

PROJECT PRESENTATION


DATA SCIENCE PORTFOLIO

Hafizh Maulana

DATA ENTHUSIAST AND ACCOUNTANT

 hafizhmaul18.github.io/porto


 [hafizhmaul18](https://www.linkedin.com/in/hafizhmaul18)

 hafizhmaul18@gmail.com

My Background



Data Enthusiast with 2 years of experience as a Finance Accountant. Recently transitioned into data science through an internship at BukitVista and completion of a Data Science Bootcamp. With strong analytical skill, Actively seeking a Data Scientist role to apply business insights to impactful projects.



Work Experienced

01.

Finance Accountant — Yayasan Masjid Raya Rahmatan lil Alamin (Dec 2022 – Present)

Managed and organized the foundation's financial database to support fundraising programs, Delivered strategic financial insights through monthly and annual reports, and effectively segmented congregational groups using CRM tools to optimize communication and donor engagement.

02.

Class Instructor — Robotics Education Center (Jun – Dec 2022)

Taught visual programming through the STEAM-based Scratch platform, while developing lesson plans and evaluating progress using Google Classroom to enhance engagement and digital literacy.

03.

Data Science Intern — Bukit Vista (Jan - Mar 2024)

Collecting, organizing, and optimizing data from website sources to construct a valuable structured knowledge information for companies in decision making.

Education & Certification

Education

01. **Dibimbing Bootcamp**
Data Science (July – Des 24)
02. **Institut PTIQ Jakarta**
Bachelor of Social Da'wah (2018 – 2022)

Certification

01. **Google Data Analytics, Google**
02. **Level 3 Diploma in Computing, NCC Education**
03. **Python and Problem Solving Basic, Hackerrank**
04. **Data Analytics with Excel, Halotech Academy**
05. **Python Dasar dan Lanjutan, Skilvul**



Top Project Portfolio

01.

Startup Failure Analysis

Conducted an in-depth analysis of startup failure patterns using machine learning models

02.

Flight Customer Clustering

Developed a customer segmentation model for airline passengers using clustering techniques

03.

Tokopedia Review Analysis

Topic and Sentiment analysis using LDA and BERT

More Portfolio : [hafizhmaul18.github.io/porto](https://github.com/hafizhmaul18/porto)

TOKOPEDIA REVIEW ANALYSIS

Understanding customer feedback through NLP & topic modeling

“Sepatu Kanky”
ONE OF THE BIGGEST LOCAL SHOES

Business Question?

Main Question!

what people are saying about
our brand?

Deep Dive Question

01. Which product get the most positive sentiment?
Which features are being praised the most?
02. Is there a recurring sentiment trend related to
delivery time or packaging quality?
03. How does sentiment for our product compare to
competitors in the same category?
04. What are the most common concerns from 1-star
reviewers?

Objectives



Analyze brand sentiment from customer reviews to understand overall public perception.



Identify top-performing products based on positive sentiment and features praised.



Spot recurring issues related to operations such as delivery delays or poor packaging.



Compare sentiment between our products and competitor products in the same category

Pre-Processing



Scrapping Data

Get real data from tokopedia using Selenium and BeautifulSoup4

Cleaning Text

Clean the text from symbol and convert every word into lowercase

Add Stopwords

import stopwords from NLTK and update with ours to make it better.

Create Tokens

Split every review and remove every word that in the stopwords.

