

EXCEL FOR OPERATIONAL ANALYSTS

This is a short self-study aid for individuals who want to use Excel 2007 to analyse operational data.

It is designed to provide individuals with some structure for learning about some of the more useful data analysis functions in Excel that can make analysis of large data sets more straightforward and yielding of insight.

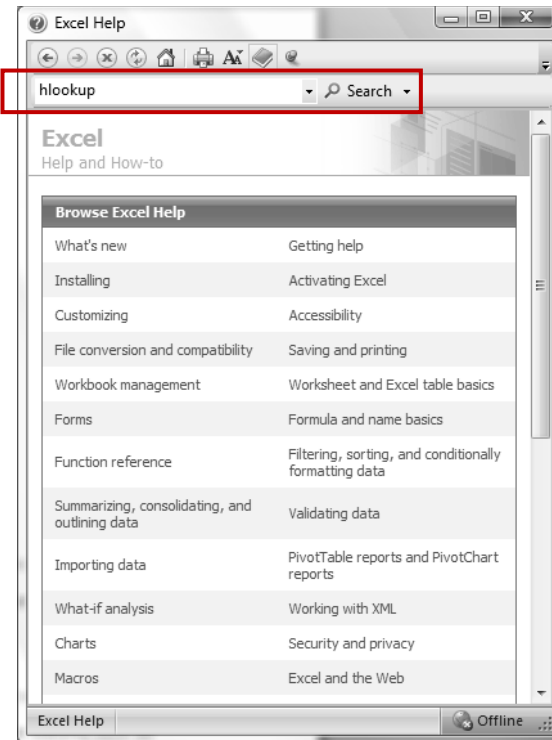


Saigei Teamtrain teaches the operational and organisational development techniques practiced by Saigei teams to staff from health and social care organizations.

THE BASICS

EXCEL HELP

This is an extraordinary resource which can be accessed by striking the F1 key. Once Help has launched you can search for the feature or function you need more support on. Here you will find explanations and examples.

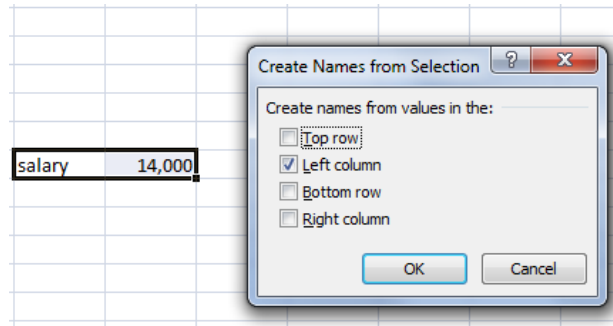


NAMING CELLS & RANGES

Excel allows you to name cells, ranges of cells (rows or columns) and tables. This means that rather than writing formula like $=L20-(L20*L21)$ you can write $=\text{salary}-(\text{salary}*\text{tax_rate})$.

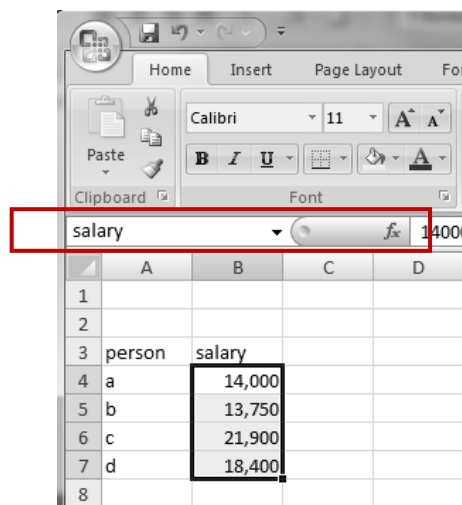
You really should name all values in your spreadsheets. This makes your work altogether easier to change, audit and develop which will in turn make it less prone to error.

To name a cell type the name next to the cell with the value in it you want to name and strike **CTRL+SHIFT+F3**.



Now, in this example, whenever you type =salary Excel will return £14,000.

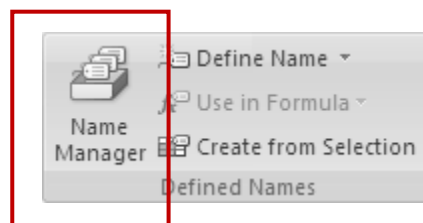
To name a range select the range and type the name into the name box to the left of the formula bar and press enter.



Now, =salary will return a different value for each row.

Once you have assigned a name you can only make changes to it using the **Name Manager**. Here you can delete the name; change the cells to which it refers; or in complex workbooks find out what the name refers to.

You can open the Name Manager using the Excel Ribbon. Click on the Formulas tab, and in the Defined Names group, click Name Manager.

