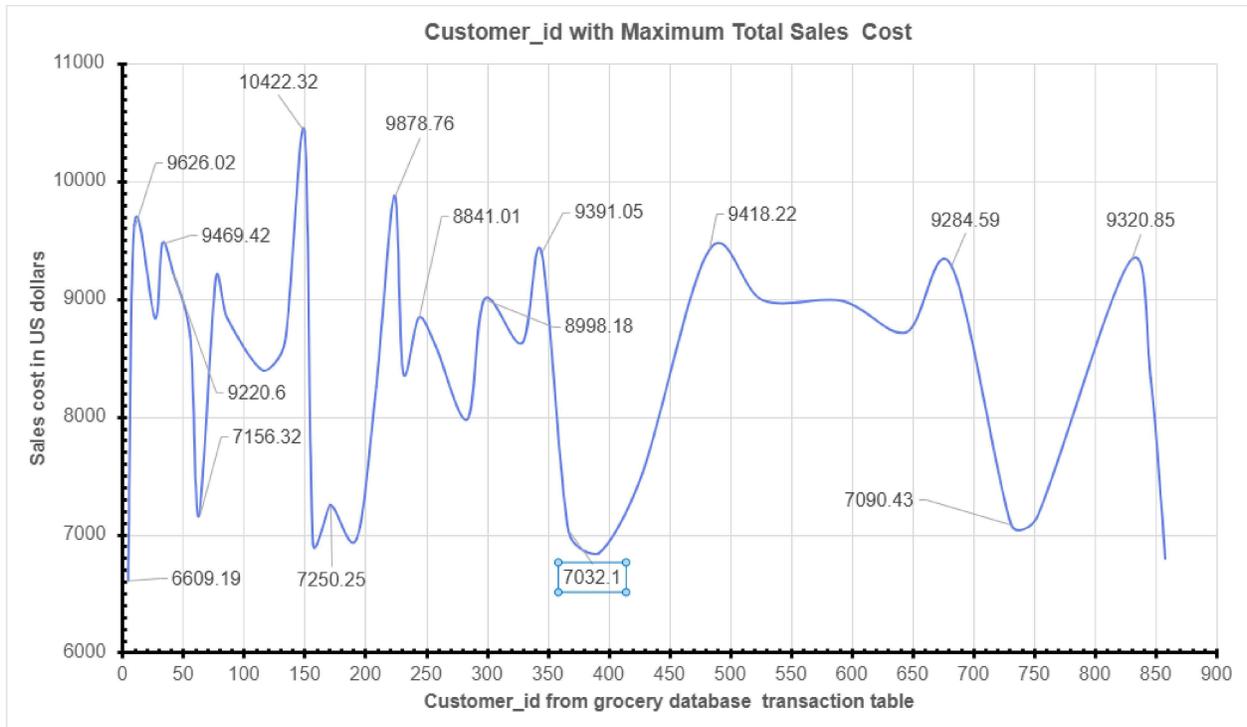
## **Results:**

customer_id	customer_total_sales_cost
224	9878.76
593	8982.19
193	6967.07
150	10422.32
10	9626.02

### **4.1. Methodology of Study 10: A graph showing the total sales cost for the top five customers**

Since the transaction table contains a total of 870 unique customers, we focused our analysis on identifying and visualizing the top five customers based on total sales cost. To place their performance in context, we plotted these five customers against the full customer ranking distribution. Using Microsoft Excel (see **Figure 1**), we generated a graph that highlights how these top performers stand out relative to the broader customer base. The results show that the top five customers are associated with the following customer IDs and total sales costs:

- Customer ID 10 with a total sales cost of \$9,626.02
- Customer ID 593 with a total sales cost of \$8,982.19
- Customer ID 193 with a total sales cost of \$6,967.0
- Customer ID 224 with a total sales cost of \$9,878.76
- Customer ID 150 with a total sales cost of \$10,422.32

