

### Result Modification - Lab Exercises Guide

#### Overview

Welcome to the Splunk Education lab environment. These lab exercises will test your knowledge of various commands to manipulate results and normalize data.

#### Scenario

You will use data from the international video game company, Buttercup Games. A list of source types is provided below.

**NOTE:** This is a lab environment driven by data generators with obvious limitations. This is not a production environment. Screenshots approximate what you should see, not the **exact** output.

Index	Type	Sourcetype	Interesting Fields
web	Online sales	access_combined	action, bytes, categoryId, clientip, itemId, JSESSIONID, price, productId, product_name, referer, referer_domain, sale_price, status, user, useragent
security	Active Directory	winauthentication_security	LogName, SourceName, EventCode, EventType, User
	Badge reader	history_access	Address_Description, Department, Device, Email, Event_Description, First_Name, last_Name, Rfid, Username
sales	Business Intelligence server	sales_entries	AcctCode, CustomerID, TransactionID
	Retail sales	vendor_sales	<pre>categoryId, product_name, productId, sale_price, Vendor, VendorCity, VendorCountry, VendorID, VendorStateProvince</pre>
network	Email security data	cisco_esa	dcid, icid, mailfrom, mailto, mid
	Web security appliance data	cisco_wsa_squid	action, cs_method, cs_mime_type, cs_url, cs_username, sc_bytes, sc_http_status, sc_result_code, severity, src_ip, status, url, usage, x_mcafee_virus_name, x_wbrs_score, x_webcat_code_abbr
	Firewall data	cisco_firewall	<pre>bcg_ip, dept, Duration, fname, IP, lname, location, rfid, splunk_role, splunk_server, Username</pre>



#### **Common Commands and Functions**

These commands and statistical functions are commonly used in searches but may not have been explicitly discussed in the course. Please use this table for quick reference. Click on the hyperlinked SPL (Search Processing Language) to be taken to the Search Manual for that command or function.

SPL	Type	Description	Example			
<u>sort</u>	command	Sorts results in descending or ascending order by a specified field. Can limit results to a specific number.	Sort the first 100 src_ip values in descending order    sort 100 -src_ip			
<u>where</u>	command	Filters search results using eval-expressions.	Return events with a count value greater than 30    where count > 30			
<u>rename</u>	command	Renames one or more fields.	Rename SESSIONID to 'The session ID'  rename SESSIONID as "The session ID"			
<u>fields</u>	command	Keeps (+) or removes (-) fields from search results.	Remove the host field from the results    fields - host			
<u>stats</u>	command Calculates aggregate statistics over the results set.		Calculate the total sales, i.e. the sum of price values.    stats sum(price)			
<u>eval</u>	command	Calculates an expression and puts the resulting value into a new or existing field.	Concatenate first_name and Last_name values with a space to create a field called "full_name"    eval full_name=first_name." ".last_name			
<u>table</u>	command	Returns a table.	Output vendorCountry, vendor, and sales values to a table    table vendorCountry, vendor, sales			
sum()	statistical function	Returns the sum of the values of a field. Can be used with stats, timechart, and chart commands.	Calculate the sum of the bytes field    stats sum(bytes)			
<pre>count or count()</pre>	statistical function	Returns the number of occurrences of all events or a specific field. Can be used with stats, timechart, and chart commands.	Count all events as "events" and count all events that contain a value for action as "action"    stats count as events, count(action) as action			

Refer to the <u>Search Reference Manual</u> for a full list of commands and functions.



### Lab Exercise 1 – Manipulating Output

#### Description

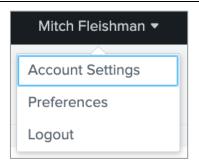
Configure the lab environment user account. Then, use the **xyseries untable**, and **bin** commands to manipulate output.

#### Steps

#### Task 1: Log into Splunk and change the account name and time zone.

Set up your lab environment to fit your time zone. This also allows the instructor to track your progress and assist you if necessary.

- 1. Log into your Splunk lab environment using the username and password provided to you.
- You may see a pop-up window welcoming you to the lab environment.
   You can click Continue to Tour but this is not required. Click Skip to dismiss the window.
- 3. Click on the username you logged in with (at the top of the screen) and then choose **Account Settings** from the drop-down menu.
- 4. In the **Full name** box, enter your first and last name.
- Click Save.
- 6. Reload your browser to reflect the recent changes to the interface. (This area of the web interface will be referred to as *user name*.)



After you complete step 6, you will see your name in the web interface.

**NOTE**: Sometimes there can be delays in executing an action like saving in the UI or returning results of a search. If you are experiencing a delay, please allow the UI a few minutes to execute your action.

- 7. Navigate to *user name* > Preferences.
- 8. Choose your local time zone from the **Time zone** drop-down menu.
- 9. Click Apply.
- 10. (Optional) Navigate to *user name* > Preferences > SPL Editor > Search auto-format and click on the toggle to activate auto-formatting. Then click Apply. When the pipe character is used in search, the SPL Editor automatically begins the pipe on a new line.



Search auto-format enabled

11. Navigate to the **Search & Reporting** app in in the **Apps** dropdown.



Scenario: Networking needs to know the percentage of HTTP server errors that occurred on the e-commerce servers over the previous week.

#### Task 2: Use xyseries to create a chart-like search.

12. This search outputs the percentage of HTTP server errors that occurred on the e-commerce servers (index=web sourcetype=access\_combined) grouped into 1-day time spans. Run this search over the Previous week.

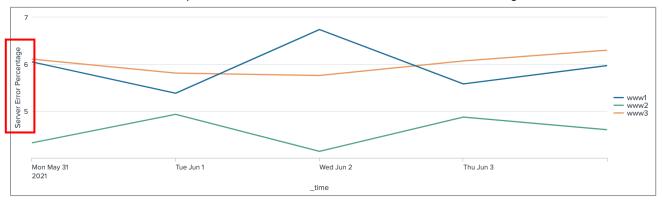
```
index=web sourcetype=access_combined
| bin _time span=1d
| stats count as total, count(eval(status>=500 AND status<=600)) as errors
  by host, _time
| eval percent = round((errors/total)*100,2)</pre>
```

host \$	1	_time \$	total 🗢 🥒	errors 🗢 🧪	percent 🗢 🖊
www1		2021-05-31	2843	172	6.05
www1		2021-06-01	2827	152	5.38
www1		2021-06-02	2895	195	6.74
www1		2021-06-03	3387	189	5.58
www1		2021-06-04	2678	160	5.97
www2		2021-05-31	2616	113	4.32
www2		2021-06-01	2657	131	4.93

13. Use the **xyseries** command to manipulate the **\_time**, **host**, and **percent** fields into chart-like output. The resulting table should be formatted like the screenshot:

_time \$	www1	www2 🕏 🖋	www3 🌲 🖋
2021-05-31	6.05	4.32	6.11
2021-06-01	5.38	4.93	5.81
2021-06-02	6.74	4.14	5.76
2021-06-03	5.58	4.87	6.07
2021-06-04	5.97	4.60	6.30

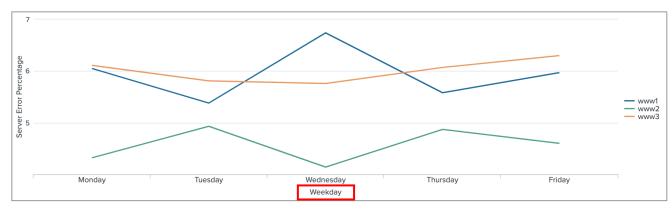
14. Navigate to the **Visualization** tab and visualize results as a **Line chart**. Click **Format > Y-axis** and choose **Custom** from the **Title** drop-down menu. Label the **Y-axis** "Server Error Percentage".



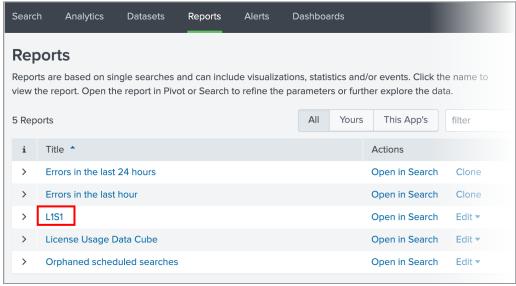


**NOTE**: Step 15 is optional and requires knowledge of **rename** and **eval** commands. You can skip these steps and follow step 16 to save your search as a report.

15. Rename \_time to "Weekday" and then format values to display the full day of the week. (Hint: The format variable for a full weekday is %A.)



- 16. Save your search as a report with the name **L1S1**.
  - a. Click Save As > Report
  - b. For **Title**, enter L1S1.
  - c. Save.
  - d. You can **View** your report or exit out of the **Your Report Has Been Created** window by clicking the **X** in the upper-right corner.
  - e. You can access your saved reports using the **Reports** tab in the application bar.
  - f. Re-initialize the search window by clicking **Search** in the application bar.



Your recently saved L1S1 report will be visible in the Reports tab.



Scenario: IT would like to know the average bytes consumption of TCP\_MISS requests from the websecurity appliance data per website usage type over the last 24 hours.

#### Task 3: Convert chart-like output into a stats-like report.

17. This search finds events where an object had to be downloaded from an origin server (action=TCP\_MISS) and calculates the average bytes consumed during this process for each website usage type. Run this search over the Last 24 hours. Are the values of averages sorted in descending order?

```
index=network sourcetype=cisco_wsa_squid action=TCP_MISS
| chart avg(sc_bytes) as averages over action by usage
| sort -averages
```

18. Modify the search so that the values of averages can be sorted. Then sort averages in descending order.

action \$	/	usage \$	1	averages \$
TCP_MISS		Personal		25747.24
TCP_MISS		Business		20156.25
TCP_MISS		Borderline		18246.2
TCP_MISS		Unknown		320

NOTE: Step 19 is optional. You can skip this step and continue to step 20.

19. Round the values of averages and list values by adding the following eval and stats commands to your search.

```
| eval averages = round(averages,0)
| stats list(usage) as usage, list(averages) as averages by action
```

action \$	/	usage \$	/	averages 🕏 🖊
TCP_MISS		Personal		25747
		Business		20156
		Borderline		14854
		Unknown		320

20. Save your search as a report with the name **L1S2**.

Scenario: Sales wants to know the 5-worst selling products last week and, of those products, the 3 most active customers by IP address.

#### Task 4: Use untable and xyseries to complete the search scenario.

21. This first objective of this scenario (find the 5 worst-selling products on the e-commerce servers by IP address) has been completed for you. Run this search over the **Previous week**.

```
index=web sourcetype=access_combined action=purchase status=200
| chart sum(price) by product_name, clientip limit=0
| addtotals
| sort 5 Total
| fields - Total
```



22. To find the top 3 most-active buyers of these products, you'll need to reorganize the results so that the clientip and product\_name values are flipped. This allows you to calculate statistics about each customer's spending and find the top 3 most active clientips. Begin by converting your table into stats-like output with columns of data in this order: product\_name, clientip, sum.



23. Transform your **stats-**like output into chart-like output. Each row of data should represent the sales events associated with one **clientip**. The resulting table should be formatted like the screenshot:



24. Now, the data is structured so you can get a total sum for each **clientip**. Repeat lines 3 – 5 but modify the **sort** command so that it finds the top 3 **clientips**.



25. Save your search as a report with the name L1S3.



# Lab Exercise 2 – Modifying Results Sets

#### Description

Modify result sets by adding data with the **appendpipe** command and calculating statistics with the **eventstats** and **streamstats** commands.

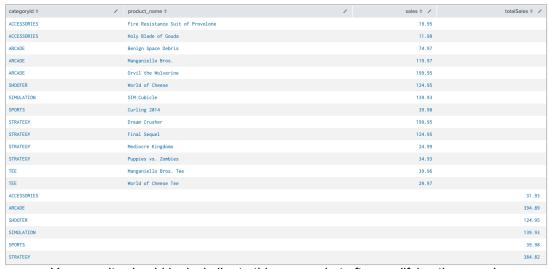
#### Steps

Scenario: Sales Ops wants a table showing retail sales over the last 24 hours by category and product name with total sales for each category.

#### Task 1: Complete this search with the appendpipe command.

1. Modify this search so that the results of the first stats command are preserved and the totalSales for each categoryId is listed at the end of the search. Run the search over the Last 24 hours.

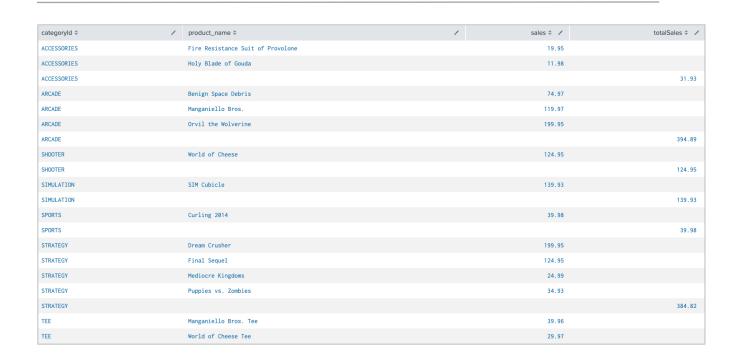
index=sales sourcetype=vendor\_sales
| stats sum(price) as sales by categoryId, product\_name
| stats sum(sales) as totalSales by categoryId



Your results should look similar to this screenshot after modifying the search.

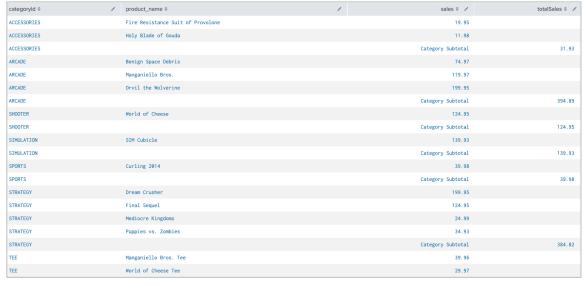
2. Sort results in ascending order by **categoryId** so that **totalSales** values appear as the last event for each **categoryId**.

# splunk>



**NOTE**: Step 3 is optional and requires knowledge of the **eval** command. You can skip this step and follow step 4 to save your search as a report.

3. Modify your results by adding a description in the sales column that displays, "Category Subtotal".



4. Save your search as a report with the name **L2S1**.



Scenario: The retail sales manager wants to identify the retail products with lower-than-average sales across all products for the previous week.

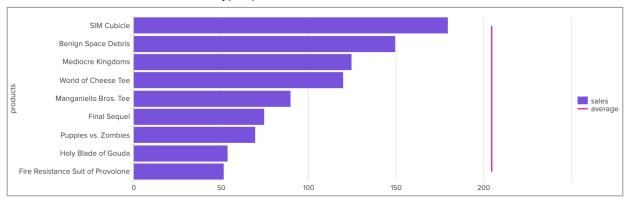
#### Task 2: Fill in the missing portions of a search.

- 5. Complete the <missing> portions of the search so that:
  - a. The **stats** command calculates the total sales of each product using the **price** field.
  - b. The eventstats command finds the average total sales across all products.
  - c. The where command filters search results for those events where the average total sales of all products was greater than the total sale of a specific product (Hint: You will compare the values of two fields.)
  - d. The **sort** command sorts results in descending order by **sales**.
  - e. Run the search over the **Previous week**.

```
index=sales sourcetype = vendor_sales
| stats <missing> as sales by product_name
| eventstats <missing> as average
| where <missing>
| sort -sales
```

Benign Space Debris 4448.22 6967.80142857142 Curling 2014 4417.79 6967.80142857142 Manganiello Bros. Tee 3746.25 6967.80142857142 World of Cheese Tee 3196.80 6967.80142857142 Puppies vs. Zombies 2105.78 6967.80142857142 Holy Blade of Gouda 2000.66 6967.80142857142				
Curling 2014       4417.79       6967.80142857142         Manganiello Bros. Tee       3746.25       6967.80142857142         World of Cheese Tee       3196.80       6967.80142857142         Puppies vs. Zombies       2105.78       6967.80142857142         Holy Blade of Gouda       2000.66       6967.80142857142	product_name \$	1	sales 🗢 🖊	average 🗢 🖊
Manganiello Bros. Tee 3746.25 6967.80142857142 World of Cheese Tee 3196.80 6967.80142857142 Puppies vs. Zombies 2105.78 6967.80142857142 Holy Blade of Gouda 2000.66 6967.80142857142	Benign Space Debris		4448.22	6967.801428571429
World of Cheese Tee 3196.80 6967.80142857142  Puppies vs. Zombies 2105.78 6967.80142857142  Holy Blade of Gouda 2000.66 6967.80142857142	Curling 2014		4417.79	6967.801428571429
Puppies vs. Zombies 2105.78 6967.80142857142 Holy Blade of Gouda 2000.66 6967.80142857142	Manganiello Bros. Tee		3746.25	6967.801428571429
Holy Blade of Gouda 2000.66 6967.80142857142	World of Cheese Tee		3196.80	6967.801428571429
•	Puppies vs. Zombies		2105.78	6967.801428571429
Fire Resistance Suit of Provolone 1584.03 6967.80142857142	Holy Blade of Gouda		2000.66	6967.801428571429
	Fire Resistance Suit of Provolone		1584.03	6967.801428571429

6. Switch to the Visualization tab and display the results as a Bar Chart. Display average as a chart overlay by clicking Format > Chart Overlay and typing in or choosing the average field for Overlay. Click X-Axis and choose Custom for Title then type "products".



7. Save your search as a report with the name L2S2.



Scenario: HR wants a report of the 3 most active users over the last month who were using the network for non-business activities. The report should include the username, event count by usage, total event count by user, and a ranking of their usage in descending order.

#### Task 3: Use streamstats to rank employees by their network activity.

8. This search finds all non-business events from the web security appliance data. The **stats** command counts how many events are associated with each unique combination of **username** and **usage**. The calculations generated by the **stats** command are listed in the **count** column and then the results are sorted in descending order by the **sort** command. Run this search over the **Last 30 days**.

index=network sourcetype=cisco\_wsa\_squid usage!=Business
| stats count by username, usage
| sort -count

username \$	1	usage \$	1	count 🕏 🥒
rerde		Personal		62
pleuchs		Personal		42
tzielinski		Personal		41
hsham		Personal		39
myavatkar		Personal		33
hsham		Unknown		26

9. Use the streamstats command to create a field called "rank" that tallies each time a username value is encountered. Re-execute your search. You should notice how the rank values change as you scroll through the results. For example, in the screenshot below you will see that acurry is assigned two rank values. The streamstats command assigned a rank value of 1 the first time it encountered acurry. When the streamstats command encountered acurry a second time, it assigned a rank value of 2. (Scroll through results to see the value for rank change for different values of username.)

username \$	6	usage \$	1	count 🗢 🖌	rank \$	1
mkemmerer		Personal		145		1
gfacello		Personal		114		1
acurry		Unknown		102		1
tzielinski		Personal		94		1
cganttchart		Personal		93		1
acurry		Personal		91		2
blu		Personal		82		1
cberztiss		Personal		80		1

**NOTE**: Step 10 is optional and requires knowledge of the **stats** and **sort** commands. You can skip this step and follow step 11 to save your search as a report.

# splunk>

10. Pipe search results to the **stats** command and list **rank**, **usage**, and **count** values and a "total" calculation using **sum(count)** by **username**. Use the **sort** command on the **total** field to limit your results to the top 3 **usernames**. (Hint: Use the **as** clause to keep column field names consistent.)

username \$	1	rank 🕏 🖋	usage \$	1	count 🕏 🖌	total 🗢 🖌
acurry		1 2 3	Unknown Personal Violation		102 91 2	195
mkemmerer		1 2 3 4	Personal Borderline Unknown Violation		145 1 1 1	148
tzielinski		1 2 3	Personal Unknown Borderline		94 47 1	142

11. Save your search as a report with the name **L2S3**.



## Lab Exercise 3 – Modifying Field Values

#### Description

Use the **foreach** command and **eval** text functions to modify multiple fields at once.

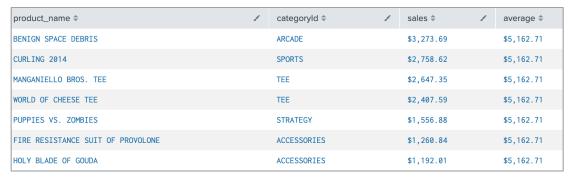
#### Steps

Scenario: The retail sales manager wants you to edit a sales report so that all product names are uppercase and all numerical values match the format \$x,xxx.

#### Task 1: Use the foreach command to format multiple fields at once.

- 1. Edit this search so that:
  - a. All **product\_name** values are uppercase.
  - b. All average and sales values are in the USD currency format \$x,xxx.
  - c. Run the search over the **Previous week**.

index=sales sourcetype=vendor\_sales
| stats sum(price) as sales by product\_name, categoryId
| eventstats avg(sales) as average
| where average > sales
| sort -sales



2. Save your search as a report with the name L3S1.



## Lab Exercise 4 – Normalizing with eval

#### Description

Use eval functions to normalize data.

#### Steps

Scenario: SecOps wants to know the 5 employees who were most active on the network during the previous business week.

Task 1: The employee username data from the network index is not normalized. Use the eval command to normalize fields from two different sourcetypes.

- 1. Complete the <missing> portions of the search so that:
  - a. Values of username and Username are normalized under a new field called "User"
  - b. The stats command counts events by User.
  - c. Results are sorted so that the 5 most active users are listed.
  - d. Run the search over the **Previous business week**.

index=network sourcetype=cisco\_firewall OR sourcetype=cisco\_wsa\_squid)
| <missing>
| stats <missing>
| <missing>



2. Save your search as a report with the name L4S1.

Scenario: Security Operations believes that an individual without legitimate credentials may have gained access to the premises. Therefore, they would like to verify the usernames of everyone who has badged into the building over the last 7 days.

Task 2: Create a new field from two string fields using the substr and lower functions. Then, use the case function to assign "Good Username" and "Bad Username" values.

3. This search finds all badge reader events from the **security** index. Then, the **where** command limits results to events where the **Last\_Name** field is not null and the **dedup** command removes duplicate **Username** values in the events. Use the **table** command to output a table with **First\_Name**, **Last\_Name**, and **Username** values. Run the search over the **Last 7 days**.

index=security sourcetype=history\_access

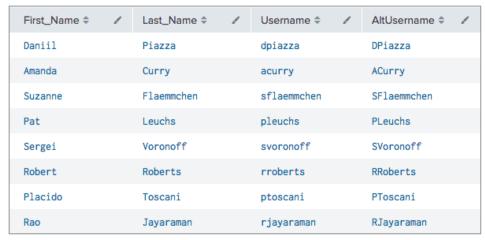


| where isnotnull(Last\_Name) | dedup Username



**NOTE**: Ignore any search reference cycle error messages that may appear when this search is run. The correct data will still be displayed.

4. Create a new field called "AltUsername". Use the **substr** function to create this field. The resulting values will contain the first letter of **First\_Name** concatenated with the full value of **Last\_Name**.



5. Convert AltUsername values to lowercase. (Note: This can be done in the same pipe.)



6. Use the eval command to create a new field called "CheckUsername" and move the table command:

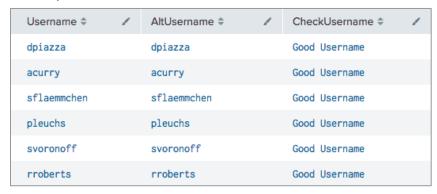


a. Pipe results to the eval command and the case function to create a field called **CheckUsername** that will check the values of **AltUsername** against the values of **Username**:

```
| eval CheckUsername = case(AltUsername=Username, "Good Username",
   AltUsername!=Username, "Bad Username")
```

If the values are the same, a value of "Good Username" is assigned, if the values are not the same, a value of "Bad Username" is assigned.

b. Move and revise the **table** command so that the resulting table contains **Username**, **AltUsername**, and **CheckUsername** in this order.



7. Sort the table so that any events with a bad username appear at the top.



8. Save your search as a report with the name **L4S2**.



**NOTE**: The following challenge exercise is optional and requires knowledge of the **stats**, **eval**, and **fields** commands.

CHALLENGE: The retail sales manager wants to know the number of sales, the average sale price, and the total sales for each category over the previous business week in retail stores.

9. Run the search below over the **Previous business week**. Notice that **appendpipe** is not generating results. Why is this happening? (Hint: Use the **typeOf** function to see how Splunk is interpreting the data. For information on how to use this function, see the Search Reference Manual.)

```
index=sales sourcetype=vendor_sales
| stats count(price) as NumberofSales, avg(price) as AverageSales, sum(price) as
TotalSales by categoryId
| eval AverageSales = "$".tostring(AverageSales,"commas"),
TotalSales="$".tostring(TotalSales,"commas")
| appendpipe
[stats sum(TotalSales) as Total
| eval TotalSales = "$".tostring(Total,"commas")
| eval AverageSales = "Grand Total"
| fields - Total]
```

With the information gained from the previous step, rewrite the search to fulfill the scenario request.

categoryld \$	1	NumberofSales	AverageSales \$	1	TotalSales \$
ACCESSORIES		758	\$4.76		\$3,608.42
ARCADE		722	\$36.67		\$26,472.78
SHOOTER		455	\$24.99		\$11,370.45
SIMULATION		449	\$19.99		\$8,975.51
SPORTS		246	\$19.99		\$4,917.54
STRATEGY		1550	\$23.34		\$36,184.50
TEE		659	\$9.99		\$6,583.41
			Grand Total		\$98,112.61

11. Save your search as a report with the name **L4X**.