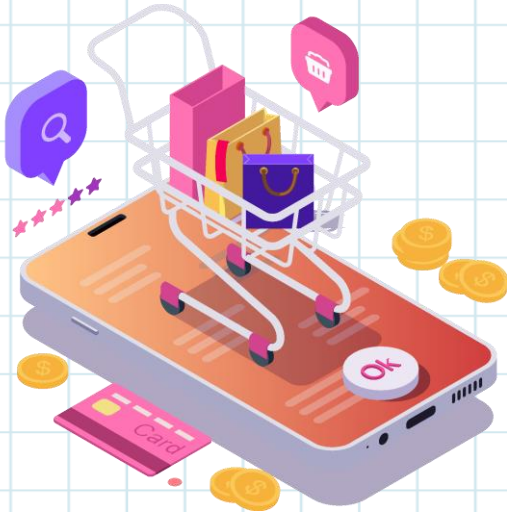




# CUSTOMER SEGMENTATION

AMARA KAWEEKATE 6310412014

# Customer Segment คืออะไร



การที่เราอยากจะแบ่งกลุ่มลูกค้า  
ของออกเป็นกลุ่มย่อย ๆ โดยที่กลุ่มย่อย ๆ  
เราคาดหวังว่าเขาจะมีคุณลักษณะที่  
เหมือน ๆ กัน เช่น ชอบซื้อของแพง ใช้เงิน  
ไม่อื่น ไม่แคร์ส่วนลด

**ข้อดีคือ** ถ้าเราเข้าใจลูกค้าเวลาทำ  
Marketing แคมเปญหรือทำ Promotion  
เราก็จะสามารถ Target ให้มันตรงกลุ่ม  
ตรงความต้องการของลูกค้าได้มาก  
ยิ่งขึ้น

## ทำไม Segment ถึงเป็นเรื่องจำเป็น ?

ถ้าเรามองลูกค้าทั้งหมดพร้อม ๆ กัน  
สิ่งที่เกิดขึ้นคือมันจะมีลูกค้าที่แตกต่างกันมาก ๆ  
ที่ปะปนกัน Insight ก็จะออกมาไม่ชัด  
เราก็จะหาไม่เจอสักที ไม่ Take Action  
ทางธุรกิจยาก



# Customer Segment Analysis

with Python

Load Data

“Supermarket Data”



Prepare Data

Group Data Supermarket  
เป็น Single View เพื่อเอา  
เข้า K-means Clustering

Calculate  
Features

จากนั้นจะได้ Data  
ที่พร้อมเข้า K-means

Compare Model  
Performance

โดยการรันโมเดลทุก ๆ ตัว  
และเก็บค่าเมตริกชี้ ออกมา  
และดูว่าโมเดลไหนดีที่สุด

Cluster  
Customers

ขั้นตอนนี้ใช้ PyCaret  
setup environment

K-Means  
Clustering

จากการจัดกลุ่มด้วย K-means  
ดีกว่า Spectral Clustering  
จากการ Plot กราฟ Elbow ค่า  
Silhouette ผ่านทุกตัว เมื่อจัดกลุ่ม  
แล้วจะได้ k=4 คือค่า K ที่เหมาะสม

