Prediksi persentase Perokok di 34 Provinsi Indonesia pada Tahun 2025 dan Klasterisasi Berdasarkan Tingkat Keparahan

Farras-Vincent

2024-12-05

```r
# Load required libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────── tidyverse 2.0.0 ──
## ✓ forcats   1.0.0      ✓ stringr   1.5.1
## ✓ lubridate 1.9.3      ✓ tibble    3.2.1
## ✓ purrr     1.0.2      ✓ tidyr     1.3.1
## ✓ readr     2.1.5
```

```
## ── Conflicts ──────────────────────────────── tidyverse_conflicts() ──
## ✘ dplyr::filter() masks stats::filter()
## ✘ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
come errors
```

```r
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.4.2
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(MLmetrics)
```

```
## Warning: package 'MLmetrics' was built under R version 4.4.2
```

```
##
## Attaching package: 'MLmetrics'
##
## The following objects are masked from 'package:caret':
##
##     MAE, RMSE
##
## The following object is masked from 'package:base':
##
##     Recall
```

```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.4.2
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
# Load datasets
file1 <- read.csv("data_perokok_15-17.csv")
file2 <- read.csv("data_perokok_18-20.csv")
file3 <- read.csv("data_perokok_21-23.csv")
file4 <- read.csv("data_perokok_24.csv")

# Merge datasets by "Provinsi"
data <- file1 %>%
  inner_join(file2, by = "Provinsi") %>%
  inner_join(file3, by = "Provinsi") %>%
  inner_join(file4, by = "Provinsi")

# Jika nama kolom sudah berubah, hapus prefix 'X'
names(data) <- gsub("^X", "", names(data))

# Lanjutkan proses transform
data[, 2:11] <- lapply(data[, 2:11], function(x) as.numeric(as.character(x)))
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
# Hapus baris dengan Provinsi "Indonesia"
data <- data %>%
  filter(Provinsi != "INDONESIA")

# View merged data
data
```

```
##                    Provinsi  2015  2016  2017  2018  2019  2020  2021  2022  2023
## 1                      ACEH 29.82 28.16 28.85 31.76 28.70 28.06 28.30 27.58 28.66
## 2             SUMATERA UTARA 29.15 27.88 28.47 31.10 27.46 27.28 27.24 25.32 26.28
## 3             SUMATERA BARAT 32.41 30.59 31.71 35.32 30.75 30.08 30.50 30.27 30.42
## 4                      RIAU 31.21 29.61 29.34 32.72 29.04 28.06 28.34 26.86 27.76
## 5                     JAMBI 30.82 29.18 29.18 28.21 28.54 28.01 27.47 28.62 28.67
## 6           SUMATERA SELATAN 33.13 31.57 32.46 33.07 30.91 30.56 30.65 30.49 30.91
## 7                  BENGKULU 33.68 33.15 33.41 35.53 33.14 32.31 33.17 32.16 31.86
## 8                   LAMPUNG 34.12 33.39 33.75 35.95 34.39 33.43 34.07 33.81 34.08
## 9       KEP. BANGKA BELITUNG 30.70 29.32 29.67 32.32 29.18 28.23 28.16 26.84 27.33
## 10                 KEP. RIAU 29.18 29.25 29.98 29.67 27.59 26.16 26.17 23.08 25.49
## 11               DKI JAKARTA 27.31 26.42 24.72 30.77 26.04 25.75 24.44 21.25 22.60
## 12                JAWA BARAT 33.82 32.67 33.19 35.78 32.97 32.55 32.68 32.07 32.78
## 13               JAWA TENGAH 28.57 27.19 27.69 30.79 27.40 27.70 28.24 28.72 28.55
## 14            DI YOGYAKARTA 24.12 23.11 22.92 25.80 22.87 22.64 24.54 23.97 24.82
## 15                JAWA TIMUR 29.03 28.16 27.69 30.66 27.93 27.78 28.53 28.51 28.83
## 16                    BANTEN 32.95 31.64 31.77 34.93 31.69 31.58 31.76 31.21 29.34
## 17                      BALI 22.96 21.62 22.22 26.05 20.96 20.50 19.58 17.91 18.90
## 18       NUSA TENGGARA BARAT 31.60 30.88 30.59 33.92 30.49 30.58 32.71 33.20 32.79
## 19       NUSA TENGGARA TIMUR 25.47 24.91 27.31 31.30 27.33 26.14 27.22 26.76 26.64
## 20            KALIMANTAN BARAT 29.35 28.09 28.84 30.92 28.50 27.49 27.93 26.64 26.96
## 21           KALIMANTAN TENGAH 30.53 29.21 29.24 32.64 29.84 28.89 29.33 26.54 27.24
## 22           KALIMANTAN SELATAN 25.76 25.34 25.03 27.18 23.95 23.83 24.51 21.89 22.24
## 23            KALIMANTAN TIMUR 25.59 25.23 24.69 29.17 24.52 24.42 23.37 22.21 22.97
## 24            KALIMANTAN UTARA 28.61 28.38 28.18 29.82 27.63 25.66 27.46 24.23 25.36
## 25              SULAWESI UTARA 29.31 29.23 29.27 32.80 28.41 27.95 27.87 25.29 26.96
## 26             SULAWESI TENGAH 32.56 31.88 32.18 35.57 31.64 30.64 29.77 29.04 28.28
## 27             SULAWESI SELATAN 25.49 25.13 25.44 29.51 25.59 24.89 24.91 23.76 24.24
## 28            SULAWESI TENGGARA 28.49 27.60 29.22 31.46 26.80 25.77 25.85 23.35 24.66
## 29                  GORONTALO 33.93 31.71 34.46 36.56 32.37 30.30 30.50 30.38 30.69
## 30             SULAWESI BARAT 28.29 27.36 26.59 29.41 27.06 26.85 27.17 25.36 25.30
## 31                    MALUKU 27.19 25.68 27.46 32.74 27.09 26.18 27.90 26.80 28.04
## 32              MALUKU UTARA 31.14 30.23 30.57 35.29 31.18 29.83 29.84 28.82 28.82
## 33                PAPUA BARAT 29.28 26.18 27.60 32.73 28.67 25.80 27.07 24.80 25.30
## 34           PAPUA BARAT DAYA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 35                      PAPUA 26.67 24.04 27.28 28.97 26.05 26.97 24.91 22.22 22.30
## 36             PAPUA SELATAN    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 37              PAPUA TENGAH    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 38          PAPUA PEGUNUNGAN    NA    NA    NA    NA    NA    NA    NA    NA    NA
##     2024
## 1  28.61
## 2  26.69
## 3  31.45
## 4  28.02
## 5  29.76
## 6  31.01
## 7  32.96
## 8  33.84
## 9  27.97
## 10 25.46
## 11 22.56
## 12 32.98
## 13 29.13
## 14 25.18
## 15 28.72
```

```
## 16 31.31
## 17 19.22
## 18 32.40
## 19 27.51
## 20 28.06
## 21 28.02
## 22 23.07
## 23 23.99
## 24 25.93
## 25 26.93
## 26 28.98
## 27 24.66
## 28 24.80
## 29 31.61
## 30 26.99
## 31 29.43
## 32 29.72
## 33 25.90
## 34 25.00
## 35 22.11
## 36 32.28
## 37 19.48
## 38 19.54
```

```r
# Check for missing values
cat("Missing Values pada dataset awal \n")
```

```
## Missing Values pada dataset awal
```

```r
colSums(is.na(data))
```

```
## Provinsi     2015     2016     2017     2018     2019     2020     2021
##        0        4        4        4        4        4        4        4
##     2022     2023     2024
##        4        4        0
```

```r
# Salin data asli ke variabel baru untuk proses pembersihan
data_cleaned <- data

# 1. Menggabungkan Papua Barat dan Papua Barat Daya
papua_barat <- subset(data_cleaned, grepl("Papua Barat$", Provinsi, ignore.case = TRUE))
papua_barat_daya <- subset(data_cleaned, grepl("Papua Barat Daya", Provinsi, ignore.case = TRUE))

# Menghitung rata-rata untuk Papua Barat dan Papua Barat Daya
papua_barat_mean <- colMeans(rbind(papua_barat[2:11], papua_barat_daya[2:11]), na.rm = TRUE)

# 2. Menggabungkan Papua, Papua Selatan, Papua Tengah, dan Papua Pegunungan
papua_selatan <- subset(data_cleaned, grepl("Papua Selatan", Provinsi, ignore.case = TRUE))
papua_tengah <- subset(data_cleaned, grepl("Papua Tengah", Provinsi, ignore.case = TRUE))
papua_pegunungan <- subset(data_cleaned, grepl("Papua Pegunungan", Provinsi, ignore.case = TRUE))
papua <- subset(data_cleaned, grepl("Papua$", Provinsi, ignore.case = TRUE))

# Menghitung rata-rata untuk semua wilayah Papua
papua_mean <- colMeans(rbind(papua_selatan[2:11], papua_tengah[2:11], papua_pegunungan[2:11],
papua[2:11]), na.rm = TRUE)

# Mengupdate data Papua Barat dengan rata-rata
data_cleaned[grepl("Papua Barat", data_cleaned$Provinsi,ignore.case = TRUE), 2:11] <- papua_barat_mean

# Mengupdate data Papua dengan rata-rata
data_cleaned[grepl("Papua$", data_cleaned$Provinsi, ignore.case = TRUE), 2:11] <- papua_mean

# Menghapus baris Papua Barat Daya
data_cleaned <- data_cleaned[!grepl("Papua Barat Daya", data_cleaned$Provinsi, ignore.case = TRUE), ]

# Menghapus baris Papua Selatan, Papua Tengah, dan Papua Pegunungan
data_cleaned <- data_cleaned %>%
  filter(!grepl("Papua (Selatan|Tengah|Pegunungan)", Provinsi, ignore.case = TRUE))

# Mengecek nilai yang hilang (missing values)
cat("\nMissing Values pada dataset setelah cleaning data \n")
```

```
## 
## Missing Values pada dataset setelah cleaning data
```

```r
colSums(is.na(data_cleaned))
```

```
## Provinsi     2015     2016     2017     2018     2019     2020     2021
##        0        0        0        0        0        0        0        0
##     2022     2023     2024
##        0        0        0
```

```r
# Output hasil
data_cleaned
```

```
##                    Provinsi  2015  2016  2017  2018  2019  2020  2021  2022  2023
## 1                      ACEH 29.82 28.16 28.85 31.76 28.70 28.06 28.30 27.58 28.66
## 2             SUMATERA UTARA 29.15 27.88 28.47 31.10 27.46 27.28 27.24 25.32 26.28
## 3             SUMATERA BARAT 32.41 30.59 31.71 35.32 30.75 30.08 30.50 30.27 30.42
## 4                      RIAU 31.21 29.61 29.34 32.72 29.04 28.06 28.34 26.86 27.76
## 5                     JAMBI 30.82 29.18 29.18 28.21 28.54 28.01 27.47 28.62 28.67
## 6           SUMATERA SELATAN 33.13 31.57 32.46 33.07 30.91 30.56 30.65 30.49 30.91
## 7                  BENGKULU 33.68 33.15 33.41 35.53 33.14 32.31 33.17 32.16 31.86
## 8                   LAMPUNG 34.12 33.39 33.75 35.95 34.39 33.43 34.07 33.81 34.08
## 9       KEP. BANGKA BELITUNG 30.70 29.32 29.67 32.32 29.18 28.23 28.16 26.84 27.33
## 10                 KEP. RIAU 29.18 29.25 29.98 29.67 27.59 26.16 26.17 23.08 25.49
## 11               DKI JAKARTA 27.31 26.42 24.72 30.77 26.04 25.75 24.44 21.25 22.60
## 12                JAWA BARAT 33.82 32.67 33.19 35.78 32.97 32.55 32.68 32.07 32.78
## 13               JAWA TENGAH 28.57 27.19 27.69 30.79 27.40 27.70 28.24 28.72 28.55
## 14             DI YOGYAKARTA 24.12 23.11 22.92 25.80 22.87 22.64 24.54 23.97 24.82
## 15                JAWA TIMUR 29.03 28.16 27.69 30.66 27.93 27.78 28.53 28.51 28.83
## 16                    BANTEN 32.95 31.64 31.77 34.93 31.69 31.58 31.76 31.21 29.34
## 17                      BALI 22.96 21.62 22.22 26.05 20.96 20.50 19.58 17.91 18.90
## 18       NUSA TENGGARA BARAT 31.60 30.88 30.59 33.92 30.49 30.58 32.71 33.20 32.79
## 19       NUSA TENGGARA TIMUR 25.47 24.91 27.31 31.30 27.33 26.14 27.22 26.76 26.64
## 20           KALIMANTAN BARAT 29.35 28.09 28.84 30.92 28.50 27.49 27.93 26.64 26.96
## 21         KALIMANTAN TENGAH 30.53 29.21 29.24 32.64 29.84 28.89 29.33 26.54 27.24
## 22        KALIMANTAN SELATAN 25.76 25.34 25.03 27.18 23.95 23.83 24.51 21.89 22.24
## 23          KALIMANTAN TIMUR 25.59 25.23 24.69 29.17 24.52 24.42 23.37 22.21 22.97
## 24          KALIMANTAN UTARA 28.61 28.38 28.18 29.82 27.63 25.66 27.46 24.23 25.36
## 25            SULAWESI UTARA 29.31 29.23 29.27 32.80 28.41 27.95 27.87 25.29 26.96
## 26           SULAWESI TENGAH 32.56 31.88 32.18 35.57 31.64 30.64 29.77 29.04 28.28
## 27          SULAWESI SELATAN 25.49 25.13 25.44 29.51 25.59 24.89 24.91 23.76 24.24
## 28         SULAWESI TENGGARA 28.49 27.60 29.22 31.46 26.80 25.77 25.85 23.35 24.66
## 29                 GORONTALO 33.93 31.71 34.46 36.56 32.37 30.30 30.50 30.38 30.69
## 30           SULAWESI BARAT 28.29 27.36 26.59 29.41 27.06 26.85 27.17 25.36 25.30
## 31                    MALUKU 27.19 25.68 27.46 32.74 27.09 26.18 27.90 26.80 28.04
## 32              MALUKU UTARA 31.14 30.23 30.57 35.29 31.18 29.83 29.84 28.82 28.82
## 33               PAPUA BARAT 29.28 27.60 28.67 27.07 25.30 29.28 27.60 28.67 27.07
## 34                     PAPUA 26.67 24.04 27.28 28.97 26.05 26.97 24.91 22.22 22.30
##       2024
## 1  28.6100
## 2  26.6900
## 3  31.4500
## 4  28.0200
## 5  29.7600
## 6  31.0100
## 7  32.9600
## 8  33.8400
## 9  27.9700
## 10 25.4600
## 11 22.5600
## 12 32.9800
## 13 29.1300
## 14 25.1800
## 15 28.7200
## 16 31.3100
## 17 19.2200
## 18 32.4000
## 19 27.5100
```

```
## 20 28.0600
## 21 28.0200
## 22 23.0700
## 23 23.9900
## 24 25.9300
## 25 26.9300
## 26 28.9800
## 27 24.6600
## 28 24.8000
## 29 31.6100
## 30 26.9900
## 31 29.4300
## 32 29.7200
## 33 25.3000
## 34 23.3525
```

