**Ujian Akhir Semester Data Mining**

**Devina Dewi Faustina Basam (20220801250)**

**Eneng Mila Lestari wijayadi (20220801264)**

Carilah data riil dari suatu bisnis (industri nya bebas) minimal 1.000 row.

* Deskripsikan data yang akan anda analisis sehingga bisa tergambar masalah apa yang ingin anda selesaikan.
  + Gunakanlah salah satu tool data mining (misal rapid miner) untuk melakukan hal-hal berikut:
* Lakukan cleansing atas data tersebut sehingga data siap untuk dianalisis
* Lakukan analisis atas data tersebut dengan **minimal dua algoritma** data mining
* Interpretasikan hasilnya

Delivery yang harus dikumpulkan adalah:

* + Data riil minimal 1000 row dan deskripsi detail kasus
  + File hasil pengolahan dari Rapid Miner atau yang lainnya. Jika menggunakan selain Rapid Miner harus menyertakan installer nya atau coding-nya
  + Hasil cleansing (pre-processing) atas data dan penjelasan proses cleansing nya
  + Hasil analisis dengan algoritma yang dipilih yang dilengkapi dengan penjelasan hasil nya
  + Interpretasi atas hasil analisis

1. **Deskripsi Kasus**

Kami menggunakan dataset penjualan di amazon melalui Kaggle dan terdapat lebih dari 1000 row data pada dataset yang kami gunakan.

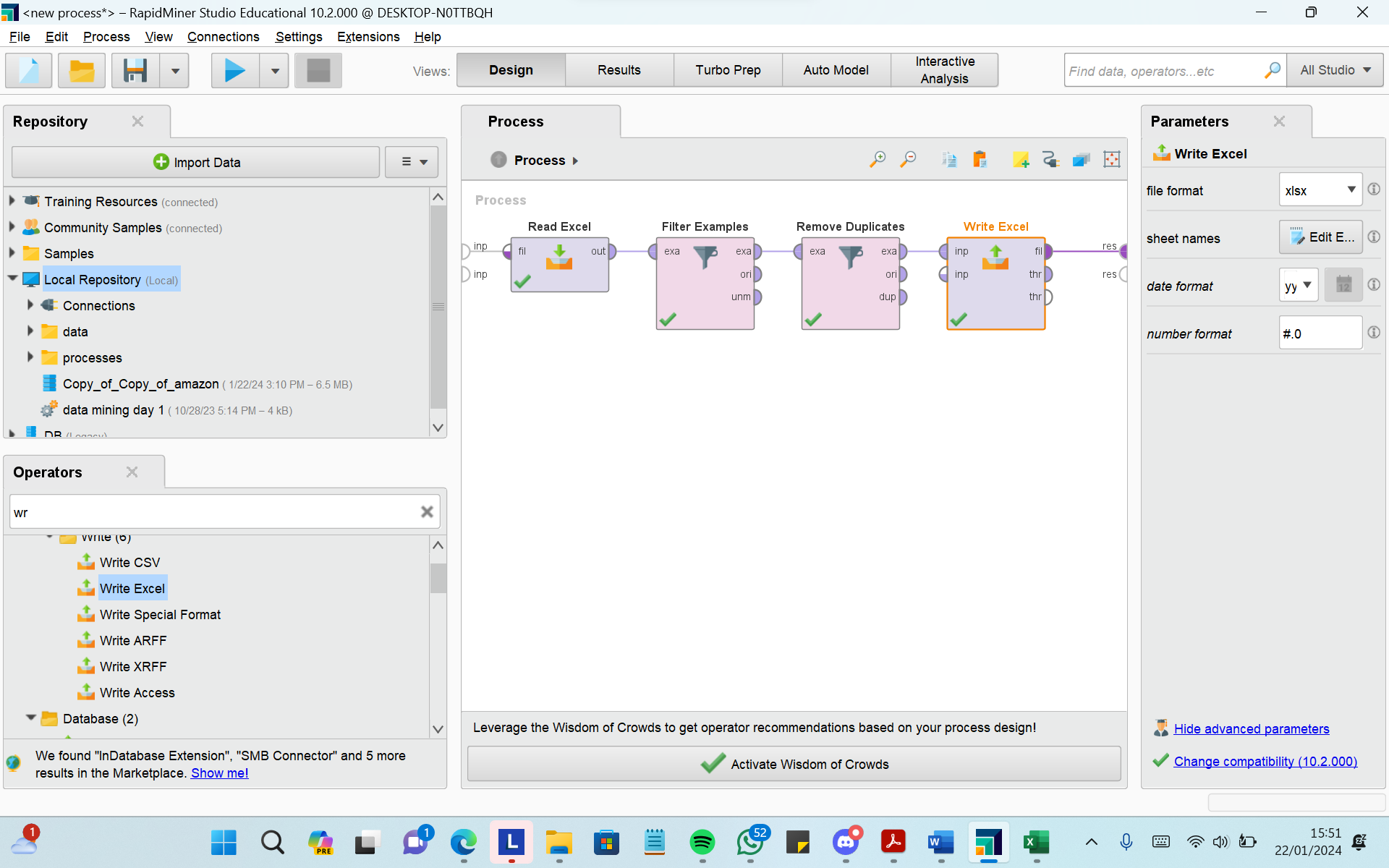
Kami mengambil 2 algoritma yaitu analisis sentiment untuk mendapatkan komentar positif, negatif, dan netral bagaimana kepuasan pelanggan setelah berbelanja di amazon untuk peningkatan layanan. Dan algortima clustering untuk mengelompokkan barang-barang yang telah terjual berdasarkan kategorinya agar memudahkan kita megetahui prduk kategori mana yang lebih banyak terjual.

