



# **Session 23**

## **Intermediate Statistics**



# Table of Content

## What will We Learn Today?

1. Correlation and Causality
2. Probability and Distribution
3. Statistical Plot





# Correlation vs Causality





# Correlation

- Korelasi (*Correlation*) adalah ukuran statistik (dinyatakan sebagai angka) yang menggambarkan ukuran dan arah hubungan antara dua variabel atau lebih.
- Namun, korelasi antar variabel tidak secara otomatis berarti bahwa perubahan dalam satu variabel adalah penyebab perubahan nilai variabel lainnya.
- Misalnya, untuk dua variabel "jam kerja" dan "pendapatan yang diperoleh" .







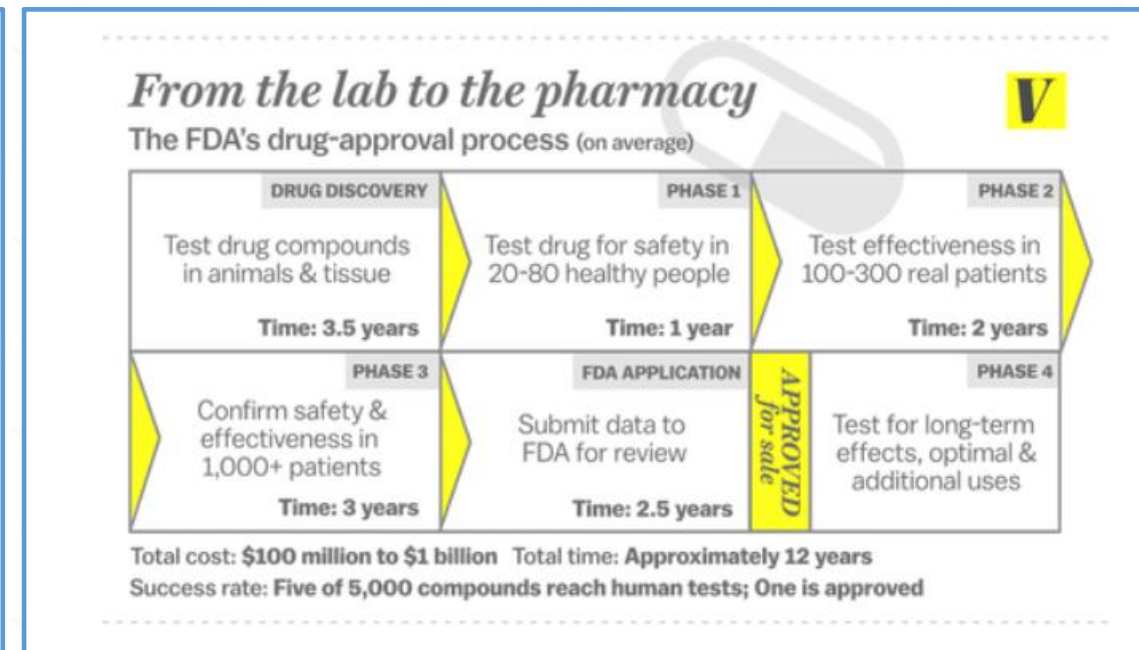
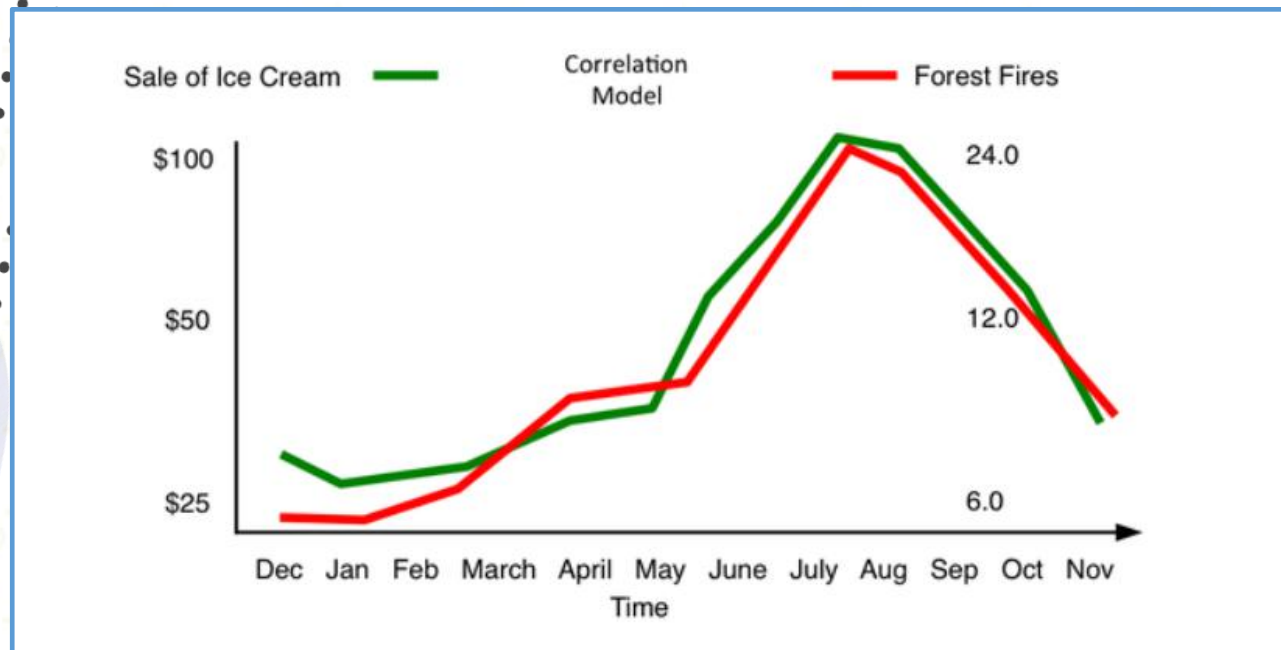
# Causality

- Kausalitas (*Causality*) menunjukkan bahwa satu peristiwa adalah hasil dari terjadinya peristiwa lainnya.
- Yaitu ada hubungan sebab akibat antara dua peristiwa.
- Kausalitas juga disebut sebagai sebab dan akibat.
- Misalnya merokok menyebabkan peningkatan risiko kanker paru-paru.



# Korelasi vs Kausalitas

- Korelasi != Kausalitas



<https://www.decisionskills.com/blog/how-ice-cream-kills-understanding-cause-and-effect>



# Mengapa korelasi dan kausalitas itu penting?

- Untuk mengidentifikasi sejauh mana satu variabel terkait dengan variabel lain.
- Sebagai contoh:
  - Apakah ada hubungan antara tingkat pendidikan seseorang dan kesehatannya?
  - Apakah memelihara hewan peliharaan menyebabkan hidup lebih lama?
  - Apakah marketing campaign sebuah perusahaan mampu meningkatkan penjualan produk mereka?
- Jika ada korelasi → panduan penelitian lebih lanjut untuk menyelidiki apakah ada kasualitas.
- Dengan memahami korelasi dan kausalitas, memungkinkan kebijakan dan program yang bertujuan untuk mencapai hasil yang diinginkan agar lebih tepat sasaran.





