**Exercise for Module 1: Introduction to Excel for Data Analysis**

Objective: Clean the provided sales data to be accurate and ready for analysis.

Instructions:

* Navigating the Excel Interface and Entering Data:
  + Format the data types for each column (e.g., date, currency).
* Calculate Basic Descriptive Statistics:
  + Calculate the mean (average), median, mode (if applicable), minimum, and maximum for both the 'Cost (USD)' and 'Quantity' columns.
  + Calculate the range of costs and quantities to understand the spread of the data.
* Analyze Supplier Ratings:
  + Calculate the average supplier rating.
  + Determine the mode of the supplier ratings to identify the most frequent rating given.
* Cost Efficiency Analysis:
  + Calculate the total cost of procurement by multiplying the cost by the quantity for each item and summing up the results.
  + Identify the item with the highest cost efficiency, defined as the lowest cost per unit (Cost/Quantity). Use small square at bottom right to view options for formatting.
* Visualization:
  + Create a column chart to visualize the distribution of supplier ratings.
  + Create a scatter plot with 'Cost (USD)' on the x-axis and 'Quantity' on the y-axis to visualize any correlation between cost and quantity procured.
* Insights and Recommendations:
  + Based on the descriptive statistics calculated, write a brief analysis discussing the efficiency of the procurement process.
  + Provide recommendations for future procurement strategies, focusing on cost efficiency and supplier selection.

**Exercise for Module 2: Data Cleaning and Preparation**

Objective: Clean the provided sales data to be accurate and ready for analysis.

Instructions:

* Missing Values:
  + Identify and fill in the missing "Sales Quantity" for Gadget C in January with the average sales quantity of Gadget C from other months.
* Data Type Corrections:
  + Correct the "Sales Revenue" entry for Widget A in February to reflect an integer value instead of text. Assume a revenue of 8500.
* Error Handling:
  + Address the negative "Sales Quantity" for Widget B in March. Replace negative values with the average quantity sold of Widget B in other months.
* Creating a Cleaned Dataset:
  + Create a new table with the corrected data, ensuring all data types are consistent (e.g., all "Sales Revenue" entries are numbers).

**Exercise for Module 3: Advanced Data Analysis Functions**

Objective: Use Excel's advanced functions to analyze the employee performance data and derive insights.

Hands-on Exercises: Analyzing Sales Data to Identify Trends

* Using Logical Functions:
* Calculate bonus eligibility: Use the IF function to determine if salespersons are eligible for a bonus (criteria: more than $5,000 in sales in a month)
  + Add a new column next to "Total Sales" and label it "Bonus Eligibility."
  + Use the IF function in the "Bonus Eligibility" column:
  + Formula: =IF(G2>5000, "Eligible", "Not Eligible") (Assuming G2 is the first cell in the "Total Sales" column).
  + Drag the fill handle down to apply this formula to all cells in the "Bonus Eligibility" column.
* Applying Lookup Functions:
  + Use VLOOKUP to find the unit price for a given product and month from a lookup table you set up in another sheet.
  + Use INDEX and MATCH to find the total sales for a specific region and month.
* Utilizing Statistical Functions:
  + Calculate total units sold across all regions using SUMIF for a specific product.
  + Determine the average sales per region using AVERAGEIF.
  + Count how many months exceeded a sales target of $3,500 using COUNTIF.
* Analyzing Data:
  + Write a brief analysis of the findings, answering questions such as:
  + Are there any trends or insights that could inform future management strategies?

Detail: **Calculating Efficiency**

1. **Add a New Column for Efficiency:**
   * Open your spreadsheet containing the employee data.
   * Add a new column titled "Efficiency."
   * In the first cell under "Efficiency" (assume it's cell C2), enter the formula to calculate the number of projects completed per hour worked. Assuming "Projects Completed" is in column A and "Hours Worked" is in column B, you would enter: =A2 / B2
   * Drag the fill handle (small square at the bottom right of the selected cell) down to copy the formula for all other employees in the list.