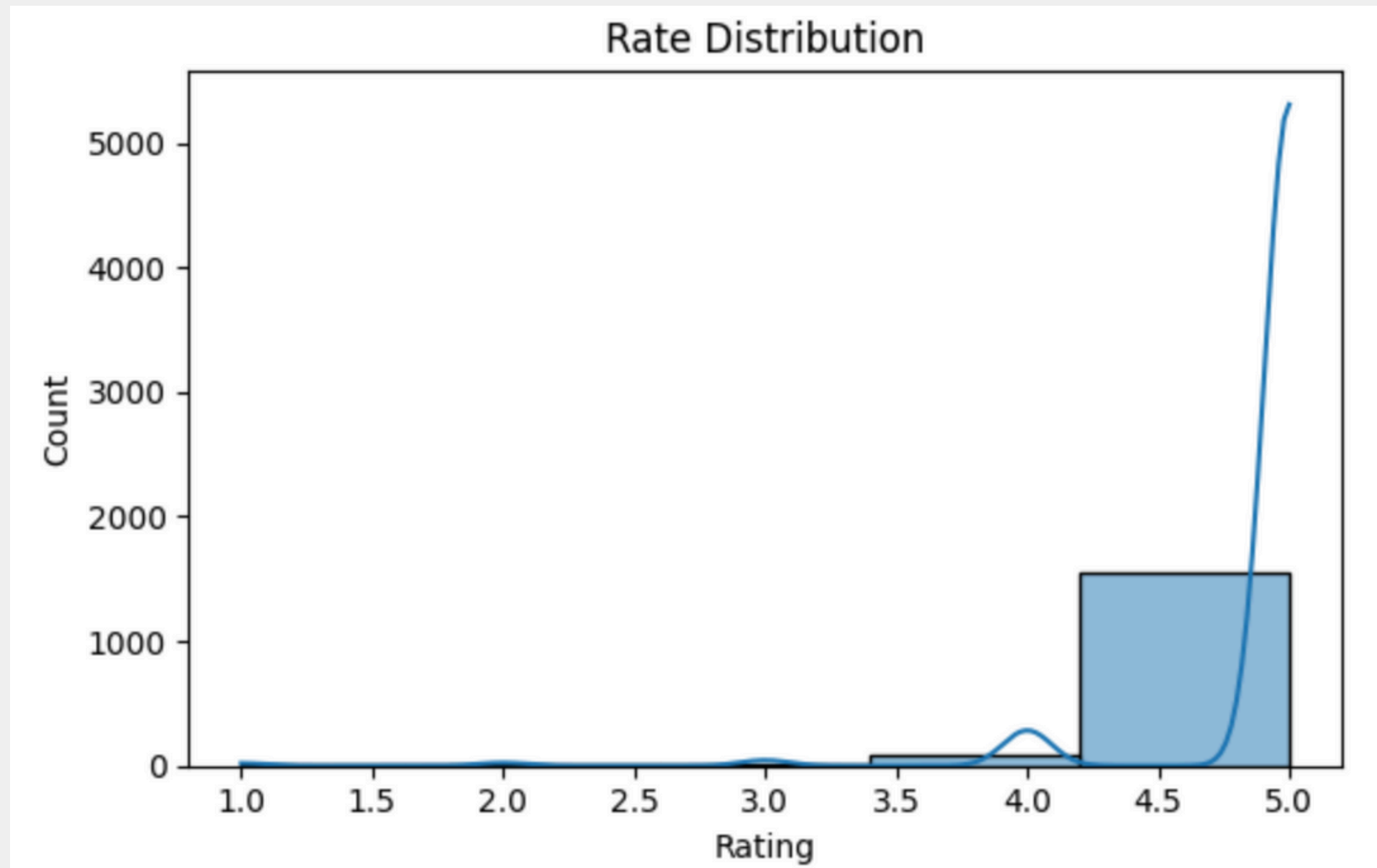
Data Information

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1871 entries, 0 to 1870
Data columns (total 3 columns):
#   Column   Non-Null Count  Dtype
---  -
0   Produk   1871 non-null   object
1   Rating   1871 non-null   int64
2   Ulasan   1871 non-null   object
dtypes: int64(1), object(2)
memory usage: 44.0+ KB
```

- This Dataset only have 3 columns:
Produk, Rating, Ulasan.
- with total **1871 Rows.**
- **Clear from Null**
- **Have 66 Duplicate values** (I decided to Remove it)



Rate Distribution

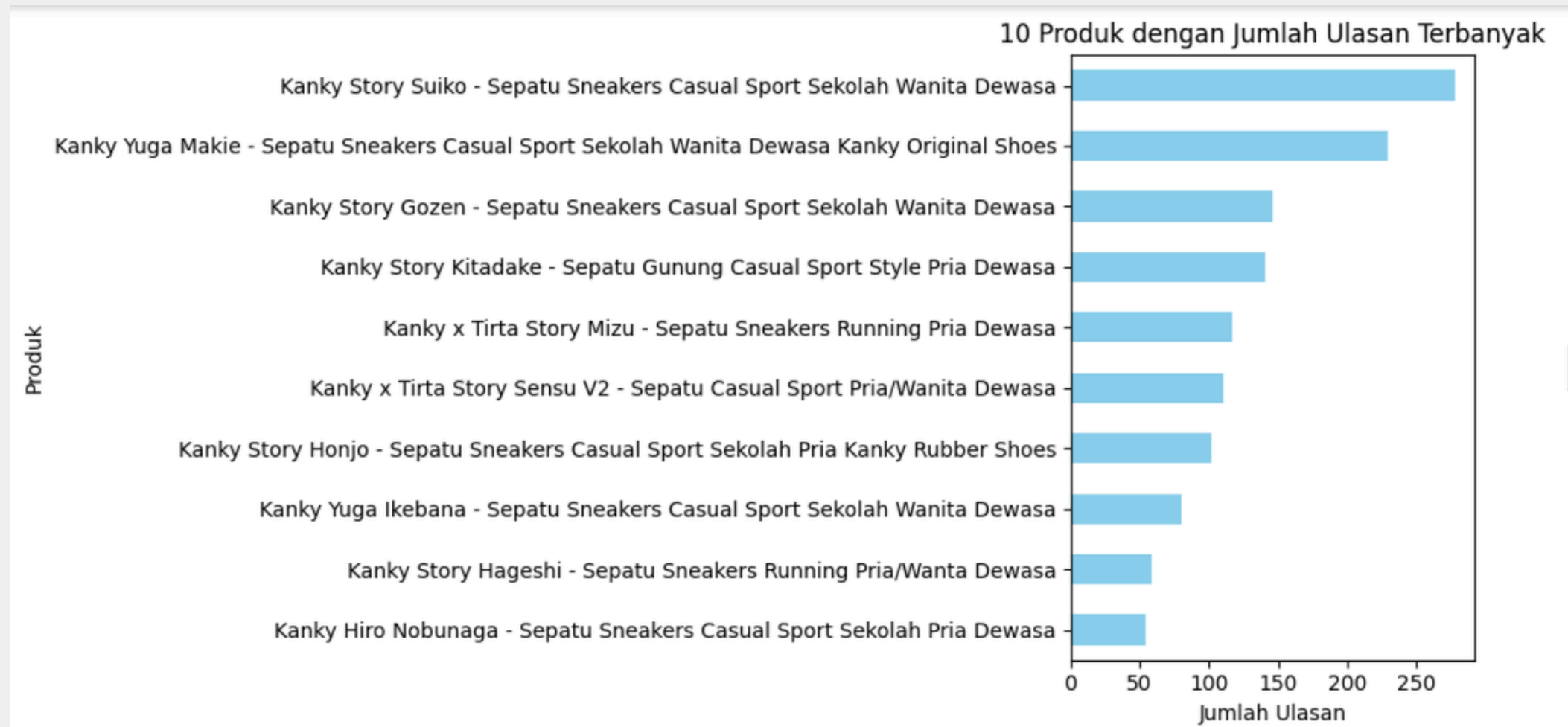