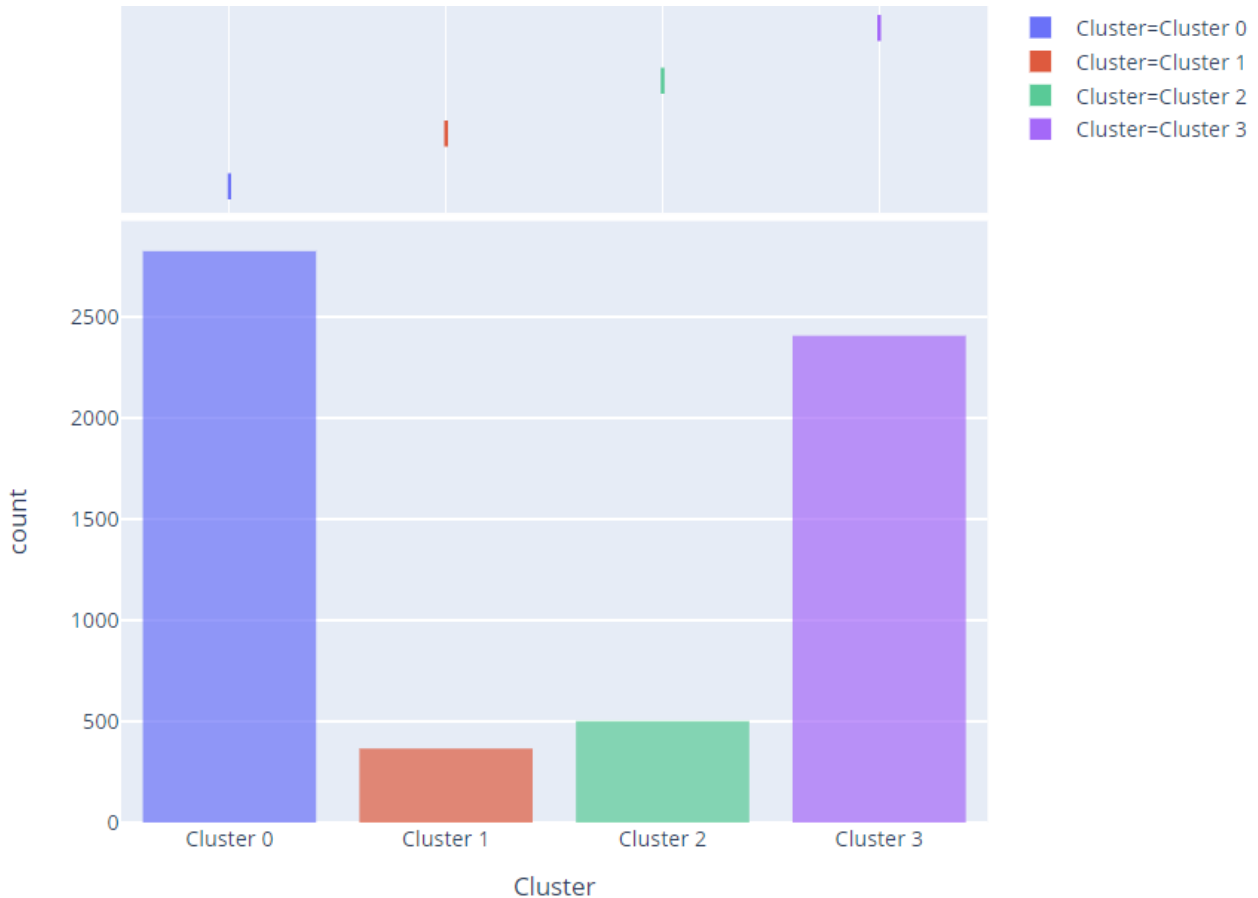


# FIND FEATURE



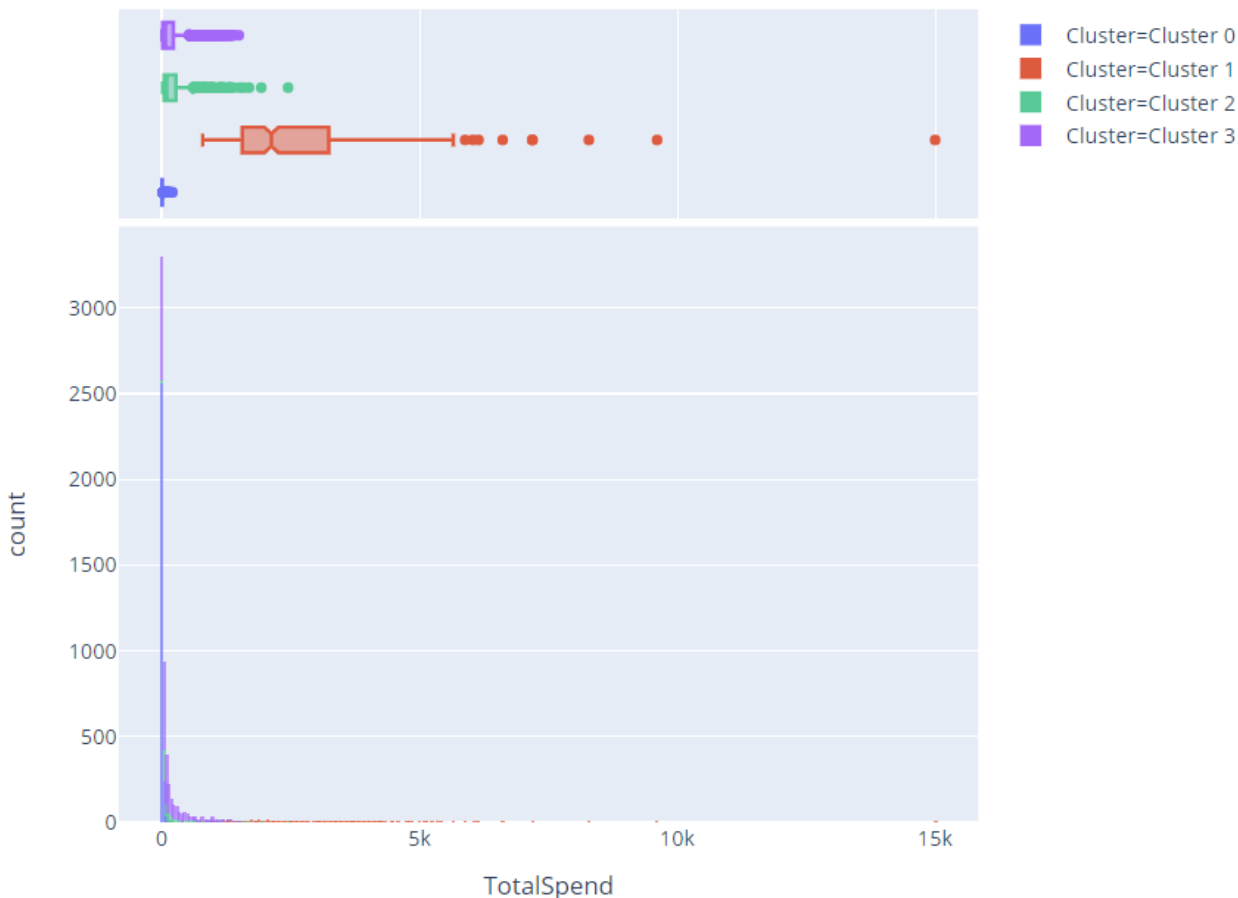
AVG. quantity total visit

# Result by K-mean Clustering



