Highline Excel 2016 Class 18: Clean & Transform Data: Replace, Flash Fill, Text To Columns, Formulas

Table of Contents

Clean Raw Data

1) Fix unusable raw data so that it can be used to perform data analysis.
   - Some Examples seen in this video:
     1. Replace Characters using Find & Replace
     2. Split First and Last Names using Flash Fill
     3. Inserting Characters using Flash Fill
     4. Split First and Last Names using Formulas
     5. Converting ISO Dates to Serial Number Dates with Text To Columns
     6. Converting Text Numbers to Number with Text To Columns

Transform Data Sets

1) Fix unusable data set so that it can be used to perform data analysis.
   - Some Examples seen in this video:
     1. Split Delimiter Data into Multiple Columns using Text To Columns.

Find & Replace feature (Clean Data)

1) What Find and Replace does:
   - Replaces one or more characters with a new set of one or more characters.
2) When to use Find and Replace:
   - Find and Replace is good when you have a simple set of characters you want to replace.
3) Find and Replace does not update when source data changes. If you want the solution to update, use Formulas or Power Query (Get & Transform).
4) Keyboard = Ctrl + H.
5) Steps:
   1. Highlight cells with data

Product Code
Carlota-3039
Aspen-3905
Yanaki-2309
FlatTop-3062
Carlota-3039

   -
   2. Home Ribbon Tab, Edit group, Find & Select drop-down, Find (keyboard = Ctrl + H)

   -
   3. Enter what to find and replace into Dialog Box. If you are sure that you want all dashes replace, click Replace All, otherwise click Find Next and examine each dash one at a time before Replace:
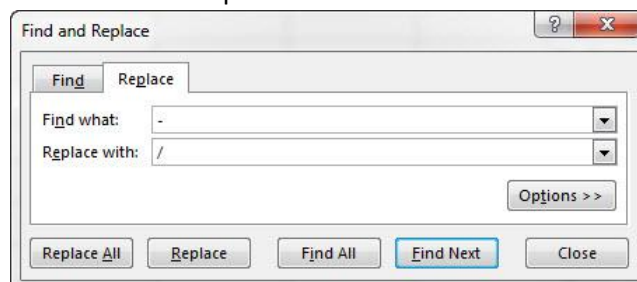
   -
   4. In this example, the only dashes used are between Product Name and number, so we use Replace All button and we get:

Product Code
Carlota/3039
Aspen/3905
Yanaki/2309
FlatTop/3062
Carlota/3039

   -

Flash Fill: Clean Data

1) What does Flash Fill do?
   - Can clean or transform data based on examples you give it next to the data set. Mostly Flash Fill is used to Clean Data.
   - Flash Fill = "Program by example" (term Microsoft uses to describe the feature).
     1. If you give it an example, behind the scenes Flash Fill builds a program to execute what you want to do to the data.
   - Flash Fill is not linked to source data.
   - Flash Fill does not update when source data changes.
2) BIG KEY:
   - You must know what sort of data is in the column and what sort of patterns there are in the data before you can give Flash Fill an accurate example.
3) When to use Flash Fill:
   - When you have a quick one-time data cleaning task.
   - If you need to have solution update when source data changes, use Formulas or Power Query (Get & Transform).
4) Flash Fill is in the Data Tools group in the Data Ribbon Tab
5) Flash Fill keyboard: Ctrl + E
6) Steps to using Flash Fill:
   1. You must know what sort of data is in the column and what sort of patterns there are in the data before you can give Flash Fill an accurate example.
   2. You give Flash Fill one or more examples next to the data set.
   3. Flash Fill sees your "Pattern", it will perform the action
7) Flash Fill actions include: Combine, Extract, Insert, Reverse
8) Flash Fill works on: Text, Numbers, Dates, Times

Example when there is consistent data and the one example we provide is unambiguous (Clean Data)

1. Data that needs to be cleaned or transformed. We see that the data always has 10 numbers that represent phone numbers.

