Explanation of the Dataset

I have **three datasets**: **CUSTOMERS (People who buy), ORDERS (Sales), and PRODUCTS (Items sold).** These datasets are connected through **Customer ID (who buys) and Product ID (what they buy).** The goal is to **understand customer behavior, find sales trends, and see which products are the most popular.**

What's Inside Each Dataset?

CUSTOMERS (People who buy)

- **Customer ID**: A unique number for each customer (**used to match with ORDERS**).
- **Customer Name**: The customer's name.
- **Email & Phone Number**: Contact details of the customer.
- **Address, City, Country & Postcode**: Where the customer lives (**useful for location-based sales**).
- **Loyalty Card**: Does the customer have store membership? (**useful for customer loyalty analysis**).

ORDERS (Sales data)

- **Order ID**: A unique number for each purchase.
- **Order Date**: When the purchase was made (**help track sales trends over time**).
- **Customer ID**: Who made the purchase (**connects to the CUSTOMERS dataset**).
- **Product ID**: What was bought (**connects to the PRODUCTS dataset**).
- **Quantity**: How many of those products were bought.

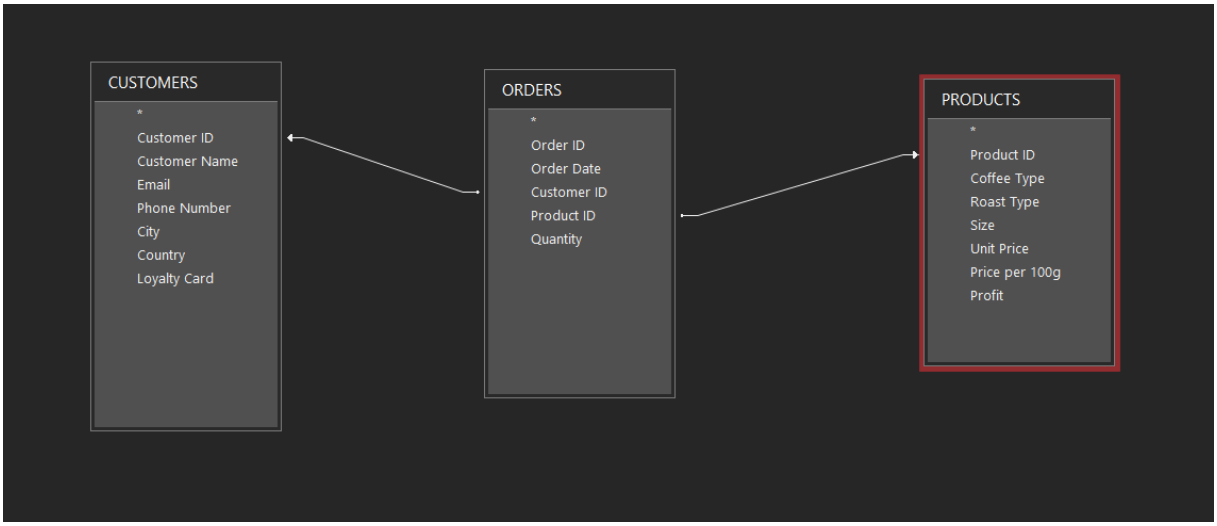
PRODUCTS (Items sold)

- **Product ID**: A unique number for each item (**used to match with ORDERS**).
- **Coffee Type**: The type of coffee (**like Arabica, Robusta, etc.**).
- **Roast Type**: How dark or light the coffee is roasted (**Light, Medium, Dark**).
- **Size**: The package size in grams.
- **Unit Price**: How much one unit of the product costs.
- **Price per 100g**: The price for 100 grams (**helps compare different sizes**).
- **Profit**: How much money does the company makes from selling one unit.

