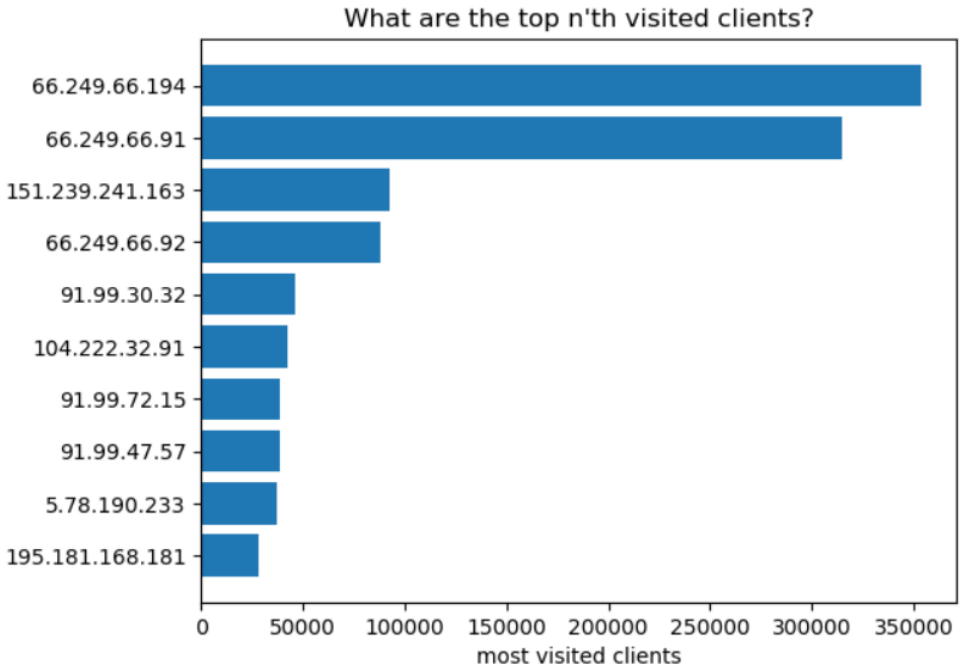
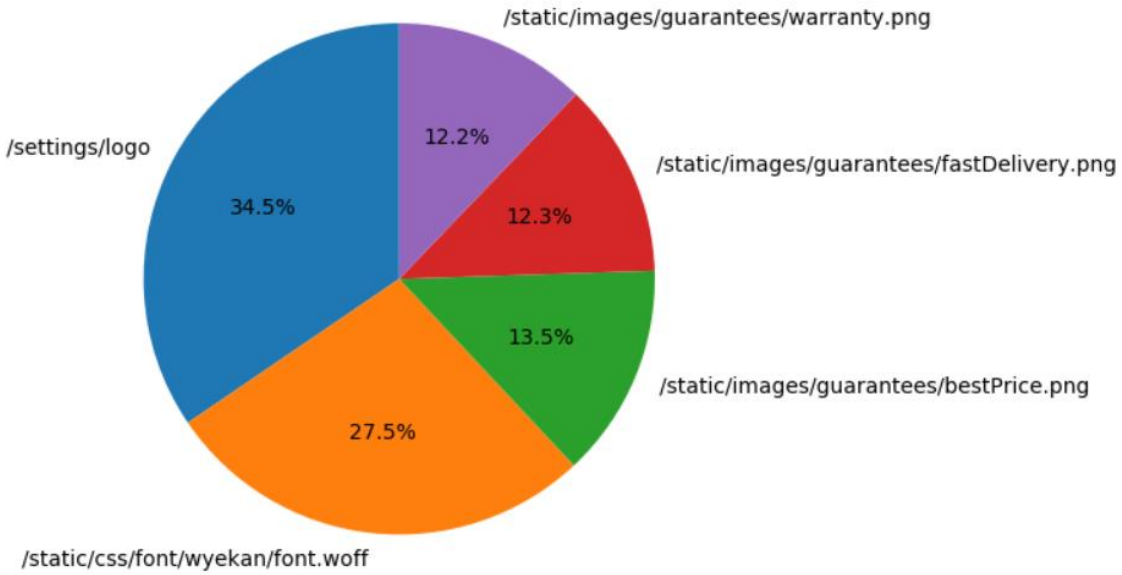
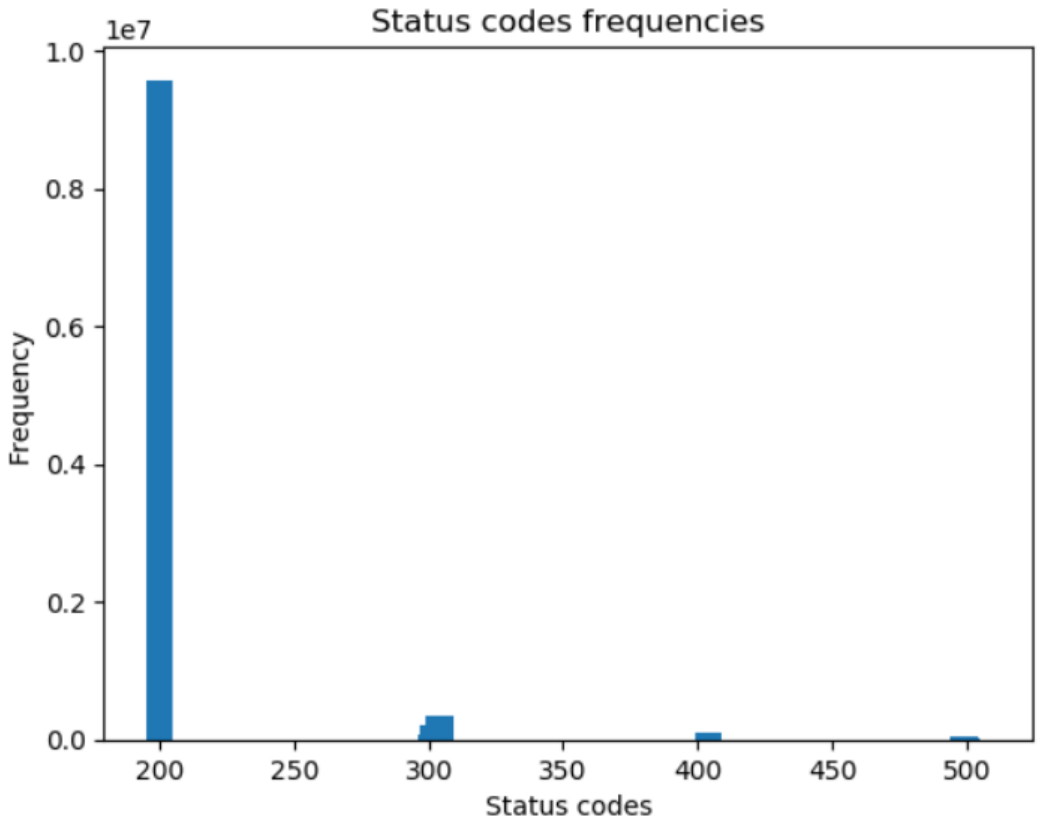
نمودار سوال اول)



