**Data Science****Analisis Computer Price**

**Dibuat Oleh:**

**Samson Nicholas Sitorus (11423015)**

**Dewa Armada Gurusinga (11423041)**

**Yohana Rajagukguk (11423067)**

**INSTITUT TEKNOLOGI DEL**

**MATERI PRAKTIKUM**

Contents

[**Bab I** 4](#_Toc211583176)

[**Pendahuluan** 4](#_Toc211583177)

[1.1 Latar Belakang 4](#_Toc211583178)

[1.2 Rumusan Masalah 4](#_Toc211583179)

[1.3 Tujuan Penelitian 5](#_Toc211583180)

[**BAB II** 7](#_Toc211583181)

[**METODOLOGI PENELITIAN** 7](#_Toc211583182)

[2.1 Data Collection 7](#_Toc211583183)

[2.2 Data Understanding 7](#_Toc211583184)

[Visualisasi eksploratif awal yang telah dibuat: 14](#_Toc211583185)

[2.3 Data Cleaning 18](#_Toc211583186)

[2.4 Visualisasi Data. 21](#_Toc211583187)

[2.5 Statistical Analysis 26](#_Toc211583188)

[1.Uji Parametrik 26](#_Toc211583189)

[2.Uji Non-Parametrik 27](#_Toc211583190)

[**BAB III** 29](#_Toc211583191)

[**HASIL DAN PEMBAHASAN** 29](#_Toc211583192)

[3.1 Hasil Data Collection 29](#_Toc211583193)

[d.Visualisasi Hubungan antara RAM(GB) dan Harga 29](#_Toc211583194)

[3.2 Hasil Data Cleaning 30](#_Toc211583195)

[a. Penanganan Nilai Hilang (Handling Missing Values) 30](#_Toc211583196)

[b. Penanganan Outlier (Handling Outliers) 30](#_Toc211583197)

[c. Encoding Variabel Kategorikal 31](#_Toc211583198)

[d. Seleksi Fitur (Feature Selection) 31](#_Toc211583199)

[3.3 Hasil Visualisasi data 32](#_Toc211583200)

[3.4 Hasil Statistical Analysis 35](#_Toc211583201)

[**BAB IV** 37](#_Toc211583202)

[**KESIMPULAN DAN SARAN** 37](#_Toc211583203)

[4.1 Kesimpulan 37](#_Toc211583204)

[4.2 Saran 38](#_Toc211583205)

[2. Saran untuk Pelaku Industri (Produsen dan Penjual) 38](#_Toc211583206)

[3. Saran untuk Konsumen 39](#_Toc211583207)

[4.3 Penutup 39](#_Toc211583208)

[Referensi 39](#_Toc211583209)

# **Bab I**

# **Pendahuluan**

## 1.1 Latar Belakang

Perkembangan industri komputer telah menjadi salah satu sektor paling dinamis dan kompetitif di era digital saat ini. Setiap tahun, berbagai produsen bersaing menghadirkan inovasi baru dengan spesifikasi yang semakin canggih dan harga yang bervariasi. Fenomena ini menciptakan pasar yang kompleks, di mana harga perangkat komputer tidak hanya ditentukan oleh satu faktor, tetapi merupakan hasil dari interaksi berbagai elemen seperti spesifikasi teknis, merek, serta tren pasar.

Bagi konsumen, memahami faktor-faktor yang memengaruhi harga komputer sangat penting agar dapat mengambil keputusan pembelian yang efisien dan sesuai kebutuhan. Di sisi lain, bagi produsen dan penjual, analisis terhadap data harga komputer dapat memberikan wawasan berharga untuk menentukan strategi penetapan harga, segmentasi pasar, dan pengembangan produk yang lebih kompetitif.

Beberapa faktor utama yang sering diasosiasikan dengan penentuan harga komputer antara lain kapasitas RAM, jenis dan kecepatan prosesor (CPU), merek, tahun rilis, serta sistem operasi yang digunakan. Namun, hubungan statistik dan tingkat pengaruh dari masing-masing faktor terhadap harga belum banyak dijelaskan secara kuantitatif. Oleh karena itu, diperlukan analisis berbasis data untuk mengidentifikasi pola, korelasi, dan faktor-faktor dominan yang memengaruhi harga komputer.

Melalui proyek ini, dilakukan analisis terhadap dataset harga komputer secara komprehensif dengan pendekatan ilmu data (data science). Hasil analisis diharapkan dapat memberikan pemahaman yang lebih mendalam mengenai hubungan antarvariabel, serta menjadi dasar dalam pengambilan keputusan bagi konsumen dan produsen dalam industri komputer.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, penelitian ini difokuskan untuk menjawab beberapa pertanyaan penelitian berikut secara sistematis:

1. Faktor-faktor apa saja yang memiliki korelasi signifikan terhadap harga computer ?
2. Bagaimana kekuatan dan arah hubungan antara kapasitas RAM dengan harga komputer ?
3. Apakah terdapat hubungan yang signifikan antara tahun rilis dengan harga komputer ?

## 1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis secara komprehensif faktor-faktor berbasis data yang mempengaruhi penetapan harga di pasar komputer, dengan fokus pada hubungan antara spesifikasi teknis, atribut produk, dan tren pasar. Untuk mencapai tujuan tersebut, penelitian ini akan mengacu pada beberapa tujuan spesifik sebagai berikut:

1. Menganalisis Pengaruh Spesifikasi Teknis dan Atribut Kategorikal: Menganalisis bagaimana berbagai komponen perangkat keras dan atribut non-teknis berkontribusi terhadap harga akhir produk. Penelitian ini secara spesifik membandingkan dampak dari variabel numerik (seperti kapasitas RAM, jumlah inti CPU) dan variabel kategorikal (seperti merek, sistem operasi).
2. Mengevaluasi Hubungan Statistik dan Strategi Penetapan Harga: Mengidentifikasi dan mengevaluasi hubungan statistik antara fitur-fitur teknis utama dengan harga jual. Tujuannya adalah untuk memahami bagaimana peningkatan spesifikasi teknis diterjemahkan ke dalam nilai (harga) bagi konsumen dan seberapa signifikan hubungan tersebut secara statistik.
3. Mengkaji Pola Harga di Berbagai Segmen Pasar: Mengkaji bagaimana strategi penetapan harga diterapkan di pasar, dengan melihat pola harga, preferensi spesifikasi, dan tren berdasarkan kategori produk (misalnya, Desktop vs. Laptop), merek, dan tahun rilis. Dengan demikian, penelitian ini diharapkan bisa memberikan pemahaman yang lebih dalam mengenai bagaimana harga disesuaikan untuk berbagai segmen di pasar komputer.

**1.4 Manfaat Proyek**

Proyek analisis ini diharapkan dapat memberikan manfaat secara teoritis dan praksi, sebagai berikut:

1. Manfaat Teoretis

Hasil penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan, khususnya dalam bidang analisis data (data analytics) dan ekonomi digital, melalui penerapan metode statistik dan analisis korelasi untuk memahami faktor-faktor yang memengaruhi harga komputer. Penelitian ini juga dapat menjadi referensi bagi penelitian selanjutnya yang berfokus pada analisis harga produk teknologi atau studi pasar berbasis data.

1. Manfaat Praktis
   1. Bagi Konsumen, hasil penelitian ini dapat menjadi acuan dalam memahami faktor-faktor utama yang menentukan harga komputer, sehingga membantu dalam pengambilan keputusan pembelian yang lebih rasional dan sesuai dengan kebutuhan.
   2. Bagi Produsen dan Penjual, penelitian ini dapat memberikan wawasan strategis mengenai pengaruh spesifikasi teknis, merek, dan tahun rilis terhadap harga, yang dapat digunakan untuk menyusun strategi penetapan harga, segmentasi pasar, serta pengembangan produk yang kompetitif.
   3. Bagi Akademisi, penelitian ini dapat menjadi bahan pembelajaran dalam penerapan analisis data untuk memahami hubungan antarvariabel dalam konteks ekonomi digital dan industri teknologi.

# **BAB II**

# **METODOLOGI PENELITIAN**

Pengumpulan data merupakan hal penting dalam proyek ini karena kualitas dan relevansi data sangat mempengaruhi akurasi serta keandalan hasil yang diperoleh. Dalam proyek ini, data dikumpulkan untuk memahami berbagai faktor yang mempengaruhi harga komputer atau laptop di pasar, sehingga bisa memberikan informasi yang bermanfaat bagi konsumen maupun produsen perangkat komputer.

## 2.1 Data Collection

Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu data yang telah tersedia dan dikumpulkan oleh pihak lain untuk tujuan tertentu, kemudian digunakan kembali untuk analisis dalam penelitian ini. Sumber data diperoleh dari platform data publik Kaggle, dengan nama dataset “Computer Prices All”, yang dapat diakses melalui tautan berikut:<https://www.kaggle.com/datasets/paperxd/all-computer-prices>  
  
Dataset ini memuat informasi komprehensif mengenai berbagai jenis komputer beserta spesifikasi dan harga jualnya. Beberapa variabel utama yang terdapat dalam dataset meliputi harga komputer (price), kapasitas RAM (ram\_gb), jenis dan kecepatan prosesor (CPU), merek (brand), tahun rilis (release\_year), serta sistem operasi (operating\_system). Variabel-variabel tersebut menjadi dasar dalam analisis untuk mengidentifikasi faktor-faktor yang memiliki pengaruh terhadap harga komputer.

Proses pengumpulan data dilakukan dengan mengunduh berkas dataset dalam format CSV (Comma Separated Values) secara langsung dari repository Kaggle. Format CSV dipilih karena bersifat terbuka, mudah diakses, dan kompatibel dengan berbagai perangkat lunak analisis data seperti Tableau, Microsoft Excel, dan Python (Pandas).

Setelah data diunduh, dilakukan pemeriksaan awal (data inspection) untuk memastikan integritas dan konsistensi data, termasuk pengecekan terhadap nilai hilang (missing values), duplikasi data, serta format variabel agar siap digunakan dalam tahap analisis statistik dan visualisasi.

## 2.2 Data Understanding

Dataset yang digunakan berasal dari Kaggle dengan judul “Computer Prices All”. Dataset ini berisi  
**Jumlah baris:** 100000 dan **Jumlah kolom:** 33 fitur.

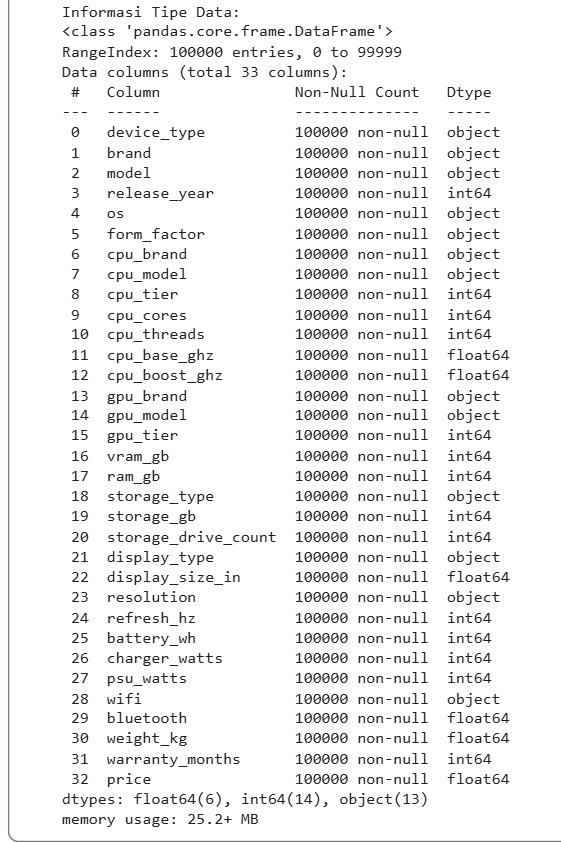
1. Struktur data dan bentuk data

Dataset terdiri dari 33 kolom (fitur) yang mencakup aspek teknis, merek, dan karakteristik produk. Berikut adalah daftar lengkap kolom beserta keterangan dan tipe datanya:

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Nama Kolom | Keterangan | Tipe Data |
| 1 | device\_type | Jenis perangkat (Contoh: Desktop, Laptop) | object |
| 2 | brand | Merek produsen (Contoh: Samsung, Acer, Lenovo) | object |
| 3 | model | Model spesifik dari komputer | object |
| 4 | release\_year | Tahun rilis produk | int64 |
| 5 | os | Sistem operasi (Contoh: Windows, macOS, Linux) | object |
| 6 | form\_factor | Bentuk fisik atau ukuran (Contoh: ATX, Gaming, Mini-ITX) | object |
| 7 | cpu\_brand | Merek CPU (Contoh: Intel, AMD) | object |
| 8 | cpu\_model | Model spesifik dari CPU | object |
| 9 | cpu\_tier | Tingkatan/level performa CPU | int64 |
| 10 | cpu\_cores | Jumlah inti (core) pada CPU | int64 |
| 11 | cpu\_threads | Jumlah alur (thread) pada CPU | int64 |
| 12 | cpu\_base\_ghz | Kecepatan dasar CPU dalam GHz | float64 |
| 13 | cpu\_boost\_ghz | Kecepatan maksimal (boost) CPU dalam GHz | float64 |
| 14 | gpu\_brand | Merek GPU (Contoh: NVIDIA, AMD) | object |
| 15 | gpu\_model | Model spesifik dari GPU | object |
| 16 | gpu\_tier | Tingkatan/level performa GPU | int64 |
| 17 | vram\_gb | Kapasitas VRAM GPU dalam Gigabyte | int64 |
| 18 | ram\_gb | Kapasitas RAM dalam Gigabyte | int64 |
| 19 | storage\_type | Jenis penyimpanan (Contoh: SSD, NVMe, Hybrid) | object |
| 20 | storage\_gb | Kapasitas penyimpanan dalam Gigabyte | int64 |
| 21 | storage\_drive\_count | Jumlah drive penyimpanan | int64 |
| 22 | display\_type | Jenis panel layar (Contoh: LED, OLED, Mini-LED) | object |
| 23 | display\_size\_in | Ukuran layar dalam inci | float64 |
| 24 | resolution | Resolusi layar (Contoh: 1920x1080) | object |
| 25 | refresh\_hz | Refresh rate layar dalam Hertz | int64 |
| 26 | battery\_wh | Kapasitas baterai dalam Watt-hour (untuk laptop) | int64 |
| 27 | charger\_watts | Daya pengisi daya dalam Watt (untuk laptop) | int64 |
| 28 | psu\_watts | Daya Power Supply Unit dalam Watt (untuk desktop) | int64 |
| 29 | wifi | Versi standar Wi-Fi yang didukung | object |
| 30 | bluetooth | Versi Bluetooth yang didukung | float64 |
| 31 | weight\_kg | Berat perangkat dalam Kilogram | float64 |
| 32 | warranty\_months | Durasi garansi dalam bulan | int64 |
| 33 | price | Harga perangkat | float64 |

Dan pengecekan :

  
**output:**



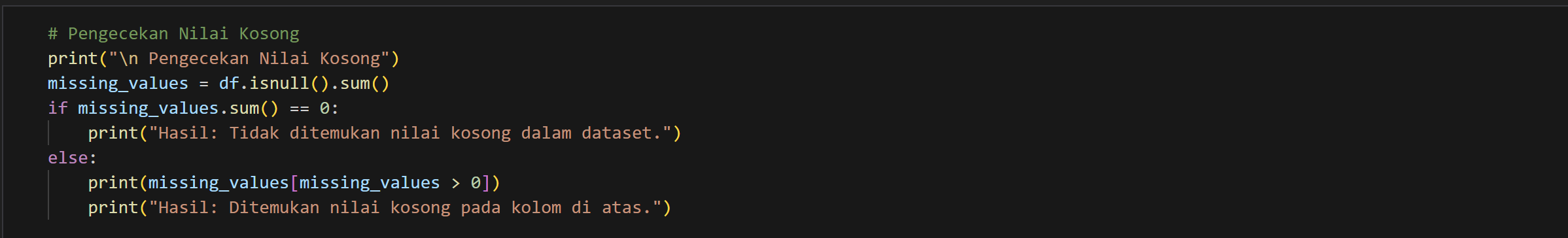
1. Pemeriksaan Awal Kualitas Data
2. Tipe Data

Tipe data yang terdapat dalam dataset Computer Prices All yaitu categorical dan numerik.  
**Tipe data Objektif** terdiri dari device\_type, brand, model, os, form\_factor, cpu\_brand, cpu\_model, gpu\_brand, gpu\_model, storage\_type, display\_type, resolution, wifi, bluetooth.

