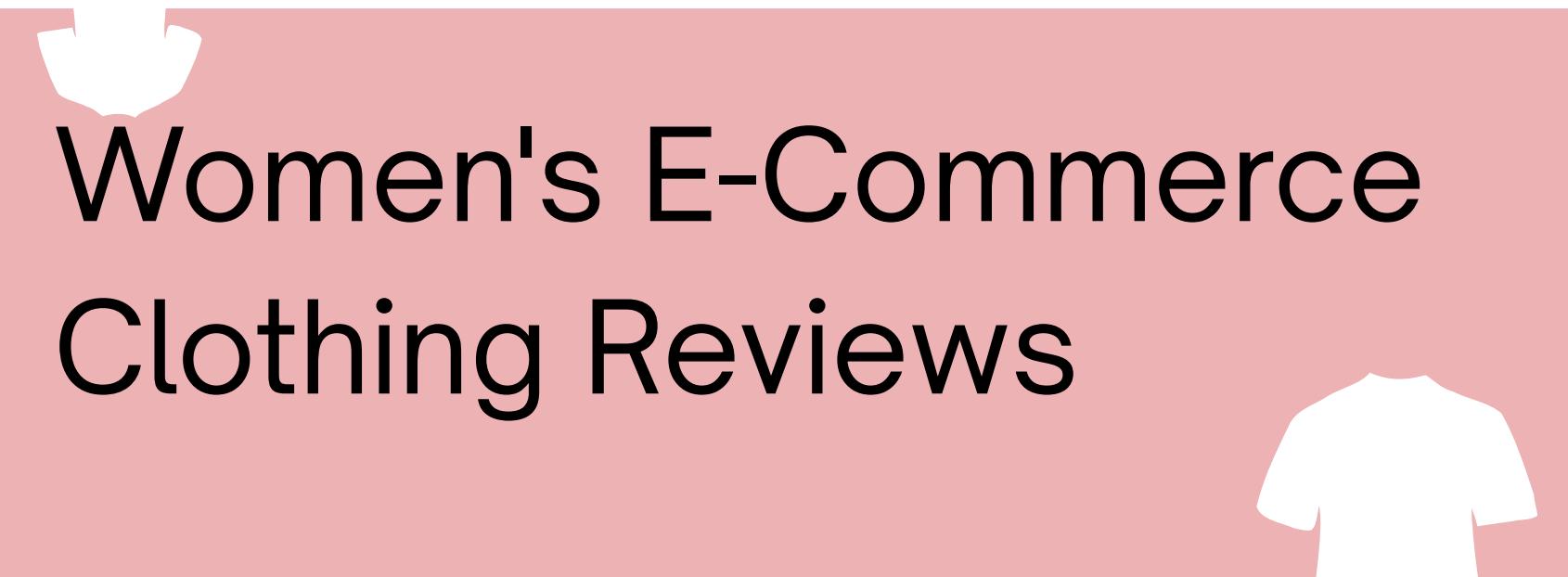


Final Project Data Analyst

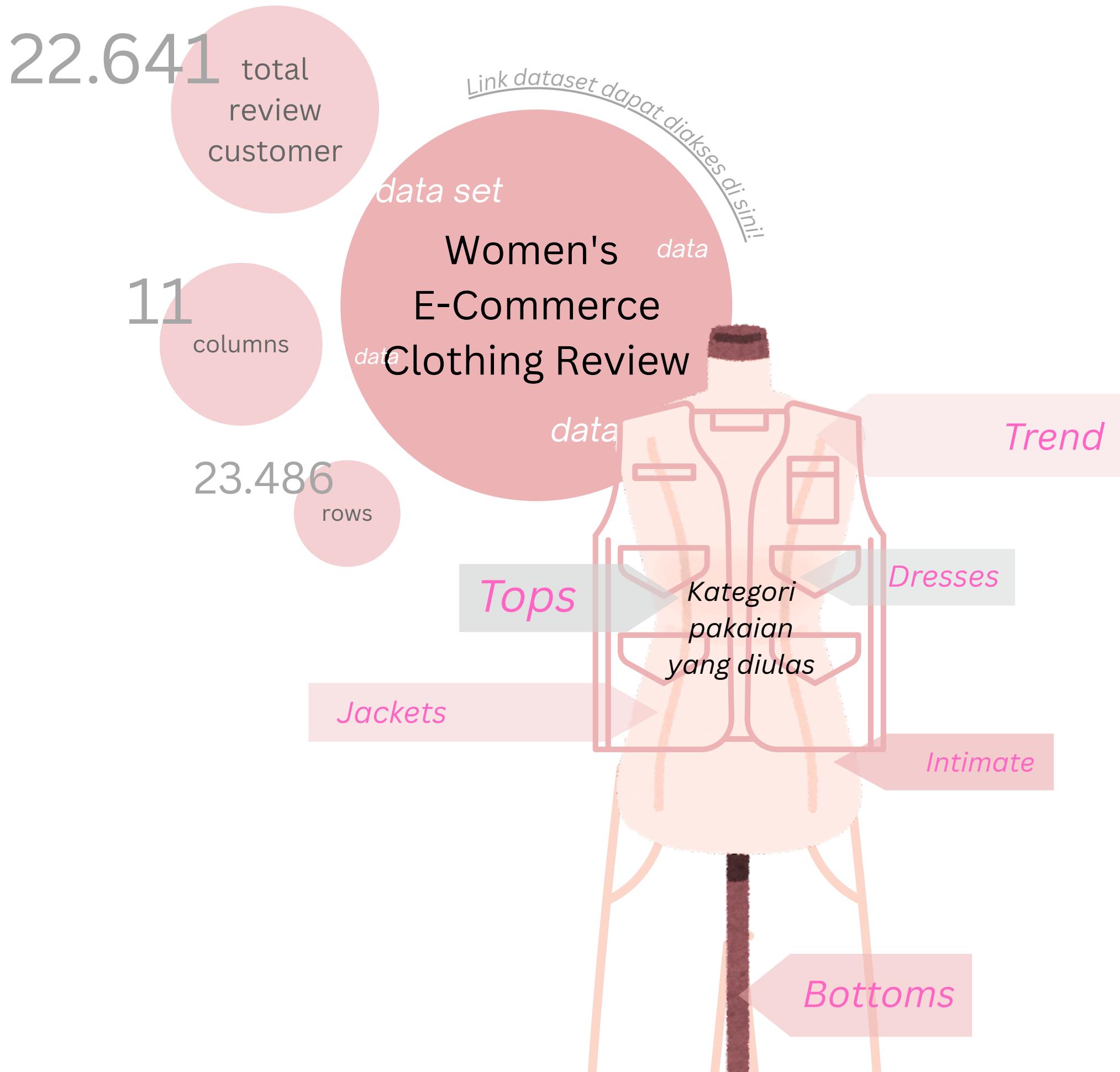


# Perkenalkan, nama saya **Qonitah Sarah**

Memiliki latar belakang pendidikan sarjana di bidang Arsitektur yang menaruh minat pada *decision-making* berbasis data. Untuk berkenalan lebih jauh, selengkapnya bisa mengunjungi profil [LinkedIn](#) dan tulisan saya di [Medium](#). Semoga sebelum berpisah, ada oleh-oleh kebaikan yang bisa dibawa pulang.