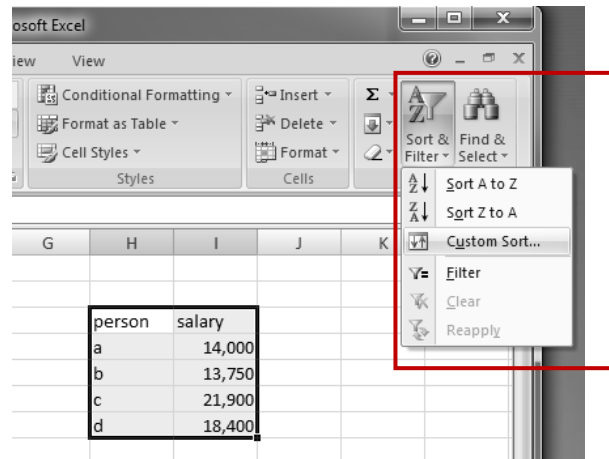


SORTING AND FILTERING

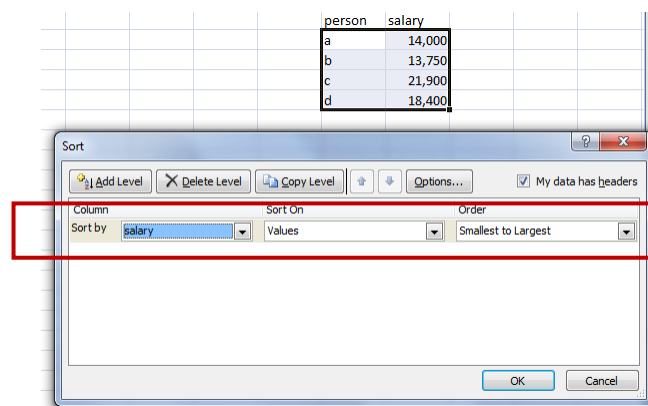
Any analyst should visually inspect and have a strong feel for the data they are analysing. The Sort and Filter tools help in this regard.

SORTING

To sort an array of data select the data; click on the Sort & Filter button and select Custom Sort.



From here you can define which column is to be sorted and in what order. Note how the names have come through to the Sort By dialogue. If you have a large array of data, name every column and sorting will be much easier to perform.



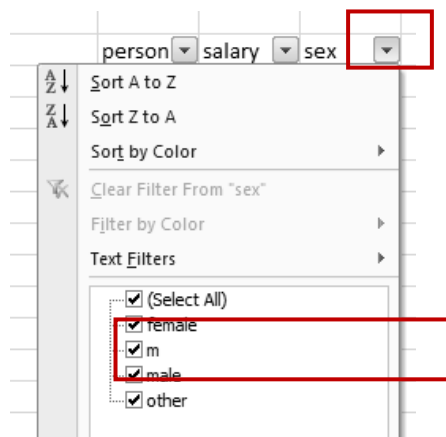
FILTERING

Using AutoFilter to filter data is a quick and easy way to find and work with a subset of data in a range of cells or table column.

Select your array. Click on the Sort & Filter button. Select Filter. Little arrow buttons will appear in the cells of the top row of your data.

person	salary	sex
a	14,000	male
b	13,750	female
c	21,900	male
d	18,400	other
e	21,000	

Your first step is to simply click on the arrow. It will list for you all the values in that column. This is a quick way of highlighting different but duplicate values eg “m” and “male”.



You can then use the Replace function to set all instances of “m” to “male” so when you query your data you get the right answer.

You can define what rows in your array are visible by ticking and unticking the boxes to the left of the values in the list. Filtered data displays only the rows with the values you tick. The other rows are still there, they are just hidden.

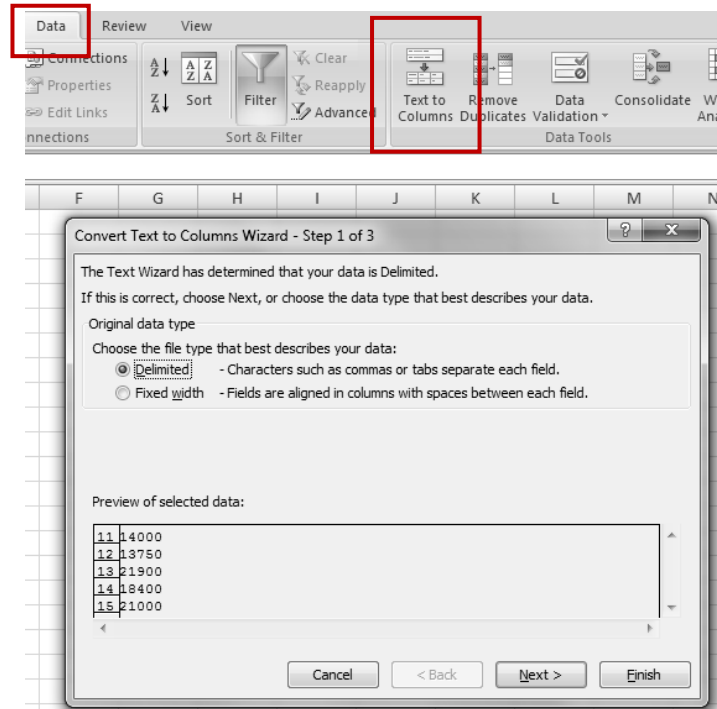
After you filter data, you can copy, find, edit, format, chart, and print the subset of filtered data without rearranging or moving it.

