

1. Download and open the file “[vgsales_dirty.xlsx](#)” in Excel.

- Inspect the data. Do you see anything in need of cleaning?
- In your “Answers” sheet, write up your initial thoughts about problems that might emerge if this data isn’t cleaned up.

2. Copy the data to a new sheet, “Sheet2.” Focus on the NA values in the data set—there are a lot!

- Which columns contain NAs?
- What do you think they represent?

3. Copy the data to a new sheet, “Sheet3.” Remove any duplicate records from the data set.

4. Copy the data to a new sheet, “Sheet4.” Find and remove columns that don’t contain any useful information (e.g., those that have all 0 values).

5. Copy the data to a new sheet, “Sheet5.” Isolate any columns that are partially empty. Use mean (average) imputation to handle these partially empty columns.

6. Copy the data to a new sheet, “Sheet6.” Focus on the rows. Are any rows completely empty? If so, remove them.

7. Copy the data to a new sheet, “Sheet7.” Perform any additional cleaning that you feel would benefit the data (such as column formatting). Write down any actions you take on your “Answers” sheet