

Name : Aryan tyagi

Course Name : Data analytics with GenAi

Topic : Excel with copilot

Email ID : [aryantyagi5432@gmail.com](mailto:aryantyagi5432@gmail.com)

### Question1

Use Copilot to identify and summarize missing values in the database.

Answer

[Assigment 5.xlsx](#)

 Step 1: Highlight Missing Values

- Select your dataset.
- Go to **Home** → **Conditional Formatting** → **New Rule** → **Format only cells that contain** → **Blanks**.
- Apply a color fill (e.g., light red).  
This will visually highlight all missing cells.

 Step 2: Count Missing Values per Column

For each column (say Region), use the formula:

=COUNTBLANK(B2:B1000)

- Replace **B2 :B1000** with the actual range of your column.
- Do this for **Region** and **Category** to see how many blanks exist.

Formula for percentage:

=COUNTBLANK(B2:B1000)/COUNTA(B2:B1000)

 Step 3: Use Filters

- Apply **Filters** (**Data** → **Filter**).
- For each column, filter by blanks to quickly see which rows are missing values.

 Step 4: Optional Pivot Table Summary

- Insert a Pivot Table.
- Drag **Column Names** to Rows.
- Use **Count of Values** and **Count of Blanks** to get a quick summary.

 This process will give you a clear **overview of missing values across columns**.

Would you like me to show you a **ready-to-use Excel formula set** that automatically builds a missing-value summary table for all your columns in one place? That way, you won't need to check each column separately.

## Question2

**use Copilot to fill missing values in the Regions and category columns**

Answer 2

[Assignment 5.xlsx](#)

### Strategies for Filling Missing Values

#### 1. Fill Region Based on Country

- **Logic:** Each country usually belongs to a single region (e.g., "France → Europe").
- **Approach:** Look at other rows where the same country has a region filled, then copy that region into the missing cells.

#### 2. Fill Category Based on Similar Orders

- **Logic:** If the same **Order ID** or **Country** has a known category, use that to fill missing ones.
- **Approach:** Use majority or most frequent category for that country/order.

#### 3. Use Helper Columns for Consistency

- Create a reference table mapping **Country** → **Region** and **Order ID** → **Category**.
- Use lookup formulas (like **VLOOKUP** or **INDEX-MATCH**) to automatically fill missing values.

### Manual Steps in Excel

Step 1: Identify Missing Values

- Select the dataset.
- Use **Filter** → filter blanks in Region or Category columns.

## Step 2: Create a Reference Table

- Insert a new sheet.
- Build a table like: Similarly, if needed, create a table for **Order ID** → **Category**.

## Step 3: Use Lookup Formulas

- In the Region column (where blank), enter:
- =VLOOKUP([@Country], CountryRegionTable, 2, FALSE)
  - Replace **CountryRegionTable** with the actual table range.
  - This pulls the region based on country.
- For Category:

=VLOOKUP([@OrderID], OrderCategoryTable, 2, FALSE)

- This fills category based on similar orders.

## Step 4: Fill Down and Replace

- Copy the formulas down for all rows.
- Once filled, **Copy** → **Paste Values** to replace formulas with actual text.

## Step 5: Handle Multiple Matches

- If a country has multiple categories, use **Pivot Table**:
- Insert Pivot Table with **Country** as rows and **Category** as values.
- See which category is most frequent.
- Use that as the default fill.

### Extra Tips

- Use **Conditional Formatting** to highlight blanks before and after filling.
- If you want automation: Excel's **Power Query** can merge tables and fill missing values more robustly.
- For large datasets, consider using **Flash Fill** or **Data Validation** to keep consistency.

 Would you like me to also show you a **step-by-step example with sample data** (like a mini dataset and the exact formulas applied), so you can copy-paste directly into Excel?

### Question 3

Fill missing numeric values in the unitsold and unitprice columns.

Answer

[Assignment 5.xlsx](#)

Great question  — handling missing **numerical values** (like *Units Sold* and *Unit Price*) requires a different approach than categorical data, because you can't just copy text from similar rows. Let's go step by step:

#### Common Strategies for Missing Numerical Data

##### 1. Simple Imputation

- **Fill with Zero:** If missing means "no sale" or "not applicable," replace with `0`.
- **Fill with Average/Median:** Use the column's average or median to replace blanks.
- Median is often better because it's less affected by extreme values.

##### 2. Group-Based Imputation

- Calculate average/median **per Country, Region, or Category**.
- Fill missing values with the typical value for that group.  
Example: If "India → Electronics" usually has Unit Price = 50, use 50 for missing entries in that group.

##### 3. Use Previous/Next Values

- If data is time-series (e.g., monthly sales), fill missing values with the previous or next available value (called *forward fill* or *backward fill*).

