jupyter _EDA_Loan Last Checkpoint: 6 days ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[1]:
```
# lib for extraction ,manipulation,analysis
import numpy as np
import pandas as pd
# for visualtion
import matplotlib.pyplot as plt
import seaborn as sns
# for stats
import scipy.stats
from scipy.stats import shapiro, chi2, normaltest, kstest, zscore
# train test split
from sklearn.model_selection import train_test_split

# Logistic regression
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score,classification_report,multilabel_confusion_matrix
# for vif
from statsmodels.stats.outliers_influence import variance_inflation_factor
# regression evaluation metrics
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

[2]:
```
# importing dataset
data=pd.read_csv(r"C:\Users\deshm\OneDrive\Desktop\Python\loan_approval_dataset.csv")
data
```

[2]:

	loan_id	no_of_dependents	education	self_employed	income_annum	loan_amount	loan_term	cibil_score	residential_assets_value	commercial_assets_value	luxury_
0	1	2	Graduate	No	9600000	29900000	12	778	2400000	17600000	
1	2	0	Not Graduate	Yes	4100000	12200000	8	417	2700000	2200000	
2	3	3	Graduate	No	9100000	29700000	20	506	7100000	4500000	
3	4	3	Graduate	No	8200000	30700000	8	467	18200000	3300000	
4	5	5	Not Graduate	Yes	9800000	24200000	20	382	12400000	8200000	
...	...	...	...	...	...	...	...	...	...	...	
4264	4265	5	Graduate	Yes	1000000	2300000	12	317	2800000	500000	
4265	4266	0	Not Graduate	Yes	3300000	11300000	20	559	4200000	2900000	
4266	4267	2	Not Graduate	No	6500000	23900000	18	457	1200000	12400000	
4267	4268	1	Not Graduate	No	4100000	12800000	8	780	8200000	700000	
4268	4269	1	Graduate	No	9200000	29700000	10	607	17800000	11800000	

4269 rows × 13 columns

[3]:
```
# Remove spaces from column names
data.columns = data.columns.str.replace(" ", "")
print("\nUpdated column names:")
print(data.columns)
```

```
Updated column names:
Index(['loan_id', 'no_of_dependents', 'education', 'self_employed',
       'income_annum', 'loan_amount', 'loan_term', 'cibil_score',
       'residential_assets_value', 'commercial_assets_value',
       'luxury_assets_value', 'bank_asset_value', 'loan_status'],
      dtype='object')
```

[4]:
```
#EDA for each data analysis
def eda(data):
    print("Shape:",data.shape)
    print("- "*50)
    print("Size:",data.size)
    print("- "*50)
    print("INFO:",data.info)
    print("- "*50)
    print("Describe:",data.describe())
    print("- "*50)
    print("Dtype:",data.dtypes)
    print("* "*50)
    print("Checking Null Values:",data.isnull().sum())

eda(data)
```

```
Shape: (4269, 13)
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Size: 55497
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
INFO: <bound method DataFrame.info of       loan_id  no_of_dependents      education self_employed  income_annum  \
0           1                2       Graduate           No       9600000
1           2                0   Not Graduate          Yes       4100000
2           3                3       Graduate           No       9100000
3           4                3       Graduate           No       8200000
4           5                5   Not Graduate          Yes       9800000
...       ...              ...            ...           ...           ...
```

```
4264     4265                5       Graduate           Yes       1000000
4265     4266                0   Not Graduate           Yes       3300000
4266     4267                2   Not Graduate            No       6500000
4267     4268                1   Not Graduate            No       4100000
4268     4269                1       Graduate            No       9200000

      loan_amount  loan_term  cibil_score  residential_assets_value  \
0        29900000         12          778                   2400000
1        12200000          8          417                   2700000
2        29700000         20          506                   7100000
3        30700000          8          467                  18200000
4        24200000         20          382                  12400000
...           ...        ...          ...                       ...
4264      2300000         12          317                   2800000
4265     11300000         20          559                   4200000
4266     23900000         18          457                   1200000
4267     12800000          8          780                   8200000
4268     29700000         10          607                  17800000

      commercial_assets_value  luxury_assets_value  bank_asset_value  \
0                    17600000             22700000           8000000
1                     2200000              8800000           3300000
2                     4500000             33300000          12800000
3                     3300000             23300000           7900000
4                     8200000             29400000           5000000
...                       ...                  ...               ...
4264                   500000              3300000            800000
4265                  2900000             11000000           1900000
4266                 12400000             18100000           7300000
4267                   700000             14100000           5800000
4268                 11800000             35700000          12000000

     loan_status
0       Approved
1       Rejected
2       Rejected
3       Rejected
4       Rejected
...          ...
4264    Rejected
4265    Approved
4266    Rejected
4267    Approved
4268    Approved

[4269 rows x 13 columns]>
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Describe:           loan_id  no_of_dependents  income_annum   loan_amount    loan_term  \
count  4269.000000      4269.000000  4.269000e+03  4.269000e+03  4269.000000
mean   2135.000000         2.498712  5.059124e+06  1.513345e+07    10.900445
std    1232.498479         1.695910  2.806840e+06  9.043363e+06     5.709187
min       1.000000         0.000000  2.000000e+05  3.000000e+05     2.000000
25%    1068.000000         1.000000  2.700000e+06  7.700000e+06     6.000000
50%    2135.000000         3.000000  5.100000e+06  1.450000e+07    10.000000
75%    3202.000000         4.000000  7.500000e+06  2.150000e+07    16.000000
max    4269.000000         5.000000  9.900000e+06  3.950000e+07    20.000000

       cibil_score  residential_assets_value  commercial_assets_value  \
count  4269.000000              4.269000e+03             4.269000e+03
mean    599.936051              7.472617e+06             4.973155e+06
std     172.430401              6.503637e+06             4.388966e+06
min     300.000000             -1.000000e+05             0.000000e+00
25%     453.000000              2.200000e+06             1.300000e+06
50%     600.000000              5.600000e+06             3.700000e+06
75%     748.000000              1.130000e+07             7.600000e+06
max     900.000000              2.910000e+07             1.940000e+07

       luxury_assets_value  bank_asset_value
count         4.269000e+03      4.269000e+03
mean          1.512631e+07      4.976692e+06
std           9.103754e+06      3.250185e+06
min           3.000000e+05      0.000000e+00
25%           7.500000e+06      2.300000e+06
50%           1.460000e+07      4.600000e+06
75%           2.170000e+07      7.100000e+06
max           3.920000e+07      1.470000e+07
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Dtype: loan_id                    int64
no_of_dependents             int64
education                   object
self_employed               object
income_annum                 int64
loan_amount                  int64
loan_term                    int64
cibil_score                  int64
residential_assets_value     int64
commercial_assets_value      int64
luxury_assets_value          int64
bank_asset_value             int64
loan_status                 object
dtype: object
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
Checking Null Values: loan_id                    0
no_of_dependents            0
education                   0
self_employed               0
income_annum                0
loan_amount                 0
loan_term                   0
cibil_score                 0
residential_assets_value    0
commercial_assets_value     0
luxury_assets_value         0
bank_asset_value            0
loan_status                 0
dtype: int64
```

dtype: int64

[5]:
```
# Analysis for Numerical Columns
def Num_col(data, col):
    mean=data[col].mean()
    median=data[col].median()
    mode=data[col].mode()[0]
    var =data[col].var()
    std=data[col].std()
    skew=data[col].skew()
    Min=data[col].min()
    Max=data[col].max()
    Range=Max-Min
    print("Numerical Columns Analysis:")
    print(f"mean:{mean}\nmedian:{median}\nmode:{mode}\nvar:{var}\nstd:{std}\nskew:{skew}\nMIN:{Min}\nMAAX:{Max}\nRange:{Range}")
```

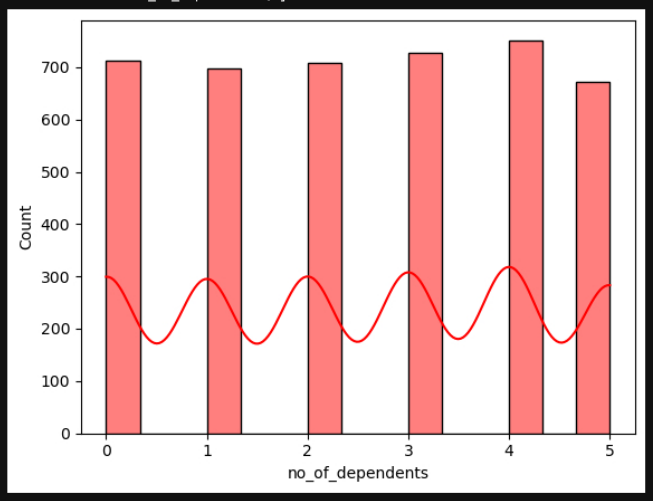
[6]:
```
data.drop(columns=["loan_id"],inplace=True)
```

[7]:
```
Num_col(data,"no_of_dependents")
```

```
Numerical Columns Analysis:
mean:2.4987116420707425
median:3.0
mode:4
var:2.876111273203152
std:1.695910160711101
skew:-0.01797054296784964
MIN:0
MAAX:5
Range:5
```

[8]:
```
sns.histplot(data=data,x="no_of_dependents",bins=15,kde=True,color="red")
```

[8]: <Axes: xlabel='no_of_dependents', ylabel='Count'>

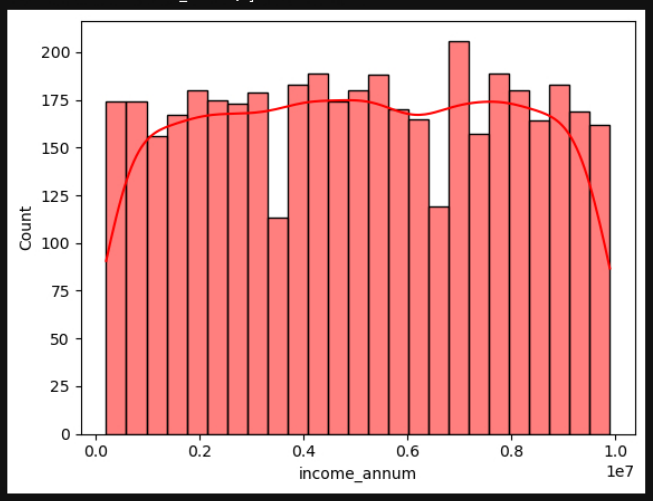


[9]:
```
Num_col(data,"income_annum")
```

