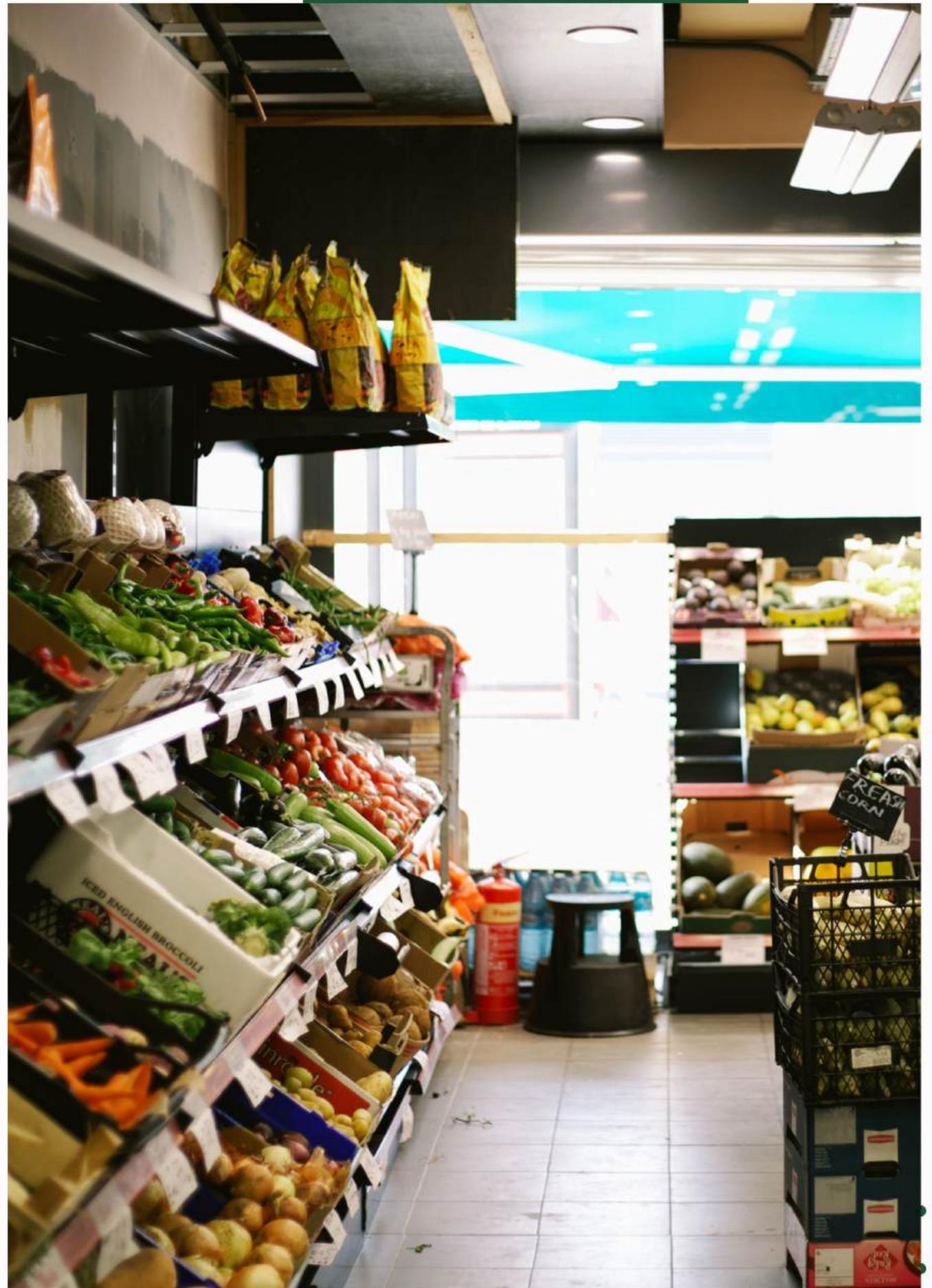


PHÂN TÍCH DỮ LIỆU GIỎ HÀNG CỦA SIÊU THỊ

Nhóm 9



Nội dung

01

Tổng quan dữ liệu

02

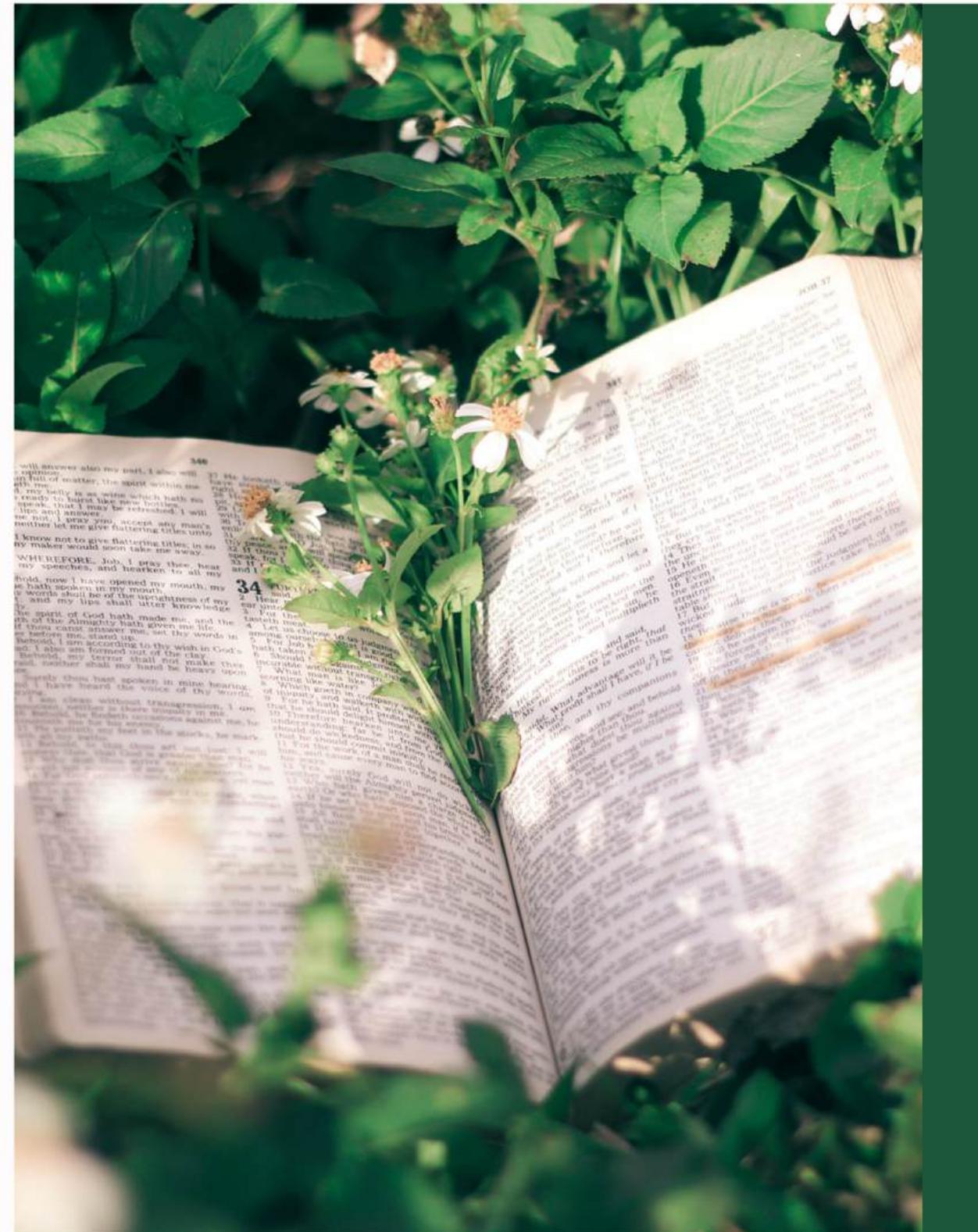
Phân tích khám phá dữ liệu

03

Xây dựng mô hình

04

Tổng kết



The background image shows a spacious, modern office or lobby area. Large windows on the left look out onto a green landscape. The room is filled with various types of green plants, including hanging vines and potted plants on desks. There are several wooden tables and chairs, some with black office chairs. The ceiling is high with exposed pipes and lighting fixtures. A large, stylized title 'TỔNG QUAN DỮ LIỆU' is overlaid in the center.

TỔNG QUAN DỮ LIỆU

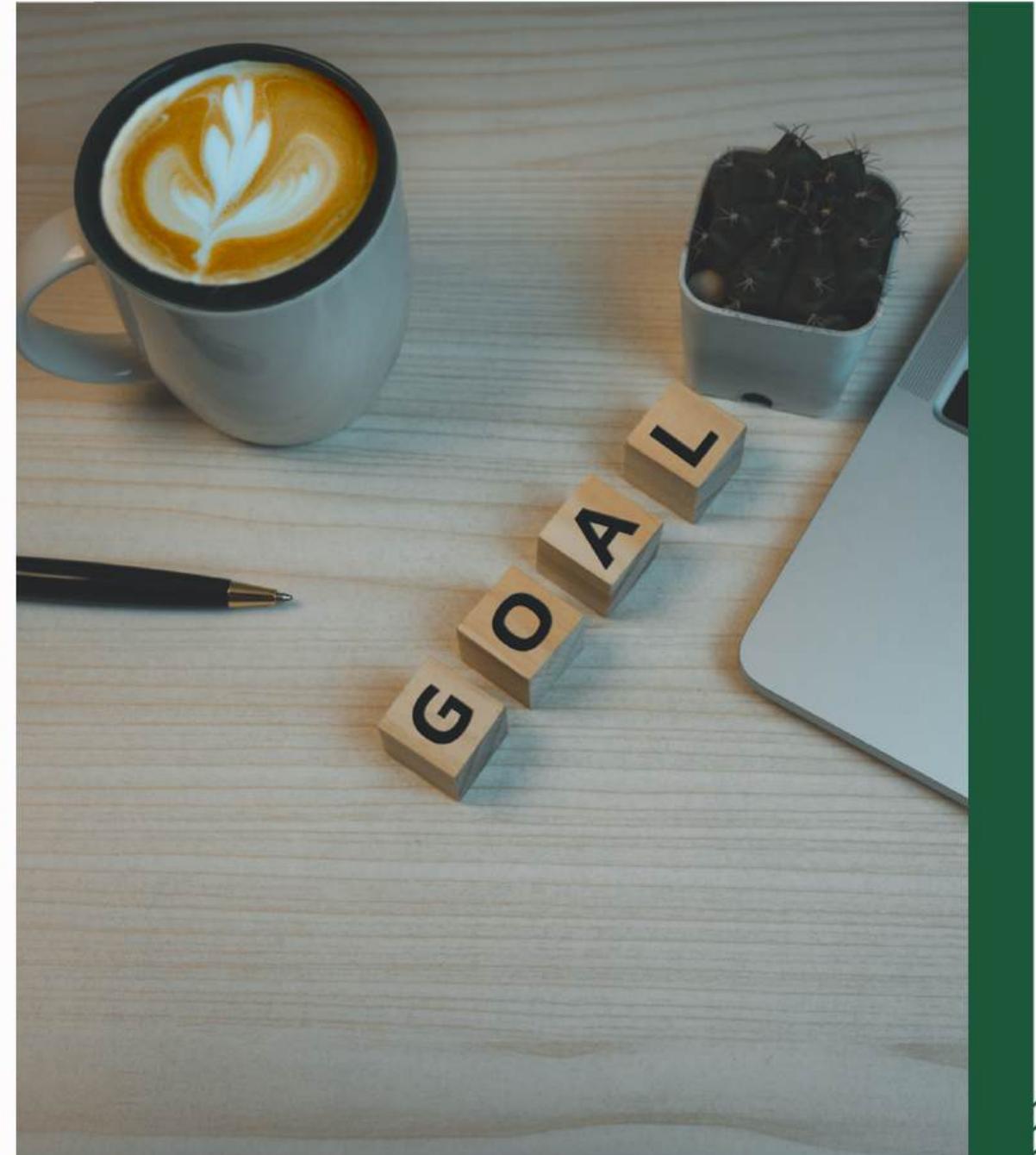
Giới thiệu bộ dữ liệu

Siêu thị Hunter là một chuỗi siêu thị **đa quốc gia** hoạt động cả trực tiếp và trực tuyến, với sự hiện diện tại hơn **10 lãnh thổ** trên thế giới. Hunter đã vượt qua **nhiều khủng hoảng lớn** như đại dịch Covid-19, chiến tranh Ukraine và khủng hoảng thiếu hụt khí đốt, nhờ vào chiến lược kinh doanh **linh hoạt và khả năng thích ứng cao** trong môi trường biến động. Siêu thị Hunter có thể tự tin mang đến cho khách hàng **trải nghiệm mua hàng tốt nhất**.



Mục tiêu

Mục tiêu của bài phân tích này là nghiên cứu và phân tích **hành vi mua sắm** của khách hàng tại siêu thị Hunter. Bài phân tích sẽ tập trung vào việc khám phá các **yếu tố ảnh hưởng** đến **quyết định mua sắm** của khách hàng, xu hướng tiêu dùng, tần suất mua sắm, và loại sản phẩm ưa thích. Kết quả từ phân tích này sẽ giúp hiểu rõ hơn về **hành vi tiêu dùng**, từ đó đưa ra các **quyết định chiến lược** nhằm tối ưu hóa trải nghiệm khách hàng và cải thiện hiệu quả kinh doanh của siêu thị.



Kiểm tra dữ liệu

```
[31] data.shape  
→ (2015280, 14)
```

Dữ liệu có hơn 2 triệu dòng và 14 cột

```
▶ data.dtypes  
→  
order_id      int64  
user_id       int64  
order_number   int64  
order_dow     int64  
order_hour_of_day    int64  
days_since_prior_order float64  
product_id    int64  
add_to_cart_order int64  
reordered     int64  
department_id int64  
department    object  
product_name   object
```

Các cột của bộ dữ liệu bao gồm:

- order_id: ID của đơn hàng
- user_id: ID người mua
- order_number: Số lượng đơn hàng
- order_dow: Thời gian mà đơn được tạo tính theo ngày trong tuần
- order_hour_of_day: Thời gian mà đơn được tạo tính theo giờ
- days_since_prior_order: Lần mua cuối cùng
- product_id: ID sản phẩm
- add_to_cart_order: Số lượng sản phẩm được đưa vào giỏ hàng
- reordered: Xem liệu sản phẩm có được mua lại không
- department_id: ID quầy
- department: Tên quầy
- product_name: Tên sản phẩm

Kiểm tra dữ liệu

KIỂM TRA TYPE DỮ LIỆU

Type của dữ liệu đã ở đúng định dạng

```
[34] data.info()  
  
→ <class 'pandas.core.frame.DataFrame'>  
Index: 2015280 entries, 0 to 2019500  
Data columns (total 14 columns):  
 #   Column           Dtype     
---  --  
 0   order_id         int64  
 1   user_id          int64  
 2   order_number     int64  
 3   order_dow        int64  
 4   order_hour_of_day int64  
 5   days_since_prior_order float64  
 6   product_id       int64  
 7   add_to_cart_order int64  
 8   reordered        int64  
 9   department_id    int64  
 10  department        object  
 11  product_name     object  
 12  order_number_lower float64  
 13  order_number_upper float64  
dtypes: float64(3), int64(9), object(2)  
memory usage: 230.6+ MB
```

Kiểm tra dữ liệu

KIỂM TRA DỮ LIỆU BỊ THIẾU

Có tổng cộng 124342 hàng bị thiếu ở cột
days_since_prior_order

order_id	0
user_id	0
order_number	0
order_dow	0
order_hour_of_day	0
days_since_prior_order	124342
product_id	0
add_to_cart_order	0
reordered	0
department_id	0
department	0
product_name	0

FILL DỮ LIỆU BỊ THIẾU

Fill các hàng bị thiếu
bằng -1

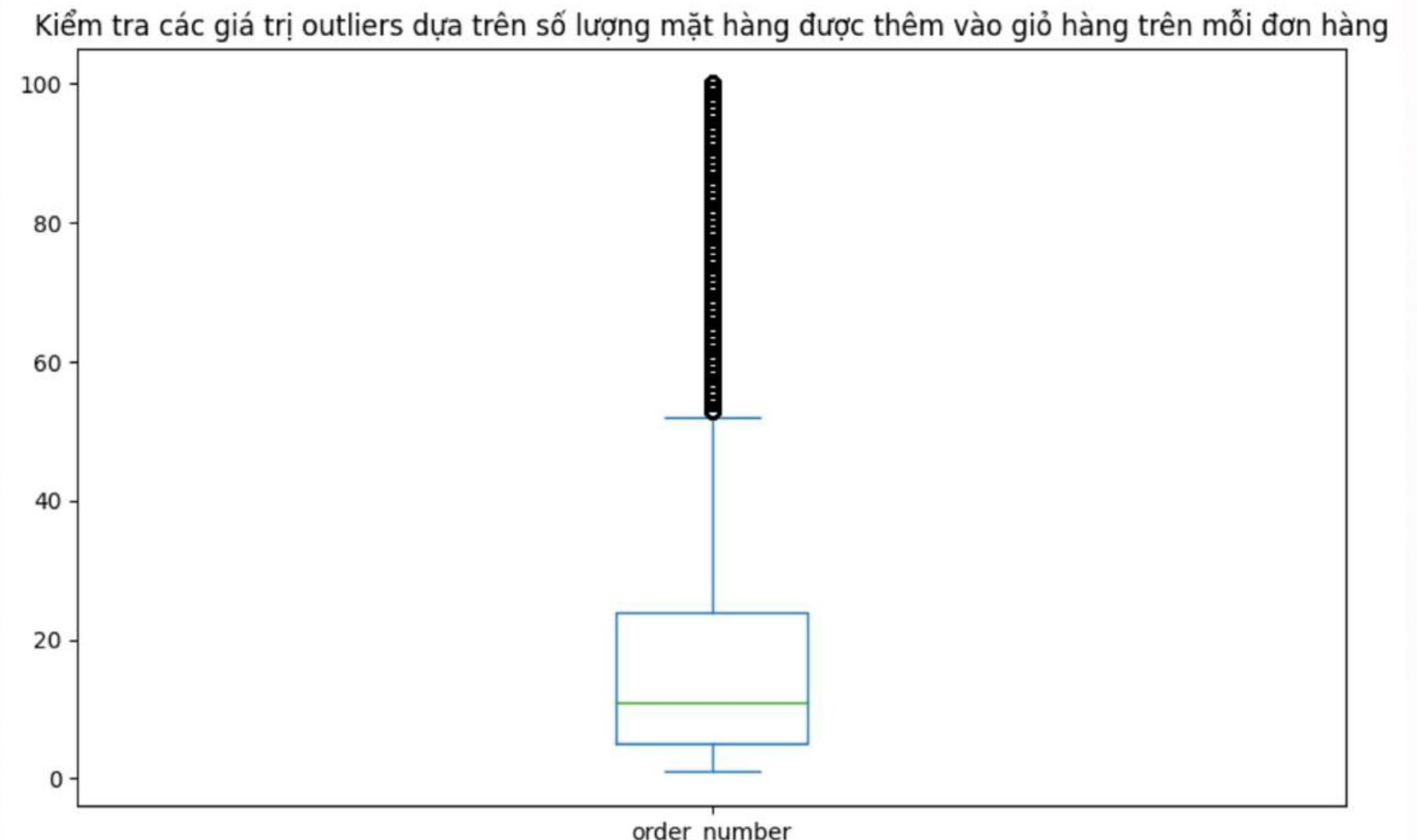
```
data['days_since_prior_order'].fillna(-1, axis = 0)
```

KIỂM TRA LẠI

Không còn dữ liệu bị thiếu

order_id	0
user_id	0
order_number	0
order_dow	0
order_hour_of_day	0
days_since_prior_order	0
product_id	0
add_to_cart_order	0
reordered	0
department_id	0
department	0
product_name	0

Kiểm tra Outliers



Kiểm tra Outliers

BẢNG TỔNG HỢP OUTLIER:
Kiểm tra các giá trị outliers
dựa trên số lần mua hàng của
mỗi khách hàng

user_id	order_id	order_number	order_dow	order_hour_of_day	days_since_prior_order	product_id	add_to_cart_order	reordered	department_id	
10	8.339500e+04	4.000000	3.000000	15.000000		14.000000	60.000000	15.500000	0.433333	10.400000
133	2.415920e+05	11.000000	3.000000	13.000000		5.000000	68.133333	15.500000	0.400000	9.033333
197	2.357070e+05	7.000000	6.000000	14.000000		6.000000	74.206897	15.000000	0.793103	11.551724
216	1.500908e+06	5.000000	5.000000	15.000000		24.000000	68.448276	15.000000	0.172414	13.655172
264	3.022479e+06	29.000000	2.000000	9.000000		9.000000	79.551724	15.000000	0.827586	11.103448
...
205789	3.418993e+06	13.000000	6.000000	12.000000		7.000000	63.900000	15.500000	0.600000	9.466667
205926	2.385091e+06	11.000000	1.000000	15.000000		6.000000	73.828571	18.000000	0.514286	10.171429
205956	4.562900e+05	3.000000	3.000000	14.000000		30.000000	69.151515	17.000000	0.303030	10.909091

BẢNG TỔNG HỢP OUTLIER:
Kiểm tra các giá trị outliers
dựa trên số lượng mặt hàng
được thêm vào giỏ hàng trên
mỗi đơn hàng

order_id	user_id	order_number	order_dow	order_hour_of_day	days_since_prior_order	product_id	add_to_cart_order	reordered	department_id	department	product_name	order_number_lower	order_number_upper	
2538	677735	175561	2	3	12	16.0	32	1	0	4	produce	packaged produce	7.467505	52.722971
8764	2018358	185514	22	4	17	17.0	128	1	0	3	bakery	tortillas flat bread	-5.770319	18.847242
8765	2018358	185514	22	4	17	17.0	54	2	1	17	household	paper goods	-5.770319	18.847242
15536	2545752	185279	9	5	9	12.0	84	1	0	16	dairy eggs	milk	10.001293	28.306399
16390	1365995	136421	1	3	10	-1.0	88	1	0	13	pantry	spreads	1.544766	74.812377
...
2010592	1097536	97028	5	0	16	7.0	24	1	1	4	produce	fresh fruits	-0.146102	4.574673
2010593	1097536	97028	5	0	16	7.0	24	2	0	4	produce	fresh fruits	-0.146102	4.574673
2015806	1581072	93695	5	4	17	30.0	97	1	0	13	pantry	baking supplies decor	3.364384	4.740879
2016905	1373400	139870	11	2	12	30.0	116	1	0	1	frozen	frozen produce	-2.906531	10.137300
2016906	1373400	139870	11	2	12	30.0	32	2	0	4	produce	packaged produce	-2.906531	10.137300

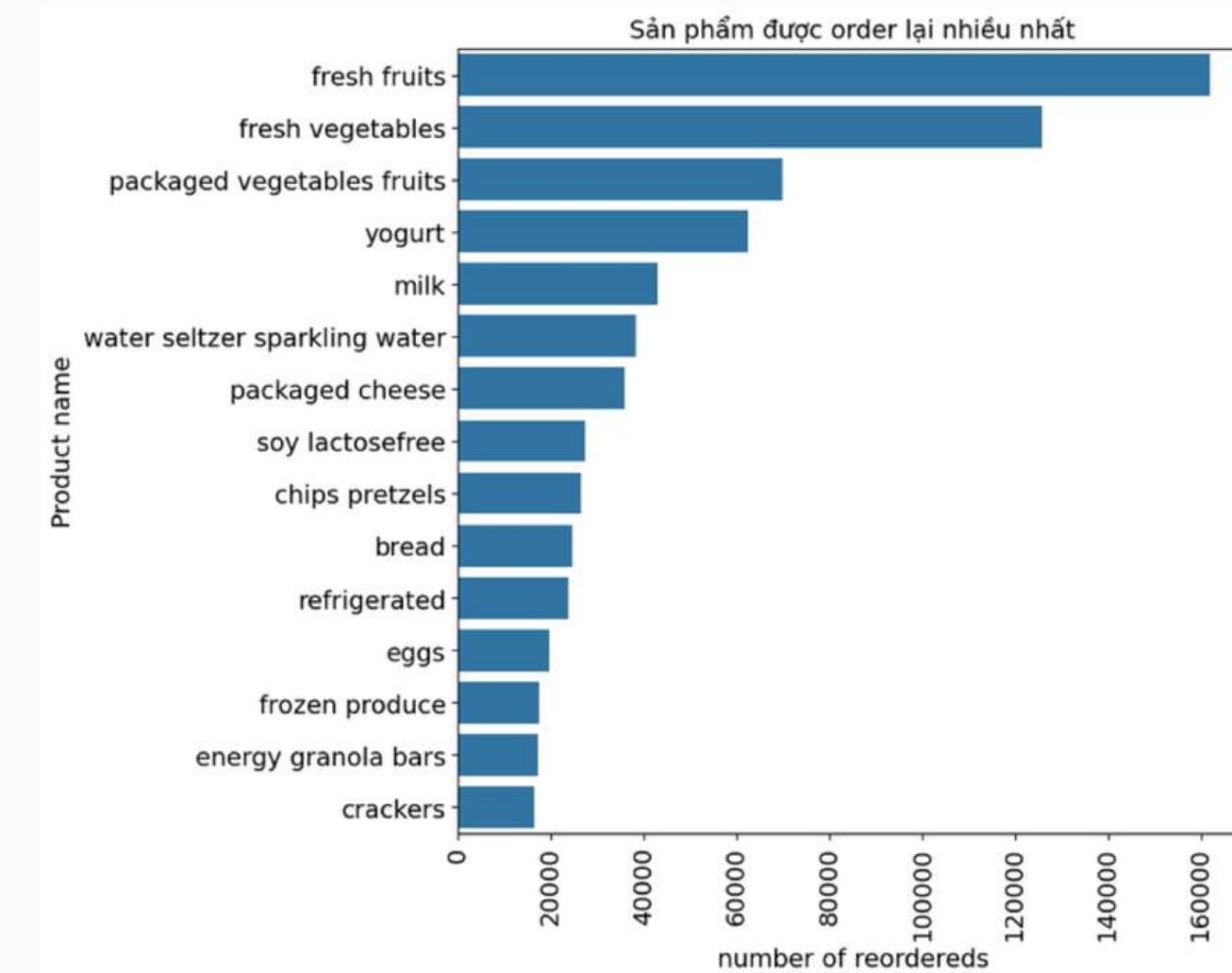
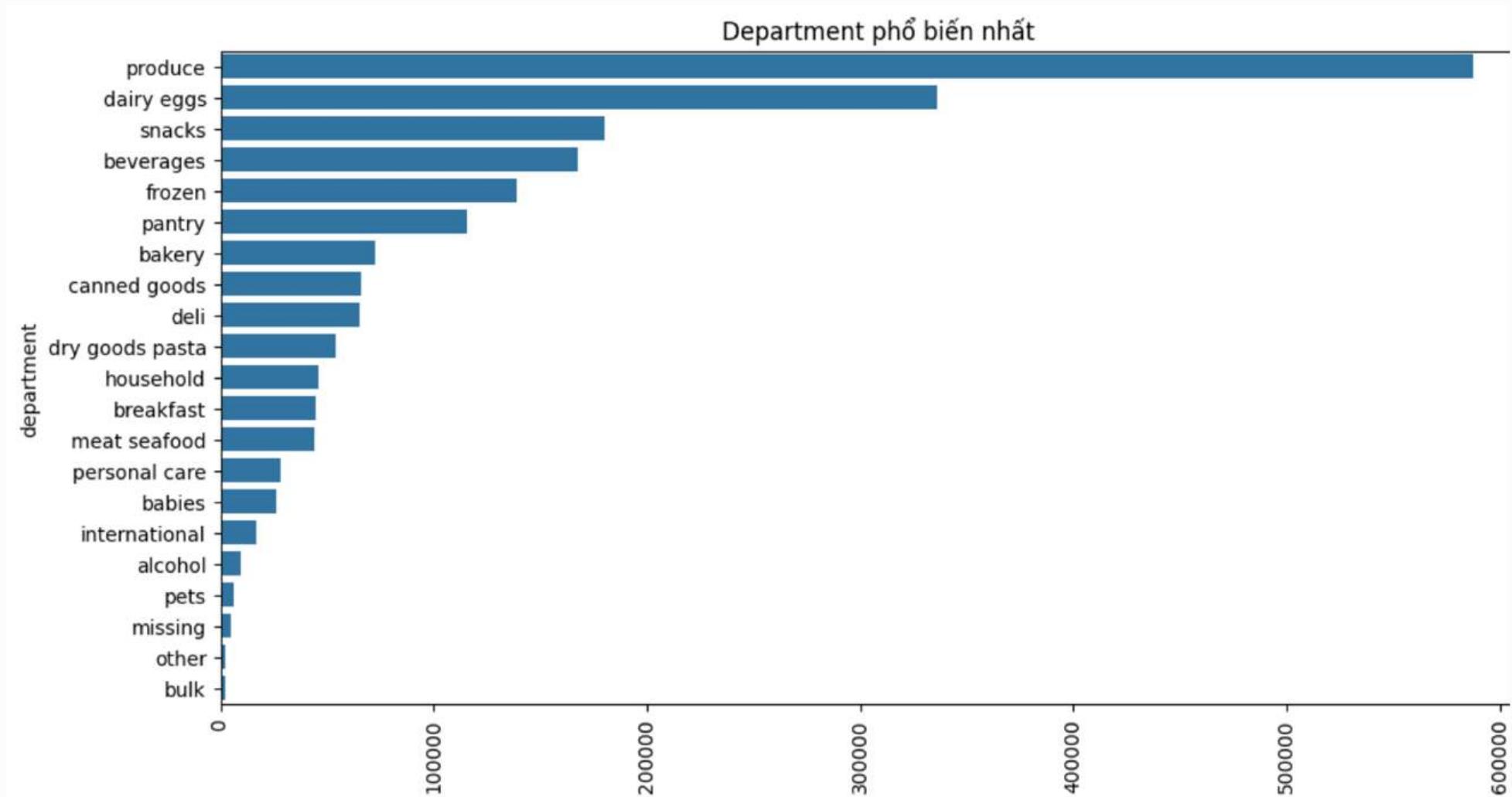
Xóa outliers

```
outliers_orders= outliers_orders.drop([21882, 85422, 88725, 156668], axis=0)
data = data.drop(outliers_cart.index)
data = data.drop(outliers_orders.index)
```

