**Cleaning Summary**

Survey Data

1. The 2022 file has an extra column "OPEID" which had to be stripped to match the column set of the other files
2. The 2007-2009 SAS files have an error. We will have to read these in as .xlsx or .csv files.
3. There are multiple labels for track and field corresponding to the same codes. Therefore, we will use the code as the primary key in the Sports table to enforce referential integrity. We also notice a label for "all track combined". On investigation, this presented no duplication issues. Schools with this code/label had only this code/label for track.

Win-Loss Data

1. The school names in this dataset were inconsistent with the survey dataset. We created a lookup table to translate the names.
2. The data source lacked data from 2009-2010 across all sports we pulled data for. Women's volleyball and soccer both additionally lacked 2010-2011 data. We will decide as we do our analysis to leave the data as is, scope our project around the missing data, or look for supplemental data sources to fill in the gaps

4/06

1. Merged all s