

**THE UNIVERSITY OF DANANG**

**VN-UK INSTITUTE FOR RESEARCH AND EXECUTIVE EDUCATION**



**FINAL REPORT**

A report submitted to the Department of Business and Economics

in partial fulfilment of the requirements for the Business Capstone Project

|  |  |
| --- | --- |
| **Institutional Affiliation:** | VN-UK Institute for Research and Executive Education, the University of Danang |
| **Module name:**  **Supervisor:** | Data Driven Decision Making  Dr. Tran The Vu |
| **Student Name:** | Nguyen Thi Kieu Nhu |
| **Group name:** | Profolio |

**Confidentiality Statement**

This document contains confidential information that must not be disclosed to anyone other than the instructor and the student unless authorised to do so.

**January, 2026**

Table of Contents

[**CHAPTER 1: PROBLEM DEFINITION & TEAM SETUP** 3](#_Toc218746193)

[1. **Group SWOT Analysis** 3](#_Toc218746194)

[2. **Project Overview** 4](#_Toc218746195)

[**3.** **WBS (Work Breakdown Structure)** 4](#_Toc218746196)

[**4.** **Show your GitHub or Workspace Setup** 6](#_Toc218746197)

[**CHAPTER 2: DATA UNDERSTANDING & SQL/COLAB FOUNDATION** 7](#_Toc218746198)

[**1.** **Dataset Overview Data source** 7](#_Toc218746199)

[**3.** **SQL Data Exploration** 8](#_Toc218746201)

[**4. Decision Questions (At least 3)** 9](#_Toc218746203)

[**CHAPTER 3: SQL MODELLING & DATA PREPARATION FOR POWER BI** 10](#_Toc218746209)

[1. **SQL Scripts for:** 10](#_Toc218746210)

[**2.** **Data Model Planning (Power BI)** 16](#_Toc218746211)

[**3.** **Preview of first Power BI visuals (Draft)**  18](#_Toc218746212)

[**CHAPTER 4: POWER BI DASHBOARD & AI AGENT INSIGHTS** 19](#_Toc218746213)

[**1.** **Power BI Dashboard Development** 19](#_Toc218746214)

[**2.** **Insight Extraction** 20](#_Toc218746215)

[3. **Comparison against decision questions** 21](#_Toc218746216)

[**CHAPTER 5: FINAL INSIGHTS, DECISION RECOMMENDATIONS & REPORTING** 22](#_Toc218746217)

[**1.** **Final Dashboard Presentation** 22](#_Toc218746218)

[**4.** **Final GitHub Submission** 24](#_Toc218746219)

[5. **Reflection** 24](#_Toc218746220)

# **CHAPTER 1: PROBLEM DEFINITION & TEAM SETUP**

## **Group SWOT Analysis**

|  |  |
| --- | --- |
| **Strengths**  • Basic understanding of data-driven thinking and how data supports business decisions.  • Familiar with fundamental tools such as Excel, Google Sheets, or basic analytics dashboards.  • Able to combine data insights with marketing or business context.  • High learning agility and openness to new data tools and technologies. | **Weaknesses**  • Limited hands-on experience with real-world datasets.  • Knowledge of advanced analytics tools (SQL, Python, BI tools) is still basic.  • Tendency to rely heavily on guidance or predefined frameworks.  • Data interpretation and storytelling skills need further development. |
| **Opportunities**  • Strong market demand for data-driven roles across industries.  • Abundant online courses, certifications, and AI tools to accelerate learning.  • Companies increasingly value data-supported decisions over intuition.  • Opportunities to practice through internships, case studies, and freelance projects.  • Rapid changes in data technologies require continuous upskilling. | **Threats**  • High competition from candidates with stronger technical data skills.  • Rapid changes in data technologies require continuous upskilling.  • Risk of focusing too much on tools rather than business impact.  • Data quality issues can limit the accuracy of beginner-level analyses. |

## **Project Overview**

What is your project?

