Code:

```
sed 's/ /,/g; s/-,-,//; s/,+0000]//; s/\[//; s/Apr/04/;s/\/04\//-04-/; s/:/,/; s/"//g; s/,HTTP\/[0-9,X].[0-9,X]//;s/GET,//'  wc_day91_1.log > wc_day91_1.csv
log_df = pd.read_csv("/home/datascience/wc_day91_1.csv", names=['ClientID', 'Date', 'Time', 'URL', 'ResponseCode', 'Size'], na_values=['-'])
```

Code for running commands from Q3

```
----------------------------------------------
is200df = log_df[log_df['ResponseCode'] == '200']
finaldf = is200df[is200df['URL'].str.endswith('jpg') | is200df['URL'].str.endswith('jpeg') | is200df['URL'].str.endswith('gif')]
finaldf['Size'].mean()
finaldf['Size'].std()
```

Result from running commands from Q3

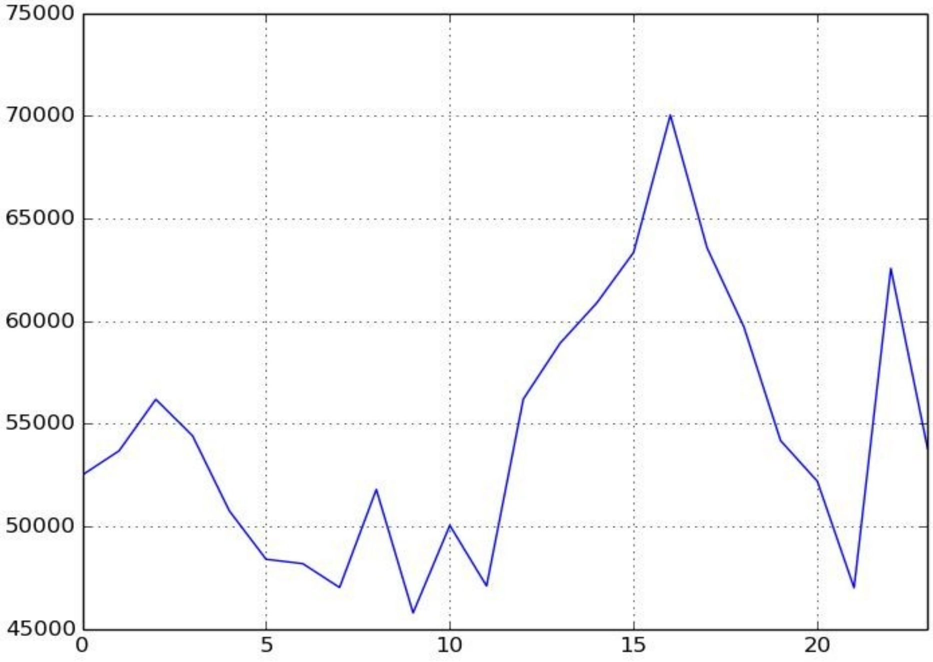
```
----------------------------------------------
3219.9428340117138
6302.9825349485855
```

Conclusion : The results are different from the previous sample; i.e. the one in Q3.

Code for running commands from Q4

```
---------------------------------------------
log_df['DateTime'] = pd.to_datetime(log_df.apply(lambda row: row['Date'] + ' ' + row['Time'], axis=1))
hour_grouped = log_df.groupby(lambda row: log_df['DateTime'][row].hour)
hour_grouped.size().plot()
show()
```

Result from running commands from Q4



Conclusion : The structure of the graph is similar to that of previous sample; i.e. the one in Q4.