



OCTOBER 2023

TEST LOTTE

LƯƠNG TRƯỜNG PHƯỚC



TABLE

of contents

1. Task 1

01. Task 1

2. Task 2

01. Task 1

Task 1

<u>Số liệu 2019</u>					
	DNI	BDG	PTT	VTU	NTG
% member invoice	19	32	28	29	31
Total invoice	1.039.749	1.113.524	1.230.567	1.716.055	1.675.375
Frequency member	2,8	2,4	2,6	2,3	2,3
Retail sale	199.713.608.000	273.475.670.541	269.679.017.259	411.168.253.862	653.842.825.094

Từ dữ liệu trên phân loại 5 cửa hàng thành 2 nhóm dựa trên các số liệu về:

- % member invoice
- Total invoice
- Frequency member
- Retail sale



Cách làm

- Nhận định

Phân loại dữ liệu không có nhãn, không được phân loại cho trước => sử dụng phương pháp phân loại không giám sát

- Thuật toán sử dụng

K-means cluster

Các bước thực hiện

1

Nhập các thư viện cần thiết và tải dữ liệu
bằng Pandas

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from sklearn.preprocessing import StandardScaler
from yellowbrick.cluster import KElbowVisualizer
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_samples, silhouette_score
```

```
path = r'C:\Users\truon\OneDrive\Documents\TestLotte\Testing.xlsx'
df = pd.read_excel(path, "Data 1", header = 1, index_col=0)
```

Các bước thực hiện

2

Tinh chỉnh dữ liệu

```
df= df.transpose()  
df
```

	% member invoice	Total invoice	Frequency member	Retail sale
DNI	18.531254	1039749.0	2.8	1.997136e+11
BDG	31.906212	1113524.0	2.4	2.734757e+11
PTT	27.875456	1230567.0	2.6	2.696790e+11
VTU	28.704826	1716055.0	2.3	4.111683e+11
NTG	30.565428	1675375.0	2.3	6.538428e+11

Các bước thực hiện

3

Chuẩn hóa dữ liệu

```
scaler = StandardScaler()  
scaled_data_zscore = scaler.fit_transform(df)
```

```
df_zscore = pd.DataFrame(scaled_data_zscore, columns=df.columns, index = df.index)  
df_zscore
```

	% member invoice	Total invoice	Frequency member	Retail sale
DNI	-1.908407	-1.106261	1.650274	-1.002727
BDG	0.932303	-0.847418	-0.412568	-0.545775
PTT	0.076210	-0.436768	0.618853	-0.569295
VTU	0.252360	1.266587	-0.928279	0.307222
NTG	0.647534	1.123860	-0.928279	1.810576

Các bước thực hiện

4

Thực hiện phân cụm:

```
model = KMeans(n_clusters = 2, init = "k-means++",  
max_iter = 300, n_init = 10, random_state = 0) # Tạo mô hình K-Means với số cụm là 2  
y_kmeans = model.fit(df_zscore) # Thực hiện phân cụm  
labels = model.labels_ # Lấy nhãn của từng điểm dữ liệu  
centroids = model.cluster_centers_  
# Lấy các điểm trung tâm của cụm
```


Các bước thực hiện

5

In tọa độ của tâm 2 cụm

```
print(y_kmeans)
print(" Our cluster centers are as follows")
print(centroids)
```

```
KMeans(n_clusters=2, n_init=10, random_state=0)
Our cluster centers are as follows
[[ 0.44994692  1.19522338 -0.92827912  1.05889888]
 [-0.29996461 -0.79681559  0.61885275 -0.70593259]]
```

Các bước thực hiện

6

Kết quả phân cụm

```
df_res = df_zscore.copy()
df_res['cluster']=model.fit_predict(df_zscore)
df_res
```

```
c:\Users\truon\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:1382:
warnings.warn(
```

	% member invoice	Total invoice	Frequency member	Retail sale	cluster
DNI	-1.908407	-1.106261	1.650274	-1.002727	1
BDG	0.932303	-0.847418	-0.412568	-0.545775	1
PTT	0.076210	-0.436768	0.618853	-0.569295	1
VTU	0.252360	1.266587	-0.928279	0.307222	0
NTG	0.647534	1.123860	-0.928279	1.810576	0

Các bước thực hiện

7

Kiểm tra

