

# HOMework

## UNSUPERVISED LEARNING

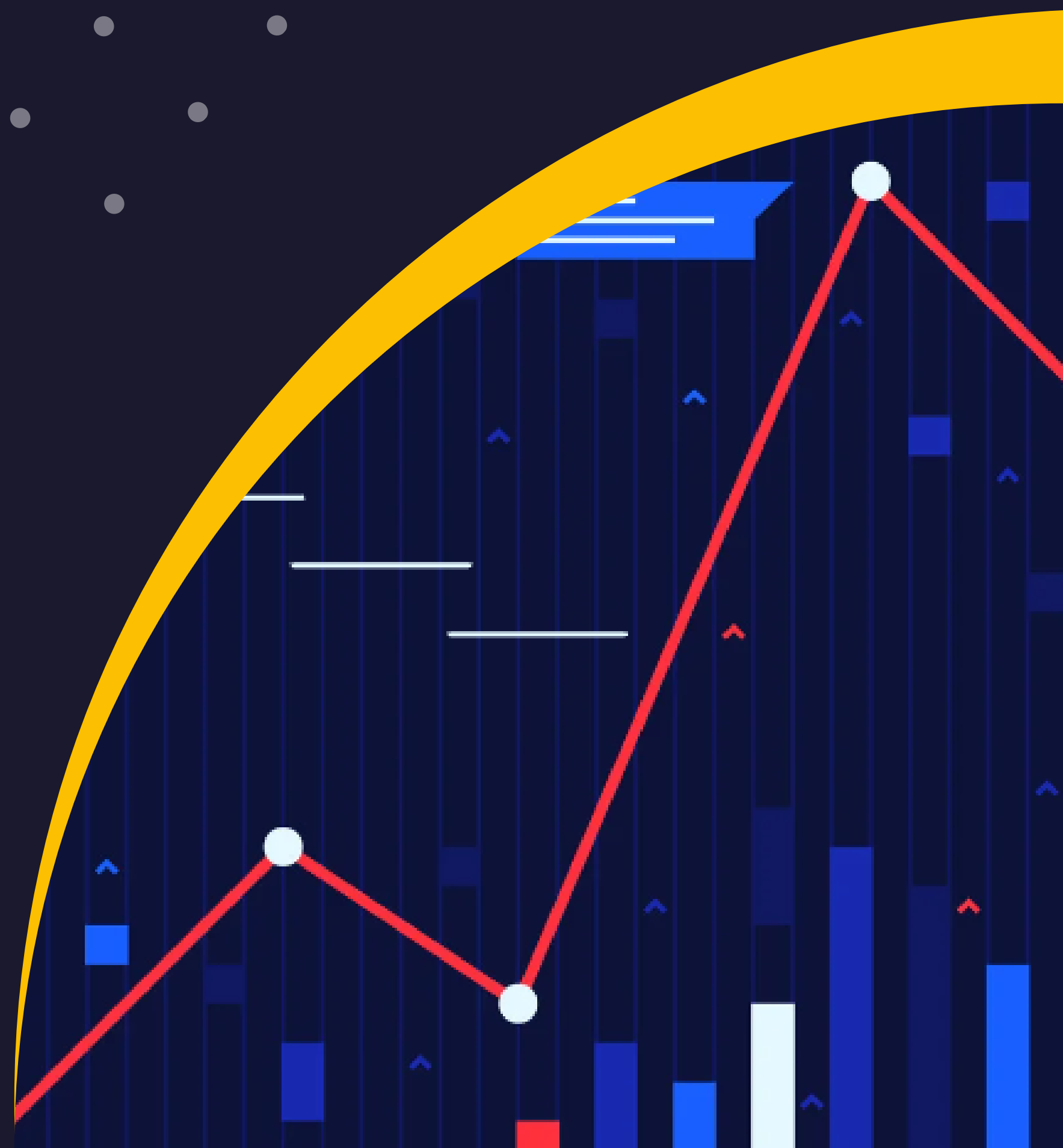
# KELOMPOK BER7AN

INDRA - JULDAN - HIMAWAN - TEGUH  
ZIFA - NADYA - EVINDA



# About DATASET

- 23 KOLOM
- 62.987 BARIS
- 7 KOLOM MISSING VALUE
- 0 DUPLICATE DATA





