

Capstone Project - Module 2

Diberikan dataset 'Northwind' kepada siswa yang bertujuan untuk melakukan analisa pada dataset tersebut, dimana para siswa dimintai Penjelasan tentang data dan konteks permasalahan yang hendak dianalisis. Dalam projek ini, siswa akan memposisikan diri sebagai data analyst dalam sebuah perusahaan. Siswa akan membuat serangkaian pertanyaan sesuai dengan fokus analisis, pertanyaan dibuat berdasarkan keperluan bisnis dan masalah yang mungkin dihadapi oleh perusahaan terkait sesuai dengan fokus analisisnya.

Database Information

Sumber Database :

https://drive.google.com/drive/folders/1fTHrwh_gcLsOFKXHnUzUGEu_APxLoD9i

General Question

1. Bagaimana konteks bisnis berdasarkan dengan data yang telah diberikan?
2. Ada berapa banyak tabel yang disediakan oleh database? Jabarkan setiap tabelnya!

SQL

1. Apakah tabel customers, orders, ordersdetail, products dan categories dapat digabungkan menjadi 1 tabel? Jika memungkinkan, tampilkan tabel yang memuat informasi transaksi yang melibatkan kelima tabel tersebut.
2. Berapa banyak jenis line produk yang tersedia? Berapa banyak jumlah line produk untuk setiap jenis produknya? Serta berapa total quantity jenis line produk nya?

Data Manipulation

1. Apakah terdapat anomaly berupa missing values, data duplicate, kesalahan data formatting, dan/atau inconsistency typing? Jika ada tunjukan serta lakukan penanganan pada anomaly tersebut.
2. Apakah terdapat tipe data yang berupa datetime pada data? Jika iya, apakah tipe data yang berupa datetime tersebut dapat dicari tahu selisihnya? Silahkan tampilkan hasilnya, berikan insight yang sesuai (jika memungkinkan).
3. Ada berapa banyak transaksi untuk setiap status barang yang di berhentikan produksinya? Lalu, bagaimana rata-rata dari UnitPrice untuk setiap status pengiriman tersebut? Jabarkan dan berikan insight.
4. Buatlah tabel yang melihat keadaan berapa banyak product yang di Discontinued oleh perusahaan beserta dengan kolom UnitsOnOrder dan ReorderLever. Berikan Insight tentang tabel tersebut!
5. Restock adalah sebuah variabel baru dimana merupakan hasil dari selisih antara UnitsInStock dan UnitsOnOrder. Apakah terdapat anomali data pada variabel tersebut? Berikan cara penanganan/insight.
6. Apakah terdapat outlier pada variable UnitsInStock? Jika ada, tunjukan di data keberapa.

Data Visualisation & Statistics

1. Apakah terdapat pelanggan yang loyal? Jika ada, bagaimana treatment dan strategi yang bisa dimanfaatkan dengan melihat adanya pelanggan setia tersebut?
2. Negara mana saja yang menjadi tujuan pemesan produk terbanyak? Tampilkan jabarkan dan buatlah analisisnya.
3. Bagaimana persentase status Level Reorder dengan Produksi Barang? Apakah strategi yang diambil oleh perusahaan sudah menempuh jalan yang benar?
4. Jenis product apa yang paling laku dan paling banyak terjual? Berikanlah analisisnya jika ada
5. Pada tahun berapa keuntungan terbesar perusahaan didapatkan?
6. Tampilkanlah Visualisasi Data Outliers dari variabel UnitsInStock.
7. Ujilah Perbedaan Quantity tiap Product pada dataset. Apakah data berdistribusi normal atau tidak?
8. Ujilah perbandingan jumlah quantity antar setiap product, serta berikan hipotesisnya!
9. Apakah terdapat hubungan antara 'UnitsOnOrder', 'UnitsInStock', 'UnitPrice', 'ReorderLever' dan Freight? Jika iya, variable mana saja yang paling mempengaruhi satu sama lain? Gambarkan bentuk hubungannya.