データサイエンスすいすい会

第12回「ECの売上データに対して自動機械学習を用いた顧客行動予測」

GRI

ForecastFlow

すいすい会の紹介

- データサイエンスの領域に関して知見を共有しあい、みんなで実践的な理解度を高めたい

- 自分の知見などを自由に発言してください
  （呼び水のテーマ紹介をします）

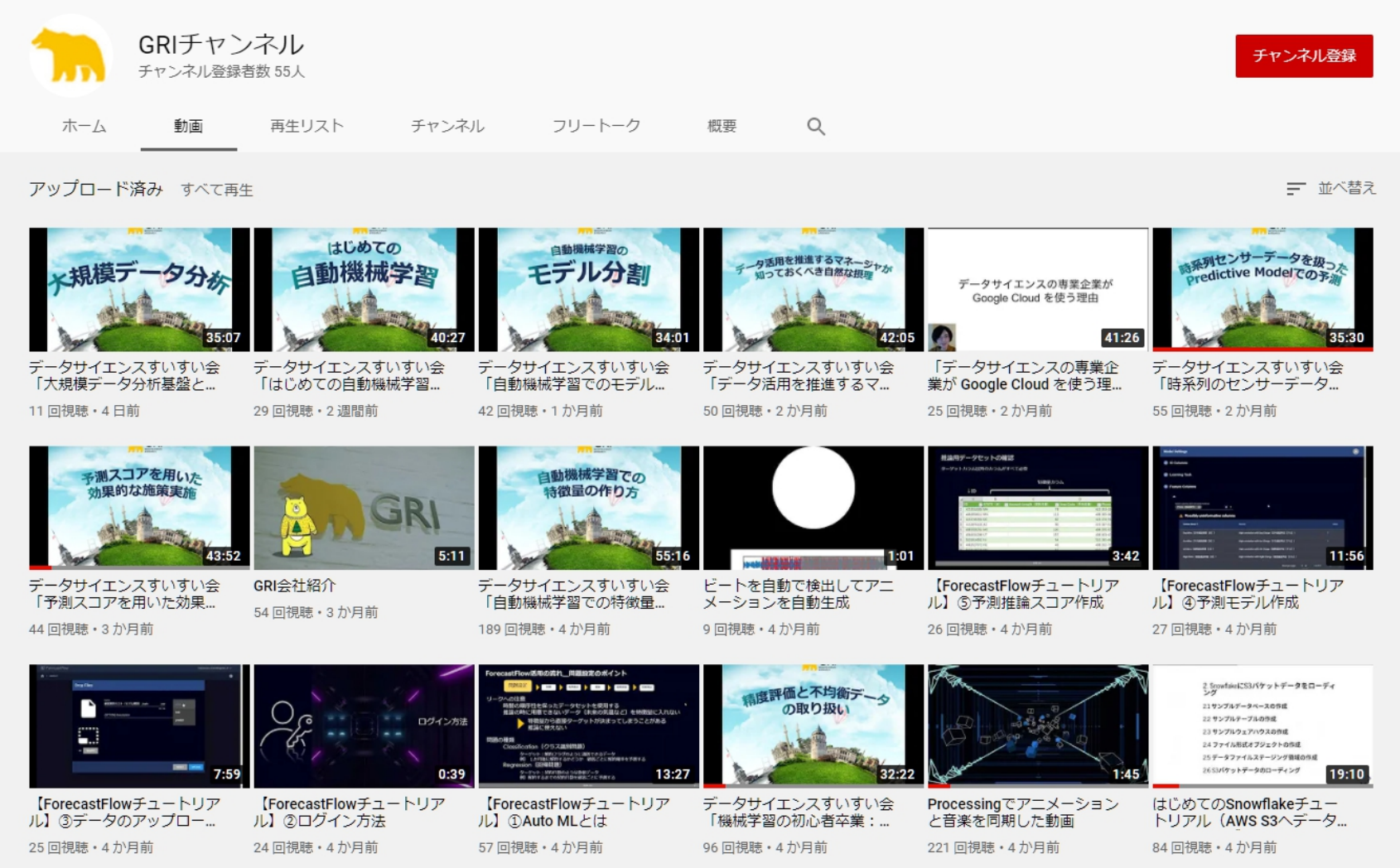
- しばらくの期間は、自動機械学習を中心にお話をさせてください

- 資料：　GRIホームページ　https://gri.jp/news/12924
- Slack　ForecastFlowチャンネル
  https://join.slack.com/t/forecastflowusers/shared_invite/enQtNTgyMjcxOTg0NzcxLTBkOWEzYWMwNDJmNTkyMDQzYmIxYWU0YWI4ZmU3ZDU0ZTMxNDUwODAxMWFmYmU1YjJiZGI0MjRhYWYyYTNlZTQ
- YouTube: GRIチャネル　https://www.youtube.com/channel/UCDVGqf-dgczYHMfPzpQ0jNQ

©GRI Inc.

過去のすいすい会の動画（YouTube）

https://www.youtube.com/channel/UCDVGqf-dgczYHMfPzpQ0jNQ/videos



本日のアジェンダ

- EC(eコマース)分野での自動機械学習を利用する典型的な分析テーマ例の整理
- あるイギリスのECの売上データを用いたForecastFlow（自動機械学習）のデモの共有
  - 売上データを用いた予測モデル構築から得られる施策提言の例を共有します
  - 自動機械学習とTableauの使いどころの違いを、予測モデル構築時の可視化分析（Tableau）やデータ前処理（Tableau Prep）の業務を通して説明します
  - 大学生が60時間かけると、”初見のデータに対してここまで実現できる”ということが実感できる内容になっております

©GRI Inc.

ECの特徴について再考

- 24時間 x 毎日、どこからでも買える
- 顧客の特定ができる、ウェブ滞在状況が分かる
- 実店舗に行く必要がないが配送料が必要
- 消費者は商品を手に取れない
- 気軽に店員に質問できない
  - 購入方法や商品に関する不明点
- 店主の心意気を店構えに表現するのが難しく、ぱっと見、お店の雰囲気を消費者に伝えにくい（画面で伝えられる情報量の制約）

©GRI Inc.

【商品視点】第１に品揃えや購買プロセスに関する分析

- 売上推移
  - 商品売上ランキング
  - デシル分析
  - RFM分析（Recency Frequency Monetary）
  - LTV分析
- アクセス解析
  - アクセス数、デバイス、ブラウザ、新規／リピータ、流入経路、キーワード、コンバージョン、SEO
- かご落ち分析
- 送料分析
- キャンペーン分析
- レコメンド
  - 関連性、新規性、偶発性、多様性

©GRI Inc.

【顧客視点】さらなる成長のため、優良顧客の育成と維持

- 実店舗と同じように、良い消費体験による、リピータの拡大が鍵

- 初回ユーザとリピータの違い
  - 購買金額
  - 購入率（サイト来訪時）
  - かご落ち
  - 返品
  - ユーザ獲得費用
  - 個人情報の登録

©GRI Inc.

