

Soal : Data Mining dan Business Intelligence  
Sifat : Online  
Waktu : 72 Jam

**Baca dengan seksama** soal berikut dan petunjuk pengerjaannya, dan pastikan anda download file dataset excel (Online Retail-unclean.xlsx) yang disediakan:

File Online Retail-unclean.xlsx berisi tentang transaksi barang retail oleh customer pada berbagai negara dan waktu. Di dalam file tersebut ada beberapa missing data, lakukan data cleaning pada missing data tersebut, dan jawab pertanyaan-pertanyaan dibawah:

1. Jawab pertanyaan berikut:
  - a. Pada kolom apa saja ada missing data?
  - b. Pada negara apa saja ada missing data?
2. Lakukan dan jawab:
  - a. Lakukan data cleaning pada missing data tersebut
  - b. Jelaskan bagaimana Anda melakukan data cleaning?
3. Jika Anda mengisi missing data dengan data baru, jelaskan bagaimana anda mendapatkan setiap value untuk data baru tersebut masing-masing!
4. Lakukan Agregasi untuk setiap negara dan total unit price setiap bulannya!
5. Jika Anda akan melakukan sampling pada data tiap negara, berapa jumlah masing-masing data sample di tiap negara jika  $\alpha = 0,03$  dan  $0,01$

**Petunjuk Pengerjaan:**

1. Jawaban no. 1 (a dan b), no. 2 (b), dan jawaban no. 3 ditulis pada file Word tersendiri, beri nama **Jawaban.docx**
2. Soal no 2 (a), setelah melakukan data cleaning, simpan hasilnya dalam file Excel baru dengan nama **Online-Retail-Clean.xlsx**
3. Soal no 4, buatlah sebuah file Excel baru dengan nama **Agregasi.xlsx** yang berisi table hasil dari agregasi yang dilakukan.
4. Soal no 5, buatlah sebuah file Excel baru dengan nama **Sampling.xlsx** yang berisi jumlah sample data tiap negara untuk masing-masing nilai  $\alpha$ .
5. Upload semua file word dan excel ke repository Github anda, buat repository baru khusus untuk UTS ini
6. Upload link repository github anda pada bagian UTS di Sinau

**Catatan:**

Anda boleh melakukan proses data cleaning, agregasi dan sampling menggunakan aplikasi python atau Jupyter Notebook, sertakan file-nya jika ada. Poin akan ditambahkan 70% lebih banyak dari hasil poin anda, dan poin dapat digunakan sebagai poin tugas.

**Peringatan:**

Segala bentuk plagiasi **tidak akan ditoleransi!** Jawaban harus dikerjakan sendiri. Jika ditemukan jawaban yang sama persis (copy+paste) pada semua soal dari mahasiswa lain maka nilai akan langsung diberikan D