

In [1]:

```
import pandas as pd
```

In [2]:

```
matplotlib inline
```

In [3]:

```
# Read data and take a look at it
data = pd.read_csv('/Users/admin/Desktop/Awamo/transactions.csv')
data.head()
```

/Users/admin/anaconda3/lib/python3.7/site-packages/IPython/core/interactiveshell.py:3057: DtypeWarning: Columns (12,13,32) have mixed types. Specify dtype option on import or set low_memory=False.

```
interactivity=interactivity, compiler=compiler, result=result)
```

Out [3]:

	End of Period	Loan Number	Region	Country Code	Country	Borrower	Gu
0	2011-04-30T00:00:00.000	IBRD00010	EUROPE AND CENTRAL ASIA	FR	France	CREDIT NATIONAL	
1	2011-05-31T00:00:00.000	IBRD00010	EUROPE AND CENTRAL ASIA	FR	France	CREDIT NATIONAL	
2	2011-06-30T00:00:00.000	IBRD00010	EUROPE AND CENTRAL ASIA	FR	France	CREDIT NATIONAL	
3	2011-07-31T00:00:00.000	IBRD00010	EUROPE AND CENTRAL ASIA	FR	France	CREDIT NATIONAL	
4	2011-08-31T00:00:00.000	IBRD00010	EUROPE AND CENTRAL ASIA	FR	France	CREDIT NATIONAL	

5 rows x 33 columns

In [4]:

```
# Take a look at the columns  
data.columns
```

Out[4]:

```
Index(['End of Period', 'Loan Number', 'Region', 'Country Code', 'Country',  
      'Borrower', 'Guarantor Country Code', 'Guarantor', 'Loan Type',  
      'Loan Status', 'Interest Rate', 'Currency of Commitment', 'Project ID',  
      'Project Name', 'Original Principal Amount', 'Cancelled Amount',  
      'Undisbursed Amount', 'Disbursed Amount', 'Repaid to IBRD',  
      'Due to IBRD', 'Exchange Adjustment', 'Borrower's Obligation',  
      'Sold 3rd Party', 'Repaid 3rd Party', 'Due 3rd Party', 'Loans Held',  
      'First Repayment Date', 'Last Repayment Date', 'Agreement Signing Date',  
      'Board Approval Date', 'Effective Date (Most Recent)',  
      'Closed Date (Most Recent)', 'Last Disbursement Date'],  
      dtype='object')
```

In [5]:

```
# Take another look to find all data types  
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 854942 entries, 0 to 854941  
Data columns (total 33 columns):  
End of Period      854942 non-null object  
Loan Number       854942 non-null object  
Region            854942 non-null object  
Country Code      854738 non-null object  
Country           854942 non-null object  
Borrower          849043 non-null object  
Guarantor Country Code 821613 non-null object  
Guarantor         794171 non-null object
```

Loan Type	854942	non-null	object
Loan Status	854942	non-null	object
Interest Rate	827914	non-null	float64
4			
Currency of Commitment	0	non-null	float64
Project ID	821645	non-null	object
Project Name	654093	non-null	object
Original Principal Amount	854942	non-null	float64
4			
Cancelled Amount	854942	non-null	float64
4			
Undisbursed Amount	854942	non-null	float64
4			
Disbursed Amount	854942	non-null	float64
4			
Repaid to IBRD	854942	non-null	float64
4			
Due to IBRD	854942	non-null	float64
4			
Exchange Adjustment	854942	non-null	float64
4			
Borrower's Obligation	854942	non-null	float64
4			
Sold 3rd Party	854942	non-null	float64
4			
Repaid 3rd Party	854942	non-null	float64
4			
Due 3rd Party	854942	non-null	float64
4			
Loans Held	854942	non-null	float64
4			
First Repayment Date	850822	non-null	object
Last Repayment Date	850924	non-null	object
Agreement Signing Date	844133	non-null	object
Board Approval Date	854942	non-null	object
Effective Date (Most Recent)	849864	non-null	object
Closed Date (Most Recent)	854148	non-null	object
Last Disbursement Date	480299	non-null	object

dtypes: float64(14), object(19)
memory usage: 215.2+ MB

In [6]:

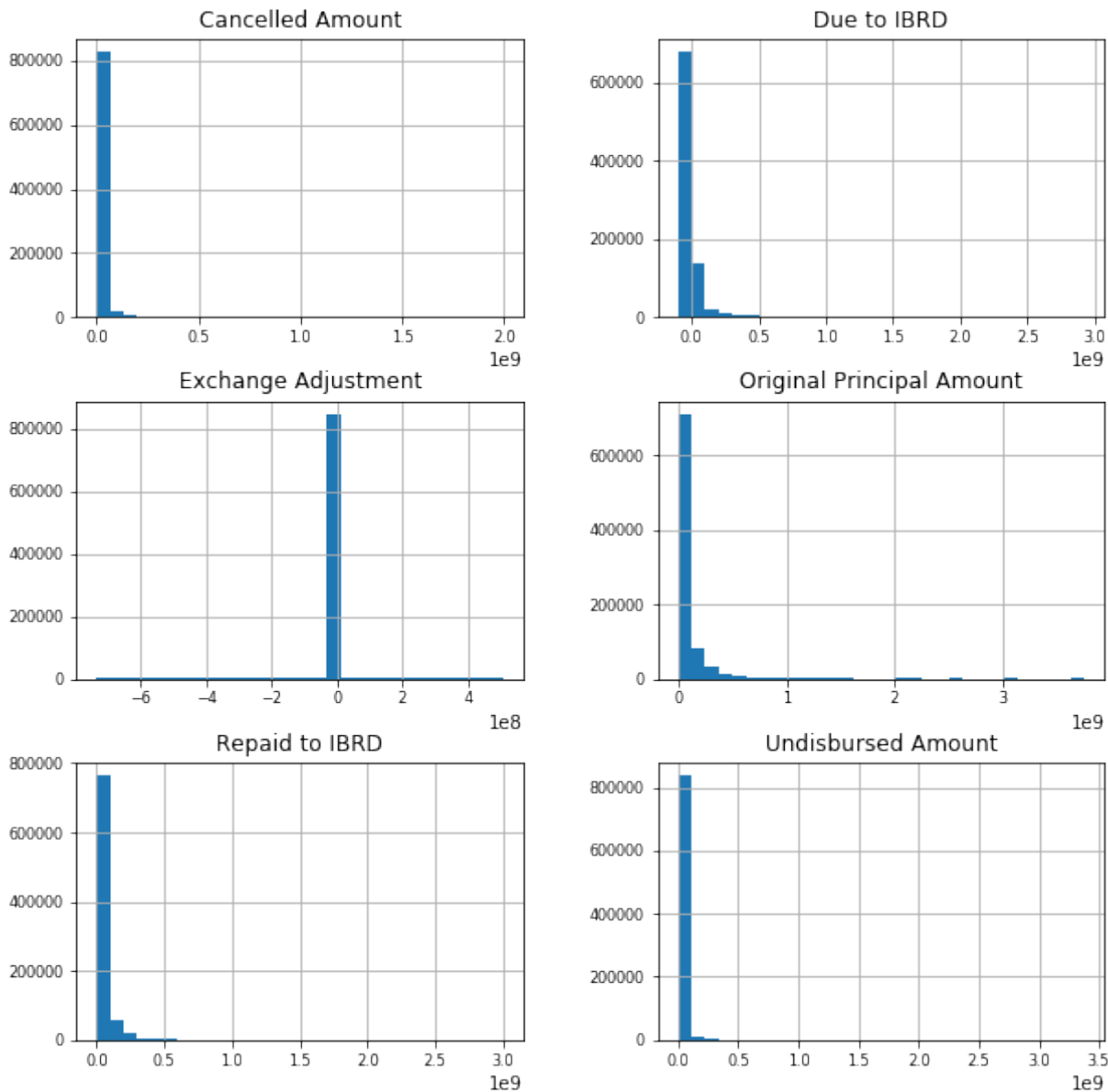
```
# Take another look using describe
data.describe()
```

Out [6]:

	Interest Rate	Currency of Commitment	Original Principal Amount	Cancelled Amount	Undisbur Amc
count	827914.000000	0.0	8.549420e+05	8.549420e+05	8.549420e
mean	4.705753	NaN	7.723954e+07	9.714845e+06	7.680076e
std	3.384497	NaN	1.490977e+08	4.478135e+07	5.391557e
min	0.000000	NaN	0.000000e+00	0.000000e+00	-1.00000e
25%	0.830000	NaN	1.000000e+07	0.000000e+00	0.000000e
50%	5.580000	NaN	3.000000e+07	2.551600e+02	0.000000e
75%	7.350000	NaN	8.521814e+07	3.517629e+06	0.000000e
max	17.000000	NaN	3.750000e+09	1.995000e+09	3.379011e

In [7]:

```
df = data[['Original Principal Amount', 'Cancelled Amount', 'Repaid  
df.hist(figsize=(10, 10), bins=30, xlabelsize=8, ylabelsize=8);
```



In []:

In [1]:

In [2]:

In [3]:

```
/Users/admin/anaconda3/lib/python3.7/site-packages/IPython/core/interactiveshell.py:3057: DtypeWarning: Columns (12,13,32) have mixed types. Specify dtype option on import or set low_memory=False.  
    interactivity=interactivity, compiler=compiler, result=result)
```

Out [3]:

	End of Period	Loan Number	Region	Country Code	Country	Borrower	Guarantor Country Code	Guarant
0	2011-04-30T00:00:00.000	IBRD00010	EUROPE AND CENTRAL ASIA	FR	France	CREDIT NATIONAL	FR	Fran
1	2011-05-31T00:00:00.000	IBRD00010	EUROPE AND CENTRAL	FR	France	CREDIT NATIONAL	FR	Fran

In [4]:

Out [4]:

```
Index(['End of Period', 'Loan Number', 'Region', 'Country Code', 'Country',  
      'Borrower', 'Guarantor Country Code', 'Guarantor', 'Loan Type',  
      'Loan Status', 'Interest Rate', 'Currency of Commitment', 'Project ID',  
      'Project Name ', 'Original Principal Amount', 'Cancelled Amount',  
      'Undisbursed Amount', 'Disbursed Amount', 'Repaid to IBRD',  
      'Due to IBRD', 'Exchange Adjustment', 'Borrower's Obligation',  
      'Sold 3rd Party', 'Repaid 3rd Party', 'Due 3rd Party', 'Loans Held',  
      'First Repayment Date', 'Last Repayment Date', 'Agreement Signing Date',  
      'Board Approval Date', 'Effective Date (Most Recent)'].
```


In [5]:

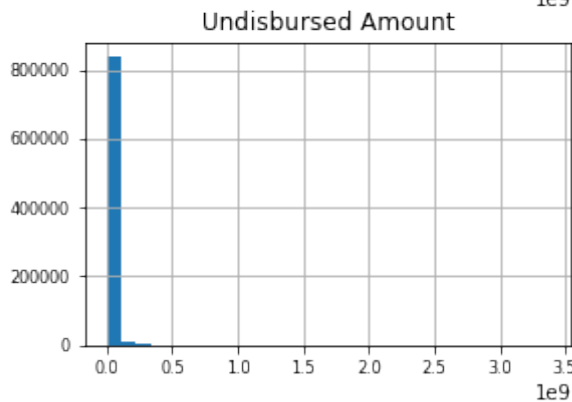
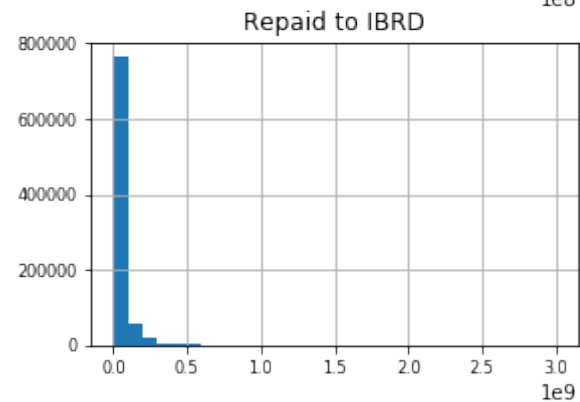
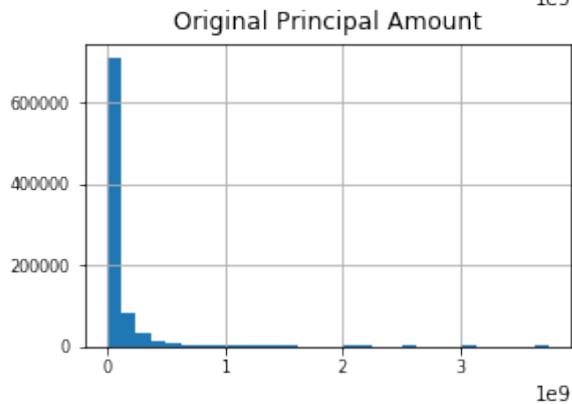
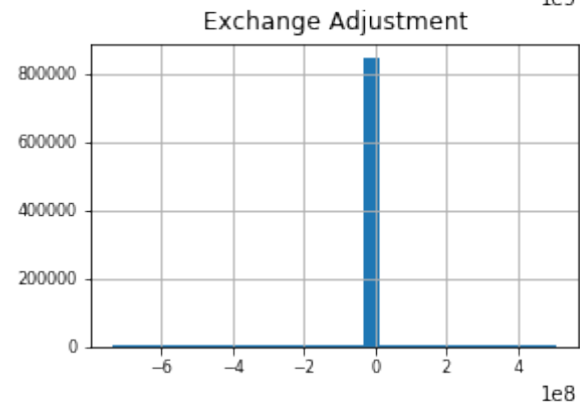
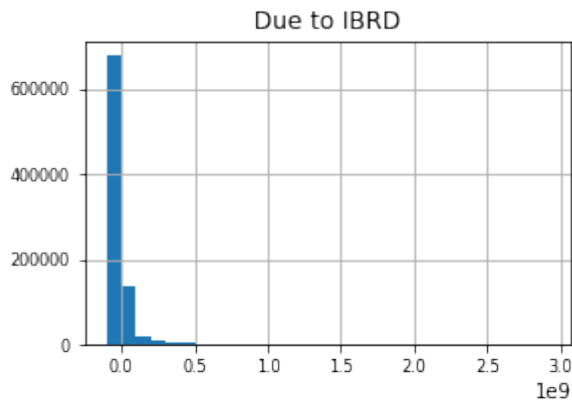
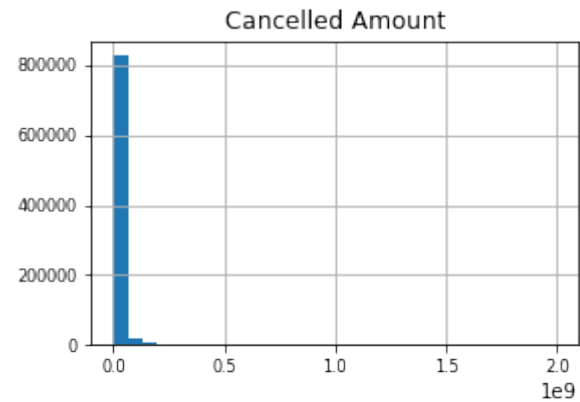
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 854942 entries, 0 to 854941
Data columns (total 33 columns):
End of Period                854942 non-null object
Loan Number                  854942 non-null object
Region                       854942 non-null object
Country Code                 854738 non-null object
Country                      854942 non-null object
Borrower                     849043 non-null object
Guarantor Country Code      821613 non-null object
Guarantor                    794171 non-null object
Loan Type                    854942 non-null object
Loan Status                  854942 non-null object
Interest Rate                827914 non-null float64
4
Currency of Commitment       0 non-null float64
Project ID                   821645 non-null object
Project Name                  654093 non-null object
Original Principal Amount    854942 non-null float64
^
```

In [6]:

Out [6]:

	Interest Rate	Currency of Commitment	Original Principal Amount	Cancelled Amount	Undisbur Amc
count	827914.000000	0.0	8.549420e+05	8.549420e+05	8.549420e
mean	4.705753	NaN	7.723954e+07	9.714845e+06	7.680076e
std	3.384497	NaN	1.490977e+08	4.478135e+07	5.391557e
min	0.000000	NaN	0.000000e+00	0.000000e+00	-1.00000e
25%	0.830000	NaN	1.000000e+07	0.000000e+00	0.000000e
50%	5.580000	NaN	3.000000e+07	2.551600e+02	0.000000e
75%	7.350000	NaN	8.521814e+07	3.517629e+06	0.000000e
max	17.000000	NaN	3.750000e+09	1.995000e+09	3.379011e

In [7]:



In []: