**FILE 1 – Soal Analisis Data (Tanpa Tantangan)**

1. **Analisis Tren Waktu (Time Series) per Produk**  
   Untuk tiap produk, identifikasi pola musiman (seasonality) atau tren jangka panjang pada metrik penjualan (asumsi: tambahkan kolom monthly\_sales pada product\_launch.csv). Apakah ada periode ‘aman’ atau ‘berisiko’ untuk meluncurkan produk di masing-masing negara?
2. **Clustering Negara Berdasarkan Respons Produk**  
   Cluster negara menggunakan gabungan fitur: rata-rata rating pelanggan, GDP per kapita, penetrasi internet, dan selisih peluncuran kompetitor. Berapa banyak cluster optimal? Apa karakteristik masing-masing cluster?
3. **Feature Importance pada Model Prediksi “Success”**  
   Bangun model klasifikasi (misal Random Forest) untuk memprediksi kolom success. Gunakan semua fitur numerik & kategorikal (termasuk engineered features seperti jumlah kata ulasan positif vs negatif, intensitas campaign per hari sejak launch). Fitur apa yang paling penting menurut model?
4. **Analisis Sentimen Ulasan Pelanggan Over Time**  
   Lakukan sentiment analysis pada teks ulasan (customer\_reviews.csv) lalu plot evolusi rata-rata skor sentiment tiap bulan untuk produk sukses vs gagal. Apakah sentimen bisa menjadi *leading indicator*?
5. **Deteksi Anomali pada Budget Kampanye**  
   Identifikasi produk dengan campaign\_budget yang secara signifikan lebih tinggi/rendah dibanding produk lain dengan tingkat sales (asumsi: tambahkan kolom monthly\_sales). Apakah ada outlier, dan bagaimana performanya?
6. **Analisis Competitor Pressure**  
   Uji korelasi antara competitor\_launch\_gap\_days dan probabilitas sukses. Apakah semakin dekat peluncuran kompetitor, semakin besar risiko kegagalan?
7. **Segmentasi Produk berdasarkan Price Sensitivity**  
   Cluster produk berdasarkan price dan rata-rata rating pelanggan. Identifikasi segmen harga: high-end, mid-range, budget. Bagaimana masing-masing segmen ini tersebar di negara?

**FILE 2 – Tantangan dan Tips Teknis Penyelesaian**

1. **Analisis Tren Waktu**
   * Time series decomposition
   * Rolling average
   * Seasonal outliers (Z-score/IQR)
   * Visualisasi tren dan musiman per produk
2. **Clustering Negara**
   * Feature scaling
   * Elbow & Silhouette method
   * Interpretasi cluster & centroid
   * Visualisasi PCA/TSNE
3. **Feature Importance (Random Forest)**
   * Encoding fitur kategorikal
   * TF-IDF/Count Vectorizer
   * SMOTE untuk imbalance
   * Interpretasi feature importance
4. **Analisis Sentimen Review**
   * Lexicon-based atau pretrained model
   * Agregasi skor bulanan
   * Perbandingan sukses vs gagal
   * Visualisasi temporal
5. **Deteksi Anomali Budget**
   * Z-score/IQR budget vs sales
   * Scatterplot outlier detection
   * Evaluasi performa produk outlier
6. **Competitor Pressure Analysis**
   * Scatterplot & regresi
   * Stratifikasi gap
   * Chi-square / regresi logistik
   * Interpretasi odds ratio
7. **Segmentasi Produk (Price Sensitivity)**
   * Clustering 2D: price & rating
   * Visualisasi scatter + hull
   * Crosstab negara vs segmen