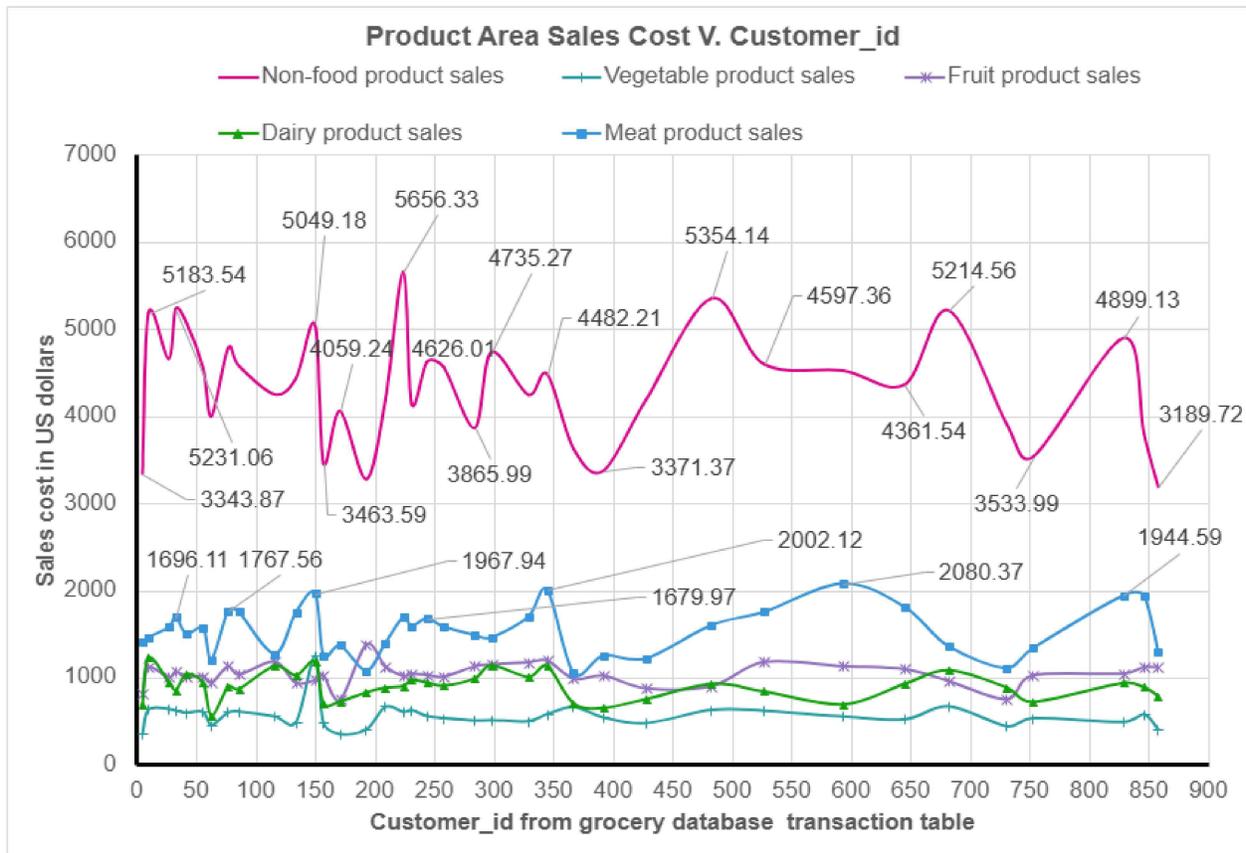


**Figure 1.** The graph displays the total sales cost for the top five customers, plotted along with an X-axis that represents all 870 customers.

