

2023年1月31日 全6頁

# 人事データ分析を始めよう

コンサルティング第三部  
コンサルタント 内野 慈

## [要約]

- 人的資本可視化の足掛かりとして、人事関連データから始められる簡易的な人事データ分析を紹介する。
- データ分析は1回分析して完了ではなく、何度もデータ分析のサイクルを実行することで、より効果を発揮する。スモールスタートで素早く分析を開始し、そこから得られた結果などをもとに、分析範囲を広げることも有効な手法となる。

## 1. 人的資本の有効活用に向けて

人的資本などの非財務情報の有効活用に向け、人的資本可視化の流れが加速している。内閣官房「新しい資本主義実現会議」内の非財務情報可視化研究会より2022年8月30日に「人的資本可視化指針」が発表されており、今後も人的資本可視化の重要性は増していくことが想定される。本稿では、人的資本可視化の足掛かりとして、人事関連データから始められる簡易的な人事データ分析を紹介する。データ分析に向け必要なステップや出力結果の見方に焦点を当て、データ分析という言葉に感じるハードルを少しでも下げられればと考えている。

## 2. 人事データの分析

始めに準備すべき人事データの種類について紹介する。準備すべきデータは大きく、「従業員属性情報データ」と「エンゲージメントデータ」に分けられる。「従業員属性情報データ」とは、年齢、性別、役職など従業員個人ごとの属性を表すデータである。「エンゲージメントデータ」とは、仕事へのやりがい、職場環境への考え等、一般的にはエンゲージメント調査、アンケート、自己申告などで取得しているデータである。それぞれのデータ例を図表1に示す。「従業員属性情報データ」と「エンゲージメントデータ」は両方揃っていることが望ましいが、それぞれ単独のデータでも分析は可能である。

(図表 1) 従業員属性情報データ、エンゲージメントデータ例

## 従業員属性情報データ例

・年齢	・勤続年数
・役職	・性別
・採用（新卒/中途）	・最終学歴（高卒/大卒など）
・職種（総合/エリア職など）	・人事評価
・所属組織（事業所）	・残業時間
・基本給与	・人事異動回数
・休暇取得状況	

## エンゲージメントデータ例

・仕事のやりがい	・職場環境
・理想のキャリア像	・給与形態への考え
・評価についての考え	・会社への愛着度
・尊敬する上司の有無	・ストレス状況

出所：大和総研作成

『可視化から始める人的資本分析 基本的な人事データ可視化手法 2021年11月10日』<sup>1</sup>では人事データに対する分析手法としての5ステップを紹介している。本稿ではこの5ステップに則り、人事データの分析手法を紹介する。

#### ① 分析目的の明確化・仮説の構築：目的、ゴールの明確化

本稿ではデータ分析の目的を「従業員の退職要因把握」に設定する。分析に利用するデータより、退職要因の把握及びリテンション施策に繋げることを最終ゴールとする。

#### ② データ収集：従業員属性情報データ、エンゲージメントデータ等

分析データについては、kaggle<sup>2</sup>上に公開されている『IBM HR Analytics Employee Attrition & Performance』<sup>3</sup>を用いる（図表2のような従業員属性情報データ及びエンゲージメントデータ合計34項目を持つ1470人分の従業員情報）。本データは、分析方法を紹介するために用いるものであり、日本企業の退職傾向を示すものでないことを注意されたい。

<sup>1</sup> 大和総研コンサルティングレポート 以下 URL 参照

[https://www.dir.co.jp/report/consulting/hr/20211110\\_022632.pdf](https://www.dir.co.jp/report/consulting/hr/20211110_022632.pdf)

<sup>2</sup> Google 社が運営しているデータ分析コンペティションプラットフォーム

<sup>3</sup> 以下 URL 参照

<https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>

(図表 2) 分析で利用するデータ

## 従業員属性情報データ

- ・Attrition (退職状況)
- ・Age (年齢)
- ・Gender (性別)
- ・TotalWorkingYears (勤続年数)
- ・MonthlyIncome (月給)
- ・OverTime (残業)
- など

## エンゲージメントデータ

- ・WorkLifeBalance (ワークライフバランス)
- ・EnvironmentSatisfaction (職場環境満足度)
- ・JobSatisfaction (職務満足度)
- など

出所：『IBM HR Analytics Employee Attrition & Performance』のデータセットをもとに大和総研作成

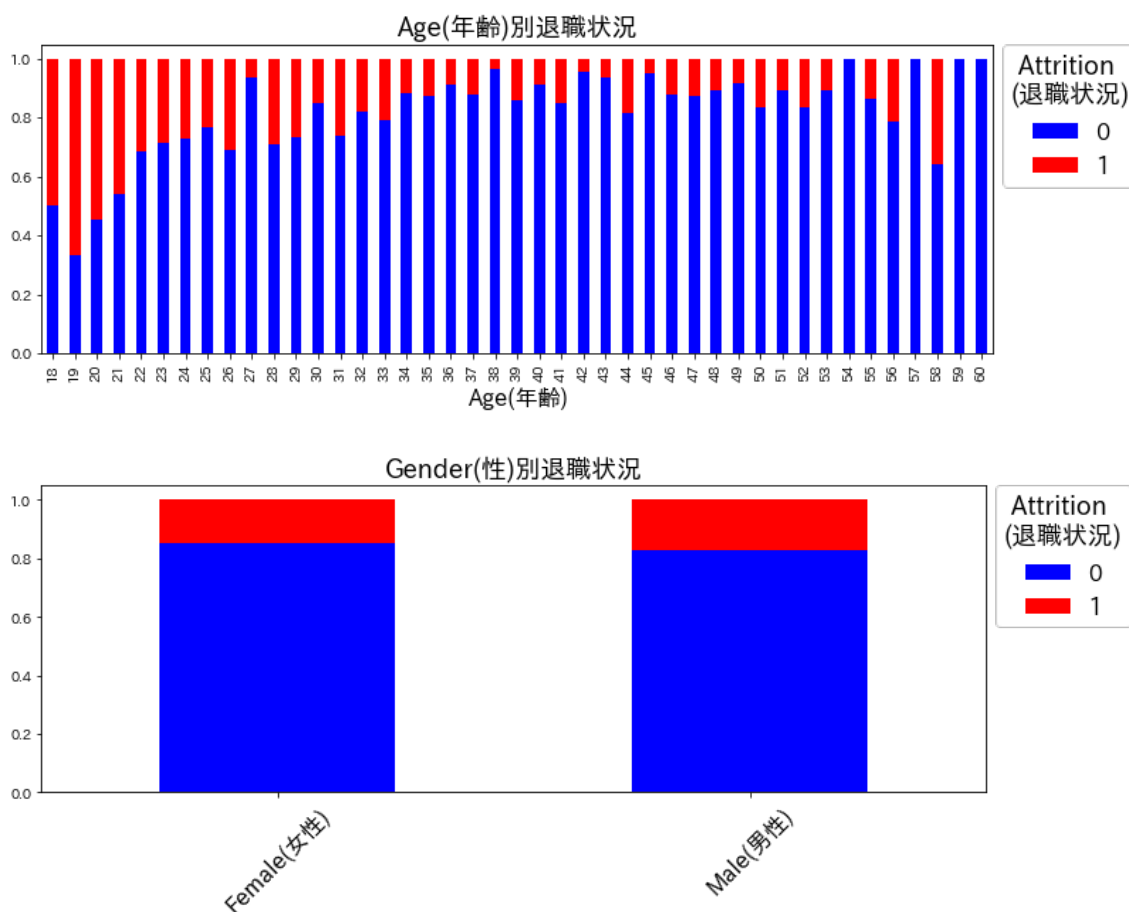
## ③ 現状の可視化：ダッシュボード化、グラフ化等

分析目的は「従業員の退職要因の把握」であるため、各属性と退職状況の関係性の可視化が有効となる。各属性と退職状況を可視化する際は 100%積み上げ棒グラフ(図表 3) が視覚的にとらえやすい。図表 3 の上部グラフの横軸は Age (年齢)、縦軸は各年齢での退職割合を 100%として、青色が在職割合、赤色が退職割合を表している。グラフより年齢が低いほど退職割合が高く、年齢が高くなるにつれてその割合が下がる傾向が読み取れる。図表 3 の下部グラフの横軸は Gender (性別)、縦軸は各性別での在籍割合、退職割合を表している。グラフより male (男性) と female (女性) で退職割合に大きな差がないことが読み取れる。この 2 つのグラフを比べると、Gender (性別) より Age (年齢) の方が退職に影響を与えていると考えられる。他属性でも同様の可視化及び比較を行うことで、退職に影響を与えている属性を調査できるが、見た目だけでは判別できないケースも想定される。次のステップでは可視化に留まらず、各属性と退職状況の関係性の分析手法について紹介する。

## ④ データ分析：統計学的手法

各属性と退職状況の関係性の分析として相関分析を紹介する。相関分析とは 2 つの属性間にある関係性を「-1~1」の数値(以下、本数値を相関係数と記載する)で表し、関係性の強さを図る手法である。相関係数が正の値を取るとき正の相関があると言い、値が 1 に近いほど各属性と退職との関係性は強くなる。退職と正の相関がある属性としては、OverTime (残業) が挙げられる。一般的に OverTime (残業) が長いほど、退職する傾向がみられる。相関係数が負の値を取る場合は、負の相関があると言い、値が-1 に近いほど、退職との関係性は逆の動きとなる。退職と負の相関がある属性としては、Age (年齢) が挙げられる。一般的に Age (年齢) が高いほど、退職せず、在籍し続ける傾向がみられる。

(図表 3) 各属性の退職状況の可視化 (100%積み上げ棒グラフ)

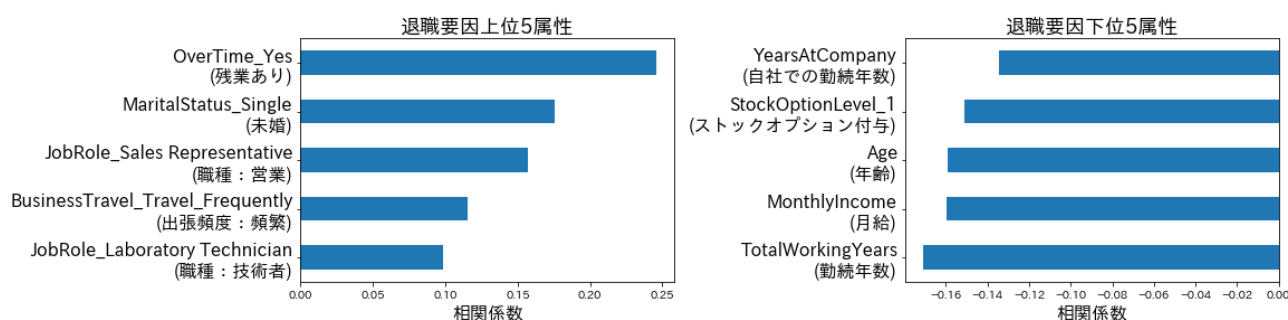


(注) Attritionの「1 (赤色)」が退職、「0 (青色)」が在職を表す。

出所：『IBM HR Analytics Employee Attrition & Performance』のデータセットをもとに大和総研作成

今回分析対象としているデータでの退職上位、下位の5属性を図表4にまとめる。図より、退職に正の相関がある属性として、「OverTime\_Yes (残業あり)」、「MaritalStatus\_Single (未婚)」、「JobRole\_Sales Representative (職種：営業)」、退職に負の相関がある属性として「TotalWorkingYears (勤続年数)」、「MonthlyIncome (月給)」、「Age (年齢)」が上位であることが分かった。退職と各属性の相関係数をグラフ化することで、退職にどの属性が影響を与えているかを視覚的に確認することができる。ここまでくれば、目的とした退職要因の洗い出しはひとまず完了である。

(図表 4) 相関分析による退職要因上位/下位属性

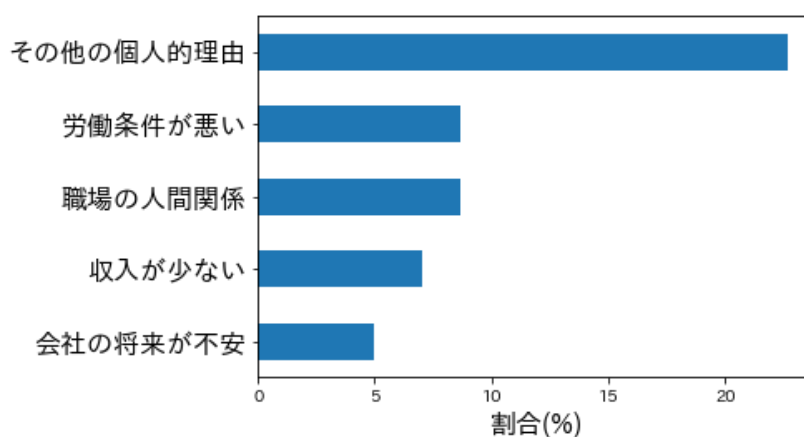


出所：『IBM HR Analytics Employee Attrition & Performance』のデータセットをもとに大和総研作成

次の分析としてはいくつか考えられる。1つは厚生労働省の雇用動向調査<sup>4</sup>でまとめられている退職理由をベンチマークとして、自社の退職要因上位属性と比較することで、自社特有の退職要因を洗い出すことである。図表 5 に令和 3 年の雇用動向調査結果の勤め先を辞めた理由（個人的理由）上位 5 属性をまとめる。図表 5 の属性と自社の退職要因上位属性を比較し、自社のみにある属性があれば、その属性が自社特有の退職要因の可能性があるので、優先してリテンション施策を立てる候補としてピックアップできる。

他には、従業員属性情報データとエンゲージメントデータだけでなく、営業成績データなど別の切り口のデータを追加して再度、要因分析を実施することなども考えられる。また、発展的な分析としては、退職者予測を行うことなども考えられるが、本稿では次のステップでの簡単な紹介に留める。

(図表 5) 厚生労働省「令和 3 年雇用動向調査結果」の退職理由上位 5 属性



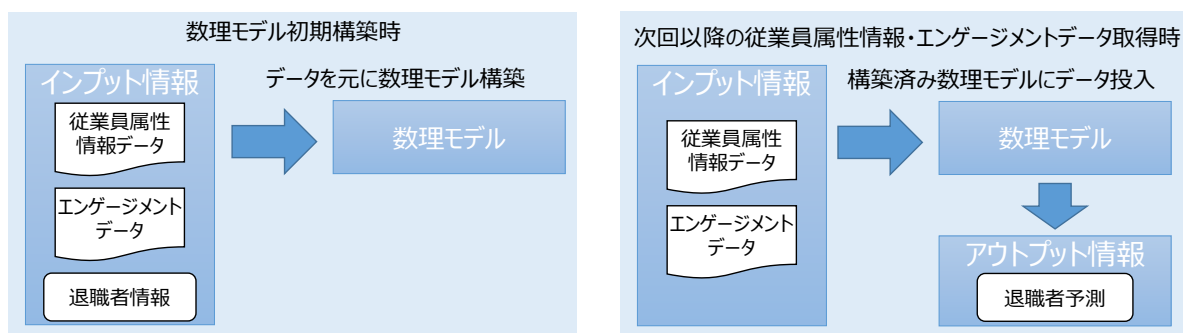
出所：厚生労働省「令和 3 年雇用動向調査結果」をもとに大和総研作成

<sup>4</sup> 厚生労働省 雇用動向調査  
<https://www.mhlw.go.jp/toukei/list/9-23-1.html>

### ⑤ データ予測：統計学的手法、機械学習手法による将来シミュレーション

機械学習などの手法を用いて退職に関する数理モデルを構築することで、退職者予測（退職リスクが高い従業員の洗い出し）を行うことができる。具体的な手段としては、従業員属性情報データ、エンゲージメントデータ、退職者の情報をインプットとして、Python や R 言語を用いて数理モデルを構築する（図表 6 左のイメージ図参照）。一般的に退職予測の数理モデル構築時は従業員の退職者情報をインプット情報とする必要があることに注意されたい。そのため、初期構築時は過去の退職者データ等を用いて構築していく。そして、次に従業員・エンゲージメントデータを取得するタイミングで、構築した数理モデルを用い、これから退職する可能性が高い従業員を予測していく（図表 6 右のイメージ図参照）。退職者予測が行えるようになることで、退職リスクが高い従業員に対して、個別にアプローチなどを行うなどのリテンション施策が取れるようになる。

（図表 6）数理モデル構築による退職者予測のイメージ図



出所：大和総研作成

## 3. おわりに

本稿では人事データ分析として、退職要因の分析に焦点を当てたデータ分析手法を紹介した。今回紹介した 5 つの分析ステップを 1 度実行して完了ではないことに注意いただきたい。データ分析は 1 サイクルの分析で完了ではなく、何度もデータ分析のサイクルを実行することで、より効果を発揮する。またデータを集める、整備することに時間を掛けすぎず、スモールスタートで素早く分析を開始し、そこから得られた結果などをもとに、分析範囲を広げることも有効な手法となる。

—以上—