The proposed project examines the record of transactions with the Maven Roasters, a hypothetical coffee shop with three branches in New York City. The data consists of the date of transaction, time, store location, and the information about the product at the level of detail.

This analysis aims to apply data-driven decision making to find out the trends of sales, trends of customer behavior and the performance of products with the aim of supporting operational and strategic business decisions.

Why did you choose this dataset/business problem?

The choice of this data is based on the fact that it is a realistic retail business scenario in which data is essential in the day to day business decision making. The coffee shops are a big user of the knowledge of time-based demand, location performance, and product profitability to realize maximum staffing, inventory, and marketing policies.

Also, the dataset is not too complex to be used by data-driven analysis beginners since it enables the exploration of:

-Time-series sales trends

- Purchasing behavior of customers.

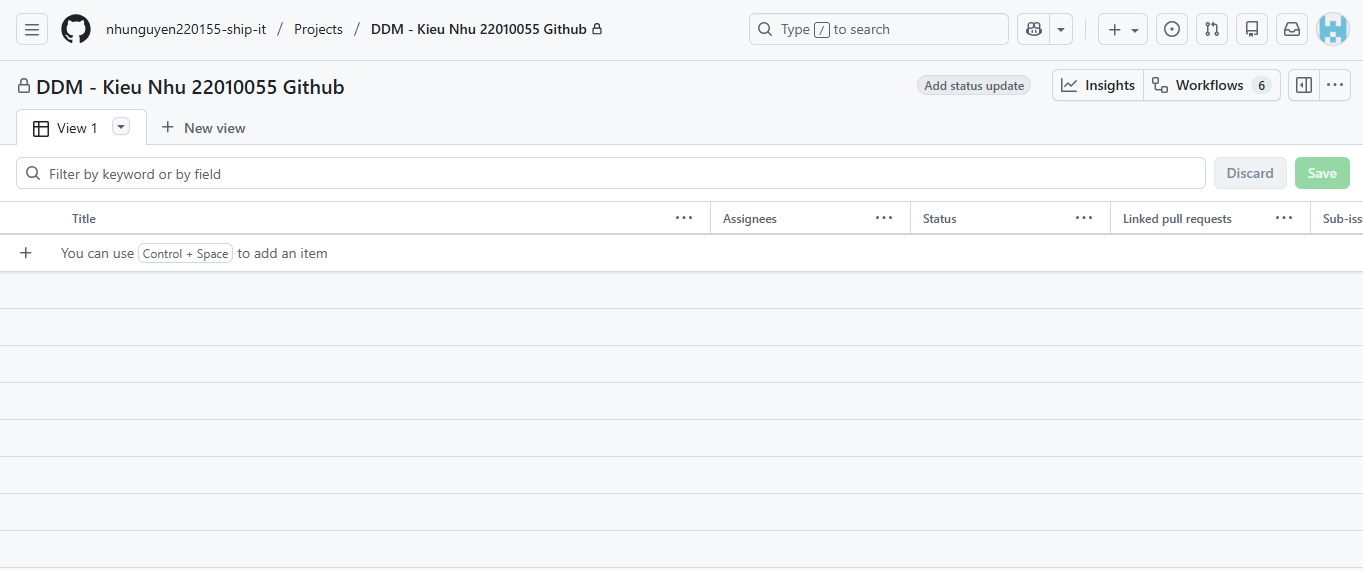
- Product-wise contributions to revenue

- Operational intelligence in various places.

