South African Retail Sales Analysis (2008–2024)

An End-to-End Data Analytics Project Using SQL and Power BI

Analyzing long-term retail performance, economic shocks, and inflation-adjusted growth in South Africa.

Prepared by: Naledi Motaung
Tools: Excel, SQL Server, Power BI, DAX
Data Source: Statistics South Africa (Stats SA)
Year: 2025

TABLE OF CONTENTS

1.Executive Summary

This project presents a comprehensive analysis of South African retail trade sales from 2008 to 2024 using official data from Statistics South Africa (Stats SA). The objective was to uncover long-term growth trends, category performance, seasonal patterns, and the impact of inflation on retail sales.

The analysis followed a full end-to-end analytics workflow, including data cleaning and transformation, SQL-based analysis, and interactive visualization using Power BI. Key metrics such as Year-on-Year (YoY) growth, Compound Annual Growth Rate (CAGR), rolling 12-month sales, and inflation-adjusted performance were calculated to provide meaningful business insights.

The results reveal a resilient retail sector that has weathered significant disruptions, including the COVID-19 lockdowns and ongoing load shedding, while maintaining steady long-term growth. The final Power BI dashboard enables stakeholders to explore these insights dynamically across time, categories, and price types.

2.Project Overview

The goal of this project was to demonstrate practical data analytics skills by analyzing national retail sales data and translating raw figures into actionable insights. Specifically, the project aimed to:

- Analyze long-term retail sales growth trends (2008–2024)
- Identify leading and lagging retail categories
- Examine seasonal sales patterns
- Measure the impact of inflation using current vs. constant prices
- Build an interactive Power BI dashboard for exploration and storytelling

This project was designed as a **professional portfolio project**, showcasing skills in data preparation, SQL analysis, DAX modeling, and business-focused visualization.

3.Data description and Preparation

The raw dataset was sourced from Statistics South Africa (Stats SA) retail trade sales data, covering the period 2002–2025. Since the project goal was to analyze growth trends, category contributions, and seasonal patterns, the dataset required significant cleaning and restructuring before analysis.

1. Initial Cleaning in Excel

- The raw Excel file contained multiple tables and repeated headers.
- Using Power Query in Excel, the data was transformed into a tabular format with consistent fields.
- The following standardized fields were created:
- Category – retail sector (e.g., General dealers, Food & Tobacco, etc.)
- Price_Type – Current Prices / Constant Prices
- Value_Type – Actual / Seasonally Adjusted values
- Year – four-digit year
- MonthNum – numeric month (1–12)
- Month_Name – month label (January–December)
- Date – constructed as the first day of each month
- Sales_R_Millions – sales values in millions of Rands
- Invalid or redundant rows (e.g., repeated headers, footnotes, or blank values) were removed.
- Since early years (2002–2007) contained significant missing values, they were excluded to maintain data quality. The final dataset begins from 2008 onwards.

2. Saving Cleaned Dataset

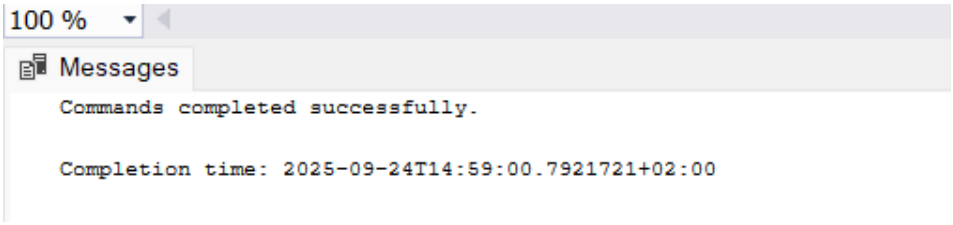
- The cleaned dataset was exported from Excel as a CSV UTF-8 (Comma delimited) file to ensure compatibility with SQL Server.
- Special care was taken to verify that:
- No extra blank rows existed at the end of the file.
- Categories with commas were properly wrapped in quotes (e.g., "Retailers of food, beverages and tobacco in specialised stores").
- Dates were in ISO format (YYYY-MM-DD), making them directly convertible to SQL DATE.