پای چارت سوال دو)



نمودار سوال چهارم)



	status_code	frequency
0	200	9579824
1	304	340228
2	302	199835
3	404	105011
4	301	67552
5	499	50852
6	500	14266
7	403	5634
8	502	798
9	401	323
10	400	318
11	408	112
12	504	103
13	405	6
14	206	3

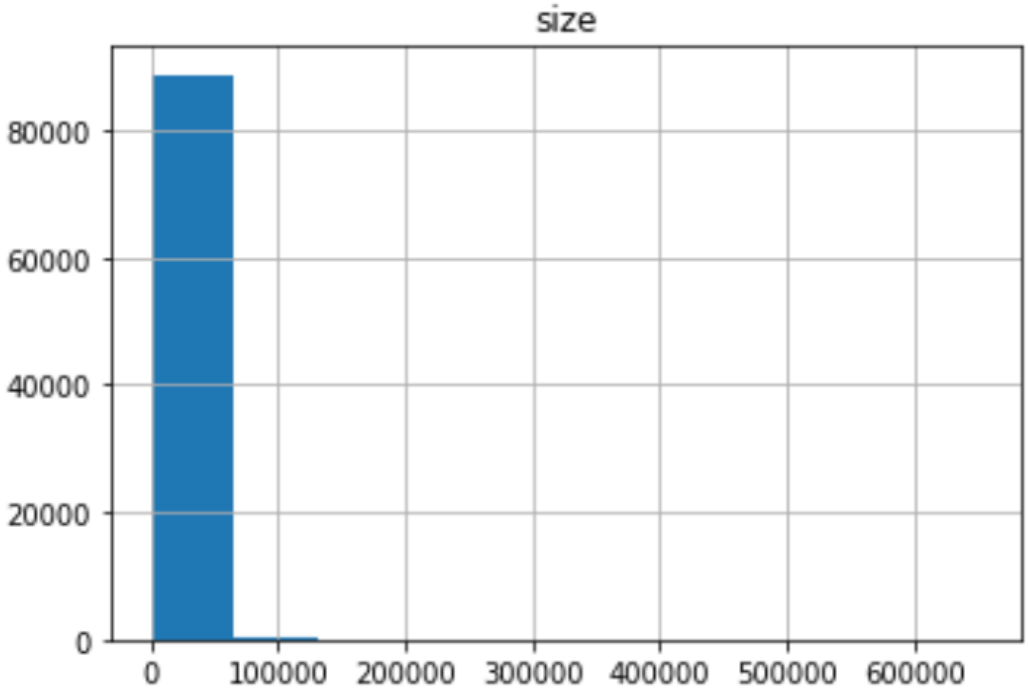
چارت سوال چهارم قسمت دوم)

hour	4xx	5xx
0	6021	17
1	4215	10
2	2394	8
3	1388	26
4	1871	3
5	1683	10
6	2036	3
7	2923	6
8	4985	76
9	7574	220
10	9117	124
11	10102	255
12	10395	429
13	10095	528
14	10584	597
15	10513	581
16	9823	621
17	9351	202
18	8979	4816
19	8302	6550
20	7758	36
21	7019	6
22	7342	22
23	7786	21

دیتا فریم سوال 5)

	Browser family	OS family	Is_bot	Is_pc
0	IE	Windows	False	True
1	IE	Windows	False	True
2	Chrome	Windows	False	True
3	IE	Windows	False	True
4	Chrome	Windows	False	True
...	...	...	...	...
10364836	Chrome Mobile	Android	False	False
10364840	Chrome Mobile	Android	False	False
10364842	Chrome Mobile	Android	False	False
10364845	Chrome Mobile	Android	False	False
10364857	Chrome Mobile	Android	False	False

سوال ششم)



قسمت دوم سوال ششم) با نگاه کردن به لیست ریکوئست ها، فهمیدیم بیشترین زیر رشته ای که تکرار شده کلمه filter بوده است.

سوال هفتم) (مرجه فایل pdf ای که در نوتبوک قرار داشت.)

1- Click rate: از روی تعداد کلیک می‌توان تشخیص داد . چون کرالر در یک سشن می‌تواند نسبت به یک انسان تعداد بسیار زیادی ریکوئست بفرستد(همان کلیک کردن).
2- HTML-to-Image Ratio: این نسبت(نسبت درخواست برای صفحه html به درخواست فایل های تصویری) برای یک انسان بیشتر است نسبت به کرالر. چون در حالت عادی، تصاویر اهمیت چندانی برای کرالر ندارد. و در یک سشن این نسبت برای انسان بیشتر است.
3- Percentage of PDF/PS file requests: این عدد نسبت درخواست برای فایل های PDF/PS در یک سشن است که در برعکس ویژگی قبلی این عدد برای کرالر بیشتر از انسان است.
4- Percentage of 4xx error responses: به دلیل این که کرالر ها می‌توانند تعداد درخواست های بیشتری برای صفحه های دیلیت شده بدهند و ارور بیشتری دریافت کنند، این نسبت هم برای کرالر ها بیشتر است.
5- Percentage of HTTP requests of type HEAD: کرالر ها برای اینکه حجم کمتری از داده را در درخواست هایشان داشته باشن از درخواست هایی استفاده می‌کنند که از نوع HEAD است. ولی انسان ها از نوع GET استفاده می‌کنند. پس اگر تعداد درخواست های HEAD بیشتر باشد احتمالا کرالر است.
6- Percentage of requests with unassigned referrers: اکثر اوقات کرالر ها درحواست هایی دارند که در آن مرجع نامشخص است. پس اگر این مقدار بالا باشد احتمالا کرالر است.
7- Robot.txt' file request: مقداری وجود دارد که 0 و 1 است به این معنی که اگر درخواستی برای فایل Robot.txt داده شد، آن مقدار 1 است. هر وبسایتی آدرسی دارد که در کنار آن یک آدرسی هم برای این فایل است که توسط دولوپر همان سایت درست شده است. اصولا انسان درخواستی برای دیدن این فایل نمی‌دهند و اکثرا از سوی ربات ها این درخواست رخ می‌دهد. پس یک را تشخیص هم این است.
8- Standard deviation of requested page's depth: مقداری وجود دارد به نام انحراف عمق درخواست که اگر زیاد باشذ به احتمال زیاد از سوی کرالر است. برای مثال
'/cshome/courses/index.html'، انحراف عمق 3 دارد و '/cshome/index.html'، انحراف عمق 2.
9- Percentage of consecutive repeated HTTP requests: این ویژگی تعداد درخواست تکراری برای یک دایرکتوری وبسایت را نشان می‌دهد که اگر هرچقد مقدار بیشتر داشته باشد احتمالا از سوی کرالر است.
10- اصولا انسان دایرکتوری هایی که انتخاب می-کنند پشت سر هم است. و یکهو به دایرکتوری ای نمی‌رود که آدرسش با آدرس فعلی خیلی فرق بکند ولی کرالر ها امکان دارد به صورت تصادفی این همل را انجام دهند.

سوال هشتم) با توجه به تعاریفی که در سوال قبل شد و نتایجی که در این سوال دیدیم، می‌توانیم نتیجه بگیریم که ربات هستند.