A sharp spike at 5.0, indicating this is the most common rating (almost 2000 reviews).

Most of the customer has already satisfied with this seller.



Rate Distribution



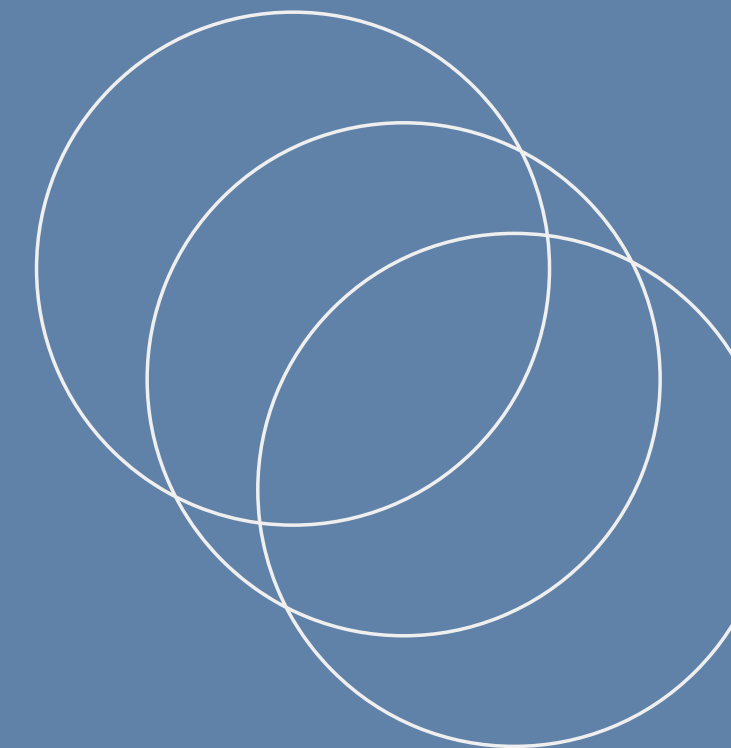
Top 3 most reviewed product :

- 1.Kanky Story Suiko
- 2.Kanky Yuga Makie
- 3.Kanky Story Gozen

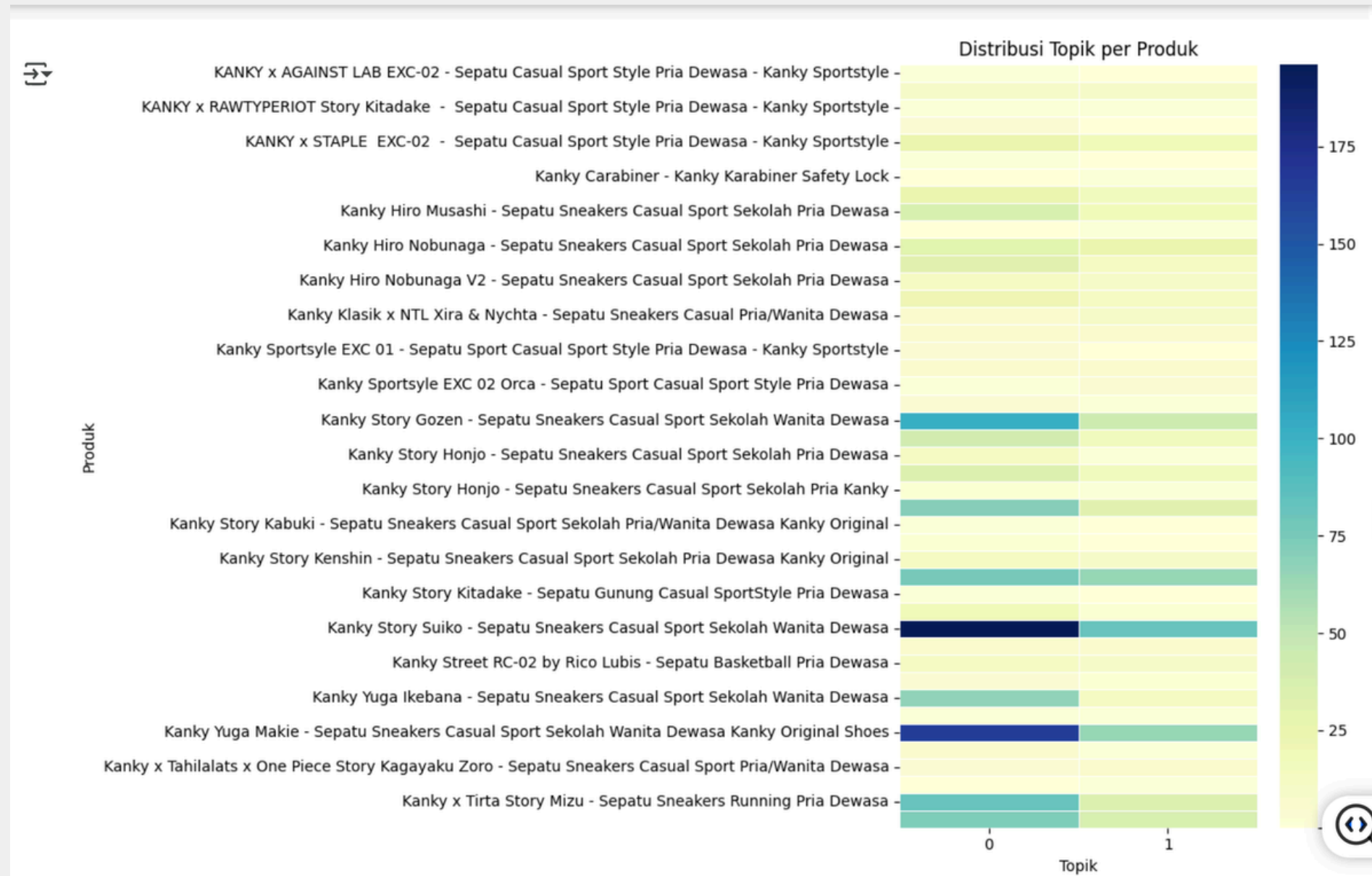
count	
Sentimen	
positive	239
neutral	9
negative	4

Kanky Story Suiko Review

Most Positive Sentiment product



Produk	Sentimen		
	negative	neutral	positive
Kanky Story Suiko - Sepatu Sneakers Casual Sport Sekolah Wanita Dewasa	4	9	239
