

時系列アソシエーション分析を用いた 購買行動に関する特徴分析

専修大学大学院

鈴木 元也
梅田 宙
大竹 恒平

- 背景
- 目的
- データ概要
- 分析の流れ
- 顧客要因の抽出
- クラスタ分析
- 購買データの加工
- アソシエーション分析
- まとめ

- 近年データマイニングツールが普及したことによって、データマイニングを手軽に用いることが可能となっている。
- アソシエーション分析に関してもよく用いられる手法である。
- しかし、アソシエーション分析はたくさんのルールを抽出することができるが、その解釈は容易ではない。
- 解釈を容易に行えるようにするのは今後のデータマイニングで重要な課題である。

目的

- クラスタリングする際よく用いられる、デモグラフィック要因を用いず、顧客の行動要因を用いて、クラスタリングを行う。
- また、クラスター別にアソシエーション分析を行うことで、より詳細な行動パターンを抽出する。
- その際、クラスター間の購買パターンの関係を考慮した解釈と、特徴的な商品を取り出すことで購買パターンの解明を試みる。

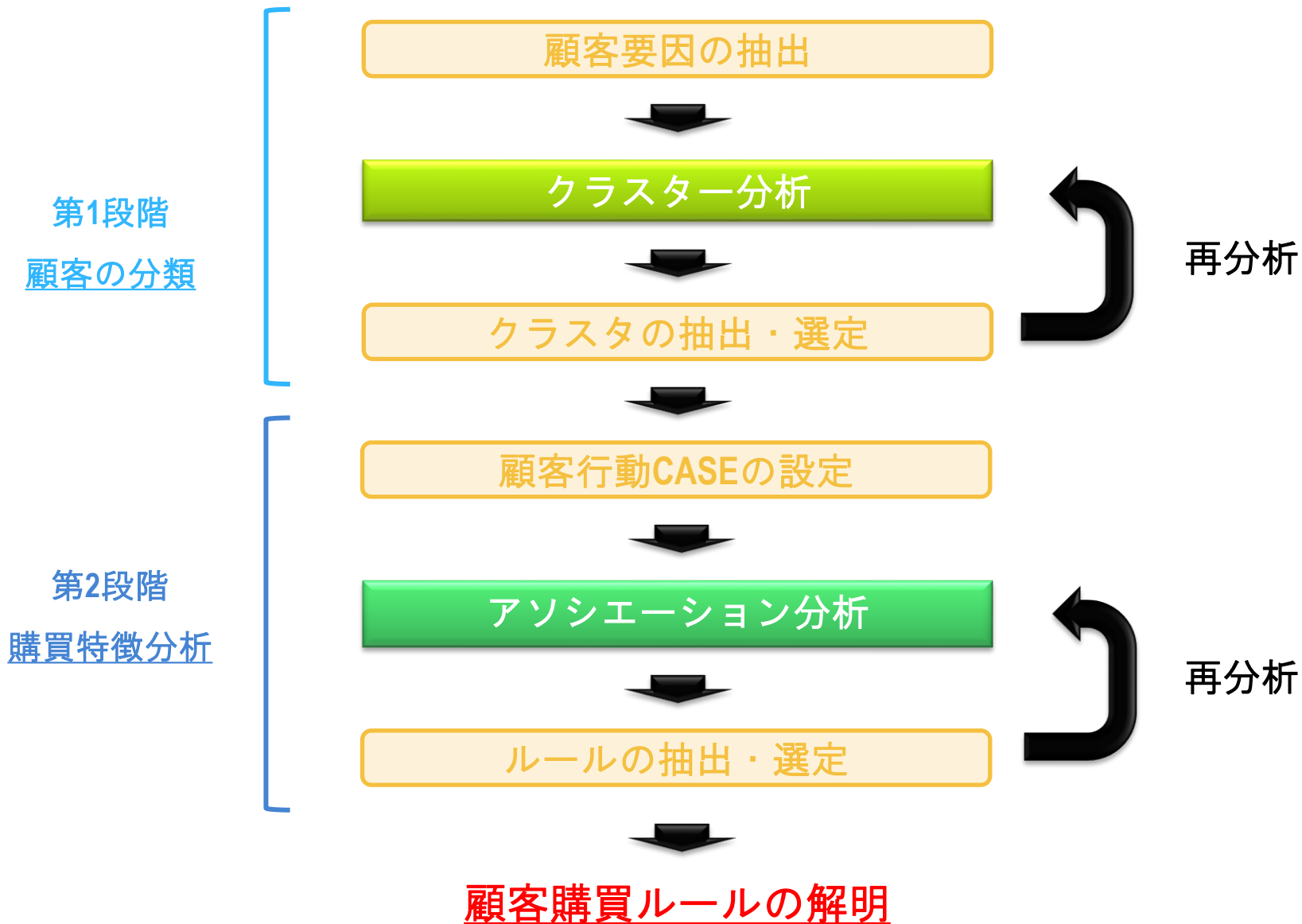
データ概要

- データ種類 : ID付POSデータ
- 小売業態 : スーパーマーケット（地方）
- データ期間 : 2001年9月1日 ~ 2006年6月30日
- 対象会員数 : 24,737人
- データ項目 : 顧客ID番号、購買日、購買時間、商品カテゴリ(3段階)、商品名、JANコード、商品単価、購買点数、購買金額、性別、年齢、世帯構成

※POSデータ例

店コード	年月日	時間	FSPコード	部門	大分類	中分類	JAN	商品名	単価	金額	点数
35*	20060718	1718	30303****	1	1	2	49303*****	*	280	280	1
35*	20060819	1458	30303****	1	1	2	49303*****	*	480	480	1
35*	20060707	1651	30303****	1	2	3	49303*****	*	98	98	1
35*	20060207	2306	30303****	1	2	3	49303*****	*	98	98	1
35*	20060303	1838	30303****	1	2	5	49303*****	*	198	198	1
35*	20060626	1226	30303****	1	1	3	49303*****	*	158	158	1
35*	20060305	1420	30303****	1	1	3	49303*****	*	198	198	1

