Managing and Analysing Data using Microsoft Excel

DATA SKILLS: Part 1

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

> **Note:** Deakin's Office 365 subscription enables Deakin students to install the Office 365 suite onto five computers for free whilst enrolled at Deakin. See:
>
> IT Help - How do I download and install Microsoft Office? (Windows)
>
> IT Help - How do I get Microsoft Office for free through Deakin University? (Mac)
>
> Accessing Office 365 online
>
> Digital Essentials – Students Guide
>
> *If you want to receive the certificate of course completion from LinkedIn Learning, 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.

STRUCTURE DATA FOR OPTIMUM USAGE

1. Remove missing row and column:

Open the file **CH_01**

Browse through the dataset, then answer the following questions:

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

- Can you spot any issues? Empty rows? Empty columns?

1.1. Remove empty row:

First, that might strike your eye is that empty row 13. Now, again, with a small list, it won't cause any huge problems or anything like that. But as a simple example here, what if the list is substantially more extensive and there's an empty row in 100 or 300? What happens if you click within the data here and say , "I want to sort the data"? Now, with the apparent row showing here, immediately we'll see what can go wrong. **Right-click row 13 in the example here, and delete it.**

1.2 Remove hidden column: Column E

We don't want empty rows or empty columns within your data. Right-click and delete this.

2. Restructure data using 'Flash fill' and 'Text to Columns': (5:06' in structure data for optimal usage)

Data quality issues:

We've got some other problems; columns A, B, and to a lesser extent column C all have the same problem. And that is too much data in the same column. It doesn't necessarily look like a lot of data, but the way this is structured here, we cannot sort this list by last name. Now, if it's only five, 10, 15, that's not a big deal. We can probably find the names pretty quickly, but at some point, or other, you're likely to want to sort this list by last name. And you cannot do that with the names looking the way they are right now.

The suggestion here is to do one of two things:

Either split these names into two separate columns, one called first name and one called last name. In some cases, you may see middle initials in the names as well, so it may take three columns.

- Another approach is to restructure these names so that we see the last name, comma, space, and first name—for example, Wagner, comma, space, and Max. Now, there are multiple ways to clean up data. We can use '*Flash fill*' or '*Text to Columns.*'

2.1. Flash fill to clean up data in Column A.

- Insert a new column next to Column A: **right click column B** and **insert a new column**.

- In the new column, type Wagner, comma, space, Max. Then, **press control enter** (Ctrl + Enter) so the active cell doesn't move.

- **Point to this flash fill feature** (or a key stroke shortcut **control E**)

- Click **flash fill**. We see a corrected entry.

Now, have a look at column 'City, State Zip',

- What is the problem? Can we sort this list by state? No, we can't.

- Can we sort it by zip code? No, we can't.

Once again, we've got the issue of too much data in a column.

So, either using **flash fill** or **text to columns,** we could clean up that data. In other words, there is too much data in each successive cell. We can't sort our list by state or zip code like this. We need to split this data into three separate columns using '**Text to columns.**'

2.2 Text to column:

- Insert three new columns next to Column B

- **Select** the column ‘City, State Zip’.

- On the Data tab, in the Data Tools group, click **Text to Columns**.

- Choose Delimited in the dialog box that appears, and click Next.

- Clear all the check boxes under Delimiters except for the ‘Comma’ and ‘Space’ check box.

- Click Finish.

Now discuss the following questions:

- What is the issue of the data?

- How can you fix it?

- What is the issue now?

- What is the best solution?

3. Reformatting data:

Column ‘*Social security number*’.

- Press the column ‘Office phone’:

- On the home tab in the ribbon, far right button, click ‘Find & Select’

- Choose ‘Replace’.

- Enter ‘-‘ in the box ‘Find what’,

- Leave the box “**Replace with**” empty,

- Click ‘**Replace all**’.

We have replaced the dash with nothing.

Now… Repeat the same process, but replace the right and left parenthesis and space with nothing.

So, several different techniques here on this worksheet are called poor design. So, when setting up data, remember some of the tips we have seen in looking at this list. We want to avoid all the problems if we don't structure our data efficiently.

Let your tutor know if you need assistance.

Now, discuss the following q**uestions:**

Q1. City, State and Zip data should be entered in separate column?

Yes/No?

Q2. Question 4 of 5 (screen shot the spreadsheet)

In the spreadsheet shown below, the cells in column B contain different types if data, making the data hard to work with. What tool can help you re-structure this data for optimum usage? (select 2 or more)

A. Sort by Column

B. Flash Fill

C. Ungroup

D. Filter

E. Text to Columns

SORTING DATA

Open the file: CH_02.xlsx.; Sheet: MultipleKeySorting

Tasks:

1. Watch to the lesson "Multiple-key sorting" (https://www.linkedin.com/learning/excel-2016-managing-and-analyzing-data/multiple-key-sorting?u=2104084)