	A	B
20	Phone	Insert
21	2065874545	
22	2536987455	
23	2159874439	
24	4156559933	
25	5105557845	

2. Next to the data set, we give flash fill an unambiguous example. Since the data is consistent and the pattern can be unambiguously demonstrated in one example, we only have to give it one example.

	A	B
20	Phone	Insert
21	2065874545	(206) 587-4545
22	2536987455	
23	2159874439	
24	4156559933	
25	5105557845	

3. We hit Enter.
4. Then either:
   i. Ctrl + E
   ii. Click Flash Fill button in the Data Tools group in the Data Ribbon Tab
   iii. Give it a second example and when we see the "Ghost" Flash Fill drop down, then we hit Enter.

8) When do we give it only one example?
- When the data is consistent and the example we provide is unambiguous.
- Many examples are straight forward and only need one example:
  1. Extracting first name from a column of first and last names.
  2. Combining first and last names from a column of first and last names.
  3. Inserting a lead apostrophe.

Example when there is NOT consistent data and therefore we need to give it more than one example (Clean Data)

1) Data that needs to be cleaned or transformed. We see that the data sometimes has a middle name and sometimes does not. If we want first initial for first name and last name, we must give it two examples:

Names	Initial for First & Last - give it enough examples
Sioux A. Ripper	S Ripper
Jon JJ Mac	
Gigi Tran	G Tran
Chin Smith	
Han Von Min	
Tyrone Binger	
Anne Sindy Van	

-
2) After we are sure that we know the data and we gave Flash Fill enough examples so that the pattern is unambiguous, we use the keyboard to get Flash Fill to work: Ctrl + E. The data should look like this:

Names	Initial for First & Last - give it enough examples
Sioux A. Ripper	S Ripper
Jon JJ Mac	J Mac
Gigi Tran	G Tran
Chin Smith	C Smith
Han Von Min	H Min
Tyrone Binger	T Binger
Anne Sindy Van	A Van

-

Using Ghost Flash Fill List

1) If there is a clear pattern that Flash Fill can see, after you type two examples, a "Ghost List" of possible solutions will appear. If you hit "Enter" Flash Fill will Fill the column as seen in these two steps:

Step 1: Type Two Examples and you see "Ghost List":

	Names	First
10		
11	MO DIM	Mo
12	GIGI TRAN	Gigi
13	ALTHEA BERTIN	Althea
14	KELLY GAMBLIN	Kelly
15	CHANDRA VALENZULA	Chandra
16	CODY CASTILLON	Cody
17	TYRONE BRAZIER	Tyrone
18	ALTHEA BUHL	Althea
19	DOLLIE MUNSEY	Dollie
20	ALLYSON PHOU	Allyson

Step 2: Hit Enter and the Ghost List fills the column:

	Names	First
10		
11	MO DIM	Mo
12	GIGI TRAN	Gigi
13	ALTHEA BERTIN	Althea
14	KELLY GAMBLIN	Kelly
15	CHANDRA VALENZULA	Chandra
16	CODY CASTILLON	Cody
17	TYRONE BRAZIER	Tyrone
18	ALTHEA BUHL	Althea
19	DOLLIE MUNSEY	Dollie
20	ALLYSON PHOU	Allyson

Text To Columns (TTC): Clean or Transform Data

1) What Text To Columns does:
   1. Split a single column of text into multiple columns based on Delimiter or Fixed Width:
   2. Convert Text Dates to Serial Number Dates
   3. Convert Text Numbers to Numbers
2) When to use:
   1. Flash Fill and Power Query mostly replace what Text To Columns (TTC) used to do.
   2. But there are still a few good uses for TTC such as:
      i. Convert Text Dates to Serial Number Dates
      ii. Convert Text Numbers to Numbers
      iii. Split text from one column into multiple columns and you do not need the output data to be linked to source data.
3) Text To Columns Keyboard: Alt, A, E
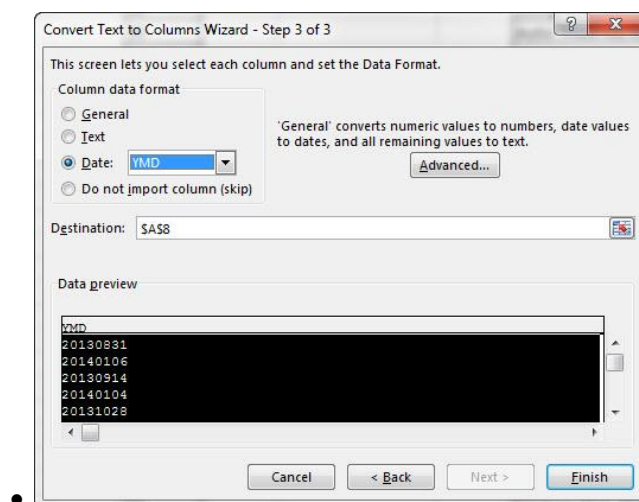4) Text To Column is found in Data Ribbon Tab, Data Tools group, Text To Columns button:
   -
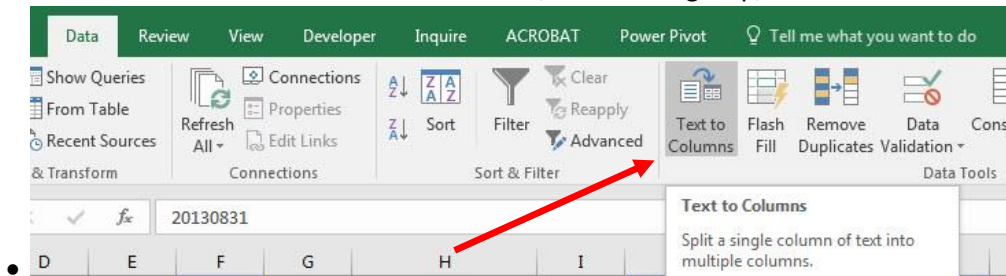
TTC to convert ISO Text Dates to Serial Number Dates (Clean Data):

1. Highlight column of data:
   -

2. Open Text To Columns 3-step dialog box with: Alt, A, E.
3. Click the Next button two times to get to step 3.
4. Select Date "YMD":
   -

5. Click Finish.

6. ISO Dates are now Serial Number Dates:

	A	B	C
15	**Dates**	**Sales Reps**	**Units**
16	8/31/2013	Chin Smith	33
17	1/6/2014	Gigi Tran	51
18	9/14/2013	Tyrone Binger	49
19	1/4/2014	Gigi Tran	39
20	10/28/2013	Gigi Tran	53

TTC to convert Text Numbers back to Numbers (Clean Data)

1. Highlight Text Numbers
2. Use Keyboard: Alt, A, E, Alt + F
   i. Mouse method would be: 1) Highlight Text Numbers, click Text To Columns button in Data Tools group in Data Ribbon Tab, then click Finish button in step 1 of the Text To Columns wizard.
3. Done!
4. Note the reason this works is because the default Data Type ("Column data format") in step 3 of the wizard is "General", which will automatically convert all text numbers back to real numbers.

TTC to split data apart into multiple columns based on a delimiter (Transform Data Set)

1. Goal:
   i. Use TTC to split data based on a delimiter, skip Products Column (2nd column).
   ii. Split text from one column into multiple columns and you do not need the output data to be linked to source data.
2. Highlight column of data:

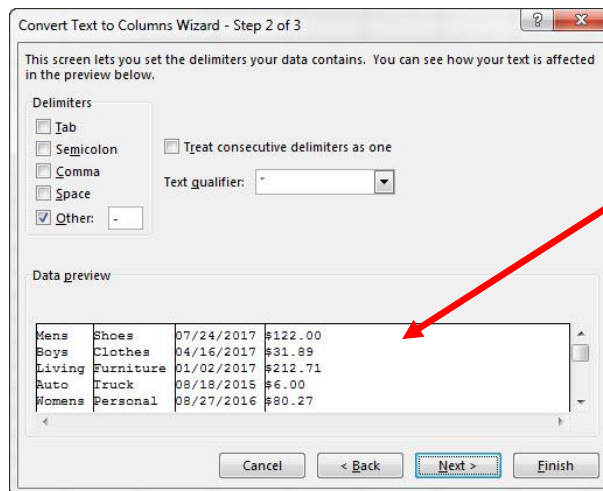|	J	K	L	M	N	O
13	Use TTC to split data based on a delimiter, skip Products Column (2nd column)					
14						
15	**Description**		**Department**	**Date**	**Amount**	
16	Mens-Shoes-07/24/2017-$122.00					
17	Boys-Clothes-04/16/2017-$31.89					
18	Living-Furniture-01/02/2017-$212.71					
19	Auto-Truck-08/18/2015-$6.00					
20	Womens-Personal-08/27/2016-$80.27					
21	Mens-Misc-04/03/2015-$75.13					

4. Alt, A, E, to open Text To Columns Wizard
5. Step 1 of Text To Columns wizard:
   i. Delimited = character that separates each field
   ii. Fixed Width = fields that have same number of characters (such as State Abbreviation has only two characters).
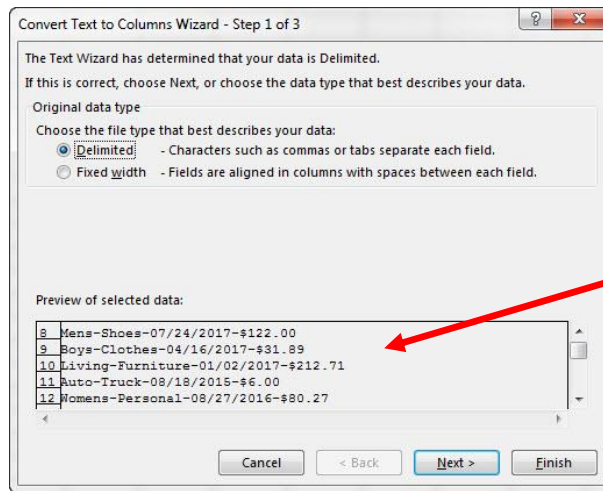
Notice preview of data.

4. Step 2 of Text To Columns wizard: list the delimiter (in our case a dash):

Notice preview of split data.

5. Step 3 of Text To Columns wizard:

Notice destination cell.

Notice skipped column.

6. Finished Data Set:

	Description		Department	Date	Amount
13	Use TTC to split data based on a delimiter, skip Products Column (2nd column)				
14					
15	**Description**		**Department**	**Date**	**Amount**
16	Mens-Shoes-07/24/2017-$122.00		Mens	7/24/2017	$122.00
17	Boys-Clothes-04/16/2017-$31.89		Boys	4/16/2017	$31.89
18	Living-Furniture-01/02/2017-$212.71		Living	1/2/2017	$212.71
19	Auto-Truck-08/18/2015-$6.00		Auto	8/18/2015	$6.00
20	Womens-Personal-08/27/2016-$80.27		Womens	8/27/2016	$80.27
21	Mens-Misc-04/03/2015-$75.13		Mens	4/3/2015	$75.13
22	Girls-Toys-08/03/2015-$66.00		Girls	8/3/2015	$66.00
23	Womens-Garments-08/09/2015-$20.00		Womens	8/9/2015	$20.00

Formulas to Clean or Transform Data

1) When to use formulas to clean and transform your data:
    - The data is in Excel already (you don't need to import it).
    - You do not have big data. When formulas have to work over very large ranges of cells, calculation time may slow down.
    - You want the solution to automatically update without having to use the Refresh button.
2) Text functions and formulas like LEFT, RIGHT, MID, SEARCH and others help to clean and transform data.
    - This topic was covered previously in this class in video:
        1. Highline Excel 2016 Class 08: Text Formulas and Text Functions to Join and Extract Data
        2. https://www.youtube.com/watch?v=rlGLP3qzNnw
3) Lookup functions and formulas like VLOOKUP, LOOKUP, INDEX and MATCH can be used to transform data sets by adding new columns to tables or flipping tables of data.
    - This topic was covered previously in this class in video:
        1. Highline Excel 2016 Class 11: Lookup Functions & Formulas, Comprehensive Lessons, 20 Examples
        2. https://www.youtube.com/watch?v=HqXEcu22EaY