##### 4. Interpolation

- For continuous data (like prices over time), use linear interpolation between known values.

##### 5. Flag Missing Values

- Add a helper column (e.g., "Price Missing?") with `1` for missing and `0` otherwise.  
This helps track which values were imputed.

#### Manual Steps in Excel

Step 1: Identify Missing Values

- Use **COUNTBLANK**:

=COUNTBLANK(C2:C1000)

- (for Units Sold column)

## Step 2: Fill with Average or Median

- Insert a cell with:
  - =AVERAGE(C2:C1000)
    - or
    - =MEDIAN(C2:C1000) Copy this value into missing cells.
- Step 3: Group-Based Filling
  - Create a **Pivot Table**:
    - Rows: Country/Category
    - Values: Average of Unit Price
  - Use these averages to fill missing values manually (or with **VLOOKUP**).
- Step 4: Interpolation (if time-based)
  - Sort data by date.
  - For missing values, use:
    - =(PreviousValue + NextValue)/2 Or use Excel's **Fill Series → Linear Trend**.
- Step 5: Document Changes
  - Add a helper column noting which values were filled and how (e.g., "Filled with median").
- ⚡ Best Practices
  - **Never blindly fill with averages** if the dataset is diverse — always check by group.
  - **Keep a record** of which method you used (important for analysis transparency).
  - **Consider the meaning of missingness**:
    - If missing means "not recorded," imputation makes sense.
    - If missing means "not applicable," use **0** or leave blank.

👉 Do you want me to prepare a **ready-to-use Excel formula set** (like **IF** + **VLOOKUP** + **AVERAGEIF**) that automatically fills missing *Units Sold* and *Unit Price* based on country/category averages? That way you won't need to do it manually row by ro

## Question4.

**Create a new calculated column total revenue**

## Answer

### [Assignment 5.xlsx](#)

Great question 👍 — handling missing **numerical values** (like *Units Sold* and *Unit Price*) requires a different approach than categorical data, because you can't just copy text from similar rows. Let's go step by step:

### ✖ Common Strategies for Missing Numerical Data

## 1. Simple Imputation

- **Fill with Zero:** If missing means "no sale" or "not applicable," replace with **0**.
- **Fill with Average/Median:** Use the column's average or median to replace blanks.
- Median is often better because it's less affected by extreme values.

## 2. Group-Based Imputation

- Calculate average/median **per Country, Region, or Category**.
- Fill missing values with the typical value for that group.  
Example: If "India → Electronics" usually has Unit Price = 50, use 50 for missing entries in that group.

## 3. Use Previous/Next Values

- If data is time-series (e.g., monthly sales), fill missing values with the previous or next available value (called *forward fill* or *backward fill*).

## 4. Interpolation

- For continuous data (like prices over time), use linear interpolation between known values.

## 5. Flag Missing Values

- Add a helper column (e.g., "Price Missing?") with **1** for missing and **0** otherwise.  
This helps track which values were imputed.



### Manual Steps in Excel

#### Step 1: Identify Missing Values

- Use **COUNTBLANK**:

=COUNTBLANK(C2:C1000)

- (for Units Sold column)

#### Step 2: Fill with Average or Median

- Insert a cell with:
  - =AVERAGE(C2:C1000)
    - or
    - =MEDIAN(C2:C1000) Copy this value into missing cells.
- Step 3: Group-Based Filling
  - Create a **Pivot Table**:
    - Rows: Country/Category
    - Values: Average of Unit Price
  - Use these averages to fill missing values manually (or with **VLOOKUP**).
- Step 4: Interpolation (if time-based)

- Sort data by date.
  - For missing values, use:
  - $=(\text{PreviousValue} + \text{NextValue})/2$  Or use Excel's **Fill Series → Linear Trend**.
- Step 5: Document Changes
  - Add a helper column noting which values were filled and how (e.g., "Filled with median").
- ⚡ Best Practices
  - **Never blindly fill with averages** if the dataset is diverse — always check by group.
  - **Keep a record** of which method you used (important for analysis transparency).
  - **Consider the meaning of missingness:**
    - If missing means "not recorded," imputation makes sense.
    - If missing means "not applicable," use **0** or leave blank.

## Question 5

Use Copilot to find the total price of Order ID = 'ORD0123'

### Answer

[Assignment 5.xlsx](#)

✖ Common Strategies for Missing Numerical Data

#### 1. Simple Imputation

- **Fill with Zero:** If missing means "no sale" or "not applicable," replace with **0**.
- **Fill with Average/Median:** Use the column's average or median to replace blanks.
- Median is often better because it's less affected by extreme values.

#### 2. Group-Based Imputation

- Calculate average/median **per Country, Region, or Category**.
- Fill missing values with the typical value for that group.  
Example: If "India → Electronics" usually has Unit Price = 50, use 50 for missing entries in that group.

