# Exploration and Cleaning of Video Game Sales Project

***Excel Work***

* Today begins my journey into the exploration of Video Game Sales. My first step was to try viewing the data in Excel. This step is mainly for browsing through this medium-sized dataset for possible situations where the data could fail on the import step in MySQL.
* Already upon initial viewing and filtering through each column, I discovered that there were multiple cases where the Year values for certain rows were written as “N/A”. This certainly would need to be cleaned or at least replaced with another value. I could have easily removed those entire rows, but that is usually bad practice and since the rest of the row values were included, I decided that I would replace the “N/A” in these rows with a value of 0.
* I also noticed that one of the Platforms was filtered as 2600. I know that this platform is actually the Atari 2600 (verified by researching the names of the games online), so I modified those rows as well.
* I searched for any special characters I could filter by (such as copyright and trademark symbols, paragraph symbols, accented letters, etc) and replaced or removed those characters.

***My SQL work***

* I decided to do the rest of exploration and cleaning after I’d imported the data into MySQL. During that round of data cleaning, I first wanted to remove any duplicate rows that Excel failed to catch (where the data doesn’t exactly match but is still inaccurate). I created a CTE that appended a Row Number column to the data which would partition over a few columns to identify dupes. Then I would select from the CTE to find any row numbers over 1 (any above 1 would be a duplicate).

A screenshot of a computer

Description automatically generated

* From the result, we can see that Madden NFL 13 and Wii de Asobu: Metroid Prime were duplicate rows. To double check, I used a select statement to find all the rows that had those Names.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* The Wii de Asobu result already shows a completely duplicate row, so we will automatically remove that one.
* Now I’ll standardize the data by fixing a few potential issues such as data whitespace trimming.