Principalmente para los datasets que se utilizaron se encontraron varios problemas. Principalmente errores de escritura, ya sea los nombres no capitalizados, mal escritos etc. En paralelo los valores numéricos en algunos casos se encontraban como negativos, o con simbología y otros no. Por lo que se limpiaron estos datos capitalizando todo con la función 'capitalize'. Se reemplazaron los valores '$' por un espacio en blanco para estandarizar todas las medidas y en otros casos el uso de "desconocido" para etiquetas.
Una parte importante fue como eliminar los valores NaN o los valores que fueran 0. Para el caso de los NaN, algunos fue prudente utilizar su mediana ya que dado los valores en general, su uso fue el con mayor criterio así no perder datos posiblemente valiosos. Por otro lado, para los valores de cantidad, simplemente se decidió eliminar algunas filas con valor 0 ya que, esa información no opacaba el análisis en general, considerando que fueron exactamente 3 filas eliminadas que contenían información redundante (0 ventas por día y con un valor monetario, el cual no tiene como explicarse).

Posteriormente luego de limpiar completamente ambas tablas, se hizo un análisis con el fin de ver algunos criterios, por ejemplo: Para los estudiantes se pudo generar una instancia de revisión de asistencia en donde si el alumno cumplía con el criterio, este aparecía como aprobado, en revisión o reprobado. Para los datos de ventas, se generó una columna con un apartado de valor total, considerando el precio unitario de cada producto por la cantidad comprada.

Los usos posteriores, pueden ser diferentes para cada dataset. Por ejemplo, para el dataset de ventas se generó una tabla dinámica la cual indica qué productos se compraron por día y el valor total de venta. Para los estudiantes se menciona el criterio de asistencia, pero el mismo criterio se puede utilizar para una revisión de las notas finales.