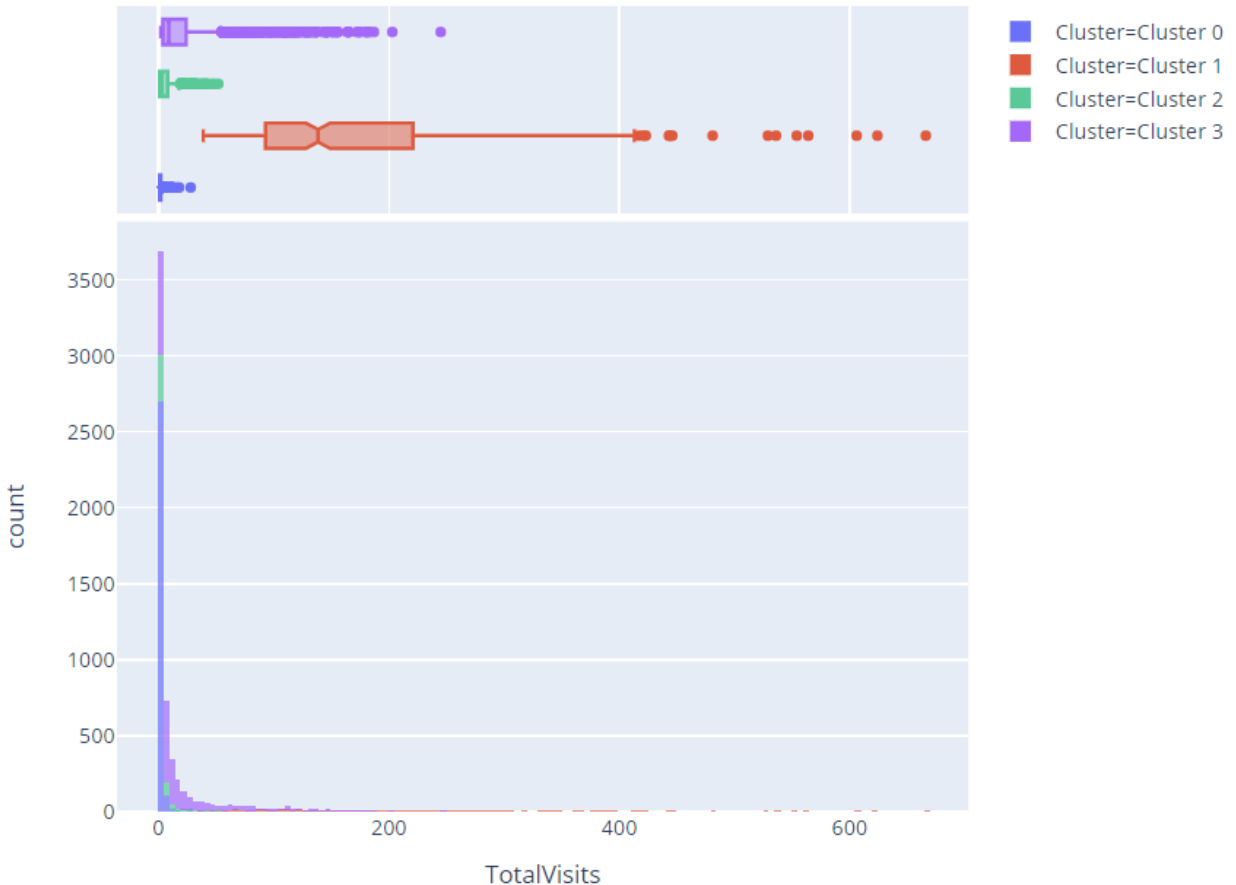
แสดงจำนวนลูกค้า Supermarket ในแต่ละ Cluster โดยมีทั้งหมด 4 Cluster โดยจะเห็นได้ว่า Cluster 0 และ 3 มีจำนวนลูกค้าสูงสุด