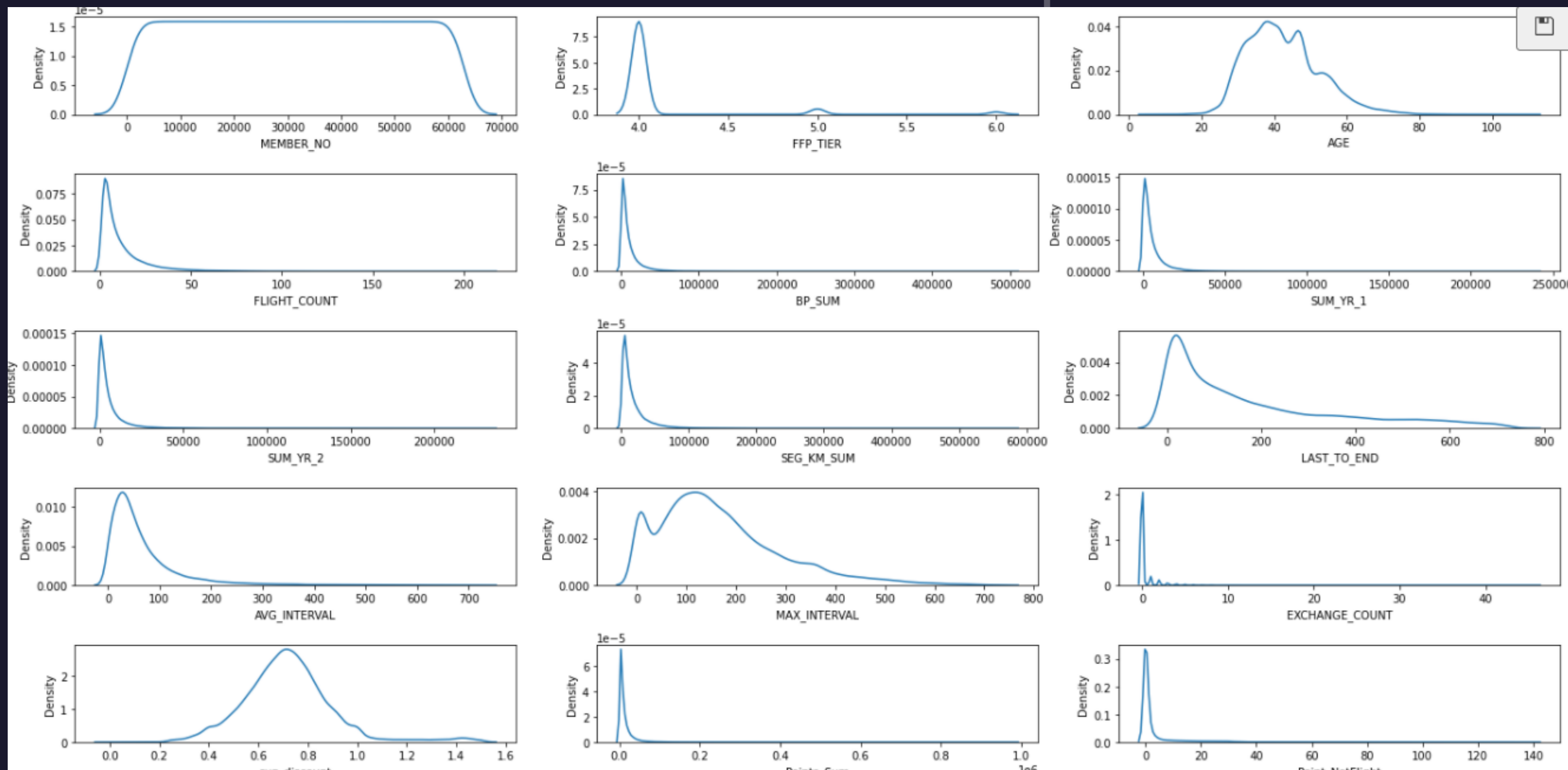
Kami memilih untuk menggunakan metode RFM dalam membuat segmentasi customer. Hal ini karena metode RFM adalah salah satu metode segmentasi customer yang cukup sering digunakan di industri.

