DATA ANALYSIS

Excel Version: Microsoft Excel 2004
for Microsoft 365 MSO
(16.0.12730.20188) 64-bit

Contents

Question 1: Importing and Cleaning Data from Multiple Files

Step 1: Files in one folder

Download all comma separated value (CSV) files and put it in **one folder**. TIP: Create a folder on the Desktop.

Step 2: Selecting folder via Excel

1. Open a Blank workbook on Excel and then navigate to the **Data** menu.
2. Select **Get Data**.
3. Select **From File**.
4. Select **From Folder**.
5. Browse for the **folder** where all the CSV files are held in and then click **OK**.
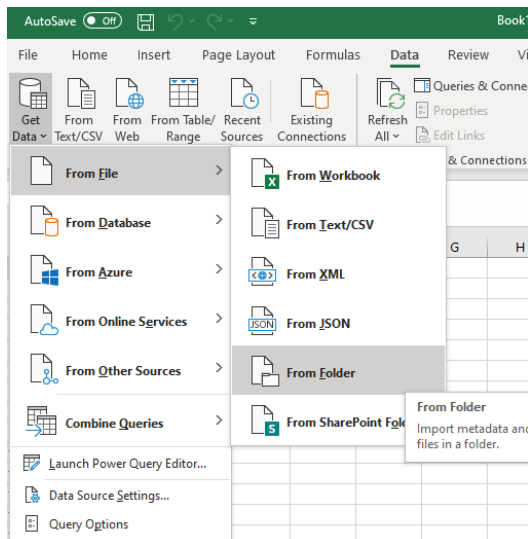
Step 3: Combing and opening Editor

1. Click **Combine** drop-down menu.
2. Select **Combine and Transform Data**.

Step 4: Data – Power Query Editor

1. Make sure the Sample File is sampling the First file.
2. Delimiter is Comma.
3. Data Type Detection is Based on first 200 rows and then Click **OK.**

Now Data – Power Query Editor window will open.

Step 5: Changing Data Type in Data - Power Query Editor

1. Select **Column1** (Column1 will be highlighted).

2. Change the Data Type to **Text**.

3. A window will appear, Change Column Type and click **Replace Current**.

Step 6: Appearing on Worksheet

1. Select **Close and Load**   .

Top Left of the Data – Power Query window.

Step 7: Copying and Pasting Data
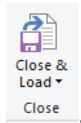
1. **Copy** all the columns (A:C)

2. Click the New Sheet icon or Paste it on the Sheet1 worksheet.

3. Right click on cell A1, go to **Past Special** and **Paste Values** only on the new worksheet.

4. **Adjust** each column to a reasonable size.

Step 8: Deleting and Renaming Worksheet

1. **Right click** on the old worksheet (Sheet2) **Sheet tab**.

2. Click **Delete**.

3. Click **Delete** again.

This will permanently remove that old worksheet.

4. Rename a worksheet by **Right clicking** on the new worksheets (Sheet1) **Sheet tab**

5. Click **Rename**.

6. Sheet1 will be highlighted, rename worksheet to **Question 1**.

Step 9: Rename Columns

1. Click on cell A1 and rename it to **Sensor ID**.

2. Click on cell B1 and rename it to **Time Stamp.**

3. Click on cell C1 and rename it to **Temperature.**

Step 10: Replace Values

1. Select the entire **Sensor ID** column(A:A). This can be done by just clicking the **A column tab** in the Editing area, which will highlight the entire column.
2. From the **Home** menu select **Find & Select** drop down menu.
3. The click **Replace**.

A window will appear as shown below:
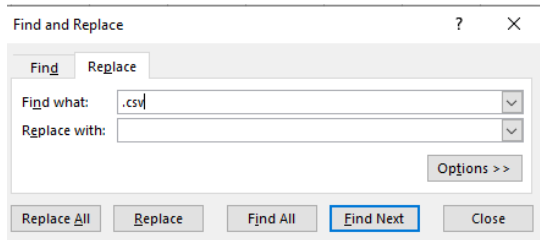
4. In the Find what: section, type **.csv**
5. In the Replace with: section, have it empty (like the picture above). Then select "Replace All".