# Result by K-mean Clustering



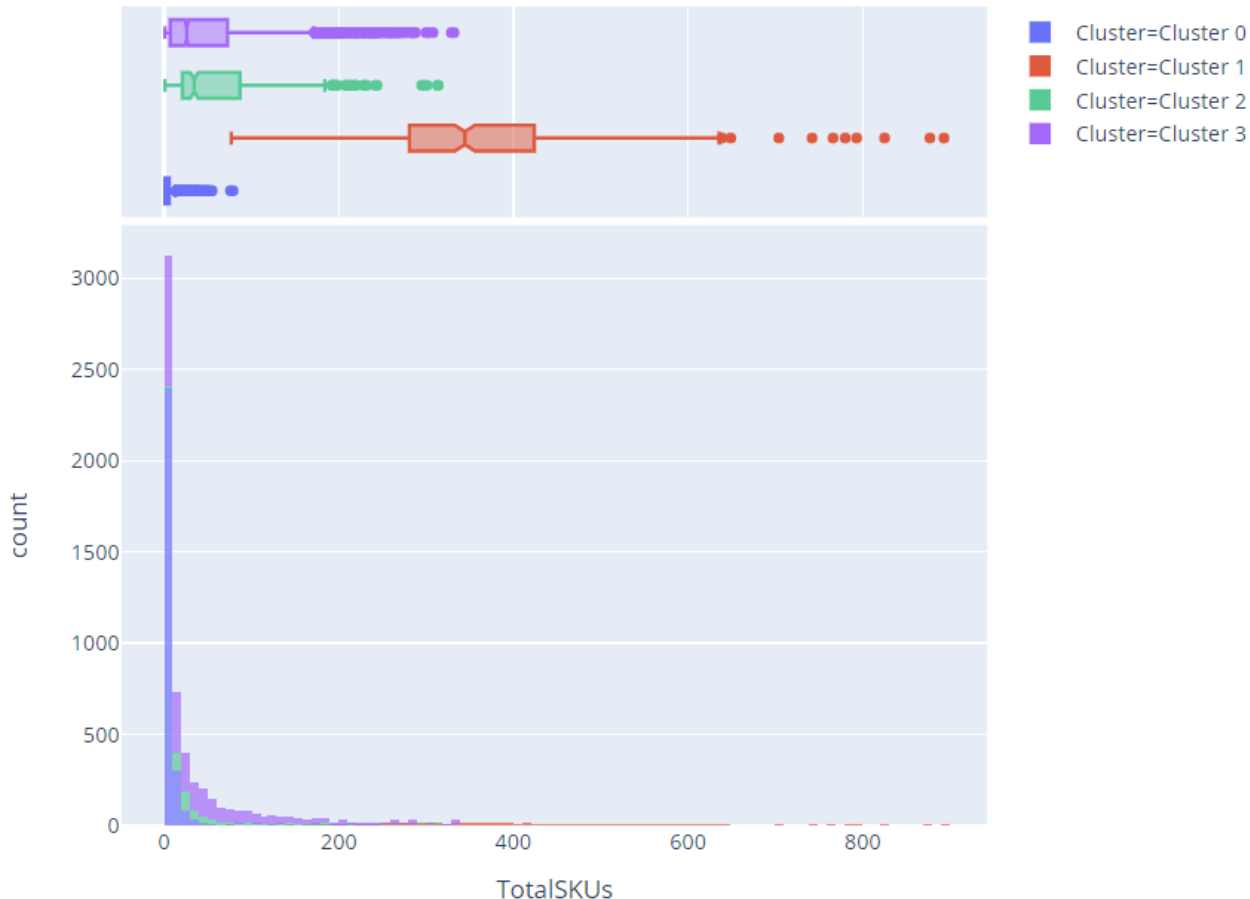
จากกราฟ Boxplot จะเห็นว่า Cluster 1 มี total spend สูง  
ในขณะที่ Cluster อื่น ๆ total spend ค่อนข้างน้อย

# Result by K-mean Clustering



จากกราฟ Boxplot จะเห็นว่า Cluster 1 มา Supermarket บ่อยมาก ส่วน Cluster 3 ก็มาบ้าง ในขณะที่ Cluster 0 และ 2 นาน ๆ ครั้งถึงจะมา

