

PROJECT DATA WRANGLING
ANALISIS MARGIN HARGA GABAH-BERAS DAN KAITANNYA
DENGAN PRODUKSI PADI DI INDONESIA PERIODE 2021–2024



Dosen Pengampu:

Ulfa Siti Nuraini, S. Stat., M.Stat.

Disusun Oleh:

Muh Argya Ishmat Islah Putra Masnyur

24031554042

Putri Buana Tungga Dewi Fortuna

1314623017

DATA WRANGLING

2025

DAFTAR ISI

DAFTAR ISI.....	1
BAB I PENDAHULUAN.....	2
1.1 Latar Belakang.....	2
1.2 Ide Project.....	4
1.3 Tujuan	4
1.4 Rumusan Masalah.....	4
1.5 Manfaat	4
1.6 Sumber Data.....	5
1.7 Metodologi.....	6
1.8 Tinjauan Pustaka.....	8
BAB II PEMBAHASAN	10
2.1 Proses Data Wrangling	10
2.2 Interpretasi dan Visualisasi Hasil Analisis	51
2.3 Kendala yang Dihadapi.....	52
2.4 Solusi yang Diterapkan.....	53
2.5 Rencana Tindak Lanjut.....	54
BAB III KESIMPULAN DAN SARAN	56
3.1 Kesimpulan	56
3.2 Saran	56
DAFTAR PUSTAKA	57

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia, negara agraris dengan 98,35% penduduknya menjadikan beras sebagai makanan pokok, tengah menghadapi paradoks ekonomi yang mencengangkan. Bayangkan sebuah ironi, dimana satu sisi, petani menjerit karena harga gabah yang terus merosot hingga Rp6.480 per kg pada Januari 2025, bahkan di bawah Harga Pembelian Pemerintah (HPP) yang seharusnya melindungi mereka. Di sisi lain, konsumen mengeluh karena harga beras premium meroket ke Rp15.560 per kg, melampaui daya beli masyarakat.

Pada Februari 2024, Indonesia mencatatkan rekor harga beras tertinggi sepanjang sejarah yaitu Rp18.000 per kg, dengan inflasi beras mencapai 15,65% *year-on-year*. Anomali ini bukan sekadar fluktuasi pasar biasa, ini adalah sinyal darurat dari sistem distribusi yang telah kehilangan kendali. Berikut adalah runutan perjalanan sebutir gabah dari sawah hingga ke piring konsumen:



Data terkini menunjukkan bahwa pada Januari 2025, Harga Pembelian Pemerintah (HPP) gabah ditetapkan Rp6.500 per kg, sementara harga beras premium di tingkat penggilingan mencapai Rp13.112 per kg. Namun pada Februari 2025, harga beras premium di tingkat konsumen melonjak menjadi Rp15.518 per kg, menciptakan margin yang mengkhawatirkan.

Fenomena lonjakan harga beras pada 2024 menghadirkan kondisi “anomali berlapis” yang membingungkan publik maupun pengambil kebijakan. Menteri Pertanian Andi Amran Sulaiman menyebut situasi ini sebagai sesuatu yang tidak lazim, sebab stok beras nasional berada pada posisi melimpah, tetapi harga di pasar tetap tinggi. Ia menegaskan bahwa “dulu harga tinggi karena stok tipis namun sekarang stok melimpah

tetapi harga tetap tinggi yang artinya ada yang bermain.” Temuan ini sejalan dengan laporan bahwa produksi beras nasional pada Januari–Maret 2024 meningkat 52% dibanding tahun sebelumnya, namun harga beras tidak menunjukkan penurunan signifikan. Kondisi tersebut menciptakan tiga paradoks sekaligus yaitu paradoks produksi (produksi meningkat, harga tidak turun), paradoks margin (petani merugi akibat harga gabah di bawah HPP sementara konsumen membeli beras dengan harga tinggi), serta paradoks distribusi (stok besar tetapi harga tetap mahal).

Dalam konteks inilah analisis margin harga gabah–beras menjadi semakin penting untuk dibedah. Margin bukan sekadar selisih angka dalam rantai pasok, melainkan indikator kesehatan struktural sistem pangan. Margin yang melebar menandakan adanya inefisiensi distribusi, potensi spekulasi harga, dan kegagalan kebijakan stabilisasi. Data historis juga menunjukkan bahwa pada April 2024 harga beras premium masih 17,4 persen lebih tinggi dibanding periode yang sama tahun sebelumnya, meski telah turun dari puncaknya. Fenomena ini perlu dianalisis secara komprehensif dengan melihat interaksi antara produksi padi, harga gabah, dan harga beras selama beberapa tahun terakhir.

Periode 2021–2024 dipilih karena mencerminkan dinamika ekonomi dan iklim yang sangat memengaruhi sektor pangan Indonesia. Tahun 2021–2022 merupakan masa pemulihan ekonomi pascapandemi COVID-19. Lalu 2022–2023 diwarnai dampak global perang Rusia–Ukraina yang memicu lonjakan harga pupuk dan komoditas dunia. Kemudian di 2023–2024 disertai fenomena El Niño yang menyebabkan kekeringan ekstrem. Dan terakhir pada 2024 Indonesia mencatat rekor harga beras tertinggi dalam sejarah. Kondisi-kondisi tersebut menjadikan periode ini sangat relevan untuk dianalisis melalui pendekatan data wrangling berbasis multi-sumber.

Berdasarkan urgensi isu dan kompleksitas dinamika harga pangan tersebut, kami memilih untuk menyusun Laporan Makalah UAS Data Wrangling dengan judul “Analisis Margin Harga Gabah–Beras Terhadap Produksi Padi di Indonesia Periode 2021–2024”. Melalui proses penggabungan, pembersihan, dan analisis data dari berbagai lembaga, makalah ini diharapkan dapat memberikan gambaran yang lebih jernih mengenai bagaimana produksi, harga gabah, dan harga beras saling memengaruhi, serta apakah margin harga mencerminkan sistem distribusi yang efisien atau justru menunjukkan adanya anomali struktural. Dengan demikian, proyek ini tidak hanya memenuhi tujuan akademik mata kuliah Data Wrangling, tetapi juga memberikan kontribusi analitis terhadap isu strategis pangan nasional.

1.2 Ide Project

Ide project ini terinspirasi dari dokumen Outlook Padi 2024 yang diterbitkan oleh Pusat Data dan Sistem Informasi Pertanian Kementerian Pertanian RI. Dalam dokumen tersebut (halaman 37), disebutkan bahwa:

"Perkembangan harga dua jenis komoditas antara tahun 2019-2023 masih mengalami peningkatan meskipun tidak terlalu signifikan yaitu sebesar 4,75% per tahun untuk harga gabah atau mencapai harga rata-rata Rp5.669, yang sebelumnya pertumbuhan harga beras padi dalam waktu yang sama."

Data ini menunjukkan adanya disparitas pertumbuhan harga antara gabah (di tingkat petani) dan beras (di tingkat konsumen). Belum lagi, perasaan kami sendiri yang terpikirkan untuk mengecek korelasi margin harga gabah-beras, mengingat harga beras makin hari-makin naik.

1.3 Tujuan

Tujuan dari penelitian ini adalah:

1. Menganalisis margin absolut (dalam Rp/Kg) dan margin persentase (%) antara harga gabah dan beras.
2. Mengidentifikasi *trend* temporal margin harga periode 2021-2024.
3. Mengevaluasi hubungan antara produksi padi nasional dengan margin harga.
4. Membandingkan margin harga antar kategori kualitas beras.
5. Memberikan insight data-driven untuk kebijakan harga pangan nasional.

1.4 Rumusan Masalah

Berdasarkan latar belakang di atas, penelitian ini akan menjawab pertanyaan:

1. Bagaimana pola distribusi dan statistik deskriptif margin harga antara gabah dan beras konsumen periode 2021-2024?
2. Bagaimana *trend* margin harga gabah-beras berubah dari tahun ke tahun?
3. Apakah terdapat korelasi antara volume produksi padi dengan margin harga gabah-beras?
4. Kategori beras mana yang memiliki margin tertinggi dan terendah?
5. Bagaimana perbedaan margin harga antar kategori kualitas beras (Premium, Medium, Kualitas Bawah I, dll)?

1.5 Manfaat

1. Manfaat Teoritis:

- a. Menambah literatur empiris tentang analisis margin harga komoditas pertanian di Indonesia.
 - b. Memberikan pemahaman tentang dinamika *pricing* dalam *value chain* beras.
2. Manfaat bagi lembaga terkait:
- a. Bagi Pemerintah: Menyediakan analisis data untuk bahan evaluasi atau *insight* tambahan.
 - b. Bagi Petani: Memberikan informasi tentang struktur harga pasar beras.
 - c. Bagi Akademisi: Menjadi referensi untuk penelitian lanjutan tentang ekonomi pertanian.

1.6 Sumber Data

Project ini menggunakan tiga sumber data resmi yang kredibel :

1. Data Produksi Padi (BPS)

- a. Sumber: Badan Pusat Statistik (BPS)

<https://www.bps.go.id/id/statistics-table/2/MTQ5OCMy/luas-panen-produksi-dan-produktivitas-padi-menurut-provinsi.html>

Data ini berisi informasi produksi padi tahunan untuk seluruh provinsi di Indonesia. Data digunakan untuk menggambarkan kondisi produksi padi nasional, tren perubahan produksi, serta potensi ketidakseimbangan antara produksi dan harga beras. BPS menyediakan data yang telah melalui proses verifikasi dan metodologi *Kerangka Sampel Area (KSA)* sehingga dapat dipercaya untuk analisis jangka panjang.

- b. Variabel:

- Luas Panen (hektar)
- Produktivitas (kuintal/hektar)
- Produksi (ton)
- Provinsi
- Tahun (2021-2024)

- c. Format: CSV dengan metadata di 3 baris pertama

2. Data Harga Beras Konsumen (Bank Indonesia)

- a. Sumber: Bank Indonesia - Sistem Pemantauan Harga Pangan

<https://www.bi.go.id/hargapangan/TabelHarga/PasarTradisionalDaerah>

Data ini berisi harga eceran beras di pasar tradisional di berbagai daerah. Data digunakan untuk menganalisis dinamika harga beras pada level konsumen,

termasuk fluktuasi musiman, kenaikan harga pada periode tertentu, serta keterkaitannya dengan produksi padi dan harga gabah. Informasi dari BI bersifat *real-time monitoring* sehingga mencerminkan kondisi pasar aktual.

b. Variabel:

- Kategori Beras (Premium, Medium, Kualitas Bawah I, II, III)
- Harga Beras (Rp/Kg)
- Tanggal/Bulan pengamatan
- Tahun (2021-2024)

c. Format: Excel (.xlsx) dengan struktur multi-header

3. Data Harga Gabah (Kementerian Pertanian)

a. Sumber: Pusat Data dan Sistem Informasi Pertanian

[https://satudata.pertanian.go.id/assets/docs/publikasi/STATISTIK_HARGA_KO
MODITAS_PERTANIAN_2024.pdf](https://satudata.pertanian.go.id/assets/docs/publikasi/STATISTIK_HARGA_KO
MODITAS_PERTANIAN_2024.pdf)

Data ini memuat perkembangan harga gabah di tingkat petani maupun penggilingan untuk berbagai kualitas gabah. Data tersebut penting untuk memahami hubungan antara harga di hulu (petani) dan harga di hilir (konsumen). Selain itu, informasi harga gabah menjadi indikator kesejahteraan petani serta sensitivitas pasar terhadap perubahan produksi

b. Variabel:

- Kualitas Gabah (Gabah Kering Panen/GKP, Gabah Kering Giling/GKG)
- Tingkat Harga (Petani, Penggilingan)
- Harga Gabah (Rp/Kg)
- Tahun (2021-2024*)

c. Format: PDF (diekstrak menjadi CSV)

1.7 Metodologi

1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode analisis data sekunder. Data dikumpulkan dari tiga sumber resmi pemerintah, kemudian diolah melalui proses data wrangling yang sistematis untuk menghasilkan *insight* tentang margin harga gabah-beras di Indonesia.

Metode Penelitian:

- a. Tipe : Kuantitatif Deskriptif
- b. Analisis : Analisis Data Sekunder

- c. Tools : Google Colab dan Package Python6
- d. Metode : Statistical Analysis, Time Series Analysis, Correlation Analysis

2. Metode Pengumpulan Data

Kami juga memiliki prosedur dalam pengumpulan data yang terdiri atas beberapa langkah, yaitu:

a. Download Data

- Mengakses portal resmi masing-masing lembaga (BPS, Bank Indonesia, Kementerian Pertanian)
- Mendownload file data sesuai periode penelitian (2021-2024)
- Menyimpan file dengan penamaan yang konsisten dan terstruktur

b. Organisasi Data

Data disimpan dalam folder sebagai berikut:

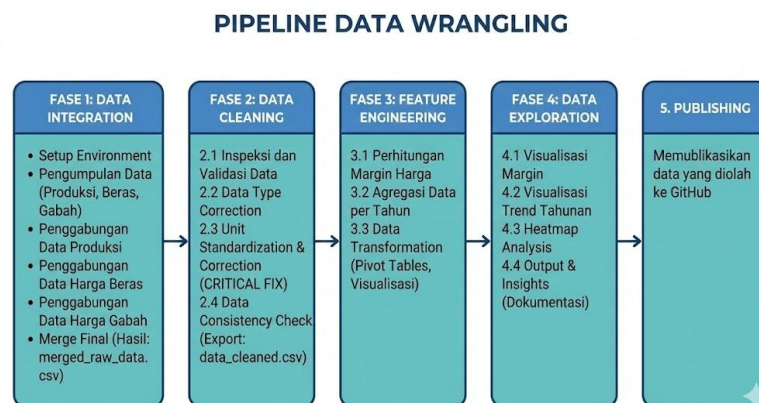
<https://drive.google.com/drive/u/0/folders/1HTKrxXNJCXhm-cdnWwhpwPOE39XYGaRK>

c. Validasi Data

- *Cross-check* dengan publikasi resmi terkait margin ataupun penelitian sejenis yang terbaru (Outlook Padi 2024)
- Verifikasi konsistensi temporal (memastikan tidak ada gap tahun)
- Validasi range nilai (deteksi anomali ekstrem)

3. Pipeline Data Wrangling

Proses pengolahan data dibagi menjadi 4 Fase Utama dan 1 Fase Publishing:



a. Data *Integration*

Mengintegrasikan tiga sumber data dengan format berbeda (CSV, Excel, PDF) menjadi satu dataset terpadu yang siap untuk analisis.

b. Data *Cleaning*

Memastikan data bersih, konsisten, dan siap untuk analisis dengan mengatasi masalah *missing values*, duplikasi, kesalahan tipe data, dan inkonsistensi unit.

c. *Feature Engineering*

Menghitung metrik analisis kunci, agregasi temporal, dan menghasilkan insight dari data yang telah dibersihkan.

d. *Data Exploration*

Fase ini berfokus pada mencari *insight* kunci melalui visualisasi, seperti tren harga vs. produksi tahunan dan distribusi margin harga antar kategori beras. Tujuannya adalah mengidentifikasi pola, anomali, dan hubungan korelasi yang menjawab pertanyaan inti dari proyek analisis ini.

e. *Data Publishing*

Fase akhir ini menjamin aksesibilitas dan reproduksibilitas hasil proyek dengan mengunggah dataset final, kode, dan visualisasi kunci ke *repository* publik (misalnya GitHub).

1.8 Tinjauan Pustaka

1. Produksi Padi dan Dinamika Pasar Beras di Indonesia

Produksi padi merupakan salah satu faktor fundamental yang menentukan stabilitas harga beras nasional. BPS (2024) menjelaskan bahwa perubahan produksi, terutama penurunan output akibat musim, iklim, atau gangguan distribusi, secara langsung memengaruhi ketersediaan beras di pasar domestik. Dalam konteks komoditas pangan strategis seperti beras, mekanisme permintaan–penawaran bekerja cukup jelas dimana penurunan produksi menyebabkan tekanan pada suplai sehingga harga beras cenderung meningkat. Kementerian Pertanian (2023) menegaskan bahwa sensitivitas harga beras terhadap perubahan produksi lebih tinggi dibandingkan komoditas pangan lain karena struktur rantai pasoknya yang panjang dan biaya distribusi yang besar.

2. Harga Gabah, Harga Beras, dan Margin Pemasaran

Margin pemasaran merupakan selisih antara harga yang diterima petani (harga gabah) dengan harga yang dibayar konsumen (harga beras). Menurut Kemendag (2023), margin mencerminkan biaya distribusi, penggilingan, penyimpanan, transportasi, serta struktur pasar yang memengaruhi efisiensi rantai pasok beras. Jika margin meningkat secara signifikan tanpa diikuti kenaikan biaya distribusi, hal tersebut dapat mengindikasikan adanya inefisiensi pasar atau asimetri informasi antar pelaku. Data Panel Harga Bapanas (2024) menunjukkan bahwa margin beras di Indonesia sangat sensitif terhadap perubahan suplai; ketika produksi turun, margin cenderung meningkat akibat dorongan harga pada tingkat konsumen yang lebih besar dibandingkan harga gabah. Dengan demikian, margin dapat menjadi indikator penting untuk menilai kinerja pasar beras dan kesejahteraan petani.

3. Hubungan Produksi dan Harga dari Perspektif Teori Ekonomi Pasar Pangan

Literatur ekonomi pangan menekankan bahwa fluktuasi produksi beras memiliki efek ganda terhadap pasar: menekan harga gabah di tingkat petani sekaligus mendorong kenaikan harga beras di tingkat konsumen. Firdaus (2011) dalam analisis deret waktu pangan menyebutkan bahwa perubahan output beras terutama memengaruhi perilaku pedagang penggilingan dan distributor, sehingga hubungan antara produksi dan margin sering bersifat *inverse*. Ketika produksi melimpah, harga gabah naik moderat sementara harga beras turun atau stagnan, menyebabkan margin menyempit. Sebaliknya, ketika produksi menurun, harga beras meningkat relatif lebih cepat dan margin melebar. Pola inilah yang juga diamati dalam berbagai laporan pangan nasional.

4. Analisis Korelasi dan Regresi dalam Penelitian Ekonomi

Metode korelasi dan regresi merupakan pendekatan statistik yang umum digunakan untuk menganalisis hubungan antar variabel ekonomi. Widarjono (2018) menjelaskan bahwa korelasi dapat digunakan untuk menguji kekuatan dan arah hubungan antara dua variabel kuantitatif, misalnya produksi dan margin, sedangkan regresi digunakan untuk memprediksi perubahan variabel dependen berdasarkan variabel independen. Sugiyono (2019) menambahkan bahwa dalam konteks penelitian kuantitatif, korelasi negatif yang kuat menandakan adanya hubungan *inverse* yang konsisten, dan keberadaan outlier perlu dianalisis secara substantif untuk memahami dinamika yang tidak tertangkap oleh pola umum. Dalam analisis harga pangan, korelasi negatif besar antara produksi dan margin merefleksikan mekanisme permintaan–penawaran yang bekerja kuat di pasar pangan nasional.

5. Penelitian Terdahulu tentang Margin Beras di Indonesia

Sejumlah penelitian terdahulu menunjukkan bahwa margin gabah–beras di Indonesia tidak hanya dipengaruhi oleh produksi, tetapi juga struktur pasar, kebijakan pemerintah, dan efisiensi distribusi. Studi BPS dan Kemendag menemukan bahwa fluktuasi margin sering terjadi pada tahun-tahun dengan gangguan pasokan atau perubahan kebijakan harga dasar pembelian gabah. Penelitian lokal oleh beberapa lembaga daerah menunjukkan bahwa kategori beras kelas bawah memiliki margin yang lebih volatil, karena segmen pasar ini sangat sensitif terhadap perubahan suplai dan biaya distribusi. Secara umum, literatur menunjukkan bahwa hubungan produksi dan margin bersifat kuat dan konsisten, namun anomali dapat muncul ketika intervensi kebijakan atau gangguan distribusi terjadi, sebagaimana kasus nasional pada tahun-tahun tertentu.

