Mata Kuliah Big Data & Data Analytics

Laporan Praktikum Tugas 1: Retrive, Cleansing, and Manipulating

Dosen Pengampu: Agus Suhendar.S.T,.M.Eng

Disusun oleh:

Lathif Ramadhan (5231811022)

PROGRAM STUDI SAINS DATA

FAKULTAS SAINS DAN TEKNOLOGI

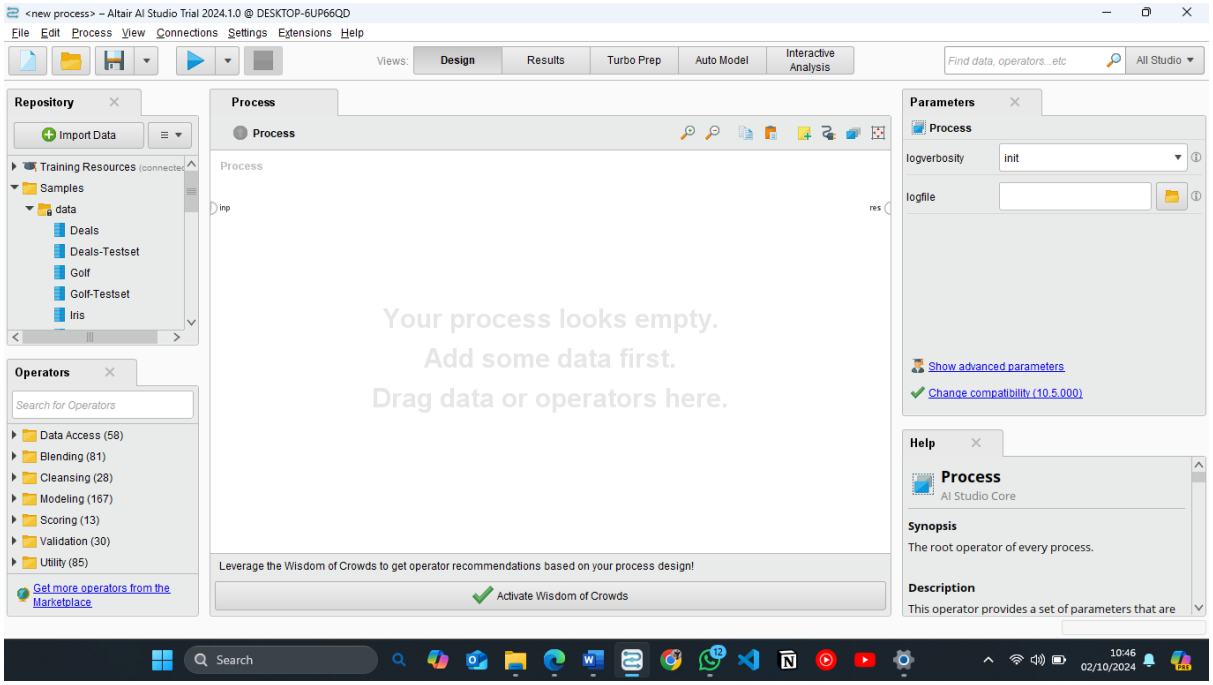
UNIVERSITAS TEKNOLOGI YOGYAKARTA

YOGYAKARTA

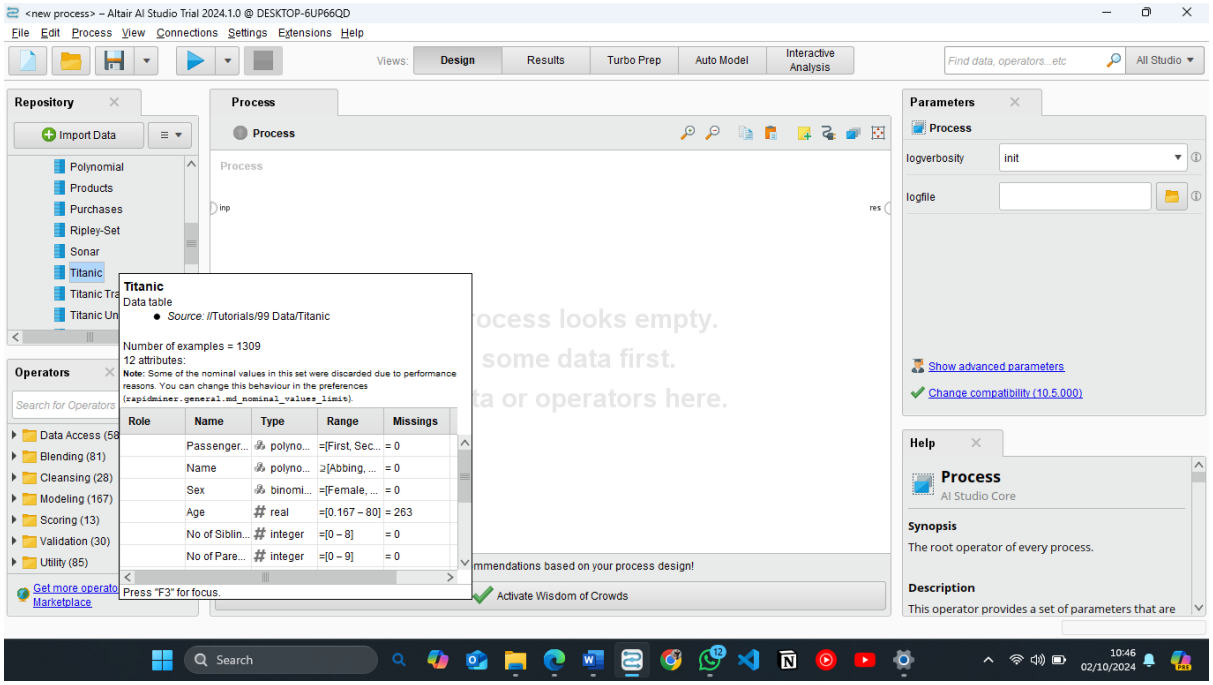
2024