## **WBS (Work Breakdown Structure)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Week** | **Phase** | **Key Activities** | **Deliverables** |
| **Week 1** | Problem Definition & Team Setup | • Define business problem & project scope  • Dataset selection & justification  • Group SWOT analysis  • Team roles & communication plan  • GitHub / Workspace setup | • Clear problem statement  • Project overview slide  • SWOT slide  • Team communication plan  • GitHub repository structure |
| **Week 2** | Data Understanding & SQL Foundation | • Dataset overview (source, schema, key fields)  • Business relevance analysis  • Data quality checks (missing values, duplicates, errors)  • SQL exploration (SELECT, WHERE, GROUP BY)  • Identify initial KPIs  • Define ≥3 decision questions | • Data understanding summary  • SQL exploration queries  • Initial KPI list  • Clear business decision questions |
| **Week 3** | SQL Modelling & Data Preparation | • Data cleaning in SQL  • Table joins & relationships  • KPI calculations in SQL  • Create views/tables for Power BI  • Power BI data model planning (Star/Snowflake schema)  • DAX & metric design planning  • Draft Power BI visuals | • Clean, modelled SQL dataset  • SQL scripts & views  • Data model diagram  • Draft Power BI visuals |
| **Week 4** | Power BI Dashboard & AI Insights | • Implement Power BI data model  • Build DAX measures  • Add filters, slicers, drill-down  • Complete dashboard visuals  • Use AI/Copilot for insight summaries  • Analyze trends, patterns, correlations | • Full draft Power BI dashboard  • Key insights & interpretations  • AI-generated narratives (optional) |
| **Week 5** | Final Insights & Reporting | • Final dashboard polishing  • KPI & trend validation  • Answer all decision questions  • Business recommendations & limitations  • Storytelling with data  • Final GitHub submission (SQL, PBIX, README)  • Team reflection | • Final dashboard  • Decision recommendations  • Storytelling presentation  • Complete GitHub submission |

## **Show your GitHub or Workspace Setup**



# **CHAPTER 2: DATA UNDERSTANDING & SQL/COLAB FOUNDATION**

## **Dataset Overview Data source**

-Data: Coffee Shop Sales Transactions.

-Basic format: Relational table (coffee\_shop\_sales)

-Granularity: There is one row per transaction item.

-Time period: Daily transactions (example of focus: April-May 2023)

## **Schema & Key fields**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| transaction\_id | INT | Unique transaction identifier |
| transaction\_date | DATE | Date of transaction |
| transaction\_time | TIME | Time of transaction |
| store\_location | VARCHAR | Store branch/location |
| product\_category | VARCHAR | Product category |
| product\_type | VARCHAR | Specific product |
| unit\_price | DECIMAL | Price per unit |
| transaction\_qty | INT | Quantity sold |

**Business relevance**  
This data allows the analysis of the retail performance of a business of a coffee shop by allowing:

* + Trend of demand (daily, hourly, weekday/weekend).
  + Comparison of products and stores performance.
  + Operational (staffing, inventory, promotions) decision-making.

1. **Initial Data Quality Check**  
    Data Cleaning & Preparation
   * Converted transaction\_date from string → DATE
   * Converted transaction\_time from string → TIME
   * Renamed corrupted column ï»¿transaction\_id → transaction\_id
   * Verified data types using DESCRIBE

Data Quality Assessment

* + No missing values in key KPI fields
  + No duplicate transaction IDs detected
  + Date & time formats standardized for time-based analysis

## **SQL Data Exploration**

## **Basic SELECT, WHERE, GROUP BY**

* SELECT, WHERE, GROUP BY, ORDER BY
* Aggregate functions: SUM, COUNT, AVG
* Window functions: LAG() for MoM analysis
* Date & time functions: MONTH(), DAY(), DAYOFWEEK(), HOUR()

**Identifying key KPIs**

|  |  |
| --- | --- |
| **KPI** | **Description** |
| Total Sales | SUM(unit\_price × transaction\_qty) |
| Total Orders | COUNT(transaction\_id) |
| Total Quantity Sold | SUM(transaction\_qty) |
| MoM Growth (%) | Month-over-Month performance |
| Average Daily Sales | Sales trend benchmark |
| Sales by Hour | Peak time identification |
| Sales by Store | Location performance |
| Sales by Product | Best & worst sellers |