**Identifying Top Performers Using Conditional Formatting**

1. **Apply Conditional Formatting:**
   * Select the column that contains the "Performance Rating" data.
   * Go to the **Home** tab and click on **Conditional Formatting**.
   * Choose **Top/Bottom Rules** and then select **Top 10 Items…**.
   * Change the number from 10 to 1 to highlight the top item, or adjust according to the number of top performers you want to highlight.
   * Choose a formatting style to apply (such as a fill color) to make the top performers stand out.

**Calculate Average Performance Rating and Efficiency**

1. **Calculate Averages:**
   * Click on the cell where you want to display the average Performance Rating (say, beneath the last cell in the Performance Rating column).
   * Enter the formula to calculate the average: =AVERAGE(range)

Replace **range** with the actual range of cells in the Performance Rating column.

* + Repeat the process for the Efficiency column.

**Using Lookup Functions to Categorize Performance**

1. **Create a Lookup Table:**
   * On the same sheet or a new sheet, create a small table that maps performance ratings to descriptive categories.
   * For example, in column E rows 1 to 5, enter the numeric ratings (1 to 5). Next to each, in column F, enter the corresponding categories ("Poor", "Below Average", "Average", "Above Average", "Excellent").
2. **Apply the VLOOKUP Function:**
   * Next to each employee's performance rating, add a new column called "Rating Category."
   * In the first cell of this new column, enter the VLOOKUP formula:

=VLOOKUP(lookup\_value, table\_array, col\_index\_num, [range\_lookup])

* + - **lookup\_value** is the cell with the performance rating.
    - **table\_array** is the range of your lookup table.
    - **col\_index\_num** is the column number of the output value in the table array; for this example, it would be 2 since the categories are in the second column.
    - **[range\_lookup]** should be FALSE to ensure an exact match.
  + Copy this formula down the column to apply it to other cells.

By following these steps, you'll be able to efficiently calculate and analyze employee data, highlight top performers, and categorize performance ratings using a lookup table. These techniques are essential for managing HR metrics and can greatly aid in decision-making processes.

**Exercise for Conditional Formatting:**

Step 1: Open Your Dataset in Excel

* Ensure your dataset is formatted as a table or is neatly organized in Excel. This should include columns for Date, SalesAmount, ProductCategory, and Region as you have described.

Step 2: Apply Conditional Formatting to Highlight High and Low Sales

Select the Sales Amount Column:

* Click the header of the column that contains the sales amounts (in this case, the column labeled "SalesAmount") to select the entire column.

Go to the Conditional Formatting Menu:

* Navigate to the “Home” tab on the Excel ribbon.
* Click on “Conditional Formatting” in the Styles group.

Choose a Rule to Highlight Cell Rules:

* From the dropdown menu, select “Highlight Cells Rules.”
* Choose from options such as “Greater Than…”, “Less Than…”, or “Between…”
* For example, to highlight sales greater than $1500, choose “Greater Than…” and enter “1500” in the box. Select a formatting style (like a fill color) and click "OK."

Apply a Different Rule for Low Sales:

* Similarly, to highlight sales below $800, go back to “Highlight Cells Rules,” select “Less Than…”, and enter “800.” Choose a different color to make these stand out.

Step 3: Color-Code by Product Category

Select the Product Category Column:

* Click the header of the column that contains the product categories to select the entire column.

Use Conditional Formatting for Each Category:

* Go to “Conditional Formatting” > “New Rule…” > “Format only cells that contain.”
* Change the rule description to “Cell Value” -> “equal to” and type “Furniture” in the next box.
* Click “Format…” and choose a fill color, such as light blue. Click "OK" to apply.
* Repeat this process for “Clothing” and “Electronics” using different colors for each category.

Step 4: Apply Data Bars or Color Scales

Highlight Sales Trends Visually:

* Select the SalesAmount column again.
* Go to “Conditional Formatting” > “Data Bars” or “Color Scales” for a graphical representation of values. Data Bars are particularly useful for seeing relative sales size directly in the cells.

Step 5: Clear or Manage Rules

Adjust or Remove Conditional Formats:

* If you need to adjust or remove any rules, go to “Conditional Formatting” > “Manage Rules.”
* Here you can see all the rules applied to the workbook. You can edit, delete, or change the order of rules as needed.