Oleh karena itu, pada dataset ini, kami menggunakan feature 'LAST\_TO\_END' sebagai recency, 'FLIGHT\_COUNT' sebagai frequency, dan 'SEG\_KM\_SUM' sebagai perwakilan monetary.



# ... EDA

## KDE PLOT

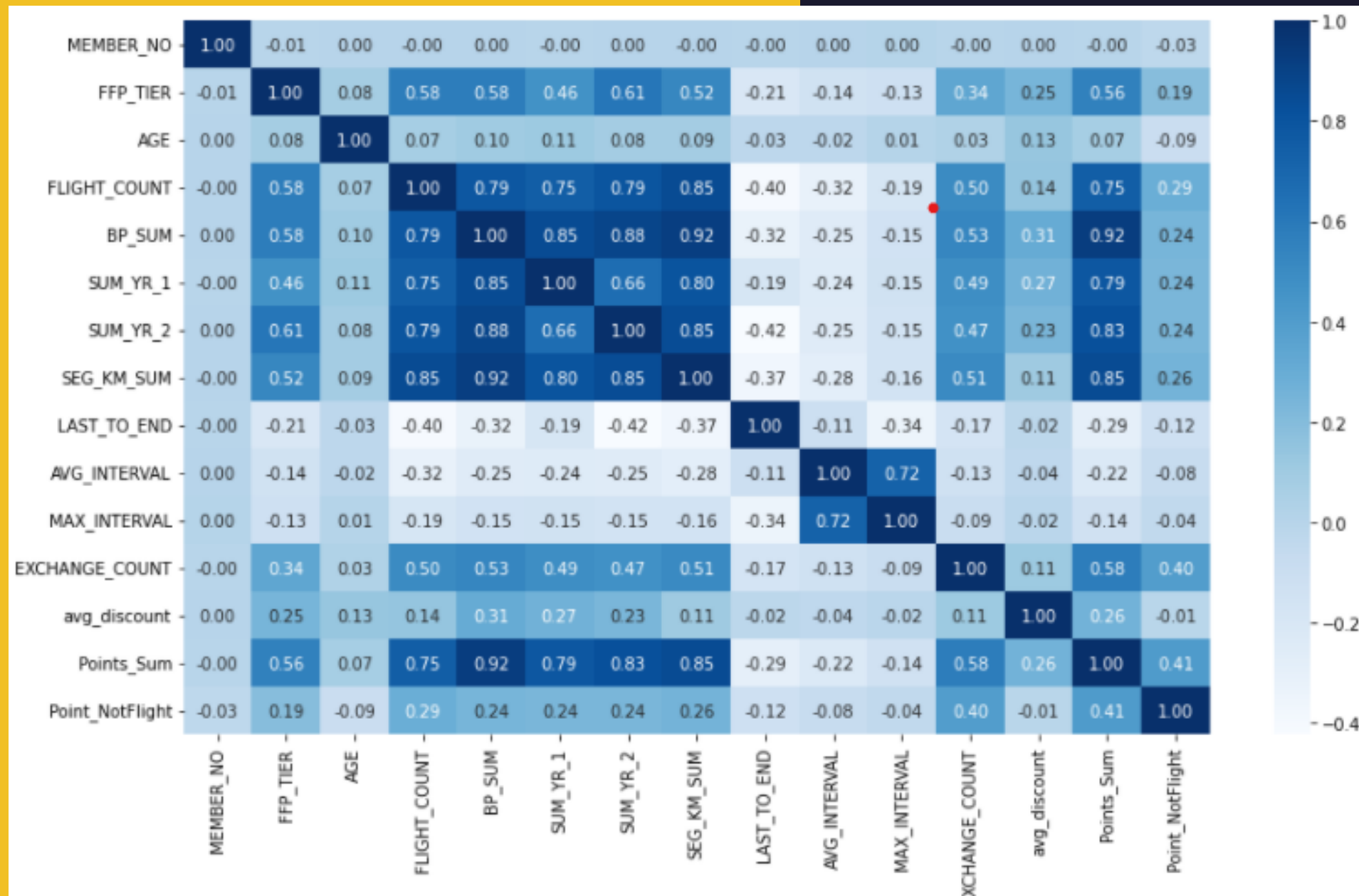


## HASIL PENGAMATAN

Kolom `avg\_discount`, `FFP\_TIER`, `AGE` memiliki distribusi data yang normal. Selain itu, kolom lainnya memiliki distribusi right skewed.

# ●●● CORRELATION

## HASIL PENGAMATAN



- Kolom `FLIGHT\_COUNT` memiliki korelasi yang kuat  $\geq 0.75$  dengan beberapa kolom seperti kolom `Points\_Sum`, `SEG\_KM\_SUM`, `SUM\_YR\_2`, `SUM\_YR\_1`, `BP\_SUM`.
- Kolom `BP\_SUM` memiliki korelasi yang kuat  $\geq 0.88$  dengan beberapa kolom seperti kolom `Points\_Sum`, `SEG\_KM\_SUM`, `SUM\_YR\_2`, `SUM\_YR\_1`.
- Kolom `SUM\_YR1` memiliki korelasi yang kuat  $\geq 0.79$  dengan kolom `Points\_Sum` dan `SEG\_KM\_SUM`.
- Kolom `SUM\_YR2` memiliki korelasi yang kuat  $\geq 0.83$  dengan kolom `Points\_Sum` dan `SEG\_KM\_SUM`.
- Kolom `SEG\_KM\_SUM` memiliki korelasi yang kuat  $\geq 0.85$  dengan kolom `Points\_Sum`.

# Pre-processing



## Feature untuk Clustering:

- Recency (R) : 'LAST\_TO\_END'
- Frequency (F) : 'FLIGHT\_COUNT'
- Monetary (M) : 'SEG\_KM\_SUM'

