

Analisis Peramalan Produksi Padi

Ardika

2023-10-31

Load Library/Packages

```
# library manipulasi data  
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.4      v readr      2.1.5  
## v forcats    1.0.0      v stringr   1.5.1  
## v ggplot2    3.5.1      v tibble    3.2.1  
## v lubridate  1.9.3      v tidyr     1.3.1  
## v purrr      1.0.2  
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)  
# library untuk manipulasi data tanggal  
library(lubridate)  
# library untuk menampilkan hasil statistik deskriptif  
library(summarytools)
```

```
##  
## Attaching package: 'summarytools'  
##  
## The following object is masked from 'package:tibble':  
##  
##      view
```

```
# library untuk membaca data dari file Excel (.xlsx)  
library(readxl)  
# library model VAR  
library(vars)
```

```
## Loading required package: MASS  
##  
## Attaching package: 'MASS'  
##
```

```
## The following object is masked from 'package:dplyr':
##
##   select
##
## Loading required package: strucchange
## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
##
## Loading required package: sandwich
##
## Attaching package: 'strucchange'
##
## The following object is masked from 'package:stringr':
##
##   boundary
##
## Loading required package: urca
## Loading required package: lmtest
```

```
# library untuk model Random Forest
library(randomForest)
```

```
## randomForest 4.7-1.1
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
##
## The following object is masked from 'package:dplyr':
##
##   combine
##
## The following object is masked from 'package:ggplot2':
##
##   margin
```

```
# library untuk model SVR
library(e1071)
# library Untuk partisi dataset
library(caret)
```

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

```
# library untuk evaluasi model
library(Metrics)
```

```
##
## Attaching package: 'Metrics'
##
## The following objects are masked from 'package:caret':
##
##   precision, recall
```

```
# library untuk model XGBoost
library(xgboost)
```

```
##
## Attaching package: 'xgboost'
##
## The following object is masked from 'package:dplyr':
##
##   slice
```

```
# library untuk model Prophet
library(prophet)
```

```
## Loading required package: Rcpp
## Loading required package: rlang
##
## Attaching package: 'rlang'
##
## The following object is masked from 'package:Metrics':
##
##   ll
##
## The following objects are masked from 'package:purrr':
##
##   %%, flatten, flatten_chr, flatten_dbl, flatten_int, flatten_lgl,
##   flatten_raw, invoke, splice
```

1. Obtain

a. import dataset

```
# Dataset padi indonesia
indo <- read_excel("dataset/pangan_Indonesia.xlsx")

# Dataset padi sumatera utara
sumut <- read_excel("dataset/padi-sumut_update.xlsx")
```

b. Menampilkan data

- Dataset indonesia

```
# Menampilkan 6 data teratas  
head(indo)
```

```
## # A tibble: 6 x 6  
##   Tahun Provinsi      Jenis_Komoditi Produksi Produktivitas Luas_Panen  
##   <dbl> <chr>          <chr>          <chr>      <chr>      <chr>  
## 1  1993 ACEH          Padi          1299699.00 40.17      323589.00  
## 2  1993 SUMATERA UTARA Padi          2918152.00 38.67      754569.00  
## 3  1993 SUMATERA BARAT Padi          1806424.00 45.80      394412.00  
## 4  1993 RIAU          Padi          436297.00  29.86      146133.00  
## 5  1993 JAMBI          Padi          607529.00  30.46      199431.00  
## 6  1993 SUMATERA SELATAN Padi          1409559.00 32.04      439895.00
```

```
# Menampilkan 6 data terbawah  
tail(indo)
```

```
## # A tibble: 6 x 6  
##   Tahun Provinsi      Jenis_Komoditi Produksi Produktivitas Luas_Panen  
##   <dbl> <chr>          <chr>          <chr>      <chr>      <chr>  
## 1  2022 GORONTALO     Ubi Kayu       1742       215.9      81  
## 2  2022 SULAWESI BARAT Ubi Kayu       27454      229.4     1197  
## 3  2022 MALUKU        Ubi Kayu       62450      278       2246  
## 4  2022 MALUKU UTARA   Ubi Kayu       4863      370.9     131  
## 5  2022 PAPUA BARAT    Ubi Kayu       8665      223.1     388  
## 6  2022 PAPUA          Ubi Kayu       74189      378.6    1960
```

```
# Menampilkan struktur data  
glimpse(indo)
```

```
## Rows: 7,140  
## Columns: 6  
## $ Tahun      <dbl> 1993, 1993, 1993, 1993, 1993, 1993, 1993, 1993, 1993, 1993, 1~  
## $ Provinsi    <chr> "ACEH", "SUMATERA UTARA", "SUMATERA BARAT", "RIAU", "JA~  
## $ Jenis_Komoditi <chr> "Padi", "Padi", "Padi", "Padi", "Padi", "Padi", "Padi", ~  
## $ Produksi    <chr> "1299699.00", "2918152.00", "1806424.00", "436297.00", ~  
## $ Produktivitas <chr> "40.17", "38.67", "45.80", "29.86", "30.46", "32.04", "~  
## $ Luas_Panen   <chr> "323589.00", "754569.00", "394412.00", "146133.00", "19~
```

- Dataset sumatera utara

```
# Menampilkan 6 data teratas  
head(sumut)
```

```
## # A tibble: 6 x 5  
##   Tahun 'Kabupaten Kota' 'Rata-rata produksi' Produksi 'Luas Panen'
```

```
##      <dbl> <chr>          <chr>          <chr>      <chr>
## 1  2010 Asahan            47.81            81685      17087
## 2  2010 Batu Bara         48.52            166063     34224
## 3  2010 Binjai           47.74            19247      4032
## 4  2010 Dairi            40.99            96612     23567
## 5  2010 Deli Serdang     50.32            427104     84875
## 6  2010 Gunungsitoli    40.700000000000003  7387       1815
```

```
# Menampilkan 6 data terbawah
tail(sumut)
```

```
## # A tibble: 6 x 5
##   Tahun 'Kabupaten Kota' 'Rata-rata produksi' Produksi 'Luas Panen'
##   <dbl> <chr>          <chr>          <chr>      <chr>
## 1  2022 Tanjungbalai    50.22            381.29     75.930000000000007
## 2  2022 Tapanuli Selatan 50              103326.65  20801.6500000000001
## 3  2022 Tapanuli Tengah 38.92            47547.98   12219.78
## 4  2022 Tapanuli Utara   46.07            137822.43  29920.67
## 5  2022 Tebing Tinggi   50              2237.85    447.99
## 6  2022 Toba Samosir    62.87            107139.29  17043.740000000002
```

```
# Menampilkan struktur data
glimpse(sumut)
```

```
## Rows: 442
## Columns: 5
## $ Tahun          <dbl> 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2~
## $ 'Kabupaten Kota' <chr> "Asahan", "Batu Bara", "Binjai", "Dairi", "Deli S~
## $ 'Rata-rata produksi' <chr> "47.81", "48.52", "47.74", "40.99", "50.32", "40.~
## $ Produksi       <chr> "81685", "166063", "19247", "96612", "427104", "7~
## $ 'Luas Panen'    <chr> "17087", "34224", "4032", "23567", "84875", "1815~
```

