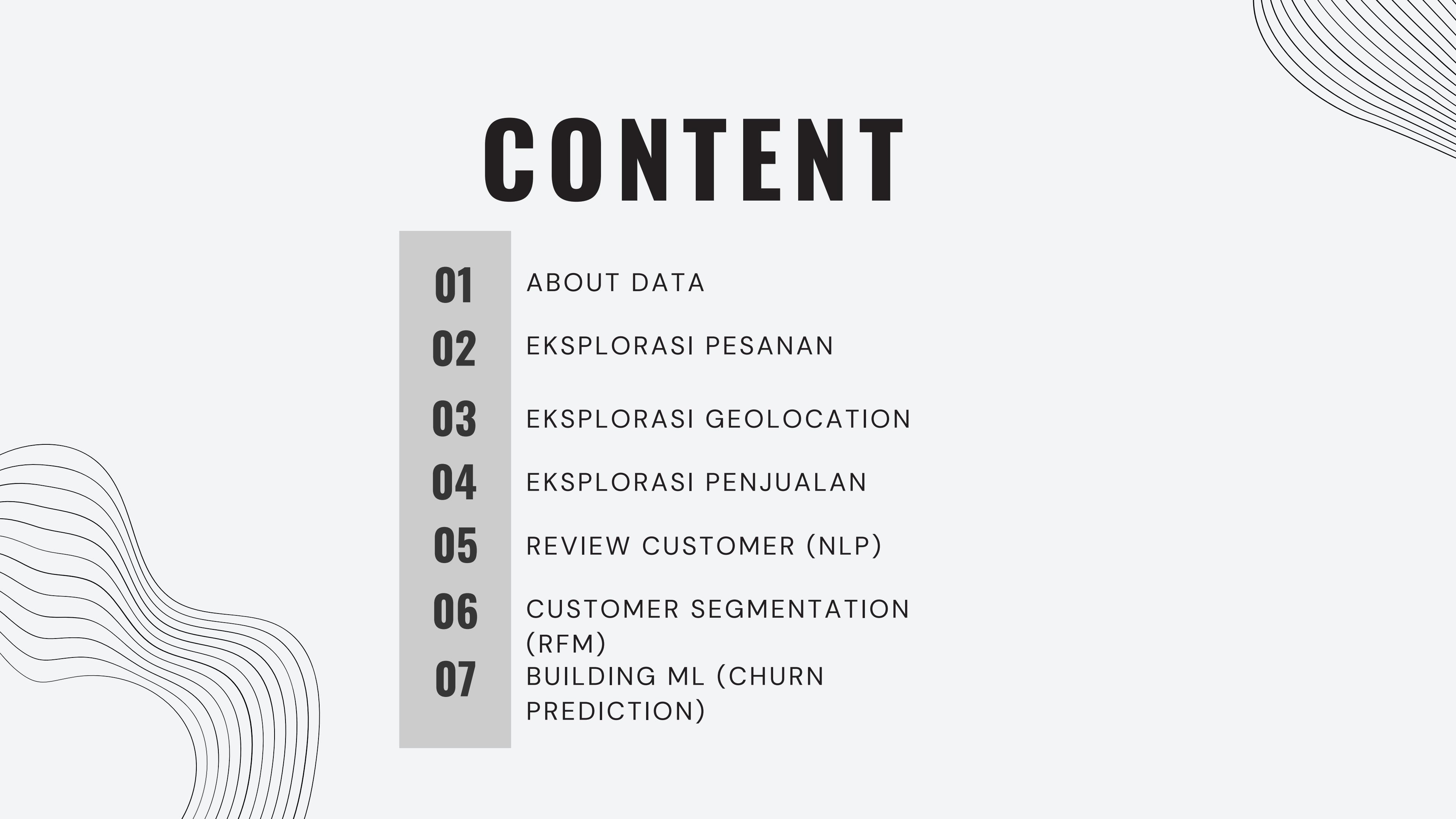


**PERQARA  
ASSESSMENT  
DATA SCIENTIST**

**RIJAL AKHDAN KHAIRULAH**

# CONTENT

- 
- 01** ABOUT DATA
  - 02** EKSPLORASI PESANAN
  - 03** EKSPLORASI GEOLOCATION
  - 04** EKSPLORASI PENJUALAN
  - 05** REVIEW CUSTOMER (NLP)
  - 06** CUSTOMER SEGMENTATION  
(RFM)
  - 07** BUILDING ML (CHURN  
PREDICTION)

# ABOUT DATA



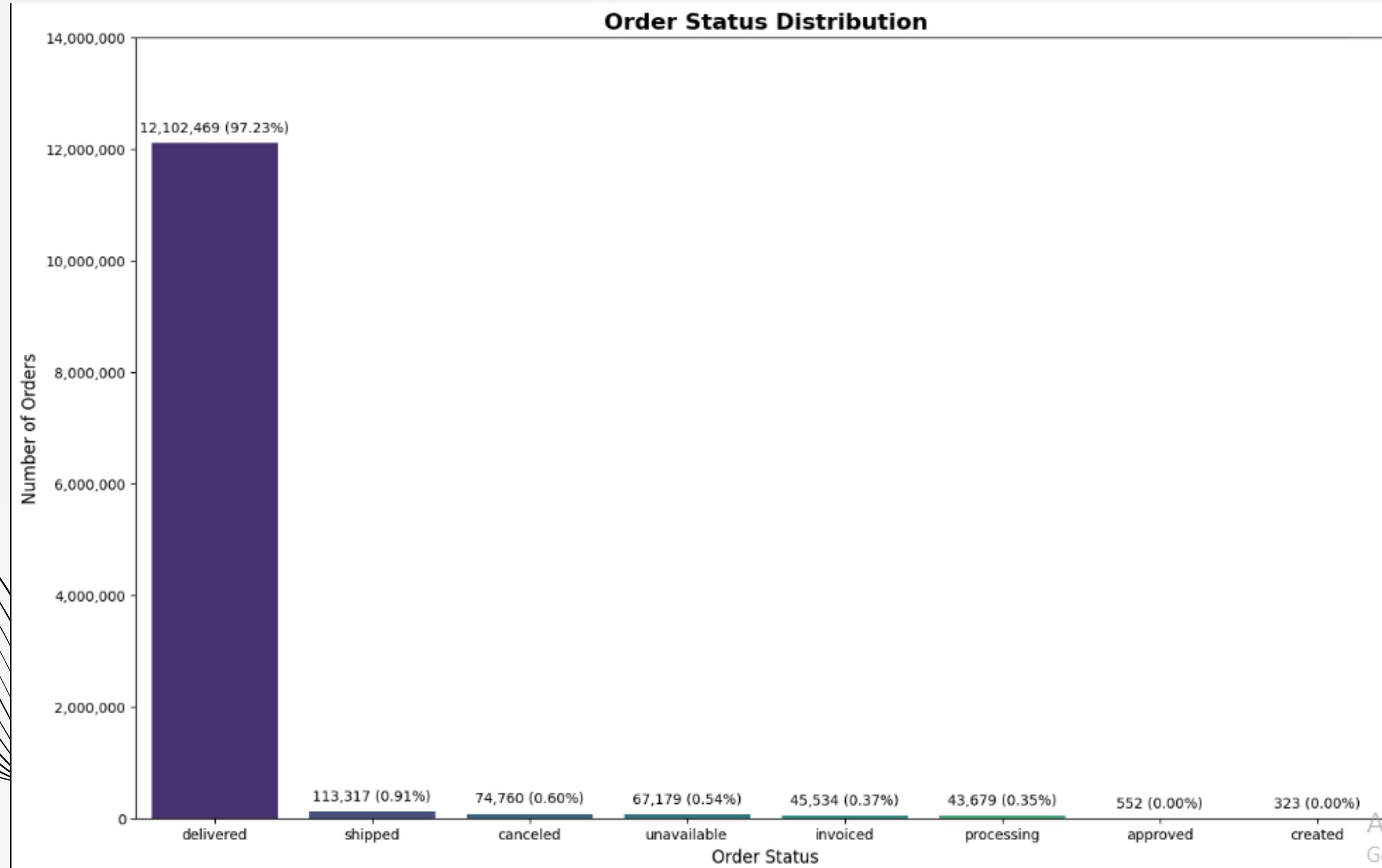
Data yang diberikan meliputi beberapa tabel yaitu tabel customer, order, geolocation, seller, dan products



Data yang diberikan merupakan data dari sebuah ecommerce yang ada di brazil terlihat dari pemesanannya berfokus pada negara tersebut

