# データサイエンスすいすい会

## 「自動機械学習でのモデル分割」

GRI

ForecastFlow

# ナビゲータ

- 古幡征史(PhD)
- 株式会社GRI 取締役
- 経歴
  - 筑波大学　　第三学群社会工学
  - 筑波大学大学院　　経営・政策科学研究科修了(MS)
  - University of Western Sydney, PhD in Computer Science（コンピューターサイエンス博士号）
  - Universite Toulouse 1 Capitole, Doctorat en Informatique（コンピューターサイエンス博士号）

  - 丸文株式会社/KPMGコンサルティング/北陸先端科学技術大学院大学/University of Southern California/ドワンゴ
- 分野
  - データ分析＆シミュレーション
  - データ分析基盤構築
  - AI開発
  - BI構築
  - マーケティング、サプライチェーン、物流管理、配送管理、営業、経営、人事管理、財務管理、オークション、制度設計
  - 東京ガス、読売新聞、ソニーネットワークコミュニケーションズ、ドワンゴ、アミューズ、ジブラルタ生命保険、野村総研、東京証券取引所、スパークス、米国運輸省、米国海軍研究所、NASAを含む50社以上

# 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
- Slack　ForecastFlowチャンネル
- YouTube: GRIチャネル

©GRI Inc.

# 過去のすいすい会の動画（YouTube）

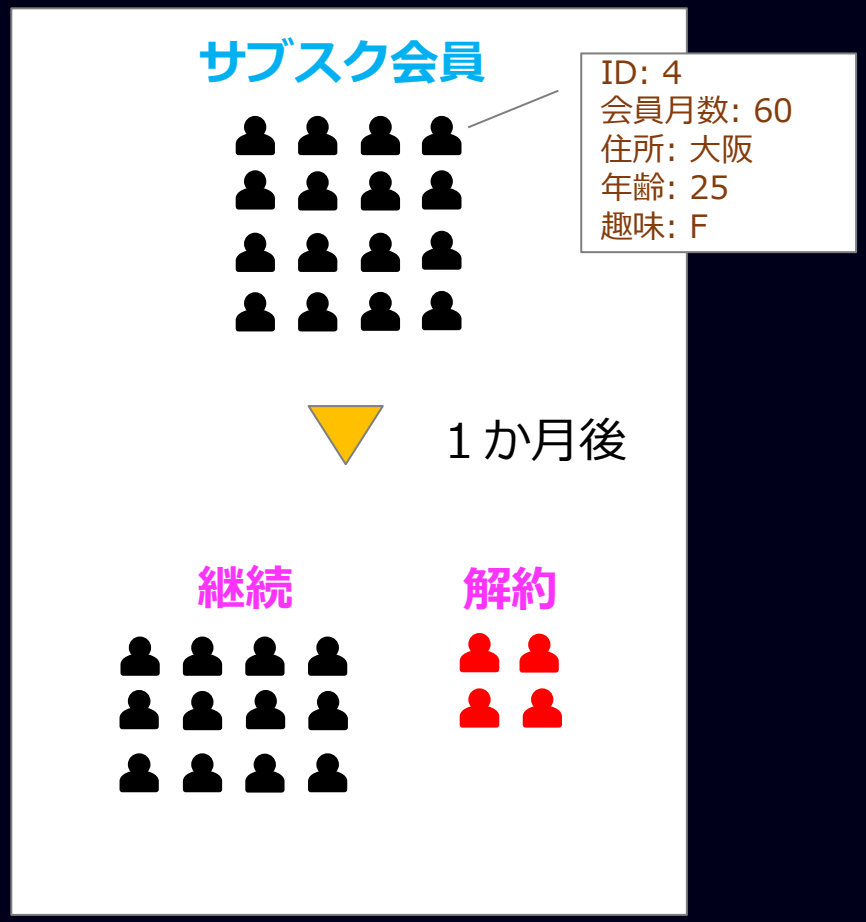
- Youtubeチャンネル：
  - https://www.youtube.com/channel/UCDVGqf-dgczYHMfPzpQ0jNQ/playlists

- 「ビジネスでAIを活用するための問題設定」
- 「アンケートと機械学習」
- 「分類問題の精度評価手法と不均衡データ」
- 「自動機械学習での特徴量の作り方」
- 「予測スコアを用いた効果的な施策実施」
- 「時系列センサーデータをを扱ったPredictive Modelでの予測」
- 「データ活用を推進するマネージャが知っておくべき自然な摂理」

©GRI Inc.

# 表形式の自動機械学習（AutoML）とは

■ 一枚表さえ用意したら、自動で機械学習をさせることができるシステム




機械学習：　ターゲットを特徴量だけで予測すること

| ID | ターゲット | 特徴量１ | 特徴量２ | 特徴量３ | 特徴量４ |
|---|---|---|---|---|---|
| 1 | Yes | 120 | 東京 | 18 | A |
| 2 | No | 80 | 千葉 | 18 | B |
| 3 | Yes | 95 | 神奈川 | 20 | B |
| 4 | Yes | 60 | 大阪 | 25 | F |
| 5 | Yes | 86 | 東京 | 26 | G |
| 6 | No | 92 | 埼玉 | 22 | C |
| 7 | No | 105 | 兵庫 | 15 | C |
| 8 | Yes | 72 | 埼玉 | 19 | D |
| 9 | Yes | 88 | 東京 | 28 | A |

©GRI Inc.

# 本日の話題：

- 予測モデルの分割とは

- ForecastFlowでのモデル分割の設定方法
  - 米国電話会社の解約データの例
  - ADK生活者総合調査での海外旅行ユーザの予測

©GRI Inc.

# 入力データセットと予測モデルの関係

■構築された予測モデルは、入力データセットの範囲で訓練を行った結果である

# 全体で予測モデルを作成する際の違和感がある例

- 性別
- 年代
- 地域
- サービス経験レベル（新規とロイヤルカスタマーの違い）
- 入会経路
- 支払方法
- キャンペーン有無
- 契約サービス種別

©GRI Inc.

