MY PORTFOLIO

MARY MWANGI

UNIVERSITY OF NAIROBI

BACHELOR OF ECONOMICS AND STATISTICS

This portfolio includes my complete Stata-based analysis of the Kenya Health Demographic Survey (KDHS) 2022, conducted as my primary contribution to the as Applied Econometrics (STA432) Group project, where I served as group leader

Email: nyaruiru41@gmail.com

Phone Contact: +254707083163

The Kenya Demographic and Health Survey 2022 (KDHS 2022) is a nationally representative microdata dataset designed to provide comprehensive information on the health, demographic, and socioeconomic characteristics of households and individuals in Kenya. The survey was conducted to inform policies and programs by tracking key indicators related to public health, fertility, reproductive health, child nutrition, and economic behaviors.

Following clear instructions in the assessment test, out of this data, we were supposed to derive new variables from existing data in the survey document to carry out analyses on their relation.

These are the variables we created:

- ☑ **Age of the respondent** (age_respondent)
- ☑ **Sex** (sex)
- ☑ **Marital Status** (marital_status)
- ☑ **Household COVID 19 Vaccination** (hh_vaccinated_covid)
- ☑ **Ownership of agricultural land** (owns_agri_land)
- ☑ **Mobile money use** (mobile_money_use)

Below is a detailed description of how we coded and labeled these variables:

- **age_respondent**

Replaced data from an already existing variable (hv105 – age of household members)

- **sex**

Replaced data from an already existent variable (hv104 – sex of household member)

Created labels for the data sets: 1 “male”, 2 “female”

Named the label ‘sex_lbl’

- **marital_status**

Replaced data from an already existent variable (hv115 – current marital status)

Created labels for the data sets: 0 “never_married”, 1 “married”, 3 “widowed”, 4 “divorced”, 5 “not_living_together”

Named the label ‘marital_status_lbl’

- **hh_vaccinated_covid**

Replaced data from an already existent variable (sh135l – how many members vaccinated against covid).

- **owns_agri_land**

Replaced data from an already existent variable (hv244 – owns land usable for agriculture).

Created labels for the dataset: 0 “no”, 1 “yes”

Named the label ‘owns_agri-land_lbl’

- **mobile_money_use**

Replaced data from an already existent variable (hv263 – mobile phone used financial transactions)

Created numeric labels for the data: 0 “no”, 1 “yes”

On verifying and inspecting the variables with a ‘codebook’ it was noted that the variable ‘marital_status’ had a couple of missing values which were all dropped.

Significance of the new variables

These variables are critical because they capture key demographic, socioeconomic, and behavioral factors that can explain differences in health and financial outcomes. Here's a breakdown of why each is important:

- **Age of the respondent (age_respondent)**

Age influences many aspects of behavior and risk. For instance, older individuals might be more inclined to get vaccinated because they are at higher risk for COVID-19 complications. Age can also affect how comfortable someone is with newer financial technologies like mobile money. In essence, age is a proxy for life-cycle effects, risk exposure, and potentially technological adoption.

- **Sex (sex)**

Men and women often exhibit different health behaviors and financial practices, influenced by both biological and social factors. Differences in vaccination uptake might emerge due to varying perceptions of risk or access to healthcare. Similarly, financial decision-making and access to mobile money services may differ between sexes because of cultural norms and economic opportunities.

- **Marital status (marital_status)**

This variable reflects an important aspect of an individual's social context. Being married or in a relationship might provide social support, which can affect health-seeking behavior (such as vaccination) or influence financial practices (like managing household finances with mobile money). Marital status might also be linked to stability and shared decision-making, impacting these outcomes.

- **Household COVID-19 Vaccination (hh_vaccinated_covid)**

As one of your major outcome variables, this indicates whether any member of the household has been vaccinated against COVID-19. It's a direct measure of the public health outcome your analysis is addressing and helps in understanding the determinants of vaccination uptake.

- **Ownership of Agricultural Land (owns_agri_land)**

This variable often serves as a proxy for socioeconomic status, particularly in settings where landownership can signal wealth, stability, and rural versus urban residency. It may affect financial behaviors, like the adoption of mobile money, due to differences in financial infrastructure availability and economic opportunities.

