# **Data Analysis in Splunk**

Reference: <https://asecuritysite.com/cyberdata/ch13_1>

1. First, open up the Buttercup games Splunk site and search for:

post status=200 action=purchase

| top categoryId

Q. Using the Buttercup games dataset, answer the following:

|  |  |
| --- | --- |
| **Question** | **Answer** |
| 1. Modify the filter so that it displays the top productID? Which is the top Product ID and how many times was it included in a successful purchase action? |  |
| 2. Modify the filter so that it displays the top clientip? Which is the top Client IP and how many times was it included in a successful purchase action? |  |
| 3. Modify the filter so that it displays the top refer? Which is the top refer and how many times was it included in a successful purchase action? |  |

2. SPL has a number of functions we can use. Examples of aggregation functions are: avg(X); count(X); dc(X); max(X);mean(X); median(X); min(X); mode(X); range(X); stdev(X); sum(X); sumsq(X); var(X). Let's use the dc() function, and which is distinct count:

sourcetype=access\_\*

| stats dc(status),dc(productId),dc(categoryId)

Q. Using the Buttercup games dataset, answer the following:

|  |  |
| --- | --- |
| **Question** | **Answer** |
| 1. Modify the filter so that it displays the count of the number of distinct status code values. What the number of distinct values?  |  |
| 2. Modify the filter so that it displays the count of the number of status code values. What the count, and why does it differ from the number of distinct values? |  |
| 3. Modify the filter so that it displays the maximum value of the status codes. What the maximum value? |  |
| 4. Modify the filter so that it displays the minimum value of the status codes. What the minimum value? |  |
| 5. Modify the filter so that it displays the range of status codes. What the range? |  |

3. Now give the columns of our data a name for this we use the "AS" modifier:

sourcetype=access\_\*

| stats dc(status) as Status,dc(productId) as "Product ID",dc(categoryId) as "Category ID"

Q. Using the Buttercup games dataset, answer the following:

|  |  |
| --- | --- |
| **Question** | **Answer** |
| 1. Modify the filter so that it displays columns with the names "Buttercup Status", "Buttercup Product IDs" and "Buttercup Category IDs". |  |
| 2. Now modify the filter so that is provides a table for the number of distinct client IPs (clientIP), status codes (status), product IDs (productId), category IDs (categoryId) and referers (refer). |  |

4. Now let's list for an action of "purchase" and then dc() for the client IP address: [here]

sourcetype=access\_\* action=purchase

| stats dc(clientip) BY categoryId

Q. Using the Buttercup games dataset, answer the following:

|  |  |
| --- | --- |
| **Question** | **Answer** |
| 1. Modify the filter so that it displays the number of distinct IP addresses for productID. Which product ID has the most distinct IP addresses for a product, and which product is it? |  |
| 2. Which useragent has the highest number of distinct IP addresses? |  |

5. We can now search for a given status code using the eval() function, and then count the return values: [here]

sourcetype=access\_\*

| stats count(eval(status="404")) AS count\_status BY sourcetype

Q. Using the Buttercup games dataset, answer the following:

|  |  |
| --- | --- |
| **Question** | **Answer** |
| 1. Modify the filter so that it displays a table of the count for the status codes of 200, 400, 403, 404, 408, 500, and 503. What are the count on these? |  |