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]: <bound method DataFrame.info of       country  year      uniqueid bank_
         account location_type  \
         0       Kenya  2018     uniqueid_1          Yes         Rural
         1       Kenya  2018     uniqueid_2           No         Rural
         2       Kenya  2018     uniqueid_3          Yes         Urban
         3       Kenya  2018     uniqueid_4           No         Rural
         4       Kenya  2018     uniqueid_5           No         Urban
         ...       ...   ...            ...          ...           ...
         23519  Uganda  2018  uniqueid_2113           No         Rural
         23520  Uganda  2018  uniqueid_2114           No         Rural
         23521  Uganda  2018  uniqueid_2115           No         Rural
         23522  Uganda  2018  uniqueid_2116           No         Urban
         23523  Uganda  2018  uniqueid_2117           No         Rural

               cellphone_access  household_size  age_of_respondent  \
         0                  Yes               3                 24
         1                   No               5                 70
         2                  Yes               5                 26
         3                  Yes               5                 34
         4                   No               8                 26
         ...                ...             ...                ...
         23519              Yes               4                 48
         23520              Yes               2                 27
         23521              Yes               5                 27
         23522              Yes               7                 30
         23523              Yes              10                 20

               gender_of_respondent relationship_with_head           marital_stat
         us  \
         0                    Female                 Spouse  Married/Living togeth
         er
         1                    Female      Head of Household                  Widow
         ed
         2                      Male         Other relative     Single/Never Marri
         ed
         3                    Female      Head of Household  Married/Living togeth
         er
         4                      Male                  Child     Single/Never Marri
         ed
         ...                    ...                    ...                      .
         ..
         23519                Female      Head of Household       Divorced/Seperat
         ed
         23520                Female      Head of Household     Single/Never Marri
         ed
         23521                Female                 Parent                  Widow
         ed
         23522                Female                 Parent       Divorced/Seperat
         ed
         23523                  Male                  Child     Single/Never Marri
         ed

                                education_level                job_type
         0                  Secondary education           Self employed
         1                  No formal education    Government Dependent
         2      Vocational/Specialised training           Self employed
         3                    Primary education  Formally employed Private
         4                    Primary education     Informally employed
         ...                                ...                     ...
         23519              No formal education            Other 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 [9]: data.columns
```

```
Out[9]: Index(['country', 'year', 'uniqueid', 'bank_account', 'location_type',
               'cellphone_access', 'household_size', 'age_of_respondent',
               'gender_of_respondent', 'relationship_with_head', 'marital_statu
        s',
               'education_level', 'job_type'],
              dtype='object')
```

```
In [10]: data.describe
```

Out[10]:

```
<bound method NDFrame.describe of       country  year      uniqueid ban
k_account location_type  \
0       Kenya  2018     uniqueid_1          Yes         Rural
1       Kenya  2018     uniqueid_2           No         Rural
2       Kenya  2018     uniqueid_3          Yes         Urban
3       Kenya  2018     uniqueid_4           No         Rural
4       Kenya  2018     uniqueid_5           No         Urban
...       ...   ...            ...          ...           ...
23519  Uganda  2018  uniqueid_2113           No         Rural
23520  Uganda  2018  uniqueid_2114           No         Rural
23521  Uganda  2018  uniqueid_2115           No         Rural
23522  Uganda  2018  uniqueid_2116           No         Urban
23523  Uganda  2018  uniqueid_2117           No         Rural

      cellphone_access  household_size  age_of_respondent  \
0                  Yes               3                 24
1                   No               5                 70
2                  Yes               5                 26
3                  Yes               5                 34
4                   No               8                 26
...                ...             ...                ...
23519              Yes               4                 48
23520              Yes               2                 27
23521              Yes               5                 27
23522              Yes               7                 30
23523              Yes              10                 20

      gender_of_respondent relationship_with_head           marital_stat
us  \
0                   Female                 Spouse  Married/Living togeth
er
1                   Female      Head of Household                  Widow
ed
2                     Male         Other relative     Single/Never Marri
ed
3                   Female      Head of Household  Married/Living togeth
er
4                     Male                  Child     Single/Never Marri
ed
...                    ...                    ...                      .
..
23519               Female      Head of Household       Divorced/Seperat
ed
23520               Female      Head of Household     Single/Never Marri
ed
23521               Female                 Parent                  Widow
ed
23522               Female                 Parent       Divorced/Seperat
ed
23523                 Male                  Child     Single/Never Marri
ed

                       education_level                   job_type
0                  Secondary education              Self employed
1                  No formal education       Government Dependent
2      Vocational/Specialised training              Self employed
3                    Primary education  Formally employed Private
4                    Primary education        Informally employed
...                                ...                        ...
23519              No formal education               Other 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]:

```
<bound method NDFrame.tail of       country  year       uniqueid bank_ac
count location_type  \
0       Kenya  2018     uniqueid_1          Yes         Rural
1       Kenya  2018     uniqueid_2           No         Rural
2       Kenya  2018     uniqueid_3          Yes         Urban
3       Kenya  2018     uniqueid_4           No         Rural
4       Kenya  2018     uniqueid_5           No         Urban
...       ...   ...            ...          ...           ...
23519  Uganda  2018  uniqueid_2113           No         Rural
23520  Uganda  2018  uniqueid_2114           No         Rural
23521  Uganda  2018  uniqueid_2115           No         Rural
23522  Uganda  2018  uniqueid_2116           No         Urban
23523  Uganda  2018  uniqueid_2117           No         Rural

      cellphone_access  household_size  age_of_respondent  \
0                  Yes               3                 24
1                   No               5                 70
2                  Yes               5                 26
3                  Yes               5                 34
4                   No               8                 26
...                ...             ...                ...
23519              Yes               4                 48
23520              Yes               2                 27
23521              Yes               5                 27
23522              Yes               7                 30
23523              Yes              10                 20

      gender_of_respondent relationship_with_head           marital_stat
us  \
0                   Female                 Spouse  Married/Living togeth
er
1                   Female      Head of Household                  Widow
ed
2                     Male         Other relative     Single/Never Marri
ed
3                   Female      Head of Household  Married/Living togeth
er
4                     Male                  Child     Single/Never Marri
ed
...                    ...                    ...                      .
..
23519               Female      Head of Household       Divorced/Seperat
ed
23520               Female      Head of Household     Single/Never Marri
ed
23521               Female                 Parent                  Widow
ed
23522               Female                 Parent       Divorced/Seperat
ed
23523                 Male                  Child     Single/Never Marri
ed

                       education_level                   job_type
0                  Secondary education              Self employed
1                  No formal education       Government Dependent
2      Vocational/Specialised training              Self employed
3                    Primary education  Formally employed Private
4                    Primary education        Informally employed
...                                ...                        ...
23519              No formal education               Other 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 [48]:

```
#printing the miisssing the missig values
print("showing the missing values",train.isnull().sum())
print("showing the null values",test.isnull().sum())
```

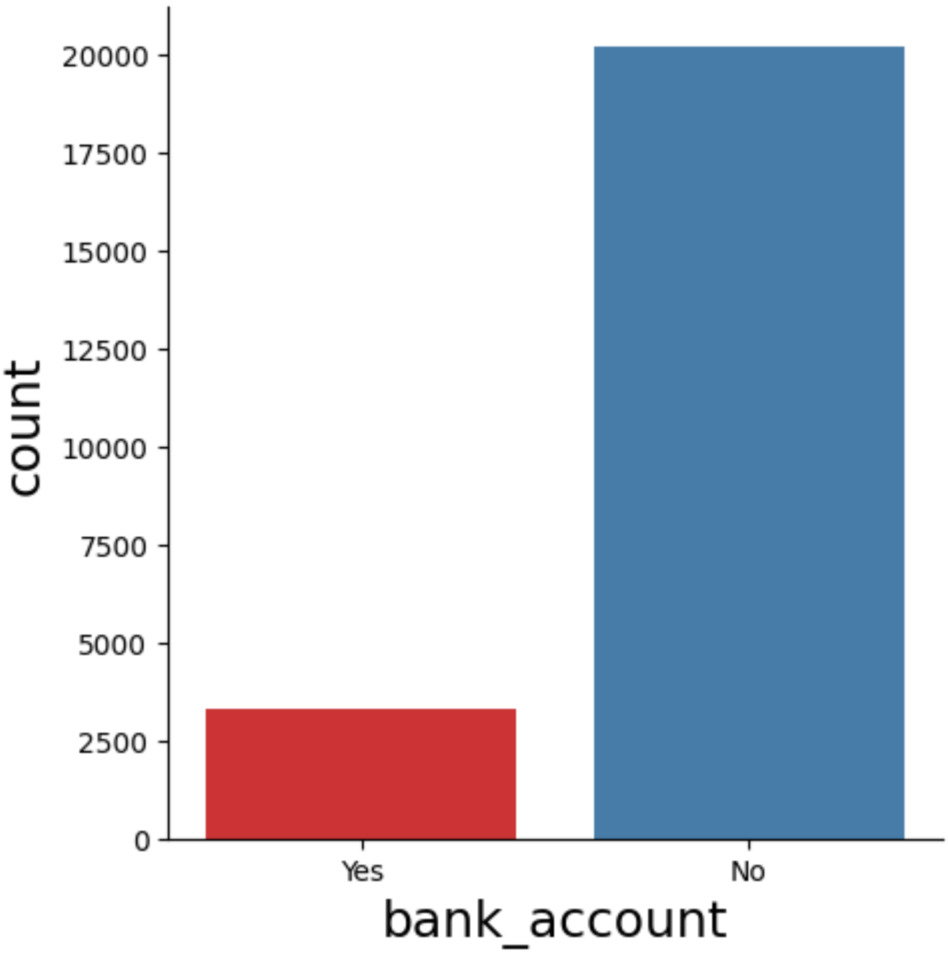
```
showing the missing values country                  0
year                      0
uniqueid                  0
bank_account              0
location_type             0
cellphone_access          0
household_size            0
age_of_respondent         0
gender_of_respondent      0
relationship_with_head    0
marital_status            0
education_level           0
job_type                  0
dtype: int64
showing the null values country                  0
year                      0
uniqueid                  0
location_type             0
cellphone_access          0
household_size            0
age_of_respondent         0
gender_of_respondent      0
relationship_with_head    0
marital_status            0
education_level           0
job_type                  0
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]:
```
print (train.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23524 entries, 0 to 23523
Data columns (total 13 columns):
 #   Column                  Non-Null Count  Dtype 
---  ------                  --------------  ----- 
 0   country                 23524 non-null  object
 1   year                    23524 non-null  int64 
 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
 9   relationship_with_head  23524 non-null  object
 10  marital_status          23524 non-null  object
 11  education_level         23524 non-null  object
 12  job_type                23524 non-null  object
dtypes: int64(3), object(10)
memory usage: 2.3+ MB
None
```

In [54]:

```
variables.T
```

Out[54]:

	0	1	2	3	
Variable Definitions	country	year	uniqueid	location_type	cellphone_acc
Unnamed: 1	Country interviewee is in.	Year survey was done in.	Unique identifier for each interviewee	Type of location: Rural, Urban	If intervie has access t cellphone: Y

In [55]:

```
train['bank_account'].value_counts()
```

Out[55]:

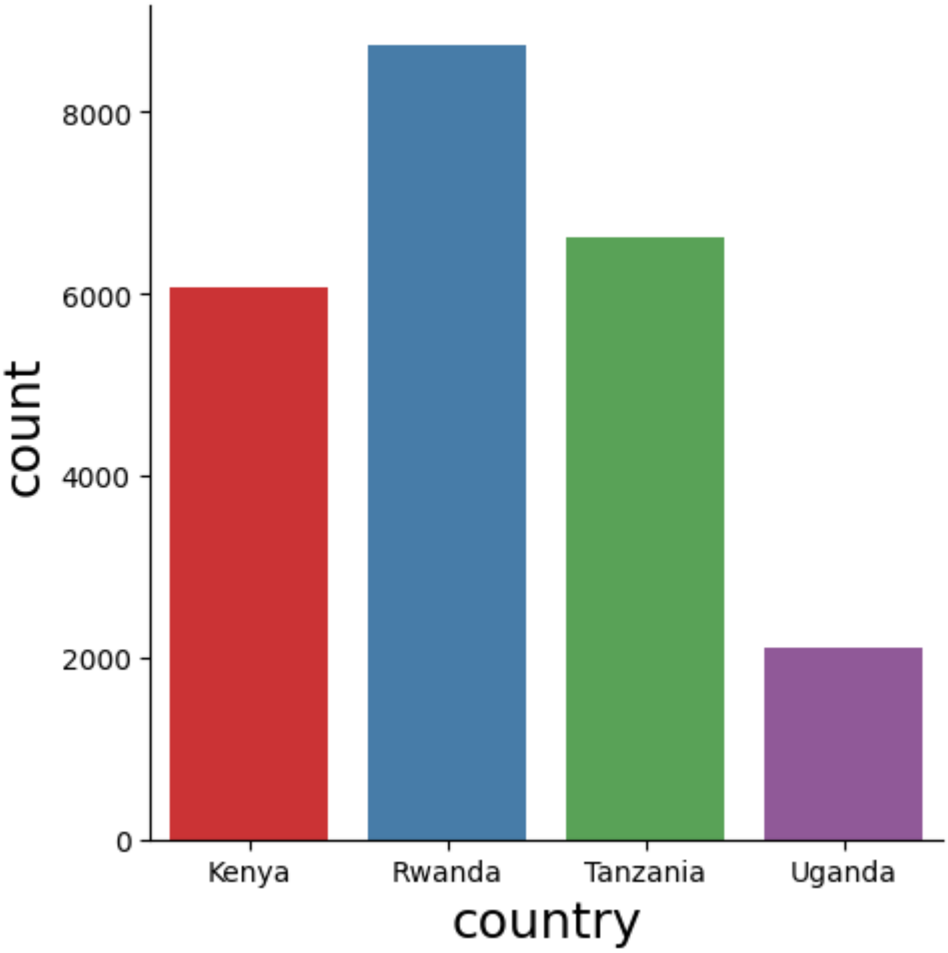
```
bank_account
No     20212
Yes     3312
Name: count, dtype: int64
```

In [56]:

```
sns.catplot(x="country", kind="count", data=train, palette="Set1")
```

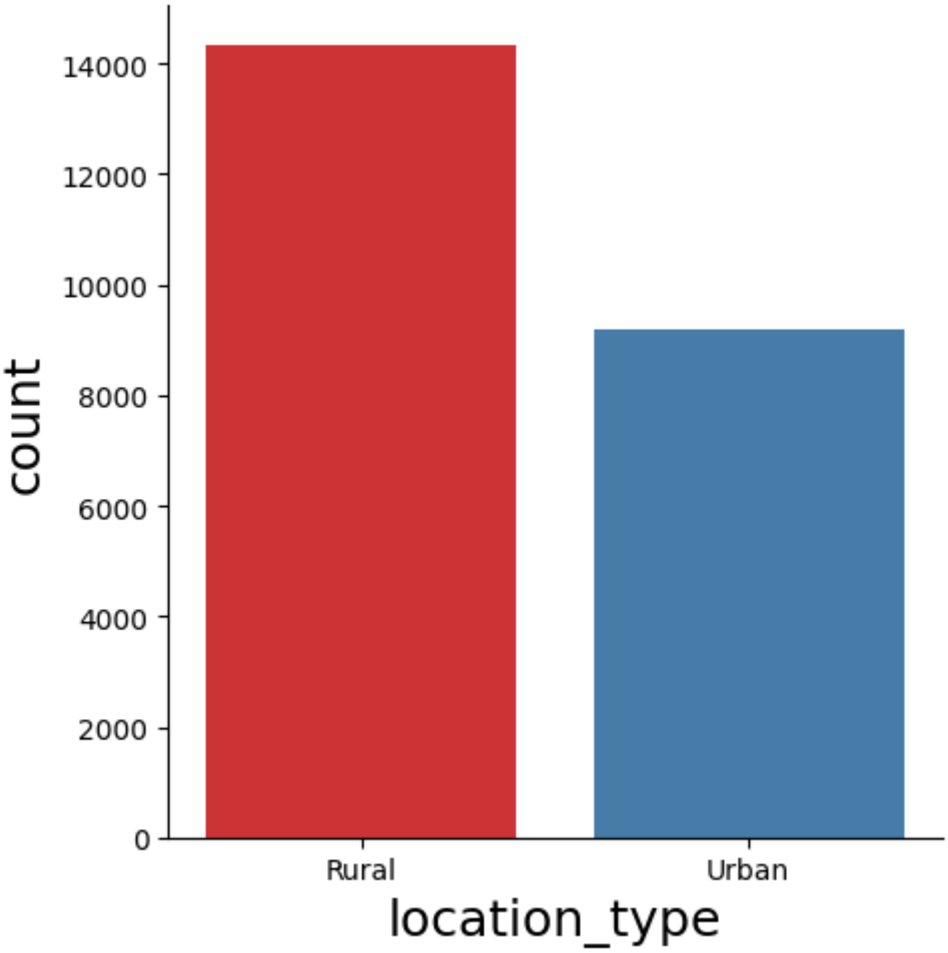
Out[56]:

```
<seaborn.axisgrid.FacetGrid at 0x7ec53d141ff0>
```

