UNIVERSITY OF NAIROBI

DEPARTMENT OF COMPUTER SCIENCE

FACULTY OF SCIENCE AND TECHNOLOGY

ANALYTICS AND AUTOMATION OF IFRS 9 MODELING AND REPORTING: A CASE STUDY OF KENYAN BANKS

BY

IBRAHIM STAMILI SAMI

A REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN COMPUTATIONAL INTELLIGENCE OF THE UNIVERSITY OF NAIROBI

JUNE, 2022

DECLARATION

This thesis report is my original work and has not been submitted for examination at any university for any award.

Signed   Date   24th June 2022

Ibrahim Stamili

P52/37735/2021

This thesis report has been submitted for examination with my approval as the university supervisor.

Signed   Date 24th June 2022

Prof. Elisha Opiyo

University of Nairobi

DEDICATION

I dedicate this work to my mother, Nancy Cheptoo. Thank you for everything! Shine on mom.

ACKNOWLEDGEMENT

I would like to acknowledge my mentor Prof. Elisha Opiyo for his input and guidance.

ABSTRACT

Becoming IFRS 9 continues to prove a challenge for Kenyan banks because of a lack of mature credit scoring system which most models rely on. A 2018 report by ICPAK report shows that 90% of banks rely on enterprise spreadsheet models built on MS Excel for their financial modeling and reporting functions. The primary disadvantages of building a financial model on excel is lack of error control, lack of reusability, little to no automation of common tasks, poor integration with existing data sources, limited scalability and poor maintainability. An error-prone output potentially results in the misrepresentation of a bank's financial position.

Existing automated implementations, such as the one proposed by Volarevic and Varovic (2018), rely heavily on credit scoring for Probability of Default estimations and are not suitable for local markets. ERP providers also offer automated IFRS 9 modeling solutions. However, these are vendor-specific and have high implementation costs. For instance, Surecomp a financial services solutions provider, offers a cloud-based IFRS 9 solution called IMEX at $300,000 as a flat-rate, one-time payment (Capterra, 2022).

This study proposes an automated IFRS 9 model built on Alteryx that uses multi-state Markov (MSM) probability analysis to estimate the Probability of Default (PD). The probability modeling approach used relies only on historical loan information making it suitable in the local context. It also uses a low-code development platform to ensure ease of development, use and maintenance while addressing the pain points of spreadsheet modeling. The solution integrates with an existing database instance for automated data input. It also integrates with a reporting and visualization platform that summarizes the key drivers of Expected Credit Loss and inform management decisions or overlays.

The model was built for one of the banks using anonymized loan data. For testing, results from the model of the model were compared to the output of the reported financials. An allowable variance of 2% was applied as advised by the bank. The results observed were congruent with the reported financials. The highest Expected Credit Loss variance was observed in the Credit Card sector at 0.45%. An evaluation of the packaged solution (Alteryx app) was then done by administering the

System Usability Measurement Inventory (SUMI) questionnaire to potential users. Generally, the system received positive reviews on the following metrics – Efficiency, Learnability, Affect and Control. The system was also evaluated on improving efficiency of the current approach.

TABLE OF CONTENTS

ix

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

IAS	International Accounting Standards
IASB	International Accounting Standards Board
IFRS	International Financial Reporting Standards
ECL	Expected Credit Loss
PD	Probability of Default
LGD	Loss Given Default
EAD	Exposure at Default
FLI	Forward Looking Information
OMV	Open Market Value
FSV	Forced Market Value
MSM	Multi-state Markov

CHAPTER 1: INTRODUCTION

1.1 Background to the study

The 2008 financial crisis was one of the biggest jolts to the global economy in recent history. It pushed some of the world's biggest financial systems to the brink of collapse. (Carmassi, Gros, & Micossi, 2009). At the heart of the crisis was a speculative lending process and an "after-the-fact" provision practice by banks. The then governing standard on loan impairment was the International Accounting Standard 39 (IAS 39). Impairment of a loan refers to the reduction of the amount the lender can recover from the loan due to its deterioration (Krayn, Oleksza, Canamero, & Gea-Carrasco, 2015). Under IAS 39, banks were required to provide for impaired loans only after these loans went into default. This type of provision meant that in the event of a crisis similar to the one in 2008, banks did not have the cushion from provisions and government bail outs were necessary.

In response, the International Accounting Standards Board (IASB), introduced the International Financial Reporting Standard 9 (IFRS 9) to replace the IAS 39 on 12th November 2009. The new standard adopts a forward-looking, probabilistic model for predicting impairment of loans. (Krayn, Oleksza, Canamero, & Gea-Carrasco, 2015). IFRS 9 proposes a forward-looking approach of predicting an asset's Expected Credit Loss (ECL).

In the local context, adoption of IFRS 9 has proven a challenge for Kenyan banks. A report by the Institute of Certified Public Accountants of Kenya (ICPAK) shows that systems and automation is the primary challenge of IFRS 9 adoption leading up to the 2018 deadline (ICPAK, 2018). In the survey, 73% of top management in financial institutions did not believe they had the infrastructure and skills needed to make a seamless transition (ICPAK, 2018). At the time of its publication, the ICPAK report also notes that 90% of banks rely on enterprise spreadsheet models built on MS Excel for their modeling and reporting functions. (ICPAK, 2018).

Building financial model on spreadsheets is time consuming because of multiple iterations that have to be made in order to get accurate output. Poor integration with existing data sources such as core banking systems means that data requests have to be made, approved, and the needed data sets extracted for each run (Ho & Sang, 2004). Lack of error tracking and error control in spreadsheet models also stretches out the modeling process as the bank iterates over erroneous outputs. Spreadsheet models also break easily because they require a lot of human input due to

lack of automation of common tasks. A change in a cell or column can break the logic of the model leading to errors and further iterations over the output (Ho & Sang, 2004). Therefore, spreadsheet models are not suitable for building flexible models especially in the case of IFRS 9 which is a standard that is continuously updated (ICPAK, 2018). Another challenge of spreadsheet modeling is manual error checking and validation of data inputs after data extraction. Erroneous data inputs have to be re-extracted and re-inserted into the model after manual data quality checks are done (Ho & Sang, 2004). The result is a complex model, that is laden with errors and frequently breaks when making updates. These compounding factors make the whole reporting process time-consuming.

1.2 Problem statement

The analysis and reporting process of becoming IFRS 9 compliant is time-consuming. Banks iterate over the credit risk modeling process because of:

1. Data requests taking too long to be fulfilled due to poor integration with existing data sources.
2. Errors in the data input.
3. Lack of automation of common tasks.
4. Errors in intermediary outputs.

These challenges point to the need for an automated financial modeling solution that addresses the challenges of enterprise spreadsheet models and also meets the reporting requirements of IFRS 9 for banks in Kenya.

1.3 Overall goal and objectives

1.3.1 Overall goal
To develop an automated IFRS 9 modeling and reporting solution for banks in Kenya.

1.3.2 Specific objectives
i. To **analyze** the bottlenecks in the IFRS 9 modeling and reporting process for banks in Kenya.
ii. To **design** an automated IFRS 9 modeling and reporting solution suitable for banks in Kenya.
iii. To **develop** an automated IFRS 9 modeling and reporting solution suitable for banks in Kenya.
iv. To **test** and **evaluate** the developed solution.

1.4 Significance

The project highlights the challenges that Kenyan banks face in meeting compliance levels of international financial reporting standards. It analyzes the bottlenecks in the current approach used by banks in Kenya. It also demonstrates the potential for automation of financial modeling and reporting in Kenyan banks. Automation will benefit practitioners within banks by helping them reduce reporting time and meet audit and regulatory requirements. Additionally, the automation approach proposed in the study can be extended to other financial models and standards. These models are useful especially for credit risk teams, internal audit teams and external audit teams when verifying the veracity of the financial statements reported by banks in Kenya.

1.5 Scope

To analyze, design, develop, test and evaluate an end-to-end automation solution for IFRS 9 modeling. The sample population selected for this study is limited to listed banks in Kenya reporting in the December 2020 period.

1.6 Assumptions and limitations

A primary limitation of the study is that it only focuses on IFRS 9 modeling and reporting. Financial institutions are governed by several IAS and IFRS standards depending on the nature of the business and region of operation. Another limitation is the inability to generalize to other types of financial institutions affected by IFRS 9 adoption such as SACCOs and Insurance Companies. The nature and requirements of financial reporting for these are different.

However, the general automation approach proposed by the study could be extended for other financial models and other financial institutions. The goal is to show the potential for automation and propose a more streamlined modeling approach that will help banks in their efforts to become IFRS 9 compliant

1.7 Definition of important terms

i. **International Accounting and Standards Board (IASB)** – An independent organization responsible for developing financial accounting and reporting standards that bring accountability, transparency and efficiency to global financial markets.

ii. **International Accounting Standards (IAS)** – A predefined suite of accounting standards defined by IASB used to describe an entity's financial performance in an interpretable and comparable manner across financial markets.

iii. **International Financial Reporting Standard (IFRS)** – Similar to IAS, but updated to reflect current financial markets and drivers.

iv. **IFRS 9** – A reporting standard that outlines how an entity should classify and measure financial instruments.

v. **Low-code environment** – A platform that is used to build application software through graphical user interface navigation instead of traditional coding approaches.

vi. **Alteryx** – A proprietary low-code platform that allows teams to build self-service analytics solutions using a workflow approach.

vii. **KNIME** – An open-source low-code platform that allows teams to build self-service analytics solutions using a workflow approach.

viii. **Data Warehouse** – A store of data aggregated from different sources and organized to optimize business needs.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter discusses the parameters and data required for IFRS 9 modeling. It also investigates the impact of the new standard and challenges facing traditional modeling and reporting approaches. It then looks into automation and the emerging trends within the field of automation such as democratization of technology, workflow automation and low-code/no-code automation. It also reviews existing IFRS 9 solutions and summarizes the gap therein.

2.2 Parameters of IFRS 9

The primary goal of IFRS 9 is to replace the "after-the-fact" provision of IAS 39 with forward-looking, probabilistic provisions. An output called Expected Credit Loss (ECL) is computed for each loan in a bank's book (Krayn, Oleksza, Canamero, & Gea-Carrasco, 2015). The ECL of a loan facility is determined by 3 parameters:

i. Probability of Default (PD) – the likelihood of impairment to default (i.e., the credit risk). The main parameter to be modeled. It provides the forward-looking aspect of IFRS 9.
ii. Loss Given Default (LGD) – upon impairment, what percentage of the asset can be recovered given collateral and historical collection information.
iii. Exposure At Default (EAD) – the outstanding amount at time of impairment.

ECL is then calculated as **PD * LGD * EAD** (Krayn, Oleksza, Canamero, & Gea-Carrasco, 2015). The next sections discuss the three parameters in detail and the approaches to calculating each within IFRS 9 guidelines.

2.3 Probability of Default (PD)

The probability of default is the main parameter that introduces the forward-looking aspect when computing Expected Credit Loss. It is a measure of the likelihood of a facility going into default within a given time period. This section compares 3 general PD modeling approaches – credit rating approach, machine learning approach and statistical approach – detailing the merits and demerits of each within the context of IFRS 9 and within the context of Kenyan banks.

2.3.1 Credit rating approach

In mature markets, PD is a factor of a borrower's credit score. Recognized credit rating and scoring agencies assign scores to borrowers. Fair, Isaac and Company (FICO) is a US based credit rating company focusing on consumer credit risk (Allen, Delong, & Saunders, 2003). FICO ratings are a fixture of consumer lending in the US. A customer's rating is congruent with their PD. Standard and Poor's (S&P) is a global rating body for bond ratings and is used to estimate the PD for bond facilities. An S&P rating of AAA means that the borrowing entity has an "extremely strong" capacity to repay. A rating of CCC means that the borrowing entity is "currently highly vulnerable" with an "extremely weak capacity to repay" (S&P, 2017). Other rating agencies focus on corporations and governments such as Fitch and Moody's (Allen, Delong, & Saunders, 2003).

The drawback to this method is that it is not suitable for the local context in estimating PD because of the lack of credit risk information especially for personal consumer loans. The situation is getting better locally. According to a 2016 survey by FinAccess, an organization that champions for financial inclusion, only 23% of the Kenyan population were unbanked at the time of the survey. 90% of the population have access to mobile money (FinAccess, 2016). There is need for a robust local credit rating approach that utilizes mobile money information to fill in the gaps in consumer loan sector and for local SMEs.

2.3.2 Machine learning approach

PD can be estimated using common machine learning approaches such as neural networks, regression models and classification tress. The predictors in this case are attributes about the borrower such as the sector to which they belong, their history of default if any, and current financial position. This approach leverages data at the customer level to make predictions of default. In a demonstration of ML prediction for estimating PD, Clinton Chee (2021) proposes a linear regression model for PD estimation using 170 features as predictors.

A major challenge of using machine learning for PD prediction is the lack of explainability. Explainability is important for IFRS 9 and any other financial model because the bank needs to justify its approach and outputs during audit (Chee, 2021). Machine learning approaches for estimating credit risk and PD are be useful in other contexts such as informing lending practices and loan interest rates but less so in IFRS 9 because of poor explainability and reproducibility during audit (Chee, 2021).

2.3.3 Statistical approach

This is the most widely used approach in modeling Probability of Default because models in this approach are mathematically sound, derived from data, and are easily explainable and reproducible (Chee, 2021). As such they have been the preferred modeling approaches used by banks in financial reporting.

However, each statistical approach is suited for a different case. For instance, the Merton Model - which is an analysis model used to gauge whether or not a company can service its debts - only works for publicly listed companies (Tudela & Young, 2003). Merton (1974) proposed a quantitative credit risk model that infers the credit risk of a borrowing company based on its market valuation, current stock pool, overall company liabilities and overall company assets (Tudela & Young, 2003). The Merton model is highly effective for publicly listed companies. In a test implementation, Tudela & Young (2003), noted that the it was 87.3% accurate for modeling PD over a one-year period. The main challenge of the Merton model and its variations is that it is only suitable for publicly traded companies and does not cover other sectors such as the consumer loan market in Kenya. Consequently, a bank would need an additional PD modeling approach in place.

Another statistical model is the Cox proportional-hazards model. It has its origins in medical research for investigating the relationship between identified predictors and the chance of patient survival. The basis of the Cox proportional hazards model is to investigate the likelihood of an unwanted event happening at a particular point in time. In the medical context this is death or infection. In the financial modeling context for IFRS 9, this is default (Lin & Wei, 1989). The estimated rate is defined as the **hazard rate** and is determined by the **hazard function $h(t)$**. In the context of IFRS 9 the hazard rate is the **default rate**, $h(t)$ is defined as the risk of defaulting at time t:

$$h(t) = h0(t) \times exp\ (b1x1 + b2x2 + ... + bnxn)$$

where:

- t – is the point in time survival determined by how far a loan is in its lifetime.
- $x1...xn$ – the set of covariates (i.e., predictor factors) that impact the hazard function
- $b1...bn$ – corresponding weights assigned to the covariates.
- $h0(t)$ – baseline hazard rate. That is, a standard default rate given that all covariates $x1...xn$ are zero. It is an adjustment factor to ensure that no facility is assigned zero PD (Lin & Wei, 1989).

The primary drawback of using a predictor based statistical approach such as the Cox proportional-hazards model is that it has arbitrary choices when it comes to the covariates and the weight that each covariate carries. Hazard rate predictive models are statistical regression models (Lin & Wei, 1989). This ties in to poor explainability and reproducibility as in the machine learning model.

The multistate Markov (MSM) model is a statistical approach that relies only on prior default history. In this approach, borrowers are grouped into rating grades called stages – usually stage 1, 2, 3, and EXIT per sector. The stages are defined by the days past due of the loan. (Meira-Machado, Una-Alvarez, Carso-Suarez, & Andersen, 2009). The PD is the likelihood of a borrower going into the default stage (Chee, 2021). Defaults are determined by transition matrices computed at sector level. MSM models have two key components – states and transitions between the states. States can either be transient or absorbing (Meira-Machado, Una-Alvarez, Carso-Suarez, & Andersen, 2009). A facility can transition in and out of transient states. Example – a loan can move from stage 1 to stage 2 and back to stage 1. Absorbing states are final. In the context of PD modeling this refers to an EXIT state where the facility is fully paid off or written off. A multi-state Markov model captures movements between n states where the probability of moving away from the current state depends on the previous state (Meira-Machado, Una-Alvarez, Carso-Suarez, & Andersen, 2009). The MSM model is preferred because it does not depend on any predictors, covariates or weights to determine PD. Its input is a facility's current state and it relies solely on history of default for the facility's sector to determine it next state. This makes the model easily explainable, reproducible and defendable during audit.

2.4 Loss Given Default (LGD)

The LGD parameter is the percentage of the loan disbursed that a lender will lose on default. In practice, there are two broad approaches to determining the LGD of a facility. The first is collateral LGD that is applied for facilities backed by collateral. Under collateral LGD, the valuation of the collateral is done at the time of facility origination. The amount from this initial valuation is the collateral's Open Market Value (OMV) – that is what a seller would gain from selling the collateral under normal market conditions (Spall, 2018). However, in most cases, the lender auctions off collateral and sells it below the OMV. Therefore, a percentage is subtracted from the OMV to account for this. This value is called the haircut value and is dependent on the type of collateral. The resultant value after subtracting the haircut is called the Forced Sale Value (FSV).

Additionally, a recovery cost is then deducted from the FSV. This deduction is meant to cover typical recovery costs such as advertising of auctions and the amount needed to fund the auctioneering or recovery process (Spall, 2018). It also depends on the type of collateral. The final FSV amount is then discounted to the present time value of money. The collateral LGD is then determined from the discounted FSV amount and the outstanding balance of the facility.

For non-collateralized loans, the LGD is calculated using data about historical collections from previously defaulted facilities. This forms what is called the collections LGD. It is computed at sector level. For instance, to determine the collections LGD in the agriculture sector, we look at previously defaulted facilities in the agriculture sector at least 3 years back. What was the total outstanding balance that was lost on default? What was the total amount collected from the defaulted facilities? The collections LGD is the ration of the amount collected in a sector to the total amount defaulted in that sector – for a given time period. (Spall, 2018).

2.5 Exposure at Default (EAD)

The final parameter is the EAD which is defined as the amount that the lender stands to lose in case of default. It is a present-valued amount of the outstanding balance at the time of reporting. The ECL is then calculated by multiplying the three parameters – EAD, PD and LGD (Spall, 2018).

It is important to note that certain nuances and details of model design decisions are open to interpretation based on the lender's business and mode of operations. However, all the model design decisions adopted should be justifiable and defendable to an external audit.

Typical model design decision left open-ended by the standard include:

i.   Definition of time to default.
ii.   Segmentation rationale and the number of segments to use.
iii.   The amount of historical data that feeds into PD and collections LGD computation. At least 3 years.
iv.   Staging criteria for facilities.

2.6 Impact and challenges of IFRS 9 modeling

A 2004 study carried out by the International Federation of Accountants (IFAC) investigated the impact of international reporting standards on the reporting process and the challenges arising. The

study revealed that the complexity of these standards negatively affected their adoption rates. Stakeholders found the data requirements to be demanding and stretching to their current technology infrastructures (Wong, 2004). Participants also cited the constantly changing nature of IFRS standards as a challenge that adds to the complexities of staying compliant. They called for a balance between continuous improvement of the standards and the window within which organizations are required to become compliant (Wong, 2004).

Respondents in the IFAC study that represented small and medium sized enterprises (SMEs) noted an imbalance between the cost of adoption of international standards and the benefits. They highlighted a lack of representation of SMEs when developing international standards (Wong, 2004). SME representatives in the study mentioned that their issues are largely ignored when developing these standards. This underrepresentation coupled with the cost of adoption makes it unattractive for SMEs to embrace and implement these standards (Wong, 2004).

Other challenges highlighted include the technological and data infrastructure demands of IFRS. Companies, especially SMEs, cited this as one of the primary reasons for late adoption. Only 11% of the respondents that represented SMEs had fully adopted IFRS and were using international standards for internal management reporting. 94% had not identified the missing data points and put in place processes to ensure that data within their systems is ready for IFRS implementation (Wong, 2004). The respondents also highlighted a human resource skill gap. 55% of respondents noted that their current in-house human resource capacity could not accommodate the demands of reporting under the new IFRS standards (Wong, 2004).

From the study, the challenges of adopting IFRS include:

   i. The complexity of the standards
   ii. Updates to the standards further increase their complexity
   iii. Technological challenges on data and infrastructure
   iv. Human resource skill gap

In a separate study, (Amoako, 2010) investigated the level of IFRS 7 compliance for banks in Ghana. In his research he looks into some of the challenges banks in Ghana face when automating IFRS 7 modeling and reporting. The results show that 80% of the banks that participated in the research were yet to fully automate their IFRS 7 reporting process. They cited lack of automation

as a major bottleneck in the process of becoming compliant (Amoako, 2010). The study also showed that the cost of automation was the major challenge especially for the smaller banks. For larger banks with international backing, complexity of existing solutions and lack of skilled human resource were identified as the key challenges in automation. (Amoako, 2010). The banks also cited the dynamic and flexible nature of IFRS 7 as a blocker. A reporting standard with changing requirements demands an equally flexible and adaptable implementation solution. However, all the banks involved in the research acknowledged the permanent nature of IFRS and the need to adopt automation in their financial reporting process (Amoako, 2010).

Additionally, a 2016 report by Deloitte UK showed that adoption of an automated financial modeling and reporting approach reduces audit-material errors by 87%. The report also showed that automation of financial modeling reduces report preparation time by 50% in year one and 25% in subsequent reporting periods. On average, the return-on-investment for automation in financial modeling is realized in year 5 when the efficiencies and cost savings match up to the initial investment (Deloitte Touché Tohmatsu Limited, 2016). Automation of financial modeling has four immediate benefits. Reduction of common errors that stem from manual input and analysis. Automation improves the turnaround time for the financial reporting preparation. It also frees up human resources and allows them to focus on value-adding activities such as drawing actionable insights from the output (Amoako, 2010). In the case of IFRS 9, the human resources shift focus from calculating and validating ECL output to understanding the key drivers for ECL. Such insights will help drive business decisions with regards to lending practices and product structure.