2. Scrub

a. Merubah nama variabel khusus untuk dataset padi sumatera utara

```
# Merubah nama variabel Kabupaten Kota, Rata-rata Produksi, dan Luas Panen
names(sumut)[names(sumut) == "Kabupaten Kota"] <- "Kabupaten_Kota"
names(sumut)[names(sumut) == "Rata-rata produksi"] <- "Rata_rata_produksi"
names(sumut)[names(sumut) == "Luas Panen"] <- "Luas_Panen"
glimpse(sumut)
```

```
## Rows: 442
## Columns: 5
## $ Tahun          <dbl> 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 201~
## $ Kabupaten_Kota <chr> "Asahan", "Batu Bara", "Binjai", "Dairi", "Deli Ser~
## $ Rata_rata_produksi <chr> "47.81", "48.52", "47.74", "40.99", "50.32", "40.70~
## $ Produksi       <chr> "81685", "166063", "19247", "96612", "427104", "738~
## $ Luas_Panen     <chr> "17087", "34224", "4032", "23567", "84875", "1815",~
```

b. Menghasilkan daftar unik dari provinsi dan kabupaten_kota yang tercatat dalam dataset indo dan sumut, tanpa ada duplikat.

```
# Dataset padi indonesia
unique(indo$Provinsi)
```

```
## [1] "ACEH" "SUMATERA UTARA" "SUMATERA BARAT"
## [4] "RIAU" "JAMBI" "SUMATERA SELATAN"
## [7] "BENGKULU" "LAMPUNG" "KEP. BANGKA BELITUNG"
## [10] "KEP. RIAU" "DKI JAKARTA" "JAWA BARAT"
## [13] "JAWA TENGAH" "DI YOGYAKARTA" "JAWA TIMUR"
## [16] "BANTEN" "BALI" "NUSA TENGGARA BARAT"
## [19] "NUSA TENGGARA TIMUR" "KALIMANTAN BARAT" "KALIMANTAN TENGAH"
## [22] "KALIMANTAN SELATAN" "KALIMANTAN TIMUR" "KALIMANTAN UTARA"
## [25] "SULAWESI UTARA" "SULAWESI TENGAH" "SULAWESI SELATAN"
## [28] "SULAWESI TENGGARA" "GORONTALO" "SULAWESI BARAT"
## [31] "MALUKU" "MALUKU UTARA" "PAPUA BARAT"
## [34] "PAPUA"
```

```
# Dataset padi sumatera utara
unique(sumut$Kabupaten_Kota)
```

```
## [1] "Asahan" "Batu Bara" "Binjai"
## [4] "Dairi" "Deli Serdang" "Gunungsitoli"
## [7] "Humbang Hasundutan" "Karo" "Labuanbatu Utara"
## [10] "Labuhan Batu" "Labuhanbatu Selatan" "Langkat"
## [13] "Mandailing Natal" "Medan" "Nias"
## [16] "Nias Barat" "Nias Selatan" "Nias Utara"
## [19] "Padang Lawas" "Padang Lawas Utara" "Padangsidempuan"
## [22] "Pakpak Bharat" "Pematangsiantar" "Samosir"
## [25] "Serdang Bedagai" "Sibolga" "Simalungun"
## [28] "Sumatera Utara" "Tanjungbalai" "Tapanuli Selatan"
## [31] "Tapanuli Tengah" "Tapanuli Utara" "Tebing Tinggi"
## [34] "Toba Samosir"
```

c. Merubah type data

- Dataset padi indonesia

```
# Merubah type data karakter menjadi numeric
indo$Produksi <- as.numeric(indo$Produksi)
indo$Produktivitas <- as.numeric(indo$Produktivitas)
indo$Luas_Panen <- as.numeric(indo$Luas_Panen)
glimpse(indo)
```

```
## Rows: 7,140
## Columns: 6
## $ Tahun      <dbl> 1993, 1993, 1993, 1993, 1993, 1993, 1993, 1993, 1993, 1~
## $ Provinsi   <chr> "ACEH", "SUMATERA UTARA", "SUMATERA BARAT", "RIAU", "JA~
```

```
## $ Jenis_Komoditi <chr> "Padi", "Padi", "Padi", "Padi", "Padi", "Padi", "Padi",~
## $ Produksi <dbl> 1299699, 2918152, 1806424, 436297, 607529, 1409559, 356~
## $ Produktivitas <dbl> 40.17, 38.67, 45.80, 29.86, 30.46, 32.04, 32.49, 38.03,~
## $ Luas_Panen <dbl> 323589, 754569, 394412, 146133, 199431, 439895, 109807,~
```

```
# Merubah type Tahun menjadi Date
indo$Tahun <- make_date(indo$Tahun)
glimpse(indo)
```

```
## Rows: 7,140
## Columns: 6
## $ Tahun <date> 1993-01-01, 1993-01-01, 1993-01-01, 1993-01-01, 1993-0~
## $ Provinsi <chr> "ACEH", "SUMATERA UTARA", "SUMATERA BARAT", "RIAU", "JA~
## $ Jenis_Komoditi <chr> "Padi", "Padi", "Padi", "Padi", "Padi", "Padi", "Padi",~
## $ Produksi <dbl> 1299699, 2918152, 1806424, 436297, 607529, 1409559, 356~
## $ Produktivitas <dbl> 40.17, 38.67, 45.80, 29.86, 30.46, 32.04, 32.49, 38.03,~
## $ Luas_Panen <dbl> 323589, 754569, 394412, 146133, 199431, 439895, 109807,~
```

- Dataset padi sumatera utara

```
# Merubah type data karakter menjadi numeric
sumut$Rata_rata_produksi <- as.numeric(sumut$Rata_rata_produksi)
```

```
## Warning: NAs introduced by coercion
```

```
sumut$Produksi <- as.numeric(sumut$Produksi)
```

```
## Warning: NAs introduced by coercion
```

```
sumut$Luas_Panen <- as.numeric(sumut$Luas_Panen)
```

```
## Warning: NAs introduced by coercion
```

```
glimpse(sumut)
```

```
## Rows: 442
## Columns: 5
## $ Tahun <dbl> 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 201~
## $ Kabupaten_Kota <chr> "Asahan", "Batu Bara", "Binjai", "Dairi", "Deli Ser~
## $ Rata_rata_produksi <dbl> 47.81, 48.52, 47.74, 40.99, 50.32, 40.70, 45.17, 38~
## $ Produksi <dbl> 81685, 166063, 19247, 96612, 427104, 7387, 87205, 8~
## $ Luas_Panen <dbl> 17087, 34224, 4032, 23567, 84875, 1815, 19307, 2320~
```

```
# Merubah type Tahun menjadi Date
sumut$Tahun <- make_date(sumut$Tahun)
glimpse(sumut)
```

```
## Rows: 442
## Columns: 5
## $ Tahun          <date> 2010-01-01, 2010-01-01, 2010-01-01, 2010-01-01, 20~
## $ Kabupaten_Kota <chr> "Asahan", "Batu Bara", "Binjai", "Dairi", "Deli Ser~
## $ Rata_rata_produksi <dbl> 47.81, 48.52, 47.74, 40.99, 50.32, 40.70, 45.17, 38~
## $ Produksi       <dbl> 81685, 166063, 19247, 96612, 427104, 7387, 87205, 8~
## $ Luas_Panen     <dbl> 17087, 34224, 4032, 23567, 84875, 1815, 19307, 2320~
```

d. Menghapus data kosong atau NA's

- Dataset padi indonesia

```
# Menampilkan variabel dengan baris kosong
colSums(is.na(indo))
```

```
##      Tahun      Provinsi Jenis_Komoditi      Produksi Produktivitas
##      0          0          0          906          906
##      Luas_Panen
##      905
```

```
# Menghapus data NA's pada variabel dengan baris kosong
indo_clean <- na.omit(indo)
colSums(is.na(indo_clean))
```

```
##      Tahun      Provinsi Jenis_Komoditi      Produksi Produktivitas
##      0          0          0          0          0
##      Luas_Panen
##      0
```

```
# Menampilkan ringkasan statistik deskriptif
summary(indo_clean)
```

```
##      Tahun      Provinsi      Jenis_Komoditi      Produksi
## Min.   :1993-01-01 Length:6232 Length:6232 Min.   :      1
## 1st Qu.:2002-01-01 Class :character Class :character 1st Qu.:    3945
## Median :2009-01-01 Mode  :character Mode  :character Median :    29790
## Mean   :2008-06-20                      Mean   :  462420
## 3rd Qu.:2016-01-01                      3rd Qu.:  176021
## Max.   :2022-01-01                      Max.   :13633701
## Produktivitas      Luas_Panen
## Min.   : 0.06 Min.   :      1
## 1st Qu.: 11.74 1st Qu.:   1773
## Median : 29.60 Median :   6984
## Mean   : 58.40 Mean   :  89842
## 3rd Qu.: 87.44 3rd Qu.:  46869
## Max.   :483.00 Max.   :2285232
```