3. Import into SQL Server

A target table was pre-created in SQL Server with the following schema:

```
CREATE TABLE dbo.Retail_Sales (
    Category NVARCHAR(200),
    Price_Type NVARCHAR(50),
    Value_Type NVARCHAR(50),
    Year INT,
    MonthNum INT,
    Month_Name NVARCHAR(20),
    [Date] DATE,
    Sales_R_Millions DECIMAL(18,2)
);
```

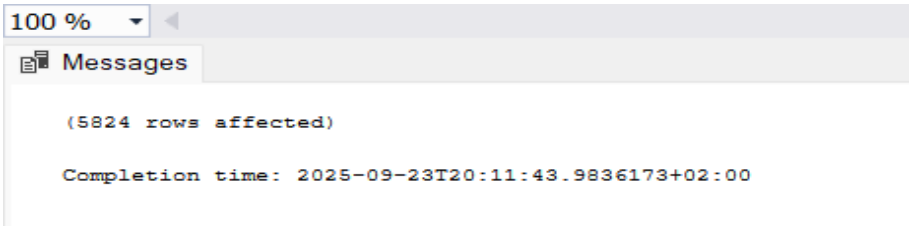
Outcome:



- **The dataset was imported into SQL Server using BULK INSERT, bypassing wizard errors caused by type conversion:**

```
BULK INSERT dbo.Retail_Sales
FROM "C:\Users\naled\Downloads\RETAIL TRADE SALES 2008 ONWARDS2.csv"
WITH (
    FORMAT = 'CSV',
    FIRSTROW = 2,              -- skip the header row
    FIELDTERMINATOR = ',',     -- CSV uses commas
    ROWTERMINATOR = '\n',      -- each line = new row
    TABLOCK
);
```

Outcome:



4. Validation

Post-import validation was performed to confirm data integrity:

- Record count matched the CSV file.
- Year range: 2008–2025.
- MonthNum values: 1–12 only.
- Sales_R_Millions contained only valid numeric values.
- Categories aligned with Stats SA retail sector definitions.

With the dataset now structured and validated, it was ready for SQL querying, analysis, and Power BI visualization.

4. SQL Data Modeling & Analysis

After data cleaning and successful import of the dataset into SQL Server, several analytical queries were executed to align with the project objectives. The analysis focused on overall sales growth, category performance, seasonal patterns, and the effect of inflation (current vs. constant prices). The following summarizes the key SQL procedures and findings.

1. Data Verification and Sanity Checks

- Before starting the analysis, SQL queries were used to confirm the integrity of the dataset.
- Row count check: Verified that all 5,825 records from the CSV were imported correctly.
- Date range: The dataset spans from January 2008 to April 2025.
- Validity of MonthNum: Confirmed that all values fall within the range 1–12.
- Missing values: No null values were found in the Sales_R_Millions column.

This confirmed that the dataset was complete and consistent.

2. Annual Sales Growth

Total yearly sales were calculated using an aggregate query:

```
SELECT Year, SUM(Sales_R_Millions) AS TotalSales
FROM dbo.Retail_Sales
WHERE Price_Type='At constant prices'
  AND Value_Type='Actual values'
  AND Year < 2025   -- exclude 2025 as incomplete year
GROUP BY Year
ORDER BY Year;
```

Findings:

- Retail sales showed a strong upward trajectory from 2008 to 2024, with only a few disruptions.
- 2020 marked a sharp contraction due to the COVID-19 pandemic and lockdown restrictions.
- 2021–2022 saw a robust rebound as restrictions lifted and pent-up demand was released.
- 2023 registered a slight dip of approximately 1.5%, reflecting the combined impact of rising interest rates, high inflation, and severe load shedding.
- 2024 returned to positive growth, reaffirming long-term resilience in the retail sector.

