

**PENERAPAN TEKNIK WRANGLING DALAM ANALISIS STATISTIK JUMLAH
PEGAWAI NEGERI SIPIL SUMATERA BAGIAN SELATAN 2024**
KOMPUTASI STATISTIK

Kelompok 5

Anggota:

- | | |
|-----------------------------|-----------|
| 1. Keren Marito Lumban Gaol | 123450020 |
| 2. Rosalia Siregar | 123450036 |
| 3. Enggli Rahmadhani | 123450043 |
| 4. Akeyla Fairuz Shafi | 123450119 |



PROGRAM STUDI SAINS DATA FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
2025

Abstrak

Penelitian ini bertujuan untuk menerapkan teknik data wrangling pada data jumlah Pegawai Negeri Sipil (PNS) di wilayah Sumatera bagian Selatan tahun 2024 guna meningkatkan kualitas data sehingga lebih terstruktur dan mudah dianalisis. Data yang digunakan merupakan data sekunder dari publikasi Badan Pusat Statistik (BPS) dan dianalisis menggunakan pendekatan komputasi statistik dengan jenis penelitian kuantitatif deskriptif. Proses data wrangling dilakukan melalui enam tahapan, yaitu menciptakan variabel baru, subsetting, sorting, recoding, merging, dan reshaping untuk memperbaiki inkonsistensi, menambah informasi, serta mengubah format data sesuai kebutuhan analisis. Hasil pengolahan menunjukkan bahwa distribusi PNS sangat bervariasi antarinstansi, dengan Dinas Pendidikan sebagai instansi terbesar yang menyerap hampir 60% dari total PNS provinsi, sedangkan mayoritas instansi lainnya tergolong kecil dengan jumlah pegawai di bawah 100 orang. Analisis juga mengungkap bahwa komposisi gender di 46,3% instansi telah seimbang, sementara dominasi pegawai wanita lebih banyak ditemukan pada sektor pelayanan publik seperti pendidikan dan kesehatan. Kesimpulan dari penelitian ini adalah bahwa data wrangling berperan penting dalam menghasilkan dataset akhir yang lebih bersih, akurat, dan informatif sehingga dapat digunakan sebagai dasar pengambilan keputusan kepegawaian yang lebih tepat dan mendukung perencanaan kebijakan publik berbasis data.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pegawai Negeri Sipil (PNS) adalah garda terdepan dalam menjalankan pemerintahan dan memberikan pelayanan kepada masyarakat. Keberadaan PNS yang cukup, tersebar dengan baik, dan bekerja secara profesional sangat penting untuk menciptakan pelayanan publik yang berkualitas. Di era modern ini, data kepegawaian bukan hanya sekadar catatan administratif biasa, tetapi sudah menjadi informasi penting yang dapat membantu pemerintah dalam membuat keputusan dan kebijakan yang tepat (Badan Kepegawaian Negara, 2023).

Provinsi Sumatera Selatan memiliki ribuan PNS yang tersebar di berbagai dinas dan instansi pemerintah. Data PNS tahun 2024 yang mencatat jumlah pegawai di setiap dinas dan membedakan antara pegawai laki-laki dan perempuan menyimpan banyak informasi berharga. Dari data ini, kita dapat mengetahui berbagai hal penting seperti: Apakah jumlah pegawai di setiap instansi sudah sesuai dengan kebutuhan? Apakah ada kesenjangan antara jumlah pegawai laki-laki dan perempuan? Instansi mana yang kekurangan atau kelebihan pegawai? Informasi-informasi ini sangat berguna untuk membuat perencanaan kepegawaian yang lebih baik.

Namun kenyataannya, data yang dikumpulkan dari berbagai sumber biasanya masih dalam kondisi mentah dan belum siap untuk dianalisis. Data mentah ini sering mengalami berbagai masalah seperti ada data yang hilang atau tidak lengkap, ada pencatatan yang sama tertulis dua kali, penulisan nama instansi yang tidak seragam, kesalahan dalam memasukkan data, serta bentuk penyajian data yang belum sesuai untuk diolah lebih lanjut (Fitriani & Nugroho, 2023). Jika data seperti ini langsung dianalisis tanpa dibersihkan terlebih dahulu, hasilnya bisa menyesatkan dan tidak akurat.

Untuk mengatasi masalah ini, diperlukan proses pengolahan data yang disebut dengan teknik munging atau wrangling. Sederhananya, ini adalah proses membersihkan dan mengubah data mentah yang masih berantakan menjadi data yang rapi, teratur, dan siap untuk dianalisis (Kusuma, 2022). Proses ini meliputi enam langkah penting yang saling berkaitan. Langkah pertama adalah menciptakan variabel baru, yaitu membuat informasi tambahan dari data yang sudah ada, misalnya menghitung perbandingan jumlah pegawai laki-laki dan perempuan (Hidayat, 2023). Langkah kedua adalah subsetting data, yaitu memilih bagian tertentu dari data yang ingin difokuskan untuk analisis (Lestari, 2022). Langkah ketiga adalah sorting data, yaitu mengurutkan data berdasarkan kriteria tertentu agar mudah dibaca dan dipahami. Langkah keempat adalah recoding data, yaitu memperbaiki dan menyeragamkan penulisan yang tidak konsisten (Anggraini, 2022). Langkah kelima adalah merging data, yaitu menggabungkan beberapa data dari sumber berbeda menjadi satu kesatuan (Nurhayati, 2023). Langkah terakhir

adalah reshaping data, yaitu mengubah bentuk penyajian data agar sesuai dengan kebutuhan analisis.

Penerapan teknik pengolahan data yang baik dan terdokumentasi dengan jelas pada data PNS Provinsi Sumatera Selatan sangat diperlukan untuk menghasilkan informasi yang dapat dipercaya. Informasi ini nantinya akan menjadi dasar bagi pemerintah daerah dalam membuat berbagai keputusan penting seperti merencanakan kebutuhan pegawai baru, memindahkan pegawai ke tempat yang lebih membutuhkan, merencanakan program pelatihan, serta menerapkan kebijakan kesetaraan antara pegawai laki-laki dan perempuan. Selain itu, penelitian ini juga penting untuk meningkatkan keterbukaan dan pertanggungjawaban dalam pengelolaan kepegawaian. Dengan data yang sudah diolah dengan baik, pemerintah daerah dapat membuat kebijakan berdasarkan fakta dan angka yang jelas, bukan hanya berdasarkan perkiraan atau kebiasaan. Hasil pengolahan data juga dapat digunakan sebagai alat untuk memantau dan mengevaluasi kinerja manajemen kepegawaian secara objektif.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, penelitian ini ingin menjawab pertanyaan-pertanyaan berikut:

1. Apa tahapan yang digunakan pada teknik wrangling yang diperlukan untuk mengolah data jumlah Pegawai Negeri Sipil di wilayah Sumatera bagian Selatan tahun 2024?
2. Bagaimana proses penerapan setiap tahapan teknik wrangling (seperti data cleaning, data transformation, data integration, dan data formatting) terhadap data jumlah PNS di Sumatera bagian Selatan tahun 2024?
3. Bagaimana hasil akhir data yang telah melalui proses teknik wrangling, terutama dalam hal struktur, format, dan keteraturan datanya?

1.3. Tujuan Penelitian

1. Mengaplikasikan teknik data wrangling dalam menganalisis data jumlah Pegawai Negeri Sipil di wilayah Sumatera bagian selatan tahun 2024 untuk menghasilkan informasi yang lebih terstruktur dan bermakna.
2. Menciptakan variabel baru dari data PNS Sumatera bagian selatan untuk memperkaya informasi analisis, seperti variabel kategori usia, masa kerja, atau rasio PNS per penduduk.
3. Melakukan subsetting data PNS berdasarkan kriteria tertentu seperti provinsi, golongan, jenis kelamin, atau unit kerja untuk memfokuskan analisis pada segmen data yang relevan.
4. Mengurutkan (sorting) data PNS berdasarkan variabel-variabel penting seperti jumlah pegawai, wilayah, atau karakteristik demografis untuk memudahkan identifikasi pola dan tren.

5. Melakukan recoding data untuk mengkategorikan ulang atau menyederhanakan variabel-variabel seperti tingkat pendidikan, golongan kepangkatan, atau rentang usia menjadi kategori yang lebih mudah dianalisis.
6. Menggabungkan (merging) data PNS dari berbagai sumber atau dataset untuk mendapatkan gambaran yang lebih komprehensif mengenai kondisi kepegawaian di Sumatera bagian selatan.
7. Mengubah struktur (reshaping) data dari format wide ke long atau sebaliknya untuk memfasilitasi analisis statistik dan visualisasi data yang lebih efektif.
8. Menghasilkan insight statistik mengenai distribusi, karakteristik, dan pola data PNS di wilayah Sumatera bagian selatan tahun 2024 sebagai dasar pengambilan keputusan kebijakan kepegawaian.

1.4 Manfaat Penelitian

Penelitian ini diharapkan memberikan berbagai manfaat, baik secara teoritis maupun praktis, khususnya dalam pengelolaan data kepegawaian di Provinsi Sumatera Selatan.

1.4.1 Manfaat Teoritis

1. Pengembangan Ilmu Pengetahuan
Penelitian ini memberikan kontribusi pada bidang manajemen data dan administrasi publik, khususnya terkait penerapan teknik munging atau data wrangling dalam pengolahan data kepegawaian pemerintah daerah.
2. Referensi Akademik
Hasil penelitian dapat menjadi rujukan bagi peneliti lain dalam mengembangkan metode serupa, baik untuk analisis data kepegawaian maupun data pemerintahan lainnya.
3. Peningkatan Pemahaman Metodologis
Penelitian ini memperjelas tahapan praktis dalam membersihkan, menggabungkan, dan menata ulang data sehingga dapat digunakan sebagai contoh dalam pengajaran atau pelatihan analisis data.

1.4.2 Manfaat Praktis

1. Bagi Pemerintah Daerah

- Memberikan dasar informasi yang akurat dan terstruktur untuk pengambilan keputusan kepegawaian.
- Membantu perencanaan kebutuhan pegawai baru dan evaluasi distribusi pegawai pada masing-masing instansi.
- Mendukung kebijakan pemerataan dan kesetaraan gender melalui informasi yang lebih jelas dan terukur.
- Memperkuat sistem manajemen kepegawaian yang transparan dan akuntabel.

2. Bagi Pengelola Data Kepegawaian

- Menyediakan pedoman praktis mengenai cara membersihkan, menyederhanakan, serta menyiapkan data agar siap dianalisis.
- Meningkatkan kemampuan staf dalam melakukan manajemen data modern yang lebih efisien dan minim kesalahan.
- Membantu membangun standar operasional prosedur (SOP) pengolahan data kepegawaian sehingga proses pengelolaan data lebih konsisten.

3. Bagi Peneliti dan Akademisi

- Memberikan contoh kasus nyata mengenai pengolahan data PNS yang dapat digunakan dalam penelitian lanjutan.
- Menjadi dasar untuk penelitian lebih lanjut terkait analisis kinerja instansi, kebutuhan formasi pegawai, atau manajemen SDM aparatur.

4. Bagi Masyarakat dan Publik

- Menyediakan informasi yang lebih terbuka terkait kondisi kepegawaian pemerintah daerah.
- Meningkatkan kepercayaan publik terhadap pemerintah melalui penyajian data yang akurat dan dapat dipertanggungjawabkan.

BAB II

LANDASAN TEORI

2.1 Data Wrangling

Data wrangling merupakan proses fundamental dalam pengolahan data yang memainkan peran penting dalam analisis data modern. Data wrangling adalah proses pembersihan, penyusunan, dan pengayaan data mentah menjadi format yang lebih terstruktur dan siap digunakan untuk analisis (Raharjo, 2023). Proses ini juga dikenal dengan istilah data munging atau data preparation yang merupakan cara untuk mengatasi masalah kualitas data seperti nilai yang hilang, duplikat, outlier, dan inkonsistensi format. Pratama dan Sari (2022) menambahkan bahwa data wrangling adalah tahapan mengubah dan mengorganisasi data mentah yang tidak terstruktur menjadi data bersih yang dapat diolah lebih lanjut. Dalam konteks analisis data kepegawaian, data wrangling menjadi tahap yang sangat penting karena data PNS seringkali berasal dari berbagai sumber dengan format yang berbeda-beda, sehingga memerlukan proses pengolahan agar dapat dianalisis secara efektif.

Tujuan utama dari proses data wrangling adalah mengubah data mentah yang tidak terstruktur atau bermasalah menjadi kumpulan data bersih yang dapat dianalisis secara efektif. Menurut Santoso (2023), tujuan dari proses data wrangling adalah memangkas waktu analisis karena menggunakan metode yang lebih efisien serta memastikan semua data yang diinput merupakan data yang benar-benar diperlukan. Proses ini memiliki beberapa manfaat utama, yaitu menghasilkan data akhir yang lebih akurat, memaksimalkan tingkat akurasi dan kualitas data untuk analisis lebih lanjut, serta memudahkan proses interpretasi dan visualisasi data. Wijaya (2021) menyebutkan bahwa seorang analis data dapat menghabiskan sebagian besar waktunya untuk melakukan data wrangling sebelum masuk ke tahap analisis dan pemodelan, yang menunjukkan betapa pentingnya proses ini dalam keseluruhan siklus analisis data.

Proses data wrangling mencakup beberapa tahapan penting yang harus dilakukan secara sistematis. Tahapan pertama adalah penemuan data (discovery), yaitu mengidentifikasi dan memahami karakteristik data yang akan diolah termasuk struktur, pola, dan masalah yang ada dalam data tersebut (Kusuma, 2022). Pada tahap ini dilakukan eksplorasi awal untuk mengenali jenis variabel, distribusi data, dan potensi masalah yang mungkin dihadapi. Tahapan kedua adalah strukturisasi data (structuring), yaitu proses mengatur data dengan cara yang lebih koheren dan bermanfaat sehingga memiliki format yang sesuai untuk kebutuhan analisis. Tahapan ketiga adalah pembersihan data (cleaning), merupakan fase kritis di mana setiap bagian data diperiksa dengan cermat, entri yang berlebihan atau tidak relevan dihapus, nilai yang hilang diperbaiki, dan data distandarisasi untuk meningkatkan kualitasnya (Fitriani dan Nugroho, 2023). Tahapan keempat adalah pengayaan data (enriching), yaitu proses menambahkan informasi tambahan dari sumber lain untuk memperkaya dataset yang sudah ada. Tahapan kelima adalah validasi data (validating), merujuk pada proses verifikasi untuk memastikan data sudah konsisten

dan berkualitas tinggi. Tahapan terakhir adalah publikasi data (publishing), yaitu menyiapkan data yang telah diolah untuk digunakan dalam analisis, visualisasi, atau pelaporan.

2.2 Teknik-Teknik Data Wrangling

2.2.1 Menciptakan Variabel Baru (Variable Creation)

Menciptakan variabel baru merupakan proses membentuk variabel tambahan dari variabel yang sudah ada untuk memperkaya informasi dalam dataset. Teknik ini dilakukan untuk mengkategorikan data numerik menjadi kategori tertentu seperti kategori usia atau masa kerja, menghitung rasio atau proporsi dari variabel yang ada, membuat indikator berdasarkan kondisi tertentu, serta menggabungkan informasi dari beberapa variabel menjadi satu variabel baru (Hidayat, 2023). Pembuatan variabel baru merupakan salah satu aktivitas penting dalam data wrangling yang memungkinkan analis untuk mengekstrak informasi tambahan dari data yang sudah ada. Dalam konteks data PNS, variabel baru dapat berupa kategori golongan, rentang usia pegawai, atau rasio PNS per penduduk di suatu wilayah, yang semuanya memberikan perspektif analisis yang lebih kaya dan mendalam untuk keperluan perencanaan kebijakan kepegawaian.

2.2.2 Subsetting Data

Subsetting data adalah proses pemilihan sebagian data dari dataset lengkap berdasarkan kriteria tertentu. Menurut Lestari (2022), subsetting berguna untuk memfokuskan analisis pada bagian data yang paling relevan dengan tujuan penelitian. Teknik subsetting bermanfaat untuk memfokuskan analisis pada segmen data yang relevan, mengurangi kompleksitas data dengan hanya mengambil variabel atau observasi yang diperlukan, memfilter data berdasarkan kondisi tertentu seperti PNS di provinsi tertentu atau golongan tertentu, serta meningkatkan efisiensi komputasi dengan bekerja pada subset data yang lebih kecil. Dengan mengeliminasi data yang tidak penting, analis dapat berfokus pada informasi yang paling bermakna untuk tujuan penelitian mereka, sehingga proses analisis menjadi lebih efisien dan terarah.

2.2.3 Sorting Data

Sorting atau pengurutan data adalah proses mengurutkan observasi dalam dataset berdasarkan satu atau lebih variabel. Manfaat sorting meliputi kemudahan dalam identifikasi nilai maksimum dan minimum, membantu dalam deteksi pola dan tren dalam data, mempermudah proses pencarian dan referensi data, serta mempersiapkan data untuk teknik analisis tertentu yang memerlukan data terurut (Prabowo, 2023). Proses pengurutan data memfasilitasi analisis yang lebih sistematis dan memungkinkan peneliti untuk dengan cepat mengidentifikasi karakteristik ekstrem dalam dataset, seperti provinsi dengan jumlah PNS tertinggi atau terendah. Sorting juga mempermudah dalam membuat rangking dan perbandingan antar kategori data.

2.2.4 Recoding Data

Recoding adalah proses mengubah nilai variabel menjadi nilai atau kategori baru. Teknik ini digunakan untuk menyederhanakan kategori yang terlalu detail menjadi kategori yang lebih umum, mengubah skala pengukuran variabel, menggabungkan kategori dengan frekuensi rendah, serta membuat kategori baru berdasarkan logika analisis tertentu (Anggraini, 2022). Contohnya adalah mengubah variabel tingkat pendidikan yang detail menjadi kategori sederhana seperti "Menengah", "Sarjana", dan "Pascasarjana". Proses recoding sangat penting untuk membuat data lebih mudah dianalisis dan diinterpretasikan, terutama ketika berurusan dengan variabel kategorikal yang memiliki terlalu banyak level atau kategori yang dapat menyulitkan proses analisis dan interpretasi hasil.

2.2.5 Merging Data

Merging atau penggabungan data adalah proses menggabungkan dua atau lebih dataset berdasarkan variabel kunci yang sama. Nurhayati (2023) menjelaskan bahwa penggabungan data dari berbagai sumber menjadi satu kumpulan data yang terintegrasi memungkinkan analisis yang lebih komprehensif. Teknik merging meliputi inner join yang menggabungkan hanya data yang memiliki kecocokan di kedua dataset, left join yang mengambil semua data dari dataset kiri dan data yang cocok dari dataset kanan, right join yang mengambil semua data dari dataset kanan dan data yang cocok dari dataset kiri, serta outer join yang menggabungkan semua data dari kedua dataset. Merging sangat berguna ketika data PNS tersebar dalam beberapa file atau database terpisah yang perlu digabungkan untuk analisis komprehensif, misalnya menggabungkan data kepegawaian dengan data kinerja atau data demografis wilayah.

2.2.6 Reshaping Data

Reshaping adalah proses mengubah struktur atau format data tanpa mengubah informasi yang terkandung di dalamnya. Terdapat dua jenis utama reshaping menurut Setiawan (2022): pertama adalah wide to long, yaitu mengubah data dari format lebar dengan banyak kolom menjadi format panjang dengan lebih sedikit kolom tetapi lebih banyak baris, di mana format ini cocok untuk analisis time series atau data panel. Kedua adalah long to wide, yaitu mengubah data dari format panjang menjadi format lebar yang memudahkan untuk membaca dan membandingkan nilai antar kategori. Reshaping memfasilitasi analisis statistik dan visualisasi data yang lebih efektif sesuai dengan kebutuhan analisis, karena struktur data yang tepat dapat sangat mempengaruhi efisiensi dan kemudahan proses analisis selanjutnya serta kesesuaian dengan metode statistik yang akan digunakan.

2.3 Pegawai Negeri Sipil (PNS)

Pegawai Negeri Sipil (PNS) adalah warga negara Indonesia yang memenuhi syarat tertentu, diangkat sebagai pegawai Aparatur Sipil Negara secara tetap oleh pejabat pembina kepegawaian

untuk menduduki jabatan pemerintahan, sebagaimana diatur dalam Undang-Undang Nomor 5 Tahun 2014 tentang Aparatur Sipil Negara. Menurut Peraturan Pemerintah Nomor 11 Tahun 2017, PNS adalah warga negara Indonesia yang memenuhi syarat tertentu, diangkat sebagai Pegawai ASN secara tetap oleh Pejabat Pembina Kepegawaian untuk menduduki jabatan pemerintahan. PNS bertugas memberikan pelayanan publik dan menjalankan tugas pemerintahan sesuai dengan peraturan perundang-undangan yang berlaku. Ramadhani (2023) menjelaskan bahwa PNS berperan sebagai perencana, pelaksana, dan pengawas penyelenggaraan tugas umum pemerintahan dan pembangunan nasional melalui pelaksanaan kebijakan dan pelayanan publik yang profesional, bebas dari intervensi politik, serta bersih dari praktik korupsi, kolusi, dan nepotisme.

Data statistik PNS mencakup berbagai informasi kepegawaian yang penting untuk perencanaan dan evaluasi kebijakan. Informasi ini meliputi jumlah dan distribusi PNS berdasarkan wilayah, komposisi PNS berdasarkan jenis kelamin, usia, dan tingkat pendidikan, distribusi berdasarkan golongan dan jabatan, serta sebaran PNS menurut instansi dan unit kerja (Badan Kepegawaian Negara, 2023). Data statistik PNS menjadi dasar penting dalam perencanaan kebijakan kepegawaian, evaluasi komposisi aparatur, dan pengambilan keputusan terkait pengembangan sumber daya manusia pemerintahan. Oleh karena itu, kualitas data PNS yang baik melalui proses data wrangling yang tepat menjadi sangat krusial untuk menghasilkan analisis dan kebijakan yang akurat serta dapat dipertanggungjawabkan dalam perencanaan pembangunan daerah maupun nasional.

2.4 Analisis Statistik

Analisis statistik adalah proses pengumpulan, pengolahan, penyajian, dan interpretasi data untuk menghasilkan informasi yang bermakna. Dalam konteks data PNS, analisis statistik digunakan untuk mendeskripsikan karakteristik kepegawaian di suatu wilayah, mengidentifikasi pola dan tren dalam distribusi PNS, membandingkan kondisi kepegawaian antar wilayah atau periode, serta mendukung pengambilan keputusan berbasis data dalam manajemen kepegawaian (Suhendra, 2022). Ilmu statistika membantu analisis dalam memahami pola dan tren dalam data setelah proses data wrangling selesai, dan statistika juga digunakan untuk melakukan analisis deskriptif, inferensial, dan probabilistik guna mendukung pengambilan keputusan berbasis data. Analisis statistik yang baik memerlukan data yang berkualitas tinggi, sehingga teknik data wrangling menjadi tahap krusial sebelum melakukan analisis lebih lanjut. Tanpa data yang bersih dan terstruktur dengan baik, hasil analisis statistik dapat menjadi menyesatkan dan menghasilkan kesimpulan yang tidak akurat, yang pada akhirnya dapat mempengaruhi kualitas pengambilan keputusan dalam konteks manajemen kepegawaian dan perencanaan pembangunan aparatur sipil negara.

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan deskriptif dan eksperimental-komputasi (eksploratif). Dikategorikan sebagai penelitian kuantitatif karena fokusnya adalah pada analisis data numerik, yaitu angka-angka diskrit mengenai jumlah Pegawai Negeri Sipil (PNS) di Sumatera Bagian Selatan. Pendekatan utama yang digunakan adalah deskriptif, bertujuan untuk memetakan dan menggambarkan distribusi serta karakteristik komposisi data kepegawaian yang diteliti. Selain itu, metodologi ini diperkaya dengan aspek eksperimental-komputasi (eksploratif) melalui penerapan teknik *data wrangling* yang sistematis untuk mengubah data mentah yang kompleks menjadi data bersih yang siap diolah, sebelum dilakukan analisis statistik akhir.

3.2 Jenis Data

Jenis data yang dianalisis dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS). Data ini diklasifikasikan sebagai kuantitatif diskrit karena menyajikan enumerasi (*cacahan*) jumlah Pegawai Negeri Sipil (PNS) berdasarkan jenis kelamin dan total per instansi. Dimensi waktu yang digunakan adalah *cross-sectional*, di mana observasi difokuskan pada perbandingan karakteristik kepegawaian antar berbagai instansi pemerintah Provinsi Sumatera Selatan pada satu titik waktu, yaitu tahun 2024.

3.3 Teknik pengumpulan data

Teknik pengumpulan data yang diterapkan dalam penelitian ini adalah studi dokumentasi (studi literatur). Metode ini sangat relevan mengingat penelitian ini sepenuhnya mengandalkan data sekunder yang telah dikumpulkan dan dipublikasikan oleh lembaga resmi. Proses ini melibatkan penelusuran dan perekaman data rekapitulasi jumlah Pegawai Negeri Sipil (PNS) berdasarkan instansi pemerintah dan komposisi jenis kelamin untuk periode 2024, yang secara spesifik diperoleh dari dokumen atau basis data publikasi Badan Pusat Statistik (BPS). Data yang didokumentasikan meliputi variabel-variabel kuantitatif diskrit: jumlah PNS pria, jumlah PNS wanita, dan total PNS per instansi.

3.4 Variabel yang diamati

Tabel 1. Rincian Variabel dan Indikator

Variabel	Definisi	Simbol	Satuan	Skala
Jumlah Pegawai Negeri Sipil Pria	Jumlah pegawai negeri sipil berjenis kelamin pria di suatu instansi.	Pria	Jumlah orang	Rasio
Jumlah Pegawai Negeri Sipil Wanita	Jumlah pegawai negeri sipil berjenis kelamin wanita di suatu instansi.	Wanita	Jumlah orang	Rasio
Total Pegawai Negeri Sipil	Jumlah seluruh pegawai negeri sipil (pria + wanita) pada suatu instansi.	Total	Jumlah orang	Rasio
Instansi	Unit atau lembaga pemerintahan tempat data pegawai diperoleh; digunakan untuk subsetting data.	Instansi	Nama Instansi	Nominal

Tabel 2. Dataset Penelitian

No	Instansi	X1 (Pria)	X2 (Wanita)	X3 (Total)
1	Setda Provinsi Sumatera Selatan	185	149	334
2	Sekretariat DPRD Provinsi Sumatera Selatan	55	34	89
3	Inspektorat Provinsi Sumatera Selatan	50	46	96
4	Satuan Polisi Pamong Praja	129	9	138

5	RS.Dr. Ernaldi Bahar	69	216	285
6	Dinas Pendidikan Provinsi Sumatera Selatan	2728	5047	7775
7	Dinas Kesehatan Provinsi Sumatera Selatan	189	621	810
8	Dinas Pengelolaan Sumber Daya Air Provinsi Sumatera Selatan	53	34	87
9	Dinas PU Bina Marga dan Tata Ruang Provinsi Sumatera Selatan	140	84	224
10	Dinas Perumahan dan Kawasan Permukiman Provinsi Sumatera Selatan	63	38	101
11	Dinas Pemuda dan Olahraga Provinsi Sumatera Selatan	43	22	65
12	Dinas Koperasi, Usaha Kecil dan Menengah Provinsi Sumatera Selatan	25	21	46
13	Dinas Tenaga Kerja dan Transmigrasi Provinsi Sumatera Selatan	93	43	136
14	Dinas Perhubungan Provinsi Sumatera Selatan	94	38	132
15	Dinas Sosial Provinsi Sumatera Selatan	46	36	82
16	Dinas Kebudayaan dan Pariwisata Provinsi Sumatera Selatan	41	47	88

17	Dinas Kelautan dan Perikanan Provinsi Sumatera Selatan	56	35	91
18	Dinas Pertanian Tanaman Pangan dan Hortikultura Provinsi Sumatera Selatan	122	113	235
19	Dinas Lingkungan Hidup dan Pertanahan Provinsi Sumatera Selatan	35	30	65
20	Dinas Perkebunan Provinsi Sumatera Selatan	31	25	56
21	Dinas Kehutanan Provinsi Sumatera Selatan	196	77	273
22	Dinas Energi dan Sumber Daya Mineral Provinsi Sumatera Selatan	63	34	97
23	Dinas Perdagangan Provinsi Sumatera Selatan	22	20	42
24	Dinas Perindustrian Provinsi Sumatera Selatan	16	21	37
25	Dinas Kependudukan dan Pencatatan Sipil Provinsi Sumatera Selatan	16	15	31
26	Dinas Pemberdayaan Perempuan dan Perlindungan Anak Provinsi Sumatera Selatan	16	19	35
27	Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu	30	34	64
28	Dinas Perpustakaan Provinsi Sumatera Selatan	28	32	60

29	Dinas Pemberdayaan Masyarakat dan Desa Provinsi Sumatera Selatan	14	15	29
30	Dinas Komunikasi dan Informatika Provinsi Sumatera Selatan	26	23	49
31	Dinas Ketahanan Pangan dan Peternakan Provinsi Sumatera Selatan	44	62	106
32	Dinas Kearsipan Provinsi Sumatera Selatan	15	25	40
33	Badan Perencanaan Pembangunan Daerah Provinsi Sumatera Selatan	47	53	100
34	Badan Kesatuan Bangsa dan Perlindungan Masyarakat Provinsi Sumatera Selatan	24	10	34
35	Dinas Penelitian, Pengembangan Daerah Provinsi Sumatera Selatan	23	32	55
36	Badan Pengembangan SDM Daerah Provinsi Sumatera Selatan	37	39	76
37	Badan Kepegawaian Daerah Provinsi Sumatera Selatan	31	37	68
38	Badan Pelaksana Penanggulangan Bencana Daerah	22	37	59
39	Badan Pengelola Keuangan dan Aset Daerah	41	44	85

40	Badan Penghubung Provinsi Sumatera Selatan	20	16	36
41	Badan Pendapatan Daerah Provinsi Sumatera Selatan	239	179	418

3.5 Diagram Alir

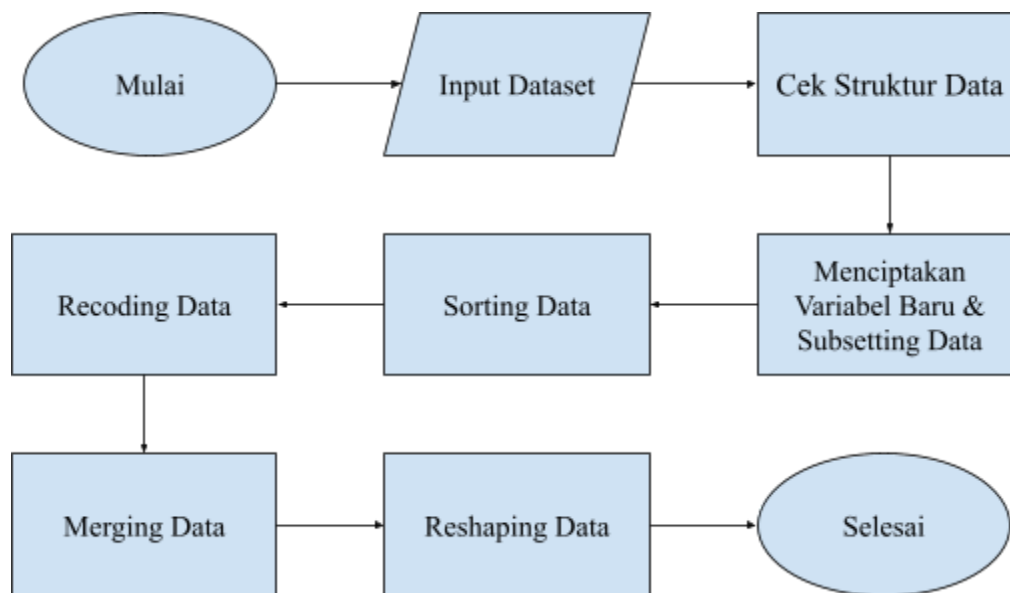


Diagram alir tersebut menggambarkan alur kerja pengolahan data jumlah PNS di setiap instansi di wilayah Sumbagsel melalui serangkaian tahapan wrangling. Proses dimulai dari Mulai, kemudian dilanjutkan dengan Input Dataset, yaitu memasukkan data mentah yang berisi informasi jumlah PNS dari berbagai instansi di provinsi-provinsi Sumbagsel ke dalam perangkat lunak pengolahan data. Setelah itu dilakukan Cek Struktur Data untuk memahami bagaimana data disusun—misalnya memeriksa apakah nama instansi, jumlah PNS, kode provinsi, dan tahun tercatat dengan format yang konsisten. Tahap berikutnya adalah Menciptakan Variabel Baru dan Subsetting Data, di mana peneliti dapat membuat variabel tambahan seperti total PNS per provinsi, kategori instansi, atau rasio jumlah PNS terhadap populasi; serta memilih subset data tertentu, misalnya hanya data tahun 2024 atau hanya instansi pemerintah daerah. Selanjutnya dilakukan Sorting Data, yaitu mengurutkan data berdasarkan kriteria tertentu, seperti mengurutkan instansi berdasarkan jumlah PNS terbesar atau mengurutkan provinsi secara alfabetis.

Setelah data terurut, proses dilanjutkan ke Recoding Data, yaitu memperbaiki atau mengubah kategori variabel agar seragam, seperti menyeragamkan nama instansi, mengubah kode wilayah, atau mengelompokkan instansi ke dalam kategori struktural dan fungsional. Tahap selanjutnya adalah Merging Data, yakni menggabungkan dataset jumlah PNS dengan dataset lain bila diperlukan, misalnya dataset jumlah penduduk, kode wilayah BPS, atau klasifikasi instansi, untuk memperkaya informasi analisis. Setelah data selesai digabungkan, dilakukan Reshaping Data, yaitu mengubah bentuk data agar lebih mudah dianalisis, misalnya mengubah data dari format lebar (wide) menjadi panjang (long), atau sebaliknya. Seluruh rangkaian proses wrangling ini berakhir pada tahap Selesai, yang menandakan bahwa dataset jumlah PNS di setiap instansi di Sumbagsel telah bersih, terstruktur, dan siap digunakan untuk analisis lebih lanjut seperti visualisasi, perbandingan antarinstansi, atau penyusunan laporan statistik.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data sekunder mengenai Jumlah Pegawai Negeri Sipil (PNS) Menurut Dinas/Instansi Pemerintah dan Jenis Kelamin pada Tahun 2024, dengan cakupan wilayah Provinsi Sumatera Selatan. Data ini diperoleh dari Badan Pusat Statistik (BPS) dan mencakup 41 instansi pemerintah provinsi dengan total 13.049 PNS yang terdiri dari 4.877 PNS pria (37,4%) dan 8.172 PNS wanita (62,6%).

Dataset awal memiliki struktur dengan 41 observasi (baris) dan 4 variabel (kolom), yaitu: Instansi (nama dinas/lembaga), Pria (jumlah PNS laki-laki), Wanita (jumlah PNS perempuan), dan Total (jumlah keseluruhan PNS). Data awal berada dalam format wide (lebar) yang perlu ditransformasi untuk keperluan analisis yang lebih mendalam.

Tabel 3. Statistik Deskriptif

Statistik	X1 (Pria)	X2 (Wanita)	X3 (Total)
Jumlah Instansi (n)	41	41	41
Total Keseluruhan	5698	8081	13779
Rata-rata (\bar{X})	139.0	197.1	336.1
Nilai Minimum (min)	14	9	29
Nilai Maksimum (max)	2728	5047	7775
Median (\tilde{X})	41	37	89
Standar Deviasi (σ)	446.74	802.73	1250.60

4.2. Proses Data Wrangling

4.2.1. Menciptakan Variabel Baru (Variable Creation)

Pada tahap menciptakan variabel baru, dilakukan pembuatan variabel tambahan dari data yang sudah ada untuk memperkaya informasi analisis. Dari dataset awal yang hanya memiliki variabel jumlah PNS pria, wanita, dan total per instansi, dibuat beberapa variabel baru seperti persentase gender, rasio pria terhadap wanita, dan kategori dominasi gender. Variabel kategori dominasi gender dibuat dengan kriteria tertentu: jika persentase pria atau wanita melebihi 60% maka dikategorikan sebagai "Dominasi Pria" atau "Dominasi Wanita", sedangkan jika persentase keduanya berkisar 40-60% maka dikategorikan sebagai "Seimbang". Berdasarkan variabel baru ini, ditemukan bahwa dari 41 instansi pemerintah Provinsi Sumatera Selatan tahun 2024, sebanyak 19 instansi (46,3%) memiliki komposisi gender yang seimbang, 13 instansi (31,7%) didominasi oleh pegawai wanita, dan 9 instansi (22,0%) didominasi oleh pegawai pria. Analisis lebih lanjut menggunakan boxplot menunjukkan bahwa instansi dengan dominasi wanita cenderung merupakan instansi besar dengan median jumlah pegawai sekitar 300-400 orang, terutama di sektor pendidikan dan kesehatan seperti Dinas Pendidikan (7.775 PNS) dan Dinas Kesehatan (810 PNS). Sementara instansi dengan dominasi pria memiliki median sekitar 100 pegawai, umumnya di sektor keamanan dan kehutanan seperti Satpol PP (138 PNS) dan Dinas Kehutanan (273 PNS). Adapun instansi dengan komposisi seimbang cenderung berukuran kecil hingga menengah dengan median 60-80 pegawai. Penciptaan variabel baru ini sangat bermanfaat untuk mengidentifikasi pola segregasi pekerjaan berdasarkan gender dan memberikan insight yang lebih mendalam untuk perencanaan kebijakan kepegawaian yang mendorong kesetaraan gender di seluruh sektor pemerintahan.

4.2.2. Subsetting Data

Proses subsetting dilakukan untuk memfokuskan analisis pada segmen data yang relevan dengan memfilter dataset berdasarkan kriteria tertentu. Pertama, dilakukan subset berdasarkan ukuran instansi untuk mengidentifikasi instansi besar yang memiliki jumlah PNS ≥ 200 orang, dan ditemukan 8 instansi yang termasuk kategori ini, yaitu Dinas Pendidikan (7.775 PNS), Dinas Kesehatan (810 PNS), Badan Pendapatan Daerah (418 PNS), Setda Provinsi (334 PNS), RS Dr. Ernaldi Bahar (285 PNS), Dinas Kehutanan (273 PNS), Dinas Pertanian Tanaman Pangan dan Hortikultura (235 PNS), dan Dinas PU Bina Marga dan Tata Ruang (224 PNS). Kedua, subset berdasarkan dominasi gender dilakukan dengan memfilter instansi yang memiliki persentase pria atau wanita lebih dari 60%, dimana ditemukan 9 instansi dengan dominasi pria termasuk Satuan Polisi Pamong Praja (93,5%), Sekretariat DPRD (61,8%), dan Dinas Kehutanan (71,8%), serta 13 instansi dengan dominasi wanita termasuk RS Dr. Ernaldi Bahar (75,8%), Dinas Kesehatan (76,7%), dan Dinas Pendidikan (64,9%). Ketiga, dilakukan subset untuk mengidentifikasi instansi kecil dengan memfilter instansi yang memiliki jumlah PNS kurang dari 50 orang, dan ditemukan 8 instansi yang termasuk kategori ini dengan Dinas Pemberdayaan Masyarakat dan Desa memiliki jumlah PNS paling sedikit yaitu 29 orang. Proses subsetting ini sangat bermanfaat

untuk mengidentifikasi instansi yang mungkin memerlukan penambahan pegawai atau redistribusi tenaga kerja, serta untuk menganalisis pola distribusi gender pada instansi dengan karakteristik tertentu sebagai dasar perencanaan kebijakan kepegawaian yang lebih targeted dan efektif.

4.2.3. Sorting Data

Pengurutan data dilakukan untuk memudahkan identifikasi pola dan karakteristik ekstrem dalam dataset PNS Provinsi Sumatera Selatan. Pertama, dilakukan sorting berdasarkan total PNS secara descending (dari terbesar ke terkecil) untuk mengidentifikasi 10 instansi dengan jumlah pegawai terbanyak sebagaimana ditampilkan pada Tabel 3. Hasil pengurutan menunjukkan bahwa Dinas Pendidikan menempati posisi teratas dengan 7.775 PNS (2.728 pria dan 5.047 wanita), diikuti oleh Dinas Kesehatan dengan 810 PNS (189 pria dan 621 wanita), Badan Pendapatan Daerah dengan 418 PNS (239 pria dan 179 wanita), Setda Provinsi dengan 334 PNS (185 pria dan 149 wanita), dan RS Dr. Ernaldi Bahar dengan 285 PNS (69 pria dan 216 wanita). Instansi lainnya dalam 10 besar meliputi Dinas Kehutanan (273 PNS), Dinas Pertanian Tanaman Pangan dan Hortikultura (235 PNS), Dinas PU Bina Marga dan Tata Ruang (224 PNS), Satuan Polisi Pamong Praja (138 PNS), dan Dinas Tenaga Kerja dan Transmigrasi (136 PNS).

Kedua, sorting dilakukan secara ascending (dari terkecil ke terbesar) untuk mengidentifikasi 10 instansi dengan jumlah PNS paling sedikit sebagaimana ditampilkan pada Tabel 4. Hasilnya menunjukkan bahwa Dinas Pemberdayaan Masyarakat dan Desa memiliki jumlah pegawai paling sedikit dengan 29 PNS (14 pria dan 15 wanita), diikuti oleh Dinas Kependudukan dan Pencatatan Sipil (31 PNS), Badan Kesatuan Bangsa dan Perlindungan Masyarakat (34 PNS), Dinas Pemberdayaan Perempuan dan Perlindungan Anak (35 PNS), dan Badan Penghubung Provinsi (36 PNS). Instansi lainnya dalam 10 terbawah meliputi Dinas Perindustrian (37 PNS), Dinas Kearsipan (40 PNS), Dinas Perdagangan (42 PNS), Dinas Koperasi, Usaha Kecil dan Menengah (46 PNS), dan Dinas Komunikasi dan Informatika (49 PNS). Ketiga, dilakukan sorting berdasarkan rasio gender untuk mengidentifikasi instansi dengan ketimpangan gender terbesar, dimana rasio tertinggi menunjukkan dominasi pria ditemukan pada Satuan Polisi Pamong Praja dengan rasio 14,33 (pria:wanita), sedangkan rasio terendah menunjukkan dominasi wanita ditemukan pada RS Dr. Ernaldi Bahar dengan rasio 0,32 (pria:wanita). Proses sorting ini sangat berguna untuk mengidentifikasi instansi yang memerlukan perhatian khusus dalam hal penambahan atau redistribusi pegawai, serta untuk mengevaluasi keseimbangan gender di berbagai instansi pemerintahan sebagai dasar penyusunan kebijakan kepegawaian yang lebih adil dan merata.

Tabel 4. Instansi dengan Jumlah PNS Terbanyak

Ranking	Instansi	Pria	Wanita	Total
1	Dinas Pendidikan	2.728	5.047	7.775
2	Dinas Kesehatan	189	621	810
3	Badan Pendapatan Daerah	239	179	418
4	Setda Provinsi	185	149	334
5	RS. Dr. Ernaldi Bahar	69	216	285
6	Dinas Kehutanan	196	77	273
7	Dinas Pertanian Tanaman Pangan dan Hortikultura	122	113	235
8	Dinas PU Bina Marga dan Tata Ruang	140	84	224
9	Satuan Polisi Pamong Praja	129	49	178
10	Dinas Tenaga Kerja dan Transmigrasi	93	43	136

Tabel 5. Instansi dengan Jumlah PNS Paling Sedikit

Ranking	Instansi	Pria	Wanita	Total
1	Dinas Pemberdayaan Masyarakat dan Desa	14	15	29
2	Dinas Kependudukan dan Pencatatan Sipil	16	15	31

3	Badan Kesatuan Bangsa dan Perlindungan Masyarakat	24	10	34
4	Dinas Pemberdayaan Perempuan dan Perlindungan Anak	16	19	35
5	Badan Penghubung Provinsi	20	16	36
6	Dinas Perindustrian	21	16	37
7	Dinas Kearsipan	15	25	40
8	Dinas Perdagangan	22	20	42
9	Dinas Koperasi, Usaha Kecil dan Menengah	25	21	46
10	Dinas Komunikasi dan Informatika	26	23	49

4.2.4. Recoding Data

Proses recoding dilakukan untuk menyederhanakan dan mengkategorikan variabel agar lebih mudah dianalisis dan diinterpretasikan. Pertama, dilakukan kategorisasi ukuran instansi dengan merecode variabel Total PNS menjadi empat kategori berdasarkan jumlah pegawai, yaitu "Sangat Besar" untuk instansi dengan total lebih dari 1000 PNS, "Besar" untuk instansi dengan 500 hingga 1000 PNS, "Sedang" untuk instansi dengan 100 hingga kurang dari 500 PNS, dan "Kecil" untuk instansi dengan kurang dari 100 PNS. Hasil recoding menunjukkan bahwa mayoritas instansi di Provinsi Sumatera Selatan termasuk kategori kecil dengan 33 instansi (80,6%), diikuti oleh kategori sedang dengan 6 instansi (14,6%), kategori besar dengan 1 instansi (2,4%), dan kategori sangat besar dengan 1 instansi (2,4%). Distribusi ini mengindikasikan bahwa sebagian besar instansi pemerintah provinsi memiliki jumlah pegawai yang relatif sedikit, dengan hanya Dinas Pendidikan yang masuk kategori sangat besar dan Dinas Kesehatan yang masuk kategori besar. Kedua, dilakukan kategorisasi status gender dengan merecode berdasarkan persentase gender, dimana instansi dikategorikan sebagai "Dominasi Pria" jika persentase pria lebih dari

60%, "Dominasi Wanita" jika persentase wanita lebih dari 60%, dan "Seimbang" jika persentase keduanya tidak melebihi 60%.

Hasil recoding menunjukkan bahwa 19 instansi (46,3%) memiliki komposisi gender yang seimbang, 13 instansi (31,7%) didominasi oleh wanita, dan 9 instansi (22,0%) didominasi oleh pria. Temuan ini menunjukkan bahwa hampir setengah dari instansi pemerintah Provinsi Sumatera Selatan telah mencapai keseimbangan gender yang baik, meskipun masih terdapat sejumlah instansi dengan ketimpangan gender yang signifikan. Ketiga, dilakukan penyeragaman nama instansi dengan mempersingkat nama-nama instansi yang terlalu panjang untuk memudahkan proses visualisasi dan analisis data, seperti mengubah "Dinas Pendidikan Provinsi Sumatera Selatan" menjadi "Dinas Pendidikan" dan "Setda Provinsi Sumatera Selatan" menjadi "Setda Provinsi". Proses recoding ini sangat penting untuk menyederhanakan struktur data dan memudahkan interpretasi hasil analisis, serta memberikan gambaran yang lebih jelas tentang karakteristik kepegawaian di Provinsi Sumatera Selatan yang dapat digunakan sebagai dasar perencanaan dan pengambilan keputusan kebijakan kepegawaian yang lebih efektif dan targeted.

4.2.5. Merging Data

Untuk memperkaya analisis dan memberikan konteks yang lebih komprehensif, dilakukan penggabungan dataset PNS dengan data tambahan dari berbagai sumber. Pertama, dataset PNS digabungkan dengan data jumlah penduduk Provinsi Sumatera Selatan untuk menghitung rasio PNS per 10.000 penduduk menggunakan formula:

$$\text{Rasio_PNS_Penduduk} = (\text{Total_PNS_Provinsi} / \text{Jumlah_Penduduk}) \times 10.000.$$

Penggabungan ini memungkinkan analisis mengenai kecukupan jumlah PNS relatif terhadap jumlah penduduk yang dilayani, sehingga dapat memberikan gambaran apakah jumlah pegawai di suatu instansi sudah memadai untuk melayani populasi atau masih memerlukan penambahan formasi. Kedua, dilakukan merging dengan dataset klasifikasi instansi berdasarkan fungsi dan bidang tugas untuk memudahkan analisis berdasarkan kategori layanan. Instansi-instansi dikategorikan menjadi empat kelompok utama: kategori Pelayanan Langsung yang meliputi Dinas Pendidikan, Dinas Kesehatan, dan RS Dr. Ernaldi Bahar yang bertugas memberikan layanan langsung kepada masyarakat; kategori Infrastruktur & Pembangunan yang meliputi Dinas PU, Dinas Perumahan, dan Dinas Pengelolaan Sumber Daya Air yang fokus pada pembangunan fisik dan infrastruktur; kategori Ekonomi & Kesejahteraan yang meliputi Dinas Perdagangan, Dinas Perindustrian, dan Dinas Koperasi yang bertugas mengembangkan perekonomian daerah; serta kategori Administrasi & Pengelolaan yang meliputi Setda, Sekretariat DPRD, dan Inspektorat yang menjalankan fungsi koordinasi dan pengawasan pemerintahan. Proses merging ini sangat bermanfaat untuk melakukan analisis komparatif antar kelompok instansi, mengidentifikasi pola distribusi pegawai berdasarkan fungsi layanan, serta mengevaluasi alokasi sumber daya manusia di berbagai sektor pemerintahan. Hasil

penggabungan data ini memberikan insight yang lebih kaya untuk perencanaan kebijakan kepegawaian yang lebih strategis dan berbasis pada kebutuhan nyata masing-masing sektor pelayanan publik di Provinsi Sumatera Selatan.

4.2.6. Reshaping Data

Proses reshaping dilakukan untuk mengubah struktur data agar sesuai dengan kebutuhan analisis dan visualisasi tertentu. Pertama, dilakukan transformasi dari format wide ke format long, dimana data awal yang memiliki kolom terpisah untuk setiap jenis kelamin (Pria dan Wanita) diubah menjadi format panjang dengan satu kolom untuk jenis kelamin dan satu kolom untuk jumlah pegawai. Sebagai contoh, data Setda Provinsi yang awalnya dalam format wide dengan struktur: Instansi "Setda Provinsi", Pria "185", Wanita "149", Total "334", diubah menjadi format long dengan dua baris data: baris pertama berisi Instansi "Setda Provinsi", Jenis_Kelamin "Pria", Jumlah "185", dan baris kedua berisi Instansi "Setda Provinsi", Jenis_Kelamin "Wanita", Jumlah "149" sebagaimana ditampilkan pada Tabel 6 dan Tabel 7. Format long ini sangat memudahkan pembuatan visualisasi perbandingan gender menggunakan grouped bar chart, line plot, atau grafik lainnya yang memerlukan data dalam format panjang, serta memfasilitasi analisis statistik yang lebih kompleks seperti analisis varians atau regresi dengan jenis kelamin sebagai variabel kategorikal. Kedua, untuk keperluan analisis komparatif antar instansi dan pelaporan yang lebih ringkas, format long dapat diubah kembali ke format wide dengan menambahkan variabel kalkulasi tambahan seperti persentase gender, rasio pria terhadap wanita, dan kategori dominasi gender. Proses reshaping ini memberikan fleksibilitas dalam mengolah dan menyajikan data sesuai dengan kebutuhan analisis spesifik, memastikan bahwa struktur data selalu optimal untuk setiap jenis analisis yang dilakukan, dan memudahkan integrasi dengan berbagai tools visualisasi data modern yang masing-masing memiliki preferensi format data yang berbeda.

Tabel 6. Format Wide (Awal)

Instansi	Pria	Wanita	Total
Setda Provinsi	185	149	334

Tabel 7. Format Long (Hasil Reshaping)

Instansi	Jenis_Kelamin	Jumlah
Setda Provinsi	Pria	185
Setda Provinsi	Wanita	149

4.3. Analisis Statistik Deskriptif

4.3.1. Statistik

Berdasarkan hasil analisis statistik deskriptif terhadap data PNS Provinsi Sumatera Selatan, ditemukan beberapa karakteristik penting yang menggambarkan pola distribusi kepegawaian di 41 instansi pemerintahan. Pertama, distribusi data menunjukkan pola yang tidak normal dengan kecenderungan right-skewed (miring ke kanan), dimana nilai mean yang jauh lebih besar dari median mengindikasikan bahwa sebagian besar instansi memiliki jumlah pegawai yang relatif kecil, namun terdapat beberapa instansi besar yang menjadi outlier dan menarik nilai rata-rata ke atas. Pola distribusi ini mencerminkan realitas bahwa mayoritas instansi pemerintah provinsi merupakan instansi dengan skala kecil hingga menengah, sementara hanya segelintir instansi seperti Dinas Pendidikan dan Dinas Kesehatan yang memiliki jumlah pegawai sangat besar karena sifat layanan mereka yang langsung menyentuh masyarakat luas. Kedua, variabilitas data yang sangat tinggi terlihat dari nilai standar deviasi yang sangat besar relatif terhadap nilai mean, menunjukkan adanya heterogenitas yang signifikan dalam ukuran instansi pemerintahan. Hal ini mengindikasikan bahwa kebutuhan pegawai di setiap instansi sangat bervariasi tergantung pada fungsi, cakupan layanan, dan kompleksitas tugas yang diemban oleh masing-masing instansi, sehingga tidak dapat digeneralisasi dengan menggunakan satu standar yang sama untuk semua instansi. Ketiga, dari perspektif komposisi gender, analisis menunjukkan bahwa secara keseluruhan terdapat dominasi PNS wanita dengan nilai mean sebesar 199,32 pegawai per instansi, jauh lebih tinggi dibandingkan PNS pria yang memiliki mean sebesar 118,95 pegawai per instansi, menghasilkan rasio 1,68:1. Dominasi wanita ini terutama terjadi pada instansi-instansi besar di sektor pendidikan dan kesehatan yang secara tradisional memang menjadi bidang dengan partisipasi perempuan yang tinggi. Temuan-temuan statistik deskriptif ini memberikan pemahaman mendalam tentang karakteristik distribusi kepegawaian di Provinsi Sumatera Selatan dan menjadi dasar penting untuk perencanaan kebijakan kepegawaian yang lebih tepat sasaran, termasuk dalam hal redistribusi pegawai, penambahan formasi, dan upaya peningkatan kesetaraan gender di berbagai sektor pemerintahan.

4.3.2. Analisis Distribusi Berdasarkan Kategori

Tabel 8. Distribusi Instansi Berdasarkan Ukuran

Kategori	Jumlah Instansi	Persentase	Total PNS	% Total PNS
Sangat Besar	1	2,4%	7.775	59,6%
Besar	1	2,4%	810	6,2%
Sedang	6	14,6%	1.733	13,3%
Kecil	33	80,6%	2.731	20,9%
Total	41	100%	13.049	100%

Mayoritas instansi, yaitu sekitar 80,6%, tergolong sebagai instansi berukuran kecil, namun kelompok ini hanya menampung 20,9% dari total PNS. Sebaliknya, Dinas Pendidikan menjadi satu-satunya instansi yang masuk kategori sangat besar dan menyerap hampir 60% dari seluruh PNS di tingkat provinsi. Hal ini menunjukkan adanya konsentrasi yang sangat tinggi dalam penempatan PNS pada sektor pendidikan dibandingkan sektor lainnya.

Tabel 9. Distribusi Instansi Berdasarkan Status Gender

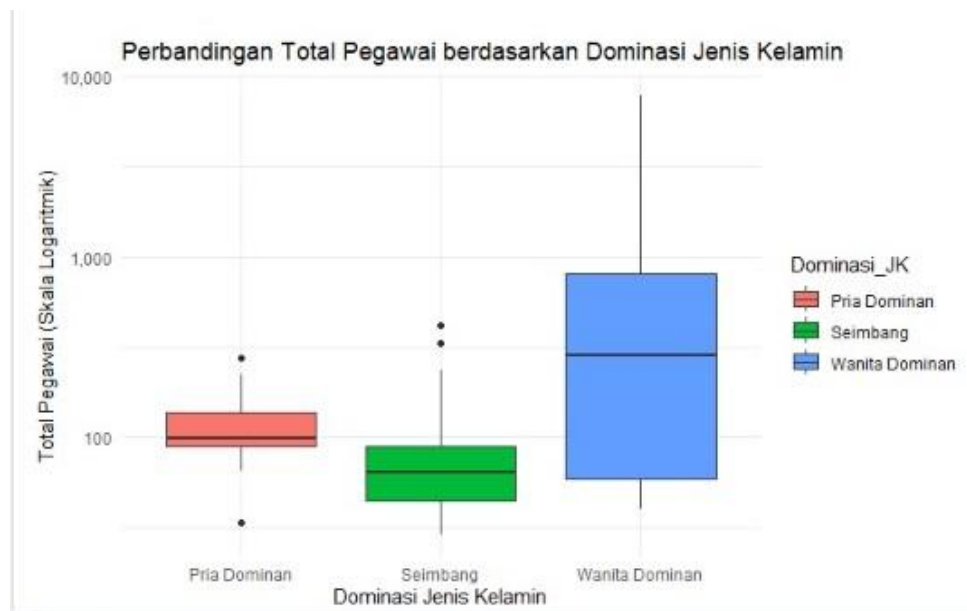
Status Gender	Jumlah Instansi	Persentase
Dominasi Pria	9	22,0%
Dominasi Wanita	13	31,7%
Seimbang	19	46,3%

Total	41	100%
-------	----	------

Analisis gender menunjukkan bahwa hampir setengah dari seluruh instansi (46,3%) memiliki komposisi gender yang seimbang. Selain itu, jumlah instansi yang didominasi wanita (31,7%) lebih besar dibandingkan instansi yang didominasi pria (22,0%). Temuan ini mengindikasikan adanya kemajuan dalam upaya mencapai kesetaraan gender di sektor kepegawaian publik.

4.4. Visualisasi Data

Visualisasi ini bertujuan untuk melihat bagaimana ukuran instansi (diukur dari Total Pegawai) didistribusikan pada kelompok-kelompok berdasarkan komposisi gendernya. Sumbu Y menggunakan Skala Logaritmik (Log-10) untuk meredam perbedaan ekstrem dalam jumlah pegawai dan memudahkan perbandingan distribusi.



Gambar 1. Perbandingan Total Pegawai berdasarkan Dominasi Jenis Kelamin (Skala Logaritmik)

Hasil visualisasi Boxplot ini mengungkapkan perbedaan distribusi yang jelas pada ukuran instansi di tiga kategori dominasi gender: Pria Dominan, Seimbang, dan Wanita Dominan.

A. Perbandingan Ukuran Pusat (Median)

Nilai median (garis tebal di dalam kotak) menunjukkan kecenderungan ukuran instansi pada kelompok masing-masing. Terlihat bahwa:

- Wanita Dominan menunjukkan nilai median tertinggi, berada di kisaran 300–400 pegawai. Ini mengindikasikan bahwa mayoritas instansi yang didominasi oleh perempuan cenderung memiliki jumlah total pegawai yang paling besar.
- Sebaliknya, kelompok Pria Dominan dan Seimbang memiliki nilai median yang jauh lebih rendah, yaitu di kisaran 80–100 pegawai. Hal ini menunjukkan bahwa instansi-instansi tersebut secara umum berukuran lebih kecil.

B. Variabilitas dan Sebaran Data

Jangkauan Interkuartil (IQR), yang diwakili oleh panjang kotak, menunjukkan sebaran 50% data tengah dari jumlah pegawai:

- Kelompok Wanita Dominan memiliki IQR yang paling lebar. Ini menyiratkan bahwa, selain memiliki rata-rata pegawai yang tinggi, instansi-instansi dalam kelompok ini juga memiliki variasi ukuran yang paling besar, mulai dari instansi yang relatif sedang hingga instansi yang sangat masif.
- Kelompok Pria Dominan dan Seimbang menunjukkan IQR yang lebih sempit, menandakan bahwa ukuran instansi dalam kelompok ini lebih seragam atau homogen.

C. Kehadiran Pencilan (Outlier)

Keberadaan titik-titik pencilan di atas whisker menunjukkan adanya instansi-instansi ekstrem yang memiliki jumlah pegawai jauh di atas kuartil ketiga (Q3) dari kelompoknya. Hal ini mengindikasikan bahwa meskipun sebagian besar instansi cenderung kecil atau sedang, ada sejumlah kecil instansi yang sangat besar di semua kategori dominasi gender.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil kajian terhadap data Pegawai Negeri Sipil (PNS) di wilayah Sumatera bagian Selatan tahun 2024 serta penerapan teknik data wrangling, dapat disimpulkan bahwa proses pengolahan data merupakan tahap yang sangat penting sebelum melakukan analisis statistik maupun penyusunan kebijakan kepegawaian. Data PNS yang diperoleh dari berbagai instansi masih memiliki sejumlah masalah seperti ketidakkonsistenan penulisan, data ganda, nilai hilang, serta format data yang tidak seragam. Jika tidak dibersihkan, kondisi ini berpotensi menghasilkan analisis yang bias dan keputusan yang keliru.

Melalui penerapan enam tahapan teknik data wrangling—variable creation, subsetting, sorting, recoding, merging, dan reshaping—data mentah berhasil diubah menjadi dataset yang lebih rapi, terstruktur, dan siap dianalisis. Proses ini memungkinkan peneliti untuk mendapatkan gambaran yang lebih jelas mengenai karakteristik kepegawaian, seperti distribusi pegawai berdasarkan instansi, proporsi pegawai laki-laki dan perempuan, hingga potensi ketidakseimbangan jumlah pegawai antar unit kerja.

Hasil akhir menunjukkan bahwa data yang telah melalui proses wrangling memiliki kualitas yang jauh lebih baik, baik dalam hal keteraturan, keutuhan, maupun keseragaman format. Dataset yang bersih ini dapat digunakan sebagai dasar yang kuat bagi pemerintah daerah untuk merumuskan kebijakan yang berbasis data, seperti perencanaan kebutuhan pegawai baru, pemerataan distribusi pegawai, penyusunan program pelatihan, serta penguatan kebijakan kesetaraan gender dalam sektor kepegawaian.

Secara keseluruhan, penelitian ini menegaskan bahwa penerapan teknik data wrangling bukan hanya sekadar kegiatan teknis, tetapi merupakan langkah strategis dalam meningkatkan kualitas informasi yang digunakan dalam manajemen kepegawaian dan pengambilan keputusan publik.

5.2. Saran

1. Standarisasi dan Perbaikan Sistem Pengelolaan Data

Pemerintah daerah perlu menetapkan standar penulisan, format, dan struktur data kepegawaian yang seragam di seluruh instansi supaya proses pengolahan data lebih mudah, cepat, dan minim kesalahan.

2. Peningkatan Kapasitas Pengelola Data

Perlu dilakukan pelatihan rutin bagi staf kepegawaian mengenai teknik data wrangling, manajemen data modern, serta penggunaan perangkat lunak analisis agar kualitas data semakin baik dan dapat langsung digunakan untuk pengambilan keputusan.

3. Pemanfaatan Data untuk Perencanaan Kebijakan yang Berkelanjutan

Data PNS yang telah dibersihkan harus digunakan secara aktif untuk evaluasi kebutuhan pegawai, pemerataan distribusi, penyusunan formasi baru, serta penyusunan kebijakan kesetaraan gender dan peningkatan kinerja aparatur.

DAFTAR PUSTAKA

- Anggraini, D. (2022). Teknik Recoding Data dalam Pengolahan Data Statistik. *Jurnal Analisis Data*, 5(2), 112-125.
- Badan Kepegawaian Negara. (2023). *Statistik PNS Indonesia*. Jakarta: BKN.
- Fitriani, R., & Nugroho, A. (2023). Pembersihan Data dalam Proses Data Wrangling. *Jurnal Teknologi Informasi*, 8(1), 45-58.
- Hidayat, R. (2023). Pembuatan Variabel Baru dalam Analisis Data. *Jurnal Statistika Terapan*, 6(3), 201-215.
- Kusuma, H. (2022). Tahapan Data Wrangling dalam Analisis Big Data. *Prosiding Seminar Nasional Teknologi Informasi*, 156-163.
- Lestari, P. (2022). Subsetting Data untuk Efisiensi Analisis Statistik. *Jurnal Komputasi Statistika*, 4(2), 78-89.
- Nurhayati, S. (2023). Penggabungan Data dari Berbagai Sumber untuk Analisis Terintegrasi. *Jurnal Sistem Informasi*, 9(1), 34-47.
- Peraturan Pemerintah Nomor 11 Tahun 2017 tentang Manajemen Pegawai Negeri Sipil.
- Prabowo, A. (2023). Pengurutan Data dalam Proses Analisis Statistik. *Jurnal Matematika dan Statistika*, 7(2), 134-145.
- Pratama, W., & Sari, D. (2022). *Data Wrangling: Konsep dan Implementasi*. Bandung: Informatika Press.
- Raharjo, B. (2023). *Pengantar Data Wrangling untuk Analisis Data Modern*. Jakarta: Media Komputindo.
- Ramadhani, F. (2023). Peran Pegawai Negeri Sipil dalam Pelayanan Publik. *Jurnal Administrasi Publik*, 11(2), 89-102.
- Santoso, I. (2023). Efisiensi Pengolahan Data Melalui Data Wrangling. *Jurnal Teknologi dan Sistem Informasi*, 5(1), 23-36.
- Setiawan, D. (2022). Reshaping Data: Transformasi Struktur Data untuk Analisis. *Jurnal Ilmu Komputer*, 10(3), 167-180.
- Suhendra, A. (2022). Analisis Statistik Deskriptif untuk Data Kepegawaian. *Jurnal Statistika dan Aplikasi*, 6(2), 145-158.

Undang-Undang Nomor 5 Tahun 2014 tentang Aparatur Sipil Negara.

Wijaya, T. (2021). Pentingnya Data Wrangling dalam Siklus Analisis Data. *Jurnal Sains Data*, 3(1), 12-25.