



**IBM Developer  
SKILLS NETWORK**

## Peer-graded Assignment Part 1 – Clean and Prepare the Data:

**Estimated time needed:** 45 minutes

In this hands-on lab you will open a CSV file in Excel for the web, convert it to an Excel format, and then clean and prepare the data.

### Software Used in this Assignment

The instruction videos in this course use the full Excel Desktop version as this has all the available product features, but for the hands-on labs we will be using the free ‘Excel for the web’ version as this is available to everyone.

Although you can use the Excel Desktop software if you have access to this version, it is recommended that you use Excel for the web for the hands-on labs as the lab instructions specifically refer to this version, and there are some small differences in the interface and available features.

### Dataset Used in this Assignment

The dataset used in this lab comes from the following source: <https://www.kaggle.com/kyanyoga/sample-sales-data> under a **Public Domain license**.

We are using a modified subset of the dataset for the lab. In order to follow the lab instructions successfully, please use the dataset provided with the lab, rather than the dataset from the original source.

### Guidelines for Submission

Download the file [sales\\_data\\_sample.CSV](#).

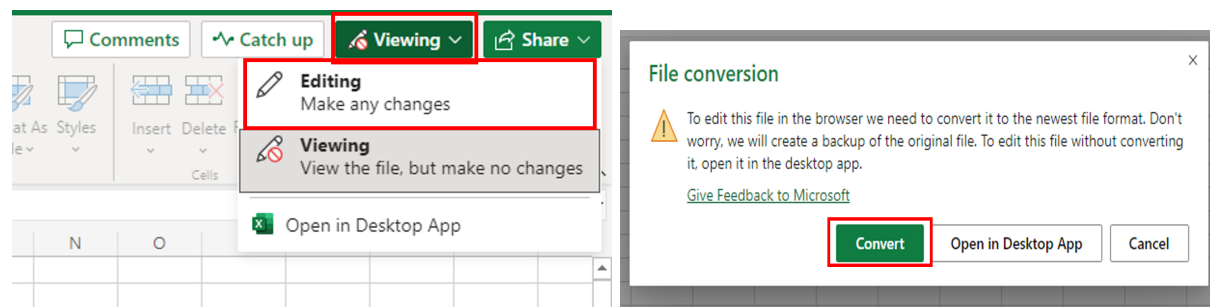
Upload and open the file with Excel for the web and convert it to an .XLSX file. Then, clean the data as detailed below.

Use the course videos from Module 3 and the Lab: Hands-on Lab 5: Cleaning Data to help you complete these tasks.

### Tasks to perform:

1. **Save the CSV file as an XLSX file:** Change the ‘Viewing’ in the ToolTip to ‘Editing’ in order to save

the file as an XLSX file. The file is converted when you click 'Convert' in the prompt.



2. **Column widths:** Adjust the widths of all columns so that the data is clearly visible in all cells.
3. **Empty rows:** Use the Filter feature to look for blanks and remove all empty rows from the data.
4. **Duplicate records:** Use either the Conditional Formatting or Remove Duplicates feature to look for and remove any duplicated records from the data based on the contents of Column A.
5. **Sorting:** Sort the data in Column A from smallest to largest.
6. **Format Cells:** In column C, set the type as Number and format it to remove the decimals. In column E, set the type as Number and format it to show up to two decimals.
7. **Whitespace:** Use the Find and Replace feature to remove all double-spaces from the data.
8. **Contact Name:** The contact name is divided into two columns as CONTACTLASTNAME (Column W) and CONTACTFIRSTNAME (Column X). Use Flash Fill to reduce the names to just one column, and then remove any unnecessary columns. Show the column values as FirstName LastName.
9. **Download your workbook:** Use 'Save As' and select 'Download a copy' to download your completed workbook as **sales\_data\_sample\_PART1.XLSX**.

## Author(s)

Rav Ahuja

## Other Contributor(s)

Malika Singla, Himanshu Birla, Simran Garg

## Changelog

Date	Version	Changed by	Change Description
2023-05-11	1.2	Anita Verma	Updated step 1 and step 8
2021-03-16	1.0	Simran	Created the version on GitLab

© IBM Corporation 2020. All rights reserved.