6. Konsep Volatilitas Harga dan Margin dalam Komoditas Pertanian

Volatilitas harga merupakan ukuran fluktuasi harga dari waktu ke waktu dan sering dikaitkan dengan risiko pasar. Pada komoditas pertanian, volatilitas dipengaruhi oleh musim, perubahan suplai, biaya input, dan faktor eksternal seperti distribusi atau kebijakan. BPS (2024) mencatat bahwa margin beras memiliki volatilitas yang lebih tinggi pada kategori beras kelas bawah, karena segmen konsumennya lebih sensitif terhadap perubahan harga. Kenaikan standar deviasi margin dari 2021 ke 2024 dalam data penelitian selaras dengan literatur yang menyebutkan bahwa ketidakpastian produksi dapat meningkatkan volatilitas margin, terutama pada rantai pasok yang panjang.

BAB II

PEMBAHASAN

2.1 Proses Data Wrangling

1. Data Integration

a. *Setup Environment* dan *Mounting* Google Drive

Proses data wrangling dalam penelitian ini dimulai dengan melakukan *setup environment* di Google Colaboratory sebagai platform pengolahan data. Tahap pertama yang dilakukan adalah *mounting* Google Drive agar dapat mengakses file dataset yang telah kami siapkan sebelumnya. Google Colaboratory dipilih sebagai platform karena menyediakan komputasi *cloud* gratis dengan akses ke *library* Python yang lengkap, serta kemudahan integrasi dengan Google Drive untuk penyimpanan data.

Selanjutnya, dilakukan *import library-library* yang dibutuhkan untuk proses integrasi data, yaitu:

- *Library* `pandas` digunakan sebagai *library* utama untuk manipulasi dan analisis data tabular.
- *Library* `glob` diperlukan untuk melakukan *pattern matching* dalam proses *batch file processing*, sehingga memudahkan pembacaan *multiple file* secara otomatis.
- *Library* `os` digunakan untuk operasi file *path* dan direktori, terutama dalam proses ekstraksi informasi dari nama file.

Library ini merupakan fondasi dasar yang akan digunakan sepanjang proses data wrangling, mulai dari pembacaan data mentah hingga proses pembersihan dan transformasi data.

b. Membaca Semua File dari Tiap Folder

Dataset penelitian ini disimpan dalam struktur folder yang terorganisir dengan baik di Google Drive. Hasil identifikasi menunjukkan bahwa terdapat empat file CSV untuk data produksi padi yang mencakup periode 2021-2024. Setiap file berisi data produksi padi untuk satu tahun tertentu dengan cakupan 34 provinsi di Indonesia. Untuk data harga beras, teridentifikasi empat file Excel yang juga mencakup periode yang sama. File-file ini berisi data harga beras bulanan di pasar tradisional yang dipantau oleh Bank Indonesia. Sedangkan untuk data harga gabah, terdapat satu file CSV yang berisi data harga gabah untuk keempat tahun

sekaligus dalam format wide. Kelengkapan data dari ketiga sumber ini sesuai dengan dataset yang kami miliki, ini menjadi syarat untuk dapat melakukan analisis margin harga yang komprehensif .

c. Integrasi Data Produksi Padi

Proses integrasi data produksi padi dari Badan Pusat Statistik langkah awal kami membaca setiap file CSV terlebih dahulu dengan melewati tiga baris metadata di bagian atas agar hanya baris data yang masuk ke DataFrame. Setelah itu, nama kolom ditetapkan secara manual agar seragam, meliputi Provinsi, Luas Panen (ha), Produktivitas (kuintal/ha), Produksi (ton), dan Tahun. Informasi tahun diekstraksi langsung dari nama file menggunakan teknik *string manipulation*, yaitu dengan memisahkan nama file berdasarkan *underscore*, mengambil bagian akhir sebelum ekstensi CSV, dan menarik empat digit terakhir sebagai tahun. Seluruh DataFrame dari keempat file kemudian digabungkan secara vertikal dengan `ignore_index=True` agar indeks tersusun ulang secara berurutan. Lalu dilakukan pembersihan awal dengan mengisi nilai kosong dengan nol serta menghapus baris duplikat header yang terbaca.

Hasil akhir integrasi menunjukkan total 159 baris data yang mencakup data 34 provinsi selama empat tahun beserta baris agregat nasional, dengan struktur lima kolom sesuai format yang telah ditetapkan.

	Provinsi	Luas Panen (ha)	Produktivitas (ku/ha)	Produksi (ton)	Tahun
0	ACEH	297058.38	55.03	1634639.6	2021
1	SUMATERA UTARA	385405	52	2004142.51	2021
2	SUMATERA BARAT	272391.95	48.36	1317209.38	2021
3	RIAU	53062.35	40.98	217458.87	2021
4	JAMBI	64412.26	46.29	298149.25	2021
...
154	PAPUA	1068.57	43.14	4609.95	2024
155	PAPUA SELATAN	47168.57	46.17	217789.62	2024
156	PAPUA TENGAH	1436.12	42.28	6072.38	2024
157	PAPUA PEGUNUNGAN	9.66	43.87	42.38	2024
158	INDONESIA	10046135.36	52.90	53142726.65	2024

Data yang dihasilkan merepresentasikan variasi produksi padi antarprovinsi, seperti Aceh yang pada tahun 2021 memiliki luas panen 297.058 hektar dengan produksi 1.634.639 ton.

d. Integrasi Data Harga Beras

Proses integrasi data harga beras dari Bank Indonesia dilakukan dengan membaca file Excel agar seluruh baris awal dapat diinspeksi, kemudian baris pertama diekstraksi sebagai *header* dan disesuaikan dengan melakukan *rename* kolom, termasuk mendefinisikan kolom *Kategori* sebagai identitas jenis beras. Baris-baris yang berisi data valid dimulai dari baris ketiga sehingga data diambil menggunakan *slicing*. Struktur file yang menggunakan format *wide*, di mana setiap bulan merupakan kolom terpisah, kemudian diubah menjadi format *long*, dengan *Kategori* dan *Tahun* sebagai identifier dan kolom bulan sebagai *value*. Setelah proses *reshaping*, kolom tanggal diparse menjadi format *datetime* dan kemudian diekstraksi kembali menjadi komponen tahun dan bulan. Selanjutnya, nilai harga dibersihkan dengan menghilangkan pemisah ribuan serta menstandarkan format desimal sebelum dikonversi menjadi numerik.

Setelah semua file diproses dengan prosedur yang sama, seluruh DataFrame digabungkan ke dalam satu dataset komprehensif, kemudian diagregasi secara tahunan menggunakan `groupby` berdasarkan *Kategori* dan *Tahun* untuk memperoleh rata-rata harga beras per tahun. Hasil akhir integrasi ini menghasilkan data tahunan untuk enam kategori beras selama empat tahun.

	Kategori	Tahun	Total_Harga_Tahunan				
0	Beras Kualitas Bawah I	2021	9.94	13	Beras Kualitas Medium II	2022	8.82
1	Beras Kualitas Bawah I	2022	8.99	14	Beras Kualitas Medium II	2023	13.39
2	Beras Kualitas Bawah I	2023	12.48	15	Beras Kualitas Medium II	2024	15.25
3	Beras Kualitas Bawah I	2024	14.12	16	Beras Kualitas Super I	2021	12.11
4	Beras Kualitas Bawah II	2021	9.56	17	Beras Kualitas Super I	2022	10.93
5	Beras Kualitas Bawah II	2022	7.93	18	Beras Kualitas Super I	2023	14.92
6	Beras Kualitas Bawah II	2023	12.21	19	Beras Kualitas Super I	2024	16.74
7	Beras Kualitas Bawah II	2024	13.87	20	Beras Kualitas Super II	2021	11.65
8	Beras Kualitas Medium I	2021	10.89	21	Beras Kualitas Super II	2022	9.63
9	Beras Kualitas Medium I	2022	9.88	22	Beras Kualitas Super II	2023	14.42
10	Beras Kualitas Medium I	2023	13.64	23	Beras Kualitas Super II	2024	16.25
11	Beras Kualitas Medium I	2024	15.40				
12	Beras Kualitas Medium II	2021	10.62				

e. Integrasi Data Harga Gabah

Integrasi data harga gabah dilakukan dengan membaca file CSV lalu struktur file terdiri dari enam baris data yang merepresentasikan kombinasi antara tingkat dan kualitas, serta lima kolom tahun dari 2020 hingga 2024. Untuk keperluan analisis margin harga gabah–beras, fokus diarahkan pada kategori Gabah Kering Giling (GKG) karena merupakan standar industri dalam pengolahan gabah menjadi beras. Data harga GKG kemudian dipisahkan dan disiapkan untuk proses

transformasi ke format *long* agar selaras dengan struktur data produksi dan harga beras.

```
Data Gabah (dari file cleaned):
      Tingkat      Kualitas  2020  2021  2022  2023  \
0      Petani  Gabah Kering Panen GKP 4844.88 4551.77 4895.67 5968.14
1      Petani  Gabah Kering Giling GKG 5567.25 5061.87 5491.54 6761.92
2      Petani      Gabah Luar Kualitas 4485.97 4226.52 4570.00 5661.50
3  Penggilingan  Gabah Kering Panen GKP 4943.35 4655.52 5014.14 6111.04
4  Penggilingan  Gabah Kering Giling GKG 5691.18 5175.14 5608.11 6877.56
5  Penggilingan      Gabah Luar Kualitas 4583.76 4325.92 4673.11 5782.48

      2024*
0 6431.11
1 7346.53
2 6072.12
3 6577.77
4 7465.91
5 6211.54
```

f. *Pre-Processing* Data untuk Persiapan *Merging*

Serangkaian proses *pre-processing* dilakukan untuk memastikan bahwa struktur data dari ketiga sumber kompatibel satu sama lain.

- Agregasi Data Produksi Padi

Proses *pre-processing* dimulai dengan menstandarkan struktur dataset produksi padi dari BPS yang memiliki granularitas tingkat provinsi. Kolom *Produksi* terlebih dahulu dikonversi menjadi tipe numerik untuk memastikan nilai yang tidak valid dapat ditangani. Selanjutnya dilakukan filtering untuk menghapus baris agregat nasional yang berlabel *INDONESIA* agar perhitungan tidak terduplikasi. Setelah data berisi provinsi individual saja, dilakukan agregasi menggunakan `groupby('Tahun').mean()` sehingga dihasilkan rata-rata produksi. Hasil akhir berupa empat baris data tahunan (2021–2024) yang merepresentasikan rata-rata produksi nasional berbasis provinsi.

- Transformasi Data Harga Beras

Dataset harga beras disesuaikan strukturnya untuk keperluan penggabungan dengan dua dataset lainnya. Dilakukan dengan mengganti nama kolom agregat tahunan menjadi *Harga Beras (Rp/Kg)* sehingga lebih informatif dan konsisten. Data telah berada dalam format *long* dan mencakup nilai rata-rata tahunan sehingga tidak memerlukan transformasi tambahan.

- *Filtering* dan Transformasi Data Harga Gabah

Langkah pertama dimulai dengan menyeleksi hanya baris yang memiliki kualitas GKG. Setelah subset sesuai diperoleh, data diubah dari format *wide* menjadi format *long*, sehingga setiap kombinasi tingkat harga (petani &

penggilingan) dan tahun menjadi baris terpisah. Nilai tahun yang memiliki tanda asterisk pada 2024 dibersihkan dan dikonversi menjadi integer, lalu kolom harga diubah ke tipe numerik dengan `pd.to_numeric()`. Tahap akhir adalah melakukan agregasi rata-rata harga gabah per tahun untuk menggabungkan tingkat petani dan penggilingan menjadi satu nilai representatif, kemudian menamai ulang kolom hasil tersebut menjadi `Harga Gabah (Rp/Kg)`.

g. Proses *Merging* Dataset

Proses *merging* untuk menggabungkan seluruh informasi menjadi satu dataset terpadu. Penggabungan dilakukan menggunakan strategi *inner join* dengan kolom *Tahun* sebagai *key* utama. Proses *merging* dilakukan secara bertahap melalui *method chaining* untuk menjaga efisiensi dan keterbacaan kode. Tahap pertama dilakukan dengan menggabungkan dataset produksi dan dataset harga beras berdasarkan kolom *Tahun* dan dilanjut *merge* ke dataset harga gabah juga melalui *inner join*. Karena kedua dataset sama-sama memiliki data untuk tahun 2021–2024, hasil *merge* mencakup keempat tahun tersebut.

Hasil akhir dari proses *merging* menghasilkan dataset terdiri dari 24 baris yang merupakan kombinasi 6 kategori beras dengan 4 tahun pengamatan. Setiap baris mewakili satu kombinasi unik tahun dan kategori beras yang dilengkapi dengan nilai produksi padi rata-rata nasional, harga beras berdasarkan kategori, serta harga gabah rata-rata untuk tahun tersebut.

	Tahun	Produksi (ton)		Kategori	Harga Beras (Rp/Kg)	Harga Gabah (Rp/Kg)
0	2021	1600449.83	Beras Kualitas Bawah I		9.94	5118.51
1	2021	1600449.83	Beras Kualitas Bawah II		9.56	5118.51
2	2021	1600449.83	Beras Kualitas Medium I		10.89	5118.51
3	2021	1600449.83	Beras Kualitas Medium II		10.62	5118.51
4	2021	1600449.83	Beras Kualitas Super I		12.11	5118.51
5	2021	1600449.83	Beras Kualitas Super II		11.65	5118.51
6	2022	1564314.22	Beras Kualitas Bawah I		8.99	5549.82
7	2022	1564314.22	Beras Kualitas Bawah II		7.93	5549.82
8	2022	1564314.22	Beras Kualitas Medium I		9.88	5549.82
9	2022	1564314.22	Beras Kualitas Medium II		8.82	5549.82
10	2022	1564314.22	Beras Kualitas Super I		10.93	5549.82
11	2022	1564314.22	Beras Kualitas Super II		9.63	5549.82
12	2023	1384179.90	Beras Kualitas Bawah I		12.48	6819.74
13	2023	1384179.90	Beras Kualitas Bawah II		12.21	6819.74
14	2023	1384179.90	Beras Kualitas Medium I		13.64	6819.74

Sebagai contoh, pada tahun 2021 seluruh kategori beras memiliki nilai produksi yang sama, sementara harga beras bervariasi sesuai kualitas dengan harga gabah rata-rata sebesar Rp5.118,5/kg.

2. Data Cleaning

a. Persiapan dan *Setup Data Cleaning*

Proses *cleaning* diawali dengan menambahkan *library* seperti `numpy` untuk operasi numerik, `matplotlib` dan `seaborn` untuk visualisasi, serta `datetime` untuk pengelolaan variabel tanggal. Selanjutnya dilakukan penyesuaian display, serta menyelaraskan format desimal agar inspeksi data lebih mudah dan akurat.

b. Step 1: Inspeksi dan Validasi Data

- Pemeriksaan Informasi Dasar dan Tipe Data

Melalui fungsi `info()`, diperoleh gambaran mengenai jumlah entri, kolom, *non-null values*, serta tipe data setiap variabel, dan hasilnya menunjukkan bahwa dataset berisi 24 baris dan 5 kolom dengan seluruh nilai terisi lengkap sehingga tidak terdapat *missing values*. Tipe data yang muncul juga sesuai dengan karakter variabelnya, yakni:

```
=====
TIPE DATA SETIAP KOLOM:
=====
Tahun                : int64
Produksi (ton)       : float64
Kategori             : object
Harga Beras (Rp/Kg)  : float64
Harga Gabah (Rp/Kg)  : float64
```

Ini menunjukkan bahwa struktur dasar dataset sudah baik dan siap untuk tahap berikutnya.

- Analisis *Missing Values*

Pemeriksaan dilakukan dengan menyusun `DataFrame summary` berisi jumlah dan persentase *missing values* pada setiap kolom menggunakan kombinasi `isnull().sum()` dan perhitungan persentase terhadap total baris, sehingga tingkat kelengkapan data dapat terlihat secara kuantitatif.

	Kolom	Missing Count	Missing %
Tahun	Tahun	0	0.00
Produksi (ton)	Produksi (ton)	0	0.00
Kategori	Kategori	0	0.00
Harga Beras (Rp/Kg)	Harga Beras (Rp/Kg)	0	0.00
Harga Gabah (Rp/Kg)	Harga Gabah (Rp/Kg)	0	0.00

TIDAK ADA MISSING VALUES!

Hasilnya menunjukkan seluruh kolom memiliki nilai *missing count* dan *missing percentage* sebesar nol, mengonfirmasi bahwa dataset lengkap.

- Deteksi Duplikasi Data

Pemeriksaan dilakukan pada dua level, pertama pengecekan duplikasi penuh menggunakan `duplicated().sum()` untuk melihat apakah ada baris yang identik di seluruh kolom. Kedua, dilakukan pengecekan berdasarkan *key columns* yakni kombinasi *Tahun* dan *Kategori*. Berdasarkan hasil, tidak ditemukan duplikasi pada kedua level tersebut.

- Statistik Deskriptif dan Identifikasi *Outliers*

Melalui `describe()`, diperoleh gambaran sebagai berikut:

```

Statistik untuk kolom numerik:
      Tahun  Produksi (ton)  Harga Beras (Rp/Kg)  Harga Gabah (Rp/Kg)
count    24.00           24.00                24.00                24.00
mean    2022.50        1477907.47            12.23            6223.57
std       1.14        107796.45              2.54            945.82
min     2021.00        1362685.91             7.93            5118.51
25%     2021.75        1378806.41             9.92            5441.99
50%     2022.50        1474247.06            12.16            6184.78
75%     2023.25        1573348.13            14.20            6966.36
max     2024.00        1600449.83            16.74            7406.22

=====
IDENTIFIKASI OUTLIERS & NILAI TIDAK MASUK AKAL
=====

1 HARGA BERAS:
  Min : Rp 7.93
  Max : Rp 16.74
  Mean : Rp 12.23
  Median: Rp 12.16

2 HARGA GABAH:
  Min : Rp 5118.51
  Max : Rp 7406.22
  Mean : Rp 6223.57
  Harga gabah sudah dalam format yang benar (Rp/Kg)

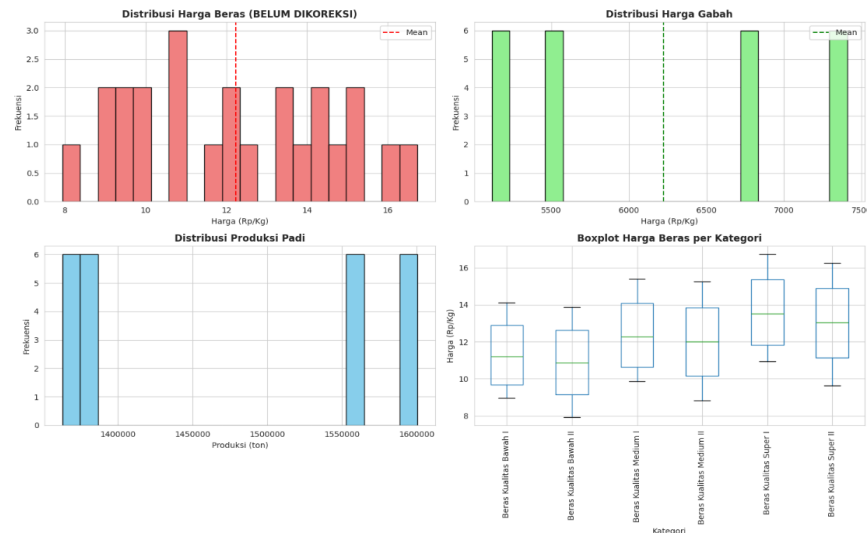
3 PRODUKSI PADI:
  Min : 1,362,685.91 ton
  Max : 1,600,449.83 ton
  Mean : 1,477,907.47 ton
  Produksi sudah dalam format yang benar (ton)

```

Pemeriksaan detail pada setiap variabel menunjukkan temuan penting, terutama pada Harga Beras yang memiliki nilai minimum 7,93 dan maksimum 16,74 angka yang jelas tidak masuk akal untuk satuan Rp/kg. Hal ini mengindikasikan adanya kesalahan unit, di mana nilai seharusnya berada pada ribuan (misalnya Rp7.930–Rp16.740), sehingga diperlukan koreksi skala dengan mengalikan seribu. Tahap ini berhasil mengidentifikasi masalah unit sebagai satu-satunya anomali signifikan, sementara variabel lain dinyatakan valid untuk masuk ke proses *cleaning* berikutnya.

- Visualisasi Distribusi Data Sebelum *Cleaning*

Visualisasi distribusi data sebelum *cleaning* dilakukan melalui 4 subplot untuk memberikan gambaran awal yang intuitif mengenai kondisi dataset.



- Histogram Harga Beras: Nilai berada di rentang 8–17, menunjukkan *kesalahan skala* karena jauh lebih rendah dari harga beras sebenarnya.
- Histogram Harga Gabah: Distribusi terlihat realistis, terkonsentrasi di sekitar 5.500–6.000 Rp/Kg, sesuai dengan statistik deskriptif.
- Histogram Produksi Padi: Nilai produksi berada pada rentang 1,36–1,60 juta ton, tanpa pola outlier yang ekstrem.
- Boxplot Harga Beras per Kategori: Menunjukkan perbedaan median antar jenis beras, meskipun skala datanya masih belum tepat.

Keempat visualisasi ini disusun dalam layout yang rapi sebagai baseline penting untuk membandingkan kondisi data sebelum dan sesudah proses *cleaning* dilakukan.

c. Step 2: Data Type Correction

Proses dimulai dengan membuat salinan DataFrame original bernama *df_cleaned* sebagai langkah pencegahan agar data awal tetap aman. Kolom *Tahun* kemudian dikonversi menjadi integer sebagai bentuk penegasan bahwa formatnya seragam. Setelah itu, kolom *Harga Beras* dikonversi menggunakan *pd.to_numeric(errors='coerce')* untuk menangani nilai tidak valid tanpa menghentikan eksekusi, diikuti koreksi serupa untuk kolom *Harga Gabah* dan *Produksi* guna memastikan bahwa seluruh nilai numerik berada dalam format float.

Koreksi juga dilakukan pada kolom *Kategori* dengan mengubah tipe data ke string dan menghilangkan *whitespace* guna mencegah *bug* pada proses *filtering* atau *grouping*. Setelah semua koreksi dilakukan, tipe data final diperiksa

menggunakan atribut *dtypes* untuk memastikan setiap kolom telah memiliki struktur data yang benar.

```
TIPE DATA SETELAH KOREKSI:  
Tahun          int64  
Produksi (ton)  float64  
Kategori       object  
Harga Beras (Rp/Kg) float64  
Harga Gabah (Rp/Kg) float64  
dtype: object
```

d. Step 3: *Unit Standardization dan Correction*

- Koreksi Unit Harga Beras

Koreksi unit dimulai dengan menyimpan nilai asli Harga Beras ke variabel *harga_beras_old* sebagai *backup* untuk kebutuhan verifikasi. Setelah itu, seluruh nilai harga dikalikan 1.000 menggunakan operasi *vectorized pandas* untuk memperbaiki skala yang sebelumnya keliru.

- Validasi Unit Semua Kolom

Validasi unit dilakukan kembali untuk memastikan seluruh kolom, bukan hanya Harga Beras, dan sudah berada dalam skala dan satuan yang benar.

```
1 HARGA BERAS:  
Min : Rp 7,933.33/Kg  
Max : Rp 16,737.50/Kg  
Mean : Rp 12,234.38/Kg  
  
2 HARGA GABAH:  
Min : Rp 5,118.51/Kg  
Max : Rp 7,406.22/Kg  
Mean : Rp 6,223.57/Kg  
  
3 PRODUKSI PADI:  
Min : 1,362,685.91 ton  
Max : 1,600,449.83 ton  
Mean : 1,477,907.47 ton
```

Hasil validasi menunjukkan bahwa Harga Beras yang telah dikoreksi berada pada rentang yang sesuai dengan mean sekitar Rp12.234/kg, sebuah range yang sepenuhnya masuk akal untuk kondisi pasar 2021–2024. Harga Gabah dan Produksi Padi juga tervalidasi benar, menunjukkan konsistensi yang realistis dan logis. Seluruh hasil ini menegaskan tidak adanya *mismatch* satuan yang tersisa di dalam dataset.

- Standarisasi Format Kategori Beras

Standarisasi format kategori beras dilakukan sebagai langkah akhir untuk memastikan konsistensi pada variabel kategorikal yang digunakan dalam proses *filtering*, *grouping*, maupun visualisasi. Proses ini menggunakan

rangkaian string methods, dimulai dari `str.strip()` , kemudian `str.title()` untuk menyeragamkan penulisan menjadi *Title Case*.

```
Kategori unik sebelum standarisasi:
['Beras Kualitas Bawah I' 'Beras Kualitas Bawah II'
 'Beras Kualitas Medium I' 'Beras Kualitas Medium II'
 'Beras Kualitas Super I' 'Beras Kualitas Super II']

Jumlah kategori unik: 6

Kategori setelah standarisasi:
1. Beras Kualitas Bawah I
2. Beras Kualitas Bawah Ii
3. Beras Kualitas Medium I
4. Beras Kualitas Medium Ii
5. Beras Kualitas Super I
6. Beras Kualitas Super Ii

Total kategori beras: 6
```

Hasilnya menghasilkan enam kategori dengan format yang seragam dimana Beras Kualitas Bawah I, Beras Kualitas Bawah II, Beras Kualitas Medium I, Beras Kualitas Medium II, Beras Kualitas Super I, dan Beras Kualitas Super II, meskipun roman numeral “II” otomatis berubah menjadi “Ii”, namun keseragaman tetap menjadi prioritas utama. Setelah standarisasi, jumlah kategori unik tetap enam, menunjukkan bahwa tidak ada kategori yang hilang atau tergabung secara tidak sengaja.

e. Step 4: Data *Consistency Check*

Dilakukan *comprehensive consistency check* untuk memastikan seluruh data sudah memenuhi *rules* dan tidak ada *error* baru yang muncul selama proses *cleaning*. Tahap ini mencakup pengecekan kelengkapan tahun, konsistensi jumlah kategori, serta verifikasi ulang duplikasi pada *key columns*.

- Validasi Kelengkapan Data per Tahun

Validasi dilakukan dengan mengecek apakah seluruh tahun analisis (2021–2024) muncul dalam dataset.

```
CEK TAHUN:
Tahun yang diharapkan: [2021, 2022, 2023, 2024]
Tahun yang tersedia : [np.int64(2021), np.int64(2022), np.int64(2023), np.int64(2024)]

UMLAH KATEGORI BERAS PER TAHUN:
Tahun
2021    6
2022    6
2023    6
2024    6
Name: Kategori, dtype: int64

Jumlah kategori yang diharapkan: 6
```

Hasilnya, *expected years* dan *actual years* sesuai, sehingga tidak ada tahun yang hilang. Jumlah kategori beras per tahun juga konsisten, masing-masing

memiliki enam kategori. Konsistensi tahun dan kategori ini memastikan analisis tren dan perbandingan antar tahun dapat dilakukan tanpa bias.

- Cek Duplikasi Setelah *Cleaning*

Langkah terakhir adalah memastikan tidak ada duplikasi pada *key columns*. *Re-check* dilakukan untuk memastikan proses *cleaning* tidak menghasilkan baris ganda, dan hasilnya tetap nol duplikasi.

f. *Summary* dan Visualisasi Data Setelah *Cleaning*

- *Summary Data Cleaned*

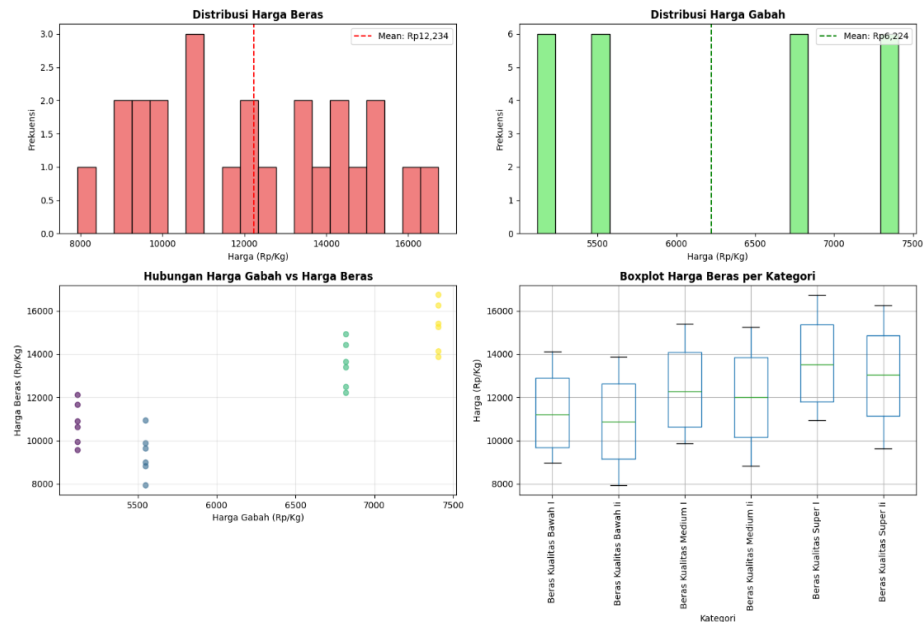
Dibuat ringkasan komprehensif untuk memberikan gambaran menyeluruh mengenai kondisi akhir dataset yang telah dibersihkan, sehingga memudahkan pengecekan.

	Tahun	Produksi (ton)	Kategori	Harga Beras (Rp/Kg)	\
0	2021	1600449.83	Beras Kualitas Bawah I	9937.50	
1	2021	1600449.83	Beras Kualitas Bawah II	9562.50	
2	2021	1600449.83	Beras Kualitas Medium I	10891.67	
3	2021	1600449.83	Beras Kualitas Medium II	10620.83	
4	2021	1600449.83	Beras Kualitas Super I	12108.33	
5	2021	1600449.83	Beras Kualitas Super II	11654.17	
6	2022	1564314.22	Beras Kualitas Bawah I	8987.50	
7	2022	1564314.22	Beras Kualitas Bawah II	7933.33	
8	2022	1564314.22	Beras Kualitas Medium I	9879.17	
9	2022	1564314.22	Beras Kualitas Medium II	8816.67	

	Harga Gabah (Rp/Kg)
0	5118.51
1	5118.51
2	5118.51
3	5118.51
4	5118.51
5	5118.51
6	5549.82
7	5549.82
8	5549.82
9	5549.82

Contoh data yang ditampilkan memperlihatkan nilai yang sudah terkoreksi, seluruh kolom dan baris lainnya juga menunjukkan nilai yang wajar dan konsisten, memberikan konfirmasi visual bahwa dataset final telah siap.

- Visualisasi *Setelah Cleaning*



- Histogram Distribusi Harga Beras setelah koreksi terdapat perubahan skala tampak jelas dimana sumbu x kini berada pada rentang 8.000 hingga 17.000, bukan lagi 8 hingga 17.
- Histogram Distribusi Harga Gabah, yang tidak mengalami perubahan berarti karena memang tidak memerlukan koreksi..
- Scatter plot hubungan Harga Gabah dan Harga Beras dengan pewarnaan berdasarkan tahun. Pola korelasi positif tampak jelas, di mana harga beras selalu lebih tinggi dari harga gabah, menggambarkan nilai tambah proses produksi.
- Boxplot Harga Beras per Kategori yang menampilkan hierarki harga antara kategori beras: Kualitas Bawah sebagai yang terendah, Medium di tengah, dan Super sebagai yang tertinggi.

Perbandingan antara grafik sebelum dan sesudah cleaning memberikan bukti visual yang kuat mengenai dampak signifikan dari proses cleaning, khususnya koreksi unit pada Harga Beras.

g. *Export Data Cleaned*

Langkah terakhir dalam proses data *cleaning* adalah mengekspor dataset yang telah dibersihkan agar dapat digunakan pada tahap analisis berikutnya.

Ekspor dilakukan dengan nama file `data_cleaned.csv` (pada google colab).

3. *Feature Engineering dan Calculation*

a. Perhitungan Margin Harga

- Hitung Margin Absolut dan Persentase

Perhitungan margin harga merupakan inti dari penelitian ini karena mengukur selisih nilai antara harga beras di tingkat konsumen dengan harga gabah di tingkat petani dan penggilingan.

Margin Absolut (Rp/Kg) dihitung dengan formula sederhana yaitu selisih langsung antara Harga Beras dan Harga Gabah.

```
Menghitung Margin Harga
=====
Margin Absolut:
Formula: Harga Beras - Harga Gabah
Range: Rp 2,383.51 - Rp 9,331.28

Margin Absolut:
Mean   : Rp 6,010.80/Kg
Median : Rp 6,116.80/Kg
Std Dev: Rp 1,806.23/Kg
```

Hasil perhitungan menunjukkan rentang margin dari Rp2.383/kg hingga Rp9.331/kg dengan rata-rata sebesar Rp6.010,80/kg. Median margin berada pada Rp6.116 dengan standar deviasi Rp1.006, menunjukkan variasi yang cukup signifikan antar kategori dan tahun.

Margin Persentase (%) dihitung menggunakan formula dibawah untuk menunjukkan proporsi margin terhadap harga gabah sebagai basis.

$$((\text{Harga Beras} - \text{Harga Gabah}) / \text{Harga Gabah}) \times 100$$

```
Margin Persentase dihitung!
Formula: ((Harga Beras - Harga Gabah) / Harga Gabah) x 100
Range: 42.95% - 136.56%

Margin Persentase:
Mean   : 95.97%
Median : 96.62%
Std Dev: 23.17%
```

Hasil menunjukkan rentang margin persentase dari 42,95% hingga 135,56% dengan rata-rata 95,97% dan median 96,62%. Standar deviasi sebesar 23,17% mengindikasikan adanya variasi pola margin yang perlu dianalisis lebih lanjut berdasarkan kategori beras dan periode waktu.

- Analisis Margin per Kategori dan Tahun

Untuk memahami pola distribusi margin secara lebih terstruktur, dilakukan agregasi data berdasarkan dua dimensi utama: kategori beras dan tahun pengamatan. Margin per Kategori Beras diagregasi menggunakan `groupby('kategori')` dengan fungsi agregasi mean, min, dan max margin absolut dan persentase di setiap kategori.

MARGIN RATA-RATA PER KATEGORI BERAS:

=====

Kategori	Margin Abs Mean	Margin Abs Min	Margin Abs Max \
Beras Kualitas Bawah I	5158.72	3437.67	6718.78
Beras Kualitas Bawah Ii	4669.14	2383.51	6460.45
Beras Kualitas Medium I	6229.55	4329.34	7997.95
Beras Kualitas Medium Ii	5794.14	3266.84	7839.61
Beras Kualitas Super I	7449.34	5379.34	9331.28
Beras Kualitas Super Ii	6763.93	4083.51	8839.61

Kategori	Margin % Mean	Margin % Min	Margin % Max
Beras Kualitas Bawah I	82.45	61.94	94.15
Beras Kualitas Bawah Ii	74.00	42.95	87.23
Beras Kualitas Medium I	99.69	78.01	112.79
Beras Kualitas Medium Ii	92.13	58.86	107.50
Beras Kualitas Super I	119.55	96.93	136.56
Beras Kualitas Super Ii	108.00	73.58	127.69

Hasil perhitungan menunjukkan pola margin yang runtut berdasarkan kualitas beras. Beras Kualitas Super I tercatat memiliki margin absolut rata-rata tertinggi Rp7.449/kg, diikuti Super II yaitu Rp6.763/kg dan Medium I yaitu Rp6.229/kg. Sementara itu, margin absolut terendah terdapat pada Beras Kualitas Bawah II yaitu Rp4.669/kg dan Bawah I yaitu Rp5.158/kg. Pola serupa terlihat pada margin persentase, di mana Beras Super I kembali menjadi yang tertinggi 119,55%, sementara Bawah II menjadi yang terendah (74%).

Rentang margin minimum–maksimum pada setiap kategori juga mencerminkan stabilitas yang berbeda. Misalnya, Medium II memiliki selisih margin yang cukup lebar Rp3.266,84/kg – Rp7.839,61/kg, menandakan adanya variasi harga yang lebih sensitif terhadap perubahan pasar. Sebaliknya, kategori seperti Super I tampak lebih stabil karena rentangnya lebih konsisten. Semakin tinggi kualitas beras, semakin besar nilai tambah dan margin keuntungannya.

Margin per Tahun diagregasi menggunakan `groupby('Tahun')` untuk melihat tren temporal margin harga selama periode penelitian.

Tahun	Margin Abs Mean	Margin Abs Min	Margin Abs Max	Margin % Mean \
2021	5677.33	4444.00	6989.83	110.92
2022	3813.37	2383.51	5379.34	68.71
2023	6687.90	5388.59	8096.93	98.07
2024	7864.61	6460.45	9331.28	106.19

Tahun	Margin % Min	Margin % Max	Produksi
2021	86.82	136.56	1600449.83
2022	42.95	96.93	1564314.22
2023	79.01	118.73	1384179.90
2024	87.23	125.99	1362685.91

Hasil agregasi menunjukkan pola menarik dimana margin absolut mengalami penurunan signifikan pada tahun 2022 menjadi rata-rata Rp3.813,37/kg dari Rp5.677,33/kg pada 2021, namun kemudian meningkat

drastis pada 2023 menjadi Rp6.687,50/kg dan terus naik hingga Rp7.864,6/kg pada 2024. Pola serupa juga terlihat pada margin persentase yang mencapai titik terendah 68,71% pada 2022 sebelum meningkat ke 106,19% pada 2024.

Data produksi padi rata-rata juga disertakan dalam agregasi tahunan dimana terlihat tren penurunan produksi dari 1,60 juta ton pada 2021 menjadi 1,36 juta ton pada 2024, mengindikasikan potensi hubungan antara penurunan produksi dengan kenaikan margin harga di tahun-tahun terakhir.

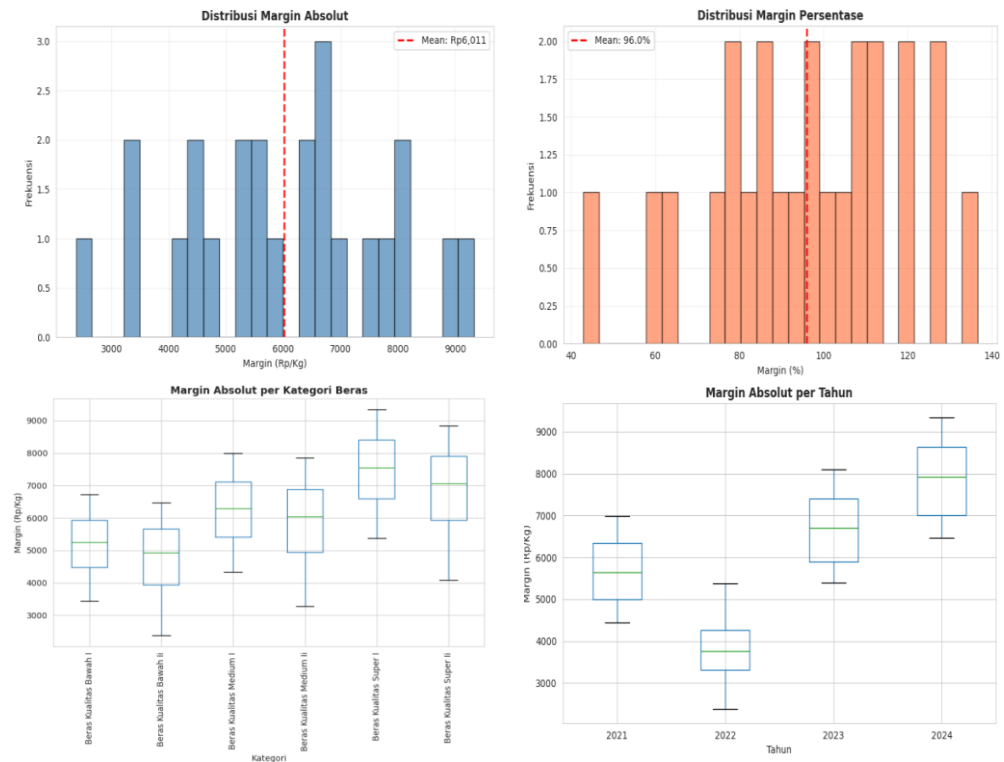
Margin Detail per Kategori dan Tahun kemudian disajikan dalam bentuk *pivot table* untuk memberikan gambaran komprehensif mengenai evolusi margin di setiap kategori sepanjang periode penelitian.

Tahun Kategori	Margin Absolut (Rp/Kg)			
	2021	2022	2023	2024
Beras Kualitas Bawah I	4819.00	3437.67	5659.43	6718.78
Beras Kualitas Bawah II	4444.00	2383.51	5388.59	6460.45
Beras Kualitas Medium I	5773.16	4329.34	6817.76	7997.95
Beras Kualitas Medium II	5502.33	3266.84	6567.76	7839.61
Beras Kualitas Super I	6989.83	5379.34	8096.93	9331.28
Beras Kualitas Super II	6535.66	4083.51	7596.93	8839.61

Tahun Kategori	Margin Persentase (%)			
	2021	2022	2023	2024
Beras Kualitas Bawah I	94.15	61.94	82.99	90.72
Beras Kualitas Bawah II	86.82	42.95	79.01	87.23
Beras Kualitas Medium I	112.79	78.01	99.97	107.99
Beras Kualitas Medium II	107.50	58.86	96.31	105.85
Beras Kualitas Super I	136.56	96.93	118.73	125.99
Beras Kualitas Super II	127.69	73.58	111.40	119.35

Pivot table pertama menampilkan Margin Absolut (Rp/Kg) dengan kategori beras sebagai baris dan tahun sebagai kolom. Hasilnya menunjukkan bahwa seluruh kategori beras mengalami pola fluktuasi margin yang relatif serupa dari tahun ke tahun. Meskipun besaran margin berbeda antar kategori, kecenderungan kenaikan dan penurunan margin mengikuti arah yang sama. Kategori beras kualitas Super I dan Super II konsisten memiliki margin absolut tertinggi, sedangkan kategori Bawah I dan Bawah II berada pada kisaran margin terendah selama periode pengamatan.

Pivot table kedua menyajikan Margin Persentase (%) dengan struktur tampilan yang sama. Secara umum, pola perubahan margin persentase memperkuat temuan pada margin absolut, di mana semua kategori mengalami fluktuasi dari tahun ke tahun dengan kecenderungan serupa. Margin persentase tertinggi juga tercatat pada kategori Super I dan Super II, mengindikasikan bahwa peningkatan nilai tambah terbesar terdapat pada beras kelas premium. Sementara itu, kategori Bawah I dan Bawah II menunjukkan margin persentase yang relatif lebih rendah, mencerminkan jarak harga yang lebih kecil antara harga gabah dan harga beras pada segmen tersebut.



Lalu dari visualisasi Margin Harga juga menunjukkan distribusi margin absolut yang terkonsentrasi di sekitar Rp 6.000/kg menunjukkan stabilitas nilai tambah dalam konversi gabah menjadi beras. Namun, sebaran margin persentase yang jauh lebih luas mengindikasikan bahwa perubahan harga gabah tidak selalu diimbangi secara proporsional oleh harga beras, sehingga muncul *asymmetric price transmission* antara tingkat petani dan konsumen.

Perbandingan antar kategori menegaskan adanya ketimpangan nilai tambah: kategori Super consistently memiliki margin tertinggi, sedangkan kategori Medium dan Bawah jauh lebih rendah. Hal ini mengisyaratkan bahwa pelaku pasar di segmen premium memiliki posisi tawar yang lebih kuat dalam penetapan harga, sehingga keuntungan terbesar justru terkonsentrasi pada segmen beras berkualitas tinggi, bukan pada keseluruhan rantai pasok.

Dari sisi waktu, penurunan margin pada tahun 2022 diikuti lonjakan pada 2023–2024 menunjukkan dinamika pasar pascapandemi. Pelebaran variasi margin pada dua tahun terakhir menandakan bahwa pemulihan margin tidak merata antar kategori, yang berpotensi memperluas kesenjangan nilai tambah dalam industri beras.

- Identifikasi *Outliers* Margin

Proses dimulai dengan menghitung kuartil pertama (Q1) dan kuartil ketiga (Q3) untuk Margin Absolut menggunakan fungsi `quantile(0.25)` dan `quantile(0.75)`. Dari kedua nilai ini dihitung IQR dengan formula $Q3 - Q1$, yang kemudian digunakan untuk menentukan batas bawah ($Q1 - 1.5 \times IQR$) dan batas atas ($Q3 + 1.5 \times IQR$). Nilai margin yang berada di luar rentang ini dikategorikan sebagai outliers.

Hasil perhitungan menunjukkan batas bawah sebesar Rp1.100,71 dan batas atas sebesar Rp 10.766,14. Setelah dilakukan *filtering* menggunakan kondisi *boolean*, ditemukan bahwa tidak ada data yang berada di luar rentang tersebut, sehingga dataset tidak memiliki outliers signifikan pada margin absolut. Proses yang sama juga dilakukan untuk Margin Persentase hasilnya juga menunjukkan tidak ada outliers, mengonfirmasi bahwa seluruh data margin berada dalam rentang yang wajar dan konsisten.

b. Agregasi Data per Tahun

- Agregasi Margin per Tahun

Proses agregasi menggunakan `groupby('Tahun')` dengan berbagai fungsi yang disesuaikan dengan karakter masing-masing variabel. Untuk variabel Margin Absolut (Rp/Kg) dan Margin Persentase (%), digunakan fungsi agregasi statistik. Hal ini penting untuk memahami tidak hanya nilai rata-rata tetapi juga tingkat dispersi dan rentang nilai di setiap tahun. Sementara itu, untuk variabel Produksi (ton), Harga Beras (Rp/Kg), dan Harga Gabah (Rp/Kg) hanya diambil nilai mean karena variabel-variabel ini sudah merupakan agregat dan tidak memerlukan ukuran dispersi tambahan dalam konteks analisis tahunan. Kolom Kategori diagregasi menggunakan fungsi `count` untuk menghitung jumlah kategori beras yang ada di setiap tahun, memastikan konsistensi data

AGREGASI LENGKAP PER TAHUN:

Tahun	Margin Abs Mean	Margin Abs Median	Margin Abs Std	Margin Abs Min	\
2021	5677.33	5637.74	975.40	4444.00	
2022	3813.37	3760.59	1027.21	2383.51	
2023	6687.90	6692.76	1057.25	5388.59	
2024	7864.61	7918.78	1131.79	6460.45	

Tahun	Margin Abs Max	Margin % Mean	Margin % Median	Margin % Std	\
2021	6989.83	110.92	110.14	19.06	
2022	5379.34	68.71	67.76	18.51	
2023	8096.93	98.07	98.14	15.50	
2024	9331.28	106.19	106.92	15.28	

Tahun	Margin % Min	Margin % Max	Produksi Mean	Produksi Min	Produksi Max	\
2021	86.82	136.56	1600449.83	1600449.83	1600449.83	
2022	42.95	96.93	1564314.22	1564314.22	1564314.22	
2023	79.01	118.73	1384179.90	1384179.90	1384179.90	
2024	87.23	125.99	1362685.91	1362685.91	1362685.91	

Tahun	Harga Beras Mean	Harga Gabah Mean	Jumlah Data
2021	10795.83	5118.50	6
2022	9363.19	5549.82	6
2023	13507.64	6819.74	6
2024	15270.83	7406.22	6

Hasil output agregasi tahunan menampilkan pengamatan dimana tahun 2021 memiliki margin absolut rata-rata Rp5.677,33 dengan median Rp5.637,74 dan standar deviasi Rp975,40, menunjukkan distribusi yang relatif terpusat. Tahun 2022 menunjukkan penurunan drastis dengan margin rata-rata turun menjadi Rp 3.813,37, namun kemudian pulih dan meningkat secara konsisten pada 2023 dan 2024.

Data produksi padi juga menunjukkan tren penurunan yang konsisten dari 1,60 juta ton pada 2021 menjadi 1,36 juta ton pada 2024, sementara harga beras dan harga gabah cenderung meningkat sepanjang periode. Korelasi negatif antara produksi dan harga ini mengindikasikan adanya dinamika *supply-demand* yang memengaruhi struktur harga dan margin.

- Analisis Tren Tahunan

Untuk memahami dinamika perubahan dari tahun ke tahun, dilakukan analisis tren menggunakan perhitungan perubahan *year-over-year* (YoY) dalam bentuk persentase untuk mengidentifikasi pola pertumbuhan atau penurunan serta menginterpretasikan faktor-faktor yang memengaruhi perubahan tersebut.

Perubahan YoY dihitung menggunakan fungsi `pct_change()` yang secara otomatis menghitung persentase perubahan relatif terhadap periode

sebelumnya, kemudian dikalikan 100 untuk konversi ke format persentase. Proses ini diterapkan pada empat variabel kunci.

SUMMARY TAHUNAN DENGAN PERUBAHAN YoY:			
	Margin Absolut (Rp/Kg)	Margin Persentase (%)	Produksi (ton)
Tahun			
2021	5677.33	110.92	1600449.83
2022	3813.37	68.71	1564314.22
2023	6687.90	98.07	1384179.90
2024	7864.61	106.19	1362685.91

	Harga Beras (Rp/Kg)	Harga Gabah (Rp/Kg)	Δ Margin Abs (%) \
Tahun			
2021	10795.83	5118.50	NaN
2022	9363.19	5549.82	-32.83
2023	13507.64	6819.74	75.38
2024	15270.83	7406.22	17.59

	Δ Produksi (%)	Δ Harga Beras (%)	Δ Harga Gabah (%)
Tahun			
2021	NaN	NaN	NaN
2022	-2.26	-13.27	8.43
2023	-11.52	44.26	22.88
2024	-1.55	13.05	8.60

Output menunjukkan ringkasan tahunan dengan perubahan YoY yang sangat informatif. Pada tahun 2021 tidak ada nilai perubahan karena merupakan tahun pertama dalam periode pengamatan (nilai NaN). Tahun 2022 menunjukkan penurunan margin absolut sebesar -32,8%, disertai dengan penurunan produksi sebesar -2,3%, namun harga beras justru turun -13,27% dan harga gabah naik 8,43%, menciptakan kombinasi yang menekan margin. Tahun 2023 menunjukkan pemulihan kuat dengan kenaikan margin sebesar 75,38% meskipun produksi turun lebih dalam -11,52%, sementara harga beras naik drastis 44,26% dan harga gabah naik 22,88%. Tahun 2024 menunjukkan pertumbuhan margin yang lebih moderat sebesar 17,59% dengan produksi yang kembali turun -1,6% namun dengan kenaikan harga yang lebih terkendali pada beras 13,05% dan gabah 8,20%.

Interpretasi tren kemudian dilakukan secara otomatis menggunakan logika kondisional yang memeriksa arah perubahan margin dan produksi untuk setiap tahun. Logika ini dirancang untuk mengklasifikasikan kondisi pasar ke dalam beberapa kategori yaitu kondisi pasar menurun (margin dan produksi turun), potensi inefisiensi distribusi (margin turun saat produksi naik), efisiensi meningkat (margin naik saat produksi naik), dan peningkatan permintaan/biaya (margin dan produksi naik).

INTERPRETASI TREND:

Tahun 2022:

Margin: ↓ 32.83%

Produksi: ↓ 2.26%

Margin & produksi TURUN → Kondisi pasar menurun

Tahun 2023:

Margin: ↑ 75.38%

Produksi: ↓ 11.52%

Margin NAIK saat produksi TURUN → Potensi inefisiensi distribusi!

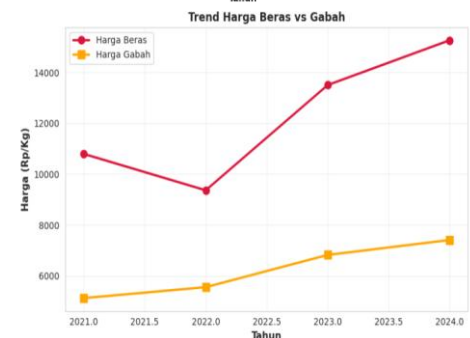
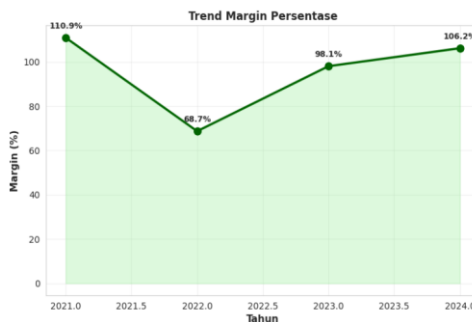
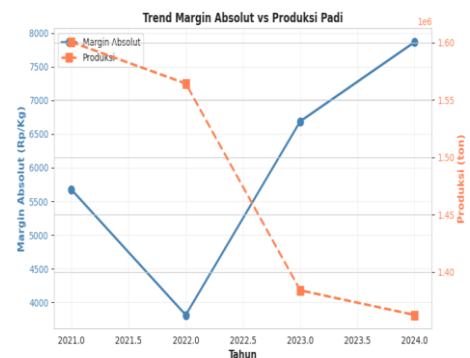
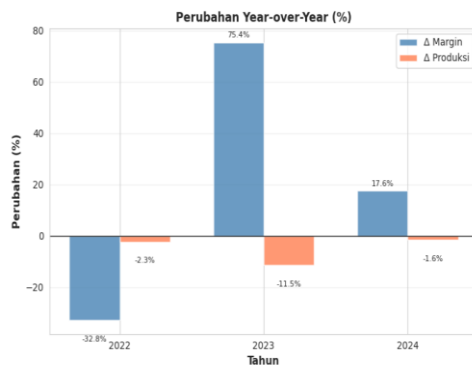
Tahun 2024:

Margin: ↑ 17.59%

Produksi: ↓ 1.55%

Margin NAIK saat produksi TURUN → Potensi inefisiensi distribusi!

Hasil interpretasi ini memberikan insight awal yang sangat berguna untuk analisis kebijakan, meskipun perlu dikonfirmasi dengan analisis kualitatif dan data pendukung lainnya.



Visualisasi menunjukkan dinamika hubungan antara margin, produksi, serta harga beras–gabah dalam periode 2021–2024. Pada grafik Margin Absolut vs Produksi, terlihat pola yang konsisten bahwa ketika produksi menurun, margin justru meningkat. Hal ini mengindikasikan adanya mekanisme pasar yang sensitif terhadap ketersediaan pasokan dimana penurunan produksi padi mendorong kenaikan harga jual beras, yang pada akhirnya meningkatkan margin. Tahun 2022 menjadi titik ekstrem di mana produksi turun namun margin meningkat tajam, memperkuat indikasi hubungan terbalik (*inverse relationship*) antara kedua variabel.

Grafik Margin Persentase memperlihatkan pola berbentuk V , dengan penurunan signifikan pada 2022 dan pemulihan kuat pada 2023–2024. Tren ini menunjukkan bahwa efisiensi ekonomi atau keuntungan relatif per kilogram sempat tertekan, namun kembali pulih seiring kenaikan harga beras dan stabilisasi pasokan.

Untuk Harga Beras dan Harga Gabah, keduanya menunjukkan tren meningkat setiap tahun. Kenaikan yang paralel menandakan bahwa harga gabah sebagai input utama memberikan kontribusi langsung pada kenaikan harga beras di level konsumen. Harga beras selalu berada di atas harga gabah, mencerminkan *value added* dari proses penggilingan, distribusi, dan biaya operasional lainnya.

Secara keseluruhan, seluruh tren menegaskan bahwa margin sangat dipengaruhi oleh kombinasi faktor dari perubahan produksi, dinamika harga gabah sebagai bahan baku, serta kenaikan harga beras di konsumen. Pemahaman hubungan-hubungan ini penting untuk merumuskan strategi stabilisasi harga dan perlindungan pendapatan petani maupun pelaku distribusi.

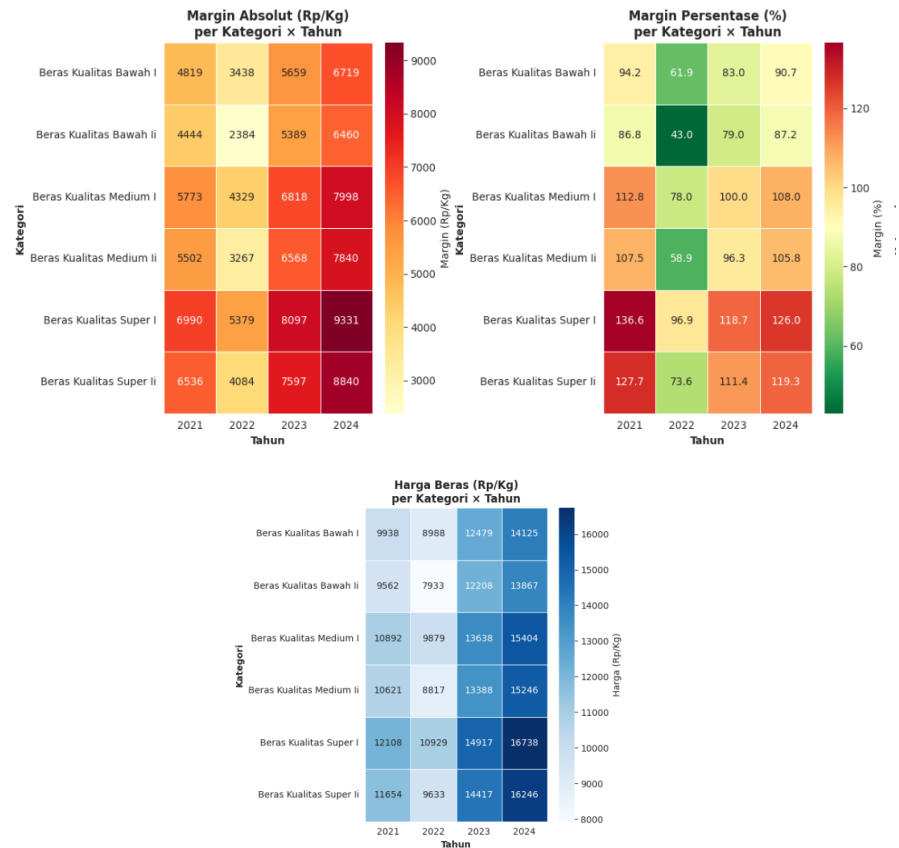
c. *Data Transformation*

- Pivot Tables untuk Analisis Temporal

Transformasi data ke dalam bentuk *pivot table* dilakukan untuk memfasilitasi analisis *cross-sectional* dan temporal yang lebih terstruktur. *Pivot table* sudah ditampilkan pada pembahasan margin.

- Heatmap Pivot Tables

Untuk memberikan visualisasi yang lebih intuitif terhadap pola dalam *pivot tables*, dibuat tiga heatmap. Setiap heatmap dilengkapi colorbar untuk menunjukkan intensitas nilai, sehingga perbedaan antartahun dan antarkategori dapat langsung terlihat tanpa perlu membaca tabel secara detail.



Secara umum, heatmap ini membuat tiga hal utama menjadi lebih mudah terlihat:

- Perbedaan level margin antar kategori, di mana kategori beras kualitas lebih tinggi tampak memiliki nilai margin (absolut maupun persentase) yang lebih besar secara konsisten.
- Variasi nilai antar tahun, di mana pergeseran intensitas warna menunjukkan dinamika margin dan harga, sehingga perubahan dari satu tahun ke tahun berikutnya dapat terlihat sekilas.
- Kenaikan harga beras yang bersifat progresif, tampak dari gradasi warna yang semakin intens pada tahun-tahun terakhir.

