UNIVERSITY OF CAPE COAST

THE PREFERENCE OF CASH FOR TRANSACTIONS AMONG MOBILE MONEY USERS IN KENYA: A MACHINE LEARNING ANALYSIS

SAMUEL ANSU GYAN

2024

UNIVERSITY OF CAPE COAST

THE PREFERENCE OF CASH FOR TRANSACTIONS AMONG MOBILE MONEY USERS IN KENYA: A MACHINE LEARNING ANALYSIS

BY

SAMUEL ANSU GYAN

Dissertation submitted to the Department of Data Science and Economic Policy of the School of Economics, College of Humanities and Legal Studies, University of Cape Coast, in partial fulfilment of the requirements for the award of Master of Science in Data Management and Analysis.

JULY 2024

DECLARATION

Candidate's Declaration

I hereby declare that this dissertation is the result of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature: ………………….. Date: 22nd November, 2024

Name: Samuel Ansu Gyan

Supervisor's Declaration

I hereby declare that the preparation and presentation of the dissertation were supervised in accordance with the guidelines on supervision of project work laid down by the University of Cape Coast.

Supervisor's Signature: …………………… Date: 22nd November, 2024

Name: Dr. Joshua Sebu

ABSTRACT

Digital finance platforms like Mobile Money have been a game changer in many African economies, gaining widespread uptake and improving the narrative of financial inclusion. Research shows that in Kenya, where the Mobile Money market is the largest in Africa, hundreds of thousands of lives have been improved through access to the service. At the same time, there exists evidence suggesting a rather broad preference of cash for retail transactions. Using open data from a field survey conducted in 2019, I empirically test the level of cash preference among Mobile Money users in Kenya and predict same with 11 user experience variables on 818 users in a simple classification model. Results indicate that cash preference is invariably high even among Mobile Money users, and that it is associated with negative user experiences. I recommend that the Kenyan government and service providers work together to improve infrastructural security, among other factors, to mitigate the rate of negative user sentiments that could lead to a decline in the usage of the service.

KEY WORDS

Cash Preference

Financial Inclusion

Kenya

Logistic Regression

Mobile Money

Technology Acceptance

ACKNOWLEDGEMENTS

I express sincere gratitude to my supervisor, Dr. Joshua Sebu of the Department of Data Science and Economic Policy for his professional guidance, constructive advice and goodwill, and for encouraging me to see and do beyond what I thought were my limits.

I am also grateful to Dr. William Godfred Cantah, who first introduced me to machine learning, and to Dr. Carl Hope Korkpoe, who taught me how to program in Python; your lessons were the bedrock of this paper.

I thank my mother and father, Mercy Yeboaa Bekoe and Emmanuel Gyan, for their constant prayers and unflinching support; my siblings Ebenezer Bekoe Gyan and Liticia Gyan, for their support and encouragement; and my good friend Foster Owusu Appiah for his unwavering companionship.

DEDICATION

To the inaugural class of the MSc Data Management and Analysis programme; and faculty of the Department of Data Science and Economic Policy.

To my friend Emmanuel Yaw Gyimah, without whose efforts none of this would have started in the first place.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

Mobile Money is a digital financial service that has permeated and grown to claim a considerable part of, if not dominate, the Kenyan monetary transaction space, as well as those of other African countries for over a decade now. It enables users to send money to one another via text messages and store it on their phones. Suri and Jack (2016) found that access to mobile money in Kenya had changed users' financial behaviours, increased financial resilience and savings, and lifted some 194,000 households out of poverty. Creemers, Murugavel, Boutet, Omary and Oikawa (2020) discovered that mobile money transactions were the equivalent of 87% of GDP in Kenya and 82% of GDP in Ghana. They projected Africa's mobile payments revenue to be some $20 billion by 2025. Mobile money, as said before, has grown to address the need for convenience in financial transactions. But actors in monetary transactions – most of whom have either previously been, currently are, or are potential mobile money customers – have been found to prefer cash to mobile money. This study is an empirical analytical project that uses secondary data on a section of the Kenyan mobile money user community to test various hypotheses in an attempt provide an appreciation of how user experiences with mobile money in Kenya may contribute to a preference of cash for transactions.

1.1 Background to the Study

For over a decade, Africa has seen the transformative power of mobile money in charting a pathway to higher financial inclusion on the continent. Parekh and Hare (2020) observed that users in Sub-Saharan Africa (SSA) alone accounted for two-thirds of global transactions as of the last quarter of 2018. The Global System for Mobile

Communications Association's (GSMA) '2017 State of the Industry Report on Mobile Money' shows that in SSA, mobile money had been implemented in 135 distinct territories with 338 million accounts. With presence in 40 African countries ('Mobile Money', 2023), the innovation is enabling Africans to edge up the financial services value chain, shifting the continent closer to achieving the Sustainable Development Goals of no poverty (Goal 1), and economic growth (Goal 8) (UNDESA, n.d.). Mobile payment adoption has allowed customers in SSA access to mobile banking facilities and other services as they open accounts, take personal or business loans, and transact with ease. (Sy, Maino, Massara, Perez-Siaz, & Sharma 2019).

Additionally, mobile money has advanced gender equality (Goal 5) by giving women greater control over their finances and lifting the barrier of dependence on male relatives for financial transactions (UNSGSA, 2018). Mobile money has bridged the gap between rural and urban populations via financial inclusion, reducing the rather stark inequality (Goal 10) in accessing financial services (Mpofu, 2022). It also has the potential to scale access to economic opportunities, as well as to essential services like healthcare and education.

The project that birthed mobile money started as a coaction between Vodafone and the UK Department for International Development (DfID) in the early 2000s, at a time when the DfID was struggling with getting funds across to Kenyans in rural farm areas (McGath, 2018). Following a presentation by Vodafone at the World Summit for Sustainable Development in Johannesburg in 2002, urging corporations to support poor countries' development, the DfID proposed to invest £1 million in a "mobile phones for microfinance" project, provided Vodafone made the same commitment (Harford, 2017).

Africa’s first cell phone-based money transfer system, M-Pesa, was introduced by Safaricom in Kenya and Vodacom in Tanzania. Both Vodafone subsidiaries, they are the largest telcos in their respective countries. ‘M’ stands for mobile, and Pesa is the Kiswahili word for money. As of 2019, M-Pesa boasts of 30 million users in 10 countries (Sy et al., 2019).

According to Harford (2017), the project was piloted in 2005 as a system by which microfinance clients could take and repay their loans via SMS. But it leapfrogged into a medium for several other uses like peer-to-peer transfers, and a means to carry otherwise risky amounts of money over long distances, to name a few.

In the late 2000s, similar digital and mobile financial services (D/MFS) began to emerge in different forms across other African countries. Among the notable innovations that marked this period were the Ghana Interbank Payment and Settlement System (GhIPSS) in 2007; the M-Pesa mobile payments system itself that went mainstream in Kenya in 2007; and the development of ‘SmartCash’ by Ugandan software developer Roland Egesa of Mobitrix Uganda in 2008, to mention a few (‘Mobile Money’, 2023). Other innovations followed these and began to take shape together with the early few.

In Ghana, the GhIPSS, which was introduced to help transform Ghana into a cashless society, rolled out e-Zwich – a smart card connected to all Ghanaian banks for the settlement of transactions – in 2008 (Breckenridge, 2010). It, however, suffered significant declines due to mechanical issues that hindered transactions. Nonetheless, the GhIPSS sustained its mandate, coming back a decade later with the Mobile Money

Interoperability system which facilitates seamless cross-platform mobile money transfers (Ifeanyi-Ajufo, 2022).

In Nigeria, which had Africa's largest unbanked population, Orekoya (2017) reports that "the Mobile Money Transfer programme was jointly launched by the GSM Association (GSMA) and Western Union in October 2007". A year after that, the volume of mobile money transactions in Nigeria stood at 3.2 million with an accompanying value of 700 million Naira. These figures grew to a volume of 15.8 million transactions valuing about 142.8 billion Naira in 2013; and to 47 million transactions, 756.89 billion Naira in 2016 (Central Bank of Nigeria, 2017).

Shortly after M-Pesa was first deployed by Safaricom in Kenya, where the mobile money market is the largest on the continent, the GSMA provided support for Safaricom's development of a mobile-based payment project, M-Pesa, by way of a grant 2009. By the end of that year, M-Pesa had over 8 million users in Kenya. This figure would double over the next three years to 16 million customers with more than 30,000 agents in Kenya (Piper, 2020).

The transition from cash to mobile money in Kenya was primarily a response to the constraints of financial exclusion. Research shows that before M-Pesa, less than a fifth of the largely rural population had access to any type of formal financial service (Song, 2016). This was due to factors such as high costs of bank account ownership, the physical remoteness of banks, and a lack of formal identification to facilitate access to banking services. Adoption of the service was also significantly enabled by mobile phone penetration in Kenyan society. (Oteri et al., 2015) note that when M-Pesa went live in 2007, 30% of Kenyans had access to mobile devices, and that they expected mobile penetration to reach 70% in 2016.

The innovation of mobile payments would thus democratize access to financial services as a low-cost, secure, and convenient alternative to traditional banking—which was mostly confined to the wealthy. M-Pesa leveraged these insights to establish a network of intermediary agents who allowed users to deposit and withdraw cash easily. This model proved especially beneficial for those living in remote areas where banks were scarce. As users became familiar with the convenience and security offered by mobile transactions, M-Pesa gained popularity, leading to a substantial increase in its user base and transaction volumes (Ng'weno & Bill & Melinda Gates Foundation, 2010).

More recently, extraneous factors such as the coronavirus pandemic further accelerated the shift towards mobile money as governments encouraged cashless transactions to mitigate virus transmission risks. During this period, the Central Bank of Kenya waived fees associated with mobile money transactions, which further incentivized users to adopt mobile money services over cash (Sehlolo, 2022). The convenience of sending money instantly and securely transformed not only individual financial practices but also broader economic activities, enabling businesses to operate more efficiently and families to manage their finances better.

As a result, mobile money services like M-Pesa have become integral to Kenya's economy, significantly improving poverty alleviation and economic resilience. Oxford Business Group (2018) reports that the value of mobile money transactions in Kenya grew from $1.6 billion in 2008 to $35.3 billion in 2017, attributing this to Safaricom's

M-Pesa. The group reports that M-Pesa claimed over 81% of Kenya’s mobile money space with some 29.1 million subscribers as of 2018. M-Pesa is regarded as the key factor behind the rapid decrease in Kenya’s financial exclusion rate, which went from 73.3% to 24.7% over the decade ending 2016. With over two thirds of its population registered onto mobile money services, the country has the highest rate of penetration in East Africa, and the 10th highest in SSA (GSM Association, 2017). But while mobile money services have gained widespread adoption in Kenya, offering a digital alternative to traditional cash transactions, some users may still prefer cash for various reasons.

1.2 Statement of the Problem

Multiple studies point to a broad preference of cash to mobile money for retail transactions in Kenya. Collins et al. (2012) discovered that despite the country’s reputation for leadership in digital payments, 99% of the retail transactions they sampled were done in cash. The Central Bank of Kenya's (2016) FinAccess household survey reported that just about 5 in 100 business owners primarily used digital payments, with likewise low levels of digital payment use among casual (5.3%) and agricultural workers (7.2%). As of recently as 2022, cash withdrawals remain the topmost use case for the country’s leading mobile money service, M-Pesa, and 90% of transactions are still done in cash (Ajene, 2022; Nnamani, 2022).

These findings raise questions, especially given that the proportion of Kenyans with access to mobile payments is about 93% (McGath, 2018). Several reasons have been said to account for this narrative. Some key findings include the following:

Flood (2018) observed that the majority of labour, which is especially informal, is compensated in cash; and that since transactions using mobile money require multiple steps, some people find it easier to use cash for small transactions. This would corroborate Mas and Ng'weno's (2012) attribution of the situation to the fact that suppliers and other actors across every trade prefer cheques or cash for smaller payments to avoid unfamiliar technical problems. Russon (2019) also highlights a lack of trust in mobile money technology, as well as a large and financially excluded informal sector, as drivers of the apparent preference for cash.

These studies provide a broad view of the situation. But they fail to show, for instance, if the reported level of cash preference is the same for both the banked and unbanked; i.e., in this context, between those with and those without mobile money accounts. There is also insufficient knowledge as to whether, in addition to the attributions made above, the user experiences among those with mobile money accounts contribute to whatever preference for cash there may be.

1.3 Purpose of the Study

In this study, I sought to ascertain the level of cash preference among mobile money users in Kenya; and determine if the preferences are predictable based on their user experiences, using predictive analysis.

1.4 Research Questions

As such, the following questions guided the study:

- What is the level of cash preference among mobile money users in Kenya?

- Are users' preferences for cash predictable based on their usage experiences with mobile money services?

1.5 Significance of the Study

Existing literature presents various tones of the evolution of mobile payments in Africa, with little to learn about the intricacies of its usage. Therefore, this study is a significant diverse contribution to knowledge about the mobile money market, with respect to common user experiences that have accompanied the growth of the service.

Mobile payments have improved access to financial services among millions in Kenya, in Africa, and around the world. As a result, extreme poverty conditions, as well as the unemployment rate, have reduced significantly over time. Knowledge and improvement of customer experiences with mobile money are therefore pivotal in sustaining the service for continued economic growth.

This study therefore helps current mobile money service providers across the value chain to identify and address critical issues within the prevailing ecosystem. The findings and recommendations herein also provide significant insights for various state and private actors, as indicated before, to consider during research and development in economies that have yet to catch on to the mobile money aspect of digital finance.

In summary, this study helps:

- Identify user segments with a higher preference for cash, helping service providers tailor their offerings.
- Inform the design of targeted marketing strategies to promote mobile money usage.

- Offer insights to policymakers for interventions aimed at reducing cash dependency and promoting financial inclusion.

1.6 Delimitations of the Study

- Economically and geographically, the study is delimited to mobile money users in Kenya. Data was collected based on the usage of and experiences with mobile money services.

1.7 Organisation of the Study

Chapter one presents an informative opening to the study by way of a brief introduction, followed by a concise background to the study, a statement of the research problem, questions guiding the study, significance of the study, delimitations of the study, limitations, definition of terms and organisation of the study. Chapter two presents a review of relevant literature, surveying the theoretical and empirical approaches that have guided past studies on mobile money and consumer cash preference to digital payments in Africa. Chapter three presents the methodology employed to conduct the analysis. Chapter four presents the results of the study and discussion of the findings. Finally, a summary of the study, conclusion and recommendations for current and potential mobile money market actors are presented in Chapter five.

CHAPTER TWO: LITERATURE REVIEW

This study is geared towards ascertaining the level of cash preference among mobile money users in Kenya, as well as the association between that and their negative user experiences, by way of predictive modelling. In this chapter, I present the theoretical rationale for the hypothesized association between mobile money user experiences and cash preference and delve into the empirical body of literature that have explored the drivers of cash preference in Kenya and other African countries.

2.1 Theoretical Framework

2.1.1 Evolution of the Theory of Technology Acceptance

The growth of mobile money has been driven mainly by its property of being a convenient alternative for transferring money. Its innovation and adaptation have transformed economies and societies across Africa. As the continent rapidly undergoes digitization, the choice between cash and digital payments remains a growing focal point in financial discourse, given how important it is for consumers to fully embrace digital payment systems like mobile money to achieve, among other things, greater financial inclusion. The theoretical significance of user experiences, either as likely determinants of sustained mobile money adoption or possible reasons for cash preference, dependent respectively on whether these experiences are positive or negative, is indicated in the stream of research that has hypothesized consumer behaviour to formulate constructs for understanding technology acceptance.

The foremost Technology Acceptance Model (TAM), as propounded by Davis (1985), lays a fundamental framework for understanding consumer adoption of technology. It helps to appreciate the cognitive and affective elements that lie between system characteristics and technology acceptance on the premise that a potential user's actual use of a particular system is significantly determined by their overall attitude toward the system. Davis's (1985) sought to apply design and implementation theories to enhance appreciation of user acceptance of information systems; and to establish the theoretical basis for a pragmatic approach to "user acceptance testing" that would enable pre-implementation evaluation of proposed new systems. Originally, the technology acceptance model was conceptualized as shown in Figure 1. X_1, X_2 and X_3 represent design features of alternative systems, Perceived Usefulness and Perceived Ease of Use are hypothesized as cognitive response, Attitude Toward Using the technology is the affective response, and Actual System Use, the behavioural response. The arrows represent the causal relationships between these variables.

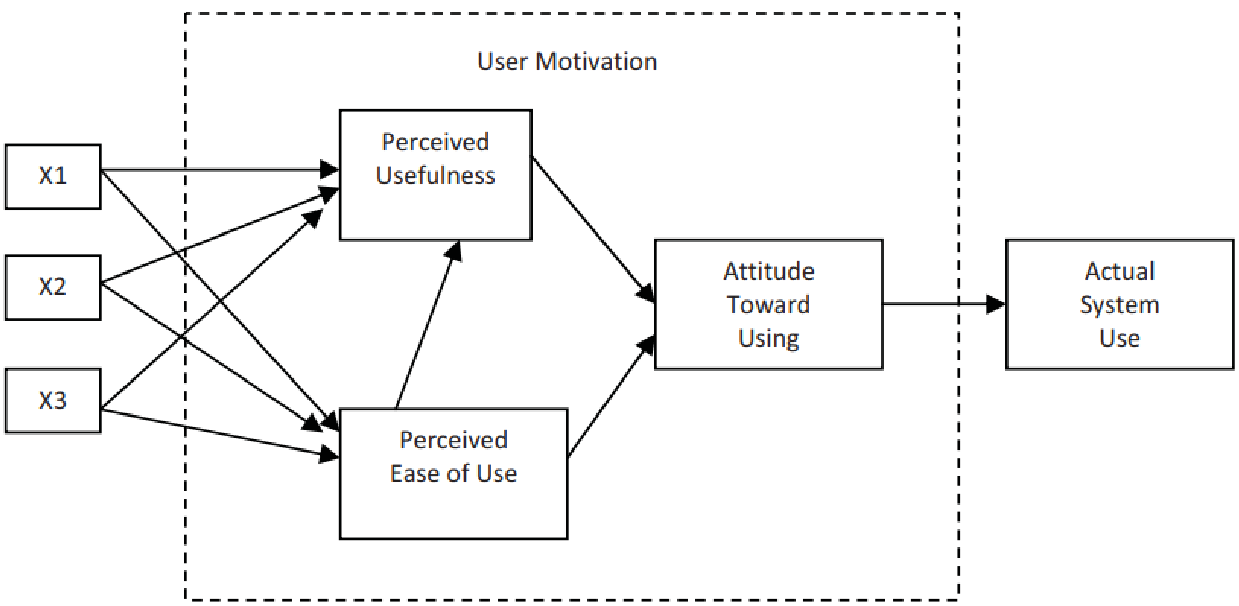


Figure 1: Technology Acceptance Model (Davis, 1985, p.24)

Again, the underlying assumptions of TAM are that an individual's actual use of a system is ultimately determined by their attitude toward using that system, which is in turn influenced by two key notions: *perceived usefulness* and *perceived ease of use*; that perceived usefulness is a direct causal result of perceived ease of use; and that the design features of alternative systems, being external variables, have a direct influence on the two notions.

The concept of technology acceptance has since evolved through further research. Davis (1989) refined his hypothesis of technology acceptance to develop and validate better measures for predicting and explaining technology use. He hypothesized perceived usefulness and perceived ease of use as the primary causal factors of actual system use in an empirical study. He posits that perceived usefulness is “the extent to which an individual considers using a particular system advantageous”; and that perceived ease of use is the extent to which they find using a particular system to be effortless. The study used correlation and regression analyses to examine the relationship between these measures and self-reported indicators of system use. Results show that both variables were significantly correlated with self-reported current usage and self-predicted future usage; and that perceived ease of use may actually be a causal antecedent to perceived usefulness and not a determinant of system use. In other words, the study found that in order for consumers to consider a particular system useful, they may first need to find it easy to use.

