

2018 年度 NTT データ数理システム  
学生研究奨励賞(Visual Mining Studio)応募用

# ECサイトにおける 優良顧客の特徴分析

中央大学 理工学部  
経営システム工学科 蘇悦

# 目次

01

はじめに

02

データ概要

03

分析

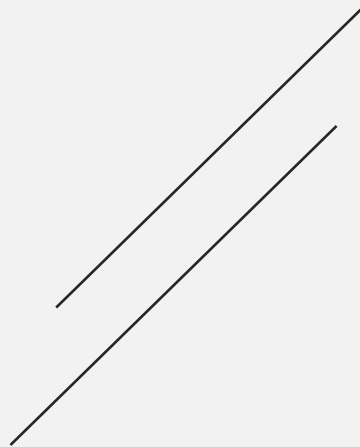
04

まとめ



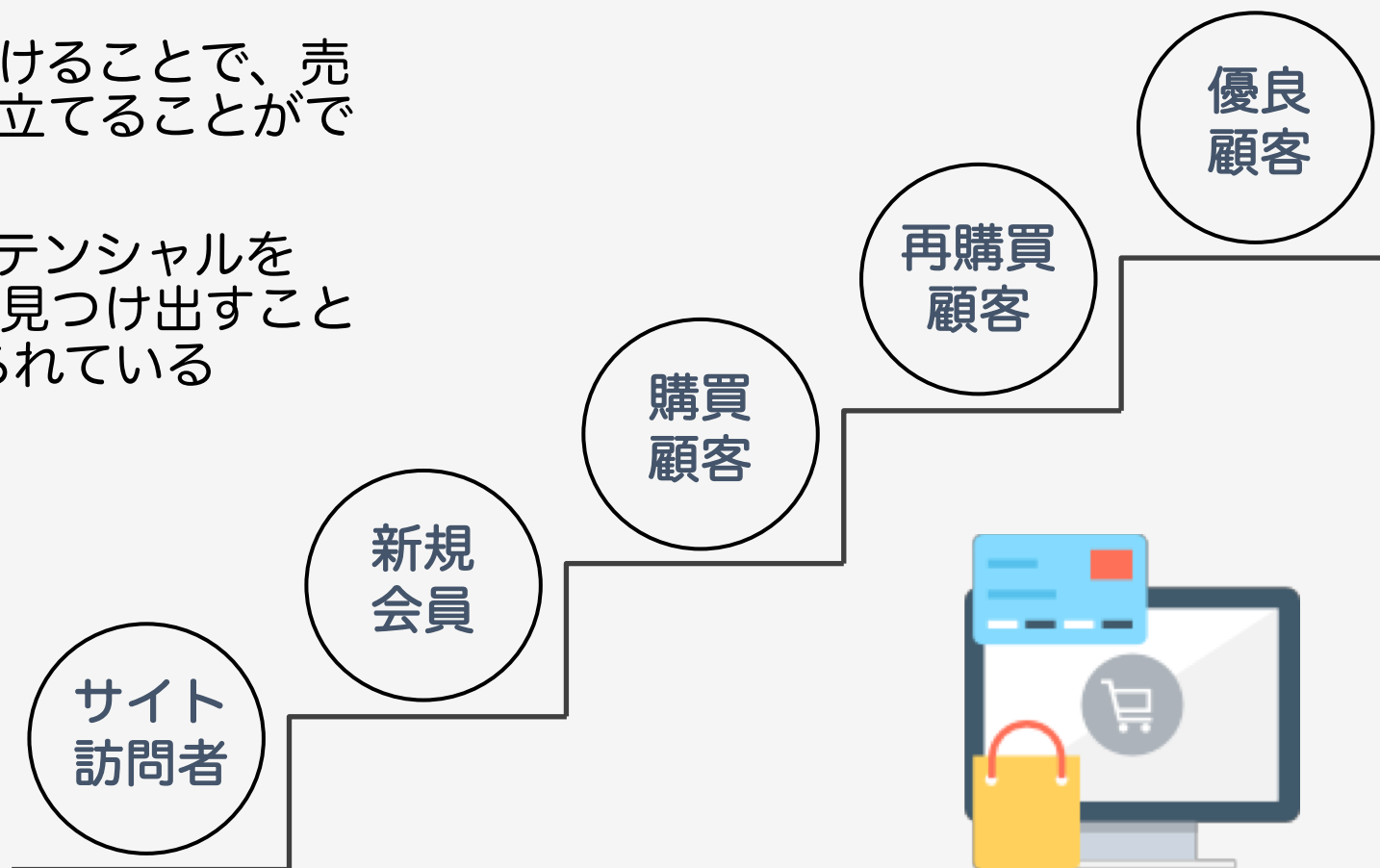
01

# はじめに

- 背景
  - 目的
- 

## ▶ 背景

- 近年，ECサイト市場の拡大を背景に，顧客獲得競争が激化していくことが予想される
- 購買意欲の高い顧客を見つけることで，売上の増加や経費の削減に役立てることができる
- 優良顧客となってくれるポテンシャルを持った新規顧客をいち早く見つけ出すことがECサイト事業者に求められている



## ▶ 目的

本研究では、

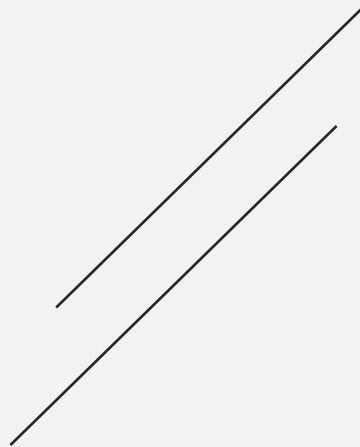
初回購買から1年間に  
再購買を行った顧客を対象に、  
ロイヤルティが高い顧客の  
初回購買時の特徴を明らかにする

