**Exemplar: Standardising text-based data**

**Overview**

In the exercise *Standardizing text-based data* you were asked to put into practice what you have learned about text functions in Microsoft Excel.

Your task in this exercise was to create formulas to tidy the data in the spreadsheet. You had to use these formulas to complete the following action items:

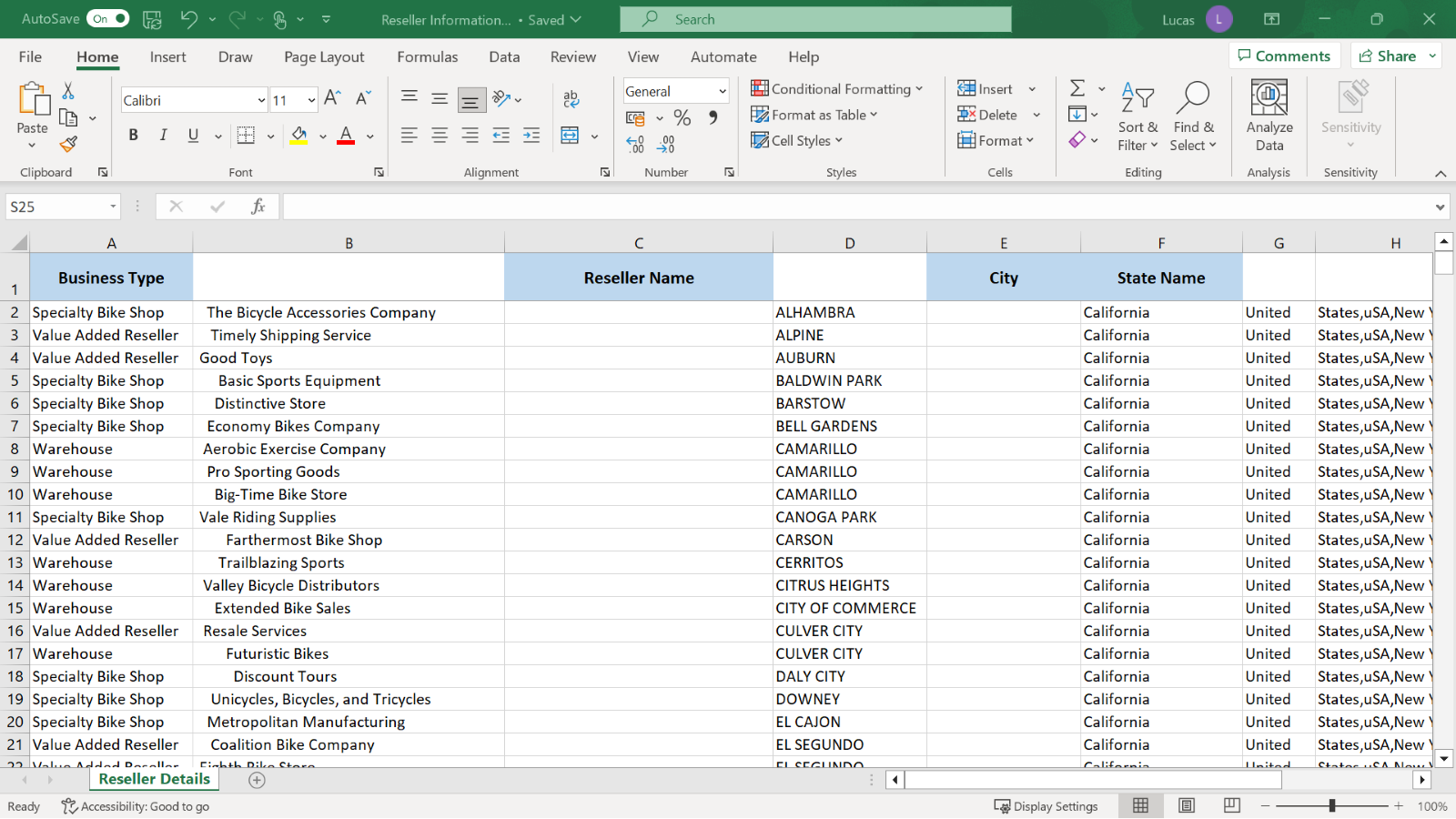
* Remove unnecessary spaces in entries.
* Change the case of entries.
* Extract portions of an entry to display it in a different column.
* Combine entries from different columns.

This reading provides you with a step-by-step guide for identifying these results. It also includes screenshots that you can compare against your own work.

You can review the formula creation techniques for this exercise in the reading *Functions that can be used to standardize text data***.**

**Step 1: Download the file**

* Download and open the Excel workbook **Reseller Information.xlsx.** The workbook contains one worksheet called **Reseller Details**that includes the names and locations of resellers.