# 入力データセットを分割し予測モデルを分割させる

■入力データセットの範囲を分割すれば、それぞれ予測モデルを構築できる

# ForecastFlow内の予測モデルの自動分割方法

- モデル構築時にSecondary IDを指定すると、そのカラム値に従い自動分割することができる

# ForecastFlow内の予測モデルの自動分割方法（デモの紹介）

■デモで実例

©GRI Inc.

# サブスクリプション・ビジネス（解約防止モデルの例）

■米国の電話会社の例

- ユーザの解約を説明できる予測モデルが欲しい
- 解約者数が急増した

# サブスクリプション・ビジネス（解約防止モデルの例）

| Name ↑ | Type |
|---|---|
| Account Length（契約月数） | NUM |
| Area Code（市外局番） | NUM |
| Churn?（解約した？） | CAT |
| CustServ Calls（カスタマーサービス問い合わせ回数） | NUM |
| Day Calls（日中通話回数） | NUM |
| Day Charge（日中通話料金【ドル】） | NUM |
| Day Mins（日中通話時間【分】） | NUM |
| Eve Calls（夕方通話回数） | NUM |
| Eve Charge（夕方通話料金【ドル】） | NUM |
| Eve Mins（夕方通話時間【分】） | NUM |
| ID | NUM |
| Int'l Plan（国際通話契約） | CAT |
| Intl Calls（国際通話回数） | NUM |
| Intl Charge（国際通話料金【ドル】） | NUM |
| Intl Mins（国際通話時間【分】） | NUM |
| Night Calls（夜間通話回数） | NUM |
| Night Charge（夜間通話料金【ドル】） | NUM |
| Night Mins（夜間通話時間【分】） | NUM |
| Phone（電話番号） | UNIQUE |
| STATE（州） | CAT |
| VMail Message（ボイスメールメッセージ数） | NUM |
| VMail Plan（ボイスメールプラン） | CAT |

■データ数：　3,333

■解約率：　14.5%

©GRI Inc.

# サブスクリプション・ビジネス（解約防止モデルの例）

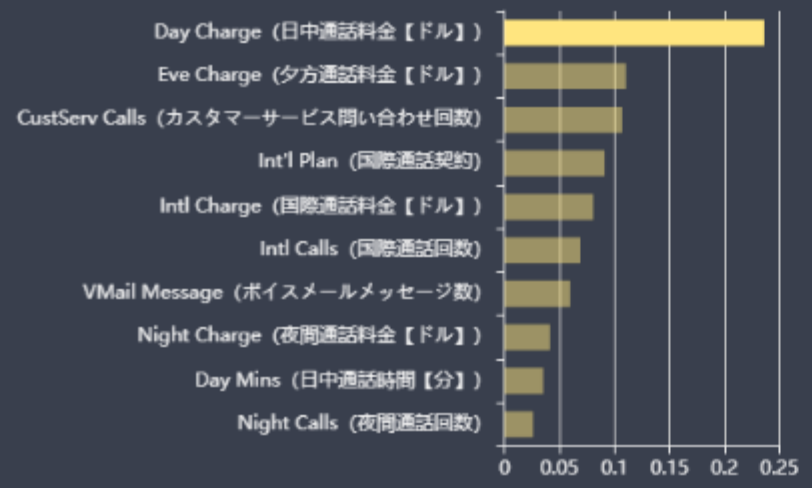
■全顧客

## Performance Report

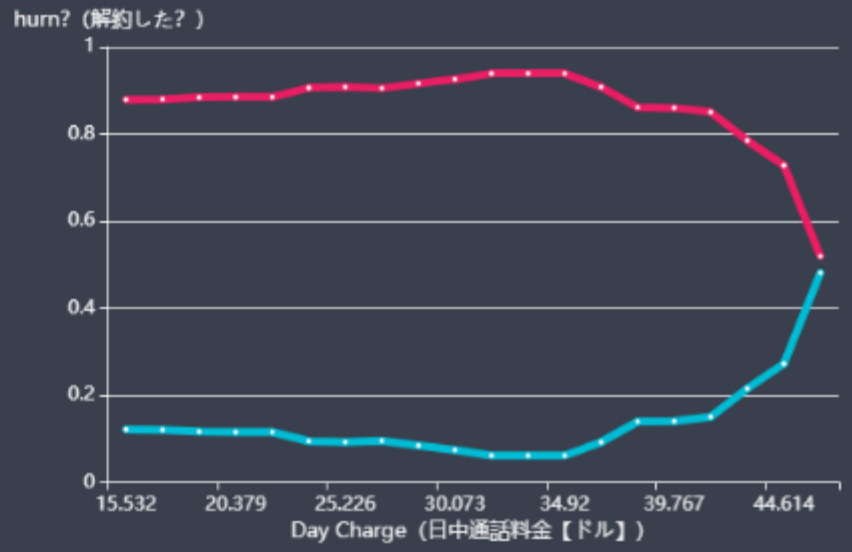
| Label Name | F1 Score | Precision Score | Recall Score | Support |
|---|---|---|---|---|
| False. | 0.968 | 0.947 | 0.989 | 855 |
| True. | 0.778 | 0.916 | 0.676 | 145 |

