

Google Sheets Normalization Exercise

1. Go to: <https://docs.google.com/spreadsheets/d/1f8GrvrHw-FAhD0xyGKf96hawYtIFwonihlpsd6jiwgl/edit?usp=sharing>
 - a. This link opens a Google Sheets document called "Normalization Exercise."
2. In "Normalization Exercise" document, click File - Make a copy.
 - a. Otherwise we'll all be making changes to the same spreadsheet!
 - b. You need to be signed in to Google, because it needs somewhere to save the document.
 - c. Give it a new name.
3. Task: **Trimming White Spaces in the Article_Title Column**

There is extra white space in our Article Titles. This can happen when the data was extracted or copied from another location. We're going to use the TRIM formula to remove the white space.

 - a. paste the formula =**TRIM(Sheet1!A2)** into Sheet 2, cell A2.
 - b. Copy cell A2. Select the remaining rows in column A (rows 3-20).
 - c. Go to Edit --> Paste Special --> Paste Formula Only.
 - d. Column A should now be populated with your article titles, with whitespace removed.
4. Task: **Change date format**
 - a. Copy the date column from sheet 1 and paste into sheet 2, as column B
 - b. In sheet 2, select the entire column B.
 - c. In the menu bar, go to Format --> Number --> and then select Date
 - d. To customize your date format, go to Format --> Number --> More Formats --> "More date and time formats"
 - e. Note that some of the dates did not change; will need to clean that up by hand.
5. Task: **Correct Capitalization for Author Names**
 - a. type formula =**PROPER(Sheet1!C2)** into Sheet 2 C2
 - b. copy cell C2, and go to Edit --> Paste Special --> Formula Only into the remaining rows of column C (rows 3-20)
6. Task: **Data Validation**

There are multiple ways to validate data in your spreadsheet. Here, we'll create a list of items, and use that to validate the spreadsheet entries.

- a. Select the column D - Institution.
- b. Copy column to Sheet 2
- c. Select Data --> Data Validation
- d. In the pop-up window, select Criteria: "List of items."
- e. Enter the following comma-separate list:
 - i. University of Minnesota, University California Berkeley, University of Michigan, Indiana University Bloomington, University of North Carolina at Chapel Hill, Michigan State University, University of Kansas, University of Texas at Austin, Stanford, Georgia Institute of Technology

- f. Check "show dropdown list in cell" and select Show warning. Save.
 - g. Go through the cell rows and select the appropriate value for the field.
7. Task: **Correct Capitalization for Place Names - Uppercase**

- a. Use UPPER function to make all letters uppercase :
 - i. In Sheet 2, Cell E2, type the formula: **=UPPER(Sheet1!E2)**.
 - ii. copy the cell and paste special - formula only into cells E3-E20.

8. Task: **Splitting Text**

If your data is structured the same in each column, splitting your data into multiple columns will be simple. If it is structured differently you'll need to perform some manual clean-up.

- a. Select Column E (Institution_Location) in Sheet 1, copy and paste into Column E of Sheet 2
 - b. With Column selected, go to the menu and click Data - Split text to columns...
 - c. In the drop-down menu that pops up on the spreadsheet, choose your separator - in this case, it will work automatically because all data is structured the same.
 - d. Delete your original Column E (Institution Location) and name your new columns INST_CITY and INST_STATE
9. Task: **Generate Geocodes inside of Google Sheets**
- a. In Sheet 2, click Add-ons --> "Get add-ons" on the top menu
 - b. Search for "Geocode by awesome table" and add it to your sheets.
 - c. Select the City/State columns and go to Add-ons --> Geocode by Awesome Table --> Start Geocoding
 - d. Click "Are your addresses in multiple columns?" and select the City and State fields. It will concatenate the fields back to one column. That's okay, you can delete that later if you want separate columns.
 - e. Click Geocode! It will take a few moments.
 - f. When it's done, close the Geocode box and you will see the new Latitude and Longitude columns.
10. Task: **Add a Primary Key**
11. Finally, tidy it up using your domain knowledge