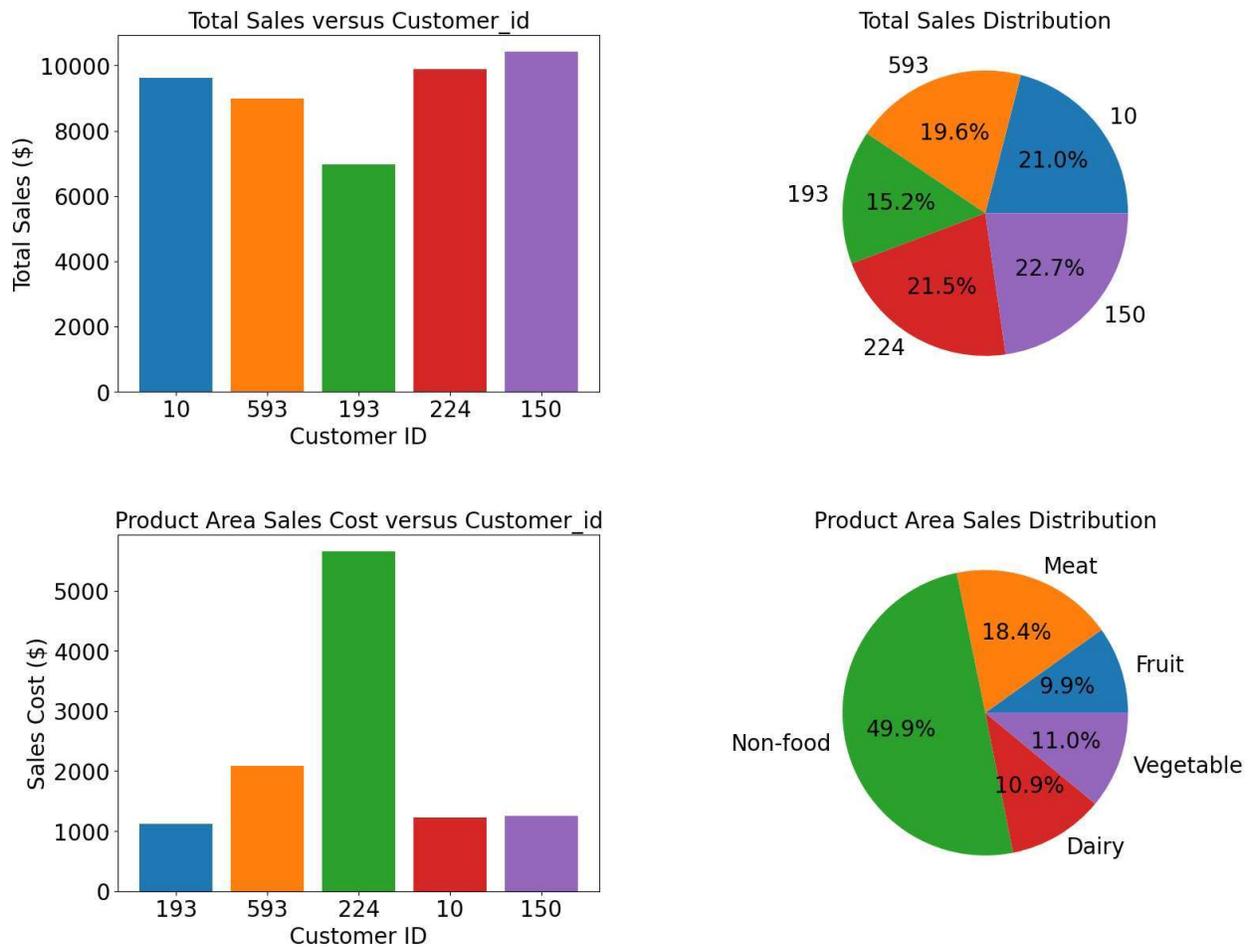
#### 4.2. Methodology of Study 11: A graph showing top sales cost for each product area

The graph in **Figure 2** show that the top selling product area are associated with the following customer IDs and total sales costs:

- Non-food: Customer id 224 with a sales cost of \$5,656.33.
- Meat: Customer 593 id with a sales cost of \$2,080.37.
- Vegetable: Customer id 150 with sales cost of \$1,247.21.
- Dairy: Customer id 10 with a sales cost of \$1,231.42.
- Fruit: Customer id 193 with a sales cost of \$1,116.95.



In **Figure 3** there is a detailed interpretation of what the histograms and pie charts reveal customer spending patterns in both scenarios after running the prompt in smart (GPT-5).



