

# RevoGrocers Sales Performance Analysis

Fadila Utami | RevoU FSDA OCT 25 | Intermediate SQL | [Linkedin](#)

MYQUERY | Nov 11, 2025



# Company Overview

**RevoGrocers** is a fictional grocery retailer operating multiple locations with a broad range of everyday products. Facing performance challenges across several categories, the company aims to reallocate resources from low-growth segments to higher-performing ones using data-driven insights to optimize revenue contribution and support strategic decision-making.

## Disclaimer:

- This analysis is based on a public dataset from Kaggle and does not represent real-world business or financial results.  
[Kaggle Grocery](#)
- RevoGrocers is a fictional company created solely for training and educational purposes.
- All insights and recommendations in this analysis are for learning only and should not be used as real business decisions.

A photograph of a man and a woman shopping in a grocery store. The man, wearing glasses and a dark jacket, is looking at a product on the shelf. The woman, wearing a light blue sweatshirt, is standing next to him. They are in an aisle filled with various cereal boxes. The background shows more shelves and products.

## Problem Statement

- How to RevoGrocers improve category performance by 10% next quarter?
- How to the company identify top-performing categories and the key factors driving revenue, including pricing impact?

A photograph showing a long row of grocery store shelves filled with various cereal boxes. The shelves are well-stocked, and the boxes are arranged in an organized manner.

## Objective

- Identify top and low performing product categories based on revenue and sales performance.
- Analyze key revenue drivers including units sold, customer activity, and pricing impact across all categories.

# Methodology

## 1. Data Understanding & Preparation

- Tables used: categories, products, sales
- Included only records with valid product–category relationships

## 2. Data Cleaning & Validation

## 3. SQL Queries

- Query explanation
- Formatted SQL query
- Screenshot of output

## 4. Descriptive Analytics

- Revenue vs. demand
- Pricing vs. buyer reach
- Category percentage contribution to total revenue

## 5. Customer Behavior Analysis

- Purchase frequency analysis using repeat purchase rate
- Focus on categories with the strongest buyer retention



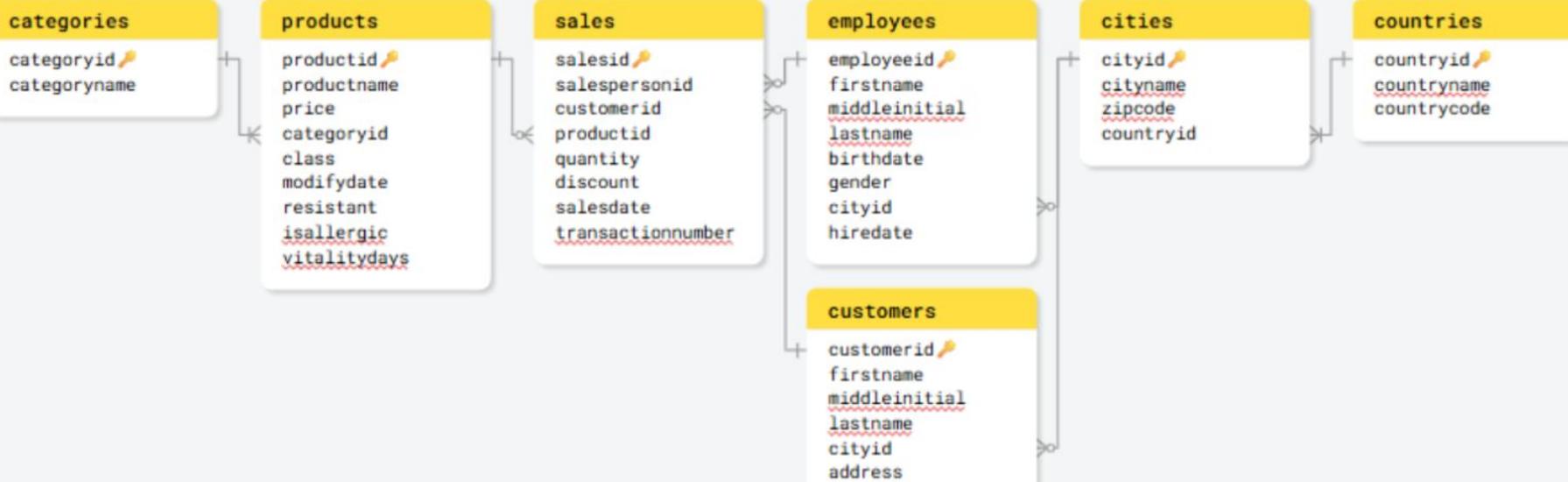
# DATA OVERVIEW

The dataset includes all key retail entities needed for comprehensive analysis, covering sales transactions, product and category details, and demographic information such as customers, employees, cities, and countries.

