The Use of Data Mining Techniques for Competitive Advantage in the Turkish Consumer Credit Sector

Chapter 1. Introduction

Background of the Study

Turkey has experienced a rapid expression of consumer credit over the past decade. However, with signs of a slowdown in the Turkish economy, and possible saturation of the market, now is an appropriate time to ask how Turkish banks are attempting to identify and convert new credit customers in order to gain competitive advantage. Now that the low-hanging fruit has been picked, there is an open question as to how data analytics might be helping banks find and convert the Turkish consumers who have not yet sought credit, or who are willing to switch credit providers.

The literature indicates that, since around the beginning of the millennium, Turkey has experienced an expansion in credit (Srnec et al., 2013, Ozturk and Acaravci, 2013, Uzkurt et al., 2013, Kasman, 2012). Credit was virtually unknown in Turkey in the mid-1990s and beforehand. However, by 2015, Turkey has come to be enormously reliant on credit, particularly consumer credit. The expansion of consumer credit has radically expanded the number of Turks who could own cars and houses, and has become a cornerstone of the economy.

In recent years, Turkey has undergone some small-scale economic shocks and a devaluation in its currency. Currently, the economic situation continues to be fluid and volatile. For banks, the search for new consumers to whom to extend credit has become more difficult (Srnec et al., 2013, Ozturk and Acaravci, 2013, Uzkurt et al., 2013, Kasman, 2012). Despite the fact that the Turkish banking sector is one of the most innovative and dynamic financial sectors in Europe (Coopers, 2012, Omurgonulsen, 2009), there is limited empirical research on the strategy employed by Turkish banks. This gap is of practical importance given that the strategies pursued by Turkish banks in the attempt to extend domestic credit expansion might prove to be an important factor in the continued growth of the Turkish economy.

Data analytics are a key component of banks' consumer credit strategies (Hitt et al., 2009). In developed countries, data analytics have been used by banks to (a) identify consumers who have not yet obtained credit and are therefore new customers, (b) upsell or cross-sell existing customers, and (c) package new products for the credit market. The advantage provided by analytics in all of these contexts is that of being able to generate actionable, strategic information from large bodies of data (Rosenberger and Nash, 2009).

In Turkey, the science of data analytics remains in its infancy. The relative novelty of domestic credit, and of the marketing sector in general, means that businesses lack detailed data on consumers (Srnec et al., 2013, Ozturk and Acaravci, 2013, Uzkurt et al., 2013, Kasman, 2012). Turkish businesses have only recently begun to invest in powerful forms of enterprise software, including data analytics software, and to apply it to business problems.

Given that data analytics have been used extensively in the financial sector, it is likely that Turkey's banks, at least the larger banks, already have robust data analytics solutions in place. However, because of gaps in the empirical literature, little is known of the prevalence of data analytics in Turkey, much less of how such solutions are being adapted and applied to contemporary business problems. This gap in the literature, combined with the importance of continued consumer credit expansion in the context of Turkey's economic growth, provide ample justification for the research topic.

Financial Background

Table 1 contains the profitability data for the entire Turkish banking sector from Q4, 2002 to Q1, 2015:

Table 1

Turkish Banking Sector Profitability by Quarter, Q4, 2002-Q1, 2015

Quarter	Profitability
2002q4	1.1
2003q1	0.3
2003q2	1.1
2003q3	2
2003q4	2.2
2004q1	0.4
2004q2	0.9
2004q3	1.6
2004q4	2.1
2005q1	0.7
2005q2	1.2
2005q3	1.1
2005q4	1.4
2006q1	0.7
2006q2	1.1
2006q3	1.8
2006q4	2.3
2007q1	0.7
2007q2	1.5
2007q3	2.2
2007q4	2.6
2008q1	0.6
2008q2	1.3
2008q3	1.6
2008q4	1.8
2009q1	0.7
2009q2	1.4
2009q3	2
2009q4	2.4
2010q1	0.7
2010q2	1.4
2010q3	1.8

Quarter	Profitability
2010q4	2.2
2011q1	0.5
2011q2	0.9
2011q3	1.2
2011q4	1.6
2012q1	0.5
2012q2	0.9
2012q3	1.3
2012q4	1.7
2013q1	0.5
2013q2	0.9
2013q3	1.2
2013q4	1.4
2014q1	0.3
2014q2	0.7
2014q3	1
2014q4	1.3
2015q1	0.3

These profitability data are publicly disclosed by the Banks Association of Turkey , (BAOT, 2015) which has three measures of profitability. The ratio in Table 1 represents [(net profit / losses) / total assets]. There is a clear seasonal trend in the data, but not necessarily a directional trend:

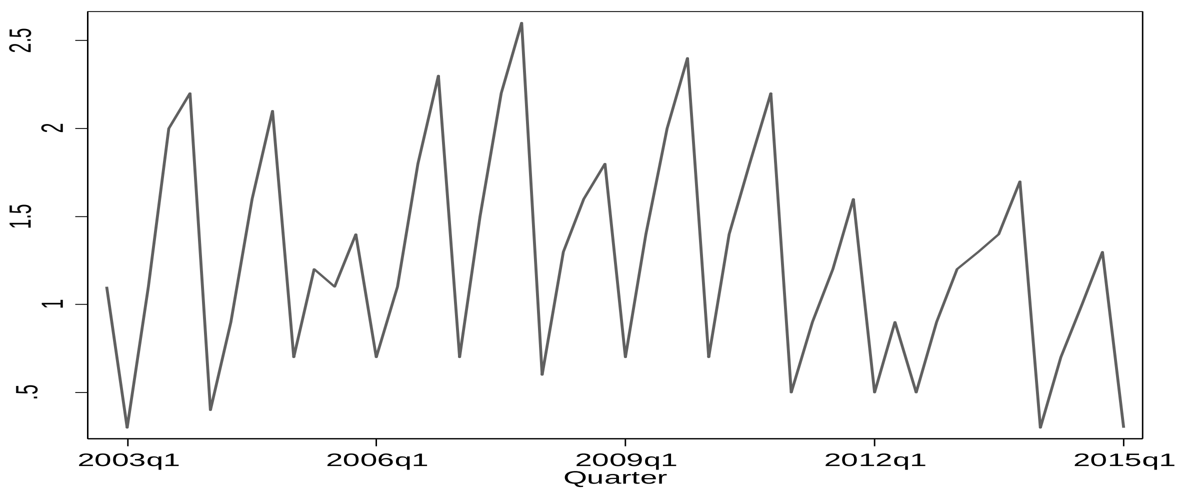

Figure 1. Line Graph, Turkish Banking Sector Profitability by Quarter, Q4/2002-Q1/2015

This trend indicates that Turkish bank profitability is not growing, which is the core of the business problem addressed in this study. Using the least squares method of regression, the adjusted $R2$ of this model is negative ($R2$ = -0.001394, t = 14.53501). This regression was performed by regressing year (the X variable) on profitability (the Y variable). The existence of a negative $R2$ indicates a bad fit for the OLS model, which, in simpler terms, means that Turkish banking profitability is neither increasing not decreasing over time. The inclusion of a time trend in the least squares regression does not substantially improve the model, which, in the adjusted version, takes on an adjusted $R2$ of 0.013922, t = 8.555830. These statistics indicate that, whether accounting for the seasonality clearly present in Figure 1 or conducting a least squares regression without the seasonal trend include, Turkish bank profitability has been essentially stable since 2002, which, for a growing economy, is not a good indicator. In fact, this finding is extremely surprising, given that the Turkish economy as a whole has experienced tremendous growth since 2002.

Moreover, these results stay substantially similar when the dependent variable is changed from [(net profit / losses) / total assets] to the other two measures of profitability reported by the Banks Association of Turkey, namely [(net profit / losses) / shareholders' equity] and [(profit / losses before taxes after continuing operations) / total assets]. By any of these three measures, there has not been a significant upwards movement in Turkish banks' profitability over the past 13 years.

