Excelling with Excel: Advanced Excel Functions for Collection Analysis

Denise Pan
University of Colorado Denver, dpan@uw.edu

Gabrielle Wiersma
University of Colorado Boulder

Follow this and additional works at: https://docs.lib.purdue.edu/charleston

Part of the Library and Information Science Commons

An indexed, print copy of the Proceedings is also available for purchase at: http://www.thepress.purdue.edu/series/charleston.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences.

http://dx.doi.org/10.5703/1288284315327

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Excelling with Excel: Advanced Excel Functions for Collection Analysis

Denise Pan, Associate Director, Technical Services, University of Colorado Denver
Gabrielle Wiersma, Head of Collection Development and Assessment, University of Colorado Boulder

Abstract

Microsoft Excel offers useful features and formulas that potentially allow acquisitions and collection development librarians to work smarter, not harder. Using journal cancellations as a workplace scenario, the presenters will provide attendees with ideas on how to organize data and complete basic calculations with Excel. The presenters provide examples on how to use several advanced Excel functions including PivotTables, VLOOKUP, and select formulas. They describe the steps for importing and exporting data, combing and comparing data from different sources, and formatting data to communicate more effectively.

Introduction

At the University of Colorado (CU), librarians from CU Denver and CU Boulder campuses collaborate to purchase shared collections. Through their joint acquisitions initiatives, they have found many opportunities to save time and improve efficiencies by applying Excel functionality to their work. The Charleston Conference preconference and proceedings are an opportunity to share their knowledge and experiences with a wider audience. Their goal is to highlight Excel features and functionality with a hypothetical scenario of a serials cancellation project with ordering and usage data. This scenario was selected because it is an activity that most libraries have completed or may need to do so in the future. The step for analyzing journal titles could be applied to other projects such as assessing patron-driven purchases, year-end expenditures, or serials renewals.

Specifically, these proceedings describe the usefulness of the following Excel features and functions:

- Importing Text File Data
- Inserting Tables
- Highlighting and Removing Duplicates
- VLOOKUP and IFERROR
- Conditional Formatting
- PivotTables and PivotCharts

Assessment Goals

The most important first step in any assessment project is defining goals and objectives in order to measure progress. To demonstrate Excel, the project goal is to identify which journal titles are candidates for cancellation from one publisher package. The objective is to cancel approximately $7,000, or 8%, of the journals. This will be accomplished by showing readers how to create a spreadsheet with information from various data sources, calculate cost per use, and analyze a set of journals based on usage. With this knowledge, readers could apply the same techniques to other data, such as impact factor or faculty input.

Gathering and Importing Data

The next step is considering what types of data to collect. For a serials review project, data sources may include payment information from the Integrated Library System (ILS) and Journal Citation reports for impact factor in text format (.txt) and usage data or price lists from publishers that can be downloaded in Excel format (.xls). If the data are only available as a text file, it will need to be imported into Excel using the Text Import Wizard.

1. To get it into Excel, click on the Data tab
2. Get External Data → From Text
3. Select the file you want to import
4. The Text Import Wizard will open after you select a file
Inserting Tables

An easy way to format your data is to convert it to a table. Even though it seems like data entered in an Excel spreadsheet is already a table, converting data into a table is actually a separate process.

By converting data into a table, Excel automatically creates a relationship between the data and automatically adds several features that will make it easier to format and analyze the worksheet. Tables eliminate the need to manually change the look and feel of spreadsheets, or adding filters to the top row. Tables also create Total Rows in the last row of a table which are useful for inserting many simple formulas such as subtotals (sum), average, counts, minimum (min), and maximum (max).

1. To create a table in Excel, click on the Insert tab ➔ Table
2. Select the cells that have data

Highlight and Remove Duplicates

Data sources may have duplicate values that need to be identified and removed. In the ILS data, the journal title, order record numbers, and ISSNs should be unique values versus the subject area, assigned collection development librarian, fund, and even cost may be repeated. Use Conditional Formatting to highlight duplicate values within a column.

1. Highlight the title and order record columns.
2. Select Conditional formatting ➔ Highlight Cell Rules ➔ Duplicate Values and select a color scheme.
3. Manually delete the duplicate values. Once deleted, the unique value will no longer be highlighted. Alternatively, if there are many duplicates, use the Remove Duplicates function in Excel.
VLOOKUP

For a journal cancellation project you often want to connect data from different sources. For example, a title list with pricing information from a publisher or ILS data with COUNTER usage stats. VLOOKUP is a formula that you can use to look up values in a list or table. It searches for a value in a table and returns a corresponding value in another column in the same row. VLOOKUP works best if there is a unique identifier to match data from different spreadsheets. Fortunately, most ILS data, title lists from publishers, and COUNTER usage reports include ISSNs which can be used as a match point to connect the data. Using the ISSN as the unique lookup value, VLOOKUP can find and enter the YTD usage statistics from a COUNTER JR1 report into a spreadsheet with your ILS data. In Excel, click on the Formulas tab ➔ Lookup & Reference ➔ VLOOKUP

1. **lookup_value** is the unique identifier that is in both spreadsheets (ISSN)
2. **Table_Array** tells Excel where to look for the data (drag to select the cells to search)
3. **Col_Index_Number** is the number of the column you want to pull data from and display in your spreadsheet (count the columns from the ISSN [1] to the YTD total [14])
4. **Range lookup**
   a. **TRUE** = approximate match (this is useful for titles or other fields that might have slight variation)
   b. **FALSE** = exact match (this is useful if you only want an exact match and works well if you are using a numerical match point like ISSN)

VLOOKUP could also be used to connect additional data like impact factor or a numerical ranking from faculty (e.g., 1 = Essential/core, do not cut; 2 = Important, only cut if absolutely
Calculating Cost Per Use with IFERROR

Cost-per-use is one metric that can be useful for evaluating journal packages. Excel can calculate cost per use by dividing the subscription costs by the YTD usage statistics. However, if a journal had 0 use, Excel will display an error message, #DIV/0! because it cannot divide the subscription costs by 0. This error message can be interpreted as a title with zero use, but it is also possible to use another Excel formula called IFERROR to calculate cost per use and change the display of the error message. For example, IFERROR can display another value like n/a, –, or the subscription cost instead of displaying #DIV/0!.

