

Assignment Guidance: Data Cleaning

Data Science Bootcamp

Periode Pembelajaran

Objectives

1. Student mampu melakukan Data Cleaning terkait variasi dataset yang disediakan
2. Student mampu melakukan drop duplikat
3. Student mampu untuk melakukan tiap plot distribusi dan lakukan looping di tiap kolom yang ingin dianalisa sebelum melakukan handling outlier sehingga mampu menginvestigasi masing-masing nilai sebelum handling dan setelah handling outlier, mampu melakukan handling outlier untuk tujuan machine learning dan juga exploratory data analysis.
4. Student mampu melakukan pengecekan persentase missing value dan melakukan investigasi missing value sebelum di handling, serta mampu melakukan metode yang tepat untuk handling missing value dalam tujuan machine learning dan juga exploratory data analysis..
5. Student mampu mengkombinasikan encoding baik secara label encoder, mean encoding, one hot encoding yang disesuaikan dengan keberadaan data dan utarakan pendapat teman-teman dalam handling datanya untuk tujuan machine learning dan juga exploratory data analysis..
6. Setiap proses yang dilakukan pada poin 1–5 diatas harap insight yang didapatkan.

Deskripsi Assignment

Case study Data Cleaning

Detail Assignment

- Berikut untuk poin-poin assignment, harap dibaca dengan saksama sesuai keterangan yang ada pada soal Data Cleaning.
Link template code :
https://drive.google.com/file/d/1q9JhXn9NV--60_q5otsb3RI788t_NGc4/view?usp=sharing
- Download template code : (**di download atau Save a copy in drive, jangan dikerjakan di file template**)

3. Link Dataset :
<https://drive.google.com/drive/folders/1ZDQcn9Df-RIS05wm4luesWygVSIE3dyl?usp=sharing>

Tools

Google Collaboratory / Python / Jupyter notebook / Anaconda Navigator

Pengumpulan Assignment

Deadline :

Maksimal H+7 Kelas (Pukul 23.30 WIB)

Details :

Dikumpulkan dalam bentuk Link Google Collaboratory, secara INDIVIDU, di LMS

Notes :

Bilamana terdapat **salah pengumpulan file, private link pada google drive dengan alasan apapun tidak diterima dan akan diberikan nilai 0**, selain itu hal ini berlaku jika jawaban yang diberikan tidak sesuai dengan yang diminta pada parameter indikator penilaian / pada file template assignment.

Indikator Penilaian

No.	Aspek Penilaian	Parameter	Bobot Maksimal
1	Data Cleaning	Student Mampu menampilkan plot distribusi dan looping tiap kolom yang dibutuhkan untuk analisa + insight untuk tujuan machine learning dan juga exploratory data analysis.	25
2		Student mampu mengecek batas atas dan bawah dan melakukan outlier handling memakai IQR + insight untuk tujuan machine learning dan juga exploratory data analysis.	25
3		Student mampu menginvestigasi nilai missing value, dan handling missing value dengan metode yang tepat (median atau modus) + insight untuk tujuan machine learning dan juga exploratory data analysis.	25
4		Mampu mengimplementasikan Label Encoder dengan tepat dan mean encoding juga + insight untuk tujuan machine learning dan juga exploratory data analysis.	25

Referensi/Template

Link **template code** (download atau save di drive masing-masing dan aksesnya jangan private ketika mengumpulkan tugas ke LMS dalam bentuk link) :
https://drive.google.com/file/d/1q9JhXn9NV--60_q5otsb3RI788t_NGc4/view?usp=s_haring

Ketentuan Pencapaian Nilai:

Nilai minimum Lulus Penyaluran Kerja: 75

Nilai minimum Lulus Bootcamp: 65

Ketentuan Penilaian:

Mengumpulkan Assignment tepat waktu: Sesuai dengan nilai yang diberikan mentor

Mengumpulkan Assignment 12 jam setelah deadline: - 3 dari nilai yang diberikan mentor

Mengumpulkan Assignment 1 x 24 Jam setelah deadline: - 6 dari nilai yang diberikan mentor

Mengumpulkan Assignment 2 x 24 Jam setelah deadline: - 12 dari nilai yang diberikan mentor

Mengumpulkan Assignment 3 x 24 Jam setelah deadline: - 18 dari nilai yang diberikan mentor

Mengumpulkan Assignment 4 x 24 Jam setelah deadline: - 24 dari nilai yang diberikan mentor

Mengumpulkan Assignment 5 x 24 Jam setelah deadline: - 30 dari nilai yang diberikan mentor

Mengumpulkan Assignment 6 x 24 Jam setelah deadline: - 36 dari nilai yang diberikan mentor

Mengumpulkan Assignment 7 x 24 Jam setelah deadline: - 42 dari nilai yang diberikan mentor