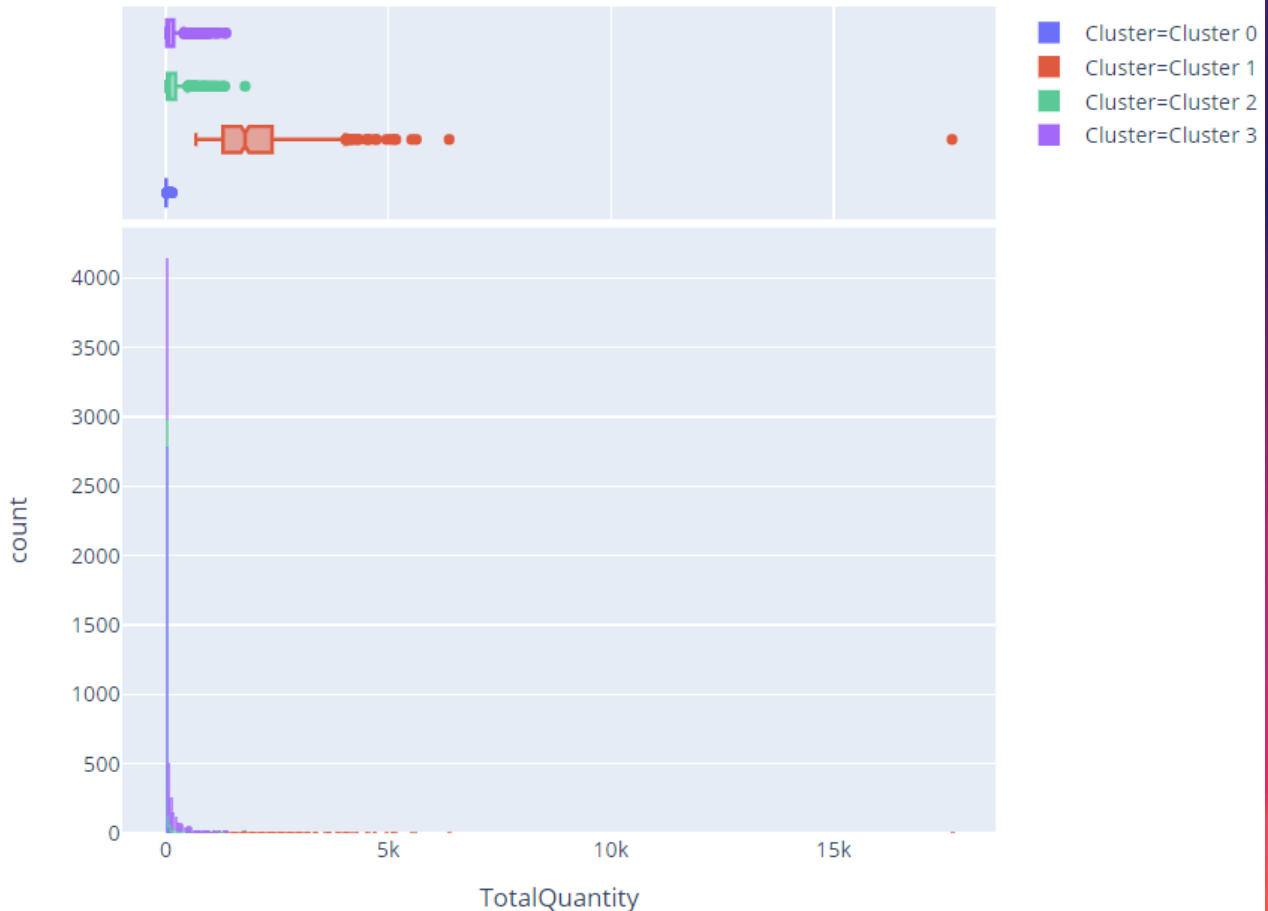
# Result by K-mean Clustering



จากกราฟ Boxplot จะเห็นว่า Cluster 1 จำนวนชนิด product ที่ซื้อหลากหลาย รองลงมา ก็จะเป็น Cluster 2 และ 3 ในขณะที่ Cluster 0 น้อยมาก

