Si separamos nuestros títulos de las columnas, algunos quizá por el poco contexto, puedan parecer solo datos, sin más, sin embargo, una vez que se le da contexto y se juntan ya tenemos información, información que utilizaremos para analizar nuestros datos y así podemos compararlos como lo hemos pensado. Tenemos los siguientes datos:

FECHA: Esta es nuestra primera columna, a pesar de que no dice como tal que es una fecha, lo deducimos por el cuerpo que tiene y por el que usualmente es representado, que es día, mes y año, representado por la siguiente notación: dd/mm/aaaa. Podemos tomar como valores de día, del 1 al 30 o 31, en el mes del 1 al 12 y en el año, solamente 2021.

DÍA: Esta columna se refiere únicamente al día del mes que toman valores del 1 al 30 o 31 dependiendo el mes. Hace referencia a la afluencia de pasajeros que abordan las estaciones por día.

MES: Esta columna se refiere al mes del año que toman valores del 1 al 12. Hace referencia a la afluencia de pasajeros que abordan las estaciones por bloque en un mes.

AÑO: Esta columna se refiere al año, toma el valor de 1. Hace referencia a la afluencia de pasajeros que abordan las estaciones por bloque en un año.

LÍNEA: Esta columna se refiere a la cantidad de líneas que contiene este medio de transporte. Estas las utilizaremos para graficar. Puede tomar los valores del 1 al 9, el 12, la A y la B.

ESTACIÓN: Cada una de estas tiene un nombre estipulado para poder diferenciarlas y así podernos mover a través de la ciudad. Sin embargo, puede llegar a ser tedioso buscar algunas de estas y no poder encontrarla debido a que quizá en la base de datos no está escrita de ese modo, por lo que decidimos agregar un diccionario de datos, en el que a cada nombre de estación le pusimos un número específico y único, por lo que será mucho menos difícil encontrar alguna. Estos datos pueden tomar el valor del 1 al 162.

AFLUENCIA: Estos son datos grandes, muy grandes por la cantidad de gente que utiliza este transporte. Decidimos limpiar nuestra base quitando los datos en 0 y menos a 1,000; debido a que se nos hizo irreal tener 0 o menos de 1,000 personas en una estación cunado millones de usuarios día por día abordan cada una de estas estaciones. Estos datos pueden tomar de 1001 a 10,000.