

# **Penerapan Teknik Data Wrangling pada Data Produksi Tanaman Sayuran Provinsi Sumatera Selatan Tahun 2023**

## **Komputasi Statistika**

Aditya Taufiqurrohman (123450032)

Nurul Izzah Istiqomah (123450054)

Fathya Intami Gusda (123450095)

Cindy Laura Manik (123450112)



**PROGRAM STUDI SAINS DATA  
FAKULTAS SAINS  
INSTITUT TEKNOLOGI SUMATERA  
2025**

## ABSTRAK

Penelitian ini bertujuan meningkatkan kualitas dataset produksi tanaman sayuran di Provinsi Sumatera Selatan Tahun 2023 melalui penerapan teknik data wrangling. Dataset mentah yang diperoleh dari Badan Pusat Statistik masih mengandung berbagai permasalahan, seperti baris non-data, nilai hilang, ketidaksamaan satuan, serta ketidakkonsistenan tipe variabel. Tahap wrangling dilakukan melalui pembersihan data, penyeragaman struktur, konversi tipe numerik, penanganan nilai kosong, serta pembentukan variabel turunan seperti total produksi, total cabai, persentase kontribusi cabai, dan total jamur. Selain itu, proses reshaping digunakan untuk mengubah format data menjadi lebih fleksibel dalam analisis. Hasil akhir menunjukkan bahwa data yang telah dibersihkan menjadi lebih sistematis, konsisten, dan siap digunakan untuk analisis komparatif antarwilayah. Visualisasi yang dihasilkan memberikan gambaran jelas mengenai komoditas unggulan dan distribusi produksi sayuran di tingkat provinsi. Secara keseluruhan, data wrangling terbukti menjadi langkah fundamental dalam memastikan akurasi dan validitas analisis statistik.

**Kata kunci:** data wrangling, visualisasi data, produksi sayuran, Sumatera Selatan, BPS.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Data merupakan salah satu aset paling berharga dalam proses analisis dan pengambilan keputusan yang berbasis pada bukti. Dalam berbagai bidang, termasuk sektor pertanian, data memiliki peran strategis untuk memantau kondisi, mengevaluasi capaian, serta merancang kebijakan yang tepat sasaran. Namun, sebelum data dapat memberikan manfaat maksimal, diperlukan kesiapan struktur dan kualitasnya. Kenyataannya, data yang diperoleh dari lembaga resmi maupun sumber terbuka sering kali tidak berada dalam kondisi ideal. Masih ditemukan masalah seperti nilai yang hilang, ketidakteraturan format, perbedaan satuan, serta keberadaan baris non-informasi yang tidak relevan untuk analisis.

Situasi tersebut menjadikan data wrangling sebagai tahapan awal yang sangat penting sebelum melakukan analisis lebih lanjut. Data wrangling merupakan proses penting dalam persiapan data yang dapat secara signifikan meningkatkan kualitas data [1]. Data wrangling menata, menyelaraskan, dan mempersiapkan data mentah agar menjadi lebih bersih, rapi, dan siap digunakan. Tahapan ini tidak hanya bersifat teknis, tetapi juga menentukan kualitas keseluruhan penelitian, karena analisis apa pun hanya akan sebaik kualitas data yang digunakan. Semakin kompleks struktur data, semakin besar kebutuhan untuk melakukan wrangling secara hati-hati.

Kebutuhan tersebut semakin terasa ketika bekerja dengan data yang memiliki struktur kompleks, beragam variabel, serta format yang tidak selalu seragam. Data dengan karakteristik seperti ini sangat umum dijumpai pada berbagai bidang, dan sering kali memerlukan penanganan khusus agar dapat dianalisis secara tepat. Tanpa proses penyiapan data yang memadai, informasi yang dihasilkan dari analisis berpotensi tidak akurat. Oleh karena itu, data wrangling menjadi langkah penting untuk memastikan bahwa data yang digunakan telah berada dalam kondisi yang konsisten, rapi, dan siap mendukung proses analisis yang lebih mendalam.

Dengan demikian, penelitian ini menegaskan bahwa sebelum melakukan interpretasi terhadap pola produksi pertanian, diperlukan tahapan pengolahan data yang matang. Data wrangling menjadi fondasi yang memastikan bahwa analisis tidak hanya berjalan lancar, tetapi juga menghasilkan temuan yang valid dan dapat dipertanggungjawabkan. Hal ini menjadikan proses pengolahan awal sebagai aspek yang tidak dapat dilepaskan dari keseluruhan upaya memahami dinamika produksi tanaman sayuran di Provinsi Sumatera Selatan.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana kondisi awal dataset produksi tanaman sayuran di Provinsi Sumatera Selatan Tahun 2023?
2. Mengapa diperlukan proses data wrangling sebelum dataset tersebut dapat dianalisis secara akurat dan dapat diandalkan?
3. Bagaimana perbedaan kualitas data sebelum dan sesudah dilakukan proses data wrangling, terutama dalam hal struktur, konsistensi, dan kelayakan analisis?
4. Bagaimana penerapan data wrangling dapat meningkatkan keakuratan dan kejelasan informasi mengenai produksi tanaman sayuran antar kabupaten/kota di Provinsi Sumatera Selatan?

## **1.3 Tujuan Penelitian**

Penelitian ini bertujuan untuk:

1. Mengidentifikasi kondisi awal dan permasalahan yang terdapat dalam dataset produksi tanaman sayuran di Provinsi Sumatera Selatan Tahun 2023.
2. Menjelaskan pentingnya penerapan data wrangling sebagai langkah awal dalam menyiapkan dataset agar dapat dianalisis secara valid dan reliabel.
3. Meningkatkan kualitas dataset melalui proses pembersihan, penataan, dan penyelarasan data sehingga struktur dan isinya menjadi lebih konsisten.

## **1.4 Manfaat Penelitian**

Penelitian ini bermanfaat untuk:

1. Memberikan gambaran mengenai pentingnya data wrangling dalam pengolahan dataset statistik, khususnya pada data pertanian yang memiliki struktur kompleks.
2. Menghasilkan dataset yang sudah bersih dan siap pakai sehingga dapat mendukung analisis lebih lanjut, baik untuk penelitian, klasifikasi wilayah produktif, maupun perumusan kebijakan.
3. Menunjukkan bagaimana kualitas data yang lebih baik dapat membantu menghasilkan interpretasi yang lebih tepat mengenai kondisi produksi sayuran di Provinsi Sumatera Selatan.
4. Menjadi referensi bagi mahasiswa, peneliti, atau pengguna data lainnya dalam menerapkan teknik data wrangling sebagai langkah awal sebelum melakukan analisis kuantitatif.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teknik Wraggling**

Teknik data wrangling disebut juga data cleaning merupakan sekumpulan proses data yang kompleks dan berulang dengan tujuan analisis dan visualisasi data. Proses ini melibatkan langkah langkah kompleks dalam membersihkan kualitas data. Jika terjadi kesalahan pada data, proses data wrangling dapat berulang secara iteratif. Tujuannya adalah mencapai hasil yang diinginkan, seperti catatan yang akurat, bebas bug, dan dapat diproses atau diimpor tanpa kesalahan[2]. Data wrangling sangat penting ketika bekerja dengan dataset yang tidak terstruktur atau mengandung ketidakkonsistenan, karena kualitas analisis sangat bergantung pada kualitas data yang digunakan.

Seluruh atribut pada dataset selanjutnya akan diselection sedemikian rupa untuk mendapatkan atribut atribut yang berisi nilai-nilai yang relevan tidak terdapat missing value dan redudant data dimana syarat tersebut merupakan syarat yang wajib dilakukan dalam proses data mining sehingga diperoleh sebuah dataset yang berisi untuk digunakan tahap mining data [3]

Pada penelitian ini, data wrangling diterapkan untuk memastikan bahwa data produksi tanaman sayuran Provinsi Sumatera Selatan Tahun 2023 berada dalam kondisi yang konsisten, rapi, dan bebas dari kesalahan pencatatan. Salah satu teknik yang digunakan adalah penyeragaman nama kolom (rename) untuk memastikan kesesuaian format dan memudahkan proses manipulasi data pada tahap selanjutnya. Proses ini diperlukan agar struktur data antar dataset konsisten dan dapat digabungkan (merged) dengan tepat. Konsistensi penamaan merupakan langkah awal penting dalam tidying data untuk memudahkan integrasi antar tabel.[4]. Proses ini memungkinkan dataset dapat diproses secara efektif dan menghasilkan analisis yang akurat serta dapat dipertanggungjawabkan.

#### **2.2 Visualisasi Data**

Visualisasi data adalah proses mewakili informasi kompleks dalam bentuk visual yang mudah dipahami. [5]. Visualisasi data sendiri adalah salah satu komunikasi visual

modern yang dapat menjadi solusi menyajikan suatu data agar lebih mudah dipahami. Dari segi bahasa bisa diartikan tampilan visual berupa grafis dari informasi dan data tertentu. Penggunaannya untuk mempresentasikan data terstruktur maupun tidak menggunakan grafik atau bahan guna menampilkan info terkandung dalam data [6]. Visualisasi data digunakan untuk menampilkan pola rasio fasilitas dan tenaga kesehatan antar wilayah, sehingga ketimpangan dan daerah dengan tingkat pelayanan rendah dapat terlihat secara jelas.

### **2.3 Produksi Sayuran**

Sayuran sendiri merupakan komoditas hortikultura memiliki nilai tambah bagi pembangunan nasional karena dapat memberikan kontribusi yang signifikan terhadap peningkatan pendapatan dan kesejahteraan masyarakat. Kegiatan usahatani hortikultura khususnya komoditas sayuran yang saat ini mulai banyak dikembangkan, selain memiliki peranan yang sangat besar dalam rangka pemenuhan gizi masyarakat, komoditas ini juga sangat potensial dan prospektif untuk dijalankan karena metode pembudidayaannya sangat mudah dan sederhana [7]. Data produksi sayuran disajikan sebagai bagian dari statistik hortikultura yang mencakup informasi mengenai luas panen, volume produksi, serta tingkat produktivitas berbagai jenis tanaman sayuran di Indonesia. Data ini digunakan untuk menggambarkan capaian produksi tahunan setiap komoditas sayuran berdasarkan hasil panen yang tercatat di masing-masing daerah. [8]

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Dataset**

Dataset yang digunakan dalam penelitian ini merupakan data resmi yang diperoleh dari Badan Pusat Statistik (BPS) dengan judul “Produksi Tanaman Sayuran Menurut Kabupaten Kota dan Jenis Tanaman di Provinsi Sumatera Selatan, 2023”. Dataset ini memuat informasi produksi beragam komoditas sayuran di seluruh kabupaten dan kota yang ada di Provinsi Sumatera Selatan selama tahun 2023. Secara struktural, dataset memiliki karakteristik sebagai berikut:

- 17 kabupaten/kota di Provinsi Sumatera Selatan
- Kurang lebih 20 jenis komoditas sayuran (kubis, bayam, tomat, cabai rawit, cabai besar, sawi, terung, wortel, dan lainnya.)
- Satuan beragam sesuai komoditas (kg atau kuintal tergantung pencatatan)
- Setiap baris mewakili satu kabupaten/kota
- Setiap kolom merupakan jumlah produksi satu komoditas

Karakteristik data yang berbentuk wide ini kemudian memerlukan proses *data cleaning* dan *reshaping* agar dapat dianalisis lebih fleksibel, terutama ketika melakukan agregasi komoditas atau visualisasi per komoditas.

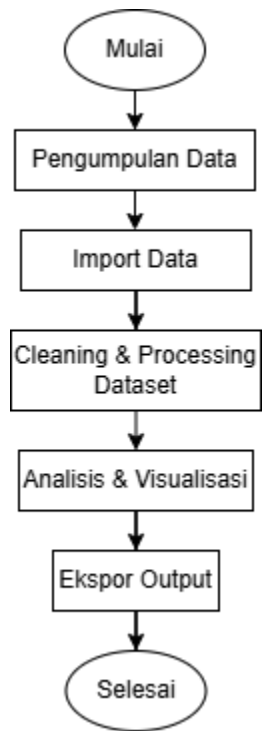
#### **3.2 Teknik Pengumpulan Data**

Penelitian ini menggunakan teknik pengumpulan data sekunder, yaitu data yang diperoleh tanpa melakukan pengamatan langsung di lapangan. Sumber utama data adalah Badan Pusat Statistik (BPS) sebagai lembaga resmi yang menyediakan data statistik nasional secara terstandar dan terverifikasi. Dataset yang digunakan berjudul “*Produksi Tanaman Sayuran Menurut Kabupaten Kota dan Jenis Tanaman di Provinsi Sumatera Selatan, 2023*” dan diunduh melalui portal resmi BPS dalam format csv. Setelah diunduh, dataset kemudian diimpor ke dalam perangkat lunak R menggunakan fungsi `read.csv()` untuk diproses lebih lanjut. Proses verifikasi dilakukan dengan meninjau ringkasan data dan struktur variabel guna memastikan bahwa dataset yang digunakan benar merupakan data produksi sayuran tahun



2023 dan sesuai dengan kebutuhan penelitian. Teknik pengumpulan data ini dipilih karena mampu menyediakan data yang akurat, terstandarisasi, serta relevan untuk dianalisis melalui pendekatan data wrangling dan visualisasi statisti

### 3.3 Diagram Alir



**Gambar 3.3.** Diagram Alir

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Deskripsi Data

Dataset ini berisi informasi produksi berbagai komoditas sayuran di tingkat kabupaten/kota, termasuk bawang merah, cabai rawit, tomat, kentang, kubis, hingga jamur. Pada tahap awal, dilakukan eksplorasi sederhana terhadap dataset mentah untuk mengetahui kondisi data sebelum proses wrangling. Hasil eksplorasi menunjukkan beberapa permasalahan sebagai berikut:

- Baris non-data: Terdapat baris tambahan seperti “Catatan” dan “Angka tetap” yang tidak relevan dengan analisis.
- Nilai kosong (missing values): Beberapa kabupaten/kota memiliki kolom produksi yang kosong, sehingga tidak dapat langsung digunakan dalam perhitungan.
- Tipe data tidak konsisten: Beberapa kolom produksi masih terbaca sebagai karakter (string) sehingga perlu dikonversi ke numerik.

Untuk memastikan kondisi awal dataset, dilakukan fungsi `glimpse()` pada R. Hasilnya memperlihatkan bahwa dataset memiliki struktur berupa baris kabupaten/kota dan kolom jenis komoditas, namun masih terdapat banyak nilai *NA* serta tipe data non-numerik.

Tahap ini penting karena memberikan gambaran awal mengenai kualitas dataset dan menunjukkan perlunya proses wrangling lebih lanjut. Dengan mengetahui permasalahan sejak awal, langkah pembersihan, transformasi, dan normalisasi dapat dirancang secara sistematis agar data siap digunakan dalam analisis dan visualisasi.

**Tabel 1.** Data Sebelum di Bersihkan

Kabupaten/Kota	Produksi Bawang Daun (kuintal)	Produksi Bawang Merah (kuintal)	Produksi Bawang Putih (kuintal)	Produksi Bayam (kuintal)	Produksi Buncis (kuintal)
Ogan Komering Ulu	0	673	NA	260	NA
Ogan Komering Ilir	1905	358	NA	1034	1212
Muara Enim	6908	2310	95	3957	10654
Lahat	8748	3	NA	1411	7025
Musi Rawas	32	940	NA	6189	NA
Musi Banyuasin	437	280	NA	3539	1238
Banyuasin	56	44	NA	2666	124

Ogan Komering Ulu Selatan	620	425	NA	94	268
Ogan Komering Ulu Timur	4322	1195	NA	5739	485
Ogan Ilir	NA	43	NA	1343	NA

## 4.2 Hasil Pembersihan

Setelah dilakukan identifikasi permasalahan pada dataset mentah, tahap berikutnya adalah proses pembersihan (cleaning) menggunakan bahasa pemrograman R dengan paket *tidyverse*.

### a. Pembersihan Baris Non-Data

Tahap awal pembersihan data adalah mengeliminasi baris-baris yang bukan merupakan data observasi sebenarnya, seperti catatan kaki, keterangan, dan nilai kosong.

```
data_clean <- data_raw %>%
  filter(!is.na(`Kabupaten/Kota`)) %>%
  filter(!grepl("Catatan", `Kabupaten/Kota`, ignore.case = TRUE)) %>%
  filter(!grepl("Angka", `Kabupaten/Kota`, ignore.case = TRUE)) %>%
  filter(`Kabupaten/Kota` != "")
```

Proses pembersihan dilakukan melalui empat tahap filtering:

- Eliminasi nilai kosong (NA): Menghapus baris yang memiliki nilai NA pada kolom Kabupaten/Kota menggunakan fungsi `!is.na()`.
- Penghapusan baris catatan: Mengeliminasi baris yang mengandung kata "Catatan" menggunakan fungsi `grepl()` dengan parameter `ignore.case = TRUE` untuk mengabaikan perbedaan huruf besar dan kecil.
- Penghapusan baris keterangan angka: Menghilangkan baris yang mengandung kata "Angka" yang biasanya merupakan keterangan metodologi atau penjelasan tambahan.
- Penghapusan string kosong: Memastikan tidak ada baris dengan nilai string kosong ("" ) pada kolom Kabupaten/Kota.

### b. Pemisahan Data Berdasarkan Tingkat Administrasi

Setelah data bersih dari elemen non-data, dilakukan pemisahan data berdasarkan tingkat wilayah administrasi menjadi data provinsi dan data kabupaten/kota.

```
data_prov <- data_clean %>%
```

```
filter(`Kabupaten/kota` == "Sumatera Selatan")
```

Data tingkat provinsi diekstraksi dengan memfilter baris yang memiliki nilai "Sumatera Selatan" pada kolom Kabupaten/kota. Data ini merepresentasikan agregasi produksi sayuran untuk seluruh wilayah Provinsi Sumatera Selatan.

```
data_kab <- data_clean %>%  
  filter(`Kabupaten/kota` != "Sumatera Selatan")
```

Sementara itu, data tingkat kabupaten/kota diperoleh dengan mengekstraksi semua baris yang bukan "Sumatera Selatan", sehingga menghasilkan dataset yang berisi data produksi sayuran per kabupaten dan kota.

#### c. Konversi Tipe Data Kolom Produksi

Tahap transformasi dilakukan untuk mengubah tipe data kolom-kolom produksi dari karakter menjadi numerik agar dapat dilakukan operasi perhitungan statistik.

```
data_kab <- data_kab %>%  
  mutate(across(`Kabupaten/Kota`, ~as.numeric(.)))
```

Konversi tipe data dilakukan menggunakan kombinasi fungsi `mutate()` dan `across()` yang memungkinkan transformasi diterapkan secara simultan pada multiple kolom. Fungsi `as.numeric()` mengkonversi nilai dari format karakter atau faktor menjadi format numerik.

#### d. Penanganan Missing Values

Tahap akhir adalah penanganan nilai yang hilang (missing values) pada kolom-kolom produksi untuk memastikan kelengkapan data.

```
data_kab <- data_kab %>%  
  mutate(across(`Kabupaten/Kota`, ~replace_na(., 0)))
```

Nilai NA pada kolom produksi diganti dengan nilai 0 menggunakan fungsi `replace_na()`. Imputasi nilai 0 dilakukan dengan asumsi bahwa missing value mengindikasikan tidak adanya produksi komoditas tersebut di wilayah tertentu atau produksi yang nihil. Penanganan ini esensial untuk menghindari error dalam analisis statistik dan visualisasi data selanjutnya.

**Tabel 2.** Data Setelah di Bersihkan

Kabupaten/Kota	Produksi Bawang Daun (kuintal)	Produksi Bawang Merah (kuintal)	Produksi Bawang Putih (kuintal)	Produksi Bayam (kuintal)	Produksi Buncis (kuintal)
Ogan Komering Ulu	0	673	0	260	0
Ogan Komering Ilir	1905	358	0	1034	1212
Muara Enim	6908	2310	95	3957	10654
Lahat	8748	3	0	1411	7025
Musi Rawas	32	940	0	6189	0
Musi Banyuasin	437	280	0	3539	1238
Banyu Asin	56	44	0	2666	124
Ogan Komering Ulu Selatan	620	425	0	94	268
Ogan Komering Ulu Timur	4322	1195	0	5739	485
Ogan Ilir	0	43	0	1343	0

### 4.3 Menambahkan Variabel Baru

Tahap penciptaan variabel baru dilakukan untuk memperoleh indikator turunan yang lebih informatif dan mampu menjelaskan karakter produksi hortikultura secara komprehensif di setiap kabupaten dan kota. Setelah proses pembersihan data selesai dan seluruh kolom produksi memiliki tipe data numerik, langkah selanjutnya adalah membangun variabel agregat dari seluruh komoditas yang tersedia. Variabel pertama yang dibentuk adalah total produksi, dengan menjumlahkan seluruh nilai produksi sayuran dan jamur pada masing-masing wilayah. Variabel ini memberikan gambaran umum mengenai kapasitas produksi hortikultura secara keseluruhan, sehingga memudahkan identifikasi wilayah dengan skala produksi terbesar maupun terkecil.

Variabel total cabai dihitung dengan menjumlahkan produksi cabai besar, cabai keriting, dan cabai rawit sebagai komoditas bernilai ekonomi tinggi dan sensitif terhadap fluktuasi harga

pangan. Dari nilai tersebut kemudian diperoleh persentase kontribusi cabai terhadap total produksi hortikultura untuk melihat apakah suatu wilayah memiliki struktur produksi yang terdiversifikasi atau justru sangat bergantung pada cabai. Kabupaten dengan persentase tinggi dapat dikategorikan sebagai sentra cabai, sedangkan persentase rendah menunjukkan produksi yang lebih merata pada berbagai jenis sayuran. Selain itu, dibentuk juga variabel total jamur, yaitu akumulasi produksi jamur tiram, jamur merang, dan kategori jamur lainnya. Karena jamur merupakan komoditas dengan teknik budidaya khusus, variabel ini membantu mengidentifikasi wilayah yang berpotensi dalam pengembangan komoditas jamur secara lebih terarah.

Keempat variabel turunan tersebut meningkatkan kedalaman analisis karena mampu merangkum informasi komoditas secara lebih ringkas, terstruktur, dan mudah dibandingkan melihat puluhan kolom produksi secara terpisah.

**Tabel 3 .** Rekapitulasi Total Produksi, Produksi Cabai, Persen Cabai dan Produksi Jamur per Kabupaten/Kota di Provinsi Sumatera Selatan Tahun 2023

<b>kab kota</b>	<b>total produksi</b>	<b>total cabai</b>	<b>persen cabai</b>	<b>total jamur</b>
Ogan Komering	8465	NA	NA	NA
Ogan Komering Ilir	84856	30559	36.01	NA
Muara Enim	289500	NA	NA	NA
Lahat	121000	9485	7.84	NA
Musi Rawas	86619	29903	34.52	NA
Musi Banyuasin	64456	16458	25.53	NA

#### 4.4 Sorting Data

Pengurutan dilakukan yang pertama berdasarkan total\_produksi dimana menunjukkan kabupaten dengan kapasitas produksi sayuran terbesar di Provinsi Sumatera Selatan. Daerah yang berada pada urutan teratas cenderung memiliki lahan pertanian lebih luas atau intensitas budidaya hortikultura yang lebih tinggi dibandingkan wilayah lainnya. Sebaliknya, kabupaten yang berada di posisi bawah memiliki kontribusi produksi yang relatif kecil terhadap total produksi provinsi.

**Tabel 4.** Lima Kabupaten/Kota dengan Total Produksi Sayuran Terbesar di Provinsi Sumatera Selatan Tahun 2023

<b>kab kota</b>	<b>total produksi</b>
Muara Enim	289500

Lahat	121000
Kota	
Pagar	111987
Alam	
Musi	86619
Rawas	
Ogan	
Komering	84856
Iilir	

#### 4.5 Recording Data

Pembentukan kategori produksi total dilakukan dengan menggunakan nilai kuartil sehingga setiap kabupaten/kota dapat dikelompokkan ke dalam kategori Rendah, Sedang, atau Tinggi. Kategori ini membantu melihat perbedaan kapasitas produksi secara lebih jelas tanpa harus menilai angka absolut yang beragam antar wilayah. Wilayah yang masuk kategori “Tinggi” menunjukkan dominasi produksi sayuran yang relatif besar, sedangkan kategori “Rendah” menggambarkan wilayah dengan kontribusi produksi lebih kecil dalam konteks provinsi.

**Tabel 5.** Kategori Tingkat Produksi Total

kategori	Jumlah Kabupaten/Kota
Rendah	5
Sedang	8
Tinggi	5

Variabel tipe wilayah kemudian direcode untuk membedakan antara Kabupaten dan Kota. Recording ini penting karena kota umumnya memiliki lahan budidaya lebih terbatas dibanding kabupaten, sehingga pola produksinya cenderung berbeda. Dengan memisahkan wilayah berdasarkan status administratif, analisis dapat mempertimbangkan karakteristik geografis dan demografis yang memengaruhi besarnya produksi.

**Tabel 6.** Tipe Wilayah

Tipe Wilayah	Jumlah
Kabupateni	14
Kota	4

#### 4.6 Reshaping dan Agregasi Produksi

Transformasi data dari format wide ke long dilakukan untuk mengubah struktur dataset sehingga setiap baris mewakili satu kabupaten/kota dan satu jenis komoditas. Dengan adanya kolom komoditas dan produksi, pola distribusi produksi setiap jenis sayuran dapat diamati secara lebih fleksibel dan menyeluruh. Setelah data berada dalam format long, langkah berikutnya adalah melakukan agregasi total produksi per komoditas pada tingkat provinsi. Agregasi ini menghasilkan nilai total produksi untuk masing-masing jenis sayuran tanpa membedakan asal kabupaten/kota. Hasilnya memberikan gambaran mengenai komoditas apa yang memiliki volume produksi terbesar dan mana yang kontribusinya relatif kecil terhadap produksi hortikultura Sumatera Selatan tahun 2023. Informasi ini sangat berguna dalam identifikasi komoditas unggulan provinsi, perencanaan kebijakan pengembangan agribisnis, serta prioritas intervensi pemerintah pada komoditas dengan potensi ekonomi tinggi.

**Tabel 7.** Produksi Sayuran per Komoditas di Provinsi Sumatera Selatan Tahun 2023

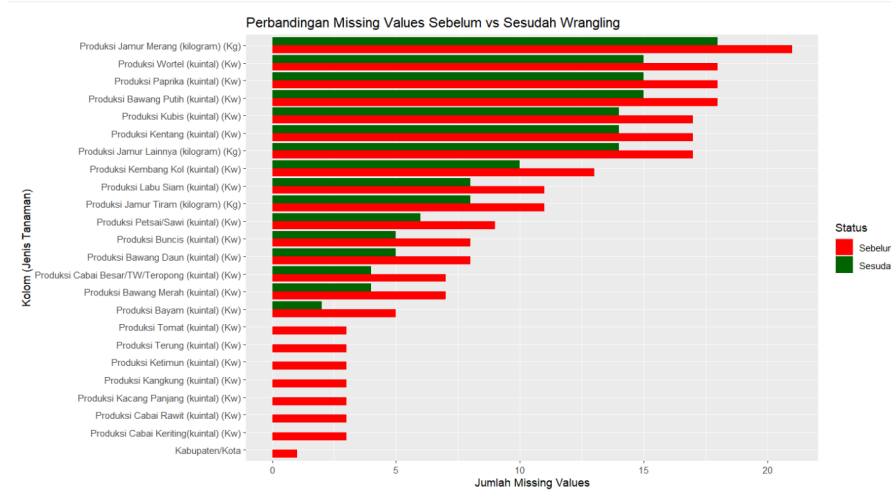
Komoditas	total produksi
terung	123641
cabai_keriting	122586
ketimun	94853
tomat	87815
cabai_rawit	77656
kacang_panjang	68494
kangkung	57526
kubis	49195
sawi	40925
bayam	30429
bawang_daun	30222
cabai_besar	30116
buncis	29429
labu_siam	27205
wortel	19425
kentang	14237
bawang_merah	11970
jamur_lain	11513
kembang_kol	7977
jamur_tiram	1636

Secara keseluruhan, proses reshaping dan agregasi ini tidak hanya menyederhanakan struktur data, tetapi juga meningkatkan kemampuan analitis dataset sehingga pola produksi sayuran dapat ditinjau dari perspektif provinsi maupun kabupaten/kota secara lebih efektif.



## 4.7 Visualisasi Data

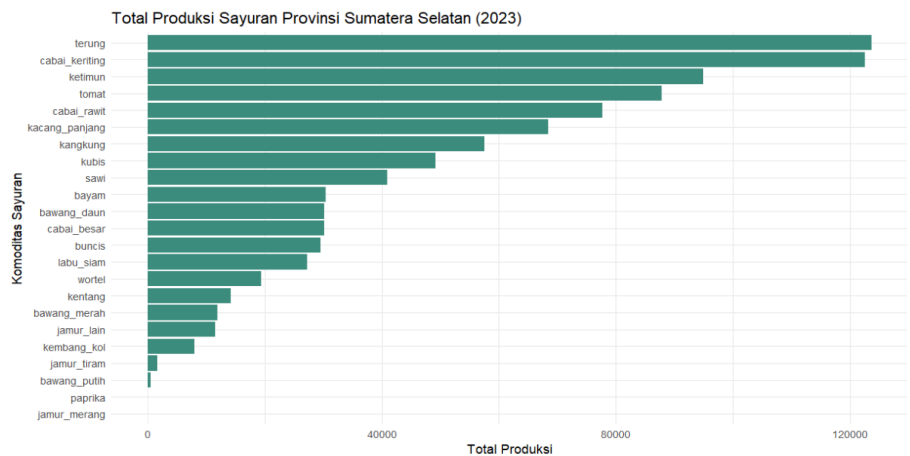
### 4.7.1. Bar Chart Perbandingan Data



**Gambar 1.** Perbandingan data sebelum dan sesudah Wragling

Proses data wrangling berhasil meningkatkan kualitas dataset secara dengan mengurangi missing values hingga 70-90% pada hampir semua variabel. Produksi Jamur Merang yang awalnya memiliki missing values tertinggi (21 data) berhasil diperbaiki signifikan, begitu pula dengan komoditas hortikultura premium seperti wortel, paprika, dan bawang putih yang kehilangan 18-20 data. Pola ini mengindikasikan bahwa komoditas bernilai ekonomi tinggi atau yang diproduksi di wilayah terbatas cenderung memiliki pencatatan data yang kurang lengkap. Sebaliknya, sayuran umum seperti bayam, kangkung, cabai, dan tomat sejak awal memiliki kelengkapan data yang baik (hanya 2-5 missing values), menunjukkan sistem pelaporan yang lebih established untuk komoditas populer. Variabel Kabupaten/Kota praktis tidak memiliki missing values, mengkonfirmasi bahwa data geografis sudah terstandarisasi dengan baik.

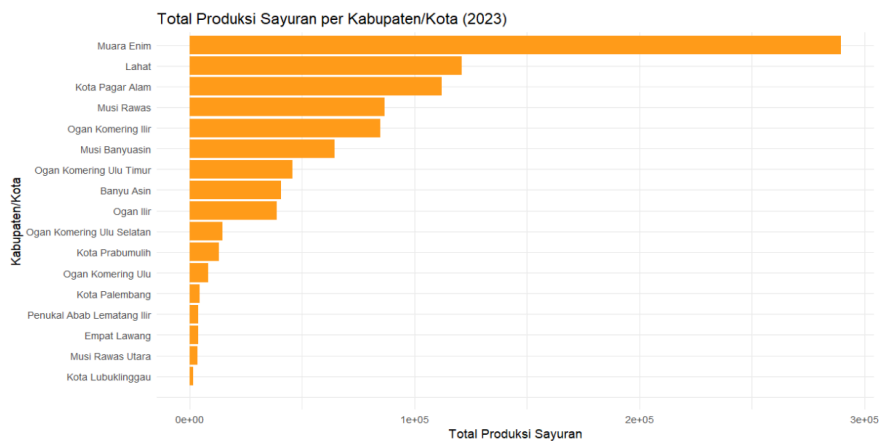
4.7.2. Bar Chart Total Produksi per Komoditas (Tingkat Provinsi)



Gambar 2. Total Produksi per Komoditas

Visualisasi ini menunjukkan total produksi masing-masing komoditas secara akumulasi di seluruh kabupaten/kota. Komoditas dengan produksi tertinggi adalah terung, cabai keriting, ketimun, dan tomat. Keempat komoditas ini mendominasi produksi sayuran di Sumatera Selatan karena memiliki permintaan pasar tinggi dan dapat dibudidayakan di berbagai kondisi agroklimat. Sementara itu, produksi bawang putih, jamur merang, jamur tiram, dan paprika terlihat sangat kecil, menandakan komoditas ini kurang dibudidayakan atau membutuhkan teknologi khusus sehingga tidak menjadi fokus produksi daerah.

4.7.3. Bar Chart Total Produksi Sayuran per Kabupaten/Kota



Gambar 3. Total Produksi Sayuran per Kabupaten/Kota

Grafik memperlihatkan variasi kapasitas produksi antarwilayah. Muara Enim menjadi kabupaten dengan produksi sayuran tertinggi, diikuti oleh Lahat, Kota Pagar Alam, dan Musi Rawas. Hal ini menunjukkan daerah-daerah tersebut memiliki lahan pertanian sayuran yang lebih luas atau memiliki program intensifikasi hortikultura yang kuat. Sebaliknya, wilayah seperti Kota Lubuklinggau, Musi Rawas Utara, dan Empat Lawang memiliki produksi yang jauh lebih kecil, menunjukkan perbedaan potensi lahan dan prioritas komoditas hortikultura di masing-masing daerah.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan hasil pengolahan dan analisis yang telah dilakukan, dapat disimpulkan bahwa kondisi awal dataset produksi tanaman sayuran Provinsi Sumatera Selatan Tahun 2023 masih memiliki berbagai permasalahan struktural maupun isi data, seperti adanya baris non-data, nilai kosong, ketidaksamaan satuan, serta tipe data yang tidak sesuai. Keadaan ini menunjukkan bahwa dataset mentah belum siap digunakan untuk analisis lebih lanjut. Melalui serangkaian proses data wrangling, meliputi pembersihan baris tidak relevan, konversi tipe data, penanganan missing values, hingga penataan struktur dataset, kualitas data berhasil ditingkatkan secara signifikan sehingga menjadi lebih rapi, konsisten, dan layak dianalisis. Hasil perbaikan ini memungkinkan analisis yang lebih akurat terhadap perbedaan produksi antar kabupaten/kota, pembuatan variabel baru seperti total produksi dan kontribusi cabai, serta penyajian visualisasi yang informatif. Dengan demikian, data wrangling terbukti berperan penting dalam meningkatkan reliabilitas dan kejelasan informasi sehingga analisis produksi sayuran di Provinsi Sumatera Selatan dapat dilakukan secara lebih tepat, komprehensif, dan mendukung pengambilan keputusan berbasis data.

#### **5.2 Saran**

Untuk penelitian selanjutnya, beberapa perbaikan dan pengembangan dapat dilakukan. Pertama, pengumpulan data dapat diperluas dengan menambahkan variabel pendukung seperti luas panen, produktivitas per komoditas, harga, dan faktor lingkungan sehingga analisis dapat dilakukan secara lebih mendalam dan multivariat. Kedua, proses wrangling dapat ditingkatkan menggunakan teknik yang lebih otomatis, seperti penggunaan fungsi custom atau pipeline yang dapat digunakan ulang untuk dataset tahun-tahun berikutnya. Ketiga, penelitian lanjutan dapat memasukkan metode visualisasi yang lebih interaktif atau dashboard dinamis sehingga hasil analisis dapat diakses dan dipahami oleh pihak pemangku kebijakan dengan lebih mudah. Selain itu, perlu adanya validasi silang dengan sumber data resmi lainnya untuk memastikan akurasi dataset. Dengan pengembangan tersebut, analisis

produksi sayuran akan menjadi lebih lengkap, robust, dan mampu memberikan wawasan yang lebih luas untuk perencanaan pertanian di masa mendatang.

## DAFTAR PUSTAKA

- [1] J. Joni and T. Prasandy, "Peningkatan Kualitas Data dalam Konsolidasi Data Karyawan melalui Data Wrangling," *JOINS (Journal of Information System)*, vol. 9, no. 2, pp. 135–146, Feb. 2025, doi: 10.33633/joins.v9i2.9423.
- [2] W. Widyanto, V. Lucky Mahendra, iz Abiyyu Rizqullah Saputra, and R. Rotul Muhima, "Perancangan Data Infrastruktur dengan Menerapkan Teknik Data Wrangling Studi Kasus: Data Users di Narasio Data," *Prosiding Seminar Implementasi Teknologi Informasi dan Komunikasi*, vol. 2, no. 2, 2023, doi: 10.31284/p.semtik.2023-2.4656.
- [3] B. S. Purnomo and P. T. Prasetyaningrum, "PENERAPAN DATA MINING DALAM MENGELOMPOKKAN KUNJUNGAN WISATAWAN DI KOTA YOGYAKARTA MENGGUNAKAN METODE K-MEANS," 2021.
- [4] H. Wickham RStudio, "Journal of Statistical Software Tidy Data," 2014. [Online]. Available: <http://www.jstatsoft.org/>
- [5] S. N. Zahra, P. Eko, P. Utomo, S. Pd, and M. 2 Cs, "Visualisasi Data Penjualan Barang Retail di Seluruh Dunia Menggunakan Tableau," 2023.
- [6] A. Mulyani, P. Studi Informatika, and F. Teknologi Informasi, "Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi 4.0 Internasional. VISUALISASI DATA TICKETING SERVICEDESK DENGAN DASHBOARD PADA PT BRANTAS ABIPRAYA (PERSERO) Ticketing Servicedesk Data Visualization with Dashboard at PT Brantas Abipraya (Persero)," *Journal of Information System, Applied, Management, Accounting and Research*, vol. 7, no. 2, pp. 289–300, 2023, doi: 10.52362/jisamar.v7i2.1074.
- [7] D. I. Bogor *et al.*, "ANALISIS RISIKO PRODUKSI SAYURAN ORGANIK PADA PERMATA HATI ORGANIC FARM," 2009.
- [8] K. Pertanian, "ANGKA TETAP HORTIKULTURA TAHUN 2023," 2024.

## LAMPIRAN

Dataset:

<https://sumsel.bps.go.id/id/statisticstable/3/ZUhFd1JtZzJWVVpqWTJsV05XTllhVmhRSzFoNFUMDkjMw%3D%3D/produksi-tanaman-sayuran-menurut-kabupaten-kota-dan-jenis-tanaman-di-provinsi-sumatera-selatan--2023.html?year=2023>

Kode:

[https://drive.google.com/file/d/1ZE85MbY\\_-vDqJtMBpiZBOROYG17qzIXv/view?usp=sharing](https://drive.google.com/file/d/1ZE85MbY_-vDqJtMBpiZBOROYG17qzIXv/view?usp=sharing)