Kanky Yuga Makie - Sepatu Sneakers Casual Sport Sekolah Wanita Dewasa Kanky Original Shoes	2	4	209
Kanky Story Gozen - Sepatu Sneakers Casual Sport Sekolah Wanita Dewasa	0	4	129
Kanky Story Kitadake - Sepatu Gunung Casual Sport Style Pria Dewasa	4	4	120
Kanky x Tirta Story Mizu - Sepatu Sneakers Running Pria Dewasa	0	5	103



Topic Distribution by Product

- Kanky Story Suiko & Yuga Makie has the most common topic
- most of the product have yellow color, it means the topic is spreading.



Hot Topic

To Evaluate the Product



From all Review
bagus, mantap, keren, nyaman, bentuk,
ukuran, size, sesuai, empuk, ringan, kanku

From the most reviewed product
bagus, ringan, nyaman, sesuai,
warna, empuk, bentuk, ukuran
→ could be good for the next shoes
to have this standard.

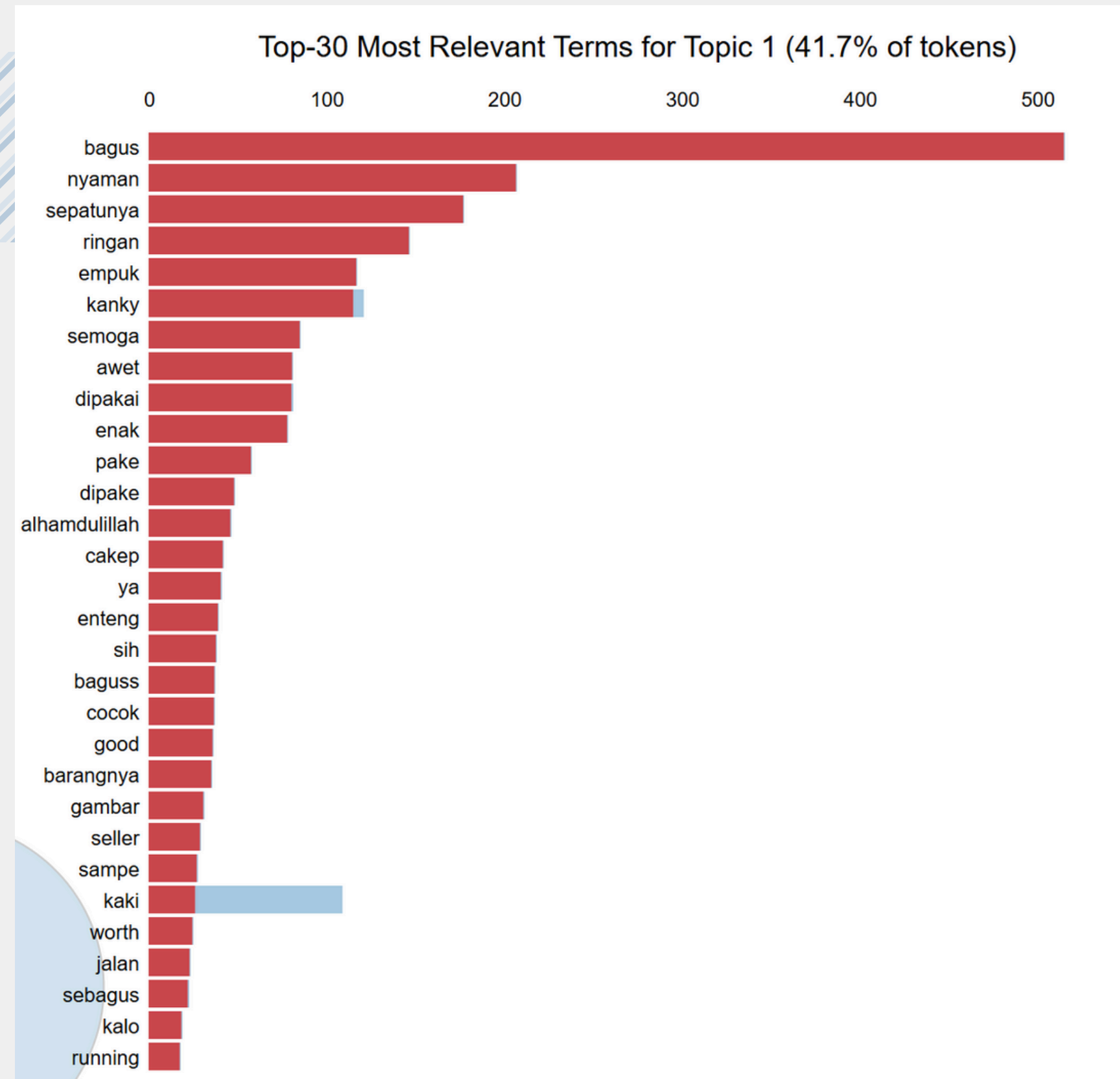


LDA from topic 1

Displayed using pyLDAvis

■ Topic 1 (41.7% of tokens) – Product Quality, Comfort and Brand

- bagus, nyaman, empuk, ringan, awet, sesuai and enak.
- Indicates strong satisfaction with the product's physical attributes and performance.
- Kanky is Mentioned! Our brand is in our costumer's mind

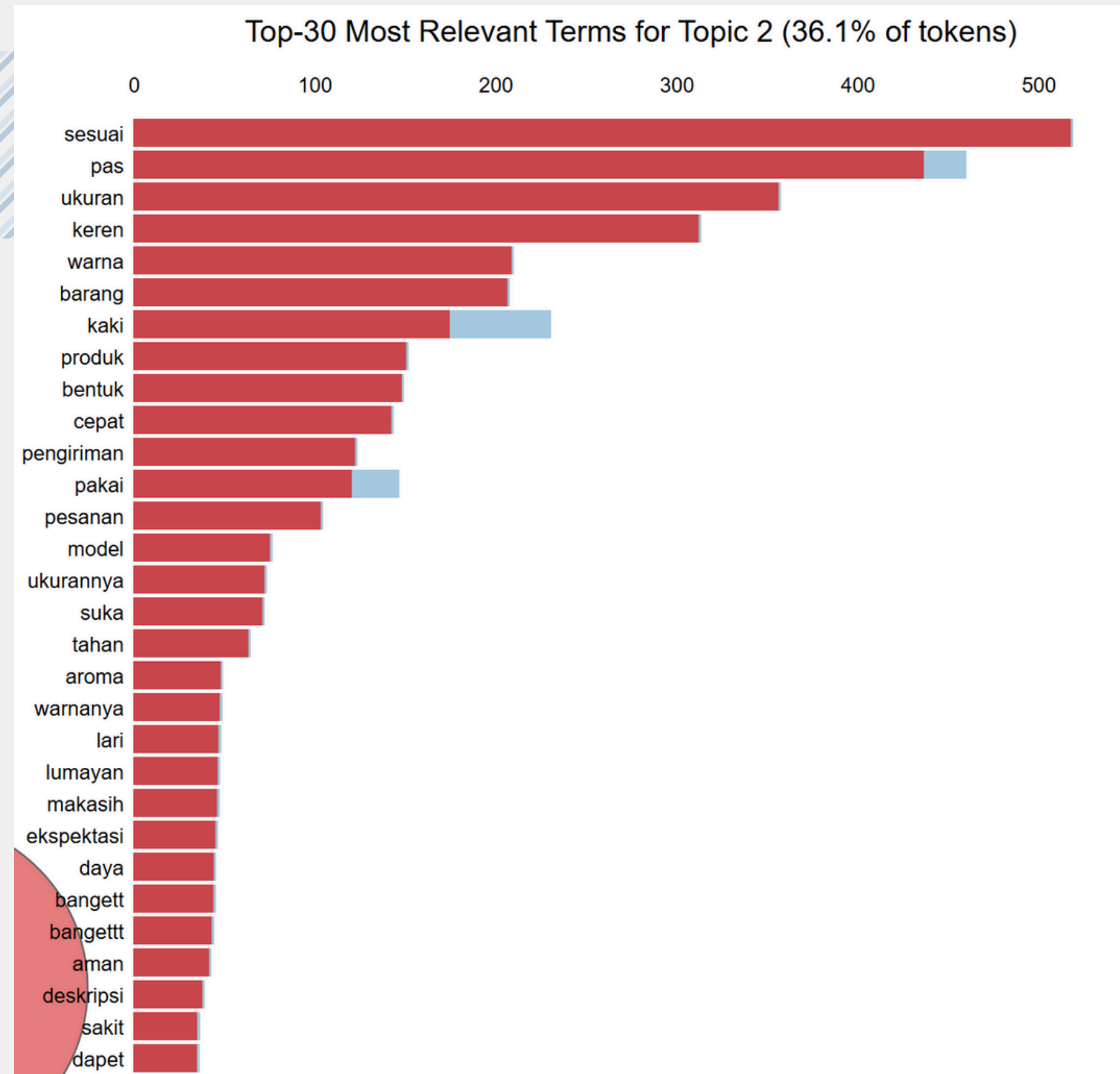




LDA from topic 2