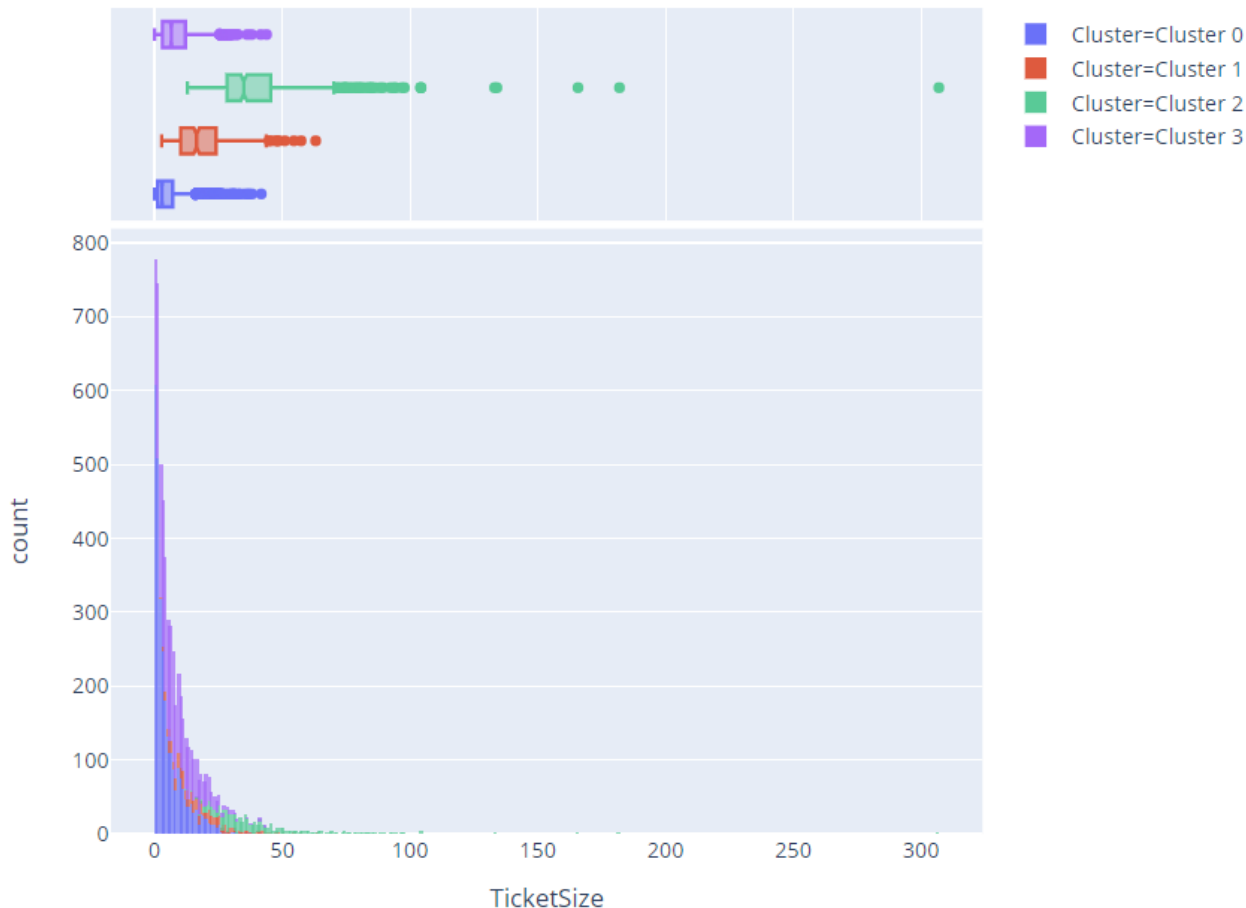


# Result by K-mean Clustering



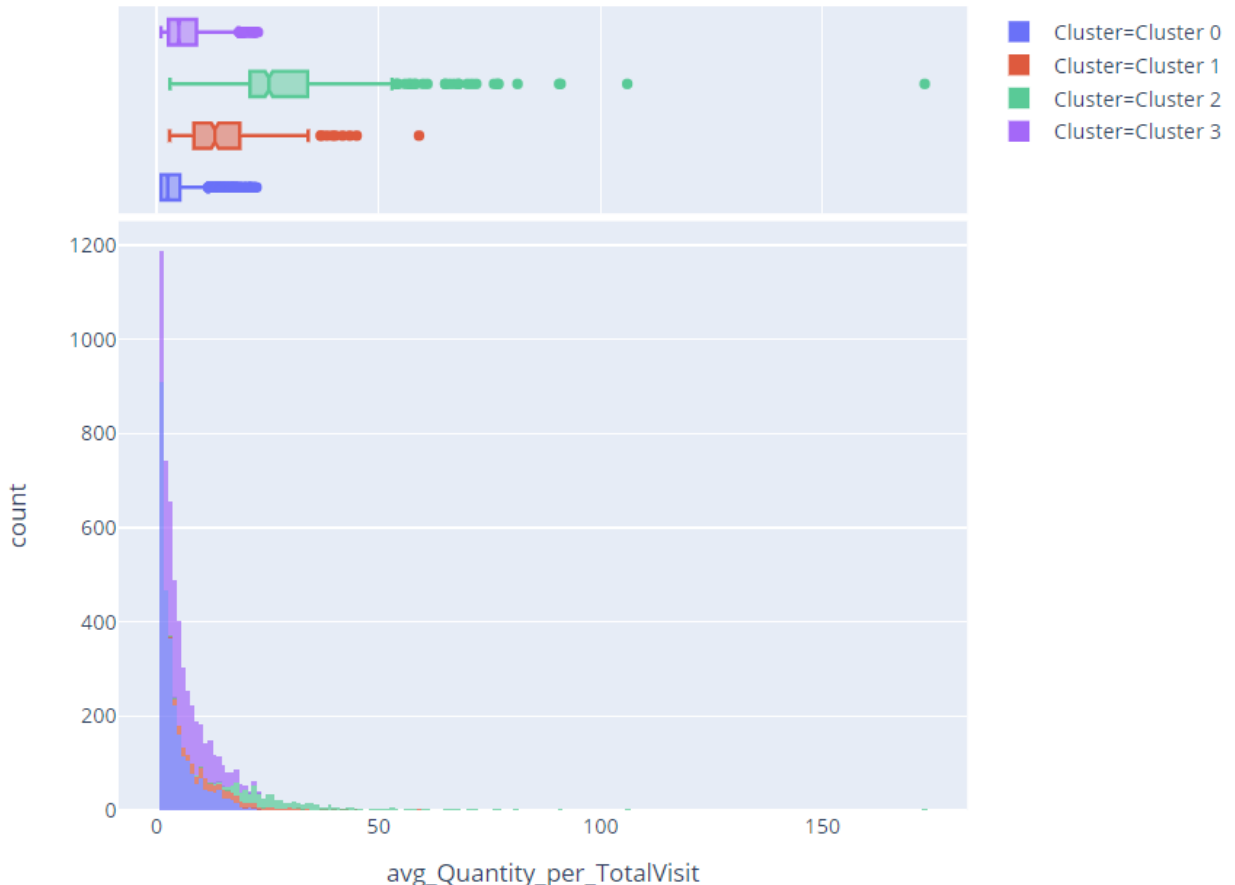
จากกราฟ Boxplot จะเห็นว่า Cluster 1 ชื่อของทีละจำนวนมาก  
ในขณะที่ Cluster อื่นๆ ชื่อทีละน้อยๆ

# Result by K-mean Clustering



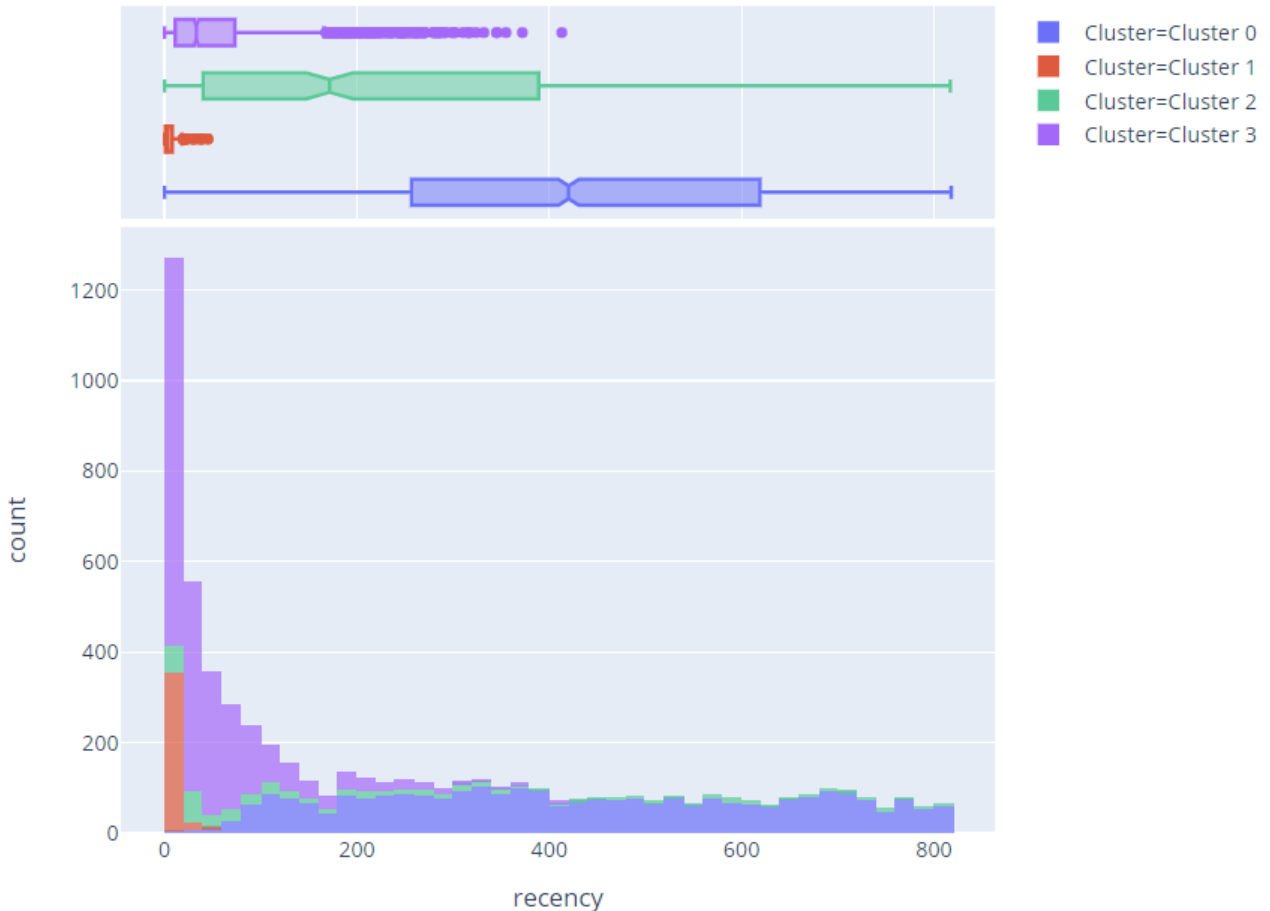
เมื่อพิจารณา Ticket Size (ยอดขายรวมใน 1 ใบเสร็จ) พบว่า Cluster 2  
มียอดใช้จ่ายต่อ 1 ใบเสร็จเยอะที่สุด รองลงมาคือ Cluster 1

# Result by K-mean Clustering




เมื่อพิจารณาค่าเฉลี่ยจำนวนของที่ซื้อต่อจำนวนครั้งที่มา พบว่า Cluster 2  
มีค่าเฉลี่ยสูงสุด รองลงมาคือ Cluster 1

# Result by K-mean Clustering



เมื่อพิจารณา Recency (ล่าสุดที่ลูกค้ามาจนถึงปัจจุบัน) พบว่า Cluster 0 เป็นลูกค้าที่หายไปนานมาก รองลงมาคือ Cluster 2 ส่วน Cluster 1 มาบ่อยสม่ำเสมอ

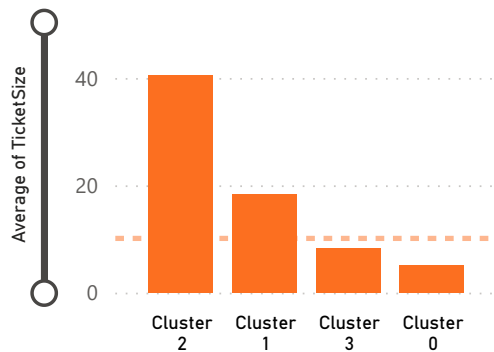
# Summary and Action Plan

	Cluster 0	Cluster 1	Cluster 2	Cluster 3
 <b>TICKET SIZE</b>	น้อย	สูง	สูงที่สุด	น้อย
 <b>SPEND</b>	น้อย	ใช้จ่ายสูง	น้อย	น้อย
 <b>QUANTITY</b>	น้อย	เยอะ	น้อย	น้อย
 <b>VISIT</b>	มานาน ๆ ครั้ง	มาถี่	มาไม่บ่อย	มานาน ๆ ครั้ง
 <b>SKUs</b>	น้อยมาก	เยอะ	น้อย	ไม่เยอะมาก
 <b>Day</b>	สั้น	นาน	ค่อนข้างนาน	นาน
 <b>Recency</b>	หายไปนานมาก	มาบ่อย สม่ำเสมอ	หายไป ค่อนข้างนาน	ยังมาแต่ไม่บ่อยนัก
 <b>Action</b>	<b>Lost Customer</b> จัดโปรแรง ๆ เช่น วันคนโสด หรือจัดโปรหนัก ๆ มากกระตุ้นให้กลับมาสนใจซื้ออีกครั้ง	<b>Champions</b> รักษาลูกค้ากลุ่มนี้ไว้ให้อยู่กับเราให้นานที่สุด	<b>Should not Lose</b> โทรหาพวกเขาโดยตรงเพื่อให้เห็นว่าเราให้ความสำคัญให้โปรพิเศษสำหรับพวกเขาเท่านั้น	<b>Need Attention</b> แนะนำสินค้ายอดนิยมในเวลานั้น หรือมอบส่วนลดบางอย่างให้เพื่อกระตุ้นให้พวกเขากลับมาซื้ออีกครั้ง