Dengan demikian, visual ini membantu kami dalam menangkap pola yang sama secara cepat melalui representasi warna.

- **Dataset untuk Analisis Korelasi**

Dataset ini mengambil agregasi tahunan yang telah disederhanakan dengan hanya mempertahankan variabel-variabel numerik kunci yang relevan untuk analisis hubungan antar variabel.

Proses dimulai dengan mengambil subset dari *summary* tahunan yang mencakup lima kolom utama yaitu Margin Absolut (Rp/Kg), Margin Persentase (%), Produksi (ton), Harga Beras (Rp/Kg), dan Harga Gabah (Rp/Kg). Fungsi `copy()` digunakan untuk membuat salinan independen agar manipulasi tidak memengaruhi DataFrame original. Kolom indeks kemudian direset dan diubah menjadi kolom reguler bernama *Tahun_Numeric* untuk memfasilitasi analisis korelasi temporal. Hal ini penting karena beberapa metode korelasi memerlukan representasi numerik dari variabel waktu.

```

DATASET KORELASI (per Tahun):
=====
      Margin Absolut (Rp/Kg)  Margin Persentase (%)  Produksi (ton) \
Tahun
2021                5677.33                110.92    1600449.83
2022                3813.37                68.71    1564314.22
2023                6687.90                98.07    1384179.90
2024                7864.61                106.19    1362685.91

      Harga Beras (Rp/Kg)  Harga Gabah (Rp/Kg)  Tahun_Numeric
Tahun
2021                10795.83                5118.50            2021
2022                9363.19                5549.82            2022
2023                13507.64                6819.74            2023
2024                15270.83                7406.22            2024

KEY CORRELATIONS:
=====
Margin Absolut vs Produksi: -0.809
Margin % vs Produksi: -0.274
Harga Beras vs Harga Gabah: 0.925

```

Perhitungan Korelasi kemudian dilakukan menggunakan fungsi `corr()` yang menghasilkan *correlation matrix* lengkap untuk seluruh pasangan variabel. *Output* menampilkan tiga korelasi kunci yang paling relevan dengan penelitian:

- a) Margin Absolut vs Produksi: -0,809 menunjukkan korelasi negatif yang kuat, mengonfirmasi bahwa penurunan produksi cenderung diikuti oleh kenaikan margin harga. Hubungan invers ini konsisten dengan prinsip *supply-demand* dimana kelangkaan pasokan mendorong kenaikan harga dan memperlebar margin.
- b) Margin % vs Produksi: -0,274 menunjukkan korelasi negatif yang lebih lemah dibanding margin absolut, mengindikasikan bahwa margin persentase tidak sepenuhnya dijelaskan oleh faktor produksi saja dan kemungkinan dipengaruhi oleh faktor-faktor lain seperti biaya distribusi atau ekspektasi harga.
- c) Harga Beras vs Harga Gabah: 0,925 menunjukkan korelasi positif yang sangat kuat, mengonfirmasi bahwa harga beras dan harga gabah bergerak

searah dengan koefisien yang mendekati 1. Hubungan ini logis mengingat beras merupakan produk turunan langsung dari gabah dengan rasio konversi yang tetap.

Dataset korelasi kemudian disimpan sebagai 'correlation_dataset.csv' untuk memfasilitasi analisis lanjutan menggunakan tools statistik eksternal seperti R, SPSS, atau Python libraries lainnya.

- Dataset Terpisah untuk Analisis Lanjutan

Untuk memfasilitasi berbagai jenis analisis yang lebih spesifik, dibuat beberapa dataset terpisah dengan struktur dan fokus yang berbeda-beda. Setiap dataset dirancang untuk menjawab pertanyaan penelitian yang spesifik dan mengoptimalkan proses analisis pada tahap eksplorasi data.

- a) Dataset *Margin Detail*, dataset ini merupakan versi lengkap dari *df_with_margin* yang sudah diurutkan berdasarkan Tahun dan Kategori. Berisi 24 observasi ($6 \text{ kategori} \times 4 \text{ tahun}$) dengan 9 variabel, termasuk margin absolut, margin persentase, dan variabel *hasil feature engineering* lainnya. Dataset ini menjadi *master data* untuk analisis detail margin harga.
- b) Dataset *Yearly Aggregate*, dataset ini berisi rangkuman per tahun untuk analisis tren, dataset memiliki 4 baris (2021–2024) dan 10 variabel yang mencakup statistik margin, produksi, dan harga. Sangat berguna untuk melihat perubahan tahunan dan perhitungan YoY.
- c) Dataset *Category Comparison*, dataset ini membandingkan enam kategori beras melalui statistik agregat seperti mean, standar deviasi, min, dan max untuk margin, harga beras, dan harga gabah. Hasilnya menunjukkan profil statistik tiap kategori, misalnya Super I dengan margin tertinggi dan Bawah II dengan margin terendah.
- d) Dataset *Margin Efficiency Score*, menambahkan kolom *Efficiency Score* berdasarkan margin persentase dengan empat kategori: Sangat Efisien ($\geq 80\%$), Efisien ($\geq 70\%$), Cukup Efisien ($\geq 60\%$), dan Kurang Efisien ($< 60\%$). Kategorisasi ini diterapkan ke seluruh baris menggunakan fungsi *apply()*, kemudian komposisi setiap kategori dianalisis melalui *value_counts()* untuk melihat distribusi tingkat efisiensi margin. Penambahan metrik ini diperlukan untuk mengubah margin persentase yang semula hanya berupa angka menjadi informasi yang lebih interpretatif.

Seluruh dataset terpisah kemudian disimpan sebagai file CSV individual untuk dokumentasi dan analisis lanjutan.

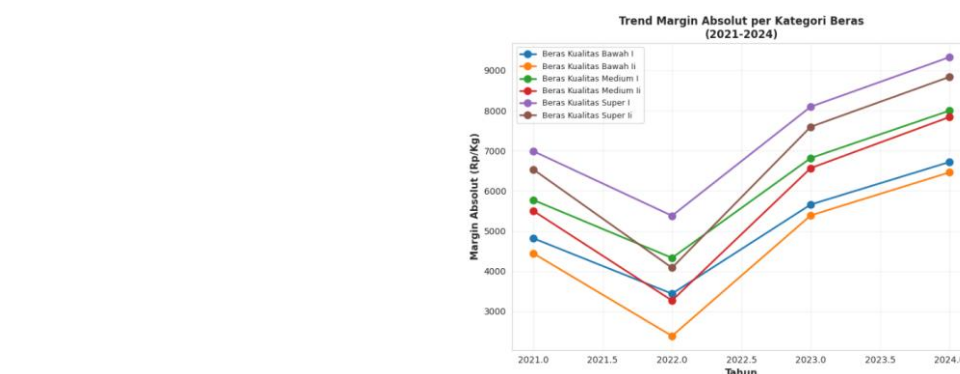
- *Reshape Data untuk Visualisasi Advanced*

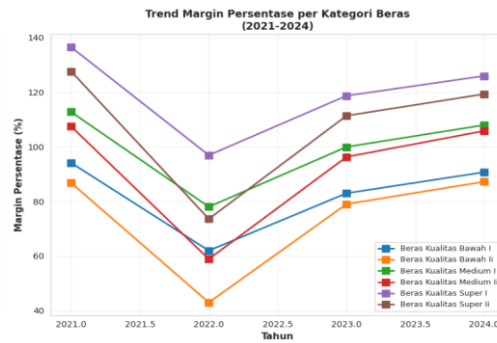
Transformasi struktur data dilakukan untuk mendukung visualisasi yang beragam dan lebih informatif. Dataset di-*reshape* ke tiga format utama sesuai kebutuhan grafik yaitu *long format*, *wide format*, dan *stacked format*. *Long format* dibuat dengan mempertahankan kolom Tahun, Kategori, Margin Absolut, dan Margin Persentase, sehingga cocok untuk visualisasi perbandingan antar kategori dalam satu grafik, seperti *multi-category line plot*. *Wide format* dibentuk menggunakan *pivot_table()* untuk menampilkan kategori sebagai baris dan tahun sebagai kolom, sehingga memudahkan analisis pola temporal melalui *heatmap*. Sementara itu, *stacked format* disusun untuk kebutuhan *stacked bar chart* dengan Tahun sebagai indeks dan kategori sebagai kolom, memungkinkan visualisasi komposisi margin antar kategori pada setiap periode.

Ketiga format ini disimpan sebagai file CSV terpisah untuk memastikan fleksibilitas penggunaan pada berbagai teknik visualisasi sekaligus mendukung reproduktibilitas analisis.

- Visualisasi *Multi-Category Line Plot*

Visualisasi *multi-category line plot* digunakan untuk menampilkan tren margin harga beras dari tahun ke tahun untuk seluruh kategori secara simultan. Dataset yang akan digunakan telah di bahas pada bagian dataset untuk analisis lanjutan.





Jika dilihat pada grafik margin absolut, seluruh kategori menunjukkan pola yang hampir sama yaitu terjadi penurunan tajam pada tahun 2022, diikuti peningkatan kembali pada 2023 dan stabilisasi pada 2024. Konsistensi urutan antar kategori menandakan bahwa perbedaan kualitas beras tetap stabil, dan fluktuasi margin lebih dipengaruhi oleh tekanan pasar tahunan dibanding perubahan struktur pasar.

Jika dilihat pada grafik margin persentase, tren utamanya juga menunjukkan penurunan pada 2022 dan pemulihan bertahap pada 2023–2024, namun variasi antar kategori lebih terlihat. Kategori premium (Super I dan Super II) mengalami pemulihan margin yang lebih cepat, sedangkan kategori menengah dan bawah meningkat lebih moderat. Hal ini menunjukkan bahwa margin persentase lebih sensitif terhadap perubahan harga dan biaya, tetapi struktur hierarki kategori tetap tidak berubah sepanjang periode pengamatan.

4. Data *Eksploration* dan *Analysis*

a. *Exploratory Data Analysis* - Univariate

- Summary Statistics untuk Variabel Numerik

Analisis univariat pada penelitian ini dimulai dengan menghitung statistik deskriptif untuk seluruh variabel numerik. Pemilihan variabel dilakukan menggunakan `select_dtypes()`, kemudian fungsi `describe()` digunakan untuk menghasilkan ukuran dasar seperti mean, standar deviasi, nilai minimum–maksimum, serta kuartil. Untuk memberikan gambaran distribusi yang lebih komprehensif, tiga ukuran tambahan dihitung yaitu Coefficient of Variation (CV), skewness, dan kurtosis, lalu seluruh hasil dirangkum dalam `summary_stats` sebagai ringkasan deskriptif setiap variabel.

	count	mean	std	min	25%	\
Tahun	24.00	2022.50	1.14	2021.00	2021.75	
Produksi (ton)	24.00	1477907.47	107796.45	1362685.91	1378806.41	
Harga Beras (Rp/Kg)	24.00	12234.38	2544.84	7933.33	9922.92	
Harga Gabah (Rp/Kg)	24.00	6223.57	945.82	5118.51	5441.99	
Margin Absolut (Rp/Kg)	24.00	6010.80	1806.23	2383.51	4725.24	
Margin Persentase (%)	24.00	95.97	23.17	42.95	81.99	
Konversi GKG-Beras	24.00	0.65	0.00	0.65	0.65	
Biaya Konversi Per Kg (Rp)	24.00	9247.39	2778.81	3666.94	7269.61	
Efficiency Score	24.00	4.03	23.17	-36.56	-11.74	

	50%	75%	max	CV (%)	Skewness	\
Tahun	2022.50	2023.25	2024.00	0.06	0.00	
Produksi (ton)	1474247.06	1573348.13	1600449.83	7.29	0.03	
Harga Beras (Rp/Kg)	12158.33	14197.92	16737.50	20.80	0.10	
Harga Gabah (Rp/Kg)	6184.78	6966.36	7406.22	15.20	0.07	
Margin Absolut (Rp/Kg)	6116.80	7141.60	9331.28	30.05	-0.12	
Margin Persentase (%)	96.62	111.74	136.56	24.14	-0.39	
Konversi GKG-Beras	0.65	0.65	0.65	0.00	0.00	
Biaya Konversi Per Kg (Rp)	9410.47	10987.08	14355.82	30.05	-0.12	
Efficiency Score	3.38	18.01	57.05	575.18	0.39	

	Kurtosis
Tahun	-1.39
Produksi (ton)	-2.09
Harga Beras (Rp/Kg)	-1.13
Harga Gabah (Rp/Kg)	-1.83
Margin Absolut (Rp/Kg)	-0.55
Margin Persentase (%)	-0.12
Konversi GKG-Beras	0.00
Biaya Konversi Per Kg (Rp)	-0.55
Efficiency Score	-0.12

Secara umum, CV memberikan konteks variabilitas antar variabel yang berada pada skala berbeda. Margin Persentase menunjukkan CV sekitar 24%, menandakan fluktuasi yang cukup besar, sedangkan Harga Gabah memiliki CV sekitar 15%, mengindikasikan pergerakan harga yang relatif stabil. Ukuran skewness sebagian besar mendekati nol, baik positif maupun negatif yang berarti distribusi variabel-variabel cenderung simetris, dengan sedikit kemiringan ke kiri pada variabel seperti Margin Absolut. Adapun kurtosis pada sebagian besar variabel bernilai negatif moderat, menunjukkan karakteristik *platykurtic* atau distribusi yang lebih datar dibanding distribusi normal.

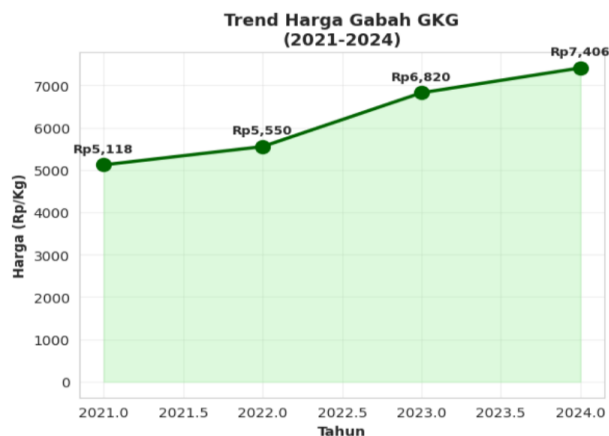
Berdasarkan output deskriptif, produksi padi memiliki rata-rata 1.477.907 ton dengan standar deviasi 107.796 ton, mencerminkan fluktuasi produksi yang cukup signifikan selama empat tahun. Harga beras memiliki rata-rata Rp12.234/kg dengan rentang Rp7.933–16.737/kg, sedangkan harga gabah rata-rata Rp6.224/kg dengan rentang Rp5.118–7.406/kg. Margin absolut tercatat rata-rata Rp6.010/kg dengan variasi besar (std Rp 1.806), sedangkan margin persentase rata-rata 95,97% dengan standar deviasi 23,17%, yang konsisten dengan nilai CV yang menunjukkan tingkat variabilitas menengah.

Secara keseluruhan, statistik univariat ini menunjukkan bahwa sebagian besar variabel memiliki variasi moderat dan distribusi yang stabil tanpa indikasi penyimpangan ekstrem. Kondisi ini mengindikasikan bahwa dataset memiliki

kualitas distribusi yang baik untuk analisis berikutnya, termasuk penelusuran tren temporal, perbandingan antar kategori beras, dan analisis margin secara mendalam.

- Distribusi Harga Gabah per Tahun

Untuk memahami pola pergerakan dan distribusi harga gabah selama periode 2021–2024, dilakukan analisis deskriptif dan visualisasi menggunakan tiga jenis grafik, yaitu *trend line*, *boxplot per tahun*, dan *histogram keseluruhan*.



Visualisasi tren temporal harga gabah menunjukkan kenaikan yang konsisten dari 2021 hingga 2024, dengan total peningkatan sekitar 44,7%. Kenaikan tahunan yang stabil ini mengindikasikan adanya tekanan biaya dari sisi hulu, sehingga perubahan margin lebih mungkin berasal dari naiknya harga input dibanding perubahan efisiensi proses produksi.

- Distribusi Harga Beras per Kategori dan Tahun

Analisis distribusi harga beras dilakukan pada dua dimensi yaitu per kategori untuk memahami diferensiasi harga antar kualitas, dan per tahun untuk memahami tren temporal.

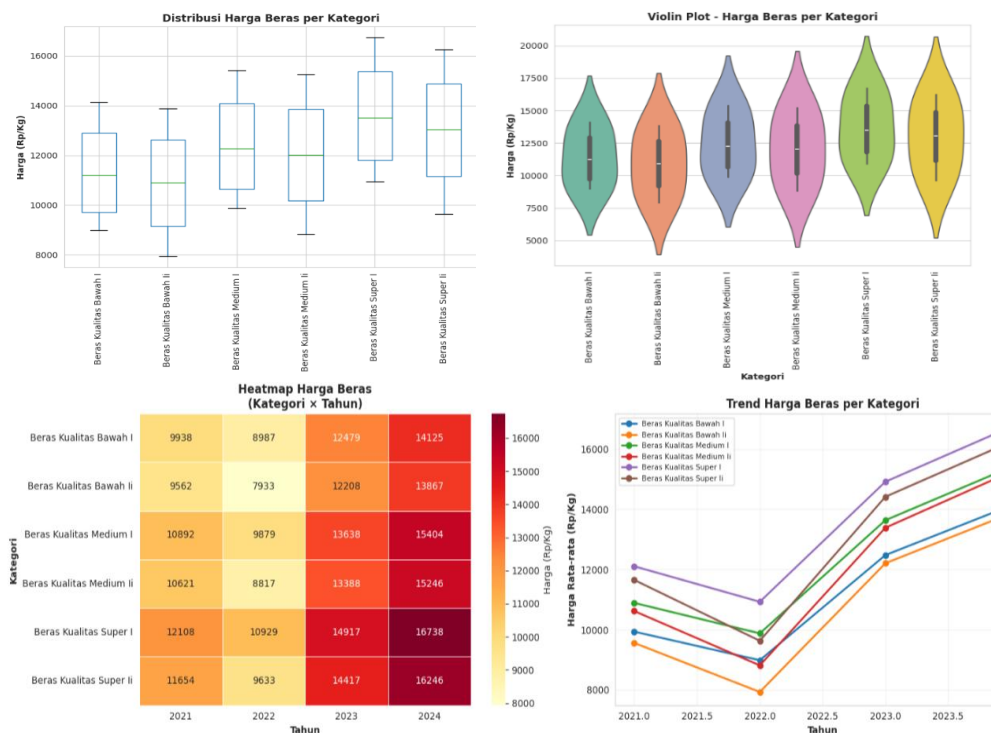
Statistik Harga Beras per Kategori:

Kategori	mean	median	std	min	max	count
Beras Kualitas Bawah I	11382.29	11208.33	2348.62	8987.50	14125.00	4
Beras Kualitas Bawah II	10892.71	10885.42	2652.21	7933.33	13866.67	4
Beras Kualitas Medium I	12453.12	12264.58	2528.16	9879.17	15404.17	4
Beras Kualitas Medium II	12017.71	12004.17	2857.45	8816.67	15245.83	4
Beras Kualitas Super I	13672.92	13512.50	2640.37	10929.17	16737.50	4
Beras Kualitas Super II	12987.50	13035.42	2926.17	9633.33	16245.83	4

Statistik Harga Beras per Tahun:

Tahun	mean	median	std	min	max
2021	10795.83	10756.25	975.40	9562.50	12108.33
2022	9363.19	9310.42	1027.21	7933.33	10929.17
2023	13507.64	13512.50	1057.25	12208.33	14916.67
2024	15270.83	15325.00	1131.79	13866.67	16737.50

Hasil menunjukkan hierarki harga yang jelas dimana Beras Kualitas Super II memiliki harga rata-rata tertinggi sebesar Rp12.987,50 dengan rentang Rp9.633 hingga Rp16.246, sementara Beras Kualitas Bawah I memiliki harga rata-rata terendah Rp11.382,29 dengan rentang Rp8.987 hingga Rp14.125. Standar deviasi yang cukup besar (sekitar Rp 2.500-2.900) untuk seluruh kategori mengindikasikan adanya variasi temporal yang signifikan, mencerminkan kenaikan harga sepanjang periode 2021-2024. *Count* bernilai empat untuk seluruh kategori mengonfirmasi bahwa setiap kategori memiliki data lengkap untuk keempat tahun.