Data set info	
categories	☆
cities	☆
countries	☆
customers	☆
employees	☆
products	☆
sales	☆
Data set ID	fsda-sql-01.grocery_dataset
Created	18 Feb 2025, 23:07:57 UTC+7
Default table expiry	60 days
Last modified	18 Feb 2025, 23:07:57 UTC+7
Data location	US
Description	
Default collation	
Default rounding mode	ROUNDING_MODE_UNSPECIFIED
Time travel window	7 days
Case insensitive	false
Labels	
Tags	

**NOTE:** This dataset is synthetic and is intended solely for testing, evaluation, and exploratory analysis purposes.

# ENTITY RELATIONSHIP DIAGRAM



The background of the image is a blurred photograph of a city street at night. In the lower-left foreground, a dark metal railing runs diagonally across the frame. The background is filled with out-of-focus lights from buildings and street lamps, creating a bokeh effect.

# BUSINESS INSIGHTS

# Highest Revenue Category After Discount

```
SELECT
    categoryName
    , CAST(ROUND(SUM(t2.quantity * t3.price * (1 - t2.discount)), 0) AS
INT64) AS total_revenue_after_discount
FROM `fsda-sql-01.grocery_dataset.categories` AS t1
INNER JOIN `fsda-sql-01.grocery_dataset.products` AS t3
    ON t1.categoryid = t3.categoryid
INNER JOIN
    `fsda-sql-01.grocery_dataset.sales` AS t2
    ON t3.productid = t2.productid
GROUP BY
    t1.categoryName
ORDER BY
    total_revenue_after_discount DESC
LIMIT 5;
```

Row	categoryName	total_revenue_aft...
1	Confections	556930717
2	Meat	492888845
3	Poultry	440025565
4	Cereals	427393432
5	Snails	372084885

- Confections is the top revenue generator: **\$556.9M**
- This indicates that Confections plays a strategically significant role in driving overall sales performance and should be prioritized in sales planning, inventory allocation, and targeted promotional initiatives.

## Revenue after discount formula:

Original Price × Quantity × (1 – Discount Percentage)

# Relationship Between Revenue After Discount And Total Units Sold

```
SELECT
    categoryName
    ,CAST(ROUND(SUM(t2.quantity * t3.price * (1 - t2.discount)), 0) AS
INT64) AS total_revenue_after_discount
    ,SUM(t2.quantity) AS total_units_sold
FROM `fsda-sql-01.grocery_dataset.categories` AS t1
INNER JOIN `fsda-sql-01.grocery_dataset.products` AS t3
    ON t1.categoryid = t3.categoryid
INNER JOIN `fsda-sql-01.grocery_dataset.sales` AS t2
    ON t3.productid = t2.productid
GROUP BY categoryName
ORDER BY total_revenue_after_discount DESC
LIMIT 5;
```

Row	categoryName	total_revenue_after_discount	total_units_sold
1	Confections	556930717	11078474
2	Meat	492888845	9719292
3	Poultry	440025565	9159847
4	Cereals	427393432	8735296
5	Snails	372084885	7199409

**Confections & Meat** are the main revenue drivers (cash cows), where high sales volume directly drives high net revenue.

**Snails shows a pricing advantage**, generating higher revenue than Produce despite lower volume, indicating strong margin potential.