- **Mobile Money Use (mobile_money_use)**

As another key outcome in the study, mobile money usage is an indicator of financial inclusion and technology adoption. Understanding which groups are more likely to use mobile money can highlight disparities in access to financial services, particularly in environments where traditional banking is less accessible.

These variables, therefore, help paint a full picture of the factors influencing both health-related decisions (like COVID-19 vaccination uptake) and economic behaviors (like mobile money usage). They are chosen not only because they are foundational demographic indicators but also because they help control for important differences between individuals when estimating the effects of each on the outcomes of interest.

Data Analysis Outputs: Tabular and Graphical Representation (from STATA)

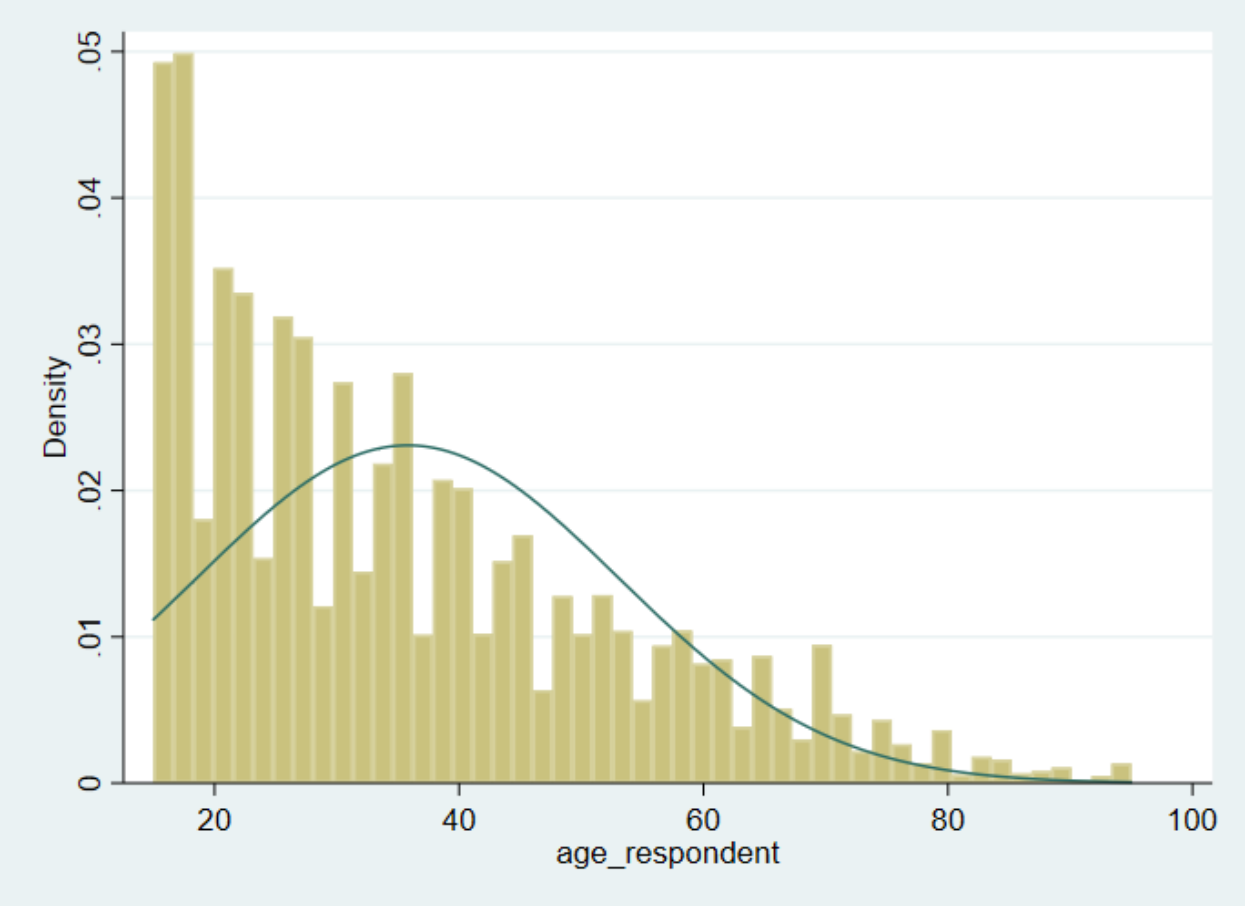
Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
age respondent	90788	35.806	17.28	15	95
sex	90788	1.521	.5	1	2
hh vaccinated covid	90788	1.613	5.896	0	98
owns agri land	90788	.573	.495	0	1
marital status	90788	1.007	1.203	0	5
mobile money use	90788	.91	.286	0	1