```
Numerical Columns Analysis:
mean:5059123.9166081045
median:5100000.0
mode:7000000
var:7878349841482.69
std:2806839.831818462
skew:-0.012814425096650109
MIN:200000
MAAX:9900000
Range:9700000
```

[10]:
```
sns.histplot(data=data,x="income_annum",bins=25,kde=True,color="red")
```

[10]: <Axes: xlabel='income_annum', ylabel='Count'>

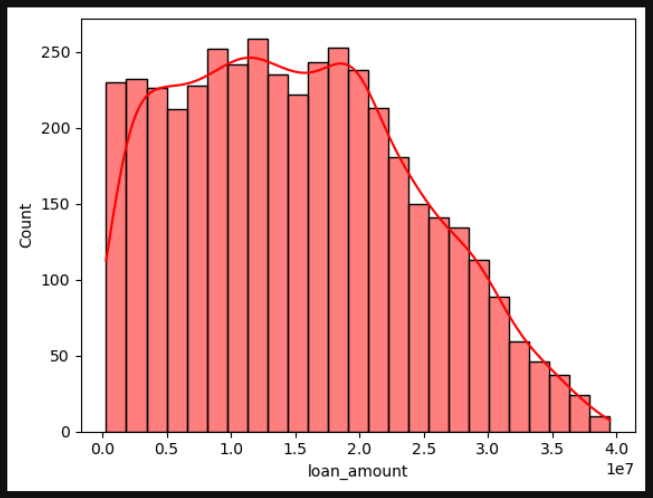


[11]:
```
Num_col(data,"loan_amount")
```

```
Numerical Columns Analysis:
mean:15133450.456781447
median:14500000.0
mode:10600000
var:81782414075625.84
std:9043362.984842854
skew:0.30872388482299223
MIN:300000
MAAX:39500000
Range:39200000
```

[12]:
```
sns.histplot(data=data,x="loan_amount",bins=25,kde=True,color="red")
```

[12]: <Axes: xlabel='loan_amount', ylabel='Count'>



[13]:
```
data.dtypes
```

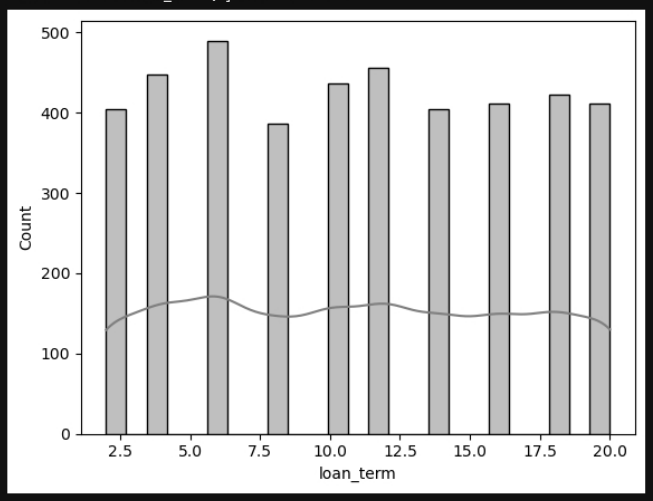
[13]:
```
no_of_dependents            int64
education                  object
self_employed              object
income_annum                int64
loan_amount                 int64
loan_term                   int64
cibil_score                 int64
residential_assets_value    int64
commercial_assets_value     int64
luxury_assets_value         int64
bank_asset_value            int64
loan_status                object
dtype: object
```

[14]:
```
Num_col(data,"loan_term")
```

```
Numerical Columns Analysis:
mean:10.900445069102835
median:10.0
mode:6
var:32.594819389495214
std:5.7091872792452
skew:0.036358907356478495
MIN:2
MAAX:20
Range:18
```

[15]:
```
sns.histplot(data=data,x="loan_term",bins=25,kde=True,color="grey")
```

[15]: <Axes: xlabel='loan_term', ylabel='Count'>



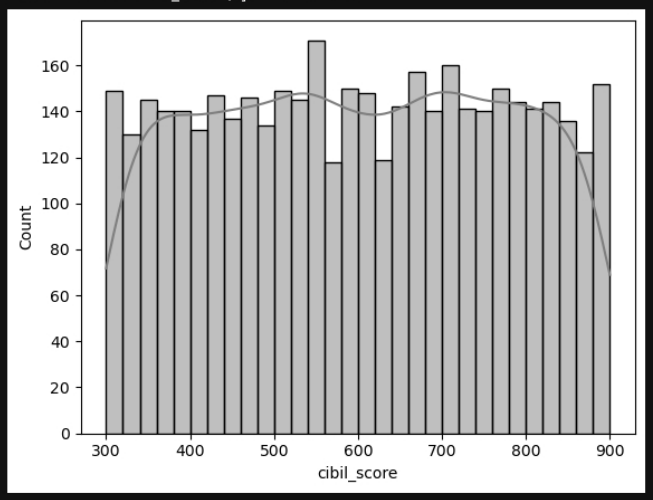
[16]:
```
Num_col(data,"cibil_score")
```

```
Numerical Columns Analysis:
mean:599.9360505973295
median:600.0
mode:348
```

```
var:29732.243097894454
std:172.43040073575904
skew:-0.009039277330065707
MIN:300
MAAX:900
Range:600
```

[17]:
```
sns.histplot(data=data,x="cibil_score",bins=30,kde=True,color="grey")
```

[17]:
```
<Axes: xlabel='cibil_score', ylabel='Count'>
```

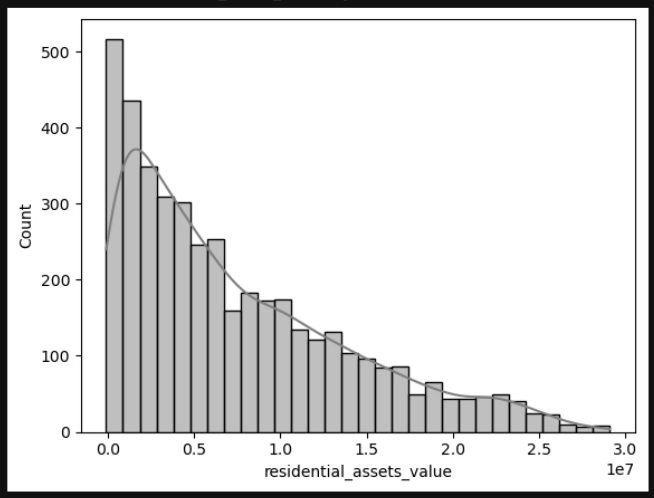


[18]:
```
Num_col(data,"residential_assets_value")
```

```
Numerical Columns Analysis:
mean:7472616.537830873
median:5600000.0
mode:400000
var:42297288864403.16
std:6503636.587664101
skew:0.9784505965115631
MIN:-100000
MAAX:29100000
Range:29200000
```

[19]:
```
sns.histplot(data=data,x="residential_assets_value",bins=30,kde=True,color="grey")
```

[19]:
```
<Axes: xlabel='residential_assets_value', ylabel='Count'>
```

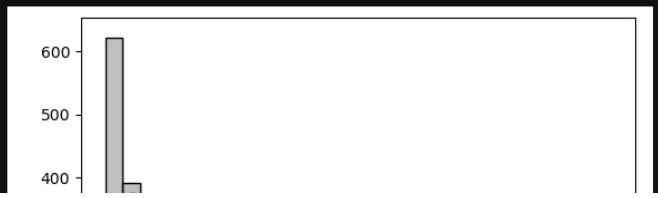


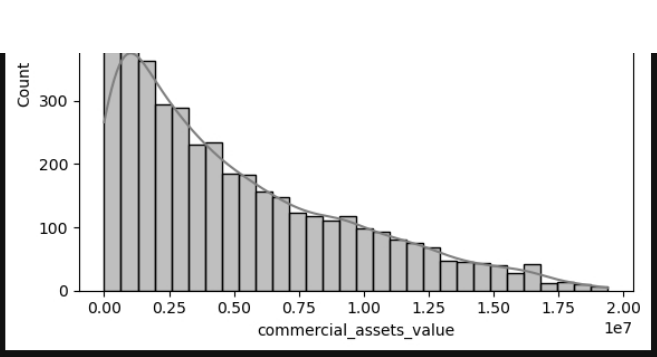
[20]:
```
Num_col(data,"commercial_assets_value")
```

```
Numerical Columns Analysis:
mean:4973155.3056922
median:3700000.0
mode:0
var:19263023335996.324
std:4388966.089638461
skew:0.9577908874986114
MIN:0
MAAX:19400000
Range:19400000
```

[21]:
```
sns.histplot(data=data,x="commercial_assets_value",bins=30,kde=True,color="grey")
```

[21]:
```
<Axes: xlabel='commercial_assets_value', ylabel='Count'>
```



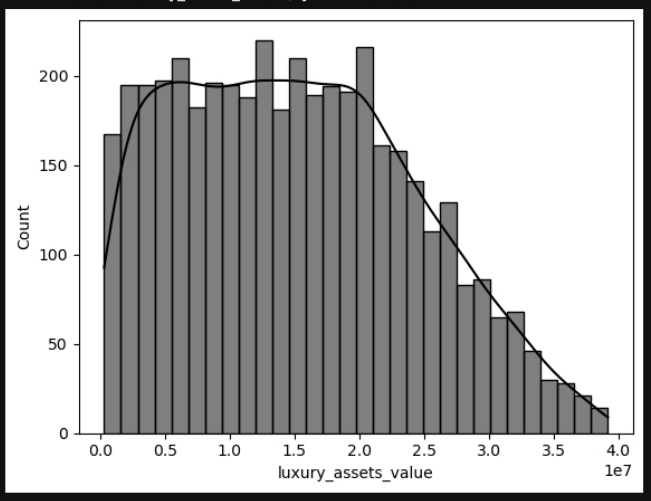


[22]:
```
Num_col(data,"luxury_assets_value")
```