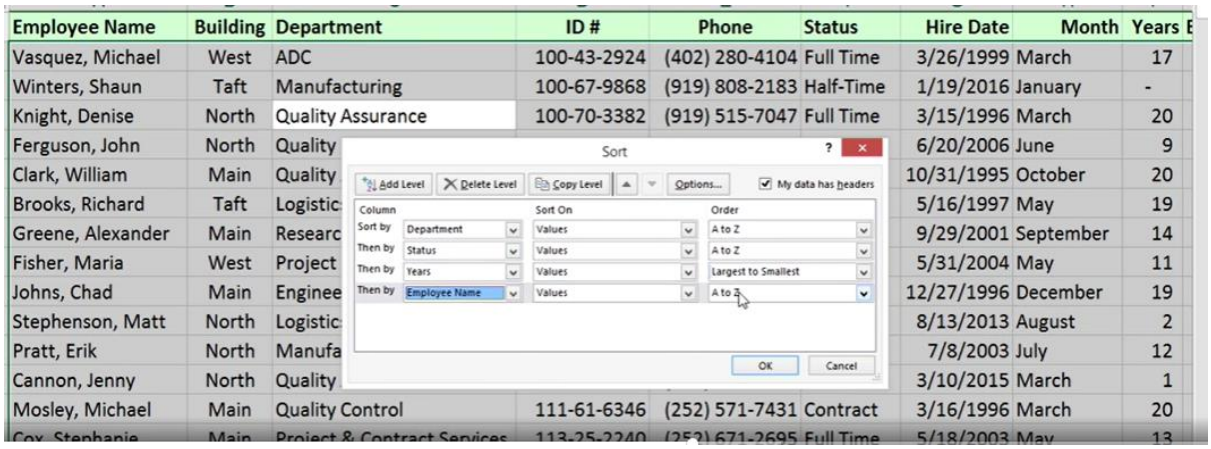
2. Sort data by the following multiple level as in the picture below.

Sort by "*Department*',

Then by *'Status'* ,

Then by *'Years'* of service

Then by *'Employee Name'*



Examine the results, then answer the following questions:

Q1. How are Full-time people sorted in order?

By year of service (level 3)

Q2. What is the major grouping?

Department (column C)

CREATING AUTOMATIC SUBTOTALS IN SORTED LISTS

Open the file CH_03.xlsx

Tasks:

1. Watch and follow the instructions in the lesson https://www.linkedin.com/learning/excel-2016-managing-and-analyzing-data/setting-up-single-and-multiple-level-subtotals?u=2104084

2. Let your tutor know if you need assistance.

3. Now, answer the questions on the Learning site.

Q1. Filtering can cause a loss of_____, so make sure you save your list before applying a filter.

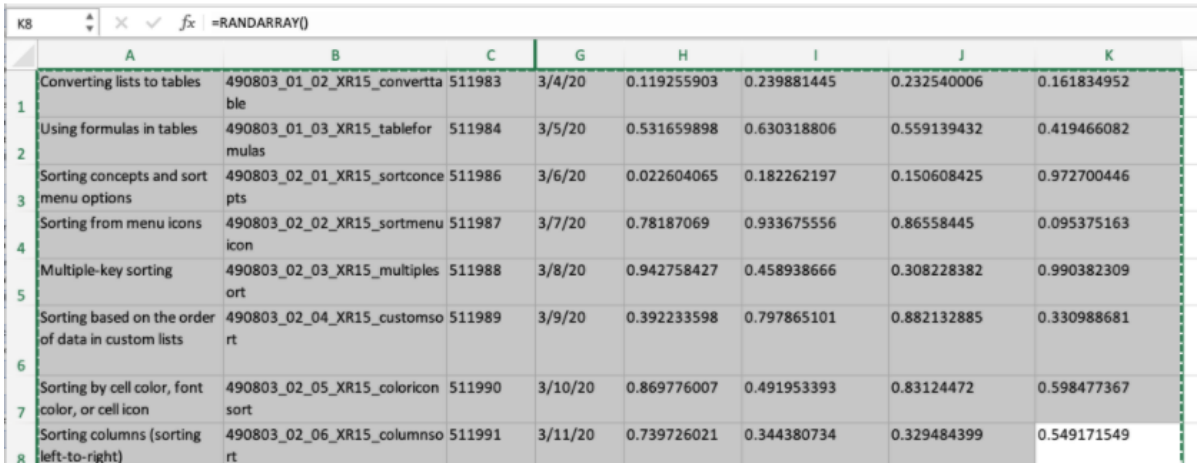
A. Headers

B. None of these answers

C. Data

D. Formatting

Q2. In the table below, you have selected the list data for cells A1:K8. You want to copy and paste this information to a new worksheet, excluding the hidden data in columns D-F. What is the most efficient way to do this?



K8 =RANDARRAY()

	A	B	C	G	H	I	J	K
1	Converting lists to tables	490803_01_02_XR15_convertable	511983	3/4/20	0.119255903	0.239881445	0.232540006	0.161834952
2	Using formulas in tables	490803_01_03_XR15_tableformulas	511984	3/5/20	0.531659898	0.630318806	0.559139432	0.419466082
3	Sorting concepts and sort menu options	490803_02_01_XR15_sortconcepts	511986	3/6/20	0.022604065	0.182262197	0.150608425	0.972700446
4	Sorting from menu icons	490803_02_02_XR15_sortmenuicon	511987	3/7/20	0.78187069	0.933675556	0.86558445	0.095375163
5	Multiple-key sorting	490803_02_03_XR15_multiplesort	511988	3/8/20	0.942758427	0.458938666	0.308228382	0.990382309
6	Sorting based on the order of data in custom lists	490803_02_04_XR15_customsort	511989	3/9/20	0.392233598	0.797865101	0.882132885	0.330988681
7	Sorting by cell color, font color, or cell icon	490803_02_05_XR15_coloriconsort	511990	3/10/20	0.869776007	0.491953393	0.83124472	0.598477367
8	Sorting columns (sorting left-to-right)	490803_02_06_XR15_columnsort	511991	3/11/20	0.739726021	0.344380734	0.329484399	0.549171549

A. Use Paste Special -> Value & Number Formatting

B. Unhide columns D-F and delete them. Then copy and Paste the new A1: H8 selection.

C. Use the Copy only Visible Cells keyboard shortcult (Alt/Opt +;)

D. You don't need to do anything. In list data, hidden cells are excluded by default when Copying and Pasting.

Q3. To get sensible results when using the Subtotal feature with a list, take this action:

A. Delete all formula columns

B. Sort the list first

C. Convert all formulas to their results.

D. Remove the title row first.

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