**Figure 3:** The histogram and pie chart displays customer id and product area percentage of total sales cost.

**First Scenario, Total Sales Cost Across All Product Areas:**

- Customer 150 stands out as the highest spender with \$10,422.32, followed closely by customer 224 and 10.
- Customer 193 has the lowest total sales cost among the top five, though still significantly above the average customer in the full dataset of 870 customers.
- The pie chart visually reinforces the dominance of customers 150, 224, and 10, who together make up the majority of the total sales cost in this scenario.
- The histogram shows a relatively tight cluster of high-value customers, indicating that the top five customers are all strong contributors to overall revenue.

**Key Insight:** The top five customers contribute disproportionately to total sales, suggesting a high-value customer segment that should be prioritized for retention strategies, loyalty programs, or targeted marketing.

**Second Scenario, Sales Cost by Product Area:**

- Customer 224 dominates this scenario with \$5,656.33 in the non-food product area far exceeding the others.
- Customers 10, 150, and 193 show much smaller contributions in their respective product categories.
- The pie chart makes this contrast even clearer: customer 224 occupies nearly half of the entire distribution.
- The histogram shows a steep drop-off after customer 224, indicating that spending in specific product areas varies widely among the same customers.

**Key Insight:** Customer 224 is a category-specific power buyer, especially in non-food products. This suggests opportunities for category-focused promotions, cross-selling, or bundling strategies targeted at this customer.

**Combined Interpretation**

- In the first scenario, spending is even distributed among the top customers.
- In the second scenario, spending becomes highly concentrated, with one customer dominating a single product area.
- This contrast highlights the importance of analyzing both overall customer value and category-specific behavior to build a complete customer profile.

**4.4. Methodology of Study 13: A table with total sales cost from each product area**

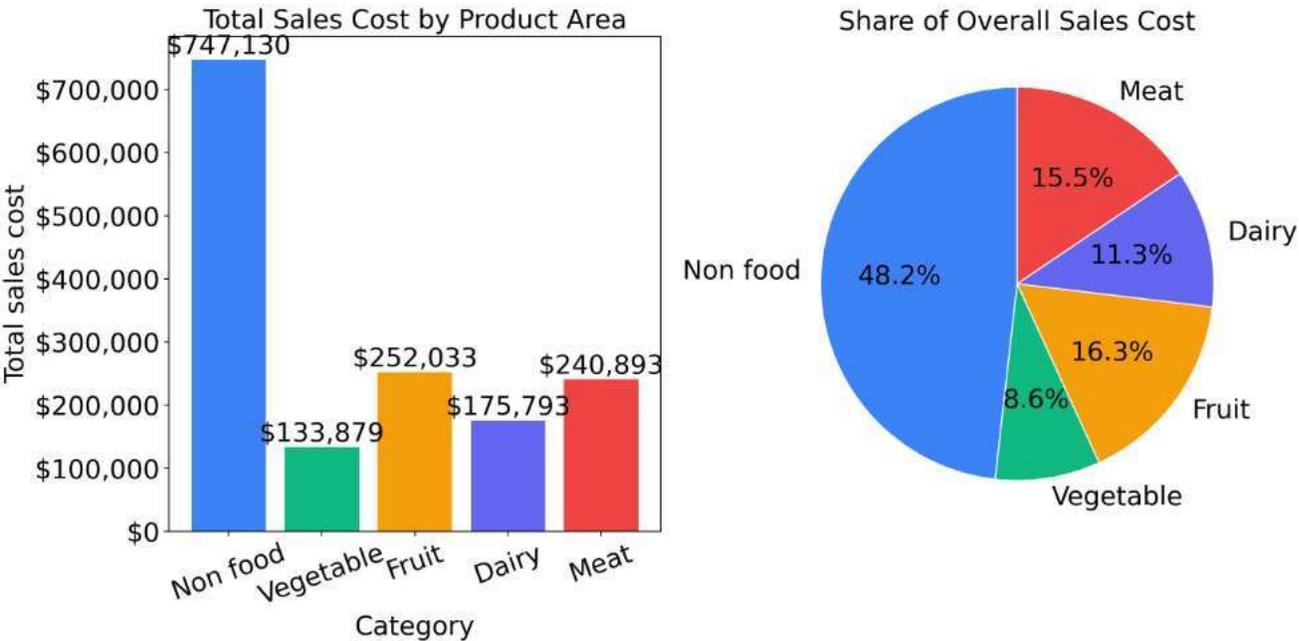
After identifying the best five customers by their identity, an SQL code was then used to estimate total sales from each product area (see section 2.4). The sum was then reported as the overall sales cost of the grocery store. The results are displayed in **Table 1** below.