```
Numerical Columns Analysis:
mean:15126305.926446475
median:14600000.0
mode:2900000
var:82878330797671.08
std:9103753.665256497
skew:0.3222075028955774
MIN:300000
MAAX:39200000
Range:38900000
```

[23]:
```
sns.histplot(data=data,x="luxury_assets_value",bins=30,kde=True,color="black")
```

[23]: <Axes: xlabel='luxury_assets_value', ylabel='Count'>

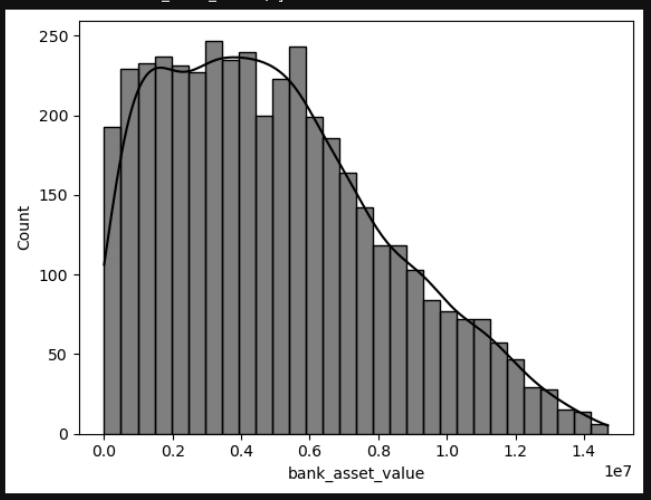


[24]:
```
Num_col(data,"bank_asset_value")
```

```
Numerical Columns Analysis:
mean:4976692.433825252
median:4600000.0
mode:1400000
var:10563704521360.266
std:3250185.3056957023
skew:0.5607250089736816
MIN:0
MAAX:14700000
Range:14700000
```

[25]:
```
sns.histplot(data=data,x="bank_asset_value",bins=30,kde=True,color="black")
```

[25]: <Axes: xlabel='bank_asset_value', ylabel='Count'>



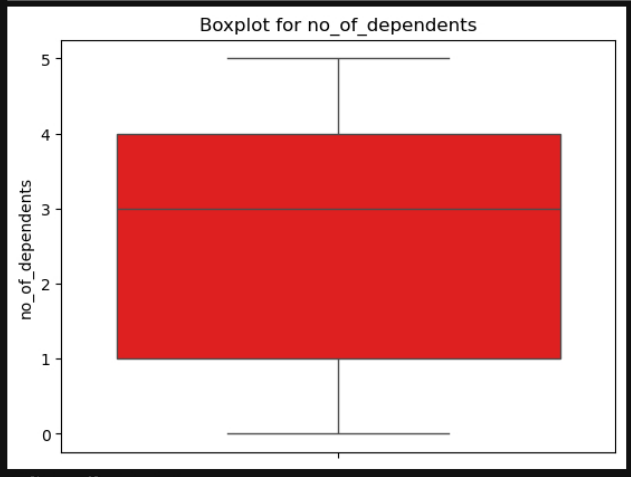
[26]:
```
# checking and handling of outliers

def Checking_and_Handling_Of_Outliers(data, col):
    sns.boxplot(data[col], color = "Red")
```

```
    plt.title(f"Boxplot for {col}")
    plt.show()

    q1 = data[col].quantile(0.25)
    q3 = data[col].quantile(0.75)

    iqr = q3 - q1

    LowerTail = q1 - 1.5*iqr
    UpperTail = q3 + 1.5*iqr

    print(f"25% Quantile q1 = {q1}\n75% Quantile q3 = {q3}\nIQR = {iqr}\n")
    print("-"*80)
    print(f"Lower Tail = {LowerTail}\nUpper Tail = {UpperTail}")
    print("-"*80)

    # Checking for Outliers
    Outliers = data[(data[col] < LowerTail) | (data[col] > UpperTail)]
    print("\nOutliers :\n",Outliers)
    print("-"*80)
#Handling of Outliers :
    data.loc[data[col] < LowerTail, col] = LowerTail # all outliers less than lowertail, assigned by lowertail value
    data.loc[data[col] > UpperTail, col] = UpperTail # all outliers greater than uppertail, assigned by uppertail value

    print("After handling of Outliers data:\n")
    print(data.head())
```

[27]:
```
data.dtypes
```

[27]:
```
no_of_dependents            int64
education                  object
self_employed              object
income_annum                int64
loan_amount                 int64
loan_term                   int64
cibil_score                 int64
residential_assets_value    int64
commercial_assets_value     int64
luxury_assets_value         int64
bank_asset_value            int64
loan_status                object
dtype: object
```

[28]:
```
Checking_and_Handling_Of_Outliers(data, "no_of_dependents")
```

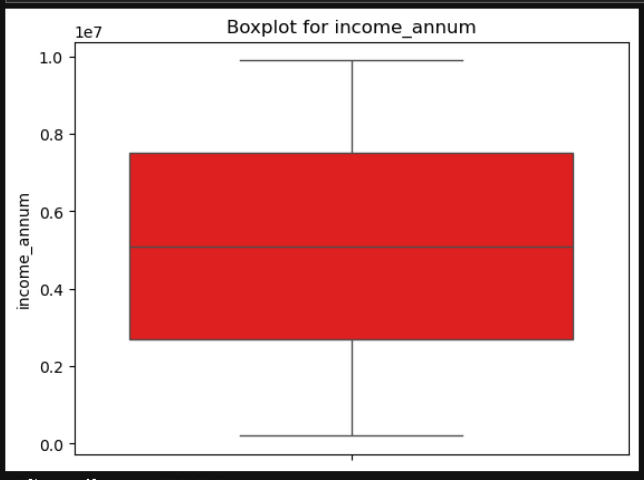



```
25% Quantile q1 = 1.0
75% Quantile q3 = 4.0
IQR = 3.0

--------------------------------------------------------------------------------
Lower Tail = -3.5
Upper Tail = 8.5
--------------------------------------------------------------------------------

Outliers :
 Empty DataFrame
Columns: [no_of_dependents, education, self_employed, income_annum, loan_amount, loan_term, cibil_score, residential_assets_value, commercial_assets_value, luxury_assets_value, bank_asset_value, loan_status]
Index: []
--------------------------------------------------------------------------------
After handling of Outliers data:

   no_of_dependents      education self_employed  income_annum  loan_amount  \
0               2.0       Graduate           No       9600000     29900000
1               0.0   Not Graduate          Yes       4100000     12200000
2               3.0       Graduate           No       9100000     29700000
3               3.0       Graduate           No       8200000     30700000
4               5.0   Not Graduate          Yes       9800000     24200000

   loan_term  cibil_score  residential_assets_value  commercial_assets_value  \
0         12          778                   2400000                 17600000
1          8          417                   2700000                  2200000
2         20          506                   7100000                  4500000
3          8          467                  18200000                  3300000
4         20          382                  12400000                  8200000

   luxury_assets_value  bank_asset_value loan_status
0             22700000           8000000    Approved
1              8800000           3300000    Rejected
```

```
2            33300000          12800000    Rejected
3            23300000           7900000    Rejected
4            29400000           5000000    Rejected
```

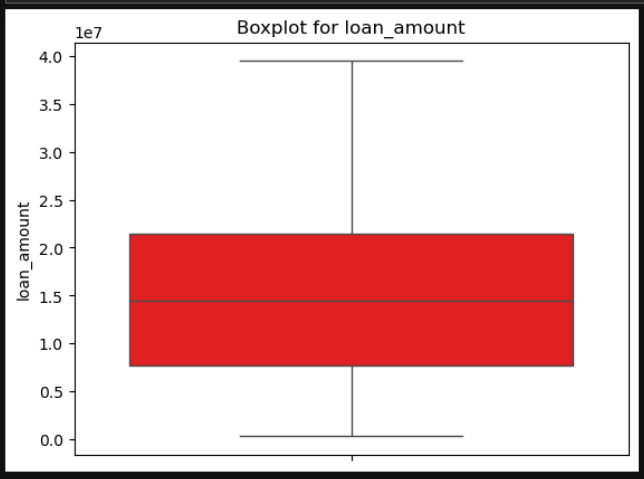
```
C:\Users\deshm\AppData\Local\Temp\ipykernel_12040\3306476731.py:26: FutureWarning: Setting an item of incompatible dtype is deprecated and will raise an error
in a future version of pandas. Value '-3.5' has dtype incompatible with int64, please explicitly cast to a compatible dtype first.
  data.loc[data[col] < LowerTail, col] = LowerTail # all outliers less than lowertail, assigned by lowertail value
```

[29]:
```
Checking_and_Handling_Of_Outliers(data, "income_annum")
```




```
25% Quantile q1 = 2700000.0
75% Quantile q3 = 7500000.0
IQR = 4800000.0

--------------------------------------------------------------------------------
Lower Tail = -4500000.0
Upper Tail = 14700000.0
--------------------------------------------------------------------------------

