Data Integration Capabilities and Performance of Commercial Banks in Kenya

JOHN WAMAI*, ROSEMARY JAMES and JOSHUA TUMUTI

Department of Management Science, School of Business, Kenyatta University, Kenya

Abstract
The purpose of this paper is to present an empirical evaluation of the effect of data integration capabilities on the performance of commercial banks in Kenya. An explanatory non-experimental design was applied to conduct a census of the 43 commercial banks in Kenya. A semi-structured questionnaire supported in collection of primary data. Secondary data from commercial banks annual financial results for period 2013 to 2019 was compiled using data collection sheet. SPSS tool was applied for analysis of data. The regression results disclosed that data integration capabilities positively and significantly affect performance of commercial banks. It was recommended that the managers of commercial banks and other stakeholders should continue investing and adopting data integration capabilities as it is an effective performance enhancement strategy. Further, the policy makers and regulators should employ policies that support data integration capabilities adoption to augment performance.

Introduction
Commercial banks constitute essential segment of economy as a key agent of financial intermediation.1 Their primary function of intermediation involves reallocating excess funds between deficit and surplus units of economy to boost economic growth.2,3 They also contribute to the economy by facilitating financial inclusion as well as making funds available for investors to borrow.4 Through intermediation, commercial banks also mobilize and facilitate efficient allocation of national resources hence increases investment quantum and thereby increase in national output.5 According to,6 poor performance of commercial banks raises fears among planners and policy designers due to high possibility of collapsing the national economy. Faced with all the challenges, commercial banks have implemented business intelligence with support of data integration capabilities to support decision making, meet strategy implementation and to improve on performance.7
Commercial banks generate millions of transactional data everyday through multiple channels.\(^8\) Data integration capabilities relies on data patterns for reliable prediction of outcomes. As a result, banks are leveraging on data for decision making, segmentation and tracking of customers’ transactional behaviours with aim of improving performance through service differentiation, customer targeted marketing and maintenance of loyal customer base.\(^9\)

Data integration capabilities are the characteristics and features of business intelligence data aspect that enables data integration into one single source of organization’s information. This is the main factor that guarantees continued advantages of business intelligence to an organization.\(^40\) Accuracy of data integration is a catalyst for boosting performance of organizations which increases firm’s revenue, return on asset and profitability.\(^27\) Data analytics allows analysis of raw data to make conclusion. It complements value for existing data.\(^13\) Data management entails the whole process of acquiring data. In this study, borrowing from previous studies, data integration capability was operationalized as Data analytics (DA), Data Management (DM) and Data quality (DQ).\(^41,42,43\)

**Statement of the Problem**

Besides promoting economic progress and national development, performance of commercial banks is vital in order to reward shareholders for their investment.\(^10\) Despite the increase in investment in data integration capabilities, performance of commercial banks in Kenya has been declining compared to the respective investment.\(^11\) The profitability in terms of return on asset has declined since 2013 from average of 2.3 to 0.76 in 2019.\(^11,12\) Researchers and scholars have argued that investment in data analytics and integration intelligence leads to improved firm performance.\(^13\) However, quantifying and measuring technology innovation contribution to an organization has always been a concern.\(^14\) The heavy investment in data integration and the declining performance of commercial banks in Kenya point to inconsistencies worth investigating.

Consequently, the study undertook to explore how data integration capabilities influences the performance of commercial banks in Kenya. It is significant in that it will inform managers of commercial banks the intrinsic of association of data integration capability and performance. It will guide in the implementation of strategies that result to better financial performance through efficient operations. Further, the policy makers stand to benefit from the outcome of the study.

**Literature Review**

**Theoretical Review**

Two major models supported the study: The Resource based view (RBV) and DeLone and McLean’s Information Systems success model. RBV was originated by Penrose (1959). This view is intra-organizational focused with performance resulting from unique capability of the firm. Its tenets contends that exclusive capabilities and resources of a firm is the basic source of competitive advantage and higher performance.\(^15,16,17\) Resources that accord a firm superior performance enabled through competitive advantage must be non-substitutable, imperfectly imitable, valuable and very rare to provide sustainable competitive advantage that yields better or higher performance.\(^17,18\)

RBV emphasizes that to attain competitive advantage, firms should deploy both intangible and tangible assets in form of organizational capability, human and physical assets.\(^19,20\) Data integration capability is a technological intangible resource that supports business decision making.\(^21,22\) recommend that VRIN resources should be derived through synergy between technological innovations and other organization resources.