```
sil_score = silhouette_score(df_zscore, labels, metric='euclidean')
print('Silhouette Score: %.3f' % sil_score)
from yellowbrick.cluster import SilhouetteVisualizer
model = KMeans(2)
visualizer = SilhouetteVisualizer(model)
visualizer.fit(df_zscore)
visualizer.poof()
```

```
Silhouette Score: 0.382
```


Task 2

Phân tích hiệu suất của các cửa hàng từ 2
nhóm đã phân tách

Period	Store	CAT2	SKU code	SKU name	Sale price (normal)	Total sale	Sale qty	Inventory (amt)	Meter for display product
1st 2019	1003	ACCESSORIES	1-006366-000	HOP XA BONG 760	15.500	-	-	-	0
1st 2019	1003	ACCESSORIES	1-007008-000	THAU TAM LON	65.800	2.003.364	33	2.948.000	0
1st 2019	1003	ACCESSORIES	1-007460-000	HOP XA BONG 106	6.100	326.909	58	508.400	0
1st 2019	1003	ACCESSORIES	1-007732-000	VI NIA NO VI 2 HPB06	10.500	-	-	-	0
1st 2019	1003	ACCESSORIES	1-007903-000	THUNG RAC DUY TAN DAP NHI	33.300	878.545	32	2.204.600	0
1st 2019	1003	ACCESSORIES	1-007904-000	THUNG RAC DUY TAN DAP TRUNG	99.200	1.101.819	12	6.565.000	0
1st 2019	1003	ACCESSORIES	1-007943-000	GANG TAY MELODY 9 SIZE L	13.800	1.336.363	105	4.466.000	0

[Click to see dashboard](#)

Khám phá dữ liệu

Period: thời kì bán hàng, có 2 thời kì trong dữ liệu

Store: có 5 cửa hàng tương ứng với 5 cửa hàng đã phân loại

CAT2: Loại sản phẩm

Mã sản phẩm

SKU code: Tên sản phẩm

SKU name: Giá danh nghĩa của sản phẩm

Sale price (normal):

Total sale: Tổng doanh số

Sale qty: Số lượng bán ra

Inventory (amt): Giá trị của hàng trong kho

Meter for display
Dung lượng chiếm của hàng trong kho

product:

Khám phá dữ liệu

=> Không có giá trị bị lặp

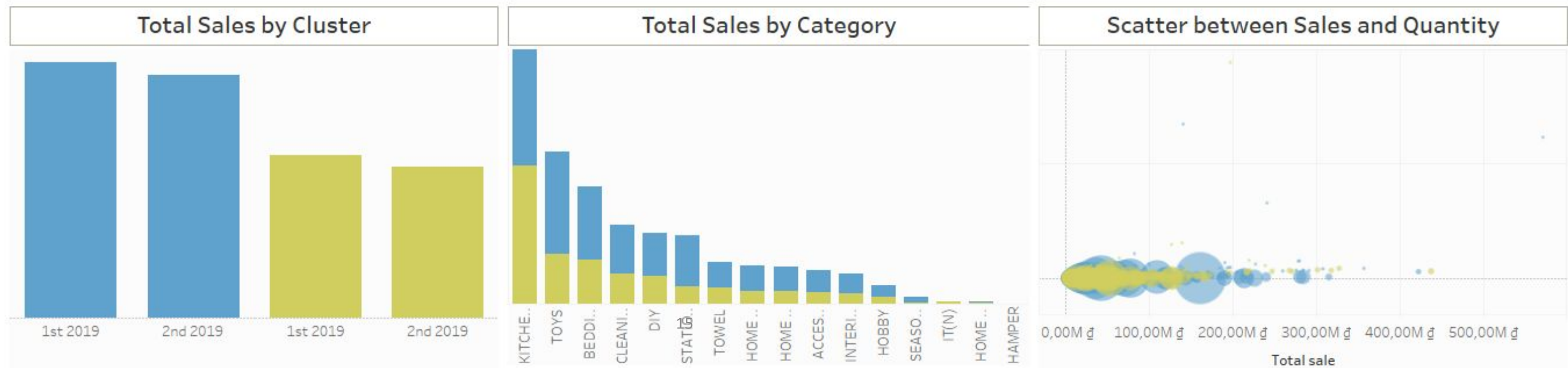
=> Giá trị bị thiếu không thể tác động

=> Một tên sản phẩm có thể có nhiều hơn 1 SKU