Step 11: Changing Data Type in Worksheet

1. Select the entire **Sensor ID** column. This can be done by just clicking the A column tab.
2. From the **Home** menu find the **number** section (in Toolbar section).
3. Click the drop-down menu and select the Data Type to **Number**, as seen in the picture below. And then decrease the decimal to 0 decimal place.
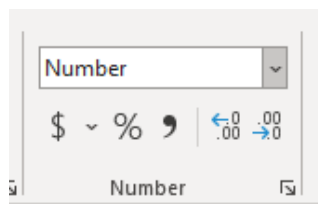
4. Same process should be applied for the **Time Stamp** column. But the Data type should be **Text** and there is no need to change the decimal placing.
5. The **Temperature** column's data type should be to **Number**.

Step 12: Add Filters

1. Highlight row 1, this can be done by click on the **1** in the editing area.
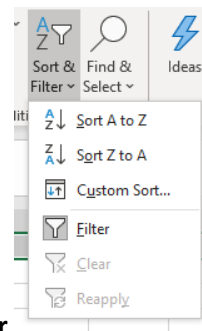
2. From the Home menu, in the editing section, select **Sort & Filter**

3. And then click **Filter**.

Step 13: Sorting Data

1. Select cell **A1**.

2. From the **Home** menu, in the editing section, Select **Sort & Filter**.

3. Select **Custom Sort**.

4. Condition to be applied is as follows below:
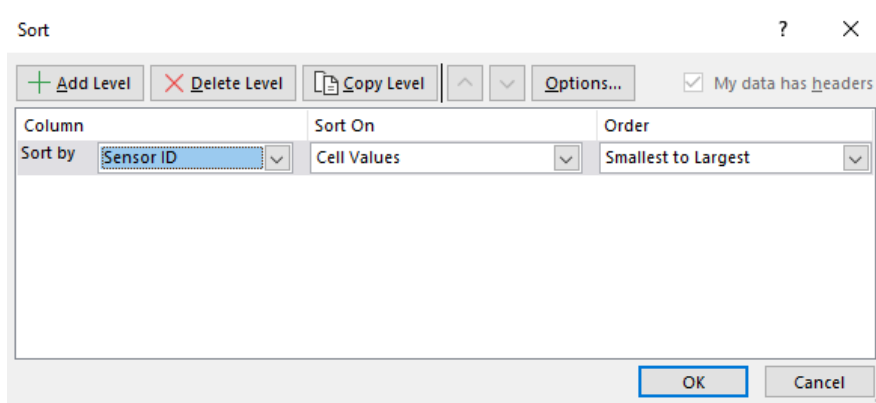
Question 2: Normalise the Data

Step 1: Copying Data from Existing Worksheet to another Worksheet

1. **Right click** on the Question 1 worksheet Sheet tab.
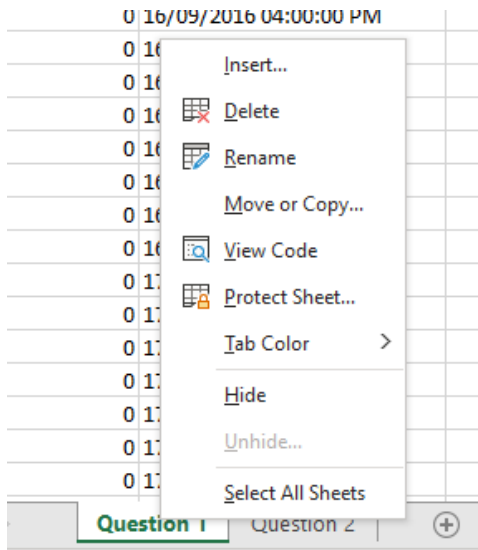
2. Click **Move or Copy…**

3. On the Move or Copy window. Select **(move to end)** in the Before sheet section.

4. Check the **Create a copy** checkbox, as shown below.
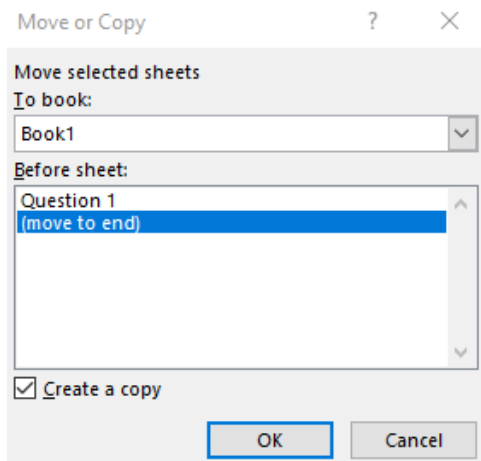
5. Then rename the **Question 1 (2)** worksheet to **Question 2**. (Refer to Step 8 in Question 1 Guideline).