# Bagaimana menghitung korelasi

- Pearson correlation coefficient ( $r$ ) = mengukur kekuatan dan arah hubungan linier antara dua variabel.
- Nilai numerik koefisien berkisar dari -1.0 hingga +1.0.
- Jika nilai koefisien = negatif, menunjukkan hubungan negatif antar variable
- Jika nilai koefisien = positif, menunjukkan hubungan positif antar variable

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

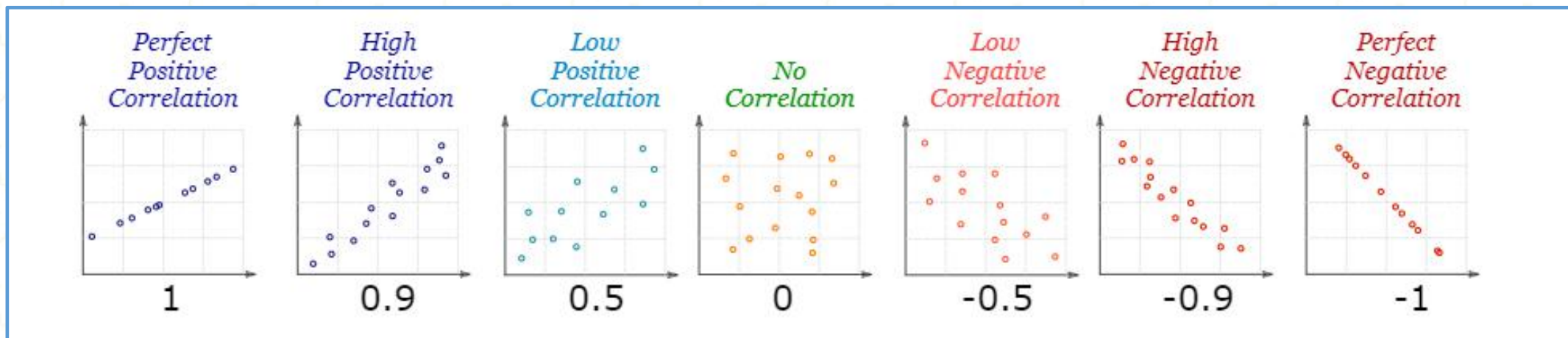






# Bagaimana menghitung korelasi

- Pearson correlation coefficient ( $r$ ) = mengukur kekuatan dan arah hubungan linier antara dua variabel.
- Korelasi dapat memiliki nilai:
  - 1 adalah korelasi positif sempurna
  - 0 tidak ada korelasi (tidak ada hubungan)
  - -1 adalah korelasi negatif sempurna





# Bagaimana mengukur kausalitas?

- Penggunaan percobaan (experimental studies) adalah cara paling efektif untuk melihat hubungan sebab akibat antar variabel.
  - Dalam percobaan, sampel atau populasi dibagi menjadi dua secara random, dengan kedua kelompok dapat dibandingkan dengan kondisi yang sama.
  - Kedua kelompok kemudian menerima perlakuan (treatment) yang berbeda, dan hasil dari masing-masing kelompok dinilai.
  - Kita dapat menyimpulkan bahwa treatment (misalnya pengobatan) mempunyai efek tertentu jika kelompok memiliki hasil yang sangat berbeda.





# Probability and Distribution







# Probability

- Probabilitas (probability) adalah deskripsi numerik tentang seberapa besar kemungkinan suatu peristiwa terjadi.
- Biasanya dilambangkan dengan P (event)
- $P(A)$ : Probabilitas kejadian A
- $P(A) = \text{kemungkinan kejadian A} / \text{semua kemungkinan}$







# Probability

A ♣	2 ♣	3 ♣	4 ♣	5 ♣	6 ♣	7 ♣	8 ♣	9 ♣	10 ♣	J ♣	Q ♣	K ♣
A ♥	2 ♥	3 ♥	4 ♥	5 ♥	6 ♥	7 ♥	8 ♥	9 ♥	10 ♥	J ♥	Q ♥	K ♥
A ♠	2 ♠	3 ♠	4 ♠	5 ♠	6 ♠	7 ♠	8 ♠	9 ♠	10 ♠	J ♠	Q ♠	K ♠
A ♦	2 ♦	3 ♦	4 ♦	5 ♦	6 ♦	7 ♦	8 ♦	9 ♦	10 ♦	J ♦	Q ♦	K ♦

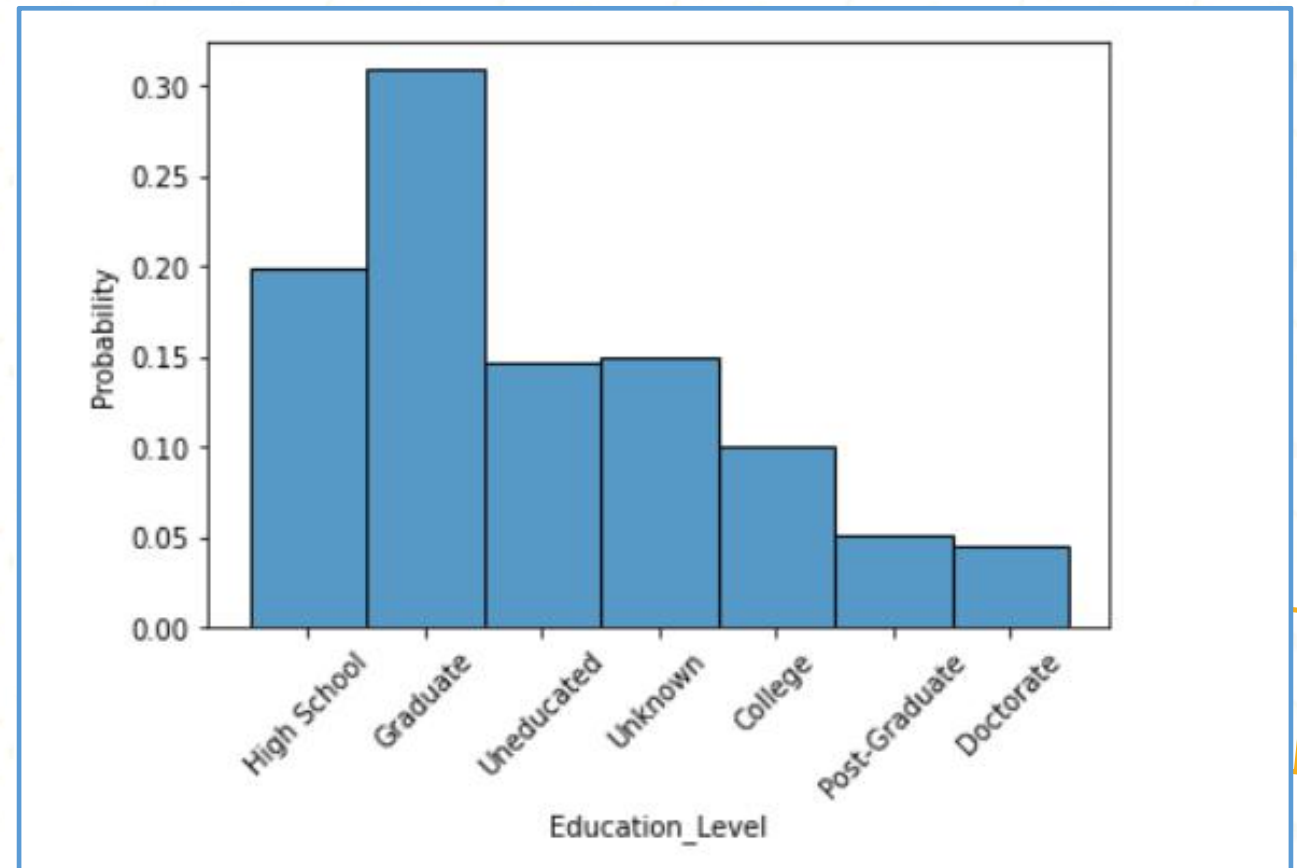
- $P(\text{King}) = 4/52$
- $P(\text{Red}) = 26/52$
- $P(\text{Queen or 3}) = 8/52$





