Addis Ababa University
College of Business and Economics

Department of Management

The Use of Banking Transactional Datasets for Humanitarian Aid Beneficiary Targeting

By:-Tesfahiwot Tefera

Advisor:- Mesfin Fikre (PhD)

A Thesis Submitted to Addis Ababa University, College of Business and Economics in Partial Fulfillment of the Requirements for the Degree of Master of Science in Management

Addis Ababa, Ethiopia

November, 2022

Addis Ababa University

College Of Business and Economics

Department Of Management

The Use of Banking Transactional Datasets for Humanitarian Aid Beneficiary

By

Tesfahiwot Tefera

A Thesis Submitted to Addis Ababa University, College of Business and Economics in Partial Fulfillment of the Requirements for the Degree of Master of Science in Management

Advisor

Mesfin Fikre (PhD)

Addis Ababa, Ethiopia

june, 2023

Declaration

I, the undersigned, declare that this thesis is my original work and all the sources of materials used for the thesis have been appropriately accredited.

Name: Tesfahiwot Tefera

Signature: ____________________

Date of submission: November, 2022

Place of submission: Addis Ababa, Ethiopia

Approval by Examining Board

Members of board of examiners approve that this examination thesis in titled, '**The Use of Banking Transactional Datasets for humanitarian Aid Beneficiary Targeting**' undertaken by Mr. Tesfahiwot Tefera Deginet the necessities for the reward of MSC in management in the department of management, School of Business and Economics, Addis Ababa University and is acceptable with regards to the standards and regulations of the University.

Approval: Board of Examiners

Name of External Examiners	**Signature**	**Date**
Dr. Mesfin Gezahegn		19 July, 2023

Name internal Examiners	**Signature**	**Date**
Dr. Demeke Chimdessa		19 July, 2023

Name of Advisor	**Signature**	**Date**
Mesfin T.		July 17/23

Acknowledgement

I would like to give my heartfelt thanks to my advisor Dr. Mesfin Fikre who made this work possible, nothing would be done without his inspiration, guidance and support. Warm thanks also to the MIS Department director and staffs of the Commercial Bank of Ethiopia for their will and patience to give me such a huge and meaningful data.

Dear God I shall knell down in front of you for giving me such an inspiring and supportive advisor who encourages to get the best out of me regarding this thesis and above all my almighty for your willingness and help to make me pass all through these and achieve the objectives.

Table of Contents

List of Figures

List of Tables

Abstract

Humanitarian aid organizations and governments are faced with the challenge of locating and identifying the poorest individuals for efficient and effective intervention. The old ways of poverty assessment through survey is expensive and have limitations. Surveys are conducted once in many years, are prone to biases by local surveyors' and supervisors, respondents give biased and wrong responses specifically if they know data is collected for aid purpose. As census is unlikely, sample based data collection is not as accurate as what one can do with latest targeting approaches. Currently researchers are exploring the use of night time satellite imagery and mobile phone data analysis for targeting purpose. These approaches are also not without limitations. In developing countries the rural poor are not connected to electric grid and hence less feasible to be used for targeting. In this paper we explore the use of bank transaction activities of individuals in aid prone counties in Ethiopia. Based on IOM's 2021 humanitarian aid seekers, 25 counties in Amhara Region of Ethiopia are selected. Over 23 million daily transaction records with an amount of over 18bln birr for 2020 and 2021 is acquired. The result is quite interesting. By analyzing how bank transaction datasets like deposits, withdrawals, and transfers etc. trends over time, it is possible to prioritize aid seeking places accurately and objectively. Such trends or seasonality can be used to derive other measurement indexes, to explore how these kinds of datasets can be used to improve humanitarian beneficiary targeting. Finally, adding bank transaction data analytics indexes into targeting aid beneficiaries' models can improve predictions. The knowledge gained from this study could provide valuable insights into strategies for utilizing banking datasets to increase efficiency and effectiveness of targeting.

Keywords: *targeting, humanitarian, data analytics, forecasting, time series, machine learning, bank account transactions*

CHAPTER ONE

1. INTRODUCTION

1.1. Background of the study

Aid organizations rely on government and international agencies supplied datasets and information. These datasets and or information are usually census data that governments collect many years in advance and are misleading and outdated. When such datasets are collected by government officers, and it includes financial matters of citizens, respondents do not give accurate information (data) as they suspect government is going to levy more tax on them. If they also know data or information is gathered for aid purpose, respondents usually exaggerate their problem and demand more aid. This leads to wrong decisions (wrong targets). Besides these challenges, survey datasets are expensive to collect, subject to biasness, is based on sample representatives.

An alternative solution to this challenge, machine learning and artificial intelligence scholars start exploring the use of datasets like night time lights[1] of places; mobile phone datasets (Blumenstock, 2016). Developments in artificial intelligence, machine learning, and data science are becoming useful tools for policy analysis, targeting, human behavior capturing, and analysis. For example, study by (Abebe & Goldner, 2018) demonstrates the use of algorithms to curb social inequalities due to poor and simple judgmental decision making by officials. Algorithms can play significant role in identifying targets to provide aid. Unlike human officers, computer algorithms can continuously track and capture the activities (behavior) of individuals.

In developing countries, the use of night time light satellite imagery appears to be less applied. For example, majority of Ethiopian rural areas who need aid do not have access to electricity. They are outside of the national electric grid. It is also impossible to claim places without electricity have poor individuals. In rural areas, there are farmers who have no access to electricity but still manage average life. Satellite imagerymight not be applied in poor countries.

[1] Understanding poverty in the Philippines with artificial intelligence - Thinking Machines Data Science

The biasness of the survey method and inevitable limitations to use night time satellite imagery intensifies the necessity of a comprehensive input for humanitarian aid beneficiary targeting.

1.2. Statement of the problem

The task of determining who is eligible and who is ineligible for humanitarian aid is a major source of inefficiency in beneficiary targeting program administration. Typically, aid programs in developing countries make targeting decisions based on survey-based measures of assets or consumption. In many developing countries, however, reliable data for targeting do not exist and would be prohibitively expensive to collect. Over the past several years, a handful of studies have shown that the classical survey method used to target beneficiaries which can be easily biased by the surveyors as well as the supervisors.

The recent phenomenon of using night time satellite imagery to predict humanitarian aid seekers is less applied in countries like Ethiopia due to the fact that most of the rural society is not accessible for electricity. Thus, Governments and aid organizations face persistent challenges in targeting social welfare programs to accurately identify and reach intended beneficiaries.

According to behavioral economics literature like (Karlan, Ratan, and Zinman, 2014), the poor can save at banks for different reasons like: to avoid temptation to expend it on unnecessary things, and not being stolen by close friends and individuals (Banerjee and Duflo, 2006; Hoos, 2010; Reyers, 2019). Such savings are used to cope up with different shocks. Emergency savings provide buffer against financial shocks, mainly among low-income households. Reyers, (2019), found that financial capability is highly associated with saving for emergencies. Those with higher levels of financial capability and having access to bank account are more likely to have emergency savings compared with those with lower levels of financial capability. During emergencies and hardship times, individuals do adjust their saving (lower saving amounts) (Eskander et al 2018. In other words, during disasters and hardship times, financial institutions like banks do face more withdrawals by their clients (Brei, Mohan, & Strobl, 2019) from their both saving and mobile money accounts. This is evidence that if there are different life shocks and events around counties one can see a reduction in deposits and an increase in withdrawals.

Individuals' bank account activities like deposits, withdrawals, transfers etc are good indicative of wealth/ poverty of counties. By looking at every bank transaction datasets and their trends and seasonality over time, this study documented interesting insights that can help decision makers prioritize counties for targeting. Even better than satellite night time light imagery datasets, bank account financial activities of individuals can reveal the reality on the ground in a real time and granular manner.

This paper will try to show the extent to which the problem of identifying accurate aid seekers which are eligible to be benefited from aid programs can be resolved through projecting bank transaction of individuals as an alternative method of predicting beneficiaries. The research will explore the use of bank transactions of individuals for humanitarian aid beneficiary targeting.

1.3. Objective of the study

1.3.1. General Objective

The general objective of the research is to explore the use of bank transaction dataset for humanitarian aid beneficiary targeting.

1.3.2. Specific Objective

The specific objectives are:

- To assess the use of big bank transaction data for humanitarian aid beneficiary targeting.
- To show how bank transaction of individuals exhibit to identify areas with the high seek of humanitarian aid.
- To indicate the levels to what extent bank transactions of individuals can help in beneficiary targeting.
- To show how bank transactions of individuals can used as an alternative assessment for beneficiary targeting.

1.4. Significance of the study

The research can be significant in such a way that it confirms bulk bank transaction data can become a comprehensive assessment method in humanitarian aid beneficiary targeting; moreover the research results will have substantial pioneer contribution in suggesting alternative method of measuring aid seeking areas targeting beneficiaries.

1.5. Limitation of the study

The very immediate and major constraint of this study can be the existence of no studies under gone in connection with using bank transactions of individuals in humanitarian aid beneficiary targeting. It becomes impossible to get a study conducted locally in this regard. Hence, the researcher will not compare and contrast his study vis a vis to others and let readers view various perspectives.

1.6. Scope of the study

The research encompasses 25 woredas of Amhara national regional state. These woredas suffer high humanitarian crisis in the past two years due to the internal conflict. Some of the woredas were real war fields and the others were temporary shelters for the vast number of war expatriates.

1.7. Methodology

The research uses secondary data i.e. deposit, Withdrawals, account to account transfers and other bank transactions of individuals as a variable to visualize their trend in a time series model and deploys a predictive analytics to forecast percentage of aid seekers in respective counties.

1.8. Organization of the Study

The study will have five chapters: the first chapter deals with background of the study or introductory part of the paper followed by the Chapter Two, which is dedicated to literature review. Chapter three will be all about the Methodology to be employed in this study. Data presentation, analysis, and interpretation will be included in chapter four of the study. The final, chapter can be dedicated to conclusions and recommendations.

CHAPTER TWO

2. LITERATURE REVIEW

The traditional methods of beneficiary targeting, census and survey are very expensive and will bring inaccurate data due to the fact that the information will be biased by the local surveyor's as well as the wrong information of the respondents and the current beneficiary assessment method which is the use of night time satellite imaginary and mobile phone date are not feasible in areas like the rural regions of Ethiopia which has poor power and network facilities. Hence this research aims to show the most convenient and cost effective way of assessing humanitarian aid seeking areas can be the use and interpretation of big bank transaction data the literature mainly focuses on the importance of Big Data, Data Analytics, and Data Driven Decision Making alongside Business Intelligence and beneficiary targeting.

2.1. Big data

Big data is a combination of structured, semi structured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modelling and other advanced analytics applications.

Systems that process and store big data have become a common component of data management architectures in organizations, combined with tools that support big data analytics uses. Big data is often characterized by the three V's:

- The large volume of data in many environments;
- The wide variety of data types frequently stored in big data systems; and
- The velocity, at which much of the data is generated, collected and processed.

These characteristics were first identified in 2001 by Doug Laney, then an analyst at consulting firm Meta Group Inc.; Gartner further popularized them after it acquired Meta Group in 2005. More recently, several other V's have been added to different descriptions of big data, including veracity, value and variability.

2.1.1. Importance of big data

Companies use big data in their systems to improve operations, provide better customer service, create personalized marketing campaigns and take other actions that, ultimately, can increase revenue and profits. Businesses that use it effectively hold a potential competitive advantage over those that don't because they're able to make faster and more informed business decisions.

For example, big data provides valuable insights into customers that companies can use to refine their marketing, advertising and promotions in order to increase customer engagement and conversion rates. Both historical and real-time data can be analyzed to assess the evolving preferences of consumers or corporate buyers, enabling businesses to become more responsive to customer wants and needs.

Big data is also used by financial services firms for risk management and real-time analysis of market data.

2.1.2. Benefits of big data

- Better customer insight
- Improve operations
- More insightful market intelligence
- Agile supply chain management
- Data driven innovation
- Smarter recommendations and targeting

Big data comes from myriad sources some examples are transaction processing systems, customer databases, documents, emails, medical records, internet click stream logs, mobile apps and social networks. It also includes machine-generated data, such as network and server log files and data from sensors on manufacturing machines, industrial equipment and internet of things devices.

In addition to data from internal systems, big data environments often incorporate external data on consumers, financial markets, weather and traffic conditions, geographic information, scientific research and more. Images, videos and audio files are forms of big data, too, and many big data applications involve streaming data that is processed and collected on a continual basis.

Big data also encompasses a wide variety of data types, including the following:

- structured data, such as transactions and financial records;
- unstructured data, such as text, documents and multimedia files; and
- Semi-structured data, such as web server logs and streaming data from sensors.

Various data types may need to be stored and managed together in big data systems. In addition, big data applications often include multiple data sets that may not be integrated upfront. For example, a big data analytics project may attempt to forecast sales of a product by correlating data on past sales, returns, online reviews and customer service calls.

Velocity refers to the speed at which data is generated and must be processed and analyzed. In many cases, sets of big data are updated on a real- or near-real-time basis, instead of the daily, weekly or monthly updates made in many traditional data warehouses. Managing data velocity is also important as big data analysis further expands into machine learning and artificial intelligence (AI), where analytical processes automatically find patterns in data and use them to generate insights.

2.1.3. Big data analytics

To get valid and relevant results from big data analytics applications, data scientists and other data analysts must have a detailed understanding of the available data and a sense of what they're looking for in it. That makes data preparation, which includes profiling, cleansing, validation and transformation of data sets, a crucial first step in the analytics process.

Once the data has been gathered and prepared for analysis, various data science and advanced analytics disciplines can be applied to run different applications, using tools that provide big data analytics features and capabilities. Those disciplines include machine learning and its deep learning offshoot, predictive modelling, data mining, statistical analysis, streaming analytics, text mining and more.

Using customer data as an example, the different branches of analytics that can be done with sets of big data include the following:

- **Comparative analysis.** Examines customer behavior metrics and real-time customer engagement in order to compare a company's products, services and branding with those of its competitors.
- **Social media listening.** Analyzes what people are saying on social media about a business or product, which can help identify potential problems and target audiences for marketing campaigns.
- **Marketing analytics.** Provides information that can be used to improve marketing campaigns and promotional offers for products, services and business initiatives.
- **Sentiment analysis.** All data gathered on customers can be analyzed to reveal how they feel about a company or brand, customer satisfaction levels, potential issues and how customer service could be improved.

2.1.4. Big data challenges

In connection with the processing capacity issues, designing big data architecture is a common challenge for users. Big data systems must be tailored to an organization's particular needs, a DIY undertaking that requires IT and data management teams to piece together a customized set of technologies and tools. Deploying and managing big data systems also require new skills compared to the ones that database administrators and developers focused on relational software typically possess.

Both of those issues can be eased by using a managed cloud service, but IT managers need to keep a close eye on cloud usage to make sure costs don't get out of hand. Also, migrating on-premises data sets and processing workloads to the cloud is often a complex process.

2.2. Data analytics

Data analytics is the process of analyzing raw data in order to draw out meaningful, actionable insights**.** These insights are then used to inform and drive smart business decisions. So, a data analyst will extract raw data, organize it, and then analyze it, transforming it from incomprehensible numbers into coherent, intelligible information. Having interpreted the data, the data analyst will then pass on their findings in the form of suggestions or recommendations about what the company's next steps should be.

Data analytics is a form of business intelligence, used to solve specific problems and challenges within an organization. Data analytics helps to have a sense of the past and to predict future trends and behaviours; rather than basing decisions and strategies on guesswork, companies made informed choices based on what the data exhibits. Armed with the insights drawn from the data, businesses and organizations are able to develop a much deeper understanding of their audience, their industry, and their company as a whole and, as a result, are much better equipped to make decisions and plan ahead.

2.2.1. Types of data analysis

2.2.1.1. Descriptive analytics

Descriptive analytics is a simple, surface-level type of analysis that looks at what has happened in the past**.** The two main techniques used in descriptive analytics are data aggregation and data mining so, the data analyst first gathers the data and presents it in a summarized format (that's the aggregation part) and then "mines" the data to discover patterns. The data is then presented in a way that can be easily understood by a wide audience (not just data experts). It's important to note that descriptive analytics doesn't try to explain the historical data or establish cause and effect relationships.

2.2.1.2. Diagnostic analytics

While descriptive analytics looks at the "what", diagnostic analytics explores the "why"**.** When running diagnostic analytics, data analysts will first seek to identify anomalies within the data that is, anything that cannot be explained by the data in front of them. For example: If the data shows that there was a sudden drop in sales for the month of March, the data analyst will need to investigate the cause. To do this, they'll embark on what's known as the discovery phase, identifying any additional data sources that might tell them more about why such anomalies arose. Finally, the data analyst will try to uncover causal relationships for example, looking at any events that may correlate or correspond with the decrease in sales. At this stage, data analysts may use probability theory, regression analysis, filtering, and time-series data analytics.

2.2.1.3 Predictive analytics

Predictive analytics tries to predict what is likely to happen in the future. This is where data analysts start to come up with actionable, data-driven insights that the company can use to inform their next steps. Predictive analytics estimates the likelihood of a future outcome based on historical data and probability theory, and while it can never be completely accurate, it does eliminate much of the guesswork from key business decisions. Predictive analytics can be used to forecast all sorts of outcomes from what products will be most popular at a certain time, to how much the company revenue is likely to increase or decrease in a given period. Ultimately, predictive analytics is used to increase the business's chances of hitting the mark and taking the most appropriate action.

2.2.1.4. Prescriptive analytics

Building on predictive analytics, prescriptive analytics advises on the actions and decisions that should be taken. In other words, prescriptive analytics shows how to take advantage of the outcomes that have been predicted. When conducting prescriptive analysis, data analysts will consider a range of possible scenarios and assess the different actions the company might take. Prescriptive analytics is one of the more complex types of analysis, and may involve working with algorithms, machine learning, and computational modelling procedures. However, the effective use of prescriptive analytics can have a huge impact on the company's decision-making process and, ultimately, on the bottom line.

2.2.2. Purpose of data analytics

Data analytics is used;

- To predict future sales and purchasing behaviors

- To evaluate the effectiveness of marketing and advertising campaigns
- To optimize marketing efforts through more accurate targeting and personalization
- To identify and eliminate bottlenecks within a certain process
- To boost customer acquisition and retention
- To boost customer engagement on social media
- To develop risk management solutions
- To identify opportunities for innovation and create new revenue streams
- To increase supply chain efficiency.

2.3. Data-driven decision-making

Data-driven decision-making is a data-driven approach for making decisions including business decisions. This data can come from data analysis, data visualization, or other data resources. Data-driven decision-makers use data in their decision process and they make decisions based on the actionable insights generated from the data. The goal is to make informed decisions while ensuring transparency across the stakeholders.

2.3.1. Types of decisions

2.3.1.1 Programmed decisions

Programmed decisions are the decisions that are taken based on some hypothesis and then the decisions are tracked over a period of time for their effectiveness. These are long-term decisions that can span over years and often form part of business strategy. Programmed decisions can leverage **descriptive, diagnostic & prescriptive analytics** tools & methodologies for measuring the impact based on KPIs & benchmarks. Descriptive analytics is a form of analytics that aims to describe the current state of the business. In simpler words, descriptive analytics helps understand what has happened. Similarly, diagnostic analytics is a form of analytics that helps in diagnosing the problem. It helps to answer question such as why does the problem exist? What are the levers that can be pulled to solve the problems? Also, prescriptive analytics can be used to identify the levers that could impact the outcome of decisions. These levers can become part of decision which can be tracked for a given time period as part of simulation exercise. Accordingly, different outputs of optimization models can be used to create optimal impact. The primary output of the descriptive and diagnostic analytics is dashboards representing actionable insights.

The dashboards can be built by consolidating data from different sources including internal and external sources. The dashboards help track actions using KPIs & benchmarks. The dashboards can be institutionalized later into organization appropriately. Programmed decisions can also be termed as policy decisions at times as these results in long-term policy definition and implementation.

The following represents different forms of programmed decisions:

A. **Marketing decisions:** is a type of programmed decisions which can be made for a given time period and the impact can be measured by tracking KPIs in order to measure the impact on sales.

B. **Manufacturing decisions:** Manufacturing decisions can also be seen as one type of programmed decision. These decisions can leverage output of prescriptive analytics to arrive at the decision points.

2.3.1.2. Un-programmed decisions

These are short-term decisions that need data to be analyzed in order to make a decision. Un-programmed decisions, at times, can also be termed as tactical and/or operational decisions. Un-programmed decision can leverage **predictive** and **prescriptive analytics**.

2.3.1.3. Organizational decisions

These are data-driven decisions that need data to arrive at the correct decision. These data can be used by multiple people in an organization and they will make similar types of un-programmed or programmed decisions based on them.

2.3.1.4. Strategic decisions: These data-driven decisions are used to make long-term planning in the organization.

2.3.2. Data-driven decision-making Approach

Data-driven decision-making is an approach to made business decisions based on the actionable insights derived from the data. Identifying the business problems and appropriate data analytics solutions holds the key to effective data-driven decision-making. In data-driven decision making, one uses data in the decision process and the decisions are made based on the data collected from one or more data sources. The goal is to make informed decisions rather than quick ones. Data-driven decision-making requires different systems to be put in place, such as data collection, data processing, data visualization, etc.

