



# FINAL PROJECT

**PROGRAM DATA, BUSINESS ANALYTICS & OPERATIONS**

***KELOMPOK DBA-78***

***House Mentoring DBA- P***



# MEET THE TEAM



**Richardo Z. Damarjanaan**

DBA2215795  
Teknik Informatika  
Universitas 17 Agustus 1945 Surabaya

**Departemen**  
**Human Resources**



**Aritama Putra R.**

DBA2302440  
Statistika  
Universitas Brawijaya

**Departemen**  
**Sales**



**Detiya Nurfatwa**

DBA2205302  
Manajemen  
Universitas Gunadarma

**Departemen**  
**Product**



**Moch. Arief Tri Laksono**

DBA2080619  
Teknik Fisika  
Institut Teknologi Sepuluh November

**Departemen**  
**Marketing**

# Daftar Isi



**Latar Belakang**



**Metodologi**



**Hasil & Pembahasan**



**Kesimpulan dan Saran**

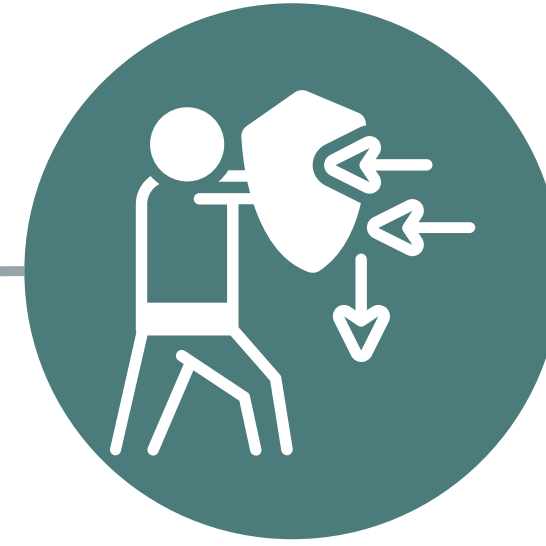
# LATAR BELAKANG



Industri 5.0



Persaingan yang  
lebih kompetitif

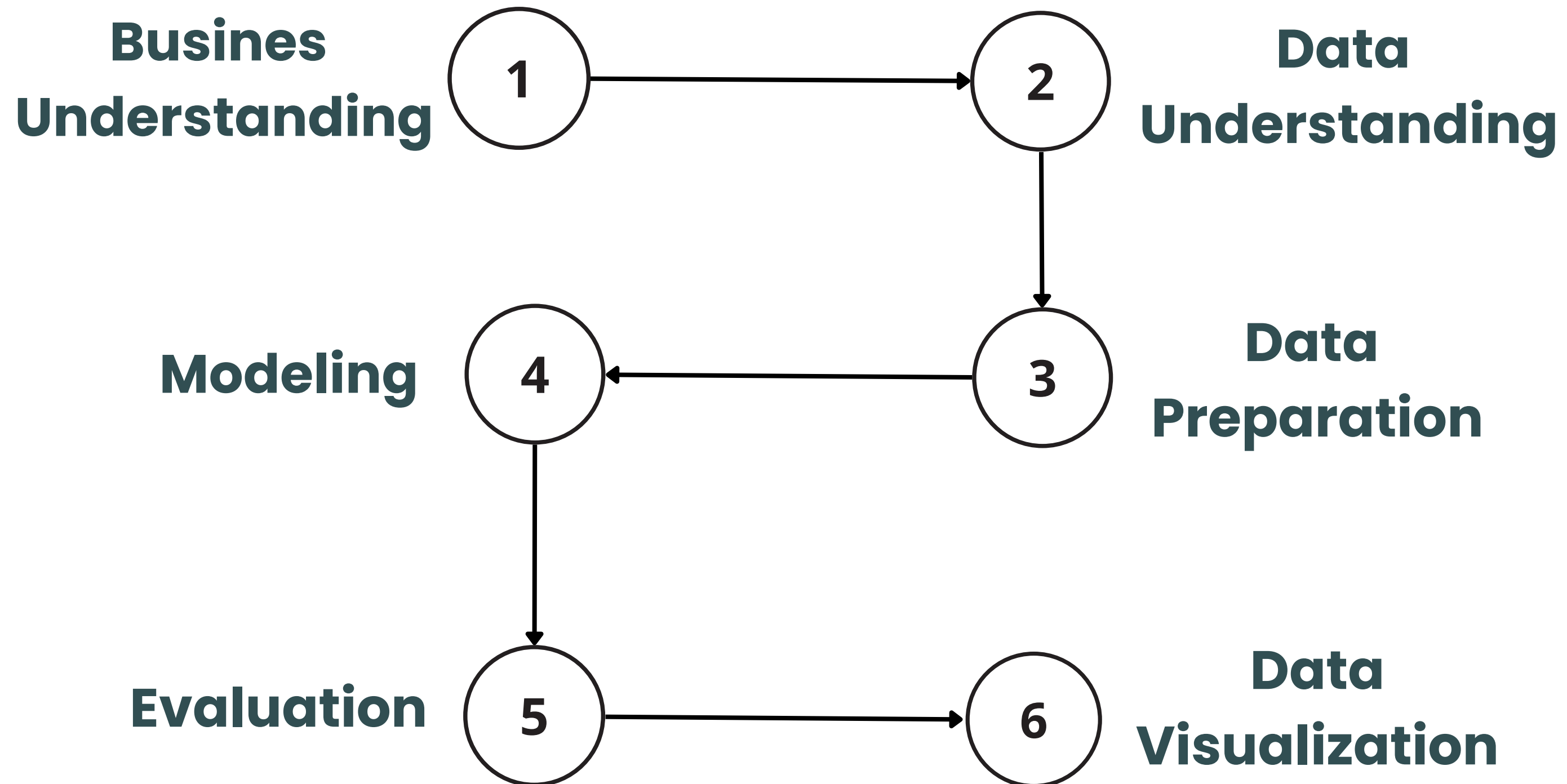


Mempertahankan  
Daya Saing



Inovasi dan  
Ekspansi Pasar

# METODOLOGI



# METODOLOGI

1

## Business Understanding

- Memahami kebutuhan serta tujuan dari sudut pandang bisnis
- Pendefinisian masalah dengan **5W+1H**
- Menentukan rencana, sasaran atau tujuan bisnis
- Menentukan strategi untuk mencapai tujuan



# METODOLOGI

## 2 Data Understanding

Dilakukan untuk Memahami Karakteristik tiap Data, dimulai dengan :

- Mengumpulkan Data
- Memahami Hubungan antar Dataset
- Mengidentifikasi Masalah Kualitas Data
- Beberapa Statistik Despkriptif

Tool yang digunakan :



### Product

Inventory\_items  
Distribution\_center\_id

### Sales

order\_items  
orders

### Marketing

Users  
order\_item  
orders

### HR

Employee  
Distribution\_center\_id



## 3 Data Preparation

- Melakukan Handling :
  - Missing Value dengan `dropna()` (jika perlu)
  - Duplicate Data dengan `drop_duplicates()` (jika perlu)
  - Outliers ( IQR Method)
- Melakukan EDA (Exploratory Data Analysis)
- Melakukan Transformasi dan Normalisasi menggunakan `OneHotEncoder()` , `LabelEncoder()` , dan `StandarScaler()`
- Menambahkan Variabel Baru
- Melakukan Splitting Data (Jika perlu)

### Product

Training : Testing  
90 : 10

### Sales

Training : Testing  
90 : 10

### Marketing

Principal  
Component  
Analysis

### HR

Training : Testing  
90 : 10

Tools yang digunakan :





# METODOLOGI

4

## Modelling

- Dilakukan metode Machine Learning
- Menentukan algoritma data mining yang akan digunakan

Tools yang digunakan :



### Product

- Classification

### Sales

- Regresi Logistik Biner

### Marketing

- K-Means Clustering

### HR

- Regresi Logistik Biner

## 5 Evaluation

- Melihat **tingkat performa** dari pola yang dihasilkan oleh algoritma.
- Parameter yang digunakan untuk Evaluasi komparasi algoritma pada tiap department seperti tabel disamping.

Tools yang digunakan :



### Product

Confusion Matrix,  
Accuracy,  
Precision, Recall

### Sales

Confusion Matrix,  
Accuracy,  
Precision, Recall

### Marketing

Shilouette score

### HR

Confusion Matrix,  
Accuracy,  
Precision, Recall

6

## Data Visualization

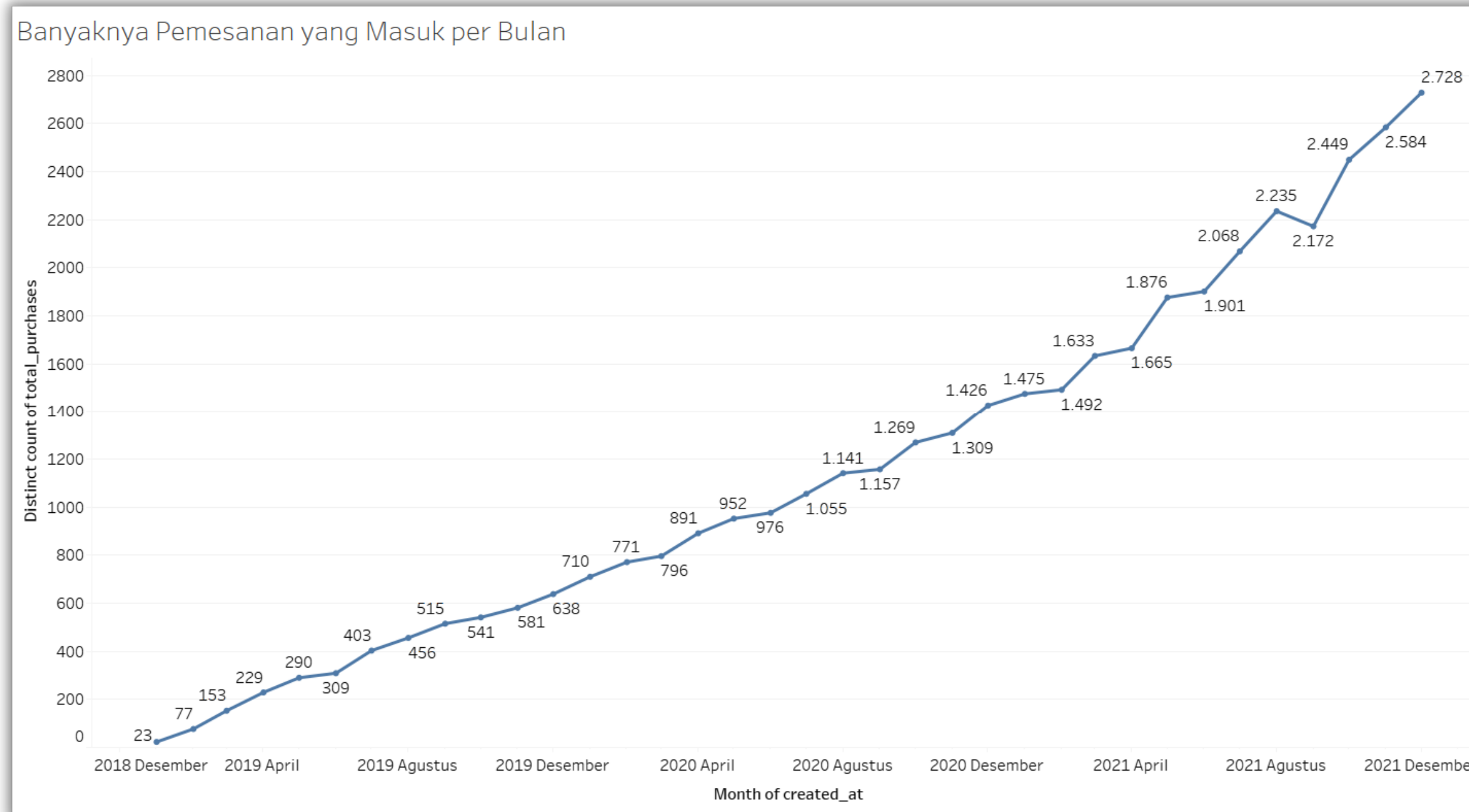
Dilakukan **visualisasi data** untuk **menyajikan data** agar lebih mudah dipahami serta **membuat *dashboard***