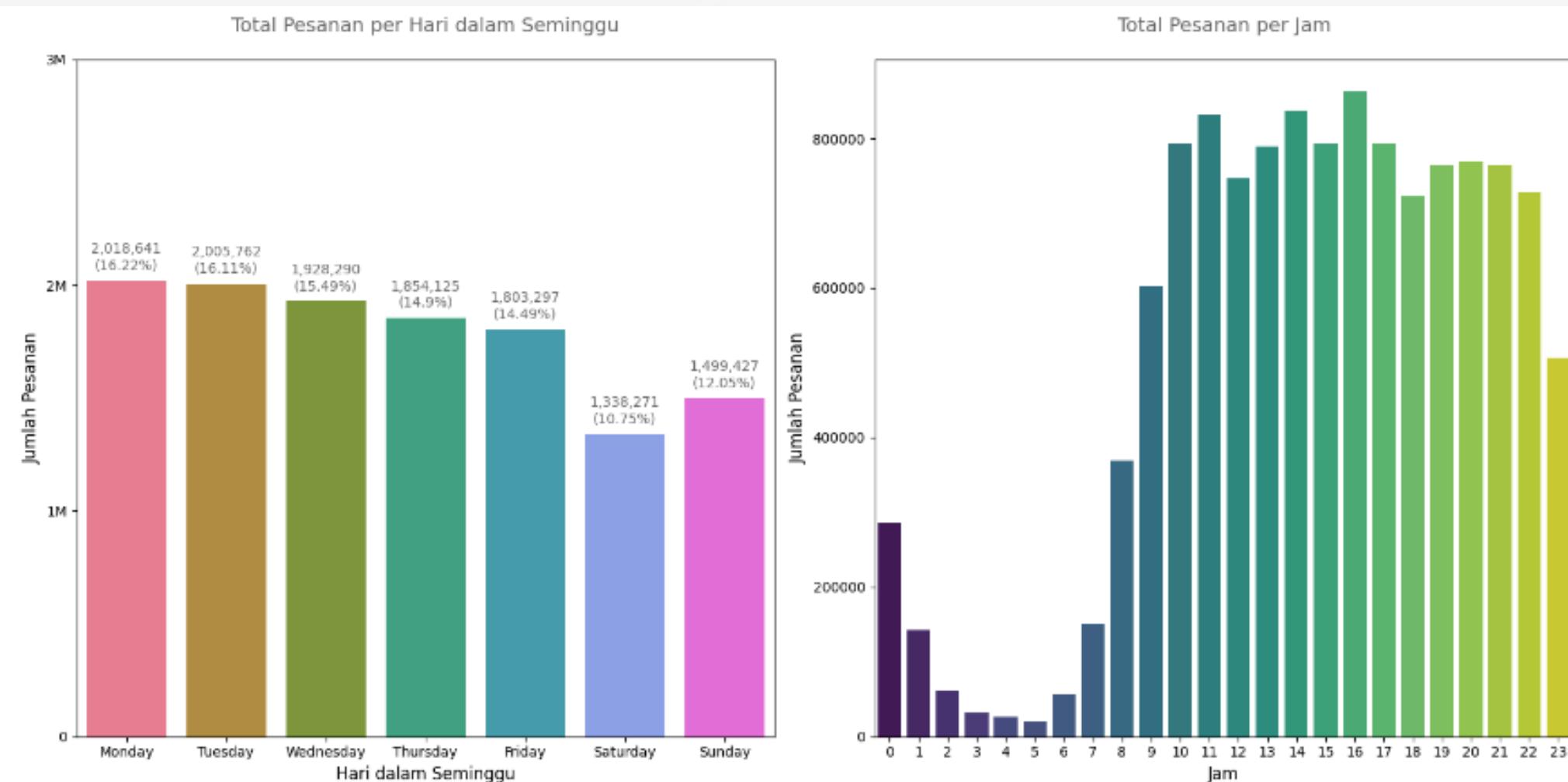
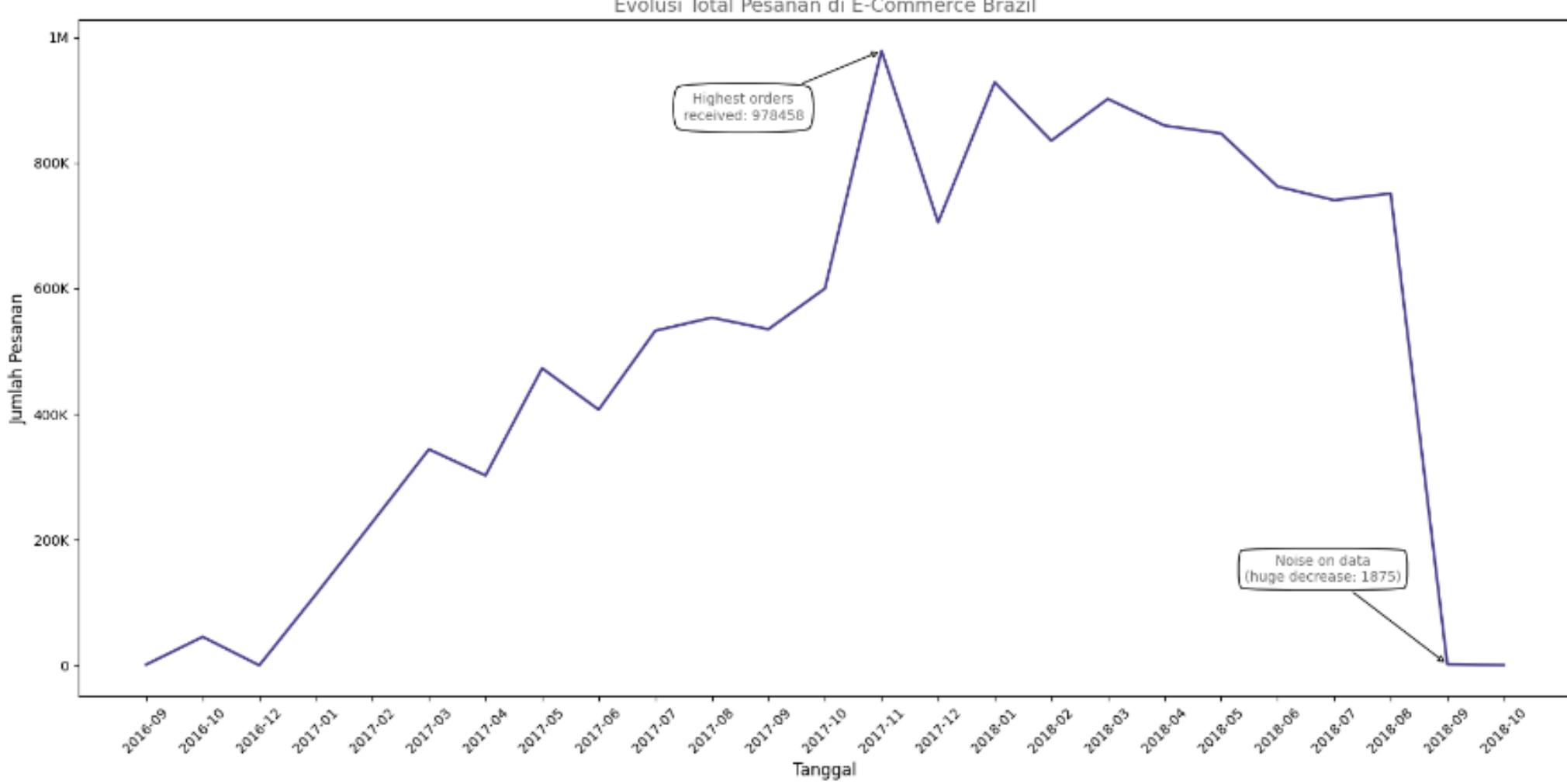


# EKSPLORASI DATA PESANAN



- Pada saat dataset ini dibuat, jumlah pesanan tertinggi berasal dari order berstatus delivered. Hanya 3% dari semua order berasal dari status lainnya.

# EKSPLORASI DATA PESANAN



Berdasarkan grafik di samping, kita dapat menyimpulkan:

- E-commerce di Brazil menunjukkan tren pertumbuhan yang signifikan dari waktu ke waktu. Meskipun terdapat beberapa periode dengan lonjakan pesanan pada bulan-bulan tertentu, secara keseluruhan terlihat bahwa belanja online semakin populer di kalangan konsumen.
- Hari Senin dan Selasa merupakan hari favorit bagi pelanggan Brazil, dengan lonjakan pembelian yang paling tinggi terjadi pada hari tersebut. Selain itu, pembelian cenderung meningkat pada sore hari.
- terdapat penurunan signifikan pada Agustus 2018 ke bulan September 2018, ini kemungkinan terdapat kesalahan dalam pengambilan data atau memang terdapat masalah yang belum diketahui pada periode tersebut

# EKSPLORASI DATA PESANAN

2,953,630

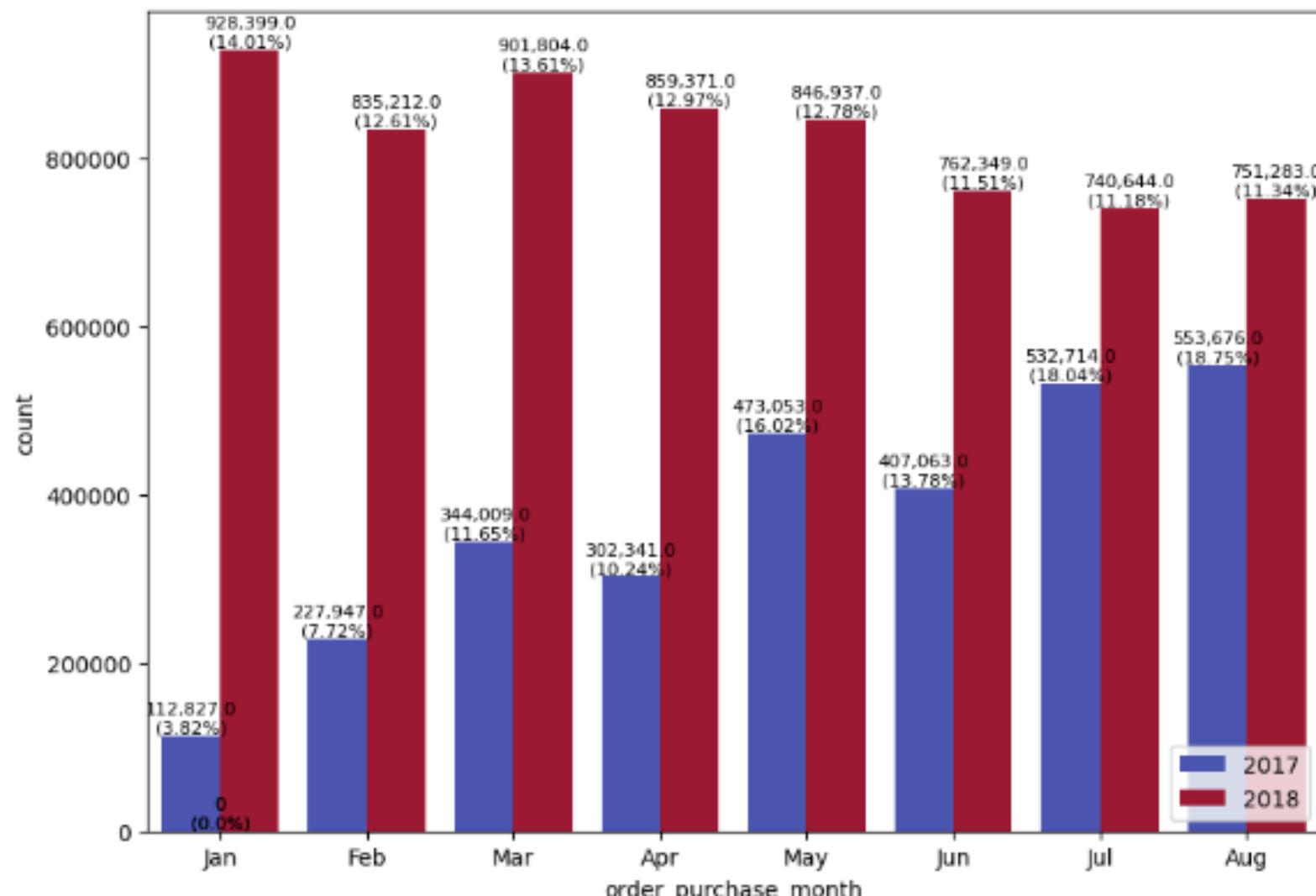
Pesanan terdaftar pada 2017  
antara Januari dan Agustus

6,625,999

Pesanan terdaftar pada 2018  
antara Januari dan Agustus

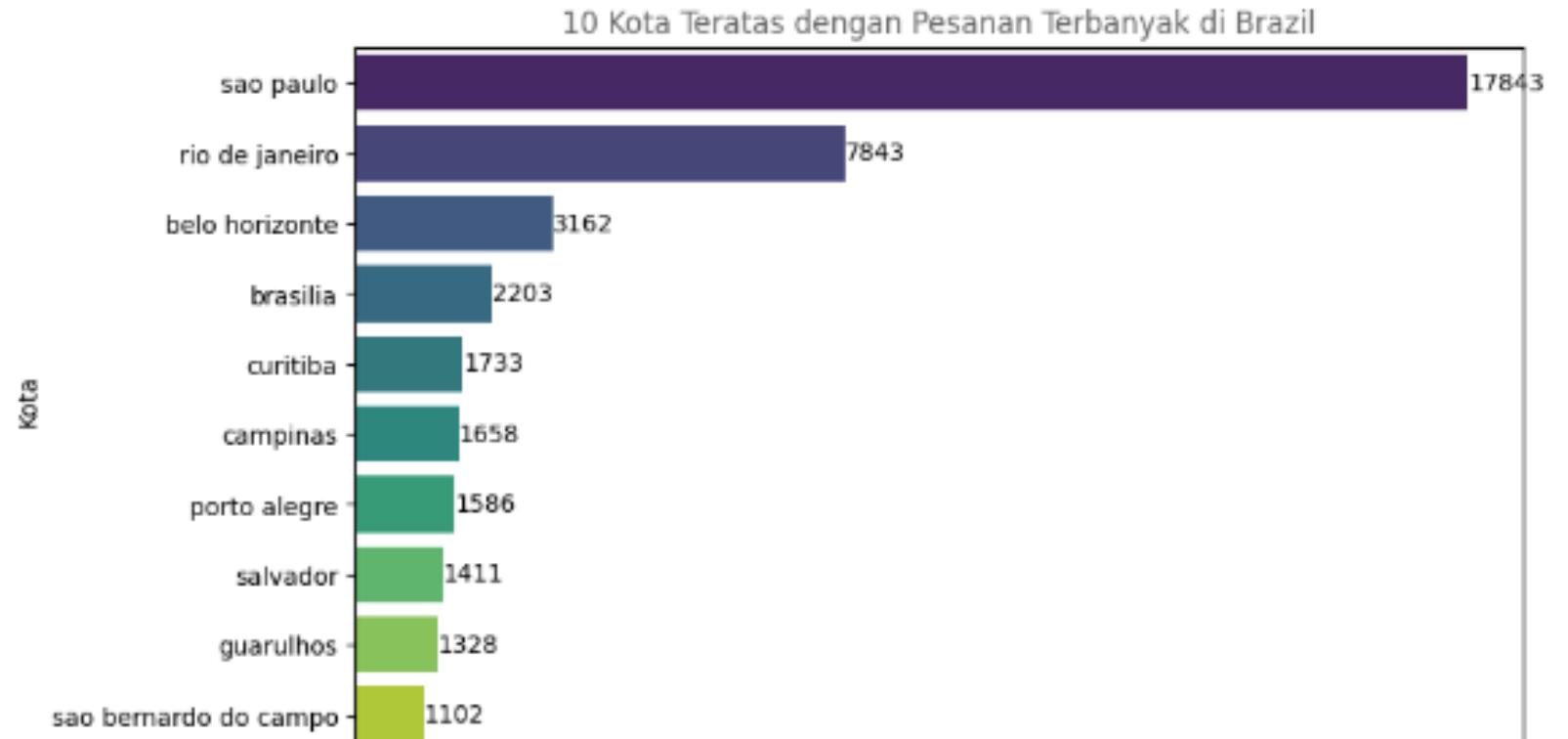
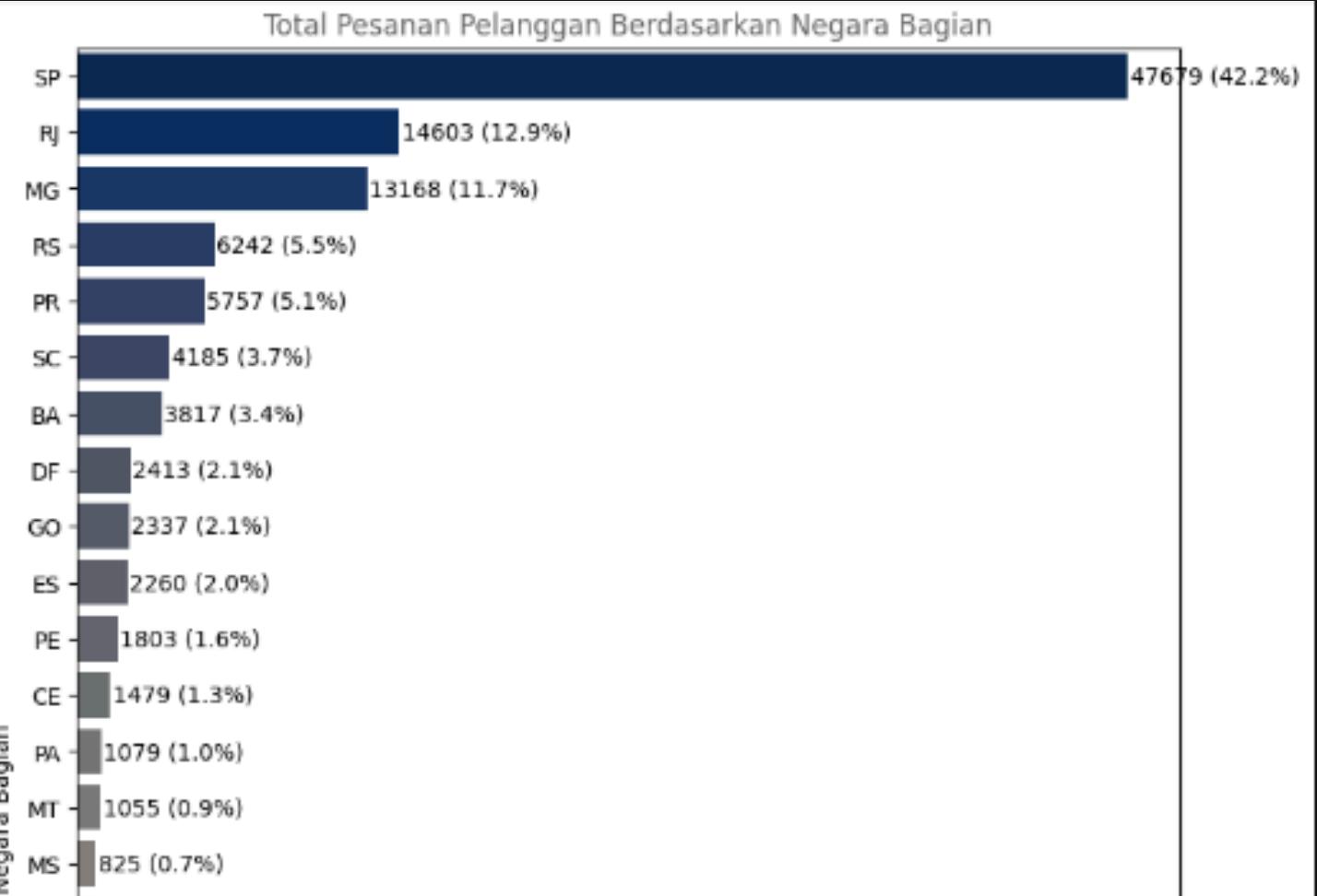
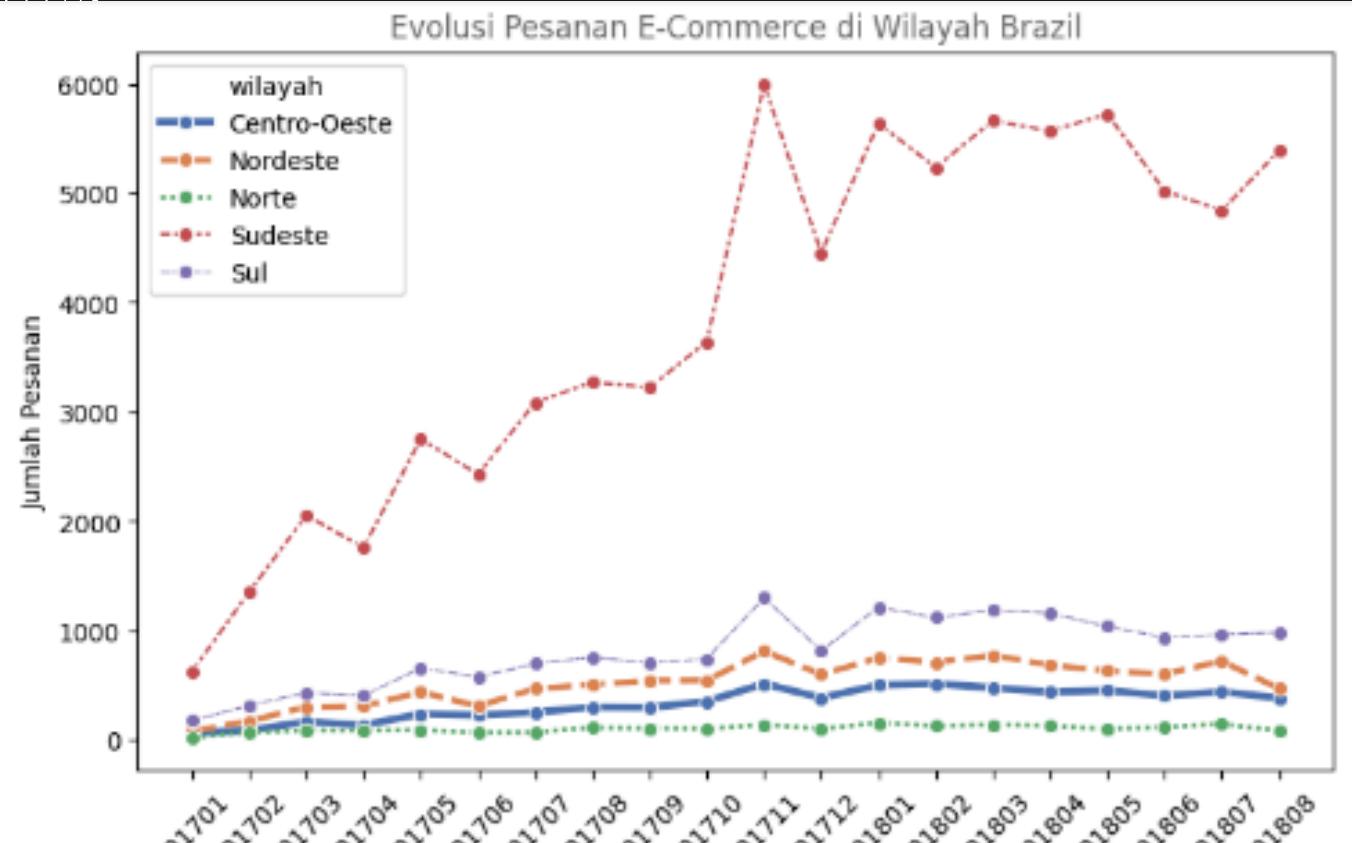
+145%