ことを目的とする。



02

## データ概要

- 使用データ
  - 分析対象顧客
  - 基礎集計
- 

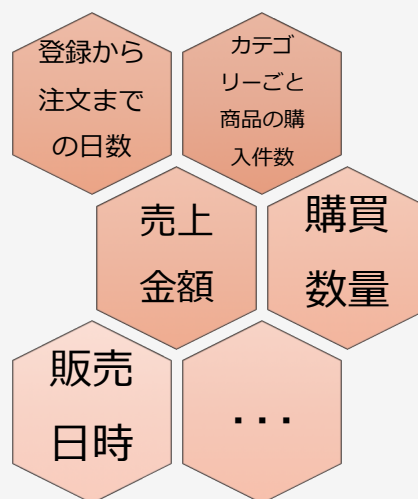
## ▶ 使用データ

- サイト：ゴルフ用品を取り扱うECサイト
- 使用データ：

### 会員顧客情報データ



### 購買データ



### アクセスログデータ

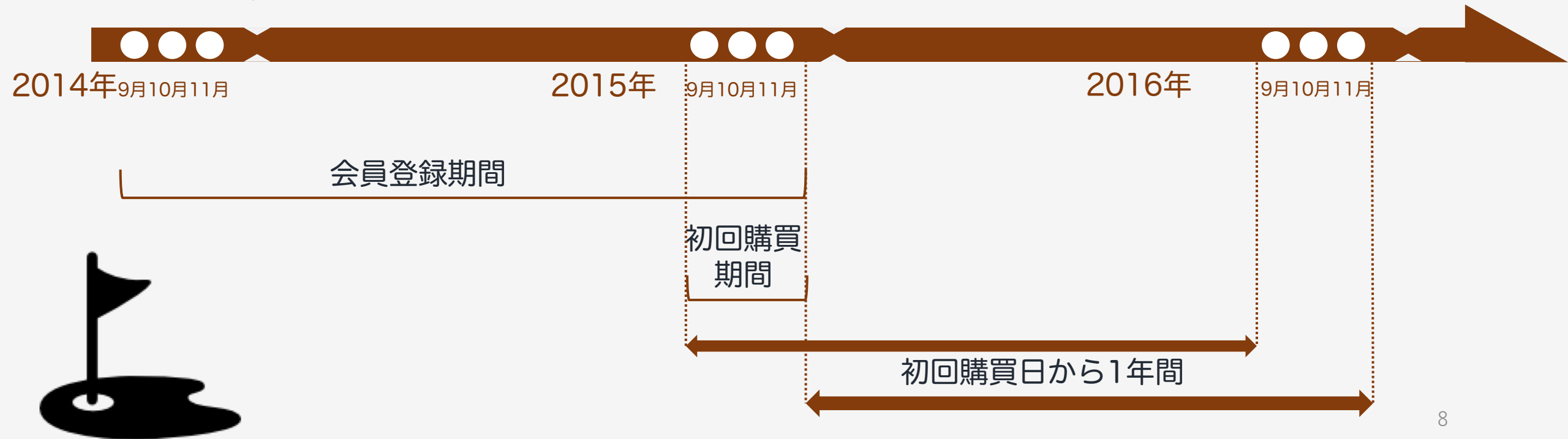


## ▶ 分析対象

- 分析ユーザ：  
2015年9月1日～2015年11月30日に初回購買を行い、1年間以内に再購買を行った顧客

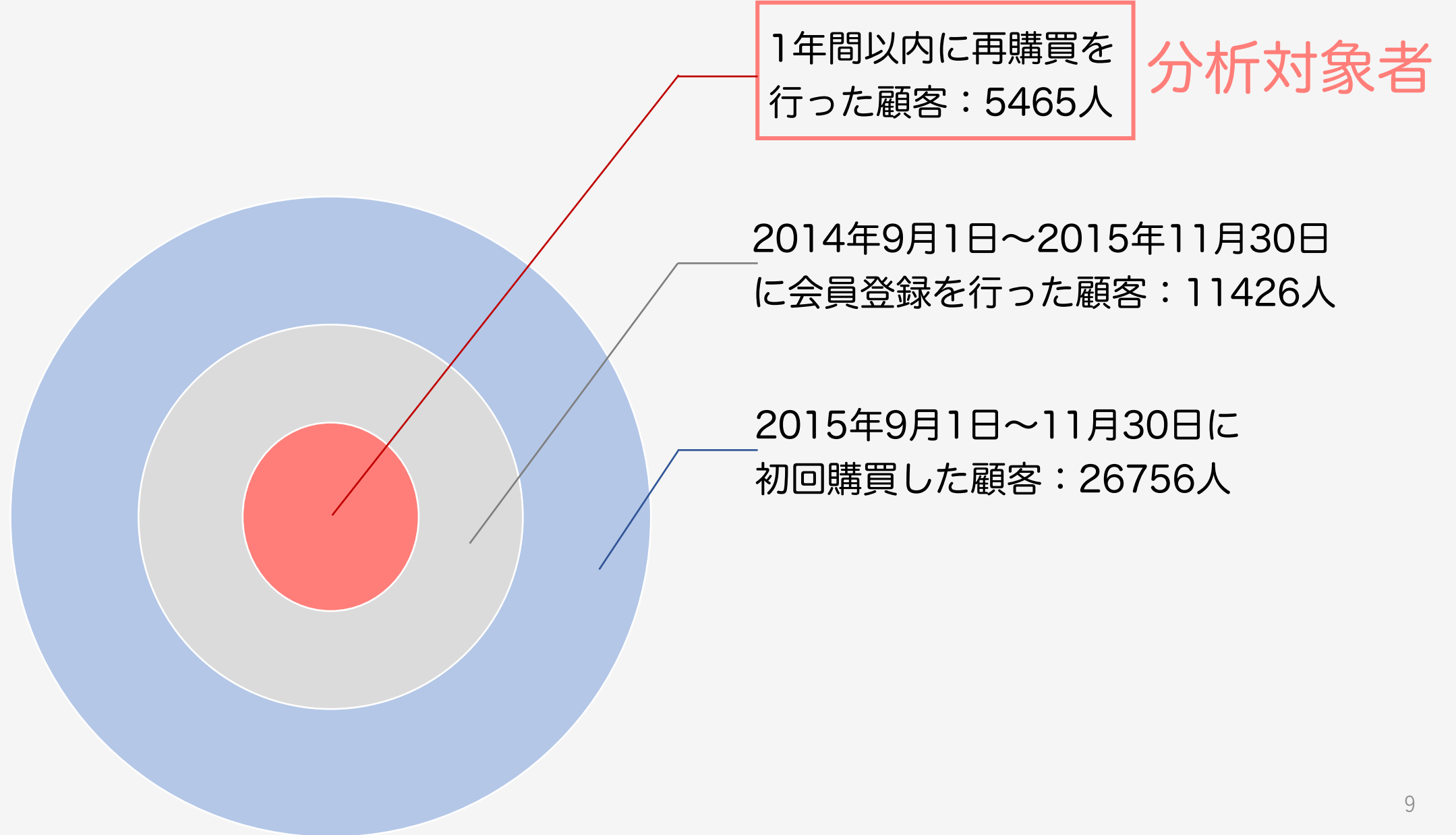
※上記の内、会員登録日から1年以上経過しているユーザは除く  
(2014年9月1日～2015年11月30日に会員登録を行った顧客)

