

## **MICROSOFT EXCEL PROJECT REPORT**

In this project, is a messy data set, and my job is to clean and organize the data using Microsoft Excel so it's ready for analysis.

The dataset contains inconsistent entries, extra spaces, missing or incorrect values, duplicate values and more.

I am to apply Excel cleaning techniques I've learnt to perform the tasks below, from text functions to filtering, sorting, and logical formulas.

Key Tasks to perform include:

1. Autofit Columns and Rows.
2. Identify and Remove Duplicates.
3. Trim Extra Spaces.
4. Eliminate Blank Cells.
5. Convert Data into Table.
6. Use Find and Replace to correct errors.
7. Validate data to be sure it is thoroughly clean.

This is an image of the dirty dataset.

**Excel\_DirtyData [Protected View] - Excel**

File Home Insert Draw Page Layout Formulas Data Review View Developer

**PROTECTED VIEW** Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View.

K10

	A	B	C	D	E	F	G	H
1	Date	ID	Name	Region	Rating	Product	Quantity	Price Per Unit
2	#####	1	John Smith	North	Good	Magic Wand	10	\$20.00
3	#####	2	Jane Doe	East	Excellent	Unicorn Horn	15	\$10.00
4	#####	3	Mike Tyson	West	Poor	Boxing Gloves	0 inf	
5	#####	4	Anna Belle	South	Average	Fairy Dust	25	\$10.00
6	#####	5	Chris P. Bacon	East	Good	Bacon Scented Candle	30	\$16.67
7	#####	6	Peter Parker		Excellent	Web Shooter	0 inf	
8	#####	7	Mary Jane	West	Poor	Potent Potion	35	\$10.00
9	#####	8	Bruce Wayne	South	Average	Bat Signal	40	\$15.00
10	#####	9	Clark Kent	East	Good	Glasses with X-ray Vision	45	\$12.22
11	#####	10	Diana Prince	North	Excellent	Lasso of Truth	50	\$14.00
12	#####	11	Tony Stark	West	Poor	Iron Man Suit	5	\$160.00
13	#####	12	Steve Rogers	South	Average	Captain America Shield	20	\$45.00
14	#####	13	Natasha Romanoff	East	Good	Black Widow's Bite	0 inf	
15	#####	14	Bruce Banner		Excellent	Gamma Radiation Serum	30	\$36.67
16	#####	4	Anna Belle	South	Average	Fairy Dust	25	\$10.00
17	#####	5	Chris P. Bacon	East	Good	Bacon Scented Candle	30	\$16.67
18	#####	6	Peter Parker		Excellent	Web Shooter	0 inf	
19	#####	15	Nick Fury	West	Poor	Eye Patch	35	\$34.29
20	#####	16	Phil Coulson		Average	Agent ID Card	0 inf	
21	#####	17	Peggy Carter	East	Good	Vintage Pistol	40	\$35.00
22	#####	18	Howard Stark	North	Excellent	Arc Reactor	45	\$33.33
23	#####	19	Hank Pym	West	Poor	Ant-Man Suit	50	\$32.00

Data

## 1. Autofit Columns and Rows.

To Autofit Columns and Rows, I clicked on the home tab → format → Under the cell size, I clicked on the Autofit Row height and Autofit Column width, as shown below.