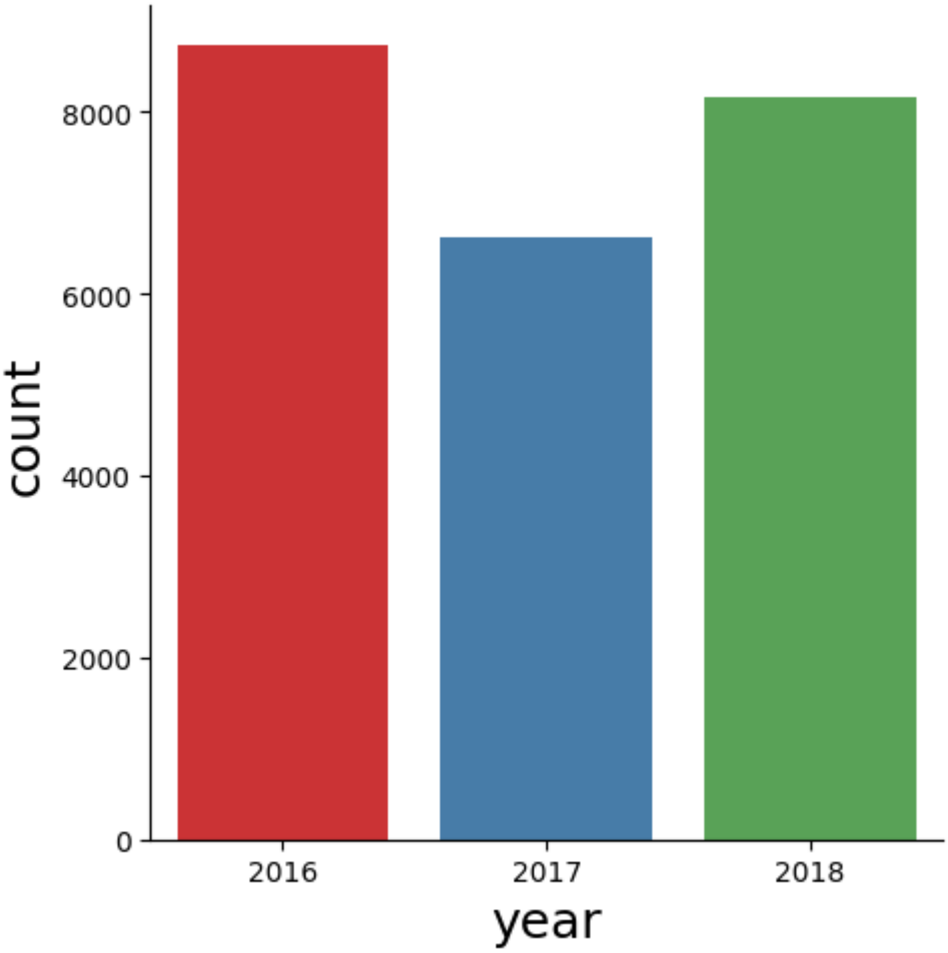


```
In [57]: sns.catplot(x="location_type", kind="count", data=train, palette="Set1")
```

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

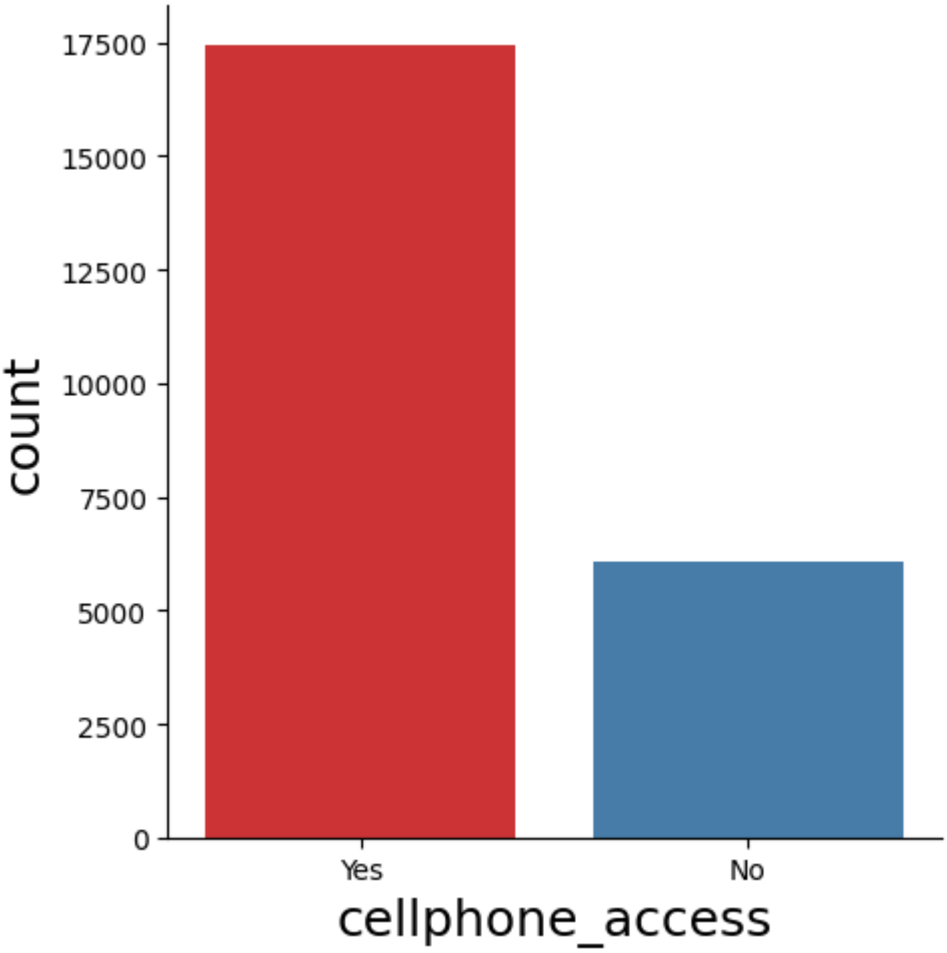


```
In [58]: sns.catplot(x="year", kind="count", data=train, palette="Set1")

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



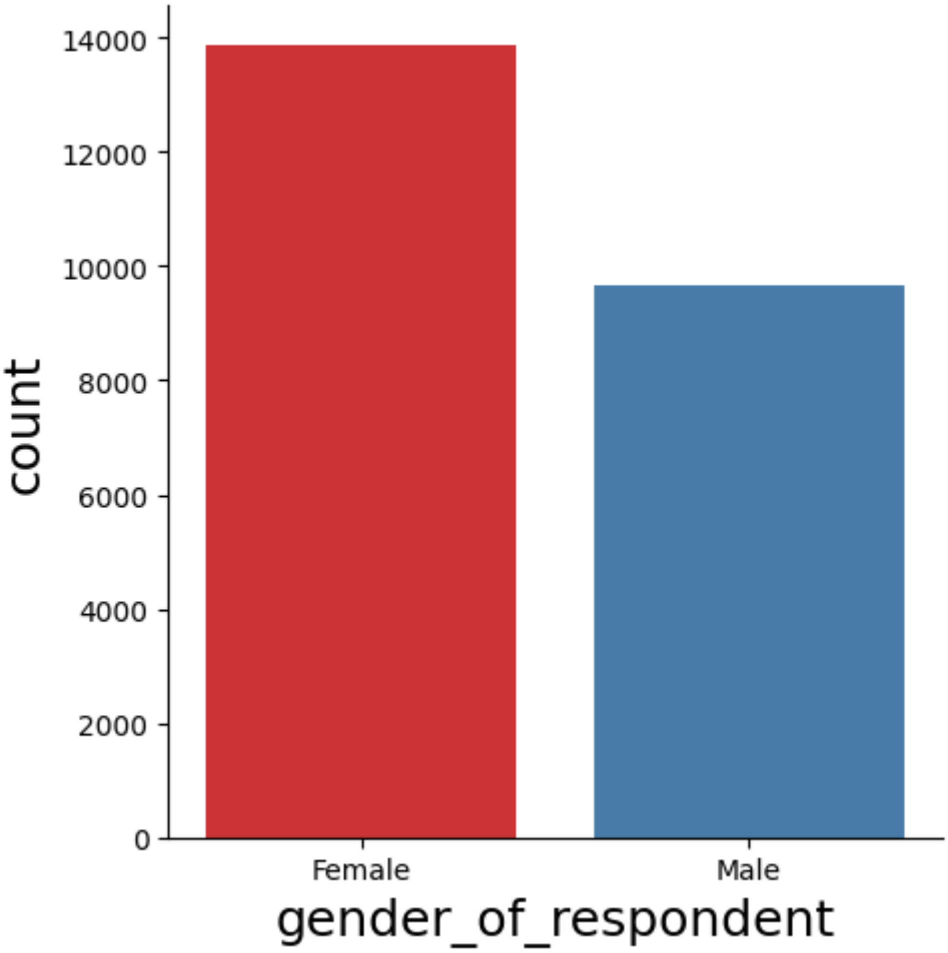
In [59]:
```
sns.catplot(x="cellphone_access", kind="count", data=train, palette="Set1
