

EFFECTIVE DATA MINING AND ANALYSIS FOR SME BANKING



By Mary Miller and Esther Nyauncho

> INTRODUCTION

Data mining may be described as the automatic or semi-automatic analysis of large quantities of data, to extract previously unknown interesting patterns such as groups of data records, unusual records and dependencies. However, for lay users such as banks, the term also encompasses data collection, analysis (e.g. exploratory, regression, panel, etc.) and review. The information may be sourced internally in the bank or externally in the market place.

This Technical Note looks at data mining both on individual SME accounts and on the SME portfolio as a whole, and suggests how and why the extraction and review of data should be undertaken. At a time when banks are starting to offer SME banking services that are distinct from products offered to corporate and retail customers, data mining is a good way to both understand the current customer base and to assess

what the SME portfolio could look like. Data mining is an important element of what can more generally be termed business intelligence.

As data mining is pursued more actively, it is important for banks to consider customer privacy. This is both a legal issue, but is also a fiduciary or trust issue for banks with respect to client relationships. Customer privacy may put a limit on the amount of data mining that can be done. In this paper there are discussions of using data mining techniques at an individual customer level, to identify needs and prospects; when using data mining to tease out population characteristics the actual identities of individuals and businesses are not needed.

> DATA MINING IN KENYA

In Kenya data mining is considered an in-house activity, although "big data" sets are becoming more available. Data mining tasks are considered to include:

About GrowthCap

Over the past few years FSDK has been at the forefront of SME banking development through conducting market assessments and studies in areas such as trade finance and SME equity funds, as well as supporting development of the credit reference bureau. Through its partnerships with its Action Research Partners (ARPs), FSDK's GrowthCap initiative is supporting adoption of SME best practices by individual financial service providers.

This paper is part of a series of Technical Notes and Resource kits that are being developed out of work with the ARPs. These provide detailed information about the best practices and are intended for use by financial service providers and those supporting such institutions which are entering the SME market.

Abstract

This Technical Note looks at data mining both on individual SME accounts and on the SME portfolio as a whole, and suggests how and why extraction and review of data should be undertaken. At a time when banks are starting to offer SME banking services that are distinct from products offered to corporate and retail customers, data mining is a good way to both understand the current customer base and to assess what the SME portfolio could look like. Data mining is an important element of what can more generally be termed business intelligence.

About the authors

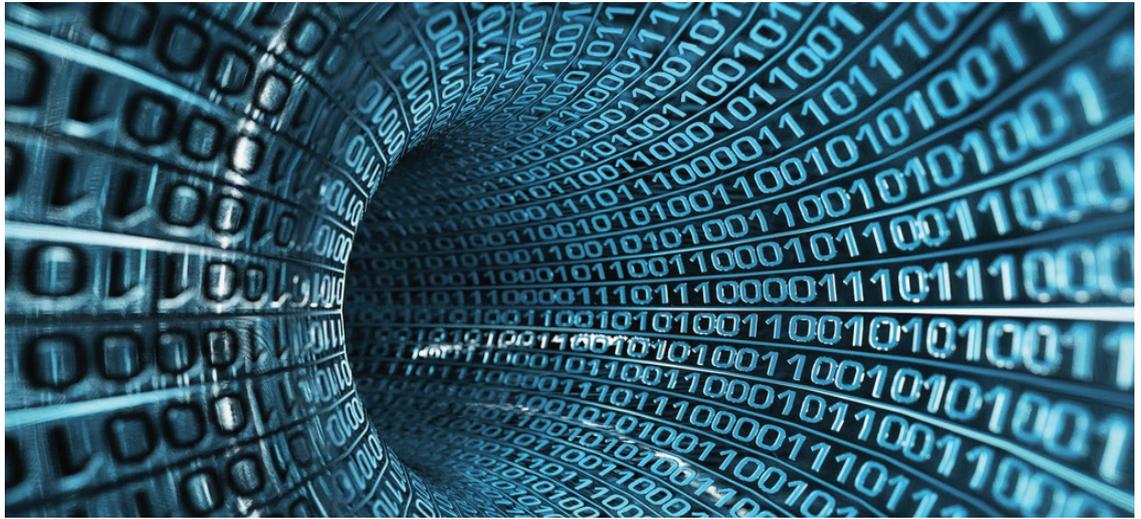


Mary M. Miller is an international expert to GrowthCap



Esther Nyauncho is an associate at GrowthCap

Financial institutions use a data warehouse to collate all data from within the financial institution, and BI systems then 'sit' on top of the data warehouse to mine data and add intelligence to it for analytics and decision making.