## Missing Value

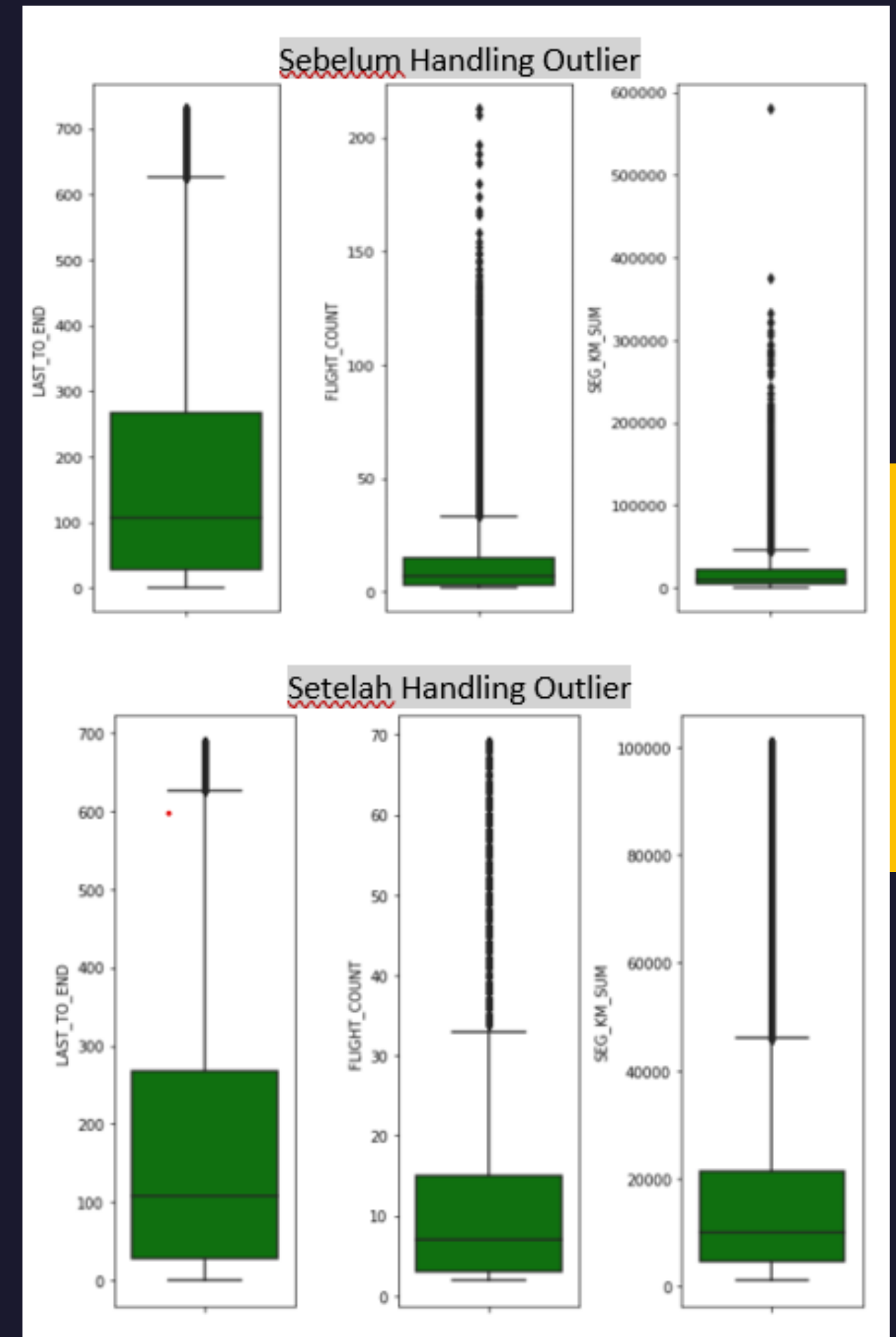
- dilakukan drop pada kolom yang memiliki missing value
- jumlah baris menjadi 62.988