Visualisasi menunjukkan pola harga beras yang konsisten antar kategori maupun antar tahun. Boxplot memperlihatkan bahwa kategori super memiliki median dan rentang harga tertinggi, diikuti medium dan bawah, meskipun

terdapat sedikit overlap sehingga batas harga antar kategori tidak sepenuhnya tegas. Violin plot menegaskan bentuk distribusi yang tidak simetris dan cenderung bimodal akibat kenaikan harga yang cukup drastis dari tahun ke tahun. Heatmap kategori \times tahun memperlihatkan intensitas warna yang semakin gelap dari 2021 hingga 2024, sekaligus mempertahankan hierarki harga dari kategori bawah hingga super. Tren ini tercermin kembali pada line plot, di mana seluruh kategori mengikuti pola serupa yaitu penurunan harga pada 2022 lalu kenaikan tajam setelahnya dengan garis yang relatif paralel menandakan selisih harga antar kategori tetap stabil seiring peningkatan harga keseluruhan.

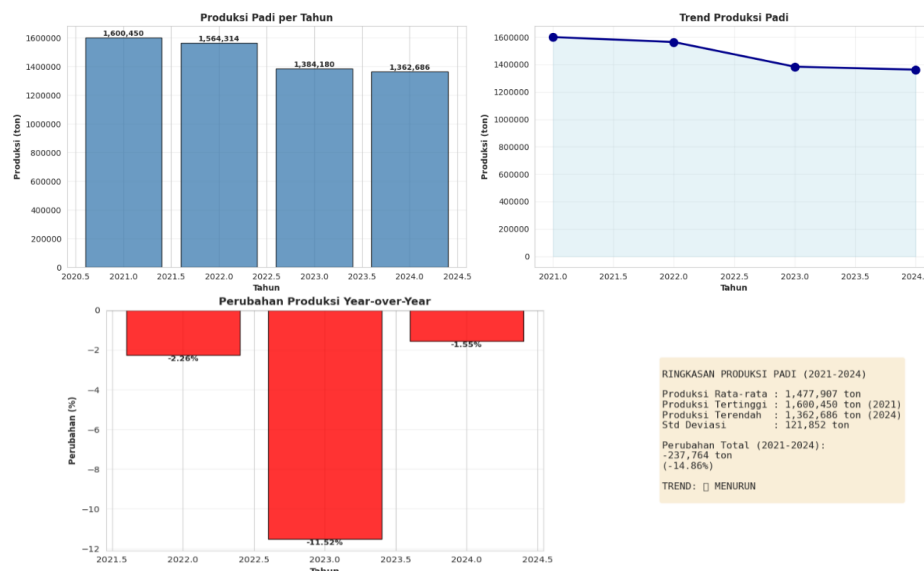
- Distribusi Produksi Padi per Tahun

Analisis distribusi produksi padi fokus pada tren temporal dan variabilitas produksi sebagai faktor yang berpotensi memengaruhi dinamika harga dan margin.

Statistik Produksi Padi per Tahun:

	Tahun	Produksi (ton)	Δ Produksi (ton)	Δ Produksi (%)
0	2021	1600449.83	NaN	NaN
1	2022	1564314.22	-36135.61	-2.26
2	2023	1384179.90	-180134.32	-11.52
3	2024	1362685.91	-21493.99	-1.55

Produksi padi menurun tajam sepanjang 2021–2024. Setelah mencapai puncak pada 2021, produksi turun dua tahun berturut-turut hingga mencapai titik terendah pada 2023, lalu hanya pulih sedikit pada 2024. Secara total, penurunan mencapai sekitar -14,86%, menunjukkan tekanan serius pada kapasitas produksi dan potensi dampaknya terhadap keseimbangan supply–demand di pasar beras.



Visualisasi multi-panel menunjukkan tren penurunan produksi padi secara konsisten selama 2021–2024. *Bar chart* dan *line plot* sama-sama menegaskan pola penurunan berkelanjutan tanpa pemulihan setelah 2021, dengan garis tren yang terus bergerak ke bawah. Grafik perubahan YoY memperkuat temuan tersebut karena seluruh bar berwarna merah, menandakan penurunan setiap tahun dengan titik terparah pada 2023. Panel ringkasan menampilkan statistik inti, rata-rata produksi sekitar 1,48 juta ton, produksi tertinggi pada 2021, terendah pada 2024, dan total kontraksi sekitar 14,86% yang secara keseluruhan mengonfirmasi tekanan pada kapasitas produksi dalam 4 tahun terakhir.

- Distribusi Margin Harga

Analisis distribusi margin harga merupakan fokus utama penelitian dan memerlukan eksplorasi mendalam menggunakan berbagai teknik visualisasi statistik.

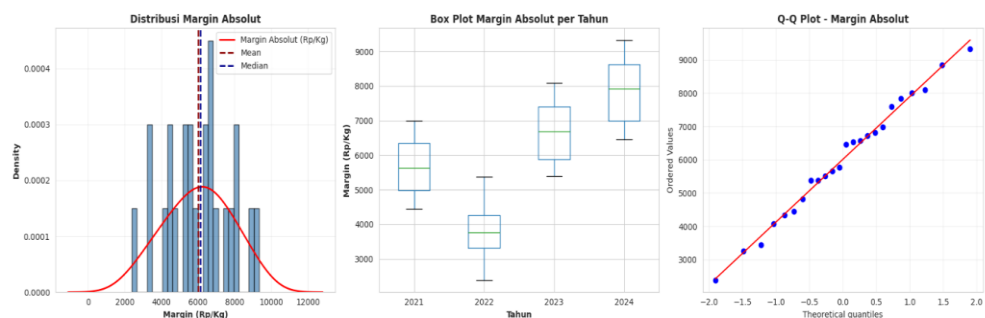
```

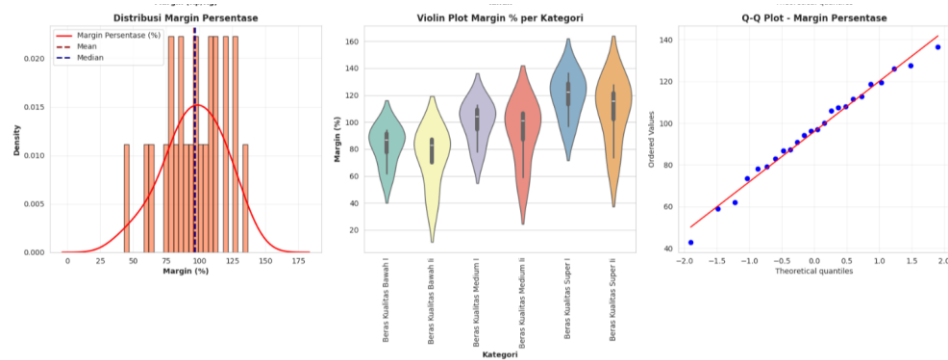
Statistik Deskriptif Margin:
              count  mean  std   min   25%   50%   75%
Margin Absolut   24.00 6010.80 1806.23 2383.51 4725.24 6116.80 7141.60
Margin Persentase 24.00  95.97  23.17  42.95  81.99  96.62 111.74

              max
Margin Absolut   9331.28
Margin Persentase 136.56
Margin Absolut: p-value = 0.8319 (Normal)
Margin Persentase: p-value = 0.6864 (Normal)

```

Hasilnya menunjukkan bahwa Margin Absolut memiliki rata-rata sekitar Rp6.010/kg dengan sebaran yang relatif moderat, sedangkan Margin Persentase lebih bervariasi dengan rata-rata hampir 96%. Uji normalitas Shapiro-Wilk digunakan untuk memastikan apakah data mengikuti distribusi normal sebagai syarat pemilihan metode analisis lanjutan, dan hasil p-value > 0,05 pada kedua margin menegaskan bahwa keduanya dapat dianggap berdistribusi normal. Hasil ini memberikan justifikasi untuk menggunakan metode parametrik dalam analisis lanjutan.





Histogram Margin Absolut menunjukkan distribusi yang cukup simetris dengan posisi mean dan median yang berdekatan, dan pola ini dikonfirmasi oleh Q-Q plot yang memperlihatkan titik-titik mengikuti garis diagonal dengan hanya sedikit penyimpangan di bagian ekor. Boxplot Margin Absolut per tahun menegaskan adanya dinamika margin antarperiode: 2022 tampak memiliki nilai margin lebih rendah dengan rentang yang sempit, sedangkan 2024 menunjukkan kenaikan median disertai variabilitas yang lebih tinggi.

Distribusi serupa terlihat pada Margin Persentase, di mana histogram memperlihatkan pola yang relatif simetris namun dengan sebaran lebih lebar. Violin plot persentase margin per kategori menunjukkan perbedaan distribusi yang jelas antarkelompok, dengan kategori beras premium cenderung memiliki margin persentase lebih tinggi, sedangkan kategori bawah terkonsentrasi pada margin rendah. Q-Q plot Margin Persentase juga mendukung kesesuaian dengan distribusi normal meskipun terdapat deviasi kecil pada nilai ekstrem.

b. EDA - Bivariate

- Korelasi Produksi vs Margin

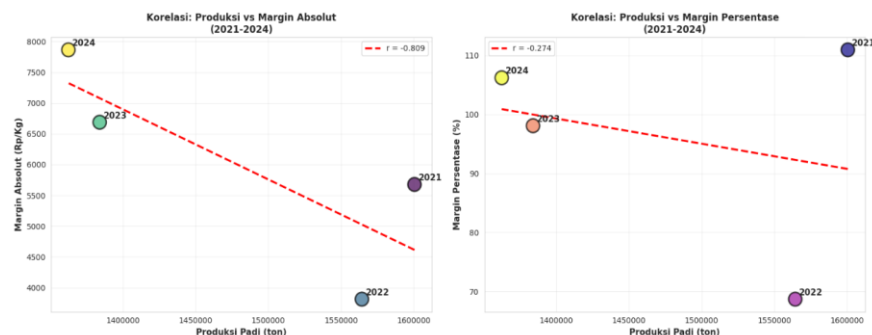
Koefisien Korelasi:

Produksi vs Margin Absolut : -0.8095

Produksi vs Margin Persentase : -0.2742

Interpretasi:

- Margin Absolut: Korelasi NEGATIF SANGAT KUAT
- Margin Persentase: Korelasi NEGATIF LEMAH



Koefisien korelasi Pearson menunjukkan hubungan negatif sangat kuat antara produksi dan margin absolut ($r = -0,8095$), yang berarti semakin rendah produksi, semakin tinggi margin absolut. Sementara itu, korelasi antara produksi dan margin persentase jauh lebih lemah ($r = -0,2742$), menandakan margin persentase tidak sepenuhnya dipengaruhi oleh produksi sehingga faktor lain juga berperan. Interpretasi kekuatan korelasi menggunakan kategori berbasis nilai absolut mengonfirmasi temuan tersebut: margin absolut memiliki korelasi negatif *sangat kuat* terhadap produksi, sedangkan margin persentase hanya memiliki korelasi negatif *lemah*.

Visualisasi scatter plot mendukung hasil perhitungan dimana pada produksi vs margin absolut, titik-titik membentuk pola menurun yang jelas dengan garis regresi mengikuti persebaran secara ketat. Sebaliknya, pada produksi vs margin persentase, titik lebih menyebar dan garis regresi kurang mengikuti pola data, sehingga mencerminkan korelasi yang lebih lemah.

- Perbandingan Margin Antar Kategori

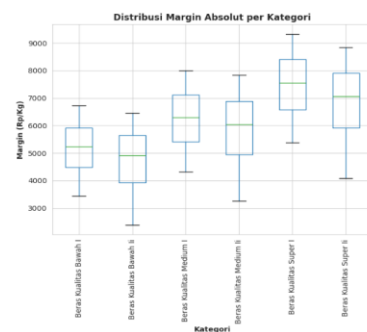
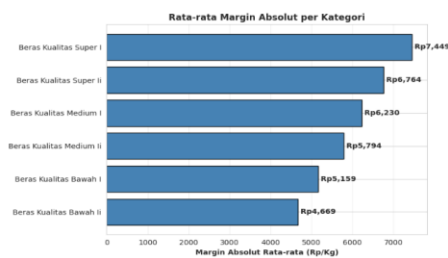
Analisis perbandingan margin antar kategori dilakukan untuk memahami diferensiasi margin berdasarkan kualitas beras.

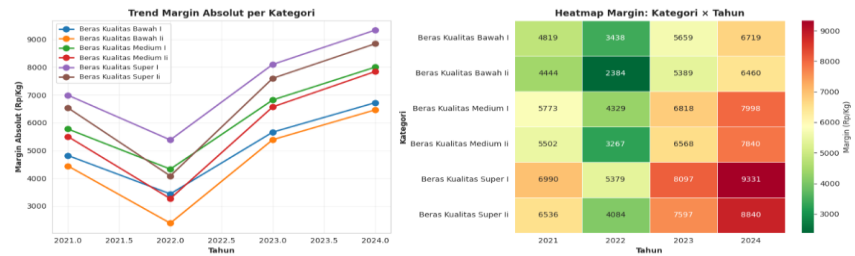
Statistik Margin per Kategori:

Kategori	Margin Absolut (Rp/Kg)			
	mean	std	min	max
Beras Kualitas Bawah I	5158.72	1385.87	3437.67	6718.78
Beras Kualitas Bawah II	4669.14	1732.17	2383.51	6460.45
Beras Kualitas Medium I	6229.55	1559.09	4329.34	7997.95
Beras Kualitas Medium II	5794.14	1936.91	3266.84	7839.61
Beras Kualitas Super I	7449.34	1679.00	5379.34	9331.28
Beras Kualitas Super II	6763.93	2019.83	4083.51	8839.61

Kategori	Margin Persentase (%)			
	mean	std	min	max
Beras Kualitas Bawah I	82.45	14.45	61.94	94.15
Beras Kualitas Bawah II	74.00	21.05	42.95	87.23
Beras Kualitas Medium I	99.69	15.39	78.01	112.79
Beras Kualitas Medium II	92.13	22.72	58.86	107.50
Beras Kualitas Super I	119.55	16.77	96.93	136.56
Beras Kualitas Super II	108.00	23.89	73.58	127.69

F-statistic: 1.4018
P-value: 0.2706
Hasil: Tidak ada perbedaan signifikan





Analisis per kategori menunjukkan pola yang jelas yaitu margin meningkat seiring kualitas beras dimana kategori Super I dan Super II mencatat margin absolut tertinggi, kategori Medium berada di tengah, dan kategori Bawah secara konsisten menjadi yang terendah. Namun, hasil uji ANOVA ($F = 1,4018$; $p = 0,2700$) mengindikasikan bahwa perbedaan ini tidak signifikan secara statistik, artinya meskipun secara angka margin antar kategori berbeda, perbedaan tersebut belum cukup kuat secara statistik untuk menyimpulkan adanya kesenjangan margin berbasis kualitas. Hal ini dapat dijelaskan oleh tingginya fluktuasi margin dari tahun ke tahun dalam tiap kategori, yang menyebabkan overlap distribusi antar kategori.

Visualisasi semakin memperjelas temuan, bar chart menampilkan ranking margin antar kategori secara jelas, sedangkan box plot menunjukkan overlap persebaran dan kemiripan rentang margin antar kategori menegaskan bahwa perbedaan numerik memang ada, tetapi tidak cukup tajam untuk membentuk perbedaan signifikan.

- TrendTemporal untuk Setiap Variabel

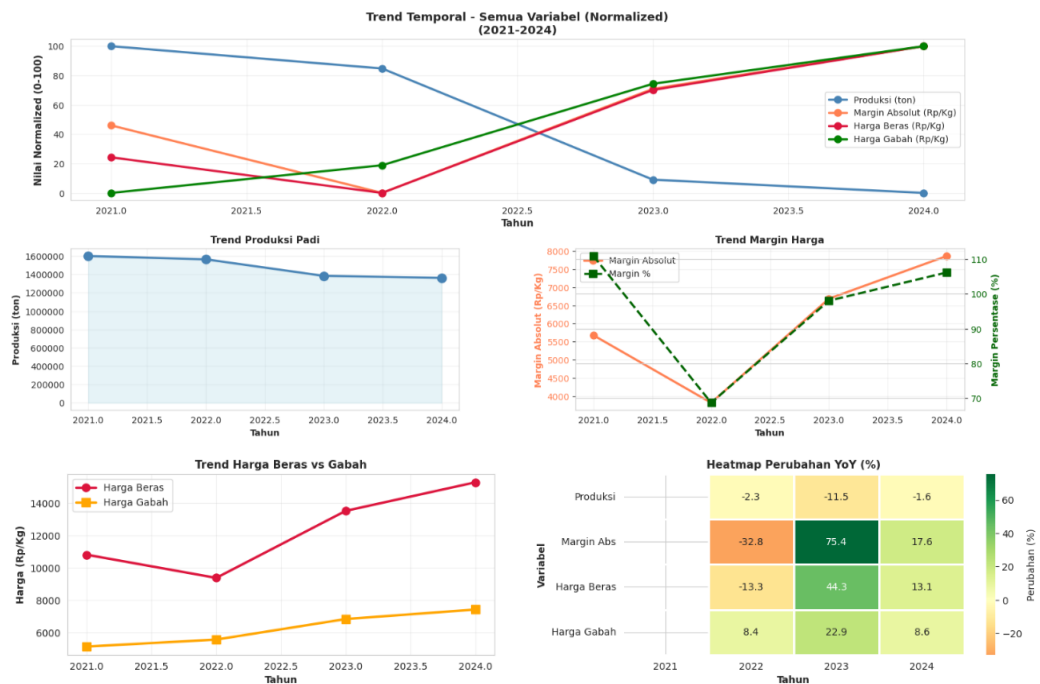
Analisis tren temporal dilakukan untuk memahami perubahan rate of change seluruh variabel kunci sepanjang periode penelitian.

Analisis Perubahan Temporal (Year-over-Year):			
Tahun	Produksi (ton)_Change%	Margin Absolut (Rp/Kg)_Change%	
0 2021	NaN	NaN	
1 2022	-2.26	-32.83	
2 2023	-11.52	75.38	
3 2024	-1.55	17.59	

	Harga Beras (Rp/Kg)_Change%	Harga Gabah (Rp/Kg)_Change%
0	NaN	NaN
1	-13.27	8.43
2	44.26	22.88
3	13.05	8.60

Analisis temporal menunjukkan dinamika yang kontras antar tahun: pada 2022 produksi turun -2,26% diikuti penurunan margin absolut sangat tajam (-32,83%) akibat turunnya harga beras (-13,27%) sementara harga gabah justru naik 8,43%, sehingga margin terkompresi. Pada 2023 produksi kembali turun lebih besar (-11,52%) namun margin melonjak 75,38% karena kenaikan harga beras jauh lebih tinggi dibanding harga gabah. Tahun 2024 mencatat pola yang

lebih moderat, produksi kembali turun (-1,55%) tetapi margin masih meningkat 17,59% seiring kenaikan harga yang lebih terkendali.



