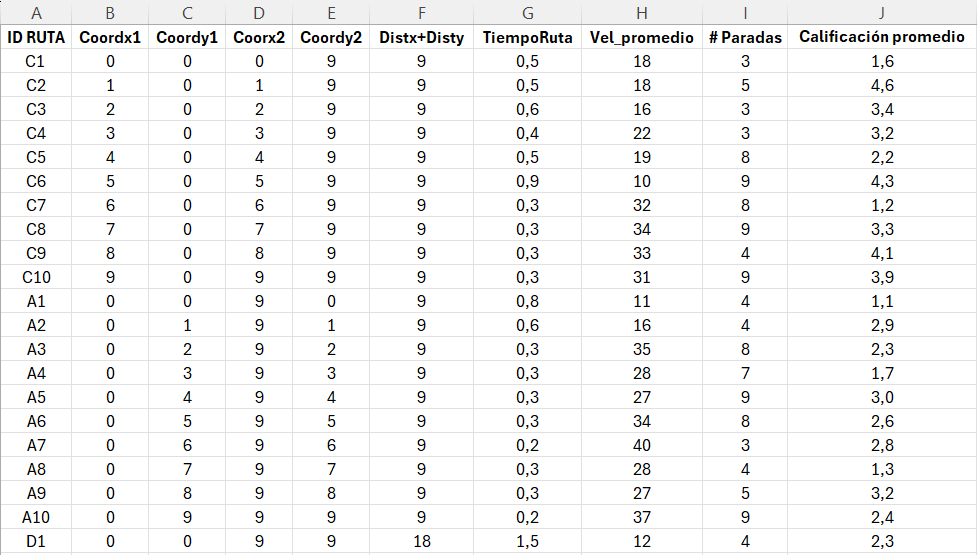
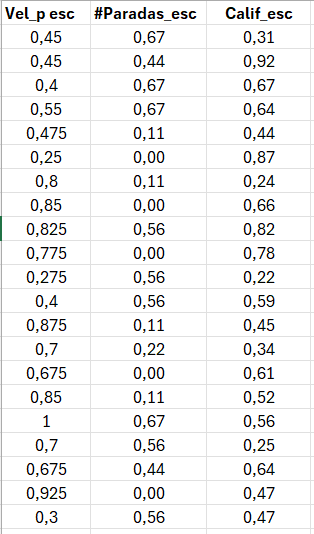
**Descripción de los datos**



EN el archivo dataset1 Se tuvo en cuenta los datos presentados en la anterior tabla

las coordenadas del nodo inicial y final permitieron determinar la distancia heurística como la distancia en el eje x sumada con la distancia en el eje y, el tiempo total de la ruta sirvió para determinar una velocidad promedio, también se tuvo en cuenta el número de paradas que realiza la ruta y una calificación promedio de los usuarios que se puede obtener de encuestas o de aplicaciones móviles del servicio público.

El dataset para este ejercicio fue generado. La consulta de datos en internet solo sirvió para determinar el tipo de dato que se podía extraer en relación con un sistema de transporte



Para generar los datos de entrada escalaron las características para que tomaran valores entre 0 y 1, en el caso de la tercera característica se usó la siguiente ecuación

Esta forma de escalado permitió que las rutas con muchas paradas recibieran una puntuación baja para esta característica y las que tuvieran pocas recibieran un puntaje cercano a 1

Se hará una clasificación no supervisada de los ejemplos del archivo dataset2.xlsx con la intención de que el código clasifique en dos grupos, rutas eficientes y rutas ineficientes