You can filter by more than one column. Filters are additive which means that each additional filter is based on the current filter and further reduces the subset of data.

CONVERTING “TEXT TO COLUMNS”

Sometimes IT systems provide data as comma separated strings. On other occasions you will want to separate a value in one cell into two (eg name and surname or a date-time stamp into a date and a time).

Select the cell or column containing the value you want to split and then click on the **Convert Text to Columns** button in the ribbon under the **Data** tab. The wizard is self explanatory. If you need more help strike **F1** and search for Text to Columns.

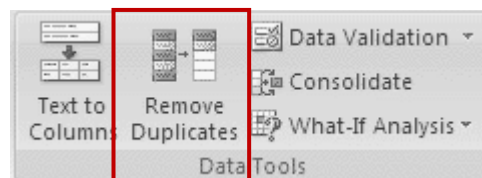


REMOVING DUPLICATES

NB Use this utility only once you know each of the duplicated records and you are sure that they are not simply the result of other data anomalies. Visual inspection of data using the Filter function is invaluable in this regard. Even then, Saigei analysts save a copy of the pre-deletion data so they can always go backward.

It is not uncommon for data-sets to contain duplicate values. Excel has a useful utility for automatically and permanently deleting duplicate records.

Select the range of cells that you wish to purge of duplicates. On the **Data** tab, in the **Data Tools** group, click **Remove Duplicates**. If you need more help hit F1.

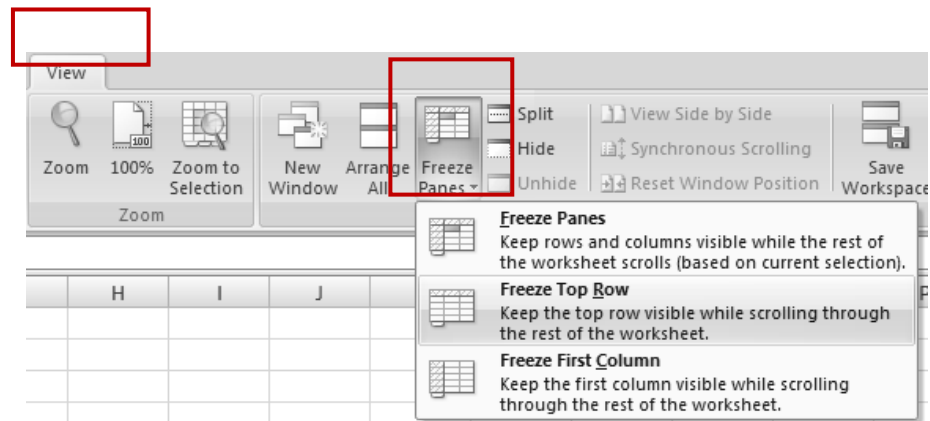


A message is displayed indicating how many duplicate values were removed and how many unique values remain, or if no duplicate values were removed.

FREEZE PANES

Sometimes your data array will be so broad and so long that you will want to keep some of it visible as you scroll around. For example you may want to retain the titles of columns.

Excel allows you to lock specific rows or columns in place by **Freezing or Splitting Panes**. This utility can be found on the View tab of the Excel Ribbon.



To freeze multiple rows, click the cell below the last row you want to freeze. To lock multiple columns, select the column to the right of the columns you want to keep visible. To lock both rows and columns, click the cell below and to the right of the rows and columns that you want to keep visible when you scroll.