3. Compound Annual Growth Rate (CAGR)

The CAGR was calculated between 2008 (first complete year of the dataset) and 2024 (most recent complete year). The result was:

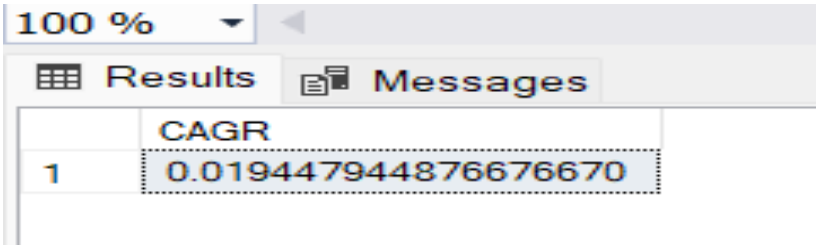
Compound Annual Growth (CAGR)=1.94%

```
DECLARE @startYear INT = 2008;
DECLARE @endYear   INT = 2024;

DECLARE @startVal DECIMAL(18,2) = (
    SELECT SUM(Sales_R_Millions)
    FROM dbo.Retail_Sales
    WHERE Year=@startYear AND Price_Type='At constant prices'
);
DECLARE @endVal DECIMAL(18,2) = (
    SELECT SUM(Sales_R_Millions)
    FROM dbo.Retail_Sales
    WHERE Year=@endYear AND Price_Type='At constant prices'
);

SELECT POWER(@endVal*1.0/@startVal, 1.0/(@endYear-@startYear)) - 1 AS
CAGR;
```

Result:



Interpretation:

- On average, retail sales in South Africa increased by ~1.9% per year over the 16-year period.
- This moderate growth highlights the resilience of the retail sector despite significant disruptions such as the COVID-19 downturn in 2020 and the economic contraction in 2023.
- Growth has been steady but not exponential, suggesting that the sector is mature and sensitive to external shocks (e.g., interest rates, energy crises).

4. Category Contribution Analysis

Category sales were aggregated to identify leading sectors:

```
SELECT Category, SUM(Sales_R_Millions) AS TotalSales
FROM dbo.Retail_Sales
WHERE Price_Type='At current prices'
  AND Value_Type='Actual values'
GROUP BY Category
      ORDER BY TotalSales DESC;
```

Results:

Results | Messages

	Category	TotalSales
1	General dealers	7504428.00
2	Retailers in textiles,clothing,footwear and leather go...	2456260.00
3	All other retailers	2166002.00
4	Retailers in hardware,paint and glass	1491804.00
5	Retailers of food, beverages and tobacco in speciali...	1321549.00
6	Retailers in pharmaceutical and medical goods,cos...	1213175.00
7	Retailers in household furniture,appliances and equ...	708938.00

Findings on retail sales by category:

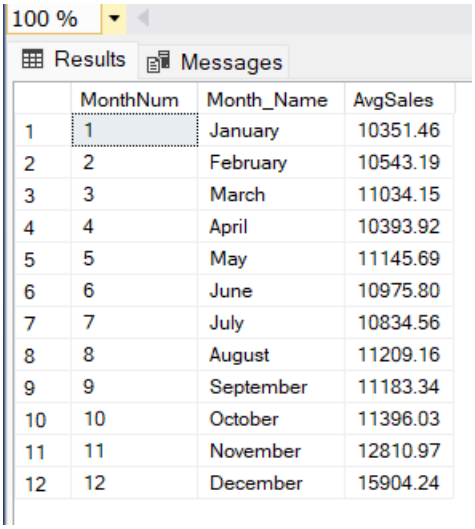
- The category of "General dealers" is the undisputed market leader, generating total sales of R 7.5 million.
- "Retailers in textiles, clothing, footwear, and leather goods" (R 2.46 million) and "All other retailers" (R 2.17 million) represent the second tier, demonstrating significant consumer demand for specialized and diversified goods.
- "Hardware, paint, and glass" (R 1.49 million), "Food, beverages, and tobacco in specialized stores" (R 1.32 million), and "Pharmaceutical and medical goods" (R 1.21 million). These categories reflect stable, essential consumer spending.
- Household furniture, appliances, and equipment" (R 0.71 million). While smaller, it still represents a substantial segment of the market, completing the long-tail distribution of retail sales.