# Tentang Data

**Women's E-Commerce Clothing Review** adalah sebuah dataset yang diperoleh dari Kaggle. Data terdiri atas sekumpulan review dari pelanggan setelah melakukan pembelian produk pakaian wanita. Identitas nama pada data telah dianonimkan karena merupakan data yang valid. Secara umum, setiap review disertai dengan variabel **identitas pengulas, rating, dan detail pakaian yang dibeli**.



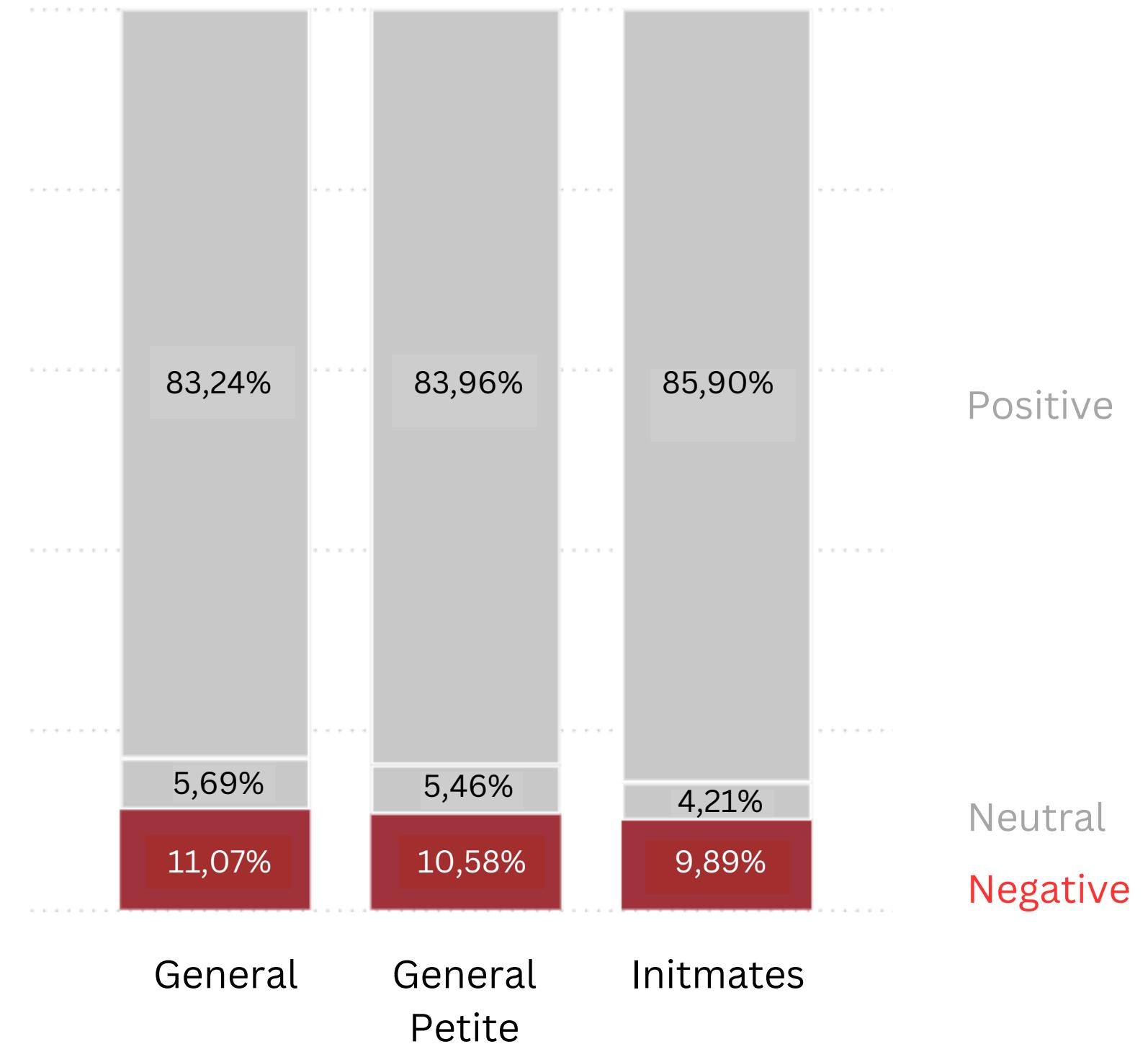
#### Problem Statement

Review positif mendominasi lebih dari 80%, namun terdapat komentar negatif yang lebih unggul dari yang netral

Jumlah review negatif hampir mencapai **2 kali lipat** dari review netral yang mengindikasikan bahwa masih terdapat beberapa hal yang mesti diperbaiki.

Walaupun berdasarkan angka di tiap divisi didominasi oleh review yang positif, **review negatif mampu mempengaruhi persepsi calon pelanggan hingga berpotensi memberikan citra yang buruk kepada merek.**

Jumlah Sentimen Pelanggan Berdasarkan Divisi Pakaian



# Tujuan

Berdasarkan pernyataan masalah sebelumnya, tujuan yang ingin dicapai adalah menganalisis hal-hal yang menjadi penyebab munculnya masalah dan melihat masalah secara lebih mendalam di berbagai sisi sehingga diperoleh rekomendasi yang tepat untuk memperbaiki pengalaman buruk pelanggan

## Tujuan Spesifik

Mengukur tingkat kepuasan dan sentimen pelanggan mengenai produk

Memperoleh insight atau wawasan dari analisis

Memberikan *actionable insight* atau rekomendasi untuk mewujudkan tujuan



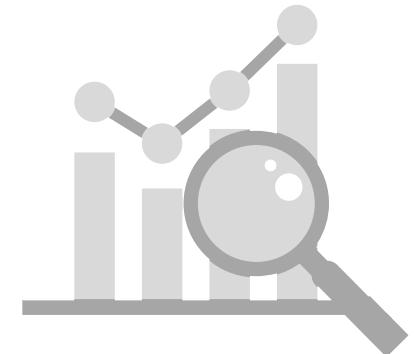
### Data preparation

- Mengecek tipe data
- Read data dalam format csv



### Data Cleaning

- Tidak ditemukan nilai duplikat
- Terdapat Missing value bernilai NaN yang tetap dipertahankan
- Outlier teridentifikasi di beberapa kolom namun tetap digunakan



### Exploratory Data Analysis & Feature Engineering

- Review rate
- Customer Satisfaction (CSAT) scoring
- Analisis sentiment dengan Twitter roberta base sentiment
- Tokenize Text
- Finalisasi data untuk pelaporan



### Data Visualization

- Customer sentiment dashboard
- Customer satisfaction dashboard

Cuplikan visualisasi dapat dilihat di bawah ini:

- [Appendix 2: Dashboard Customer Satisfaction](#)
- [Appendix 3: Dashboard Customer Sentiment](#)

