

データ活用の取組みへの第一歩 ～内閣府「景気ウォッチャー調査」のテキスト分析の実例～

筑波総研株式会社 企画調査部 研究員 金田 憲一

はじめに

データやデジタル技術を駆使して、ビジネスに変革をもたらすデジタルトランスフォーメーション（DX）がビジネス用語として定着するなかで、企業におけるデータの活用や分析の重要性がこれまで以上に高まっています。

ビジネスにおいてデータの活用が本格化したのは、インターネットの普及以降です。検索・閲覧履歴にもとづいた商品広告にはじまり、現在ではデータを活用し需要に応じて価格を変動させる「ダイナミックプライシング」などの価格戦略も登場しています。また、スマートフォンや時計型デバイスなどの登場で、私たちの生活利便性はさらに向上していますが、一部の企業では、これらのデバイスから得られる位置情報や生体・生活情報などを活用して収益を上げるなど、様々なデータの経済的価値が高まっています。

一方で、商用的な活用以外に、企業経営においてもデータの活用が重要であると言われてしています。

不確実性や複雑性が高まる現代では、経験則に基づいた判断が通用しにくく、事業環境の未来予測が難しくなっています。そしてこの状況は、コロナ禍でさらに加速しました。こうしたなか、事業を成功に導いていくカギとなるのがデータです。データを分析することによって得られた客観的指標に基づき、自社の課題や事業環境を「見える化」し、把握することが、将来の事業計画を策定していく上で大きなヒントになります。

しかし、分析の仕方がわからない、効果がイメージできない、データ人材がいない、といった理由から、本格的なデータの活用に踏み出せている企業は一部にとどまっています。

データ分析のソースには、売上・客数などの数値以外にもテキスト、画像などがあります。本稿では、文字列から有益な情報を抽出する「テキストマイニング」と呼ばれる手法を用いて、内閣府「景気ウォッチャー調査」のコメントをもとに、データ分析の仕方や効果の一例を示したいと思います。

1. わが国はデータの活用が遅れている

スイスの国際経営開発研究所（IMD）が毎年公表している「世界主要各国のデジタル競争力ランキング」によれば、2021年における日本の順位は64か国・地域中28位と、前年から1ランク下がり過去最低順位を更新しています。

ランキングでは、デジタル競争力に影響を与える3つの要因（「知識」「技術」「将来への備え」）が定められていますが、日本の順位が特に低いのが、「知識」の項目における「人材」（47位）、「将来への備え」の項目における「ビジネスの俊敏性」（53位）となっています（図1）。

さらに詳細に内訳をみると、「デジタル技術スキル」や「企業の俊敏性」、「ビッグデータの分析と活用」などの項目が特に低い順位となっており、わが国ではデータの活用が遅れています（表1）。

図1 2021年の日本のデジタル競争力ランキング

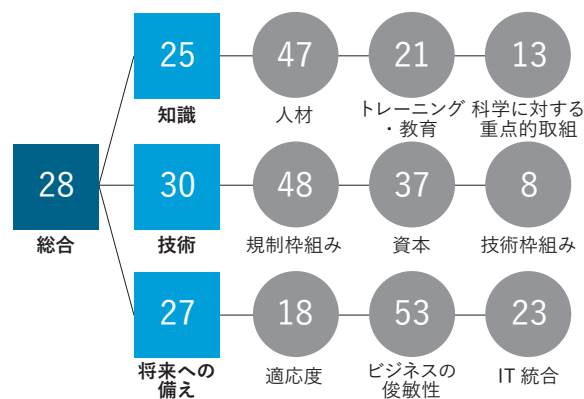


表1 「人材」「ビジネスの俊敏性」の内訳

	順位		順位
人材	47	ビジネスの俊敏性	53
教育評価	5	機会と脅威	62
国際経験	64	世界のロボット分布	2
外国人高度技術者	49	企業の俊敏性	64
都市管理	15	ビッグデータの分析と活用	63
デジタル技術スキル	62	知識移転	40
留学生	26	起業家の失敗への恐れ	33

出所：IMD「IMD WORLD DIGITAL COMPETITIVENESS RANKING 2021」より筆者作成

わが国ではこうした課題を解決するため、デジタルの力を活用して地方活性化を図る「デジタル田園都市国家構想」の基本方針のなかで、政府の各種施策を通じて、デジタル人材を2022年度から26年度までの5年間で230万人育成するとしています。

