Managing and Analysing Data using Microsoft Excel

DATA SKILLS: Part 2

In this class, we focus only on selected and advanced exercises that will help to develop essential data analysis skills for BI work.

> **Note:** If you want to receive the certificate of course completion from LinkedIn, please make sure you complete all activities of the course (https://www.linkedin.com/learning/excel-2016-managing-and-analyzing-data/welcome?u=2104084) at your own pace and in your own time.

1. FILTER BY USING ADVANCED CRITERIA.

If the data you want to filter requires complex criteria (such as Job Status = "Contract" OR Years = "8"), you can use the Advanced Filter dialog box under the Data tab.

To open the Advanced Filter dialog box, click the **Data** tab -> **Advanced.**

TASK:

- Open the file **CH_05.xlsx,** worksheet **Complex OR**

- **Follow** the instructions in the lesson 'Using Advanced Filter for complex 'OR' criteria' https://www.linkedin.com/learning/excel-2016-managing-and-analyzing-data/using-advanced-filter-for-complex-or-criteria?resume=false&u=2104084

- Let your tutor know if you need assistance.

- You can now answer the questions on the Learning site.

Question. With most Advanced filter uses, what kind of range do you need to set up?

- Structured

- Criteria

- Sorted

- slicer

2. ELIMINATING DUPLICATE DATA

Sometimes duplicate data makes it harder to understand your data. Use **conditional formatting** to find and highlight duplicate data. Or you can use **remove duplicates** command.

Note: Using **conditional formatting**, you can review the duplicates and decide whether to remove them. Using the **remove duplicates** command, the duplicate data will be permanently deleted. Before you delete the duplicates, copying the original data to another worksheet is a good idea so you don't accidentally lose any data.

2.1 Using conditional formatting.

TASK:

1. Open the file **CH_06.xlsx,** worksheet **List with Duplicates**

Have a look at the dataset, then answer the following questions:

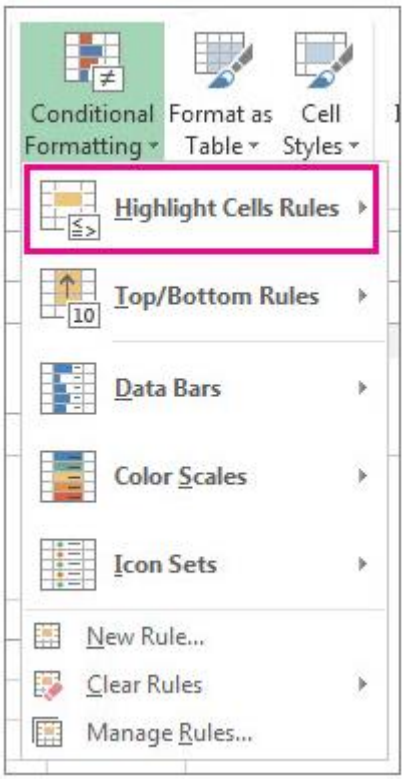
- What do you think about the quality of the dataset?

- Can you spot any issues?

2. Select the cells you want to check for duplicates.

Note: Excel can't highlight duplicates in the Values area of a PivotTable report.

3. Click **Home** > **Conditional Formatting** > **Highlight Cells Rules** > **Duplicate Values**.



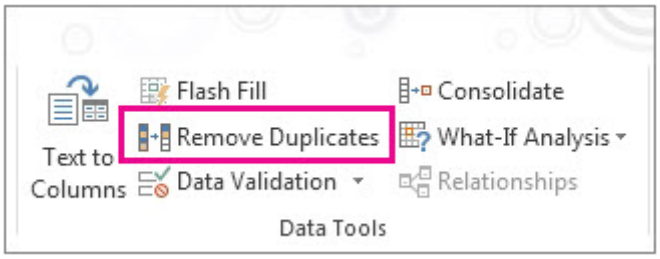
4. In the box next to **values**, pick the formatting you want to apply to the duplicate values, and then click **OK.**

Using the remove duplicates command:

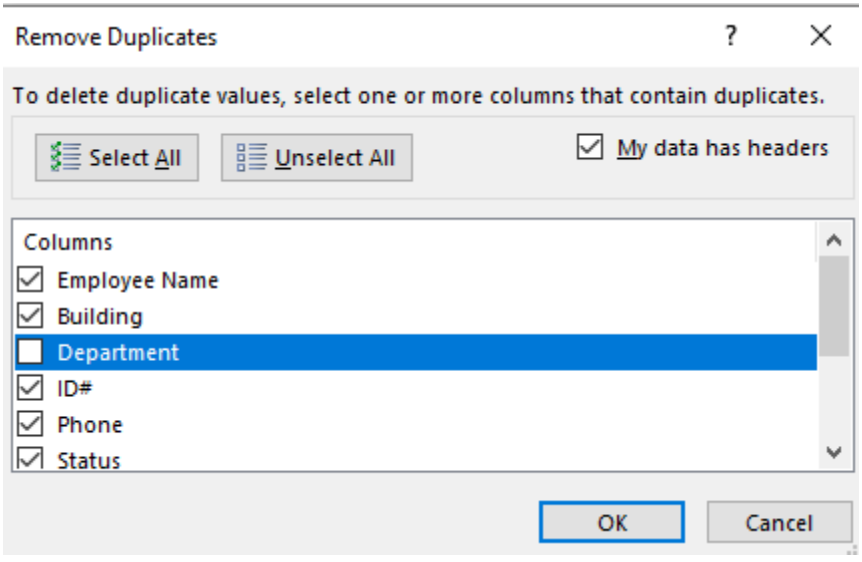
Using **the remove duplicates command, duplicate data will be permanently deleted. Before you delete the duplicates, it is a good idea to copy the original data to another worksheet so you don't accidentally** lose any data.

- Select the cell range with duplicate values you want to remove.

- Click **Data** > **Remove Duplicates**, and then Under **Columns**, check or uncheck the columns where you want to remove the duplicates.



For example, uncheck **Department** in the **Remove Duplicates** box.



- Click OK

Questions to discuss: Why do we need to remove duplicate data?

3. DATA ANALYSIS TOOLS

3.1. Quick data analysis using SUMIF, COUNTIF, and related functions

The SUMIF function is to sum the values in a range that meets the *criteria that you specify. For example, suppose that in a column containing numbers, you want to sum only the values larger than 8.* You can use the following formula: =SUMIF(B2:B25,">8")

Syntax: SUMIF (range, criteria, [sum_range]). For example: =SUMIF(A2:A5,300000,B2:B5)

Range: Required. The range of cells that you want to be evaluated by criteria. Cells in each range must be numbers or names, arrays, or references that contain numbers. Blank and text values are ignored. The selected range may contain dates in standard Excel format (examples below).

Criteria: Required. The criteria are in the form of a number, expression, a cell reference, text, or a function that defines which cells will be added. Wildcard characters can be included - a question mark (?) to match any single character and an asterisk (*) to match any sequence of characters. If you want to find an actual question mark or asterisk, type a tilde (~) preceding the character. For example, criteria can be expressed as 8, ">8", C5, "3?", "orange*", "*~?", or TODAY().

sum_range: Optional. If you want to add cells other than those specified in the range argument, the actual cells to add. If the **sum_range** argument is omitted, Excel adds the cells specified in the range argument (the same cells to which the criteria is applied). **Sum_range** should be the same size and shape as range. If it isn't, performance may suffer, and the formula will sum a range of cells that starts with the first cell in sum_range but has the same dimensions as range. For example:

range	***sum_range***	**Actual summed cells**
A1:A5	B1:B5	B1:B5
A1:A5	B1:K5	B1:B5

COUNTIFS *function applies criteria to cells across multiple ranges and counts the number of times all criteria are met.*

Syntax: COUNTIFS(criteria_range1, criteria1, [criteria_range2, criteria2]…)

criteria_range1: Required. The first range is in which to evaluate the associated criteria.

criteria1 Required. The criteria in the form of a number, expression, cell reference, or text that defines which cells will be counted. For example, criteria can be expressed as 32, ">32", B4, "apples", or "32".

criteria_range2, criteria2, ... Optional. Additional ranges and their associated criteria. Up to 127 range/criteria pairs are allowed.

AVERAGEIFS *function returns the average (arithmetic mean) of all cells that meet multiple criteria*

Syntax: AVERAGEIFS(average_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)

Average_range Required. One or more cells to average, including numbers or names, arrays, or references that contain numbers.

Criteria_range1, criteria_range2, … Criteria_range1 is required, subsequent criteria_ranges are optional. 1 to 127 ranges in which to evaluate the associated criteria.