Key Analysis Questions:

1. **Which coffee types are the most popular based on total sales?**: (Identifies best-selling coffee varieties.)
2. **Which customer segments spend the most money?**: (Compares spending habits by country and loyalty card holders.)
3. **How do sales change over time?**: (Examines seasonal or monthly trends in sales.)
4. **Which countries/regions have the highest number of orders?**: (Analyzes geographical distribution of customers.)
5. **What is the percentage of coffee sales for each coffee type?**: (Assesses whether customers prefer premium or budget-friendly options.)

Preparing of Database:



I combined three datasets: **CUSTOMERS, ORDERS, and PRODUCTS**. The **ORDERS** table connects with **CUSTOMERS** using **Customer ID** (so I know who made the order) and with **PRODUCTS** using **Product ID** (so I know what was bought). I cleaned the data by fixing missing values, changing dates to the correct format, and removing unnecessary columns like full addresses. Now, I have one clean table that shows **who bought what, when, and from where**. This will help us analyze sales, customers, and product trends easily.

Explanation of Visualizations and Method Descriptions

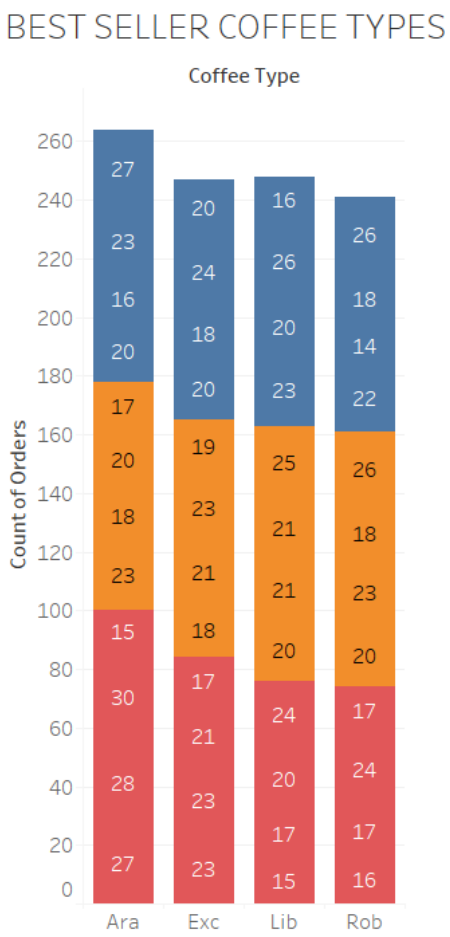


Figure 1

What is Figure 1 about?

This chart shows which coffee types are the most popular based on the number of orders. It helps the company understand which coffee types sell the most so they can manage stock better and focus on customer favorites. Within each cluster, coffee sizes increase from top to bottom.

How was it made?

Stacked Bar Chart is used to show different coffee types.
X-axis (Coffee Type): Shows different coffee types.
Y-axis (Number of Orders): Shows how many times each coffee type was ordered.
Colors: The bars are divided by Roast Type (D, L, M), so I can compare how different roast types sell.
Filters:

- Size Filter: You can select different coffee sizes.
- Roast Type Filter: You can see which roast type is more popular.

This visualization answers "Which coffee types are the most popular based on total sales?" question.

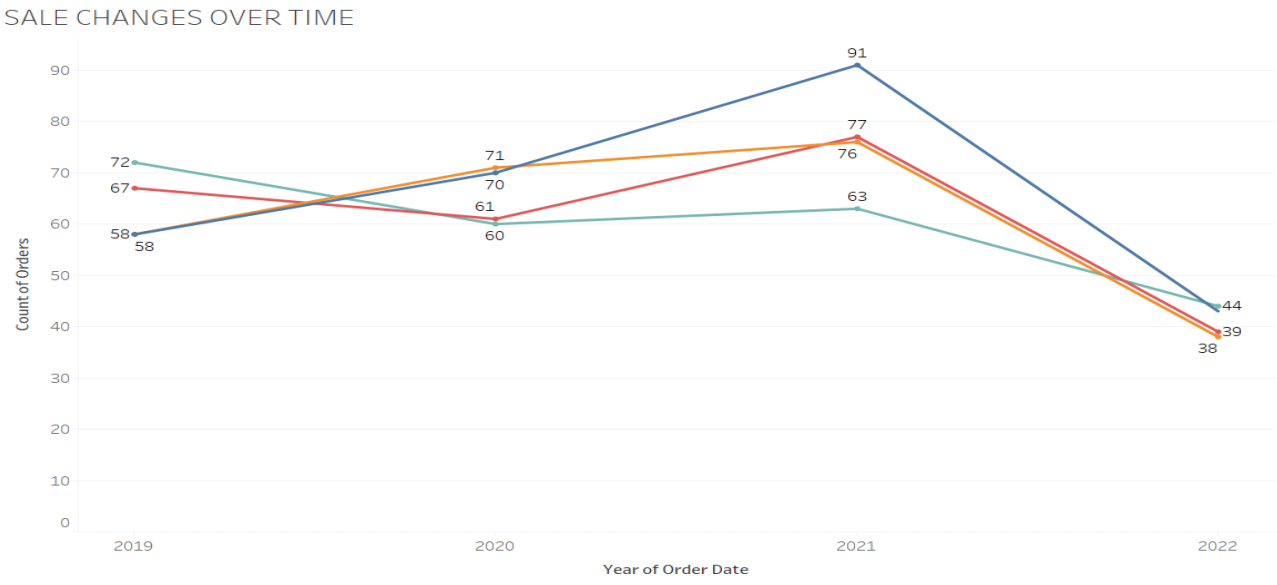


Figure 2

What is Figure 2 about?

This chart shows how coffee sales change over the years. It helps the company understand sales trends and see if sales are increasing or decreasing over time.

How was it made?

Line Chart is used to show sales trends.
X-axis (Year of Order Date): Shows the years (2019, 2020, 2021, 2022).
Y-axis (Count of Orders): Shows how many orders were placed each year.
Colors: Each coffee type has a different color to compare their sales trends.

Filters:

- **Year Filter:** You can select specific years.
- **Month Filter:** You can check monthly trends.
- **Roast Type Filter:** You can see trends for different roast types.

This visualization answers **"How do sales change over time?**" question.

Figure 3

What is Figure 3 about?

This map shows where most coffee orders come from. It helps the company understand which countries and cities buy the most coffee and focus on high-demand locations.

How was it made?

Map Visualization was used to display order locations.
Latitude & Longitude are automatically generated to position orders on the map.
Size of the circles: The bigger the circle, the more orders from that city.
Color intensity: Darker colors show locations with more orders.

Filters:

- **Year Filter:** You can select specific years.
- **Month Filter:** You can check seasonal sales trends.
- **Country & City Filter:** You can focus on specific locations.
- **Coffee Type & Roast Type Filters:** To analyze which coffee types are popular in different locations.

This visualization answers **"Which countries/regions have the highest number of orders?**" question.

Figure 4

What is Figure 4 about?

This chart shows how much customers spend in different countries and whether they have a loyalty card or not. It helps the company understand which customer groups bring the most sales and profit.

How was it made?

Bar Chart was used to compare spending.
X-axis (Country & Loyalty Card): Groups customers by country and whether they have a loyalty card.
Y-axis:

- **Top Chart:** Shows **Profit** (how much the company earns).
- **Bottom Chart:** Shows **Total Sales** (total amount spent by customers).
  Color: Each country is shown in a different color.

Filters:

- **Year Filter:** You can see spending trends by year.
- **Month Filter:** To analyze seasonal spending.
- **Loyalty Card Filter:** To compare customers with and without a loyalty card.

Custom Calculation I Created: "Total Sales"

I created **a new field called "Total Sales"** to calculate the total amount customers spent. Since the dataset didn't have this field, I created it using:

Total Sales = Quantity × Unit Price

Custom Set I Created: "Loyalty Card Holders"

I created a new set to use it as the "Loyalty Card Holders" filter. With this filter feature, I can analyze loyalty card's effect to the spending habits.

Custom Parameter I Created: "Select Metric"

I created a parameter named "Select Metric", allowing users to switch between Total Sales, Profit, and Order Count. After creating it, I selected "Show Parameter" to make it visible. Then, I created a calculated field called "Selected Metric", using a CASE statement to dynamically update values. Finally, I dragged "Selected Metric" into Columns/Rows, making the visualization interactive.

This visualization answers **"Which customer segments spend the most money?**" question.

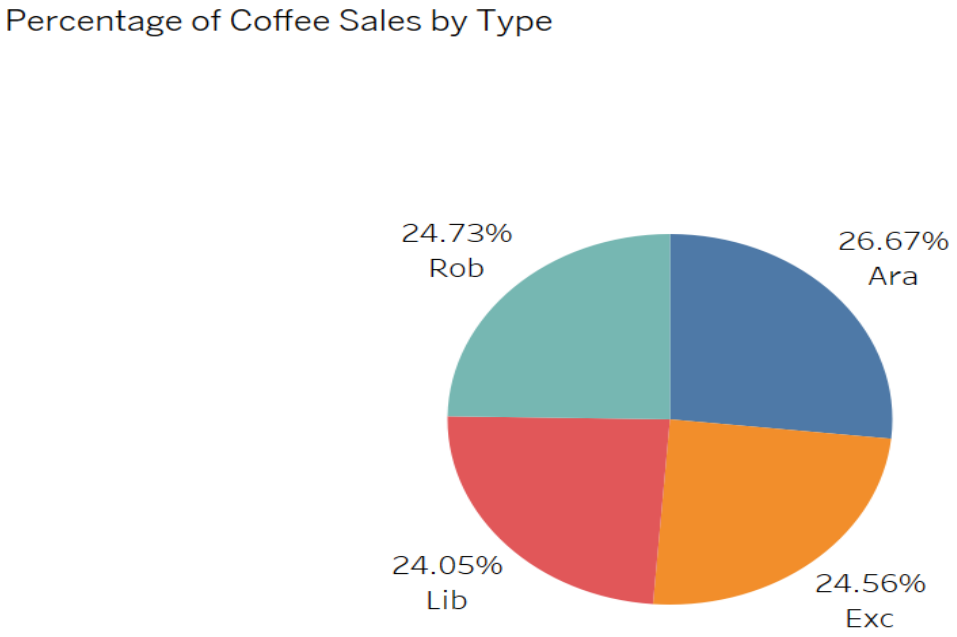


Figure 5

What is Figure 5 about?

This **pie chart** shows the **percentage of total coffee sales** for each coffee type. It helps the company see **which coffee types are more popular** compared to others.

How was it made?

Pie Chart was used to display coffee sales as percentages.
Slices of the Pie: Represent different coffee types.
Labels: Show the percentage of total sales for each coffee type.

Filters:

- **Year Filter:** Allows selecting sales data for specific years.
- **Month Filter:** Helps analyze seasonal trends.

Custom Calculation for Dynamic Percentage Updates

One important feature of this pie chart is that **the percentage updates dynamically when filters are applied**.
To achieve this, we created a calculated field:

$$\text{Percentage of Total (Filtered)} = \frac{\text{Total Sales for Coffee Type}}{\text{Total Series for Selected Filters}} x\ 100$$

Why is this important?

- If no filters are applied, the total is **100% across all years and months**.
- If a filter is applied (for instance its only 2022 sales), the percentages **recalculate based on the selected data**.
- This ensures that the total always adds up to **100%**, no matter which filters are active.

This visualization answers **"What is the percentage of coffee sales for each coffee type?**" question.