- Dataset padi sumatera utara


```
# Menampilkan variabel dengan baris kosong
colSums(is.na(sumut))
```

```
##           Tahun      Kabupaten_Kota Rata_rata_produksi      Produksi
##           0           0              13              14
##      Luas_Panen
##           13
```

```
# Menghapus data NA's pada variabel dengan baris kosong
sumut_clean <- na.omit(sumut)
colSums(is.na(sumut_clean))
```

```
##           Tahun      Kabupaten_Kota Rata_rata_produksi      Produksi
##           0           0              0              0
##      Luas_Panen
##           0
```

```
# Menampilkan ringkasan statistik deskriptif
summary(sumut_clean)
```

```
##      Tahun      Kabupaten_Kota      Rata_rata_produksi      Produksi
## Min.   :2010-01-01 Length:428      Min.   :28.40      Min.   : 258
## 1st Qu.:2013-01-01 Class :character 1st Qu.:42.40      1st Qu.: 19344
## Median :2016-01-01 Mode  :character Median :48.13      Median : 64848
## Mean   :2015-12-27      Mean   :47.89      Mean   : 198171
## 3rd Qu.:2019-01-01      3rd Qu.:52.54      3rd Qu.: 137248
## Max.   :2022-01-01      Max.   :71.00      Max.   :5136186
##      Luas_Panen
## Min.   : 68
## 1st Qu.: 3957
## Median : 13534
## Mean   : 39349
## 3rd Qu.: 27771
## Max.   :988068
```

e. Filter data padi

- Dataset padi indonesia

```
indo_padi <- indo_clean[indo_clean$Jenis_Komoditi == "Padi", ]
indo_padi
```

```
## # A tibble: 930 x 6
##   Tahun      Provinsi      Jenis_Komoditi Produksi Produktivitas Luas_Panen
##   <date>      <chr>      <chr>      <dbl>      <dbl>      <dbl>
## 1 1993-01-01 ACEH      Padi      1299699      40.2      323589
## 2 1993-01-01 SUMATERA UTARA Padi      2918152      38.7      754569
## 3 1993-01-01 SUMATERA BARAT Padi      1806424      45.8      394412
## 4 1993-01-01 RIAU      Padi      436297      29.9      146133
```

```
## 5 1993-01-01 JAMBI Padi 607529 30.5 199431
## 6 1993-01-01 SUMATERA SELATAN Padi 1409559 32.0 439895
## 7 1993-01-01 BENGKULU Padi 356709 32.5 109807
## 8 1993-01-01 LAMPUNG Padi 1646900 38.0 433078
## 9 1993-01-01 DKI JAKARTA Padi 28488 47.9 5945
## 10 1993-01-01 JAWA BARAT Padi 10820862 50.3 2152592
## # i 920 more rows
```

- Dataset padi sumatera utara

```
sumut_padi <- sumut_clean[sumut_clean$Kabupaten_Kota != "Sumatera Utara",]
sumut_padi
```

```
## # A tibble: 415 x 5
##   Tahun      Kabupaten_Kota Rata_rata_produksi Produksi Luas_Panen
##   <date>      <chr>          <dbl>      <dbl>      <dbl>
## 1 2010-01-01 Asahan          47.8      81685      17087
## 2 2010-01-01 Batu Bara        48.5     166063     34224
## 3 2010-01-01 Binjai          47.7      19247       4032
## 4 2010-01-01 Dairi           41.0      96612     23567
## 5 2010-01-01 Deli Serdang     50.3     427104     84875
## 6 2010-01-01 Gunungsitoli     40.7       7387       1815
## 7 2010-01-01 Humbang Hasundutan 45.2      87205     19307
## 8 2010-01-01 Karo            38.5      89302     23203
## 9 2010-01-01 Labuanbatu Utara   48.2     198284     41165
## 10 2010-01-01 Labuhan Batu     48.2     111260     23065
## # i 405 more rows
```

