**Soal Prioritas 1**

1. Sebutkan dan jelaskan berbagai tools yang dapat digunakan untuk meningkatkan kualitas dari sebuah data!

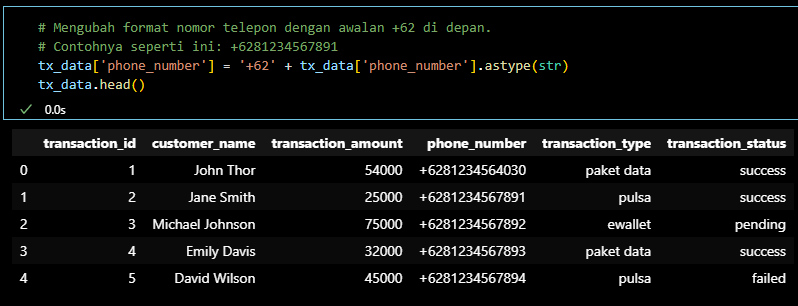
* Data Cleansing Tools: membantu membersihkan dan memperbaiki data yang kotor, tidak lengkap, atau tidak konsisten, seperti OpenRefine, DataCleaner, dan Trifacta Wrangler.
* Data Validation Tools: membantu memvalidasi data berdasarkan aturan bisnis atau persyaratan tertentu, seperti Talend Data Quality, Informatica Data Quality, dan IBM InfoSphere.
* Data Matching and Deduplication Tools: digunakan untuk mencocokkan dan menghilangkan duplikasi data, seperti FairlyMatch, Tamr, dan IBM InfoSphere MDM.
* Data Profiling Tools: membantu menganalisis struktur, isi, dan kualitas data, seperti Erwin Data Profiler, Talend Data Preparation, dan Microsoft SQL Server Integration Services.
* Metadata Management Tools: membantu mengelola metadata, seperti definisi data, asal usul data, dan aturan bisnis, seperti Alation Data Catalog, Collibra Data Governance, dan IBM Information Governance Catalog.

1. Terdapat berbagai cara dalam meningkatkan kualitas dari sebuah data (data quality) seperti data governance, data profiling dan data matching. Dari berbagai cara tersebut, sebutkan berbagai tantangan yang perlu diperhatikan!

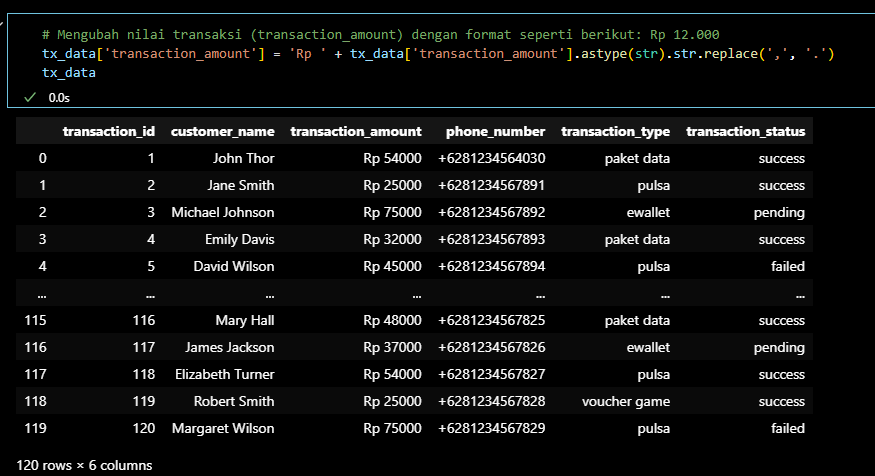
* Tata kelola data yang sulit diimplementasikan secara efektif dan mendapatkan komitmen dari seluruh organisasi.
* Kesulitan mengumpulkan dan menganalisis metadata yang komprehensif serta menginterpretasikan hasil profiling data.
* Mencocokkan dan menghapus duplikasi data yang tidak konsisten, serta menentukan ambang batas yang tepat.
* Keterbatasan sumber daya, kompleksitas data, dan resistensi budaya organisasi terhadap perubahan praktik pengelolaan data.

**Soal Prioritas 2**

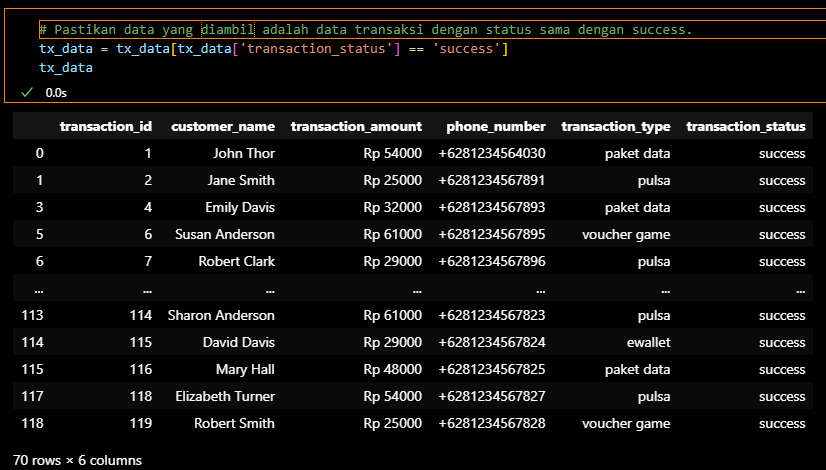
1. Lakukan proses data cleaning dengan kriteria sebagai berikut:
2. Mengubah format nomor telepon dengan awalan +62 di depan. Contohnya seperti ini: +6281234567891



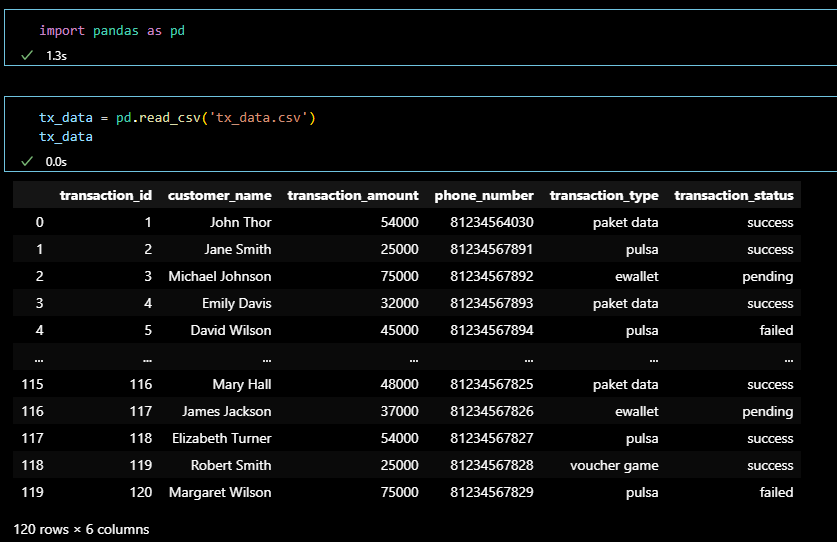
1. Mengubah nilai transaksi (transaction\_amount) dengan format seperti berikut: Rp 12.000



1. Pastikan data yang diambil adalah data transaksi dengan status sama dengan success.



1. Proses data cleaning dapat menggunakan library Pandas.



1. Hasil akhir data cleaning disimpan di dalam DataFrame.

