

業務データを活用したレポートシステムの研究

株式会社メイテツコム

■ 執筆者 Profile ■



渡辺 心

2001 年 (株)メイテツコム入社

2014 年 現在 社会情報ソリューション部担当

■ 論文要旨 ■

昨今注目されているビッグデータの活用は、システム構築や統計学の知識などが壁になり、実践できないことが多い。本来データ分析で重要なことは、データから新たな価値を見出すことである。この研究では、交通系 IC カードである manaca のデータから新たな価値を見出すことを目的としたレポートシステムを構築した。構築にあたり①目的を達成できる機能、②システムを用いたワークフロー、③ソフトウェア、④分析に適したデータ設計、⑤価値を見出すためのレポート、以上を検討した。特にレポートについては様々なアプローチで作成、分析までを実践している。結果、IT ベンダーだけで分析を行い、新しい価値を求めることは厳しいが、ユーザの協力を得ることで成功に近づけることがわかった。ユーザとともにレポートシステムを作り上げていくことで新たな価値を見出すことを実現したい。

■ 論文目次 ■

<u>1. はじめに</u>	《 4》
<u>2. レポートシステムの目的</u>	《 4》
2. 1 名古屋鉄道の分析業務調査	
2. 2 レポートシステムの機能	
<u>3. レポートシステムの構築</u>	《 6》
3. 1 ソフトウェアの検討	
3. 2 レポートシステムについて	
<u>4. データ設計</u>	《 10》
<u>5. レポートの検討</u>	《 11》
5. 1 異なる項目の組み合わせ	
5. 2 レポート化と分析	
5. 3 レポートの表現	
5. 3. 1 散布図を利用したmanaca利用傾向の分析	
5. 3. 2 バルーンチャートを利用したバス乗降の分析	
<u>6. 考察</u>	《 18》
6. 1 データ分析の考察	
6. 2 レポートシステムの考察	
<u>7. 今後に向けて</u>	《 19》
<u>8. 最後に</u>	《 20》

■ 図表一覧 ■

図 1	レポートシステムのワークフロー	《 5》
図 2	RStudio-serverの画面	《 8》
図 3	システム構成	《 8》
図 4	レポートシステム画面構成	《 9》
図 5	Shinyプログラムの構成図	《 9》
図 6	分析対象データ	《 10》
図 7	分析用データ定義のイメージ	《 11》
図 8	駅時間帯別降車推移（曜日）	《 12》
図 9	駅時間帯別降車推移（年代）	《 13》
図 10	1カ月間のmanaca利用傾向の散布図	《 14》
図 11	1カ月間のmanaca利用傾向の分類	《 15》
図 12	系統別曜日ごとの乗車量	《 17》
図 13	系統別曜日ごとの降車量	《 17》
図 14	乗車人員数推移結果	《 19》
表 1	データの分類	《 11》
表 2	項目の組み合わせ	《 12》
表 3	事業者にとっての価値	《 12》
表 4	変動係数の例	《 15》
表 5	manaca利用傾向の分類に対する分析	《 16》

1. はじめに

昨今、IT 業界で注目されているキーワードと言えば「ビッグデータ」を挙げる方が多いだろう。このビッグデータという言葉自体には明確な定義は存在しないが、膨大な量のデータを複数のサーバで処理し、分析することをイメージするのではないだろうか。インターネット、スマートフォン、IC カードなどで得られた大量のデータを使って分析し、得られたものをビジネスに活用していくことがビッグデータの目的である

しかし、ビッグデータの活用をいざ実践しようにも、大規模なシステム構築、統計学の知識などが壁になり、実践できないことが多いのではないだろうか。確かに何テラバイトにも及ぶ大量のデータをリアルタイムに処理するのであれば大規模なシステム構築が必要になり、統計学の知識がないとできない分析もある。しかし、これらは手段、手法の一部であり、本来データ分析において最も重要なことは、データから新たな価値を見出し、ビジネスに活用していくことである。

メイテツコムでは 2012 年度から 2013 年度にかけて、NEDO（独立行政法人新エネルギー・産業技術総合開発機構）から支援を受け、交通事業者向けレポートシステムの研究を行った。この研究では、名古屋鉄道株式会社並びに名鉄バス株式会社の協力を得て、過去に発生した交通系 IC カードである manaca の乗降情報と購買履歴などを分析するレポートを作成している。また、作成したレポートを表示するシステムをオープンソースのみで構築し、データ分析を実践した。その中で、組み合わせたデータを可視化することにより、新たな価値を創出でき、そこから新しいビジネスが見えてくることを経験できた。

本論文では、レポートシステム研究で得た、データ分析の実践経験と、データ分析を行う環境作りの共有を目的とし、研究で実施した取り組みについて述べる。

2. レポートシステムの目的

本研究では静的属性・過去の購買結果に加えて、移動情報を組み合わせた価値あるレポートを作成し、提供することを目的とした。レポートシステムに求められているものは、実際に事業者が見て、理解し、次の行動につなげる事が出来る、いわゆる「PDCA (Plan Do Check Action)」サイクルを実現できるものである。この要件を満たすために、レポートシステムは IT に不慣れな人でも抵抗感なく利用できる事が求められる。そのため、これまで事業者内で分析などを行う際に利用されている一般的な形式や、広く認識されている形式である事が標準的なレポート機能として求められると想定した。

まず初めに、既に統計解析ツールを導入している名古屋鉄道に、分析業務の現状をヒアリングし、レポートシステムに必要な要素を調査する。

2. 1 名古屋鉄道の分析業務調査

統計解析ツールは主に、グラフィック、集計・検定・統計分析、レポートといった機能を備えている。統計解析ツールが提供している機能の中で、主にどの機能を利用しているかをヒアリングしたところ、「統計解析ツールには様々な機能があるが、その結果を次につなげることが難しい。毎月定例的に参照しているのは、manaca の購買情報の集計のみで

ある。グラフィックには、Excel を用いて皆が見てわかる集計情報のグラフを作成している」との情報を得た。統計解析ツールは豊富な機能を備えているが、使いこなすことが難しい。そのため、価値を見出す、もしくは提供するために最適なものを選択し都度提供していけるシステムを目指すことにした。この結果を踏まえ、次にレポートシステムの機能を定義する。

2.2 レポートシステムの機能

(1) グラフィック

Excel や統計解析ツールなどで提供されている一般的な機能を標準的なレポート機能とする。レポートの表現としては、表形式、各種グラフ、地図へのマッピングを扱い、PDF 形式、CSV 形式、画像ファイル(PNG 形式など) で出力できるようにする。

(2) レポート出力までの流れ

統計解析ツールは敷居が高く、ユーザ側にもある程度のスキルが求められる。このシステムではユーザが必要としているレポートを簡単に出力できるような流れを実現する。

以上の2つを実現できるレポートシステムを構築する。更に、先に述べた「価値を見出す、もしくは提供するのに最適なものを選択し都度提供していくシステム」を考慮しなければならない。この要件を満たすためのワークフローを検討した。ワークフローを図1に示す。

