

共同研究「経営実務データを用いたデータサイエンティスト育成方法の研究」報告書

2017年3月15日

国立大学法人滋賀大学データサイエンス教育研究センター長

竹村 彰通

## 1. 共同研究の趣旨

本共同研究の研究題目は「経営実務データを用いたデータサイエンティスト育成方法の研究」であり、共同研究の目的は、企業の経営実務データを用いて販売予測や適性な在庫管理など、これからのデータサイエンティストに求められる能力を育成するための方法を研究することである。共同研究の期間は、平成28年10月26日から平成29年3月31日までである。

特定非営利活動法人ビュー・コミュニケーションズは実際の売り上げデータを教育目的のために適切に処理をおこなった上で国立大学法人滋賀大学に提供し、滋賀大学データサイエンス教育研究センターはデータをプロジェクト型の演習の形に整備する作業をおこなうことで、データサイエンティスト育成方法に関する共同研究をおこなった。

ビュー・コミュニケーションズからは、3種類の販売データが提供された。いずれのデータも滋賀大学に平成29年4月1日に開設される日本初のデータサイエンス学部の入学生にも理解できる実務データであり、演習のために非常に有用なデータであった。その点で本共同研究の意味は大きい。

## 2. 滋賀大学データサイエンス学部の育成人材像と演習の重視

滋賀大学データサイエンス学部では、データサイエンスの基礎要素技術としてのデータエンジニアリング(情報工学)及びデータアナリシス(統計学)に加えて、データからの価値創造すなわちデータから新たな知見を得てビジネスや地方公共団体が直面する課題を解決できる能力の育成を重視している。

この価値創造の能力を育成するためには、実際のデータを用いた試行錯誤や成功体験が必要であり、さらに教科書的な理論と実際のデータとの乖離などについても経験を積む必要がある。データエンジニアリングとデータアナリシスの要素技術のみならず、価値創造のスキルも有する人材をデータサイエンティストとよび、滋賀大学データサイエンス学部では、このようなデータサイエンティストを輩出することを目標としている。ビュー・コミュニケーションズから提供された3種類の販売データは、データ分析の試行錯誤を経験するためにも、また理論と実際の乖離を経験するためにも有用である。理論と実際の乖離については、ビュー・コミュニケーションズ副理事長の最近の著書「なぜあなたの予測は外れるのかーAIが起こすデータサイエンス革命」においても多くの例とともに示されており、データサイエンス教育において実際のデータを扱うことが重要であることがわかる。

ビッグデータ時代を迎え、データサイエンティストの不足が広く認識されるようになっており、滋賀大学に続いて他の大学でもデータサイエンス教育が充実される状況である。

### 3. 滋賀大学データサイエンス学部における演習の設計

滋賀大学データサイエンス学部における演習の設計は「PPDAC サイクルを回す」こと  
にという考え方に基づいている。PPCAC サイクルとは、問題解決における各段階を  
Problem (問題)、Plan (調査の計画)、Data (データ)、Analysis (分析)、Conclusion (結  
論)に分けるという考え方である。ビュー・コミュニケーションズ提供のデータに基づく  
演習においても、データをそのまま学生に示すのではなく、例えば売り上げと天気の関係  
をどのように考えるか(Problem)、その問題を考えるときにどのようなデータが必要か  
(Plan)なども考えさせることとしている。これについては以下の5節に例示する。

データの分析手法や結論の導き方については、2年次後半で履修する時系列解析入門や3  
年次前半で履修する時系列解析など手法を学んだ後のほうが、深い分析が可能となる。し  
かしながら、滋賀大学データサイエンス学部では、まず手法を学ぶのではなく、初年次か  
ら演習を重視しデータから出発してその分析の必要のために手法を学ぶこととしている。  
このため、ビュー・コミュニケーションズ提供のデータについても、本共同研究では、主  
に2017年度の新入生向けの演習の作成をおこなった。

### 4. ビュー・コミュニケーションズより提供のデータ

ビュー・コミュニケーションズから提供されたデータは以下の3種類のデータである。

1. 冷やし麺の販売実績及び気象の日次データ
  - ・冷やし麺は対象エリア 10 店舗の合計で、単品別及び品目計
  - ・気象は、最高気温、最低気温、降水量、降雪量、天気
  - ・対象期間: 2015/6/1~2016/5/31 (一年間)
2. アイスクリームの販売実績及び気象の日次データ条件
  - ・アイスクリームは対象エリア 10 店舗の合計で、プレミアムカップアイスとステイ  
ックアイスを区別し、単品別及び品目計
  - ・気象は、最高気温、最低気温、降水量、降雪量、天気
  - ・対象期間: 2015/6/1~2016/5/31 (一年間)
3. ペットフードの販売実績の週次データ
  - ・特定の店舗における 13 種類のペットフードの週次販売データ
  - ・対象期間: 2013 年 11 月 25 日の週~2016 年 11 月 14 日の週 (三年間)

最初の二つのデータについては、気象のデータも同時に提供されていることから、時系列  
データ間の相関についてもさまざまに調べることができる。また 3 番目のペットフードの  
販売データは、3 年間にわたることから、予測モデルの作成と検証に適したデータである。

## 5. 提供データを用いた演習課題の例

ここでは、アイスクリームの販売実績および冷やし麺類の販売実績データについて、初年次生の演習課題を示す。ペットフードの販売実績については、時系列解析などのある程度の手法を学んだ上で分析するほうがより適切と考えられるため、2年次の演習で利用することを考えている。

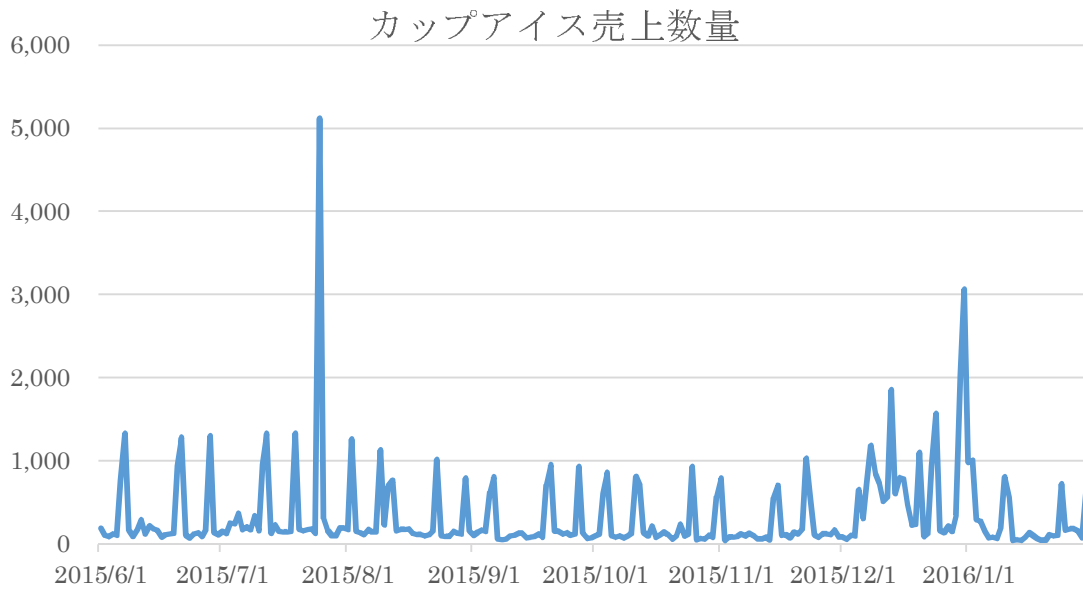
まず PPDAC サイクルの最初のステップである Problem(問題)の例としては、冷やし麺類やアイスクリームの売り上げはどのような気候データに左右されるか、気温なのか天候なのか、という問題を学生に投げかける。どのようなデータが必要で、どのような分析ができるかをグループワークで考察させる。

計画(Plan)の段階では、どのような手法を用いてどのような分析を行うか考察させる。記述統計量、代表値(平均など)、二変量データの数値表現、散布図、相関係数、データの種類(数値か、数値でないか)などを適宜説明する。例えば、冷やし麺では、月別推移やグラフ表現(棒、折れ線)、売り上げデータと気温データの間関係を調べるためには、散布図、相関係数などを用いることを考えさせる。

データ(Data)の段階では、CSV データを与え、データの基本的な処理方法を学ばせる。Excel を用いる。ファイルの扱いやデータの基本的な扱い方、各種の集計方法、欠損値の処理方法などを理解させる。またデータの選択方法や取得方法も順に理解させる。例えば、冷やしそば売り上げデータの月別推移を折れ線グラフで表現したり、売り上げデータと気温データを要素とする散布図を作らせる。

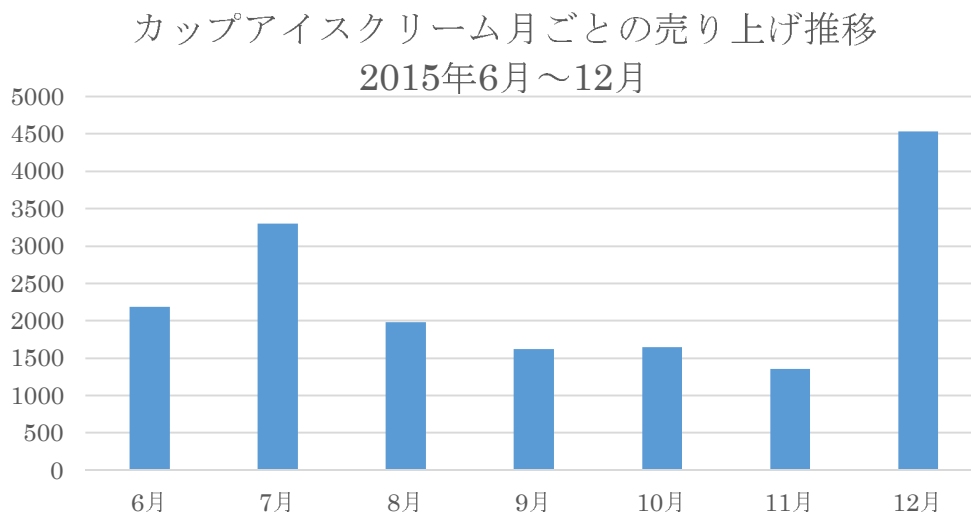
そして分析(Analysis)の段階では、「計画」に基づき、「データ」の分析を行い、「問題」として立てた予想に基づいて分析する。例えば、以下に図で示すように、プレミアムカップアイスクリームの売り上げは気温とほとんど相関がないことがわかる。そこで、他の要因を考えるために、曜日別週別売り上げ棒グラフなどを試させる。

例えば、プレミアムカップアイスクリームの売り上げの図は以下のようなになる。



日次では、かなり激しい変動をしていることがわかる。

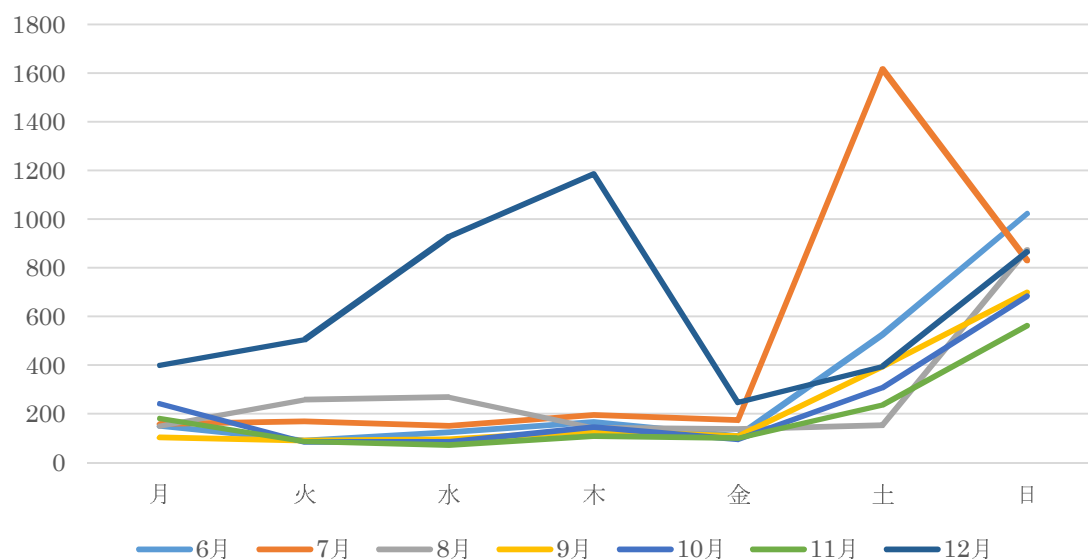
これを2015年6月から12月までの月ごとに集計すると、月ごとの売り上げの推移は次のようなグラフとなる。



このような図を書くことによって、学生はカップアイスクリームの売り上げが、天候のみによって影響されるわけではないことが理解できる。

また曜日別の売り上げのパターンを図示してみると、曜日の効果があり、また季節の影響もあることがわかる。

## カップアイスクリーム 2015年6月－12月の曜日別 売り上げ



### 6. まとめ

以上に示したように、ビュー・コミュニケーションズから提供されたデータは、さまざまな観点から分析が可能であり、初年次生のうちから、さまざまにデータを試行錯誤的に分析することができ、教育的価値が非常に高いものである。

滋賀大学データサイエンス学部としては、今後もさらに興味深い演習課題をビュー・コミュニケーションズとの共同研究の形で開発して行きたいと考えている。