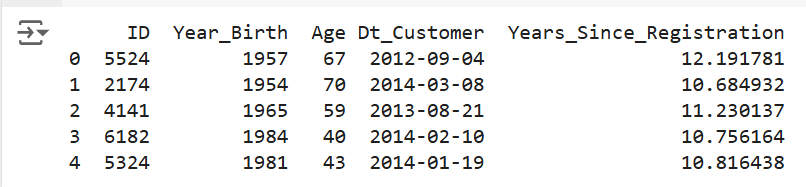
**Nama**  : Carmelita Cynthia Dagur

**Kelompok** : Empat ( 4 )

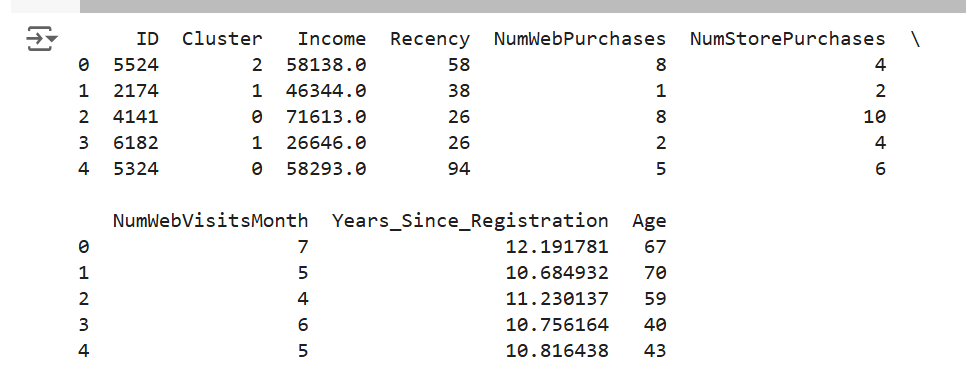
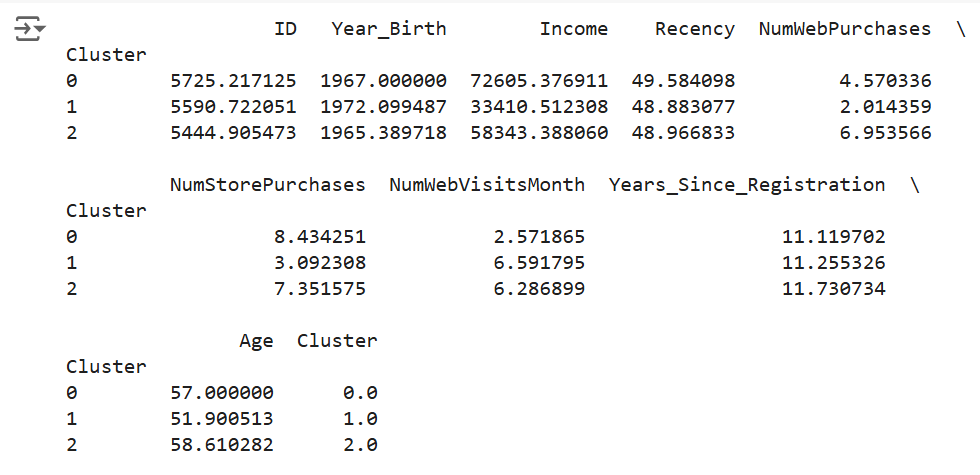
**Link Google Colab** : [Carmelita Cynthia Dagur\_\_Kelompok4\_Praktik2\_Modul4.ipynb](https://colab.research.google.com/drive/190v_1YpKNa1bLFR9ehxULIMcd7RIHUq5?usp=sharing)