Tabulation of age_respondent

	Freq.	Percent	Cum.
15	3514	3.87	3.87
16	3788	4.17	8.04
17	3687	4.06	12.10
18	3707	4.08	16.19
19	2671	2.94	19.13
20	2998	3.30	22.43
21	2218	2.44	24.87
22	2706	2.98	27.86
23	2258	2.49	30.34
24	2278	2.51	32.85
25	2591	2.85	35.71
26	2132	2.35	38.05
27	2170	2.39	40.44
28	2346	2.58	43.03
29	1793	1.97	45.00
30	2709	2.98	47.99
31	1350	1.49	49.47
32	2143	2.36	51.83
33	1498	1.65	53.48
34	1738	1.91	55.40
35	2420	2.67	58.06
36	1733	1.91	59.97
37	1503	1.66	61.63
38	1803	1.99	63.61
39	1268	1.40	65.01
40	2150	2.37	67.38
41	835	0.92	68.30
42	1512	1.67	69.96
43	1146	1.26	71.23
44	1100	1.21	72.44
45	1521	1.68	74.11
46	989	1.09	75.20
47	942	1.04	76.24
48	1110	1.22	77.46
49	783	0.86	78.33
50	1506	1.66	79.98
51	765	0.84	80.83
52	1140	1.26	82.08
53	792	0.87	82.95
54	748	0.82	83.78
55	838	0.92	84.70
56	775	0.85	85.56
57	617	0.68	86.23
58	852	0.94	87.17
59	694	0.76	87.94
60	1209	1.33	89.27
61	531	0.58	89.85
62	723	0.80	90.65
63	569	0.63	91.28
64	559	0.62	91.89
65	728	0.80	92.70
66	345	0.38	93.08
67	406	0.45	93.52
68	441	0.49	94.01
69	363	0.40	94.41
70	1036	1.14	95.55
71	243	0.27	95.82
72	451	0.50	96.31

73	321	0.35	96.67
74	306	0.34	97.00
75	332	0.37	97.37
76	229	0.25	97.62
77	163	0.18	97.80
78	203	0.22	98.03
79	141	0.16	98.18
80	396	0.44	98.62
81	83	0.09	98.71
82	156	0.17	98.88
83	109	0.12	99.00
84	109	0.12	99.12
85	127	0.14	99.26
86	98	0.11	99.37
87	67	0.07	99.44
88	54	0.06	99.50
89	41	0.05	99.55
90	117	0.13	99.68
91	24	0.03	99.70
92	50	0.06	99.76
93	23	0.03	99.78
94	30	0.03	99.81
95	168	0.19	100.00
Total	90788	100.00	



Interpretation: The respondents vary widely in age, with a central tendency around 35 years. The spread suggests diversity in age groups. Bigger population exists between ages 15 – 45.

Tabulation of sex

	Freq.	Percent	Cum.
male	43511	47.93	47.93

female	47277	52.07	100.00
Total	90788	100.00	

Interpretation: The majority gender is female as depicted by the frequency and percentages.

Tabulation of hh_vaccinated_covid

	Freq.	Percent	Cum.
0	34260	37.74	37.74
1	22499	24.78	62.52
2	19363	21.33	83.85
3	7776	8.57	92.41
4	4007	4.41	96.82
5	1531	1.69	98.51
6	654	0.72	99.23
7	232	0.26	99.49
8	80	0.09	99.57
9	25	0.03	99.60
10	21	0.02	99.63
12	20	0.02	99.65
98	320	0.35	100.00
Total	90788	100.00	

Interpretation: The number of covid vaccinations is higher in the ages 0 – 2, as it gradually decreases, however there is leap at the oldest age,98. Not a big leap but statistically significant.