■ Topic 2 (36% of tokens) – Fit, Size, Delivery, Local product.

- Sesuai, pas, ukuran, cepat, pengiriman, mantap, and bentuk.
- Suggests that many customers are concerned with whether the shoes fit well and arrive on time.

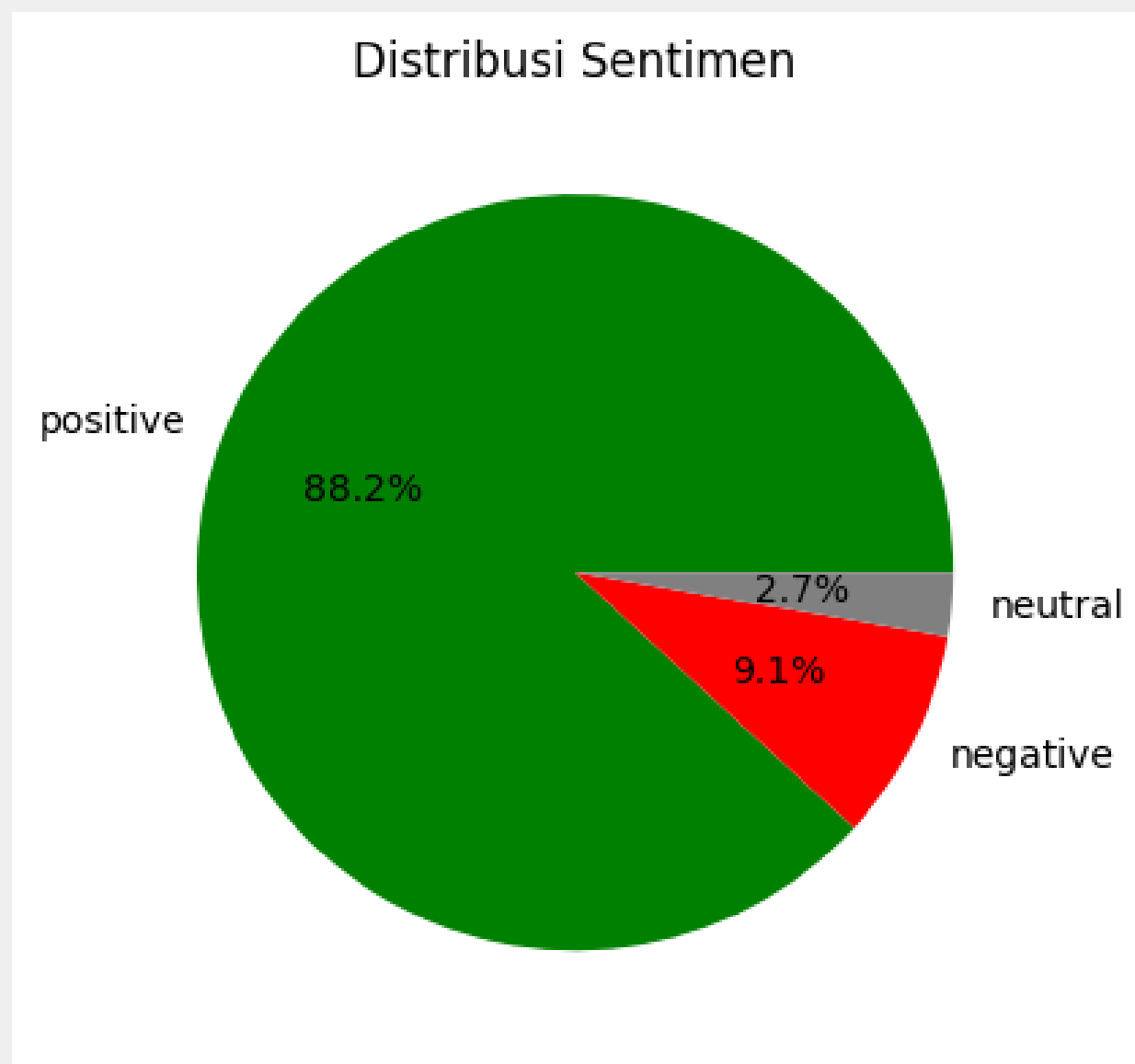


Competitor Analysis

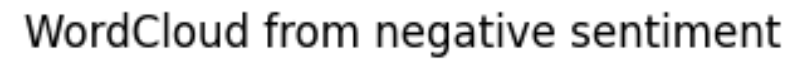
Dimension	Kanky	Ventela	Compass
Top Word	bagus	bagus	bagus
Brand Mentioned	Kanky	Ventela	Compass
Autheticity	-	original, ori, real	original, lokal
Delivery	pengiriman, cepat (topic 2)	pengiriman, kurir (topic 2)	pengiriman, paket, respon (topic 2)
Fit/Comfort	nyaman, empuk, ringan, enteng	empuk, cocok	Nyaman, sesuai
Other's Good Topic	Cakep, Awet	Keren, Berkualitas	Kualitas



Sentiment Distribution



- ✓ High Positive Perception, It validates the themes from Topics 1 and 2: quality, comfort, fit, and fast delivery are valid.
- ✓ The 9.1% negative feedback is small but not negligible.
- 😐 Low Neutral Sentiment, it shows high engagement and emotional response to the brand.



PROJECT PRESENTATION

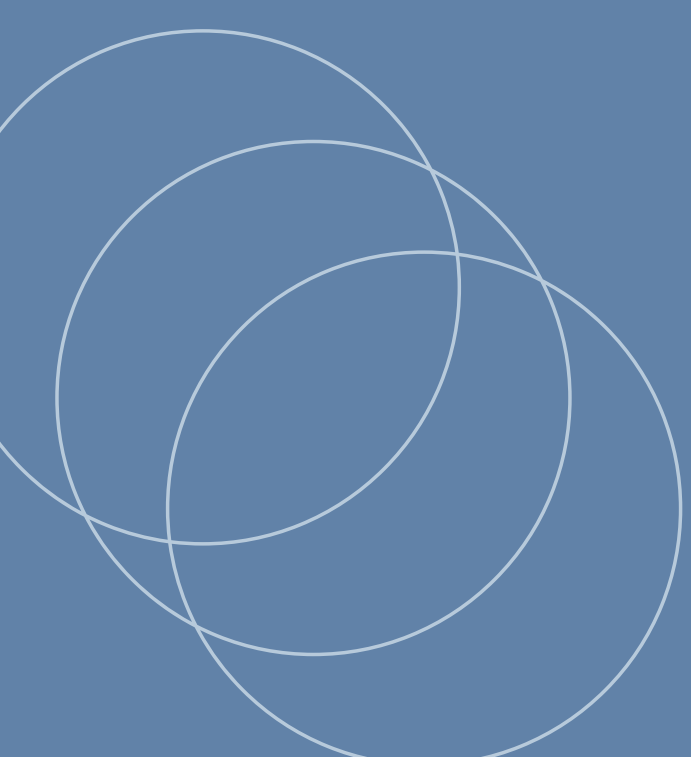


Conclusion and Recommendation

Use actual customer phrasing in ads
Like "Nyaman dipake"

Box and Packaging the reason behind 1
star. QC must give much attention in it

