Journal of Finance & Banking Studies 7(2), 2018: 01-08

IJFBS
Finance & Banking Studies

Finance & Banking Studies

IJFBS, VOL 7 NO 2 ISSN: 2147-4486
Contents available at *www.ssbfnet.com/ojs*

https://doi.org/10.20525/ijfbs.v7i2.877

Data Quality in Banking System: Case of Azerbaijan

Ali Dadashzade

University of Bath, Bath, UK

Abstract

Data quality in banking and financial sector is one of the most researched topics nowadays. With the increasing regulatory burden and increased importance of targeted sales, data quality directly influences funds and performance of banking system. In this paper, the author is aiming to define universal reasons and causes of data quality problem and apply the case to local Azerbaijani banks taking into account local managers' personal view based on their banking experience. Key finding of the research is that unintegrated software, wrong data insertion, aging of data with the growing speed of market, corporate governance and inability to calculate true costs of low data quality to the local banks are the reasons of data quality issue in the local banks. Moreover, main costs of the data quality issue are time and money, appearance of hidden data factories, obstacles to apply and measure KPIs, uncorrelations in sensitivity analysis and ineffective marketing strategies.

Keywords: *Data Quality; Business Administration; Banking; Azerbaijani Banks*

JEL classification: *M100* *Submitted: 24.09.2018* *Accepted: 05.11. 2018*

Introduction

Banks have many reasons to be concerned about the quality of their data. Main reasons for the problem are that customer-based banking services have become more diverse, much rapid that, rate at which data is becoming obsolete has drastically increased. Therefore, banks need more sophisticated methods to store, proceed and employ the data available.

Today many banking organizations struggle with the changing methods of aggregating and managing vast amounts of data. Not being able to receive and store best data possible, Banks lose both financially and non-financially. Organizations are discovering that data quality deficiencies have a significant impact on their most strategic business initiatives and these deficiencies do not let banks to achieve the growth, agility and competitiveness they aim. Due to compliance and transparency pressures, it is no longer acceptable to ignore flaws in data and organizations must prove the accuracy of information that they report to auditors, regulators and the public.

This paper is aiming to describe the reasons for data quality problem, its financial and non-financial causes for banks. To find out the roots of the problem and to understand stakeholders' view of the problem, regional banks' managers were interviewed face-to-face and through questionnaire.

Reasons for Data Quality Problem

Most IT and business people understand that poor data quality is a business problem. Study by Ecosystems Insight (2009) show that up to 25% of data in an average bank's CIF/CIS is incorrect (Ecosystems, Insight, 2009). Errors in customer data lead to numerous issues that impact a bank's earnings. Data quality problems start a chain reaction in a bank's business processes.

Problems with data quality are costly to an enterprise. When facing the potential for missed opportunities, uninformed decision-making, non-compliance sanctions, and low customer satisfaction, today's business leaders are making data quality a priority in their organizations' data management programs.

Understanding the reasons behind common data quality issues is an essential first step in an effective data quality initiative. In this part of the paper, main sources of data quality problem are listed and explained. Moreover, local managers' views are also considered to understand importance and regularity of reasons.

Having incompetent software. Are existing packages integrated?

For most data-driven organizations, seeing where your business is heading means understanding where you've been. In other words, these businesses need to perform analytics on historical performance data. By aggregating and structuring transactional data from various sources used in the business's operations, traditional data warehouses serve as a "single source of truth" for reporting and analysis (Oracle, 2017). McKinsey recently shed light on the use of analytics in the banking industry to maximize business performance. In fact, analytics help banks improve retention, optimize discounting, and segment their customer base to effectively acquire new customers (Garg, Grande, Miranda, & Sporleder, 2017). Banks run different applications and have various systems in place. If all financial data was kept in just one system, it would be much faster to gather and analyse data.

Banks make huge investments in software packages to gather and analyse the data. But in many cases, the software invested in can be incompetent to gather the data required to analyse. For instance, core banking systems are different than tools employed for analytics. Data warehouses usually gather not just from the core system, but from various other software which might not be fully integrated. Trying to understand data across multiple systems can be a challenge. The typical banking professional won't understand the databases housing the data and the calculations necessary to extract and transform data. With the growing amount of data—and spread across systems—it's more vital now than ever to bring it together in one spot.

It is also worth to mention that most of the top data warehousing vendors have their own suit of solutions/products. Vendors tend to promote their own solution rather than advocating what is best suited for the customer (Neri, 2012).