This is to ensure visibility of the data in the cells

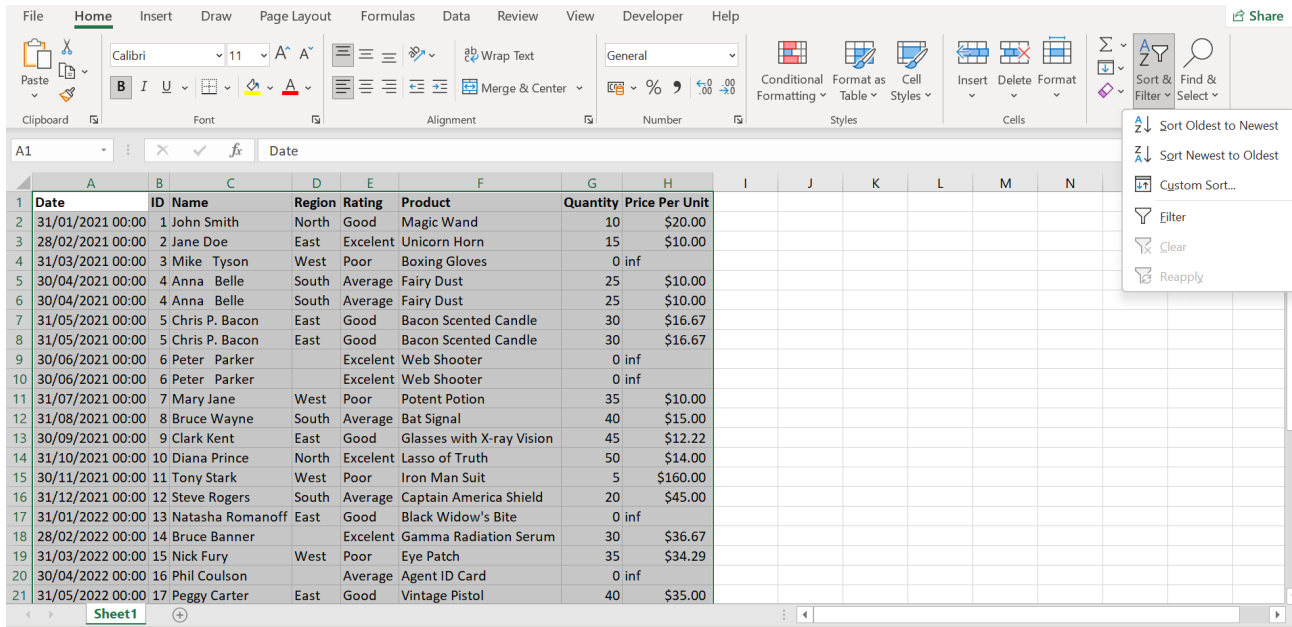
The screenshot shows the Microsoft Excel interface with the 'Home' tab selected. The 'Format' dropdown menu is open, and the 'Cell Size' submenu is visible. The 'Autofit Row Height' and 'Autofit Column Width' options are highlighted. The spreadsheet data is visible in the background, showing columns A through M and rows 1 through 21. The data includes dates, names, regions, ratings, products, quantities, and prices.

Date	ID	Name	Region	Rating	Product	Quantity	Price Per Unit
31/01/2021 00:00	1	John Smith	North	Good	Magic Wand	10	\$20.00
28/02/2021 00:00	2	Jane Doe	East	Excelent	Unicorn Horn	15	\$10.00
31/03/2021 00:00	3	Mike Tyson	West	Poor	Boxing Gloves	0 inf	
30/04/2021 00:00	4	Anna Belle	South	Average	Fairy Dust	25	\$10.00
31/05/2021 00:00	5	Chris P. Bacon	East	Good	Bacon Scented Candle	30	\$16.67
30/06/2021 00:00	6	Peter Parker		Excelent	Web Shooter	0 inf	
31/07/2021 00:00	7	Mary Jane	West	Poor	Potent Potion	35	\$10.00
31/08/2021 00:00	8	Bruce Wayne	South	Average	Bat Signal	40	\$15.00
30/09/2021 00:00	9	Clark Kent	East	Good	Glasses with X-ray Vision	45	\$12.22
31/10/2021 00:00	10	Diana Prince	North	Excelent	Lasso of Truth	50	\$14.00
30/11/2021 00:00	11	Tony Stark	West	Poor	Iron Man Suit	5	\$160.00
31/12/2021 00:00	12	Steve Rogers	South	Average	Captain America Shield	20	\$45.00
31/01/2022 00:00	13	Natasha Romanoff	East	Good	Black Widow's Bite	0 inf	
28/02/2022 00:00	14	Bruce Banner		Excelent	Gamma Radiation Serum	30	\$36.67
30/04/2021 00:00	4	Anna Belle	South	Average	Fairy Dust	25	\$10.00
31/05/2021 00:00	5	Chris P. Bacon	East	Good	Bacon Scented Candle	30	\$16.67
30/06/2021 00:00	6	Peter Parker		Excelent	Web Shooter	0 inf	
31/03/2022 00:00	15	Nick Fury	West	Poor	Eye Patch	35	\$34.29
30/04/2022 00:00	16	Phil Coulson		Average	Agent ID Card	0 inf	
31/05/2022 00:00	17	Peggy Carter	East	Good	Vintage Pistol	40	\$35.00

