ETL PROJECT_1
Pentaho Data Integration

csvi-Cars → ur_Duplicates → SV-dtype_correction → TO-sql_dba

Pentaho Transformations

Pentaho Data Integration(PDI) provides the Extract, Transform, and Load (ETL) capabilities that facilitate the process of capturing, cleansing, and storing data using a uniform and consistent format that is accessible and relevant to end users and IoT technologies.

This is an ETL process from Pentaho Data Integration tool. It consist of 4 steps:

1. Extraction. Getting the data from local to the tool to be able to explore, trim, check the data type and finally preview the imported data.
2. Duplicates. In this step, I'll drop all duplicated rows to clean it more.
3. Select Values. Next, it's the data preparation to correct all data type before loading to a sql database or data warehouse.
4. Table Output. Finally, it's the creation of the landing table in the database. After the data is ready for further analysis.

By running the transformation, a green tick will appear if everything is correct as pictured above.