Perbandingan Total Pesanan antara 2017 dan 2018 (Januari hingga Agustus)



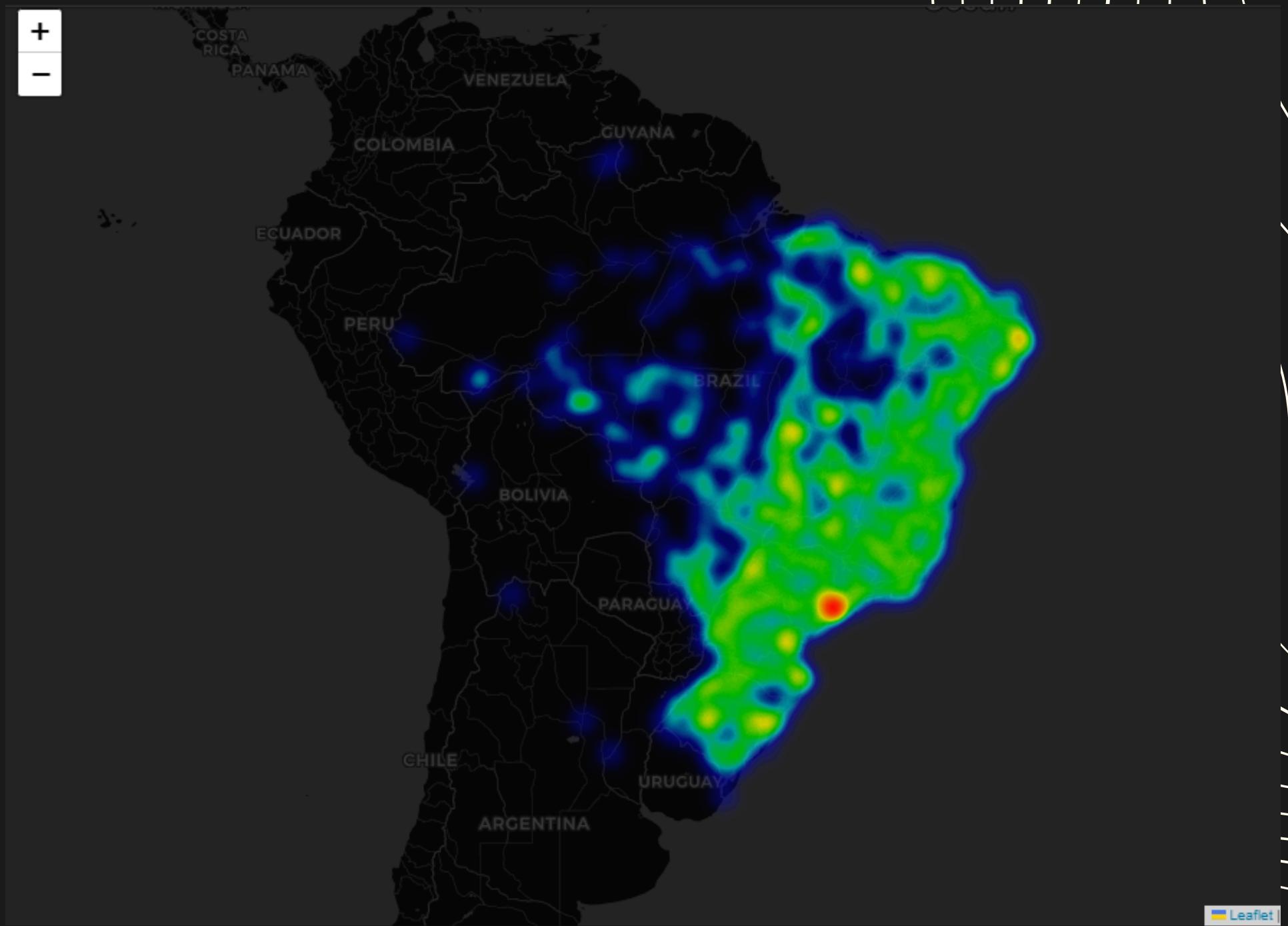
- Grafik disamping menunjukkan perbandingan pemesanan pada tahun 2017 dengan tahun 2018, dimana pemesanan mengalami peningkatan sebanyak 145% dibanding tahun sebelumnya

# EKSPOLORASI GEOLOCATION



- Grafik di disamping mengindikasikan bahwa pertumbuhan pesanan di wilayah tenggara Brazil (sudeste) mengalami peningkatan yang sangat signifikan dibandingkan dengan wilayah lainnya.
- Selain itu, grafik lainnya menunjukkan bahwa kota dengan jumlah pesanan e-commerce terbanyak adalah São Paulo.

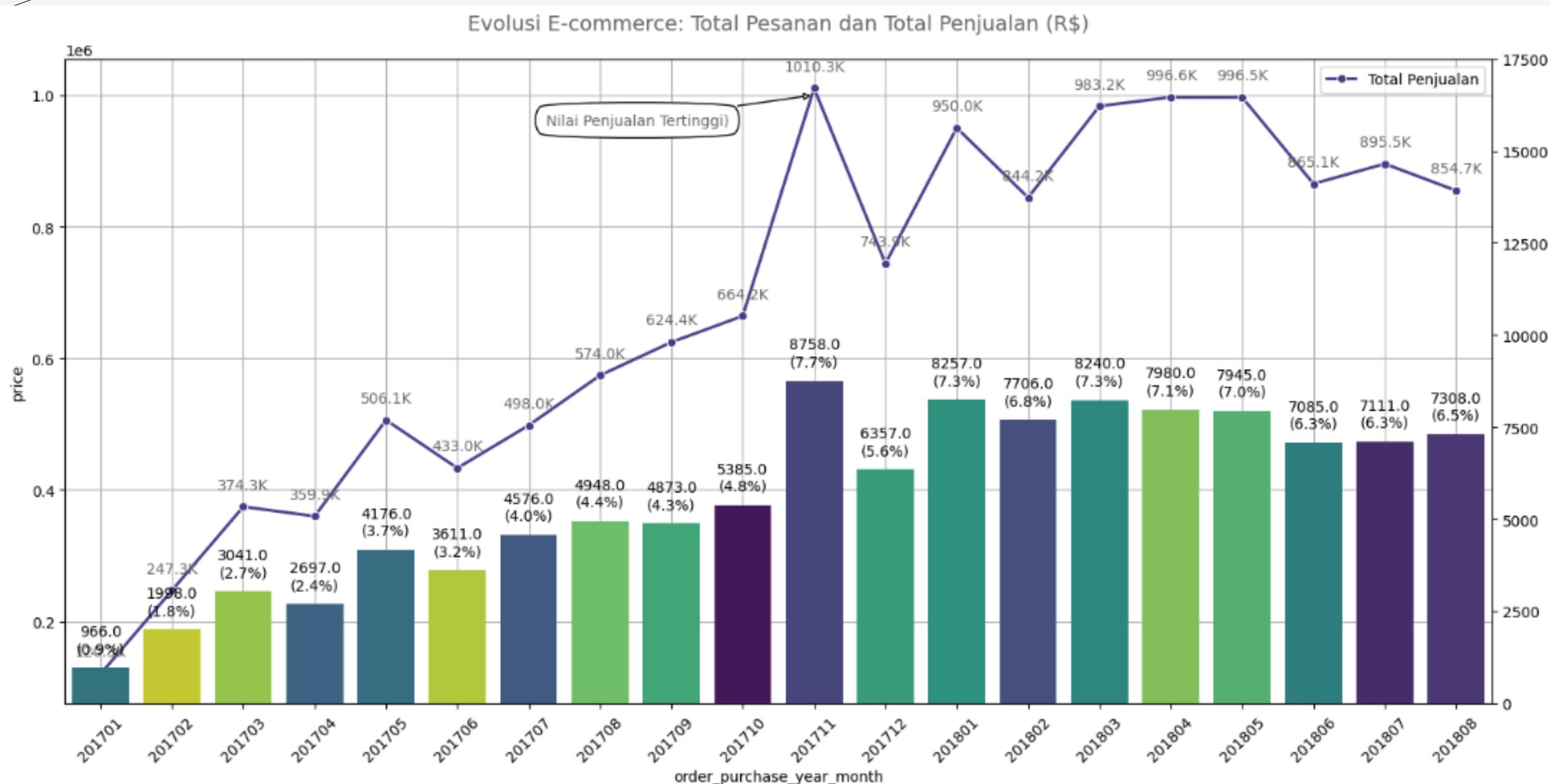
# GEOLOCATION MAP



Berdasarkan peta yang ditampilkan di atas, kita dapat melihat bahwa wilayah tenggara Brasil memiliki jumlah pesanan e-commerce tertinggi. Terlebih lagi, area São Paulo menunjukkan kepadatan pesanan yang paling tinggi, yang terlihat dari warna yang paling intens pada heatmap.

# EKSPOLASI PENJUALAN E-COMMERCE

01 02 03 04



- Grafik di atas menunjukkan bahwa total pesanan dan penjualan mengalami pertumbuhan yang signifikan dari tahun 2017 hingga pertengahan 2018. Pada bulan November 2017, terdapat puncak penjualan tertinggi dengan nilai mencapai R 1010.3K. Selama periode ini, penjualan meningkat sebesar +137.26% dari tahun 2017 ke 2018.

# EKSPOLASI PENJUALAN E-COMMERCE

R\$3.11M

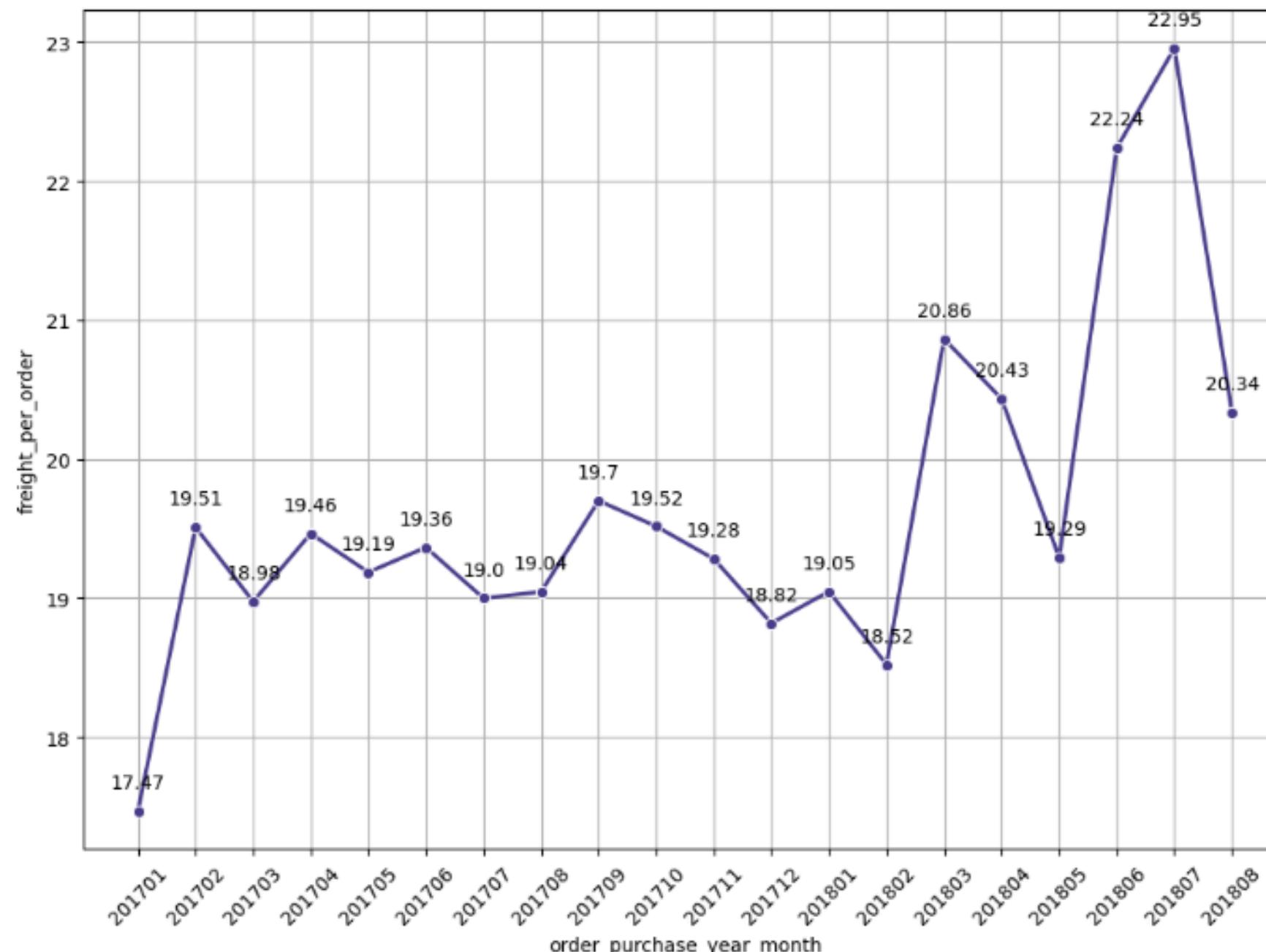
Total penjualan pada 2017  
antara Januari dan Agustus

R\$7.39M

Total penjualan pada 2018  
antara Januari dan Agustus

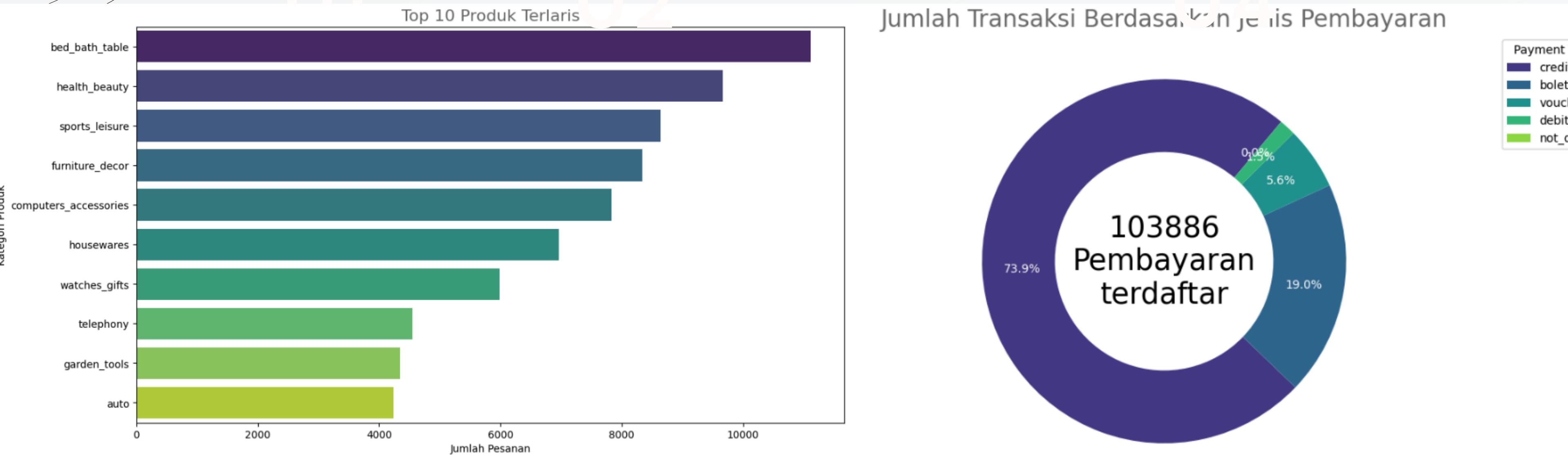
+137.26%

Evolusi Rata-Rata Nilai Pengiriman (R\$) yang Dibayar Pelanggan



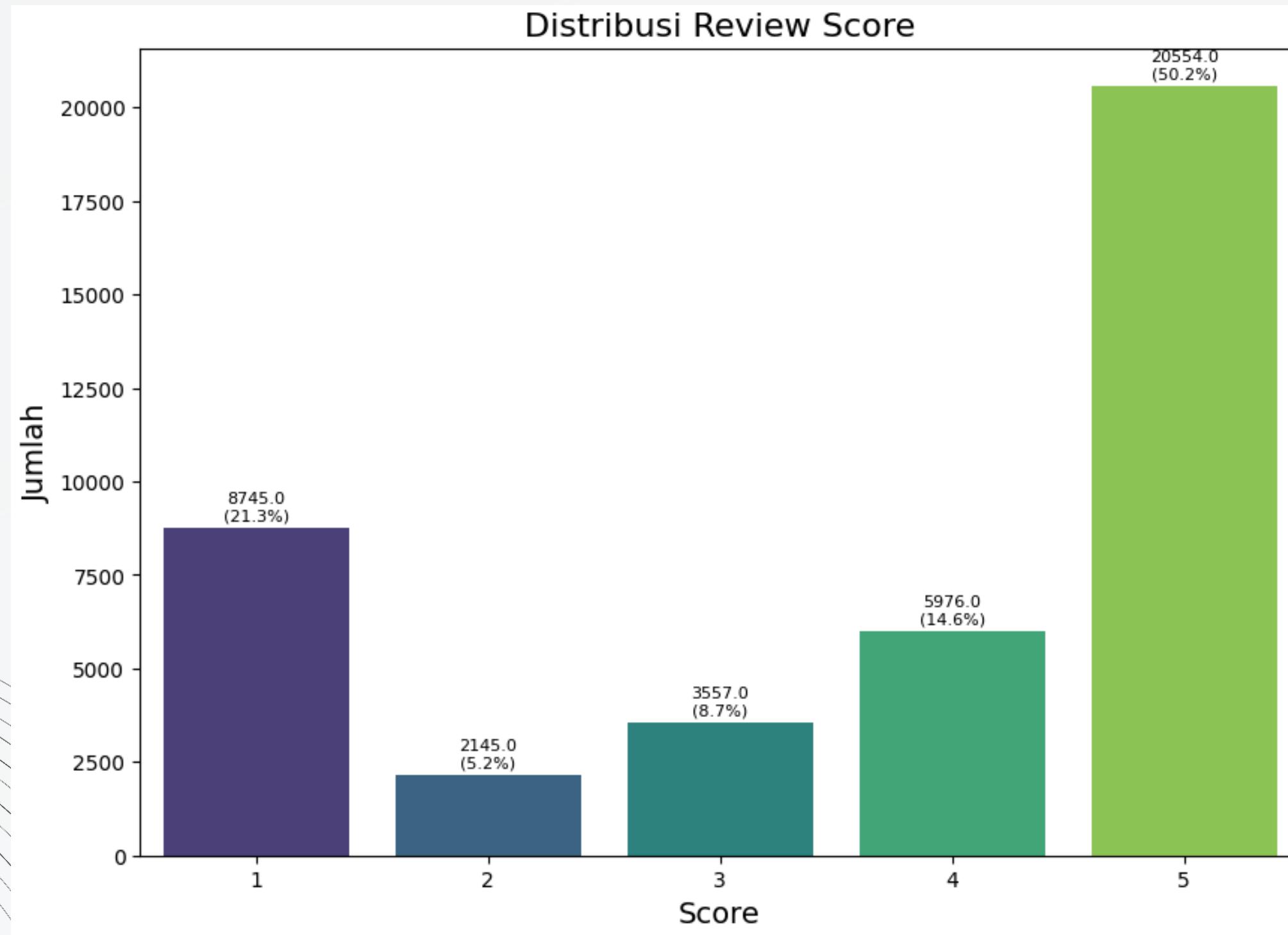
- Grafik juga menunjukkan fluktuasi dalam biaya pengiriman yang dibayar oleh pelanggan, dengan nilai tertinggi tercatat pada bulan July 2018, yaitu sekitar R 22.95 per pesanan. Setelah mencapai puncaknya pada July 2018, biaya pengiriman mengalami penurunan, tetapi tetap lebih tinggi dibandingkan dengan periode awal tahun 2017.

# EKSPOLASI PENJUALAN E-COMMERCE



E-commerce bisa fokus pada kategori populer seperti "bed\_bath\_table" dan "health\_beauty," sementara kategori lain seperti "auto" dan "garden\_tool" membutuhkan strategi pemasaran tambahan. Mendukung metode pembayaran yang paling umum tetap penting untuk meningkatkan pengalaman pelanggan.

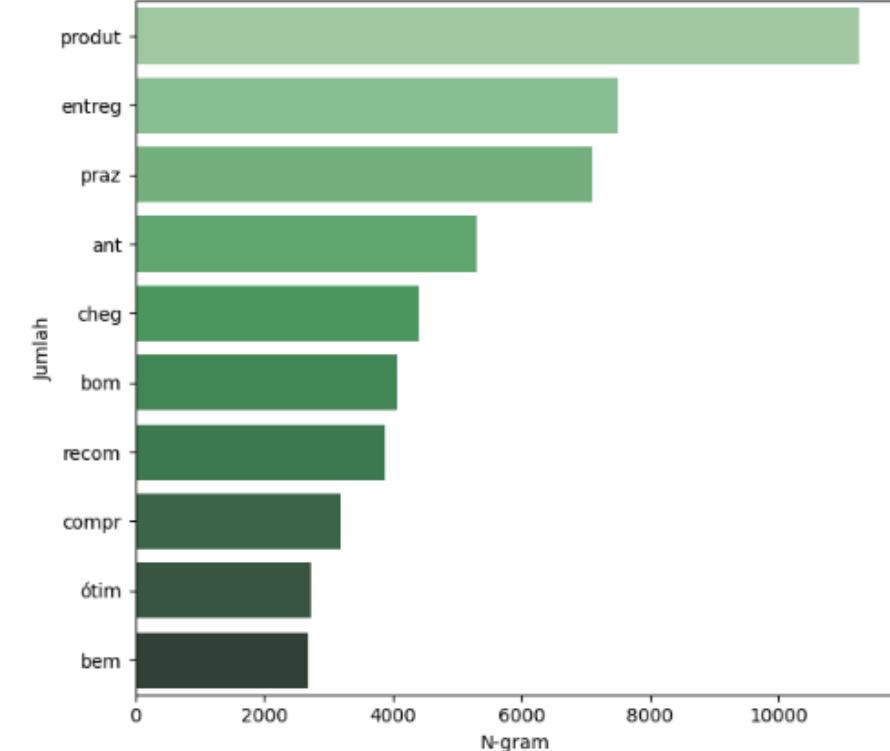
# REVIEW CUSTOMER (NLP)



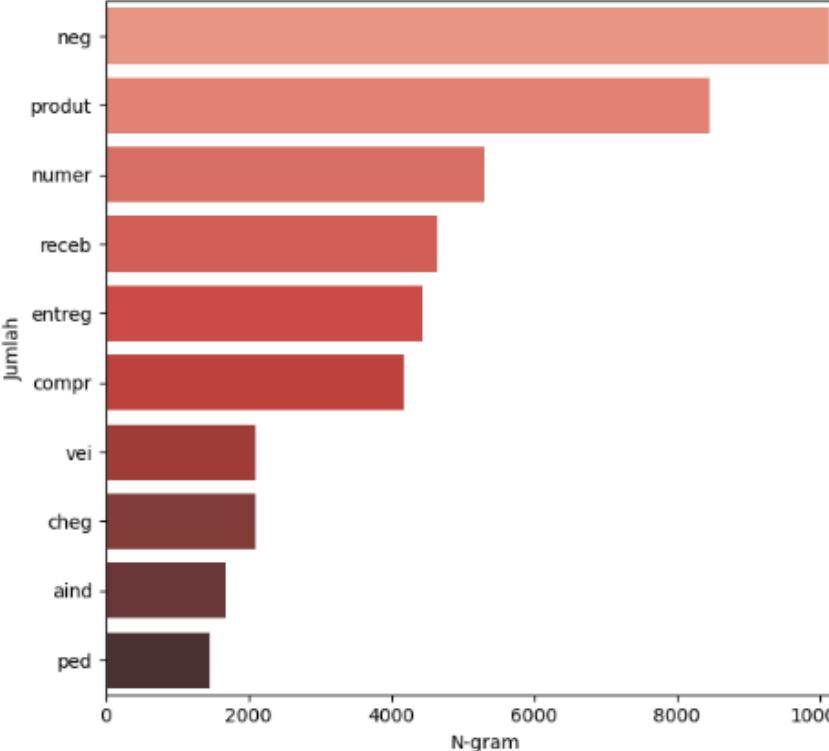
Mayoritas komentar di platform e-commerce bernilai 5, dengan 20.554 komentar yang menunjukkan kepuasan pelanggan yang tinggi. Meski begitu, ada 8.745 komentar dengan rating 1 yang menunjukkan ketidakpuasan. Ini mengindikasikan bahwa, meski umpan balik umumnya positif, masih ada aspek layanan yang perlu diperbaiki untuk meningkatkan pengalaman pelanggan secara keseluruhan.

# REVIEW CUSTOMER (NLP)

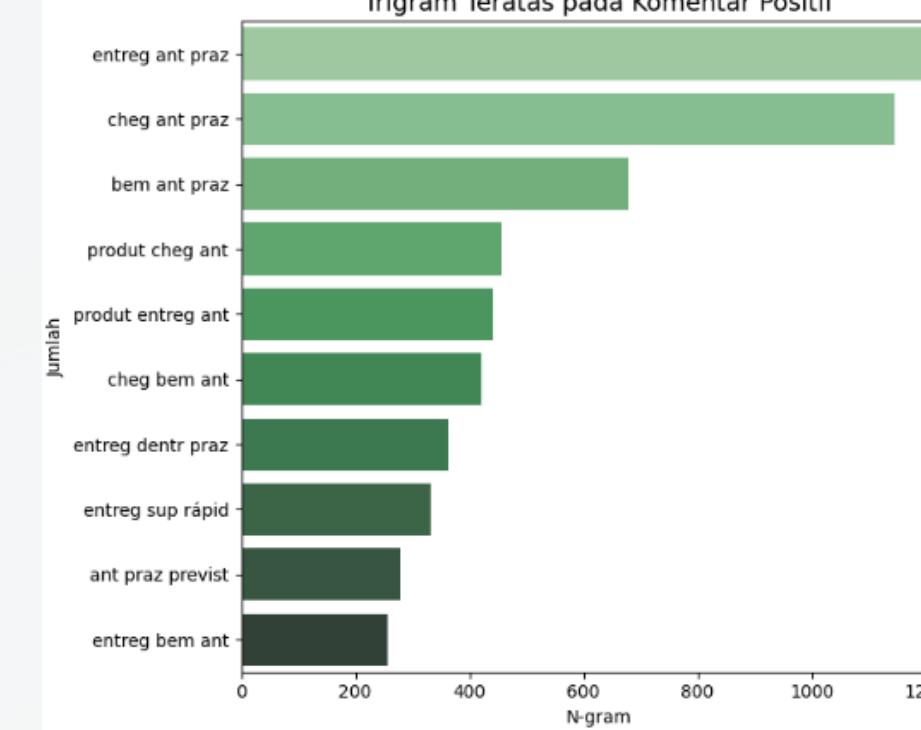
Unigram Teratas pada Komentar Positif



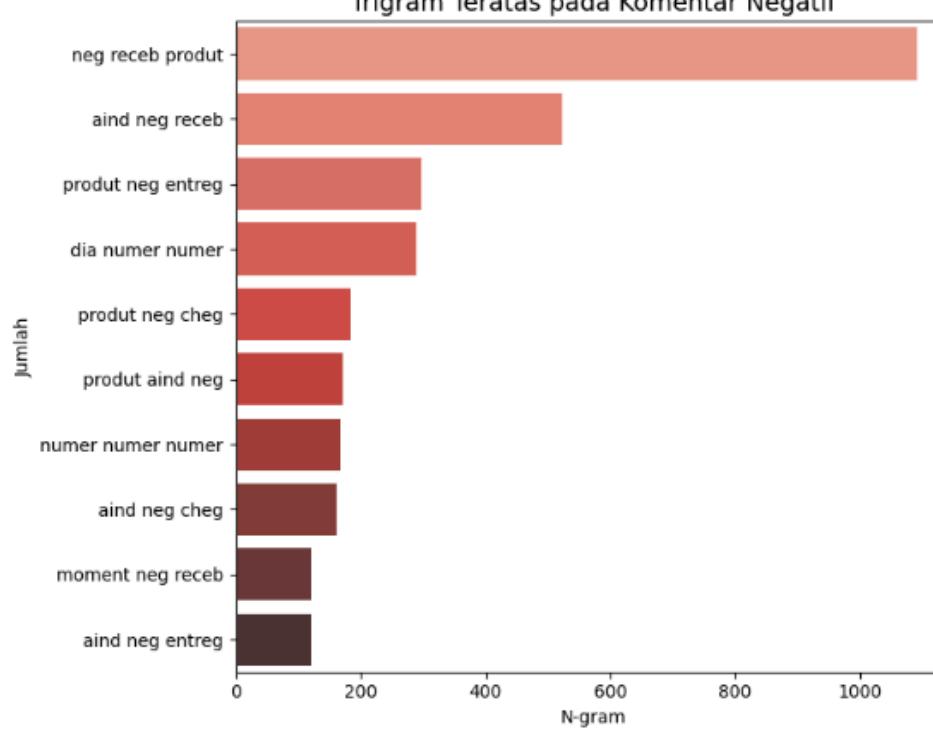
Unigram Teratas pada Komentar Negatif



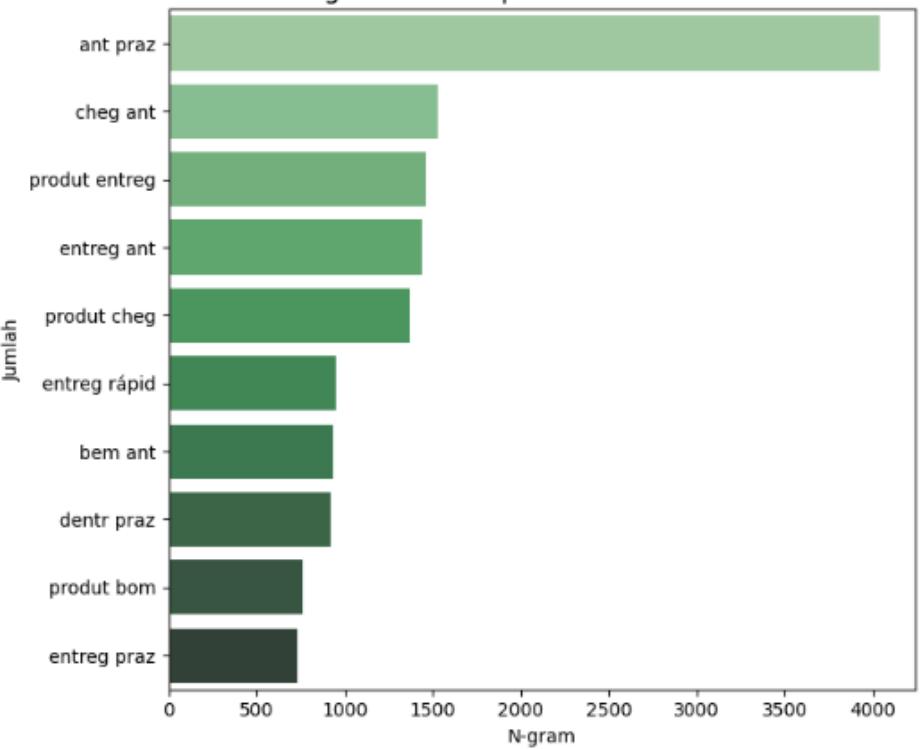
Trigram Teratas pada Komentar Positif



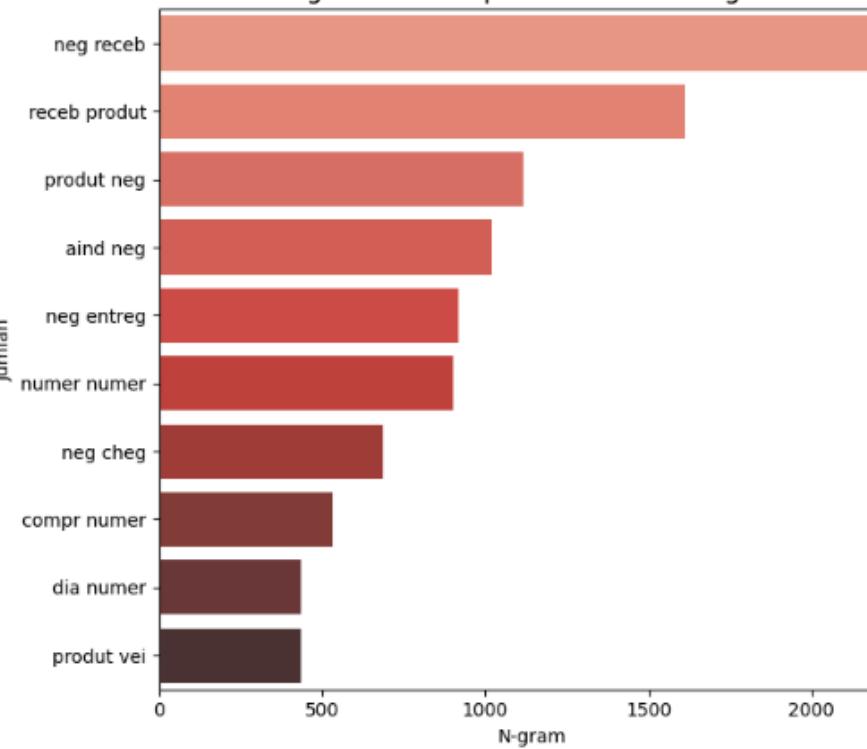
Trigram Teratas pada Komentar Negatif



Bigram Teratas pada Komentar Positif



Bigram Teratas pada Komentar Negatif



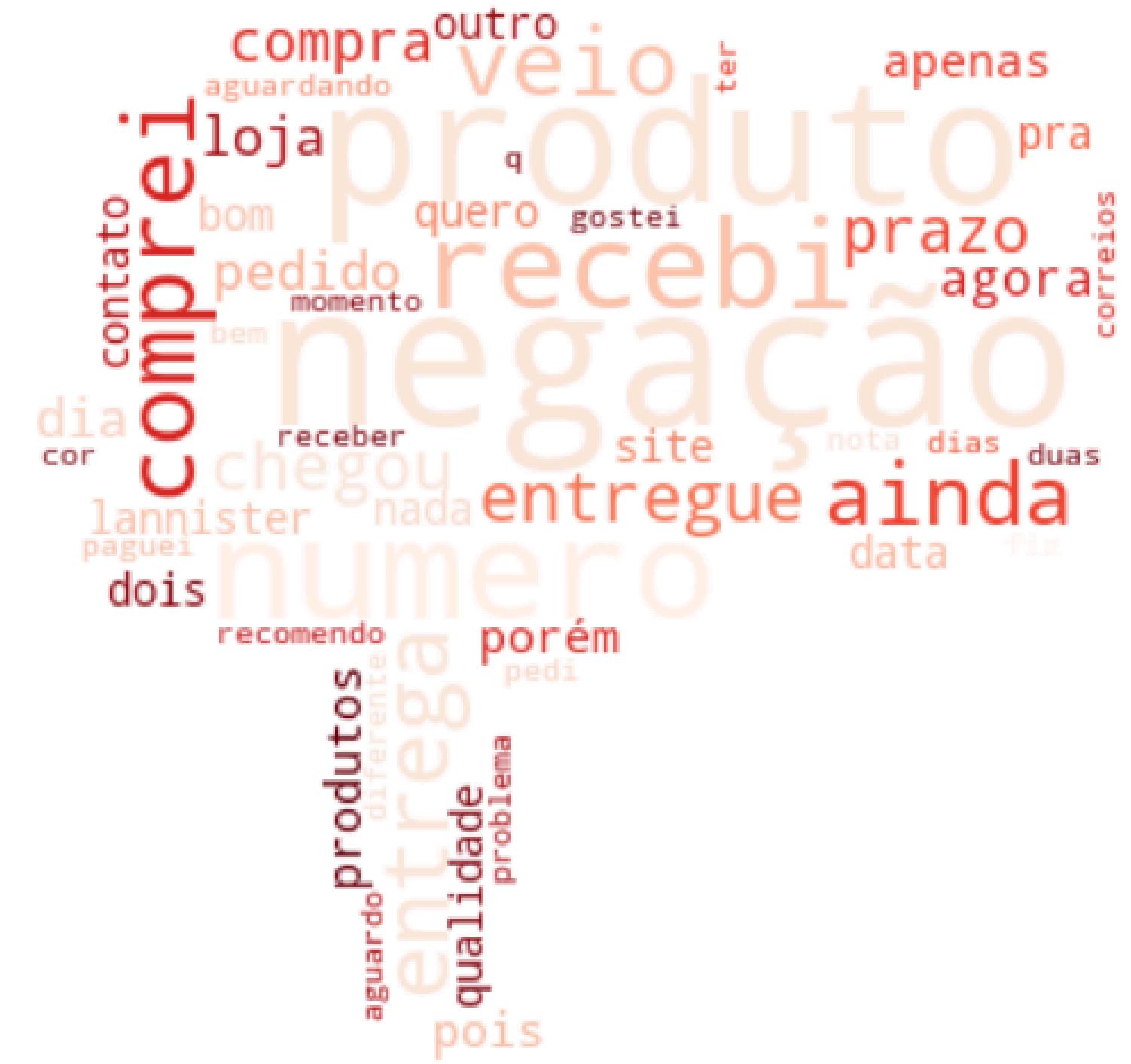
Secara keseluruhan, komentar positif menunjukkan bahwa pelanggan sangat menghargai kecepatan dan ketepatan waktu pengiriman, dengan kata-kata seperti "antes," "prazo," dan "entrega" sering muncul. Sebaliknya, komentar negatif sering menyebutkan masalah dengan produk yang diterima dan proses pembelian, terutama terkait tenggat waktu dan penerimaan produk. E-commerce dapat memanfaatkan wawasan ini untuk fokus pada peningkatan pengiriman tepat waktu dan kualitas produk.

