1. Statistics:
   Unique fields

time	7542133
duration	385323
client_address	2214
result_code	77
bytes	484519
request_method	220
URL	196762
user	1818
hierarchy_code	36926
type	162

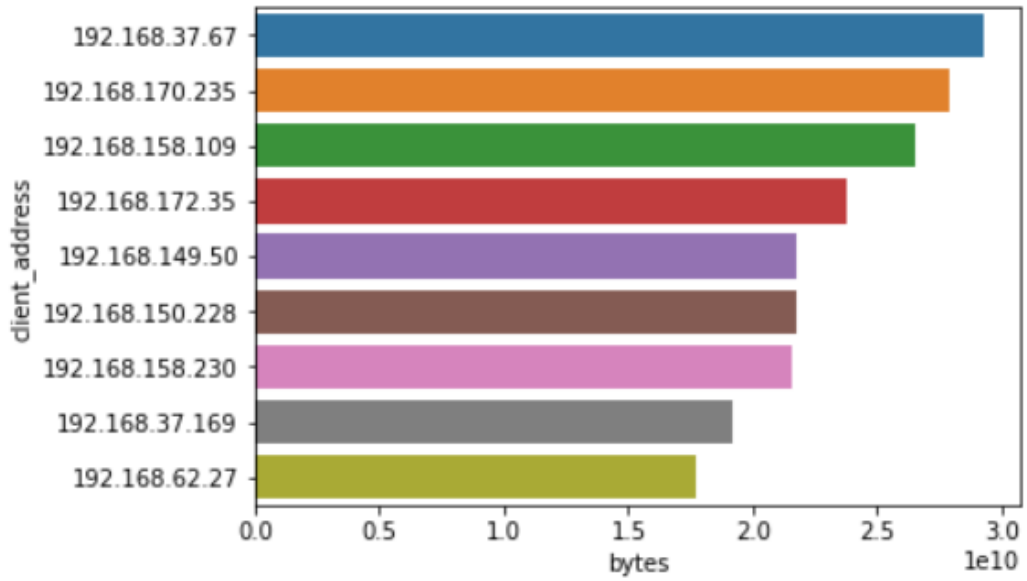
   Observations

	duration	**bytes**
count	8892475	8892475
mean	29747.26	299204.5
std	147602.4	12231316
min	0	0
25%	0	403
50%	26	3084
75%	3825	4119
max	86399993	5.26E+09

   These values indicate that out of the 2214 unique client addresses 25% didn't connect to the network.

   75% of the client addresses had a max transaction of 4119 bytes only.

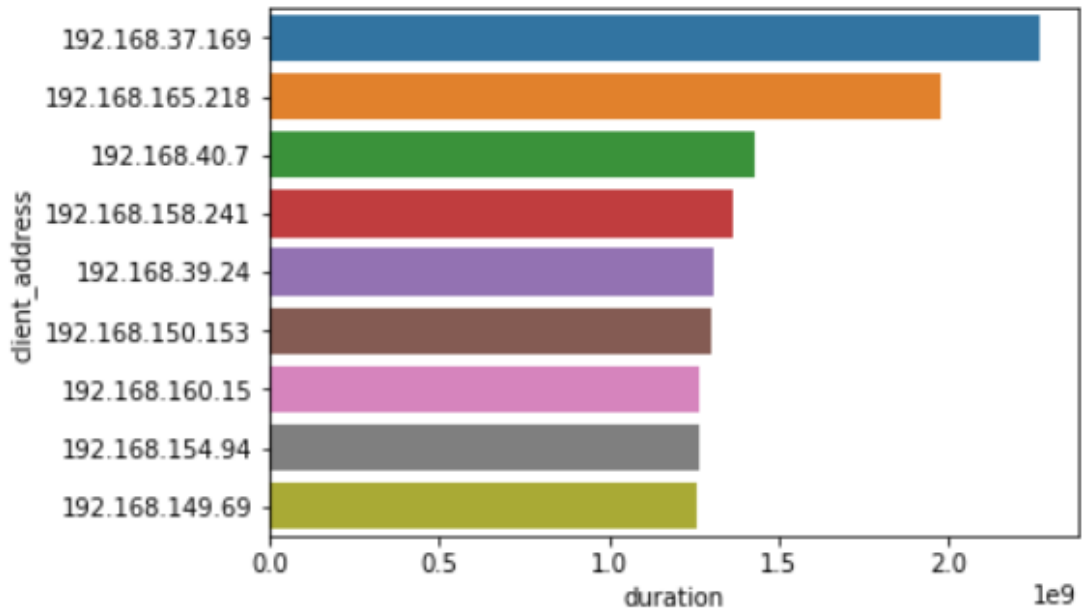
2. Client address vs amount of data transacted by each client. We can see that **192.168.37.67** has the highest amount of data transacted.