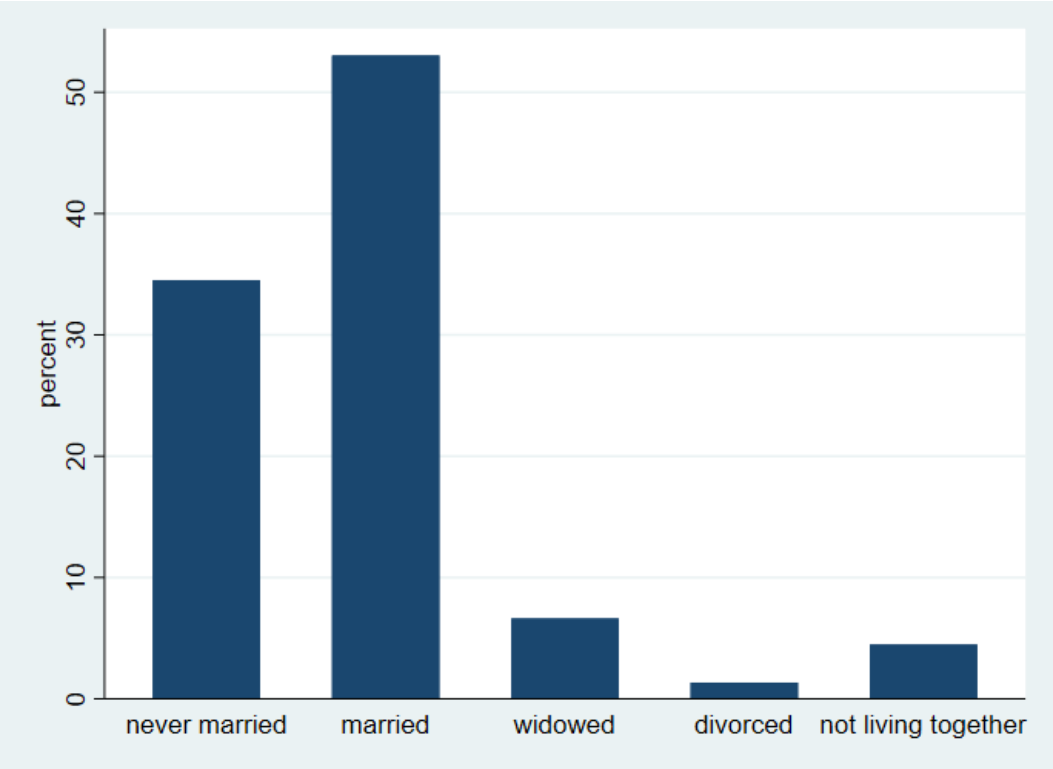
Tabulation of owns_agri_land

	Freq.	Percent	Cum.
no	38759	42.69	42.69
yes	52029	57.31	100.00
Total	90788	100.00	

Interpretation: about 57.31% of the population are in possession of agricultural land.

Tabulation of marital_status

	Freq.	Percent	Cum.
never married	31337	34.52	34.52
married	48153	53.04	87.56
widowed	6016	6.63	94.18
divorced	1217	1.34	95.52
not living together	4065	4.48	100.00
Total	90788	100.00	



Interpretation:

Never Married
Frequency: 31,337 respondents
Percentage: 34.52% of the total sample
More than a third of the respondents have never been married, suggesting a youthful population or a trend toward delayed marriage.

Married
Frequency: 48,153 respondents
Percentage: 53.04%
Over half of the respondents are married, making this the dominant marital category in the sample.

Widowed
Frequency: 6,016 respondents
Percentage: 6.63%
This group represents a smaller proportion, likely consisting of older individuals.

Divorced
Frequency: 1,217 respondents
Percentage: 1.34%
Divorce rates appear relatively low within the sample.

Not Living Together
Frequency: 4,065 respondents
Percentage: 4.48%
This could indicate separated individuals who aren't formally divorced or those in non-cohabiting relationships or couples separated by distance due to work related issues.

Tabulation of mobile_money_use

	Freq.	Percent	Cum.
no	8171	9.00	9.00
yes	82617	91.00	100.00
Total	90788	100.00	

Interpretation:
Respondents Who Do Not Use Mobile Money
Frequency: 8,171 respondents
Percentage: 9.00% of the total sample
Cumulative Percentage: 9.00% (only these respondents so far)
Interpretation: A small proportion of individuals do not use mobile money services, possibly due to lack of access, digital literacy barriers, or personal preferences for cash transactions.

Respondents Who Use Mobile Money
Frequency: 82,617 respondents
Percentage: 91.00%
Cumulative Percentage: 100.00% (all respondents accounted for)
Interpretation: The vast majority (91%) utilize mobile money services, suggesting **high financial inclusion** and a strong preference for digital transactions

Question One
Using the downloaded dataset, estimate the effect of age, sex, and marital status on uptake of COVID 19 vaccinating (20 marks) hint: present descriptive, diagnostics and regression results

Answers (Question One)

Logistic regression

hh_vaccinated_covid	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
age_respondent	1.007	0	16.27	0	1.007	1.008	***
sex	1.125	.016	8.41	0	1.094	1.156	***
marital_status	.871	.006	-21.55	0	.86	.882	***
Constant	1.221	.032	7.60	0	1.16	1.286	***

Mean dependent var	1.613	SD dependent var	5.896
Pseudo r-squared	0.004	Number of obs	90788
Chi-square	539.331	Prob > chi2	0.000
Akaike crit. (AIC)	119809.659	Bayesian crit. (BIC)	119847.324

*** p<.01, ** p<.05, * p<.1

Understanding the Regression Model

Dependent Variable: hh_vaccinated_covid (Binary: 0 = Not vaccinated, 1 = Vaccinated).
Independent Variables:
Age (age_respondent)
Sex (sex)
Marital status (marital_status)

Each variable includes:
Coefficient (Coef.): Shows how much the variable influences vaccination likelihood.
Standard Error (St.Err.): Measures the variability of the coefficient.
t-Value: Indicates the strength of the relationship.
p-Value (Sig.): Tests if the effect is statistically significant.
Confidence Interval (95% Conf.): Range where the true coefficient likely falls.

Key Findings & Interpretation

- ☑ **Age (age_respondent)**

Coefficient: 1.125 (greater than 1)
p-Value: 0.000 (highly significant)
Interpretation:
Older respondents are more likely to be vaccinated.

The coefficient greater than 1 indicates that as age increases, the likelihood of vaccination also increases significantly.

☑ **Sex (sex)**

Coefficient: 0.871 (less than 1)
p-Value: 0.000 (highly significant)
Interpretation:
The negative coefficient suggests that males are less likely to be vaccinated compared to females.
-This may reflect gender-based disparities in vaccine access, beliefs, or healthcare-seeking behavior.

☑ **Marital Status (marital_status)**

Coefficient: 1.221 (greater than 1)
p-Value: 0.000 (highly significant)
Interpretation:
Being married (or certain categories of marital status) increases the likelihood of vaccination.
This could be due to household decision-making, family influence, or better access to healthcare.

☑ **Constant (Intercept)**

Coefficient: 1.221
Interpretation: Represents the baseline probability of vaccination when all independent variables are zero.

☑ **Why Is This Happening?**

✓ Age plays a strong role, likely because older individuals are at higher risk and thus more motivated to get vaccinated.
✓ Men appear less likely to be vaccinated, possibly due to lower healthcare-seeking behavior.
✓ Marital status influences vaccination, which may reflect household decision-making or partner encouragement.

☑ **Model Performance Indicators**

- Pseudo R-Squared: 0.004 - Low, meaning this model explains only a small portion of vaccination behavior.
- Chi-Square (Prob > chi2): 0.000 - The model is statistically significant overall.
- Akaike Information Criterion (AIC) & Bayesian Information Criterion (BIC):
- Lower values indicate a better model fit.

Matrix of correlations

Variables	(1)	(2)	(3)	(4)
(1) hh_vaccinated_~d	1.000			
(2) age_respondent	-0.005	1.000		
(3) sex	-0.002	0.026	1.000	
(4) marital_status	-0.024	0.460	0.172	1.000

Understanding Correlation Coefficients

Range: Correlation values range from -1 to +1:

✓ +1 - Perfect positive correlation (as one variable increases, the other increases).
✓ -1 - Perfect negative correlation (as one variable increases, the other decreases).
✓ 0 - No correlation (no relationship between the variables).

Interpretation of Key Relationships

hh_vaccinated vs. Age (age_respondent) → -0.005

✓ Very close to zero, indicating almost no relationship between age and household vaccination status.
✓ This suggests that age does not strongly influence whether a household is vaccinated.

hh_vaccinated vs. Sex (sex) → -0.002

✓ Another value close to zero, meaning gender has no strong correlation with household vaccination.
✓ In practical terms, men and women appear to have similar vaccination rates.

hh_vaccinated vs. Marital Status (marital_status) → -0.024

✓ Slightly negative, but still close to zero.
✓ This suggests marital status has very little effect on whether a household is vaccinated.

Age vs. Sex (age_respondent & sex) - 0.026

- ✓ Slightly positive, meaning older individuals may be slightly more likely to be a certain gender (possibly reflecting demographic trends).

Age vs. Marital Status (age_respondent & marital_status) - 0.460

- ✓ Moderate positive correlation, suggesting older respondents are more likely to be married.
- ✓ This aligns with expectations, as marital status generally changes with age.

Sex vs. Marital Status (sex & marital_status) - 0.172

- ✓ Weak positive correlation, meaning gender has some influence on marital status.
- ✓ This could indicate differences in marriage rates between men and women.

Why Are These Relationships Happening?

- Vaccination rates appear unrelated to age, sex, or marital status because they may depend more on policy, availability, or health awareness rather than individual demographics.
- Age and marital status are positively correlated since older individuals are naturally more likely to be married.
- Sex has a weak relationship with marital status, possibly reflecting societal patterns in marriage trends.

Question Two

Using the downloaded dataset, estimate the effect of age, sex, and marital status, ownership of agricultural land on uptake of usage of mobile money (20 marks) hint: present descriptive, diagnostics and regression results.

Answers (Question Two)

Logistic regression

mobile_money_use	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
age_respondent	.984	.001	-22.69	0	.983	.985	***
sex	.99	.024	-0.43	.666	.944	1.037	
marital_status	.929	.009	-7.32	0	.911	.948	***
owns_agri_land	2.092	.05	30.81	0	1.996	2.192	***
Constant	14.099	.648	57.54	0	12.884	15.429	***

Mean dependent var	0.910	SD dependent var	0.286
Pseudo r-squared	0.030	Number of obs	90788
Chi-square	1652.571	Prob > chi2	0.000
Akaike crit. (AIC)	53291.407	Bayesian crit. (BIC)	53338.489

*** p<.01, ** p<.05, * p<.1

Understanding the Regression Model

The dependent variable is mobile_money_use (binary: 0 = No, 1 = Yes).
The independent variables included are:

- Age of respondent (age_respondent)
- Sex (sex)
- Marital status (marital_status)
- Ownership of agricultural land (owns_agri_land)

Each variable has:

- **Coef. (Coefficient):** The direction and strength of influence.
- **St.Err. (Standard Error):** How precise the estimate is.
- **t-value:** Indicates how strongly the variable affects mobile money use.
- **p-value:** Tests statistical significance (whether the effect is meaningful).
- **95% Confidence Interval:** The range where the true effect likely falls.
- **Significance Level (Sig):** Shows statistical importance (p<.01, p<.05, p<.1).

Key Findings & Interpretation

- ☑ **Age (age_respondent)**

Coefficient: 0.984
p-value: 0.000 (highly significant)
Interpretation: Older respondents are slightly less likely to use mobile money (since the coefficient is below 1). This suggests mobile money adoption might be higher among younger users, possibly due to tech familiarity.