Outliers :
 Empty DataFrame
Columns: [no_of_dependents, education, self_employed, income_annum, loan_amount, loan_term, cibil_score, residential_assets_value, commercial_assets_value, lu
xury_assets_value, bank_asset_value, loan_status]
Index: []
--------------------------------------------------------------------------------
After handling of Outliers data:

   no_of_dependents      education self_employed  income_annum  loan_amount  \
0               2.0       Graduate            No       9600000     29900000
1               0.0   Not Graduate           Yes       4100000     12200000
2               3.0       Graduate            No       9100000     29700000
3               3.0       Graduate            No       8200000     30700000
4               5.0   Not Graduate           Yes       9800000     24200000

   loan_term  cibil_score  residential_assets_value  commercial_assets_value  \
0         12          778                   2400000                 17600000
1          8          417                   2700000                  2200000
2         20          506                   7100000                  4500000
3          8          467                  18200000                  3300000
4         20          382                  12400000                  8200000

   luxury_assets_value  bank_asset_value loan_status
0             22700000           8000000    Approved
1              8800000           3300000    Rejected
2             33300000          12800000    Rejected
3             23300000           7900000    Rejected
4             29400000           5000000    Rejected
```

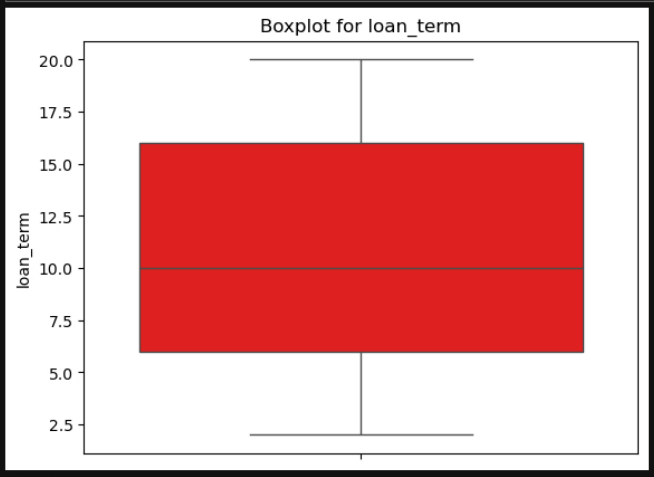
[30]:
```
Checking_and_Handling_Of_Outliers(data, "loan_amount")
```




```
25% Quantile q1 = 7700000.0
75% Quantile q3 = 21500000.0
IQR = 13800000.0

--------------------------------------------------------------------------------
Lower Tail = -13000000.0
Upper Tail = 42200000.0
--------------------------------------------------------------------------------
```

```
Outliers :
 Empty DataFrame
Columns: [no_of_dependents, education, self_employed, income_annum, loan_amount, loan_term, cibil_score, residential_assets_value, commercial_assets_value, luxury_assets_value, bank_asset_value, loan_status]
Index: []
--------------------------------------------------------------------------------
After handling of Outliers data:

   no_of_dependents      education self_employed  income_annum  loan_amount  \
0               2.0       Graduate           No       9600000     29900000
1               0.0   Not Graduate          Yes       4100000     12200000
2               3.0       Graduate           No       9100000     29700000
3               3.0       Graduate           No       8200000     30700000
4               5.0   Not Graduate          Yes       9800000     24200000

   loan_term  cibil_score  residential_assets_value  commercial_assets_value  \
0         12          778                   2400000                 17600000
1          8          417                   2700000                  2200000
2         20          506                   7100000                  4500000
3          8          467                  18200000                  3300000
4         20          382                  12400000                  8200000

   luxury_assets_value  bank_asset_value loan_status
0             22700000           8000000    Approved
1              8800000           3300000    Rejected
2             33300000          12800000    Rejected
3             23300000           7900000    Rejected
4             29400000           5000000    Rejected
```

[31]:
```
Checking_and_Handling_Of_Outliers(data, "loan_term")
```



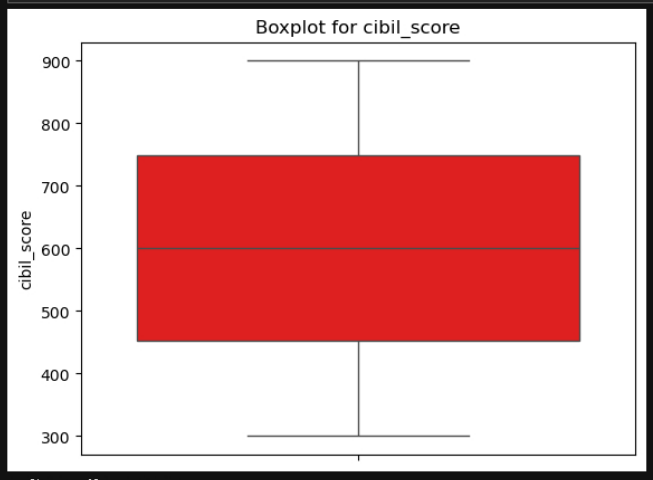

```
25% Quantile q1 = 6.0
75% Quantile q3 = 16.0
IQR = 10.0

--------------------------------------------------------------------------------
Lower Tail = -9.0
Upper Tail = 31.0
--------------------------------------------------------------------------------

Outliers :
 Empty DataFrame
Columns: [no_of_dependents, education, self_employed, income_annum, loan_amount, loan_term, cibil_score, residential_assets_value, commercial_assets_value, luxury_assets_value, bank_asset_value, loan_status]
Index: []
--------------------------------------------------------------------------------
After handling of Outliers data:

   no_of_dependents      education self_employed  income_annum  loan_amount  \
0               2.0       Graduate           No       9600000     29900000
1               0.0   Not Graduate          Yes       4100000     12200000
2               3.0       Graduate           No       9100000     29700000
3               3.0       Graduate           No       8200000     30700000
4               5.0   Not Graduate          Yes       9800000     24200000

   loan_term  cibil_score  residential_assets_value  commercial_assets_value  \
0         12          778                   2400000                 17600000
1          8          417                   2700000                  2200000
2         20          506                   7100000                  4500000
3          8          467                  18200000                  3300000
4         20          382                  12400000                  8200000

   luxury_assets_value  bank_asset_value loan_status
0             22700000           8000000    Approved
1              8800000           3300000    Rejected
2             33300000          12800000    Rejected
3             23300000           7900000    Rejected
4             29400000           5000000    Rejected
```

[32]:
```
data.dtypes
```

[32]:
```
no_of_dependents            float64
education                    object
self_employed                object
income_annum                  int64
loan_amount                   int64
loan_term                     int64
cibil_score                   int64
residential_assets_value      int64
commercial_assets_value       int64
luxury_assets_value           int64
```

```
bank_asset_value             int64
loan_status                 object
dtype: object
```

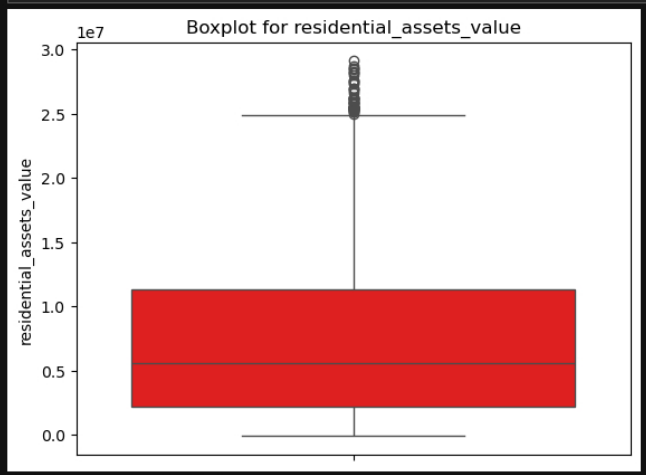
[33]:
```
Checking_and_Handling_Of_Outliers(data, "cibil_score")
```




```
25% Quantile q1 = 453.0
75% Quantile q3 = 748.0
IQR = 295.0

--------------------------------------------------------------------------------
Lower Tail = 10.5
Upper Tail = 1190.5
--------------------------------------------------------------------------------

Outliers :
 Empty DataFrame
Columns: [no_of_dependents, education, self_employed, income_annum, loan_amount, loan_term, cibil_score, residential_assets_value, commercial_assets_value, luxury_assets_value, bank_asset_value, loan_status]
Index: []
--------------------------------------------------------------------------------
After handling of Outliers data:

   no_of_dependents      education self_employed  income_annum  loan_amount  \
0               2.0       Graduate            No       9600000     29900000
1               0.0   Not Graduate           Yes       4100000     12200000
2               3.0       Graduate            No       9100000     29700000
3               3.0       Graduate            No       8200000     30700000
4               5.0   Not Graduate           Yes       9800000     24200000

   loan_term  cibil_score  residential_assets_value  commercial_assets_value  \
0         12        778.0                   2400000                 17600000
1          8        417.0                   2700000                  2200000
2         20        506.0                   7100000                  4500000
3          8        467.0                  18200000                  3300000
4         20        382.0                  12400000                  8200000

   luxury_assets_value  bank_asset_value loan_status
0             22700000           8000000    Approved
1              8800000           3300000    Rejected
2             33300000          12800000    Rejected
3             23300000           7900000    Rejected
4             29400000           5000000    Rejected
```

```
C:\Users\deshm\AppData\Local\Temp\ipykernel_12040\3306476731.py:26: FutureWarning: Setting an item of incompatible dtype is deprecated and will raise an error in a future version of pandas. Value '10.5' has dtype incompatible with int64, please explicitly cast to a compatible dtype first.
  data.loc[data[col] < LowerTail, col] = LowerTail # all outliers less than lowertail, assigned by lowertail value
```

[34]:
```
Checking_and_Handling_Of_Outliers(data, "residential_assets_value")
```