Step 2: Creating more Columns

For this Question, everything will be done on the Question 2 worksheet.

1. Select cell D1 and type **Format**.
2. Select cell E1 and type **Year**.
3. Select cell F1 and type **Month**.
4. Select cell G1 and type **Day**.
5. Select cell H1 and type **Time**.
6. Select cell I1 and type **Normalised Time**.
7. Select cell J1 and type **Fahrenheit**.
8. Select cell K1 and type **Celsius**.
9. Adjust each cells width to a reasonable amount.

Step 3: Format Column

Please type formula, do not copy and paste

1. In cell D2, the formula should read as: **=IFERROR(IF(DAY(B2),"AU","US"),"US")**
2. Then populate the entire column with the same formula by double clicking the bottom right of cell D2 (a black t will appear).

The logic behind this is that in Australia the formatting for a Date is that the day is in the beginning and for the US it is in the middle. Therefore, this formula will identify if the formatting is AU, then it will record AU. But if it is not there will be an error. To counter that error, use IFERROR and set it to US, if an error occurs.

Step 4: Year Column

Please type formula, do not copy and paste

1. In cell E2, the formula should read as: **=IFERROR(YEAR(B2),MID(B2,7,4))**
2. Then populate the entire column with the same formula by double clicking the bottom right of cell E2.

Step 5: Month Column

Please type formula, do not copy and paste

1. In cell F2, the formula should read as: **=IFERROR(MONTH(B2),MID(B2,2,1))**
2. Then populate the entire column with the same formula by double clicking the bottom right of cell F2.

This is because the American format has the Month in the beginning of their Date format.

Step 6: Day Column

Please type formula, do not copy and paste

1. In cell G2, the formula should read as: **=IFERROR(DAY(B2),MID(B2,4,2))**
2. Then populate the entire column with the same formula by double clicking the bottom right of cell G2.

This is because the American format has the Day in the middle of their Date format.

Step 7: Time Column

Please type formula, do not copy and paste

1. In cell H2, the formula should read as: **=TEXT(RIGHT(B2,11),"hh:mm:ss")**
2. Then populate the entire column with the same formula by double clicking the bottom right of cell H2.

Step 8: Normalised Time Column

Please type formula, do not copy and paste

1. In cell I2, the formula should read as:

   =DATE(E2,F2,G2)+TIME(LEFT(H2,2),MID(H2,4,2),RIGHT(H2,2))

If followed correctly, the cells of Day, Month, Year and Time are all allocated in the correct cells.

2. Then populate the entire column with the same formula by double clicking the bottom right of cell I2.

3. Then, select the entire **Normalised Time** column (I column)

4. From the **Home** menu go to the number section and select the drop-down menu, that will present a list of all formatting types.

5. Select **More Number Formats.**

6. From the category list, select **Custom**. And fill in the Type box, type, **yyyy/mm/dd hh:mm,** then click **OK**. As shown below.

Step 9: Fahrenheit Column

Please type formula, do not copy and paste

1. In cell J2, the formula should read as:

    =IF(C2<MINIFS(C:C,D:D,"US"),CONVERT(C2,"C","F"),C2)

This is saying IF temperature in the temperature column is less than (<) the minimum temperature of the data from the US manufacturer, convert the temperature from Celsius to Fahrenheit but if it isn't that means that temperature is already converted to Fahrenheit.