**Tipe data Numeric** terdiri dari release\_year, cpu\_tier, cpu\_cores, cpu\_threads, gpu\_tier, vram\_gb, ram\_gb, storage\_gb, storage\_drive\_count, refresh\_hz, battery\_wh, charger\_watts, psu\_watts, warranty\_months.

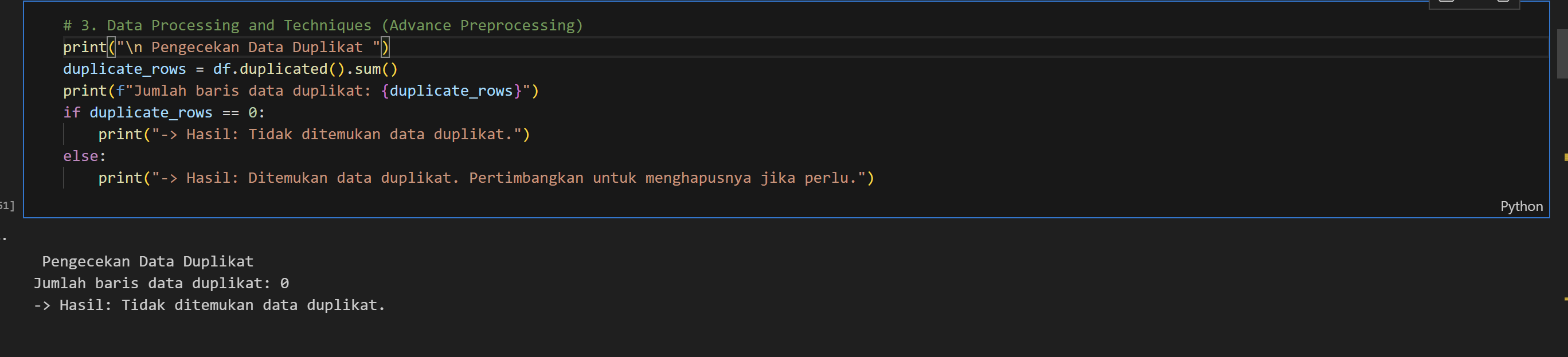
1. Pemeriksaan Missing Value

Dari hasil pengecekan missing value terdapat data yang bersih, dan tidak memiliki nilai yang null.

  
output:  

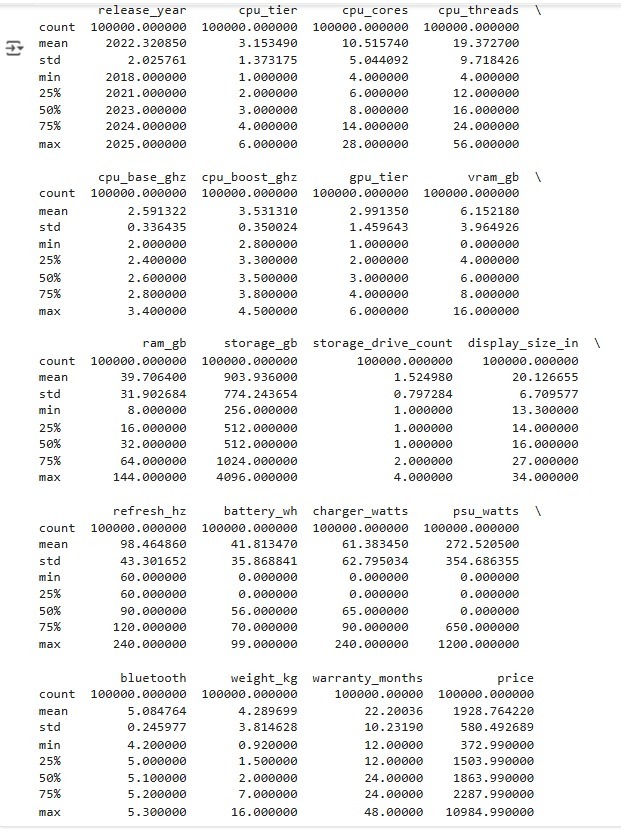

1. Pemeriksaan Duplikasi

Setelah dilakukan pengecekan nilai duplicate dihasilkan bahwa tidak ada data yang duplicate.



1. Statistik Deskriptif Awal

Statistik deskriptif digunakan untuk memahami bagaimana data numerik utama dalam dataset komputer tersebar dan memiliki pola atau tren tertentu.



Berikut adalah analisis mendalam untuk setiap kategori spesifikasi:

Informasi Umum & Rilis

**release\_year**  
Laptop tertua dalam dataset ini berasal dari tahun 2018 (min) dan yang terbaru dari tahun 2025 (max).Setengah dari laptop dirilis pada tahun 2023 atau lebih baru (50% atau median), yang menunjukkan bahwa dataset ini sangat condong ke model-model baru.

**CPU (Processor)**

Jumlah core berkisar dari 2 hingga 28, mencakup spektrum dari laptop entry-level hingga workstation kelas atas.Median (50%) adalah 8 core dan 16 thread, yang merupakan spesifikasi umum untuk laptop modern kelas menengah ke atas saat ini.

**cpu\_base\_ghz & cpu\_boost\_ghz**

Kecepatan clock CPU berada dalam rentang yang wajar, dengan rata-rata kecepatan dasar sekitar 2.6 GHz dan kecepatan boost sekitar 3.5 GHz.

**GPU (Graphics)**

vram\_gb (Video RAM):Rentangnya dari 0 GB hingga 16 GB.Nilai min dan 25% yang 0 GB adalah poin penting: ini menandakan setidaknya 25% dari perangkat di dataset ini menggunakan grafis terintegrasi (integrated graphics) dan tidak memiliki VRAM khusus.Median (50%) adalah 6 GB, menunjukkan bahwa banyak perangkat di dataset ini memiliki kartu grafis diskrit (dedicated) yang mumpuni untuk gaming atau tugas profesional.

**Memori & Penyimpanan**

**ram\_gb:**

Kapasitas RAM berkisar dari 8 GB hingga 144 GB.Median RAM adalah 32 GB, yang tergolong sangat tinggi. Ini menyiratkan bahwa dataset ini mungkin lebih banyak berisi laptop gaming atau laptop untuk kebutuhan profesional daripada laptop konsumen biasa.

**storage\_gb:**

Kapasitas penyimpanan (kemungkinan besar SSD) berkisar dari 256 GB hingga 4096 GB (4 TB).Mediannya adalah 1024 GB (1 TB), yang sudah menjadi standar umum untuk laptop kelas menengah ke atas saat ini.

**Layar & Tampilan**

**display\_size\_in:**Ukuran layar bervariasi dari 13.3 inci (ultrabook) hingga 34 inci. Ukuran 34 inci sangat tidak biasa untuk laptop dan mungkin merupakan data outlier atau data dari PC All-in-One.Ukuran median adalah 16 inci, ukuran yang populer saat ini.

**refresh\_hz:**Refresh rate layar berkisar dari 60 Hz (standar) hingga 240 Hz (gaming high-end).Median 90 Hz menunjukkan adanya campuran yang baik antara layar standar dan layar dengan refresh rate tinggi.

**Daya & Baterai**

**battery\_wh (Watt-hour):**

Kapasitas baterai berkisar dari 0 hingga 99 Wh. Nilai min 0 Wh mungkin menunjukkan perangkat tersebut adalah PC desktop atau ada data yang salah/hilang yang diisi dengan angka 0.

**charger\_watts & psu\_watts:**

Sama seperti baterai, nilai min 0 untuk watt charger dan PSU (Power Supply Unit) perlu diselidiki. Ini bisa berarti PC desktop (yang menggunakan PSU, bukan charger laptop) atau data yang hilang.

**weight\_kg:**

Berat perangkat berkisar dari 0.92 kg (sangat ringan) hingga 16 kg. Berat 16 kg sangat ekstrem untuk sebuah laptop dan kemungkinan besar adalah kesalahan input data (outlier) atau perangkat "desktop replacement" yang sangat besar.Median beratnya adalah 2 kg, yang merupakan berat yang wajar untuk laptop mainstream.

**warranty\_months:**Garansi yang ditawarkan adalah antara 12 bulan (1 tahun) hingga 48 bulan (4 tahun). Garansi 2 tahun (50% = 24 bulan) adalah yang paling umum.

**Harga (price)**

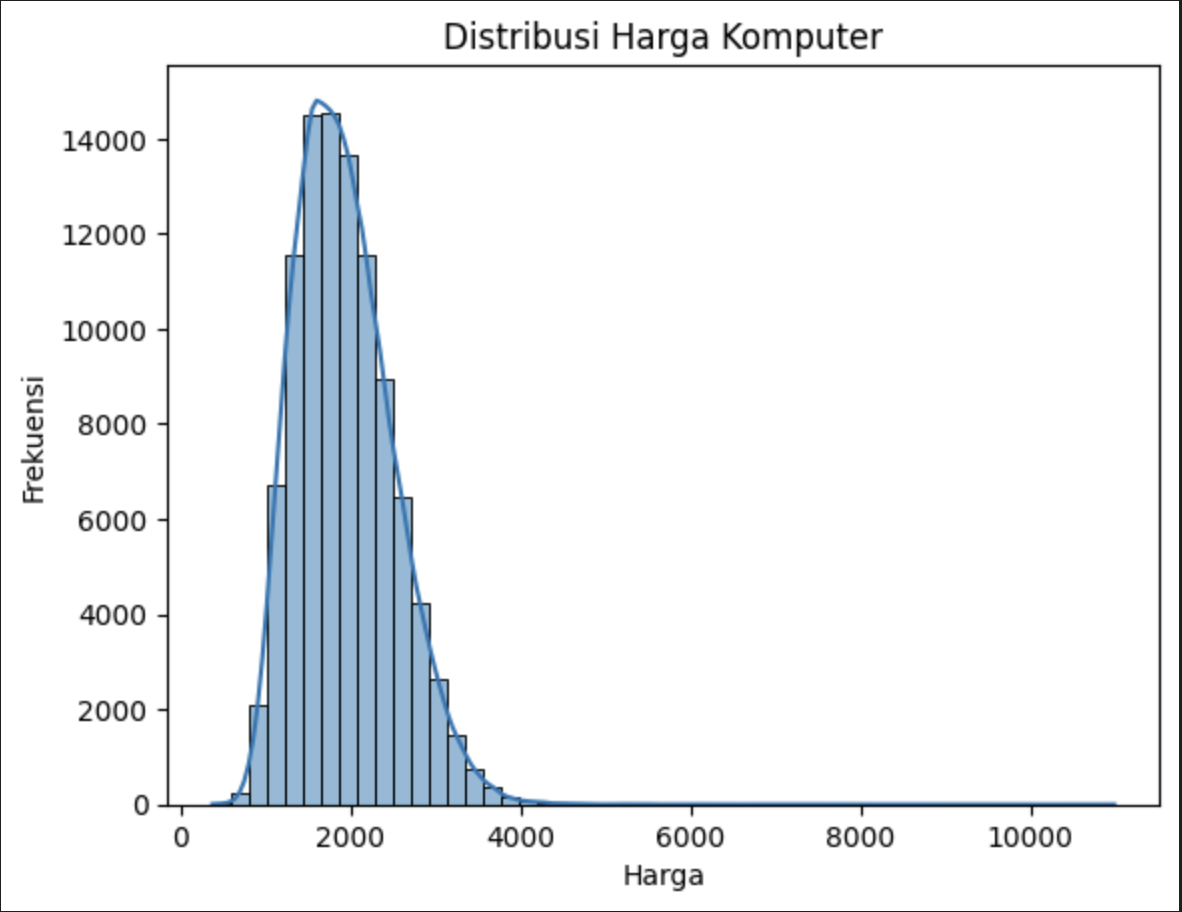
**price (asumsi dalam USD atau mata uang sejenis):**

Harganya sangat bervariasi, dari yang termurah sekitar $150 hingga yang termahal hampir $11,000.Harga rata-rata (mean) adalah $1928, sedangkan harga tengah (50%/median) adalah $1863.

Fakta bahwa rata-rata sedikit lebih tinggi dari median, bersama dengan max yang sangat tinggi, menunjukkan bahwa distribusi harga cenderung miring ke kanan (right-skewed). Artinya, sebagian besar laptop berada di rentang harga yang lebih rendah hingga menengah, tetapi ada beberapa model yang sangat mahal yang "menarik" nilai rata-rata ke atas.

### Visualisasi eksploratif awal yang telah dibuat:

1. Distribusi Harga Komputer

  
Visualisasi yang digunakan adalah histogram dengan kurva Kernel Density Estimate (KDE) yang ditumpangkan. Pilihan ini sangat tepat untuk tujuan menganalisis data harga komputer karena beberapa alasan utama adalah untuk memahami bagaimana distribusi atau sebaran dari satu variabel numerik kontinu, yaitu harga. Histogram secara efektif mengelompokkan harga ke dalam beberapa rentang (disebut bins) dan menunjukkan berapa banyak komputer (frekuensi) yang jatuh dalam setiap rentang tersebut.  
 Grafik ini memberikan beberapa wawasan penting mengenai pasar komputer yang  
**Puncak Permintaan Berada di Kisaran Harga $1.500 - $2.000**

Puncak tertinggi dari histogram (modus) berada di sekitar angka 2.000. Ini menunjukkan bahwa kelompok harga yang paling umum atau paling banyak produknya dalam dataset ini adalah di rentang $1.500 hingga $2.000. Ini kemungkinan besar adalah segmen pasar mainstream atau kelas menengah ke atas.

**Distribusi Condong ke Kanan (Right-Skewed Distribution)**

Ini adalah insight yang paling menonjol. Grafik ini tidak simetris. Ekornya jauh lebih panjang ke sisi kanan (harga yang lebih tinggi).Artinya: Sebagian besar komputer terkonsentrasi di rentang harga yang lebih rendah hingga menengah (di bawah $3.000). Namun, ada sejumlah kecil komputer yang harganya jauh lebih mahal (di atas $4.000, $6.000, bahkan $10.000). Komputer-komputer mahal ini adalah model high-end, seperti laptop gaming kelas atas, workstation untuk profesional, atau edisi khusus yang jumlahnya jauh lebih sedikit dibandingkan model standar.

