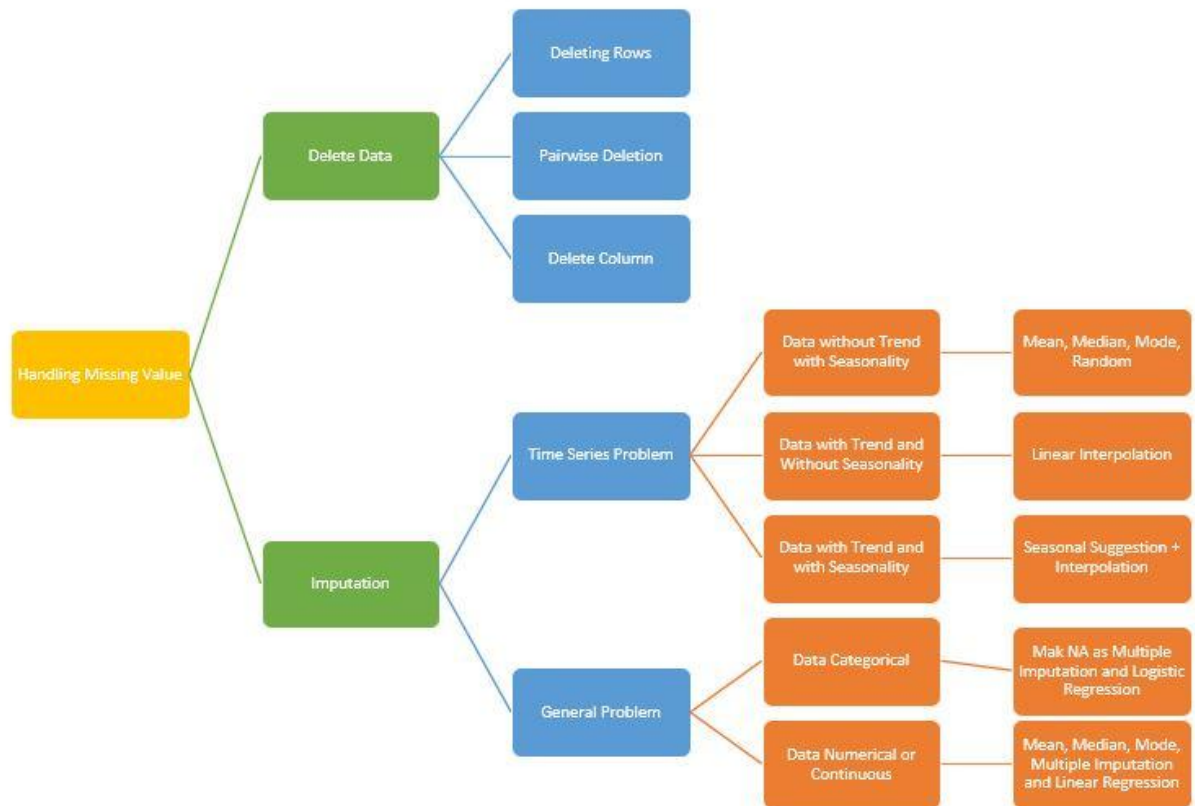


Solusi mengisi Missing Value

Sebelum mengenal lebih banyak mengenai Solusi yang biasa ada pada kasus-kasus data science.

Skema yang biasanya sering dalam pengelolaan data.



Dalam diagram diatas, perlu diketahui bahwa kasus kehilangan data bisa diatasi dengan berbagai cara. Bahkan, melakukan penghapusan data juga merupakan solusi yang bisa menjadi pilihan apabila jika dirasa mengisi nilai kosong akan memberikan pengaruh yang kurang bagus terhadap analisa, atau apabila pertimbangan data yang dihapus atau data yang hilang sedikit dan tidak memberikan terlalu banyak sumbangsih untuk analisa yang akan dilakukan. Penghapusan data bisa langsung pada baris data tersebut atau langsung satu kolom data. Pada solusi kedua yaitu menggunakan imputation (pengisian data yang kosong) bisa tergantung dari permasalahannya. Khusus untuk masalah yang berhubungan forecasting atau peramalan tergantung dari data yang ada (lebih lengkap bisa dilihat pada gambar). Khusus untuk general problem tergantung jenis datanya. Jika yang hilang data kategorikal atau bersifat string bisa menggunakan relasi antar kolom dengan Logistic Regression, jika numerical bisa menggunakan statistik sederhana dan linear regression. Pada sesi kali ini kita akan mencoba menangani data hilang dengan statistik sederhana, Mean dan Median.