Blank Data and Mandatory Fields

For data to be existing either the system must be capable enough to catch and store it or the data must be imputed manually. Companies usually prioritise data by adding mandatory fields which are obstacle for leaving the field blank. More often mandatory fields are defined to comply with regulatory rules. By defining mandatory fields by their regulatory importance, other fields are left blank which brings in data quality problem (Steerman, 2016). For banks it is important to define mandatory fields in accordance to their business needs. Due to large number of customers and operations, employees are eager to leave non-mandatory fields blank, as they think of them as "unnecessary bureaucratic information". There are plenty of users filling data to the system, but the users themselves are a part of big business process which they are usually not aware of. Employee of a banking branch located away from the head-office has limited knowledge of what can be consequences of not-filling the data.

Often automated processes are used to fill in large volumes of similar data in batches, as this saves effort and time. The systems pushing this bulk amount of data may end up inserting huge amounts of wrong data as well. This can be quite disastrous, especially when data travels down several databases in series.

Not All Data Filled in is Correct

Having wrong data might have lethal influence on decision making. Therefore, having data fields filled in does not mean it creates any value for the company. Risks for having wrong data is even higher when human error risk is substantial (World Bank, n.d.). For example, a bank having retail sale campaign for young entrepreneurs aged between 25-35, then it becomes a must to have the correct data of customers to understand who fills in the category. If that data is filled in wrongly, the Bank will not direct its sales to targeted group.

Several motives may be counted as motives for inserting wrong data. It can be both intentional or unintentional. If employees who fill in data does not have standardised approach on how and what to fill in, error is inevitable. Of course, employees do not enter bad data into systems intentionally, but inevitably their level of attention is directly related to their motivation. A sales representative taking an order cares a lot about the customer details and payment details, since his commission probably depends on that. But is he quite so fussed about the credit reliability of the customer or the demographic background information that marketing has asked telesales workers to capture? The main question and solution lie in this question.

Part of the problem can be solved through trainings. Moreover, attitude of filling in wrong data just for filling in the fields can be addressed by adding control mechanisms that assist to standardise data.

The problem also arises when the employee thinks she has inserted correct data, whereas, in fact it is wrong. For instance, when phone number is inserted as “111-111-111” it is easy to catch that through the system. On the other hand, if phone number inserted is a real phone number, but not the customer’s, then neither the system, nor the control mechanisms can detect it.

It is also worth to mention that another reason for the wrong data problem is that, quite often customers are not too keen to share their true information. This trend is more experienced when phone numbers are requested. With the developing IT technology and social sets, people develop feeling of being monitored and controlled which motives them to “hide their customer profile”.

Problem of aging data

Most data are inserted to the system at a point of time. Data represents real-world objects which may change on their own with time, and the data representation might not be catching up with this change. Thus, this data gets automatically aged and transformed into a meaningless form. In such circumstances, continuing to rely on old data does not just fail to add value, it destroys value of fresh data. Therefore, banks and any companies, have huge incentive to use data so long as it remains productive.

For instance, banks mostly collect customer information when customers visit the bank for operations. If bank keeps sending advertisements to the same old phone numbers, the Bank will fail to advertise its product and will also bear costs. It will also affect fresh data as the customer will be categorised as one without any interest to the new product.

Aging data is not totally useless. Old data can be a good starting point for future predictions. Online retailers use old data about the customer to understand its preferences.

Corporate Governance: Demanding Information Derived from Data.

Although part of the responsibility for the data quality lies on the shoulders of data inserters, corporate governance within the business plays crucial role as well (Michelberger, 2016). Nowadays influence of culture and working environment on individuals is massive. With the increasing number of working hours, individuals become more aligned with the culture. In companies with historical culture, it takes few months for new-comers to be shaped-out (O'Reilly, Chatman, & Caldwell, 1991). The culture of “taking the correct action even if nobody detects” i.e. to enter all available data even if nobody monitors it, might be solution to the problem of data quality. Unfortunately, this culture is not widespread as years of standard long-term corporate governance is needed to create the culture.

Individuals might not be too much interested in inserting full data if there is no any governance preventing individuals from “wrong action”. The problem arises at this point. If wrong data is not seen as crucial issue for the enterprise, then data inserters will not be motivated or taught to do it right.

Of course, strict monitor and control functions might have influence on data quality problem. Sense of being monitored and controlled always works to prevent wrong action. On the other hand, strict monitoring also results in individuals hiding wrong actions (Yerby, 2013).

Cost vs. Benefit of Inserting Full Data: Is Benefit Measured?

Typically, if we can improve the quality of data, then we will have better information to support decision making. Better decisions will lead to better outcomes and will in turn be likely to have better quality data arising from them. In any organization with data quality problem, quite often managers’ typical excuse is cost to be incurred when improving data quality. Typical costs include re-designing the system to fit purpose, additional workforce and time. As these costs are direct and can be easily measured, it is easy to calculate cost of having it. But what is the cost of not having reliable data? It is hard to define the exact monetary impact of having good data over bad data, as calculation includes significant estimates. 68% of companies do not calculate the cost of data quality to the business (Hayler, 2011). Interestingly, to define approximate estimate, entities need good data itself. So, the problem is rolling back to the starting point: bad data problem. According to quality management gurus: “Data quality is free. It’s not a gift, but it’s free (Crosby, 1979). What costs money are the things without quality – all the actions that involve not getting data quality right the first time and all the actions to correct these data quality issues” (Crosby, 1979). If managers only think of cost of improving data quality and benefits are not defined or approximated reasonably, long-term benefits of good data will be ignored for the sake of short-term win of “cost control”.

Survey Results and Main Findings about Reasons of Bad Data Quality

Local banks’ middle and upper managements’ view about the problem is not one sided. Irrespective of background, almost all the individuals interviewed, agree on the fact that using more than one software that are not directly linked is one of the serious reasons of data quality problem. In the local market, all banks use more than one software which is the result of competitiveness; banks offer wider range of products. One of the managers directly coping with data quality problem think that non-standardisation of data is the key problem. If banks could exert normative standards of data inserting, no matter where information is extracted from, information could be easily integrated to systems.

Views about inserting wrong data and having mandatory fields are very diverse. Head office managers see the problem as lethal to reporting system. Finance departments struggle a lot with wrong or missing data as it is their job to report both internal and externally and quality of data directly influences reporting. Mid-level branch managers, however, think that with the mass amount of daily operations, employees are doing their best to fill in the information. Most of them could not predict where the data they fill in (or do not fill in) is needed. Moreover, managers at branch level define their job-role mostly operational i.e. double entry and documentation. Therefore, the problem is mid-level managers, apart from those managers in head-offices do not know end-users of the information, therefore they do not assess effects of non-filling.

Furthermore, interviewed managers accept that aging data creates serious problem. One of the marketing managers’ view was that, easy-access to all social and private utilities in late years allow society to be more flexible; i.e. it becomes easy to change phone number, to change residential places, to change taste and so on. As there is not a single integrated data warehouse where banks could exploit data from, it is hard to update aging data.

Local managers’ view about corporate governance problem is very diversified. Smaller banks’ managers accept that corporate governance have impact on quality of data inserted. Whereas, managers of larger banks are in the thought that, corporate culture is hard to change, and it is not too much effective tool. Interestingly, managers with previous “Big Four” experience think culture is key to success, whereas, those with only local bank experience think vice-versa. One of the managers even expressed view that, “believing that data quality problem can be solve via good corporate culture is a myth”. Manager’s point was that data quality can only be resolved via strong demand from regulatory bodies.

Finally, managers were interviewed whether any of them has assessed cost of lost opportunity due to data quality problem. General sense was that it is early for the emerging market to calculate lost opportunity where data is an issue.

Causes of Data Quality

Bad Data Costs Time and Money: Massive Worldwide Effect

$3.1 trillion, IBM's estimate of the yearly cost of poor-quality data, in the US alone, in 2016 (Redman, Bad Data Costs the U.S. $3 Trillion Per Year, 2016). According to Redman (2016), with the bad data quality, "hidden data factory" appears within the companies. Salespeople waste time dealing with erred prospect data; service delivery people waste time correcting customer orders received from sales. Data scientists spend time cleaning data; IT expends enormous effort lining up systems. Senior executives hedge their plans because they don't trust the numbers from finance. With the quality problems, some individuals turn into employees of hidden factory and their role become more about data hunting and correcting rather than analysing the information. Research by various institutes and research companies indicate that 50% of time of knowledge workers is wasted by hunting for data, finding and correcting errors (Redman, Data's Credibility Problem, 2013). Nearly one third of analysts spend more than 40 percent of their time vetting and validating their analytics data before it can be used for strategic decision-making (Redman, Data's Credibility Problem, 2013).

No reasonably well-informed external customer would pay more for these steps. Thus, the hidden data factory creates no value. But as employees in the factory know what the problem is and correct them manually, if the system is not integrated so that to correct data automatically, these employees become "stars" of their departments and managers become dependent on the information from the factory. This is a common problem of the banks experiencing data quality problem.

Setting Budget and KPIs with Bad Data: Is There Bottom-up Approach? How deep Can Managers Dig?

