

## Capstone Project - Module 2

### Data Source :

Northwind Database

### General Question :

1. Bagaimana konteks bisnis berdasarkan dengan data yang telah diberikan? Silahkan buat asumsi atau penamaan jika diperlukan.
2. Jelaskan permasalahan apa yang ingin dianalisis dari data ini?
3. Ada berapa banyak tabel yang disediakan oleh database yang tersedia? Jabarkan setiap tabelnya. Jika terdapat ERD, silahkan ditampilkan

### SQL

1. Apakah tabel employees, employeeterritories, territoriesl, dan region dapat digabungkan menjadi 1 tabel? Jika memungkinkan, tampilkan tabel yang memuat informasi yang melibatkan keempat tabel tersebut.
2. Sebagai informasi tambahan untuk analisis, tampilkan tabel data transaksi perusahaan dengan customer, unit price nya, quantity, serta kategory produk apa saja yang dijual oleh perusahaan.

### Data Manipulation

1. Dari tabel yang ditampilkan, apakah perlu dibuatkan feature baru selain feature/kolom yang sudah ada untuk memudahkan analisis? Jelaskan kenapa feature baru tersebut perlu untuk ditambahkan.
2. Apakah terdapat anomaly berupa missing values, data duplicate, kesalahan data formatting, dan/atau inconsistency typing? Jika ada tunjukan serta lakukan penanganan pada anomaly tersebut.
3. Total price adalah sebuah variabel baru yang merupakan hasil perkalian antara unitprice dan quantity. Apakah terdapat outlier pada variable baru yang bernama totalprice tersebut? Jika ada, tunjukan di data keberapa. Lalu, bagaimana penanganan terhadap outlier ini? Coba jelaskan.

## **Data Visualization & Statistics**

1. Tampilkan karyawan dengan gaji tertinggi di perusahaan
2. Parameter apa saja yang dapat menjadi tolak ukur performa karyawan berdasarkan data yang sudah ada? Coba jelaskan.
3. Berdasarkan revenue nya, coba tampilkan revenue perusahaan terbesar berasal dari kategori produk apa, dan dari region mana saja. Dan kenapa hal tersebut bisa terjadi
4. Terdapat sekitar 9 variabel revenue perusahaan dari setiap karyawan yang ada. apakah terdapat perbedaan nilai antar setiap revenue yang dihasilkan dari setiap karyawan? Buktikan dengan uji hipotesis dan grafik yang mendukung
5. Pada data ini, apakah terdapat hubungan antara salary yang diberikan dengan parameter performa yang ditentukan? Gambarkan bentuk hubungannya dalam bentuk grafik
6. Kesimpulan apa yang dapat diambil dari analisa mengenai salary dari tiap karyawan di perusahaan ini berdasarkan data yang ada? Dan strategi apa yang bisa diterapkan untuk meningkatkan revenue perusahaan kedepannya?