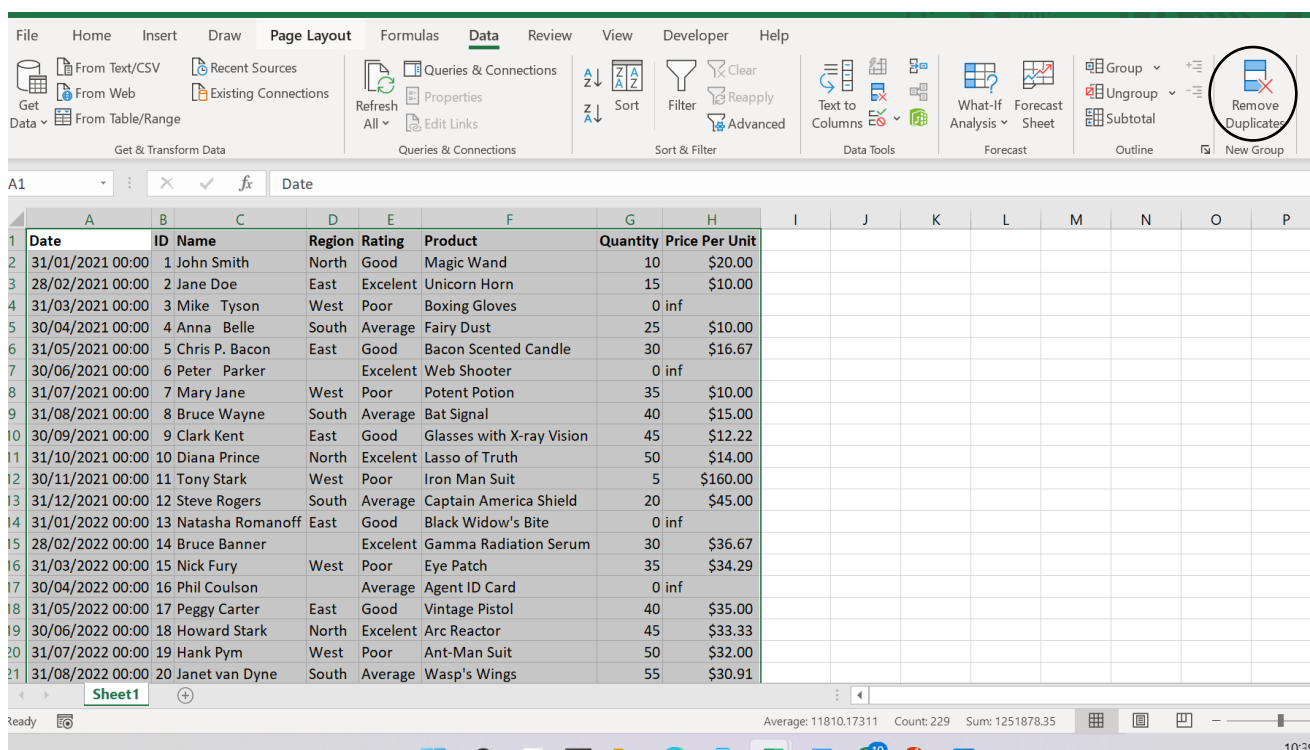
## 2. Identify and Remove Duplicates

To be able to identify and remove duplicates properly, I sorted in ascending order then removed duplicates.