In Kenya, a study by ICPAK indicated that 90% of financial models used in Kenyan banks are implemented on MS Excel. Modeling using spreadsheet models proves complex, difficult to maintain and difficult to scale. Outputs are error ridden and often difficult to interpret and defend to auditors (ICPAK, 2018). A joint report by PricewaterhouseCoopers and KPMG indicates that more than 90% of enterprise spreadsheet contain errors that result in material inconsistencies during audit (KPMG Kenya, 2016).

The challenges and opportunities validate the need for a solution that allows reporting entities to quickly develop and deploy automated financial modeling solutions. The solution should abstract the standard's complexities and be flexible enough to keep up with continuous updates to the standard. Additionally, it should not demand a large amount of upfront investment in terms of

technology and human resource input. In doing so it would address the bottlenecks of the current modeling approaches.

2.7 Automation as an enabler for business processes

IBM defines automation as a collective term for technology applications where human input is minimized throughout the process of accomplishing a task. Common automation scenarios include Business Process Automation (BPA), domotics (home automation) and IT automation (IBM Corporation, 2021). Automation is classified into 4 broad categories based on the level of complexity – Basic automation, Process automation, Integration automation and Artificial Intelligence (AI) automation.

Basic automation is at the lowest and least complex level. It involves automating simple, rudimentary tasks. At this level the most mundane tasks are centralized and automated using a shared system. An example use case is automating data entry using Robotic Process Automation (RPA) (IBM Corporation, 2021). Process automation is a more complex type of automation that manages end-to-end business process with significant increases in productivity and efficiency across the board. Examples of process automation include process mining and workflow automation (IBM Corporation, 2021). Under integration automation, machines perform tasks by following a set of predefined rules. Integration automation involves the deployment of digital workers – an application software bot trained on a specific set of skills (IBM Corporation, 2021). A common use case is in customer service bots. 75% of respondents to an IBM survey indicated that customer service is the business area where they recorded the highest return-on-investment from integration automation (IBM Corporation, 2021). The final type of automation involves inclusion of AI in the automation process. Unlike integration automation, the agents deployed do not simply follow a set of predefined rules to accomplish a task. Instead, they learn from and make informed decisions based on past experiences. Increases in accuracy and efficiency are realized 85% of the times when AI automation is deployed (IBM Corporation, 2021).

A study on the emerging technologies of automation by Lambert and Marshall (2018) highlights two key areas of focus for the next 5 years – AI automation and workflow automation. AI automation integrates robotics and AI to build autonomous machines with the ability to interact and learn from their immediate environment. It features both hardware and software automation with a wide range of applications. Workflow automation is a branch of process automation where

logic-based applications perform tasks in the business process with limited human input. According to Lambert and Marshall (2018), workflow automation is moving towards low-code and no-code workflows for automating business processes. Mixson (2021) attributes the move to low-code/no-code workflows as a response to the demand shortage of skills. For instance, the demand for skills in analytics is expected to rise by 28% over the next 7 years into 2028. To address this increasing demand, 75% of large enterprises are actively investing in low-code/no-code application tools by 2024 (Vincent, et al., 2021). Low-code/no-code development environments enable organizations to quickly build in-house capacity and develop value-adding automation software. The next section explores the concepts of low-code/no-code automation as the approach proposed for this study.

2.7.1 Low-code/no-code automation

Low-code/no-code automation is part of a shift in thinking within computer science that seeks to democratize technology. In its 2020 report on the top 10 strategic technological trends, Gartner ranks democratization as the third most important technology that will drive significant disruption while creating opportunities within the next 10 years. Democratization of technology is a concept that champions for the enabling of any individual to create technological solutions without the need for extensive or expensive training (Grand Review Research, 2021).

The market demand for technologies that promote democratization has increased exponentially in recent times. A study by Grand View Research forecasts a 40.6% Compounded Annual Growth Rate (CAGR) for democratization technologies in the next 5 years (Grand Review Research, 2021). The report also estimates that by 2028 the global market for such technologies will hit a high of $13.74 billion (Grand Review Research, 2021). This optimistic lookout indicates a shift from traditional, code-based methods of developing automation software to low-code/no-code development environments.

A report by Gartner estimates that low-code/no-code application development platforms will account for 65% of all development activity within large enterprises. The report also projects a 165% growth rate for such development platforms within the next two years. Such projections are backed up by the rise of start-ups geared towards developing these low-code/no-code platforms. One such platform, Unqork (a New York based startup), was used by various local governments in the US to develop contact tracing applications in the wake of the COVID-19 pandemic (Marcus,

2020). Apart from startups, tech giants have also invested heavily into such platforms. In November 2019, Microsoft launched its workflow automation software Power Automate. In January 2020, AppSheet, a no-code application development platform was acquired by Google for $2.7 billion (Marcus, 2020).

2.7.2 KNIME and Alteryx for low-code workflow automation

The benefits of low-code/no-code automation development include increased turnaround time, reduction of training costs and democratization of technology. This study proposes the use of two such development platforms for workflow automation – Alteryx (proprietary) and KNIME (open source).

Konstanz Information Miner (KNIME) is an open-source low-code platform that is used to build workflows that integrate data analytics and reporting. It is a very versatile tool with a wide range of applications such as data analytics, machine learning and data engineering. It provides a comprehensive set of free tools for building workflows using a drag-and-drop functionality. The KNIME platform can be grouped into two offerings. The first is the KNIME Analytics Platform which is free and open source. For scalability, organizations opt for the second offering, the KNIME Server which is offered in three tiers. The KNIME Server Small allows small and medium sized teams to collaborate and execute workflows remotely. The KNIME Server Medium provides in-browser access to workflows, collaboration, remote execution, and REST API access. The final offering is the KNIME Server Large which is tailored for large, global teams. It includes distributed execution of large workflows across interconnected networks. It also provides connectors to big data platforms such as Hadoop, Apache Spark and Databricks (KNIME Analytics Platform, 2021).
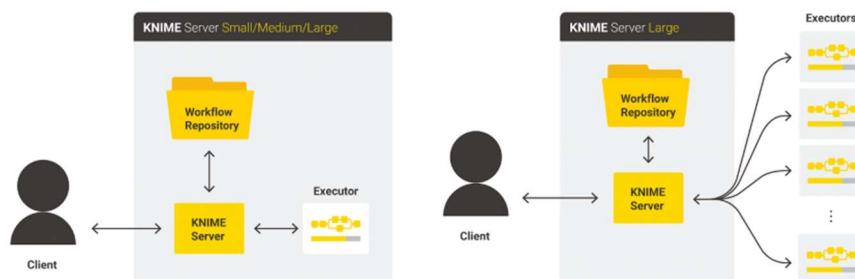

Figure 1 KNIME Server Scaling

KNIME has been used by Sport England, a non-departmental public body, to build automated workflows for Twitter sentiment analysis with PowerBI integration. Rabobank, a Dutch multinational bank, deployed KNIME workflows as part of an anti-money laundering solution that automated the process of checking and flagging transaction anomalies (KNIME Analytics Platform, 2021). KNIME also boasts an ever-growing community with support from open-source developers across the world. It is a mature workflow automation platform that should be considered for any large-scale automation project.

Similar to KNIME, Alteryx allows users to build a wide range of software applications through an intuitive drag-and-drop user interface. It provides the user with self-service analytics tools that perform a predefined set of functions. Custom functions can be developed using either the Alteryx Formula Expression Tool or by using integrative tools such as the R or Python tools. Like KNIME, Alteryx can be used to develop workflows in data science, analytics, reporting and ETL processes. The primary solutions offered by Alteryx are the Alteryx Designer and the Alteryx Server. The Alteryx Designer is the desktop application that is priced at around $5,000 for a single annual user license. Alteryx also provides a custom, subsidized pricing option for large organizations who wish to purchase multiple user licenses. It provides over 70 different connectors to data sources and over 250 data manipulation tools. It also offers over 100 reporting tools that are easily integrated with common business intelligence platforms such as PowerBI and QlikView. (Alteryx Analytics Platform, 2021). The Alteryx Server is recommended for large enterprises for scalability and improved collaboration and sharing of workflows. Other key features of Alteryx Server include remote workflow scheduling and distributed execution of large workflows across multiple nodes (Alteryx Analytics Platform, 2021). There is no predefined pricing for an Alteryx Server installation and organizations need to contact Alteryx for a quote depending on their deployment requirements. Large enterprises that have deployed Alteryx at scale report signification returns in their automation efforts. 7-Eleven, a convenience service store, reported a 60% efficiency increase in their tax reporting processes. Walmart uses Alteryx to automate their email blasts across the organization reaching over 1.2 million associates. Coca-Cola has used Alteryx to automate their reporting process for over 600 restaurants (Alteryx Analytics Platform, 2021).

The choice between the two platforms boils down to developer preference. A 2021 survey by Gartner on data science and machine learning platforms, classifies Alteryx as a challenger and

KNIME as a visionary (Gartner Inc, 2021). The survey rates the platforms on two parameters – ability to execute and completeness of vision. As a challenger, Alteryx has a very high ability to execute and a medium-to-low completeness of vision. As a visionary, KNIME has a more complete and robust long-term strategy but a limited current ability to execute.

Figure 2 Gartner 2021 Magic Quadrant for Data Science and Machine Learning Platforms

2.8 Existing IFRS 9 modeling solutions and gap

Here we look into some of the existing solutions for IFRS 9 modeling and reporting. The approaches can be classified as either internal or out-of-the-box. Internal approaches are those that are developed by the reporting entity. The reporting entity is in charge and in control of every design decision that goes into the modeling. The advantage of this approach is that the model is tailored to the business operations of the reporting entity. Additionally, understanding of the modeling nuances means that the developed model is flexible enough to handle changes and updates to the standard. The internal model approach is also less costly. However, it requires significantly higher initial investment in terms of time and human resource to develop as compared to adopting an out-of-the-box solution.

Internal solutions are mostly implemented on MS-Excel or other spreadsheet platforms. A survey by KPMG UK showed that 85% of financial institutions rely on enterprise financial models for their financial model. The diffusion of internal spreadsheet models throughout enterprise has made them assets and risk factors (Powell, Baker, & Lawson, 2008). The primary disadvantage of enterprise spreadsheet modeling is the susceptibility to errors. In 1998, a study by Panko and Sprague was used to determine the frequency and types of errors in spreadsheet models. The study required 152 undergraduate and MBA students to develop a spreadsheet model from a simple word problem. 35% of the models developed were incorrect, 24% contained significant material errors and thus did not address the problem. Code inspection only revealed errors in 15% of the incorrect models (Panko & Sprague, 1998). When applied to large complex problems such as reporting under IFRS 9, spreadsheet models feature more errors that remain undetected but have a material effect on the output.

Volarevic and Varovic (2018) proposed a simplified spreadsheet model for implementing IFRS 9. The goal of the proposed model was to automate the calculation of two of the three parameters needed for ECL computation – LGD and EAD. Data inputs were stored and updated through multiple MS Excel spreadsheets. For determining the PD, the model relied on external credit scoring approaches from reliable rating agencies. The PD is informed by a combination of ratings from the big three agencies – Moody's, Standard & Poor's (S&P) and Fitch (Volarevic & Varovic, 2018). Even though ratings from these agencies are recognized and easily explainable across markets, the factors used to obtain these ratings may not translate directly to the local context. Additionally, the ratings do not cover all entities within the local market. The internal model proposed by Volarevic and Varovic (2018) was designed for banks in Croatia and works well for markets in developed countries with mature credit scoring systems in place. Additionally, the proposed approach inherits the pitfalls of enterprise spreadsheet modeling such as lack of scalability, flexibility, maintainability and being prone to errors (Deloitte Touche Tohmatsu Limited, 2016).

Out-of-the-box solutions are offered by audit, consulting and technology firms to reporting entities. An example is Deloitte's Finevare, KPMG's Clara and SAP/R3 from SAP technologies (Volarevic & Varovic, 2018). For such solutions, reporting entities do not build any in-house capacity but instead outsource the automation process to third parties. This approach is more

expensive but requires little time and human resource investment from the reporting entity. One of its drawbacks is the inflexibility to standard changes. A lack of model transparency may also prove troublesome for the reporting entity when explaining and defending the model design to auditors (Volarevic & Varovic, 2018).

ERP providers such as Oracle and SAP offer their own modeling solutions for IFRS 9. The main incentive behind these solutions is that they leverage the existing infrastructure. An enterprise that has invested heavily on Oracle infrastructure benefits greatly by outsourcing their modeling and reporting to Oracle's IFRS 9 solutions. Oracle promises a unified data-model and a common results repository for automated reporting (Oracle Financial Services, 2016). However, adoption of these solutions may result in vendor-lock in. In economics, vendor lock-in is defined as the situation in which a customer is heavily dependent on products from one vendor and is guaranteed to incur significant costs for switching. The switching costs often do not make business sense and the customer perpetually sticks to that vendor (Opara-Martins, Sahandi, & Tian, 2016).

2.9 Gap

The proposed internal solutions still use the enterprise spreadsheet modeling approach. The level of automation is low and the disadvantages of spreadsheet modeling for enterprises arise. Additionally, the solution proposed by Volarevic and Varovic (2018) for PD modeling requires a well-established credit rating system that is not yet available in the local market. Out-of-the box solution are tied to the infrastructure deployed and demand a high initial investment. An Oracle solution requires and Oracle data and IT infrastructure. Out-of-the-box models are also black box models with challenges such as inflexibility and poor explainability.

2.10 Summary of literature review

Becoming IFRS compliant such as complexity, model changes, upskilling and technological challenges. These challenges call for financial model solution that abstracts the complexities, maintains flexibility and requires minimal investment in both technology and human resource. Internal modeling solutions based on spreadsheet models do not fully address these challenges. They are error-prone, complex and difficult to maintain and scale. The proposed simplified spreadsheet model by Volarevic and Varovic (2018) is not tailored to the local market since it uses international risk ratings to determine the PD parameter. Out-of-the-box solution abstract the

18

complexities but require heavy initial investment and are very vendor specific. The literature also explores the concept of low-code/no-code automation and democratization of technology. All projections point to a future where complexities in developing automation solutions have been abstracted. These low-code/no-code platforms do not require extensive upskilling or initial investment to build and deploy solutions. The study proposes a solution that sits in between internal and out-of-the-box solutions by employing a low-code approach to financial modeling.

CHAPTER 3: METHODOLOGY

This chapter outlines how the systems analysis, design and implementation was conducted and the key considerations for each. It also describes the testing and evaluation approaches and their considerations as well.

3.1 System analysis approach

This section describes how the systems analysis was done. The goal of the systems analysis is to understand the bottlenecks in the current IFRS 9 modeling and reporting process in Kenyan banks. This section describes how data was collected, how it was analyzed and finally how the user requirements from the analysis were converted into system requirements.

3.1.1 Data collection approach

This section describes how data needed for systems analysis was collected. Two approaches were used for collecting quantitative and qualitative data respectively:

1. Administered questionnaire
2. Interviews

Questionnaire

For collecting quantitative data. The first section of the instrument captures demographic details of the respondents:

i. The type of bank – Local or International (different regulatory and reporting requirements depending on type of bank and areas of operation)

ii. The job function of the respondent (to understand whether the bank has a dedicated team of IFRS 9 analysts or it relies on the existing credit risk and IT teams)

iii. The largest sector in the bank's portfolio (PDs for different sectors are modeled differently)

iv. The percentage increase of reporting under IFRS 9 (to understand the impact of IFRS 9 on the bank's provisions)

The second section of the instrument administered was a self-assessment of the respondents in their knowledge of the IFRS 9 standard. The final section of the instrument looks into the current approach used and asks the respondents to gauge the level of automation of the current approach.

model IFRS 9 and the level of automation in that solution. For a detailed list of the questions asked, please refer to Appendix I.

Interviews

IFRS 9 modeling is contextual for every bank. The standard offers a framework for modeling and reporting but leaves model nuances open to interpretation by the bank as long as they can defend their model decisions to an independent audit. As such, there is no single model for all banks. Therefore, one of the locally listed banks, Kenya Commercial Bank (KCB), was chosen as a use case for model implementation and to demonstrate the solution.

Interviews were conducted with the bank's credit risk stakeholders using an unstructured interview approach. Questions asked during the interviews shed light on:

1. What is the current IFRS 9 modeling and reporting process and the steps involved?
2. At each step, what are the inputs, processes, expected outputs and stakeholders?
3. At each step, what are the bottlenecks of the processes?
4. At each step, what is the average time it takes to produce the expected output?

Document analysis

Document analysis involved going over KCB's excel models with the process owners at the bank to understand the bank's current IFRS 9 implementations. Datasets were obtained for the December 2020 period and with consent from the data owners at KCB. The goal is to understand the implementation of the model, especially the open-ended issues left to interpretation by the standard. This includes issues such as segmentation approach and definition of default based on days-past-due. As mentioned, each IFRS 9 model implementation is unique. The bank was chosen as a demo to implement the solution and demonstrate its capabilities in addressing the inefficiencies defined in the problem statement.

3.1.2 Data analysis approach

This section describes how the data collected from the questionnaire and interviews was analyzed. A descriptive statistical analysis approach was used for analyzing the questionnaire data with a focus on measures of frequencies. This approach was preferred because the data collected was categorical. Responses were bucketed and the frequency of each answer category obtained.

For the analysis of the interview output, a business process analysis approach was used, to crystallize information from the interviews and discussions stakeholders at the KCB bank. Each step in the IFRS 9 modeling business process was outlined with a focus on the inputs of each step in the process, the tasks involved, and the expected outputs. Dataflow diagrams were used to represent the flow of information within the current approach for each step.

3.1.3 Approach for defining system requirements

After data collection and analysis, this section outlines the approach for arriving at systems requirements from the user requirements. Based on the information gathered and analysis done on user requirements, a list of functional and nonfunctional system requirements was defined. The goal here is to answer the question:

> *What should the system do to meet the user requirements defined in the use case diagrams and DFDs?*

A system requirements specification was then outlined that contains the user description, the external interfaces that the system should connect to, and the functional and non-functional requirements specification. This SRS forms the final output of the analysis phase.

3.2 System design approach

This section describes the design phase in terms of what is to be designed and the tools and techniques used to come up with the designs. First is the high-level design the system architecture which shows the components of the system and how they interact. Next is the system modules and finally the design of the MSM model. From this, the system has the following main components that need to be designed:

1. High-level architecture design
2. Workflow design – the actual system modules
3. User Interface design
4. Database design
5. MSM model design

The workflow design refers to the system modules that need to be implemented. That is, data input module, data quality module and impairment module. Design considerations include – what are

the inputs for each module, what are the outputs for each module and how are the modules interconnected. A DFD design approach was used because it best represents processes and workflows and movement of data through the system.

Wireframes were used for UI design. This was done using the Alteryx designer tools which provides a wireframing functionality for prototyping UI components. These UI components must be tied to an Alteryx workflow and workflows are chained to ensure smooth navigation through the system. The Alteryx workflow is the engine that runs behind the scenes and these are the modules. The user interface sits on top of these workflows.

Database design was informed by the current KCB implementation approach, the datasets they use and the key fields needed. For building the data model, a tool called Database Diagram (dbdiagram.io) was used to model the key datasets, the fields, and how these datasets relate. It is an open-source tool that allows you to define a relational database design diagram through code. It also generates the SQL DDL and DML statements for creating the database schema.

The MSM model design was represented as pseudocode with an example of one facility to show how the loans and sectors are tracked across the periods for PD computation.

3.3 System implementation approach

This section describes how the system was implemented. The first sub-section outlines the standard SDLC method used. The next sub-section shows the platform and tools used for implementation and considerations made when selecting them.

3.3.1 System development approach

An agile development methodology with Kanban for tracking tasks was used as the standard system development methodology. Agile is preferred because of its iterative approach. It allows the solution to be built, and tested in incremental stages of the project. The choice for agile was informed by the overall system architecture illustrated in Section 2.11. The section outlines a modular system design that lends itself to agile development with each module being developed separately and iteratively. Kanban is an enabler within agile development for easy tracking of work items and tasks. It allows project teams to prioritize tasks within shortened time cycles and measure throughput metrics to ensure the project remains on schedule (Radigan, 2020). The agile Kanban approach was chosen for the flexibility and simplicity it offers under project management.
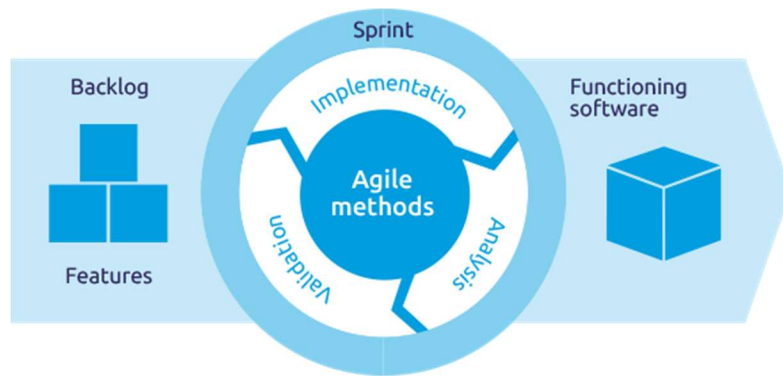
Figure 3 The agile process adopted from the Atlassian (2019)

3.3.2 Platform and tools used for workflow, database, UI and dashboard implementation

The system was implemented primarily on the Alteryx Platform. Alteryx was chosen as the platform of choice because it is a low-code application development platform. Such platforms provide a user-friendly way of building complex analytics applications. Alteryx also provides additional features such as UI design and integration, connecting to a wide range of data sources, and interfacing with popular BI tools for reporting. Alteryx R-tools was used for implementation MSM model on R. The bulk of the impairment logic sits on Alteryx workflows. The user interface was built on Alteryx designer on top of the workflows.