Pada tahapan pertama kami melakukan cleansing data menggunakan rapid miner setelah itu dilakukan analisis sentiment dan clustring menggunakan Bahasa pemrograman phyton.

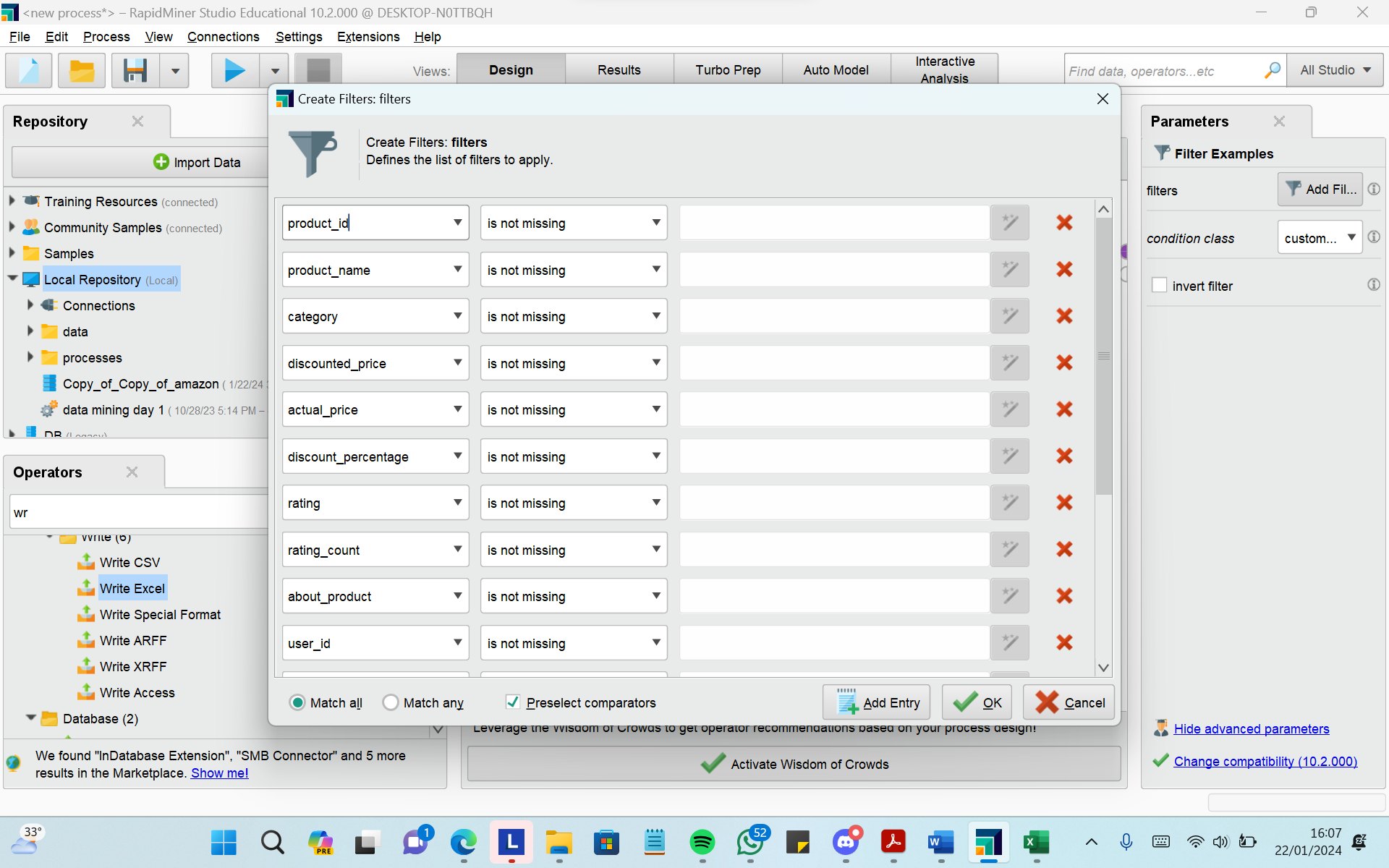
1. **Cleansing Data Menggunakan Rapid Miner**