To sort, I clicked on the home tab on the menu bar, then clicked on sort in ascending order A → Z, as shown below.



After sorting them I removed duplicates, I clicked on the data tab and looked for the remove duplicate icon to remove duplicates, as shown below.



### 3. Trim Extra Spaces.

To trim and clean extra spaces, I clicked on an empty cell, typed **=TRIM(CLEAN(C2:C29))**, it trims and clean the spaces in the **name** column, then I copied the trimmed names and **pasted as values** on the name column.

We trim to eliminate extra unwanted spaces.

The screenshot shows the Microsoft Excel interface with the Data tab selected. The formula bar displays `=TRIM(CLEAN(C2:C29))`. The spreadsheet contains the following data:

Date	ID	Name	Region	Rating	Product	Quantity	Price Per Unit	Trimmed
31/01/2021 00:00	1	John Smith	North	Good	Magic Wand	10	\$20.00	John Smith
28/02/2021 00:00	2	Jane Doe	East	Excelent	Unicorn Horn	15	\$10.00	Jane Doe
31/03/2021 00:00	3	Mike Tyson	West	Poor	Boxing Gloves	0 inf		Mike Tyson
30/04/2021 00:00	4	Anna Belle	South	Average	Fairy Dust	25	\$10.00	Anna Belle
31/05/2021 00:00	5	Chris P. Bacon	East	Good	Bacon Scented Candle	30	\$16.67	Chris P. Bacon
30/06/2021 00:00	6	Peter Parker		Excelent	Web Shooter	0 inf		Peter Parker
31/07/2021 00:00	7	Mary Jane	West	Poor	Potent Potion	35	\$10.00	Mary Jane
31/08/2021 00:00	8	Bruce Wayne	South	Average	Bat Signal	40	\$15.00	Bruce Wayne
30/09/2021 00:00	9	Clark Kent	East	Good	Glasses with X-ray Vision	45	\$12.22	Clark Kent
31/10/2021 00:00	10	Diana Prince	North	Excelent	Lasso of Truth	50	\$14.00	Diana Prince
30/11/2021 00:00	11	Tony Stark	West	Poor	Iron Man Suit	5	\$160.00	Tony Stark
31/12/2021 00:00	12	Steve Rogers	South	Average	Captain America Shield	20	\$45.00	Steve Rogers
31/01/2022 00:00	13	Natasha Romanoff	East	Good	Black Widow's Bite	0 inf		Natasha Romanoff
28/02/2022 00:00	14	Bruce Banner		Excelent	Gamma Radiation Serum	30	\$36.67	Bruce Banner
31/03/2022 00:00	15	Nick Fury	West	Poor	Eye Patch	35	\$34.29	Nick Fury
30/04/2022 00:00	16	Phil Coulson		Average	Agent ID Card	0 inf		Phil Coulson
31/05/2022 00:00	17	Peggy Carter	East	Good	Vintage Pistol	40	\$35.00	Peggy Carter
30/06/2022 00:00	18	Howard Stark	North	Excelent	Arc Reactor	45	\$33.33	Howard Stark
31/07/2022 00:00	19	Hank Pym	West	Poor	Ant-Man Suit	50	\$32.00	Hank Pym
31/08/2022 00:00	20	Janet van Dyne	South	Average	Wasp's Wings	55	\$30.91	Janet van Dyne

The screenshot shows a zoomed-in view of the spreadsheet data from the previous image. The formula bar shows the text "John Smith". The data is as follows:

	A	B	C	D	E	F	G	H
1	Date	ID	Name	Region	Rating	Product	Quantity	Price Per Unit
2	31/01/2021 00:00	1	John Smith	North	Good	Magic Wand	10	\$20.00
3	28/02/2021 00:00	2	Jane Doe	East	Excelent	Unicorn Horn	15	\$10.00
4	31/03/2021 00:00	3	Mike Tyson	West	Poor	Boxing Gloves	0 inf	
5	30/04/2021 00:00	4	Anna Belle	South	Average	Fairy Dust	25	\$10.00
6	31/05/2021 00:00	5	Chris P. Bacon	East	Good	Bacon Scented Candle	30	\$16.67
7	30/06/2021 00:00	6	Peter Parker		Excelent	Web Shooter	0 inf	
8	31/07/2021 00:00	7	Mary Jane	West	Poor	Potent Potion	35	\$10.00
9	31/08/2021 00:00	8	Bruce Wayne	South	Average	Bat Signal	40	\$15.00
10	30/09/2021 00:00	9	Clark Kent	East	Good	Glasses with X-ray Vision	45	\$12.22
11	31/10/2021 00:00	10	Diana Prince	North	Excelent	Lasso of Truth	50	\$14.00
12	30/11/2021 00:00	11	Tony Stark	West	Poor	Iron Man Suit	5	\$160.00
13	31/12/2021 00:00	12	Steve Rogers	South	Average	Captain America Shield	20	\$45.00
14	31/01/2022 00:00	13	Natasha Romanoff	East	Good	Black Widow's Bite	0 inf	
15	28/02/2022 00:00	14	Bruce Banner		Excelent	Gamma Radiation Serum	30	\$36.67
16	31/03/2022 00:00	15	Nick Fury	West	Poor	Eye Patch	35	\$34.29
17	30/04/2022 00:00	16	Phil Coulson		Average	Agent ID Card	0 inf	
18	31/05/2022 00:00	17	Peggy Carter	East	Good	Vintage Pistol	40	\$35.00
19	30/06/2022 00:00	18	Howard Stark	North	Excelent	Arc Reactor	45	\$33.33

#### 4. Eliminate Blank Cells.

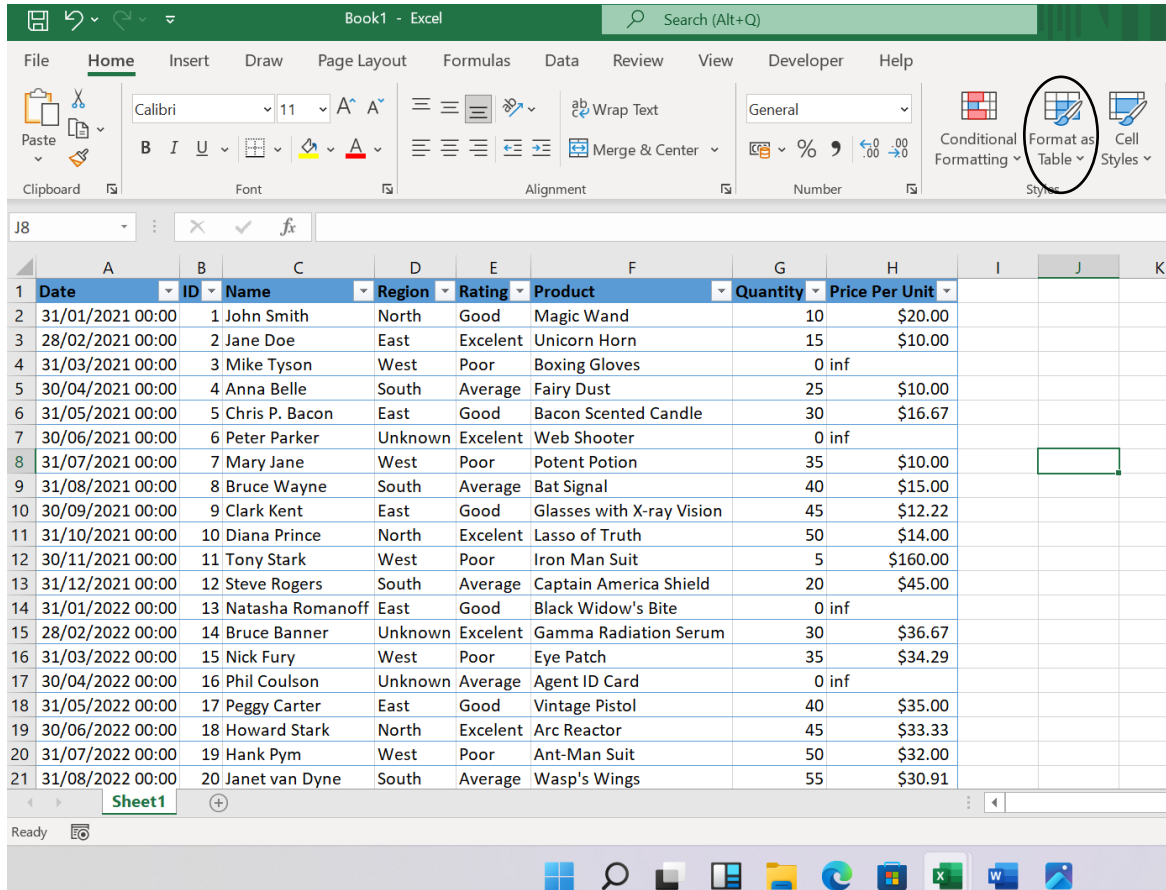
To remove blank cells on the **Region column**, I highlighted the region column, I typed (**Ctrl + F** – to find and replace) on my keyboard and it showed the icon for find and replace, to find **blank** and replace the **blank** with **unknown**, it replaced the three blank spaces with “**Unknown**”.

Book1 - Excel

</

## 5. Convert Data into Table.

I highlighted all the data; on the home tab, I clicked on **format as table**, which displayed different table designs and I chose one of the designs to create the table.



The screenshot shows the Microsoft Excel interface with the 'Home' tab selected. The 'Format as Table' button in the 'Styles' group is circled. The spreadsheet displays a table with the following data:

Date	ID	Name	Region	Rating	Product	Quantity	Price Per Unit
31/01/2021 00:00	1	John Smith	North	Good	Magic Wand	10	\$20.00
28/02/2021 00:00	2	Jane Doe	East	Excelent	Unicorn Horn	15	\$10.00
31/03/2021 00:00	3	Mike Tyson	West	Poor	Boxing Gloves	0	inf
30/04/2021 00:00	4	Anna Belle	South	Average	Fairy Dust	25	\$10.00
31/05/2021 00:00	5	Chris P. Bacon	East	Good	Bacon Scented Candle	30	\$16.67
30/06/2021 00:00	6	Peter Parker	Unknown	Excelent	Web Shooter	0	inf
31/07/2021 00:00	7	Mary Jane	West	Poor	Potent Potion	35	\$10.00
31/08/2021 00:00	8	Bruce Wayne	South	Average	Bat Signal	40	\$15.00
30/09/2021 00:00	9	Clark Kent	East	Good	Glasses with X-ray Vision	45	\$12.22
31/10/2021 00:00	10	Diana Prince	North	Excelent	Lasso of Truth	50	\$14.00
30/11/2021 00:00	11	Tony Stark	West	Poor	Iron Man Suit	5	\$160.00
31/12/2021 00:00	12	Steve Rogers	South	Average	Captain America Shield	20	\$45.00
31/01/2022 00:00	13	Natasha Romanoff	East	Good	Black Widow's Bite	0	inf
28/02/2022 00:00	14	Bruce Banner	Unknown	Excelent	Gamma Radiation Serum	30	\$36.67
31/03/2022 00:00	15	Nick Fury	West	Poor	Eye Patch	35	\$34.29
30/04/2022 00:00	16	Phil Coulson	Unknown	Average	Agent ID Card	0	inf
31/05/2022 00:00	17	Peggy Carter	East	Good	Vintage Pistol	40	\$35.00
30/06/2022 00:00	18	Howard Stark	North	Excelent	Arc Reactor	45	\$33.33
31/07/2022 00:00	19	Hank Pym	West	Poor	Ant-Man Suit	50	\$32.00
31/08/2022 00:00	20	Janet van Dyne	South	Average	Wasp's Wings	55	\$30.91

## 6. Use Find and Replace to correct errors.

On the rating column, I observed that there was a spelling error with the word “**Excellent**” which was spelt “**Excelent**”, in addition there was an error in imputing the price per unit some cells had ‘**inf**’ which was a text, it was replaced with a null value ‘**\$0.00**’. I used the **find and select tab** on the home tab, it displayed a space where I could type the what I was looking for and what to replace it with. At the Rating column, I observed that there were some entries that were made and doesn’t corelate with a rating (**Excellent, good, average, poor**) I replaced them with ‘**not given**’

After which I trimmed the price per unit column to eliminate unwanted spaces, by using the =trim(clean(H2:H29)).

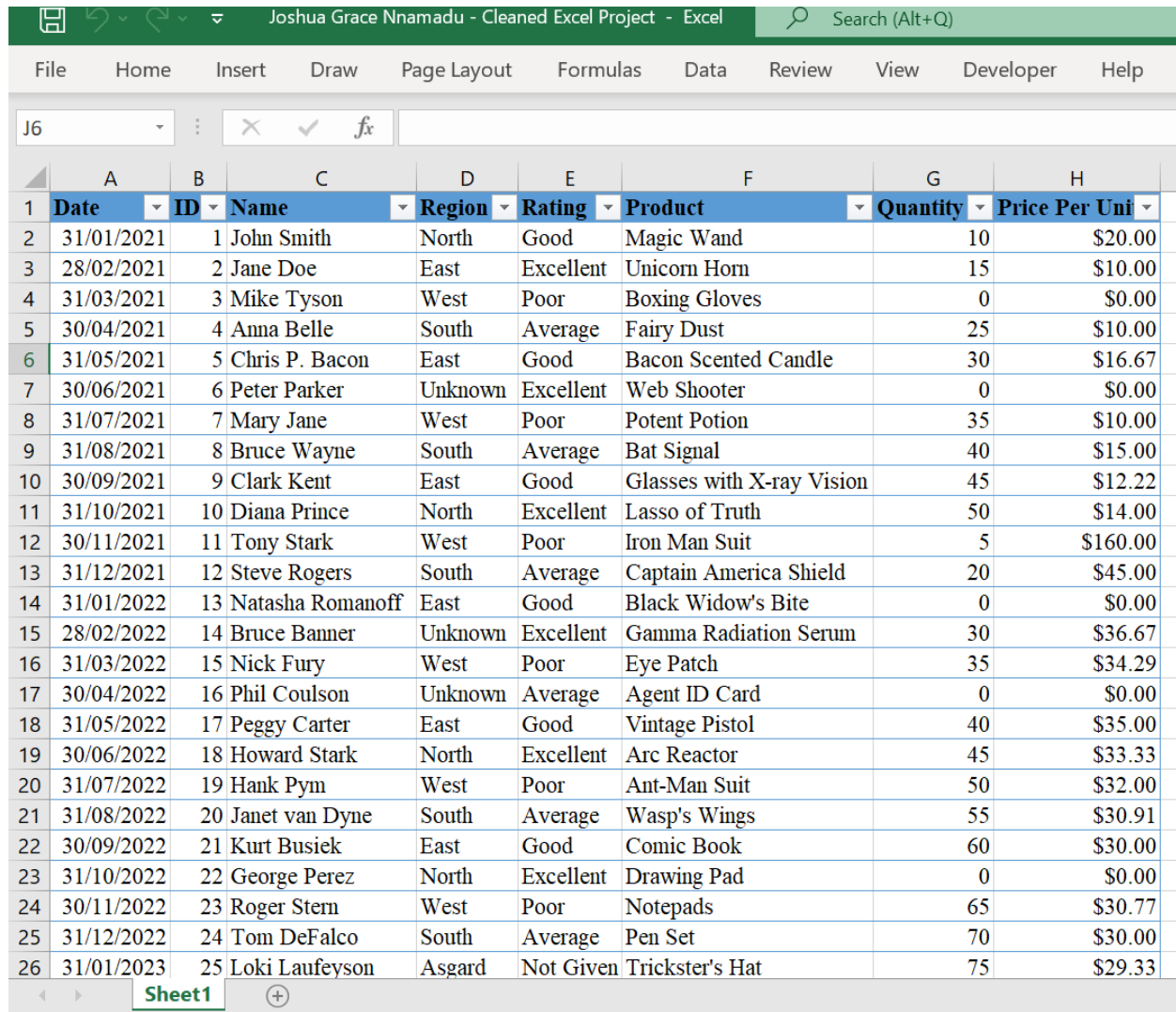
I also converted the date column to a short date format.

[illegible]



## 7. Validate data to be sure it is thoroughly clean.

I ensured all data were appropriately cleaned and contained meaningful information ready for analysis and reporting.



	A	B	C	D	E	F	G	H
1	Date	ID	Name	Region	Rating	Product	Quantity	Price Per Unit
2	31/01/2021	1	John Smith	North	Good	Magic Wand	10	\$20.00
3	28/02/2021	2	Jane Doe	East	Excellent	Unicorn Horn	15	\$10.00
4	31/03/2021	3	Mike Tyson	West	Poor	Boxing Gloves	0	\$0.00
5	30/04/2021	4	Anna Belle	South	Average	Fairy Dust	25	\$10.00
6	31/05/2021	5	Chris P. Bacon	East	Good	Bacon Scented Candle	30	\$16.67
7	30/06/2021	6	Peter Parker	Unknown	Excellent	Web Shooter	0	\$0.00
8	31/07/2021	7	Mary Jane	West	Poor	Potent Potion	35	\$10.00
9	31/08/2021	8	Bruce Wayne	South	Average	Bat Signal	40	\$15.00
10	30/09/2021	9	Clark Kent	East	Good	Glasses with X-ray Vision	45	\$12.22
11	31/10/2021	10	Diana Prince	North	Excellent	Lasso of Truth	50	\$14.00
12	30/11/2021	11	Tony Stark	West	Poor	Iron Man Suit	5	\$160.00
13	31/12/2021	12	Steve Rogers	South	Average	Captain America Shield	20	\$45.00
14	31/01/2022	13	Natasha Romanoff	East	Good	Black Widow's Bite	0	\$0.00
15	28/02/2022	14	Bruce Banner	Unknown	Excellent	Gamma Radiation Serum	30	\$36.67
16	31/03/2022	15	Nick Fury	West	Poor	Eye Patch	35	\$34.29
17	30/04/2022	16	Phil Coulson	Unknown	Average	Agent ID Card	0	\$0.00
18	31/05/2022	17	Peggy Carter	East	Good	Vintage Pistol	40	\$35.00
19	30/06/2022	18	Howard Stark	North	Excellent	Arc Reactor	45	\$33.33
20	31/07/2022	19	Hank Pym	West	Poor	Ant-Man Suit	50	\$32.00
21	31/08/2022	20	Janet van Dyne	South	Average	Wasp's Wings	55	\$30.91
22	30/09/2022	21	Kurt Busiek	East	Good	Comic Book	60	\$30.00
23	31/10/2022	22	George Perez	North	Excellent	Drawing Pad	0	\$0.00
24	30/11/2022	23	Roger Stern	West	Poor	Notepads	65	\$30.77
25	31/12/2022	24	Tom DeFalco	South	Average	Pen Set	70	\$30.00
26	31/01/2023	25	Loki Laufeyson	Asgard	Not Given	Trickster's Hat	75	\$29.33

## Summary

I was able to autofit columns and rows to ensure visibility, identify and remove duplicate values, trim extra spaces, eliminate blank cells, correct errors and to convert data into table. This was done by using the appropriate excel functions ranging from sorting and filtering, remove duplicates, text and logical functions.