```
25% Quantile q1 = 2200000.0
75% Quantile q3 = 11300000.0
IQR = 9100000.0

--------------------------------------------------------------------------------
Lower Tail = -11450000.0
Upper Tail = 24950000.0
--------------------------------------------------------------------------------
```

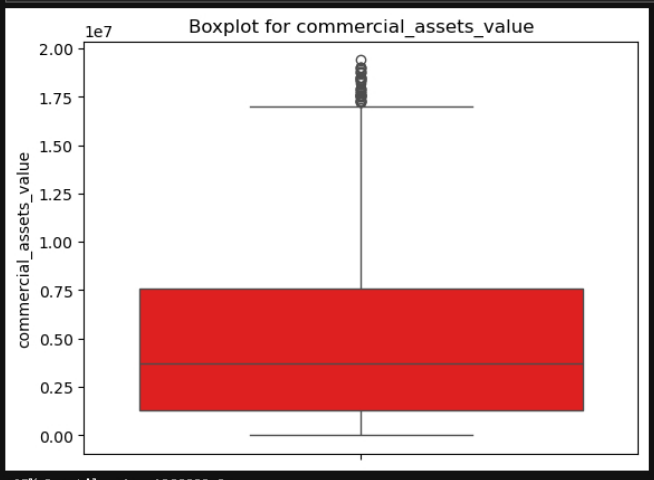
Outliers :

	no_of_dependents	education	self_employed	income_annum	loan_amount	loan_term	cibil_score	residential_assets_value
82	2.0	Not Graduate	Yes	9900000	21200000	16	363.0	25500000
98	4.0	Graduate	No	9400000	29400000	12	562.0	25900000
123	0.0	Not Graduate	Yes	9000000	18700000	18	865.0	26800000
228	1.0	Not Graduate	Yes	8700000	27000000	10	717.0	25500000
262	3.0	Graduate	No	9200000	34300000	18	523.0	25600000
693	0.0	Graduate	No	9200000	25400000	2	783.0	25400000
714	4.0	Not Graduate	No	9900000	22700000	12	567.0	28300000
781	3.0	Not Graduate	Yes	9900000	20400000	4	865.0	26300000
892	4.0	Graduate	No	9300000	24900000	14	455.0	27600000
905	5.0	Not Graduate	No	9700000	37000000	14	459.0	28700000
919	0.0	Not Graduate	Yes	8800000	25600000	10	715.0	25300000
924	0.0	Not Graduate	Yes	8600000	23600000	2	361.0	25800000
953	5.0	Graduate	Yes	8700000	28100000	10	660.0	25500000
956	0.0	Not Graduate	Yes	9100000	22000000	12	387.0	25300000
987	5.0	Graduate	No	9400000	25800000	14	625.0	28200000
997	3.0	Not Graduate	Yes	8500000	28300000	2	810.0	25200000
1002	5.0	Not Graduate	No	9600000	20700000	14	442.0	25800000
1397	1.0	Not Graduate	Yes	9200000	24500000	4	545.0	25700000
1419	2.0	Graduate	No	9100000	21700000	12	549.0	25500000
1468	3.0	Not Graduate	Yes	9500000	25500000	12	881.0	27000000
1591	1.0	Graduate	No	9500000	23000000	10	627.0	28500000
1625	2.0	Not Graduate	Yes	9900000	33200000	4	465.0	26300000
1965	0.0	Not Graduate	No	9900000	22300000	20	613.0	28400000
1997	0.0	Graduate	No	9300000	20300000	14	759.0	27000000
2185	2.0	Graduate	Yes	9200000	25300000	4	827.0	25900000
2318	2.0	Not Graduate	Yes	9600000	29800000	16	499.0	28500000
2384	3.0	Graduate	Yes	9100000	30100000	14	564.0	26600000
2412	4.0	Not Graduate	No	9700000	26200000	4	518.0	26200000
2586	1.0	Not Graduate	Yes	9800000	36000000	12	629.0	28000000
2715	4.0	Graduate	Yes	9100000	28400000	12	808.0	25800000
2818	1.0	Not Graduate	Yes	9500000	24400000	14	696.0	25500000
2828	4.0	Graduate	Yes	9200000	18600000	2	889.0	26900000
2922	0.0	Not Graduate	Yes	9600000	23800000	12	389.0	25200000
2927	4.0	Graduate	No	9400000	32200000	6	898.0	27600000
2930	3.0	Not Graduate	No	8700000	24800000	20	360.0	25300000
2940	1.0	Graduate	Yes	9600000	34100000	20	828.0	26100000
3119	4.0	Not Graduate	No	9800000	29400000	8	592.0	29100000
3157	0.0	Not Graduate	Yes	8600000	20000000	10	704.0	25400000
3234	1.0	Not Graduate	No	9900000	23400000	16	687.0	28200000
3310	5.0	Not Graduate	No	8500000	23400000	16	604.0	25100000
3498	2.0	Graduate	Yes	9100000	24200000	16	685.0	25600000
3631	3.0	Graduate	No	9800000	21000000	12	342.0	25400000
3763	3.0	Graduate	No	9600000	34500000	20	523.0	26100000
3782	5.0	Graduate	Yes	9500000	21700000	16	307.0	27500000
3860	4.0	Graduate	No	8700000	25200000	18	790.0	25000000
3868	5.0	Not Graduate	No	9600000	27100000	8	594.0	25500000
3872	0.0	Graduate	No	8800000	17500000	4	848.0	25400000
3880	3.0	Not Graduate	No	9100000	19300000	12	696.0	26200000
4027	1.0	Not Graduate	No	9100000				
4042	4.0	Graduate	No	9300000				
4074	4.0	Not Graduate	Yes	9400000				
4237	0.0	Not Graduate	No	9100000				

```
4027     24000000          20        353.0                25100000
4042     34900000           6        837.0                27400000
4074     25700000          14        845.0                27300000
4237     30100000           8        322.0                26200000

      commercial_assets_value  luxury_assets_value  bank_asset_value  \
82                   11400000             26600000           6800000
98                   15200000             36400000           7100000
123                         0             20900000          11300000
228                   8600000             17500000           9100000
262                   4000000             29100000           8800000
693                   2000000             19600000          12400000
714                   9900000             29700000           5400000
781                   3600000             36500000          12800000
892                   4700000             35100000          13900000
905                  17900000             22800000           5600000
919                   2500000             34900000           8600000
924                   5300000             24700000           7500000
953                    800000             25900000          10100000
956                  15600000             31700000           7000000
987                  15400000             30800000          11100000
997                   4300000             18500000          10700000
1002                  9800000             31800000           8200000
1397                        0             20600000          13300000
1419                 11000000             34700000          12200000
1468                 14100000             21600000           8900000
1591                 11500000             21700000           9100000
1625                 17000000             24400000           5000000
1965                  8900000             23600000           4900000
1997                 11600000             26500000          10900000
2185                 15900000             19300000          11000000
2318                  6300000             31800000           7400000
2384                  7300000             22700000           8200000
2412                        0             22500000          14000000
2586                  4100000             29800000          11200000
2715                 11800000             19300000           9700000
2818                  6700000             29500000           6300000
2828                  4200000             33300000           9700000
2922                 15800000             27000000           8300000
2927                 11800000             19500000           8600000
2930                  8700000             31100000           4300000
2940                 19000000             20800000          12400000
3119                 16700000             35900000           9000000
3157                 10000000             30800000          10300000
3234                  3700000             29600000           8200000
3310                 16500000             24500000           4800000
3498                  1900000             31100000           9100000
3631                 16600000             26900000           7000000
3763                  2600000             24300000           5300000
3782                  2700000             28500000           9700000
3860                  9100000             18400000           8400000
3868                  4600000             26800000          13300000
3872                  8200000             31600000           5100000
3880                  7700000             30100000          13300000
4027                  9200000             26100000          10900000
4042                   800000             21400000           8500000
4074                  4100000             26500000          12200000
4237                  9500000             18800000           5900000

     loan_status
82      Rejected
98      Approved
123     Approved
228     Approved
262     Rejected
693     Approved
714     Approved
781     Approved
892     Rejected
905     Rejected
919     Approved
924     Rejected
953     Approved
956     Rejected
987     Approved
997     Approved
1002    Rejected
1397    Rejected
1419    Rejected
1468    Approved
1591    Approved
1625    Approved
1965    Approved
1997    Approved
2185    Approved
2318    Rejected
2384    Approved
2412    Rejected
2586    Approved
2715    Approved
2818    Approved
2828    Approved
2922    Rejected
2927    Approved
2930    Rejected
2940    Approved
3119    Approved
3157    Approved
3234    Approved
3310    Approved
3498    Approved
3631    Rejected
3763    Rejected
3782    Rejected
3860    Approved
3868    Approved
```

```
3872    Approved
3880    Approved
4027    Rejected
4042    Approved
4074    Approved
4237    Rejected
--------------------------------------------------------------------------------
After handling of Outliers data:

   no_of_dependents      education self_employed  income_annum  loan_amount  \
0               2.0       Graduate            No       9600000     29900000
1               0.0   Not Graduate           Yes       4100000     12200000
2               3.0       Graduate            No       9100000     29700000
3               3.0       Graduate            No       8200000     30700000
4               5.0   Not Graduate           Yes       9800000     24200000

   loan_term  cibil_score  residential_assets_value  commercial_assets_value  \
0         12        778.0                   2400000                 17600000
1          8        417.0                   2700000                  2200000
2         20        506.0                   7100000                  4500000
3          8        467.0                  18200000                  3300000
4         20        382.0                  12400000                  8200000

   luxury_assets_value  bank_asset_value loan_status
0             22700000           8000000    Approved
1              8800000           3300000    Rejected
2             33300000          12800000    Rejected
3             23300000           7900000    Rejected
4             29400000           5000000    Rejected
```

[35]:
```
Checking_and_Handling_Of_Outliers(data, "commercial_assets_value")
```




```
25% Quantile q1 = 1300000.0
75% Quantile q3 = 7600000.0
IQR = 6300000.0

--------------------------------------------------------------------------------
Lower Tail = -8150000.0
Upper Tail = 17050000.0
--------------------------------------------------------------------------------