#### 3. Use Previous/Next Values

- If data is time-series (e.g., monthly sales), fill missing values with the previous or next available value (called *forward fill* or *backward fill*).

## 4. Interpolation

- For continuous data (like prices over time), use linear interpolation between known values.

## 5. Flag Missing Values

- Add a helper column (e.g., "Price Missing?") with 1 for missing and 0 otherwise. This helps track which values were imputed.



### Manual Steps in Excel

#### Step 1: Identify Missing Values

- Use COUNTBLANK:

=COUNTBLANK(C2:C1000)

- (for Units Sold column)

#### Step 2: Fill with Average or Median

- Insert a cell with:
  - =AVERAGE(C2:C1000)
    - or
    - =MEDIAN(C2:C1000) Copy this value into missing cells.
- Step 3: Group-Based Filling
  - Create a **Pivot Table**:
    - Rows: Country/Category
    - Values: Average of Unit Price
  - Use these averages to fill missing values manually (or with VLOOKUP).
- Step 4: Interpolation (if time-based)
  - Sort data by date.
  - For missing values, use:
    - =(PreviousValue + NextValue)/2 Or use Excel's **Fill Series → Linear Trend**.
- Step 5: Document Changes
  - Add a helper column noting which values were filled and how (e.g., "Filled with median").
- ⚡ Best Practices
  - **Never blindly fill with averages** if the dataset is diverse — always check by group.
  - **Keep a record** of which method you used (important for analysis transparency).
  - **Consider the meaning of missingness**:
    - If missing means "not recorded," imputation makes sense.
    - If missing means "not applicable," use 0 or leave blank.

## Question 6

. Use Copilot to generate a summary chart of average profit by region.

Answer

[Assignment 5.xlsx](#)

 I'll read the dataset and calculate the **average profit** for each region.  I'll create a **visual chart** (bar chart or column chart) showing the comparison across regions.

-  I'll highlight which regions are performing best and which are lagging.

 Please upload your Excel file, and I'll build the chart for you!

## Question 7

Create a pivot table summarizing total sales by category and region.

Answer

[Assignment 5.xlsx](#)

To create a pivot table summarizing total sales by category and region, you'll need to work directly in Excel with your dataset. Here's a clear step-by-step guide you can follow:

 Steps to Create the Pivot Table

1. Open your dataset in Excel.
  - Make sure your data has clear column headers (e.g., *Category*, *Region*, *Sales*).
2. Insert a Pivot Table.
  - Select your dataset.
  - Go to the Insert tab → click PivotTable.
  - Choose whether to place the Pivot Table in a new worksheet or the existing one.
3. Set up the Pivot Table fields.
  - Drag Category into the Rows area.
  - Drag Region into the Columns area.
  - Drag Sales into the Values area.
  - By default, Excel will summarize sales as Sum of Sales (which gives you total sales).
4. Format the Pivot Table.
  - Apply number formatting to the sales values (e.g., currency or comma style).
  - Add filters if you want to narrow down by time period or other dimensions.

## Question8

build a dashboard combining key KPIs and charts.

### Answer

[Assignment 5.xlsx](#)

Got it  — building a dashboard in Excel that combines key KPIs and charts is a powerful way to monitor performance at a glance. Here's a structured approach you can follow:

#### Step 1: Identify Key KPIs

Choose the most relevant metrics for your business. Common examples:

- Total Sales 
- Average Profit Margin 
- Sales by Region 
- Top 5 Products by Revenue 
- Year-over-Year Growth 

#### Step 2: Prepare Your Data

- Ensure your dataset has clean headers (e.g., *Date, Region, Category, Sales, Profit*).
- Convert your data into an Excel Table (*Insert → Table*) for easier management.

#### Step 3: Build Pivot Tables

Create Pivot Tables for each KPI:

- Sales by Region → Rows: Region, Values: Sum of Sales.
- Profit Margin → Values: Profit ÷ Sales.
- Top Products → Rows: Product, Values: Sum of Sales, then sort descending.

#### Step 4: Add Charts

Insert charts linked to your Pivot Tables:

- Clustered Column Chart → Sales by Region.
- Line Chart → Monthly Sales Trend.
- Pie Chart → Category Contribution.
- KPI Cards → Use simple text boxes or conditional formatting to highlight totals.

 **Step 5: Assemble the Dashboard**

- Place Pivot Charts and KPI cards neatly on one worksheet.
- Use Slicers (Insert → Slicer) for interactivity (filter by Region, Category, Year).
- Add a Title and Branding for clarity.

**Question 9**

Use Copilot to generate a summary report.

Answer 9

[Assignment 5.xlsx](#)