

```
#!/usr/bin/env python
# coding: utf-8
```

```
# In[58]:
```

```
from collections import defaultdict
```

```
# In[63]:
```

```
i = 0
dic = defaultdict(list)

with open('./full_fleet.txt') as f:
    k = f.readline()

    while k:
        i += 1
        href, timestamp, width, height, _, cookies_enabled, cookie_user_id, font1,
font2, font3, font4, font5, ip = k.split(',')[13]
        href = href.split('/')[1].split('.')[0]
        useragent = k.split(',')[13:]
        useragent = ' '.join(useragent)
        useragent = "".join(e for e in useragent if e.isalnum())

        # cluster based on all the properties using dictionary
        dic[(ip, width, height, cookies_enabled, cookie_user_id, useragent, font1,
font2, font3, font4, font5)].append(int(href))

        k = f.readline()

f.close()
print(i)
```

```
# In[64]:
```

```
ret = []
for x in dic.values():
    ret.append(sorted(x))

sorted(ret)
```

```
# In[ ]:
```