DeLone and McLean’s Information Systems success model had its initiation by.\(^23\) They identified system success drivers as, net system benefits, information quality, intention of use, system and service qualities and finally user satisfaction. According to the model, system and information qualities influences satisfaction of user as well as use of an information system. These in turn influences each other whereas both influence individual impact. Individual impact in turn influences organizational impact.

Information and system qualities have been found to be drivers of information systems use leading to individual satisfaction and hence firm performance.\(^24\) The model served in assessing the benefits of data integration capability as an innovation driver of performance of commercial banks.\(^25,26\)
Empirical Literature Review

27 Studied the influence of data quality on performance of firms in Dublin. They analysed 150 firms across major industries. Performance was operationalized into ROI, ROA, and ROE. Data quality, usability, analytics, intelligence and accessibility were the attributes of data integration capability. They assert that data quality, analytics and intelligence influences return on asset (ROA) positively. They concluded that quality data leads to improved performance.

29 In their research, confirmed a positive association of firm performance and data quality. They further assert that data quality improves movement of information and hence improved firm performance. Another study by 30 examined the influence that data has on firms in the manufacturing sector. The survey analysed data on 533 firms collected from Global Manufacturing Research Group (GMRG). The study concluded that quality and accuracy of data positively impacts firm performance through proper planning.

A study was conducted to evaluate whether Big data analytics has any impact on firms’ performance. 31 Data was collected through online survey from 297 respondents. These were experts in areas of business, big data analytics, IT management and business analysis. The study findings confirm that integration of data analytics in the firm’s decision making significantly influences the performance.

32 Concluded that multiple use of data mining and analytics maximizes opportunities, minimizes risk and supports business growth. The study further revealed that knowledge and data management maximizes business opportunities for organizations through integration of relevant information. This leads to improved firm performance.

27 Had a contextual gap as it had a broad scope in the context of all industries. Further, the study 29, 32 posed both methodological and contextual gaps as noted in conceptualization and variables of the data management component. This study extends from other researches by carrying out a census on commercial banks in Kenya - a developing country.

2.3 Conceptualization and measurement of Variables
The regressor variable was performance of commercial banks. The indicator of performance was return on asset, whose data was collected from published financial disclosures.

The independent variable was data integration capabilities with three indicators: data analytics, data management and data quality. Data management guides the process of data acquisition ensuring its availability. Data quality guarantees completeness, reliability and security of the organization data. Data analytics entails mining of data to identify useful patterns for business value. This characteristics constituted data integration capability. It was hypothesized that if commercial banks adopt data integration capability, their return on asset, the performance indicator, would improve.

Hypothesis
From the reviewed literature and subsequent conceptualization of the variables, it has been hypothesized that:

\[ H_0: \text{Data Integration capabilities have no significant effect on performance of commercial banks in Kenya.} \]

Methodology
The research, consequently, was anchored on positivism paradigm. This is because the study attempts to provide solutions to practical problems, develop law-like generalizations and discover causal relationship through statistical analysis. 33 An explanatory nonexperimental research design was adopted. This involves collection of data on quantitative variables to determine their relationship. 34 The objective was to determine the effect of data integration capability on the performance of commercial banks in Kenya. This was therefore a census comprising all commercial banks in the country. Census approach was adopted since commercial banks are relatively few and their published information is readily available. 35 Asserts that census enhances validity of the collected data as it includes more cases which provides extra information.

The respondents were drawn from banks’ management cadre comprising the heads of IT, Operations and Credit within the commercial banks. A questionnaire was shared via email to gather primary data. Secondary data was compiled from
published financial statements and CBK reports. Opinion of experts in data integration and analytics as well as literature review were widely consulted for validity of the instrument.

Descriptive and regression methods were used to analyse data. The following empirical model was used:

$$\text{ROA}_i = \beta_0 + \beta_1 \text{DIC}_i + e$$

Where

$\beta_0 =$ Intercept Constant

$\text{DIC}_i =$ Data Integration Capability of bank $i$

$\beta_i, \beta_0, \beta_1 =$ Regression coefficients of bank $i$

$i =$ Bank 1, 2, 3…40

$e =$ the error term.

$\text{ROA} =$ Return on Asset

The Cronbach’s alpha, which is mostly used, was used for reliability tests in the study.

It is expected that the value of the alpha should be greater that 0.8 for the items of variables to be acceptable as reliable.  

**Results and Discussion**

**The Reliability**

The reliability test results are as indicated in Table 1.

The presence of internal consistency is indicated by a Cronbach’s Alpha value bigger than 0.8 signifying the reliability of scale used. Results in table 1 shows a Cronbach’s Alpha values higher than 0.8 confirming construct reliability. This implies presence of internal consistency indicating that items measures in the study construct belongs to that construct.