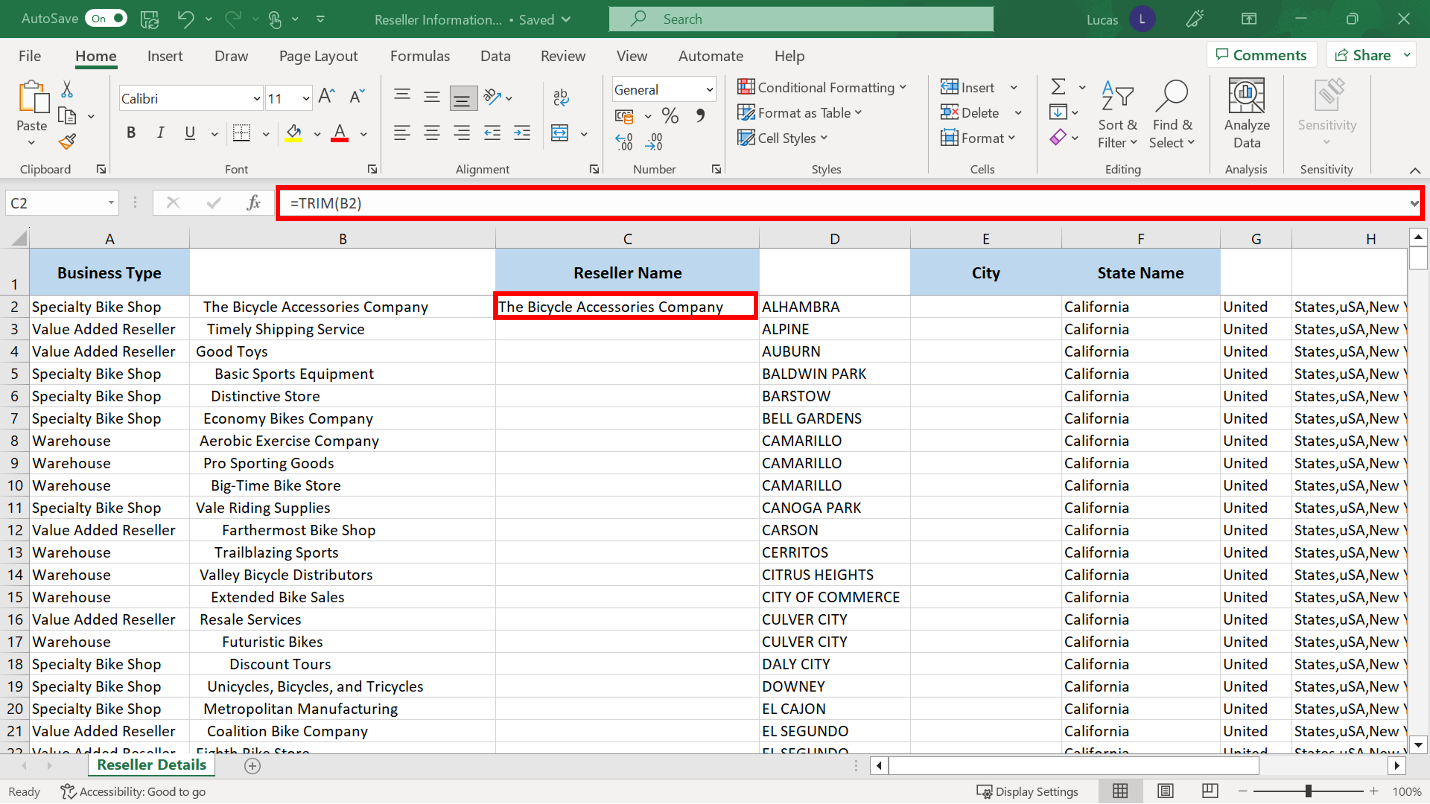
**Step 2: Creating the Calculations**

1. In **C2** you created a formula using the **TRIM** function to remove the redundant spaces before and after the entry in B2. The formula removes spaces before or after the piece of text but retains the necessary spaces between the words. The formula in **C2** should read:

**=TRIM(B2)**

The result generated in **C2** should be:

The Bicycle Accessories Company



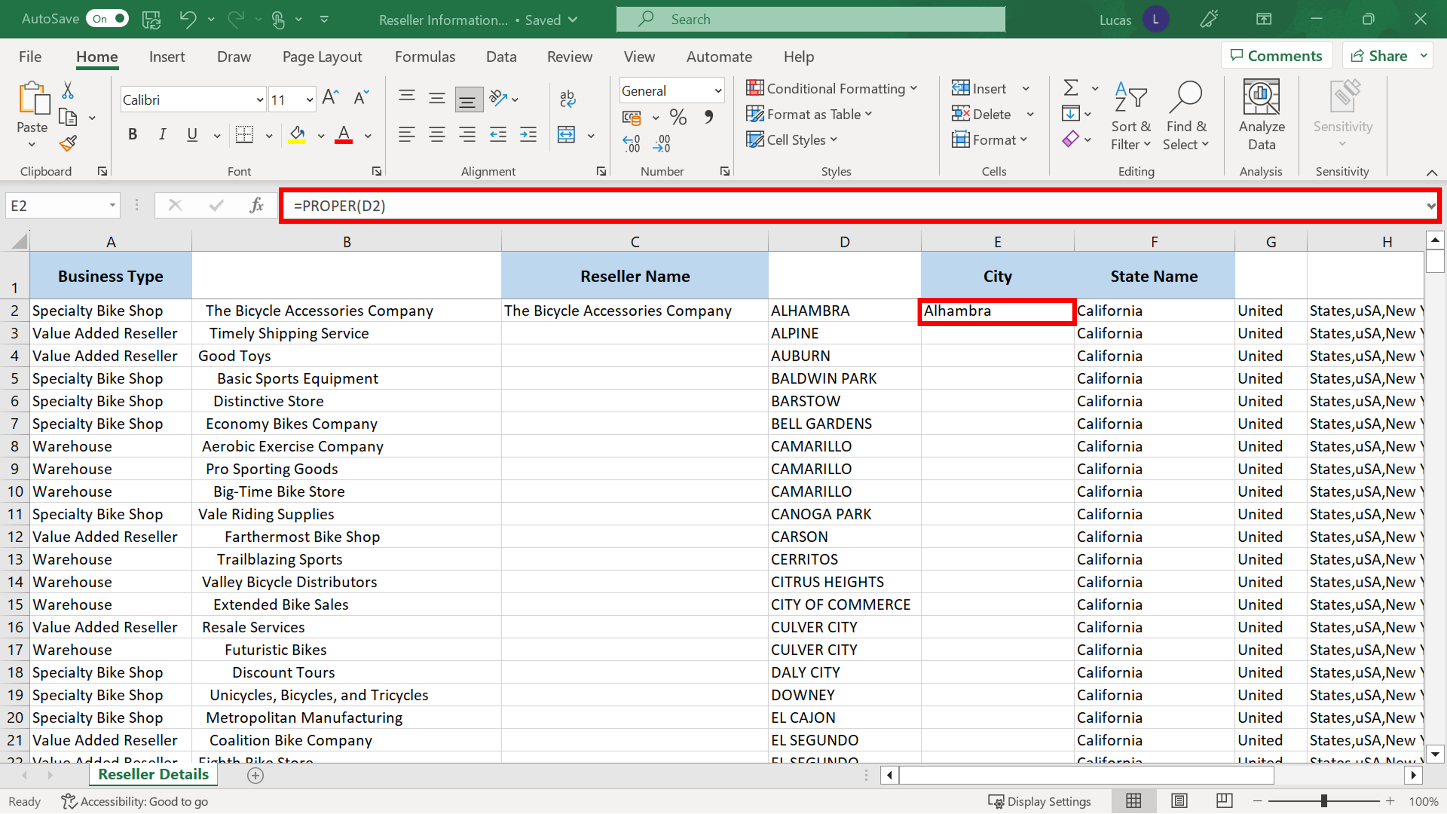
1. The city names in column **D** had all been entered using block capitals. The city name should be lowercase but still have a capital letter at the beginning of the word. The correct function to achieve this is **PROPER**.

