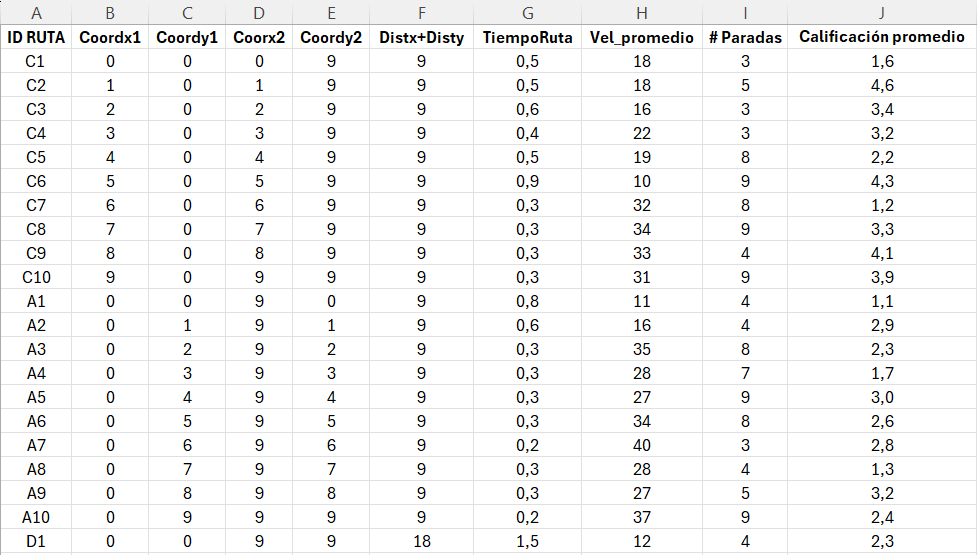
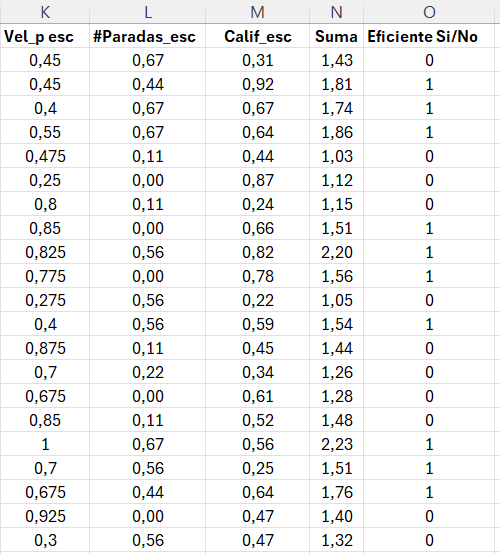
**Descripción de los datos**



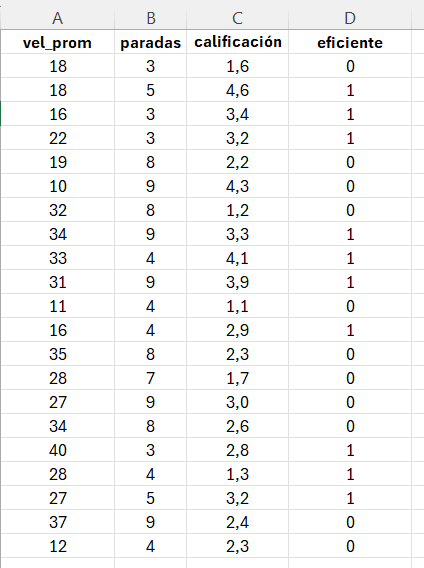
EN el archivo dataset1 Se tuvo en cuenta los datos presentados en la anterior tabla

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 el conjunto de entrenamiento del árbol de decisiones se decidió escalar las características para que tomaran valores entre 0 y 1, si la suma de las tres características escaladas daba mayor a 1,5 se consideraba que la ruta era eficiente y se le dio el valor Booleano=1. Si suma de las tres características daba menor a 1.5 se consideraba una ruta ineficiente y se le dio valor Booleano=0. Esto únicamente para los datos de entrenamiento del modelo puesto que algoritmo usado es de aprendizaje supervisado por medio de árboles de decisiones



Finalmente se organizó el conjunto de entrenamiento para ser leído por el algoritmo de Python y generar el árbol de decisiones. Las tres primeras columnas son características y la cuarta son los valores de salida del conjunto de entrenamiento