A local PostgreSQL instance was used as the primary database. It stores datasets needed for modeling, the data quality results, the impairment results and the configuration parameters for each execution. PostgreSQL was chosen because it is an open-source relational database platform with a supported JDBC connector on Alteryx. Any similar relational database can be used for this purpose.

A local PowerBI instance was used as the reporting and visualization. Key metrics reported include, the ECL, the PD, the LGD and the key drivers for each parameter. The dashboard is connected to the PostgreSQL database for automated integration. The dashboard is a value-add for the user because it helps to uncover the key drivers of the ECL parameter. It also helps the user understand the key accounts and key segments in their loan book that are driving the ECL.

24

3.4 System testing approach

This section shows how system testing of the outputs was done. The primary outputs to be tested are the sector-level PDs and the sector level ECLs. Accuracy of these two parameters indicates accuracy of the model. These outputs need to be benchmarked against some acceptable point-of-reference. Also recall that any IFRS 9 model design and implementation is unique to a bank. Therefore, the acceptable point-of -reference chosen should be within the bank's context.

Therefore, for testing, the PD and ECL output was compared to the reported financial statements for KCB bank for the period ending December 2020. As advised by the credit risk stakeholders, the bank would only adopt the new approach if the results across all sectors did not exceed a 2% variance from its current internal credit risk scoring approach. The 2% mark is an arbitrary evaluation metric depending on the risk appetite of the bank. PD and ECL outputs were compared sector-wise and outputs were considered accurate and valid if they met the set threshold.

Analysis of the test results was done on Excel by taking an absolute difference between the Alteryx output and the output reported by the bank to determine the variance. This analysis was done for each of the bank's segments and loan products.

3.5 System usability testing approach

This section outlines how usability testing was done for the new system. The metrics and questions used to evaluate the system were derived from the System Usability Measurement Inventory (SUMI) for evaluating the user's experience. Due to cost and time constraints 20 respondents were selected to the system demonstration and evaluation – the minimum number of respondents needed for representative results as outlined by the SUMI guide (Jurek & Corbett, 1993). SUMI questions are structured in such a way the user responds with either agree, undecided or disagree (Jurek & Corbett, 1993).

The system was evaluated on four metrics – efficiency, affect, learnability and control – with 5 questions for each metric. The following is a description of each metric and the questions therein:

Efficiency **-** The degree to which the user can achieve the goals of his interaction with the application in a direct and timely manner. Questions asked:

1. The instructions and prompts are helpful.

2. I sometimes don't know what to do next with this software.

3. This software has at some time stopped unexpectedly.

4. Getting data files in and out of the system is not easy.

5. Learning to operate this software initially is full of problems

Affect - How much the application captures the user's emotional responses. Questions asked:

1. I would recommend this software to my colleagues.

2. It is obvious that user needs have been fully taken into consideration.

3. There are too many steps required to get something to work.

4. This software responds too slowly to inputs.

5. The way that system information is presented is clear and understandable.

Learnability – The ease with which a user can get started and learn new features of the product. Questions asked:

1. It takes too long to learn the software functions.

2. I feel in command of this software when I am using it.

3. Tasks can be performed in a straight forward manner using this software.

4. Learning how to use new functions is not difficult.

5. I will never learn to use all that is offered in this software.

Control - The degree to which the user feels s/he, and not the software, is setting the pace. Questions asked:

1. I feel in command of this software when I am using it.

2. I prefer to stick to the functions that I know best.

3. I keep having to go back to look at the guides.

4. It is easy to make the software do exactly what you want.

5. This software occasionally behaves in a way which can't be understood

The questions were administered as an online questionnaire to potential users. Responses were collected on Google Sheets and analysis done on Python. PowerBI was used to visualize and represent the output of the analysis.

3.6 System evaluation approach

This section shows how the system was evaluated in line with the problem statement as defined in section 1.2:

The analysis and reporting process of becoming IFRS 9 compliant is time-consuming. Banks iterate over the credit risk modeling process because of:

1. Data requests taking too long to be fulfilled due to poor integration with existing data sources.
2. Errors in the data input.
3. Lack of automation of common tasks.
4. Error-prone outputs.

This section described how the system was evaluated against what it sought out to address as defined in the problem statement. A process comparison was done showing the process in the current approach vs the process in the new system. Additionally, efficiency of IFRS 9 the modeling and reporting was defined as the answer to the following question:

> ***How much time in minutes does it take from the start of the modeling process to when a valid output is availed to end-users?***

CHAPTER 4: ANALYSIS, DESIGN AND IMPLEMENTATION

This chapter details the outputs of the system analysis, design, implementation and testing steps as described in the methodology chapter 3.

4.1 Analysis

This section describes the output of the systems analysis process as described in section 3.1 Systems analysis approach.

4.1.1 Output of questionnaire

On demographics, from the results of the questionnaire administered, 76% of the respondents were employees from local banks while 24% were from international or internationally backed banks. A majority of the respondents were from the credit and risk departments.
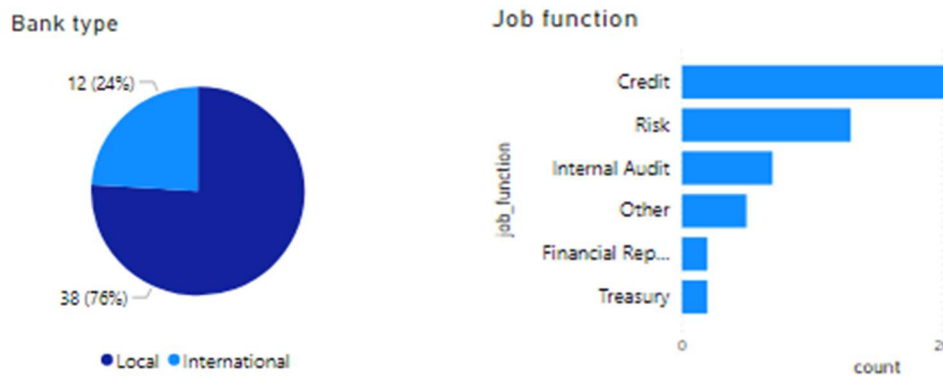

Figure 4 Demographic analysis of respondents

39% of the banks represented by the respondents have consumer as their largest segment in their loan book portfolios. Retail came a close second with 26%.
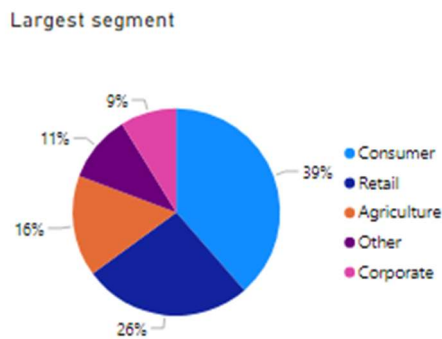

Figure 5 Largest portfolio

28

47% of the respondents noted an increase of at least 10% since they started reporting under the new standard. 7% noted a decrease in provisions while 7% recorded an increase above 40%.
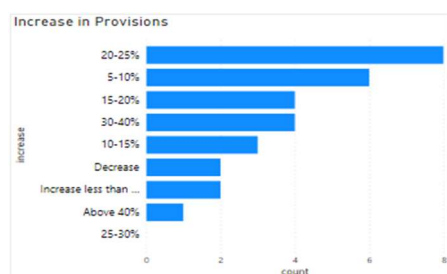

Figure 6 Increase in provisions

The second part of the instrument administered was a self-assessment of the respondents in their knowledge of the IFRS 9 standard. As seen in the literature review one of the key challenges in adoption is lack of knowledge and expertise among the bank personnel.

The following is an analysis and commentary on the responses received for the section. A majority of the respondents (39) responded recorded that they are or have been directly involved in IFRS 9 reporting at their respective banks while 11 had not been directly involved in the reporting process. The respondents were asked to rate their: understanding of the standard, proficiency in statistical modeling, excel, and analytics.

On a Scale of 1-10 Rate yourself on:

Rating	Understanding of IFRS 9	Proficiency Statistical Modeling	Proficiency Excel	Proficiency Analytics & Visualization
1	0.82%	4.08%	6.37%	3.15%
2	3.27%	8.16%	8.61%	8.56%
3	8.57%	14.29%	12.36%	21.17%
4	9.80%	12.24%	11.99%	12.16%
5	24.49%	24.49%	21.72%	22.52%
6	12.24%	10.20%	10.49%	10.36%
7	20.00%	14.29%	14.23%	13.96%
8	9.80%	6.12%	7.87%	4.95%
9	11.02%	6.12%	6.37%	3.15%

Figure 7 Respondents self-rating matrix

On average the most of the participants showed a greater understanding of the standard than the tools used to deliver on its modeling and reporting with over 70% rating themselves 5 or above. The respondents had the lowest self-ratings in analytics and visualization with only 30% giving

themselves a rating that is over 5. The proficiency in statistical modeling and excel was average with a majority of the respondents rating themselves between 4 and 6 for both questions.

Average Self Ratings

Directly Involved in IFRS 9?	IFRS 9 Understanding	Statistical Modeling	Average of Analytics and Visualization	Excel Proficiency
Yes	5.55	4.37	4.39	5.63
No	3.09	3.27	5.00	4.82

Figure 8 Self rating grouped by Directly Involved in IFRS 9?

The respondents who were directly involved in reporting had a higher average score for all categories except analytics and visualization.

Only 8.2% of the banks represented by the respondents had more than 5 certified IFRS 9 specialists within their banks. 91.2% had less than 5 with some respondents having no in house specialists at all. These were assumed to outsource all their IFRS 9 modeling and reporting needs.

How many IFRS 9 certified specialist do you have in house?

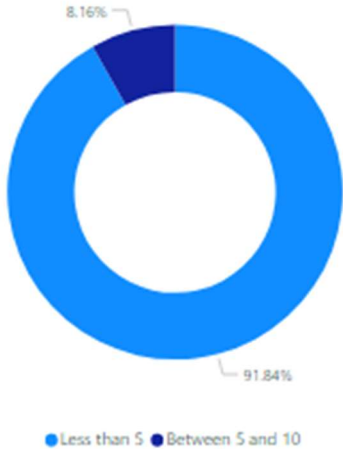*Figure 9 How many IFRS 9 certified personnel do you have?*

The majority of respondents across both categories (those that were directly involved in modeling and those that were not) stated that they would prefer an outsourced IFRS 9 modeling and implementation approach as opposed to maintaining one in-house. The respondents argued that this would free them up from the complexities of implementation every reporting cycle.
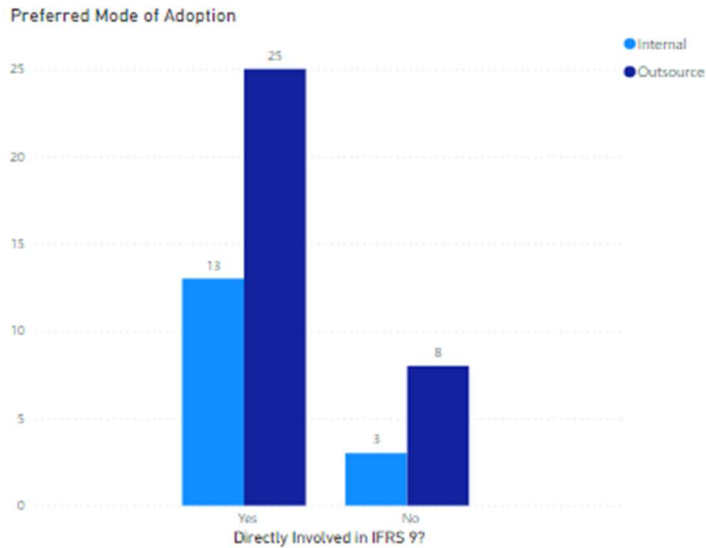
Figure 10 Preferred mode of adoption

However, despite most of the respondents championing for outsourced approaches, 79% of them were either unaware of the resources set aside by their bank for such efforts or did not believe their bank had allocated enough resources towards IFRS 9 implementation. 20% of the respondents said they believed that the resources allocated by the bank towards IFRS 9 modeling and reporting were sufficient.

Figure 11 Resources allocated to adoption

The last question in this section was only answered by the group of respondents who were involved in the IFRS 9 modeling. It inquired on the most difficult parameter to model in light of the standard and data requirements. A majority of the respondents, 41% said that PD was the most difficult

parameter to model. This was followed by EAD (20%) and LGD (16%). The respondents highlighted the fact that PD requires either a mature risk rating system in place or historical data up to 5 years. The availability of this data might be impacted especially with core banking system changes.
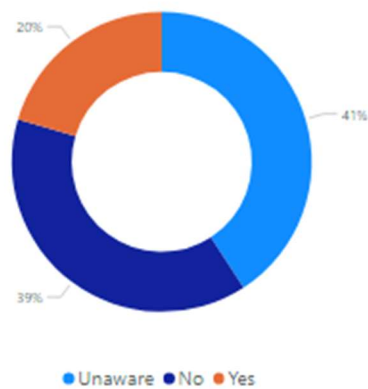
Most challenging IFRS 9 parameter to model given the standard and data requirements?
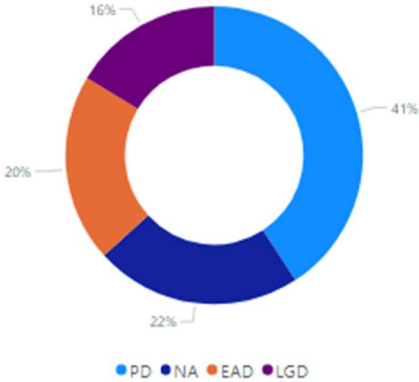
Figure 12 Most challenging IFRS parameter to model

This part of the instrument profiles the bank in terms of the current solution used to model IFRS 9 and the level of automation in that solution.

The following is an analysis and commentary on the responses received for the section. 45% of the respondents said that they use enterprise spreadsheet models in implementing IFRS 9 solutions. 31% use fully outsourced solutions (or out-of-the-box solutions) and 24% use a custom solution.

On what platform do you currently model IFRS 9?
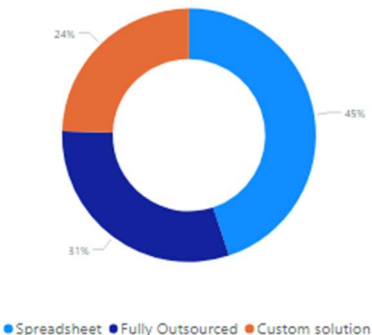
Figure 13 On what platform do you model IFRS 9

On integrated analytics 53% of the respondents said they have some sort of analytical add-on to the modeling approach. Those who relied on spreadsheet modeling solutions reported a

disproportionate lack of integrated analytics capabilities. A majority of those who used out-of-the-box and customized solutions did have analytics embedded as a feature.
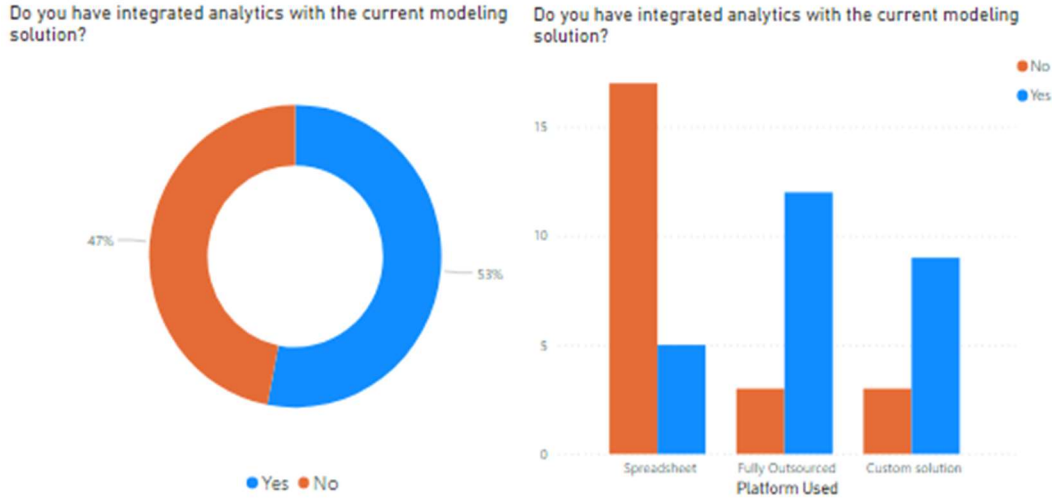
Figure 14 Do you have analytics integrated in your approach

60% of the respondents said they did not have an internal risk rating in place except for their top accounts on which they apply management overlays. All of the respondents who use a custom internal solution did not have a risk rating in place and relied solely on historical data to model PDs. A majority (63%) of the respondents who use spreadsheet models had a risk rating implemented in place of historical data. While a majority (56%) of those who use fully outsourced model relied on historical data for PD modeling and did not have a risk rating in place.
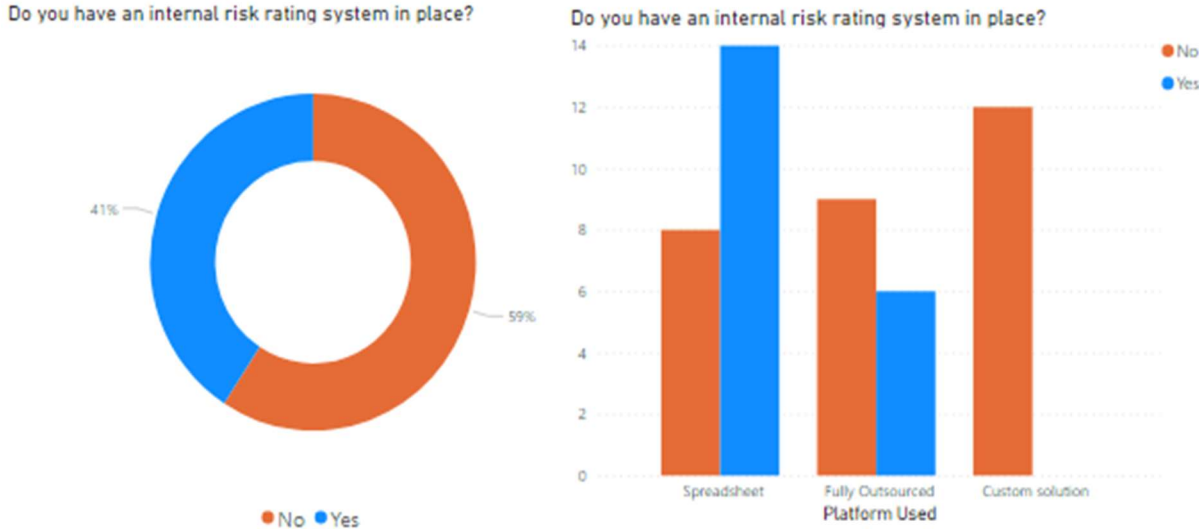
Figure 15 Do you have an internal risk rating system

The next set of questions looked into the amount of historical data available in years (for PD modeling), the number of personnel involved in modeling, the turnaround time in days and the number of audit queries raised during the end-year and interim audits. On average, all banks represented in the study had enough historical data as required by the standard for PD modeling (at least 5 years of data). Banks that relied on spreadsheets models had the highest number of average personnel involved in modeling and reporting – 4.14. Respondents from these banks also recorded the highest average turnaround time (6.95 days) and the highest number of average audit queries (4.55).

Platform Used	Average of Historical data years	Average of Personnel Involved	Average of TAT in days	Average of Audit queries
Custom solution	6.58	3.00	6.08	3.50
Fully Outsourced	6.40	2.33	4.27	4.07
Spreadsheet	6.14	4.14	6.95	4.55

Figure 16 Averages by platform used

Almost all banks in the survey reported that the most difficult element in implementing the controls framework for IFRS 9 is the controls in relation to data and data quality. Due to the complexity of the standard, the level of data inputs to the calculation process is significantly higher than under previous processes.

Figure 17 Most difficult element in implementing IFRS 9 controls

When asked to rank the features of a potential IFRS 9 solution, the respondents ranked accuracy and cost as the most important features with each scoring 8.97 and 7.56 respectively. Analytics and automation came 3rd and 4th with average scores of 7.43 and 6.56 respectively.
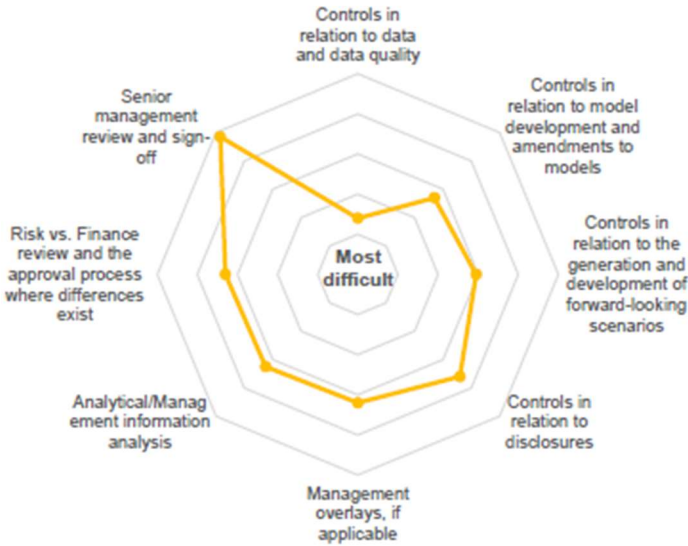

How important are the following elements in a potential IFRS 9 solution

param	avg_im
im_accuracy	8.97
im_cost	7.56
im_analytics	7.43
im_automation	6.56
im_flexibility	5.89
im_integration	5.48
im_time	4.87
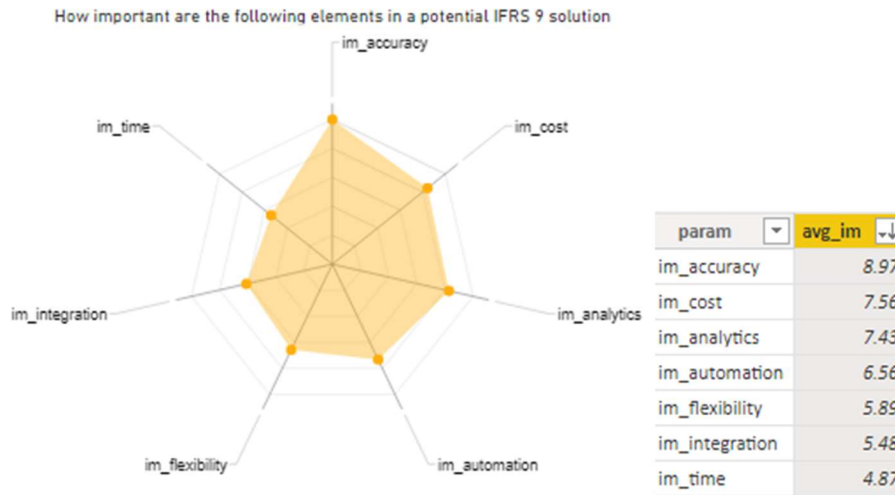
Figure 18 Feature ratings for a potential IFRS 9 solution

4.1.2 Output of interviews

This section summarizes the discussion that were had with the stakeholders from credit risk and IT and their responses to the interview questions.

What is the current IFRS 9 modeling and reporting process and the steps involved?

The first question was on the current IFRS 9 modeling process and the steps involved. From the interviews and discussions, the modeling process at the KCB bank has the following key steps – model design, data extraction, data quality checks, data input, model execution and results discussion and validation.

At the model design step, any changes to the previous model design are implemented on Excel, reviewed and signed off. Then a data request is made from the modeling team to the IT team for extraction of relevant datasets with key fields. The data is sent to the modeling team in form of excel files. The next step is the data quality checks where the modeling team runs manual sanity checks against the data and the key fields needed for modeling. If the data quality checks reveal significant errors on inconsistencies, a data quality report is sent back to the IT team and a re-extraction and re-validation is done.

Once the modeling team confirms that the data extracted is fit for modeling, they copy and paste the cleaned data on to the excel models. The modeling team maintains three separate models for the three IFRS 9 parameters – PD, EAD and LGD. Output of the three models is then copied into a final ECL model that produces the preliminary results. The credit team then checks the ECL output for validity, performs some analysis on Excel showing the key drivers for ECL.

The results of the analysis are then presented to management for review and sign off. If management is not satisfied with the output based on their understanding of the business, they suggest changes to the model or data. For example, in one execution Kenya Airways was contributing to a very large percentage of their ECL because of their deteriorating facilities. However, government bailouts were expected and management chose to apply an overlay on the customer's PD to bring their ECL down. In IFRS 9, these are called management overlays and must have a clear justification and documentation as they form part of audit queries. Management comments that demand a data change might trigger a re-extraction. This is an iterative process until valid and representative results are obtained after-which management signs off on the final numbers to be included in the financial statements.

At each step, what are the inputs, processes, expected outputs and stakeholders?

From the responses, the inputs, processes, outputs of each step are outlined below.

Model design

This is the first step of the reporting process. Any changes to the model design specifications from the prior reporting period are discussed and approved by the management team and the credit risk team. These changes are then implemented by the IFRS team and the final model design is reviewed with appropriate sign offs.

Inputs	Processes	Outputs
Model design considerations	Convert design specification to excel models	Model design document Excel models

Table 1 Model design process for chosen bank

Data extraction

The IFRS Team sends out a data request to the bank's IT team to extract reporting data such as the loan listings and collateral for the period from the core banking system. The IFRS Team also sends out a similar request to the Credit Risk team to provide updated information on things like risk ratings, facility restructure from a separate system.

Inputs	Processes	Outputs
Data request	Extract data from different database sources	Excel files containing raw extracted data

Table 2 Data extraction process for chosen bank

Data quality checks

The IFRS team performs predefined checks on the extracted datasets. This ensures consistency on key fields used in modeling such as the facility type, the segment and staging information.

Inputs	Processes	Outputs
Data quality checklist	Perform data quality and consistency checks on Excel	Data quality reports on key data sets for modeling

Table 3 Data quality process for chosen bank

Data input

This involves copy pasting data sets from the excel sheets into the model workbooks.

Inputs	Processes	Outputs
Cleaned data files on excels	Copy and pasting to update the data in the model	Populated model

Table 4 Data quality process for chosen bank

Model execution

The different excel models for the IFRS 9 parameters – PD, LGD – are run by the IFRS team. Below is a breakdown of the inputs, processes and outputs for each:

Model	Inputs	Processes	Outputs
PD	Internal qualitative credit risk	Assign customer level PDs based on the ratings	Probability of Default for customer

		Calculate the collateral	Collateral and
LGD	Collateral listing	LGD for covered	Segment level LGD
	At least 5 years of	facilities	
	collections data	Calculate the segment	
		level LGD for	
		uncovered facilities	

Table 5 Model execution process for chosen bank

<u>Results discussion and validation</u>

The IFRS team communicated preliminary results to the management and internal audit teams. It is an iterative process where the model and data are subject to changes. The final results are reviewed and disseminated after sign off from management.

Inputs	Processes	Outputs
Model results	Summarizing the results:	Summary of the model results
	Top provisions	
	Segment level summary	
	Largest movements etc.	

Table 6 Results presentation process for chosen bank

<u>*At each step, what are the bottlenecks of the process?*</u>

This question seeks to understand the inefficiencies of the current process as outlined in the first and second questions. From the responses, the first bottleneck is that the data request step. Fulfillment of the data requests take too long. Sometimes the contact person at IT is unavailable to handle IFRS 9 data requests. This further increases the time complexities of the first step. From this realization, one of the primary user requirements would be seamless extraction of key datasets needed for modeling. Recall that data requests are also iterative after management gives direction for updates to the data.

The second step is data validation. This is a manual process where the modeling team checks the key modeling fields for consistency and sanity. This includes the account numbers, customer numbers, repayment terms, restructuring history, segmentation information, product type information, maturity date, and value fields like balance and collateral value. This is a time-

consuming process done on Excel and a data quality report is generated and sent back to IT for the datasets deemed not fit for modeling.

After data quality checks and validation, the data has to be input into the models. Recall that they have 3 different models one for PD, EAD and LGD. Data input involves copying and pasting manually onto the excel sheets. This is a brittle process that breaks easily. Copying the wrong field or inconsistent field order might break the models leading to garbage output.

Model execution is core work of the modeling team. Outputs of the PD, EAD, and LGD models are fed into the final ECL spreadsheet. Once housekeeping issues of data request, extraction and validation are done, a typical run takes an analyst on average 10 mins to produce and output for all parameters. This assumes that the model does not break when copying data into it.

The next step is to distill the results for management. The modeling team does analysis of the preliminary results and summarizes the output for management consumption. This takes another 10 mins. With a fresh analysis done for each output iteration. Feedback and signoff from management takes about 2-3 days to obtain because they have to comb through the results, figure out the drivers and recommend adjustments for the next iteration. Running the model and summarizing the results in one iteration takes ~20 mins. Other internal processes around the modeling (such as data requests) take longer depending on availability of the contact person and their bandwidth to handle IFRS 9 requests. This further stretches out the modeling process.

4.1.3 Output of document analysis

As mentioned, the document analysis revealed that KCB relies on 3 models for the parameters. This means, that the credit risk team maintains three excel workbooks. Screenshots of these workbooks and documents reviewed are shown below:

Figure 19 LGD Model Input

Figure 20 PD Model Input

Figure 21 ECL Model Output
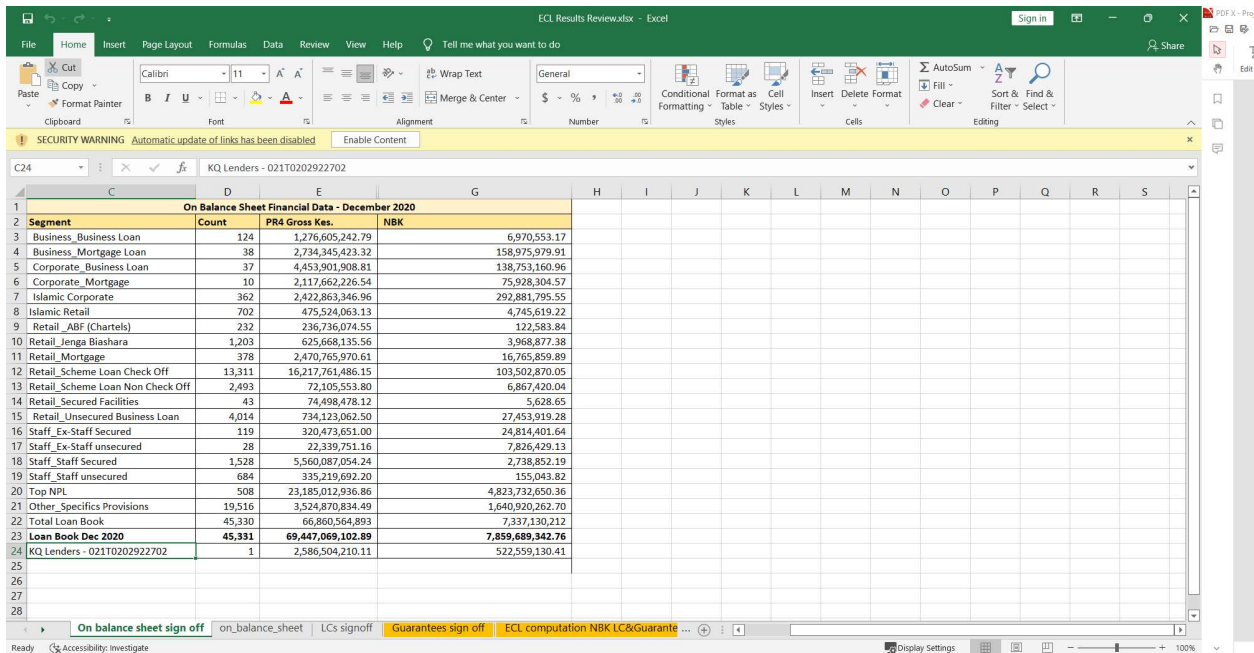
Figure 22 ECL Results Summary

A unified model on excel would be too complex and too difficult to maintain. That is why each parameter is modeled on its own workbook with a separate one for ECL results summary. Outputs are chained by copy and pasting which further increases the likelihood of making errors. This in turn increases the number of iterations making the modeling process more time-consuming.

4.1.2 Summary of user requirements

From the results of the analysis phase, it is evident that the switch to IFRS 9 has had a significant impact in the provisions reported by local banks. 47% of the respondents recorded an increase of 10% or more. 7% recorded an increase above 40% while a majority of the respondents noted that their increase is between 20-25%. The current implementation approaches in local banks are heavily spreadsheet based. From the responses (and from literature), these approaches require more human input, have a higher turnaround time and are more error prone. Also, from the responses data and data quality controls are the most difficult to implement. Lack of these controls leads to more errors and higher number of audit queries. A majority of respondents noted that using spreadsheet models lack integration. The results from the analysis revealed the need for an integrated, flexible and automated IFRS 9 solution that has accuracy top of mind and minimizes time spent and human input. Analysis of the interview responses and process review showed that getting a summarized outputs to management takes bout ~20 mins for each run. The time cost is compounded by internal processes such as data requests, manual data quality checks and data analysis for each iteration. On average, the bank goes through at least 3 iterations for each reporting period.

4.1.3 System requirements specification

This section outlines the target user description, the external interfaces for the system, the system's functional requirements and non-functional requirements based on the output of the analysis.

User description

The user is expected to be a computer-literate. They are also expected to understand the IFRS 9 standard in terms of: the inputs needed at field-level, the design for each IFRS 9 parameter and the outputs expected. The user is typically a member of the credit team or the financial modeling and reporting team (if the bank has such a dedicated department in place).

External Interfaces

The system will connect to a database that stores all datasets required for modeling. These will be views created from the core banking databases which house the raw data.

The system will also connect to a dashboard interface through an external API for automatic summarization of data quality checks and model results.

Functional requirements

This section uses use cases to represent details of the system's functional requirements.

1. Users should be able to load and update datasets needed for impairment.

Use Case Name	Load and update data
Trigger	The user accesses the application
Precondition	The external interface to the core banking database is configured
Basic Path	1. The user selects the update data module 2. The user selects which datasets to update 3. The user maps the fields from the data source(s) on the application 4. The user clicks to run the data input workflow 5. The system alerts the user on progress and update status of each dataset selected. 6. The system routes the user back to the start of the application
Alternative Paths	If in Step 2 the user does not select any dataset: The system routes the user back to the start of the application
Postcondition	The updates datasets are loaded and available for modeling
Exception Paths	The user may abandon the update process at any time The required interfaces have not been configured
Other	The system should timestamp all datasets loaded for freshness checks

Table 7 Use case for loading and updating data

2. Users should be able to run data quality checks on the uploaded datasets to ensure that the data is appropriate for modeling.

Use Case Name	Perform data quality checks
Trigger	The user selects the data quality module

Precondition	Data has been loaded at least once for every required dataset The external interface to the database has been configured The external interface to the dashboard has been configured
Basic Path	1. The user selects the data quality module 2. The user selects the datasets they would want to perform checks on 3. The user clicks to run checks against selected datasets 4. The system updates the dashboard with the data quality output 5. The system routes the user back to the start of the application
Alternative Paths	If in step 2 the user does not select any dataset: The system routes the user back to the start of the application.
Postcondition	The data quality results are loaded on the database and reflected on the dashboard
Exception Paths	The user may abandon the data quality process at any time The user has not loaded data for a selected dataset The required interfaces have not been configured
Other	The system should timestamp all results loaded for freshness checks

Table 8 Use case for performing data quality checks

Users should be able to run impairment using uploaded datasets

Use Case Name	Run impairment
Trigger	The user selects the impairment module
Precondition	Data has been loaded at least once for every required dataset. The external interface to the database has been configured The external interface to the dashboard has been configured
Basic Path	1. The user selects the impairment module 2. The user clicks to run PD, LGD, EAD, and ECL modules 3. The system updates the dashboard with the data quality

	output 4. The system routes the user back to the start of the application
Alternative Paths	N/A
Postcondition	The results are loaded on the database and reflected on the dashboard
Exception Paths	The user may abandon the impairment process at any time The required datasets have not been loaded The required interfaces have not been configured
Other	The system should timestamp all results loaded for freshness checks

Table 9 Use case for computing impairment

<u>*Non-functional requirements*</u>

The non-functional requirements describe how a system must behave defining the constraints of its functionality:

  i.   The system should be flexible to accommodate model design and data changes.
  ii.   The system should be able to integrate with existing databases as a source for impairment data.
  iii. The system should accurately model the agreed upon design.

4.2 Design

This section describes the output of the systems design process as described in section 3.2 Systems design approach. The first design is the high-level architecture of the new system. Next are workflow designs represented as DFDs for the 3 modules – data input, data quality and impairment. Then the database design and finally wireframes representing the UI and dashboard designs.

<u>*4.2.1 High-level architecture*</u>

The high-level architectural design is informed by the conceptual working of the new system. It shows the different components of the system and how they interact with each other. The first component is the user interface. Here the user should be able to update model configurations such

as the reporting entity and reporting period. These are then persisted in the database with a timestamp. The connector module controls the flow of execution for subsequent steps based on the user selection. There are three execution flows – impairment, data input and data quality. The data input module allows the user to update all or some of the datasets needed for modeling by pointing to source database. These datasets are then loaded onto the system database timestamped. The data quality module allows the user to run data quality checks against the data used for modeling. Data quality results are also stored on the database timestamped. The impairment module allows the user to run the model and produce timestamped results in the database. With the database and reporting interface configured, the latest results for both data quality and impairment should reflect on the dashboard.
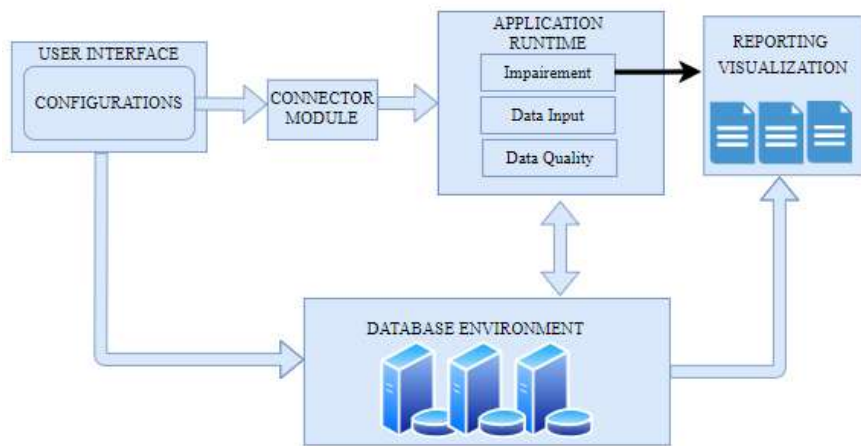
Figure 23 High-level system architecture

4.2.1 Workflow design DFDs

The workflows refer to the actual system modules that are to be implemented. These include the data input module, data quality module and impairment module. DFDs were used to represent the flow of information in the system.

Data input DFD

Here the user is updating a dataset in the system database. The input to the system is a data source pointer. This could be a configured database instance or a flat file – it is arbitrary. Next the system pulls the data and maps the fields in the dataset to those expected for modeling. The user needs to validate this mapping for a first-time connection. The mappings are persisted in the system

database for subsequent connections. The system then needs to timestamp the data and store it in the corresponding table on the system database.
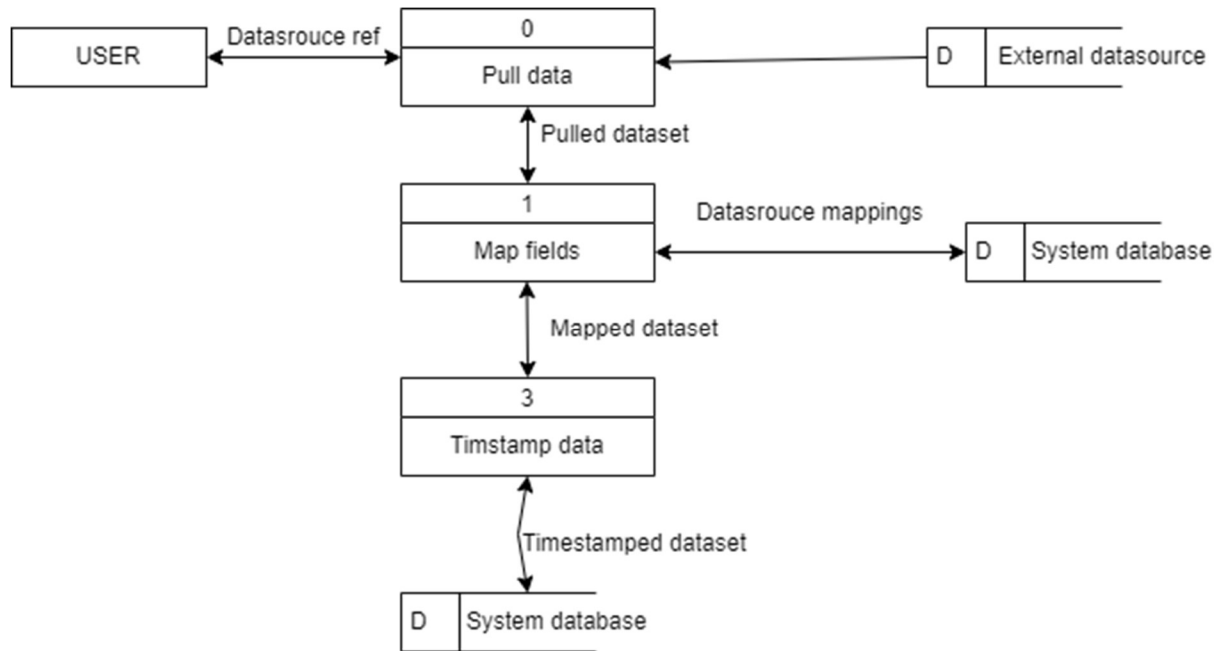
Figure 24 DFD data input

Data Quality DFD

Here the user selects the dataset on which they want to run the data quality checks. This step has 2 preconditions – the dataset needs to have been uploaded at least once and the data quality checks need to have been configured. The system the pulls the dataset from the system database. Runs the preconfigured data quality checks to produces the data quality output. Timestamps the data quality output and persists it onto the system database for dashboard updates.
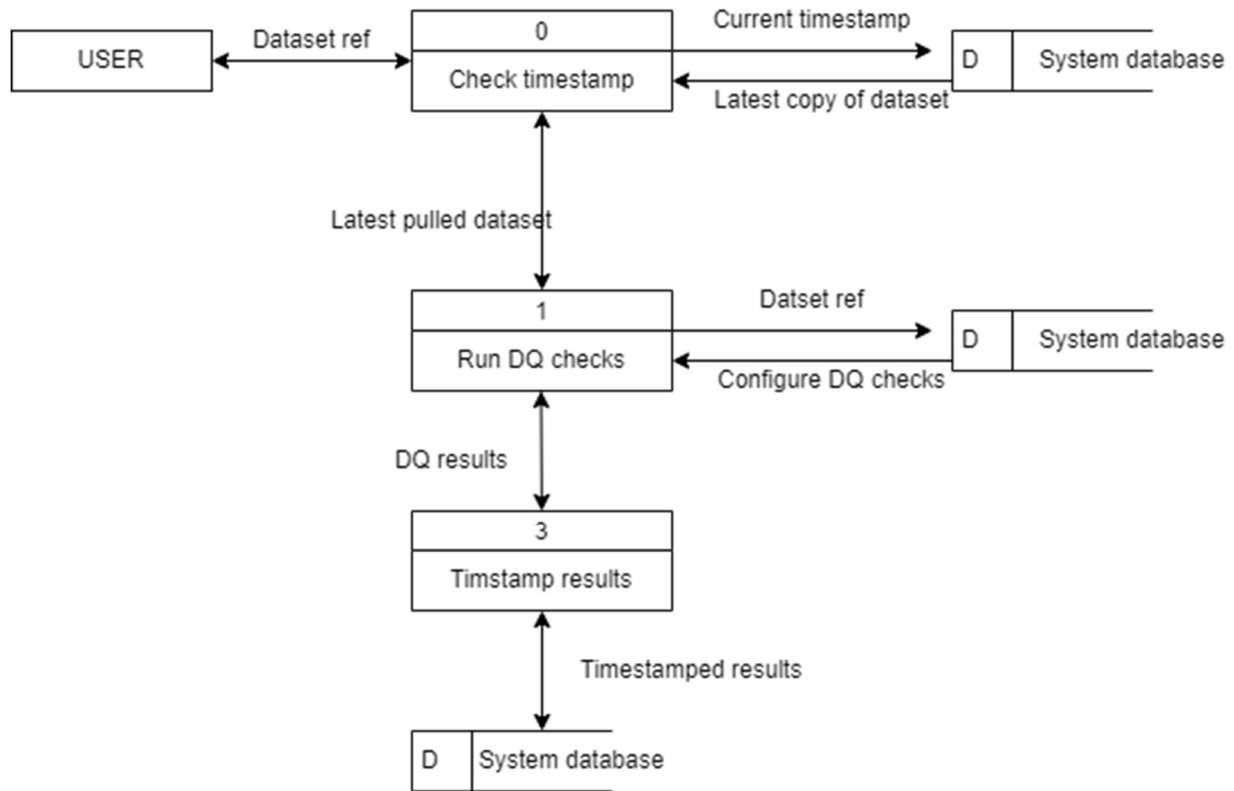
Figure 25 DFD data quality

Impairment DFD

On impairment the user inputs the name of the reporting period for which they want to run impairment. The system then queries the database and pulls the needed datasets from the system database. It runs the individual workflows in this order – PD, LGD, EAD and lastly ECL. After computing each parameter, it timestamps the results and persists them on the database.
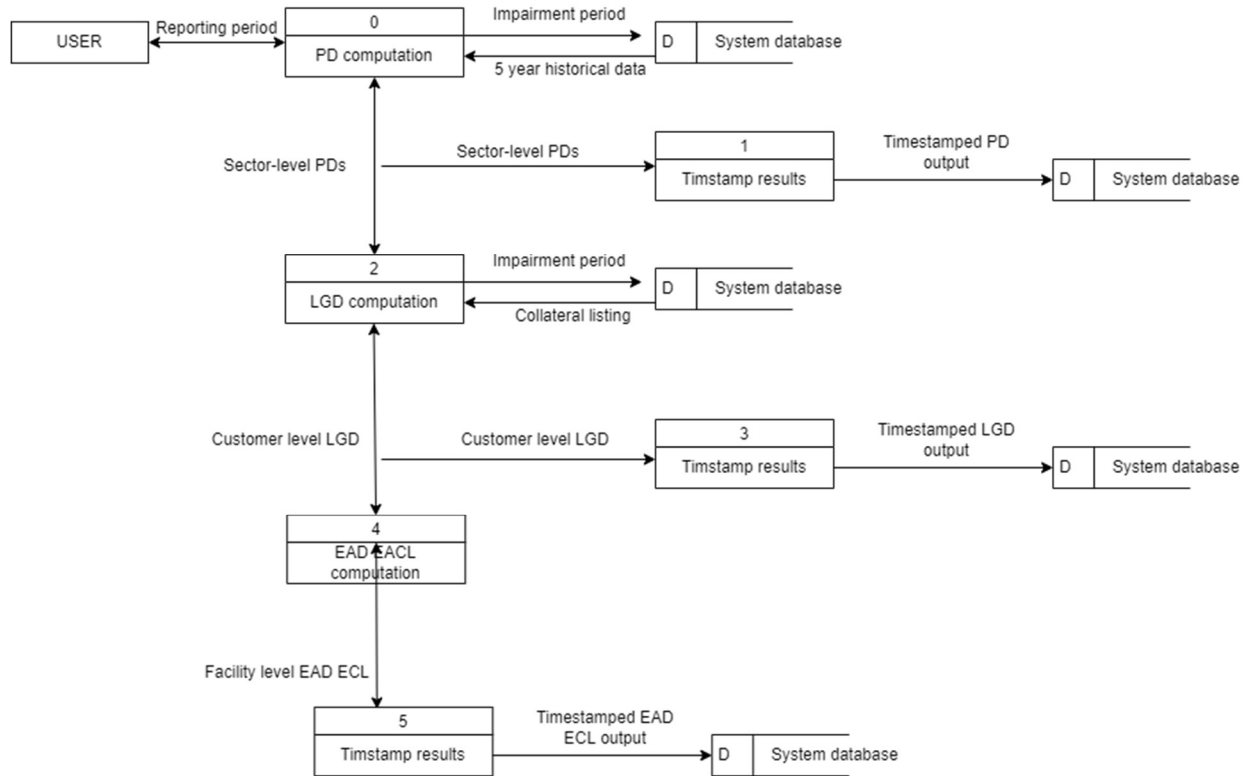
Figure 26 DFD Impairment

4.2.2 Database design ERD

A star schema approach used with the loan book being the central fact table and the other tables being dimensional tables. The reference between the loan listing and the other assets and off-balance sheet listing table is the account number. Collaterals are defined at customer level, hence the reference between the loan listing and the collateral table is the customer number on a one-to-many relationship. This means that one customer's collateral is mapped to all of their loans in the loan book. Finally, the collateral LGD is computed differently for each collateral type. That is why we have a collateral types table to get the LGD computation parameters for each collateral type. The reference between the collateral and collateral type tables is the collateral ID on a one-to-many relationship. Refer to Appendix II; Datasets Used for an anonymized snapshot of the datasets in the following diagram:
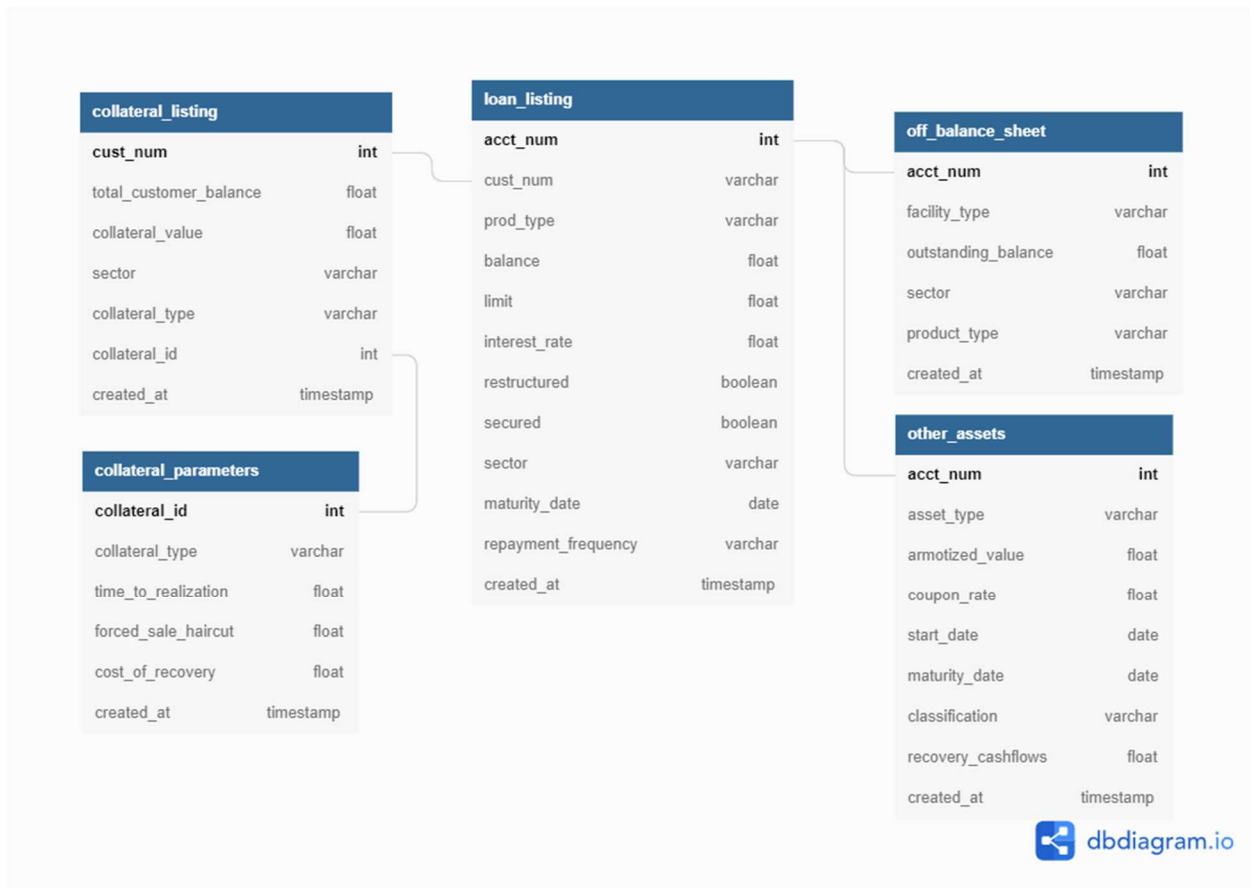
Figure 27 Input Database Schema

4.2.3 UI design wireframes

This section outlines the design of the user interface and dashboard of the proposed system. It includes wireframes done on Alteryx designer of various screens necessary for the execution of the three primary tasks – data input, data quality and impairment. It also includes wirefrmes for the dashboard to

First page – This is the landing window where the user enters the reporting entity's name and the impairment date
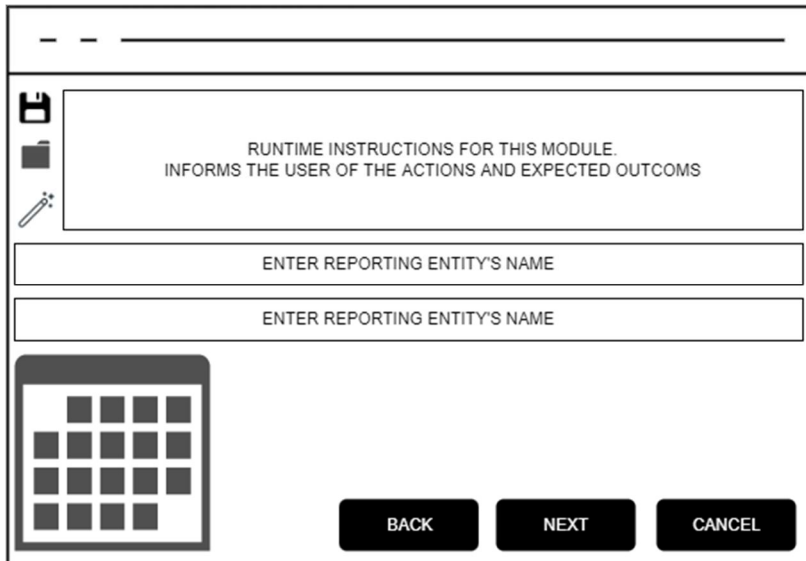
Figure 28 Landing page mockup

On the next page, the user selected the execution path to take from a dropdown list.
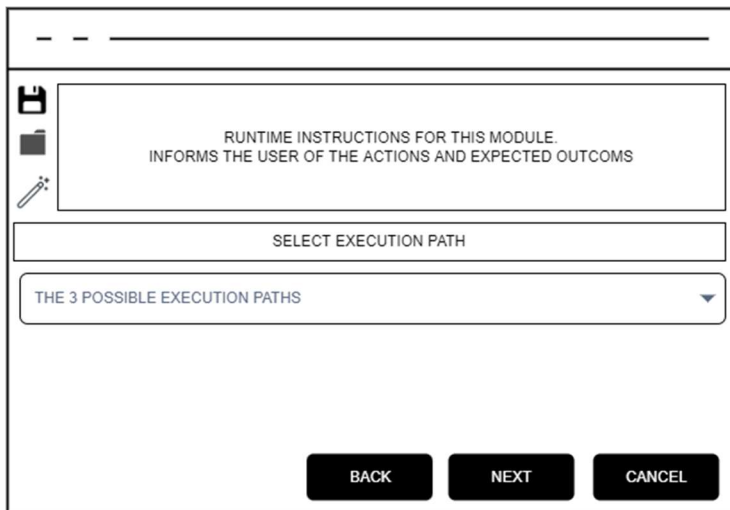
Figure 29 Select execution path page mockup

Input data – on this screen the user selects where or not they want to update a dataset. They then select the source file of that dataset and after loading it, map the fields in that source file to the expected fields for modeling.
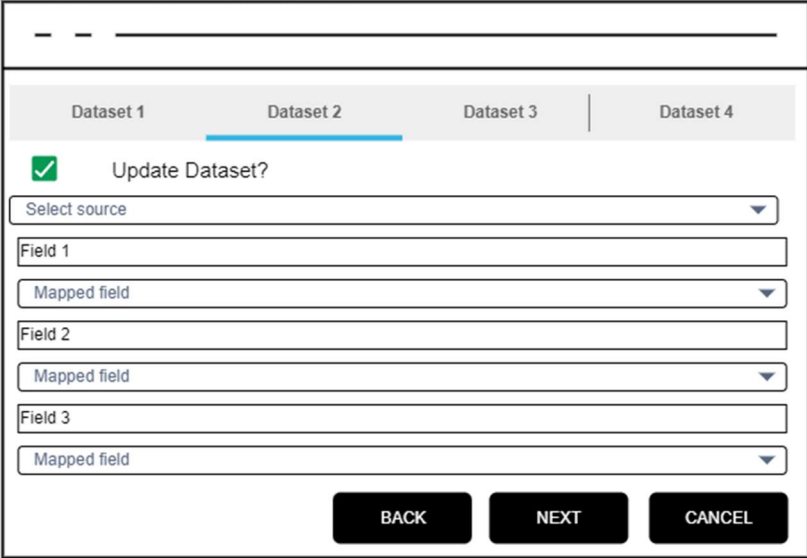
Figure 30 Update dataset page mockup

In the data quality module, the user simply selects the datasets for which they want to run the data quality module from a dropdown list of the used datasets.
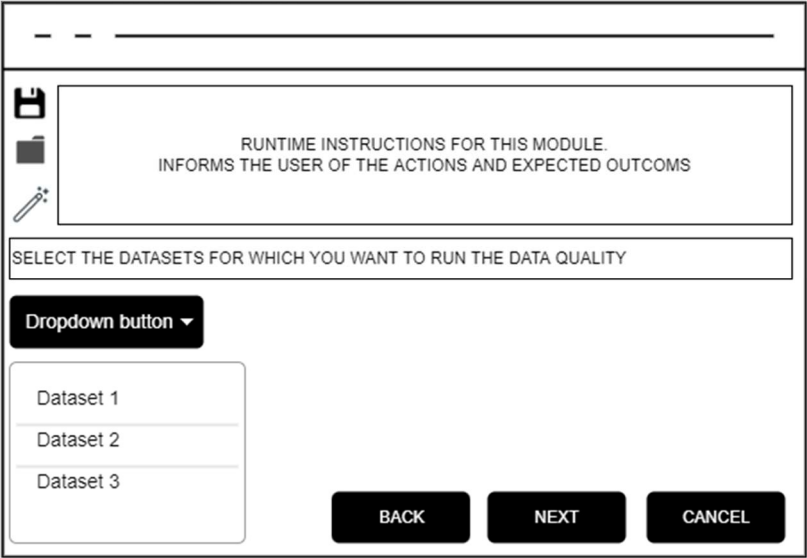
Figure 31 Data quality page mockup

4.2.4 MSM model design pseudocode and example

This section describes how the MSM model is implemented. Recall that, in IFRS 9, the probability of default (or credit risk parameter) applies the forward-looking requirement of the standard. Put simply, the standard demands an answer to the following question:

Given a facility's current segment and stage, what is the probability of it going into default in the next period?

Using a simple example, this section outlines how the MSM model was implemented for this use case.

The MSM modeling approach used applies the bank's historical default data 5 years back, grouped by segment and stage. The segment refers to the sector – Agriculture, Telecom, Consumer, Business, Islamic etc.

When using MSM for PD modeling, the states a facility can be in at any point in time are the stages and transitions are movements between states. An additional absorbing state is defined called EXIT to capture facilities that have been fully paid off or written off. The other states (i.e., stages) are informed by the days-past-due field as shown below:

i. Stage 1 – Performing (DPD <=30)
ii. Stage 2 – Impaired (DPD > 30 AND DPD < 90)
iii. Stage 3 – Defaulted (DPD > 90)
iv. EXIT – Facility has been paid off or written off

The goal of the model is to compute the likelihood of a facility in segment X moving from its current stage to stage 3 default. To better explain this let's take a Consumer Loan currently in stage 1. Say we have historical loan listing data from Jan 2015 to date. The first step is to bucket the data based on stage and segment. So, we take all Consumer loans in stage 1 in Jan 2015. We then determine the next month's (Feb 2015) probability transition matrix.

If, out of the total KES X: A remained in stage 1, B moved to stage 2, C moved to stage 3 and D exited the book in Feb 2015, then the PD between Jan 2015 and Feb 2015 is C/X. We repeat this process to obtain the marginal transitions and marginal probability between months which is an unconditional probability. We then chain these transition matrices using MMULT to obtain the cumulative probability.

Let then say our facility had an amortization schedule of 1 year based on its maturity date. We obtain the point-in-time exposures during the one-year period. To then get the point-in-time ECL:

(Point-in-time EAD) * (Point-in-time PD) * LGD

The facility's total ECL is a sum of all the point-in-time ECLs.

```
For each segment:
    For each stage (1, 2, 3)
    Find the earliest occurrence of the segment
    For each month:
        Find the value of loans in that stage
        Track the loans (by value) in the next month
        Calculate the cummulative probability using a matrix multiplication of:
            month(n) * month(n+1)

    Calculate the marginal probability of default as:
        cummulative probability of month(n+1) - cummulative probability(month)
```

Figure 32 Probability of default pseudocode

4.3 Implementation

This section describes the output of the systems implementation process as described in section 3.3 System implementation approach.

4.3.1 Workflow Implementation

This section shows the implementation of the system workflows that run the 3 modules – data input, data and impairment. The workflows were implemented primarily on the Alteryx Platform. Alteryx was chosen as the platform of choice because it is a low-code application development platform meaning it provides a user-friendly way of building data analytics applications. Note that all workflows cannot be presented in this section. For a full reference consult the project folder accompanying this report.

The first screenshot shows the main data input workflow. The workflow connects to the source database, maps the fields needed for modeling, renames them, timestamps the dataset and loads it onto the system database:
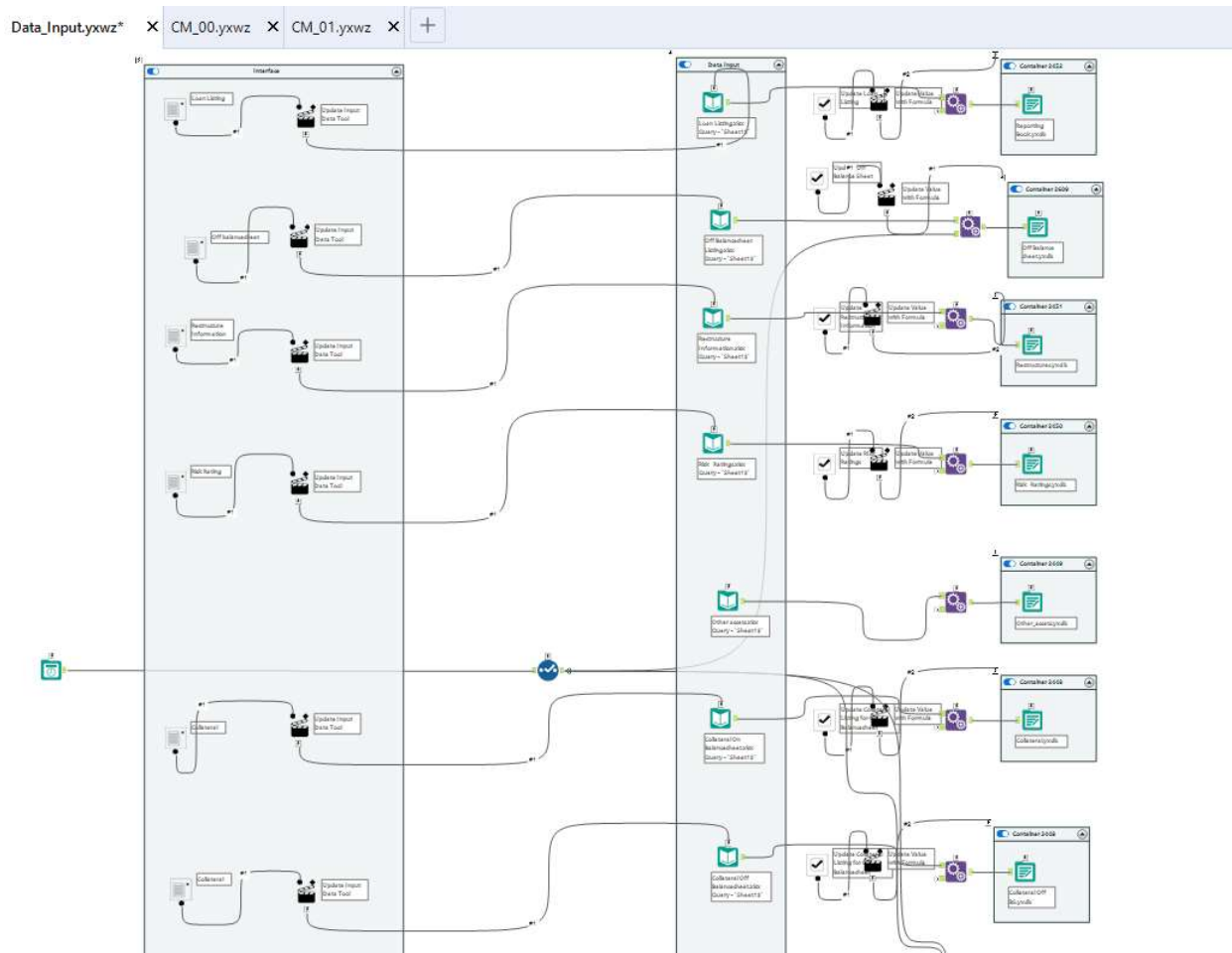
Figure 33 Alteryx data input workflow

The next screenshot shows the primary workflow for performing data quality. It loads data from the system database. Performs standard data quality and sanity checks on key field needed for modeling. Outputs the timestamped results of the data quality check to the database for feeding into the dashboard.
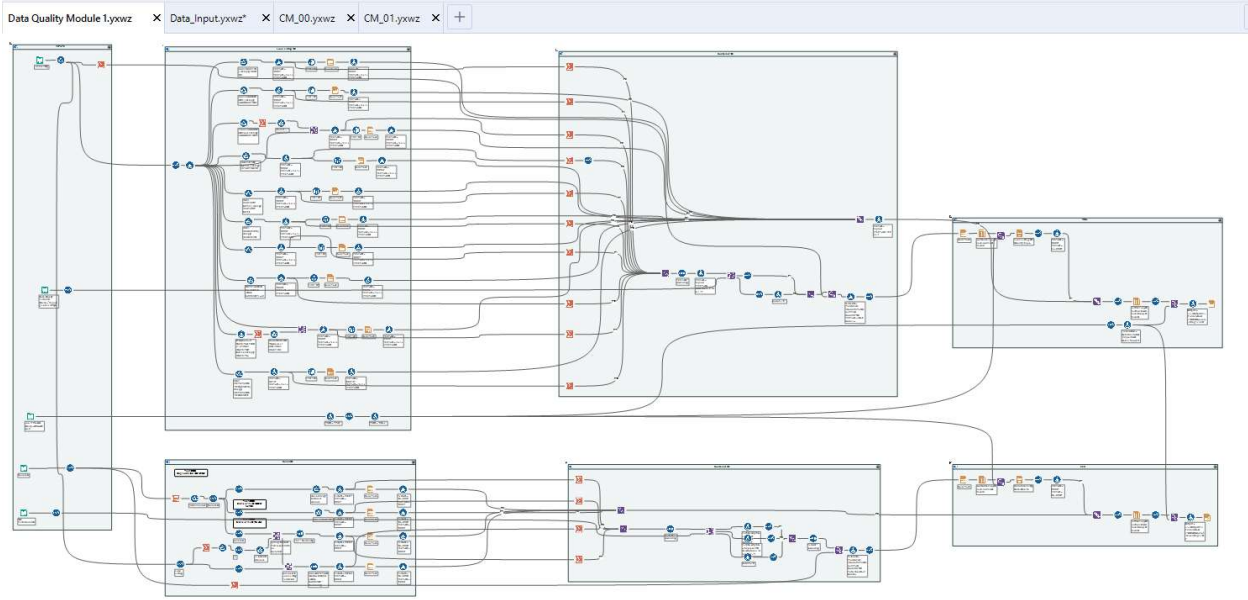
Figure 34 Alteryx data quality workflow

The next screenshot shows the impairment module workflows. Three workflows were implemented – one for PD modeling that has the R-script loaded, another for LGD computation and the final one for EAD and ECL computation. The workflows are shown in the following screenshots:
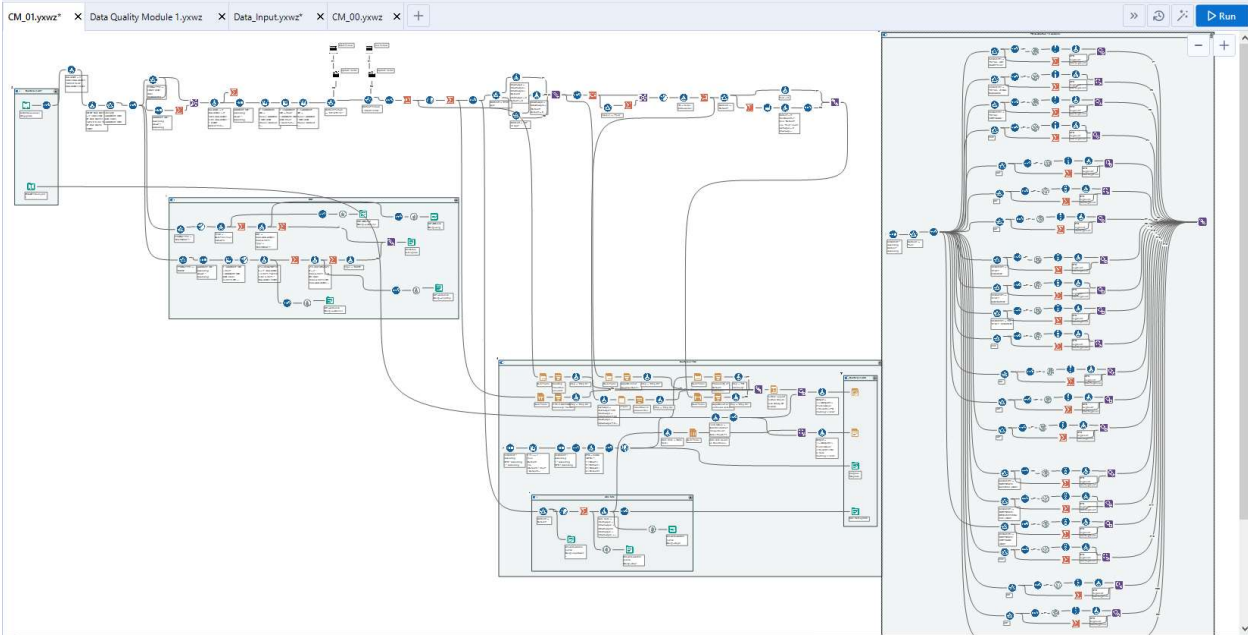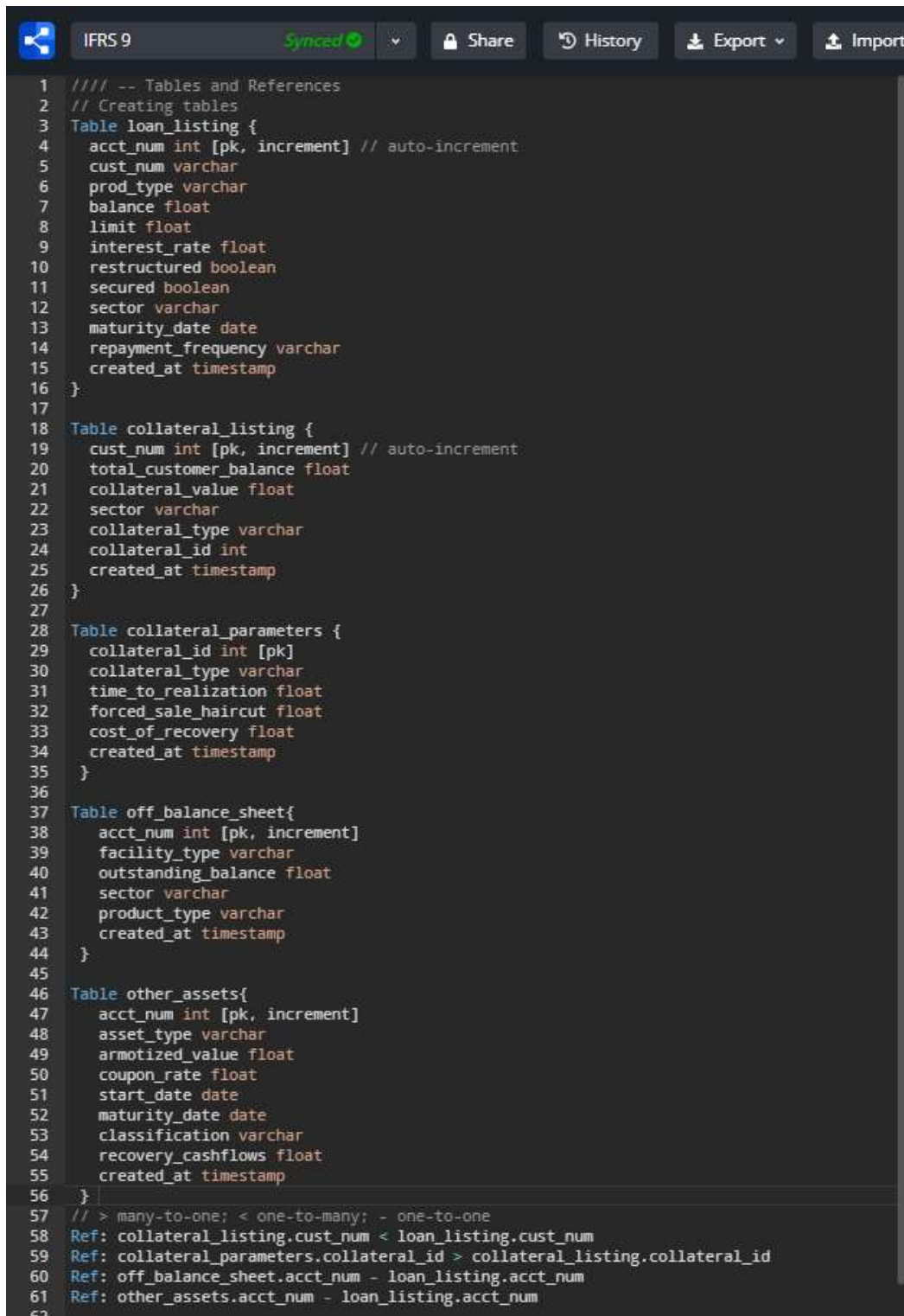
Figure 35 Alteryx PD workflow

Figure 36 Alteryx LGD workflow

Figure 37 Alteryx EAD ECL workflow

4.3.2 Database implementation

The database was implemented on a local PostgreSQL installation running on Windows. The DDL (Data Definition Language) query was generated from dbdiagram.io (a tool for designing database

schemas). DB diagram uses a special DDL syntax for defining schemas and relationships as shown below:


```
////  -- Tables and References
// Creating tables
Table loan_listing {
  acct_num int [pk, increment] // auto-increment
  cust_num varchar
  prod_type varchar
  balance float
  limit float
  interest_rate float
  restructured boolean
  secured boolean
  sector varchar
  maturity_date date
  repayment_frequency varchar
  created_at timestamp
}

Table collateral_listing {
  cust_num int [pk, increment] // auto-increment
  total_customer_balance float
  collateral_value float
  sector varchar
  collateral_type varchar
  collateral_id int
  created_at timestamp
}

Table collateral_parameters {
  collateral_id int [pk]
  collateral_type varchar
  time_to_realization float
  forced_sale_haircut float
  cost_of_recovery float
  created_at timestamp
}

Table off_balance_sheet{
  acct_num int [pk, increment]
  facility_type varchar
  outstanding_balance float
  sector varchar
  product_type varchar
  created_at timestamp
}

Table other_assets{
  acct_num int [pk, increment]
  asset_type varchar
  armotized_value float
  coupon_rate float
  start_date date
  maturity_date date
  classification varchar
  recovery_cashflows float
  created_at timestamp
}
// > many-to-one; < one-to-many; - one-to-one
Ref: collateral_listing.cust_num < loan_listing.cust_num
Ref: collateral_parameters.collateral_id > collateral_listing.collateral_id
Ref: off_balance_sheet.acct_num - loan_listing.acct_num
Ref: other_assets.acct_num - loan_listing.acct_num
```

Figure 38 DDL on dbdiagram.io

58

4.3.3 UI implementation

The user interface components were designed and built using the Alteryx interface designer. Below are some of the screenshots of the UI with a focus on the three system modules – data input, data quality and impairment. The first screenshot shows the configuration input interface where the user specifies the reporting period and the reporting entity.
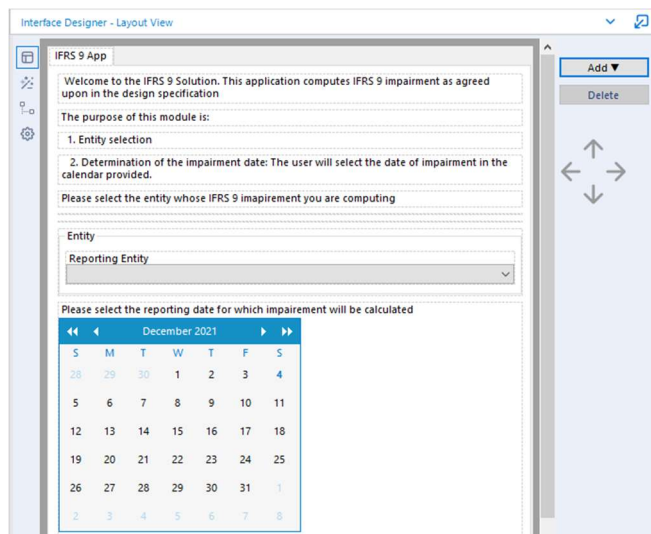
Figure 39 UI configuration input

The next screenshot shows the connector module where the user selects the execution path to follow:
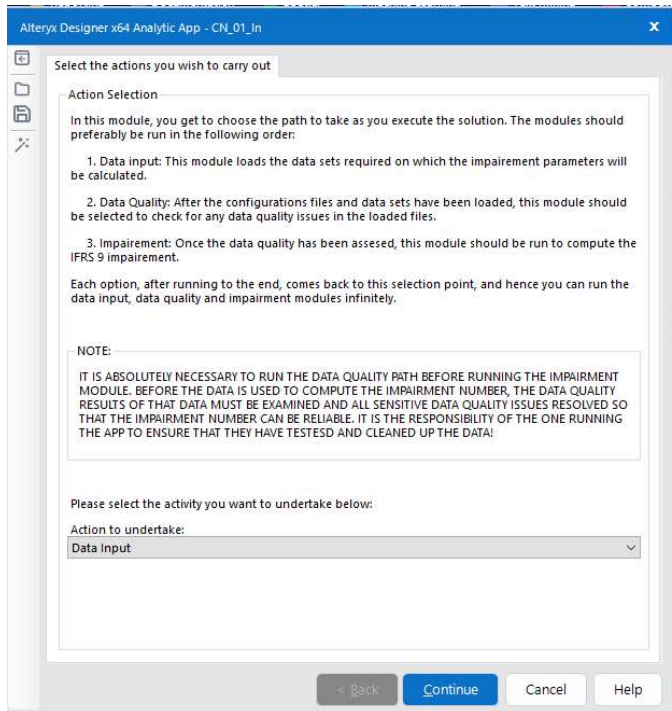
Figure 40 UI connector module

The next screenshot shows the data input UI where the user can select the data source for uploading fresh datasets onto the system database.
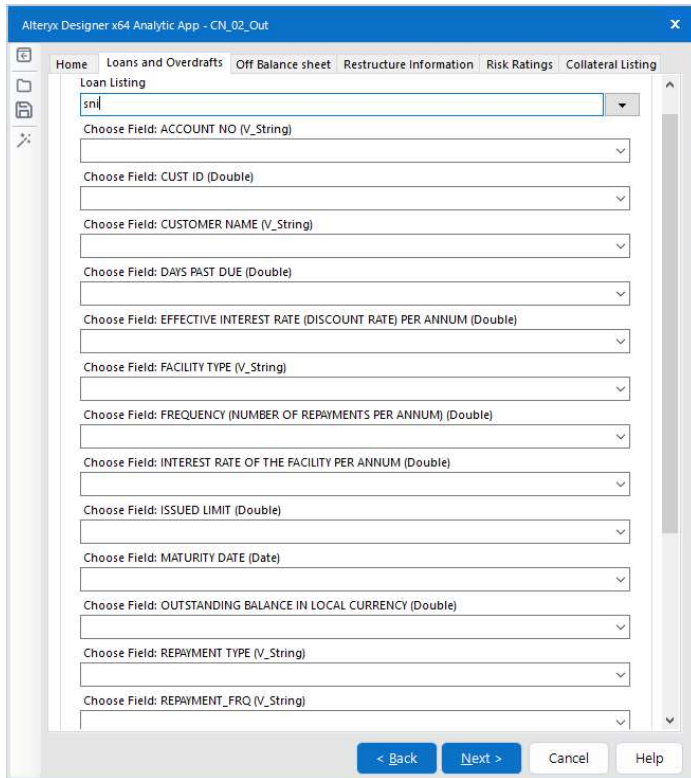
Figure 41 UI data input

The next screenshot shows the data quality UI where the user can trigger a data quality run on uploaded datasets.
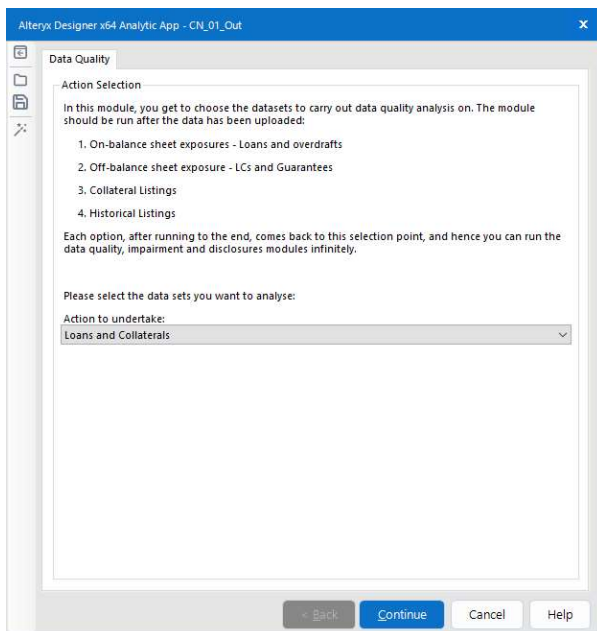
Figure 42 UI data quality

The next screenshot shows the impairment module where the user triggers an impairment computation. This runs the PD, LGD, EAD and finally ECL workflows shown in section 4.3.2 Workflow implementation.
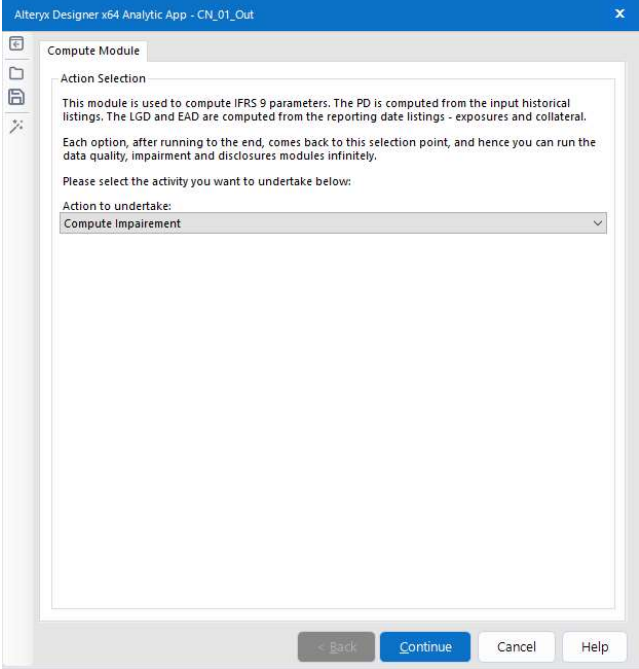
Figure 43 UI impairment

4.3.4 Dashboard Implementation

The system was connected to a PowerBI instance where results from the data quality and impairment modules automatically update the visuals on the dashboard. The dashboard is split into three main sections. The credit overview section provides a breakdown of the ECL by stage and by segment/category. It also computes the loss ratio parameter which is a percentage of the ECL against the balance. The second section provides an analysis of the PD and LGD parameters such as the average PD and LGD by segment, top uncovered exposures, and exposure flow by stage. The final section of the dashboard compares prior results to the generated results. It contains analysis such as top exposure and ECL changes and loss ratio changes by segment.

The dashboard is a value-add for the user because it helps to uncover the key drivers of the ECL parameter. It also helps the user understand the key accounts and key segments in their loan book that are driving the ECL. As such management evaluation of preliminary results is faster with a quicker turn-around time. Below are screenshots of the three dashboard sections from PowerBI.

Figure 44 Dashboard credit overview
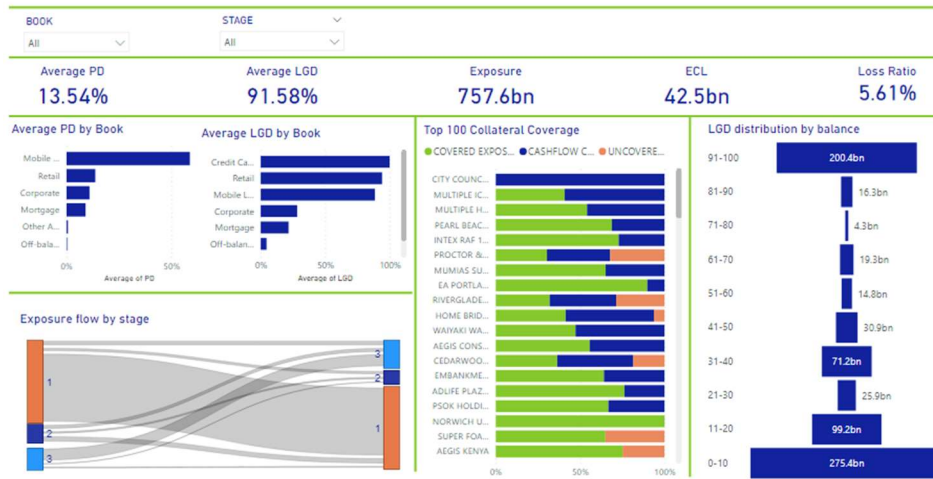
Figure 45 Dashboard PD and LGD analysis

Figure 46 Dashboard comparison

4.3.5 MSM model implementation

The first implementation was of the MSM model described in section 3.3.1. The model was implemented on R using Alteryx-R tools. Below is a screenshot of the core of the R script used to perform the matrix transitions and calculate sector-level probabilities of default.

```
Insert Code ▾  ☐ Run script when refreshed (F5)   More About R

original <- read.Alteryx("#1", mode="data.frame")
matrix.names <- names(original)
original <- matrix(as.numeric(unlist(original)),
nrow=nrow(original),ncol=ncol(original),byrow=FALSE)

lifetime.pd <- function(t,original){
    mmult.result <- original
    final.result <- original
    months <- c(1,1,1,1)
    for(i in 2:t){
      mmult.result <- original%*%mmult.result
      final.result <- rbind(final.result,mmult.result)

      months <- c(months,c(i,i,i,i))
    }
    print(length(months))
    final.result <- cbind(final.result,months)

    return(as.data.frame(final.result))
}

res <- lifetime.pd(300,original)
names(res) <- c(matrix.names,"T")
write.Alteryx(res, 2)|
```

Figure 47 MSM model implementation

CHAPTER 5: TESTING AND EVALUATION

This chapter outlines the output of the testing and evaluation approaches as described in the final sections of the methodology chapter.

5.1 System testing

This section outlines the results of the system testing. For testing the sector level PD and ECL outputs from the new system were compared with what the bank reported. An output is only valid if it varies by at most 2%. Below is the output of PD comparison:

Sector	Alteryx PD	Reported PD	Abs Difference
Business: Business Loan	14.74%	14.30%	0.44%
Business: Consumer/Scheme Loan	0.00%	0.30%	0.30%
Business: Mortgage Loan	11.96%	13.00%	1.04%
Corporate: Asset Finance	0.00%	0.00%	0.00%
Corporate: Business Loan	13.04%	14.20%	1.16%
Corporate: Consumer/Scheme Loan	8.41%	7.95%	0.46%
Corporate: Mortgage Loan	25.92%	25.83%	0.09%
Islamic: Islamic Corporate	25.15%	25.14%	0.01%
Islamic: Islamic Retail	20.35%	20.34%	0.01%
Retail: ABF (Chartels)	16.13%	16.10%	0.03%
Retail: Mortgage	22.14%	22.13%	0.01%
Retail: Secured Facilities	0.14%	0.18%	0.04%
Staff	39.26%	39.17%	0.09%

Table 10 PD Testing

Below is the output of the ECL comparison:

IFRS 9 Category	Alteryx ECL	Reported ECL	Abs Difference
ABF Corporate	115,576,229.52	115,407,371.36	0.15%
ABF Retail	1,052,563,740.45	1,050,632,779.46	0.23%
Corporate Non-Revolving	1,534,160,862.21	1,531,148,852.08	0.20%
Corporate ODs	289,223,124.10	288,779,514.95	0.19%
Credit Cards	337,684,821.82	337,137,769.42	0.45%
Ex Staff Mortgage	101,356,549.47	101,172,520.25	0.18%
Guarantees	1,004,840,727.56	1,003,492,335.58	0.13%
Mobile Fuliza	526,558,742.45	525,732,681.98	0.16%
Mobile MKCB	5,592,512,664.72	5,585,098,955.79	0.13%
Mobile Mobi	308,996,768.70	308,418,698.45	0.19%
Mortgage Check Off	1,216,057.61	1,214,760.98	0.11%

Mortgage Commercial	334,901,505.28	334,259,956.19	0.19%
Mortgage Residential	476,606,606.31	476,094,857.46	0.11%
Other assets	2,588,198,030.99	2,584,642,535.50	0.14%
Retail Checkoff	5,581,617,440.76	5,572,003,620.46	0.17%
Retail Non-Checkoff	2,088,599,817.70	2,086,159,455.53	0.12%
Retail ODs	857,945,563.98	856,269,054.84	0.20%
Retail Secured	661,665,001.82	660,820,024.58	0.13%
Staff Mortgage	17,414,598.06	17,383,882.02	0.18%
Top 100	11,055,049,807.17	11,036,658,681.34	0.17%
Unfunded LCs	4,793,426.51	4,784,738.48	0.18%

Table 11 ECL testing

All outputs were consistent with the bank's reported figures. The ECL for all sectors fell within the allowable 2% variance.

5.2 System usability testing

The system usability was evaluated using the System Usability Measurement Inventory (SUMI). Questions were asked around 4 areas – efficiency, affect, learnability and control

Efficiency questions:

1. The instructions and prompts are helpful.
2. I sometimes don't know what to do next with this software.
3. This software has at some time stopped unexpectedly.
4. Getting data files in and out of the system is not easy.
5. Learning to operate this software initially is full of problems.

Figure 48 SUMI Efficiency Analysis

Based on the results from the respondents a majority of the respondents found the system efficient with 80% indicating that instructions provided are helpful in achieving the system objectives. 60% of the respondents said that they know what to do next throughout the system's execution. However, 75% of the respondents reported that the system has at some point stopped unexpectedly.

Affect questions:

1. I would recommend this software to my colleagues.
2. It is obvious that user needs have been fully taken into consideration.
3. There are too many steps required to get something to work.
4. This software responds too slowly to inputs.
5. The way that system information is presented is clear and understandable.
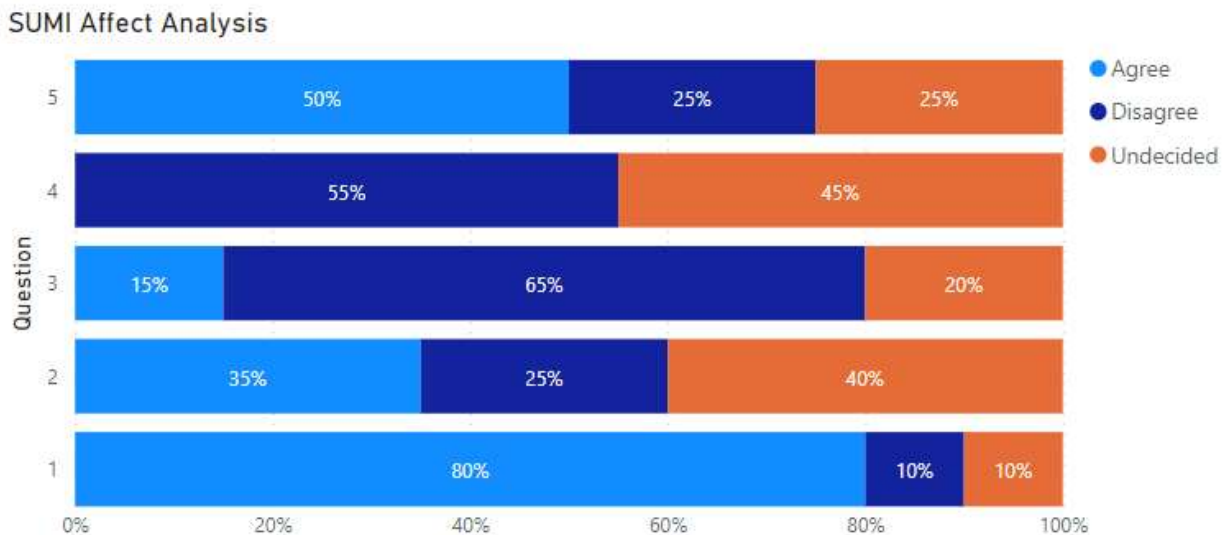
Figure 49 SUMI Affect Analysis

The respondents showed a generally positive outlook when using the system. 80% of the respondents indicated that they would recommend the system to other similar users. There were questions on the system speed with 45% of the respondents were undecided on the system's response time being too slow. Additionally, 50% of the respondents agreed that the system design and implementation take into consideration user needs.

Learnability questions:

67

1. It takes too long to learn the software functions.
2. I feel in command of this software when I am using it.
3. Tasks can be performed in a straight forward manner using this software.
4. Learning how to use new functions is not difficult.
5. I will never learn to use all that is offered in this software.

Figure 50 SUMI Learnability Analysis

From the responses, it is evident that a new user of the system can get quickly up to speed and produce the required output. 85% of the respondents disagreed that it takes too long to learn system functions. None of the respondents indicated that they will never learn to use all of the system's functionalities. 75% of the respondents recorded that system tasks can be performed in a straight forward manner using the proposed system.

Control questions:

1. I feel in command of this software when I am using it.
2. I prefer to stick to the functions that I know best.
3. I keep having to go back to look at the guides.
4. It is easy to make the software do exactly what you want.
5. This software occasionally behaves in a way which can't be understood.

SUMI Control Analysis
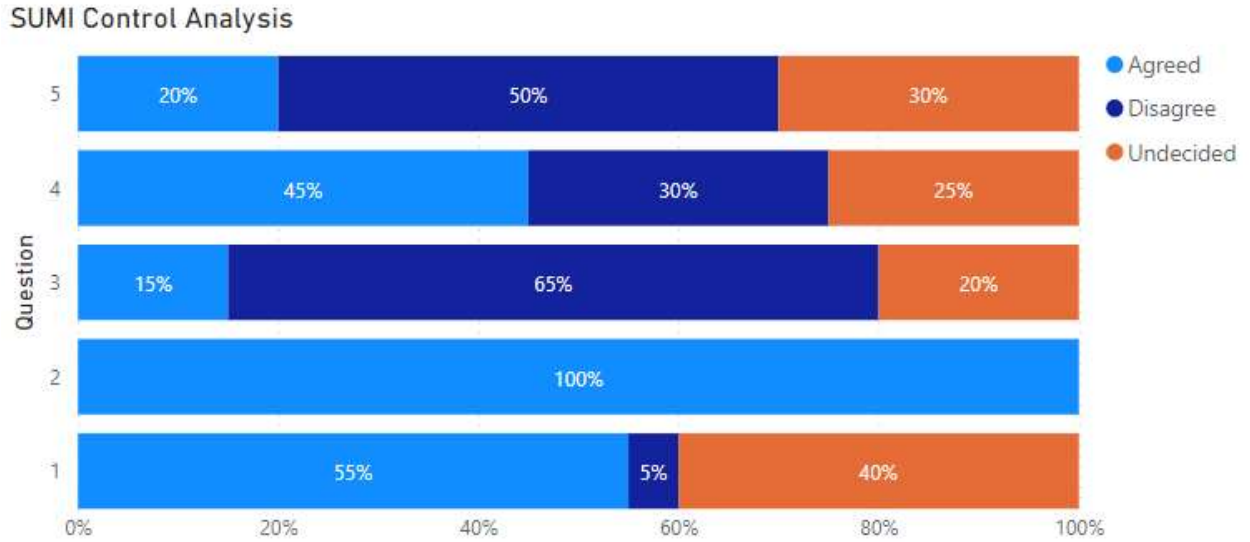
Figure 51 SUMI Control Analysis

The responses show that users of the system felt in control when using the system. It was easy to recover from errors and exceptions during system execution. 55% of the respondents said that they felt in command of the system when using. All respondents indicated that it is easy to make the software do exactly what they want and 65% said that they do not have to keep referring to guides or ask for help when using the system.

5.3 System evaluation

As discussed in section 3.6, the primary evaluation metric chosen was efficiency which is measures as the answer to the following question:

> ***How much time in minutes does it take from the start of the modeling process to when a valid output is availed to end-users?***

The goal is to determine whether the solution addresses the problem of time-consuming modeling and reporting process.

From the analysis phase, the current modeling process has the following key steps:

1. Data request
2. Data quality
3. Data input
4. Running impairment

5. Summarizing results

6. Management review and signoff

The new system cuts off the first three steps which would take 1-2 days. There is no need for a data request because the new system has an external interface to existing database sources. Additionally, the data quality checks have already been predefined and run automatically after data input.The computation of numbers on the excel model takes ~20 mins to run the numbers and summarize the results. In comparison, the new system takes ~5 seconds to run all impairment parameters and summarized results instantly reflect on the dashboard through the preconfigured interface. See below a typical runtime for the final ECL output:

Figure 52 Alteryx ECL runtime

The new system also addresses errors in the data input by automating the process of performing data quality checks process. With every data update, a corresponding data quality check is run automatically and the data quality dashboard updated.
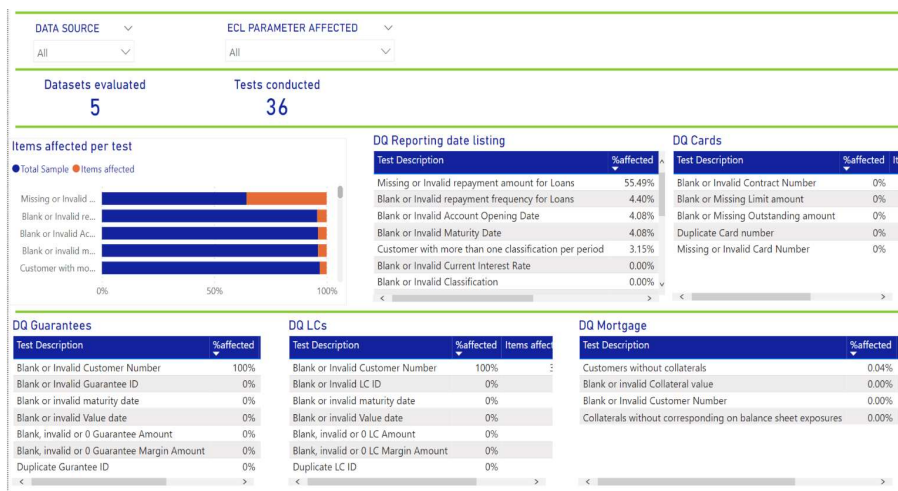
Figure 53 Data quality dashboard

CHAPTER 6: DISCUSSION, CONCLUSION AND RECOMMENDATIONS

6.1 Discussion

Testing was done to ensure accuracy of the MSM PD model and the ECL outputs. Accuracy was benchmarked against the reported numbers as reference. Evaluation was done to appraise the system as whole by presenting it to potential users and obtaining feedback on the system's usability metrics.

Results from the tests were consistent with the number reported by the bank with an allowable variance of 2%. The bank chosen for modeling reported an ECL of KES 34.47 billion from their original model. Had they used the proposed system, they would have reported KES 34.53 billion as their expected credit loss. This falls within the 2% variance – 0.0013%. The sector with the largest ECL variance was Credit Cards at 0.45%.

The PD, sector-level comparison was also done and showed that the PD for each sector computed also fell within the stipulated variance. The sector with the largest variance was Corporate: Business Loan with a difference of 1.16% between the reported and computed PD. This indicates that the multi-state Markov approach adopted for the solution mirrors the bank previous modeling approach and can be used for future reporting periods.

On usability, the system received positive feedback from potential users under the Standard Usability Measurement Inventory (SUMI). System usability was evaluated on four areas – efficiency, learnability, affect and control. Usability of such a system is key because it intends to abstract the complexities of spreadsheet models which require detailed knowledge and workings of the model to prevent breakage and erroneous output. For instance, 80% of the users reported that they would recommend the system to others and 85% responded positively on the system's efficiency questions.

The system was also evaluated against the problem statement to show that it addresses the issue of time-consuming modeling and reporting cycles. The runtime of the system after data input was ~5 seconds on average to get a summarized output on the dashboard. On the excel approach running the models and summarizing the output took ~20 mins. This shows that the system improves the efficiency of the current process.

6.2 Conclusion

The overall goal of this project was to investigate the challenges facing Kenyan banks in IFRS 9 adoption. Design, develop, test and evaluate a solution that addresses the challenges gathered in systems analysis. The table below shows how all specific objective have been met by matching and objective to its output:

Objective	Output
To investigate the challenges in IFRS 9 modeling and IFRS 9 reporting for banks in Kenya.	System requirements specification based on the output of the system analysis step
To design an automated financial modeling and reporting solution compliant to the IFRS 9 standard and suitable for banks in Kenya.	System design specifications – workflow design, process design, interface design and database design
To develop an automated financial modeling and reporting solution compliant to the IFRS 9 standard and suitable for banks in Kenya.	System modules in terms of Alteryx workflows, PostgreSQL database and PowerBI dashboard
To test and evaluate the automated analytics solution.	System testing and evaluation results.

The project revealed the challenges of the current financial modeling and reporting approaches in use by banks in Kenya. It showed that a majority of the banks still rely on enterprise spreadsheet models to fulfill their annual financial reporting duties. The disadvantages of this approach are evident in literature and from the responses during requirements gathering. Enterprise spreadsheet models are often complex, not integrated, error-prone, time consuming, and can lead to inaccurate results.

The system developed on a low-code platform Alteryx automates the tedious and repetitive tasks of the financial modeling process such as data input, data validation and data quality checks. This frees up the user to focus on model design and results interpretation. The use of a low-code platform ensures that learning curve for any design alterations remains low and the system is flexible to changes. Low-code automation technologies abstract low-level implementation details

allowing the user to focus on the application's logic. They are part of efforts to democratize technology thus allowing stakeholders to be involved in the implementation or modification of applications.

6.3 Recommendations

For future work the approach can be extended for other financial reporting standards such as IFRS 7 and 16 and bundled as a suite. Additionally, the logic of implementation can be ported to an open-source low-code development platform such as KNIME to further reduce development and deployment costs. Future implementations should allow users to update key model design value from the user interface. This applies to the LGD model in particular. Values such as the haircut applied and time to realization for each collateral type. This will further increase design flexibility since these parameters are subject to change.

REFERENCES

Albers, M. J. (2017). Quantitative Data Analysis—In the Graduate Curriculum. *Journal of Technical Writing and Communication*, 78-90.

Allen, L., Delong, G., & Saunders, A. (2003). Issues in the Credit Risk Modeling of Retail Markets. *SSRN Electronic Journal*.

Alteryx Analytics Platform. (2021, January 31). *From Data to Discoveries to Decisions — In Minutes*. Retrieved from Alteryx: https://www.alteryx.com/

Amoako, G. K. (2010). *COMPLIANCE WITH INTERNATIONAL FINANCIAL REPORTING STANDARD 7 (IFRS 7): A STUDY OF LISTED BANKS IN GHANA.* Accra: University of Cape Coast.

Berruti, F., Ross, E., & Weinberg, A. (2017). *The transformative power of automation in banking.* McKinsey Inc.

Capterra. (2022, January 24). *IMEX.* Retrieved from Capterra - Helping businesses choose software since 1999: https://www.capterra.com/p/2103/IMEX/

Carmassi, J., Gros, D., & Micossi, S. (2009). The Global Financial Crisis: Causes and Cures. *Journal of Common Market Studies*, 159-196.

Chee, C. (2021, May 19). *Probability of Default (PD) using Machine Learning (ML) Techniques.* Retrieved from LinkedIn: https://www.linkedin.com/pulse/probability-default-pd-using-machine-learning-ml-techniques-chee/

Deloitte Touche Tohmatsu Limited. (2016). *A Drain on Resources? The Impact of IFRS 9 on Banking.* London: Deloitte Touche Tohmatsu Limited.

FinAccess. (2016). *The 2016 Household Financial Access Surey.* Nairobi: Central Bank fo Kenya.

Gartner Inc. (2021). *Gartner 2021 Magic Quadrant for Data Science and Machine Learning Platforms.* Gartner Inc.

Global, S. (2017, May 23). *INTRO TO CREDIT RATINGS: A Credit Rating is an Informed Opinion.* Retrieved from S&P Global Ratings: https://www.spglobal.com/ratings/en/about/intro-to-credit-ratings

Grand Review Research. (2021). *Robotic Process Automation Market Growth & Trends.* San Fransisco: Grand View Research, Inc.

Ho, T. S., & Sang, B. L. (2004). *The Oxford Guide to financial modeling: Applications for capital markets, corporate finance, risk management and financial institutions.* Oxford: Oxford university press.

IBM Corporation. (2021). *The buyer's guide to intelligent business automation from IBM.* 2021: IBM Corporation.

ICPAK. (2018). *Enhancing IFRS 9 Compliance for the Banking Sector in Kenya.* Nairobi: ICPAK.

Jurek, K., & Corbett, M. (1993). The software usability measurement inventory. *British journal of educational technology, 24(3)*, 210-212.

Kereri, P. (2018). Financial Statement Fraud. *International Journal of Social Sciences and Information Technology*, 702-723.

KNIME Analytics Platform. (2021, January 31). *End to End Data Science*. Retrieved from KNIME - Open for Innovation: https://www.knime.com/

KPMG Kenya. (2016). *A new beginning: are banks ready for IFRS 9?* Nairobi: KPMG.

Krayn, A., Oleksza, P., Canamero, M., & Gea-Carrasco, C. (2015, May 10). *IFRS 9 Will Significantly Impact Banks' Provisions and Financial Statements.* Retrieved from MOODY'S ANALYTICS: https://www.moodysanalytics.com/risk-perspectives-magazine/risk-data-management/regulatory-spotlight/ifrs-9-will-significantly-impact-banks-information

Krejcie, R. V., & Daryle, M. W. (1970). Determining sample size for research activities. *Educational and psychological measurement 30(3)*, 607-610.

Kumar, N., & Balaramachandran, R. (2018). ROBOTIC PROCESS AUTOMATION - A STUDY OF THE IMPACT ON CUSTOMER EXPERIENCE IN RETAIL BANKING INDUSTRY. *Journal of Internet Banking and Commerce*, 50-62.

Lin, D., & Wei, L. (1989). The robust inference for the Cox proportional hazards model. Journal of the American statistical Association. *Journal of the American statistical Association*, 1074-1078.

Marcus, W. (2020). The Rise of No/Low Code Software Development—No Experience Needed? *Elsevier Public Health Emergency Collection*, 960-961.

Meira-Machado, Una-Alvarez, Carso-Suarez, & Andersen. (2009). Multi-state models for the analysis of time-to-event data. *Stat Methods Med Res.*, 195-222.

Mixson, E. (2021). *Low-Code Automation and the Democratization of Innovation.* Intelligent Automation Network.

Mugenda, O., & Mugenda, A. (1999). Research methods: Quantitative and qualitative approaches. *Acts Press*.

Nairobi Securities Exchange. (2021, January 1). *Listed Companies.* Retrieved from NSE Discover Opportunity: https://www.nse.co.ke/listed-companies/list.html

Opara-Martins, J., Sahandi, R., & Tian, F. (2016). Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective. *Journal of Cloud Computing*.

Oracle Financial Services. (2016). *Impairment Principles and Approach: Building an Expected Loss Model.* San Fransisco: Oracle Financial Services.

Owens, L. K. (2002). Introduction to survey research design. In. *SRL fall 2002 seminar series (Vol. 1).*

Panko, R. R., & Sprague, R. H. (1998). Hitting the wall: errors in developing and code inspecting a `simple' spreadsheet model. *Decision Support Systems*, 337-353.

Powell, S. G., Baker, K. R., & Lawson, B. (2008). A critical review of the literature on spreadsheet errors. *Decision Support Systems*, 128-138.

Radigan, D. (2020, January 1). *How the kanban methodology applies to software development.* Retrieved from Atlassian: https://www.atlassian.com/agile/kanban

Reid, L. G., & Moats, L. C. (1997). Critical conceptual and methodological considerations in reading intervention research. *Journal of learning disabilities, 30(6)*, 578-588.

S&P. (2017, May 23). *A Credit Rating is an Informed Opinion.* Retrieved from S&P Global Ratings: https://www.spglobal.com/ratings/en/about/intro-to-credit-ratings

Somayya, M. M., HolmukheDurgesh, K., & Jaiswal. (2019). The Future Digital Work Force: Robotic Process Automation (RPA). *Journal of Information Systems and Technology Management*, 90-102.

Spall, C. (2018). *Financial instruments – Introducing IFRS 9.* London: KPMG in the UK.

Suphat, S. (2007). Fundamentals of quantitative research. *Language Institute Chulalongkorn University, 1(3),*, 1-20.

Trobia, A., & Lavrakas, P. J. (2011). Encyclopedia of survey research methods. *Encyclopedia of Survey Research Methods. first ed. Sage Publications, Inc*, 169-170.

Tudela, M., & Young, G. (2003). *A Merton-model approach to assessing the default risk of UK public.* London: Bank of England.

Vincent, P., Wong, J., Ray, S., Jain, A., Guttridge, K., Ilijima, K., . . . Natis, Y. (2021). *Identify and Evaluate Your Next Low-Code Development Technologies.* Gartner Inc.

Volarevic, H., & Varovic, M. (2018). INTERNAL MODEL FOR IFRS 9 - EXPECTED CREDIT LOSSES CALCULATION. *EKONOMSKI PREGLED*, 269-297.

Wisniewski, M. (2017, December 4). *Betting on Alexa to change financial services.* Retrieved from Financial Planning: https://www.financial-planning.com/news/banks-add-alexa-skills-to-develop-voice-command-services

Wong, P. (2004). *Challenges and Successes in Implementing International Standards: Achieving Convergence to IFRSs and IASs.* New York: International Federation of Accountants.

APPENDIX

APPENDIX I: Systems analysis questionnaire

Respondent demographics

    i.   The type of bank – Local or International

    ii.   The job function of the respondent

    iii.   The largest segment in the bank's portfolio

    iv.   The percentage increase of reporting under IFRS 9

Knowledge of the standard

    i.   Are you directly involved in the IFRS 9 reporting processes within your bank?

    ii.   On a scale of 1-10 rate your understanding of the IFRS 9 standard.

    iii.   On a scale of 1-10, rate your proficiency in the following – Excel, Statistical modeling, Analytics and visualization

    iv.   How many certified IFRS 9 personnel are available within your bank?

    v.   What is your preferred mode of adoption between outsourcing and internal implementation?

    vi.   Do you think your bank has set aside enough resources to either outsource implementation or develop in-house capacity?

    vii.   What is the most challenging IFRS 9 parameter to model given the standard requirements and the data available at your bank?

Modeling approaches in use

    i.   On what platform do you currently model IFRS 9 parameters?

    ii.   Do you have an internal risk rating system in place?

    iii.   Do you have integrated analytics with the current modeling solution?

    iv.   How far back is historical loan listings data available for the bank?

    v.   On average, how many personnel are involved in the modeling and reporting process?

    vi.   What is the average turn-around time in days from modeling to final reporting?

    vii.   On average, how many audit queries are raised on the provisions from the bank per year?

    viii.   What are the most difficult elements in implementing an IFRS 9 controls framework?

ix.   On a scale of 1-10, how important are the following elements in a potential IFRS 9 solution: deployment costs, deployment time, automation, accuracy, integration, flexibility, analytics

APPENDIX II: Datasets Used

This section provides an anonymized snapshot of some of the datasets provided by KCB bank used for modeling. For anonymization, all customer names have been removed and a random, incremental RecordID field assigned to represent account numbers in all instances. Note that it is not possible to represent the full datasets because of sheer size. For instance, the historical loan listings used exceed the allowable excel limit as shown below:
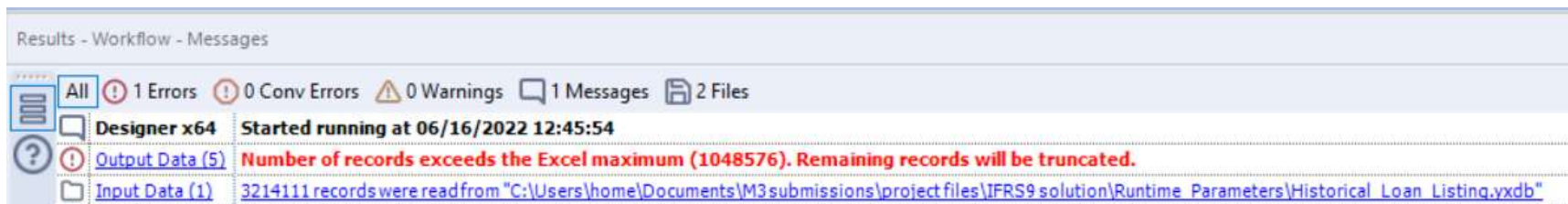
For a full list of datasets used, refer to the **"Project Files"** folder accompanying this report under **"IFRS9 Solution/Runtime Parameters"** or request for the modeling data used for the December 2020 period from KCB.

Loan listings

Key field: Account number, Sector, Product type, Balance, Effective Interest Rate, Days Past Due, Repayment Frequency, Repayment Amount, Maturity Date, Secure Flag. The same schema is adopted for the historical loan listings used to model PD.

ACCT_NUM	SECURED	PROD_TYPE	SECTOR	ISSUED_LIMIT	BALANCE	EIR	REPAYMENT FREQUENCY
22	YES	LOAN	OTHERS_NPL	0	6,880,072.17	0.00%	MONTHLY
24	YES	LOAN	TOP_NPL	0	87,545,795.06	0.00%	MONTHLY
26	YES	LOAN	ISLAMIC: ISLAMIC CORPORATE	0	1,790,876.72	8.50%	MONTHLY
42	YES	LOAN	OTHERS_NPL	0		0.00%	MONTHLY

1

					302,353.00		
43	YES	LOAN	OTHERS_NPL	0	964,163.82	0.00%	MONTHLY
44	YES	LOAN	OTHERS_NPL	0	940,435.35	0.00%	MONTHLY
67	YES	LOAN	ISLAMIC: ISLAMIC RETAIL	0	5,821,586.51	13.00 %	MONTHLY
87	YES	LOAN	ISLAMIC: ISLAMIC RETAIL	0	12,310,568.90	13.00 %	MONTHLY
89	YES	LOAN	ISLAMIC: ISLAMIC RETAIL	0	267,673.39	0.00%	MONTHLY
94	YES	LOAN	ISLAMIC: ISLAMIC CORPORATE	0	59,454,452.99	13.00 %	MONTHLY
99	YES	LOAN	ISLAMIC: ISLAMIC RETAIL	0	3,492,083.36	13.00 %	MONTHLY
4847	YES	LOAN	BUSINESS: MORTGAGE LOAN	0	9,243,113.06	13.00 %	MONTHLY
4855	YES	LOAN	RETAIL: MORTGAGE	0	7,388,457.85	13.00 %	MONTHLY
4868	YES	LOAN	RETAIL: MORTGAGE	0	39,198,567.75	13.00 %	MONTHLY
4872	YES	LOAN	RETAIL: ABF (CHARTELS)	0	472,878.28	13.00 %	TRIANNUALLY
4874	YES	LOAN	RETAIL: ABF (CHARTELS)	0	1,660,395.23	13.00 %	TRIANNUALLY
4875	YES	LOAN	RETAIL: ABF (CHARTELS)	0	4,268,226.79	13.00 %	TRIANNUALLY
5233	YES	LOAN	OTHERS_NPL	0	28,220.24	0.00%	MONTHLY
5252	YES	LOAN	TOP_NPL	0	22,117.90	13.00 %	MONTHLY
5277	YES	LOAN	OTHERS_NPL	0	2,941.83	0.00%	MONTHLY
5317	YES	LOAN	OTHERS_NPL	0		0.00%	MONTHLY

					32,337.43			
5340	YES	LOAN	RETAIL: SCHEME NONCHECKOFF		0	21,500.00	0.00%	MONTHLY
5365	YES	LOAN	OTHERS_NPL		0	11,458.87	0.00%	MONTHLY
5369	NO	LOAN	STAFF: UNSECURED		0	32,172.47	0.00%	MONTHLY

Collateral listing

It is at customer level. Key fields include the Customer Number, Collateral Value, Total Outstanding Balance, Sector and Collateral Type

CUST_NUM	TOTAL_CUSTOMER_BALANCE	COLLATERAL_VALUE	CUSTOMER_SECTOR	COLLATERAL_TYPE
1	32,039,918.08	80,000,000.00	BUSINESS: MORTGAGE LOAN	CASH COVERED
5	110,120,023.60	765,000,000.00	CORPORATE: MORTGAGE LOAN	LAND
6	42,494,664.91	90,000,000.00	CORPORATE: MORTGAGE LOAN	LAND, BUILDINGS
51	11,748,264.45	11,500,000.00	BUSINESS: MORTGAGE LOAN	VEHICLE, LAND
116	14,103,588.89	67,000,000.00	CORPORATE: MORTGAGE LOAN	CASH COVERED
150	15,380,106.30	32,000,000.00	BUSINESS: MORTGAGE LOAN	LANDED COLLATERAL
151	13,127,365.96	26,000,000.00	BUSINESS: MORTGAGE LOAN	LANDED COLLATERAL
262	16,149,979.62	30,000,000.00	CORPORATE: MORTGAGE LOAN	EXPECTED CASHFLOWS

				CASH COVERED
265	8,232,467.92	26,000,000.00	RETAIL: SECURED FACILITIES	
334	9,667,447.11	73,300,000.00	BUSINESS: CONSUMER/SCHEME LOAN	LAND
360	9,331,820.05	11,500,000.00	BUSINESS: MORTGAGE LOAN	LAND
381	42,135,069.32	219,000,000.00	CORPORATE: CONSUMER/SCHEME LOAN	LAND
390	5,819,202.87	11,500,000.00	RETAIL: SECURED FACILITIES	LAND
432	27,864,987.82	60,000,000.00	RETAIL: SECURED FACILITIES	VEHICLE, LAND, BUILDINGS
460	11,447,246.91	50,000,000.00	BUSINESS: MORTGAGE LOAN	VEHICLE, CASH COVERED

Collateral Parameters

These inform how to treat different collateral types. Key fields – Collateral ID, Collateral Type, Time to Realization, Forced Sale Haircut and Cost of Recovery. These are informed by the nature of the business and previous experience in collateral recovery

ID	Collateral Type	Time to realization	Forced Sale Haircut	Cost of recovery (% of Balance)
Collateral Type 1	Landed Properties	1	0%	4%
Collateral Type 2	Cash	0	0%	1%
Collateral Type 3	Fixed assets	0	0%	1%
Collateral Type 4	Stock	0	5%	1%
Collateral Type 5	Receivables under 60 Days	0	5%	1%
Collateral Type 6	Receivables 60 to 90 Days	0.01	5%	1%
Collateral Type 7	Receivables 91 to 120 Days	0.02	5%	1%
Collateral Type 8	Vehicle	0.5	5%	1%
Collateral Type 9	Other	0	5%	1%
Collateral Type 10	Undefined	0	5%	1%

Off balance sheet listing

Includes Customer Acceptances, Guarantees and Letters of Credit (LCs). Key fields – Account Number, Facility Type, Product Type, Sector

ACCT_NUM	FACILITY TYPE	OUTSTANDING BALANCE	SECTOR	PRODUCT TYPE
1	CUSTOMER ACCEPTANCES	29,748,264.00	ISLAMIC: ISLAMIC CORPORATE	ACCEPTANCE
2	CUSTOMER ACCEPTANCES	279,552,000.00	CORPORATE: BUSINESS LOAN	ACCEPTANCE
3	CUSTOMER ACCEPTANCES	546,000,000.00	CORPORATE: BUSINESS LOAN	ACCEPTANCE
4	GUARANTEES ISSUED - GENERA	15,524.00	RETAIL: SECURED FACILITIES	GUARANTEE
5	GUARANTEES ISSUED - GENERA	20,888.00	RETAIL: SECURED FACILITIES	GUARANTEE
6	LETTERS OF CREDIT ISSUED	21,450.00	RETAIL: SECURED FACILITIES	LC
10	GUARANTEES ISSUED - GENERA	50,000.00	RETAIL: SECURED FACILITIES	GUARANTEE
11	GUARANTEES ISSUED - GENERA	52,777.00	RETAIL: SECURED FACILITIES	GUARANTEE
12	GUARANTEES ISSUED - GENERA	54,351.00	RETAIL: SECURED FACILITIES	GUARANTEE
13	LETTERS OF CREDIT ISSUED	80,000.00	RETAIL: SECURED FACILITIES	LC
19	LETTERS OF CREDIT ISSUED	100,000.00	RETAIL: SECURED FACILITIES	LC

Other Assets

Includes treasury bills and bonds. Key fields – Account Number, Asset Type, Asset Value, Armotized Asset Value, Start Date, Maturity Date, Classification, Recovery Cashflows, Time to Realization Years

ACCT_NUM	ASSET TYPE	ASSET VALUE	ARMOTIZED ASSET VALUE	COUPON RATE	START DATE	MATURITY DATE	Classification	RECOVERY CASHFLOWS
1	Treasury Bills	10,000,000,000.00	8,931,200,000.00	13.23%	7/2/2020	7/1/2021	Sovereign	8,931,200,000
2	Treasury Bills	5,000,000,000.00	4,465,600,000.00	13.15%	9/14/2020	7/15/2021	Corporate	4,465,600,000
3	Treasury Bills	5,000,000,000.00	4,455,850,000.00	13.46%	8/13/2020	8/12/2021	Sovereign	4,455,850,000
4	Treasury Bills	1,753,700,000.00	1,567,938,530.00	12.33%	10/14/2020	10/7/2021	Corporate	1,567,938,530
5	Treasury Bills	3,000,000,000.00	2,673,090,000.00	13.02%	10/26/2020	10/21/2021	Sovereign	2,673,090,000
6	Treasury Bills	2,000,000,000.00	1,772,580,000.00	14.11%	11/6/2020	11/4/2021	Sovereign	1,772,580,000
7	Treasury Bills	5,000,000,000.00	4,447,150,000.00	13.05%	11/20/2020	11/18/2021	Sovereign	4,447,150,000
8	Treasury Bond	7,900,000,000.00	8,527,724,414.00	16.50%	9/20/2019	9/21/2021	Corporate	8,527,724,414
9	Treasury Bond	2,000,000,000.00	2,174,755,824.00	17.00%	7/10/2017	12/3/2021	Sovereign	2,174,755,824
10	Treasury Bond	1,807,300,000.00	1,869,548,897.00	14.13%	11/11/2019	7/7/2022	Corporate	1,869,548,897

APPENDIX III: Data collected

Respondent demographics

    i.   The type of bank – Local or International

Bank Type	Responses
Local	38
International	12

    ii.   The job function of the respondent

Job function	Responses
Credit	21
Financial Reporting	2
Internal Audit	7
Other	5
Risk	13
Treasury	2

    iii.   The largest segment in the bank's portfolio

Segment	Responses
Agriculture	9
Consumer	22

Corporate	5
Other	6
Retail	15

iv.   The percentage increase of reporting under IFRS 9

Segment	Responses
Agriculture	9
Consumer	22
Corporate	5
Other	6
Retail	15

Knowledge of the standard

Analytics and Visualization	clean_ad option_mode	Understanding of IFRS 9	Statistical modeling	Proficiency Excel	Preferred mode of adoption	Number certified IFRS 9	Most challenging parameter	Enough resources?	Directly Involved in IFRS 9?	clean_num_ifrs9_cert
6	Internal	2	4	5	In	4	NA	No	No	Less than 5
4	Internal	4	3	1	In	2	EAD	Yes	Yes	Less than 5
3	Internal	4	3	3	In	2	LGD	Yes	Yes	Less than 5

3	Internal	4	3	7	In	3	PD	No	Yes	Less than 5
5	Internal	5	9	12	In	3	LGD	No	Yes	Less than 5
9	Internal	5	9	5	In	4	NA	No	No	Less than 5
3	Internal	5	4	6	In	3	PD	No	Yes	Less than 5
9	Internal	6	13	12	In	7	LGD	Yes	Yes	Less than 5
6	Internal	7	6	10	In	6	EAD	Yes	Yes	Less than 5
15	Internal	7	17	15	In	7	LGD	Yes	Yes	Less than 5
7	Outsource	1	6	17	Out	5	NA	Unaware	No	Less than 5
9	Outsource	2	5	13	Out	5	NA	Unaware	No	Less than 5
4	Outsource	2	5	5	Out	4	PD	No	Yes	Less than 5
9	Outsource	3	4	1	Out	5	EAD	No	Yes	Between 5 and 10

9	Outsource	3	4	3	Out	3	EAD	No	Yes	Less than 5
5	Outsource	3	2	4	Out	5	NA	Unaware	No	Between 5 and 10
7	Outsource	3	3	5	Out	3	NA	Unaware	No	Less than 5
9	Outsource	3	5	4	Out	2	PD	No	Yes	Less than 5
6	Outsource	3	5	9	Out	3	PD	Unaware	Yes	Less than 5
2	Outsource	3	2	7	Out	3	PD	Yes	Yes	Less than 5
8	Outsource	4	3	2	Out	4	EAD	No	Yes	Less than 5
9	Outsource	4	15	19	Out	4	PD	Unaware	Yes	Less than 5
6	Outsource	5	3	3	Out	6	NA	Unaware	No	Between 5 and 10
6	Outsource	5	4	1	Out	3	NA	Unaware	No	Less than 5
4	Outsource	5	4	7	Out	4	PD	No	Yes	Less than 5

3	Outsource	5	5	3	Out	8	PD	Unaware	Yes	Between 5 and 10
14	Outsource	5	14	21	Out	12	PD	Unaware	Yes	Less than 5
8	Outsource	6	3	5	Out	2	EAD	No	Yes	Less than 5
5	Outsource	6	5	7	Out	1	PD	No	Yes	Less than 5
1	Outsource	6	3	4	Out	3	PD	Unaware	Yes	Less than 5
4	Outsource	7	3	6	Out	1	EAD	No	Yes	Less than 5
6	Outsource	7	5	7	Out	2	PD	No	Yes	Less than 5
8	Outsource	8	6	15	Out	5	EAD	No	Yes	Less than 5
3	Outsource	8	6	6	Out	3	PD	Unaware	Yes	Less than 5
7	Outsource	9	11	17	Out	7	PD	Unaware	Yes	Less than 5

Modeling approaches in use

Analytics	Audit queries	Historical data years	im_accuracy	im_analytics	im_automation	im_cost	im_flexibility	im_integration	im_time	Internal	Personnel Involved	Platform Used	TAT in days
No	5	7	8	10	7	8	6	6	1	No	2	Fully Outsourced	4
No	2	8	10	4	5	3	3	7	1	Yes	5	Spreadsheet	4
No	3	9	9	10	6	8	6	9	2	No	2	Fully Outsourced	4
No	5	7	9	8	7	5	4	10	2	No	3	Spreadsheet	6
No	9	16	26	10	19	18	12	13	2	Yes	14	Spreadsheet	20
No	6	6	7	6	5	6	4	10	3	No	4	Custom solution	7
No	10	14	17	15	14	6	9	16	3	No	8	Spreadsheet	15

No	18	18	36	17	30	27	21	14	3	Yes	17	Spreadsheet	31
No	2	8	10	5	7	7	4	4	4	No	5	Spreadsheet	12
No	4	9	7	10	8	7	5	8	4	Yes	1	Fully Outsourced	5
No	17	22	27	13	21	17	16	16	4	Yes	14	Spreadsheet	26
No	5	7	8	4	4	7	6	10	5	No	3	Custom solution	6
No	7	7	7	8	7	5	5	2	5	Yes	3	Spreadsheet	5
No	4	8	9	0	9	6	9	10	6	No	4	Custom solution	5
No	4	6	8	9	8	6	6	3	6	Yes	4	Spreadsheet	4
Yes	2	6	8	7	7	6	8	8	1	No	3	Custom solution	6
Yes	1	7	7	10	9	8	4	10	1	No	3	Fully Outsourced	3

Yes	2	3	10	1	9	7	8	6	1	Yes	2	Spreadsheet	8
Yes	7	13	15	13	13	8	19	15	2	No	5	Custom solution	10
Yes	7	5	7	8	7	6	3	5	2	No	2	Fully Outsourced	3
Yes	15	18	24	24	21	18	12	25	2	Yes	7	Fully Outsourced	11
Yes	4	13	19	10	13	8	15	17	3	No	5	Custom solution	13
Yes	5	11	18	17	11	14	11	16	3	No	4	Fully Outsourced	10
Yes	3	5	10	2	8	5	9	8	3	No	2	Spreadsheet	4
Yes	4	13	19	18	15	16	15	17	3	Yes	8	Fully Outsourced	12
Yes	10	19	20	21	18	12	21	28	4	No	8	Custom solution	22

Yes	12	10	16	18	15	16	11	15	4	No	4	Fully Outsourced	8
Yes	4	6	9	3	6	4	4	9	4	No	3	Spreadsheet	5
Yes	5	7	9	8	7	4	4	8	5	No	2	Fully Outsourced	4
Yes	9	7	8	8	5	5	2	5	5	No	4	Spreadsheet	6
Yes	8	8	9	7	6	6	3	2	6	No	7	Spreadsheet	7
Yes	4	7	8	9	8	8	8	10	7	No	4	Custom solution	4