5. Seasonality Analysis

Monthly averages across all years were computed to detect seasonal patterns:

```
SELECT
    MonthNum,
    Month_Name,
    CAST(AVG(Sales_R_Millions) AS DECIMAL(10,2)) AS AvgSales
FROM dbo.Retail_Sales
WHERE Price_Type = 'At current prices'
  AND Value_Type = 'Actual values'
  AND Year < 2025
GROUP BY MonthNum, Month_Name
ORDER BY MonthNum;
```

RESULTS:



100 %

Results | Messages

	MonthNum	Month_Name	AvgSales
1	1	January	10351.46
2	2	February	10543.19
3	3	March	11034.15
4	4	April	10393.92
5	5	May	11145.69
6	6	June	10975.80
7	7	July	10834.56
8	8	August	11209.16
9	9	September	11183.34
10	10	October	11396.03
11	11	November	12810.97
12	12	December	15904.24

Key Findings: Monthly Retail Sales Patterns

- December is the peak sales month by a substantial margin, averaging R15.90 billion - approximately 53% higher than the annual average and 24% above November's strong performance.
- November marks the beginning of the holiday surge, with average sales of R12.81 billion, indicating consumers start their festive shopping early in the penultimate month.
- A clear quarterly pattern emerges with Q4 (October-December) dominating at approximately R13.37 billion monthly average, while other quarters show more stable but lower performance around R10.8-R11.1 billion.
- January is consistently the weakest month at R10.35 billion, reflecting the typical post-holiday spending contraction as consumers recover from December expenditures.
- Sales demonstrate a general upward trajectory through the year, starting lower in Q1, stabilizing in Q2-Q3, and building momentum toward the year-end peak.

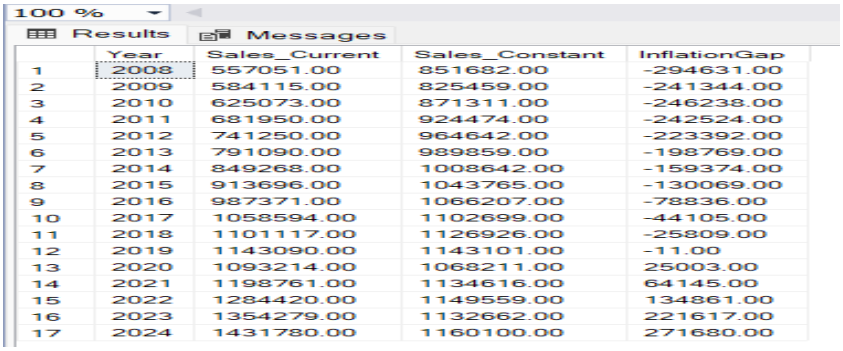
- The April dip (R10.39 billion) following strong March performance suggests potential Easter timing variations or tax-year-end effects temporarily impacting consumer spending patterns.
- The holiday season (October-December) disproportionately drives retail performance, contributing significantly to annual revenues despite representing only 25% of the calendar year.
- Month-to-month volatility is generally moderate except for the December surge, indicating relatively stable consumer spending patterns throughout most of the year with concentrated festive season intensity.

6. Inflation Impact (Current vs. Constant Prices)

To assess inflation effects, current-price and constant-price sales were compared annually:

```
SELECT Year,
  SUM(CASE WHEN Price_Type='At current prices' THEN Sales_R_Millions ELSE 0 END) AS
Sales_Current,
  SUM(CASE WHEN Price_Type='At constant prices' THEN Sales_R_Millions ELSE 0 END) AS
Sales_Constant,
  SUM(CASE WHEN Price_Type='At current prices' THEN Sales_R_Millions ELSE 0 END) -
  SUM(CASE WHEN Price_Type='At constant prices' THEN Sales_R_Millions ELSE 0 END) AS
InflationGap
FROM dbo.Retail_Sales
WHERE Value_Type='Actual values'
  AND Year < 2025   -- exclude incomplete year
GROUP BY Year
ORDER BY Year;
```

RESULTS:



100 %

Results | Messages

	Year	Sales_Current	Sales_Constant	InflationGap
1	2008	557051.00	851682.00	-294631.00
2	2009	584115.00	825459.00	-241344.00
3	2010	625073.00	871311.00	-246238.00
4	2011	681950.00	924474.00	-242524.00
5	2012	741250.00	964642.00	-223392.00
6	2013	791090.00	989859.00	-198769.00
7	2014	849268.00	1008642.00	-159374.00
8	2015	913696.00	1043765.00	-130069.00
9	2016	987371.00	1066207.00	-78836.00
10	2017	1058594.00	1102699.00	-44105.00
11	2018	1101117.00	1126926.00	-25809.00
12	2019	1143090.00	1143101.00	-11.00
13	2020	1093214.00	1068211.00	25003.00
14	2021	1198761.00	1134616.00	64145.00
15	2022	1284420.00	1149559.00	134861.00
16	2023	1354279.00	1132662.00	221617.00
17	2024	1431780.00	1160100.00	271680.00

Key Findings: Current vs. Constant Price Retail Sales Analysis (2008-2023)

- From 2008 to 2015, the InflationGap was consistently negative, averaging around -200,000. This shows that while nominal sales grew, inflation eroded much of the purchasing power, meaning real growth was far weaker than suggested by current-price figures.
- Between 2016 and 2020, the gap narrowed substantially, even approaching zero in 2019. This period reflects relative price stability and stronger real consumer demand.
- From 2021 onwards, the InflationGap turned positive, with nominal sales consistently outpacing inflation-adjusted sales. By 2023, the gap exceeded +200,000, underscoring the sharp inflationary pressures that boosted sales values without equivalent volume growth.
- Overall, this analysis highlights that much of the "growth" observed in recent years is inflation-driven, not necessarily a reflection of increased consumer spending power.

Power BI Dashboard

An interactive Power BI dashboard was developed to present the findings in a clear and intuitive manner. The dashboard allows users to filter data dynamically and explore trends across multiple dimensions.

Key Dashboard Features

Interactive Filters (Slicers):

Year (2008–2024, excluding 2025)

Retail Category

Price Type (Current vs Constant Prices)

Value Type (Actual vs Seasonally Adjusted)

Core Visualizations:

- Sales by Category: Horizontal bar chart showing category contributions
- Seasonal Distribution: Monthly sales patterns highlighting peak and low-demand periods
- Inflation Impact: Comparison of current-price vs constant-price sales
- Trend Analysis: Multi-line trend chart by category with rolling 12-month totals
- Growth Metrics: KPI cards for Total Sales, YoY Growth, CAGR, and Rolling 12-Month Sales

The dashboard enables users to clearly observe the effects of major economic events such as the 2020 COVID-19 lockdowns and periods of intensified load shedding.

Insights & Recommendation

Key Insights

- Long-term Growth: South African retail sales achieved a CAGR of approximately 1.9% between 2008 and 2024, indicating steady but moderate growth.
- COVID-19 Impact: A sharp contraction occurred in 2020, followed by a strong rebound in 2021 and 2022 driven by pent-up consumer demand.
- Inflation Effects: Recent growth in nominal sales is largely inflation-driven, with real (constant-price) growth lagging behind.
- Seasonality: December consistently records the highest sales, while January shows the weakest performance due to post-holiday spending contraction.
- Category Dominance: General dealers remain the largest contributors to total retail sales, followed by clothing and diversified retailers.

Recommendations

- Decision-makers should focus on real (inflation-adjusted) growth metrics rather than nominal sales alone.
- Retailers can optimize inventory and staffing strategies by leveraging strong seasonal demand patterns, particularly during Q4.
- Policymakers should consider the long-term economic drag of energy instability, as load shedding correlates with periods of reduced real growth.
- Future analyses should incorporate volume-based metrics to better distinguish price-driven growth from actual consumption increases.

Conclusion / Next Steps

This project demonstrates how structured data analysis and visualization can transform complex national datasets into clear, actionable insights. By combining SQL analysis with interactive Power BI reporting, the project provides a holistic view of South Africa's retail sector over nearly two decades.

Next Steps:

- Extend the analysis to include volume indices for deeper demand insights
- Integrate macroeconomic indicators (inflation, interest rates, GDP)
- Automate data refresh and dashboard updates
- Publish the Power BI dashboard using Power BI Service for stakeholder access