

# PT-272

## Power BI で実現する予測機能と Data Mining Add-in

日本マイクロソフト株式会社  
ビジネスプラットフォーム統括本部  
データプラットフォーム技術部  
シニアテクノロジー スペシャリスト

伊藤 歩

ROOM **D**

# 架空の企業: Contoso ドラッグ

## Contoso ドラッグ企業情報

---

- 1997 年に第一号店を神奈川県に開店
  - 現在、全国 200 店舗を展開する中堅ドラッグストアチェーン
  - 2007 年に東証一部に上場
  - 業界に先駆けて「ポイントカード」を導入
  - ソーシャルメディアに公式アカウントを作成して積極的に情報を発信
- 

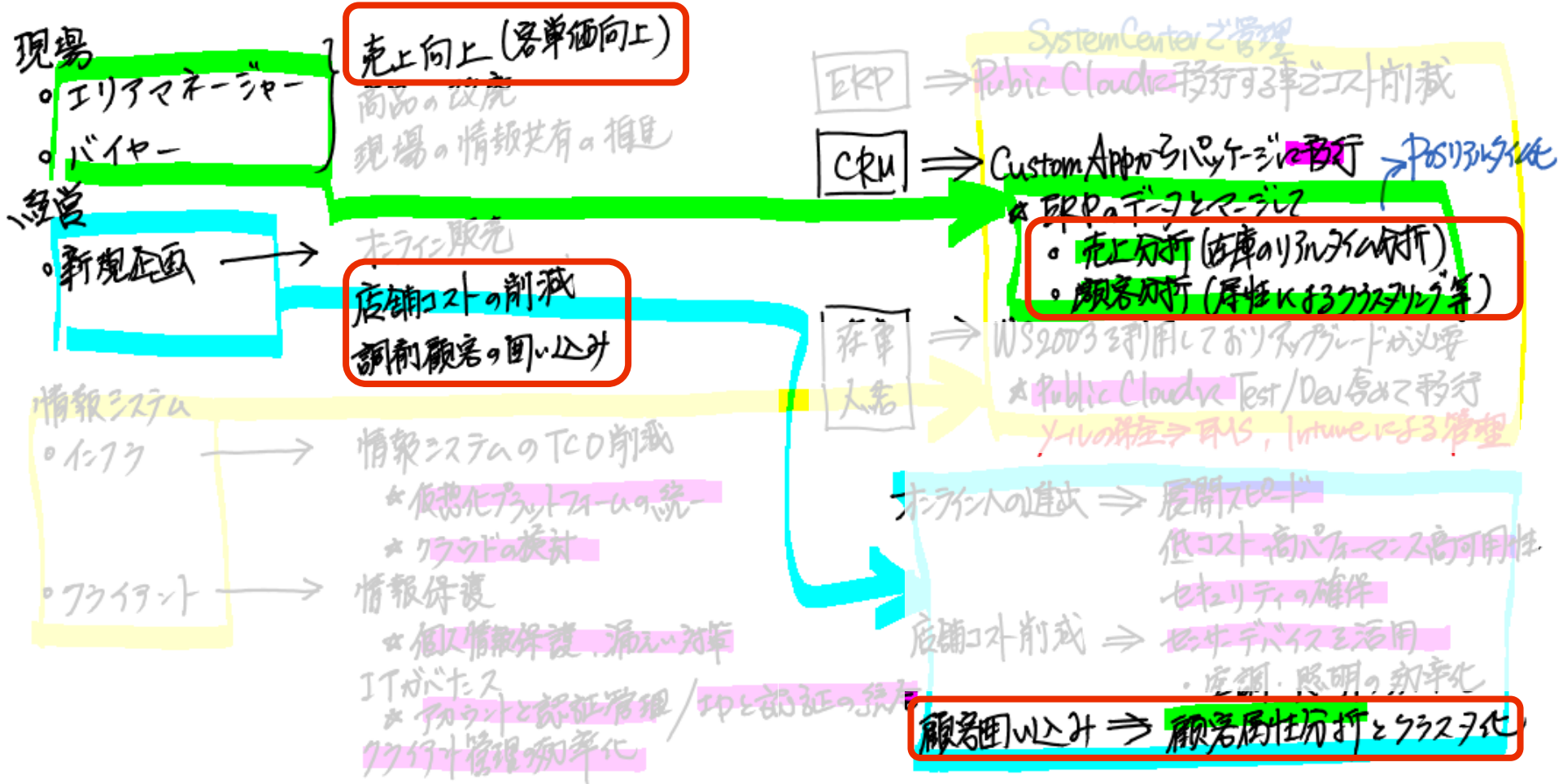


### 山良子 (38 歳)

Contoso ドラッグの経営企画部で新規企画を担当するマネージャー

- 営業部門で新規出店開発を行った後、一昨年から経営企画部で売上拡大につながる新規企画を担当

# Contoso ドラッグの問題点と対策



# 改善のための要件: 新規企画担当マネージャー 山良子の場合



経営企画部で  
新規企画を担当するマネージャー

## 課題



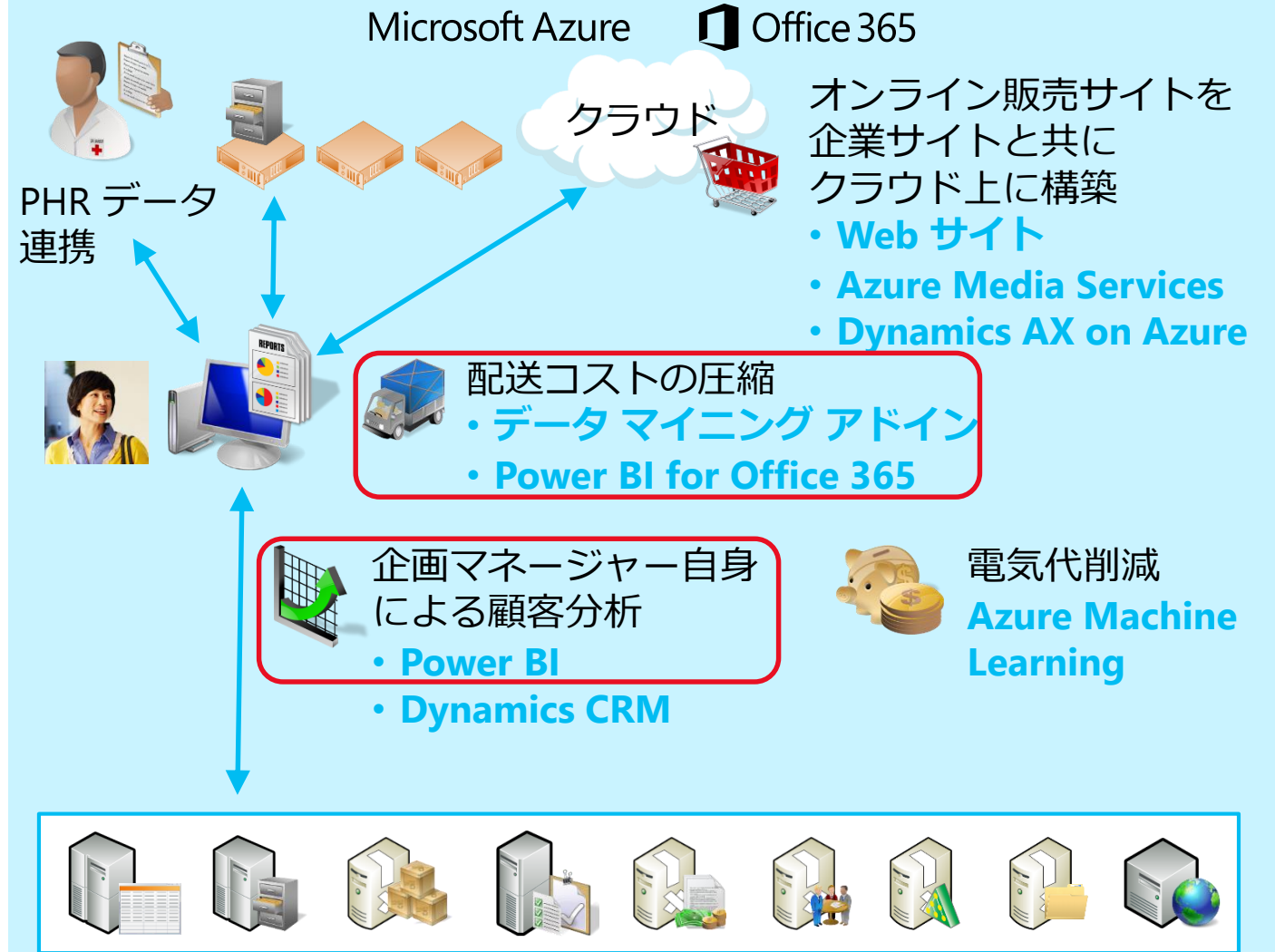
- オンライン販売の検討
- 店舗コストの削減
- 店舗の処方箋顧客囲い込み

## 改善方法



- 顧客分析: スマホなどのアクティブユーザーの把握
- 配送コストの圧縮
- 設備機器の効率的な活用による空調費や照明費などの削減 (IoT の活用)
- 血圧データの個人 PHR 収集

## マイクロソフトのソリューションとその効果



# 課題 1：配送コストの圧縮

突発的な配送手配によるコスト増



売上予測に基づく在庫管理の適正化

- 1 これまでの売上傾向を元に将来の売上を予測し、在庫管理を適正化する
- 2 予測値がどのくらい確からしいか検証の上利用すべき
- 3 過去の売上には在庫切れで機会損失があるので、修正して予測する必要がある