Process efficiency and effectiveness are two of the most important determinants of business success (Lynch & Cross, 1992). Thus, performance measurement becomes a critical business activity. As it is with the governments, companies also track and control financial results with proper budgeting. With the growing number of operations and types of products, financial performance is segmented into parts and control is sustained through budgeting. It would be a lot easy to construct budget in total numbers such as total interest income, interest expense, operating expense and so on. But that is not enough for today's businesses. With the development of businesses and increased number of applications of bonus schemes, it is vital for companies to make deep analysis to assess performance of the entity (Shahin & Mahbod, 2007). To add, setting and assessing logical KPIs also depends of quality of data. Business organisations must achieve efficient and effective processes to attain competitive positions. The performance of these processes is measurable through key performance indicators (KPIs). KPIs are very important to management decision-making and are relied upon by all levels of an organisation to measure success in achieving outcome. For instance, if the is not special product code for restructured loans and sales managers have KPI that depends on NPL, there will not be fair assessment of performance.

Risk Assessment and Sensitivity Analysis: Estimates Based on Current Data is not Always Reliable.

Credit data is the collection of records in the bank's database that describes the bank's borrowers, including details on their outstanding loans and posted collateral (World Bank, 2007). The credit risk models that are used to calculate the PD, LGD and EAD risk measures rely on good quality credit data for making reliable risk assessments. Firstly, because a bank cannot refrain from using its historical credit data during model development since it "must incorporate all relevant, material and available data, information and methods" (Hanmath & Shivaji, 2014).

Secondly, credit risk models are generally developed as statistical regression models. Such models perform better if the data used as input is completer and more accurate. Unreliable credit risk data can lead to unreliable risk measure estimates. The data must be credible enough to rely upon when making estimates of the future. For being a step ahead of competitors, it is vital for any entity to make best judgements and analysis.

Marketing and Targeted Sales: Wrong Target Problem

The amount of money spent on marketing is growing, and the way we spend is changing (Statista, 2018). Every year marketers across the world waste millions of dollars because of poor data (Haug, 2011). Poor data can not only impact your company's bottom line, but it can also devastate your brand's image. The data in your database is becoming worthless by the minute if you have no data quality strategy in place. To compete in tomorrow's business landscape, you'll need to use your marketing budget more efficiently than ever before. According to researches, 67 per cent of businesses say some of their marketing emails bounce back after a campaign 70 per cent of businesses report data quality problems in loyalty programs, 22 per cent of contact data is thought to be inaccurate (Doyle, 2015). MarketingSherpa found out that 25%-30% of all data becomes inaccurate and it has lethal monetary effect on marketing and drastic effect on marketing effectiveness (MarketingSherpa, 2007).

One of the most promising trends in marketing right now is individualized multichannel product offering. The ability to target customers and prospects both offline and then online through various devices and sites with tailored messaging based on their behaviour and unique profile is a marketer's dream. It is very common for local banks to send messages to targeted audience about special days such as "Doctors' day", "Oil-workers' day" and so on. It is used as a massive marketing tool to reach customers and supply them with information about the campaigns. But it is not for sure that Banks reach the targeted audience. Unfortunately, again due to data quality problems, these costs are not usually analytically aligned with the profit made from the campaign. Therefore, local banks do not analyse effectiveness of marketing tools.

If a business is not connecting with the right people or organizations, it's directly missing out on opportunities to find prospects and obtain stronger leads.

Bad data can send you chasing leads that aren't there, writing messages to CFOs when you need to be targeting CIOs, or advertising only to multi-billion corporations when your best customers are in the $10 million per year range. Bad data can make your business look (and feel) like it doesn't know what it's doing.

Survey Results and Main Findings about the Causes of Bad Data

It is worth to mention that local managers' understanding of causes of the problem is mostly unique. One of the managers with analytics experience shares the view that, most of his working time was spent to bring data provided into normal reportable format. Too much work done on "tiding" led to less time to analyse.

Moreover, setting KPI is accepted as a general problem due to hardness of gathering deep enough data to assess performance. For instance, one of the budgeting managers think that personal performance is assessment best for front-office workers, treasury and other profit centres, whereas back-office workers are hard to assess based on given data. Furthermore, managers also accept that KPIs are not calculated on readily available data, rather manual corrections are done to bring data into assessable form.

In the local market, problem with data quality is also experienced among risk managers. General view of the problem is that, mathematical judgements in sensitivity analysis are very common. Moreover, some more complex models cannot be employed due to lack of required data in the system. One of the risk managers shared his experience that due to judgements, most times sensitivity analysis result in uncorrelation among variables and this reduces credibility of their work.