+ a b | e a u

# HASIL DAN PEMBAHASAN

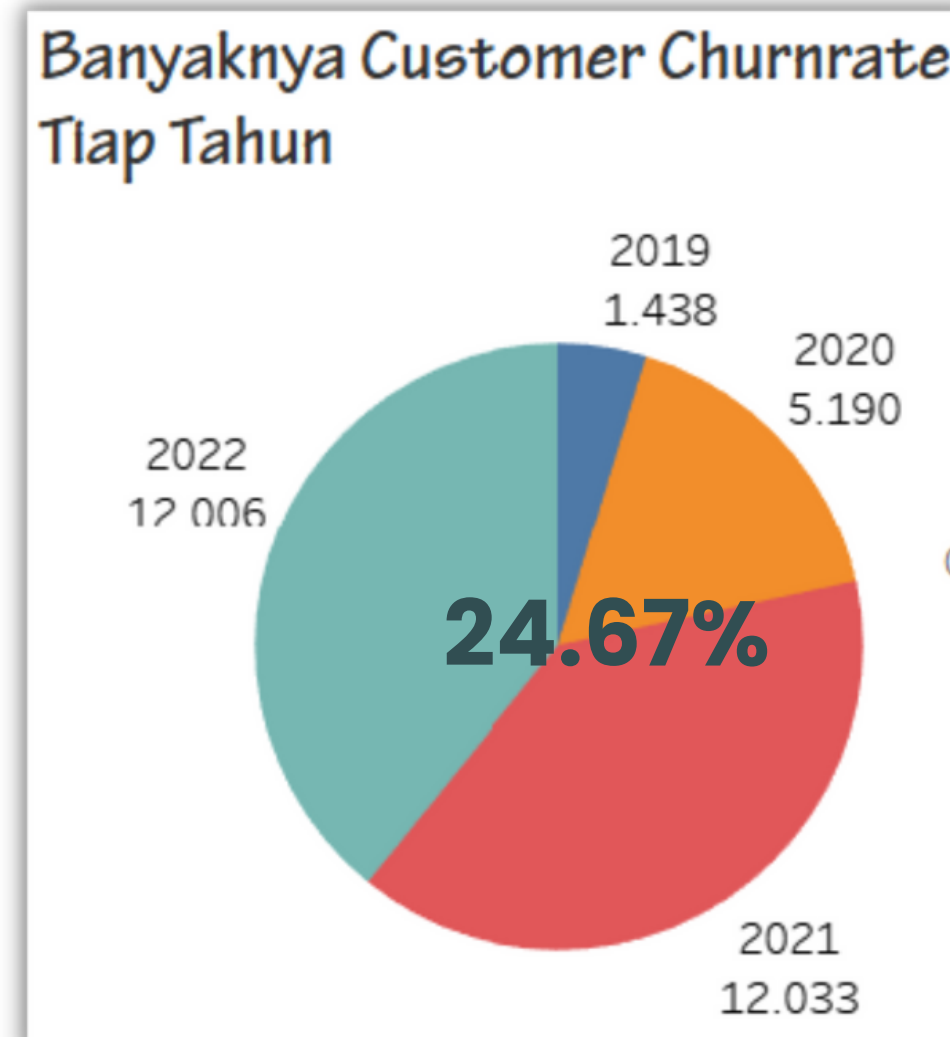
- **SALES**



Dilihat dari timeseries banyaknya pemesanan yang masuk per bulan, dapat disimpulkan bahwa jumlah pemesanan memiliki tren yang stabil naik.

## HASIL DAN PEMBAHASAN

- **SALES**



- Tingkat customer churn rate pada penjualan produk sebesar 24.67%
- Gender wanita lebih sering membatalkan/mengembalikan pemesanan

# HASIL DAN PEMBAHASAN

## • SALES

Hipotesis untuk masing-masing  $\beta_1$ ,  $\beta_2$ , dan  $\beta_3$ :

- $H_0: \beta_1 = 0$
- $H_0: \beta_2 = 0$
- $H_0: \beta_3 = 0$
- $H_1: \beta_1 \neq 0$
- $H_1: \beta_2 \neq 0$
- $H_1: \beta_3 \neq 0$

Dimana ,

$\beta_1$  : Variabel Gender

$\beta_2$  : Variabel Jumlah barang yang dipesan

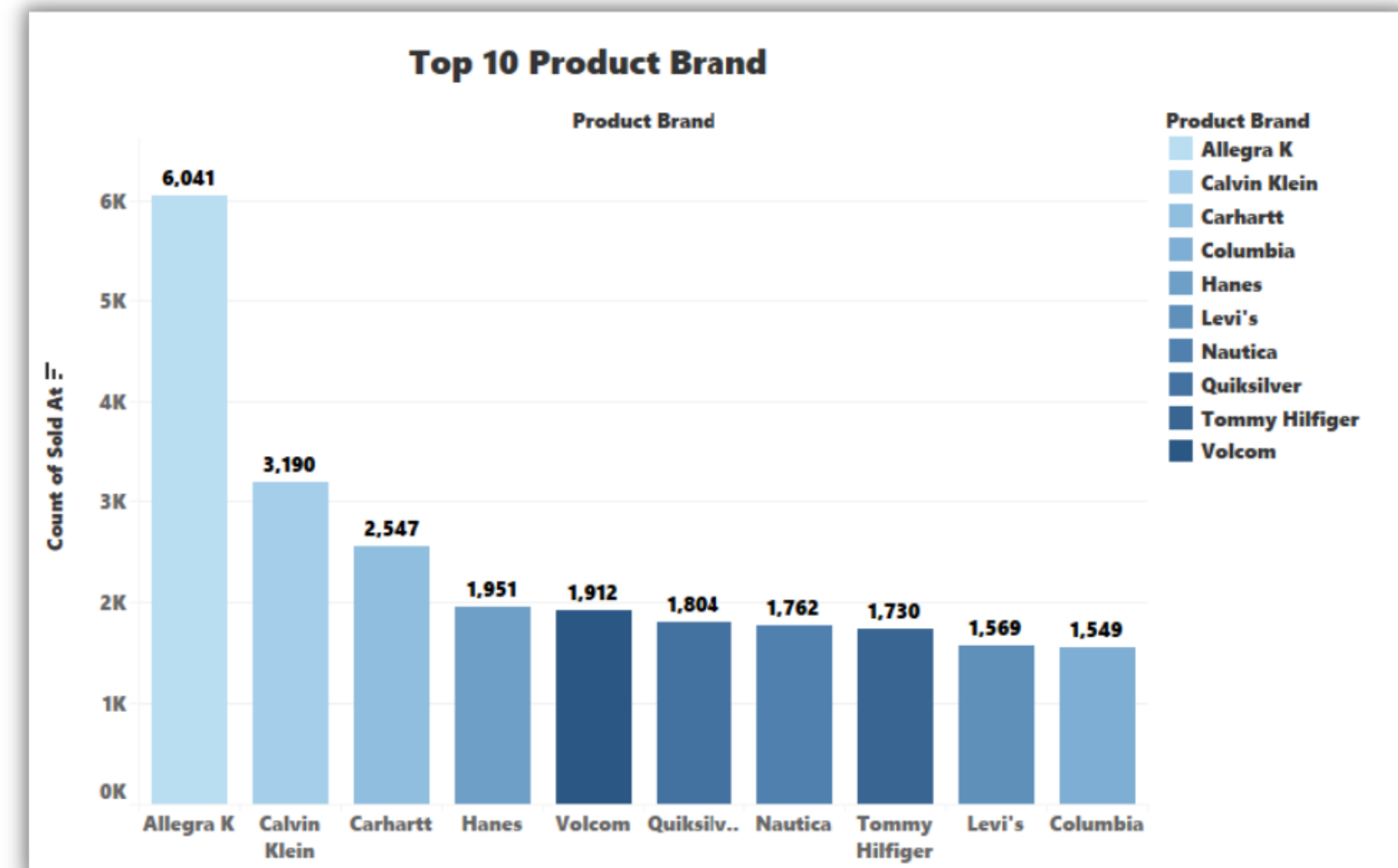
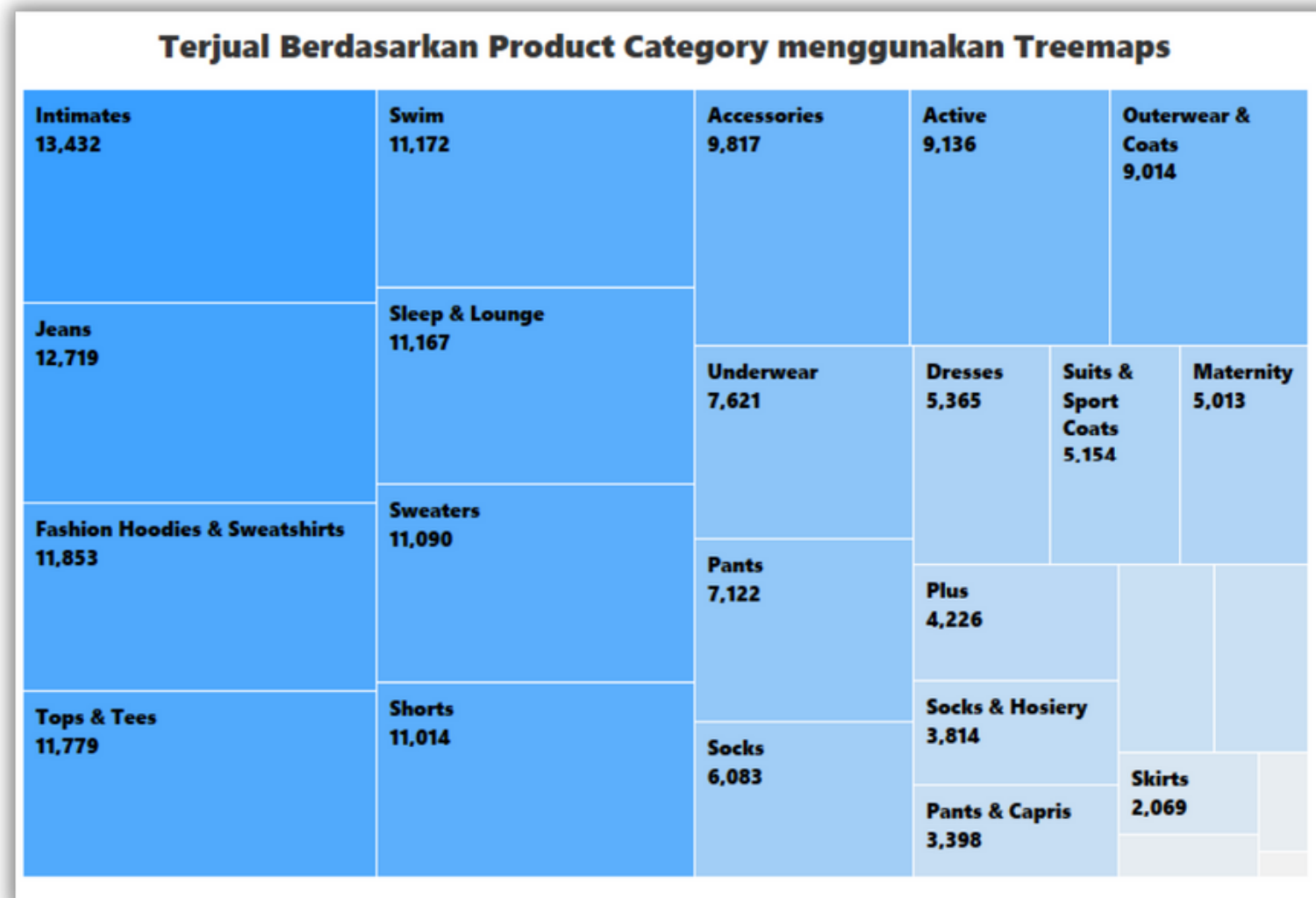
$\beta_3$  : Total harga yang dipesan