# Exploratory Data Analysis Steps

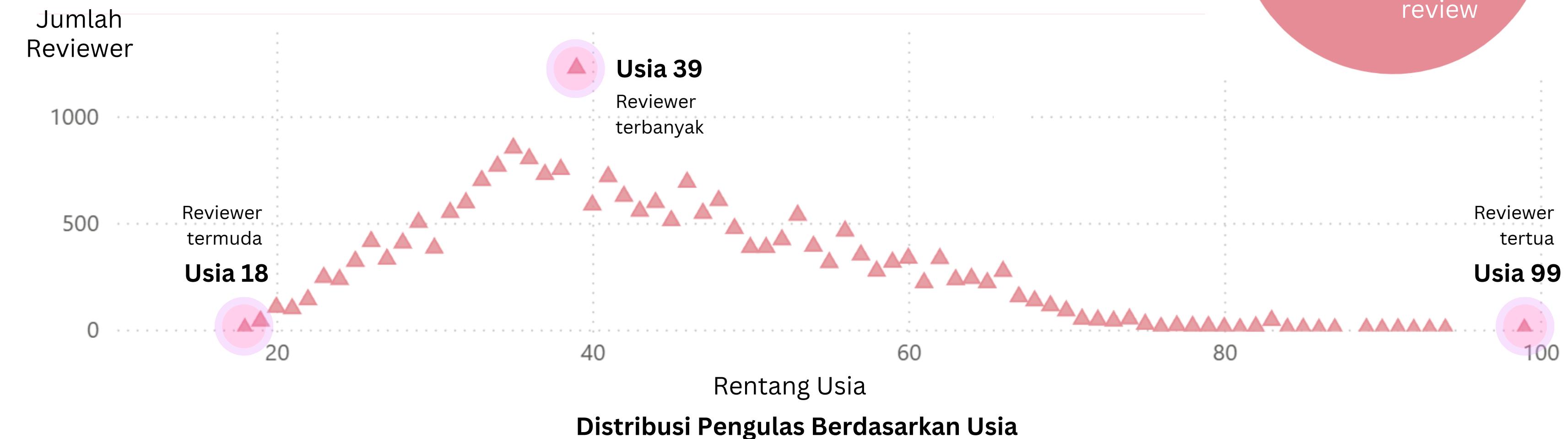
## Customer Satisfaction

# Siapa saja mereka yang menulis review?

96% dari total 23.486 pelanggan memberikan ulasan mengenai produk yang dibeli, hal tersebut menunjukkan tingginya *engagement* pelanggan hingga meluangkan waktu untuk menjelaskan pendapat mereka. Usia mereka sangat beragam, mulai dari yang sangat muda yaitu 18 tahun hingga yang tertua yaitu 99 tahun.

23.486

Total  
Pelanggan

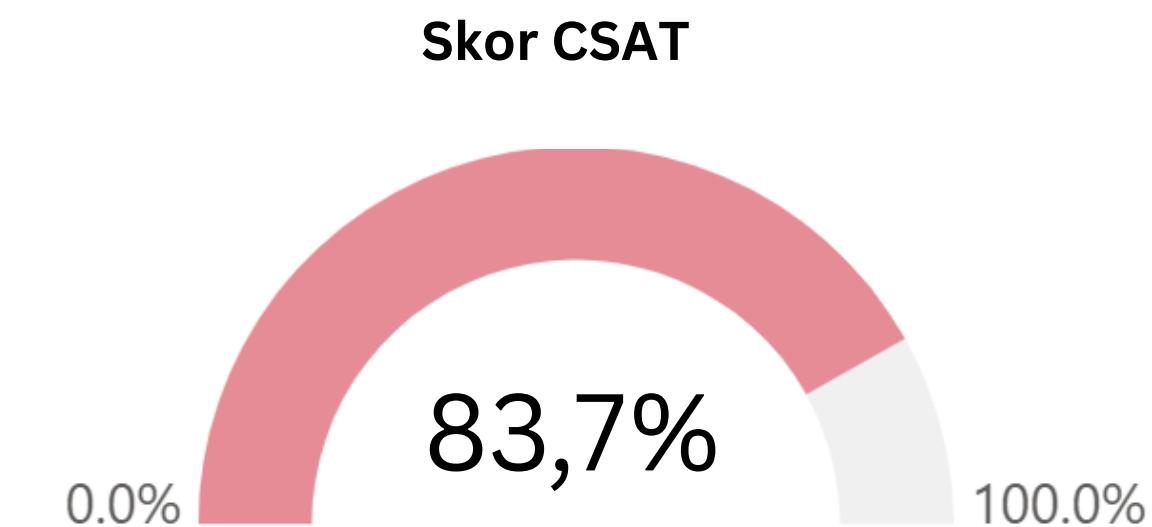
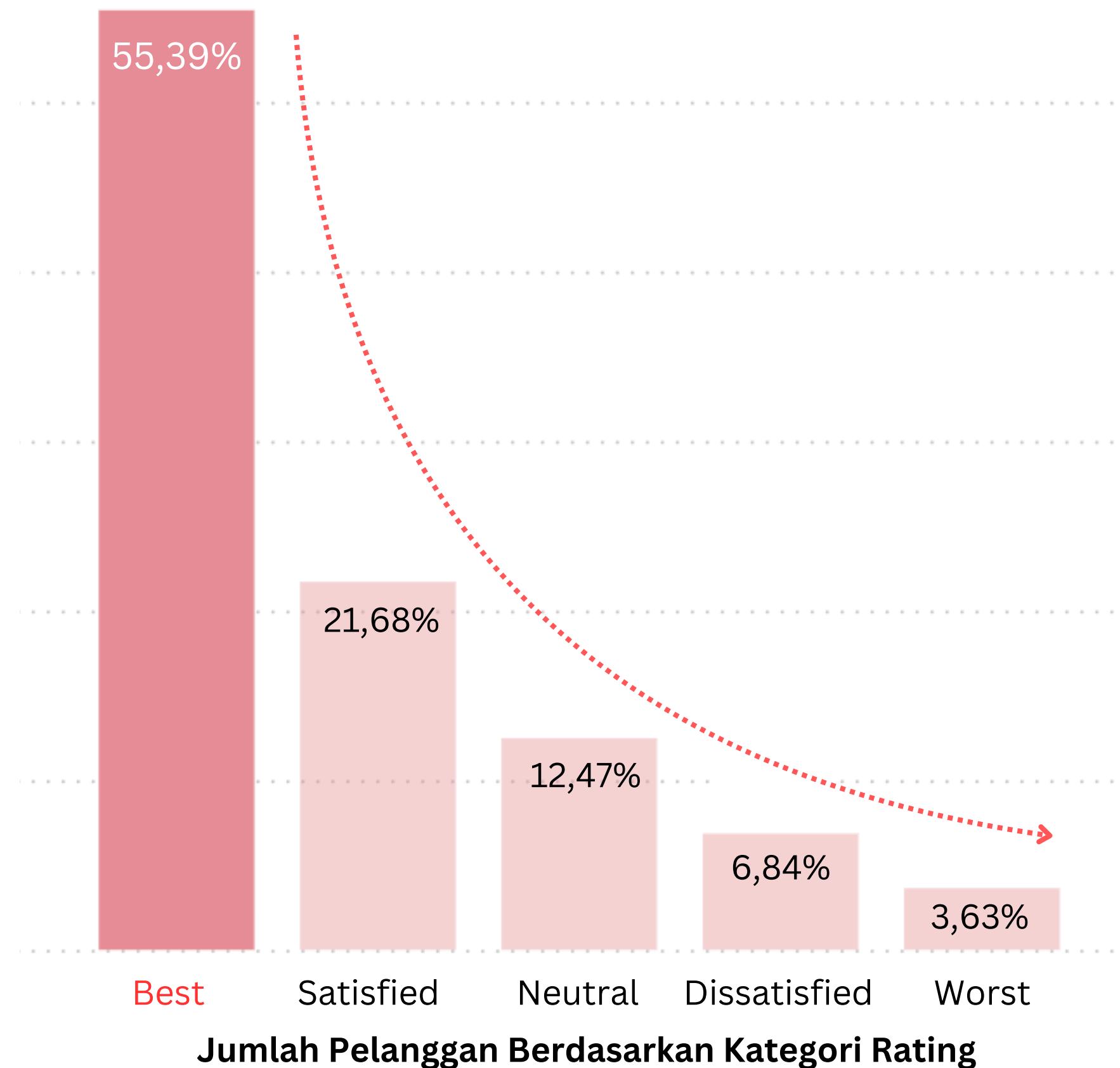


Customer Satisfaction

# Tingkat kepuasan pelanggan secara keseluruhan bernilai baik

Skor CSAT (Customer Satisfaction) diperoleh dengan menghitung nilai rating yang diberikan kepada produk. Hasilnya menunjukkan bahwa kepuasan pelanggan tergolong baik yang ditunjukkan dengan perolehan nilai 83,7%.