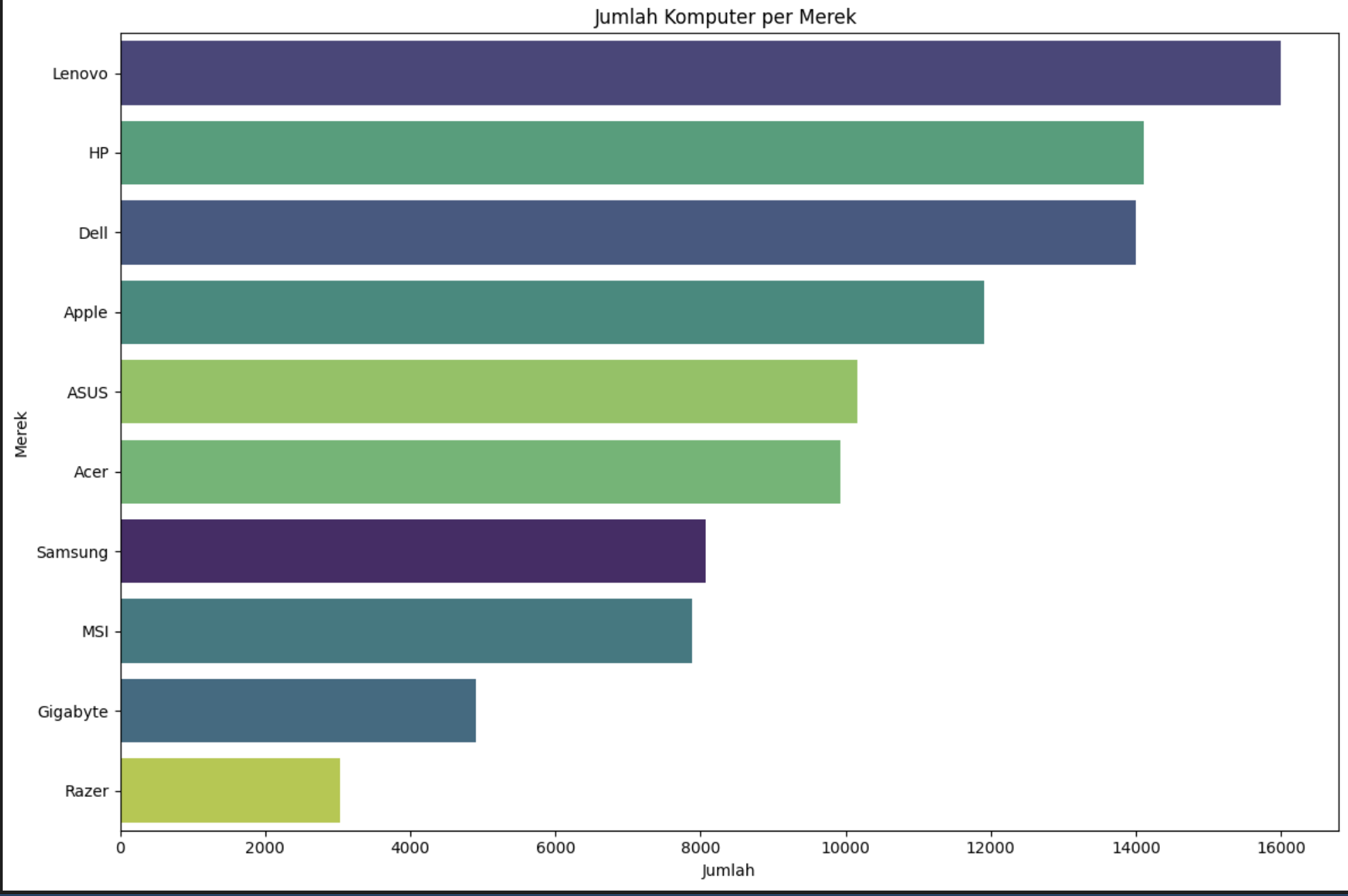
**Konfirmasi dari Statistik Deskriptif**

Pola right-skewed ini konsisten dengan analisis statistik sebelumnya, di mana nilai rata-rata (mean) harga ($1928) sedikit lebih tinggi dari nilai tengah (median) harga ($1863). Beberapa produk dengan harga yang sangat tinggi tersebut "menarik" nilai rata-rata ke arah kanan.

**Pasar Niche di Segmen Premium**

Batang histogram menjadi sangat pendek setelah melewati angka $4.000. Ini menandakan bahwa pasar untuk komputer dengan harga di atas angka tersebut adalah pasar niche (ceruk pasar). Jumlah unit yang tersedia atau terjual di segmen ini jauh lebih sedikit, tetapi mereka ada dan berkontribusi pada variasi harga keseluruhan.

1. Jumlah Komputer per Merek

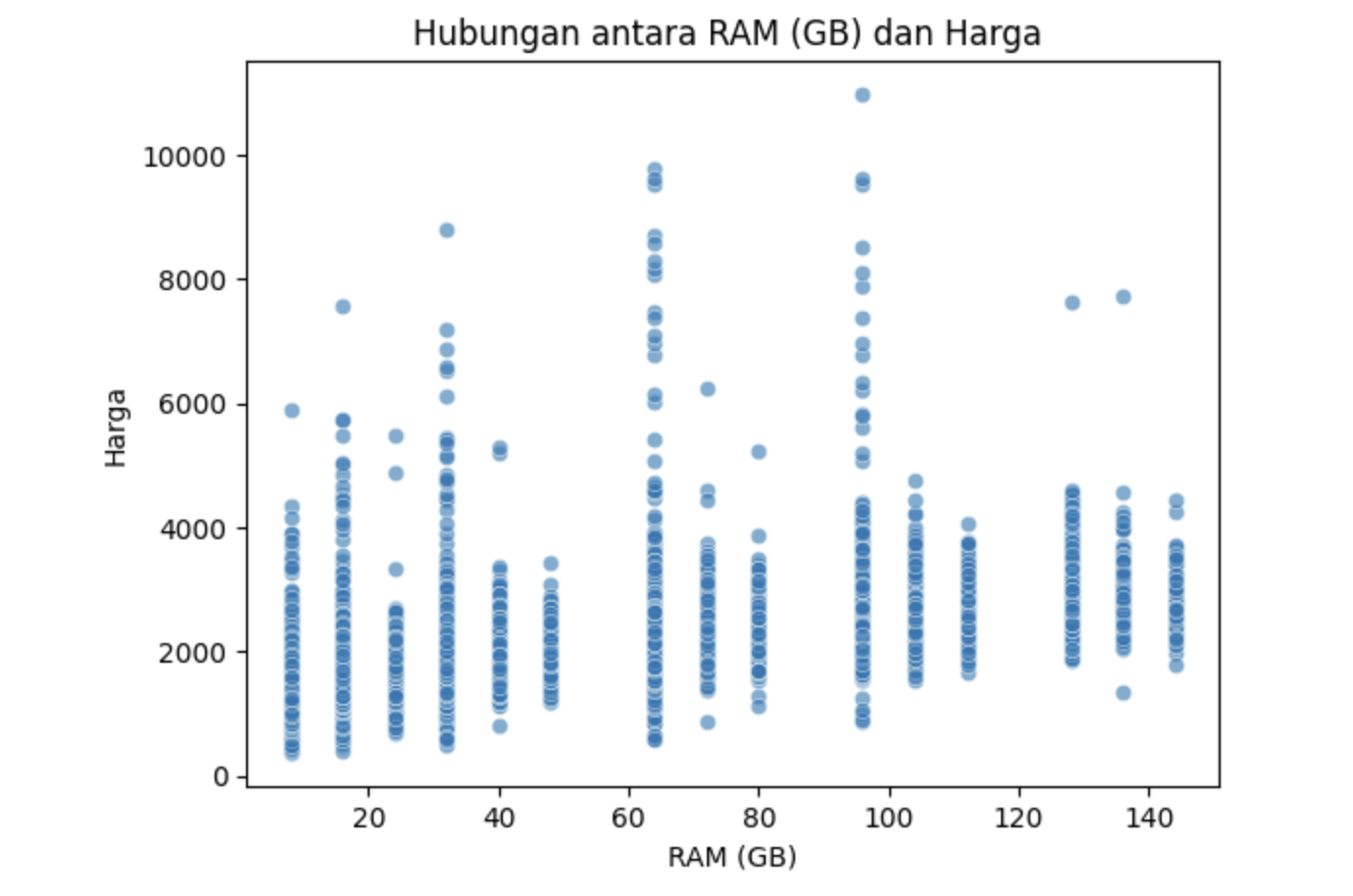


Tujuan utama dari pembuatan visualisasi data adalah untuk mengubah kumpulan data yang besar dan rumit menjadi sebuah cerita yang sederhana dan mudah dipahami. Dengan menyajikan informasi secara visual, seperti melalui diagram batang atau histogram, kita dapat dengan cepat mengidentifikasi pola-pola kunci, seperti merek mana yang paling mendominasi pasar atau di rentang harga mana kebanyakan produk terjual. Visualisasi memungkinkan kita untuk melihat gambaran besar, menemukan anomali atau hal-hal yang menonjol dalam data, dan pada akhirnya mengkomunikasikan temuan tersebut kepada orang lain secara efektif, jauh lebih cepat daripada hanya membaca tabel angka yang panjang.

ni dia insight utamanya secara sederhana:

* Ada 3 Raja Pasar: Lenovo, HP, dan Dell jelas mendominasi. Mereka punya jumlah produk paling banyak.
* Pesaing di Bawahnya: Apple, ASUS, dan Acer adalah pemain besar berikutnya, tapi jumlahnya masih di bawah tiga raja tadi.
* Pemain Niche/Khusus: Merek seperti MSI dan Razer punya porsi pasar yang jauh lebih kecil. Ini wajar karena mereka fokus pada segmen tertentu, seperti laptop gaming.

1. Hubungan antara RAM(GB) dan Harga



Scatter Plot (Diagram Tebar) Ini adalah pilihan terbaik karena tujuannya adalah untuk melihat hubungan atau korelasi antara dua hal yang sama-sama berupa angka, yaitu RAM (GB) dan Harga.Setiap titik di grafik ini mewakili satu komputer. Dengan menebar semua titik, kita bisa melihat apakah ada pola, misalnya: "Apakah jika RAM semakin besar, harganya juga pasti semakin mahal?".

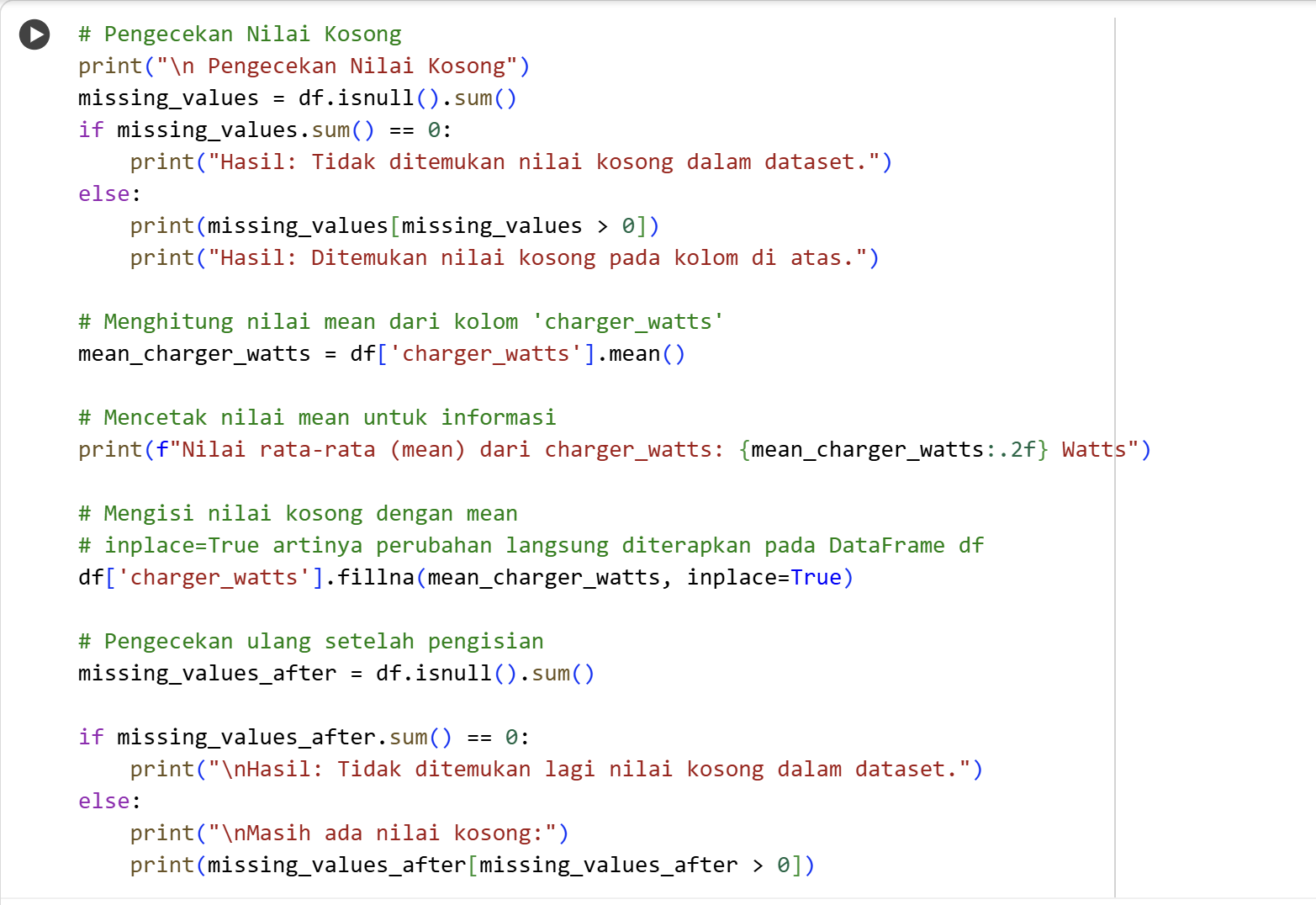
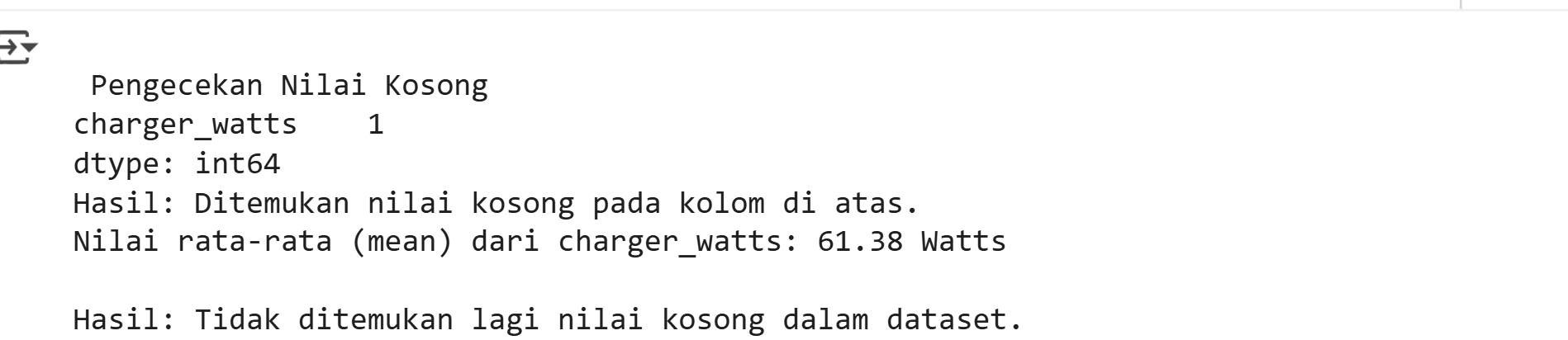
* RAM Bukan Faktor Utama Penentu Harga. Ini adalah kesimpulan paling penting. Anda tidak bisa menebak harga laptop hanya dari besaran RAM-nya.
* Pada Ukuran RAM yang Sama, Rentang Harganya Sangat Lebar.Sebuah laptop dengan RAM 32 GB bisa jadi berharga murah, namun ada juga laptop 32 GB lain yang harganya sangat mahal.
* Komponen Lain Jauh Lebih Berpengaruh. Faktor seperti kartu grafis (GPU), prosesor (CPU), kualitas layar, dan merek memiliki dampak yang jauh lebih besar pada harga akhir dibandingkan sekadar besaran RAM.