You created a formula in **E2** to achieve this. The formula should read:

**=PROPER(D2)**

The result should read:

Alhambra



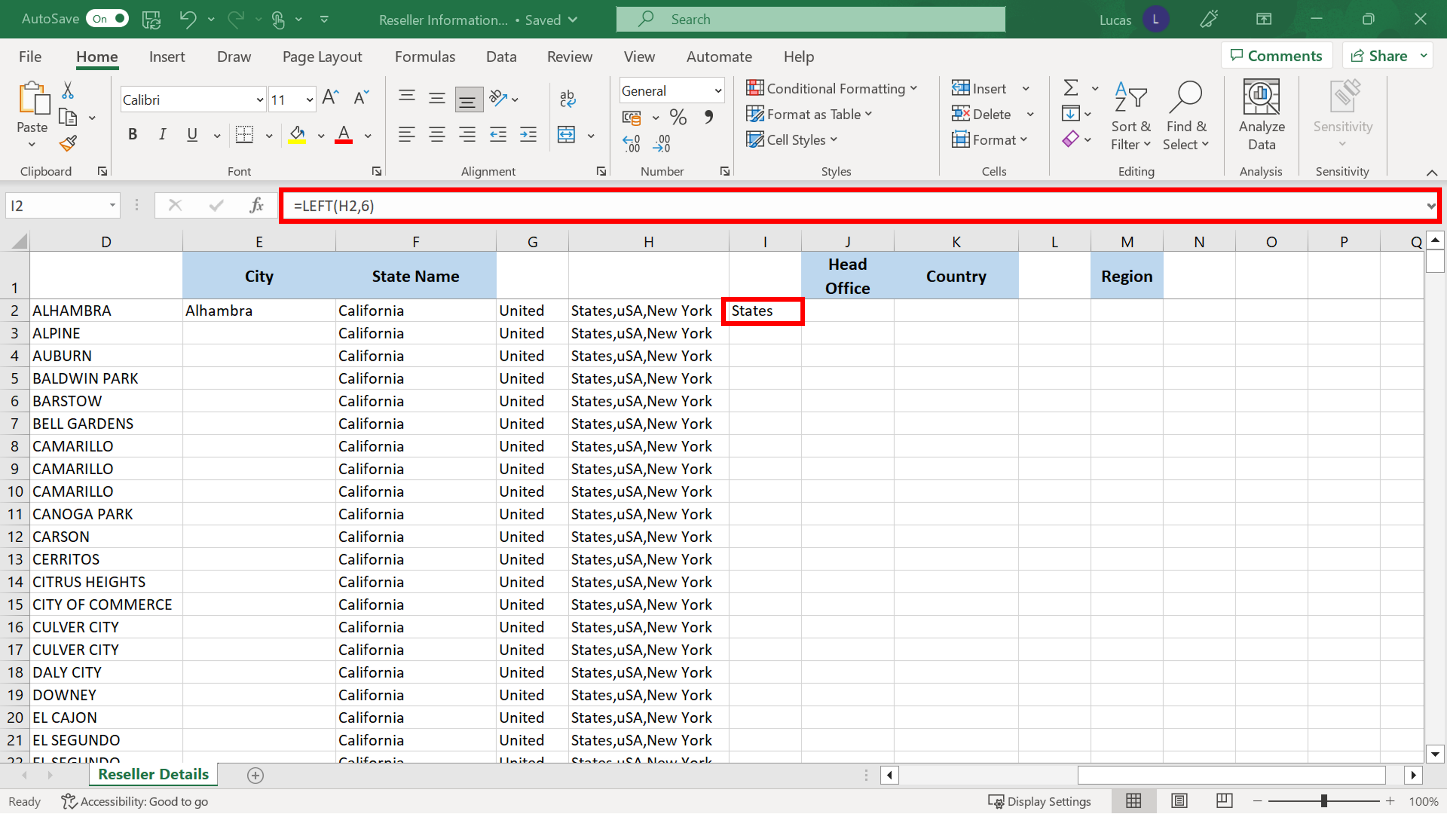
1. Column **H** in the worksheet contained text strings that were incorrectly joined together in a download process. You needed to extract the word states from the entry in **H2**. Since the word States was at the beginning of the text string, it was possible to use the **LEFT** function in a formula to extract the six characters of that word and display it in cell **I2**.

You created a formula in cell **I2** that reads:

**=LEFT(H2,6)**

The result should read:

States



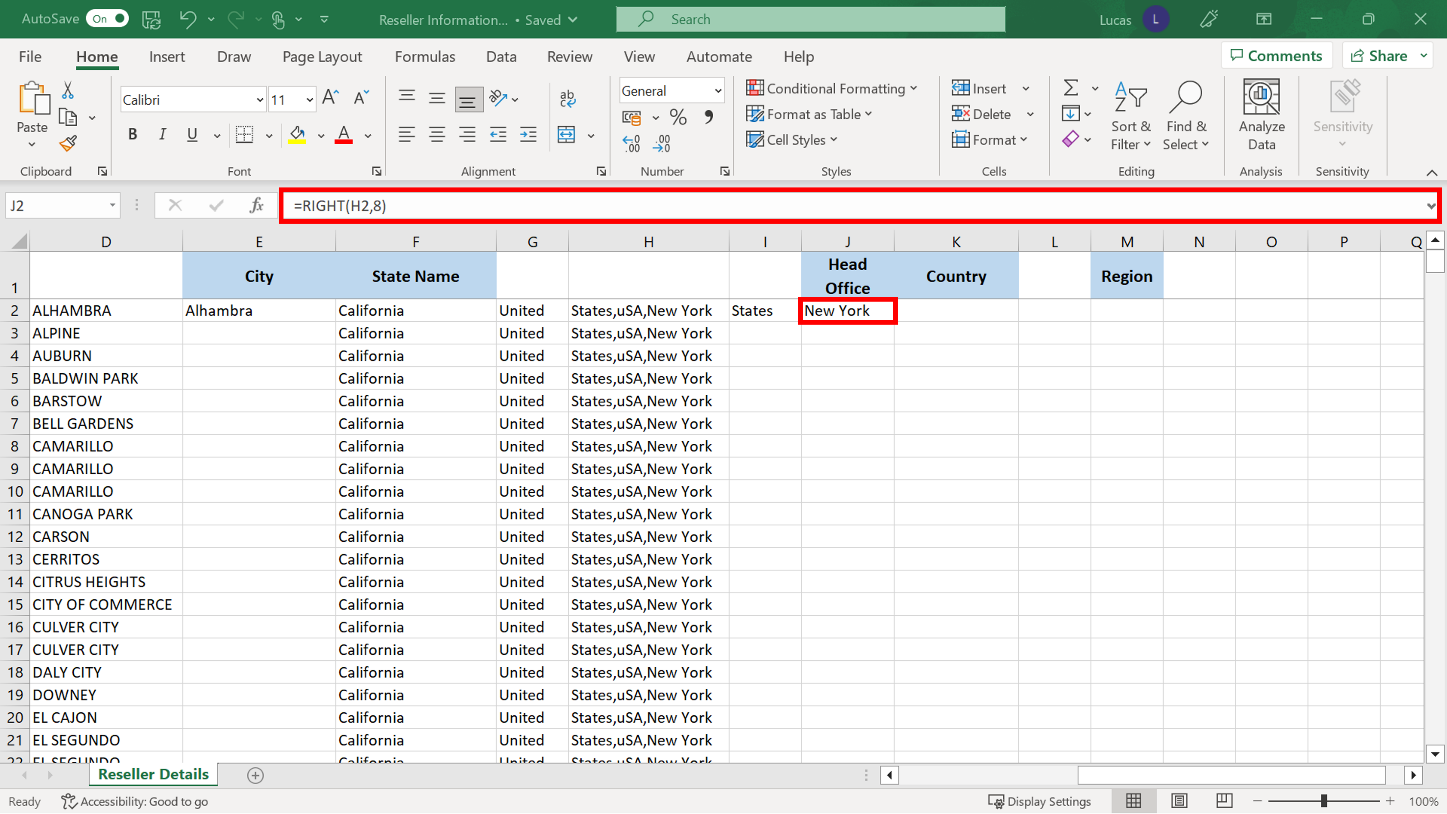
1. The next task was to extract the words New York from the entry in cell **H2**. The words New Yorkwere to the right of the cell. In **J2** you created a formula that included the **RIGHT** function to extracts 8 characters from the entry in **H2** (counting from the right side of the cell).

The formula in **J2** should read:

**=RIGHT(H2,8)**

The result should reads

New York



1. The entry in cell **G2**, United, and the entry in cell **I2**, States, needed to be combined. In **K2** you created a formula using the **CONCAT** function to join the contents of these two cells together. The **CONCAT** function requires cell references as its arguments.

However, the formula in **K2** should not have been written as:

**=CONCAT(G2, I2)**

This formula is incorrect because it generates a result where the two words are joined together with no space between them resulting in the entry:

UnitedStates

In this case, for the result to be displayed correctly, a third argument was necessary to indicate that a space should be included between the two words.