- 分析対象期間：2015年9月1日～2016年11月30日





## ▶ 分析対象



## ▶ 基礎集計(1)

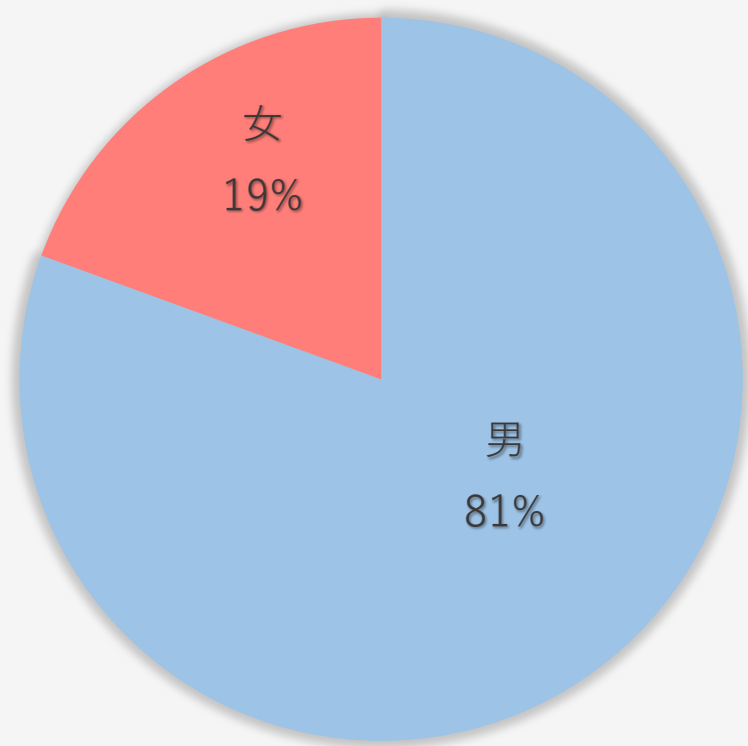


図1 男女別

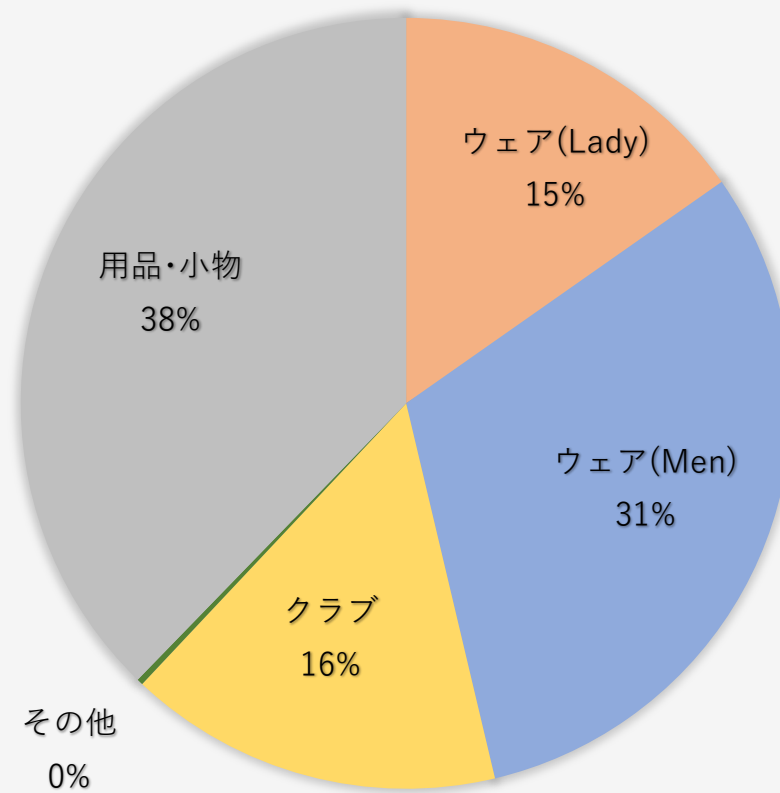


図2 初回購買内容

## ▶ 基礎集計(2)

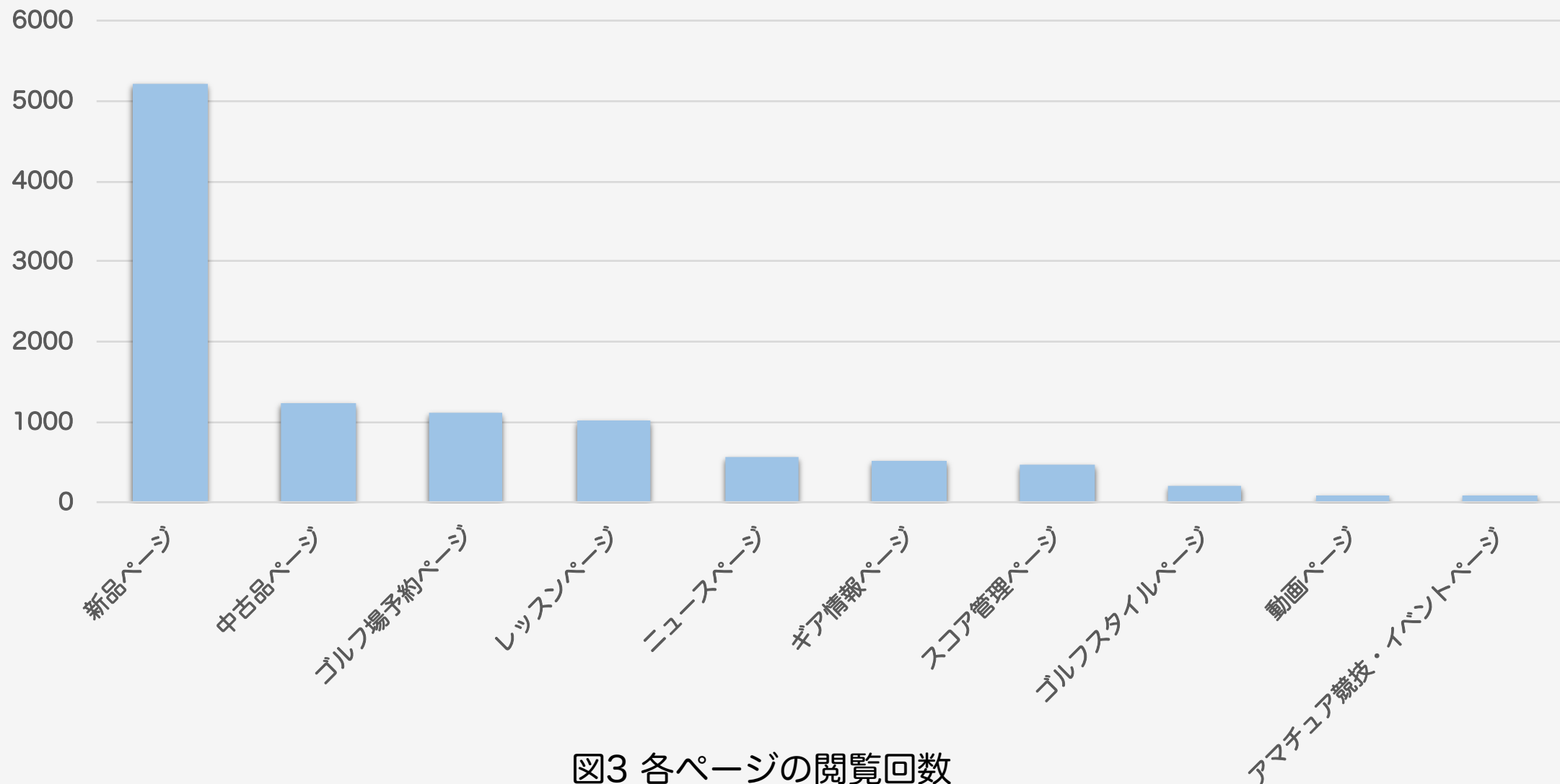
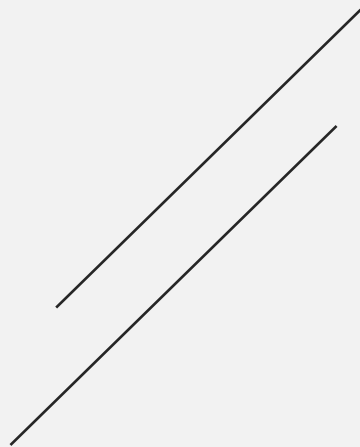