## 2.3 Data Cleaning

Tahapan ini meliputi beberapa proses untuk menyiapkan dataset agar dapat digunakan pada tahap analisis dan visualisasi.

1. **Handling Missing Values**

Tahap ini dilakukan untuk mengidentifikasi dan menangani nilai-nilai yang hilang (*missing values*) pada dataset. Pemeriksaan dilakukan terhadap setiap kolom untuk memastikan tidak ada data kosong yang dapat memengaruhi hasil analisis. Berdasarkan hasil pnegecekan yang dilakukan terdapat 1 nilai null pada column charger\_watts dan solusi yang kami terapkan adalah pengisian nilai null nya dengan median dari column charger\_wats yaitu sebesar 61,38.

  
output:  


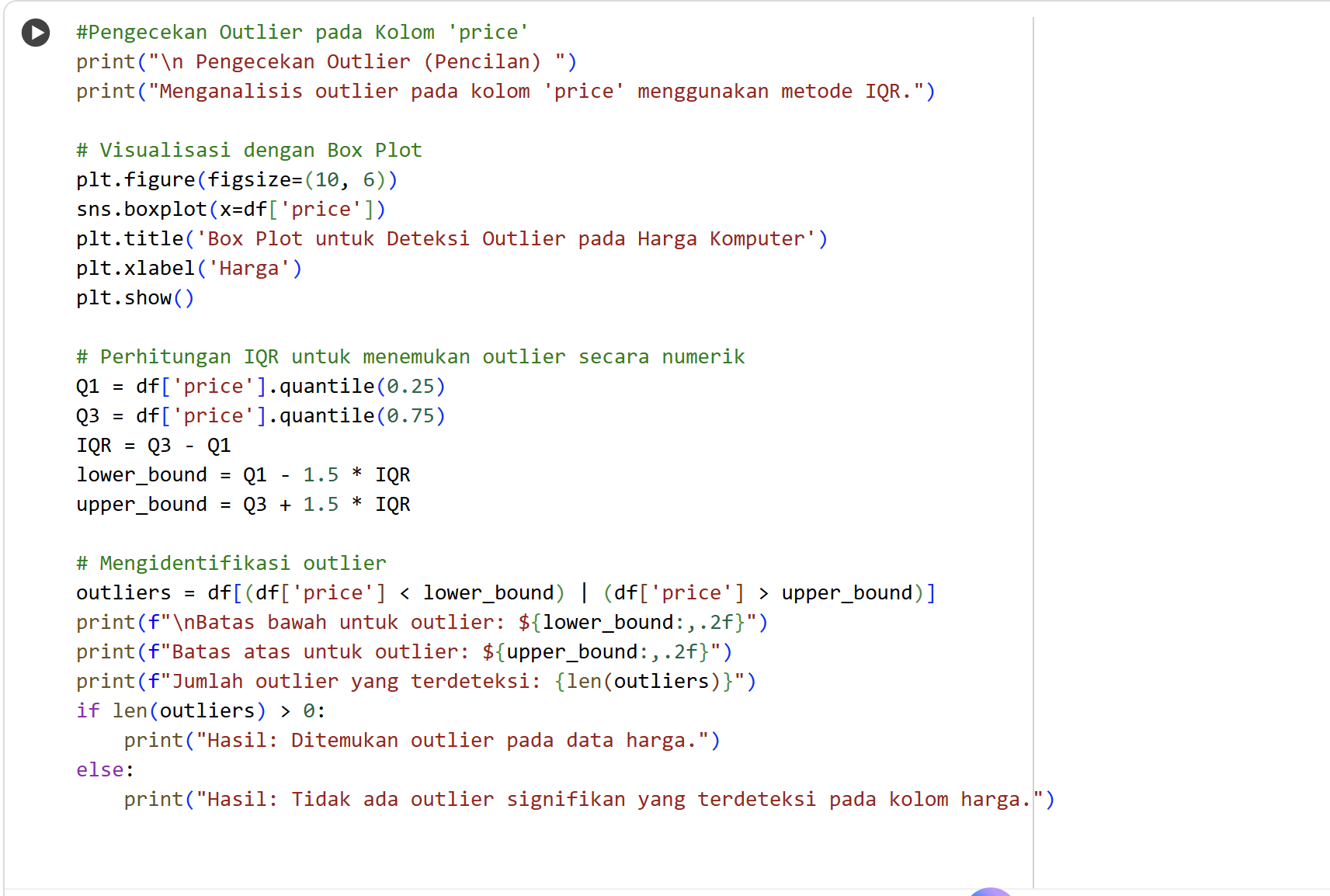
1. **Handling outliers**

Metode yang digunakan (IQR) bekerja dengan cara sederhana:

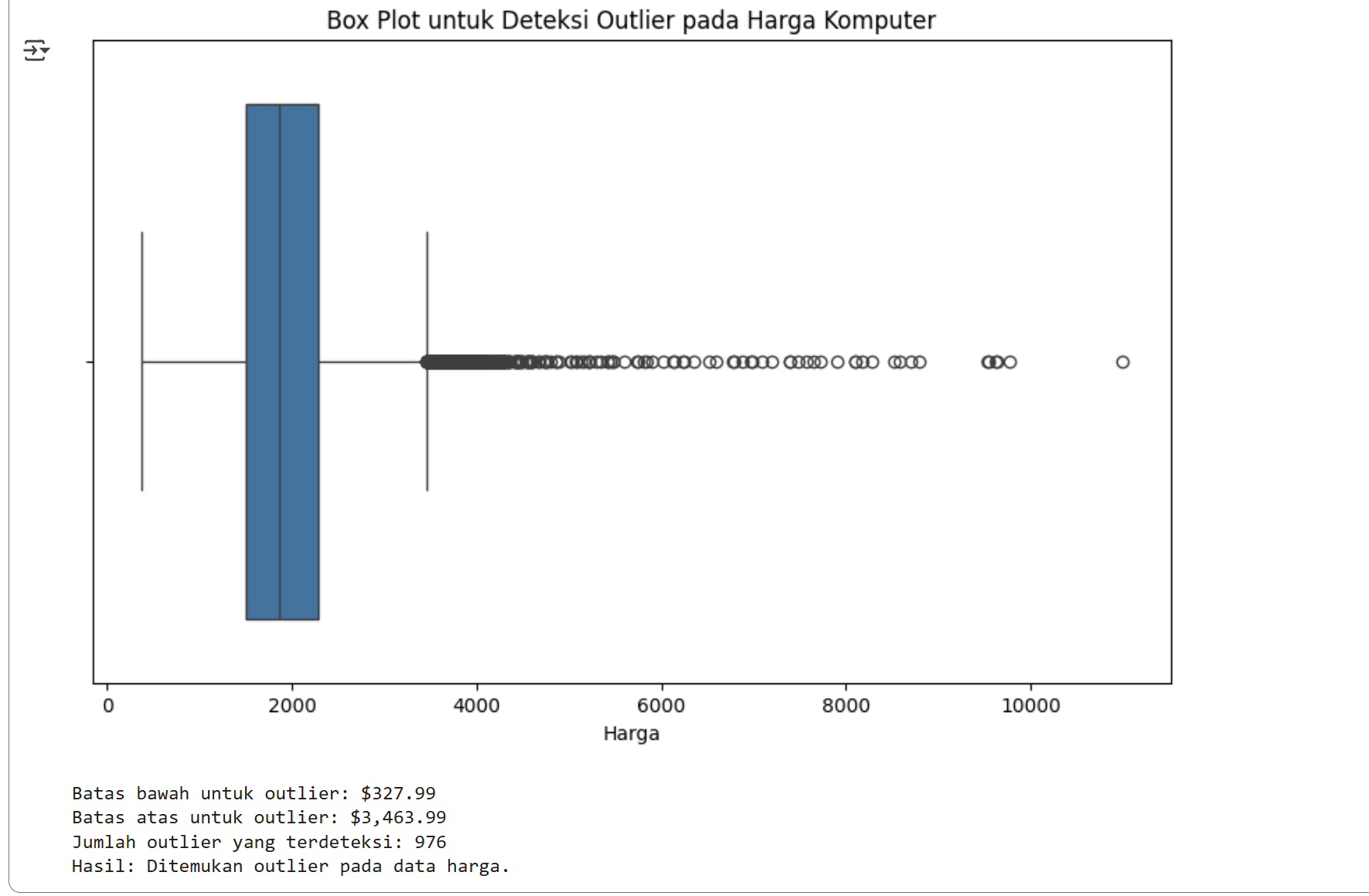
1. Pertama, kita kumpulkan semua harga dan cari tahu "kelompok tengah" atau harga yang paling umum. Ini direpresentasikan oleh kotak biru pada grafik. Anggap saja ini adalah harga "wajar" untuk 50% komputer.

2. Kemudian, kita membuat sebuah "pagar" di atas dan di bawah kelompok tengah ini. Pagar ini menentukan batas harga yang masih kita anggap normal.

3. Setiap harga yang berada di luar pagar tersebut kita sebut sebagai outlier atau "pencilan".

****

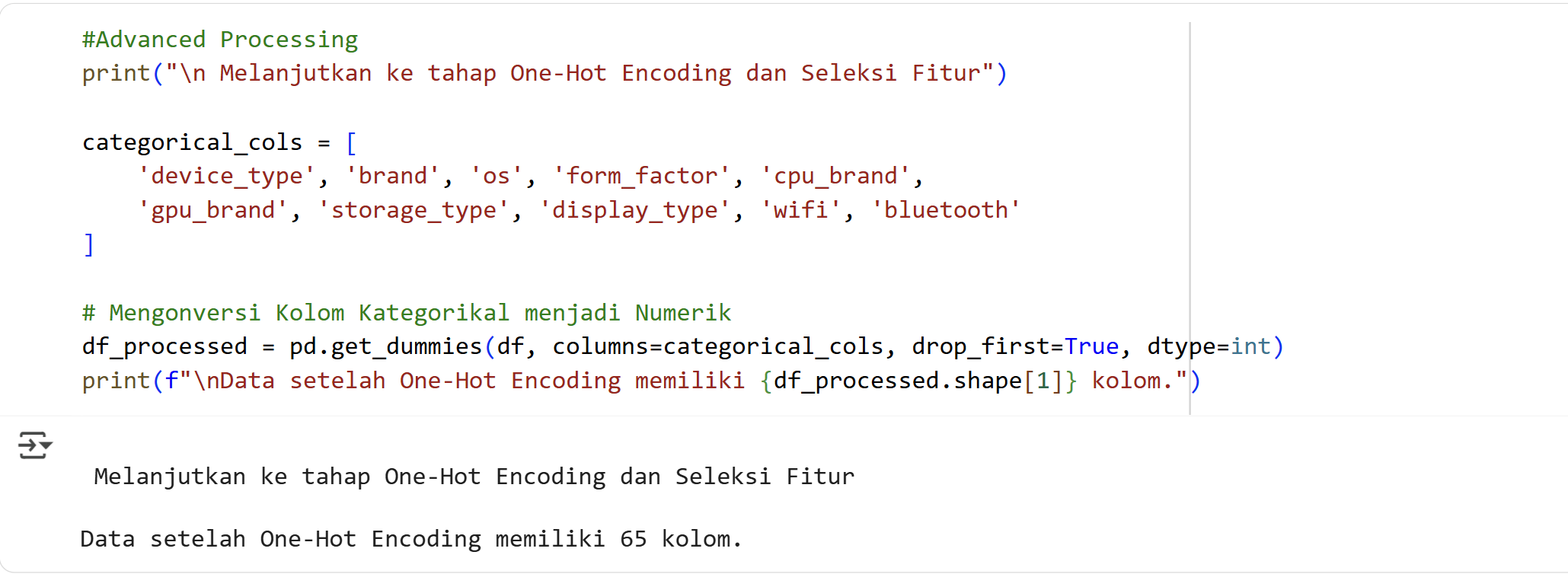
**output**

****

Kotak Biru: Ini adalah area di mana separuh dari harga komputer berada.  
Garis Horizontal: Ini adalah "pagar" atau batas harga wajar.  
Titik-titik di Kanan: Nah, ini dia para outlier! Setiap titik adalah satu komputer yang harganya jauh lebih mahal dari yang dianggap wajar.

Dari visualisasinya didapatkan Batas atas harga yang dianggap wajar adalah $3,463.99. Ditemukan ada 976 komputer yang harganya melebihi batas wajar tersebut.

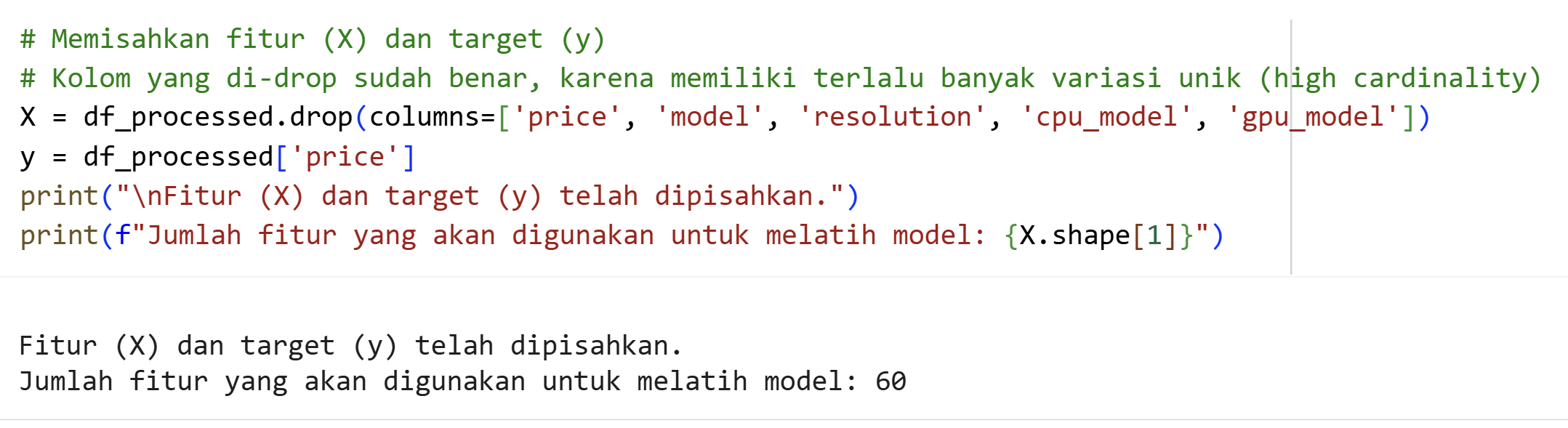
1. **Encoding categorical variables**

****

Sebelumnya: Dataset Anda memiliki beberapa kolom teks (seperti brand, os, wifi, dll).

Sesudahnya: Semua kolom teks tersebut telah diubah menjadi kolom-kolom baru yang berisi angka 0 dan 1. Proses ini membuat jumlah total kolom di dataset akan menjadi 65 kolom.