3. Explore

a. Visualisasi Data

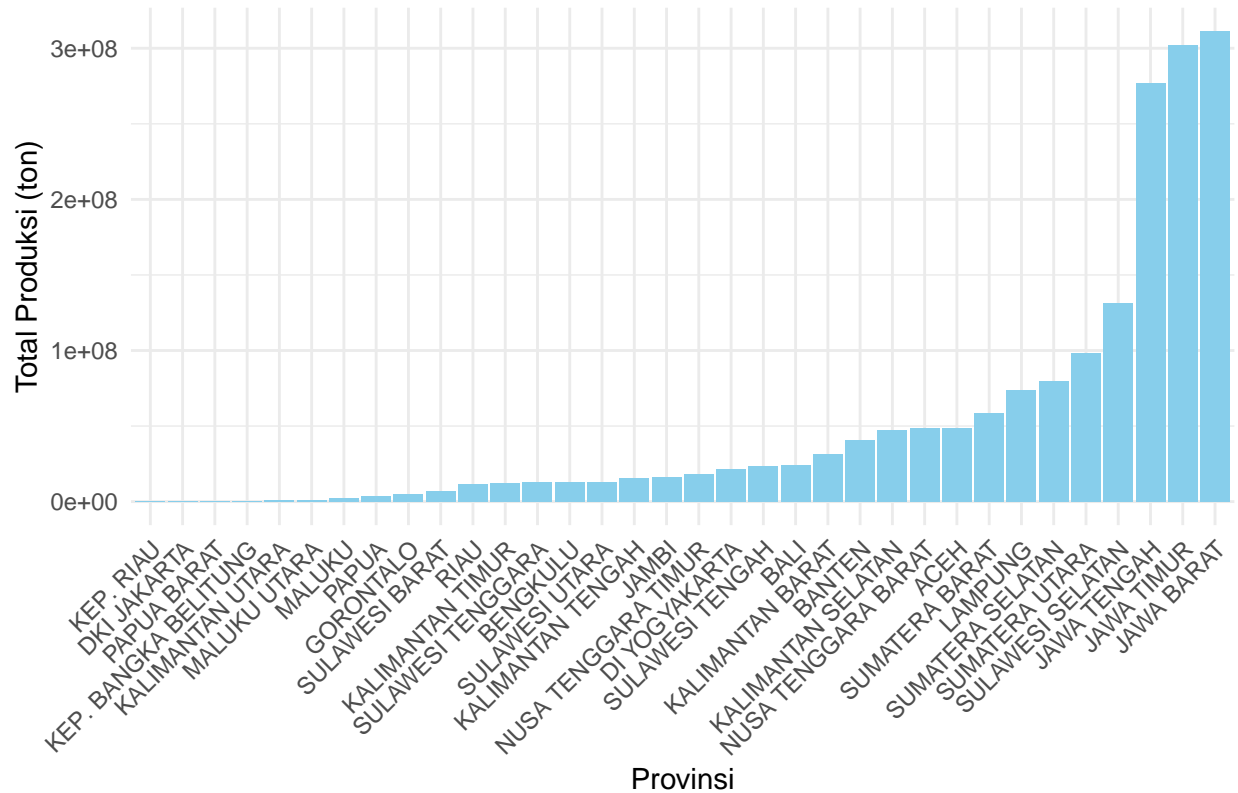
- Dataset padi indonesia

```
# - Jumlah Produksi Padi Per Provinsi (Barplot)

# Hitung total produksi padi per provinsi
total_produksi_provinsi <- indo_padi %>%
  group_by(Provinsi) %>%
  summarise(Total_Produksi = sum(Produksi))

# Visualisasi menggunakan barplot
ggplot(total_produksi_provinsi, aes(x = reorder(Provinsi, Total_Produksi), y = Total_Produksi)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Jumlah Produksi Padi Per Provinsi",
       x = "Provinsi",
       y = "Total Produksi (ton)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Jumlah Produksi Padi Per Provinsi

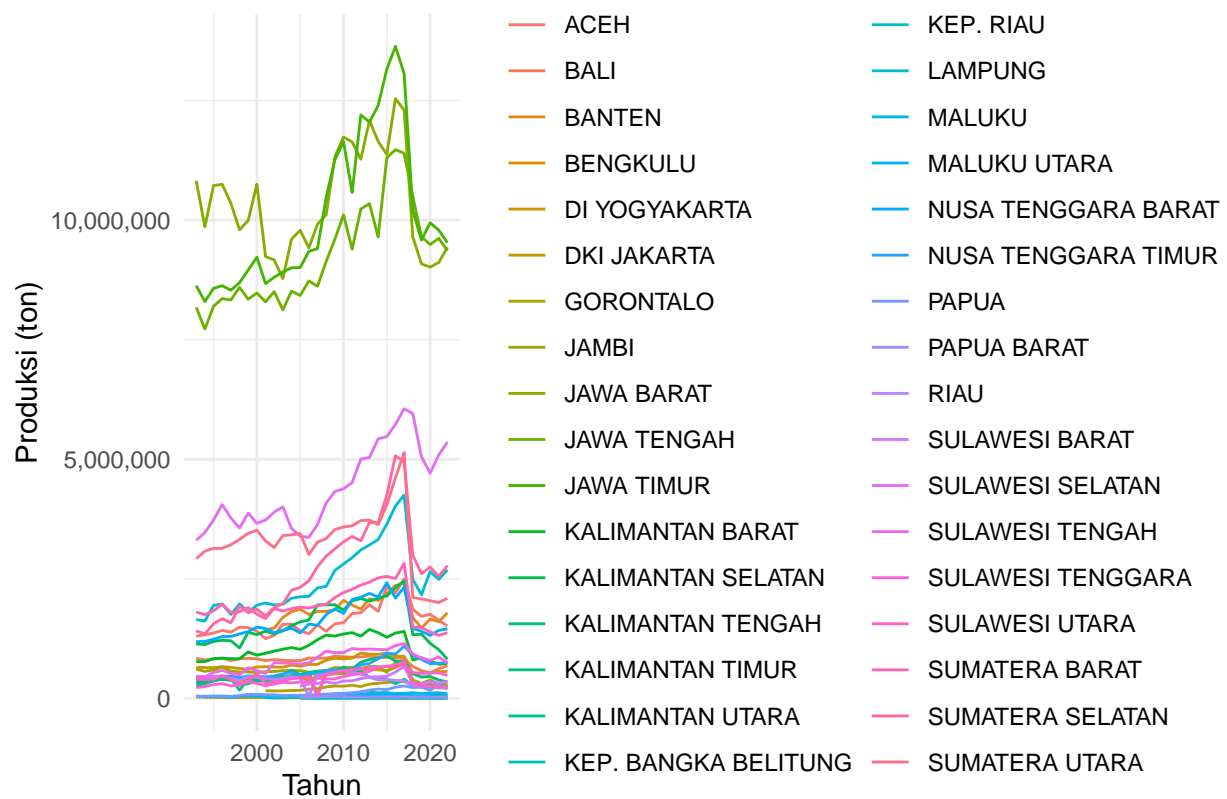


- Tren Produksi Padi Per Provinsi (Grafik Garis)

Visualisasi tren produksi padi per provinsi

```
ggplot(indo_padi, aes(x = Tahun, y = Produksi, color = Provinsi)) +
  geom_line() +
  labs(title = "Tren Produksi Padi Per Provinsi",
        x = "Tahun",
        y = "Produksi (ton)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(legend.position = "right")
```

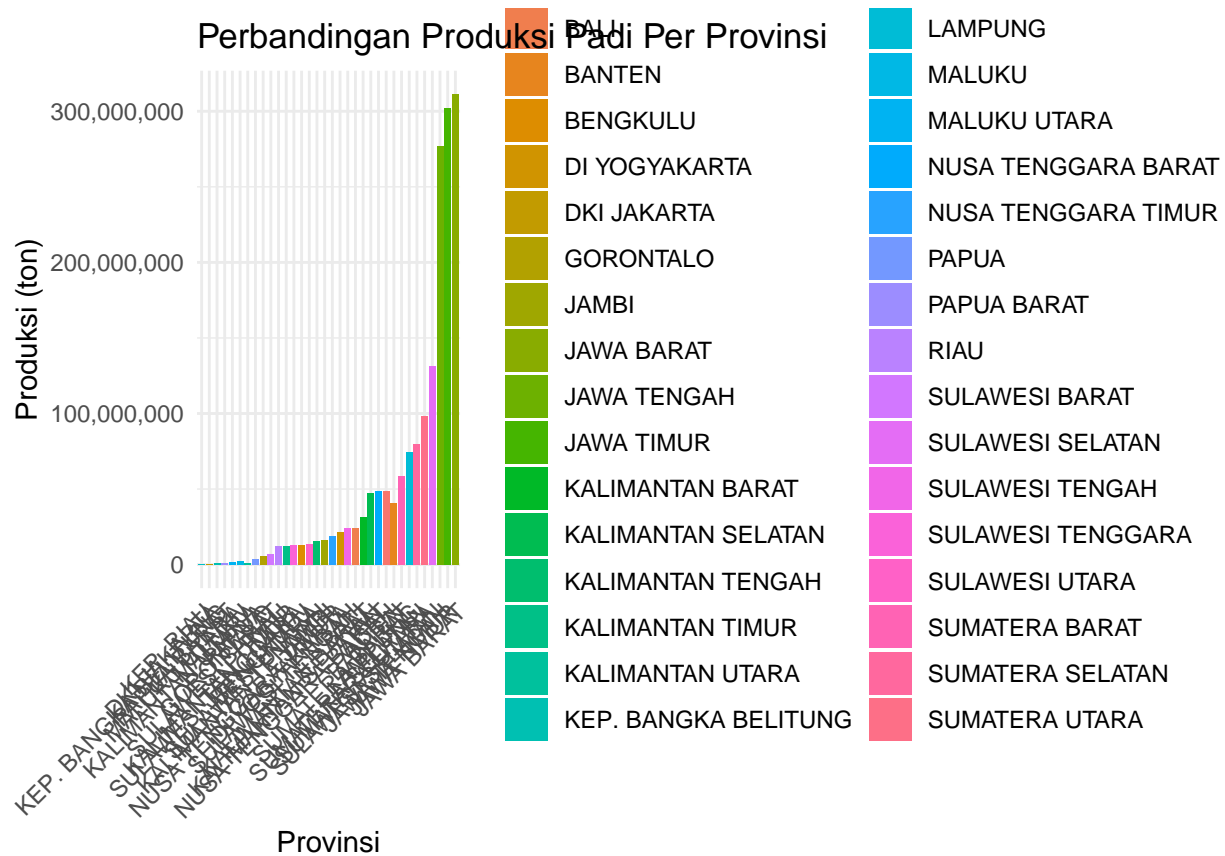
Tren Produksi Padi Per Provinsi



- Perbandingan Produksi Padi Per Provinsi (Barplot)

Visualisasi perbandingan produksi padi per provinsi