Table 1		
Category_name	Product_area_id	Total_sales_cost
Non food	1	747129.45
Vegetable	2	133879.25
Fruit	3	252033.53
Dairy	4	175792.77
Meat	5	240892.82
*****	Total overall sales cost	1,549,727.82

The data obtained in **Table 1** were then inserted into smart (GPT-5) to generate the histogram and pie chart showing total sales by product area and share of the overall sales cost in percentage as described in section 4.5. The results are displayed in **Figure 4** below.

**4.5. Methodology of Study 14: A histogram/pie chart showing total sales cost for each product area**

In Figure 4 it shows that Non-Food products generate the largest share of revenue at about 48%, making them the key sales driver but also a potential concentration risk. Fruit and Meat, contributing roughly 16% and 15.5% respectively, offer strong opportunities for seasonal promotions and cross-category bundling. Dairy accounts for about 11.3% of sales and can be further optimized through strategic pricing and emphasis on high-margin items. Vegetables hold the smallest share at 8.6%, indicating room for growth through improved assortment, pricing, placement, and promotional strategy.



**Figure 4.** Combined histogram and pie chart showing a share of overall sales costs in percentage.

**Summary and Insights**

- Nonfood accounts for ≈48% of overall sales—primary revenue driver; sustain momentum while managing category concentration risk.
- Fruit (≈16%) and Meat (≈15.5%) form a strong second tier—good targets for seasonal promos and cross-category bundles.
- Dairy (≈11.3%) contributes steadily; optimize pricing and highlight high-margin SKUs to lift performance.

- Vegetable ( $\approx 8.6\%$ ) is the smallest share; review assortment, pricing, shelf placement, and promotion cadence to unlock growth.

#### 4.6. Methodology of Study 15: Interactive Dashboards Created by Python\_AI

To create an interactive dashboard, the following prompt was created and run in smart (GPT-5):

**Prompt:** In the following description in first, and second, scenarios, create REST API Interactive dashboard using only python code in notebook, make sure the histogram and pie chart are small enough to fit in a normal page and the words are big enough to be visible easily. Place a histogram and a pie chart of first scenario on the top of histogram and pie chart of the second scenario. Make sure the font size is large enough to display clear reading, say above 20. Match the color from the histogram to that in the pie chart: In addition to that, make sure you provide insights in both scenarios.

First scenario: You are given customer id 10 then sales cost for all product area is \$ 9626.02, customer id 593 then total sales cost for all product is \$ 8982.19, customer id 193 then total sales cost for all product area is \$ 6967.07, customer id 224 then sales cost for all product id is \$9878.76, customer id 150 then sales cost for all product area is 10422.32. Provide a filter by customer 10, customer 224, customer 593, customer 193 in the interactive dashboard.

Second scenario: You are given customer id 193 then sales cost for fruit product area is \$1116.95, customer id 593 then sales cost for meat product area is \$2,080.37, customer id 224 then sales cost for non-food product area is \$5,656.33, customer id 10 then sales cost for dairy product area is \$1,231.42, customer id 150 then sales cost for Vegetable products area is \$1,247.21. Provide filters by nonfood, meat, vegetables, dairy, and fruits in the interactive dashboard.