The following represents the key aspects related to data-driven decision making:

- Identify problems/opportunities to be dealt with using analytics
    - Break down problems into sub-problems for detailed analysis
- Determine the levers in form of data that can be part of the solution to the problem. These levers when pulled can impact the business. Determine the data which can represent these levers. The data can be both internal and external.
- Come up with analytical solutions which can be used to track the hypothesis using KPIs. Dashboards are the most common form of the analytical solution.
- Track the KPIs, take the decision, and act
- Measure the outcome using KPIs

Data-driven decision-making requires defining KPIs, and benchmarks, tracking the KPIs, and driving actions appropriately. KPIs can be of the following different types:

- **Business KPIs or Value metrics** – These help in measuring the performance of data-driven decision-making systems in relation to the business impact made by data-driven decisions. For example, a revenue growth rate is an indicator of data-driven decision-making because it shows whether decisions are being made to increase sales or not. Business KPIs can also be termed as lagging indicators or lagging KPIs. These KPIs can also be called projects outcome KPIs.

- **Customer KPIs** – These help in understanding data-driven decision-making from a customer perspective. For example, if the data shows that customers are complaining about an issue and it is not being addressed quickly, then one would need to use this insight and drive actions to address the customer complaints. Customer KPIs can also be termed as leading indicators or leading KPIs.

- **Project output KPIs**: Project outputs KPIs are data-driven decision-making KPIs that are used to measure the success of projects. Examples of project output KPIs include the number of user stories completed on time, system availability, system performance, etc.

2.3.3. Systems to facilitates data-driven decision-making

- **Data collection or ingestion**: The data collection systems enable the data analysts/scientists to capture data points that are required for analysis. The data should be captured in a way that it is not biased towards any particular outcome, and can be used by anyone who has an access to data. Example of data ingestion systems includes data ingestion using data pipelines, data warehouses, etc.

- **Tagging and annotation**: Tagging allows the analysts/scientists to annotate each data point with a set of tags that provide enough details about the data point for use by other analytics professionals. The tagging system should be easy to use so as not to interrupt workflow processes across teams/organizations. At the same time, data points should be tagged with the most appropriate tags.

- **Data processing:** The data collected would need to be processed so that it can be used for decision-making. This is where data processing systems come into the picture. Example of data processing systems includes big data processing engines such as Hadoop, and data processing tools like R and Python for data analysis.

- **Data visualization**: The data analysis scientists would need to figure out the data insights required for decision-making, and how they can be used by business leaders. This is where data visualization comes in place which enables visualization experts to share insights with the stakeholders through visually appealing graphs, charts, plots, etc. This is where tools such as the Qliksense dashboard come into the picture.

- **Decision model**: Data scientists would come up with the decision models which can be used by business stakeholders to make data-driven decisions. An example of a decision model includes the deployment of predictive analytics systems for data-driven decisions.

2.3.4. Advantages of data-driven decision-making

- Helps organizations to achieve their goals by enabling stakeholders to make informed decisions in a timely manner based on actionable insights, and gain a competitive edge with greater speed, agility, and scale (SAS).
- **Faster, better & informed decisions**: Better business decisions are made by considering both data and human intuition, which leads to better results for companies. It reduces the bias introduced due to gut feeling.
- **Increased data visibility**: Data-driven decision-making helps in increasing data transparency as the data is available to multiple people and it can be easily shared.
- **Increased data accessibility**: Data-driven decision-making helps in creating data access for everyone, not just a single person or department. This leads to increased data sharing and reduces the chances of mistakes being made due to a lack of information.
- **Increased data quality**: Data-driven decision-making ensures data quality is increased by removing data redundancy and errors. This leads to more accurate data collection and higher precision in the analysis of the data.
- **Efficient use of time**: With data-driven decision-making, companies can save on time and resources as human input and associated risks are reduced significantly.
- **Increased data usage**: Data-driven decision-making ensures data is used in a better way, for example, data can be easily shared and reused. This leads to increased data value as more people use the same data which increases its worth over time.
- **Enhanced data-literacy**: Data-driven decision-making helps data professionals to improve their data literacy skills, which in turn will lead them towards better career opportunities.

- **Increased customer satisfaction**: When multiple data points are collected about customers, companies can come up with various ways to increase their overall satisfaction levels which will lead to increased retention rates and ultimately better future revenue for the company.

2.3.5. Disadvantages of data-driven decision making

- **Confirmation bias**: The business stakeholders might tend to use data points that support their decision-making and ignore the ones that may go against it. This leads to decisions driven by confirmation bias, which can lead to poor business outcomes in the future due to wrong data analysis.

- **Data leakage**: The data collected for decision-making needs to be safeguarded so that other organizations cannot get access to it and use it against business interests.

- **Data bias**: Different data sources may have data points with different levels of biases that need to be taken into account while deriving any data insights from them.

- **Unforeseen consequences**: The data analysis scientists fail to take into account other factors such as data quality, data collection errors, etc. while coming up with data insights which can lead to unforeseen consequences in the future because of wrong data analysis.

2.4. Business Intelligence

Business Intelligence (BI) is a set of processes, architectures, and technologies that convert raw data into meaningful information that drives profitable business actions. It is a suite of software and services to transform data into actionable intelligence and knowledge.

BI has a direct impact on organization’s strategic, tactical and operational business decisions. BI supports fact-based decision making using historical data rather than assumptions and gut feeling.

BI tools perform data analysis and create reports, summaries, dashboards, maps, graphs, and charts to provide users with detailed intelligence about the nature of the business.

2.4.1. Importance of BI

- Measurement: creating KPI (Key Performance Indicators) based on historic data
- Identify and set benchmarks for varied processes.
- With BI systems organizations can identify market trends and spot business problems that need to be addressed.
- BI helps on data visualization that enhances the data quality and thereby the quality of decision making.
- BI systems can be used not just by enterprises but SME (Small and Medium Enterprises)

2.4.2. Steps to implement Business Intelligence systems

1) Raw Data from corporate databases is extracted. The data could be spread across multiple systems heterogeneous systems.

2) The data is cleaned and transformed into the data warehouse. The table can be linked, and data cubes are formed.

3) Using BI system the user can ask quires, request ad-hoc reports or conduct any other analysis.

2.4.3. Types of BI users

The Following are four key users of Business Intelligence System:

1. The Professional Data Analyst:

The data analyst is a statistician who always needs to drill deep down into data. BI system helps them to get fresh insights to develop unique business strategies.

2. The IT users:

The IT user also plays a dominant role in maintaining the BI infrastructure.

3. The head of the company:

CEO or CXO can increase the profit of their business by improving operational efficiency in their business.

4. The Business Users

Business intelligence users can be found from across the organization. There are mainly two types of business users

1. Casual business intelligence user
2. The power user.

The difference between both of them is that a power user has the capability of working with complex data sets, while the casual user need will make him use dashboards to evaluate predefined sets of data.

2.4.4. Advantages of BI

Here are some of the advantages of using Business Intelligence System:

1. Boost productivity

With a BI program, It is possible for businesses to create reports with a single click thus saves lots of time and resources. It also allows employees to be more productive on their tasks.

2. To improve visibility

BI also helps to improve the visibility of these processes and make it possible to identify any areas which need attention.

3. Fix Accountability

BI system assigns accountability in the organization as there must be someone who should own accountability and ownership for the organization's performance against its set goals.

4. It gives a bird's eye view:

BI system also helps organizations as decision makers get an overall bird's eye view through typical BI features like dashboards and scorecards.

5. It streamlines business processes:

BI takes out all complexity associated with business processes. It also automates analytics by offering predictive analysis, computer modelling, benchmarking and other methodologies.

6. It allows for easy analytics.

BI software has democratized its usage, allowing even nontechnical or non-analysts users to collect and process data quickly. This also allows putting the power of analytics from the hand's many people.

2.4.5. BI System Disadvantages

1. Cost:

Business intelligence can prove costly for small as well as for medium-sized enterprises. The use of such type of system may be expensive for routine business transactions.

2. Complexity:

Another drawback of BI is its complexity in implementation of data warehouse. It can be so complex that it can make business techniques rigid to deal with.

3. Limited use

Like all improved technologies, BI was first established keeping in consideration the buying competence of rich firms. Therefore, BI system is yet not affordable for many small and medium size companies.

4. Time Consuming Implementation

It takes almost one and half year for data warehousing system to be completely implemented. Therefore, it is a time-consuming process.

2.4.6. Trends in Business Intelligence

The following are some business intelligence and analytics trends that you should be aware of.

Artificial Intelligence: Gartner' report indicates that AI and machine learning now take on complex tasks done by human intelligence. This capability is being leveraged to come up with real-time data analysis and dashboard reporting.

Collaborative BI: BI software combined with collaboration tools, including social media, and other latest technologies enhance the working and sharing by teams for collaborative decision making.

Embedded BI: Embedded BI allows the integration of BI software or some of its features into another business application for enhancing and extending its reporting functionality.

Cloud Analytics: BI applications will be soon offered in the cloud, and more businesses will be shifting to this technology. As per their predictions within a couple of years, the spending on cloud-based analytics will grow 4.5 times faster.

2.5. Targeting

In general targeting can be defined as "the act of ensuring that aid reaches people who need it at the right time and place, in the right form and quantity and at the same time does not go to people who don't need it", (Maxwell et al (2009:3). Such processes involve selection of bigger location and start narrowing down to specific location, community, and individuals. This way, scarce resources can reach those that need it the most. To achieve this, different targeting approaches and criteria were in use.

The criteria can be based on characteristics of individuals, community, and or geographic places. In the process there are many errors (inclusion and exclusion errors). Inclusion errors occur when people who should not benefit from aid program get the benefit while exclusion errors happen when individuals who would have benefited from aid program are excluded. Determining efficient and effective targeting criteria and approach that minimize targeting errors is challenging and are key[2] success factors. Lavalee et al, (2010) recommended considering factors like physical ability (disability), cultural, ethnic, religion, caste, and age discrimination, those who are already trapped by previous crisis and again hit by other shocks such as hurricane, earthquake, war, or economic recession subsequently.

[2] https://timothyschwartzhaiti.com/a-model-for-humanitarian-aid-beneficiary-targeting

Traditionally, beneficiaries are targeted by assessing geographic factors like population density (Urban, peri-urban, and rural), economic zone (low income areas or farming, pastoral, hunting, mining, logging), ecological zone (littoral dry, humid plateau, humid mountain, dry mountain), ethnic or socio-cultural/linguistic groupings, infrastructural and service conditions (availability of services such as schools or infrastructure such as water, roads and electricity), and or security situation/restrictions (conflict or high crime areas, zones of guerilla activity).

Once priority areas are identified through one or combination of the above criteria the next step is to actually identify individual beneficiaries. This can be done in a number of ways: self-targeting, administrative, or community targeting. Each of these has its own limitations and there is no one best way[3]. It is rear to get/ see justifications for proper targeting vulnerable households. Whatever the approach is, the task of targeting has always limitations (Jaspers & Shoham, 1999; Schwartz, Maass, & DeMattee). Success metrics is determined by how well the targeting mechanism is designed, implemented and reduced both inclusion and exclusion errors. Study by (Clay, Molla, & Habtewold, 1999) indicated that there is huge error of inclusion and exclusion while targeting for food aid. Besides these errors traditional targeting approaches involve huge administration costs, corruptions, and unfair governance. None of these approaches are good enough if deployed independently (Sharp, 1998).

In poor countries statistics is used less and data for statistical analysis is not captured much[4]. In much of Africa, for instance, national statistics on economic production may be off by as much as 50 percent (Jerven, 2013). Globally, in the decade between 2004 and 2013, at least 74 countries had no more than one estimate of the national distribution of poverty (Serajudding et al., 2015). This calls for a new, innovative, efficient, and effective approach.

2.5.1. Community based targeting

Targeting datasets can be gathered by establishing a dedicated data collection and curating centers or programs like (Kaba, et al, 2018; Raj & Relton, 2013, www.idpoor.gov.kh)[6], which is

[3] http://www.fao.org/3/y5702e/y5702e0a.htm#bm10
[4] https://didl.berkeley.edu/projects/

[5] Program established to identify aid beneficiaries in Cambodia

based on community participation and rating approach. Identifying trusted individuals (men and women in the community who are the center of gravity is another approach (Foster, 2021). These individuals can give their genuine and honest assessment about their community that can be considered for targeting. This approach has a 95% success rate and appears to be more successful in selecting households that are, on average, more ‘vulnerable’ than others as measured by their food (in)security status, proxies of wealth, and levels of education. But it doesn’t appear to be cost efficient. Decentralized decision making by community leaders can be misleading and crates variations from place to place. There is also no reason to believe that local notions and understandings of poverty and vulnerability at the community level will be the same as those developed by a national government agency or development partner (Conning & Kevane, 2002). Hence this lack of uniformity and objective assessment results in variation. It also lacks transparency, discriminatory practices, exclusion of the poor considered ‘undeserving’, and elite capture (McCord, 2013).

The involved costs (administrative costs, private costs, social costs, psychological costs, incentive costs, political costs) and addressing complaints and governance is also another challenges of community based targeting (Devereux et al, 2015). But this approach is valuable and appears to be legitimate by the community as it brings community knowledge to the targeting process that is inaccessible in other forms of targeting.

2.5.2. Geographic targeting

Geographical targeting is prioritizing assistance to particular locations. It looks into nature of geography like arid or semiarid. Such approach is also misleading. There are people living in arid/semiarid places and yet are not as poor as people living in other places. Study in Kenya[7] revealed 43% - 64% of the total number of acutely and chronically malnourished people and food insecure households in the country live in non-arid and semi-arid land areas and are excluded from aid scheme by design. Targeting using administrative data (international aid agencies) require having survey of the entire target population (Altındağ, et al, 2021). But with respect to cost minimization, policymakers should consider targeting methods that maximize the

[6] Program established to identify aid beneficiaries in Cambodia
[7]1524574523.Report–Assessment-of-targeting-of-asset-creation-programme-v1-1.pdf

use of existing data, and should additionally provide scope to include new features into administrative data should they be identified as beneficial for program targeting.

In poor countries statistics is used less and data for statistical analysis is not captured much[8]. In much of Africa, for instance, national statistics on economic production may be off by as much as 50 percent (Jerven, 2013). Globally, in the decade between 2004 and 2013, at least 74 countries had no more than one estimate of the national distribution of poverty (Serajudding et al., 2015). This calls for new, innovative, efficient, and effective approaches.

2.6. Artificial Intelligence and Machine Learning for Humanitarian Action

These days the use of proxy measurements is growing. For example, predicting floods by analyzing Flickr tags (Tkachenko, Jarvis, Procter, 2017), analyse business recovery from natural hazars (Eyre, De Luca, & Simini, 2020; Lam et al, 2022), analyze disaster damage by analyzing social media (Twitter datasets) (Kryvasheyeu, 2016), analyzing twitter data to understand disaster resilience (Zou et al, 2018) by analyzing spatial temporal patterns of twitter activities of people during hazards, by analyzing Facebook usage patterns developing disaster maps of people looking for aid (Maas, 2019) etc. Proxy based measurement or assessment is becoming common.

As discussed above, survey based targeting households is more accuratebut expensive. Alternative to this, researchers from UC Berkeley's Center for Effective Global Action and Facebook's Data for Good and Stanford University have explored and developed an alternative targeting approach that is based on average wealth of each location. Specifically, they developed relative wealth index of places at 2 by 2 kilometers for 135 low income countries.

Their wealth index combines publicly available survey data with nontraditional, predictive data, including high-resolution satellite imagery, topographic maps, mobile network data, and connectivity data. According to them, counties with high population densities, lower road infrastructures, fewer smart phones, Wi-Fi, lesser phone call durations, lesser mobile internet usage suggest lesser relative wealth of communities than counties wither higher values for the

[8]https://didl.berkeley.edu/projects/

previous variables. Thus, by computing and comparing these values for different counties, governments and aid agencies can target their development and humanitarian aid.

The above datasets and current processing algorithms can help government and organizations to be able to measure poverty and vulnerability in real time. Let's imagine the efficiency and effectiveness of social media platforms like Facebook and search engine firm like Google in their ad targeting; they are almost spot on! This is can be a good inspiration to explore the use of ML/AI techniques in humanitarian targeting[9]. But with respect to applying in developing countries, the use of night time satellite data and mobile data appears to be feasible. Majority of counties who seek aid are not connected to electricity grid and hence there is no night time satellite imagery. Mobile data at mobile network operator cannot be used directly as phone numbers (codes) are not earmarked against their respective geographic locations.

Humanitarian organizations should change current system from one that reacts, to one that anticipates (Arendt-Cassetta, 2021). Developments in new and emerging technologies can support this paradigm shift from reaction to anticipation by enabling earlier, faster and potentially more effective humanitarian action. Technologies like AI can facilitate analysis and interpretation of vast and complex datasets to improve projections and decision-making, mobile applications, chat bots and social media can create immediate feedback loops with people affected by humanitarian crises. Unmanned aerial vehicles (UAVs) and remote sensing can speed up the assessment, mapping and monitoring of vulnerabilities. Together these technologies can lead to better access to information, assistance and livelihoods, and facilitate stronger, more relevant needs analysis, a more prioritized and people-centered response, and more meaningful and systematic monitoring.

2.7. Related Works

As an alternative solution to the above challenge, machine learning and artificial intelligence scholars have started exploring the use of datasets like night time satellite images of places and mobile phone datasets (Blumenstock 2016). Algorithms can play a significant role in identifying targets to provide aid to. Unlike human officers, computer algorithms can continuously track and

[9] https://didl.berkeley.edu/portfolio/toward-real-time-measures-of-poverty-and-vulnerability

capture the activities (behaviour) of individuals (Abebe & Goldner 2018). Such objective behaviour tracking in a more continuous manner helps aid organizations.

Specifically, researchers from UC Berkeley's Centre for Effective Global Action, Stanford University, and Facebook's Data for Good have explored alternative targeting approaches that are based on the average wealth of each location. They developed a relative wealth index of places at 2 by 2 kilometres for 135 low-income countries. The wealth index combines data like demographic health survey (DHS)

data, satellite imagery, topographic maps, mobile network data, and connectivity data. According to them, counties with high population densities, lower road infrastructure, fewer smartphones, Wi-Fi, shorter phone call durations, and lower mobile internet usage suggest a lower relative wealth of communities than counties with higher values for these variables. By computing and comparing the values of these variables for different counties, governments and aid agencies can target their development and humanitarian aid. But in developing countries like Ethiopia, the majority of counties that seek aid are not connected to the electricity grid, and hence there is no night time satellite imagery.

Besides the use of night time satellite images and mobile phone datasets, there are other insightful proxy measurements (indirect measurements) in the areas of artificial intelligence for social goods. For example, predicting floods from Flickr data tags (Tkachenko, Jarvis, and Procter 2017); analysing business recovery from natural hazards from Facebook datasets (Eyre, De Luca, & Simini, 2020; Lam et al. 2022); analysing disaster damage from Twitter datasets (Kryvasheyeu et al. 2016); understanding disaster resilience from Twitter datasets (Zou et al. 2018); and developing disaster maps of people looking for aid from Facebook usage patterns (Maas 2019). These studies show how the use of proxy measurements (assessments) is becoming common.

In a similar analogy, behavioural economics and finance literature offer great insight and proxies to measure poverty and or wealth levels and hence assist in targeting. According to Karlan, Ratan, and Zinman (2014), the poor can save at banks to avoid the temptation to spend it on unnecessary things, and being stolen by close friends and individuals (Banerjee and Duflo 2006; Hoos 2010; Reyers 2019) and as a buffer against financial shocks. Reyers (2019) found that financial capability is highly associated with saving for emergencies. Those with greater financial capability and those with access to a bank account are more likely to have emergency savings compared with those with lower levels of financial capability.

During emergencies and hardship times, individuals have lower saving amounts (Eskander et al. 2018) and more withdrawals (Brei, Mohan, & Strobl 2019) from their accounts. This means bank transactions like deposits, withdrawals, transfers, etc. are indicative of the wealth or poverty of a country. In light of this knowledge and evidence, by looking at bank transaction datasets and their trends over time, this study documented interesting insights that can help decision makers prioritize counties for targeting. Even better than satellite night time light imagery datasets, bank account financial activities of individuals can reveal the reality on the ground in a real-time and granular manner. These datasets are also captured automatically by bank infrastructure. The following section describes the nature of bank transaction datasets and the analysis or insights derived from them.

CHAPTER THREE

3. Methodology

This section briefly presents a general description of the data and the empirical methodology used in this study. System generated Secondary data is employed in the present study for the period of two years 2020 and 2021. MIS department of the Commercial Bank of Ethiopia is the major source of data. Aid areas and summary of transaction variables used in this study are listed in the table 1 blow.

Table 1: Aid places and Bank Transaction Types

Transaction Place	AC	ATM WithD	Ck WithD	Deposit	Mobile Transfer	Withdrwal	Bus Deposit in Hierarchy	Bus Deposit out Hierarchy	Bus Transfer	LMTS Transfer	Bus transfer via web	IB_Txn	IFB LMTS Transfer	MB AgentDeposit	POS_Txn
Dejen	x	X	x	x	X	x	x	x	X	X	x	x	x		x
Beyeda	x	X	x	x	X	x	x			X	x	x			x
Dabat	x	X	x	x	X	x	x	x	X	X	x	x	x		x
East Belda	x	X		x	X	x	x	x	X	X	x				x
Metema	x	X	x	x	X	x	x	x	X	X	x	x			x
Dehana	x	X	x	x	X	x	x	x	X	X					
Mehal Meda	x	X	x	x	X	x	x	x	X		x				x
Menz	x	X	x	x	X	x							x		x
Gubalafto	x	X	x	x	X	x	x	X	X	X	x	x			x
Habiru	x	X	x	x	X	x	x	x	X	X		x	x	x	x
Meket	x	X	x	x	X	x	x	x	X	X	x	x			X
Raya	x	X	x	x	X	x	x	x	X	X		x	x	x	x
Ataye	x		x	x	X	x									
Ebinat	x	X	x	x	X	x	x	x	X	X	x				x
Wadla	x	X	x	x	X	x	x	x	X			x			X
Haik	x	X	x	x	X	x	x	x	X			X	x		x
Jamma	x	X	x	x	X	x	x	x	X	X	x	x	x	x	
Legambo	x	X	x	x	X	x	x	x	X	X		x	X		x
Borena	x	X	x	x	X	x	x	x	X		x	x	x		X
Tehuledere	X	X	x	x	X	x	x	x	X			x	x		X
Werebabu	x	X	x	x	X	x	x		X			x	x		X
Wereilu	x	X	x	x	X	x	x	x		X		x			X
Merabete	x	X	x	x	X	x	x	X		X					x
Tenta	x	X	x	x	X	x	x	x			x	x	x	x	X
Gayint	x	X	x	x	X	x	x	x			x	x			x

3.1. Novelty of research datasets

As stated in the literature almost all targeting researches use night time satellite imagery, mobile phone data, and demographic and health (DHS) datasets. With respect to developing countries, these datasets have some limitations. Thus, in this paper we resorted to use bank account transaction datasets for targeting purpose. These kinds of datasets are real time, cost efficient (cheap as they are collected by banks banking applications). Such datasets are also genuine as they are far from manipulation.

With over 37 million active customers all over the country the commercial bank of Ethiopia dominates the financial service in the Amhara National Regional State without any potential competitor and based on United Nations Office for the Coordination of Humanitarian Affairs (OCHA 2020) reports about the number of individuals seeking aid, 25 bank branches operating at the most in-need aid sites are selected. The branches made over 23 million daily transaction records, with an amount of over 18 billion Ethiopian Birr in 2020 and 2021. But, in this paper, we present findings based on datasets of these bank branches operating in the most aid seeking counties in the Amhara region of Ethiopia.

The majority of the branches have transactional datasets like AC (account to account transfer), deposit (saving), withdrawal, mobile transfer, internet banking, ATM withdrawal, cheque withdrawal, POS (Point of Sale), local money transfer (LMTS), local money transfer at interest free, and business transfer over the web. Based on literature and our own understanding about which transactions are informative about the poverty of communities around those bank branches, we analyzed deposits, withdrawals, account transfers, ATM withdrawals, mobile transfers, and POS transactions. Table 2 gives a brief explanation of why we selected these transactions and their respective meanings in the context of banking practices in Ethiopia. The following section elaborates on the steps followed in data processing.

Trans type	Transaction mode	Brief explanation
AC: Account to account transfer using transfer form	It needs physical presence of account holder at the bank's branch	The account may/may not be opened at that specific location. But it shows money transfers took place at that particular area branch. This can show more volume of AC means communities around that specific place are better than other places.
Deposit: Cash deposit using deposit voucher	It needs physical presence of the depositor at the bank's branch	The account may/may not be opened at that specific place. But it shows physical cash is deposited at that particular area branch. Higher deposit amounts means the communities around such bank branches are economically better off than others.
Cash withdrawal using Withdrawal voucher	It needs physical presence of account holder at bank's branch	The account may/may not be opened at that specific place. But it shows physical cash is withdrawn from that particular area branch. This can show higher withdrawal amount means communities around such bank branches are economically better off than others with lower withdrawal amounts.
Account to account transfer using mobile banking	It doesn't require physical presence of individuals.	It shows the account holder needs to open a bank account at that specific bank branch. But transfers can be made from anywhere in the country. The higher the volume of such transactions, the wealthier the surrounding communities are (in relative terms).
Cash withdrawal from ATM	Needs physical presence of card holder at bank	The account may/may not be opened at that bank branch. But the transaction took place at that particular area branch. This shows the higher the ATM withdrawal amount, the better the community around such bank branches.
POS: Cash Withdrawal /money transfer using POS	It needs physical presence of ATM card holder at bank branch	The account may/may not be opened at that specific branch. But cash is withdrawn or transferred from that bank branch. This shows the higher the volume of POS the better the wealth of the communities around such bank branches.
LMTS: Local money transfer	Need physical presence of a person who transfers	It doesn't touch any accounts; just physical cash is transferred from one person to another person through the bank. This means the higher the volume of LMTS the better the wealth of the community around such bank branches.

Table 2. Bank transaction types and descriptions

3.2. Data Transformation steps

3.2.1. Dataset Regularization

Data transformation is the process of converting raw data into a format or structure that would be more suitable for model building and also data discovery in general. It is an imperative step in feature engineering that facilitates discovering insights. Data transformation helps to minimize algorithm bias when the data distribution is skewed[10]. Bank transaction datasets are time series (value-time paired) that can form trends or seasonality. Observations are assumed to be continuous in time and it allows only one path. But bank transaction datasets are not continuous, as banks do not operate on holidays and Sundays. To minimize adjusted loss function and prevent over fitting or under fitting, data is regularized by using interpolation technique. By using regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it. Time Series analysis helps to extract meaningful characteristics of the data. Models assume current values will be expressed as driving in some ways from past values.

3.2.2. Calculate Percentage Change Trends

Different counties/ places do have different number of population. Thus, the amount and number of say deposit transactions vary from place to place. In order to compare trends or seasonality of transactions across different counties over the 24 months' time period, we need to transform the data. The best approach is to find the ratio of monthly transactions with respect to the first month. This should be done for all transaction types. This gives us percentage changes over the 24 months. The assumption is that if places/ communities experience life shocks like drought, wild fire, conflict, flooding etc, they will have a decreasing amount in their deposit behavior.

3.2.3. Visualize the seasonality of key bank transaction types

Once step one and two above are executed we can compare different counties for targeting purpose. Predictive Analytics

Advanced analytics and artificial intelligence has been seen being applied in the humanitarian domain. Predictive analytics, machine learning, and statistical models are used to calculate the probable characteristics of humanitarian emergencies is growing rapidly. For example, https://www.wfp.org has developed a Hunger Map showing food security in real time. The map uses ML based predictive models to monitor and assess food security in real time. In the absence of real data, the agency uses predictive models to estimate future food security. The use of

[10] 3 Common Techniques for Data Transformation | by Destin Gong | Towards Data Science

predictive analytics to streamline waste collection in refugee areas in Jordan by UNICEF[11,] to identify children who are more likely to default from their subsequent immunization visits[12,] real-time analytics, alerts and expert insight into water and energy systems[13], to predict migrations and refugees[14,] to forecast flooding[15].

These technologies are being used to forecast the likely trajectory and features of humanitarian emergencies including pandemics, famines, natural disasters and refugee movements (Hernandez & Roberts, 2020). This form of artificial intelligence is used to predict where and when disasters will unfold, what the defining characteristics of the situation will be and who will be the most affected populations. Researchers showed that predictive analytics could be used to determine which counties/ communities are most likely to face drought, flood, conflict, and or desert locust.

In the same line of arguments made above, we constructed forecasting model for bank transaction types for the respective counties. Forecast the number of aid seekers, detect aid beneficiaries that doesn't deserve (suspicious/ fraudulent individuals), predict if an individual is likely for an aid beneficiary action (probability of getting the aid), recommend if someone deserves an aid (recommendation), separate (cluster) groups of individuals for targeting aid beneficiary program (as highly, medium, low) aid seekers.

Data Description and Preprocessing: Dejen Site

Worda	Date	Trans	Count	Amount
Dejen	Fri 1 Jan 2021 00:00:00	AC	105.0	21316165.
Dejen	Fri 1 Jan 2021 00:00:00	ATM Withdrwal	797.0	806429.
Dejen	Fri 1 Jan 2021 00:00:00	Cheque Withdrwal	14.0	246621.
Dejen	Fri 1 Jan 2021 00:00:00	Deposit	372.0	2922435.

Table 3: Sample dataset format, after imported into the processing software

[11] "Smart Refugee Camps": applying the best of IoT and ICT for better camp management | by UNICEF Jordan | Medium
[12] JMIR Public Health and Surveillance - Using Predictive Analytics to Identify Children at High Risk of Defaulting From a Routine Immunization Program: Feasibility Study
[13] Oxfam (sensorinsight.io)
[14] How to Predict the Next Refugee Crisis (businessinsider.com)
[15] Google AI Blog: An Inside Look at Flood Forecasting (googleblog.com)

3.3. Handling Outliers

In financial modeling and analysis, a seemingly outlying observation may not really be an outlier, even though we think it to be so. Until we model and understand the nature of the dataset, we cannot decide it to be an outlier and eliminate from analysis. Strictly speaking, an outlier should be classified as such based on the scientific reasons, and the fact that it is "out there" is merely a manifestation of an underlying reason that disqualifies the observation from being in the "group" to begin with. And that reason should be an exogenous factor to the modeling environment.

An outlier may have the most salient information. Oftentimes it's actually the outlier itself that points you to where the problem is. Sometimes you are actively looking for outliers, finding outliers are the goal! Remember: never remove an outlier without a scientific reason. Don't remove outliers for being outliers. You need a scientific reason that justifies the removal, and that depends on the purpose of the model ("find problems" vs. "build model with homogenous data"). Sometimes you want only homogenous data, and sometimes you want to diagnose problems.

There is Wolfram inbuilt functions that help us detect outliers. Box Plot, distribution chart, and Find Anomalies. In descriptive statistics, the box plot is a method for graphically depicting groups of numerical data through their quartiles. Box plots may also have lines extending vertically from the boxes (whiskers) indicating variability outside the upper and lower quartiles, hence the terms box-and-whisker plot or box-and-whisker diagram. They are non-parametric, and they don't make any distributional assumptions. The width of the bands, and what is displayed with the whiskers, is not standardized, and there are a lot of variations. But typically, the quartiles, and the band inside the box, are the second quartile (the median) the ends of the whiskers can represent several possible alternative values.

https://en.wikipedia.org/wiki/Box_plot

It is statistical malpractice to use the box plot on skewed data! In my opinion the box plot can be thought of as a sub function of the histogram. It's like a "data summary" with most of the "middle information" removed. It is designed to accentuate extreme data, according to definitions related to quantiles / quartiles. BUT: If you know your data has a lot of observations that would exceed these quantile-based definitions, then the box plot is pointless to use, in fact, misleading.

We must always look at the histogram first (quick at-a-glance non-parametric overview)! If the data is not skewed, the box plot may be a suitable next step ("quartile summary"). If the data is skewed, don't base any decisions on any box plot-derived outliers! Distribution Chart: displays a distribution chart with a distribution symbol for each data. This is ugly, but it did its job: detect the outlier(s) (Hastie et al 2009). Find Anomalies also does a good job.

CHAPTER FOUR

4. ANALYSIS AND FINDINGS

This chapter mainly deals with the results of processed data in a way that it will be presented to show the extent to what level it will give meaningful information to support the research objective.

4.1. Deposit Trend Analysis

We can visualize and analyze each transaction for each site. As an example, a simple histogram and scatter plot visualization revealed that there are some elements which appear to be an outlier. But, unlike many machine learning projects, we do not remove such values from the dataset.

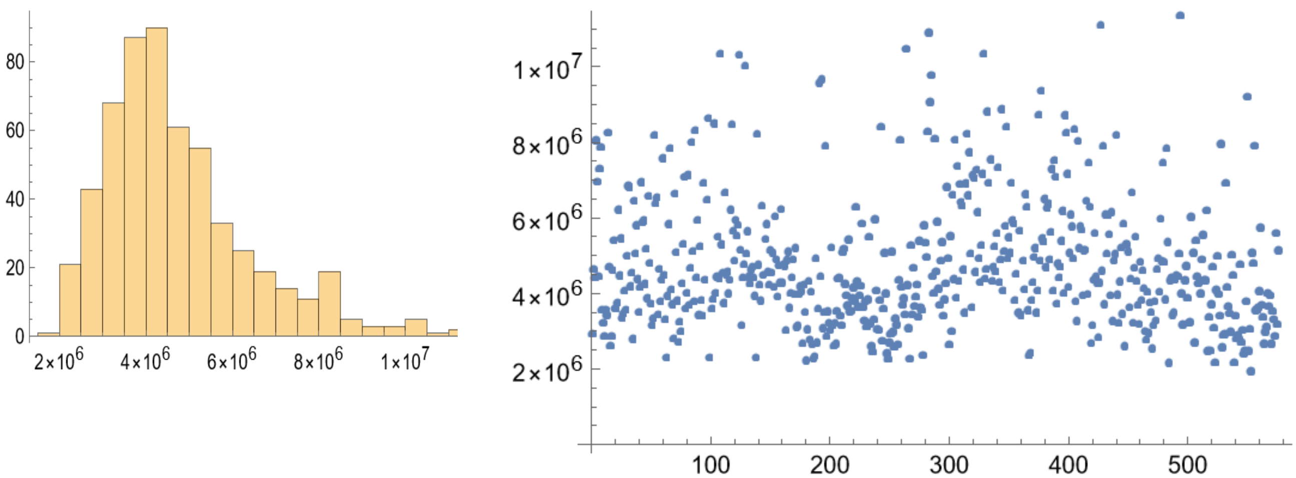


Figure 1. Exploring the distribution of the dataset Deposit transactions for Dejen Site

The next obvious step is to construct time series objects and visualize the distribution of the datasets. For all the sites, we can construct two time series data objects: for transaction counts and for transaction amounts. Then we can visualize how the number of transactions. But before, we visualize the datasets, it is important to regularize it. As it stands, data is not regular. Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent over fitting or under fitting. Using

Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it. In this paper, we regularized the datasets into days.

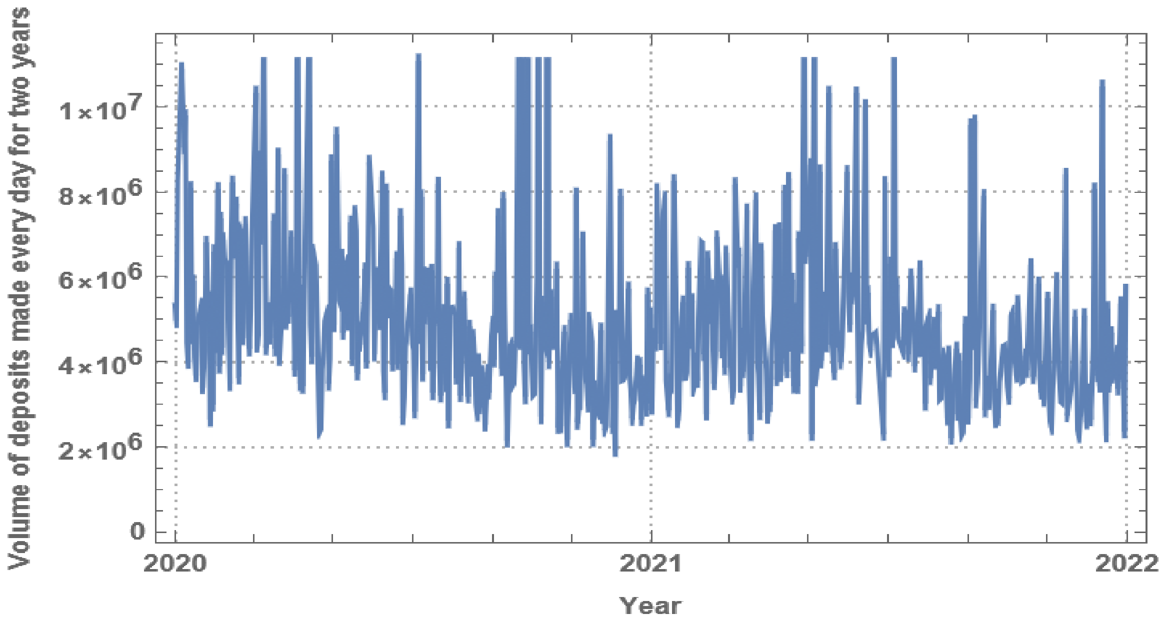


Figure 2. Volume of deposit made every day for two years

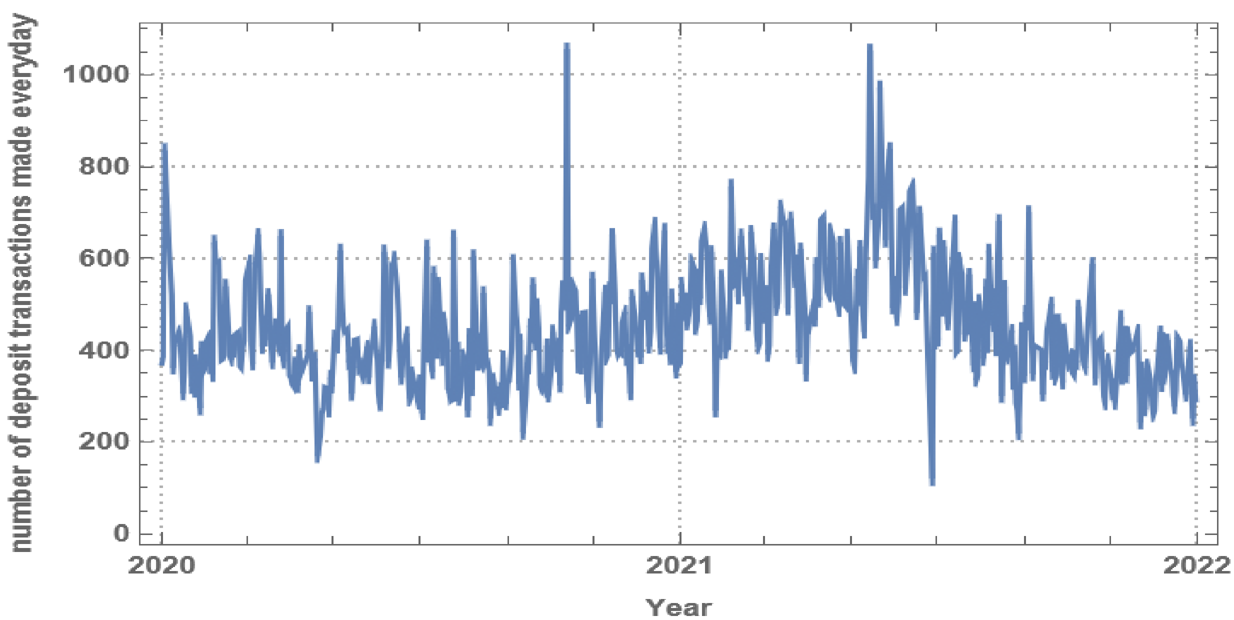


Figure 3. Number of deposit transactions made every day for two years

As it can be seen from the figures above, both the deposit amounts and its counters (frequencies) form kind of seasonal trend. In order to compare deposits across the two years with other research sites, we calculated the ratio of deposit amounts to number of transactions. Here is the result.

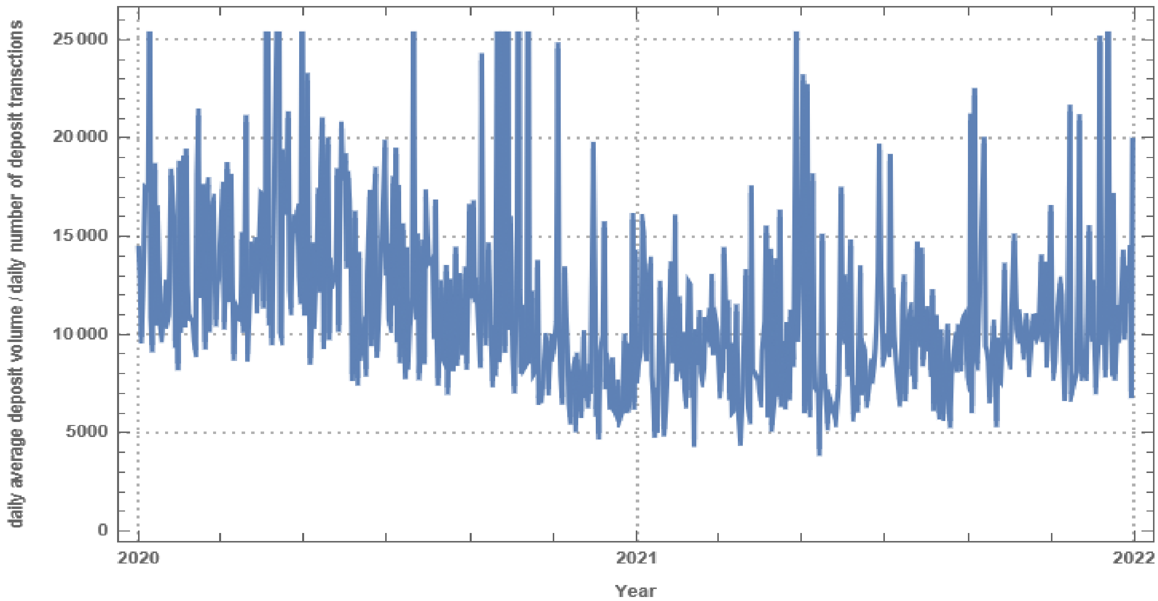


Figure 4. Ratio of daily average deposit volume per average daily transactions

The plot shows again a kind of trend. Deposit amount per transaction count shows a kind of reduction in 2020 and then an increase in 2021.

The other important ways to compare deposit of one site with deposit of another is by computing ratios. Let's divide all deposit amounts to the first amount (i.e. January 01, 2020) and see how it is trending.

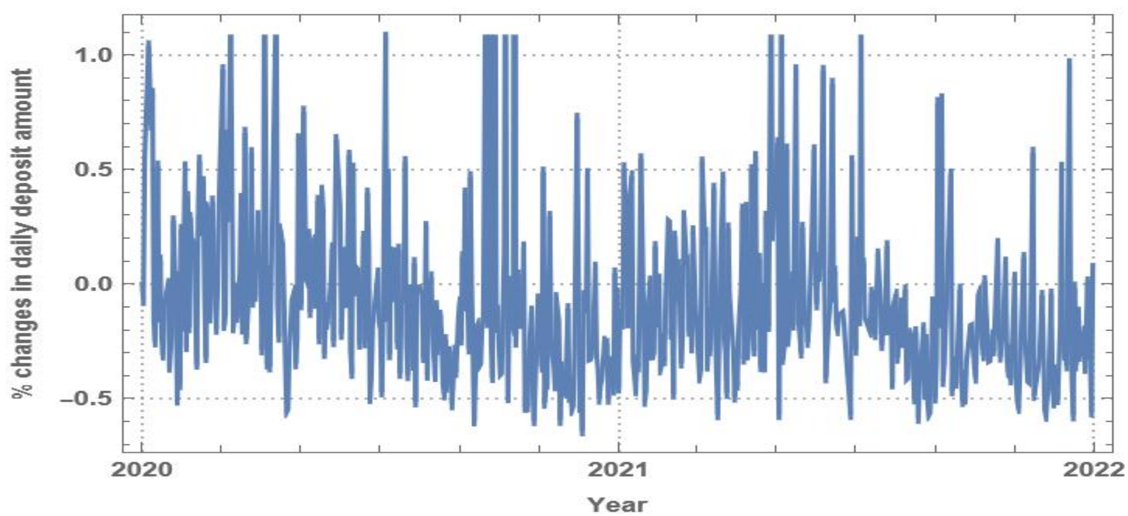


Figure 5. Percentage changes (with respect to the first day deposit) of deposits over two years' time

As can be seen from the above figure, the ratio of deposits made throughout the two years to the first day first year forms a kind of seasonality. The trend below 0.0 is an indication of decrease while those above 0.0 is an increase.

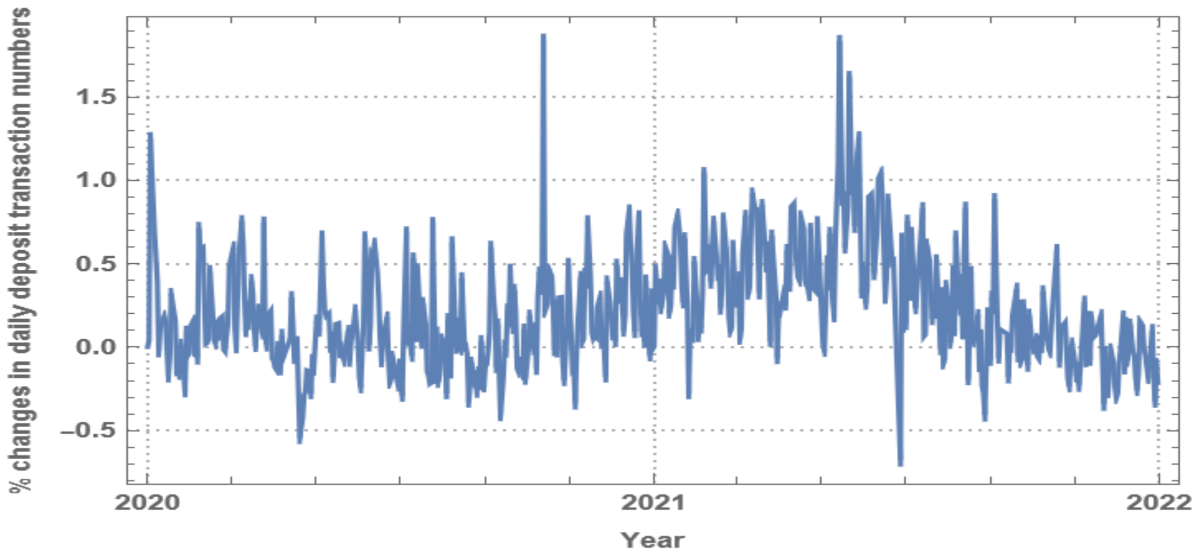


Figure 6. Percent changes in deposit transaction numbers over the two years

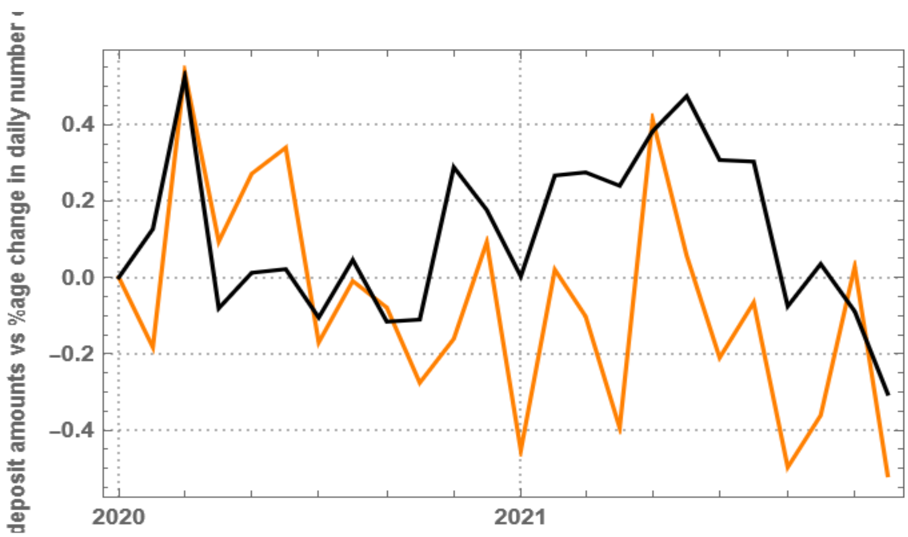


Figure 7. Percentage change trends in volume of deposit (Orange) vs percent changes in number of deposit transactions (Black)

The figure shows changes in deposit volume are lower than changes in the number of transactions. This shows as the time goes, the amount of deposit is getting lower and lower as compared to transaction counts. This means, as the times goes; amount of deposit per transaction is very less.

4.2. Withdrawal Trend Analysis

Withdrawal is the second transaction type in our dataset. Its amount and frequencies over two years' time period can reveal the financial behavior and economic stress of people proximate to this bank branch. How frequent and how many individuals withdrawal over time appears to be a great insight. It can also be used to compare many places and judge how stressful individuals are. The regularized withdrawal amount over two years (daily amounts) appears to be seasonal.

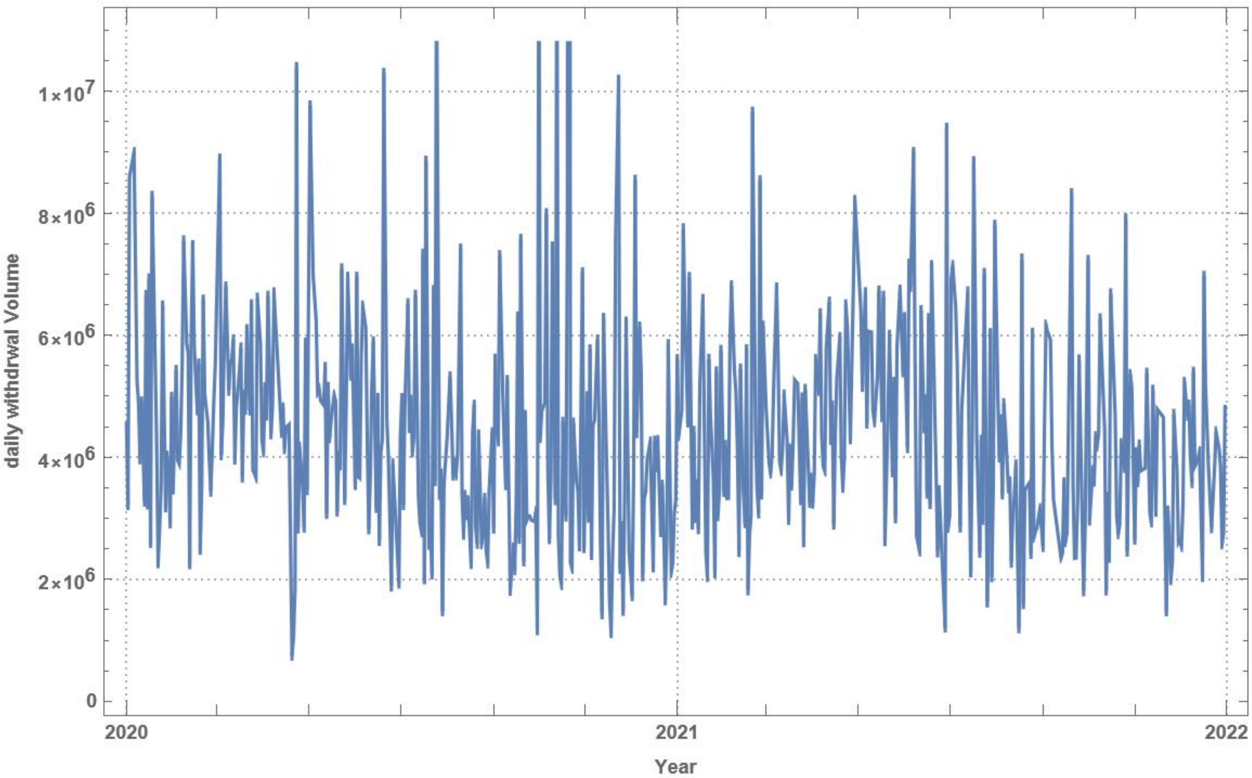


Figure 8. Regulated withdraw volume over the two years

The ratio of withdraw amounts to the number of withdraw transaction also reveal how stressful those bank customers are and can reveal economic stress of the surrounding communities of the bank branch. Thus, it is a great indicator of wealth/ poverty of those bank branch proximate people.

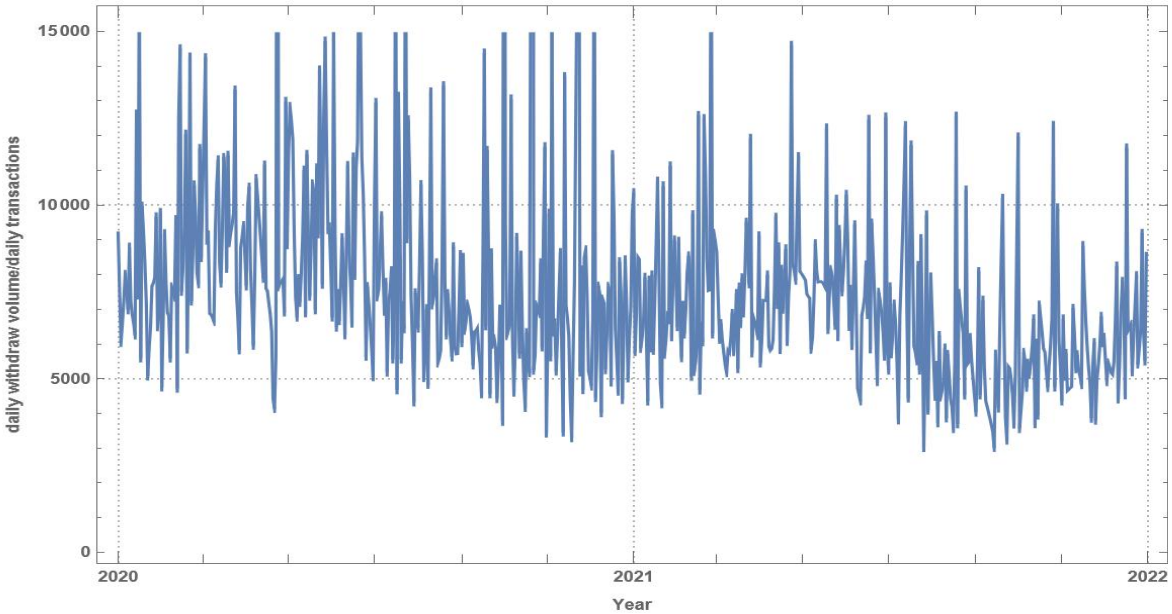


Figure 9. Ratio of amount of average of daily withdrawal to average number of withdrawal transactions.

The figure shows at this research site the average daily withdrawal to transaction is around 5000. It also forms a kind of seasonality trend.

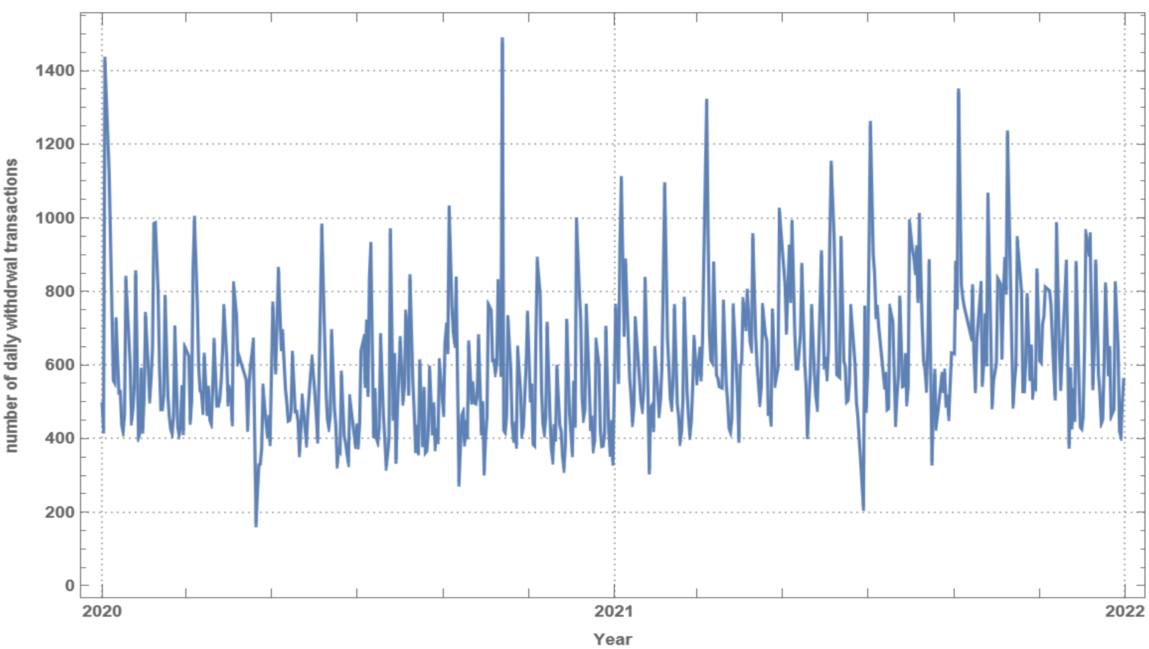


Figure10. Number of daily withdrawal transactions

What is really indicative is the ratio of withdrawals over the two years' time period to the first day withdrawal. Its trend is very informative. The following figure demonstrates this.

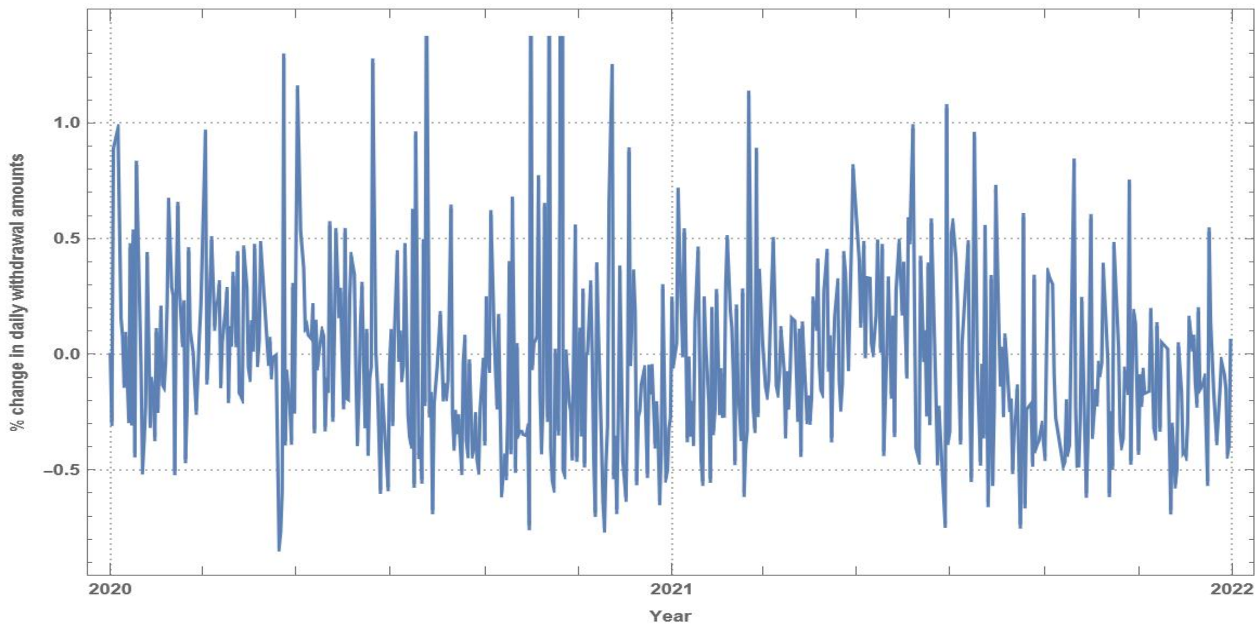


Figure 11. Percentage change (with respect to the first day, first month first year) in withdrawal over the two years' time period.

As can be seen the rate of withdrawal is decline in the first year with respect to the first base time and increase at the beginning of the second year with gradual decrease at the end.

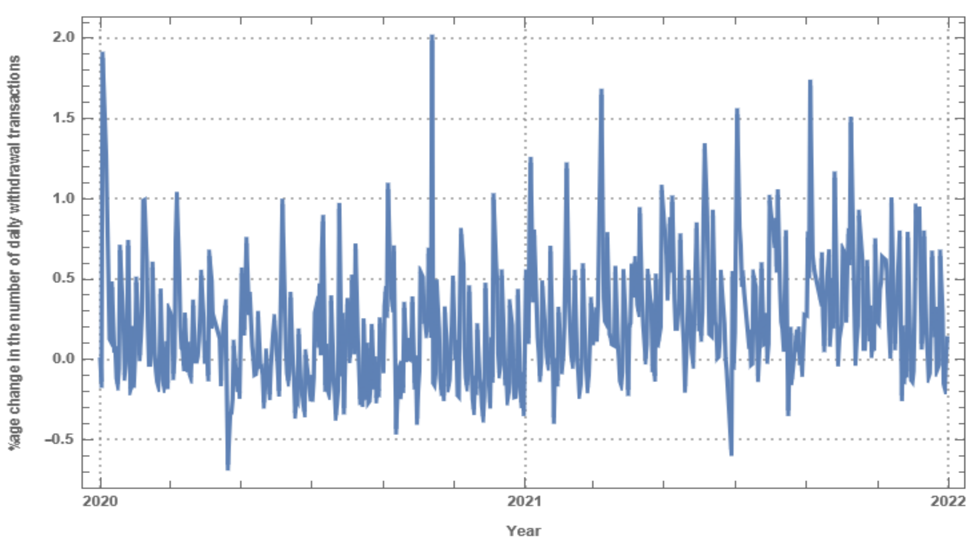
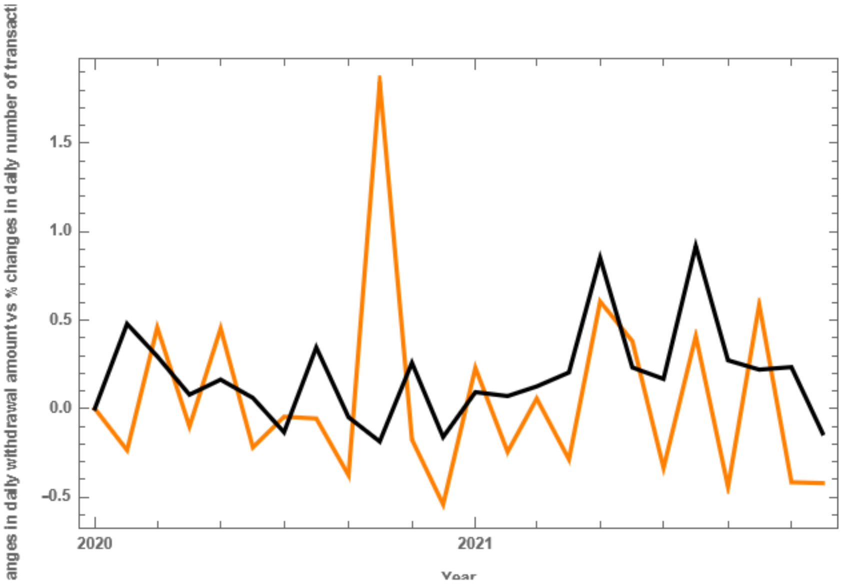


Figure 12. Percentage change in daily number of withdrawal transactions

As it can be seen above, the rate of frequency of transaction increases with respect to the first day.



Withdrawal (Orange) number of transaction (Black)

Figure 13. Percentage change of withdrawal amount vs percentage change of number of daily transactions

The rate in the volume of transaction amount outraces the frequency of the transaction only in the mid and final quarter of the first year.