tren rata rata nasional per tahun

```
# Calculate yearly mean
yearly_means <- colMeans(data_cleaned[, -1], na.rm = TRUE)

# Visualize the trend
years <- names(yearly_means)
values <- as.numeric(yearly_means)

ggplot(data = data.frame(Year = years, Value = values), aes(x = Year, y = Value, group = 1))
+
  geom_line(color = "blue") +
  geom_point(size = 2) +
  labs(title = "Rata-rata Persentase Perokok di Indonesia (2015-2024)",
       x = "Tahun", y = "Persentase (%)") +
  theme_minimal()
```

Rata-rata Persentase Perokok di Indonesia (2015–2024)
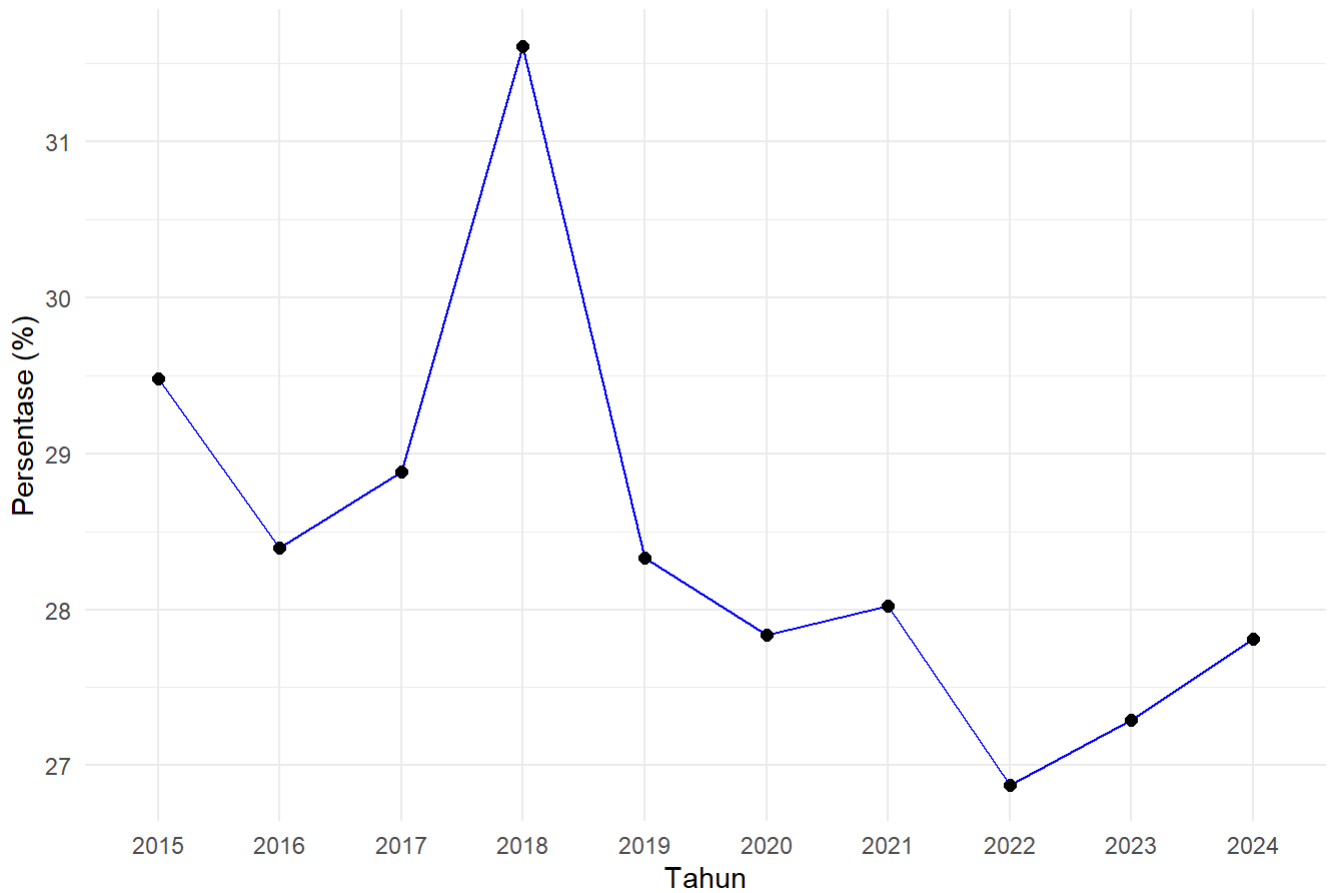
Distribusi Persentase Merokok (Setiap Tahun)

```r
# Transform data to long format for ggplot
library(tidyr)
long_data <- pivot_longer(data_cleaned, cols = -Provinsi, names_to = "Year", values_to = "Value")

# Plot the distribution for each year
ggplot(long_data, aes(x = Value)) +
  geom_histogram(binwidth = 2, fill = "skyblue", color = "black") +
  facet_wrap(~ Year, scales = "free_y") +
  labs(title = "Distribusi Persentase Merokok per Tahun",
       x = "Persentase Merokok (%)", y = "Frekuensi") +
  theme_minimal()
```

Distribusi Persentase Merokok per Tahun
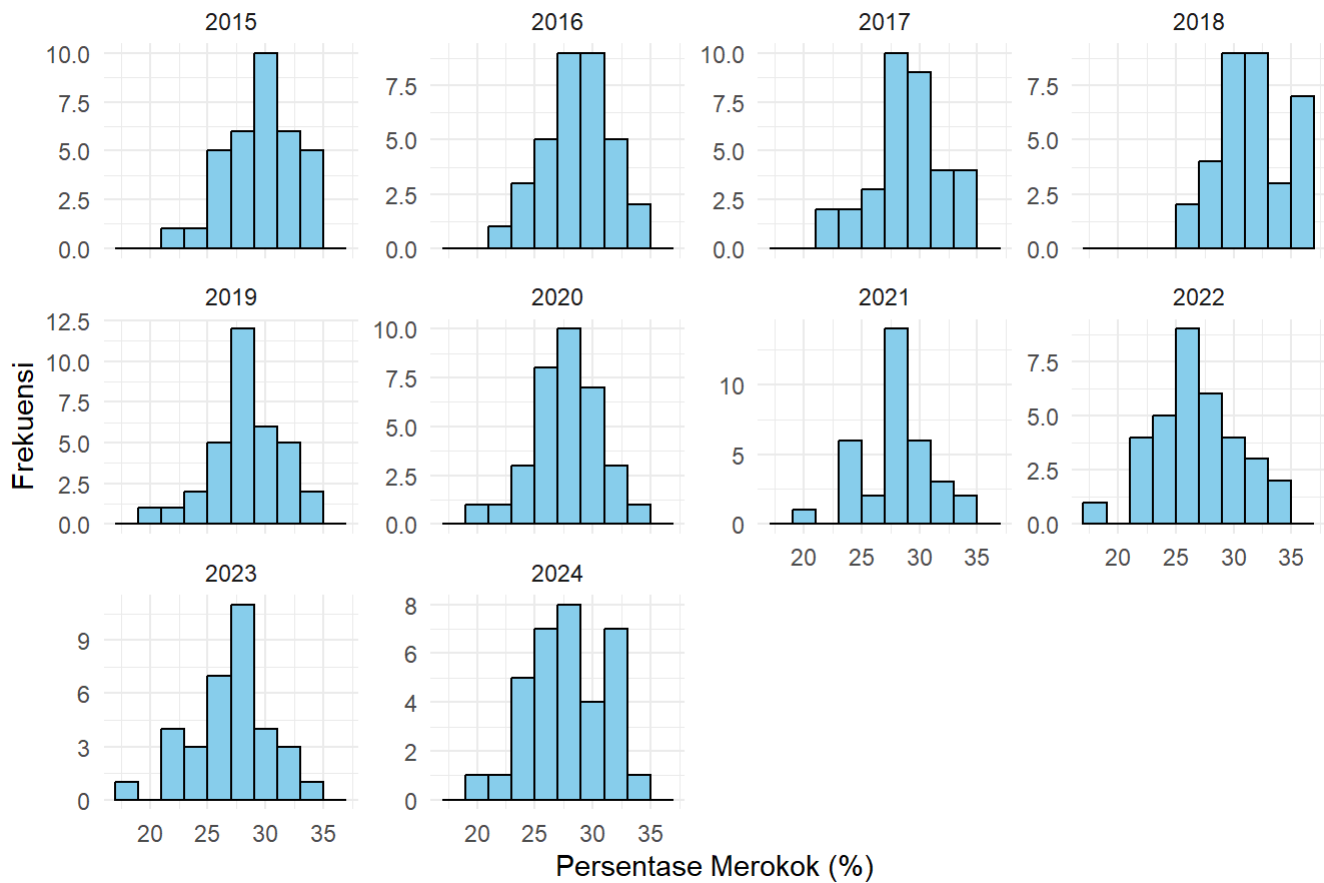

```
# Line plot for each province over the years
ggplot(long_data, aes(x = Year, y = Value, group = Provinsi, color = Provinsi)) +
  geom_line() +
  labs(title = "Perubahan Persentase Merokok per Provinsi (2015-2024)",
       x = "Tahun", y = "Persentase (%)") +
  theme_minimal() +
  theme(legend.text = element_text(size = 4),
      legend.title = element_text(size = 12, face = "bold"))
```