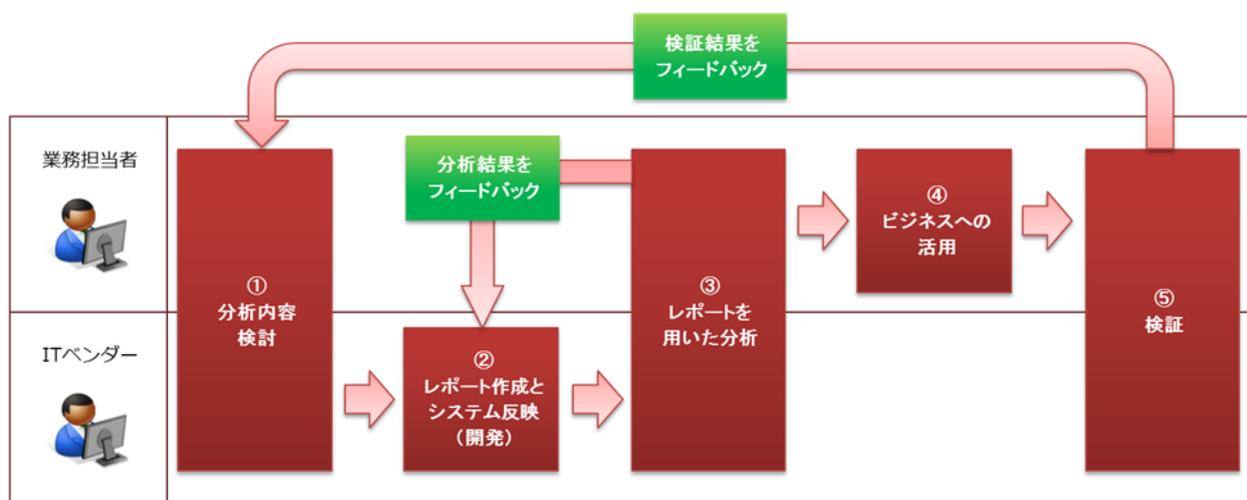


図1 レポートシステムのワークフロー

図1に示した各フェーズの中で、②、③がレポートシステムに関わる部分である。この部分はレポート作成と分析を行うわけだが、分析結果からレポートの修正、追加など、試行錯誤しながらフェーズを繰り返す必要がある。したがって、システムの機能以外に、②、③を短い期間で実施できる環境作りを行わなければならない。この部分は、次に紹介するレポートシステムの環境構築時に選択するソフトウェアで実現している。

また、⑤から①に戻る部分があるように、全体の流れも繰り返していく必要がある。分析業務は1回で終わらせることなく、ビジネスへの活用を行った結果を検証し、更に次

のビジネスにつなげることが必須だからである。世の中のトレンドと同じように、ビジネスは常に流動的で変化していくものである。その変化を捉えるために、分析を怠らないことが新たな価値の創出につながっていくだろう。

3. レポートシステムの構築

3.1 ソフトウェアの検討

今回のレポートシステムは、なるべく安価に導入できることを目指して構築した。そのため、すべてのソフトウェアにオープンソースを利用する。レポートシステムを構成するために必要と考えたソフトウェアは、「オペレーティングシステム」、「データベース」、「統計解析ソフトウェア」の3つである。次に選択したソフトウェアとその選定理由を述べる。

(1) オペレーティングシステム

オープンソースで選択するので、安定した稼働が期待できる Linux ディストリビューションを選定対象とし、その中から「Ubuntu」を選択した。Ubuntu は数ある Linux ディストリビューションの中でも、新技術を積極的に取り入れる傾向があり、半年ごとにリリースを行うなど、オペレーティングシステムとしては比較的頻繁にアップデートされている。そのため、外部リポジトリのパッケージ量は膨大である。前述したとおり、オープンソースで構成することから、ソフトウェアの選択肢を広げるという意図で選択した。

(2) データベース

今回は研究であるため、従来のリレーショナルデータベースマネジメントシステム（以下、RDBMS）と、一般的に NoSQL と呼ばれているデータベースの2つを搭載し、両方を利用できる形にした。

まず、RDBMS は「Mysql」を選択した。RDBMS のオープンソースで代表的なものは Mysql と Postgresql の2つである。双方とも商用データベースに匹敵する機能を持っており、どちらを選んでもほぼ問題はない。その中で Mysql を選択した理由は、情報量の豊富さである。チューニング次第でパフォーマンスを高くすることができ、そのための情報が多いことが決め手になった。一方の Postgresql は空間情報についての機能が充実している。空間情報を利用するのであればこちらを薦めておきたい。

次に、NoSQL である。NoSQL の場合、目的に合わせたデータストアを選択しなければならぬ。選択肢として代表的なものは、「分散 Key-Value Store」、「列指向データベース」、「ドキュメント指向データベース」などが挙げられる。今回はその中から、「ドキュメント指向データベース」に分類される「Mongodb」を選択した。ドキュメント指向データベースの特徴は、構造を定義する情報を持たないことで、データをドキュメントのように格納することである。そのため、RDBMS のように構造設計を行う必要がない。データ分析では、業務データ以外にも、アプリケーションログなどの複雑な構造をもつデータや、他業務のデータを扱うことがある。ドキュメント指向データベースであれば、これらのデータをそのまま保存することができるため、少なくとも分析環境においては高いスケーラビリティをシステムに持たせることが可能である。その中で

Mongodb 選択した理由は、レポートシステムでは検索処理が多々発生するため、RDBMS に匹敵するデータ検索機能を持ち、データベースにインデックスを設定可能であることからである。

(3) 統計解析ソフトウェア

データ分析を行うには、統計ソフトウェアを利用する。統計ソフトウェアを利用することで、分析や統計を簡単に行い、結果をグラフィカルに表示することができる。

今回は、統計ソフトウェアのデファクトスタンダードである「R 言語」（以下、R）を利用する。R はオープンソースながら豊富な統計解析機能を備えており、インターネット上での情報交換も盛んに行われている。近年では学术论文など、信頼性を要求されるところでも利用されている。

環境面についても Windows、Linux、MacOS の各 OS に対応しているため、ほとんどのすべての環境に導入することが可能である。更に、ODBC でのデータベース接続機能も備えているため、既存の業務データの活用も容易である。（当然、今回選択した Mysql、Mongodb も接続可能。）

以上、これらのソフトウェアでシステムを構築していく。また、前章で課題とした、レポート作成とシステム反映を短い期間で実施する環境と、レポートの提供方法は、統計解析ソフトウェア R のパッケージなどを利用することで簡単に実現することができた。

(1) レポート提供方法

レポートの提供には「Shiny」を利用する。Shiny は、R のパッケージで、処理結果を WEB ブラウザ上に表示できる。WEB で提供できるアプリケーションはユーザインターフェースを提供するプログラムとサーバサイドプログラムで構成されている。ユーザインターフェースは Shiny で用意されている部品を使うことで簡単にレイアウトを作成でき、サーバサイドプログラムは、R で作成したプログラムをほぼそのままの記述で WEB に公開できる。そのため、研究成果を簡単に共有することができる。

(2) レポート開発環境

レポートの研究、開発用ソフトウェアとして「RStudio-server」を導入する。RStudio-server を導入することで、WEB ブラウザから R を利用できるようになる。RStudio-server が R をそのまま利用することよりも、開発環境として優れている部分は次の通りである。

1. 作成したグラフをすぐに確認できる。
2. 変数、配列に格納した値を確認できる。
3. 入力補完機能を備えている。

まず、1 は R の欠点を補っている部分である。R はプログラムからグラフを出力すると、マシン上に画像ファイルを作成する。この画像ファイルを見るためにわざわざコマンドを発行するなどのステップを踏まなければならない。対して、RStudio-server はグラフを出力すると、そのグラフを WEB ブラウザ上にそのまま表示する。レポート開発中は、何度もグラフを表示するので、開発効率はかなり高くなるといえる。2 と 3 は