The current study addresses only modest component of the profit formula of Turkish banking, namely profit; losses, equity, assets, and shareholders' equity are not considered. One possibility, given the flat profitability of the Turkish banking sector as a whole—from which the dataset for Table 1 was drawn—is that some banks are radically better performers than others.

Additionally, data analytics might be a plausible explanation for why some banks might be able to derive higher levels of profitability than others. If so, then there would be a strong empirical case for Turkish banks to adopt data analytics as a means of enhancing their profitability and competitive standing vis-à-vis other Turkish banks.

Research Purpose and Questions

The purpose of this quantitative, correlational study is to determine the financial utility of data mining techniques in terms of their ability to drive (a) profit margins and (b) revenue growth among Turkish banks.   This study aims to answer the following research questions:

RQ1: To what extent are advanced data mining techniques used among Turkish banks seeking new consumer credit business?

RQ2: To what extent are advanced data mining techniques associated with the profit margin of consumer credit activities among Turkish banks?

RQ3: To what extent are advanced data mining techniques associated with the revenue growth of consumer credit-derived revenues among Turkish banks?

By answering these research questions, the study aims to determine the financial utility of data mining techniques in terms of their ability to drive (a) profit margins and (b) revenue growth among Turkish banks. The hypotheses of the study are as follows:

The first research question is descriptive; therefore, no hypotheses have been associated with it.

H_0: Data mining techniques are not significantly associated with the profit margin of consumer credit activities among Turkish banks.

H_A: Data mining techniques are significantly associated with the profit margin of consumer credit activities among Turkish banks.

H_0: Data mining techniques are not significantly associated with the consumer credit-derived revenue growth of consumer credit activities among Turkish banks.

H_A: Data mining techniques are significantly associated with the consumer credit-derived revenue growth of consumer credit activities among Turkish banks.

The level of significance for the study will be 0.05.

Research Objectives

The most general research objective of the study is to determine the financial utility of data analytics in terms of their association with profitability and revenue growth among Turkish banks. This objective is connected to theories of knowledge management that specify that analytics can allow banks to derive actionable knowledge from otherwise static information about consumers and market dynamics. The more specific research objective of the study is to apply inferential techniques to a sample of Turkish banks in order to determine whether panel-based analysis can account for heterogeneity in banking performance. Whereas previous analyses of Turkish banking have often drawn upon time series analysis as well as analyses of the sector as a whole, the current study will be able to distinguish performance on a bank-to-bank basis. Ultimately, the study will build both theoretical and empirical knowledge around the relationship between data analytics and banking performance in Turkey.

Chapter 2. Literature Review

Introduction

The purpose of the literature review is to discuss theories and empirical studies related to the financial utility of data analytics in the banking sector. Theories of knowledge management and customer relationship management (CRM) are the most applicable ones in this topic. Afterwards, various aspects of data analytics will be discussed. Finally, aspects of data mining and analysis directly related to the topic will be discussed.

Applicability of Knowledge Management and CRM Theory

Dixon wrote of knowledge management that:

> Information [is] data that is ‘in formation’—that, data that has been sorted, analyzed, and displayed, and is communicated through spoken language, graphic displays, or numeric tables. Knowledge, by contrast, is defined as the meaningful links people make in their minds between information and its application in action in a specific setting. (Dixon, 2000).

In business terms, the main premise of data mining is to retrieve, prepare, and comb through information in order to turn it into knowledge that can be related to strategy. Thus, for example, a company that has data on thousands of customers can only be said to have information until these data are mined to yield insights into strategic perspectives, such as (a) who the most valuable customers are, (b) what kinds of products appear to customers, and (c) what kinds of sales behaviors are most closely associated with certain successful orders.

The broader theoretical framework for this study lies in a systems definition of CRM in the business-to-consumer (B2C) domain. The larger CRM system (Hosseini et al., 2013) possesses three interrelated, sequential components: (1) formation, (2) management, and (3) outcome. The stage of formation consists of B2C relationship strategy and the B2C relationship process, both of which provide inputs to the management stage of CRM. Business-to-consumer relationship strategy encompasses a company's vision for its customer relationships, while the B2C relationship process encompasses a company's specific set of plans for how such relationships ought to be governed. For example, in terms of B2C relationship strategy, a company that a smaller portion of customers who account for a higher proportion of spend are more likely to treat relationships more strategically, including the use of rewards, incentives, and other customer-driven initiatives, which also help to determine the tenor of the B2C relationship process. In other words, the nature of the formation stage of the CRM system is dependent on the dynamics between a business and its customers and how these dynamics inform the business's strategy.

Data mining has many possible uses in the banking sector. One obvious use is in the CRM context. Strategically, CRM has been defined as consisting of three domains: (1) formation, (2) management, and (3) outcome (Hosseini et al., 2013). Data mining addresses the second and third of these domains. By turning what Dixon (2000) referred to as information into knowledge, data mining allows banks to formulate improved strategies for managing their customers, such as by better segmenting them or by identifying the selling strategies, products, and campaigns to which they respond best. By supporting these kinds of strategic engagements with the customer, data mining also supports the outcome stage of CRM.

Hosseini et al. (2013) suggested that both CRM and data mining systems can be understood in terms of their impact on (a) measuring customers' values, needs, and behaviors; (b) applying these metrics to the customization of engagement with the customer; (c) using quantitative measures to determine how well particular selling strategies succeed in closing sales; and (d) applying a continuous-improvement mindset to future engagement with the customer. Thus, both CRM and data mining can be considered as part of a systems-oriented approach that amalgamates components including databases, point-of-sale records, portals and analytics that support management decisions, and various standalone components of CRM and KM software (Uzkurt et al., 2013, Ford et al., 2011, Koen et al., 2011, Remneland-Wikhamn et al., 2011).

Technical Details of Data Mining

One of the first steps in data mining is to obtain data from where originally reside. This challenge is referred to as *extraction*, which is the first step in the extract, transform, load (ETL) aspect of data analytics. The technological details discussed in this section of the literature review are relevant to Dixon's (2000) theory of knowledge management, in that ETL is the process that prepares information to be analyzed. Without ETL, information remains unamenable to actual analysis.

In terms of extraction, one of the main challenges is that data mining projects often require gathering data from several sources that might be incompatible with each other. For example, some data might reside within a relational database while other data are on the Internet. In order to extract these data, the software used for data mining purposes must be able to achieve several functions, including parsing. Dua and Chowriappa discussed parsing as follows:

> Parsing aids in the identification and isolation of individual data elements in the source data tables or files. It enables easier correction and standardizing and matching of data, as

it allows for comparison of individual components, rather than of long complex strings of data. (Dua and Chowriappa, 2012).

There are numerous ways in which data gathered from multiple sources might be similar or dissimilar to each other. Parsing is a way of obtaining precise insight into the nature of the data in order to achieve what Dua and Chowriappa (2012) described as correction and standardization. If the extracted data cannot be corrected or standardized, the output of a data mining project is not likely to be helpful. As Silvers has argued, “To avoid ‘garbage out,’ an ETL application must avoid ‘garbage in’” (Silvers, 2008). The goal of parsing, then, is to obtain consistent, high-quality data to inform subsequent stages of the data mining project (Bhardwaj and Singh, 2015).

Parsing is an example of the larger goal of extraction, which Silvers (2012) described as identification and control. All extracted data must be identified and controlled precisely for the extraction phase of ETL to count as a success. Given the disparity of data sources, the technical challenge is to employ an ETL program with the maximum ability to meet the identification and control requirements of extraction.

The challenges of transformation are both technical and process-related. As Silvers (2012) argued, it is unlikely that extracted data will be useful in precisely the manner in which they are extracted. As a result, it is necessary to transform the data in numerous ways. For example, it might be necessary to transform the data so that certain columns are left out, certain values are added together or subtracted from each other, new variables are created from existing variables, columns are transposed into rows, and so forth. The technical challenges of transformation pertain to the ability of an ETL product to be able to bring about these kinds of transformations. Meanwhile, the process-related challenge is for the users of the data mining project to be able to specify how the data need to be transformed.

According to Casters, Bouman, and Van Dongen, the main purpose of a load strategy is how "to update fact tables and…how to accommodate for late- or early-arriving facts" (Casters et al., 2010). As such, load strategies must pay specific attention to how and when facts are refreshed (Miller and Hutchinson, 2013). Another challenge that is involved in crafting data load strategies is how to deal with large volumes of data. In this regard, Hughes (2012) pointed out that some data load strategies might have to rely on loading only a subset of the data, which in turn raises questions of how to ensure that a selection of limited data can accurately represent an entire dataset.

Hornick, Marcade, and Venkalaya (2010) stated that "more and more applications are building in data-mining based intelligence, providing industry-specific and problem-specific interfaces, without users even being aware of data mining's presence. However, there are many situation-specific problems that still warrant customized solutions" (p. 81). When setting up a retrieval strategy, it is necessary to understand the scope of the data warehouse and the requirements served by the data mart. According to Baran and Galka (2013), "The data warehouse contains such a huge quantity of data that it may prove unwieldy…Data marts contain a subset of the data in the data warehouse and allow for more efficient analysis of the relevant portion of the firm's transactions" (p. 20). Retrieval strategies will thus depend in part on the size of datasets. If the data warehouse is indeed vast in size, that the retrieval strategy ought to be focused on identifying and gathering only those data that are relevant to specific kinds of decision-making. In such a strategy, analysts ought to devote much of their effort to determining which kind of data should populate which data mart. Creating discrete, relevant, and relatively small data marts is a way of ensuring that the data mart will be genuinely useful to different audiences within the business.

Data Mining in the Context of Banking

Suh (2011) noted that “Simple, blind application of data-mining algorithms can lead to the discovery of meaningless and useless ‘knowledge’ from databases” (p. 8). While the technical components of data mining discussed in the previous section of the literature review are a necessary component of any data demining study, they are not sufficient in themselves to generate, support, and improve the kinds of analytical insights banks need to improve their engagement with customers.

The literature suggests that the quality of data mining depends on (a) the kinds of analysis that are performed with data made available through data marts and other means and (b) how well business processes support changes driven by analysis (Kantardzic, 2011). Banks, for example, often utilize sophisticated statistical models in order to be able to predict customer responses (Moro et al., 2014). ETL and other data mining tools are necessary for such predictive reasons, because, without ETL, there would be no analyzable data. In addition, CRM, data mining, and other add-on tools often support an analytics function that allows banks to generate predictions from data. However, the quality of such analytics has been questioned in the literature, not as much because of technical reasons as because of business-related reasons (Moro et al., 2014).

For example, there are many kinds of analytical software that work together with the ETL component of data mining to provide an integrated solution for obtaining and analyzing customer data (Hosseini et al., 2013, Pan et al., 2006, Trainor et al., 2014). However, the analytical tools themselves cannot take the place of human analysis. For example, banks that try to predict customer behavior must choose from several statistical models that can carry out the task of prediction. For example, predicting a customer's balance at time *x*, based on data collected from several past points in time—a common analytical problem faced by banks—can be accomplished by linear or non-linear means, and by various kinds of traditional as well as time series regression approaches. While an analytics tool or add-on might technically support all of these kinds of analysis, the tool itself cannot inform the bank of which predictive model might be most appropriate. The bank has to have its own reasons to justify the use of particular models over others, in which case the expertise of human analysts becomes indispensable (Uzkurt et al., 2013).

The sophistication of data analytics and data mining systems require a similar degree of sophistication in the user, particularly after data mining systems undergo the ETL process and hand data off to data analytics. For example, in light of the problem of predicting a customer's deposit amount *x* at time *y*, the data analytics user needs the sophistication necessary to understand and run tests of stationary on the time series data, and also to fit the right kinds of models. If the appropriate analytical steps are not taken, then the output of the data mining processes is essentially worthless to the organization, remaining what Dixon (2000) classified as information rather than becoming knowledge.

Among banks in rapidly growing economies, such as Turkey, the use of data mining is tempered by various macroeconomic conditions. The Turkish economy had been dominated by the state itself, which increasingly gave way to private industry. As part of this transition, the Turkish banking sector came to benefit from a changing legal and regulatory framework that made credit more widespread.

The rise of consumer credit in Turkey over the past several years is a fascinating story in its own right, one without which the fate of the Turkish banking industry cannot be understood. At the beginning of the 2000s, total Turkish consumer credit was negligible. In the absence of liberal lending policies, and reflecting a longtime economic and cultural commitment to cash purchases, there was a minuscule credit market in Turkey. However, by 2015, Turkey has become one of the most credit-dependent countries in the Near and Middle East. This change has coincided with the rule of the Justice and Development Party (known by its Turkish acronym as the Ak Party) in Turkey, which began at the start of the 2000s and is ongoing.

The Ak Party has presided over a sea change in the Turkish banking sector. The easing of regulations has made it much easier for private, for-profit banks to be established and to offer credit to Turkish individuals and businesses (Uzkurt et al., 2013). The various banks that have sprung up in Turkey from 2000 onwards have operated in an environment of remarkable opportunity. Such banks have had the opportunity to market credit to consumers who are essentially new to this phenomenon, at least on a large scale, representing a so-called blue ocean strategic opportunity (Uzkurt et al., 2013).

Turkey’s banks are a mixed lot, representing (a) the state bank, (b) older private banks with their roots in Ottoman and early Republican times, and (c) newer private banks that are attempting to benefit from the post-2000 environment of liberalized credit and banking (Uzkurt et al., 2013). The state bank continues to hold important advantages in terms of capitalization, visibility, and statutory association with the payments of state officials. The older private banks are also well-capitalized and have had decades to align themselves with the best practices of international banking. Finally, the newer banks have benefited from the credit glut. However, as the data presented in Table 1 indicate, the profitability of Turkish banks, considered as a whole, has been flat from late 2002 onwards.

The flat profitability of Turkish banks, which is the identified problem in the current study, does not appear to have been studied in the empirical literature. In fact, scholars have called attention to the Turkish banking sector as being a thriving and vibrant portion of the Turkish economy. Given that the profitability of Turkish banking is flat, the main identified gap in the literature is that there is no plausible explanation of why, in a country with extremely suitable macroeconomic and microeconomic conditions for banks, there is not a rising secular trend in the profitability performance of Turkish banks. One plausible explanation is that banks might vary from each other in their application of advanced technology, such as data mining, that can be utilized to seize competitive advantage. Even in a market with high demand for credit, it is still possible for a handful of larger, more sophisticated banks to outperform their peers through the use of advanced technology and aligned business processes.

Conclusion

The economy of Turkey has long underperformed given the size of the country. In the 1980s, the sluggish Turkish economy, until then dominated by state-run enterprises and regulatory obstacles to the exercise of private banking, began to generate opportunities for banks as well as other kinds of businesses. As of 2015, Turkey is, by GNP, the 17th-largest economy in the world (WB, 2015). However, there are some signs that the Turkish economic miracle is in dire straits. Earlier in 2015, Turkish elections resulted in the failure of the incumbent Ak Party to win a majority for the first time in 13 years at the polls. At the same time, hundreds of thousands of Turks defaulted on their credit obligations (Poddar et al., 2015). In this environment, it is all the more urgent for Turkish banks to be able to understand how to drive profitability and revenue growth.

As discussed in the literature review, data mining—especially when combined with data analysis in an integrated, strategic approach to both knowledge management and CRM—has the potential to give banks numerous competitive advantages. These advantages can be described as generating knowledge from what is otherwise inert information (Dixon, 2000). They can allow banks to segment customers more accurately, predict deposit levels and balance portfolios accordingly, improve selling rates, and engage in various other actions capable of increase profitability and revenue growth. Based on the current empirical literature, it is unknown whether Turkish banks routinely use data mining, and, if so, how such use is correlated with beneficial financial outcomes. Chapter 3 contains a description and defense of a quantitative model that can be used to close this gap in the literature. The resulting knowledge can be of benefit to Turkish banks that have yet to make a business case for the use of analytics as a part of their banking strategies and operations.

Chapter 3. Research Plan

Introduction

The purpose of the research plan is to describe and defend the means by which the research questions of the study will be answered. The research plan has been divided into the following components: (a) an overview of generally available approaches to research, (b) the choice of the correlational quantitative research design, (c) data collection plan, and (d) data analysis plan. Finally, a brief conclusion will summarize the research plan. Note that all data gathered in the study were hypothetical, meant only to illustrate the proposed method.

Overview of Research Methods

The two main recognized forms of research methods are quantitative and qualitative research, whose orientations are summarized in Table 1 below (McNabb, 2010).

Table 1

Differences between Quantitative and Qualitative Research

Philosophical Foundations	Qualitative Research Designs	Quantitative Research Designs
Ontology (perceptions of reality)	Researchers assume that multiple, subjectively derived realities can coexist.	Researchers assume that a single, objective world exists.
Epistemology (roles for the researcher)	Researchers commonly assume that they must interact with their studied phenomena.	Researchers assume that they are independent from the variables under study.
Axiology (researchers' values)	Researchers overtly act in a value-laden and biased fashion.	Researchers overtly act in a value-free and unbiased manner.
Rhetoric (language styles)	Researchers often use personalized, informal, and context-laden language.	Researchers most often use impersonal, formal, and rule-based text.
Procedures (as employed in research)	Researchers tend to apply induction, multivariate, and multiprocess interactions, following context-laden methods.	Researchers tend to apply deduction, limited cause-and-effect relationships, with context-free methods.

Note: Adapted from McNabb (2010).

Quantitative research is associated with an analysis of the objectively defined relationship between mathematically coded variables (Balnaves and Caputi, 2001). Because the problem addressed in this study involved explanation of a mathematically coded phenomenon (low profitability among Turkish banks) through another mathematically coded phenomenon (degree of adoption of data mining), quantitative analysis was the appropriate research methodology for the study.

Choice of Correlational Quantitative Design

The research design adopted in this study was correlational. This quantitative design has been described as follows:

> The variables included in correlational research are isolated and measured by the investigator, but they are characteristics that occur naturally in the subjects…a correlation study consists of establishing a relationship between variations in the *X* variable to variations in the *Y* variable. (Keppel et al., 1992).

In this study, the potential explanatory factor of data mining was assessed for its relationship to the explained factors of profitability and revenue growth. The characteristics of data mining, profitability, and growth occur naturally among banks. They cannot be experimentally or pseudo-experimentally isolated; they are already occurring events beyond any researcher manipulation. Hence, the correlational approach was the appropriate design for this study.

Data Collection

Data collection for the study will be discussed in terms of data for the independent, dependent, and intermediate variables for the study. The independent variable for the study is the level of data mining adoption. For purposes of this study, data mining adoption was defined through the following questions:

- Q 1: Are all customer data used by your bank in a single database?
- Q 2: If yes to Q1, is access to customer data provided through an interface used by financial analysts at your bank?
- Q 3: Do you have predictive analytics software installed at your bank?
- Q 4: If yes to Q3, do you use software from SAS, IBM, or other providers to predict customer actions in the future?
- Q 5: Do you have segmentation analysis software installed at your bank?
- Q 6: If yes to Q5, does your bank create segment profiles every quarter?
- Q 7: Do you have a Chief Information Officer position at your bank?
- Q 8: Do you use data analytics to predict customer propensities to buy?
- Q 9: Do you use data analytics for up-selling or cross-selling support?
- Q 10: Do you use data analytics to identify your bank's most or least profitable products?

There are 10 questions in the instrument. An answer of 'yes' to any question is worth 2 points; an answer of 'no, but we have plans to do so within the next 12 months' is worth 1 point; and an answer of 'no, and we have no plans to do within the next 12 months' is worth 0 points. Thus, the scale will have a maximum possible value of 20 and a minimum possible value of 0. The scale value will represent the independent variable of the study, designated as data mining sophistication (DMS).

The DMS scale is a simple scale that minimizes ambiguity of response. Such an approach has advantages over an approach in which each question is measured continuously. The latter approach would be more vulnerable to participants' misperceptions and could not be triangulated through independent means. On the other hand, merely asking participants whether they have, and are implementing, certain forms of data analytics can address the threat posed by inaccuracy. Nonetheless, the scale provided in this proposal cannot be independently triangulated. It is also possible that there might be so-called floor or ceiling effects if there is a heavily non-Gaussian distribution in the histogram of data analytics scores, as considered in Chapter 4 of the study.

The dependent variables of the study have to do with (a) year-over-year revenue growth ascribable to the consumer credit segment and (b) year-over-year profitability growth ascribable to the consumer credit segment. These variables were calculated using data gathered in the following questions posed to the banks that participated in this study:

- Q 11: What were your bank's total revenues in calendar year 2013, in Turkish Lira? What portion of these revenues can be ascribed to the consumer credit segment?
- Q 12: What were your bank's total profits in calendar year 2013, in Turkish Lira? What portion of these profits can be ascribed to the consumer credit segment?
- Q 13: What were your bank's total revenues in calendar year 2014, in Turkish Lira? What portion of these revenues can be ascribed to the consumer credit segment?
- Q 14: What were your bank's total profits in calendar year 2014, in Turkish Lira? What portion of these profits can be ascribed to the consumer credit segment?

The study also contained intermediate variables. These intermediate variables were as follows: (a) the total revenues of the bank and (b) the age of the bank.

- Q 15: When was your bank founded?
- Q 16: What were the total 2014 revenues of your bank?

Data Analysis

Data analysis will be discussed under the heading of the different research questions of the study.

RQ1 Data Analysis

The first research question was as follows: To what extent are advanced data mining techniques used among Turkish banks seeking new consumer credit business? The first research question is descriptive; therefore, no hypotheses have been associated with it. This question was answered by conducting a histogram of the DMS scores for the sample of the study and presenting associated measures of central tendency. An important aspect of this analysis was to ensure that there were no marked floor or ceiling effects in the distribution of DMS scores, meaning that the scores fell into a roughly Gaussian distribution rather than being uniformly high (ceiling effect) or low (floor effect).

RQ2 and RQ3 Data Analysis

The second research question was as follows: To what extent are advanced data mining techniques associated with the profit margin of consumer credit activities among Turkish banks? The null hypothesis associated with this research question was that data mining techniques are not significantly associated with the profit margin of consumer credit activities among Turkish banks. The alternative hypothesis associated with this research question was that data mining techniques are significantly associated with the profit margin of consumer credit activities among Turkish banks.

The third research question was as follows: To what extent are advanced data mining techniques associated with the revenue growth of consumer credit-derived revenues among Turkish banks? The null hypothesis associated with this research question was as follows: Data mining techniques are not significantly associated with the consumer credit-derived revenue growth of consumer credit activities among Turkish banks. The alternate hypothesis associated with this research question was as follows: Data mining techniques are significantly associated with the consumer credit-derived revenue growth of consumer credit activities among Turkish banks.

It was possible to use statistical techniques to combine analyses of these two research questions. These research questions were answered in two ways, one of which drew upon an ordinary least squares (OLS) regression model and the other of which drew upon multivariate analysis of covariance (MANCOVA). In the OLS model, there were two separate regressions, one of DMS score on profit margin and another of DMS score on revenue growth. However, other OLS regressions were conducted in order to determine the mediating power, if any, of the study's two intermediate variables, which were (a) the year in which the bank was founded and (b) the total 2014 revenues of the bank.