The following code was generated after running the prompt in smart (GPT-5):

```
import pandas as pd
import plotly.graph_objects as go
from ipywidgets import widgets, VBox
from IPython.display import display
# -----
# DATA
# -----
scenario1 = pd.DataFrame({
    "CustomerID": [10, 593, 193, 224, 150],
    "SalesCost": [9626.02, 8982.19, 6967.07, 9878.76, 10422.32]
})
scenario2 = pd.DataFrame({
    "CustomerID": [193, 593, 224, 10, 150],
    "SalesCost": [1116.95, 2080.37, 5656.33, 1231.42, 1247.21]
```

```

})
# Color palette
colors = ['#1f77b4', '#ff7f0e', '#2ca02c', '#d62728', '#9467bd']
# -----
# FUNCTION TO UPDATE DASHBOARD
# -----
def update_dashboard(scenario_choice):
    if scenario_choice == "Scenario 1":
        df = scenario1
        insights = """
        • Customer 150 has the highest total sales cost ($10,422.32).
        • Customer 193 has the lowest total sales cost ($6,967.07).
        • Customers 150 and 224 dominate the total sales volume.
        """
    else:
        df = scenario2
        insights = """
        • Customer 224 has the highest product-area sales cost ($5,656.33).
        • Customer 193 has the lowest product-area sales cost ($1,116.95).
        • Scenario 2 shows a more uneven distribution, with customer 224 contributing the most.
        """

    # Create figure with histogram + pie chart
    fig = make_subplots(
        rows=1, cols=2,
        specs=[[{"type": "bar"}, {"type": "pie"}]],
        subplot_titles=[f"{scenario_choice} Histogram", f"{scenario_choice} Pie Chart"]
    )
    # Histogram
    fig.add_trace(
        go.Bar(
            x=df["CustomerID"],
            y=df["SalesCost"],
            marker_color=colors
        ),
        row=1, col=1
    )
    # Pie chart

```

```

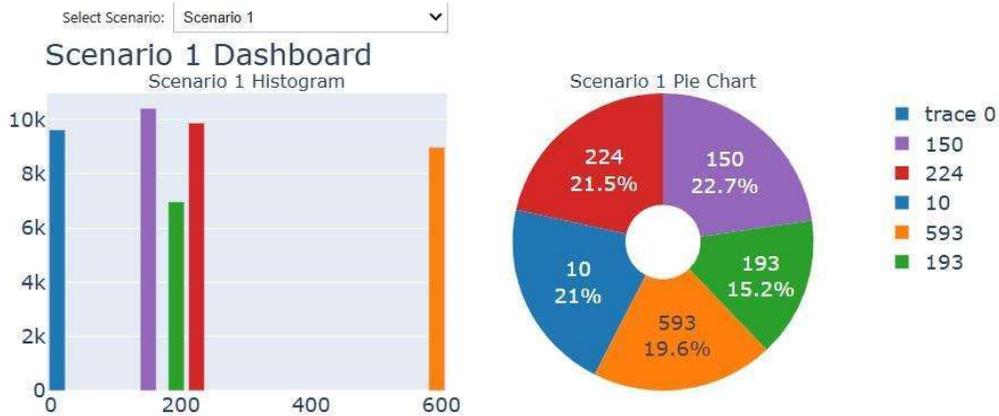
fig.add_trace(
    go.Pie(
        labels=df["CustomerID"],
        values=df["SalesCost"],
        marker=dict(colors=colors)
    ),
    row=1, col=2
)

fig.update_layout(
    height=450,
    width=1000,
    font=dict(size=20),
    title_text=f"{scenario_choice} Dashboard",
    title_font_size=28
)

fig.show()
print("INSIGHTS")
print(insights)
# -----
# DROPDOWN WIDGET
# -----
dropdown = widgets.Dropdown(
    options=["Scenario 1", "Scenario 2"],
    value="Scenario 1",
    description="Select Scenario:",
    style={'description_width': '150px'},
    layout=widgets.Layout(width='400px')
)
# -----
# LINK DROPDOWN TO FUNCTION
# -----
widgets.interact(update_dashboard, scenario_choice=dropdown)

```

The code above was then run in jupyter notebook to generate interactive dashboards shown in Figure 5(a) and 5(b). Figure 5(b) is a typical REST API display after the code modification in python notebook; the modified code is not shown in this report.

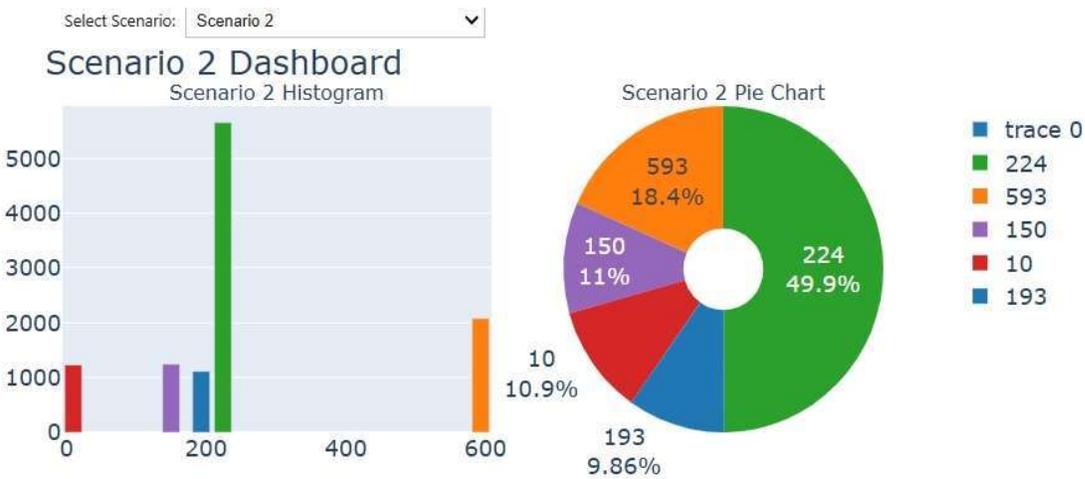


#### Insights

Customer 150 has the highest total sales cost at 10,422.32. *Customer 193 has the lowest total sales cost at 6,967.07.*

Customers 150 and 224 contribute the largest share of total sales.

```
] : <function __main__.update_dashboard(scenario_choice)>
```



#### Insights

Customer 224 has the highest product-area sales cost at 5,656.33. *Customer 193 has the lowest product – area sales cost at 1,116.95.*

Scenario 2 shows a more uneven distribution, with customer 224 contributing the most.

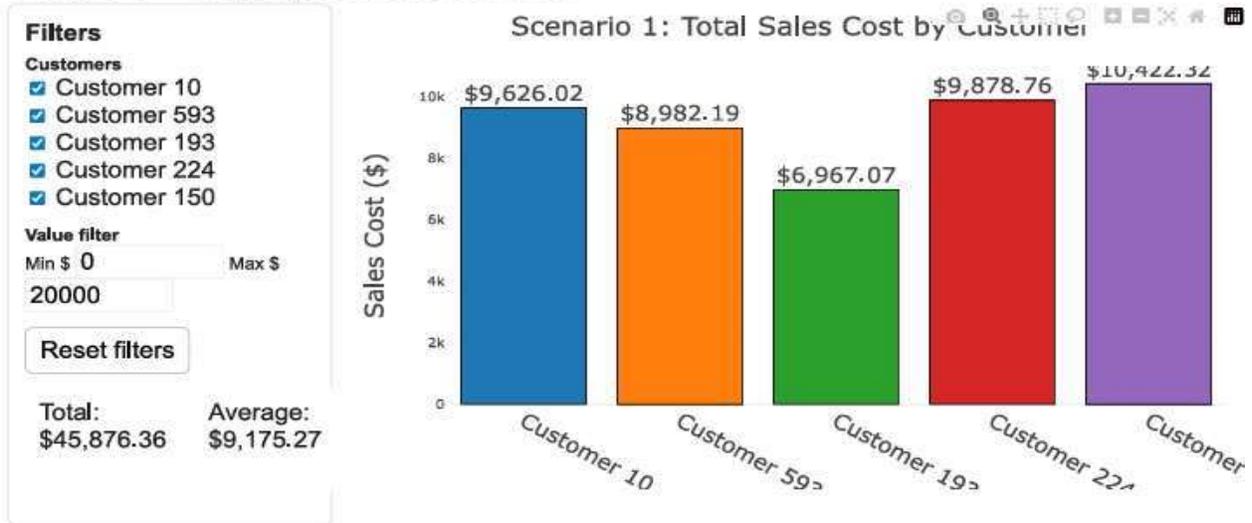
```
] : <function __main__.update_dashboard(scenario_choice)>
```

**Figure 5(a)** is an interactive dashboard created using Python-AI for scenario 1 and scenario 2.

## Interactive Sales Dashboard: Scenario 1 (Top) and Scenario 2 (Bottom)

Use the filters to update histograms and pie charts. Fonts are  $\geq 22$ pt and colors are matched between each histogram and its pie chart.

### Scenario 1 — Total Sales Cost by Customer



### Scenario 2 — Sales Cost by Product Area (Customer shown)



Tip: If you run the REST API (python app.py), this page will auto-load data from <http://localhost:8000>; otherwise it uses the embedded dataset.

**Figure 5(b)** is an interactive dashboard scenario 1(top) and scenario 2 (bottom) created using Python-AI.

## 5. Conclusions

The analysis of the grocery store's transaction database provides a clear and comprehensive understanding of both product-area performance and customer purchasing behavior. By examining the structure of the dataset, validating its completeness, and applying SQL-based aggregation techniques, the study establishes a reliable foundation for evaluating sales trends across the business. Calculations of minimum, average, and maximum sales costs, combined with total sales by product area and customer, reveal meaningful patterns that guide deeper interpretation. Visualizations created through Excel smart (GPT-5) and Python-AI further enhance clarity, allowing complex relationships such as the distribution of spending across customers and categories more transparency and easy to communicate with precision.

Across all findings, one theme emerges consistently: non-food products are the dominant revenue driver, far surpassing every food category. This category not only leads in total sales but also attracts the highest spending customers, most notably customer 224, whose purchasing behavior significantly influences category performance. Among food items, meat, fruit, and dairy form a secondary tier of contribution, while vegetables represent the smallest share of sales. These differences highlight where the store is currently strongest and where strategic attention could unlock additional growth.

The two comparative scenarios with total spending across all product areas and spending within specific categories offer valuable contrast. In the first scenario, the top five customers show relatively balanced spending levels, forming a concentrated group of high value buyers who collectively contribute a substantial portion of total revenue. This consistency suggests a stable core customer segment that would benefit from loyalty initiatives and retention focused strategies. In the second scenario, however, spending becomes highly uneven, with customer 224 overwhelmingly dominating the non-food category. This sharp concentration underscores the importance of understanding not only who the top customers are, but what they buy and where their preferences lie.

These insights point to several actionable opportunities. Strengthening lower-performing categories such as dairy and vegetables may require adjustments in pricing, assortment, or promotional cadence. Meanwhile, high performing areas like non-food and meat can be leveraged through targeted cross selling, bundling, and personalized offers aimed at customers who already show strong engagement. The contrast between overall customer value and category with specific behavior reinforces the need for a dual lens approach to customer analytics and to consider both total contribution and purchasing specialization.

Overall, the analysis delivers a data-driven perspective that equips management with the information needed to refine inventory planning, optimize category strategies, and tailor marketing efforts. By understanding which product areas excel, which require support, and how top customers behave across categories, the store is positioned to enhance sales performance, strengthen customer satisfaction, and make informed decisions that support long-term growth.

### **Acknowledgment**

I would like to express my sincere gratitude to Andrew Jones, Team Leader and Instructor at Data Science Infinity (DSI), for granting me access to the DSI Grocery Database. His support has been instrumental in enabling me to complete the requirements for the Certificate in Data Analytics Using SQL.

I also extend my heartfelt appreciation to Andrew Ng, Founder of DeepLearning.AI and Co-founder of Coursera, for his continued support. His generosity in allowing me ongoing access to Coursera's AI tools, even after completing my DeepLearning.AI training and certification, has been invaluable to my learning journey.