Requirement

Using the tutorial dataset, write queries that: 1. Generate a list of usernames that were used in failed SSH login attempts from 27.0.0.0/8 2. Generate a list of web pages that have been indexed by Google (“Googlebot”) 3. Generate a table consisting of [Timestamp, Client IP, URL] where the HTTP Server’s response was a 50X error code.

Answers for requirement 1

Below command generates a list of usernames that were used in failed SSH login attempts from 27.0.0.0/8

```
ssh* error OR *fail* OR severe |rex field=_raw "(?<val_ignore_1>.*)\ for
(?<raw_user>.*)\ from (?<IP>.*)\ port (?<val_ignore_2>.*)" | rex field=IP
"27\.0\.0\.(?<range>\d{1,3})" | where range >=0 AND range<=8 | eval
User=replace(raw_user, "invalid user ", "") | table User IP
```

Explanation

```
ssh* error OR *fail* OR severe
```

Filters the dataset for ssh errors

```
rex field=_raw "(?<val_ignore_1>.*)\ for (?<raw_user>.*)\ from (?<IP>.*)\ port
(?<val_ignore_2>.*)"
```

Using regex on `_raw` field this will scarp out the raw_user and IP fields

```
rex field=IP "27\.0\.0\.(?<range>\d{1,3})" | where range >=0 AND range<=8
```

CIDR match for IP 27.0.0.0/8

```
eval User=replace(raw_user, "invalid user ", "")
```

Data massaging

```
table User IP
```

Prints table with User and IP columns

Answers for requirement 2

Below is the command that returns the list of web pages that have been indexed by Google (“Googlebot”).

```
sourcetype=access_* useragent=*google* AND useragent=*Bot* | dedup file | table
uri_path file | rename file as File | rename uri_path AS Webpages
```

Explanation

Google indexes webpages using thier spiders a.k.a google bots. These bots crawl through a webpage gets all the children webpage and index all of them. User agent field in the http server access log carries information about the incoming requests user agent. Splunk is intelligent to parse that access log and map the data to appropriate fields. So when a google bot request a webpage, the web server which renders that page to bot will log the bots user agent More information about the user agents for google bot are present here

```
sourcetype=access_* useragent=*google* AND useragent=*Bot*
```

Search for record whose useragent field containing strings "google" and "Bot".

```
dedup file
```

Gets unique file names. `uri_domain` field which is supposed to carry the domain name is empty and so I dint print it

```
table uri_path file | rename file as File | rename uri_path AS Webpages
```

Prints table with File and Webpages column

Answers for requirement 3

Below command generates a table consisting of [Timestamp, Client IP, URL] where the HTTP Server’s response was a 50X error code.

```
sourcetype=access* status=50*| table req_time clientip uri_domain uri | rename
req_time as Timestamp, uri AS URI, uri_domain AS Domain, clientip AS "Client IP"
```

Explanation

```
sourcetype=access* status=50*
```

Searches for 50X error in status field

```
table req_time clientip uri_domain uri | rename req_time as Timestamp, uri AS URI,
uri_domain AS Domain, clientip AS "Client IP"
```

Prints table with Timestamp URI Domain Client IP fields