Cleansing data ini dilakukan untuk membersihkan data dari simbol – simbol yang tidak diperlukan dan mengkonversi teks kedalam format yang sesuai untuk analisis.

Dalam Cleansing Data kami menggunakan tools sebagai berikut:

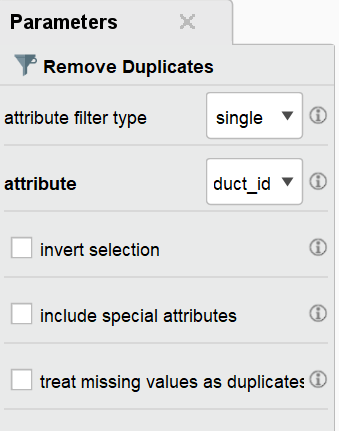


1. **Read Excel :** Digunakan untuk membaca file dataset excel.
2. **Filter Example:** Digunakan untuk menghapus missing value yang ada pada dataset, dengan menambah filter “is no missing” .



1. **Remove Duplicate:** Digunakan untuk menghapus duplicate yang terjadi di

product\_id, dengan mengganti attribute filter type “single” lalu pilih attribute “product\_id”.

****

1. **Write Excel:** Digunakan untuk menyimpan file excel yang telah selesai di cleansing.
2. **Analisis data Menggunakan Analisis Sentiment pada Google Colab**

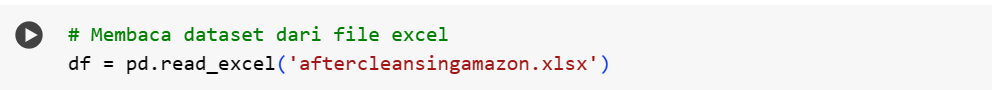
Analisis sentimen adalah proses untuk menilai atau mengevaluasi sentimen, opini, atau perasaan yang terkandung dalam teks, ulasan, atau data lainnya. Ini sering digunakan dalam berbagai konteks, termasuk bisnis, media sosial, survei pelanggan, dan analisis pendapat. Dalam data penjualan amazon ini kami menggunakan analisis sentiment untuk menampilkan hasil positif, negative, dan netral.

Implementasi Coding Phyton yang kami gunakan adalah sebagai berikut;