The background image shows a modern office environment. It features large windows on one side, allowing natural light to fill the space. The interior is decorated with numerous green plants of various types and sizes, creating a lush, tropical atmosphere. There are several wooden desks arranged in rows, each equipped with a computer monitor and keyboard. The office has a high ceiling with exposed pipes and ductwork, giving it an industrial-chic feel. In the background, there's a seating area with large, comfortable green sofas and armchairs.

PHÂN TÍCH KHÁM PHÁ DỮ LIỆU

SẢN PHẨM PHỔ BIẾN

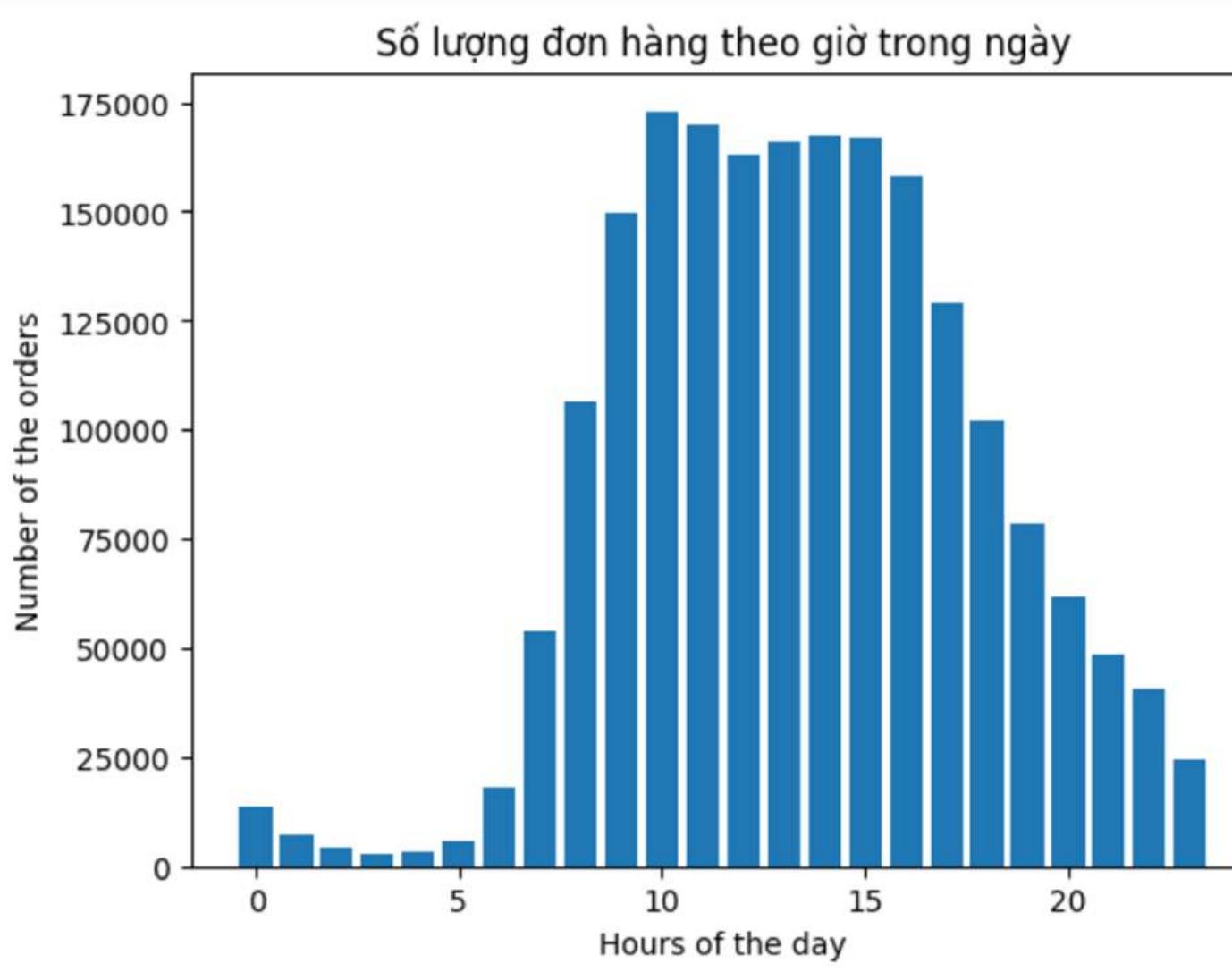
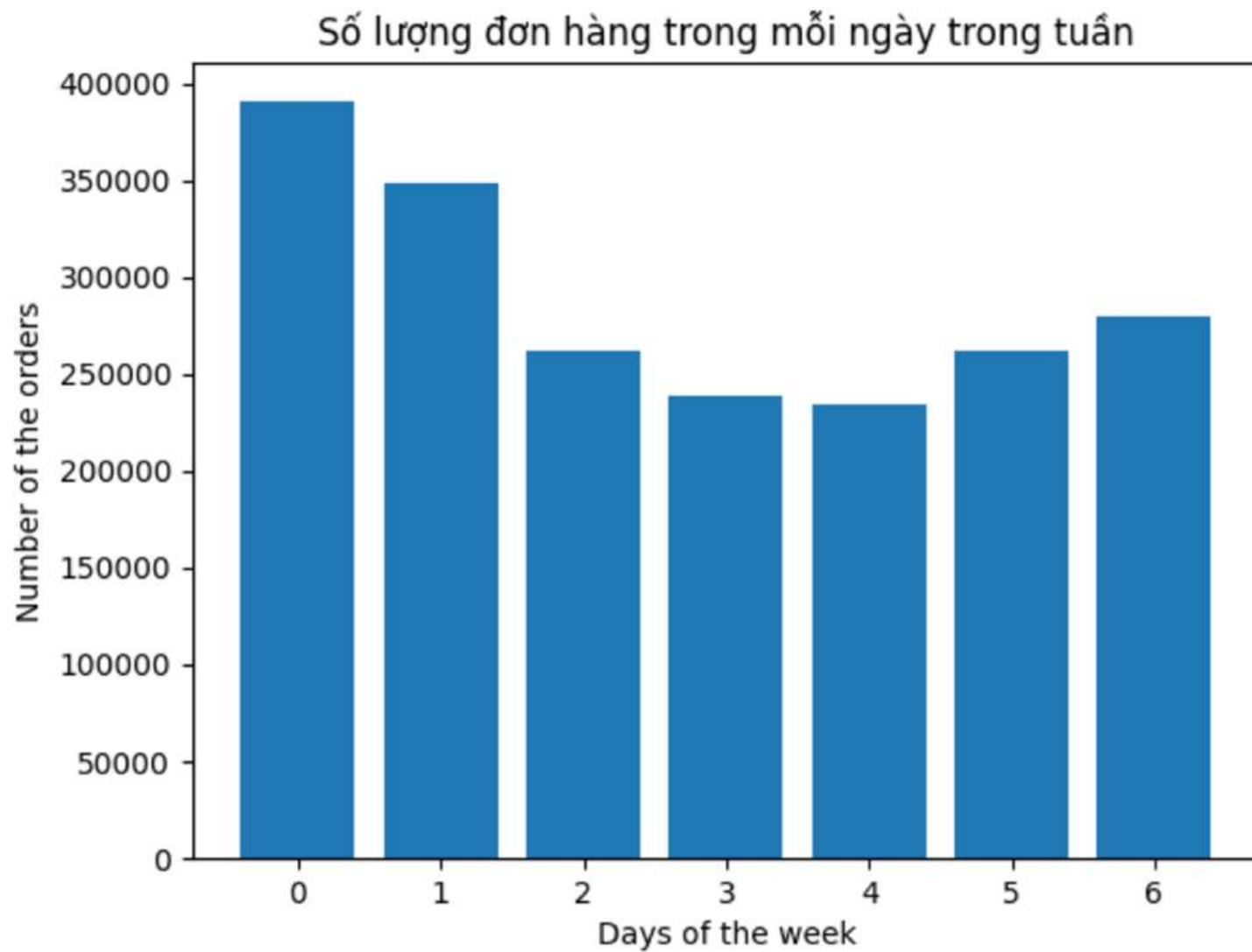


- Produce (trái cây và rau củ) là department có số lượng sản phẩm lớn nhất, xếp sau đó là dairy eggs (các sản phẩm được làm từ sữa và trứng), snacks, bererages, và đồ đông lạnh
- Sản phẩm được order lại nhiều nhất là các loại rau củ trái cây, các thực phẩm từ trứng sữa như là sữa chua, sữa, phô mai,... Còn có thêm cả các thực phẩm khác làm từ tinh bột như bánh mì, bánh quy



Các sản phẩm được mua nhiều nhất ở siêu thị chủ yếu là các loại rau củ, thực phẩm từ sữa, thực phẩm đông lạnh, nước uống, đây cũng là những sản phẩm được mua lại thường xuyên, siêu thị nên giữ nguồn cung ổn định của các mặt hàng này.