For this aspect of the RQ2 analysis, the three-step procedure for mediation will be utilized, as also recommended by other scholars who have studied mediation (MacKinnon et al., 2000). In this procedure, the sequential steps are as follows, bearing in mind that Y = dependent variable, X = independent variable, M = mediating variable:

1. Regress X on Y. In this step, the existence of an effect that is mediated is established.
2. Regress X on M. In this step, the link between the independent variable and the mediator is established.
3. Regress X and M on Y. In this step, X is controlled when testing the effect of M on Y.

A MANCOVA was used to triangulate the findings in regression and also to study the effect of a profit * revenue growth interaction. In OLS regression, only one dependent variable at a time can be studied. However, in a MANCOVA, more than one dependent variable can be entered into the model ay once, allowing the interactions between dependent variables to be measured and thus providing a statistically viable means of combining the analyses for RQs 2 and 3. In addition, a MANCOVA allows for the placement of covariates into the model.

In terms of the covariates, bank age was treated as a dummy variable, with banks founded before 1990 being coded as 1, and banks founded in 1990 or afterwards coded as 0. This approach made it possible to discern a general effect of bank age, which might not have been detectable if bank age had been treated as a continuous variable. In Turkey, economic liberalism / neoliberalism dates from the late 1980s, so the 1990 cutoff appeared to be an appropriate way of determining the potential effects of bank age on the variables examined in the study.

Finally, because year-over-year profit and revenue growth might have been better measures of 'heat' than of more enduring profit and revenue growth, 5-year measures of these values were also taken and incorporated as dependent variables in robustness testing. Data for these variables were obtained from the 17 respondents to the study.

Conclusion

The purpose of the research plan was to describe and defend the means by which the research questions of the study will be answered. The research plan was divided into the following components: (a) an overview of generally available approaches to research, (b) the choice of the correlational quantitative research design, (c) data collection plan, and (d) data analysis plan. Because the problem addressed in this study involved explanation of a mathematically coded phenomenon (low profitability among Turkish banks) through another mathematically coded phenomenon (degree of adoption of data mining), quantitative analysis was deemed to be the appropriate research methodology for the study. Because the characteristics of data mining, profitability, and growth occur naturally among banks and cannot be manipulated experimentally or pseudo-experimentally, a correlational design was adopted for this study. Data were collected through the use of a custom-designed measure of data mining sophistication disseminated to 17 Turkish banks and analyzed through a combination of OLS regression and MANCOVA. The results associated with this research plan are presented in Chapter 4 of the study.

Chapter 4. Data Collection, Findings, and Analysis

Introduction

The purpose of this chapter of the study is to present and analyze the findings associated with the study. The findings will be presented and analyzed in the order of the research questions. Afterwards, a brief conclusion will summarize the significant findings and appropriate descriptive and inferential statistics of the study. Note that all data gathered in the study were hypothetical, meant only to illustrate the proposed method.

Findings and Analysis: RQ1

The first research question was as follows: To what extent are advanced data mining techniques used among Turkish banks seeking new consumer credit business? The first research question is descriptive; therefore, no hypotheses have been associated with it. This question was answered by conducting a histogram of the DMS scores for the sample of the study and presenting associated measures of central tendency. An important aspect of this analysis was to ensure that there were no marked floor or ceiling effects in the distribution of DMS scores.The first step in the analysis was to construct a histogram for the DMS distribution, after which key measures of central tendency could be reported upon.

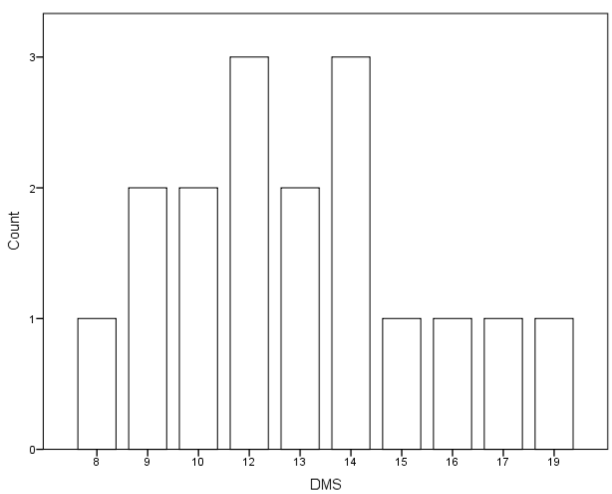

Figure 1. Histogram, DMS Scores.

The frequency table for DMS scores has been provided in Table 2 below, with chosen measures of central tendency following in Table 3. Note that there were 17 respondents in the study.

Table 2

DMS Scores: Frequency Distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	8	1	5.9	5.9	5.9
	9	2	11.8	11.8	17.6
	10	2	11.8	11.8	29.4
	12	3	17.6	17.6	47.1
	13	2	11.8	11.8	58.8
	14	3	17.6	17.6	76.5
	15	1	5.9	5.9	82.4
	16	1	5.9	5.9	88.2
	17	1	5.9	5.9	94.1
	19	1	5.9	5.9	100.0
	Total	17	100.0	100.0	

Table 3

Measures of Central Tendency, DMS Score

N	Valid	17
	Missing	0
Mean		12.76
Std. Error of Mean		.730
Median		13.00
Mode		12[a]
Std. Deviation		3.011
Variance		9.066
Skewness		.287
Std. Error of Skewness		.550
Kurtosis		-.322
Std. Error of Kurtosis		1.063
Range		11
Minimum		8
Maximum		19
Percentiles	25	10.00
	50	13.00
	75	14.50

Both skewness (*Skewness* = 0.287, *S.E.* = 0.550) and kurtosis (*Kurtosis* = -0.322, *S.E.* = 1.063) were tolerably close to 0, and the histogram displayed a roughly Gaussian structure. There did not appear to be any floor or ceiling effects in the DMS distribution. Given that the median value of the DMS scale was 10 (the midpoint between the minimum possible score of 0 and the maximum possible score of 20), the opportunity was taken to perform a one-sample *t* test to determine whether DMS score was significantly greater than 10. It was found that the mean value of DMS ($M = 12.76$, $S.D. = 3.011$) was indeed significantly greater than 10, $t(16) = 3.786$, $p = 0.002$. The one-sample *t* test indicated that the banks in the sample had a slightly greater than moderate level of data mining sophistication, at least as measured on the scale used in the study.

Findings and Analysis: RQ2 and RQ3

As described in Chapter 3, the findings for RQ2 and RQ3 were obtained by collapsing the variables for these two studies into a single model, which allowed for the use of MANCOVA. However, because OLS regression only admits one dependent variable at a time, separate OLS models were used to test the second and third research questions.

The second research question was as follows: To what extent are advanced data mining techniques associated with the profit margin of consumer credit activities among Turkish banks? The null hypothesis associated with this research question was that data mining techniques are not significantly associated with the profit margin of consumer credit activities among Turkish banks. The alternative hypothesis associated with this research question was that data mining techniques are significantly associated with the profit margin of consumer credit activities among Turkish banks.

The third research question was as follows: To what extent are advanced data mining techniques associated with the revenue growth of consumer credit-derived revenues among

Turkish banks? The null hypothesis associated with this research question was as follows: Data mining techniques are not significantly associated with the consumer credit-derived revenue growth of consumer credit activities among Turkish banks. The alternate hypothesis associated with this research question was as follows: Data mining techniques are significantly associated with the consumer credit-derived revenue growth of consumer credit activities among Turkish banks.

The first regression carried out was of DMS score on year-over-year profit margin growth, with correlation results presented in Table 4 below and regression results following in Table 4 below.