実績に基づく予測

予測の検証

過去データを  
修正して予測

# Demo 1

## 売上予測

# Power View 予測

1

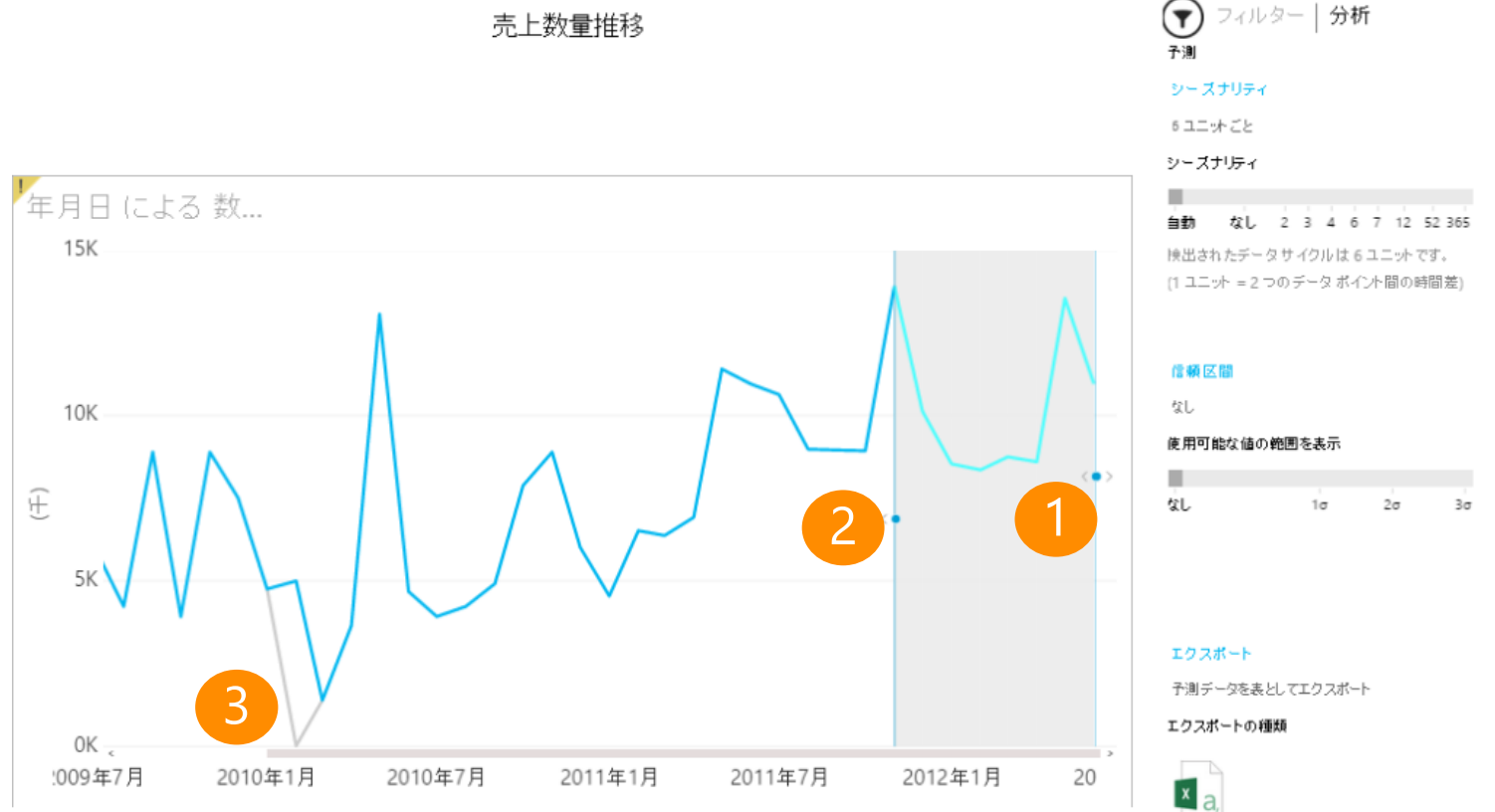
Forecast

2

Hindcast

3

what-if analysis



# 予測：分析ウィンドウ

## A シーズナリティ

- データの周期性 (データサイクル)
- 自動検出するが、明示的に指定も可能

## B 信頼区間

- 標準偏差  $\sigma$  (シグマ) を設定し、予測値 (水色の折れ線) の 68%/95%/99% ( $1\sigma/2\sigma/3\sigma$ ) がシェードエリア (水色の網掛部分) 内にあることを示す

## C エクスポート

- 元のデータ、調整データ、予測データをCSV形式でエクスポート



フィルター | 分析

予測

シーズンナリティ **A**

12 ユニットごと

シーズンナリティ

自動 なし 2 3 4 6 7 12 52 365

検出されたデータサイクルは 12 ユニットです。  
(1 ユニット = 2 つのデータポイント間の時間差)

信頼区間 **B**

期待される値の 95%

使用可能な値の範囲を表示

なし 1σ 2σ 3σ

期待される値の 95% (2 の標準偏差) がシェード  
エリア内にあります。

エクスポート **C**

予測データを表としてエクスポート

エクスポートの種類



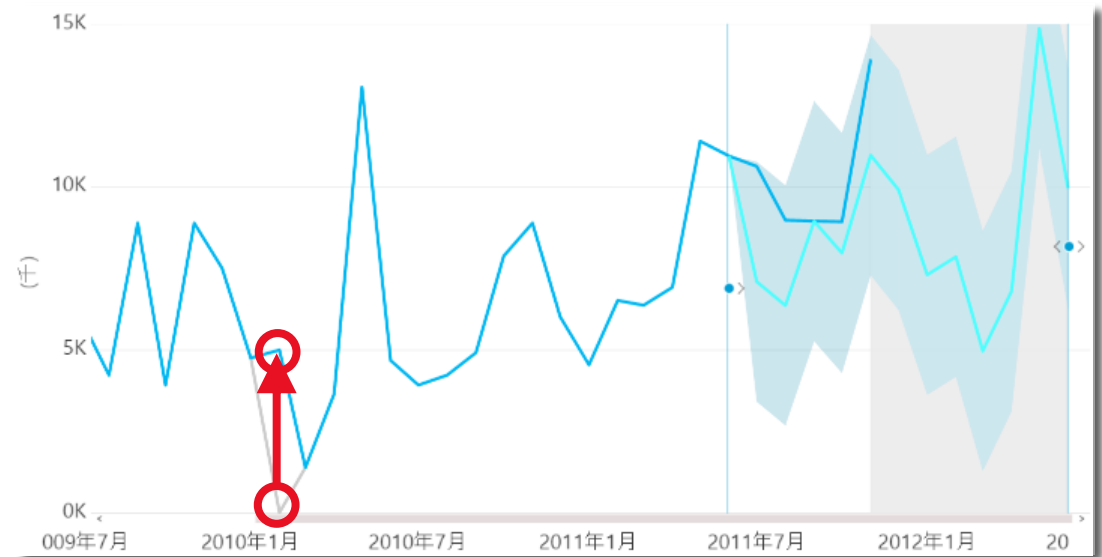
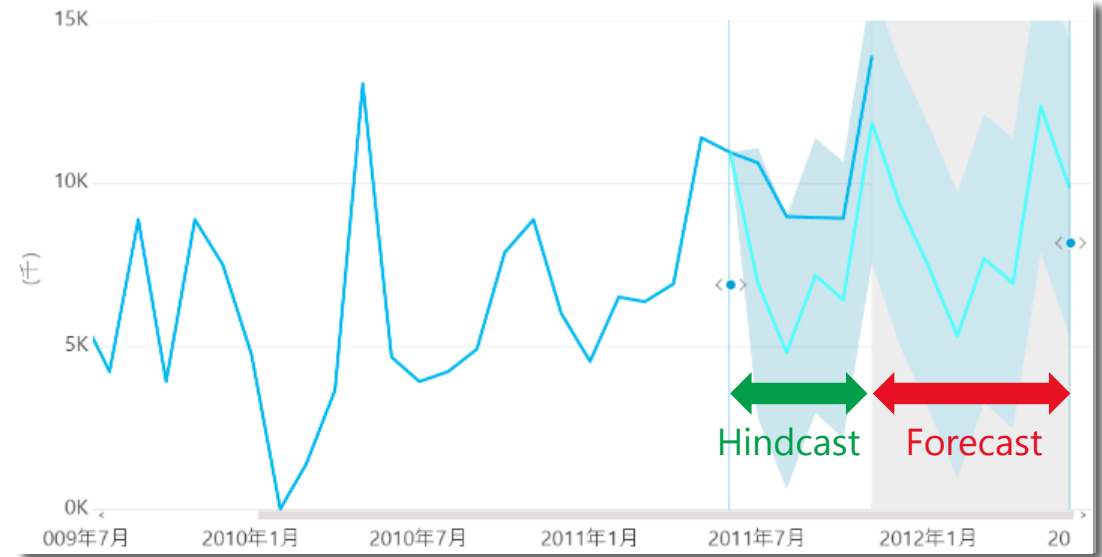
# グラフ上での操作

- Hindcast

- 予測モデルを使用して、過去データを予測し、モデルの予測精度を検証
- 水色の線が予測、青い線が実際の値
- Power BI アプリでは動作しない (2014/10/17 現在)

- what-if analysis

- 折れ線の頂点をドラッグにより修正し、突発的な事象による影響を排除
- 複数の頂点を変更可能
- 予測と信頼区間が再計算される



# 課題 2 : 店舗の処方箋顧客 囲い込み

処方箋 (調剤) 利用客を増やしたい



顧客を分類し、適切なキャンペーンを実施

- 1 売上履歴から、調剤の利用有無に注目して顧客を分類
- 2 複数の方法 (アルゴリズム) を用いて分類
- 3 複数の分類結果を比較して、より効果的であろう顧客ターゲットに対しキャンペーンを実施