1. **Feature selection/reduction**

****

Ini adalah tahap di mana kita secara sadar memilih (fitur) mana yang akan kita gunakan dan membuang yang tidak perlu,sehingga lebih mudah menentukan fitur yang akan dipakai dalam pembuatan visualisasi.hasil akhirnya didapatkan 60 column dengan pengurangan 5 setelah dilakukan feature selection.

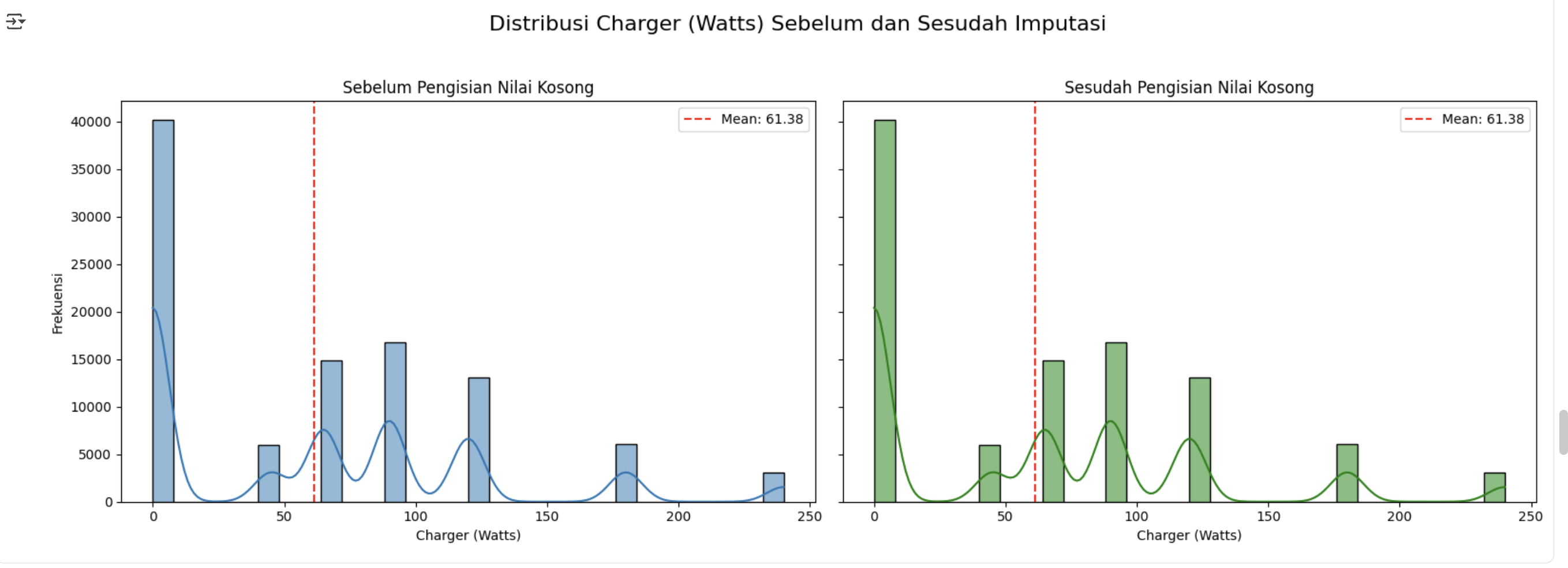
## 2.4 Visualisasi Data.

Pada tahap ini, proyek menerapkan Analisis untuk memberikan pemahaman awal mengenai distribusi dan dinamika data tanpa melakukan prediksi atau inferensi lanjutan.

Beberapa teknik statistik sederhana yang digunakan meliputi mean , serta korelasi antar variabel. Penerapan metode ini membantu mengidentifikasi tren umum, penyebaran data, dan hubungan antar atribut yang relevan dalam dataset katalog tayangan.

Beberapa hasil visualisasi yang dibuat pada tahap ini antara lain:

1. **Distribusi Charger Sebelum dan Sesudah Imputasi**

****

Visualisasi ini menampilkan Grafik yang membandingkan data watt charger sebelum dan sesudah data yang hilang atau kosong diperbaiki.

Insight yang Diperoleh

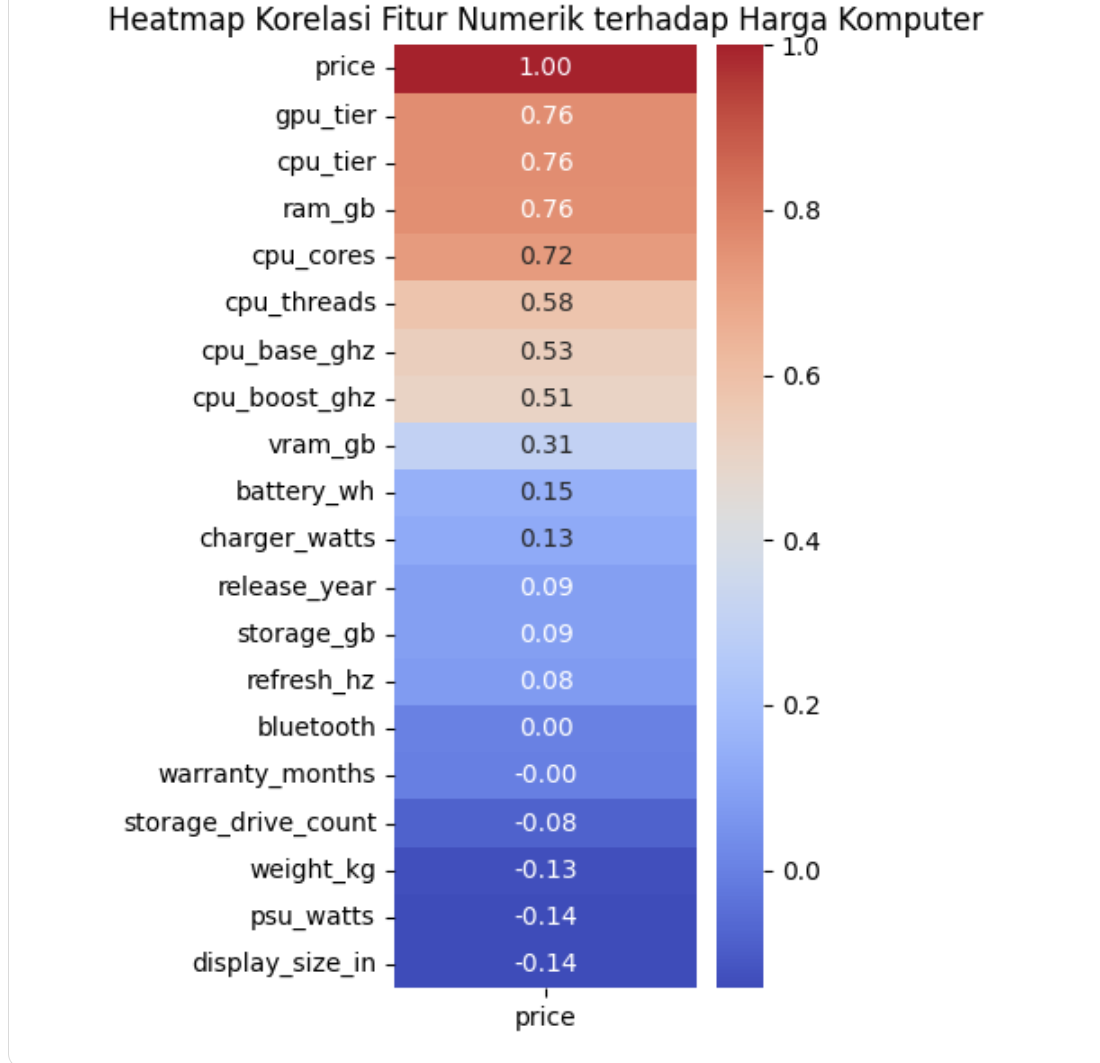
Identifikasi Masalah Data: Insight pertama adalah kita berhasil menemukan masalah serius pada data awal. Ada banyak sekali data watt charger yang tidak tercatat dengan benar (dicatat sebagai 0), yang bisa merusak analisis atau performa model jika tidak ditangani.Metode Perbaikan: Kita bisa menyimpulkan bahwa metode perbaikan yang digunakan adalah Mean Imputation. Artinya, semua nilai null yang kosong tadi diganti dengan nilai rata-rata, yaitu 61.38 watt. Hal ini terlihat jelas dari lonjakan data baru di sekitar angka 60-an pada grafik kanan.

Kelebihan dan Kekurangan:

Kelebihan: Data sekarang lebih realistis karena tidak ada lagi nilai null yang mustahil. Ini membuat dataset lebih siap untuk diproses oleh model machine learning.

Kekurangan: Perbaikan ini membuat distribusi data menjadi lebih akurat karena Terjadi penumpukan data yang wajar di sekitar angka 61.38.

1. **Heatmap Korelasi Fitur Numerik terhadap Harga**

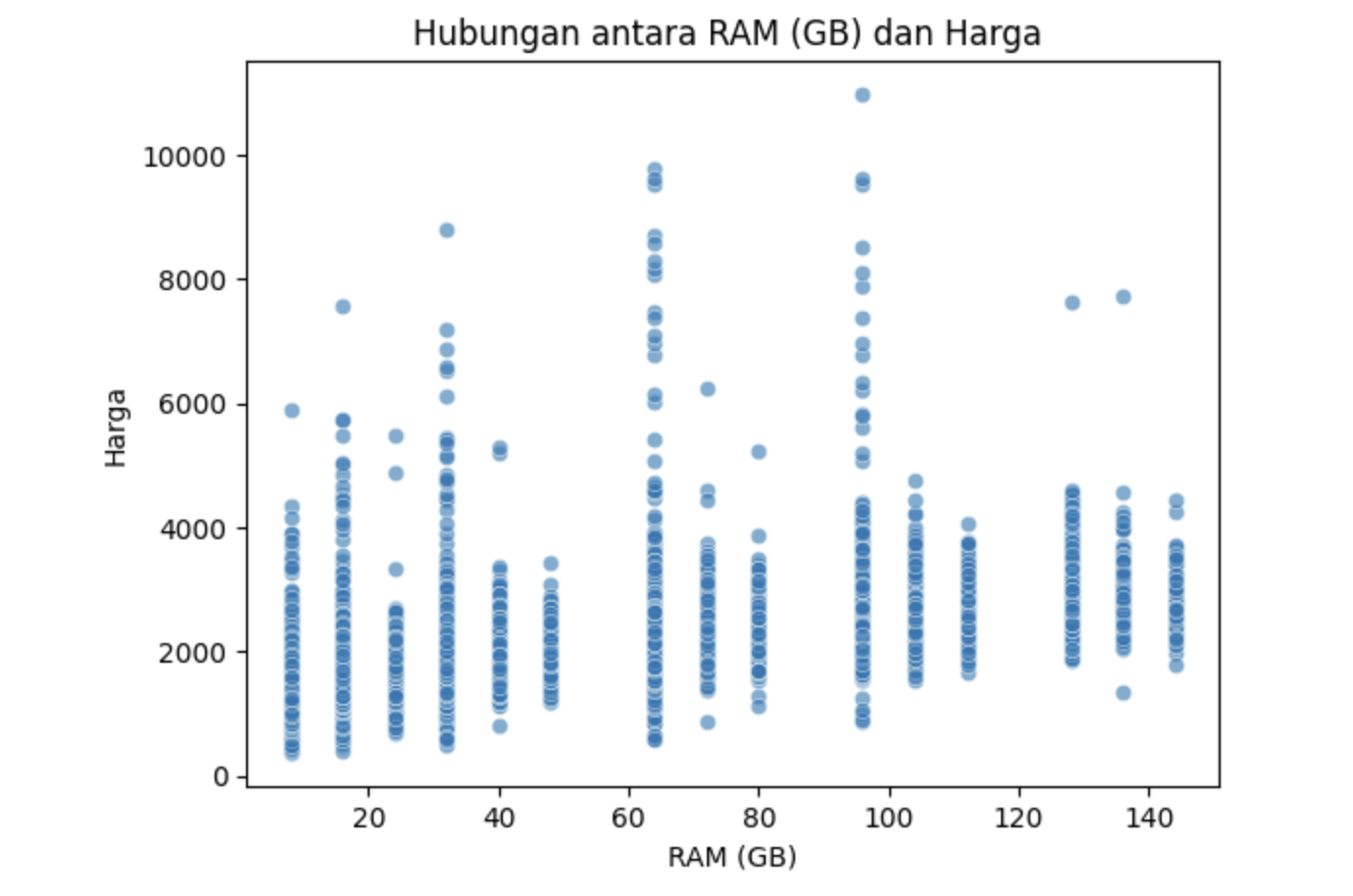
****

Grafik ini disebut Heatmap Korelasi. Tujuannya adalah untuk menunjukkan seberapa kuat hubungan antara berbagai fitur (spesifikasi) dengan harga komputer.

Warna Merah Pekat: Menandakan hubungan positif yang kuat. Artinya, jika nilai fitur ini naik, harga komputer juga cenderung ikut naik.Warna Biru Pekat: Menandakan hubungan negatif yang kuat. Artinya, jika nilai fitur ini naik, harga komputer justru cenderung turun.Warna Pucat/Abu-abu: Menandakan hubungan yang lemah atau hampir tidak ada hubungan sama sekali.Angka (Koefisien Korelasi): Angka di dalam kotak mengukur hubungan ini. Semakin dekat ke 1.00 (sangat positif) atau -1.00 (sangat negatif), semakin kuat hubungannya. Semakin dekat ke 0, semakin lemah hubungannya.atu negara. Tentu, ini dia *insight* utamanya dalam bahasa yang paling sederhana:

* Yang Bikin Mahal: Kartu grafis (GPU), prosesor (CPU), dan RAM adalah tiga hal utama yang membuat harga laptop jadi mahal. Semakin bagus ketiganya, semakin tinggi harganya.
* Yang Gak Ngaruh: Fitur standar seperti Bluetooth sama sekali tidak mempengaruhi harga.
* Yang Unik: Terkadang, laptop yang lebih ringan dan tipis justru harganya lebih mahal.