2. Then populate the entire column with the same formula by double clicking the bottom right of cell J2.

Step 10: Celsius Column

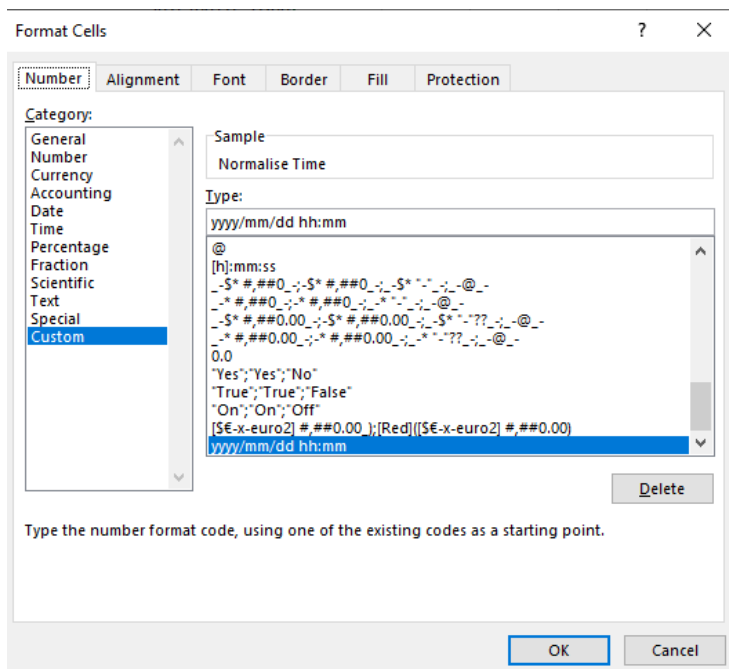*Please type formula, do not copy and paste*

1. In cell K2, the formula should read as:

    =IF(C2<=MAXIFS(C:C,D:D,"AU"),C2,CONVERT(C2,"F","C"))

This is saying IF temperature in the temperature column is less than or equal (<=) to the maximum temperature of the data from the AU manufacturer, then the temperature is already in Celsius but if it isn't convert the temperature from Fahrenheit to Celsius.

2. Then populate the entire column with the same formula by double clicking the bottom right of cell K2.

Step 11: Formatting/Presentation

1. Add **Filter** to row 1. This may take the filter away if you click it the first time. If it does repeat Add **Filter steps**. (Refer to Step 12 of Question 1 Guideline).
2. After this, click on cell A1, then go to the **Insert menu**.
3. Click **Table**, it will ask about Where is the data for your table?, this will be "**=$A$1:$K$9026**" and have the **My table has headers** box checked, then click **OK.**
4. Default layout will be selected automatically but can be changed via **Table Design** menu.

5. Adjust each column appropriately, so that all values are displayed clearly.

Question 3: Overlap

Step 1: Create a Pivot Table

1. From the Question 2 worksheet, go to the **Insert** menu and click **Pivot Table**.

2. The **Select a table or range** box should be checked.

3. Select the entire table including the headers, **Table/Range:** 'Question 2'!$A:$K

4. **New Worksheet** box should be checked and then click **OK**.

5. From the Pivot Table Fields section, drag **Sensor ID** to the **Columns** section.

6. From the Pivot Table Fields section, drag **Normalised Time** to **Rows** section.

7. From the Pivot Table Fields section, drag **Fahrenheit** to the **Values** section.

8. In the **Values** section, click the drop-down menu for **Sum of Fahrenheit.**

9. Click **Value Field Settings**…

10. Change the **Summarize Values By** box to **Average** by selecting average and then click **OK**.

Step 2: Displaying Correlation Table Sensor ID overlaps

1. From **Data** menu, click **Data Analysis**
2. From the Data Analysis window, select **Correlation** and then click **OK**.
3. **Input Range:** $B$4:$Z$21.

4. **Grouped by:** Columns.

5. Check the **Labels in First Row** checkbox.

6. Check the **New Worksheet Ply:** checkbox and then click **OK.**

7. On the new worksheet created, highlight cells from B2:Z26.

8. Then from the **Home** menu, select the **Conditional Formatting** drop-down box.

9. Click **New Rule…**

10. Select a Rule Type: **Format only cells that contain**

11. In the Format only cells with: section, have **Cell Value between 0.99 and 1.**

12. Click **Format**, then **Fill** and Choose any color as the **Background Color**, then click **OK**, then **OK** again.

Step 3: Identifying Sensor ID overlaps

Using the correlation table, identifying Locations (Sensors) that overlap will be clear. The Sensor ID's are displayed along row 1 and column A. To identify overlaps, find the data that has a strong correlation, in this case it will be correlations that are between 0.99 and 1. A correlation of 1 will suggest that that the locations are identical. However, since some values are not identical, correlation between locations will be the ones that are greater than 0.99 or simply closest to 1. Hence, the locations that have the strongest correlation will be highlighted. Ignore the values that are 1 as they are correlated with their own sensors.