Logit Regression Results						
=====						
Dep. Variable:	status_0	No. Observations:	112060			
Model:	Logit	Df Residuals:	112057			
Method:	MLE	Df Model:	2			
Date:	Tue, 21 Jun 2022	Pseudo R-squ.:	inf			
Time:	13:55:19	Log-Likelihood:	-2.0845e+07			
converged:	True	LL-Null:	0.0000			
Covariance Type:	nonrobust	LLR p-value:	1.000			
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
gender_1	-0.4041	0.012	-32.386	0.000	-0.429	-0.380
num_of_item	-0.5383	0.008	-69.377	0.000	-0.553	-0.523
total_purchases	0.0002	8.91e-05	1.936	0.053	-2.17e-06	0.000
=====						

- Diperoleh nilai p-value masing-masing variabel dari analisis regresi logistik untuk mengetahui faktor yang mempengaruhi status pemesanan.

# HASIL DAN PEMBAHASAN

## • PRODUCT

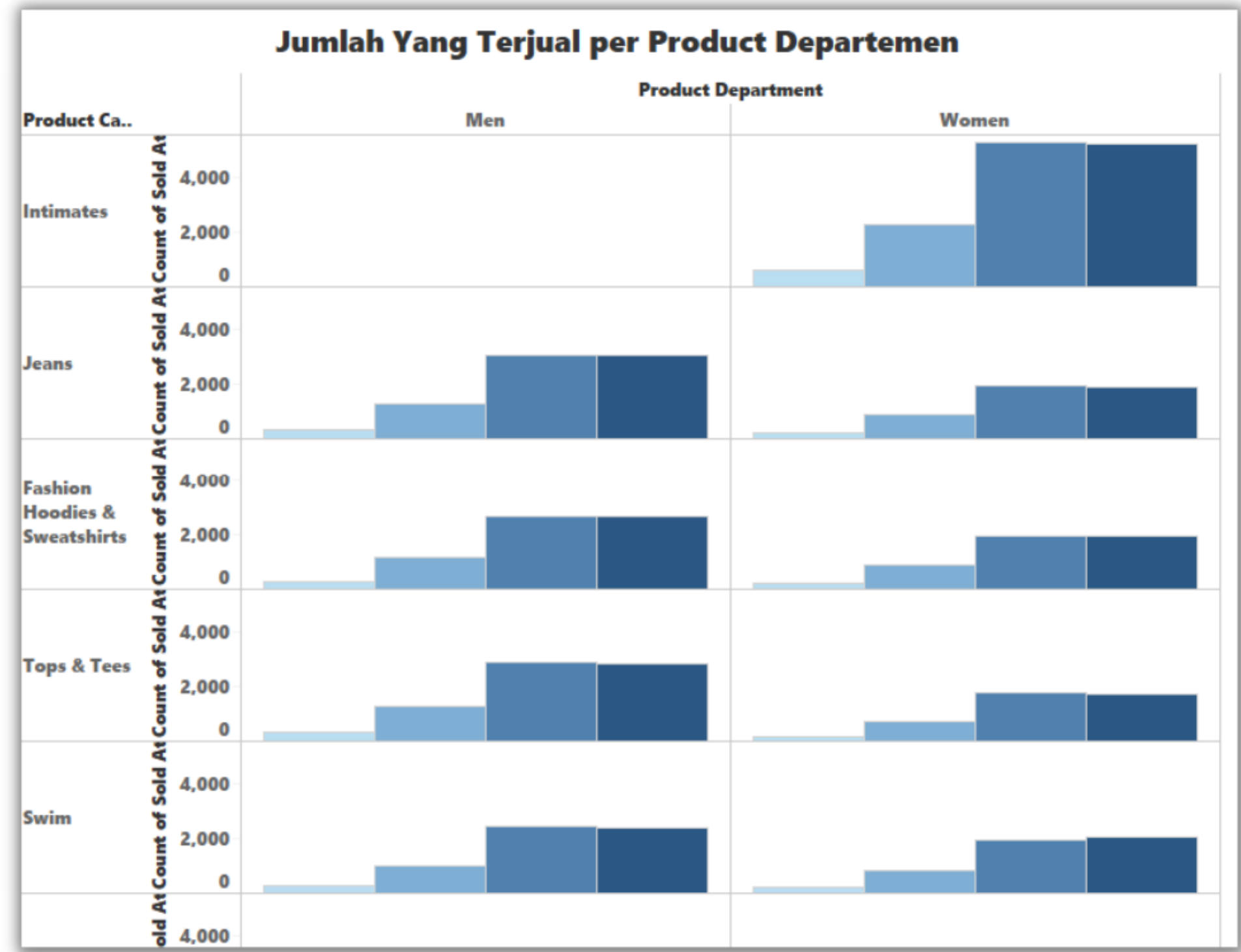


- Berdasarkan data diatas dapat diambil kesimpulan bahwa product yang paling laris terjual pada product category adalah intimates dengan jumlah terjual 13.432/ 4 tahun. sedangkan pada Product Brand Allegra K menempati urutan pertama dengan total penjualan 6,041



# HASIL DAN PEMBAHASAN

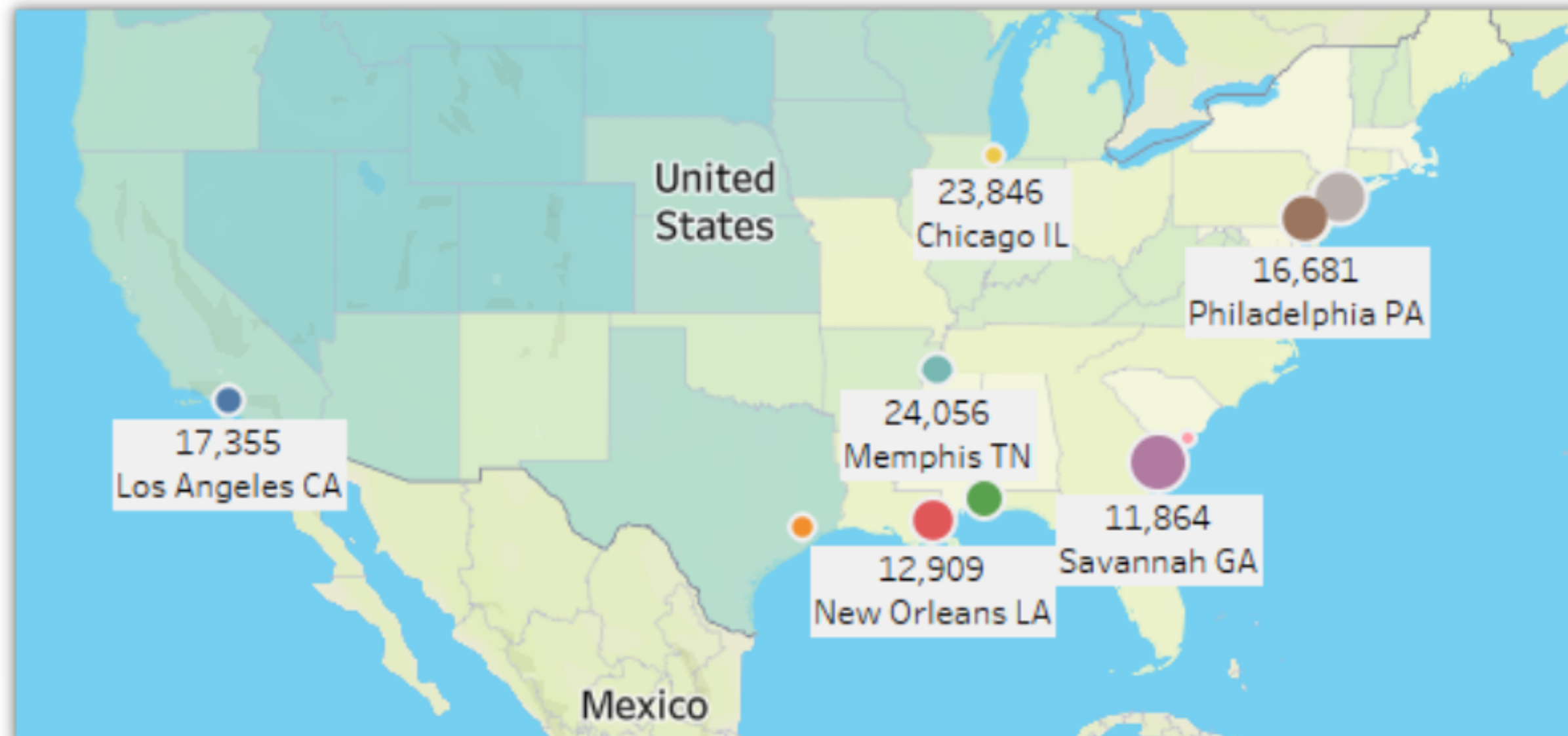
- PRODUCT



Dapat disimpulkan bahwa **laki-laki cenderung lebih banyak membeli barang dibandingkan dengan wanita.**

## HASIL DAN PEMBAHASAN

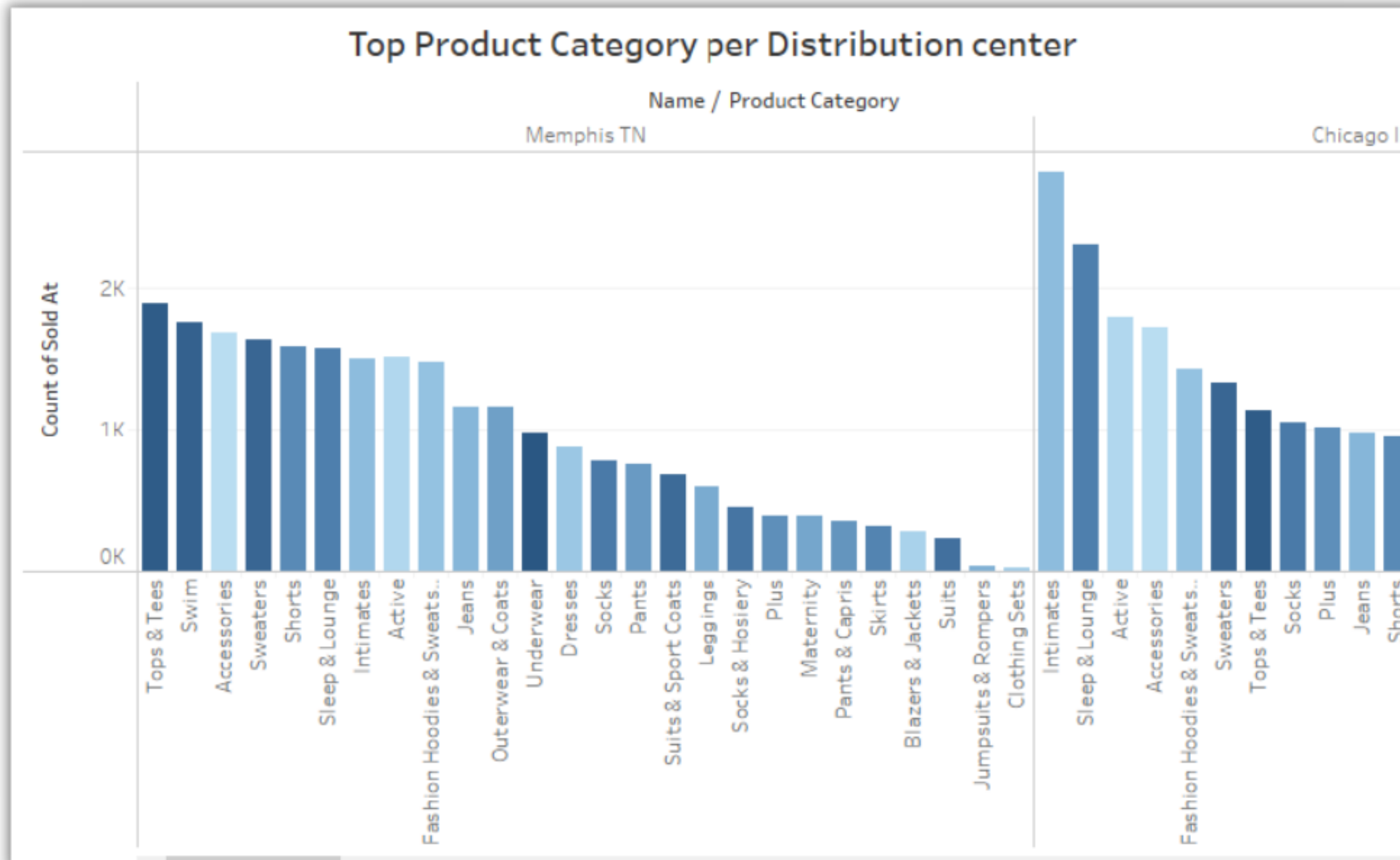
- **PRODUCT**



Berdasarkan data yang di analisis. Mamphis menempati urutan pertama dengan total penjualan selama 4 tahun yaitu 24,056 sehingga dapat disimpulkan bahwa memphis membutuhkan stock produk lebih banyak dibanding dengan distribusi center lain

# HASIL DAN PEMBAHASAN

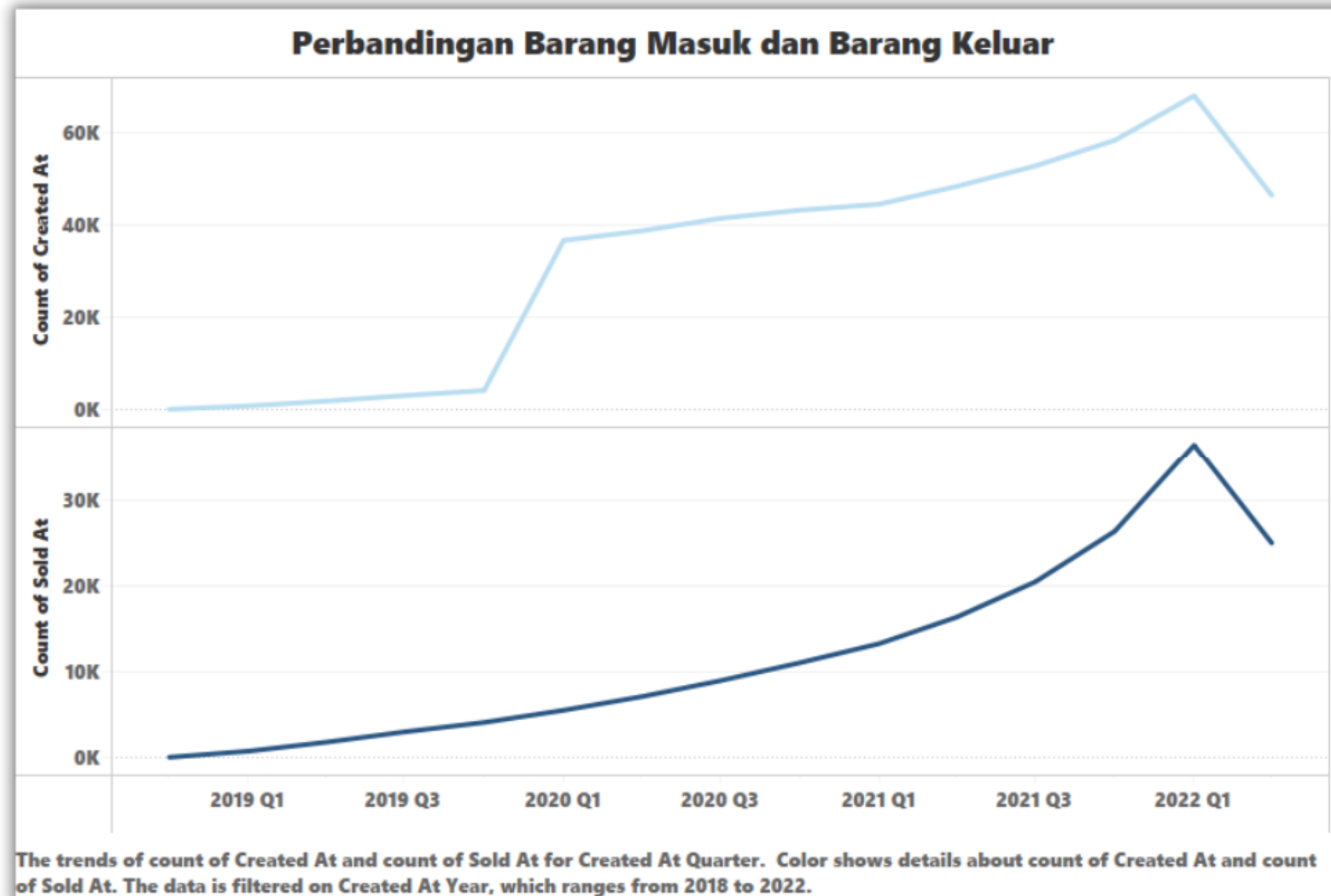
## • PRODUCT



Pada distribution center rata-rata barang yang paling diminati adalah **Tops and Tees, Intimates dan jeans.** sedangkan barang yang kurang diminati adalah **Clothing sets**

# HASIL DAN PEMBAHASAN

- **PRODUCT**



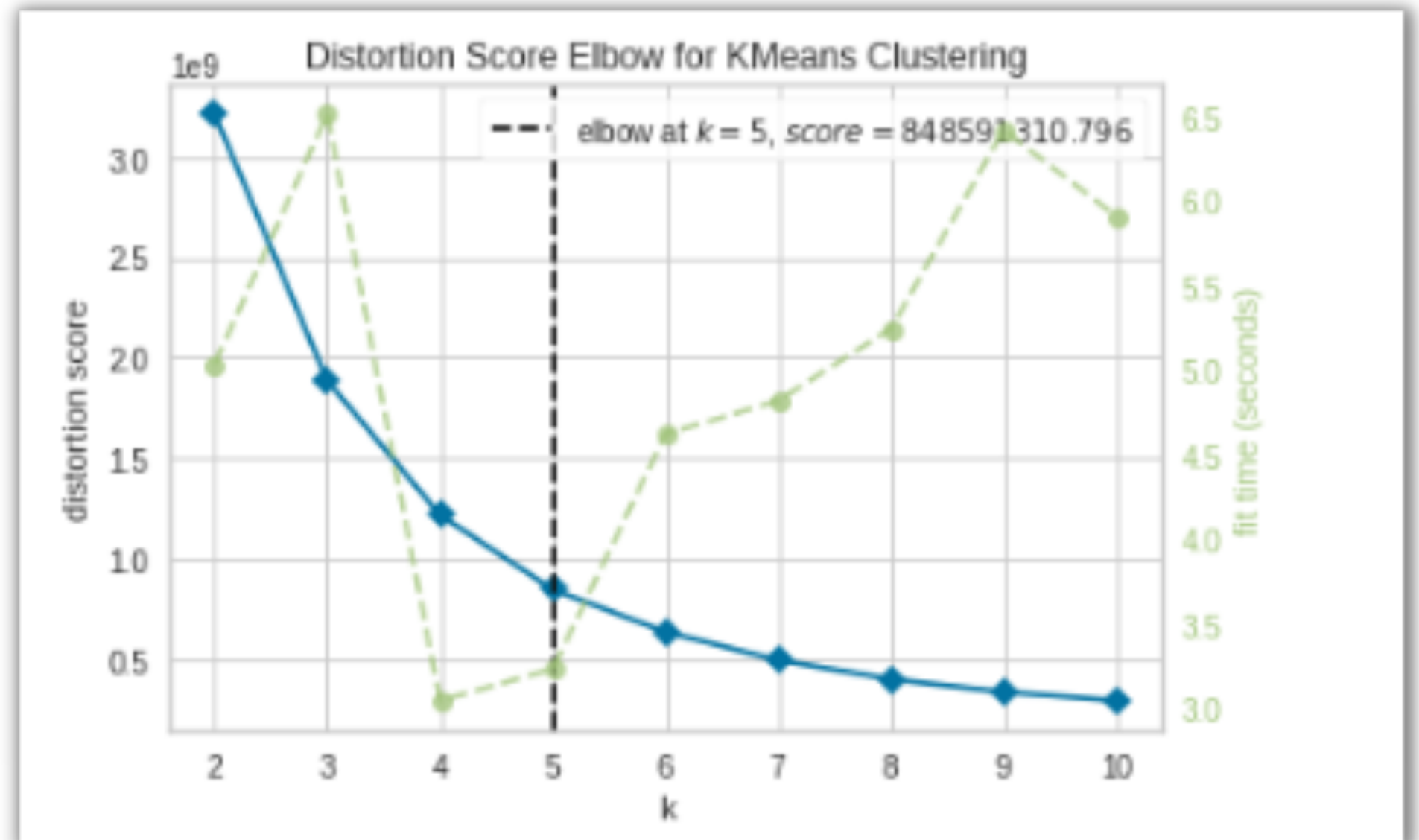
Data penjualan produk pada tahun 2019 **Q1 -Q4** sangat baik yaitu **100%**. kemudian **mengalami penurunan** pada Q1 tahun 2020 dengan total **penurunan 75%**. kemudian pada tahun **2022 Q1** penjualan kembali naik dan **mencapai 53%**