# Probability mass function (pmf)

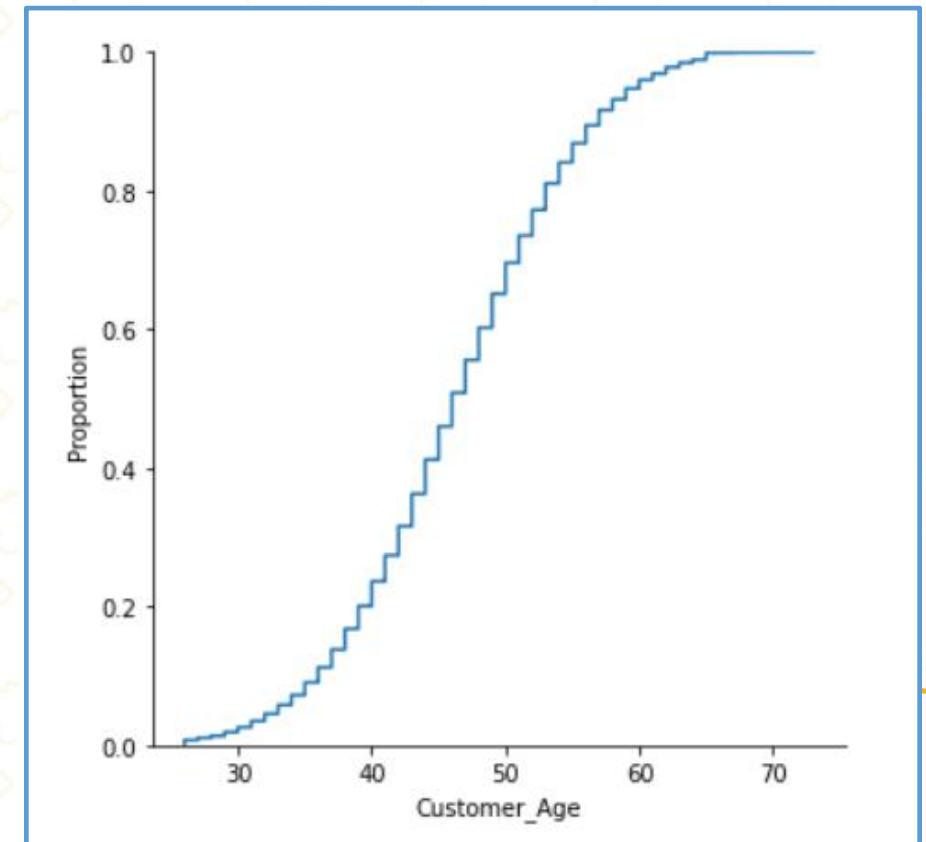
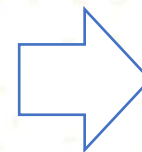
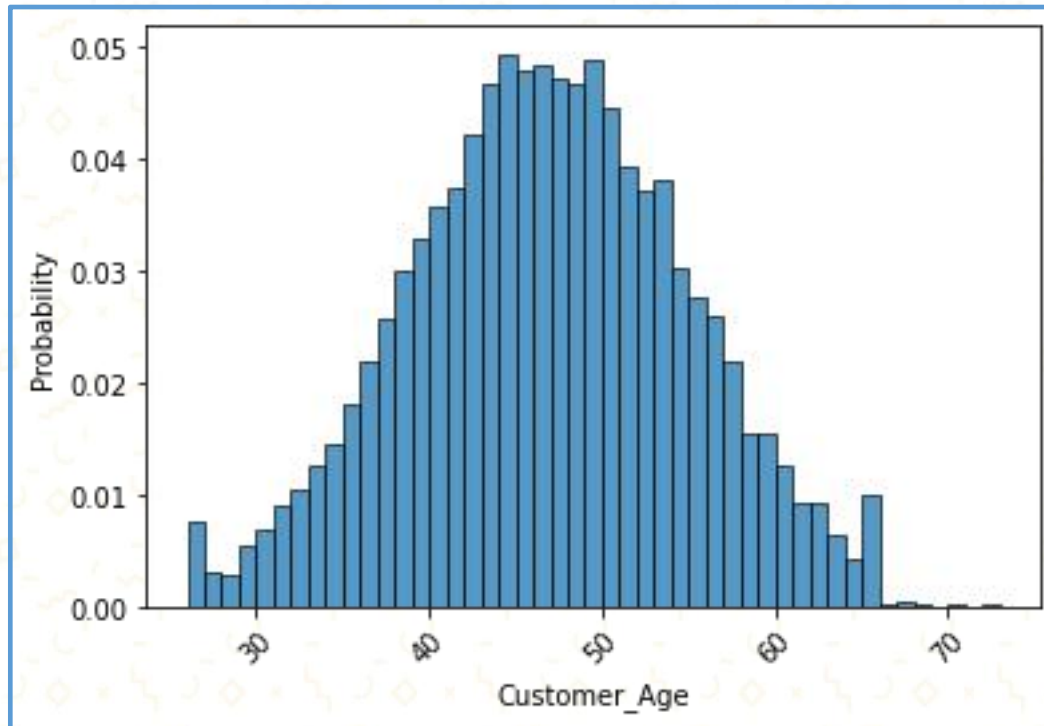
- Fungsi matematika yang memberikan kemungkinan dari variabel acak untuk memiliki suatu nilai.
- Hanya digunakan untuk variabel discrete atau categorical





# Cummulative density function (cdf)

- Fungsi yang menjumlahkan nilai kemungkinan sampai suatu kejadian tertentu.

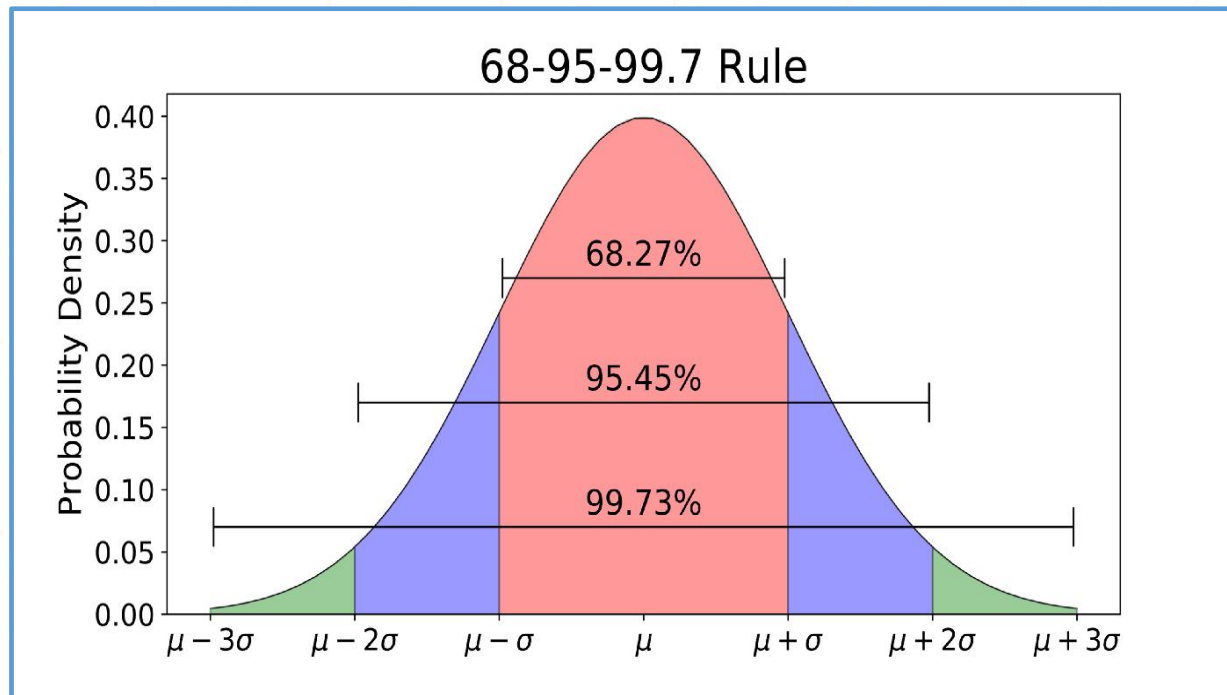






# Normal Distribution

- Distribusi normal (normal distribution) adalah distribusi probabilitas kontinu (continuous probability distribution) yang dicirikan oleh kurva berbentuk lonceng simetris.