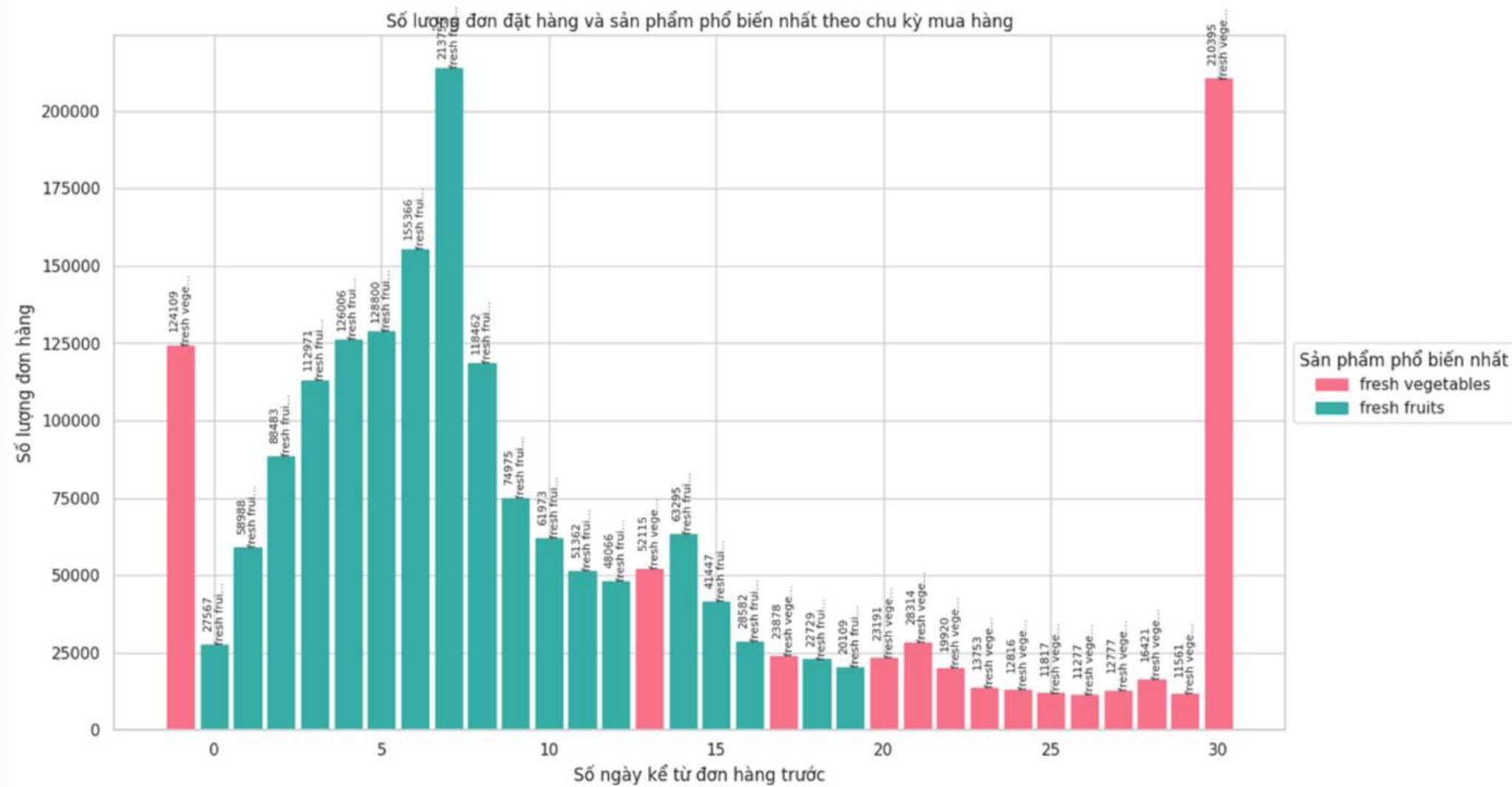
GIỜ MUA CAO ĐIỂM



- Chủ nhật (Ngày 0) là ngày có số lượng đơn hàng lớn nhất với hơn 350k đơn hàng, xếp sau nó là thứ Hai với số lượng đơn hàng thấp hơn một chút so với Chủ Nhật xếp thứ ba là thứ 7
- Bắt đầu từ 6h sáng số lượng đơn hàng tăng nhanh, cao điểm nhất từ 10h - 15h. Người mua có xu hướng mua vào buổi sáng sớm

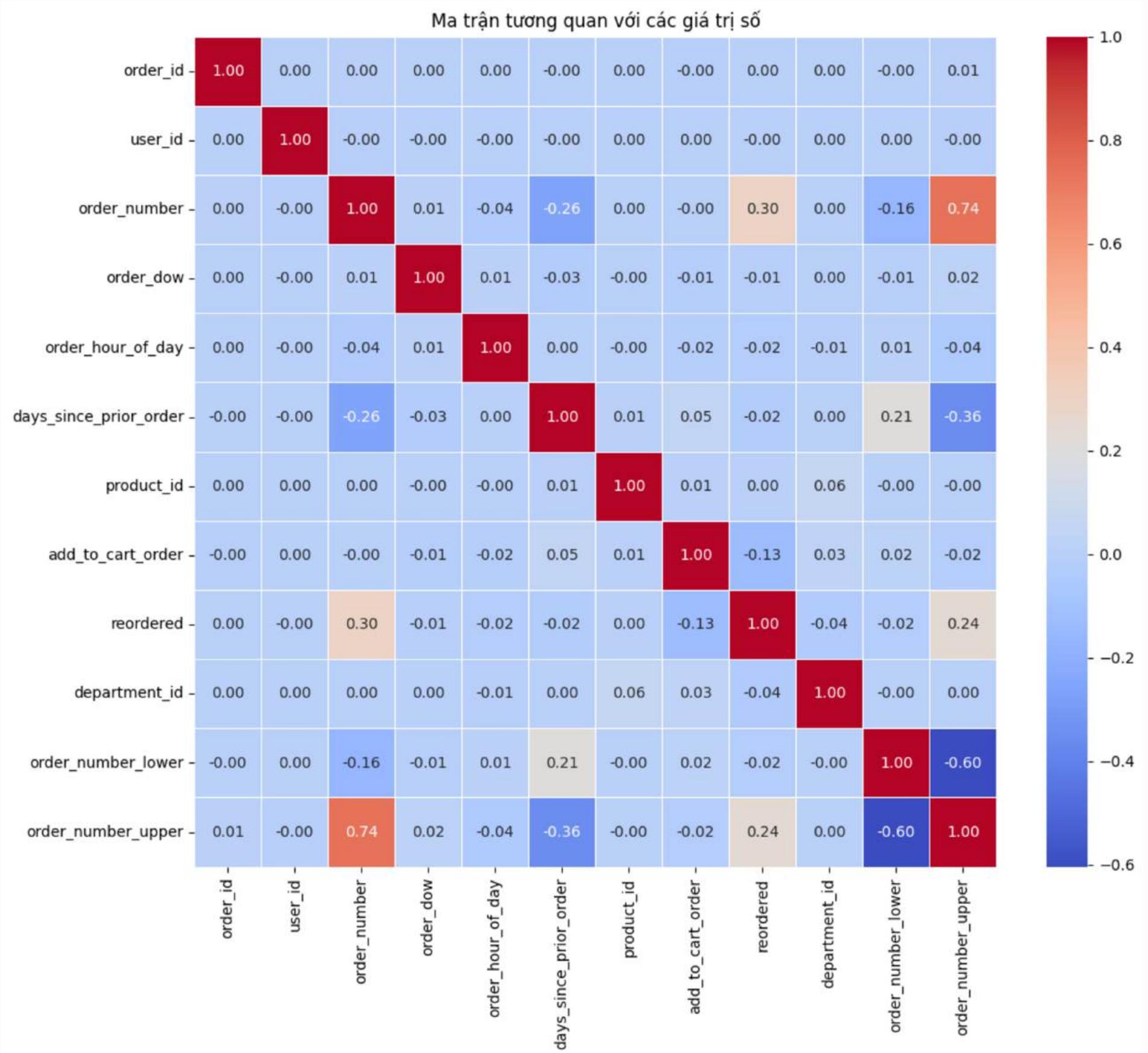
→ Có xu hướng mua hàng vào những ngày đầu tuần và cuối tuần nhằm tích trữ lương thực cho cả tuần, người mua có xu hướng mua vào lúc sáng sớm, khả năng là do khung giờ này là khung giờ thực phẩm giữ được độ tươi mới

CHU KỲ MUA VÀ SẢN PHẨM



- Biểu đồ thể hiện chu kỳ mua hàng của mọi người cao nhất là 7 ngày vào 30 ngày, cũng có nhiều khách hàng mua nhiều lần trong một ngày. Sản phẩm phổ biến nhất đối với chu kỳ 7 ngày là fresh fruit, còn đối với chu kỳ 30 và -1 thì sản phẩm được mua nhiều nhất sẽ là fresh vegetables
- Khách hàng mua sẽ có xu hướng tích trữ trong vòng 1 tuần hoặc 1 tháng nhằm tích trữ, khách hàng mua nhiều lần trong một ngày có thể là do họ không có kế hoạch mua sắm rõ ràng hoặc do họ quên mua sản phẩm nào đó.
- Khách hàng có chu kỳ 7 ngày mua nhiều nhất là hoa quả tươi, đây là khoảng thời gian mà hoa quả vẫn giữ được độ tươi ngon. Khách hàng có chu kỳ 30 ngày thì sẽ mua nhiều nhất là rau củ, khả năng là những người này sẽ mua các loại rau củ tích trữ được lâu ngày

BẢNG TƯƠNG QUAN



Bảng tương quan của bộ dữ liệu
thể hiện các biến có mức độ
tương quan nhẹ với nhau, có một
số biến tương quan ở mức trung
bình nhẹ và trung bình. Có thể
dẫn đến sự trùng lặp thông tin
khi Clustering

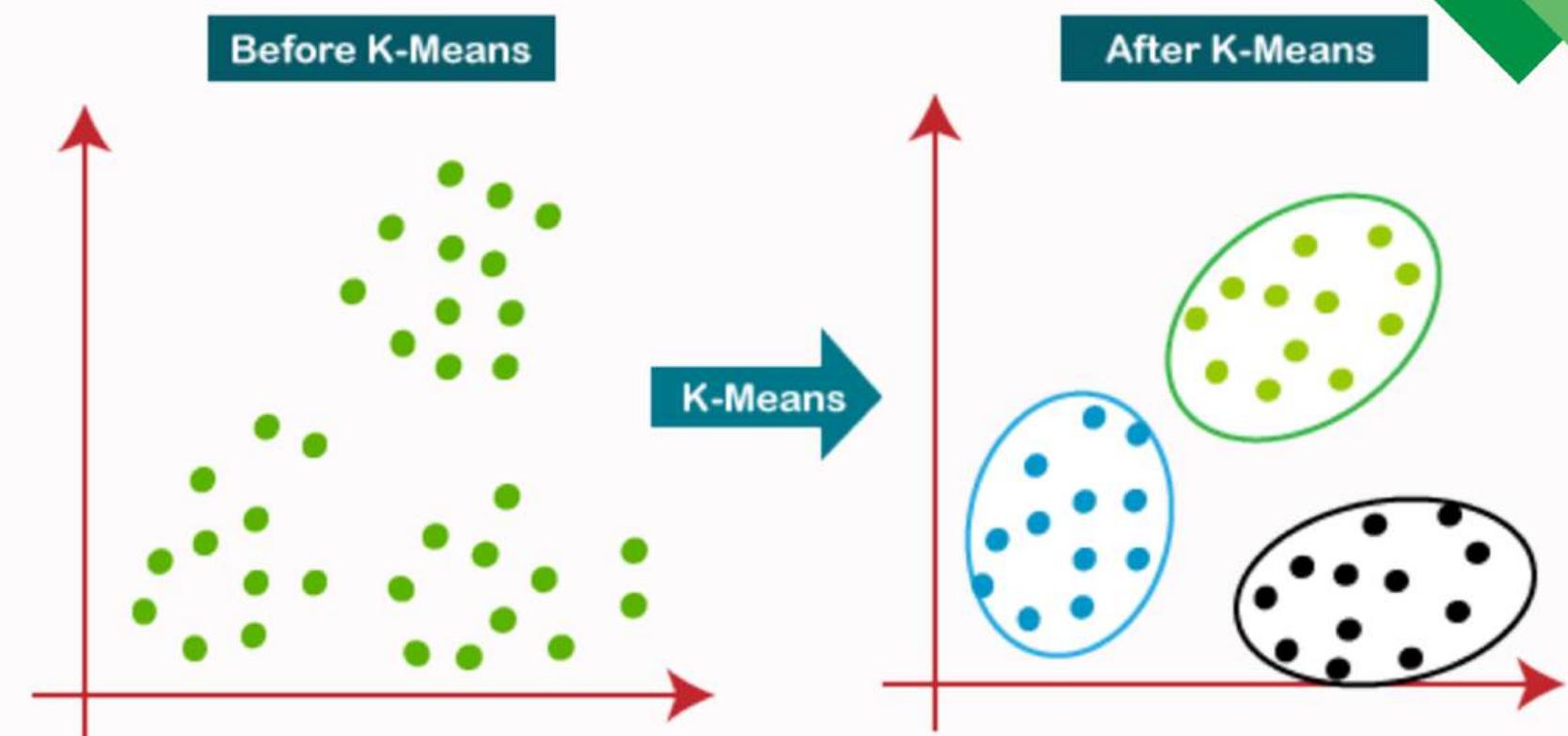


XÂY DỰNG MÔ HÌNH

GIỚI THIỆU MÔ HÌNH: LÝ DO LỰA CHỌN

K-means Clustering (Phân cụm K-means)