# REVIEW CUSTOMER (NLP)

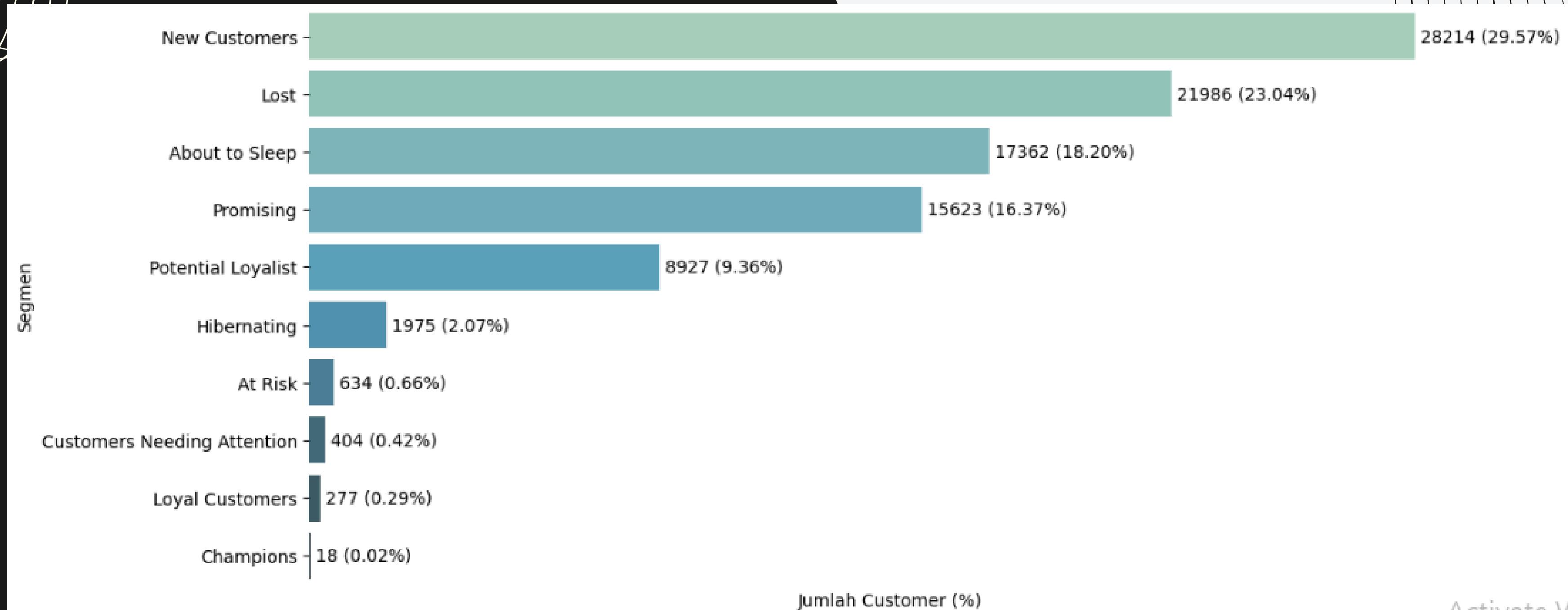
## WordCloud for Positive Words in Reviews



## WordCloud for Negative Words in Reviews



# SEGMENTASI CUSTOMER



Mayoritas pelanggan berada dalam segmen New Customers dan Lost, yang berarti bahwa E-commerce ini baik dalam menarik pelanggan baru tetapi perlu meningkatkan retensi pelanggan. Ada segmen yang menjanjikan seperti Promising dan Potential Loyalist yang bisa menjadi fokus untuk dijadikan loyalis. Loyal Customers dan Champions adalah segmen kecil tetapi sangat penting untuk dipertahankan karena mereka merupakan pelanggan yang paling berharga.

# **BUILDING MODEL FOR CHURN PREDICTION**

# FEATURE ENGINEERING



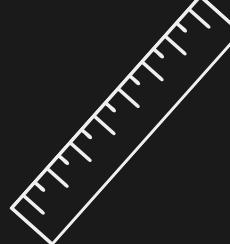
Mengisi nilai yang hilang dengan rata-rata, median, modus, atau menggunakan model untuk imputasi.

## HANDLING MISSING VALUE



Mengubah fitur kategori menjadi fitur numerik. Misalnya, One-Hot Encoding, Label Encoding.

## ENCODING CATEGORICAL FEATURES



Menyetarakan skala fitur yang memiliki rentang nilai yang berbeda, agar model tidak bias terhadap fitur dengan rentang nilai yang lebih besar.

## FEATURE SCALING



Mengatasi ketidakseimbangan data kelas dengan teknik oversampling (SMOTE) atau undersampling.

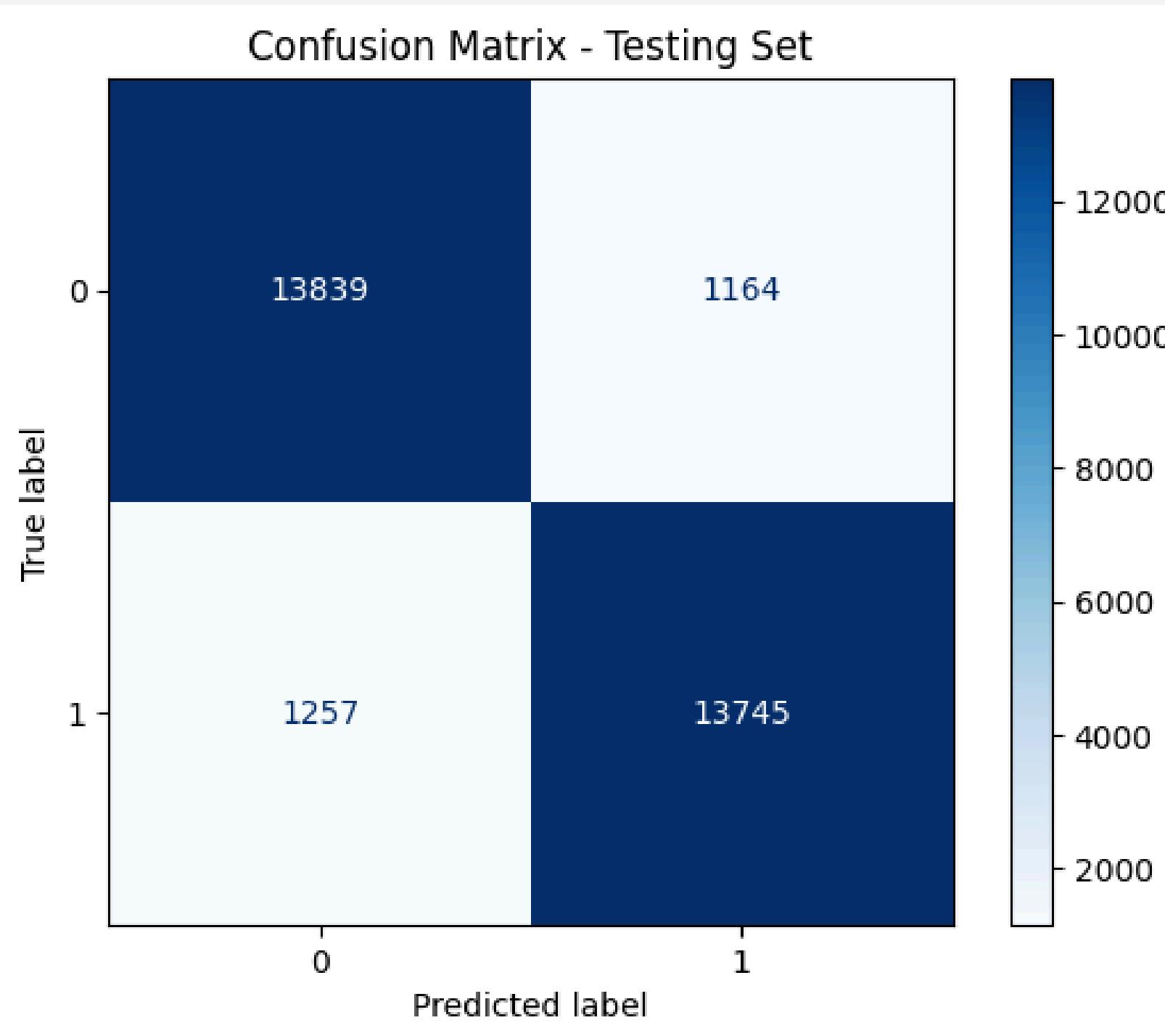
## HANDLING IMBALANCED DATA

# MODEL

| NO | Model Machine Learning                      | Score |
|----|---|-------|
| 1  | Random Forest (RF)                          | 0.92  |
| 2  | Gradient Boosting (GB)                      | 0.91  |
| 3  | Logistic Regression (LR)Decission Tree (DT) | 0.91  |
| 4  | Logistic Regression (LR)                    | 0.90  |
| 5  | Linear Discriminant Analysis (LDA)          | 0.90  |

# EVALUATION

Confusion Matrix - Testing Set



Evaluation on Testing

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.92      | 0.92   | 0.92     | 15003   |
| 1            | 0.92      | 0.92   | 0.92     | 15002   |
| accuracy     |           |        | 0.92     | 30005   |
| macro avg    | 0.92      | 0.92   | 0.92     | 30005   |
| weighted avg | 0.92      | 0.92   | 0.92     | 30005   |

Secara keseluruhan, model Random Forest ini cukup andal dan dapat diandalkan dalam melakukan prediksi dengan akurasi yang tinggi pada data yang belum pernah dilihat sebelumnya (data testing). Performa yang sangat baik pada data pelatihan mungkin mengindikasikan sedikit overfitting, tetapi hasil pada data testing menunjukkan bahwa model masih generalizable dengan baik.

# CONCLUTION

**Berdasarkan evaluasi terhadap data testing, model Random Forest memberikan kinerja terbaik dengan akurasi 91.9%. Penggunaan SMOTE untuk menangani data yang tidak seimbang juga membantu meningkatkan kinerja model, terutama dalam hal recall pada kelas minoritas.**

**Implementasi dari hasil prediksi ini dapat membantu e-commerce untuk mengurangi tingkat churn, meningkatkan retensi pelanggan, dan pada akhirnya meningkatkan profitabilitas perusahaan.**

# THANK'S FOR WATCHING

*Rijal Akhdan Khairulah*

