

ANALISIS DAN OPTIMASI KUALITAS DATASET PRODUKSI DURIAN PROVINSI LAMPUNG (2021–2024) MENGGUNAKAN TEKNIK MUNGING DAN WRANGLING DALAM KOMPUTASI STATISTIK



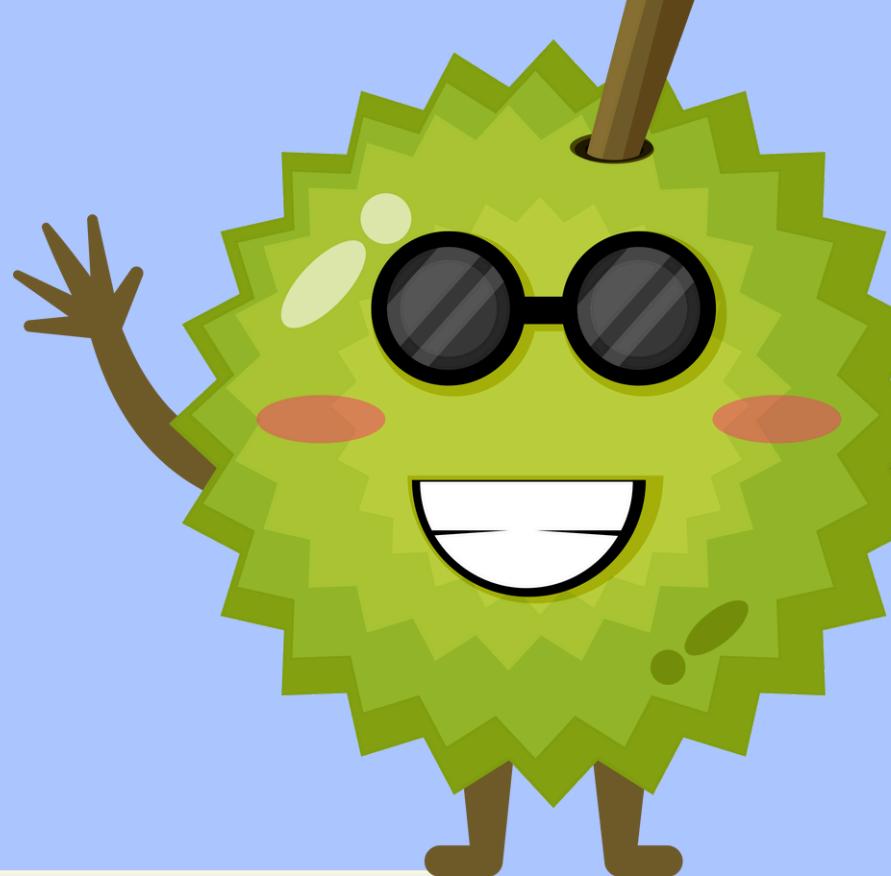
TUGAS BESAR KOMPUTASI STATISTIK



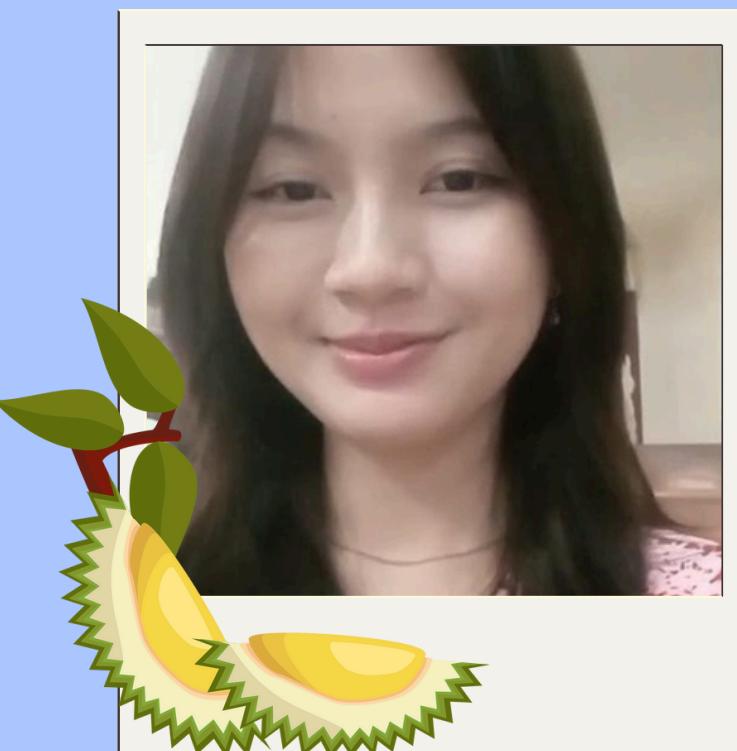
Oleh Kelompok 9 RB



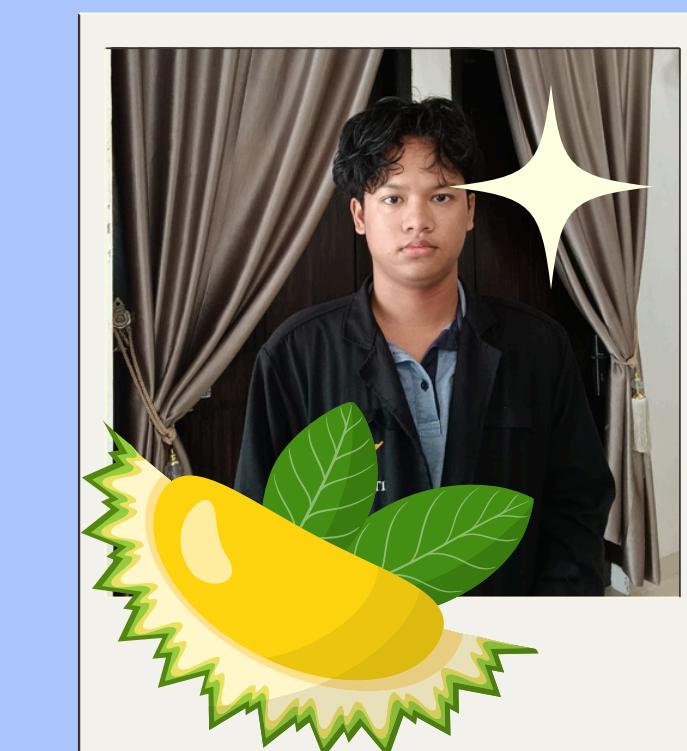
ANGGOTA KELOMPOK KAMI



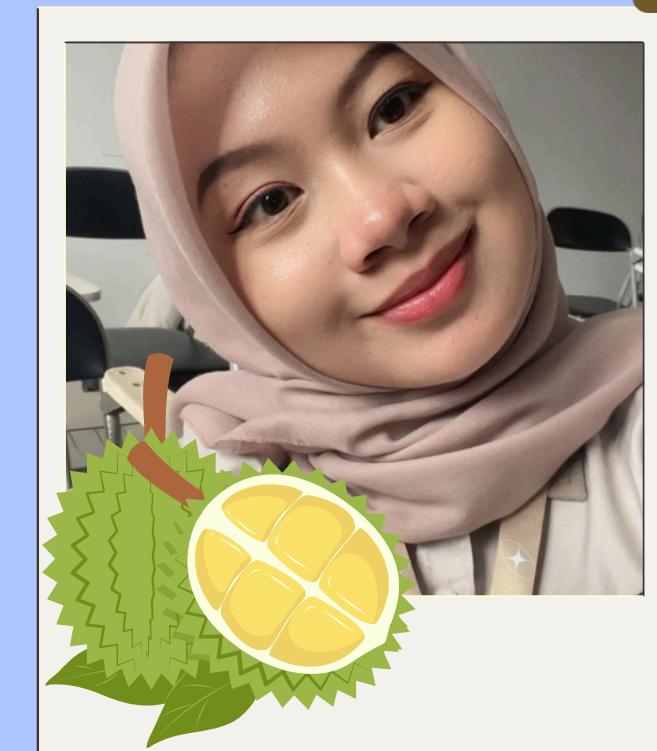
DEVYNA SONYA PALUPI S.
123450007



KEN GRACYA WAOMA
123450045



ALI ARISTO M. PARISI
123450088



HANIFAH INAYA SANI
123450123

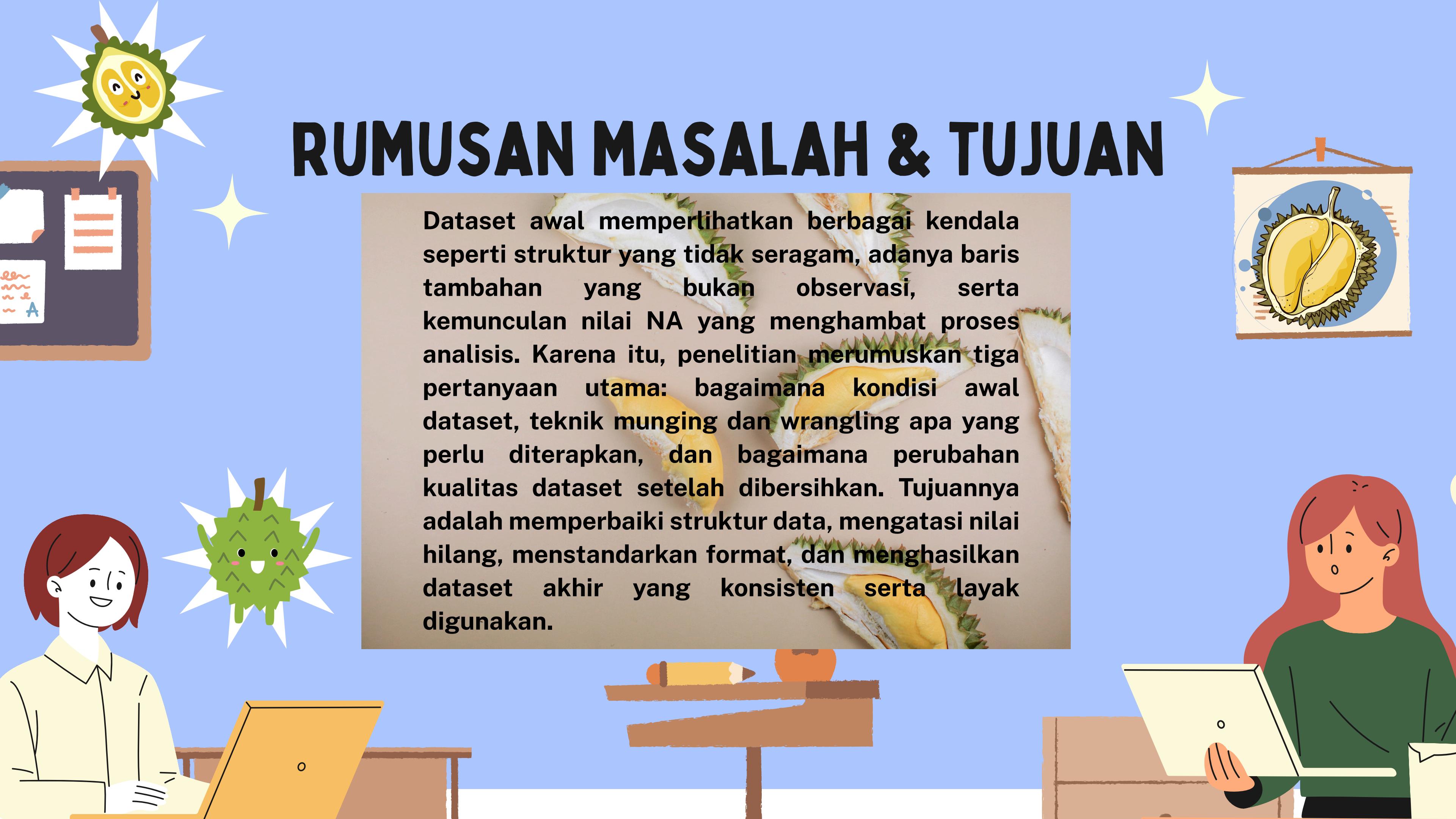
LATAR BELAKANG

Produksi durian di Lampung terus meningkat dan menjadi komoditas penting bagi sektor hortikultura daerah. Namun, data produksi yang tersedia masih banyak bermasalah: terdapat nilai hilang, perbedaan format antar tahun, baris catatan yang tidak relevan yang mengganggu akurasi analisis. Karena kondisi ini dapat menyebabkan hasil analisis menjadi keliru, penelitian dilakukan untuk membersihkan dan menata ulang dataset produksi durian 2021–2024 agar lebih konsisten, rapih.



RUMUSAN MASALAH & TUJUAN

Dataset awal memperlihatkan berbagai kendala seperti struktur yang tidak seragam, adanya baris tambahan yang bukan observasi, serta kemunculan nilai NA yang menghambat proses analisis. Karena itu, penelitian merumuskan tiga pertanyaan utama: bagaimana kondisi awal dataset, teknik munging dan wrangling apa yang perlu diterapkan, dan bagaimana perubahan kualitas dataset setelah dibersihkan. Tujuannya adalah memperbaiki struktur data, mengatasi nilai hilang, menstandarkan format, dan menghasilkan dataset akhir yang konsisten serta layak digunakan.



DATASET

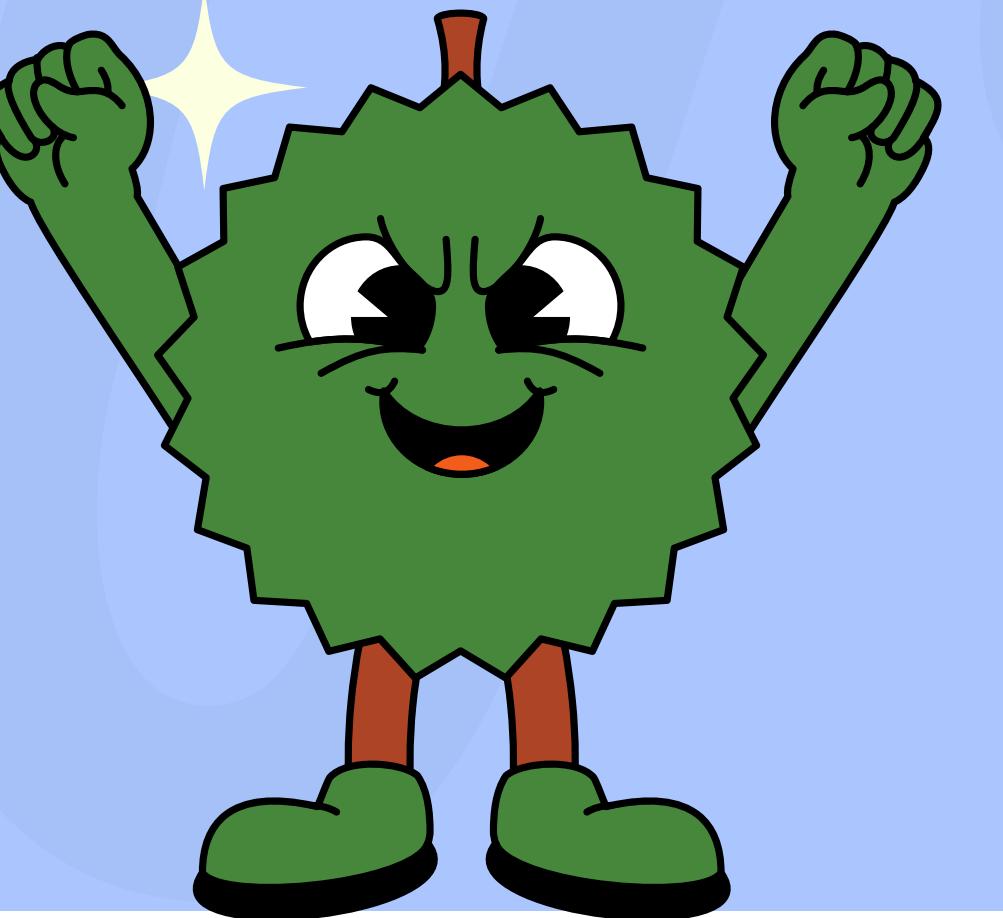
Dataset penelitian berasal dari BPS Provinsi Lampung dan berisi informasi produksi durian per kabupaten/kota untuk periode 2021 hingga 2024. Pada kondisi awal, dataset masih bercampur dengan baris yang tidak relevan seperti "Provinsi Lampung", catatan tambahan, angka tetap, serta baris komoditas lain seperti melinjo dan petai yang tidak berkaitan dengan durian. Selain itu, masih ditemukan nilai NA dan format yang berbeda antar tahun, sehingga sebelum dianalisis dataset perlu dipilih ulang, dibersihkan, dan diseragamkan terlebih dahulu.