- Cluster profile analysis
- Dependency analysis
- Concept description
- Deviation detection
- Data visualisation
- Predictive analytics/modelling
- Sequential/categorical analysis

Many of the banks in Kenya are at an in-between stage in the process of data collection. While banks do have core banking systems that can do some of the work of extracting data on portfolio sets and analysis of income and costs, they are not frequently appropriate for the analysis of dependent relationships or determining causative behaviour. In banking customer relationship management (CRM) systems are used to capture comprehensive data on clients, which can be used for this type of analysis and predictions. CRM is intended to help the bank acquire new customers, retain existing customers, and maximise the value of the customers to the bank. It will identify the bank's most profitable customers, and prospects, so that the bank can prioritise service and calling. The information contained in a CRM system will be drawn from and serve the sales, service, and marketing functions of the bank.

As CRM systems are being introduced the quality of information that is available for analysis will depend on requesting and uploading the right information from customers, which initially will call for major input from the branches and relationship managers (RMs). Individual offices and RMs typically have

stores of information available about their customers that may be written (but is more likely just mentally stored) but which is not collected in any systematic way, and is likely to be lost if the RM leaves the bank or branch management changes. An important objective of a successful CRM installation is to codify this information, and to keep it accurately updated thereafter.

The next stage of data mining capability for a bank can be described as a business intelligence (BI) system. This mines data from many sources within the bank, including core banking and CRM, card systems, and other. Financial institutions (FI) can use a data warehouse to collate all data from within the financial institution and BI systems then 'sit' on top of the data warehouse to mine data and add intelligence to it for analytics and decision making. The primary reasons for using a data warehouse are to combine data from different systems, and to have data available from a source where the primary data storage cannot be affected by the analysis process.

> REASONS FOR DATA MINING

Why would a financial institution want to do data mining and data analysis?

The key purpose of a bank undertaking data mining is to enable it to take a more proactive stance when it pursues customers, rather than just evaluating customer-initiated loan requests. For instance, rather than just make annual update calls on SME clients the bank can plan to look for loan renewals and expansion requests, ask for family business, and provide industry information, as a planned part of such a call.

Typical opportunities arising from the use of data-mined information include:

- Identifying cross-selling opportunities for retail business (from SMEs) and to SMEs (from retail business, typically high net worth segment, also business entities operating from retail accounts).
- Determining what products are the logical follow-on from an existing product set.
- Estimating wallet size and profit potential.
- Prioritising calling for a branch or relationship management team.
- Determining success of marketing strategies, advertising, calling plans.
- Predicting behaviour, e.g., whether a customer is likely to leave.
- Determining sector and industry concentrations.
- Segmenting customers for targeted value proposition/marketing.
- Looking for upstream and downstream clients – the SME's buyers and clients.

➤ KEY FORMS OF DATA MINING FOR SME BANKING

Data mining exercises can be used to focus attention on SME customers at the individual level, on SMEs as segments, or SMEs as a collective portfolio.

4.1 DATA MINING INDIVIDUAL CUSTOMER RELATIONSHIPS

If well-designed, a CRM system should be able to pull together the type of customer relationship evaluation described below. In the absence of a CRM system or as information is added to the CRM system for the first time, RMs and branch staff can do a manual data mining of individual SME customer relationships. This will involve bringing together all account information on a particular SME to evaluate its total financial wallet, including both the part of the wallet captured by the bank and the part captured by competitors. This can then form the basis of planning marketing calls, and can also be fed into a CRM system.

A sample information collection format and instructions are shown in Annex 1. The form may provide some background information on the customer including the number of years in operation, the number of employees, and the subsector that the business is operating in. The main purpose of the form, however, is to determine which types of accounts the customer has with the bank, what accounts the customer has with competitors, and who the competitors are.

As the bank implements a new customer outreach it should be doing this type of data mining on all its customers. While it will start with the best known customers, it should finally be investigating all accounts, including dormant accounts, on the books as they may lead to potential new business. It is important to remember that any existing banking relationship is more of a lead than a cold call, and even contacting a dormant account holder is leveraged by potential to re-introduce the bank and to let the customer know about new bank services and products.

