

Nama: Jafar Shodiq

Study Case:

MigWare adalah salah satu layanan hardware manage service yang dimiliki oleh MIG. Pada aktivitas hardware managed service terdapat dua aktivitas utama yaitu sewa dan juga perawatan. Sebelum memasukan data tersebut ke data warehouse perlu dilakukan pre- processing sehingga nantinya data dapat diolah dan nantinya dapat digunakan untuk mendapatkan insight, dan juga menghasilkan report yang dapat meningkatkan layanan.

1. Bagaimana data quality dari data excel yang diberikan dan bagaimana meningkatkan data quality dari data tersebut.
2. Lakukan data cleansing dari data tersebut, jelaskan tahapan tahapannya. (dapat melampirkan file jupyter notebook, google colab, atau sejenisnya)
3. Insight apa yang temukan dari data tersebut.

Format Jawaban:

Pertanyaan

Sheet: (nama sheet excel)

- (jawaban)

Jawaban:

Bagaimana data quality dari data excel yang diberikan dan bagaimana meningkatkan data quality dari data tersebut?

Sheet: SEWAATM

Tanggapan:

- Dari 65780 baris, hanya 1247 baris yang memiliki nilai
- Banyak missing value atau baris yang tidak berisi nilai (NaN)
- Kolom Tagihan tidak konsisten terutama untuk format tanggal/periode. Sebagian menggunakan format YYYY-mm-dd, sebagian menggunakan MM'yy. Sehingga cukup sulit dan akan memakan banyak waktu untuk dapat membersihkannya.
- Pada kolom Tagihan juga terdapat subkolom JT48 yang semua value-nya kosong.
- Pada kolom Nama LOKASI dan KOTA terdapat inkonsistensi untuk penggunaan huruf besar atau huruf kecil (Misal. Jkt dengan JKT).
- Pada kolom Nama LOKASI, terdapat banyak comment atau notes-notes kecil.

- Status seperti cancel, RPC, atau off diletakkan begitu saja di cell yang sama dengan kolom KOTA.

Solusi:

- Untuk pre-test ini, saya tidak menggunakan semua kolom Tagihan dikarenakan inkonsistensi pada kolom-kolom ini. Untuk benar-benar bisa melakukan cleaning pada kolom-kolom ini, saya membutuhkan waktu yang lebih.
- Semua tanggal pada kolom Tanggal Request, Delivery, dan Jatuh Tempo, saya pisahkan untuk tanggal, bulan, dan tahun ke kolom baru yang berbeda.
- Karena tidak konsistennya penggunaan huruf besar dan huruf kecil, saya mengubah semua kata menjadi huruf kecil.
- Nama kolom juga saya ubah dengan alasan yang sama.
- Saya menambahkan kolom baru dengan nama 'status' yang memiliki value cancel, RPC, off, atau yang lainnya yang menunjukkan status yang sebelumnya ada pada kolom KOTA.

Sheet: Data Aktivitas

Tanggapan:

- Sheet ini jauh lebih tersusun dan rapi dibanding SEWAATM
- Banyak value yang berupa space kosong, bukan NaN
- Kolom Status dan Pilihan tidak berisi value

Solusi:

- Value yang berupa space kosong, saya ganti dengan karakter '-' atau '0'
- Tidak mengikutkan kolom Status dan Pilihan

Lakukan data cleansing dari data tersebut, jelaskan tahapan tahapannya. (dapat melampirkan file jupyter notebook, google colab, atau sejenisnya)

Sheet: SEWAATM

- (terlampir)

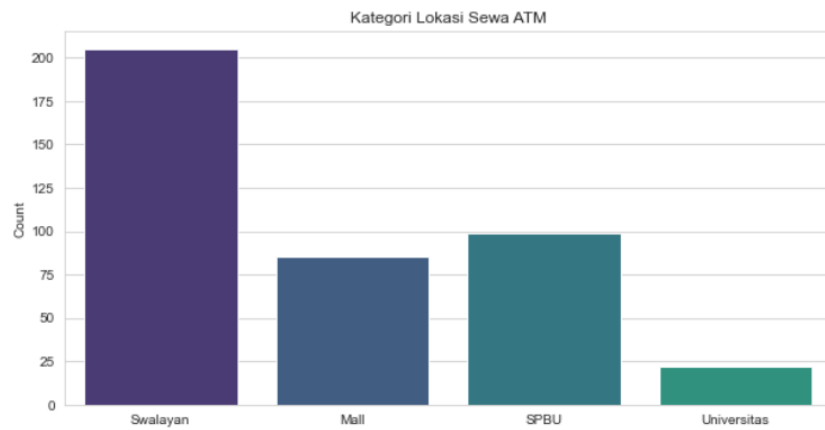
Sheet: Daftar Aktivitas

- (terlampir)

Insight apa yang temukan dari data tersebut?