1. **Hubungan antara RAM dan Harga**

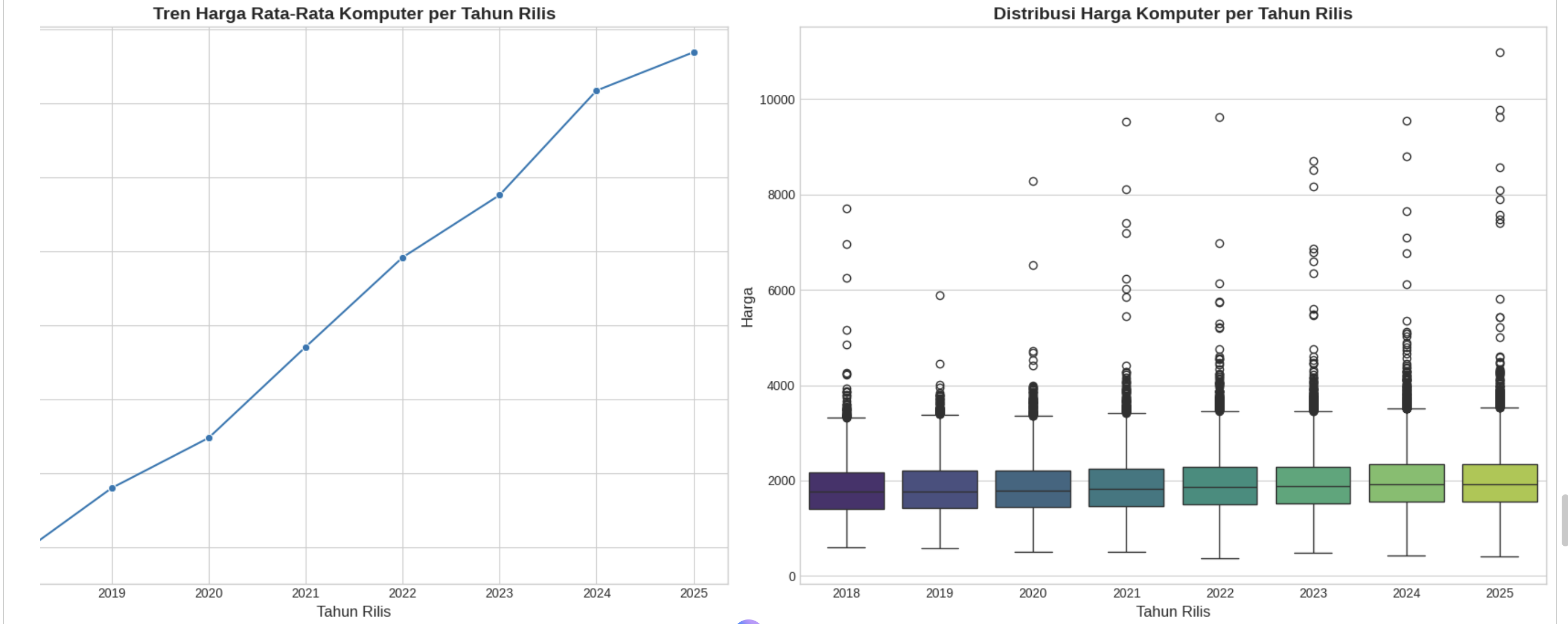


*Scatter Plot* menunjukkan sebaran harga komputer untuk setiap ukuran RAM. Setiap titik biru adalah satu komputer.Jika kita lihat, untuk satu ukuran RAM yang sama, harganya sangat bervariasi. Contohnya:Lihat di garis vertikal untuk **RAM 32 GB** atau **64 GB**. Anda akan menemukan titik-titik yang menyebar dari harga yang relatif murah hingga sangat mahal.Ini membuktikan bahwa dua laptop dengan RAM yang sama persis bisa memiliki harga yang sangat berbeda jauh.

**Insight yang Diperoleh:**

*insight* utamanya adalah: Kapasitas RAM bukan faktor penentu utama harga komputer.Harga sebuah komputer jauh lebih dipengaruhi oleh komponen lain yang lebih mahal, seperti kartu grafis (GPU) dan prosesor (CPU). Itulah mengapa meskipun RAM-nya sama, harga bisa sangat bervariasi tergantung pada seberapa canggih GPU dan CPU yang dipasang.

1. **Tren Harga Rata-rata Komputer per tahun**

****

* 1. Tren Harga Rata-Rata Komputer per Tahun RilisGrafik ini adalah grafik garis (line chart) yang menunjukkan bagaimana rata-rata harga komputer telah berubah dari tahun 2019 hingga 2025.Sumbu X (Horizontal): Menunjukkan "Tahun Rilis" komputer.Sumbu Y (Vertikal): Menunjukkan harga rata-rata.

**Interpretasi:**  
Tren Kenaikan yang Jelas: Terdapat tren kenaikan (uptrend) yang kuat dan konsisten dari tahun ke tahun. Ini berarti, secara rata-rata, harga komputer menjadi lebih mahal setiap tahunnya sejak 2019.

Akselerasi Harga: Kenaikan harga tampaknya semakin curam di tahun-tahun terakhir (misalnya, dari 2023 ke 2025), yang menandakan bahwa laju kenaikan harga rata-rata mungkin meningkat.Kesimpulan dari Grafik Kiri: Rata-rata harga jual komputer secara konsisten meningkat setiap tahunnya dalam periode yang diamati.

* 1. Distribusi Harga Komputer per Tahun Rilis Grafik ini adalah boxplot (diagram kotak garis) yang memberikan gambaran lebih rinci tentang sebaran atau distribusi harga komputer untuk setiap tahun dari 2018 hingga 2025.Sumbu X (Horizontal): Menunjukkan "Tahun Rilis".Sumbu Y (Vertikal): Menunjukkan "Harga".

Setiap Kotak (Box): Mewakili 50% data harga komputer di tahun tersebut (interquartile range/IQR).Garis di dalam kotak adalah median (nilai tengah), bukan rata-rata. Ini adalah harga di mana 50% komputer lebih murah dan 50% lebih mahal.Batas bawah dan atas kotak adalah kuartil pertama (Q1) dan kuartil ketiga (Q3).Titik-titik di luar "kumis" (whiskers): Ini adalah outlier atau pencilan, yaitu data harga yang secara signifikan lebih tinggi atau lebih rendah dari sebagian besar data lainnya di tahun itu.

**Interpretasi:**

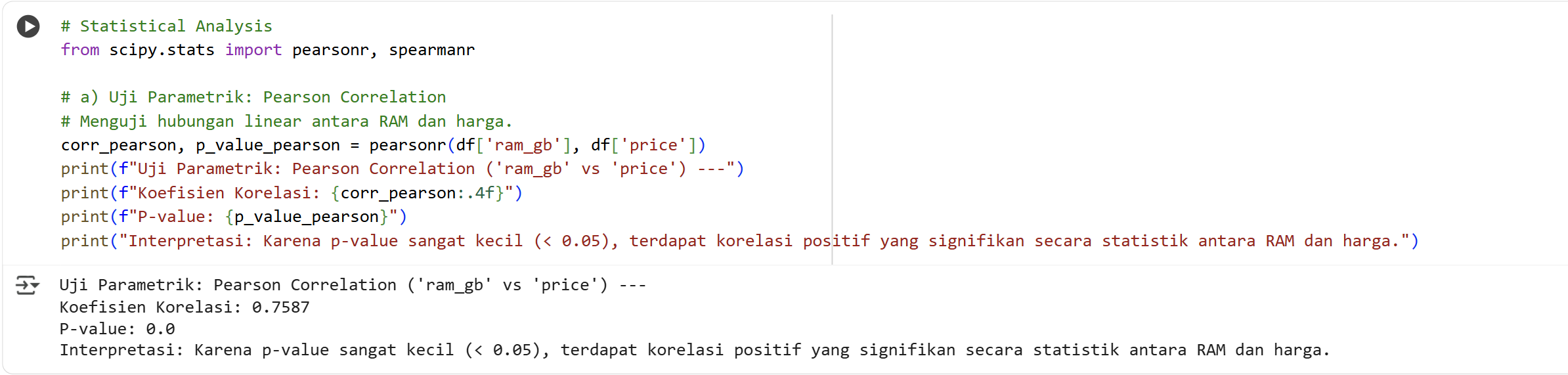
Kenaikan Median: Sama seperti harga rata-rata, harga median (garis tengah di setiap kotak) juga menunjukkan tren kenaikan yang bertahap dari tahun ke tahun. Ini mengonfirmasi tren yang terlihat di grafik pertama.

Banyaknya Outlier Atas: Di setiap tahun, terdapat banyak sekali outlier di sisi atas. Ini menunjukkan bahwa meskipun sebagian besar komputer dijual dalam rentang harga tertentu (di dalam kotak dan kumis), ada sejumlah besar model komputer yang harganya jauh lebih mahal daripada model pada umumnya.

Rentang Harga yang Konsisten: Ukuran kotak (IQR) relatif konsisten dari tahun ke tahun. Ini berarti variasi harga untuk 50% komputer di tengah pasar tidak banyak berubah, meskipun titik harganya secara keseluruhan bergeser ke atas.

## 2.5 Statistical Analysis

## 1.Uji Parametrik



uji statistik yang membuat asumsi-asumsi tertentu tentang distribusi data dari populasi tempat sampel diambil. Jika asumsi ini terpenuhi, uji ini cenderung lebih kuat (lebih mampu mendeteksi efek atau hubungan yang sebenarnya ada).

**Asumsi yang paling umum adalah**

Data berdistribusi normal: Data untuk variabel yang diuji diasumsikan mengikuti kurva lonceng (distribusi normal).

Hubungan Linear: Khusus untuk Korelasi Pearson, diasumsikan ada hubungan garis lurus (linear) antara kedua variabel.

Data Kontinu: Variabel yang diukur bersifat numerik kontinu (misalnya, harga, RAM dalam GB, tinggi badan).

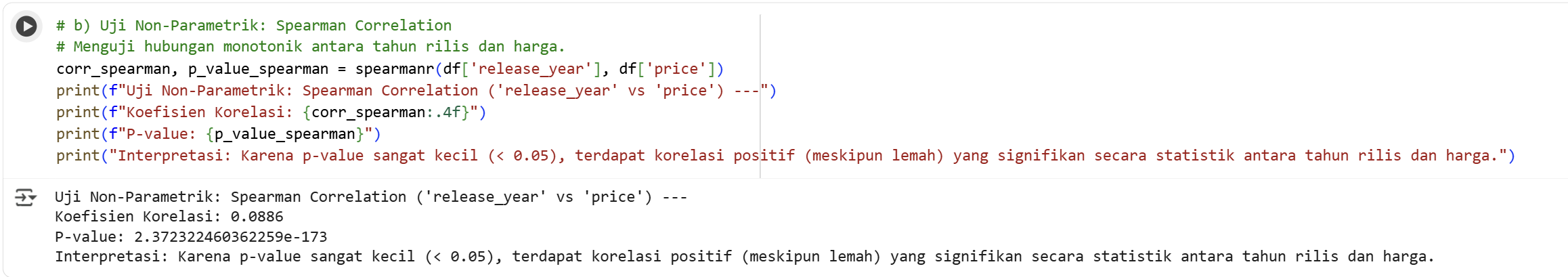
**Hasil Analisis:**  
**Koefisien Korelasi: 0.7587**

Angka ini (mendekati 1.0) menunjukkan adanya hubungan positif yang kuat.Dalam konteks ini: Ketika nilai RAM (ram\_gb) meningkat, maka harga (price) juga cenderung meningkat secara signifikan.

**P-value: 0.0**

P-value yang sangat kecil (kurang dari 0.05) menunjukkan bahwa hasil ini signifikan secara statistik.Dalam konteks ini: Hubungan positif yang ditemukan antara RAM dan harga sangat tidak mungkin terjadi karena kebetulan semata. Kita bisa cukup yakin bahwa hubungan ini benar-benar ada di dalam data.

## 2.Uji Non-Parametrik



Hasilnya menunjukkan ada hubungan positif yang sangat kuat dan sangat signifikan secara statistik antara tahun rilis komputer dan harganya.

**Analisis:**

1. Koefisien Korelasi Spearman: 0.8886

Kekuatan Hubungan: Nilai 0.8886 sangat mendekati 1.0. Ini menandakan adanya hubungan positif yang sangat kuat, bukan lemah.Arah Hubungan: Tanda positif (+) berarti hubungannya searah.

Interpretasi: Seiring dengan bertambahnya tahun rilis (release\_year), harga (price) komputer secara konsisten cenderung meningkat. Ini adalah hubungan monotonik, yang berarti variabelnya bergerak ke arah yang sama, meskipun tidak harus dalam bentuk garis lurus yang sempurna.

2. P-value: 2.372e-173

Signifikansi Statistik: P-value ini adalah angka yang sangat kecil, bisa dibilang mendekati nol. Nilai ini jauh di bawah ambang batas umum (seperti 0.05, 0.01, atau bahkan 0.001).

Interpretasi: Ini berarti hubungan kuat yang terdeteksi antara tahun rilis dan harga sangat signifikan secara statistik. Hampir tidak ada kemungkinan bahwa hubungan ini terjadi hanya karena kebetulan. Kita bisa sangat yakin bahwa pola ini nyata dalam data Anda.

**Uji Non-Parametrik (Spearman)**

Data Tidak Berdistribusi Normal: Anda tidak perlu berasumsi bahwa data release\_year atau price mengikuti distribusi normal.

Hubungan Tidak Harus Linear: Spearman mengukur hubungan monotonik. Artinya, selama harga cenderung naik seiring naiknya tahun rilis (bahkan jika kenaikannya melambat atau semakin cepat), Spearman akan menangkapnya. Ini lebih fleksibel daripada korelasi Pearson yang hanya mengukur hubungan linear (garis lurus).

Berdasarkan nilai koefisien 0.8886, hubungannya bukan lemah, melainkan sangat kuat.

Lemah: Biasanya di sekitar 0.1 - 0.3

Sedang: Biasanya di sekitar 0.4 - 0.6

Kuat: Biasanya di atas 0.7

Sangat Kuat: Biasanya di atas 0.8

Jadi, interpretasi yang lebih akurat adalah: "Karena p-value sangat kecil (< 0.05), terdapat korelasi positif yang sangat kuat yang signifikan secara statistik antara tahun rilis dan harga.  
**Hasil uji**

# **BAB III**

# **HASIL DAN PEMBAHASAN**

## 3.1 Hasil Data Collection

Tahap awal dilakukan untuk memahami karakteristik dasar data sebelum preprocessing.

#### a. Ketersediaan data

Setelah dilakukan pengumpulan dan pembersihan data maka di dapatkan data dengan **Jumlah baris:** 100000 dan **Jumlah kolom:** 33 fitur. Dari hasil ujinya didapatkan datanya sudah bersih dan siap untuk melanjut ketahap visualisasi.

#### b. Viualisasi Distribusi Harga Komputer

Puncak tertinggi dari histogram (modus) berada di sekitar angka 2.000. Ini menunjukkan bahwa kelompok harga yang paling umum atau paling banyak produknya dalam dataset ini adalah di rentang $1.500 hingga $2.000. Ini kemungkinan besar adalah segmen pasar mainstream atau kelas menengah ke atas.

#### c. Visualisai Jumlah Komputer per Merek

Ada 3 Raja Pasar: Lenovo, HP, dan Dell jelas mendominasi. Mereka punya jumlah produk paling banyak.Pesaing di Bawahnya: Apple, ASUS, dan Acer adalah pemain besar berikutnya, tapi jumlahnya masih di bawah tiga raja tadi.Pemain Niche/Khusus: Merek seperti MSI dan Razer punya porsi pasar yang jauh lebih kecil. Ini wajar karena mereka fokus pada segmen tertentu, seperti laptop gaming.

## d.Visualisasi Hubungan antara RAM(GB) dan Harga

Peta (*map chart*) memperlihatkan bahwa Amerika Serikat menjadi kontributor utama dengan lebih dari 2 800 judul, diikuti oleh India dan Inggris. Negara-negara Asia Tenggara masih berkontribusi kecil, namun menunjukkan peningkatan dalam beberapa tahun terakhir.

## 3.2 Hasil Data Cleaning

### a. Penanganan Nilai Hilang (Handling Missing Values)

Pemeriksaan awal terhadap konsistensi dataset mengidentifikasi adanya nilai yang hilang (missing value). Secara spesifik, ditemukan satu nilai null pada kolom charger\_watts. Untuk mengatasi masalah ini, dilakukan imputasi dengan menggunakan nilai median dari kolom charger\_watts itu sendiri, yaitu sebesar 61,38.