1. In Excel, click on the Formulas tab ➔ Logical ➔ IFERROR
   a. Value include the formula =subscription costs/use
   b. Value_if_error include any value (–, n/a) or select the subscription cost

Conditional Formatting

Conditional formatting can also be used to highlight and graphically present data. For example, Excel can create data bars that represent how much each journal title costs compared to the total subscription costs. Conditional formatting also contains various icons that can be used to visually represent data intervals. For example, red, yellow, and green icons can be added to cost per use to indicate high (>\$30+), medium ($11–29), and low values (<$10) compared to average interlibrary loan or pay-per-view article costs. It can also highlight the top/bottom which is useful for identifying the most expensive journals and the journals with the highest cost-per-use, which would be likely candidates for cancellation. Adding graphs and icons to data helps quickly identify patterns and interpret the data without having to create separate graphs or charts.
Figure 7. Conditional Formatting

**PivotTables**

A PivotTable report is an interactive way to summarize large amounts of data very quickly. Some notable examples include:

- Subtotal and summarize data by categories and subcategories
- Expand and collapse levels of data to focus results
- Drill down to details from the summary data

Before building a PivotTable report, however, data should be in an Excel worksheet in a list or table format. It is essential that the data are set up correctly. Otherwise, the PivotTables will not work. The presenters suggested creating an “All Data” worksheet that combined multiple years of COUNTER JR1 usage data that shows the number of successful full-text article requests for journals into one Excel spreadsheet. Please keep the following tips in mind before creating an “All Data” worksheet with multiple years of JR1 usage data.

- Keep a copy of original source data in workbook
- Differentiate data by adding “Year” column
- Do not copy “Total for all journals” row—infates usage
- Name each column in row 1 (create header)
- Mind your column headings (carefully copy and paste)
- No blank rows or column—must provide column names

With PivotTables, researchers are able to see comparisons, patterns, and trends to answer unanticipated questions about their data. Using the scenario of a journal cancellation project, the presenters demonstrated how to use PivotTables to summarize multiple years of usage data and analyze usage trends. Specifically, PivotTables can be created to answer these questions:

- What is the total usage for all journals by year?
- What is the use for each journal title by year?
- Which titles had zero uses?
- Which titles were used the most?

**Summary by Year PivotTable**

The first example, Summary by Year PivotTable, can be used to answer the question “What is the total usage for all journals by year.” This PivotTable summarizes usage data for each year. The years can be expanded to see the individual journal titles and their usage for that year. For example, in the sample data there was no use for Aerospace Engineering in 2010. In that same year, the title Art Review had 96 total uses, 16 were HTML and 80 were for PDF.
The Use by Title PivotTable can answer the question “What is the use for each journal title by year?” Start by making a copy of the Summary by Year worksheet and renaming it as Use by Title. By pivoting or flipping the year and title, to title and year, 4 years of usage can be displayed for Art Review.

Zero Use Titles PivotTable

To answer the question, “Which journal titles were not used?” try creating a Zero Use Titles PivotTable. Make a copy of the report you just made and rename it as Zero Use Titles. Use a report filter to select journal titles with 0 uses by clicking on the report filter icon. Deselect all, check 0, and click OK. In the sample data, the journal title Chemical Reaction had 0 uses in 2 years, 2010 and 2013.

Top 10 Titles PivotTable

The fourth PivotTable, Top 10 Titles PivotTable, answers the question “Which journal titles were used the most?” Value filters can be used to identify the Top 10 journal titles by total number of successful full-text article requests by YTD total. Create a copy of the Zero Use Title report and rename as Top 10 Titles. Use the report filter to select all except 0 and blank. Click on the filter icon on the row labels, which displays a drop down for Value filters. Top 10 is at the bottom of the list. In the Top 10 filter window, change the drop down from Count of Title to Sum of YTD Total.
As the saying goes, “A picture is worth a thousand words.” Charts provide a visual representation of data. They show big-picture trends and relationships between different series of data in a graphical format. Similarly, PivotChart can help you see comparisons and patterns from PivotTable report summary data. To demonstrate PivotTables, we analyzed the usage data from a fictitious journal package.

When we create a PivotTable, Excel automatically creates a PivotChart.

1. Click the PivotTable to display the PivotTable Tools, and the Options tab.

2. Click PivotChart.

3. In the Insert Chart dialog box, click the Column chart type and click OK. The Column chart is useful for showing data changes over a period of time or for illustrating comparisons.
The source data for a PivotChart is the PivotTable. Any changes to the field layout and data in the associated PivotTable report is immediately reflected in the PivotChart report. In addition, a PivotChart can be created manually from scratch the same way you create PivotTable.

**Conclusion**

Far from a comprehensive description or instructions on Excel features and functions, the “Excelling with Excel: Advanced Excel Functions for Collection Analysis” proceedings are intended to highlight the application of Excel in acquisitions and collection development activities. For more details on procedures, see the slides posted on the online program page for this presentation at http://sched.co/15Azgp.