3. Client address vs duration of connectivity including every transaction.

192.168.37.169 was connected for the longest period of time.

```
<matplotlib.axes._subplots.AxesSubplot at 0x1e92ddb8dd8>
```

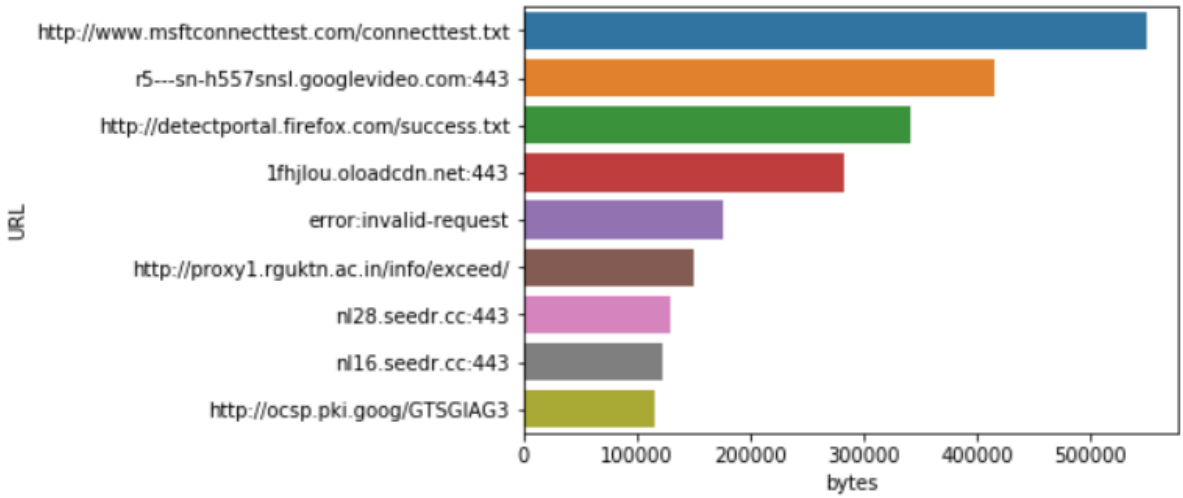


4. Correlation matrix between duration and bytes. We can see that they are not correlated very much. So we can say that the duration of connectivity doesn't mean that there will be high amount of data transaction.

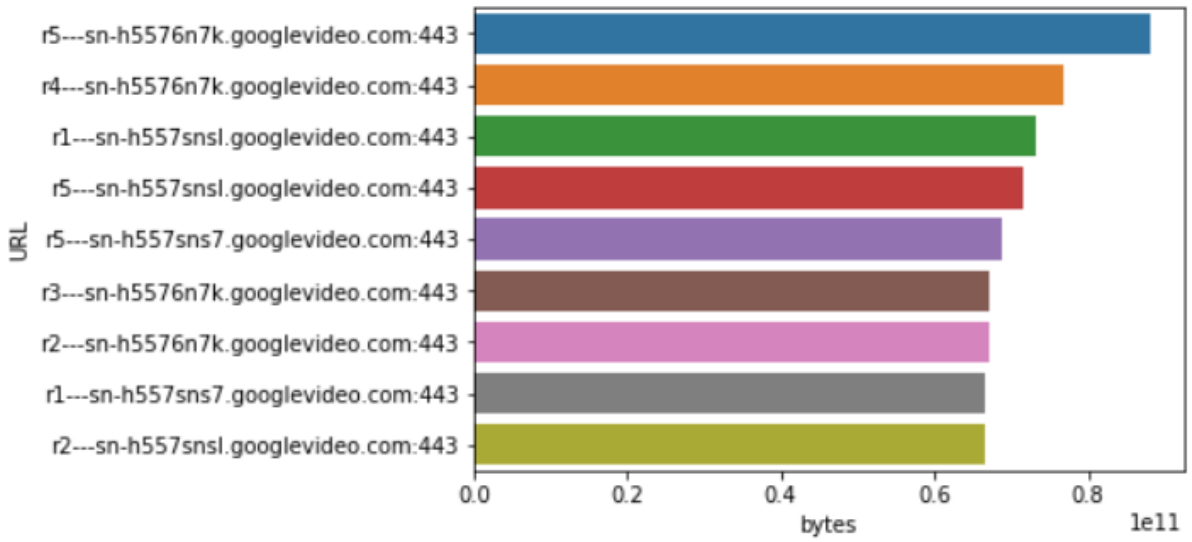
	duration	bytes
duration	1	0.110386
bytes	0.110386	1

5. URL analysis –

   - Top URL hits.
     Msftconnecttest.com has the highest hits by the count of bytes.

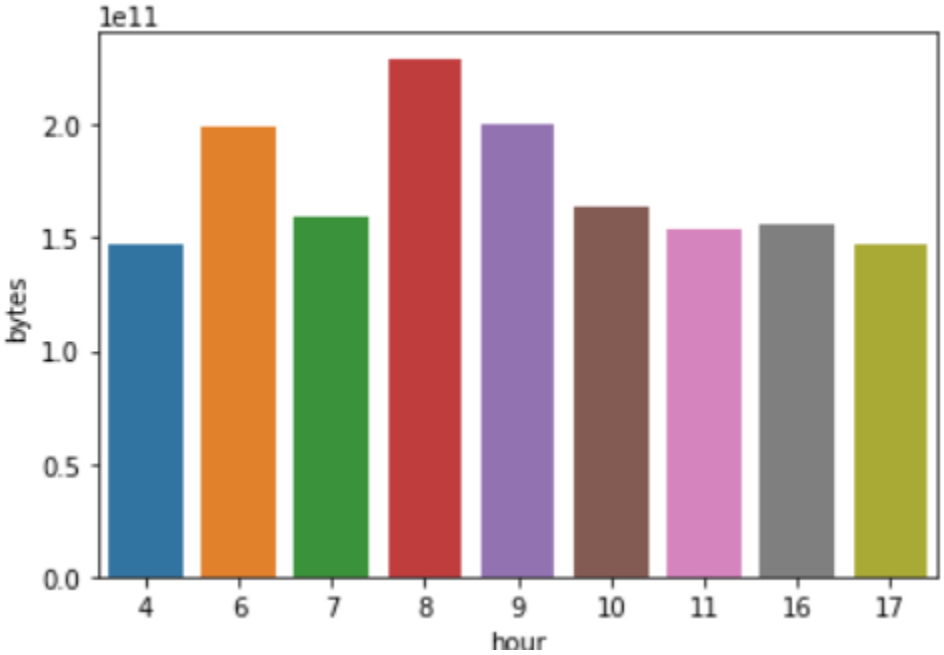


- Top URL data
  googlevideo.com consumes the highest amount of data



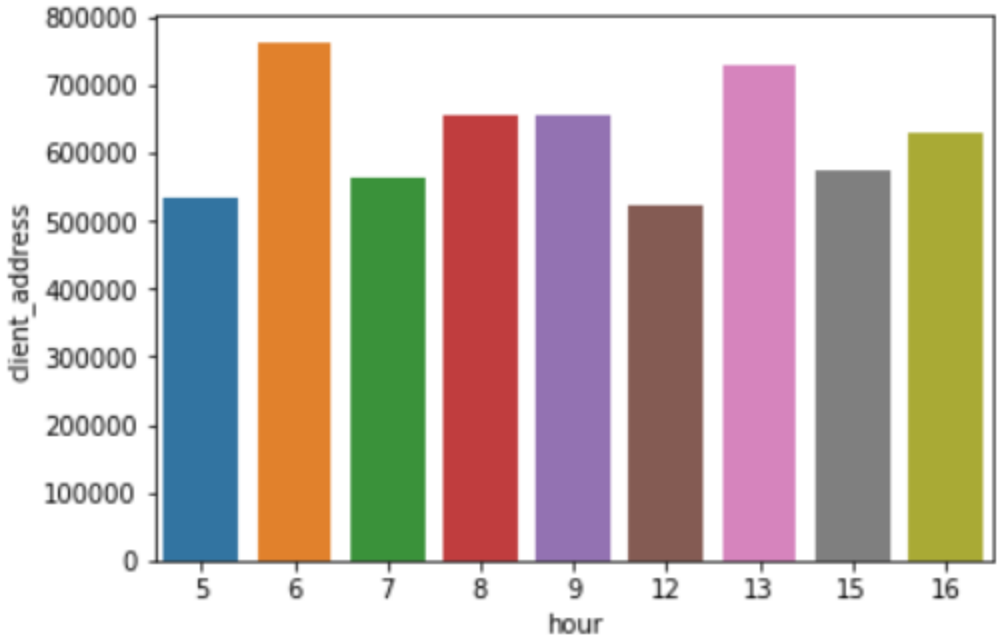
6. Maximum Usage on the basis of time. Hour in 24 hour format

   8 am and 9 am are the peak times for data usage(sum)

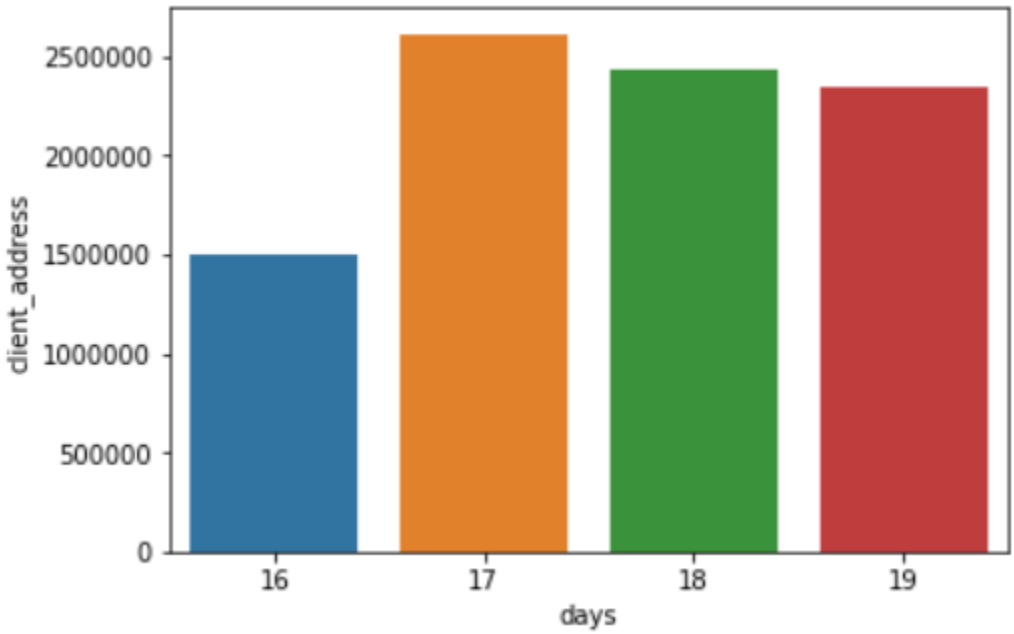


7. Hour vs Client_address(count)

   The highest hits of traffic come at around 6 am

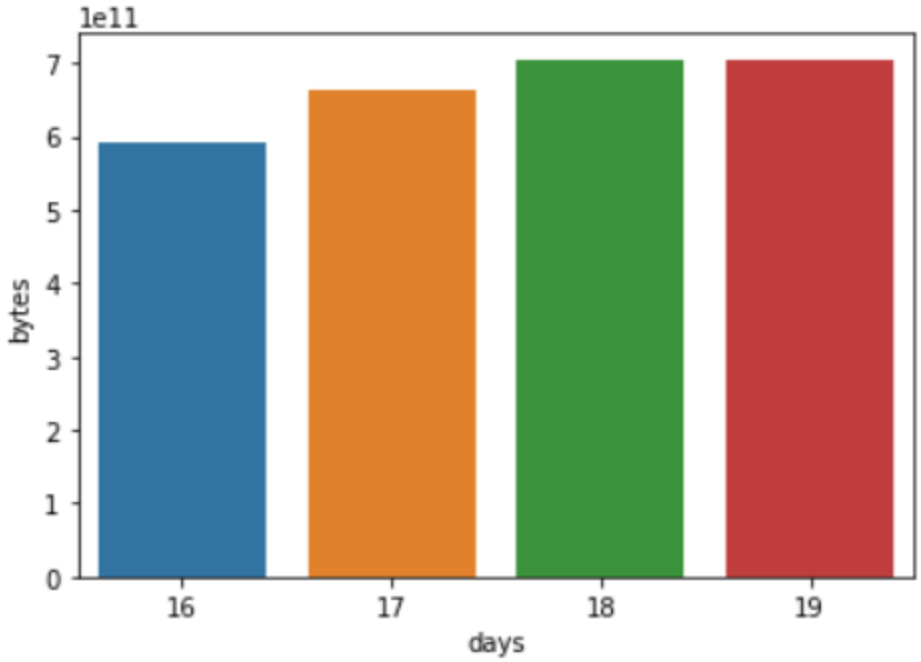


8. Top day usage by hits(count)
   17th day a month has the highest hits of client address. Meaning more traffic



9. Data usage by day wise

   This shows that the 18th and 19th Days have the highest data usage.


1e11
7
6
5
4
3
2
1
0
bytes
16
17
18
19
days