

Bagian A

Anda telah diberikan kumpulan data yang berisi 1.000 catatan penjualan benih dan pupuk dari cabang regional penyedia perlengkapan pertanian. Setiap catatan mencakup informasi berikut:

- **Sales ID** (menunjukkan faktur penjualan tertentu)
- **Customer ID** (menunjukkan siapa pelanggan yang melakukan pembelian, dengan total 4 pelanggan)
- **SalesRep ID** (menunjukkan perwakilan penjualan yang melakukan penjualan, dengan 8 perwakilan berbeda)
- **Product ID** (menunjukkan produk spesifik yang dijual, dengan total 6 produk berbeda)
- **Sales Quantity** (jumlah produk yang dibeli dari setiap penjualan, dalam ton)
- **Sales Amount** (nilai moneter setiap penjualan, dalam US\$)
- **Date** (tanggal faktur penjualan, Tahun, Bulan, Hari)

Tugas:

1. Analisis Deskriptif:

- a. Hitung total penjualan tahunan untuk setiap produk selama periode tertentu.
- b. Tentukan rata-rata penjualan per pelanggan berdasarkan karakteristiknya.
- c. Identifikasi 3 perwakilan penjualan teratas berdasarkan total penjualan.

2. Visualisasi Data:

Menggunakan Power BI atau aplikasi serupa, buat visualisasi berikut:

1. Total penjualan untuk setiap produk yang disegmentasi berdasarkan jenis pelanggan.
2. Proporsi total penjualan menurut jenis pelanggan untuk setiap bulan dalam setahun. Visualisasi ini harus menyoroti tren musiman penjualan di seluruh jenis pelanggan.
3. Kinerja penjualan dari tiga perwakilan penjualan teratas selama periode pelaporan tiga tahun. Sertakan penanda untuk perubahan atau peristiwa signifikan yang mungkin berdampak pada kinerja penjualan mereka.

3. Analisis dan Pelaporan:

Tulis analisis singkat (200-300 kata) yang menginterpretasikan hasil analisis deskriptif dan visualisasi Anda. Jelaskan *insights* apa yang dapat diambil dari analisis dan visualisasi ini. Diskusikan tren, anomali, atau *insights* apa pun yang dapat menginformasikan strategi penjualan benih dan pupuk di masa depan. Pertimbangkan bagaimana sertifikasi perwakilan penjualan dapat mempengaruhi kinerja penjualan.

Petunjuk/Catatan:

- Pastikan bagan tersebut diberi label dengan jelas dan menyertakan *legend* jika perlu.
- Jelaskan *insights* apa yang dapat diambil dari visualisasi ini.
- Sertakan anotasi atau *insights* yang relevan mengenai tren.
- Tekankan pada penyampaian cerita sebagai bagian penting dari analisis

Bagian B.

Perusahaan Anda, Loans 'R Us, adalah pemimpin dalam industri kartu kredit dan pinjaman yang dipercaya oleh pelanggan di seluruh dunia. Meskipun kinerja perusahaan baik, Anda merasa perusahaan dapat berbuat lebih baik dengan mengurangi jumlah penipuan kartu kredit yang meningkat akhir-akhir ini. Salah satu karyawan Anda menyarankan agar Anda memanfaatkan kekuatan analisis data dan menganalisis data yang Anda miliki di database. Beberapa pertanyaan/kekhawatiran yang ingin Anda temukan jawabannya:

1. Buat model untuk memprediksi jumlah tagihan kartu kredit menggunakan variabel di dalam kumpulan data. Anda bebas menggunakan variabel apa pun yang tersedia di dataset (kami sarankan menggunakan data numerik). Apakah Anda mampu membuat model yang akurat dan andal? Variabel manakah yang (atau tidak) relevan dalam memprediksi jumlah tagihan kartu kredit?
2. Buat model clustering dengan tiga (3) cluster menggunakan variabel berikut. Jelaskan karakteristik masing-masing cluster. Dari mana sebagian besar pelanggan Pinjaman 'R Us berasal/berlokasi? Rekomendasi apa yang dapat Anda berikan dari klaster tersebut untuk meningkatkan jumlah pelanggan Loans 'R Us?
 - a. amt: Jumlah Transaksi
 - b. lat: Lintang Lokasi Pembelian
 - c. long: Bujur Lokasi Pembelian

- d. `city_pop`: Populasi Kota Pemegang Kartu Kredit
 - e. `merch_lat`: Garis Lintang Lokasi Pedagang
 - f. `merch_long` : Lokasi Pedagang di Bujur
3. Untuk membantu pekerja *frontline* dalam menilai penipuan kartu kredit, Anda perlu membuat model klasifikasi (K nearest neighbor, Naïve Bayes, dll.) berdasarkan kumpulan data yang tersedia untuk Anda. Anda bebas menggunakan variabel apa pun dalam dataset. Salah satu karyawan Anda menyarankan untuk membagi kumpulan data menjadi dua (pelatihan dan pengujian) untuk membuat model klasifikasi ini. Karena ketidak rataan data, pendekatan bertingkat untuk membagi kumpulan data mungkin diperlukan.
- a. Jelaskan variabel apa yang Anda gunakan dari kumpulan data, metode klasifikasi yang digunakan, dan bagaimana kumpulan data tersebut dibagi menjadi data pelatihan dan pengujian.
 - b. Seberapa bagus model klasifikasi Anda?
 - i. Berapa banyak yang diprediksi sebagai penipuan sebenarnya adalah penipuan
 - ii. Berapa banyak yang digadang-gadang sebagai penipuan sebenarnya bukanlah penipuan
 - iii. Berapa banyak yang digadang-gadang bukan penipu ternyata bukan penipu
 - iv. Berapa banyak yang digadang-gadang bukan penipu, ternyata penipuan
 - c. Jelaskan bagaimana Anda akan menggunakan model yang baru dibuat ini untuk membantu pekerja *frontline* mengambil keputusan berdasarkan prediksi yang dibuat oleh model tersebut.

Info Kolom

`trans_date_trans_time`: Tanggal Waktu Transaksi

`merchant`: Nama Pedagang

`category`: Kategori Pedagang

`amt`: Jumlah Transaksi

city: Kota Pemegang Kartu Kredit

state: Negara Pemegang Kartu Kredit

lat: Lintang Lokasi Pembelian

long: Bujur Lokasi Pembelian