

# 01

# CLV Dashboard

Dataset : ชุดข้อมูล Supermarket ที่มี Transaction  
การซื้อของลูกค้า 956K Recode

Requirement : สร้าง CLV Dashboard จาก Dataset  
ด้วย PowerBI



# Simple Calculation Relation to CLV

Method 1 :

$$\text{ARPU} = \frac{\text{TR}}{\text{CQ}}$$

(Average Revenue Per User) (Total Revenue for a chosen period) (Number of Customer for a chosen period)

Method 2 :

$$\text{CLV} = \frac{\text{T} \times \text{AOV} \times \text{AGM} \times \text{ALT}}{\text{Number of customer for the period}}$$

(Customer Lifetime Value)

- **T** is the average number of transactions
- **AOV** is the average value of an order (Total revenue / Number of order)
- **AGM** the average gross margin
- **ALT** represents the average lifespan of a customer (1 / Churn rate %) ; Churn rate = (Customer at the beginning - Customer at the end) / Customer at the beginning

# Dashboard

## SUPERMARKET DASHBOARD

YEAR

2007

MEMBER

All

BASKET DOMINANT

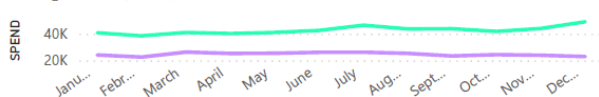
All

Total Sales

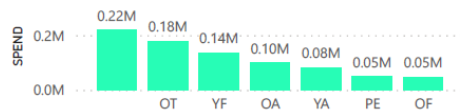
0.81M

Sales/Month

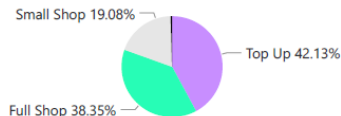
STORE\_REGION ● E02 ● W01



Sales by Customer Lifestage



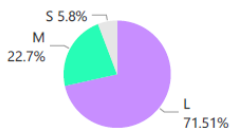
Sales by Basket Type



Total Transactions

435K

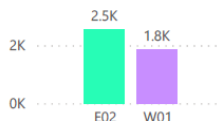
Transactions by Basket Type



Total Customer

4366

Customer/Store Region



Transactions by Day & Hour

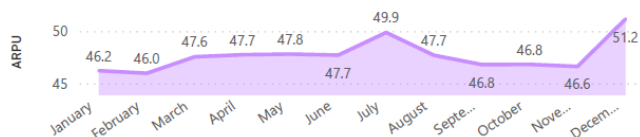
SHOP_WEEK DAY	8	9	10	11	12	13	14	15	16	17	18	19	20	21	Total
1	4.27%	3.54%	5.14%	6.52%	9.00%	9.97%	9.42%	9.25%	8.07%	8.04%	7.00%	5.92%	4.79%	9.07%	100.00%
2	4.14%	3.20%	4.60%	7.04%	8.39%	8.90%	9.19%	8.87%	8.68%	7.69%	7.49%	6.35%	5.13%	10.31%	100.00%
3	3.91%	3.83%	4.74%	6.58%	8.22%	9.98%	9.59%	10.16%	8.49%	7.67%	6.97%	5.50%	4.40%	9.98%	100.00%
4	3.58%	3.37%	4.70%	6.97%	8.65%	9.70%	9.63%	9.64%	8.28%	7.64%	7.04%	6.09%	4.67%	10.05%	100.00%
5	3.67%	3.12%	4.36%	6.41%	8.44%	9.99%	9.78%	8.57%	8.14%	8.85%	7.28%	6.25%	5.24%	9.90%	100.00%
6	3.45%	3.42%	4.85%	6.14%	8.27%	9.63%	10.02%	9.33%	8.51%	8.85%	6.62%	6.34%	4.92%	9.65%	100.00%
7	4.03%	2.92%	4.91%	6.66%	8.31%	9.85%	10.97%	8.98%	9.04%	7.31%	6.59%	6.06%	4.90%	9.48%	100.00%
Total	3.86%	3.34%	4.76%	6.62%	8.47%	9.71%	9.80%	9.25%	8.46%	8.01%	7.00%	6.07%	4.87%	9.78%	100.00%

## CUSTOMER LIFETIME VALUE

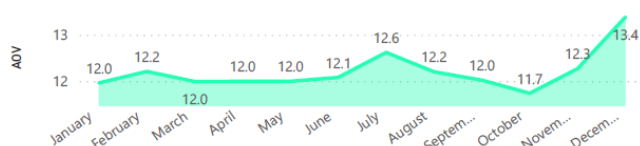
AVG. GROSS MARGIN

50.00%

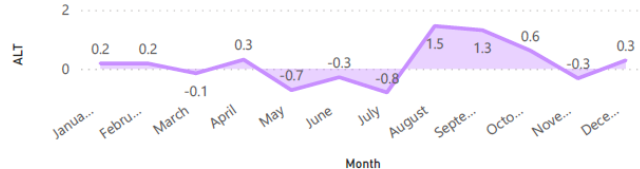
ARPU/Month



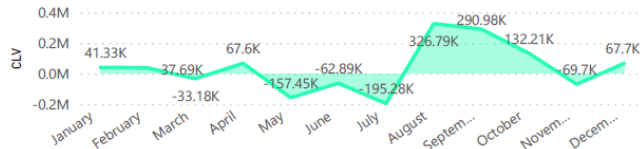
AOV/Month



ALT/Month



CLV/Month





# 02

## Customer Segmentation

Dataset : ชุดข้อมูล Supermarket ที่มี Transaction  
การซื้อของลูกค้า 956K Record

Requirement : Customer Segmentation Using  
K-Means Clustering

# Define features



## Spend and Quantity

- Total Spend
- Total Visits
- Ticket Size
- Total



## Date and Time

- First Date
- Max Date
- Total Days
- Recency
- Day Shopping
- Night Shopping

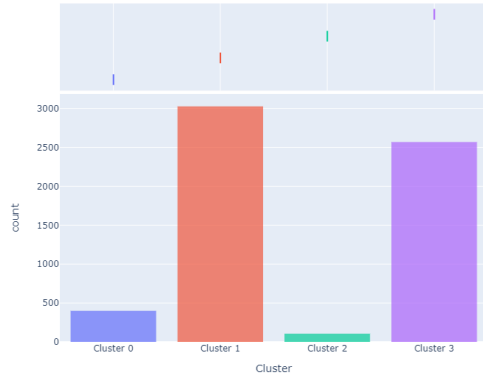


## Basket Dominant Mission

- No. of Fresh
- No. of Grocery
- No. of Mixed
- No. of Nonfood
- No. of XX

# Result From K-mean Clustering

## Number of customer in each clusters



## Average of each feature in each clusters

	Count	TotalSpend	TotalVisits	TotalQuantity	DayShopping	NightShopping	N_Fresh	N_Grocery	N_Mixed	N_Nonfood	N_XX	TicketSize	TotalDays	Recency
Cluster														
Cluster 0	400	1577.091100	121.017500	1273.515000	713.535000	143.472500	431.967500	79.870000	334.020000	9.647500	1.502500	17.499000	778.167500	7.797500
Cluster 1	3026	14.539400	1.711200	11.168200	6.153000	1.572000	3.458700	1.093900	2.920000	0.226700	0.025800	8.252600	43.898500	438.383300
Cluster 2	106	4053.438700	259.594300	3401.971700	1707.490600	443.367900	1068.698100	177.056600	885.886800	17.094300	2.122600	19.550800	808.226400	3.773600
Cluster 3	2568	154.336600	15.995300	119.056100	65.233300	16.044800	38.139000	10.416700	30.648000	1.850500	0.223900	10.750900	581.418600	59.447000

# Interpretation

Cluster	Name	Character	Action
0	ตัวสำรอง	มีจำนวนน้อยเป็นอันดับ 2, ซื้อม่อย ซ้อยเยอะ จ่ายเยอะเป็นอันดับ 2 และมีราคาเฉลี่ยต่อการ ซื้อ1ครั้งสูงเป็นอันดับ 2	รักษาความสัมพันธ์ที่ดีเอาไว้รวมถึงทำความเข้าใจลูกค้าแต่ละ บุคคลมากขึ้น, นำเสนอโปรโมชั่น, สิทธิพิเศษเพิ่มเติมหากมียอด การซื้อเพิ่มขึ้น พยายามทำให้ลูกค้ากลุ่มนี้กลายเป็น 'ตัวจริง'
1	คนที่เคยเจอ	มีจำนวนเยอะที่สุด, โดยเฉลี่ยเข้ามาซื้อสินค้า 1-2 ครั้ง และมีราคาเฉลี่ยต่อการซื้อ1 ครั้งต่ำ ที่สุด	ทำอะไรมากไม่ได้ แต่เนื่องจากเป็นลูกค้าที่มีจำนวนมากที่สุด หากสามารถดึงลูกค้ากลุ่มนี้มาได้จะเป็นการเพิ่มยอดขายได้เป็น อย่างมาก อาจจะเสนอขายสินค้าลดราคา-ราคาพิเศษ
2	ตัวจริง	มีจำนวนน้อยที่สุด, ซื้อม่อย-ซ้อยเยอะ-จ่ายเยอะ ที่สุด และมีราคาเฉลี่ยต่อการซื้อ1ครั้งสูงที่สุด	รักษาความสัมพันธ์ที่ดีเอาไว้, ทำให้ลูกค้ากลุ่มนี้รู้สึกเป็นคน พิเศษแตกต่างจากคนอื่น ๆ เช่น ส่งของขวัญปีใหม่, ที่จอตรก พิเศษ
3	คนรู้จัก	มีจำนวนเยอะเป็นอันดับ 2, โดยเฉลี่ยเข้ามาซื้อ สินค้า 16 ครั้ง และมีราคาเฉลี่ยต่อการซื้อ1ครั้ง ต่ำเป็นอันดับ 2	เช่นเดียวกับลูกค้ากลุ่ม 'คนที่เคยเจอ' เนื่องจากเป็นกลุ่มที่มี ปริมาณมากพอสมควร ดังนั้นควรพยายามทำให้ลูกค้ากลับมา ซื้อสินค้าบ่อยขึ้น เช่นการส่งโปรโมชั่น เสนอสินค้าราคาพิเศษ ทำการ Upsell-Cross sell

เพิ่มเติม : ส่วนใหญ่ลูกค้าจะเข้ามาซื้อสินค้าในช่วง 8.00-18.00 น. , เข้ามาซื้อในแผนก Fresh และ Mixed มากที่สุด

# 03

## Product Recommendation

Dataset : แบบสอบถามจากนักศึกษาในชั้นเรียนว่าเคยซื้อสินค้าอะไรบ้างหรือไม่ โดยมีสินค้าให้เลือกทั้งหมด 27 รายการและได้รับคำตอบจากนักศึกษา 29 คำตอบ

Requirement : Product Recommendation โดยใช้ Market Basket Analysis และ Collaborative Filtering





# Market Basket Analysis

**Market Basket Analysis** : การวิเคราะห์ตะกร้าตลาด เป็นการวิเคราะห์เพื่อค้นหา กลุ่มสิ่งของที่น่าจะปรากฏร่วมกันในตะกร้าซื้อขาย โดยผลลัพธ์ที่ได้จะแสดงในรูปแบบของ กฎที่บ่งบอกถึงความเป็นไปได้ในการซื้อสินค้าต่าง ๆ ร่วมกัน

**Association rule** : กฎของความสัมพันธ์ ระหว่างกลุ่มของสินค้า 2 กลุ่ม

ตัววัดประสิทธิภาพที่ใช้ได้แก่

- **Support** : ตัวเลขที่แสดงสัดส่วนของจำนวนตะกร้าการซื้อขายที่มี **itemset** ต่อจำนวนตะกร้าการซื้อขายทั้งหมด
- **Confidence** : ความน่าจะเป็นที่สินค้า '**consequents**' จะถูกหยิบเข้าตะกร้าด้วยหลังจากที่สินค้า '**antecedents**' ถูกหยิบเข้าตะกร้าไป (**Confidence** ที่สูงจะแสดงความมั่นใจว่าในตะกร้ามีสินค้า '**antecedents**' อยู่แล้ว สินค้า '**consequents**' ก็จะถูกหยิบไปด้วย)
- **Lift** : เปรียบเทียบความน่าจะเป็นที่จะพบสินค้า '**consequents**' ในตะกร้าที่มีการซื้อสินค้า '**antecedents**' กับ ความน่าจะเป็นที่จะพบ สินค้า '**consequents**' ในตะกร้า (ยิ่งค่า **Lift** มีค่ามากเท่าไร ความน่าจะเป็นที่ลูกค้าจะหยิบสินค้า '**consequents**' ติดไปในตะกร้าที่มี '**antecedents**' อยู่แล้วยิ่งมีค่ามากขึ้น)

ref. : <https://bigdata.go.th/big-data-101/data-science/what-is-association-rule/>  
ref. : [http://rasbt.github.io/mlxtend/user\\_guide/frequent\\_patterns/association\\_rules/](http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/)

# Collaborative Filtering

**Collaborative Filtering** : อัลกอริทึมที่ใช้สร้าง **Patterns** หรือ **Rules** ในการแนะนำสินค้า ซึ่งในที่นี้จะใช้แบบ **Item-Based Filtering**

**Item-Based Filtering** : การแนะนำโดยหาสินค้าที่ถูกซื้อด้วยลูกค้ากลุ่มเดียวกัน

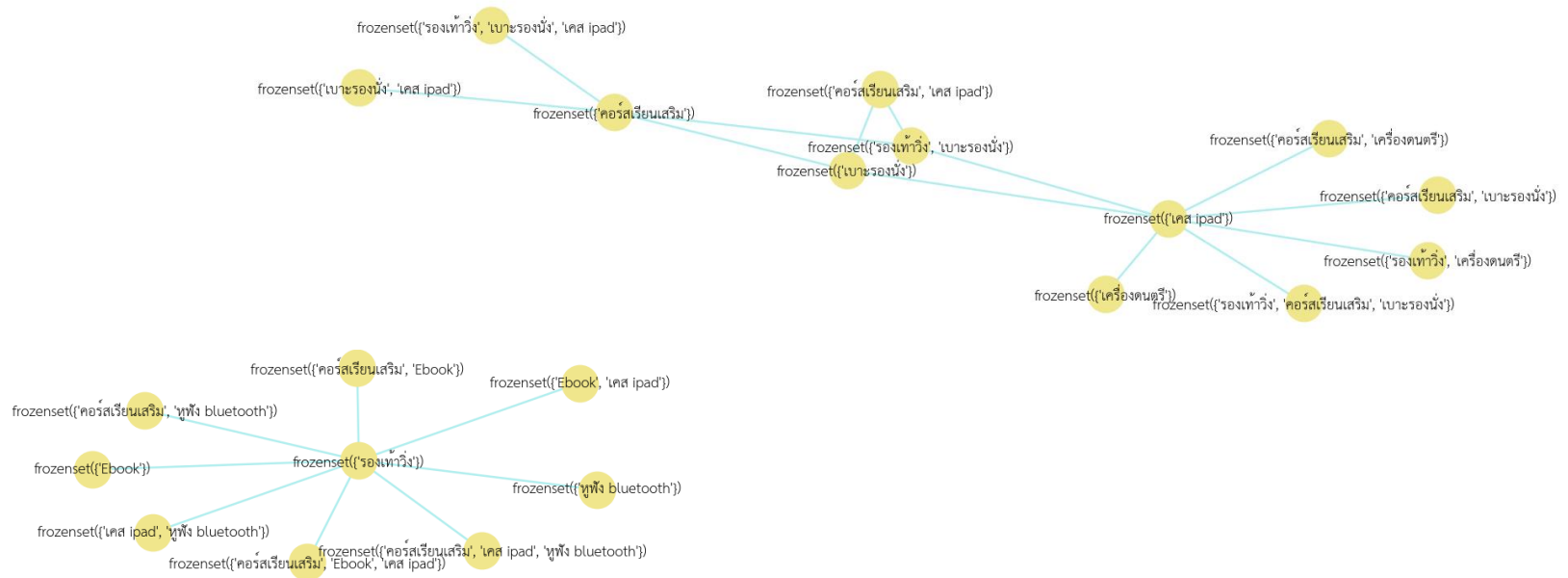
**Cosine Similarity** : ใช้ในการหาความเหมือนกันของสินค้า (**Cosine Similarity** ที่สูงจะแสดงว่าสินค้านั้นมีความเหมือนกันสูง ตัวอย่างการใช้งานเช่น เมื่อลูกค้าซื้อสินค้า **A** เราจะแนะนำสินค้า **B** ที่มีค่า **Cosine Similarity** กับสินค้า **A** สูง)

$$\text{Cosine Similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

ref. : <https://thanachart-rit.medium.com/%E0%B8%AA%E0%B8%A3%E0%B9%89%E0%B8%B2%E0%B8%87-recommendation-engine-collaborative-filtering-%E0%B9%82%E0%B8%94%E0%B8%A2-python-e81fbb774fc1>

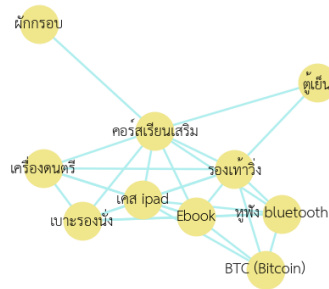
# Network Graph : Market Basket Analysis

**Association rule : Filter** คู่สินค้าที่มี **Lift**  $\geq 1$  และ **Confidence**  $\geq 0.5$  และเลือกนำผลมา **Plot Graph** เฉพาะคู่สินค้าที่มี **Confidence = 1**



# Network Graph : Collaborative Filtering

Filter rule : Filter คู่สินค้าที่มี  $\text{Cosine Similarity} \geq 0.6$

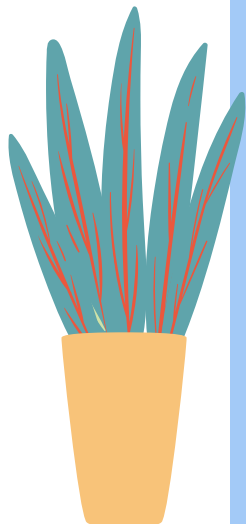


# 04

## Campaign Response Model

Dataset : ชุดข้อมูล Retail\_Data\_Response และ Retail\_Data\_Transactions

Requirement : Improve classification model โดยใช้ค่า AUC ในการเปรียบเทียบ



# Process

1. Load Data

2. Clean Data

3. Feature Engineer

- create data set with RFM variables
- create data set with CLV variables

4. Split Train Test Data

5. Fixing Imbalanced Data (Train Set)

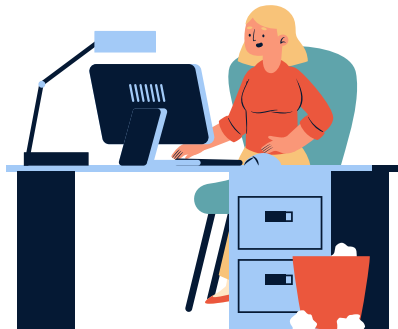
- Under Sampling
- Over Sampling
- SMOTE

6. Model Training

- Logistic Regression
- XGBoost

7. Measure Model Performance

- AUC Train Set
- AUC Test Set
- AUC Different



# Model Performance

Classification Algorithm	Resaming	Data set	AUC Train Set	AUC Test Set	AUC Different
Logistic Regression	Under Sampling	RFM	0.7300	0.7202	0.0099
Logistic Regression	Under Sampling	CLV	0.7314	0.7301	0.0013
Logistic Regression	Over Sampling	RFM	0.7181	0.7256	0.0075
Logistic Regression	Over Sampling	CLV	0.7174	0.7333	0.0159
Logistic Regression	SMOTE	RFM	0.7256	0.7079	0.1771
Logistic Regression	SMOTE	CLV	0.7402	0.7010	0.0391
XGBoot	Under Sampling	RFM	0.8026	0.7303	0.0723
XGBoot	Under Sampling	CLV	0.8067	0.7373	0.0694
XGBoot	Over Sampling	RFM	0.7633	0.7388	0.0244
XGBoot	Over Sampling	CLV	0.7754	0.7457	0.0297
XGBoot	SMOTE	RFM	0.7754	0.7304	0.0451
XGBoot	SMOTE	CLV	0.7754	0.7147	0.0607

**Conclusion** : เลือก Model XGBoot ที่ใช้ CLV data set ที่ผ่านการ Fixing imbalanced data ด้วยเทคนิค Over sampling เนื่องจากโมเดลให้ AUC Different ไม่เกิน 0.03 อีกทั้ง AUC train และ AUC test มีค่าสูงที่สุด



# 05

## A/B Testing

Dataset : แบบสอบถามเกี่ยวกับการเลือกจองห้องพักจากป้ายโฆษณา  
โดยได้รับคำตอบจากผู้ทำแบบสอบถามทั้งหมด 32 คน

Requirement : สร้างแบบสอบถาม เก็บข้อมูล จากนั้นนำข้อมูลวิเคราะห์

# Banner A



จ่ายก่อน  
นอนทีหลัง

พิเศษเพียง  
**999.-**  
1,998.-

สามารถเข้าพักได้ตั้งแต่ วันที่ - 30 พ.ย. 64 (เลื่อนวันเข้าพักได้)  
\*Maximum 2 Nights Stay (ราคาสำหรับผู้ใหญ่ 2 ท่าน)



One Bedroom Suite  
Pool Access

ห้องพักเดี่ยว 1 ห้องนอน 1 ห้องรับแขก



A La Carte Breakfast  
for 2 person

อาหารเช้าตามสั่ง 2 ท่าน  
(ราคาอาหารเช้า ท่านละ 250.-)



Complimentary  
2 Drinks with Pool/room

ฟรี! Soft Drink หรือ Mocktail  
จำนวน 2 ท่าน / ห้อง

จองและชำระเงินได้ตั้งแต่วันที่ - 15 พ.ย. 64



# Banner B



จ่ายก่อน  
นอนทีหลัง

ลดสูงสุด  
**50%**  
เหลือเพียง  
**999.-**

สามารถเข้าพักได้ตั้งแต่ วันที่ - 30 พ.ย. 64 (เลื่อนวันเข้าพักได้)  
\*Maximum 2 Nights Stay (ราคาสำหรับผู้ใหญ่ 2 ท่าน)



One Bedroom Suite  
Pool Access

ห้องพักเดี่ยว 1 ห้องนอน 1 ห้องรับแขก



A La Carte Breakfast  
for 2 person

อาหารเช้าตามสั่ง 2 ท่าน  
(ราคาอาหารเช้า ท่านละ 250.-)



Complimentary  
2 Drinks with Pool/room

ฟรี! Soft Drink หรือ Mocktail  
จำนวน 2 ท่าน / ห้อง

จองและชำระเงินได้ตั้งแต่วันที่ - 15 พ.ย. 64

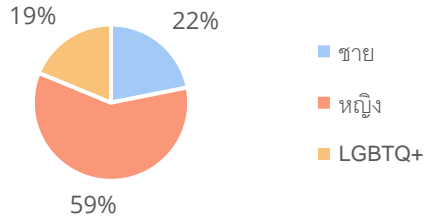




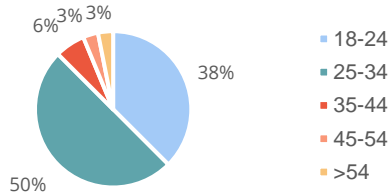
# Survey Results

## Descriptive

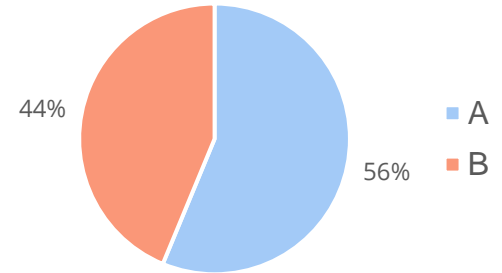
เพศ



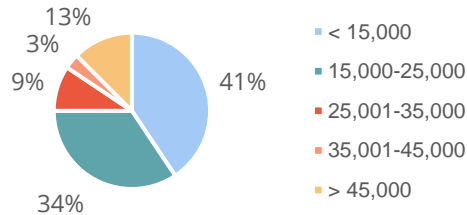
อายุ



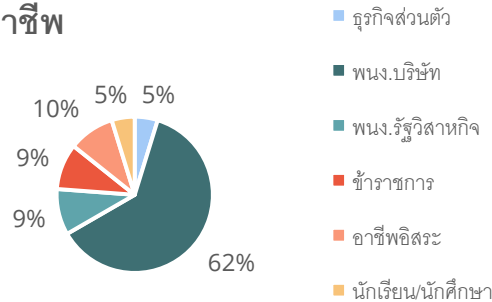
ป้ายโฆษณา



รายได้



อาชีพ



**Conclusion :** ป้ายโฆษณาแบบ A มีผลต่อการเลือกจองห้องพักรมมากที่สุด

# Survey Results

## Chi-Square Test

Variable	Banner A	Banner B	Chi-Square	P-Value
เพศ			0.91	0.823
ชาย	4	3		
หญิง	10	9		
LGBTQ+	3	2		
อายุ			2.878	0.578
18-24	7	5		
25-34	10	6		
35-44	1	1		
45-54	0	1		
> 54	0	1		
รายได้			2.381	0.666
< 15,000	8	5		
15,000-25,000	7	4		
25,001-35,000	1	2		
35,001-45,000	0	1		
> 45,000	2	2		
อาชีพ			3.582	0.611
ธุรกิจส่วนตัว	1	0		
พณง.บริษัท	8	5		
พณง.รัฐวิสาหกิจ	0	2		
ข้าราชการ	1	1		
อาชีพอิสระ	1	1		
นักเรียน/นักศึกษา	7	5		

**Conclusion :** เพศ อายุ รายได้ และอาชีพ ไม่มีผลต่อการเลือกของห้องพักรักจากป้ายโฆษณา ที่ระดับนัยสำคัญ 0.05

# 06

## Customer Movement Analysis

Dataset : ชุดข้อมูล Supermarket ที่มี Transaction การซื้อของลูกค้า 956K Recode

Requirement : Query ข้อมูลด้วย BigQuery จากนั้นนำไปสร้าง Dashboard ด้วย Data Studio



# Customer Status

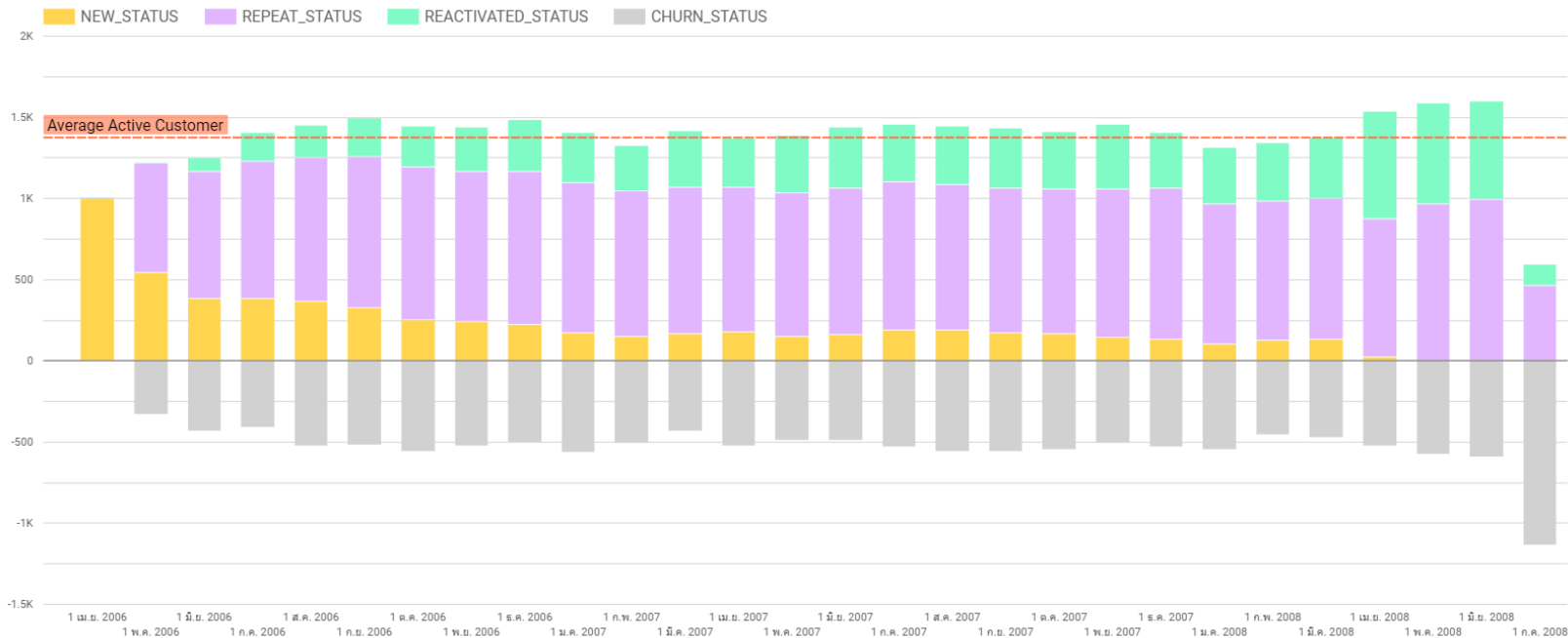


พิจารณาสถานะของลูกค้าโดยใช้เกณฑ์ดังนี้

	Current	Last month	Before Last month
New Customer	✓	✗	✗
Repeat Customer	✓	✓	(Ignore)
Chun Customer	✗	✓	(Ignore)
Reactivated Customer	✓	✗	✓

# Dashboard

# Customer Movement Analysis



# 07

## Voice of Customer Analysis

Dataset : ชุดข้อมูล Wongnai Reviews

Requirement : Voice of Customer Analysis Using K-Means Clustering and Agglomerative clustering



# Process

1. install and import dependencies
2. Document embedding and dimension reduction
3. Document clustering using **K-Means** Clustering
  - define a function to tokenize a sentence into words : “removed\_words”, “new\_words”
4. Document clustering using **Agglomerative** Clustering with cosine similarity
5. Result discussion



**Note:** K-Means Clustering VS Agglomerative Clustering

- **K-Means** : เป็นการแบ่งกลุ่มของข้อมูลเกิดจากจุดข้อมูลที่กระจุกตัวใกล้กับจุดกึ่งกลางของกลุ่มมากที่สุด (Centroid-based Clustering)
- **Agglomerative** : การแบ่งกลุ่มประเภทนี้จะสร้างให้เกิดต้นไม้ (แบบล่างขึ้นบน) ของกลุ่มข้อมูลขึ้น (Hierarchical Clustering)

ref. : <https://bigdata.go.th/big-data-101/4-types-of-clustering/>

# Result Discussion

ร้านกาแฟ, คาเฟ่



รีวิวเป็นไปในทางที่ดี มีหัวข้อที่รีวิว  
เกี่ยวกับรสชาติ, ราคา, เมนู

รูปภาพจาก : <https://www.pngegg.com/th/>

ร้านอาหาร, ร้านอาหารไทย



รีวิวเป็นไปในทางที่ดี มีหัวข้อที่รีวิวเกี่ยวกับ  
รสชาติ, ส้มตำ, น้ำแตงโมปั่น

รูปภาพจาก : <https://baansomtum.com/menu/>

ร้านชานมไข่มุก



รีวิวเป็นไปในทางที่ไม่ดี (Keywords : "แย่มาก")  
มีหัวข้อที่รีวิวเกี่ยวกับที่ตั้งของร้าน

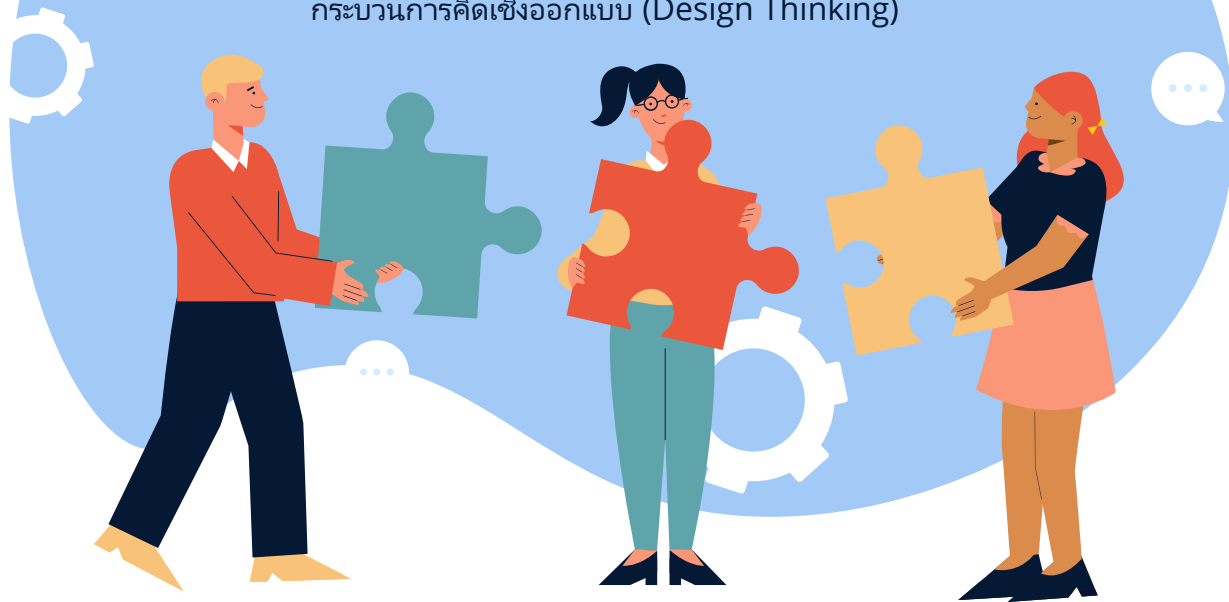
รูปภาพจาก : <https://today.line.me/th/v2/>



# 08

## Design Thianking

Requirement : สัมภาษณ์เพื่อนเรื่องปัญหาในการเรียนออนไลน์ จากนั้นนำไปเข้ากระบวนการคิดเชิงออกแบบ (Design Thinking)



# Profile



**Name : Ploy**

- เพศหญิง
- อายุ 24 ปี
- จบปริญญาตรีจากม.เกษตรศาสตร์
- อยู่คนเดียวที่หอพัก
- มีรถยนต์ส่วนตัว

## Empathize : Lifestyle

- ชอบฟังเพลง
- ติดตามศิลปิน K-pop
- ชอบถ่ายรูป
- เวลารว่าง วันหยุดมักจะไปคาเฟ่
- ชอบอ่านนิยาย
- สนใจศิลปะ
- ดื่มชา
- มักจะพบปะสังสรรค์กับเพื่อนๆ

## Define : ปัญหาในการเรียนออนไลน์

- งานเยอะขึ้น
- จัดสรรเวลาได้แย่ง
- รู้สึกเหงากว่าปกติเพราะไม่เจอเพื่อน
- ไม่โฟกัสกับการเรียนเท่าที่ควร

# Ideate

- **Mobile App Planning** คอยแจ้งเตือนรายการที่ต้องทำ ที่เชื่อมต่อกับระบบการเรียน การส่งงานออนไลน์ของทางมหาวิทยาลัย ซึ่งถ้าหากสามารถทำตาม Plan, ส่งงานได้สำเร็จ จะได้รับคูปองส่วนลดในการสั่งอาหารเครื่องดื่ม Delivery, ส่วนลดในการชื้อนียายออนไลน์ หรือส่วนลดการใช้แอปพลิเคชันฟังเพลง เช่น Spotify
- ระบบเรียนออนไลน์แบบ **Virtual** ที่ผู้เรียนสามารถสนุกกับการสร้าง Avatar ของตัวเอง สามารถเรียนแบบได้พบเจอเพื่อนๆ และอาจารย์ผ่าน Avatar ของแต่ละคน ซึ่งจะช่วยให้ผู้เรียนรู้สึกมีส่วนร่วม, กระตือรือร้น และมีสมาธิในการเรียนมากยิ่งขึ้น

# Prototype



- ระบบเรียนออนไลน์แบบ Virtual ที่มีแควส/ภารกิจในการเรียนให้ผู้เรียนได้ทำ เมื่อทำสำเร็จผู้เรียนจะสามารถได้รับ Item พิเศษในการแต่งตัว Avatar ของตนเอง หรือได้รับคูปองส่วนลดในการสั่งอาหารเครื่องดื่ม Delivery

# Test

## Like

- น่าจะได้สนุกกับการแต่งตัว Avatar
- ช่วยลดความเหงาจากการไม่ได้เจอเพื่อนๆ ได้
- ชอบที่ได้รับส่วนค่าอาหาร

## Dislike

- อาจจะได้รับความคิด ความกดดันเพิ่มจากการทำแควส
- อาจจะทำให้ต้องเสียค่าใช้จ่ายในการเรียนมากขึ้น

## Question

- เทคโนโลยีตอนนี้พร้อมรองรับหรือไม่
- อุปกรณ์ที่ใช้ในการเรียนที่เพิ่มเติมขึ้นมาจะซื้อจากที่ไหน
- จะเซตห้องที่บ้านให้ตรงกับlayoutของห้องเรียนในโลกเสมือนได้อย่างไร

## Idea

- มหาวิทยาลัย/ภาครัฐสนับสนุนอุปกรณ์เพิ่มเติมในการเรียน
- แควสที่ให้ทำไม่เป็นการบังคับ (ที่ไม่ใช่เป็นการส่งงาน)