Fase 3: Preparación de los datos

La fusión de datos no fue necesario en el punto de vista que se sólo se usó una aplicación para la recopilación de datos y que en el procesos se usaron los mismos estándares de evaluación para el registro de los valores de los alimentos, por otra parte se podría considerar que la aplicación Samsung Healt generó diferentes archivos csv para registrar los datos por lo que en el proceso se tuvo que relacionar los datos de los diferentes archivos para poder registrar los alimentos con sus respectivos valores de información nutrimental ya que todo la información estaba con claves y se tuvo que procesar todos los datos.

No hubo necesidad de realizar subconjuntos ya que toda la información que se presenta es de la misma agrupación de datos y toda está en la misma sintonía para realizar este trabajó por lo que lo justifico en que el proyecto necesita datos del registro nutrimental como el alimento calorías, carbohidratos, sodio, lípidos, proteínas y todos esos datos son parte de la información de los alimentos por lo que los datos con los que se trabajan son del mismo conjunto de datos ya que es alimento con información del alimento.

En cuanto a los datos se puede agregar más datos ya que requiere un registro de alimentos diarios por lo que se puede adicionar la cantidad de datos con su respectiva información cuanto sea necesario ya que una ventaja de esto es que va a adquirir una mayor precisión nuestro análisis al poder tener más datos con que trabajar porque tienen más información para analizar.

No hubo necesidad en realizar cambios en atributos de los datos ya que todos contienen la información necesaria para trabajar y todos los atributos de todos los datos fueron obtenidos de la misma aplicación con los mismos entandares por lo que se asegura una relación correcta entre los atributos con los datos.

No hubo necesidad ordenar los datos para el análisis en general a excepción en la obtención de la moda y mediana estadística donde sí se necesitaba tener el orden de mayor a menor de cada uno de los atributos de los datos.

Los valores en blanco que se tuvieron que eliminar fueron con respecto al agua ya que este no proporcionaba información que ayudara al análisis de los datos por lo que quedó fuera de la realización del trabajo.