Dimension of Data Quality

1. Accuracy

* Do data objects accurately represent the “real world” values they are expected to model?

[SEWAATM]

* Setelah dilakukan exploring data, memang data penyewaan ATM ini ada di lokasi-lokasi seperti di swalayan (indomaret, alfamart, dsb), mall, SPBU, dan universitas.
* Isi datanya pun terdapat informasi status seperti cancel request. Meskipun digabungkan ke dalam kolom LOKASI dan KOTA
* Karena ini merupakan data sewa ATM, maka akurat untuk ada tanggal request, tanggal delivery, dan tanggal jatuh tempo untuk pelunasan, serta tagihan-tagihan.

[Data Aktifitas]

* Karena ini merupakan data untuk aktifitas perlakuan barang, informasi seperti lokasi, tanggal, nomor id, jenis layanan, laporan kerusakan, jenis sparepart, dsb memang diperlukan
* Are there incorrect spellings of product or person names, addresses, and even untimely or not current data?

[SEWAATM]

* Ada, tetapi tidak terlalu significant. Contohnya alfa mart dengan alfamart atau univ dengan universitas
* Penginputan data lebih konsisten dibanding sheet SEWAATM

1. Completeness

* Is all the requisite information available? Do any data values have missing elements? Or are they in an unusable state?

[SEWAATM]

* Meskipun kolom yang dibutuhkan lengkap, akan tetapi tidak semua kolom memiliki value yg lengkap, bahkan ada kolom yang sama sekali tidak memiliki value
* Masing-masing kolom Tagihan memiliki subkolom JT48 yg semuanya tidak memiliki value
* Tidak ada kolom status yg menunjukkan status dari pemesanan

[Data Aktifitas]

* Banyak missing value
* Bisa ditambahkan kolom tanggal selesai

1. Consistency

* Are data values the same across the data sets?
* Secara keseluruhan, value data konsisten pada kolom satu dengan yg lain
* Untuk value-nya, tanggal delivery dan jatuh tempo konsisten 3 tahun, tetapi ada yg 3 tahun 1 bulan, 3 tahun 2 bulan, atau 3 tahun 3 bulan (ini tergantung policy perusahaan?)

1. Timeliness

* Is the information available when needed?

[SEWA ATM] dan [Data Aktifitas]

* Ketika data di-store ke dalam bentuk Excel, data tersebut hanya bisa di-update oleh satu orang tiap satu waktu. Akan tetapi, apabila data di-store ke dalam database yg ada di server, data tersebut dapat diakses oleh banyak orang dalam satu waktu, sehingga data selalu tersedia. Atau dengan google spreadsheet
* Data yg sudah saya lakukan cleaning sudah dipisahkan untuk tiap komponen tanggalnya. Jadi ketika suatu saat dibutuhkan untuk review atau laporan tahunan/bulanan dapat langsung diakses.
* Data berupa ekstensi excel biasanya akan berantakan ketika dicoba dibaca oleh pandas, terutama apabila pada file tersebut terdapat custom judul atau multiple-cell header, sehingga membutuhkan waktu lebih untuk merapihkan format awal.

1. Validity

* Is information in a specific format, does it follow business rules, or is it in an unusable format?

[SEWAATM]

* Untuk format, banyak data yg tidak konsisten valuenya:
* Pertanggalan tidak disamakan formatnya. Ada yg memakai nama bulan dalam bahasa Indonesia, ada yg 3 huruf bahasa Indonesia (cth Agt), ada yg 3 huruf bahasa inggris (cth Aug)
* Pemisahan tahun ada yg memakai dash (-) atau tanda petik (‘)
* Penggunaan huruf besar dan huruf kecil yg tidak konsisten
* Di dalam kolom LOKASI dan KOTA terdapat status dari pemesanan dan notes2 kecil dari penginput data.
* Di dalam kolom LOKASI dan KOTA terdapat juga kata2 yg di-bold.

[Data Aktifitas]

* Sheet ini lebih rapi dan konsisten untuk penginputan kategori

1. Uniqueness

* Is this the only instance in which this information appears in the database?

[SEWAATM]

* Instance di khususnya kolom ‘nama LOKASI’ akan menampilkan bahwa hampir tidak ada duplikasi value karena adanya tambahan string status seperti cancel, off, rpc, dan sebagainya. Misal: indomaret monas dan indomaret monas cancel. Pandas akan membaca kedua instance ini sebagai dua instance yg unik.

[Data Aktifitas]

* Kolom Nomor memiliki value yg berbeda sebagai id. Jadi tiap transaksi adalah unique

[SEWAATM]

(+) Accuracy, Consistency

(-) Completeness, Timeliness, Validity, Uniqueness

[Data Aktifitas]

(+) Accuracy, Consistency, Validity, Uniqueness

(-) Completeness, Timeliness