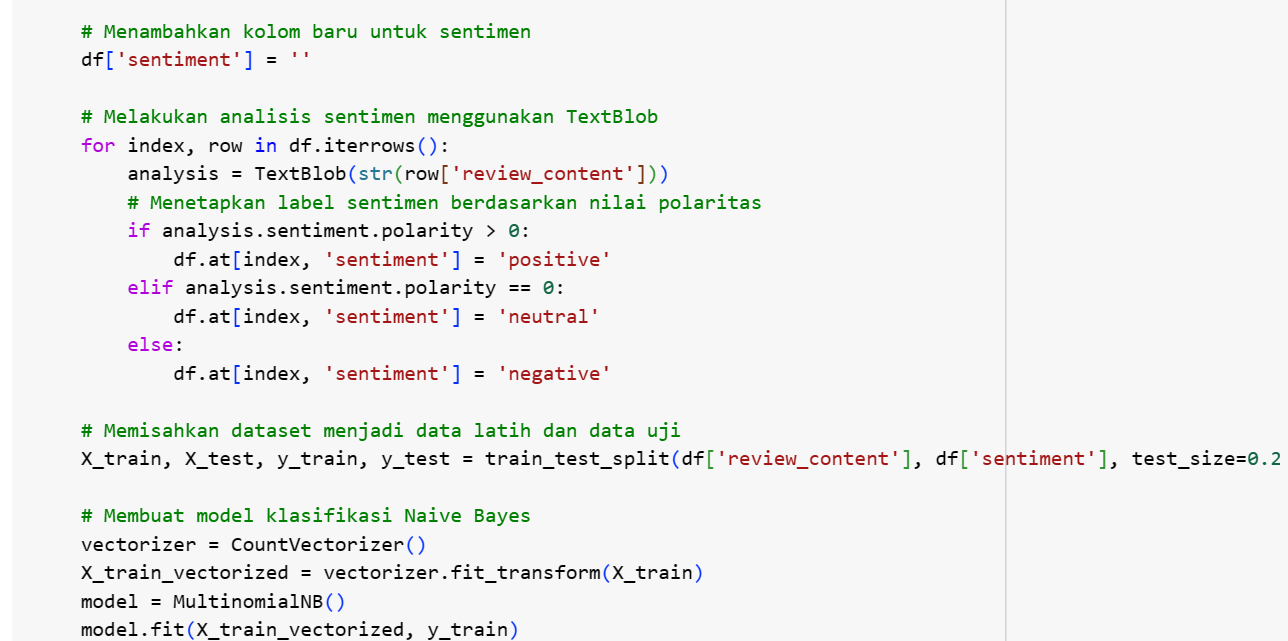
1. Import library

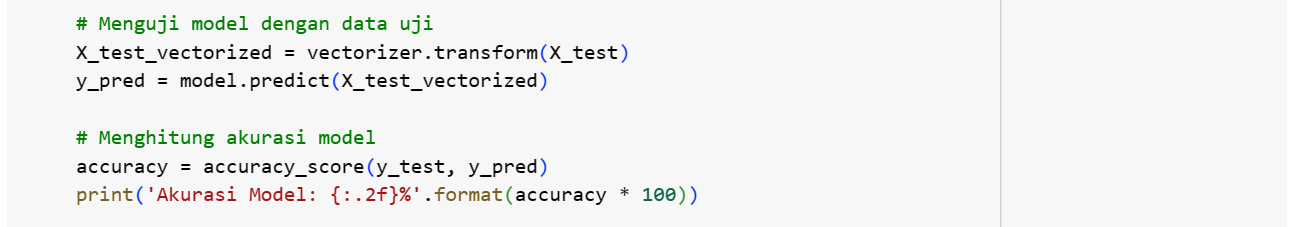


1. Memasukan data yang dibaca dari data excel ‘aftercleansingamazon.xlsx’ kedalam data frame.

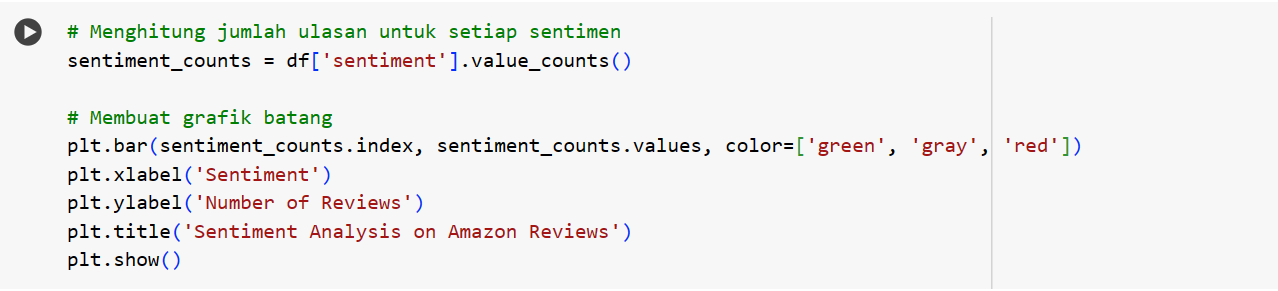


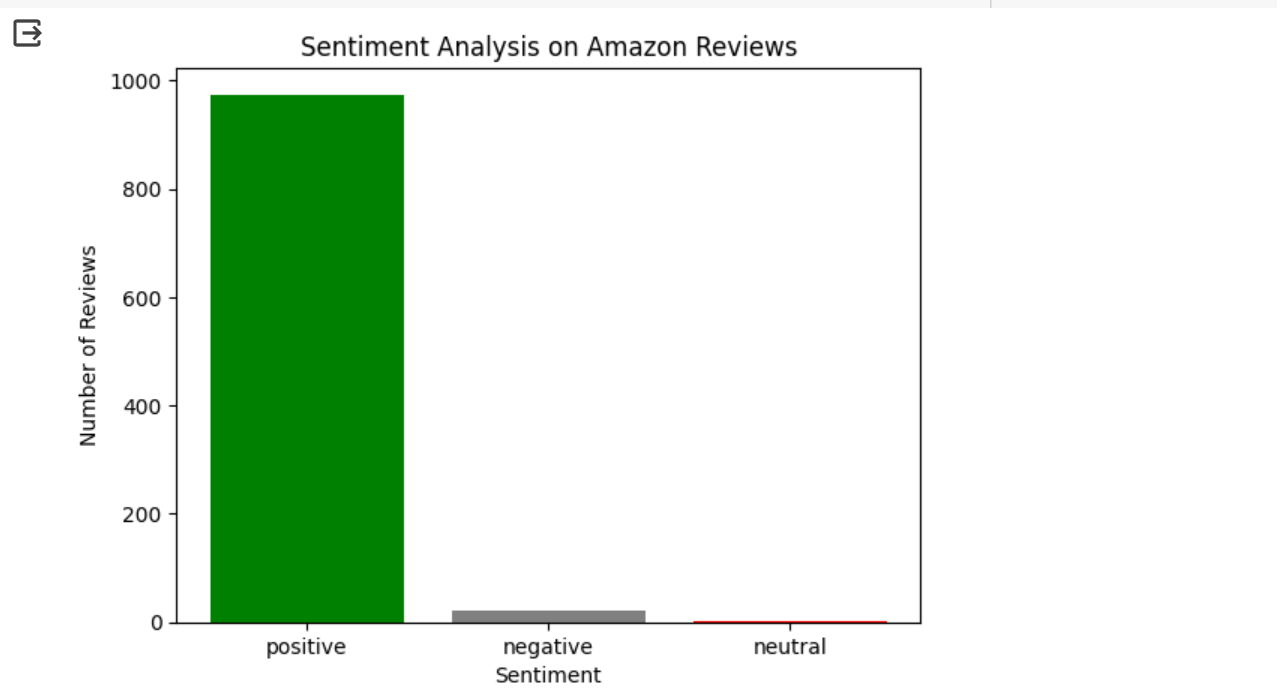
1. Implementasi untuk melakukan analisis sentiment.





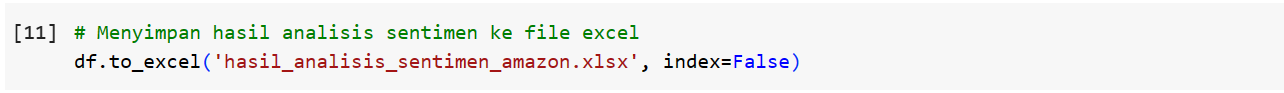
1. Menghitung jumlah ulasan dan memunculkan diagram batang





Dari hasil analisis sentiment diatas menunjukan bahwa pelanggan puas terhadap penjualan product amazon ini ditunjukan dengan diagram bersentimen positive lebih tinggi.

1. Menyimpan hasil analisis sentiment kedalam file excel.



1. Implementasi dari Codingan diatas dapat diakses pada file terlampir dengan nama analisis\_sentimen\_amazon.py dan juga link berikut:

<https://colab.research.google.com/drive/1Q3ucvrWKGLthkKCcN2BX33J7yfgprpA9?usp=sharing>

1. Kesimpulan yang dapat diambil :

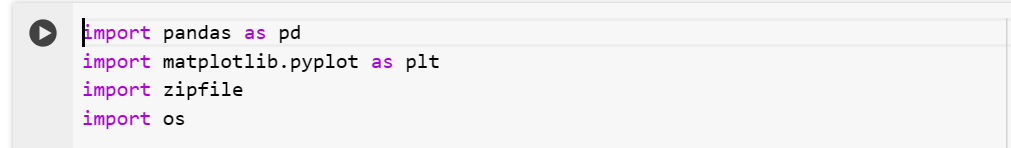
Dari hasil analisis sentiment diatas didapat bahwa pelanggan yang puas lebih banyak dari yang kurang puas.

