



CASE STUDY

Data Science Competition



DATA SCIENCE COMPETITION GELAR RASA 2025

AKSES DATASET

Dataset : FMCG Personal Care – Synthetic Dataset

Sumber : bit.ly/DatasetDSCGelarRasa2025

DESKRIPSI DATASET

Dataset ini mensimulasikan data transaksi penjualan produk kategori Personal Care (Fast Moving Consumer Goods) selama periode 1 Januari 2020 hingga 31 Desember 2025. Data bersifat sintetis dan dirancang menyerupai pola nyata di industri FMCG, mencakup atribut seperti produk, harga, jumlah penjualan, wilayah, saluran distribusi, dan waktu transaksi.

TUJUAN

Analisis terhadap dataset ini diarahkan untuk:

1. Innovation Radar

Mengidentifikasi produk dengan potensi pertumbuhan tinggi atau inovasi yang menarik minat konsumen.

2. Trend Forecasting

Memprediksi tren penjualan dan preferensi konsumen di masa mendatang.

3. Product Cannibalization Analysis

Mengevaluasi apakah peluncuran produk baru mengurangi penjualan produk lain dalam kategori yang sama.

DATA SCIENCE COMPETITION

GELAR RASA 2025

FILE DI PAKET

- **products.csv** : Master produk (product_id, product_name, brand, type, size_ml, base_price, launch_date)
- **marketing.csv** : Kampanye pemasaran (campaign_id, product_id, spend, channel, engagement_rate)
- **reviews.csv** : Sampel review pelanggan (~10k rows)
- **sales.csv** : Transaksi penjualan (~1,000,000 rows) -- kolom-kolom penting ada di bawah.

DETAIL

- **transaction_id** : Unique UUID untuk setiap transaksi.
- **date** : Tanggal transaksi (YYYY-MM-DD).
- **product_id** : Kode produk (PC001.. PC015).
- **region** : Kota/region tempat transaksi (Jakarta, Bandung, ...).
- **channel** : Channel penjualan (Shopee, Tokopedia, Official Store, Alfamart, Indomaret, Hypermarket).
- **units_sold** : Jumlah unit terjual dalam transaksi tersebut.
- **avg_price** : Harga rata-rata per unit setelah diskon (IDR).
- **discount_pct** : Persentase diskon yang diberikan pada transaksi.
- **revenue** : units_sold * avg_price.
- **days_since_launch** : Selisih hari antara tanggal transaksi dan tanggal launch produk (negatif jika sebelum launch).

DATA SCIENCE COMPETITION

GELAR RASA 2025

PERATURAN PENGERJAAN

1. Waktu penggerjaan proyek dimulai dari tanggal 1 November hingga 15 November 2025.
2. Peserta dilarang menggunakan *dataset* selain yang diberikan oleh panitia.
3. Peserta menggunakan bahasa pemrograman Python.
4. Tidak ada batasan dalam penggunaan *library* dan *tools*.
5. Panitia tidak bertanggung jawab atas *link* penggerjaan yang tidak dapat diakses akibat pengaturan akses yang terkunci.
6. Hasil akhir proyek dikumpulkan dalam bentuk file .ipynb yang berisi:
 - Kode (syntax)
 - *Output*
 - Interpretasi atau penjelasan
7. Sanksi keterlambatan:
 - Setiap keterlambatan pengumpulan akan dikenakan sanksi pengurangan 2 poin per jam (berlaku kelipatan).
 - Peserta yang tidak mengumpulkan hingga H+1 akan didiskualifikasi.



DATA SCIENCE COMPETITION GELAR RASA 2025

TEKNIS PENGUMPULAN

- Batas pengumpulan file adalah 15 November 2025 pukul 23.59 WIB
- Pengumpulan dilakukan oleh ketua tim atau perwakilan.
- File dikumpulkan melalui link berikut:
bit.ly/PengumpulanProyekDSCGelarRasa2025 dengan format nama file: DSC2025_NamaTim_NamaKetua.ipynb.

DATA SCIENCE COMPETITION

GELAR RASA 2025

KRITERIA PENILAIAN

Detail Aspek	Rentang Poin
<p>Uji asumsi dan evaluasi dari model yang dibangun</p> <ul style="list-style-type: none"> Model memenuhi uji asumsi klasik dan residual. MSE, RMSE, MAPE atau evaluasi lainnya.= 	15
<p>Kualitas data <i>preprocessing</i> dan <i>feature engineering</i></p> <ul style="list-style-type: none"> Kreativitas dalam memilih <i>features</i>. Efektivitas pembersihan data dalam penanganan <i>missing value</i>, <i>outlier</i>, <i>duplicate</i>, dan lain-lain. 	20
<p>Inovatif</p> <ul style="list-style-type: none"> Menggunakan teknik terkini atau kombinasi dari beberapa metode untuk meningkatkan performa model. Inovasi dalam menangani <i>imbalanced dataset</i>. 	30
<p>Visual</p> <ul style="list-style-type: none"> Penyajian hasil yang menarik dan mudah dipahami. 	25
<p>Dokumentasi dan reproduksibilitas</p> <ul style="list-style-type: none"> Dokumentasi kode yang jelas dan lengkap. Kemudahan dalam menjalankan ulang kode yang dikirim dengan hasil yang sesuai pada presentasi. 	10
Total	100