**Exercise for USEFUL FUNCTIONS:**  
Functions to Practice

* INDEX MATCH:
  + Use to find the total sales of a specific transaction ID.
* IF combined with AND / OR:
  + Check if a transaction exceeds a certain amount in a specific region (e.g., Total Sales > 100 in the "North" region).
* OFFSET combined with SUM or AVERAGE:
  + Calculate the average of total sales for the first five transactions.
* CHOOSE:
  + Select between multiple columns based on an index (e.g., Choose between Product Name, Units Sold, and Total Sales).
* XNPV and XIRR:
  + Calculate the net present value and internal rate of return of cash flows from the sales transactions, assuming hypothetical cash flow dates and rates.
* SUMIF and COUNTIF:
  + Sum total sales or count transactions where certain conditions are met (e.g., all sales by a particular sales representative).
* PMT and IPMT:
  + Assume a loan amount for the total of all transactions and calculate the monthly payment and interest payment.
* LEN and TRIM:
  + Calculate the length of product names and clean any that have extra spaces.
* CONCATENATE (or CONCAT in newer versions):
  + Combine the Sales Representative’s name with the transaction ID for a unique identifier.
* CELL, LEFT, MID, and RIGHT:
  + Extract specific parts of strings like the first letter of each product name or the last two digits of a Product ID.

**Exercise for Module 4: Data Analysis Toolpak**

Dataset Description:

* CustomerID: A unique identifier for each customer.
* Age: Age of the customer.
* Income: Annual income of the customer normally distributed around $50,000 with a standard deviation of $15,000.
* Region: Geographic region of the customer.
* SalesLastYear: Sales attributed to each customer last year.
* MarketingSpend: Amount spent on marketing per customer last year.
* SupplierRating: Rating of suppliers on a scale from 1 to 5.
* ProcurementCost: Cost associated with procurement for each customer.

Instructions:

Descriptive Statistics:

* Use the dataset to calculate mean, median, mode, and standard deviation for Income, SalesLastYear, and ProcurementCost.

Regression Analysis:

* Perform linear regression to investigate the relationship between MarketingSpend and SalesLastYear, adjusting for variables like Age and Income.

ANOVA:

* Use ANOVA to compare average ProcurementCost across different regions to see if there is a statistically significant difference in costs.

**Breakdown of the ANOVA Table**

Summary Table:

* Groups: Geographic regions (East, North, South, West).
* Count: Number of observations in each group.
* Sum: Total of some measure (e.g., sales, costs) for each group.
* Average: Mean of the measure for each group.
* Variance: Variance of the measure within each group, indicating variability of data points around the mean.

ANOVA Table:

* SS (Sum of Squares): Measures the variability of the data points.
* Between Groups: Variability due to differences between group means.
* Within Groups: Variability due to differences within each group.
* df (Degrees of Freedom): Number of independent values that can vary in the data sample.
* Between Groups: Number of groups minus one (3 in this case).
* Within Groups: Total number of observations minus the number of groups (246 in this case).
* MS (Mean Square): Average of squares (SS divided by df).
* MS for Between Groups: SS Between / df Between.
* MS for Within Groups: SS Within / df Within.
* F: Ratio of MS Between to MS Within, which tests if the group means are significantly different.
* P-value: Probability that the observed F-statistic would occur if the null hypothesis were true (i.e., if there were no differences among group means).
* F crit: Critical value of F at a given significance level. If F > F crit, you reject the null hypothesis.

Interpretation:

* F-Statistic (0.160680562): Very low, indicating that the variance between the groups is not significantly greater than the variance within the groups.
* P-value (0.922708683): Much higher than the common alpha level of 0.05, indicating strong evidence that there is no statistically significant difference between the group means at the conventional 5% level. This means you fail to reject the null hypothesis.
* F crit (2.641295667): The F statistic is far less than the F critical value, further supporting the failure to reject the null hypothesis.

Conclusion:

The ANOVA analysis suggests that there are no significant differences in the averages of the measure (possibly sales or costs) across the four geographic regions (East, North, South, West). This implies that the regional factor does not significantly affect the measure being analyzed, and other variables might need to be considered to explain variations in the data.

