

Homework EDA

Final Project - Stage 1



Estimasi Waktu Pengerjaan

 **3 - 5 jam**

Jumlah Soal

 **5 Soal**

Total Point

 **100 poin**

Teknis Pengerjaan

1. Pekerjaan dilakukan secara **berkelompok, sesuai kelompok Final Project**
2. **Masing-masing anggota kelompok tetap perlu submit ke LMS** (jadi bukan perwakilan)
3. File yang perlu dikumpulkan:
 - File **jupyter notebook** (.ipynb) yang berisi source code. Tuliskan juga insights yang ditemukan ke notebook tersebut sebagai text markdown.
 - File **laporan homework** (.pdf) yang berisi rangkuman dari insight-insight yang diperoleh, beserta rekomendasinya (rekomendasi pre-processing untuk EDA, dan rekomendasi bisnis untuk business insight).
4. Upload hasil pengerjaanmu melalui LMS.
 - Masukkan semua file ke dalam **1 file** dengan format **ZIP**.
 - Nama File:
EDA - <Nama Kelompok>.zip

1. Descriptive Statistics (15 poin)

Gunakan function **info** dan **describe** pada dataset final project kalian. Tuliskan hasil observasinya, seperti:

- A. Apakah ada kolom dengan tipe data kurang sesuai, atau nama kolom dan isinya kurang sesuai?
- B. Apakah ada kolom yang memiliki nilai kosong? Jika ada, apa saja?
- C. Apakah ada kolom yang memiliki nilai summary agak aneh?
(min/mean/median/max/unique/top/freq)

* Untuk masing-masing jenis observasi, tuliskan juga jika tidak ada masalah, misal untuk A: "Semua tipe data sudah sesuai"

2. Univariate Analysis (25 poin)

Gunakan visualisasi untuk melihat distribusi masing-masing kolom (feature maupun target). Tuliskan hasil observasinya, misalnya jika ada suatu kolom yang distribusinya menarik (misal skewed, bimodal, ada outlier, ada nilai yang mendominasi, kategorinya terlalu banyak, dsb). Jelaskan juga apa yang harus di-follow up saat data pre-processing.

3. Multivariate Analysis (15 poin)

Lakukan multivariate analysis (seperti correlation heatmap dan category plots, sesuai yang diajarkan di kelas). Tuliskan hasil observasinya, seperti:

- A. Bagaimana korelasi antara masing-masing feature dan label. Kira-kira feature mana saja yang paling relevan dan harus dipertahankan?
- B. Bagaimana korelasi antar-feature, apakah ada pola yang menarik? Apa yang perlu dilakukan terhadap feature itu?

* Tuliskan juga jika memang tidak ada feature yang saling berkorelasi

4. Business Insight (30 poin)

Selain EDA, lakukan juga beberapa analisis dan visualisasi untuk menemukan suatu business insight. Tuliskan minimal 3 insight, dan berdasarkan insight tersebut jelaskan rekomendasinya untuk bisnis.

5. Git (15 poin)

Upload project teman-teman di sebuah repository git. Berkolaborasi di Git jika ada perubahan version dari waktu ke waktu.

- A. Buat Repository Git
- B. Upload file notebook atau file pengerjaan lainnya pada repository tersebut

Untuk file README, dapat merupakan summary insight yang telah didapatkan dari EDA.

Selamat Mengerjakan!