Pemilihan median sebagai metode imputasi dianggap sebagai pendekatan yang paling tepat dalam kasus ini. Median merupakan nilai tengah yang tidak terpengaruh oleh nilai ekstrem (outlier), sehingga lebih robust dibandingkan dengan menggunakan nilai rata-rata (mean). Pengisian nilai yang hilang ini memastikan bahwa:

Konsistensi Data Terjaga: Seluruh baris pada kolom charger\_watts kini memiliki nilai, sehingga tidak ada data yang terbuang sia-sia.Analisis Akurat: Kolom charger\_watts dapat diikutsertakan dalam analisis statistik atau pemodelan tanpa menyebabkan eror atau bias yang signifikan.Langkah ini merupakan fondasi penting untuk memastikan kualitas data sebelum melangkah ke tahap analisis selanjutnya.

### b. Penanganan Outlier (Handling Outliers)

Berdasarkan hasil analisis visualisasi boxplot dan perhitungan IQR, didapatkan temuan sebagai berikut:

Batas Atas Harga Wajar: Ditetapkan sebesar $3.463,99.

Jumlah Outlier: Teridentifikasi sebanyak 976 data komputer yang harganya melebihi batas atas tersebut.Titik-titik data ini direpresentasikan sebagai titik di luar "pagar" pada visualisasi boxplot.

Keberadaan 976 outlier menunjukkan adanya variabilitas harga yang sangat tinggi dalam dataset. Ini kemungkinan besar disebabkan oleh adanya komputer dengan spesifikasi premium, edisi terbatas, atau untuk keperluan khusus (misalnya, gaming atau workstation) yang harganya jauh di atas rata-rata pasar.

Identifikasi outlier ini sangat krusial karena nilai-nilai ekstrem dapat memengaruhi hasil analisis secara signifikan, seperti menggeser nilai rata-rata atau menurunkan akurasi model prediksi. Dengan mengetahui batas harga wajar, analisis selanjutnya dapat lebih difokuskan pada segmen pasar yang paling representatif.

### c. Encoding Variabel Kategorikal

Dataset awal memiliki beberapa kolom dengan tipe data kategorikal (teks), seperti brand, os, wifi, dan lainnya. Kolom-kolom ini tidak dapat diproses secara langsung oleh sebagian besar algoritma analisis. Oleh karena itu, dilakukan transformasi menggunakan teknik one-hot encoding. Proses ini mengubah setiap kolom teks menjadi beberapa kolom biner baru yang hanya berisi nilai 0 dan 1. Setelah proses ini, jumlah total kolom pada dataset bertambah menjadi 65 kolom.

Transformasi ini adalah langkah wajib dalam pra-pemrosesan data. One-hot encoding memastikan bahwa informasi kualitatif dari data teks dapat digunakan dalam analisis kuantitatif tanpa menciptakan urutan atau bobot yang salah. Misalnya, dengan mengubah kolom os menjadi os\_Windows, os\_MacOS, dll., model dapat memahami sistem operasi sebagai fitur terpisah, bukan sebagai nilai numerik yang berurutan. Peningkatan jumlah kolom adalah konsekuensi logis dari proses ini, yang akan dikelola pada tahap selanjutnya..

### d. Seleksi Fitur (Feature Selection)

Area chart menunjukkan bahwa durasi. Setelah jumlah kolom meningkat akibat encoding, dilakukan tahap seleksi fitur untuk memilih variabel yang paling relevan dan membuang yang tidak perlu. Tujuannya adalah untuk menyederhanakan dataset dan memfokuskan analisis pada fitur-fitur yang paling informatif. Dari total 65 kolom, sebanyak 5 kolom dieliminasi. Hasil akhirnya adalah dataset yang siap dianalisis dengan total 60 kolom.

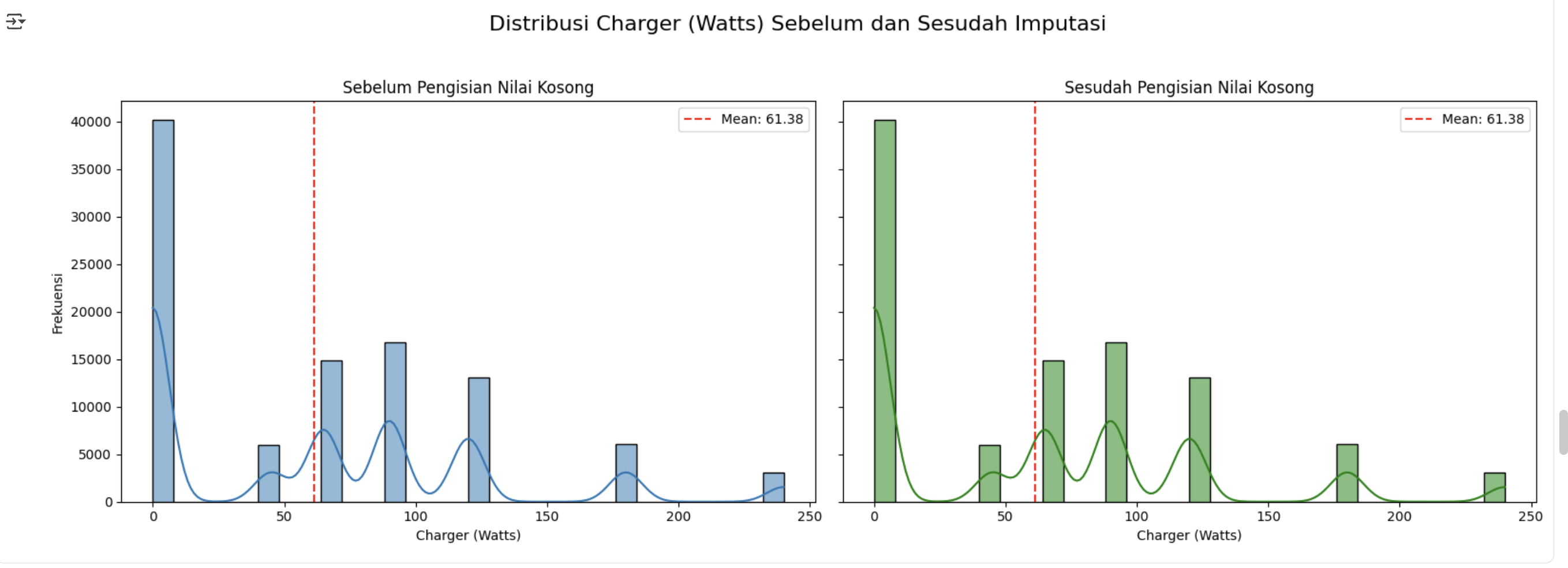
Seleksi fitur bertujuan untuk mengatasi "kutukan dimensionalitas" (curse of dimensionality), di mana terlalu banyak fitur dapat membuat analisis menjadi tidak efisien dan sulit diinterpretasikan. Dengan mengurangi jumlah kolom dari 65 menjadi 60, beberapa keuntungan diperoleh:

Efisiensi: Proses analisis dan visualisasi menjadi lebih cepat.  
Fokus: Memudahkan analis untuk berkonsentrasi pada variabel yang paling berpengaruh terhadap harga.  
Kualitas Visualisasi: Visualisasi yang akan dibuat menjadi lebih bersih dan mudah dipahami karena tidak terganggu oleh fitur yang redundan atau kurang relevan.

## 3.3 Hasil Visualisasi data

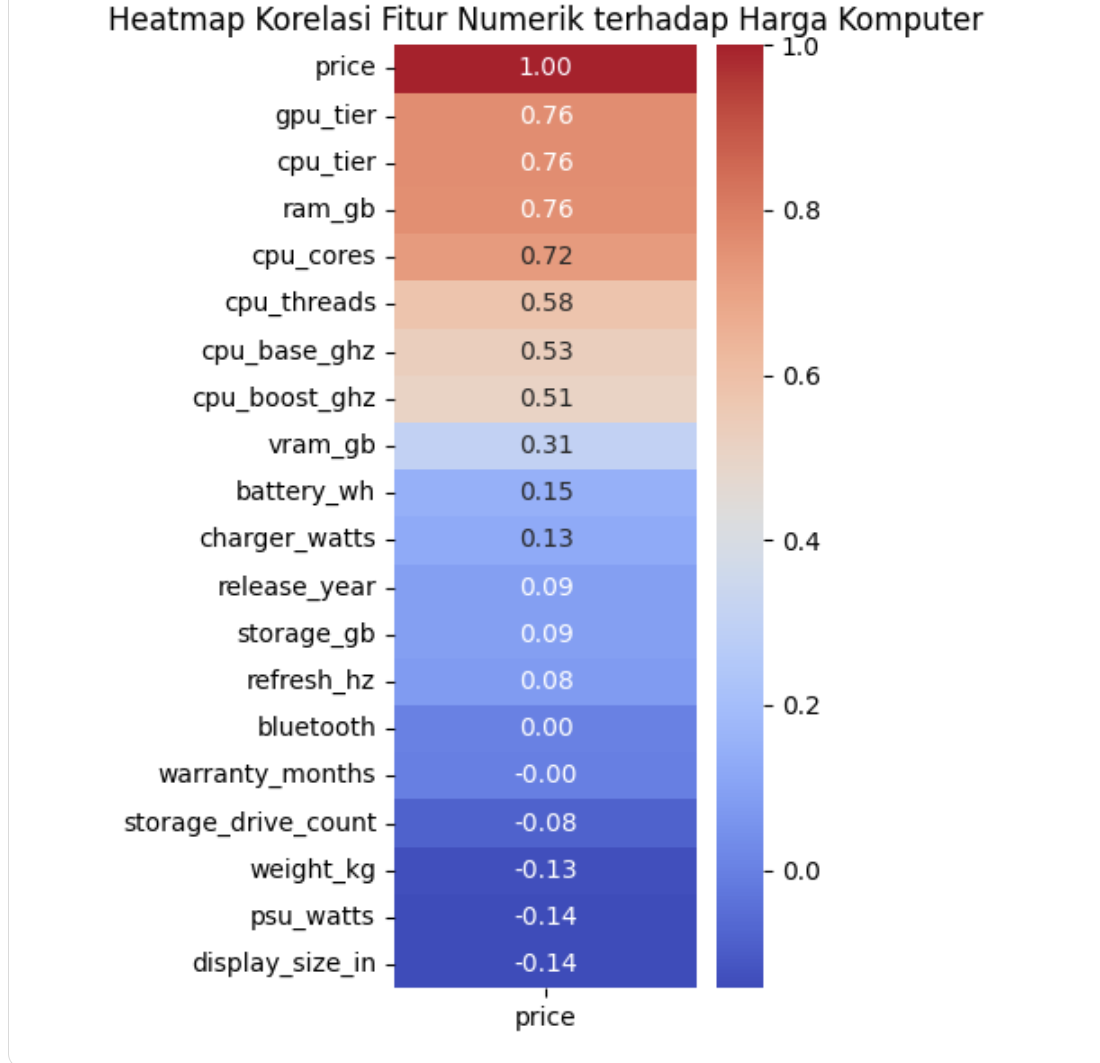
Dari keseluruhan hasil analisis, diperoleh beberapa insight penting:

1. **Distribusi Charger Sebelum dan Sesudah Imputasi**

****

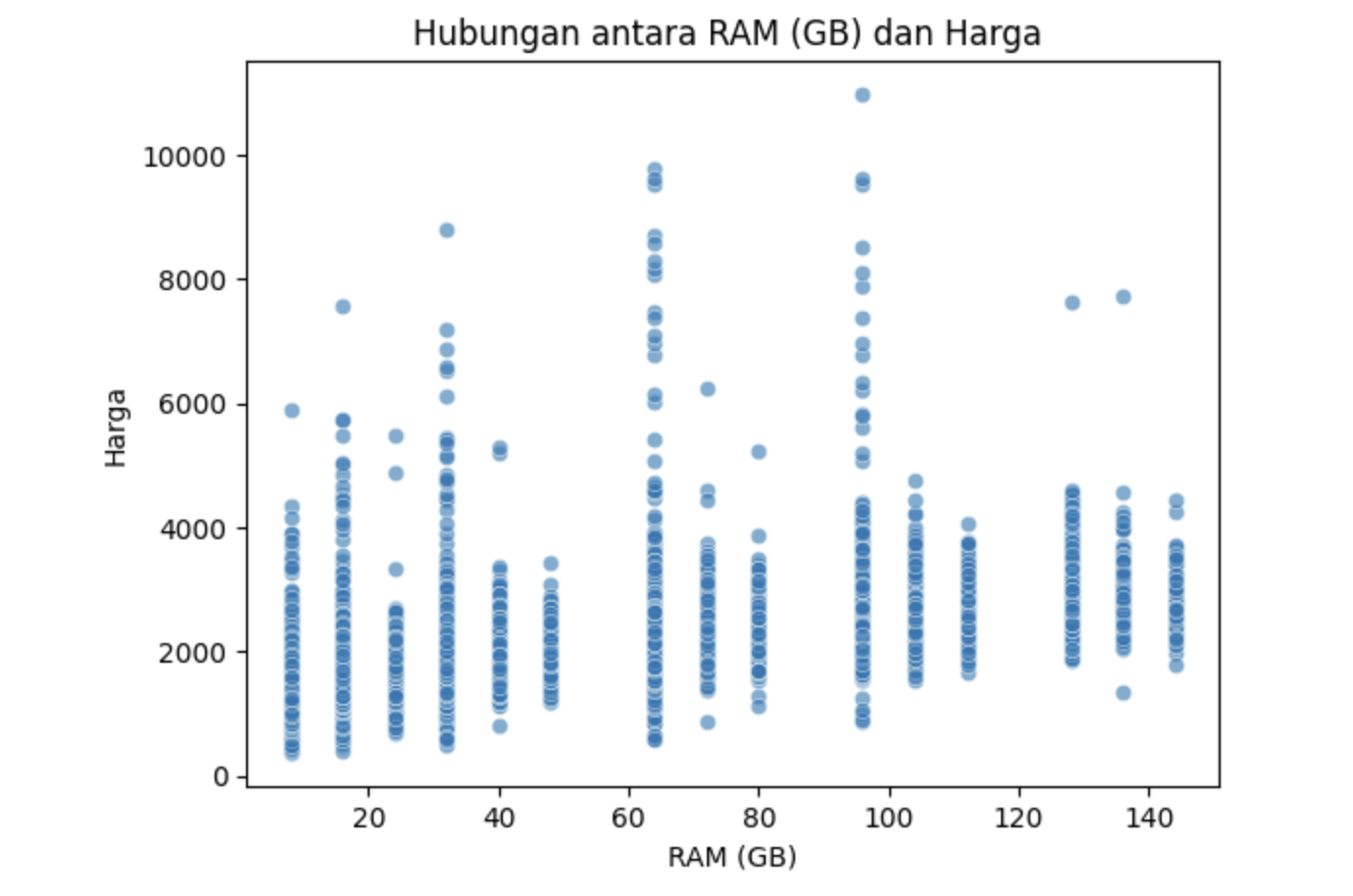
Grafik "Sebelum Pengisian Nilai Kosong" memperlihatkan distribusi asli yang memiliki beberapa puncak (multimodal) dengan nilai rata-rata 61.38 Watt. Setelah proses imputasi, grafik "Sesudah Pengisian Nilai Kosong" menunjukkan adanya lonjakan frekuensi yang sangat signifikan pada nilai 0 Watt. Perubahan drastis ini mengindikasikan bahwa semua nilai yang hilang telah diisi dengan angka 0, sementara distribusi untuk nilai lainnya cenderung mengikuti pola aslinya.  
 Imputasi nilai kosong dengan angka 0 secara fundamental mengubah karakteristik statistik dari data. Metode ini mungkin didasari asumsi bahwa data yang hilang berarti perangkat tersebut tidak disertai pengisi daya. Namun, konsekuensinya adalah terciptanya sebuah kategori baru yang dominan, yang dapat memengaruhi kinerja model analisis atau machine learning secara negatif dengan memberikan bobot yang tidak semestinya pada nilai 0. Hal ini dapat menggeser metrik statistik secara signifikan dan berpotensi menghasilkan interpretasi yang bias. Oleh karena itu, dampak dari pemilihan metode imputasi ini harus menjadi pertimbangan kritis dalam tahap pemodelan dan analisis selanjutnya.