Recommendation:

Given that the regions do not differ significantly in terms of the measure analyzed, it might be useful to look into other factors that could affect the measure, or possibly refine the grouping or measure used in the analysis. Additionally, if there are underlying assumptions of ANOVA (such as homogeneity of variances, normality of data) that have not been checked, they should be verified to ensure the validity of the ANOVA results.

**Exercise for Module 4: Data Visualization**

Hands-on Exercise: Visualizing Sales Trends Over Time

Objective:

To create a Line Chart to visualize sales trends over a year and apply advanced features to enhance understanding of the data

Instructions:

* Creating the Line Chart:
  + Enter the sales data into an Excel spreadsheet
  + Select the data range
  + Navigate to Insert > Charts > Line and choose the Line Chart that best fits your data visualization needs

A Line Chart will be created, showing sales trends over the months

* Customizing the Chart:
  + Add Chart Elements like Titles, Axis Labels, and Legends for clarity
  + Customize the chart style and color to enhance visual appeal
  + Adjust the axis scales if necessary to better display the data
* Adding a Trendline:
  + Click on the data series to select it
  + Choose Add Chart Element > Trendline > Linear
  + Customize the trendline color and style as needed
* Applying Conditional Formatting:
  + Highlight the data table
  + Navigate to Home > Conditional Formatting
  + Select a rule, for example, “Color Scales” to visually differentiate high and low sales months.
* Interpretation:
  + Analyze the chart to identify any patterns, trends, or outliers in the sales data

**Exercise for Module 6: Macro Tutorial: Automating Weekly Loan Report Preparation**

Step 1: Enabling the Developer Tab

Open Excel and go to the File menu.

Click on Options at the bottom of the sidebar.

In the Excel Options dialog box, select Customize Ribbon on the left pane.

On the right-hand side, check the Developer checkbox under the "Main Tabs" list to add it to your ribbon.

Click OK to close the dialog box and see the Developer tab in the ribbon.

Step 2: Recording the Macro

Go to the Developer tab.

Click on Record Macro.

In the Record Macro dialog box:

Type ‘weekly loan report’ in the Name field.

Optionally, add a description such as "Macro used to format the weekly loan report."

Click OK to start recording.

Step 3: Formatting the Dataset

Change the font of the entire dataset to Arial or Calibri by selecting all data and choosing the desired font from the Home tab.

Step 4: Format the table New Column

Press Ctrl + A to select all connected data.

Choose the desired table format by choosing a style from the Format as Table options.

Step 5: Format currency columns:

Select columns such as loan amount, installment, annual income, total payment, and the last payment amount.

Apply currency format by selecting the ‘$’ symbol from the Number group on the Home tab.

Step 6: Modifying Columns

Delete the homeownership column as it is not used:

Right-click on the column header and select Delete.

Step 7: Insert New Column

Insert a new column ‘Debt to Income’ ratio in Column C deriving it from

Formula: =F2/(I2/12)

Change the format to percentage using the Percent Style button in the Number group on the Home tab.

Sort data by the Debt to Income ratio:

Go to Data -> Sort.

Choose Debt to Income and set order to Largest to Smallest.

Step 8: Further Customizations

Use Find and Replace to update terminology in the loan loss column:

Press CTRL + F, select the Replace tab, type Delinquent in ‘Find what’ and Charged Off in ‘Replace with’, then click Replace All.

Apply conditional formatting:

Go to the Home tab, select Conditional Formatting -> Highlight Cell Rules -> Equal to ‘Charged Off’.

Step 9 Auto-fit columns to ensure all data is visible:

Select all columns and double-click the boundary between two column headers. Or auto fit the columns and rows, by selecting ALT-H-O-A and ALT-H-I

Step 10: Finalizing the Macro

Stop recording by clicking ‘Stop Recording’ under the Developer tab.

To view or edit the macro, click Macros, select ‘weekly loan report’, and click Edit.

In the VBA editor, add comments to your code by typing an apostrophe (') followed by your comment text.

Save your workbook as an Excel Macro-Enabled Workbook (\*.xlsm).

Step 11: Running the Macro on New Data

Open a new worksheet or tab with the new week’s data.

Go to the Developer tab, click Macros.

Select weekly loan report from the list and click Run.