```

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



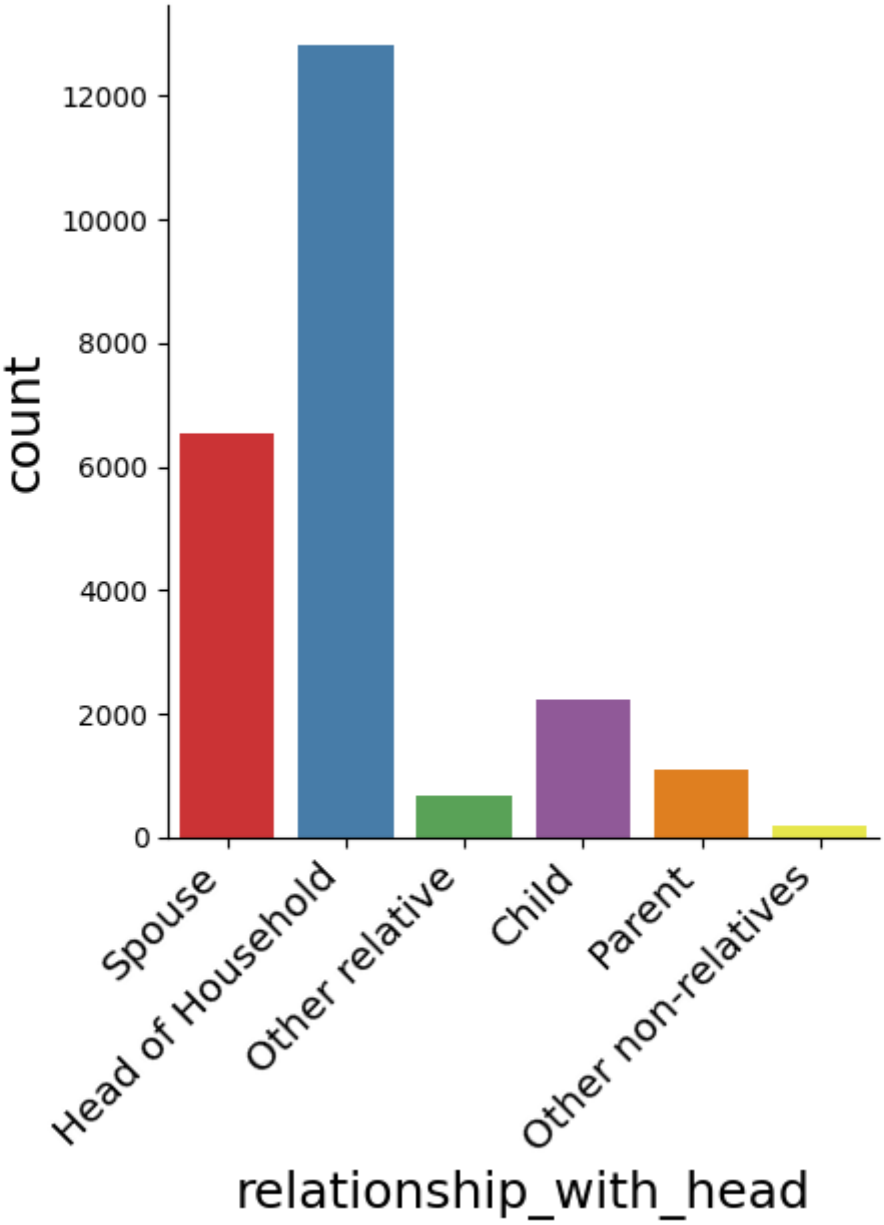
In [64]:
```
sns.catplot(x="gender_of_respondent", kind="count", data=train, palette="
```

Out[64]:
```
<seaboorn.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')])
```

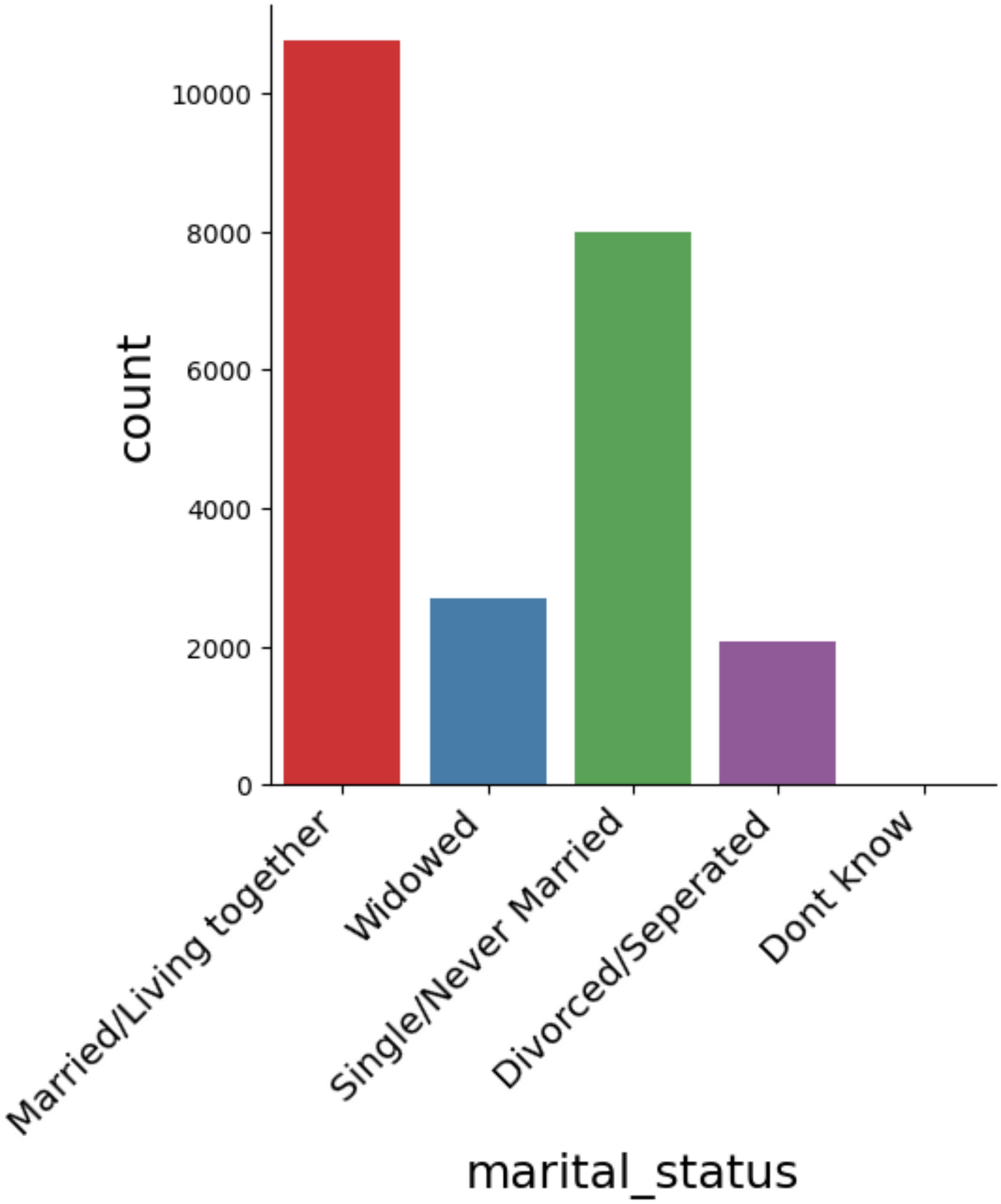


In [68]:

```
sns.catplot(x="marital_status", kind="count", data=train, palette="Set1")

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

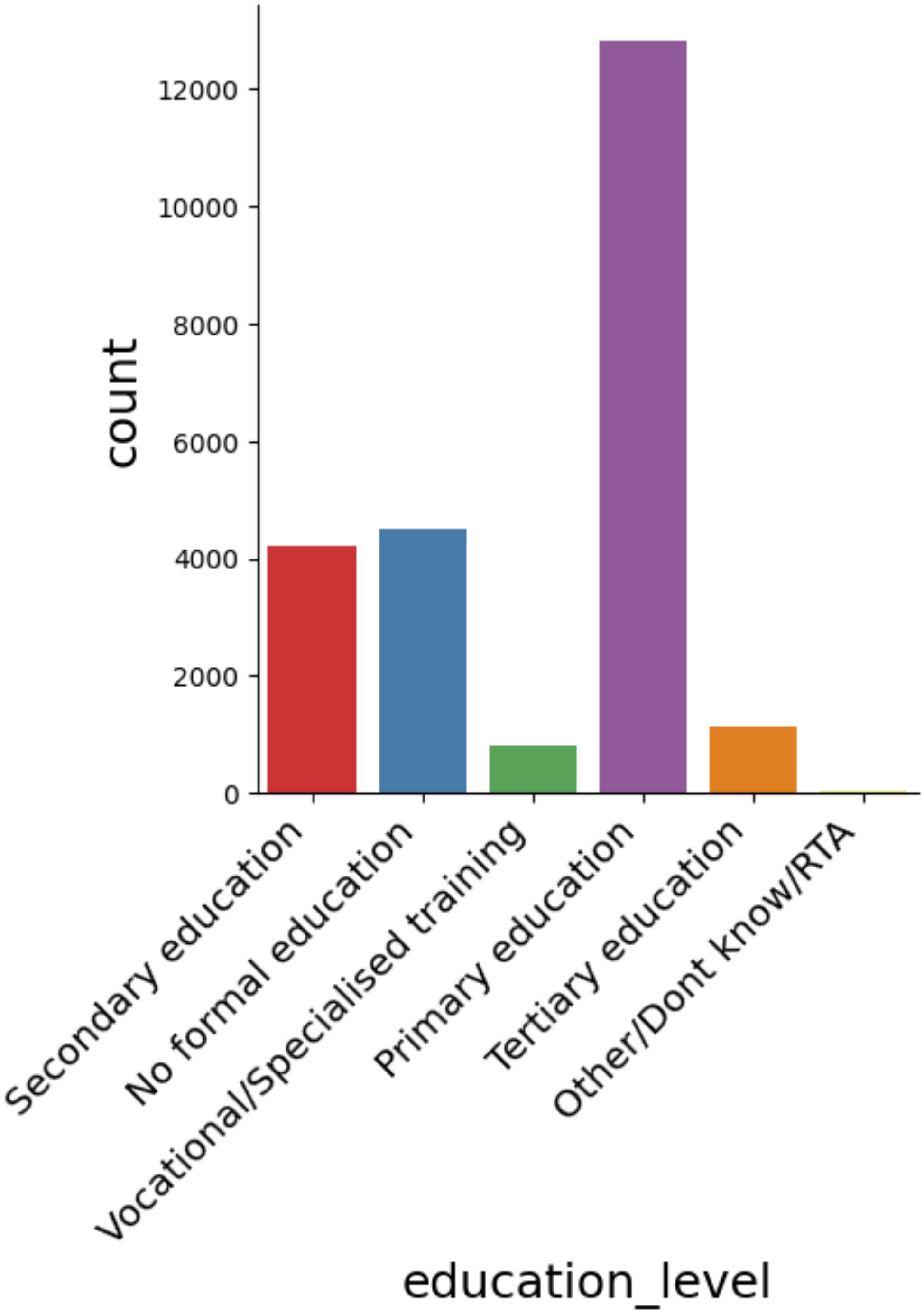
Out[68]:

```
([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')])
```



```
In [69]: sns.catplot(x="education_level", kind="count", data=train, palette="Set1"

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

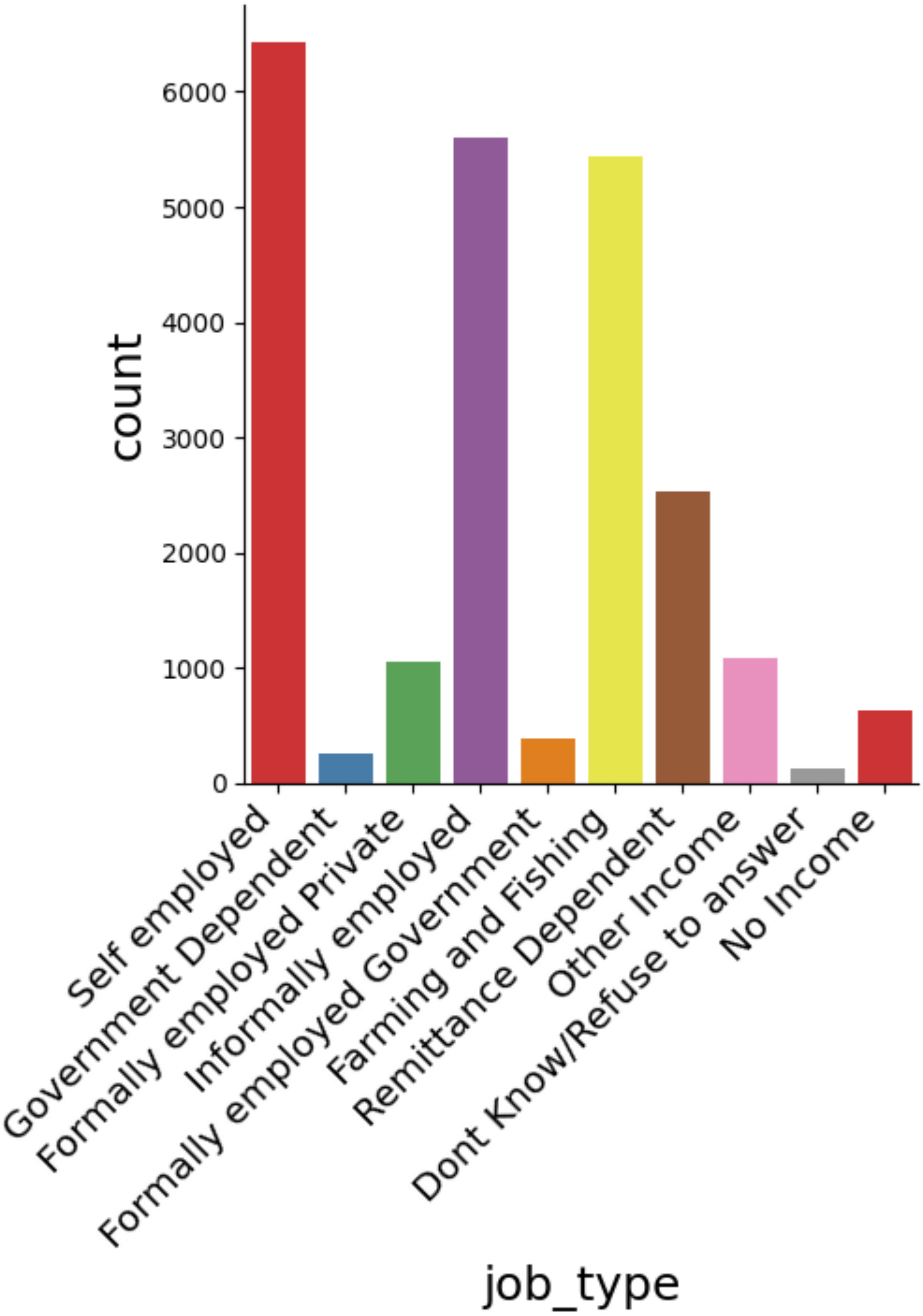
```
Out[69]: ([0, 1, 2, 3, 4, 5],
          [Text(0, 0, 'Secondary education'),
           Text(1, 0, 'No formal education'),
           Text(2, 0, 'Vocational/Specialised training'),
           Text(3, 0, 'Primary education'),
           Text(4, 0, 'Tertiary education'),
           Text(5, 0, 'Other/Dont know/RTA')])
```



In [70]:

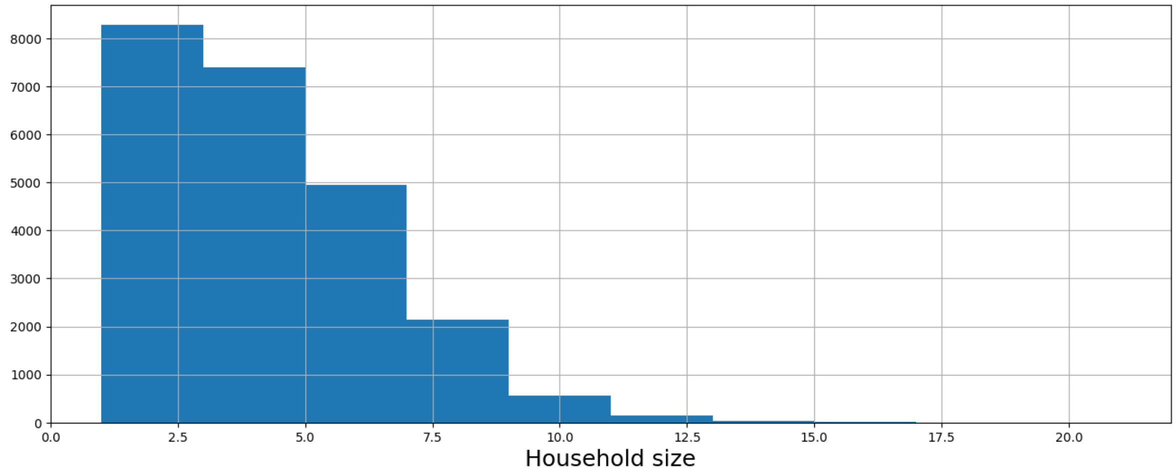
```
sns.catplot(x="job_type", kind="count", data=train, palette="Set1")

