Answer6:

```
cat wc_day91_2.log | 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_2.csv
```

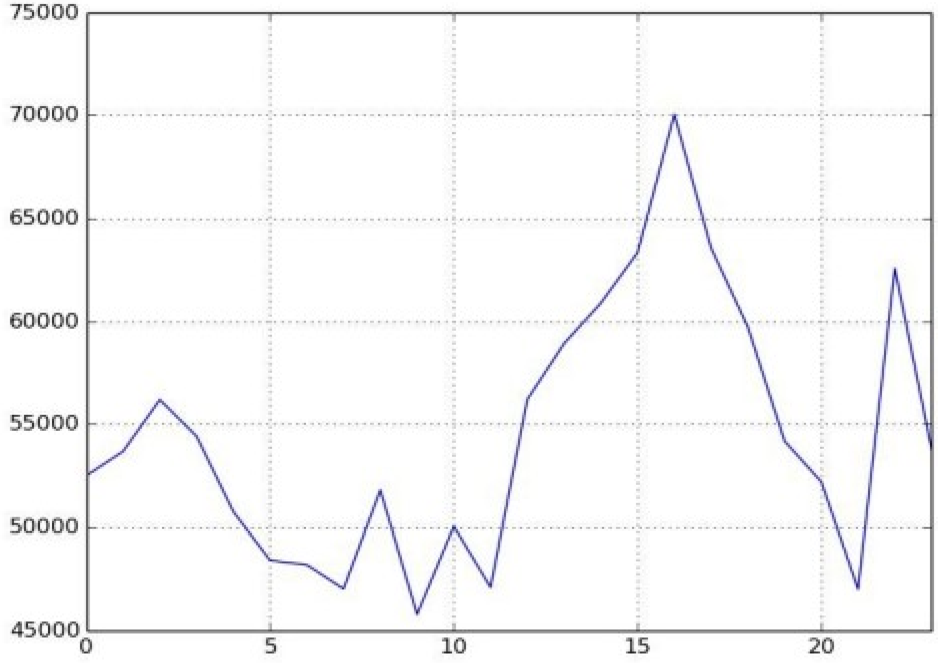
```
from pylab import *
import pandas as pd
log_df = pd.read_csv("/home/datascience/Downloads/wc_day91_2.csv", names=['ClientID', 'Date', 'Time', 'URL', 'ResponseCode', 'Size'], na_values=['-'])
```

```
******************Commands to run Question3 *************************************
resp200df = log_df[log_df['ResponseCode'] == '200']
resultdf = resp200df[resp200df['URL'].str.endswith('jpg') | resp200df['URL'].str.endswith('jpeg') | resp200df['URL'].str.endswith('gif')]
resultdf['Size'].mean()
resultdf['Size'].std()

***********************
```

Mean: 3219.9428340117138
Standard Deviation: 6302.9825349485855

The mean and Standard deviation is different than the previous data set.

```
******************Commands to run Question4 on this dataset***************************

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: If we look from here both the graph are similar. The only difference is sample size.