分析の流れ



分析（顧客要因の抽出）

■顧客要因と考えられる仮説一覧

仮説1



セールに反応するような特徴的な顧客が存在するのでは？

仮説2



朝来店する顧客と夕方来店する顧客の購買行動は異なるのでは？

仮説3



直近での購買情報によって顧客の購買行動は異なるのでは？

仮説4



顧客の過去の購買によって顧客の購買行動は異なるのでは？

仮説5



顧客登録期間によって顧客の購買行動は異なるのでは？



クラスター分析

- ✓ 顧客をセグメントする際、一般的なデモグラフィックな要因ではなく、顧客の行動ベースでの要因を上記にあるよう考え、詳細を見る。

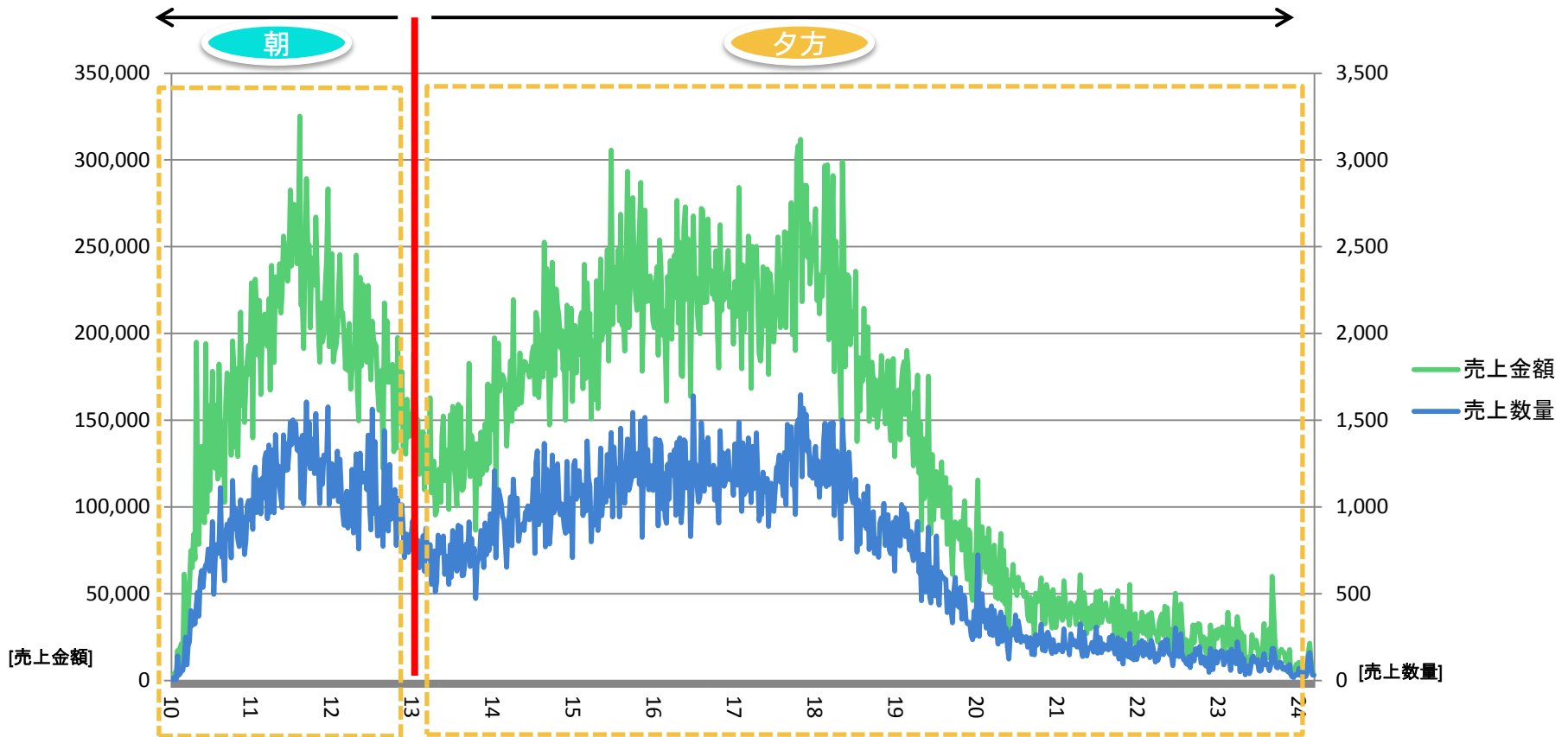
分析（顧客要因の抽出）

仮説2



朝来店する顧客と夕方来店する顧客の購買行動は異なるのでは？

■時間別推移(2006年2月)



✓ 11時30分ごろと、17時～18時ごろにピークがある。

分析（顧客要因の抽出）

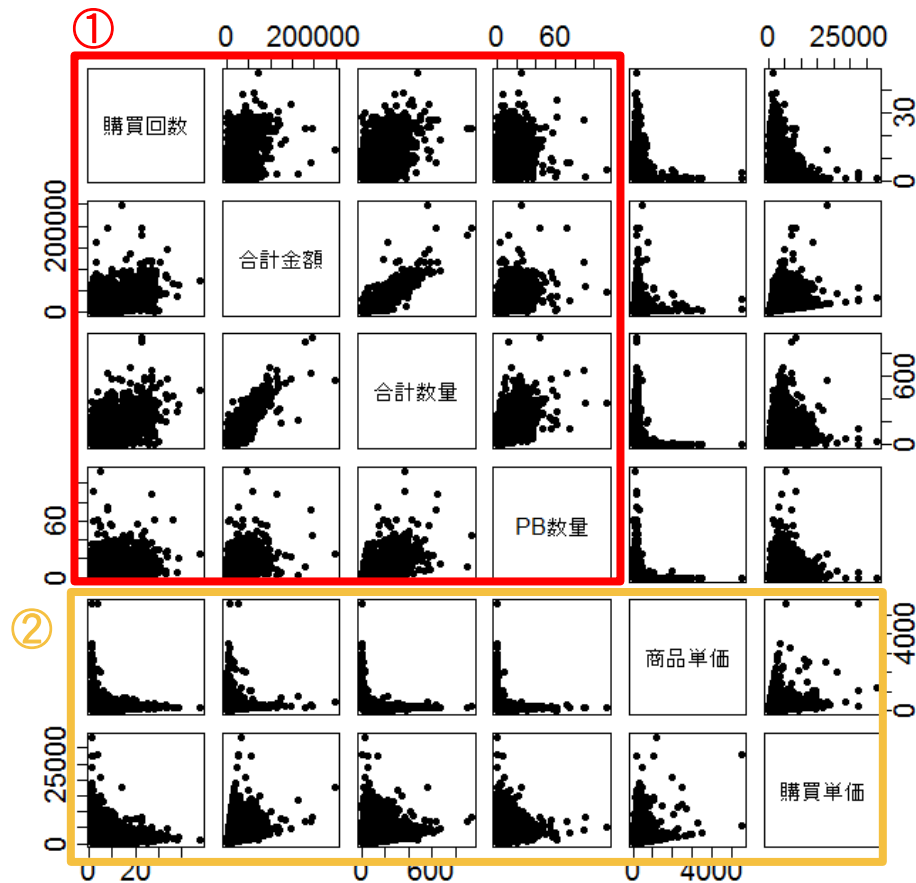
仮説3



直近での購買情報によって顧客の購買行動は異なるのでは？

■顧客別対象月の購買情報(2006年2月)

- 購買回数：1か月あたりの購買を行った回数の総和
- 合計金額：1か月あたりの購買金額の総和
- 合計数量：1か月あたりの購買数量の総和
- PB数量：1か月あたりのプライベートブランド購買数量の総和
- 商品単価：合計金額／合計数量
- 購買単価：合計金額／購買回数



