

Tugas Mandiri: Analisis Data E-Commerce untuk Pertumbuhan Bisnis

Konteks Bisnis

Anda baru saja bergabung sebagai Data Engineer pertama di "RetailKita", sebuah startup e-commerce yang sedang berkembang. Manajemen ingin membuat keputusan yang lebih didorong oleh data untuk meningkatkan penjualan dan memahami perilaku pelanggan. Namun, data saat ini tersebar di beberapa file CSV dan belum ada infrastruktur data terpusat.

Tugas Anda adalah mengambil data mentah ini, mengubahnya menjadi format yang dapat dianalisis, mengekstrak wawasan kunci, dan mengusulkan langkah selanjutnya yang lebih canggih.

Tugas Anda

Tugas ini dibagi menjadi tiga bagian yang mencerminkan peran hibrida yang akan Anda jalani.

Bagian 1: Rekayasa dan Persiapan Data

Tujuan dari bagian ini adalah untuk mengambil data mentah yang terpisah dan membangun fondasi yang bersih dan terstruktur untuk analisis.

1. **Proses ETL Sederhana:** Tulis sebuah skrip (kami sarankan menggunakan Python) yang melakukan hal berikut:
 - Membaca tiga file CSV.
 - Membersihkan data.
 - Menggabungkan data dari ketiga file tersebut menjadi satu tabel analitik tunggal yang komprehensif.
 - Membuat kolom baru yang berguna, seperti `total_price` (`quantity * price`).
2. **Memuat ke Database:** Muat tabel analitik yang sudah bersih tersebut ke dalam database. Untuk tugas ini, silahkan menggunakan database lokal pilihan ada (cth : seperti **SQLite**).

Bagian 2: Analisis dan Wawasan

Sekarang setelah data Anda bersih dan terstruktur, gunakan **SQL** untuk menjawab pertanyaan bisnis kunci berikut:

1. **Produk Terlaris:** Tulis kueri SQL untuk mengidentifikasi **5 produk terlaris** berdasarkan total pendapatan (`total_price`).
2. **Pelanggan Paling Berharga:** Tulis kueri SQL untuk menemukan **10 pelanggan teratas** berdasarkan total uang yang mereka habiskan.

3. **Tren Penjualan Bulanan:** Tulis kueri SQL untuk menghitung total pendapatan per bulan.
4. **Visualisasi Kunci:** Buat satu visualisasi (misalnya, grafik batang atau garis) yang paling efektif untuk menunjukkan tren penjualan bulanan kepada tim manajemen. Anda bisa menggunakan library Python seperti Matplotlib/Seaborn atau alat lain yang Anda sukai.

Bagian 3: Pemikiran Strategis

Tim Pemasaran ingin membuat program untuk meningkatkan retensi pelanggan. Mereka meminta Anda untuk membantu mengidentifikasi pelanggan yang berisiko *churn* (berhenti bertransaksi).

Anda boleh membangun model *machine learning* untuk tugas ini. Jelaskan pendekatan Anda dalam sebuah ringkasan singkat (1-2 paragraf):

1. Pendekatan Pemodelan:

- Bagaimana Anda akan mendefinisikan *churn* dalam konteks data ini?.
- Jenis model *machine learning* apa (misalnya, klasifikasi, regresi) yang akan Anda gunakan untuk memprediksi *churn*? Jelaskan secara singkat mengapa Anda memilih model tersebut sebagai titik awal.
- Fitur apa saja yang akan Anda buat dari data yang ada untuk melatih model Anda? .

Hasil yang Diharapkan

Harap kumpulkan semua hasil Anda dalam sebuah repositori Git (misalnya, di GitHub atau GitLab) dan kirimkan tautannya kepada kami. Repositori tersebut harus berisi:

1. Kode Sumber:

- Skrip Python untuk proses ETL Anda (dari Bagian 1).
- File .sql yang berisi semua kueri SQL Anda (dari Bagian 2).

2. Hasil Analisis:

- File CSV atau teks yang berisi jawaban untuk pertanyaan 1 & 2 dari Bagian 2.
- Gambar visualisasi Anda (dari Bagian 2, pertanyaan 4).

3. Presentasi Singkat:

- Sebuah dokumen singkat (misalnya, PDF, Google Slides, atau Markdown) yang merangkum:
 - Temuan utama Anda dari analisis di Bagian 2.
 - Penjelasan pendekatan sains data Anda dari Bagian 3.
 - Asumsi apa pun yang Anda buat selama mengerjakan tugas.

4. Dokumentasi (README.md):

- File README.md yang jelas di repositori Anda yang menjelaskan proyek ini dan memberikan instruksi sederhana tentang cara menjalankan kode Anda.
-

Kriteria Evaluasi

Kami akan mengevaluasi kiriman Anda berdasarkan:

- **Kualitas dan Kejelasan Kode:** Apakah kode Anda terstruktur dengan baik, efisien, dan mudah dibaca?
- **Akurasi Analisis:** Apakah kueri SQL dan analisis Anda secara akurat menjawab pertanyaan yang diajukan?
- **Pemecahan Masalah:** Bagaimana Anda menangani potensi masalah dalam data (jika ada)?
- **Komunikasi:** Seberapa jelas Anda dalam mempresentasikan temuan dan ide-ide Anda dalam dokumen ringkasan?
- **Pemahaman Bisnis:** Apakah analisis dan rekomendasi Anda relevan dengan konteks bisnis yang diberikan?

Terima kasih atas waktu dan usaha Anda. Kami sangat menantikan untuk melihat hasil kerja Anda!