```
In [29]: import pandas as pd
         import numpy as np
         import seaborn as sns
         import matplotlib.pyplot as plt
         from lightgbm import LGBMClassifier
         from sklearn.model_selection import train_test_split
         from sklearn.metrics import accuracy score
         plt.rcParams["axes.labelsize"] = 18
         import warnings
         warnings.filterwarnings('ignore')
         # %matplotlib inline is used in Jupyter Notebooks to display plots inline
         %matplotlib inline
In [32]: train = pd.read_csv('Train.csv')
         test =pd.read csv('Test.csv')
         ss= pd.read_csv('SampleSubmission.csv')
         variables=pd.read_csv('VariableDefinitions.csv')
In [44]: print ('train data shape:',train.shape)
         print ('test data shape:',test.shape)
         print('ss data set shape:',ss.shape)
        train data shape: (23524, 13)
        test data shape: (10086, 12)
        ss data set shape: (33610, 2)
In [45]: #checking the information
         data.info
```

Out[45]:		d method			of	C	ountry	yea	r uniqueid bank_
	o O	nt locati Kenya	2018		ueid 1		Yes	=	Rural
	1	Kenya	2018		ueid 2		No		Rural
	2	Kenya	2018	•	ueid 3		Yes		Urban
	3	Kenya		•	ueid 4		No		Rural
	4	Kenya	2018	uniq	ueid_5		No)	Urban
									;
	23519	Uganda		uniquei	_		No		Rural
	23520	Uganda	2018	uniquei	_		No		Rural
	23521 23522	Uganda Uganda	2018 2018	uniquei uniquei	_		No No		Rural Urban
	23523	Uganda	2018	uniquei	_		No		Rural
		-							
	0	cellphon	_		ehold_si		age_of_	_res	pondent \
	0 1			es No		3 5			24 70
	2			es		5			26
	3			es		5			34
	4			No		8			26
	23519		Υ	es		4			48
	23520			es		2			27
	23521			es		5			27
	23522			es		7			30
	23523		Y	es	-	10			20
	c \	gender_o	f_resp	ondent r	elationsh	nip_	_with_he	ead	marital_stat
	us \ 0			Female			Spor	ISA	Married/Living togeth
	er			T Cilia CC			Spot	,,,,	narried, Living togeth
	1			Female	Head	of	Househo	old	Widow
	ed								
	2			Male	01	the	r relati	ive	Single/Never Marri
	ed 3			Female	Head	of	Househo	blo	Married/Living togeth
	er					•			a = ca, == : : : : : : : : : : : : : : : : : :
	4			Male			Chi	ild	Single/Never Marri
	ed								
	• • •								•
	23519 ed			Female	Head	of	Househo	old	Divorced/Seperat
	23520			Female	Head	of	Househo	old	Single/Never Marri
	ed 23521			Female			Pare	ent	Widow
	ed 23522			Female			Pare	ent	Divorced/Seperat
	ed 23523			Male			Chi	ild	Single/Never Marri
	ed								
	0 1 2 3 4	Vocatio	N	econdary o formal ecialise Primary	educatio	on on ng on	Formal	ly e	job_type Self employed nment Dependent Self employed mployed Private rmally employed
	23519		N	o formal	educatio	on			Other Income

```
23520
                       Secondary education
                                                  Other Income
                        Primary education
                                                  Other Income
       23521
       23522
                       Secondary education
                                                 Self employed
                       Secondary education
                                                     No Income
       23523
       [23524 rows x 13 columns]>
In [9]: data.columns
'gender_of_respondent', 'relationship_with_head', 'marital_statu
       sΊ,
             'education_level', 'job_type'],
            dtype='object')
In [10]: data.describe
```