```
ggplot(indo_padi, aes(x = reorder(Provinsi, Produksi), y = Produksi, fill = Provinsi)) +
  geom_bar(stat = "identity") +
  labs(title = "Perbandingan Produksi Padi Per Provinsi",
       x = "Provinsi",
       y = "Produksi (ton)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
# - Perbandingan Produktivitas dengan Produksi Padi (Barplot)
```

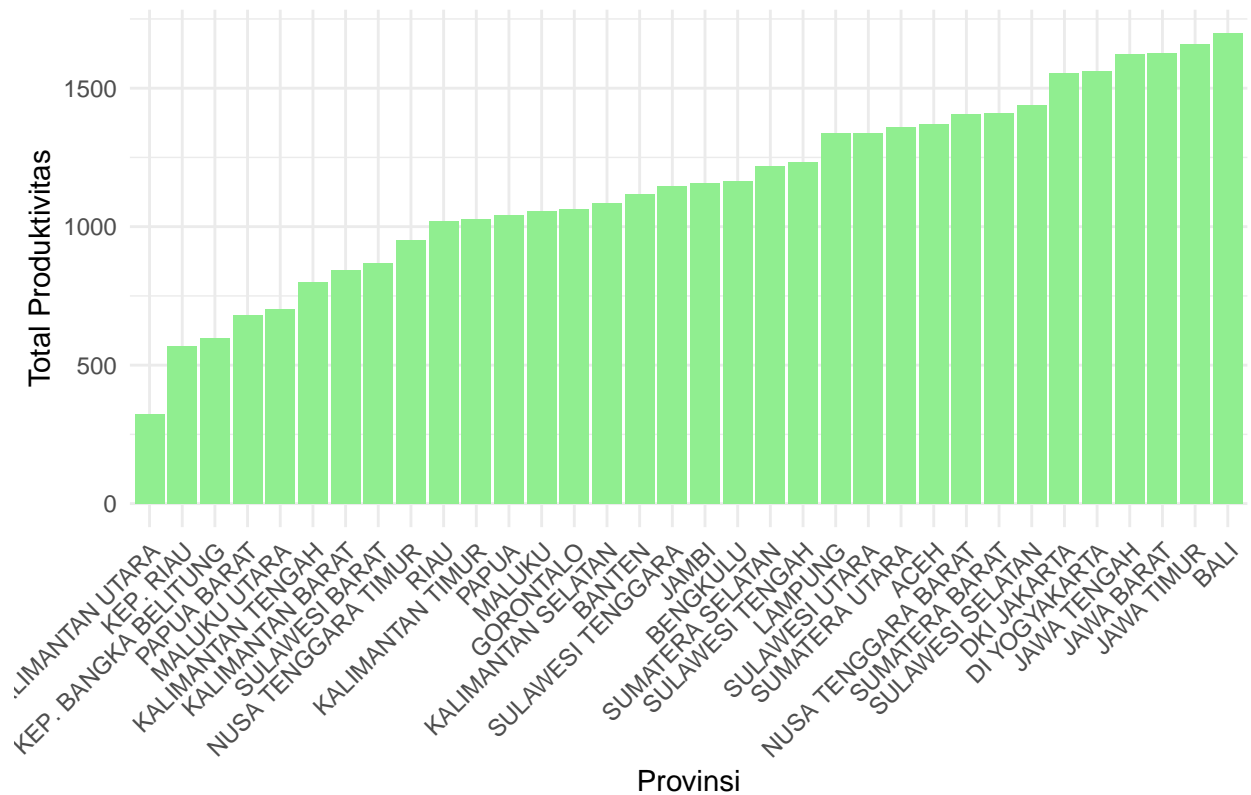
```
# Hitung total produktivitas padi per provinsi
```

```
total_produkktivitas_provinsi <- indo_padi %>%
  group_by(Provinsi) %>%
  summarise(Total_Produktivitas = sum(Produktivitas))
```

```
# Visualisasi perbandingan produktivitas dengan produksi padi
```

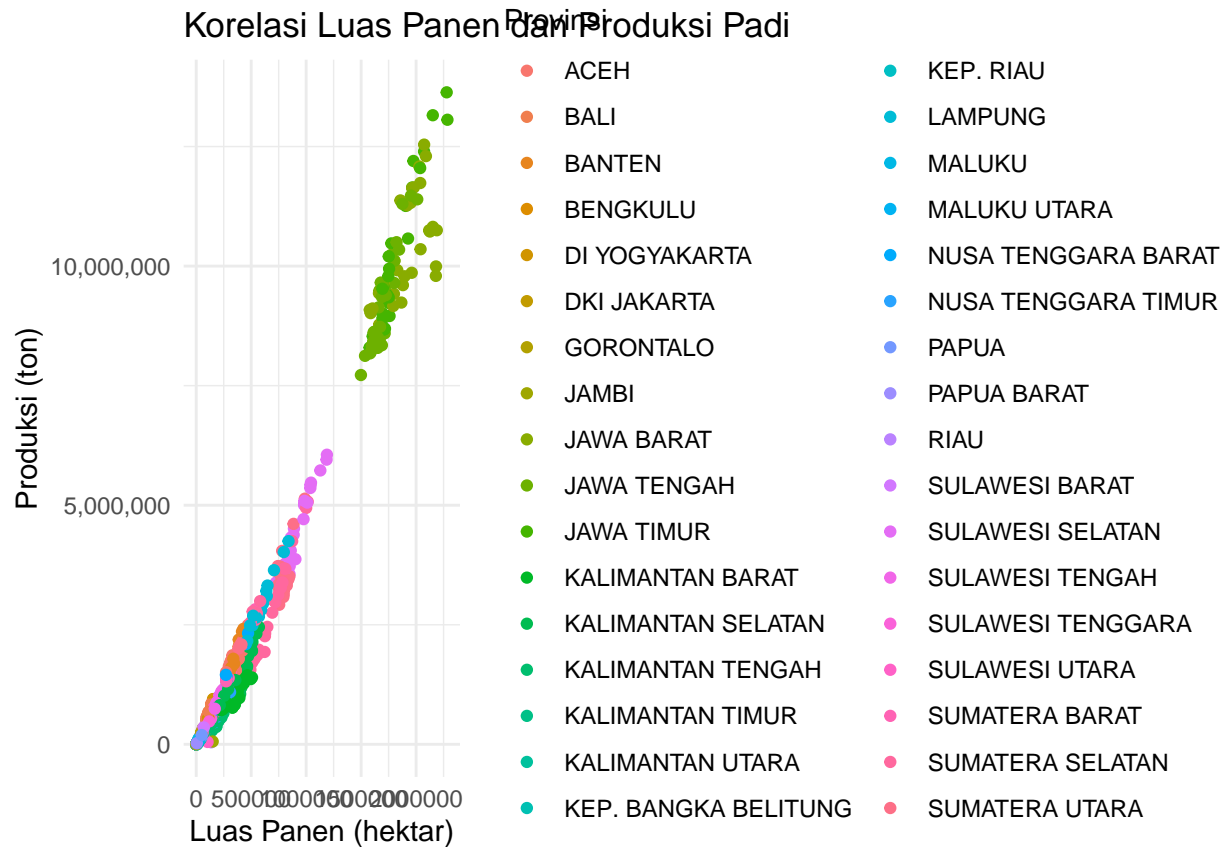
```
ggplot(total_produkktivitas_provinsi, aes(x = reorder(Provinsi, Total_Produktivitas), y = Total_Produktivitas)) +
  geom_bar(stat = "identity", fill = "lightgreen") +
  labs(title = "Perbandingan Produktivitas dengan Produksi Padi",
       x = "Provinsi",
       y = "Total Produktivitas") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Perbandingan Produktivitas dengan Produksi Padi



```
# - Korelasi Luas Panen dan Produksi Padi (Scatterplot)

# Visualisasi korelasi luas panen dan produksi padi
ggplot(indo_padi, aes(x = Luas_Panen, y = Produksi, color = Provinsi)) +
  geom_point() +
  labs(title = "Korelasi Luas Panen dan Produksi Padi",
       x = "Luas Panen (hektar)",
       y = "Produksi (ton)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(legend.position = "right")
```



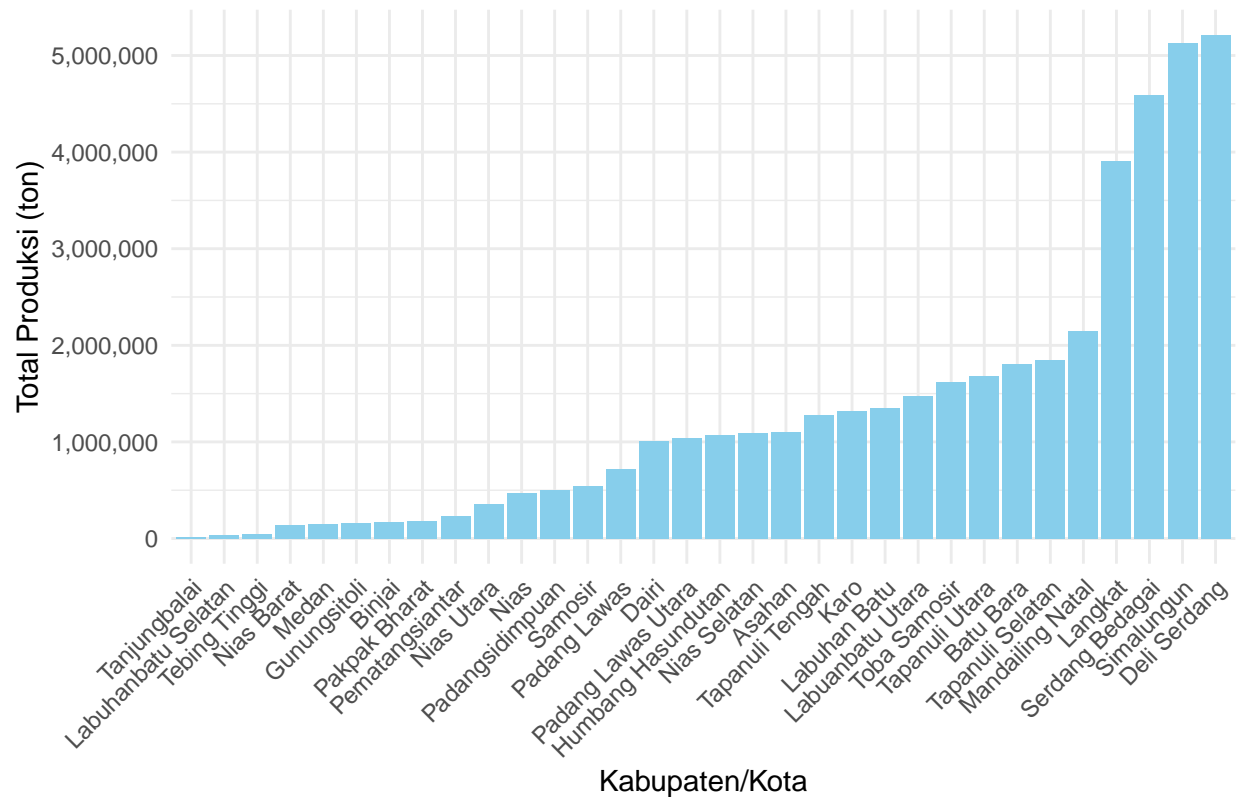
- Dataset padi sumatera utara

```
# - Jumlah Produksi Padi per Kabupaten/Kota (Barplot)

# Hitung total produksi padi per kabupaten/kota
total_produksi_kabkota <- sumut_padi %>%
  group_by(Kabupaten_Kota) %>%
  summarise(Total_Produksi = sum(Produksi))

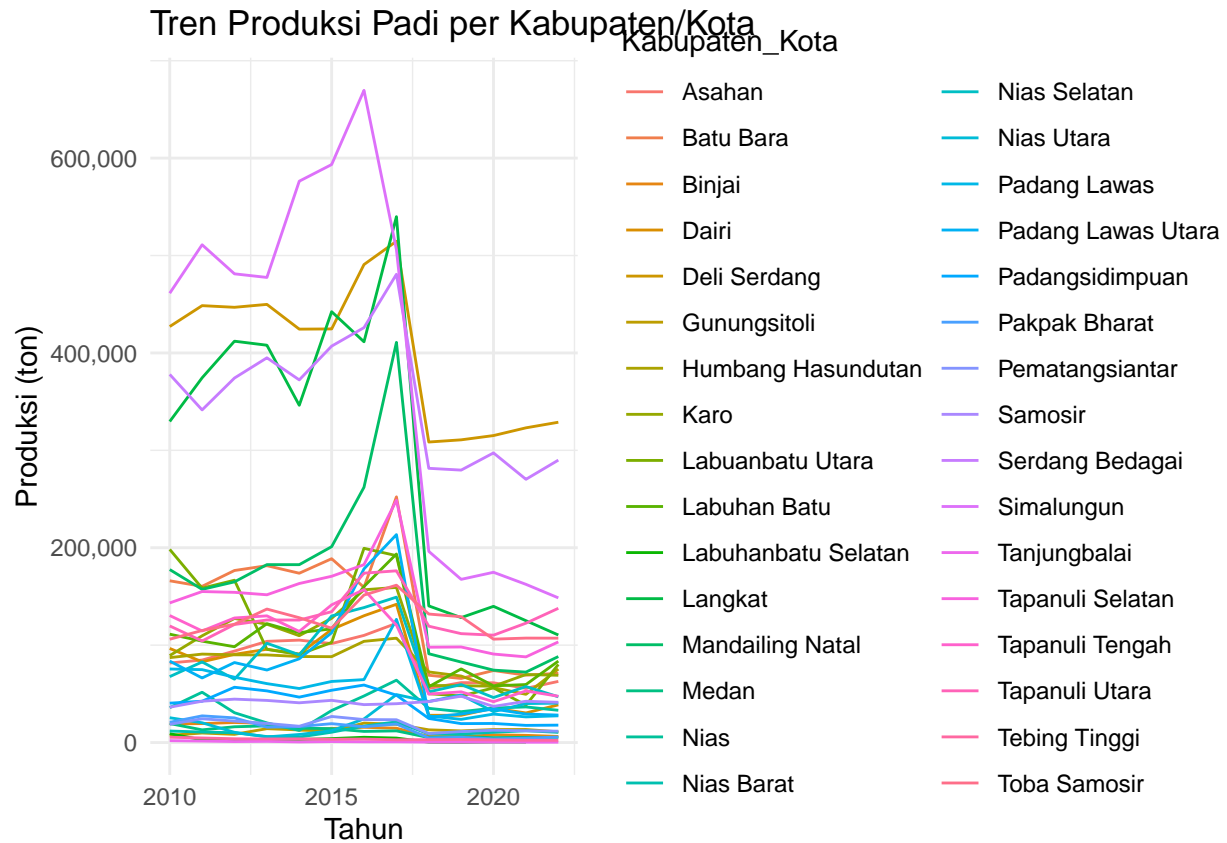
# Visualisasi menggunakan barplot
ggplot(total_produksi_kabkota, aes(x = reorder(Kabupaten_Kota, Total_Produksi), y = Total_Produksi)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Jumlah Produksi Padi per Kabupaten/Kota",
       x = "Kabupaten/Kota",
       y = "Total Produksi (ton)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Jumlah Produksi Padi per Kabupaten/Kota



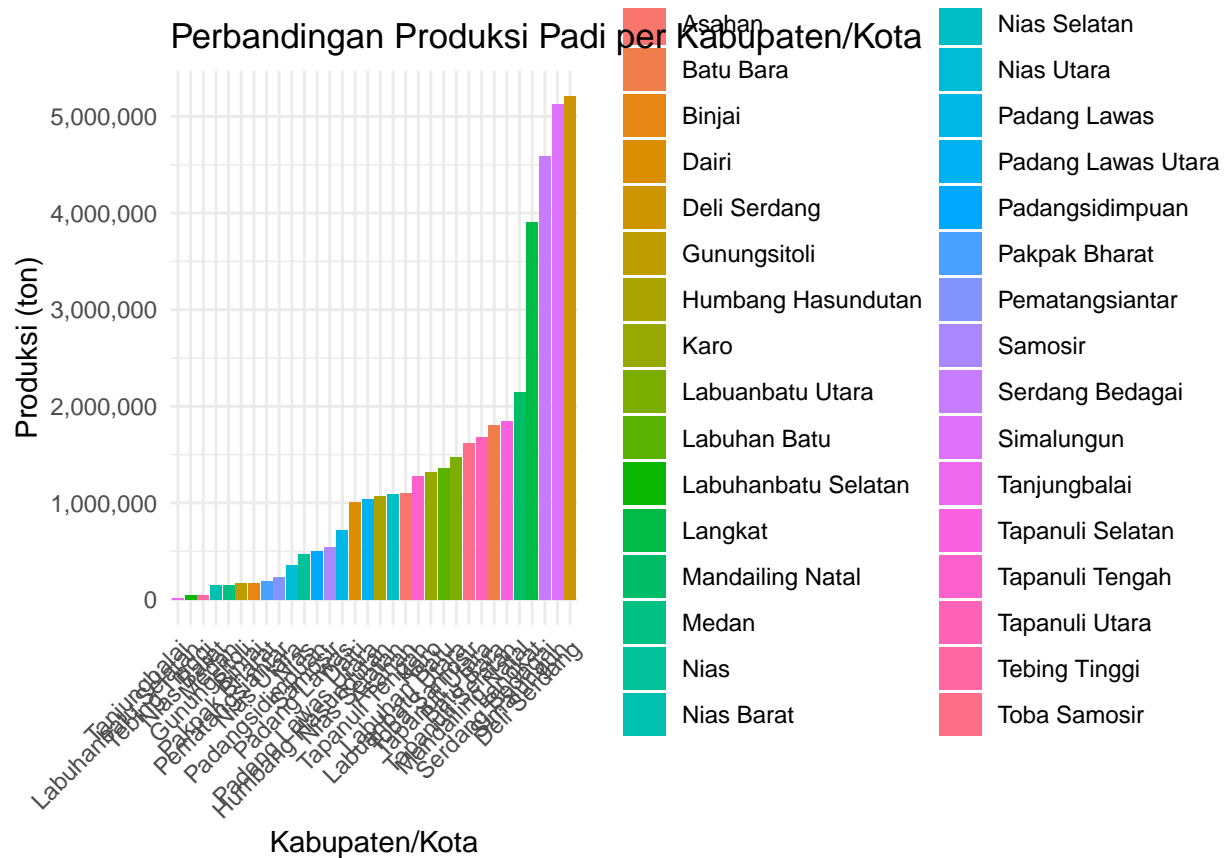
```
# - Tren Produksi Padi per Kabupaten/Kota (Grafik Garis)

# Visualisasi tren produksi padi per kabupaten/kota
ggplot(sumut_padi, aes(x = Tahun, y = Produksi, color = Kabupaten_Kota)) +
  geom_line() +
  labs(title = "Tren Produksi Padi per Kabupaten/Kota",
       x = "Tahun",
       y = "Produksi (ton)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(legend.position = "right")
```

```
# - Perbandingan Produksi Padi per Kabupaten/Kota (Barplot)

# Visualisasi perbandingan produksi padi per kabupaten/kota
ggplot(sumut_padi, aes(x = reorder(Kabupaten_Kota, Produksi), y = Produksi, fill = Kabupaten_Kota)) +
  geom_bar(stat = "identity") +
  labs(title = "Perbandingan Produksi Padi per Kabupaten/Kota",
       x = "Kabupaten/Kota",
       y = "Produksi (ton)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



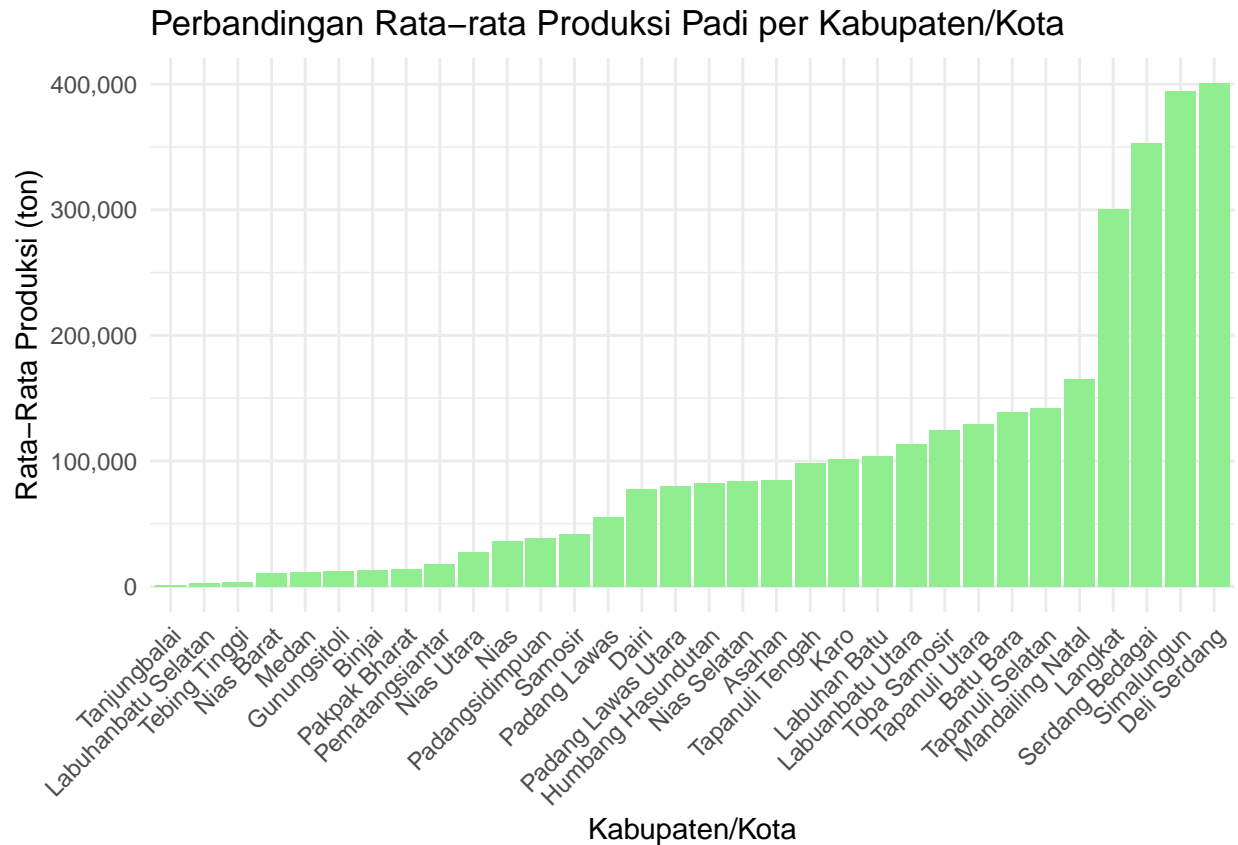
```
# - Perbandingan Rata-rata Produksi Padi per Kabupaten/Kota (Barplot)
```

```
# Hitung rata-rata produksi padi per kabupaten/kota
```

```
rata_rata_produksi_kabkota <- sumut_padi %>%
  group_by(Kabupaten_Kota) %>%
  summarise(Rata_Rata_Produksi = mean(Produksi))
```