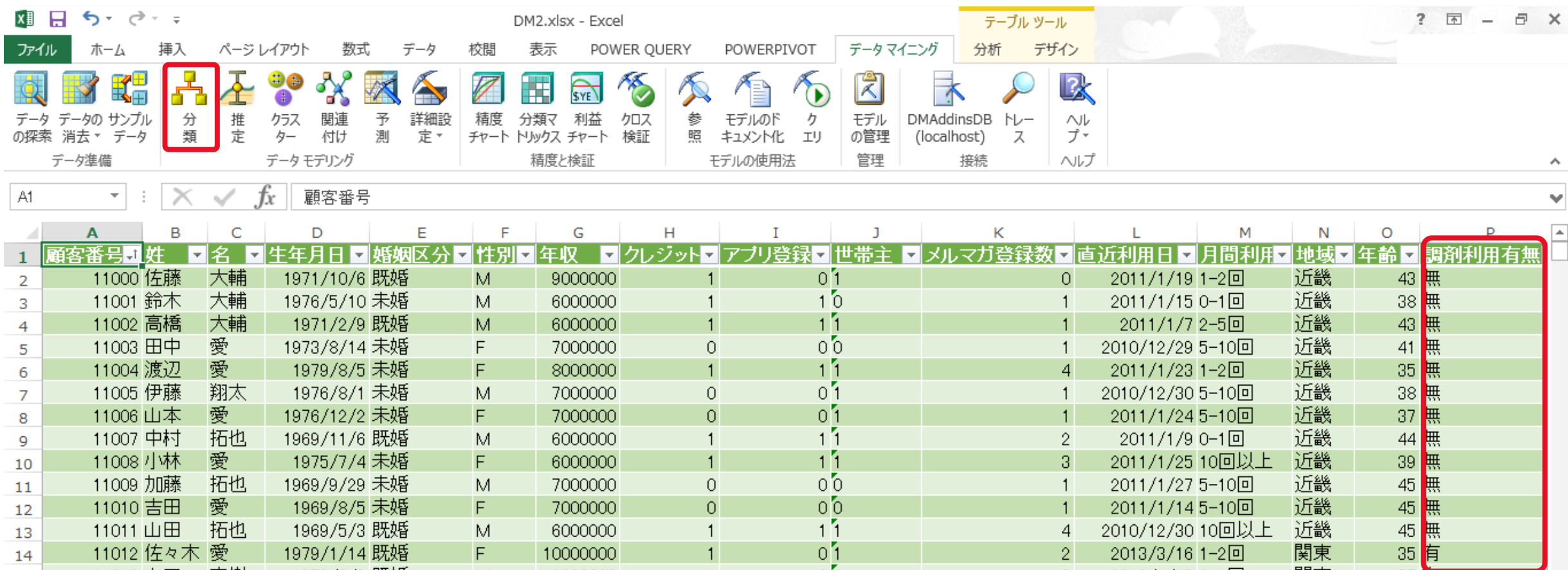
実績に基づく予測

予測の検証

CRM との連携

# 調剤利用有無との相関性で分類

性別、年齢、年収、婚姻区分、クレジットカード有無、メルマガ登録数、アプリ登録有無、RFM の各値等との相関性に基づく分類を行う



DM2.xlsx - Excel

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 POWER QUERY POWERPIVOT データマイニング テーブルツール (分析 デザイン)

データ探索 データの消去 サンプルデータ データ準備

分類 推定 クラスター 関連付け 予測 詳細設定

精度チャート 分類マトリックス 利益チャート クロス検証

参照 モデルのドキュメント化 クエリ

モデルの管理 DMAddinsDB (localhost) トレース ヘルプ

A1 : X ✓ fx 顧客番号

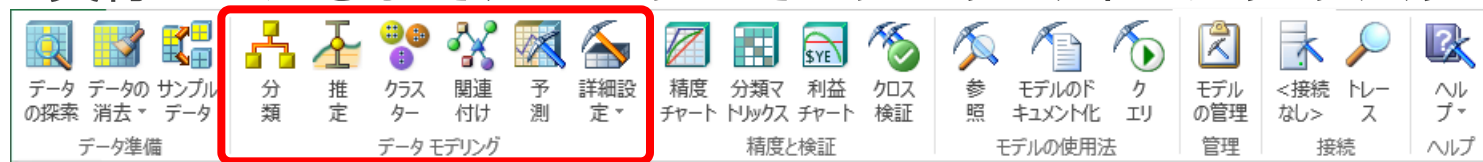
|    | A     | B   | C  | D         | E    | F  | G        | H     | I     | J   | K       | L          | M     | N  | O  | P      |
|----|-------|-----|----|-----------|------|----|----------|-------|-------|-----|---------|------------|-------|----|----|--------|
| 1  | 顧客番号  | 姓   | 名  | 生年月日      | 婚姻区分 | 性別 | 年収       | クレジット | アプリ登録 | 世帯主 | メルマガ登録数 | 直近利用日      | 月間利用  | 地域 | 年齢 | 調剤利用有無 |
| 2  | 11000 | 佐藤  | 大輔 | 1971/10/6 | 既婚   | M  | 9000000  | 1     | 0     | 1   | 0       | 2011/1/19  | 1-2回  | 近畿 | 43 | 無      |
| 3  | 11001 | 鈴木  | 大輔 | 1976/5/10 | 未婚   | M  | 6000000  | 1     | 1     | 0   | 1       | 2011/1/15  | 0-1回  | 近畿 | 38 | 無      |
| 4  | 11002 | 高橋  | 大輔 | 1971/2/9  | 既婚   | M  | 6000000  | 1     | 1     | 1   | 1       | 2011/1/7   | 2-5回  | 近畿 | 43 | 無      |
| 5  | 11003 | 田中  | 愛  | 1973/8/14 | 未婚   | F  | 7000000  | 0     | 0     | 0   | 1       | 2010/12/29 | 5-10回 | 近畿 | 41 | 無      |
| 6  | 11004 | 渡辺  | 愛  | 1979/8/5  | 未婚   | F  | 8000000  | 1     | 1     | 1   | 4       | 2011/1/23  | 1-2回  | 近畿 | 35 | 無      |
| 7  | 11005 | 伊藤  | 翔太 | 1976/8/1  | 未婚   | M  | 7000000  | 0     | 0     | 1   | 1       | 2010/12/30 | 5-10回 | 近畿 | 38 | 無      |
| 8  | 11006 | 山本  | 愛  | 1976/12/2 | 未婚   | F  | 7000000  | 0     | 0     | 1   | 1       | 2011/1/24  | 5-10回 | 近畿 | 37 | 無      |
| 9  | 11007 | 中村  | 拓也 | 1969/11/6 | 既婚   | M  | 6000000  | 1     | 1     | 1   | 2       | 2011/1/9   | 0-1回  | 近畿 | 44 | 無      |
| 10 | 11008 | 小林  | 愛  | 1975/7/4  | 未婚   | F  | 6000000  | 1     | 1     | 1   | 3       | 2011/1/25  | 10回以上 | 近畿 | 39 | 無      |
| 11 | 11009 | 加藤  | 拓也 | 1969/9/29 | 未婚   | M  | 7000000  | 0     | 0     | 0   | 1       | 2011/1/27  | 5-10回 | 近畿 | 45 | 無      |
| 12 | 11010 | 吉田  | 愛  | 1969/8/5  | 未婚   | F  | 7000000  | 0     | 0     | 0   | 1       | 2011/1/14  | 5-10回 | 近畿 | 45 | 無      |
| 13 | 11011 | 山田  | 拓也 | 1969/5/3  | 既婚   | M  | 6000000  | 1     | 1     | 1   | 4       | 2010/12/30 | 10回以上 | 近畿 | 45 | 無      |
| 14 | 11012 | 佐々木 | 愛  | 1979/1/14 | 既婚   | F  | 10000000 | 1     | 0     | 1   | 2       | 2013/3/16  | 1-2回  | 関東 | 35 | 有      |

# Demo 2

## 顧客の分類

# データマイニングクライアント

Analysis Services マイニング エンジン機能を Excel からウィザードで利用  
データのクリーニングとサンプリングをはじめ、モデルの構築、モデルの精度のテスト、クロス検証  
の実行にいたるまで、ほぼすべてのデータマイニングタスクのためのツール



| データモデリングの種類          | 説明  |
|----------------------|---|
| 分類                   | モデル内の他の列の値に基づいて、ある 1 つの列の値を予測する分類モデルを構築   |
| 推定                   | データパターンを抽出し、そのパターンを使用して、連続する数値、日付、または時間の値を予測する推定モデルを構築  |
| クラスター                | 類似の特性を共有する行のグループを検出するクラスターモデルを構築  |
| 関連付け                 | 買い物かご分析などの複数のトランザクションに現れるアイテム間の関連付けを検出する、アソシエーションモデルを構築   |
| 予測                   | 一連のセルのパターンを検出する予測モデルを構築し、追加の値を予測  |
| 詳細設定<br>– マイニング構造の作成 | SQL Server Analysis Services データマイニングアルゴリズムを使用し、Excel に格納されたデータに基づいて新しいデータマイニングモデルを構築。強力な新しいクエリエディターによって、パラメーターをカスタマイズし、データマイニング拡張機能 (DMX) ステートメントを構築 |
| 詳細設定<br>– 構造へのモデルの追加 | 新しい関連モデルを作成することにより、既存のデータ構造を変更<br>これらの新しいモデルにより、同じデータを別のデータマイニングアルゴリズムで分析   |

# データモデリング - 分類

**分類ウィザード**

作業の開始 (データ分類ウィザード)

**説明**  
データ分類ウィザードでは、Excel テーブル、Excel 範囲、または外部データソースにある既存のデータに基づいて、分類モデルを構築できます。分類モデルは、ある列の値を、他の列の値に基づいて予測するパターンを抽出します。

**操作**  
このウィザードでは、対象列および必要な入力列を選択できます。

**注**  
データ分類ウィザードを使用するには、SQL Server Analysis Services データベースに接続しておく必要があります。このウィザードで作成するモデルは、保存するか、データの処理中にのみ存在させることができます。ただし、一時的なモデルを作成するには、サーバーの構成で一時的なマイニング モデルを許可する必要があります。この操作を許可する設定については、サーバー管理者に問い合わせてください。また、このウィザードでは、モデルにトレーニング データを保存してそのデータをパターンと関連付ける (ドリルスルーを有効にする) かどうかを指定することもできます。

**ソースデータの選択**

ソースデータの選択

☒ テーブル(T): '顧客分析'テーブル39  
☐ データ範囲(R):  
☒ データ範囲にヘッダーが含まれている(H)  
☐ 外部データソース(E):  
データソース名(S):  
クエリ(Q):

**分類**

分析する列(C): 調剤利用有無  
入力列(I):

| <input checked="" type="checkbox"/> | 列名       |
|-------------------------------------|----------|
| <input type="checkbox"/>            | 顧客番号     |
| <input type="checkbox"/>            | 姓        |
| <input checked="" type="checkbox"/> | 名        |
| <input type="checkbox"/>            | 生年月日     |
| <input checked="" type="checkbox"/> | 婚姻区分     |
| <input checked="" type="checkbox"/> | 性別       |
| <input checked="" type="checkbox"/> | 年収       |
| <input checked="" type="checkbox"/> | クレジットカード |
| <input checked="" type="checkbox"/> | アプリ登録    |
| <input checked="" type="checkbox"/> | 世帯主      |
| <input checked="" type="checkbox"/> | メルマガ登録数  |
| <input checked="" type="checkbox"/> | 直近利用日    |
| <input checked="" type="checkbox"/> | 月間利用頻度   |
| <input checked="" type="checkbox"/> | 地域       |
| <input checked="" type="checkbox"/> | 年齢       |
| <input type="checkbox"/>            | 調剤利用有無   |
| <input type="checkbox"/>            | 住所1      |
| <input type="checkbox"/>            | 住所2      |
| <input type="checkbox"/>            | 電話番号     |
| <input type="checkbox"/>            | Email    |

パラメーター(P)... < 戻る(B) 次へ(N) > キャンセル

**データをトレーニングセットとテストセットに分割**

テスト用データの割合(P): 30  
最大行数(M):  
**説明**  
このウィザードでは、マイニング構造のデータがトレーニング セットとテスト セットに分割されます。データマイニングエンジンでは、既定で、トレーニング セットを使用してマイニング モデルをトレーニングし、テスト セットを使用してモデルの精度をテストします。  
テスト用として確保する必要がある入力データ量を指定するには、このページのオプションを使用します。両方のオプションを設定した場合は、両方の制限が使用されます。たとえば、最大行数がテスト用データの割合より小さい場合は、最大行数がテストに使用されます。最大行数を 0 に設定すると、この制限は使用されません。  
このウィザードで作成されるパーティションはランダムなので、典型的なトレーニング セットとテスト セットを確保できます。パーティションはマイニング構造に作成されるため、ソース データには影響しません。

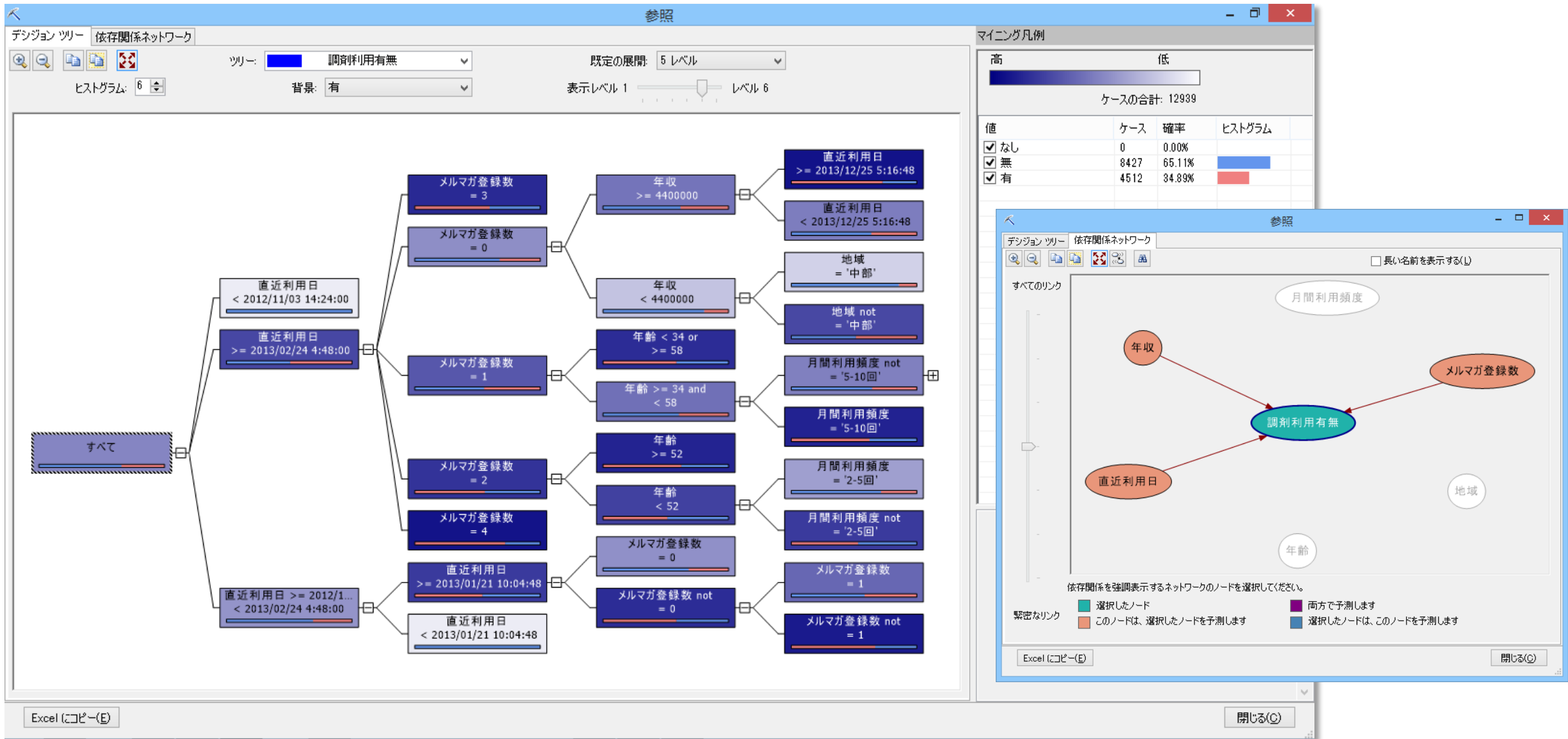
< 戻る(B) 次へ(N) > キャンセル

**完了**

構造名(S): テーブル39 構造  
構造の説明(D): DM\_Dataxlsx ワークブック内の 顧客分析 シードにある テーブル39 テーブルの構造  
モデル名(M): 分類 ディビジョンツリー  
モデルの説明(Q): テーブル39 テーブルに基づいて構築された Microsoft\_Ddecision\_Trees モデル  
オプション  
☒ モデルを参照する(R) ☒ ドリルスルーを有効にする(E)  
☐ 一時的なモデルを使用する(U)

< 戻る(B) 完了(F) キャンセル

# モデルの参照 – ディシジョン ツリー



# データの準備・調査

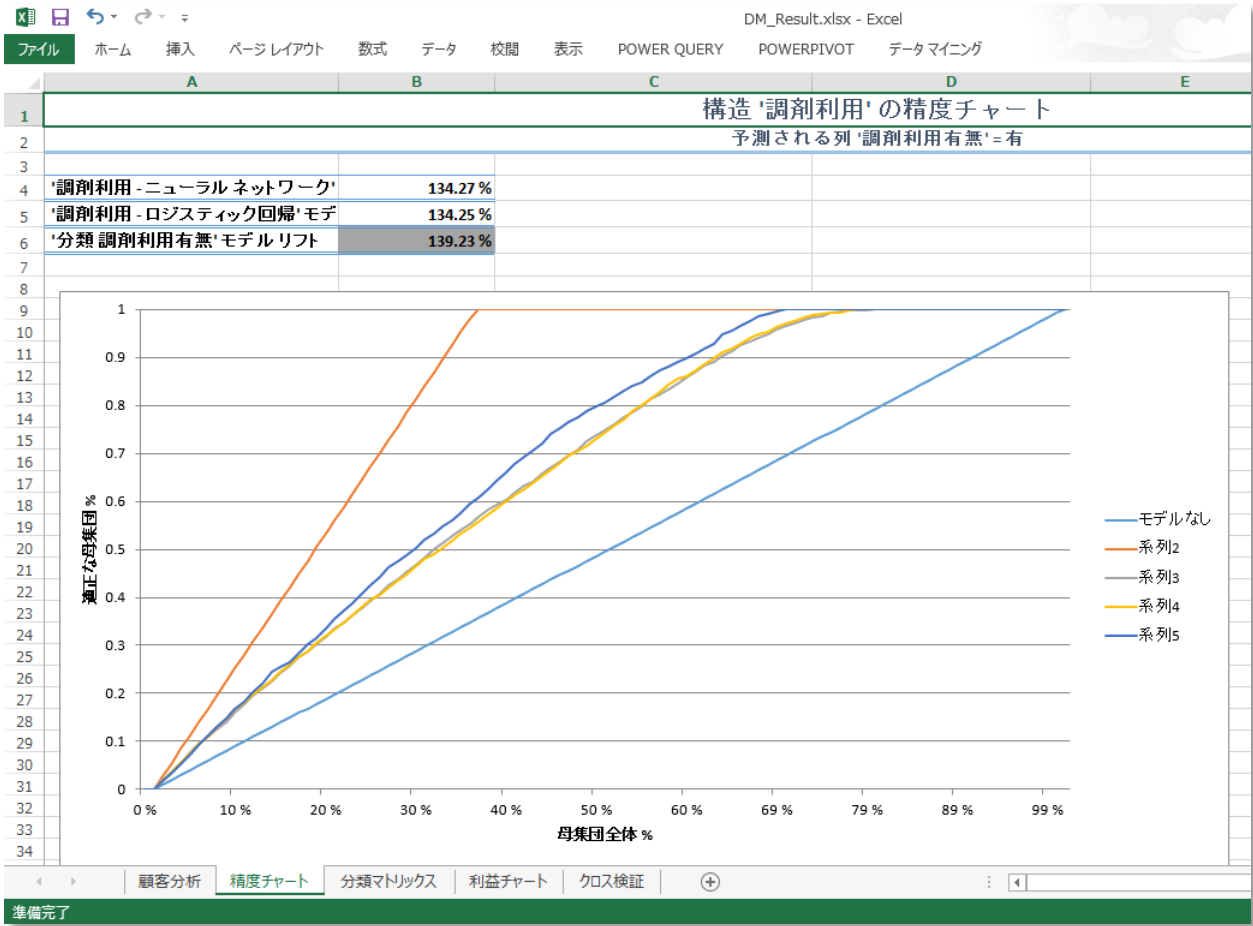
- 分析元データの内容の確認
- モデル作成のために使うトレーニングデータと、評価用のデータの準備



- データの消去
  - 外れ値の処理
  - ラベルの変更
- サンプルデータ (サンプリング)
  - ランダム
  - データ分布を均衡化



# 精度と検証



## 分類マトリックス

DM\_Result.xlsx - Excel

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 POWER QUERY POWERPIVOT データマイニング

構造 '調剤利用' の正しい分類と不適切な分類の数  
予測される列 '調剤利用有無'

列は実際の値に対応します  
行は予測される値に対応します

| モデル名:                                 | 調剤利用 - ニューラル ネットワーク | 調剤利用 - ニューラル ネットワーク | 調剤利用 - ロジスティック回帰 |
|---------------------------------------|---------------------|---------------------|------------------|
| 正しい分類の合計:                             | 69.05 %             | 3829                |                  |
| 誤分類の合計:                               | 30.95 %             | 1716                |                  |
| モデル '調剤利用 - ニューラル ネットワーク' の結果 (パーセント) |                     |                     |                  |
|                                       | 有 (実際)              | 無 (実際)              |                  |
| 有                                     | 47.66 %             | 19.21 %             |                  |
| 無                                     | 52.34 %             | 80.79 %             |                  |
| 正しい分類                                 | 47.66 %             | 80.79 %             |                  |
| 誤分類                                   | 52.34 %             | 19.21 %             |                  |
| モデル '調剤利用 - ニューラル ネットワーク' の結果 (カウント)  |                     |                     |                  |
|                                       | 有 (実際)              | 無 (実際)              |                  |
| 有                                     | 936                 | 688                 |                  |
| 無                                     | 1028                | 2893                |                  |
| 正しい分類                                 | 936                 | 2893                |                  |
| 誤分類                                   | 1028                | 688                 |                  |



# テーブル分析ツール

Analysis Services マイニング エンジンを利用して Excel 上でマイニングを行う



- Excel テーブルのデータを使用する簡易かつ強力なツール
- アルゴリズムによって適切なパラメーターが検出され、モデルの理解に必要なグラフとレポートをすべて生成

| テーブル分析ツール名 |             | 説明  |
|------------|-------------|---|
| 主要な影響元の分析  |             | 選択した特定の値または値列に最も大きな影響を及ぼすデータ列を特定                                |
| カテゴリの検出    |             | 類似の特性を共有する行を検出  |
| 自動推論       |             | 選択した列内で不足しているデータ値を検出し、データのパターンに基づいて新しい値を推論                      |
| 例の全体適用     |             | ある列に部分的に入力されている例を全体に適用する  |
| 予測         |             | 与えられた一連のデータから将来の値を予測  |
| 例外の強調表示    |             | データの列から、検出されたパターンと一致しない値を検出                                     |
| シナリオ分析     | ゴール シーク ツール | ターゲット値を指定すると、そのターゲットを満たすために変更する必要がある基本要因が特定される                  |
|            | What-If ツール | 既存の値を操作し、その変更が結果に与えると見られる影響を調べることができる                           |
| 予測計算       |             | 過去のパターンから導き出された基準に基づいてケースを評価するための対話型ワークシートを作成し、意思決定に使用できるスコアを計算 |
| 買い物かご分析    |             | 併せて購入される頻度が高い複数の製品を見つけ出して、クロスセルやアップセルで利用できるパターンを特定              |



Microsoft

# Appendix

# Demo 1 に必要な環境

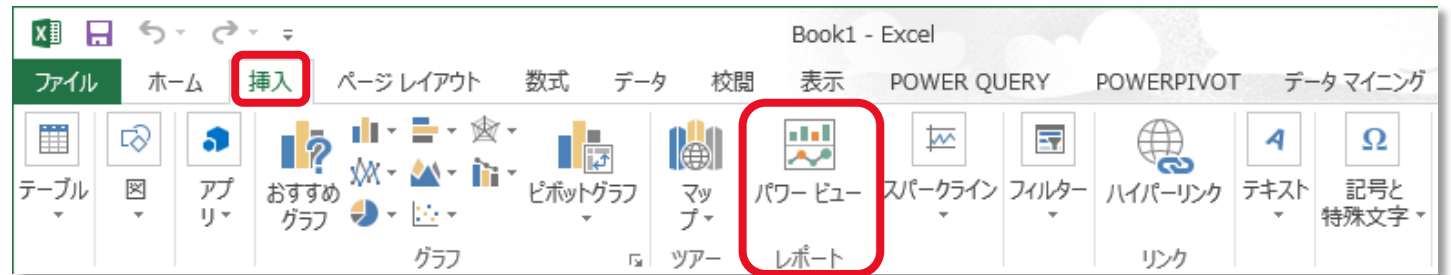
- サービス提供側：Power BI for Office 365, Power View を含む Excel ファイル
- 利用者：Internet Explorer (10 or 11) または Power BI Windows ストア アプリ

- 下記サイトより Power BI for Office 365 のトライアルを申込み

<http://www.microsoft.com/ja-jp/office/2013/business/powerbi/pricing/default.aspx>

- Power View を含む Excel ファイルの作成には以下のいずれかが必要

- Office 365 ProPlus
- Office Professional Plus 2013
- Excel 2013 スタンドアロン



※ Windows RT PC 上の Office では利用不可

※ Silverlight 5.0 のインストールと 2 GB 以上の RAM が必要

# Demo1 予測用レポートの作り方

- Excel 2013 で Power View シートを作成する

<http://office.microsoft.com/ja-jp/excel-help/HA102899553.aspx>

- 予測用レポート (折れ線グラフ) 作成における注意点
  - X 軸は日付または時間書式か、均一に増える整数のデータ (文字や小数は不可)
  - 折れ線グラフは1本のみ  
(複数折れ線は動作しない)
  - 折れ線グラフに含むことのできる  
データの数は 1,000 未満
  - 直近の 4 つの X 軸データは、均一に増加
  - X 軸のデータ間隔は最小で 1 日



# Demo 2 に必要な環境 (サーバー)

- SQL Server 2014 の評価版をダウンロード

<http://technet.microsoft.com/ja-jp/evalcenter/dn205291.aspx>

- SQL Server Analysis Services (SSAS) のインストール

- 既定値を使用して Analysis Services をインストールすると、多次元データベースとデータマイニングモデルを実行するインスタンスがインストールされる
- インストールの詳細は下記参照

<http://msdn.microsoft.com/ja-JP/library/ms143708.aspx>



# Demo 2 に必要な環境 (クライアント)

- Office 2010/2013 用のデータ マイニング アドインをダウンロード  
<http://www.microsoft.com/ja-jp/download/details.aspx?id=35578>
- SQL Server Analysis Services (SSAS) への接続

リボンメニューの [データ マイニング] にて [ヘルプ] → [作業の開始] で接続先の SSAS を指定すると構成ウィザードが開始されるので、接続先の SSAS にデータマイニング用のデータベースを作成する



# (参考) 構成ウィザード

Office 2010 データマイニング アドイン用に SQL Server を構成します

▶ **SQL Server 2012 データマイニング アドイン構成ウィザードへようこそ**

このウィザードを使用すると、SQL Server 2012 Analysis Services インスタンスを構成して、Office 2010 用 SQL Server 2012 データマイニング アドインに必要なサーバー機能を有効にできます。

SQL Server 2012 データマイニング アドインは、組織内の Excel 2010 および Visio 2010 ユーザーに対して、SQL Server 2012 Analysis Services に組み込まれた強力なデータマイニング機能を使用する魅力的な分析機能を提供します。

SQL Server 2012 Analysis Services インスタンスの選択および構成を続行するには [次へ] をクリックし、ウィザードを終了するには [キャンセル] をクリックします。構成しようとしている Analysis Services インスタンスに対してサーバー管理者の権限が必要であることに注意してください。

Office 2010 データマイニング アドイン用に SQL Server を構成します

**手順 1/4: Analysis Services データベースに接続する**  
Office 2010 用データマイニング アドインをサポートするように構成するサーバー インスタンスを選択します

サーバー名(E): localhost

認証(I): Windows 認証

Office 2010 データマイニング アドイン用に SQL Server を構成します

**手順 2/4: 一時的なマイニング モデルの作成を許可する**

Excel 2010 用テール分析ツール アドインは、一時的なモデルの作成を許可する SQL Server 2008 (またはそれ以降の) Analysis Services インスタンスに接続しないと機能しません。一時的なモデルは、ユーザーが接続を閉じたときにインスタンスから削除されます。

サーバーは既にセッションのマイニング モデルを許可しています。

現在の設定を変更して一時的なマイニング モデルの作成を禁止する場合は、以下のチェック ボックスをオフにしてください。

☒ 一時的なマイニング モデルの作成を許可する(C)

Office 2010 データマイニング アドイン用に SQL Server を構成します

**手順 3/4: アドイン ユーザー用のデータベースを作成する**  
アドイン ユーザーが自由に使用できる独立したデータベースを作成または選択します

Office 2010 用データマイニング アドインのユーザー用に新しいデータベースを作成すると、ユーザーは一時的なモデルや永続的なモデルを自由に作成できます。また、他のデータベースに影響を及ぼすことなくこれらのユーザーのオブジェクト権限を管理できるようになります。

☐ 新しいデータベースを作成する(C)

データベース名(D): DMAddinsDB

☒ 既存のデータベースを使用する(U)

データベース(I): DMAddinsDB

Office 2010 データマイニング アドイン用に SQL Server を構成します

**手順 4/4: アドイン ユーザーに適切な権限を設定する**  
アドイン ユーザーがサーバー データベースに永続的なモデルを作成できるようにします

サーバーで永続的なモデルを作成するには、ユーザーはデータベース管理者である必要があります。

特定のデータベース内のデータベース管理者権限をユーザーに許可すると、そのユーザーは、新しいオブジェクトを作成するだけでなく、既存のオブジェクト(ユーザーがそのデータベース内で作成したオブジェクトを含む)を変更および削除することもできます。ただし、ユーザーは、他のデータベース内のオブジェクトに影響を与えることはできません。権限を分離するために、追加で各ユーザーに個別のデータベースを作成することもできます。

前の手順で既存のデータベースを選択した場合、選択したユーザーがデータベース内の既存のオブジェクトに加え、他のユーザーが作成した新しいオブジェクトも管理できることが適切かどうかを確認してください。

☒ アドイン ユーザーにデータベース管理権限を与える(G)

データベース名: DMAddinsDB

追加するユーザーまたはグループを指定します(S):

APRIL2010user

追加(D)... 削除(R)