# Relationship Between Revenue After Discount and The Number of Unique Customers

```
SELECT
    categoryName
    ,CAST(ROUND(SUM(t2.quantity * t3.price * (1 - t2.discount)), 0)
AS INT64) AS total_revenue_after_discount
    ,COUNT(DISTINCT t2.customerid) AS number_of_unique_customers
FROM `fsda-sql-01.grocery_dataset.categories` AS t1
INNER JOIN `fsda-sql-01.grocery_dataset.products` AS t3
    ON t1.categoryid = t3.categoryid
INNER JOIN `fsda-sql-01.grocery_dataset.sales` AS t2
    ON t3.productid = t2.productid
GROUP BY categoryName
ORDER BY total_revenue_after_discount DESC
LIMIT 5;
```

Row	categoryName	total_revenue_after_discount	number_of_unique_customers
1	Confections	556930717	98743
2	Meat	492888845	98701
3	Poultry	440025565	98679
4	Cereals	427393432	98651
5	Snails	372084885	98376

**Confections** leads in revenue and customer base, showing strong demand and consistent sales performance.

**Meat, Poultry, and Cereals** also deliver high revenue and wide customer reach, making them key strategic categories.

# Calculate the average unit price for each product category

```
SELECT
    t1.categoryName
    , ROUND(AVG(t3.price), 2) AS average_price_per_unit
FROM
    `fsda-sql-01.grocery_dataset.categories` AS t1
INNER JOIN
    `fsda-sql-01.grocery_dataset.products` AS t3
    ON t1.categoryid = t3.categoryid
GROUP BY categoryName
ORDER BY average_price_per_unit DESC
LIMIT 5;
```

Row	categoryName	average_price_per_unit
1	Grain	61.4
2	Dairy	53.56
3	Snails	53.2
4	Meat	52.27
5	Confections	51.85

The top five product categories show that Grain has the highest average price per unit, followed by Dairy, Snails, Meat, and Confections, indicating varied pricing levels among categories even within a limited view of the dataset

# Calculate the average unit price for each product category

```
WITH AvgPrice AS (
    SELECT
        t1.categoryid,
        t1.categoryName,
        ROUND(AVG(t2.price), 2) AS average_price_per_unit
    FROM `fsda-sql-01.grocery_dataset.categories` AS t1
    JOIN `fsda-sql-01.grocery_dataset.products` AS t2
        ON t1.categoryid = t2.categoryid
    GROUP BY t1.categoryid, t1.categoryName
),
UniqueCustomers AS (
    SELECT
        t2.categoryid,
        COUNT(DISTINCT t1.customerid) AS number_of_unique_customers
    FROM `fsda-sql-01.grocery_dataset.sales` AS t1
    JOIN `fsda-sql-01.grocery_dataset.products` AS t2
        ON t1.productid = t2.productid
    GROUP BY t2.categoryid
)
SELECT
    t1.categoryName,
    t1.average_price_per_unit,
    t2.number_of_unique_customers
```

```
FROM AvgPrice AS t1
INNER JOIN UniqueCustomers AS t2
    ON t1.categoryid = t2.categoryid
ORDER BY t1.average_price_per_unit DESC
limit 5;
)
```

Row	categoryName	average_price_per_unit	number_of_unique_customers
1	Grain	61.4	97335
2	Dairy	53.56	98308
3	Snails	53.2	98376
4	Meat	52.27	98701
5	Confections	51.85	98743

There is a slight inverse relationship between average unit price and unique customers, where lower-priced categories like Confections attract more buyers compared to the higher-priced Grain category. However, the customer gap is minimal, indicating low price sensitivity among consumers.

# Categories Contribute the Most to Total Revenue after discount (percentage-wise)

```
SELECT
    categoryName
    ,CAST(ROUND(SUM(t2.quantity * t3.price * (1 - t2.discount)), 0) AS INT64) AS total_revenue_after_discount
    ,CONCAT(CAST(ROUND(SUM(t2.quantity * t3.price * (1 - t2.discount)) / SUM(SUM(t2.quantity * t3.price * (1 - t2.discount))) OVER () * 100,
2) AS STRING), '%') AS revenue_contribution_percentage
FROM `fsda-sql-01.grocery_dataset.categories` AS t1
INNER JOIN `fsda-sql-01.grocery_dataset.products` AS t3
    ON t1.categoryid = t3.categoryid
INNER JOIN `fsda-sql-01.grocery_dataset.sales` AS t2
    ON t3.productid = t2.productid
WHERE
    t2.quantity > 0
    AND t3.price > 0
    AND t2.discount BETWEEN 0 AND 1
    AND t2.SalesDate IS NOT NULL
GROUP BY categoryName
ORDER BY total_revenue_after_discount DESC
LIMIT 5;
```