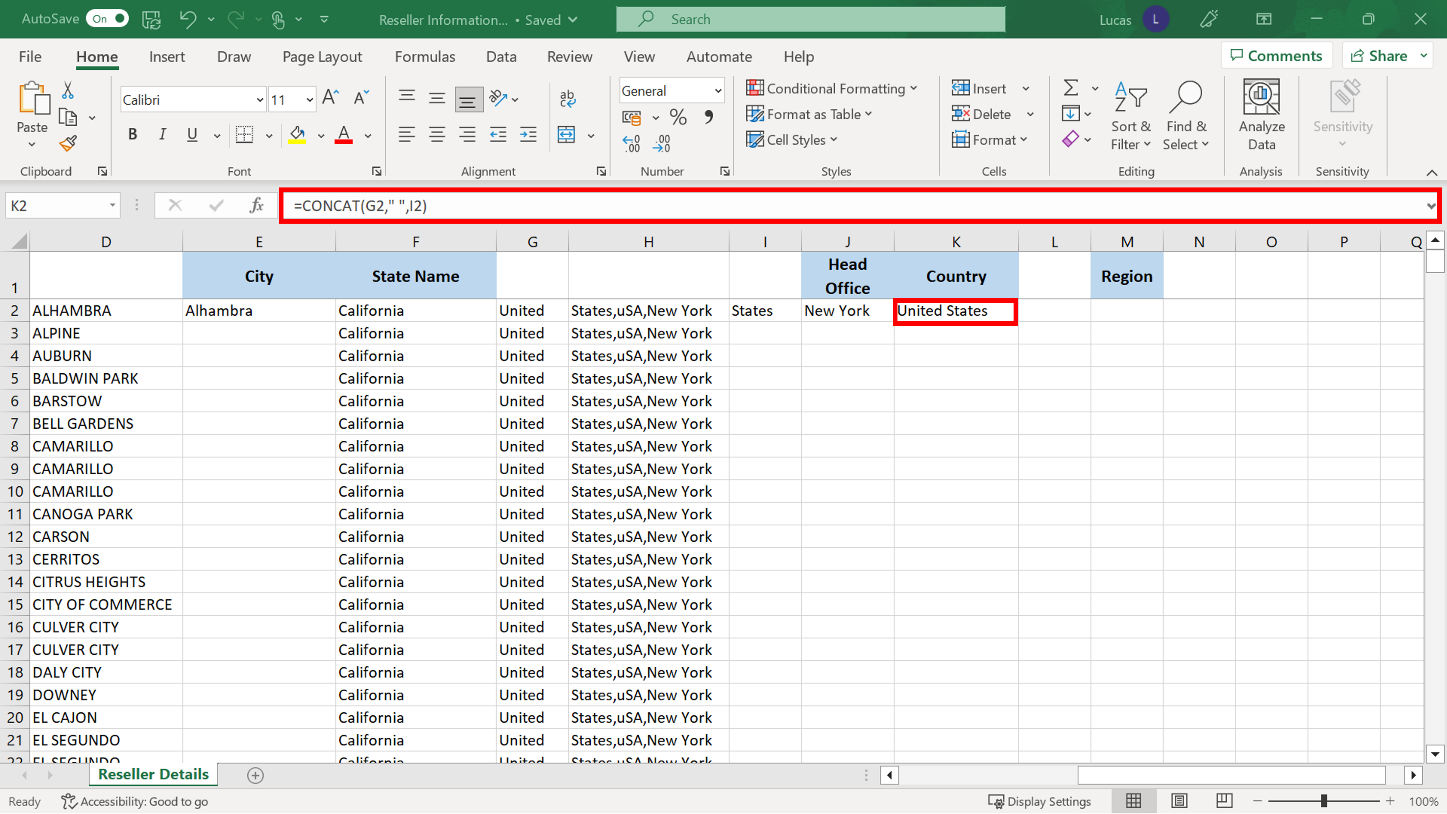
So, the formula in **K2** should instead read as:

**=CONCAT(G2,” “,I2)**

The double quote characters before and after the space character indicate that it is text data which must be included.

The result in **K2** should read:

United States



1. The text string in **H2** had the letters **uSA** in the middle of the longer text string. You created a formula in **L2** using the **MID** function to extract this.

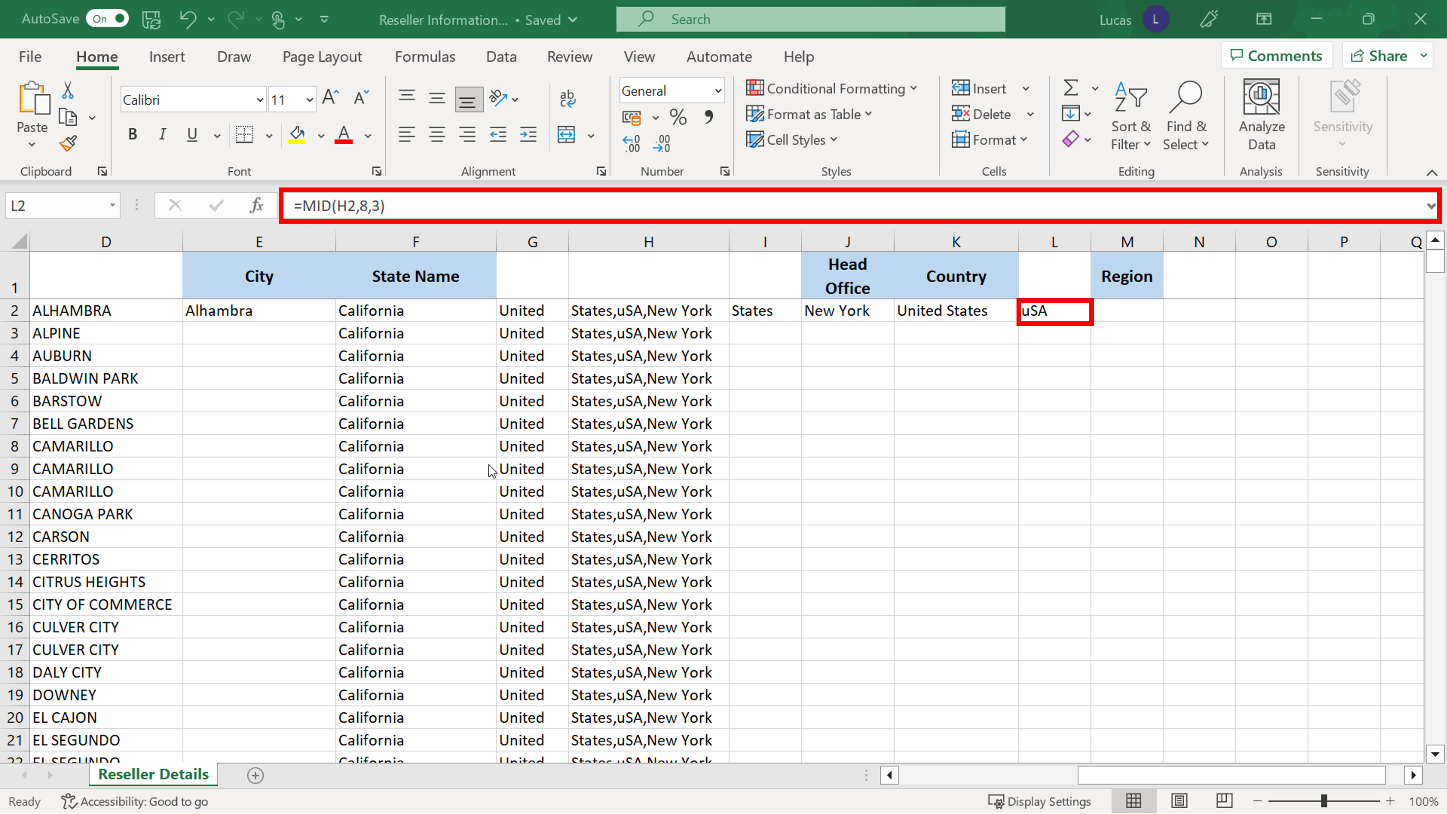
The formula arguments specify the source cell. They also specify the number of characters from the left at which to begin the extraction process. And how many characters to the right of the position to take.