code ở cùng một cửa hàng, trong cùng kì

Sales Performance

SALES PERFORMANCE

Total Sales	Sales quantity	Number of prduct types	Category	Meter for display
197.863,20M ₺	3.799,68K	19.943	16	4.294

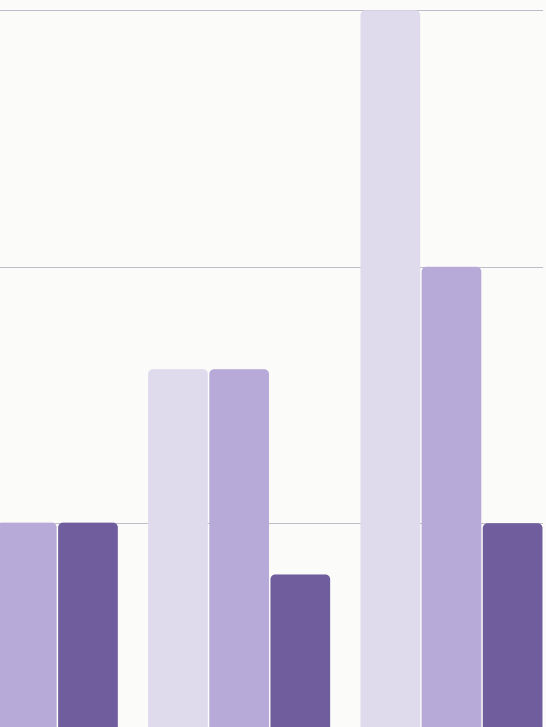


Top SKU by Sales Quantity

Cluster	S...	KITCHEN EQUIPMENT	STATIONARY	CLEANING	TOYS	DIY	ACCESSORI..	HOBBY	TOWEL	BEDDING	INTERIOR	HOME APPLIANCE..	HOME APPLI.. ANCES..	SEASONING ELECTRONIC	HOME ELECTRONIC	HAMPER	IT(N)
0	1013	500,56K	212,38K	116,73K	84,70K	56,57K	56,16K	53,03K	47,67K	40,27K	19,16K	10,10K	4,17K	3,92K	1,43K	0,01K	
	1009	406,15K	162,42K	83,66K	81,46K	46,46K	39,71K	25,96K	37,98K	53,94K	18,38K	8,63K	4,09K	1,78K	3,37K	0,00K	
1	1005	247,94K	127,53K	51,72K	35,80K	32,01K	20,18K	21,55K	26,13K	40,42K	11,79K	5,75K	2,61K	0,52K	3,10K	0,00K	
	1006	235,38K	97,96K	57,13K	33,05K	29,10K	28,72K	26,60K	25,75K	26,98K	8,91K	5,40K	1,66K	0,45K	2,71K	0,01K	
	1003	159,48K	85,14K	34,62K	23,86K	20,04K	19,63K	25,84K	15,17K	12,27K	5,15K	3,55K	1,54K	0,39K	5,05K		0,28K

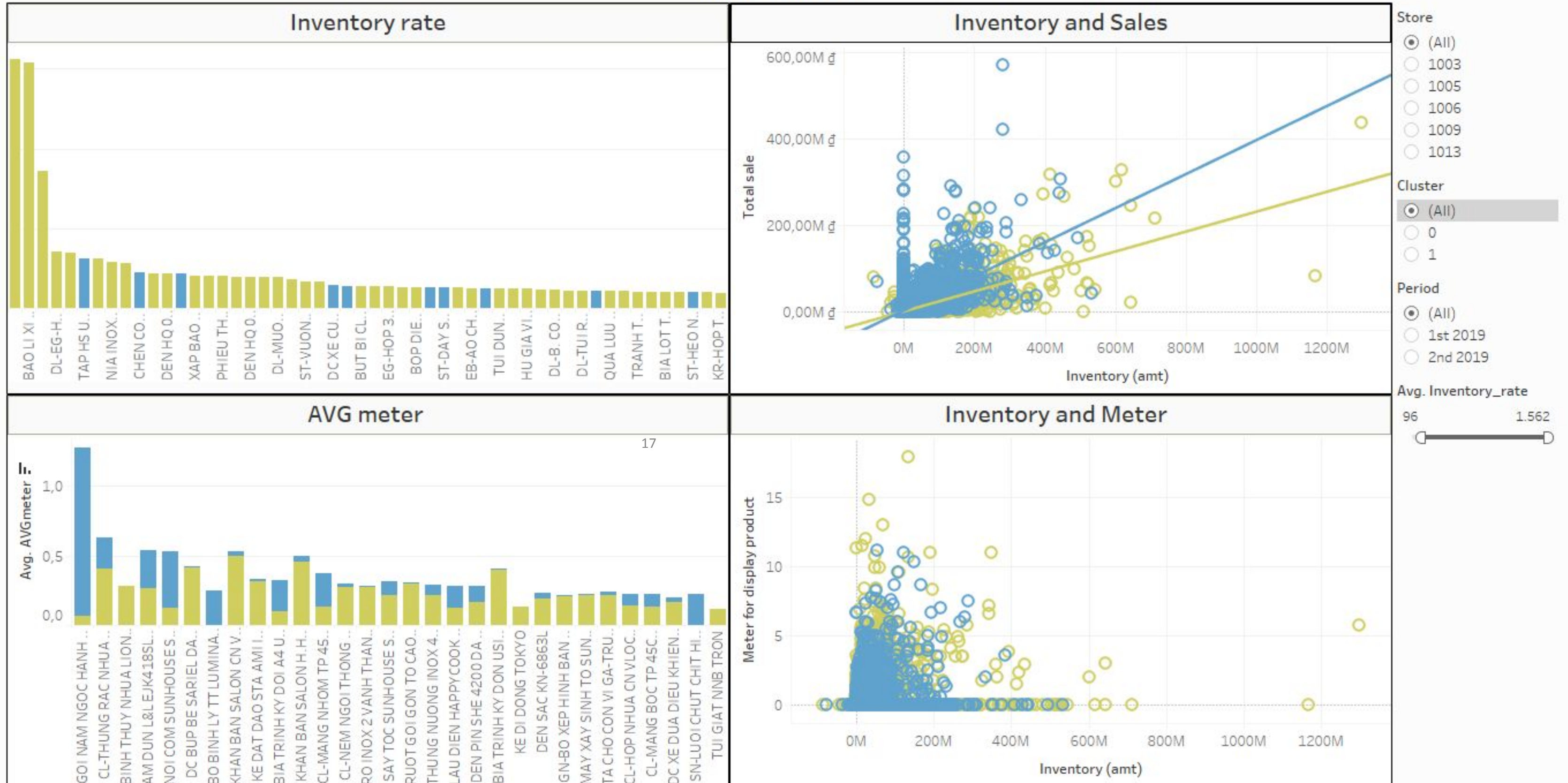
Kết quả

- ❑ Các cửa hàng ở nhóm thứ nhất có doanh thu gấp rưỡi doanh thu của các cửa hàng ở nhóm thứ 2
- ❑ Các loại sản phẩm đóng góp chính vào doanh của hai nhóm cửa hàng là tương đồng nhau, chủ yếu nằm ở các nhóm:
 - Kitchen equipment
 - Toys
 - Bedding
 - Cleaning
- ❑ Các nhóm cửa hàng thứ nhất bán những mặt hàng cao cấp hơn, trong khi đó các cửa hàng ở nhóm thứ 2 bán những mặt hàng rẻ hơn, và cũng ít hơn về số lượng



Inventory Performance

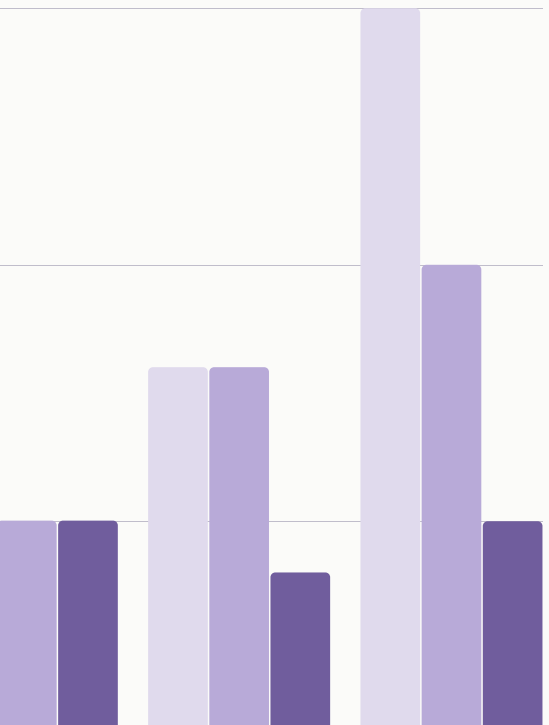
INVENTORY



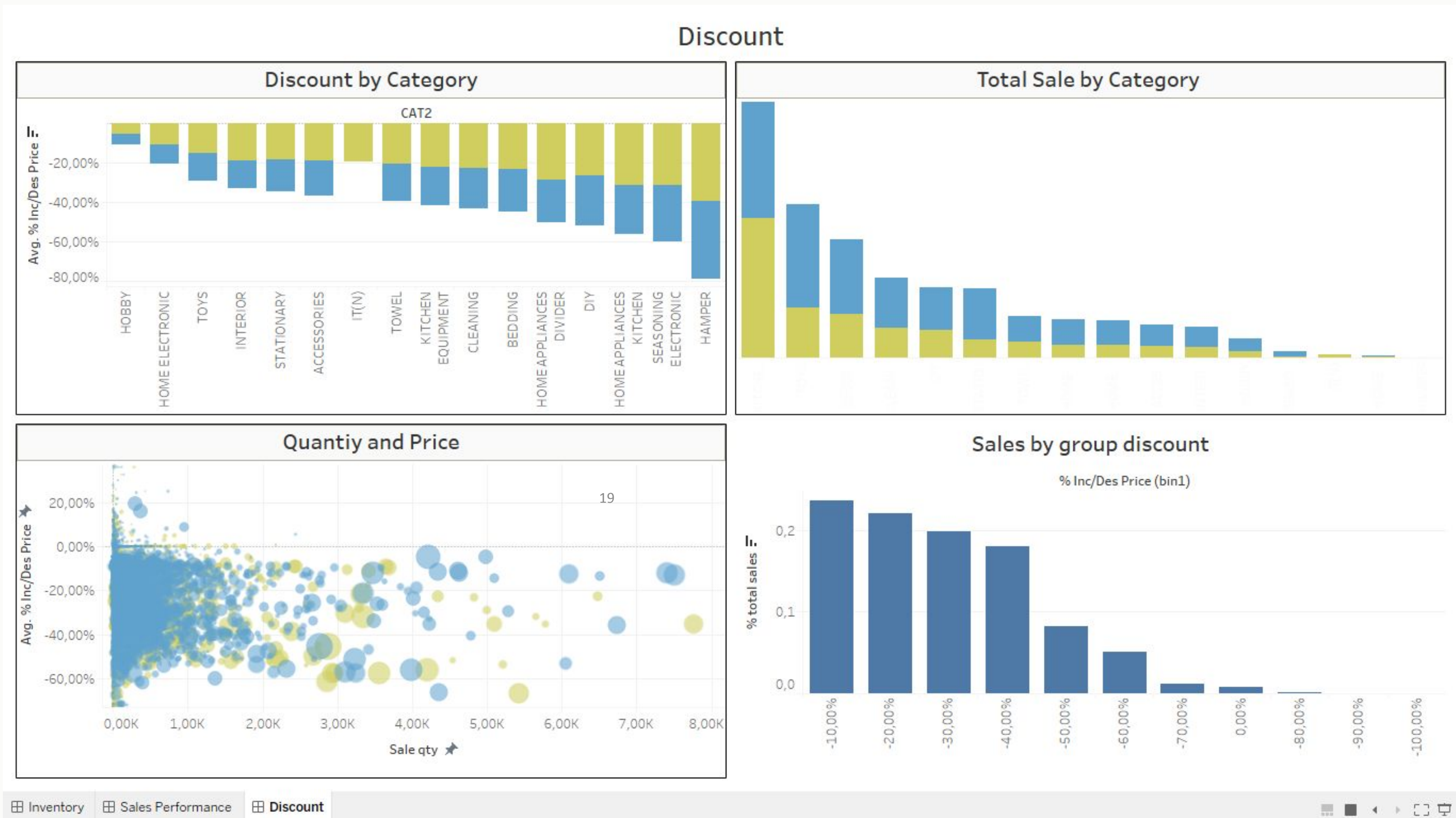
Kết quả

- Các mặt hàng có tỉ lệ tồn kho thuộc top cao ở hai nhóm của hàng đều không đóng góp vào doanh thu rất ít , ngoại trừ gói trang trí cái loại của của hàng 1005 ở kì 1
- Các mặt hàng ở nhóm của hàng thứ nhất có xu hướng chiếm ít dung lượng kho hơn so với cửa hàng thứ 2,
- Giá trị kho của nhóm 1 có xu hướng ít hơn nhóm 2, và có hệ số biến thiên cao hơn với doanh số bán ra

=> NHÓM 1 QUẢN LÝ HÀNG TỒN KHO TỐT HƠN NHÓM 2

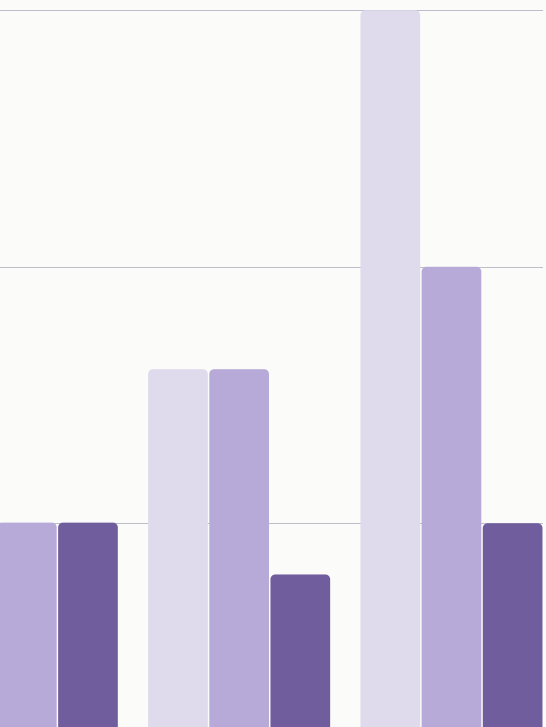


Discount Performance



Kết quả

- Các mặt hàng có giá trị giảm giá bình quân cao nhất ở 2 nhóm của hàng giống nhau bao gồm: Hamper, Seasoning Electronic, Home appliance kitchen
- Những mặt hàng trên không đóng góp quá nhiều vào doanh thu
- Những mặt hàng giảm giá ít hơn như lại đạt được hiệu quả doanh thu: Toys, cleaning, Kitchen equipment, Stationary
- Mức giảm giá hiệu quả từ 10% -20% chiếm tới 45 % doanh số trong những sản phẩm được giảm giá





**THANK FOR
WATCHING**