## **4. Decision Questions (At least 3)**

## How have total sales, orders, and quantities changed month-over-month?

## Which days and hours generate the highest sales, and how should staffing be optimized?

## Which store locations and product categories drive the most revenue?

## Are weekend sales significantly different from weekday sales?

## Which products should be prioritized or discounted based on performance?

# **CHAPTER 3: SQL MODELLING & DATA PREPARATION FOR POWER BI**

## **SQL Scripts for:**

* + Data cleaning  
    Data cleaning

UPDATE [Coffee Shop Sales]

SET transaction\_date = CONVERT(date, transaction\_date, 105);

ALTER TABLE [Coffee Shop Sales]

ALTER COLUMN transaction\_date DATE;

UPDATE [Coffee Shop Sales]

SET transaction\_time = CONVERT(time, transaction\_time);

ALTER TABLE [Coffee Shop Sales]

ALTER COLUMN transaction\_time TIME;

* + Joins & relationships



**Relationship Structure**

* A Many-to-One (✱:1) relationship was established between the Transactions table and the Date Table.
* Each transaction record is linked to one specific date
* One date can be associated with multiple transactions

**Key Relationship Definition**

Primary link:

* Transactions[transaction\_date] → Date Table[Date]
* This relationship enables consistent time-based analysis across the entire dataset.

Filter Direction

* Single-direction filtering was applied:
* The Date Table filters the Transactions table
* This approach follows Power BI best practices by:
* Reducing ambiguity
* Improving model performance
* Ensuring predictable filtering behavior
* Join Logic Applied

Although Power BI relationships were used instead of physical SQL joins, the logical join behavior aligns with:

INNER JOIN

* Used for core analysis where only matched date records are required
* LEFT JOIN (conceptual)
* Used during validation to ensure data completeness and identify potential missing dates

**Business & Analytical Impact**

This relationship design enables: Accurate sales trend analysis over time

Consistent aggregation by:

* Product category
* Store location
* Day, month, and other time dimensions
* Seamless integration with DAX measures and Power BI visuals

**KPI calculations**

**TOTAL SALES**

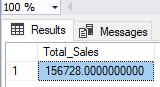
SELECT

ROUND(SUM(unit\_price \* transaction\_qty), 0) AS Total\_Sales

FROM [Coffee Shop Sales]

WHERE transaction\_date >= '2023-05-01'

AND transaction\_date < '2023-06-01';



**TOTAL SALES KPI - MOM DIFFERENCE AND MOM GROWTH**

WITH monthly\_sales AS (

SELECT

YEAR(transaction\_date) AS year,

MONTH(transaction\_date) AS month,

SUM(unit\_price \* transaction\_qty) AS total\_sales

FROM [Coffee Shop Sales]

WHERE transaction\_date >= '2023-04-01'

AND transaction\_date < '2023-06-01'

GROUP BY YEAR(transaction\_date), MONTH(transaction\_date)

)

SELECT

year,

month,

ROUND(total\_sales, 0) AS total\_sales,

ROUND(

(total\_sales - LAG(total\_sales) OVER (PARTITION BY year ORDER BY month))

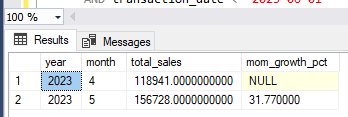
/ LAG(total\_sales) OVER (PARTITION BY year ORDER BY month) \* 100,

2

) AS mom\_growth\_pct

FROM monthly\_sales

ORDER BY month;



**TOTAL ORDERS by MONTHS**

SELECT

YEAR(transaction\_date) AS year,

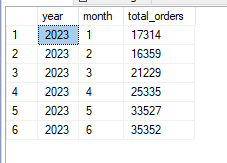
MONTH(transaction\_date) AS month,

COUNT(\*) AS total\_orders

FROM [Coffee Shop Sales]

GROUP BY YEAR(transaction\_date), MONTH(transaction\_date)

ORDER BY year, month;