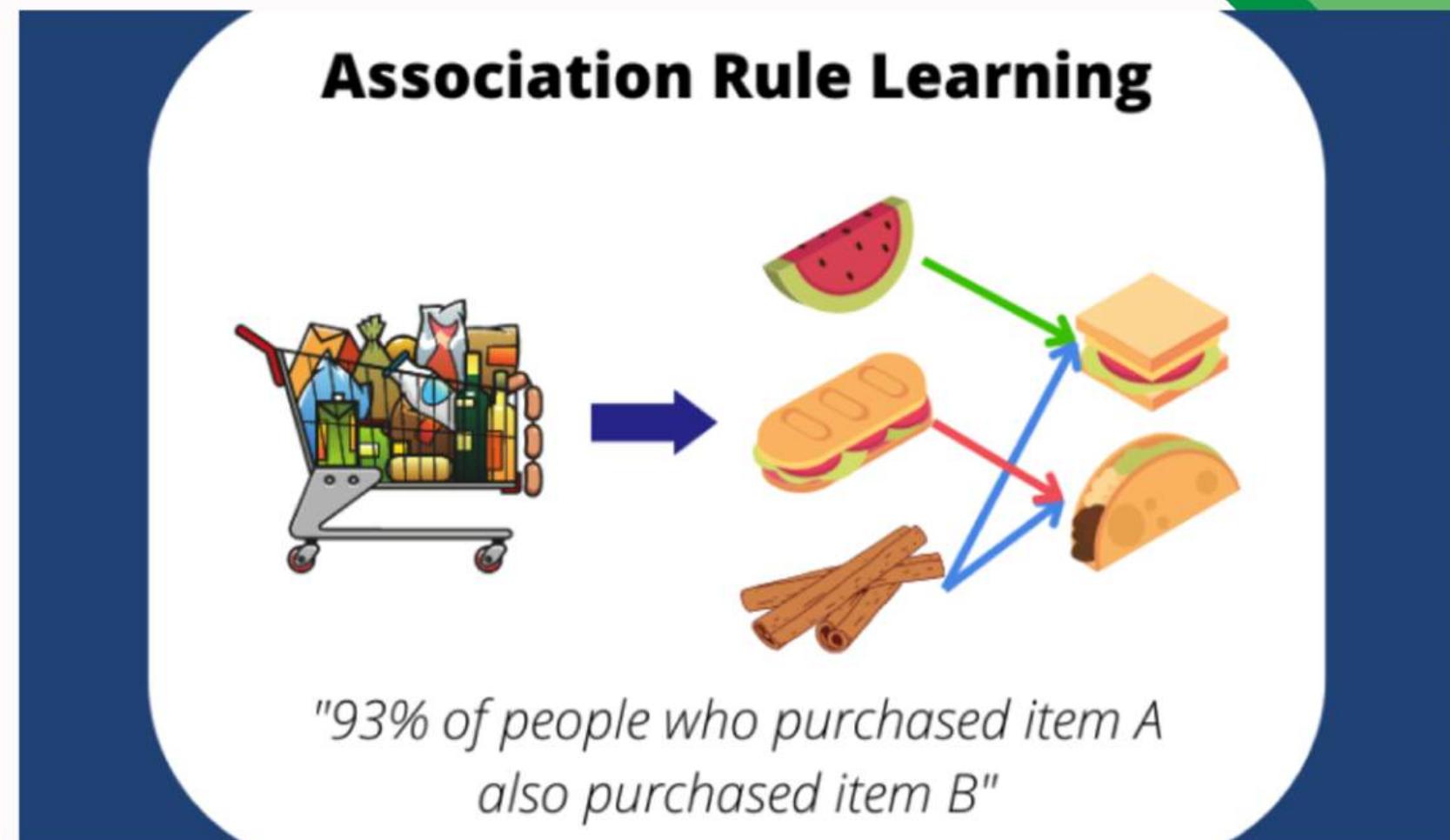
- Mô hình K-means Clustering giúp phân nhóm khách hàng hoặc sản phẩm dựa trên các **đặc điểm tương đồng**. Mục tiêu của thuật toán này đối với Hunter là **nhóm các sản phẩm theo hành vi mua sắm**.
- Mô hình giúp Hunter hiểu sâu hơn về khách hàng và sản phẩm, từ đó cải thiện **hiệu quả kinh doanh**, tối ưu hóa tiếp thị, và tăng **sự hài lòng** của khách hàng.



GIỚI THIỆU MÔ HÌNH: LÝ DO LỰA CHỌN

Association Rules (Quy tắc liên kết)

- Thuật toán Association Rules giúp phát hiện các **mối quan hệ ẩn** giữa các mặt hàng được mua cùng nhau trong dữ liệu giao dịch của Hunter. Mục đích chính là **tối ưu hóa bán chéo (cross-selling)** và **cải thiện** sắp xếp sản phẩm trong cửa hàng.
- Hiểu rõ **mối quan hệ giữa các mặt hàng** giúp **tăng doanh thu** và cải thiện **trải nghiệm khách hàng** thông qua các chiến lược bán chéo và khuyến mãi hợp lý.



CLUSTERING: XỬ LÝ DỮ LIỆU

Trước khi xây mô hình K-Mean Clustering thì nhóm sẽ xử lý dữ liệu như sau:

- Encode dữ liệu

```
encoder = LabelEncoder()
object_col = df.select_dtypes(include='object').columns

for i in object_col:
    df[i] = encoder.fit_transform(df[i])
```

- Scale dữ liệu

```
ss = StandardScaler()
df = pd.DataFrame(ss.fit_transform(df))
```

CLUSTERING: CHỌN SỐ CLUSTER

Elbow Method

Nhóm sẽ sử dụng Elbow Method để chọn số cluster tối ưu cho mô hình.

- Vòng lặp từ 1 đến 10 cụm: Mỗi lần chạy, thuật toán K-means được huấn luyện với số cụm khác nhau.
- Giá trị WCSS (Tổng bình phương khoảng cách trong cụm) được tính và lưu lại.

Elbow Method

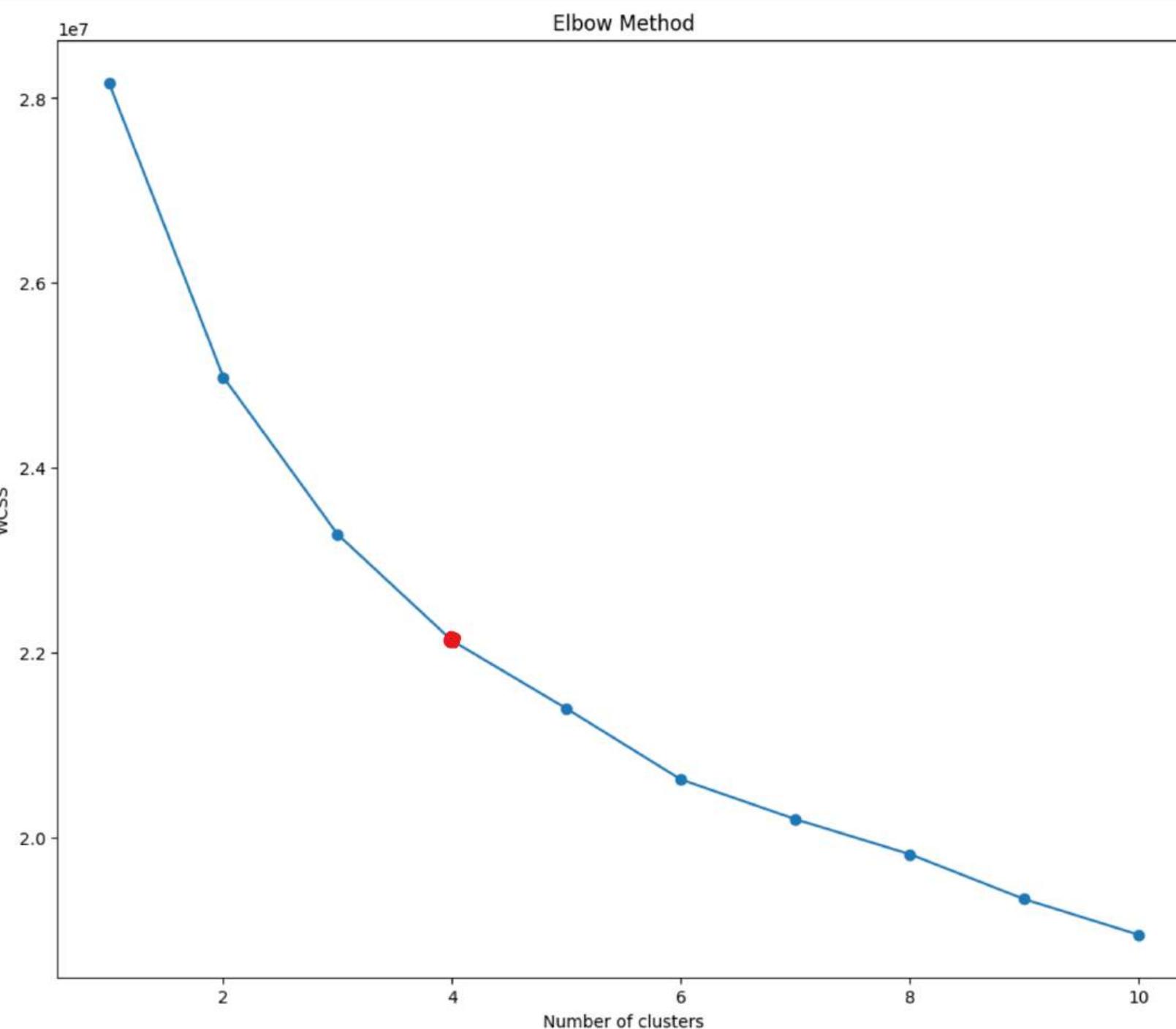
```
[ ] wcss = {'wcss_score':[],'cluster':[]}
for i in range(1,11):
    kmeans = KMeans(n_clusters = i, random_state = 42)
    kmeans.fit(df)
    wcss['wcss_score'].append(kmeans.inertia_)
    wcss['cluster'].append(i)

[ ] wcss_df = pd.DataFrame(wcss)
wcss_df
```

CLUSTERING: CHỌN SỐ CLUSTER

```
[ ] from kneed import KneeLocator  
k1 = KneeLocator(wcss_df['cluster'], wcss_df['wcss_score'], curve='convex', direction='decreasing')  
print(f'Number of cluster:{k1.elbow}')
```

Ở đây số Cluster tối ưu sẽ là
4



CLUSTERING: THỰC HIỆN CLUSTER

Gắn K vào mô
hình

```
[95] kmeans = KMeans(n_clusters = 4, random_state = 42)
     y_pred = kmeans.fit_predict(df)

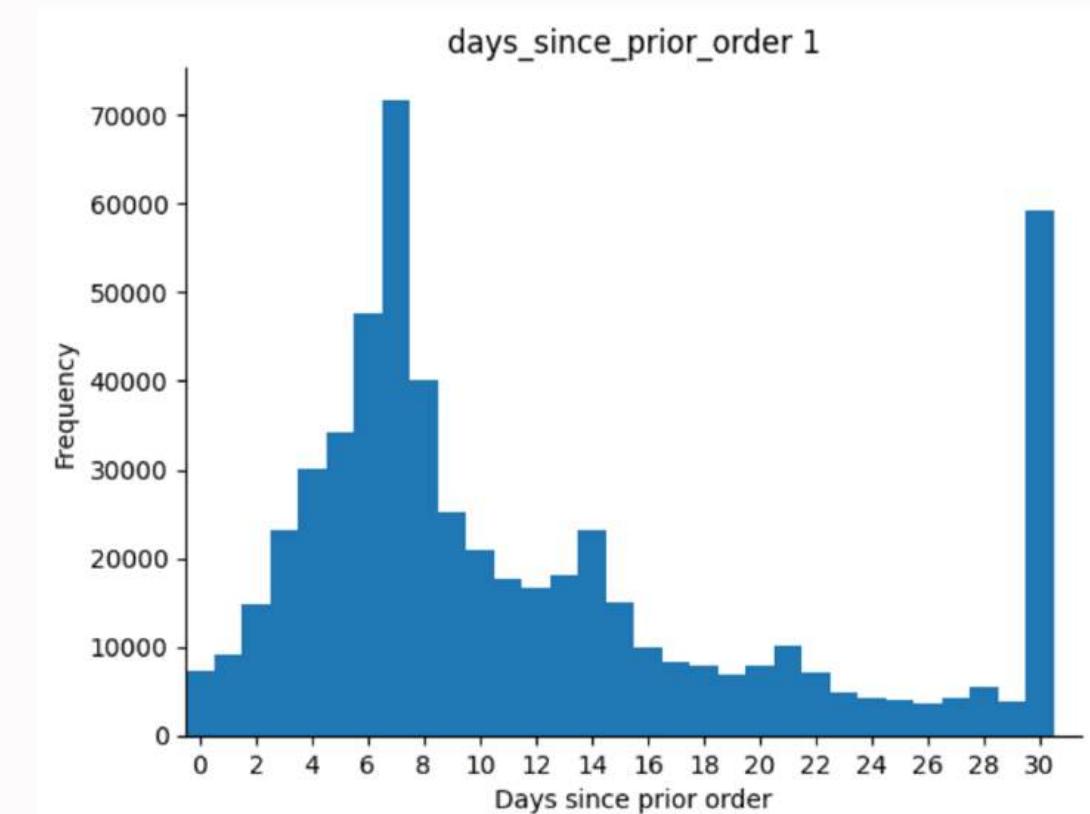
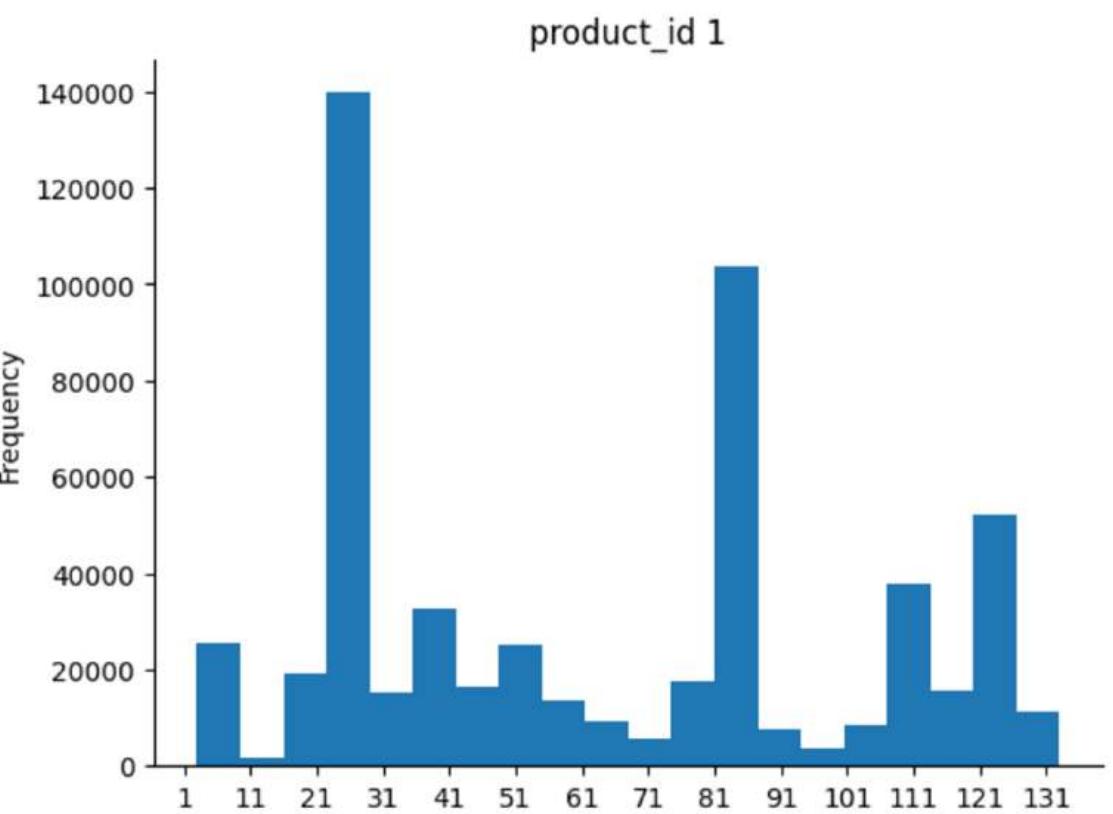
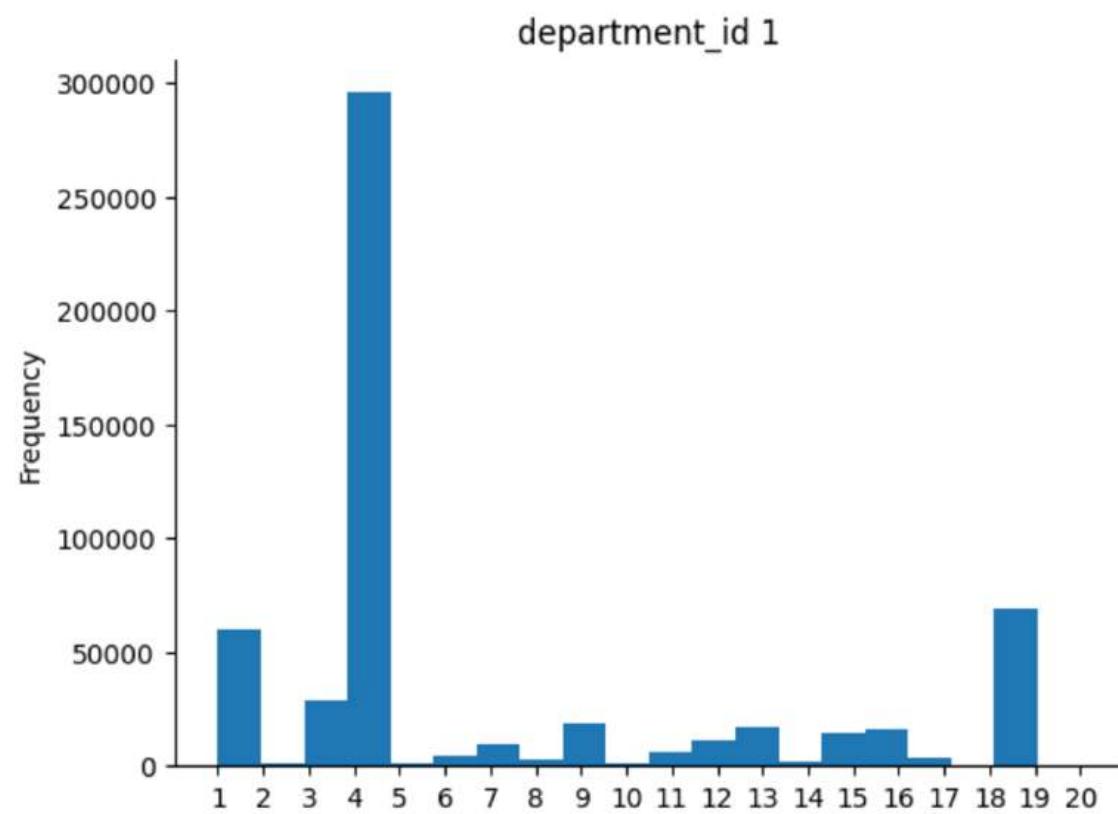
[97] df['Cluster'] = y_pred

df['Cluster'].value_counts()

count
cluster
1    614142
0    562483
3    484967
2    350384

dtype: int64
```

NHÓM 1

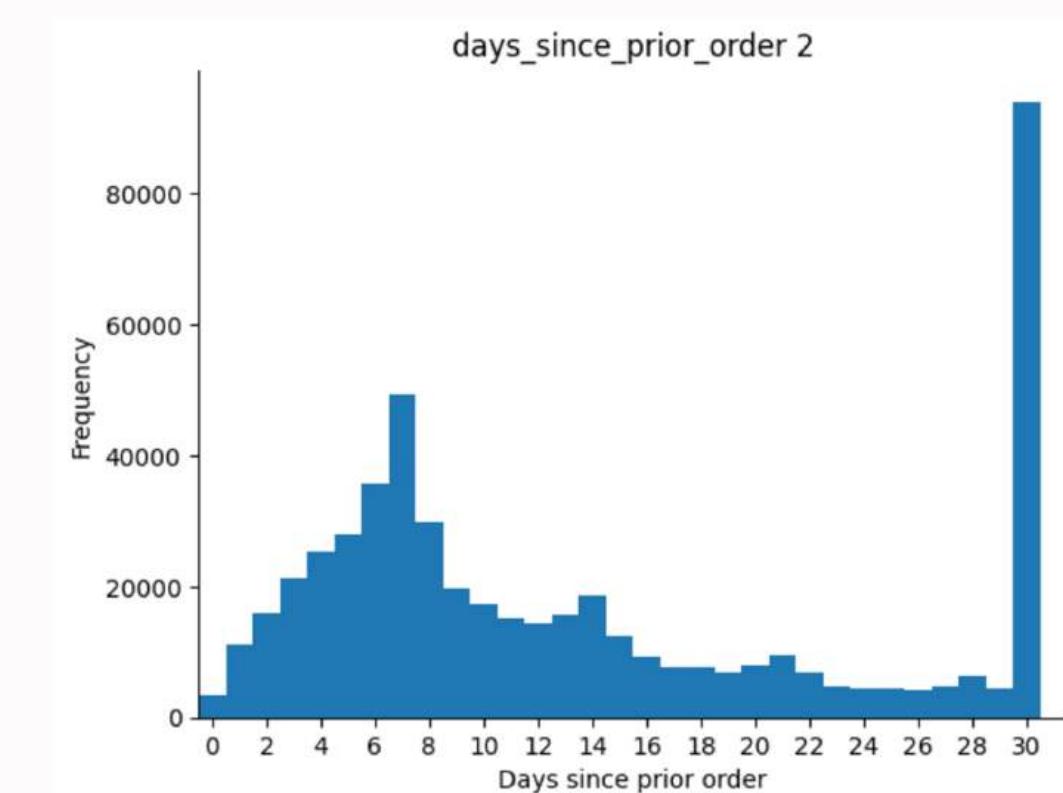
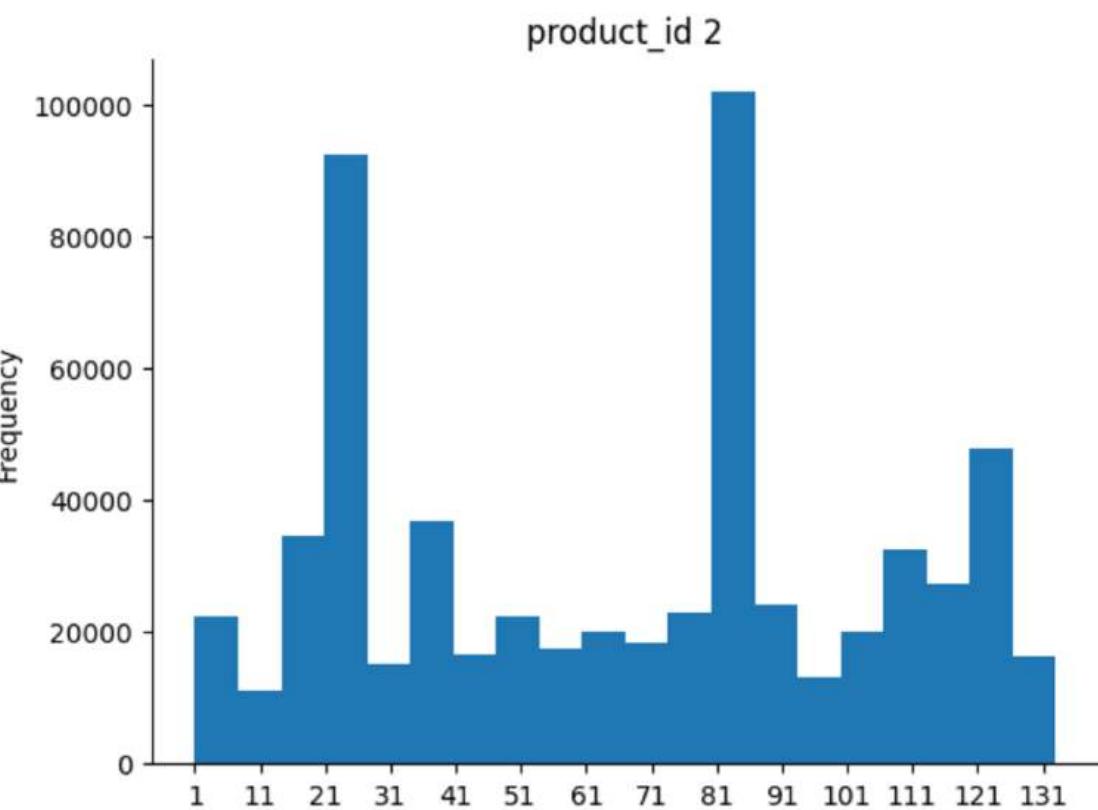
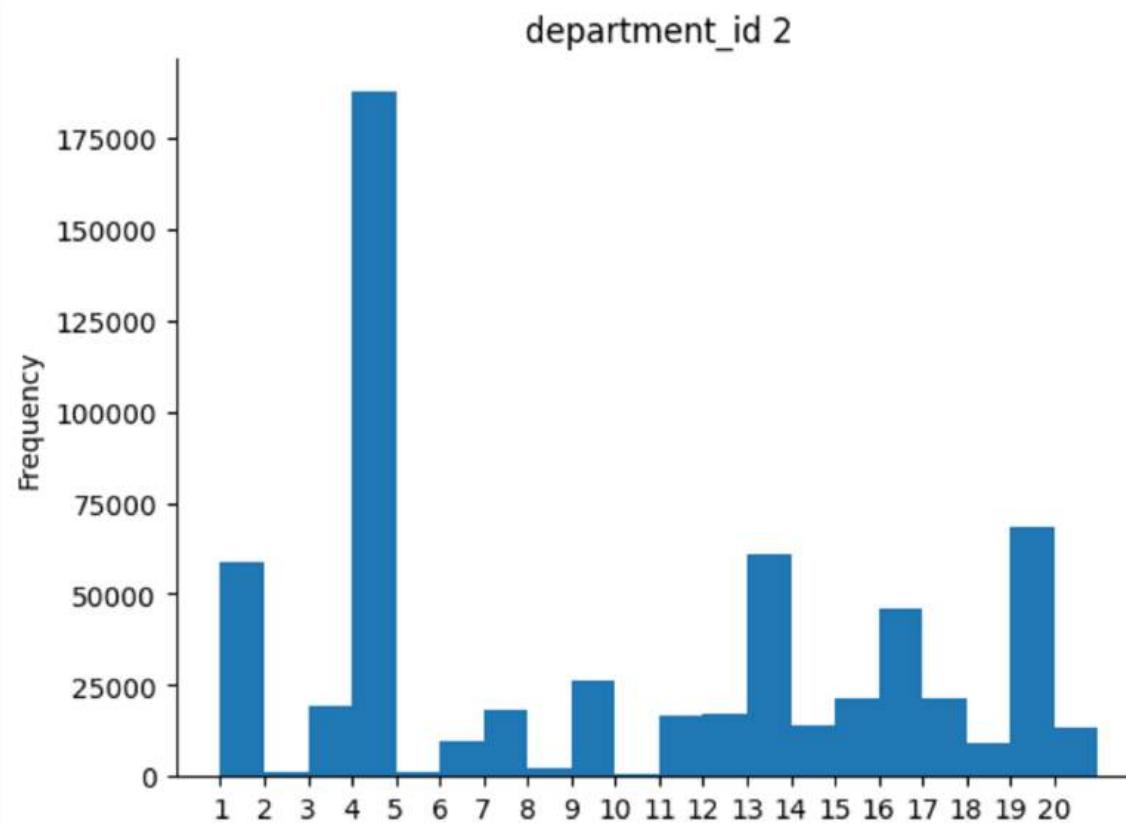


Quầy được mua nhiều nhất ở nhóm 1 là quầy số 4, đây là quầy produce (quầy nông sản tươi). Xếp sau đó là quầy 19 và 1, đây là quầy snacks và冻品 (thực phẩm đông lạnh)

Sản phẩm được mua nhiều nhất là 21 và 81, hai sản phẩm này là phô mai và đồ ăn đóng hộp

Chu kỳ mua hàng của nhóm này cao nhất là ở 7 ngày và xếp sau đó là 30 ngày

NHÓM 2

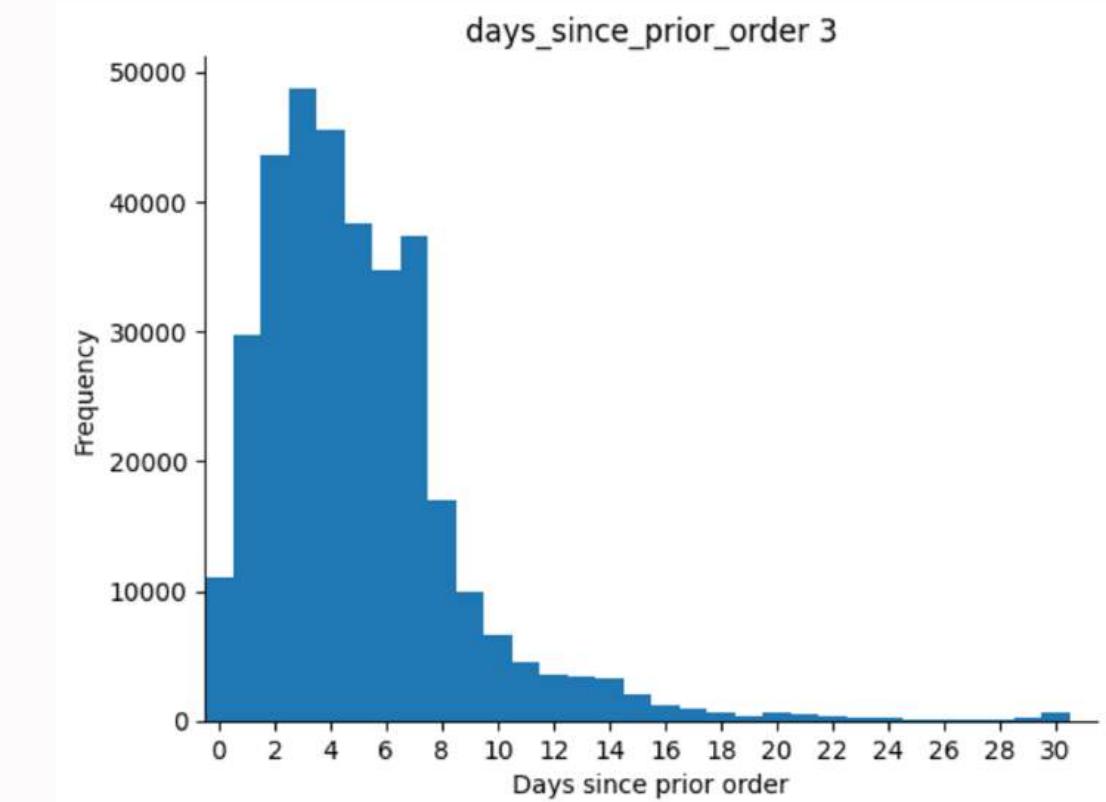
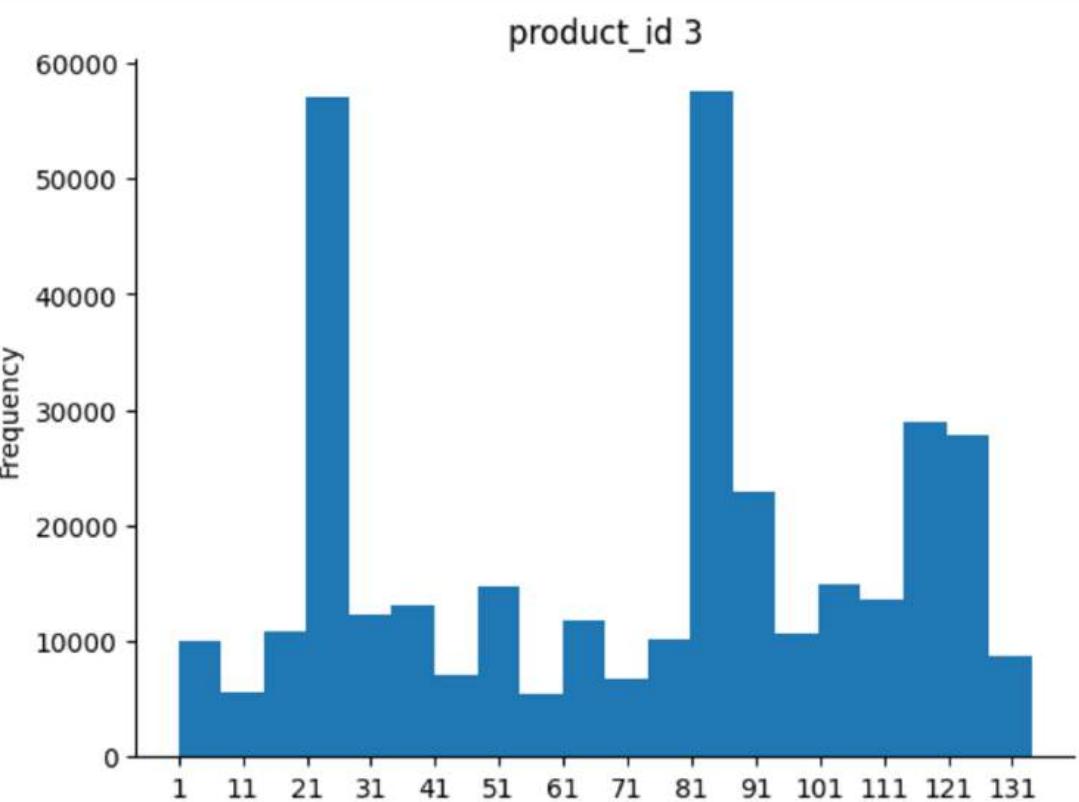
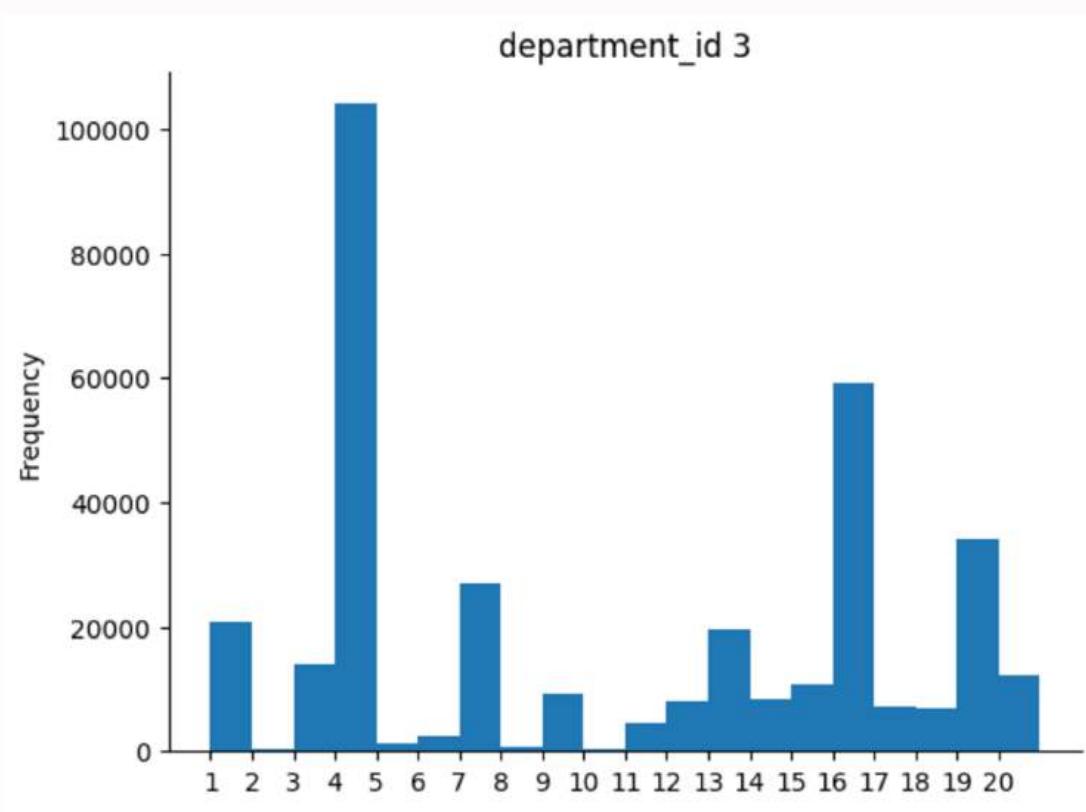


Quầy được mua nhiều nhất ở nhóm 2 là quầy số 4, đây là quầy produce (quầy nông sản tươi). Xếp sau đó là quầy 19 và 1, đây là quầy snacks và frozens (thực phẩm đông lạnh), có thêm cả quầy breakfast đứng vị trí thứ tư

Sản phẩm được mua nhiều nhất là 21 và 81, hai sản phẩm này là phô mai và đồng ăn đóng hộp, các sản phẩm khác cũng được mua nhiều hơn nhóm 1

Chu kỳ mua hàng của nhóm này cao nhất là ở 30 ngày và xếp sau đó là 7 ngày

NHÓM 3

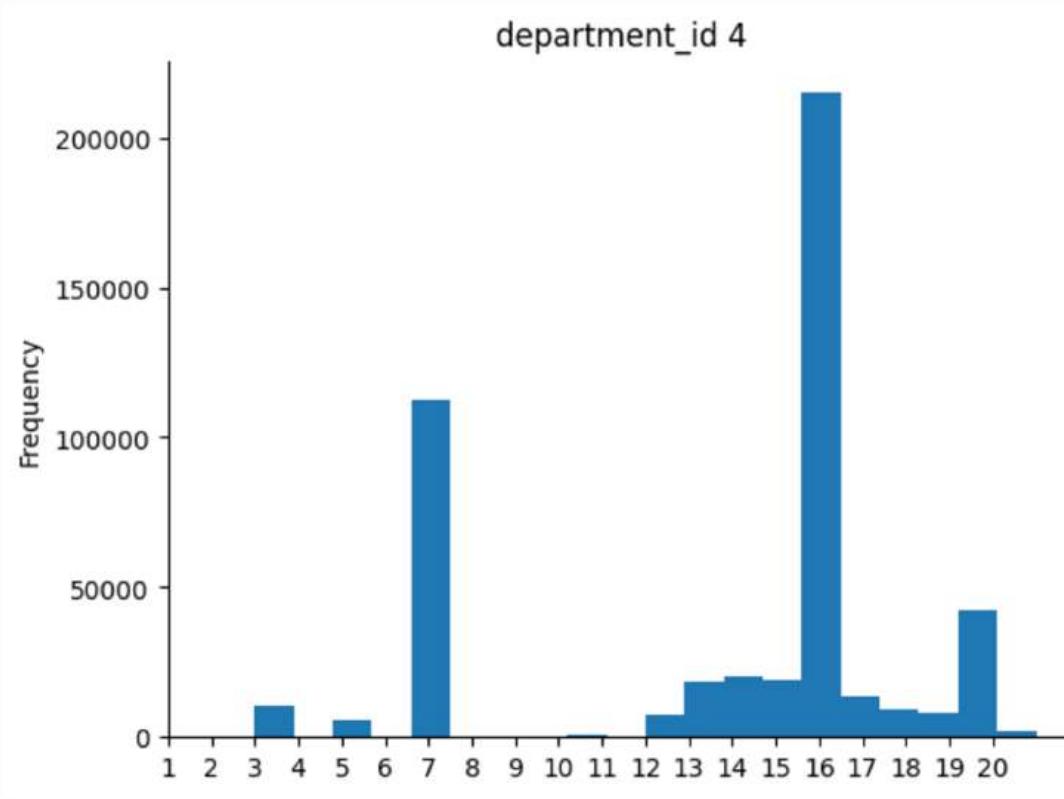


Quầy được mua nhiều nhất ở nhóm này là quầy số 4 (nông sản tươi) và quầy số 16 (thực phẩm từ trứng và sữa)

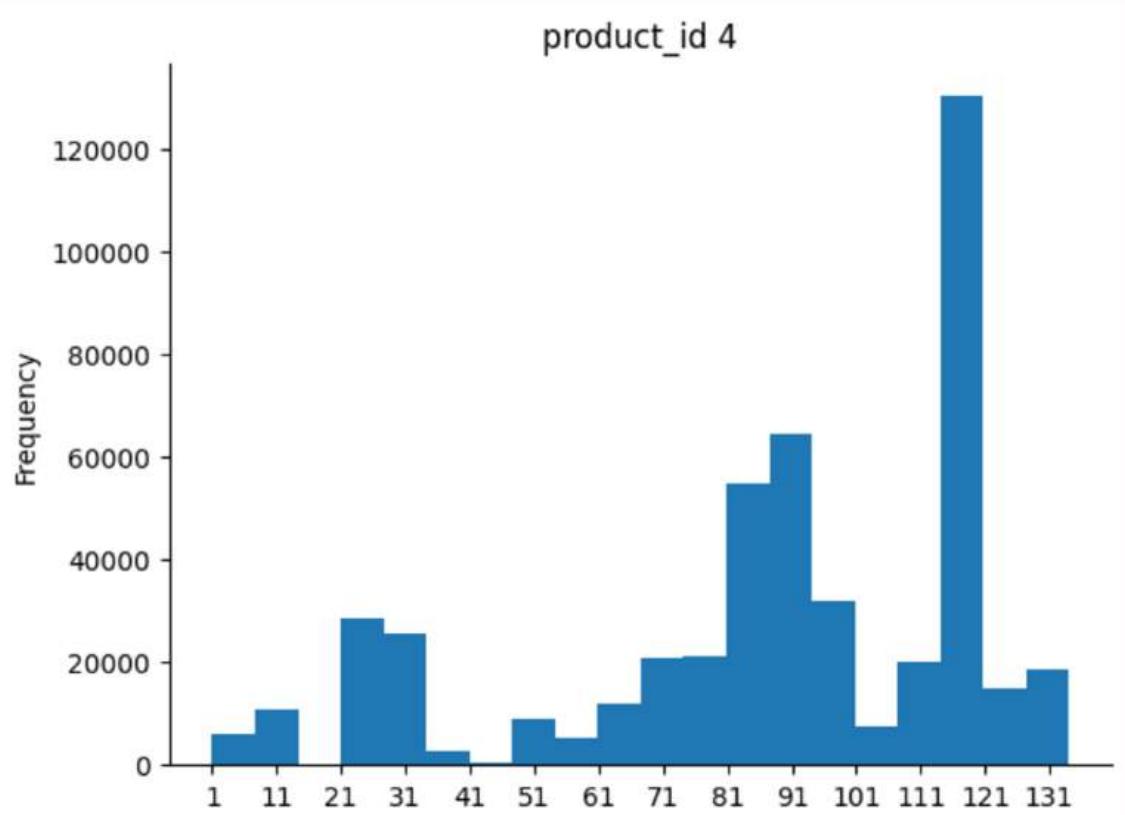
Sản phẩm được mua nhiều nhất là 21 và 81, hai sản phẩm này là phô mai và đồng ăn đóng hộp. Xếp thứ ba là sản phẩm 121 (ngũ cốc)

Chu kỳ mua hàng của nhóm này cao nhất nằm trong khoảng từ 1-7 ngày

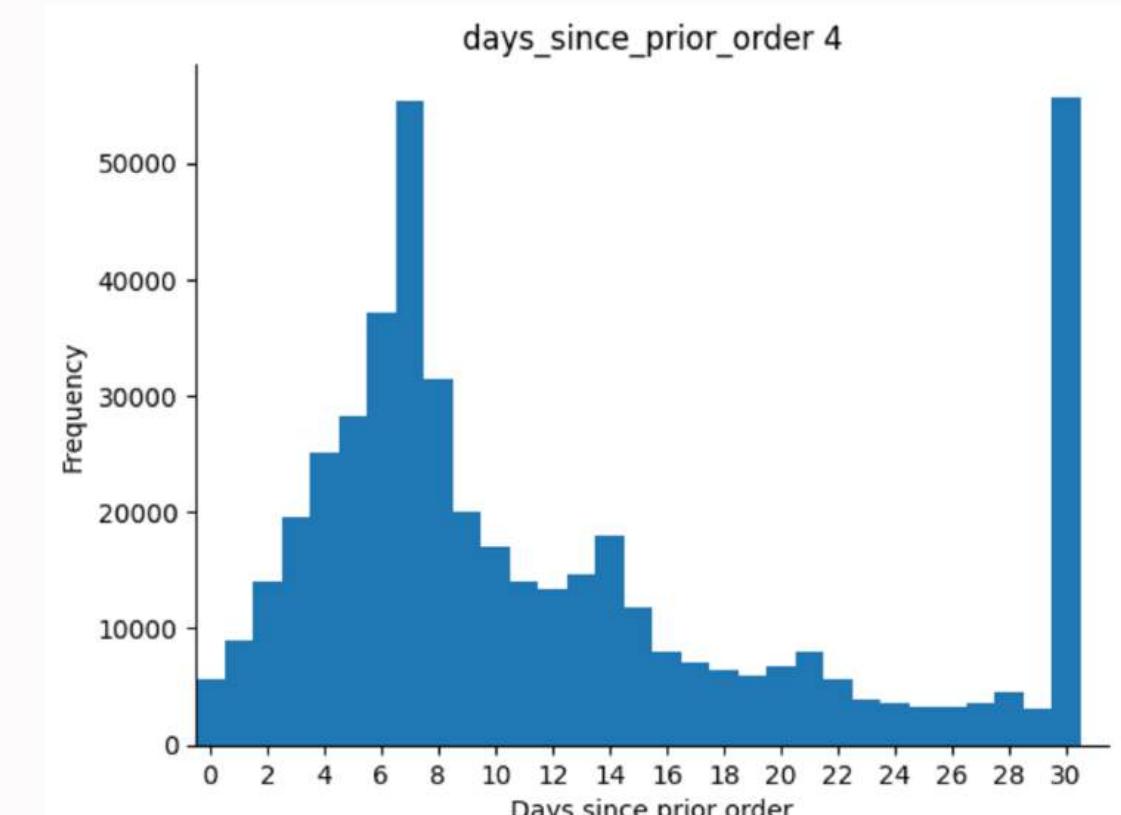
NHÓM 4



Quầy được mua nhiều nhất ở nhóm này là quầy số 16 và 7, 2 quầy này là dairy eggs (sản phẩm từ trứng và sữa) và beverages (đồ uống)



Sản phẩm được mua nhiều nhất là 121 (ngũ cốc), xếp sau đó là 21 và 81 (phô mai và đồ ăn đóng hộp)



Chu kỳ mua hàng của nhóm này cao nhất nằm ở cả 7 và 30 ngày

ASSOCIATION RULES: XỬ LÝ SỐ LIỆU

```
▶ df_apriori = df[["order_id", "product_name"]].sort_values(by="order_id", ascending = False)
df_apriori
```

	order_id	product_name
1739207	3421080	ice cream ice
1739214	3421080	butter
1739206	3421080	milk
1739208	3421080	juice nectars
1739209	3421080	fresh vegetables
...
1067804	10	fresh fruits
1067803	10	fresh herbs
1067802	10	fresh vegetables
1067811	10	fresh vegetables
1067801	10	fresh fruits

2011976 rows × 2 columns

ASSOCIATION RULES: DANH SÁCH GIAO DỊCH

```
# create transaction list
transactions = [a[1]['product_name'].tolist() for a in list(df_apriori.groupby(['order_id']))]

[ ] model_encoder = TransactionEncoder()
ls_trans = model_encoder.fit_transform(transactions)
df2 = pd.DataFrame(ls_trans, columns = model_encoder.columns_)
df2.replace([True, False], [1,0], inplace = True)
df2
```

	air fresheners candles	asian foods	baby accessories	baby body care	baby food formula	bakery desserts	baking ingredients	baking supplies decor	beauty	beers coolers	spreads
0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
...
195925	0	0	0	0	0	0	0	0	0	0	0	0
195926	0	0	0	0	0	0	0	0	0	0	0	0
195927	0	0	0	0	0	0	0	0	0	0	0	0
195928	0	0	0	0	0	0	0	0	0	0	0	0
195929	0	0	0	0	0	0	0	0	0	0	0	0

195930 rows x 134 columns

ASSOCIATION RULES: DANH SÁCH GIAO DỊCH

```
model_encoder = TransactionEncoder()
ls_trans = model_encoder.fit_transform(transactions)
df2 = pd.DataFrame(ls_trans, columns = model_encoder.columns_)
df2.replace([True, False], [1,0], inplace = True)
# Min Support: Để loại bỏ những itemset lẻ ít xuất hiện
df_support = apriori(df2, min_support = 0.1, use_colnames = True)
df_support.sort_values("support", ascending = False)
from mlxtend.frequent_patterns import association_rules
df_associate = association_rules(df_support, metric = "lift", min_threshold = 1)
# antecedents: A (xuất hiện trước)
# consequents: B (Xuất hiện sau) A-> B
df_associate.sort_values(["support", "confidence", "lift"], ascending = False)
```

Xây dựng Association Rules để tìm sản phẩm thường được mua cùng nhau

ASSOCIATION RULES: CÁC CHỈ SỐ

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
4	(fresh vegetables)	(fresh fruits)	0.452177	0.565202	0.323784	0.716056	1.266903	0.068213	1.531282	0.384565
5	(fresh fruits)	(fresh vegetables)	0.565202	0.452177	0.323784	0.572864	1.266903	0.068213	1.282551	0.484533
11	(packaged vegetables fruits)	(fresh fruits)	0.371868	0.565202	0.275027	0.739583	1.308529	0.064847	1.669621	0.375371
10	(fresh fruits)	(packaged vegetables fruits)	0.565202	0.371868	0.275027	0.486599	1.308529	0.064847	1.223474	0.542281
23	(packaged vegetables fruits)	(fresh vegetables)	0.371868	0.452177	0.239055	0.642849	1.421677	0.070905	1.533871	0.472202
...
33	(milk, fresh fruits)	(fresh vegetables)	0.167305	0.452177	0.101444	0.606345	1.340947	0.025793	1.391634	0.305344
35	(milk)	(fresh vegetables, fresh fruits)	0.246996	0.323784	0.101444	0.410712	1.268475	0.021471	1.147514	0.281077
34	(fresh vegetables, fresh fruits)	(milk)	0.323784	0.246996	0.101444	0.313309	1.268475	0.021471	1.096568	0.312995
36	(fresh vegetables)	(milk, fresh fruits)	0.452177	0.167305	0.101444	0.224347	1.340947	0.025793	1.073541	0.464125
37	(fresh fruits)	(milk, fresh vegetables)	0.565202	0.127076	0.101444	0.179483	1.412411	0.029621	1.063871	0.671554

Các Associate Rules mạnh nhất:

- (fresh vegetables) → (fresh fruits)
- (fresh fruits) → (fresh vegetables)
- (packaged vegetables fruits) → (fresh fruits)
- (fresh fruits) → (packaged vegetables fruits)
- (fresh fruits, milk) → (fresh vegetables)
- (milk) → (fresh fruits, fresh vegetables)

TỔNG KẾT

Phần EDA:

- Produce (rau củ) và Dairy eggs (sữa và trứng) là hai department có số lượng đơn hàng cao nhất, vì vậy cửa hàng có thể tập trung vào việc giữ nguồn cung ổn định cho những mặt hàng này, đồng thời sử dụng chúng để thu hút khách hàng quay lại thường xuyên hơn.
- Từ các sản phẩm được reorder nhiều nhất theo chu kỳ như rau củ và thực phẩm từ trứng sữa, siêu thị có thể triển khai chương trình đặt hàng định kỳ và khuyến mãi.
- Ngoài các sản phẩm rau củ quả, snacks (đồ ăn nhẹ) cũng nằm trong top department có số lượng đơn hàng lớn. Siêu thị có thể tận dụng upselling và cross-selling khi khách hàng mua sắm thực phẩm thiết yếu bằng cách đề xuất các sản phẩm đồ ăn nhẹ phù hợp

Phần Clustering:

- Nhóm 1: Đây có thể là nhóm gia đình, thể hiện ở việc họ có thể mua sắm hàng tuần cho thực phẩm tươi và hàng tháng cho thực phẩm dự trữ.
- Nhóm 2: Có thể là nhóm gia đình lớn hơn hoặc có thu nhập cao hơn nhóm 1. Họ quan tâm đến bữa sáng, có thể do lối sống bận rộn. Xu hướng mua sắm định kỳ hàng tháng rõ rệt.
- Nhóm 3: Có thể là nhóm người trẻ, sống độc thân hoặc cặp đôi. Họ ưa chuộng thực phẩm tiện lợi (ngũ cốc) nhưng vẫn quan tâm đến dinh dưỡng (sữa, trứng). Tần suất mua sắm thường xuyên hơn, có thể do lối sống linh hoạt.
- Nhóm 4: Có thể là nhóm gia đình có con nhỏ hoặc người cao tuổi. Họ quan tâm đến sản phẩm giàu canxi (sữa) và đồ uống. Xu hướng mua sắm kết hợp giữa hàng tuần và hàng tháng, cho thấy sự cân bằng giữa tiện lợi và dự trữ.

➡ Mặc dù các nhóm không có sự khác biệt lớn về phân khúc sản phẩm, nhưng những khác biệt nhỏ kết hợp với sự khác nhau giữa chu kỳ mua hàng của các nhóm đã giúp tạo nên một bức tranh chân dung khách hàng rõ nét.

TỔNG KẾT

Phản Association Rules:

- Rau củ quả đóng gói có mối liên hệ mạnh với trái cây tươi, cho thấy khách hàng thường kết hợp cả hai loại trong giỏ hàng.
 - Sữa thường được mua cùng với trái cây tươi và rau củ tươi, nhưng mối liên hệ không mạnh bằng giữa các loại rau củ quả với nhau.
 - Lift > 1 trong tất cả các trường hợp, cho thấy các sản phẩm này có xu hướng được mua cùng nhau nhiều hơn so với khi mua riêng lẻ.
-
- Cửa hàng nên sắp xếp rau củ, trái cây tươi và sữa xếp gần nhau
 - Tạo combo hoặc khuyến mãi kết hợp các sản phẩm có liên quan mạnh như là combo "bữa ăn lành mạnh" gồm rau củ, trái cây tươi và sữa.
 - Khi khách hàng mua sữa, gợi ý thêm trái cây và rau củ tươi để tăng giá trị đơn hàng.
 - Tạo chương trình khuyến mãi "mua 2 tặng 1" cho bộ ba sản phẩm: rau củ tươi, trái cây tươi và rau củ quả đóng gói.

THANK YOU



Thank You!!