**1. Tabel Segmentasi Pelanggan**

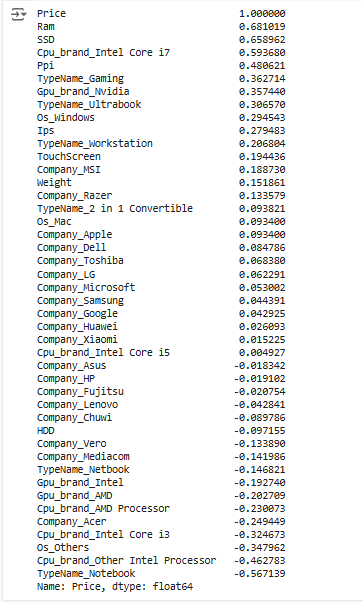
1. ****

**Pembahasan :**

Syntax ini dimulai dengan mengimpor pustaka yang diperlukan, termasuk `pandas` untuk manipulasi data, `KMeans` dari `sklearn` untuk klasterisasi, `StandardScaler` untuk normalisasi data, serta beberapa pustaka tambahan seperti `numpy` dan `datetime`. Data pelanggan yang sudah dibersihkan dimuat dari file CSV (`customer\_data\_cleaned.csv`). Kolom 'Dt\_Customer' yang berisi tanggal registrasi pelanggan diubah menjadi tipe data datetime dengan format tertentu. Selanjutnya, kolom 'Years\_Since\_Registration' dihitung berdasarkan selisih antara tanggal saat ini dan tanggal registrasi, kemudian dibagi 365 untuk mendapatkan hasil dalam tahun. Kolom 'Age' dihitung berdasarkan tahun kelahiran pelanggan yang terdapat dalam kolom 'Year\_Birth' dan tahun saat ini. Terakhir, kode menampilkan lima baris pertama dari kolom-kolom tertentu, seperti ID, tahun kelahiran, usia, tanggal registrasi, dan lama registrasi dalam tahun.

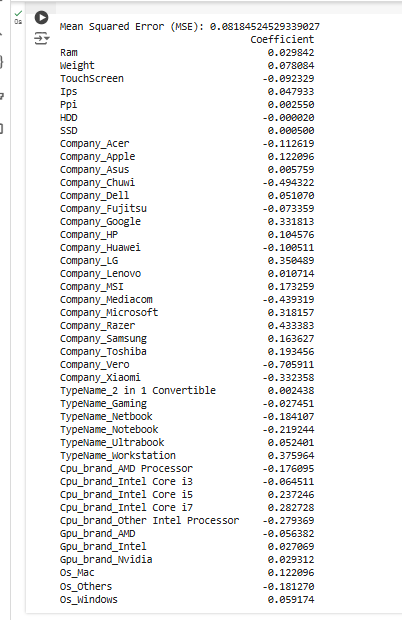
1. **  
   **

**Pembahasan :**Syntax ini dimulai dengan memilih beberapa fitur dari data yang telah dibersihkan, seperti `Income`, `Recency`, `NumWebPurchases`, dan lainnya, yang akan digunakan untuk analisis klaster. Fitur-fitur tersebut kemudian dinormalisasi menggunakan `StandardScaler` agar skala variabel menjadi seragam sebelum analisis lebih lanjut. Selanjutnya, metode KMeans digunakan untuk mengelompokkan data menjadi 3 klaster berdasarkan fitur yang sudah dinormalisasi, dengan hasil klaster yang ditambahkan sebagai kolom baru ('Cluster') dalam data. Kode kemudian menampilkan lima baris pertama data dengan ID, klaster, dan fitur-fitur lainnya. Setelah itu, kode melanjutkan dengan menghitung rata-rata dari kolom numerik untuk masing-masing klaster menggunakan `groupby` dan `mean`, yang memberikan gambaran karakteristik tiap klaster. Hasilnya kemudian dicetak untuk dianalisis lebih lanjut.