同時に、2021年12月に開催された同構想実現会議のなかで、「民間企業等においても、政府の施策を活用しつつ、独自のデジタル推進人材の育成・確保の取組みを進めることが期待される」としており、政府の取組みと同時に、民間企業においてもデジタル化への取組みを進めることを期待しています。

2. 本稿での計量テキスト分析の概要

データの活用と言われても、分析の仕方がわからない、得られる効果についてもイメージが湧きにくいと思います。

そこで今回は、データ分析の一例として、視覚的にも効果がわかりやすい、テキストマイニングの一種である計量テキスト分析を紹介したいと思います。

以下では、今回の分析に用いた、内閣府「景気ウォッチャー調査」の概要とテキスト分析のメリットについて整理します。

(1) 内閣府「景気ウォッチャー調査」の概要

内閣府が毎月公表している「景気ウォッチャー調査」は、地域ごとの景気動向を的確かつ迅速に把握し、景気動向判断の基礎資料とすることを目的とした経済統計です。調査は毎月25日～月末に実施しており、その結果は翌月上旬に公表されるなど、速報性があります。

調査方法は、全国の小売業の販売員やタクシー運転手など、景気の動きを観察できる業種の就業者（景気ウォッチャー）が、現状と先行きの景況感を、「1良い」「2やや良い」「3どちらとも言えない」「4やや悪い」「5悪い」の5段階で評価します。それぞれの回答は指数（DI）に集約され、DIが50を上回れば「景気が良い」、下回れば「景気が悪い」と回答した人が多いことを意味します。

同調査では、数値の結果に加え、「景気判断理由集」（表2）として、回答者からのコメントを公表しており、現場のリアルな声を確認することができます。ただし、そのコメント数は単月分で1,000を超えるため、内容の定量的な把握は簡単なことではありません。

表2 分析に使用する景気判断理由集の一部

分野	景気の状態判断	業種・職種	判断の理由	追加説明及び具体的状況の説明
家計動向関連	◎	一般小売店 [土産] (経営者)	販売量の動き	3月でまん延防止等重点措置が全国的に解除されたことに加えて、春らしい陽気となったことで前年の2倍ほどの人出がみられ、売上もコロナ禍前の50%を上回る日がみられるところまでばん回している。4月の売上は前年比で261%、前々年比で441%、2019年比で25.3%となっている。
	○	商店街 (代表者)	お客様の様子	まん延防止等重点措置が解除されたことで徐々に人出が増え、回復基調に入ってきた。ただし、新規感染者数が高止まりしていることもあって、満足な購買状況にはなく、回復までには相当な時間が掛かりそう。そのようななか、経済回復を早期に実現しなければならぬことから、感染症対策を万全にした上で4月28日からイベントを開催することになった。行政に頼るだけではなく、民間側がイベントを安全に実施する見本をみせていくことで、消費者の閉ざされた気持ちを明るくしていきたい。これからもそのようなイベントを多数実施していきたい。
	○	商店街 (代表者)	販売量の動き	りんごの出荷が増え始めた。気温上昇に伴って当社商材の出荷も増えている。
	○	一般小売店 [酒] (経営者)	販売量の動き	少しずつコロナ禍前の売上に近づいているが、まだまだ予断を許さない状況にある。
	○	百貨店 (売場主任)	来客数の動き	前月、前々月は週によって来客数の増えた週もみられたが、新型コロナウイルスの影響を受けていた。4月は新規感染者数が増えても、来客数にそれほど影響がみられなくなり、前月や前々月と比べて来客数が安定している。
	○	百貨店 (販売促進担当)	来客数の動き	気温の上昇とともに、駅周辺の来街者、店舗の来客数が増加している。それに伴って買上客数、売上も増加傾向にある。
	○	スーパー (店長)	それ以外	新型コロナウイルスの感染状況によるよしあしがみられる。3か月前はまん延防止等重点措置の適用で客が大きく落ち込んだが、4月はまん延防止等重点措置が解除され、小康状態にあるため、景気がやや良くなっている。
	○	スーパー (店長)	お客様の様子	1月末～3月中旬までまん延防止等重点措置が適用されていたことを踏まえても、4月の客の消費金額は確実に増えてきている。コロナ禍前ほどではないが、衣料品の売上も上向きになっており、今後の政府の施策次第で一層の伸長も期待できる。

