

Before we start analyzing the data, we should initially look at them. By `describe()` method we can make initial descriptive statistics. In our case, it tells us, that we have $1.3e5$ records.

Descriptive statistics and analysis

First of all it is important to see the general statistics of the transactions dataset

```
] : transactions.describe()
```

```
] :
```

	client_id	code	type	sum
count	1.300390e+05	130039.000000	130039.000000	1.300390e+05
mean	5.086859e+07	5594.629996	2489.372135	-1.812909e+04
std	2.872854e+07	606.087084	2253.296578	5.584445e+05
min	2.289900e+04	742.000000	1000.000000	-4.150030e+07
25%	2.577174e+07	5211.000000	1030.000000	-2.244916e+04
50%	5.235837e+07	5641.000000	1110.000000	-5.502490e+03
75%	7.506302e+07	6010.000000	2370.000000	-1.122960e+03
max	9.999968e+07	9402.000000	8145.000000	6.737747e+07

It tells us, that we have $1.3e5$ records. Also we can get some insights about the amount of transactions. It is obvious, that since the median (50% percentile) is negative, more transactions lead to cash outflows.

Also, we can get some insights into the number of transactions. It is obvious, that since the median (50% percentile) is negative, more transactions lead to cash outflows. As expected from the insights in the Descriptive statistics and analysis section, most transactions are money outflow. The overall distribution of transaction sums can be viewed using a boxplot.

```
: data['sum'].plot.box(figsize=(20,20))  
: <matplotlib.axes._subplots.AxesSubplot at 0x7fd5b15c50>
```