Table 4

Correlation, Profit Growth and DMS Score

		Profit Growth	DMS
Pearson Correlation	Profit Growth	1.000	.725
	DMS	.725	1.000
Sig. (1-tailed)	Profit Growth	.	.000
	DMS	.000	.
N	Profit Growth	17	17
	DMS	17	17

Profit growth and DMS score were strongly and significantly correlated, $R = 0.725$, $p < 0.001$. The coefficient of determination could be calculated from R as being 0.525, meaning that 52.5% of the variation in year-over-year growth profit growth could be explained through variation in DMS score.

Table 5

OLS Regression of DMS Score on Profit Growth

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.725[a]	.526	.494	1.52038	1.548

a. Predictors: (Constant), DMS
b. Dependent Variable: Profit Growth

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	38.453	1	38.453	16.635	.001[b]
	Residual	34.673	15	2.312		
	Total	73.126	16			

a. Dependent Variable: Profit Growth
b. Predictors: (Constant), DMS

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-4.516	1.653		-2.732	.015	-8.039	-.992
	DMS	.515	.126	.725	4.079	.001	.246	.784

a. Dependent Variable: Profit Growth

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-.3967	5.2668	2.0565	1.55027	17
Residual	-2.86275	2.50752	.00000	1.47210	17
Std. Predicted Value	-1.582	2.071	.000	1.000	17
Std. Residual	-1.883	1.649	.000	.968	17

a. Dependent Variable: Profit Growth

The regression was significant, $p = 0.01$. The regression equation was as follows:

Profit Growth = (DMS Score)(0.515) - 4.516

Hence, for every 1-point increase in the 20-point DMS scale, Turkish banks in the sample were observed to experience 0.515% greater year-over-year profit growth.

The second regression carried out was of DMS score on year-over-year revenue growth, with correlation results presented in Table 6 below and regression results following in Table 7.

Table 6

Correlation, Revenue Growth and DMS Score

		Revenue Growth	DMS
Pearson Correlation	Revenue Growth	1.000	.587
	DMS	.587	1.000
Sig. (1-tailed)	Revenue Growth	.	.007
	DMS	.007	.
N	Revenue Growth	17	17
	DMS	17	17

Profit growth and DMS score were moderately and significantly correlated, $R = 0.587$, $p = 0.007$. The coefficient of determination could be calculated from R as being 0.345, meaning that 34.5% of the variation in year-over-year growth revenue growth could be explained through variation in DMS score.

The regression was significant, $p = 0.013$. The regression equation was as follows:

Revenue Growth = (DMS Score)(0.593) - 4.320

Hence, for every 1-point increase in the 20-point DMS scale, Turkish banks in the sample were observed to experience 0.593% greater year-over-year revenue growth.

Table 7

OLS Regression of DMS Score on Revenue Growth

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.587[a]	.345	.301	2.54083	1.109

a. Predictors: (Constant), DMS
b. Dependent Variable: Revenue Growth

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	50.970	1	50.970	7.895	.013[b]
	Residual	96.838	15	6.456		
	Total	147.808	16			

a. Dependent Variable: Revenue Growth
b. Predictors: (Constant), DMS

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-4.320	2.762		-1.564	.139	-10.208	1.569
	DMS	.593	.211	.587	2.810	.013	.143	1.042

a. Dependent Variable: Revenue Growth

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.4226	6.9431	3.2470	1.78483	17
Residual	-3.89357	3.82077	.00000	2.46015	17
Std. Predicted Value	-1.582	2.071	.000	1.000	17
Std. Residual	-1.532	1.504	.000	.968	17

a. Dependent Variable: Revenue Growth

In order to test the mediation effects, if any, of bank age and total revenue, these variables were next regressed individually on profit growth and revenue growth (Tables 8-11). Afterwards, these intermediate variables were added to the OLS regression of DSM on profit growth and revenue growth (Tables 12 and 13, respectively). These regressions allowed the mediating power of the intermediate variables to be measured.

Table 8

OLS Regression of Bank Age on Profit Growth

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.271[a]	.073	.011	2.12562	1.104

a. Predictors: (Constant), Age
b. Dependent Variable: Profit Growth

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.353	1	5.353	1.185	.294[b]
	Residual	67.774	15	4.518		
	Total	73.126	16			

a. Dependent Variable: Profit Growth
b. Predictors: (Constant), Age

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	2.727	.803		3.394	.004	1.015	4.440
	Age	-1.140	1.048	-.271	-1.088	.294	-3.373	1.093

a. Dependent Variable: Profit Growth

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.5870	2.7271	2.0565	.57839	17
Residual	-3.92714	2.49300	.00000	2.05812	17
Std. Predicted Value	-.812	1.160	.000	1.000	17
Std. Residual	-1.848	1.173	.000	.968	17

a. Dependent Variable: Profit Growth

The regression of bank age on year-over-year profit growth was not significant, $p = 0.294$. It remained to determine whether age would become significant when added to the regressions of (a) DMS score on profit growth and (b) DMS score on revenue growth.

Table 9

OLS Regression of Bank Age on Revenue Growth

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.197[a]	.039	-.025	3.07768	.990

a. Predictors: (Constant), Age
b. Dependent Variable: Revenue Growth

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.726	1	5.726	.605	.449[b]
	Residual	142.081	15	9.472		
	Total	147.808	16			

a. Dependent Variable: Revenue Growth
b. Predictors: (Constant), Age

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	3.941	1.163		3.388	.004	1.461	6.420
	Age	-1.179	1.517	-.197	-.778	.449	-4.412	2.053

a. Dependent Variable: Revenue Growth

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2.7614	3.9407	3.2470	.59824	17
Residual	-5.07140	4.11133	.00000	2.97995	17
Std. Predicted Value	-.812	1.160	.000	1.000	17
Std. Residual	-1.648	1.336	.000	.968	17

a. Dependent Variable: Revenue Growth

The regression of bank age on year-over-year revenue growth was not significant, $p = 0.449$. Having conducted the regressions of the intermediate variable of bank age on profit and revenue growth, the next two regressions added the variable of revenue to that of age. These were the last two regressions before the combined model.

Table 10

OLS Regression of Bank Age and Revenue on Profit Growth

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.508[a]	.258	.152	1.96848	1.295

a. Predictors: (Constant), Revenue, Age
b. Dependent Variable: Profit Growth

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18.878	2	9.439	2.436	.124[b]
	Residual	54.249	14	3.875		
	Total	73.126	16			

a. Dependent Variable: Profit Growth
b. Predictors: (Constant), Revenue, Age

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-2.477	2.883		-.859	.405	-8.660	3.707
	Age	.986	1.496	.234	.660	.520	-2.221	4.194
	Revenue	.456	.244	.663	1.868	.083	-.067	.979

a. Dependent Variable: Profit Growth

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.7892	3.9060	2.0565	1.08621	17
Residual	-3.73827	3.06287	.00000	1.84134	17
Std. Predicted Value	-1.167	1.703	.000	1.000	17
Std. Residual	-1.899	1.556	.000	.935	17

a. Dependent Variable: Profit Growth

The regression of bank age ($\beta = 0.986$, $p = 0.520$) and revenue ($\beta = 0.456$, $p = 0.083$) on year-over-year profit growth was not significant, $p = 0.124$. The same regression was next conducted with the dependent variable or revenue growth in place of the dependent variable of profit growth.

Table 11

OLS Regression of Bank Age and Revenue on Revenue Growth

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.456[a]	.208	.095	2.89214	1.077

a. Predictors: (Constant), Revenue, Age
b. Dependent Variable: Revenue Growth

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	30.705	2	15.352	1.835	.196[b]
	Residual	117.103	14	8.364		
	Total	147.808	16			

a. Dependent Variable: Revenue Growth
b. Predictors: (Constant), Revenue, Age

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-3.131	4.236		-.739	.472	-12.216	5.954
	Age	1.711	2.197	.285	.778	.449	-3.002	6.423
	Revenue	.620	.359	.634	1.728	.106	-.149	1.389

a. Dependent Variable: Revenue Growth

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.6772	5.8282	3.2470	1.38530	17
Residual	-4.71600	4.13305	.00000	2.70535	17
Std. Predicted Value	-1.133	1.863	.000	1.000	17
Std. Residual	-1.631	1.429	.000	.935	17

a. Dependent Variable: Revenue Growth

The regression of bank age ($\beta = 1.711$, $p = 0.449$) and revenue ($\beta = 0.620$, $p = 0.106$) on year-over-year revenue growth was not significant, $p = 0.196$. Having conducted all of the pertinent intermediate regressions, the next two regressions included both DMS and the intermediate variables in the same model.

Table 12

OLS Regression of Bank Age, Revenue, and DMS Score on Profit Growth

Model Summary[b]

Model	R	R Square	Adjusted R Square	S.E.	Durbin-Watson
1	.730[a]	.534	.426	1.61971	1.477

a. Predictors: (Constant), DMS, Age, Revenue
b. Dependent Variable: Profit Growth

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	39.021	3	13.007	4.958	.016[b]
	Residual	34.105	13	2.623		
	Total	73.126	16			

a. Dependent Variable: Profit Growth
b. Predictors: (Constant), DMS, Age, Revenue

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	S.E.	Beta			Lower Bound	Upper Bound
1	(Constant)	-4.838	2.521		-1.919	.077	-10.283	.608
	Age	.172	1.265	.041	.136	.894	-2.562	2.905
	Revenue	.095	.239	.138	.397	.698	-.422	.612
	DMS	.468	.169	.659	2.771	.016	.103	.832

a. Dependent Variable: Profit Growth

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-.4309	5.0928	2.0565	1.56168	17
Residual	-3.01928	2.44542	.00000	1.45999	17
Std. Predicted Value	-1.593	1.944	.000	1.000	17
Std. Residual	-1.864	1.510	.000	.901	17

a. Dependent Variable: Profit Growth

The regression of bank age ($\beta = 0.172$, $p = 0.894$), revenue ($\beta = 0.095$, $p = 0.698$), and DMS score ($\beta = 0.468$, $p = 0.016$) was significant, $p = 0.016$. Note that the β coefficient of DMS score declined from 0.515 in the original model (see Table 5) to 0.468 when age and revenue were included, which means that age and revenue were slight moderators of the effect, but not the significance, of DMS score on profit growth.

Table 13

OLS Regression of Bank Age, Revenue, and DMS Score on Revenue Growth

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.602[a]	.362	.215	2.69289	1.063

a. Predictors: (Constant), DMS, Age, Revenue
b. Dependent Variable: Revenue Growth

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	53.536	3	17.845	2.461	.109[b]
	Residual	94.272	13	7.252		
	Total	147.808	16			

a. Dependent Variable: Revenue Growth
b. Predictors: (Constant), DMS, Age, Revenue

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-5.645	4.191		-1.347	.201	-14.698	3.408
	Age	.843	2.103	.141	.401	.695	-3.701	5.387
	Revenue	.235	.398	.241	.592	.564	-.624	1.095
	DMS	.498	.281	.493	1.774	.099	-.108	1.104

a. Dependent Variable: Revenue Growth

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.4051	6.4036	3.2470	1.82921	17
Residual	-3.95055	3.95209	.00000	2.42734	17
Std. Predicted Value	-1.554	1.726	.000	1.000	17
Std. Residual	-1.467	1.468	.000	.901	17

a. Dependent Variable: Revenue Growth

The regression of bank age ($\beta = 0.843$, $p = 0.695$), revenue ($\beta = 0.235$, $p = 0.564$), and DMS score ($\beta = 0.498$, $p = 0.099$) was not significant, $p = 0.109$. Note that the β coefficient of DMS score declined from 0.593 in the original model (see Table 7) to 0.498 when age and revenue were included. In addition, DMS lost its significance when age and revenue were included, which means that age and revenue were moderators of both the effect and the significance of DMS score on profit growth.

Finally, a MANCOVA was conduced with the following specifications. The independent variable, also known as the fixed factor, was DMS score, whereas the dependent variables were both profit and revenue growth. Bank age and revenue were specified as random factors. The model is presented in Table 14 below:

Table 14

MANCOVA Results, RQ2 and RQ3

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	Profit Growth	64.329[a]	11	5.848	3.324	.098
	Revenue Growth	119.080[b]	11	10.825	1.884	.251
Intercept	Profit Growth	6.435	1	6.435	3.657	.114
	Revenue Growth	14.170	1	14.170	2.466	.177
Age	Profit Growth	2.233	1	2.233	1.269	.311
	Revenue Growth	4.113	1	4.113	.716	.436
Revenue	Profit Growth	4.573	1	4.573	2.599	.168
	Revenue Growth	9.795	1	9.795	1.705	.248
DMS	Profit Growth	45.452	9	5.050	2.870	.129
	Revenue Growth	88.375	9	9.819	1.709	.288
Error	Profit Growth	8.797	5	1.759		
	Revenue Growth	28.728	5	5.746		
Total	Profit Growth	145.021	17			
	Revenue Growth	327.037	17			
Corrected Total	Profit Growth	73.126	16			
	Revenue Growth	147.808	16			

a. R Squared = .880 (Adjusted R Squared = .615)
b. R Squared = .806 (Adjusted R Squared = .378)

Overall, the MANCOVA demonstrated that, when both profit and revenue growth were treated as dependent variables, DMS lost its predictive power, with an observed *p* of 0.129 for profit growth and 0.288 for revenue growth. This finding, understood in conjunction with the findings for the OLS regressions conducted earlier in the chapter, indicates that data mining sophistication appears to be a more powerful predictor of profit growth than of revenue growth. Another point of note in the MANCOVA is that the adjusted $R2$ of profit growth was much higher (0.615) than the equivalent figure for revenue growth, indicating that DMS, bank age, and revenue are far weaker predictors of revenue growth than of profit growth.

In the interests of robustness testing, the MANCOVA was repeated for 5-year average of profit margin and year-over-year revenue growth, with results presented in Table 15 below.

Table 15

Robustness Testing of MANCOVA Tests

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	Profit Margin 5y	424.507[a]	11	38.592	.774	.665
	Revenue Growth 5y	28.480[b]	11	2.589	.727	.695
Intercept	Profit Margin 5y	116.662	1	116.662	2.339	.187
	Revenue Growth 5y	.099	1	.099	.028	.874
Age	Profit Margin 5y	14.310	1	14.310	.287	.615
	Revenue Growth 5y	.140	1	.140	.039	.851
Revenue	Profit Margin 5y	33.776	1	33.776	.677	.448
	Revenue Growth 5y	.758	1	.758	.213	.664
DMS	Profit Margin 5y	384.599	9	42.733	.857	.605
	Revenue Growth 5y	23.520	9	2.613	.734	.677
Error	Profit Margin 5y	249.338	5	49.868		
	Revenue Growth 5y	17.811	5	3.562		
Total	Profit Margin 5y	8301.610	17			
	Revenue Growth 5y	335.350	17			
Corrected Total	Profit Margin 5y	673.845	16			
	Revenue Growth 5y	46.291	16			

a. R Squared = .630 (Adjusted R Squared = -.184)
b. R Squared = .615 (Adjusted R Squared = -.231)

In this MANCOVA, DMS score was a weaker predictor, and the appearance of negative values for adjusted $R2$ indicated a bad model. In order to further explore these findings, two further OLS regressions were conducted. In the first regression, DMS score, bank age, and revenue were regressed on 5-year profit margin average. In the second regression, DMS score, bank age, and revenue were regressed on 5-year revenue growth average.

The regression of bank age ($\beta = -5.510$, $p = 0.287$), revenue ($\beta = -1.625$, $p = 0.107$), and DMS score ($\beta = 1.091$, $p = 0.123$) on 5-year profit margin average was not significant, $p = 0.337$. In addition, the regression of bank age ($\beta = 0.994$, $p = 0.502$), revenue ($\beta = 0.172$, $p = 0.517$), and DMS score ($\beta = 0.133$, $p = 0.477$) on 5-year revenue growth average was not significant, $p = 0.557$. These robustness checks indicate that DMS has no significant bearing on longer-term measures of profitability and revenue growth, even though DNS appears to be associated with shorter-term measurements of these qualities. The implications of this finding will be discussed further in Chapter 5 of the study.

Summary

The following findings were obtained in the study. First, it was observed that the distribution of DMS scores was roughly Gaussian. There did not appear to be any floor or ceiling effects in the DMS distribution. Additionally, it was found that the mean value of DMS ($M = 12.76$, $S.D. = 3.011$) was significantly greater than 10, $t(16) = 3.786$, $p = 0.002$, meaning that the banks in the sample had a slightly greater than moderate level of data mining sophistication as measured by the scale created for the study.

Profit growth and DMS score were strongly and significantly correlated, $R = 0.725$, $p < 0.001$. The coefficient of determination could be calculated from R as being 0.525, meaning that 52.5% of the variation in year-over-year growth profit growth could be explained through variation in DMS score. The regression equation was as follows: *Profit Growth = (DMS*

Score)(0.515) - 4.516. Hence, for every 1-point increase in the 20-point DMS scale, Turkish banks in the sample were observed to experience 0.515% greater year-over-year profit growth.

Revenue growth and DMS score were moderately and significantly correlated, $R = 0.587$, $p = 0.007$. The coefficient of determination could be calculated from R as being 0.345, meaning that 34.5% of the variation in year-over-year growth revenue growth could be explained through variation in DMS score. The regression equation was as follows: *Revenue Growth = (DMS Score)(0.593) - 4.320.* Hence, for every 1-point increase in the 20-point DMS scale, Turkish banks in the sample were observed to experience 0.593% greater year-over-year revenue growth.

The regression of bank age ($\beta = 0.172$, $p = 0.894$), revenue ($\beta = 0.095$, $p = 0.698$), and DMS score ($\beta = 0.468$, $p = 0.016$) was significant, $p = 0.016$. The β coefficient of DMS score declined from 0.515 in the original model to 0.468 when age and revenue were included, which means that age and revenue were slight moderators of the effect, but not the significance, of DMS score on profit growth. The regression of bank age ($\beta = 0.843$, $p = 0.695$), revenue ($\beta = 0.235$, $p = 0.564$), and DMS score ($\beta = 0.498$, $p = 0.099$) was not significant, $p = 0.109$. The β coefficient of DMS score declined from 0.593 in the original model to 0.498 when age and revenue were included. In addition, DMS lost its significance when age and revenue were included, which means that age and revenue were moderators of both the effect and the significance of DMS score on profit growth.

The effect of DMS weakened in a MANCOVA model in which revenue and profit growth were included. In addition, DMS had no effect when the dependent variables in OLS regression were changed to 5-year averages of profit margin and revenue growth.

References

BAOT. 2015. *Statistical reports* [Online]. Available: http://www.tbb.org.tr/en/banks-and-banking-sector-information/statistical-reports/march--2015---quarterly-statistics-by-banks-%28all-periods%29---usd/1053.

BHARDWAJ, A. K. & SINGH, M. 2015. Data mining-based integrated network traffic visualization framework for threat detection. *Neural Computing and Applications,* 26**,** 117-130.

ÇARKOGLU, A. & KALAYCıOĞLU, E. 2007. *Turkish Democracy Today: Elections, Protest and Stability in an Islamic Society,* London, I.B. Tauris.

CASTERS, M., BOUMAN, R. & VAN DONGEN, J. 2010. *Pentaho Kettle solutions: building open source ETL solutions with Pentaho Data Integration,* New York, NY, John Wiley & Sons.

CIDDI, S. 2009. *Kemalism in Turkish Politics: The Republican People's Party, Secularism and Nationalism* New York, Routledge.

COOPERS, P. 2012. *New Turkish Commercial Code* [Online]. Available: http://www.pwc.com/tr/tr/publications/assets/TTK-A_blueprint_for_the_future.pdf.

DIXON, N. M. 2000. *Common knowledge: How companies thrive by sharing what they know,* Cambridge, MA, Harvard Business Press.

DUA, S. & CHOWRIAPPA, P. 2012. *Data mining for bioinformatics*, New York, NY, CRC Press.

FORD, D., GADDE, L.-E., HÅKANSSON, H. & SNEHOTA, I. 2011. *Managing business relationships*, New York, NY, Wiley.

HITT, M. A., IRELAND, R. D. & HOSKISSON, R. 2009. *Strategic management: Competitiveness and globalization*, Mason, South-Western.

HOSSEINI, M., NEMATI, B. & SADEGHI, N. 2013. An investigation on relationship between CRM and organizational learning through knowledge management: A case study of Tehran travel agency. *Management Science Letters,* 3**,** 999-1006.

KANTARDZIC, M. 2011. *Data mining: concepts, models, methods, and algorithms,* New York, NY, John Wiley & Sons.

KASMAN, A. 2012. Cost efficiency, scale economies, and technological progress in Turkish banking. *Central Bank Review,* 2**,** 1-20.

KOEN, P. A., BERTELS, H. M. & ELSUM, I. R. 2011. The three faces of business model innovation: challenges for established firms. *Research-Technology Management,* 54**,** 52-59.

MILLER, S. & HUTCHINSON, W. 2013. *Oracle business intelligence applications,* New York, NY, McGraw-Hill.

MORO, S., CORTEZ, P. & RITA, P. 2014. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems,* 62**,** 22-31.

OMURGONULSEN, M. O. U. 2009. Critical thinking about creative accounting in the face of a recent scandal in the Turkish banking sector. *Critical Perspectives on Accounting,* 20**,** 651-673.

OZTURK, I. & ACARAVCI, A. 2013. The long-run and causal analysis of energy, growth, openness and financial development on carbon emissions in Turkey. *Energy Economics,* 36**,** 262-267.

PAN, S.-L., TAN, C.-W. & LIM, E. T. 2006. Customer relationship management (CRM) in e-government: a relational perspective. *Decision Support Systems,* 42**,** 237-250.

PODDAR, A., ELLIS, C. M. & OZCAN, T. 2015. Imperfect recall: The impact of composite spending information disclosure on credit card spending. *Journal of Consumer Policy,* 38**,** 93-104.
REMNELAND-WIKHAMN, B., LJUNGBERG, J., BERGQUIST, M. & KUSCHEL, J. 2011. Open innovation, generativity and the supplier as peer: The case of iphone and android. *International Journal of Innovation Management,* 15**,** 205-230.
ROSENBERGER, L. E. & NASH, J. 2009. *The Deciding Factor: The Power of Analytics to Make Every Decision a Winner*, John Wiley & Sons.
ŞEN, M. 2010. Transformation of Turkish Islamism and the Rise of the Justice and Development Party. *Turkish Studies,* 11**,** 59-84.
SILVERS, F. 2008. *Building and maintaining a data warehouse,* New York, NY, CRC Press.
SRNEC, K., NERADOVÁ, A. & HES, T. 2013. Socio-economic profile of village bank member in suburban Ankara: microcredit mannequin and assumptions on microfinance market of Turkey. *International Letters of Social and Humanistic Sciences***,** 55-75.
TRAINOR, K. J., ANDZULIS, J. M., RAPP, A. & AGNIHOTRI, R. 2014. Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM. *Journal of Business Research,* 67**,** 1201-1208.
UZKURT, C., KUMAR, R., SEMIH KIMZAN, H. & EMINOGLU, G. 2013. Role of innovation in the relationship between organizational culture and firm performance: A study of the banking sector in Turkey. *European Journal of Innovation Management,* 16**,** 92-117.
WB. 2015. *Development indicators* [Online]. Available: http://data.worldbank.org.