In a subsequent study, Davis, Bagozzi and Warshaw (1989) modified the model to include behavioural intention. The guiding principle here was that a potential user might formulate a firm intention to use a system once they had perceived it useful, without necessarily developing an attitude toward it, coming up with an altered adaptation of TAM as shown in Figure 2.

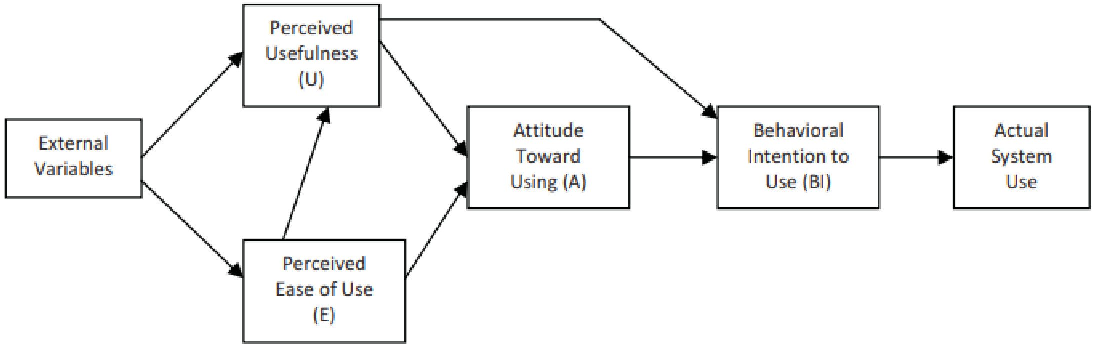


Figure 2: Modified TAM with Behavioural Intention Variable (Davis et al., 1989, p. 985)

Davis et al. (1989) employed the model in Figure 2 to predict people's technology acceptance from their intentions to use a particular system, measured after an hour's exposure to said system, and again after 14 weeks. They found perceived usefulness and perceived ease of use to have significant effects on behavioural intention, which was in turn highly correlated with system use. Perceived ease of use, however, had a relatively small effect which diminished eventually. But the bottom line was that both constructs directly influenced behavioural intention and obviated the need for an 'attitude toward use' variable.

Davis (1993) also reviewed the relationships between the original variables in Figure 1, hypothesizing new direct relationships between perceived usefulness and actual system use, as well as between system characteristics and attitude toward using systems, as shown in Figure 3. These hypotheses followed the discovery that against what had initially been purported in Davis (1985), perceived usefulness could also directly affect actual system use, and that system characteristics could directly affect a potential user's attitude toward using the system without the intervention of perceptions.

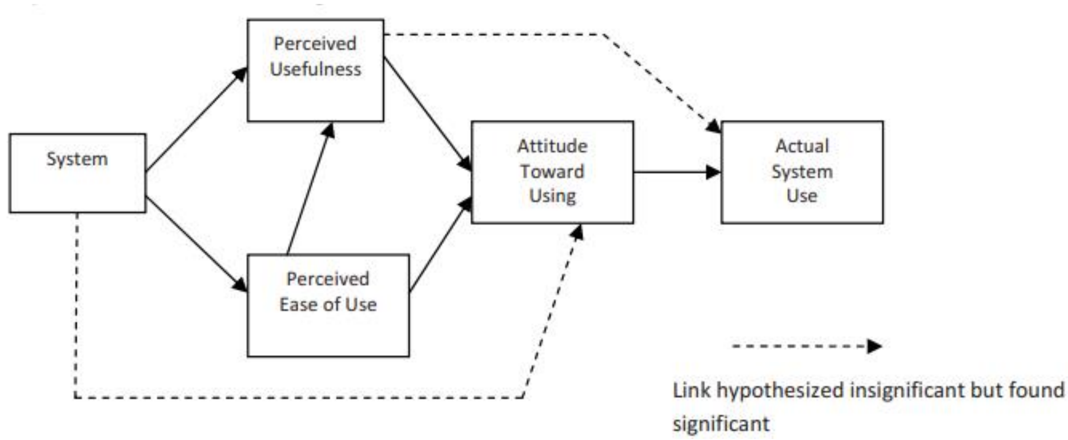


Figure 3: Hypothesis of new relationships (Davis, 1993, p.481)

Using a field study on 112 users of two end-user systems to test the model, Davis (1993) reports that perceived usefulness actually determined usage, with system characteristics also significantly accounting for over a third of the variance in same. But the effect of system characteristics on attitude towards use remained unexplained. Davis and Venkatesh (1996), following Davis et al. (1989) and Davis (1993) thus replaced the construct of attitude toward using with behavioural intention in what Chuttur (2009) describes as the "final version of TAM" (p.10) as shown in Figure 4. Moreover, to improve the determinative power of the model, the researchers considered how, in addition to system characteristics, "external" factors like user education and participation in system design, as well as how a system is implemented, could also inform individual beliefs toward system use. As a consequence, Davis (1993) was able to determine the effect of perceived usefulness on system use, and the hitherto unexplained effect of system characteristics on the attitude variable (see Figure 3) was eliminated (Davis & Venkatesh, 1996).

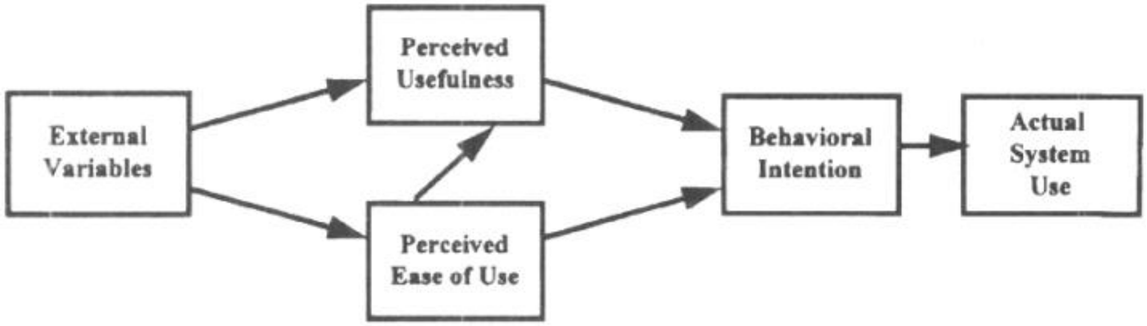


Figure 4: Technology Acceptance Model (Davis & Venkatesh, 1996, p. 453)

Researchers have since extensively applied TAM to study technology adoption in different professional, social, geographical, and economic settings. These include healthcare, education, digital finance, and telecommunications in Africa, the Middle East and Asia, with variables having different levels of influence on consumer behaviour towards accepting and using technologies (Burgess & Worthington, 2021). In the next section, I review how TAM has been applied to study mobile money acceptance in Kenya and other African jurisdictions, the insights that have emerged thereby, and the gap(s) thereof which I seek to address in the current study.

2.2 Empirical Review

2.2.1 Mobile Money Adoption: TAM's Primary Constructs in Context

In the context of mobile money in Kenya and Africa at large, TAM has served as a foundational theory to investigate the drivers and barriers influencing the adoption of digital finance systems. Kenya and other key African economies have witnessed significant advancements in digital finance, yet cash transactions remain more prominent than digital payments. This continuous existence of cash transactions amid a plethora of mobile money services underscores the complexity of technology acceptance. Various studies have applied TAM to scrutinise the adoption of mobile money across Africa. Among other findings, they show how the perception that mobile payments enhance financial transactions and accessibility (perceived usefulness) influences users' intentions to embrace digital finance, and how the notion of simplicity and user-friendliness of mobile money platforms (perceived ease of use) shapes users' attitudes towards adopting these systems.

Lule, Omwansa and Waema (2012) applied TAM to study the determinants of mobile banking adoption in Kenya. The study found that among those of self-efficacy and credibility, perceptions of usefulness and ease of use significantly influenced consumers' attitudes towards using mobile banking systems. Similarly, in Togo, Gbongli, Xu and Amedjoneku (2019) used a structural equation modelling-artificial neural network framework to study the adoption of mobile-based money services "for financial inclusion and sustainability" (p. 1). They report that TAM adequately mediates consumer attitudes and intentions, and that perceived usefulness and ease of use affect adoption decisions with the latter being the most significant factor affecting consumer attitudes. Kelly and Palaniappan (2023) likewise posit that among other perceptive variables like risk, cost and trust, usefulness and ease of use affect users' attitudes towards mobile money, which in turn influence their decisions to use the service. Other studies found usefulness and ease of use to have significant influences on other constructs of technology acceptance than attitude towards use. Elnaiem's (2019) modification of TAM to understand women's adoption of mobile money in Zambia produces the narrative that positive perceptions about usefulness, ease of use and trust in mobile money service results increase women's acceptance and use of same. Lubua and Semlambo (2017) assessed the influence of the same constructs on the adoption of mobile money services by small and medium enterprise (SME) owners in Tanzania. Their findings were that these variables significantly influenced SME owners' behavioural intention to use the service. Tobbin and Kuwornu (2011) also arrived at the conclusion that perceived ease of use and perceived usefulness were key factors in the formation of intention to use mobile transfers in Ghana, as compared to trust, trialability and risk.

The different applications of TAM in distinct contexts yielding the expected causal relationships between its primary constructs affirm the potency of the model in understanding consumer adoption of mobile money technology. Perhaps what is even more striking to note is that against the backdrop of a "final" model which eliminated the attitudinal variable (Davis & Venkatesh, 1996), these studies reintroduce the construct of consumer attitudes into the empirical discourse of technology acceptance and use. It bears mentioning, therefore, that consumers' psychological postures and the contributing perceptions thereof need to be treated as structural imperatives in advancing empirical efforts toward understanding technology acceptance. However, these variables are only primary constructs which are almost generic to any system. It is binding, per the paradigm of the current study, to consider factors that are more peculiar to money, consumers' spending behaviours and their varying economic and social settings, which influence them to either accept mobile money technology or maintain cash as their mode of transaction.

Studies that distinctively investigate cash preference and the reasons thereof are in short supply. But the subject appears in some literature that study the motivations behind the acceptance and adoption of mobile money. Research has it that the acceptance of mobile money in Kenya and other African countries involves a more complex interaction beyond TAM's primary constructs. From trust and security concerns to low literacy levels and infrastructural limitations, payment preferences evolve at the intersection of sociocultural, behavioural, socioeconomic, and technological factors. In the following section, I review those factors that have been argued to influence mobile payment adoption or underpin the inclination towards cash, highlighting the challenges and opportunities for fostering greater adoption of mobile payment methods in Kenya and across other African economies.

2.2.2 Sociocultural and Behavioural Factors

Culturally, the integration of mobile finance into daily life is essential for influencing payment habits and changing consumer behaviour. As mobile payments become the norm for everyday transactions, consumers may become more comfortable using it as a payment method, reducing the preference for and reliance on cash. Yet instances and experiences that undermine the reliability and stability of digital payment systems have kept cash afloat in the Kenyan transactional space. A significant body of literature highlights the cultural and behavioural inclinations that often underpin the use of cash in Kenya. Flood (2018) reports in The Guardian that despite Kenyans preferring to make utility bill payments and urban-rural money transfers through digital payments, 80% of transactions are still done in cash. This observation is attributed to what could be described as the cumbersome nature of the M-Pesa payment process. Other reasons, she writes, are high transaction fees and the informal nature of the country’s workforce. Estimating digitally paid salaries to be just about 10%, Flood (2018) highlights the extra step it takes to move money into the digital environment as a possible deterrent to mobile and digital payment adoption. Likewise, surveys by Rolfe (2019) and Shirono, Chhabra, Das, Fan & Villanova (2021) show that norms, including a preference for tangible transactions and trust in physical currency, are key factors perpetuating the reliance on cash in retail settings. These studies recognize the potential of mobile payments to revolutionize retail transactions and improve economic growth in Kenya, but also concur that when it comes to broadly incorporating mobile payments into everyday life, Kenyan society is just not there yet. Businesses and consumers alike have yet to adopt mobile payments as the primary method of facilitating transactions, leaving cash still deeply in grained in business culture. A similar narrative exists about South Africa, as a survey by Nteta (2017) attributes the low adoption of mobile payments for informal trade in Cape Town to a well-established cash-only-basis cultural norm sustained among traders. The study further illustrates how this norm produces a form of cultural capital because traders and customers either proudly or reluctantly identify with the culture of cash-only transactions entrenched in the informal economy.

It is acknowledged in literature that no single set of behavioural factors can determine mobile money (non)adoption and/or cash use (Khan & Blumenstock, 2016; Suri, 2017). However, the presence of these cultural elements influences how consumers behave with payments in different economic contexts alike. Behavioural factors influencing the preference for cash over mobile payments include knowledge and familiarity, trust, convenience and personal control (Baganzi & Lau, 2017; Davidson & McCarty, n.d.; Dzokoto, Appiah, Chitwood & Imasiku, 2016). Baganzi and Lau (2017) examine trust and risk in mobile money acceptance in Uganda using a field survey. Their work accentuates the importance of the relative confidence people have in cash transactions that has sustained its use. They stress the relevance of trust in mobile money uptake by iterating that even in the face of increasing cases of fraud, there is still a deficiency of guarantee that mobile money agents and telecom employees will not take undue advantage of mobile money users. It goes without saying that as people are naturally protective of their money, this vulnerability heightens risk perception and discourages the use of mobile payments. Studying the factors that drive customer usage of mobile payments in sub-Saharan Africa, Davidson and McCarty (n.d.) posit that trust is perhaps the most necessary precondition for onboarding consumers from unawareness to regular use of mobile money. They explain that trust must be high because for many users, their first transaction with a mobile money service will be to hand over cash to an agent, which agreeably demands that the customer is able to exercise a significant level of trust in both the service and its immediate provider. Dzokoto et al. (2016) write of payment preferences in Ghana and Zambia that choices are influenced by the materiality of the status-quo, which in this context is cash, as opposed to the uncertainty associated with “risky novel” alternatives. They note the trajectory of consumer money moving from physical cash to digital payment methods, but also concede that mobile money has not reached its full potential due to the sentimental attachments consumers have towards the perceived comparative advantage of materiality and tangibility that cash offers.

Similarly, Russon (2019) and Batiz-Lazo, Maixé-Altés and Peon (2023) agree that familiarity with and trust in cash transactions, the reliance on the convenience of not having to worry about network issues, and low financial literacy—for which reason consumers may find it more challenging to understand and use digital payment methods—contribute to the preference of cash over digital payments in Kenya. Batiz-

Lazo et al. (2023) further add that consumers prefer cash for the fact that it provides immediate liquidity, a sense of control, and privacy, which may not be fully replicated by digital alternatives. Mwangi (2017) and Okenyuri & Ouma (2019) likewise emphasize the significance of cash as a symbol of security and familiarity within Kenyan communities, thus contributing to its continued prevalence in everyday transactions.

2.2.3 Socioeconomic and Technological Factors

Research has also delved into the socioeconomic and technological challenges associated with digital finance adoption in Kenya. Factors such as limited access to digital and financial infrastructure in rural areas, concerns about transaction security, and a lack of digital literacy, have been cited in several studies as substantial barriers to the widespread acceptance of digital payment systems in Kenya and other African countries. In a review of literature on mobile payments in sub-Saharan Africa, Boateng and Sarpong (2019) discovered that transaction costs and technological drawbacks including cybercrime, resistance to change, unfamiliarity with procedures, and device (in)compatibility had been found by multiple studies to underlie the preference for cash to mobile payments. Risk perception, ease of use, price transparency, cultural values, and social influence have also emerged as factors that drive consumer adoption of same (Boateng & Sarpong, 2019; Mas & Morawczynski, 2009; Suri, 2017). It stands to reason then that in any assessment of options that does not significantly place mobile payments ahead of traditional cash transactions in terms of these factors, the rational consumer would stick to cash.

Kingiri and Fu (2020) applied a technological innovation system framework in a case study to understand the diffusion and adoption of M-Pesa in Kenya. Among other things, they discover that the uptake of mobile money is attributable, among other things, to the spread of and improved access to ICT-enabled technologies. Soutter, Ferguson and Neubert (2019), studying impact factors and mass adoption of digital payments in Kenya, Nigeria and South Africa likewise found that enabling economic and technological environments significantly drive mobile money adoption, and cause difficult implementation frameworks to evolve. Li, Alhassan, Reddy and Duppati (2019) used data from the Global Findex 2014 to examine how formal financial inclusion affects informal financial intermediation and cash preference in Africa. They found that cash use is attributable to the informal nature of developing economies, convenience for low income earners, low financial education, and a lack of formal financial infrastructure – which limits choices and provides no alternatives to cash usage. Suri (2017), citing Heyer and Mas (2009), also emphasizes the importance of high telecom (technology) penetration in the success of M-Pesa in Kenya.

In Uganda, Wamuyu (2014) uses a survey questionnaire as well as two focus group discussions to examine how contextual factors influence the uptake and continuance of mobile money usage. The literature presents insights on the money transfer practices that predated mobile money and how they may have influenced its uptake; as well as the infrastructural characteristics of mobile money that influence consumers' intentions to continue its use. The study results show that the inconveniences that come with informal money transfer methods, like the anxiety of having to wait hours for money to arrive via public transport, influenced the rapid uptake of mobile money, and may sustain its usage. Yet between these themes, the study highlights transaction costs and system volatilities that enable fraud and provide little to no room for redeeming wrongly made transactions, as usage challenges that hinder the complete adoption of mobile money as the primary spending mechanism. Simione and Muehlschlegel (2023) show comparatively that in Uganda, while mobile money users are more likely than non-users to perceive cash as risky and therefore dislike carrying large amounts of cash, it remains a viable option due to a lack of interoperability infrastructure and high transfer costs. Age, educational attainment, gender, settlement type, income, and access to banks were also found to be significantly associated with mobile money usage. According to the study, individuals with a minimum of upper secondary education are more likely to use mobile money than those without; more men than women use mobile money; and rural dwellers are less likely to use the service relative to urban folks. High-income individuals and those with bank accounts were also found to be more likely to use mobile money than low-income individuals and financially excluded people. It is deducible from these studies that in rural, less developed and low-income settings, where access to technology remains nearly non-existent at best, transactions may understandably be done in cash. Evidently, cash preference and use are as cultural and behavioural as they are economic and infrastructural in terms of motivations. But while cash remains deeply entrenched due to these factors, the potential of digital finance to transform the retail landscape is evident, contingent on addressing existing barriers and fostering enabling environments for the adoption and utilization of digital payments. Efforts towards achieving this should seek to identify and address these circumstantial challenges faced by both merchants and consumers within their respective geographical and economic contexts.

2.3 The Knowledge Gap and the Current Study

Available literature thus provides insights on the drivers of mobile payment adoption or otherwise across jurisdictions in Africa. This review shows that among other things, perceived usefulness and ease of use influence consumer attitudes towards mobile money (Gbongli et al., 2019; Kelly & Palaniappan, 2023; Lule et al., 2012), behavioural intention to use mobile money (Lubua & Semlambo, 2017; Tobbin & Kuwornu, 2011), and actual acceptance and use of same (Elnaiem, 2019). Yet while TAM provides a foundational framework for understanding technology acceptance, studies reveal the need for a comprehensive approach that accounts for the multifaceted (external) factors peculiarly shaping the adoption (or rejection) of mobile money in Africa. Studies by Flood, 2018, Nteta, 2017, Rolfe, 2019, and Shirono et al., 2021 confirm the role of sociocultural norms and values in sustaining cash transactions in Kenyan and other societies, which factors also account for the formation of certain consumer behaviours (Baganzi & Lau, 2017; Batiz-Lazo et al., 2023; Davidson & McCarty, n.d.; Dzokoto et al., 2016; Mwangi, 2017; Okenyuri & Ouma, 2019; Russon, 2019). Economic development, education, and technological penetration have also been found to have a direct relationship with mobile money technology acceptance and use in different parts of Africa (Alhassan et al., 2019; Boateng & Sarpong, 2019; Kingiri & Fu, 2020; Mas & Morawczynski, 2009; Simione & Muehlschlegel, 2023; Soutter et al., 2019; Suri, 2017; Wamuyu, 2014).

The synthesis however exposes a discernible gap in how the construct of technology acceptance has been applied to mobile money adoption and use, given that the existing body of knowledge mainly discusses pre-adoption variables and their influence on initial adoption of the system. There is a paucity of empirical evidence as to whether usage experiences could lead consumers to form attitudes that potentially translate to discontinued usage. This study therefore does not fully employ the TAM model, as the results on the primary constructs would be redundant. Rather, yet still with the theory of technology acceptance in mind, I suppose cash preference to be a post-adoption 'attitude toward use' dependent variable, possibly influenced by the experiences and resultant notions that mobile money users form about the system. The role of predictive modelling in this hypothesis is therefore to classify users as either having a preference for cash or not. This will be based on the significance of the association between self-reported negative experiences with, as well as sentiments about, mobile money technology, and self-reported preferences for cash to mobile money for transactions or vice versa. Ultimately, if cash preference proves to be associated significantly with negative user experiences, most users who identified as preferring cash will be correctly classified as such, and, as part of conclusions, such experiences may be situated as factors contributing to the formation of attitudes that threaten the continued acceptance and use of mobile money technology. This study is potentially a substantive contribution of literature to the evolving discourse on financial inclusion and mobile money adoption in Africa, and may break ground for a reimagination of technology acceptance hypotheses to consider the inclusion of post-actual system use variables.

2.4 Definition of Terms

Algorithm – A well-defined, step-by-step procedure that a computer follows to perform a specific task or solve a particular problem.

Classification – A supervised learning task where the goal is to assign a predefined label or category to a given input data point based on its features.

Features – The independent variables in the regression estimation.

Labels – The values of a target variable in a classification problem.

Supervised Learning – A machine learning estimation in which there is a target variable.

Target – The dependent variable in the regression estimation.

Unsupervised Learning – A machine learning estimation in which there is no target variable.

CHAPTER THREE: RESEARCH METHODS

3.1 Research Design

Nonlinear regression models in machine learning, like linear models, rely on probability and statistics to mediate relationships between variables. In order to determine that an observation belongs to one of two (or more) classes against the null hypothesis that no relationship exists between a predictor and a dependent variable, an algorithm computes a probabilistic measure of association between predictor variables and the dependent (target) variable. Then the algorithm determines which label to assign to said observation based on a given probability threshold. Following these generically hypothetical and computational demands of the nonlinear regression, I adopted a quantitative experimental approach to the current study.

An experimental research approach is well-suited for this study because it allows for the systematic evaluation of how specific factors may jointly influence the preference for cash versus mobile money. In a machine learning framework that classifies users based on their responses to carefully chosen independent variables, this approach effectively simulates a controlled environment where the relationships between variables can be analysed. In the current study, the experimental nature of classification models allows for testing hypotheses about the predictors of cash preference and extracting the signal that variables such as trust, service reliability, and security, give about the incidence of user inclination towards cash.

Additionally, while this study does not manipulate variables directly (as in a randomized control trial), it systematically tests how a combination of variables predicts the target outcome, mirroring an experimental design in its rigor and ability to uncover patterns. This method is particularly valuable in gauging user behaviour from reported experience, and may provide reliable and actionable insights into the drivers of cash preference among mobile money users in Kenya.

3.1.1 Supervised Learning - An Experimental Approach

A classification framework best fits the problem in this study because the target variable is binary. Classification is a supervised machine learning method where a model tries to correctly label some given input data (*Classification in Machine Learning*, n.d.). Supervised learning involves splitting the dataset into two parts namely: the training set and the test set. Cross-sectional data is used to build predictive models that can generalize to new, unseen data. The data collected includes feature variables that are used to predict a target variable—which could either be a class label (which would require classification as is the case in this study) or a numerical value (which would require linear regression).

In classification, the model is fully trained using the training data, then it is evaluated on test data before being used to perform prediction on new unseen data. Both the training and test sets comprise random samples of the features (user experiences) and the target variable (cash preference). The entire training set is used to train the machine learning model. Thus, both the features and the target variable of the training set are fed to the model to enable it to learn from patterns in the features and associate them to the respective labels in the target variable.

Like most machine learning projects, as concluded by Kamiri and Mariga (2021), this study employs an experimental approach to supervised learning. The element of experimentation in supervised learning manifests in testing, evaluating, and refining where necessary, the trained model (Hayduk, 2022; Langley, 1988). At this stage in the current study, only the features of the test set were introduced to the trained model for label prediction, accordingly, based on what had been learnt from the training set. The actual labels in the target variable of the test set were subsequently compared to those predicted by the model to assess the model's accuracy, among other metrics of evaluation.

I used the logistic regression algorithm in this study because it is the standard framework for assessing the relationship between a binary target and its predictors (Marschner, 2015); and it is the suitable algorithm to use when the target variable is dichotomous (Gupta, 2022). Moreover, it has certain significant advantages over the other classification algorithms, especially the decision tree, random forest, and support vector machine, which are also common. These advantages mainly include its simplicity and interpretability, and the fact that it is less prone to overfitting (Varghese, 2019). The logistic regression algorithm also has higher computational efficiency (Kraan, 2020); and handles imbalanced data significantly better than other classifiers (Gupta, 2022). This became evident in testing, as the model showed a significant ability to discriminate between classes (see Figure 24 in Chapter 4).

The logistic regression notation is of the form: $y = \frac{e^{\beta 0+ \beta 1X}}{1 + e^{\beta 0+ \beta 1X}}$, where 'y' is a binary indicator for whether or not a respondent prefers cash, β_0 is the constant of the model, and β_1 is the coefficient of the regressor(s) 'X' on which 'y' is regressed. 80% of the data was used to train the model, and the remaining 20% was used to test it. Post evaluation techniques and processes like model deployment, monitoring and maintenance, as well as iteration and improvement, are out of the scope of this study.

My sole objective was to evaluate and report on the performance of the classifier strictly for academic purposes.

3.2 Data Source and Description

I sourced the dataset for this study from a publicly available repository on GitHub with appropriate permissions granted for its use. The dataset was collected as part of a broader survey aimed at understanding the transactional behaviours and preferences of mobile money users in Kenya. It features 2442 observations over 29 variables. It covers responses from a diverse demographic, offering valuable insights into the factors influencing users' preference for cash over mobile money. I selected the variables to include in answering the research questions of this study based on their relevance to transactional experiences, trust, understanding of mobile money systems, and perceptions of service reliability. Each of these variables is binary, measured using 'yes' or 'no' responses. This allowed for clear and structured analysis using machine learning classification techniques. The timestamp column of the dataset shows that data collection ended in November 2019.

3.2.1 Variable Definitions, Modes of Measurement, and Justifications for Inclusion

For the machine learning analysis, I used 12 variables: One target variable and 11 predictors. In the following paragraphs, I define what the variables are, how they were measured, and why I included them in the study.

Experience with a mobile money agent being low on cash:

- **Definition**: This variable assesses whether the respondent has encountered situations where an agent could not dispense sufficient cash due to limited funds.

- **Measurement**: Binary response ('yes' or 'no').
- **Justification**: Such experiences may lead users to perceive mobile money systems as unreliable in emergencies, driving a preference for cash as a readily available alternative.

Transaction failure experience:

- **Definition**: Whether the respondent has faced failed mobile money transactions.
- **Measurement**: Binary response ('yes' or 'no').
- **Justification**: Repeated transaction failures can diminish confidence in mobile money systems and reinforce the preference for cash, which is not subject to technological interruptions.

Issue resolution experience:

- **Definition**: Whether the respondent has successfully reported an issue and had it resolved.
- **Measurement**: Binary response ('yes' or 'no').
- **Justification**: Efficient issue resolution builds trust and satisfaction, encouraging continued use of mobile money services instead of reverting to cash.

Understanding of personal data collection:

- **Definition**: Whether the respondent is aware of the types of personal data collected by mobile money providers.
- **Measurement**: Binary response ('yes' or 'no').

- **Justification**: Transparency about data practices can impact trust. Lack of understanding may foster scepticism and influence users to favour cash for privacy reasons.

Possession of terms and conditions:

- **Definition**: Whether the respondent has a copy of the terms and conditions for using mobile money.
- **Measurement**: Binary response ('yes' or 'no').
- **Justification**: Having access to terms and conditions may reflect user awareness and trust in the service, factors that could influence their transactional choices.

Experience with network issues:

- **Definition**: Whether the respondent has experienced service disruptions due to network problems.
- **Measurement**: Binary response ('yes' or 'no').
- **Justification**: Network reliability is crucial for mobile money usage efficiency. Frequent issues may prompt users to retain cash as a backup.

Fraud victimization:

- **Definition**: Whether the respondent has ever been a victim of fraud while using mobile money.
- **Measurement**: Binary response ('yes' or 'no').
- **Justification**: Fraud incidents can significantly erode trust in mobile money and lead users to prefer cash for safety.

Trust in the mobile money service:

- **Definition**: Whether the respondent trusts the mobile money service provider.
- **Measurement**: Binary response (‘yes’ or ‘no’).
- **Justification**: System trust, as discussed in Chapter Two, is a fundamental factor influencing the adoption of digital financial services. Therefore, low trust in any system may drive a preference for the alternative, which in the case of mobile money is cash.

Knowledge of complaint channels:

- **Definition**: Whether the respondent knows the appropriate channels to lodge complaints regarding the service.
- **Measurement**: Binary response (‘yes’ or ‘no’).
- **Justification**: Awareness of complaint mechanisms may enhance users' sense of control and security, which may promote reliance on mobile money services.

Understanding of terms of use:

- **Definition**: Whether the respondent comprehends the terms and conditions of mobile money services.
- **Measurement**: Binary response (‘yes’ or ‘no’).
- **Justification**: A clear understanding of service terms fosters trust and confidence in using mobile money systems.

Trust in mobile money agents:

- **Definition**: Whether the respondent trusts agents facilitating mobile money transactions.
- **Measurement**: Binary response ('yes' or 'no').
- **Justification**: Mobile money agents act as intermediaries between users and the system. Trust in agents is essential for a positive user experience and adoption of mobile money.

Cash Preference (Target Variable)

- **Definition**: The target variable indicates whether a respondent prefers cash over mobile money for transactions. This reflects their inclination to rely on physical currency despite having access to mobile money services.
- **Measurement**: This variable is also binary, measured as 'yes' if the respondent primarily prefers cash for transactions and 'no' if they primarily use mobile money.
- **Justification**: The cash preference variable is the central focus of the study. It provides the outcome to be analysed in the classification framework. Understanding why some users maintain a preference for cash despite the widespread availability of mobile money is crucial for identifying barriers to digital financial inclusion.

The selected variables represent critical dimensions of user experience, service reliability, trust, and understanding of mobile money systems. Each variable addresses potential barriers or enablers of mobile money usage. This allows for the computation of a plausibly actionable joint signal of factors driving cash preference. The binary nature of the variables ensures variable type uniformity and eliminates the need for linear transformations, which facilitates effective classification.

I also explored five demographic variables to understand respondents' age distribution, gender, their settlement types and educational backgrounds, as well as the types of financial accounts they owned. These variables capture essential dimensions of user context, including geographical access (settlement type), financial behaviour and resources (financial account type), generational patterns (age), societal roles and disparities (gender), as well as cognitive and technological engagement (educational background). They provide a fair understanding of financial service consumer segmentation in Kenya, and hence of the demographic and socio-economic factors that may influence cash preference among mobile money users. Thus, through the analysis, I uncovered patterns and trends that shape transactional preferences, and insights into how diverse factors may interact to inform the choice between cash and mobile money. These variables also help identify potential areas for targeted interventions to promote digital financial inclusion. Their descriptions are as follows:

Settlement Type

- **Definition:** Settlement type captured whether the respondent resides in an urban or rural area, reflecting the geographical context of their daily activities and access to financial services.
- **Measurement:** This variable is binary, classified as a 'yes' or 'no' response to whether the respondent resides in an urban area.

- **Justification:** Settlement type is a significant factor in financial behaviour, as rural areas often face challenges such as limited agent networks, poor or even absent infrastructure, and a generally lower digital literacy among dwellers. Comparing urban and rural preferences provides insights into how geographical disparities may influence the use of cash versus mobile money.

Financial Account Type

- **Definition:** Financial account type indicates whether the respondent primarily uses a mobile money account, a bank account, or some other type of account for their financial transactions.

- **Measurement:** This variable is categorical, recorded as 'Mobile Money,' 'Bank Account,' 'Village Savings and Loans Associations,' 'Savings and Credit Corporative Organizations,' 'Savings and Credit Corporative Organizations,' or 'Online Bank Account'.

- **Justification:** The type of financial account a respondent uses can shape their transactional habits. Those with access only to mobile money may rely on it out of necessity, while bank users might use cash for certain transactions. Exploring this variable helps identify the interplay between different financial services and cash usage.

Age

- **Definition:** Respondent's chronological age at the time of data collection.

- **Measurement:** This is a discrete numerical variable, measured in years.

- **Justification:** Age often correlates with technology adoption and financial habits. Younger individuals may be more comfortable using mobile money due to greater digital familiarity, while older people might prefer cash due to unfamiliarity, established habits or distrust of digital platforms. Analysing age could help uncover generational trends in cash preference.

Gender

- **Definition:** Gender captures whether the respondent identifies as male or female.
- **Measurement:** This variable is categorical, recorded as 'male' or 'female' based on the respondent's self-identification.
- **Justification:** Gender can influence financial behaviour, with societal roles, responsibilities, and access to resources often differing between men and women. For example, women in some contexts may face more barriers to digital financial literacy, impacting their cash or mobile money preferences. Including gender provides a lens to examine such disparities.

Educational Background

- **Definition:** The highest level of formal education completed by the respondent, reflecting their academic attainment and potential familiarity with technology.
- **Measurement:** This variable is categorical, classified into the following levels: 'primary 1' to 'primary 6', 'secondary 1' to 'secondary 6', 'Technical Vocational Education and Training (tvet)', 'tertiary', and 'other' based on the respondent's self-reported educational qualifications.

- **Justification:** Education influences digital literacy, financial decision-making, and the ability to understand and navigate mobile money systems. Higher educational attainment is often associated with greater familiarity with technology and trust in digital platforms, which may reduce reliance on cash. Analysing educational background helps identify how varying levels of education correlate with transactional preferences.

3.3 Exploratory Data Analysis – Demographic Information

As indicated before, respondents in this survey had different types of financial accounts namely: Mobile Money, Traditional Bank Accounts, Village Savings and Loans Association (VSLA) Accounts, Savings and Credit Cooperative Organisation (SaCCO) Accounts, and Online Bank Accounts. For the purpose of the current study, which is focused on only mobile money users, I used the code in Figure 5 to sift out observations with mobile money accounts only. They were originally 827 out of 2442, representing 33.9% of the total number of observations.

```
momo = (data[data["account_type"] == "Mobile Money"]).drop(["start_time", "end_time"], axis=1)
momo.head(3)
```

Figure 5: Code for sub-setting mobile money user data

The variables used at this stage of EDA, i.e., age, gender, educational background, and settlement type, either had no or minimal missingness, and were ‘clean’ for analysis. Hence there was no need for extensive data preprocessing at this stage. Statistical summaries were obtained using python’s ‘describe()’ method. For all variables, this method provides summaries such as the total number of observations, the number of unique values there are, the most frequently occurring value, and the number of times it occurs. The method further provides the mean, median, standard deviation, the upper and lower quartiles, as well as the minimum and maximum values for numerical variables.

```
momo.describe(include = "all").style.background_gradient(cmap="summer")
```

Figure 6: Code for statistical summaries

To show which age groups the users fell into the most, a boxplot was used to visualize respondents' age distribution. Aside the 25th, 50th and 75th percentiles, the boxplot also shows, for instance, the presence and location of outliers. The boxplot was created using the seaborn visualization package as shown in Figure 7.

```
# Users' age distribution

# Subsetting age variable
ages = momo['age']

# Creating the boxplot
sns.boxplot(ages)

# Labelling and showing the plot
plt.title("Boxplot of Respondents' Ages")
plt.xlabel("Ages")
plt.show()
```

Figure 7: Code for age distribution boxplot

Analyses of respondents' educational backgrounds, gender and types of settlement, were done using annotated bar graphs which were created using seaborn's 'countplot()' method and pandas' bar graph subpackage respectively. Figures 8 and 9 are snippets of the codes used to create the annotated graphs for both variables. Seaborn's countplot method creates a simple bar graph of the variable it is called on. The 'for' loop in Figure 9 iterates the annotation function over the bars of the countplot.

```python
# Bar graph of users' educational backgrounds

# Subsetting the grade completed variable
grades = momo['highest_grade_completed']

# Instantiating the countplot
ax = sns.countplot(data=momo, x=grades)

# Calculating percentage values
total_grades = len(grades)

for p in ax.patches:
    percentage_grades = '{:.1f}%'.format(100 * p.get_height() / total_grades)
    x = p.get_x() + p.get_width() / 2
    y = p.get_height()
    ax.annotate(percentage_grades, (x, y), ha='center', va='bottom')
```

Figure 8: Code for the annotated bar graph of respondents' educational backgrounds

```python
# Stacked bar graph of users' gender by settlement type

# Creating the stacked bar plot
gen_set = pd.crosstab(momo['gender'], momo['urban'])
ax = gen_set.plot(kind="bar", stacked=True, title="Respondents' Gender by Settlement Type")

# Calculating percentage values
total_set = gen_set.sum(axis=1)

# Annotating the bars with percentage values
for i, category in enumerate(gen_set.index):
    bottom = 0
    for j, value in enumerate(gen_set.columns):
        percentage_set = (gen_set.loc[category, value] / total_set.loc[category]) * 100
        ax.annotate(f"{percentage_set:.1f}%", (i, bottom + gen_set.loc[category, value] / 2),
                    ha='center', fontsize='10')
        bottom += gen_set.loc[category, value]
```

Figure 9: Code for the annotated stacked bar graph of respondents' gender by settlement type

Since the stacked bar graph visualizes a cross-tabulation of respondents' gender by the types of settlement they lived in, the 'for' loops in Figure 10 iterate the annotation function over the respective stacks in each bar.

Research Question One: What is the level of cash preference among mobile money users in Kenya?

I built an annotated bar graph of the dichotomous cash preference variable to address this question. The 'for' loop in Figure 10 iterates the annotation function across the two resulting bars in the graph.

```
# Visualizing cash preference

pref = momo['prefer_cash']

ax = sns.countplot(data=momo, x=pref)

# Calculating percentage values
total_pref = len(pref)

for p in ax.patches:
    percentage_pref = '{:.1f}%'.format(100 * p.get_height() / total_pref)
    x = p.get_x() + p.get_width() / 2
    y = p.get_height()
    ax.annotate(percentage_pref, (x, y), ha='center', va='bottom')
```

Figure 10: Code for the bar graph showing cash preference among mobile money users

Research Question Two: Are users' preferences for cash predictable based on their usage experiences with mobile money services?

To answer this question, I used a subset of the mobile money dataset which contained only the variables representing user experiences, to model cash preference. To this study, users' knowledge of terms and conditions of usage, as well as sentiments towards the mobile money service, are also considered experiences.

3.4 Data Preprocessing

An anomaly emerged during the summary statistics stage of the exploratory data analysis. Two user experience variables, 'understand_terms' and 'agent_trust', which are dichotomous and hence, were expected to have unique value counts of two each, rather had unique value counts of three as seen in Table 1.

Table 1: A section of the descriptive statistics table

understand_terms	agent_trust	taken_loan	network_issues	clear_about_fees
776	642	262	814	814
3	3	2	2	2
yes	no	no	yes	yes

Further analysis showed that instead of either ‘yes’ or ‘no’, nine out of the 827 observations had the entry ‘-97’ for the variables above, resulting in them showing three instead of two unique values. These observations were dropped because it was not clear what the cause of the anomaly was and dropping them was not going to cause any significant problems in the analysis. This reduced the number of valid observations to 818. The grid in Figure 11 is a matrix created using the ‘missingno’ package. The ‘missingno’ package is used to visualise variable missingness in Python. It is ideal for understanding the depth of missingness in large datasets such as this one.

