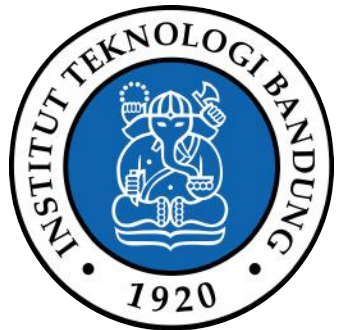
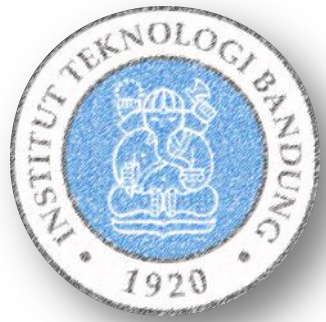


# Latihan Soal (Tool: Python)

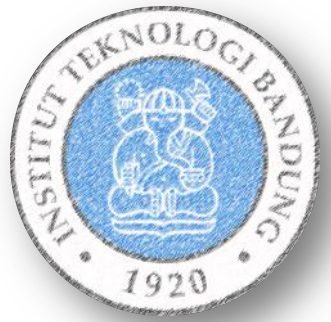
Tim Penyusun Materi Pengenalan Komputasi  
Institut Teknologi Bandung © 2019





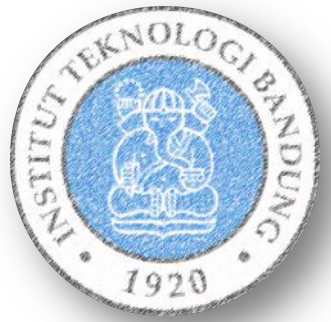
# Sebelum dimulai

- Siapkan notebook yang telah terinstall Python Pandas
- Download file financial.csv
  - Data berisi penjualan dari sebuah perusahaan
- Untuk setiap latihan berikut, tuliskan:
  - Hasilnya (termasuk tampilan visualisasinya)
  - Program/script untuk mendapatkannya.



# Financial Dataset

- Attributes:
  - Segment
  - Country
  - Product
  - DiscountBand
  - UnitsSold
  - ManufacturingPrice
  - GrossSales
  - Discount
  - Sales
  - COGS (Cost of goods sold)
  - Profit
  - Date
  - MonthNumber
  - MonthName
  - Year

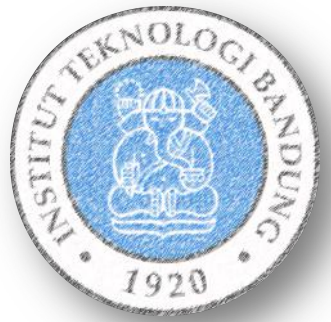


# Impor Data

- Load data financial.csv
- Simpan sebagai sebuah dataframe misalnya bernama **dffin**

# Data Understanding

- Ada berapa banyak baris data sales?
- Ada berapa kolom dan apa saja namanya?
- Apa tipe data tiap kolom: numerik? teks? kuantitatif? ordinal?
- Perhatikan data pada tiap kolom
  - Nilai-nilai apa saja yang terdefinisi untuk atribut bertipe ordinal
  - Range untuk tiap atribut numerik
- Apakah ada data yang “kotor”?
  - Data kosong?
  - Data berisi nilai yang “aneh”?
  - Apa yang perlu dilakukan terhadap data kotor? Diberikan nilai? Dibiarkan?



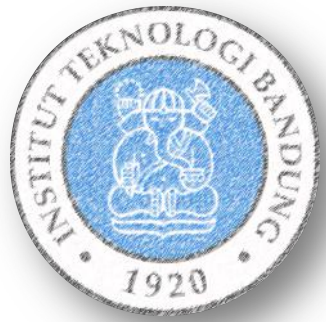
# Statistics

- Amati dan jelaskan statistik untuk data tiap atribut numerik:
  - Nilai ekstrim: minimum, maksimum
  - Nilai rata-rata (mean) dan deviasi standar
  - Percentile 25%, 50%, 75%



# Data exploration

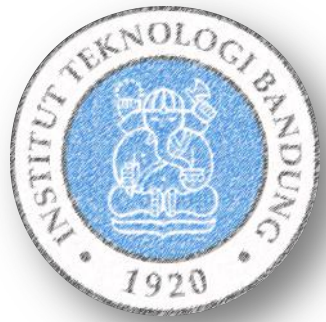
- Tampilkan data pada MonthName December Year 2014
- Tampilkan data untuk Product Paseo dengan UnitsSold > 1000
- Tampilkan 10 data dengan nilai kolom Sales tertinggi
- Tampilkan data Country dengan total Profit
- Tampilkan frekuensi data per Segment pada tahun 2013
- Tampilkan data terurut berdasarkan Year dan MonthNumber
- Tampilkan data terurut berdasarkan Product



# Visualisasi (1)

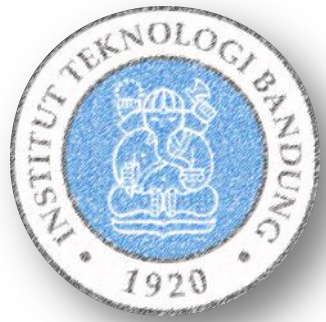
- Untuk tiap hal berikut:
  - pilihlah satu saja jenis grafik yang tepat untuk menggambarkan fenomenanya,
  - jelaskan mengapa Anda memilih grafik tersebut,
  - buatlah grafiknya
- Menampilkan distribusi data total UnitsSold untuk tiap bulan (month) pada tahun (Year) 2014.
- Membandingkan total Sales dan total Profit untuk tiap jenis Product pada tahun (Year) 2014.





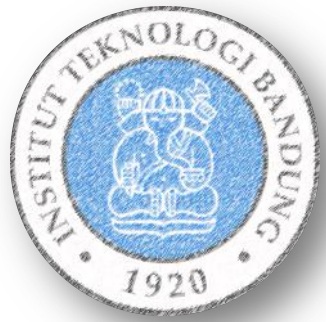
## Visualisasi (2)

- Membandingkan rata-rata ManufacturingPrice dan SalePrice untuk penjualan dari negara (Country) United States.
- Menampilkan komposisi (persentase) berapa total UnitsSold untuk tiap jenis Product.
- Menampilkan komposisi (persentase) total UnitsSold per Country.
- Menampilkan total perkembangan jumlah Sales dari bulan ke bulan selama tahun (Year) 2013 untuk negara (Country)



# Hubungan antara Discounts dan Profit

- Perusahaan ingin mengetahui hubungan antara data Discounts dan Profit.
  - Buatlah grafik *scatter plot* yang menunjukkan hubungan antara kolom Discounts (sebagai sumbu X) dan Profit (sebagai sumbu Y).
  - Hitunglah bagaimana nilai korelasi antara kedua atribut.
  - Berdasarkan grafik dan nilai korelasi tersebut, sebutkan dan jelaskan bagaimana korelasi antara kedua kolom tersebut.



# 3 product paling menguntungkan per Segment

- Perusahaan ingin mengetahui 3 (tiga) Product yang memberikan total Profit paling tinggi untuk masing-masing Segment
- Berikan usulan teknik analisis data yang dapat digunakan untuk menghasilkan pengetahuan tersebut.