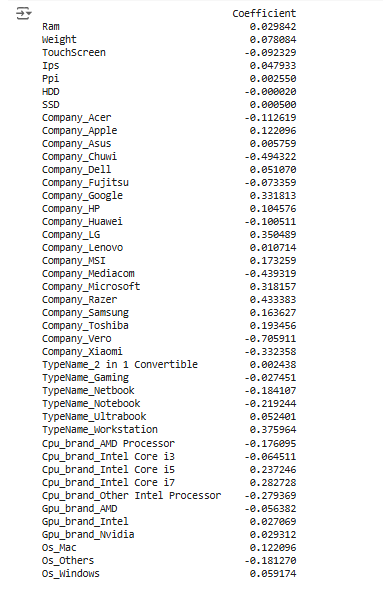
**2. A. **

**Pembahasan :**

Syntax ini dimulai dengan menghitung matriks korelasi antara semua kolom dalam data yang telah diencoding (`df\_encoded`) menggunakan fungsi `.corr()`. Matriks korelasi ini menunjukkan hubungan statistik antara variabel numerik dalam dataset. Kemudian, kolom korelasi untuk variabel `Price` dipilih dan diurutkan berdasarkan nilai korelasi secara menurun (dari yang paling kuat hingga yang paling lemah) dengan menggunakan `.sort\_values(ascending=False)`. Hasilnya, kolom-kolom yang memiliki korelasi paling tinggi dengan harga (`Price`) akan ditampilkan terlebih dahulu. Kode ini membantu untuk mengidentifikasi variabel-variabel yang berpotensi mempengaruhi harga dan dapat digunakan dalam analisis lebih lanjut atau model prediktif.

**b. **

**Pembahasan :**Syntax ini dimulai dengan memisahkan dataset yang telah diencoding menjadi dua bagian: `X` yang berisi fitur-fitur input (semua kolom kecuali `Price`), dan `y` yang berisi variabel target (`Price`). Selanjutnya, data dibagi menjadi data pelatihan (80%) dan data pengujian (20%) menggunakan fungsi `train\_test\_split`. Kemudian, model regresi linier (`LinearRegression`) diinisialisasi dan dilatih pada data pelatihan (`X\_train` dan `y\_train`). Setelah model terlatih, prediksi harga dilakukan pada data pengujian (`X\_test`) dan nilai prediksi disimpan dalam `y\_pred`. Untuk mengevaluasi kinerja model, Mean Squared Error (MSE) dihitung dengan membandingkan nilai prediksi (`y\_pred`) dengan nilai aktual (`y\_test`), dan hasilnya ditampilkan. Terakhir, koefisien regresi dari model dipresentasikan dalam bentuk dataframe untuk menunjukkan seberapa besar pengaruh masing-masing fitur terhadap prediksi harga.

1. ****

**Pembahasan :**Syntax ini dimulai dengan memuat dataset yang telah dibersihkan (`laptop\_data\_cleaned.csv`) ke dalam sebuah DataFrame menggunakan `pandas`. Beberapa kolom kategorikal, seperti `TouchScreen` dan `Ips`, diubah menjadi format numerik menggunakan `LabelEncoder`, yang mengonversi nilai kategorikal menjadi label numerik. Selanjutnya, kolom-kolom kategorikal lainnya seperti `Company`, `TypeName`, `Cpu\_brand`, `Gpu\_brand`, dan `Os` diubah menjadi variabel dummy (kolom biner) menggunakan `pd.get\_dummies`. Setelah proses encoding selesai, variabel target (`Price`) dipisahkan dari fitur lainnya, dengan `X` berisi fitur input dan `y` berisi harga. Dataset dibagi menjadi data pelatihan dan data pengujian menggunakan `train\_test\_split`. Model regresi linier (`LinearRegression`) dilatih pada data pelatihan, kemudian digunakan untuk memprediksi harga pada data pengujian (`y\_pred`). Kinerja model diukur dengan menghitung Mean Squared Error (MSE) antara prediksi dan nilai aktual. Terakhir, koefisien regresi yang menggambarkan kontribusi setiap fitur terhadap harga dipresentasikan dalam bentuk dataframe.