The formula should read:

=**MID(H2,8,3)**

The result should read:

uSA



1. When the letters **uSA** had been extracted to **L2**, they were displayed in mixed casing exactly as they appeared in the text string. They should all be upper case.

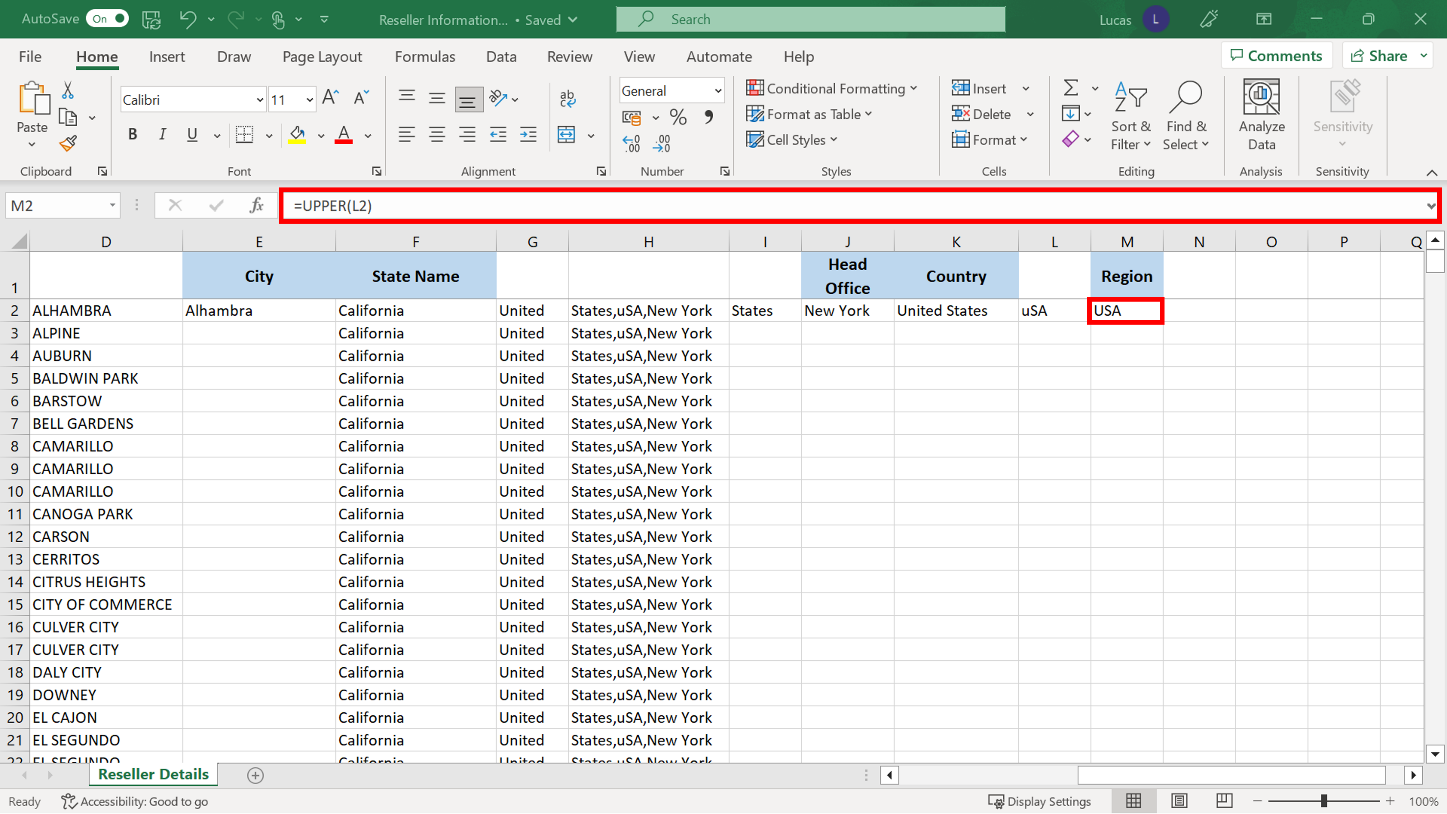
You created a formula in **M2** using the **UPPER** function to transform them into upper case.

The formula should read:

**=UPPER(L2)**

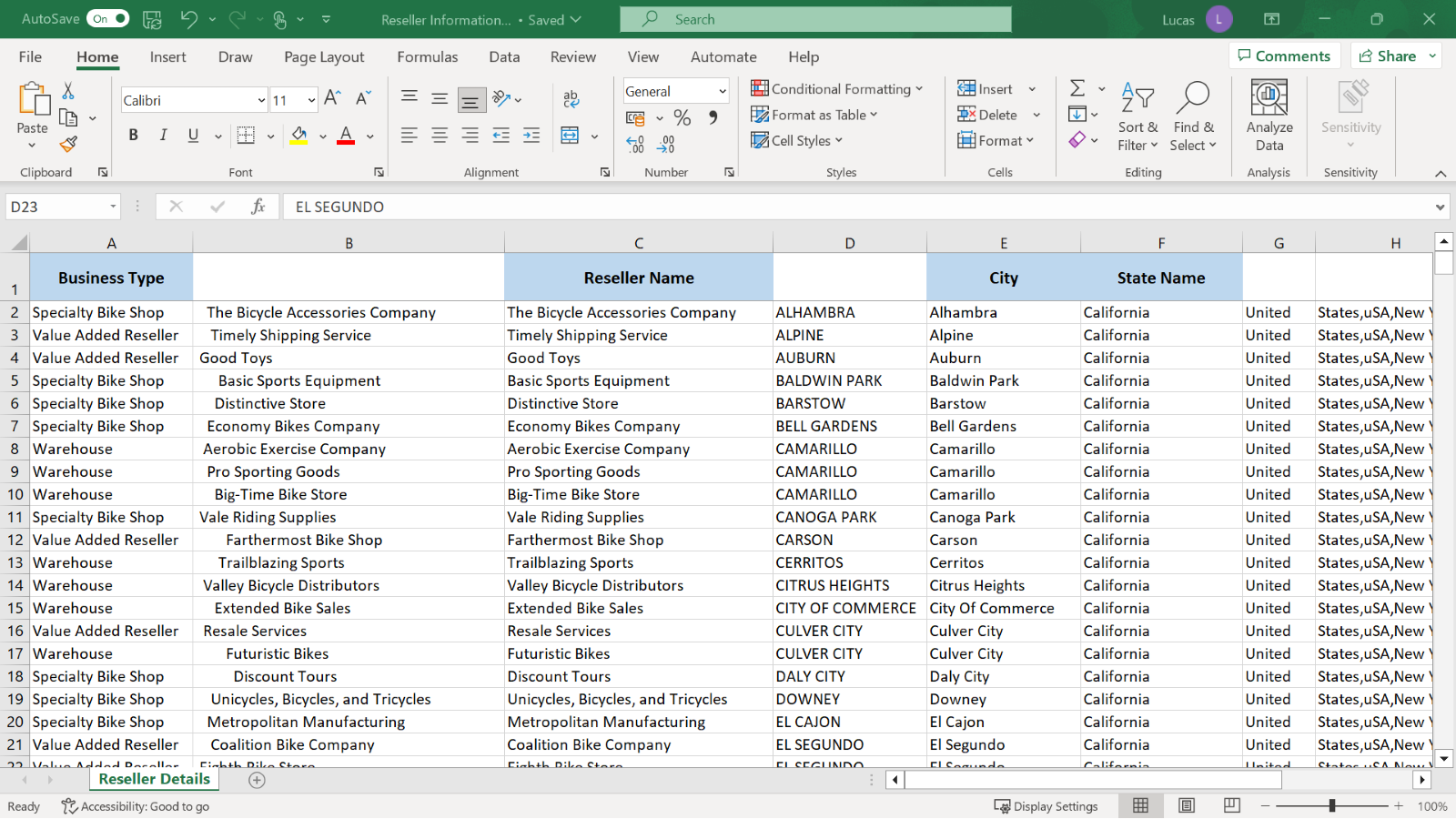
The result should read:

USA



1. Now that all the formulas were created, the **Autofill** shortcut could be used to copy them down their respective columns.

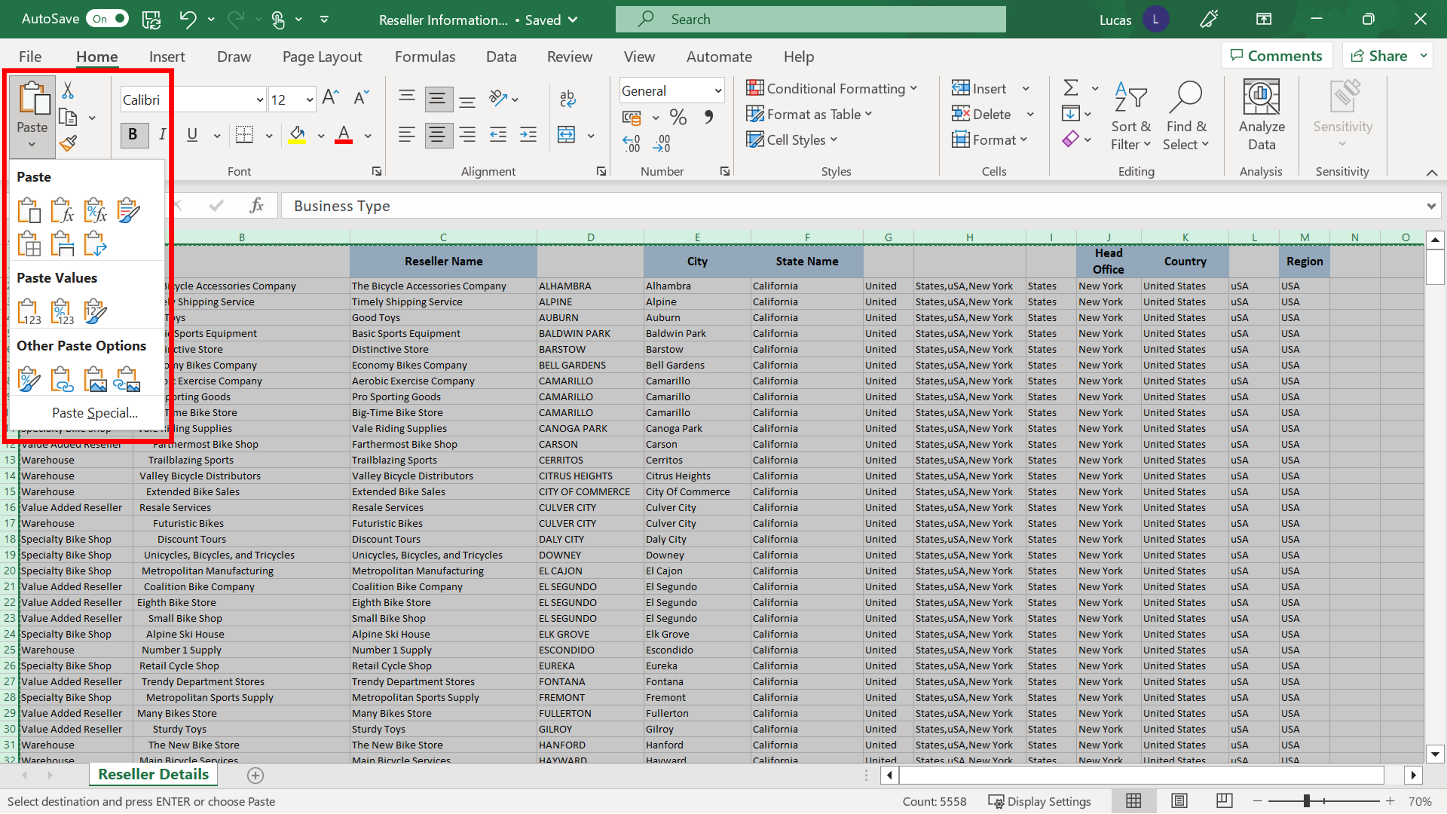
If you began with column **C** then there was a block of data to the left that Excel used as a reference. You could use the double-click shortcut method to copy the formulas down to row **428**. If you worked left to right, you were able to use the double-click shortcut on cells **E2**, **I2**, **J2**, **K2**, **L2**, and **M2**.