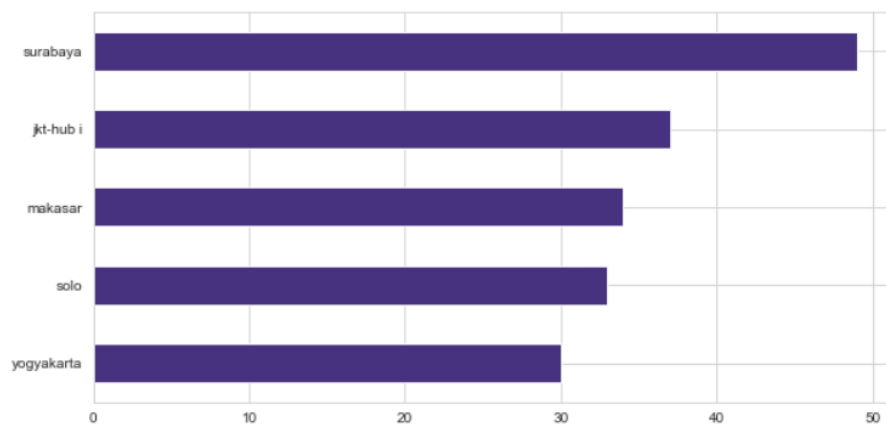
Sheet: SEWAATM

Swalayan adalah penyewa ATM terbanyak.



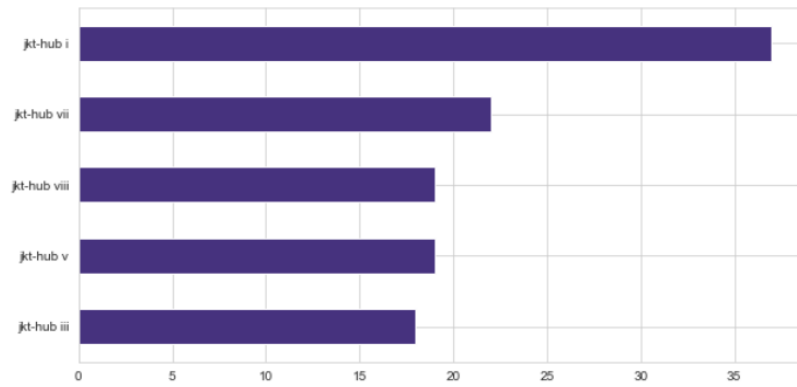
	Cat	Count
0	Swalayan	205
1	Mall	85
2	SPBU	99
3	Universitas	22

Surabaya adalah penyewa ATM terbanyak.



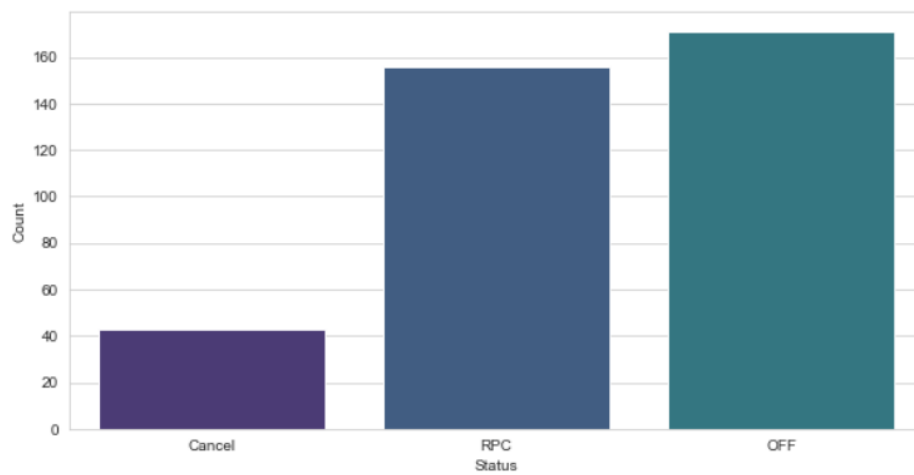
surabaya	49
jkt-hub i	37
makasar	34
solo	33
yogyakarta	30

Di Jakarta, jkt-hub I adalah penyewa terbanyak.