Outliers :
      no_of_dependents      education self_employed  income_annum  \
0                  2.0       Graduate            No       9600000
157                5.0   Not Graduate           Yes       9900000
231                2.0       Graduate            No       9800000
258                0.0       Graduate            No       9800000
323                3.0   Not Graduate           Yes       9500000
367                1.0   Not Graduate           Yes       9400000
554                3.0   Not Graduate           Yes       9500000
791                4.0       Graduate           Yes       9300000
895                0.0       Graduate            No       9400000
905                5.0   Not Graduate            No       9700000
1029               2.0   Not Graduate           Yes       9500000
1131               0.0   Not Graduate            No       9900000
1194               4.0   Not Graduate           Yes       9000000
1254               5.0   Not Graduate           Yes       9800000
1272               5.0   Not Graduate            No       9900000
1304               2.0       Graduate           Yes       9400000
1609               5.0   Not Graduate           Yes       9800000
1761               2.0   Not Graduate           Yes       9800000
1768               4.0   Not Graduate            No       9900000
1812               5.0   Not Graduate            No       9000000
2004               2.0   Not Graduate            No       9900000
2302               5.0       Graduate           Yes       9700000
2349               0.0       Graduate            No       9200000
2914               3.0   Not Graduate           Yes       9900000
2933               0.0       Graduate            No       9600000
2940               1.0       Graduate           Yes       9600000
2976               3.0   Not Graduate           Yes       9400000
3000               1.0   Not Graduate           Yes       9900000
3439               5.0   Not Graduate           Yes       9700000
3541               3.0   Not Graduate            No       8800000
3790               1.0   Not Graduate            No       9600000
3827               1.0       Graduate           Yes       9600000
3882               1.0       Graduate           Yes       9700000
3949               1.0   Not Graduate           Yes       9100000
4010               3.0       Graduate           Yes       9000000
4120               3.0       Graduate            No       9100000
4205               2.0   Not Graduate            No       8900000
```

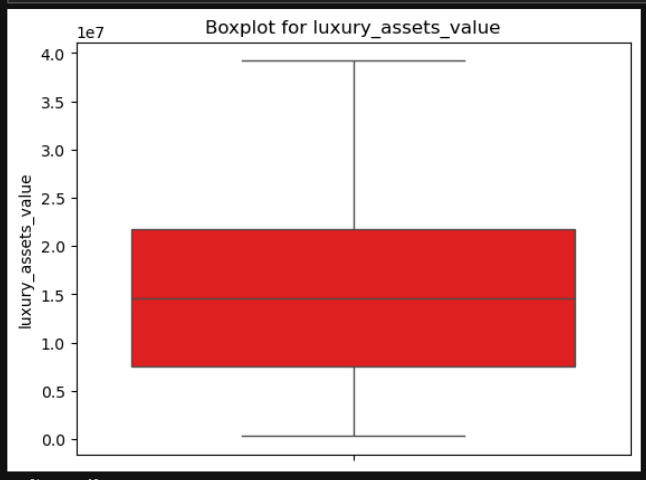
	loan_amount	loan_term	cibil_score	residential_assets_value	\
0	29900000	12	778.0	2400000	
157	33300000	6	364.0	5300000	
231	32400000	2	543.0	24200000	
258	38800000	8	652.0	23200000	
323	24200000	8	879.0	3100000	
367	29800000	10	377.0	24100000	
554	22800000	6	319.0	200000	
791	35500000	6	547.0	2800000	
895	27900000	6	456.0	21400000	
905	37000000	14	459.0	24950000	
1029	37000000	8	887.0	21800000	
1131	27000000	16	566.0	3500000	
1194	21500000	14	508.0	16600000	
1254	35700000	10	726.0	9400000	
1272	35800000	20	470.0	9500000	
1304	24800000	14	513.0	24200000	
1609	23200000	10	416.0	11000000	
1761	32700000	2	304.0	4700000	
1768	29100000	18	772.0	24100000	
1812	35200000	2	437.0	18400000	
2004	31600000	16	892.0	4200000	
2302	30100000	12	434.0	19600000	
2349	22500000	10	317.0	14200000	
2914	36500000	16	562.0	17300000	
2933	23200000	18	389.0	1000000	
2940	34100000	20	828.0	24950000	
2976	30900000	20	873.0	22800000	
3000	25700000	2	340.0	15600000	
3439	27400000	2	328.0	0	
3541	24900000	16	668.0	16700000	
3790	32400000	20	447.0	15400000	
3827	35900000	8	534.0	24500000	
3882	34300000	20	756.0	15700000	
3949	23600000	20	535.0	2800000	
4010	18600000	16	518.0	18400000	
4120	19900000	2	711.0	2200000	
4205	17800000	20	603.0	3300000	

	commercial_assets_value	luxury_assets_value	bank_asset_value	\
0	17600000	22700000	8000000	
157	18700000	27800000	8300000	
231	17500000	25700000	7700000	
258	19000000	29700000	5400000	
323	17200000	26400000	12700000	
367	18500000	30900000	11700000	
554	18800000	34100000	9400000	
791	17800000	27000000	13900000	
895	18500000	35600000	7800000	
905	17900000	22800000	5600000	
1029	18300000	19200000	11700000	
1131	17300000	36900000	11500000	
1194	17300000	24300000	13100000	
1254	18900000	31200000	12000000	
1272	18400000	39100000	14700000	
1304	18200000	37200000	7500000	
1609	17300000	21600000	8500000	
1761	19000000	24500000	8600000	
1768	19400000	26800000	11700000	
1812	17800000	18300000	10400000	
2004	17500000	31400000	12700000	
2302	17900000	21000000	6900000	
2349	17400000	28500000	10700000	
2914	18500000	22300000	9800000	
2933	17600000	22200000	8800000	
2940	19000000	20800000	12400000	
2976	17700000	19300000	9000000	
3000	18800000	29500000	11800000	
3439	18400000	22000000	7900000	
3541	17200000	30000000	10500000	
3790	18400000	19800000	4700000	
3827	17700000	31900000	5400000	
3882	18500000	26800000	14100000	
3949	17600000	26400000	4500000	
4010	17600000	28900000	10300000	
4120	17900000	24300000	13000000	
4205	17600000	29000000	13100000	

	loan_status
0	Approved
157	Rejected
231	Approved
258	Approved
323	Approved
367	Rejected
554	Rejected
791	Rejected
895	Rejected
905	Rejected
1029	Approved
1131	Approved
1194	Rejected
1254	Approved
1272	Rejected
1304	Rejected
1609	Rejected
1761	Approved
1768	Approved
1812	Approved
2004	Approved
2302	Rejected
2349	Rejected
2914	Approved
2933	Rejected
2940	Approved
2976	Approved

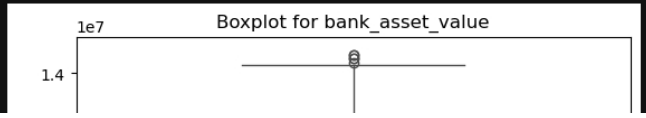
```
3000    Rejected
3439    Rejected
3541    Approved
3790    Rejected
3827    Rejected
3882    Approved
3949    Rejected
4010    Rejected
4120    Approved
4205    Approved
--------------------------------------------------------------------------------
After handling of Outliers data:

   no_of_dependents      education self_employed  income_annum  loan_amount  \
0               2.0       Graduate            No       9600000     29900000
1               0.0   Not Graduate           Yes       4100000     12200000
2               3.0       Graduate            No       9100000     29700000
3               3.0       Graduate            No       8200000     30700000
4               5.0   Not Graduate           Yes       9800000     24200000

   loan_term  cibil_score  residential_assets_value  commercial_assets_value  \
0         12        778.0                   2400000                 17050000
1          8        417.0                   2700000                  2200000
2         20        506.0                   7100000                  4500000
3          8        467.0                  18200000                  3300000
4         20        382.0                  12400000                  8200000

   luxury_assets_value  bank_asset_value loan_status
0             22700000           8000000    Approved
1              8800000           3300000    Rejected
2             33300000          12800000    Rejected
3             23300000           7900000    Rejected
4             29400000           5000000    Rejected
```

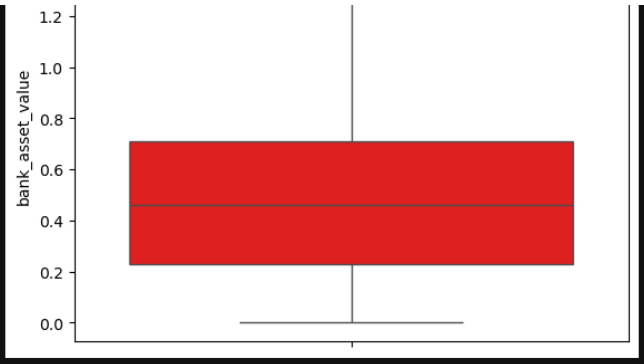
[36]:
```
Checking_and_Handling_Of_Outliers(data, "luxury_assets_value")
```




```
25% Quantile q1 = 7500000.0
75% Quantile q3 = 21700000.0
IQR = 14200000.0

--------------------------------------------------------------------------------
Lower Tail = -13800000.0
Upper Tail = 43000000.0
--------------------------------------------------------------------------------

Outliers :
 Empty DataFrame
Columns: [no_of_dependents, education, self_employed, income_annum, loan_amount, loan_term, cibil_score, residential_assets_value, commercial_assets_value, lu
xury_assets_value, bank_asset_value, loan_status]
Index: []
--------------------------------------------------------------------------------
After handling of Outliers data:

   no_of_dependents      education self_employed  income_annum  loan_amount  \
0               2.0       Graduate            No       9600000     29900000
1               0.0   Not Graduate           Yes       4100000     12200000
2               3.0       Graduate            No       9100000     29700000
3               3.0       Graduate            No       8200000     30700000
4               5.0   Not Graduate           Yes       9800000     24200000

   loan_term  cibil_score  residential_assets_value  commercial_assets_value  \
0         12        778.0                   2400000                 17050000
1          8        417.0                   2700000                  2200000
2         20        506.0                   7100000                  4500000
3          8        467.0                  18200000                  3300000
4         20        382.0                  12400000                  8200000

   luxury_assets_value  bank_asset_value loan_status
0             22700000           8000000    Approved
1              8800000           3300000    Rejected
2             33300000          12800000    Rejected
3             23300000           7900000    Rejected
4             29400000           5000000    Rejected
```

[37]:
```
Checking_and_Handling_Of_Outliers(data, "bank_asset_value")
```






```
25% Quantile q1 = 2300000.0
75% Quantile q3 = 7100000.0
IQR = 4800000.0

--------------------------------------------------------------------------------
Lower Tail = -4900000.0
Upper Tail = 14300000.0
--------------------------------------------------------------------------------