- ✓ ① . . . 購買回数、金額、数量、PBは相関している。
- ✓ ② . . . 商品単価が高い商品を買いつけることは少ない。

分析（顧客要因の抽出）

仮説4



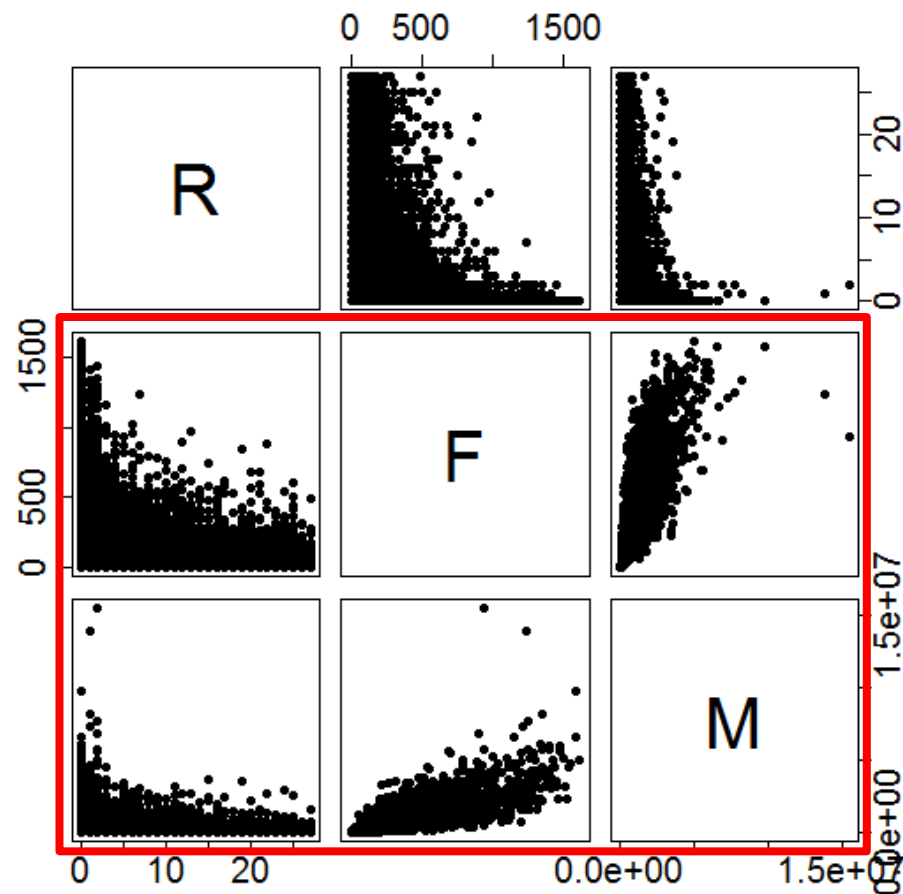
顧客の過去の購買によって顧客の購買行動は異なるのでは？

■ 顧客別全期間の購買情報 (2001年9月～2006年2月)

RFM

R (recency : 最新購買日) 最近いつ買ったか
F (frequency : 累計購買回数) どのくらいの頻度
M (monetary : 累計購買金額) いくら使っているか

- R : 2006年2月28日時点での最新購買日
- F : 2006年2月28日までの購買回数
- M : 2006年2月28日までの購買金額



✓ 購買回数Fに比べて購買金額Mは変化にあまり差がない。

分析（顧客要因の抽出）

仮説5



顧客登録期間によって顧客の購買行動は異なるのでは？

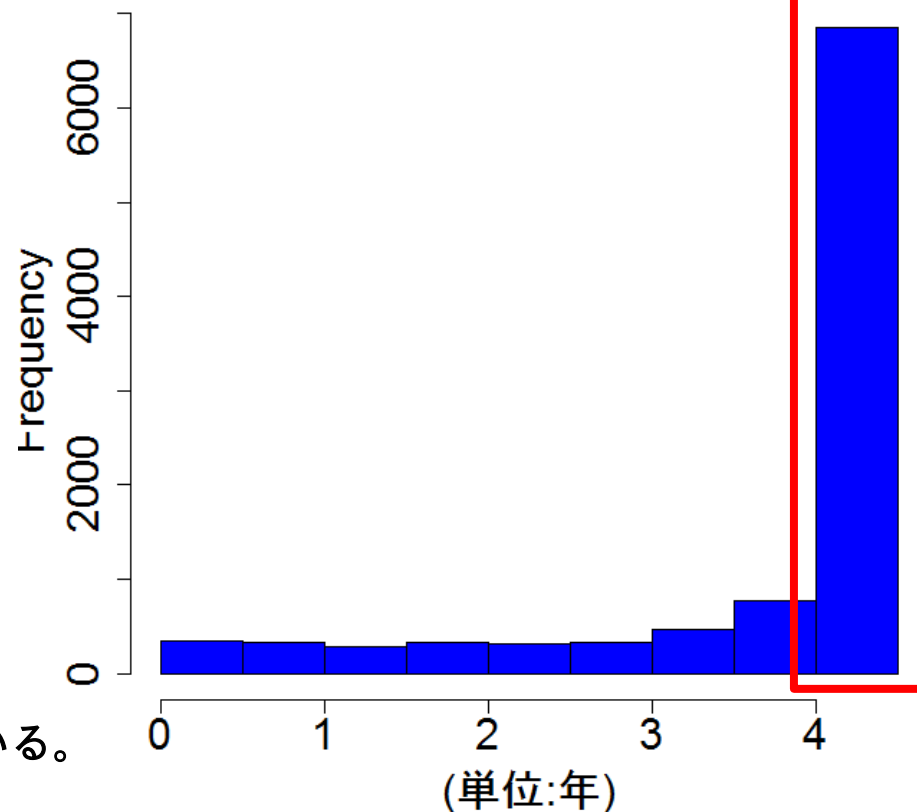
■顧客別登録期間(2006年2月28日時点)

□ 顧客登録日：最新購買日 - 最古購買日

※顧客登録日の情報がなかったため、最古購買日を顧客登録日とした。

- ✓ 利用期間が4年以上の顧客がほとんど。
- ✓ スーパー利用顧客は長期間利用されている。

顧客登録期間



分析（第1段階：クラスタ分析）

■ 分析手法：k-menas法

- ・ ・ ・ 非階層型クラスタリングの手法の1つ。クラスタを決めその中心から近い点をクラスタリングする。
- ・ ・ ・ 階層型と比較し計算にかかる処理がすくないことや、視覚的にわかりやすいことからこの手法を選択した。

利用データ

- ・ 仮説該当項目

ツール

- ・ R
- ・ パッケージ(cclust)

クラスタ決定方法

- ・ `clustIndex`関数
(指標：calinski)

```
>library(cclust)
```

```
>cclust (x, centers, iter.max=100, verbose=FALSE, dist="euclidean", method= "kmeans",  
rate.method="polynomial", rate.par=NULL) #k-menas法
```

```
>clustIndex ( y, x, index = "all" ) #クラスタ決定指標
```

分析（第1段階：クラスター分析）

■ クラスタ数の決定

cluster	calinski	hartigan	ratkowsky	db	scott	friedman	rubin	xuindex
k-means(2クラスター)	14,843.67	0.39	0.28	706,863.30	9,402.23	1.56	2.55	94.48
k-means(3クラスター)	17,132.03	1.23	0.27	782,873.48	16,686.73	3.88	5.27	91.14
k-means(4クラスター)	15,659.51	1.54	0.25	784,178.00	20,735.77	5.65	7.89	89.78
k-means(5クラスター)	17,046.23	1.92	0.23	828,665.08	24,543.11	7.90	11.54	87.95
k-means(6クラスター)	17,001.56	2.14	0.21	857,544.83	27,539.72	10.19	15.55	86.87