1. **Analisis data Menggunakan Clustering pada Google Colab**

Clusterin adalah suatu Teknik mengelompokan data ke dalam kelompok-kelompok atau cluster berdasarkan kesamaan karakteristik tertentu. Dalam dataset penjualan amazon ini kami menggunakan clustering untuk mengelompokan barang barang yang telah terjual kedalam beberapa kategori.

Implementasi Coding Phyton yang kami gunakan adalah sebagai berikut.

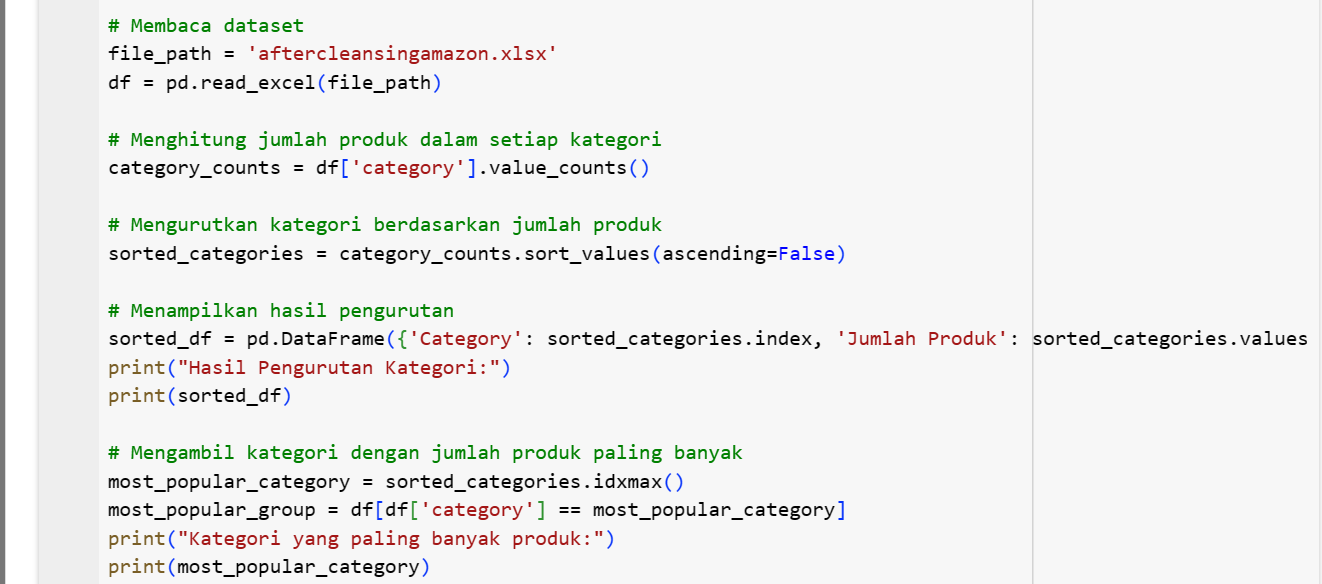
1. Import library



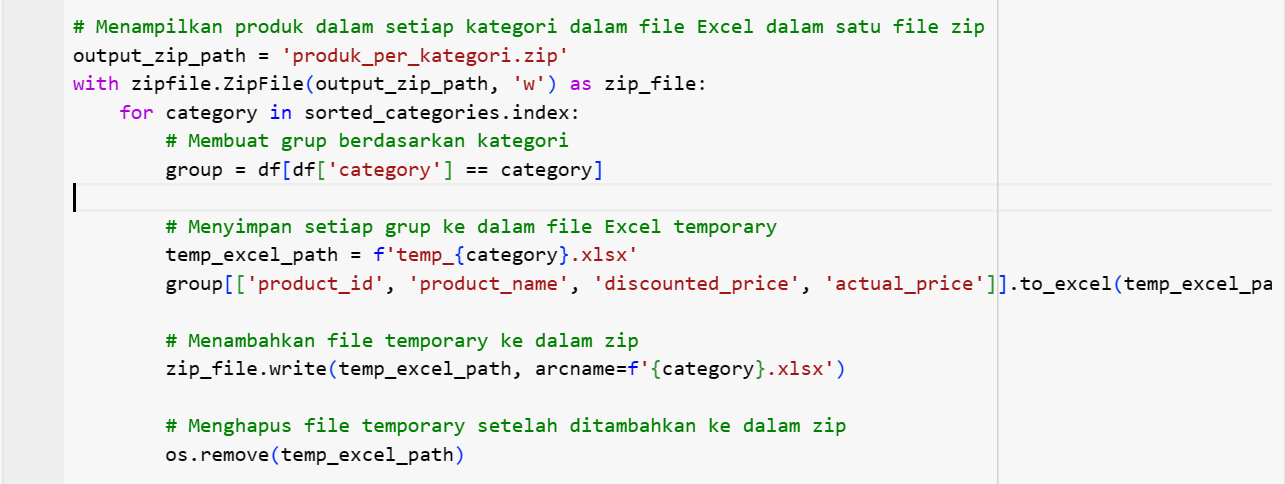
1. Memasukan data yang dibaca dari data excel ‘aftercleansingamazon.xlsx’ kedalam data frame.