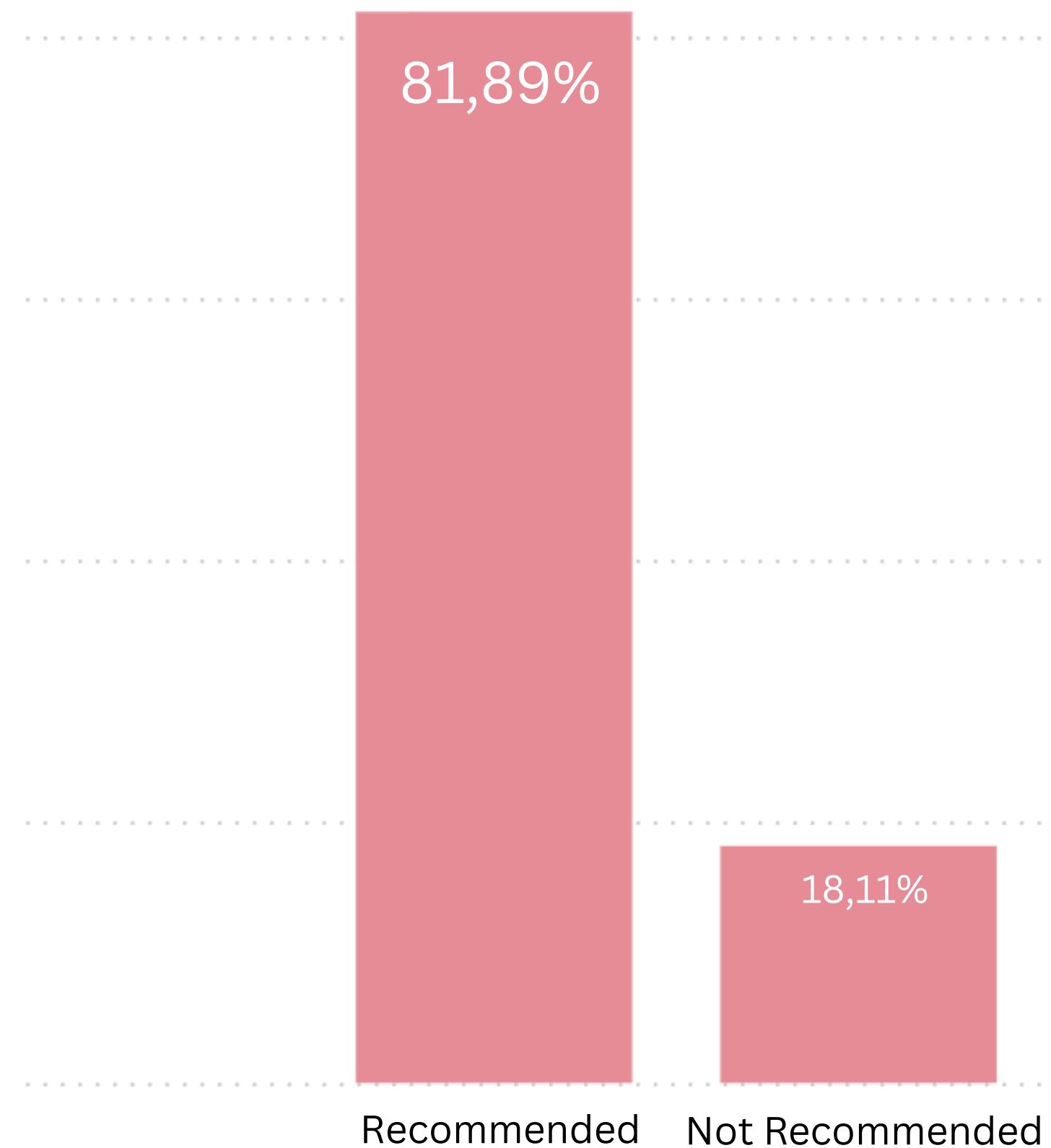
Hasil tersebut didominasi 55,39% oleh rating ‘Best’ yang merupakan tingkat tertinggi. Penurunan jumlah pelanggan sejalan dengan menurunnya tingkatan rating.



# Sebagian besar pelanggan merekomendasikan produk

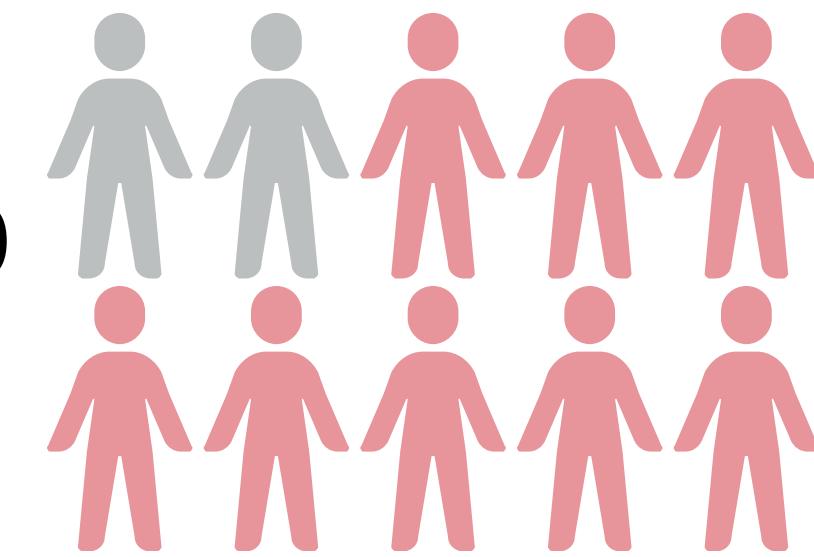
Pelanggan yang merekomendasikan produk tidak seluruhnya memberikan rating yang tinggi, terdapat sebagian kecil pelanggan yang tetap merekomendasikan produk dengan penilaian yang buruk.

Mereka suka, bahkan memuji produk. Nilai yang rendah muncul dengan alasan yang bervariasi, namun secara garis besar hal ini lantaran produk yang tiba tidak sesuai seperti yang diharapkan. Kebanyakan dari mereka berada di rentang usia antara 40-60 tahun.



8 dari 10

pelanggan yang  
melakukan review  
merekomendasikan  
produk



## Customer Sentiment

# Bagaimana sentimen yang tercermin dari review mereka?

Merupakan kabar yang baik bahwa review positif adalah sentimen yang paling banyak ditulis. Rentang usia 30-an menjadi usia dengan jumlah sentimen positif sekaligus negatif terbanyak.

Jumlah Reviewer

1000

500

0

Usia 35 tahun  
menulis review  
negatif terbanyak

20

39 tahun adalah usia dengan review  
positif terbanyak, yaitu berjumlah 1.039

40

60

80

100

Rentang Usia

Distribusi Pengulas Berdasarkan Usia

Berapa Total Persentase  
di Tiap Sentimen?

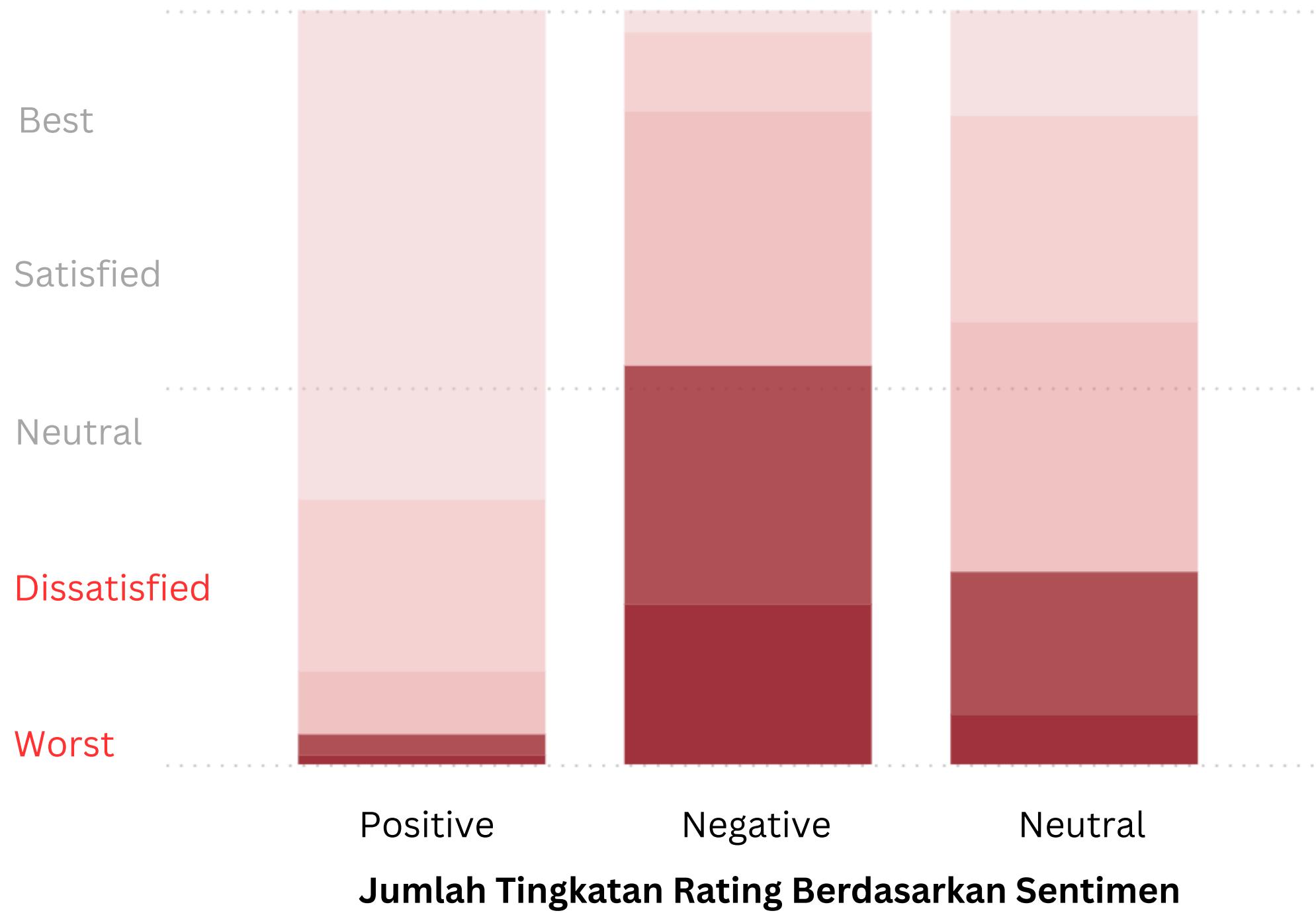
5,5% Neutral

Negative 10,8%

83,7%

Positive

99 tahun yang merupakan usia  
tertua, seluruhnya meninggalkan  
review yang positif



**Customer Sentiment**

## Review yang positif tidak menjamin produk telah sempurna

Level rating dan sentimen saling mempengaruhi.

Semakin tinggi penilaian, maka akan semakin banyak pula yang memberikan review yang positif. Meskipun demikian, review positif tidak luput dari kepuasan yang buruk dan sentimen negatif bukan berarti akan buruk sepenuhnya.

## Customer Sentiment

Setiap sentimen kompak memberikan lima kelas produk serupa yang masih memerlukan perbaikan

### Top 3 Clothing Id dengan penilaian yang buruk

Dresses	Knits	Blouses	Sweater	Pants
• 1078	• 862	• 829	• 936	• 1059
• 1094	• 868	• 850	• 927	• 1056
• 1081	• 867	• 836	• 941	• 1068

Kelas produk terbagi secara lebih mendetail dengan id clothing. Urutan tiga teratas dipilih untuk menentukan prioritas sehingga dapat fokus menyelesaikan terlebih dahulu masalah yang paling banyak dialami oleh pelanggan.

### Negatif

Dresses

Knits

Sweaters

Blouses

Pants

### Positif

Dresses

Knits

Blouses

Dresses

Blouses

Knits

Pa... Sw...

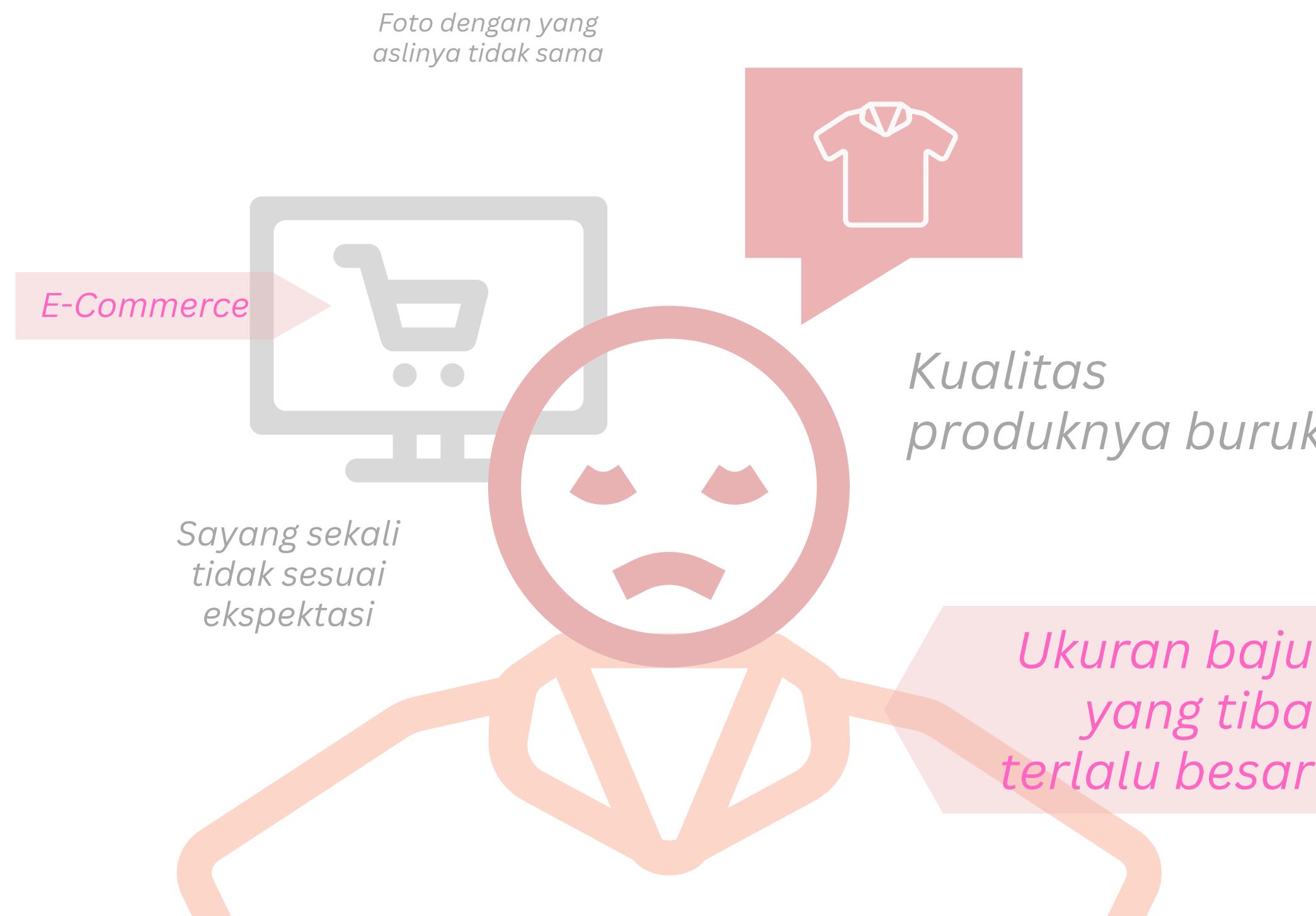
### Netral

Untuk selengkapnya, topik dan hal-hal yang sering dibicarakan bisa dilihat lebih jauh di bawah ini:

- [Appendix 4: Bigram & Word Cloud - Sentimen Positif](#)
- [Appendix 5: Bigram & Word Cloud - Sentimen Netral](#)
- [Appendix 6: Bigram & Word Cloud - Sentimen Negatif](#)

Customer Sentiment

## Hal-hal apa saja yang mereka keluhkan?



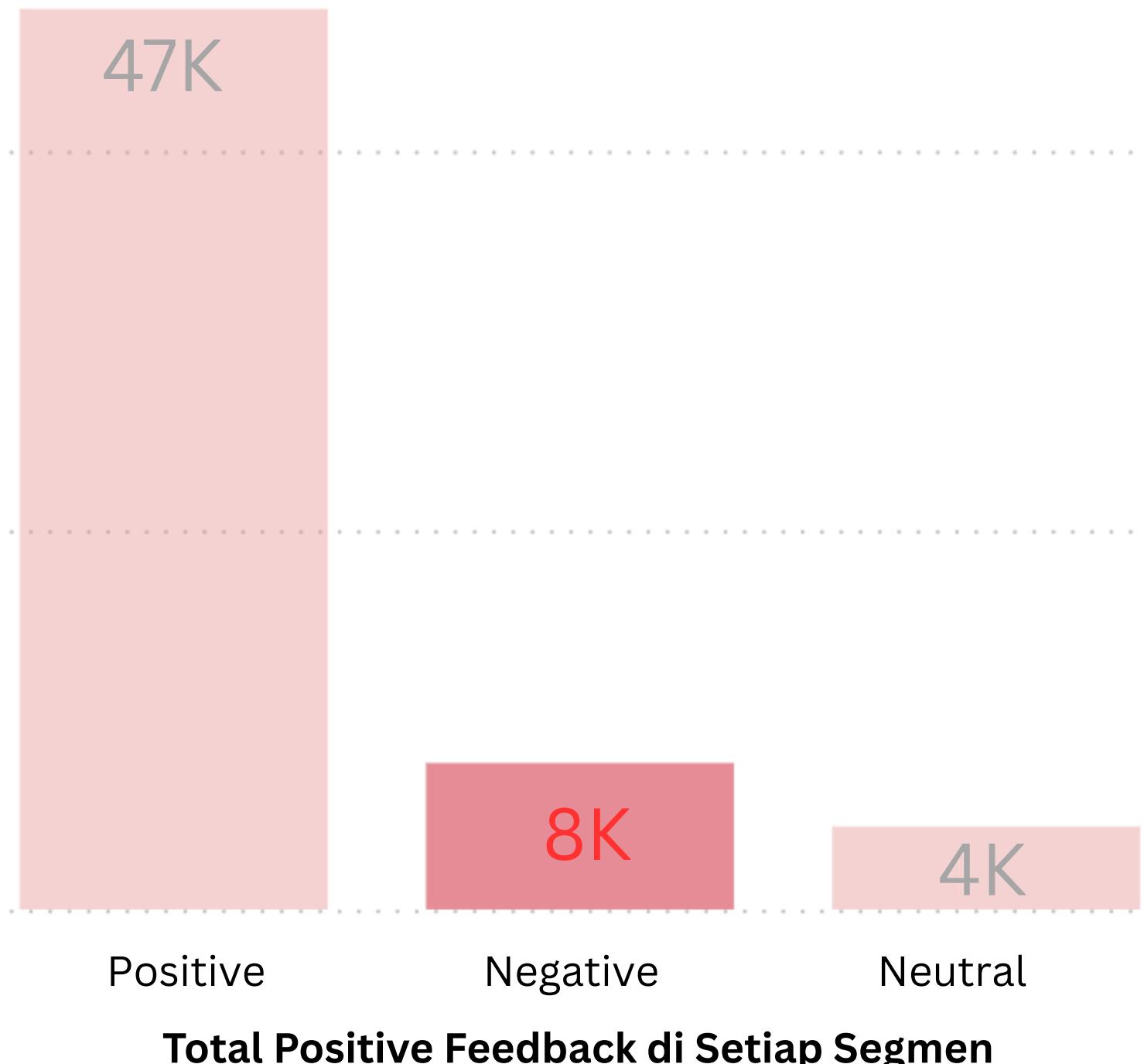
Kehadiran belanja dengan sistem online menawarkan sejumlah kentungan, salah satunya adalah memberikan kemudahan dalam berbelanja. Karena produk hanya tampil di sebuah layar, keterbatasan ini menimbulkan masalah dimana produk yang digambarkan kemungkinan tidak sesuai dengan rupa aslinya dan hal ini yang dikeluhkan oleh banyak pelanggan. Kualitas produk juga yang terkait dengan aspek fisik seperti ukuran, warna, dan material masih menjadi PR yang harus diperbaiki. Produk yang tidak memuaskan alhasil membuat beberapa pelanggan tidak segan untuk melakukan pengembalian.

**Nilai Positive  
Feedback Tertinggi  
di Tiap Sentimen**

122  
user

65  
user

108  
user



**Customer Sentiment**  
**Jumlah user yang  
merasa terbantu  
dengan review negatif  
relatif sedikit, namun  
angka tertingginya  
lebih dari setengah  
review positif**

Komentar positif sangat banyak membantu user yang lain. Jumlahnya signifikan lebih tinggi, mendominasi 79,7%. Namun ulasan yang negatif tetap harus menjadi perhatian karena nilai maksimum tergolong cukup tinggi. Mengetahui positif feedback akan sangat berguna karena menggambarkan beberapa kemungkinan, yaitu user yang membaca merasa ulasan tersebut bermanfaat atau setuju karena mengalami hal yang serupa.