Perubahan Persentase Merokok per Provinsi (2015–2024)
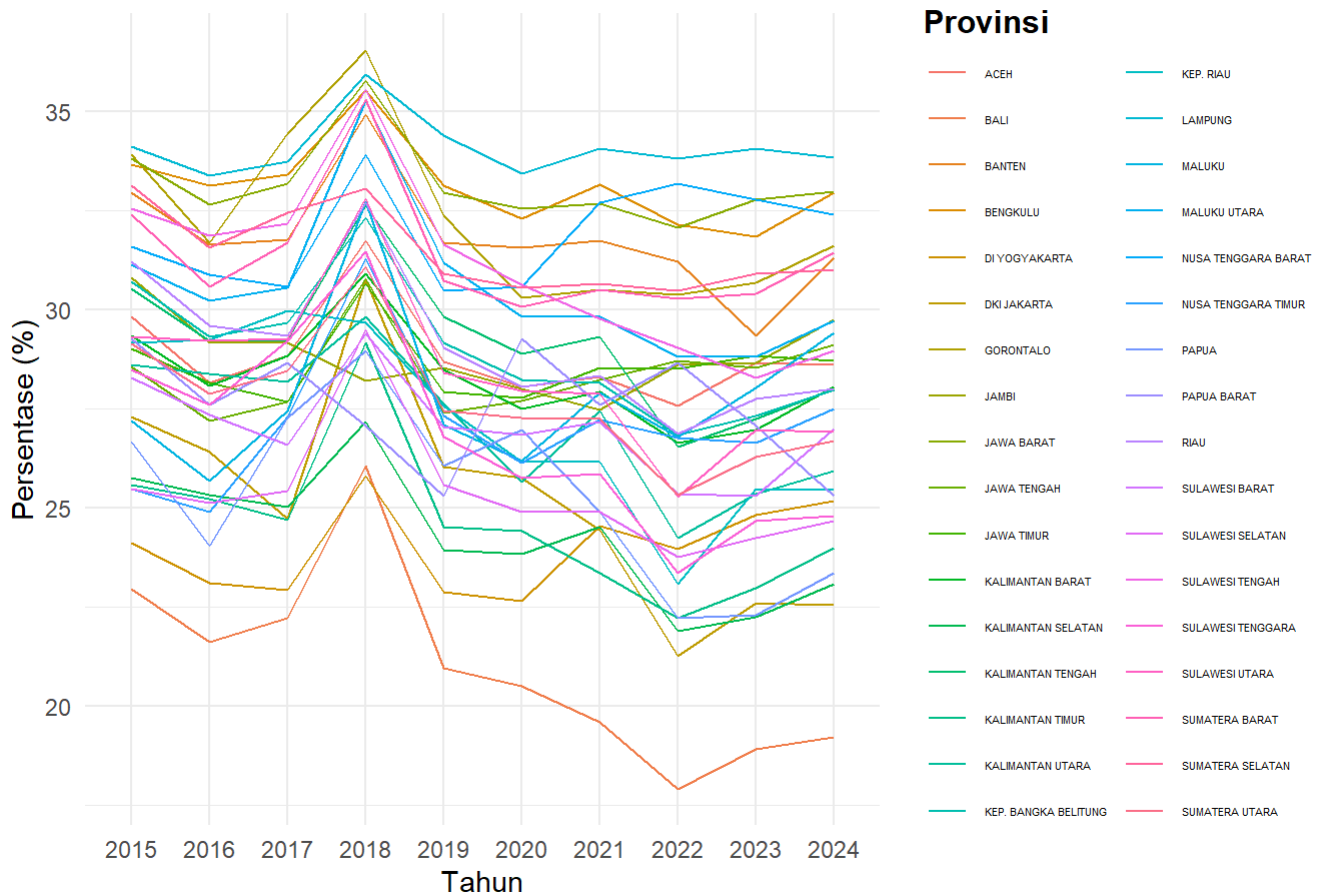
```
# Prepare data for heatmap
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 4.4.2
```

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
heatmap_data <- melt(data_cleaned, id.vars = "Provinsi", variable.name = "Year", value.name = "Value")

# Create heatmap
ggplot(heatmap_data, aes(x = Year, y = reorder(Provinsi, Value), fill = Value)) +
  geom_tile() +
  scale_fill_gradient(low = "yellow", high = "red") +
  labs(title = "Heatmap Persentase Merokok (2015–2024)",
       x = "Tahun", y = "Provinsi", fill = "Persentase (%)") +
  theme_minimal()
```

Heatmap Persentase Merokok (2015–2024)
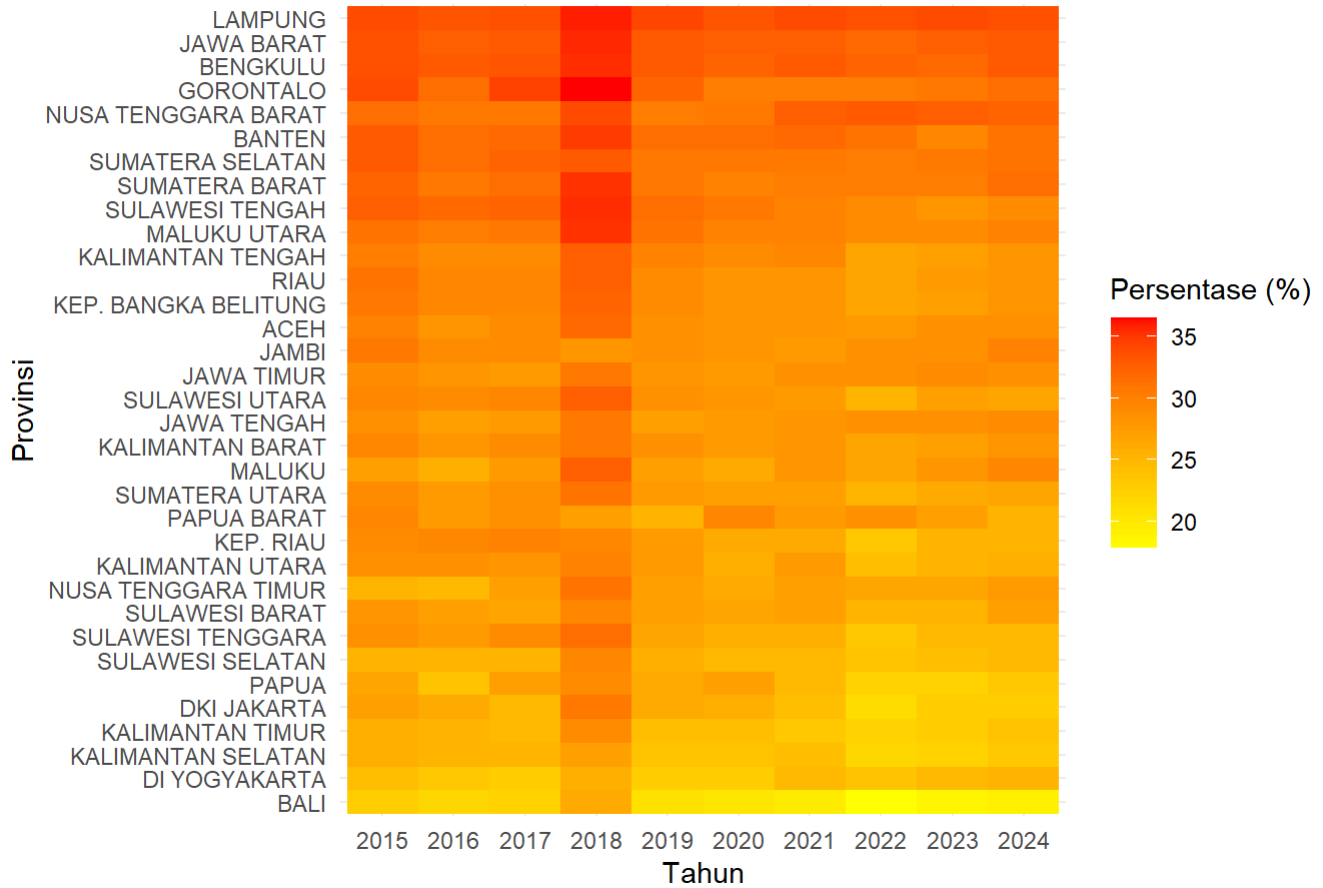
data preparation

```r
# Transformasikan data ke format time series

data_ts <- data_cleaned %>%
  select(-Provinsi) %>%
  t() %>%
  as.data.frame()
colnames(data_ts) <- data_cleaned$Provinsi
rownames(data_ts) <- as.character(2015:2024)

data_ts
```

```
##      ACEH SUMATERA UTARA SUMATERA BARAT  RIAU JAMBI SUMATERA SELATAN BENGKULU
## 2015 29.82          29.15          32.41 31.21 30.82             33.13    33.68
## 2016 28.16          27.88          30.59 29.61 29.18             31.57    33.15
## 2017 28.85          28.47          31.71 29.34 29.18             32.46    33.41
## 2018 31.76          31.10          35.32 32.72 28.21             33.07    35.53
## 2019 28.70          27.46          30.75 29.04 28.54             30.91    33.14
## 2020 28.06          27.28          30.08 28.06 28.01             30.56    32.31
## 2021 28.30          27.24          30.50 28.34 27.47             30.65    33.17
## 2022 27.58          25.32          30.27 26.86 28.62             30.49    32.16
## 2023 28.66          26.28          30.42 27.76 28.67             30.91    31.86
## 2024 28.61          26.69          31.45 28.02 29.76             31.01    32.96
##      LAMPUNG KEP. BANGKA BELITUNG KEP. RIAU DKI JAKARTA JAWA BARAT JAWA TENGAH
## 2015   34.12                30.70     29.18       27.31      33.82       28.57
## 2016   33.39                29.32     29.25       26.42      32.67       27.19
## 2017   33.75                29.67     29.98       24.72      33.19       27.69
## 2018   35.95                32.32     29.67       30.77      35.78       30.79
## 2019   34.39                29.18     27.59       26.04      32.97       27.40
## 2020   33.43                28.23     26.16       25.75      32.55       27.70
## 2021   34.07                28.16     26.17       24.44      32.68       28.24
## 2022   33.81                26.84     23.08       21.25      32.07       28.72
## 2023   34.08                27.33     25.49       22.60      32.78       28.55
## 2024   33.84                27.97     25.46       22.56      32.98       29.13
##      DI YOGYAKARTA JAWA TIMUR BANTEN  BALI NUSA TENGGARA BARAT
## 2015         24.12      29.03  32.95 22.96                31.60
## 2016         23.11      28.16  31.64 21.62                30.88
## 2017         22.92      27.69  31.77 22.22                30.59
## 2018         25.80      30.66  34.93 26.05                33.92
## 2019         22.87      27.93  31.69 20.96                30.49
## 2020         22.64      27.78  31.58 20.50                30.58
## 2021         24.54      28.53  31.76 19.58                32.71
## 2022         23.97      28.51  31.21 17.91                33.20
## 2023         24.82      28.83  29.34 18.90                32.79
## 2024         25.18      28.72  31.31 19.22                32.40
##      NUSA TENGGARA TIMUR KALIMANTAN BARAT KALIMANTAN TENGAH KALIMANTAN SELATAN
## 2015              25.47            29.35             30.53              25.76
## 2016              24.91            28.09             29.21              25.34
## 2017              27.31            28.84             29.24              25.03
## 2018              31.30            30.92             32.64              27.18
## 2019              27.33            28.50             29.84              23.95
## 2020              26.14            27.49             28.89              23.83
## 2021              27.22            27.93             29.33              24.51
## 2022              26.76            26.64             26.54              21.89
## 2023              26.64            26.96             27.24              22.24
## 2024              27.51            28.06             28.02              23.07
##      KALIMANTAN TIMUR KALIMANTAN UTARA SULAWESI UTARA SULAWESI TENGAH
## 2015            25.59            28.61          29.31            32.56
## 2016            25.23            28.38          29.23            31.88
## 2017            24.69            28.18          29.27            32.18
## 2018            29.17            29.82          32.80            35.57
## 2019            24.52            27.63          28.41            31.64
## 2020            24.42            25.66          27.95            30.64
## 2021            23.37            27.46          27.87            29.77
## 2022            22.21            24.23          25.29            29.04
## 2023            22.97            25.36          26.96            28.28
## 2024            23.99            25.93          26.93            28.98
```

