Bitácora del Capitán

Introducción:

Bienvenidos a la bitácora del capitán, lo que vendría a ser mi forma ultra super hyper MEGA original de documentar mi progreso en el aprendizaje del análisis de datos. Espero que lo disfruten tanto como yo :D.

28/12/2023, jueves 10:24am:

Luego de renegar como un pelotudo para averiguar como transformar correctamente un .csv a tabla sql de SSMS, después de tanto tiempo entendí que lo que estaba haciendo mal era Tasks>Import Data en vez de Tasks>Import Flat File. Voy a proseguir con la tarea de mapear todas las tablas (la base de datos es sobre la formula 1).

12:49:

Parece que todas las columnas ‘points’ no están del todo limpias, ya que hay muchos valores que deberías ser int que están como string. Por el momento los guardé como varchar(50), para futura manipulación lo tendré en cuenta y deberé hacer una limpieza posterior.

16:15

Tuve un problema con una de las claves primarias en tabla lap\_times. Tiene que ser compuesta para que sea única (formada por el id del corredor, el id de la carrera, y la vuelta). El problema fue que al principio solo había puesto los ids como pk y estaba mal.

16:20

Otra consideración para limpieza, en la tabla pit\_stops columna *duration* es redundante, debo eliminarla.

Lo hice antes, modifico en excel directamente, la borro, exporto como .csv y listo.

Tuve que borrar varias columnas de la tabla ***races****,* no aportaban información lo suficientemente relevante para el modelo y estaban bastant incompletas (por ejemplo: *fp1\_date*, 94% valores nulos).