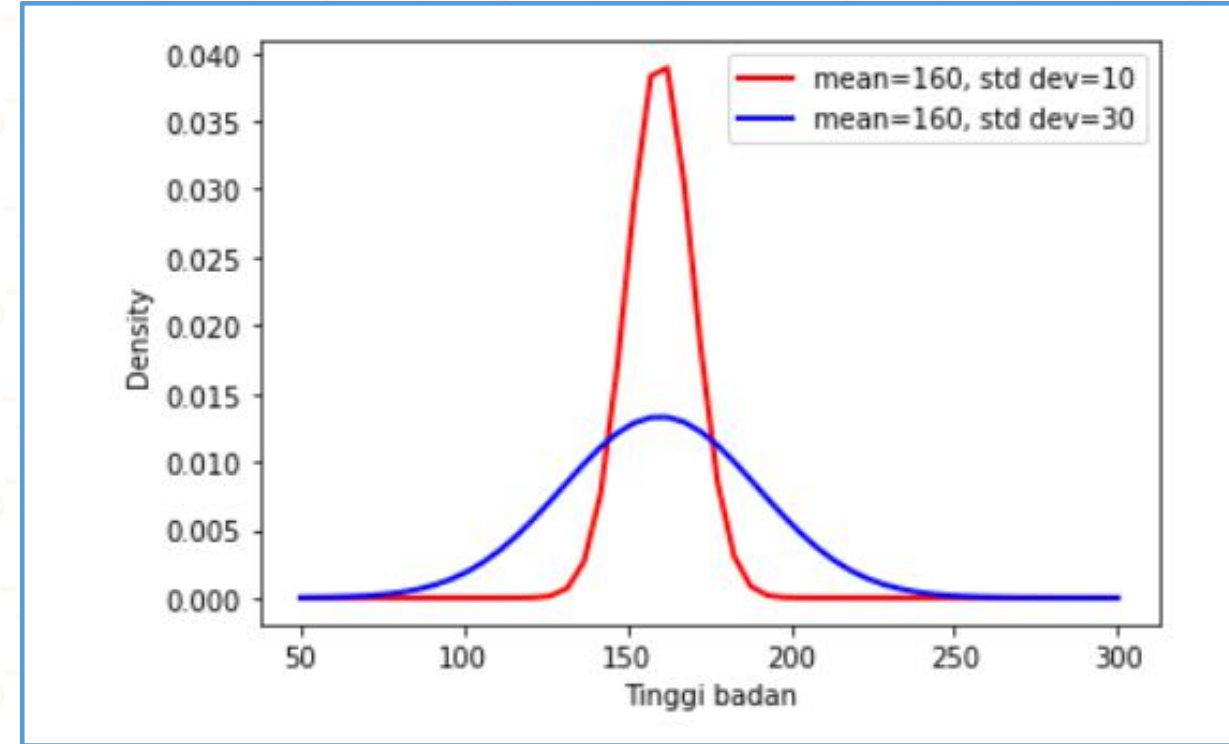
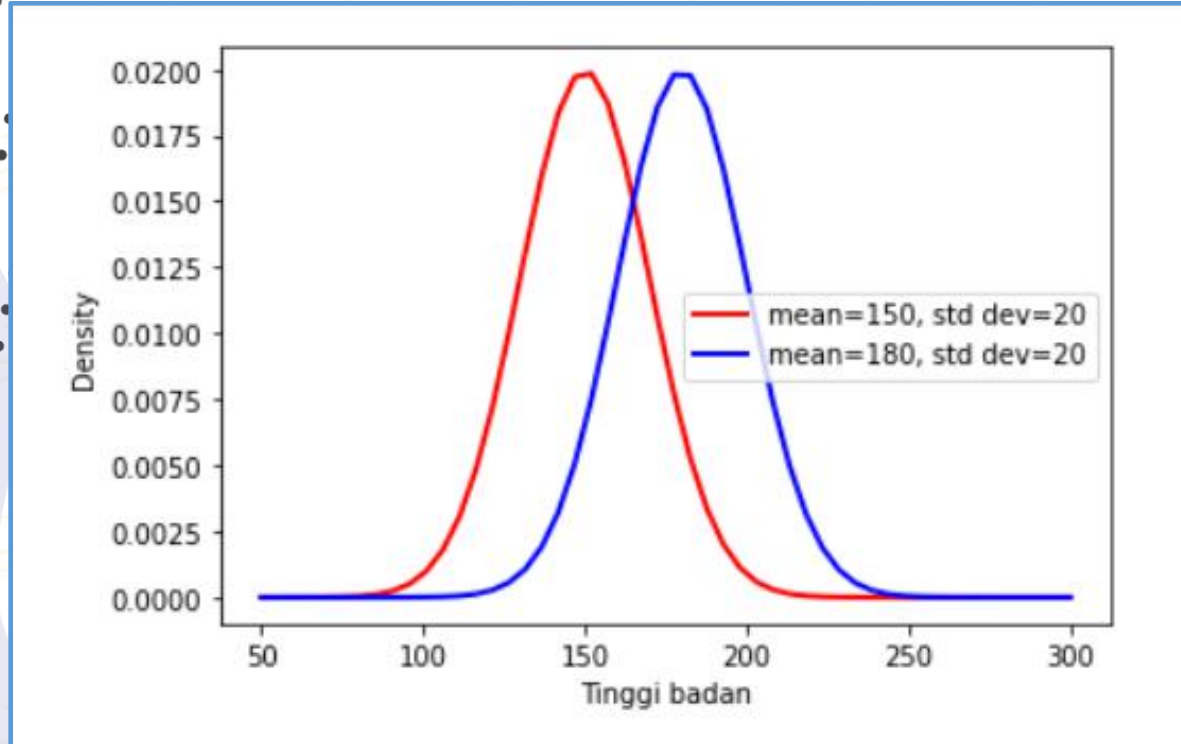