```
##        SULAWESI SELATAN SULAWESI TENGGARA GORONTALO SULAWESI BARAT MALUKU
## 2015             25.49             28.49     33.93           28.29  27.19
## 2016             25.13             27.60     31.71           27.36  25.68
## 2017             25.44             29.22     34.46           26.59  27.46
## 2018             29.51             31.46     36.56           29.41  32.74
## 2019             25.59             26.80     32.37           27.06  27.09
## 2020             24.89             25.77     30.30           26.85  26.18
## 2021             24.91             25.85     30.50           27.17  27.90
## 2022             23.76             23.35     30.38           25.36  26.80
## 2023             24.24             24.66     30.69           25.30  28.04
## 2024             24.66             24.80     31.61           26.99  29.43
##      MALUKU UTARA PAPUA BARAT    PAPUA
## 2015         31.14       29.28 26.6700
## 2016         30.23       27.60 24.0400
## 2017         30.57       28.67 27.2800
## 2018         35.29       27.07 28.9700
## 2019         31.18       25.30 26.0500
## 2020         29.83       29.28 26.9700
## 2021         29.84       27.60 24.9100
## 2022         28.82       28.67 22.2200
## 2023         28.82       27.07 22.3000
## 2024         29.72       25.30 23.3525
```

```r
# List untuk menyimpan hasil prediksi
prediksi_2025 <- c()

data_prediksi = data_ts

# Loop untuk tiap provinsi
for (prov in colnames(data_prediksi)) {
  ts_data <- ts(data_prediksi[[prov]], start = 2015, frequency = 1)
  model_arima <- auto.arima(ts_data)
  pred <- forecast(model_arima, h = 1)$mean
  prediksi_2025 <- c(prediksi_2025, as.numeric(pred))
}

# Gabungkan hasil prediksi ke data asli
data_prediksi["2025", ] <- prediksi_2025
data_prediksi <- as.data.frame(t(data_prediksi))
data_prediksi <- cbind(Provinsi = rownames(data_prediksi), data_prediksi)
rownames(data_prediksi) <- NULL

data_prediksi
```

```
##                   Provinsi  2015  2016  2017  2018  2019  2020  2021  2022  2023
## 1                      ACEH 29.82 28.16 28.85 31.76 28.70 28.06 28.30 27.58 28.66
## 2            SUMATERA UTARA 29.15 27.88 28.47 31.10 27.46 27.28 27.24 25.32 26.28
## 3            SUMATERA BARAT 32.41 30.59 31.71 35.32 30.75 30.08 30.50 30.27 30.42
## 4                      RIAU 31.21 29.61 29.34 32.72 29.04 28.06 28.34 26.86 27.76
## 5                     JAMBI 30.82 29.18 29.18 28.21 28.54 28.01 27.47 28.62 28.67
## 6          SUMATERA SELATAN 33.13 31.57 32.46 33.07 30.91 30.56 30.65 30.49 30.91
## 7                  BENGKULU 33.68 33.15 33.41 35.53 33.14 32.31 33.17 32.16 31.86
## 8                   LAMPUNG 34.12 33.39 33.75 35.95 34.39 33.43 34.07 33.81 34.08
## 9      KEP. BANGKA BELITUNG 30.70 29.32 29.67 32.32 29.18 28.23 28.16 26.84 27.33
## 10                KEP. RIAU 29.18 29.25 29.98 29.67 27.59 26.16 26.17 23.08 25.49
## 11              DKI JAKARTA 27.31 26.42 24.72 30.77 26.04 25.75 24.44 21.25 22.60
## 12               JAWA BARAT 33.82 32.67 33.19 35.78 32.97 32.55 32.68 32.07 32.78
## 13              JAWA TENGAH 28.57 27.19 27.69 30.79 27.40 27.70 28.24 28.72 28.55
## 14            DI YOGYAKARTA 24.12 23.11 22.92 25.80 22.87 22.64 24.54 23.97 24.82
## 15               JAWA TIMUR 29.03 28.16 27.69 30.66 27.93 27.78 28.53 28.51 28.83
## 16                   BANTEN 32.95 31.64 31.77 34.93 31.69 31.58 31.76 31.21 29.34
## 17                     BALI 22.96 21.62 22.22 26.05 20.96 20.50 19.58 17.91 18.90
## 18     NUSA TENGGARA BARAT 31.60 30.88 30.59 33.92 30.49 30.58 32.71 33.20 32.79
## 19     NUSA TENGGARA TIMUR 25.47 24.91 27.31 31.30 27.33 26.14 27.22 26.76 26.64
## 20         KALIMANTAN BARAT 29.35 28.09 28.84 30.92 28.50 27.49 27.93 26.64 26.96
## 21       KALIMANTAN TENGAH 30.53 29.21 29.24 32.64 29.84 28.89 29.33 26.54 27.24
## 22       KALIMANTAN SELATAN 25.76 25.34 25.03 27.18 23.95 23.83 24.51 21.89 22.24
## 23        KALIMANTAN TIMUR 25.59 25.23 24.69 29.17 24.52 24.42 23.37 22.21 22.97
## 24        KALIMANTAN UTARA 28.61 28.38 28.18 29.82 27.63 25.66 27.46 24.23 25.36
## 25           SULAWESI UTARA 29.31 29.23 29.27 32.80 28.41 27.95 27.87 25.29 26.96
## 26          SULAWESI TENGAH 32.56 31.88 32.18 35.57 31.64 30.64 29.77 29.04 28.28
## 27         SULAWESI SELATAN 25.49 25.13 25.44 29.51 25.59 24.89 24.91 23.76 24.24
## 28        SULAWESI TENGGARA 28.49 27.60 29.22 31.46 26.80 25.77 25.85 23.35 24.66
## 29                GORONTALO 33.93 31.71 34.46 36.56 32.37 30.30 30.50 30.38 30.69
## 30          SULAWESI BARAT 28.29 27.36 26.59 29.41 27.06 26.85 27.17 25.36 25.30
## 31                   MALUKU 27.19 25.68 27.46 32.74 27.09 26.18 27.90 26.80 28.04
## 32             MALUKU UTARA 31.14 30.23 30.57 35.29 31.18 29.83 29.84 28.82 28.82
## 33             PAPUA BARAT 29.28 27.60 28.67 27.07 25.30 29.28 27.60 28.67 27.07
## 34                    PAPUA 26.67 24.04 27.28 28.97 26.05 26.97 24.91 22.22 22.30
##        2024    2025
## 1  28.6100 28.8500
## 2  26.6900 26.6900
## 3  31.4500 31.3500
## 4  28.0200 28.0200
## 5  29.7600 28.8460
## 6  31.0100 31.0100
## 7  32.9600 33.1370
## 8  33.8400 34.0830
## 9  27.9700 27.9700
## 10 25.4600 25.4600
## 11 22.5600 22.5600
## 12 32.9800 33.1490
## 13 29.1300 28.3980
## 14 25.1800 23.9970
## 15 28.7200 28.5840
## 16 31.3100 31.8180
## 17 19.2200 19.2200
## 18 32.4000 31.9160
## 19 27.5100 27.0590
```

```
## 20 28.0600 28.0600
## 21 28.0200 28.0200
## 22 23.0700 23.0700
## 23 23.9900 24.6160
## 24 25.9300 25.9300
## 25 26.9300 26.9300
## 26 28.9800 28.9800
## 27 24.6600 25.3620
## 28 24.8000 24.8000
## 29 31.6100 31.6100
## 30 26.9900 27.0380
## 31 29.4300 27.8510
## 32 29.7200 30.5440
## 33 25.3000 27.5840
## 34 23.3525 23.3525
```

```r
# Membuat data untuk 2024 dan 2025
data_hasil_long <- data_prediksi %>%
  select(Provinsi, `2024`, `2025`) %>%
  pivot_longer(cols = c("2024", "2025"), names_to = "Year", values_to = "Persentase")

# Membuat bar plot perbandingan 2024 dan 2025 dengan warna berbeda
ggplot(data_hasil_long, aes(x = reorder(Provinsi, Persentase), y = Persentase, fill = Year))
+
  geom_bar(stat = "identity", position = "dodge") +  # 'dodge' untuk meletakkan bar secara be
rdampingan
  coord_flip() +
  scale_fill_manual(values = c("2024" = "lightblue", "2025" = "lightcoral")) +  # Menetapkan
warna untuk 2024 dan 2025
  labs(title = "Perbandingan Persentase Merokok per Provinsi (2024 vs 2025)",
       x = "Provinsi", y = "Persentase (%)", fill = "Tahun") +
  theme_minimal()
```

Perbandingan Persentase Merokok per Provinsi (2024 vs 2025)
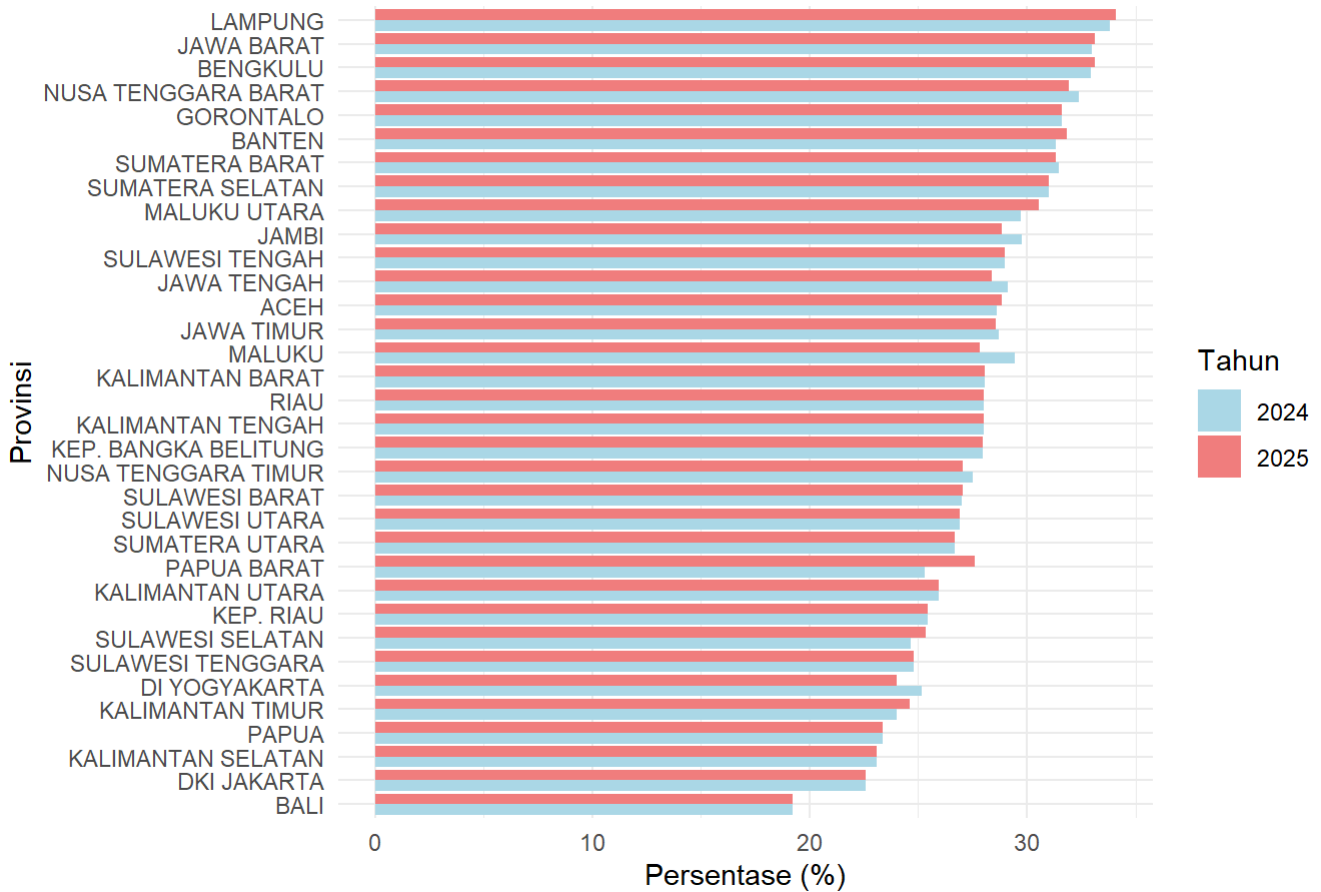