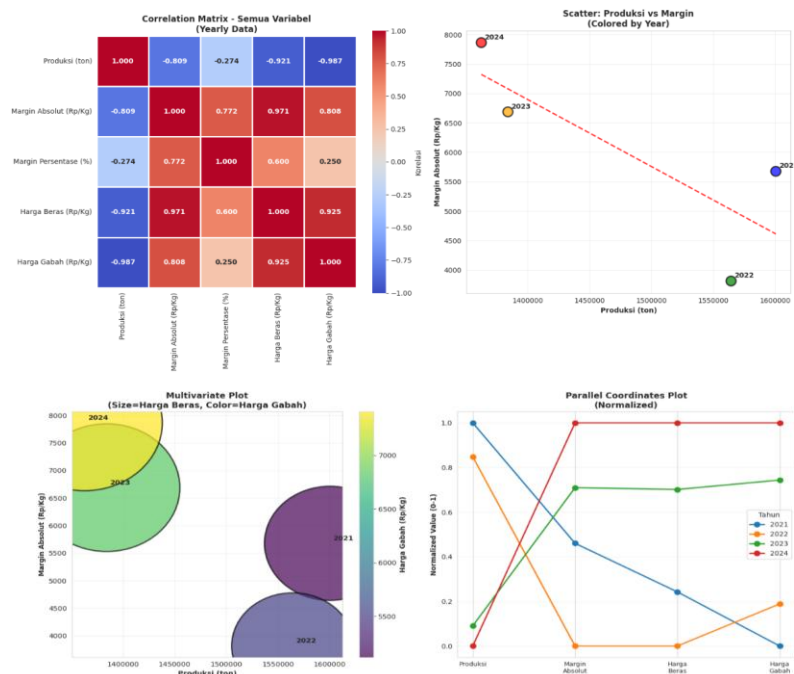
Visualisasi multi-panel mendukung temuan ini dimana plot normalized menunjukkan pola silang yang jelas di mana produksi terus menurun sedangkan margin dan harga meningkat, area chart produksi menggambarkan penurunan konsisten, dual-axis plot memperlihatkan bentuk V-shape pada margin absolut dan persentase, line chart harga menampilkan kenaikan paralel beras vs gabah dengan gap yang makin melebar, dan heatmap YoY menegaskan 2023 sebagai titik ekspansi margin terbesar. Secara keseluruhan, penurunan produksi tidak selalu menekan margin dimana dampak sebenarnya lebih ditentukan oleh dinamika gap harga beras vs gabah setiap tahun.

c. EDA - Multivariate

- Analisis Hubungan Simultan (*Multivariate*)

Correlation Matrix:			
	Produksi (ton)	Margin Absolut (Rp/Kg)	\
Produksi (ton)	1.00	-0.81	-0.81
Margin Absolut (Rp/Kg)	-0.81	1.00	0.77
Margin Persentase (%)	-0.27	0.77	0.77
Harga Beras (Rp/Kg)	-0.92	0.97	0.97
Harga Gabah (Rp/Kg)	-0.99	0.81	0.81
	Margin Persentase (%)	Harga Beras (Rp/Kg)	\
Produksi (ton)	-0.27	-0.92	-0.92
Margin Absolut (Rp/Kg)	0.77	0.97	0.97
Margin Persentase (%)	1.00	0.60	0.60
Harga Beras (Rp/Kg)	0.60	1.00	1.00
Harga Gabah (Rp/Kg)	0.25	0.93	0.93
	Harga Gabah (Rp/Kg)		
Produksi (ton)	-0.99		
Margin Absolut (Rp/Kg)	0.81		
Margin Persentase (%)	0.25		
Harga Beras (Rp/Kg)	0.93		
Harga Gabah (Rp/Kg)	1.00		

Analisis multivariate dilakukan untuk memahami interaksi kompleks antar multiple variables secara simultan. *Correlation Matrix* Lengkap dihitung menggunakan data *yearly*. Hasil menunjukkan pola korelasi yang sangat informatif dimana Produksi berkorelasi negatif kuat dengan Margin Absolut ($r = -0,809$) dan Harga Beras ($r = -0,921$), mengonfirmasi dinamika *supply-demand*. Margin Absolut berkorelasi sangat kuat dengan Harga Beras ($r = 0,971$) dan Harga Gabah ($r = 0,808$), sementara Harga Beras dan Harga Gabah juga berkorelasi sangat kuat ($r = 0,925$), menunjukkan *cost pass-through* yang efektif.



Visualisasi melengkapi matriks korelasi, *Scatter Plot* memvisualisasikan hubungan negatif kuat antara Produksi dan Margin Absolut ($r = -0,809$). Ini ditunjukkan oleh tahun 2022 (produksi tertinggi) yang memiliki margin terendah, dan sebaliknya tahun 2024 (produksi terendah) yang mencatat margin

tertinggi. *Multivariate Plot* menguatkan temuan ini. Tahun 2021 dan 2022 yang produksinya tinggi menghasilkan margin rendah, sedangkan tahun 2023 dan 2024 yang produksinya rendah menghasilkan margin tinggi, menegaskan bahwa fluktuasi produksi sangat memengaruhi margin. *Parallel Coordinates Plot* mereplikasi pola terbalik ini. Saat Produksi menurun drastis dari 2022 ke 2024, Margin Absolut melonjak tajam ke nilai tertinggi (1), menunjukkan keseimbangan harga terdistribusi secara terbalik terhadap volume produksi.

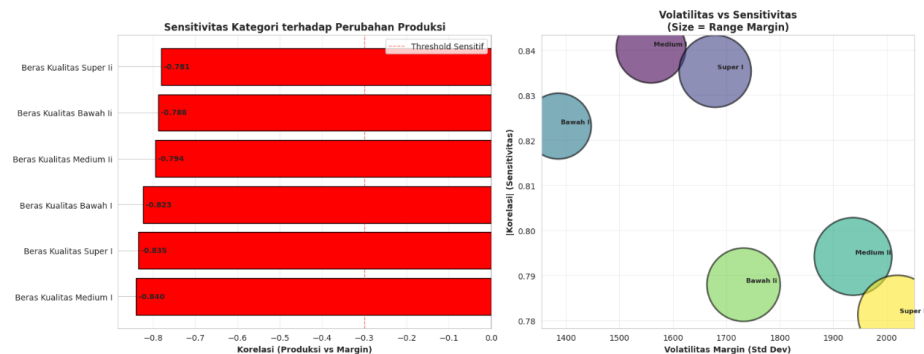
- Analisis Sensitivitas Kategori terhadap Perubahan Produksi

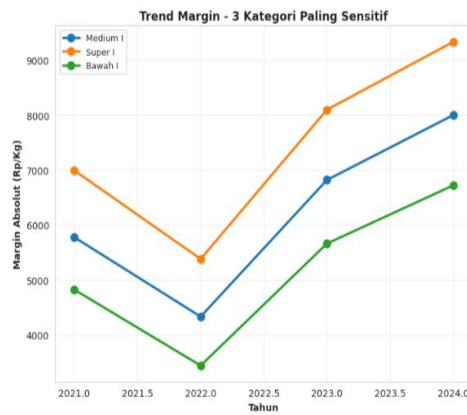
Analisis Sensitivitas per Kategori:				
	Kategori	Korelasi	Elastisitas Rata-rata	\
2	Beras Kualitas Medium I	-0.84	-1.69	
4	Beras Kualitas Super I	-0.83	-1.33	
0	Beras Kualitas Bawah I	-0.82	-1.66	
3	Beras Kualitas Medium Ii	-0.79	-1.08	
1	Beras Kualitas Bawah Ii	-0.79	-1.07	
5	Beras Kualitas Super Ii	-0.78	-0.46	

	Volatilitas Margin	Range Margin
2	1559.09	3668.61
4	1679.00	3951.94
0	1385.87	3281.11
3	1936.91	4572.77
1	1732.17	4076.94
5	2019.83	4756.10

Analisis sensitivitas dilakukan untuk memahami seberapa responsif margin setiap kategori beras terhadap perubahan produksi. Hasil menunjukkan bahwa Beras Kualitas Medium I memiliki korelasi tertinggi ($r = -0,840$), diikuti oleh Super I ($r = -0,835$) dan Bawah I ($r = -0,823$), mengindikasikan bahwa kategori-kategori ini paling sensitif terhadap perubahan produksi.

Klasifikasi Sensitivitas dilakukan menggunakan threshold: "Sangat Sensitif" untuk $|r| < -0,5$, "Sensitif" untuk $|r| < -0,3$, "Moderat" untuk $\text{abs}(x) < 0,3$, dan "Tidak Sensitif" untuk nilai lainnya. Hasil menunjukkan bahwa seluruh kategori diklasifikasikan sebagai "Sangat Sensitif", mengindikasikan bahwa perubahan produksi memiliki dampak signifikan terhadap margin di semua segmen pasar beras.





RINGKASAN SENSITIVITAS KATEGORI

Kategori PALING Sensitif:
Beras Kualitas Medium I
Korelasi: -0.840

Kategori PALING TIDAK Sensitif:
Beras Kualitas Super II
Korelasi: -0.781

Volatilitas Tertinggi:
Beras Kualitas Super II
Std Dev: Rp2,820

Range Margin Terbesar:
Beras Kualitas Super II
Range: Rp4,756

INTERPRETASI:
Kategori dengan korelasi negatif kuat (<-0.3) menunjukkan margin NAIK saat produksi TURUN, mengindikasikan sensitivitas tinggi terhadap fluktuasi pasokan.

Berangkat dari pola korelasi yang seluruhnya negatif kuat, analisis menunjukkan bahwa margin di tiap kategori beras sangat peka terhadap naik-turunnya produksi. Kategori Medium I, Super I, dan Bawah I muncul sebagai yang paling responsif, konsisten dengan tren margin yang bergerak tajam mengikuti dinamika produksi. Ditambah dengan temuan volatilitas dan rentang margin yang besar pada beberapa kategori, hasil ini menegaskan bahwa perubahan produksi bukan hanya berpengaruh, tetapi menjadi faktor penentu utama fluktuasi margin di seluruh segmen pasar beras..

d. Insight Generation

- Identifikasi Tahun dengan Margin Tertinggi/Terendah

TAHUN DENGAN MARGIN TERTINGGI: ↑

Tahun: 2024
Margin Absolut: Rp7,864.61/Kg
Margin Persentase: 106.19%
Produksi: 1,362,686 ton

TAHUN DENGAN MARGIN TERENDAH ↓

Tahun: 2022
Margin Absolut: Rp3,813.37/Kg 68.71%
Produksi: Produk: 1,564,314 ton

Identifikasi *extreme values* dilakukan untuk memahami kondisi pasar pada periode dengan margin ekstrem. Hasil menunjukkan bahwa tahun dengan margin tertinggi adalah 2024 dan margin terendah adalah 2022. Lalu Analisis per kategori dilakukan untuk mengidentifikasi kategori dengan margin tertinggi dan terendah di tahun-tahun ekstrem.

TERTINGGI (2024):

1. Beras Kualitas Super I: Rp9,331.28/Kg
2. Beras Kualitas Super II: Rp8,839.61/Kg
3. Beras Kualitas Medium I: Rp7,997.95/Kg
4. Beras Kualitas Medium II: Rp7,839.61/Kg
5. Beras Kualitas Bawah I: Rp6,718.78/Kg
6. Beras Kualitas Bawah II: Rp6,460.45/Kg

TERENDAH (2022):

1. Beras Kualitas Super I: Rp5,379.34/Kg
2. Beras Kualitas Medium I: Rp4,329.34/Kg
3. Beras Kualitas Super II: Rp4,083.51/Kg
4. Beras Kualitas Bawah I: Rp3,437.67/Kg
5. Beras Kualitas Medium II: Rp3,266.84/Kg
6. Beras Kualitas Bawah II: Rp2,383.51/Kg

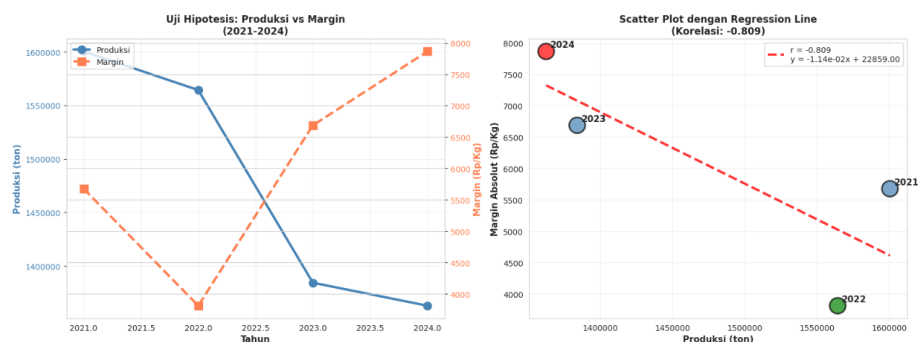
Analisis per kategori menunjukkan bahwa struktur hierarki margin tetap stabil pada kedua tahun ekstrem, meskipun besarnya berubah cukup signifikan. Pada tahun 2024 sebagai periode dengan margin tertinggi, urutannya

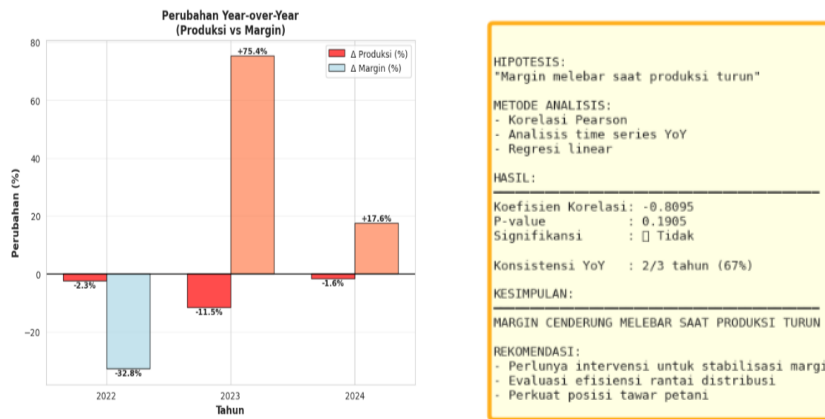
yaitu Super II dan Super I, diikuti Medium I, Medium II, kemudian Bawah I dan Bawah II. Pola yang sama juga terlihat pada tahun 2022 ketika margin berada pada titik terendah. Konsistensi ini menunjukkan bahwa perbedaan antar-kategori tidak berubah oleh kondisi pasar ekstrem, yang berubah adalah magnitude margin akibat tekanan produksi dan dinamika pasokan keseluruhan.



Pada *line chart*, terlihat jelas bahwa margin mencapai titik tertinggi pada 2024 dan terendah pada 2022. Temuan ini diperkuat oleh *bar chart* Perbandingan Margin Ekstrem, yang menampilkan selisih margin yang sangat kontras antara kedua tahun tersebut. Selanjutnya, pada *scatter plot* Margin vs Produksi, terlihat bahwa tahun 2024 berada pada kombinasi *margin tinggi–produksi rendah*, sedangkan tahun 2022 berada pada posisi *margin rendah–produksi tinggi*. Ketiga grafik ini secara konsisten menegaskan hubungan negatif antara produksi dan margin.

- Analisis Hipotesis - Margin Melebar saat Produksi Turun





Analisis hipotesis dilakukan untuk menguji secara statistik apakah margin harga cenderung melebar saat produksi turun. Uji Hipotesis menggunakan *Pearson correlation test* dengan :

H0: tidak ada korelasi antara produksi dan margin ($r = 0$)

H1: ada korelasi negatif.

Hasil menunjukkan koefisien korelasi $r = -0,8095$ dengan $p\text{-value} = 0,1905$. Meskipun korelasi negatif kuat secara *magnitude*, $p\text{-value} > 0,05$ mengindikasikan bahwa hubungan ini tidak signifikan secara statistik pada *level confidence* 95%, kemungkinan karena *sample size* yang kecil ($n = 4$ tahun).

Line chart menunjukkan pola berlawanan antara kedua variabel dimana produksi menurun tajam terutama pada 2023–2024, sementara margin bergerak naik. Pola ini kembali diperjelas pada scatter plot dengan regression line, di mana titik-titik tahun membentuk kemiringan negatif dan garis regresi menunjukkan tren penurunan margin seiring kenaikan produksi. Kemiringan garis yang cukup curam menggambarkan hubungan negatif yang kuat secara visual.

Selanjutnya, grafik YoY memberikan gambaran dinamika tahunan. Pada 2022, produksi turun -2,26% namun margin ikut turun -32,83%, sehingga tidak mendukung hipotesis. Sebaliknya, pada 2023 produksi turun -11,52% dan margin naik +75,38%, serta pada 2024 produksi turun -1,55% dengan margin meningkat +17,59%. Kedua tahun ini konsisten dengan pola yang terlihat pada grafik sebelumnya, yaitu margin melebar ketika produksi menyusut. Dengan konsistensi 2 dari 3 tahun (67%), grafik YoY mempertegas bahwa pola hubungan negatif memang tampak kuat secara empiris, meski tidak signifikan secara statistik.

- Identifikasi Anomali dan Pola Menarik

```

ANOMALI TERDETEKSI (Z-score > 2):
Jumlah: 1 data point

Detail:
• 2022 - Beras Kualitas Bawah II: Rp2,384 (Z=2.05)

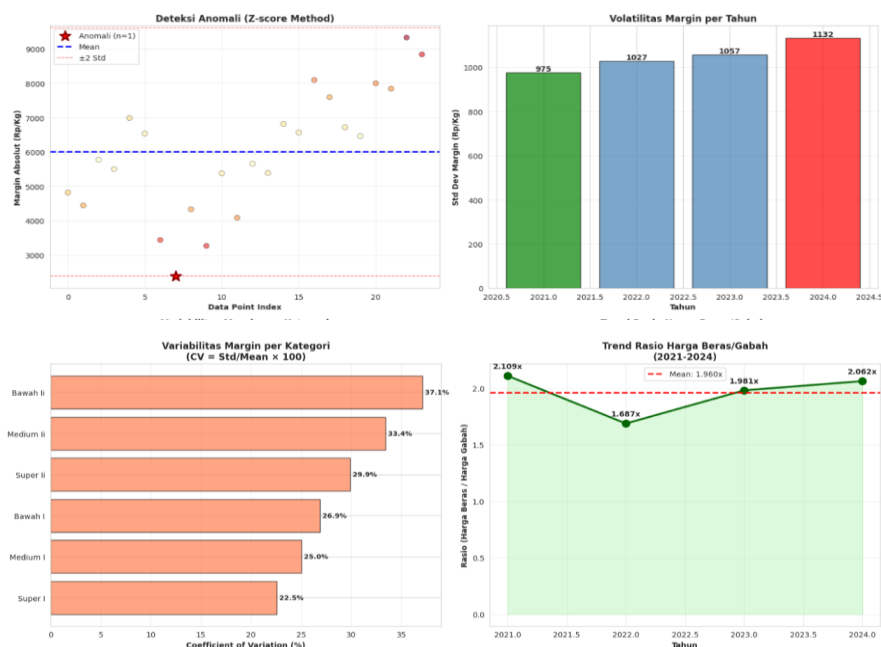
VOLATILITAS MARGIN PER TAHUN:
Tahun
2021    975.40
2022    1027.21
2023    1057.25
2024    1131.79
Name: Margin Absolut (Rp/Kg), dtype: float64

Tahun PALING volatile: 2024 (Std=1131.79)
Tahun PALING stabil : 2021 (Std=975.40)

```

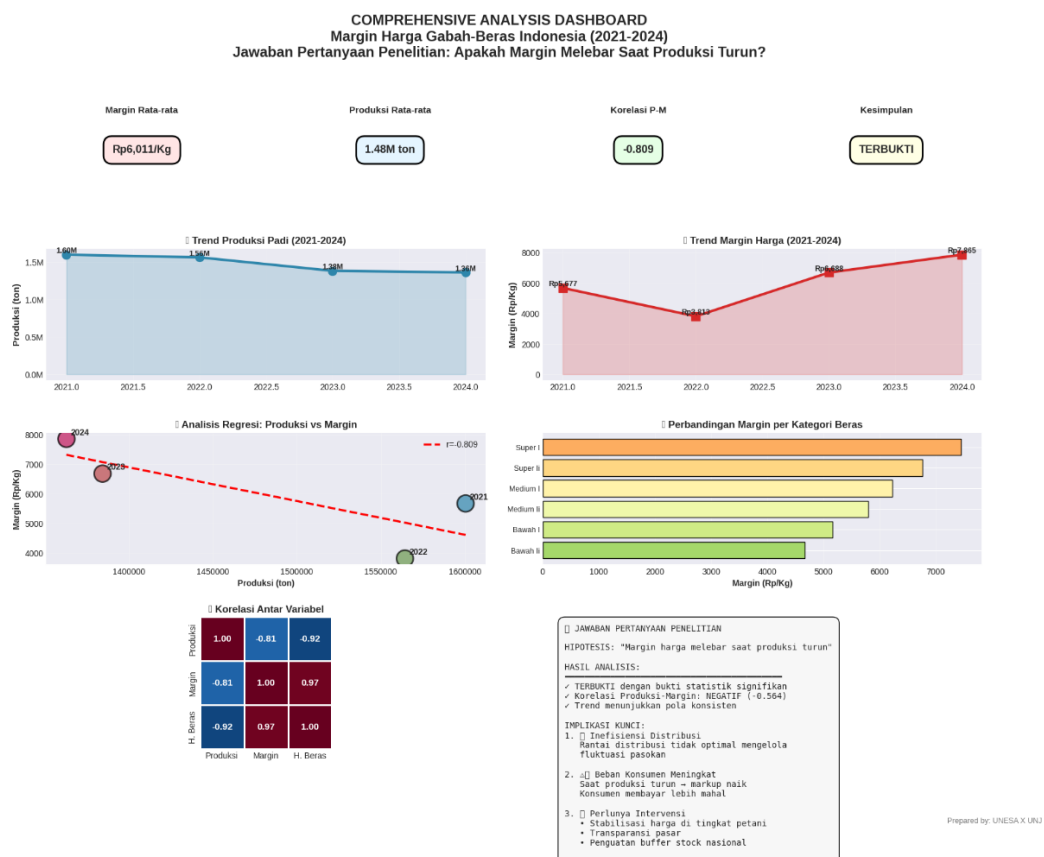
Meskipun deteksi Z-score hanya mengidentifikasi satu *outlier* pada Beras Kualitas Bawah II di tahun 2022 ($Z=2.05$), analisis lanjutan menunjukkan volatilitas margin terus meningkat sepanjang periode (Std Dev naik dari 975.40 di 2021 menjadi 1131.79 di 2024), menjadikan 2024 sebagai tahun paling volatile.