# HASIL DAN PEMBAHASAN

- **MARKETING**

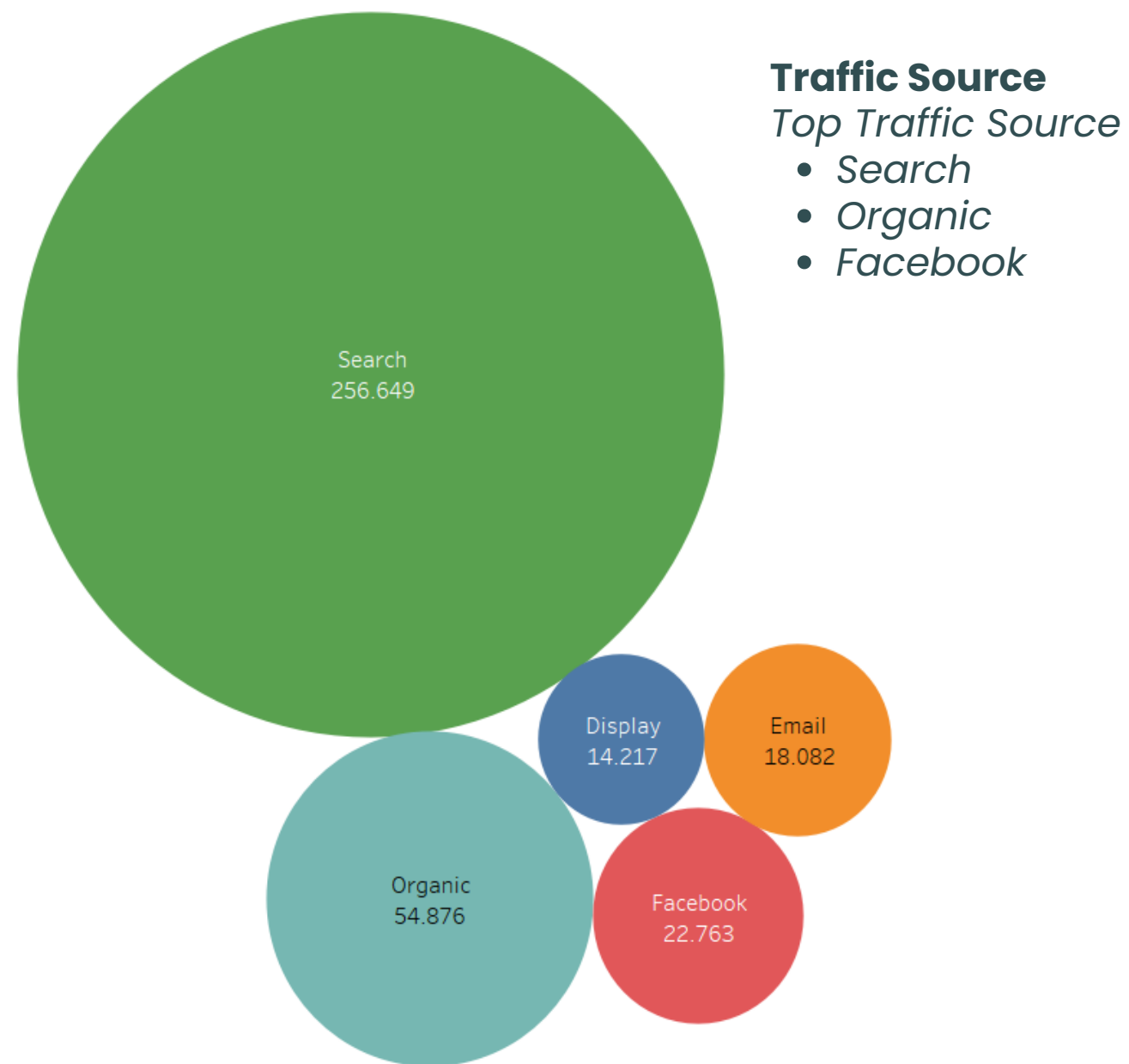
setelah dilakukan tahapan metodologi, didapat bahwa K Means Cluster dengan Elbow Methode didapat 5 buah kluster dengan nilai shiluite score 70%. Variabel yang terkluster meliputi

- **Umur**
- **Jumlah Barang yang dibeli**
- **Harga Barang**

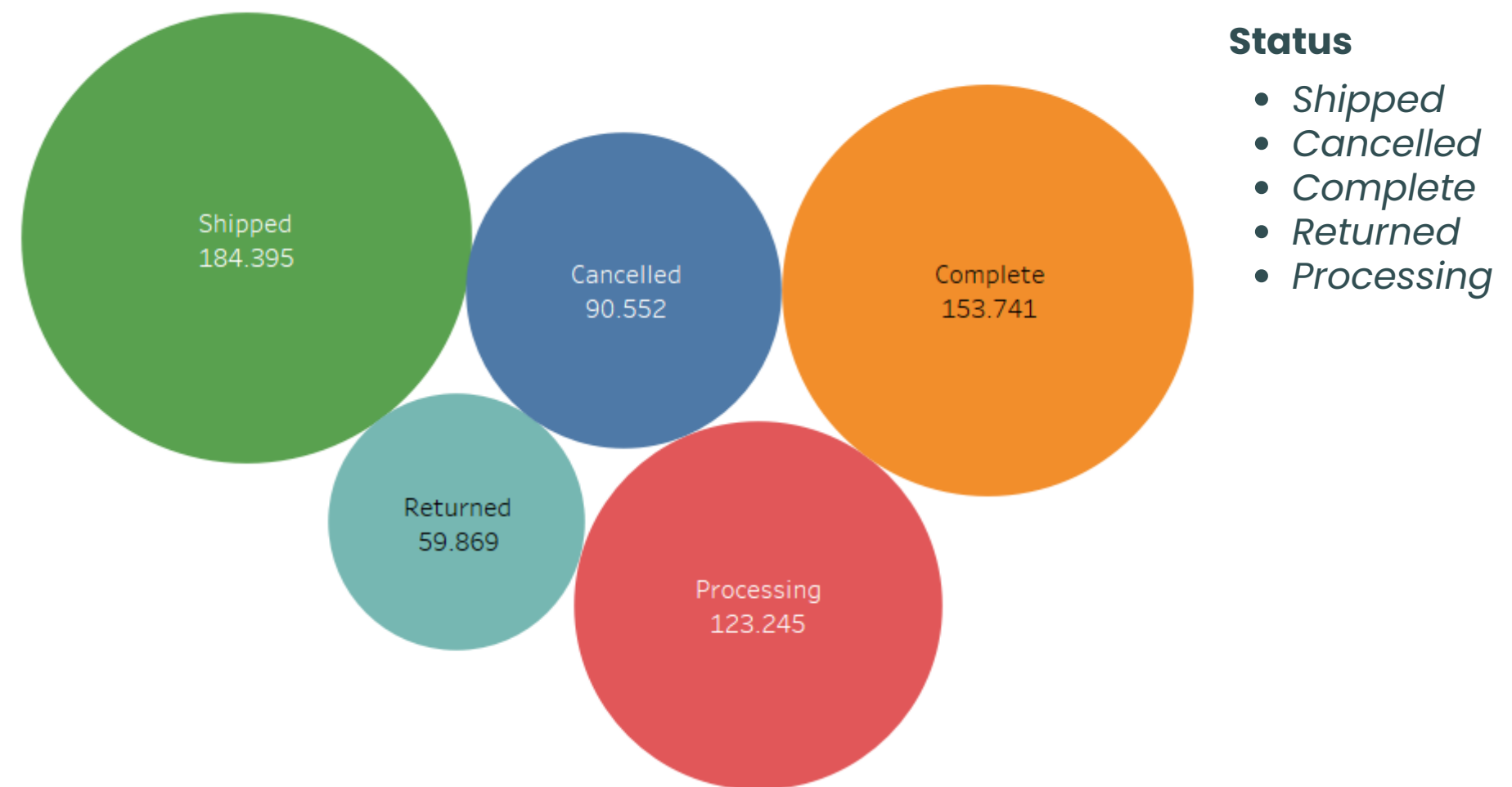


# HASIL DAN PEMBAHASAN

## ● MARKETING



Pengguna mengakses ecommerce paling banyak menggunakan **Search, Organic, dan Facebook.**

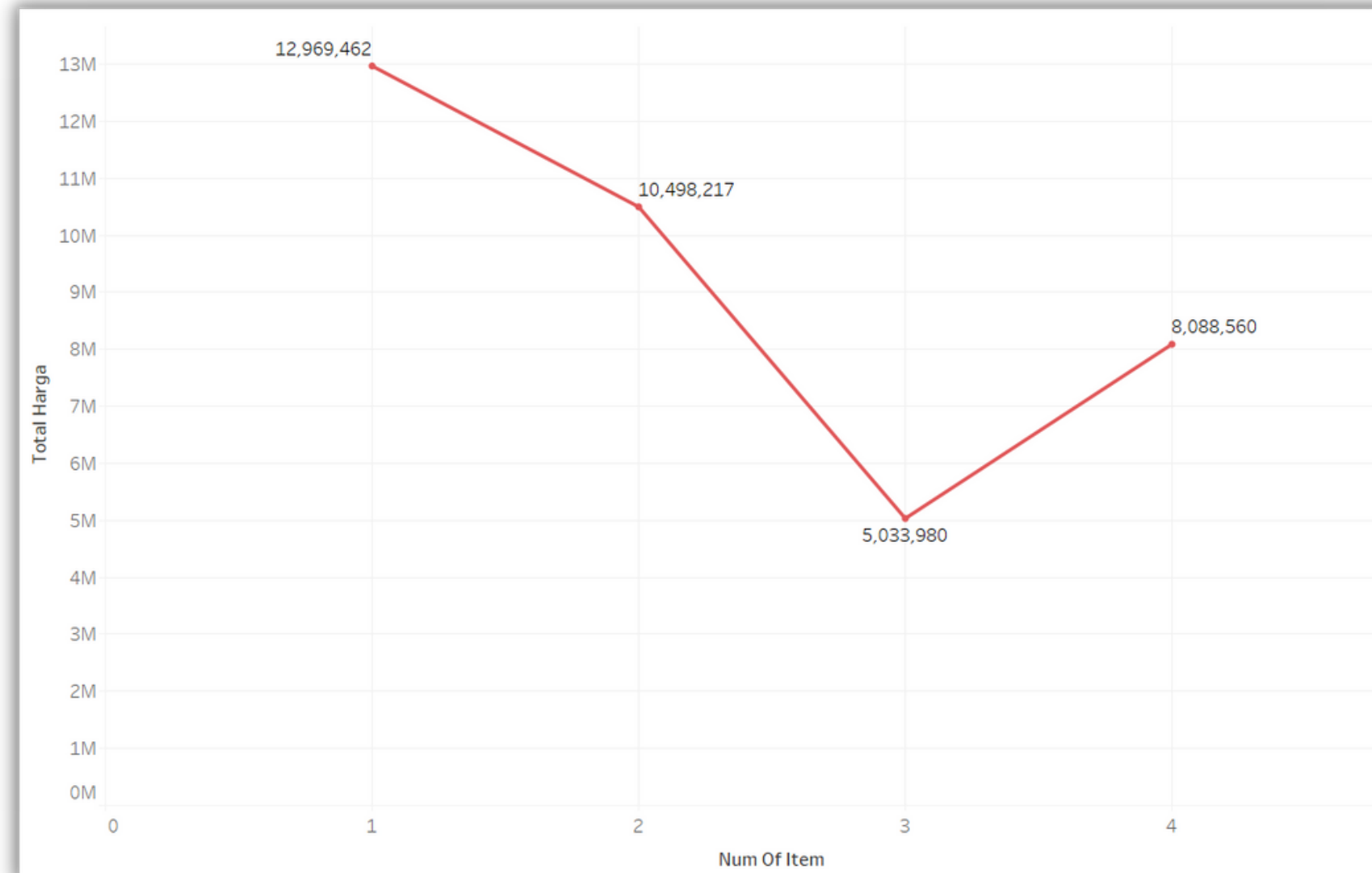


Barang yang *complete*, shipped, dan *processing* berjumlah **398.162** buah.  
Total barang keseluruhan berjumlah **548.613** buah



# HASIL DAN PEMBAHASAN

## ● MARKETING



### Jumlah barang dan total Harga

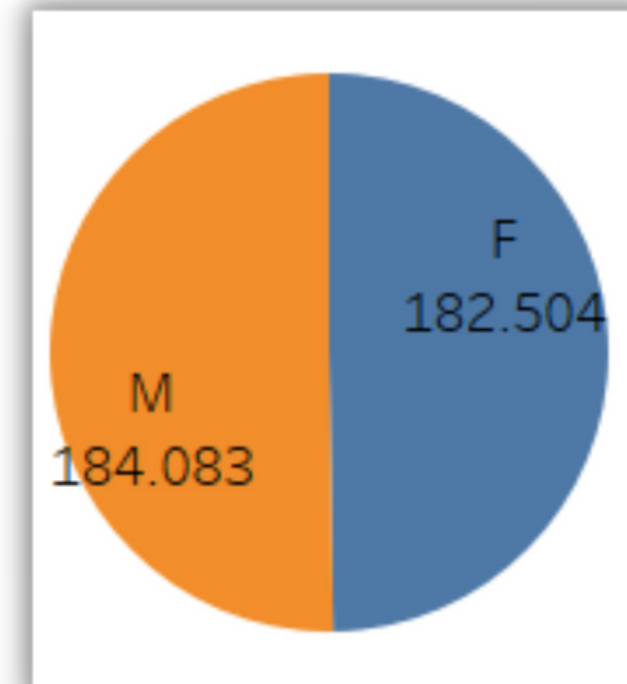
Barang

1 : 12.980.462 (Barang: 217.055)

2 : 10.498.217 (Barang: 175.250)

3 : 5.033.980 (Barang: 84.343)

4 : 8.088.560 (Barang: 135.104)



### Gender

Male : 184.083

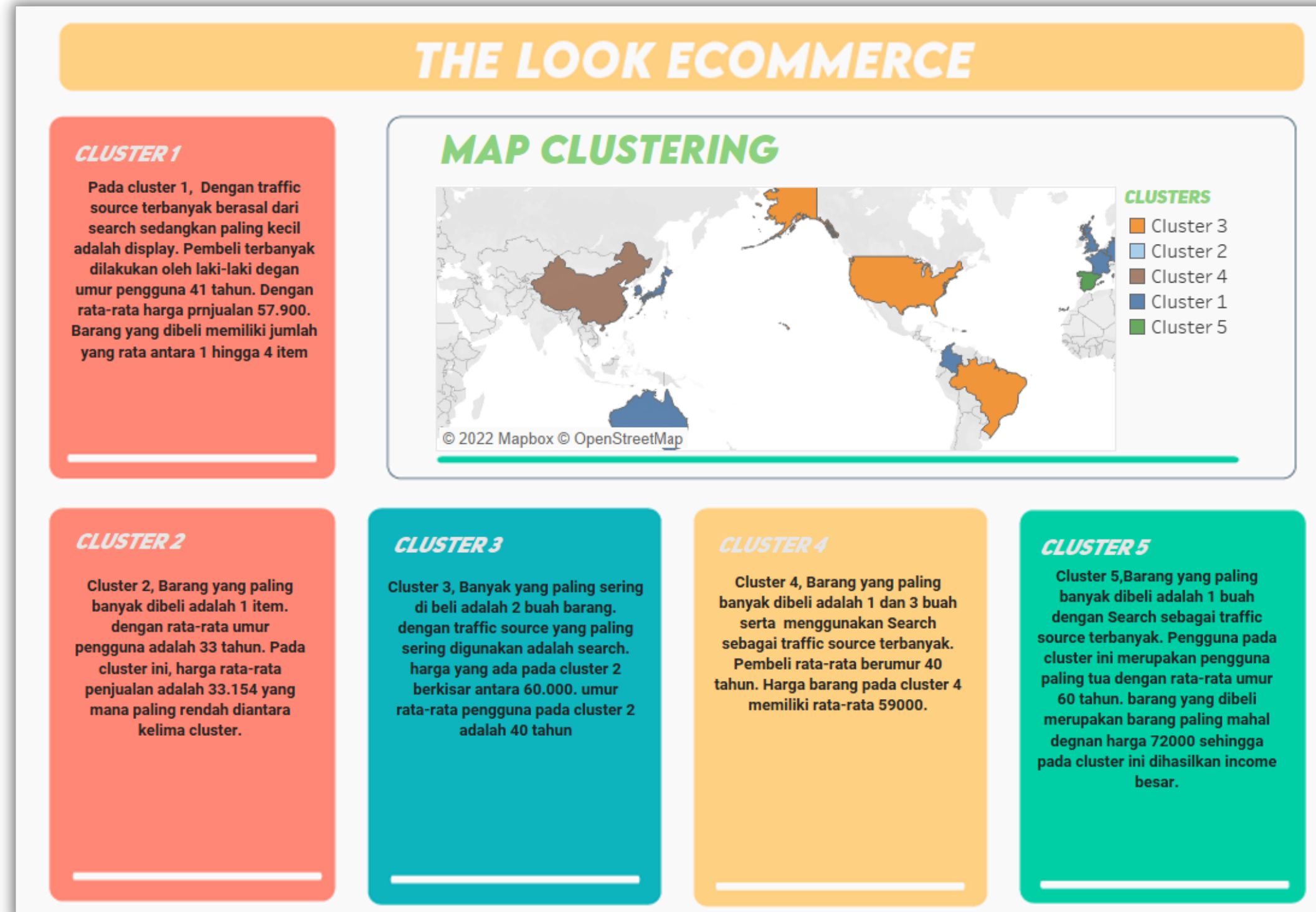
Female : 182.504

Pembelian yang dilakukan oleh *user* lebih banyak membeli dengan kuantitas barang berjumlah **1 buah**. dengan total transaksi paling tinggi hingga **12 juta**. sedangkan paling sedikit pada jumlah barang 3 buah yang mana total transaksi sejumlah 5 juta USD.



# HASIL DAN PEMBAHASAN

## MARKETING



# HASIL DAN PEMBAHASAN

## • HUMAN RESOURCES

### Hypotesis Testing

$\beta_1$  = Age

$\beta_2$  = length\_service

$\beta_3$  = Gender

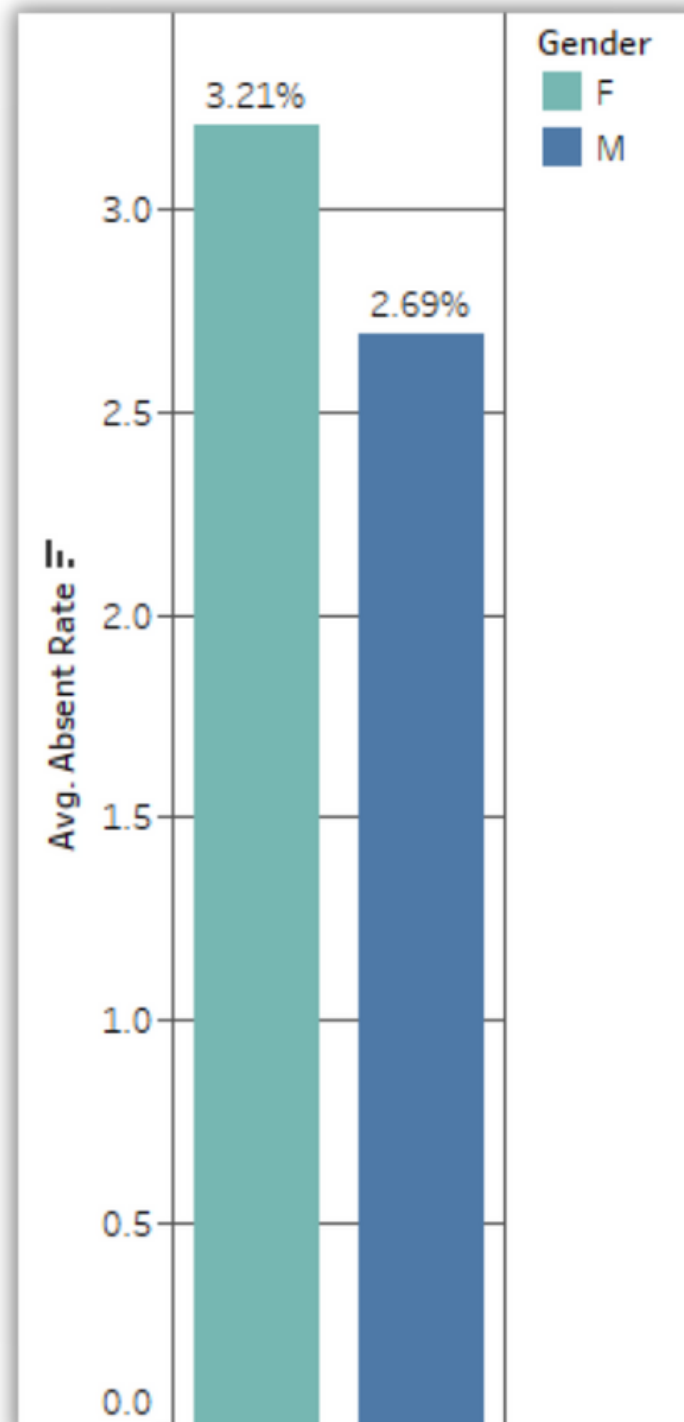
- $H_0: \beta_1 = 0$
- $H_1: \beta_1 \neq 0$
- $H_0: \beta_2 = 0$
- $H_1: \beta_2 \neq 0$
- $H_0: \beta_3 = 0$
- $H_1: \beta_3 \neq 0$

Logit Regression Results						
=====						
Dep. Variable:	absent_hour	No. Observations:	7502			
Model:	Logit	Df Residuals:	7499			
Method:	MLE	Df Model:	2			
Date:	Tue, 21 Jun 2022	Pseudo R-squ.:	0.02869			
Time:	18:24:08	Log-Likelihood:	-3946.6			
converged:	False	LL-Null:	-4063.1			
Covariance Type:	nonrobust	LLR p-value:	2.394e-51			
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
age	1.5057	0.038	39.659	0.000	1.431	1.580
length_service	-0.0609	0.028	-2.208	0.027	-0.115	-0.007
gender_F	0.0993	2.19e+06	4.53e-08	1.000	-4.3e+06	4.3e+06
gender_M	-0.0993	2.19e+06	-4.53e-08	1.000	-4.3e+06	4.3e+06
=====						

- Berdasarkan tabel diatas dengan nilai P-Value tersebut maka dapat disimpulkan bahwa **Usia dan Lama Bekerja Staff berpengaruh terhadap Jam Absensi Staff.** sedangkan untuk **Jenis Kelamin tidak berpengaruh sama sekali.**

# HASIL DAN PEMBAHASAN

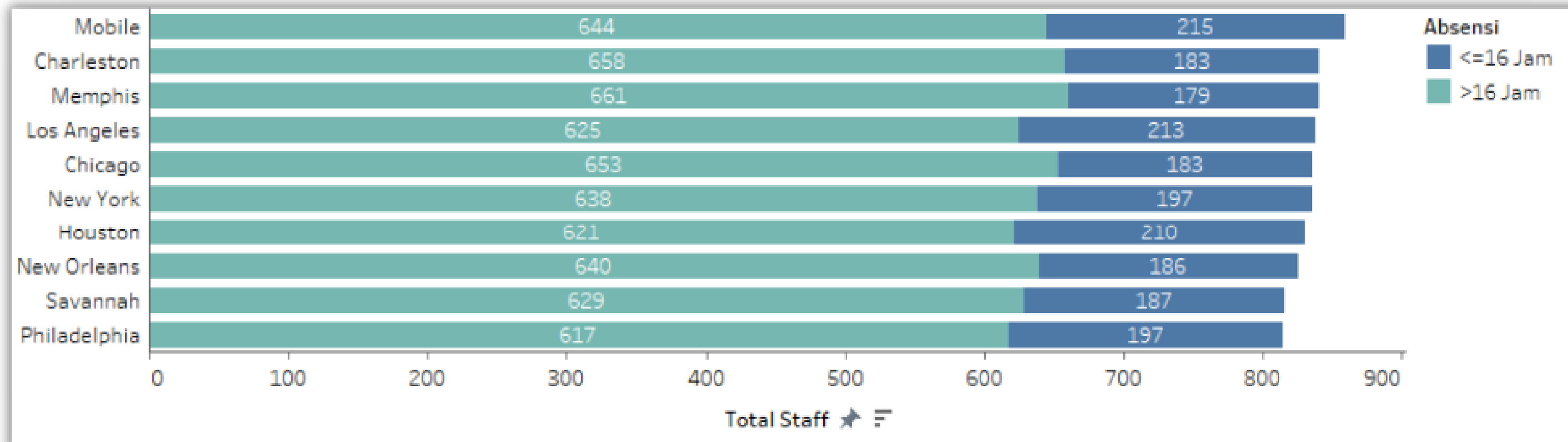
## • HUMAN RESOURCES



- Tingkat Absensi (Absent Rate) Staff Wanita lebih tinggi dibandingkan Pria. Hal ini menggambarkan bahwa **Tingkat Produktivitas Staff Pria Lebih tinggi** dibandingkan Staff Wanita
- Tingkat Absensi (Absent Rate) Pada Tempat distribusi Memphis **lebih tinggi** jika dibandingkan dengan tempat distribusi yang lain karena Distribusi (Mempis) memiliki **rata-rata Usia karyawan yang lebih tinggi yaitu 42.752 Tahun** dengan Rata-rata lama staff bekerja yaitu 4.7 Tahun. Hal ini berpengaruh terhadap produktivitas pada tempat distribusi tersebut.

# HASIL DAN PEMBAHASAN

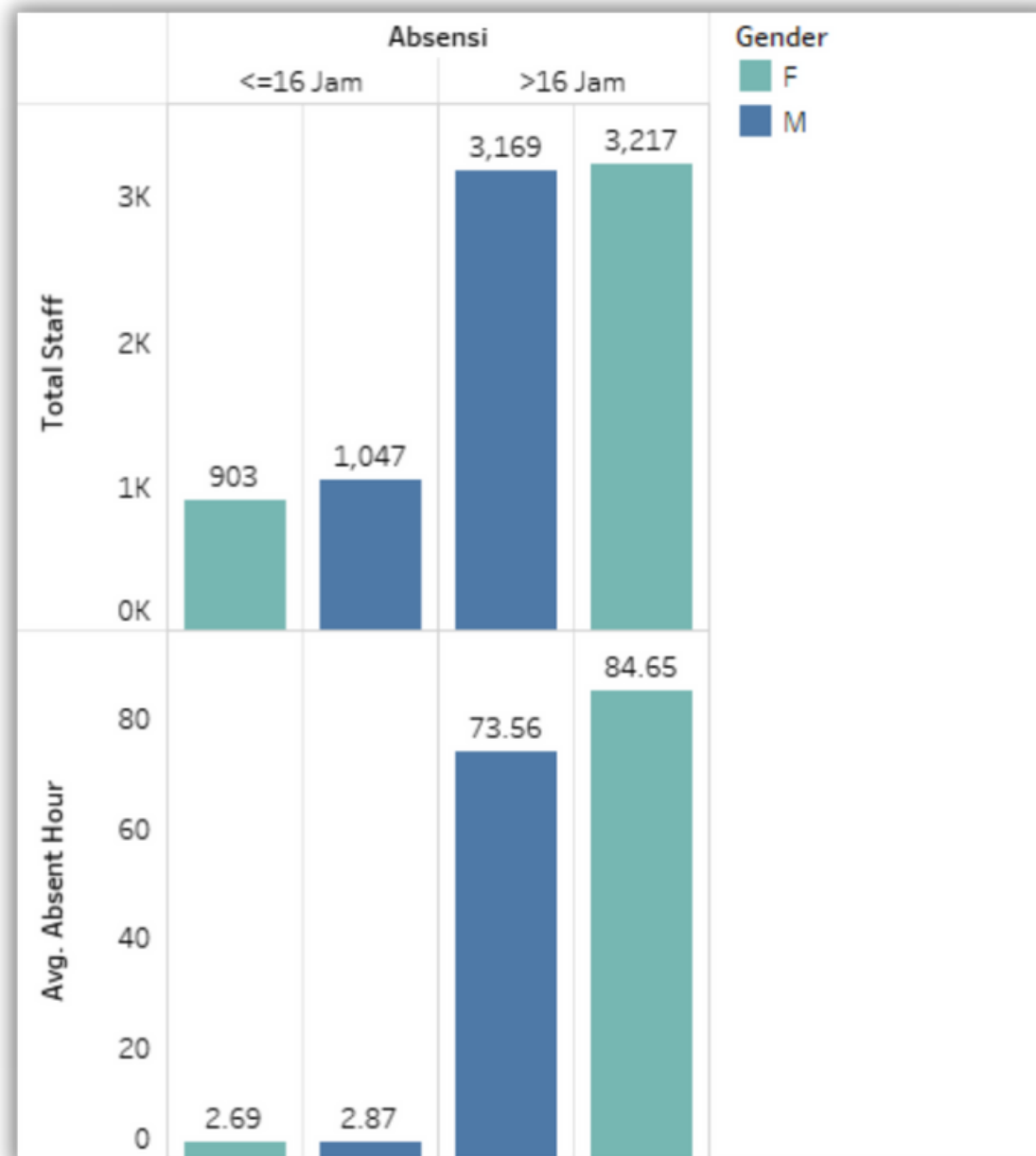
- **HUMAN RESOURCES**



- Rata-rata karyawan masih banyak yang **absen lebih dari 2 hari kerja (> 16 jam) dalam setahun** pada semua distribusi. Hal ini tentunya sangat berdampak pada produktivitas perusahaan

# HASIL DAN PEMBAHASAN

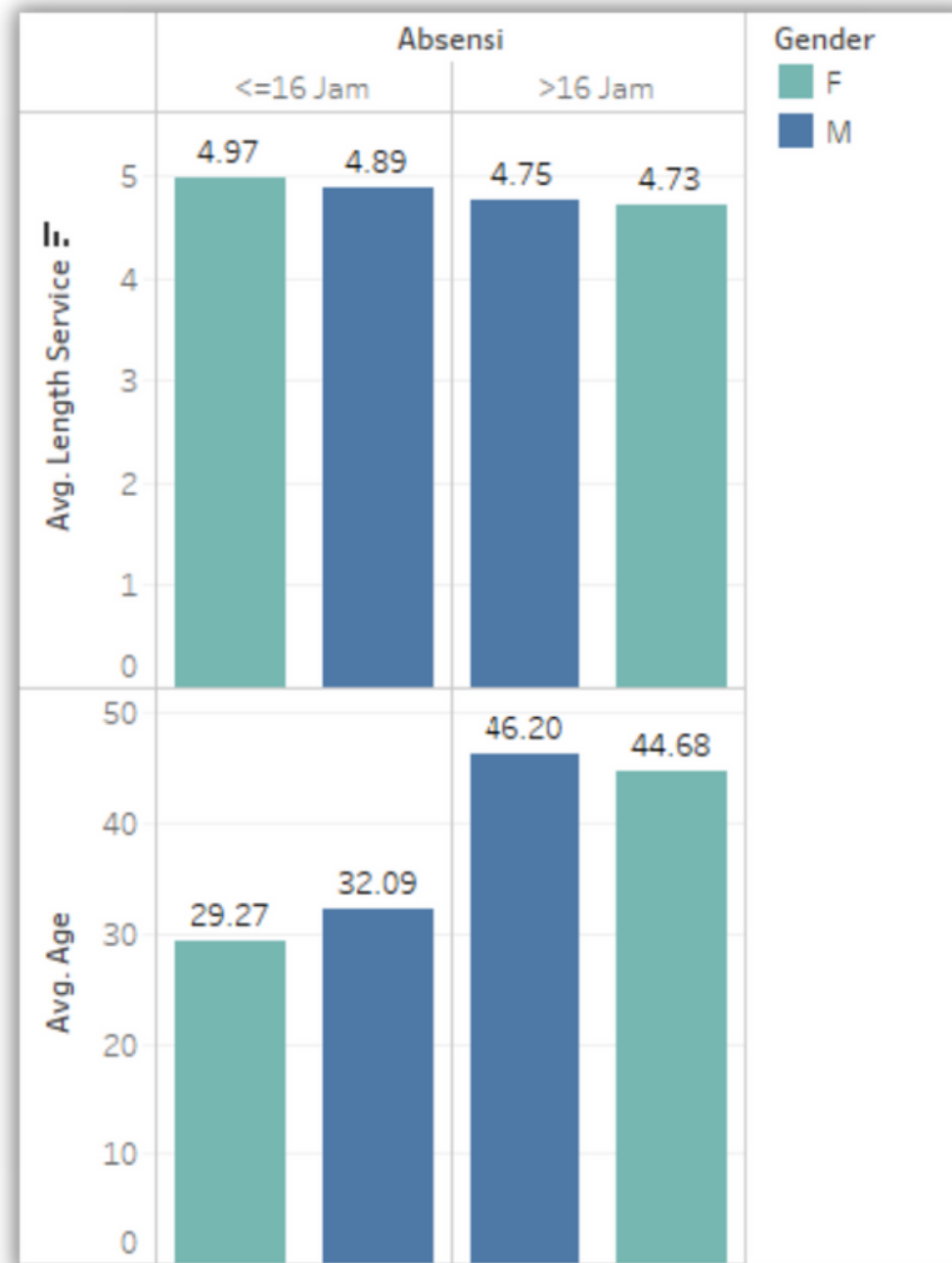
## • HUMAN RESOURCES



- Staff yang absen >16 jam (>2 hari) kebanyakan adalah **Staff wanita** dengan rata-rata absensi yaitu **84.65 jam (10 Hari)**
- Staff yang absen ≤16 jam (≤2 hari) juga kebanyakan adalah staff wanita dengan rata-rata absensi yaitu **2.69 jam (<1 hari)**

# HASIL DAN PEMBAHASAN

## • HUMAN RESOURCES



- Staff Pria dan Wanita yang absen >16 jam (>2 hari) sudah bekerja lebih dari 4 tahun (senior/lama) dengan rata-rata usia 44 & 46 tahun yang dimana **jauh lebih tua usianya** dibandingkan staff pria dan wanita yang absensinya <=16 jam (<=2 hari) yaitu 32 & 28 tahun



# KESIMPULAN

## • SALES

Berdasarkan data, didapatkan kesimpulan :

- Jumlah banyaknya pemesanan yang masuk memiliki **tren yang stabil** naik per bulannya.
- Tingkat customer churn rate pada penjualan produk terjadi **sebesar 24.67%**.
- Tingkat customer churn rate **terbesar** terjadi pada tahun 2021.
- Gender wanita lebih sering **membatalkan/mengembalikan** pemesanan.

Berdasarkan hasil analisis regresi logistik, didapatkan kesimpulan :

- Secara nyata variabel total harga penjualan produk gagal tolak  $H_0$ , sehingga tidak mempengaruhi status pemesanan.
- Secara nyata variabel gender dan total produk yang dipesan adalah tolak  $H_0$ , sehingga secara signifikan kedua variabel tersebut mempengaruhi status pemesanan.



# KESIMPULAN

## • MARKETING

Berdasarkan kluster, didapatkan kesimpulan :

- Laki-laki lebih banyak membeli barang daripada wanita.
- Pembelian paling banyak dilakukan melalui *traffic source* **Search, organic, dan facebook.**
- **Cluster 2** merupakan kelompok *young age* (33 tahun) dengan pembelian barang berjumlah 1 buah dengan harga yang murah. (33.000)
- **Cluster 5** merupakan kelompok *old age* (50 tahun) dengan pembelian barang berjumlah 1 item dengan harga barang yang mahal (72.000)
- **Cluster 1, 3, dan 4** merupakan kelompok *middle age* (40 tahun) dengan Pembelian barang berjumlah 1 dan 2 dengan harga 60.000.

# KESIMPULAN

- **PRODUCT**

Berdasarkan data yang telah dianalisis didapatkan kesimpulan bahwa

- **Pada gender wanita** barang yang paling diminati adalah product category intimates dengan total penjualan **13.432/ 4 tahun**.
- **Pada gender laki-laki** barang yang paling diminati adalah product category Jeans dengan total penjualan **7.764/ 4 tahun**.
- Gender laki-laki memiliki **kecenderungan** membeli banyak barang daripada wanita, seperti : Tops and Tees, swim dan sweeter.
- Brand yang paling laris adalah **Allegra K** dengan jumlah terjual 6.041
- Pada Distribution center produk yang paling diminati adalah Tops and Tees, Intimates, Jeans, Swims, fashion Hoodies dan Underwear. sedangkan produk yang kurang diiminati adalah Clothing sets

# KESIMPULAN

- **HUMAN RESOURCES**

Berdasarkan data yang telah dianalisis didapatkan

- **Usia dan Lama Bekerja Staff berpengaruh terhadap Jam Absensi Staff.** sedangkan untuk **Gender tidak berpengaruh sama sekali** (*berdasarkan nilai P-Value*)
- Staff Wanita **lebih sering asben** dibandingkan Staff Pria
- Tempat distribusi Mempis memiliki staff yang lebih sering absen dibandingkan tempat lainnya Hal ini dikarenakan **rata-rata usia staff yang lebih tinggi yaitu 42.7 Tahun**.
- Rata-rata Staff pada setiap distribusi center masih memiliki **jam absensi diatas 2 hari** dalam setahun.
- Staff Wanita dan Pria yang jam absensinya diatas 2 hari (> 16 jam) memiliki **rata-rata usia 40 & 45 tahun** dimana **lebih tua** dibandingkan yang absensi dibawah 2 hari yaitu **29 & 32 tahun**

# SARAN

- **SALES**

Untuk **menurunkan tingkat customer churn rate**, alangkah baiknya The Look e-commerce :

- Meningkatkan intensitas dalam memberikan **diskon atau *gift*** bagi **customer yang sering berbelanja** kembali di The Look e-commerce (*customer retention*).
- Meningkatkan promosi penjualan dengan melihat customer yang akan dituju.

# SARAN

## • MARKETING

Diberikan Saran sebagai berikut :

- **Cluster 2** merupakan kelompok *young age* (33 tahun) dengan pembelian barang berjumlah 1 dengan harga yang murah. (33.000). sebaiknya **diberikan diskon** supaya user membeli barang dengan jumlah yang lebih banyak lagi. Pada usia **young age cenderung tertarik dengan promo-promo yang diberikan.**
- **Cluster 5** merupakan kelompok *old age* (50 tahun) dengan pembelian barang berjumlah 1 buah dengan harga barang yang mahal (72.000). sebaiknya diberikan juga **opsi paket bundling (terdiri paket produk)**. Karena umur **old age diasumsikan orang yang telah berkeluarga** sehingga juga membeli barang untuk anak dan istrinya. Serta memberikan slogan banner pada laman ecommerce seperti "**Baju Baru Anak Senang**". Hal itu mendorong orang untuk mengingat anak dan membelikan baju.
- **Cluster 1, 3, dan 4** merupakan kelompok *middle age* (40 tahun) dengan pembelian barang berjumlah 1 dan 2 dengan harga 60.000. **Perlu diberikan promo menarik** terkait dengan bundling **2 barang sehingga lebih antusias dalam membeli.**
- Pada Pembelian, **ditambahkan proteksi pengiriman barang** supaya barang tidak rusak dan memperkecil terjadinya **returned.**
- Melakukan Marketing digital dengan kuantitas promosi yang lebih massif dan kontinyu pada **top traffic source (search, organic, dan facebook)** sehingga akan lebih banyak yang membeli karena sifat impulsif yang dimiliki oleh user.

# SARAN

- **PRODUCT**

- Untuk meningkatkan penjualan dan profit disarankan perusahaan the look e-commerce **lebih memperbanyak stok** pada Category Produk yang paling diminati
- Pada **Mamphis** diharapkan lebih memperbanyak **stok Tops and Tees**
- Pada **Chicago dan Houston** lebih memperbanyak **stok Intimates**
- Pada **Mobile AL, Philedelphia dan New Orleans** lebih memperbanyak **stok Jeans**
- Pada **New Jersey dan Charleston** lebih memperbanyak **Fashion Hoodies and Sweeters**
- di **Los Angles** lebih memperbanyak **Stok Produk Swim**
- dan di **Savanah** lebih memperbanyak stok **Underwear**
- kemudian untuk departemen sales diharapkan **dapat melakukan bundling** terhadap produk yang kurang diminati
- Untuk bagian marekting diharapkan melakukan **promosi yang lebih menarik** terkait barang di berbagai platform untuk menarik perhatian konsumen



# SARAN

- **HUMAN RESOURCES**

- Untuk meningkatkan produktivitas pada perusahaan maka :
  - Perlu dilakukan **recruitment** agar mendapatkan staff dengan **usia yang lebih produktif (20–40 Tahun)**.
  - Perlu juga dilakukan **penyempuraan kebijakan** yang mencakup prosedur yang harus diikuti **jika ketidakhadiran staffnya berlebihan (>16 jam atau 2 hari)**.
  - **Motivasi dan Apresiasi** juga perlu diberikan kepada staff karena tidak hanya membuat staff merasa **lebih percaya diri**, tetapi juga **meningkatkan produktivitas** di tempat kerja.



# THANK YOU.

*Data, Business Analytics and Operations Program  
@Ruangguru*