	method			be of		country	year	uniqueid ban
к_acco 0	unt loca Kenya	tion_t; 2018		ueid 1		Yes		Rural
1	Kenya	2018		ueid 2		No		Rural
2	Kenya	2018		ueid_3		Yes		Urban
3	Kenya			ueid 4		No		Rural
4	Kenya	2018	unic	ueid_5		No		Urban
23519	Uganda	2018	uniquei	_		No		Rural
23520	Uganda	2018	uniquei	_		No		Rural
23521 23522	Uganda Uganda	2018 2018	uniquei	_		No No		Rural Urban
23523	Uganda	2018	uniquei uniquei	_		No		Rural
	o gamaa		9					
	cellphon	_		ehold_s	ize	age_of_r	espond	ent \
0			es		3			24
1			No		5			70
2 3			es		5 5			26 34
4			es No		8			26
23519			es		4			48
23520		Υ	es		2			27
23521		Υ	es		5			27
23522		Υ	es		7			30
23523		Y	es		10			20
us \	gender_o	f_resp	ondent r	elation	ship_	_with_hea	d	marital_stat
os (0 er			Female			Spous	e Mar	ried/Living togeth
1 ed			Female	Hea	d of	Househol	d	Widow
2 ed			Male		0the	r relativ	е	Single/Never Marri
3			Female	Неа	d of	Househol	d Mar	ried/Living togeth
er 4			Male			Chil	d	Single/Never Marri
ed			riacc			CHILC	u	Single/Never Harri
 23519			Female	Hea	d of	Househol	d	Divorced/Seperat
ed 23520			Female	Hea	d of	Househol	d	Single/Never Marri
ed			Гот-1 -			D		
23521 ed			Female			Paren	τ	Widow
23522 ed			Female			Paren	t	Divorced/Seperat
23523			Male			Chil	d	Single/Never Marri
ed								
0 1 2	Vocatio	N	educa econdary o formal ecialise	educat	ion ion	Gov	ernmen	job_type lf employed t Dependent lf employed
3 4	2 2 2 2 2 2	-, -,	Primary	educat educat	ion	-	emplo	yed Private ly employed
23519		N	o formal					ther Income

23520	Secondary education	Other Income
23521	Primary education	Other Income
23522	Secondary education	Self employed
23523	Secondary education	No Income