The information collected through such a data mining exercise will include estimates and speculations, thus may not be totally accurate. It will reflect the RM's best guess or understanding about a customer, and the CRM system, or the form, can be updated as information is verified or if it changes.

Good data mining of this type is going to be a tedious and mundane task (which is why CRM systems may not always be as up to date and accurate as they should be). However, a bank that is serious about marketing and cross-selling a full financial wallet will recognise the need to go through this process. If a CRM system has not yet been put in place by the bank but is still under consideration, this type of work can inform what types of data will be useful to collect.



A bank that is serious about marketing and cross-selling a full financial wallet, will recognise the need to go through the data mining process



Benefits of individual SME customer data mining

On an individual customer basis the main benefit of this work is to be well-prepared for customer marketing calls and to be able to anticipate customer needs, rather than just reacting to customer requests. This can enable RMs to plan their calls with results in mind, e.g. collect information for the expected request for new equipment financing; introduce bank's payroll product to support service to customer's growing employee base. However, other benefits can be realised for more than just single active customer relationships. Other possibilities include:

- Finding customers whose relationship can be revived and expanded.
- Increasing knowledge of which financial institutions are competitors. Banks generally know who their competition but this type of information helps pinpoint, for example, who is dominant in payroll services, and who provides equipment finance in a particular sector.
- Prioritising customers, establishing segments for service – assuming that the data mining is accompanied by a customer profitability analysis, including estimates of potential wallet size, analysis of this type – will inform the bank about account volumes and business characteristics that fall into different segment and sub-segment categorisations.
- Understanding whom the SME customers do business with, including whom they pay and who pays them. A fairly detailed profile of a customer's business can be build around this information, cyclic patterns, where the suppliers and buyers are situated and where they bank.

4.2 PORTFOLIO ANALYSIS

Portfolio analysis is the process of reviewing account data to better understand the customers and accounts that make up the portfolio. It is another form of data mining, used to determine characteristics of the whole or parts of the loan or deposit portfolio. Information may be drawn from the core banking systems, usually exported to an Excel spreadsheet, and can be sorted by using filters and pivot tables. The analytical capabilities of the core loan and deposit systems may also be able to do data sorts. Sorting current data indicates groupings or concentrations by sector, legal form, and geographical information, and calculations can determine the values of numbers of accounts, averages, and means for these groupings. This quantification of the portfolio can confirm or clarify expectations about account sizes.

The following types of questions about the bank's SME term loans might be addressed through a portfolio analysis:

- What is the average original term loan size?
- What is the average outstanding of term loans?
- What are the mean values (the "middle" value – of 100 items ranked in size order, the mean would be the value of the 50th item)?
- How many loans are equal to or above the bank's stated cap for SME loans?
- How many loans (original value) are underneath the stated floor value?
- What are the typical ranges for the loans?

For example, shown below is the portfolio of term loans for a bank showing the original amounts of the loans. The bank's policy is that SME loans have a maximum size of 100 and a preferred floor of 20:

Table 1: Sample term loan portfolio

Loan Number	Original Amount	Excluding Outliers
1	185	
2	150	
3	90	90
4	45	45
5	40	40
6	38	38
7	35	35
8	30	30
9	29	29
10	28	28
11	26	26
12	25	25
13	25	25
14	24	24
15	22	22
16	21	21
17	20	20
18	20	20
19	15	15
20	10	10
Total	878	543
Count	20	18
Average:	43.9	30.2
Mean:	28.0	26.0



The initial analysis shows that the average loan size is 43.9, and that the total portfolio value is 878 for 20 loans. However, this array indicates that loans 1 and 2 do not appear to belong in the portfolio, and that they should be reclassified as corporate loans. When they are excluded the average loan size drops to 30.2. While there are a few larger loans, and one very large loan (no. 3), loans tend to be for amounts between 20 and 30, suggesting that overall the bank's portfolio is concentrated in a high volume of smaller loans.

For this portfolio the possible segments by size were grouped by boxing off those loans of like size. Note that there are two loans that are smaller than the preferred minimum. If there were lots of these smaller loans the bank should realistically conclude that its market is concentrated in credits that are smaller than its preferred minimum. It can either acknowledge that the market is there, or enforce the minimum as a part of a more up-market strategy.

For overdrafts the bank would want to look both at original commitments and outstanding amounts, to get an idea of how many of the overdraft lines are used and average usage compared to available amounts. For deposits it would be preferable to compare primary current accounts rather than all accounts, as many secondary accounts are used for transfers and the like, and are swept or otherwise have low or zero balances.

Depending on information fields that are available in the systems, the sorting can also provide information about operating sectors, geography, and legal status, and can be an impetus to accurately select codes for these fields.¹

¹ Too frequently codes go unused, or the "other" choice is selected. The data mining work may include a step to re-code or accurately code customer accounts so that future portfolio analysis can be more accurate and informative.

Table 2: Sorting by sectors

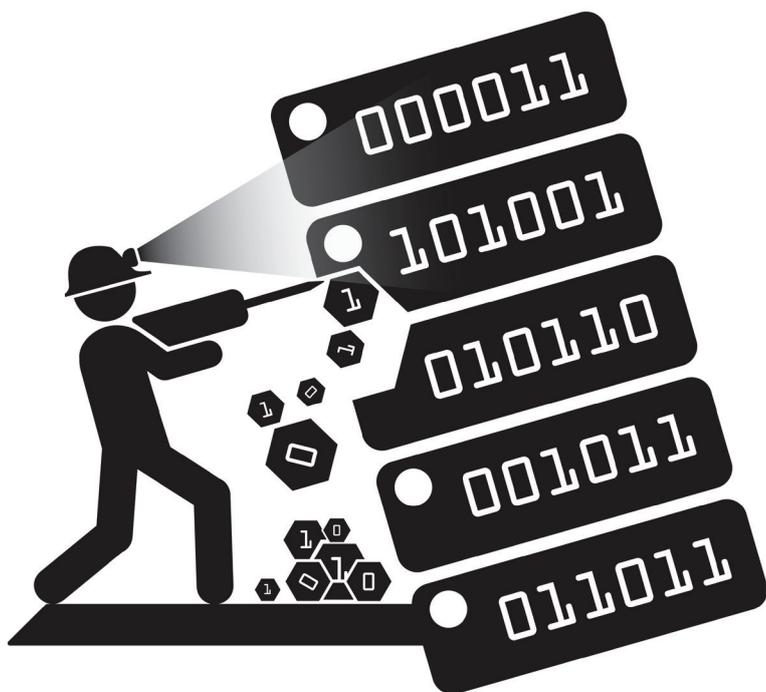
The Central Bank of Kenya requires that banks collect information by eleven major sectors, as follows:

- Agriculture
- Building and Construction
- Energy and Water
- Financial Services
- Manufacturing
- Mining and Quarrying
- Personal Household
- Real Estate
- Tourism, Restaurants, and Hotels
- Trade
- Transportation and Communication

Using these categories and eliminating any that would not be SME (i.e. Financial Services and Personal Household) can be a start to understanding the groupings or concentrations within the portfolio.

In Kenya banks are using some data mining techniques on deposit account data bases to relate account balance, account transactions (debit and credit), nature of transactions, source/ location of transactions, dates, etc., to better understand customers and customer behaviour. This started with personal banking departments with a focus on high net worth customers, but is increasingly spreading to SME departments. Oreko databases are the mostly commonly used in by banks

In Kenya banks are using some data mining techniques on deposit account data bases to better understand customers and customer behaviour.



and telecommunication companies in Kenya². Other sources of data that are being used include the Credit Reference Bureau, consumption data from utility companies, court records, supermarket chains and petrol stations (using loyalty cards), other club memberships, economic data from the Central Bureau of Statistics, and social media, which is growing in significance. Kenyans in particular are among the largest users of social media in the world.

Benefits of portfolio analysis

Doing this type of portfolio analysis has two specific benefits. First, it gives the bank a realistic idea of its current portfolio, and can help the bank tailor services appropriately. Second it informs future planning, helping to select targets for performance, e.g. the bank would like to move the average account size up, and so overall the bank would focus on clients that are larger than its current portfolio. The bank's focus can be sharpened further when the portfolio analysis is also supported by a profitability analysis.

² This is a dataset technology company commonly used by Kenyan banks and telcos, which provides databases and storage.

> BIG DATA IN KENYA

Big Data might be described as a database of customer information that can be easily accessed, for a price, to inform the bank about customer trends and relationships. It is a comprehensive term for collections of large and complex data sets that are difficult to process using traditional data processing applications. Put simply, very large amounts of data are generated in our world on a day to day basis including in the financial sector. Systems have been specifically developed that can handle manage and analyse this volume of data points.

Many big technology firms such as IBM and Oracle are pitching big data analysis as a new business in the region, as neither Kenyan businesses nor other potential users in East Africa has given much thought or focus to this area. Sources in Kenya include large retail outlets, particularly Safaricom, which does extensive data mining on its database.

M-pesa department uses SAS³ for all M-pesa data analytics including customer segmentation based on usage data, data visualisation to identify outliers for possible hidden strategic opportunities, basic link analyses among customers to understand the essence of social networks, and trend analysis to compare actual business performance against forecast. Safaricom also uses STATA and SPSS⁴. VISA is mining card network information and is a key source of big data. IT departments are now moving into the mobile and person to person payments as well.

Information being collected in this way to date does not appear to be sorted in such a way that SME segment information can be isolated. This is because the data is more applicable to individuals than to businesses. Over time, however, this information may be useful for credit scoring development, or for if-then assumptions (if the SME customer has this product, then they are likely to next need this service).

³ Statistical Analysis System is a software suite developed by SAS Institute for advanced analytics, business intelligence, data management, and predictive analytics.

⁴ STATA and SPSS are software packages used for statistical analysis.

> KEY CHALLENGES WITH EFFECTIVE DATA MINING FOR SME BANKING

The biggest challenge for effective data mining is for banks to take a pro-active stance and actually start doing it. Some of this is occurring in analysis of portfolios, but individual account mining necessarily starts with laborious manual reviews that cannot be computerized and stored in a CRM system until data is generated. Other challenges to data mining and management of data sets include:

- Poor quality of data in Kenyan banks serving SMEs – issues with storage and cleanliness. Raw data inaccuracies include errors on names, business details, industry types and classifications, caused by lack of attention to detail, poor records, and a poor certification and validation process.
- Consumer protection is a concern. Without proper data security measures and infrastructure it will not be easy to establish effective data mining practices.
- The lack of well trained staff who can manage complex and heterogeneous data sets.
- Any data mining algorithm has to be amenable to lend itself to scalability with the growing volumes of data for an SME portfolio.

The rewards are worth it, however:

- The availability of data prepped for entry into a CRM system that in turn will be able to automatically generate reports and information
- The basis for informed active engagement with clients rather than simply providing reactive services.
- Much better insight into the activities of competitors.

> REFERENCES

Aher, Ashok T., and D. D. Bhakkad, *CRM in the Banking Sector: Impact and Reality*, *Research Link 93 Vol. X* December 2011.

Henschel, Jim, and Chris Menzer, *Time to Overhaul Your Analytics Machine?*, ABeam Consulting USA, BAI Banking Strategies, 24 July 2012.

Onut, Semih, Ibrahim Erdem, Bora Hosver, *Customer Relationship Management in Banking Sector and A Model Design for Banking Performance Enhancement*, Yildiz Technical University and Garanti Bank, undated (approx. 2002).

Stuart, Bill, Rich Data, *Hiding in Plain Sight*, ABeam Consulting USA, BAI Banking Strategies, 9 December 2013.

Umang, Tzar C., *Business Intelligence for SMEs with Data Analytics*, June 2013.

Wendel, Charles, 2014: *The Year Big Data Becomes the Secret Sauce in Commercial Banking*, FIC Newsletter, January 2014.

Acknowledgements: The authors wish to acknowledge the feedback given by several local GrowthCap consultants in helping to shape this final document and the final editing undertaken by Dr Pat Richardson, as an international adviser to GrowthCap.

> ANNEXES

ANNEX 1: CUSTOMER DATA MINING

1.1 SCHEDULE OF TASKS

It is important to remember that data mining is focused on collecting and centralising information on separate customer relationships, but also is providing data for the bank to understand its own portfolio and make marketing decisions. In all cases there should be staff assignments for responsibility of completion of tasks (or ongoing monitoring assignments) and deadlines as a part of developing a good data mining plan.

- 1. Set task schedule** - Fill in the dates and assignments below and commit to all responsibilities and reporting deadlines.

Senior Officer Responsible: _____

Complete By Date: _____

- 2. Data mining assigned accounts** - each relationship manager should review his/her SME portfolio against a checklist - what services does the customer take from the bank, what services could the customer take, where else does the customer get services. (This should be done with an eye to future planning - what products or services should the bank be offering the client.) This should be done by a set deadline, so that the relationship managers can start calling on the customers and sharpening up their client acquisition plans.

Senior Officer Monitoring: _____

Complete By Date: _____

Account Officers Reporting:

Follow-up: Calls on existing clients for cross-sales, collection of additional information; set acquisition/cross-sale plans – calls should be planned and be proactive rather than just annual reviews, “get to know you”, and anticipate certain results – e.g., sell a new loan, get financial information for file documentation, introduce point of sale services.

- 3. Data mining of inactive accounts** - probably will be done on a branch by branch basis, and branch managers/staff who have been around for a while will likely know accounts. This review does not need to be as “in-depth” as the data mining – the objective is more to do a yes/no review, and then to spend more time reviewing some of the accounts. Many of these will not be prospects, but some will - establish a calling programme to “re-introduce” the bank, let the customer know what is new at the bank, etc. While customers may be in other settled banking relationships, this can open the door - promote the idea that all customers should have at least two banks, etc. This should also have more opportunities, because the bank is already known, than cold-calling.

Senior Officer Monitoring: _____

Complete By Date: _____

Branch Managers/Account Officers Reporting:

Follow-up: Assign inactive accounts to relationship managers and establish a calling programme to reintroduce the bank to certain clients and build relationships.

- 4. Analysing product data.** From the data mining exercises start determining what products are being used by what type of clients - this can be gathered onto a table like the illustrative product table previously provided. This helps to confirm where the bank is competitive, also where products are little-used - should they be dropped or would it be a good idea to look for more clients who could use them, if SME Bank has an opportunity to be the market leader.

Senior Officer Responsible: _____

Complete By Date: _____

Bank personnel identified to do analysis:

Follow-up: Evaluation of product mix, opportunities for cross-sales.

- 5. Analysing sector data. From the data-** mining exercises, determine what industries/sectors being served – should these be emphasised more? The bank may wish to designate industry experts within the bank or the SME department who start to follow the industry, get to know the industry association and its members, etc.

Senior Officer Responsible: _____

Complete By Date: _____

Bank personnel identified to determine industry concentrations (may be same staff assignments as #4):

Follow-up: Identify industry/sector champions to do industry studies, follow industry progress.

- 6. Tracking competitors: Start tracking the product offerings of the other banks** - some of this data will come from website scans and mystery shopping and some from what you are hearing from staff and customers. This does not need to be done for all banks, only those who are regular competitors. Start with a decision of which banks to track.

Senior Officer Responsible: _____

Complete Selection of Assignments By Date: _____

Assignments for:

Reviewing website offerings

Mystery shopping

Follow-up: Design collection system to track information passed on from staff and customers.

Progress may be monitored with the use of a tracking sheet for each client that specifies target products to introduce and results.

1.2 ILLUSTRATIVE INFORMATION COLLECTION SHEET

Data Mining - Current Portfolio		ILLUSTRATIVE													
Account Officer:															
Client Name	Industry	Business Location (nearest branch)	Line of credit	Term loan	Transaction - deposit acct	Time Deposit	Business Loan	Guarantees	Domestic Transfers	International Transfers	X	X	X	X	Comments
Key:															
Service with SME Bank	Y														
Service not applicable	X														
Service with another bank	B														
(show bank name in Comments)															
Not Known	?														

The Kenya Financial Sector Deepening (FSD) programme was established in early 2005 to support the development of financial markets in Kenya as a means to stimulate wealth creation and reduce poverty. Working in partnership with the financial services industry, the programme's goal is to expand access to financial services among lower income households and smaller enterprises. It operates as an independent trust under the supervision of professional trustees, KPMG Kenya, with policy guidance from a Programme Investment Committee (PIC). In addition to the Government of Kenya, funders include the UK's Department for International Development (DFID), the World Bank, the Swedish International Development Agency (SIDA), Agence Française de Développement (AFD) and the Bill and Melinda Gates Foundation.



info@fsdkenya.org • www.fsdkenya.org
 FSD Kenya is an independent trust established to support the development of inclusive financial markets in Kenya
 5th Floor KMA Centre • Junction of Chyulu Road and Mara Road, Upper Hill
 P.O. Box 11353, 00100 Nairobi, Kenya,
Tel +254 (20) 2718809, 2718814 • **Cell** +254 (724) 319706, (735) 319706