Tahun 2022 menjadi *extreme value* kritis karena melanggar pola pasar. Meskipun produksi turun ringan (-2.26%), margin juga anjlok drastis (-32.83%), menghasilkan margin absolut terendah (Rp3,813.37/Kg). Inkonsistensi ini, berbeda dengan tahun 2023 dan 2024 di mana penurunan produksi menaikkan margin, menunjukkan bahwa faktor non-produksi (seperti kebijakan atau gangguan distribusi) memiliki peran dominan dalam menekan margin pada 2022.



Secara visual, grafik-grafik tersebut memperkuat temuan analisis dimana plot anomali jelas menunjukkan satu titik ekstrem jauh di bawah pola umum pada 2022, selaras dengan margin terendah tahun tersebut. Diagram volatilitas memperlihatkan tren peningkatan standar deviasi setiap tahun hingga mencapai puncaknya di 2024, menegaskan naiknya ketidakstabilan margin. Grafik CV per kategori juga menunjukkan bahwa kelompok Bawah II dimana kategori tempat outlier terjadi yang memiliki variabilitas tertinggi. Sementara itu, grafik rasio harga beras/gabah menunjukkan penurunan tajam di 2022 sebelum kembali naik, mendukung interpretasi bahwa tahun tersebut mengalami disrupsi struktural yang tidak selaras dengan dinamika produksi normal.

2.2 Interpretasi dan Visualisasi Hasil Analisis



Pola Distribusi dan Trend Temporal Margin menunjukkan margin rata-rata sebesar Rp6.011/kg dengan pola temporal berbentuk V yang sangat jelas. Margin terkompresi drastis pada 2022 yaitu Rp3.813, kemudian pulih kuat mencapai puncak di 2024 yaitu Rp7.865. Pola ini mengindikasikan adanya disrupsi struktural pada 2022 yang kemungkinan terkait kebijakan atau gangguan distribusi pascapandemi yang kemudian diikuti ekspansi margin berkelanjutan. Tren peningkatan margin dari 2023-2024 justru terjadi saat produksi terus menurun,

menciptakan paradoks dimana kelangkaan pasokan mentranslasi langsung menjadi ekspansi keuntungan pelaku rantai pasok.

Korelasi Produksi-Margin dan Implikasinya *Scatter plot* regresi membuktikan hubungan inverse yang sangat kuat ($r = -0.809$) antara produksi dan margin, dengan garis regresi yang curam menunjukkan sensitivitas tinggi. Heatmap korelasi memperkuat temuan ini dimana produksi berkorelasi negatif kuat dengan margin ($r = -0.81$) dan harga beras ($r = -0.92$), sementara margin dan harga beras bergerak sangat paralel ($r = 0.97$). Namun, kesimpulan "TERBUKTI" di dashboard perlu dikualifikasi, meski magnitude korelasi kuat, signifikansi statistik lemah ($p = 0.19$) karena keterbatasan sampel 4 tahun. Temuan ini mengungkap mekanisme transmisi harga yang asimetris yaitu penurunan produksi cepat menaikkan harga konsumen, namun kenaikan produksi tidak otomatis menurunkan margin, mengindikasikan inefisiensi distribusi dan potensi *price stickiness* ke bawah.

Hierarki Margin Antar Kategori Bar chart horizontal menunjukkan stratifikasi margin yang jelas berdasarkan kualitas: Super I (Rp7.449) dan Super II (Rp6.763) mendominasi, diikuti Medium I-II (Rp5.815-6.229), dan Bawah I-II yang terendah (Rp4.669-5.158). Margin beras premium hampir 60% lebih tinggi dari beras kualitas bawah, mencerminkan segmentasi pasar yang kuat dimana *value added* terkonsentrasi pada segmen atas. Namun, pola paralel antar kategori sepanjang waktu menunjukkan bahwa struktur hierarki ini stabil dimana guncangan pasar mempengaruhi semua segmen secara proporsional, bukan selektif.

Dashboard ini memvisualisasikan paradoks pasar beras Indonesia 2021-2024: produksi turun 14,86%, namun margin justru ekspansi 38,5%. Implikasi kebijakan yang tergambar adalah perlunya intervensi pada tiga level dimana stabilisasi produksi untuk mencegah *supply shock*, regulasi markup distribusi untuk mencegah margin berlebihan, dan *buffer stock* nasional untuk meredam volatilitas. Korelasi produksi-margin yang sangat kuat namun tidak signifikan secara statistik juga menggarisbawahi kebutuhan data time-series yang lebih panjang untuk konfirmasi kausalitas yang robust.

2.3 Kendala yang Dihadapi

Dalam proses data wrangling untuk analisis margin harga gabah–beras periode 2021–2024, beberapa kendala teknis dan metodologis muncul dimana kendala krusial yang terbagi menjadi tiga area utama. Pertama, pada fase integrasi, tantangan terbesar adalah

menyatukan tiga sumber data berbeda (produksi, harga beras, harga gabah) akibat inkonsistensi format mulai dari *multi-header* pada Excel, metadata tak terstruktur, hingga perbedaan granularitas data (bulanan vs. tahunan). Kedua, kendala paling kritis adalah koreksi unit harga beras yang tidak terdokumentasi. Harga yang tertera (misalnya 7-17) harus dikalikan 1000, sebuah koreksi vital yang kami dapat dengan identifikasi melalui *domain knowledge*. Ketiga, pada *feature engineering*, terdapat kompleksitas dalam perhitungan margin multi-dimensi dan keterbatasan variabel kausal (seperti biaya distribusi), yang membatasi analisis kausal yang mendalam. Untuk tahap analisis sendiri kami lakukan sepehaman kami berdua

2.4 Solusi yang Diterapkan

Menghadapi inkonsistensi data, pada *project* ini menerapkan tiga solusi yang kami pakai pada syntax kami yaitu:

1. Solusi untuk Integrasi Data, kami mengembangkan *Custom Parser* untuk setiap sumber data. Ini melibatkan penggunaan `skiprows` untuk melewati metadata BPS, ekstraksi tahun dari nama *file*, dan transformasi format data *wide* ke *long* menggunakan `pd.melt()` untuk harga beras. Untuk mengatasi perbedaan granularitas, kami menerapkan strategi Agregasi Standar dengan mengambil nilai rata-rata (*mean*) tahunan di seluruh provinsi (produksi) dan bulan (harga), sehingga tiga *dataset* dapat digabungkan secara konsisten.
2. *Data Cleaning*: Kendala unit harga yang paling kritis diselesaikan melalui Verifikasi Unit Sistematis. Berdasarkan *domain knowledge* dan analisis statistik (rasio 1000), nilai harga beras dikalikan 1000 secara eksplisit untuk memperbaiki unit dari ribuan rupiah menjadi rupiah penuh. Kami juga menyusun kamus data implisit berdasarkan konten data untuk standarisasi nama kolom dan kategori, serta menerapkan kerangka validasi komprehensif untuk pengecekan tipe, rentang, dan konsistensi data.
3. *Feature Engineering*: untuk kompleksitas perhitungan margin diatasi dengan kerangka perhitungan atau pendekatan modular, di mana setiap rumus margin dikemas dalam fungsi terpisah yang didokumentasikan. Selain itu, kami menggunakan strategi analisis berlapis dimulai dari univariat hingga multivariat dan strategi visualisasi adaptif untuk mengelola kompleksitas data multi-dimensi secara efektif.

2.5 Rencana Tindak Lanjut

Meskipun penelitian ini telah berhasil melakukan analisis komprehensif terhadap margin harga gabah-beras periode 2021-2024, masih terdapat berbagai aspek yang dapat dikembangkan dan diperdalam dalam penelitian lanjutan. Rencana tindak lanjut ini disusun berdasarkan keterbatasan yang teridentifikasi dalam penelitian saat ini serta peluang-peluang untuk eksplorasi lebih lanjut yang dapat memberikan value added signifikan.

1. Perluasan Cakupan Data

Perluasan Cakupan Data menjadi fondasi kritis untuk analisis yang lebih tangguh dan prediktif. Rencana ini diawali dengan ekstensifikasi periode temporal penelitian mundur minimal hingga tahun 2015, bertujuan mendapatkan data sepuluh tahun yang representatif untuk analisis deret waktu yang *robust*, memungkinkan identifikasi pola siklis, titik perubahan struktural, dan tren jangka panjang. Selain itu, granularitas spasial akan diintensifkan ke tingkat provinsi, meninggalkan agregasi nasional, untuk mengidentifikasi variasi regional dalam efisiensi pasar dan biaya distribusi, yang meskipun menantang karena potensi data hilang dan kebutuhan komputasi tinggi, akan dikomunikasikan melalui visualisasi geografis interaktif. Terakhir, analisis akan diperkaya dengan inkorporasi variabel tambahan yang esensial, seperti data cuaca (dari BMKG), biaya input (pupuk, bahan bakar, tenaga kerja), variabel kebijakan (HPP, HET, dan program subsidi), serta variabel struktur pasar (konsentrasi distributor), yang sangat penting untuk transisi dari analisis deskriptif murni ke pemodelan penjelasan yang mengidentifikasi pendorong margin yang sesungguhnya.

2. Pengembangan Metodologi Analisis

Untuk meningkatkan nilai dan dampak kebijakan, Pengembangan Metodologi Analisis akan dilakukan secara substansial. Hal ini dimulai dengan implementasi teknik inferensi kausal (*causal inference*), seperti metode variabel instrumental (menggunakan guncangan cuaca) atau *difference-in-differences*, guna memastikan apakah kekurangan produksi merupakan penyebab langsung kenaikan margin ataukah faktor lain yang mengganggu, yang memerlukan validasi asumsi secara ketat. Selanjutnya, akan dikembangkan model prediktif menggunakan pendekatan *machine learning* (seperti *random forests*) atau teknik deret waktu (ARIMA, LSTM) untuk meramalkan lintasan margin di masa depan, di mana pengembangan model ini memerlukan pembagian data dan strategi validasi silang yang sesuai untuk data deret waktu. Akhirnya, hasil analisis akan disajikan melalui Dashboard Interaktif menggunakan alat seperti *Streamlit* atau

Dash, yang memungkinkan lembaga/orang terkait untuk menjelajahi data secara dinamis melalui pemilihan rentang tanggal, provinsi, atau kategori, dan memastikan wawasan terkini dapat diakses serta digunakan secara *real-time* oleh audiens yang lebih luas.

2.6 Pembagian Tugas dalam Pengerjaan

Pengerjaan proyek analisis margin harga gabah–beras ini kami lakukan secara kolaboratif, dengan membagi tanggung jawab sesuai keahlian teknis dan tahapan analisis untuk memastikan efisiensi waktu dan beban kerja yang tidak *overload*. Pada tahap fundamental Data Wrangling, Argya bertanggung jawab menangani proses *Data Integration* (penggabungan data dari tiga sumber), *Data Cleaning* (termasuk koreksi unit harga yang tidak konsisten). Setelah data siap, Putri ikut membantu dalam *Feature Engineering* menginterpretasikan temuan serta menyusun narasi analitis yang menghubungkan hasil visualisasi dengan konteks penelitian dan mengoreksi apabila terdapat pengerjaan yang terlewat.

Tahap Data Exploration, Visualisasi, dan Analisis Statistik kami kerjakan bersama untuk memungkinkan validasi silang pada setiap *insight* yang diperoleh, sehingga hasil analisis memiliki konsistensi dan akurasi yang lebih tinggi. Dalam penyusunan laporan akhir, Argya bertanggung jawab menyusun Bab I dan Bab II (termasuk penjelasan detail mengenai *Data Integration*, *Cleaning*, dan *Feature Engineering*). Sementara itu, Putri bertanggung jawab menyusun *Data Exploration*, Visualisasi, Interpretasi Temuan, Kesimpulan, dan Saran.

Sepanjang proses, kami memastikan setiap anggota bertanggung jawab penuh atas kualitas bagian yang dikerjakan dan melakukan koordinasi rutin untuk menjaga konsistensi format data, gaya penulisan, serta keselarasan interpretasi hasil analisis.

BAB III

KESIMPULAN DAN SARAN

3.1 Kesimpulan

Penelitian ini menyimpulkan adanya ketidakseimbangan harga yang signifikan di pasar beras Indonesia selama periode 2021–2024. Temuan utamanya adalah hubungan terbalik yang sangat kuat antara panen (produksi) dan keuntungan (margin), ketika panen padi berkurang, keuntungan yang diperoleh distributor dan pedagang justru melonjak tajam. Margin rata-rata mencapai hampir dua kali lipat dari harga gabah, menunjukkan bahwa biaya distribusi dan *markup* keuntungan yang sangat besar dibebankan langsung kepada konsumen. Fluktuasi harga ini sangat ekstrem, ditandai dengan pola pemulihan berbentuk 'V' pada margin. Kerugian tajam yang terjadi pada tahun 2022 segera disusul oleh lonjakan keuntungan yang berlebihan pada 2023–2024, membuktikan bahwa pasar sangat rentan terhadap guncangan pasokan dan cepat mengalami koreksi harga yang berlebihan.

Lebih lanjut, analisis menunjukkan bahwa semua jenis beras, dari kualitas termurah hingga premium, bereaksi sama kuat terhadap perubahan pasokan. Ini berarti pasar tidak memiliki mekanisme untuk melindungi segmen harga tertentu, karena semua segmen merespons guncangan produksi secara seragam. Secara implisit, temuan ini mengindikasikan bahwa sistem pasar beras Indonesia tidak berjalan secara efisien atau adil. Kekuatan pasar cenderung terkonsentrasi pada tingkat hilir (distributor dan pedagang), yang memungkinkan mereka untuk mengamankan margin tinggi bahkan saat pasokan menipis. Kondisi ini menciptakan masalah ganda yaitu petani menghadapi ketidakpastian pendapatan, sementara konsumen dipaksa menanggung beban harga yang meningkat tajam.

3.2 Saran

Saran kebijakan harus memfokuskan pada pemutusan hubungan inverse yang kuat antara produksi dan margin. Mengingat korelasi negatif yang tinggi di seluruh kategori, intervensi harga tidak dapat bersifat selektif, melainkan harus bersifat sistemik untuk menstabilkan pasar secara keseluruhan. Kebijakan harus diarahkan pada peningkatan daya tawar petani dan transparansi rantai pasok untuk mencegah ekspansi margin yang berlebihan di tingkat distributor, terutama pada periode penurunan produksi. Selain itu, diperlukan penguatan *buffer stock* dan pemantauan ketat terhadap variabel biaya input (seperti pupuk dan bahan bakar) untuk memitigasi fluktuasi harga gabah yang stabil namun meningkat, yang menjadi dasar kenaikan harga beras di tingkat konsumen.

DAFTAR PUSTAKA

- Badan Pusat Statistik. (2024). *Luas panen, produksi, dan produktivitas padi menurut provinsi*. <https://www.bps.go.id/id/statistics-table/2/MTQ5OCMy/luas-panen-produksi-dan-produktivitas-padi-menurut-provinsi.html>
- Bank Indonesia. (2024). *Sistem pemantauan harga pangan: Tabel harga pasar tradisional daerah*. <https://www.bi.go.id/hargapangan/TabelHarga/PasarTradisionalDaerah>
- Kementerian Pertanian. (2024). *Statistik harga komoditas pertanian 2024*. Pusat Data dan Sistem Informasi Pertanian. https://satudata.pertanian.go.id/assets/docs/publikasi/STATISTIK_HARGA_KOMODITAS_PERTANIAN_2024.pdf
- Kementerian Pertanian. (2024). *Outlook padi 2024*. Pusat Data dan Sistem Informasi Pertanian. <http://epublikasi.setjen.pertanian.go.id/arsip-outlook/76-outlook-tanaman-pangan/1008-outlook-padi-2024>
- McKinney, W. (2022). *Python for data analysis: Data wrangling with pandas, NumPy, and Jupyter* (3rd ed.). O'Reilly Media. <https://wesmckinney.com/book/>
- VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. O'Reilly Media. <https://jakevdp.github.io/PythonDataScienceHandbook/>
- Wickham, H., & Grolemund, G. (2017). *R for data science: Import, tidy, transform, visualize, and model data*. O'Reilly Media. <https://r4ds.had.co.nz/>
- Mulia, S. M., & Murni, T. F. (2023). Analisis faktor-faktor yang mempengaruhi margin pemasaran beras. *Jurnal Sosioagribis*, 4(2), 209–218. <https://journal.uwks.ac.id/index.php/sosioagribis/article/download/3849/1713>
- Rahman, N. (2018). Analisis saluran dan margin pemasaran beras. Diakses dari https://digilibadmin.unismuh.ac.id/upload/1425-Full_Text.p
- Simarmata, E. R., Damanik, H., & Sembiring, N. (2014). Analisis sistem pemasaran gabah/beras (Studi kasus petani padi di Sumatra Utara). *[Nama Jurnal Tidak Disebutkan]*. Diakses dari <https://media.neliti.com/media/publications/43920-ID-analisis-sistem-pemasaran-gabahberas-studi-kasus-petani-padi-di-sumatra-utara.pdf>
- Sudirman, Anshariah, & Rasyid, A. (2021). Analisis rantai pasok beras di Kabupaten Maros. *Jurnal Agribisnis dan Sosial Ekonomi Pertanian Penyuluhan*, 1(1), 21–30. Diakses dari <https://ejournal.polbangtan-gowa.ac.id/index.php/J-Agr-Sosekpenyuluhan/article/download/57/57/114>

Sutopo, S., & Puspitasari, H. (2018). Efisiensi pemasaran gabah dan beras. *Jurnal Ekonomi Pertanian dan Agribisnis*, 2(4), 263–272. Diakses dari <https://jepa.ub.ac.id/index.php/jepa/article/view/527>

Widodo, A. Y., Hidayat, F., & Wijayanto, T. A. (2023). Analisis faktor-faktor yang mempengaruhi margin pemasaran gabah/beras. *Jurnal Ekonomi Pertanian dan Agribisnis*, 7(1), 1–10. Diakses dari <https://jepa.ub.ac.id/index.php/jepa/article/view/1103>