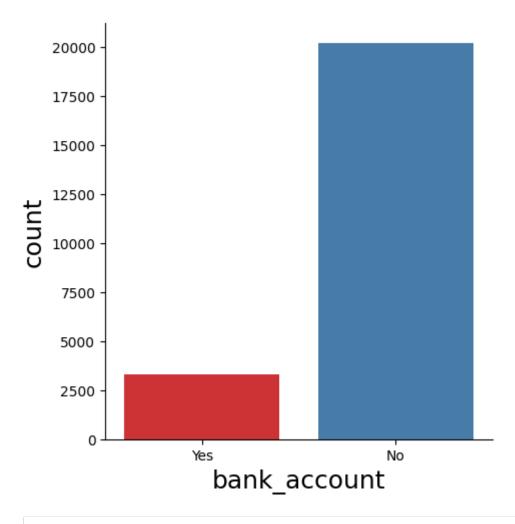
[23524 rows x 13 columns]>

In [11]: data.tail

Out[11]:		l method location		e.tail d	f	coui	ntry	year	ur	niqueid	bank_ac
	0	Kenya	2018		ueid_1			Yes	Ru	ural	
	1	Kenya	2018		ueid_2			No		ural	
	2	Kenya	2018		ueid_3			Yes		rban	
	3	-	2018		ueid_4			No		ıral	
	4	Kenya	2018	unic	ueid_5			No	Uı	rban	
	23519	Uganda	2018	uniquei	_			No		ıral	
	23520	Uganda	2018	uniquei	_			No		ıral 	
	23521	Uganda	2018	uniquei	_			No		ıral	
	23522 23523	Uganda Uganda	2018 2018	uniquei uniquei	_			No No		rban ural	
	23323	oganua	2010	unitquei	.u_2117			NO	IXC	arac	
	0	cellphon	_	ss hous es	ehold_s:	ize 3	age_	of_res	pondent 24	\	
	1			No		5			70		
	2		Υ	es		5			26		
	3		Υ	es		5			34		
	4			No		8			26		
	23519		Υ	es		4			48		
	23520		Υ	es		2			27		
	23521			es		5			27		
	23522			es		7			30		
	23523		Υ	es		10			20		
	us \	gender_o	f_resp	ondent r	elations	ship_	_with	_head		marit	al_stat
	0 er			Female			S	pouse	Married	d/Living	g togeth
	1 ed			Female	Head	d of	Hous	ehold			Widow
	2 ed			Male	(the	r rel	ative	Sing	gle/Neve	er Marri
	3 er			Female	Head	d of	Hous	ehold	Married	d/Living	g togeth
	4 ed			Male				Child	Sino	gle/Neve	er Marri
	23519 ed			Female	Неас	d of	Hous	ehold	Di	ivorced/	'Seperat
	23520 ed			Female	Неас	dof	Hous	ehold	Sino	gle/Neve	er Marri
	23521 ed			Female			Р	arent			Widow
	23522 ed			Female			Р	arent	Di	ivorced/	'Seperat
	23523 ed			Male				Child	Sing	gle/Neve	er Marri
	0 1 2 3 4	Vocatio	N	econdary o formal ecialise Primary	education_levelocation_levelocation_levelocation	ion ion ing ion	Form	ally e	Self e nment De	employed Private	1 - 1
	23519		N	o formal	. educat:	Lon			0the	r Income	9

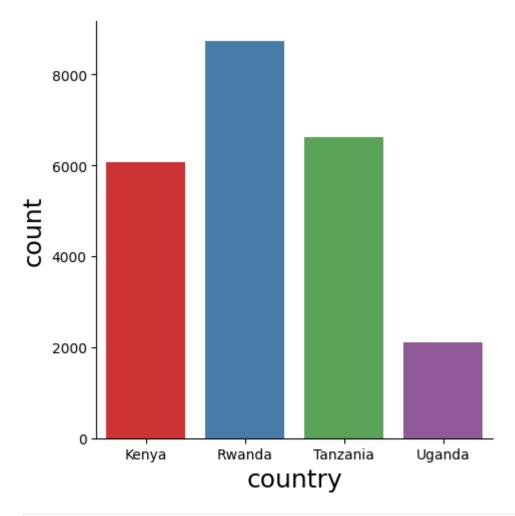
```
23520
                             Secondary education
                                                                Other Income
                                                                Other Income
                               Primary education
          23521
          23522
                             Secondary education
                                                               Self employed
                                                                   No Income
          23523
                             Secondary education
          [23524 rows x 13 columns]>
In [48]: #printing the misssing the missig values
         print("showing the missing values",train.isnull().sum())
         print("showing the null values",test.isnull().sum())
        showing the missing values country
        year
        uniqueid
                                   0
                                   0
        bank account
        location type
                                   0
        cellphone access
                                   0
        household size
                                   0
        age of respondent
                                   0
        gender of respondent
                                   0
        relationship with head
                                   0
                                   0
        marital status
        education level
                                   0
                                   0
        job type
        dtype: int64
        showing the null values country
                                                           0
                                   0
        uniqueid
        location type
                                   0
        cellphone access
                                   0
        household size
        age of respondent
        gender of respondent
                                   0
        relationship with head
                                   0
        marital status
        education level
                                   0
                                   0
        job type
        dtype: int64
In [49]: sns.catplot(x="bank account", kind="count", data=train, palette="Set1")
```