Therefore, Locations that are Identical are as follows:

Sensor 0 (AU) and Sensor 14 (US)
Sensor 5 (AU) and Sensor 9 (US)
Senor 8 (US) and Sensor 22 (AU)
Sensor 15 (US) and Sensor 18 (AU)
Sensor 16 (US) and Sensor 21 (AU)

Step 4: Rename Worksheets

1. Rename worksheet Sheet1 to **Question 3 Pivot Table**
2. Rename worksheet Sheet2 to **Question 3 Overlaps**
3. OPTIONAL**:** Drag sheets to order them to your liking.

Question 4 (a): Min, Max and Average of each Region

1. From the Question 2 worksheet, go to the **Insert** menu and click **Pivot Table**
2. The **Select a table or range** box should be checked. And select the entire table including the headers, **Table/Range:** 'Question 2'!$A:$K
3. **New Worksheet** box should be checked and then click **OK**.
4. A new worksheet will appear.
5. From the Pivot Table Fields, drag **Celsius** to the values area and change it to **Minimum** from the Value Field Setting (refer to Steps 8-10 in Question 3 Create a Pivot Table).
6. From the Pivot Table Fields, drag **Celsius** to the values area and change it to **Maximum**.
7. From the Pivot Table Fields, drag **Celsius** to the values area and change it to **Average**.
8. From the Pivot Table Fields, drag **Sensor ID** to the rows area.
9. Rename worksheet to **Question 4 A** and re-order worksheet.
10. OPTIONAL: Select the Row Labels filter and uncheck Blanks.

Question 4 (b): Trends

1. From the Question 2 worksheet, go to the **Insert** menu and click **Pivot Chart**
2. The **Select a table or range** box should be checked.
3. Select the entire table including the headers, **Table/Range:** 'Question 2'!$A:$K
4. **New Worksheet** box should be checked and then click **OK**.
5. A new worksheet will appear.
6. From the Pivot Table Fields, drag **Celsius** to the values area and change it to **Minimum**.
7. From the Pivot Table Fields, drag **Celsius** to the values area and change it to **Maximum**.
8. From the Pivot Table Fields, drag **Celsius** to the values area and change it to **Average**.
9. From the Pivot Table Fields, drag **Sensor ID** to the filters area.
10. From the Pivot Table Fields, drag **Normalised Time** to the rows area.
11. Right click on the graph, then select **Change Series Chart Type** and choose the first **Line** graph.
12. Click on the graph, from the **Design** menu, click **Add Chart Element**, then **Trendline**, then **Linear**, highlight **Average of Celsius** and then **OK**. (OPITIONAL: repeat for Max and Min of Celsius).
13. Select the **Row Labels** filter and uncheck <15/09/2016 and >1/10/2016.
14. The Sensor ID filter will appear on the top of the pivot table, this is where Locations (Sensor ID) can be filtered to check which locations (Sensor IDs) are experiencing a raising or declining trend in temperature (Celsius).
15. Rename worksheet to **Question 4 B** and re-order worksheet.

Question 4 (c): Additional Information – Average and Standard Deviation

1. From the Question 2 worksheet, go to the **Insert** menu and click **Pivot Chart**
2. The **Select a table or range** box should be checked.
3. Select the entire table including the headers, **Table/Range:** 'Question 2'!$A:$K
4. **New Worksheet** box should be checked, then **OK**.
5. A new worksheet will appear.
6. From the Pivot Table Fields, drag **Celsius** to the values area and change it to **Average**.
7. From the Pivot Table Fields, drag **Celsius** to the values area and change it to **StdDev**.
8. From the Pivot Table Fields, drag **Normalised Time** to the rows area.
9. From the Pivot Table Fields, drag **Format** to the filters area.
10. Right click on the graph, then select **Change Series Chart Type** and choose the first **line** graph.
11. Click on the graph, then go to the **Design** menu, then click **Add Chart Element**, then **Trendline**, then **Linear**, highlight **Average of Celsius** and then **OK**.
12. Rename worksheet to Question 4 C.
13. Filter to which Format (manufacturer) you want to look at to see that has been recording a raising or declining trend in temperature.