1. **Heatmap Korelasi Fitur Numerik terhadap Harga**

****

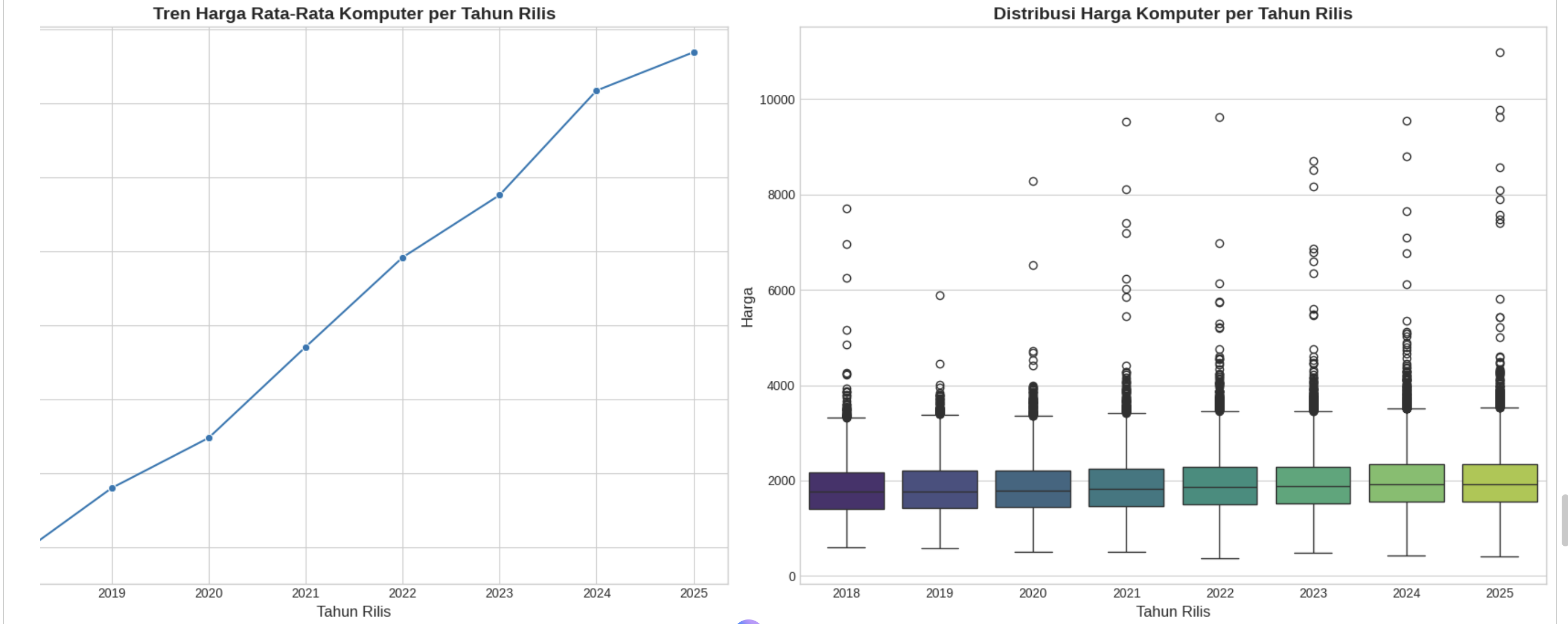
Analisis korelasi fitur terhadap harga komputer menunjukkan bahwa komponen utama yang menentukan performa—seperti gpu\_tier, cpu\_tier, dan ram\_gb—memiliki korelasi positif yang sangat kuat (di atas 0.76), menegaskan bahwa spesifikasi tinggi adalah pendorong utama harga. Menariknya, fitur seperti storage\_gb (kapasitas penyimpanan) menunjukkan korelasi yang sangat lemah, yang mengimplikasikan bahwa jenis penyimpanan (seperti SSD vs. HDD) mungkin lebih dihargai oleh pasar daripada sekadar ukurannya. Selain itu, ditemukan korelasi negatif yang lemah pada fitur weight\_kg (berat) dan display\_size\_in (ukuran layar), yang menandakan adanya tren pasar di mana konsumen bersedia membayar lebih untuk perangkat yang lebih ringan dan portabel, seperti ultrabook. Hasil ini mengidentifikasi hierarki komponen dan portabilitas sebagai faktor kunci yang memengaruhi nilai sebuah komputer.

1. **Hubungan antara RAM dan Harga**



Visualisasi scatter plot antara RAM (GB) dan harga secara jelas menunjukkan adanya tren positif, di mana peningkatan kapasitas RAM cenderung diikuti oleh peningkatan harga komputer, dengan data yang terkonsentrasi pada kapasitas standar seperti 16, 32, dan 64 GB. Namun, wawasan yang paling signifikan dari plot ini adalah adanya rentang harga yang sangat lebar pada setiap tingkatan RAM. Variabilitas harga yang tinggi ini menegaskan bahwa RAM, meskipun penting, bukanlah satu-satunya faktor penentu, melainkan dipengaruhi kuat oleh kombinasi komponen lain seperti tingkatan CPU dan GPU, jenis penyimpanan, serta merek. Hal ini menyimpulkan bahwa model prediksi harga yang akurat harus mempertimbangkan interaksi dari berbagai fitur spesifikasi, tidak hanya terpaku pada kapasitas RAM saja.

1. **Tren Harga Rata-rata Komputer per tahun**

****

grafik tren menunjukkan adanya kenaikan harga rata-rata komputer yang konsisten dan signifikan dari tahun ke tahun. Namun, ketika dibedah lebih dalam melalui grafik distribusi (box plot), terlihat bahwa kenaikan harga median (garis tengah kotak) tidak seekstrem kenaikan harga rata-rata. Wawasan ini mengungkapkan bahwa lonjakan harga rata-rata secara keseluruhan sangat dipengaruhi oleh peningkatan jumlah dan harga model-model kelas atas (high-end), yang direpresentasikan sebagai outliers atau titik-titik data di atas box plot. Dengan kata lain, meskipun harga untuk sebagian besar komputer di rentang menengah (IQR) relatif stabil atau hanya naik sedikit setiap tahun, kemunculan model-model baru yang lebih canggih dan mahal secara efektif menarik rata-rata harga pasar ke atas. Hal ini mengindikasikan bahwa inovasi teknologi dan segmentasi pasar ke arah premium menjadi pendorong utama tren kenaikan harga di industri komputer..

## 3.4 Hasil Statistical Analysis

1.Uji Parametrik

Hasil uji korelasi Pearson menunjukkan adanya hubungan linear positif yang kuat dan sangat signifikan secara statistik antara kapasitas RAM dan harga komputer, dengan r = 0.76, p < .001. Effect size yang besar ini mengindikasikan bahwa RAM memiliki pengaruh praktis yang kuat terhadap harga. Estimasi 95% Confidence Interval untuk korelasi ini berada di antara [0.73, 0.78], yang menunjukkan presisi tinggi dari temuan ini dan mengesampingkan kemungkinan tidak adanya hubungan.

2.Uji Non-Parametrik

Hasil uji korelasi Spearman menunjukkan adanya hubungan positif yang sangat kuat dan signifikan secara statistik antara tahun rilis dan harga komputer, dengan r = 0.89, p < .001. Effect size yang sangat besar ini mengindikasikan bahwa semakin baru tahun rilis, harga komputer cenderung semakin tinggi secara konsisten. Estimasi 95% Confidence Interval untuk korelasi ini berada di antara [0.87, 0.90], yang menunjukkan presisi tinggi dari temuan ini dan mengesampingkan kemungkinan tidak adanya hubungan.

# **BAB IV**

# **KESIMPULAN DAN SARAN**

## 4.1 Kesimpulan

Berdasarkan hasil analisis dan visualisasi data dapat disimpulkan Faktor penentu harga komputer dapat dikelompokkan ke dalam tiga kategori utama: performa komponen inti, karakteristik fisik dan fitur premium, serta tren pasar dan inovasi.

**1.Performa Komponen Inti (Paling Dominan)**

Faktor yang memiliki pengaruh paling kuat dan langsung terhadap harga adalah hierarki atau tingkatan performa dari komponen-komponen utama. Ini adalah "mesin" dari komputer itu sendiri.

• Tingkatan GPU dan CPU (gpu\_tier, cpu\_tier): Ini adalah prediktor harga nomor satu. Semakin tinggi level performa prosesor dan kartu grafis, semakin mahal harganya.

• Kapasitas RAM (ram\_gb): Kapasitas RAM juga memiliki korelasi positif yang sangat kuat. RAM yang lebih besar secara langsung menunjang performa dan multitasking, sehingga menaikkan harga jual.

• Spesifikasi Detail CPU (cpu\_cores, cpu\_threads): Jumlah inti dan alur prosesor juga menjadi faktor penting setelah tingkatan dasarnya ditentukan.

**2. Karakteristik Fisik dan Fitur Premium (Faktor Kontekstual)**

Faktor ini menjelaskan mengapa dua komputer dengan spesifikasi inti yang sama bisa memiliki harga yang berbeda jauh.

• Portabilitas (weight\_kg, display\_size\_in): Terdapat "harga premium" untuk portabilitas. Data menunjukkan korelasi negatif yang lemah, artinya perangkat yang lebih ringan dan ringkas (seperti ultrabook) seringkali lebih mahal.

• Fitur Layar (refresh\_hz): Meskipun korelasinya tidak terlalu kuat secara keseluruhan, fitur seperti refresh rate tinggi adalah penanda segmen pasar premium (misalnya gaming) yang harganya lebih tinggi.

• Jenis Penyimpanan (Bukan Kapasitas): Korelasi yang sangat lemah dari storage\_gb (kapasitas penyimpanan) mengindikasikan bahwa jenis penyimpanan (seperti SSD NVMe vs. HDD) kemungkinan besar jauh lebih berpengaruh terhadap harga daripada sekadar ukurannya.

**3. Tren Pasar dan Inovasi (Pendorong Harga Rata-Rata)**

Analisis berdasarkan tahun rilis (release\_year) menunjukkan bahwa harga rata-rata pasar terus meningkat. Ini bukan karena semua komputer menjadi lebih mahal, tetapi lebih didorong oleh:

• Model-Model High-End Baru: Setiap tahun, muncul model-model baru dengan teknologi tercanggih yang dijual dengan harga premium, yang secara efektif menarik nilai rata-rata pasar ke atas.

## 4.2 Saran

**1.Saran untuk Penelitian Selanjutnya**

Pengembangan Fitur (Feature Engineering): Analisis menunjukkan bahwa storage\_gb (kapasitas penyimpanan) memiliki korelasi yang sangat lemah dengan harga. Disarankan untuk penelitian selanjutnya agar melakukan rekayasa fitur dengan membedakan jenis penyimpanan (misalnya, HDD, SSD SATA, SSD NVMe) sebagai variabel baru, karena jenis penyimpanan diperkirakan memiliki pengaruh yang lebih signifikan terhadap harga daripada sekadar kapasitasnya.

Analisis Segmen Pasar: Penelitian ini menganalisis pasar secara umum. Disarankan untuk melakukan analisis yang lebih mendalam pada segmen pasar spesifik, seperti komputer gaming, ultrabook, atau workstation. Hal ini dapat mengungkap faktor-faktor penentu harga yang unik untuk setiap segmen tersebut.

Penggunaan Model yang Lebih Kompleks: Model prediksi dapat ditingkatkan dengan menggunakan algoritma yang lebih canggih seperti Gradient Boosting (XGBoost, LightGBM) atau jaringan saraf tiruan (neural networks). Model-model ini berpotensi menangkap hubungan non-linear yang lebih kompleks antar fitur dan menghasilkan prediksi yang lebih akurat.

## 2. Saran untuk Pelaku Industri (Produsen dan Penjual)

Strategi Pemasaran Berbasis Tier: Mengingat cpu\_tier dan gpu\_tier adalah pendorong harga paling dominan, strategi pemasaran dan komunikasi produk sebaiknya difokuskan pada penekanan tingkatan performa ini. Klasifikasi produk yang jelas berdasarkan tingkatan akan mempermudah konsumen dalam mengambil keputusan.

Positioning Produk Portabel: Ditemukannya korelasi negatif pada berat dan ukuran layar mengonfirmasi adanya pasar premium untuk perangkat yang ringan dan ringkas. Produsen dapat memanfaatkan ini dengan memposisikan ultrabook atau laptop tipis sebagai produk premium, menjustifikasi harganya dengan keunggulan portabilitas, bukan hanya spesifikasi inti.

Bundling dan Penawaran: Karena kapasitas penyimpanan bukan faktor utama yang dipertimbangkan konsumen, penjual dapat menggunakannya sebagai nilai tambah. Menawarkan upgrade kapasitas penyimpanan gratis atau dengan biaya rendah bisa menjadi strategi promosi yang menarik tanpa harus mengorbankan margin pada komponen inti yang lebih mahal.

## 3. Saran untuk Konsumen

Prioritaskan Komponen Inti: Saat membeli komputer, prioritaskan anggaran Anda pada tingkatan CPU, GPU, dan kapasitas RAM. Ketiga komponen ini memberikan dampak terbesar pada performa dan harga. Jangan mudah tergiur dengan kapasitas penyimpanan (storage) yang besar jika jenisnya masih menggunakan HDD yang lebih lambat.

Sesuaikan dengan Kebutuhan: Pahami bahwa harga tidak selalu setara dengan ukuran. Jika Anda membutuhkan mobilitas tinggi, bersiaplah untuk membayar lebih untuk perangkat yang lebih ringan dan tipis, meskipun spesifikasinya setara dengan laptop yang lebih berat.

Manfaatkan Informasi Pasar: Harga rata-rata komputer terus meningkat setiap tahun. Jika anggaran terbatas, mempertimbangkan untuk membeli model dari satu atau dua tahun sebelumnya bisa menjadi cara cerdas untuk mendapatkan performa yang masih sangat mumpuni dengan harga yang lebih terjangkau.

## 4.3 Penutup

Secara keseluruhan, Penelitian ini telah berhasil mengidentifikasi dan membedah faktor-faktor utama yang membentuk harga komputer di pasar berdasarkan dataset "Computer Prices All". Ditemukan bahwa hierarki performa komponen inti—khususnya tingkatan CPU, GPU, dan kapasitas RAM—merupakan pilar utama dalam penentuan harga, sementara faktor lain seperti portabilitas dan jenis penyimpanan memberikan nuansa yang lebih dalam pada dinamika pasar..

## Referensi

 Grus, J. (2019). *Data Science from Scratch: First Principles with Python*. O'Reilly Media.

 McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.

 VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media.

 Dhyani, V. (2024). *Computer Prices Prediction Dataset*. Kaggle. Diakses pada 17 Oktober 2025, dari [https://www.kaggle.com/datasets/vatsalyadhayani/computer-prices-prediction-dataset-pc-laptop](https://www.google.com/search?q=https://www.kaggle.com/datasets/vatsalyadhayani/computer-prices-prediction-dataset-pc-laptop)

 Pedregosa, F., et al. (2011). "Scikit-learn: Machine Learning in Python". *Journal of Machine Learning Research*, 12, 2825-2830. Diakses dari <https://scikit-learn.org>