IDE によくある機能で、R プログラムの作成をサポートしてくれる。次に RStudio-server の画面を図 2 に示す。

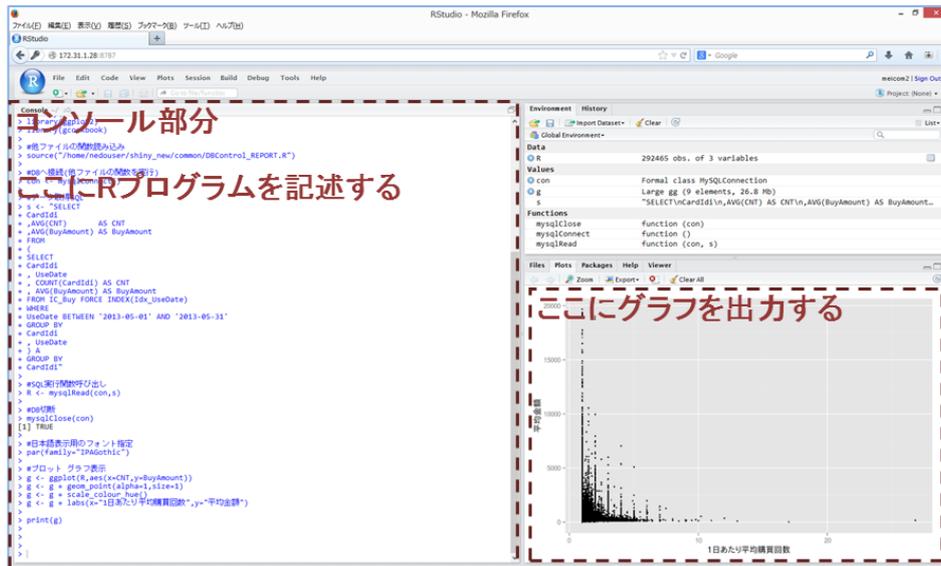


図 2 RStudio-server の画面

以上のソフトウェアを利用することで、「RStudio-server」でレポート開発を行い、出来上がったレポートを「Shiny」を利用して公開するという流れを確立した。最終的なシステム構成を図 3 に示す。

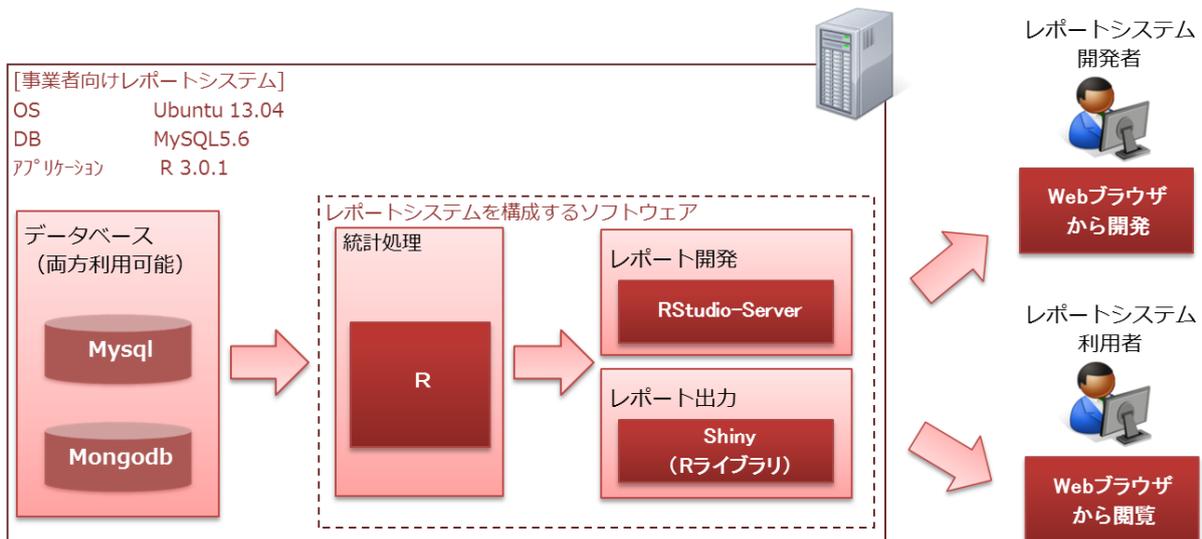


図 3 システム構成

3.2 レポートシステムについて

次に構築したレポートシステムを紹介する。レポートシステムの画面構成を図4に示す。

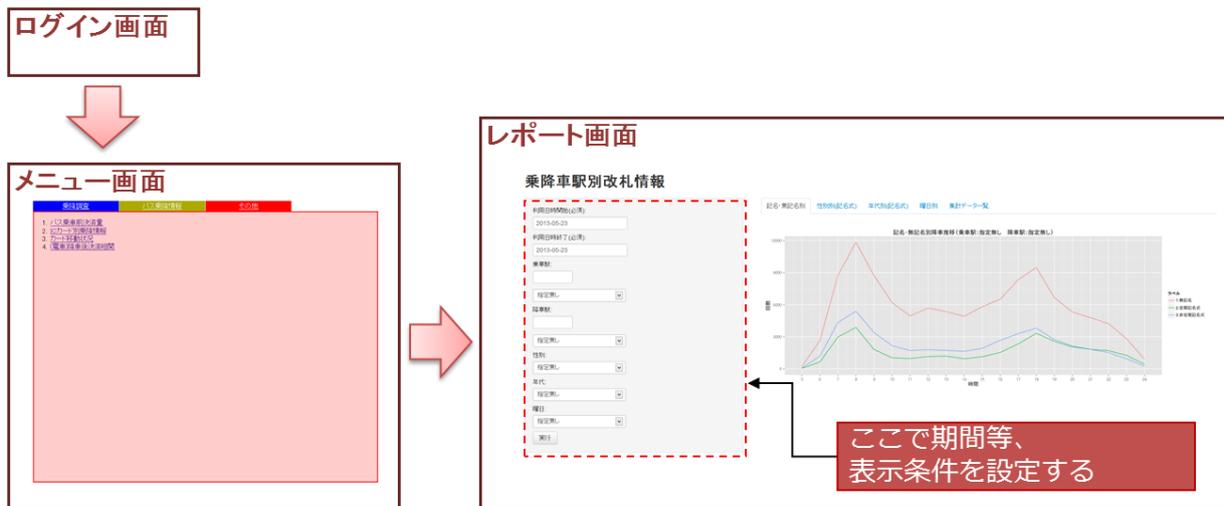
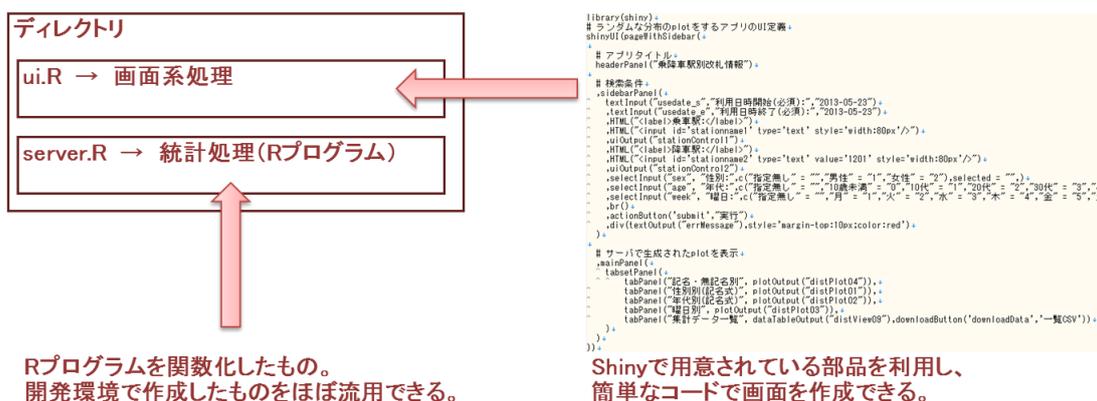