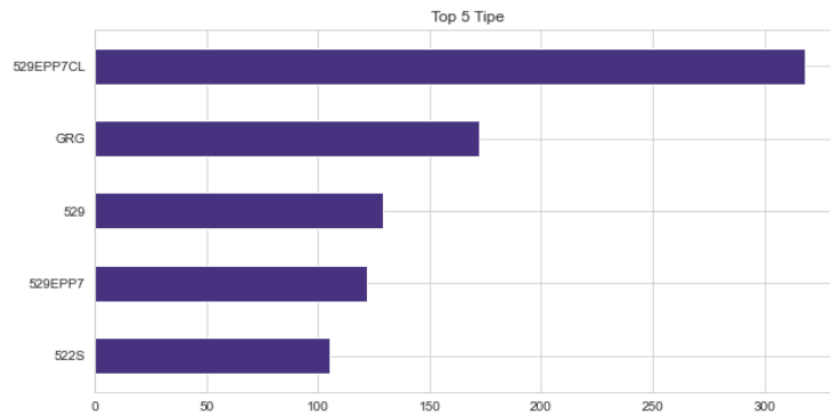
jkt-hub i	37
jkt-hub vii	22
jkt-hub v	19
jkt-hub viii	19
jkt-hub iii	18

Status total



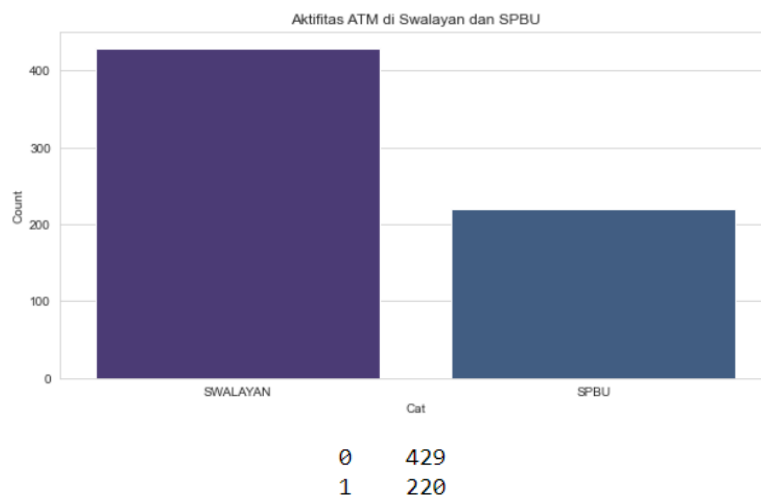
Status	Count
0 Cancel	43
1 RPC	156
2 OFF	171

Top 5 tipe

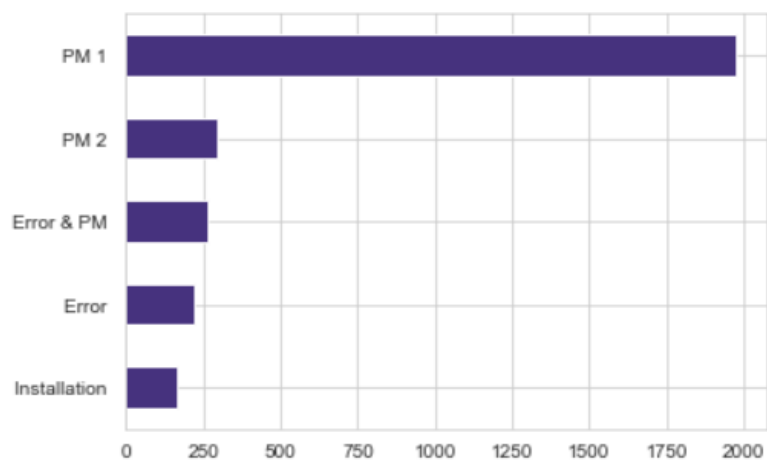


Sheet: Daftar Aktifitas

Aktifitas di swalayan lebih banyak dibanding SPBU



Jenis layanan terbanyak yaitu PM1



Laporan kerusakan terbanyak yaitu preventive maintenance

