

## LATIHAN CLUSTERING

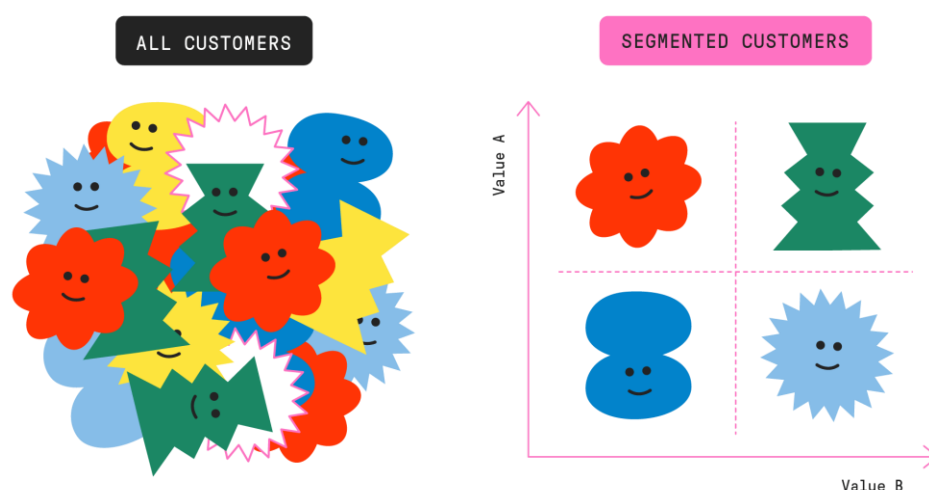
### WORKSHOP R 2020

#### “Retail Marketing”

#### Business Understanding

PT X adalah sebuah perusahaan retail yang saat ini berusaha untuk menjadi salah satu perusahaan retail terbesar di negara Y. Penjualan yang dilakukan oleh PT X cenderung bagus, namun terdapat tidak efektifnya pengeluaran pada divisi pemasaran. Didapati bahwa hasil penjualan yang dihasilkan dari usaha pemasaran tidak sebanding.

Mengetahui hal tersebut, divisi pemasaran PT X ingin memaksimalkan jumlah penjualannya melalui membuat kampanye-kampanye pemasaran yang *personalized*. Divisi pemasaran PT X kemudian meminta data historis pembeli di PT X untuk dianalisis lebih lanjut agar bisa membuat kampanye pemasaran yang *personalized*. Menurut divisi pemasaran PT X, kampanye pemasaran dapat dibuat *personalized* dengan melakukan segmentasi konsumen dari data historis pembeli. Dengan begitu, kampanye pemasaran yang dilakukan dapat lebih terarah.



Sumber: intercom.com

#### Data Understanding

Dataset berupa data historis pembeli yang digunakan oleh divisi pemasaran PT X berisi data-data demografis pembeli dan data transaksi pembeli. Data demografis berupa umur, jenis kelamin, status, tempat tinggal (jauh atau tidaknya dan mempunyai rumah sendiri atau menyewa), jumlah anak, dan gaji. Data transaksional yang digunakan adalah data *history*, katalog produk yang dibeli, dan jumlah biaya transaksi yang dilakukan.

Proses segmentasi konsumen tidak dapat langsung digunakan begitu saja karena beberapa jenis data masih bersifat kategorikal. Untuk mempermudah proses segmentasi konsumen, data kategorikal dapat diubah menjadi data numerical, sehingga pengelompokkan konsumen dapat lebih mudah dan cepat dilakukan.

### Your Job

Sebagai seorang *data analyst*, Anda diminta bantuan oleh PT. X membuat *cluster* konsumen berdasarkan seluruh variabel yang ada pada dataset. Gunakanlah metode *clustering* secara khusus algoritma K-means untuk proses *clustering* yang dilakukan.