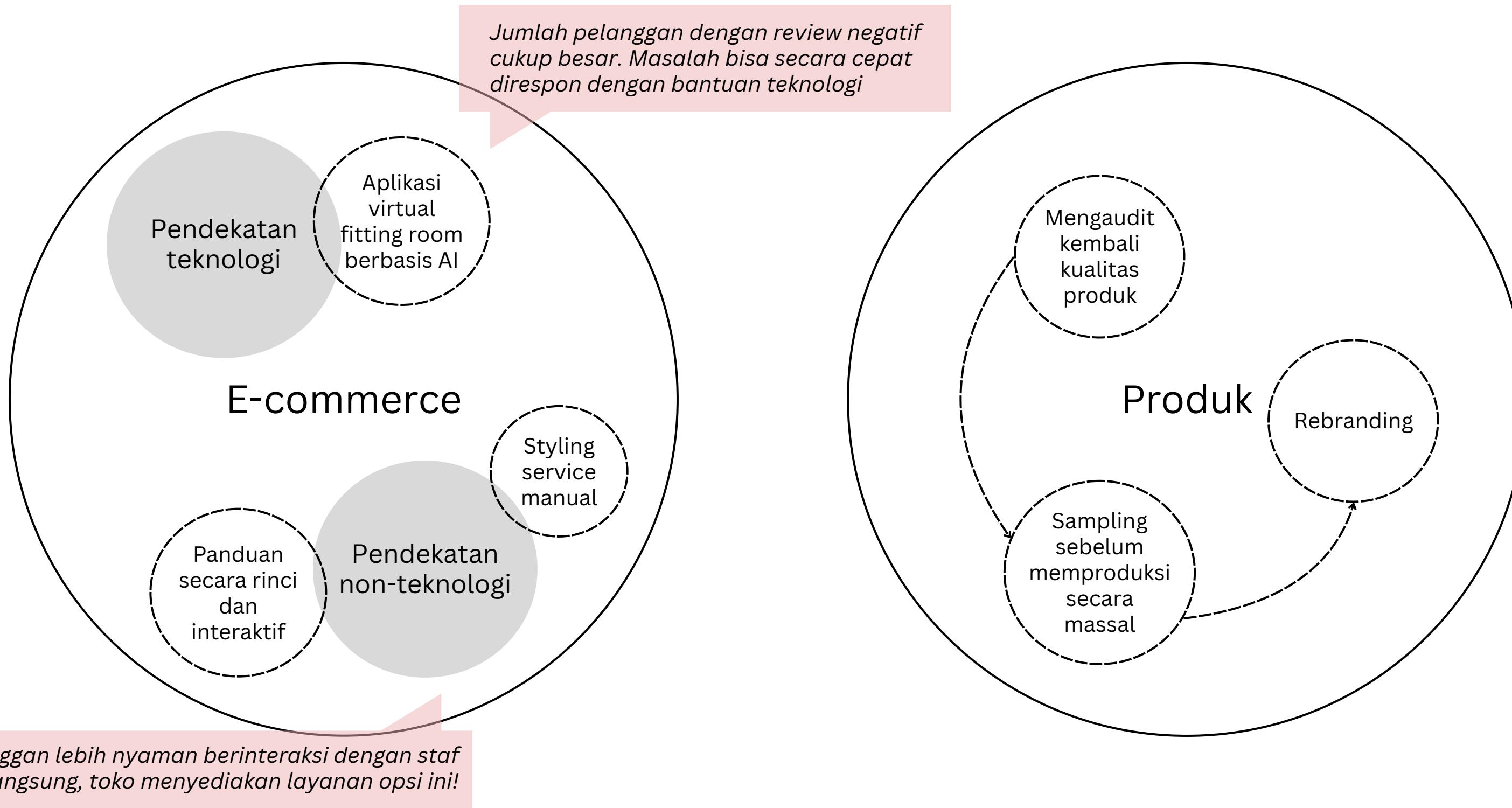
## Knits - Tops (65 Positive Feedback)

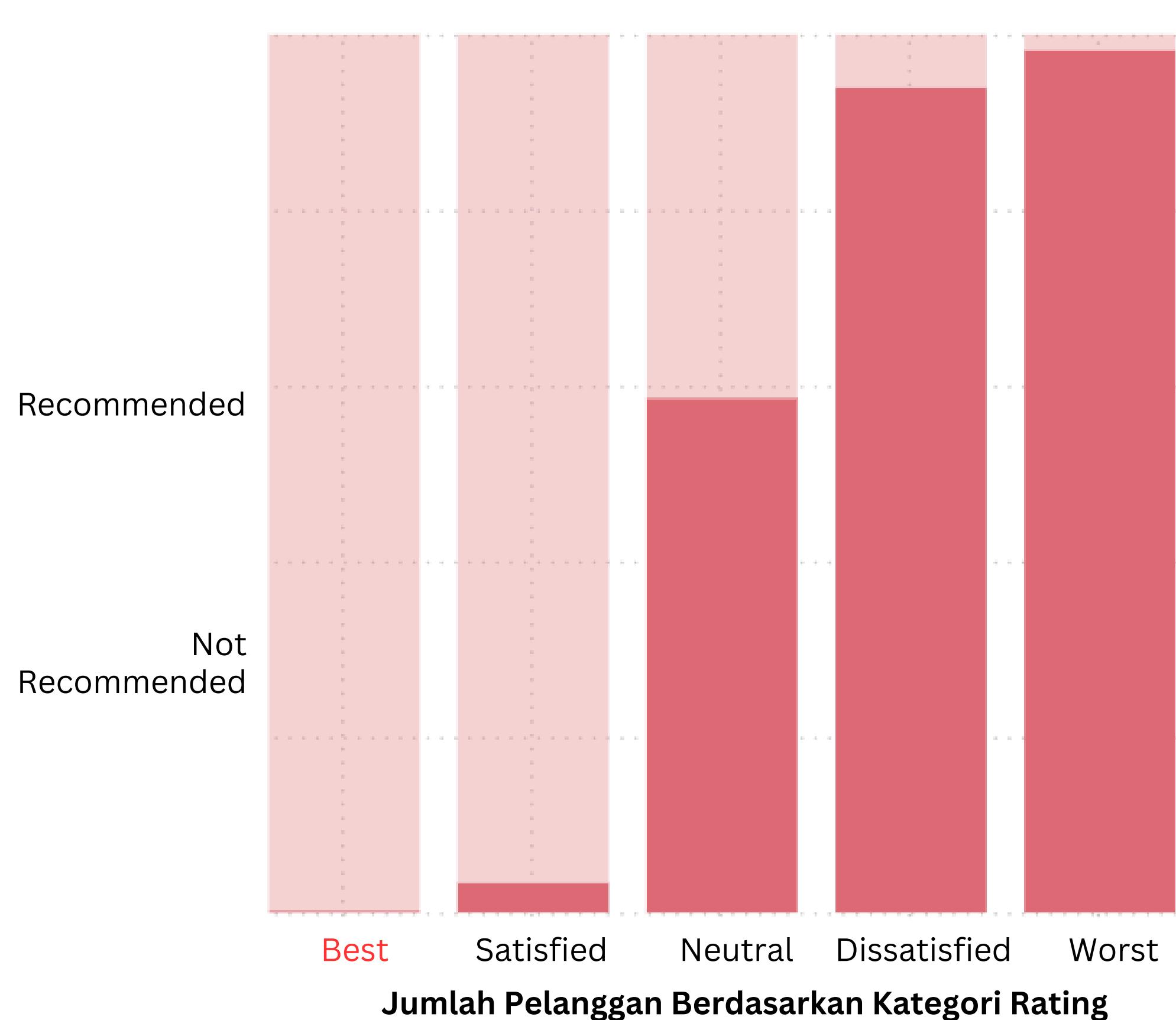
*"What drew me to this shirt was the beautiful silver and gold embroidery on the front. the shirt was folded at the store, which makes sense because if it were hung it up, fewer people would try it on. this shirt is sheer from the front and more sheer in the back.*

