

## Dokumentasi Project Data Engineer

Note: Jika Anda mengalami masalah dalam mengakses project yang saya zip, Anda bisa mendapatkan project ini di GitHub saya (<https://github.com/Ella-Raputri/MiniProject-Advance/tree/main/DE>). Terima kasih.

Penjelasan mengenai file dan folder yang ada di folder ini:

- Folder FilterDataset

Folder FilterDataset berisi Jupyter Notebook dan beberapa file csv hasil pemecahan dari netflix\_titles.xlsx. Saya menggunakan beberapa file tersebut untuk melakukan pengecekan pada data yang ada di file.xlsx. Saya menemukan bahwa terdapat beberapa row yang kurang konsisten, di mana 1 row terpecah menjadi 2 row di file Excel tersebut. Setelah melakukan sedikit cleaning dan cek tersebut, saya menyimpan hasil akhirnya di folder fixed/netflix\_titles.xlsx di dalam folder FilterDataset. File Excel di folder fixed ini lah yang akan saya gunakan untuk dimasukkan ke proyek SSIS.

- Folder Excel\_to\_DB

Folder ini berisi proyek SSIS yang berisi operasi insert data dari Excel ke database dan join data-data tersebut menjadi 1 tabel.

- create\_table.sql

File SQL ini berisi query yang saya gunakan untuk create table yang akan digunakan pada proyek SSIS.

- table\_fact\_dim.sql

File SQL ini berisi query yang saya gunakan untuk create table Fact dan beberapa table Dim (Dimension) dari proyek SSIS dan DWH.

- EllaRaputri\_Soal\_Project\_Data\_Warehouse.pdf

File pdf ini berisi jawaban saya untuk soal-soal di proyek DWH.