

# Data Cafe "Data Scientist" Qualification Challenge

## Supermarket Analysis

By

Wachirawit Iedchoo



## Supermarket Analysis

การรู้จักลูกค้าของตัวเองไม่ใช่แค่เป็นสิ่งสำคัญแต่เป็นสิ่งจำเป็นสำหรับการประกอบธุรกิจ การรู้จักลูกค้าทำให้ผู้ประกอบการสามารถวางแผนธุรกิจได้มีประสิทธิภาพมากขึ้น ทำให้ผู้ประกอบการจะรู้ว่าควรโฟกัสที่ลูกค้าประเภทไหน และควรวางแผนการตลาดให้เหมาะสมกับลูกค้าแต่ละประเภทอย่างไร จากประโยชน์ข้างต้นทำให้ผมเลือกตั้งคำถามในการวิเคราะห์ข้อมูลดังนี้

### “ใครคือลูกค้าหลักของธุรกิจ?”

โดยประโยชน์ที่ธุรกิจจะได้รับจากการตอบคำถามนี้คือ

1. ทำให้รู้จักลูกค้ามากขึ้น
2. ทำให้รู้ว่าควรโฟกัสไปที่ลูกค้ากลุ่มไหน
3. ทำให้Marketerรู้ว่าควรวางแผนการตลาดให้เหมาะสมกับลูกค้ากลุ่มนี้ได้  
อย่างไร
4. ทำให้รู้ความต้องการและพฤติกรรมของลูกค้ากลุ่มนี้

### วิธีที่ใช้ในการวิเคราะห์

จากคำถามข้างต้นจึงเลือกใช้การทำ Customer Segmentation โดยใช้เทคนิคการวิเคราะห์ข้อมูล RFM Analysis มาช่วยในการค้นหาลูกค้าหลัก

โดยวิธีการวิเคราะห์แบบ RFM จะแบ่งลูกค้าด้วยเกณฑ์ 3 เกณฑ์ คือ

1. Recency คือ จำนวนวันที่ลูกค้ามาซื้อครั้งล่าสุดนับจากปัจจุบัน
2. Frequency คือ จำนวนครั้งของการซื้อสินค้า
3. Monetary คือ ยอดรวมที่ลูกค้าเคยซื้อ

จากนั้นทำการคำนวณค่า Recency, Frequency และ Monetary ของลูกค้าแต่ละดังนี้

	CUST_CODE	monetary	frequency	recency
0	CUST0000000107	0.800000	1	103
1	CUST0000000369	178.763940	126	1
2	CUST0000001388	3.814286	4	86
3	CUST0000002302	153.479111	71	3
4	CUST0000002637	8.821894	7	37

จากนั้นทำการคำนวณ R,M,F score โดยวิธีการคำนวณ R,M, F score ทำเหมือนกัน โดยการเปรียบเทียบข้อมูล R,M,F ของลูกค้าแต่ละคนกับ 3rd Quantile ของแต่ละเกณฑ์โดยมีวิธีการกำหนด R,M,F score ดังนี้

- M,F score: ถ้าข้อมูลมากกว่า **3rd Quantile** จะมี **M,F score** เท่ากับ **1** แต่ถ้าน้อยกว่าจะเท่ากับ **2**
- R score: ถ้าข้อมูลน้อยกว่า **3rd Quantile** จะมี **R score** เท่ากับ **1** แต่ถ้ามากกว่าจะเท่ากับ **2**

ได้ผลดังนี้

	CUST_CODE	monetary	frequency	recency	m_score	r_score	f_score	rmf
0	CUST0000000107	0.800000	1	103	2	1	2	122
1	CUST0000000369	178.763940	126	1	1	1	1	111
2	CUST0000001388	3.814286	4	86	2	1	2	122
3	CUST0000002302	153.479111	71	3	1	1	1	111
4	CUST0000002637	8.821894	7	37	2	1	2	122