**CALENDAR TABLE – DAILY SALES, QUANTITY and TOTAL ORDERS**

SELECT

transaction\_date,

ROUND(SUM(unit\_price \* transaction\_qty), 0) AS total\_sales,

SUM(transaction\_qty) AS total\_quantity\_sold,

COUNT(\*) AS total\_orders

FROM [Coffee Shop Sales]

WHERE transaction\_date = '2023-05-18'

GROUP BY transaction\_date;



**SALE TREND PERIOD**

SELECT

YEAR(transaction\_date) AS year,

MONTH(transaction\_date) AS month,

ROUND(SUM(unit\_price \* transaction\_qty), 0) AS monthly\_sales

FROM [Coffee Shop Sales]

GROUP BY YEAR(transaction\_date), MONTH(transaction\_date)

ORDER BY year, month;

WITH monthly\_sales AS (

—

SELECT

YEAR(transaction\_date) AS year,

MONTH(transaction\_date) AS month,

SUM(unit\_price \* transaction\_qty) AS total\_sales

FROM [Coffee Shop Sales]

GROUP BY YEAR(transaction\_date), MONTH(transaction\_date)

)

SELECT

year,

month,

ROUND(total\_sales, 0) AS total\_sales,

ROUND(

(total\_sales - LAG(total\_sales) OVER (PARTITION BY year ORDER BY month))

\* 100.0

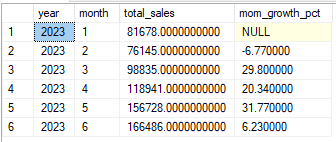
/ NULLIF(LAG(total\_sales) OVER (PARTITION BY year ORDER BY month), 0),

2

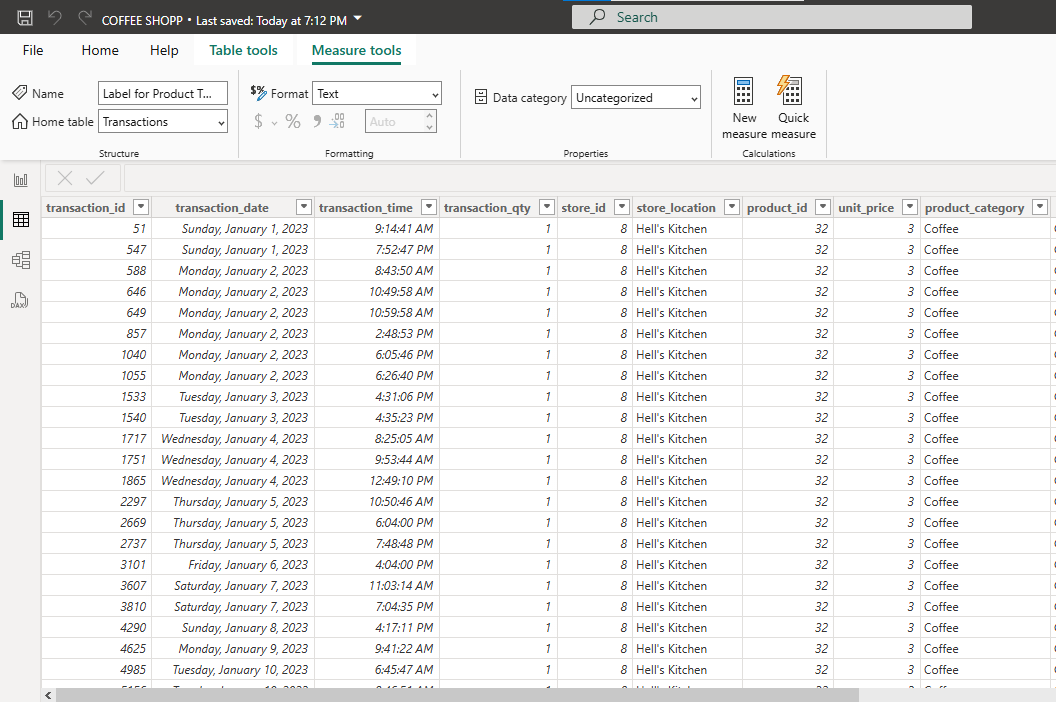
) AS mom\_growth\_pct

FROM monthly\_sales

ORDER BY year, month;



* + Views or tables for Power BI import



## **Data Model Planning (Power BI)**

Schema Design: Star Schema

The Coffee Shop Sales data model is designed using a **Star Schema**, which is widely adopted in analytical and BI systems due to its simplicity and performance efficiency.