Row	categoryName	total_revenue_after_discount	revenue_contribution_percentage
1	Confections	556930717	12.87%
2	Meat	492888845	11.39%
3	Poultry	440025565	10.17%
4	Cereals	427393432	9.88%
5	Snails	372084885	8.6%

**Confections** contributes the largest share of total revenue (12.87%), followed by **Meat (11.39%)**, **Poultry (10.17%)**, **Cereals (9.88%)**, and **Snails (8.60%)**.

Together, these top five categories generate over **52.91% of total revenue**, confirming them as the company's primary revenue drivers.

# The product categories with the highest repeat-purchase rate

```
WITH CustomerRepeatIdentifier AS (
    SELECT
        categoryName,
        customerid,
        COUNT(t1.salesid) AS transaction_count
    FROM `fsda-sql-01.grocery_dataset.sales` AS t1
    INNER JOIN `fsda-sql-01.grocery_dataset.products` AS t2 ON
t1.productid = t2.productid
    INNER JOIN `fsda-sql-01.grocery_dataset.categories` AS t3 ON
t2.categoryid = t3.categoryid
    GROUP BY 1, 2
),
CategoryRPRMetrics AS (
    SELECT
        cri.categoryName,
        COUNT(DISTINCT cri.customerid) AS total_unique_customers,
        COUNTIF(cri.transaction_count >= 2) AS repeat_buyers_count
    FROM
        CustomerRepeatIdentifier AS cri
    GROUP BY
        cri.categoryName
)
```

```
SELECT
    categoryName,
    total_unique_customers,
    repeat_buyers_count,
    CONCAT(CAST(ROUND((CAST(repeat_buyers_count AS FLOAT64) /
total_unique_customers) * 100 , 2) AS STRING), '%') AS
repeat_purchase_rate_percentage
FROM CategoryRPRMetrics
ORDER BY CAST(REPLACE(repeat_purchase_rate_percentage, '%', '') AS
FLOAT64) DESC
limit 5;
```

Row	categoryName	total_unique_customers	repeat_buyers_count	repeat_purchase_rate_percentage
1	Confections	98743	98598	99.85%
2	Meat	98701	98318	99.61%
3	Poultry	98679	98122	99.44%
4	Cereals	98651	97867	99.21%
5	Produce	98601	97550	98.93%

**Confections** has the highest repeat purchase rate (99.85%), showing extremely strong customer loyalty.

**Meat, Poultry, Cereals, and Produce** also demonstrate high retention, making these categories ideal targets for focused retention and revenue growth strategies.

# Cumulative Transaction Amount Top User

```
WITH CustomerRevenue AS (
    SELECT
        customerid,
        SUM(price * quantity * (1 - discount)) AS
    total_customer_revenue
    FROM `fsda-sql-01.grocery_dataset.sales` AS t1
    JOIN `fsda-sql-01.grocery_dataset.products` AS t2 ON
t1.productid = t2.productid
    GROUP BY customerid
),
RankedCustomerRevenue AS (
    SELECT
        customerid,
        total_customer_revenue,
        RANK() OVER (ORDER BY total_customer_revenue DESC) AS
    revenue_rank,
        SUM(total_customer_revenue) OVER (ORDER BY
total_customer_revenue DESC) AS cumulative_revenue
    FROM CustomerRevenue
)
SELECT
    customerid,
    total_customer_revenue,
```

```
        revenue_rank,
        cumulative_revenue
FROM
    RankedCustomerRevenue
WHERE
    revenue_rank = 1
ORDER BY
    revenue_rank;
```

Row	customerid	total_customer_r...	revenue_rank	cumulative_revenue
1	94800	130324.1589599...	1	130324.1589599997

**Customer ID 94800** is the top customer, generating the highest total revenue (130,324.16), making them the most valuable buyer in the dataset.

This customer should be prioritized with loyalty programs, premium engagement, and exclusive offers to sustain and increase future revenue contribution.

# Overall Sales Performance Summary

- **Confections** is the top-performing category ( $\approx 551M$  revenue, **12.87% share, 99.85% repeat rate**), making it the most profitable and strategically critical segment. Together with **Meat, Poultry, and Cereals**, these core categories contribute over 50% of total revenue, supported by strong repeat purchasing.
- Revenue is strongly driven by **repeat-purchase behavior**, while similar unique customer counts across categories suggest relatively **price-inelastic demand**. **Grain** shows a pricing anomaly (highest average price but low volume), whereas **Snails** demonstrates premium potential by generating strong revenue despite lower volume.
- **Customer ID 94800** is the highest-value buyer (**130,324.16** total revenue), highlighting the importance of premium retention and loyalty strategies to maximize lifetime value.

# Business Implications (Summary)

- **Dual Revenue Structure**  
Revenue is driven by both **high-volume categories** (Confections, Meat) and **high-margin products** (Snails), meaning profitability depends on balancing sales scale and premium pricing.
- **Profitability Prioritization**  
Snails' high Average Unit Price (AUP) shows strong margin potential, indicating growth should focus not only on volume expansion but also on maximizing high-margin categories.
- **Customer Loyalty Efficiency**  
Higher Repeat Purchase Rate (RPR) proves returning customers are more cost-efficient, so strengthening retention and increasing basket size in high-RPR categories is more profitable than aggressive new customer acquisition.

# BUSINESS RECOMMENDATIONS (OBIPR)

Objective Issue	Business impact	Isolation	Prioritization	Recommendation
Profit Maximization	A substantial increase in Net Profit Margin achieved without relying solely on large volume increases.	The Snails category exhibits a very high Average Unit Price (AUP), pointing to superior Profit Per Unit (Insight). This revenue is volume independent (Behavior).	VERY HIGH	Monetize Snails' Margin: Maintain the high AUP and immediately develop an <i>upselling</i> strategy focused on the premium value and quality of these high-margin products.
Loyalty Efficiency	Reduction in Customer Acquisition Cost (CAC) and a clear uplift in Average Order Value (AOV) from existing loyal customers.	Key categories (e.g., Confections/Meat) show a high Repeat Purchase Rate (RPR) and high RPR spikes on specific dates (Behavior). These customers represent the most stable CLV base (Insight).	HIGH	Drive Loyal Basket Size: Launch specific incentives designed solely to encourage higher <i>basket sizes</i> among customers who already have a proven high RPR history.
Revenue Stability	Preserve the Net Profit Margin of the core revenue drivers by mitigating operational losses.	Confections & Meat are the largest revenue drivers by volume. <i>Waste</i> in these categories significantly erodes overall profit (Behavior).	MEDIUM	Operational Investment: Allocate capital for upgrading storage and <i>supply chain</i> technology to minimize product waste and ensure product quality for the high-volume core categories.

# Key Findings (Summary)

- **Dual Revenue Drivers (Profit Structure)**  
Revenue is powered by high-volume cash cows **confections, meat** and high-margin premium products (**Snails**) with significantly higher AUP.
- **Customer Loyalty as a Growth Lever**  
Strong repeat-purchase behavior increases customer lifetime value and cost efficiency, with transaction timing insights enabling more targeted marketing.
- **Profit Vulnerability in High-Volume Segments**  
Heavy reliance on volume-driven categories increases operational risk—inefficiencies or stockouts in Confections or Meat could significantly impact overall profitability.

# Thanks!



# Appendix

Photo by [Thomas Le](#) on [Unsplash](#)

Photo by [Eduardo Soares](#) on [Unsplash](#)

Photo by [Eduardo Soares](#) on [Unsplash](#)

Photo by [Mitchell Luo](#) on [Unsplash](#)

Photo by [Vlad Frolov](#) on [Unsplash](#)

Photo by [Marie-Michèle Bouchard](#) on [Unsplash](#)

Photo by [No Revisions](#) on [Unsplash](#)

Photo by [sydney Rae](#) on [Unsplash](#)

Photo by [Gabrielle Ribeiro](#) on [Unsplash](#)

Photo by [Erol Ahmed](#) on [Unsplash](#)

Photo by [Wilhelm Gunkel](#) on [Unsplash](#)