図3 各ページの閲覧回数

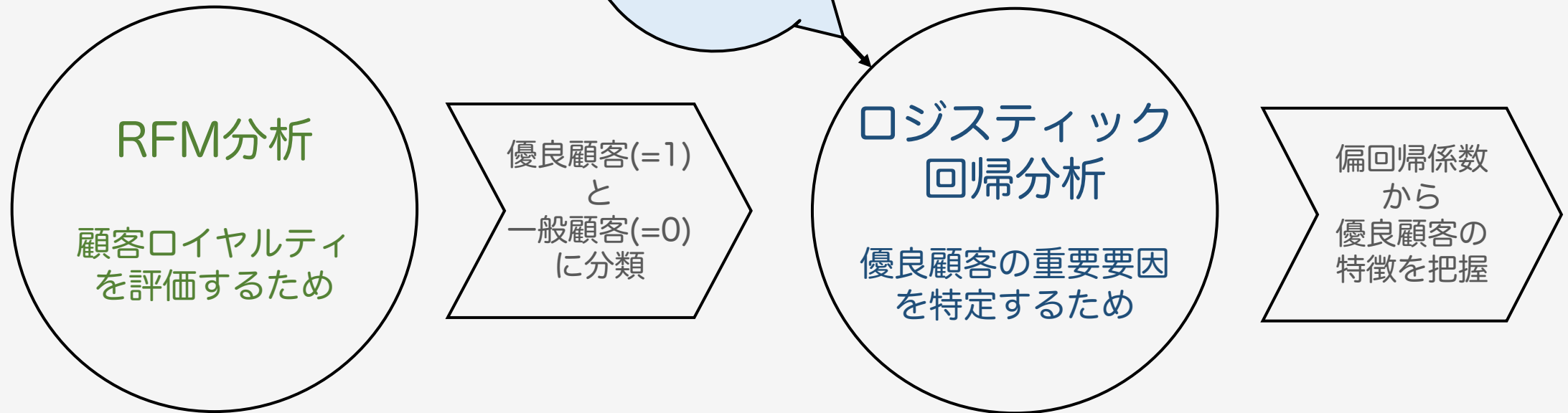


03

## 分析

- 分析手法
    - RFM分析
    - ロジスティック回帰分析
  - 分析の流れ
  - 分析結果
- 

# ▶ 分析手法



初回購買の内容や探索行動に関する変数



初回購買日  
(ロジスティック  
回帰分析)



初回購買日の翌日から1年間 (RFM分析)

## ▶ RFM分析

- R、F、M とはそれぞれ

**R**ecency(最新の購買日)

**F**requency(頻度)

**M**onetary(購入総額)

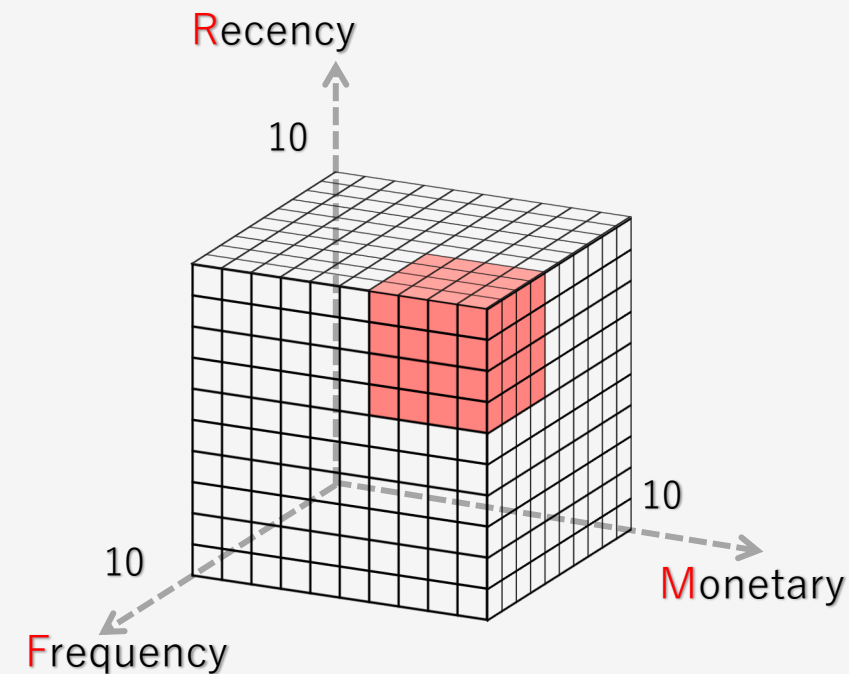
の略で、R、F、M の値を基に優良顧客を見つけるための分析を行うためのアイコンです。

- RFM 分析によって、顧客をランク付けした上でグループ化する。

## ▶ RFM分析

目的：顧客ロイヤルティを評価するため

- **R**ecency：直近の購買からの経過期間（日数）
- **F**requency：1年以内の  $\left\{ \begin{array}{l} \text{① 購買件数} \\ \text{② セッション数*} \end{array} \right.$
- **M**onetary：1年以内の総購買金額

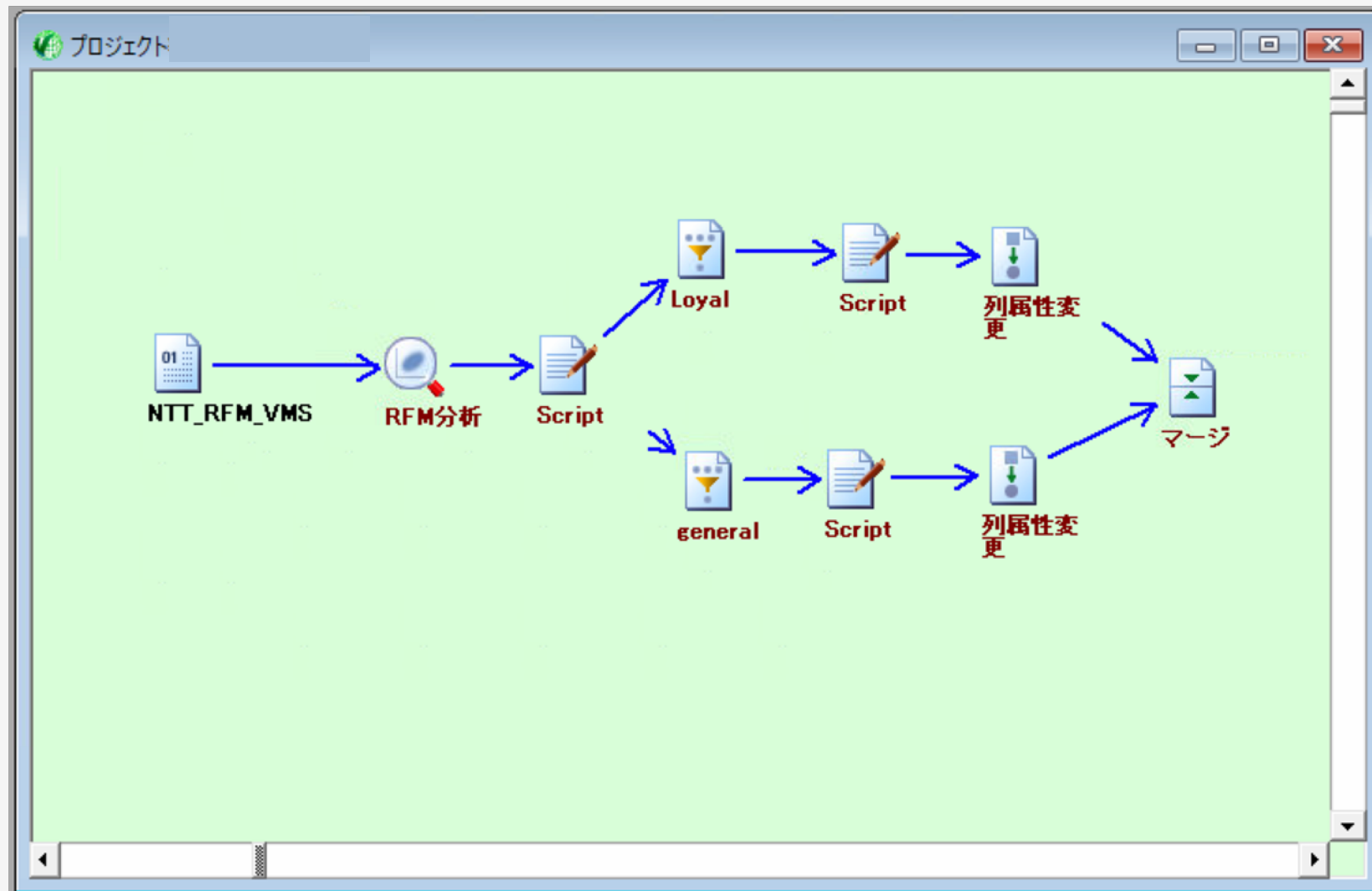


R,F,M三つの指標ごとに10等分によって、ランクを付けた。

本研究では、RFM分析で、Fの評価指標を①購買件数と②セッション数2つのパターンで行った。

\*セッション数はウェブページの閲覧回数

## ▶ RFM分析



株式会社NTTデータ数理システムから提供していただいた“Visual Mining Studio(VMS)”を使用。  
これにより分析を行う。

図2 VMSの例



## ▶ ロジスティック回帰分析

$$p_i = \frac{\exp\{\beta_0 + \sum_{i=1}^n \beta_i x_i\}}{1 + \exp\{\beta_0 + \sum_{i=1}^n \beta_i x_i\}}$$

$x_i$  : 説明変数とする  
 $\beta_i$  : パラメータとする

- 目的変数 : 優良顧客(=1), 一般顧客(=0)
- 説明変数 : 顧客の会員情報  
初回購買時の購買行動  
初回購買日のウェブ閲覧行動
- 変数選択 : ステップワイズ法
- 判別精度確認 : クロスバリデーション法(2-fold)

## ▶ ロジスティック回帰分析

- ロジスティック回帰分析において、説明変数が過剰の場合、回帰式の解釈が困難になる場合や、目的変数の予測の汎用性が下がる場合がある。そこで本研究では、モデルの選択基準として赤池情報量基準(AIC) によるステップワイズ法を使用した。
- モデルの判別精度を確認するために、ロジスティック回帰分析で使ったデータを2グループ (Group A, Group B) に分割し、クロスバリデーション法を行った。クロスバリデーション法は、テストデータが学習データに含まれないため、未知のデータに対する検知・分類がそのモデルでどの程度できているかを測る有効な検証法である。

## ▶ ロジスティック回帰分析

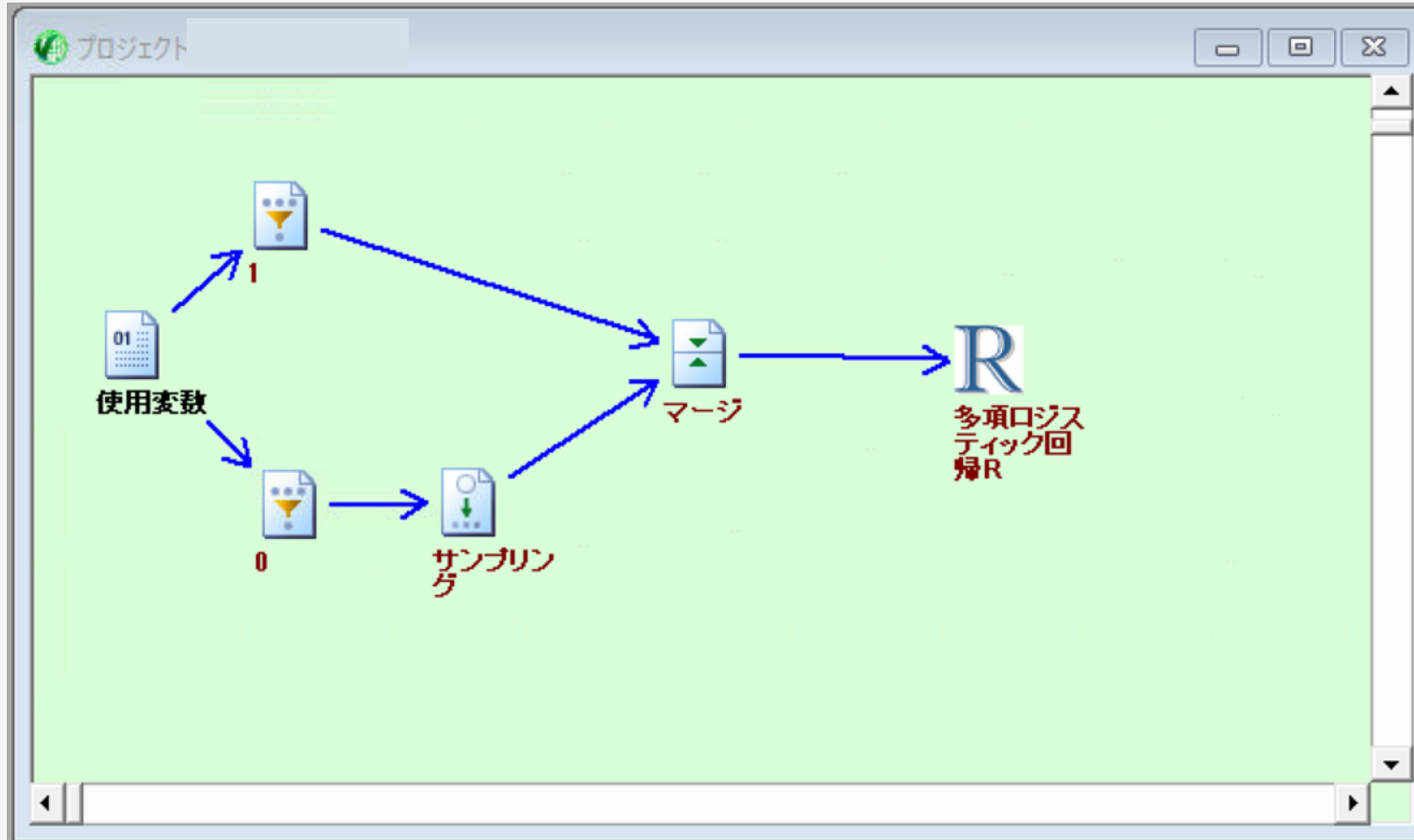


図3 VMSの例

# ▶ 使用変数



表1.1 モデル構築に使用した変数 (1)

		変数名	データ形式
目的変数		優良顧客(=1), 一般顧客(=0)	0,1
説明変数	会員情報変数	性別 (男=1, 女=0)	0,1
		年齢	整数
		首都圏(=1), 非首都圏(=0)	0,1
	購買行動変数	会員登録から注文までの日数	整数
		購買数量	整数
		平均売上金額	整数
		ウェア(Lady)の購買件数	0,1
		ウェア(Men)の購買件数	0,1
		クラブの購買件数	0,1
		用品・小物の購買件数	0,1
		その他の購買件数	0,1
		新品の購買件数	0,1
		中古商品の購買件数	0,1
		セール商品の購買件数	0,1



例)

user_id	seibetu	age	capital	date	amt	qty	mean_amt	goods	wear_Men
158778	1	50	0	0	23944	1	23944	0	0
159734	1	63	0	0	6078	2	3039	2	0
730617	1	41	1	40	95972	8	11996	0	0
737672	1	57	0	6	3920	1	3920	0	1
738826	1	44	1	0	29000	2	14500	0	2
740998	1	31	0	0	66741	3	22247	3	0
746543	1	41	1	11	4673	2	2336	0	0
764802	1	34	0	0	9034	2	4517	0	0
772381	1	38	1	0	14800	1	14800	0	1
790699	1	23	0	0	3946	2	1973	2	0
796152	1	67	1	0	5200	1	5200	0	1
796428	1	48	0	5	10000	1	10000	0	1
797970	1	52	1	0	6943	1	6943	0	0
799140	1	26	0	0	18858	1	18858	0	0
799149	0	35	1	0	6160	2	3080	2	0
799156	1	33	1	0	3511	1	3511	1	0
799184	1	39	1	12	12200	2	6100	0	2
799227	0	36	0	0	27400	3	9133	3	0
799260	1	65	0	14	3133	1	3133	0	0
799300	1	51	0	0	7912	1	7912	0	0

\*合計購買金額は購買数量との関係係数が高いため変数に使えなかった

## ▶ 使用変数



表1.2 モデル構築に使用した変数 (2)

		変数名	データ形式
説明変数	閲覧行動変数	セッション平均時間	整数
		セッション数	整数
		来訪区分(初来訪=0,再来訪=1)	0,1
		平均ページビュー数	整数
		レッスンページの閲覧有無	0,1
		ゴルフ場予約ページの閲覧有無	0,1
		動画ページの閲覧有無	0,1
		ニュースページの閲覧有無	0,1
		ゴルフスタイルページの閲覧有無	0,1
		新品ページの閲覧有無	0,1
		中古品ページの閲覧有無	0,1
		ギア情報ページの閲覧有無	0,1
		イベントページの閲覧有無	0,1
		スコア管理ページの閲覧有無	0,1

表1.3 モデル構築に使用した変数 (3)

		変数名	データ形式
説明変数	閲覧行動変数	直接アクセス	0,1
		メルマガ	0,1
		Yahoo!ニュースからの流入有無	0,1
		Yahoo!ニュース以外のYahoo!コンテンツからの流入有無	0,1
		Twitterからの流入有無	0,1
		Facebookからの流入有無	0,1
		外部イト_その他からの流入有無	0,1
		検索エンジン(SEO)_Googleからの流入有無	0,1
		検索エンジン(SEO)_Yahooからの流入有無	0,1
		検索エンジン(SEO)_その他からの流入有無	0,1
		検索エンジン(SEM)_Googleからの流入有無	0,1
		検索エンジン(SEM)_Yahooからの流入有無	0,1
		検索エンジン(SEM)_その他からの流入有無	0,1
		コンテンツマッチからの流入有無	0,1
		セッション切れの有無	0,1
		RSS経由の有無	0,1



## ▶ 分析の流れ

RFM分析で、Fの評価指標を①購買件数と②セッション数2つのパターンで行った。



\*セッション数はウェブページの閲覧回数

## ▶ ①RFM分析(F:購買件数の場合)

ランク	Rスコア 直近の購買からの経過時間	Fスコア 1年以内の購買件数	Mスコア 1年以内の総購買金額
10	0 ~ 11日	16 ~	110486 ~
9	12 ~ 30日	10 ~ 15件	63860 ~ 110480円
8	31 ~ 58日	7 ~ 9件	41000 ~ 63833円
7	59 ~ 95日	5 ~ 6件	28400 ~ 40890円
1~6	...	...	...

R,F,M各スコア7以上、合計スコア23以上の顧客を「優良顧客」とし、それ以外の顧客を「一般顧客」とした。

	人数(人)
優良顧客	1082
一般顧客	4383

# ▶ ①RFM分析(F:購買件数の場合)

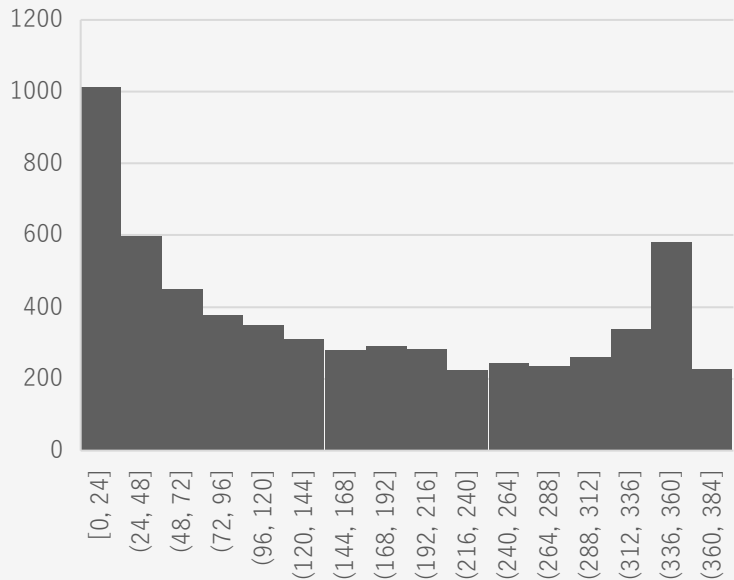


図4.1 R:直近の購買からの経過期間/日

各顧客の初回購買から1年後の日付けを基準とし、1年以内の最終的な購買は何日行われたか

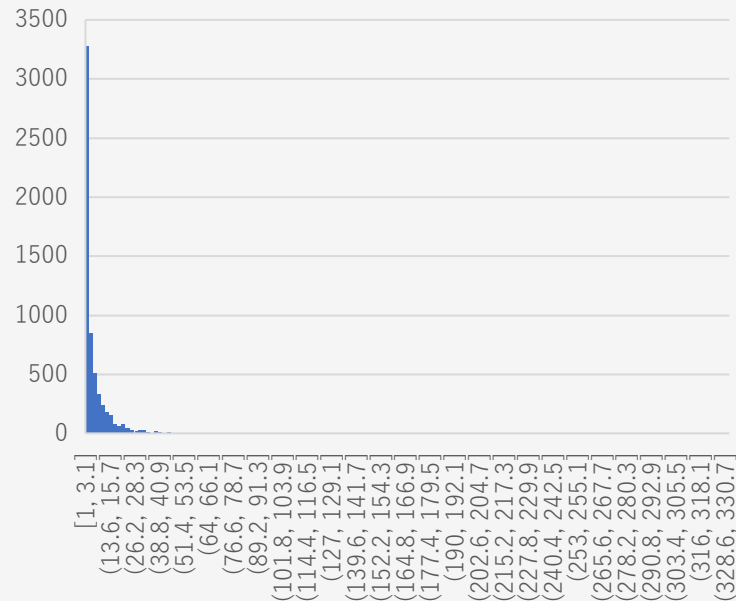


図4.2 F:1年以内の購買件数/件

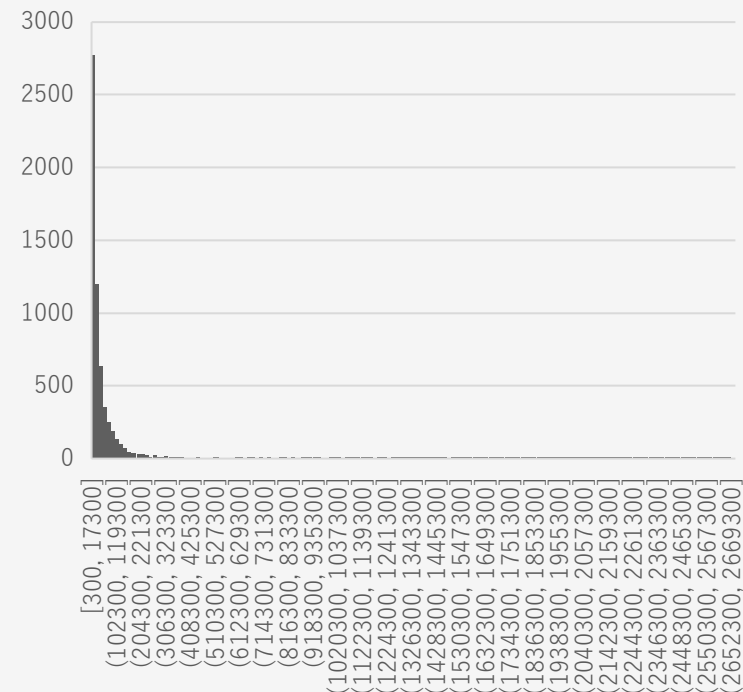


図4.3 M:総購買金額/円

- Rは「購買行動において、特徴的な購買期間が見られた：短期で再購買を行い、その後購買を行わない」
- Fは「ほとんどが1~3回しか買っていない顧客だが、頻度の多い人は極めて少ない」
- Mは「100万円以上を使った顧客もいたが、ほとんどが合計300~17300円使っていた」



## ▶ ①ロジスティック回帰分析結果(RFM分析でF:購買件数の場合)

2グループ (Group A, Group B) の一方に対して学習を行い, もう一方に対して判別精度の検証を行った結果:

	group.A	group.B	平均
正答率	84.2%	86.4%	85.3%

表2 選択された偏回帰係数の推定値

変数	係数
切片	-6.9370***
会員更新日	0.6248*
購買件数	0.3289***
平均購買金額	0.0001***
ウェア(Men)の購買有無	0.6600**
クラブの購買有無	0.8443***
用品・小物の購買有無	0.6662*
セール商品の購買有無	1.0930***
ゴルフスタイルページの閲覧有無	0.8528*

\*\*\*  $p < 0.001$  \*\*  $p < 0.01$  \*  $p < 0.05$ .  $p < 0.1$

- 会員更新を行う顧客、
- たくさんの商品を購入していた顧客、
- メンズウェア、クラブ、用品・小物、セール商品を購入していた顧客、
- ゴルフスタイルページを閲覧している顧客

は優良顧客になりやすいことがわかった。

## ▶ ②RFM分析(F:セッション回数)

ランク	Rスコア 直近の購買からの経過時間	Fスコア 1年以内のセッション回数	Mスコア 1年以内の総購買金額
10	0 ~ 11日	482 ~	110486 ~
9	12 ~ 30日	317 ~ 481回	63860 ~ 110480円
8	31 ~ 58日	224 ~ 316回	41000 ~ 63833円
7	59 ~ 95日	160 ~ 223回	28400 ~ 40890円
1~6	...	...	...

R,F,M各スコア7以上の顧客を「優良顧客」とし、それ以外の顧客を「一般顧客」とした。

	人数(人)
優良顧客	898
一般顧客	4567

## ▶ ②RFM分析(F:セッション数)

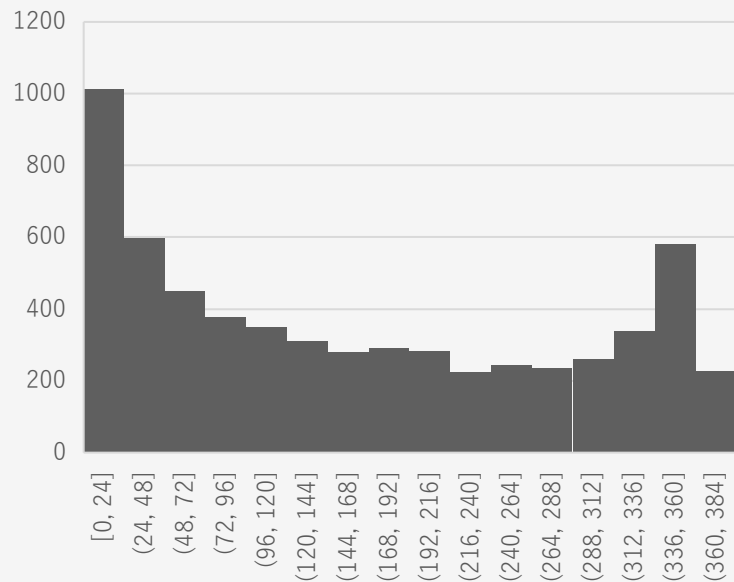


図5.1 R:直近の購買からの経過期間/日

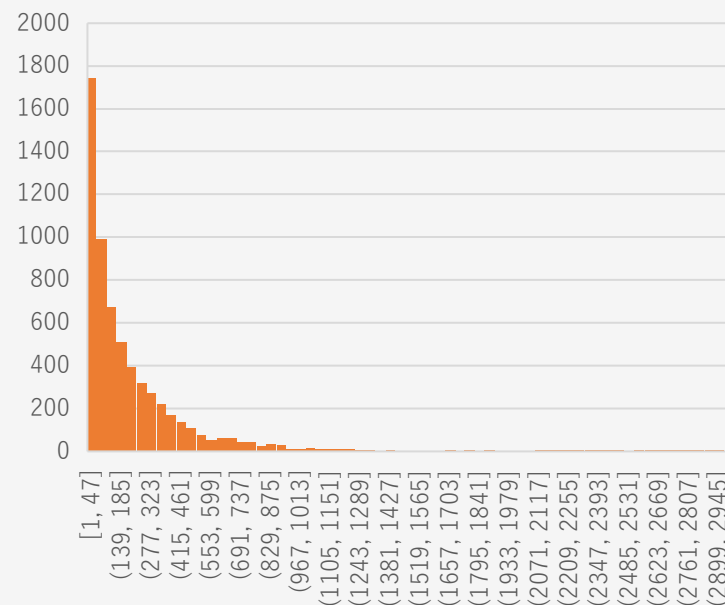


図5.2 F:1年以内のセッション数

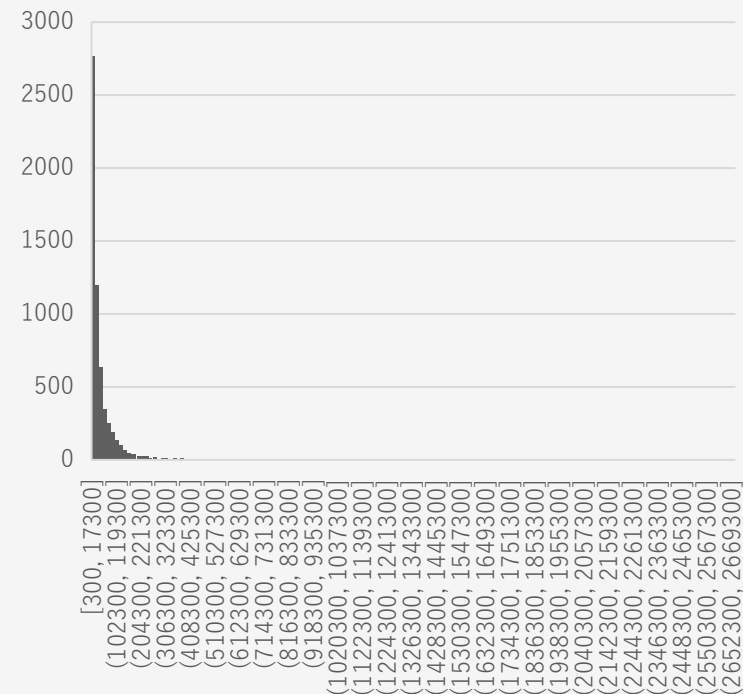


図5.3 M:総購買金額

- RとMはパターン① (F:購買件数) と同じ
- F「パターン① (F:購買件数) より、データの散らばりが大きくなった」

## ▶ ②ロジスティック回帰分析結果(RFM分析でF:購買件数の場合)

2グループ (Group A, Group B) の一方に対して学習を行い, もう一方に対して判別精度の検証を行った結果:

	group.A	group.B	平均
正答率	82.2%	80.1%	81.1%

表3 選択された偏回帰係数の推定値

変数	係数
切片	-19.730
会員更新日	0.986**
購買件数	0.209***
ウェア(Lady)の購買有無	0.559*
ウェア(Men)の購買有無	0.645**
クラブの購買有無	0.782***
用品・小物の購買有無	0.835**
直接アクセス	0.574**
メルマガからの流入有無	0.690***

- 会員更新を行う顧客、
- たくさんの商品を購入していた顧客、
- ウェア(Men)、ウェア(Lady)、クラブ、用品・小物を購入していた顧客、
- 直接アクセスした顧客
- メルマガから流入した顧客

は優良顧客になりやすいことがわかった。

## ▶ 結果の比較

- RFMでF指標の評価基準は、  
購買件数とセッション数の比較

- 結果で共通している変数は

会員更新日
購買件数
ウェア(Men)の購買有無
クラブの購買有無
用品・小物の購買有無

- 共通していない変数は

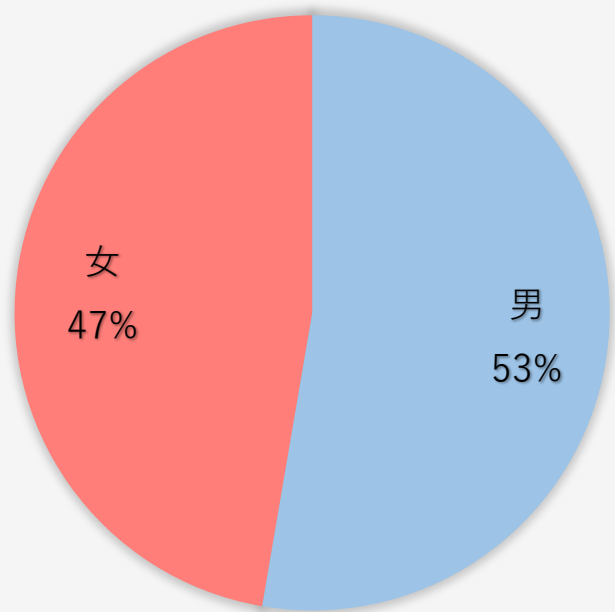
F: 購買件数	セール商品の購買有無
	ゴルフスタイルページの閲覧有無
F: セッション数	ウェア(Lady)の購買有無
	直接アクセス
	メルマガからの流入有無

## ▶ 考察

- F指標の評価基準に関係なく、優良顧客の特徴：
  - 会員更新を行う顧客
  - 初回購買で購買件数が多かった顧客
  - 特定の商品カテゴリーを最初に購買した顧客
- F:購買件数の基準で、優良顧客となる特徴：
  - セール商品を購買していた顧客
  - ゴルフスタイルページを閲覧していた顧客
- F:セッション数の基準で、優良顧客となる特徴：
  - ウェア(Lady)の購買を最初に購買した顧客
  - 直接ホームページをアクセスした顧客、つまりホームページをお気に入りに追加した顧客
  - メルマガをよくチェックしていた顧客

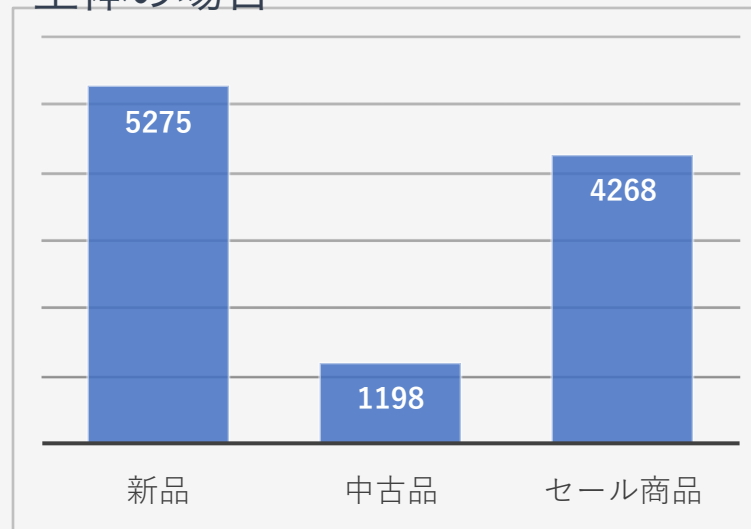
## ▶ 結果から元のデータに戻る

Fはセッション数の基準で、  
ウェア(Lady)を購入していた  
優良顧客の男女割合：

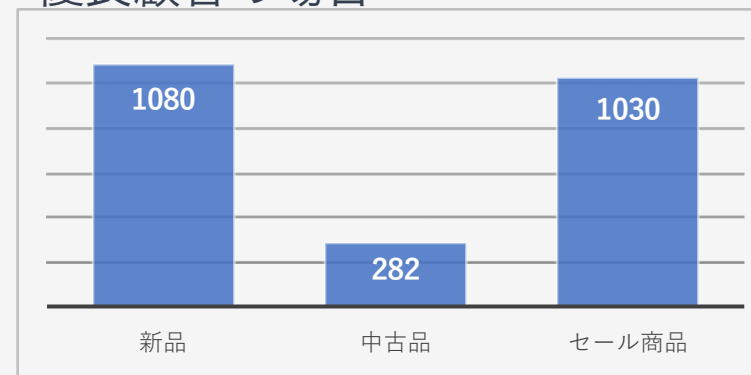


新品、中古品、セール商品を  
を購入していた顧客数：

全体の場合



Fは購買件数の基準で、  
優良顧客の場合



## ▶ 考察

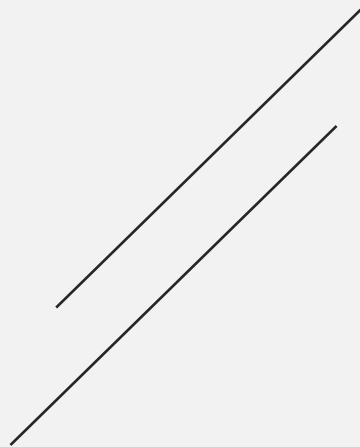
- 全体的に、女性の割合(19%)が低いが、ウェア(Lady)を購入している顧客の内男女の割合は近い、Fはセッション数で評価しているとき、ウェア(Lady)の購買を最初に購買した顧客は優良顧客になりやすいことから、女性顧客も大事だと考えられる。
- セール商品をよく購買していた顧客は、バーゲンハンターと思われるが、売上の拡大や収益の向上にも大切だと考えられる。





04

## まとめ

- まとめ
  - 今後の課題
- 

## ▶ 7.まとめと今後の課題

### まとめ

- 優良顧客を予測するモデルを構築した。
- ロイヤルティが高い顧客の初回購買行動を把握することができた。
- RFM分析でFの指標の違いによって、分析結果の比較も行った。

### 今後の課題

- 分析対象期間を延ばすことで、対象者を増やし、季節性も考えた上で、より精度の高い優良顧客予測モデルを構築できると考える。

## ▶ 参考文献

- 生田目崇, 鈴木元也, “ECサイトにおけるサイト閲覧行動と購買行動の分析,” 経営情報学会2013年春季全国研究発表大会要旨集, 274-277, 2013
- 山下遥, 鈴木秀男, “セール品に注目した顧客の購買行動の解析—2 値データのクラスタリングを考慮した ロジスティック回帰分析—,” オペレーションズ・リサーチ, 60 (2), 81-88, 2015
- 佐藤由将, 大竹恒平, 生田目崇, “ECサイトにおけるリピート顧客の特徴の分析,” 情報処理学会第79回全国大会講演論文集, 507-508, 2017
- 阿部誠. “RFM指標と顧客生涯価値:階層ベイズモデルを使った非契約型顧客関係管理における消費者行動,” 日本統計学会誌37 (2), 239-259, 2008