```
# Data untuk klasterisasi (2015-2024)

data_cluster = data_prediksi

data_kmeans <- data_cluster[, as.character(2015:2024)]

# Menentukan jumlah klaster (misalnya 3: rendah, sedang, tinggi)
set.seed(123)  # Untuk reproducibility
kmeans_model <- kmeans(data_kmeans, centers = 3)

# Tambahkan label klaster ke data
data_cluster$Cluster <- kmeans_model$cluster

print(data_cluster)
```

```
##                    Provinsi 2015  2016  2017  2018  2019  2020  2021  2022  2023
## 1                     ACEH 29.82 28.16 28.85 31.76 28.70 28.06 28.30 27.58 28.66
## 2            SUMATERA UTARA 29.15 27.88 28.47 31.10 27.46 27.28 27.24 25.32 26.28
## 3            SUMATERA BARAT 32.41 30.59 31.71 35.32 30.75 30.08 30.50 30.27 30.42
## 4                     RIAU 31.21 29.61 29.34 32.72 29.04 28.06 28.34 26.86 27.76
## 5                    JAMBI 30.82 29.18 29.18 28.21 28.54 28.01 27.47 28.62 28.67
## 6          SUMATERA SELATAN 33.13 31.57 32.46 33.07 30.91 30.56 30.65 30.49 30.91
## 7                  BENGKULU 33.68 33.15 33.41 35.53 33.14 32.31 33.17 32.16 31.86
## 8                  LAMPUNG 34.12 33.39 33.75 35.95 34.39 33.43 34.07 33.81 34.08
## 9    KEP. BANGKA BELITUNG 30.70 29.32 29.67 32.32 29.18 28.23 28.16 26.84 27.33
## 10                KEP. RIAU 29.18 29.25 29.98 29.67 27.59 26.16 26.17 23.08 25.49
## 11              DKI JAKARTA 27.31 26.42 24.72 30.77 26.04 25.75 24.44 21.25 22.60
## 12               JAWA BARAT 33.82 32.67 33.19 35.78 32.97 32.55 32.68 32.07 32.78
## 13              JAWA TENGAH 28.57 27.19 27.69 30.79 27.40 27.70 28.24 28.72 28.55
## 14            DI YOGYAKARTA 24.12 23.11 22.92 25.80 22.87 22.64 24.54 23.97 24.82
## 15               JAWA TIMUR 29.03 28.16 27.69 30.66 27.93 27.78 28.53 28.51 28.83
## 16                   BANTEN 32.95 31.64 31.77 34.93 31.69 31.58 31.76 31.21 29.34
## 17                     BALI 22.96 21.62 22.22 26.05 20.96 20.50 19.58 17.91 18.90
## 18   NUSA TENGGARA BARAT 31.60 30.88 30.59 33.92 30.49 30.58 32.71 33.20 32.79
## 19   NUSA TENGGARA TIMUR 25.47 24.91 27.31 31.30 27.33 26.14 27.22 26.76 26.64
## 20        KALIMANTAN BARAT 29.35 28.09 28.84 30.92 28.50 27.49 27.93 26.64 26.96
## 21      KALIMANTAN TENGAH 30.53 29.21 29.24 32.64 29.84 28.89 29.33 26.54 27.24
## 22      KALIMANTAN SELATAN 25.76 25.34 25.03 27.18 23.95 23.83 24.51 21.89 22.24
## 23       KALIMANTAN TIMUR 25.59 25.23 24.69 29.17 24.52 24.42 23.37 22.21 22.97
## 24       KALIMANTAN UTARA 28.61 28.38 28.18 29.82 27.63 25.66 27.46 24.23 25.36
## 25         SULAWESI UTARA 29.31 29.23 29.27 32.80 28.41 27.95 27.87 25.29 26.96
## 26        SULAWESI TENGAH 32.56 31.88 32.18 35.57 31.64 30.64 29.77 29.04 28.28
## 27       SULAWESI SELATAN 25.49 25.13 25.44 29.51 25.59 24.89 24.91 23.76 24.24
## 28      SULAWESI TENGGARA 28.49 27.60 29.22 31.46 26.80 25.77 25.85 23.35 24.66
## 29               GORONTALO 33.93 31.71 34.46 36.56 32.37 30.30 30.50 30.38 30.69
## 30         SULAWESI BARAT 28.29 27.36 26.59 29.41 27.06 26.85 27.17 25.36 25.30
## 31                   MALUKU 27.19 25.68 27.46 32.74 27.09 26.18 27.90 26.80 28.04
## 32            MALUKU UTARA 31.14 30.23 30.57 35.29 31.18 29.83 29.84 28.82 28.82
## 33             PAPUA BARAT 29.28 27.60 28.67 27.07 25.30 29.28 27.60 28.67 27.07
## 34                    PAPUA 26.67 24.04 27.28 28.97 26.05 26.97 24.91 22.22 22.30
##       2024    2025 Cluster
## 1  28.6100 28.8500       1
## 2  26.6900 26.6900       1
## 3  31.4500 31.3500       2
## 4  28.0200 28.0200       1
## 5  29.7600 28.8460       1
## 6  31.0100 31.0100       2
## 7  32.9600 33.1370       2
## 8  33.8400 34.0830       2
## 9  27.9700 27.9700       1
## 10 25.4600 25.4600       1
## 11 22.5600 22.5600       3
## 12 32.9800 33.1490       2
## 13 29.1300 28.3980       1
## 14 25.1800 23.9970       3
## 15 28.7200 28.5840       1
## 16 31.3100 31.8180       2
## 17 19.2200 19.2200       3
## 18 32.4000 31.9160       2
## 19 27.5100 27.0590       1
```

```
## 20 28.0600 28.0600        1
## 21 28.0200 28.0200        1
## 22 23.0700 23.0700        3
## 23 23.9900 24.6160        3
## 24 25.9300 25.9300        1
## 25 26.9300 26.9300        1
## 26 28.9800 28.9800        2
## 27 24.6600 25.3620        3
## 28 24.8000 24.8000        1
## 29 31.6100 31.6100        2
## 30 26.9900 27.0380        1
## 31 29.4300 27.8510        1
## 32 29.7200 30.5440        2
## 33 25.3000 27.5840        1
## 34 23.3525 23.3525        3
```

```r
ggplot(data_cluster, aes(x = reorder(Provinsi, `2025`), y = `2025`, fill = as.factor(Cluste
r))) +
  geom_bar(stat = "identity", color = "black") +  # Membuat bar plot
  coord_flip() +  # Membalik sumbu agar nama provinsi di sisi vertikal
  labs(
    title = "Persentase Tahun 2025 Berdasarkan Provinsi",
    x = "Provinsi",
    y = "Persentase (%)",
    fill = "Cluster"
  ) +
  scale_fill_manual(values = c("red", "blue", "green")) +  # Warna klaster
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),  # Rotasi teks sumbu X
    legend.position = "top"
  )
```

Persentase Tahun 2025 Berdasarkan Provinsi

Cluster ■ 1 ■ 2 ■ 3
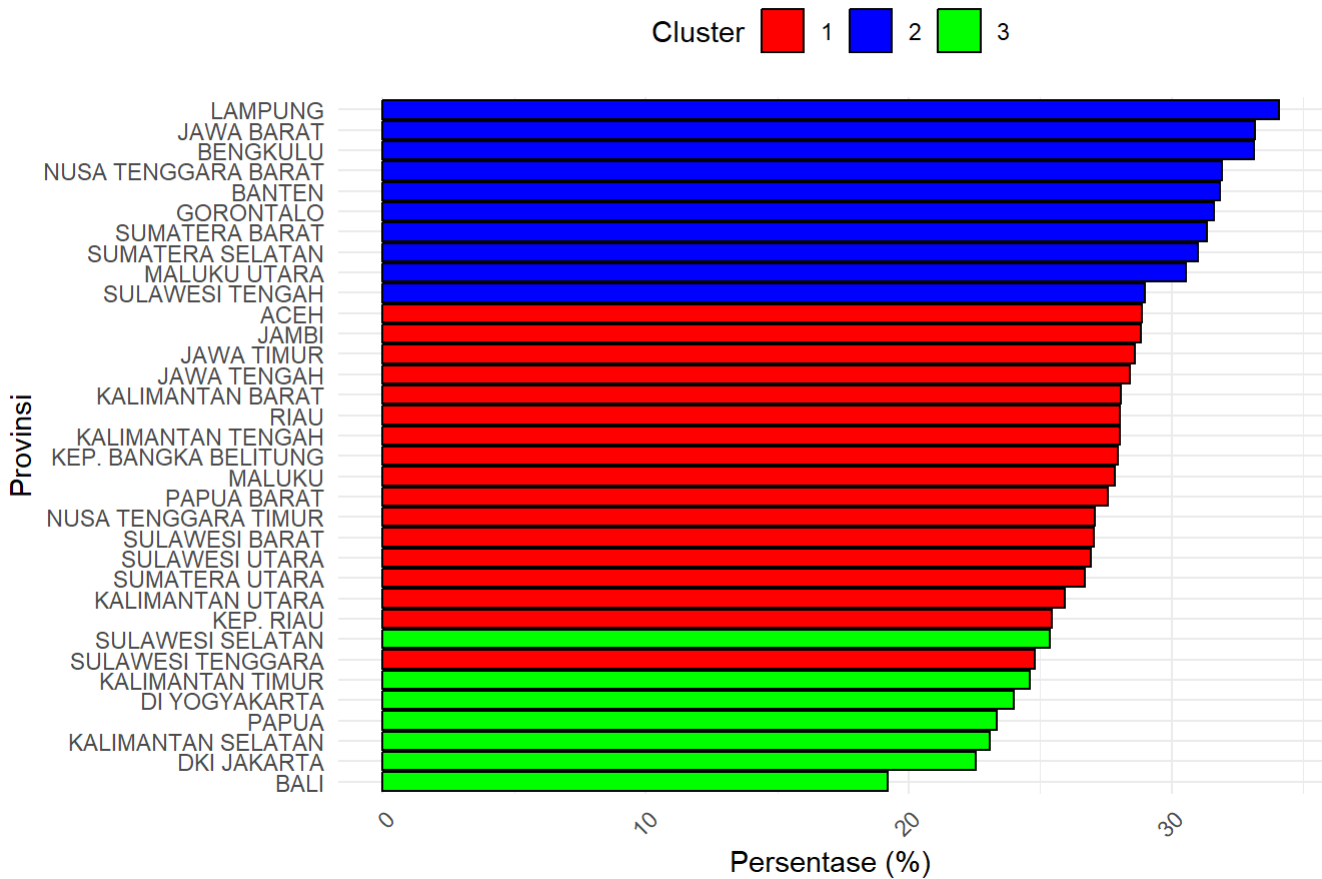