**Exercise module 7: Data cleaning in Power Query and data analysis with visualization**

Let’s effectively prepare our dataset for deeper analysis and how to extract meaningful insights from it.

Part 1: Data Preparation

Step 1: Load Data

Navigate to the ‘Data’ tab on the Excel ribbon.

Click on ‘Get Data’ to expand the menu options.

Step 2: Launching Power Query Editor

From the ‘Get Data’ options, select ‘Launch the Power Query Editor’.

If this option is not directly visible, you may need to select ‘From Other Sources’ and then ‘Blank Query’, which will open the Power Query Editor.

Step 3: Choosing the Data Source

Within the Power Query Editor, navigate to the ‘Home’ tab.

Click on ‘New Source’ to open a list of available data sources.

Step 4: Selecting File Type

From the list of data sources, choose ‘File’.

Select the specific type of file you wish to import (e.g., Excel, CSV, XML, JSON, etc.).

Step 5: Navigating to the Dataset

Once you've selected the file type, a file dialog will appear.

Navigate through your system’s directories to locate the dataset you want to import.

Select the file and click ‘Open’ to import it into the Power Query Editor.

Step 6: Editing and Loading the Data

After the dataset is loaded into the Power Query Editor, you can perform any necessary data transformations or cleaning tasks.

Once your data is ready, click ‘Close & Load’ to load your transformed data into an Excel worksheet.

Step 7: Removing Duplicates:

Navigate to your dataset.

Identify the column 'orderID'.

Right-click on the column header.

Select 'Remove Duplicates'.

This step ensures we have unique records for analysis.

Step 8: Correcting Date Formats

Identify columns formatted as numbers that should be dates.

Select the column, then go to the ‘Data Type’ dropdown.

Choose ‘Date’ from the options.

This automatically converts number formats to date formats.

Step 9: Merging Columns

Select the columns for 'First Name' and 'Last Name'.

Hold the CTRL key to select both columns simultaneously.

Right-click and choose ‘Merge Columns’.

Set a separator if needed, and rename the new column to ‘Name’.

Confirm by clicking ‘OK’ to see the combined names in a new column.

Step 10: Converting Decimal to Whole Numbers

Locate the 'Units' column which is recorded in decimal format.

Select the column, navigate to ‘Transform’ tab.

Choose ‘Rounding’ -> “Rounding Up” to convert values to whole numbers.

Step 11: Formatting the Salesperson Column

Go to the ‘Transform’ tab.

Select ‘Format’, then ‘Capitalize Each Word’ to capitalize the first letters.

To remove extra spaces, return to ‘Format’ and select ‘Trim’.

Step 12: Applying Changes

Return to the Excel main interface by selecting ‘Home’.

Click ‘Close and Load’.

Your changes are applied and visible in a new tab.

Part 2: Data Analysis

Step 13: Editing Queries and Connections

Right-click on ‘Queries and Connections’ if adjustments are needed.

Select ‘Edit’ to modify any steps.

Step 14: Analyzing Revenue

Use ‘Data Analysis’ tool from the Data ribbon.

Select ‘Summary Statistics’.

Set output range and confirm to view descriptive statistics of revenue.

Step 15: Revenue by Salesperson

Insert a Pivot Table.

Drag ‘Salesperson’ to rows and ‘Revenue’ to values.

This displays revenue aggregated by each salesperson.

Step 16: Demographic Analysis

In a new Pivot Table, drag ‘Gender’ to rows and both ‘Revenue’ and ‘Units’ to values.

Adjust units to display as ‘% of Grand Total’ for better comparison.

Step 17: Using Excel’s AI Feature

Select ‘Analyze Data’ under ‘Data’.

Type ‘Revenue by Salesperson’.

Excel provides a breakdown similar to the Pivot Table.

Part 3: Visualization Techniques

Step 18: Conditional Formatting in Pivot Table

Use conditional formatting to highlight data points.

Add error bars to the Pivot Table.

Step 19: Analyzing Age and Revenue

Select columns for age and revenue.

Insert a scatter plot.

Add a trendline and display the equation to analyze correlation.

Step 20: Identifying Outliers with Boxplot

Select the revenue column.

Insert a box and whisker plot from the ‘Charts’ section.

Analyze the plot to identify outliers and extreme values.