出所：内閣府「景気ウォッチャー調査」より抜粋

(2) テキスト分析のメリット

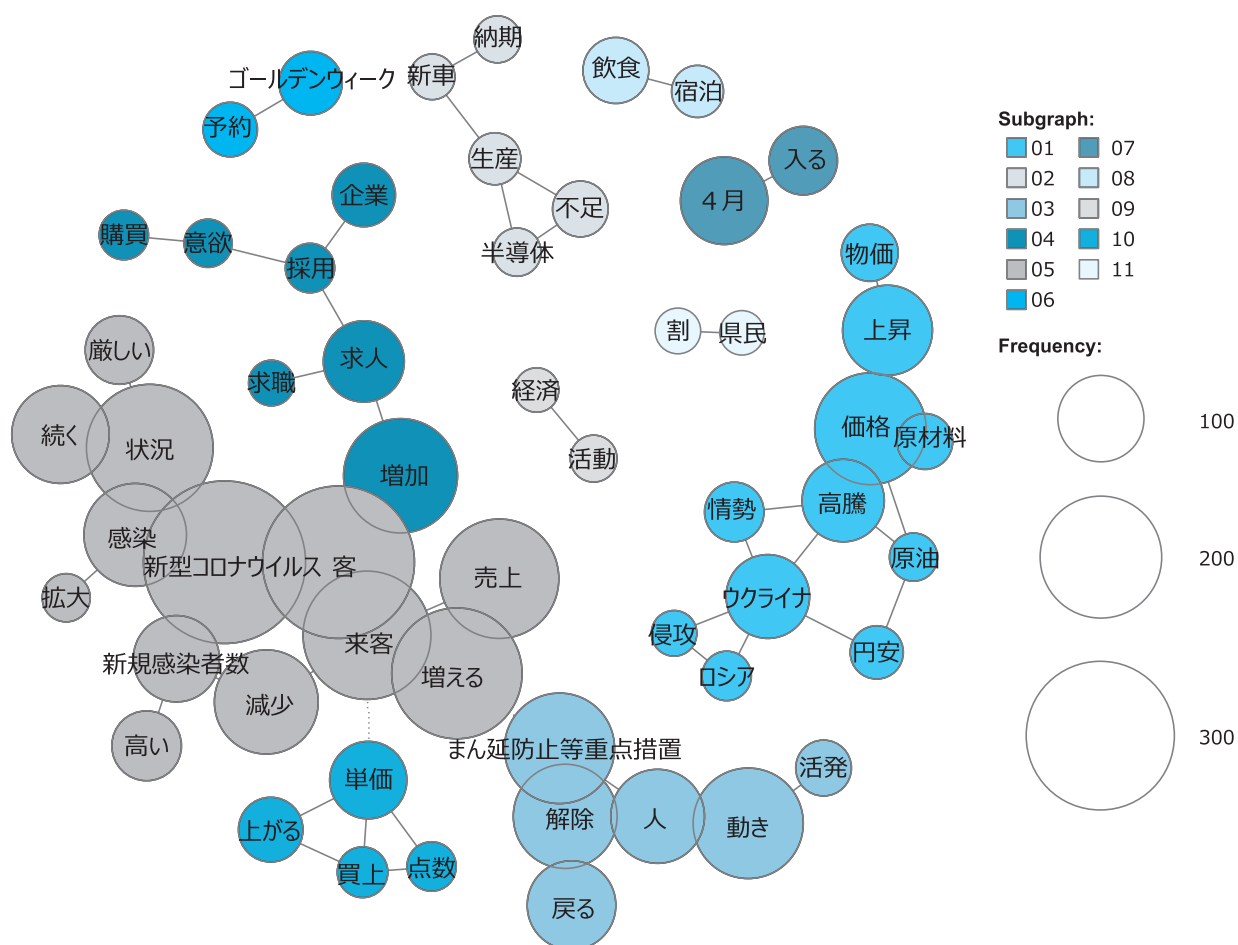
1,000を超えるコメントすべてに目を通したとしても、その全体像を捉えることは難しいうえに、そこから何が言えるかをまとめることは容易ではありません。仮にコメントを読んで何かしらの傾向を言えたとしても、自分にとって印象の強かった少数のコメントに左右されているなどバイアス（偏り）が生じている可能性もあります。

また、「景気が良いと回答した人はどのような語を多く使用しているか」や「業種によってどのような語の使用数が多いか」など、変数とテキストの関係性を捉えることはさらに困難になります。

こうした課題を解決してくれるのが、計量テキスト分析です。計量テキスト分析を行うメリットとして、第一に、分析をする人自身が正確にデータを捉え、データをより理解できるようになります。また、統計的な情報を使って、読む価値のある特徴的なコメントを検索することも可能になります。第二に、分析結果が客観的データで示されるため、結果の共有のしやすさや信頼性が向上します。

本稿の分析では、フリーソフトウェアである「KH Coder」を使用しました。同ソフトを用いた研究事例は5,000件を超えており、多くの官公庁や企業でも利用されています。

図4 共起ネットワーク (2022年4月・現状判断)



出所：内閣府「景気ウォッチャー調査」より筆者作成

の多くを占めていることがわかります。1月は、年始以降にオミクロン株の拡大により新型コロナの新規感染者数が急激に増加し、中旬以降に各地で「まん延防止等重点措置」が適用されました。「外出・自粛」といった語のグループや「宿泊・予約・キャンセル・相次ぐ」「来客・売上・減少」などのグループからもわかるように、マイナスの話題が多くを占めています。

そのほか、半導体など部品不足に伴う企業の生産活動への影響、原材料やガソリン価格の高騰などに関連する語が確認できますが、その数は「新型コロナウイルス」に比べると多くありません。

次に、図4の4月調査の結果をみると、「新型コロナウイルス」は300語台まで減り、話題が分散している様子が確認できます。

この間の国内の様子を振り返ると、新型コロナの新規感染者数が減少するなかで、3月22日から全国的にまん延防止等重点措置が解除されました。それに伴い、人の動きの活発化やゴールデンウィークの

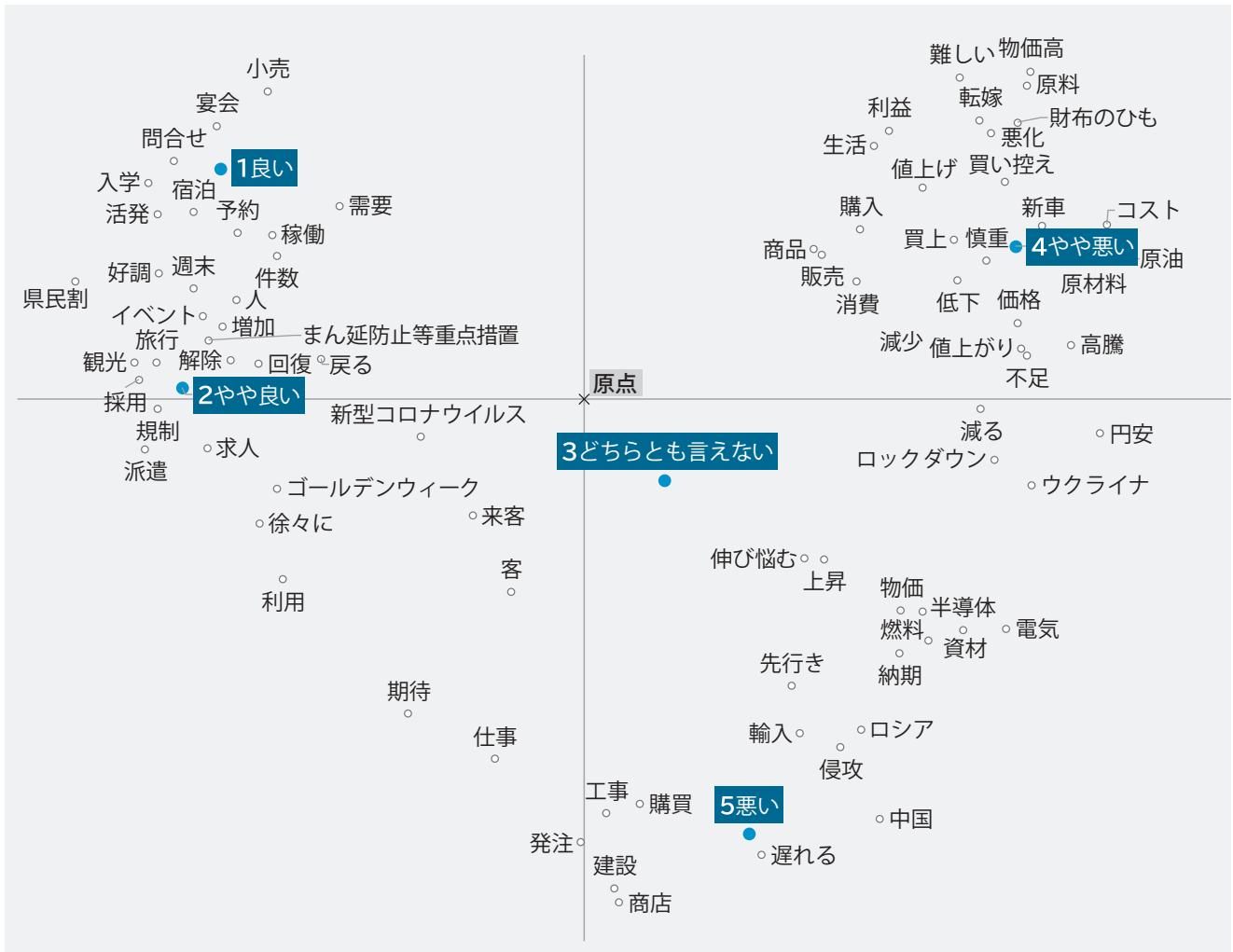
予約に関するプラスの話題が多く上がっています。

一方で、感染者数の水準は過去に比べると高止まりしていたため、業況の厳しい状況が続いている様子も確認できます。図3の同調査のDI値が、好不況の分かれ目である50前後であることから、判断が分かれる結果となっています。

また、2月24日に開始したロシアによるウクライナへの軍事侵攻が新たな話題として上がっています。それに伴うさらなる原材料・原油価格の上昇による物価高騰、さらに、半導体不足による新車の納期遅れなど、経済にとって下押しのリスクとなるような語が多くみられます。

このように共起ネットワークからは、1,000を超えるコメントの全体像の把握が可能であり、図を並べることで時系列の比較がビジュアルでも確認できます。また、すべてのコメントをもとにして作られたデータであるため、自分で数千にわたるコメントを読んでそれぞれの結果の特徴を説明するよりも、説明に客観性を持たせることが可能です。

図5 景気判断を変数に使用した対応分析の結果(2022年4月・現状判断)



出所：内閣府「景気ウォッチャー調査」より筆者作成

(2) 景気判断別にみた対応分析

共起ネットワークからは、コメントで取り上げられている主な話題を確認することができましたが、多く出現している語が良い意味で使われているのか、あるいは、悪い意味で使われているのかまでは読み取ることはできません。例えば、P17の図4で「ゴールデンウィーク」「予約」という語だけをみても、ゴールデンウィークの予約が多くて景気が「良い」と判断しているのか、少なくて景気が「悪い」と判断しているのかを確認することはできません。

そこで、必要となるのが「対応分析」です。対応分析は、コメントの内容と変数の関係を明らかにすることが可能で、今回は変数に「良い」～「悪い」の5段階の景気判断を使用しています。対応分析を行うことで、たとえば、「景気が良いと判断をした人はこんな語を使用している」といった特徴がわかります。なお、分析の際は、より多くの語を抽出できるように設定すると、データの特

徴をつかみやすくなります。

図5が4月調査のコメントから作成した対応分析の結果です。図では、中央にある原点からみて、どの語がどの変数の方向にあるかをチェックすることで、その変数でよく使用されている特徴的な語を読み取ることができます。そして、原点からの距離が遠いほど強い特徴を持っています。一方で、原点に近い位置にある語はあまり特徴がなく、変数に関係なく満遍なく出現している語になります。

先ほど例で示した、「ゴールデンウィーク」という語をみると「やや良い」の方向にあることがわかります。また、「良い」の方向も含めてみると、「宿泊」「予約」「県民割」「観光」「旅行」「宴会」といった飲食・宿泊などに関連が深い語もあり、これらの語が景気判断の好材料として使用されていることがわかります。厳密に言うと、景気が良いと判断した方のコメントのなかで、これらの語が多く使用されている、ということになります。

一方で、「4 やや悪い」の方向をみてみると、「物

価高」「値上げ」「買い控え」「慎重」「財布のひも」などの語が確認できます。

こうした結果からは、飲食や宿泊業などでは行動制限措置の解除に伴い景気が「良い」という判断が多くみられた一方で、一部の小売業では、物価高による買い控えが景気判断の際の悪材料になっている、といった推察ができます。

また、「5悪い」の方向も含めてみると、「中国」「ロックダウン」や「建設」「工事」などの語が確認できます。3月末からの中国の一部都市における都市封鎖による工場稼働停止・物流網混乱が与える影響や、木材などの原材料価格高騰による建設業の業況悪化などが推察できます。

(3) コンセプトを使用した詳細な分析

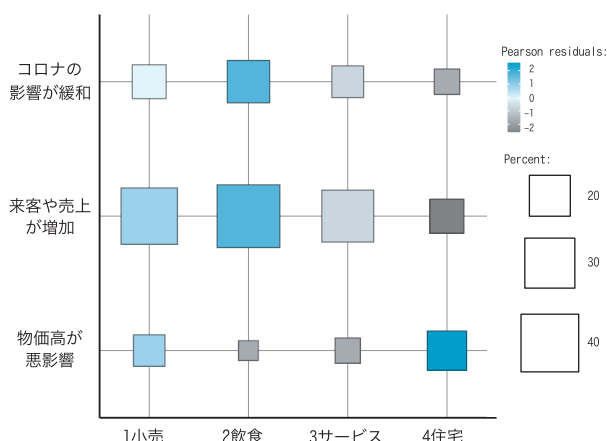
これまでの分析で、データ全体のおおよその特徴をつかむことができましたが、分析者自身が「コンセプト」を定めることでさらに詳細な分析が可能です。今回は、家計消費に関連する4つの業種「小売、飲食、サービス、住宅」に着目して分析したいと思います。

コンセプトは、意味の似た単語（同義語）の処理と、単語同士の組み合わせにより作成します。詳細な説明は省きますが、今回は、「コロナの影響が緩和」「来客や売上が増加」「物価高が悪影響」といったコンセプトを作成して分析しています。

図6の四角の大きさは、分析対象の業種毎のコメントが、作成したコンセプトにどのくらい当てはまっているかの割合を示しています。

結果をみると、「コロナの影響が緩和」のコンセプトでは、「飲食」が多く該当していますが、業種間の差はそれほど大きくありません。

図6 コンセプトを使用した分析



出所：内閣府「景気ウォッチャー調査」より筆者作成

次に「来客や売上が増加」では、「飲食」「小売」の順に割合が高い一方で、「住宅」では低くなっています。また、マイナスのコンセプトである「物価高が悪影響」では、「住宅」「小売」の順に割合が高く、とくに住宅での悪影響が大きくなっています。

3. (2) の対応分析の最後で、小売業における物価高による買い控えや原材料費高騰による建設業の業況悪化の可能性を示しましたが、コンセプトを用いた分析の結果からも、同様の結果が導き出せそうです。

4. おわりに

本稿では、景気ウォッチャー調査のコメントを用いて計量テキスト分析の実例を示しました。景気の「良い」・「悪い」の判断は、普段よく目にする経済ニュースでも把握ができるため、分析結果は目新しいものではなかったかもしれませんが、分析の流れや効果の一部を示せたと思います。

本稿での分析は以上となりますが、実際のデータ分析は、分析結果をもとに新たな仮説を設定して分析を繰り返すことが重要であり、さらに分析を進めていくことで新しい知見が得られるかもしれません。

企業におけるテキスト分析の活用先としては、ホテル・旅館などにおける利用者アンケートなどが上げられます。性別や年代別にデータ分析をすることで、自社の強み・弱みを認識し、改善につなげることが可能となります。

一方で、少し視点を変えると、社内にあるテキストデータからも分析は可能です。議事録や営業日誌など社内に蓄積されているデータを分析することで、自社の課題や業務の問題点など新たな発見があるかもしれません。また、作業報告書などのデータを分析して暗黙知になっている個人のノウハウを抽出することができれば、個人に依存していたスキルや経験を社内に共有することが可能となります。

今回使用したフリーソフトのKH Coderは、操作がわかりやすく結果も理解しやすいため、データ分析の第一歩として、分析の仕方や効果を実感できる優れたツールであると思います。

本稿が、データ活用の取組みへの第一歩としての一助となれば幸いです。

【参考文献】

樋口耕一 2004「テキスト型データの計量的分析 -2つのアプローチの峻別と統合-」『理論と方法』(数理社会学会) 19(1): 101-115
https://www.jstage.jst.go.jp/article/ojams/19/1/19_101/_pdf