Note: เพิ่ม column rmf เพื่อให้ง่ายต่อการจัดกลุ่ม

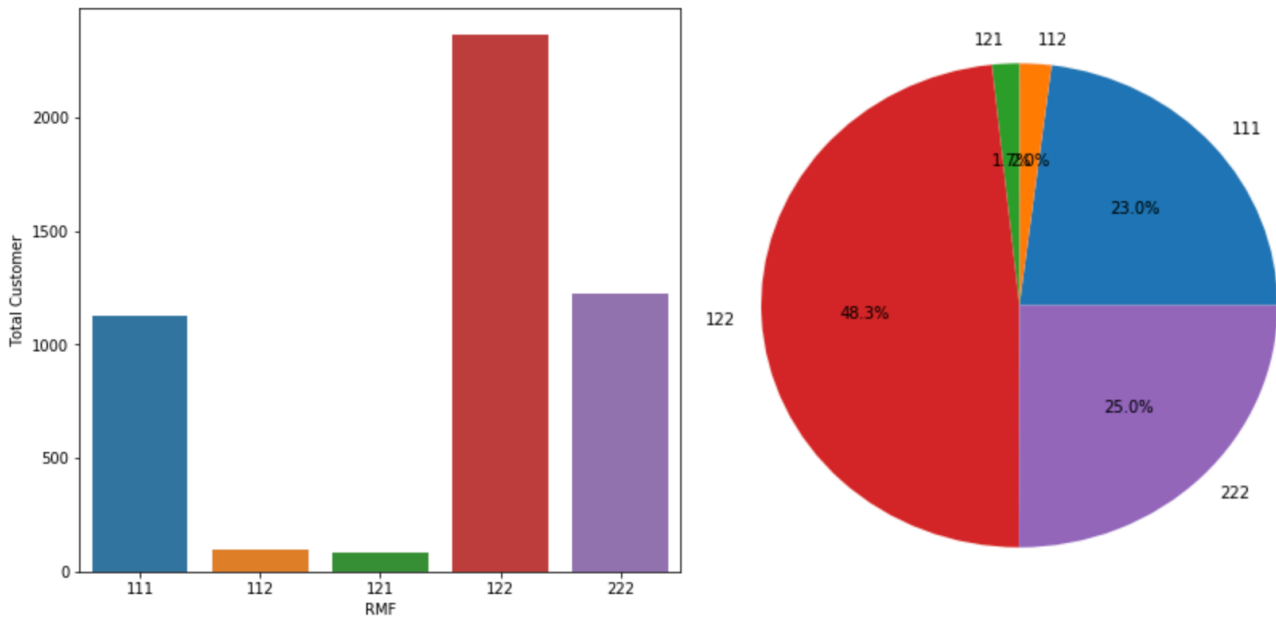
### **ความหมายของ RMF score**

จากข้อมูลลูกค้าทั้งหมดสามารถจัดกลุ่มลูกค้าตามRMF score ได้ 5 กลุ่ม ดังนี้

- 111 คือ ลูกค้าชั้นดีที่เพิ่งซื้อ มียอดมาก และ ซื้อหลายครั้ง
- 112 คือ ลูกค้าที่เพิ่งซื้อ มียอดมาก และ ซื้อน้อยครั้ง
- 121 คือ ลูกค้าที่เพิ่งซื้อ มียอดน้อย และ ซื้อหลายครั้ง
- 122 คือ ลูกค้าที่เพิ่งซื้อ มียอดน้อย และ ซื้อน้อยครั้ง
- 222 คือ ลูกค้าที่ชื้อนานแล้ว มียอดน้อย และ ซื้อน้อยครั้ง

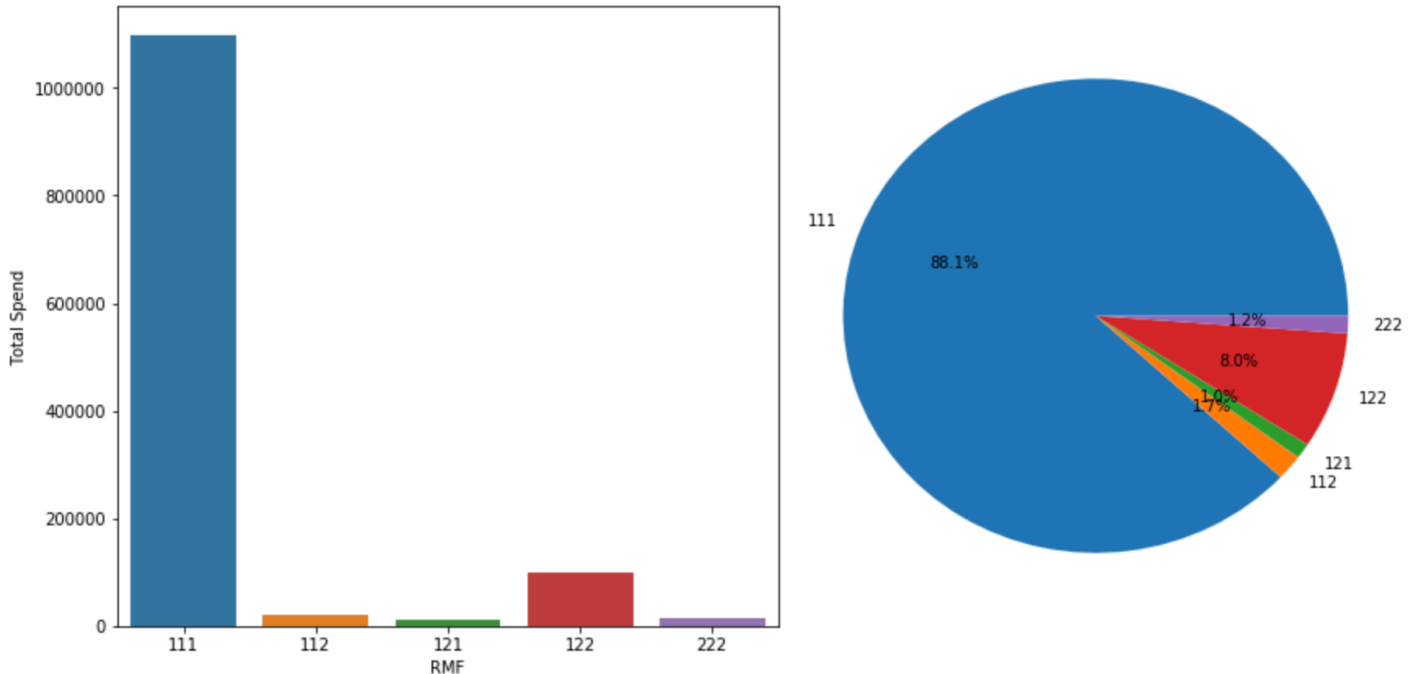
## วิเคราะห์ข้อมูลลูกค้าโดยการแบ่งกลุ่มด้วย RMF score

จำนวนลูกค้าในแต่ละกลุ่ม



จากกราฟจะพบว่าลูกค้าส่วนใหญ่เป็นลูกค้าประเภท 122 มีประมาณ 48.3% รองลงมาคือ 222 และ 111 ที่คิดเป็น 25.0%, 23.0%. ตามลำดับ

## สำรวจยอดขายจากลูกค้าในแต่ละกลุ่ม



แม้ว่าจำนวนลูกค้าในกลุ่ม 122 และ 222 จะมีจำนวนมากกว่าลูกค้ากลุ่ม 111 แต่ยอดขายส่วนใหญ่มากกว่า 88% มาจากลูกค้ากลุ่ม 111 ซึ่งเป็นลูกค้าชั้นดี สอดคล้องกับทฤษฎี 80/20\* ที่กล่าวว่า 80% ของยอดขายจะมาจากลูกค้าแค่ 20% ดังนั้นจึงตอบคำถาม “ใครคือลูกค้าหลักของธุรกิจ?” ได้ว่ากลุ่มลูกค้าหลักที่ผู้ประกอบการควรโฟกัสคือ กลุ่มลูกค้าชั้นดี (RMF=111) ที่ถือเป็นลูกค้าที่ทำให้เกิดยอดขายได้มากที่สุด

คำถามต่อไป...

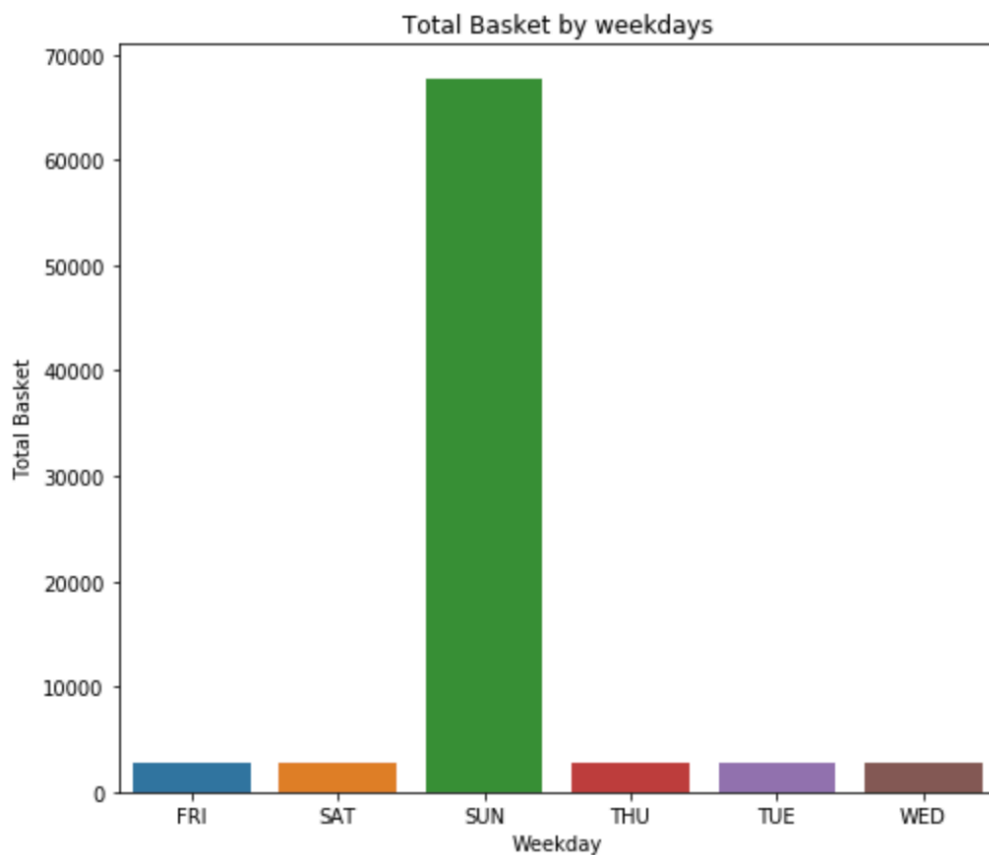
**“ลูกค้าชั้นดีเป็นอย่างไร?”**

\*read more about 80/20 rule: <https://marketinginsidergroup.com/strategy/marketing-80-20-rule-take-advantage/>

## First Tier Customers Analysis

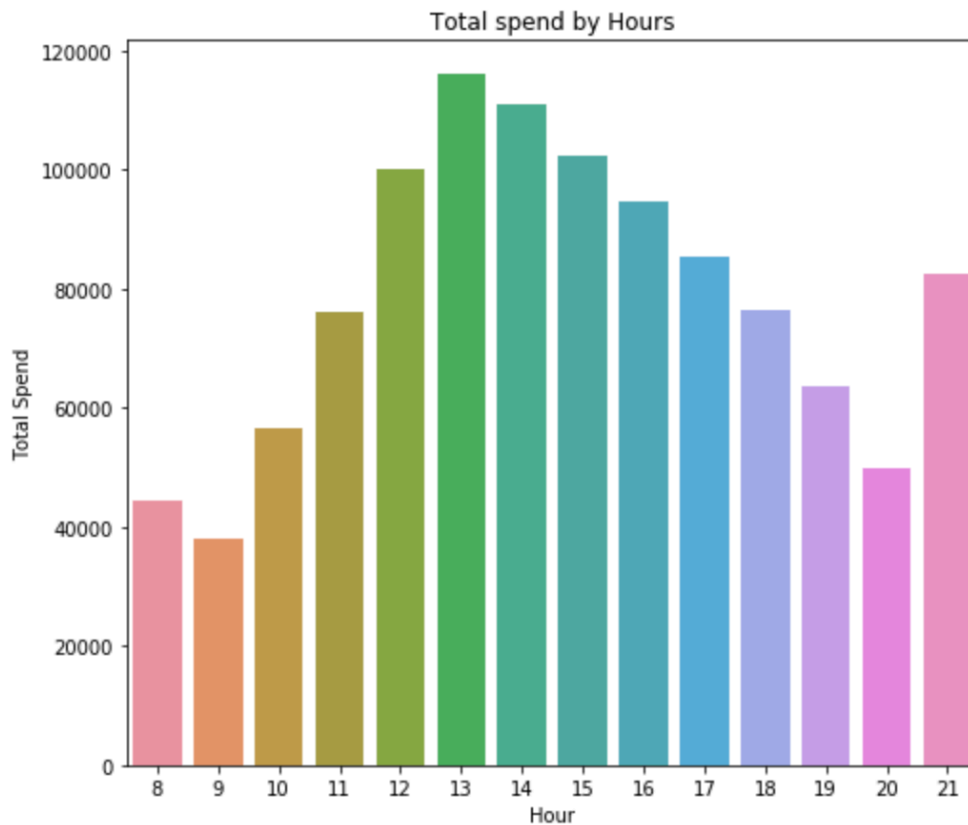
ต่อไปจะเป็นการวิเคราะห์ลูกค้ากลุ่ม 111 เพื่อหา insights ต่างๆที่เกี่ยวข้องกับลูกค้ากลุ่มนี้ โดยต่อไปนี้จะเรียกลูกค้ากลุ่มนี้ว่ากลุ่ม “ลูกค้าหลัก”

ลูกค้าหลักชอบไปซื้อของวันไหน?