1. Pertama-tama, kita buka dulu aplikasi **Rapidminer** atau **Altair AI Studio**.

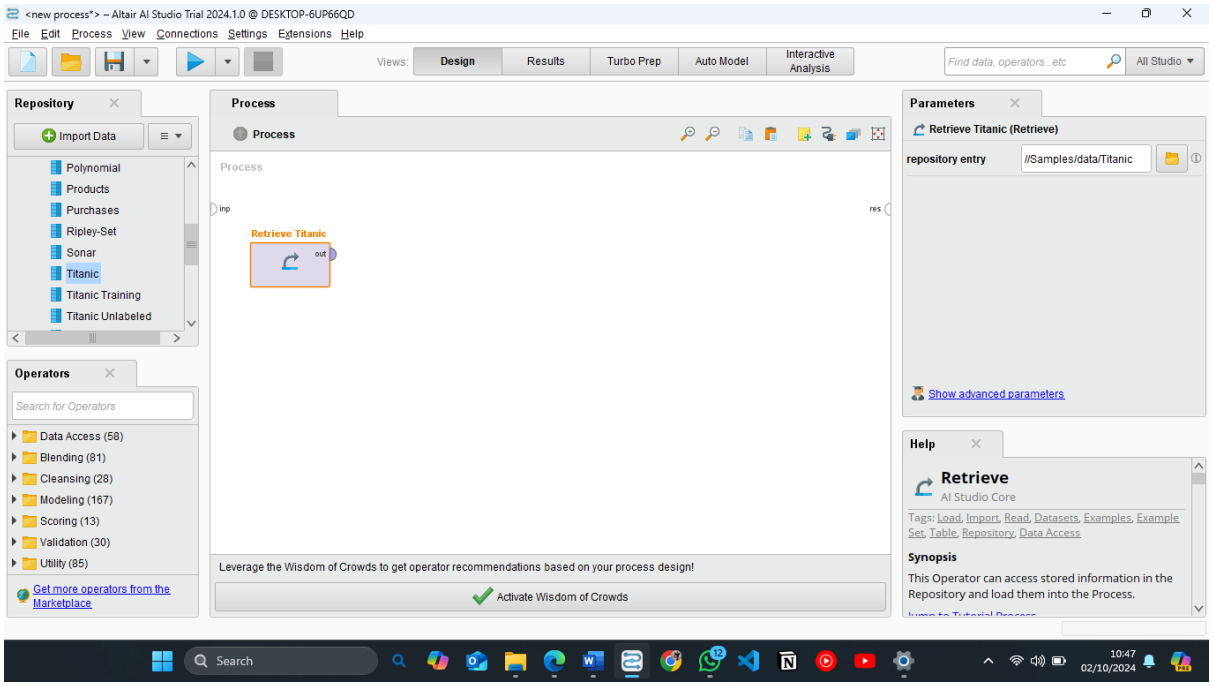
2. Setelah itu, kita buka pada ***Repository*** > ***Samples*** > ***data*** > cari data ***Titanic***!!



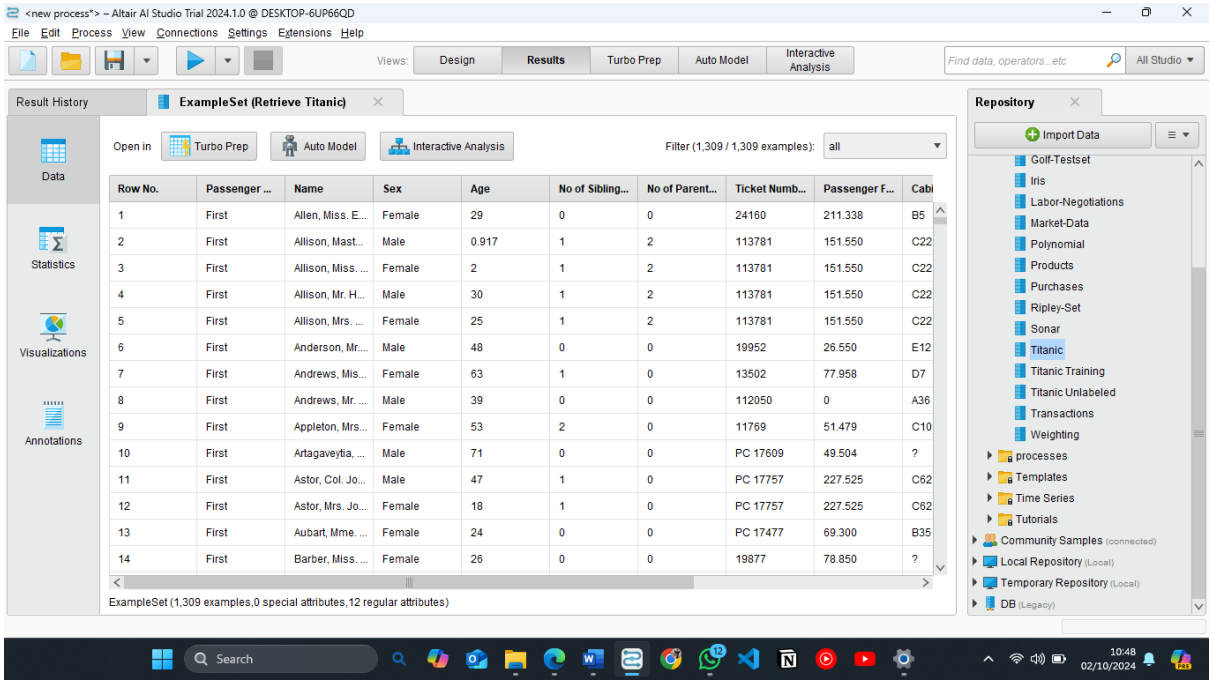
3. Nah, kalau dataset ***Titanic*** telah kita temukan, maka kita *drag* dataset tersebut ke dalam ***Process***.