* A **central Fact table (Fact\_Transactions)** stores quantitative transactional data.
* Multiple **Dimension tables** provide descriptive business context.
* This structure minimizes join complexity and improves query performance in Power BI.

**Primary keys, foreign keys**

Purpose

The fact table captures **measurable business events**, representing each sales transaction.

Key Fields

* Transaction\_ID (Primary Key)
* Date\_Key (Foreign Key → Dim\_Date)
* Product\_ID (Foreign Key → Dim\_Product)
* Store\_ID (Foreign Key → Dim\_Store)

Measures

* Sales Amount
* Transaction Quantity
* Order Count

Dimension Tables

Dim\_Date

* **Date\_Key** *(Primary Key)*
* Date
* Day
* Month
* Month Name
* Year

**Role**: Enables consistent time-based analysis and trend comparison.

Dim\_Product

* Product\_ID (Primary Key)
* Product Category
* Product Type
* Unit Price

Dim\_Store

**Store\_ID (Primary Key)**

**Store Location**

**Feature/Metric Design  
DAX planning**  
***DAX Planning Overview***

* DAX measures are planned before report development
* Business logic is centralized in measures
* Base measures are reused to build advanced metrics
* Calculations are dynamic and context-aware
* Time intelligence relies on a dedicated Date Table

For example:

Total Sales = SUM(Fact\_Transactions[sales\_amount])

MoM Growth (%) =

DIVIDE(

[Total Sales] - [Previous Month Sales],

[Previous Month Sales])

## **Preview of first Power BI visuals (Draft)**

# **CHAPTER 4: POWER BI DASHBOARD & AI AGENT INSIGHTS**

## **Power BI Dashboard Development**

Data model implemented

A Star Schema architecture was implemented, with Transactions serving as the central Fact table. This design supports efficient analytical querying and aligns with Power BI best practices.

The Fact\_Transactions table stores core quantitative business metrics, including:

* Sales Amount
* Transaction Quantity
* Transaction ID

To enable robust time-based analysis, a dedicated Date Dimension was integrated, providing multiple levels of temporal granularity such as Month, Week Number, and Day.

A Many-to-One (✱:1) relationship was established between:

Transactions[transaction\_date] → Date Table[Date]

Single-directional filtering was applied from the Date Dimension to the Fact table, ensuring data accuracy, eliminating ambiguity, and optimizing model performance during DAX evaluation.

**DAX measures**

Several analytical DAX measures were developed to support performance tracking and comparison:

* Month-to-Date (MTD) Sales were calculated using TOTALMTD combined with CALCULATE.
* Previous Month (PM) Sales were derived using DATEADD(…, -1, MONTH) to enable period-over-period comparison.
* Month-over-Month (MoM) Variance Metrics include:  
  - Absolute variance (Month Difference) to measure sales change in value.  
  - Percentage growth (MoM %) to evaluate relative performance changes.
* A dynamic trend indicator (▲ / ▼) was implemented using DAX variables to visually highlight positive or negative performance trends within the dashboard.

**Filters, slicers, drill-down**  
To enhance interactivity and analytical flexibility, the dashboard incorporates multiple filtering and navigation features:

* + Global slicers for Year and Month, driven by the Date Dimension.
  + Hierarchical drill-down enabled across:  
    - Month → Week → Day
  + Date granularity attributes were created using:  
    - WEEKNUM
    - FORMAT
    - Day number logic

## **Insight Extraction**

**What does the data tell us?**

The analysis reveals that sales performance is strongly driven by time-based patterns, product mix, and store location. Revenue generation is not evenly distributed across days, hours, or product categories, indicating clear opportunities for operational and strategic optimization.

Overall, the business shows positive Month-over-Month (MoM) growth, suggesting improving customer demand and effective sales performance during the analyzed period.

**Patterns, trends, correlations**

#### Time-Based Trends

* Peak sales occur during morning hours, particularly between 8:00–10:00 AM, aligning with commuter and workday coffee consumption behavior.
* Weekdays consistently outperform weekends in total sales, indicating that the business is highly dependent on regular workday traffic.
* Sales tend to be above average during mid-week (Tuesday–Thursday), while weekends show lower transaction volumes but similar average order sizes.

#### Product Performance

* A small number of top-selling product categories account for a significant share of total revenue, demonstrating a Pareto (80/20) effect.
* High-volume products are typically lower-priced items, while some higher-priced products contribute more to revenue despite lower quantities sold.

#### Store Location Impact

* Sales performance varies noticeably across store locations, with certain branches consistently generating higher revenue and order volumes.
* High-performing locations also exhibit stronger MoM growth, suggesting better customer footfall or more effective local demand.

#### Correlations Observed

#### There is a positive correlation between transaction quantity and total sales, as expected.

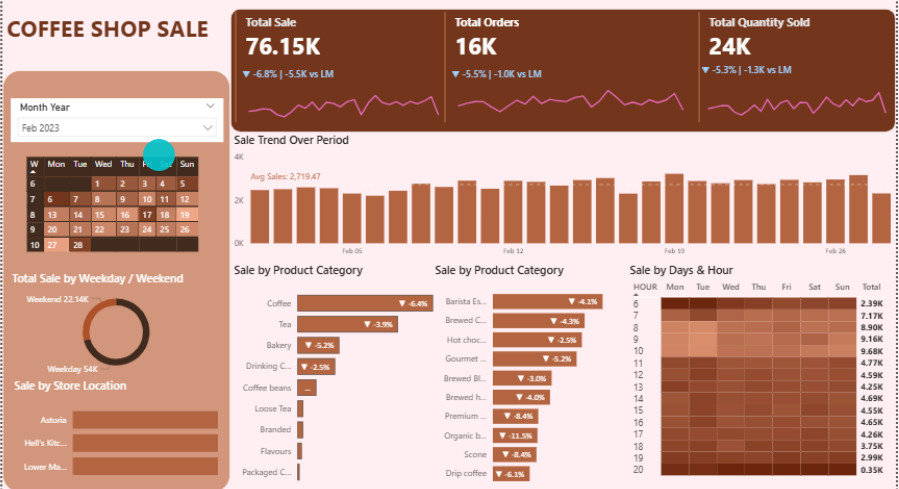
#### Peak sales hours coincide with higher order counts, rather than higher unit prices, indicating volume-driven revenue growth.

## 3. **Comparison against decision questions**

|  |  |
| --- | --- |
| Decision Question | Insight Outcome |
| How have sales, orders, and quantities changed MoM? | All key KPIs show positive MoM growth, confirming improving performance. |
| Which days and hours are busiest? | Weekday mornings (8–10 AM) are the busiest and most profitable periods. |
| Which products drive the most revenue? | A limited number of core product categories dominate revenue contribution. |
| Do weekends perform better than weekdays? | No. Weekdays generate higher sales due to workday traffic. |
| Which store locations perform best? | Performance is uneven, with specific locations consistently outperforming others. |

# **CHAPTER 5: FINAL INSIGHTS, DECISION RECOMMENDATIONS & REPORTING**

## **Final Dashboard Presentation**

****

1. **Decision Recommendations**

|  |  |  |  |
| --- | --- | --- | --- |
| Business Question | Key Findings (Data Evidence) | Business Explanation | Managerial Implication |
| How have sales trended over time? | Sales, orders, and quantity declined approximately 5–7% Month-over-Month | The decline is consistent across all KPIs, indicating a demand slowdown rather than operational inefficiency | Focus on demand-stimulation strategies (promotions, bundles, campaigns) instead of cost-cutting |
| Which days are busiest? | Weekdays contribute ~70% of total sales | Coffee consumption is strongly linked to work routines and commuting behavior | Prioritize staffing, inventory, and promotions on weekdays |
| What times of day are most popular? | Sales peak during 8–11 AM and 4–6 PM | Demand is driven by morning commutes and afternoon work breaks | Optimize staff scheduling and introduce time-based promotions |
| Does this time-based trend hold across locations? | Similar hourly demand patterns observed across all stores | Customer behavior is influenced more by daily routines than store location | Apply time-based strategies consistently across all branches |
| Which products sell the most? | Coffee, Espresso, Brewed Coffee dominate transactions | Core fresh beverages are the primary reason customers visit | Prioritize core beverage products in marketing and availability |
| Which products sell the least? | Packaged coffee, loose tea, flavored add-ons | These are low-frequency, supplementary items | Reduce inventory emphasis and avoid over-promotion of low performers |
| Which products drive the most revenue? | Core beverage categories | High transaction volume combined with steady demand | Revenue growth should focus on enhancing beverage sales performance |

1. **Storytelling with Data**

The approach of this report is based on a structured storytelling method that makes the results of the dashboard turn into obvious and usable business insights. The story takes some logical sequence of overall performance, customer behavior and then to decision-making and insights are not based on assumptions but have a visual evidence.

The story starts with the high level performance overview based on the key KPIs. The total Sales, total orders and total quantity sold has a moderate decline month over month. The fact that this decline affects all metrics with the help of KPI comparison visuals and sales trend charts show that the demand slows down in the short term, but not operational inefficiencies are a problem.

This is followed by the story of why this trend takes place. Visual representations of time, like the comparison of weekdays and weekends, and heatmaps per hour indicate that customer demand is highly routine based. The highest number of sales is observed during weekdays and also in the mornings and the late-afternoon hours. Such images support the decision-making process that is based on staffing optimization and promotional strategies that are time-related.

Subsequently, product level pictures give an understanding of where the revenue is generated. The charts with categories are very clear in demonstrating that core beverage products, which include coffee and espresso, constitute the largest share in sales volume and contribution to revenues. Retail-type products, on the other hand, have low contribution. Such visual data assists in strategic planning to focus on the main products and restrict the investment into the categories that do not give results.

Lastly, the story finishes with a linkage of knowledge to practice. All these suggestions, such as promotions, staffing, or product focus are directly accompanied by a visual representation in the dashboard. This provides the transparency of the decision making and makes the analysis more credible.

On the whole, the dashboard shows that the logical and interesting storytelling can be performed by using effective data visualization. The report goes beyond a descriptive analysis through the connection of business questions, visual evidence, and managerial decisions to help inform and make data-driven strategy.

## **Final GitHub Submission**

* + SQL scripts
  + Power BI (.pbix)
  + Documentation / README
  + AI Agent outputs (optional but encouraged)

## **Reflection**

What worked well

The dashboard was able to convert the transactional sales data to meaningful business insights. With the help of well-defined KPIs, it was possible to gain an overview of the overall performance within a short time, whereas time-based and product-level visualization helped to see the pattern of customer behaviour. Interactive capabilities facilitated logical browsing of the data and enhanced decision-making on the basis of data.

What could be improved

The limitation of the analysis lies in the lack of cost, profitability, and customer demographic data that could not be used to provide further financial and behavioral enlightenment. Moreover, trend and seasonality analysis is restricted by the relatively short time span. Predictive analytics and more detailed customer-level information could be added to the approach in the future to increase the strategic value.

Insights in the industry.

The implications can be of great use to the coffee retail and quick-service food business where demand is greatly affected by daily habits and time. Results of peak hours, weekday dominance and revenue concentration in core products have direct relations with operational decisions of staffing, promotions and product prioritization in similar businesses.