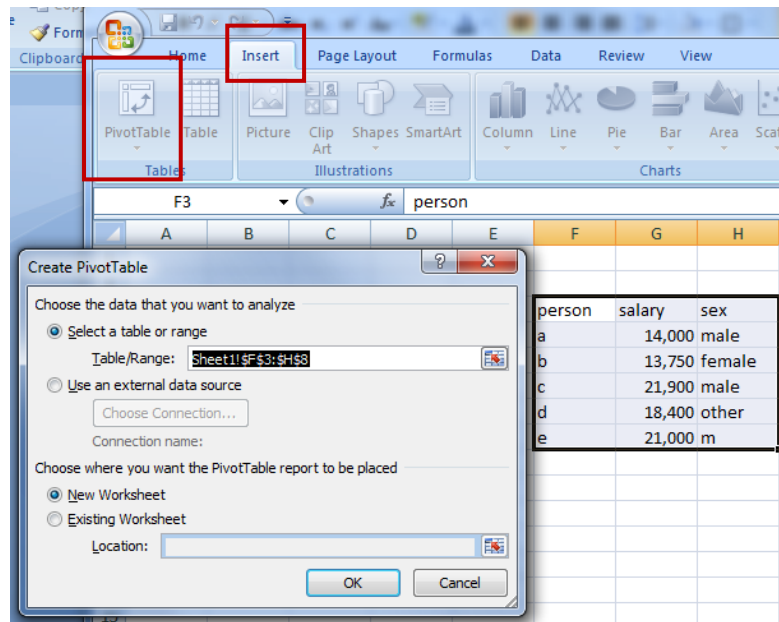
PIVOT TABLES

There is a mystic around pivot tables but they really aren't that difficult to use. They let you summarise and present an overview of large datasets. Their genius though is that you can drill into the table to explore the detail beneath the summary.

For example: you may have a dataset containing two year's worth of surgical data. Using a Pivot Table you could present the total number of procedures by speciality in each month. You can then drill into this overview to show another layer of detail say: the minimum, mean and maximum number of procedures for each day of the week or for each consultant.

Pivot Tables are worth exploring in depth. The best way to learn is to just start and play around.

Select all of the fields in your data set. Do make sure you include *all* of the fields. On the Insert Tab of the ribbon (not the data tab) select the Pivot Table button. From the Table Group. Select the option to create the table in a **New Worksheet**.



Now, click on the empty table and the field list appears. Drag names across from the field list and drop them into the table.

Just try it and see what happens. Playing really is the best way to get to grips with this facility. After you have had a go, read the Excel Help. It should make some more sense now.

NB: this utility embeds a copy of your data in the pivot table so any changes you make to the original data will not appear in your pivot table. Remember this because your pivot table will probably reveal errors in the data you need to correct. Having corrected them you will need to recreate the Pivot Table.

INTERMEDIATE EXCEL FUNCTIONS

In order to create insight from data, very often you have to generate values or correlate data to draw a validated conclusion. Below are the functions Saigei team members find most useful.

Use the Excel Help system (F1) to get detailed help with examples on how to use them.

VLOOKUP	Works like a grid reference to find a value ie it searches the first column of a table you have named (eg Boys), for a <i>value</i> you specify (eg John), then returns another value from the <i>column</i> you specify on the same row (eg birthday).
HLOOKUP	Works like Vlookup but across the first row in an array rather than down the first column.
COUNT/COUNTIF	The COUNT function counts the number of cells in a defined range that contain numbers. The COUNTIF function counts the number of cells in a defined range that meet the criterion you specify eg all the cells that start with a the letter "E" or that contain a number larger than a number you specify.
MAX/MIN	MAX Returns the largest value in a set of values. MIN Returns the smallest number in a set of values.
IF	IF() tests to see if a condition you have set is true. It then performs whatever instruction you have specified depending on the outcome. This is a very sophisticated function that is worth learning about in detail.
STDEV	Estimates standard deviation based on a selected sample. The standard deviation is a measure of how widely values are dispersed around the average value (the mean).
HOURL	Returns the hour of a time value. The hour is given as an integer, ranging from 0 (12:00 A.M.) to 23 (11:00 P.M.). Essential for mapping demand by hour.
DAY	Returns the day of a date, represented by a serial number. The day is given as an integer ranging from 1 to 31. Essential for mapping demand by day.
MONTH	Returns the month of a date represented by a serial number. The month is given as an integer, ranging from 1 (January) to 12 (December).
YEAR	Returns the year corresponding to a date. The year is returned as an integer in the range 1900-9999. This function is crucial if you want to calculate the number of years between two dates, eg calculating age on admission for a patient.

Today	Returns the serial number of the current date. The serial number is the date-time code used by Excel for date and time calculations. If the cell format was General before the function was entered, Excel changes the cell format to Date. The TODAY function is useful when you need to have the current date displayed on a worksheet, regardless of when you open the workbook. It is also useful for calculating intervals.
Indirect	This function is very useful if you have to perform a calculation first in order to know what value you need. It looks in a cell to see what value is in the cell and then treats this value as a name. So indirect(answer) looks in cell you specify. If the value is "salary_peter", it will then look up and return the value for salary_peter ie "£21,000".