จากกราฟจะพบว่าจำนวนการซื้อของกลุ่มลูกค้าหลักส่วนใหญ่เกิดขึ้นในวันอาทิตย์ คิดเป็น **82.9%**

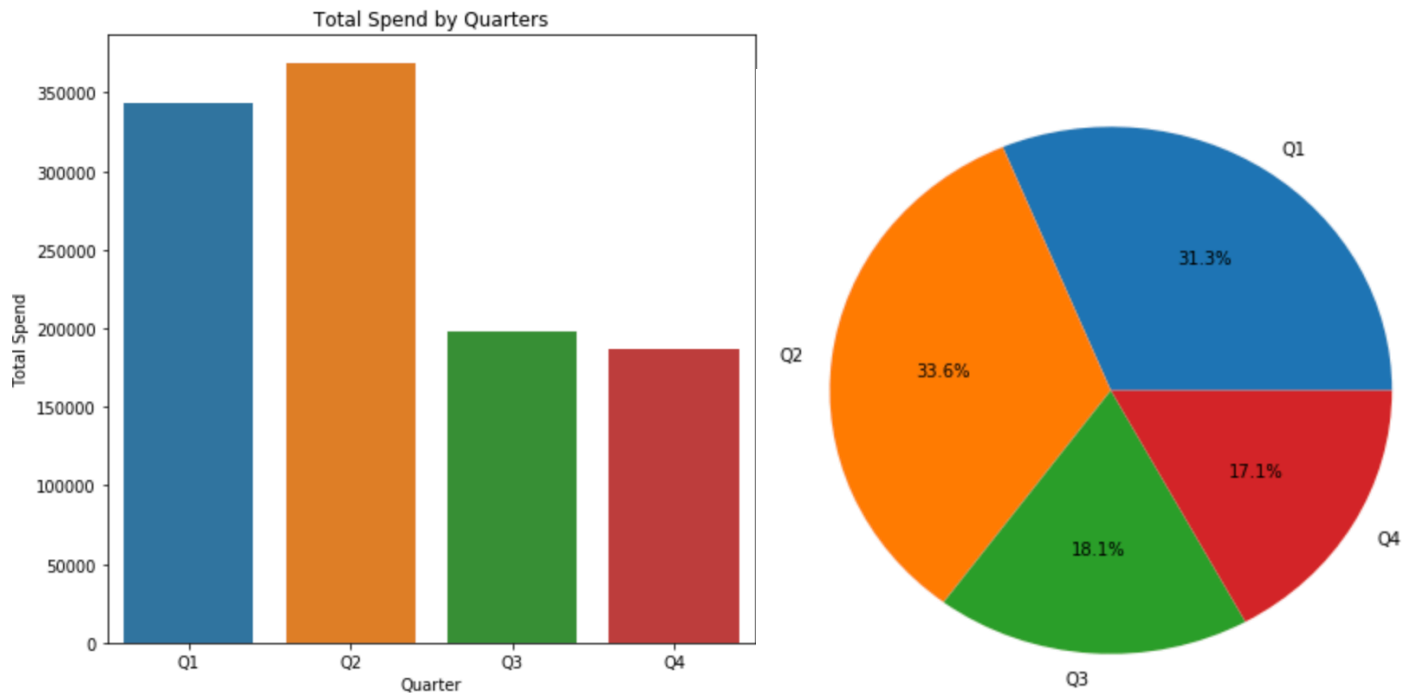
ลูกค้าหลักชอบไปซื้อของเวลาไหน?



จากกราฟพบว่ายอดจากลูกค้าหลักจะค่อยๆเพิ่มขึ้นในช่วงเช้า **8:00 - 13:00** และค่อยๆลดลงในช่วงบ่ายถึงค่ำเวลา **14:00 - 20:00** และจะมียอดพุ่งเพิ่มมาอีกครั้งเวลา **21:00 - 22:00**

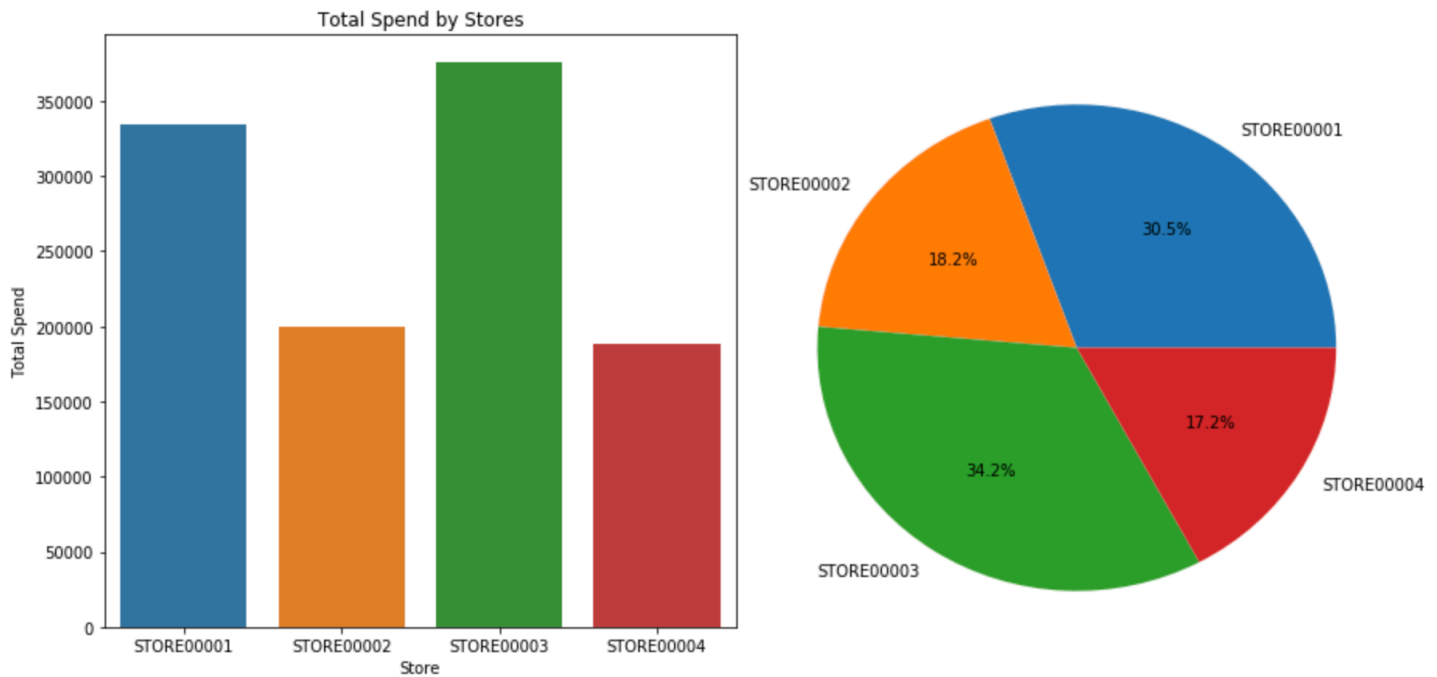


**ยอดจากลูกค้าหลักส่วนใหญ่มาจากไตรมาสไหน?**



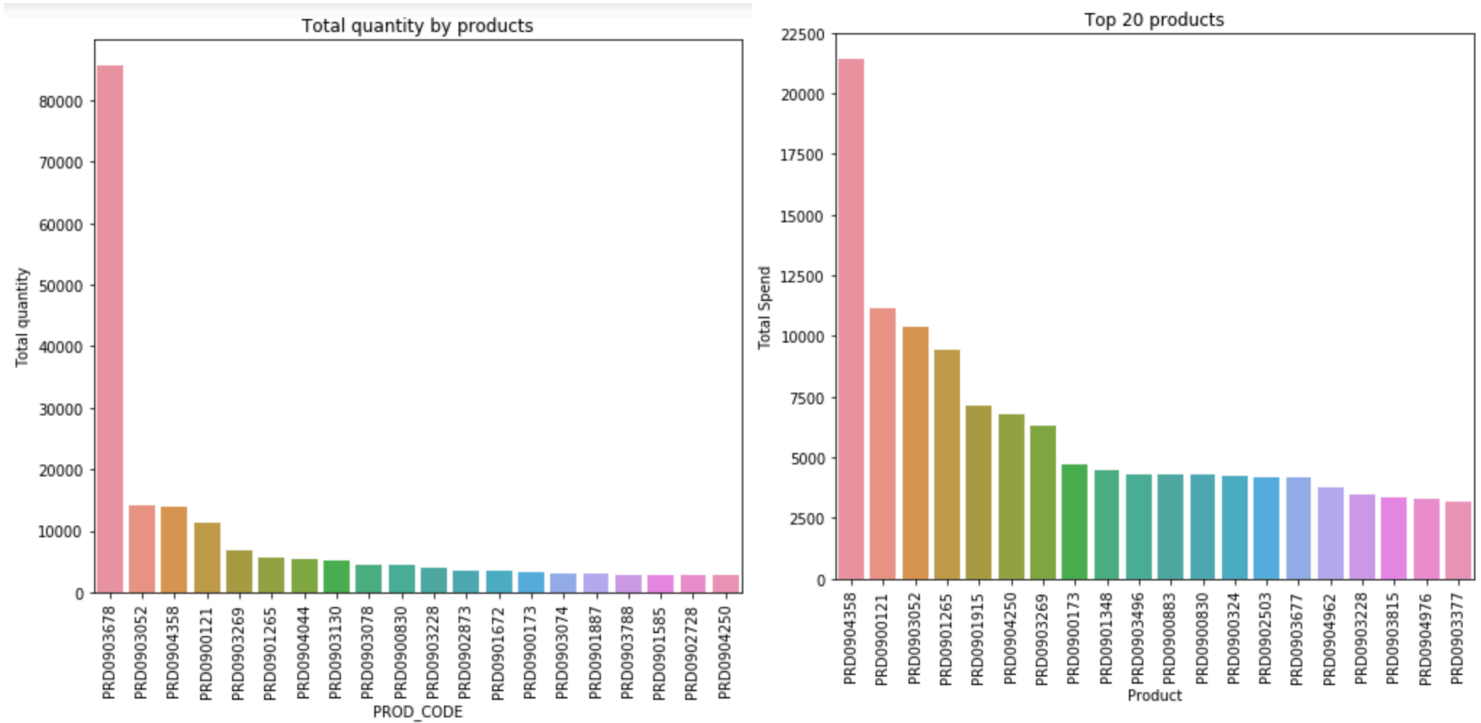
**ยอดจากลูกค้าหลักส่วนใหญ่มาจากครึ่งแรกของปีคิดเป็น 64.7% และ ยอดจากครึ่งหลังของปีคิดเป็น 35.1% จะสังเกตได้ว่ายอดในครึ่งหลังของปีหายไปเกือบครึ่ง**

ยอดจากลูกค้าหลักส่วนใหญ่มาจากสาขาไหน?



ยอดขายส่วนใหญ่มาจากร้าน **STORE00001** และ **STORE00003** คิดเป็น **64.7%** ส่วนที่เหลือประมาณ **35.3%** มาจากร้าน **STORE00002** และ **STORE00004**

## ลูกค้าหลักชอบซื้ออะไร?

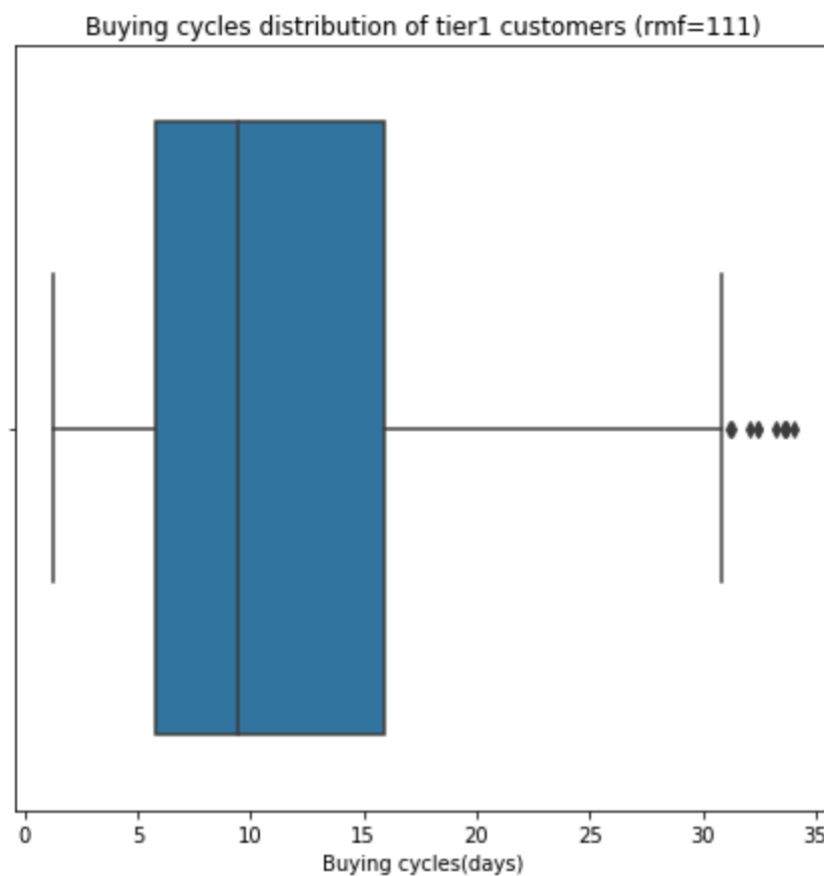


เห็นว่าสินค้าที่ขายดีที่สุด(ปริมาณเยอะที่สุด) คือ **PRD0903678** ซึ่งขายได้มากกว่าอันดับ 2 ถึง 6 เท่า แต่ถ้าเรียงตามยอดขาย **PRD0903678** จะอยู่ที่อันดับ **202**

ในขณะที่สินค้าที่ทำยอดขายได้มากที่สุดคือ **PRD0904358** ซึ่งเป็นสินค้าขายดีอันดับ 3 โดย **PRD0904358** ทำยอดขายได้มากกว่าอันดับที่ 2 เกือบ 2 เท่า

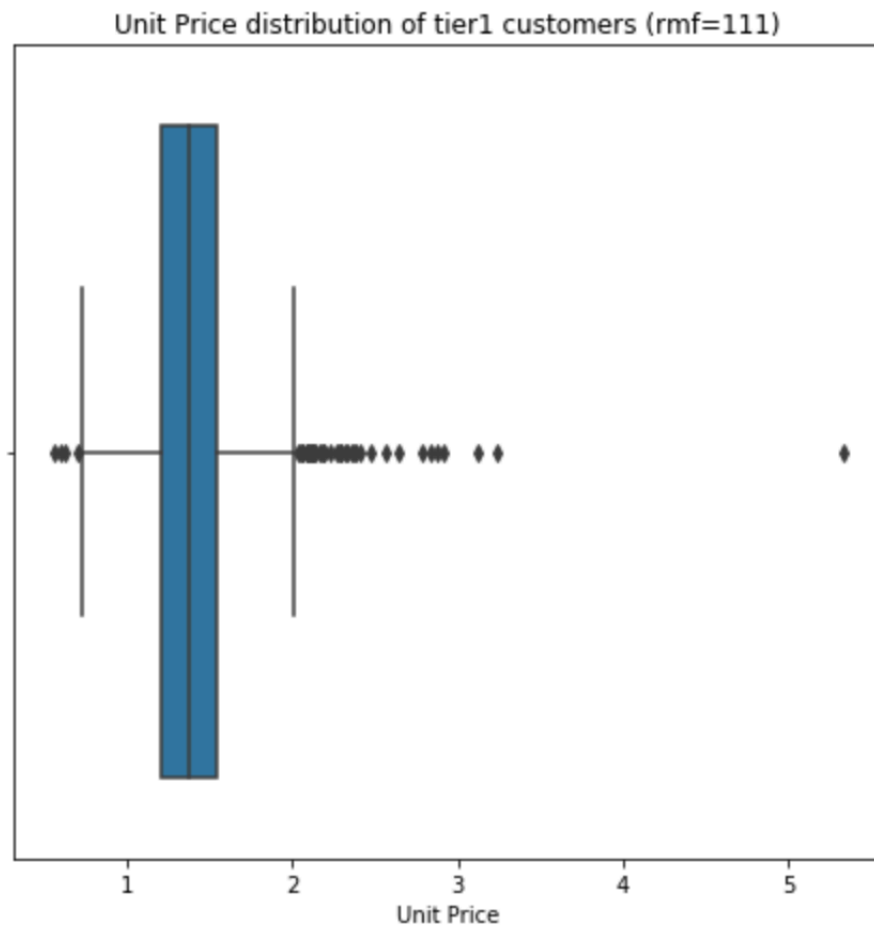
โดยสินค้าที่มียอดขายสูงสุด 20 ตัวแรก มียอดขายรวมกันเท่ากับ **124,437** คิดเป็น **11.35%** ของยอดขายทั้งหมด(ที่เกิดจากกลุ่มลูกค้าหลัก)

## ลูกค้าหลักซื้อบ่อยแค่ไหน?



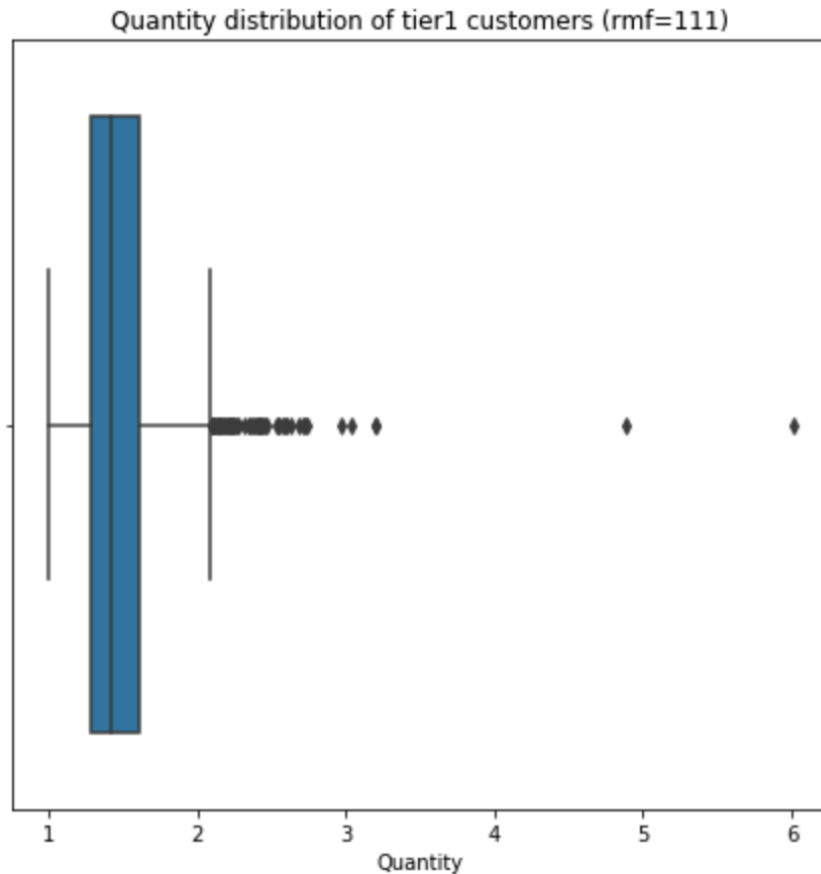
โดยเฉลี่ยลูกค้าหลักของเราจะซื้อของทุกๆ **11.4** วัน โดย **99%** ของลูกค้าหลักจะซื้อของทุกๆ น้อยกว่าหรือเท่ากับ **30.38** วัน (**99th percentile=30.38**)

ลูกค้าหลักชอบซื้อของถูกหรือแพง?



โดยเฉลี่ยแล้วลูกค้าหลักมักจะซื้อของที่มีราคาต่อชิ้นเฉลี่ยอยู่ที่ 1.4 โดย 99% ของลูกค้าหลักจะซื้อของที่มีราคาต่อชิ้นน้อยกว่าหรือเท่ากับ 2.38 (99th percentile=2.38)

ลูกค้าหลักชอบที่จะซื้อครั้งละมากๆหรือน้อยๆ ?

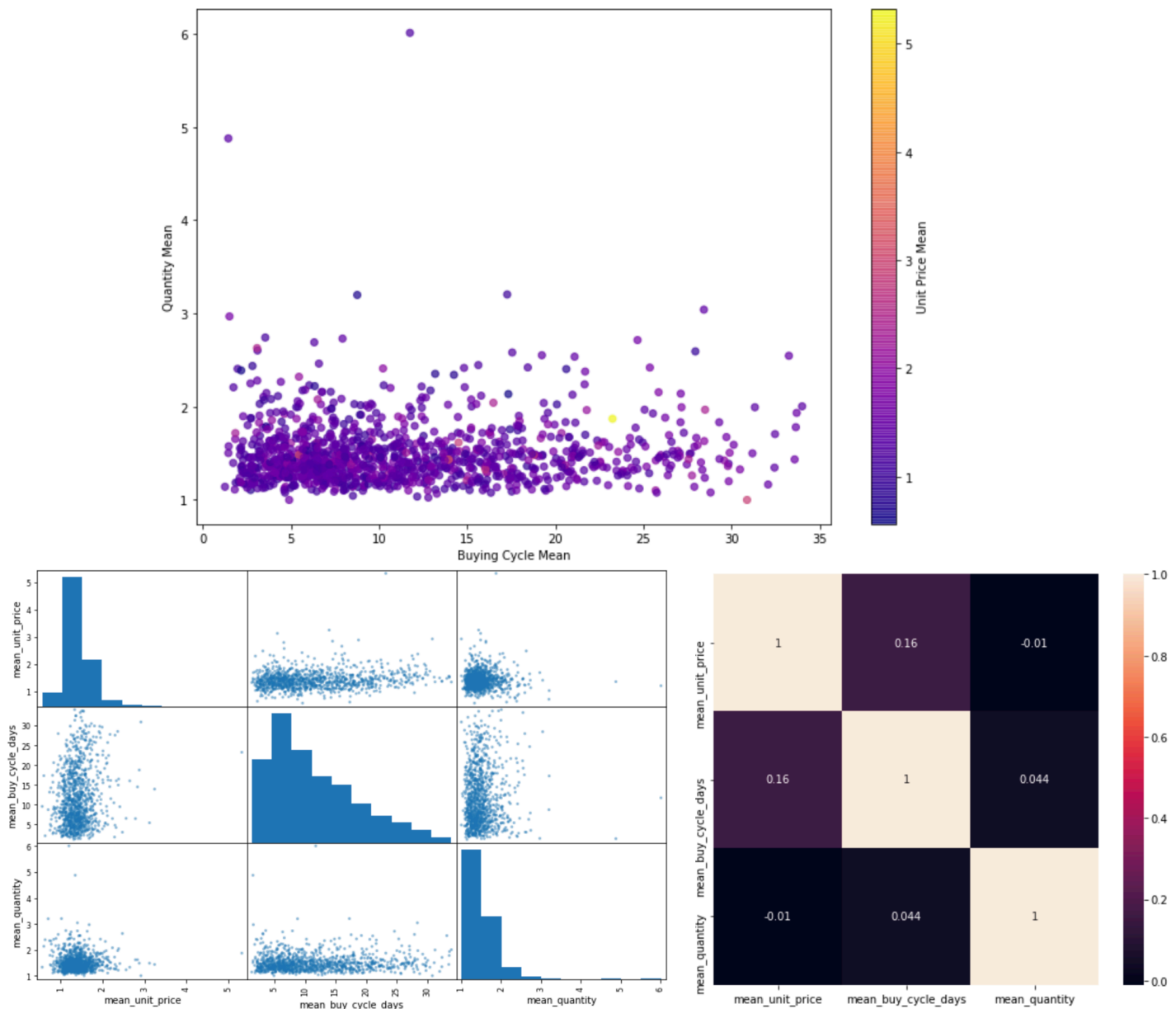


โดยเฉลี่ยแล้วลูกค้าหลักมักจะซื้อของครั้งละ 1.5 ชิ้น โดย 99% ของลูกค้าจะซื้อครั้งละน้อยกว่าหรือเท่ากับ 1.9 ชิ้น (99th percentile=1.9)

สรุป

ลูกค้าหลักส่วนใหญ่มักจะชอบซื้อสินค้าที่มีราคาถูก ซื้อครั้งละน้อยชิ้น แต่มาซื้อค่อนข้างบ่อย

## Buying Cycle vs. Quantity vs. Unit Price



จากการวิเคราะห์ค่า **Correlation** เพื่อหาความสัมพันธ์ระหว่างค่า ความถี่ในการซื้อ ราคาต่อชิ้นเฉลี่ย และ ปริมาณสินค้าที่ซื้อ พบว่าตัวแปรเหล่านี้มีค่า **correlation coefficient** ระหว่างกันน้อยมาก สรุปได้ว่าตัวแปรเหล่านี้ไม่มีความสัมพันธ์เชิงเส้นต่อกัน กล่าวคือตัวแปรเหล่านี้ไม่ได้มีผลซึ่งกันและกันอย่างมีนัยสำคัญ

## Summary

จากการวิเคราะห์ข้อมูลการซื้อสินค้าของลูกค้าจำนวน 4,891 คน ที่มีซื้อสินค้าจาก Supermarket จำนวน 4 สาขา โดยข้อมูลถูกเก็บตั้งแต่ 1/1/2007 ถึง 6/7/2008 จากการทำ customer segmentation ด้วยเทคนิค RMF Analysis พบว่าเจ้าของธุรกิจควรโฟกัสไปยังกลุ่มลูกค้าหลัก(RMF = 111)เนื่องจากเป็นกลุ่มที่มีสร้างยอดขายได้มากที่สุด และจากการวิเคราะห์ข้อมูลลูกค้าเหล่านี้พบ Insights ที่น่าสนใจดังนี้

- ยอดรวมจากลูกค้าหลักมีประมาณ 88% ของยอดรวมจากการขายทั้งหมด
- ลูกค้าหลักชอบมาซื้อของวันอาทิตย์
- ยอดขายจากลูกค้าหลักจะค่อยๆเพิ่มขึ้นในตอนเช้าไปถึงเที่ยงและจะค่อยๆลดลงระหว่างช่วงเที่ยงถึงช่วงค่ำ ก่อนจะกลับมาเพิ่มขึ้นอย่างมีนัยสำคัญอีกครั้งในช่วง 21:00 - 22:00
- ครึ่งหนึ่งของยอดขายที่เกิดจากลูกค้าหลักเกิดขึ้นในครึ่งปีแรกและเกิดขึ้นที่ร้าน STORE00001 STORE00002
- สินค้าขายดีที่สุดคือ PRD0903678 แต่สินค้าที่มียอดสูงสุดคือ PRD090435
- ลูกค้าหลักส่วนใหญ่มักจะชอบซื้อสินค้าที่มีราคาถูก ซื้อครั้งละน้อยชิ้น แต่มาซื้อค่อนข้างบ่อย
- ความถี่ในการซื้อ ราคาต่อชิ้นเฉลี่ย และ ปริมาณสินค้า ไม่มีผลต่อกัน

โดย Code ทั้งหมดที่ใช้ในการวิเคราะห์ข้อมูลสามารถเข้าผ่านลิงค์ข้างล่างนี้

[https://github.com/joei1994/ML-Projects/blob/data-cafe/mini-projects/data-cafe-supermarket/customer\\_segmentation.ipynb](https://github.com/joei1994/ML-Projects/blob/data-cafe/mini-projects/data-cafe-supermarket/customer_segmentation.ipynb)