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

```
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')])
```



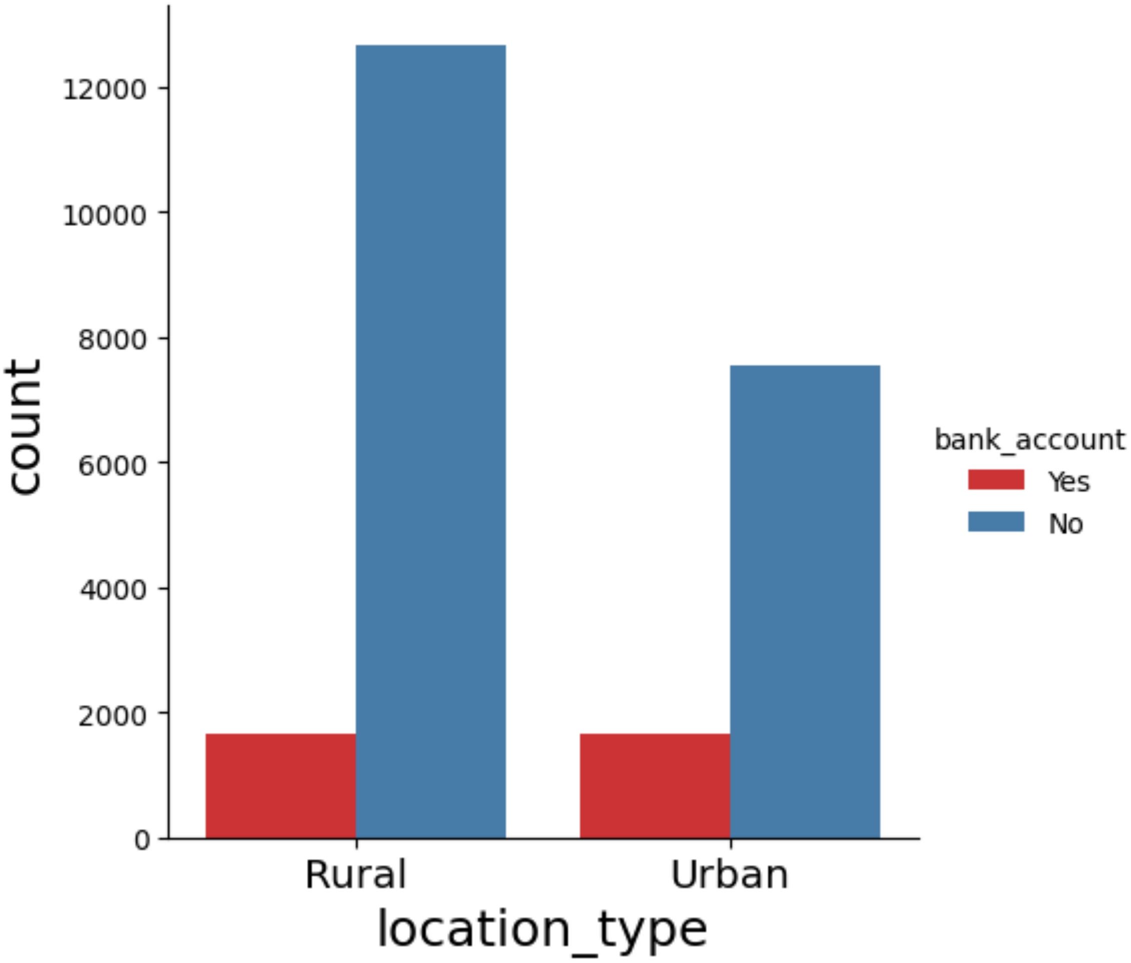
```
In [71]: #16 is for width and y for height
         plt.figure(figsize=(16,6))
         train.household_size.hist()
         plt.xlabel('Household size')
```

```
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 [ ]: