
Report

Coffee Analysis

29-09-2023

Abdullah Alessa

Overview of the Data:

The starting point is examining the data and columns, inspecting and cleaning them of any **null** values or undesirable symbols so that SQL can read them correctly. This also includes renaming columns to avoid repetition with SQL functions and changing data types, specifically for time and date columns, to ensure accurate reading.

Understanding the data and its nature, what each column contains, and the purpose of collecting data from the source is essential to form a quick and comprehensive picture.

I began by **formulating 40 questions** and writing them on an external sheet, progressing from simple to deep analytical questions. I structured these questions based on a gradient from **general to specific** and from **simple to complex** across every aspect of the analysis.

The first issue I faced was: **Where should I start?** The data is large and encompasses all aspects. One could analyze revenue, orders, product types, or categories. Every part could yield a specialized analysis without touching on other data facets. This brings us back to the fundamental idea of analysis: **you must know what is required of you as an analyst before starting the analysis.** Examples include:

- The company's sales are low.
- Quantities do not always meet customer needs; there is a shortage in certain **categories**.
- The cost of some products is burdensome, and the price is unsuitable for sale.

- The number of employees in the branches is unbalanced and needs better organization.

There is always a **main idea** from which you start, serving as a roadmap to achieve precise analysis. I disregarded this idea and relied on a **comprehensive analysis** of the general data structure.

Report Slides (Analysis Structure)

Slide 1: Presents the objectives I aim to achieve.

Slide 2: A miniature view of the data table and columns.

Slide 3: Presents the **daily working hours** for each branch, their locations, and the total number of working days, weeks, and months, showing the start and end dates of operations. This is where a problem in the data became clear: **8 weeks are missing** in the year, as follows:

- Weeks **29, 30, 34, 37, 38, 39, 41, 43**.
- July 10 to August 6 (Summer period).
- August 14 to August 28 (Summer period).
- September 4 to October 2 (Autumn).
- October 2 to October 30 (Autumn).

This indicates the following:

- July working period: 10 days.
- August working period: 17 days.
- September working period: 4 days.
- October working period: 2 days.

This means **four months** from Q3 to Q4 are missing. Despite this, it will later become clear that the **lowest revenue was in July**.

The loss of data for these weeks is confusing and significantly impacts the data, particularly concerning the variance in averages and totals. I believe its absence from the study is due to an **import error**, as it is highly unlikely the store was completely closed during these periods. Studying the neighborhoods where the branches are located and the demographics, it is unnatural for the store to be operating during these weeks without at least one invoice. Therefore, I attribute this to an error during data import or entry.

This highlights the importance of knowing the **actual number of working days** when calculating averages in the future.

Therefore:

- What is the actual number of working days and hours (Month – Week – Day – Hour)?
- How many branches does the company own, and where are they located?
- What are the daily working hours for each branch?

Slide 4:

- What is the **Total Revenue**?
- What is the **Total Orders**?
- What is the **Total Quantity**?
- What is the **Total Products**?
- What is the **Total Number of Branches**?

These comprehensive numbers are crucial for comparison in the upcoming steps and for adhering to

the core analysis idea of moving from **general to specific**.

Slide 5:

- What is the **Total Revenue, Orders, and Quantity for each branch**, and how does it compare to the overall total?
- What percentage does each **category** constitute of the total revenue, orders, and quantity?

This reveals the **best-performing branches** and the **best-performing categories**. It shows that **Store 5** and the **Packaged Chocolate category** are **underperforming** compared to their counterparts, while **Store 8** is one of the best-performing branches, and the **Coffee & Tea category** dominates the rest of the categories.

Slide 6:

- What is the **main reason for Store 5's low performance**?

A quick comparison reveals the problem was in the order quantity and revenue for the **Coffee Beans category**. To confirm this, I compared this category to the best-performing store (Store 8).

*Note: To measure performance, I relied on **Total Revenue**, as there is a clear positive correlation: as orders increase, revenue increases.*

Slide 7:

- What is the **best-performing quarter**?
- What is the **best-performing quarter for each branch**?

This highlights the issue I mentioned earlier: the **Third Quarter's performance was very low**, followed by the Fourth Quarter. The **Second Quarter was the best-performing**, which is typical given the branch locations, as this time (April-May-June) is spring and sees high tourist density.

Comparing the stores across each quarter, we observe that Q2 is still the best. We also note a slight variation between branches in Q3 and Q4, where **Store 8's revenue in the last quarter was worse** compared to the other branches.

Slide 8:

- What are the **best-performing months**?

We note that the months of Q2 were the best performing, followed by the months of Q1. The **worst-performing month was July**, followed by December.

Slide 9:

- What are the **best 3 revenue months for each branch**?
- What are the **worst 3 revenue months for each store**?

In the top three best months, all months agree. In the worst three months, there is variation between stores.

Store 8 performed well in July compared to other branches. Also, its performance was good compared to others when looking at the revenue size, which explains why **Branch 8 is superior** to the rest.

Slide 10:

- What is the **best-performing week**?

This was a very important question, specifically the one that revealed the **missing weeks in the data**. We note that the best-performing weeks are highlighted in green, and the weak weeks are in red.

Slide 11:

- What are the **best-performing days of the week** for all stores, compared to Store 8?

We note that the best day was **Monday**, the start of the week and the return to routine and work. The worst day was **Saturday**, which is natural as Saturday is the beginning of the weekend and usually a day for family at home or outings outside the city. I included the comparison for Store 8 because it is the best, but all stores can be compared to understand the nature of each branch, its patrons, and customer purchasing behavior, and to act upon the best points. (A side analytical note for clarification only).

Slide 12:

- What are the **best-performing hours** during operation?

It is clear here that **peak hours are in the morning** and decline in the evening.

- What is the **revenue volume during peak hours** for each branch?

This analysis can be used for several purposes benefiting the branches, including **balancing employees** in each branch, directing **marketing campaigns** during this time, and a **packaging strategy** (bundles/deals) that we will discuss later to boost

demand for lower-demand products. We also note that revenue during peak hours constitutes a very large proportion of the total revenue, and here, strategies can be studied to **increase peak activity throughout all hours** of the day.

Slide 13:

- What is the **average product price**?
- What is the **average quantity per order**?
- What is the **average revenue per invoice**?
- What is the **average revenue per invoice for each branch**, compared to the general average?
- What is the **average quantity per invoice for each branch**, compared to the general average?

We note that **Store 5 exceeds the general average** in terms of revenue and quantity. Herein lies the importance of this analysis, as it shows that **Store 5 sells larger quantities per single invoice**. This could be due to customer behavior in this branch (e.g., groups of friends, families, or office orders for colleagues), leading to a higher single invoice value. This point can be leveraged in **packaging strategies** or **offers**, and the experience can be transferred to other branches. Studying this behavior is crucial for replicating it in other branches.

Slides 14 – 16 – 18:

- What is the average of **Revenue - Orders - Quantities** (Quarterly – Monthly – Weekly – Daily)?
- A general overview of the average Revenue – Orders - Quantities.

Slides 15 – 17 – 19:

- What is the average of **Revenue - Orders - Quantities** (Quarterly – Monthly – Weekly – Daily) **for each branch?**
- A general overview of each branch's average.

The benefit is to form a general picture for analyzing each store and comparing it to the company's overall performance, identifying which of these stores achieves the highest percentage.

Slide 20:

- Which months had an **average (Income – Orders) better than the general average?**

This compares the average income of each month with the overall monthly average, showing which months performed better. We note that the months of Q1 and Q2 outperform the general average.

Slide 21:

- What is the **best revenue date?**
- What is the **worst revenue date?**

The comparison was made with the daily average income, and the best performance was on **June 19, 2023**. The worst day was **January 28, 2023**. This is the beginning of the year, and we notice that sales at the start of the year are initially poor due to residents returning from the holiday period and adjusting their budgets. This can be utilized by promoting products and experimenting with offers to increase revenue.

Slide 22:

- What is the **average Revenue and Orders per working day**?
- What is the **average Revenue and Orders for Branch 8**? (Any branch can be added using the same query).

Slide 23:

- What is the **average Orders per hour**?
- What is the **average Revenue per hour**?

Comparing the general average with the average of each hour, we note that the **early morning hours perform better** than the general average.

Slide 24:

- What is the **average quantity per order** from each category?

Slide 25:

- What is the **total revenue for each category in each store**?

Slide 26:

- What are the **best and worst three performing categories** in each quarter?

We note that the categories remain **constant** across all quarters in terms of the best and worst performers.

Slide 27:

- What are the **top 5 revenue categories during morning peak hours**?
- What are the **top 5 revenue categories during off-peak hours** in the evening?

Slide 28:

- In the product analysis, it became clear to me that the **Details column is the same as the Product Number**, where the number represents the details. Therefore, I avoided analyzing the details as they convey the same meaning, but one can refer to the Details column to explain the number.
- What are the **top and bottom 5 products by order volume**?
- What are the **top and bottom 5 products by revenue**?

Slide 29:

- What are the **top 3 best-selling products** in each store?
- What is the **best product** in each quarter in terms of orders and revenue?

Slide 30:

- What is the **Product of the Month** for all branches?
- What is the **Product of the Month** in the best 3 months for each branch?

Slide 31:

- What is the **best product in each quarter for Branch 3**? (The query can be applied to every store).

Slide 32:

- What is the **most expensive product**?
- What is the **cheapest product**?
- How many products have a price exceeding \$18?

Through this query, one can identify **scattered values in the averages** and exclude them so they do not affect the total value, and then perform a separate analysis for these **outliers**. This analysis can also be used for **executive decisions** regarding price review to align with customer behavior and average income—either by reducing the price or discontinuing some costly, low-demand products. Conversely, the price of some essential and high-demand products could be raised to increase income.

Slide 33:

- What are the **top 5 revenue products during peak hours**?
- What are the **top revenue products during off-peak hours**?

Slide 34:

- Dividing the day into three periods: **Morning, Noon, and Evening**.
- What are the **top 3 selling products** in each period of the day?
- What are the **bottom 3 selling products** in each period of the day?

This can be used for the **packaging strategy** and directing offers to align with each period. Here, I noted that I added the **type from each category** because some categories may perform well overall, but this specific type was not doing well.

Slide 35:

- What are the **best-performing types by revenue**?

This analysis helps in making decisions about product types, either by discontinuing them or intensifying marketing for a specific type.

Slide 36:

- What are the **top 5 most requested types** in each store?

Slide 37:

- What are the **top two most requested types** in each store during the morning and evening periods?

Slide 38:

- What are the **top 3 most requested types** in each quarter?

Slide 39:

- Dividing the year into 3 seasons: **Summer, Winter, and Transitional Phase**.
- What are the **most requested types** in each season?

This segmentation is important for identifying the most requested types, such as hot or cold beverages. It can also be used to propose a **seasonal drink** to increase sales.

Slides 40 – 41 – 42 – 43:

- What is the **most and least growing quarter**?

- What is the **most and least growing store** in each quarter?
- What are the **most and least growing months**?
- What is the **most and least growing store** in each month?

Slide 44:

- **Peak time analysis** is important for correctly **distributing employees**.
- Directing and intensifying **offers** during off-peak times to increase sales.
- It can be used to make executive decisions regarding **closing early** or **starting late** at the beginning of the day.
- **Reducing costs** related to salaries, incentives, and energy ultimately means **profit**.
- What is the **highest peak day** during a work week?

Slide 45:

- What are the **highest peak days** during the working week for each branch?

Slide 46:

- What are the **highest peak hours** for each branch?

Slide 47:

- Dividing the day into **two main shifts** according to the common practice in the labor market: the **Morning Shift** and the **Evening Shift**, and determining which one has a higher peak.
- What are the **peak times between the shifts**?

We note that **Branch 3 is balanced** in its peak activity across both periods. **Branches 5 and 8 show a clear**

variation between the two shifts, which gives an indication of customer behavior in these two branches and may be due to the branch's proximity to a work area. The difference between the two shifts reaches almost **50%**.

Conclusion (Final Geographical Analysis)

At the end of the slides, I conducted a **geographical analysis** of the population's nature and composition, as well as tracking the annual holidays, celebrations, and events in the area. This was done for **Branch 5 as an example** (and can be applied to all branches to increase sales in the coming years).

I studied the **weakest months** for this branch and presented some solutions based on the extracted data and the **demographics** surrounding this branch.

Key Points:

- A **critical missing column** in the data is the **Cost** column, which is essential for calculating **Profit**.
- The **Product** column is identical to the **Details** column, so I avoided analyzing it as it would yield the same result. The **Details** column can be referenced for clarification or explanation if needed.

End of Report.