Extract First and Last Name automatically when source data changes with Formula (Clean Data):

	A	B	C	D	E	F	G	H	I	J
8	**Data Cleaning Goal:**									
9	Data Dump always gives us first and last together									
10	Build template that will automatically extract first and last name.									
11										
12	**Name**	**First**	**Last**	**Quiz 1**	**Quiz 2**	**Quiz 3**	**Test 1**	**Test 2**	**Test 3**	**Total**
13	Sioux Radcoolinator	Sioux	Radcoolinator	22	12	17	84	61	26	222
14	Gigi G. Pham	Gigi	Pham	14	27	14	73	83	93	304
15	Tyrone Smith	Tyrone	Smith	22	29	29	90	91	90	351
16	Billy Bobo Banner	Billy	Banner	23	29	26	91	27	62	258
17										
18	Formula in B13: =LEFT(A13,SEARCH(" ",A13)-1)									
19	Formula in C13: =TRIM(RIGHT(SUBSTITUTE(A13," ",REPT(" ",LEN(A13))),LEN(A13)))									

1) Esc Key:
    i. Closes Backstage View (like Print Preview).
    ii. Closes most dialog boxes.
    iii. If you are in Edit mode in a Cell, Esc will revert back to what you had in the cell before you put the Cell in Edit mode.
2) **F2 Key** = Puts formula in Edit Mode and shows the rainbow colored Range Finder.
3) SUM Function: **Alt + =**
4) **Ctrl + Shift + Arrow** = Highlight column (Current Region).
5) **Ctrl + Backspace** = Jumps back to Active Cell
6) **Ctrl + Z** = Undo.
7) **Ctrl + Y** = Undo the Undo.
8) **Ctrl + C** = Copy.
9) **Ctrl + X** = Cut.
10) **Ctrl + V** = Paste.
11) **Ctrl + PageDown** =expose next sheet to right.
12) **Ctrl + PageUp** =expose next sheet to left.
13) **Ctrl + 1** = Format Cells dialog box, or in a chart it opens Format Chart Element Task Pane.
14) **Ctrl + Arrow**: jumps to the bottom of the "**Current Region**", which means it jumps to the last cell that has data, right before the first empty cell.
15) **Ctrl + Home** = Go to Cell A1.
16) **Ctrl + End** = Go to last cell used.
17) Alt keyboards are keys that you hit in succession. Alt keyboards are keyboards you can teach yourself by hitting the Alt key and looking at the screen tips.
    i. Create PivotTable dialog box: **Alt, N, V**
    ii. Page Setup dialog box: **Alt, P, S, P**
    iii. Keyboard to open Sort dialog box: **Alt, D, S**
18) **ENTER** = When you are in Edit Mode in a Cell, it will put thing in cell and move selected cell DOWN.
19) **CTRL + ENTER** = When you are in Edit Mode in a Cell, it will put thing in cell and keep cell selected.
20) **TAB** = When you are in Edit Mode in a Cell, it will put thing in cell and move selected cell RIGHT.
21) **SHIFT + ENTER** = When you are in Edit Mode in a Cell, it will put thing in cell and move selected cell UP.
22) **SHIFT + TAB** = When you are in Edit Mode in a Cell, it will put thing in cell and move selected cell LEFT.
23) **Ctrl + T** = Create Excel Table (with dynamic ranges) from a Proper Data Set.
    i. Keyboard to name Excel Table: **Alt, J, T, A**
    ii. **Tab** = Enter Raw Data into an Excel Table.
24) **Ctrl + Shift + ~ (`)** = General Number Formatting Keyboard.
25) **Ctrl + ;** = Keyboard for hardcoding today's date.
26) **Ctrl + Shift + ;** = Keyboard for hardcoding current time.
27) **Arrow Key** = If you are making a formula, Arrow key will "hunt" for Cell Reference.
28) **Ctrl + B** = Bold the Font
29) **Ctrl + * (on Number Pad)** or **Ctrl + Shift + 8** = Highlight Current Table.
30) **Alt + Enter** = Add Manual Line Break (Word Wrap)
31) **Ctrl + P =** Print dialog Backstage View and Print Preview
32) **F4 Key** = If you are in Edit mode while making a formula AND your cursor is touching a particular Cell Reference, F4 key will toggle through the different Cell References:
    i. **A1** = Relative
    ii. **A1** = Absolute or "Locked"