Out[49]: <seaborn.axisgrid.FacetGrid at 0x7ec5411b32b0>



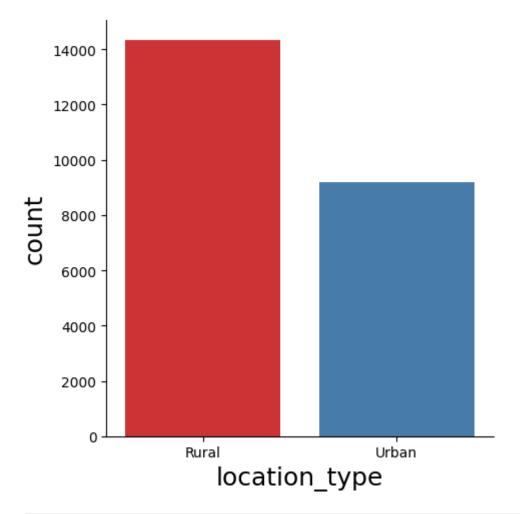
In [51]:	ss.head()		
Out[51]:	unique_id	bank_account	
	0 uniqueid_1 x Kenya	0	
	1 uniqueid_2 x Kenya	0	
	2 uniqueid_3 x Kenya	0	
	3 uniqueid_4 x Kenya	0	
	4 uniqueid_5 x Kenya	0	
In [52]:	<pre>print (train.info())</pre>		

```
<class 'pandas.core.frame.DataFrame'>
        RangeIndex: 23524 entries, 0 to 23523
        Data columns (total 13 columns):
                                     Non-Null Count Dtype
             Column
            ----
        - - -
                                     -----
         0
             country
                                     23524 non-null object
         1
                                     23524 non-null int64
             year
         2
             uniqueid
                                     23524 non-null object
         3
             bank_account
                                     23524 non-null object
         4
             location_type
                                     23524 non-null object
         5
             cellphone access
                                     23524 non-null object
         6
             household size
                                     23524 non-null int64
         7
             age_of_respondent
                                     23524 non-null int64
         8
             gender of respondent
                                     23524 non-null object
             relationship with head 23524 non-null object
         9
         10 marital_status
                                     23524 non-null object
         11 education level
                                     23524 non-null object
            job_type
         12
                                     23524 non-null object
        dtypes: int64(3), object(10)
        memory usage: 2.3+ MB
        None
In [54]: variables.T
Out[54]:
                                                    2
                                                                   3
                                0
                                        1
            Variable
                                             uniqueid location_type cellphone_acc
                          country
                                     year
         Definitions
                                     Year
                                               Unique
                                                                        If intervie
                                                             Type of
                          Country survey
                                           identifier
                                                                       has access t
          Unnamed: 1 interviewee
                                     was
                                                           location:
                                                                       cellphone: Y
                                             for each
                           is in.
                                     done
                                                        Rural, Urban
                                          interviewee
                                      in.
        train['bank account'].value_counts()
In [55]:
Out[55]: bank account
         No
                20212
                 3312
         Yes
         Name: count, dtype: int64
In [56]: | sns.catplot(x="country", kind="count", data=train, palette="Set1")
Out[56]: <seaborn.axisgrid.FacetGrid at 0x7ec53d141ff0>
```



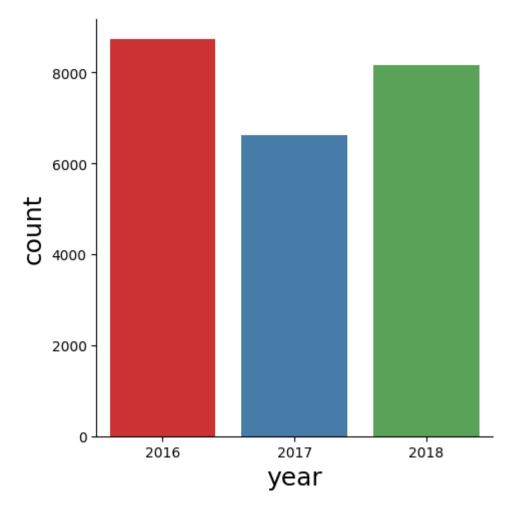
In [57]: sns.catplot(x="location_type", kind="count", data=train, palette="Set1")