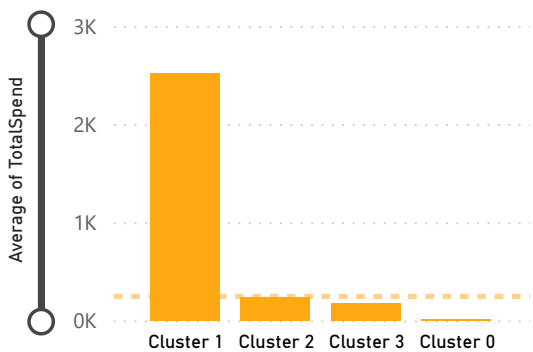
# Interpret Results by Customer Segment

Cluster	Count of Cluster	Average of TicketSize	Average of TotalSpend	Average of TotalQuantity	Average of TotalVisits	Average of TotalSKUs	Average of total_days	Average of recency
Cluster 0	2825	5.21	9.65	7.66	1.73	5.14	47.08	433.48
Cluster 1	366	18.31	2,519.41	2071.14	172.18	367.94	792.38	5.65
Cluster 2	502	40.49	244.23	182.66	6.67	61.05	288.96	248.75
Cluster 3	2407	8.27	178.17	140.50	21.16	50.98	604.40	54.51
Total	6100	10.11	246.04	198.29	20.03	49.60	331.62	243.07

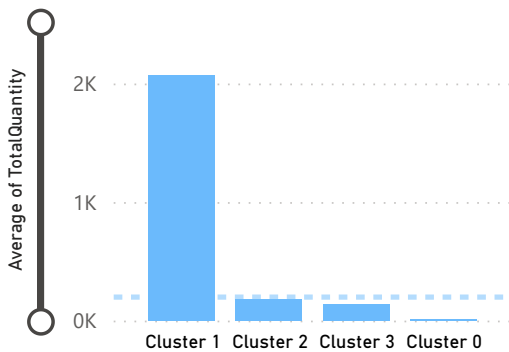
Average of TicketSize by Cluster



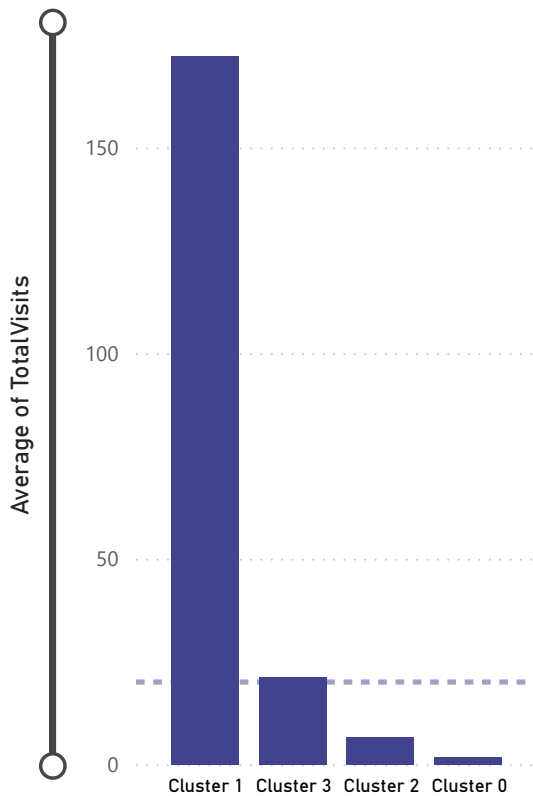
Average of TotalSpend by Cluster



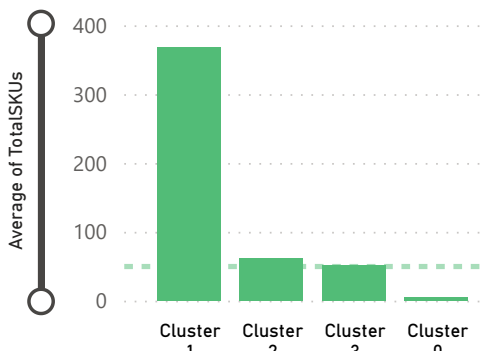
Average of TotalQuantity by Cluster



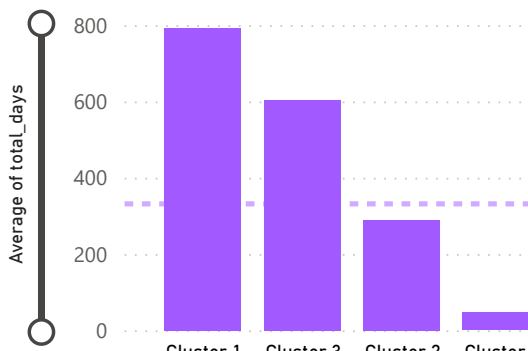
Average of TotalVisits by Cluster



Average of TotalSKUs by Cluster



Average of total\_days by Cluster



Average of recency by Cluster