Finally, managers also expressed their view about targeted sales programmes. Marketing strategies in the market are almost identical. Managers rely on the data in the system and send promotion messages only to those customers who the system recognises as "targeted" for their campaign. Unfortunately, most banks in the market do not capitalise or deeply analyse marketing expense, therefore, they do not assess real impact of marketing on their sales.

Conclusion

Data quality is not a local issue for banking industry rather the issue is more systematic and gets even harder to tackle with in the growing market. Unintegrated systems, non-standard data inputs, falsification of data by inputter, culture and difficultness in estimating cost of wrong data are main reasons that result in data quality to deteriorate. Main causes of data quality to banks are mostly monetary, but again, estimates make it difficult to calculate exact monetary effect in the local market. Local banking sector need rapid solutions and more investment to improve quality of data.

References

Crosby, P. B. (1979). *Quality is Free: The Art of Making Quality Certain.* McGraw-Hill.

Debbarma, N., Nath, G., & Das, H. (2013). Analysis of Data Quality and Performance Issues in Data Warehousing and Business Intelligence. *International Journal of Computer Applications, 79*(15), 20-26.

Doyle, M. (2015). *Is bad data wasting your marketing efforts*. Retrieved 09 10, 2018, from https://www.b2bmarketing.net/en-gb/resources/blog/bad-data-wasting-your-marketing-efforts

Ecosystems, Insight. (2009). *Data Quality and Integration in Banking*. Retrieved 08 17, 2018, from https://www.insightecosystems.com/wp-content/uploads/2015/01/Data-Quality-And-Integration.pdf

Garg, A., Grande, D., Miranda, G. M.-L., & Sporleder, C. (2017, April). *McKinsey & Company.* Retrieved 08 28, 2018, from https://www.mckinsey.com/industries/financial-services/our-insights/analytics-in-banking-time-to-realize-the-value

Hanmath, M., & Shivaji, W. (2014). *Risk Management in Banks - Regulatory Prespective.*

Haug, A. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management, 4*(2), 168-193.

Hayler, A. (2011). *Poor corporate data quality strategy impedes business progress.* Retrieved 09 03, 2018, from https://www.computerweekly.com/opinion/Poor-corporate-data-quality-strategy-impedes-business-progress

Lynch, R. L., & Cross, K. F. (1992). *Measure Up!: The Essential Guide to Measuring Business Performance.* Mandarin.

MarketingSherpa. (2007). *E-commerce Benchmark Guide 2007.* MarketingSherpa.

Masayna, V., Koronios, A., Gao, J., & Gendron, M. (n.d.). *Data quality and KPIs: A link to be established.*

Michelberger, K. (2016). Corporate governance effect on firm performance; A Literature Review. *Regional Formation and Development Studies, 3*(20), 86-94.

Neri, M. (2012). Data quality in banking : regulatory requirements and best practices. *Journal of risk management in financial institutions, 5*, 146-161.

Oracle. (2017). *Database Data Warehousing Guide*. Retrieved 08 20, 2018, from https://docs.oracle.com/database/121/DWHSG/concept.htm#DWHSG8075

O'Reilly, C. A., Chatman, J., & Caldwell, D. F. (1991). People and organizational culture: A profile comparison approach to assessing person - organization fit. *Academy of Management Journal, 34*(3), 487-516.

Redman, T. C. (2013). Data's Credibility Problem. *Harvard Business Review*.

Redman, T. C. (2016). Bad Data Costs the U.S. $3 Trillion Per Year. *Harvarad Business Review*.

Shahin, A., & Mahbod, M. A. (2007). Prioritization of key performance indicators, An integration of analytical hierarchy process and goal setting. *International Journal of Productivity and Performance Management, 56*(3), 226-240.

Singh, R., & Singh, K. (2010). A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing. *International Journal of Computer Sciences Issues, 3*(2), 41-50.

Statista. (2018). *The Statistics Portal*. Retrieved 09 21, 2018, from https://www.statista.com/statistics/282197/global-marketing-spending/

Steerman, C. (2016). Data Quality – Overcoming the Biggest Hurdle for Compliance and Risk Management. ComplySci.Inc.

World Bank. (2007). *Doing Business 2007: How to perform.* Washington: World Bank.

World Bank. (n.d.). *Operational Risk - Risk Assessment*. Retrieved 09 05, 2018, from http://siteresources.worldbank.org/EXTFINANCIALSECTOR/Resources/282884-1239831335682/6028531-1239831365859/G-Operational_Risk_Assessment-Fulton.pdf

Yerby, J. (2013). Legal and ethical issues of employee monitoring. *Online Journal of Applied Knowledge Management, 1*(2), 44-55.