Out[57]: <seaborn.axisgrid.FacetGrid at 0x7ec54090f4f0>



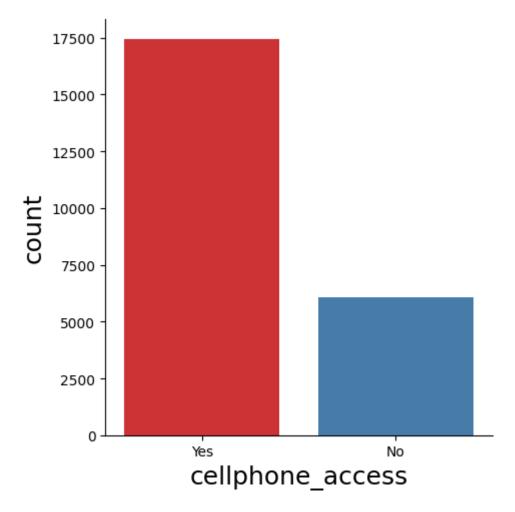
In [58]: sns.catplot(x="year", kind="count", data=train, palette="Set1")

Out[58]: <seaborn.axisgrid.FacetGrid at 0x7ec53d188670>



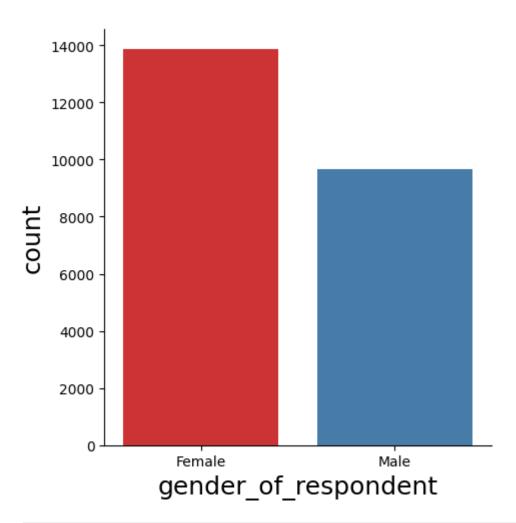
In [59]: sns.catplot(x="cellphone_access", kind="count", data=train, palette="Set1

Out[59]: <seaborn.axisgrid.FacetGrid at 0x7ec53d00e9e0>



In [64]: sns.catplot(x="gender_of_respondent", kind="count", data=train, palette="

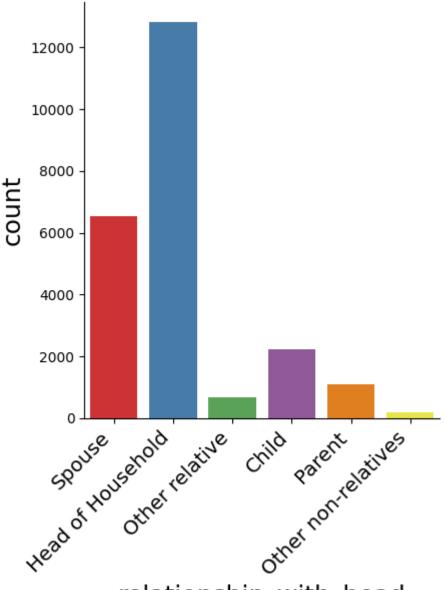
Out[64]: <seaborn.axisgrid.FacetGrid at 0x7ec53cdd7b50>



```
In [67]: sns.catplot(x="relationship_with_head", kind="count", data=train, palette

plt.xticks(
    rotation=45,
    horizontalalignment ='right',
    fontweight='light',
    fontsize = 'x-large'
)

Out[67]: ([0, 1, 2, 3, 4, 5],
    [Text(0, 0, 'Spouse'),
    Text(1, 0, 'Head of Household'),
    Text(2, 0, 'Other relative'),
    Text(3, 0, 'Child'),
    Text(4, 0, 'Parent'),
    Text(5, 0, 'Other non-relatives')])
```