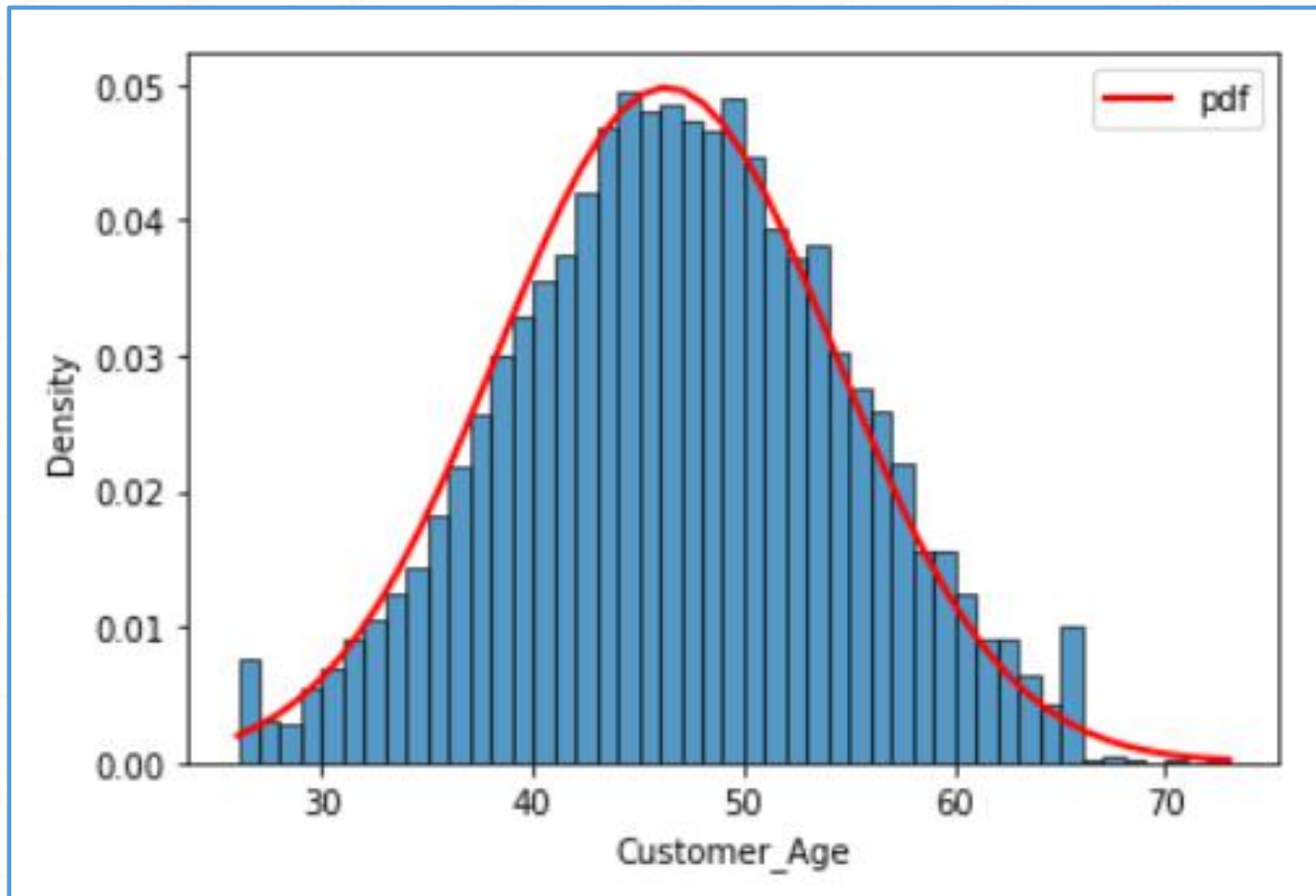
# Normal Distribution

- Selain itu, distribusi normal ditentukan oleh pusatnya (mean) dan penyebarannya (standar deviasi)





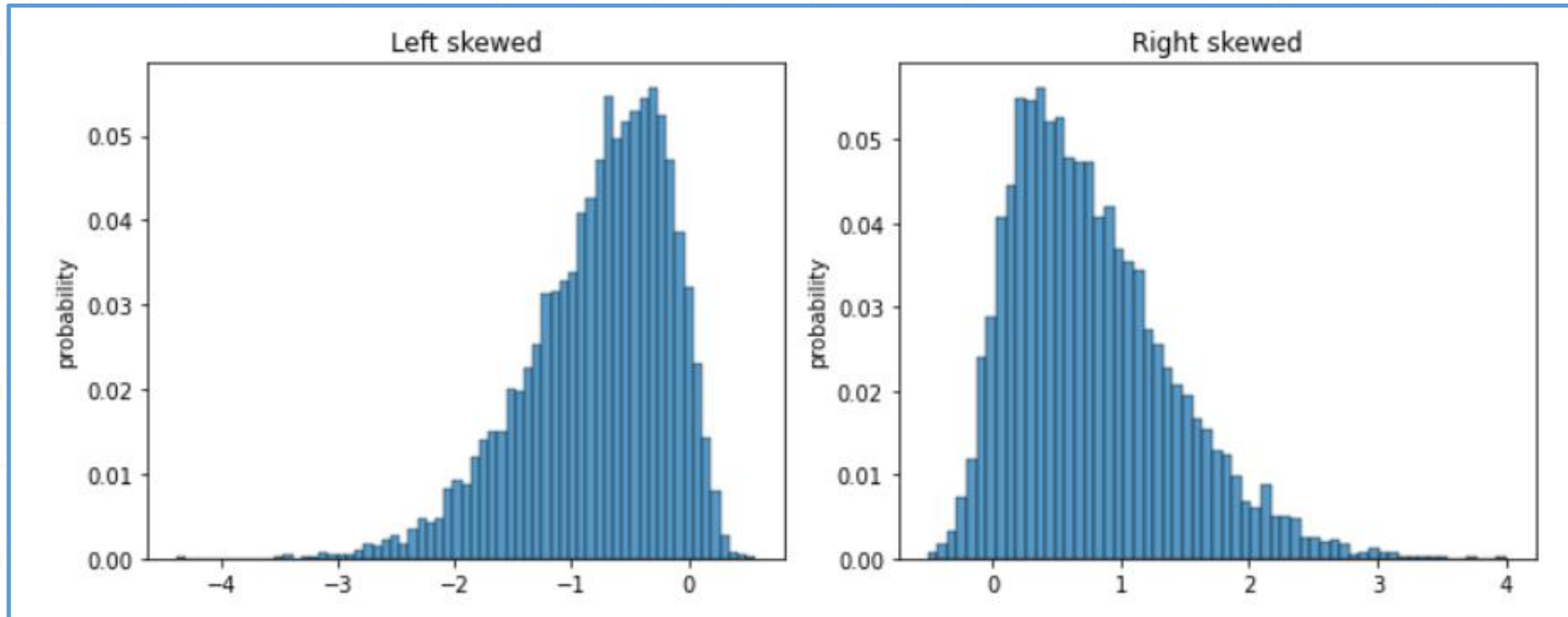
# Normal Distribution





# Skewness

- Skewness adalah derajat asimetri dari suatu distribusi.
  - Right Skewness: ketika nilai memanjang jauh ke kanan
  - Left Skewness: ketika nilai memanjang jauh ke kiri





# Statistical Plot

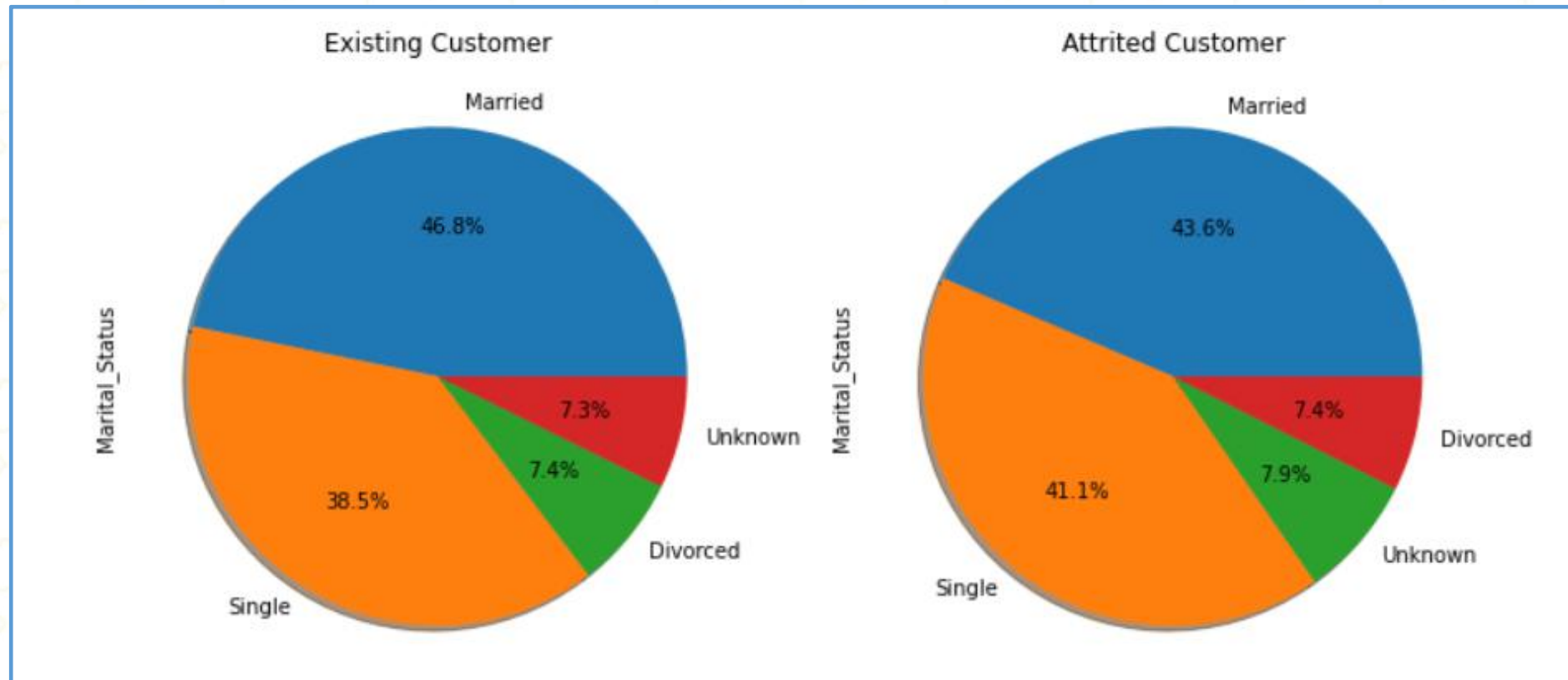






# Pie chart

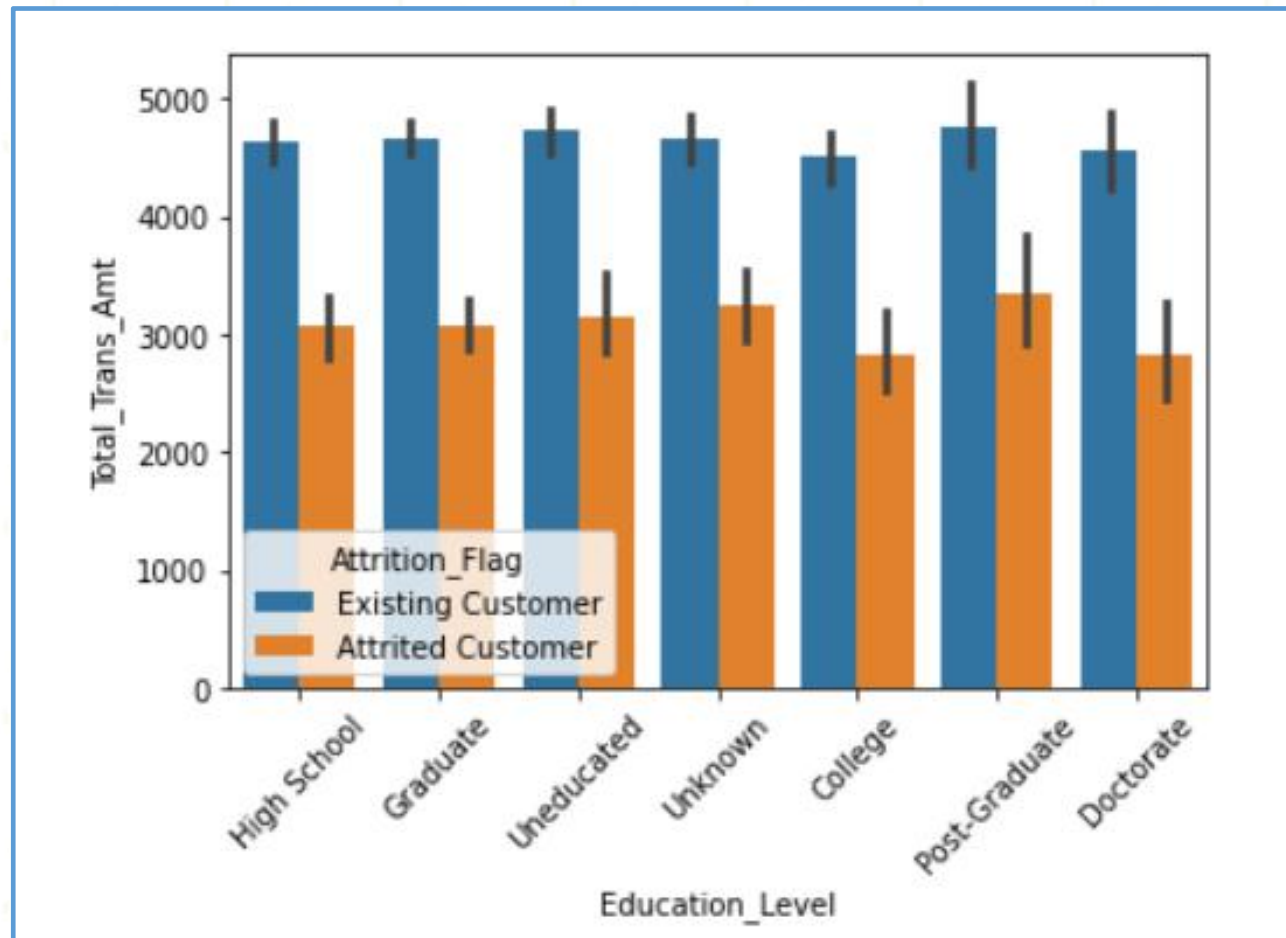
- Lingkaran yang dibagi menjadi beberapa irisan.
- Setiap irisan mewakili jumlah atau persentase data.





# Bar plot

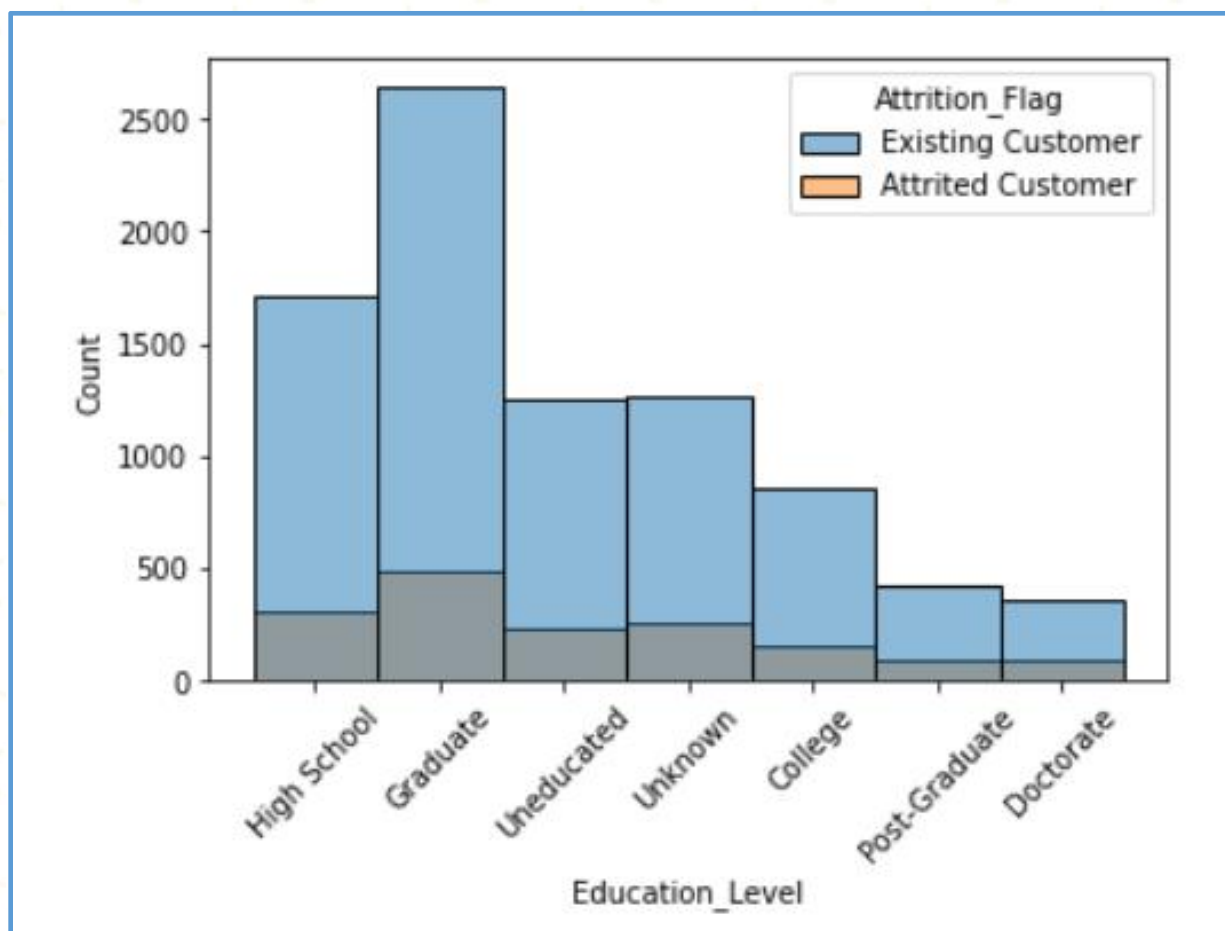
- Digunakan untuk menampilkan hubungan antara variabel numerik dan kategoris





# Histogram

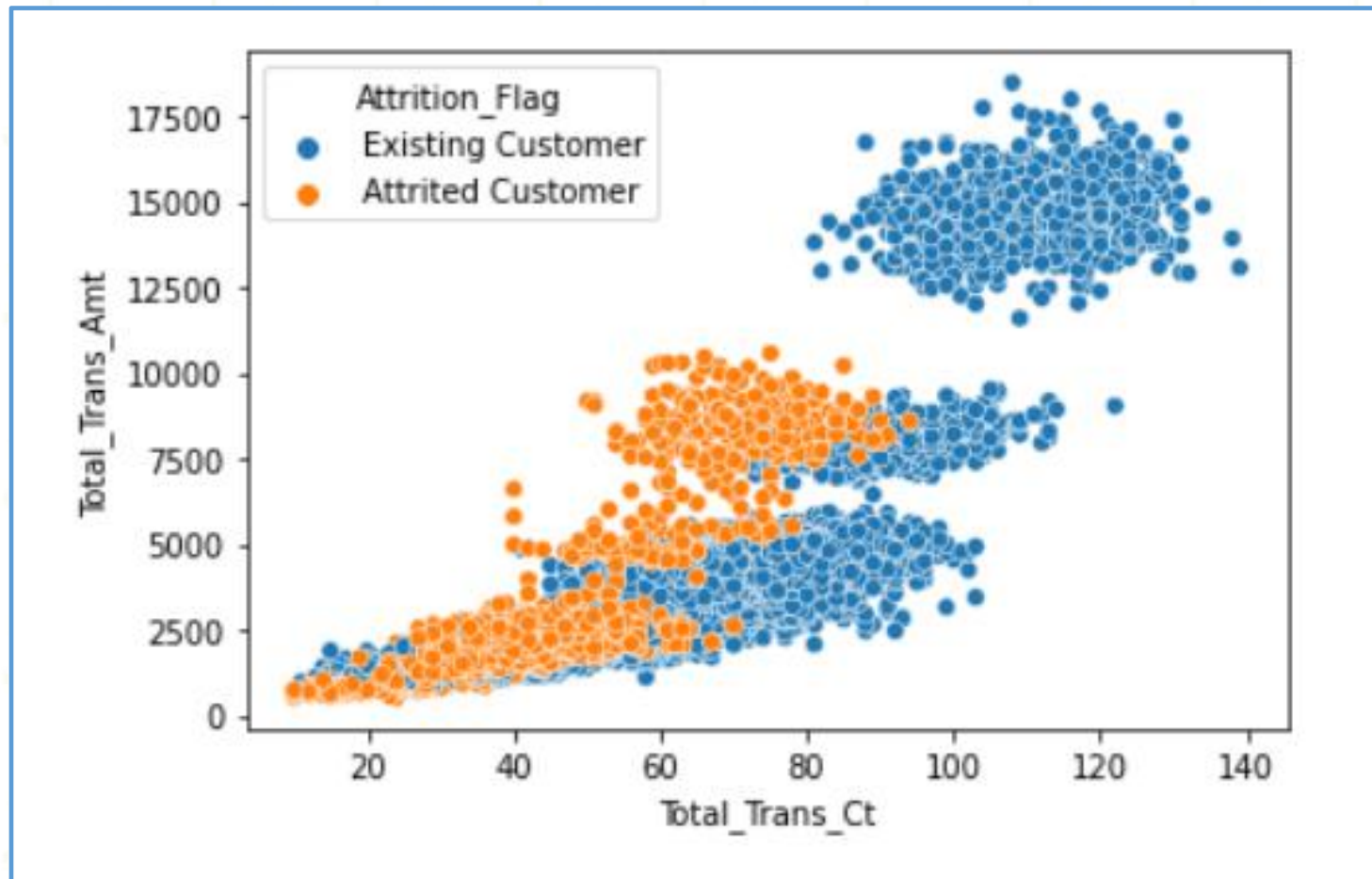
- Grafik yang paling umum digunakan untuk menunjukkan distribusi frekuensi





# Scatter plot

- Menggunakan titik untuk mewakili nilai untuk dua variabel numerik yang berbeda

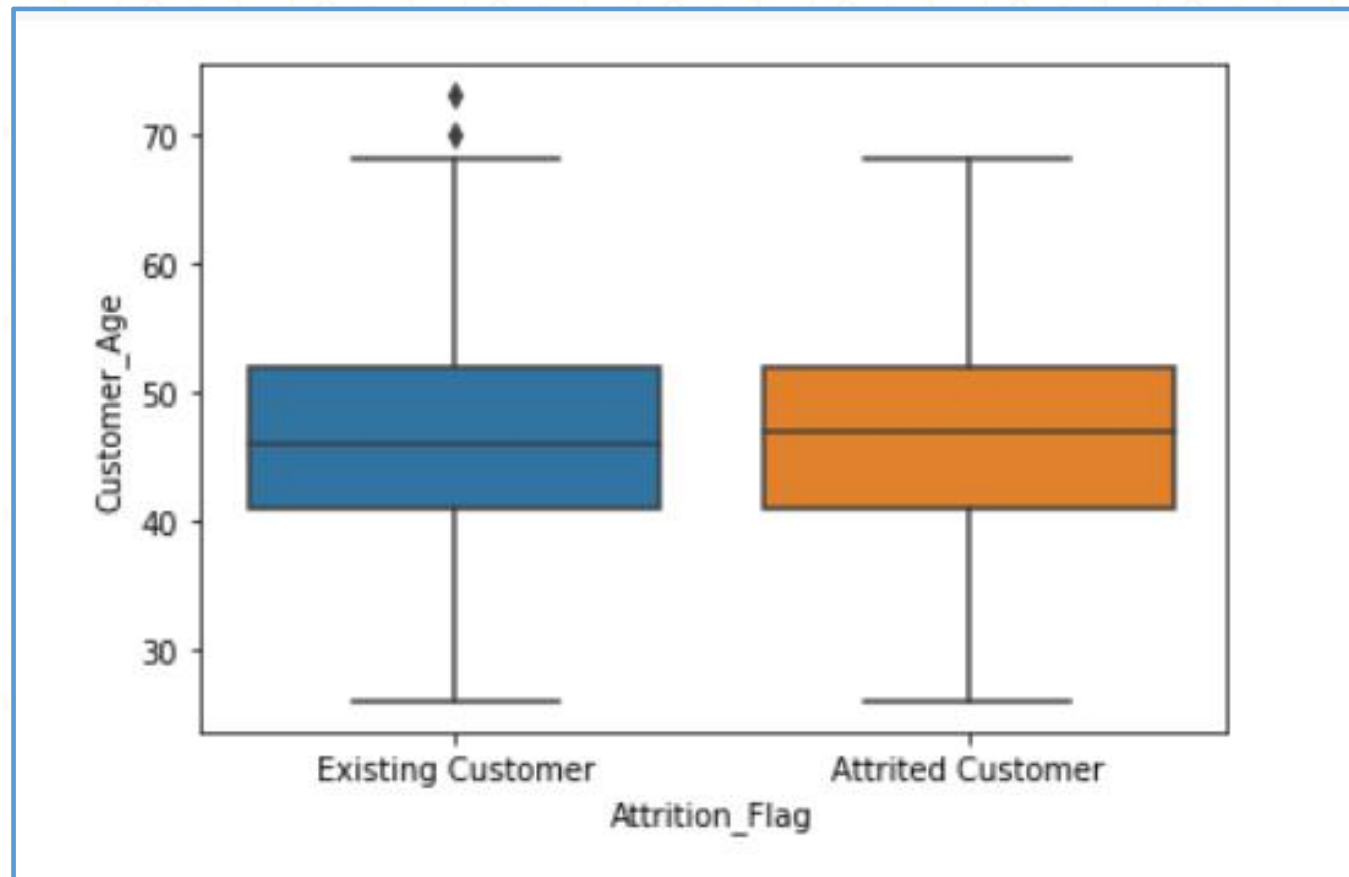






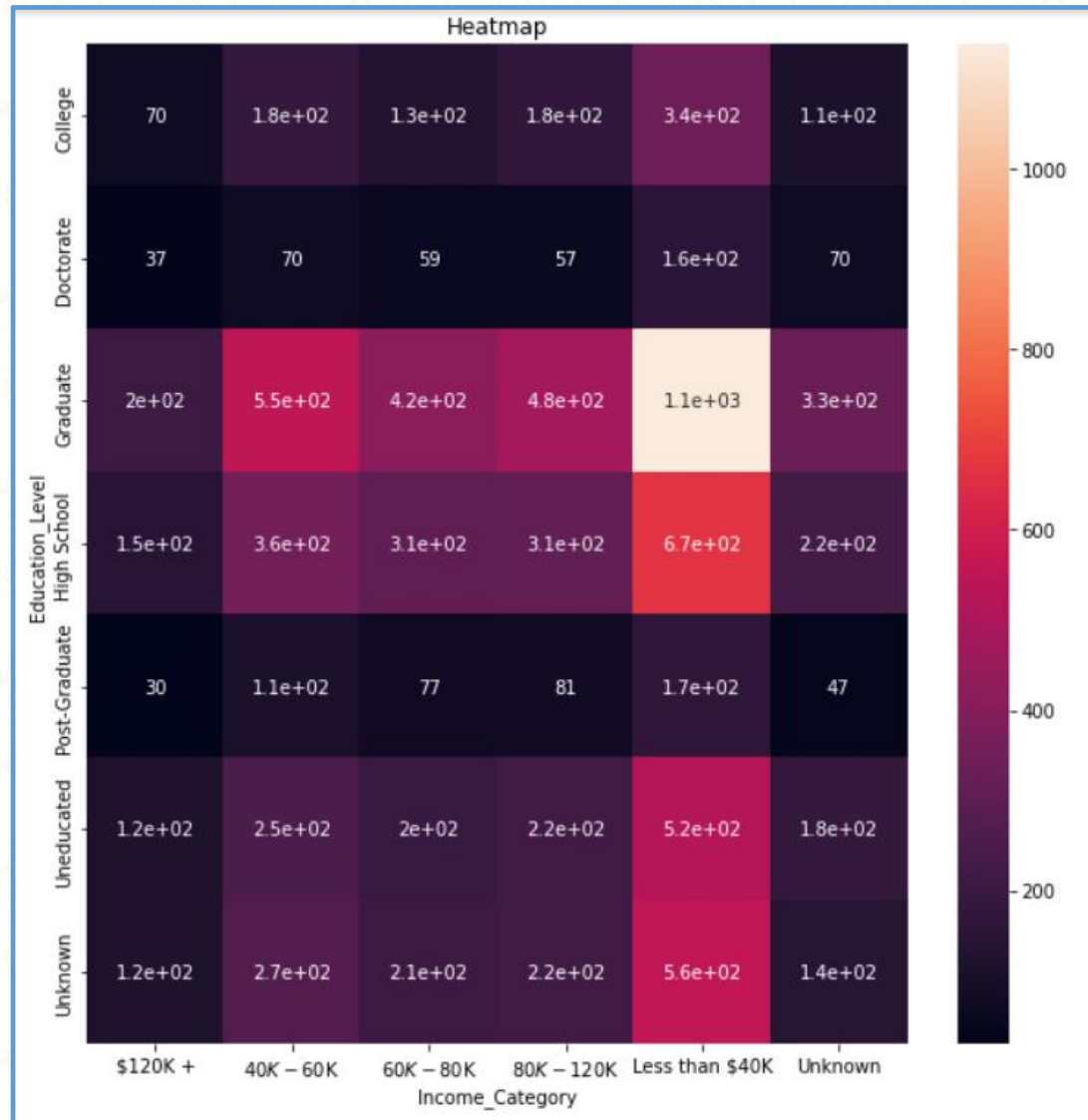
# Box plot

- Metode untuk menggambarkan kelompok data numerik secara grafis berdasarkan kuartilnya.





# Heatmap

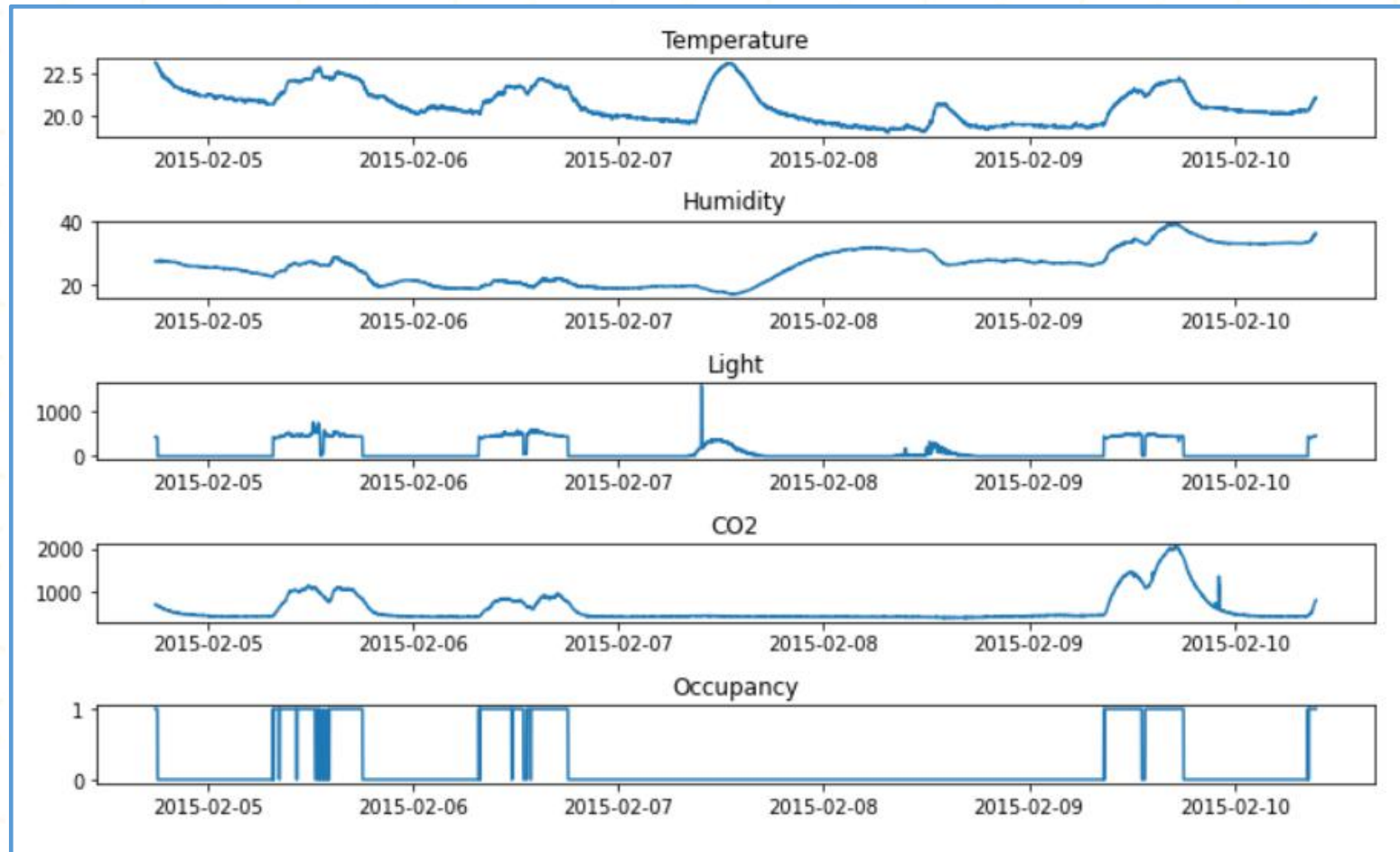


- Representasi grafis dari data yang menggunakan sistem kode warna untuk mewakili nilai yang berbeda.



# Line plot

- Menampilkan informasi sebagai serangkaian titik data yang dihubungkan oleh garis lurus



Thank  
YOU