VARIABEL

Variabel utama yang digunakan dalam penelitian ini adalah nama wilayah (kabupaten/kota) dan jumlah produksi durian untuk masing-masing tahun 2021, 2022, 2023, dan 2024. Pada proses wrangling, variabel produksi diseragamkan menjadi satu kolom bernama Produksi, sementara variabel tahun ditambahkan untuk menandai periode masing-masing data. Dengan penataan ulang ini, seluruh variabel memiliki struktur yang konsisten sehingga memudahkan analisis tren, perbandingan antar wilayah, serta proses visualisasi.

KONSEP MUNGING & WRANGLING

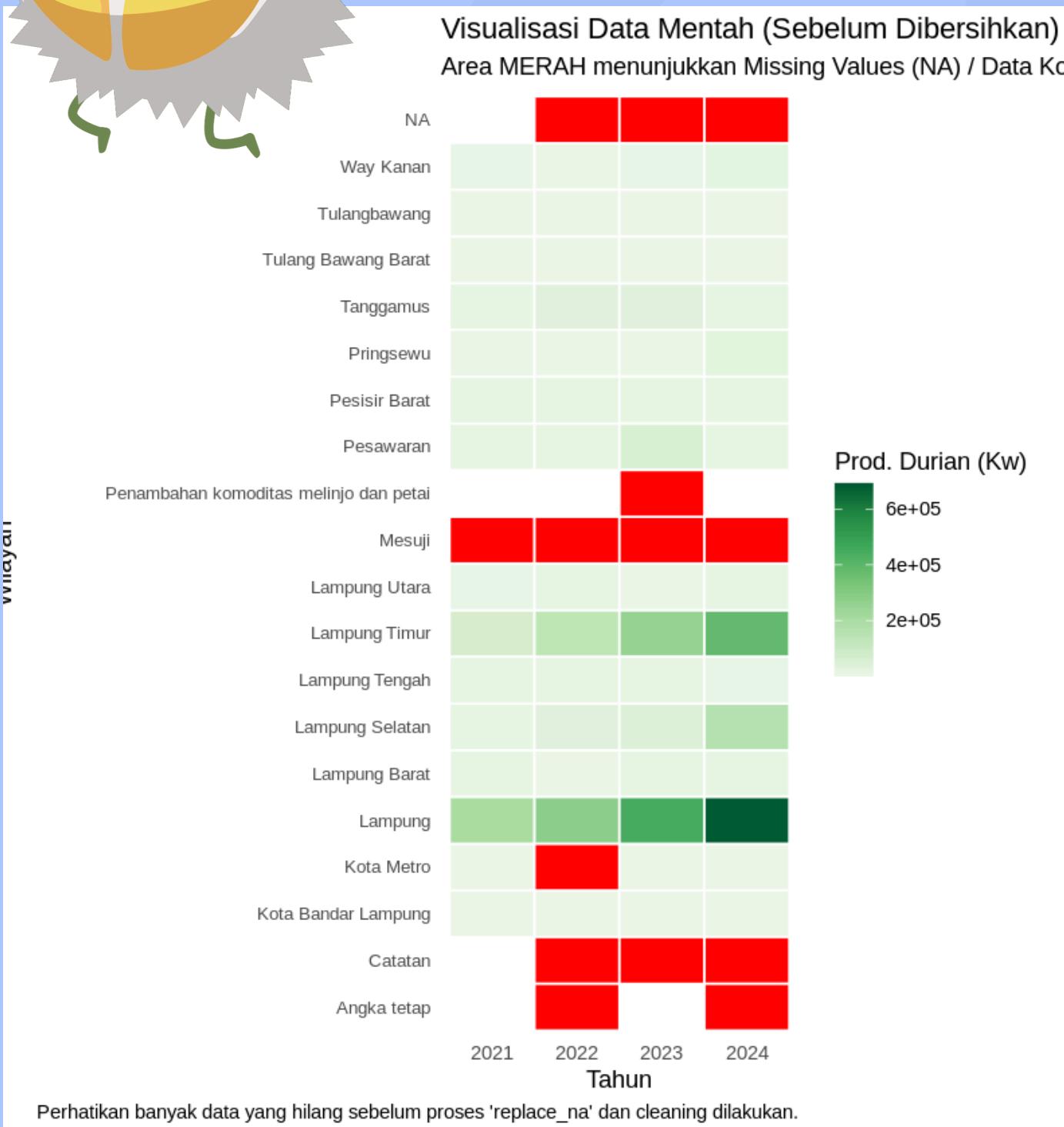
Proses munging dan wrangling dilakukan untuk merapikan empat dataset produksi durian tahun 2021–2024 yang awalnya masih berisi baris tidak relevan, format berbeda, dan nilai NA. Melalui fungsi bersihkan_data(), kolom yang diperlukan dipilih, nama variabel diseragamkan, baris yang tidak penting dibuang, dan nilai kosong diganti dengan 0. Setelah seluruh dataset dibersihkan, data digabungkan sehingga menghasilkan dataset akhir yang rapih.



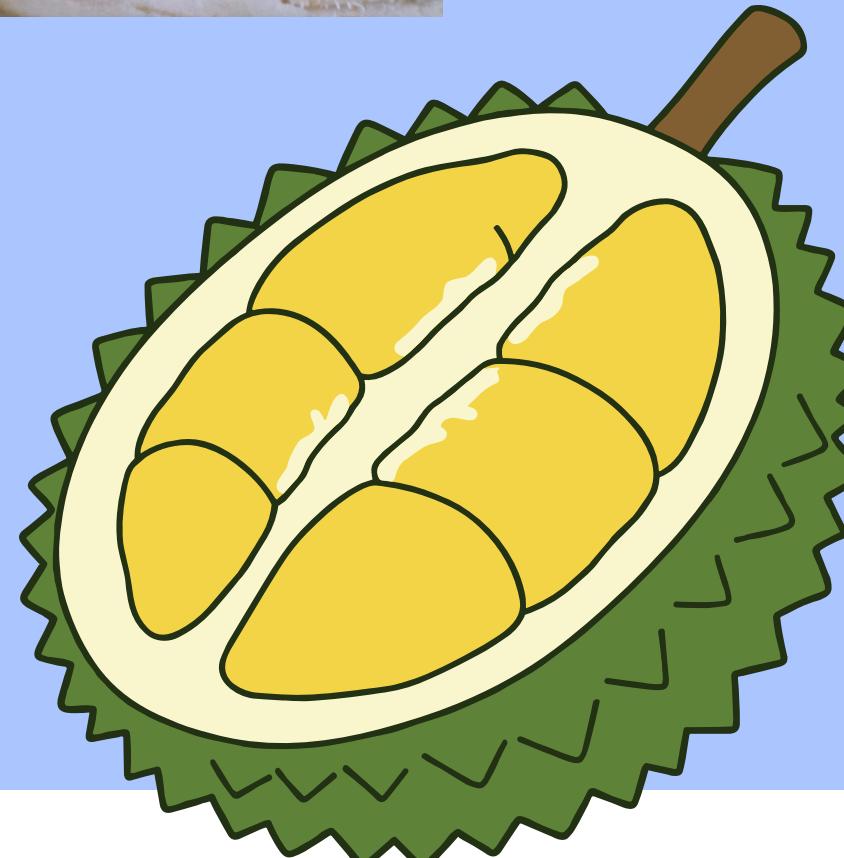
VISUALISASI SEBELUM WRANGLING



wrangling

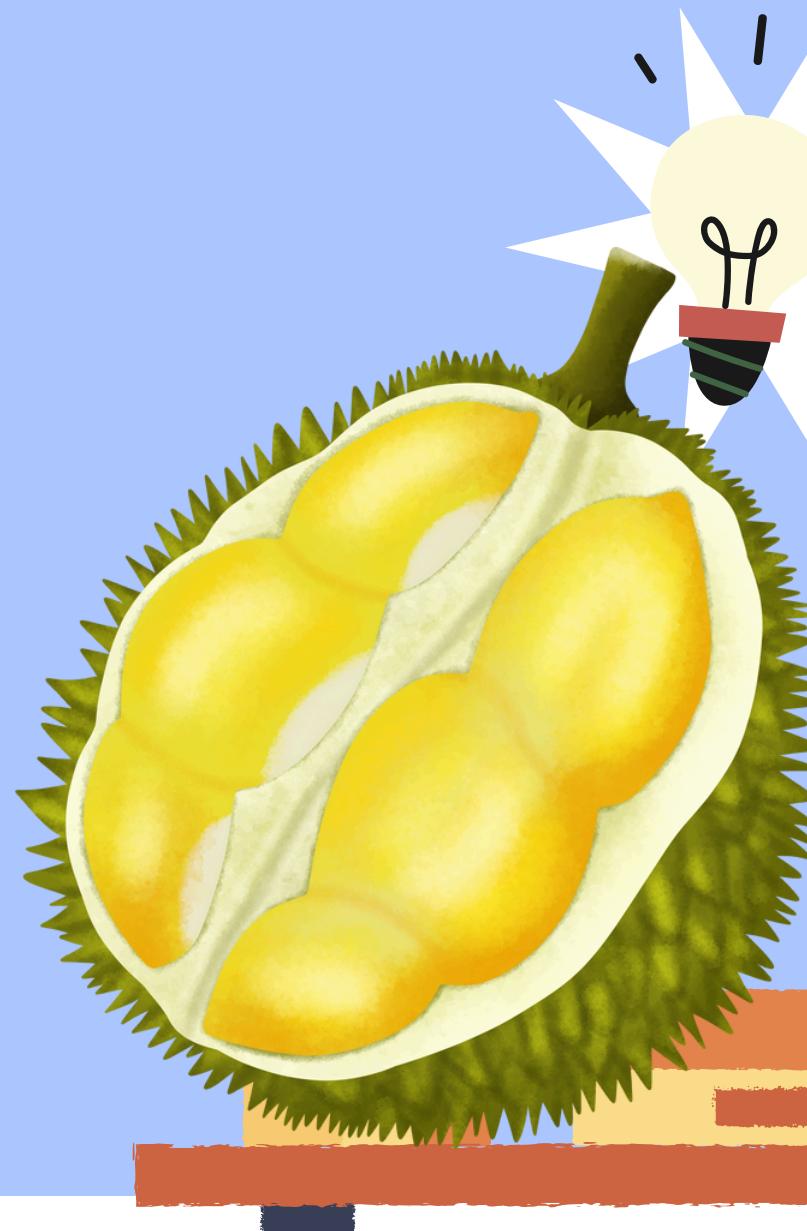


Visualisasi sebelum wrangling dilakukan untuk melihat kondisi awal data apa adanya, sebelum melalui proses pembersihan. Tahap ini membantu mengidentifikasi masalah seperti nilai kosong, baris yang tidak relevan, atau format yang tidak konsisten. Dengan memahami kondisi data sejak awal, proses munging dan wrangling dapat dijalankan dengan lebih tepat dan terarah.



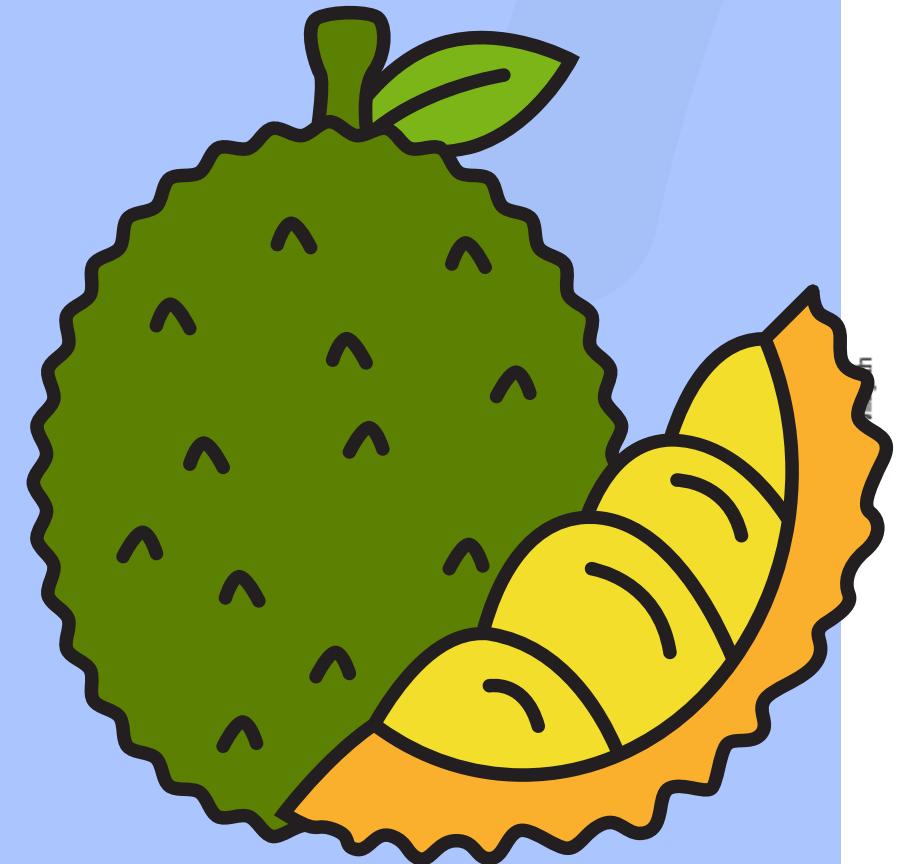
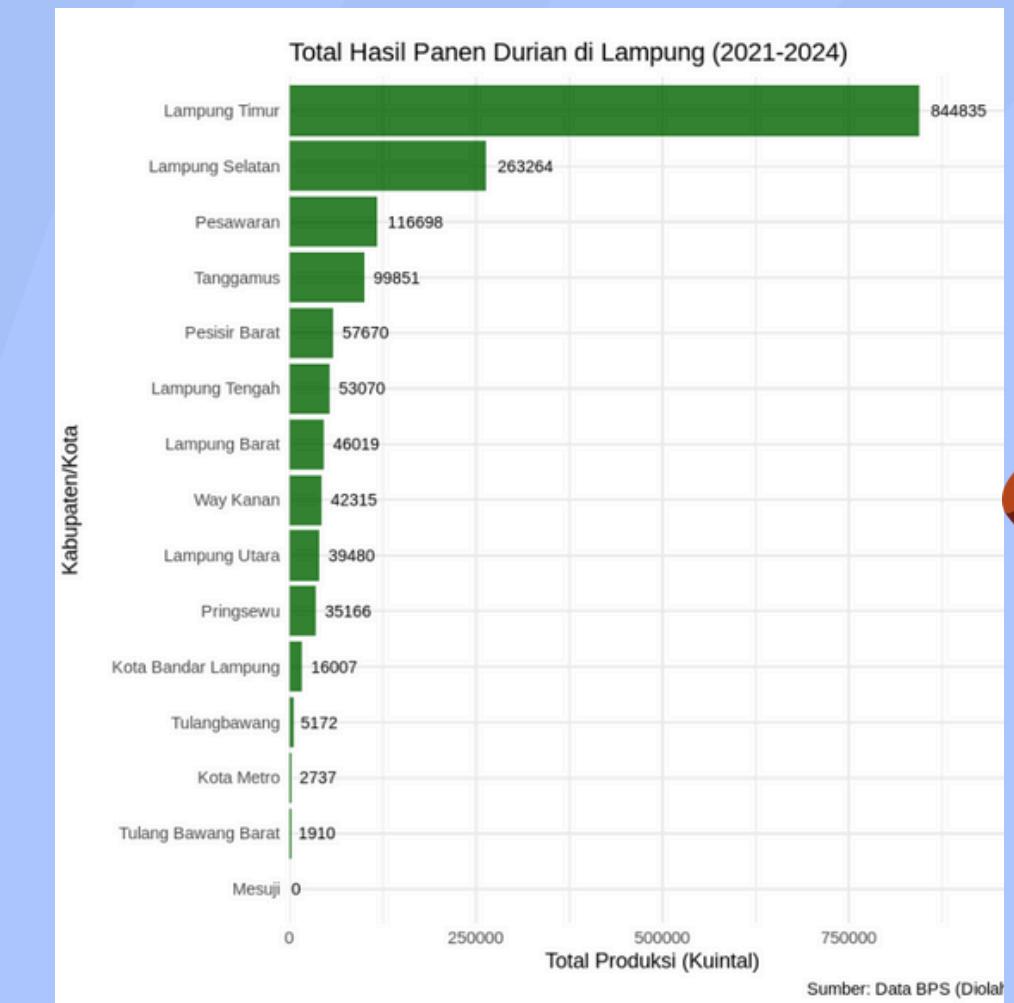
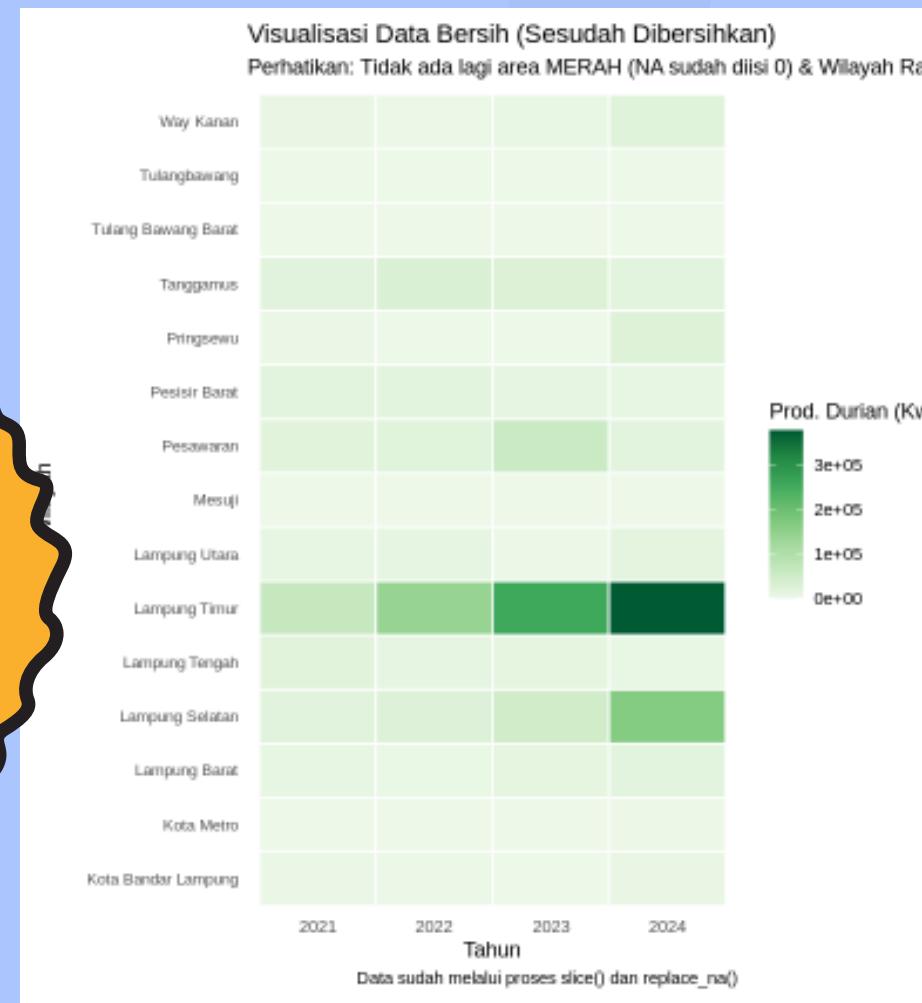
PROSES MUNGING & WRANGLING

Visualisasi sebelum wrangling dilakukan untuk melihat kondisi awal data apa adanya, sebelum melalui proses pembersihan. Tahap ini membantu mengidentifikasi masalah seperti nilai kosong, baris yang tidak relevan, atau format yang tidak konsisten. Dengan memahami kondisi data sejak awal, proses munging dan wrangling dapat dijalankan dengan lebih tepat dan terarah.

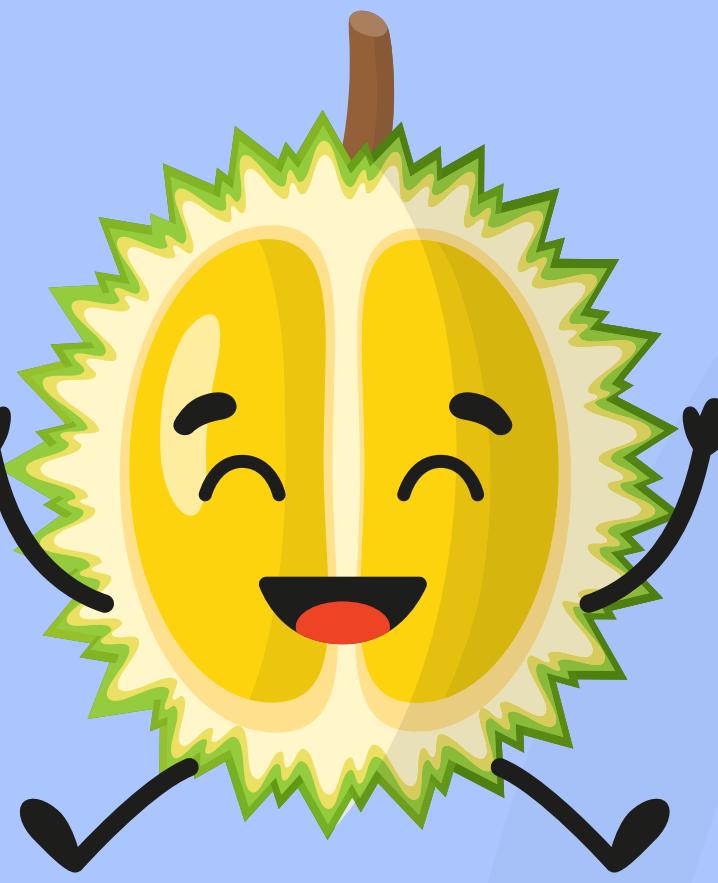


VISUALISASI SESUDAH & PENJELASAN HASIL

Visualisasi sesudah wrangling digunakan untuk menampilkan data yang sudah bersih, rapi, dan siap dianalisis. Pada tahap ini, grafik atau tabel yang dihasilkan memberikan gambaran yang lebih jelas tentang pola produksi, perbandingan antar wilayah, maupun perubahan dari tahun ke tahun. Penjelasan hasil dibutuhkan untuk menginterpretasikan temuan tersebut agar pembaca dapat memahami makna dan insight yang diperoleh dari proses analisis.



Produksi durian di Lampung sepanjang 2021–2024 terlihat sangat timpang, dengan Lampung Timur menjadi penyumbang terbesar jauh melampaui daerah lain. Lampung Selatan dan Pesawaran masih menunjukkan kapasitas produksi yang cukup kuat, sementara kabupaten seperti Tulang Bawang Barat, Kota Metro, dan terutama Mesuji berada pada level yang sangat rendah. Kondisi ini menunjukkan bahwa aktivitas budidaya durian masih terkonsentrasi di wilayah tertentu.



TERIMA KASIH