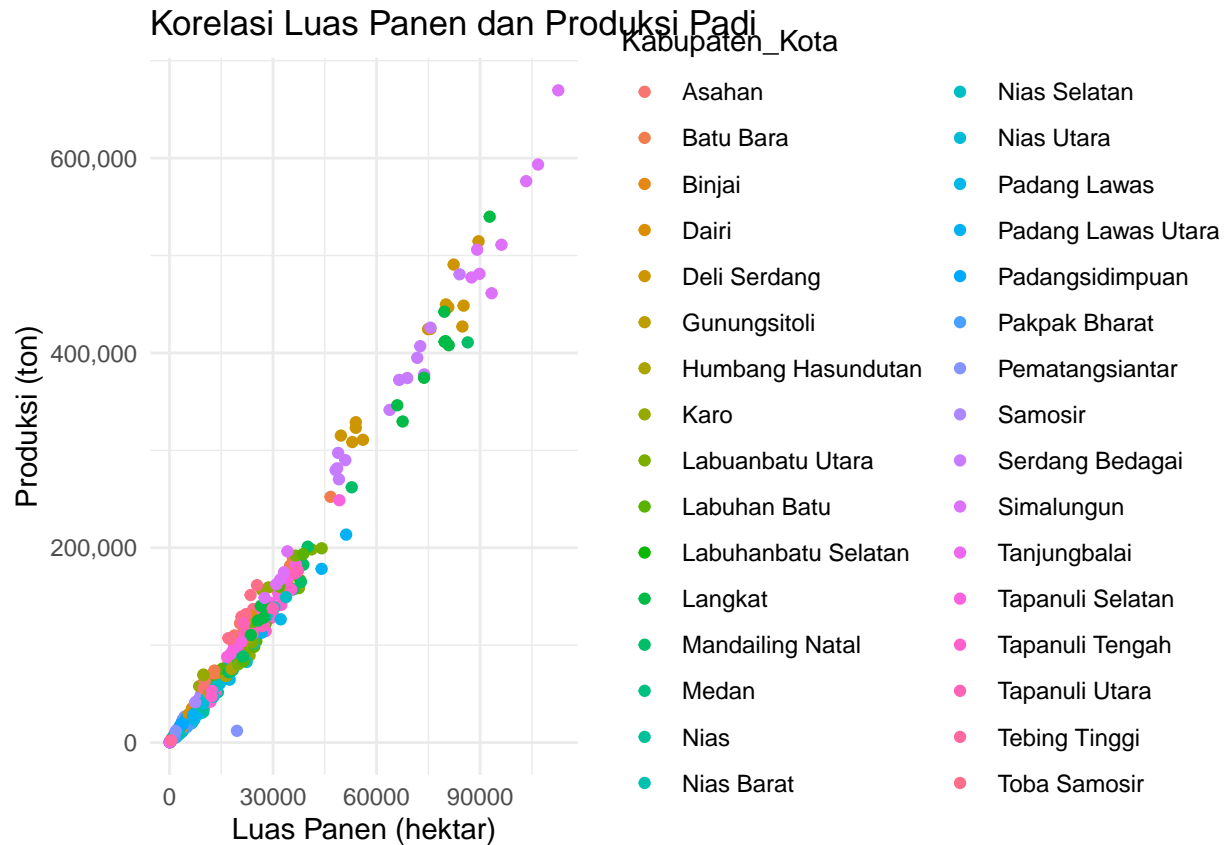
```
# Visualisasi perbandingan rata-rata produksi padi per kabupaten/kota
```

```
ggplot(rata_rata_produksi_kabkota, aes(x = reorder(Kabupaten_Kota, Rata_Rata_Produksi), y = Rata_Rata_Produksi)) +
  geom_bar(stat = "identity", fill = "lightgreen") +
  labs(title = "Perbandingan Rata-rata Produksi Padi per Kabupaten/Kota",
       x = "Kabupaten/Kota",
       y = "Rata-Rata Produksi (ton)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
# - Korelasi Luas Panen dan Produksi Padi (Scatterplot)

# Visualisasi korelasi luas panen dan produksi padi
ggplot(sumut_padi, aes(x = Luas_Panen, y = Produksi, color = Kabupaten_Kota)) +
  geom_point() +
  labs(title = "Korelasi Luas Panen dan Produksi Padi",
       x = "Luas Panen (hektar)",
       y = "Produksi (ton)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(legend.position = "right")
```



b. Analisis Deskriptif

- Hitung Pearson Correlation Coefficient

```
# - Dataset padi indonesia

# Hitung korelasi antara luas panen dan produksi padi
correlation_coefficient <- cor(indo_padi$Luas_Panen, indo_padi$Produksi)

# Nilai correlation coefficient
cat("Koefisien Korelasi Pearson Luas Panen dan Produksi:", correlation_coefficient)
```

```
## Koefisien Korelasi Pearson Luas Panen dan Produksi: 0.9910544
```

```
# - Dataset padi sumatera utara

# Hitung korelasi antara luas panen dan produksi padi
correlation_coefficient <- cor(sumut_padi$Luas_Panen, sumut_padi$Produksi)

# Nilai correlation coefficient
cat("Koefisien Korelasi Pearson Luas Panen dan Produksi:", correlation_coefficient)
```

```
## Koefisien Korelasi Pearson Luas Panen dan Produksi: 0.9919934
```

Berdasarkan grafik diatas bisa lihat bahwa ada korelasi antara luas panen dan produksi. Hubungan korelasi adalah dengan nilai korelasi 0,991. Jadi hubungan luas panen dan produksi padi bisa disimpulkan bahwa luas panen yang besar akan meningkatkan jumlah produksi padi.

4. Model