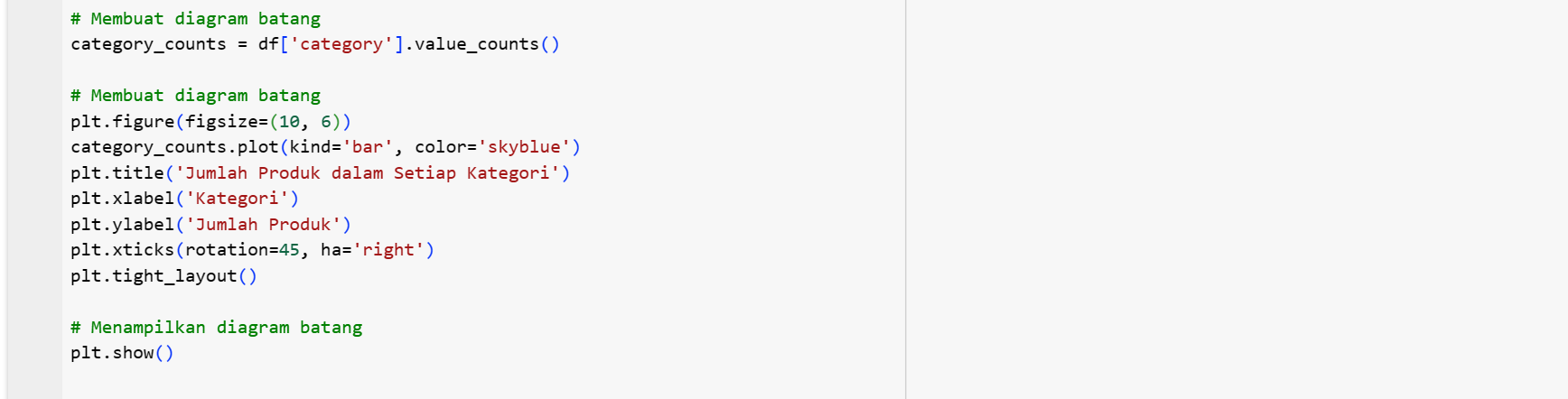
1. Membuat grup berdasarkan kategori.

