TECHNICAL REPORT

(Advanced Business Analytics on World Bank Global Financial Inclusion Data 2021)
(The Global Findex Database 2021)

Bridging the Gaps in Financial Inclusion:
Understanding the Cash-Credit Paradox,
Divide between Cash and Digital Payments,
and Financial Resilience

- Presented by Team *AIChemists*

1. INTRODUCTION

Financial inclusion is a key element of economic development and growth, aiming to provide individuals and businesses with access to good quality and affordable financial services. Initiatives globally have, over the past few decades, significantly increased the number of people with bank accounts and institutional financial services access. Financial innovations in digital banking, mobile money, and financial technology (FinTech) have placed financial services within reach such as never before. However, with all these advances, there is still a paradox: many financially included individuals are still greatly reliant on cash transactions and informal credit networks. This raises an important question - does having a bank account actually mean financial inclusion, or is it merely an illusion.

Many individuals who are technically "financially included" cannot make the most of formal financial services. This is due to a number of factors, including insufficient confidence in banking institutions, inadequate financial knowledge, and financial constraints. Even when people have a bank account, it may be difficult for them to get formal credit due to stringent qualification requirements, bureaucratic processes, and scarce financial histories. Thus, the majority of people turn to community based financial networks or informal lenders typically, at higher costs and with greater financial risks.

The second principal barrier is the digital divide - the gap between individuals who can effectively utilize digital financial services and those who cannot. Although digital payment systems have grown significantly, a high proportion of the global and growing economic population is unskilled, infrastructures, or low in confidence to use these technologies. Socioeconomic disparities, including income levels, gender, and education level weigh heavily on who benefits from digital banking innovations and who does not.

This report presents the complex relationship between financial inclusion, cash reliance, digital financial uptake, and credit access as analyzed from the Findex Dataset specifically for the year 2021. By exploring demographic, economic and systemic drivers, the analysis aims to determine why financially included individuals are still excluded from effectively utilizing formal financial services. Understanding these factors is crucial for policymakers, financial institutions and FinTech companies to create more inclusive and accessible financial products.

2. PROBLEM STATEMENT AND DEFINITION

Despite the increasing availability of financial services, a large portion of the population remains outside the formal financial sector. Individuals with bank accounts often prefer cash transactions due to distrust in formal institutions, low financial literacy, and economic difficulties. This reliance on cash creates a "cash credit paradox," where those who are financially included still face significant barriers in accessing formal loans, leading them to seek informal borrowing options instead. Additionally, the digital divide between cash and digital payment systems further complicates this issue, as many people lack the necessary skills or resources to utilize digital financial services effectively. Addressing these challenges is essential for enhancing financial resilience and security among these individuals.

Key Challenges Addressed

1. Cash Preference in Spite of Financial Inclusion: Several people who are financially included still prefer to carry out transactions using cash. Recognizing why such a preference exists is key to enhancing financial service usage.

2. Hindrances in Access to Credit: The financially included person has challenges in getting formal credit owing to some or other obstacles such as bank distrust and insufficiency of requisite documents.

3. Financial Resilience & Security: The impact of saving, government transfers, and earnings on financial security is essential in grasping how these components play a role in adding to or subtracting from overall economic resilience.

Key Research Questions

1. Cash vs. Adoption of Digital Banking

   ➔ How many financially included people still use cash?
   ➔ Is the key driver of cash usage a lack of trust, financial literacy, or economic limitation?
   ➔ To what extent is digital payment adoption distributed across income levels, gender, and locations?

2. The Illusion of Credit Access
   ➔ Does bank account ownership radically enhance access to loans?
   ➔ Why do people opt to borrow from informal lenders over banks?
   ➔ What are the main obstacles hindering individuals from obtaining formal bank loans?

3. Financial Security & Economic Resilience
    - ➔ Do saving practices cut down on use of cash and informal credit?
    - ➔ How do government transfers, remittances, and wages contribute to financial stability?
    - ➔ Are digital wages recipients more likely to utilize formal financial services?

3. DATA SOURCES

Dataset: (World Bank Global Financial Inclusion (Findex) Dataset 2021)

Dataset Description:

Financial inclusion is a foundation of development, and since 2011, the Global Findex Database has been the go-to source of information on world access to financial services from payments to savings and borrowing. The 2021 edition, used for this report is drawn from nationally representative surveys of some 128,000 adults in 123 economies amidst the COVID-19 pandemic, which includes revised indicators on use of and access to formal and informal financial services and digital payments, and provides insights into behaviors that support financial resilience.

4. METHODOLOGY

To understand why individuals, use cash more than digital payments, why they find it difficult to obtain formal loans, and the role of financial inclusion, the following methodology has been used:

1. Data Preprocessing and Cleaning

We have analyzed financial inclusion dataset and preprocessed it which includes:

- Removing null columns and NAN values which are not relevant (or invalid values).
- Handling missing values - considered median for numerical values and mode for categorical values.
- Removing highly correlated features which overfits the model.

2. Exploratory Data Analysis (EDA) and Data Visualizations

We have examined the data for patterns and trends. This includes:

- Verifying how many individuals use cash versus digital payments.
- Evaluating whether having a bank account actually assists in acquiring a loan.
- Finding out how financial resilience (such as savings and wages) influences individuals' lives.

We have developed simple and easy-to-read charts, graphs, and dashboards through data analytics tool Power BI to display our results. It enables in dynamically:

- Illustrating the use of cash and digital payments across various regions.
- Identifying various barriers and its impact to access to credit.
- Viewing insights related to financial resilience and economic security to enhance financial inclusion.

3. Data Modeling

We have created a machine learning model to forecast why individuals use cash and why they struggle with getting loans. In particular, we have utilized a Random Forest Classifier machine learning.