## Handling Outlier

- menggunakan pendekatan IQR

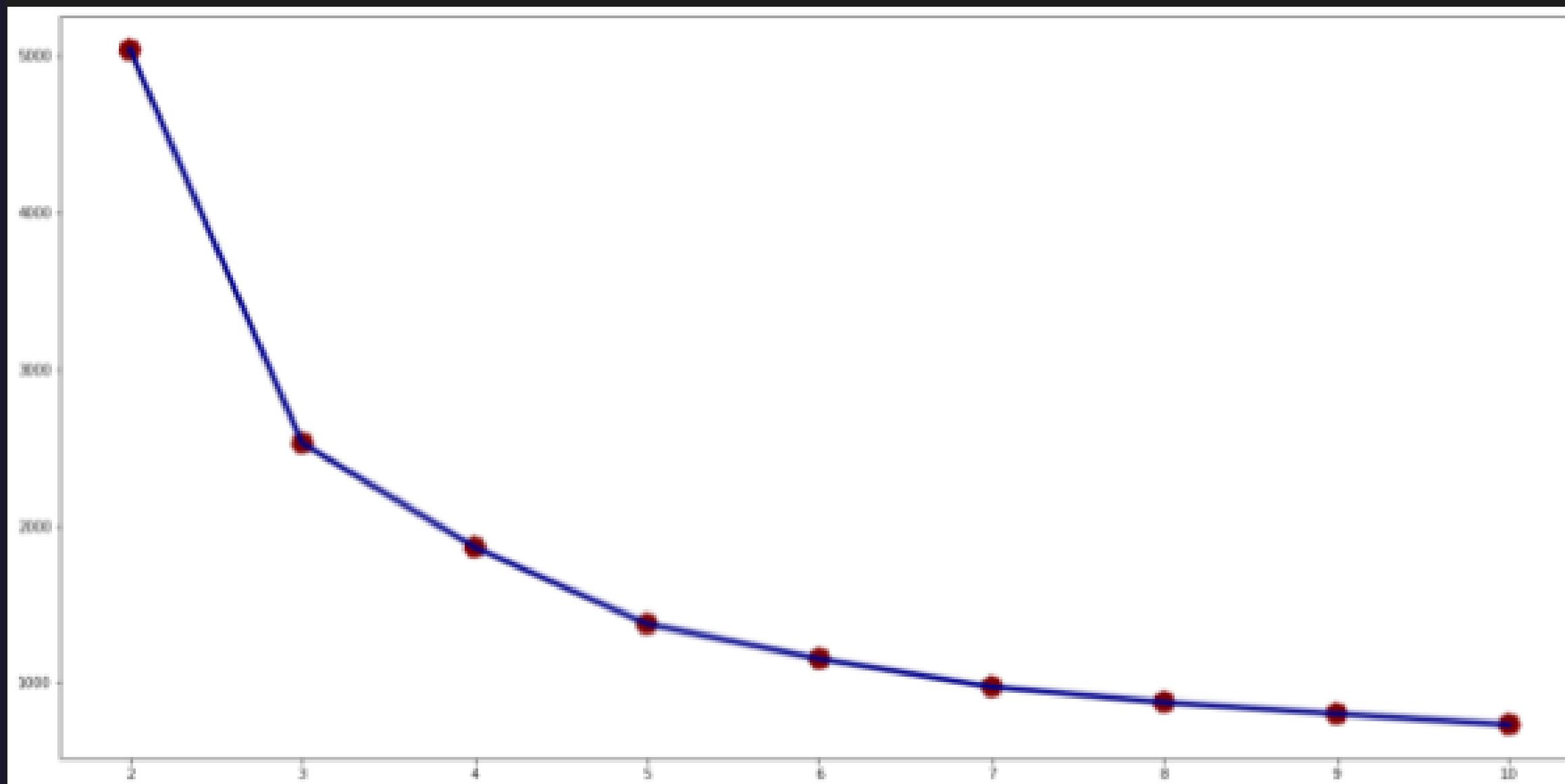
## Standardisasi

- menggunakan pendekatan StandarScaler



# ... MODELLING

## ELBOW METHOD



## HASIL

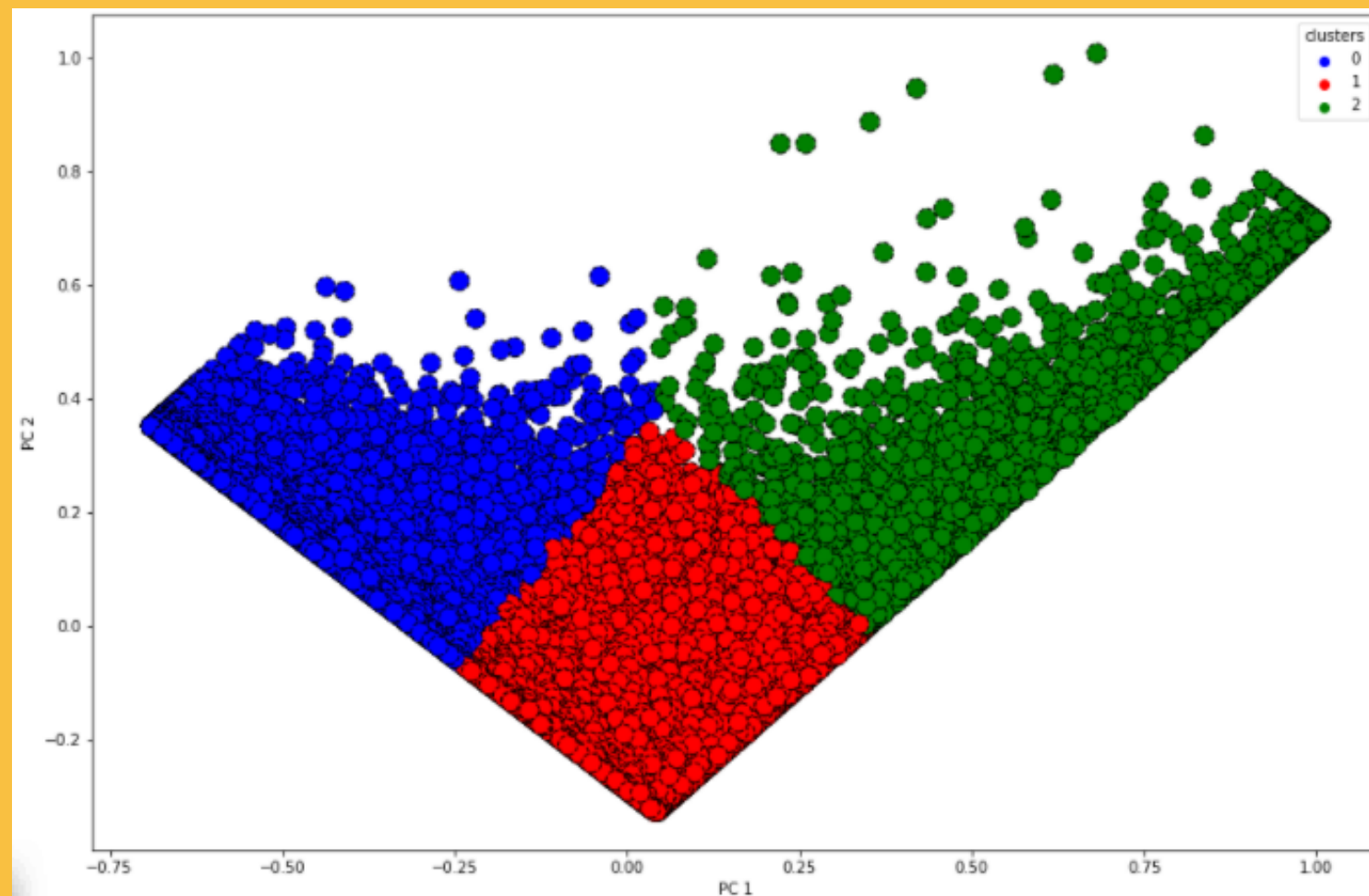
### + PENGAMATAN

+ Kolom `avg\_discount`,  
+ `FFP\_TIER`, `AGE`  
+ memiliki distribusi data  
+ yang normal. Selain itu,  
+ kolom lainnya memiliki  
+ distribusi right skewed.

## K-MEANS CLUSTERING

	MEMBER_NO	recency	frequency	monetary	clusters
0	54993	1	69	100841.28	2
1	28065	7	69	100841.28	2
2	55106	11	69	100841.28	2

## VISUALISASI : PCA





# ●●● REKOMENDASI BISNIS

Untuk customer Low Value dan Potential (clusters 0 dan clusters1) :

- Kami merekomendasikan untuk diberikan promo berupa cash back atau voucher yang menarik sehingga mereka dapat kembali untuk melakukan transaksi. Hal ini karena biaya untuk mempertahankan customer 5 - 25 kali lebih murah dibandingkan untuk mencari customer baru
- Kami merekomendasikan adanya akumulasi poin dari setiap penerbangan yang dilakukan customer selama satu tahun, lalu customer dengan poin-poin tertinggi(customer Important) akan diberikan undian hadiah. Hal ini bertujuan untuk membuat mereka tetap loyal.





# THANKS



+123-456-7890

hello@reallygreatsite.com

www.reallygreatsite.com