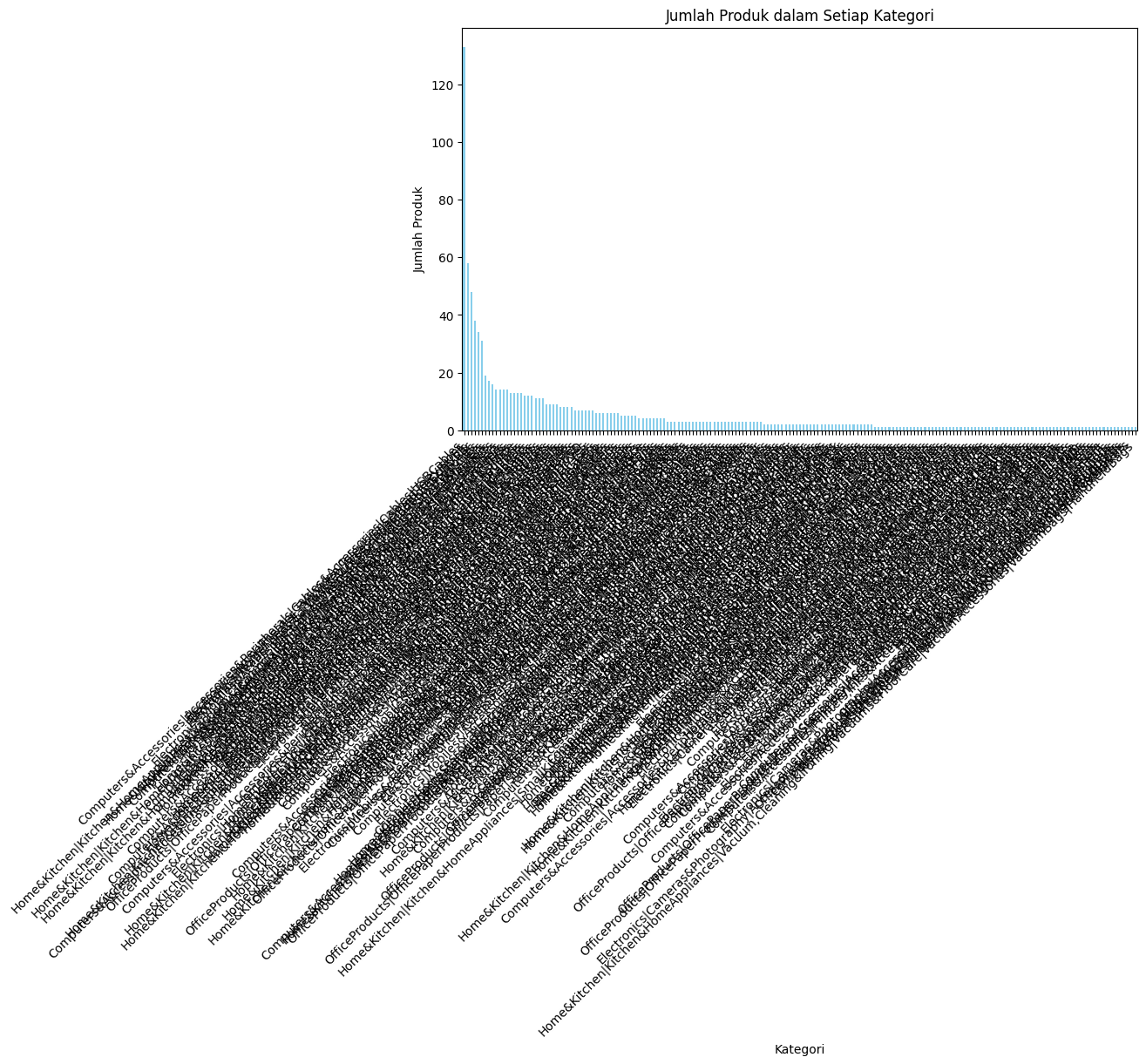


1. Menyimpan file excel perkategori kedalam filezip.



1. Membuat dan menampilkan diagram batang.





Dari hasil clustering diatas menunjukan bahwa ada 60 produk yang telah terjual masuk kedalam 180 kategori.

1. Implementasi dari Codingan diatas dapat diakses pada file terlampir dengan nama clustering\_amazon\_product.py dan juga link berikut: <https://colab.research.google.com/drive/1aGqbQFxYC1DadNOnNMWWsA4-tWT9aVxV?usp=sharing>
2. Kesimpulan :

Kategori data yang memiliki produk paling banyak terjual adalah Computers&Accessories|Accessories&Peripherals|Cables&Accessories|Cables|USBCables dengan jumlah 133 produk.

1. Intepretasi

* Dengan Analisis Sentiment untuk menentukan apakah layanan yang diberikan oleh amazon sudah baik berdasarkan review pelanggan.

Dari hasil analisis terlihat bahwa rating konsumen lebih banyak membawa rating positif terhadap penjualan produk diamazon. Dengan rating yang baik tersebut amazon dapat terus meningkatkan layanan menjadi lebih baik dan memperkuat pemasaran.

* Dengan Clustering untuk menentukan kategori mana yang diminati oleh pelanggan.

Berdasarkan analisis yang dilakukan dapat ditarik kesimpulan bahwa ada 133 produk dari 180 kategori dengan penjualan tertinggi yaitu computer & accessories, accesseories & pheriperals, cables & accessories, Cabbles, dan USBCables dengan jumlah 133 produk. Dengan adanya data clustering ini diharapkan amazon terus menyediakan produk tersebut dan meningkatkan variasi dari produk tersebut.