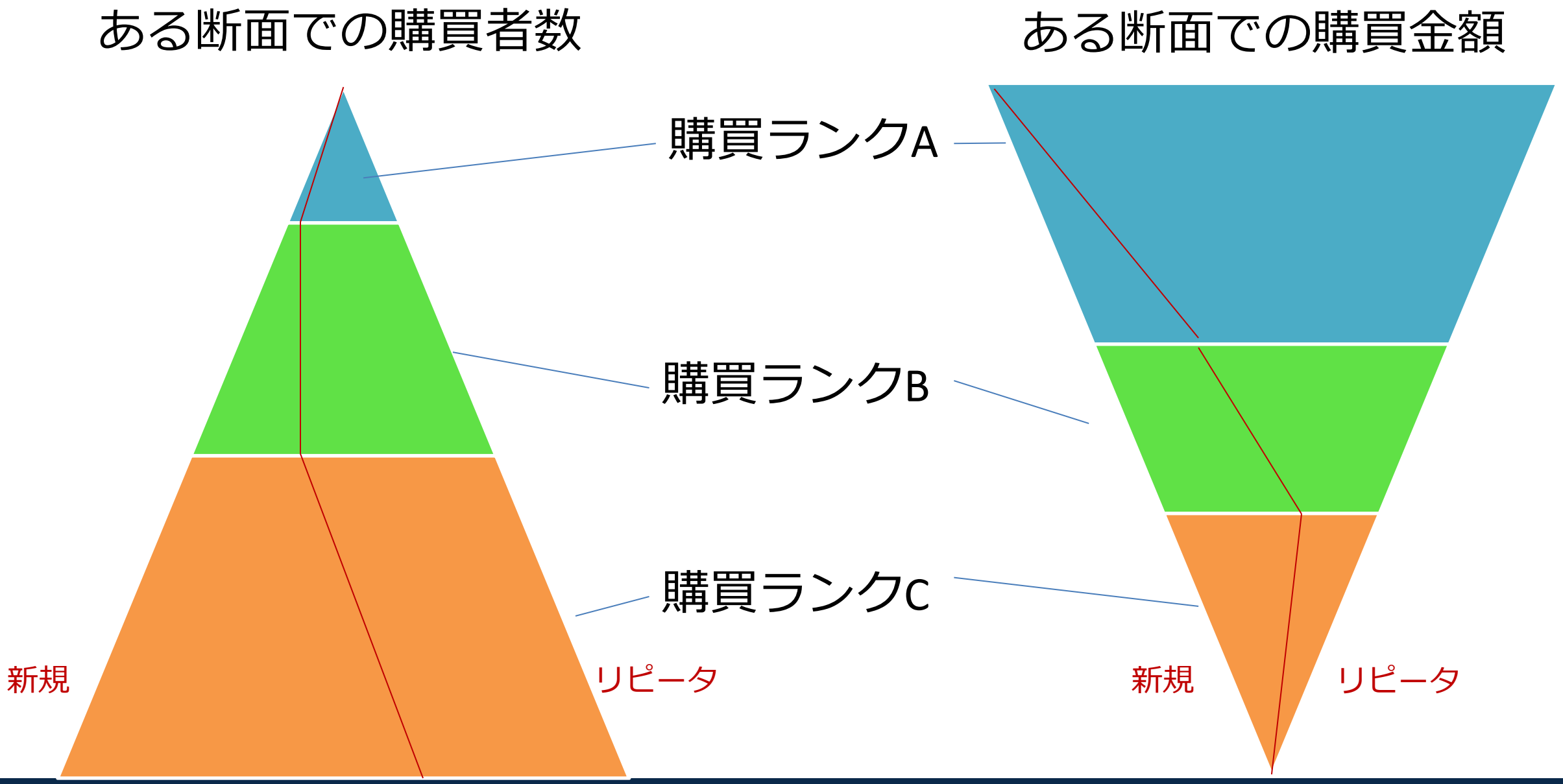
ECを成長させるためにチェックする典型的な指標

訪問数↑
会員登録数↑
会員獲得コスト↓

リピート率↑
優良顧客数↑

購入率↑
購入額↑
回遊率↑
滞在時間↓↑

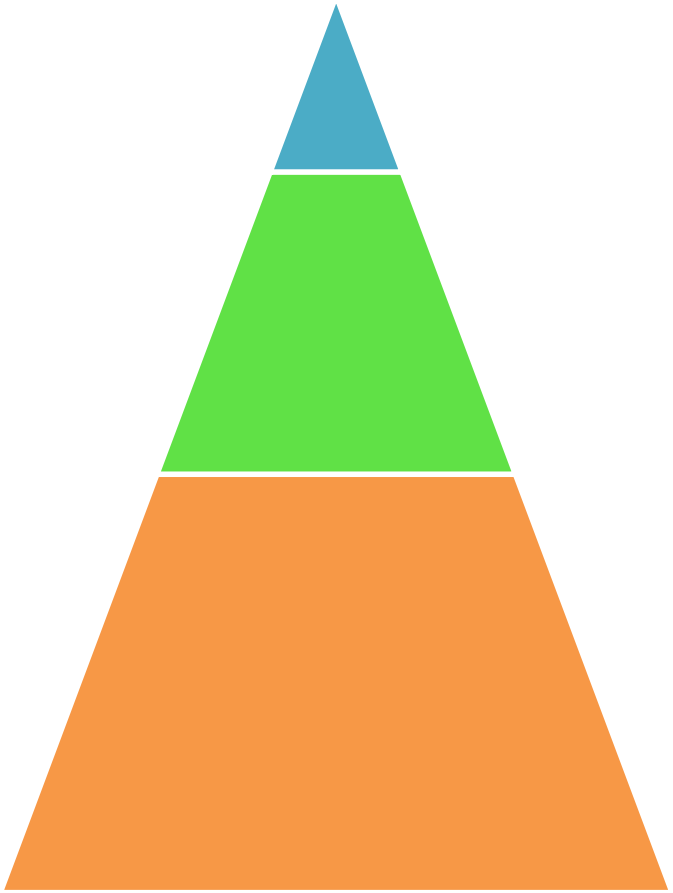
優良顧客（購買ランクA）の重要性

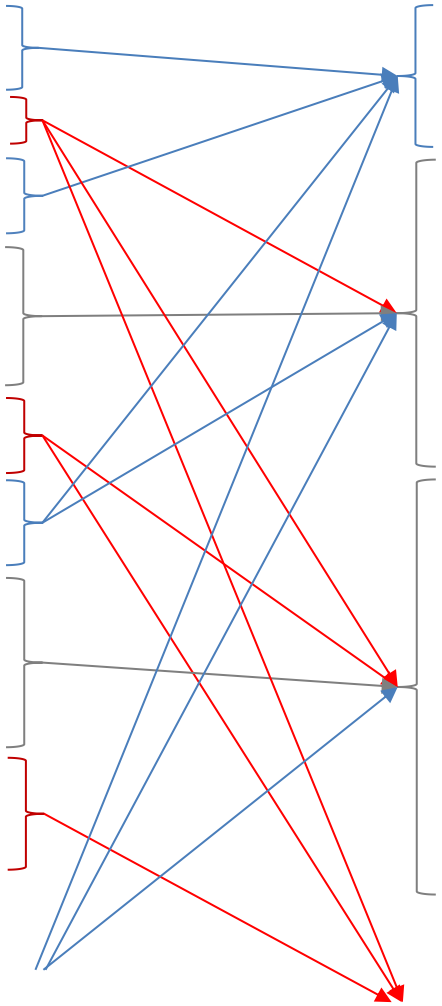


©GRI Inc.

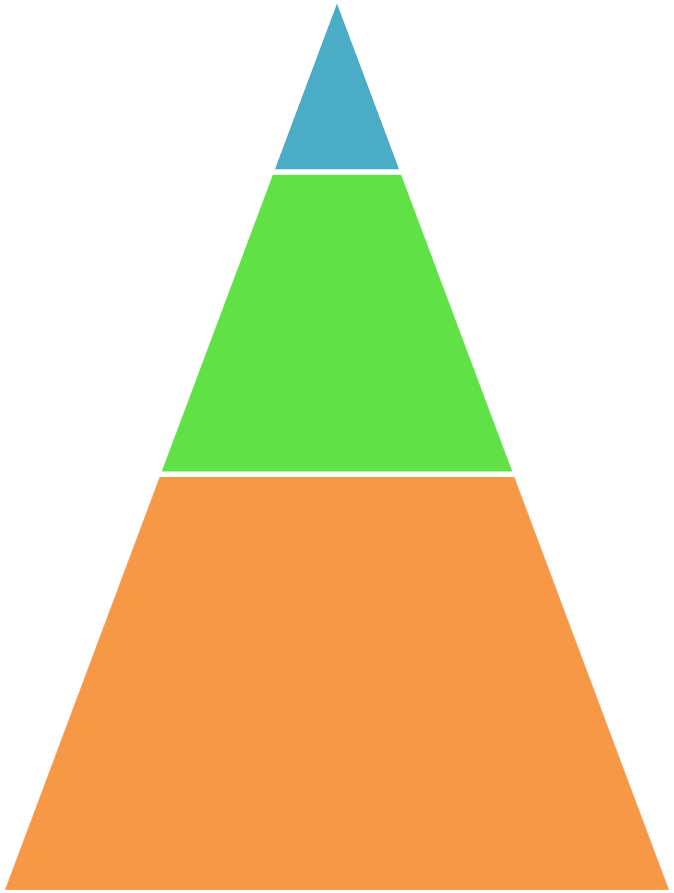
優良顧客の月次遷移

N月での購買ランク





N+1月での購買ランク



©GRI Inc.

自動機械学習で顧客の行動変容予測

問題設定（予測モデル）

- ターゲット：　　各ユーザのランクが変動するか否か（行動変容）
- 特徴量：
  - ユーザ属性（デモグラ）
  - 行動属性（買い方、ウェブ回遊の仕方）

ForecastFlow

【処方的アナリティクス】

- 自動機械学習で行動変容の違いを特徴量だけで説明できるか？
- どの特徴量がどの程度、説明力を持っているか？

【予期的推論】

- ユーザに望ましい行動をしてもらうために、どのような施策を打つべきか？
- 施策に対する効果検証をより正しく行う

©GRI Inc.

EC分析のデモ

GRIでのインターンを通して福岡大の高松さんにECの分析を依頼

©GRI Inc.

Tableau × ForecastFlow

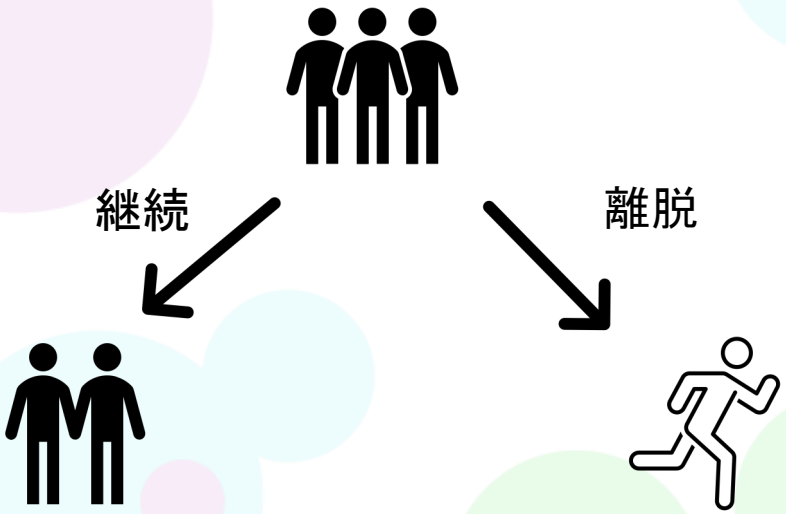
高松郁人

アジェンダ

- 自己紹介
- 今回やったこと
- データ整備
- 特徴量作成
- エラー
- 結果
- まとめ

今回やったこと

EC顧客の継続予測



目的

リピーターとそうでない人の
特徴の違いを明らかにする。

なぜ違い明らかにするか

リピーターは超重要だから！！

新規のお客様に初回購買してもらうまでには
長い道のりと大きなコストが必要。

しかし、リピーターはハードルが低い。
比較的コストをかけず
購買をしてくれる可能性が高い集団。

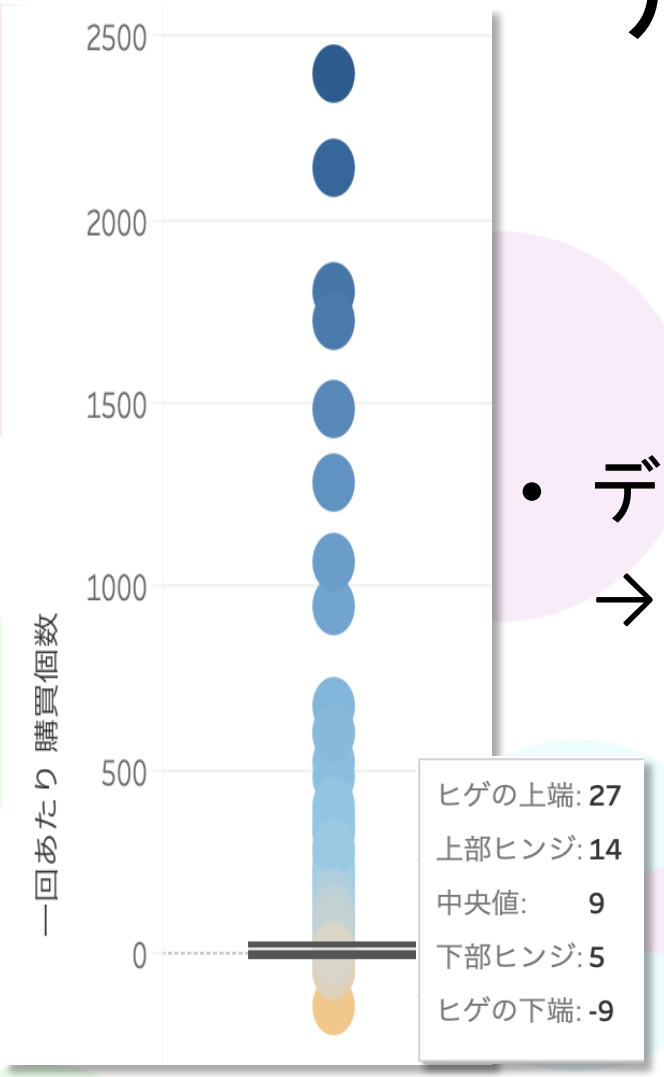
リピーターを予測する流れ

データ確認

データ概要

- イギリスの小物雑貨ECサイトのデータ
- 一行一商品の購買データ
- 406,829行
- 顧客は全員で4,372人
- 期間は 2010/12/01~2011/12/09 の<u>約一年間</u>

データ概要



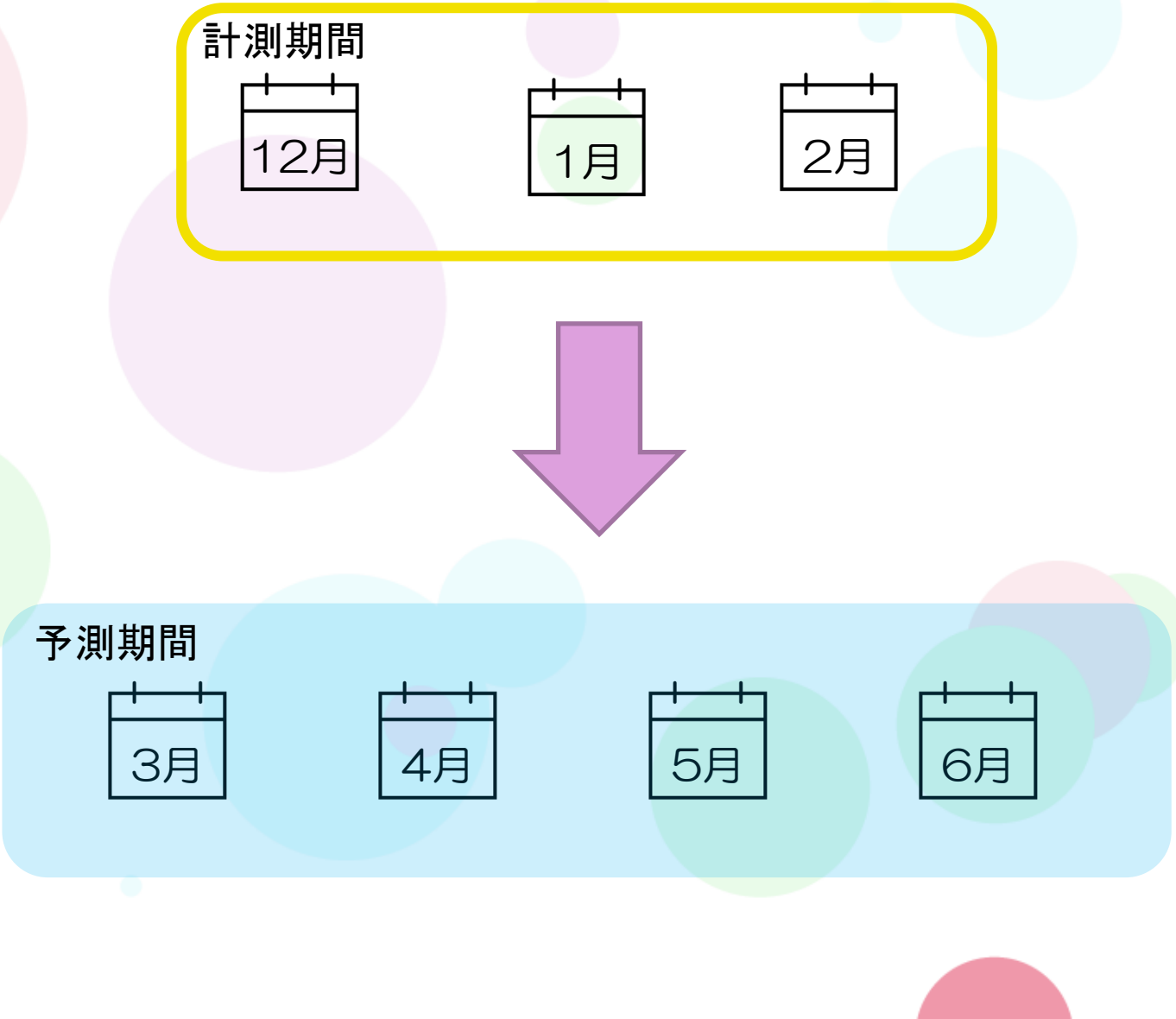
▲一購買あたりの個数

- データは BtoB 寄りのEC。
  →一回あたりの購買個数が多いため。

モデル構築

モデル

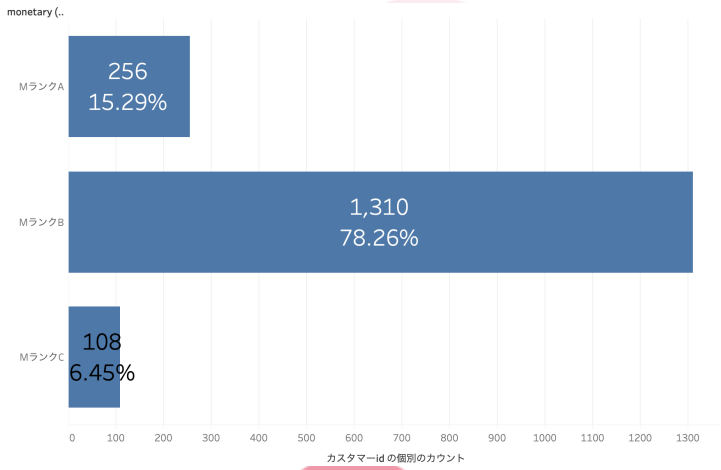
- 3ヶ月2010/12/01~2011/02/28の行動を計測。
- その後4ヶ月2011/03/01~2011/06/30を予測範囲として設定する。
- この4ヶ月以内に、一度でも購買があれば<u>継続</u>とする


計測期間
12月
1月
2月
予測期間
3月
4月
5月
6月

母集団定義

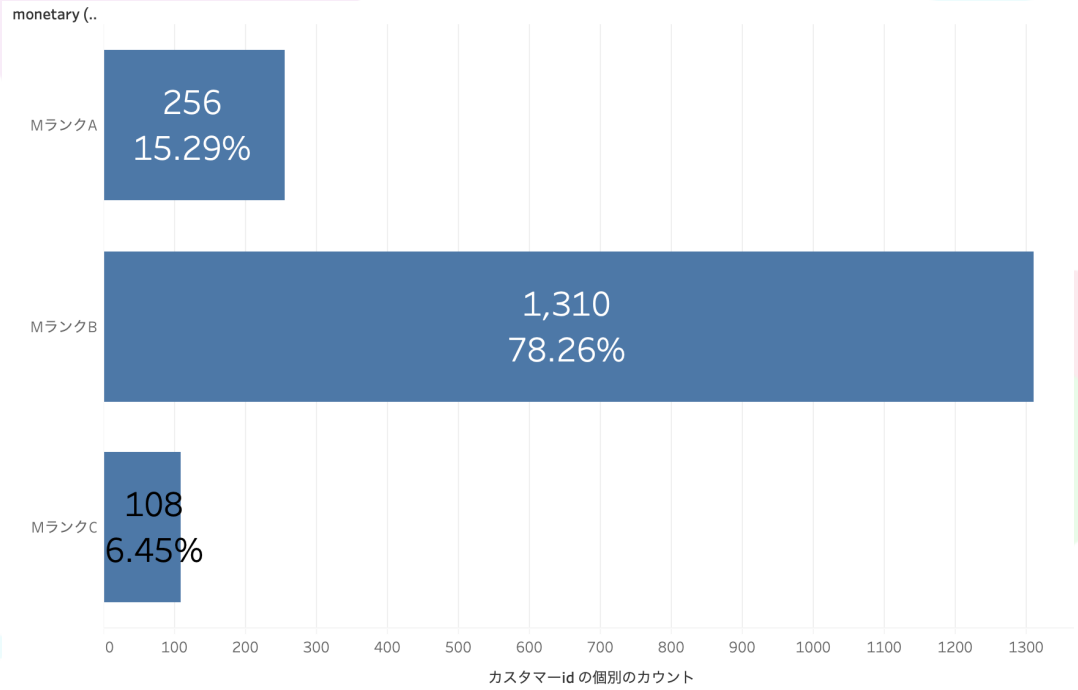
- 計測期間にデータを絞り、RFM分析を行う。

- Monetary15,000～160,000円に
該当する人（MのランクBとする）が
1,310人いることをTableauで確認

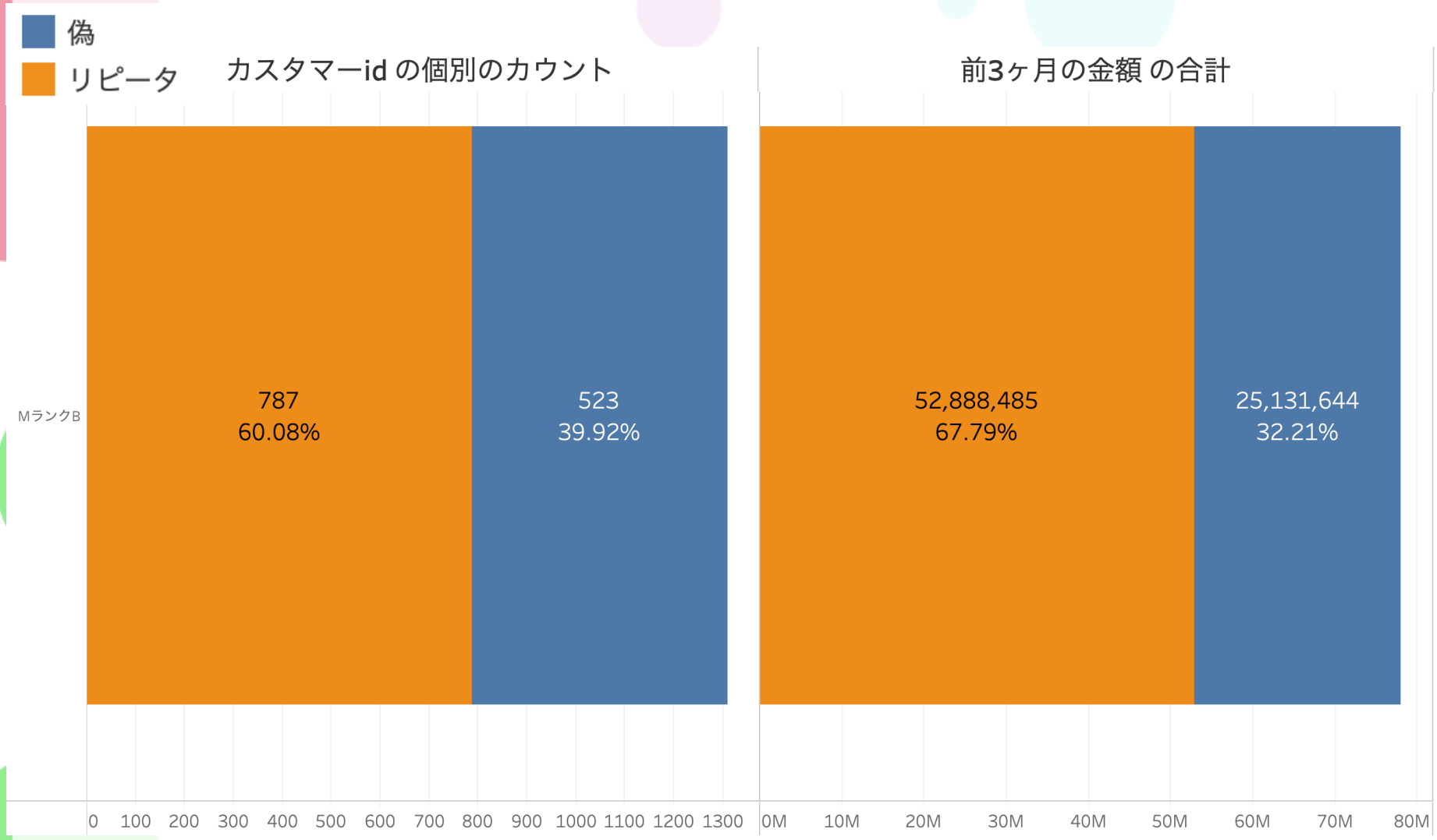


母集団定義

- Tableau PrepでMの ランクBの
  1,310人を母集団として抽出。



MのランクBが継続した割合



可視化から

- MのランクB 1,310人中、787人がリピートすることがわかった。(リピート率60%)

このリピーターを予測！

もう一度モデルの確認

モデル

- 母集団1,310人。
- 定義は期間内に15,000~160,000円購入者
- 期間2010/12/01~2011/02/28の行動を計測。
- その後4ヶ月2011/03/01~2011/06/30を予測範囲として設定する。
- 継続（リピート）したかどうかを予測する。

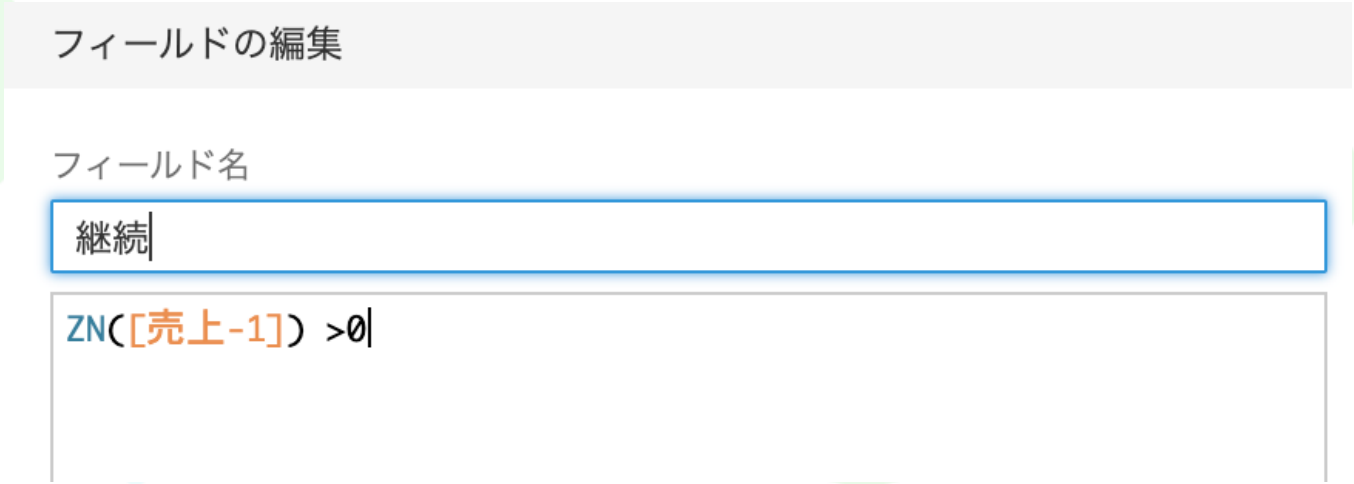
目的

リピーターとそうでない人の
特徴の違いを明らかにする。

Tableau Prepで変数作成

目的変数(Target)

- 継続（予測期間内に再び購買したか否か）
- 真/偽のブール型としてフィールド作成



特徴量作成

①frequency

②recency

③monetary

④請求時間朝、昼、夜

⑤購買単価

⑥初回購買までの日数

⑦カテゴリごとの購買回数

①frequency

- 購買頻度のこと。
- 一人の顧客が2010/12/01～2011/2/28の間に何回購入したか。
- 個数ではなく回数。

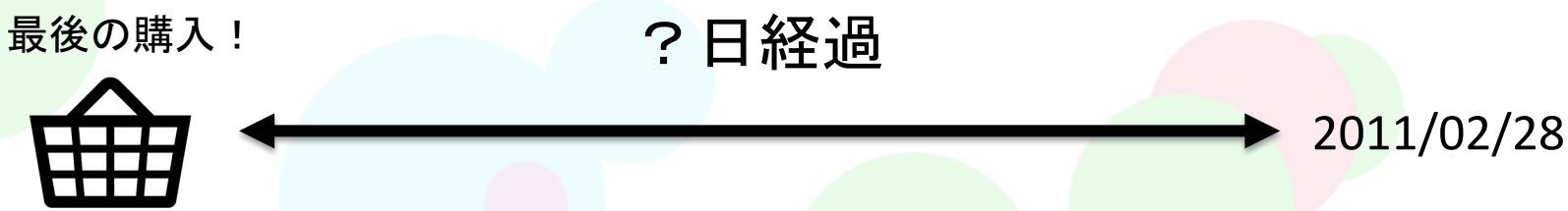
各月の平均購買回数

請求日時 の月

2010年12月	2011年1月	2011年2月
3.1013	3.3321	3.2927

②Recency

- 最終購買から 2011/02/28 までの日数。
- 空白の日数がわかる。



③monetary

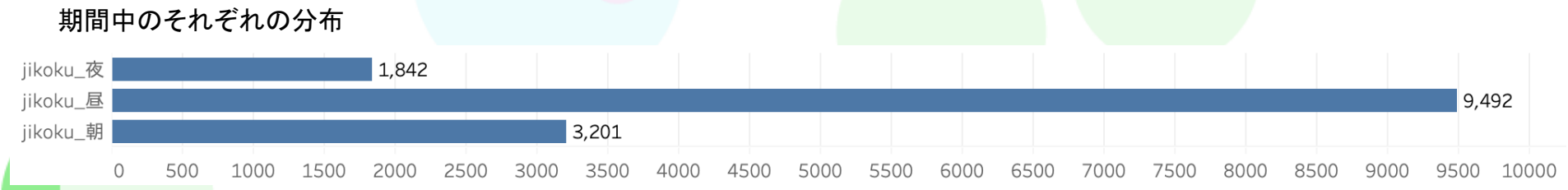
- 購買金額のこと。
- 期間は2010/12/01～2011/2/28。

各月の平均購買金額

2010年12月	2011年1月	2011年2月
156,108	180,139	178,223

④購買回数 朝、昼、夜

- 請求時間6時～20時を三つに区切る
- 6～10時台が“朝”
- 11～15時台が“昼”
- 16～20時台が“夜”
- それぞれの時間帯の購買回数をカウント。
- Frequencyを細かく分けたもの。



⑤購買単価

- 一人の一購買あたりの単価を求める。
- =一人の売上の合計÷一人の購買回数合計。
- 期間は2010/12/01～2011/2/28

全体の購買単価_平均	43,748

⑥計測期間内の初回購買までの日数

- 2010/12/01～そのユーザーが購買するまでにかかった日数。

Tableau Prep

グループ化したフィールド

タイプ	グループのタイプ	フィールド名
Abc	グループ	カスタマーID

集計フィールド

タイプ	集計	フィールド名
#	SUM	売上
#	MIN	recency
Abc	MAX	カスタマー国籍
#	CNTD	frequency
📅	MIN	初回請求日時
#	SUM	購買個数
#	AVG	商品単価（ポンド）

フィールド名

20101201~初回請求までの経過日数

```
DATEDIFF('day',#2010-12-01#,[初回請求日時])
```

⑦カテゴリごとの購買回数

- 27種のカテゴリを列にする（アンピボット）
- それぞれ購買があれば1がつく。
- 一個購買につき1がつく。

カスタマーID	CUSHION	HANGER	FRAME	STAND	CARD	BUNTING	LANTERN	WALLET	DRAWER	SCALE	MUG	BANK	DOORMAT	CHRISTMAS
18283	0	0	0	0	0	0	0	0	0	0	8	3	0	0
18269	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18260	0	0	0	3	0	1	0	0	0	0	0	0	0	0
18259	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18257	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18250	0	0	0	1	0	0	0	0	0	0	0	0	0	0
18245	0	0	1	0	5	1	0	0	0	0	0	0	1	2
18239	1	0	0	1	3	1	3	0	0	0	0	2	0	11
18233	0	0	0	0	0	0	0	0	1	0	0	0	0	0
18231	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18230	0	1	1	0	0	0	0	0	0	0	0	0	1	0

以上の作成した特徴量＋@を一行一人の表にして出力

カスタマーID	最も購買が多い時間帯	請求時間__朝(6~10)	請求時間__昼（11~15）	請求時間__夜（16~20）	1回あたり購買単価	20101201~初回請求までの経過日数	recency	継続	monetary	カスタマー国籍	frequency	購買個数	CUSHION
12348	不規則	23	0	23	78416.8	15	34	真	156833.6	Finland	2	1855	0
12350	夕	0	0	17	46816	63	26	偽	46816	Norway	1	197	0
12352	昼	0	15	0	41510	77	12	真	41510	Norway	1	98	0
12361	昼	0	10	0	26586	86	3	偽	26586	Belgium	1	91	0
12362	午前	29	29	0	33194	78	4	真	66388	Belgium	2	227	0
12365	昼	0	69	0	14965.533333333	82	7	偽	44896.6	Cyprus	3	173	0
12372	昼	0	20	0	55890.8	77	12	真	55890.8	Denmark	1	224	0
12373	昼	0	14	0	51044	62	27	偽	51044	Austria	1	197	0
12386	午前	10	10	0	28133	7	53	偽	56266	Australia	2	354	0
12388	昼	0	50	0	38874.5	47	21	真	77749	Australia	2	241	0
12393	午前	22	22	0	44765	41	45	真	89530	Australia	2	298	0
12395	昼	0	141	94	25021.64	2	10	真	125108.2	Belgium	5	1011	0
12402	昼	0	11	0	31584	50	39	偽	31584	Belgium	1	94	0
12410	朝	56	28	0	31783.733333333	65	17	偽	95351.2	Switzerland	3	209	0
12413	夕	26	0	52	19380.666666667	36	47	真	58142	France	3	176	0
12414	昼	0	7	0	21526.4	63	26	真	21526.4	Austria	1	145	0
12417	午前	30	30	30	27198.733333333	16	34	真	81596.2	Belgium	3	163	0
12421	昼	0	11	0	22043	54	35	真	22043	Spain	1	104	0
12422	朝	13	0	0	33390	49	40	真	33390	Australia	1	113	0
12423	朝	62	0	0	30116.8	20	7	真	60233.6	Belgium	2	382	0
12427	朝	26	0	0	17258.5	2	67	偽	34517	Germany	2	70	0
12437	昼	0	117	78	25989.88	42	5	真	129949.4	France	5	698	0
12441	朝	11	0	0	24297	7	82	偽	24297	France	1	121	0
12458	朝	19	0	0	61868.8	89	0	偽	61868.8	Switzerland	1	270	0
12462	昼	0	88	0	51891.7	70	10	偽	103783.4	Spain	2	283	1
12464	昼	0	18	0	25469.5	64	0	真	50939	Belgium	2	36	0
12468	昼	0	12	0	30882.6	57	32	真	30882.6	Germany	1	127	0
12474	昼	53	265	0	19696.366666667	5	3	真	118178.2	Germany	6	405	0
12476	昼	76	152	0	15506.866666667	12	14	真	93041.2	Germany	6	171	0
12480	昼	0	42	0	114275	37	52	真	114275	Germany	1	367	0
12483	昼	0	28	0	35529.2	34	55	真	71058.4	Sweden	2	292	0
12484	昼	0	98	0	159413.8	37	52	真	159413.8	Spain	1	630	0
12489	昼	0	15	0	46890.2	37	52	偽	46890.2	France	1	105	0
12490	朝	43	0	0	120990.8	44	45	真	120990.8	France	1	548	0
12494	午前	72	72	0	5630	8	6	真	39410	France	7	118	0
12500	午前	10	10	0	50324.4	40	11	真	100648.8	Germany	2	546	0
12502	朝	31	0	0	86415	64	25	真	86415	Spain	1	171	0
12507	朝	13	0	0	62462.4	75	14	偽	62462.4	Spain	1	219	0
12509	昼	0	7	0	24710	89	0	偽	24710	France	1	55	0
12510	昼	0	29	0	95347	48	41	偽	95347	Spain	1	294	0
12515	昼	0	16	0	53718	20	69	偽	53718	Italy	1	131	0
12523	午前	15	15	0	17244.5	42	7	真	34489	France	2	121	1
12524	朝	56	0	0	60391.1	40	7	真	120782.2	Germany	2	663	0
12527	午前	15	15	0	19944.4	48	10	偽	39888.8	Germany	2	221	0

ForecastFlowにて予測

結果の画面をお見せします

ForecastFlow

結果からわかったミス

請求時間__昼の計算式は間違っていた！

ForecastFlow

結果からわかったミス

請求時　なぜわかった？　っていた！

ForecastFlow

Frequency＞請求時間__昼のはず

＝全時間の購買回数

計算式が間違っていた

Frequency＞昼間の購買回数 のはず

＝全時間の購買回数

発見の一つ！

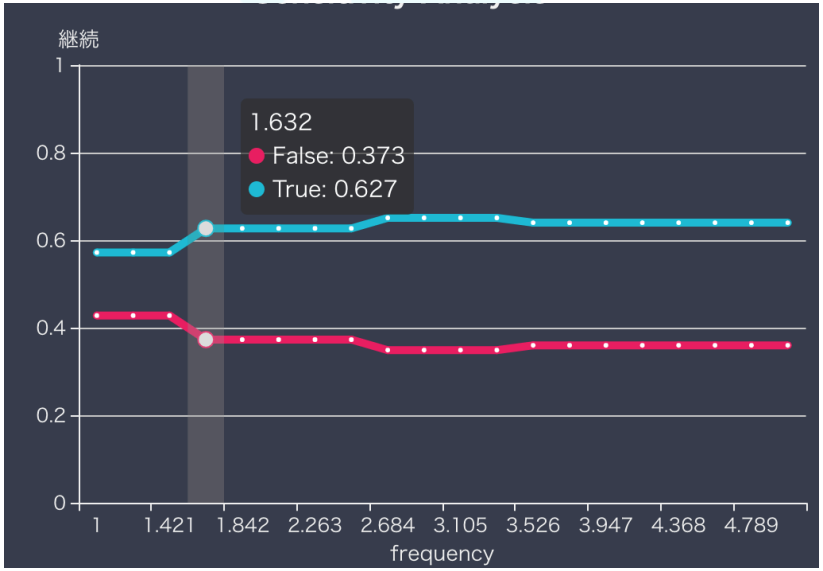
計算式が間違っていた

提案を3つ

これらの結果から提案①

Frequency（購買回数）

１～２回、２～３回に変化がある。



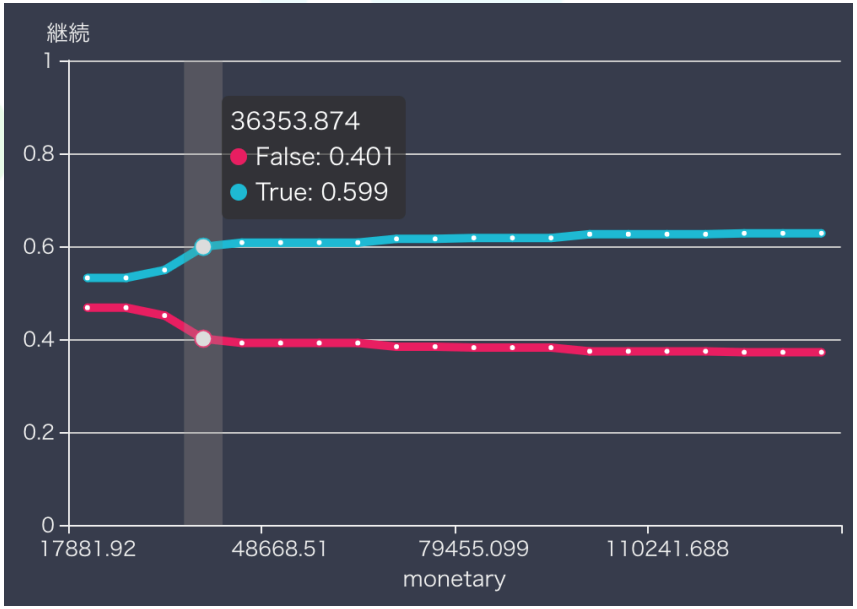
提案

初回利用後に
2回目、3回目に使えるクーポンを発行して、
3回目の利用まで誘導する。

これらの結果から提案②

Monetary（購買金額）

¥30,000~¥36,000の間に
大きな変化がある。



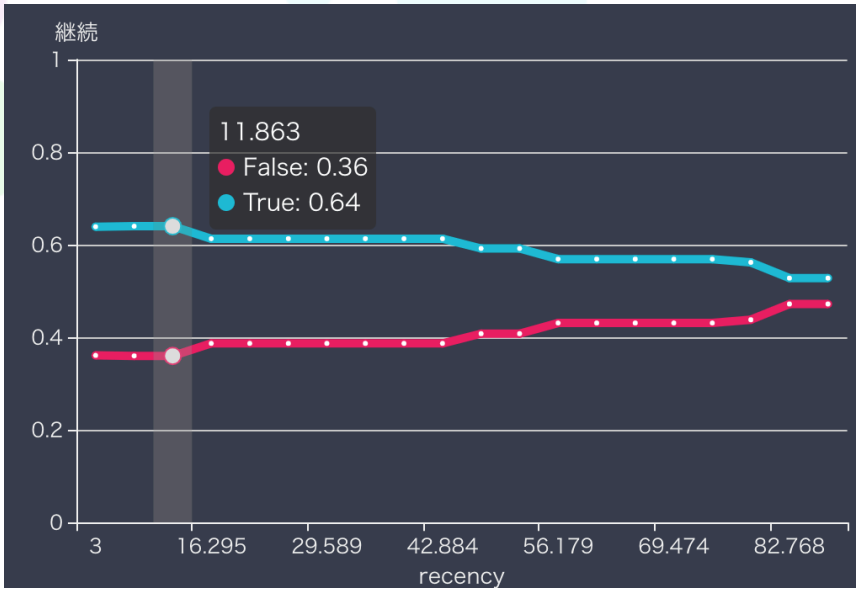
提案

3ヶ月以内に35,000円以上購入者は
会員ランクをプラチナとして<u>優遇</u>

これらの結果から提案③

Recency（離脱日数）

11~16日の間、42~60日の間に変化がある。



提案

- 2週間購買がないユーザーと
  2ヶ月間購買がないユーザーに対して通知を送る
- アプリではない場合、メールを送信

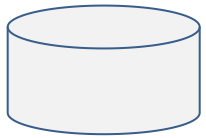
まとめ

- ミスからも発見があることを知った。
- これまで「なんとなく」で決めていたことが明確な根拠を持って決めることができる。
- 学生が60時間かければ、ForecastFlowを用いてデータに基づいた施策提言ができる。

ご静聴ありがとうございました

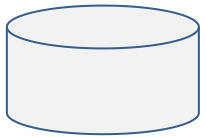
ECのデータで機械学習するために使うデータの典型例

受注実績／出荷実績



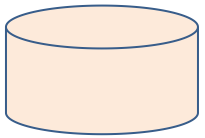
※ヘッダーと明細、返品、キャンセルの扱い

在庫データ

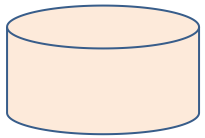


※在庫切れの状態

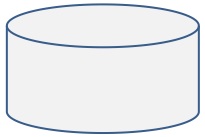
顧客マスタ



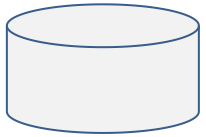
クーポンマスタ



ウェブアクセス

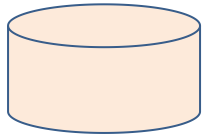


キャンペーン実績

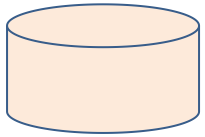


※ウェブ、TVCM、チラシ等

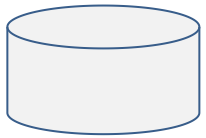
商品マスタ



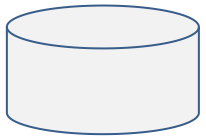
商品階層マスタ



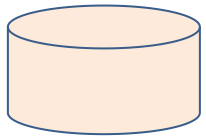
ポイント保有



ポイント利用

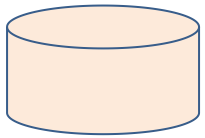


配送の価格表

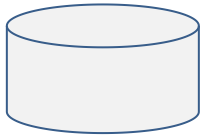


※配送業者、地域、サイズ、種別

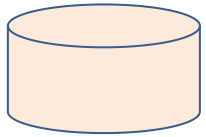
祝日マスタ



SNS



地域マスタ



©GRI Inc.

次回すいすい会の内容予定

- 21/3/10（水）12:00~
- 予測のサイクル（未来の施策と機械学習の実行タイミングの関係性）
- 特徴量エンジニアリング（特徴量のNullと0の違いの意味）

©GRI Inc.

残っている質問（今後、説明していきたい内容）

- その他のビジネスの実践例を知りたい
- どのくらい正解データの数があるべきか
- 未来の施策と機械学習の実行タイミングの関係性
- ダウンサンプリングとProbability Calibration
- 効果検証
- 様々な自動機械学習ツールの比較
- 様々なETL/ELTツールの比較
- Feature Store、分析基盤、CDPの説明

©GRI Inc.

無料トライアル・アカウントの配布

■ トライアル・アカウントの詳細ページ

- https://forecastflow.jp/free/

■ 申し込みメールを下記へ

- forecastflow@gri.jp

- ForecastFlow無料アカウント申請フォーム
  - お名前：
  - 所属名：
  - 部署名：
  - ForecastFlowで利用するメールアドレス：

©GRI Inc.