In [3]:
```
import pandas as pd
%matplotlib inline
```

In [4]:
```
mydata = pd.read_csv("marketing_data.csv")
```

In [5]:
```
mydata
```

Out[5]:

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FREQUENCY	PURCHASES_INSTALLM
0	C10001	40.900749	0.818182	95.40	0.00	95.40	0.000000	0.166667	0.000000	
1	C10002	3202.467416	0.909091	0.00	0.00	0.00	6442.945483	0.000000	0.000000	
2	C10003	2495.148862	1.000000	773.17	773.17	0.00	0.000000	1.000000	1.000000	
3	C10004	1666.670542	0.636364	1499.00	1499.00	0.00	205.788017	0.083333	0.083333	
4	C10005	817.714335	1.000000	16.00	16.00	0.00	0.000000	0.083333	0.083333	
...	...	...	...	...	...	...	...	...	...	
8945	C19186	28.493517	1.000000	291.12	0.00	291.12	0.000000	1.000000	0.000000	
8946	C19187	19.183215	1.000000	300.00	0.00	300.00	0.000000	1.000000	0.000000	
8947	C19188	23.398673	0.833333	144.40	0.00	144.40	0.000000	0.833333	0.000000	
8948	C19189	13.457564	0.833333	0.00	0.00	0.00	36.558778	0.000000	0.000000	
8949	C19190	372.708075	0.666667	1093.25	1093.25	0.00	127.040008	0.666667	0.666667	

8950 rows × 18 columns

In [6]:
```
mydata[:5]
```

Out[6]:

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FREQUENCY	PURCHASES_INSTALLMENT
0	C10001	40.900749	0.818182	95.40	0.00	95.4	0.000000	0.166667	0.000000	
1	C10002	3202.467416	0.909091	0.00	0.00	0.0	6442.945483	0.000000	0.000000	
2	C10003	2495.148862	1.000000	773.17	773.17	0.0	0.000000	1.000000	1.000000	
3	C10004	1666.670542	0.636364	1499.00	1499.00	0.0	205.788017	0.083333	0.083333	
4	C10005	817.714335	1.000000	16.00	16.00	0.0	0.000000	0.083333	0.083333	

In [7]:
```
type(mydata)
```

Out[7]:
```
pandas.core.frame.DataFrame
```

In [33]:
```
len(mydata)
```

Out[33]:
```
8950
```

In [8]:
```
mydata.isnull()
```

Out[8]:

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FREQUENCY	PURCHASES_INSTALLMENT
0	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	
8945	False	False	False	False	False	False	False	False	False	
8946	False	False	False	False	False	False	False	False	False	
8947	False	False	False	False	False	False	False	False	False	
8948	False	False	False	False	False	False	False	False	False	
8949	False	False	False	False	False	False	False	False	False	

8950 rows × 18 columns

In [9]:
```
mydata.MINIMUM_PAYMENTS.isnull()
```

Out[9]:
```
0       False
1       False
2       False
3        True
4       False
        ...
8945    False
8946     True
8947    False
8948    False
8949    False
Name: MINIMUM_PAYMENTS, Length: 8950, dtype: bool
```

In [10]:
```
mydata[mydata.MINIMUM_PAYMENTS.isnull()]
```

Out[10]:

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FREQUENCY	PURCHASES_INSTALLM
3	C10004	1666.670542	0.636364	1499.00	1499.00	0.0	205.788017	0.083333	0.083333	
45	C10047	2242.311686	1.000000	437.00	97.00	340.0	184.648692	0.333333	0.083333	
47	C10049	3910.111237	1.000000	0.00	0.00	0.0	1980.873201	0.000000	0.000000	
54	C10056	6.660517	0.636364	310.00	0.00	310.0	0.000000	0.666667	0.000000	
55	C10057	1311.995984	1.000000	1283.90	1283.90	0.0	0.000000	0.250000	0.250000	
...	...	...	...	...	...	...	...	...	...	
8919	C19160	14.524779	0.333333	152.00	152.00	0.0	0.000000	0.333333	0.333333	
8929	C19170	371.527312	0.333333	0.00	0.00	0.0	1465.407927	0.000000	0.000000	
8935	C19176	183.817004	1.000000	465.90	0.00	465.9	0.000000	1.000000	0.000000	
8944	C19185	193.571722	0.833333	1012.73	1012.73	0.0	0.000000	0.333333	0.333333	
8946	C19187	19.183215	1.000000	300.00	0.00	300.0	0.000000	1.000000	0.000000	

313 rows × 18 columns

In [11]:
```
print(mydata.CUST_ID[0])
```

```
C10001
```

In [12]:
```
mydata
```

Out[12]:

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FREQUENCY	PURCHASES_INSTALLM
0	C10001	40.900749	0.818182	95.40	0.00	95.40	0.000000	0.166667	0.000000	
1	C10002	3202.467416	0.909091	0.00	0.00	0.00	6442.945483	0.000000	0.000000	
2	C10003	2495.148862	1.000000	773.17	773.17	0.00	0.000000	1.000000	1.000000	
3	C10004	1666.670542	0.636364	1499.00	1499.00	0.00	205.788017	0.083333	0.083333	
4	C10005	817.714335	1.000000	16.00	16.00	0.00	0.000000	0.083333	0.083333	
...	...	...	...	...	...	...	...	...	...	
8945	C19186	28.493517	1.000000	291.12	0.00	291.12	0.000000	1.000000	0.000000	
8946	C19187	19.183215	1.000000	300.00	0.00	300.00	0.000000	1.000000	0.000000	
8947	C19188	23.398673	0.833333	144.40	0.00	144.40	0.000000	0.833333	0.000000	
8948	C19189	13.457564	0.833333	0.00	0.00	0.00	36.558778	0.000000	0.000000	
8949	C19190	372.708075	0.666667	1093.25	1093.25	0.00	127.040008	0.666667	0.666667	

8950 rows × 18 columns

In [13]:
```
print(mydata.MINIMUM_PAYMENTS[3])
```

```
nan
```

In [14]:
```
meta=pd.read_csv("bank_transactions.csv",encoding='utf-8')
```

In [15]:
```
meta
```

Out[15]:

	TransactionID	CustomerID	CustomerDOB	CustGender	CustLocation	CustAccountBalance	TransactionDate	TransactionTime	TransactionAmount (INR)
0	T1	C5841053	10/1/94	F	JAMSHEDPUR	17819.05	2/8/16	143207	25.0
1	T2	C2142763	4/4/57	M	JHAJJAR	2270.69	2/8/16	141858	27999.0
2	T3	C4417068	26/11/96	F	MUMBAI	17874.44	2/8/16	142712	459.0
3	T4	C5343380	14/9/73	F	MUMBAI	866503.21	2/8/16	142714	2060.0
4	T5	C9031234	24/3/88	F	NAVI MUMBAI	6714.43	2/8/16	181156	1762.5
...	...	...	...	...	...	...	...	...	...
1048562	T1048563	C8020229	8/4/90	M	NEW DELHI	7635.19	18/9/16	184824	799.0
1048563	T1048564	C6459278	20/2/92	M	NASHIK	27311.42	18/9/16	183734	460.0
1048564	T1048565	C6412354	18/5/89	M	HYDERABAD	221757.06	18/9/16	183313	770.0
1048565	T1048566	C6420483	30/8/78	M	VISAKHAPATNAM	10117.87	18/9/16	184706	1000.0
1048566	T1048567	C8337524	5/3/84	M	PUNE	75734.42	18/9/16	183222	1166.0

1048567 rows × 9 columns

In [27]:
```
meta.CustomerID.value_counts()
```

Out[27]:
```
C5533885    6
C6624352    6
C1736254    6
C6735477    6
C1113684    6
           ..
C8716232    1
C4474828    1
C3341662    1
C6821983    1
C6420483    1
Name: CustomerID, Length: 884265, dtype: int64
```

In [22]:
```
import numpy as np
import datetime
```

In [24]:
```
date1 = pd.Series(pd.date_range('2012-1-1 12:00:00', periods=7, freq='M'))
df = pd.DataFrame(dict(date_given=date1))
print(df)
```

```
  date_given
0 2012-01-31 12:00:00
1 2012-02-29 12:00:00
2 2012-03-31 12:00:00
3 2012-04-30 12:00:00
4 2012-05-31 12:00:00
5 2012-06-30 12:00:00
6 2012-07-31 12:00:00
```

In [26]:
```
df['day_of_week_in_number'] = df['date_given'].dt.dayofweek
```

In [27]:
```
df
```

Out[27]:

	date_given	day_of_week_in_number
0	2012-01-31 12:00:00	1
1	2012-02-29 12:00:00	2
2	2012-03-31 12:00:00	5
3	2012-04-30 12:00:00	0
4	2012-05-31 12:00:00	3
5	2012-06-30 12:00:00	5
6	2012-07-31 12:00:00	1

In [28]:
```
from dateutil.relativedelta import relativedelta
from datetime import date
```

In [29]:
```
date1 = pd.Series(pd.date_range('2012-1-1 12:00:00', periods=7, freq='M'))
date2 = pd.Series(pd.date_range('2013-3-11 21:45:00', periods=7, freq='W'))
df = pd.DataFrame(dict(Start_date = date1, End_date = date2))
```

In [30]:
```
df
```

Out[30]:

	Start_date	End_date
0	2012-01-31 12:00:00	2013-03-17 21:45:00
1	2012-02-29 12:00:00	2013-03-24 21:45:00
2	2012-03-31 12:00:00	2013-03-31 21:45:00
3	2012-04-30 12:00:00	2013-04-07 21:45:00
4	2012-05-31 12:00:00	2013-04-14 21:45:00
5	2012-06-30 12:00:00	2013-04-21 21:45:00
6	2012-07-31 12:00:00	2013-04-28 21:45:00

In [31]:
```
df['diff_seconds'] = df['End_date'] - df['Start_date']
df['diff_seconds']=df['diff_seconds']/np.timedelta64(1,'s')
```

In [32]:
```
df
```

Out[32]:

	Start_date	End_date	diff_seconds
0	2012-01-31 12:00:00	2013-03-17 21:45:00	35545500.0
1	2012-02-29 12:00:00	2013-03-24 21:45:00	33644700.0
2	2012-03-31 12:00:00	2013-03-31 21:45:00	31571100.0
3	2012-04-30 12:00:00	2013-04-07 21:45:00	29583900.0
4	2012-05-31 12:00:00	2013-04-14 21:45:00	27510300.0
5	2012-06-30 12:00:00	2013-04-21 21:45:00	25523100.0
6	2012-07-31 12:00:00	2013-04-28 21:45:00	23449500.0

In [33]:
```
df['diff_hours'] = df['End_date'] - df['Start_date']
df['diff_hours']=df['diff_hours']/np.timedelta64(1,'h')
```

In [34]:
```
df
```

Out[34]:

	Start_date	End_date	diff_seconds	diff_hours
0	2012-01-31 12:00:00	2013-03-17 21:45:00	35545500.0	9873.75
1	2012-02-29 12:00:00	2013-03-24 21:45:00	33644700.0	9345.75
2	2012-03-31 12:00:00	2013-03-31 21:45:00	31571100.0	8769.75
3	2012-04-30 12:00:00	2013-04-07 21:45:00	29583900.0	8217.75
4	2012-05-31 12:00:00	2013-04-14 21:45:00	27510300.0	7641.75
5	2012-06-30 12:00:00	2013-04-21 21:45:00	25523100.0	7089.75
6	2012-07-31 12:00:00	2013-04-28 21:45:00	23449500.0	6513.75

In [39]:
```
import nbconvert
```

In [ ]:
```
```