**COFFEE BEAN SALES DATA ANALYSIS**

**Step – 1: Data Overview**

Data on coffee bean sales, divided into 3 separate sheets**:**

* **Orders:** Consisting of all the information relating to the orders, i.e., Order ID (Primary Column), Order Date, Customer ID, Product ID and quantity.
* **Customers:** Consists of all the customers information, i.e., Customer ID (Primary Column), Customer Name, Email, Phone Number, Address Line 1, City, Country, Postcode, and Loyalty Card.
* **Products:** All the information about the product can be seen, i.e., Product ID (Primary Column), Coffee Type, Roast Type, Size, Unit Price, Price Per 100g and Profit.

**Step – 2: Data Gathering and Cleaning**

To create a comprehensive table encompassing all details related to product orders, and to enhance the usability of the data for subsequent analysis, various columns from the Customers and Products tables have been combined within the Orders table. These combined columns facilitate a more cohesive representation of the information pertaining to product orders, allowing for more effective analytical processes.

**Column selected are:**

|  |
| --- |
| **Customers Table** |
| Customer Name |
| Email |
| Country |

|  |
| --- |
| **Products Table** |
| Coffee Type |
| Roast Type |
| Size |
| Unit Price |

The techniques employed to retrieve data from the Customers Table involve the utilization of the VLOOKUP formula. Meanwhile, a combination of the INDEX and MATCH functions is applied for extracting information from the Products Table. This approach ensures that a uniform formula can be adapted and applied across every column within the table, enabling the extraction of data in a consistent manner.

**Column Treatment and Adjustments:**

1. **Email Column:** The missing values in the Email column were replace with an empty cell using an IF statement.
2. **Sales Column:** To populate the sales column, a multiplication between Quantity and Unit Price was done
3. **Coffee Type Name Column:** Recognizing that the "Coffee Type" column exclusively contains abbreviations such as Rob, Exc, Ara, and Lib, a supplementary column titled 'Coffee Type Name' has been established. To map the abbreviations to their respective full names, a Nested IF function has been implemented. This function systematically converts "Rob" to "Robusta", "Exc" to "Excelsa", "Ara" to "Arabica", and "Lib" to "Liberica", thereby enhancing the comprehensibility of the coffee types.
4. **Roast Type Name Column:** A similar approach was extended to the "Roast Type Name" column. Employing a comparable technique, abbreviations were replaced with their corresponding full roast type names. Specifically, "M" was transformed to "Medium", "L" denoted "Light", and "D" was indicative of "Dark".
5. **Order Date Column:** To mitigate potential confusion arising from varying date formats across different regions, a strategic adjustment was made. The month component, previously represented as a numeric value, has been transformed into a categorical value.
6. **Size Column:** The "Size" column lacked metric value notation. As a solution, the unit "kg" was uniformly appended to each row within this column.
7. **Unit Price and Sales Column:** Changed to currency and $ symbol was added.
8. **Loyal Card:** Added a new column which checks weather the customer has a loyalty card or not.

* Examined for duplicate variables across all columns in the Data Tab, and no instances of duplicate entries were detected.
* Before moving to plotting Pivot charts and tables, a step was taken to convert the entire data range into a table format named “Orders”. This transformation holds the advantage of ensuring that any future alterations in the dataset will seamlessly reflect in the pivot table and graphs, eliminating the need for manual adjustments. This dynamic synchronization enhances the efficiency and accuracy of data analysis processes.

**Step – 3: Data Analysis and Visualisation**

**Pivot Tables and Charts:**

* **Total Sales Sheet:**
  + A pivot table titled "TotalSales" was generated with the purpose of visually representing the aggregate sales across different time periods. However, an issue emerged wherein the data was initially grouped by years. To address this, a modification was made by reorganizing the data into monthly and yearly groups. This adjustment enables a more detailed visualization of sales trends over time, facilitating a finer granularity for analysis.
  + Proceeded by adding the "Coffee Type Name" and "Sales" and columns to the column list and values respectively of the pivot table. These columns will be utilized along the Y-axis for plotting against the date. This aims to provide insights into the quantity of sales recorded on specific dates, while also considering the different coffee types involved.
  + Inserted a 2D Line Chart and formatted it to my satisfaction.
  + Inserted a Timeline feature from the PivotChart Analyzer. This Timeline tool offers the capability to analyse specific segments of the timeline under consideration. By utilizing this feature, one can dynamically narrow down the timeframe of interest within the pivot chart, thereby enhancing the precision of the analysis and facilitating the exploration of particular time intervals in a more focused manner.
  + Inserted 3 Slicers, Size, Roast Name Type, and Loyalty Card.
* **Country Bar Chart Sheet:**
  + Added a new bar graph to represent the countries with the highest sales figures in descending order.
* **Top 5 Customers:**
  + Added a new bar graph and added customer names to the axis category, sorted the whole data in descending order and then applied the top 5 filter.
* **Bar Chart Sheet:**
  + Created a new sheet named “Dashboard”, where are the interactive visuals will be presented.

Lastly, established connections between all the slicers and the various visual elements present. This linkage ensures that whenever filters are applied through the slicers, the resultant changes are seamlessly reflected across all associated graphs and visual representations.

**End of Project**