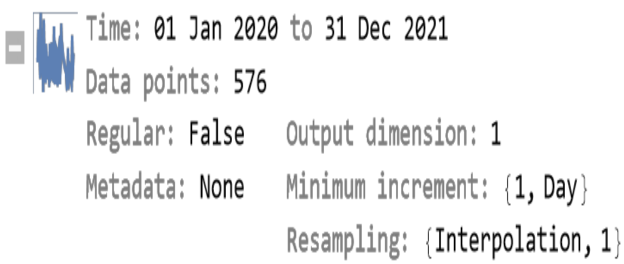
4.3. Time Series models

As can be seen, the dataset has a time stamp. This helps us to approach this study as a time series problem. We can create a time series object for selected transaction types (deposit, account transfer, withdrawal, mobile transfer, POS transactions, ATM withdrawals, and local money transfers) for the selected branches as shown in figure 1(a): Wolfram Mathematica's in-built Time Series function was used.

This kind of data transformation (process of converting raw data into a format or structure is suitable for model building and data discovery in general). It helps to minimize algorithm bias when the data distribution is skewed (Gong, 2021), which is the case in this study. Such dataset observations need to be continuous in time and allow only one path.

But bank transaction datasets are not continuous, as banks do not operate on holidays and Sundays. But some transactions, like ATM withdrawals and mobile transfers, do happen during such days. To minimize the adjusted loss function and prevent over-fitting or under-fitting, we regularized each dataset by using an interpolation technique.

In time series analysis, models assume current values will be expressed as driven in some ways from past values. Each dataset is re-sampled by using the Time Series Resample function. This function uses (linear interpolation methods by default). To make the visualization more visual for easy comparison, datasets are re-sampled into months (figure 1 (b)) except for POS transactions, which are re-sampled into weeks.



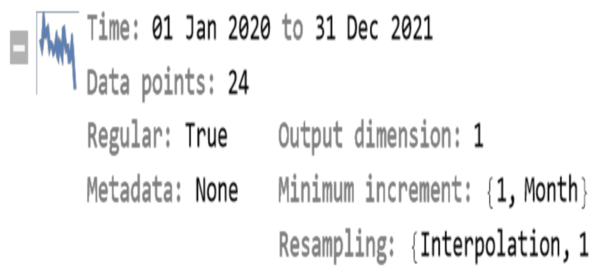


Figure 14 a: Sample time series data object (not Regularized) *b: Sample regularized time series data object*

Then, to analyze how poverty or wealth has trended over the two years for communities around selected branches, we have computed percentage changes in the following subsequent months with respect to the values of the first month. Figure 2 shows a sample time series object showing the percent change.

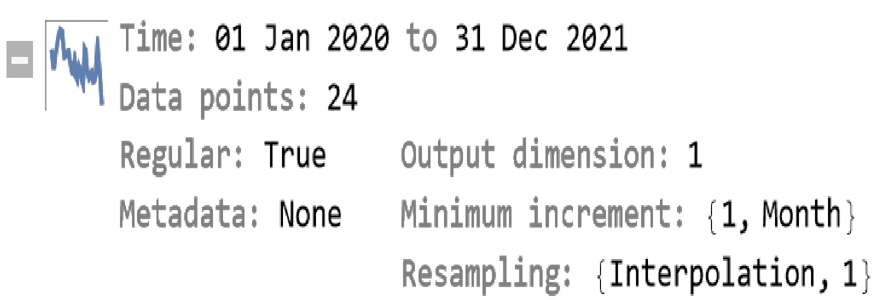


Figure 15. Sample time series data object of percentage changes with respect to the first month

Depending on the type of transaction, we can gain insights about how the wealth or poverty of communities around those bank branches is trending. For example, places with better economic situations will reveal an increase in the deposit as compared to other places. Contrary to this, if the economic situation of a place gets worse as compared to the first month, then the ratio will be lower (below zero). Those with zero values for this ratio mean there is no up or down with respect to the value of month one. The following subsection reveals the visualization and hence wealth or poverty trends.

4.4 Visualize seasonality and trends of selected transactions

As shown in figure 2, once the dataset objects are prepared, we can visualize the patterns and contrast the wealth of the local communities surrounding those bank branches for particular transactions. This is covered in more detail below.

4.4.1. Trends in Percentage Changes in Deposit Transaction

Close observation of cash deposits by communities using those branches reveals that there are huge ups and downs. As can be seen from figure 16, none of the study sites reveal a uniform positive or negative trend. Counties like Werebabu, Dehana, and Jamma reveal relatively different behaviors.

Compared to the first month, there are months where they have more than three time deposit amounts. But, the majority of the counties show a negative, which means their subsequent deposit amount, is lower than the first month. Some counties have the least and decreasing deposit amounts, though the majority of the months. For example, Wereilu appears to have the least of all others, followed by Tehuleder, Habiru, and Wadla counties. One can also easily observe which county gets the lowest and in which month. This is an interesting insight for targeting aid seekers. As discussed in the literature, counties with lower saving (deposit) amounts are relatively poorer than others. Another interesting insight about these transaction types is that each country has ups and downs at almost the same time (month), irrespective of how high or low the volumes are.

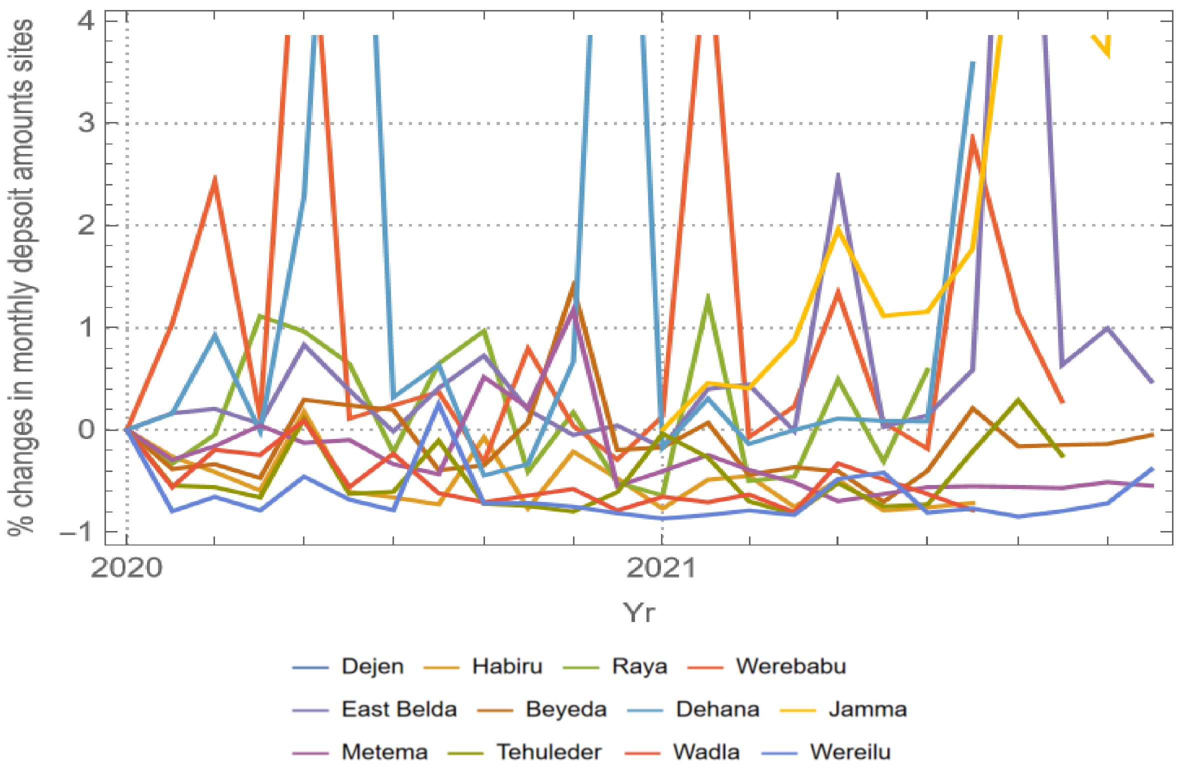


Figure 16. Trends in percentage changes in deposit

4.4.2 Trends in Percentage Changes in Withdrawal Transaction

Regarding withdrawal, figure 17 shows that majority of the study sites have a below zero (negative) trend compared to their respective values at the beginning of the month. From this figure, we learned that there are counties (East Belesa, Habiru, Raya Kobo, and Metema) whose withdrawal is reduced by over 50% with respect to the first month. Of course, there is an up and down with time. Similar to deposit transactions, withdrawal transactions also do not have uniform ups or downs. But, in general, there

is a tendency towards negative secular trends. It can also be seen which county has the least withdrawal amount as compared to other counties and can inform targeting counties for aid.

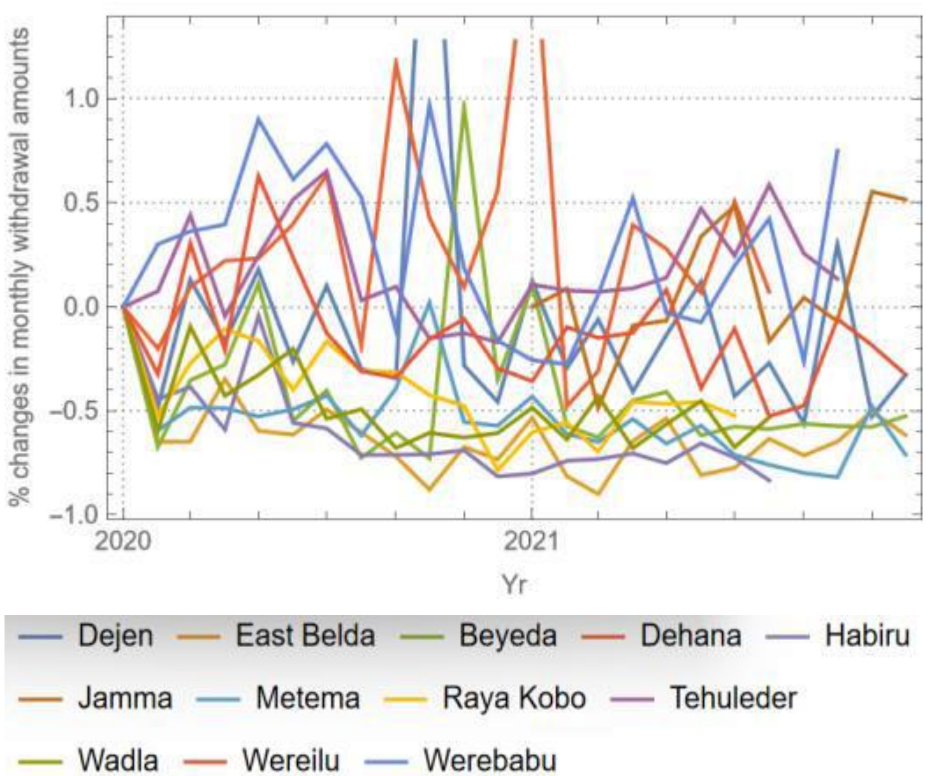


Figure 17. Trends in percentage changes in Withdrawal

4.4.3 Trends in Percentage changes in Account to account transfer

Account to account money transfer figure 18 shows a relatively positive secular trend for a few of the counties. Similar to deposit and withdrawal transactions, account to account transfer transactions do not show a uniform up or down trend. Again, this transaction has highs and lows for the majority of branches (counties) almost at the same time (months), similar to deposit and withdrawal. It should be understood that the highs and lows are in relative terms. As with poverty or wealth analysis, even though many branches show a decreasing trend (negative-below zero) in 2020, branches like (East Belesa and Dejen) exhibited improvement and a positive trend in 2021. One branch (Beyeda) shows the least across the two years' period. This insight adds value to development and humanitarian aid agencies in their targeting tasks.

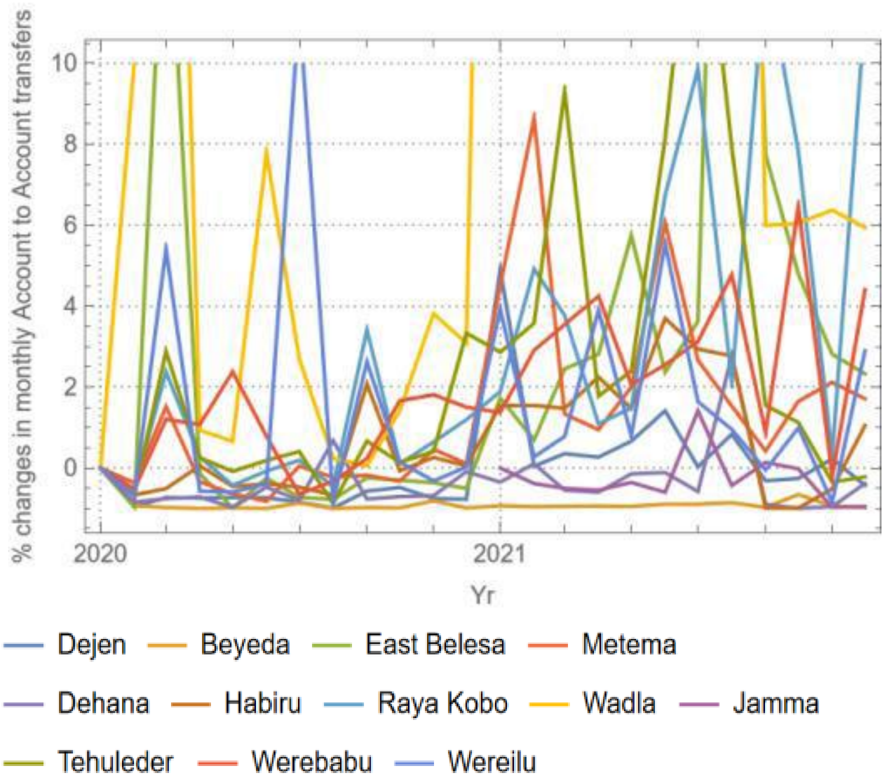


Figure 18: Trends in percentage changes in AC

4.4.4 Trends in Percentage Changes in ATM withdrawal Transaction

Unlike other transaction types, the percentage changes in ATM withdrawal amounts at the different bank branches oscillate between an increase and a decrease of 50% of the value in the first month (figure 19). At one branch (Wadla), there is a clear outlier. This branch has a significant high percentage increase (50%-100%). From the dataset we cannot tell why this happens. This branch has no data for 2021. Again, the majority of the branches have highs and lows at almost similar time periods. From the figure, one can also tell which branch has encountered the least ATM withdrawal amount and when. For example, Dehana, Werebabu, and Beyeda had the least withdrawal changes over the two years' time (with respect to the first month).

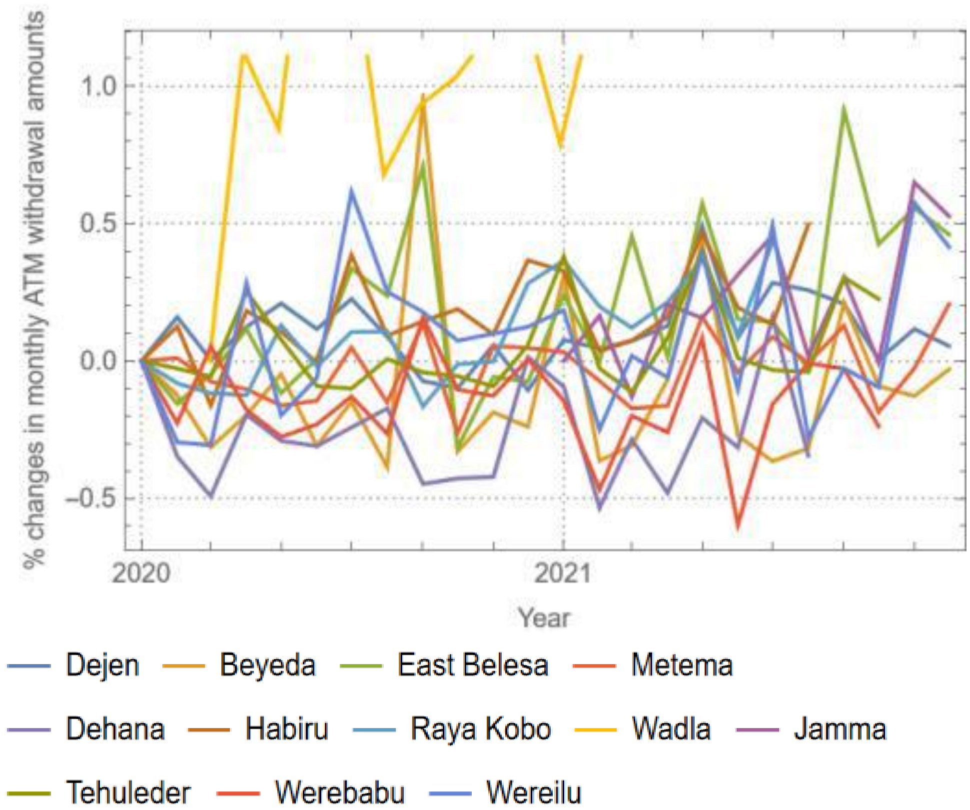


Figure 19. Trends in percentage changes in ATM

4.4.5 Trends in Percentage Changes in Mobile Transfer Transaction

Generally, mobile transfer transactions show a decreasing trend (20). There is a clear difference between the two years. In 2021, almost all branches showed a decrease in the amount of mobile transfers, while in 2020 there were both increase and decrease trends. Close observation of individual bank branches reveals Raya Kobo is the least of all in 2021, while Werebabu, East Belesa, and Tehuleder have shown the least in 2020. It is also clear when and which branch has got the least.

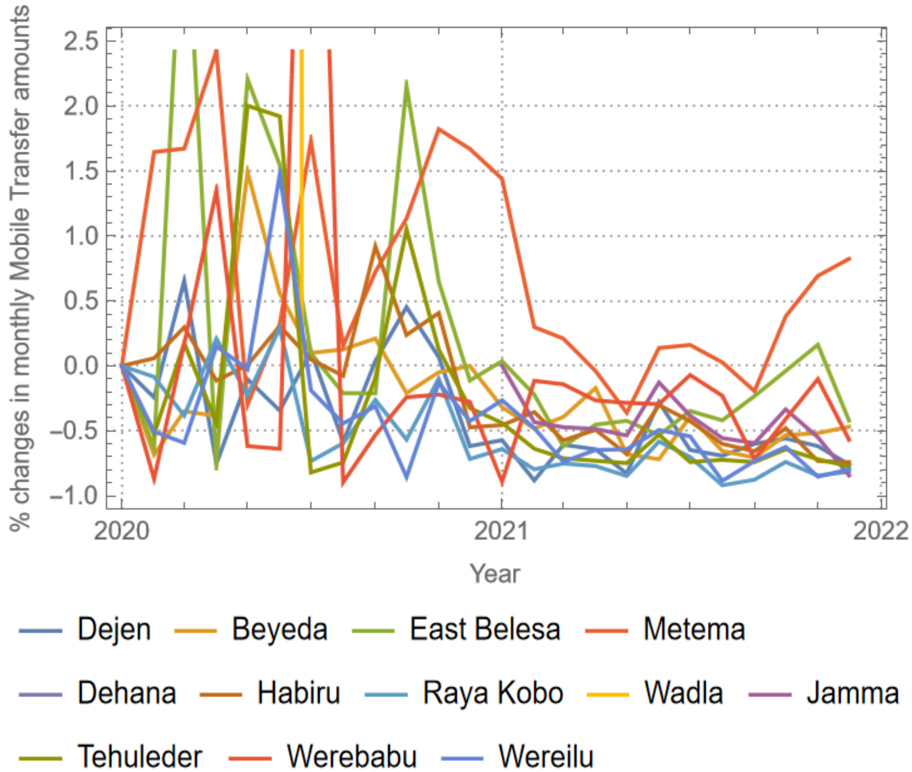


Figure 20. Trends in changes in Mobile transfer

4.4.6 Trends in Percentage Changes in Point of Sale Transaction

Unlike the other datasets, the number of data points for the point of sale transaction type is small (in relative terms). Thus, this dataset is resampled into weeks. Even though this dataset reveals both positive and negative changes (increase and decrease), the amount of increase is significantly high. As can be seen by figure 21, the percentage increase in POS transactions for the majority of the branches is very high, from 200% to 850%. This is quite unusual compared to other transaction types discussed above. This might be due to the war in the region; people might resort to using digital money rather than cash. Still, there are branches with the least percentage changes across the two years. For example, branches like Raya Kobo, Habiru, Dejen, Werebabu, and Tehuleder show decreasing trends for POS transaction types.

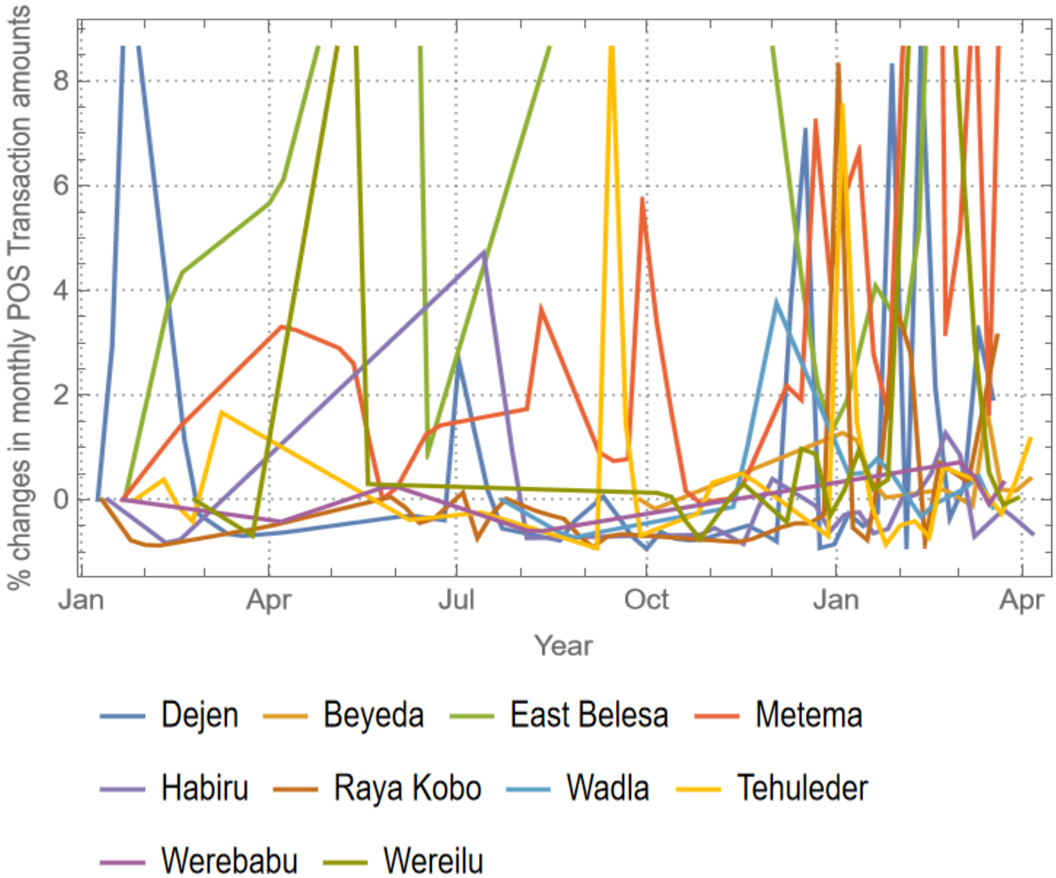


Figure 21. Trends in percentage in POS transaction

The variability of these transactions can be attributed to different factors like internal conflicts and displacements, natural disasters, COVID-19, and others. Whatever the factors might be what is important for aid agencies is to look at into how financial capabilities of those communities trend over time. Relative comparison matters a lot for such decisions.

4.5. Further Analysis

For comparison purpose the following sites are selected for further analysis.

4.5.1. East Belesa Site

Deposit

- Deposit amount Trend

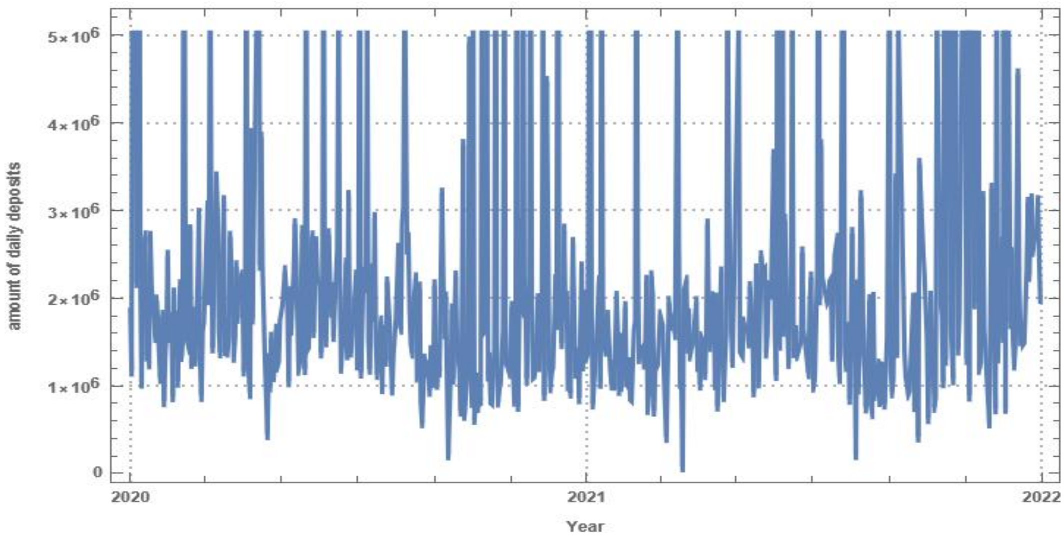


Figure 22. Amount of daily deposit

The amount of deposit in East-Belesa shows a normal trend as it lies between one and three million except increase at the end of each year.

- Deposit %age change trend

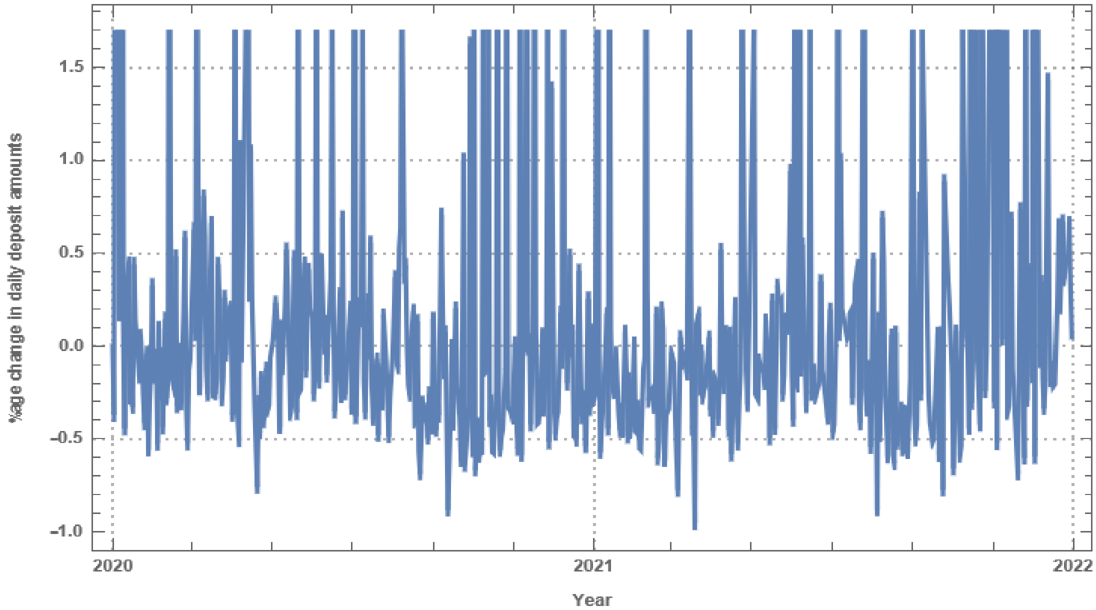


Figure 23. Percentage change in daily deposit amount

The rate of deposit amount shows an increase decrease trend in a specific path except in the final months of each year.

- Deposit Transaction count trends

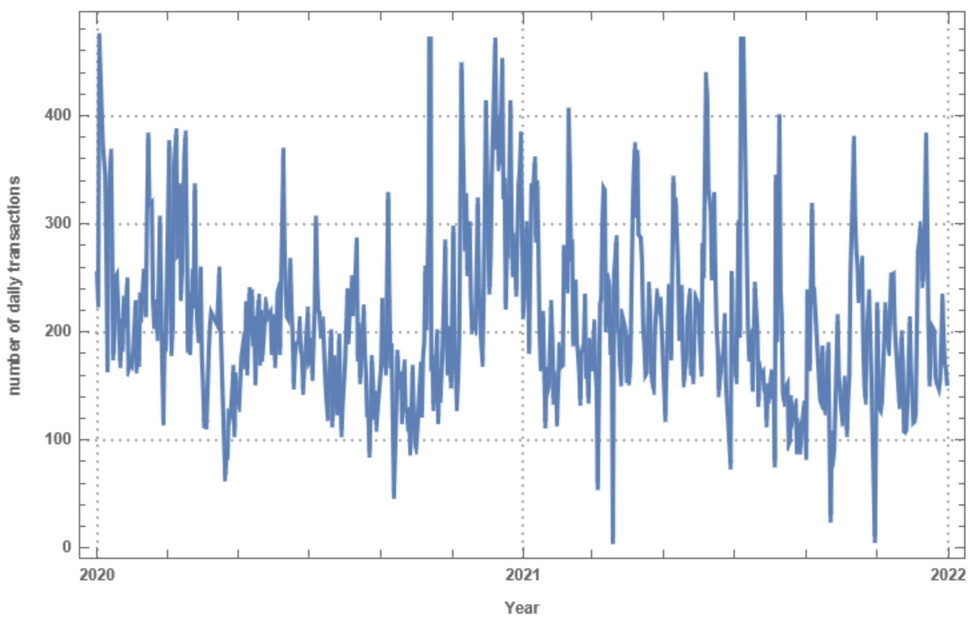


Figure 24. Number of daily deposit transaction

The first year shows seasonal increase and decrease in the frequency of transaction while the second year exhibits a total decline in the deposit frequency.

- Deposit Transaction percentage change trends

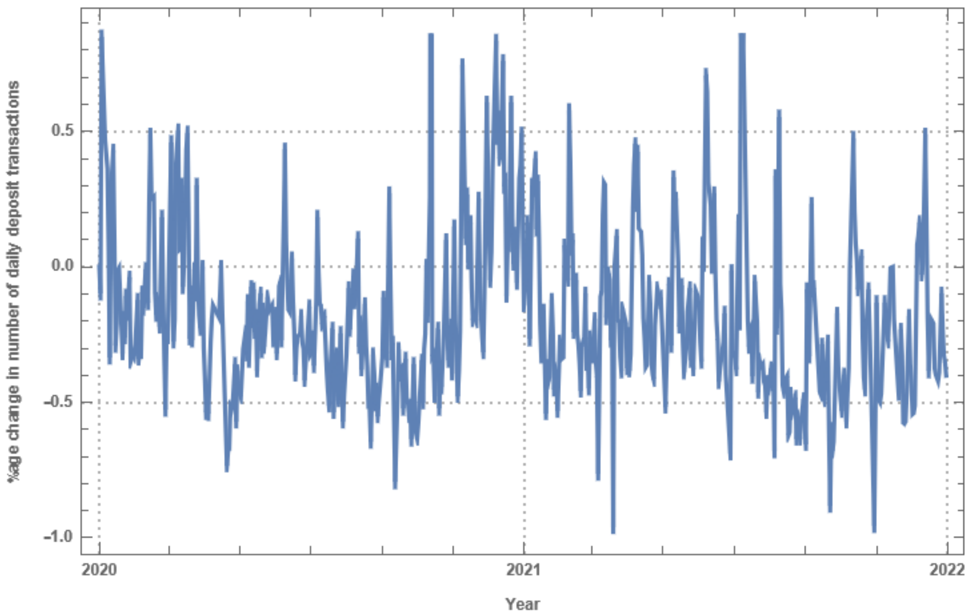


Figure 25. Percentage change in number of transactions

The rate of deposit frequency is decreased with respect to the first day.

- Deposit/ transaction count ratios

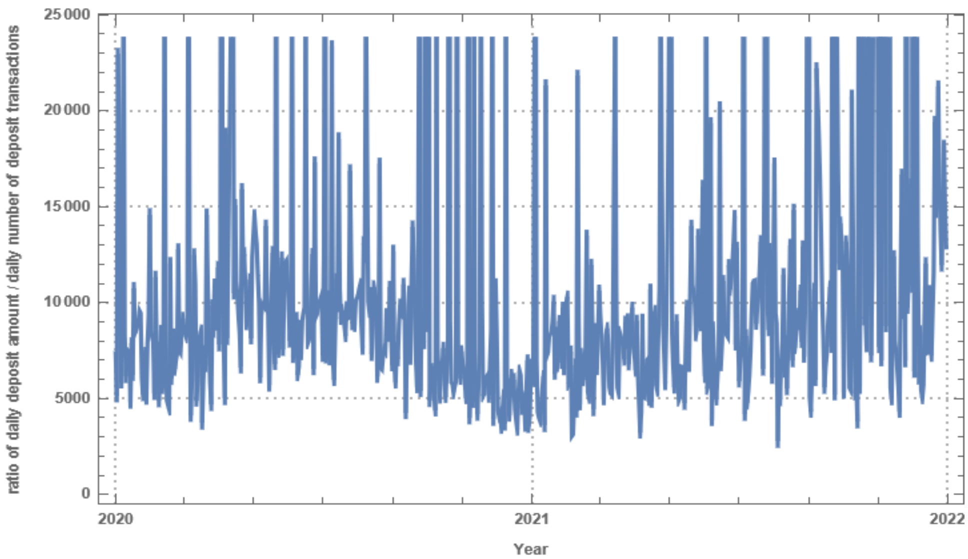


Figure 26. Ratio of daily deposit amount to transaction frequency

The ratio shows a trend of high increase particularly at the end of each year.

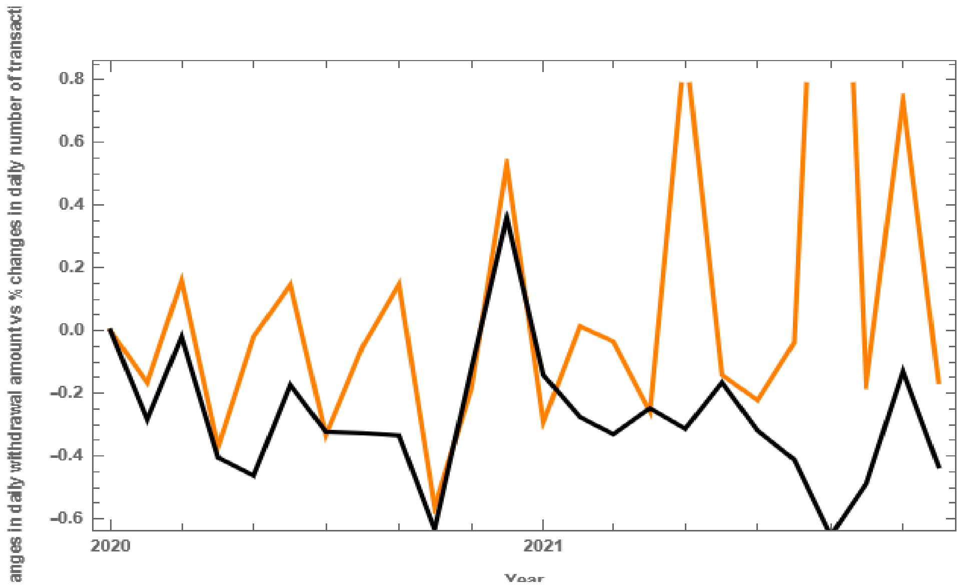


Figure 27. Percentage changes in deposit vs %age changes in deposit transaction counts

Deposit frequency percentage declines with respect to the deposit amount.

Withdrawal

➢ Withdrawal Amount

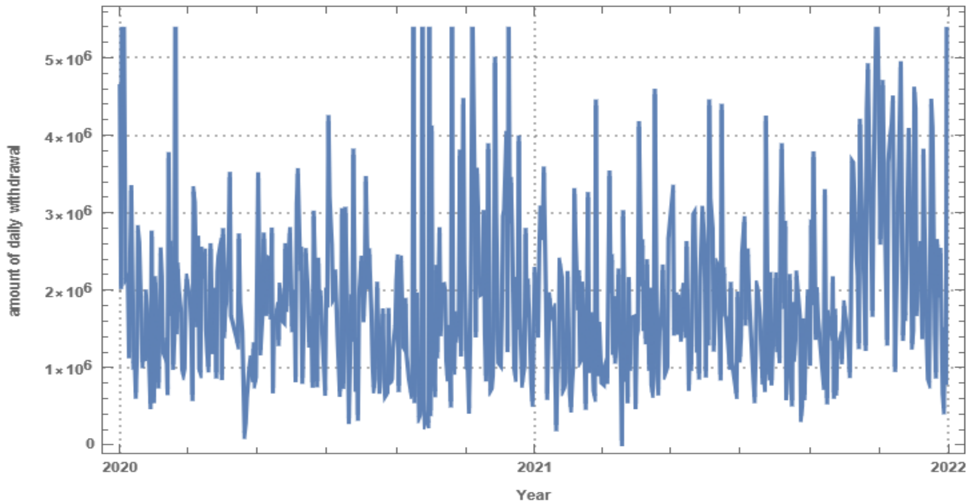


Figure 28. withdrawal amount

The amount of withdraw shows a trend to decrease every time.

➢ percentage change in withdraw amount

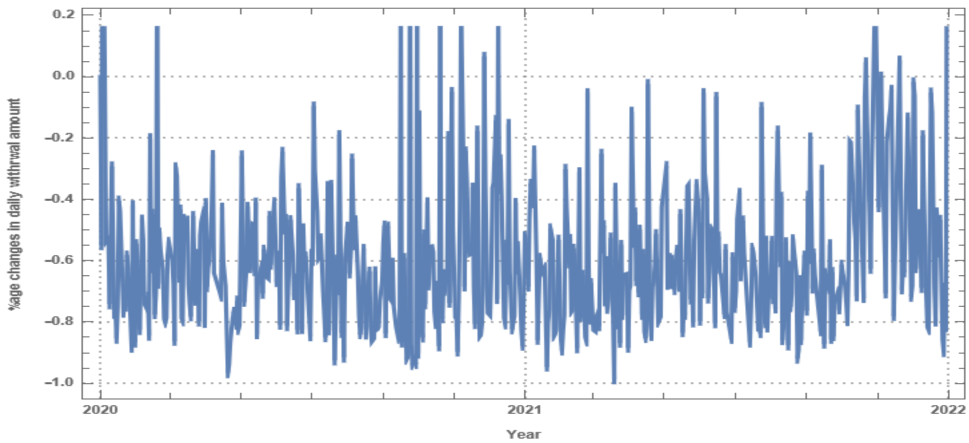


Figure 29. Percentage change in withdraw amount declines

As it shows in the figure above percentage change in withdrawal amount still lies blow the first date.

➢ Withdrawal Frequency

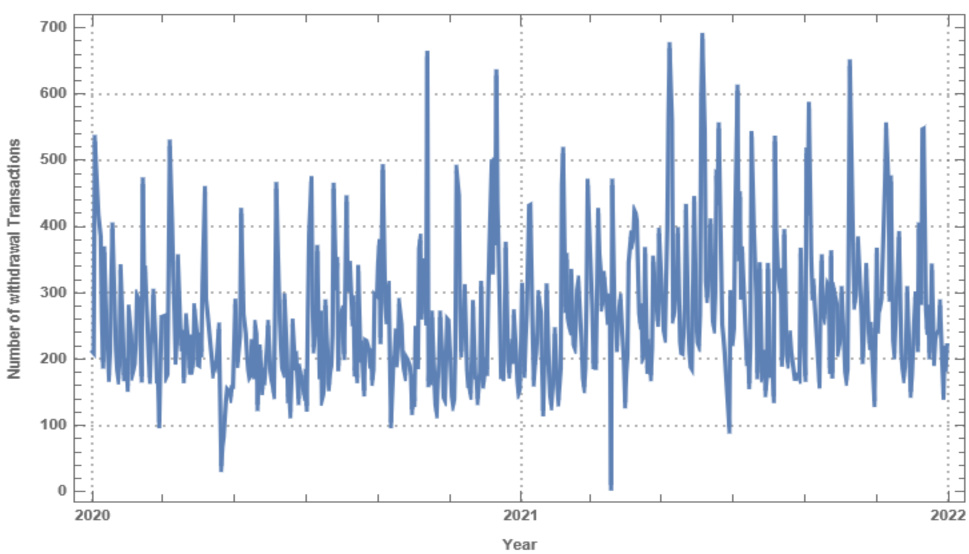


Figure 30. Number of withdrawal transactions

The frequency of withdrawal transactions shows a trend gradual increase

➢ percentage change in withdraw transaction frequency

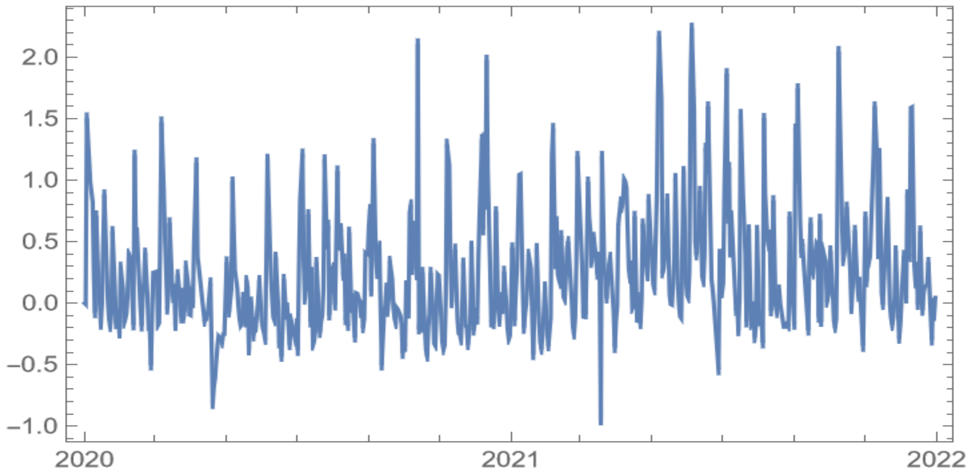


Figure 31. Percentage change in number of withdrawal

Withdrawal frequency is at high percentage rate with respect to first date.

➢ Ratio of withdraw amount / number of withdraw transaction

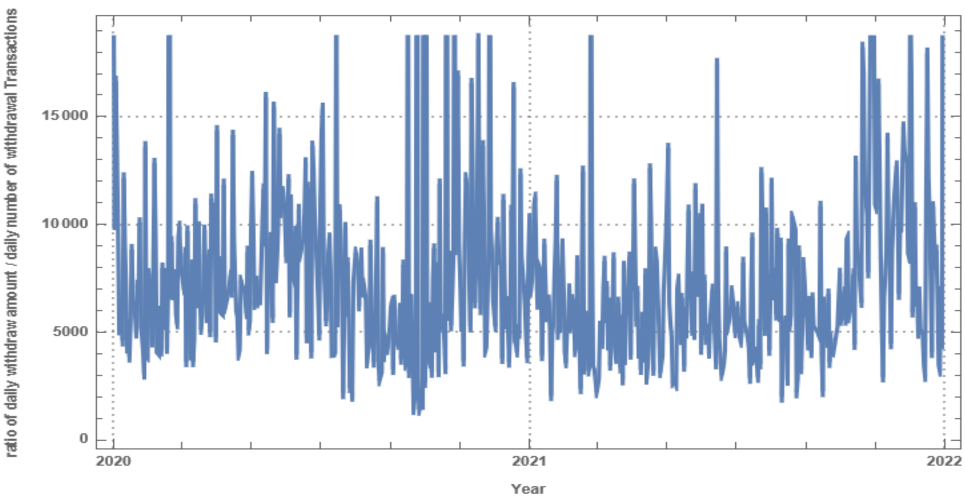


Figure 32. Withdrawal amount ratio to transaction frequency

It shows a decline in the amount while the frequency shows seasonal increase.

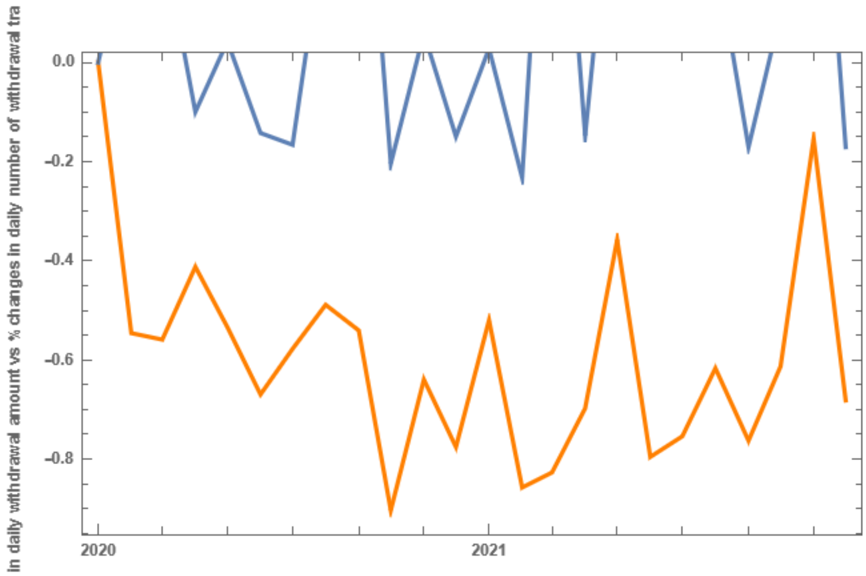


Figure 33. Percentage change in withdraw amount vs %age change in withdraw transactions

As the figure above shows, there is a high percentage decrease in the amount of transaction with respect to the percentage of withdrawal frequency.

4.5.2. Beyeda Site

Deposit

- Deposit amount Trend

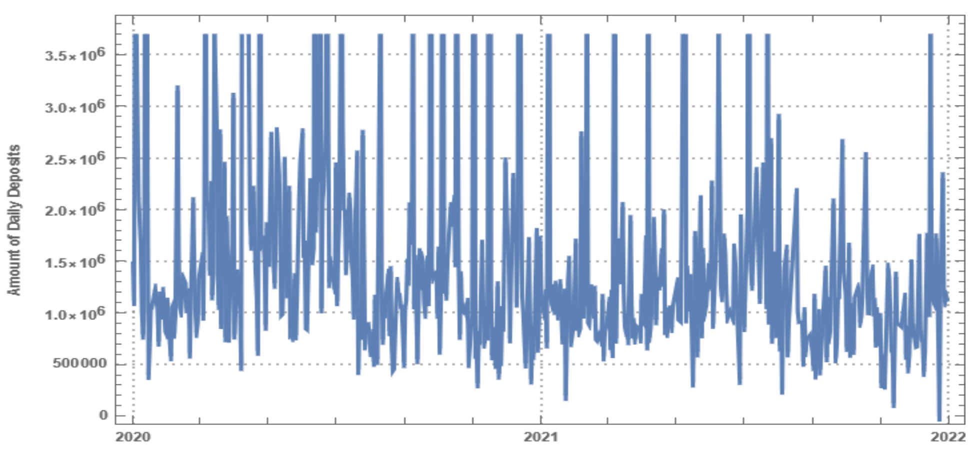


Figure 34. Amount of daily deposit

The amount of daily deposit shows an increase in the first year while it shows a decrease in the second year.

➢ Deposit percentage change trend

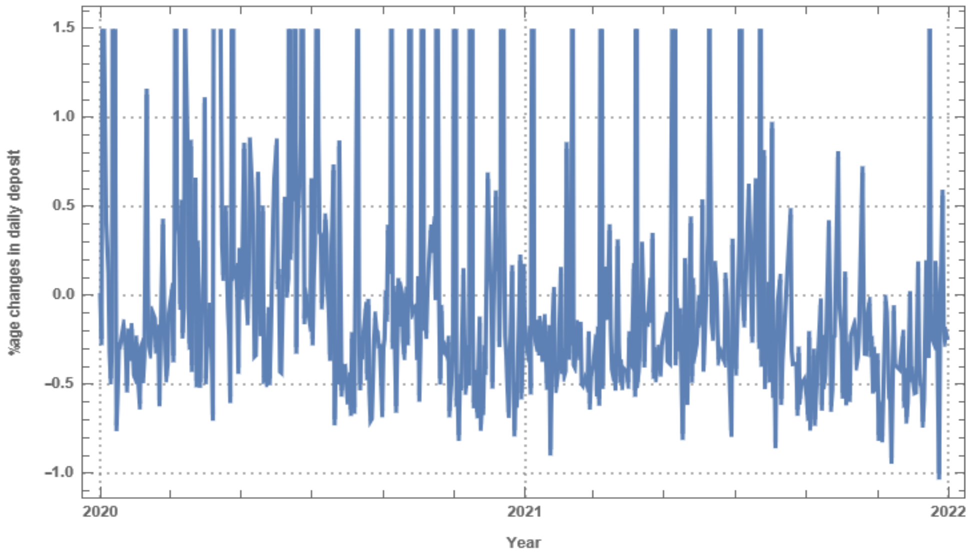


Figure 35. Percentage deposit change

There was an increase trend in the first year with respect to the first date while it was dominated by the decline in the next year.

➢ Deposit Transaction count trends

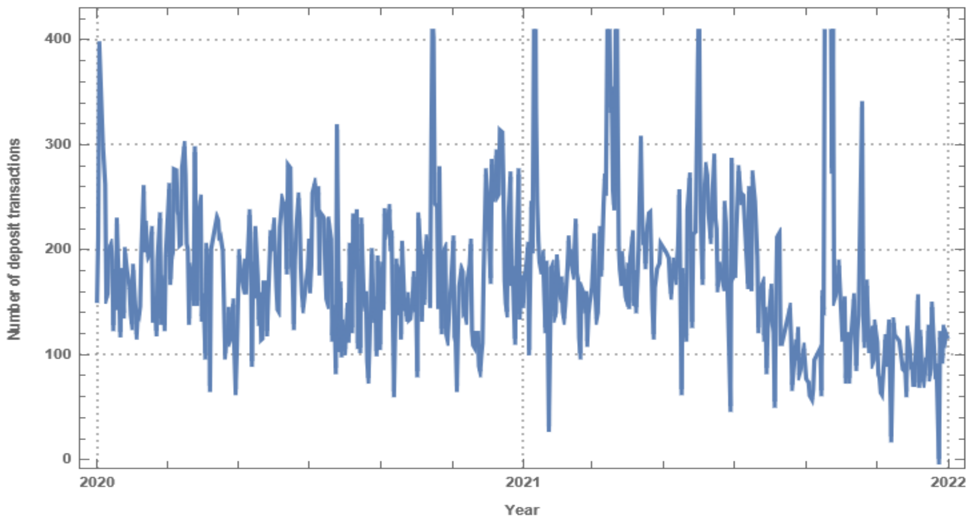


Figure 36. Number of deposit transaction

The frequency of deposit transaction shows a trend to increase at the beginning and to decline throughout the rest of the year.

➢ Deposit Transaction percentage change trends

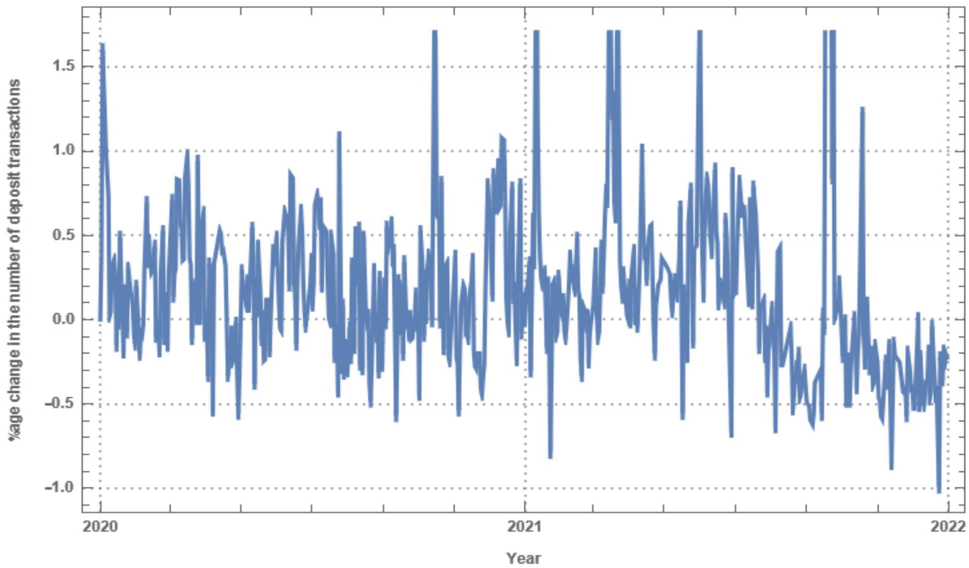


Figure 37. Percentage change in frequency of transaction

The frequency always tend to decline while as it is depicted in the above figure.

➢ Deposit/ transaction count ratios

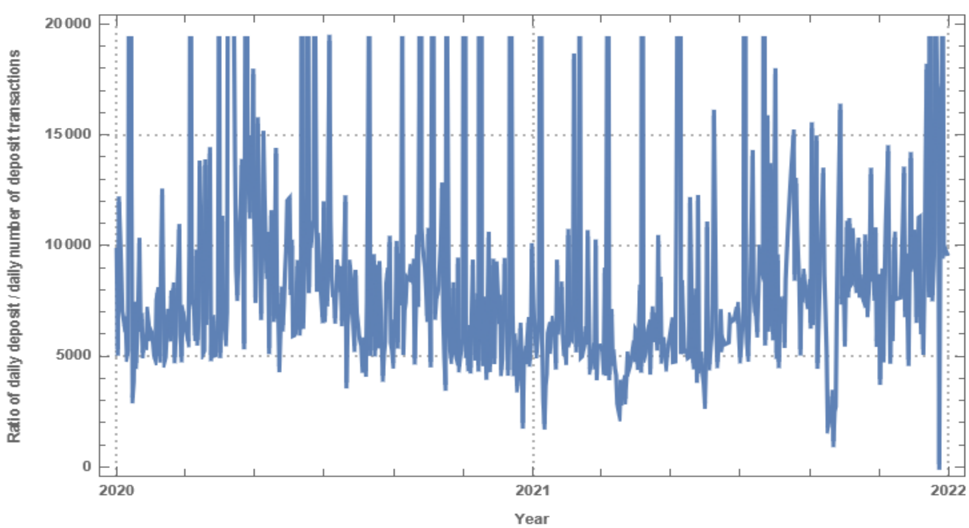


Figure 38. Deposit amount to transaction frequency ratio

The amount of deposit outraces the number of transaction.

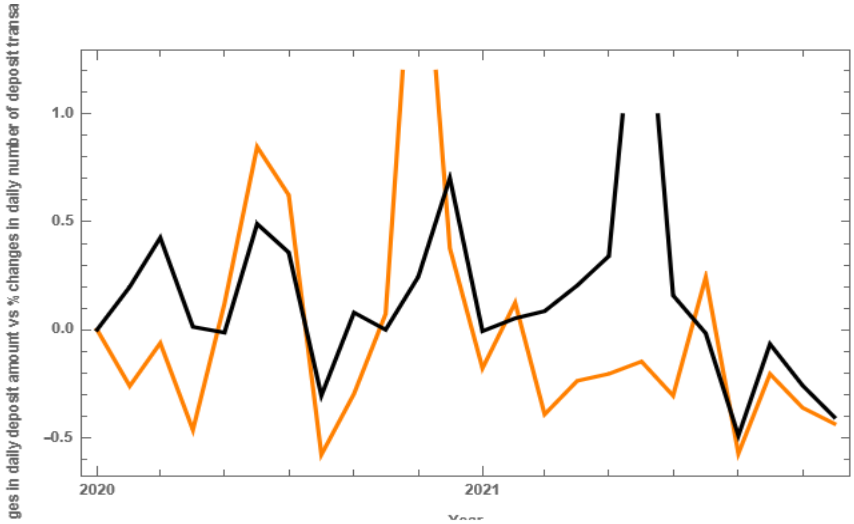


Figure 39. Percentage changes in deposit vs %age changes in deposit transaction counts Deposit (Black) and transaction (Orange)

The deposit amount rate shows an increase with respect to the frequency except at the end of the second year where both declines.

Withdrawal

- Daily withdraw amount

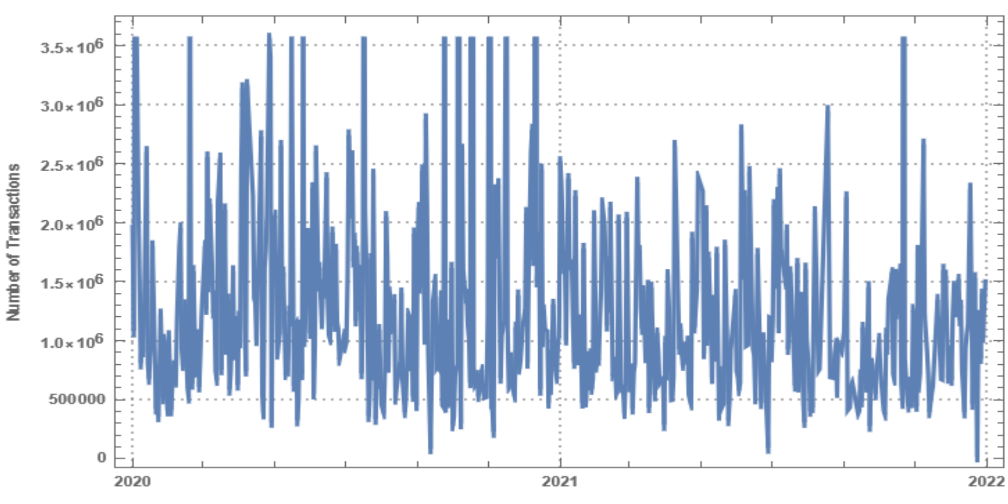


Figure 40. Daily amount of withdrawal

The amount of withdrawal always declines as it shows above.

- Daily withdraw percentage change

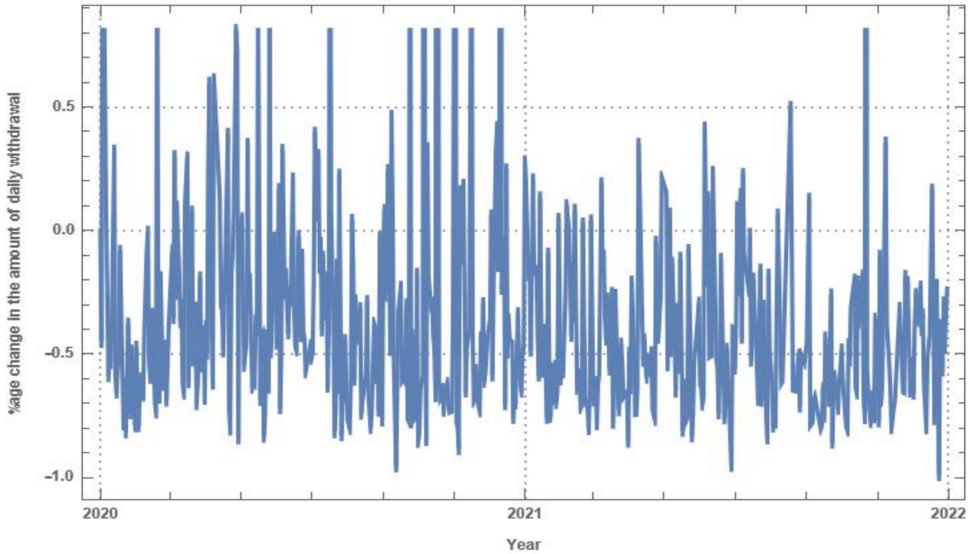


Figure 41. Percentage change in daily withdraw

With respect to the first date the daily withdrawal amount declines throughout the season

- Daily withdrawal number of transactions

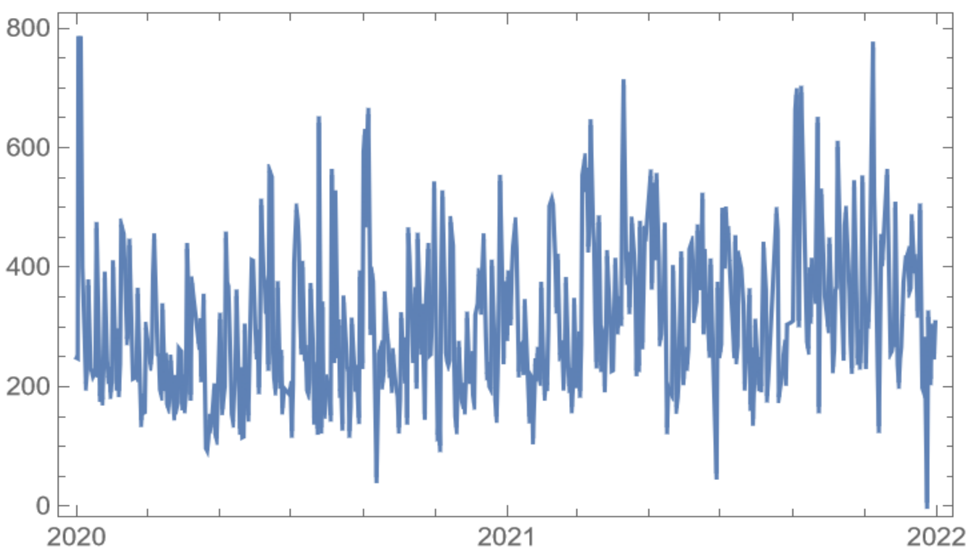


Figure 42. Daily number of withdrawal transactions

The daily withdrawal frequency was increased most of the time

➢ Daily number of transactions percentage change

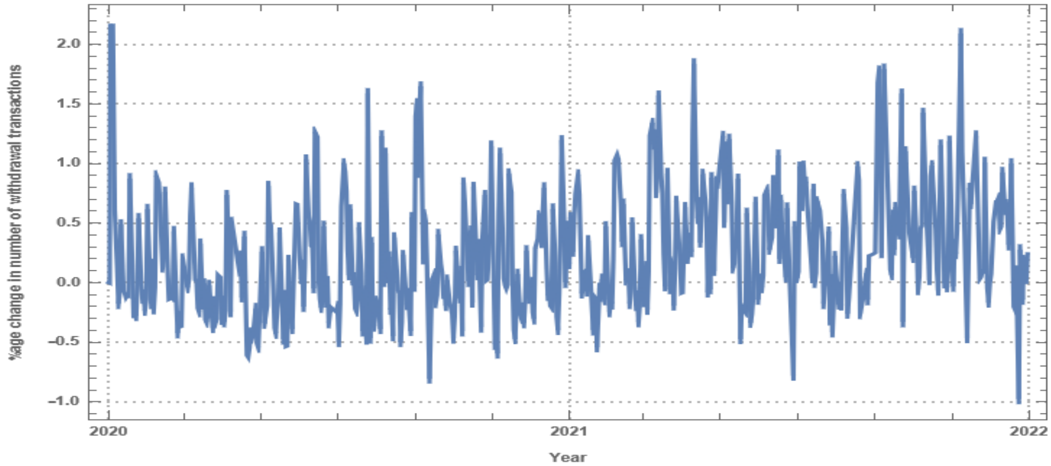


Figure 43. Percentage change in daily number of withdrawal

The frequency highly increased in the second year

➢ Ratio of withdrawal amount to number of withdrawal

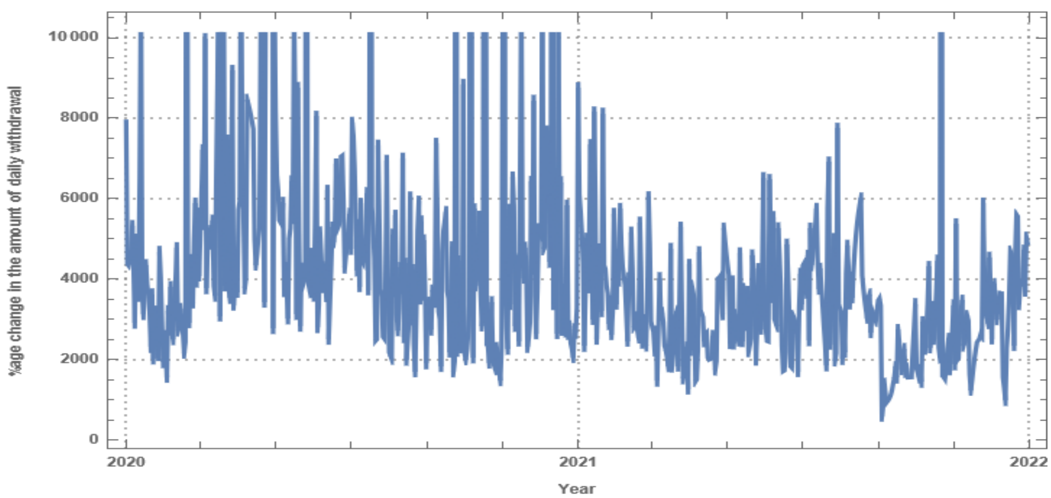
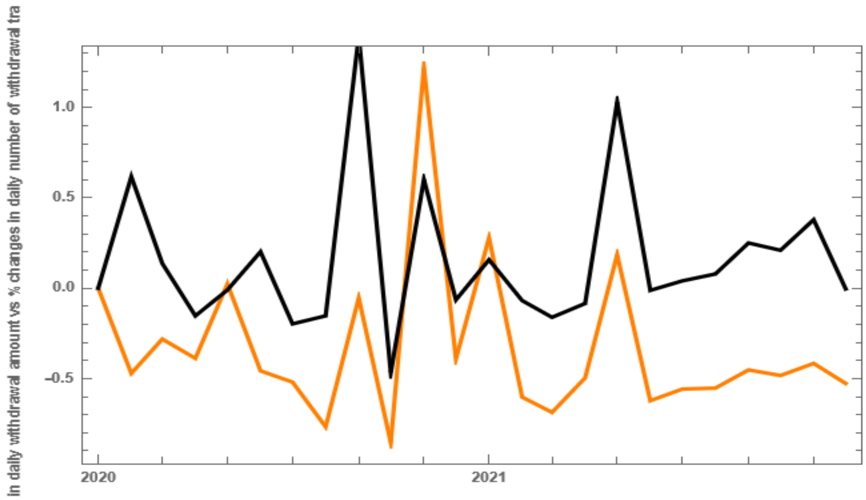


Figure 44. Ratio of withdrawal amount to the number of transaction

The amount tend to decrease while the transaction increases



Withdrawal (orange) and transaction (black)

Figure 45. Percentage changes in daily withdraw amount vs % changes in number of daily transactions

The percentage amount of withdrawal declines while the rate of transaction increases.

4.5.3. Metema Site

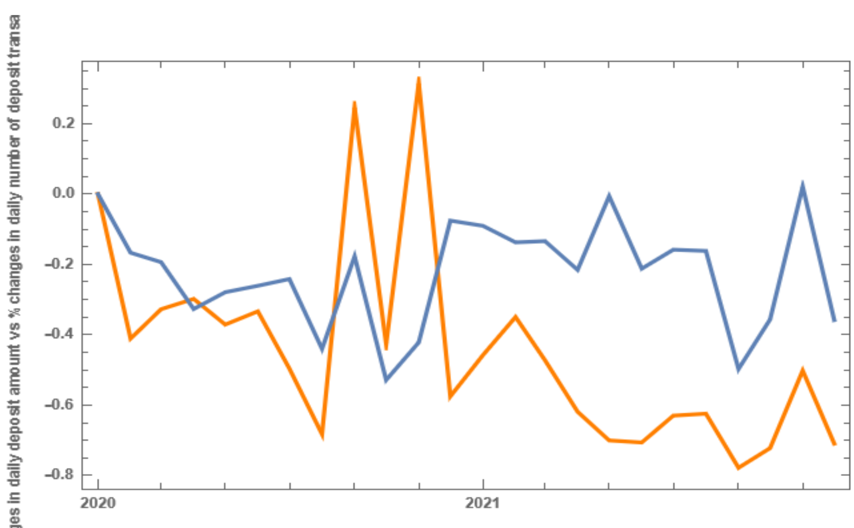


Figure 46. Deposit:39 %age changes in deposit vs %age changes in deposit transaction counts Deposit (orange) and transaction (blue)

As the above figure shows both amount and frequency of deposit are declined with respect to the first date where the deposit amount rapidly declined below the number of transaction.

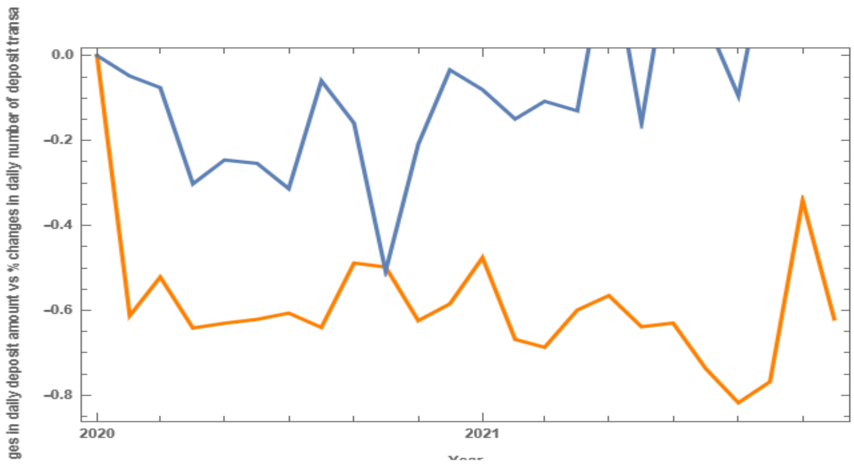
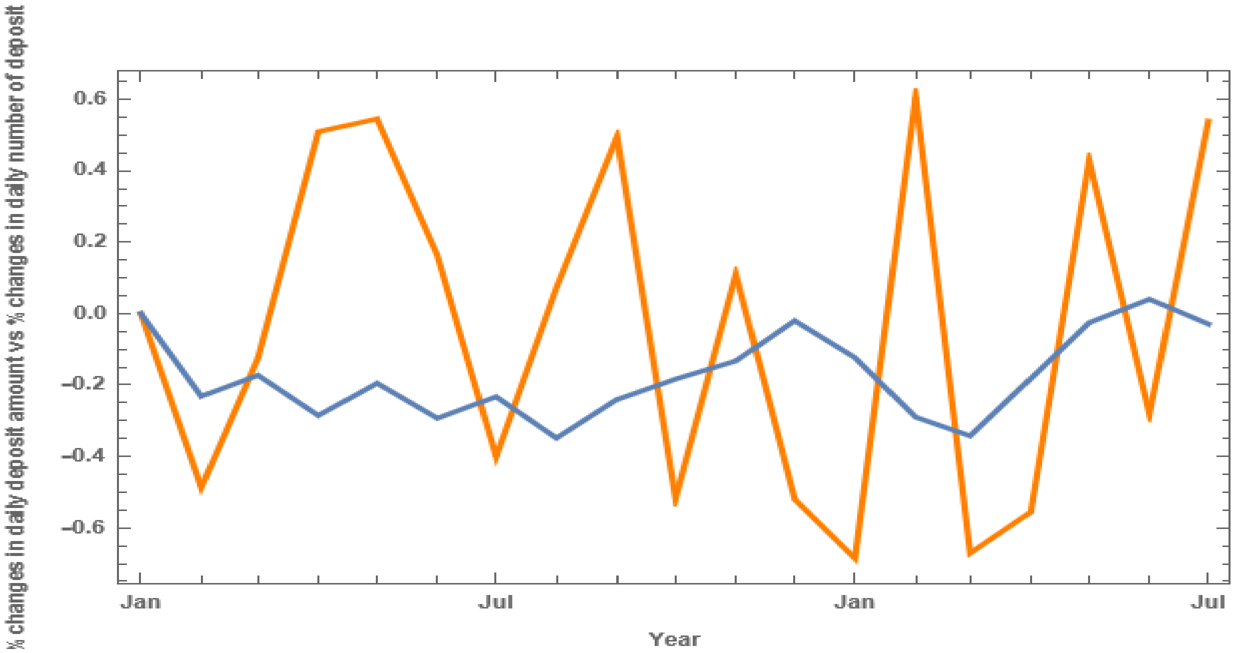


Figure 47. Percentage change in withdrawal amount vs withdrawal transactions

Withdrawal (orange) and transaction (blue)

As the above figure shows both amount and frequency of withdrawal are declined below zero with respect to the first date where the withdrawal amount rapidly declined below the number of transaction.

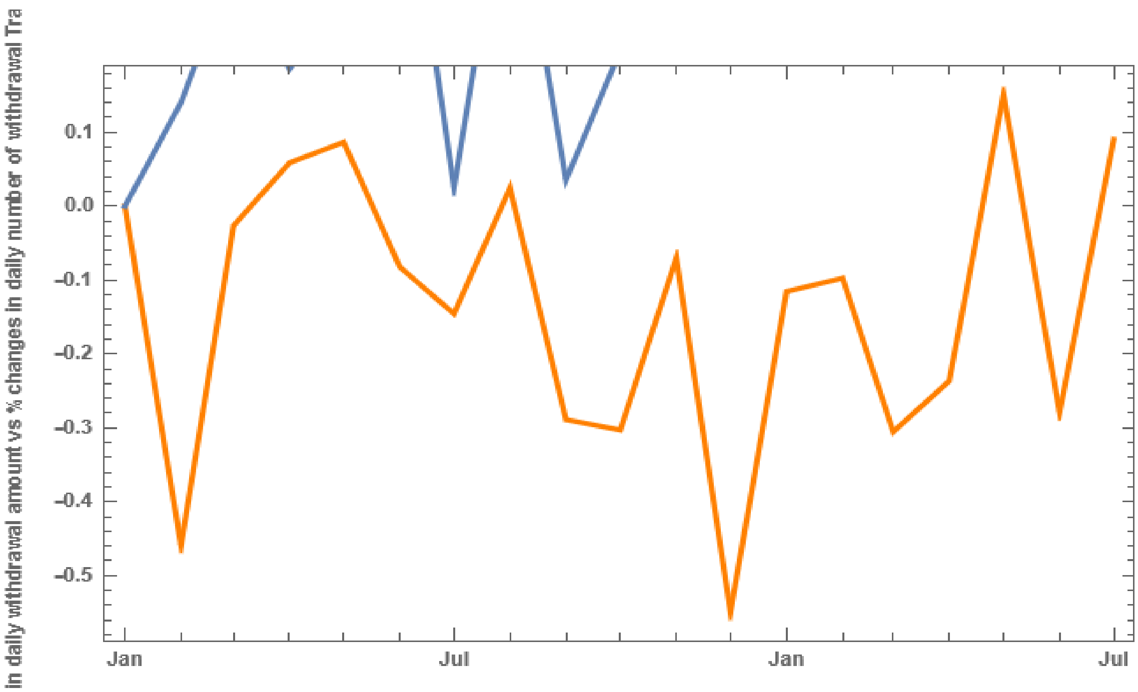
4.5.4. Raya Site



Deposit (orange) and transaction (blue)

Figure 48. Percentage changes in deposit vs %age changes in deposit transaction counts

The amount of deposit percentage change shows ain increase with respect to the nuber of deposit which shows a decline.



Withdrawal (orange) and transaction (blue)

Figure 49. Percentage change in withdrawal amount vs %age change withdrawal transactions

As the above figure shows both amount and frequency of withdrawal are declined below zero with respect to the first date where the withdrawal amount rapidly declined below the number of transaction.

4.6 Fitting Transaction Data Points

Time series models are used to forecast the future based on previous observed or collected datasets at regular time intervals (Engineering Statistics Handbook, 2010). By using the Time Series Model Fit function, we try to fit the datasets with different process models (AR, MA, ARMA, ARIMA, SARIMA, SARMA, FARIM, and ARCH). Table 4 summarizes the quality of selected process models for each transaction type for sample branches. The quality of the process model is evaluated by looking at the values of AIC (Akaike information criterion), BIC (Bayesian information criterion), and SBC (Schwartz-bayes information criterion). Models with smaller AIC and BIC are considered the better models.

Branch	Criteria	Deposit FARIM	Withdraw SARMA	Acct to acct transfer SARMA	ATM withdrawal FARIM	Mobile Transfer ARMA	POS SARMA
Dejen	AIC	-121.783	-22.6525	11.869	-140.502	54.0882	112.934
	AICc	-118.583	-20.0811	15.069	-137.302	58.1935	115.134
	BIC	-114.206	-25.2849	9.38105	-133.958	60.245	114.689
	SBC	-119.427	-21.4745	14.2251	-138.146	57.6224	115.077
		FARIM	ARCH	AR	SARMA	SARMA	SARMA
Beyeda	AIC	-41.8762	-71.4633	24.4644	56.737	90.1902	-47.3722
	AICc	-38.6762	-67.358	27.6644	59.3084	92.7616	-44.8922
	BIC	-42.6739	-65.6659	21.0774	54.6305	92.4991	-45.9967
	SBC	-39.5201	-679291	26.8205	57.915	91.3682	-46.04
				SARMA	FARIM	SARMA	SARMA
East beresa	AIC	30.347	-84.788	94.1364	-74.1634	6.66051	269.06
	AICc	32.9184	-82.2166	96.7078	-70.9634	9.86051	271.775
	BIC	27.7338	-83.9599	93.5034	-70.5708	5.42077	273.675
	SBC	31.5251	-83.61	95.3144	-71.8073	9.01661	275.393
		FARIM	FARIM	FARIM	FARIM	ARMA	SARMA
metema	AIC	-58.4665	-78.7363	36.1437	-192.275	71.4305	171.741
	AICc	-55.2665	-78.1652	38.7151	-189.075	74.6305	173.944
	BIC	-54.8644	-80.5385	36.8225	-181.408	75.5158	173.425
	SBC	-56.1104	-82.2705	37.3218	-189.918	73.7866	173.868

Table 4: Data fitting process models for selected branches

As can be seen from table 4, every transaction type fits different process models. FARIM (which represents an autoregressive fractionally integrated moving-average process) and SARMA (which represents weakly stationary seasonal autoregressive moving-average) models better fit the majority of datasets as compared to the other process models. The ups and downs (seasonality nature of most of the dataset is also captured by SARMA). Bank transaction types (first row of table 4) are distinctive of one another. Each of these transaction types are better to be represented as independent data models per bank branches.

As bank transaction datasets come in a stream every minutes, day, month, or year, there is seasonality in the trends. This is reflected on how majority of the transaction types fit with SARMA model than other models. These models are not same every month or year. By nature, bank transaction datasets come in a stream, which suggests continuous update on the type of models. The online learning (a modified form of supervised) machine learning paradigm is a good fit to serve this purpose. It gives more importance to recent datasets than older datasets. Figure 6 depicts its schema. According to this figure, as datasets come in a stream, new models should be trained based on the newly incoming datasets. Similarly, trend analysis of aid places should be done continuously, giving more emphasis to recent datasets. Unlike many traditional machine learning problems, poverty targeting problems need to be based on recent datasets and should be updated continuously.

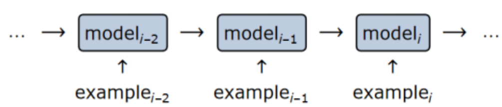


Figure 50: schema of online machine learning paradigm (source: Bernard pp. 18)

CHAPTER FIVE

5. CONCLUSION AND CONTRIBUTIONS

5.1. Conclusions

This study explored the use of machine learning/ artificial intelligence techniques to identify higher and lower level financial capability assessment that can help aid agencies and government agencies make better and informed decisions to prioritize and target beneficiaries for aid. Trend analysis of transactions at each bank branch serving the surrounding community (proxy approach) can reveal how wealth or poverty trends in a more granular manner. These datasets are automatically captured by the banks' IT infrastructure. They are accurate, captured in real time, detailed, and not manipulated directly to the benefit of others and thus they can be used to identify the neediest communities for aid.

By developing the necessary data sharing policies with banks, governments and aid agencies can monitor poverty and wealth in real time across the nation.

It is obvious that accurate advance prediction enables the timely positioning of humanitarian aid suppliers. With this regard using bank transactions data to target beneficiary is inevitable rather than being an alternative.

5.2. Contribution

As stated in the literature transaction types like bank deposits (saving) and withdrawals can signify a lot about financial/ wealth capabilities of individuals. In this study, we found that trend analysis of individuals' bank account transaction datasets like deposit, withdrawals, and transfers can signify how

poverty/ wealth trend at specific location or community. So far, these datasets have not been analysed and used for decision-making by either banks or other institutions.

5.3. Future Directions

In the future, we intend to scale up this approach. From all the banks and their branches operating in the nation, we will collect transaction datasets and develop a national digital poverty map accessible by governments and aid agencies on the web. The national digital poverty map will be updated every month through an online machine learning model. To make the map more valuable to users, we will develop an online machine learning model that feeds and updates the map every month for better decision making.

.

References

Altındağ, O., O'Connell, S. D., Şaşmaz, A., Balcıoğlu, Z., Cadoni, P., Jerneck, M., & Foong, A. K. (2021). Targeting humanitarian aid using administrative data: Model design and validation. *Journal of Development Economics*, *148*, 102564.

Arendt-Cassetta, L. (2021). From digital promise to frontline practice: new and emerging technologies in humanitarian action. *New York: OCHA*.

Banerjee, A. V. (2006). The Economic Lives of the Poor Abhijit V. Banerjee and Esther Duflo October 2006. Africa. pp 164.

Brei, M., Mohan, P., & Strobl, E. (2019). The impact of natural disasters on the banking sector: Evidence from hurricane strikes in the Caribbean. *The Quarterly Review of Economics and Finance*, *72*, 232-239.

Clay, D. C., Molla, D., & Habtewold, D. (1999). Food aid targeting in Ethiopia: A study of who needs it and who gets it. *Food policy*, *24*(4), 391-409.

Devereux, S., Masset, E., Sabates-Wheeler, R., Samson, M., te Lintelo, D., & Rivas, A. M. (2015). Evaluating the targeting effectiveness of social transfers: A literature review.

Eskander, S., Fankhauser, S., Jha, S., Batool, S., & Qaisrani, A. (2018). Do natural disasters change savings and employment choices? Evidence from Pakistan.

Eyre, R., De Luca, F., & Simini, F. (2020). Social media usage reveals recovery of small businesses after natural hazard events. *Nature communications*, *11*(1), 1-10.

Foster, S. (2021). Alyans Peyizan Kafé D’Ayiti: A Copowerment Model for the Producers and Consumers of Coffee in Haiti.

Gallant, J. (2016). *Financial Decision Making in the Foxhole: The effects of Conflict on saving Behavior in Kenya* (Doctoral dissertation, Queen’s University).

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

Hernandez, K., & Roberts, T. (2020). Predictive analytics in humanitarian action: a preliminary mapping and analysis.

Hoos, K. (2010). Urban People: Too Poor to Save Money? The Case of Micro Deposit in the Philippines. *Philippine quarterly of culture and society*, *38*(1), 75-83.)

Jaspers, S., & Shoham, J. (1999). Targeting the vulnerable: a review of the necessity and feasibility of targeting vulnerable households. *Disasters*, *23*(4), 359-372.

Karlan, D., Ratan, A. L., & Zinman, J. (2014). Savings by and for the Poor: A Research Review and Agenda. *Review of Income and Wealth*, *60*(1), 36-78.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science advances*, *2*(3), e1500779.

Lam P., Lyer S., Giraudy E., Pompe A., (2022). Facebook Business Activity Trends for Disaster Recovery. 272137999_3000199660246151_8348540010599272876_n.pdf (fbcdn.net)

Maas, P., Iyer, S., Gros, A., Park, W., McGorman, L., Nayak, C., & Dow, P. A. (2019, May). Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery. In *KDD* (Vol. 19, No. 3173).

McCord, A. (2013). Community-based targeting in the Social Protection sector. *Working Paper No. 514*.

Reyers, M. (2019). Financial capability and emergency savings among South Africans living above and below the poverty line. *International Journal of Consumer Studies*, *43*(4), 335-347.

Schnitzer, P. (2019). How to target households in adaptive social protection systems? Evidence from humanitarian and development approaches in Niger. *The Journal of Development Studies*, *55*(sup1), 75-90.

Schwartz, T., Maass, H., & DeMattee, A. COORDINATION NATIONALE DE LA SÉCURITÉ ALIMENTAIRE.

Sharp, K. (1998). *Between relief and develpoment: targeting food aid for disaster prevention in Ethiopia*. Overseas Development Institute.

Tkachenko N, Jarvis S, Procter R (2017) Predicting floods with Flickr tags. PLoS ONE 12(2): e0172870. https://doi.org/10.1371/journal.pone.0172870

Zou, L., Lam, N. S., Cai, H., & Qiang, Y. (2018). Mining Twitter data for improved understanding of disaster resilience. *Annals of the American Association of Geographers*, *108*(5), 1422-1441.