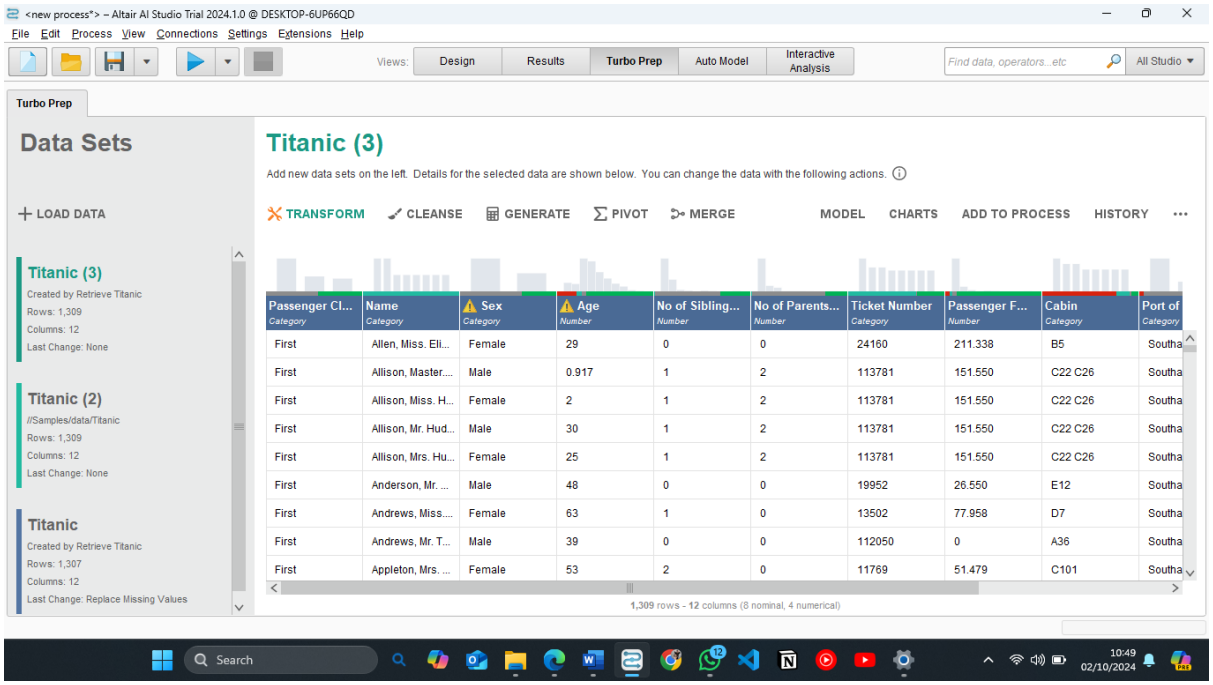
4. Ketika sudah di *drag*, tarik ***out*** ke ***res***. Lalu, kita ***play***/***run***.



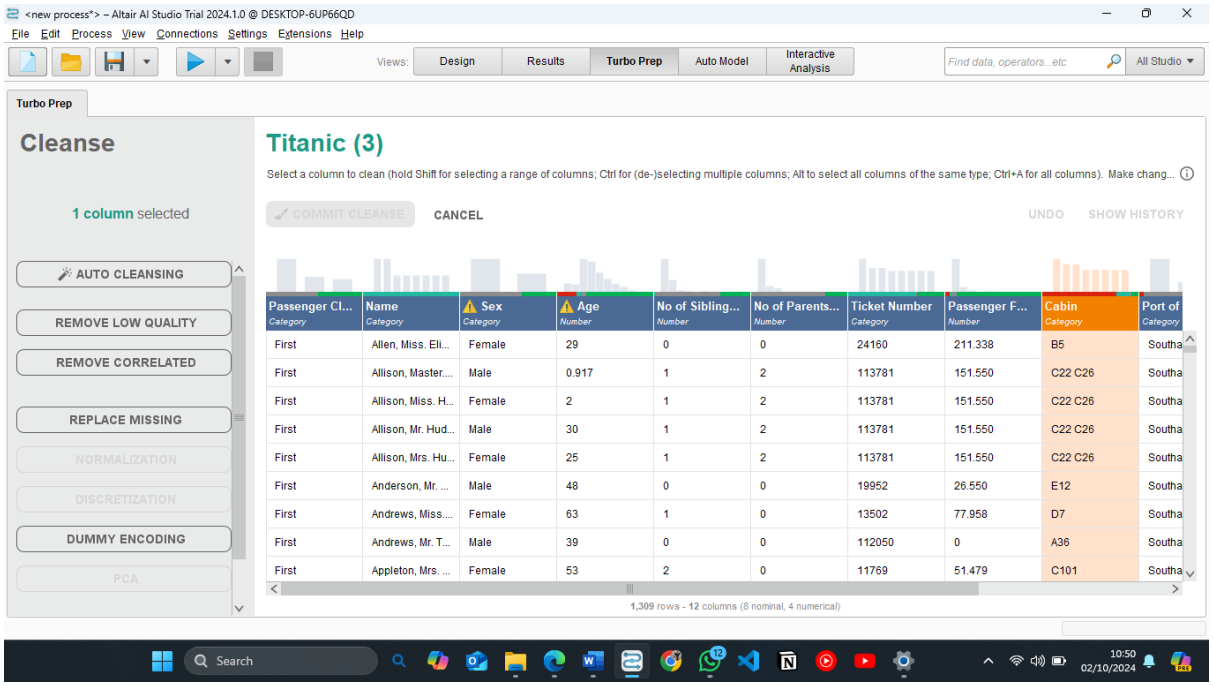
5. Maka, setelah di ***play***, hasilnya akan seperti ini.