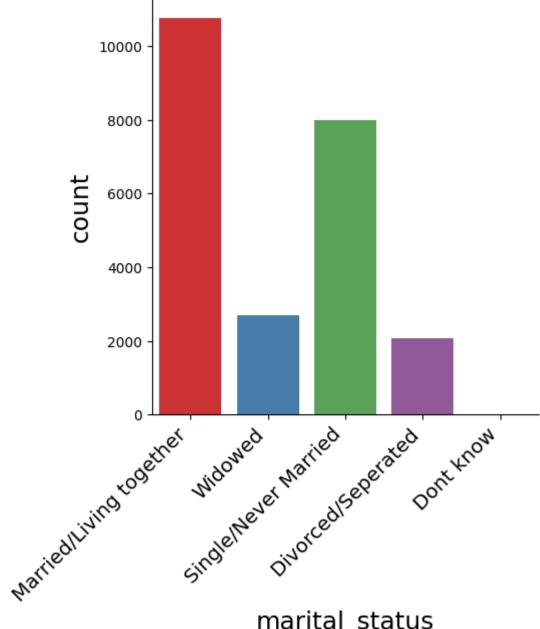


relationship_with_head

```
In [68]: sns.catplot(x="marital_status", kind="count", data=train, palette="Set1")

plt.xticks(
    rotation=45,
    horizontalalignment ='right',
    fontweight='light',
    fontsize = 'x-large'
)

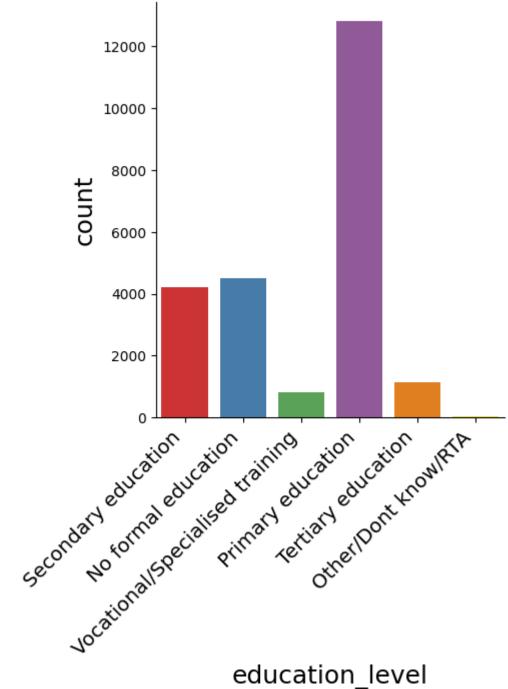
Out[68]: ([0, 1, 2, 3, 4],
    [Text(0, 0, 'Married/Living together'),
    Text(1, 0, 'Widowed'),
    Text(2, 0, 'Single/Never Married'),
    Text(3, 0, 'Divorced/Seperated'),
    Text(4, 0, 'Dont know')])
```



marital_status

```
In [69]: sns.catplot(x="education_level", kind="count", data=train, palette="Set1"
         plt.xticks(
             rotation=45,
             horizontalalignment = 'right',
             fontweight='light',
             fontsize = 'x-large'
Out[69]: ([0, 1, 2, 3, 4, 5],
           [Text(0, 0, 'Secondary education'),
            Text(1, 0, 'No formal education'),
           Text(2, 0, 'Vocational/Specialised training'),
           Text(3, 0, 'Primary education'),
            Text(4, 0, 'Tertiary education'),
            Text(5, 0, 'Other/Dont know/RTA')])
```

7/23/24, 12:13 16 of 20



education_level

```
In [70]: sns.catplot(x="job type", kind="count", data=train, palette="Set1")
         plt.xticks(
             rotation=45,
             horizontalalignment = 'right',
             fontweight='light',
             fontsize = 'x-large'
         )
```

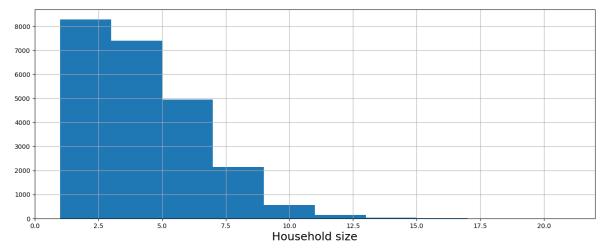
7/23/24, 12:13 17 of 20

```
Out[70]: ([0, 1, 2, 3, 4, 5, 6, 7, 8, 9],
                                                  [Text(0, 0, 'Self employed'),
                                                       Text(1, 0, 'Government Dependent'),
                                                       Text(2, 0, 'Formally employed Private'),
                                                       Text(3, 0, 'Informally employed'),
                                                       Text(4, 0, 'Formally employed Government'),
                                                      Text(5, 0, 'Farming and Fishing'),
                                                       Text(6, 0, 'Remittance Dependent'),
                                                       Text(7, 0, 'Other Income'),
                                                       Text(8, 0, 'Dont Know/Refuse to answer'),
                                                       Text(9, 0, 'No Income')])
                                                                                             6000
                                                                                             5000
                                                                                             4000
                                                                                            3000
                                                                                             2000
                                   Government Dependent are employed private of and rish of the private of the priva
                                                                                                                                                    Remittance Dependent
                                                                                                                                                                     Don't Know Refuse to answer
                                                                                                                                           Wed Juvening and Fishing,
                                                                                                                                                                                                                                                                          NO Income
```

job_type

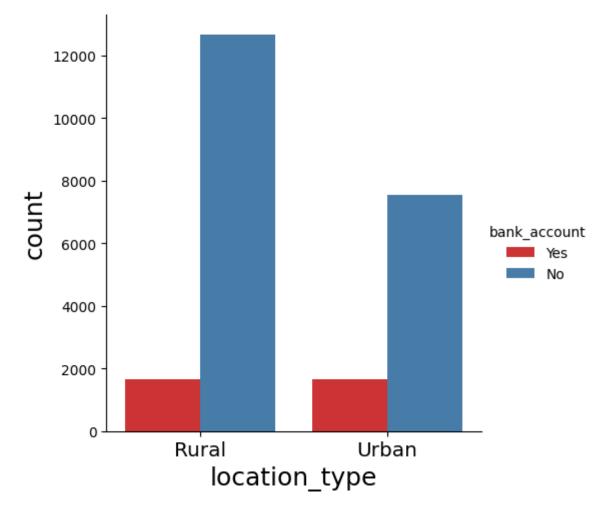
```
In [71]: #16 is for width and y for height
         plt.figure(figsize=(16,6))
         train.household size.hist()
         plt.xlabel('Household size')
```

```
Out[71]: Text(0.5, 0, 'Household size')
```



```
In [75]: sns.catplot(x="location_type", hue = "bank_account", kind="count", data=
    plt.xticks(
        fontweight='light',
        fontsize = 'x-large'
)
```

Out[75]: ([0, 1], [Text(0, 0, 'Rural'), Text(1, 0, 'Urban')])



In [89]: #import preprocessing module
from sklearn.preprocessing import LabelEncoder
from sklearn .preprocessing import MinMaxScaler

```
#convert target labael to numerical ddata
        le =LabelEncoder()
        train['bank_account'] = le.fit_transform(train['bank_account'])
        #separated training features from targetabs
        x_train=train.drop(['bank_account'],axis=1)
        y_train=train['bank_account']
        print(y_train)
       0
                1
       1
                0
       2
                1
       3
                0
       4
                0
       23519
               0
       23520
               0
       23521
               0
       23522
               0
       23523
               0
       Name: bank_account, Length: 23524, dtype: int64
In [ ]:
```