図4 レポートシステム画面構成

レポートシステムの画面構成は、ありふれた WEB システムと同様である。ログイン画面で認証を行った後、メニュー画面からレポートを選択する流れである。レポート画面は左側に期間などの条件を設定し、実行すると右側に結果が表示されるという流れである。ユーザは集計結果を、簡単な条件を設定するだけで閲覧することが可能である。

このシステムの特徴は2点ある。一つ目はレポートの動的出力である。様々な条件でレポートを出力することで、新しい価値を見つける作業をサポートする。2つ目は、Shinyの機能を活かした、スムーズなレポート追加である。次に、Shinyプログラムの構成を図5に示す。



Rプログラムを関数化したもの。
開発環境で作成したものをほぼ流用できる。

Shinyで用意されている部品を利用し、
簡単なコードで画面を作成できる。

図5 Shinyプログラムの構成

server.RにRのコードを記載する。開発環境で作成したRプログラムを関数化するだけでよい。ui.Rは簡単なコードでユーザーインターフェースを構築することができる。ちなみ

に、図5のプログラム例は、図4の画面を構成するコードである。このように、Shiny を利用すると、RStudio-server で作成したレポートを短時間でレポートシステムに反映することができる。なお、画面例以外にも、表形式の画面出力や、PDF、CSV の各ファイルを出力することも Shiny の機能を使って実装可能である。

4. データ設計

データ設計で最初に行うべきことは、提供されたデータを理解し、分析に適した形で設計を行うことである。そのために行うことは、業務を理解すること、実データとテーブル定義を見比べながら、時間によるデータの変化を捉えることである。各項目の持つ性質と意味を理解することが最も重要であるため、時間をかけるべき部分であるといえる。

データを理解できたら設計を行う。分析に用いるデータは、業務システムのデータ構造のように、第3正規形まで正規化したものではなく、正規化を崩した状態のほうが理解しやすい。データはリソース系データ（マスタ）とイベント系データ（随時発生するデータ）の2つに分けられるが、分析の対象となるデータのほとんどはイベント系データである。正規化されたイベント系データは、コードの羅列になっており、そのコードが示す名称などはマスタに格納されている。分析中に1レコードに着目することもあるが、コードの羅列では一目みただけではデータを理解しづらい。また、データ量が増えると、正規化されたデータを処理するたびにテーブルの結合が発生するため、処理コストが増大してしまう。そのため、データ分析に用いるデータは、正規化を崩した状態にすることが望ましい。

以上を踏まえて設計を行う。まずは、今回の分析対象データを図6に示す。なお、駅、停留所、店舗の情報はマスターデータと統合し、正規化を崩した状態として扱う。

マスタ(リソース系)	分析対象データ(イベント系)																								
<table border="1"> <tr><th>記名式manaca情報</th></tr> <tr><td>manacaidi</td></tr> <tr><td>性別</td></tr> <tr><td>生年月日</td></tr> <tr><td>定期券情報</td></tr> </table>	記名式manaca情報	manacaidi	性別	生年月日	定期券情報	<table border="1"> <tr><th>manaca電車乗降情報</th></tr> <tr><td>manacaidi</td></tr> <tr><td>乗車駅</td></tr> <tr><td>降車駅</td></tr> <tr><td>利用金額</td></tr> <tr><td>乗降日時</td></tr> </table>	manaca電車乗降情報	manacaidi	乗車駅	降車駅	利用金額	乗降日時	<table border="1"> <tr><th>manacaバス乗降情報</th></tr> <tr><td>manacaidi</td></tr> <tr><td>乗車停留所</td></tr> <tr><td>降車停留所</td></tr> <tr><td>利用金額</td></tr> <tr><td>乗降日時</td></tr> </table>	manacaバス乗降情報	manacaidi	乗車停留所	降車停留所	利用金額	乗降日時	<table border="1"> <tr><th>manaca決済情報</th></tr> <tr><td>manacaidi</td></tr> <tr><td>購入店舗</td></tr> <tr><td>購入金額</td></tr> <tr><td>購入日時</td></tr> </table>	manaca決済情報	manacaidi	購入店舗	購入金額	購入日時
記名式manaca情報																									
manacaidi																									
性別																									
生年月日																									
定期券情報																									
manaca電車乗降情報																									
manacaidi																									
乗車駅																									
降車駅																									
利用金額																									
乗降日時																									
manacaバス乗降情報																									
manacaidi																									
乗車停留所																									
降車停留所																									
利用金額																									
乗降日時																									
manaca決済情報																									
manacaidi																									
購入店舗																									
購入金額																									
購入日時																									

図6 分析対象データ

図6の通り、今回の分析対象は大きく分けて4つのデータで構成されており、「manacaidi」が共通項目である。電車乗降情報、バス乗降情報、決済情報はイベント系データ、記名式manaca情報のみがマスタにあたる。

これらのデータを分析していくわけだが、イベント系データだけでは、乗降件数、決済件数の集計ぐらいしかできない。しかし、記名式manaca情報を加えることで、「どの年代が何時に利用しているか」、「男性と女性で傾向に差があるのか」などを分析することが

できる。前述したように、正規化されていないほうが、分析に利用しやすいため、各トランザクションデータと記名式 manaca 情報がすでに結合されている状態のデータを定義する。図 7 に分析用に決済情報を定義するまでのイメージを示す。

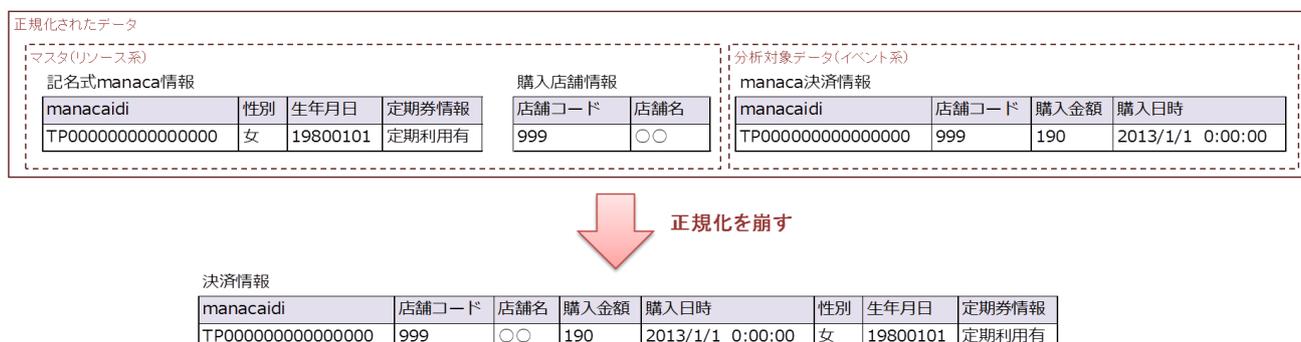


図 7 分析用データ定義のイメージ

5. レポートの検討

データ分析で一番重要なことは、分析からデータに新たな価値を見出し、ビジネスの活用につなげることであり、今回は新たな価値を見出せるレポートを作成、提供することが目的である。しかし、新たな価値は設計したデータを眺めるだけでは見えてこない。人間が分析を行うには可視化することが必要だが、やみくもに可視化しても上手くいかないことが多い。本章では、価値を見出すために行ったアプローチと実際に作成したレポートを紹介する。

5. 1 異なる項目の組み合わせ

今回のデータは、「移動情報」、「購買情報」、「個人属性」の3つに分類することができる。分析対象データを当てはめると以下表 1 のようになる。

分類	データ名
移動情報	manaca電車乗降情報
	manacaバス乗降情報
購買情報	manaca決済情報
個人属性	記名式manaca情報

表 1 データの分類

これらが持つ項目、移動情報であれば乗車駅、降車駅などを組み合わせることで、新しい価値を見出すことができるのではないかと想定し、まずは、移動情報、購買情報、個人属性の各項目を組み合わせで可視化されるものを検討し、そこから生まれる要素を定義した。その中の一部分を表 2 に示す。

No.	乗降情報	購買情報	個人属性	可視化されるもの	要素名
1	降車駅 降車時刻		性別 曜日 年代	時間帯毎の降車傾向	駅時間帯別降車推移
2	乗車駅 降車駅			バス停留所の乗り場単位での乗降数と乗車密度	系統毎乗降状況
3	乗車駅 降車駅		ICカード	1カ月間の停留所利用回数の割合	停留所毎利用回数割合
4		店舗 利用時刻		指定期間内に一度も決済されていないmanaca加盟店	未決済加盟店
5	降車駅 利用時刻	店舗 利用時刻		電車降車時間と降車後購買までの経過時間の相関	駅降車後決済相関

表2 項目の組み合わせ

次に、表2で組み合わせた要素で可視化されるものについて、事業者にとっての価値を検討しレポート化する。検討結果を表3に示す。

No.	要素名	事業者にとっての価値
1	駅時間帯別降車推移	性別、曜日、年代毎の降車傾向をつかむことができる
2	系統毎乗降状況	乗り場単位での利用者数、停留所間の乗車密度を知ることができる
3	停留所毎利用回数割合	停留所の利用頻度と、ユーザ層をつかむことができる
4	未決済加盟店	ICカードが利用されていない加盟店を知ることができる
5	駅降車後決済相関	電車降車時間と降車後購買までの経過時間の相関をつかむことができる

表3 事業者にとっての価値

5.2 レポート化と分析

事業者にとっての価値についてレポート化し、分析を行った。次に実際の分析結果の1例を挙げる。前章で定義した要素「駅時間帯別降車推移」のレポート(図8)と分析結果を示す。

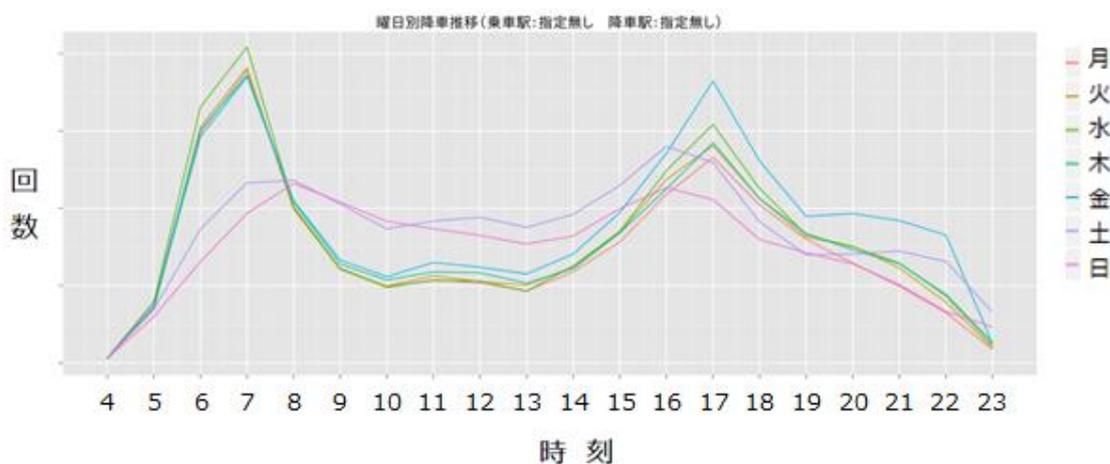


図8 駅時間帯別降車推移(曜日)

○ 分析

平日（月～金）と休日（土、日）では明らかに異なる傾向を示している。平日に関しては通勤、通学といった定期的な利用が多いと推測されるため、休日の利用促進を促すほうことに効果があると考えられる。ただし、金曜日の夜だけ他の平日に比べて明らかに降車数が多い。翌日が休日という人が多いため、降車した後、どこかに訪問しているものと推測できる。休日及び休日の前日にプロモーションを行うと効果が高いと推測する。

このようにデータを組み合わせて可視化することで、データに新たな価値を見出すことができる。このレポートでは「降車駅」、「降車時刻」、「曜日」を組み合わせている。次に「曜日」を「年代」に変えたレポート（図9）と分析結果を示す。

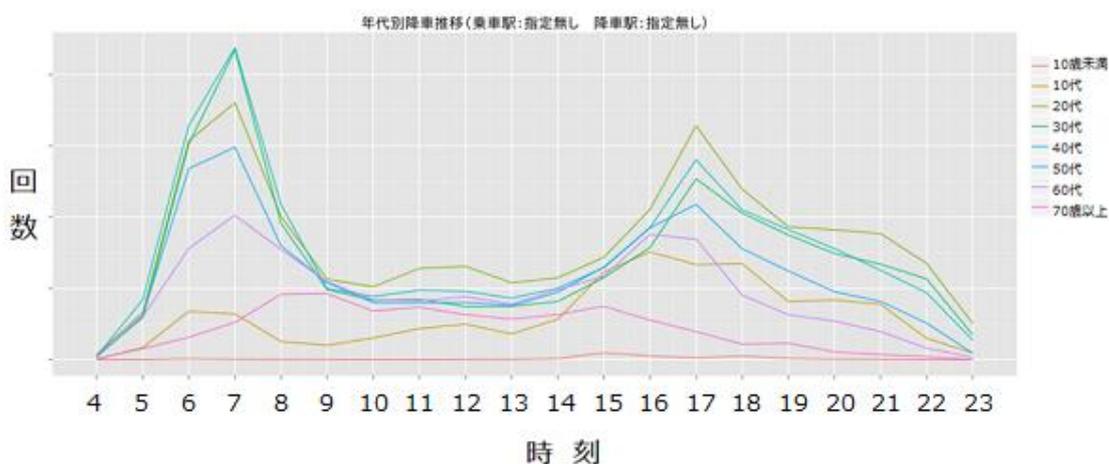


図9 駅時間帯別降車推移 (年代)

○ 分析

年代によって利用者数が大きく異なる。

10代 : 学生が多いと考えられる。夕方に降車が多いのは、下校後、都市部にある予備校や進学塾などに通うことが想像される。

20～50代 : 就業している人と考えられる。

60代 : 人数が少ないが、20～50代と同じ傾向があるため、就業している人と考えられる。

70代 : 緩やかなカーブを描いている。仕事をリタイアした人が多いと考えられる。9～16時の利用が多いのは通院している人と想像される。

「曜日」を「年代」に変えることで、全く異なる分析が可能になる。更に、要素を変えるだけでなく、加えることで、更なる価値を求めることができる。図8に年代の要素を「加える」ことで、「年代ごとの曜日別降車傾向」をつかむことができる。図8で見出した価値を、「休日及び休日の前日にプロモーションを行うと効果が高い」とすると、年代

を加えることで、「〇〇代の人を対象に、休日及び休日の前日にプロモーションを行うと効果が高い」という更なる価値を得ることができる。この場合、更に精度が高いプロモーションを行えると推測する。

5.3 レポートの表現

前章では、折れ線グラフを使ったレポートを紹介したが、表現方法によってまた異なる価値を見出すこともできる。今回作成したレポートの中から、分析によく使われる散布図を使った分析を紹介する。

5.3.1 散布図を利用した manaca 利用傾向の分析

散布図は2項目の分布を把握でき、傾向を分析することに適している。図10に散布図を示す。

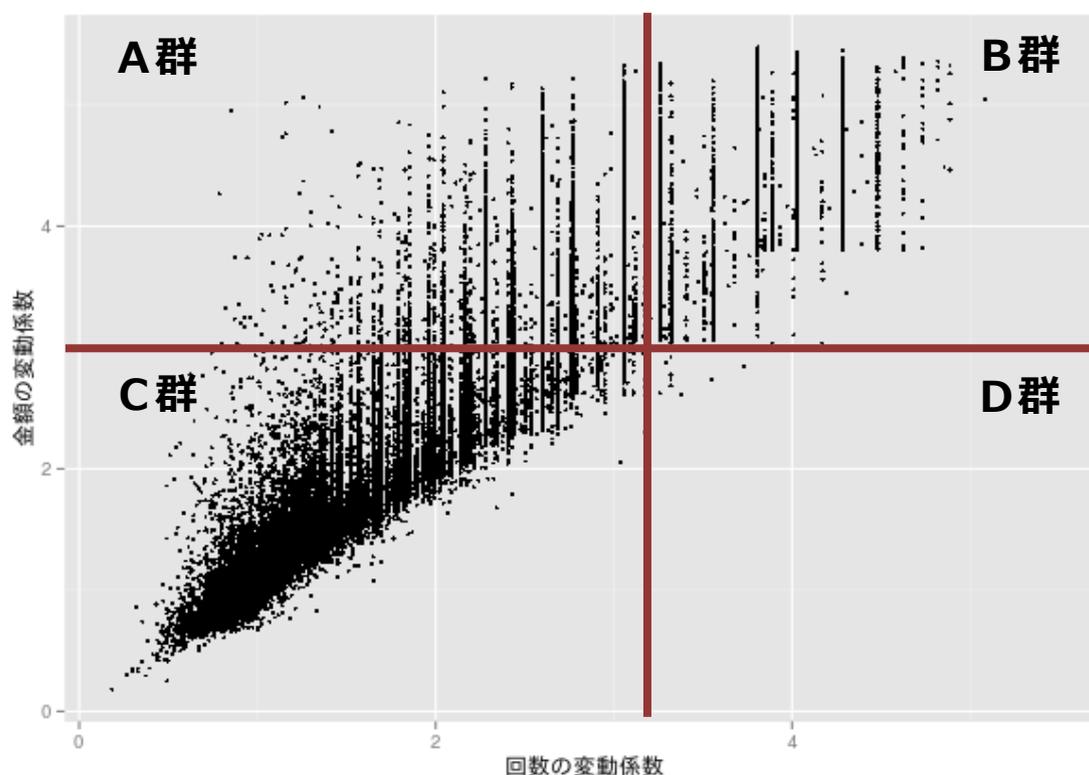


図10 1カ月間の manaca 利用傾向の散布図

この散布図は、manaca ユーザが1カ月間に manaca を利用した傾向を分析するために、利用金額の変動係数と利用回数の変動係数を軸にデータをプロットしたものである。プロットした一つの点は1人の manaca ユーザを示している。分析に入る前に散布図の軸とした、変動係数について解説する。まずは、表4に変動係数の例を示す。

日付	利用者1		利用者2		利用者3		利用者4	
	回数	金額	回数	金額	回数	金額	回数	金額
5月1日	1	100	3	120	1	200	0	0
5月2日	1	100	1	100	0	0	0	0
5月3日	1	100	2	130	0	0	0	0
5月4日	1	100	0	0	1	100	0	0
5月5日	1	100	0	0	0	0	0	0
5月6日	1	100	1	100	2	400	1	700
5月7日	1	100	5	250	0	0	0	0
合計	7	700	12	700	4	700	1	700
平均	1	100	1.7	100	0.6	100	0.1	100
変動係数	0	0	1	0.8	1.2	1.4	3	2.4

表4 変動係数の例

表1では、1週間のmanaca利用回数と利用金額を表にし、それぞれの変動係数（標準偏差÷平均）を求めている。表中の4つは、合計の利用金額は同じだが、1回あたりの利用金額と利用回数にはそれぞれ特徴を持たせている。

- 利用者1 → 毎日同じ回数、金額を利用している。
- 利用者2 → 利用頻度は高いが回数にバラつきがある。
- 利用者3 → 利用頻度が少ない。
- 利用者4 → たまにしか利用しない。

表4の通り、利用傾向が極端になるほど、変動係数が大きくなる。そのことから、変動係数を使うことで、利用傾向を分析することが可能になる。以上のことから、図10の散布図を使って、manaca利用傾向を分類することができる。図11に利用傾向の分類を、表5に利用傾向の分類に対する分析を示す。

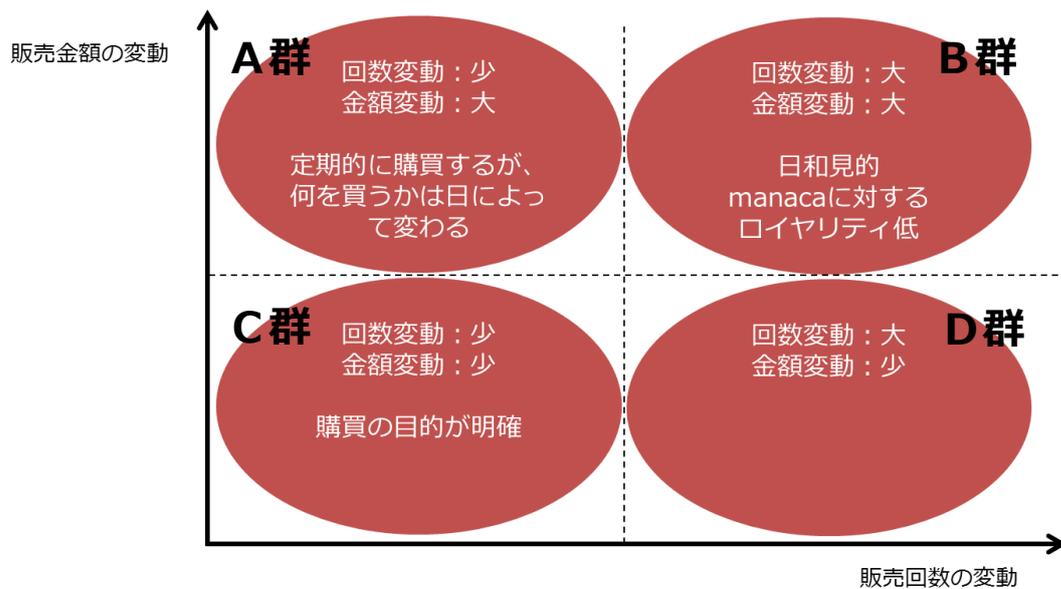


図11 1カ月間のmanaca利用傾向の分類

分類	分析
A群	決済手段として日常的にmanacaを使用しているが、利用金額の変動が大きい。 manacaで決済した場合にプレミアムをつけるなど、お得感を演出することでアップセル（より上位のものを購入してもらうこと）やクロスセル（別の商品を購入してもらうこと）を期待できる。
B群	決済手段として必ずしもmanacaを使用するとは限らない。 manacaに対するロイヤリティが低い客。 manaca決済の動機付けが必要。 交通利用と絡めて、manaca利用を促進するキャンペーンが有効だと考えられる。
C群	決済手段として日常的にmanacaを使用しているかつ金額の変動が小さい。 日常のルーチンとして決まったものを購入していると考えられるため、この層に商品やサービスのキャンペーンを実施しても、決済量増につながりにくいと考えられる。 この層に対しては、商品のサービスやプロモーションより、manacaを中心とした様々な情報を提供し、永続的にmanacaを利用してもらい、この群に留まってもらうことが大切と考えられる。
D群	あまり想定できない。実際に集計した結果でも少数しか存在しなかった。

表5 manaca 利用傾向の分類に対する分析

この結果を元に様々な施策を考え出すことができる。例えばA群のようなユーザを増やしたいということであれば、B群に対してプロモーションをかけ、プロモーション前にB群にいたユーザがどの程度A群に移動したかを検証すればよい。検証するには、実際にプロモーションを行う必要があるが、プロモーション前後での比較を実施することで効果の有無を確認することは可能である。

また、ABテストの結果分析にも活用できる。一つの点が1人であるため、プロモーションをかけた人、かけてない人の利用傾向をプロモーション前後で比較し、効果を検証することが可能である。このように、散布図を使うことで、データの分類を行い、傾向を分析することができる。

5. 3. 2 バルーンチャートを利用したバス乗降の分析

バルーンチャート（またはバブルチャート）は散布図の1種で、散布図の要素に加えて、プロットする1点のデータ量を円の大きさで表現するものである。そのため3つのデータの関係性を一つのグラフで見ることができる。次に示す図はあるバス路線の1系統について、曜日ごとの乗車量（図12）と降車量（図13）をバルーンチャートで表現したものである。

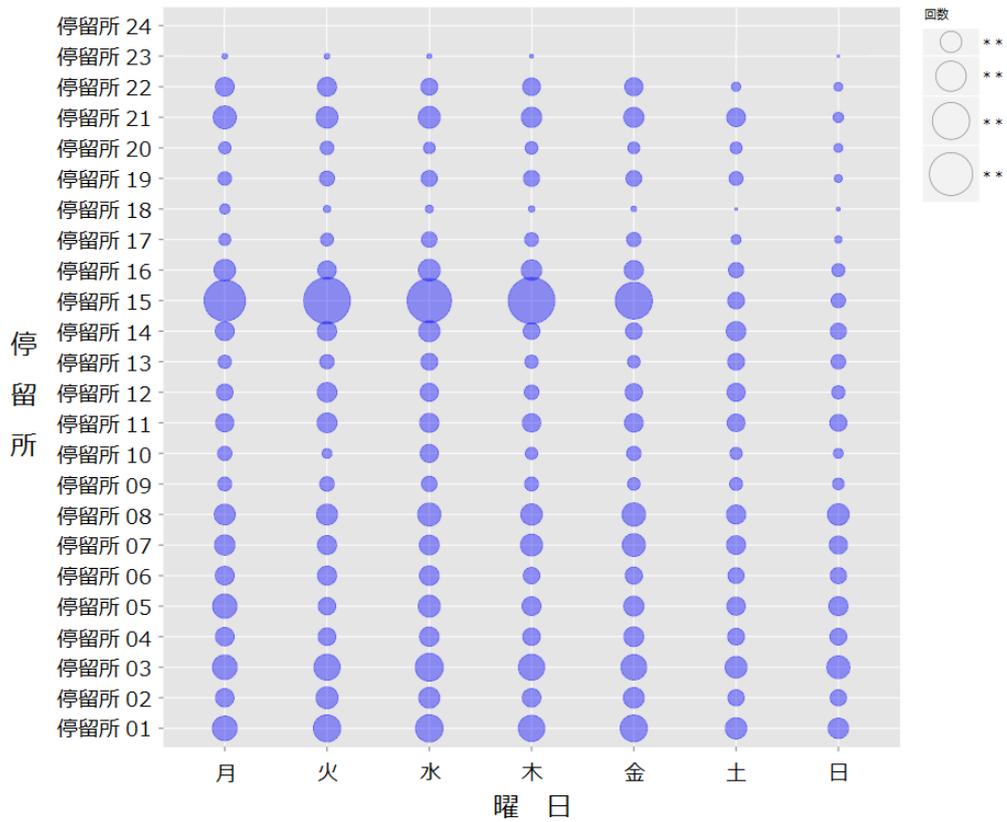


図 1 2 系統別曜日ごとの乗車量

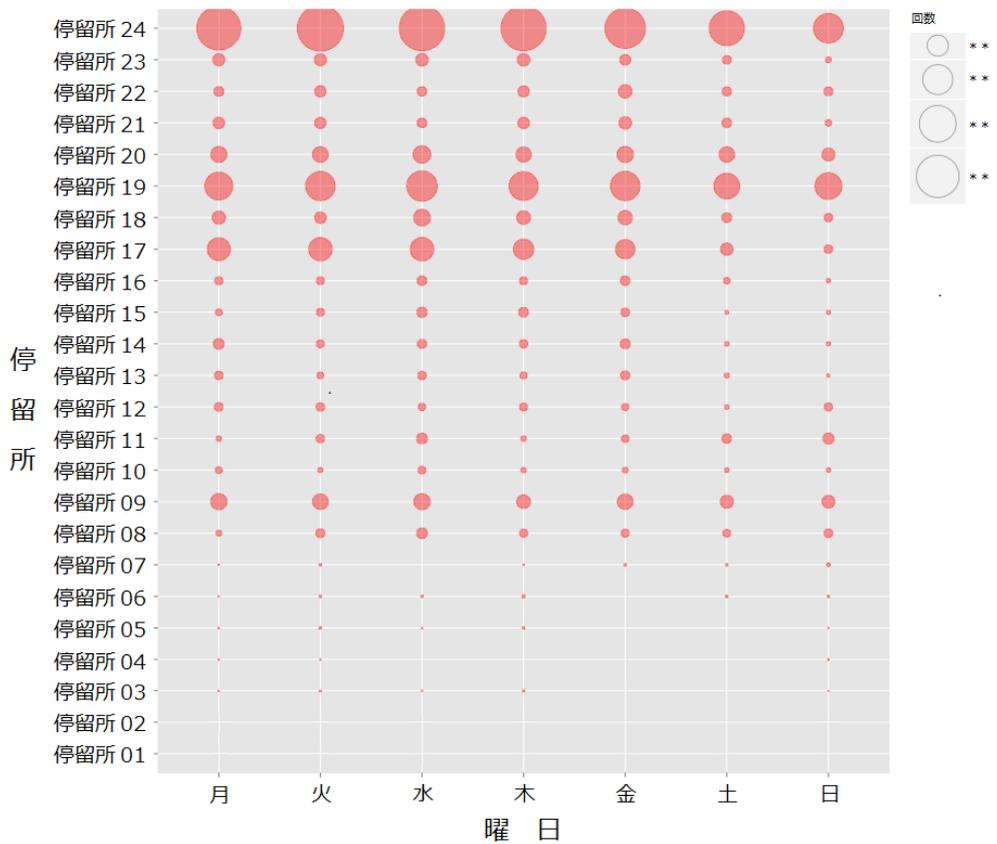


図 1 3 系統別曜日ごとの降車量

この系統は、始発点が停留所01、終点が停留所24で、一つのプロットが、その停留所を利用した人数となっている。そのため、停留所01で降車する人と停留所24で乗車する人は存在しない。バルーンチャートの特徴は、データの大小を1目で判断できることである。この2つの図であれば、平日に停留所15で乗車する人が多いことと、終点で降車する人が多いことが1目でわかる。この表現を使うことで、曜日ごとのダイヤ数を検討する、曜日の軸を時間に変えて、時間ごとのダイヤ数を検討するといったことに活用することができる。

6. 考察

6.1 データ分析の考察

今回はビジネスプランに辿り着くことができなかつたため、2.2で示したワークフローのビジネスへの活用及び検証を実践できず、データが示す事象が何故発生するのかという原因特定や、キャンペーンなどの施策を実施した時の効果を検証するまでには至らなかつた。分析結果を踏まえたビジネスへの活用と、活用した上での検証については今後も各事業者とともに、実践を重ねていく必要がある。

今回分析した結果は過去のある時点における分析結果であり、時間の経過とともにどう変化していくのかを引き続きウォッチしていくことが重要である。

6.2 レポートシステムの考察

レポートシステムの構築は、概ね満足のいく結果ではあつた。しかし、不足している点としては NoSQL を有効活用できなかったことが挙げられる。今回は過去3年分のデータを扱つたため、RDBMS で十分なレスポンスを得ることができてしまった。しかし、本格的に運用することになればデータ量は膨大になり、次第に RDBMS ではレスポンスが低下する可能性がある。そういった場面での選択肢が NoSQL の多くが持つ機能であるため、有効活用できるような研究を続けていくことが必要である。

7. 今後に向けて

繰り返しではあるが、今回はビジネスプランにたどり着くことができていないが、名古屋鉄道並びに名鉄バスからは、今回のレポートシステムを継続して利用したいという意見を得た。利用目的はビジネスプラン企画の材料として使いたいというものである。事業者と行ったディスカッションにて、次に示す乗車人員数推移結果（図14）を活用し、停留所間の乗車数からダイヤの調整を行うという意見があった。

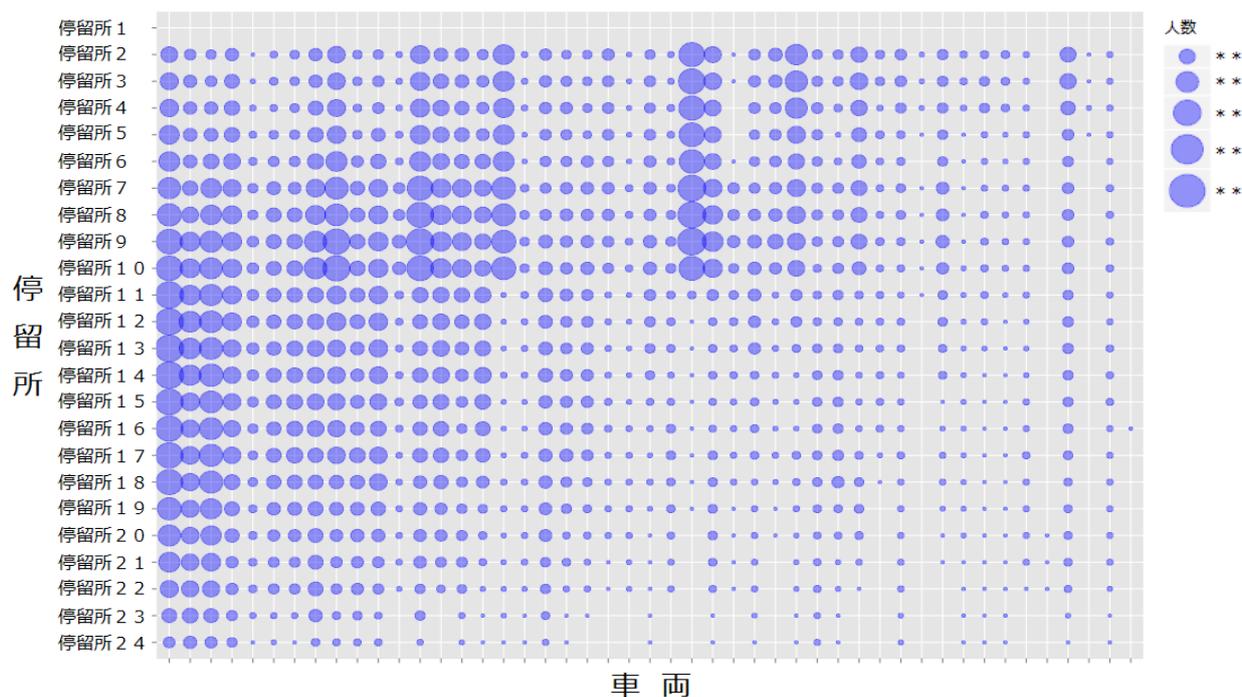


図14 乗車人員数推移結果

この図の1点（円）は、停留所時点でのバスに乗車している人数を表しており、車両（便）別の乗車人数数の推移が一目でわかる。ある時間の便を増やす、減らすといったことの裏付けとして利用できる。事業者がビジネスプランを企画する場合、企画に対する裏付け、根拠を明確にしなければならない。しかし、裏付けの材料として必要な、現状の集計や分析に苦勞しているのが現状である。レポートシステムが可視化するものは、その裏付け作りに有効であり、新しいアクションを起こすための叩き台として扱うことができる。

この結果から、レポートシステムを用いた分析を継続していくことでビジネスプランを計画し、ビジネスへの活用と検証を行っていくことは十分可能であると判断できる。当然、このような意見、要望を得て価値を見出すレポートを作成するためには事業者の協力が不可欠である。

今後もレポートシステムを用いた分析を継続し、検討を重ねてビジネスへの活用を目指していく予定である。時間はかかってしまうものの、2.2で示したワークフローを繰り返していくことによりデータの価値を見出すノウハウを取得し、成果を挙げたいと考えている。

8. 最後に

今後、IT ベンダーがデータ分析業務に携わることが増えていくことが予想される。我々 IT ベンダーだけでデータ分析を行い、新しい価値を求めることは厳しい。ユーザの協力を得ることと、データの特徴を知ることが、成功に近づく必須条件である。ワンストップサービスを意識し、ユーザとともにシステムを作り上げていくことでこの条件を満たすことはできる。

データ分析の敷居は高いように思えるが、実践しなければ何一つ始まらない。今回のレポートシステム構築ノウハウが、この敷居を少しでも低くできれば幸いである。