2クラスター

3クラスター

4クラスター

5クラスター

5クラスター

1cluster N=1383

1cluster N=6976

1cluster N=442

1cluster N=4042

1cluster N=164

2cluster N=8653

2cluster N=2513

2cluster N=2926

2cluster N=227

2cluster N=1536

3cluster N=547

3cluster N=1742

3cluster N=1031

3cluster N=2420

4cluster N=4926

4cluster N=1974

4cluster N=728

5cluster N=2763

5cluster N=2070

6cluster N=3113

分析（第1段階：クラスター分析）

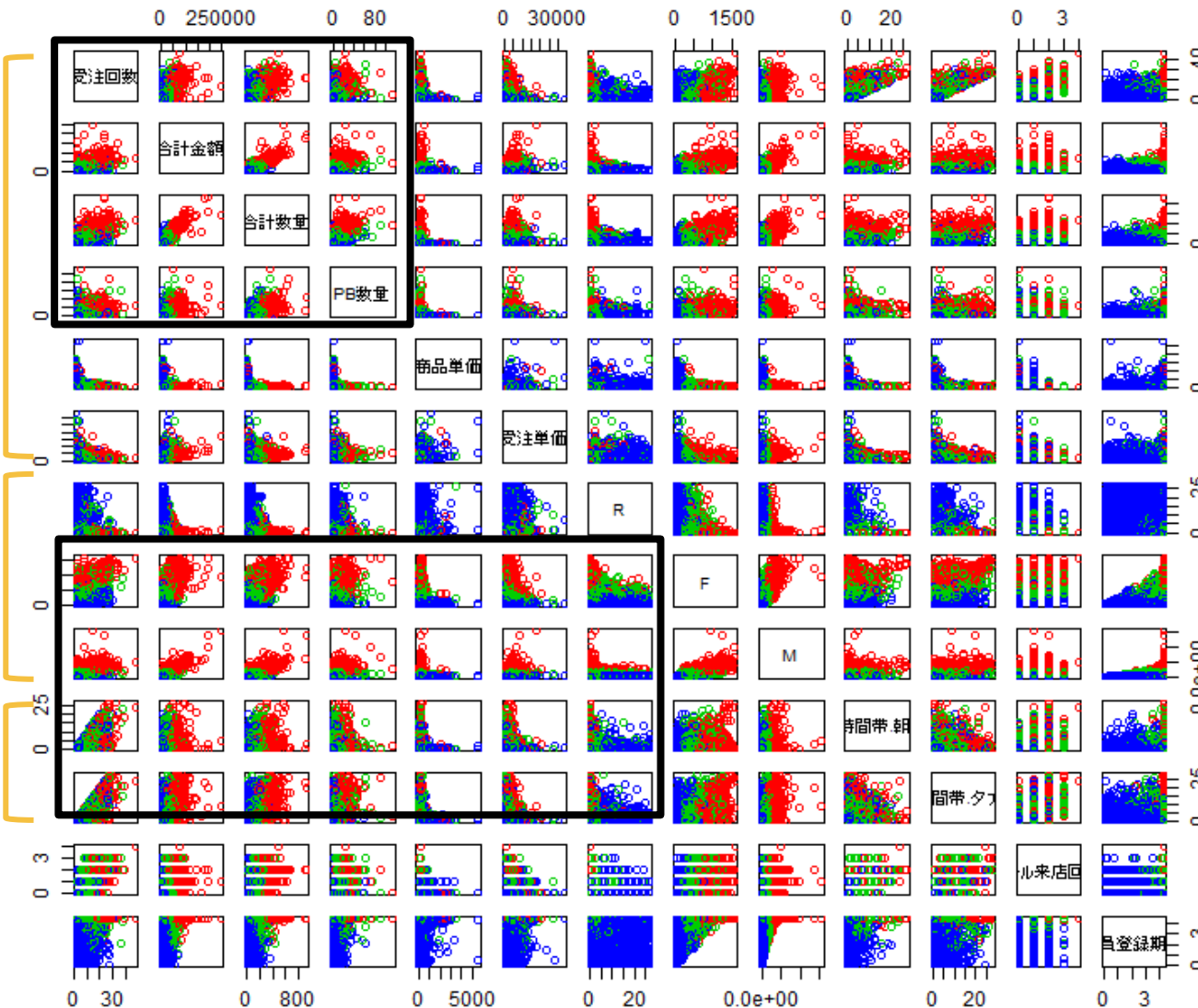
仮説③

仮説④

仮説②

①

⑤



青：N=6976人
 緑：N=2513人
 赤：N=547人



青：cluster1
 (購買回数が少ない)
 緑：cluster2
 (PBは購入しているが金額が
 高くない。)
 赤：cluster3
 (購入金額が高い)

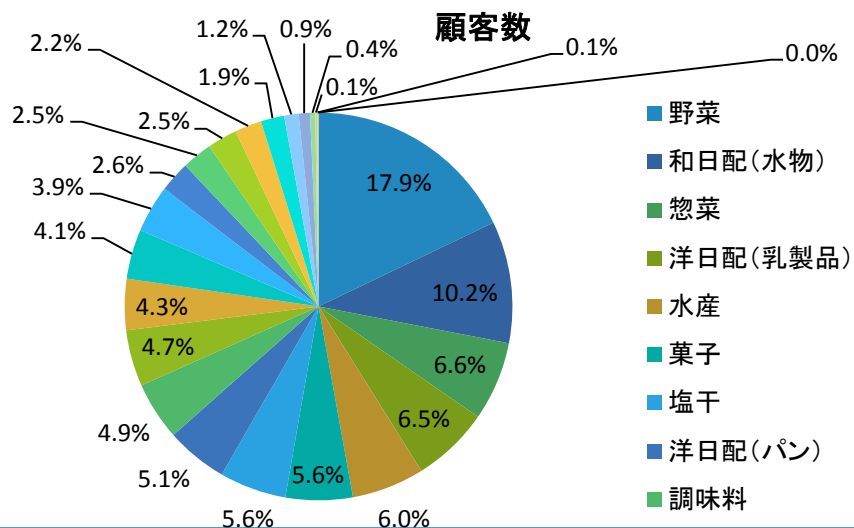
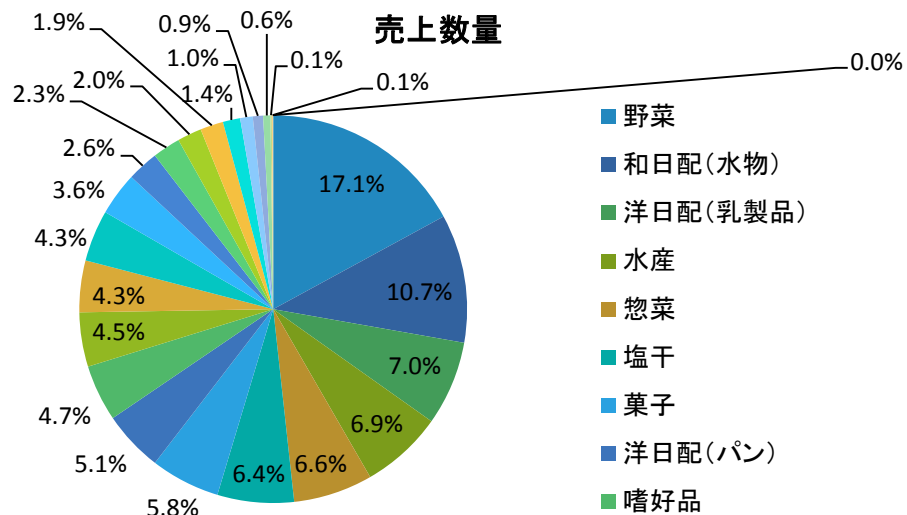
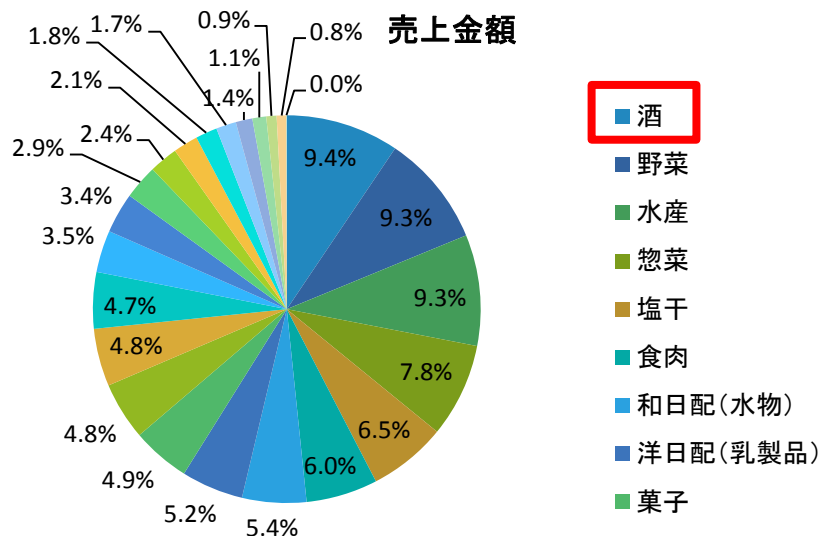


Cluster1：ライトユーザ
 Cluster2：ミディアムユーザ
 Cluster3：ヘビーユーザ

顧客優良度別の
 クラスターに分類

分析（顧客行動CASEの設定）

■商品カテゴリ別購買情報



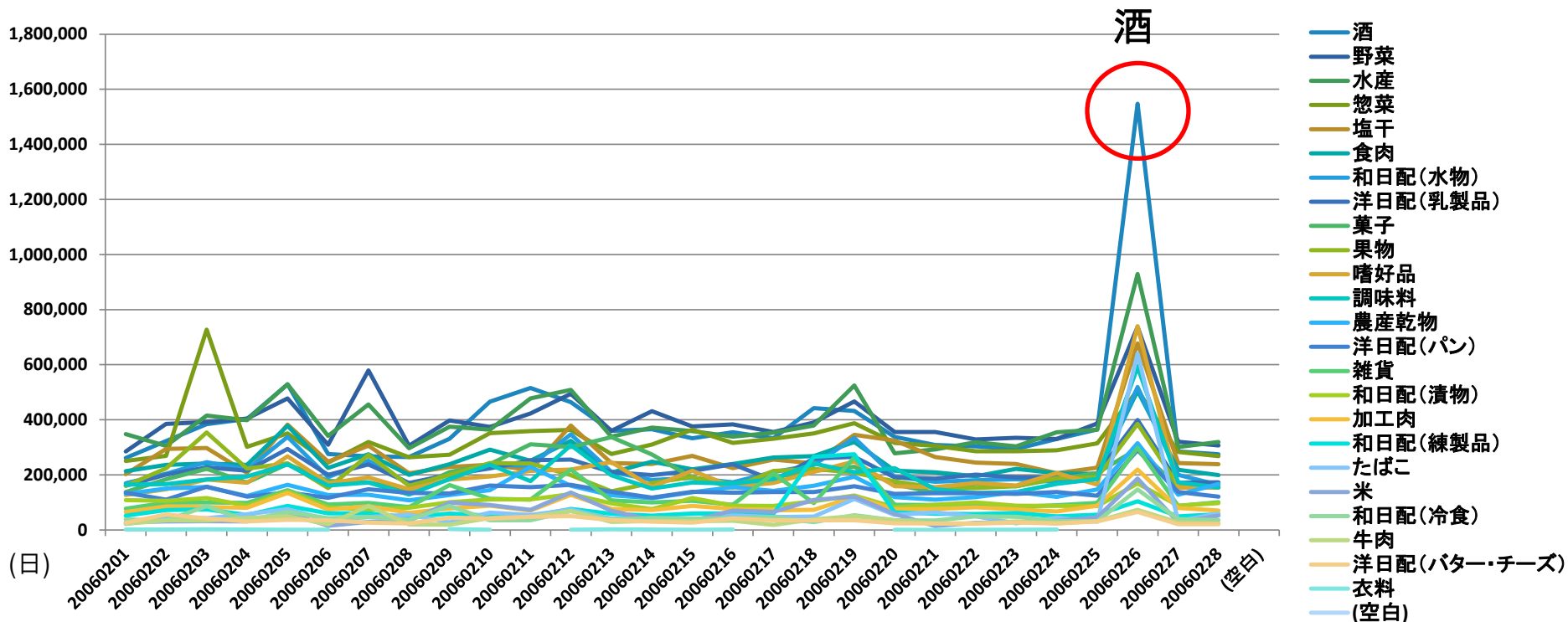
- ✓ 酒の売上が高い(特に単価が)
- ✓ 全体の野菜の購入割合が高い



時系列に傾向を把握する

分析（顧客行動CASEの設定）

■商品カテゴリ別売上推移



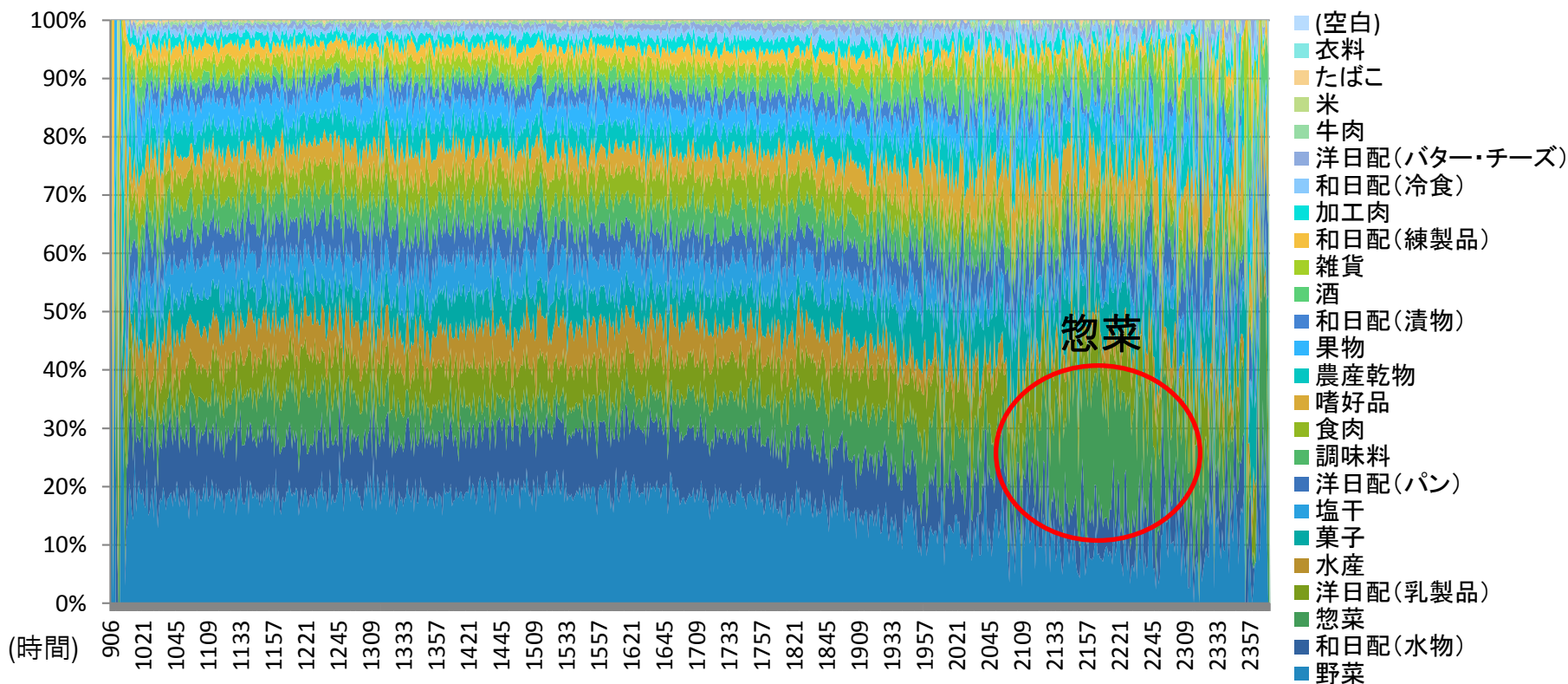
✓ 一定して売上の高い「酒」がある特定の日（セール時）に極端に売上が増加する



「酒」購入顧客の購買行動を解明することは店にとって有益である。

分析（顧客行動CASEの設定）

商品カテゴリ別顧客数推移(構成比率)



✓ ほとんど同様に变化する商品構成比だが、特定の時間の「葱菜」の占める割合が高い



他商品と異なる購買をする顧客の購買を解明することは店にとって有益である。

分析（顧客行動CASEの設定）

■顧客行動CASEの設定

顧客優良度別のクラスタ分類より

CASE 1

顧客育成のための顧客購買パターンの解明

売上情報の推移より

CASE 2

特徴的商品の顧客行動パターンの解明



対象商品

- ・ 酒 . . . 一部の顧客が定期的に購入するが売上
の上位を占める商品
- ・ 惣菜 . . . 夕方から夜にかけて売れる商品

分析（第2段階：アソシエーション分析）

■分析手法：時系列アソシエーション分析

- ・ ・ ・ アソシエーション分析とは、多量のデータの中から価値のある相関関係(ルール)を時系列に見つけ出すこと。
- ・ ・ ・ 購買パターンを抽出する代表的手法であり、時系列を考慮することが可能であるためこの手法を選択した。

ツール

- ・ VMS (時系列アソシエーション分析)

利用データ

- ・ キー列：FSPコード
- ・ 時系列：年月日
- ・ 対象列：部門、大分類

評価指標の閾値

- ・ 信頼度 > 0%
- ・ サポート >= 3%
- ・ リフト >= 1

中分類、商品名



加工

分析（第2段階：アソシエーション分析）

■商品カテゴリの大小関係

部門

>

大分類

>

中分類

>

商品名

大きい

大きい

少し大きい

細かい

中分類

キャベツ

キャベツ

キャベツ

商品名

キャベツ半分

春キャベツ1/4

キャベツ

新商品名

キャベツ_小

キャベツ_小

キャベツ_大

※大きさ、量、
質を考慮

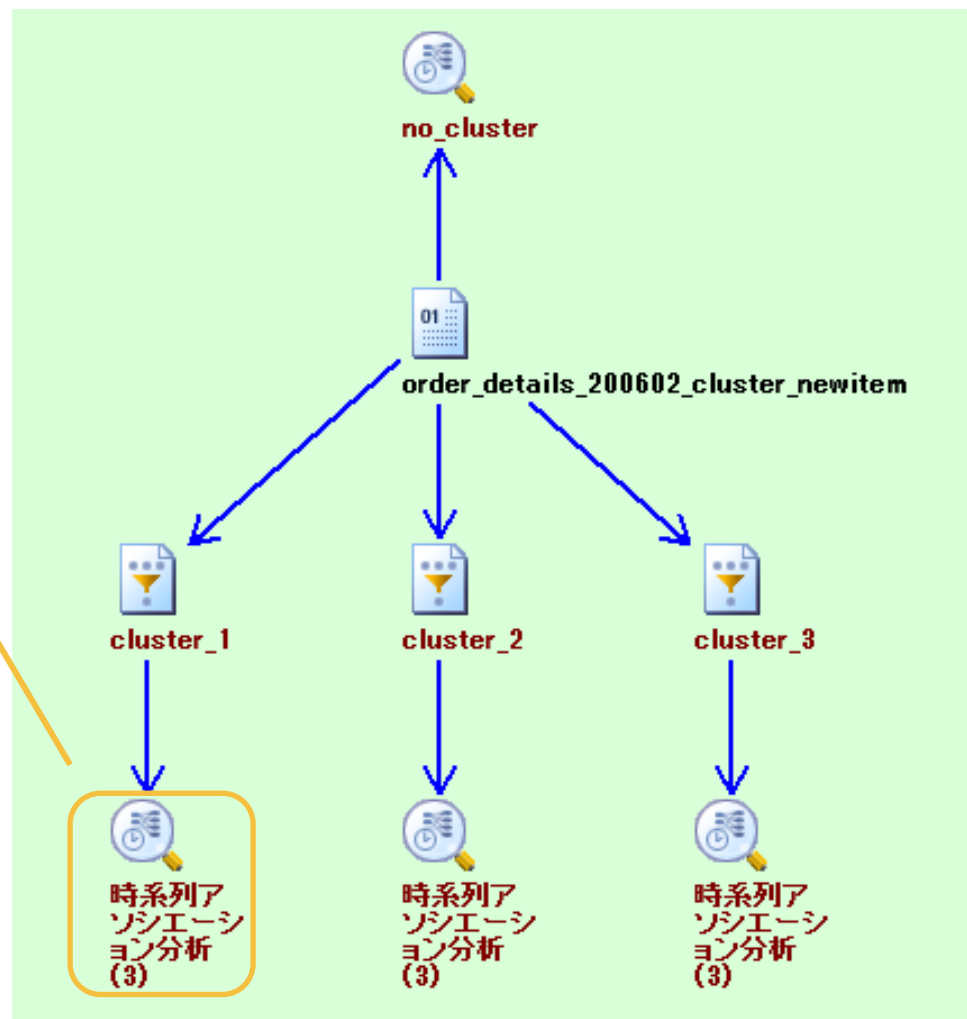
分析（第2段階：アソシエーション分析）

CASE 1

顧客育成のための顧客購買パターンの解明

アウトプットデータ

前提	結論	サポート	信頼度	1.リフト
24-弁当	24-弁当	3.842	28.122	2.059
24-プレーンヨーグルト	24-プレーンヨーグルト	4.358	25.741	1.520
24-牛乳->24-丸納豆	24-牛乳	3.326	48.333	1.487
24-食パン	24-食パン	5.433	27.664	1.409
24-牛乳->24-木綿豆腐	24-牛乳	3.426	45.437	1.398
24-丸納豆+24-牛乳	24-牛乳	4.114	44.496	1.369
24-丸納豆->24-牛乳	24-牛乳	3.225	44.466	1.368
24-牛乳->24-牛乳	24-牛乳	5.677	44.444	1.368
24-牛乳+24-食パン	24-牛乳	3.426	43.853	1.349
24-木綿豆腐->24-牛乳	24-牛乳	3.469	43.060	1.325
24-豚肉薄切り	24-豚肉薄切り	3.297	20.890	1.324
24-木綿豆腐+24-牛乳	24-牛乳	4.472	42.857	1.319
24-牛乳	24-木綿豆腐+24-牛乳	4.458	13.719	1.315
24-牛乳	24-牛乳+24-普通卵	4.659	14.336	1.306
24-牛乳->24-菓子パン	24-牛乳	3.512	42.169	1.298
24-木綿豆腐->24-木綿豆腐	24-木綿豆腐	3.455	37.893	1.284
24-牛乳	24-牛乳+24-ホウレンソウ	3.211	9.881	1.284
24-菓子パン->24-菓子パン	24-菓子パン	4.917	43.974	1.279
24-牛乳->24-普通卵	24-牛乳	3.369	41.228	1.269
24-牛乳	24-丸納豆+24-牛乳	3.770	11.601	1.255
24-デニッシュペストリー	24-菓子パン	3.326	42.804	1.245
24-丸納豆+24-木綿豆腐	24-丸納豆	3.426	31.614	1.241
24-普通卵->24-牛乳	24-牛乳	3.139	39.818	1.225



分析（第2段階：アソシエーション分析）

CASE 1

顧客育成のための顧客購買パターンの解明

■ 評価方法

1. 有効とされたルールの結論部にあたる、種類の割合が前後のセグメントとどの程度変わるのか。
2. 変化の大きい結論部の前提部を取り出し傾向をみる。

※具体例

セグメント1	セグメント2
結論部	結論部
豆腐 30%	豆腐 50%
牛乳 20%	牛乳 10%
味噌 5%	味噌 15%
その他 45%	その他 25%



各セグメントの「豆腐」の前提部の傾向を比較する。

※全体の特徴を捉えるため傾向を比較する際の基準にサポートを利用する

分析（第2段階：アソシエーション分析）

■ サポート値が上位10位のルールを抽出

Cluster_1

結論部	結論数	割合
24-牛乳	15	26.8%
24-木綿豆腐	5	8.9%
24-菓子パン	4	7.1%
24-普通卵	4	7.1%
24-木綿豆腐+24-普通卵	2	3.6%
24-牛乳+24-普通卵	2	3.6%
24-海藻	2	3.6%
24-木綿豆腐+24-牛乳	2	3.6%
24-丸納豆	2	3.6%
24-丸納豆+24-木綿豆腐	2	3.6%

洋生菓子のルールの割合が多い。



Cluster_2

結論部	結論数	割合
24-牛乳	288	12.7%
24-洋生菓子	129	5.7%
24-食パン	127	5.6%
24-プレーンヨーグルト	73	3.2%
24-菓子パン	67	3.0%
24-バナナ	54	2.4%
24-もやし大	53	2.3%
24-胡瓜中	41	1.8%
24-普通卵	41	1.8%
24-さけ・ます（塩蔵魚）	36	1.6%

Cluster_3

結論部	結論数	割合
24-人参中	4,213	3.6%
24-たこ	2,235	1.9%
24-牛乳	2,094	1.8%
24-プレーンヨーグルト	2,083	1.8%
24-木綿豆腐	1,785	1.5%
24-食パン	1,783	1.5%
24-さけ・ます（塩蔵魚）	1,542	1.3%
24-煮物野菜	1,531	1.3%
24-洋生菓子	1,529	1.3%
24-胡瓜中	1,486	1.3%

人参のルールの割合が多い。



分析（第2段階：アソシエーション分析）

■結論部が一致した前提部の一部を抽出

Cluster_2 [結論：洋生菓子]

前提	結論	サポート
24-菓子パン	24-洋生菓子	15.32033
24-菓子パン->24-菓子パン	24-洋生菓子	9.789097
24-チョコレート	24-洋生菓子	8.396339
24-プレーンヨーグルト	24-洋生菓子	8.356546
24-木綿豆腐->24-菓子パン	24-洋生菓子	7.958615



Cluster1の顧客に菓子パン、やお菓子などを購入させて「洋生菓子」を購入させることでcluster2へ促進することができるのでは？

Cluster_3 [結論；人參中]

前提	結論	サポート
24-木綿豆腐	24-人參中	15.90494
24-普通卵	24-人參中	15.17367
24-もやし大	24-人參中	14.80804
24-ホウレンソウ	24-人參中	14.80804
24-丸納豆	24-人參中	14.80804
24-木綿豆腐->24-木綿豆腐	24-人參中	14.44241



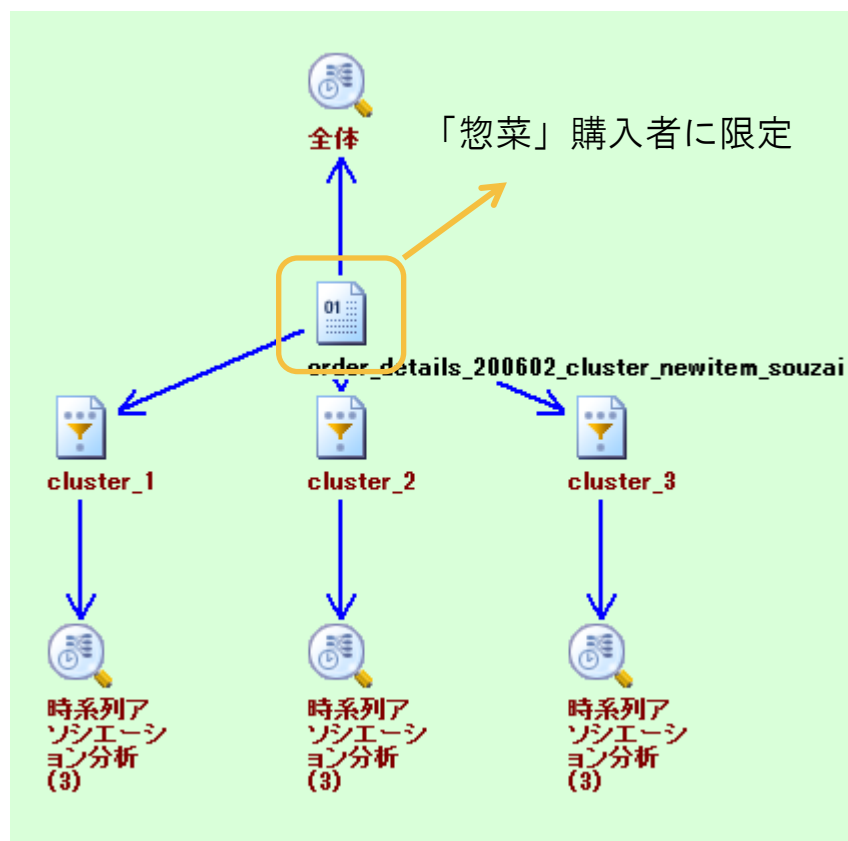
Cluster2の顧客に木綿豆腐などを購入させて「人參中」を購入させることでcluster3へ促進することができるのでは？

分析（第2段階：アソシエーション分析）

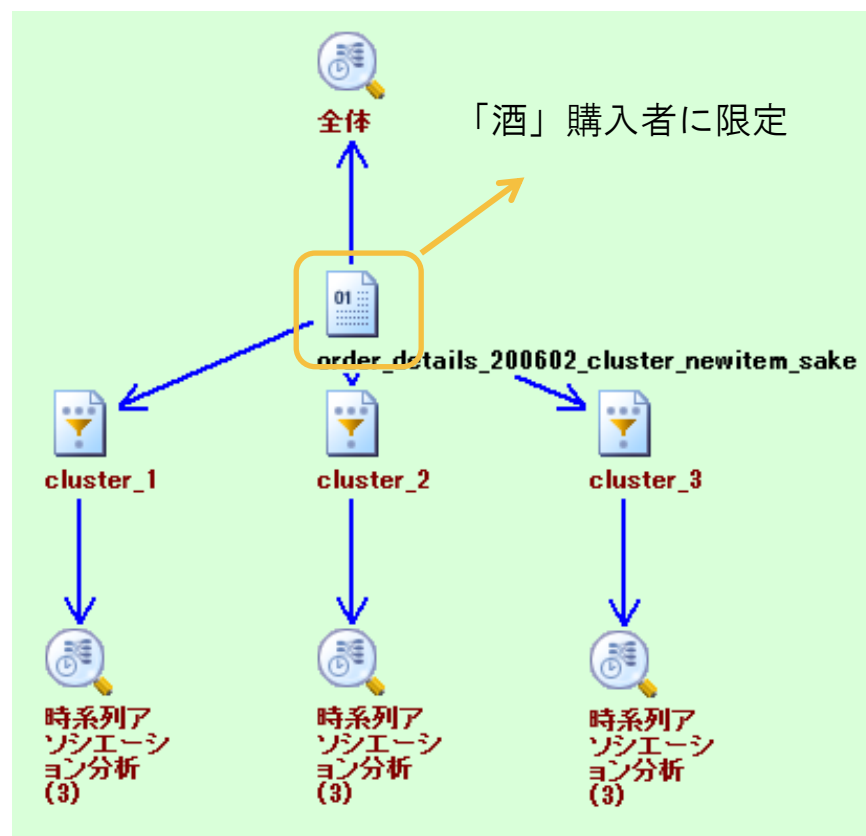
CASE 2

特徴的商品の顧客行動パターンの解明

「惣菜」購入者モデル



「酒」購入者モデル



分析（第2段階：アソシエーション分析）

CASE 2

特徴的商品の顧客行動パターンの解明

■ 評価方法

- ・ 特徴的商品が結論部に存在する場合の前提条件の傾向をみる。

※具体例

※前提部を重視するため傾向を比較する際の基準にリフトを利用する

豆腐を
購入する傾向

前提部	結論部	リフト
豆腐	発泡酒	2.23
牛乳	発泡酒	2.11
豆腐	その他発泡酒	1.99
なす	発泡酒	1.88

分析（第2段階：アソシエーション分析）

■特徴商品「酒」

no_Cluster

前提	結論	リフト
24-発砲酒->24-発砲酒	24-発砲酒	2.2
24-その他雑酒->24-その他雑酒	24-その他雑酒	2.2
24-発砲酒	24-発砲酒+24-普通卵	2.2
24-木綿豆腐+24-発砲酒	24-発砲酒	2.1
24-発砲酒	24-木綿豆腐+24-発砲酒	2.1
24-発砲酒	24-牛乳+24-発砲酒	2.0
24-発砲酒->24-即席麺	24-発砲酒	1.9
24-発砲酒->24-木綿豆腐	24-発砲酒	1.8
24-即席麺->24-発砲酒	24-発砲酒	1.8
24-木綿豆腐->24-発砲酒	24-発砲酒	1.8

- ✓ 発砲酒もしくは木綿豆腐を買う傾向

Cluster_1

前提	結論	リフト
24-その他雑酒->24-その他雑酒	24-その他雑酒	1.9
24-発砲酒	24-発砲酒	1.4
24-その他雑酒	24-その他雑酒	1.0

- ✓ お酒を前回購入している傾向

Cluster_2

前提	結論	リフト
24-その他雑酒->24-その他雑酒	24-その他雑酒	2.3
24-発砲酒->24-発砲酒	24-発砲酒	2.3
24-発砲酒	24-ハウレンソウ+24-	2.3
24-木綿豆腐+24-発砲酒	24-発砲酒	2.2
24-即席麺+24-発砲酒	24-発砲酒	2.2
24-発砲酒	24-即席麺+24-発砲酒	2.2
24-発砲酒	24-牛乳+24-発砲酒	2.1
24-発砲酒	24-発砲酒+24-菓子バ	2.1
24-牛乳+24-発砲酒	24-発砲酒	2.0
24-発砲酒	24-発砲酒+24-普通卵	1.9

- ✓ 発砲酒もしくは即席麺、木綿豆腐を買う傾向

Cluster_3

前提	結論	リフト
24-清酒金印->24-もやし大	24-清酒金印	4.0
24-その他雑酒+24-ちくわ	24-その他雑酒	3.6
24-その他雑酒->24-デニツ	24-その他雑酒	3.5
24-その他雑酒->24-砂糖	24-その他雑酒	3.4
24-鍋物セット大->24-その他雑酒	24-その他雑酒	3.4
24-生珍味+24-その他雑酒	24-その他雑酒	3.4
24-その他雑酒->24-板こん	24-その他雑酒	3.4
24-その他雑酒	24-その他雑酒+24-催	3.4
24-近海魚+24-発砲酒	24-発砲酒	3.3
24-発砲酒	24-しめじ小+24-発砲	3.3

- ✓ ごはんのついでに買う傾向

分析（第2段階：アソシエーション分析）

■特徴商品「酒」

no_Cluster

前提	結論	リフト
24-惣菜（練）	24-惣菜（練）	1.7
24-ちくわ	24-惣菜（練）	1.4
24-牛乳	24-惣菜（練）+24-牛乳	1.4
24-ピーマン	24-惣菜（練）	1.4
24-菓子パン->24-もやし	24-惣菜（練）	1.3
24-牛乳->24-もやし大	24-惣菜（練）	1.3
24-板なしかま	24-惣菜（練）	1.3
24-菓子パン->24-丸納豆	24-惣菜（練）	1.3
24-菓子パン->24-油揚	24-惣菜（練）	1.3
24-木綿豆腐+24-菓子パン	24-惣菜（練）	1.2

✓ 菓子パンをを購入する傾向

Cluster_1

✓ ルールなし

Cluster_2

前提	結論	リフト
24-ちくわ	24-惣菜（練）+24-牛乳	1.1
24-ちくわ	24-惣菜（練）	1.0
24-菓子パン	24-惣菜（練）+24-菓子パン	1.1
24-菓子パン->24-ちくわ	24-惣菜（練）	1.0
24-牛乳	24-惣菜（練）+24-牛乳	1.1
24-牛乳->24-ちくわ	24-惣菜（練）	1.0
24-惣菜（練）	24-惣菜（練）	1.2
24-惣菜（練）->24-木綿豆腐	24-惣菜（練）	1.2
24-木綿豆腐	24-惣菜（練）+24-海藻	1.0
24-木綿豆腐+24-ちくわ	24-惣菜（練）	1.1

✓ 練り物惣菜購入者は練り物を購入する傾向

Cluster_3

前提	結論	リフト
24-肉惣菜ハンバーグ->24-肉惣菜	24-肉惣菜ハンバーグ	3.9
24-中華加工調味料->24-肉惣菜	24-肉惣菜ハンバーグ	3.9
24-肉惣菜ハンバーグ->24-惣菜	24-肉惣菜ハンバーグ	3.5
24-肉惣菜ハンバーグ->24-油揚	24-肉惣菜ハンバーグ	3.3
24-肉惣菜->24-スープ	24-肉惣菜ハンバーグ	3.2
24-肉惣菜->24-しょうゆ漬	24-肉惣菜ハンバーグ	3.2
24-即席麺->24-肉惣菜	24-肉惣菜	3.0
24-コロッケ	24-木綿豆腐+24-肉惣菜	3.0
24-即席麺->24-肉惣菜	24-肉惣菜ハンバーグ	2.9
24-油揚->24-肉惣菜	24-肉惣菜ハンバーグ	2.8

✓ 過去に肉惣菜を購入している傾向

まとめ

- 顧客の行動ベースでクラスタリングをすることが可能である。
- 顧客の行動ベースでのクラスタリングすると、顧客の購買パターンは異なることがわかる。
- 購買パターンに、商品の大きさ、量、質的要因を入れることによってより詳細な購買パターンを見つけることができる。
- 顧客のクラスター毎の行動を解明することで顧客育成のためのアプローチの発見に繋げることができる。