Delivery speed is part of brand
perception.



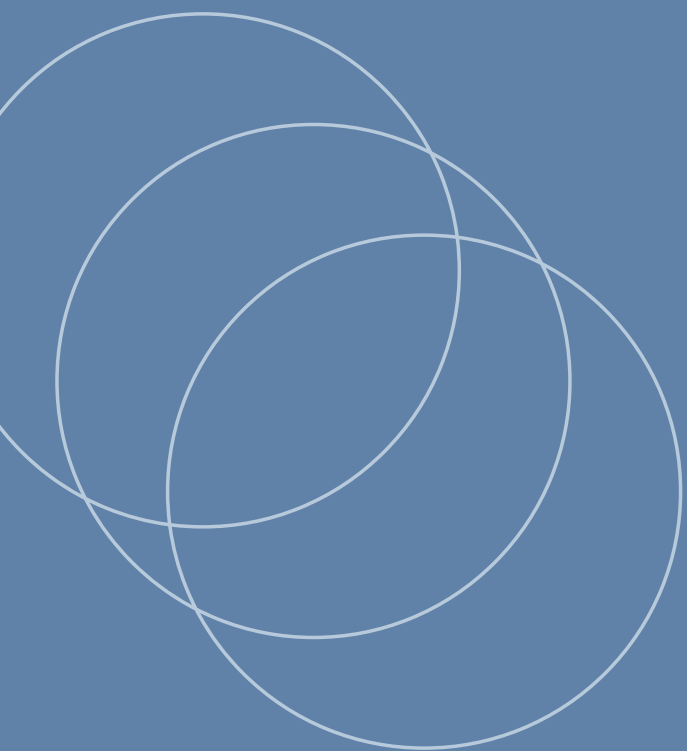


Conclusion and Recommendation

Overall, our brand has already satisfied our customer. We have to strengthen our good sentiment to be our branding.

Differentiation to Competitor : Nyaman, Ringan and Awet.

Build our Authenticity by adding Hologram/ any sign.





Next Step

01.

Dashboard

Develop interactive dashboards with Streamlit or Tableau.

02.

Classified


Classify the product : Collab, Running, Woman




Thank You



 hafizhmaul18.github.io/porto

 [hafizhmaul18](#)

 hafizhmaul18@gmail.com