**Descriptive Statistics**

The study sought the level of agreement of the respondents on statements regarding data integration capability and performance of commercial banks. A five-point Likert scale was used with the findings as indicated in Table 2.

<table>
<thead>
<tr>
<th>Table 1: Test of Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable(s)</strong></td>
</tr>
<tr>
<td>BI Data integration capability</td>
</tr>
</tbody>
</table>

Source: Research Data (2021)

**Table 2: Data integration capability and performance**

<table>
<thead>
<tr>
<th>Statement</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data analytics has supported segmentation of our customers for targeted engagement</td>
<td>82</td>
<td>4.09</td>
<td>.79</td>
</tr>
<tr>
<td>With analytics, we are able to customize products for our customers</td>
<td>82</td>
<td>4.26</td>
<td>.72</td>
</tr>
<tr>
<td>Using analytics, we can predict customer spending behaviour for cross selling decision making</td>
<td>82</td>
<td>4.50</td>
<td>.53</td>
</tr>
<tr>
<td>Our data governance measure ensures good data is clean for better decision making</td>
<td>82</td>
<td>3.94</td>
<td>.73</td>
</tr>
<tr>
<td>We use data analytics to derive the real identity of our customers to ensure data security reducing loss due to fraud.</td>
<td>82</td>
<td>4.01</td>
<td>.79</td>
</tr>
<tr>
<td>With proper data security, customer data is safe which reduces loss of income due to data loss.</td>
<td>82</td>
<td>4.29</td>
<td>.73</td>
</tr>
<tr>
<td>Data is classified to ensure proper use of data for benefit of the bank.</td>
<td>82</td>
<td>3.89</td>
<td>.90</td>
</tr>
<tr>
<td>We ensure company data is cleansed for proper decision making</td>
<td>82</td>
<td>4.38</td>
<td>.54</td>
</tr>
<tr>
<td>High quality data supports informed decision making.</td>
<td>82</td>
<td>4.51</td>
<td>.53</td>
</tr>
<tr>
<td>Average</td>
<td>82</td>
<td>4.21</td>
<td>.69</td>
</tr>
</tbody>
</table>

Source: Research (2021)
The output in Table 2 revealed that the highest number of respondents, with a mean of 4.09 agreed with the first statement with a very low variation as shown by standard deviation value of 0.79. On the statement that with analytics, the bank is able to customize products for its customers, a majority of customers agreed and strongly agreed with a mean of 4.26. A low deviation affirmed this with standard deviation value of 0.72. Regarding statement whether by using analytics, it is possible to predict customer spending behaviour for cross selling, a majority (mean=4.5), were in concurrence. A low variance value of 0.53 further confirms this findings. With the statement which sought to confirm if the bank’s data governance measure ensures good data is clean for better decision making, a majority of respondents concurred by a mean of 3.94. there was a low deviation of 0.73.

Another statement sought to find whether data analytics is applied to derive the real identity of bank customers to ensure data security reducing loss due to fraud. A majority of managers concurred with a mean of 4.01 and a low standard deviation value of 0.79. Regarding the statement that with proper data security customer data is safe which reduces loss of income due to data loss, the mean was 4.29 while the standard deviation was 0.73. This implied that most managers strongly agreed with the statement. Most managers agreed that data classification guarantees proper use of data to benefit the bank.

This is presented by a mean and standard deviation of 0.9 respectively.

It was also established that respondents strongly agreed with the statement that the bank ensures that its data is cleansed for proper decision making (mean=4.38, SD=1.3) and that high quality data supports informed decision making (mean=4.51, SD=0.53). Table 2 shows that data integration capability had an inclusive average score of 4.2 with a low deviation of 0.69. This implies that most managers strongly supported the view that data integration capability enhances performance. The data integration capabilities that were identified to be adopted to improve performance included data analytics, data management and data quality.

The findings in this section conform with results who asserted that quality data leads to increased financial performance. also confirmed a positive association of firm performance and data quality with the latter improving movement of information and hence improved firm performance. On their part found out that quality and accuracy of data positively impacts firm performance through proper planning. The study collected data for a seven-year period (2013-2019) on the return on assets for the commercial banks. Return on asset (ROA) trend was generated to indicate variation between the years as shown in figure 1.

Figure 1 indicates that the average ROA had a steady decreasing trend from the year 2013 to the year 2019. It can be seen that from an average value of 2.354 in the year 2013, the only insignificant increases were recorded in the years 2015 and 2018 where the commercial banks recorded average ROA of 1.9737 and 0.8333 respectively. For the rest of the period, there has been a notable decrease in ROA to a low value of 0.7667 that was recorded in the year 2019. The findings are consistent with capital authority reports.

![Fig.1: Trend Analysis of Commercial Banks ROA](Source: Published Financial Statement for the period 2013-2019)
Diagnostic Tests results
To test for deviations from normality, Shapiro-Wilk normality test was considered. Its null hypothesis ($H_0$) supposes that the sample follows normal distribution. Table 3 shows the results.

### Table 3: Normality Test Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Shapiro-Wilk Stat</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI Data integration capability</td>
<td>0.992</td>
<td>82</td>
<td>0.905</td>
</tr>
</tbody>
</table>

* This forms the lower bound of the true significance.

a Lilliefors Significance Correction

Source: Research (2021)

The results in Table 3 shows that $p$-value (0.905) was greater than 0.05 with a $P=0.992$. Consequently, the null hypothesis failed to be rejected. This follows the Shapiro Wilk normality test’s null hypothesis that the residuals do not significantly deviate from normal distribution. The conclusion was that residuals were normally distributed.

![Normal P-P Plot of Regression Standardized Residual](image)

Fig. 2: ROA vs Data Integration Capability

Source: Research (2021)

A scatter plot was produced to highlight the kind of linear relationship that exists between data integration capability and performance of commercial banks.

The results in Figure 2 indicated existence of a positive linear association of the independent variable and ROA (financial performance). This therefore implies that financial performance (ROA) increases as independent variables increase.

The Durbin-Watson test was used to test for autocorrelation where $1.5<d<2.5$ is a threshold for autocorrelation interpretation. Table 4 displays the results.

### Table 4: Durbin Watson Autocorrelation Test Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Std. Error</th>
<th>Durbin-Watson Stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.487</td>
<td>1.036</td>
</tr>
</tbody>
</table>

a Predictors: (Constant), Organizational Capability, Infrastructure capability, Data integration capability

b Dependent Variable: Performance

Source: Research Data (2021)
The results in Table 4 show Durbin-Watson statistics of 1.036. According to values below 1 or more than 3 are a definite cause for concern due to autocorrelation. The value of 1.036, therefore, implied that the data could be used to carry out regression analysis.

For heteroscedasticity, Breusch Pagan Godfrey test was used with the null hypothesis that residuals are homoscedastic. Since the prob > Chi2 value was 0.078 which is greater than 0.05, the null hypothesis failed to be rejected at five percent level.

### Hypothesis Testing
To determine the significance of the association of data integration capability with the performance of commercial banks, the study carried out a regression analysis on the variables. The subsequent outputs are presented in Table 5.

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>Std. Error</td>
</tr>
<tr>
<td></td>
<td>Data integration capability</td>
<td>0.643</td>
<td>0.297</td>
</tr>
<tr>
<td></td>
<td>a Dep. Var: Performance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Research (2021)

Table 5 shows that the coefficient of data integration capabilities was 0.643 and the matching p-value as 0.034. Given that 0.034 < 0.05, the null hypothesis was rejected. The indication was that data integration capabilities significantly and positively influences performance of commercial banks in Kenya. This discovery is supported by whose study on improvement of enterprise performance through big data analytics capability found that data analytics capability positively affects organisational performance. The results further confirms the findings that data analytics augments performance.

### Summary and Conclusion
The study had an objective of assessing the effect of data integration capability on performance of commercial banks in Kenya. The empirical results revealed that data integration capabilities bears a positive but significant influence on the performance of commercial banks on Kenya. From the attributes of data integration capability, it was revealed that most managers agreed that data integration capability influences the performance of commercial banks. Consequently, it is concluded that data integration capability leads to enhanced performance of commercial banks.

### Recommendations of the Study
The study recommends that commercial bank managers should continue adopting data integration techniques through data analytics, data management and quality. This stems from the fact that data integration capability positively augments performance. The management must therefore invest and promote use of data integration capabilities to enhance their financial performance. Further, the government should spearhead the laws and policies that encourage and enable use of data integration capabilities across organizations. This will promote a culture of data-based decision making.

### Suggestions for Further Research
The study centered on the effect data integration capability has on performance of commercial banks. A further research involving other financial institutions can be carried out in future to examine how data integration capabilities impacts on their performance.

### Acknowledgement
I convey my earnest gratitude to my supervisors, Dr. Rosemary James and Dr. Joshua Tumuti for their insurmountable support and encouragement during the formulation of this study.
Funding
This study was self-funded.

Conflict of Interest
There is no conflict of interest.

References