**3. Insight dan Rekomendasi :   
Pernyataan 1 :**

* **Insight :**

Melalui analisis klasterisasi, pelanggan dapat dikelompokkan berdasarkan karakteristik yang serupa, seperti pendapatan, frekuensi pembelian online, dan waktu terakhir kali mereka bertransaksi. Setiap klaster memiliki pola yang berbeda, yang dapat memberikan gambaran tentang jenis pelanggan yang lebih sering berbelanja atau memiliki kecenderungan untuk membeli produk tertentu. Misalnya, klaster dengan pendapatan lebih tinggi mungkin lebih sering melakukan pembelian dalam jumlah besar atau produk premium, sementara klaster dengan pendapatan lebih rendah bisa menunjukkan pembelian yang lebih terarah pada barang-barang dengan harga lebih terjangkau.

* **Rekomendasi:**

Berdasarkan hasil klasterisasi ini, perusahaan dapat menyesuaikan strategi pemasaran untuk masing-masing klaster. Untuk klaster dengan pendapatan tinggi, strategi yang bisa digunakan adalah menawarkan produk premium atau diskon eksklusif. Untuk klaster dengan pembelian lebih jarang atau pendapatan lebih rendah, perusahaan bisa memfokuskan upaya pada promosi dan program loyalitas untuk meningkatkan frekuensi pembelian. Selain itu, dengan mengetahui waktu terakhir pembelian, kampanye pemasaran yang tepat waktu bisa dirancang untuk menarik kembali pelanggan yang telah lama tidak bertransaksi.

**Pernyataan 2 :**

* **Insight**

Berdasarkan analisis data, beberapa faktor yang paling mempengaruhi harga laptop adalah \*\*RAM\*\*, \*\*SSD\*\*, dan \*\*PPI (Pixels per Inch)\*\*. Laptop dengan \*\*RAM\*\* yang lebih besar (seperti 16GB) dan \*\*SSD\*\* yang lebih besar (512GB atau 1TB) cenderung memiliki harga yang lebih tinggi, karena keduanya sangat berpengaruh terhadap kinerja dan kecepatan akses data. Selain itu, laptop dengan layar berkualitas tinggi, yang diukur dengan \*\*PPI\*\* lebih besar, memiliki harga yang lebih tinggi, karena layar dengan resolusi tinggi memberikan kualitas visual yang lebih tajam dan jelas. \*\*Weight\*\* laptop juga mempengaruhi harga, di mana laptop yang lebih ringan, seperti ultrabook, sering kali dijual dengan harga premium karena kepraktisan dan mobilitas yang ditawarkan. Meskipun \*\*CPU brand\*\* dan \*\*GPU brand\*\* memiliki korelasi yang lebih kecil terhadap harga, mereka tetap memberikan pengaruh dalam segmen pasar tertentu yang memprioritaskan kinerja grafis.

* **Rekomendasi:**

Untuk meningkatkan penjualan, perusahaan sebaiknya fokus pada laptop dengan \*\*RAM besar\*\* dan \*\*SSD\*\* besar, terutama untuk menarik konsumen yang mencari performa tinggi. Laptop dengan kapasitas RAM 16GB atau lebih dan SSD 512GB atau 1TB akan lebih menarik bagi pengguna yang membutuhkan kinerja cepat, seperti profesional atau pelajar. Selain itu, memasarkan laptop dengan \*\*PPI tinggi\*\* bisa menjadi strategi untuk menarik konsumen yang mengutamakan kualitas tampilan visual, seperti desainer grafis atau pembuat konten. Untuk segmen yang mengutamakan portabilitas, perusahaan bisa mempromosikan laptop \*\*ringan\*\* (seperti ultrabook) kepada profesional dan pelajar yang sering bepergian. Meskipun merek seperti \*\*Apple\*\*, \*\*HP\*\*, dan \*\*Dell\*\* tidak terlalu mempengaruhi harga secara langsung, mereka tetap penting dalam strategi pemasaran, karena banyak konsumen yang memilih laptop berdasarkan merek yang terpercaya. Perusahaan juga dapat menawarkan bundle produk dengan aksesori untuk meningkatkan daya tarik, serta menyusun penawaran yang lebih tersegmentasi berdasarkan spesifikasi untuk memenuhi kebutuhan pasar yang berbeda.