## Insight

### Feature Importance



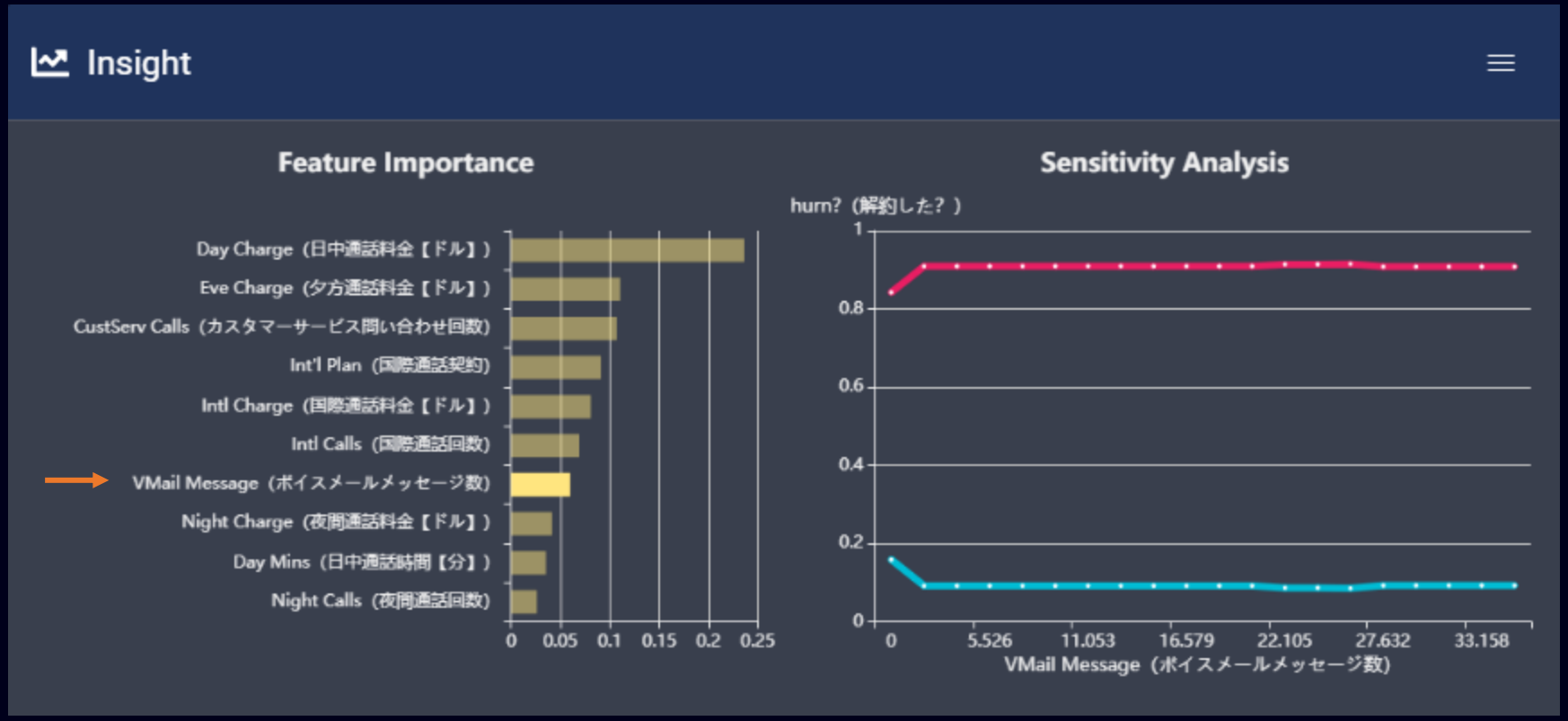

### Sensitivity Analysis




©GRI Inc.

# サブスクリプション・ビジネス（解約防止モデルの例）

■全顧客




ボイスメール契約：　922（契約率27.7%）
ボイスメール契約の顧客はロイヤルカスタマーに近い可能性が

# 予測モデル自動分割の指定方法

■Models > CREATE > Model Settings

モデル設定画面でSecondary IDを指定
（ボイスメールプランの値ごとにモデルを自動分割）

※カテゴリ値が15個までが指定可能

# 予測モデル自動分割の結果

■自動分割されたモデルの結果の確認

Models

Search

CREATE

| Name | Train Status | Desc | Type | Target | Metric | Created At | Actions |
|---|---|---|---|---|---|---|---|
| 解約予測（ボイスメールプラン分割）@ VMail Plan（ボイスメールプラン）= no | Completed | | Classification | Churn?（解約した？） | f1 | 2021-01-10 16:45:49 | |
| 解約予測（ボイスメールプラン分割）@ VMail Plan（ボイスメールプラン）= yes | Completed | | Classification | Churn?（解約した？） | f1 | 2021-01-10 16:45:49 | |

モデル名 @サブモデル名

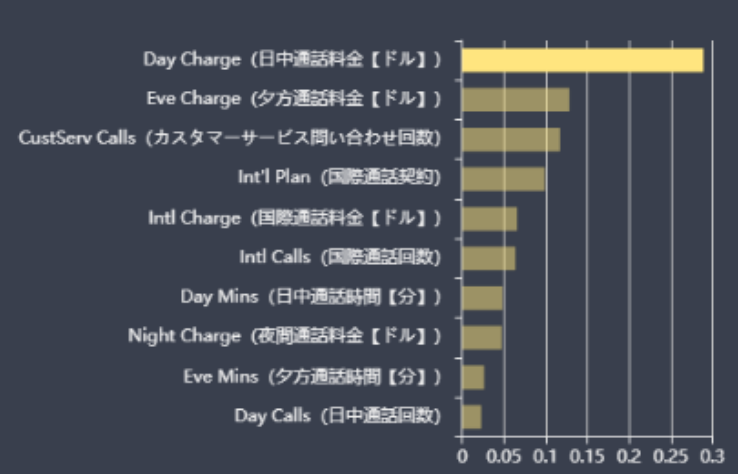
# 予測モデルの結果（ボイスメールに契約していない顧客）

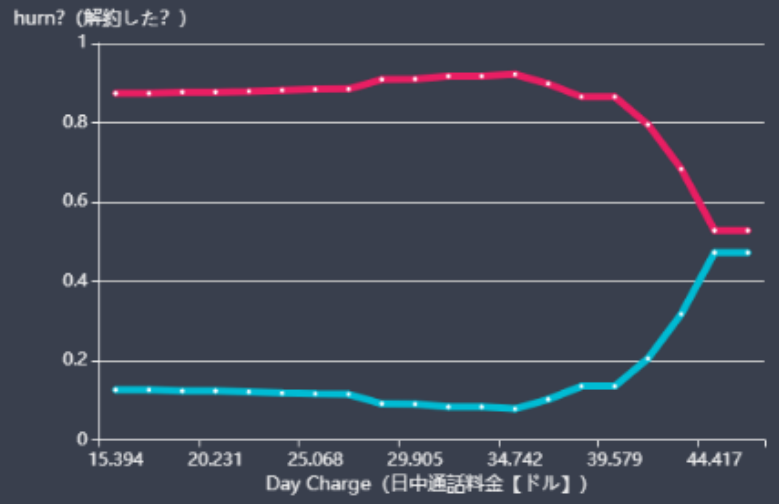
■ ボイスメールに契約していない顧客（全体の72.3%）は、全顧客の予測モデルと類似した結果

Performance Report

| Label Name | F1 Score | Precision Score | Recall Score | Support |
|---|---|---|---|---|
| False. | 0.969 | 0.95 | 0.988 | 602 |
| True. | 0.826 | 0.928 | 0.744 | 121 |

Insight



# 予測モデルの結果（ボイスメールに契約している顧客）

■ ボイスメールに契約してる顧客（全体の27.7%）は、全顧客の予測モデルと異なる結果

Performance Report

| Label Name | F1 Score | Precision Score | Recall Score | Support |
|---|---|---|---|---|
| False. | 0.964 | 0.93 | 1 | 253 |
| True. | 0.345 | 1 | 0.208 | 24 |

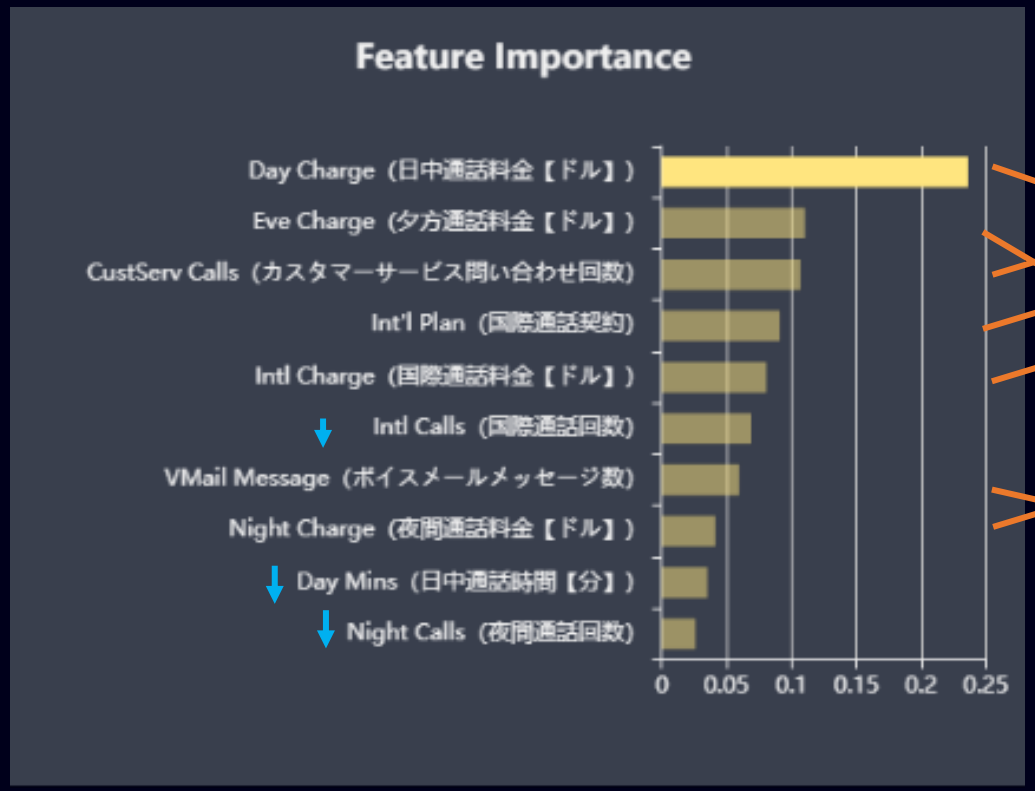
<u>予測精度の低下</u>

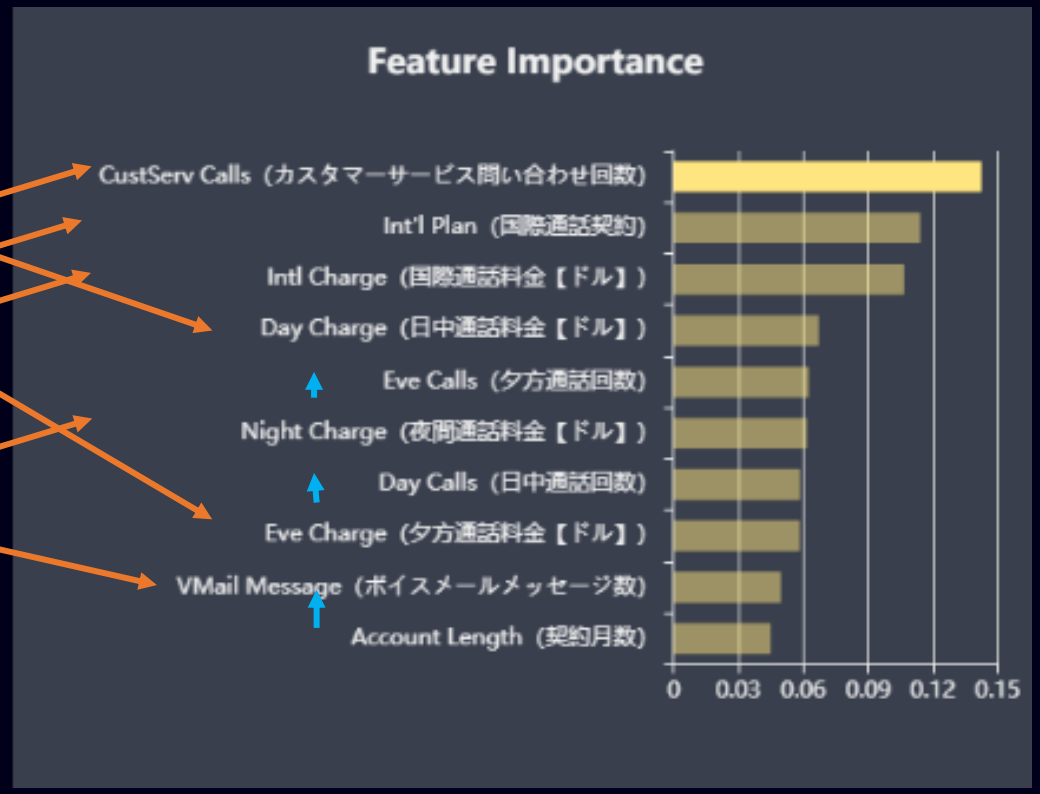
- サンプル数が減少
- 解約率の偏り度合の増加
  - 顧客全体の解約率　14.5%
  - ボイスメール契約顧客の解約率　8.6%

©GRI Inc.

# 予測モデルの結果（ボイスメールに契約している顧客）

■ ボイスメールに契約してる顧客（全体の27.7%）は、全顧客の予測モデルと異なる結果



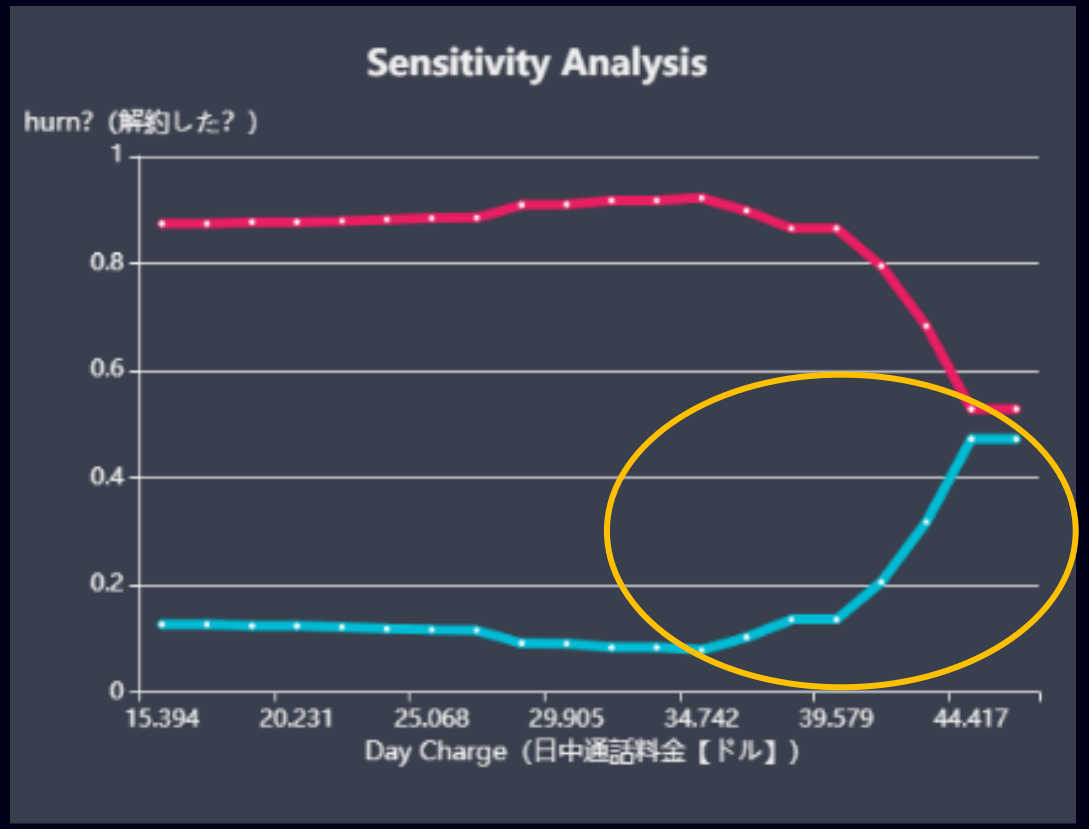




©GRI Inc.

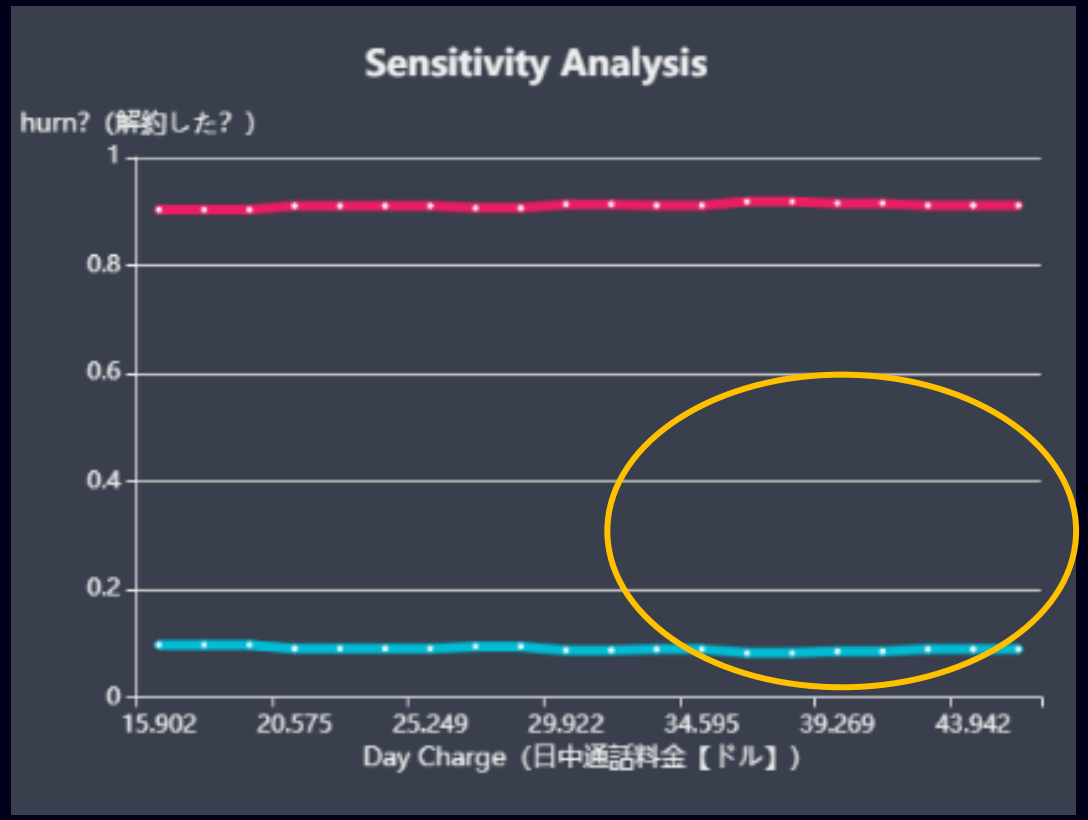
# 予測モデルの結果（ボイスメールに契約している顧客）

■ ボイスメール契約は、優良顧客流出の歯止めになっている

ボイスメール未契約の顧客の
日中通話料金ごとの解約確率




ボイスメール契約の顧客の
日中通話料金ごとの解約確率




©GRI Inc.

# ボイスメール契約有無で予測モデル分割による知見

■多くの優良顧客（日中の通話時間が長い高課金ユーザ）が解約している中、ボイスメール契約ユーザは、その影響が薄い

→　ボイスメール契約は顧客にとってのスイッチングコストが高くなる機能で、顧客のロイヤル化に貢献

→　単純に価格を安くするだけでなく、価値の高いオプション機能開発の重要性を示唆（メリットはオプションのクロスセルだけではない）

©GRI Inc.

# ADK生活者総合調査を用いた海外旅行するユーザの予測

海外旅行するユーザの特性が分かると…

- オンライン広告の対象を絞り込むことができ、少しでも効率的な広告配信できる
- キャンペーンの中味を考えるときに、相性の良いテーマを訴求できる

©GRI Inc.

# ADK生活者総合調査とは

## ADKが企画・運営しているオリジナル調査です。

生活者をライフスタイル・消費行動・メディア接触など多面的側面から捉えることで、より深いターゲットの実像を導き出すことを目的としたオリジナル調査です。

どんな価値観を持っている人が、どんな生活行動をし、どんな商品を所有したり、買いたいと考えているのか？彼らに効率よく到達し、好感を持って接触してもらえるメディアは何か？などを明らかにすることができます。

© 2021 ADK Marketing Solutions Inc.  ALL RIGHTS RESERVED

# ADK生活者総合調査の特徴

| | |
|---|---|
| ① 大規模調査 | 男女 15歳から79歳まで約15,000人を調査<br>関東（1都6県）と関西（2府4県）を調査し、日本の人口の約50%をカバー |
| ② 多数の調査項目による<br>シングルソースデータ | 6,000を超える多様な質問項目で、生活者の意識・行動からメディア接触まで<br>一貫した分析が可能 |
| ③ 毎年、定期的に調査 | 現在の形で調査を始めた2007年以来、毎年調査を重ねている<br>意識・行動の時系列変化を追うことができる |
| ④ 追跡調査が可能 | 生活者総合調査の回答者に対して追跡調査が可能<br>絞り込んだ条件での調査も可能で、全部の調査データとクロス集計も可能 |
| ⑤ 外部調査データとの連携 | 一部サンプルについて、インテージ社「i-SSP」と「SCI」と回答者が重なっており、<br>生活者総合調査と「i-SSP」「SCI」とのクロス集計が可能 |

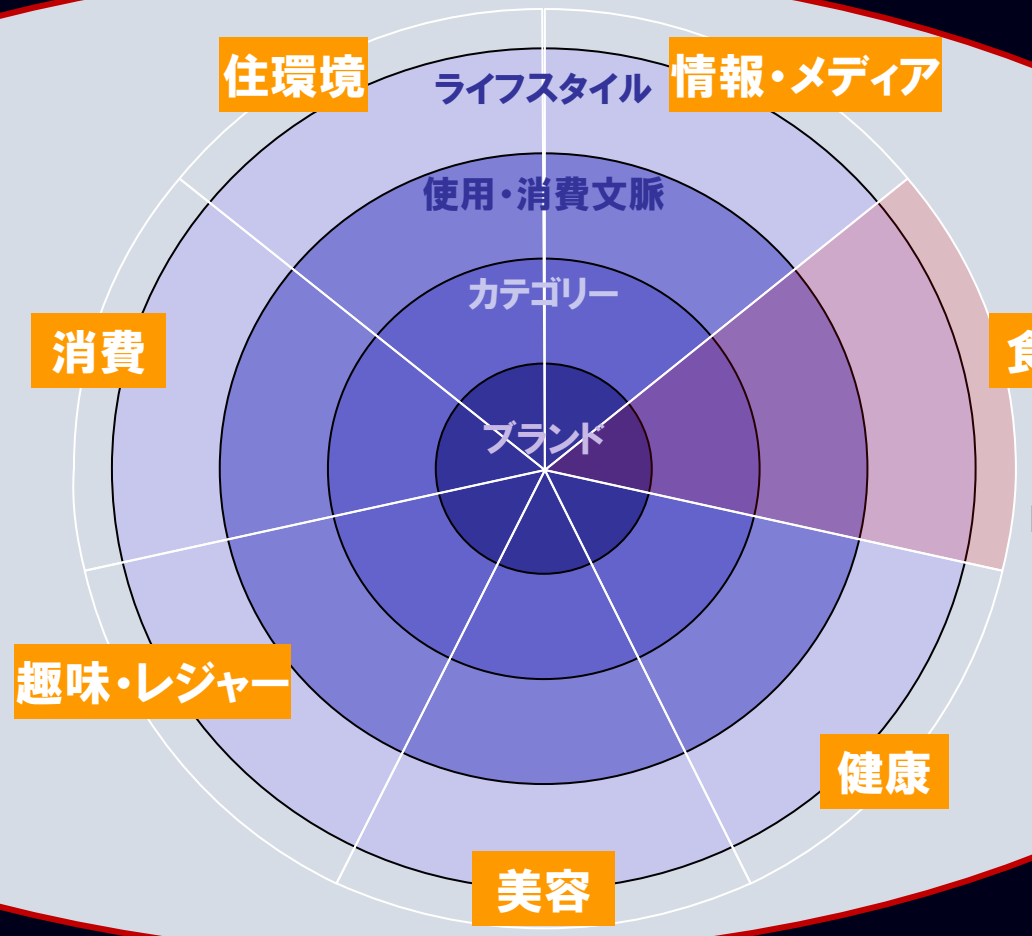
© 2021 ADK Marketing Solutions Inc. ALL RIGHTS RESERVED

# 『ADK生活者総合調査』の設計コンセプト

生活者が最終的に行動（商品の購入など）に至るまでには、その人のベースとなる価値観や意識、性格、嗜好などがあり、影響している。それらは、生活の分野別にも存在し、それぞれ意識や行動が形成されている。

**生活者『基本部分』**

生活習慣・生活行動
メディア
価値観・性格・嗜好
ライフスタイル
基本特性　etc.




『食』に関する．．．．
- ライフスタイル
- 使用・消費文脈
- 消費欲求
- カテゴリー

ADK Marketing Solutions Inc.  ALL RIGHTS RESERVED

# 主な調査項目

## 生活篇

1）日々の行動
平日・休日×生活時間帯別の行動実態
生活場所・情報接点・行為・買い物行動（店）
居住地・通勤通学地・訪問地・利用路線・駅・商業施設等
趣味・旅行・スポーツイベント・好きなタレント　等

2）意識・価値観
生活価値観・生活のよりどころ・日頃大切にしている言葉
景況感・幸福度、家族観
情報とメディアについての意識・食に関する意識・健康に関する意識・
住環境に関する意識・美容や身だしなみに関する意識

3）消費・購買行動
消費意識・金銭感覚、消費行動・消費欲求、ファッション意識・実態
利用商品・今後利用したい商品、利用ブランド

4）デジタルライフ
インターネット利用実態
ソーシャルメディア／動画サイト／ショッピングサイトに関する実態

5）基本特性
性別・年齢・職業・未既婚・家族構成・収入
就業形態・転職経験・転職意向、こづかい使途・ペット　　等

## ビークル篇

1）テレビ番組（地上・BS）・衛星テレビ局（BS・CS）・ラジオ局（AM・FM）
番組視聴・聴取実態（視聴頻度、リアルタイム／録画状況、視聴集中度）
評価・イメージ・接触後の行動

2）テレビ／ラジオ 局別曜日別接触時間

3）新聞（電子版を含む一般紙・スポーツ誌・経済誌）
閲読実態（閲読曜日）
接触後の行動

4）雑誌・フリーペーパー/マガジン
閲読実態（閲読号、購読状況）
評価・イメージ・接触後の行動

5）インターネットサイト（一般サイト・企業サイト）
サイト接触実態（利用デバイス、接触頻度-日数／時間、利用形態）
接触後の情報発信・行動（一般サイトのみ）
サイトイメージ（一般サイトのみ）
接触のきっかけ・目的（企業サイトのみ）

6）アプリ（ゲーム／一般）
認知・ダウンロード・利用状況

ADK Marketing Solutions Inc.  ALL RIGHTS RESERVED

# 調査概要　ADK生活者総合調査2020年

| | |
|---|---|
| **調査目的** | 生活者の生活行動・価値観・メディア接触を多面的に把握する |
| **調査方法** | WEB調査 |
| **調査対象** | 15～79歳の男女（中学生は除く） |
| **調査地域** | 関東 (東京都・神奈川県・埼玉県・千葉県・群馬県・栃木県・茨城県)<br>関西 (大阪府・京都府・兵庫県・奈良県・和歌山県・滋賀県) |
| **対象者抽出方法** | 調査会社(株式会社インテージ)が有する調査協力者パネルより抽出 |
| **調査期間** | 生活編:　2020/7/1(水) ~ 7/17(金)<br>ビークル編: 2020/10/30(金) ~ 11/25(水)　※生活編回答者への追跡調査として実施 |

※ 例年、5月下旬~6月上旬に生活編・ビークル編を同時に実施しておりましたが、2020年は新型コロナの影響で同期間の生活者の行動やメディア接触が「非常時」であったため、時期を延期かつ分割して実施しております。

© 2021 ADK Marketing Solutions Inc.　ALL RIGHTS RESERVED

# 海外旅行に行く消費者の特徴を明らかに

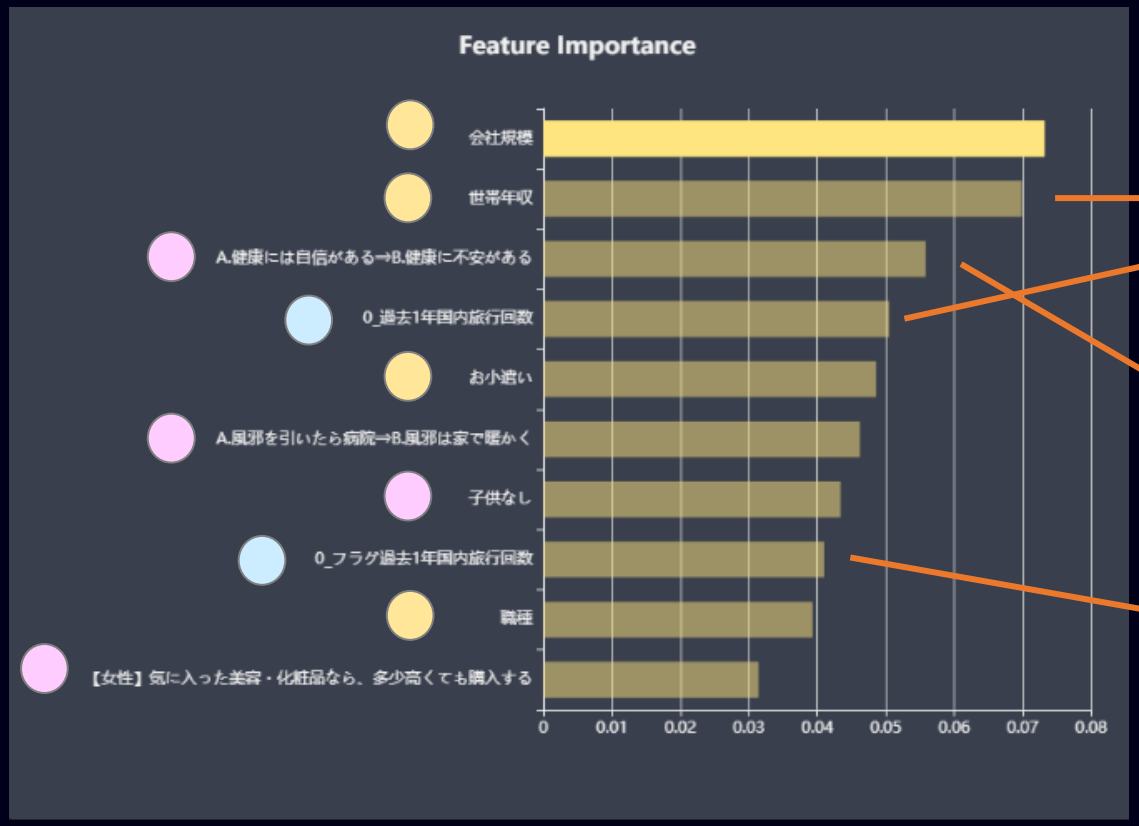
- 調査ID数：　　10,000以上
- ターゲット：　　３年以内に海外旅行に行ったか？

©GRI Inc.

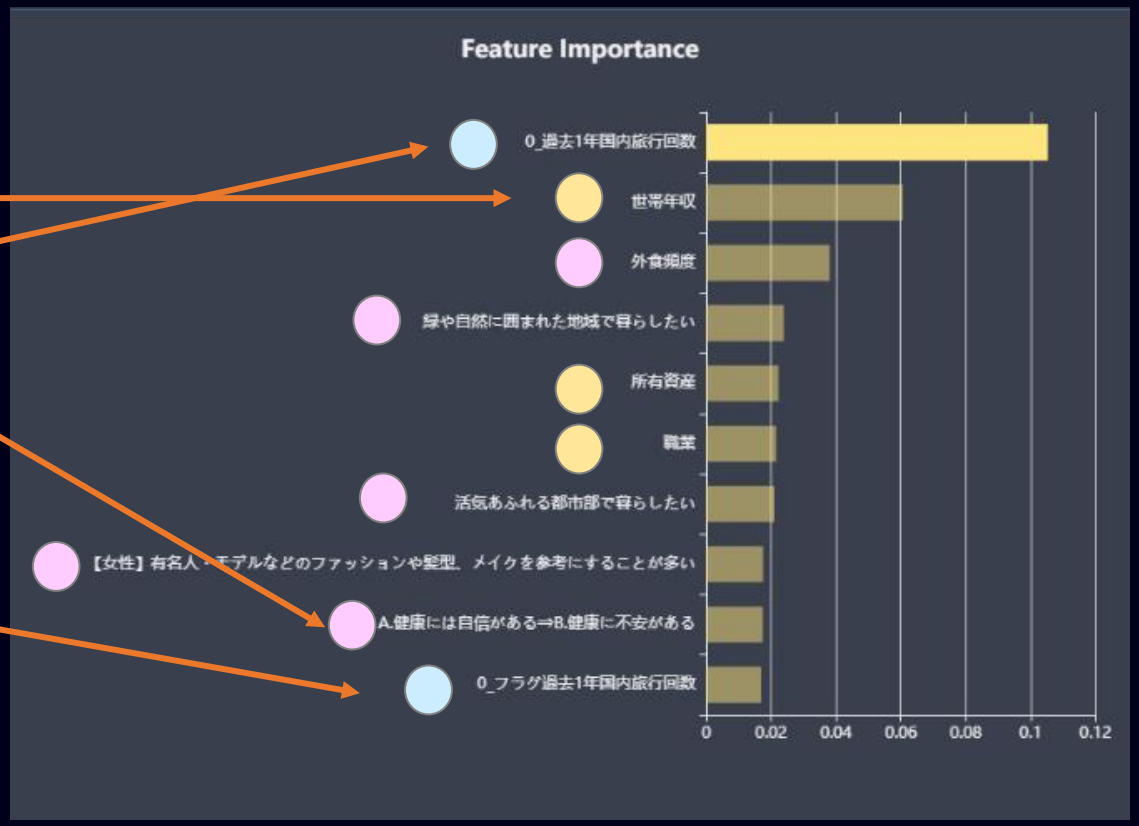
# 海外旅行に行く消費者の特徴を明らかに（性年代ごとに）

■予測モデルの自動分割で、簡単に性年代ごとに12個の予測モデルを構築

**30代女性**




**50代女性**



# 予測モデルの分割のまとめ

- 予測モデルをカテゴリカル特徴量で分割すると、全体では見えなかった知見を見つける可能性がある

- ただし、予測精度が下がる可能性がある（トレードオフ）

- ForecastFlowで自動分割機能が使えるのは、カテゴリカル特徴量で値の種類が15以下の場合
  - Secondary IDで指定する

©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
- 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.

# ForecastFlowトライアル・アカウントの配布

- 訓練する際、最大1,000行になります
- 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：
    - イベントパスワード： 2020

©GRI Inc.