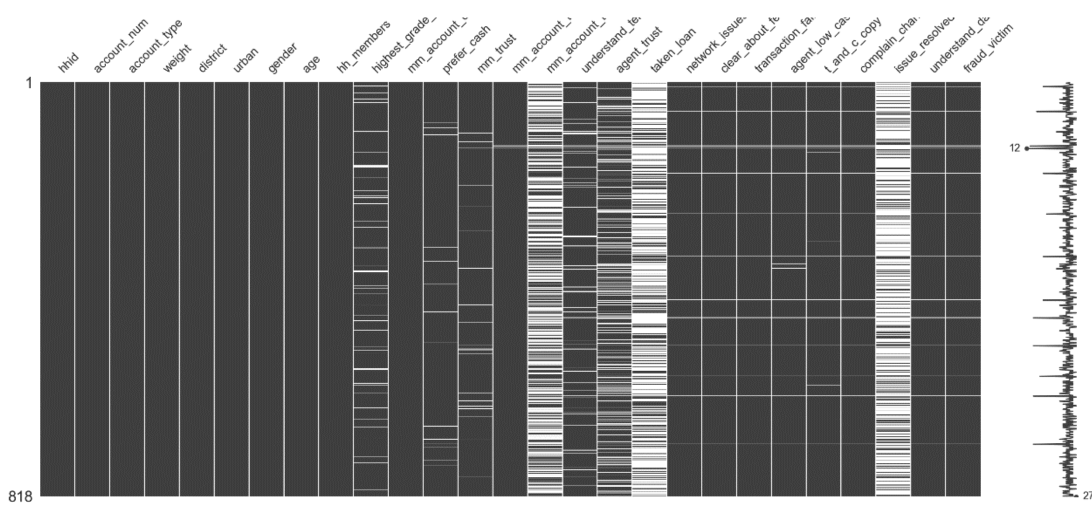


Figure 11: A matrix of the mobile money DataFrame visualizing missingness

All the white spaces in the matrix represent missing values. Due to such excessive missingness, and the fact that dropping that many observations with missing values would have compromised data volume and integrity, I performed a mode imputation using scikit-learn's 'SimpleImputer' package. I used mode imputation because the variables with missingness were categorical, and so measures like the average and median would have been incalculable and inappropriate. Figure 12 shows the code for the imputation process.

```
# Importing the imputer package
from sklearn.impute import SimpleImputer

# Defining the imputation strategy for mode imputation
imputer = SimpleImputer(strategy='most_frequent')

# Fitting the imputer to the data
momo_imputed = imputer.fit_transform(momo)
```

Figure 12: Code for mode imputation using simple imputer

3.5 Exploratory Data Analysis – Feature Selection

The chi-square test for independence was used to identify and select (or omit) features for analysis. I used this test because it is best suited for testing independence between categorical features and a categorical target variable. As an advantage, the chi-square test helps reduce dimensionality of the dataset by removing irrelevant or weakly associated features, which can lead to more efficient and faster model training (Biswal, 2023).

The underlying notation for the chi-squared test is $x^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$, where x^2 is the chi-squared test statistic, Σ denotes the summation over all categories in a contingency table, O_i and E_i represent the observed and expected frequencies in a particular cell of the contingency table. A contingency table is used to compare the observed and expected frequencies of the combinations of categories of the feature and target variable. A larger chi-square test statistic indicates a stronger association between the feature and the target variable. The p-value associated with the test statistic is the criterion that tells the statistical significance of this association. The general threshold for the statistical significance of association is a p-value below 0.05.

Figure 13 shows the code for calculating the chi-squared statistic and the p-value for each variable.

```
from scipy.stats import chi2_contingency

# Create an empty DataFrame to store the chi-squared test results
results = pd.DataFrame(columns=['Feature', 'Chi2', 'P-Value'])

# Calculate the chi-squared statistic and p-value for each binary feature
for feature_name in user_exp.columns:
    contingency_table = pd.crosstab(user_exp[feature_name], user_exp['prefer_cash_yes'])
    chi2, p, _, _ = chi2_contingency(contingency_table)
    results.loc[feature_name] = [feature_name, chi2, p]

# Sort the results by p-value (ascending order)
sorted_results = results.sort_values(by='P-Value')

# Print or analyze the sorted_results DataFrame to identify significant features
print(sorted_results)
```

Figure 13: Code for chi-squared statistic and p-value for each variable

Notwithstanding the potency of the chi-square test for feature selection, another factor that informed this stage of the process was domain knowledge. Thus, beyond selecting features that met the rule-of-thumb criteria of a high value test statistic and a p-value below 0.05, other variables were either added or omitted on the basis that they could or could not plausibly explain cash preference among users. An example of such a variable is whether the user has taken a mobile money loan. Intuitively, this condition is incapable of predicting cash preference, as it does not indicate the occurrence of an unfavourable experience with the service.

3.6 Feature Engineering

Using pandas' 'get_dummies()' method, I encoded the variables as dummies to enable compatibility with scikit-learn, since the library does not accept non-numerical values. The resulting DataFrame showed '1' in place of 'yes' for every variable, and '0' in place of 'no'. The instances of positive responses '1' for cash preference were the class of interest.

3.7 Model Trialling and Selection

Exploring the target variable, I encountered the problem of class imbalance such that a large percentage (95.5%) of users answered 'yes' to preferring cash to mobile money for transactions. This development necessitated the employment of appropriate strategies to cater for the minority class, while testing various models to determine and select which one does the best job differentiating between both classes.

I performed a short experiment with five algorithms namely, Logistic Regression, Decision Tree, Random Forest, Support Vector Machines and Gradient Boost, to find out which algorithm would work best with the data. To identify the most fitting classification algorithm among the five that were tested, I undertook an extensive process which incorporated minority oversampling techniques. Each model was rigorously assessed across various metrics, including accuracy, precision, recall, F1 and the Area under the Receiver Operating Characteristic Curve (ROC-AUC) scores. The ROC-AUC score is an overriding metric because in context, a score close to 1 shows that the algorithm is able to tell positive and negative classes apart very well. Thus, if an algorithm has a score of 0.5 its performance is considered equivalent to an outcome of random chance. In other words, notwithstanding impressive accuracy, precision, recall and F1 scores, a ROC-AUC score of or closer to 0.5 implies that the model is effectively useless. So, despite some algorithms demonstrating slightly better accuracy, precision, recall and F1 scores, the logistic regression algorithm stood out for achieving the most acceptable ROC-AUC score. This holistic process highlighted its comparative advantage at handling imbalanced data, making it suitable for the task at hand, and was the deciding factor in its selection.

Figure 14 shows the importation of the various packages for the exercise. 'RandomOverSampler' was used to oversample the minority class to create some distributional balance in the target variable for model training. The 'train_test_split' package was used to set the percentage of the data to be used for testing (and by extension training) the model. The package also provides for the indication (and maintenance) of a standard of randomization during the data split. The packages imported from 'sklearn.metrics' were used to evaluate the model at different levels.

```
# imbalanced learn for oversampling
from imblearn.over_sampling import RandomOverSampler

# train_test_split is used to split the data into training and test sets
from sklearn.model_selection import train_test_split

# Importing the logistic regression classifier
from sklearn.linear_model import LogisticRegression

# importing the model evaluation metrics
from sklearn.metrics import (classification_report, confusion_matrix,
                             accuracy_score, roc_curve, auc)
```

Figure 14: Code for module importation

3.8 Model Training

Model training is a fundamental stage in classification where the algorithm learns patterns and relationships within the dataset to predict outcomes in the target variable. During model training, the algorithm adjusts its internal parameters based on the input data and the target. The goal of training is to build a model that can extrapolate learnt patterns to make accurate predictions outside of the given sample.

Training the model first required splitting the data into training and test sets. Again, I used 80% of the data to train the model, and 20% for testing as indicated by the 'test_size' argument in Figure 15. The 'random_state' argument ensures that the same data points are assigned to the same sets in different runs of data splitting; 'ros' is an instantiation of the 'RandomOverSampler' which was used to resample the training set; 'logreg' is an instantiation of the logistic regression classifier which can take specific parameter indications; and 'y_pred' is a series of labels predicted by the model using the test features.

The next step was to train the model using the resampled training set (X_train_resampled and y_train_resampled), but then evaluate the model using the original test features. This is because by oversampling the minority class, my aim was to create a more representative training set that enables the model to learn from a more balanced distribution of classes. However, during evaluation, it was crucial to use the actual test set—the unused data—because this set reflects real-world scenarios and provides an accurate assessment of how well the model generalizes to unseen data. So while the resampled training set helped in training the model effectively, the separate test set ensured an unbiased evaluation of its performance on new, authentic data points. This approach helps gauge the model's real-world predictive capabilities beyond the trained dataset.

```
# Creating training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state=42)

# Instantiating the Oversampler
ros = RandomOverSampler(random_state=42)

#Fitting the resampled values to the data
X_train_resampled, y_train_resampled = ros.fit_resample(X_train, y_train)

# Instantiating the classifier
logreg = LogisticRegression()

# Model fitting
logreg.fit(X_train_resampled, y_train_resampled)

# Predicting the labels of the test set
y_pred = logreg.predict(X_test)
```

Figure 15: Code for model training

3.9 Model Evaluation

As with all classification problems, a confusion matrix was used to diagnose the model's predictions (Kamiri & Mariga, 2021). The confusion matrix for binary classification is a two-by-two matrix that summarizes a model's prediction performance. It shows the number of instances that are either correctly (truly) or wrongly (falsely) classified. Three main metrics were derived from the confusion matrix to make this assessment. These are Accuracy, F1, and the ROC-AUC scores. The primary metric for model evaluation was accuracy. However, to cater for class imbalance in the target variable, the more robust F1 and ROC-AUC score metrics were used to better understand the model's performance.

The accuracy score is simply the fraction of classified instances that were correct. Thus, in this project, it shows the overall percentage of users who were correctly labelled by the model as preferring cash or not.

Accuracy is calculated as: $\frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positives + False\ Negatives}$.

The F1 score is a scalar value that integrates the precision and recall scores into one metric to give a better understanding of model how the model operates. It is particularly useful to this study because users with a preference for cash far outnumbered those without, posing the problem of class imbalance in the target variable. F1 score differs from accuracy score in that beyond the mere number of incorrect predictions, it considers whether the reported errors are false positive or false negative. As both the precision and the recall metrics are rates, F1 score is computed as their harmonic mean. Thus, F1 score = $2\left(\frac{Precision \times Recall}{Precision + Recall}\right)$. The F1 score value ranges from 0 to 1, with 1 being the best value.

Precision (also known as the Positive Predictive Value) is a measure of the fraction of the total number positive predictions that are truly positive. A high precision score means the classifier has a low false positive rate. That is, not many users who prefer cash are classified as though they do not. Precision is calculated as: $\frac{True\ Positives}{True\ Positives + False\ Positives}$.

Recall (or hit rate) is the ratio of truly positive predictions to the total number of actual positive instances. It is calculated as: $\frac{True\ Positives}{True\ Positives + False\ Negatives}$. A high recall indicates that the model rightly predicted most users as preferring cash. These metrics were tabulated using scikit-learn’s classification report library.

The Receiver Operating Characteristic Curve and the Area Under the Curve (ROC–AUC).

Again, due to class imbalance in the target variable, it could not be concluded that the model was useful based on accuracy and F1 scores only. This is because high accuracy and F1 scores could still be achieved even if the model predicted the majority (positive) class all the time. A useful metric for model evaluation in this situation was the area under the ROC curve – a line plot of the ratio of true positives to false positives there are at different thresholds (Mandrekar, 2010).

A false positive classification is an event where an instance which does not belong to the positive class is classified as belonging to it. Thus, with respect to this study, an event where a user without a preference for cash is classified as preferring cash. A model's ability to classify negative instances as negative is known as specificity, or true negative rate. Therefore the false positive rate (FPR), which can be obtained by subtracting the percentage of true negatives from the total rate of negative instances, is also known as one minus specificity (1–specificity) (Bhalla, n.d.). FPR is estimated as:

$$\frac{False\ Positives}{True\ Positives + False\ Positives}.$$

The ROC curve is helpful for understanding the trade-offs between a model's sensitivity and 1–specificity as the threshold for classification is adjusted. Both the true positive rate (TPR) and the FPR values range from 0 to 1, with 1 indicating the highest rate in either metric.

The AUC value quantifies the probability that the model correctly classifies a random positive instance higher than it classifies a random negative instance. In other words, it helps understand the nuances of a model's performance, providing insights into its ability to discriminate between classes without being sensitive to a specific threshold.

The AUC is estimated as follows: $\sum \frac{(FPR_{(i+1)} - FPR_i) \times (TPR_i + TPR_{(i+1)})}{2}$, where FPR_i and $FPR_{(i+1)}$ are the False Positive Rate values at points *i* and (*i* + 1) on the ROC curve; and TPR_i and $TPR_{(i+1)}$ are the True Positive Rate values at points *i* and (*i* + 1) on the ROC curve (Mandrekar, 2010). AUC ranges in value from 0 to 1. A model with 100% wrong predictions has an AUC of score of 0; one with 100% correct predictions has an AUC of 1. Generally, "an AUC score of 0.5 suggests no discrimination, between 0.7 and 0.8 is considered acceptable, between 0.8 and 0.9 is considered excellent, and more than 0.9 is considered outstanding" (Mandrekar, 2010, p. 1316). Figure 16 shows the code for plotting the curve and computing the score.

```
# Computing predicted probabilities: y_pred_prob
y_pred_prob = logreg.predict_proba(X_test)[:,1]

# Generating ROC curve values: fpr, tpr, thresholds
fpr, tpr, thresholds = roc_curve(y_test, y_pred_prob)

# Calculating the area under the curve
roc_auc = auc(fpr, tpr)

plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, color='red', lw=2, label='ROC curve (area = %0.2f)' % roc_auc)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('The Receiver Operating Characteristic (ROC) Curve')
plt.legend(loc='lower right')
plt.show()
```

Figure 16: Code for plotting the ROC curve and computing the AUC score

Predicted probabilities are the probabilities that a binary classification model assigns to each instance to belong to the positive class. They can be any real number between 0 and 1, where 0 points to a certainty that the instance belongs to the negative class, and 1, the positive class. A threshold is a decision boundary of cutoff values that determine how predicted probabilities are converted into class labels, i.e., yes, or no; positive, or negative. Instances with predicted probabilities greater than the threshold are predicted as positive, and those below, negative (Bhalla, n.d.). The common default threshold – which was used in this study – is 0.5.

3.10 Chapter Summary

This chapter featured the processes and methodologies adopted to transform raw data into actionable insights. It began with brief descriptions of the types of machine learning, and why Supervised Learning was the ideal machine learning type for this problem. Then followed variable description and sub-setting of mobile money data for analysis. Exploratory data analysis was conducted to understand respondents' demographic information using visualizations and statistical summaries.

The chapter further delved into the processes of data cleaning, handling of missing values and dropping observations where necessary to ensure the quality and integrity of the dataset towards machine learning. Feature selection and engineering emerged as key steps, where relevant variables were selected for model building on the basis of their associations with the target variable and encoded for compatibility with the machine learning library.

As part of model evaluation, various metrics were used to assess the model. Ultimately, the ROC-AUC score helped to understand the trade-offs between true and false positives in delineating mobile money users with a preference for cash from those without.

CHAPTER FOUR: RESULTS AND DISCUSSION

The purpose of this study was to determine the level of cash preference among mobile money users in Kenya, and to find out if their preferences could be based on their experiences with the service. This chapter elaborates the results obtained from the logistic regression classification model trained to predict the preference for cash. The final dataset comprised 818 observations, with a 20% test size, resulting in a test set of 164 samples (158 for the positive class and 6 for the negative class).

4.1 Exploratory Data Analysis

Originally, the survey captured a total of 827 mobile money users across three districts. Nine observations were dropped following the emergence of anomalies. This reduced the number of valid mobile money observations to 818. Most respondents lived in rural settlements as indicated in Table 2.

Table 2: Descriptive table of key demographic variables

	hhid	account_num	account_type	weight	district	urban	gender	age	hh_members
count	827.000000	827.000000	827	827.000000	827	827	827	827.000000	827.000000
unique	nan	nan	1	nan	3	2	2	nan	nan
top	nan	nan	Mobile Money	nan	District_A	Rural	female	nan	nan
freq	nan	nan	827	nan	318	601	425	nan	nan
mean	1609.573156	1.000000	nan	355.787488	nan	nan	nan	36.383313	4.719468
std	349.088024	0.000000	nan	279.530173	nan	nan	nan	12.257107	2.050488
min	1001.000000	1.000000	nan	14.582491	nan	nan	nan	18.000000	1.000000
25%	1309.500000	1.000000	nan	188.686040	nan	nan	nan	27.000000	3.000000
50%	1617.000000	1.000000	nan	283.233020	nan	nan	nan	34.000000	5.000000
75%	1912.500000	1.000000	nan	443.208890	nan	nan	nan	43.000000	6.000000
max	2205.000000	1.000000	nan	3157.224700	nan	nan	nan	85.000000	18.000000

The 'hhid' variable indicates respondents' unique household identifier, while 'account_num' represents the number of financial accounts an individual had. The average household (hh_members) had 5 members, with some having as many as 18. For all these users, mobile money was the only form of access to financial services they had. The average respondent was middle-aged, with the youngest and oldest being 18 and 85 respectively. Figure 17 visualizes users' age distribution.

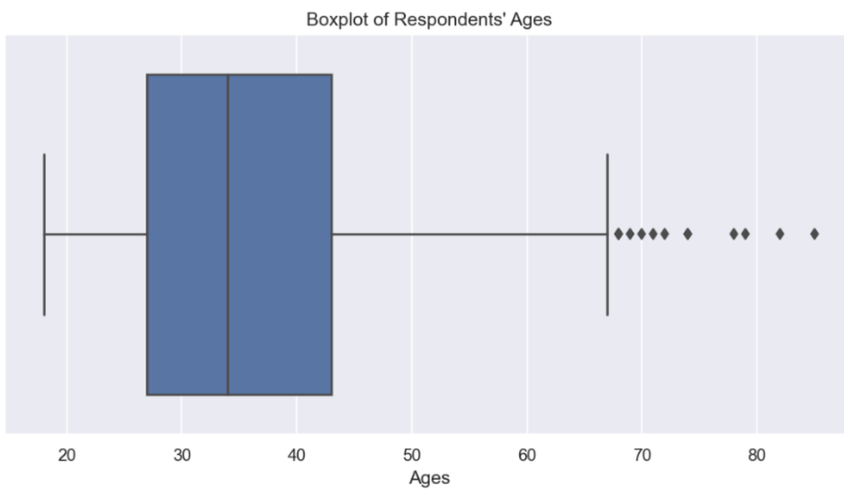


Figure 17: Age distribution

Studies on mobile money in Kenya have been more focused on access and adoption rates with only few giving attentions to age groups who take up the service. Even with those, the focus has largely been on how young adults (mostly aged 18 – 25) use the service (Mintz-Roth, 2018).

The evidence here shows that although skewed towards young people, mobile money usage in Kenya well includes people of all age groups. This emergence leaves a need for further inquiry into how use cases and usage patterns among other age groups may converge or otherwise, relative to those of young people as shown in prior studies.

Ajayi and Ross (2020) showed that mobile money users in Kenya tend to be more educated. They as well reported an association between free primary education and formal financial service use in Kenya, especially via mobile banking. The bar graph in Figure 18 visualizes respondents' highest school grade/level completed.

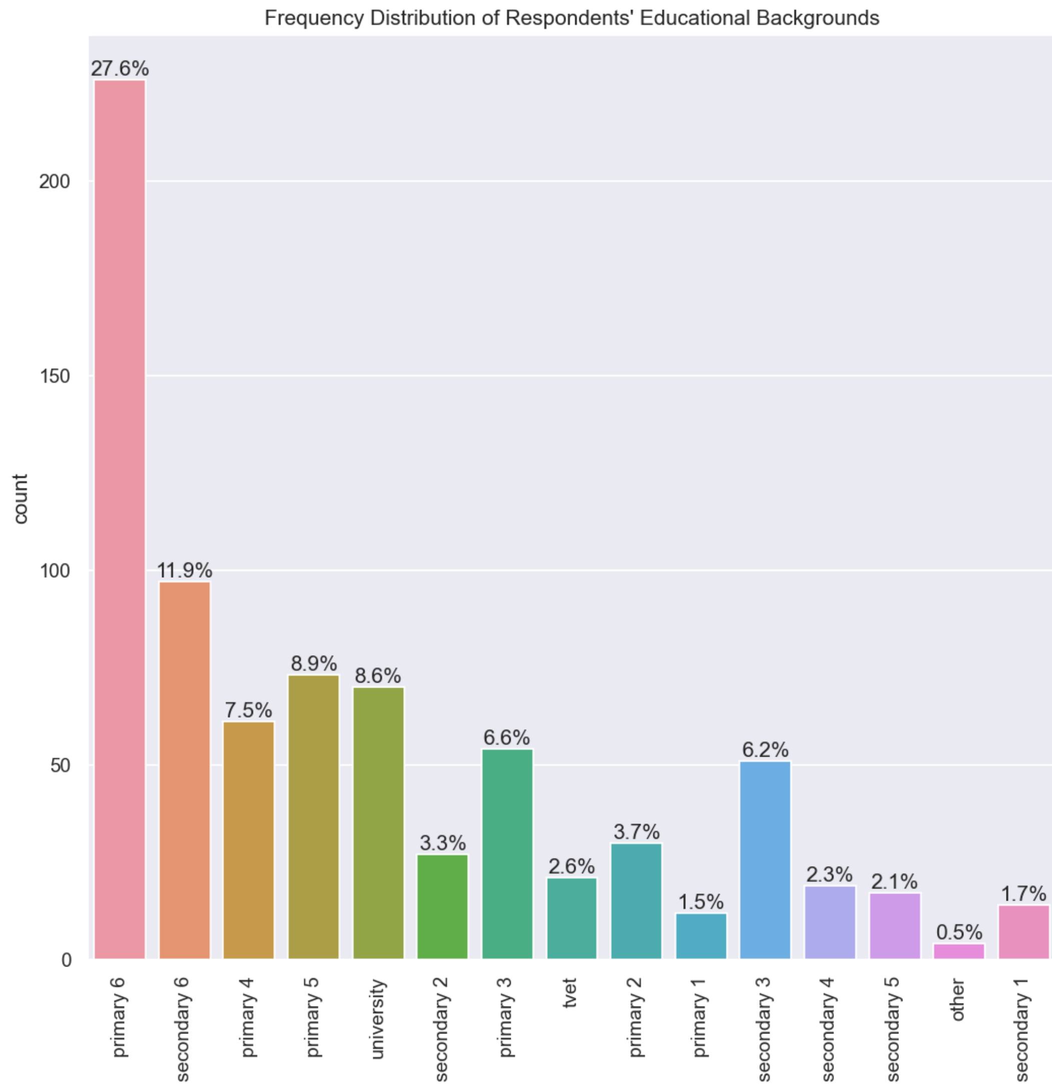


Figure 18: Respondents' educational backgrounds

As of the time of data collection, most respondents had only primary education, with most of those respondents having schooled up to primary six. This lends credence to Ajayi and Ross' (2020) suggestion that mobile money users in Kenya have at least primary education.

A lot of other respondents had also been to secondary school, most of whom had finished secondary six. According to Wikipedia, the secondary school system in Kenya is similar to Ghana's high school system: Students spend three years each in junior and senior secondary school. It is hence perceivable that in Figure 16, secondary one to three would refer to junior secondary school, while four to six represent senior secondary school. Only about nine percent of respondents had attained university education.

Respondents largely lived in rural areas, with barely a third living in urban Kenya, as shown in Figure 19.

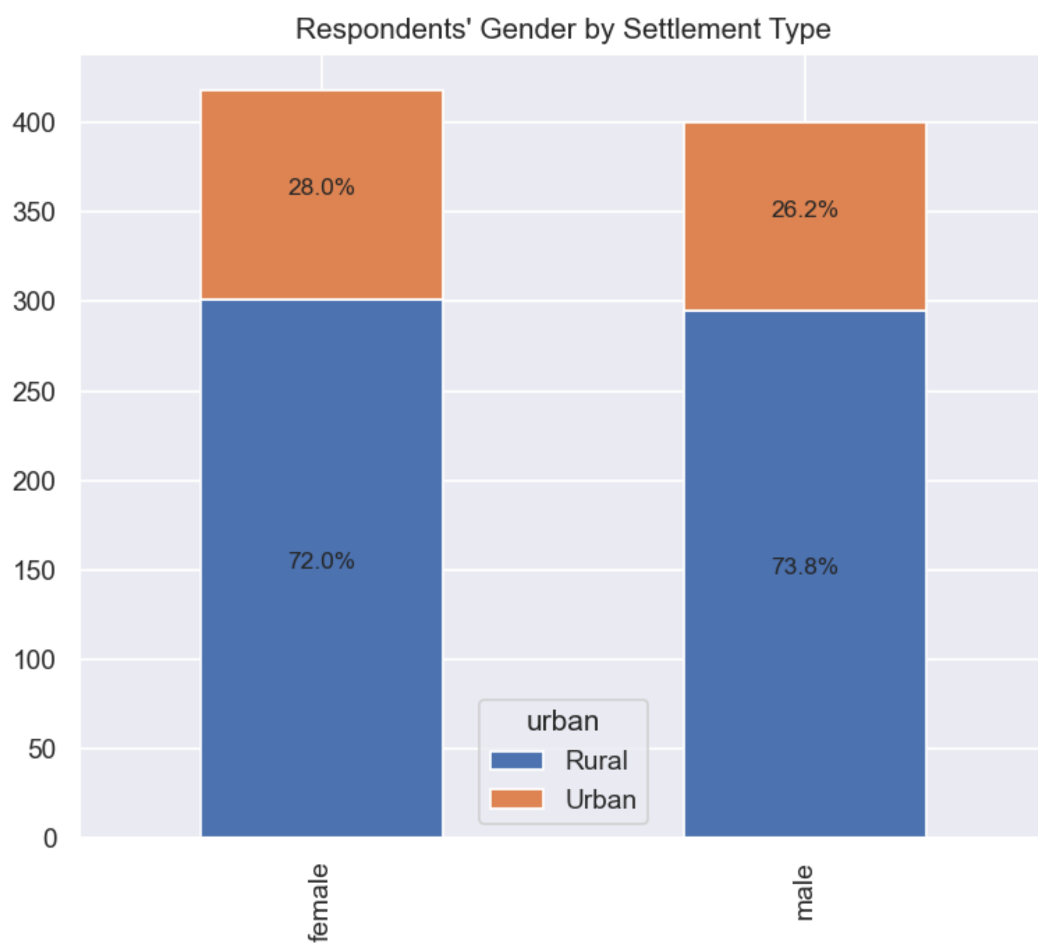


Figure 19: Respondent's gender and settlement types

This corresponds with existing evidence although data collection was done in only three out of the 47 districts (counties) in Kenya. Several sources confirm that more than two-thirds (between 70 and 73 percent) of the Kenyan population is rural (*Kenya Rural Population 1960-2023 | MacroTrends*, n.d.; *Kenya Rural Population, Percent - Data, Chart*, n.d.; *Kenya - Rural Population (% of Total Population)*, n.d.; *Kenya - Rural Population - 2023 Data 2024 Forecast 1960-2022 Historical*, n.d.). Thus, the sample is geographically representative of the population.

Additionally, the evidence shows equality between genders on the usage of mobile money, with females even making up the (narrow) majority. Efforts should go into making this the reality as recent studies on financial inclusion report relatively lower rates for Kenyan women (Johnen & Mußhoff 2022; World Bank, n.d.).

4.2 Results on Research Questions

Research Question One: What is the level of cash preference among mobile money users in Kenya?

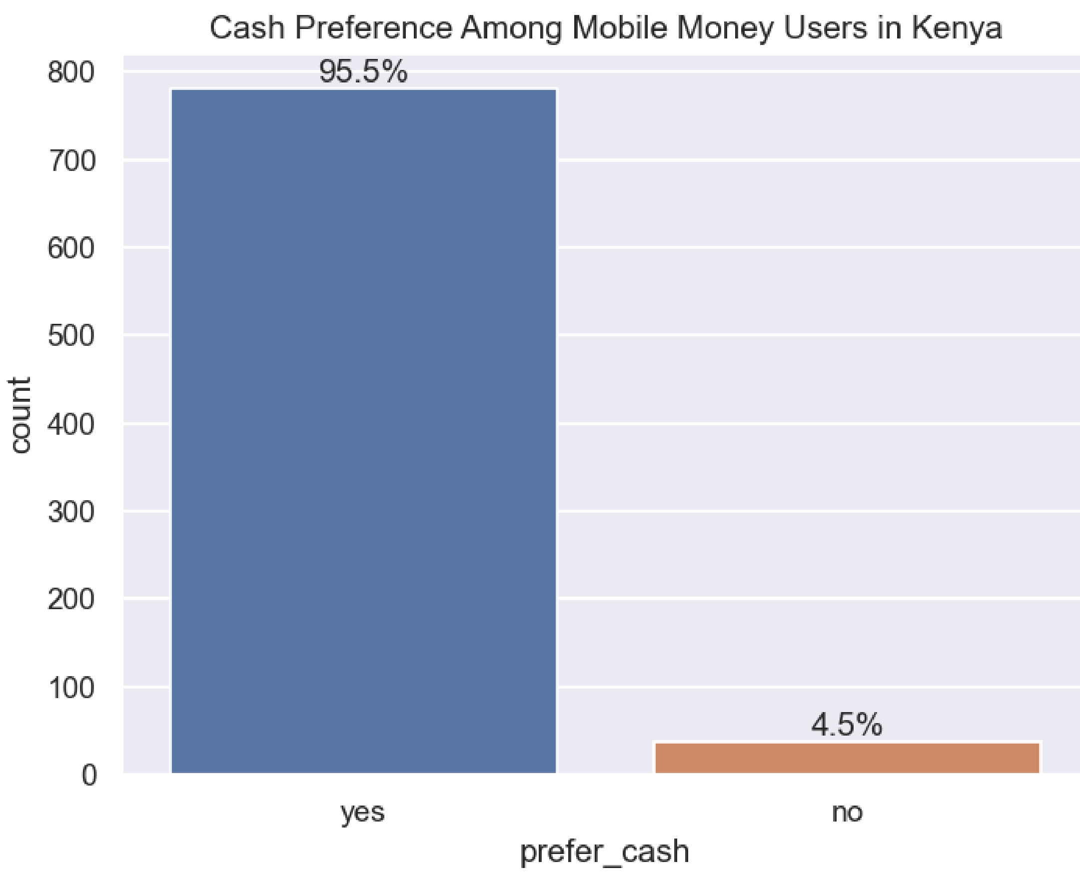


Figure 20: Cash preference among mobile money users in Kenya

Research Question Two: Are users' preferences for cash predictable based on their usage experiences with mobile money services?

```
[[  5   1]
 [ 54 104]]
```

Figure 21: The Confusion Matrix

Table 3: The classification report

	precision	recall	f1-score	support
0	0.08	0.83	0.15	6
1	0.99	0.66	0.79	158
accuracy			0.66	164
macro avg	0.54	0.75	0.47	164
weighted avg	0.96	0.66	0.77	164

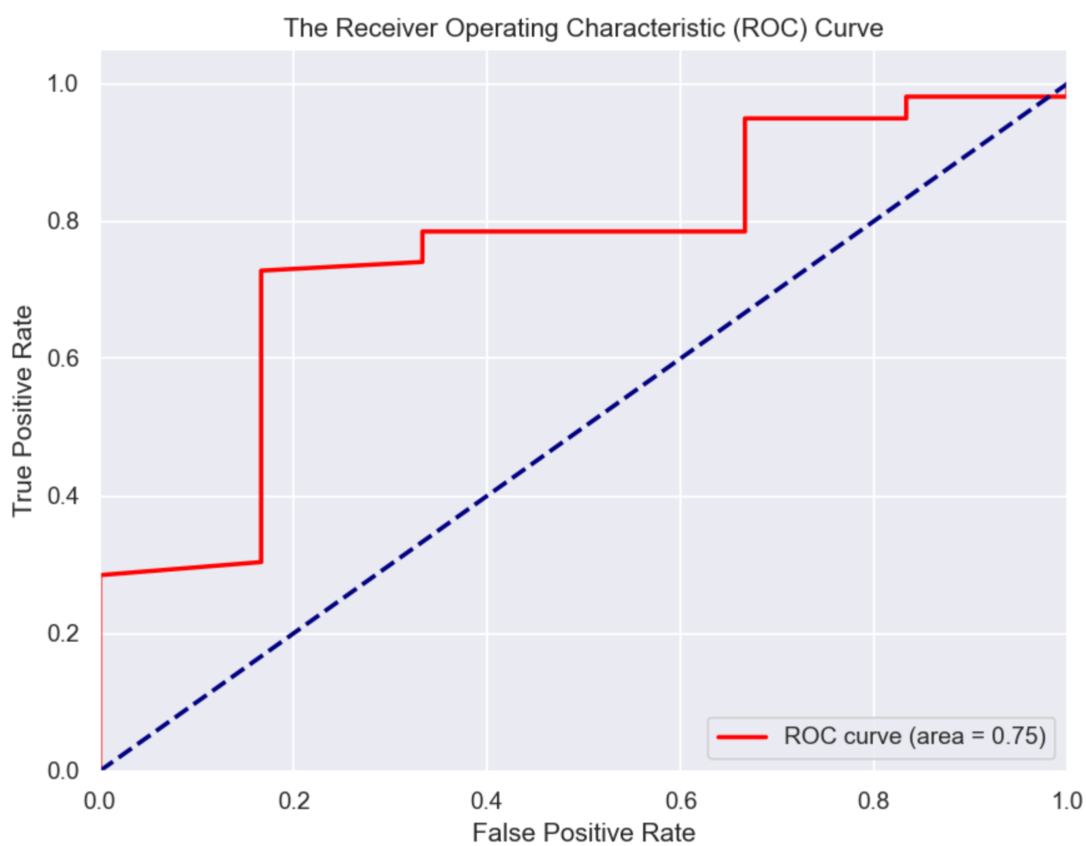


Figure 22: The ROC curve

4.3 Discussion

Out of 818 mobile money users, 95.5% said they preferred cash to mobile money. This rate of cash preference is similar, if not identical, to those in the claims by Nnamani (2022), as well as Oxford Business Group (2018), that cash preference is high in Kenya despite widespread uptake of mobile money services.

The confusion matrix obtained from the model's predictions on the test set had the following values:

- True Positives (TP): 104 users were correctly classified as preferring cash to mobile money.
- False Positives (FP): One user was incorrectly classified as preferring cash when they did not.
- False Negatives (FN): 54 users were incorrectly classified as not preferring cash when they did.
- True Negatives (TN): Five users were correctly classified as not preferring cash.

Review of Evaluation Metrics from the Classification Report:

Macro average, weighted average, and support

Macro average is a simple averaging of the performance metrics (precision, recall, F1-score) calculated for each class in a classification problem. It measures the overall performance of the model across all classes without respect to class distribution, treating each class as of equal weight in the average. As such it is observed that the macro averages for all the performance metrics do not reflect the disproportionate distribution of classes in the target.

Macro average is computed as the arithmetic mean of the individual performance metric across classes.

Weighted average conversely considers class distribution as it assigns a weight for each class based on the number of instances in that class. It is computed by multiplying each class's metric by the proportion of instances belonging to that class in the dataset, then summing up the weighted metrics across classes. The weighted average is useful in projects with imbalanced data such as this because it gives more importance to classes with more instances, while ensuring that the performance on the minority classes contributes proportionally to the overall score. As is evident from Table 7, the weighted averages across metrics reflect the class distribution in the target variable.

The support value indicates the number of samples in the test set. In this case, the support is 164, which corresponds to the total number of observations in the test set (158 positive class and 6 negative class).

Accuracy

The model achieved an accuracy score of 0.66. This implies that the model was effective in classifying users with a preference for cash 66% of the time. However, accuracy alone is the most informative metric, especially since the dataset exhibits significant class imbalance with a much larger number of positive class instances (preference for cash) compared to the negative class instances (no preference for cash). As such, a high accuracy score may be partly attributed to correctly classifying the majority class only. Therefore, it's crucial to consider additional metrics to evaluate model performance more comprehensively.

F1 score

The F1 score, which combines precision and recall, is a balanced measure of performance. The implication with an F1 score of 0.75 is that the model performs reasonably well in suppressing the occurrence of false positives (precision) and false negatives (recall). Thus, the model moderately captures users' preferences for cash while similarly minimizing misclassifications. But how does the model discriminate between cash preference and otherwise at different classification thresholds? The AUC score is an effective metric for measuring that.

Discriminatory power

The AUC score quantifies the model's ability to tell classes apart. While an AUC score of 1.0 represents perfect discrimination, the AUC score of 0.75 suggests that the model exhibits reasonable discriminative power; thus, the achieved score indicates that the model has the capacity to distinguish preferences for cash from that for mobile money effectively. But the ROC curve shows a tradeoff between TPR (sensitivity) and FPR (1-specificity). So, a moderate score could mean that while the model performs well in identifying users who prefer cash, it may have limitations in distinguishing users who do not.

4.4 Practical Implications

The findings herein significantly buttress the current narrative that cash preference for transactions is pervasive despite widespread uptake of mobile phone-based financial services in Kenya. It provides an evidential basis to empirically assert the previously speculated notions that most mobile money users prefer cash to mobile money; and that this preference is associated with their negative experiences using the service. These outcomes have significant imports for service providers and policymakers in the country.

Businesses face the risk of customer churn if these negative experiences are not addressed. The possible ripple effects of this include a loss of competitive advantage leading to losses in market share, higher customer support costs, and diminished brand equity due to broadly negative customer sentiments. To the Kenyan economy, mobile money has been grounds for increased investments (Barasa, 2021); a means of poverty reduction (Matheson, 2016; Suri & Jack, 2016); and an overall source of economic growth (Bill & Melinda Gates Foundation, 2021). Therefore, prolonged regulatory inaction to address the issues that accentuate user experience points to potential declines in these indicators. Understanding user preferences can inform the development of tailored digital financial services and targeted marketing strategies, ultimately promoting financial inclusion and economic growth. The next chapter presents specific recommendations that industry actors can consider to improve the narrative for the better.

4.5 Chapter Summary

I this chapter, I exhaustively dissected the results of the analysis. It showed that cash preference among mobile money users in Kenya is extremely high. It also showed that based on user experiences, the logistic regression classifier effectively classified users as preferring cash, although there were limitations when it came to classifying users who do not.

CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

The widespread preference for cash to mobile money in Kenya, as established by existing research, is concerning for mobile money service providers, development actors working towards financial inclusion, and the national economy alike. However, it was not categorically clear if the reported level of cash preference in broader Kenya was the same for mobile money users in the country.

Considering that, this study sought to address two objectives: Find out what the level of cash preference is among mobile money users in Kenya; and ascertain if their preferences are predictable based on their experiences with the mobile money service using logistic regression. The evidence emphatically establishes that a vast majority of mobile money users in Kenya prefer cash for transactions. The logistic regression classification model also showed strong performance in classifying users with a preference for cash.

5.1 Conclusions

Based on the results in this study, I conclude that cash preference is invariably high among mobile money users in Kenya; and that the sentiment is associated with negative user experiences. While the results are promising, it is essential to acknowledge potential limitations. The dataset size, particularly the small number of observations for the negative class, i.e., users who do not prefer cash, highlights the potential issue of how limited representation in the dataset could impact the model's generalizability. That notwithstanding, this study is an empirical reference point for ascertaining the level of cash preference among mobile money users in Kenya. It helps a great deal to know what issues are faced by mobile money customers and how these issues may influence negative sentiments towards the service that may result in eventual churn. Additionally, it serves as a call for industrial adjustments and further research.

5.2 Recommendations

Service providers in Kenya need to reengineer system infrastructure to mitigate, if not eliminate, loopholes that expose users to fraudulent activities. This will boost customer confidence in the security of mobile banking services and encourage more non-users to patronize the service. This effort should be accompanied by an improvement and expansion of mobile network coverage to make mobile money a more competitive option for transactions. Regulators must prioritize a business environment that enforces liability on service providers to be more transparent to users about the terms of usage of mobile money services. This may involve breaking down transactional costs for users to fully understand the costs involved, providing specific details on the type of user data collected by service providers, and improving the efficacy of complain channels available to the users in case of any issues. These efforts towards service improvement have the potential to deliver results like more thorough uptake, which may reflect in, for example, direct payments being a major use case for mobile money in Kenya.

5.3 Suggestions for Further Research

This study, while insightful, leaves room for further inquiry based on a few reasons: It is not clear what the sentiments are on the side of merchants or businesspeople who have experienced receiving payments through mobile money. Perhaps the high preference for cash among respondents is associated with other variables not included in this dataset. While the dataset appears geographically representative of the Kenyan population, 818 users are not quite numerically representative of the over 30 million active mobile money users in the country (Taylor, 2023). It is also necessary for industrial adoption that other classification models be tested for optimal generalization.

In view of these, data collection in future research efforts should prioritize increasing the sample and doing so over a much wider geographical area to make findings more representative and generalizable. Studies should also include the user experiences and sentiments of merchants, so that resulting insights give a more vivid and thorough illustration of issues that come with usage of the service on either side of the transaction divide. Attention should also be given to obtaining a more balanced cash preference variable to avoid encountering analytical challenges that result from class imbalance.

Additionally, collecting data on other variables like usage patterns and use cases, and engineering new features from existing one to obtain more peculiar information for analysis could provide valuable insights for industry actors. Further research must experiment with other classification algorithms and employ more advanced modeling techniques like cross validation and hyperparameter tuning, and to enhance predictive accuracy and model generalizability for potential industrial application.

In summary, the following are factors that should go into further research efforts:

- Collect additional data, i.e., include more users who do not prefer cash, to balance the class distribution and increase the number of negative class instances in the target variable.
- Data collection must be done over a much wider geographical area to make the resulting evidence more representative of the population.
- Observations should proportionately include merchant users to give a more actual illustration of issues associated with the use of mobile money.
- More variables should be added to enrich the data, build a more robust model, and improve predictive performance.
- Incorporate feature engineering to include more relevant variables that could influence users' preferences.
- Explore different classification algorithms to compare and potentially improve model performance and generalization.

5.4 Limitations of the Study

- It is important to establish that while insights about the mobile money market in Kenya that arose in the conduction of this study may advise both state and private actors in other countries, the findings herein are not generalizable to other populations or countries. This study makes inferences, conclusions about Kenya’s mobile money market only.
- The study may have faced selection problems due to factors such as education level, which is shown to be low, and other factors not categorically included in

the data such as wealth, and changes in technology preference. These factors may have influenced both the adoption of mobile money and the preference of cash for transactions.

- I omitted the sampling weights from the model since it was unclear which sampling procedure was used to collect the data, and which methodology was used in estimating the weights.

REFERENCES

Ajayi, K. F., & Ross, P. H. (2020). The Effects of Education on Financial Outcomes: Evidence from Kenya. *Economic Development and Cultural Change*, *69*(1), 253–289. https://doi.org/10.1086/702996

Ajene, E. (2022, December 4). Afridigest Fintech Review: Tingo's acquisition is complete [Substack newsletter]. *Afridigest*. https://afridigest.substack.com/p/afr-week-48-2022

Alhassan, A., Li, L., Reddy, K., & Duppati, G. (2019). The impact of formal financial inclusion on informal financial intermediation and cash preference: Evidence from Africa. *Applied Economics*, *51*(42), 4597–4614. https://doi.org/10.1080/00036846.2019.1593316

Baganzi, R., & Lau, A. K. W. (2017). Examining Trust and Risk in Mobile Money Acceptance in Uganda. *Sustainability*, *9*(12), Article 12. https://doi.org/10.3390/su9122233

Barasa, L. N. (2021, July 5). Mobile money and the business economy in Kenya. *Globaldev Blog*. https://globaldev.blog/mobile-money-and-business-economy-kenya/

Batiz-Lazo, B., Maixé-Altés, J. C., & Peon, D. (2023). Behavioural Drivers of Cashless Payments in Africa. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4521894

Bhalla, D. (n.d.). *A Complete Guide to Area Under Curve (AUC)*. Retrieved 7 October 2023, from https://www.listendata.com/2014/08/learn-area-under-curve-auc.html

Bill & Melinda Gates Foundation. (2021, April). *ImpactofMobileMoneyonPoverty_ResearchBrief.pdf*. https://docs.gatesfoundation.org/Documents/ImpactofMobileMoneyonPoverty_ResearchBrief.pdf

Biswal, A. (2023, August 9). *What is a Chi-Square Test? Formula, Examples & Uses | Simplilearn*. Simplilearn.Com. https://www.simplilearn.com/tutorials/statistics-tutorial/chi-square-test

Boateng, R., & Sarpong, M. (2019). A Literature Review of Mobile Payments in Sub-Saharan Africa. In Y. Dwivedi, E. Ayaburi, R. Boateng, & J. Effah (Eds.), *International Working Conference on Transfer and Diffusion of IT (TDIT): Vol. AICT-558* (pp. 128–146). Springer International Publishing. https://doi.org/10.1007/978-3-030-20671-0_9

Breckenridge, K. (2010). The World's First Biometric Money: Ghana's E-Zwich and the Contemporary Influence of South African Biometrics. *Africa*, *80*(4), 642–662. https://doi.org/10.3366/afr.2010.0406

Central Bank of Kenya. (2016). *736331048_FinAccess Household 2016 Key Results Report.pdf*.

https://www.centralbank.go.ke/uploads/financial_inclusion/736331048_FinAccess%20%20Household%202016%20Key%20Results%20Report.pdf

Central Bank of Nigeria. (2017). *Cbn economic report fourth quarter 2017 published.pdf*. https://www.cbn.gov.ng/out/2018/rsd/cbn%20economic%20report%20fourth%20quarter%20%20%20%202017%20published.pdf

Chuttur, M. (2009). *Overview of the Technology Acceptance Model: Origins, Developments and Future Directions*.

Classification in Machine Learning: A Guide for Beginners. (n.d.). Retrieved 12 November 2024, from https://www.datacamp.com/blog/classification-machine-learning

Collins, D., Zollmann, J., & Fleming, P. (2012, November 14). *Is M-PESA Replacing Cash in Kenya? | Blog | CGAP*. https://www.cgap.org/blog/is-m-pesa-replacing-cash-in-kenya

Creemers, T., Murugavel, T., Boutet, F., Omary, O., & Oikawa, T. (2020, August 12). *Five Strategies for Mobile-Payment Banking in Africa*. BCG Global. https://www.bcg.com/publications/2020/five-strategies-for-mobile-payment-banking-in-africa

Davidson, N., & McCarty, M. Y. (n.d.). *Driving Customer Usage of Mobile Money for the Unbanked*.

Davis, F. D. (1985, December 20). *A TECHNOLOGY ACCEPTANCE MODEL FOR EMPIRICALLY TESTING NEW END-USER INFORMATION SYSTEMS: THEORY AND RESULTS*. MIT.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioural impacts. *International Journal of Man-Machine Studies*, *38*(3), 475–487. https://doi.org/10.1006/imms.1993.1022

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: Three experiments. *International Journal of Human-Computer Studies*, *45*(1), 19–45. https://doi.org/10.1006/ijhc.1996.0040

Dzokoto, V. A., Appiah, E. N., Chitwood, L. P., & Imasiku, M. L. (2016). The Changing Face of Money: Preferences for Different Payment Forms in Ghana and Zambia. *Journal of Applied Business and Economics*, *18*(4), 66–81.

Elnaiem, T. (2019). *The Role of Trust and Gender in Mobile Money Adoption in Lusaka, Zambia: An Analysis Using the Technology Acceptance Model* [University of Guelph]. http://hdl.handle.net/10214/17441

Flood, Z. (2018, February 22). Zimbabwe and Kenya lead the way in Africa's dash from cash. *The Guardian*. https://www.theguardian.com/world/2018/feb/22/kenya-leads-way-mobile-money-africa-shifts-towards-cash-free-living

Gbongli, K., Xu, Y., & Amedjonekou, K. M. (2019). Extended Technology Acceptance Model to Predict Mobile-Based Money Acceptance and Sustainability: A Multi-Analytical Structural Equation Modeling and Neural Network Approach. *Sustainability*, *11*(13), Article 13. https://doi.org/10.3390/su11133639

GSM Association. (2017). 2017 State of the Industry Report on Mobile Money. *Mobile for Development*. https://www.gsma.com/mobilefordevelopment/resources/2017-state-industry-report-mobile-money/

Gupta, Y. (2022, March 24). Logistic Regression in 'Simple English'. *Data Science Simplified*. https://medium.com/dssimplified/logistic-regression-in-simple-english-a4edda1eb05f

Harford, T. (2017, February 13). Money via mobile: The M-Pesa revolution. *BBC News*. https://www.bbc.com/news/business-38667475

Hayduk, C. (2022, January 22). *The Role of Experimentation in Fundamental Machine Learning Research | LinkedIn*. https://www.linkedin.com/pulse/role-experimentation-fundamental-machine-learning-research-hayduk/

Ifeanyi-Ajufo, N. (2022). *Digital Financial Inclusion and Security: The Regulation of Mobile Money in Ghana*. Carnegie Endowment for International Peace. https://carnegieendowment.org/2022/09/19/digital-financial-inclusion-and-security-regulation-of-mobile-money-in-ghana-pub-87949

Johnen, C., & Mußhoff, O. (n.d.). *Digital credit and the gender gap in financial inclusion: Empirical evidence from Kenya—Johnen—2023—Journal of International Development—Wiley Online Library*. Retrieved 7 October 2023, from https://onlinelibrary.wiley.com/doi/full/10.1002/jid.3687

Kamiri, J., & Mariga, G. (2021). *Research Methods in Machine Learning: A Content Analysis*. *10*(2).

Kelly, A. E., & Palaniappan, S. (2023). Using a technology acceptance model to determine factors influencing continued usage of mobile money service transactions in Ghana. *Journal of Innovation and Entrepreneurship*, *12*(1), 34. https://doi.org/10.1186/s13731-023-00301-3

Kenya Rural Population 1960-2023 | MacroTrends. (n.d.). Retrieved 7 October 2023, from https://www.macrotrends.net/countries/KEN/kenya/rural-population

Kenya Rural population, percent—Data, chart. (n.d.). TheGlobalEconomy.Com. Retrieved 7 October 2023, from https://www.theglobaleconomy.com/Kenya/rural_population_percent/

Kenya—Rural population (% of total population). (n.d.). Retrieved 7 October 2023, from https://www.indexmundi.com/facts/kenya/indicator/SP.RUR.TOTL.ZS

Kenya—Rural Population—2023 Data 2024 Forecast 1960-2022 Historical. (n.d.). Retrieved 7 October 2023, from https://tradingeconomics.com/kenya/rural-population-percent-of-total-population-wb-data.html

Khan, M. R., & Blumenstock, J. E. (2016). Predictors without Borders: Behavioural Modeling of Product Adoption in Three Developing Countries. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 145–154. https://doi.org/10.1145/2939672.2939710

Kingiri, A. N., & Fu, X. (2020). Understanding the diffusion and adoption of digital finance innovation in emerging economies: M-Pesa money mobile transfer service in Kenya. *Innovation and Development*, *10*(1), 67–87. https://doi.org/10.1080/2157930X.2019.1570695

Kraan, R. (2020, May 5). *Demystifying decision trees, random forests & gradient boosting*. Medium. https://towardsdatascience.com/demystifying-decision-trees-random-forests-gradient-boosting-20415b0a406f

Langley, P. (1988). Machine Learning as an Experimental Science. *Machine Learning*, *3*(1), 5–8. https://doi.org/10.1023/A:1022623814640

Lubua, E., & Semlambo, A. (2017). The influence of the ease of use and perceived usefulness to the adoption of mobile money services in SMEs in Tanzania. *The Information Technologist (1597-4316)*, *14*, 131–141.

Lule, I., Omwansa, T. K., & Waema, T. M. (2012). *Application of Technology Acceptance Model (TAM) in M-Banking Adoption in Kenya*. *6*(1).

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Marschner, I. C. (2015). Relative Risk Regression for Binary Outcomes: Methods and Recommendations. *Australian & New Zealand Journal of Statistics*, *57*(4), 437–462. https://doi.org/10.1111/anzs.12131

Mas, I., & Morawczynski, O. (2009). *Designing Mobile Money Services: Lessons from M-PESA* (SSRN Scholarly Paper 1552753). https://doi.org/10.2139/ssrn.1552753

Mas, I., & Ng'weno, A. (2012). *Why Doesn't Every Kenyan Business Have a Mobile Money Account?* (SSRN Scholarly Paper 2044550). https://papers.ssrn.com/abstract=2044550

Matheson, R. (2016, December 8). *Study: Mobile-money services lift Kenyans out of poverty*. MIT News | Massachusetts Institute of Technology. https://news.mit.edu/2016/mobile-money-kenyans-out-poverty-1208

McGath, T. (2018, September 21). *M-PESA: How Kenya revolutionized mobile payments*. Medium. https://mag.n26.com/m-pesa-how-kenya-revolutionized-mobile-payments-56786bc09ef

Mintz-Roth, M. (2018, March 7). *Women, men and youth and the adoption of mobile money in Kenya*. Financial Sector Deepening Kenya. https://www.fsdkenya.org/blogs-publications/blog/women-men-and-youth-and-the-adoption-of-mobile-money-in-kenya/

Mobile Money. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Mobile_Money&oldid=1148469712

Mpofu, F. Y. (2022). Industry 4.0 in Financial Services: Mobile Money Taxes, Revenue Mobilisation, Financial Inclusion, and the Realisation of Sustainable Development Goals (SDGs) in Africa. *Sustainability*, *14*(14), Article 14. https://doi.org/10.3390/su14148667

Mwangi, I. W. (2017). *Dynamics Of Financial Inclusion And Welfare In Kenya*.

Ng'weno, A., & Bill & Melinda Gates Foundation. (2010, November). *How Mobile Money has Changed Lives in Kenya*. https://docs.gatesfoundation.org/Documents/mobile-money.pdf

Nnamani, C. (2022, October 11). Mobile money in Africa surges as more markets open. *TechCabal*. https://techcabal.com/2022/10/11/mobile-money-is-surging-in-africa-as-more-disruption-occur/

Nteta, Z. (2017). *(In)formality in Africa: Exploring the social and cultural factors influencing mobile payment non-adoption amongst informal traders in Cape Town*.

Okenyuri, L., & Ouma, R. (2019). *Mobile Phone Banking and its Convenience in Making Transactions The Case of M-PESA Customers in Kisii Couny, Kenya*. *17*(2).

Orekoya, S. (2017). *MOBILE MONEY AND MONETARY POLICY IN NIGERIA.*

Oteri, O. M., Kibet, L. P., & N, N. E. (2015). Mobile Subscription, Penetration and Coverage Trends in Kenya's Telecommunication Sector. *International Journal of Advanced Research in Artificial Intelligence (Ijarai)*, *4*(1), Article 1. https://doi.org/10.14569/IJARAI.2015.040101

Oxford Business Group. (2018, October 31). *Digital transactions are gaining popularity, but cash may remain the preferred option for many Kenyans—Africa 2018—Oxford Business Group*. https://oxfordbusinessgroup.com/reports/kenya/2018-report/economy/mobile-money-although-digital-transactions-are-gaining-popularity-market-fragmentation-could-see-cash-remain-the-preferred-option

Parekh, N., & Hare, A. (2020, October 22). *The rise of mobile money in sub-Saharan Africa: Has this digital technology lived up to its promises?* The Abdul Latif Jameel Poverty Action Lab (J-PAL). https://www.povertyactionlab.org/blog/10-22-20/rise-mobile-money-sub-saharan-africa-has-digital-technology-lived-its-promises

Piper, K. (2020, September 11). *What Kenya can teach its neighbors—And the US — about improving the lives of the 'unbanked'*. Vox. https://www.vox.com/future-perfect/21420357/kenya-mobile-banking-unbanked-cellphone-money

Rolfe, A. (2019, January 25). *Mobile money transactions equivalent of half of Kenya's GDP*. Payments Cards & Mobile. https://www.paymentscardsandmobile.com/mobile-money-transactions-half-of-kenyas-gdp/

Russon, M.-A. (2019, February 19). *The battle between cash and mobile payments in Africa—BBC News*. https://www.bbc.com/news/business-47292757

Sehlolo, M. (2022, December 6). *Kenya reinstates mobile money fees*. https://www.connectingafrica.com/mobile-money/kenya-reinstates-mobile-money-fees

Shirono, K., Chhabra, E., Das, B., Fan, Y., & Villanova, H. C. (2021). Is Mobile Money Part of Money? Understanding the Trends and Measurement. *IMF Working Papers, 2021*(177). https://doi.org/10.5089/9781513587851.001.A001

Simione, F., & Muehlschlegel, T. (2023). *Mobile Money, Perception About Cash, and Financial Inclusion: Learning from Uganda's Micro-Level Data* (SSRN Scholarly Paper 4641124). https://doi.org/10.5089/9798400259944.001

Song, K. (2016, November 17). M-Pesa: Transforming Kenya with Mobile Money. *Technology and Operations Management*. https://d3.harvard.edu/platform-rctom/submission/m-pesa-transforming-kenya-with-mobile-money/

Soutter, L., Ferguson, K., & Neubert, M. (2019, July 12). *Digital Payments: Impact Factors and Mass Adoption in Sub-Saharan Africa by Leigh Soutter, Kenzie Ferguson, Michael Neubert: SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3418628

Suri, T. (2017). Mobile Money. *Annual Review of Economics*, *9*(1), 497–520. https://doi.org/10.1146/annurev-economics-063016-103638

Suri, T., & Jack, W. (2016). The long-run poverty and gender impacts of mobile money. *Science, 354*(6317), 1288–1292. https://doi.org/10.1126/science.aah5309

Sy, A., Maino, R., Massara, A., Perez-Siaz, H., & Sharma, P. (2019, February 14). *Fintech in Sub-Saharan Africa: A Potential Game Changer*. IMF. https://www.imf.org/en/Blogs/Articles/2019/02/14/fintech-in-sub-saharan-africa-a-potential-game-changer

Taylor, P. (2023, May 2). *Topic: Mobile money in Africa*. Statista. https://www.statista.com/topics/6770/mobile-money-in-africa/

Tobbin, P., & Kuwornu, J., K. M. (2011). Adoption of Mobile Money Transfer Technology: Structural Equation Modeling Approach. *European Journal of Business and Management*, *3*.

UNDESA. (n.d.). *Goal 8 | Department of Economic and Social Affairs*. Retrieved 8 October 2023, from https://sdgs.un.org/goals/goal8

UNSGSA. (2018, September). *Financial Inclusion: Technology, Innovation, Progress*. https://sdgs.un.org/sites/default/files/publications/2649unsgsa.pdf

Varghese, D. (2019, May 10). *Comparative study on Classic Machine learning Algorithms*. Medium. https://towardsdatascience.com/comparative-study-on-classic-machine-learning-algorithms-24f9ff6ab222

Wamuyu, P. K. (2014). The Role of Contextual Factors in the Uptake and Continuance of Mobile Money Usage in Kenya. *THE ELECTRONIC JOURNAL OF INFORMATION SYSTEMS IN DEVELOPING COUNTRIES*, *64*(1), 1–19. https://doi.org/10.1002/j.1681-4835.2014.tb00457.x

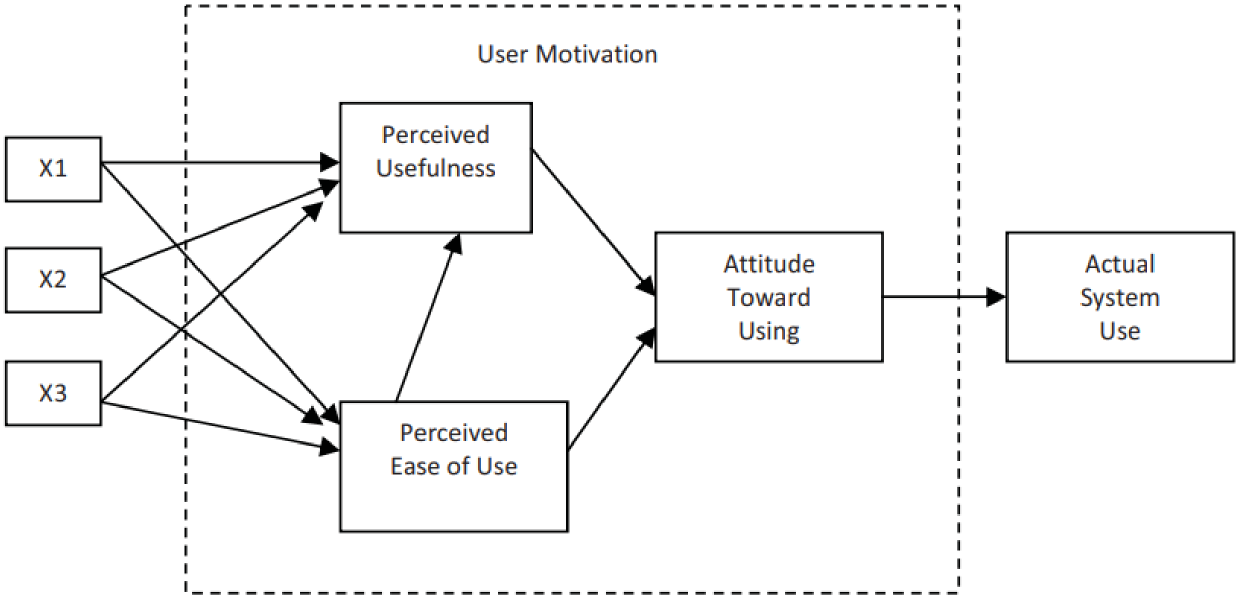
Worthington, A. K., & Burgess, G. L. (2021). *Technology Acceptance Model*. https://ua.pressbooks.pub/persuasiontheoryinaction/chapter/technology-acceptance-model/

APPENDICES

APPENDIX I

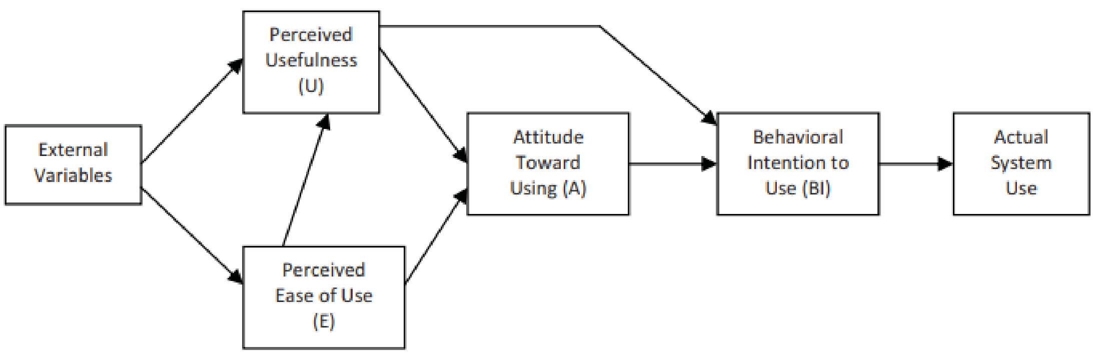
Theoretical Framework

Figure 1: Technology Acceptance Model



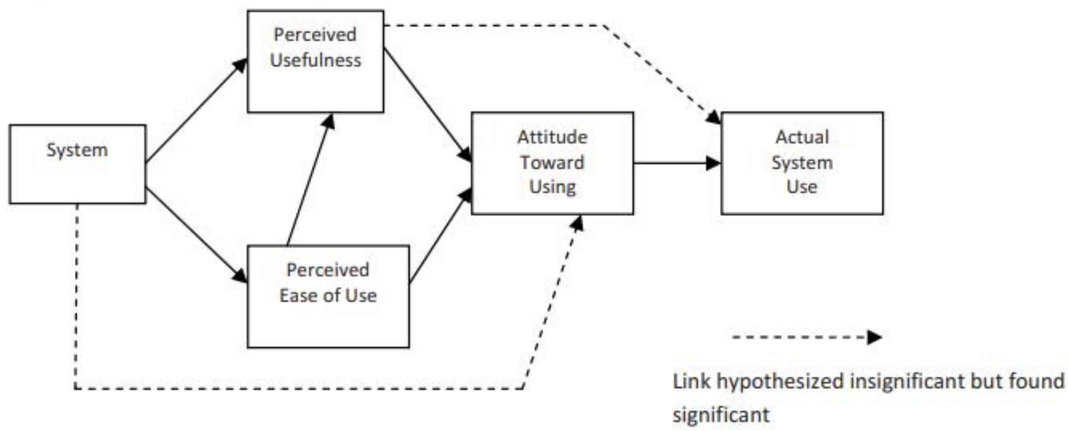
Source: Davis, (1985, p.24)

Figure 2: Modified TAM with Behavioural Intention Variable



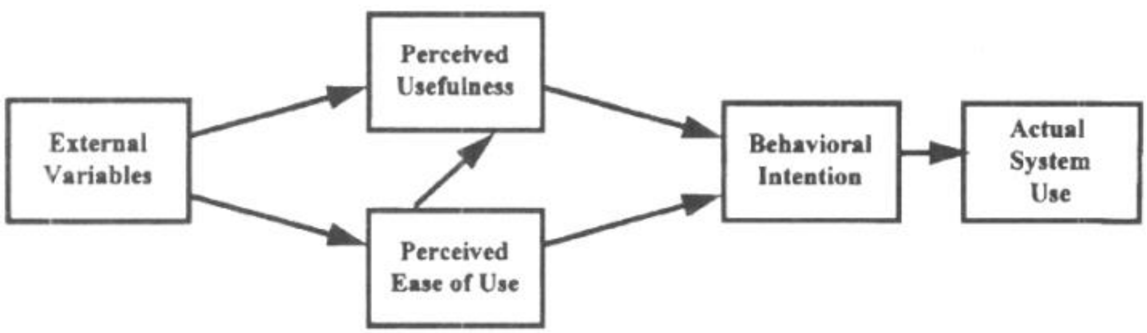
Source; Davis et al. (1989, p.985)

Figure 3: Hypothesis of new relationships



Source: Davis (1993, p.481)

Figure 4: Technology Acceptance Model



Source: Davis & Venkatesh (1996, p. 453)

APPENDIX II

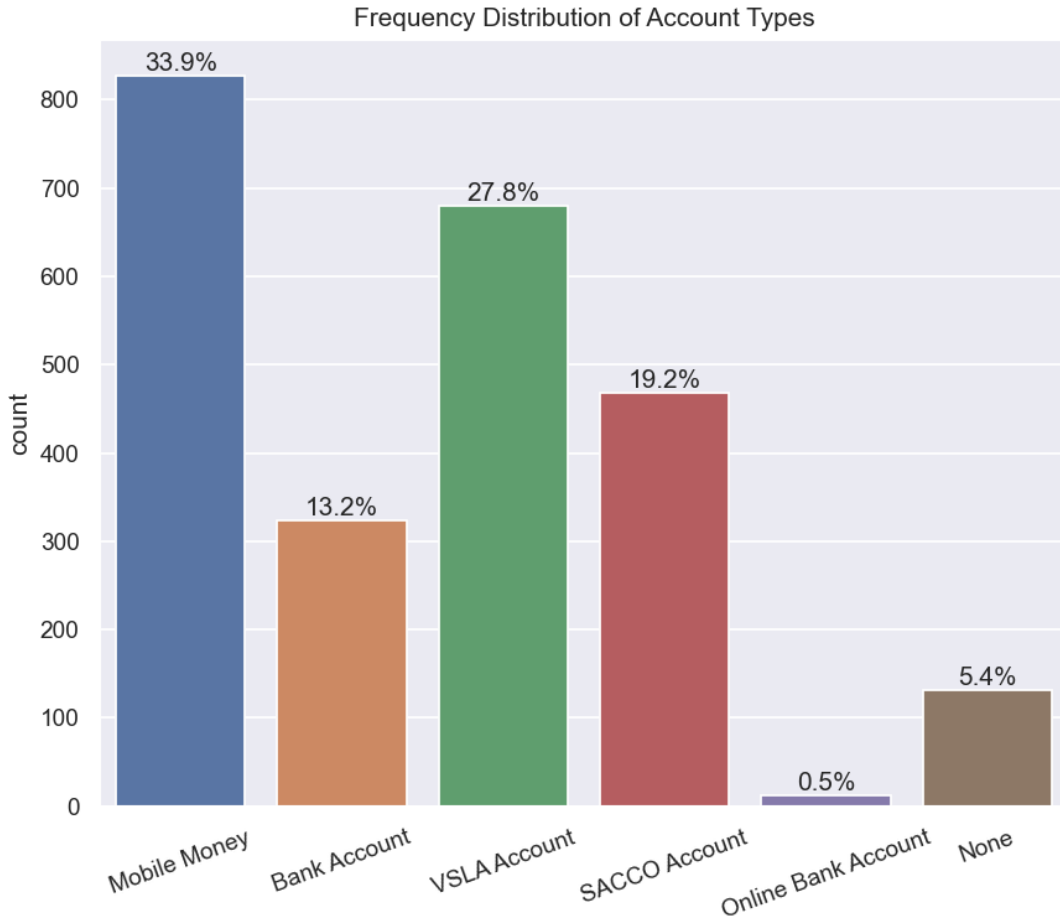
Table 1: Variable description

```
RangeIndex: 2442 entries, 0 to 2441
Data columns (total 29 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   start_time              2442 non-null   object 
 1   end_time                2442 non-null   object 
 2   hhid                    2442 non-null   int64  
 3   account_num             2442 non-null   int64  
 4   account_type            2442 non-null   object 
 5   weight                  2442 non-null   float64
 6   district                2442 non-null   object 
 7   urban                   2442 non-null   object 
 8   gender                  2442 non-null   object 
 9   age                     2442 non-null   int64  
 10  hh_members              2442 non-null   int64  
 11  highest_grade_completed 2235 non-null   object 
 12  mm_account_cancelled    2442 non-null   object 
 13  prefer_cash             2395 non-null   object 
 14  mm_trust                2314 non-null   object 
 15  mm_account_telco        1964 non-null   object 
 16  mm_account_telco_main   1011 non-null   object 
 17  understand_terms        1848 non-null   object 
 18  agent_trust             1634 non-null   object 
 19  taken_loan              643 non-null    object 
 20  network_issues          2146 non-null   object 
 21  clear_about_fees        2155 non-null   object 
 22  transaction_failed      2155 non-null   object 
 23  agent_low_cash          2106 non-null   object 
 24  t_and_c_copy            2101 non-null   object 
 25  complain_channel        2146 non-null   object 
 26  issue_resolved          575 non-null    object 
 27  understand_data         2146 non-null   object 
 28  fraud_victim            2155 non-null   object 
dtypes: float64(1), int64(4), object(24)
```

Source: Field survey (2019)

APPENDIX III

Figure 5: Types of financial accounts held by respondents



Source: Field survey (2019)

APPENDIX IV

Figure 6: Code for sub-setting mobile money user data

```
momo = (data[data["account_type"] == "Mobile Money"]).drop(["start_time", "end_time"], axis=1)
momo.head(3)
```

	hhid	account_num	account_type	weight	district	urban	gender	age	hh_members	highest_grade_completed
0	1001	1	Mobile Money	145.94444	District_A	Urban	male	32	1	primary 6
5	1003	1	Mobile Money	760.46191	District_A	Urban	male	30	8	secondary 6
7	1004	1	Mobile Money	433.96402	District_A	Rural	male	68	4	primary 6

Source: Python workflow

APPENDIX V

Figure 7: Code for statistical summaries

```
momo.describe(include = "all").style.background_gradient(cmap="summer")
```

Source: Python workflow

APPENDIX VI

Figure 8: Code for age distribution boxplot

```
# Users' age distribution

# Subsetting age variable
ages = momo['age']

# Creating the boxplot
sns.boxplot(ages)

# Labelling and showing the plot
plt.title("Boxplot of Respondents' Ages")
plt.xlabel("Ages")
plt.show()
```

Source: Python workflow

APPENDIX VII

Figure 9: Code for the annotated bar graph of respondents' educational backgrounds

```
# Bar graph of users' educational backgrounds

# Subsetting the grade completed variable
grades = momo['highest_grade_completed']

# Instantiating the countplot
ax = sns.countplot(data=momo, x=grades)

# Calculating percentage values
total_grades = len(grades)

for p in ax.patches:
    percentage_grades = '{:.1f}%'.format(100 * p.get_height() / total_grades)
    x = p.get_x() + p.get_width() / 2
    y = p.get_height()
    ax.annotate(percentage_grades, (x, y), ha='center', va='bottom')
```

Source: Python workflow

APPENDIX VIII

Figure 10: Code for the annotated stacked bar graph of respondents' gender by settlement type

```
# Stacked bar graph of users' gender by settlement type

# Creating the stacked bar plot
gen_set = pd.crosstab(momo['gender'], momo['urban'])
ax = gen_set.plot(kind="bar", stacked=True, title="Respondents' Gender by Settlement Type")

# Calculating percentage values
total_set = gen_set.sum(axis=1)

# Annotating the bars with percentage values
for i, category in enumerate(gen_set.index):
    bottom = 0
    for j, value in enumerate(gen_set.columns):
        percentage_set = (gen_set.loc[category, value] / total_set.loc[category]) * 100
        ax.annotate(f"{percentage_set:.1f}%", (i, bottom + gen_set.loc[category, value] / 2),
                    ha='center', fontsize='10')
        bottom += gen_set.loc[category, value]
```

Source: Python workflow

APPENDIX IX

Figure 11: Code for the bar graph showing cash preference among mobile money users

```
# Visualizing cash preference

pref = momo['prefer_cash']

ax = sns.countplot(data=momo, x=pref)

# Calculating percentage values
total_pref = len(pref)

for p in ax.patches:
    percentage_pref = '{:.1f}%'.format(100 * p.get_height() / total_pref)
    x = p.get_x() + p.get_width() / 2
    y = p.get_height()
    ax.annotate(percentage_pref, (x, y), ha='center', va='bottom')
```

Source: Python workflow

APPENDIX X

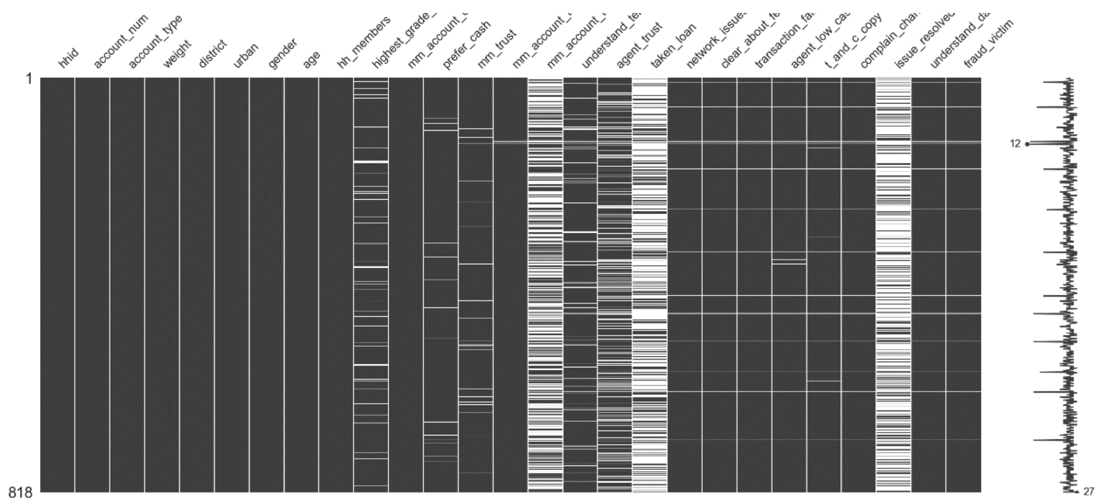
Table 9: A section of the descriptive statistics table

understand_terms	agent_trust	taken_loan	network_issues	clear_about_fees
776	642	262	814	814
3	3	2	2	2
yes	no	no	yes	yes

Source: Python workflow

APPENDIX XI

Figure 12: A matrix of the mobile money DataFrame visualizing missingness



Source: Python workflow

APPENDIX XII

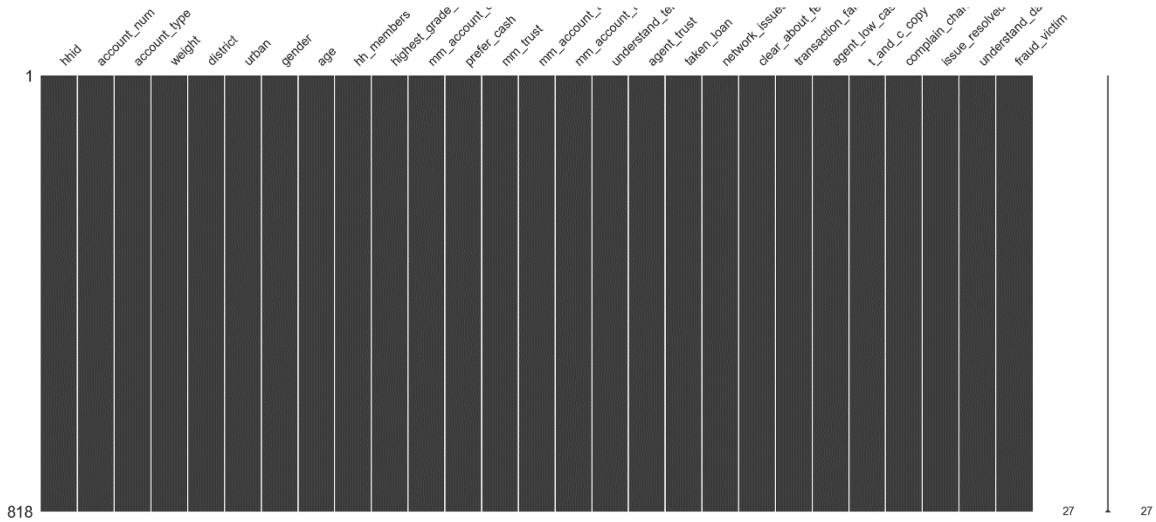
Figure 13: Code for mode imputation using simple imputer

```python
# Importing the imputer package
from sklearn.impute import SimpleImputer

# Defining the imputation strategy for mode imputation
imputer = SimpleImputer(strategy='most_frequent')

# Fitting the imputer to the data
momo_imputed = imputer.fit_transform(momo)
```

Source: Python workflow

APPENDIX XIII

Figure 14: A matrix of the imputed mobile money DataFrame



Source: Python workflow

APPENDIX XIV

Figure 15: Code for chi-squared statistic and p-value for each variable

```
from scipy.stats import chi2_contingency

# Create an empty DataFrame to store the chi-squared test results
results = pd.DataFrame(columns=['Feature', 'Chi2', 'P-Value'])

# Calculate the chi-squared statistic and p-value for each binary feature
for feature_name in user_exp.columns:
    contingency_table = pd.crosstab(user_exp[feature_name], user_exp['prefer_cash_yes'])
    chi2, p, _, _ = chi2_contingency(contingency_table)
    results.loc[feature_name] = [feature_name, chi2, p]

# Sort the results by p-value (ascending order)
sorted_results = results.sort_values(by='P-Value')

# Print or analyze the sorted_results DataFrame to identify significant features
print(sorted_results)
```

Source: Python workflow

APPENDIX XV

Table 3: Chi-squared test results

	Feature	Chi2	P-Value
prefer_cash_yes	prefer_cash_yes	795.008385	6.566476e-175
agent_low_cash_yes	agent_low_cash_yes	6.781228	9.212151e-03
transaction_failed_yes	transaction_failed_yes	5.150923	2.323403e-02
issue_resolved_yes	issue_resolved_yes	4.605238	3.187442e-02
understand_data_yes	understand_data_yes	4.348555	3.704037e-02
t_and_c_copy_yes	t_and_c_copy_yes	3.172703	7.487843e-02
network_issues_yes	network_issues_yes	2.974093	8.460751e-02
fraud_victim_yes	fraud_victim_yes	2.049683	1.522379e-01
mm_trust_yes	mm_trust_yes	1.943487	1.632904e-01
complain_channel_yes	complain_channel_yes	1.114158	2.911798e-01
understand_terms_yes	understand_terms_yes	1.038316	3.082134e-01
agent_trust_yes	agent_trust_yes	0.603890	4.370976e-01
taken_loan_yes	taken_loan_yes	0.117510	7.317507e-01
clear_about_fees_yes	clear_about_fees_yes	0.000000	1.000000e+00

Source: Python workflow

APPENDIX XVI

Table 4: Machine learning variables

Target variable	*Description*
prefer_cash	Indicating whether a user prefers cash to mobile money.
Features	
mm_trust	Indicating whether a user trusts mobile money.
understand_terms	Indicating whether a user understood the terms and conditi ons when they registered for a mobile money account.
agent_trust	Does a user trust mobile money agents?
network_issues	Has the user ever had an issue with the network being unavailable for mobile money transactions?
transaction_failed	Has a transaction ever failed to go through?
agent_low_cash	Has an agent they have dealt with ever not had enough cash or electronic cash available?
t_and_c_copy	Does the user have a copy of the mobile money terms and conditions?
complain_channel	Do you understand how and where to complain if you have an issue with mobile money?
issue_resolved	Has the user had an issue successfully resolved after making a complaint?
understand_data	Does the user understand what data mobile money providers collect about them?
fraud_victim	Has the user been a victim of fraud?

Source: Dataset documentation

APPENDIX XVII

Table 5: Dummy variables

	prefer_cash_yes	mm_trust_yes	understand_terms_yes	agent_trust_yes	taken_loan_yes	network_issues_yes
0	1	0	1	0	0	1
1	1	0	1	0	0	0
2	1	0	0	0	0	0
3	1	0	1	0	0	1
4	1	0	1	0	0	1

Source: Python workflow

APPENDIX XVIII

Table 6: Evaluation metrics for all trialled models (N = Negative Class, P = Positive class)

Model	*Accuracy*	*Precision*	*Recall*	*F1*	*ROC-AUC*
Logistic Regression	66%	N – 0.08 P – 0.99	N – 0.83 P – 0.66	N – 0.15 P – 0.79	0.75
Decision Tree	76%	N – 0.05 P – 0.97	N – 0.33 P – 0.78	N – 0.09 P – 0.86	0.54
Random Forest	77%	N – 0.03 P – 0.96	N – 0.17 P – 0.78	N – 0.05 P – 0.86	0.53
Support Vector Machines	76%	N – 0.03 P – 0.96	N – 0.17 P – 0.78	N – 0.05 P – 0.86	0.69
Gradient	76%	N – 0.05	N – 0.33	N – 0.09	0.57

Boost	P – 0.97	P – 0.78	P – 0.86

Source: Data analysis Jupyter notebook

APPENDIX XIX

Figure 16: Code for module importation

```
# imbalanced learn for oversampling
from imblearn.over_sampling import RandomOverSampler

# train_test_split is used to split the data into training and test sets
from sklearn.model_selection import train_test_split

# Importing the logistic regression classifier
from sklearn.linear_model import LogisticRegression

# importing the model evaluation metrics
from sklearn.metrics import (classification_report, confusion_matrix,
                             accuracy_score, roc_curve, auc)
```

Source: Python workflow

APPENDIX XX

Figure 17: Code for model training

```
# Creating training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state=42)

# Instantiating the Oversampler
ros = RandomOverSampler(random_state=42)

#Fitting the resampled values to the data
X_train_resampled, y_train_resampled = ros.fit_resample(X_train, y_train)

# Instantiating the classifier
logreg = LogisticRegression()

# Model fitting
logreg.fit(X_train_resampled, y_train_resampled)

# Predicting the labels of the test set
y_pred = logreg.predict(X_test)
```

Source: Python workflow

APPENDIX XXI

Figure 18: Code for plotting the ROC curve and computing the AUC score

```
# Computing predicted probabilities: y_pred_prob
y_pred_prob = logreg.predict_proba(X_test)[:,1]

# Generating ROC curve values: fpr, tpr, thresholds
fpr, tpr, thresholds = roc_curve(y_test, y_pred_prob)

# Calculating the area under the curve
roc_auc = auc(fpr, tpr)

plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, color='red', lw=2, label='ROC curve (area = %0.2f)' % roc_auc)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('The Receiver Operating Characteristic (ROC) Curve')
plt.legend(loc='lower right')
plt.show()
```

Source: Python workflow

APPENDIX XXII

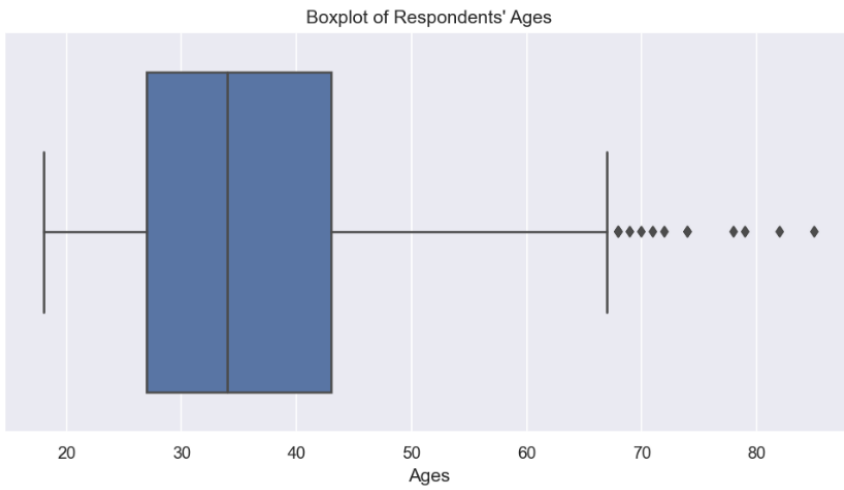
Table 7: Descriptive table of key demographic variables

	hhid	account_num	account_type	weight	district	urban	gender	age	hh_members
count	827.000000	827.000000	827	827.000000	827	827	827	827.000000	827.000000
unique	nan	nan	1	nan	3	2	2	nan	nan
top	nan	nan	Mobile Money	nan	District_A	Rural	female	nan	nan
freq	nan	nan	827	nan	318	601	425	nan	nan
mean	1609.573156	1.000000	nan	355.787488	nan	nan	nan	36.383313	4.719468
std	349.088024	0.000000	nan	279.530173	nan	nan	nan	12.257107	2.050488
min	1001.000000	1.000000	nan	14.582491	nan	nan	nan	18.000000	1.000000
25%	1309.500000	1.000000	nan	188.686040	nan	nan	nan	27.000000	3.000000
50%	1617.000000	1.000000	nan	283.233020	nan	nan	nan	34.000000	5.000000
75%	1912.500000	1.000000	nan	443.208890	nan	nan	nan	43.000000	6.000000
max	2205.000000	1.000000	nan	3157.224700	nan	nan	nan	85.000000	18.000000

Source: Field survey (2019)

APPENDIX XXIII

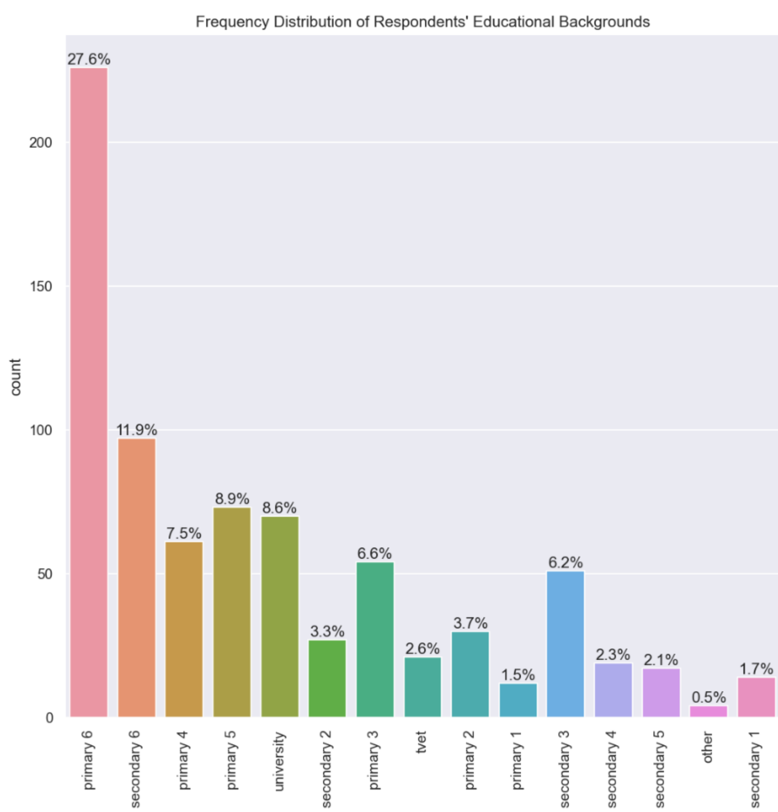
Figure 19: Age distribution



Source: Field survey (2019)

APPENDIX XXIV

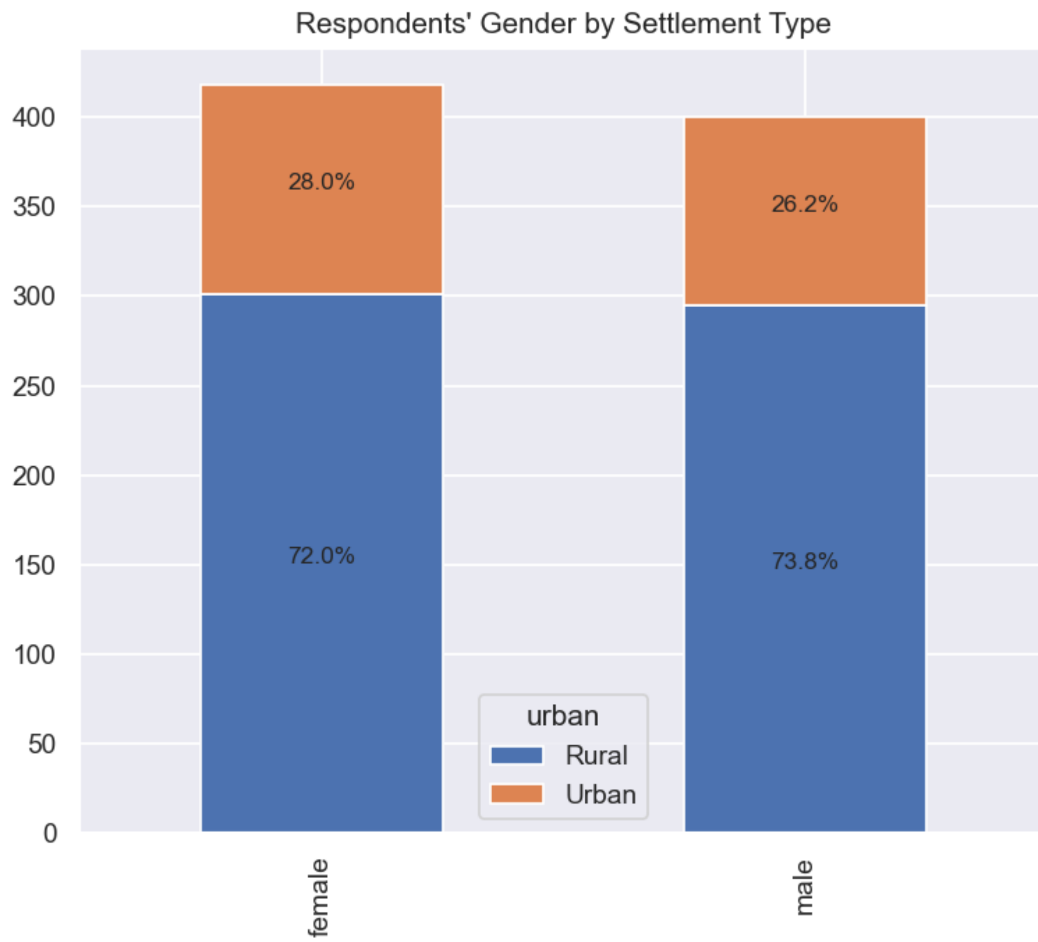
Figure 20: Respondents' educational backgrounds



Source: Field survey (2019)

APPENDIX XXV

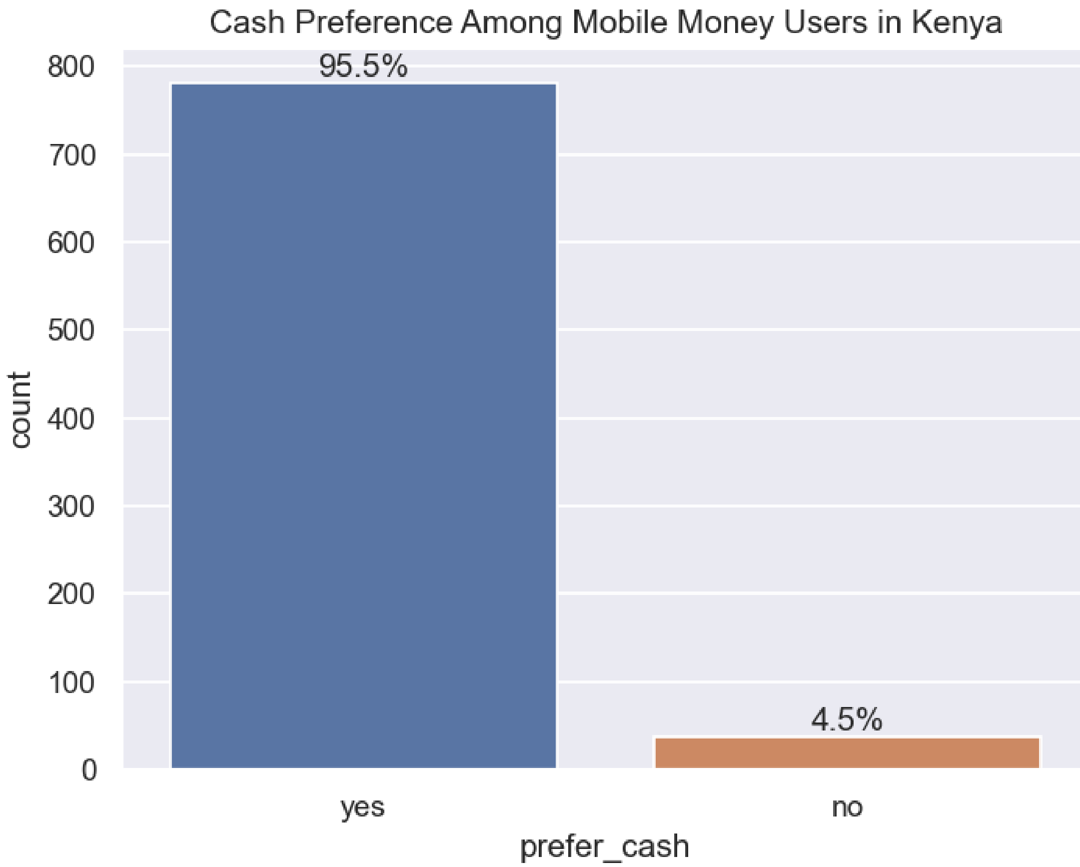
Figure 21: Respondents' gender and settlement types



Source: Field survey (2019)

APPENDIX XXVI

Figure 22: Cash preference among mobile money users in Kenya



Source: Field survey (2019)

APPENDIX XXVII

The evaluation metrics

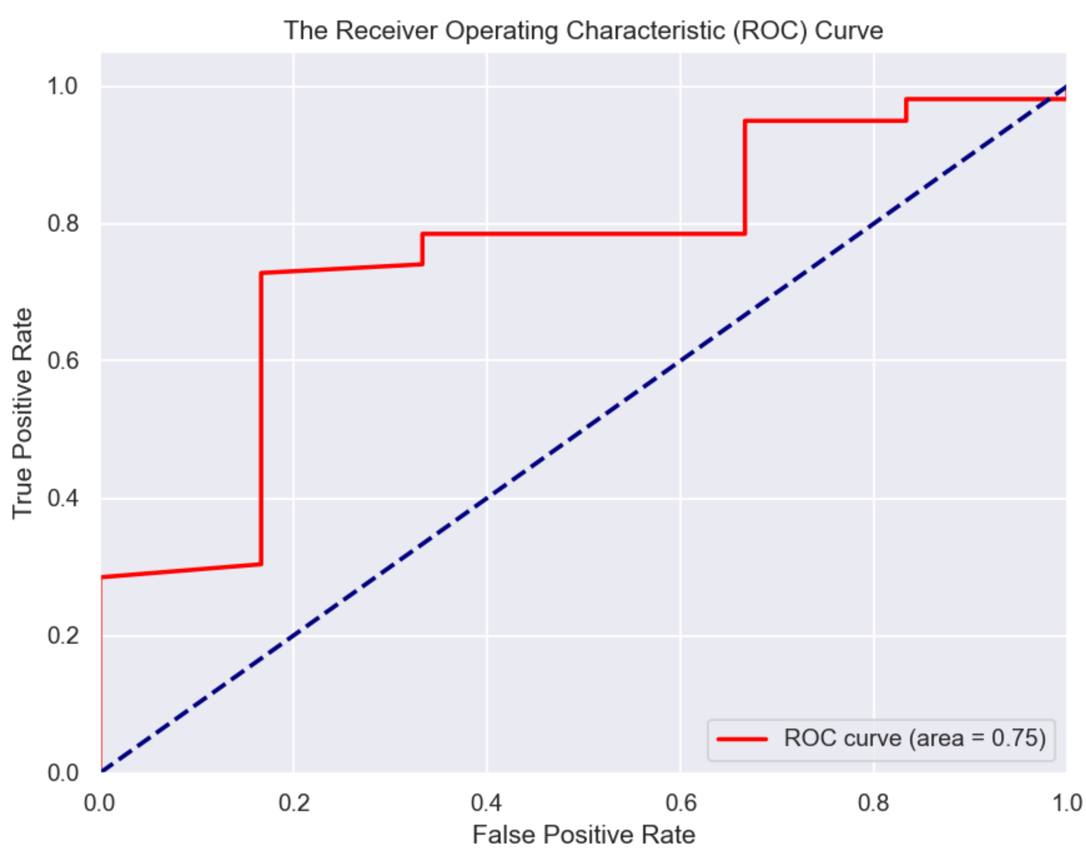
Figure 23: The Confusion Matrix

```
[[  5   1]
 [ 54 104]]
```

Table 8: The Classification Report

	precision	recall	f1-score	support
0	0.08	0.83	0.15	6
1	0.99	0.66	0.79	158
accuracy			0.66	164
macro avg	0.54	0.75	0.47	164
weighted avg	0.96	0.66	0.77	164

Figure 24: The ROC curve



Source: Data analysis Jupyter notebook; Field survey (2019)