- ☑ **Sex (sex)**

Coefficient: 0.99
p-value: 0.666 (not statistically significant)
Interpretation: Gender does not significantly impact mobile money usage in this model. Men and women appear to use mobile money at similar rates.

- ☑ **Marital Status (marital_status)**

- Coefficient: 0.929
- p-value: 0.000 (highly significant)
- Interpretation: Individuals in certain marital status categories (such as married respondents) are less likely to use mobile money than the reference group. This could be due to household financial structures or differing payment preferences.

- ☑ **Ownership of Agricultural Land (owns_agri_land)**

Coefficient: 2.092
p-value: 0.000 (highly significant)
Interpretation: Owning agricultural land strongly increases the likelihood of using mobile money. This is likely due to financial transactions associated with farming, such as subsidies, payments, or supply purchases.

- ☑ **Constant (Intercept)**

Coefficient: 14.099
Interpretation: Represents the baseline probability of using mobile money when all independent variables are set to zero.

- ☑ **Model Performance Indicators**

Pseudo R-Squared: 0.030 - Low, meaning predictors explain only a small portion of mobile money usage.
Chi-Square (Prob > chi2): 0.000 - Model is statistically significant overall.
Akaike Information Criterion (AIC) & Bayesian Information Criterion (BIC): Lower values indicate better model fit.

Why Is This Happening?

- ✓ Age slightly reduces mobile money use, possibly due to tech barriers for older individuals.
- ✓ Gender does not significantly influence mobile money adoption—usage is fairly balanced.
- ✓ Marital status affects adoption, likely influenced by household financial management.
- ✓ Agricultural land ownership strongly promotes mobile money use, possibly due to farming transactions.

Matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)
(1) mobile_money_use	1.000				
(2) age_respondent	-0.083	1.000			
(3) sex	-0.011	0.026	1.000		
(4) marital_status	-0.067	0.460	0.172	1.000	
(5) owns_agri_land	0.095	0.107	-0.023	-0.016	1.000

Interpretation of Key Relationships

Mobile Money Use & Other Variables

- ☑ **Mobile money use vs. Age (age_respondent): -0.083**

✓ **Slight negative correlation** → Older respondents are less likely to use mobile money.

✓ This could suggest that younger individuals are more comfortable with digital financial tools.

- ☑ **Mobile money use vs. Sex (sex): -0.011**

✓ **Near zero correlation** → Mobile money use is **not significantly influenced by gender**.

✓ Usage rates appear to be **similar** between men and women.

☑ **Mobile money use vs. Marital Status (marital_status): -0.067**

✓ **Weak negative correlation** → Certain marital categories may use mobile money **less frequently**.

✓ This could reflect differences in household financial management.

☑ **Mobile money use vs. Agricultural Land Ownership (owns_agri_land): 0.095**

✓ **Weak positive correlation** → Individuals who own agricultural land are **slightly more likely** to use mobile money.

✓ This might be because farmers often rely on digital transactions for supply purchases or financial services.

Age (age_respondent) & Other Variables

☑ **Age vs. Sex (sex): 0.026**

✓ **Near zero correlation** → Age and gender are **mostly independent** in this dataset.

☑ **Age vs. Marital Status (marital_status): 0.460**

✓ **Moderate positive correlation** → **Older individuals are more likely to be married**.

✓ This is expected, as people tend to marry as they age.

☑ **Age vs. Agricultural Land Ownership (owns_agri_land): 0.107**

✓ **Weak positive correlation** → Older individuals **slightly more likely** to own agricultural land.

Sex (sex) & Other Variables

☑ **Sex vs. Marital Status (marital_status): 0.172**

✓ **Weak positive correlation** → Gender has a **minor influence** on marital status.

✓ This could reflect gender differences in marriage patterns.

☑ **Sex vs. Agricultural Land Ownership (owns_agri_land): -0.023**

✓ **Near zero correlation** → Land ownership is **not strongly influenced by gender**.

Why Are These Relationships Happening?

✓ **Mobile money use is higher among younger individuals**, likely due to better tech adoption.

✓ **Older individuals are more likely to be married**, as expected from societal trends.

✓ **Agricultural land ownership has a minor connection to mobile money use**, possibly reflecting rural economic activity.