Practice Set-I Questions (Answer Key)

1. **Assign the serial number in the products table.**

    - Insert a column Sr. No.
    - Type 1 in the first cell, e.g. A2.
    - Type formula in cell A3 >>> A2 + 1 and drag it down the column.

2. **Remove any duplicate entries based on the Email column in the customer sheet. How many unique customers remain after removing duplicates?**

    - To check the presence of duplicate values, select the column >>> Go to conditional formatting >>> Highlight cell rules >>> duplicate values
    - You will find that there are no duplicate values.
    - In case, if you find duplicate values, go to click Data >>> Data Tools >>> Remove Duplicates

3. **Identify and correct any email addresses that do not follow the standard email format (e.g., missing "@" or domain).**

    - The standard format for an email should include:
        - An "@" symbol.
        - A domain part after the "@" symbol (e.g., ".com", ".net"). Here, “.” in “.com” should appear at the minimum second position after @.

    - Insert a new column next to the email addresses column. This column will be used to indicate whether the email addresses are valid or not.
    - Use the following formula in the new column to check if the email addresses follow the standard format:

    =IF(AND(ISNUMBER(FIND("@", D2)), ISNUMBER(FIND(".", D2, FIND("@", D2) + 2))), "Valid", "Invalid")

    - In this example, replace `D2` with the cell containing the email address.
    - Here’s what each part of the formula does:
    - **FIND("@", D2):** Checks for the presence of "@" in the email address.
    - **FIND(".", D2, FIND("@", D2) + 2):** Checks for a dot (".") after the "@" symbol, ensuring there's at least one character after "@".
    - **AND(ISNUMBER(...), ISNUMBER(...)):** Ensures both conditions are met.

- **IF(..., "Valid", "Invalid"):** Labels the email as "Valid" or "Invalid" based on the conditions.

- Copy the formula down the new column to check all email addresses.
- Filter the new column to show only Invalid entries.

4. **Find the average price of products in each category. Which category has the highest average price?**

    - Insert Pivot table.
    - Drag and drop categories on the rows section.
    - Keep price in the values section.
    - Change the aggregate function from sum of values to average.

5. **(a) Increase the price of all "Widgets" by 10%.**
   (b) What is the new total price for all "Widgets"?

    - Select a new cell in any new column.

    - Add IF function. =IF(C2="Widgets",D2*1.1,D2)

    - Use SUMIF to calculate the new total price.
      =SUMIF(D2:D101,"Widgets",J2:J101)

6. **Refer sales sheet. Create a pivot table to summarize total sales (Quantity) by Region and ProductID. Which region has the highest total sales?**

    - Insert Pivot table.
    - Drag and drop ProductID on the rows section.
    - Drag and drop Region on the columns section.
    - Keep total sales in the values section.
    - Grand total at the bottom suggests that the **eastern region** has the highest total sales.

7. **Plot a line chart showing the total sales (Quantity) trend over time. Identify any months with a significant sales drop.**

    - Insert Pivot table. Select date and quantity data from the sales sheet.
    - Drag and drop Date on the rows section.
    - Deselect quarters in the pivot table so that we will have a month over the rows.

- Drag and drop Quantity on the values section.

8. **Identify the top 5 products with the highest total sales (Quantity). What are their IDs and total quantities sold?**

    - Insert Pivot table.
    - Drag and drop ProductID on the rows section.
    - Keep total sales in the values section.
    - By clicking the sort option inside the pivot table, click on more sort options.
    - Enable descending.
    - Select Sum of Quantity values for descending order.
    - By clicking the sort option again, click on top 10. Change top 10 to top 5.

9. **Create a summary table showing the total quantity of products ordered by each customer. Which customer placed the largest number of orders?**

    - Insert Pivot table using customer data.
    - Drag and drop CutomerID on the rows section.
    - Keep Quantity in the values section.
    - By clicking the sort option inside the pivot table, click on more sort options.
    - Enable descending.
    - Select Sum of Quantity values for descending order.
    - The customer with ID 93 has placed the highest number of orders, i.e., 24.

10. **Calculate the average number of days between orders for each customer. Which customer has the shortest average time between orders?**

    Step 1: Calculate number of days between the orders.
    Select the data in the order sheet.
    Sort all the entries from smallest to largest based on customerID so that we can analyze orders placed by each customer.
    Use TEXT and IF as follows:
    =TEXT(IF(B3=B2,D2," "),"dd/mm/yyyy")

    If the customerID in B3 is the same as in B2, then return the value in D2.
    Otherwise return " ".

    Step 2: List down unique customerID in L column.

    =UNIQUE(B2:B101)

    Step 3: Find the average number of days.
    Use the `AVERAGEIF` function in Excel which calculates the average (arithmetic mean) of all cells in a range that meet a given condition.

The syntax for the `AVERAGEIF` function is as follows:

AVERAGEIF(range, criteria, [average_range])

- `range`: The range of cells that you want to apply the criteria to.
- `criteria`: The condition that must be met for a cell to be included in the average.
- `[average_range]`: (Optional) The actual set of cells to average. If omitted, Excel averages the cells in the `range`.

In the formula `AVERAGEIF($B$2:B$101,L3,J$2:J$101)`, here's what each part means:

`$B$2:$B$101` contains customerID, `L3` contains the unique customerID, and `J$2:J$101` contains the difference in number of days.

- The formula will look through the range `$B$2:B$101` for cells that match the value in `L3`.
- For each matching cell, it will take the corresponding cell in `J$2:J$101` and include it in the average calculation.
- The result will be the average of the days in `J$2:J$101` for the corresponding customerID.

Customer with ID 83 has the shortest average time between orders, i.e., 1 day.

11. **Using VLOOKUP to find product prices, calculate the total value of each order. Which order has the highest total value?**

    To get price of each product, use VLOOKUP as follows:
    =VLOOKUP(C2, Product!$B$2:$E$101,4,0)

    Total value of each order (Bill amount) = price x quantity

12. **Calculate the percentage contribution of each product to the total sales (Quantity). Which product contributes the most to total sales?**

    - Use the result obtained in question 11.
    - Construct a pivot table having productID on rows and bill amount as values.
    - Change the values to percentage of grand total.
    - Sort the pivot table data in descending order according to percentage.
    - Product ID 196 contributes the most to total sales, i.e., 4.75%

13. **Apply conditional formatting to highlight sales quantities greater than 7. How many sales records are highlighted?**
    - Select the data in the quantity column of the sales sheet.
    - Apply conditional formatting as greater than 7.
    - Apply filter on quantity column.
    - Select filter by colour.
    - After filtering, the count shown in the status bar at the bottom is 24.

Practice Set-II Questions (Answer Key)

1. **Create a column as totals to showcase revenue generated over 7 months**.

   Use SUM function as follows: =SUM(C3:I3)

2. **Create an Average, Min and Max column for the revenue.**

   Use AVERAGE function as follows: =AVERAGE(C3:I3)

   Use MIN function as follows: =MIN(C3:I3)

   Use MAX function as follows: =MAX(C3:I3)

3. **Create a month over month column for the latest month.**

   Use following formula to calculate MoM: =I3/H3 -1

4. **Create a conditional format for the MoM column.**

   Conditional formatting >> Highlight cells rules >> More rules…

   Select a rule type >> Format all cells based on their values

   Format all cells based on their values >> Format style >> Data bar

   Set minimum to -1 and maximum to 1.

   Negative Value and axis >> Cell midpoint

   Set colour to green.

5. **Sort the data by Totals in descending order.**
   Select the complete table.
   Select sort and filter
   Select column of Totals

6. **Create a new column that has “above average” or “below average” text depending on the average of Total value.**
   Compute the average of all values in the total column.
   =IF(J2>P1, "Above Average", "Below Average")

   Use $ to lock the average value placed in cell P1.
   =IF(J2>$P$1, "Above Average", "Below Average")

7. **Join the distributor from the subsheet part 2. This new column should be the second column in subsheet 1 (next to movie).**

   Use VLOOKUP.
   Use =VLOOKUP(A3,'Part 2'!$A$2:$B$17,2,0)

8. **Create a pivot table showing the sum and average values for distributors.**
   Insert Pivot table.
   Keep values on columns.
   Keep rows on distributors.
   Use sum of values for values of pivot table.

9. **Create a horizontal bar chart showing the totals by Movie.**
   Drag and select the values of two columns: Movie and Totals
   Go to the insert tab.
   Click on the recommended chart.
   Select the chart type where categorical data of the movie is on Y-axis and the data of totals on X-axis.

Stretch the image horizontally to adjust.

Give the chart title “Total revenue generated by Movies”

10. **Create a vertical bar chart showing the average value by Genre.**

    Drag and select the values of pivot table: Genre vs Average of average

    Go to the insert tab.

    Click on the recommended chart.

    Select the column chart type.

    Give the chart title “Average Value by Genre”

11. **Create a pie chart showing the average value by Distributor.**

    Select the values of pivot table: Distributor vs Average of average

    Go to the insert tab.

    Click on the pie chart.

    Give the chart title “Average value by Distributor”