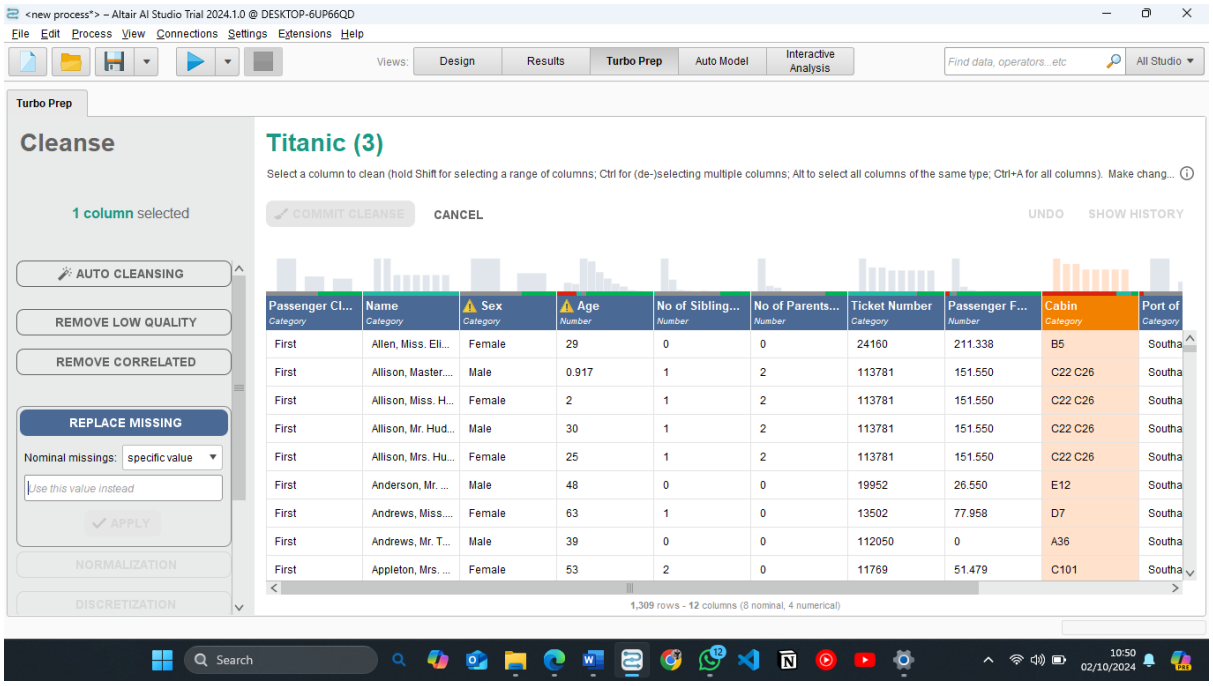
6. Lalu, kita pergi ke bagian ***Turbo Prep*** untuk memulai ***Manipulating Data*** dan ***Cleansing Data***. Kita fokus pada kolom-kolom yang diatas terdapat garis merahnya karena semakin panjang atau banyak garis merahnya, maka ada data yang kosong.

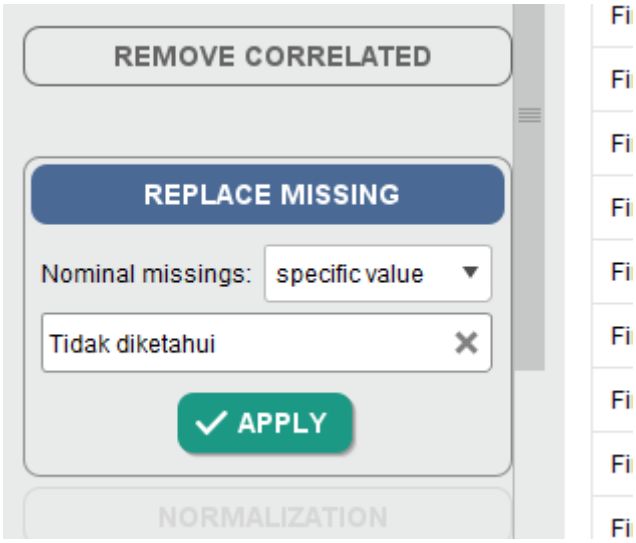


7. Untuk kolom pertama, kita pilih kolom ***Cabin***. Maka, seluruh data beserta nama kolom ***Cabin*** akan bewarna orange seperti ini.

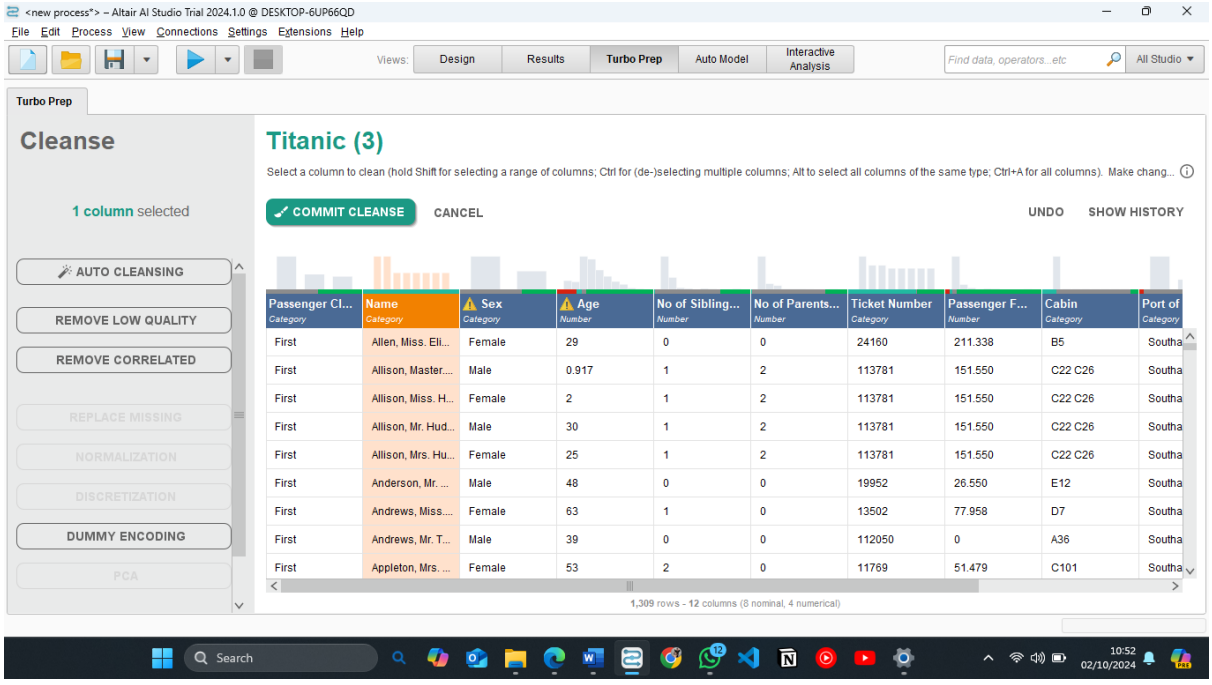


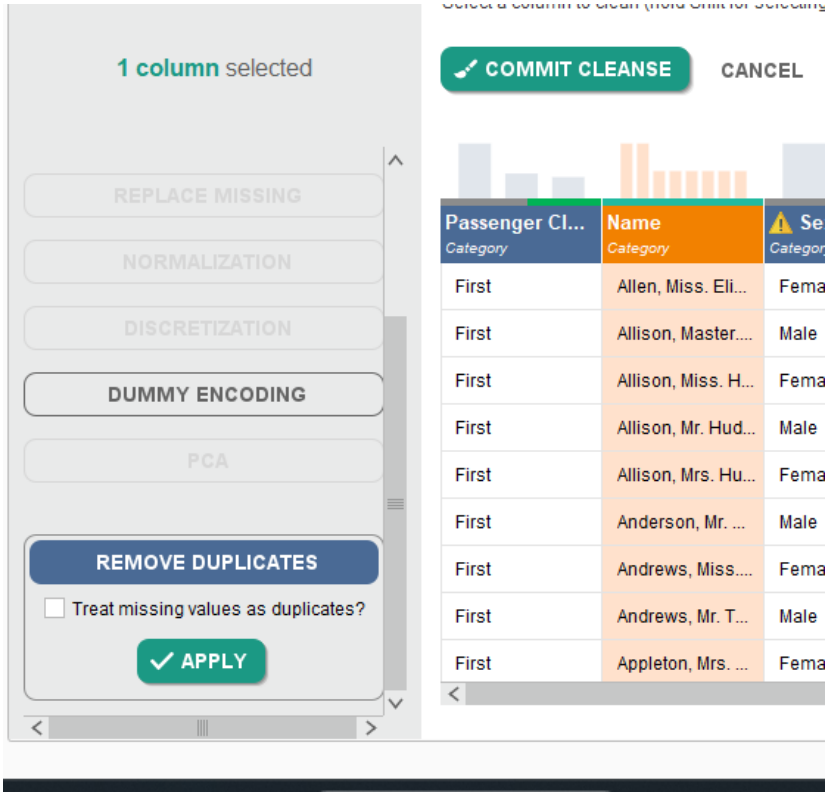
8. Kemudian, kita pergi ke ***Cleanse*** > ***Replace Missing*** > untuk ***nominal missing***, pilih ***spesific value*** > isi dengan kata “Tidak diketahui” > lalu klik ***Apply***.