iii. **A$1** = Mixed with Row Locked (Relative as you copy across the columns AND Locked as you copy down the rows)

iv. **$A1** = Mixed with Column Locked (Relative as you copy down the rows AND Locked as you across the columns)

33) **Ctrl + Shift + 4** = Apply Currency Number Formatting

34) **Tab key** = When you are selecting a Function from the Function Drop-down list, you can select the function that is highlighted in blue by using the Tab key.

35) **F9 Key** = To evaluate just a single part of formula while you are in edit mode, highlight part of formula and hit the F9 key.

  i. If you are creating an Array Constant in your formula: Hit F9.

  ii. If you are evaluating the formula element just to see what that part of the formula looks like, REMEMBER: to Undo with Ctrl + Z.

36) **Alt, E, A, A** = Clear All (Content and Formatting)

37) Evaluate Formula One Step at a Time Keyboard: **Alt, M, V**

38) Keyboard to open Sort dialog box: **Alt, D, S**

39) **Ctrl + Shift + L** = Filter (or **Alt, D, F, F**) = Toggle key for Filter Drop-down Arrows

40) **Ctrl + N** = Open New File

41) **F12** = Save As (Change File Name, Location, File Type)

42) Import Excel Table into Power Query Editor: **Alt, A, P, T**

43) **Ctrl + 1 (When Chart element in selected):** Open Task Pane for Chart Element

44) **F4 Key** = If you are in Edit mode while making a formula AND your cursor is touching a particular Cell Reference, F4 key will toggle through the different Cell References:

  i. **A1** = Relative

  ii. **A1** = Absolute or "Locked"

  iii. **A$1** = Mixed with Row Locked (Relative as you copy across the columns AND Locked as you copy down the rows)

  iv. **$A1** = Mixed with Column Locked (Relative as you copy down the rows AND Locked as you across the columns)

45) Keyboard to open Scenario Manager = **Alt, T, E**

46) **Ctrl + Tab** = Toggle between Excel Workbook File Windows

47) **Ctrl + Shift + F3** = Create Names From Selection

48) **Ctrl + F3** = open Name Manager

49) **F3** = Paste Name or List of Names

50) **Alt + F4** = Close Active Window

51) **Window Key + Up Arrow** = Maximize Active Window

52) **Ctrl + Shift + Enter** = Keystroke to enter Array Formulas that: 1) have a function argument that requires it, or 2) whether or not you are entering the Resultant Array into multiple cells simultaneously.

53) **Ctrl + /** = Highlight current Array

54) Data Validation Dialog Box: **Alt, D, L**

55) **F11** = Create Chart on a new sheet

56) **Alt + F11** = Create Chart on currently selected sheet.

57) New Format Rule dialog box: **Alt, H, L, N**

58) Delete conditional Formatting Rule: **Alt, O, D, D**

59) Manage Rule dialog box keyboard: **Alt, O, D**

60) "Format values where this formula is true": **Alt, H, L, N, PageDown, Tab**

61) Shift + F11 = Insert a New Sheet

62) Ctrl + F1 = Toggle Ribbon Tabs on and off

63) Ctrl + Alt + F5 = Refresh All Data in Excel Workbook.

64) Zoom to Selection = **Alt, W, G**

New Keyboards in This Video:

65) Ctrl + F = Find
66) Ctrl + H = Find and Replace