**this is what bugs me about so many "clothes" that are made these days: they are so cheaply made, so thin with cloth, that you are forced to layer with a camisole or something else in order to wear it in public. i can't wear sheer clothes"**

## Customer Sentiment

# Mulai berbaikan, merealisasikan ekspektasi pelanggan yang belum sempat terwujud





Pelanggan adalah sosok yang detail, kritis, dan peduli. Mereka cenderung merekomendasikan produk yang memuaskan

Kualitas barang yang baik dengan kepuasan tertinggi tidak hanya memperkuat branding toko melainkan dapat memperluas jangkauan produk semakin dikenal dan dibeli.

# Appendix 1 : Dokumentasi Proses Analisis Data

## **Link Google Collab:**

[Final Project DA - EDA](#)

[Final Project DA - Tokenize Text](#)

## **Link Dashboard:**

[Final Project DA - Dashboard Power BI](#)

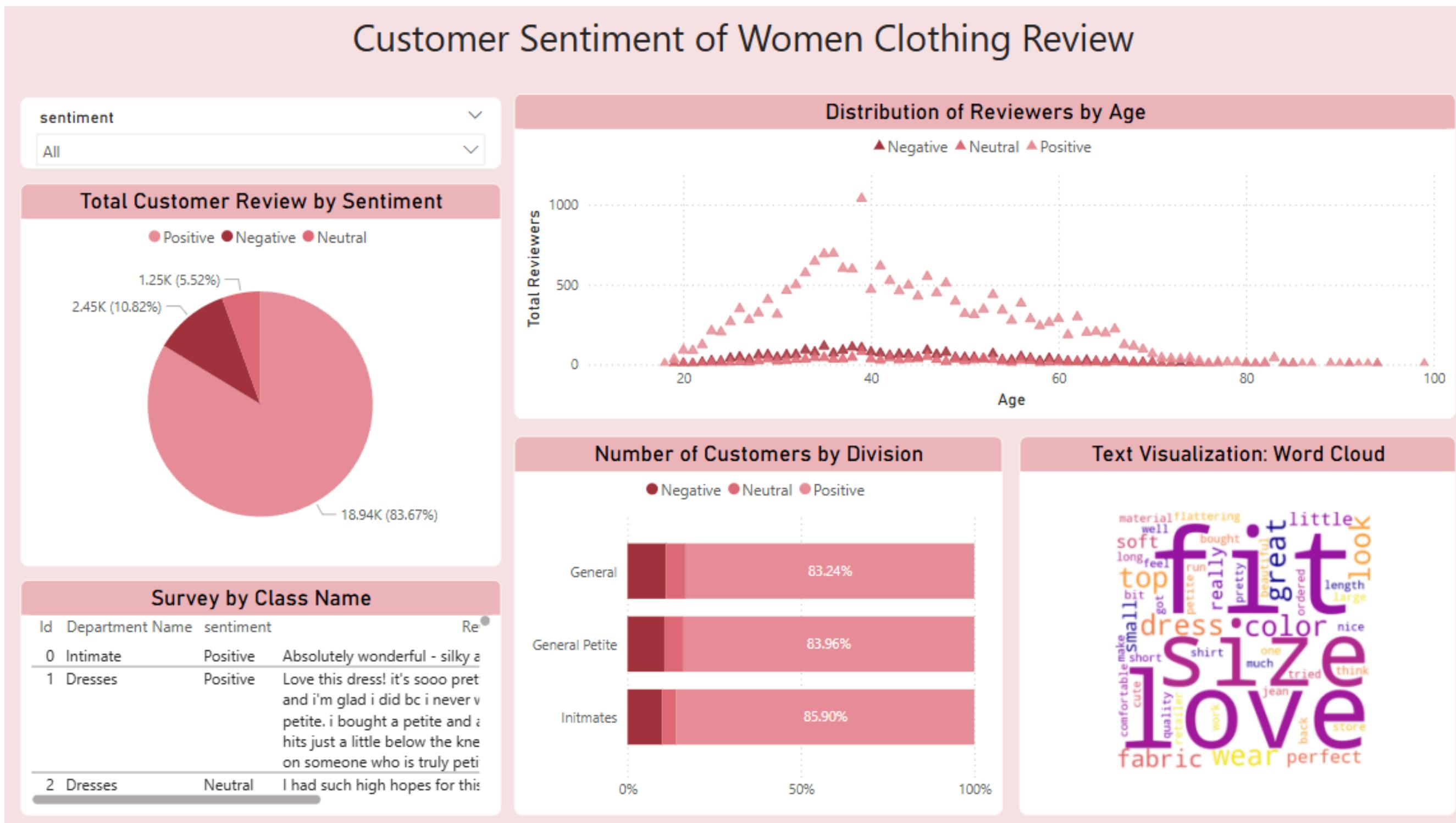
[Kembali ke halaman sebelumnya](#)

# Appendix 2: Dashboard Customer Satisfaction



[Kembali ke halaman sebelumnya](#)

# Appendix 3: Dashboard Customer Sentiment



[Kembali ke halaman sebelumnya](#)

# Appendix 4: Bigram & Word Cloud - Sentimen Positif

Kedua visualisasi menunjukkan kata-kata (word cloud) dan kombinasi kata (bigram) yang sering ditulis sebagai review pada lima kelas produk teratas dengan penilaian yang buruk (rating ‘dissatisfied’ dan ‘worst’) di review positif. Hal tersebut dapat memberikan gambaran secara umum dan konteks secara spesifik mengenai topik yang sering dibicarakan.

50 Bigram Teratas pada Sentimen Positif



50 Kata (Word Cloud) Teratas pada Sentimen Positif



[Kembali ke halaman sebelumnya](#)

# Appendix 5: Bigram & Word Cloud - Sentimen Netral

Kedua visualisasi menunjukkan kata-kata (word cloud) dan kombinasi kata (bigram) yang sering ditulis sebagai review pada lima kelas produk teratas dengan penilaian yang buruk (rating ‘dissatisfied’ dan ‘worst’) di review netral. Hal tersebut dapat memberikan gambaran secara umum untuk mengevaluasi aspek-aspek yang masih memerlukan perbaikan.

50 Bigram Teratas pada Sentimen Netral



50 Kata (Word Cloud) Teratas pada Sentimen Netral



[Kembali ke halaman sebelumnya](#)

# Appendix 6: Bigram & Word Cloud - Sentimen Negatif

Kedua visualisasi menunjukkan kata-kata (word cloud) dan kombinasi kata (bigram) yang sering disebutkan pada review pada lima kelas produk dengan sentimen negatif tertinggi di seluruh tingkatan rating.

**50 Bigram Teratas pada Sentimen Negatif**



**50 Kata (Word Cloud) Teratas pada Sentimen Negatif**

