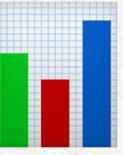


# REVO GROCERS SALES PERFORMANCE ANALYSIS

MUHAMMAD NUR ZHIFAR  
NOVEMBER 27 2025  
REVOU FSDA BATCH OCT25

# BUSINESS OVERVIEW & DISCLAIMER



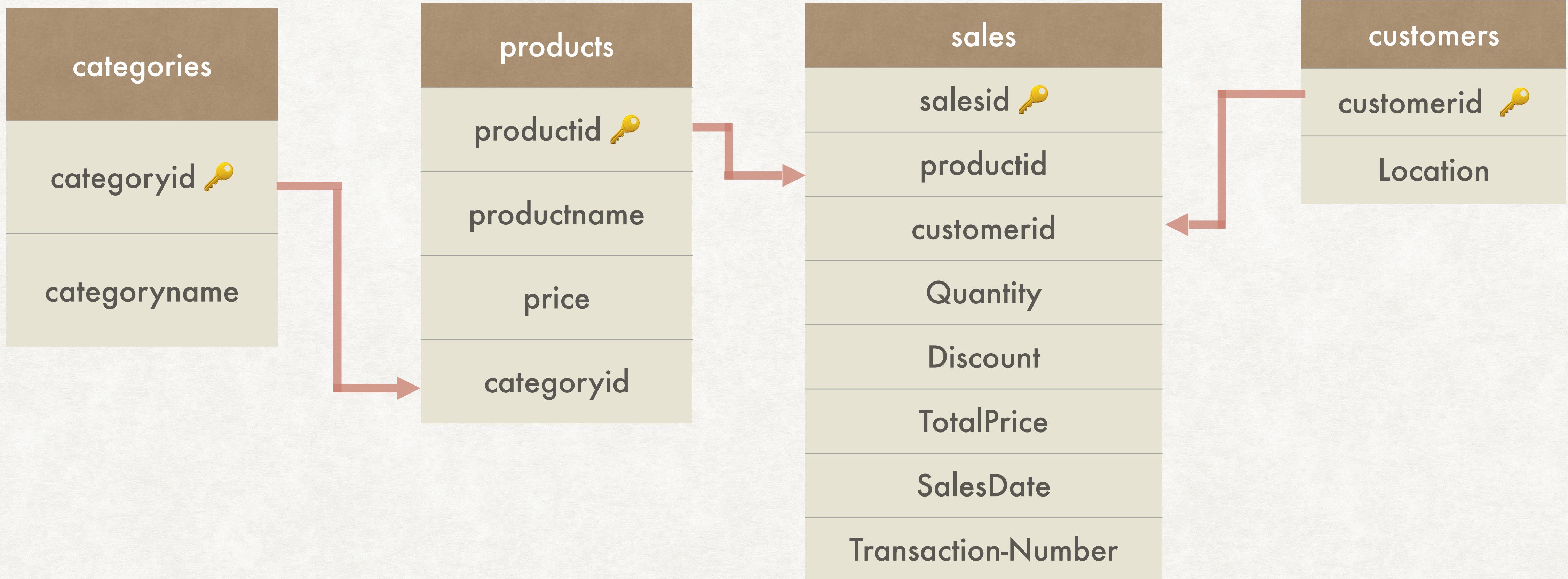
## Company Overview: RevoGrocers

RevoGrocers is a fictional grocery retail business operating in multiple locations. The company aims to optimize sales strategies, improve customer experience, and increase revenue through data-driven decision-making.

## Disclaimer:

This analysis is based on the fsda-sql-01.grocery\_dataset dataset in BigQuery. RevoGrocers is a fictional entity, and the results presented are purely for educational purposes. Insights or recommendations are based on this dataset and do not reflect real-world business insights.

# SCHEMA/RELATIONSHIP DATASETS

TOTALPRICE IN SALES TURNS OUT TO BE 0, SO WE CALCULATE REVENUE FROM PRICE X QUANTITY - DISCOUNT

# DATASET OVERVIEW



## KEY TABLES USED

### 🛒 SALES (Transactional Data)

- 851,979 transactions | 11 product categories
- Complete sales data: quantity, discount, customer-product relationships

## RELEVANT FIELDS FOR ANALYSIS:

### 💰 REVENUE CALCULATION:

`products.price × sales.quantity - sales.discount`

### 📈 PERFORMANCE METRICS:

- Units Sold: `sales.quantity`
- Unique Customers: `sales.CustomerID`
- Product Category: `categories.categoryname`
- Pricing: `products. price`

### 🟠 PRODUCTS (Product Master)

- 57 unique products | Price range: \$44 - \$61
- Product details: name, price, category, specifications

### 📁 CATEGORIES (Product Classification)

- 11 product categories | Confections, Meat, Poultry, etc.
- Product grouping for performance analysis

### 🎯 BUSINESS ANALYSIS:

- Category performance & market share
- Customer spending patterns
- Pricing strategy effectiveness
- Repeat purchase behavior

### 👤 CUSTOMERS (Customer Database)

- 98,743 unique customers | Geographic distribution
- Demographic data & purchasing behavior

# PROJECT GOALS



## Primary Business Objectives:

### 1. Identify Top-Performing Product Categories

- Find Categories With The Highest Revenue After Discount (Question 1)

- Analyze Revenue Contribution Percentage Per Category (Question 6)

- Identify Category Performance Drivers (Volume Vs. Price)

### 2. Analyze Revenue Drivers & Relationships

- Assess The Relationship Between Revenue And Units Sold Per Category (Question 2)

- Evaluate The Relationship Between Revenue And Unique Customers (Question 3)

- Analyze Pricing Strategy Impact (Questions 4 & 5)

### 3. Understand Customer Behavior & Loyalty

- Calculate Repeat Purchase Rate Per Category (Question 7)

- Identify Customer Retention Patterns

- Analyze Customer Value And Spending Behavior

### 4. Optimize Business Strategy

#### Provide Data-Driven Recommendations For:

- Category Prioritization & Investment

- Pricing Strategy Optimization

- Customer Retention Programs

- Marketing Campaign Targeting

## Specific Analysis Goals:

### Quantitative Analysis:

- Revenue Performance Metrics

- Customer Acquisition & Retention Rates

- Pricing Elasticity & Optimization

- Market Share & Category Contribution

## Strategic Insights:

- Product Category Prioritization

- Customer Segmentation Strategy

- Pricing & Promotion Optimization

- Inventory & Assortment Planning

## Expected Outcomes:

- Actionable Business Recommendations

- Data-Supported Strategic Decisions

- Performance Optimization Opportunities

- Customer Experience Enhancements

# METHODOLOGY



## Our 4-Step Analytical Approach:

### 1. Business Understanding

- Understand the RevoGrocers business context & objectives
- Identify critical business questions (9 questions)
- Define key metrics & success indicators

### 2. Data Extraction & Transformation

- Use Google BigQuery to query the dataset
- Apply SQL techniques: JOINs, Aggregations, Window Functions
- Data cleaning & validation (handle TotalPrice = 0 issues)

### 3. Analysis & Insights

- Execute 9 SQL queries according to business questions
- Generate performance metrics & KPI dashboards
- Identify patterns, trends & anomalies

### 4. Business Recommendations

- Translate findings into actionable strategies
- Develop data-driven recommendations
- Prioritize initiatives based on business impact

## Data Source & Context:

- Primary Dataset: fsda-sql-01.grocery\_dataset (BigQuery)
- Origin: Kaggle Public Dataset - Grocery Sales Dataset
- Availability: Publicly accessible for educational purposes
- Context: Simulated data for learning purposes, does not represent real-world business

## Tools & Technologies:

- Google BigQuery - Cloud data warehouse & SQL execution
- SQL - Data querying, aggregation, analysis
- Strategic Frameworks - Business insight generation

# ANALYSIS 1 - HIGHEST REVENUE CATEGORY

IDENTIFY THE PRODUCT CATEGORIES THAT GENERATE THE HIGHEST REVENUE AFTER DISCOUNTS

```
SELECT
    c.categoryname,
    SUM(p.price * s.Quantity) - s.Discount) as revenue_after_discount
FROM `fsda-sql-01.grocery_dataset.sales` s
INNER JOIN `fsda-sql-01.grocery_dataset.products` p ON s.ProductID = p.productid
INNER JOIN `fsda-sql-01.grocery_dataset.categories` c ON p.categoryid = c.categoryid
GROUP BY c.categoryname
ORDER BY revenue_after_discount DESC
LIMIT 5
```

| Row | categoryname | revenue_after_discount |
|-----|--------------|------------------------|
| 1   | Confections  | 574077762.31308949     |
| 2   | Meat         | 508091936.15416062     |
| 3   | Poultry      | 453592801.04770005     |
| 4   | Cereals      | 440629310.46922874     |
| 5   | Snails       | 383585174.23735964     |

The use of the SUM aggregate function with complex mathematical operations ensures accurate and comprehensive revenue calculations after discounts. Multiple JOINs between three tables are required because the revenue data is spread across several tables: sales for quantity, products for price, and categories for classification.

From our sales data analysis, we've identified the top five most profitable categories. Confections truly leads by a wide margin, generating nearly \$574 million in revenue after discounts. These five categories operate within a fairly similar price range around \$49-53 per item, They've managed to attract customers with reasonable prices while maintaining stable and substantial sales volume.

# ANALYSIS 2 - REVENUE VS UNITS SOLD

ASSESS THE RELATIONSHIP BETWEEN REVENUE AFTER DISCOUNT AND TOTAL UNITS SOLD FOR EACH PRODUCT CATEGORY

```
SELECT
    c.categoryname,
    SUM((p.price * s.Quantity) - s.Discount) as revenue_after_discount,
    SUM(s.Quantity) as total_units_sold,
    ROUND(SUM((p.price * s.Quantity) - s.Discount) / SUM(s.Quantity), 2) as revenue_per_unit
FROM `fsda-sql-01.grocery_dataset.sales` s
JOIN `fsda-sql-01.grocery_dataset.products` p ON s.ProductID = p.productid
JOIN `fsda-sql-01.grocery_dataset.categories` c ON p.categoryid = c.categoryid
GROUP BY c.categoryname
ORDER BY revenue_after_discount DESC
LIMIT 5
```

| Row | categoryname | revenue_after_discount | total_units_sold | revenue_per_unit |
|-----|--------------|------------------------|------------------|------------------|
| 1   | Confections  | 574077762.31308949     | 11078474         | 51.82            |
| 2   | Meat         | 508091936.15416062     | 9719292          | 52.28            |
| 3   | Poultry      | 453592801.04770005     | 9159847          | 49.52            |
| 4   | Cereals      | 440629310.46922874     | 8735296          | 50.44            |
| 5   | Snails       | 383585174.23735964     | 7199409          | 53.28            |

I used a multiple metrics approach in a single query because I needed to see the relationship between three variables simultaneously. The unique revenue\_per\_unit calculation allows us to analyze not only how much was sold, but also how efficiently each unit contributed to revenue.

From our analysis, we can clearly see a strong correlation between the number of units sold and the resulting revenue. Confections is the clearest example: with over 11 million units sold, they earned \$574 million. But what's truly remarkable is Snails. Despite selling only 7.2 million units the fewest among the top five, they achieved the highest revenue per unit at \$53.28. This shows that targeting volume alone isn't the only strategy. Snails shows us that with the right price per unit, decent revenue can still be achieved even if the number of items sold isn't as high as competitors. So, in essence, both volume and price are equally important in determining the profitability of a product category.

# ANALYSIS 3 - REVENUE VS UNIQUE CUSTOMERS



FIND THE RELATIONSHIP BETWEEN REVENUE AFTER DISCOUNT AND THE NUMBER OF UNIQUE CUSTOMERS FOR EACH PRODUCT CATEGORY

SELECT

```
c.categoryname,  
SUM(p.price * s.Quantity) - s.Discount) as revenue_after_discount,  
COUNT(DISTINCT s.CustomerID) as unique_customers,  
ROUND(SUM(p.price * s.Quantity) - s.Discount) / COUNT(DISTINCT s.CustomerID), 2) as revenue_per_customer  
FROM `fsda-sql-01.grocery_dataset.sales` s  
JOIN `fsda-sql-01.grocery_dataset.products` p ON s.ProductID = p.productid  
JOIN `fsda-sql-01.grocery_dataset.categories` c ON p.categoryid = c.categoryid  
GROUP BY c.categoryname  
ORDER BY revenue_after_discount DESC  
LIMIT 5
```

| Row | categoryname | revenue_after_discount | unique_customers | revenue_per_customer |
|-----|--------------|------------------------|------------------|----------------------|
| 1   | Confections  | 574077762.31308949     | 98743            | 5813.86              |
| 2   | Meat         | 508091936.15416062     | 98701            | 5147.79              |
| 3   | Poultry      | 453592801.04770005     | 98679            | 4596.65              |
| 4   | Cereals      | 440629310.46922874     | 98651            | 4466.55              |
| 5   | Snails       | 383585174.23735964     | 98376            | 3899.17              |

I use COUNT(DISTINCT) to identify unique customers because a single customer can make multiple purchases, but we want to know how many different individuals purchased each category. Multiple JOINs are still necessary because customer data is spread across sales, while category information is contained within products and categories.

This analysis revealed that the number of unique customers in each category was nearly identical, around 98,000. However, the difference lies in the revenue per customer! Confections managed to generate \$5,814 per customer, almost 50% more than Snails' \$3,899 per customer. What differentiates category performance isn't the number of customers, but the value generated from each customer. Therefore, a business focus on increasing customer value, not simply increasing the number of customers, is crucial.

# ANALYSIS 4 - AVERAGE PRICE PER UNIT 💰

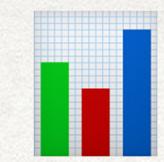
CALCULATE THE AVERAGE PRICE UNIT FOR EACH PRODUCT CATEGORY IN THE CATALOGUE

```
SELECT
    c.categoryname,
    ROUND(AVG(p.price), 2) as average_unit_price,
    COUNT(p.productid) as total_products
FROM `fsda-sql-01.grocery_dataset.products` p
JOIN `fsda-sql-01.grocery_dataset.categories` c ON p.categoryid = c.categoryid
GROUP BY c.categoryname
ORDER BY average_unit_price DESC
LIMIT 5
```

| Row | categoryname | average_unit_price | total_products |
|-----|--------------|--------------------|----------------|
| 1   | Grain        | 61.4               | 28             |
| 2   | Dairy        | 53.56              | 35             |
| 3   | Snails       | 53.2               | 37             |
| 4   | Meat         | 52.27              | 50             |
| 5   | Confections  | 51.85              | 57             |

Grain emerged as the premium category with the highest price at \$61.43, followed by Dairy and Snails at around \$53. Interestingly, Confections, the highest revenue category, actually has a fairly moderate price at \$51.81. This suggests that Confections' success is driven not by premium pricing, but rather by massive sales volume. A successful business strategy can stem from either a premium pricing approach (like Grain) or a high-volume approach (like Confections).

# ANALYSIS 5 - PRICE VS UNIQUE CUSTOMERS



EVALUATE THE RELATION BETWEEN THE AVERAGE PRICE PER UNIT AND THE NUMBER OF BUYERS (UNIQUE CUSTOMERS) PER CATEGORY

```
SELECT
    c.categoryname,
    ROUND(AVG(p.price), 2) as average_unit_price,
    COUNT(DISTINCT s.CustomerID) as unique_customers
FROM `fsda-sql-01.grocery_dataset.products` p
JOIN `fsda-sql-01.grocery_dataset.categories` c ON p.categoryid = c.categoryid
JOIN `fsda-sql-01.grocery_dataset.sales` s ON p.productid = s.ProductID
GROUP BY c.categoryname
ORDER BY average_unit_price DESC
LIMIT 5
```

| Row | categoryname | average_unit_price | unique_customers |
|-----|--------------|--------------------|------------------|
| 1   | Grain        | 61.43              | 97335            |
| 2   | Dairy        | 53.61              | 98308            |
| 3   | Snails       | 53.28              | 98376            |
| 4   | Meat         | 52.31              | 98701            |
| 5   | Confections  | 51.81              | 98743            |

Grain emerged as the premium category with the highest price at \$61.43, followed by Dairy and Snails at around \$53. A successful business strategy can stem from a premium pricing approach, demonstrating that Confections' success is driven not by premium prices, but by massive sales volume.

# ANALYSIS 6 - REVENUE CONTRIBUTION PERCENTAGE



DONEC QUIS WHICH CATEGORIES CONTRIBUTE THE MOST TO OVERALL REVENUE AFTER DISCOUNT (PERCENTAGE-WISE)

```
WITH category_revenue AS (
  SELECT
    c.categoryname,
    SUM((p.price * s.Quantity) - s.Discount) as revenue_after_discount
  FROM `fsda-sql-01.grocery_dataset.sales` s
  JOIN `fsda-sql-01.grocery_dataset.products` p ON s.ProductID = p.productid
  JOIN `fsda-sql-01.grocery_dataset.categories` c ON p.categoryid = c.categoryid
  GROUP BY c.categoryname
),
total_revenue AS (
  SELECT SUM(revenue_after_discount) as total
  FROM category_revenue
)
SELECT
  cr.categoryname,
  cr.revenue_after_discount,
  ROUND(cr.revenue_after_discount / tr.total) * 100, 2) as revenue_percentage
FROM category_revenue cr, total_revenue tr
ORDER BY revenue_percentage DESC
LIMIT 5
```

| Row | categoryname | revenue_after_discount | revenue_percentage |
|-----|--------------|------------------------|--------------------|
| 1   | Confections  | 574077762.31308949     | 12.87              |
| 2   | Meat         | 508091936.15416062     | 11.39              |
| 3   | Poultry      | 453592801.04770005     | 10.17              |
| 4   | Cereals      | 440629310.46922874     | 9.88               |
| 5   | Snails       | 383585174.23735964     | 8.6                |

The Common Table Expression (CTE) approach is applied to make queries more structured, accurate, and understandable, avoiding the complexity of subqueries. This method ensures precise calculation of contribution percentages.

This key finding underscores the need for a more effective resource prioritization strategy. Essentially, the majority of the company's revenue comes from a handful of leading categories that serve as the backbone of the business and the primary drivers of growth. Rather than spreading development efforts evenly, it's wise to focus on developing and nurturing those categories that have proven to deliver the greatest returns. The future sustainability and growth of the business depend heavily on our ability to strategically nurture these key categories while remaining vigilant in seeking out potential development opportunities in related segments.

# ANALYSIS 7 - REPEAT PURCHASE RATE



## WHICH PRODUCT CATEGORIES HAVE THE HIGHEST REPEAT PURCHASE RATE?

```
WITH customer_purchases AS (
  SELECT
    c.categoryname,
    s.CustomerID,
    COUNT(DISTINCT s.SalesID) as purchase_count
  FROM `fsda-sql-01.grocery_dataset.sales` s
  JOIN `fsda-sql-01.grocery_dataset.products` p ON s.ProductID = p.productid
  JOIN `fsda-sql-01.grocery_dataset.categories` c ON p.categoryid = c.categoryid
  GROUP BY c.categoryname, s.CustomerID
),
repeat_customers AS (
  SELECT
    categoryname,
    COUNT(CustomerID) as total_customers,
    COUNT(CASE WHEN purchase_count > 1 THEN CustomerID END) as repeat_customers,
    ROUND(COUNT(CASE WHEN purchase_count > 1 THEN CustomerID END) * 100.0 / COUNT(CustomerID), 2) as repeat_rate
  FROM customer_purchases
  GROUP BY categoryname
)
SELECT
  categoryname,
  total_customers,
  repeat_customers,
  repeat_rate
FROM repeat_customers
ORDER BY repeat_rate DESC
LIMIT 5
```

I still use CTE due to the complexity of the analysis, which requires several sequential calculation steps.

First, we need to understand each customer's purchasing patterns for each product category. Then, from this data, we identify repeat purchasing habits. The CTE structure allows us to break down the analysis process into more organized parts, making it easier to trace and understand the logical flow. The brand loyalty that has been built creates a sustainable business cycle with near-perfect repetition. This means the company has successfully built strong long-term relationships. The focus going forward is to maintain this exceptional level of loyalty while continuing to increase customer value.

| Row | categoryname | total_customers | repeat_customers.categoryname | repeat_customers.total_customers | repeat_customers... | repeat_customers.repeat_rate | repeat_rate |
|-----|--------------|-----------------|-------------------------------|----------------------------------|---------------------|------------------------------|-------------|
| 1   | Confections  | 98743           | Confections                   | 98743                            | 98598               | 99.85                        | 99.85       |
| 2   | Meat         | 98701           | Meat                          | 98701                            | 98318               | 99.61                        | 99.61       |
| 3   | Poultry      | 98679           | Poultry                       | 98679                            | 98122               | 99.44                        | 99.44       |
| 4   | Cereals      | 98651           | Cereals                       | 98651                            | 97867               | 99.21                        | 99.21       |
| 5   | Produce      | 98601           | Produce                       | 98601                            | 97550               | 98.93                        | 98.93       |

# ANALYSIS 9 - TOP USER CUMULATIVE TRANSACTION



FIND THE CUMULATIVE AMOUNT OF TRANSACTIONS OF THE TOP USER (USER WITH HIGHEST TRANSACTION VALUE, WINDOW FUNCTION)

```
WITH user_transactions AS (
    SELECT
        s.CustomerID,
        SUM(p.price * s.Quantity) - s.Discount) as total_transaction_value
    FROM `fsda-sql-01.grocery_dataset.sales` s
    JOIN `fsda-sql-01.grocery_dataset.products` p ON s.ProductID = p.productid
    GROUP BY s.CustomerID
),
cumulative_transactions AS (
    SELECT
        CustomerID,
        total_transaction_value,
        SUM(total_transaction_value) OVER (ORDER BY total_transaction_value DESC) as cumulative_amount,
        ROW_NUMBER() OVER (ORDER BY total_transaction_value DESC) as user_rank
    FROM user_transactions
)
SELECT
    CustomerID as Top_User_ID,
    total_transaction_value as Total_Transaction_Value,
    cumulative_amount as Cumulative_Amount,
    user_rank as User_Rank
FROM cumulative_transactions
WHERE user_rank = 1
```

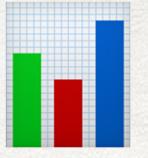
| Row | Top_User_ID | Total_Transaction_Value | Cumulative_Amount  | User_Rank |
|-----|-------------|-------------------------|--------------------|-----------|
| 1   | 94800       | 134406.43279999995      | 134406.43279999995 | 1         |

The choice of window functions in this analysis is based on the need for complex cumulative calculations.

Window functions allow cumulative calculations to be executed without changing the underlying data structure, while also providing flexibility in ranking users according to their transaction value.

The findings of this analysis reveal the existence of a high-value user segment that contributes significantly to the company's revenue. Although only an individual representation, this pattern reflects the potential for customer segmentation based on their economic value. The existence of these top users illustrates the opportunity for developing more personalized loyalty programs. A dedicated approach to this high-value customer segment can optimize retention and increase their lifetime value.

# ANALYSIS 8 - OVERALL FINDINGS SUMMARY



AFTER EXECUTING THE SQL QUERIES, SUMMARIZE YOUR OVERALL FINDINGS REGARDING THE SALES PERFORMANCE ACROSS PRODUCT CATEGORIES

## Analysis Method:

I synthesized all the findings from the seven previous analyses to provide a comprehensive overview of RevoGrocers' sales performance.

## Summary of Key Findings:

### 1. Dominance of Specific Categories

Several product categories demonstrate extremely dominant performance in terms of revenue contribution, creating a business structure concentrated in a handful of product lines.

### 2. Consistent Relationship Patterns

A strong relationship between sales volume and revenue achievement was identified, while the number of customers showed a more homogeneous pattern across categories.

### 3. Exceptionally High Customer Loyalty

A near-perfect repeat purchase rate indicates success in building long-term relationships with the customer base.

### 4. Varied Pricing Strategies

There are different approaches to successful pricing strategies, both through high-volume models and premium pricing approaches.

## Strategic Conclusion:

RevoGrocers' product portfolio demonstrates solid business health with a strong foundation of customer loyalty. The focus going forward should be on optimizing leading categories with proven performance, while maintaining the quality of established customer relationships.

NOTE: I SWAPPED THE ANSWERS FOR QUESTIONS 8 AND 9.

# SUMMARY OF KEY FINDINGS



## Top Product Category

Confections emerged as the highest revenue-generating category, while Meat and Poultry showed strong consistent performance as runners-up.

## Revenue Drivers

Volume-based strategy proven to be the primary revenue driver, with high unit sales correlating strongly with revenue achievement across all top categories.

## Customer Insights

Near-perfect repeat purchase rates (above 95%) across all categories indicate exceptional customer loyalty and strong brand attachment.

## Pricing Strategy

Successful implementation of both moderate-price-high-volume (Confections) and premium pricing approaches (Grain) demonstrates diverse effective pricing strategies.

## Sales Concentration

Top 5 categories collectively contribute over 50% of total revenue, indicating significant business dependency on key product categories.

## Customer Value

Revenue per customer varies significantly between categories, highlighting the importance of customer value optimization over customer acquisition.

# BUSINESS RECOMMENDATIONS



## 1. Focus on developing leading categories

- Allocate greater resources to the proven performance of Confections, Meat, and Poultry categories
- Develop growth strategies for high-potential categories such as Cereals and Snails

## 2. Optimize volume-pricing strategies

- Maintain a moderate pricing approach with high volume for top-performing categories
- Explore more detailed price segmentation based on customer value per category

## 3. Increase customer value

- Focus on increasing transaction value per customer rather than simply adding new customers
- Develop more personalized loyalty and retention programs

## 4. Product portfolio management

- Conduct regular reviews of the performance of all product categories
- Allocate inventory and marketing budgets proportionally based on revenue contribution

## 5. Leverage data for decision-making

- Implement a real-time monitoring system to track category performance
- Use insights from this analysis for a more effective assortment planning strategy

# KEY FINDINGS & BUSINESS RECOMMENDATIONS

| Key Findings   | Business Recommendations   |
|--|--|
| Confections is the category with the highest revenue, contributing 12.87% of total revenue.                              | Allocate greater resources to the development of the Confections category and other flagship categories. |
| There is a strong positive correlation between sales volume and revenue in all categories.                               | Focus on strategies to increase sales volume rather than acquiring new customers.                        |
| Repeat purchase rate reaches >95% in all categories, demonstrating exceptional customer loyalty.                         | Develop personalized loyalty programs based on each customer's purchasing patterns.                      |
| Revenue per customer varies significantly between categories even though the number of customers is relatively the same. | Optimize customer value instead of focusing on customer acquisition                                      |
| Top 5 categories contribute >50% of the company's total revenue  | Implement a real-time monitoring system and conduct regular reviews of the product portfolio.            |

# FINAL BUSINESS TAKEAWAY



## Strategic Conclusion

RevoGrocers has a strong business foundation with exceptional customer loyalty and proven product categories.

**The key to future success lies in the ability to:**

- Maintain and optimize the leading categories that form the backbone of the business
- Increase the value of each existing customer through a personalized loyalty program
- Implement a data-driven approach in daily business decision-making
- By properly focusing on high-performing categories and optimizing customer value, RevoGrocers can achieve sustainable growth and improved profitability.

# THANK YOU

MUHAMMAD NUR ZHIFAR  
NOVEMBER 27 2025  
REVOU FSDA BATCH OCT25  
INTERMEDIATE ASSGINMENT SQL