Outliers :
      no_of_dependents      education self_employed  income_annum  \
200               4.0   Not Graduate          Yes       9800000
1272              5.0   Not Graduate           No       9900000
1633              0.0       Graduate          Yes       9800000
1674              1.0   Not Graduate           No       9800000
1805              5.0   Not Graduate           No       9900000

      loan_amount  loan_term  cibil_score  residential_assets_value  \
200      21200000         20        355.0                  22000000
1272     35800000         20        470.0                   9500000
1633     23200000         10        573.0                  23900000
1674     21300000          2        356.0                  23000000
1805     33600000         20        580.0                   7000000

      commercial_assets_value  luxury_assets_value  bank_asset_value  \
200                   8900000             31800000          14400000
1272                 17050000             39100000          14700000
1633                  3300000             23600000          14600000
1674                 12300000             25700000          14600000
1805                 10800000             37400000          14700000

     loan_status
200     Rejected
1272    Rejected
1633    Approved
1674    Rejected
1805    Approved
--------------------------------------------------------------------------------
After handling of Outliers data:

   no_of_dependents      education self_employed  income_annum  loan_amount  \
0               2.0       Graduate           No       9600000     29900000
1               0.0   Not Graduate          Yes       4100000     12200000
2               3.0       Graduate           No       9100000     29700000
3               3.0       Graduate           No       8200000     30700000
4               5.0   Not Graduate          Yes       9800000     24200000

   loan_term  cibil_score  residential_assets_value  commercial_assets_value  \
0         12        778.0                   2400000                 17050000
1          8        417.0                   2700000                  2200000
2         20        506.0                   7100000                  4500000
3          8        467.0                  18200000                  3300000
4         20        382.0                  12400000                  8200000

   luxury_assets_value  bank_asset_value loan_status
0             22700000           8000000    Approved
1              8800000           3300000    Rejected
2             33300000          12800000    Rejected
3             23300000           7900000    Rejected
4             29400000           5000000    Rejected
```

[38]:
```
# Analysis of Categorical Columns(Variable)
def Cat_col(data, col):
    unique_values = data[col].unique()  # Fixed typo: renamed to unique_values
    value_counts = data[col].value_counts()
    mode = data[col].mode()[0]  # Fixed mode access by adding parentheses

    # Enhanced string formatting for clarity
    print(f"Unique Values in '{col}':\n{unique_values}\n")
    print(f"Value Counts in '{col}':\n{value_counts}\n")
    print(f"Mode of '{col}': {mode}\n")


    data[col].value_counts().plot.pie(autopct="%1.1f%%")
    plt.title(f"data-{col} (pie chart)")
    plt.show
```

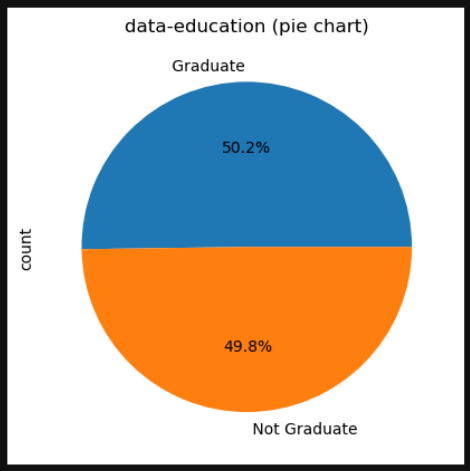
[39]:
```
data.dtypes
```

[39]:
```
no_of_dependents           float64
education                   object
self_employed               object
income_annum                 int64
loan_amount                  int64
loan_term                    int64
cibil_score                float64
```

```
residential_assets_value      int64
commercial_assets_value       int64
luxury_assets_value           int64
bank_asset_value              int64
loan_status                  object
dtype: object
```

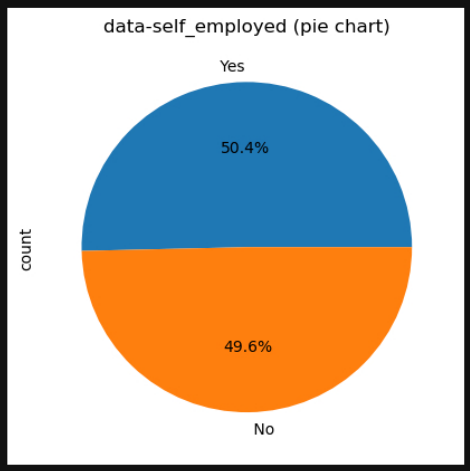
[40]:
```
Cat_col(data,"education")
```

```
Unique Values in 'education':
[' Graduate' ' Not Graduate']

Value Counts in 'education':
education
Graduate        2144
Not Graduate    2125
Name: count, dtype: int64

Mode of 'education':  Graduate
```



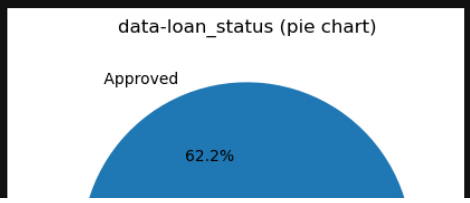
[41]:
```
Cat_col(data,"self_employed")
```

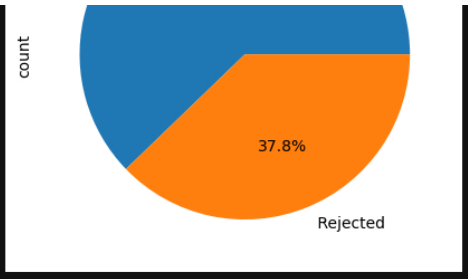
```
Unique Values in 'self_employed':
[' No' ' Yes']

Value Counts in 'self_employed':
self_employed
Yes    2150
No     2119
Name: count, dtype: int64

Mode of 'self_employed':  Yes
```



[42]:
```
Cat_col(data,"loan_status")
```

```
Unique Values in 'loan_status':
[' Approved' ' Rejected']

Value Counts in 'loan_status':
loan_status
Approved    2656
Rejected    1613
Name: count, dtype: int64

Mode of 'loan_status':  Approved
```

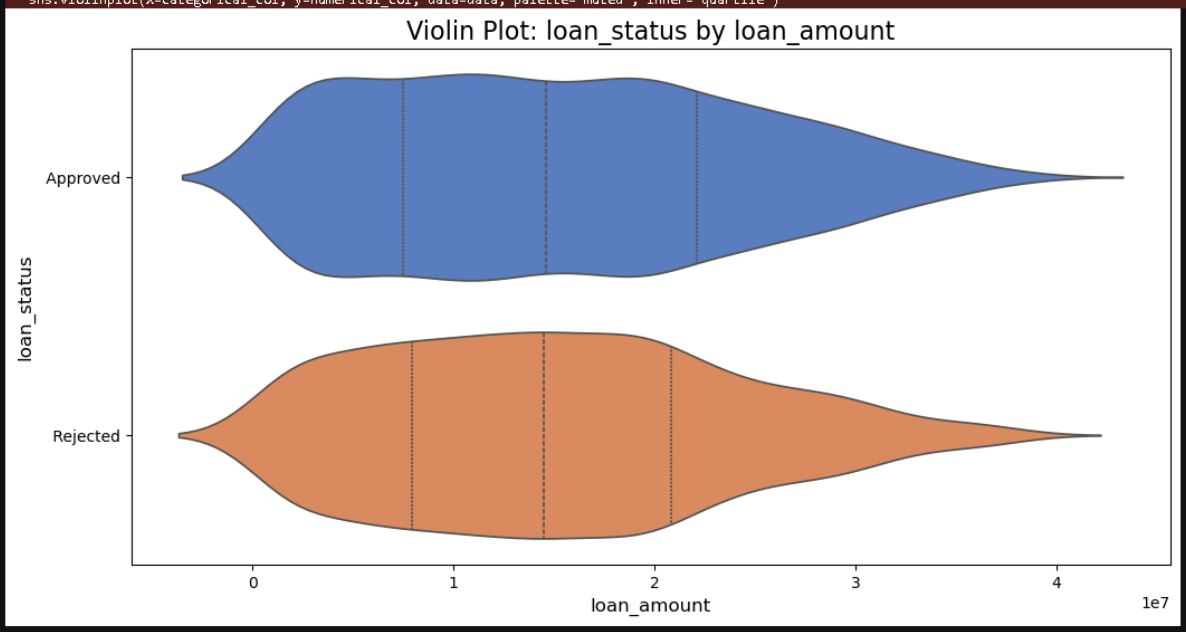




[43]:
```
# Bivariate analysis of columns
def bivariate_violin_plot(data, categorical_col, numerical_col):
    plt.figure(figsize=(12, 6))
    sns.violinplot(x=categorical_col, y=numerical_col, data=data, palette="muted", inner="quartile")
    plt.title(f'Violin Plot: {numerical_col} by {categorical_col}', fontsize=16)
    plt.xlabel(categorical_col, fontsize=12)
    plt.ylabel(numerical_col, fontsize=12)
    plt.show()
```

[44]:
```
bivariate_violin_plot(data,"loan_amount" ,"loan_status")
```

```
C:\Users\deshm\AppData\Local\Temp\ipykernel_12040\3825122554.py:4: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

  sns.violinplot(x=categorical_col, y=numerical_col, data=data, palette="muted", inner="quartile")
```



Encoding

[46]:
```
data.dtypes
```

[46]:
```
no_of_dependents            float64
education                    object
self_employed                object
income_annum                  int64
loan_amount                   int64
loan_term                     int64
cibil_score                 float64
residential_assets_value      int64
commercial_assets_value       int64
luxury_assets_value           int64
bank_asset_value              int64
loan_status                  object
dtype: object
```

[47]:
```
data["education"].unique()
```

[47]:
```
array([' Graduate', ' Not Graduate'], dtype=object)
```

[48]:
```
data["education_encoded"]=data["education"].replace({'Graduate':0, 'Not Graduate':1})
```

[49]:
```
data["self_employed"].unique()
```

[49]:
```
array([' No', ' Yes'], dtype=object)
```

[50]:
```
data["self_employed_encoded"]=data["self_employed"].replace({'No':0, 'Yes':1})
```

[51]:
```
data["loan_status"].unique()
```

[51]:
```
array([' Approved', ' Rejected'], dtype=object)
```

[52]:
```
data["loan_status_encoded"]=data["loan_status"].replace({'Approved':0, 'Rejected':1})
```

[53]:
```
data.dtypes
```

[53]:
```
no_of_dependents            float64
education                    object
self_employed                object
income_annum                  int64
loan_amount                   int64
loan_term                     int64
cibil_score                 float64
residential_assets_value      int64
commercial_assets_value       int64
luxury_assets_value           int64
bank_asset_value              int64
loan_status                  object
education_encoded            object
self_employed_encoded        object
loan_status_encoded          object
dtype: object
```

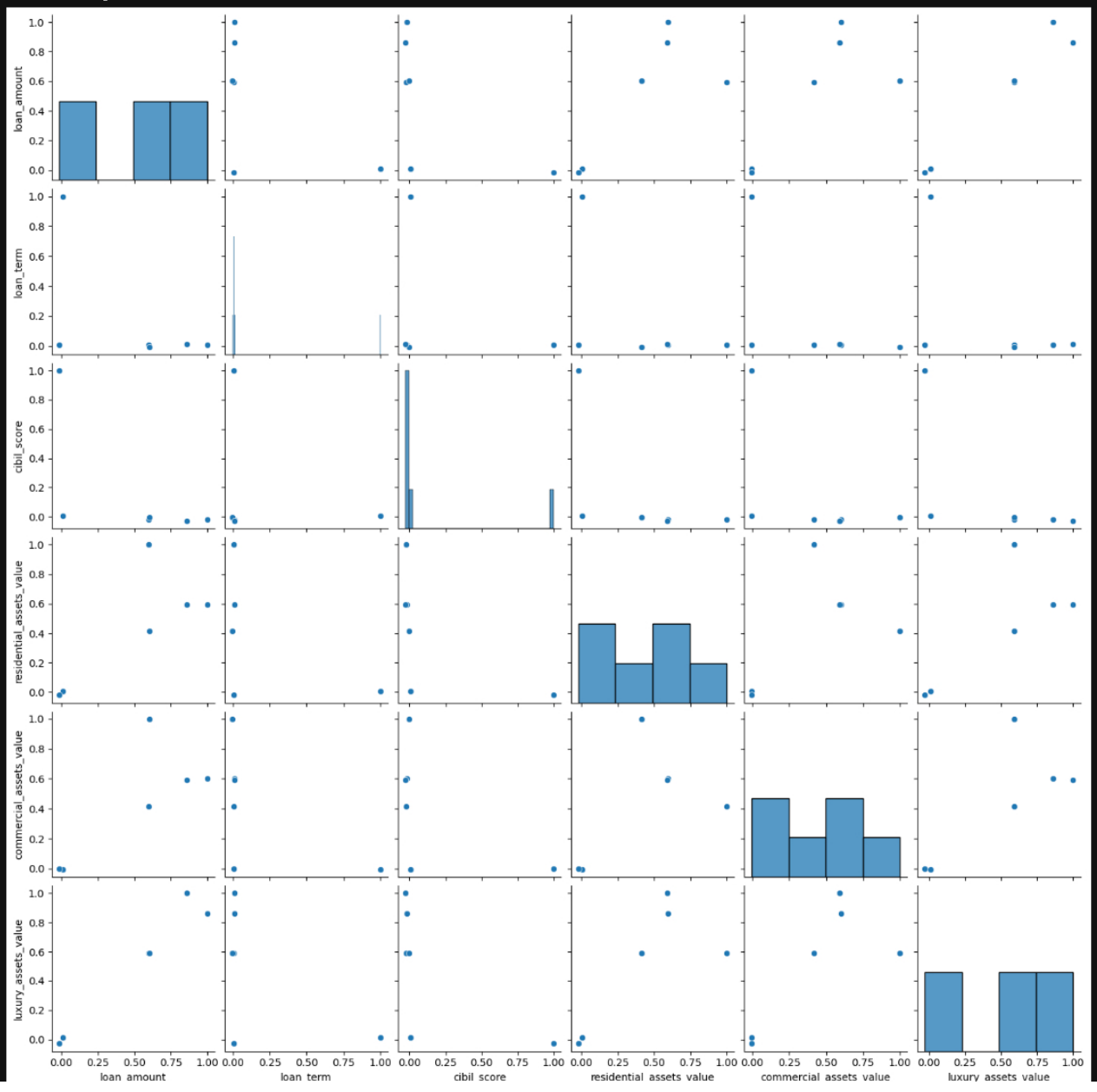
[54]:
```
data.head()
```

[54]:

loan_amount	loan_term	cibil_score	residential_assets_value	commercial_assets_value	luxury_assets_value	bank_asset_value	loan_status	education_encoded	self_employe
29900000	12	778.0	2400000	17050000	22700000	8000000	Approved	Graduate	
12200000	8	417.0	2700000	2200000	8800000	3300000	Rejected	Not Graduate	
29700000	20	506.0	7100000	4500000	33300000	12800000	Rejected	Graduate	
30700000	8	467.0	18200000	3300000	23300000	7900000	Rejected	Graduate	
24200000	20	382.0	12400000	8200000	29400000	5000000	Rejected	Not Graduate	

[63]:
```
corr=data.iloc[:,4:10].corr()
sns.pairplot(corr)
```

[63]:
```
<seaborn.axisgrid.PairGrid at 0x273f0674a70>
```



```
sns.heatmap(corr)
```

2. No multicolinearity/ Independance

```
# Take a dataframe for independent columns
df = data.iloc[:, :11]
df
```

```
VIF_df = pd.DataFrame()

VIF_df["Independant Columns"] = df.columns

VIF_df
```

```
a= df.shape[1]
a
```

```
vif_list = []

for i in range(a):
    vif = variance_inflation_factor(df.to_numpy(), i)
    vif_list.append(vif)

VIF_df["VIF"] = vif_list

VIF_df
```

```

```

```
# Create a loan-to-income ratio
df['loan_to_income_ratio'] = df["loan_amount"] / df["income_annum"]
```

```
df.drop(columns=["loan_amount"])
df.drop(columns=["income_annum"])
```

```
df = data.iloc[:, :10]
df
```

```
VIF_df = pd.DataFrame()

VIF_df["Independant Columns"] = df.columns

VIF_df
```

```
a= df.shape[1]
a
```

```
vif_list = []

for i in range(a):
    vif = variance_inflation_factor(df.to_numpy(), i)
    vif_list.append(vif)

VIF_df["VIF"] = vif_list

VIF_df
```

```
# Train-test split

X = data.drop(columns=['loan_status'])  # Replace 'target_column' with your target variable
Y = data['loan_status']

xtrain, xtest, ytrain, ytest = train_test_split(X, Y, test_size=0.25, random_state=42)

print(f'Shape of X_train: {xtrain.shape}')
print(f'Shape of X_test: {xtest.shape}')
print(f'Shape of y_train: {ytrain.shape}')
print(f'Shape of y_test: {ytest.shape}')
```

Model Training

```
logistic_reg=LogisticRegression()
logistic_reg
```

```
logistic_reg_model=logistic_reg.fit(xtrain,ytrain)
```

Evaluation of LR

Evalution on training data

```
ytrain_predict = logistic_reg_model.predict(xtrain)
```

```
accuracy = accuracy_score(ytrain,ytrain_predict)
print(f"Accuracy Score = {accuracy}")
print("*"*60)

con_mat = multilabel_confusion_matrix(ytrain,ytrain_predict)
print(f"Confusion Matrix : \n{con_mat}")
```

```
print("*"*60)

clss_report = classification_report(ytrain,ytrain_predict)
print(f"Classification report : \n{clss_report}")
print("*"*60)
```

[ ]:
```
## Evalution on testing data
ytest_pred = logistic_reg_model.predict(xtest)
```

[ ]:
```
accuracy = accuracy_score(ytest,ytest_pred)
print(f"Accuracy Score = {accuracy}")
print("*"*60)

con_mat = multilabel_confusion_matrix(ytest,ytest_pred)
print(f"Confusion Matrix : \n{con_mat}")
print("*"*60)

clss_report = classification_report(ytest,ytest_pred)
print(f"Classification report : \n{clss_report}")
print("*"*60)
```

Save model in pickle file

[ ]:
```
import pickle
```

[ ]:
```
with open("loanmodel.pkl","wb") as f:
    pickle.dump(logistic_reg_model,f)
```

[ ]:
```
with open("loanmodel.pkl","rb") as f:
    model = pickle.load(f)
```

[ ]:
```
df.columns
```

[ ]:
```
def Prediction( no_of_dependents, education, self_employed, income_annum, loan_amount, loan_term, cibil_score, residential_assets_value, commercial_assets_valu
    test_data = {
         "no_of_dependents": [no_of_dependents],
         "education": [education],
         "self_employed":[self_employed],
         "income_annum": [income_annum],
         "loan_amount":[loan_amount],
         "loan_term": [loan_term],
         "cibil_score": [cibil_score],
         "residential_assets_value":[residential_assets_value],
         "commercial_assets_value": [commercial_assets_value],
         "luxury_assets_value":[luxury_assets_value] }
     # Load the trained model
    with open("loanmodel.pkl", "rb") as f:
        final_model = pickle.load(f)

    # Predict charges
    predicted_loan_status = final_model.predict(data)[0]
    print(f"Predicted loan_status =  {predicted_loan_status:.2f}")
    #test_df = pd.DataFrame(test_data)
    #print(test_df)

    #return f"Species = {logistic_reg_model.predict(test_df)[0]}"
```

[ ]:
```
df.head(1)
```

[ ]:
```
Prediction(1,2,4,1,6,8,7,1,5,6)
```

[ ]: