Coffee Sales Analysis

Coffee is more than just a drink! It’s more like a ritual experience and a daily necessity. As a coffee lover, I know there are story of choices that lies behind the steaming cup and they are, the size, the roast, the blend and loyal customers who can’t start their day without it.

I have been hired as a data analyst for this coffee shop, and the mission is to dive beneath the aroma and to uncover the patterns brewing within the numbers. This data is worth brewing for success!!!!!

The dataset used for this analysis can be accessed via this link. It contains 3 sheets covering 4 years’ worth of coffee sales data across multiple countries. The key sheets in this raw datasets are:

Orders Table: Order ID, Order Date, Customer ID, Product ID, Quantity and 11null columns

Customer Table: Customer ID, Customer Names, Email, Phone Number, Address, City, Country, Postcode and Loyalty card.

Product Table: Product ID, Coffee Type, Roast Type, Size , Unit Price, Price per 100g and Profit

Tools

Microsoft Excel – Data Cleaning, Transformation, Analysis and Insight

Business Problem

The goal of this project is to explore trends in coffee sales and identify key business insights. The analysis aims specifically to:

- Find trends of total sales for each coffee type
- Identify the top 10 customers by total sales contribution
- Calculate total sales by country
- Loyalty impact on the profit
- Determine total quality ordered for each coffee type
- Build visual dashboards for insights
- Recommend the store

Methodology

Cleaning the messy data allow the coffee business to make a better data-driven decision. The dataset underwent a structured preparation process, which are **Data Cleaning, Data Merging and Data Transformation.** These were essential for correcting inconsistencies, merging related table and creating new variables for analysis.

DATA CLEANING

- **Convert data into table**
  The dataset in each sheet was converted to a table for filters to be applied to all columns for data visibility and to make it easier for sorting or reviewing the dataset while cleaning.

- **Identify Duplicates**

  To ensure consistency in the dataset, all the sheets were thoroughly examined for duplicate entries but there was no duplicate record in the sheets. Therefore, no action was required for removal.

- **Highlight Blank Cell**
  Missing data can negatively impact analysis by creating inconsistencies and incomplete insights. The customer table had approximately 204 missing values in "**email**" and 130 in "**phone number**". They were highlighted in yellow color.

- **Use Find and Replace to correct errors**

All blank cells in the **Email** and **Phone Number** columns were replaced with placeholder **"Unknown".** This method was used to retain as much customer data as possible rather than deletion for accurate report.

- **Removed Unnecessary columns**
  Some certain columns were identified as irrelevant for my analysis objectives such as (Email, Address Line, City, Phone Number and Postcode). They were all removed from the customers table to streamline the dataset and improve clarity.

DATA MERGING

- **Merged Columns**
  The customers ID from the order sheet was matched with the customers ID in the customers sheet using the XLOOKUP function and the Product ID from Order sheet was matched with the Product ID in the product sheet using XLOOKUP. The following columns were merged to the order table namely:

  Customers Sheet: Customer Name, Country, Loyalty
  Product Sheet: Coffee Type, Roast Type, Size, Unit Price and Profit

DATA TRANFORMATION

- **Date Extraction**
  The year was extracted from the **Order Date** column using YEAR () function to enable year-based analysis.

- **Expanding Coffee Type Abbreviation**
  Abbreviations in the **Coffee Type** column (**"Rob" to "Robusta")** were replaced with their full descriptive names. This transformation was implemented using the IFS () function and later converted to static values to ensure consistency.

- **Standardizing Roast Type Labels**
  The **Roast Type** column (**"L" to "Light"**) was standardized by converting single letter codes using IFS () function into their full descriptive forms to ensure uniform labelling in the dataset and enhance clarity.

- **Deriving Key Metrics**

Additional columns (**Sales and Profit per Order**) were added to the dataset to provide a foundation for analyzing revenue performance, profit contribution and identify trends.

Data visualization and Analysis

The data cleaning, merging, transformation has been done. The next step is to involve performing explanatory analysis and to create visuals to uncover meaningful insights.

- **Trend of Total Sales for Coffee Type:** Line chart was the first visual created. It illustrates the trend of totals sales for each coffee type over the 4-year period. In the **"Order date", "Year"** was extracted using the function **YEAR ()** and it was dragged into the Rows field likewise "**Coffee Type**" while the "**Sales**" was dragged into value field.

  Key Insight: *Sales across all coffee varieties grew steadily from 2019 to 2021, reflecting strong demand. However, 2022 recorded a sharp decline. On further inspection, the sales records for* ***Q4 2022 (September–December)*** *were missing, which partially explains the downturn. This highlights the importance of complete data capture for accurate performance monitoring.*

- **TOP 10 Customers by Sales Contribution:** To identify the top ten customers based on their contribution to overall sales, the Customer Name column was dragged to the row field and Sales as the value field. To show the top ten customers, a filter was first applied; however, because several customers had the same sales value as the tenth client, the filter returned fifteen consumers instead.
  The following actions were performed to precisely extract the top 10 clients:

  The **SORT()** function was applied to sort the Customer Name and Sales values in descending order based on sales.

  The **TAKE()** function was then used on the sorted data to extract only the first 10 rows, representing the top 10 customers by sales.

  The resulting dataset was copied and pasted as values to retain the top 10 list without dynamic formulas. Finally, the data was converted into a table and used to create a bar chart for clear visualization

  Key Insight: *The top 10 customers collectively contribute a significant share of total sales, with the highest individual sales recorded at $317 and the lowest within the*

group at $205. The distribution shows a comparatively small difference between the best-performing clients, indicating that this group consistently makes purchases. The gap between the highest and lowest sales values, however, points to areas where focused tactics could be used to keep top donors loyal and promote increased spending among mid-tier customers.

- **Total Sales by Country:** This chart illustrate the distribution of total sales across the coffee shop's key markets. The sales have been expressed as a percentage of the grand total allowing for a clear comparison of each country's contribution to overall revenue.

  Key Insight: *This distribution indicates that the coffee shop's market presence and customer base are heavily skewed towards the United States. The relatively small share of sales in the United Kingdom and Ireland suggests that these markets currently contribute marginally to total revenue. The visualization clearly shows a dominant market, with limited diversification across other countries.*

- **Total Quantity Ordered by Coffee Type**: Columns chart was used for this illustration. It allows for a direct comparison of customer preferences across the different coffee varieties.

  Key Insights*: The chart shows that Arabica is the most frequently ordered coffee type, reflecting a slightly stronger customer preference compared to the other varieties. Robusta, Excelsa, and Liberica display relatively similar demand patterns, indicating a fairly balanced interest in these coffee types, with no single variety dominating the market aside from Arabica.*

- **Loyalty Program Impact on Yearly Profit**: The data was visualized using a line chart converted into a lollipop chart to clearly highlight yearly profit contributions from orders with and without loyalty program participation. Profits are expressed as a percentage of total profit, providing a comparative view across years.

  Key Insights*: The lollipop chart, derived from a line chart, shows that profit contributions from loyalty program participants remain relatively stable over time, while profits from non-loyalty orders fluctuate more noticeably. Despite the loyalty program's presence, overall yearly profit varies, indicating that while the program contributes consistently, it does not appear to be the main factor influencing yearly profit growth.*

- **Key Performance Indicators:** The KPI section presents current year totals for "**Sales, Quantity and Profit**" along with their year-on-year (YoY) percentage changes. This visualization provides a quick snapshot of business performance for the current year.

  Key Insights: *All three KPIs show a significant decline in the current year. Sales and profit fell by 49%, while quantity sold decreased by 52%. These negative YoY changes indicate a contraction in overall business activity, reflecting reduced revenue and lower customer engagement compared to the previous year. However, this trend is partly influenced by the* ***missing Q4 2022 data,*** *which makes the decline appear steeper than it may be in reality.*

- **Slicers:** Year, Roast Type, Loyalty Card participation, and Coffee Size slicers allow users to filter and analyze sales, quantity, and profit by specific segments. These filters provide a dynamic and interactive view, highlighting trends over time, variations in customer preferences, and the impact of loyalty program participation and product size on business performance.

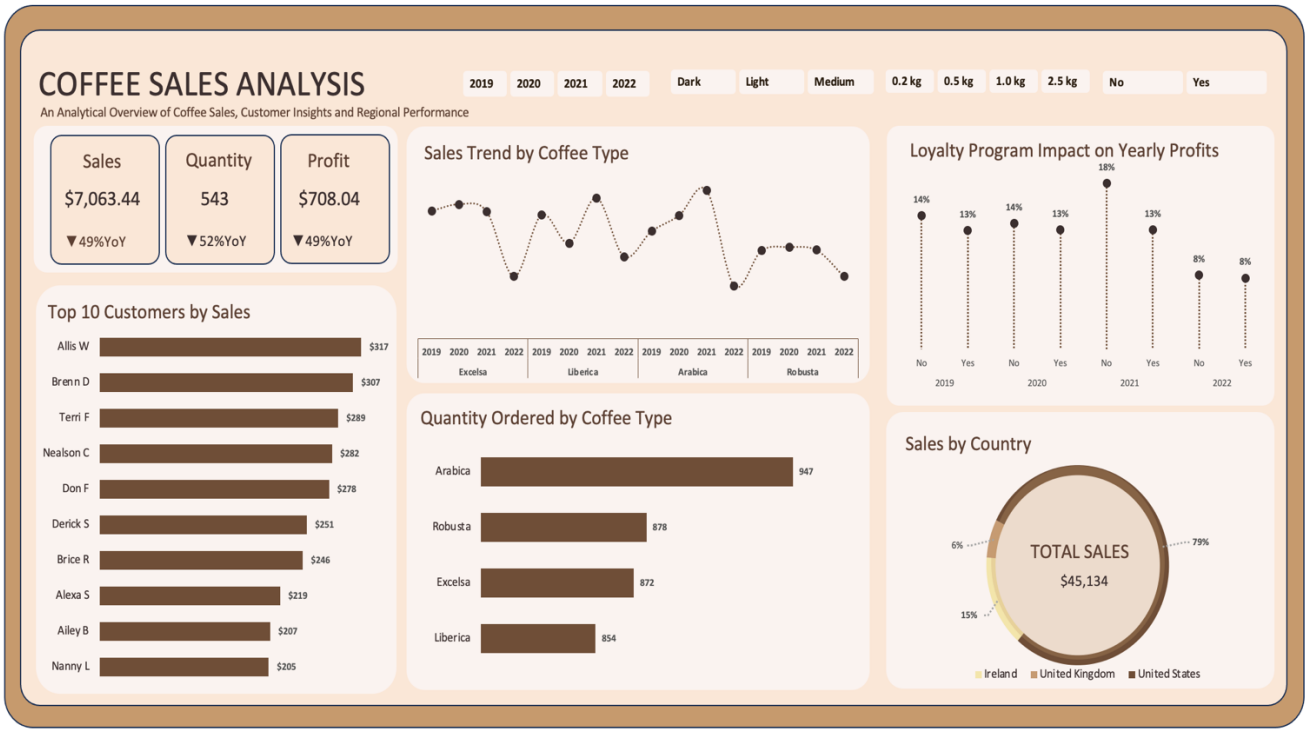

Fig 1: Coffee Shop Dashboard showing Sales, Quantity, Profit, and filters by Year, Coffee Type, Loyalty Card, and Size.

Recommendation and Implication

Based on the analysis of sales, customer behaviour, and operational KPIs, several key implications and strategic recommendations emerge for the coffee shop:

- **Sales by Country**: The United States is the company's biggest source of sales, with Ireland and the United Kingdom making only a little contribution. Although this concentration raises the possibility of market risk, it also draws attention to unexplored growth prospects in weak areas. To diversify revenue streams and lessen reliance on a single market, it is advised to increase marketing initiatives and promotional activities in Ireland and the UK.

- **Coffee Type Performance**: The most popular coffee variety is Arabica, while demand for Robusta, Excelsa, and Liberica is comparatively balanced. To maximise sales without overstocking, it is implied that marketing and inventory strategies should assist the promotion of other types while giving priority to Arabica.

- **Loyalty Program impact**: Loyalty program members generate stable and recurring revenue, but their contribution has plateaued rather than driven incremental growth. This suggests that while the program secures a reliable customer base, it lacks momentum in expanding its impact. Enhancing incentives (e.g., tiered rewards, exclusive discounts, referral bonuses) and launching targeted campaigns to convert non-loyalty customers could deepen engagement and increase lifetime value.

- **Top 10 Customers**: The top 10 customers consistently contribute a large share of total sales, reflecting loyalty and predictable purchasing behavior. However, the gap between top and mid-tier customers presents an opportunity to **increase spending among mid-tier customers** through targeted loyalty programs or personalized offers, while maintaining strong engagement with top-tier clients.

- **Sales Trends by Coffee Type**: Sales increased steadily between 2019 and 2021, demonstrating healthy demand growth. The apparent decline in 2022, however, is partly influenced by missing data for Q4 (September–December), which makes the downturn appear steeper than it may be in reality. It is recommended to validate and update the dataset to obtain a more accurate view of full-year performance. Nevertheless, management should remain proactive in analyzing potential market headwinds (e.g., competition, seasonal demand shifts) and adjust marketing, inventory, and pricing strategies accordingly.

- **Key Performance Indicators (CY with YoY)**
  Sales, quantity, and profit reflect a notable contraction in the current year. While this trend may partially result from incomplete 2022 records, it still signals the importance of strengthening customer engagement, diversifying sales channels, and improving operational efficiency. Further investigation into cost structures, pricing strategies, and regional demand variations is recommended to ensure long-term stability and growth.

Based on the overall analysis, the coffee shop should **diversify its market presence** by strengthening sales in Ireland and the UK, reducing overreliance on the U.S. market.
Enhancing **loyalty programs** and targeting mid-tier customers can drive higher engagement and spending.
Product focus should prioritize Arabica while promoting other varieties to balance sales and inventory.
Finally, addressing the recent decline in sales partly due to missing Q4 2022 data through targeted promotions, marketing, and operational improvements will be key to restoring growth and stability.

Conclusion

The analysis highlights key patterns: Arabica remains the most popular variety, sales are heavily concentrated in the U.S., and loyalty programs provide steady but modest contributions. While the top 10 customers drive significant revenue, overall sales and profit show a recent decline, partly influenced by missing Q4 2022 data. These findings emphasize the importance of market diversification, stronger customer engagement, and product optimization to sustain growth and competitiveness.