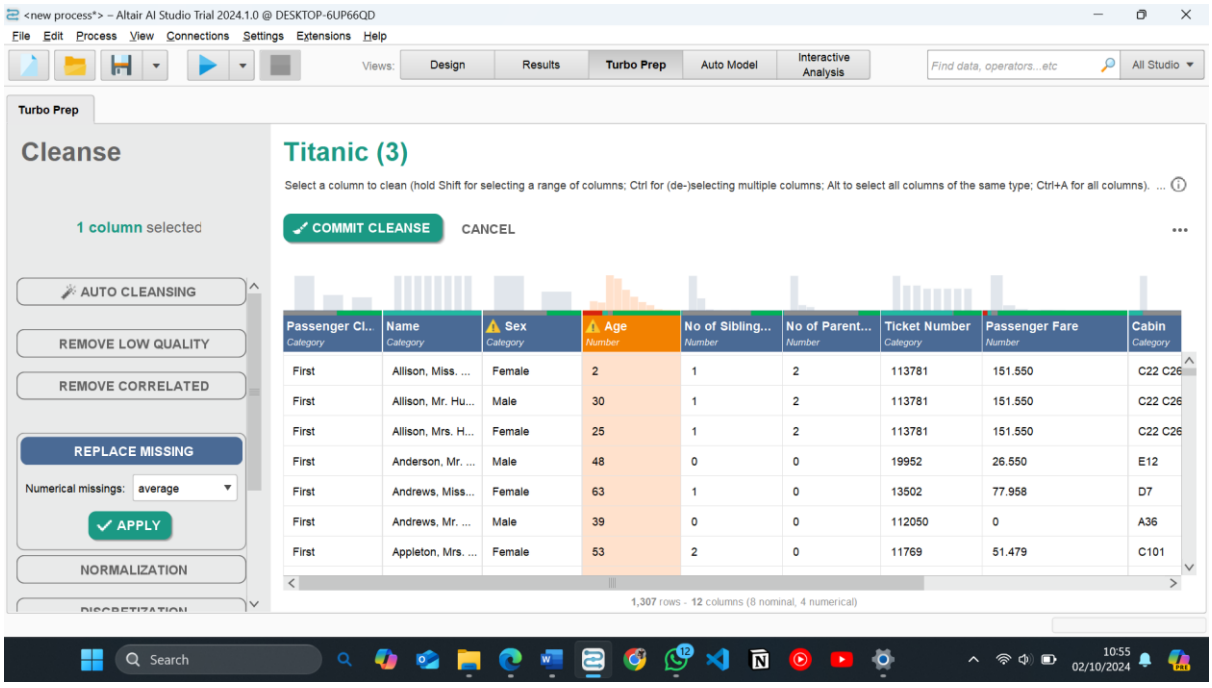


9. Untuk menghindari data nama *double*/*duplicate* atau salah input. Kita klik kolom ***Name*** > dibagian ***Cleanse*** > klik ***Remove Duplicate*** > ***Apply.***

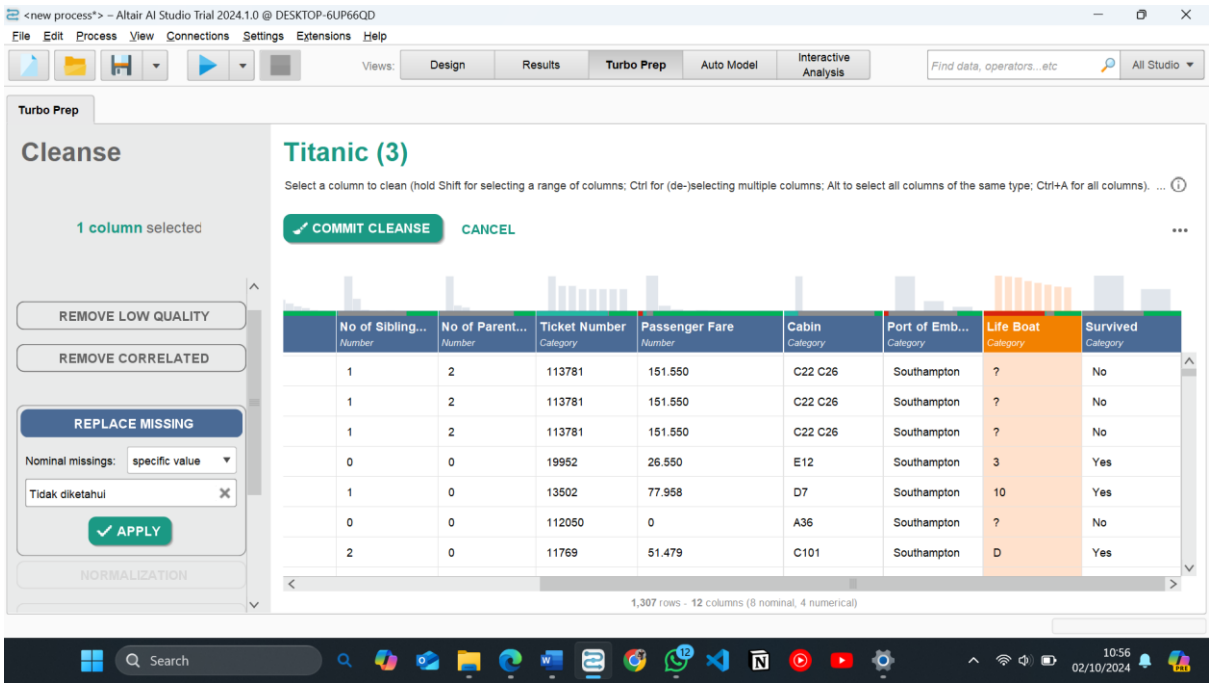




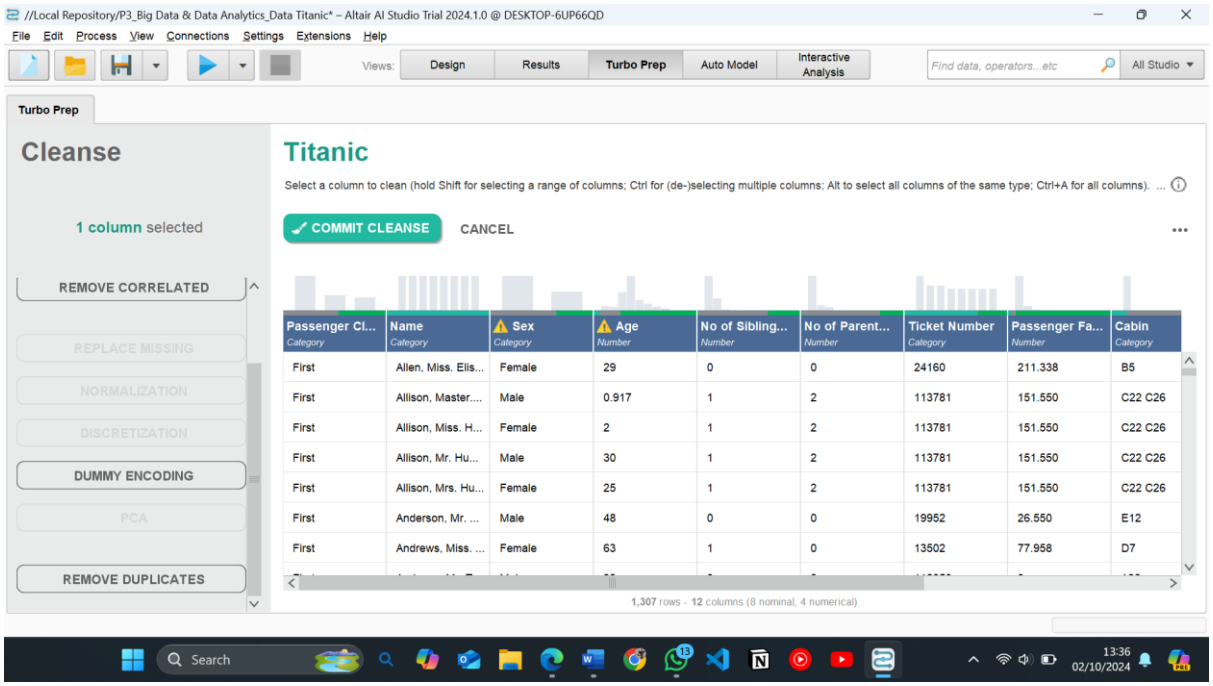
10. Karena data di kolom ***Age*** juga terdapat data yang kosong. Maka, kita klik ***Age*** > ***Replace Missing*** > ***Avarage***.



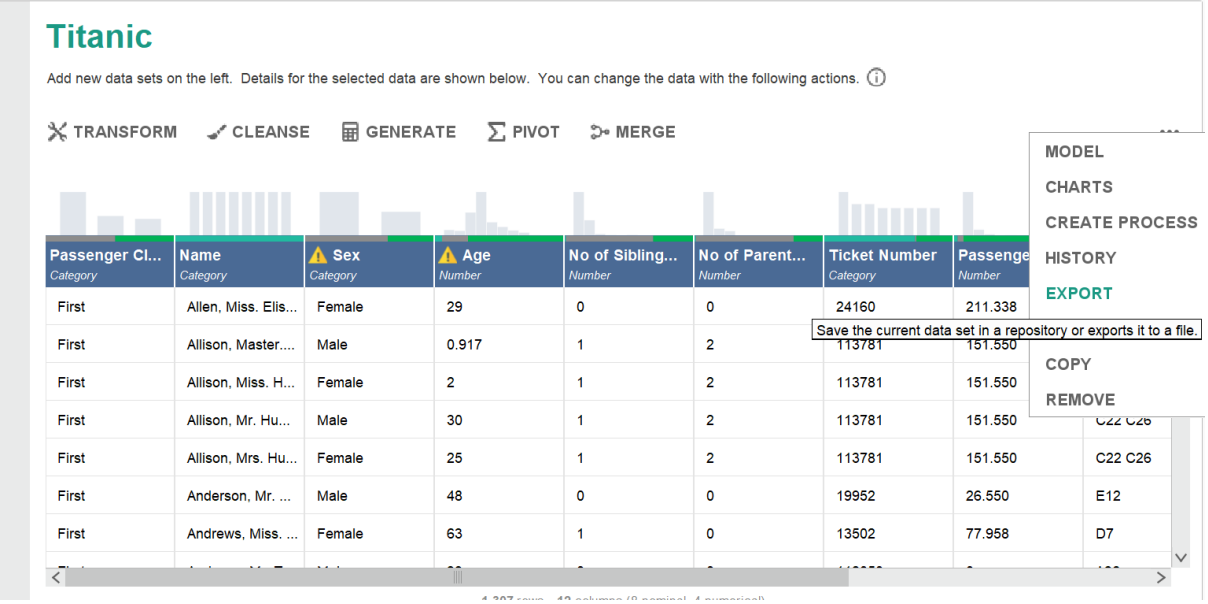
11. Selanjutnya kita olah data pada kolom ***Life Boat*** karena sebagian besar lebih banyak warna merah daripada hijau yang dimana artinya banyak data yang kosong/hilang. Maka, solusinya yaitu kita klik kolom ***Life Boat*** > ***Replace Missing*** > ***Specific Value***.



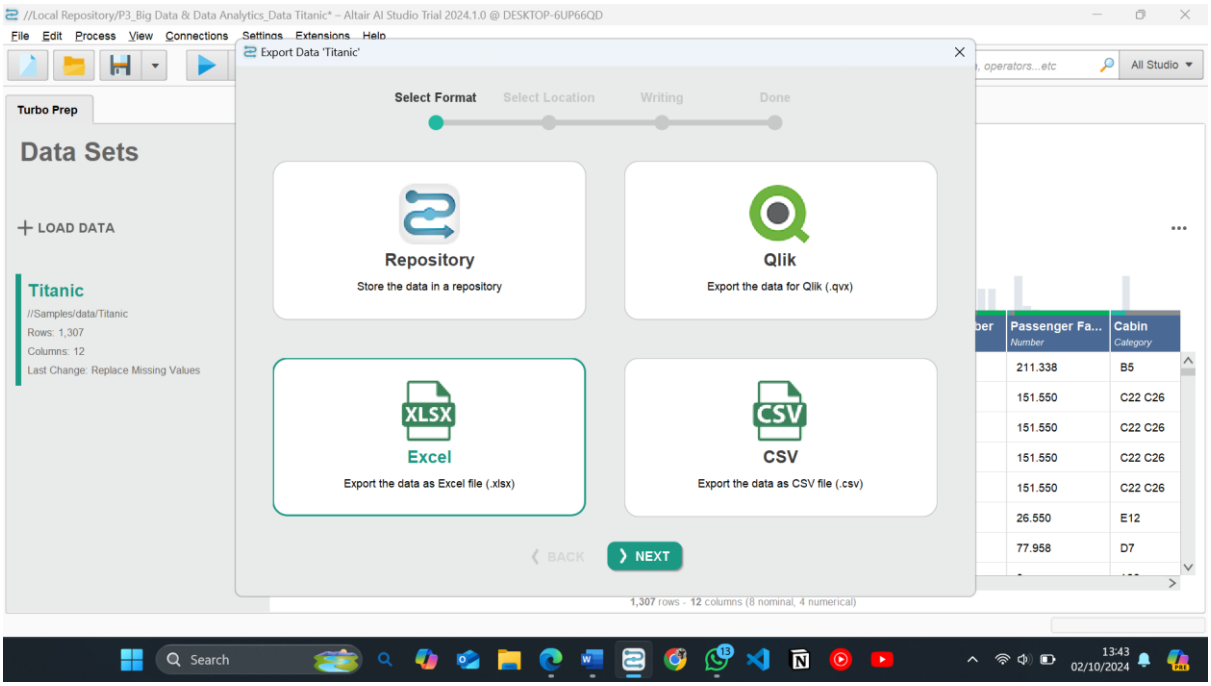
12. Setelah semua selesai, maka klik ***Commit***.



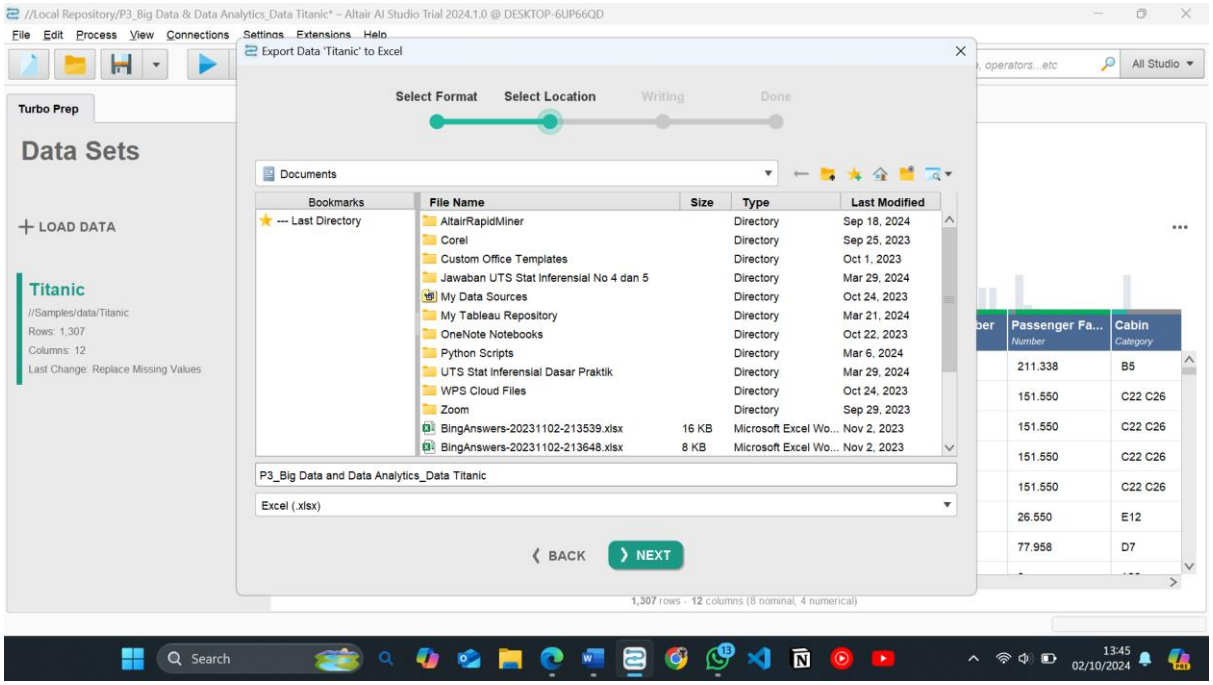
13. Setelah ***Commit***, kita bisa klik ***Export*** untuk menyimpan data yang sudah kita kerjakan.



14. Nah, disini terdapat 4 pilihan format file yang akan digunakan untuk menyimpan data kita. Contohnya, kita memilih format ***Excell*** untuk file kita, maka klik ***Excell*** > lalu ***Next***.



15. Lalu, kita beri nama file hasil pekerjaan kita dari data ***Titanic*** tadi. Kemudian pilih lokasi penyimpanan. Setelah itu klik ***Next***.



16. Nah, sudah selesai. Kita bisa ***Close***.