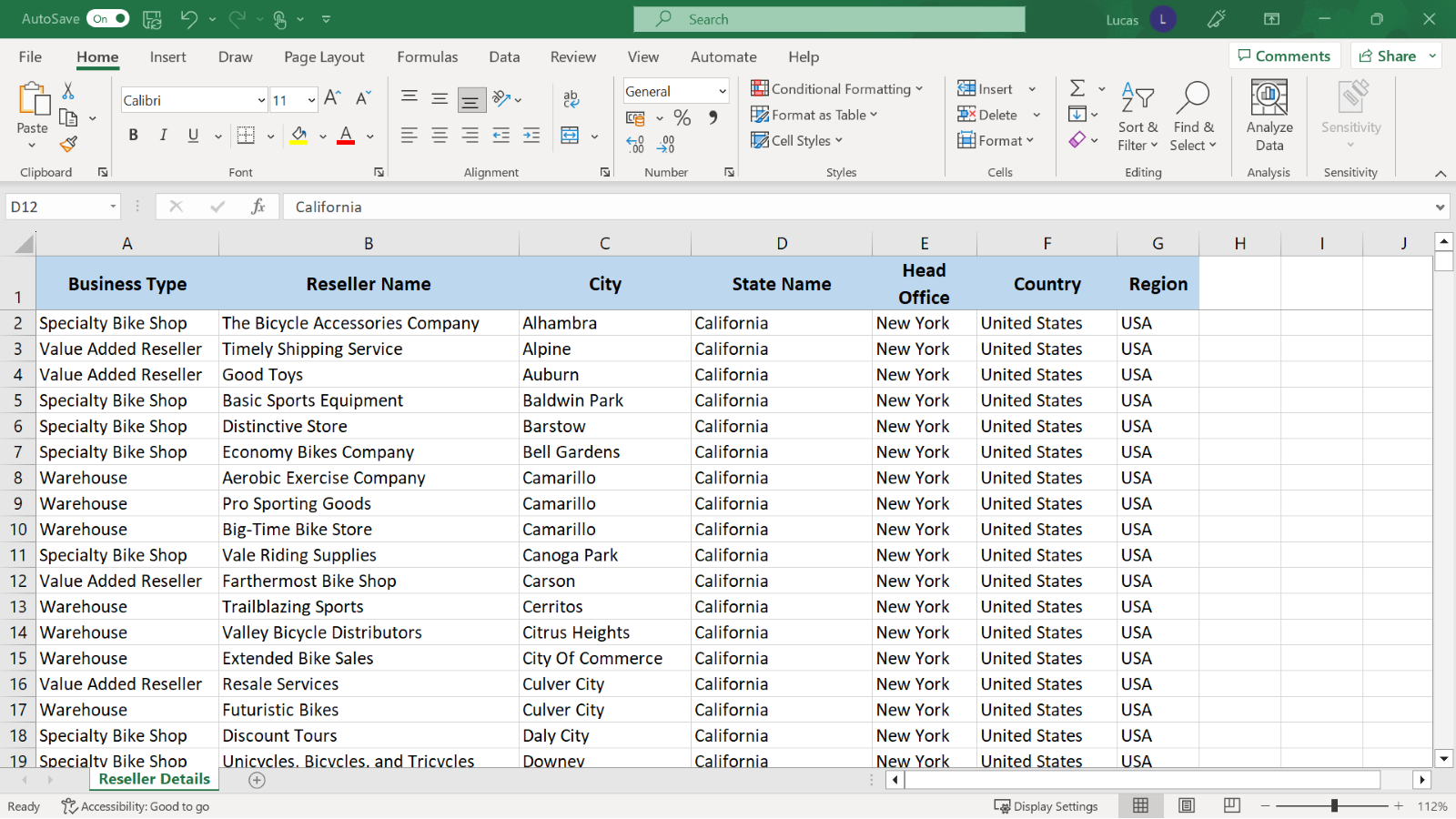
1. Now that these steps were complete, the spreadsheet had columns that contained duplicate information and columns that needed to be deleted.

First you had to copy the results of the formulas in columns **C**, **E**, **I**, **J**, **K**, **L** and **M** and paste these formulas back as values so that only the results generated remained in the spreadsheet.

To do this, you highlighted all the data and selected **Copy**. You then selected **Paste values** on the **Paste** drop-down on the **Home ribbon**.



1. Now that all the columns contain values, you could delete the original incorrect columns, columns **B**, **D**, **G**, **H**, **I** and **L** to achieve the final polished spreadsheet.



**Conclusion**

Congratulations! You have successfully finished this exercise and used a range of text function formulas to clean and standardize the data in the worksheet.