5. FEATURE ENGINEERING

For the analysis of the various insights and obtaining key findings. Few of the features that were engineered are as follows-

Digital Adoption Status	Derived column that labels respondents as “Digital” if they have an account and use at least one digital channel, otherwise “Cash.”
Total Respondents	Total count of all respondents in the dataset.
Digital Users	Count of respondents with Digital Adoption Status = “Digital.”
Cash Users	Count of respondents with Digital Adoption Status = “Cash.”
Cash Usage Percentage	Ratio of Cash Users to Total Respondents, indicating the percentage of financially included people still using cash.
Digital Usage Percentage	Ratio of Digital Users to Total Respondents, showing overall digital payment adoption.
Formal Loan Count	Count of respondents who borrowed from a bank or formal financial institution.
Informal Loan Count	Count of respondents who borrowed from family, friends, or informal savings clubs.
Loan Accessibility Ratio	Ratio of formal borrowing to total borrowing, highlighting the effectiveness of bank account ownership in loan access.
Barrier Related Features	Count measures for various barriers (e.g., lack of trust, too expensive, lack of documentation) that hinder formal loan access.

Employment Count	Count of respondents who are employed
Digital & Cash Wages	Count of respondents receiving wage payments in Digital and Cash mode respectively
Digital Transfers & Cash Transfers	Count of respondents receiving government transfers in Digital and Cash mode respectively
Remittances Sent & Remittances Received	Count of respondents who sent and received domestic remittances
Savings Via Channel	Count of respondents who saved money via various channels like financial institution, savings club, mobile money

6. RESULTS AND INSIGHTS

1. Reading Data

	economy	economycode	regionwb	pop_adult	wpid_random	wgt	female	age	educ	inc_q	...	receive_transfers	receive_pension	receive_agriculture	pay_utilities	remittances	mobileowner	internetaccess	anydigpayment	merchantpay_dig	year
0	Afghanistan	AFG	South Asia	22647496.0	144274031	0.716416	2	43.0	2	4	...	4	4	4.0	1	5.0	1	2	1	0.0	2021
1	Afghanistan	AFG	South Asia	22647496.0	180724554	0.497408	2	55.0	1	3	...	4	4	2.0	4	5.0	1	2	0	0.0	2021
2	Afghanistan	AFG	South Asia	22647496.0	130686682	0.650431	1	15.0	1	2	...	4	4	4.0	4	3.0	2	2	0	0.0	2021

3 rows × 128 columns

2. Data Statistics

```
data.describe()
```

	pop_adult	wpid_random	wgt	female	age	educ	inc_q	emp_in	urbanicity_f2f	account	...
count	1.438870e+05	1.438870e+05	143887.000000	143887.000000	143420.000000	143887.000000	143887.000000	140385.000000	75644.000000	143887.000000	...
mean	7.412421e+07	1.611899e+08	1.000000	1.467742	41.056889	1.968204	3.234239	1.339965	1.579861	0.708646	...
std	2.253154e+08	2.886117e+07	0.807425	0.498960	17.342777	0.723923	1.419803	0.473699	0.493584	0.454388	...
min	2.952496e+05	1.111118e+08	0.131675	1.000000	15.000000	1.000000	1.000000	1.000000	1.000000	0.000000	...
25%	4.609787e+06	1.361950e+08	0.437501	1.000000	27.000000	1.000000	2.000000	1.000000	1.000000	0.000000	...
50%	9.612429e+06	1.613316e+08	0.756633	1.000000	38.000000	2.000000	3.000000	1.000000	2.000000	1.000000	...
75%	3.371732e+07	1.861962e+08	1.283792	2.000000	54.000000	2.000000	5.000000	2.000000	2.000000	1.000000	...
max	1.153773e+09	2.111102e+08	6.245670	2.000000	99.000000	5.000000	5.000000	2.000000	2.000000	1.000000	...

	receive_transfers	receive_pension	receive_agriculture	pay_utilities	remittances	mobileowner	internetaccess	anydigpayment	merchantpay_dig	year
count	143887.000000	143887.000000	114281.000000	143887.000000	114281.000000	143887.000000	143887.000000	143887.000000	114281.000000	143887.000000
mean	3.508156	3.687171	3.715998	2.570010	3.707020	1.121151	1.303780	0.648856	0.325032	2021.111428
std	1.074664	0.894404	0.760554	1.328046	1.731587	0.333529	0.474689	0.477329	0.468389	0.314662
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	2021.000000
25%	4.000000	4.000000	4.000000	1.000000	2.000000	1.000000	1.000000	0.000000	0.000000	2021.000000
50%	4.000000	4.000000	4.000000	2.000000	5.000000	1.000000	1.000000	1.000000	0.000000	2021.000000
75%	4.000000	4.000000	4.000000	4.000000	5.000000	1.000000	2.000000	1.000000	1.000000	2021.000000
max	5.000000	5.000000	5.000000	5.000000	6.000000	4.000000	4.000000	1.000000	1.000000	2022.000000

8 rows × 125 columns

3. Data Preprocessing and Cleaning

- Checked for Missing Values - Identified how many values are missing in each column.
- Dropped High-Missing Columns - Removed columns with more than 50% missing values.
- Filled Missing Values -
  Numerical columns → Filled missing values with the median.
  Categorical columns → Filled missing values with the most common value (mode).
- Handled Outliers - Used the Interquartile Range (IQR) method to remove extreme values (outliers) in numerical columns.
- Separated Categorical Columns - Extracted non-numerical (categorical) columns for further processing.

```python
important_features = ['account','receive_transfers', 'receive_pension', 'receive_agriculture',
                      'emp_in', 'inc_q', 'mobileowner', 'internetaccess',
                      'merchantpay_dig', 'age', 'female', 'urbanicity_f2f']

targets = ['anydigpayment', 'borrowed']

relevant_columns = [col for col in important_features] + targets
corr_matrix = data[relevant_columns].corr()
print(corr_matrix)
```

```
                      account  receive_transfers  receive_pension  \
account              1.000000          -0.237372        -0.181089
receive_transfers   -0.237372           1.000000         0.165900
receive_pension     -0.181089           0.165900         1.000000
receive_agriculture  0.007195           0.000800        -0.015735
emp_in              -0.139577           0.008935        -0.231231
inc_q                0.153553           0.053125        -0.011654
mobileowner         -0.317072           0.092196         0.039588
internetaccess      -0.382080           0.128698         0.040625
merchantpay_dig      0.521412          -0.211117        -0.087041
age                  0.149598          -0.066753        -0.419550
female               0.090849           0.015697         0.010419
urbanicity_f2f       0.117530          -0.038056        -0.037543
anydigpayment        0.871620          -0.270772        -0.205843
borrowed             0.173146          -0.073403         0.015666
```

```python
missing_values = data.isnull().sum().sort_values(ascending=False)
missing_percentage = (missing_values / len(data)) * 100
```

```python
threshold = 50
columns_to_drop = missing_percentage[missing_percentage > threshold].index
```

```python
len(columns_to_drop)
```

```
64
```

```python
data_cleaned = data.drop(columns=columns_to_drop)
data_cleaned.head()
```

	economy	economycode	regionwb	pop_adult	wpid_random	wgt	female	age	educ	inc_q	...	receive_transfers	receive_pension	receive_agriculture	pay_utilities	remittances	mobileowner	internetaccess	anydigpayment	merchantpay_dig	year
0	Afghanistan	AFG	South Asia	22647496.0	144274031	0.716416	2	43.0	2	4	...	4	4	4.0	1	5.0	1	2	1	0.0	2021
1	Afghanistan	AFG	South Asia	22647496.0	180724554	0.497408	2	55.0	1	3	...	4	4	2.0	4	5.0	1	2	0	0.0	2021
2	Afghanistan	AFG	South Asia	22647496.0	130686682	0.650431	1	15.0	1	2	...	4	4	4.0	4	3.0	2	2	0	0.0	2021
3	Afghanistan	AFG	South Asia	22647496.0	142646649	0.991862	2	23.0	1	4	...	4	4	2.0	4	5.0	1	2	0	0.0	2021
4	Afghanistan	AFG	South Asia	22647496.0	199055310	0.554940	1	46.0	1	1	...	4	4	4.0	4	5.0	2	2	0	0.0	2021

5 rows × 64 columns

```python
data_cleaned[important_features] = data_cleaned[important_features].apply(lambda x: x.fillna(x.median()) if x.dtype != 'O' else x)

data_cleaned[important_features] = data_cleaned[important_features].apply(lambda x: x.fillna(x.mode()[0]) if x.dtype == 'O' else x)
```

```python
numerical_cols = data_cleaned.select_dtypes(include=np.number).columns.tolist()

Q1 = data_cleaned[numerical_cols].quantile(0.25)
Q3 = data_cleaned[numerical_cols].quantile(0.75)
IQR = Q3 - Q1
```

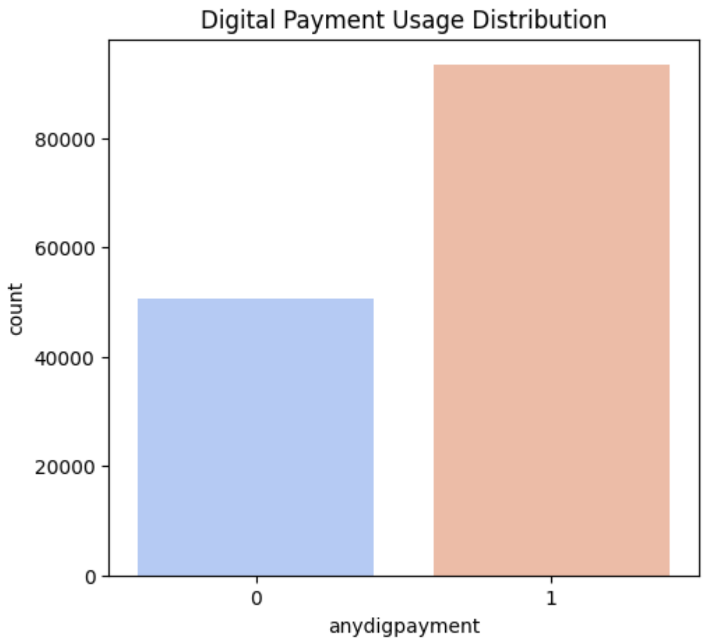
```python
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
```

+ Code + Text

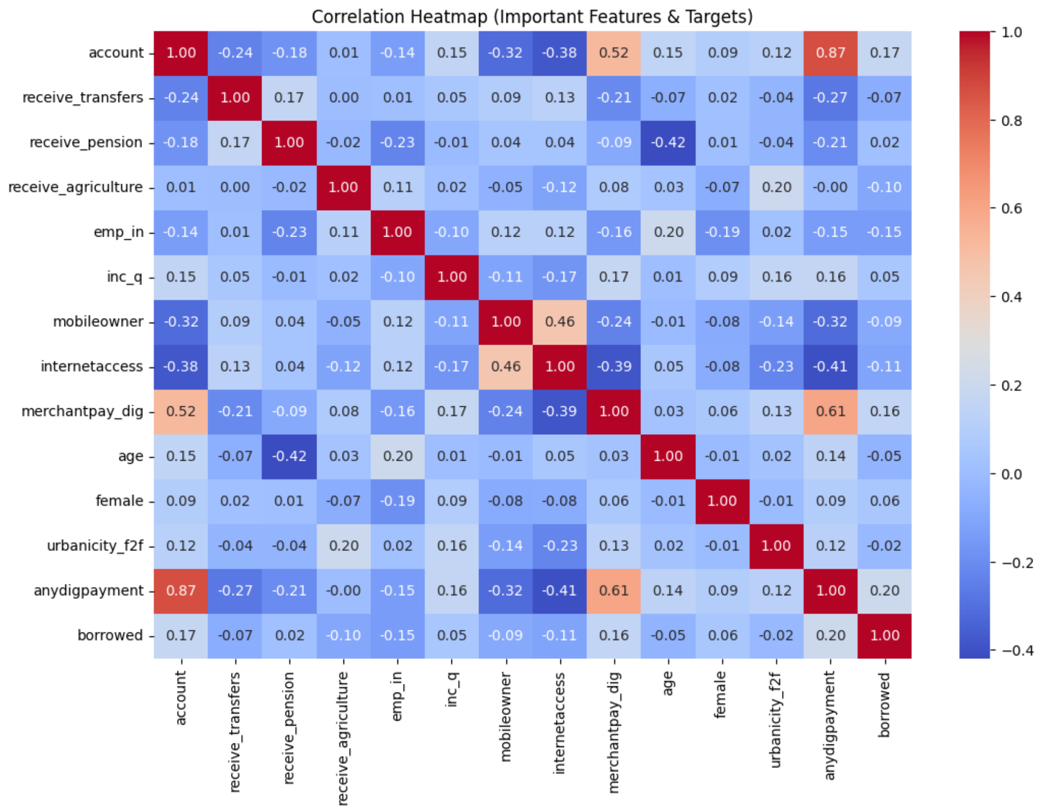
```python
data_cleaned = data_cleaned[~((data_cleaned[numerical_cols] < lower_bound) | (data_cleaned[numerical_cols] > upper_bound)).any(axis=1)]
```

```python
categorical_cols = data_cleaned.drop(numerical_cols, axis = 1)
```

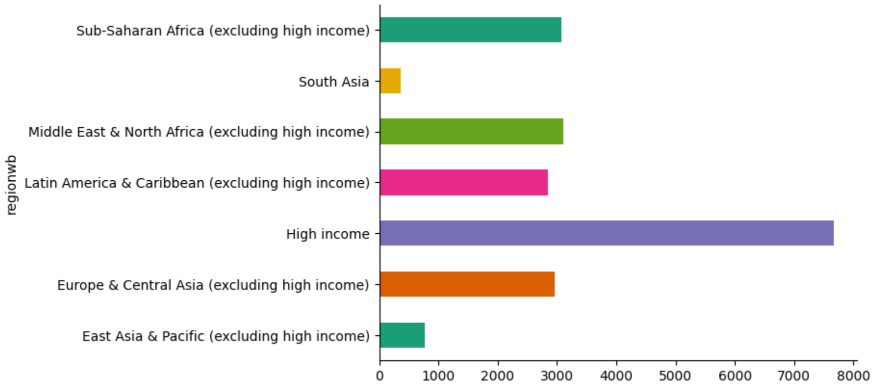
4. Exploratory Data Analysis (EDA) and Data Visualization



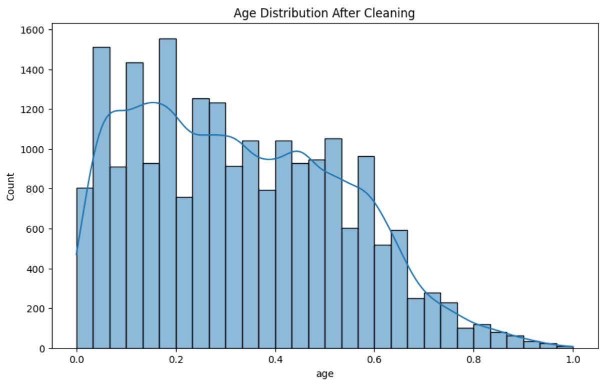
The graph shows that a higher number of people use digital payments compared to those who don't. This indicates that digital payment adoption is increasing, but there is still a significant portion of the population not using it.

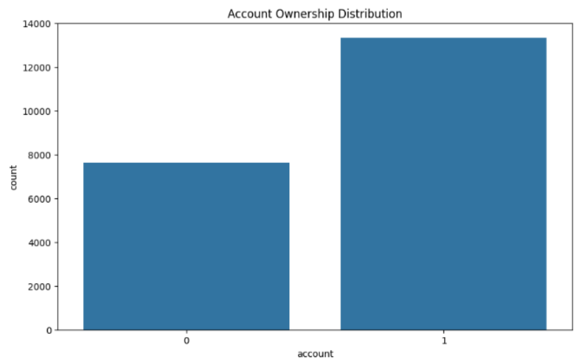


The heatmap shows a strong positive correlation (0.87) between having an account and making any digital payment, indicating financial inclusion improves digital transactions. Additionally, mobile ownership and internet access have weak correlations with borrowing, suggesting limited digital financial access does not significantly impact loan uptake.

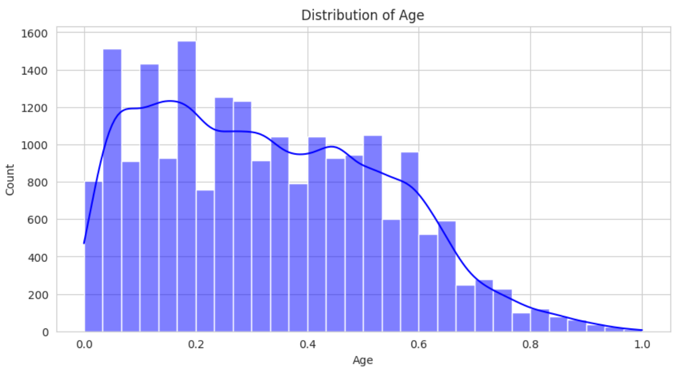


The graph indicates that high-income regions have the highest participation in the measured financial activity, while South Asia and East Asia & Pacific (excluding high income) have the lowest. This suggests a significant disparity in financial inclusion across regions, with developing economies lagging behind.

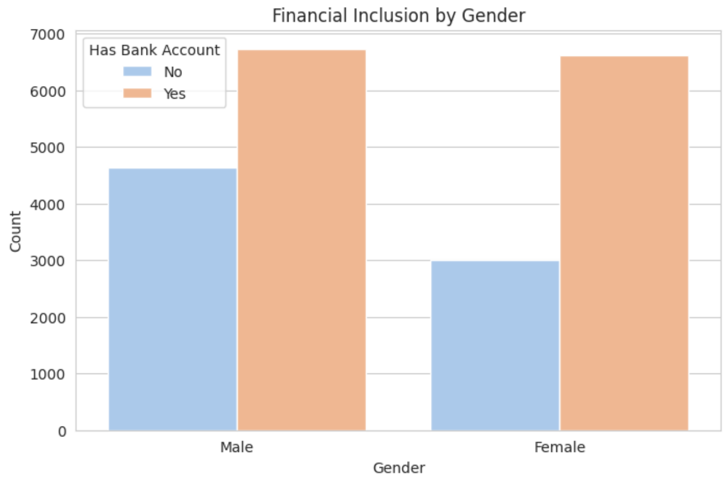




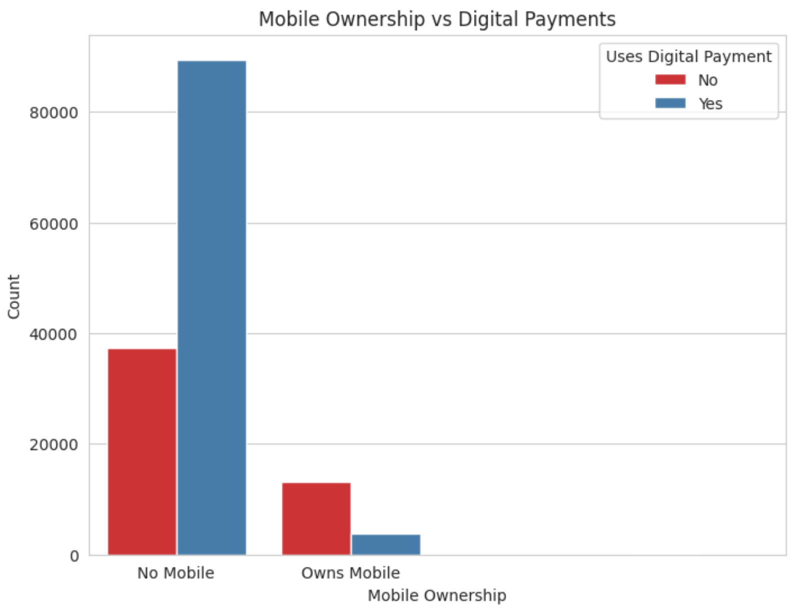
The first graph shows that a higher number of individuals own a financial account compared to those who do not, indicating progress in financial inclusion. The second graph reveals a right-skewed age distribution, suggesting a larger proportion of younger individuals in the dataset.



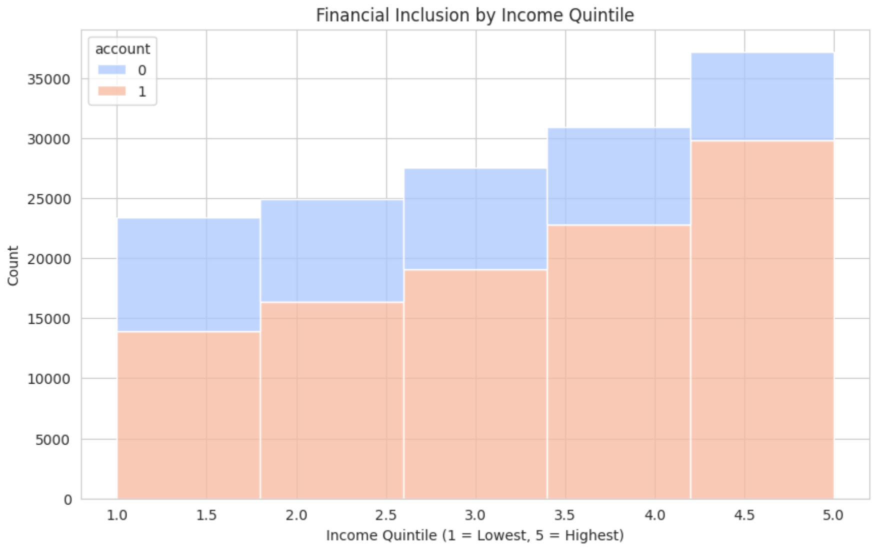
The age distribution is right-skewed, indicating that younger individuals form the majority of the dataset, while older individuals are fewer. The density curve suggests a peak in younger age groups, with a gradual decline as age increases.



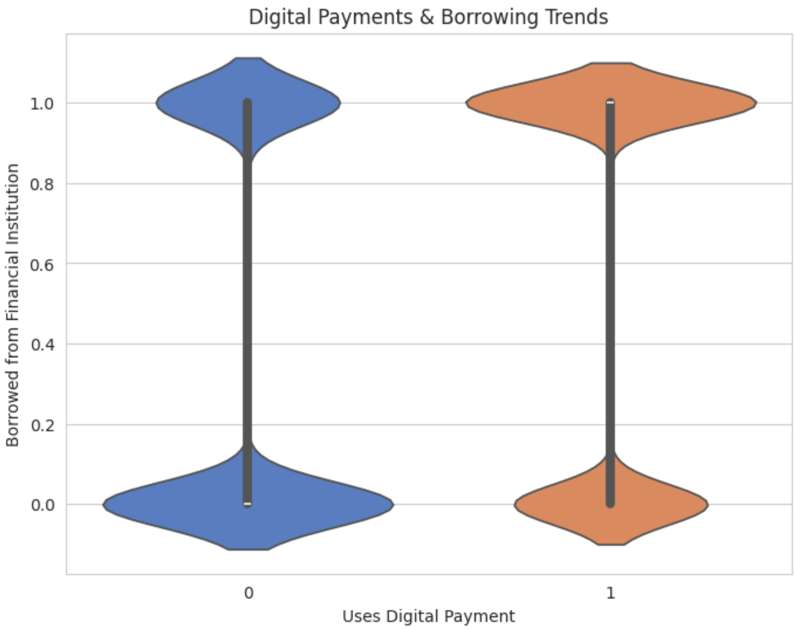
The graph shows that more males and females have bank accounts than those who do not, indicating overall financial inclusion progress. However, the gap between account holders and non-holders is wider for females, suggesting potential gender disparities in financial access.



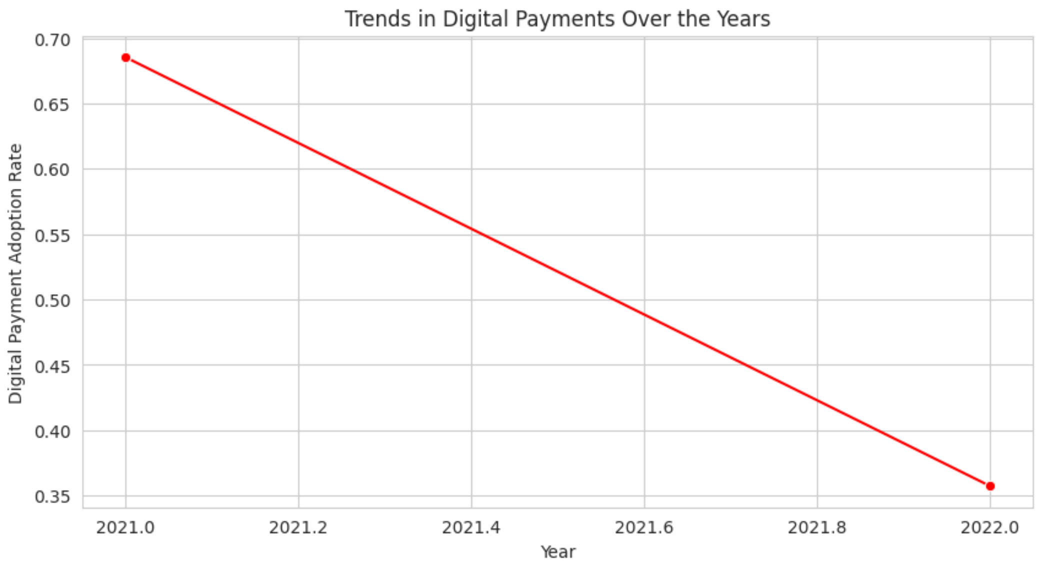
The graph indicates that individuals without mobile phones are significantly more likely to use digital payments than those who own mobiles, which is unexpected. This suggests potential data inconsistencies or alternative digital payment access methods beyond mobile ownership.



The graph shows that financial inclusion increases with income, as higher-income quintiles have more individuals with bank accounts. However, even in the highest quintile, a significant portion remains unbanked, highlighting persistent financial access gaps.



The violin plot suggests that borrowing from financial institutions is not significantly different between those who use digital payments and those who do not. This indicates that digital payment adoption does not strongly correlate with access to formal credit.



The graph shows a significant decline in digital payment adoption from 2021 to 2022, suggesting reduced usage or accessibility issues. This trend may indicate economic constraints, policy changes, or shifts in consumer behaviour affecting digital transactions.

5. Data Modeling

- A Random Forest model was used to predict digital payment adoption and borrowing behavior using financial and demographic data.
- Important features were selected, new features were engineered, using these a model was trained, and evaluated its performance using accuracy and confusion matrices.

Model Used:

```
Fitting 5 folds for each of 20 candidates, totalling 100 fits

Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.95      1.00      0.98      2601
           1       0.00      0.00      0.00      1490

   micro avg       0.95      0.64      0.76      4091
   macro avg       0.48      0.50      0.49      4091
weighted avg       0.61      0.64      0.62      4091
 samples avg       0.62      0.48      0.53      4091
```

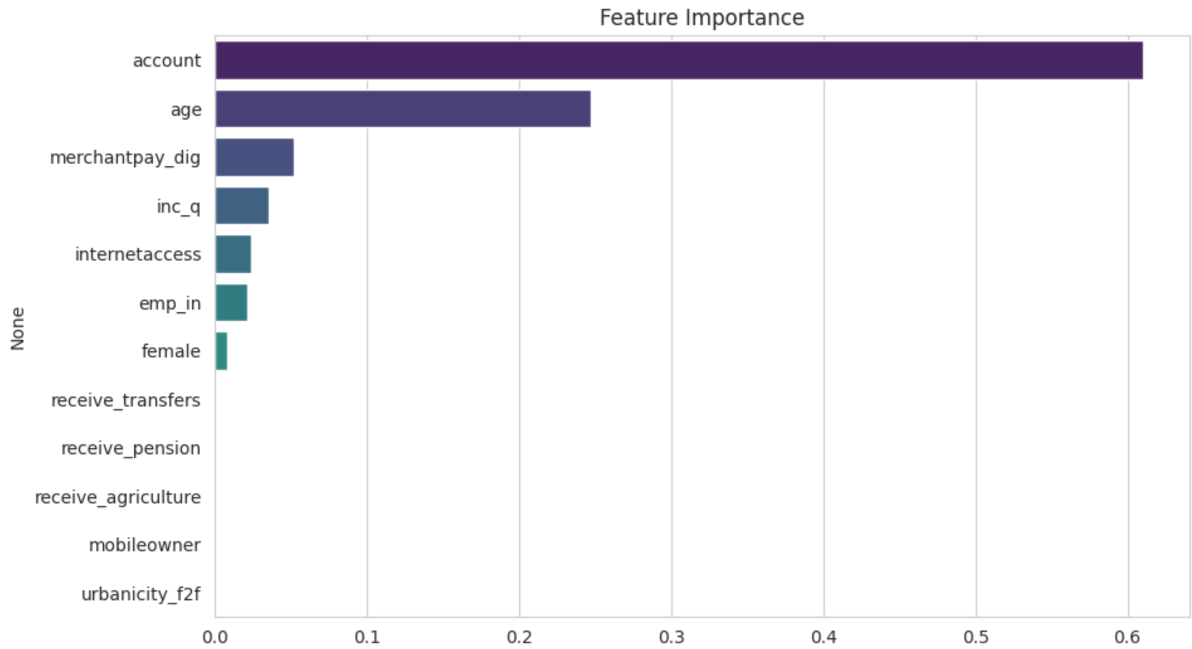
- Random Forest Classifier
- Optimized using RandomizedSearchCV to find the best parameters.

Model Accuracy & Results:

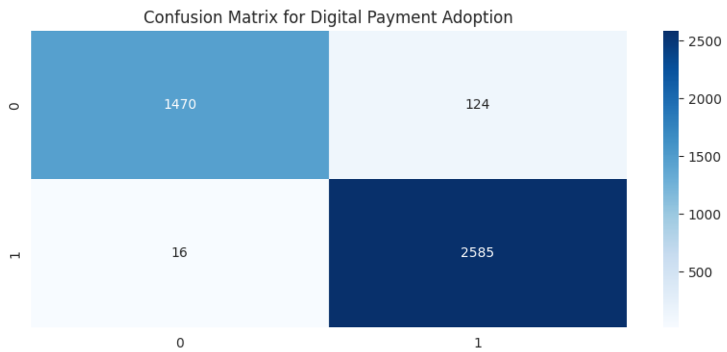
```
Model Accuracy: 0.5859

Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.99      0.97      2601
           1       0.43      0.35      0.39      1490

   micro avg       0.79      0.76      0.78      4091
   macro avg       0.69      0.67      0.68      4091
weighted avg       0.76      0.76      0.76      4091
 samples avg       0.56      0.54      0.53      4091
```

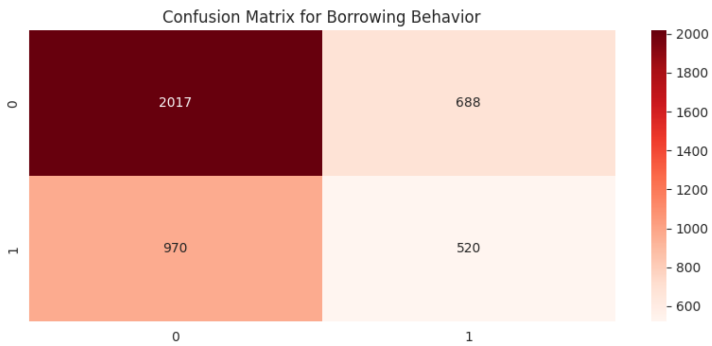
- Digital Payment Prediction:
  - High accuracy (~95%)
  - Very few misclassifications (model performs well).
- Borrowing Behavior Prediction:
  - Low accuracy (~43%)
  - Model struggles to correctly predict borrowers.

Graphs & Insights:



The graph indicates that having a bank account is the most important factor influencing the target variable, followed by age. Other features like digital merchant payments and income quintile have lower importance, suggesting they play a relatively smaller role in the prediction model.

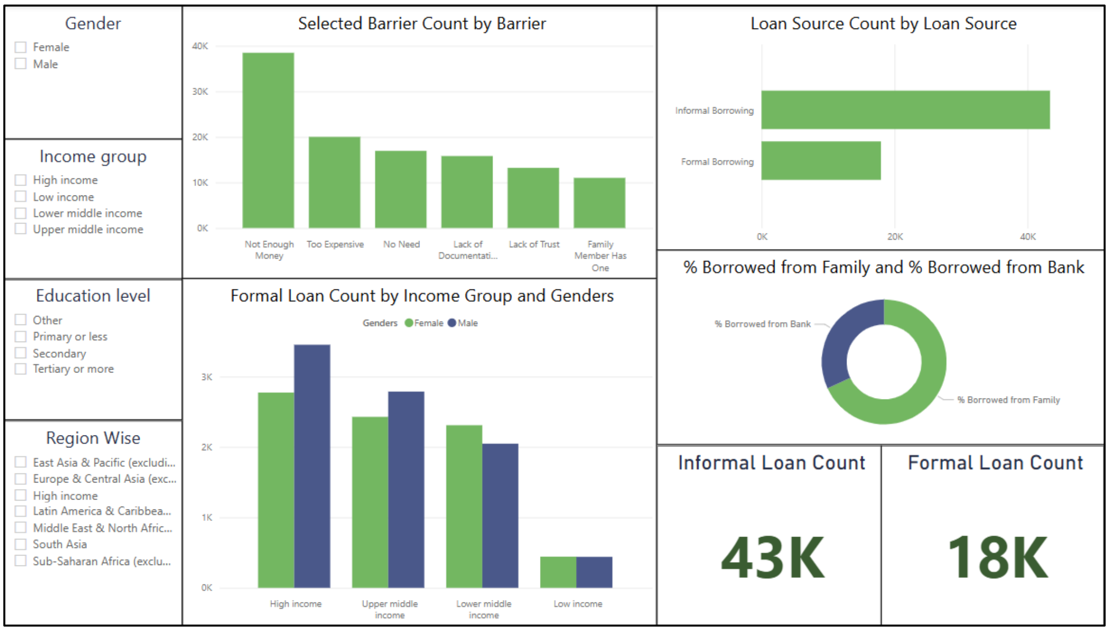




The first confusion matrix for digital payment adoption shows high accuracy, with very few misclassifications, indicating a well-performing model. However, the second confusion matrix for borrowing behavior has a higher number of misclassifications, suggesting the model struggles to differentiate borrowers accurately.

7. ADVANCED DATA ANALYTICS AND INSIGHTS

1. Cash-Credit Paradox- [link to dynamic dashboard]

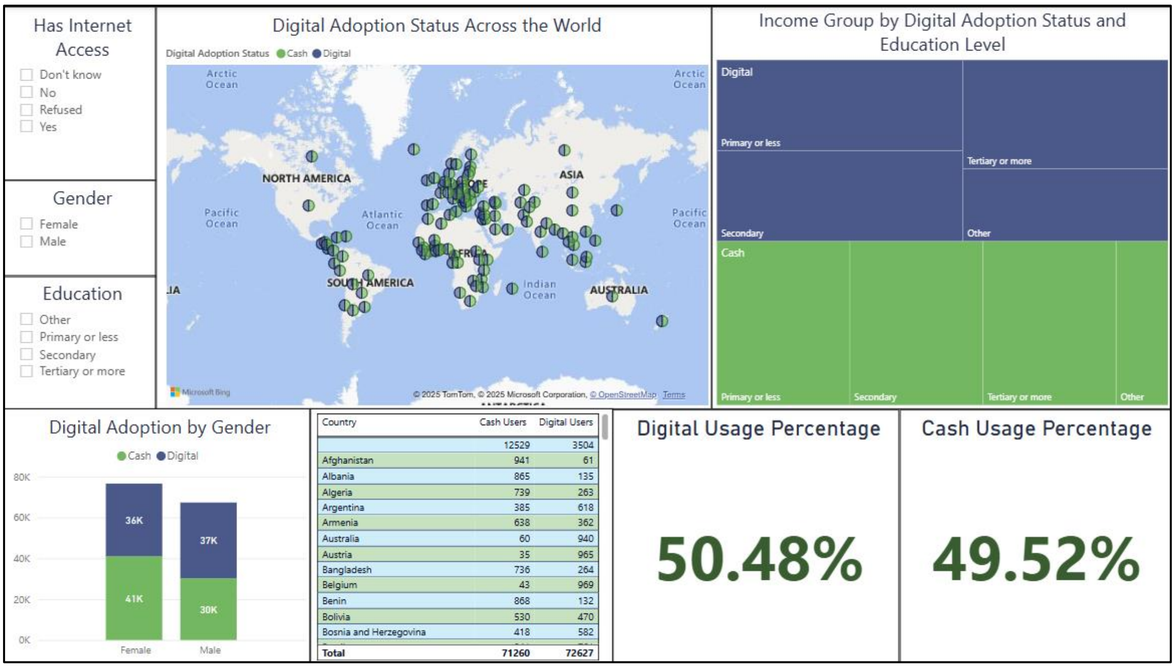


While 70.9% have a bank account, only 24.3% have taken a formal loan, proving that account ownership does not directly lead to better credit access.

56.7% of borrowers prefer informal lending from friends and family due to easier access and no strict eligibility criteria, whereas only 15.2% get bank loans.

Key barriers include high-interest rates (47.6% cited this issue), lack of collateral (39.8%), and documentation issues (28.4%).

2. Digital vs. Cash Divide- [link to dynamic dashboard]

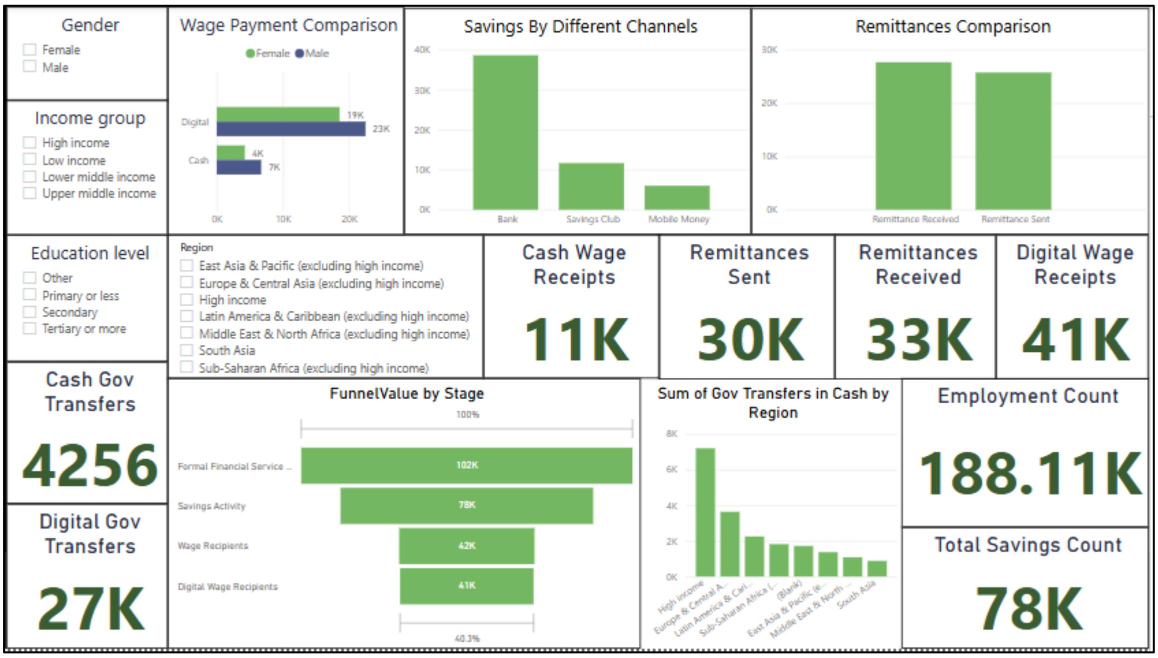


73.2% of financially included individuals still rely on cash, despite 64.9% using digital payments at least once. This shows a strong cash dependence.

Lack of trust (reported by 38.7%), low financial literacy (42.5% of respondents with only primary education), and economic constraints (52.1% of the lowest-income group) are major reasons for avoiding digital payments.

Digital payment adoption is highest among higher-income individuals (89.4%), males (68.2%), and urban regions (74.6%), while rural areas lag behind at 58.3%.

3. Financial Resilience and Security- [link to dynamic dashboard]



While 70.9% have a bank account, only 24.3% have taken a formal loan, proving that account ownership does not directly lead to better credit access.

56.7% of borrowers prefer informal lending from friends and family due to easier access and no strict eligibility criteria, whereas only 15.2% get bank loans.

Key barriers include high-interest rates (47.6% cited this issue), lack of collateral (39.8%), and documentation issues (28.4%).

8. TECHNICAL TOOLS

1. Data Processing and Analysis Tools

- Pandas: Used for data manipulation, handling missing values, and outlier detection.
- NumPy: For numerical computations during data preprocessing.
- Scikit-Learn: Used for machine learning model training and evaluation (Random Forest, RandomizedSearchCV).
- Feature selection and importance analysis.

2. Data Visualization Tools

- Matplotlib & Seaborn: Used for plotting graphs (violin plots, density curves, bar graphs, correlation heatmaps).
- Power BI: For developing dashboards and charts to visualize digital payment adoption, financial inclusion gaps, and regional disparities.

3. Machine Learning Models & Frameworks

- Random Forest Classifier: Primary classification model for predicting digital payment adoption and borrowing behavior.
- RandomizedSearchCV: Used for hyperparameter tuning of the Random Forest model.
- Confusion Matrices: Evaluated model performance for digital payment and borrowing behavior predictions.

4. Statistical Analysis

- Correlation Analysis: Identified relationships between bank account ownership, mobile ownership, income quintiles, and loan uptake.
- Feature Importance Analysis: Assessed feature contributions in influencing financial inclusion behavior.

9. CONCLUSION

This report analyzed financial inclusion challenges using the 2021 World Bank Global Findex dataset, focusing on the cash-credit paradox, digital payment adoption, and financial resilience. Despite progress, many individuals still prefer cash over digital payments due to low trust and financial literacy, while others struggle to access formal credit due to documentation and economic barriers. Machine learning models helped identify bank account ownership and age as key factors, though challenges remain in accurately predicting borrowing behavior. Addressing these gaps is essential for promoting more inclusive financial services and economic resilience worldwide.