```
ggplot(data_cluster, aes(x = Provinsi, y = `2025`, color = as.factor(Cluster))) +
  geom_point(size = 4, alpha = 0.8) +  # Membuat scatter plot
  labs(
    title = "Diagram Clustering Berdasarkan Provinsi (Tahun 2025)",
    x = "Provinsi",
    y = "Persentase Tahun 2025 (%)",
    color = "Cluster"
  ) +
  scale_color_manual(values = c("red", "blue", "green")) +  # Warna untuk setiap cluster
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 90, hjust = 1),  # Rotasi nama provinsi
    legend.position = "top"  # Letak legenda di atas
  )
```

Diagram Clustering Berdasarkan Provinsi (Tahun 2025)
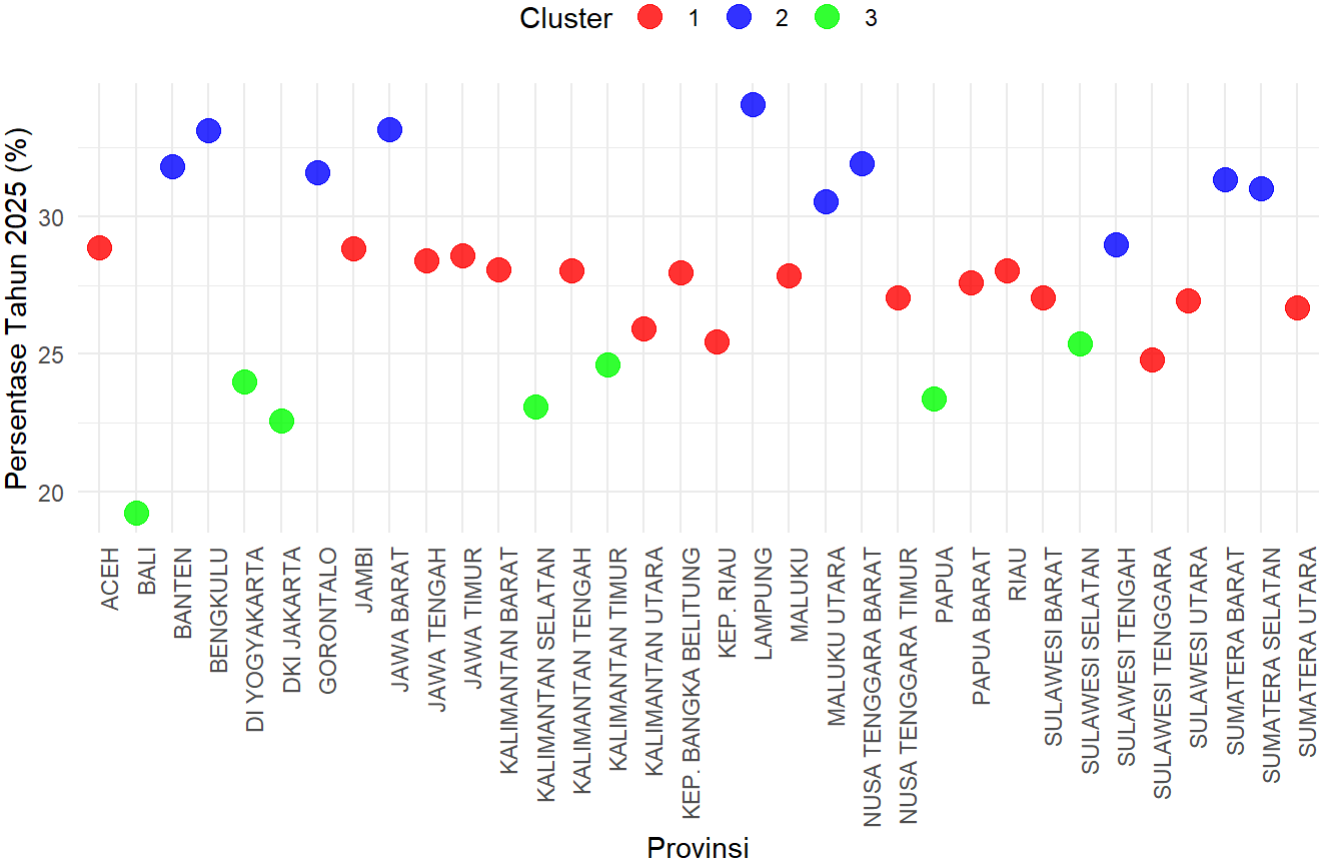

```r
# Transformasikan data menjadi format Long untuk visualisasi clustering
data_cluster_long <- data_cluster %>%
  pivot_longer(cols = as.character(2024:2025), names_to = "Year", values_to = "Persentase")
data_cluster_long
```

```
## # A tibble: 68 × 13
##    Provinsi       `2015` `2016` `2017` `2018` `2019` `2020` `2021` `2022` `2023`
##    <chr>          <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
##  1 ACEH            29.8   28.2   28.8   31.8   28.7   28.1   28.3   27.6   28.7
##  2 ACEH            29.8   28.2   28.8   31.8   28.7   28.1   28.3   27.6   28.7
##  3 SUMATERA UTARA  29.2   27.9   28.5   31.1   27.5   27.3   27.2   25.3   26.3
##  4 SUMATERA UTARA  29.2   27.9   28.5   31.1   27.5   27.3   27.2   25.3   26.3
##  5 SUMATERA BARAT  32.4   30.6   31.7   35.3   30.8   30.1   30.5   30.3   30.4
##  6 SUMATERA BARAT  32.4   30.6   31.7   35.3   30.8   30.1   30.5   30.3   30.4
##  7 RIAU            31.2   29.6   29.3   32.7   29.0   28.1   28.3   26.9   27.8
##  8 RIAU            31.2   29.6   29.3   32.7   29.0   28.1   28.3   26.9   27.8
##  9 JAMBI           30.8   29.2   29.2   28.2   28.5   28.0   27.5   28.6   28.7
## 10 JAMBI           30.8   29.2   29.2   28.2   28.5   28.0   27.5   28.6   28.7
## # i 58 more rows
## # i 3 more variables: Cluster <int>, Year <chr>, Persentase <dbl>
```

```
# Visualisasi plot clustering untuk tahun 2024 dan 2025
ggplot(data_cluster_long, aes(x = Year, y = Persentase, color = as.factor(Cluster))) +
  geom_jitter(width = 0.1, alpha = 0.6, size = 3) +  # Scatter plot dengan jitter
  scale_color_manual(values = c("red", "blue", "green"), name = "Cluster") +  # Warna klaster
  labs(
    title = "Plot Clustering untuk Tahun 2024 dan Prediksi 2025",
    x = "Tahun",
    y = "Persentase (%)"
  ) +
  theme_minimal() +
  theme(legend.position = "top")  # Menampilkan legend di atas plot
```


Plot Clustering untuk Tahun 2024 dan Prediksi 2025
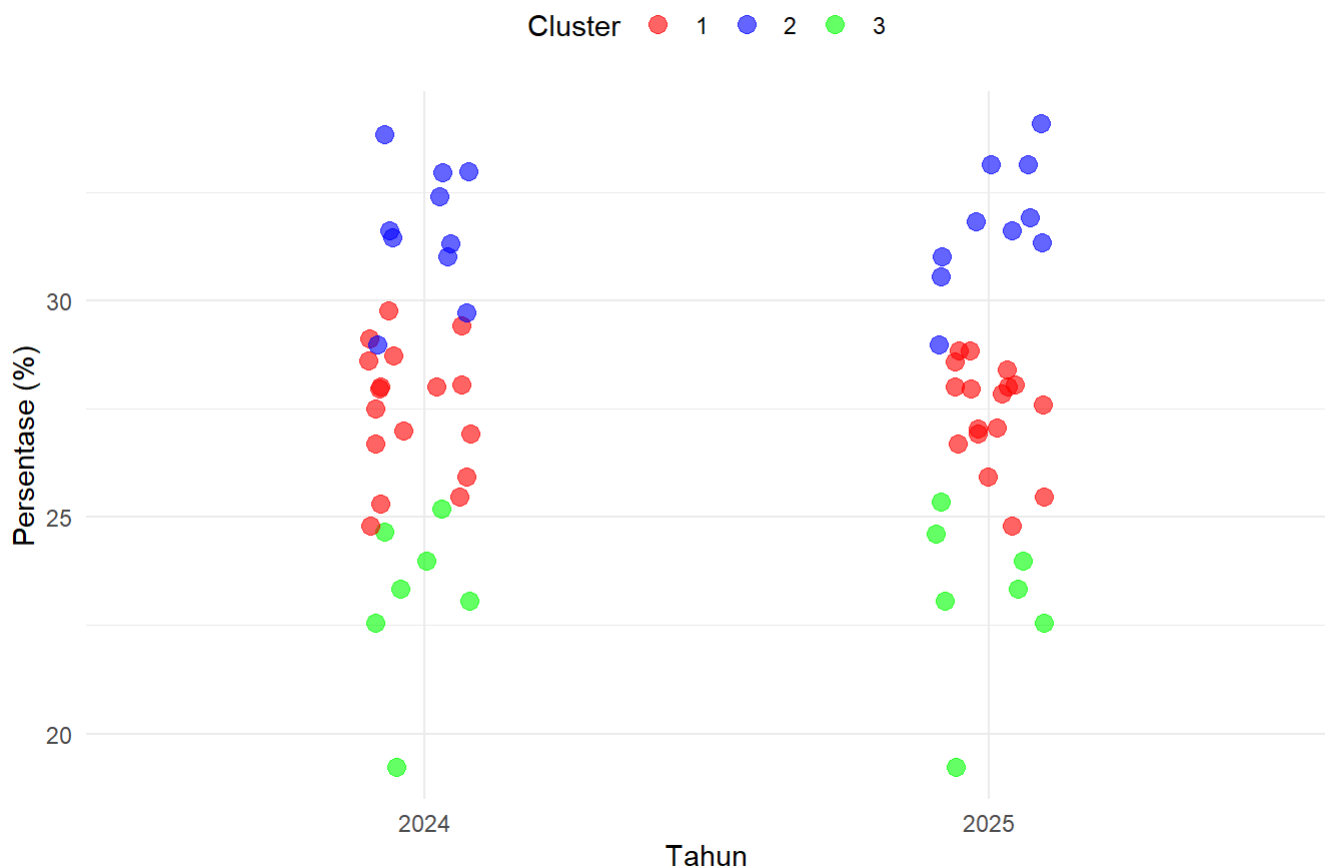
```
# Transformasikan data menjadi format long
data_cluster_long <- data_cluster %>%
  pivot_longer(cols = as.character(2015:2025), names_to = "Year", values_to = "Persentase")

# Visualisasi density plot
ggplot(data_cluster_long, aes(x = Persentase, fill = as.factor(Cluster))) +
  geom_density(alpha = 0.6) +  # Alpha mengatur transparansi
  facet_wrap(~Year, scales = "free_y") +  # Satu density plot untuk setiap tahun
  scale_fill_manual(values = c("red", "blue", "green"), name = "Cluster") +  # Warna khusus p
er klaster
  labs(
    title = "Density Plot Per Klaster (2015-2025)",
    x = "Persentase (%)",
    y = "Density"
  ) +
  theme_minimal()
```

Density Plot Per Klaster (2015-2025)
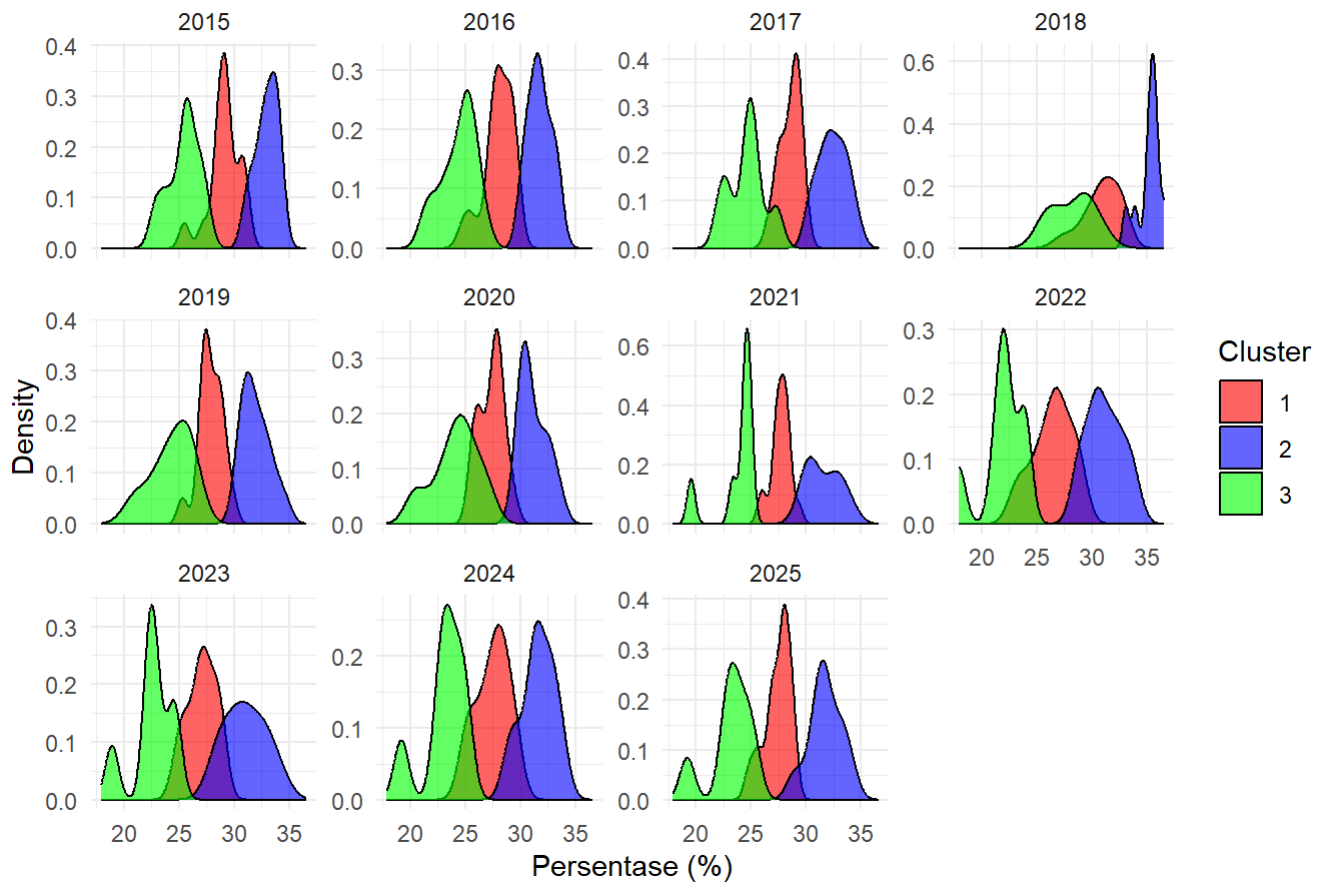

```
# 1. Plot Distribusi Klaster (Boxplot)
data_long <- data_cluster %>%
  pivot_longer(cols = as.character(2015:2024), names_to = "Tahun", values_to = "Persentase")

ggplot(data_long, aes(x = as.factor(Cluster), y = Persentase, fill = as.factor(Cluster))) +
  geom_boxplot() +
  labs(
    title = "Distribusi Persentase Perokok Berdasarkan Klaster (2015-2024)",
    x = "Klaster",
    y = "Persentase Perokok",
    fill = "Klaster"
  ) +
  theme_minimal()
```
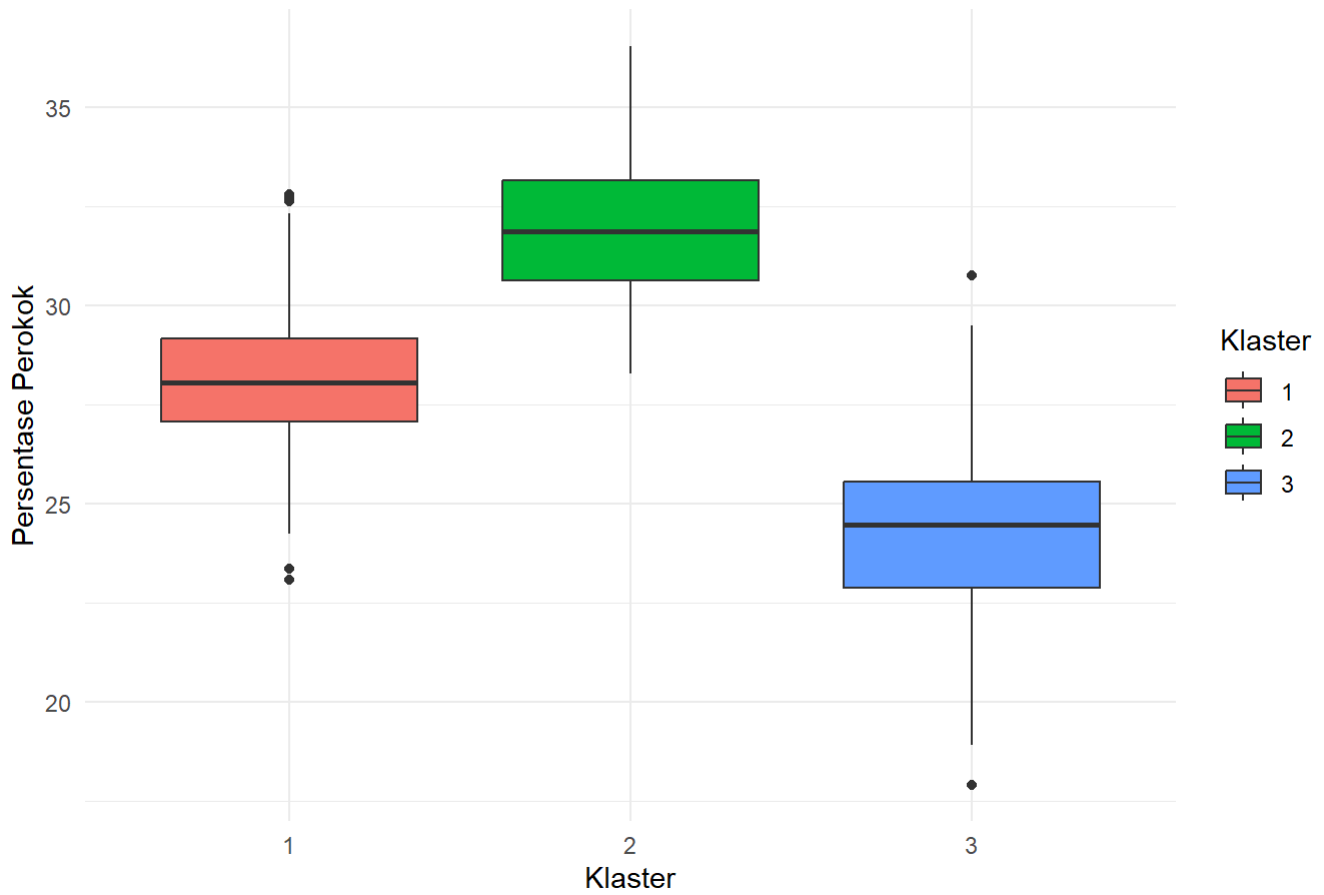
Distribusi Persentase Perokok Berdasarkan Klaster (2015-2024)


```
# Beri Label pada prediksi 2025 sesuai klaster
data_cluster <- data_cluster %>%
  mutate(Prioritas = case_when(
    Cluster == 1 ~ "Sedang",
    Cluster == 2 ~ "Tinggi",
    Cluster == 3 ~ "Rendah",
  ))
print(data_cluster)
```

```
##                 Provinsi 2015  2016  2017  2018  2019  2020  2021  2022  2023
## 1                   ACEH 29.82 28.16 28.85 31.76 28.70 28.06 28.30 27.58 28.66
## 2          SUMATERA UTARA 29.15 27.88 28.47 31.10 27.46 27.28 27.24 25.32 26.28
## 3          SUMATERA BARAT 32.41 30.59 31.71 35.32 30.75 30.08 30.50 30.27 30.42
## 4                   RIAU 31.21 29.61 29.34 32.72 29.04 28.06 28.34 26.86 27.76
## 5                  JAMBI 30.82 29.18 29.18 28.21 28.54 28.01 27.47 28.62 28.67
## 6        SUMATERA SELATAN 33.13 31.57 32.46 33.07 30.91 30.56 30.65 30.49 30.91
## 7                BENGKULU 33.68 33.15 33.41 35.53 33.14 32.31 33.17 32.16 31.86
## 8                LAMPUNG 34.12 33.39 33.75 35.95 34.39 33.43 34.07 33.81 34.08
## 9    KEP. BANGKA BELITUNG 30.70 29.32 29.67 32.32 29.18 28.23 28.16 26.84 27.33
## 10              KEP. RIAU 29.18 29.25 29.98 29.67 27.59 26.16 26.17 23.08 25.49
## 11             DKI JAKARTA 27.31 26.42 24.72 30.77 26.04 25.75 24.44 21.25 22.60
## 12              JAWA BARAT 33.82 32.67 33.19 35.78 32.97 32.55 32.68 32.07 32.78
## 13             JAWA TENGAH 28.57 27.19 27.69 30.79 27.40 27.70 28.24 28.72 28.55
## 14          DI YOGYAKARTA 24.12 23.11 22.92 25.80 22.87 22.64 24.54 23.97 24.82
## 15              JAWA TIMUR 29.03 28.16 27.69 30.66 27.93 27.78 28.53 28.51 28.83
## 16                 BANTEN 32.95 31.64 31.77 34.93 31.69 31.58 31.76 31.21 29.34
## 17                   BALI 22.96 21.62 22.22 26.05 20.96 20.50 19.58 17.91 18.90
## 18   NUSA TENGGARA BARAT 31.60 30.88 30.59 33.92 30.49 30.58 32.71 33.20 32.79
## 19   NUSA TENGGARA TIMUR 25.47 24.91 27.31 31.30 27.33 26.14 27.22 26.76 26.64
## 20       KALIMANTAN BARAT 29.35 28.09 28.84 30.92 28.50 27.49 27.93 26.64 26.96
## 21     KALIMANTAN TENGAH 30.53 29.21 29.24 32.64 29.84 28.89 29.33 26.54 27.24
## 22     KALIMANTAN SELATAN 25.76 25.34 25.03 27.18 23.95 23.83 24.51 21.89 22.24
## 23      KALIMANTAN TIMUR 25.59 25.23 24.69 29.17 24.52 24.42 23.37 22.21 22.97
## 24      KALIMANTAN UTARA 28.61 28.38 28.18 29.82 27.63 25.66 27.46 24.23 25.36
## 25         SULAWESI UTARA 29.31 29.23 29.27 32.80 28.41 27.95 27.87 25.29 26.96
## 26        SULAWESI TENGAH 32.56 31.88 32.18 35.57 31.64 30.64 29.77 29.04 28.28
## 27       SULAWESI SELATAN 25.49 25.13 25.44 29.51 25.59 24.89 24.91 23.76 24.24
## 28      SULAWESI TENGGARA 28.49 27.60 29.22 31.46 26.80 25.77 25.85 23.35 24.66
## 29              GORONTALO 33.93 31.71 34.46 36.56 32.37 30.30 30.50 30.38 30.69
## 30         SULAWESI BARAT 28.29 27.36 26.59 29.41 27.06 26.85 27.17 25.36 25.30
## 31                 MALUKU 27.19 25.68 27.46 32.74 27.09 26.18 27.90 26.80 28.04
## 32           MALUKU UTARA 31.14 30.23 30.57 35.29 31.18 29.83 29.84 28.82 28.82
## 33           PAPUA BARAT 29.28 27.60 28.67 27.07 25.30 29.28 27.60 28.67 27.07
## 34                  PAPUA 26.67 24.04 27.28 28.97 26.05 26.97 24.91 22.22 22.30
##      2024    2025 Cluster Prioritas
## 1  28.6100 28.8500       1    Sedang
## 2  26.6900 26.6900       1    Sedang
## 3  31.4500 31.3500       2    Tinggi
## 4  28.0200 28.0200       1    Sedang
## 5  29.7600 28.8460       1    Sedang
## 6  31.0100 31.0100       2    Tinggi
## 7  32.9600 33.1370       2    Tinggi
## 8  33.8400 34.0830       2    Tinggi
## 9  27.9700 27.9700       1    Sedang
## 10 25.4600 25.4600       1    Sedang
## 11 22.5600 22.5600       3    Rendah
## 12 32.9800 33.1490       2    Tinggi
## 13 29.1300 28.3980       1    Sedang
## 14 25.1800 23.9970       3    Rendah
## 15 28.7200 28.5840       1    Sedang
## 16 31.3100 31.8180       2    Tinggi
## 17 19.2200 19.2200       3    Rendah
## 18 32.4000 31.9160       2    Tinggi
## 19 27.5100 27.0590       1    Sedang
```

```
## 20 28.0600 28.0600        1      Sedang
## 21 28.0200 28.0200        1      Sedang
## 22 23.0700 23.0700        3      Rendah
## 23 23.9900 24.6160        3      Rendah
## 24 25.9300 25.9300        1      Sedang
## 25 26.9300 26.9300        1      Sedang
## 26 28.9800 28.9800        2      Tinggi
## 27 24.6600 25.3620        3      Rendah
## 28 24.8000 24.8000        1      Sedang
## 29 31.6100 31.6100        2      Tinggi
## 30 26.9900 27.0380        1      Sedang
## 31 29.4300 27.8510        1      Sedang
## 32 29.7200 30.5440        2      Tinggi
## 33 25.3000 27.5840        1      Sedang
## 34 23.3525 23.3525        3      Rendah
```

Provinsi dengan tingkat prioritas tinggi

```
# Filter data dengan prioritas tinggi
prioritas_tinggi <- data_cluster %>%
  filter(Prioritas == "Tinggi")
print(prioritas_tinggi)
```

```
##                 Provinsi  2015  2016  2017  2018  2019  2020  2021  2022  2023
## 1        SUMATERA BARAT 32.41 30.59 31.71 35.32 30.75 30.08 30.50 30.27 30.42
## 2      SUMATERA SELATAN 33.13 31.57 32.46 33.07 30.91 30.56 30.65 30.49 30.91
## 3              BENGKULU 33.68 33.15 33.41 35.53 33.14 32.31 33.17 32.16 31.86
## 4               LAMPUNG 34.12 33.39 33.75 35.95 34.39 33.43 34.07 33.81 34.08
## 5            JAWA BARAT 33.82 32.67 33.19 35.78 32.97 32.55 32.68 32.07 32.78
## 6                BANTEN 32.95 31.64 31.77 34.93 31.69 31.58 31.76 31.21 29.34
## 7   NUSA TENGGARA BARAT 31.60 30.88 30.59 33.92 30.49 30.58 32.71 33.20 32.79
## 8       SULAWESI TENGAH 32.56 31.88 32.18 35.57 31.64 30.64 29.77 29.04 28.28
## 9             GORONTALO 33.93 31.71 34.46 36.56 32.37 30.30 30.50 30.38 30.69
## 10        MALUKU UTARA 31.14 30.23 30.57 35.29 31.18 29.83 29.84 28.82 28.82
##       2024   2025 Cluster Prioritas
## 1   31.45 31.350       2    Tinggi
## 2   31.01 31.010       2    Tinggi
## 3   32.96 33.137       2    Tinggi
## 4   33.84 34.083       2    Tinggi
## 5   32.98 33.149       2    Tinggi
## 6   31.31 31.818       2    Tinggi
## 7   32.40 31.916       2    Tinggi
## 8   28.98 28.980       2    Tinggi
## 9   31.61 31.610       2    Tinggi
## 10 29.72 30.544       2    Tinggi
```

```
ggplot(prioritas_tinggi, aes(x = Provinsi, y = `2025`)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(title = "Provinsi dengan Prioritas Tinggi", x = "Provinsi", y = "Nilai Tahun 2025") +
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Provinsi dengan Prioritas Tinggi
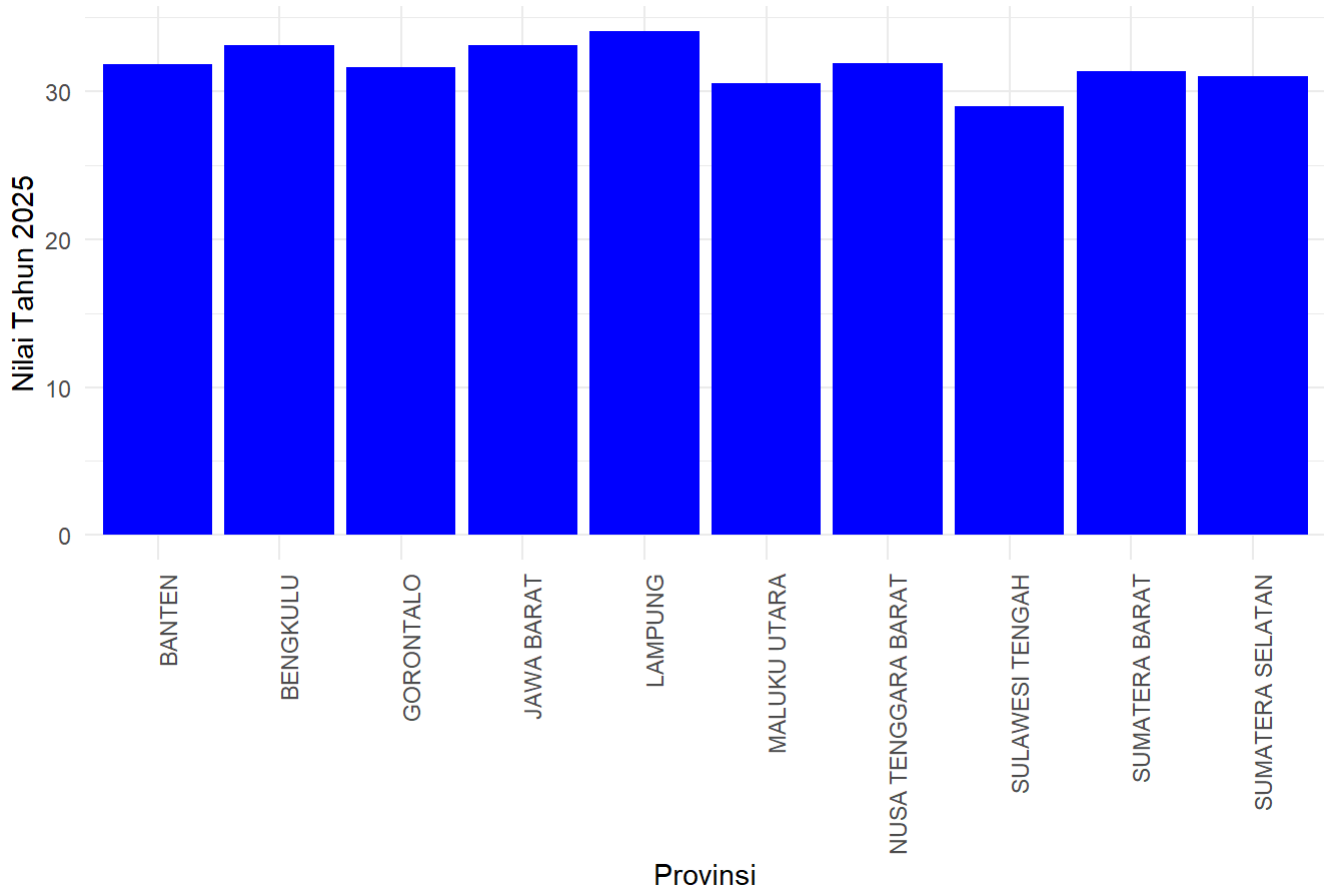

```r
# Menambahkan label berdasarkan rata-rata tertinggi
summary_cluster <- data_long %>%
  group_by(Cluster) %>%
  summarise(
    Rata_Rata = mean(Persentase, na.rm = TRUE),
    Median = median(Persentase, na.rm = TRUE),
    Min = min(Persentase, na.rm = TRUE),
    Max = max(Persentase, na.rm = TRUE),
    Jumlah_Provinsi = n_distinct(Provinsi)
  )

# Mengurutkan klaster berdasarkan rata-rata dan memberikan label
summary_cluster <- summary_cluster %>%
  arrange(desc(Rata_Rata)) %>%
  mutate(Cluster_Label = case_when(
    row_number() == 1 ~ "Tinggi",    # Klaster dengan rata-rata tertinggi
    row_number() == 2 ~ "Sedang",    # Klaster dengan rata-rata kedua
    row_number() == 3 ~ "Rendah"     # Klaster dengan rata-rata terendah
  ))

# Gabungkan label ini kembali ke data asli
data_long <- data_long %>%
  left_join(summary_cluster[, c("Cluster", "Cluster_Label")], by = "Cluster")

# Menampilkan hasil
print(summary_cluster)
```

```
## # A tibble: 3 × 7
##   Cluster Rata_Rata Median   Min   Max Jumlah_Provinsi Cluster_Label
##     <int>     <dbl>  <dbl> <dbl> <dbl>           <int> <chr>
## 1       2      32.1   31.9  28.3  36.6              10 Tinggi
## 2       1      28.1   28.0  23.1  32.8              17 Sedang
## 3       3      24.2   24.5  17.9  30.8               7 Rendah
```

Evaluasi Prediksi

```
# Menghitung MAE, MSE, dan RMSE


# Misalkan kita memiliki data aktual untuk tahun 2024, atau menggunakan data yang lebih baru
# Data aktual tahun 2024
data_aktual_2024 <- data_prediksi$`2024`

# Prediksi untuk tahun 2025
data_prediksi_2025 <- data_prediksi$`2025`

# Hitung MAE, MSE, dan RMSE
mae_value <- MAE(data_aktual_2024, data_prediksi_2025)
mse_value <- MSE(data_aktual_2024, data_prediksi_2025)
rmse_value <- RMSE(data_aktual_2024, data_prediksi_2025)

cat("Evaluasi Prediksi untuk Tahun 2025\n")
```

```
## Evaluasi Prediksi untuk Tahun 2025
```

```
cat("MAE: ", mae_value, "\n")
```

```
## MAE:  0.3352941
```

```
cat("MSE: ", mse_value, "\n")
```

```
## MSE:  0.3808036
```

```
cat("RMSE: ", rmse_value, "\n")
```

```
## RMSE:  0.6170929
```

evaluasi K Mean Clustering

```
# Silhouette Score untuk KMeans (seberapa mirip dengan kluster lain)
library(cluster)
library(clusterSim)
```

```
## Warning: package 'clusterSim' was built under R version 4.4.2
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
cat("Shilloutte Score: ")
```

```
## Shilloutte Score:
```

```
silhouette_score <- silhouette(kmeans_model$cluster, dist(data_kmeans))
mean(silhouette_score[, 3])  # Menampilkan nilai rata-rata silhouette score
```

```
## [1] 0.4992467
```

```
cat("\n=> semakin mendekati 1, berarti pemisahan kluster semakin baik")
```

```
##
## => semakin mendekati 1, berarti pemisahan kluster semakin baik
```

```
cat("\n\nInertia Score: ")
```

```
##
##
## Inertia Score:
```

```
# Inertia (Within-cluster sum of squares) (jarak kluster)
kmeans_model$tot.withinss  # Nilai inertia
```

```
## [1] 651.2174
```

```
cat("\n=>semakin rendah nilai inertia, semakin baik pemisahan kluster ")
```

```
##
## =>semakin rendah nilai inertia, semakin baik pemisahan kluster
```

```
db_index <- index.DB(data_kmeans, kmeans_model$cluster, centrotypes = "centroids")
cat("\n\nDavies-Bouldin Index:", db_index$DB)
```

```
## 
## 
## Davies-Bouldin Index: 0.7307665
```

```
cat("\n\n=>nilai DBI semakin mendekati 0, semakin baik pemisahan kluster ")
```

```
## 
## 
## =>nilai DBI semakin mendekati 0, semakin baik pemisahan kluster
```