Criteria1, criteria2, ... Criteria1 is required; subsequent criteria are optional. 1 to 127 criteria in the form of a number, expression, cell reference, or text that define which cells will be averaged. For example, criteria can be expressed as 32, "32", ">32", "apples", or B4.

MAXIFS *returns the maximum value among cells specified by a given set of conditions or criteria.*

Syntax: MAXIFS(max_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)

max_range Required. The actual range of cells in which the maximum will be determined.

criteria_range1 Required. Is the set of cells to evaluate with the criteria.

criteria1 Required. The criteria is a number, expression, or text that defines which cells will be evaluated as maximum.

criteria_range2, criteria2, ... Optional. Additional ranges and their associated criteria. You can enter up to 126 range/criteria pairs.

MINIFS *function returns the minimum value among cells specified by a given set of conditions or criteria*.

Syntax: MINIFS(min_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)

man_range Required. The actual range of cells in which the minimum will be determined.

criteria_range1 Required. Is the set of cells to evaluate with the criteria.

criteria1 Required. It is the criteria in the form of a number, expression, or text that defines which cells will be evaluated as minimum. The same criteria work for the MAXIFS, SUMIFS and AVERAGEIFS functions.

criteria_range2, criteria2, ... Optional. Additional ranges and their associated criteria. You can enter up to 126 range/criteria pairs.

TASK:

- Open file**: CH_07.xlsx,** worksheet **Countif-Sumif**

- **Follow** the instructions in the lesson 'Using SUMIF, COUNTIF, and related functions for quick data analysis.'

https://www.linkedin.com/learning/excel-2016-managing-and-analyzing-data/using-sumif-countif-and-related-functions-for-quick-data-analysis?u=2104084

- Let your tutor know if you need assistance.

3.2. Data analysis using Database functions DSUM, DAVRAGE, DMAX

DSUM *adds the numbers in a field (column) of records in a list or database that match conditions that you specify*

Syntax: DSUM(database, field, criteria)

> **Database** Required. The range of cells that make up the list or database. A database is a list of related data in which rows of related information are records and columns of data are fields. The first row of the list contains labels for each column.
>
> **Field** Required. Indicates which column is used in the function. Enter the column label enclosed between double quotation marks, such as "Age" or "Yield," or a number (without quotation marks) that represents the position of the column within the list: 1 for the first column, 2 for the second column, and so on.
>
> **Criteria** Required. Is the range of cells that contains the conditions that you specify. You can use any range for the criteria argument, as long as it includes at least one column label and at least one cell below the column label in which you specify a condition for the column.

DAVERAGE *averages the values in a field (column) of records in a list or database that match conditions you specify.*

Syntax: DAVERAGE(database, field, criteria)

> **Database** Require. is the range of cells that make up the list or database. A database is a list of related data in which rows of related information are records and columns of data are fields. The first row of the list contains labels for each column.
>
> **Field** Required. indicates which column is used in the function. Enter the column label enclosed between double quotation marks, such as "Age" or "Yield," or a number (without quotation marks) that represents the position of the column within the list: 1 for the first column, 2 for the second column, and so on.
>
> **Criteria** Required. is the range of cells that contains the conditions you specify. You can use any range for the criteria argument, as long as it includes at least one column label and at least one cell below the column label in which you specify a condition for the column.

DMAX *Returns the largest number in a field (column) of records in a list or database that matches the conditions you specify.*

Syntax: DMAX(database, field, criteria)

Database Required. The range of cells that make up the list or database. A database is a list of related data in which rows of related information are records and columns of data are fields. The first row of the list contains labels for each column.

Field Required. Indicates which column is used in the function. Enter the column label enclosed between double quotation marks, such as "Age" or "Yield," or a number (without quotation marks) that represents the position of the column within the list: 1 for the first column, 2 for the second column, and so on.

Criteria Required. The range of cells that contains the conditions that you specify. You can use any range for the criteria argument, as long as it includes at least one column label and at least one cell below the column label in which you specify a condition for the column.

TASK:

- Open file**: CH_07.xlsx.** Worksheet **DS Functions**

- **Follow** the instructions in the lesson 'Using database functions including DSUM, DAVRAGE, DMAX'

https://www.linkedin.com/learning/excel-2016-managing-and-analyzing-data/using-database-functions-including-dsum-daverage-dmax?u=2104084

- Let your tutor know if you need assistance.

- You can now answer the questions on the Learning site.

----------END----------