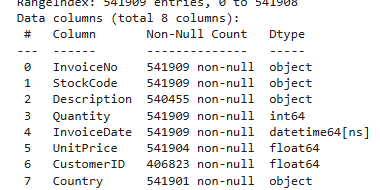
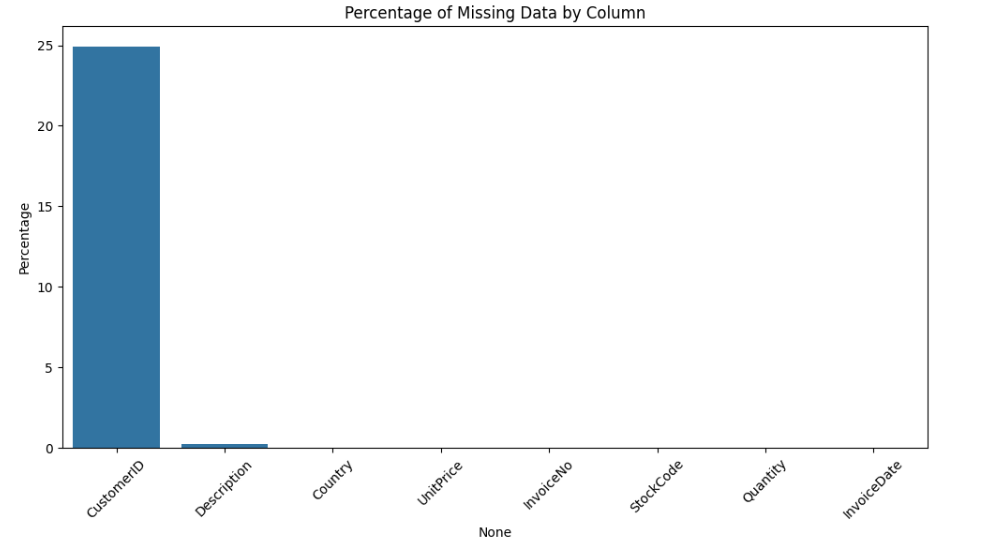
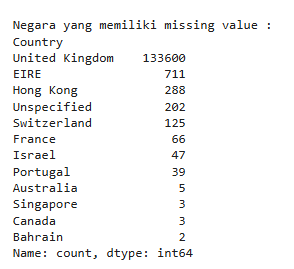
Jawaban UTS DATA MINING

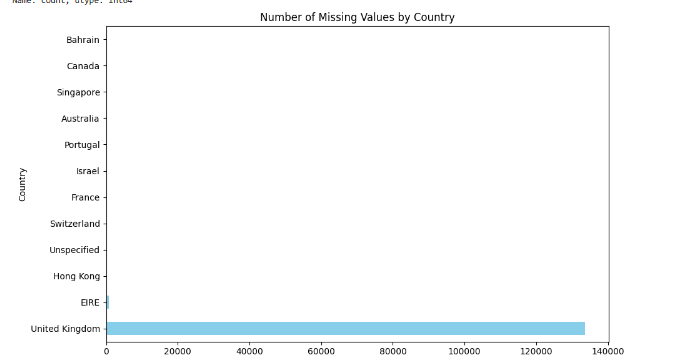
1. Jawab :
2. Kolom Apa Saja Yang Kosong





1. Negara Apa Saja Yang Kosong





1. Proses melakukan data cleaning :
2. Menganalisis Data yang Hilang:

* Persentase nilai yang hilang pada setiap kolom dihitung dan divisualisasikan untuk memahami sejauh mana masalah tersebut.
* Daftar negara dengan data yang hilang juga ditampilkan untuk analisis lebih lanjut.

1. Langkah-langkah Data Cleaning:
   * Untuk Kolom Numerik: Nilai yang hilang diganti dengan median dari kolom tersebut. Median dipilih karena lebih tahan terhadap nilai ekstrem (outlier) dibandingkan rata-rata (mean).
   * Contoh: Jika kolom UnitPrice memiliki nilai yang hilang, median harga dari semua entri yang tersedia di UnitPrice digunakan untuk mengisi kekosongan.
   * Untuk Kolom Kategorikal: Nilai yang hilang diganti dengan mode dari kolom tersebut, yang mewakili nilai yang paling sering muncul.
   * Contoh: Jika kolom Description memiliki nilai yang hilang, deskripsi produk yang paling umum digunakan untuk mengisi kekosongan.
   * Untuk Kolom CustomerID: Nilai yang hilang diganti berdasarkan mode dalam setiap negara. Pendekatan ini memastikan bahwa imputasi mencerminkan perilaku khas pelanggan di negara tersebut.
   * Jika tidak ada mode yang tersedia (misalnya, semua nilai dalam grup hilang), nilai default 0 digunakan.
2. Verifikasi:

* Setelah proses data cleaning selesai, semua kolom diperiksa untuk memastikan tidak ada nilai yang hilang tersisa.

1. Cara mengisi data yang kosong :
2. Kolom Numerik:

* Nilai median dari kolom dihitung berdasarkan data yang tersedia (tidak hilang) dan digunakan untuk mengisi entri yang hilang. Median lebih tahan terhadap nilai ekstrem atau distribusi yang miring.
* Contoh: Jika UnitPrice memiliki nilai yang hilang, nilai tersebut diisi dengan median harga, misalnya 2.5.

1. Kolom Kategorikal:

* Nilai mode (nilai yang paling sering muncul) dihitung dan digunakan untuk mengisi entri yang hilang. Mode membantu menjaga konsistensi dengan menggunakan kategori yang paling representatif.
* Contoh: Jika Description memiliki nilai yang hilang, dan deskripsi produk yang paling umum adalah White Hanging Heart T-Light Holder, maka nilai ini digunakan untuk mengisi kekosongan.

1. Kolom CustomerID:

* Nilai yang hilang pada kolom CustomerID diisi berdasarkan mode dari CustomerID dalam setiap negara.
* Ini memastikan bahwa nilai yang diimputasi relevan dengan konteks negara. Sebagai contoh, jika sebagian besar pesanan dari United Kingdom dilakukan oleh CustomerID 17850, maka nilai yang hilang untuk CustomerID di negara tersebut diisi dengan 17850.
* Jika semua nilai dalam grup negara hilang (tidak ada mode yang valid), nilai default 0 digunakan sebagai cadangan.